Institut für Geodäsie und Geoinformation der Universität Bonn

Professur für Astronomische, Physikalische und Mathematische Geodäsie

# A global drought monitoring framework using GRACE/-FO data assimilation

Dissertation

zur Erlangung des akademischen Grades

Doktorin der Ingenieurwissenschaften (Dr.-Ing.)

der Landwirtschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

von

Helena Gerdener

aus Steinfurt

Bonn 2024

| | |
|---|---|
| Referent: | Prof. Dr. Jürgen Kusche |
| Korreferent: | Prof. Dr. Annette Eicker |
| Korreferent: | Prof. Dr. Gabriëlle De Lannoy |

Tag der mündlichen Prüfung: 18.11.2024
Angefertigt mit Genehmigung der Landwirtschaftlichen Fakultät der Universität Bonn

## A global drought monitoring framework using GRACE/-FO data assimilation

Extreme climate phenomena like droughts can lead to a shortage of available water resources that in turn can cause serious consequences such as famines. Monitoring of extremes is extremely important, and past research mainly focused on fluxes like precipitation or streamflows and surface waters because of easy access to measurements. More complicated is the use of subsurface water storages like groundwater for observing water shortages, mostly monitored via in-situ observations that are spatially irregularly distributed. The stations require maintenance and access can be restricted because, e.g., of political reasons. Hydrological models provide simulations of surface and subsurface water as well but they use strong process simplifications, the forcing data are error-prone, and thus the models imperfectly represent reality.

Another possibility to derive information about water storages is via satellite observations. Since 2002, the gravity missions Gravity Recovery And Climate Experiment (GRACE) and its successor GRACE Follow-On (GRACE-FO) provide measurements of surface and subsurface water globally from space that can be represented as Total Water Storage Anomalies (TWSA). Concretely, TWSA is the change in the aggregated volume of water stored in different compartments, among others, rivers, soil moisture, and groundwater. However, due to the orbit constellation, only a spatial resolution of about 300 km is given, which is too coarse for most drought applications. A possibility to simultaneously improve the spatial resolution, disaggregate TWSA into the single water compartments, and improve the models' realism can be achieved by data assimilation. With data assimilation, the TWSA observations are integrated into a hydrological model and simulated model output is pushed towards the observations.

So far, no global framework exists on how to optimally use water storage information from GRACE/-FO within assimilation for monitoring drought and its water propagation through the hydrological water cycle, i.e. how precipitation deficits lead to a decrease in water storage that in turn affects vegetation growth. Therefore this thesis has three major goals: (1) developing a framework that allows assimilating GRACE/-FO TWSA into the Water Global Assessment and Prognosis (WaterGAP) hydrological model for the first time globally, (2) analyzing dominant hydrological signatures and identifying the signatures of precipitation - water storage - vegetation seasonal maxima and non-seasonal events, and, (3) developing a drought monitoring framework and an unprecedented prototype warning alert system under the consideration of GRACE/-FO and assimilation outputs.

The global GRACE/-FO assimilation into the WaterGAP model is set up with ensemble-based filters, tuned for global application, e.g. to include spatial observation error correlations via localization techniques and provided as Global Land Water Storage (GLWS) release 3.0 – the update of GLWS release 2.0. The analysis of dominant signatures (e.g., linear trends) shows that TWSA from assimilation outputs inherit properties from both, the GRACE/-FO observations and the simulations, thus, present a smooth transition between them. The timing of seasonal or episodic high or low precipitation leading to an increase or decrease in water storage is for most vegetation regimes, as expected, found to be shortest for soil moisture, longer for surface water, and longest for groundwater. Validation with respect to independent data sets shows an overall improvement of GLWS compared to WaterGAP simulations for TWSA and groundwater but the assimilation does not have a major impact on surface water storage and soil moisture as these storages do not significantly improve.

In drought monitoring it is common to define indices that would classify actual drought conditions as "severe", "moderate", etc., to facilitate an easy means of communication to decision makers. Therefore, in this thesis, the performance of existing GRACE drought indices is studied. As it is difficult to evaluate such indices, a synthetic study is set up that reveals how trends, accelerations, and noise in GRACE/-FO time series are biasing drought detection. Hereinafter, droughts are analyzed globally and in selected focus regions for TWSA, and subsequently, I investigate subsequent drought events from soil moisture to surface water and groundwater. An approach is developed for determining drought hazard risk maps from assimilation-derived soil moisture and the combination of surface water and groundwater. Finally, a prototype of a warning alert system is set up and analyzed to identify the spatiotemporal dynamics of droughts for hydrological basins. This system might pave the way for future implementations of TWSA or water storages from GRACE/-FO assimilation into existing or future operational early warning systems.

## Globale Überwachung von Dürren unter Verwendung von GRACE/-FO Datenassimilierung

Extreme Wetterereignisse wie Dürren können die Verfügbarkeit von Wasserressourcen stark einschränken, was zu ernsthaften Konsequenzen wie Hungersnöten führen kann. Daher ist es wichtig, solche Ereignisse kontinuierlich mit Messungen zu überwachen, wobei sich vergangene Untersuchungen zumeist auf den Fluss von Wasser wie zum Beispiel in Niederschlag oder Flussströmen beziehen, da diese einfach zu messen sind. Dagegen ist es schwieriger, Wasserressourcen im Boden oder weit unter der Erdoberfläche zu messen. Grundwasser beispielsweise wird zumeist durch lokale Bohrlöcher oder Brunnen gemessen. Diese haben jedoch den Nachteil, dass sie räumlich unregelmäßig verteilt sind, regelmäßige Wartungen benötigen und die Messungen in manchen Fällen, zum Beispiel aus politischen Gründen, nicht öffentlich zur Verfügung gestellt werden. Eine weitere Quelle aus der Information über Wasserspeicher über und unter der Erdoberfläche gewonnen werden können sind hydrologische Modelle. Da sie aber zum Teil stark vereinfachte Prozessannahmen treffen und durch fehlerbehaftete Eingangsdaten angetrieben werden, können die daraus entstehenden Simulationen die Realität nicht perfekt nachbilden.

Eine weitere Möglichkeit Informationen über Wasserspeicher zu erlangen sind Satellitenbeobachtungen. Seit 2002 erfassen die Satellitenmissionen Gravity Recovery And Climate Experiment (GRACE) und dessen Nachfolger GRACE Follow-On (GRACE-FO) globale Beobachtungen aus dem All von Oberflächengewässern und Gewässern unter der Erdoberfläche als Anomalien des Gesamtwasserspeichers. Der Gesamtwasserspeicher ist die vertikale Summe aus mehreren Wasserspeichern, neben Anderen auch aus Flüssen, Bodenfeuchte und Grundwasser. Aufgrund der Konstellation der Satellitenmissionen, haben GRACE/-FO eine grobe räumliche Auflösung von etwa 300 km, welche für das Überwachen von lokal auftretenden Dürren nicht ausreicht. Eine Möglichkeit um gleichzeitig die räumliche Auflösung zu verbessern, den Gesamtwasserspeicher in einzelne Wasserspeicher zu unterteilen und Modellsimulationen realistischer zu machen, bietet die Datenassimilierung: Mit deren Hilfe können die GRACE/-FO Beobachtungen in ein hydrologisches Modell integriert werden.

Zum jetzigen Zeitpunkt gibt es keine standardisierte Methodik, wie man global Wasserspeicheränderungen aus GRACE/-FO mit Hilfe der Datenassimilierung für die Überwachung von Dürren oder deren Verlauf durch den hydrologischen Wasserkreislauf optimal verwenden sollte, deshalb befasst sich diese Arbeit mit drei Hauptzielen: (1) Assimilierung von Gesamtwasserspeicher aus GRACE/-FO in das Water Global Assessment and Prognosis (WaterGAP) hydrologische Modell erstmals global, (2) Extrahierung von hydrologischen Signaturen aus den Daten sowie Identifizierung von Signaturen zwischen saisonalen Maxima aus Niederschlag, Wasserspeichern und Vegetation und (3) Entwicklung eines Prototypen eines neuen Warnsystems unter Verwendung der GRACE/-FO- und Assimilierungsdaten.

In der Umsetzung werden die GRACE/-FO Daten in das WaterGAP Modell mittels Algorithmen assimiliert, die auf Generierung von Ensembles basieren und für die globale Anwendung optimiert werden, zum Beispiel, indem räumliche Korrelationen zwischen Beobachtungen durch Lokalisierungstechniken einbezogen werden. Die finale Version wird als GLWS Version 3 zur Verfügung gestellt, welche eine Aktualisierung des GLWS 2.0 Datensatzes darstellen wird. Die Analyse von Signaturen (z.B., räumliche lineare Trends) zeigt, dass der Gesamtwasserspeicher aus GLWS Eigenschaften der GRACE/-FO Beobachtungen und Modellsimulationen erbt, und somit einen gleichmäßigen Übergang zwischen beidem darstellt. In den meisten Vegetationsgebieten zeigt sich, dass Niederschlag zunächst die Bodenfeuchte wieder auffüllt, danach Oberflächengewässer und schließlich das Grundwasser. Die Validierung gegen unabhängige Daten verdeutlicht, dass Gesamtwasserspeicher und Grundwasser aus GLWS realistischer ist als aus den GRACE/-FO Beobachtungen und/oder Simulationen. Für Oberflächengewässer und Bodenfeuchte waren etwaige Verbesserungen nicht signifikant.

Um diese Wasserspeicher aus GRACE/-FO und der Assimilierung für Dürremonitoring zu nutzen, werden existierende GRACE Dürreindikatoren herangezogen. Da es keine direkten Validierungsdaten für Dürreindikatoren gibt, wird eine synthetische Studie aufgesetzt. Mit deren Hilfe werden Eigenschaften wie lineare Trends, Beschleunigungen und zeitliches Rauschen aus der GRACE/-FO Zeitreihe ermittelt, die die Identifikation von Dürren beeinflussen. Danach werden Dürren im Gesamtwasserspeicher, Grundwasser, in Oberflächengewässern und in der Bodenfeuchte global und in ausgewählten Regionen analysiert. Es wird ein Konzept vorgestellt, wie man mit Bodenfeuchtedaten und der Kombination aus Grundwasser- und Oberflächengewässerdaten aus der Assimilierung eine Risikoinformation für Dürregefahr herleitet.

Abschließend wird ein Prototyp eines Warnsystems für Dürre erstellt und analysiert, um räumlich-zeitliche Dynamiken von Dürren in hydrologischen Einzugsgebieten zu bestimmen. Dieses System stellt eine Basis dar, wie man zukünftig Gesamtwasserspeicher oder andere Wasserspeicher aus GRACE/-FO und der globalen Assimilierung in bestehende oder noch zu entwickelnde operationelle Frühwarnsystem einbauen kann.

## Acknowledgements

During the years of my PhD journey, I had many great experiences, met new inspiring people, and learned a lot about myself. Now that my thesis has come to an end, I am especially grateful for all the support.

First, I want to thank Jürgen Kusche, who provided me the opportunity to join the Astronomical, Physical, and Mathematical Geodesy (APMG) group at the University of Bonn in October 2017. He gave me the chance to dive into different projects, was always supportive and our discussions and the feedback helped me to sort my thoughts. The topics we discussed and ideas to handle difficult problems highly motivated me.

Moreover, I want to thank Prof. Dr. Annette Eicker and Prof. Dr. Gabriëlle De Lannoy for being referees for my thesis. It was a pleasure to meet you at conferences, meetings, or workshops and to learn from your expertise.

I would like to thank all my colleagues I met in the context of the different projects I worked on. I started my PhD with the GlobeDrought project with colleagues from the University of Göttingen, Goethe University Frankfurt, United Nations University, University of Bonn, Remote Sensing Solutions, and Welthungerhilfe. Then I continued my work in the second phase of the GlobalCDA project with colleagues from Geothe University of Frankfurt, GeoForschungsZentrum Potsdam, HafenCity University Hamburg, Technical University Munich, University of Stuttgart, DLR, University of Luxemburg, and Tu Graz. The most recent project where I contributed to was the ESA Quantum Space Gravimetry project together with colleagues from ESA, TU Munich, HafenCity University Hamburg, TU Delft, University of Trieste and Politecnico di Milano. I am very grateful to be part of such an interdisciplinary research community on national and international levels.

I especially would like to thank Prof. Dr. Petra Döll for sharing her knowledge and asking the right questions that certainly improved our research. She and her group at Goethe University were part of the GlobalCDA and GlobeDrought projects. They did not only provide helpful model simulations and code but we also shared discussions, analyzed complex results together, and had nice conversations apart from research.

Since the thesis would not have been feasible without the assimilation software, I would like to thank Lars Nerger. He provides the software and continuously improves it. I also enjoyed our meetings at conferences or in the context of lectures at the University of Bonn.

I am very thankful for all my past and present colleagues in the office from the APMG, GES, GRV, DETECT, and TG groups. One gets always help when needed, finds an open ear just to mirror own ideas, and lunch breaks are always fun. Thanks to Anne, Kerstin, Bernd, Armin, and Christina for reading parts of my thesis and providing suggestions. It helped a lot to improve my thesis. I also would like to thank Anne and Kerstin for being such nice office mates and for always being open to any of my interruptions during your work.

A different contribution is made by my friends and family. Thank you for always listening to my problems and for having fun together which distracts me from stressful times. I am especially thankful for my parents Elke and Manfred and my brothers Frederic and Valentin, who support me whenever they can. Finally, I want to thank Moritz. The thesis would not have been possible without your endless support, patience, and loving nature. Whenever I forgot to eat while writing, you cooked a meal, when I was down, you motivated me or when I could not concentrate, you brought me back on track. Thank you for always being there for me.

# Contents

**1 Introduction** **1**
1.1 Motivation . . . . . 1
1.2 The hydrological water cycle and droughts . . . . . 2
1.3 Quantification of the water cycle . . . . . 3
1.4 Synthesizing observations and model simulation for drought monitoring . . . . . 6
1.5 Objectives of the thesis . . . . . 7
1.6 Outline of the thesis . . . . . 9

**2 Gravity measurements from GRACE and GRACE Follow-On** **11**
2.1 Overview of GRACE and GRACE Follow-On . . . . . 11
2.2 Potential theory . . . . . 14
2.2.1 Surface spherical harmonics . . . . . 15
2.2.2 Mass changes . . . . . 17
2.2.3 Vertical deformation . . . . . 18
2.3 Processing steps for deriving GRACE/-FO mass changes . . . . . 18
2.3.1 Spectral domain . . . . . 19
2.3.2 Spatial domain . . . . . 21
2.4 Uncertainty estimation . . . . . 26

**3 Monitoring drought** **29**
3.1 Overview and definitions of drought . . . . . 29
3.2 Linking variables of the water cycle . . . . . 30
3.3 Basics of probability theory . . . . . 31
3.4 Drought index computations . . . . . 33
3.4.1 Meteorological and soil moisture drought indices . . . . . 34
3.4.2 Hydrological drought indices . . . . . 35
3.4.3 Drought indices from GRACE/-FO . . . . . 36
3.5 Drought risk and drought hazard risk . . . . . 41
3.6 Drought information systems . . . . . 42

**4 Hydrological modeling** **44**
4.1 Overview . . . . . 44
4.2 WaterGAP . . . . . 45
4.3 Global water use models . . . . . 46
4.4 Water balances within WGHM . . . . . 47
4.5 Water storage limits . . . . . 50
4.6 Reanalysis fluxes . . . . . 51
4.7 Vegetation representation . . . . . 51

**5 Data assimilation** **53**
5.1 Theory of assimilation filters . . . . . 54
5.1.1 Kalman Filter . . . . . 55

5.1.2 Ensemble Kalman Filter . . . 56
5.1.3 Error Subspace Transform Kalman Filter . . . 57
5.2 Implementation . . . 59
5.2.1 Parallel Data Assimilation Framework . . . 59
5.2.2 Simulated TWS versus observed TWSA . . . 60
5.2.3 Spatial and temporal aggregation . . . 61
5.2.4 Parameter and forcing perturbation . . . 64
5.2.5 Setup initial states and mean state vector . . . 65
5.3 Challenges and tuning strategies . . . 66

**6 Analysis of the data assimilation framework 71**
6.1 Methods for extracting and relating signal pattern and delays . . . 71
6.1.1 Statistical metrics . . . 71
6.1.2 Extracting dominant patterns . . . 73
6.1.3 Outlines of hydrological regions and vegetation biomes . . . 76
6.2 Refinements of the assimilation framework . . . 78
6.2.1 Effect of non-standard settings on the assimilation . . . 79
6.2.2 Tuning and error representation of the assimilation . . . 85
6.2.3 Final release choices for GLWS3.0 and comparison to GLWS2.0 . . . 90
6.3 Properties of the global land water storage (GLWS) . . . 91
6.3.1 Dominant signal patterns of hydrological signatures . . . 92
6.3.2 Water cycle dynamics . . . 103
6.4 Validation of the global assimilation framework . . . 110
6.4.1 GNSS . . . 113
6.4.2 Groundwater wells . . . 114
6.4.3 Surface water . . . 118
6.4.4 Soil moisture . . . 118

**7 Drought monitoring with GRACE/-FO and assimilation outputs 121**
7.1 Reliability of GRACE/-FO indices . . . 121
7.1.1 Synthetic study setup . . . 122
7.1.2 Results of the synthetic study . . . 124
7.1.3 Index choice . . . 129
7.2 Monthly monitoring of hydrological drought . . . 130
7.2.1 Global drought monitoring . . . 130
7.2.2 Regional drought monitoring . . . 132
7.3 Drought monitoring system . . . 137
7.3.1 Prototype warning system and risk assessment setup . . . 138
7.3.2 Consecutive periods of drought and drought hazard risk . . . 141
7.3.3 Warning alerts . . . 145

**8 Conclusion and outlook 152**
8.1 Conclusion . . . 152
8.2 Outlook . . . 155

**A Additional results assimilation i**

**B Additional results drought monitoring xi**

**C Glossary xiii**

**D Acronyms xiv**

**List of Figures** **xvi**

**List of Tables** **xix**

# Chapter 1

# Introduction

## 1.1 Motivation

Water is the essential source of life and about 70% of global freshwater is withdrawn for irrigation in agriculture (Gleick, 2014). Thus, the use of water for the agricultural sector is likely the biggest economic sector for water use worldwide and is mainly used for growing crops and livestock farming to ensure enough food on Earth. The water footprint is a measure of how much water volume is used for the production of goods and services in relation to the population. For example, cereal products, meat, and milk products were identified as three of many consumption goods with a very high water footprint from 1996 to 2005 (Mekonnen & Hoekstra, 2011). In addition, large water volumes are also required for other sectors, for example, water supply, powerplant cooling, in households, and for the industry. Water is essential in transport such as river shipping and ensures industrial and consumer product generation.

Extreme events like droughts can strongly limit the availability of water for all systems and have disastrous consequences for the ecology, economy, and human life. They can decrease tourism and water quality or lead to wildfires, hunger, or even death. The economic sectors of water supply and agriculture bear much of the impacts of droughts, and in many countries, the agriculture is the most heavily affected sector[1]. One out of ten people are already suffering from hunger worldwide (United Nations, 2022), which could be increased by droughts.

The natural onset of droughts can occur in every part of the hydrological water cycle ranging from the rather obvious components, for example, a precipitation deficit or a decrease in lake level to less obvious components such as groundwater deficits. Further affected water fluxes or compartments are, for example, streamflow, soil moisture, and evapotranspiration. Most of them are declared as essential climate variables (e.g., World Meteorological Organization, 2021), which highlights the need for monitoring. The first two choices that need to be decided on for the monitoring are which kind of drought one is interested in and which observable should be used, for example, hydrological droughts that are defined by a shortage in water storage observed by using lake levels. Nonetheless, by considering of only one observable of a water compartment for monitoring droughts, a drought event in another water compartment can be missed. This leads to difficulties for drought management (Hisdal *et al.*, 2024) and temporally subsequent droughts in different comparments of the water cycle are disregarded.

In 2020, the Global Climate Observing System (GCOS) steering committee approved adding terrestrial water storage – Total Water Storage (TWS) – to the essential climate variables (World Meteorological Organization *et al.*, 2022). TWS is the vertical sum of all surface and subsurface water storages. As a unique observation type, the gravity satellite missions Gravity Recovery

[1] https://www.fao.org/land-water/water/drought/es/ (last accessed 24.04.2024)

And Climate Experiment (GRACE, Tapley *et al.*, 2004) and GRACE Follow-On (GRACE-FO, Kornfeld *et al.*, 2019) are until today the only missions that were sent to space for measuring anomalies in TWS globally. Observing Total Water Storage Anomalies (TWSA) from space provides the opportunity to measure large mass changes as, for example, the melting of glaciers (e.g., Jacob *et al.*, 2012; Gardner *et al.*, 2013) or anthropogenic groundwater use (e.g., Rodell *et al.*, 2009; Döll *et al.*, 2014). The missions can contribute to overcoming the challenge of missing out on a drought event in a certain water compartment. For example, choosing measurements of lake levels as observable would not capture an exceptional drought event that manifests in the groundwater while GRACE/-FO TWSA measurements would capture the drought independent of whether it manifests in lake levels or in the groundwater.

Nonetheless, observing TWSA with GRACE/-FO also brings a few challenges. The spatial resolution is limited to about 300 km and only anomalies of the TWS can be obtained. Moreover, further details on the temporal evolution of droughts through the single water compartments cannot be gained from the remote-sensed TWS anomalies as the GRACE/-FO observations require further techniques for separation of the TWSA into single water compartments. Hydrological models provide simulations of the water storage that encompass soil moisture, canopy, snow, lakes, wetlands, reservoirs, rivers, and groundwater with finer spatial detail than the GRACE/-FO missions measure TWSA. The models were combined with GRACE/-FO TWSA in the past to improve the models' realism and overcome the spatial limitations of the observations. It is therefore a major goal, to improve our knowledge about TWSA continuously.

In this thesis, my goal is to increase the current understanding of drought monitoring and drought development from a remote sensing perspective using TWSA and show the possibilities of combining GRACE/-FO observations with model simulations via data assimilation on a global scale for drought monitoring and developing a prototype of a global drought information system.

## 1.2 The hydrological water cycle and droughts

From a very general point of view, the natural phenomenon of drought is referred to as a shortage of water when compared to a "normal" for a prolonged period and thus, leading to much drier conditions than a system is used to. Drought is understood as a rare extreme event affecting the economy and ecology worldwide. All climate regimes ranging from arid to humid can experience droughts, also initiated by climate events in remote regions such as the El Niño Southern Oscillation, a reoccurring event in the Pacific Ocean (e.g., Vicente-Serrano *et al.*, 2011). The danger of drought is its slow-onset nature, blurred temporal boundaries, and lack of visible physical damage because drought is not as visible as, for example, floods (Food and Agriculture Organization of the United Nations, 2021a). The danger can therefore be underestimated (e.g., Smith *et al.*, 2024).

As motivated, the sectors of water supply and agriculture highly suffer from drought events but also other sectors such as waterborne transportation, electricity production (hydropower or cooling water), and recreation are affected by drought (e.g., Sheffield & Wood, 2011; Van Vliet *et al.*, 2012; Van Loon, 2015). With about one-quarter of the world's labor force working in the agricultural sector (Roser, 2023), the potential of drought influencing a large part of the population is crucial. For example, drought increases the production losses of the agricultural sector more than any other natural hazard (Food and Agriculture Organization of the United Nations, 2021b).

To study drought, a detailed look into water mass distribution and transport on Earth is required. The terrestrial water cycle describes the transport of water above and below land masses. In a very general view, two main groups of water variables can be viewed: (1) fluxes and (2) storages.

Flux variables are, for example, precipitation, evapotranspiration, runoff, and streamflow, and storage variables are, for example, soil moisture, snow, groundwater, and surface water in lakes, wetlands, and rivers. Most of the variables are connected and the vertical sum of all water storages is the TWS.

Describing water transport in the water cycle starts here with precipitation, which falls on the Earth's surface and refills several storages on the surface, i.e. canopy, reservoirs, lakes, wetlands, and rivers or seeps into the upper part of the soil. Outgoing from there, the water can take very different ways, for example, seeping down to lower soil layers or the groundwater storage, routing through rivers into the direction of the ocean, and many others. To close the loop, at some point in time, the water will evaporate or transpire from the current surface location, e.g. via the soil, plants, and the ocean. When much moisture gathers in clouds, precipitation might fall again. A water shortage – and thus a drought – can now occur in all the different parts of the water cycle and can be seen as an anomaly of the steady state. Furthermore, a shortage of a certain flux or water storage can lead to a shortage of another water compartment. In other words, drought can also propagate through the different water compartments of the water cycle with varying drought duration and intensity per compartment as illustrated in Fig. 1.1.

Classically, to monitor drought, direct access to observations has been possible for many years for fluxes and surface water. However, this was not always the case for subsurface water such as soil moisture and groundwater due to restricted physical access to it. Since groundwater is the major source of liquid freshwater on Earth with approximately 98%, thus a major resource for drinking water (Margat & Van der Gun, 2013), and can also contribute to vegetation growth, especially in arid regions (Glanville *et al.*, 2023), the interest in studying this part of the water cycle increased in the last few decades. The satellite-derived TWSA, which incorporate groundwater – helped to understand, for example, impacts of climate change, cyclic phenomena, and natural disasters on the surface and subsurface water storages.

As a consequence of the high natural and socio-economic impacts of droughts, drought is the disaster with the most damage (Sivakumar *et al.*, 2014) among other disasters such as floods, earthquakes, etc. Studies showed that irrigation with groundwater is expected to increase to ensure water supply (e.g., Gaye & Tindimugaya, 2019; Cobbing, 2020) and, at the same time, the occurrence of droughts is expected to increase (e.g., McGlade *et al.*, 2019; Spinoni *et al.*, 2020; IPCC, 2022). These reasons and the complexity of the water transport in the water cycle underline that a high emphasis should be placed on improving the monitoring of droughts, for example for monitoring subsequent drought events in different water storages.

## 1.3 Quantification of the water cycle

To quantify anomalies of fluxes and storages compared to the expected state within the hydrological water cycle, extreme events like droughts can be monitored, which means collecting data on the current state and analyzing them. In contrast, forecasting droughts refers to developing realistic future scenarios of possible drought events. Collecting data for drought monitoring can be categorized into observations and model simulations. For the observations, the water fluxes or storages can either be observed via local in-situ stations irregularly distributed on Earth, or via remote sensing observation from space. In-situ observations of streamflow, surface waters, and soil moisture are usually easily accessible. Nonetheless, some countries do not gather as many in-situ observations as others. In some cases, information is missing to further process the gathered data, access to the data is even restricted due to political reasons, or data are not digitalized. As a result, deriving many stations with globally and temporally consistent coverage

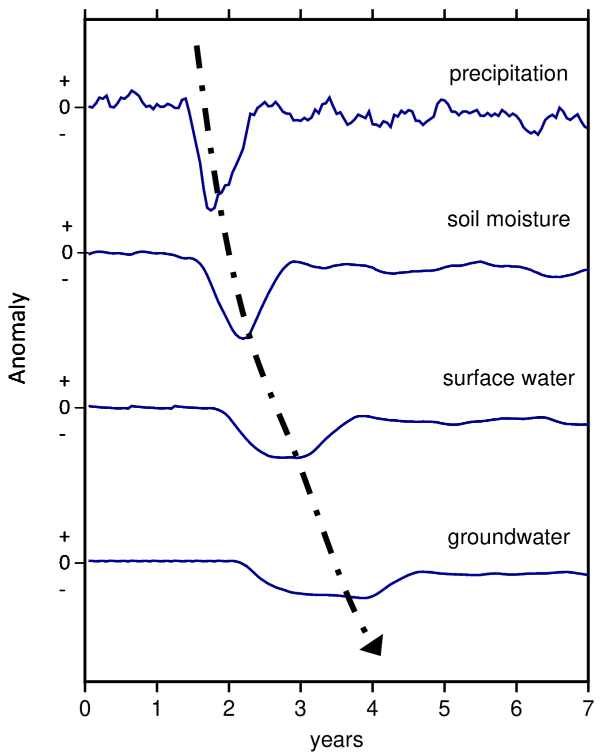


Figure 1.1: Illustration of how drought propagates through the water cycle fluxes and storages of precipitation, soil moisture, surface water, and groundwater. Adaption of Van Loon (2015).

that cover each climate regime can be very challenging.

Accessing in-situ observations of groundwater has even stronger challenges since they are technically more difficult to derive as compared to, for example, gauge stations because the groundwater well is underneath the Earth's surface and needs steady maintenance. In the last years, the Global Groundwater Monitoring Network (GGMN) initiative of the International Groundwater Resources Assessment Centre (IGRAC)[2] started to work on a global scale to gather and provide networks of groundwater station data to be able to observe drought in subsurface water. Although this is pioneering, the process of adding in-situ stations to the network all over the world is still ongoing, which incorporates securing agreements with operators or agencies, implementing automatic data transmission, etc. Many regions worldwide still do not have an optimal spatial coverage of stations in the database and some stations also provide a limited temporal coverage, e.g., observations for only a few years, not suitable for drought detection. Furthermore, transforming well data to storage anomalies requires further information about soil conditions, which are difficult to derive with good quality.

Remote sensing data provide another opportunity to measure water variables. For example, some lake levels have been measured by altimetry missions (e.g., Topex/Poseidon, Jason-1, Jason-2, and Jason-3) since the early nineties, and soil moisture data of the first few millimeter of the soil can be retrieved since the end of the seventies where no canopy is covering the soil via microwave sensors on satellites (e.g., the missions Tropical Rainfall Measuring Mission (TRMM), Soil Moisture Active Passive (SMAP), and Surface Water and Ocean Topography (SWOT). In 2002 the GRACE mission was launched into space, which was a gravity mission that observed global mass changes, including natural water mass changes, and anthropogenic contributions to

[2] https://www.un-igrac.org/ (last accessed 25.04.2024)

the water mass changes. GRACE and its successor GRACE-FO were and are unique missions that can globally capture anomalies of the TWS, and contribute to many essential climate variables directly and indirectly (Tapley *et al.*, 2019). GRACE helped derive great scientific achievements, for example, uncovering groundwater depletion and withdrawal for irrigation to an extent not known before, for example, in India, the Californian Central Valley, and the Middle East (e.g., Famiglietti *et al.*, 2011; Voss *et al.*, 2013; Rodell *et al.*, 2018).

Besides these examples, it was possible to monitor dry periods or years in the TWSA time series from GRACE/-FO all over the world, for example, the European heat wave in 2003 (Andersen *et al.*, 2005), a multi-year dry period in the Murray-Darling basin (Leblanc *et al.*, 2009) or a drought in 2011 in Texas (Long *et al.*, 2013). Abelen *et al.* (2015) could relate dry periods in monthly TWSA from GRACE to El Niño/La Niña in the La Plata basin. In addition, some studies made use of drought index computation to describe the severity, timing, and duration of a drought in the GRACE/-FO TWSA (e.g., Yirdaw *et al.*, 2008; Thomas *et al.*, 2014; Zhao *et al.*, 2017). In late 21st century, the number of drought events determined from TWSA is expected to increase in many regions of the world (Pokhrel *et al.*, 2021). In 2015, Van Loon (2015) noted that using satellite data focusing on hydrological drought monitoring is still limited and emphasized the need to integrate GRACE/-FO. This is still valid today because the number of studies that use GRACE/-FO for drought monitoring is still limited and no consensus exists on how optimal drought monitoring with GRACE drought indices should be set up and how the indices compare against each other.

Although GRACE and GRACE-FO were already successfully used for drought monitoring, four major drawbacks of the data exist that complicate the monitoring:

- Precipitation shortages can lead to a deficit in water storage and the extent of the precipitation can be very local. In many cases, GRACE/-FO is spatially too coarse with its native resolution of about 300 km.

- A one-year gap exists between the end of GRACE and the launch of GRACE-FO. In addition, some months in the data are missing because the saving of battery capacity for the satellites led to only a few observations in these months from which no monthly field could be computed. In consequence, the TWSA data are not completely temporally consistent. Single months can be interpolated but once there is a long gap interpolation methods do not perform well.

- As GRACE senses the vertical sum of water storage – the TWSA – it cannot distinguish between different water compartments like soil moisture, surface water, and groundwater.

- A large number of reviews exist that gather the advantages and disadvantages of meteorological drought indices for monitoring. For GRACE TWSA, only a few drought indices exist but it is unclear how they relate to meteorological drought indices and what signals in the GRACE TWSA might bias drought detection.

Since the hope of having a satellite mission with much higher spatial resolution than GRACE/-FO is low, there is a desire to spatially downscale the TWSA. Simple statistical downscaling approaches incorporate different higher-resolution predictor variables such as precipitation or soil moisture data to estimate spatially downscaled TWSA, e.g., a recent partial least squares regression approach in Vishwakarma *et al.* (2021) or a machine learning approach to downscale and estimate TWSA simultaneously as shown in Li *et al.* (2021). A few studies (e.g., Van Loon *et al.*, 2017; Frappart *et al.*, 2019) used variables of soil moisture or others to separate the groundwater from the TWSA, however, these approaches strongly depend on the input data sets or model simulations. For example, Van Loon *et al.* (2017) found that resulting groundwater uncertainty is largely influenced by the uncertainty of Global Land Data Assimilation System (GLDAS) soil moisture included in the separation.

A further possibility to derive information about water variables is via global hydrological model simulations. The model simulations have a higher spatial and temporal resolution than the GRACE/-FO observations and provide consistent information on the grid for the separate compartments. However, hydrological models rely on error-prone forcing data often affected by biases and do not perfectly represent reality because they are underlying assumptions and model-specific concepts that simplify or do not include all processes. For example, linear trends in hydrological models are underestimated as compared to the GRACE/-FO observations (Scanlon *et al.*, 2018), or inconsistencies exist in model simulations of groundwater due to oversimplifications (Xia *et al.*, 2017).

## 1.4 Synthesizing observations and model simulation for drought monitoring

To overcome the challenges that appear when working with observations and model simulations, the GRACE/-FO TWSA can be integrated into hydrological models via data assimilation. Data assimilation enables improving the models' realism and – at the same time – spatially downscales and vertically disaggregates the observations into single compartments. Thus, the final TWSA from assimilation represents a synthesized data set with fine spatial and consistent temporal resolution, where GRACE/-FO monthly gaps are filled. Moreover, the assimilation outputs other variables such as surface water, soil moisture, and groundwater that are relevant for drought monitoring.

Multiple regional studies integrated GRACE/-FO TWSA into hydrological models for drought monitoring. For example, Li *et al.* (2012) elaborated the potential of assimilating GRACE into land surface models for improving drought monitoring in Western and Central Europe, Schumacher *et al.* (2018) applied GRACE data assimilation to improve drought simulations for the Murray-Darling basin, and Houborg *et al.* (2012) assimilated TWSA in the United States and monitored drought via drought indicators. In contrast, a decade ago, no global data assimilation systems were available at all, but during approximately the last five years, very few groups worldwide implemented global data assimilation systems. Studies by Li *et al.* (2019) and Felsberg *et al.* (2021) globally assimilate TWSA from GRACE/-FO Mass Concentration blocks (mascons) into the Catchment Land Surface Model (CLSM, Koster *et al.*, 2000), which was developed by the National Aeronautics and Space Administration (NASA). Khaki *et al.* (2023) assimilate globally gridded TWSA from mascon solutions or assimilate K-band ranges into the World-Wide Water Resources Assessment system (W3RA) model. One of the most recent works is the study by Forootan *et al.* (2024), who assimilate also globally gridded GRACE/-FO TWSA via a new Markov Chain Monte Carlo-Data Assimilation approach into the W3RA model.

To my knowledge, the existing global assimilation systems do not incorporate observation correlations as measured from GRACE/-FO on a global scale, probably because most assimilation systems have no implementation for the correlations. They either use only the variance information from mascons (Li *et al.*, 2019; Felsberg *et al.*, 2021; Khaki *et al.*, 2023) or derive uncertainties from collocation techniques (Forootan *et al.*, 2024) applied to the gridded TWSA. However, it is important to consider the observations correlations because the specific mission constellations of

GRACE/-FO lead to strong spatial correlations between grid cells in the level 3 TWSA data.

Although multiple studies developed techniques for monitoring drought with GRACE/-FO, to my knowledge, at the time of writing, only two global assessments and one large-scale assessment (United States) of drought monitoring with data assimilation are known to the author of this thesis, i.e. Houborg *et al.* (2012), Li *et al.* (2019) and Forootan *et al.* (2024). Whereas Houborg *et al.* (2012) computed drought percentile for surface water, root zone excess, and groundwater, Forootan *et al.* (2024) and Li *et al.* (2019) only monitor droughts via indices for groundwater. However, what is still missing in the literature is how different drought indices based on GRACE/-FO or outputs from GRACE/-FO assimilation compare or how to outline a roadmap with these data to determine drought risk for various sectors of the economy. Furthermore, no study analyzes the temporal evolution of drought, which means subsequent drought in precipitation and different water compartments and how to include GRACE/-FO in a prototype of a warning alert system to uncover these temporal dependencies (Fig. 1.1).

## 1.5 Objectives of the thesis

The aforementioned challenges and limitations lead to specific research questions about integrating GRACE/-FO TWSA for drought monitoring at the global scale:

- How can GRACE/-FO TWSA optimally be assimilated globally into a hydrological model?
- How do global observations, model simulations, and assimilation outputs compare and how do hydrological signatures as, for example, linear trends compare across the three data types?
- How do seasonal maxima or non-seasonal events in precipitation propagate through the water cycle and thus recharge or drain water storage?
- Which drought index is best suitable for drought monitoring with GRACE/-FO TWSA and assimilation outputs?
- How to set up drought monitoring systems that incorporate GRACE/-FO TWSA or outputs from GRACE/-FO assimilation into hydrological models?

I published a number of studies during the last years that were meant to address these research questions and the described limitations. The updates and new results are now included in this thesis via three major work packages: first, refining a framework for globally assimilating GRACE/-FO TWSA, second, analyzing dominant patterns and signatures in the data, and third, testing the performance of drought monitoring using GRACE/-FO TWSA and assimilation outputs and develop a prototype of a warning alert system.

**1) Develop a global data assimilation framework based on GRACE and GRACE-FO TWSA:** A first version of a global data set that assimilates GRACE/-FO TWSA globally into the global hydrological model WaterGAP was published as the Global Land Water Storage (GLWS) data set release 2 on the data server PANGAEA (Gerdener *et al.*, 2023b). This thesis builds upon the GLWS, reduces shortcomings in previous in-house GRACE/-FO processing in preparation for the assimilation, and further develops the assimilation system to improve the water storage outputs. For example, an earthquake correction for GRACE was computed and uploaded to PANGAEA (Gerdener *et al.*, 2020b), contributed to Deggim *et al.* (2020), where it was shown that earthquakes in the TWSA could potentially bias drought detection and similar techniques will be tested for the input observations for the assimilation. After testing and analyzing the final assimilation system according to its robustness, a global data set of TWSA, surface storage, soil moisture, and groundwater is derived. The code for this data set will be frozen and released as GLWS release 3.

**2) Extract global signatures from the global product GLWS from assimilation:**

In Gerdener *et al.* (2022), the time of seasonal precipitation maxima (e.g., annual precipitation amplitudes) recharging water storage was analyzed at first. Second, we investigated the timing of seasonal moisture in the root zone or seasonal irrigation with groundwater and surface water that facilitates vegetation growth. Third, the same approach was applied to non-seasonal events. Hereinafter, the described propagation of precipitation replenishing water storages and leading to vegetation growth will be denoted in this thesis as precipitation – water storage – vegetation dynamics. The study was set up and tested in South Africa. Since the timing of the precipitation – water storage – vegetation dynamics strongly depends on regional conditions such as climatology, hydrology, topology, etc., this approach will be adapted and extended to a global scale. I aim to extract information on the timing of the precipitation – water storage – vegetation dynamics to learn about lead times of non-seasonal events like droughts before they get monitored in drought warning systems and to possibly use this information in the future to improve the drought monitoring systems. A special focus is here given on the non-seasonal signal of water storage, because events such as droughts are likely not contained in the fitting of seasonal modeling as they appear irregularly.

**3) Develop a global drought framework using GRACE/-FO TWSA and storage information from global assimilation:**

For this, I want to begin with extending the knowledge about detecting droughts by using observation-based TWSA and study the performance of drought indices for GRACE/-FO via a synthetic study by extending Gerdener *et al.* (2020a). After balancing the advantages and disadvantages of the GRACE/-FO drought indices, I will choose one of the indices to analyze drought detection with assimilation versus model simulations and observations and analyze drought in the storage compartments of surface water, soil moisture, and groundwater. Early results from drought monitoring with GLWS release 2 in West Africa were contributed to the review paper Dibi-Anoh *et al.* (2023) but will not again be discussed in this thesis. Instead, the analysis is globally extended and provides further detail for a few hydrological basins. With the help of the drought indices, consecutive drought events can be extracted and are used for two applications: (i) develop a prototype of a global drought monitoring system and (ii) a drought hazard risk analysis by using GRACE/-FO assimilation outputs. Hereby, I will relate to agricultural systems and water supply as sectors for which a drought hazard risk assessment system is computed.

In addition, I would like to mention here that I included many details about GRACE/-FO processing and application, assimilation tuning and application, and drought risk computation contributed in publications that I do not repeat here for the sake of brevity. The studies include the analysis of GRACE/-FO TWSA from multiple products in Germany (Güntner *et al.*, 2023), using GRACE/-FO together with machine learning (Li *et al.*, 2024) the WaterGAP modeling in restart mode (Müller Schmied *et al.*, 2023), drought risk computation in South Africa and globally with drought hazard from modeling (Meza *et al.*, 2020; Meza *et al.*, 2021), analyzing the assimilation of GRACE/-FO TWSA jointly with discharge gauge data (Schulze *et al.*, 2024) or parameter estimation (Döll *et al.*, 2024), and trend estimations with climate models (Jensen *et al.*, 2024).

In summary, this work presents a detailed framework to include GRACE/-FO in drought monitoring. Thus, I provide a contribution to improve the knowledge about the essential climate variable TWSA but also storage compartments and to reach the Sustainable Development Goals (SDG) of the United Nations. In particular, I contribute to two SDGs: SDG 2, which aims to end hunger, achieve food security, improve nutrition, and promote sustainable agriculture, and SDG 6, which aims to ensure availability and sustainable management of water and sanitation for all.

## 1.6 Outline of the thesis

The thesis starts with Chapter 2 introducing the gravity measurements from the satellite missions GRACE and GRACE-FO. I give an overview of the missions with data products, introduce the basics of potential theory used to explain the gravity field in spherical harmonics and introduce the approach to how we obtain surface mass changes from it. Then, the processing steps I implemented in my thesis are described and in addition, I show how to derive uncertainty information for the corresponding mass changes.

Chapter 3 presents an overview of monitoring drought as a natural hazard. I first give an overview of how to define drought in different compartments of the water cycle. Then, the link of precipitation to TWSA is shown to relate precipitation-based drought indices to GRACE/-FO TWSA drought indices. Thereinafter, I explain drought indices and present various indices for meteorological drought, soil moisture drought, hydrological drought, and indices for GRACE/-FO and how they can be used to investigate drought risk. The Chapter finishes with the presentation of existing regional and global drought monitoring systems.

Chapter 4 adds information about hydrological modeling in general and how this is represented in the WaterGAP model. I focus on WaterGAP only since it is a global model that has an explicit representation of surface water storages, soil moisture, and groundwater storage, and it considers anthropogenic water use. The two main components of WaterGAP are introduced, which are the WaterGAP Global Hydrology Model (WGHM) model which accounts for water balances, and the Global Irrigation Model which includes anthropogenic water use in the simulation. Next, I describe which limits for water storages are contained in the model equations because it is also relevant for adapting assimilation strategies, which forcing data are required to start and run the model, and how the vegetation is represented.

In Chapter 5, the details about the data assimilation of observations into hydrological models are presented. The theory of assimilation filter includes basics from probability theory that are required for the filter algorithms. The Kalman filter, the Ensemble Kalman Filter (EnKF), and the Error Subspace Transform Kalman Filter (ESTKF) are described in detail. The concrete implementation of the data assimilation framework for this thesis is elaborated. I explain the Parallel Data Assimilation Framework (PDAF) that is used for efficiently implementing filter algorithms in the model and show spatial and temporal aggregation techniques to match model simulations and observations. Since the filters used in this work are ensemble filters, perturbation of forcing and parameters, as well as the initialization of states is described. The chapter closes by presenting challenges and tuning strategies for assimilation.

The analysis of the data assimilation outputs is presented in Chapter 6, i.e. monthly TWSA, surface water, soil moisture, and groundwater at $0.5°$ grid cells from 2003 to 2019. I explain statistical methods to compare measurements, observations, and assimilation outputs and how to extract signal patterns from the variables of the hydrological water cycle derived from assimilation. The processing choices of the GRACE/-FO observations and the assimilation are reconsidered and refined and a final assimilation production run is then generated, termed GLWS release 3, and compared with release 2. After that, the properties of GLWS are explored in detail according to dominant signal patterns and water cycle dynamics and validated against independent data from Global Navigation Satellite System (GNSS), surface water storage based on remote sensing, and groundwater wells.

Chapter 7 presents an overall framework on how to monitor droughts with water storages from GRACE/-FO and assimilation. I first test the reliability of indices according to different biasing signatures in the data or the indices (e.g. trends, accumulation periods, etc.) via a synthetic study and choose a final index for monitoring with real data. Then, a detailed analysis of TWSA indices and the vertical disaggregation of drought into surface water, soil moisture, and groundwater is presented and uncertainty information is provided. The suggested drought warning system is then set up using fluxes and vegetation information in the indices together with the knowledge from storage drought monitoring. In a retrospective analysis, drought hazard risk and warning alerts are shown and elaborated.

The thesis will close with Chapter 8. In this Chapter the work is summarized, conclusions are drawn and an outlook on what future work could follow is presented.

# Chapter 2

# Gravity measurements from GRACE and GRACE Follow-On

## 2.1 Overview of GRACE and GRACE Follow-On

The satellite gravity mission Gravity Recovery And Climate Experiment (GRACE) was a joint mission developed by National Aeronautics and Space Administration (NASA) and German Aerospace Center (DLR). It was launched in March 2002 and ended in October 2017. It was planned as a 5-year mission but lasted 15 years in total. The mission was designed to measure the global time-variable gravity field, from which information about the mass distribution on Earth and transport within the Earth system can be obtained. This is achieved by continuously measuring the distance between a pair of satellites in a configuration where one satellite is the leading satellite and the second satellite is the following satellite. In this constellation, the two satellites experience time-delayed gravitational acceleration from the masses on Earth (Fig. 2.1). On the 12th of October 2017, it was found that on board of one of the two satellites – GRACE-B – the battery capacity was not sufficient enough to run the instruments for another measurement campaign and to send the data to Earth, which is why the end of the mission was decided[1]. GRACE Follow-On (GRACE-FO), the successor of GRACE, is still in orbit and it has a similar obit configuration. GRACE-FO was launched in May 2018 and was planned with a 5-year lifetime. At the time of writing this planned lifetime is already exceeded by more than one year.

GRACE/-FO provide unique measurements of spatial and temporal mass variations on global to regional scales. This includes mass transports in different Earth systems, for example, the atmosphere, hydrosphere, biosphere, oceans, cryosphere, and solid Earth (e.g., Wahr *et al.*, 1998). With its spatial resolution of approximately 300 km and an uncertainty of Total Water Storage Anomalies (TWSA) fields of about 2 cm (Wahr *et al.*, 2006), GRACE/-FO achieved a scientific breakthrough in the disciplines cryosphere, hydrology, and oceans (Tapley *et al.*, 2019).

A few out of many groundbreaking examples are now given: GRACE was essential for observing total ice mass losses in Greenland and Antarctica (e.g., Velicogna & Wahr, 2005; Chen *et al.*, 2006; Velicogna & Wahr, 2006) and uncovered significant changes for glaciers all over the world (e.g., Jacob *et al.*, 2012; Gardner *et al.*, 2013). The measurements helped detect regions with a decline in freshwater due to the irrigation of agriculture, for example, in India, where groundwater depletion was identified (e.g., Rodell *et al.*, 2009; Döll *et al.*, 2014). Additionally, anomalies in TWS were related to large-scale climate events like the El Niño Southern Oscillation (e.g., Phillips *et al.*, 2012; Ni *et al.*, 2018), and floods (e.g., Long *et al.*, 2014; Tangdamrongsub *et al.*, 2016). Finally, as it is of particular relevance for this thesis, GRACE enabled monitoring droughts all over the world,

[1] https://idw-online.de/de/news683751 (last accessed 15.07.2024)

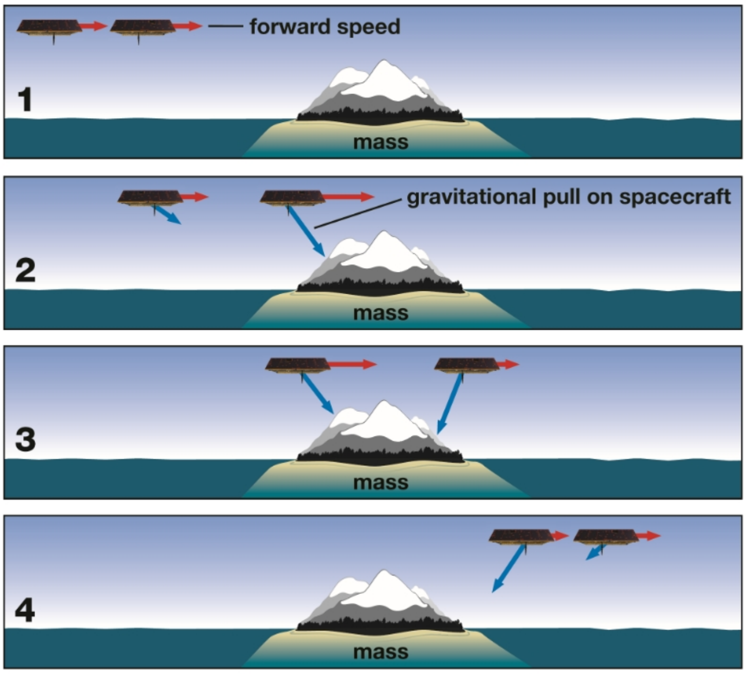


Figure 2.1: Simplified concept of how the GRACE or GRACE-FO satellite pairs react to a mass change. Image credit: NASA.

for example, for the Amazon basin, and Europe (e.g., Andersen *et al.*, 2005; Frappart *et al.*, 2012).

Studies about future missions show the value of different satellite constellations to increase the accuracy of measured gravity fields, for example by using more than one satellite pair (e.g., Wiese *et al.*, 2012; Elsaka *et al.*, 2014). The first time that one of these concepts was agreed upon to be realized is the Mass-Change and Geosciences International Constellation (MAGIC) mission: a joint mission by European Space Agency (ESA) and NASA consisting of two satellite pairs whose realization was already approved[2].

To introduce the GRACE and GRACE-FO missions in more detail, the mission's concept and available data products are presented.

**Mission design**

The GRACE mission consisted of two satellites that flew in the same orbit around the Earth with an initial distance of about 220 km. At launch, the altitude above the Earth was about 485 km, thus the satellite-to-satellite system flew in a low Earth orbit, which steadily decreased during the mission lifetime. The decay in altitude was about 30 meters per day. The mission was designed as a near-circular polar orbit with an inclination of 89.5° and an orbit period of about 90 minutes. In light of this initial orbit, it was decided to collect 30 days of data to cover the whole surface of the Earth while flying in an orbit with non-repeating ground tracks. Some months within the data are missing because of the systematic shutting down of instruments to prevent battery failure (after 2011).

An inter-satellite distance was measured continuously by a K-band microwave ranging instrument. The K-band system was set up with dual-frequency one-way phase measurements – transmitted and received by both satellites (Tapley *et al.*, 2004) – and had an accuracy better than 10 $\mu$m.

[2] https://www.eoportal.org/satellite-missions/magic#eop-quick-facts-section (last accessed 15.03.2024)

Each satellite was further equipped with a SuperSTAR accelerometer, GPS receiver/antenna, star cameras, and laser retro reflectors (Tapley *et al.*, 2004). The accelerometers measured the non-gravitational forces acting on the satellites, whereas the GPS systems provided the satellite positions above the Earth and derived time tags for the measurements. Star cameras oriented the satellites against the stellar background and laser retro reflectors enabled orbit verification and were measured from satellite laser ranging stations on Earth.

The start of its successor GRACE/-FO in June 2018 led to a gap in gravity observations between the two missions of about one year. GRACE-FO is designed in a similar setup as GRACE considering the orbit constellation and the satellite-to-satellite system. Additionally, it is equipped with laser ranging interferometry to derive more precise gravity fields, which is the first time ever of such an instrument in space (Kornfeld *et al.*, 2019). However, due to the failure of an accelerometer onboard one of the two satellites – GRACE-D –, transplantation of acceleration data from the other satellite is required, which limits the desired improvements in uncertainty relative to GRACE (Loomis *et al.*, 2020). Nonetheless, GRACE-FO is still capable of providing gravity fields that are at least as accurate as GRACE because of acceleration transplants.

**Data products**

Official GRACE/-FO data products are provided by the Center for Space Research (CSR), Jet Propulsion Laboratory (JPL), and GeoForschungsZentrum (GFZ). In addition to the official centers, multiple other institutions provide different levels of data, such as the Graz University of Technology, the Tongji University in Shanghai, and the University of Bonn. All centers provide the measurements for at least one of four processing levels, depending on the purpose of use: The Level 0 data are the originally measured data sensed from the instruments without any adaptions. Level 1 (1A and 1B) provides all necessary input processing steps to derive time-variable gravity fields. This means that the K-band ranges or – in the case of GRACE-FO – ranges from laser ranging interferometry together with measurements from the other instruments (star camera, accelerometer, GNSS, etc.) are prepared to make them publicly available for the scientific community (Case *et al.*, 2010). This means that sensor calibration factors are applied to Level 0 data to transform them to geophysical units, time tags are added, a quality check is performed, and the sampling rate is reduced. The gravity field coefficients are available in Level 2 as Spherical Harmonic Coefficients (SHC). The most user-friendly level is Level 3, the final spatially gridded TWSA. For the conversion from Level 2 to Level 3, corrections are applied. In most cases, this implies a filtering, a temporal mean removal per SHC, and the removal of the Glacial Isostatic Adjustment (GIA) (Sec. 2.3). Thus, a Level 3 product can directly be used for many research questions. However, not all applications can be pursued with these data as in some cases a different or additional processing is required.

Another possibility to directly process Level 3 data from Level 1 data is the so-called Mass Concentration block (mascon) approach. With mascons, the range-rate measurements are related to the mass concentration functions at chosen locations via partial derivatives in a regularized approach. Due to the regularization, the mascons system does not require a de-striping filter, as it is commonly done for level 2 SHCs. However, mascon approaches are not independent of hydrological model simulations or other GRACE/-FO solutions, because prior information is required. For example, mascon solutions by JPL use model-driven correlations (Watkins *et al.*, 2015). To be able to process near-real-time gravity fields, the mascons from Goddard Space Flight Center (GSFC), avoid using the model-driven correlations. Instead, they use trend and annual amplitude information from a previous forward solution – e.g., GSFC V1.1 but the authors suggest this could be any other time variable gravity product – and extrapolate them further in time (Loomis *et al.*, 2019). In fact, this method still uses regularization and is again not

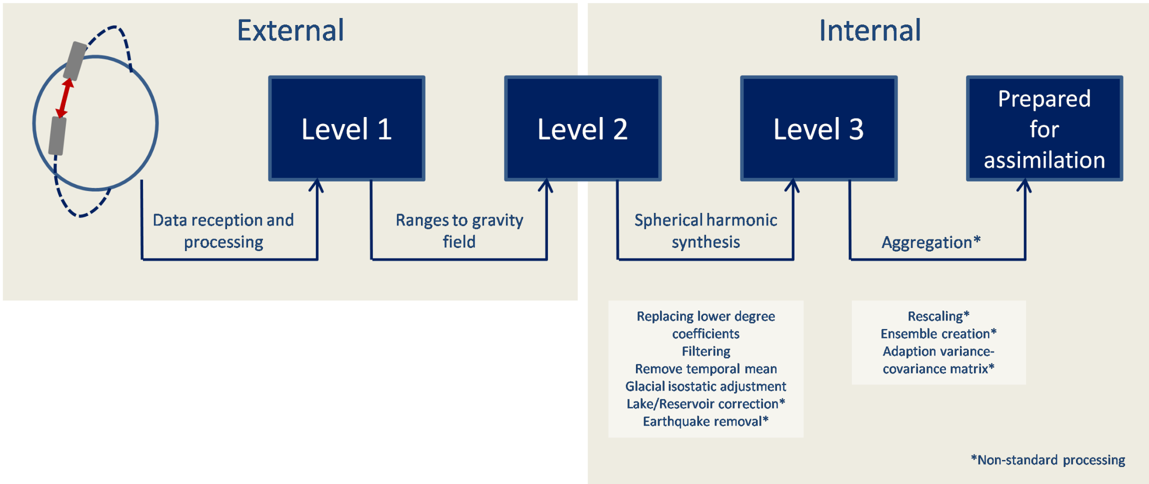


Figure 2.2: Simplified schematic illustration of GRACE/-FO data levels. At the University of Bonn, the processing from level 2 to level 3 and the final preparation for data assimilation is implemented as part of a self-developed software package.

completely independent from hydrological models. Nevertheless, Watkins *et al.* (2015) showed that after applying scale factors – accounting for signal attenuation during the processing – there is no significant difference between both processing types. To conclude, available high-resolution level 3 TWSA products based on mascon solution or SHCs are both restricted to the native GRACE/-FO spatial resolution of about 300 km and no geophysically meaningful signal below that is contained (Watkins *et al.*, 2015; Vishwakarma *et al.*, 2021).

To avoid including constraints, regularization, and additional model information in the observations and enable self-controlled processing in this thesis, external level 2 spherical harmonic solutions are downloaded and processed with the help of a self-developed software making use of potential theory via spherical harmonic synthesis to gridded fields (Fig. 2.2). The reverse transformation is the spherical harmonic analysis, which transforms spatial grids to SHCs. The potential theory for performing the transformations and processing corrections are explained in the following sections.

## 2.2 Potential theory

Potential theory is the basis for describing the gravitational potential via spherical harmonic functions and has widely been described in the literature, for example in early literature by, e.g., Sternberg & Smith (1944). Here, I mainly follow Hofmann-Wellenhof & Moritz (2006) and Ilk (2021) for the potential theory and Wahr *et al.* (1998) for details about expressing the gravitational potential in terms of mass change.

A fundamental law in potential theory is Newton's law of gravitation, which describes that two points with masses $m_1$ and $m_2$ with the distance $r$ in-between attract each other with the force

$$F = G\frac{m_1 m_2}{r^2}, \tag{2.1}$$

where $G$ is the gravitational constant. The force is acting along the direct connection between the two points. Suppose that one mass is the attracting mass and the other mass is the attracted mass, the equation can be simplified by setting the attracted mass to unity

$$F = G\frac{m}{r^2}. \tag{2.2}$$

We can now introduce the gravitational potential of a point mass by a new scalar function with

$$V = \frac{Gm}{r}. \tag{2.3}$$

Next, a solid body with density $\rho$ can be viewed as multiple point masses that are continuously distributed over the volume $v$

$$V = G\iiint_v \frac{dm}{r} = G\iiint_v \frac{\rho}{r}dv \tag{2.4}$$

with

$$\rho = \frac{dm}{dv}. \tag{2.5}$$

Upon introducing a Cartesian coordinate system, which leads to the representation of the force as a vector, the gradient of the gravitational potential is computed, i.e. $\mathbf{F} = [XYZ] = \nabla V$.

One can now describe the gravitational potential inside or outside of a body. The potential inside of a body is described using Poisson's equation. Here, the focus is on the case outside of a body described via the Laplace equation. When density becomes zero, the sum of the second derivatives (Laplace operator $\Delta$) of the potential in Cartesian coordinates becomes zero. Together with introducing the spherical coordinates this yields in

$$\Delta V(\lambda, \theta, r) = \nabla\nabla V = 0 \tag{2.6}$$

with the geocentric longitude $\lambda$ and co-latitude $\theta$ (polar distance). The solutions of the Laplace equations are called (spherical) harmonic functions and are introduced next.

### 2.2.1 Surface spherical harmonics

As mentioned before, the gravitational potential can be expanded with a polynomial function as it is harmonic because of fulfilling the Laplace equation. A converging harmonic function can be approximated for the empty exterior space ($F_e$) of a mass distribution via

$$F_e(\lambda, \theta, r) = \sum_{n=0}^{\infty} \frac{R}{r}^{n+1} Y_n(\lambda, \theta), \tag{2.7}$$

with the Earth's reference radius $R$. Assuming the harmonic function shall be referred to the Earth's surface, $r$ is equal to $R$, and the series expansion results in

$$F_e(\lambda, \theta) = \sum_{n=0}^{\infty} Y_n(\lambda, \theta), \tag{2.8}$$

with degree $n$ (integer numbers), the geocentric longitude $\lambda$, and co-latitude $\theta$ (polar distance).

The harmonic expansion of $Y_n$ is called the surface spherical harmonics consisting of a series of orthogonal functions. Every single surface spherical harmonic represents a solution to the

Laplace equation, and it can be written as a linear combination that includes the SHCs $c_{nm}$ and $s_{nm}$ of degree $n$ and order $m$ as well as the Legendre function $P_{nm}(\cos\theta)$, yields in

$$Y_n(\lambda,\theta)=\sum_{m=0}^{n}[(c_{nm}\cos(m\lambda)+s_{nm}\sin(m\lambda))P_{nm}(\cos\theta)] \tag{2.9}$$

and is inserted in Eq. 2.8. The resulting gravitational potential on the Earth's surface via surface spherical harmonic expansion is given via

$$V(\lambda,\theta)=\frac{GM}{R}\sum_{n=0}^{\infty}Y_n(\lambda,\theta) \tag{2.10}$$

and by including Eq. 2.9 in Eq. 2.10 yields the full description of the gravitational potential related to the SHCs with spherical coordinates:

$$V(\lambda,\theta)=\frac{GM}{R}\sum_{n=0}^{\infty}\sum_{m=0}^{n}[(c_{nm}\cos(m\lambda)+s_{nm}\sin(m\lambda))P_{nm}(\cos\theta)]. \tag{2.11}$$

The Legendre functions can be derived via a direct implementation via a summation formula or a recursive implementation. The direct implementation fails for medium to high degrees, thus here the Legendre polynomials are obtained in a recursive and stable procedure. This requires the introduction of normalization, which in the following is always necessary for the polynomials $\bar{P}_{nm}$ and spherical harmonic coefficients $\bar{c}_{nm}$ and $\bar{s}_{nm}$. The fully normalized Legendre polynomials of degree $n$ and order $m$ are computed as follows:

$$\bar{P}_{nn}=\sqrt{\frac{2n+1}{2n}}\;\sin(\theta)\;\bar{P}_{n-1,n-1} \tag{2.12}$$

$$\bar{P}_{n,n-1}=\sqrt{2n+1}\;\cos(\theta)\;\bar{P}_{n-1,n-1} \tag{2.13}$$

$$\begin{aligned}\bar{P}_{n,m}=\sqrt{\frac{(2n+1)}{(n-m)(n+m)}}\cdot\Big(\sqrt{2n-1}\;\cos(\theta)\;\bar{P}_{n-1,m}-\\ \sqrt{\frac{(n-m-1)(n+m-1)}{2n-3}}\;\bar{P}_{n-2,m}\Big),\end{aligned} \tag{2.14}$$

with the starting values $\bar{P}_{00}=1$ and $\bar{P}_{11}=\sqrt{3}\sin(\theta)$ (Colombo, 1981).

The expansion of potential coefficients via $Y_n$ to spatial grids is also called spherical harmonic synthesis. Reversely, the efficient computation from spatial grids back to SHCs is applied through spherical harmonic analysis. By introducing the normalized abbreviations

$$\begin{aligned}\bar{C}_{nm}(\lambda,\theta)=\bar{P}_{nm}(\cos\theta)\cos(m\lambda)\\ \bar{S}_{nm}(\lambda,\theta)=\bar{P}_{nm}(\cos\theta)\sin(m\lambda)\end{aligned} \tag{2.15}$$

the spherical harmonic analysis can be expressed in

$$\begin{aligned}c_{nm}&=\frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\lambda'=0}^{2\pi}F(\lambda',\theta)\bar{C}_{nm}d\lambda' d\theta\\ s_{nm}&=\frac{1}{4\pi}\int_{\theta=0}^{\pi}\int_{\lambda'=0}^{2\pi}F(\lambda',\theta)\bar{S}_{nm}d\lambda' d\theta.\end{aligned} \tag{2.16}$$

From the coefficients signal degree variances can be estimated, which means the variance of the Earth's potential depending on degree n by

$$\sigma_n^2=\sum_{m=0}^{n}(c_{nm}^2+s_{nm}^2). \tag{2.17}$$

In other words, the signal degree variances describe the power spectrum of the Earth's potential. As described here, the degree variances are unitless but can be expressed in geoid heights with the help of the radius

$$\sigma_n^2(N) = R^2\sigma_n^2. \tag{2.18}$$

An approximate estimate relates the spherical harmonics maximum degree $n_{max}$ to the spatial resolution on the grid as

$$n_{\max} \approx \sqrt{\frac{180°}{\Delta\theta}\frac{360°}{\Delta\lambda}}. \tag{2.19}$$

### 2.2.2 Mass changes

Following Wahr *et al.* (1998) and Wahr (2007), one can now relate the gravitational potential to mass redistribution within the Earth system. For this, a thin layer is assumed to describe the surface density redistribution caused by a mass change. Thus, within the layer, the mass changes are concentrated at the Earth's surface. We interpret that the mass changes in the atmosphere, ocean, and land water masses such as ice caps, glaciers, and surface and subsurface masses occur within such a layer. In general, the layer is assumed to have a thickness of approximately 10 to 15 km, because it refers to the atmosphere as the largest part, whereas for hydrological and oceanographic applications the layer is approximately 1 km.

As next, the surface mass density changes are defined as the radial integral of the density redistribution in the thin layer via

$$\Delta\sigma(\theta,\lambda) = \int_{\text{layer}} \Delta\rho(r,\theta,\lambda)dr. \tag{2.20}$$

Expressed in spherical harmonics, the integral is replaced by the sum of Legendre functions (Wahr, 2007), and by directly integrating Eq. 2.15, the expression yields in

$$\Delta\sigma(\theta,\lambda) = \frac{M}{4\pi R^2}\sum_{n=0}^{n_{\max}}\sum_{m=0}^{n}\frac{2n+1}{1+k_n'}(\Delta c_{nm}\cos(m\lambda)\bar{C}_{nm}(\cos\theta)+\Delta s_{nm}\sin(m\lambda)\bar{S}_{nm}(\cos\theta)), \tag{2.21}$$

where $\Delta c_{nm}$ $\Delta s_{nm}$ are anomalies of the dimensionless SHCs up to the maximum degree/order $n_{\max}$, $\theta$ and $\lambda$ are the co-latitude and longitude, and $k_n'$ are the elastic load Love numbers. The load Love numbers are an integral part of Earth models, which were developed to describe the properties of the Earth. Under the assumption of a radially symmetric Earth, they represent the gravity signal caused by solid Earth changes induced by surface mass changes and can be provided from Earth models. As can be found from the previous equation, the Stokes SHC can be related to surface mass density coefficients following

$$\Delta c_{nm}^{\sigma} = \frac{\rho_e}{3\rho_w}\frac{2n+1}{1+k_n'}\Delta c_{nm}, \tag{2.22}$$

where $\rho_e$ is the mean density of the Earth with approximately 5518 $kg/m^3$ and $\rho_w$ is the average density of water assumed to be 1025 $kg/m^3$. The change in surface mass density can be expressed in equivalent water heights as commonly used in hydrology via

$$\Delta S(\theta,\lambda) = \frac{M}{4\pi R^2\rho_w}\sum_{n=0}^{n_{\max}}\sum_{m=0}^{n}\frac{2n+1}{1+k_n'}(\Delta c_{nm}\cos(m\lambda)\bar{C}_{nm}(\cos\theta)+\Delta s_{nm}\sin(m\lambda)\bar{S}_{nm}(\cos\theta)). \tag{2.23}$$

As the geoid is another typical representation of gravitational changes, the expansion of gravity to changes in the geoid is

$$\Delta N(\theta,\lambda) = R\sum_{n=0}^{n_{\max}}\sum_{m=0}^{n}\Delta c_{nm}\cos(m\lambda)\bar{C}_{nm}(\cos\theta)+\Delta s_{nm}\sin(m\lambda)\bar{S}_{nm}(\cos\theta). \tag{2.24}$$

### 2.2.3 Vertical deformation

From the previously described gravitational potential and surface masses one can also express the gravity changes in displacements, here with a focus on vertical displacements. The vertical displacements can directly be computed from the integration of surface mass load in combination with the Green's functions $G(\psi)$ over the whole sphere $\Omega$ as

$$\Delta\nu(\theta,\lambda,t) = \int_{\Omega} \Delta\sigma(\theta',\lambda',t)G(\psi)d\Omega, \tag{2.25}$$

where the Green's function includes the load Love numbers $h'_n$ and the Legendre Polynome as

$$G(\psi) = \frac{R}{M}\sum_{n=0}^{\infty} h'_n \bar{P}(\cos\theta). \tag{2.26}$$

For applications, it is important that the Green's function converge fast to zero, which is not the case for the previous equation. In order to provide faster convergence, the Kummer transformation is applied. Following Wang *et al.* (2012), an approximation is given by

$$G(\psi) = \frac{Rh'_\infty}{M}\left(\frac{1}{2\psi} - 1\right) + \frac{R}{M}\sum_{n=1}^{\infty}(h'_n - h'_\infty)\bar{P}(\cos\theta) \tag{2.27}$$

where $\psi = \sin\frac{\theta}{2}$ is the angle between the position of (point) load and the location of interest and $h_\infty$ is the load Love number for infinity degree. In praxis, the load Love number for infinity degree is replaced by the maximum degree. For degrees higher than the maximum degree, Eq. 2.27 approaches quickly zero, as the load Love numbers approach the limit value.

One could also compute the vertical displacements by directly relating spherical harmonic surface mass coefficients to vertical displacement coefficients by

$$\Delta c^{\nu}_{nm} = \frac{3\rho_w}{\rho_e}\frac{1}{2n+1}h'_n \Delta c^{\sigma}_{nm} \tag{2.28}$$

(compare to Kusche & Schrama, 2005). The SHCs of vertical displacements are then used to derive the final vertical displacement per location with

$$\Delta\nu(\theta,\lambda) = R\sum_{n=0}^{n_{\max}}\sum_{m=0}^{n} \Delta c^{\nu}_{nm} cos(m\lambda)\bar{C}_{nm}(\cos\theta) + \Delta s^{\nu}_{nm}\sin(m\lambda)\bar{S}_{nm}(\cos\theta). \tag{2.29}$$

## 2.3 Processing steps for deriving GRACE/-FO mass changes

A starting point for processing GRACE/-FO data is to choose a specific data product. The level 2 spherical harmonic coefficients are provided by various processing centers, however, here I use the ITSG2018 GRACE solutions and ITSG operational GRACE-FO solutions provided by TU Graz (Mayer-Gürr *et al.*, 2018; Kvas *et al.*, 2019) because they are provided with full error co-variances in the form of the normal equations together with the gravity fields. The solutions are either provided up to degree and order 60, 96, or 120. The higher the degree, the more signal bands are included but at the same time, the errors in the gravity fields increase. Therefore, the solutions of degree and order 96 are utilized, as typically done for hydrological applications.

In general, the SHCs are transformed to gridded level 3 TWSA as was shown in Eq. 2.23. In the following, additional standard and advanced processing steps are described. The subdivision into processing steps in the spectral and spatial domain is set according to how it was done in this thesis, which is close to standard procedure for other processing centers. Nonetheless, all spatial corrections could also be transformed to the spectral domain – and vice versa – before applying them.

### 2.3.1 Spectral domain

The processing of GRACE/-FO data in the spectral domain includes the removal of a temporal mean per SHC, replacing lower degree coefficients, and filtering. The temporal mean is computed from SHCs for a chosen period, e.g., from 2003 to 2016. Afterwards, for each available monthly field, this temporal mean is removed to derive anomalies of the SHCs (Eq. 2.21). Since defining a time frame to which the temporal mean refers also depends on the application – here data assimilation using a hydrological model – the topic of temporal mean removal is discussed in the assimilation Sec. 5.2.5, while other application-independent processing steps are explained in the following.

#### Replacing lower degree coefficients

Due to the measurement configuration of the GRACE/-FO missions, lower degree coefficients of the spherical harmonics are measured imprecisely or not reasonably. The degree one coefficients of the GRACE/-FO gravitational potential coefficients describe the coordinates of the center of mass relative to the center of figure. As the missions are measuring regarding the center of mass, inter-satellite ranging and range rates on-board of GRACE/-FO are insensitive to these coefficients. The $c_{20}$ coefficient is related to the dynamic flattening of the Earth, i.e., the oblateness. This coefficient can be measured more precisely by other satellite missions such as satellite laser ranging. Due to the loss of the GRACE-B accelerometer data and for intermittent data for several months (Loomis *et al.*, 2020), $c_{30}$ is also considered to be unreliable and, thus, needs to be replaced. In more detail, starting in November 2016, the accelerometer of GRACE-B (GRACE) was turned off to prevent the diminished battery from complete depletion. Until the end of its lifetime and with the start of GRACE-FO, the GRACE-D accelerometer data were much more noisy than the GRACE-C accelerometer data (Loomis *et al.*, 2020).

Thus, the coefficients of degree one ($c_{10}$, $c_{11}$ and $s_{11}$), $c_{20}$ and $c_{30}$ need to be replaced by other solutions: Degree one coefficients (RL06, Technical Note 13) and $c_{20}$ coefficients (RL06) are provided by NASA/JPL [3,4] following the approaches as recommended by Swenson *et al.* (2008), Cheng *et al.* (2011), Cheng *et al.* (2013), and Sun *et al.* (2016) and $c_{30}$ (Technical Note 14) is provided by NASA/GSFC (Loomis *et al.*, 2019). The correction for $c_{30}$ is limited to the period after July 2016 as recommended by Loomis *et al.* (2020); the other coefficients are corrected for the whole period. The final solutions are then referred to the center of the surface figure. It should be noted that the load Love numbers (e.g., presented in Eq. 2.21) also need to refer to the same center of frame.

#### Filtering

The GRACE/-FO satellites fly on the same orbit of 89.5°. Due to this constellation, measurement errors, and background models, the gravity field is much more sensitive in the north-south direction than in the east-west direction, which manifests itself as striping patterns in the gravity field solutions. Wahr *et al.* (1998) showed that with increasing degree of the coefficients, the noise, and especially also the striping noise increases. Several studies introduced filtering techniques to minimize this pattern. The idea of filtering for the GRACE/-FO missions is intended to remove high-frequency noise patterns, while, at the same time, reducing signal loss.

[3] `https://grace.jpl.nasa.gov/data/get-data/geocenter/` (last accessed 15.03.2024)
[4] `https://grace.jpl.nasa.gov/data/get-data/oblateness/`, (last accessed 15.03.2024)

Two main techniques are proposed for filtering: (1) as multiplication of a filter kernel with SHC in the spectral domain and (2) as convolution of a filter kernel with a grid in the spatial domain. The filter kernels themselves differ between isotropic and anisotropic filter (Wahr *et al.*, 1998; Chen *et al.*, 2005; Fengler *et al.*, 2007). Isotropic filters filter equally for all given directions. One of the most commonly used isotropic filters is the Gaussian filter (Jekeli, 1981). This is a filter that can be applied for different filter radii and is partially uniform, which means that its size does not depend on the locations. A commonly applied filter width is 500 km. The radius describes the distance at which the Gaussian weighting function achieves half its magnitude (Chen *et al.*, 2005). The filter was proposed by Wahr *et al.* (1998) for the GRACE mission and the filter kernel was first mentioned by Jekeli (1981). In the spatial domain, the gravity signal is convoluted with a weighting function, which is a bell-shaped Gaussian function. In the spectral domain, this corresponds to a degree-dependent weighting by multiplication.

However, the Gaussian filter requires large radii to remove the north-south stripes, which strongly dampens the signal. In contrast, anisotropic filters have a filter kernel in the spatial domain that varies with latitude, and in the spectral domain varies with degree and order. Many numbers of anisotropic filters have been developed in the last decades. One example of an anisotropic filter was introduced by Swenson & Wahr (2006), who combined the Gaussian filtering with a preceding decorrelation step. Its general procedure is a moving window shifted across the spherical coefficients with a quadratic polynomial kernel. They introduced this filter because they found correlations between same-order coefficients.

Another anisotropic and widely used filter is the DDK filter (Kusche, 2007; Kusche *et al.*, 2009). The DDK filter uses a regularization of a synthetic normal equation, which includes stationary fields of synthetic error covariance matrix and signal variance derived from geophysical models. The DDK filter is a stationary filter, which means that constant error information of the signal variance is assumed. Thus, the same filter kernel is applied for all monthly fields. There exist different filter widths from DDK1 (strong smoothing) with a larger filter radius and up to DDK9 (weak smoothing) with a small filter radius. The filter kernel is designed to be tighter in the north-south direction as compared to the east-west direction, has negative side lobes, and it generally depends on the location.

Klees *et al.* (2008) has further developed the DDK filter by using non-stationary error-covariance information and signal variance but these are non-public. Therefore, the filter cannot be accessed. A recently developed filter is the VADER filter (Horvath *et al.*, 2018). The filter is similar to Klees *et al.* (2008), which means it presents a temporally non-stationary filter but instead, it can be applied to any GRACE solution and has differences in their determination of the month-to-month signal variance.

To conclude, the DDK filter and VADER filter both present useful filters but the VADER filter weights are – to my knowledge – not publicly available and require operational updates for new monthly solutions due to its temporal non-stationarity. Consequently, I keep using the DDK filter as it is already implemented into the GRACE/-FO processing chain at the Institute of Geodesy and Geoinformation at the University of Bonn.

The effect of the DDK1, DDK3, and DDK5 filters on the GRACE/-FO TWSA spatial grids is exemplarily demonstrated in Fig. 2.3 for May 2010. With a focus on the land mass, it can be noticed that DDK1 has a strong smoothing effect on the TWSA. In the north of South America, strong negative TWSA smear into the ocean, which is not as much the case for the DDK3 filter. For the DDK5 filter, the smoothing effect is much less but the striping patterns visibly persist in the data and remain largely unsmoothed. The stripes are most intense in the ocean, but are

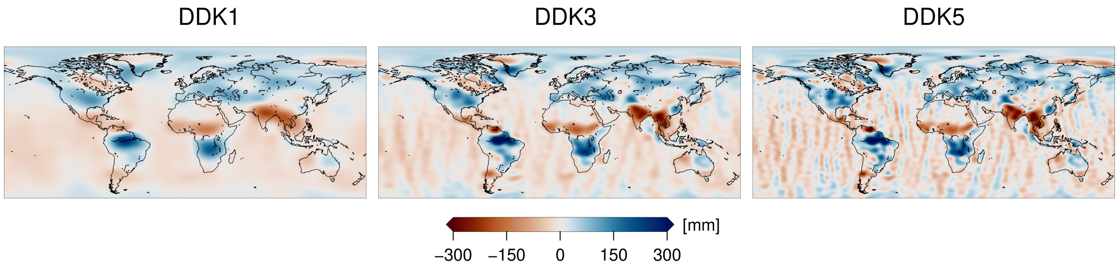


Figure 2.3: GRACE/-FO TWSA in May 2010 filtered with DDK1 (left), DDK3 (center), and DDK5 (right) with respect to the mean over 2003 to 2016.

also apparent on the land TWSA, for example, in the western part of North Africa a noise-like signal appears. To smooth as many stripes and as much noise as possible while not blurring the observed signal too much, the DDK3 filter is chosen for the global application in this thesis. For more local applications, the choice of DDK strength might be different and would need to be revised again.

### 2.3.2 Spatial domain

In the commonly agreed processing procedures, the effect of Glacial Isostatic Adjustment is accounted for at the spatial domain. Other effects one could account for in the processing are not always included, for example, by further introducing corrections for leakage or correcting for large earthquakes. I will present now how the Glacial Isostatic Adjustment (GIA), spatial leakage, and large earthquakes are considered in the processing at the spatial domain for this thesis.

#### Glacial isostatic adjustment correction

GIA is the effect of the response to ice unloading following the ice ages. The melting of the ice caps more than 11000 years ago led to a slowly developing elastic rebound of the solid Earth surface that continues nowadays. The largest effect of GIA is found over areas that are still permanently ice-covered, e.g. Greenland, but it also has large uplift effects of several millimeters to centimeters per year over Europe and Canada. Even far-field effects occur due to GIA, although these are much smaller than in Greenland, for example, the U.S. East Coast is sinking (Peltier, 2004; A *et al.*, 2012; Caron *et al.*, 2018).

Since the GRACE/-FO satellites measure changes in mass redistribution, the effect of GIA is still contained in the gravity field solutions. Depending on the application, GIA is subject of research, or its occurrence limits the analysis of other signals. For many hydrological applications and drought monitoring the GIA signal should be removed because it biases interpretations. The GIA effects occur worldwide, which emphasizes the need to correct for it globally.

To estimate the effect of GIA, multiple studies developed models to derive GIA rates in the last decades, for example, A *et al.* (2012), Caron *et al.* (2018), or Peltier *et al.* (2018). These models incorporate the ice history and the viscosity of the mantle to simulate GIA mass rates and uplift by solving a set of differential equations. Three available products of GIA rates are considered here for the use as provided for download by NASA/JPL[5]: the ICE5G (A *et al.*, 2012), the ICE6G-D (Peltier *et al.*, 2018), and the Caron-2018 trend rates. Fig. 2.4 shows the maximum minus the minimum trend from the three models that is computed per grid to detect regions where the models do not agree very well. This is denoted hereinafter as the “range” between

[5] https://grace.jpl.nasa.gov/data/get-data/gia-trends/

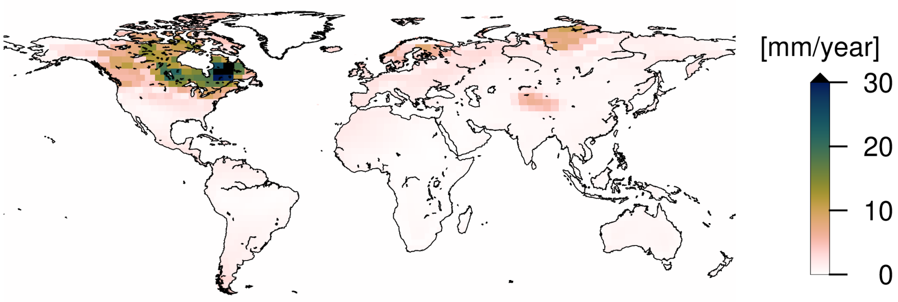


Figure 2.4: Maximum minus minimum linear trends [mm/year] for three different GIA models that can be used for correcting GRACE/-FO data.

the three models. Greenland is removed from the figure to cover only land mass cells that are also present in the modeling used in the further context of the thesis. Large ranges between the models can be found up to 30 mm or more per year in Eastern Canada. During the last ice age, a large ice dome was located in this region, which now looks different in the ice history of different GIA models. In addition, when computing the absolute difference in trends between two models globally averaged, the difference is much smaller for ICE5G-D and ICE6G-D (0.26 mm per year) as for Caron-2018 and ICE6G-D (0.38 mm per year) and Caron-2018 and ICE5G-D (0.40 mm per year). It should be recognized that these globally averaged differences in trends are rather small compared to the spatial maximum of trend differences one would derive because of the large number of cells close to zero.

Since Caron-2018 differs a lot from the other two models, and the ICE6G-D is the updated version of the ICE5G-D model, I decided to use the ICE6G-D model. Nonetheless, it should be noted that trend estimates from GRACE/-FO or GRACE/-FO assimilation outputs should be interpreted carefully in regions where the model GIA rates differ.

#### Leakage correction

To reduce the striping errors and errors of background models in the GRACE/-FO data, spatial filtering is applied as presented previously (Sec. 2.3.1). Spatial filtering introduces a problem: the filter cannot distinguish at all between noise or signal and therefore signal attenuation and signal strengthening occurs through applying a filter. This effect is very prominent for regions where neighboring grids have strong differences in signal, for example at coastlines, large lakes, and reservoirs, etc., and is called leakage (e.g., Klees *et al.*, 2007; Baur *et al.*, 2009). Considering the example of a large lake, the lake water masses signal leaks into the surrounding areas, which means the signal outside of the lake borders increases and at the same time the signal inside decreases (Fig. 2.5). Leakage is denoted as leakage-out for the regions of large signals leaking into the surrounding areas. Leakage-in then describes the leakage from an area with large signals into a region, where the signal is leaking into. Another source for the leakage effect is found in the spectral domain for the spherical harmonic synthesis that transforms the SHC to gridded TWSA. When applying the synthesis, a maximum degree $n_{\max}$ (Eq. 2.23) is chosen to which the series is expanded. With this truncation existing high frequencies in the data are mapped into the lower frequencies (e.g., Baur *et al.*, 2009).

Numerous studies developed approaches to deal with the leakage since the start of GRACE, for example, regional studies that dealt with basin or region averages spatially extended averaging kernels to incorporate signal that might leak the surrounding area (Swenson & Wahr, 2002). Since assimilation frameworks harmonize observation and model simulations, applying similar

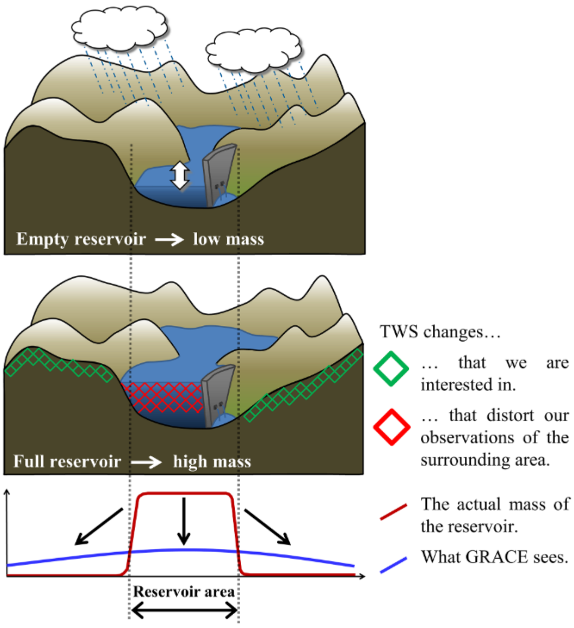


Figure 2.5: Illustration of lake leakage (taken from Deggim *et al.*, 2021).

filter techniques to both was proposed (e.g., Güntner, 2008). However, assimilation also aims to improve the coarse GRACE/-FO spatial resolution by downscaling to the model resolution. In the case that the model is filtered similarly to the data, this might introduce unwanted spatial smoothing to the assimilation outputs.

Instead, to restore the amplitudes in the filtered GRACE/-FO data on global scales, several studies compared filtered and unfiltered mass anomalies from global geophysical models (Klees *et al.*, 2007; Landerer & Swenson, 2012). This technique can also be transferred to hydrological models. One of the studies is by Landerer & Swenson (2012), who compared the difference between filtered and unfiltered total water storage anomalies from modeling. A data-driven approach by Vishwakarma *et al.* (2017) avoids incorporating models because this introduces also corresponding model uncertainties, and errors are then propagated to the data. However, the approach is limited to catchments. Dobslaw *et al.* (2020) further improved this data-driven approach by considering weaker and stronger filtered solutions in pairs (DDK filter) and applied it globally. A different way to account for leakage was provided by Deggim *et al.* (2021). They used forward modeling of water volumes from satellite altimetry and optical remote sensing from 280 lakes and reservoirs to provide a global gridded monthly leakage correction data set from 2003 to 2016. Thus, in their approach, they received measurements of the water masses from other missions than the GRACE/-FO missions, whereas the previous studies base their approaches on modeling and/or the GRACE/-FO observations.

In summary, several possibilities exist on how to account for a leakage correction. The two approaches of computing rescaling factors via a hydrological model and using the leakage correction data product for lakes and reservoirs have already been implemented in the in-house GRACE/-FO processing. Therefore, these two approaches are now described in depth. However, the global approach in Dobslaw *et al.* (2020) could be a future extension to the processing.

To calculate rescaling factors from hydrological modeling, modeled TWS is first reduced by a temporal mean, and spherical harmonic analysis is applied to derive and filter the spherical harmonic coefficients with the DDK3 filter as for the GRACE/-FO filtering. Next, filtered TWSA are computed from SHCs via spherical harmonic synthesis. To avoid including the effect that spherical analysis has, i.e. a spatial blur effect, the spherical harmonic coefficients of the model simulations are twice transformed to grids via spherical harmonic synthesis: One time with filtering and one time without filtering the SHCs. The difference between filtered and unfiltered modeled TWSA is then minimized in order to determine the rescaling factors $k$. The global rescaling factors are derived over land on a 0.5° grid and can be spatially aggregated flexibly in case of a coarser resolution than the 0.5° is required.

Long *et al.* (2015) categorized the magnitude of rescaling factors for grids into different interpretations for the large-scale mean: A factor lower than zero is out of phase, a factor equal to zero indicates an uncorrelated behavior and a factor between zero and 0.3 indicates that leakage is prominent. A regular lower amplitude at the grid level can be found for factors higher than 0.3 and lower than one, while larger amplitudes are for values higher than one and lower than three. A factor of exactly one means unfiltered and filtered TWSA model simulations have the same amplitudes. Values of larger than three are related to coastlines or strong local behavior like lakes and reservoirs. Following Long *et al.* (2015), rescaling factors lower than zero are set to zero, and factors higher than three are set to three. Landerer & Swenson (2012) discuss in detail if there is a difference if the rescaling factor is first applied, and second, a spatial mean is computed or in reversed order. They tested small and large river basins and found that computing the rescaling factors both ways leads to similar results.

The leakage correction REgional COrrections for GRACE for Lakes and Reservoirs (RECOG-LR) provides two publicly available data sets (Deggim *et al.*, 2020), one for removing the surface water estimates and one for relocating mass changes to its origin. The removal and relocation data sets are produced by forward modeling of water volumes by altimetry and optical remote sensing, and provided as monthly correction fields from 2003 to 2016. Linear trends from the RECOG-LR data set for the removal and relocation data set are shown in Fig. 2.6. As expected, the trend removal map shows a spatially smeared trend signal around large lakes because Deggim *et al.* (2021) had to apply filtering, for example, for the Great Lakes in North America. The trends for the relocation show very small-scale intense linear trends for single grid cells, thus no spatial smearing is included.

Nonetheless, the monthly RECOG-LR removal and relocation data sets do not cover the complete study period used in this thesis. To be able to include RECOG-LR in the processing up to 2019, I use the estimated linear removal and relocation trends (computed for 2003 to 2016) and compute monthly correction fields for the time period 2003 to 2019. In other words, the 2003 to 2016 trends are distributed to the 2003 to 2019 period.

#### Earthquake correction

A natural phenomenon that is contained in the GRACE/-FO gravity fields is the effect of earthquakes, which cause a redistribution of masses on Earth. The earthquake signal can mask important hydrological signals. Especially for drought detection, it is important to correct for earthquakes because the sudden jump in the time series due to an earthquake could be interpreted as a drought or flooding event. Some studies found that large earthquakes are detectable in the GRACE/-FO data such as the 2004 Sumatra-Andaman earthquake, the Maule earthquake close to Chile, and the Tohoku earthquake close to Japan (e.g., Panet *et al.*, 2007; Einarsson *et al.*, 2010; Broerse, 2014).

Typically, the provided level 3 solutions of the official centers do not account for earthquake-related water mass changes in their standard processing since it can also be a relevant signal for analysis. Gravity field solutions that are corrected for earthquakes are, for example, provided by the GFZ in Potsdam (Boergens *et al.*, 2019). To correct for large earthquakes, this section follows

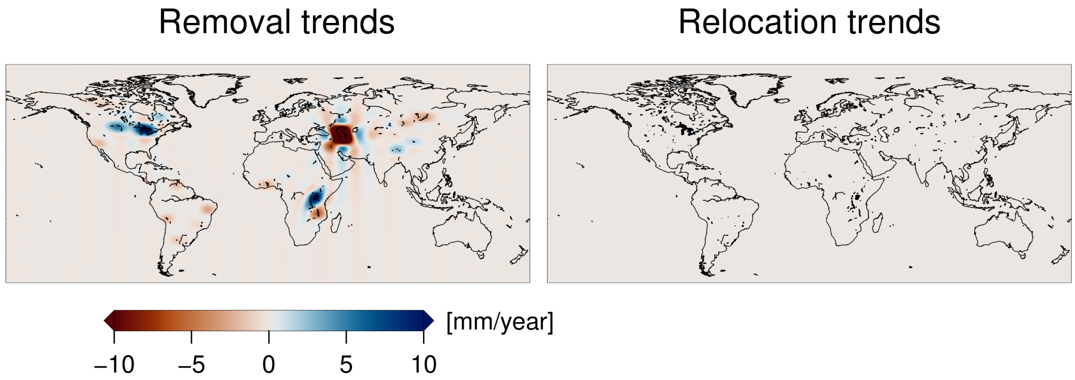


Figure 2.6: Linear trends for the lake and reservoir removal (left) and relocation (right) of RECOG-LR. The trends are given in mm per year and are computed for 2003 to 2016.

Deggim *et al.* (2021), where an approach is presented to remove earthquake signals from GRACE data by fitting a model to the temporal behavior. The approach applied for the fitting is a sampling-based Monte-Carlo approach presented in Einarsson *et al.* (2010) and Einarsson (2011). In this thesis, a correction is applied for earthquakes with a magnitude of 9.0 or higher. In fact, this includes the removal of the Sumatra-Andaman earthquake from the GRACE/-FO data, which had a magnitude of 9.1 and took place in December 2004, and the Tohoku earthquake, which had a magnitude of 9.0 and occurred in March 2011. The choice of removing earthquakes with a minimum magnitude of 9.0 is made based on Einarsson (2011): In this thesis, it was concluded, that earthquakes with a magnitude as small as the Nias earthquake (8.7 or smaller) require more refined techniques than the Monte-Carlo approach for separating the respective earthquake because of a low signal-to-noise ratio. Additionally, in the special case of the Nias earthquake, the separation is not recommended as it is in close temporal and spatial vicinity to the Sumatra-Andaman earthquake.

For including a fitting of earthquake signals into the processing of the GRACE/-FO data, it is required (after temporal mean removal and filtering of the SHC) to perform backward modeling of the SHCs to geoid changes as shown in Eq. 2.24. The geoid changes at a specific time ($t$) and location ($\theta, \lambda$) then can be seen as the combination of a bias ($\Delta N_{\text{bias}}$), linear trend ($\Delta N_{\text{trend}}$), annual signal ($\Delta N_{\text{ann}}$), semi-annual-signal ($\Delta N_{\text{semiann}}$), the S2 de-aliasing period ($\Delta N_{\text{S2}}$) and an earthquake signal ($\Delta N_{\text{EQ}_\nu}$) for the respective earthquake $\nu$ as

$$\begin{aligned} \Delta N(\theta,\lambda,t) = \Delta N_{\text{bias}}(\theta,\lambda,t) + \Delta N_{\text{trend}}(\theta,\lambda,t) + \\ \Delta N_{\text{ann}}(\theta,\lambda,t) + \Delta N_{\text{semiann}}(\theta,\lambda,t) + \Delta N_{\text{S2}}(\theta,\lambda,t) + \\ \Delta N_{\text{EQ}_\nu}(\theta,\lambda,t). \end{aligned} \tag{2.30}$$

The earthquake signal $\Delta N_{\text{EQ}_\nu}$ consists of a co-seismic and a post-seismic component. The co-seismic earthquake is modeled via a constant value beginning at the time of the respective earthquake ($t_\nu$) by a Heaviside step function $H$. The post-seismic signal also starts at the time of the respective earthquake but in contrast, has a post-seismic delay model through an exponential function that includes the relaxation time $\tau$

$$\Delta N_{\text{EQ}_\nu}(\theta,\lambda,t) = C_{\nu_{\text{co}}}(\theta,\lambda)H_{t_\nu}(t) + C_{\nu_{\text{post}}}(\theta,\lambda)H_{t_\nu}(t)\left(1 - \exp - \frac{t - t_\nu(\theta,\lambda)}{\tau(\theta,\lambda)}\right). \tag{2.31}$$

The coefficients and timing parameters of the respective earthquakes $C_{\nu_{\text{co}}}$, $C_{\nu_{\text{post}}}$, $t_\nu$ and $\tau$ are being estimated for each grid cell independently by a non-linear Bayes estimator together with the coefficients required for the other hydrological signatures (Eq. 2.30). The two time-related

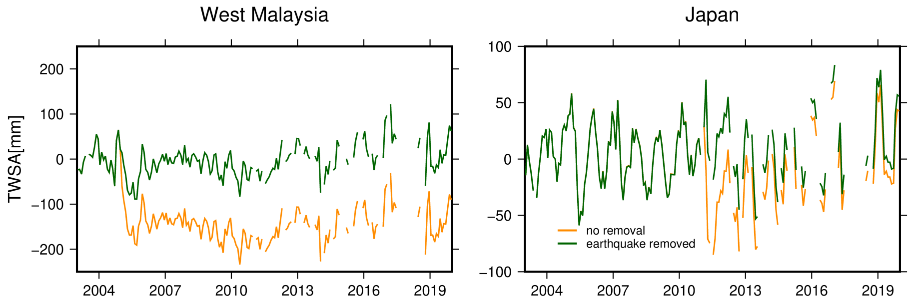


Figure 2.7: GRACE/-FO TWSA [mm] spatially averaged over West Malaysia (left) and Japan (right) with and without an earthquake correction (based on Deggim *et al.*, 2021).

variables of the respective earthquakes $t_v$ and $\tau$ are location-depending because the start or relaxation of the earthquake within the data per location may be delayed the farther the distance to the epicenter (e.g., Einarsson, 2011). The Bayes estimator solves for the coefficients in a Monte Carlo approach requiring $n$ ensemble members ($n$ is chosen to be 750 members). More information about the Bayes estimator can be found in Einarsson (2011). Finally, forward modeling transforms the geoid changes back to SHC via spherical harmonic analysis. Now, the classical processing of SHCs to gridded TWSA by performing spherical harmonic synthesis can be continued.

An earthquake correction data set was published (Gerdener *et al.*, 2020b), which is available from 2003 to 2016. The approach for estimating earthquake corrections has been extended to the full period and specific GRACE/-FO processing of this thesis as shown in Fig. 2.7 (based on Deggim *et al.*, 2020). The GRACE/-FO TWSA in West Malaysia and Japan show a clear jump at the time of the respective earthquakes (December 2004 for Sumatra-Andaman and March 2011 for the Tohoku earthquake). After applying the earthquake correction, this jump has been significantly reduced.

## 2.4 Uncertainty estimation

Errors in the gravity field coefficients can be caused by instrument noise, aliasing, and data processing (e.g., Wahr, 2007) and map into the mass fields. The aliasing occurs because the sampling rate for adequately measuring short-term variations in the gravity field is not high enough. As the observations are measured along the orbital path of the GRACE mission, monthly averages are computed to provide a global field. The infrequent sampling of the short-term variations along the orbit then leads to aliasing in the monthly averages (Wahr, 2007). De-aliasing models were developed to reduce this effect (atmospheric variability, and short-period variations in ocean bottom pressure) but the errors of the models propagate to the gravity fields.

To estimate spatially gridded error information for TWSA, the full normal equation systems provided for the gravity field computation by TU GRAZ are used (Mayer-Gürr *et al.*, 2018). Other processing centers that provide full month-to-month normal equations systems are CSR and GFZ. The underlying assumption here is that the inverse normal equation matrix represents the variance-covariance matrix of the spherical harmonic coefficients. Thus, the normal equation matrices for each time step are inverted to variance-covariance matrices and propagated to spatial TWSA variance-covariance matrices. These gridded variance-covariances matrices are important for this thesis as the assimilation that I will use (Sec. 5) requires observation error information in the filter algorithms. It is also possible to propagate the variances (i.e. variance-covariance matrix that is zero except for the diagonal elements) that are provided together with the spherical harmonic coefficients from the spectral domain to spherical grids. However, the huge advantage of using the inverted normal equations is that the correlations between the spherical harmonic coefficients are considered and can further be propagated to the spatial domain to derive more precise error behavior on the grid and spatial correlations.

The normal equation matrix $\mathbf{N}$, the right-hand side normal equations vector $\boldsymbol{n}$, and the parameters $\tilde{\boldsymbol{x}}$ form the normal equation system as

$$\mathbf{N}\tilde{\boldsymbol{x}} = \boldsymbol{n}. \tag{2.32}$$

To derive the error information in the form of the full variance-covariance matrix on the grid level, a few consecutive steps are applied. In this thesis, the maximum degree and order used is the same as for the degree and order of the provided normal equation system, i.e., degree and order 96. In case a maximum degree and order lower than 96 is required during the processing, the normal equation matrix can be reduced to the used maximum degree (e.g., $n_{\max} = 60$) by de-correlating the normal equations system below/equal to degree and order 60 and above (Niemeier, 2008). This functionality is also implemented in the GRACE/-FO processing at the Institute of Geodesy and Geoinformation and can be relevant for many applications, for example, for comparison to studies that make use of a degree and order 60 normal equations. From the (reduced) normal equations, the full variance-covariance matrix for the spherical harmonic coefficients is estimated

$$\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \sigma_0^2 \mathbf{N}^{-1}, \tag{2.33}$$

where $\sigma_0$ is the a-posteriori variance. Subsequently, the variance propagation is performed for the temporal mean removal, filtering, and spherical harmonic synthesis

$$\boldsymbol{\Sigma}_{\boldsymbol{ll}} = \mathbf{F}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\mathbf{F}^T. \tag{2.34}$$

$\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ is the input error matrix of the SHC, which can be interpreted as input observation error. $\mathbf{F}$ denotes the relationship between input errors and output errors after applying one of the processing steps. $\boldsymbol{\Sigma}_{\boldsymbol{ll}}$ yields the output error matrix of the output variable. For example, for propagating the errors for the processing step of filtering, $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ is the input variance-covariance matrix of the spherical harmonic coefficients and $\boldsymbol{\Sigma}_{\boldsymbol{ll}}$ is the output variance-covariance matrix after propagating the filter errors. Finally, the full variance-covariance matrix for TWSA is derived on the spatial grid of choice.

As was shown in Güntner *et al.* (2023), multiple Level 3 solutions of GRACE/-FO derived via the spherical harmonics slightly differ due to the different underlying Level 2 data or processing choices. In the best case, the variance-covariance matrix reflects the errors in such a way that all different GRACE/-FO solutions experience values mainly within the errors. To here quantify the variations in TWSA due to different GRACE/-FO solutions and possible adapt variance-covariance matrices for further incorporation into the assimilation framework (Sec. 5.2), the current GRACE/-FO TWSA is compared to other solutions under the consideration of the variance-covariance matrix estimated for this thesis (Eq. 2.34). This quantification aims to give reasonable advice for using the variance-covariance matrix in its original form ($1\sigma$) or multiplied with factors $2^2 = 4$ ($2\sigma$) or $3^2 = 9$ ($3\sigma$). Under the assumption of Gaussian distribution, the $1\sigma$, $2\sigma$, and $3\sigma$ standard deviation would cover approx. 68.27, 95.45, or 99.73% of the data (per grid), respectively.

Tab. 2.1 shows the percentage of CSR or GFZ months of gridded TWSA fields that fit into the error band of the TWSA processed for this thesis, here denoted as ITSG TWSA. The error band of ITSG is defined as the current TWSA plus/minus the standard deviation for the $1\sigma$, $2\sigma$, and $3\sigma$ widths per grid and per month as derived from propagating the full normal equations to the grid. Here, the evaluation is exemplarily applied on a 4° grid – due to the large amount of storage and computational resources that a full error matrix on an 0.5° grid would require, and as this is

Table 2.1: Percentage of months where either CSR or GFZ TWSA fit into the error band of the TWSA processed in this thesis, globally averaged for all 4° grid cells. The error band is estimated via using the processed TWSA plus/minus 1, 2, or $3\sigma$, where $\sigma$ is the root of the variances from the variance-covariance matrix used in this thesis. The percentage is either globally averaged (column "all") or the spatial mean across cells with a latitude ($\theta$) of higher equal 45°, lower equal -45°, and between -45 and 45° is shown.

| | **Latitudes** | $\mathbf{1\sigma}$ | $\mathbf{2\sigma}$ | $\mathbf{3\sigma}$ |
|---|---|---|---|---|
| **CSR** [%] | all | 48.07 | 78.12 | 91.24 |
| | $\theta \geq 45°$ | 40.68 | 69.67 | 85.38 |
| | $\theta \leq -45°$ | 40.73 | 71.62 | 86.91 |
| | $-45° < \theta < 45°$ | 53.50 | 84.31 | 95.54 |
| **GFZ** [%] | all | 42.50 | 71.97 | 87.36 |
| | $\theta \geq 45°$ | 34.25 | 61.19 | 78.82 |
| | $\theta \leq -45°$ | 40.29 | 70.73 | 86.62 |
| | $-45° < \theta < 45°$ | 48.52 | 79.81 | 93.57 |

one of the typical in-house grid sizes for assimilation – but the conclusions can also be drawn for other resolutions. The percentage of fits is then either globally averaged or averaged for latitudes larger/equal 45°, lower/equal to -45°, and between -45 and 45°. All in all, the higher this error calibration factor, the better fit CSR and GFZ into the error band of ITSG. A look at the latitudes reveals that especially errors for the latitudes above 45° or below -45° require a higher multiplying factor than the latitudes between -45 and 45°. In conclusion, an error calibration factor of three is applied to the variance-covariance matrix of this thesis to provide a more realistic data uncertainty.

# Chapter 3

# Monitoring drought

## 3.1 Overview and definitions of drought

Drought is a complex and multifaced natural climate disaster that represents a shortage in water with a slow-onset nature (Van Loon, 2015). Drought should not be confused with aridity which is the regional permanent absence of precipitation (Wilhite, 1994) or heat waves that describe a short dry period of a few weeks connected to high temperature (e.g., Van Loon, 2015:and therein). A shortage can only be discovered when the current conditions are compared to the "normal" conditions at a certain location. There is no unique definition for drought but generally, various definitions exist according to in which flux or compartment of the water cycle the drought is occurring. Most studies distinguish between four types of drought as frequently used (e.g, Wilhite & Glantz, 1985; Tallaksen & Van Lanen, 2004):

1. meteorological drought,
2. soil moisture drought (agricultural drought),
3. hydrological drought,
4. and socioeconomic drought.

Meteorological drought is typically defined as a precipitation deficit and an increase in evapotranspiration compared to the "normal" situation, and can more rapidly develop than other drought types. Soil moisture drought refers to a decrease in water available in the root zone. In case soil moisture drought affects agricultural cropland, the drought is referred to as agricultural drought. Thus, an agricultural drought results from the meteorological drought taking long enough to limit the plants' demand. Hydrological drought refers to the effect of precipitation shortages on surface and subsurface water storages, in low streamflow, lake levels, reservoirs, and groundwater (Tallaksen & Van Lanen, 2004). Even though a long-lasting meteorological drought has ended, it takes from months to years for a hydrological drought to end and the corresponding water storage to recover to "normal" conditions (Yihdego *et al.*, 2019). The last drought type is socio-economic drought, which describes the effect of meteorological, soil moisture, and hydrological drought on the society and economy, for example on economic goods such as a shortage of bread. Therefore, socio-economic drought is very much dependent on social aspects of a region, for example, dominant working sectors, infrastructures, population, and political decisions.

There is a temporal evolution of drought through the water cycle and the impact on the ecosystem and society. A lack of precipitation can lead to a decrease of soil water content. The decrease in soil moisture can further lead to a shortage in groundwater storage. This suggests that in the same region, there is a typical sequence of drought types: meteorological drought followed by soil moisture drought, which is then followed by hydrological drought. However, by reason of the complexity of the water cycle, drought at a location does not necessarily show a temporal evolution from fluxes to storages, for example, meteorological drought can also happen at the same time as hydrological drought. Hydrological drought might occur in only one water compartment whereas other water compartments are not affected by the drought. For example, a lake decline impacted by drought might be compensated by groundwater surplus as was the case in the study of Tweed *et al.* (2009). They found that a lake and wetland complex (28 monitored water bodies) in South East Australia experienced a decrease in water level during an Australian drought (1997 – 2006) and 36% of the lakes monitored changed from groundwater throughflow or discharge lakes to intermittent recharge lakes. In addition to vertical interactions of compartments during drought, drought can also spread spatially to other regions due to land-atmosphere feedbacks (Schumacher *et al.*, 2022).

Monitoring drought means that the location and severity of a drought event are continuously tracked (e.g., Hayes *et al.*, 2012), thus, a drought is tracked in near-real time or close to it and the tracking is updated on a regular basis. In contrast, a retrospective drought analysis is applied for droughts in the past, thus, no continuous tracking is required anymore. For drought forecasts, retrospective and near-real-time data are used to extract relevant information such as trends, seasonality, etc. As drought is a slowly developing phenomenon, the extracted information can then be used together with the drought indices to predict possible drought events for the upcoming time (mostly a few months).

Drought risk depends on the exposure of a given society. A very simple example is that a region is more exposed to droughts if the population density is higher or it is more dependent on livestock farming. It is also relevant if they can adapt during the period of the drought, which can also be interpreted as how vulnerable society is to drought. Thus, drought does not only refer to the natural hazard component but it must be viewed in relation to exposure and vulnerability, which together determine the overall drought risk (IPCC, 2014; McGlade *et al.*, 2019).

To describe the options for monitoring drought from the perspective of drought as a hazard, I first show links between variables of the water cycle. Next, I explain the basics about probability distributions as these are used to compute drought indices. Then, I introduce meteorological, soil moisture, and hydrological drought indices that one can use in monitoring drought. The two final sections introduce the concepts of drought risk computation and lists existing regional and global warning systems for drought that use drought indices.

## 3.2 Linking variables of the water cycle

This section discusses the link between two main variables of the hydrological water cycle – precipitation and TWS – to give an understanding of how drought indices can be related (based on Gerdener *et al.*, 2020a). In general, the water balance equation describes that a change in storage $\frac{dS}{dt}$ is a result of the precipitation $P$, evapotranspiration $E$, and runoff $R$

$$\frac{dS}{dt} = P - E - R. \tag{3.1}$$

One can assume that the variability of precipitation from the climatological mean ($\Delta P$) is much stronger than the variability of evapotranspiration ($\Delta E$) and runoff ($\Delta R$). As a consequence, precipitation variability represents the main component for storage variation, and evapotranspiration and runoff variations are assumed to be neglectable small ($\Delta E = 0$, $\Delta R = 0$) in the following.

Many drought indices make use of precipitation accumulated over a certain period to relate the indices to water sources. As one of probably many early studies, McKee *et al.* (1993) presented five arbitrary but typical time scales, e.g. 3, 6, 12, 24, and 48 months, for the accumulation of precipitation and related the accumulated precipitation to water compartments, for example, soil moisture, and groundwater. Thus, TWSA $\Delta S$ (referred to a long-term temporal mean) in a specific month $t$ is referred to accumulated precipitation anomalies via

$$\overline{\Delta S}(t) = \Delta t \sum_{n=t_0}^{t} \Delta P_n \ , \tag{3.2}$$

where $t_0$ is the beginning of the monitoring and $\Delta t$ is the passed time to the chosen month $t$. In other words, the change in storage is equal to the precipitation summation. To estimate the change in storage between two times $t_1$ and $t_2$, the accumulation for the precipitation variability (with respect to climatological mean) reduces to the time period:

$$\overline{\Delta S}(t_2) - \overline{\Delta S}(t_1) = \Delta t \sum_{n=t_1}^{t_2} \Delta P_n. \tag{3.3}$$

Assuming TWSA are accumulated over several months, this implies a two-fold accumulation for the precipitation variability via

$$\sum_{\tau=t_0}^{t} \overline{\Delta S}_n(t) = \Delta t \sum_{\tau=t_0}^{t} \sum_{n=t_0}^{\tau} \Delta P_n. \tag{3.4}$$

For example, for a time scale of three months, an index that uses accumulated precipitation starts with the third month. The resulting first accumulation value is derived via accumulating the first to third months (e.g., January to March) of data, the second accumulation value is derived by accumulating months two to four (e.g., February to April), etc. The procedure is done for each of the subsequent months, thus, for a total of $n$ observed months of a climate variable, the final index time series yields $n-2$ elements. The accumulation of precipitation or TWSA can be used for drought indices to relate the severity of a drought to the duration of a drought. The idea behind that is that a water deficit is only significant if it lasts for a few months. At the same time, one could interpret the accumulation as temporal smoothing; it is done in a moving average manner. The accumulation is exemplarily shown for water storage by

$$x^{+}_{i,j,q} = \sum_{k=1}^{q} \overline{\Delta s}(t_{i,j+1-k}), \tag{3.5}$$

where $q$ is the accumulation period. A similar approach can be conducted to derive a time series of storage differences, as shown, in Eq. 3.3, here now shown for a $q$ months period:

$$x^{-}_{i,j,q} = \overline{\Delta s}(t_{i,j}) - \overline{\Delta s}(t_{i,j+1-q}). \tag{3.6}$$

The differences in storage can be interpreted as accumulated precipitation between these months.

## 3.3 Basics of probability theory

The detection of drought events, independent of which variable of the hydrological water cycle is analyzed, is often computed by fitting distribution functions to the data per grid cell to derive probabilities. Therefore, flux or water storage time series that represent the entire history over several decades are either empirically or parametrically represented by a Probability Distribution Function (PDF) or Cumulative Distribution Function (CDF) for computing drought indices. For drought indices based on precipitation, two common distribution functions are the Gamma distribution and the log-logistic distribution. For example, the Standardized Precipitation Index (SPI) that will be described in Sec. 3.4.1 uses precipitation as input data and is calculated by fitting a Gamma distribution to the precipitation time series per grid cell. Afterwards, the derived probability values are transformed such that they are Gaussian distributed. In addition, the assimilation filters used for performing the global assimilation (Chap. 5) and a statistical method used for extracting dominant signal patterns (Sec. 6.1.2) in this thesis assume Gaussianity. Therefore, the Gaussian distribution, the Gamma distribution, and the log-logistic distribution will now be described.

The probability of a variable $X$ is given by the PDF function $f(x)$, which describes the chance that the realization of the variable $(x)$ takes a certain value. The integration of any PDFs up to the current realization $x$ yields in the CDF:

$$F(x) = \int_{-\infty}^{x} f(x')dx'. \tag{3.7}$$

Let us now first introduce the Gaussian distribution, which is a function with the shape of a symmetric bell, which computes the probability of a variable $x$ (e.g., TWSA) as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2}). \tag{3.8}$$

To derive the Gaussian probability values, the first two statistical moments of the variable $x$ are required, which are the mean or expected value

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx, \tag{3.9}$$

and the variance

$$\sigma^2 = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x)dx. \tag{3.10}$$

as soon as $\mu$ is zero and $\sigma$ is one, the Gaussian distribution is also called normal distribution. For Gaussian PDFs, the interval $\mu \pm \sigma$ covers about 68% of the realizations, the interval $\mu \pm 2\sigma$ covers about 95% and the interval $\mu \pm 3\sigma$ covers about 99%.

The PDF for the Gamma distribution is a two-parameter distribution function that is written as

$$f(x) = \frac{x^{\gamma-1} \exp^{\frac{-x}{\beta}}}{\Gamma(\gamma)\beta^{\gamma}}, \tag{3.11}$$

where $\gamma$ is the shape parameter, $\beta$ is a rate parameter $(\gamma, \beta > 0)$, and $\Gamma$ is the Gamma function. It is a positively skewed distribution for non-negative random variables (e.g., precipitation) that has a zero lower bound and is unlimited on the right (e.g., Thom, 1958). The mean and the variance of the Gamma distribution are computed as $\mu = \beta\gamma$ and $\sigma = \beta^2\gamma$, respectively.

At last, the log-logistic distribution is introduced, whose PDF is described in, for example, Tadikamalla (1980) via

$$f(x) = \frac{(\beta\alpha^{\beta}x^{\beta-1})}{(\alpha^{\beta} + x^{\beta})^2}, \tag{3.12}$$

where $\alpha$ is the scale parameter and $\beta$ is the shape parameter $(\alpha, \beta > 0)$. Thus, the log-logistic distribution is a two-parameter distribution. Similar to the Gamma distribution, the log-logistic distribution is used for non-negative random variables. The mean and variance of this distribution are computed by

$$\mu = \frac{\alpha\frac{\pi}{\beta}}{\sin\left(\frac{\pi}{\beta}\right)} \tag{3.13}$$

and

$$\sigma = \alpha^2 \left( \frac{2\frac{\pi}{\beta}}{\sin\left(2\frac{\pi}{\beta}\right)} - \frac{\frac{\pi}{\beta}^2}{\sin^2\left(\frac{\pi}{\beta}\right)} \right). \tag{3.14}$$

In this thesis, the log-logistic distribution is considered when the input data set is the difference of precipitation minus evapotranspiration.

## 3.4 Drought index computations

Drought indices can be used for the retrospective, for monitoring, and for forecasting. As with the definition of drought, a clear way to find suitable drought indices is not given in the literature, and many different drought indices exist, with no clear guidance on when to use which index (Herbert & Döll, 2023). Overall, drought indices determine the magnitude, severity, duration, and regional extent of a drought, thus quantifying a drought. The result that a drought index can assume is often categorized into severity classes, e.g., for triggering emergency measures such as the restriction of water use.

A widely discussed aspect of drought indices with no consensus is how they define the “normal” situation and compare the current situation against that (e.g., Wilhite & Glantz, 1985; Van Loon & Van Lanen, 2012). A local “normal” is restricted to the current situation and cannot just be transferred to other regions because conditions such as topography, vegetation, or climate regime influence the normal (e.g., Tallaksen & Van Lanen, 2004). To refer to the normal, various techniques have been used, for example, threshold, standardization, quantile, and percentile rank methods, including distribution functions, etc. (Palmer, 1965; McKee *et al.*, 1993; Thomas *et al.*, 2014; Herbert & Döll, 2023).

The magnitude of a meteorological, soil moisture, or hydrological drought is the precipitation or water storage deficit for a chosen month (referred to the chosen ’normal’), and drought severity is the accumulated drought magnitude for the drought duration. Threshold methods describe drought severity in an absolute quantitative metric because they accumulate deficit below a threshold, whereas standardized indices output the drought normalized magnitude that is inserted into drought severity classes per month. To derive the severity of a drought event over time for standardized methods, introducing an additional threshold is required. Therefore, severity classes per month should not be confused with the severity of a multi-month drought event.

Defining indices that join the drought monitoring for different drought types (e.g, meteorological and hydrological drought) can be very helpful to provide a generalized statement about water stress for the whole system. For example, Azmi *et al.* (2016) developed a composed index in three study regions of Australia, Hao & AghaKouchak (2014) combined a meteorological and a soil moisture index via copula modeling. The Moroccan Composite Drought Indicator (MCDI) was developed based on adding up and weighting four indices (Bijaber *et al.*, 2018). However, in this thesis I will not look at composite drought indices because the focus is on monitoring drought in several climate variables separately to detect a drought as early as possible, to track weaker and stronger drought phases, and also to derive important information about its course through the water cycle.

In the following sections, I do not go into detail about many drought indices because reviews and applications widely exist (e.g., Heim Jr, 2002; Keyantash & Dracup, 2002; Mishra & Singh, 2010; World Meteorological Organization & Global Water Partnership, 2016; Yihdego *et al.*, 2019). First, I present a choice of commonly used global meteorological and soil moisture indices. Then, I introduce global hydrological indices especially those that are used in combination with GRACE/-FO to focus on including subsurface water observations in the drought detection.

### 3.4.1 Meteorological and soil moisture drought indices

In the following, common drought indices that are used for monitoring meteorological drought and soil moisture drought are presented. Meteorological drought indices often rely solely on precipitation or on precipitation and temperature to include evaporation in drought monitoring. The most commonly used meteorological drought indices are the SPI and the Standardized Precipitation Evapotranspiration Index (SPEI).

The SPI is an index developed by McKee *et al.* (1993) and is widely used to detect meteorological and soil moisture drought. The index is based on monthly precipitation data and is able to calculate results on different time scales (typically 3, 6, 9, or 12 months). For example, the precipitation can be accumulated for each month and its previous five months, e.g. for SPI6. This enables us to relate precipitation deficits to a drought duration. At each location, a Gamma distribution is then fitted to the accumulated precipitation and transformed into a normal distribution by evaluating the inverse normal distribution at the corresponding probability. The output yields the probability of precipitation and is categorized into four classes: mild drought for values between -0.5 and -1, moderate drought between -1 and -1.5, severe drought between -1.5 to -2, and extreme drought for values lower than -2.

In many applications, the index is used to study meteorological drought in regions all over the world, for example, in Mongolia (Li *et al.*, 2014), Jordan (Mohammad *et al.*, 2018) and Sicily (Bonaccorso *et al.*, 2003). It can also be taken into account for monitoring soil moisture drought (and hydrological drought), but then they use longer accumulation periods. Thus, a strict classification of the SPI as an index for meteorological drought only is not correct. In this thesis, the SPI is used along with short accumulation time scales up to three months and thus refers to meteorological drought and soil moisture drought, but it should be remembered that it could also be used for monitoring hydrological drought with accumulation time scales of 6 or more months.

The SPEI was developed by Vicente-Serrano *et al.* (2010) to incorporate potential evapotranspiration into the methodology of the SPI. The calculation is similar to the SPI, but the SPEI uses temperature data additionally as input to compute the potential evapotranspiration via the Thornthwaite method (Thornthwaite, 1948). The difference between precipitation and potential evapotranspiration is then estimated. Using these differences, a probability distribution based on the log-logistic distribution is derived and transformed to a normal distribution for each location. The output is categorized into the same four severity classes as for the SPI. Since it considers more components of the water balance than the SPI, the index better covers drought in agricultural regions.

Many further indices for meteorological drought exist, ranging from very basic to more sophisticated approaches, for example, percentiles, the effective drought index (Byun & Wilhite, 1996), and the rainfall aridity index (Van Rooy, 1965) etc. The list could be extended but it should be kept in mind that the indices differ in their sensitivity towards signals contained in the input precipitation, e.g., temporal noise. As there are numerous meteorological drought indices adapted to certain study setups (e.g. the region), the World Meteorological Organization (WMO) provided an overall recommendation for the science community and policymakers of various countries for a single index and chose the SPI (Sivakumar *et al.*, 2011). For more details about various indices, I here refer to the above-mentioned reviews and collections.

In addition, further indices use precipitation and temperature data to compute soil moisture drought, for example, the Palmer Drought Severity Index (PDSI). The PDSI was introduced by Palmer (1965) and included the water balance under the consideration of demand and supply for determining soil moisture drought. It uses precipitation, temperature, and available water capacity as input data sets. In contrast to the precipitation-based indices, it is based on a period of nine months and thus cannot consider variable periods. According to Alley (1984), two of the largest limitations of PDSI are the limited performance for snow and frozen regions and that the limits for defining the severity classes are chosen arbitrarily: Extreme drought is defined below -4.0, severe drought has an upper limit of -3.0, moderate drought limit up to -2.0 and mild drought below -1.0.

As with PDSI, many more indices for soil moisture drought use precipitation and temperature (and available water capacity) as input data as compared to indices that are solely based on soil moisture (compare e.g., World Meteorological Organization & Global Water Partnership, 2016). This results from the limited availability of soil moisture data for many regions in the early years of index development. Meanwhile, remote sensing improved the availability of soil moisture data but the measurements only refer to the top few millimeters to centimeters of the soil instead of the root zone soil moisture. Two examples of indices based on soil moisture are the Standardized Soil Moisture Index (SSI) and the Soil Moisture Deficit Index (SMDI).

The SSI is transplanting the methodology of the precipitation-based SPI to soil moisture. Therefore, it fits a distribution to the soil moisture data (Hao & AghaKouchak, 2014) and transforms the data to a normal distribution so that the final index values can be compared across spatial regions under the consideration of various time scales (accumulation periods). In contrast to precipitation within the SPI, soil moisture requires a different underlying probability distribution. For example, Hao & AghaKouchak (2014) avoid choosing a probability distribution for soil moisture by proposing an empirical approach, where the marginal distribution is transformed to standardized values using the Gringorten plotting position.

The SMDI was developed by Narasimhan & Srinivasan (2005) and is computed under the consideration of median, minimum, and maximum soil water. It calculates deficits of soil moisture for four different soil depths and was initially computed by using root zone soil water model simulation from the Soil and Water Assessment Tool model. Since the index is developed for the different soil depths, the use of the index can be adapted according to the specific crops and thus, it can more precisely be classified as an index for agricultural drought. However, using all depths could include autocorrelations (compare e.g., World Meteorological Organization & Global Water Partnership, 2016). The use of the model simulations and thus their corresponding soil layer depths limits the comparison of this index to other models or observation types but one could still transfer the methodology. Further examples of soil moisture indices are the evapotranspiration deficit index (Narasimhan & Srinivasan, 2005), the crop moisture index (Palmer, 1968), and the soil moisture anomaly index (Bergman *et al.*, 1988) and can also be found in the previously mentioned collections.

According to Van Loon & Van Lanen (2012), meteorological drought indices should not be used for hydrological drought solely because the response of the system to climate events like drought is non-linear. The same is also valid for transplanting methodologies of meteorological indices to soil moisture indices. So, in the following, I focus on precipitation for meteorological indices and soil moisture for soil moisture indices. After summarizing all the above-mentioned advantages, limitations, and recommendations, the SPI is considered as an index for meteorological drought and the SSI are further considered for soil moisture drought.

### 3.4.2 Hydrological drought indices

Hydrological drought indices mostly refer to single hydrological storage or flux variables such as streamflow, runoff, surface waters, and groundwater. The choice of hydrological variable is very important because a drought might appear in a certain compartment or flow, but other compartments might be unaffected or appear to be less affected, which could lead to a misjudgment of the severity of the drought.

The main difference between runoff and streamflow is that runoff is the flow of water on surfaces or that has infiltrated the ground, whereas streamflow is the flow that is channeled in rivers. Runoff is often derived from model simulations, whereas streamflow can be simulated or measured from gauge stations. Hydrological drought indices based on these fluxes are widely distributed: Modarres (2007) and Telesca *et al.* (2012) developed the standardized streamflow indicator and Shukla & Wood (2008) the standardized runoff index, which are underlying the same methodology as SPI and exchange the precipitation by the corresponding flux. But there is no clear consensus on which underlying probability distribution should be used (e.g., Barker *et al.*, 2016). Further commonly used methods for streamflow indices are threshold-level methods (typical thresholds are the $50^{th}$ and $80^{th}$ percentile as e.g., in Herbert & Döll, 2023).

Herbert & Döll (2023) summarize and review a number of streamflow indices (Dracup *et al.*, 1980). They underline that – as with other variables of the hydrological water cycle – there is no universal streamflow index. The choice of the index strongly depends on the regional hydrological and socio-economic characteristics and the choice of, for example, the preferred threshold for the cumulative indices is always subjective.

The presented indices have been implemented very often because local streamflow observations are easy to access or long-term simulations of streamflow are easy to implement. However, a few decades ago there existed no hydrological variable that was globally observed with a spatially regular coverage that covers surface and subsurface water storage. As highlighted by e.g., Mishra & Singh (2010), especially subsurface drought in groundwater has diverse impacts: Not only lower groundwater heads and a decrease in groundwater flow to riparian areas occurs, but also a decrease in capillary rinse might reduce water storage in wetlands, lakes, and soil, which in turn could affect yield production. With the launch of GRACE/-FO, a new observable was provided that would complement in-situ surface and subsurface observations and might thus alleviate existing limitations for hydrological drought monitoring.

Two examples of indices that were not directly related to GRACE/-FO but consider surface and subsurface water storages are the standardized reservoir supply index (Gusyev *et al.*, 2015) and the standardized water-level index (Bhuiyan, 2004), both approaches are again based on the SPI calculation that uses distribution functions, using reservoir inflows and average reservoir storage volumes or groundwater well data in India, respectively, to analyze groundwater recharge. As GRACE/-FO offer another unique observable for monitoring hydrological drought, the subsequent section gives an overview of GRACE/-FO drought monitoring and provides more detailed equations for a choice of indices.

### 3.4.3 Drought indices from GRACE/-FO

To observe drought with GRACE/-FO, several studies focused on removing the seasonal cycle from TWSA and tried to identify extreme events within the non-seasonal residual signal, for example in Frappart *et al.* (2012) in the Amazon basin during the 2005 drought or in Abelen *et al.* (2015) in the La Plata basin. One important aspect is missing when using the de-seasoned signal: Results from different regions cannot be compared, because the magnitude of the non-seasonal residual signal or total values varies from region to region (Wanders *et al.*, 2010) and there is no clear distinction when the drought starts, ends, and about the severity of a drought.

The Total Storage Deficit Index (TSDI) was one of the first drought indices that utilized GRACE TWSA and was introduced by Yirdaw *et al.* (2008) as a regional index and applied in the Canadian Prairie. The index used the methodology of the SMDI (Sec. 3.4.1), where the mean, minimum, and maximum values per calendar month are used to remove seasonality, and two parameters were estimated to calibrate the index for one basin. TSDI in its original form by Yirdaw *et al.* (2008) cannot be directly used globally on the grid level since it is adapted to a basin. One could calculate such calibration factors for multiple hydrological basins but then the spatial extent of the index is only available for grids inside the available polygons that were chosen to compute the calibration factors instead of deriving a global continuous grid. Later on, a simple technique for computing the TSDI on a global scale was developed by Wanders *et al.* (2010) by including a standardization.

In 2014, Thomas *et al.* (2014) computed drought magnitude, duration, and severity by calculating GRACE TWSA deficits when there were a predetermined number of consecutive months below a certain threshold. This deficit approach provided helpful information about drought magnitude, severity, and duration of a drought. Since the deficit and severity are given in total numbers, they cannot be compared across regions. Details about the performance of Thomas *et al.* (2014) are contained in Gerdener *et al.* (2020a) and will not be repeated here. A disadvantage of the threshold methods is that no drought severity classes are calculated. Thus, in global studies for drought detection standardization is a better choice to enable better comparison (Van Loon *et al.*, 2014), and subjective choices about setting the thresholds are avoided. The GRACE Groundwater Drought Index (GGDI) by Thomas *et al.* (2017) extended the methodology. A climatology is removed from the data and the output is normalized, which provides spatial comparability.

Yi & Wen (2016) presented the GRACE-based Hydrological Drought index. In the first step, GRACE non-seasonal residual signals are computed for this index by removing the climatology (Eq. 3.17). Then, the de-seasoned signals are multiplied with a converting factor, which is computed by calibrating an empirical relationship between GRACE TWSA and long-term soil moisture storage. Thus, the index requires soil moisture data for the calibration process and is not as readily implementable as other previously mentioned indices for GRACE/-FO. Furthermore, Frappart *et al.* (2013) computes extrema of water storage per year (minimum or maximum) and standardizes the yearly extrema by removing the mean of all yearly extrema and dividing by the standard deviation of the extrema. In fact, they derive one index value per year and do not make use of monthly values.

Two very basic and easily understandable indices are the Drought Indicator (DI) developed by Houborg *et al.* (2012) and the Drought Severity Index (DSI) by Zhao *et al.* (2017). The DI used a percentile rank method and the DSI a simple standardization – both under the consideration of the respective calendar month – to determine a monthly index. Their ease of use makes them a good candidate for global application.

To summarize, the DI, DSI, TSDI, and the GGDI are described in more detail in the following sections (based on Gerdener *et al.*, 2020a) because the other indices are calibrated to a certain basin or region characteristics, are not comparable across regions, are not considering monthly values, or include other types of data sets.

GRACE-derived monthly gridded TWSA for $n$ years is referred to as

$$x_{i,j} = \overline{\Delta s}(t_{i,j}) \tag{3.15}$$

with

$$t_{i,j} = i + \left(j - \frac{1}{12}\right)\frac{1}{12} \qquad i = 1, \ldots, n \qquad j = 1, \ldots, 12\ . \tag{3.16}$$

The temporal mean for the respective calendar month $j$ can also be denoted as climatology and is computed as

$$\tilde{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j} \tag{3.17}$$

with $j$ running from January to December. Correspondingly, the monthly standard deviation yields in

$$\tilde{\sigma}_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_{i,j} - \tilde{x}_j)^2}, \tag{3.18}$$

and the minimum and maximum of the respective month $j$ are denoted as $x_j^{\min}$ and $x_j^{\max}$.

**Drought Severity Index (DSI)**

The Drought Severity Index (DSI) was developed by Zhao *et al.* (2017). For each grid cell, it uses the climatology as represented in Eq. 3.17 for a specific calendar month, for example, January, and the corresponding standard deviation for that month (Eq. 3.18) to derive a standardization of the GRACE/-FO TWSA by

$$\mathrm{DSI}_{i,j} = \frac{x_{i,j} - \tilde{x}_j}{\tilde{\sigma}_j}, \tag{3.19}$$

i.e. it represents standardized anomalies. The resulting index values are classified into drought severity classes from -2.0 (exceptional drought) to +2.0 (exceptionally wet) as shown in Tab. 3.1. Assuming that the TWSA follow a Gaussian distribution, the severity classes provide the probability of a certain month according to the probability distribution function. For example, 2.1% of the months would be exceptionally dry and in the same manner, 2.1% of the months would be exceptionally wet. One could also assume a different distribution underlying the droughts, which would change the percentage.

In Zhao *et al.* (2017), a bias correction is suggested to account for the short GRACE period. For that study, the period was from mid-2002 to 2014, which is about four and a half years shorter than what is considered in this thesis. Thus, firstly, this thesis does not include a bias correction to see how the indices perform under current conditions. In addition, the effect of the length of the TWSA time series on the drought monitoring is investigated later on.

To relate the DSI to drought duration as McKee *et al.* (1993) also did for the SPI, the accumulation (Eq. 3.5) or differences (Eq. 3.6) of TWSA are included into the methodology and yield in the DSIA and DSID, respectively, as

$$\mathrm{DSIA}_{i,j,q} = \frac{x^{+}_{i,j,q} - \tilde{x}^{+}_{j,q}}{\tilde{\sigma}^{+}_{j,q}} \tag{3.20}$$

$$\mathrm{DSID}_{i,j,q} = \frac{x^{-}_{i,j,q} - \tilde{x}^{-}_{j,q}}{\tilde{\sigma}^{-}_{j,q}}. \tag{3.21}$$

This technique is exemplarily shown for the DSI, but the same can be transferred to the other GRACE/-FO indices resulting in the DIA, DID, TSDIA, TSDID, GGDIA, and GGDID. Furthermore, the use of the indices is not restricted to TWSA only throughout the thesis. Instead, I also consider surface water, soil moisture, and groundwater from data assimilation for its computation. For the notation, TWSA is consistently used, for example for Zhao *et al.* (2017) as TWSA DSI, but in subsequent chapters the corresponding name of the used storage is included, i.e., surface water DSI for surface water, soil moisture DSI, and groundwater DSI. This is also valid for the three GRACE/-FO indices hereinafter.

Table 3.1: Drought severity levels of the DSI as described in Zhao *et al.* (2017). The values of DSI are unitless and its probability of occurrence is given in percentage (assuming a Gaussian distribution). The table is based on Gerdener *et al.* (2020a).

| | **DSI [-]** | | |
|---|---|---|---|
| **Drought Severity Level** | **Min.** | **Max.** | **Probability (Gaussian)** |
| Abnormal dry | $-0.8$ | $-0.5$ | 9.7% |
| Moderate dry | $-1.3$ | $-0.8$ | 11.5% |
| Severe dry | $-1.6$ | $-1.3$ | 4.2% |
| Extreme dry | $-2.0$ | $-1.6$ | 3.4% |
| Exceptional dry | | $-2.0$ | 2.1% |

Table 3.2: Drought severity level of the DI (Houborg *et al.*, 2012). The values of DI are given in percent. The table is based on Gerdener *et al.* (2020a).

| | **DI [%]** | |
|---|---|---|
| **Drought Severity Level** | **Min.** | **Max.** |
| Abnormal | 20 | 30 |
| Moderate | 10 | 20 |
| Severe | 5 | 10 |
| Extreme | 2 | 5 |
| Exceptional | 0 | 2 |

### Drought Indicator (DI)

The Drought Indicator (DI) by Houborg *et al.* (2012) involves the implementation of a percentile rank method. In fact, the number of months below or equal to the current-month TWSA are normalized by the total number of values for that calendar month, which is an empirical way to derive a CDF:

$$\mathrm{DI}_{i,j} = \frac{\sum_i (x_j \leq x_{i,j})}{\sum_i x_j} \cdot 100. \tag{3.22}$$

In fact, $\sum_i x_j$ is the number of all years with month $j$ and $\sum_i (x_j \leq x_{i,j})$ counts only those years, where the TWSA is lower than or equal to the TWSA of the considered month and year. The percentiles are equal to the probability with which a drought event occurs and were chosen similarly as for the U.S. Drought Monitor, for example, a percentile of lower than 2% means that this value occurs only 2% or lower of the time for the specific month. The resulting drought magnitude values are divided into drought severity classes Tab. 3.2. The severity classes here are not chosen to be set under the consideration of a Gaussian distribution but if this would be the case 0.6% of months would be detected as exceptionally dry. In addition, Houborg *et al.* (2012) adjust the CDFs to a long-term model run to account for a bias correction. This would also be feasible in this thesis but to concentrate on the GRACE/-FO data or the outputs from data assimilation, the bias adjustment is not accounted for.

### Total Storage Deficit Index (TSDI)

The Total Storage Deficit Index (TSDI) is an index whose most recent development included a two-step standardization procedure for the computation (e.g., Wanders *et al.*, 2010; Nie *et al.*, 2018). At first, the total storage deficit in percent is computed by weighting de-seasoned

Table 3.3: Drought severity levels of the TSDI as described in Nie *et al.* (2018). The values of TSDI are unitless and its probability of occurrence is given in percentage (assuming a Gaussian distribution). The table is based on the SPI severity classes as shown in McKee *et al.* (1993).

| | **TDSI [-]** | | |
|---|---|---|---|
| **Drought Severity Level** | **Min.** | **Max.** | **Probability (Gaussian)** |
| Mild dry | −1.0 | 0 | 34% |
| Moderate dry | −1.5 | −1.0 | 9.2% |
| Severe dry | −2.0 | −1.5 | 4.4% |
| Extreme dry | | −2.0 | 2.3% |

TWSA (derived by removing the monthly climatology) in a specific month by the range for the specific month

$$\mathrm{TSD}_{i,j} = \frac{x_{i,j} - \tilde{x}_j}{x_j^{max} - x_j^{min}} \cdot 100, \tag{3.23}$$

where $i$ is the year and $j$ is the calendar month. Originally, the index was implemented as a regional drought index in Yirdaw *et al.* (2008) that combines the TSD now with the previous-month TSDI and two tuning parameters estimated from a drought monograph. Instead, Wanders *et al.* (2010) included a second standardization step to provide a globally applicable drought index. Thus, the TSD values are normalized via the TSD mean and standard deviation. This means a mean is removed, and the result is normalized by the standard deviation:

$$\mathrm{TSDI}_{i,j} = \frac{\mathrm{TSD}_{i,j} - \mu_{\mathrm{TSD}}}{\sigma_{\mathrm{TSD}}}. \tag{3.24}$$

It is important here to understand that the mean and standard deviation are computed from the complete time series of TSD values, which is different from the proposed mean in Eq. 3.17 from the corresponding month. Nie *et al.* (2018) assume normal distribution and thus use the drought severity classes of the SPI analogously for the TSDI.

#### GRACE Groundwater Drought Index (GGDI)

Thomas *et al.* (2014) introduced a threshold method to monitor drought as observed by GRACE/-FO. The method considers several consecutive months below a certain TWSA threshold. In its original form, the approach uses a climatology (Eq. 3.17), to determine the threshold, but one could also think about other thresholds, for example, a fitted seasonal signal (Eq. 6.12). The anomalies with respect to this threshold are denoted as water deficit or drought magnitude (Eq. 3.25)

$$\Delta x_{i,j} = \begin{cases} 0 & \text{for } x_{i,j} \geq c \\ x_{i,j} - \tilde{x}_j & \text{for } x_{i,j} > c. \end{cases}, \tag{3.25}$$

In other words, the deficits are a measure of how much water is required to return to the normal condition. As soon as the deficit lasts several consecutive months, this is recorded as drought duration $d_{i,j}$. In Thomas *et al.* (2014), a minimum duration of three months is necessary. In the final step, the drought severity $s_{i,j}$ is determined by combining drought average magnitude with drought duration by $s_{i,j} = \Delta x_{i,j} d_{i,j}$. Different from with DSI, DI, and TSDI the methodology by Thomas *et al.* (2014) combines the amount of accumulated monthly water storage deficit with its duration instead of inserting normalized values into drought severity classes. This non-standardized approach is not directly comparable to other drought indices and also not comparable for different regions but it has the advantage, that drought severity is not a standardized measure but concretely provides the water volume deficit.

To make this approach more comparable, Thomas *et al.* (2017) presented an extension to Thomas *et al.* (2014) for the particular case of groundwater storage, which is calculated from GRACE/-FO TWSA and other fluxes. Therefore, the index name is the GRACE Groundwater Drought Index (GGDI), however, the underlying approach can also be applied to TWSA. At first, water deficits and surpluses are similarly calculated as in Thomas *et al.* (2014) by removing the climatology. Then, for a global comparison, these deviations are standardized by removing the mean and standard deviation, computed from all monthly values as similarly done for the TSDI (Eq. 3.24):

$$\mathrm{GGDI}_{i,j} = \frac{\mathrm{GSD}_{i,j} - \mu_{\mathrm{GSD}}}{\sigma_{\mathrm{GSD}}}, \tag{3.26}$$

where GSD is the groundwater deficit, similarly achieved as for Thomas *et al.* (2014) (Eq. 3.25). Unfortunately, the study does not provide any drought severity classes, but since the standardization is exactly equal to the TSDI, the severity classes from Tab. 3.3 are considered in this thesis.

## 3.5 Drought risk and drought hazard risk

Drought risk is the likelihood of losses or probability of hazardous events to affect a specific region in a certain periodic interval IPCC (2014) and is determined by considering hazard, exposure, and vulnerability (e.g., Cardona *et al.*, 2012). Hazard is referred to as a natural phenomenon, implying that flux or storage has lower water than normal. The exposure generally describes different types of physical entities on the ground, which are, for example, the people, livelihoods, ecosystems, and infrastructures that could be affected (Peduzzi *et al.*, 2009; IPCC, 2014). The vulnerability describes developmental factors of the society to adapt against drought hazards, in other words, the predisposition to be affected and a lack of coping capacity. These are, for example, the health of the population or governmental decisions. As with the exposure, the changes in vulnerability that occur in response to drought hazard occur with a temporal delay (e.g., IPCC, 2014; Meza *et al.*, 2020). In summary, drought risk depends not only on the probability of physical hazard but also on how vulnerable society is at a specific location and point in time (Carrão *et al.*, 2016). The manifestation of drought risk is the drought impact, which often has disastrous consequences. To increase a society's protection against drought impacts, it is important to estimate drought risk more precisely, and consequently to improve estimates for each of the three components hazard, exposure, and vulnerability.

In contrast to temporally dynamical drought monitoring systems (Sec. 3.6), drought risk (R) is often determined as static output computed from the combination of static maps of hazard (H), vulnerability (V), and exposure (E) that are combined via

$$\mathrm{R} = \mathrm{H} \cdot \mathrm{E} \cdot \mathrm{V}. \tag{3.27}$$

Since static maps are provided, the risk assessments can be seen as a retrospecitve drought assessment. Typically, the final risk map provides values between zero and one, where zero is no risk and one is the highest risk. For combining hazard, exposure, and vulnerability, it is required that each map is standardized. To focus on the hazard component as observed by remote sensing data and hydrological model simulations, the drought hazard risk is further described in detail, whereas more information about exposure and vulnerably can be found in some studies (e.g., Carrão *et al.*, 2016; Meza *et al.*, 2020).

Drought hazard risk can be seen as the drought frequency of a natural event with a certain drought severity for a chosen time interval (return period). The previously mentioned drought hazard indices (and others) are widely used for drought hazard risk computation because they provide information on the timing and severity of a drought. For example, McKee *et al.* (1993)

defined a drought event as soon as the SPI is continuously negative and reaches a value of -1 or less. The beginning of the drought is set as soon as the SPI falls below zero. The drought is considered as terminated as soon as the SPI gets positive. Then, the frequency of drought yields from the $N =$ number of droughts per 100 years.

Multiple studies make use of this approach, among others, Bazrafshan *et al.* (2020) for determining drought frequency over Iran or Loukas & Vasiliades (2004) in Greece. Other studies use threshold indices like for example Herbert & Döll (2023), who relate (among other techniques) the lowest 20% of an index per calendar month to a return period of five years. Carrão *et al.* (2014) suggests using the Fisher-Jenks algorithm that tries to find an optimal threshold between drought events and non-drought events. Similarly, one could identify drought periods by using segmented regression which is comparable to the breakpoint algorithm (Schwarz *et al.*, 2020). Since the standardized drought hazard indices that were previously introduced are all easily usable for the determination of drought frequency without introducing extra algorithms and at the same time can be compared for different drought types, the same methodology as in McKee *et al.* (1993) of defining a continuous drought event is used in further context.

As focused in many studies, the drought risk indices should always be developed according to the considered risk source (e.g., Lloyd-Hughes, 2014; Hagenlocher *et al.*, 2019). For example, streamflow or water storage in the upstream reservoir should be used in the case of a river as a water supply source (Herbert & Döll, 2023). A global drought risk map that combines hazard, vulnerability, and exposure can be found in, e.g., Carrão *et al.* (2016), which was the first global assessment for drought risk, but distinguishing between different drought hazard indices was not included. To more precisely monitor the affected systems Meza *et al.* (2020) distinguished between rainfed and irrigated agriculture for global drought risk on national level.

## 3.6 Drought information systems

Over the past decades, efforts have been made to assess drought risk at different spatial scales and incorporate it into dynamic drought monitoring systems operating in near-real time. Dynamical systems that operationally provide monthly or daily forecasts or monitoring maps to enable managing of drought exist on regional scales for multiple regions.

A well-known system is the North American Drought Monitor[1], which provides near-real-time drought information for Canada and Mexico and incorporates the U.S. Drought Monitor[2] via weekly updates that use various indices. Among others, this includes indices based on precipitation, soil moisture, and outputs from assimilating GRACE/-FO into hydrological models. A European system was set up with the European Drought Observatory (EDO)[3]. It uses a combination of measured precipitation, modeled soil moisture, and a vegetation variable and builds a stepping warning system 'watch-warning-alert'. In addition, the EDO also provides information on GRACE but no outputs of GRACE/-FO assimilation are integrated or provided. A German system offers daily drought information[4] using model simulations for the soil down to 1.8 m depth. Within the South Asia Drought Monitoring System[5] multiple drought indices are used ranging from integrated drought severity index (based on vegetation, temperature, and rainfall), SPI, and a soil moisture index, but the system does not use GRACE/-FO for drought

[1] `https://www.drought.gov/nadm/` (last accessed 24.03.2024)
[2] `https://droughtmonitor.unl.edu/CurrentMap.aspx` (last accessed 24.03.2024)
[3] `http://edo.jrc.ec.europa.eu/edov2/php/index.php?id=1000` (last accessed 24.03.2024)
[4] `https://www.ufz.de/index.php?en=37937` (last accessed 24.03.2024)
[5] `http://dms.iwmi.org/` (last accessed 24.03.2024)

monitoring, only for validation of a model (Saha *et al.*, 2021). The Australian Drought Monitor was developed from the Northern Australia Climate Program[6] where they use combined drought indices based on the concept of the U.S. Drought Monitor composed of SPI, soil moisture, evapotranspiration, and the Normalized Difference Vegetation Index (NDVI). The African Flood and Drought Monitor[7] was developed by the Princeton Climate Institute, University of Southampton and Princeton University and uses multiple satellite observations model simulations. Climate variables of precipitation, evapotranspiration, streamflow, and soil moisture are incorporated but deeper layer storage information is not integrated.

In the last decade, an increasing importance towards global systems was observed. Very simple systems based on a single drought index existed early, for example, the global real-time drought monitoring APCC Global Drought Monitoring[8] based on SPI or the SPEI Global Drought Monitor[9]. A more complex system is the Global Drought Observatory (GDO)[10] by the European Commission, which uses SPI, soil moisture maps and also includes GRACE/-FO, similarly as for the European system. They also provide combined hazard indices.

The described systems highlight that GRACE/-FO are currently increasingly included in operational monitoring systems. A few regional systems already include GRACE/-FO observations of outputs from assimilating GRACE/-FO into a hydrological model, nonetheless to my knowledge, the global operational drought-monitoring system (Global Drought Information Systems) [11], is the only global system that ingests information from GRACE assimilated into models. The system operationally integrates the root zone soil moisture, shallow groundwater, and surface soil moisture derived from assimilation at NASAs Goddard Space Flight Center[12]. However, the system does not look at temporally subsequent drought events in different compartments of the water cycle (time-depending warning system) as is the case for the GDO, which has a combined drought index that considers temporal delays between precipitation and soil moisture and vegetation on a global scale but does not include GRACE or GRACE/-FO assimilation products in this temporal analysis.

All in all, there is currently no global system that has a more in-depth look into subsequent drought events in the storages and drought hazard risk computation by using GRACE/-FO and assimilation outputs. Therefore, in this thesis, a detailed framework is presented for globally monitoring drought, computing drought hazard risk, and observing subsequent droughts with focusing on assimilation products.

[6] https://www.nacp.org.au/drought_monitor (last accessed 24.03.2024)
[7] https://www.unccd.int/land-and-life/drought/toolbox/african-flood-and-drought-monitor (last accessed 24.03.2024)
[8] https://www.apcc21.org/ser/global.do?lang=en (last accessed 24.03.2024)
[9] https://spei.csic.es/map (last accessed 24.03.2024)
[10] https://edo.jrc.ec.europa.eu/gdo/php/index.php?id=2001 (last accessed 24.03.2024)
[11] https://gdis-noaa.hub.arcgis.com/ (last accessed 24.03.2024)
[12] nasagrace.unl.edu (last accessed 24.03.2024)

# Chapter 4

# Hydrological modeling

## 4.1 Overview

Hydrological models aim to simulate the water storages and water flow to provide new insights into the water cycle. With the help of the models, the complex temporal evolution and spatial propagation of water via vertical and horizontal water flows can be understood and used for monitoring, near-real-time analysis, or forecasting. The knowledge derived from the simulations enables the management of water resources.

Generally, the setup of such models differs in many aspects. For example, models can be developed based on conceptual strategies or on physical relations. Conceptual models were developed for water management and rely on empirical equations and parameters. Instead, physical models are based on physical processes and make use of complex mathematical equations. The probably most obvious difference between different hydrological models is the spatial extent, which can vary from a very fine local scale (100 m and below) to a regional scale up to a coarse scale global model (about 50 km).

The local and regional models enable a more detailed insight into water processes and have high spatial resolutions but a general overview of large-scale hydrological processes is only possible with larger-scale or global models. The models – here with a focus on the global models – can be further distinguished into the two categories of Land Surface Models (LSMs) and Global Hydrology Models (GHMs).

GHMs were developed with the main purpose of simulating water storage and fluxes under the consideration of water balances. In contrast, the LSMs main purpose is simulating the energy balance at the surface. The LSMs focus on the surface and soil, thus, many LSMs have no or a limited representation of surface water flow, groundwater representation, and subsurface water bodies. A strict classification into these categories is difficult, therefore, in the following both LSMs and GHMs are referred to as GHMs only as it is also the case in other studies (e.g., Hagemann *et al.*, 2013; Döll *et al.*, 2016). An overview of GHMs can be found, for example, for 12 GHMs in Sood & Smakhtin (2015) and a comparison and evaluation for seven GHMs can be found in Scanlon *et al.* (2018).

One of the early GHMs was, among others, the Water Balance Model - Water Transport Model by Vörösmarty *et al.* (1989), which provides a link of simulating soil moisture, evapotranspiration, and surface water together with routing discharge via a linear reservoir model. The Global Water Availability Assessment model (Meigh *et al.*, 1999) simulated soil moisture, surface water in lakes, wetlands and glaciers, and groundwater using a gridded probability distribution model. Both models represent soil as a single layer. Nowadays, a large number of GHMs exist, for example, to name only a few popular models, the Catchment Land Surface Model (CLSM), the Water Global Assessment and Prognosis (WaterGAP) model, and the models contained in the Global Land Data Assimilation System (GLDAS).

CLSM aims at representing land-atmosphere interactions and has two soil layers, a surface zone and root zone (0-2 cm and 0-100 cm) layer, and simulates groundwater indirectly via catchment deficit. The CLSM does not consider groundwater withdrawals and, as mentioned in Li *et al.* (2019), was never calibrated. The CLSM model was also integrated into GLDAS. GLDAS (Rodell *et al.*, 2004) combines several offline land surface models: the CLSM, CLM, Noah, Mosaic, and VIC. The CLM is a community-based model (e.g., Dai *et al.*, 2003; Lawrence *et al.*, 2019) with a focus on ecological climatology that simulates water storage for soil moisture, surface water, and groundwater. The Mosaic model was developed by using subgrid heterogeneity of the land surface, the so-called mosaic strategy that separates each grid into a mosaic of tiles depending on the underlying vegetation and simulates surface water fluxes as well as soil moisture storage (Koster & Suarez, 1992). Noah is a model used for operational purposes and is continuously improved (Ek *et al.*, 2003) with a focus on soil moisture storage. VIC can be run as an energy and water balance model or water balance model only (Liang *et al.*, 1994). As with Mosaic and Noah, the VIC model storage simulations focus on the soil moisture.

Tab. 1 in Scanlon *et al.* (2018) summarizes the structure of models. All models mentioned earlier have a representation of soil via one or several layers, but surface water (e.g., lakes, rivers and wetlands) and groundwater are not always explicitly included as LSMs focus on the land-atmosphere interaction. For example, surface water is not directly included in CLSM but has an explicit representation of groundwater storage. The global freshwater model WaterGAP conceptually includes layers for soil moisture, surface water, and groundwater. It is one of only a few global models worldwide that quantify the anthropogenic water use of groundwater and surface water. In earlier successful applications it was shown that the model can be linked to a GRACE/-FO assimilation framework. Due to these reasons, I use WaterGAP in the further context of this thesis.

## 4.2 WaterGAP

The global freshwater model WaterGAP simulates daily global water storages and fluxes together with consumptive water use and water withdrawals on an 0.5° grid (approximately 55 km at the equator) covering the global continental area excluding Antarctica (Müller Schmied *et al.*, 2021). A special focus was laid on integrating human water use and man-made reservoirs. WaterGAP has a long version history, its development started in 1996 and is still ongoing with steady improvements, for example, in 2019, including groundwater simulation was further developed by testing new gradient-based groundwater models (Reinecke *et al.*, 2019), which is still under development. The model improved our knowledge about surface and sub-surface hydrological processes all over the world in the last decades. For example, WaterGAP contributed to analyzing droughts under future climate change and retrospective drought risk assessment (e.g., Meza *et al.*, 2020; Satoh *et al.*, 2022).

The main structural components of WaterGAP are the global water use models, the linking model Groundwater and Surface Water USE (GWSWUSE) use, and the WaterGAP Global Hydrology Model (WGHM) as shown in Fig. 4.1. In the global water use models and the linking model, anthropogenic water use is derived (Sec. 4.3) and further transmitted to WGHM. WGHM then solves for the vertical and horizontal water balance equations (Sec. 4.4). With its 0.5° spatial resolution, WaterGAP has a total number of 67420 grid cells. A cell in WaterGAP considers the continental area, i.e. the cell area minus the ocean area. The continental area is further separated

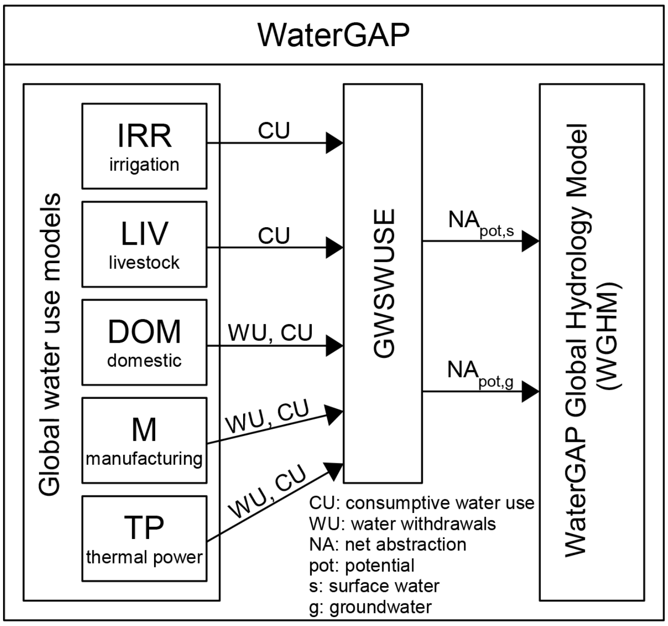


Figure 4.1: Model structure of WaterGAP. Figure taken from Müller Schmied *et al.* (2021).

into land area and area of the surface water bodies (lakes, reservoirs, and wetlands), which are updated for each time step. In addition, the model includes vegetation metrics (Sec. 4.7) in the simulation to improve the modeling of water storages and fluxes.

The model requires meteorological input forcing data, which means precipitation, temperature, and shortwave and longwave radiation (Sec. 4.6). To reduce model uncertainty in model simulations, common approaches calibrate the models against observation (e.g., Döll *et al.*, 2003). With this technique, the model parameters are adjusted so that the simulation is closer to the observations. The most commonly used variable for parameter calibration is streamflow but the community also started to use other data sets for calibration, for example, against TWS observations (Trautmann *et al.*, 2023). The WaterGAP standard configuration is set up together with a calibration against long-term annual river discharge for about 1320 large drainage basins (e.g., Müller Schmied *et al.*, 2021) for reducing model uncertainties.

Due to the uncertainty of the forcing data, model assumptions, and simplifications, the model is still under development. For example, glaciers were included by calculating glacier runoff with glacial mass change and precipitation from external models but are still not represented as storage in the model and it is planned to integrate a groundwater-gradient model in the future (Reinecke *et al.*, 2019; Müller Schmied *et al.*, 2021). Studies showed that WaterGAP long-term trends are underestimated compared to observations (Scanlon *et al.*, 2018). Here, I use the WaterGAP2.2e version (Müller Schmied *et al.*, 2023). The main improvements of the WaterGAP2.2e version to WaterGAP2.2d are an improved water abstraction algorithm and handling of inland sinks and small man-made reservoirs. There is also a WaterGAP3 version, which operates on 5' but this version is still under development and not coupled with assimilation frameworks yet.

## 4.3 Global water use models

To improve the WaterGAP model simulation with respect to anthropogenic water use, WGHM is linked to global water use models for flux and storage computation by deriving water use

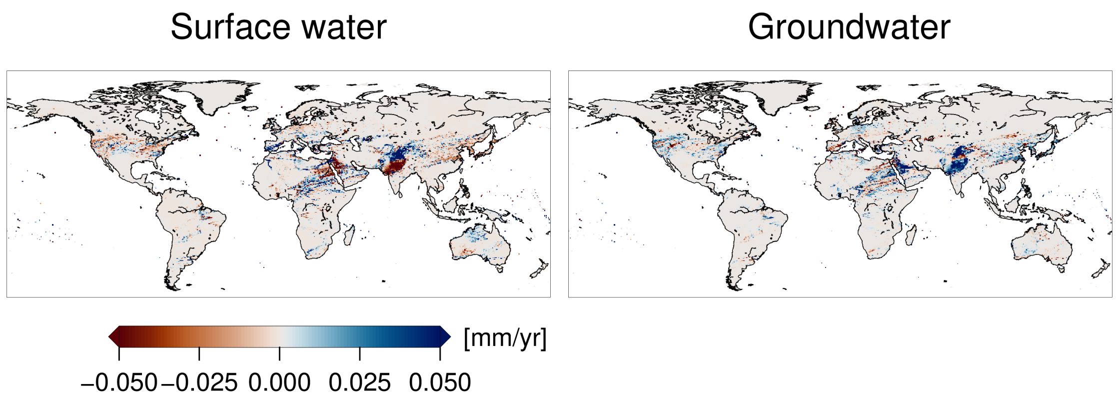


Figure 4.2: Linear trends [mm/year] for WGHM derived from GWSWUSE of net surface water abstraction (left) and net groundwater abstraction (right).

from surface water and groundwater (Müller Schmied *et al.*, 2021). The global water use models output consumptive and withdrawal water use derived from the five sectors irrigation (Döll & Siebert, 2002; Portmann, 2017), livestock, domestic, manufacturing, and cooling of thermal power plants (Flörke *et al.*, 2013), where irrigation is the strongest reason for withdrawal and consumptive water uses worldwide (Döll *et al.*, 2014). More information about how the water use models compute water abstractions and withdrawal, and consumptive uses for the five sectors can be found in Müller Schmied *et al.* (2021).

The linking model Groundwater and Surface Water USE (GWSWUSE) then reads the withdrawal and consumptive uses from the five models. An exception is the withdrawal use for irrigation, which is internally calculated in the GWSWUSE. The linking model then differentiates between surface water and groundwater use and produces the net abstraction from these storages by considering water use efficiencies for irrigation and fractions for groundwater for all sectors. Net abstraction is calculated as the difference between total water abstraction from one of the two sources and the return flow to it (Müller Schmied *et al.*, 2021).

Fig. 4.2 shows the global net abstractions of surface water (left) or groundwater (right) as provided from GWSWUSE. Positive trends indicate that water is anthropogenically abstracted in the long term, whereas negative trends indicate that water is again added back to the surface water storage or groundwater. The largest trends of net groundwater abstractions can be found in India. India has a large population with a strong dependency on the agricultural sector, which requires a large amount of drinking water extracted from groundwater and irrigation water for crop production. At the same time, the net surface water abstraction shows negative trends, which means that much water is added (Müller Schmied *et al.*, 2021). I hypothesize here that the extracted groundwater used for irrigation might first lead to an increased soil moisture and then runs off to surface water bodies.

## 4.4 Water balances within WGHM

The WaterGAP Global Hydrology Model (WGHM) simulates the fluxes and storages on a daily basis. The water storages are represented by 10 compartments: soil, snow, canopy, reservoir, river, groundwater, local and global lake, and local and global wetland. When the ten compartmens are aggregated, TWS can be derived from the model. For the three compartments canopy, soil, and snow, a vertical water balance equation is used to describe the vertical water routing of the corresponding compartment. The change in canopy storage $\frac{dS_c}{dt}$ is described via

$$\frac{dS_c}{dt} = P - P_t - E_c, \tag{4.1}$$

where canopy change is calculated from the input precipitation $P$ from the forcing data, the precipitation throughfall $P_t$, and the evaporation from the canopy $E_c$. The change in snow storage $\frac{dS_{sn}}{dt}$ results from

$$\frac{dS_{sn}}{dt} = P_{sn} - M - E_{sn}, \tag{4.2}$$

with the input throughfall snow $P_{sn}$, the snowmelt $M$, and the sublimation $E_{sn}$ and correspondingly, the change in soil storage $\frac{dS_s}{dt}$ computes from the input effective precipitation $P_{eff}$, the land runoff $R_l$, and the evaporation from the soil $E_s$:

$$\frac{dS_s}{dt} = P_{eff} - R_l - E_s. \tag{4.3}$$

As will be explained in Sec. 4.5, the soil moisture storage is defined in the effective root zone, whose depth in turn depends on the land cover of the location. Müller Schmied *et al.* (2014) provide a table that shows which land cover has which rooting depth. For example, for savanna, the rooting depth is 1.5 m. The smallest rooting depth is given for bareground with 0.1 m and the largest rooting depth is given for evergreen broadleaf forest with 4 m. This feature of the WaterGAP soil moisture depth differs from, e.g., remote-sensed soil moisture that is typically measured in the first few centimeters of the soil. Thus, the soil moisture storage in WaterGAP simulates soil moisture in a deep layer.

The water compartments of the global and local lakes, global and local wetlands, rivers, reservoirs, and groundwater are described via lateral water balances. The change of water body volume $\frac{dS_{l,res,w}}{dt}$ combines the lakes, wetlands and reservoir with

$$\frac{dS_{l,res,w}}{dt} = Q_{in} + A(P - E_{pot}) - R_{g_{l,res,w}} - NA_{l,res} - Q_{out}, \tag{4.4}$$

with the water inflow into the water body $Q_{in}$, the area of the water body $A$, precipitation $P$, potential evaporation $E_{pot}$, point groundwater recharge from lakes, reservoirs and wetlands $R_{g_{l,res,w}}$, net abstraction from lakes and reservoirs $NA_{l,res}$ and the outflow from the water body $Q_{out}$. The river volume $\frac{dS_r}{dt}$ is computed from inflow into the river $Q_{r,in}$, the outflow from water the body $Q_{out}$ and net abstraction from surface water $NA_{s,r}$:

$$\frac{dS_r}{dt} = Q_{r,in} - Q_{r,out} - NA_{s,r}. \tag{4.5}$$

The lakes, wetlands, reservoirs, and rivers can be summarized as surface water. In total, ca. 250000 small lakes and more than 5000 reservoirs with a surface area smaller than 100 km$^2$ are considered as local lakes. 1355 lakes with a surface area of more than 100 km$^2$ are considered as global lakes. The difference between global and local lakes as defined in WaterGAP is that a global lake may spread over more than one grid cell, whereas local lakes in a grid cell are aggregated and simulated as one storage. Global man-made reservoirs and regulated lakes with a maximum storage capacity of 0.5 km$^3$ are considered as reservoirs. Local lakes and wetlands receive runoff from only those cells where they are located, while global lakes and wetlands additionally receive runoff from upstream cells based on the routing algorithm. Wetlands are only given in percentage area of a grid cell instead of exact locations. The spatial extent of lakes, reservoirs, and wetlands are shown in Fig. 4.3.

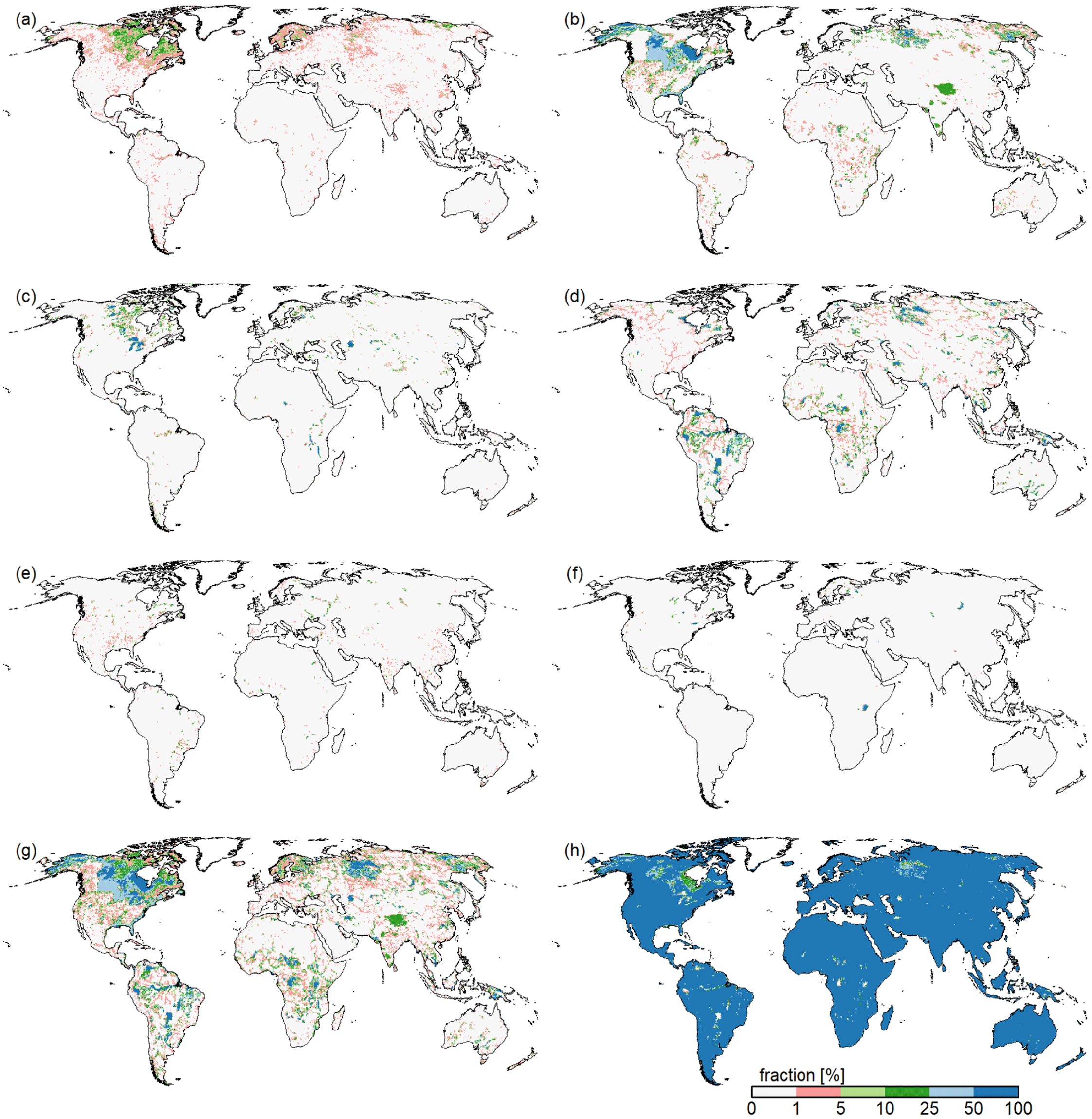


Figure 4.3: Fractions of WGHM the surface water bodies of (a) local lakes, (b) local wetlands, (c) global lakes, (d) global wetlands, (e) global reservoirs, and (f) regulated lakes. Additionally, the grid cell area covered by lakes, reservoirs and wetlands (g) and the land fraction (h) are shown. (taken from Müller Schmied *et al.*, 2021).

Table 4.1: Upper and lower boundaries for the water storage compartments in WaterGAP2.2e (pers. communication).

| **Water storage** | **Min.** | **Max.** |
|---|---|---|
| Global lake | $-S_{\text{la,global,max}}$ | $S_{\text{la,global,max}}$ |
| Local lake | $-S_{\text{la,local,max}}$ | $S_{\text{la,local,max}}$ |
| Global wetland | 0 | $S_{\text{we,global,max}}$ |
| Local wetland | 0 | $S_{\text{we,local,max}}$ |
| Reservoir | 0 | $S_{\text{re,max}}$ |
| River | 0 | - |
| Canopy | 0 | $S_{\text{ca,max}}$ |
| Soil | 0 | $S_{\text{so,max}}$ |
| Snow | 0 | (indirectly) 1000mm |
| Groundwater | - | - |

The lateral water balance for groundwater storage is

$$\frac{dS_g}{dt} = R_g + R_{g_{l,res,w}} - Q_q - NA_g, \tag{4.6}$$

where the change in groundwater storage $\frac{dS_g}{dt}$ is calculated from the diffuse groundwater recharge from soil $R_g$, point groundwater recharge from lakes, reservoirs and wetlands $R_{g_{l,res,w}}$, the groundwater discharge $Q_g$ and net abstraction from groundwater $NA_g$. The groundwater storage can get negative when net abstraction from groundwater is high and thus groundwater depletion is prevalent. The depth of groundwater below the surface is not known.

## 4.5 Water storage limits

To provide realistic simulations of the water storages, the WaterGAP model code includes several limits. In the model code, these upper and lower boundaries for storages prevent overestimation or underestimation of certain storages. Groundwater is the only storage that has no lower and upper boundary, and the river storage has no upper boundary. All remaining storages have a lower and upper limit as shown in Table 4.1 and Eq. 4.7 to 4.13.

The maximum limit for wetlands and lakes is estimated from the maximum storage capacity of the corresponding body. The storage capacity depends on the standard depth parameter, which is 2 m for wetlands and 5 m for lakes, multiplied by the lake or wetland area $A$:

$$S_{\text{la,global,max}} = A_{\text{la,global,max}} \cdot 5m \tag{4.7}$$

$$S_{\text{la,local,max}} = A_{\text{la,local,max}} \cdot 5m \tag{4.8}$$

$$S_{\text{we,global,max}} = A_{\text{we,global,max}} \cdot 2m \tag{4.9}$$

$$S_{\text{we,local,max}} = A_{\text{we,local,max}} \cdot 2m \tag{4.10}$$

The reservoir maximum limit is computed by taking 85% of the pre-defined maximum storage capacity $S_{\text{re,C}}$ of the corresponding reservoir via

$$S_{\text{re,max}} = S_{\text{re,C}} \cdot 0.85. \tag{4.11}$$

Soil and canopy water storages (Eq. 4.13 and 4.12) are limited by using the product of the total available soil water capacity (TAWC) in the 1 m top soil layer with the root depth and a third of the Leaf Area Index (LAI), respectively as

$$S_{\text{ca,max}} = 0.3 \cdot \text{LAI}, \tag{4.12}$$

$$S_{\mathrm{so,max}} = \mathrm{TAWC} \cdot d_{\mathrm{root}}. \tag{4.13}$$

The snow water storage is the only storage that has an indirect static upper limitation of 1000 mm (Tab. 5.2). This means as soon as a snow-related variable in the model reaches 1000 mm, the model prevents uncontrolled snow accumulation by not further decreasing temperature in the upper subgrids of the snow model. In addition, negative values are only allowed for groundwater and lakes.

The WaterGAP version 2.2e also includes the restriction that as soon as the water storages get numerically zero, in this case this means below $1 \times 10^{-12}$ mm (or larger $-1 \times 10^{-12}$ mm and lower $1 \times 10^{-12}$ mm for groundwater and lakes), the water storage is set to zero. This is especially important for including a data assimilation framework to avoid numerical issues. In fact, slight changes in a water compartment during an assimilation update step that is numerically zero could yield different storage simulations for the next model propagation step in time as compared to when the model would run without the assimilation.

## 4.6 Reanalysis fluxes

Forcing data enable the hydrological model to partition input forcing data sets into model fluxes. The forcing data required for WGHM need to consist of daily precipitation, temperature, longwave and shortwave radiation. In the past years, multiple forcing data sets of reanalysis and observations from the meteorological community were used for the model simulations, among others, the Global Soil Wetness Project Phase 3 (GSWP3, Kim, 2017), the Watch Forcing Dataset (WFD, Weedon *et al.*, 2011), ERA-Interim (Dee *et al.*, 2011), ERA5 (Hersbach *et al.*, 2020), or even combinations of them like the WFDEI (WFD/ERA-Interim, Weedon *et al.*, 2014), and the WFDE5 (WFD/ERA5, Cucchi *et al.*, 2020).

In the context of model simulation and data assimilation at the University of Bonn, the most recent internally used data set is the homogenized data set GSWP3-W5E5 (Lange *et al.*, 2022). GSWP3 covers the period 1900 to 1978 and W5E5 covers the period 1979 to 2019. The homogenization reduces a bias between the two forcing data sets. The W5E5 is a combination of the WATCH Forcing Data methodology and ERA5 reanalysis data. The time before 1979 is not required in the context of this thesis but is also available. Post-processing (e.g., transform to the required file format and produce water use output files) is required before the forcing data can be included in the WaterGAP model, which is done at Goethe University in Frankfurt. Thus, although some of these forcing data sets are provided in near-real-time from the meteorological community, the irregular application of post-processing currently limits the application to the year 2019 at the time of writing.

The forcing data set W5E5 is used throughout this thesis for model simulation and running the data assimilation framework, and the precipitation from this data set is used for computing meteorological drought indices.

## 4.7 Vegetation representation

To include vegetation metrics in the simulation of water storages, the WGHM model uses static land cover classes and fluxes. Global land cover maps by Moderate Resolution Imaging Spectroradiometer for the year 2004 (Müller Schmied *et al.*, 2014) are used, read into the model as a temporally static map, and are shown in Fig. 4.4. The classes range from very dry land cover classes like, for example, within the Savanna over larger parts of Southern Africa to wetter land cover classes like the evergreen broadleaf forest in, for example, the Amazon basin.

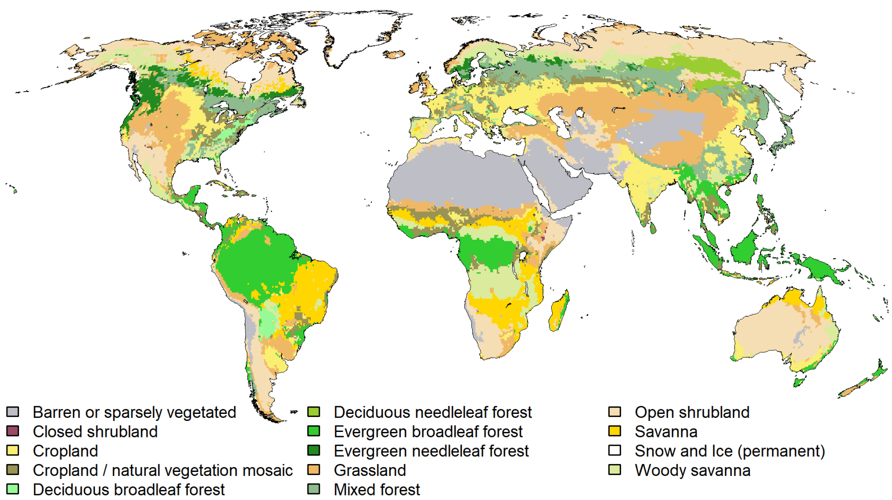


Figure 4.4: Land cover classification used in WaterGAP (taken from Müller Schmied *et al.*, 2021).

The growth of vegetation is then simulated during the growing season and starts with a temperature larger than 8°C under consideration of the classes. One of the vegetation indicators within the model is the LAI, which is computed by a function of daily precipitation and temperature. The model limits LAI values (unitless) to minimum and maximum values based on the static land cover classes, deciduous plant fractions, and factors for evergreen plants (Müller Schmied *et al.*, 2021).

As mentioned in, for example, Wei *et al.* (2017), vegetation transpiration is the largest component in global terrestrial evapotranspiration. Thus, a further metric for vegetation from the model simulations is now introduced: the Actual Evapotranspiration (AET). In WGHM, AET is derived by estimating canopy evaporation (by determining potential evapotranspiration using the Priestley-Tayler method), snow sublimation, AET from the soil, and potential evaporation of open water surfaces. The total of AET is then derived by merging the fluxes with transpiration from irrigated surfaces. According to Müller Schmied *et al.* (2021), trends of AET might not be fully reliable because of the error in the forcing data and model uncertainties.

# Chapter 5

# Data assimilation

Models do not perfectly represent reality, are based on assumptions, and thus they miss process representations and are subject to forcing errors. Observations are measured via remote sensing instruments from satellite or in-situ data and present a real picture of the current variable with measurement errors, but they can be sparse and can have gaps or missing quantities.

Data assimilation is a method for integrating real observations into model simulations and aims to improve the model's realism, while filling observational gaps, downscaling the observations spatially, and disaggregating them vertically. Thus, observations, model trajectories, initial conditions, forcing data, and parameters are considered to find the best estimation of the true state. Nonetheless, data assimilation violates the physical consistency of model runs (e.g., the model aims at the closure of the water balance equations) but that disadvantage is taken into account for improving the realism of the states.

In the past, the hydrological community implemented multiple strategies for the integration of observations into the models. In general, one can divide these strategies into two techniques: variational data assimilation and sequential data assimilation (e.g., Le Dimet *et al.*, 2009; Abaza *et al.*, 2015; Girotto *et al.*, 2017; Khaki *et al.*, 2018). Variational data assimilation uses all available observations within a time interval to improve the initial conditions. In this technique one sets up a cost function at the beginning and all observations are used simultaneously to estimate the best fit for the initial conditions. The lower atmosphere is a chaotic system and estimating model states strongly depends on the initial state, therefore, variational assimilation is widely used in the meteorological community. Sequential data assimilation uses observations as soon as they are available to be integrated. Thus, for each time step, a model prediction is simulated and – in case of existing observations – updated with the observation for that time step by considering the probability distribution of the model and observation. For the next time step, the updated state then serves as input to the next model prediction and the procedure repeats. In addition, data assimilation techniques can be divided into filter and smoother; a filter estimates from past and present observations, while the smoother estimates from past, present, and future observations.

In this study, the focus is on sequential data assimilation because the interest in finding optimal states dominates in contrast to improving initial conditions as is the case for variational schemes. The data assimilation framework is based on two sequential filters that are commonly used in the hydrology community: the EnKF and the ESTKF. Both filters are Kalman filters and use ensemble representations for model states, parameters, and forcing data. A big advantage of ensemble-based filters is that they approximate model uncertainties by Monte Carlo representation because usually, model uncertainties are unknown and not provided due to the high complexity of the models. To provide a detailed picture for understanding the filter algorithms used in this work, I first give some basic statistical theory before introducing the Kalman filter, the EnKF, the ESTKF, its concrete implementation in WaterGAP, and existing challenges and tuning strategies.

## 5.1 Theory of assimilation filters

The Bayes theorem is the basis of most data assimilation filters and smoothers. In Sec. 3.3, I introduced the basics of PDF along with the first two statistical moments, which is now used for further explaining Bayes theorem. Ensemble-based filters assume Gaussianity, which means that the first two statistical moments of mean and variance are considered in the optimality constraints that the filters satisfy but further moments like skewness and kurtosis (third and fourth central statistical moment) are not explicitly considered. Other filters, for example, the particle filter, consider the entire PDF, unlike Kalman filters. The $n$-th central statistical moment computes as

$$M_n(x) = E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - E[X])^n f(x) dx. \tag{5.1}$$

I now describe the Bayes theorem following Wikle & Berliner (2007), where observed quantities are referred to as $y$, unobservable quantities as $x$, and for PDFs I stick to the notation $f$. The Bayes theorem unifies the PDFs of observables and unobservables by

$$f(x|y) = \frac{f(x)f(y|x)}{f(y)}. \tag{5.2}$$

The posterior explicitly reads as the distribution of the unobservable given the observable/data and is the main goal to derive. In fact, different (priori-) estimators PDFs are used to estimate a posterior or analysis state distribution. $f(y|x)$ is the PDF of the observed quantities given the unobserved quantities, and can also be seen as classical likelihood estimation. To provide an example for this thesis in accordance with Wikle & Berliner (2007), $y$ could be the imperfect TWSA observations of the true unobservable TWSA and the PDF then quantifies the distribution of the measurement error of the TWSA. $f(x)$ is the prior distribution of the unobservable quantities, for example, historical information about TWSA from climatology or a forecasted TWSA. $f(y)$ is also a prior distribution, called the marginal distribution of the observable, and can be seen as a normalizing factor.

From the Bayes theorem, a proportional relation can be found: The posterior density for $x$ given the measurement $y$ is proportional to the product of the prior density of $x$ and the likelihood of $y$

$$f(x|y) \propto f(x)f(y|x). \tag{5.3}$$

The temporal evolution of the distributions is now defined by applying Markov assumptions to the prior variables of the proportional expression of the Bayes theorem by

$$f(x_{0:T}) = f(x_0) \prod_{t=1}^{T} f(x_t|x_{t-1}) \tag{5.4}$$

and

$$f(y_{1:T}|x_{0:T}) = \prod_{t=1}^{T} f(y_t|x_t), \tag{5.5}$$

where $f(x_0)$ is the initial state distribution, $f(x_t|x_{t-1})$ is the evolution distribution, and $f(y_t|x_t)$ is the distribution of the observations assumed to be independent of the true state. Inserted into the Bayesian theorem, we derive

$$f(x_{0:T}|y_{1:T}) \propto f(x_0) \prod_{t=1}^{T} f(y_t|x_t) f(x_t|x_{t-1}). \tag{5.6}$$

From this expression, it can be learned that the previous state can be updated without recalculations as soon as new data are available. This procedure reflects exactly the sequential behavior of sequential filter algorithms.

### 5.1.1 Kalman Filter

The Kalman filter is a sequential data assimilation algorithm and was first introduced by Kalman & Bucy (1961). For linear systems and under the assumption of uncorrelated errors and Gaussianity, it is the optimal sequential method, and second-order statistics are integrated to predict model forecast error statistics (Evensen *et al.*, 2009). For this linear system, the Kalman filter assumes a Gaussian distribution of error representation, and the analysis (or the update) is the result of merging two Gaussian distributions.

Before explaining the filter theory, three states are introduced: the forecast state $\boldsymbol{x}_t^f$, which is the state vector of model simulations at time $t$ containing all model grid cells for the considered model variables, the observation state $\boldsymbol{y}_t$, which contains the observation grids for the considered observation variables, and the update state $\boldsymbol{x}_t^a$, which contains the final updated states by integrating the observation into the prediction state.

A two-step procedure is now applied consisting of the consecutive execution of a forecast/prediction step and an analysis/update step. The forecast step and thus the forecast state $\boldsymbol{x}_t^f$ is derived by the propagation of the linear stochastic model for one step in time via

$$\boldsymbol{x}_t^f = \mathbf{M}_{t-1,t}\boldsymbol{x}_{t-1}^f + \boldsymbol{\eta}_t, \tag{5.7}$$

where $\mathbf{M}$ is the linear model operator and $\boldsymbol{\eta}_t$ are the model errors with the corresponding variance-covariance martix $\mathbf{Q}$. As forecasted state $\boldsymbol{x}_{t-1}^f$ serves either an initial state derived from model simulations or the previous updated state $\boldsymbol{x}_{t-1}^a$. We have to propagate the model variance-covariance matrix $\mathbf{P}^f$ in time as well by

$$\mathbf{P}_t^f = \mathbf{M}_{t-1,t}\mathbf{P}_{t-1}^f(\mathbf{M}_{t-1,t})^T + \mathbf{Q}_t \tag{5.8}$$

to derive the forecast error variance-covariance matrix of the current time step. Hereinafter, the time index will be dropped when all variables within an equations refer to the same time index. The aim is now to get as close as possible to the true state by using model forecast and observations $\boldsymbol{y}$ with variance-covariance $\mathbf{R}$ of the observation. Subsequently, a Best Linear Unbiased Estimator (Koch & Schönfeld, 1989) is set up to derive the cost function

$$J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}^f)^T(\mathbf{P}^f)^{-1}(\boldsymbol{x} - \boldsymbol{x}^f) + (\boldsymbol{y} - \mathbf{H}\boldsymbol{x})^T\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}) \tag{5.9}$$

The minimum value of the cost function ($\min(J(x)) = x^a$) is found by taking the derivative via

$$\nabla J(\boldsymbol{x}) = 2(\boldsymbol{x} - \boldsymbol{x}^f)(\mathbf{P}^f)^{-1} + 2(\boldsymbol{x} - \boldsymbol{y})\mathbf{R}^{-1} = 0. \tag{5.10}$$

From solving and reordering the cost function, we can extract the second step. The forecasted state is updated by the observations $\boldsymbol{y}$ and a corresponding weighting matrix $\mathbf{K}$ – called Kalman or gain matrix – to derive the updated state:

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \mathbf{K}(\boldsymbol{y} - \mathbf{H}\boldsymbol{x}^f), \tag{5.11}$$

where $\mathbf{H}$ is the observation operator that maps the model states into the observation space. In other words, the observations are assimilated with the forecast state to find a better estimation of the true state. The Kalman matrix $\mathbf{K}$ gives weights model errors and observation errors via:

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T(\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R})^{-1} \tag{5.12}$$

To complete the update step also for the updated errors of the model state are derived via

$$\mathbf{P}^a = (\mathbb{1} - \mathbf{KH})\mathbf{P}^f, \tag{5.13}$$

with unit matrix $\mathbb{1}$. When a nonlinear model is given, one can describe the model trajectory as

$$\boldsymbol{x}_t^f = \mathbf{m}(x_{t-1}^f). \tag{5.14}$$

Issues of the Kalman filter are that the filter works only optimally for linear systems while the models are usually non-linear and high dimensional. A linearization can be used to formulate the model propagation, which is done in the so-called Extended Kalman filter as

$$\mathbf{M} = \frac{\partial \mathbf{m}}{\partial x} \tag{5.15}$$

with $\mathbf{M}$ being the tangent linear operator (Jacobian) of $\mathbf{m}$ (Evensen, 2003). The linearized evolution can lead to unstable and poor evolution of the covariance matrix (e.g., Evensen, 1993). Further, for high-dimensional models, the propagation of the state covariance matrix gets unfeasible ($n^2$ with n of O$(10^7 - 10^9)$). Evolving and inverting large covariance matrices is extremely costly due to memory and computation time. A general aim should be to reduce costs as much as possible, for example, by a simplified error evolution, (e.g., a constant variance only), reduced rank of $\mathbf{P}$, reduced resolution of the model (at least for error propagation), or reduced model complexity.

Thus, a more stable way for large assimilation systems is to use ensemble filters, which are able to deal with non-linear systems, derive the variance-covariance matrix of model errors, and retain as much as possible of the error information for model states and observations.

### 5.1.2 Ensemble Kalman Filter

The Ensemble Kalman Filter (EnKF) was designed to cope with the non-linearity of models by applying a Monte Carlo method. The development started by Evensen (1994) and was further continued in Evensen (2003).

The ensemble represents the probability distributions of the model state, which avoids computing and propagating a large variance-covariance matrix of the model state. The assumption is set up, that the probability density functions of the model and observations are independent (Evensen, 2003). So, an ensemble of state vectors is used instead of using a single state vector as in the original Kalman filter.

The ensemble sizes can vary and thus the computational costs can be adapted depending on the given infrastructure and application, whereas small examples for the ensembles should still be statistically representative (Kalnay, 2003). With these adaptions, the filter is much more efficient than the (Extended) Kalman Filter and more suited for parallel computing. For linear solutions and increasing ensemble size, the EnKF would converge towards the Kalman filter (Evensen, 2003).

The forecast of each model state ensemble member $\boldsymbol{x}^{f(i)}$ for the current time step is the non-linear model forward integration in time

$$\boldsymbol{x}_t^{f(i)} = M\left(\boldsymbol{x}_{t-1}^{f(i)}\right) + \boldsymbol{\nu}_t, \tag{5.16}$$

where as input serves either a model state of the previous time step $\boldsymbol{x}_{t-1}^{f(i)}$ or the previous analysis step and $\boldsymbol{\nu}$ is the random model error. In total, the ensemble consists of $N_e$ ensemble members, and each ensemble member is propagated independently in time.

The mean forecast state vector $\bar{\boldsymbol{x}}^f$ results from the single ensemble members and is the basis for computing the forecast error covariance matrix $\tilde{\mathbf{P}}^f$:

$$\bar{\boldsymbol{x}}^f = \frac{\sum_{i=1}^{N_e} \boldsymbol{x}^{f(i)}}{N_e} \tag{5.17}$$

$$\tilde{\mathbf{P}}^f = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} (\boldsymbol{x}^{f(i)} - \bar{\boldsymbol{x}}^f)(\boldsymbol{x}^{f(i)} - \bar{\boldsymbol{x}}^f)^T. \tag{5.18}$$

In addition, an observation ensemble $\boldsymbol{y}^{(i)}$ is generated by

$$\boldsymbol{y}^{(i)} = \boldsymbol{y} + \boldsymbol{\epsilon}^{(i)}, \tag{5.19}$$

where $\boldsymbol{\epsilon}^{(i)}$ is the observation error. The corresponding observation error covariance matrix is denoted with $\mathbf{R}$ as similarly done for the Kalman filter. The observation vector can be related to the simulated model state vector via the observation operator $\mathbf{H}$, in other words, the observation operator transforms the forecast state vector to the observations space. The analysis/update state for each ensemble member is then derived by adding an increment to the forecasted model state, in which the observation ensemble and the forecast model ensemble get weighted by the Kalman Gain matrix $\mathbf{K}$:

$$\boldsymbol{x}^{a(i)} = \boldsymbol{x}^{f(i)} + \mathbf{K}(\boldsymbol{y}^{(i)} - \mathbf{H}\boldsymbol{x}^{f(i)}). \tag{5.20}$$

Again, as in the original Kalman filter (Eq. 5.12), the Kalman Gain matrix $\mathbf{K}$ relates the empirically determined forecast error matrix $\tilde{\mathbf{P}}^f$ and the observation error matrix $\mathbf{R}$ by weighting and is also used to estimate the update error matrix $\tilde{\mathbf{P}}^a$ (analogous to Eq. 5.13):

$$\mathbf{K} = \tilde{\mathbf{P}}^f \mathbf{H}^T (\mathbf{H}\tilde{\mathbf{P}}^f \mathbf{H}^T + \mathbf{R})^{-1}. \tag{5.21}$$

An important point to notice is that $\mathbf{R}$ is only used once in the Kalman Gain matrix, whereas the model ensemble error matrix $\tilde{\mathbf{P}}^f$ is used twice.

In addition to the notation of the EnKF in this section, the filter can also be formulated in observation space (e.g., Vetra-Carvalho *et al.*, 2018). For this, the forecast ensemble is now written as matrix via $\mathbf{X}^f = [\boldsymbol{x}^{f(1)}\boldsymbol{x}^{f(2)}...\boldsymbol{x}^{f(N_e)}]$, $\mathbf{X}'^f$ denotes the forecast ensemble matrix, where each column is reduced by the forecast ensemble mean, and $\mathbf{Y}$ is the observation ensemble matrix that holds the observations per ensemble member in each column $\mathbf{Y} = [\boldsymbol{y}^{(1)}\boldsymbol{y}^{(2)}...\boldsymbol{y}^{(N_e)}]$. As next, the analysis ensemble matrix $\mathbf{X^a}$ can be computed from

$$\mathbf{X}^a = \mathbf{X}^f + \frac{1}{N_e - 1}\mathbf{X}'^f \mathbf{S}^T \mathbf{F}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{X}^f) \tag{5.22}$$

with $\mathbf{S} = \mathbf{H}\mathbf{X}'^f$ that projects the forecast ensemble to the observation space and the innovation covariance matrix $\mathbf{F} = \mathbf{S}\mathbf{S}^T + (N_e - 1)\mathbf{R}$.

Finally, it must be mentioned that the EnKF filter assumes Gaussian distributions. It is possible to run the EnKF with prior ensembles that are non-Gaussian. However, since the analysis step expects Gaussianity, the final update represents a mixture by inheriting some of the non-Gaussianity from the prior ensemble (Evensen, 2003).

### 5.1.3 Error Subspace Transform Kalman Filter

The Error Subspace Transform Kalman Filter (ESTKF) is an ensemble-based filter and is a variant of two filters: (1) the Ensemble Transform Kalman Filter and (2) the Singular Evolutive

Interpolated Kalman filter. The Ensemble Transform Kalman Filter uses a square root analysis scheme that avoids the perturbation of observation errors as is done for the EnKF (cmp. $(\mathbf{H}\tilde{\mathbf{P}}^f\mathbf{H}^T + \mathbf{R})$ in Eq. 5.21). This means that the inverse of the observation error matrix ($\mathbf{R}^{-1}$) is directly computed and within the filter, the forecast state ensemble is explicitly transformed to the analysis state ensemble (Bishop *et al.*, 2001). The Singular Evolutive Interpolated Kalman filter applies a square root filter analysis step in the ensemble error subspace, which saves runtime. More information about the Ensemble Transform Kalman Filter and Singular Evolutive Interpolated Kalman filter can be found, for example, in Bishop *et al.* (2001), and Pham (2001).

Nerger *et al.* (2012) showed that the Singular Evolutive Interpolated Kalman can also be formulated as a square root filter and the state ensemble transformation can be analogously performed as in the Ensemble Transform Kalman Filter as a combination of analysis update and ensemble transformation. In the case that the same forecast ensembles are used in the Singular Evolutive Interpolated Kalman and the Ensemble Transform Kalman Filter, the analysis state and analysis state covariance matrix of both filters are identical. However, the ensembles from which the analysis state and its covariance matrix are computed do not necessarily need to be equal.

Nerger *et al.* (2012) then adapted the reformulated Singular Evolutive Interpolated Kalman in the way that the ensemble transformation from forecast to analysis is performed in the error subspace instead of the ensemble subspace. More information about the error subspace can be found in Nerger *et al.* (2005). The adapted filter was named the ESTKF. Due to the reformulation of the Singular Evolutive Interpolated Kalman as an ensemble square root filter and its adaption, the ESTKF has slightly less computational costs as compared to the Ensemble Transform Kalman Filter and avoids perturbation of the observation error.

As mentioned, the analysis step is now computed in the so-called error subspace. This is performed by defining a new matrix

$$\mathbf{L} = \mathbf{X}^f \mathbf{A}_{ESTKF}, \tag{5.23}$$

where $\mathbf{A}_{ESTKF}$ is the projection matrix of size $N_e \times N_e - 1$ that projects the forecast ensemble matrix $\mathbf{X}^f = [\boldsymbol{x}^{f(1)}\boldsymbol{x}^{f(2)}...\boldsymbol{x}^{f(N_e)}]$ to the error subspace. The elements of $\mathbf{A}_{ESTKF}$ are chosen in the way that the application on the forecast state implicitly subtracts the ensemble mean and subtracts a fraction of the last column from all other columns (Vetra-Carvalho *et al.*, 2018) and is assembled by

$$A_{\mathrm{ESTKF}_{i,j}} = \begin{cases} 1 - \frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & i = j, i < N_e \\ -\frac{1}{N_e}\frac{1}{\frac{1}{\sqrt{N_e}}+1} & i \neq j, i < N_e \\ -\frac{1}{N_e} & i = N_e \end{cases} \tag{5.24}$$

As a result of this concept, the columns of $\mathbf{A}_{ESTKF}$ are orthonormal and the matrix has full rank.

To perform the analysis step, Nerger *et al.* (2012) and Vetra-Carvalho *et al.* (2018) present a notation where the state ensembles are updated given a weighting matrix/weight vectors. In the following, $\mathbf{X}'$ are perturbation matrices (or vectors $\boldsymbol{x}'$) for the forecast and the analysis with a zero mean, $\bar{w}$ is the weight vector for the ensemble mean and $\mathbf{W}$ is a weight matrix for the ensemble perturbations. With these matrices, one can reformulate the expression for the ensemble mean analysis state and the perturbed analysis state with zero mean by

$$\bar{\boldsymbol{x}}^a = \bar{\boldsymbol{x}}^f + \mathbf{X}'^f \bar{\boldsymbol{w}}, \tag{5.25}$$

and

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{W}'. \tag{5.26}$$

The weight matrix $\mathbf{W}'$ in ESTKF can directly be computed from the transformation matrix $\mathbf{T}_{\text{ESTKF}}$, which yields in

$$\mathbf{T}_{\text{ESTKF}}\mathbf{T}_{\text{ESTKF}}^T = (\mathbb{1} + \frac{1}{N_e - 1}(\mathbf{HL})^T\mathbf{R}^{-1}(\mathbf{HL}))^{-1}. \tag{5.27}$$

Except for $\mathbf{L}$, all variables are also known from the EnKF and $\mathbf{L}$ introduces the transformation to the error subspace. Subsequently, $(\mathbf{T}_{\text{ESTKF}}\mathbf{T}_{\text{ESTKF}}^T)^{-1}$ is decomposed with eigenvalue decomposition to derive the transformation matrix by

$$\mathbf{T}_{\text{ESTKF}} = \mathbf{U}\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{U}^T \tag{5.28}$$

For all square root analysis filters, one could replace the singular value decomposition with another decomposition method, e.g., the Cholesky decomposition. However, Nerger *et al.* (2012) showed that when using deterministic ensemble transformations, the decomposition with Cholesky showed higher numerical errors in twin experiments than the decomposition with the symmetric square root. Finally, with the help of the symmetric root matrix $\mathbf{T}_{\text{ESTKF}}$ and the projection matrix $\mathbf{A}_{\text{ESTKF}}$ the weight matrix

$$\mathbf{W}' = \mathbf{A}_{\text{ESTKF}}\mathbf{T}_{\text{ESTKF}}\mathbf{A}_{\text{ESTKF}}^T \tag{5.29}$$

can be used to transform the ensemble analysis perturbations. The weight vector is then used to transform the ensemble analysis mean as

$$\bar{\boldsymbol{w}} = \frac{1}{\sqrt{N_e - 1}}\mathbf{A}_{\text{ESTKF}}\mathbf{U}\mathbf{\Sigma}^{-1}\mathbf{U}^T(\mathbf{HL})^T\mathbf{R}(\boldsymbol{y} - \mathbf{H}\bar{\boldsymbol{x}}^f). \tag{5.30}$$

## 5.2 Implementation

### 5.2.1 Parallel Data Assimilation Framework

The Parallel Data Assimilation Framework (PDAF, Nerger & Hiller, 2013; Nerger *et al.*, 2020) is a software environment for the implementation of data assimilation for simple to complex models in an efficient manner, developed at the Alfred Wegener Institute. Available open-source code enables the users to focus on applying data assimilation instead of implementing assimilation filters, which simplifies the implementation of assimilation systems. PDAF is implemented in the form of parallelized infrastructure with generic code where case-specific routines need to be adapted by the user depending on the chosen algorithm and provides different assimilation filters that can be used on supercomputers.

PDAF is structured in a way that model development and PDAF development can be further continued independently. A parallelization support is provided for ensemble forecasts, more precisely, the parallelization of ensemble forecasts can be implemented independently from the model, as well as the analysis step. All model and observation fields are stored in vector form, thus the state vector $x$ points in the n-dimensional space and the observation vector $y$ points in the m-dimensional space. The observation vector does not need to have the same grid resolution as the model and can also be sparse.

The coupling of the chosen model and the assimilation system is possible in two main modes (1) offline mode and (2) online mode:

During the offline mode, the model and assimilation program operate separately via two executables. These executables run consecutively for each time step. At the beginning, the model initializes required variables and reads in forcing and auxiliary data. As next, the model runs an integration step and finalizes with post-processing and saving relevant variables for the data assimilation to the storage. Then, the assimilation program loads the files, computes the analysis step, and writes out the necessary files for running the next model integration (Fig. 5.1).

The advantage of running a data assimilation framework in offline mode is its rather easy implementation. Only a few changes are necessary in the model code for writing the required files for the update step. But the downside is that reading and writing files is very costly, and limits efficiency. Furthermore, the model and PDAF need to be restarted and initialized for each time step, increasing the run time. Especially for working on a supercomputer, the offline mode limits its performance strongly, because each time step a job needs to be submitted to the user management.

In contrast to the offline mode, the online mode requires only one executable program because the model and filter are coupled into a single executable program. This means that it is not necessary to repeat the initialization of the model and PDAF successively, they are only executed once at the beginning (Fig. 5.1). Then, the executable runs over all time steps and repeats in consecutive order model integration and assimilation. In other words, the prediction step forwards model states to the update step and vice versa. Finally, the program closes with a post-processing routine. The online mode is computationally more efficient but requires more implementation work because the model and PDAF functions need to be directly linked to transfer the model states and further required fields.

During the last years, the assimilation framework for integrating GRACE/-FO TWSA into the WaterGAP model at the Institute of Geodesy and Geoinformation at the University of Bonn developed from an offline framework without PDAF for regional applications, went to an offline version with PDAF capable of regional and global applications and nowadays works in online mode for the proposed filters EnKF and ESTKF with and without localization and inflation factor variations for regional and global applications. Further filter algorithms are easily implementable and might be integrated in the future.

### 5.2.2 Simulated TWS versus observed TWSA

In each GRACE/-FO standard processing of level2 SHCs, the temporal mean per SHC for a chosen time period needs to be removed from each monthly field before transforming the SHC to gridded data. Thus, transforming the SHC to the grid yields anomalies of TWS. When comparing observations to model simulations, on the one hand, we have the absolute water storage and on the other hand, we have anomalies of that water storage. To synthesize simulation and observations, two possibilities exist: (1) the temporal mean of model simulations is removed per grid to compute anomalies or (2) a temporal mean is added to the observations per grid to compute the absolute storage. In this work, I decided to stick to the first option, which is assimilating anomalies of TWS into the model. To make the GRACE/-FO TWSA even more comparable to model simulation, a temporal mean is additionally removed on the grid level (technique 1) to adapt similar techniques as used for the model simulations to derive TWSA from TWS.

In order to study the effect of the mean removal, the Root Mean Square Difference (RMSD) (Sec. 6.1.1) between GRACE/-FO and WaterGAP is evaluated for five different referencing periods. For this, the mean is removed on SHC level for GRACE/-FO and on grid level for WaterGAP and GRACE from 2003 to 2008, 2003 to 2012, 2003 to 2016, 2009 to 2016, and 2003 to 2019. As expected, the longer the period of mean computation, the smaller is the global mean RMSD (Tab. 5.1). The reference period 2003 to 2019 shows the lowest RMSD between GRACE/-FO and WaterGAP with 66.30 mm and the reference period 2003 to 2008 shows the highest RMSD with 72.22 mm. Nonetheless, although there is a clear tendency towards an

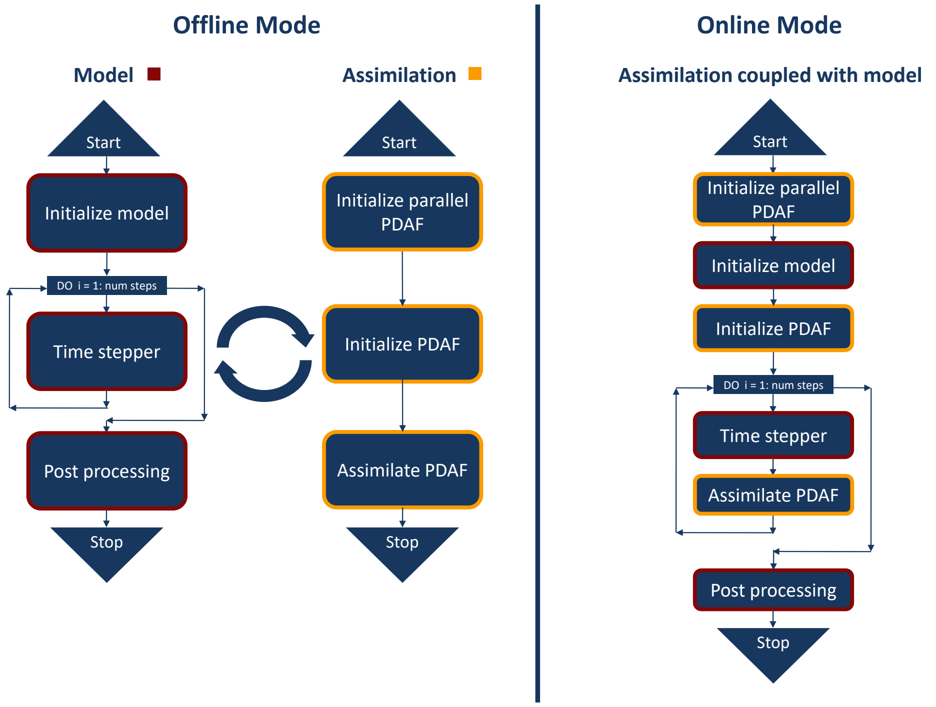


Figure 5.1: Simplified workflow for the PDAF offline and online mode. Based on Nerger *et al.* (2020).

Table 5.1: Global mean RMSD between GRACE/-FO and WaterGAP TWSA estimated for monthly data from 2003 to 2019 under the consideration of different reference periods in the GRACE/-FO processing chain and to derive simulated TWSA from TWS. The RMSD are given in mm.

| **Reference period** | **Global mean RMSD [mm]** |
|---|---|
| 2003-2008 | 72.22 |
| 2003-2012 | 69.17 |
| 2003-2016 | 67.01 |
| 2003-2019 | 66.30 |
| 2009-2016 | 70.95 |

improved RMSD with a longer reference period, the changes in RMSD are still relatively small. If the comparison is applied between GRACE/-FO and an open-loop simulation (will be introduced in Sec 5.2.5), the same conclusion can be drawn. Therefore, the final choice of reference period is the full available time span of 2003 to 2019.

### 5.2.3 Spatial and temporal aggregation

WaterGAP simulates daily water storage variations covering a total number of 67800 grid cells over land on an 0.5° grid. At the same time, the GRACE/-FO observations are given as monthly fields and cover the whole globe with ocean and land mass with a spatial resolution of about 300 km. So in fact, there is a mismatch of spatial and temporal resolution between the model simulation and the observation. It is important to understand that the spatial width of the grid on which the GRACE/-FO TWSA are calculated (e.g., 0.5° or 3°) is only chosen to solve the spherical harmonical synthesis at regular grids, but should not be confused with the native spatial resolution of GRACE/-FO of about 300 km. To relate the model grids to the observations spatially, it is required to apply four processing steps:

1. The spatial degree of the GRACE/-FO observation grids needs to be chosen and the TWSA fields processed on that grid.

2. All observation cells that do not overlap with any of the model cells need to be removed.

3. A restriction is applied that a cell of the observations is only included in the framework as soon as it is covered by at least 50 % of model simulation cells (important for coastal regions).

4. An observation operator is set up that directly relates the model spatial grids of 10 compartments to the GRACE/-FO observation grids of TWSA.

The model state vector consists of the 10 water storages given in WaterGAP (Tab. 4.1) and holds them for each cell. The observation operator aggregates the 10 compartments to TWS. Then, depending on the choice of observation grid, the model cells are aggregated to be mapped into the observation space. As a consequence, the observation operator is a matrix, that has the size $n$ x $m$, where $n$ is the number of model grids times the 10 compartments and $m$ is the number of observation grids. For the observation space, there exist different choices for the grid size. Before Eicker *et al.* (2014), it was common to compute spatial averages for basins (e.g., Houborg *et al.*, 2012; Li *et al.*, 2012). Eicker *et al.* (2014) then suggested to compute regular aggregated grids with different grid sizes for the GRACE observations and tested, which effect the grid size has on the numerical stability of the variance-covariance matrix of the observations for the Mississippi River basin under the condition that a 500 km Gaussian filtering is applied. Nonetheless, Eicker *et al.* (2014) mention that the condition of the observation error matrix strongly depends on the post-processing and the results might differ for other setups. Thus, here I try different grid sizes to see which influence it has on the numerical stability of the observation error matrix.

Various options for the spatial grid representation of GRACE/-FO TWSA for data assimilation exist. For example, one could consider hydrological regimes by calculating basin averages but this requires a threshold for minimum basin size. Very small basins represent an area where GRACE/-FO is usually not suitable and there would be many coastal regions or regions in the north, where no grid point would be included in the assimilation. There is also the possibility of providing the GRACE data on the same grid size as the model output (0.5°), however, the GRACE/-FO data are limited to the spatial resolution of 300 km and a smaller grid size would not add geophysical signal. In addition, expressing the full variance-covariance matrix on that grid requires a large amount of RAM and physical storage. To overcome limited storage capacities, one could consider the assimilation of continent-wise GRACE/-FO data. However, tests show, that the effort of running six separate assimilation runs is very high, and the variance-covariance matrices of the observations can derive a worse condition on a finer spatial scale as compared to a larger scale (e.g., on 0.5° as compared to 4°). For this, I computed the condition numbers for each continental variance-covariance matrix and temporally averaged them. The condition numbers are a ratio of the largest singular value to the smallest singular value. For example, for Europe, the temporal mean condition number for the GRACE/-FO variance-covariance matrix on 0.5° is about 4312 or for Africa 1467 – probably because of the native spatial resolution of about 300 km instead of about 55 km for an 0.5° grid – and observation error correlations between continents would be neglected.

To enable a single global framework that considers correlations worldwide, the TWSA observations and error information can be processed on coarser grids, as the model spatial resolution,

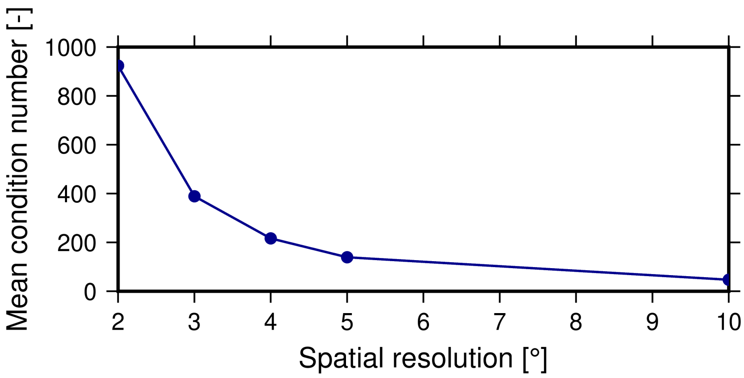


Figure 5.2: Temporal mean condition numbers of global GRACE/-FO variance-covariance matrices for TWSA over land in dependency of the chosen size of the spatial grid (2 to 10°). The condition numbers are unitless.

for example on 2, 3, 4, 5, or 10° as shown in Fig. 5.2, where the temporal average condition numbers are presented. All options show a much lower condition number as compared to the 0.5° for the continental example for Europe and Africa. It seems as if the condition number from degrees higher or equal 4° do not vary considerably, but this strongly changes for degrees 3 and finer. In summary, the finer the grid, the larger the condition number, which evolves exponentially with finer grid width and leads me to the decision of a 4° grid for the observations. It must be mentioned at this point, that the comparison of condition numbers provides only an indication that the condition of an observation variance-covariance matrix on, e.g., 2° could potentially lead to faster crashes than, e.g., a 4° matrix during assimilation. However, this is not necessarily the case and also depends on the chosen filter algorithm. To underline the choice of a 4° grid, saving only the error matrices for Africa on an 0.5° grid for the complete period (171 months) would require about 150 GB of storage versus 1.4 GB for the global 4° (matrix size is 995 x 995 entries). Currently, the observation error matrices need to be stored because the programs for processing observations and running the assimilation are separate programs. Nonetheless, if future gravity missions would provide a higher resolution of the TWSA, using the finer grid of, for example, 0.5° together with localization strategies would be a possible workaround under the consideration that enough RAM and physical storage is available.

Next, the model simulations need to be matched temporally to the GRACE/-FO observation, which means I compare daily to monthly storage information respectively. Daily GRACE/-FO solutions exist but they do not share the same spatial resolution as the monthly solutions have (e.g., Mayer-Gürr *et al.*, 2018). The literature has various possibilities to derive the link between the daily model simulations and the monthly observations. In a very general view, the EnKF adds an increment (Eq. 5.20) to the model prediction and derives the new update per time step. To derive a monthly increment, the daily simulations need to be temporally averaged. At the end of the month, a monthly TWSA value is calculated from the daily model simulations (and a mean value is removed).

A very common method to calculate the monthly TWSA value of the simulations is to take the average of all given days (e.g., Eicker *et al.*, 2014; Girotto *et al.*, 2016; Schumacher *et al.*, 2016; Tangdamrongsub *et al.*, 2017), other approaches use only specific days, for example, day five, 15, and 25 of the month (e.g, Zaitchik *et al.*, 2010; Reager *et al.*, 2015; Kumar *et al.*, 2016) to adapt to the overpasses of TWSA. However, to apply this scheme, one needs to use a smoother (increase in computational costs), and thus the model has to be forwarded twice, which strongly increases computational costs.

Some studies also compute daily increments (e.g., Girotto *et al.*, 2016; Girotto *et al.*, 2017) (use of 3D-EnKF) and one could interpolate the observations to daily values but this introduces additional errors. Girotto *et al.* (2016) found that applying daily increments led to large daily variability, and suggested applying monthly averages of the daily increments in a GRACE/-FO based assimilation system instead. I here avoid computing daily increments and compute monthly averages of the simulations because the inhouse system is already set up for monthly increments, changing to daily increments would require many adaptions in the model and PDAF codes, and the increased computational effort for daily increments is also avoided. In principle, there should only be minor differences for the monthly assimilation outputs that are analyzed within this thesis when applying daily or monthly increments. However, in the future with new missions that provide operational daily solutions of TWSA, it might be better to switch to daily increments because the assimilation outputs could then be analyzed on a daily basis.

To distribute the monthly increment to the daily model values, various studies followed different methodologies, for example, adding the full monthly increment to a specific day, first day, last day (Eicker *et al.*, 2014; Girotto *et al.*, 2016; Schumacher *et al.*, 2016; Girotto *et al.*, 2019; Felsberg *et al.*, 2021). Girotto *et al.* (2016) found that the assimilation output is closest to the GRACE observations when the increment is applied at the beginning of a month, instead of applying it incrementally for each day or at the end of a month. Nonetheless, these differences were found for groundwater and soil moisture as only minor. Other studies divide the monthly increment up into the number of days in that month and add for each day the divided part (Zaitchik *et al.*, 2010; Reager *et al.*, 2015; Tangdamrongsub *et al.*, 2017; Li *et al.*, 2019), which introduces temporal correlations into the model but a smoother is required for this consecutive updating. Finally, in the assimilation framework used in this thesis, the increment is added for the last day because of easy implementation and avoiding extra computational costs with smoothers.

### 5.2.4 Parameter and forcing perturbation

To represent the model uncertainty, typically two main input uncertainties of model variables are introduced into the standard model run: forcing uncertainty and parameter uncertainty. By integrating the perturbed forcing data and model parameters into the model and starting the model for a sample of initial states (Sec. 5.2.5), multiple runs can be performed. One model run per ensemble member is produced to derive a final uncertainty band for the model states.

To derive an uncertainty for the forcing data, the daily field of the forcing data set W5E5 as introduced in Sec. 4.6 are considered from 2000 to 2019 (see Sec. 5.2.5). Each daily field of precipitation and temperature is now perturbed, whereas long- and short-wave radiation is not perturbed following Eicker *et al.* (2014) and Schumacher *et al.* (2016). In the future one might consider the perturbation of the radiation as well, but the effect of adding the radiation perturbation on the forecast ensemble spread is expected to be small. In this framework, precipitation and temperature are perturbed with triangular distribution via a random sampling method. Precipitation has a multiplicative 10% factor for the error – thus, ranging between 0.9 and 1.1 – and the temperature has a $\pm 2°$ additive error (Schulze *et al.*, 2024).

Typically, model parameters in a standard run of WaterGAP are set as constant values by default except for the runoff coefficient because the model is calibrated against streamflow. In preparation for the data assimilation, the parameters are randomly perturbed under the consideration of ranges and probability distributions to derive uncertainty for the parameters. In contrast to the forcing data, the parameters are perturbed once. The ranges and distributions are derived from information from WaterGAP modeling experts, which provide lower and upper limits for each parameter based on multiple studies that examined the sensitivity of the parameters, among

Table 5.2: WaterGAP parameter names, their standard values, and the ranges and distributions used for the setup of the ensemble. This table is based on Döll *et al.* (2024) and internal exchange for the GlobalCDA project.

| **Parameter** (units) | **Standard WGHM value** | **Large range** | **Small range** | **Chosen distribution** |
|---|---|---|---|---|
| Precipitation multiplier | 1 | 0.5-2 | 0.77-1.3 | triangular |
| Net radiation mult. | 1 | 0.5-2 | 0.77-1.3 | triangular |
| Priestley-Taylor coeff. humid | 1.26 | 0.885-1.65 | 1-1.5 | uniform |
| Priestley-Taylor coeff. (semi)arid | 1.74 | 1.365-2.115 | 1.5-2 | uniform |
| MCWH | 0.3 | 0.1-1.4 | 0.2-0.8 | uniform |
| LAI mult. | 1 | 0.2-2.5 | 0.5-2 | triangular |
| Snow freeze temperature | 0 | -1-3 | -1-3 | uniform |
| Snow melt temperature | 0 | -3.75-3.75 | -3.75-3.75 | uniform |
| Degree-day factor mult. | 1 | 0.5-2 | 0.7-1.5 | triangular |
| Temperature gradient | 0.006 | 0.001-0.01 | 0.001-0.01 | uniform |
| Max. soil water storage mult. | 1 | 0.5-3 | 0.8-2 | triangular |
| Runoff coefficient | Variable | 0.3-3 | 0.5-2 | uniform |
| Maximum potential evapo. | 15 | 6-22 | 10-17 | uniform |
| GW recharge factor mult. | 1 | 0.3-3 | 0.5-2 | triangular |
| Max. groundwater recharge mult. | 1 | 0.3-3 | 0.5-2 | triangular |
| Critical precipitation | 12.5 | 2.5-20 | 7-13 | uniform |
| GW discharge coeff. | 0.01 | 0.001-0.02 | 0.002-0.01 | uniform |
| River roughness coeff. mult. | 3 | 1-5 | 2-4 | triangular |
| Active lake depth | 5 | 1-20 | 2-10 | uniform |
| Active wetland depth | 2 | 1-20 | 2-10 | uniform |
| Surface water discharge coeff. | 0.01 | 0.001-0.1 | 0.003-0.03 | uniform |
| Evapo. red. factor mult. | 1 | 0.33-1.5 | 0.33-1 | triangular |
| Net abstraction from SW mult. | 1 | -2-2 | 0.5-2 | triangular |
| Net abstraction from GW mult. | 1 | -2-2 | 0.5-2 | triangular |

others, Döll *et al.* (2003); Döll & Fiedler (2008); Hirabayashi *et al.* (2008); Werth *et al.* (2009); Schumacher *et al.* (2018); Hosseini-Moghari *et al.* (2020). Most of the parameters can spatially vary, whereas some parameters have the same value for each grid cell. Spatially varying parameters are introduced as multipliers and are determined to be triangularly distributed, the remaining parameters are uniformly distributed. In the case of a triangular distribution, the peak of the triangular distribution is at the standard WaterGAP value. Table 5.2 shows the standard WaterGAP values as well as the chosen distribution, small and large ranges for deriving ensemble members.

Two parameters require additional constraints to enable the model to simulate useful output: After distributing the snow parameters independently, it can happen that one ensemble member accidentally has a lower snowmelt temperature than the snow freeze temperature. In this case, the snow melt temperature is set equal to the snow freeze temperature.

### 5.2.5 Setup initial states and mean state vector

For running the data assimilation, it is necessary to derive initial states as a starting point and a temporal mean for reducing the simulated model TWS to anomalies of TWS to be comparable to GRACE/-FO observation. The initial states and a long-term mean state are derived from three consecutive phases:

1. the initialization phase,
2. the spin-up phase, and
3. the open-loop simulation.

In the initialization phase, the model is run as it is done for a standard run, which means without input files for the initial states and with unperturbed forcing data and parameters. The model then refers to internal starting values, but deriving a stable output that generates initial values takes time, which is the reason why initialization is run for multiple years. Typically, the phase lasts five or more years to simulate a realistic storage level. In this case, the initialization starts in January 1995 and ends in December 2000 (following one of the approaches in Schumacher, 2016), accordingly, its duration is six years. At the end of the initialization phase, important files are written out that are needed to restart the model with the same settings and to avoid another initialization for multiple assimilation runs. These files imply the water states, a specific variable that describes snow storage in varying elevations, and many other smaller variables, which are summarized as 'additional output input'. The additional output input stores, for example, land area fractions, water use variables, return flows, etc.

To run an assimilation via an ensemble-based filter, it is required to derive initial conditions for each of the ensemble members $N_e$. Thus, as next, the stored files after the initialization are reproduced $N_e$ times. A spin-up phase is now run for each of the $N_e$ files using the perturbed forcing data and parameters to evolve a spread in water storages for the ensemble. It needs to be long enough to derive a spread that is nearly constant in time. As a standard procedure, the spin-up phase in-house for WaterGAP is typically running from January 2001 to December 2002, which is two years of runtime. It should be noted that the spin-up phase starts with the same initial conditions for each ensemble member but ends with different conditions because of the forcing and parameter perturbation.

The final output files at the end of the spin-up phase, are used as initial conditions for the following open-loop simulation and the data assimilation. The open-loop simulation is a free model run from 2003 to 2019 without assimilating any observations but perturbing model parameters and forcing data. However, it is important to understand that it differs from the standard WaterGAP model run since it includes the perturbed forcing and parameter and starts with loading initial conditions, whereas a standard run has internal start values. Finally, a temporal mean from 2003 to 2019 is computed from the open-loop simulation as explained in Sec. 5.2.2. This mean derived from the state vector of the open-loop simulation is used for the transformation of TWS as derived in the model simulation, open-loop, and assimilation to the respective TWSA.

## 5.3 Challenges and tuning strategies

Running a data assimilation framework can be confronted by various challenges. One of the main problems is, that a large ensemble is computationally cost-intensive, which means, a large number of storage and computational resources are required. In fact, many assimilation systems are forced to use a limited number of ensemble members to approximate the forecast covariance matrix, i.e., that the number of ensemble members is much smaller than the number of state entries. This effect is called undersampling and can lead to the underestimation of errors in the analysis error matrix ($\tilde{\mathbf{P}}^a$) and further sampling errors (Evensen, 2003). Undersampling can be described as follows: the state vector can have very large dimensions (e.g., $O(10^7)$) and the ensemble size is much lower due to computational costs. In turn, undersampling can result in different problems as inbreeding, filter divergence, and the development of long-range spurious correlations.

Inbreeding means that the error covariances are systematically underestimated after an analysis step, thus, an offset in the weighting matrix Kalman Gain exists and influences the next update. Filter divergence is when the updated state has a large offset from the true state but a small ensemble spread leading the filter to be off-track (Houtekamer & Mitchell, 1998; Anderson, 2001; Hamill *et al.*, 2001). In other words, the forecasted state vector is unable to be adjusted by observations.

Inbreeding can lead to filter divergence, and it can develop over time, for example, when inbreeding decreases the analysis covariance, the spread of the ensemble is reduced. In the next step, a new model forecast based on the decreased analysis covariance is produced and weighted against the spread of the observation ensemble. Due to the decrease in forecast spread, the probability increases again that more weight is given to the forecast than the observation. Over time, the system is stuck in the filter divergence. The smaller an ensemble size, the larger the effect of inbreeding because the model space is not adequately spanned Ehrendorfer (2007).

Although they should be uncorrelated, spurious correlations in the state ensemble covariance matrix ($\tilde{\mathbf{P}}^f$) exist, which results from the above-mentioned problems of undersampling. In addition, the observation data can introduce spurious long-range correlations as well into the assimilation framework, that exist due to, for example, measurement constellations. The spurious correlations in the state ensemble covariance matrix then can lead to a small unphysical update.

A further but very case-specific and numerical problem that arose during the development of a global data assimilation framework is that the coupling of a model with the assimilation framework led to differences compared to when the model is run without the coupling. These differences were very small for a few integration steps in time but accumulated strongly for long periods.

Thus, there exist difficulties that can arise during assimilation and initiate further difficulties. To counteract these problems and tune the filter algorithm, the next sections introduce tools for stabilizing the data assimilation, i.e. the inflation factor variation, and localization technique, and introduce the adaptions that were made in the WaterGAP model for numerical stabilization.

**Localization**

To remove or reduce the spurious long-range correlations which increase the degrees of freedom for analysis and increase the size of the analysis correction, localization can be applied. Localization reduces such spurious correlations from the assimilation. In general, two techniques are used to apply localization. The first one is the covariance localization: Modifications are applied in the covariance matrix of the forecast $\tilde{\mathbf{P}}^f$ by computing the Schur-product (element-wise product of matrix entries) of $\tilde{\mathbf{P}}^f$ with a fixed smoother matrix ($\mathbf{S}$), which means the Kalman Gain matrix changes. In PDAF, the covariance localization is implemented in the way that the forecast covariance matrix is modified in the observation space. The new Kalman Gain matrix then yields in

$$\mathbf{K} = \mathbf{S} \circ (\tilde{\mathbf{P}}^f \mathbf{H}^T)(\mathbf{S} \circ (\mathbf{H}\tilde{\mathbf{P}}^f \mathbf{H}^T) + \mathbf{R})^{-1}. \tag{5.31}$$

The approach presents a nice opportunity to suppress error correlations in the state, however, covariance localization is restricted to the forecast error matrix and cannot account for spurious long-range correlations in the observation error matrix $\mathbf{R}$.

The second one is domain localization: The state update and ensemble transformation are fully local and use only observations within a distance around the region. This can also be seen as local updating of the state vector for different local subdomains of the state vector. As a local subdomain, one could, for example, define a coarse grid of 3°, 2° or take each single grid cell of the state vector. Then, the analysis step is applied for each subdomain iteratively by taking observations into account that are within a certain distance from the subdomain center as illustrated in Fig. 5.3. This means the forecast error matrix and the observation error covariance matrix $\mathbf{R}$ are both modified.

Evensen (2003) expects that to derive the same result, a global analysis requires a larger ensemble than the local analysis with domain localization but the runtime for a local analysis is longer because it iteratively searches the observations in specific subdomains. The aim is to choose a radius large enough to include significant observations but at the same time small enough to exclude the insignificant observations and thus exclude spurious measurements (Evensen, 2003). In other words, localization does not use all available information and thus is a trade-off between including as much information as possible by simultaneously excluding far-distance information from the framework. As opposed to the covariance localization, domain localization only directly reduces spurious long-range correlations for the observation error matrix, but does not directly exclude correlations from the forecast error matrix. Nonetheless, as it is required to define a subdomain for running domain localization, the correlations of the forecast error matrix are indirectly also reduced.

The observations that are limited by the considered localization radius, can also be weighted. The most simplest way is to give each observation within the radius a uniform weight. Furthermore, exponential weighting is also commonly used. The observations closest to the subdomain center have the highest weight and the weight factor very fastly decreases apart from this center.

In PDAF, the EnKF uses covariance localization and the ESTKF uses domain localization because $\tilde{\mathbf{P}}^f$ is not explicitly available. Vetra-Carvalho *et al.* (2018) emphasize that for some cases localization only does not bring enough stabilization to the assimilation system, thus inflation factors could be used together with localization (Vetra-Carvalho *et al.*, 2018) for improved performance.

#### Inflation factor

As discussed, inbreeding should be avoided or minimized. Its effect is that the variance is always underestimated because of finite ensemble size, sampling errors (unknown structure of $\tilde{\mathbf{P}}^f$), and model errors, which means, there is a decrease in model error over time when I assimilate. A possible solution is to increase the ensemble spread before the analysis via the inflation factor. As its name says, increasing the inflation factor works to inflate the forecast ensemble range again, which mathematically is a regularization.

Here, as well as in most cases, the inflation factor $\rho$ is conventionally chosen as a fixed number larger than one and is applied to the state vector and its mean via

$$\boldsymbol{x}^{(i)} = \rho(\boldsymbol{x}^{(i)} - \bar{\boldsymbol{x}}) + \bar{\boldsymbol{x}}, \tag{5.32}$$

where $\boldsymbol{x}^{(i)}$ is the $i$-th ensemble member state vector. The inverse of the squared inflation factor $\frac{1}{\rho^2}$ is denoted as the forgetting factor. Vetra-Carvalho *et al.* (2018) also list other possibilities for applying inflation facts, for example, in an additive manner (e.g., Ott *et al.*, 2004) or present studies that estimate temporally and spatially varying factors (e.g., Anderson, 2009).

The choice of the inflation factor strongly varies from assimilation framework to assimilation framework, because it depends on model dynamics, used filters, localization, and ensemble size (Vetra-Carvalho *et al.*, 2018). The PDAF provides a framework for applying a constant

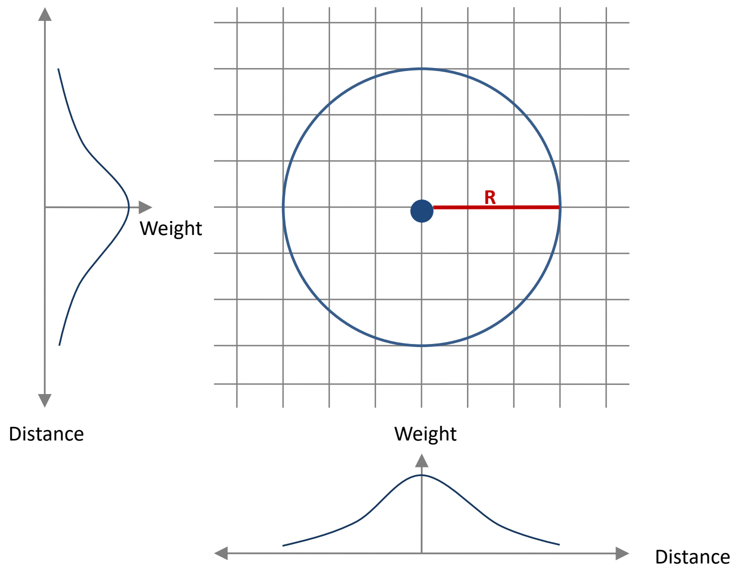


Figure 5.3: Illustration of a localization radius and weighting of observations within the radius under the consideration of the current subdomain (grid).

multiplicative forgetting factor, thus, this is mainly considered within this thesis.

**Model stabilization**

During global assimilation, more attention needs to be paid to stabilize the joint model-assimilation framework. When coupling the model to an assimilation framework, the information of the storages is communicated between both systems. However, the physical requirements of the model are not transferred to the assimilation system. It appeared that not limiting updated water storages made the data assimilation unable to stably run over the whole lifetime (2003 to 2019). At one point in time during the update, one of the storages eventually reaches a value outside of the boundaries. In some cases, a consecutive prediction step can capture this. Nonetheless, after too many cases in time, the system becomes unstable and storage values can become unrealistically large, which propagate to other storages as well. Thus, it is important to also include physical hard boundaries for the updated daily states that inform the next prediction step. In accordance with the limits within the model (Sec. 4.5), it was necessary to also apply the limits to the updated water storages after assimilation of observations and before they reenter the model for the next model prediction step.

As briefly introduced, the coupling between the model and assimilation framework can also lead to numerically small differences. In most cases, the numerical differences do not largely influence the assimilation but since a global data assimilation framework updates many state variables, in contrast to, for example, a regional approach, much more sensitivity exists. Concretely, in the case of WaterGAP it was required to further improve model code in two specific parts by desensitizing (1) if-clauses that compare according to equality of climate variables that change after assimilation, and (2) reading and writing of water states of the model when interrupted monthly. Both parts were jointly analyzed and tested together with the developers of WaterGAP from Goethe University in Frankfurt. An example of the desensitizing of if-clauses is that as soon as water storage derived from assimilation is compared to a variable from model simulation only, the differences do not need to be equal to zero but are allowed to be within a certain tolerance, e.g., $1 \times 10^{-12}$ mm. The ability to monthly interrupt the model was especially required during a phase where the coupled model-assimilation system needed to be run in offline mode for several reasons. Afterward, this technique was published in (Müller Schmied *et al.*, 2023). Nowadays, the model-assimilation framework runs in online mode, but the improvements during the previous developing phase still have a positive impact on the assimilation outputs.

# Chapter 6

# Analysis of the data assimilation framework

To further develop the assimilation framework that produced the Global Land Water Storage (GLWS) release 2, the GRACE/-FO processing and data assimilation options need to be elaborated on in detail, comprehensively assessed, and justified. I use statistical metrics (Sec. 6.1) to compare observation, model simulation, and data assimilation outputs to tune and refine the global data assimilation (Sec. 6.2). The final choice of options leads to a new release. The properties of this new data set are analyzed in detail on the spectral and spatial domains (Sec. 6.3.1). Thereinafter, relevant relations between different variables of the hydrological water cycle are derived using statistical approaches that enable extracting signal patterns and temporal lags (Sec. 6.3.2). With the help of the findings, one can analyze the temporal relations and spatial characteristics between seasonal and non-seasonal signals for different water compartments and in different spatial regions. Finally, the global assimilation is validated against independent data to uncover which parts of the assimilation already perform well with respect to a variety of metrics and which require improvements (Sec. 6.4). For all analyses, the TWSA from model simulations are derived by aggregating the single water compartments to TWS and removing the temporal mean from 2003 to 2019 of the (ensemble mean) open-loop simulation. The same step is repeated for deriving TWSA from open-loop simulation and assimilation but, in addition, an ensemble median for the 32 ensemble members is calculated to derive the water storage outputs. In case of TWS, a temporal mean needs to be removed to derive TWSA and to be able to compare the model simulations, the assimilation outputs and the observations against each other.

## 6.1 Methods for extracting and relating signal pattern and delays

### 6.1.1 Statistical metrics

To test the performance of different data sets, assimilation runs or model simulations against each other, various statistical metrics are presented here. These metrics are widely used and cover different application fields, for example, they can be a measure for similarities and phase timing in data sets or they can be used to provide range information. The metrics presented here are the Pearson correlation coefficient, the Nash-Sutcliffe efficiency (NSE), Root Mean Square Difference (RMSD), and Kling-Gupta Efficiency (KGE).

The Pearson correlation coefficient $\rho$ is a measure of how well two time series agree with each other by combining the variances of both and normalizing them via

$$\rho(x,y) = \frac{\sum_{t=1}^{n}(x_t - \mu_x)(y_t - \mu_y)}{\sqrt{\sum_{t=1}^{n}(x_t - \mu_x)^2}\sqrt{(y_t - \mu_y)^2}}, \tag{6.1}$$

where $x_t$ and $y_t$ are the two time series at time $t$, $\mu_x$ and $\mu_y$ are the respective temporal means, and $n$ is the total number of time steps. Thus, the correlation does not provide information about biases or how the amplitudes compare in magnitude. This metric can be used in a lagged manner, to estimate cross-correlations and thus provides information about the temporal delay between the two time series. The correlation values range from $-1$ to 1, where 1 is perfect correlation, 0 is no correlation at all, and $-1$ is that the time series are anti-correlated. Let us assume that we have two sinusoidal time series ($\sin \omega t$ with $\omega = 2\pi/year$), which are equal except that the second sinusoid is shifted in time by $\phi$ ($\sin \omega t + \phi$). The correlation between both time series for a shift of 30°, 60°, 90° would be 0.87, 0.5, and 0, respectively. Logically, after a full cycle, a correlation of one is given again. Assuming we have a time series of white noise, the correlation to a second equal time series would be one but as soon as a shift is introduced, the correlations would always be low (e.g., 0.2).

The NSE is a metric that was developed to describe the fit of model simulations and observations. The observation (here $y$) and simulations (here $x$) are used to derive a measure of how much the data can be explained by the simulation by computing variances via

$$\mathrm{NSE}(x, y) = 1 - \frac{\sum_{t=1}^{n}(y_t - x_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2}. \tag{6.2}$$

The values range from $-\infty$ to 1, where 1 means that observations and simulations perfectly fit together, 0 means the model simulations are equally accurate as the observations and negative values mean that the mean of the observations is a better predictor than the simulations. In contrast to the correlations, the NSE derives worse outputs in case the observations and model simulation experience different magnitudes of the amplitudes. Let us again assume that we have two sinusoidal signals ($2\pi$ per year) and shift the second time series by 30°, 60°, 90°. The NSE would then derive the values 0.73, $-1.13 \cdot 10^{-12}$, and -1, respectively. Considering the example of two white noise time series, the NSE is probably negative as soon as a shift is included.

As next, the KGE (Gupta & Kling, 2011) is introduced. It was developed to improve the performance of correlation and NSE and is calculated as

$$\mathrm{KGE}(x, y) = 1 - \sqrt{(\rho(x, y) - 1)^2 + (R_{\text{bias}}(x, y) - 1)^2 + (R_{\text{var}}(x, y) - 1)^2} \tag{6.3}$$

with $\rho$ being the Pearson correlation between simulations $x$ and observations $y$, and $R_{bias}$ and $R_{var}$ relating the mean and variance of observation and model

$$R_{\text{bias}}(x, y) = \frac{\mu_y}{\mu_x}, \quad R_{\text{var}}(x, y) = \frac{\sigma_y/\mu_y}{\sigma_x/\mu_x} \tag{6.4}$$

Döll *et al.* (2024) introduced an extra condition for the use of KGE with TWSA: $\mu_x$ and $\mu_y$ are assumed to be equal, thus, the bias ratio is set to one, and $R_{var}$ is simplified as $\frac{\sigma y}{\sigma x}$. This condition again removes the sensitivity of the metric towards biases. As model simulations, observations, and assimilation output can have strong biases due to the forcing data and ranges within the parameters are perturbed, I follow the condition for this work as well. It should further be mentioned, that this thesis does not concentrate on biases as the change in water is most interesting, thus the focus is on resulting trends, seasonality, or non-seasonal behavior instead. The KGE range is similar as for NSE ($-\infty$ to 1) with 1 indicating a perfect match of simulation and observation.

The final two metrics that will be used throughout this work are the Root Mean Square (RMS) and the Root Mean Square Difference (RMSD). The RMS is a quantity that provides information of the variability for a single data set and is computed via

$$\mathrm{RMS}(x) = \sqrt{\frac{\sum_{t=1}^{n} x_t^2}{n}}, \tag{6.5}$$

whereas the RMSD represents the variability of the difference signal between two different data sets and yields in

$$\mathrm{RMSD}(x,y) = \sqrt{\frac{\sum_{t=1}^{n}(x_t - y_t)^2}{n}}. \tag{6.6}$$

RMS and RMSD values are not normalized, thus ranges from zero to a maximum value in units of the time series occur. As already slightly discussed, it is important to understand that each metric is sensitive to certain aspects of the time series. For example, the correlation coefficients and the KGE (as explained before) are insensitive to a bias, whereas the NSE, RMS, and RMSD are sensitive to it. The correlation coefficients are insensitive against amplitude magnitudes, and all other metrics are. In addition, the RMS and RMSD are the only two metrics announced here that are not normalized.

### 6.1.2 Extracting dominant patterns

**Multi-linear regression**

Multi-linear regression is a statistical method to estimate hydrological signatures contained in the data set and describe them by linear additive operations. In geodesy and hydrology, a signal is often a composite of at least a bias ($a_0$), a linear trend ($a_1$), an annual signal ($b_1$, $b_2$), and a semi-annual signal ($c_1$, $c_2$) at grid cell $j$ and time $t$. The annual and semi-annual signal components are described via sinusoidal functions. The full signal $x_{j,t}$ then consists of the mentioned signatures by

$$\begin{aligned} x_{j,t} = a_{0_j} + a_{1_j}(t - t_0) + b_{1_j}\cos(\omega t) + b_{2_j}\sin(\omega t) \\ + c_{1_j}\cos(2\omega t) + c_{2_j}\sin(2\omega t), \end{aligned} \tag{6.7}$$

where $\omega = 2\pi$ typically represents a full year cycle. In some cases, accelerations are additionally considered. Trends and accelerations in GRACE/-FO TWSA exist due to various hydrological and anthropogenic processes, like the melting of glaciers or groundwater abstraction. Trends in the fluxes of precipitation, evapotranspiration, and runoff lead to accelerations in the GRACE/-FO TWSA (e.g., Eicker *et al.*, 2016). However, the study Gerdener *et al.* (2022) and further tests show that jointly estimating linear trends together with constant accelerations in the regression analysis could fail in meaningfully separating the two signatures. This is likely due to the short period of available GRACE/-FO observations because with longer time scales the acceleration effect is much stronger. Thus, for the purpose of extracting trends or sinusoidal signals as single signatures, the subsequent analysis excludes the accelerations. In the future, one might reconsider the joint estimation as a longer time series might enable a better separation.

In a least squares adjustment, the coefficients $a_0$ to $c_2$ can jointly be estimated. The annual amplitude and phases yield from the estimated $b$ coefficients (similar as in Phillips *et al.*, 2012; Forootan *et al.*, 2016) as

$$A_{\mathrm{ann}_j} = \sqrt{\hat{b}_{1_j}^2 + \hat{b}_{2_j}^2}, \quad \phi_{\mathrm{ann}_j} = \arctan(\frac{\hat{b}_{2_j}}{\hat{b}_{1_j}}). \tag{6.8}$$

Analogously, the semi-annual amplitudes and phases can be estimated from the $c$ coefficients as

$$A_{\mathrm{sem}_j} = \sqrt{\hat{c}_{1_j}^2 + \hat{c}_{2_j}^2}, \quad \phi_{\mathrm{sem}_j} = \arctan(\frac{\hat{c}_{2_j}}{\hat{c}_{1_j}}). \tag{6.9}$$

After applying a multi-linear regression (Eq. 6.7) for each climate variable separately, the amplitudes and phases of the two data sets can be compared to each other. For example, for relating precipitation (P) and groundwater (GW) peak times, the relative annual amplitudes ($A^{\mathrm{P,GW}}_{\mathrm{ann}_j}$) and relative annual phases ($\phi^{\mathrm{P,GW}}_{\mathrm{ann}_j}$) can be computed as

$$A^{\mathrm{P,GW}}_{\mathrm{ann}_j} = \frac{\frac{A^{\mathrm{GW}}_{\mathrm{ann}_j}}{\sigma_{A^{\mathrm{GW}}_{\mathrm{ann}}}}}{\frac{A^{\mathrm{P}}_{\mathrm{ann}_j}}{\sigma_{A^{\mathrm{P}}_{\mathrm{ann}}}}}, \quad \phi^{\mathrm{P,GW}}_{\mathrm{ann}_j} = \phi^{\mathrm{GW}}_{\mathrm{ann}_j} - \phi^{\mathrm{P}}_{\mathrm{ann}_j}, \tag{6.10}$$

where $\sigma_{A^{\mathrm{GW}}_{\mathrm{ann}}}$ and $\sigma_{A^{\mathrm{P}}_{\mathrm{ann}}}$ are the standard deviations of the groundwater and precipitation annual amplitudes computed from all grid cells. I consider only annual peak times in this thesis but the approach could also be transferred to relate semi-annual amplitudes and peak times via

$$A^{\mathrm{P,GW}}_{\mathrm{sem}_j} = \frac{\frac{A^{\mathrm{GW}}_{\mathrm{sem}_j}}{\sigma_{A^{\mathrm{GW}}_{\mathrm{sem}}}}}{\frac{A^{\mathrm{P}}_{\mathrm{sem}_j}}{\sigma_{A^{\mathrm{P}}_{\mathrm{sem}}}}}, \quad \phi^{\mathrm{P,GW}}_{\mathrm{sem}_j} = \phi^{\mathrm{GW}}_{\mathrm{sem}_j} - \phi^{\mathrm{P}}_{\mathrm{sem}_j}. \tag{6.11}$$

To extract non-seasonal residual signal of the original signal, the time series are separated into a "seasonal" and a "non-seasonal" part. To derive this separation, the estimated coefficients are used to add up the linear trend, the annual signal, and the semi-annual signal. When removing this seasonal signal from the original data, non-seasonal residual signals are obtained that include irregular events like extreme events. The seasonal and non-seasonal parts of the data are written as

$$\begin{aligned} \hat{x}_{\mathrm{sea}_{j,t}} &= \hat{b}_{1_j}\cos(\omega t) + \hat{b}_{2_j}\sin(\omega t) \\ &+ \hat{c}_{1_j}\cos(2\omega t) + \hat{c}_{2_j}\sin(2\omega t), \quad \text{and} \\ \hat{x}_{\mathrm{nonsea}_{j,t}} &= x_{j,t} - x_{\mathrm{sea}_{j,t}} - \hat{a}_{1_j}(t - t_0) - \hat{a}_{0_j}. \end{aligned} \tag{6.12}$$

Thus, the seasonal signal is the linear combination of restored annual and semi-annual signal and the non-seasonal signal contains the residuals after removing seasonality, linear trends, and the constant bias.

### Principal and Independent Component Analysis

The Principal Component Analysis (PCA) and Independent Component Analysis (ICA) are two statistical decomposition techniques that can also be used for dimension reduction (Preisendorfer, 1988; Hyvärinen, 1999). PCA is a second-order method, which means statistical moments up to the second order are considered. It aims to find dominant signatures in the data by maximizing the variance between the signatures. With this, main orthogonal modes are derived. It is common to disregard certain orthogonal modes beyond some threshold, with the understanding that these signal components contain mostly noise. In addition, the components of the data are aimed to be uncorrelated. Due to its easy implementation, the PCA is a method that is widely used and its application enabled identifying hydrological signatures contained in the data by many studies (e.g., Khaki *et al.*, 2018; Cerón *et al.*, 2020). The PCA allows one to decompose the observation or simulation $\mathbf{X}$ into

$$\mathbf{X} = \mathbf{P}\mathbf{E}^T. \tag{6.13}$$

$\mathbf{X}$ is of size $n$ x $j$, where $n$ is the number of time steps and $j$ is the number of grids (Preisendorfer, 1988). The temporal mean of the time series is removed per grid and an empirical auto-covariance matrix is obtained. The auto-covariance matrix $\mathbf{C}$ that is used to identify temporal modes can be computed as

$$\mathbf{C} = \frac{1}{n}\mathbf{X}^T\mathbf{X}. \tag{6.14}$$

The matrix $\mathbf{C}$ is a symmetric and positive (semi-) definite matrix. As the mean was removed and variances are considered in the computation of $\mathbf{C}$, the PCA includes the first two statistical moments in its computation. In addition, one could compute spatial modes but for simplicity, the following equations refer to $\mathbf{C}$ only. The auto-covariance matrix can now be decomposed into orthogonal modes by eigenvalue decomposition

$$\mathbf{CE} = \mathbf{E}\mathbf{\Lambda}^2. \tag{6.15}$$

This decomposition provides the eigenvectors $\boldsymbol{e}_i$ of the $i$-th mode – also known as Empirical Orthogonal Functions (EOFs) – stored in the eigenvalue matrix $\mathbf{E}$, and the eigenvalues $\lambda_i$ stored in the diagonal entries of the matrix $\mathbf{\Lambda}$. By projecting the observations or simulations on the EOFs with $\boldsymbol{p}_i = \mathbf{X}\boldsymbol{e}_i$ (in matrix form: $\mathbf{P} = \mathbf{XE}$), the Principal Components (PCs) are given. The final PCs can be normalized and the EOFs provided in units via

$$\boldsymbol{pc}_i = \frac{\boldsymbol{p}_i}{\sigma_{\boldsymbol{p}_i}}, \qquad \mathbf{eof} = \sigma_{\boldsymbol{p}_i}\boldsymbol{e}_i, \tag{6.16}$$

The variance fraction $\eta$ of the current mode $i$ of the total variance is derived by

$$\eta_i = \frac{100\ \lambda_i^2}{\sum_{i=1}^{n} \lambda_i^2}. \tag{6.17}$$

A typical pitfall in the interpretation of PCA is that the mathematical modes might not contain completely separated physical modes. Instead, different parts of physical modes may be contained in one mathematical mode, which, for example, Hyvärinen (1999) explained as the so-called mixing problem.

In contrast to PCA, ICA is a higher-order statistical technique that aims to derive independent statistical patterns from the data to derive physically independent processes in the mathematical modes. Thus, the ICA-independent modes are often hypothesized to better separate the physical pattern in the data as compared to PCA. Here, I follow the approach of Forootan & Kusche (2012), who define the ICA as a rotated extension of the PCA, used with singular value decomposition. Generally, the observations or simulations $\mathbf{X}$ can be seen as a combination of independent source signals with a mixing matrix

$$\mathbf{X} = \mathbf{AS}, \tag{6.18}$$

where $\mathbf{A}$ is the mixing matrix. Its pseudo-inverse can be related to $\mathbf{W}$ with $\mathbf{W} = \mathbf{A}^{\dagger}$ and $\mathbf{S}$ is a matrix that aims to contain independent source signals – or so-called independent components – in its rows. The formulation requires that the number of independent sources must be equal to or lower than the number of linear mixtures given in $\mathbf{X}$. Again, there are two options to perform ICA, either providing spatial or temporal independent sources. The spatial independent sources are determined in an iterative manner by decomposing

$$\mathbf{X} \approx \mathbf{X}_j = \bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{R}_j \mathbf{R}_j^T \mathbf{E}_j^T \tag{6.19}$$

with $j$ is the number of retained modes, $\mathbf{A} = \bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{R}_j$ and $\tilde{\mathbf{S}} = \mathbf{R}_j^T \mathbf{E}_j^T$. The matrices $\bar{\mathbf{P}}_j$, $\mathbf{\Lambda}_j$ and $\mathbf{E}_j$ are again the principal components, eigenvalue matrix and eigenvector matrix (also orthogonal base functions) determined from eigenvalue decomposition (Eq. 6.15 and description thereinafter) and $\mathbf{R}_j$ is the rotating matrix, that either rotates the principal components or base functions. $\mathbf{R}_j$ is determined from the cumulant tensor as

$$\mathbf{R} = \prod_{m=1}^{k} \mathbf{V}_m \tag{6.20}$$

where each $\mathbf{V}_m$ is a rotation matrix that diagonalizes the k-th qumulant matrix $\mathbf{Q}_m$ from a matrix set that represents the cumulant tensor ($\mathbf{Q} = \mathbf{Q}_1, \mathbf{Q}_2, ..., \mathbf{Q}_k$). More information about the cumulant tensor can be found in e.g., Cardoso & Souloumiac (1993). It is required that the rotation matrix is orthogonal given as

$$\mathbf{R}_j \mathbf{R}_j^T = \mathbb{1}_j. \tag{6.21}$$

The independent sources are computed from the rotation of $\mathbf{E}$ with $\mathbf{R}$

$$\tilde{\mathbf{S}} = \mathbf{R}_j^T \mathbf{\Lambda}_j^{-1} \bar{\mathbf{P}}_j^T \mathbf{X}_j = \mathbf{W}\mathbf{X}_j \tag{6.22}$$

To apply this decomposition for temporally independent sources, the aforementioned equations yield in

$$\begin{aligned} \mathbf{X}^T \approx \mathbf{X}_j^T &= \mathbf{E}_j \mathbf{\Lambda}_j \mathbf{R}_j \mathbf{R}_j^T \bar{\mathbf{P}}_j^T, \\ \text{with} \quad \mathbf{A} = \mathbf{E}_j \mathbf{\Lambda}_j \mathbf{R}_j, &\quad \tilde{\mathbf{S}} = \mathbf{R}_j^T \mathbf{E}_j^T \\ \mathbf{S} = \mathbf{R}_j^T \mathbf{\Lambda}_j^{-1} \mathbf{E}_j^T \mathbf{X}_j^T &= \mathbf{W}\mathbf{X}_j^T. \end{aligned} \tag{6.23}$$

**Process models**

An auto-regressive process model can help to identify temporal auto-correlations by relating some variable which is defined on a grid, at some time step, to the previous values of this variable at earlier time steps. This relation is shown hereinafter and includes the process coefficients $\phi$ up to the order $q$ of the observations or the simulations

$$x_{j,t} = \phi_{j,1} x_{j,t-1} + ... + \phi_{j,q} x_{j,t-q} + \epsilon_{j,t}, \tag{6.24}$$

where $\epsilon_{j,t}$ is the noise. The coefficients indicate how the previous time steps $t-1$ to $t-q$ are correlated to the current time step $t$ and can be computed via a least squares adjustment per grid $j$, for example by using TWSA from GRACE/-FO. The optimal order $q$ is usually selected by using information criteria. There exist several of these information criteria in the literature, for example, the Akaike and the Bayesian information criteria (Rissanen, 1978; Akaike, 1998). In general, for the computation of information criteria, a likelihood needs to be estimated that provides a measure of how good models of varying orders explain the data.

In the same manner, one can relate multiple variables in a similar scheme process model, for example, fluxes and storages. Changes in storages in a month $t$ and grid $j$

$$s_{j,t} = c_j \cdot s_{j,t-1} + d_j \cdot p_{j,t}, \tag{6.25}$$

can be referred to precipitation of the same month $p_{j,t}$ and previous month storages $s_{j,t-1}$ following Humphrey & Gudmundsson (2019) and Gerdener *et al.* (2022). The process coefficient $c_j$ is unitless, whereas the process coefficient $d_j$ relates flux and storage units. The functional relation of current month vegetation $v_j(t)$ to current month storage $s_{j,t}$ and previous month vegetation $v_{j,t-1}$ is

$$v_{j,t} = e_j \cdot v_{j,t-1} + f_j \cdot s_{j,t}. \tag{6.26}$$

Similar as with the storage, the $e_j$ coefficient is unitless and the $f_j$ coefficient relates storage to vegetation units.

### 6.1.3 Outlines of hydrological regions and vegetation biomes

Throughout the analysis of the data assimilation and the prototype drought monitoring system, there are several possibilities to obtain results for specific regions. The assimilation updates the model at grid level. Nonetheless, for analyzing the performance of the assimilation framework and the drought detection for different hydrological regions or vegetation regimes, I introduce here three ways how to choose coherent regions:

1. Polygon outlines (basins, lakes, aquifers) from HydroSHEDS
2. Vegetation biomes from the Copernicus Global Land Service
3. Coherent regions in terms of temporal correlations in GRACE/-FO TWSA from clustering

The platform HydroSHEDS[1] provides global outlines of hydrological basins (Lehner *et al.*, 2008). I use drainage basins larger or equal to an area of 90.000 km$^2$ which results in a total number of 140 basins worldwide (Fig. 6.1, left) that cover 51% of the global land area (except Greenland and Antarctica). The resolution of the provided basin polygons is 30". The top five largest basins according to the area as found in the HydroSHEDS database are the Amazon, Congo, Mississippi, Nile, and La Plata basins. HydroSHEDS also provide nationwide boundaries or lake polygons, which can be used for specific case studies.

Since hydrological basins are developed by considering the topography of watersheds but not vegetation coverage, varying vegetational and hydrological conditions can be represented within one basin. To relate attributes and peaking times of water storages to environmental conditions, ecological processes, and biotic communities, the Copernicus Global Land Service provides a cluster layer map – the Global biome cluster layer for the 100 m global land cover processing line – covering 73 biomes (Buchhorn, 2022), which is shown in Fig. 6.1 (right). In contrast to the outline of the basins, the layer covers the whole globe without including too many clusters with very small sizes. This means that especially more dry regions are included than for the hydrological basin layer, for example, large parts of Australia.

The third possibility of deriving coherent regions is via clustering and can be performed with various clustering techniques, for example, K-means, DBSCAN Agglomerative, or Expectation-Maximization (EM) clustering. In general, clustering has the advantage that it directly adapts to the data but the underlying map needs to be carefully chosen depending on the application. In this thesis, the EM clustering is used for finding regions with coherent temporal correlations in GRACE/-FO TWSA later on (Sec. 7.1).

The EM algorithm has the advantage of low computational costs per iteration step, easy implementation, and a good overall convergence, thus it is considered for use in this work. The EM is a two-step algorithm for clustering that consecutively computes the expectation step and the maximization step in an iterative procedure and is widely described and analyzed in the literature (e.g., Dempster *et al.*, 1977; Redner & Walker, 1984). Code is provided via Chen (2018) and adapted to the current framework. In general, the expectation step (E-step) assumes that the data point $x_t$ is derived from a mixture of $k$ Gaussians and one expresses the likelihood of the Gaussian mixtures as follows

$$z_{tk} = \frac{1}{Z_t}\pi_k \mathcal{N}(x|\mu_k, \Sigma_k), \tag{6.27}$$

given the Gaussian mixture model parameters of the mean $\mu_k$, covariance $\Sigma_k$, and mixing coefficients $\pi_k$. The number of Gaussians is equal to the number of clusters and needs to be

[1] https://www.hydrosheds.org/ (last accessed 16.03.2024)

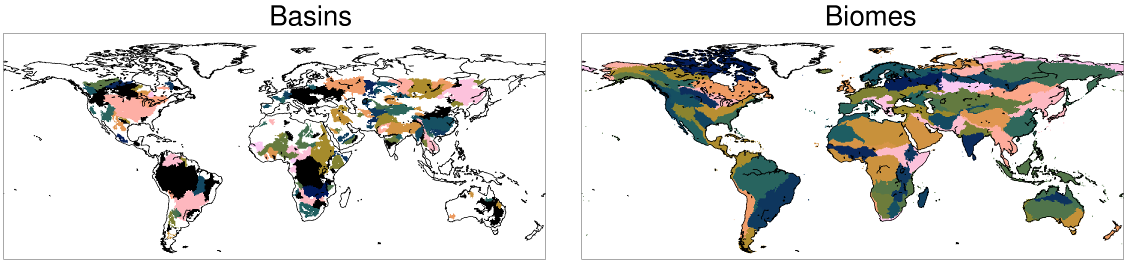


Figure 6.1: 140 drainage basins larger than 90.000 km$^2$ from the HydroSHEDS (left) and 73 classified biomes derived by the Global biome cluster layer (right).

predefined for the algorithm. In the maximization step (M-step), for each of the Gaussian mixtures a mean $\mu_k$, covariance $\Sigma_k$, $\pi_k$ are estimated via

$$\mu_k = \frac{1}{N_k} \sum_t z_{tk} x_i, \tag{6.28}$$

$$\Sigma_k = \frac{1}{N_k} \sum_t z_{tk} (x_t - \mu_k)(x_t - \mu_k)^T, \tag{6.29}$$

$$\pi_k = \frac{N_k}{N}, \tag{6.30}$$

where $N$ is the total number of data points and $N_k$ is the number of assigned data points per cluster

$$N_k = \sum_i z_{ik}. \tag{6.31}$$

## 6.2 Refinements of the assimilation framework

The GRACE/-FO processing of Level 2 SHC to Level 3 TWSA (Chap. 2) and the assimilation framework (Chap. 5) provide multiple possible settings for this work. For the standard settings of GRACE/-FO, it was possible to derive a final choice, this includes the maximum degree and order, filtering, the choice of a GIA model, the reference period, and replacement of lower degree coefficients as shown in Sec. 2. However, for a few GRACE/-FO settings that do not belong to the standard processing chain of official processing centers, it is unclear whether to include them because their influence on the assimilation is unknown. This will be further elaborated in this section.

In addition, for the assimilation, I avoid testing different ensemble sizes because the amount of storage gets very large for global assimilation (e.g., storing large sizes of forcing ensembles and updated states ensembles). Instead, the tuning options presented in Sec. 5.2 are tested, i.e., variations of inflation factors, filter algorithm, and localization techniques. This section aims to find suitable final processing for the global data assimilation framework for which the code can be frozen for further applications.

The base for the following tests are the options that were used for the release 2 GLWS. This includes defined "standard" options: no earthquake correction or leakage correction is included, the EnKF filter algorithm is applied, a small parameter range is used, and no localization of inflation is applied. In the following, I step-wise include one of the above-mentioned "non-standard" tuning options by keeping the other options as standard. One exception is the inflation factor. This factor is tested in conjunction with the ESTKF with localization as these two can be mutually dependent. The tested options are summarized in Tab. 6.1.

Table 6.1: Settings for the GLWS releases 2 and options that are step-wise tested in this thesis for creating the new GLWS release 3.

| | GLWS | |
|---|---|---|
| **Option** | **RL002** | **Test for RL003** |
| Earthquake correction | No | Yes |
| Lake/Reservoir relocation | No | Yes |
| Rescaling | No | Yes |
| Assimilation filter | EnKF | ESTKF |
| Localization | No | Yes |
| Inflation | No | Yes |

### 6.2.1 Effect of non-standard settings on the assimilation

This section now evaluates the use of three possible non-standard processing steps on the GRACE/-FO TWSA that do not belong to the standard processing: (1) the correction of earthquakes, (2) the removal and relocation of lakes and reservoirs via linear trends, and (3) rescaling of amplitudes and their effect on the assimilation output.

#### Correcting for leakage

Subsequently, the effect of reducing the leakage signal in GRACE/-FO data for data assimilation is analyzed using the two possibilities presented in Sec. 2.3.2: (1) The RECOG-LR lake and reservoir correction for leakage (Deggim *et al.*, 2021) and (2) rescaling via computing rescaling factors from hydrological modeling (e.g., Landerer & Swenson, 2012). At first, the lake and reservoir removal and relocation is in the following applied to the GRACE/-FO 4° spatial grid and the relevance for the assimilation analyzed. The 4° grid is chosen as this is exactly how the observations are assimilated into the model. It is important to remember that the lake and reservoir relocation fields cannot be removed per month as the monthly correction fields are provided only up to 2016. Instead, the trends from the removal and relocation data set (Sec. 2.3.2) are used to account for the lake and reservoir leakage effect.

The GRACE/-FO linear trends for TWSA for the Great Lakes in North America on 4° are much higher, as soon as the removal and relocation of lake leakage are included (Fig. 6.2, top right) as compared to not applying the relocation technique (top left). Thus, the relocation adds much more trend signal to the observations. However, the four-degree grid is very coarse and represents huge trends for the whole grid, although the lake signal is much more local. It only very broadly represents the outline of the lakes. When assimilating the GRACE/-FO TWSA corrected for lakes and reservoirs into the WaterGAP model (bottom right), the linear trend changes are much stronger for some lakes compared to an assimilation of uncorrected TWSA (bottom left). Thus, a very positive finding is that the assimilation transports water mainly to the intended lakes instead of removing trends for each of the 0.5° cells within one 4° cell equally. The linear trends are much higher for Lake Huron, Lake Erie, and Lake Ontario after relocation than without relocation. For Lake Superior and Lake Michigan, removal and relocation do not significantly change the trends derived from assimilation although the GRACE/-FO grid shows large trend changes nearby. Lake Superior is the largest of the five lakes and is – as well as Lake Michigan – not fed by the other lakes. Possibly, the water storage of Lake Michigan and Lake Superior is already close to their maximum capacity before assimilation. A strong change in storage is then not directly leading to water storage change but water above the maximum water capacity (Sec. 4.5) is further routed to Lake Erie, Lake Huron, and Lake Ontario.

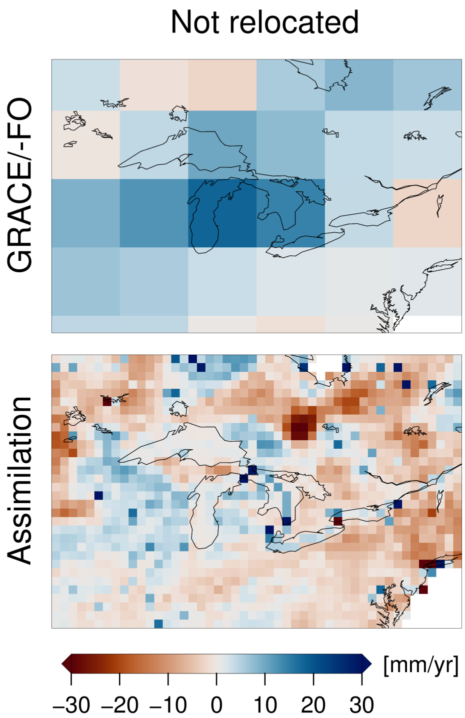


Figure 6.2: Linear trends [mm/yr] for TWSA from the 4° GRACE/-FO (top) or 0.5° assimilation (bottom) estimated for 2003 to 2019 with (right) and without (left) considering the lake and reservoir removal and relocation via linear trends.

In interim summary, the lake and reservoir relocation does indeed have a significant impact on some lakes and thus is not only valuable for correcting GRACE as shown in Deggim *et al.* (2021) but also for GRACE/-FO observations that are assimilated. Nonetheless, it should be mentioned that the relocation applied in this thesis is only done via linear trends and neglects seasonal and non-seasonal signatures. Thus, the variability of trends, seasonal, and non-seasonal residual variability, are exemplarily analzed for Lake Superior. For a 4° grid cell located in Lake Superior (the largest lake of the U.S. Great Lakes), the linear trend in the GRACE/-FO TWSA has a much larger variability than the seasonal signal and the interannual variability with RMS values of 91.41 mm, 21.61 mm, and 62.59 mm, respectively. This underlines that the linear trends can be the most dominant signal contained in the data and a corresponding correction of leakage via trends with RECOG-LR is reasonable.

There are further possibilities to integrate the relocation data set in a different way than using the trends and applying them to the 2003 to 2019 period. For example, removing monthly TWSA fields could be applied from 2003 to 2016, and removing TWSA trends could only be used for the three remaining years up to 2019. However, this procedure would complicate the interpretation of the results as the results before and after 2016 are underlying different methods. Instead, it is here suggested to further improve the relocation in the future by extending the correction data set in time to relocate the full TWSA signal. In addition, a further possibility for improving the relocation data set RECOG-LR is by integrating dynamic lake areas. for the example of the Great Lakes, strong changes in the lake area are not expected but the dynamic lake shapes could be much more important for small lakes that experience strong decreasing or increasing linear trends.

Second, the effect of rescaling on GRACE/-FO observation and the data assimilation is further elaborated. When comparing rescaling factors from six different models in Fig. 6.3 (models introduced in Sec. 4) – estimated es temporal median from 2003 to 2016 – large differences in

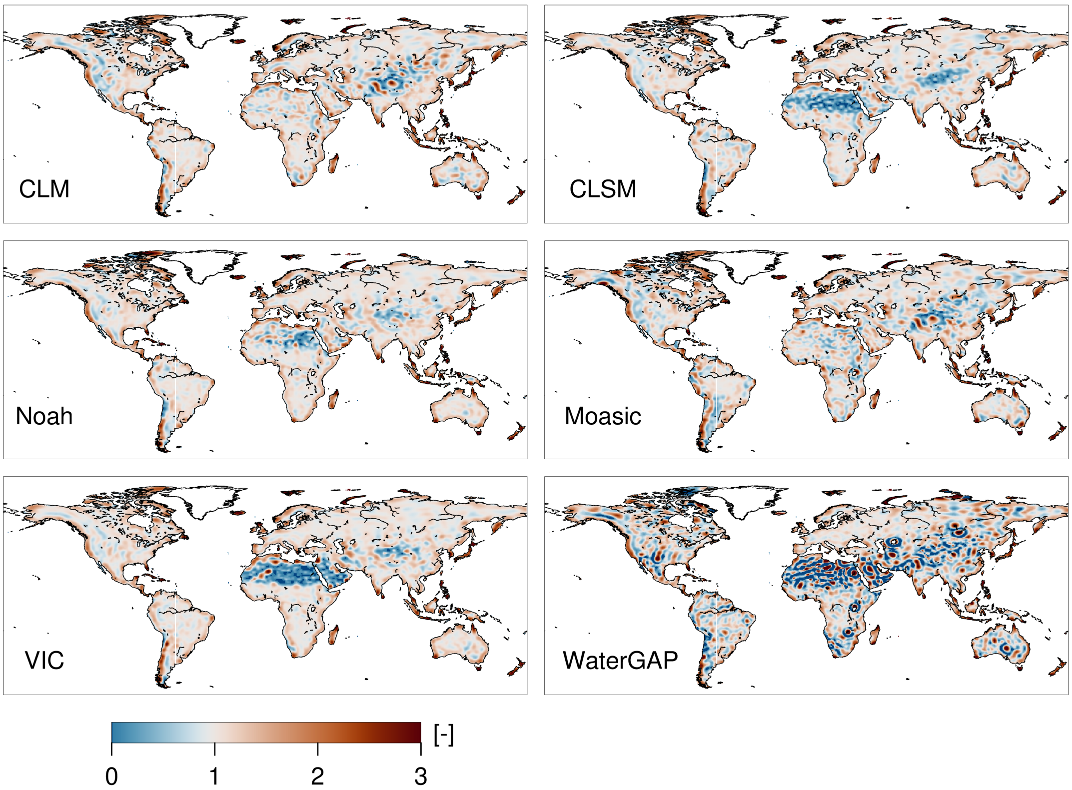


Figure 6.3: Rescaling factors from six different models: CLM, CLSM, Noah, Mosaic, VIC, and WaterGAP. The rescaling factors are unitless and are computed as median from monthly rescaling factors for 2003 to 2016 (Sec. 2.3.2.

the rescaling factors can be observed especially for surface water areas like the Amazon River and in arid regions like the Sahara. The explanation for the differences in rescaling factors for surface water might be that very different representations of water storages exist in the models. For example, some models do not have an explicit representation of groundwater or surface water storage. These different representations could lead to different TWSA simulations and thus the rescaling factors can strongly vary. The differences in rescaling factors for arid regions probably result from a low signal-to-noise ratio, thus the models might differ in these regions much more because the noise dominates.

When I modify the length of the time series of simulated TWSA, a percentage change in global mean rescaling factors between 2003 to 2008 and 2003 to 2019 of 3.47% becomes visible, exemplarily calculated for the WaterGAP model (Sec. 4.2). Only minor changes are found when comparing rescaling factors estimated for 2003 to 2016 versus 2003 to 2019 (0.61%). As this thesis is about the assimilation with WaterGAP, only this model will be used for the rescaling factor estimation to avoid including other incomplete model process representations in the rescaling. The considered period for computing rescaling factors is chosen to be 2003 to 2019 as this is in line with the considered time series for the assimilation.

Rescaling factors for TWSA were estimated from the WaterGAP model and are applied to the 4° GRACE/-FO grid. Fig. 6.4 shows the difference between TWSA annual amplitudes with and without rescaling either estimated via multi-linear regression (Sec. 6.1.2) for the 4° GRACE/-FO (left) or for the assimilation (right). The amplitude differences are further separated into positive and negative values to better distinguish between regions that increase in amplitude and regions that decrease in amplitude after rescaling. If this separation is not given, it would be difficult to distinguish positive and negative amplitude differences from each other. Positive values indicate that the annual amplitude after rescaling is larger than without rescaling and negative values indicate that the annual amplitude after rescaling is smaller as compared to no rescaling.

The positive amplitude differences (top row) show exactly what was expected: coastlines and surface water areas have a higher amplitude after rescaling in the GRACE/-FO 4° grid and this is also conveyed to the assimilation, where similar regions show an increase in amplitude after rescaling but with much fine spatial scale. The Sumatra-Andaman region, the Tohoku region, and the U.S. Great Lakes are especially interesting. For all of them, a high rescaled amplitude is identified.

For some regions, the amplitude also decreases in size after rescaling (bottom row). The reason is that the leakage effect does not only lead to a loss of signal in areas like the coast or surface waters but also, for example, the surrounding area of surface water bodies gains signal as the leakage smears the water body signal inland (leakage-out and leakage-in, Sec. 2.3.2). This effect is very prominent, for example, for the Amazon River, where the 4° grid cells that cover the river show positive differences and the surrounding cells show strong negative differences in amplitudes. It is further striking that via rescaling the amplitude in the Sahara is decreased, and thus the noise is also decreased. Therefore, the transformation of the SHC via spherical harmonic synthesis increases the noise.

At this point, a decision is required if rescaling or the leakage relocation should be applied to account for the leakage in the data and minimize its effects on assimilation. The RECOG-LR data set presents a great opportunity to correct for leakage in lakes and reservoirs. As the removal and relocation data set might be further developed for GRACE/-FO and dynamic lake areas, future studies might show even more significance for its application. In addition, the relocation data set itself without the removal can serve as an independent data set for evaluation against the surface water from assimilation. However, since this global assimilation also includes many cells of TWSA along coastlines, the rescaling is chosen instead. Nonetheless, the lake and reservoir removal and relocation data sets present a further possibility for the evaluation of surface water in this thesis, and future work could consider the monthly correction fields.

#### Earthquake correction

As was shown in Sec. 2.3.2, an earthquake signal contained in a TWSA time series is for some regions dominantly included and might bias the detection or correct analysis of other hydrological signals, for example, trends and drought events. Thus, when using GRACE/-FO only for non-earthquake application fields, one should correct the signal for large earthquakes. Nonetheless, at this point it is not clear, whether such a correction for earthquakes is also relevant for assimilating GRACE/-FO data globally.

Prepared for assimilation on a 4° grid, the GRACE/-FO TWSA spatially averaged for West Malaysia is shown in Fig. 6.5 (left; and for Japan in the appendix Fig. A.1). The observed GRACE/-FO TWSA without earthquake correction indicates a jump in the times that was also identified before when using the 0.5° spatial GRACE/-FO TWSA grid (Sec. 2.3.2). They also compare very well to the time series shown in Deggim *et al.* (2021), where the GRACE/-FO TWSA was computed from spatially averaging the 0.5° up to 2016 instead. Thus, the representation on the 4° does not significantly change the interpretation. After correcting for the earthquake signal, no co- and post-seismic jump is included in the time series. When assimilating the GRACE/-FO 4° into WaterGAP, no clear differences between earthquake uncorrected and corrected TWSA are identified (Fig. 6.5, right). A small shift between the time series is apparent

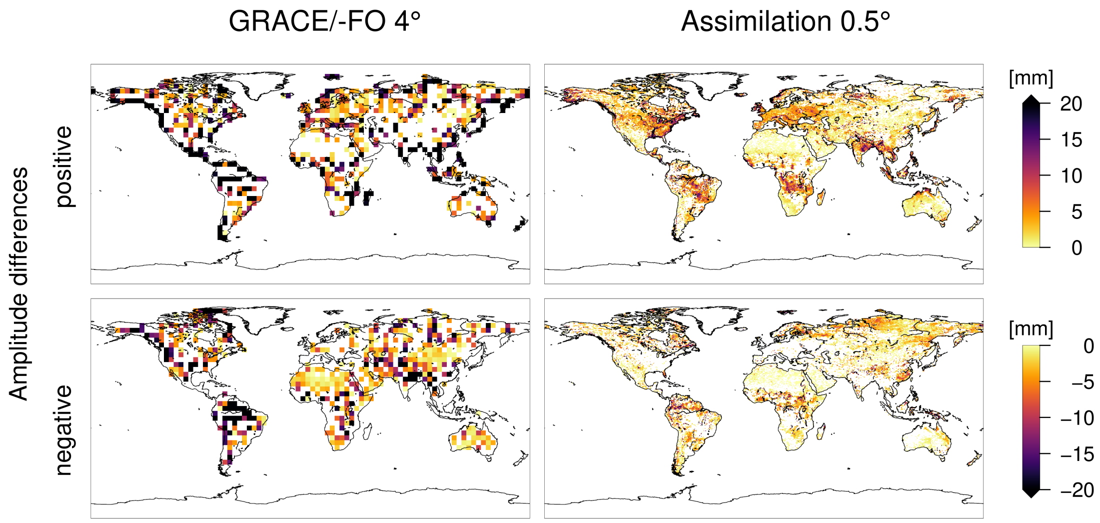


Figure 6.4: Annual amplitude differences [mm] between rescaled and not rescaled TWSA derived from the 4° GRACE/-FO (left) or the 0.5° assimilation (right). The amplitudes are estimated by multi-linear regression from 2003 to 2019. To better distinguish between regions that increase in amplitude and regions that decrease in amplitude after rescaling, the amplitude differences are separately shown for positive and negative differences, respectively.

but not significantly large. It should be noted, that the variability in TWSA is much smaller in GRACE/-FO as compared to the assimilation, especially when comparing annual amplitudes. This is not surprising because both regions are located at the coast and the leakage of the ocean into the land masses is a very common reason, why the amplitudes of the GRACE/-FO time series are much smaller than those of the model. At this point, the rescaling is not included but the analysis of the earthquake correction for the assimilation shows that for such coastal regions, it could help to improve the effect of earthquake correction for assimilation.

When including rescaling before evaluating the effect of earthquake corrections for assimilation, the small shift between uncorrected and corrected TWSA decreases slightly (Fig. A.2). This small decrease was expected, because rescaling enlarges the amplitudes, especially at the coastline, which makes the observations already closer to the larger amplitude of the model simulations. At the same time, the effect of earthquake correction is still small probably because model simulations are not biased by earthquake signal anyways. Thus, under this setup of the assimilation, only minor influences of the earthquake signal in the observations also move to the assimilation when considering the spatially averaged TWSA.

A comparison of linear trend differences for the input 4° grid of the observations with and without earthquake removal (Fig. 6.6, left) provides the opportunity to analyze the effect of earthquake removal on the spatial domain. The trend differences show clearly the regions of applied correction fields surrounding the Sumatra Andaman epicenter and the Tohoku epicenter. These trend differences do not show up in the assimilation as clearly as for the GRACE/-FO data. Instead, the trend differences in the assimilation are larger than the GRACE/-FO trend differences for many regions. It must be mentioned, that these trend differences are also found in the far distance of the earthquake-affected regions that do not show up for the observations only. Thus, it is assumed that the assimilation in general derives larger changes in trends due to numerical differences in the assimilation setup than the earthquake correction. This means that the earthquake correction might be too small in magnitude to be visible in the gridded final

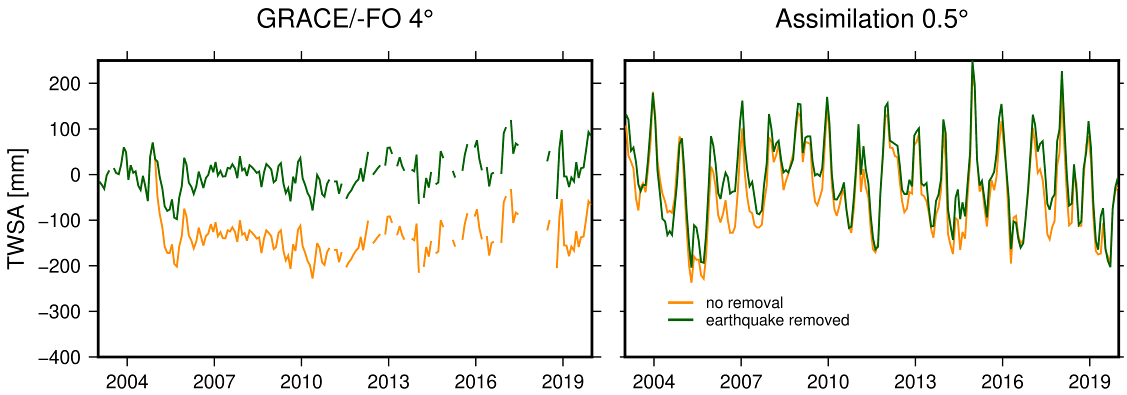


Figure 6.5: TWSA [mm] with and without earthquake correction derived for the 4° GRACE/-FO observations (left) or the 0.5° assimilation of the 4° GRACE/-FO TWSA (right) spatially averaged for West Malaysia.

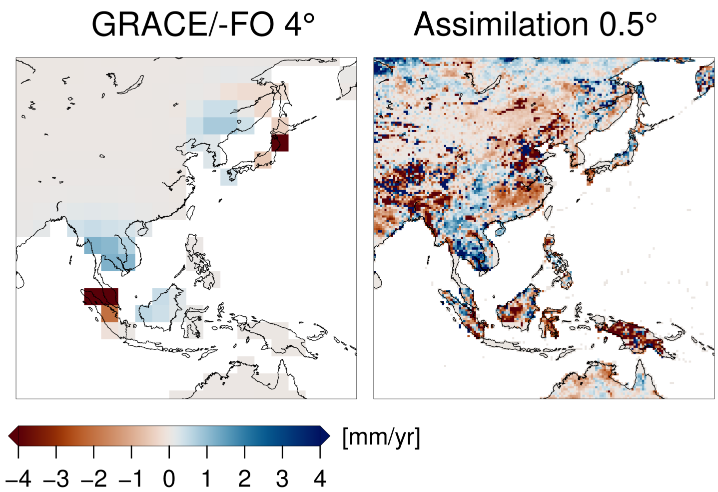


Figure 6.6: Linear trend differences [mm/yr] between uncorrected and earthquake corrected TWSA derived from the 4° GRACE/-FO (left) or 0.5° assimilation with corrected and uncorrected GRACE/-FO TWSA. The trends are estimated by multi-linear regression from 2003 to 2019.

assimilation product.

In conclusion, the earthquake removal in the GRACE/-FO does not have a significant contribution to the assimilation. However, the results might change for a different GRACE/-FO aggregated grid together with rescaling for the assimilation but running the global assimilation on, for example, a 2° spatial grid is currently not possible. The model cannot handle the updates and gets unstable (Sec. 5.2.3 and 5.3) and GRACE/-FO data have only a native resolution of 300 km, thus increasing the grid sizes will have its limitations for improvements because at some point no physical information will spatially be added. This hopefully changes in the near future, when TWSA observations from the MAGIC mission become available, which we expect to have a spatial resolution of 0.5°. Since no negative aspects of correcting for earthquakes were observed, the earthquake correction is from now on included in the processing chain.

### 6.2.2 Tuning and error representation of the assimilation

After carefully exploiting the GRACE/-FO processing choices, a close look is now taken at the tuning options for the assimilation itself. All options shown here can be categorized into two main classes, which are the options for including observation uncertainty into the assimilation framework and options for representing model prediction (or update) uncertainty. This includes a performance study on filter choice, variable observation error information, localization, parameter ranges, and inflation factor variations.

#### Effect of exchanging the filter algorithm

Before introducing further techniques for including full error information into the assimilation framework, the effect of replacing the EnKF with the ESTKF is analyzed. This enables a better understanding of which influence the choice of filter algorithm itself has and which effect further tuning options like localization might have. The ESTKF is a good choice for comparing against the EnKF because it combines two filter algorithms to derive advantages of both as described in Chap. 5 and also uses representing model error information via ensembles.

When comparing the assimilation results for TWSA from the two filters EnKF and ESTKF, the mean RMSD for the 140 basins (Sec. 6.1.3) of TWSA is found to be 17.36 mm, which is small, whereas the global mean RMSD is larger with 38.28 mm. For the global mean RMSD, all global 0.5° grid cells are averaged, therefore, all values from arid to humid are considered. This is not the case for the RMSD for basin averages, because the basins are including fewer arid regions. Thus, it is assumed that the TWSA from single arid cells contained in the globally averaged RMSD lead to the differences compared to the RMSD from basin averages when using the two filters. However, overall regional differences are small as shown by the RMSD for basin averages. Fig. 6.7 shows the differences between linear trends estimated via EnKF and trends estimated via ESTKF. For most of the global land area, the differences are below 2 mm per year, which is very low. Only for some regions, trend differences of 2 mm per year or higher can be found, mainly in areas with rivers and wetlands, for example, for a small part of the Amazon River. Nonetheless, these trend differences are nearly everywhere below 10 mm per year and are still small compared to the usual extent of trend in those regions. In median, the trend differences are usually below 38% of the full trend signal (see also subsequent results for global trend analysis in Sec. 6.3.1).

In conclusion, according to the TWSA analysis of basin averaged RMSD values and the linear trends differences, a change in the filter from EnKF to ESTKF has only an insignificant influence on the assimilation results. For these comparisons, the observation error is only considered by the main diagonal of the variance-covariance matrix as it was done for release 2 of GLWS (Gerdener *et al.*, 2023a). Further investigations with a full variance-covariance matrix are following next.

#### Integrating observation correlations

The release 2 of the global assimilation with GLWS (Gerdener *et al.*, 2023b) was produced by including observation variance per grid, but neglected correlations between grid cells (correlations in the spherical domain were included). But I hypothesize that especially for GRACE/-FO it is important to include spatial correlations in the data assimilation because neighboring cells are correlated due to the mission constellation and post-processing. Here, the effect of including observation correlations in the data assimilation is tested. The observation error information and correlations are stored in the variance-covariance matrix. An exemplary comparison between the diagonal elements of the variance-covariance versus the full variance-covariance matrix GRACE/-FO TWSA for January 2004 (Fig. 6.8) shows that including the full observation error

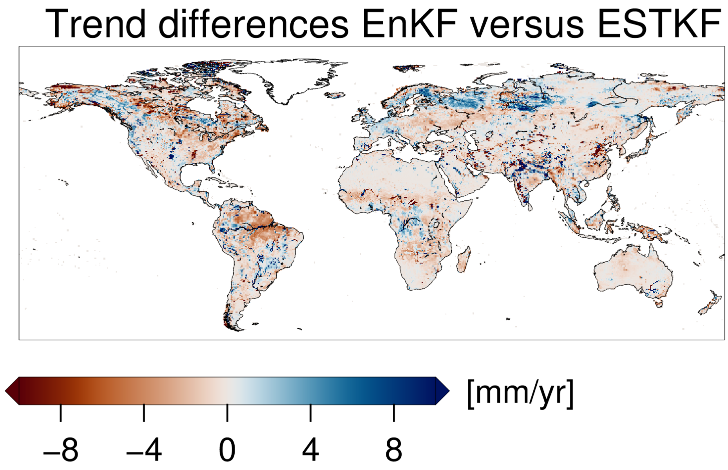


Figure 6.7: Differences between the TWSA linear trends [mm/year] estimated from EnKF and ESTKF assimilation for 2003 to 2019.

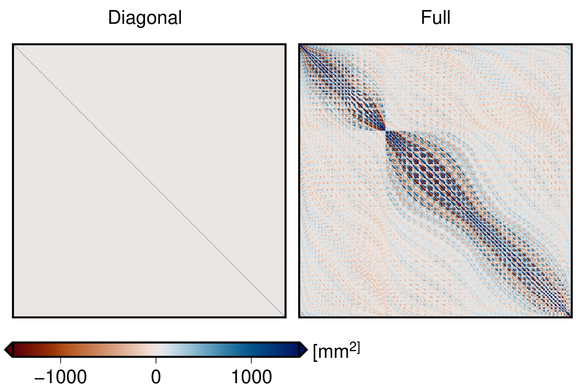


Figure 6.8: Diagonal (left) and full (right) variance-covariance matrix [$mm^2$] of the 4° GRACE TWSA grid over land for January 2004.

information stores much more information than the diagonal error information and a systematic pattern can also be identified. The correlations between neighboring grid cells in the variance-covariance matrix of the TWSA observation from January 2004 is significantly high with a spatial average value of about 0.83. When computing the correlations not for the direct neighboring but for next-neighboring grid cells, an average correlation of 0.54 is found, which is lower than the direct neighboring cells but is still high.

The EnKF was the original filter taken for the second release of the data assimilation. Assimilating the GRACE/-FO TWSA with full variance-covariance matrices into the model without localization led to strong instabilities and the output did not show realistic results. For example, linear trends derived from the assimilation with EnKF with full variance-covariance matrices (Fig. 6.9, left) show strong negative trends in the Okavango basin, although the observed TWSA trends for the basin indicate positive trends. A second example is that positive trends are shown in India, where groundwater depletion usually leads to negative trends (Rodell *et al.*, 2018). A closer look into TWSA basin averages for the Okavango basin clearly shows that the assimilation prediction and update both are off-track as compared to the open-loop simulation and GRACE/-FO. Outliers seem to first appear in the prediction step of the model and the update

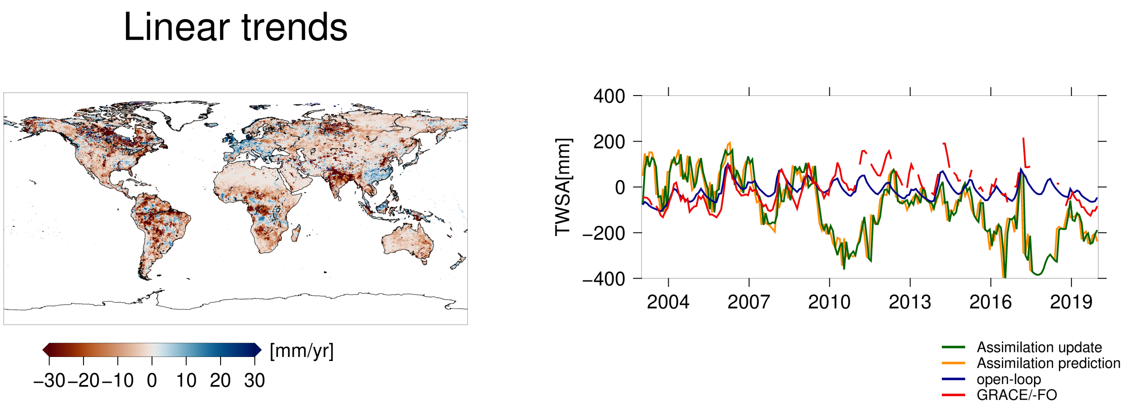


Figure 6.9: Linear trends [mm/year] of TWSA from assimilating GRACE/-FO TWSA with full variance-covariance matrices (left) and basin averages of TWSA [mm] for the Okavango basin for GRACE/-FO, the open-loop simulation, the assimilation prediction and the assimilation update. The assimilation was run with the EnKF filter without localization.

cannot adapt the states so much that they get close to more realistic values.

The possible reasons for this were introduced in Sec. 5.3, for example, spurious long-range error correlations can lead to filter divergence. The ESTKF explicitly computes the inverse of the variance-covariance matrix and avoids perturbation of the observation error. However, due to a failed decomposition of the full observation variance-covariance matrix, the ESTKF without localization failed in producing output. In contrast, the EnKF without localization does not abort because the observation variance-covariance matrix is perturbed with the empirical forecast error matrix (term $(\mathbf{H}\tilde{\mathbf{P}}^{\mathbf{f}}\mathbf{H}^T + \mathbf{R})^{-1}$ in the Kalman Gain matrix, Eq. 5.12) and is not explicitly computed as in the ESTKF (Eq. 5.27). In conclusion, neither EnKF nor ESTKF without localization produce reliable results when accounting for the full variance-covariance matrix of observations.

To stabilize the framework and include as much observation error information as possible, localization techniques (Sec. 5.3) are now tested. For reducing correlations in observations, only domain localization is suitable because it directly modifies the observation variance-covariance matrix. Opposingly, the covariance localization does not modify the observations directly but instead, the error forecast matrix is modified. As shown in Eq. 5.31, the localized forecast matrix is then added up with the observations error matrix, thus observation correlations are still part of the update and might destabilize the system. Some tests could confirm this hypothesis: Applying covariance localization with the EnKF is unsuccessful. The assimilation runs until August 2007 and aborts in September because the water storage values aggregate over time and reach extremely high values that cannot be handled anymore. For computing domain localization with local ESTKF, it is required to define a subdomain, for which subsequently the analysis step is performed. A size of 0.5° for the subdomain is chosen because larger domain sizes, e.g. 3°, lead to artifacts in the assimilation outputs that resemble the outlines of the subdomains as boxes. Nonetheless, it should be mentioned that with the finer grid size of the subdomain, the run time increases.

Fig. 6.10 shows the global mean standard deviations of TWSA when assimilating GRACE/-FO with full variance-covariance matrices via domain localization with the local ESTKF for different filter radii and weighting options. There are nearly no differences between TWSA standard deviations when considering the constant weighting of observations or exponentially decaying weighting of observations with increasing distance towards the subdomain center. Nonetheless,

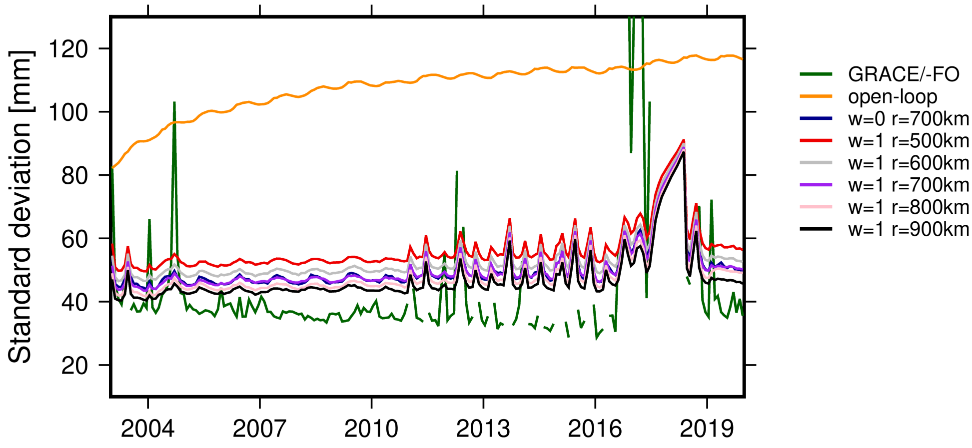


Figure 6.10: Global mean standard deviation of TWSA [mm] from GRACE/-FO, open-loop simulation and assimilation with full variance-covariance matrices using the local ESTKF for varying filter radii (r) and weighting functions (w), where zero represents a constant and one an exponential weighting function.

to avoid spatial artifacts that might occur when using a hard cut-off radius with constant weight instead of a soft cut-off with exponential weight, the analysis of the optimal radius for the localization is based on exponential weighting. Basically, all different localization radii lead to stable results. The higher the radius, the smaller is the global mean standard deviation of TWSA. With a larger radius, more correlations are included and the filter considers observations more realistically, but at some point, one must be careful and avoid destabilization. A radius of 700 km seems to be a good choice for this setup since it includes more information as, for example, a 500 km radius but does not get close to overfitting. This finding is in agreement with Springer (2019), who also chose a localization radius of 700 km for a regional GRACE/-FO assimilation assessment with local ESTKF in Europe. Localization radii much larger than the presented options, e.g., 2500 km, lead to destabilization of the framework and produce overly large outliers for storage outputs, as was found for the full variance-covariance matrix.

One could consider a third possibility for localization of observations, which is generally not included in the PDAF framework: localizing the observation variance-covariance matrix outside of PDAF manually and then applying EnKF with this localized variance-covariance matrix. In principle, this is exactly what was applied for release 2 of GLWS, where the diagonal variance-covariance matrix was used (which is the minimum localization radius), but for the new release observation correlations within a certain distance shall be included. In contrast to the local ESTKF, this method does not use subdomains and in contrast to local EnKF, it applies localization to the observation variance-covariance matrix instead of the state vector. However, although this allows the assimilation framework to run for all months, the water storage updates are found as not reasonable and contain extremely large aggregations of water over time. It seems as if correlations between model grids are also leading to destabilization as soon as observation correlations are included. Here I would like to give a reminder that correlations between model grids did not pose a problem for GLWS release 2, thus I hypothesize that the destabilization yields from the combination of both, correlations in the observation and forecast state matrices.

Finally, the domain localization via local ESTKF is chosen, because it provides the best results in terms of observation error localization and does not require an extra processing step of modifying the variance-covariance matrices outside of PDAF.

**Representing uncertainty of model states**

The model uncertainty is set up via perturbation of the forcing data and the model parameters. A concrete plan for disturbing the forcing was set, and for the parameters, two possibilities were introduced: Either the parameter range is set up between the so-called large or small parameter ranges (Sec 5.2.4). Fig. 6.11 (left) shows the global mean standard deviation of TWSA estimated from the ensembles from open-loop simulation and assimilation under the consideration of small or large parameter ranges. An expected behavior is found for the large and small range open-loop simulation standard deviation, this means the large parameter range causes a larger global mean standard deviation than the small range. In addition, the large range of the open-loop simulation varies much more with values ranging from about 150 mm up to 225 mm, while the small range varies from about 80 mm to 120 mm. Both parameter ranges lead to an increasing TWSA standard deviation with time, but this effect is much stronger for the large range.

The global mean standard deviations for TWSA from the assimilation runs are very equal and close to the GRACE/-FO standard deviation. Since the GRACE/-FO standard deviation of TWSA is much smaller than the standard deviations of the open-loop simulation, the small size of the standard deviation for the assimilation is not surprising because the assimilation result gets pulled towards the data set, which is more precise than the open-loop simulation. Further obvious are specific peaking months on the standard deviations from assimilation, for example in 2018, where the standard deviation of the assimilation is much larger than usual. In these months, no GRACE/-FO observations exist and the spread increases as would be the case for an open-loop simulation. To minimize the range difference of the ensemble between GRACE/-FO and open-loop simulation or prediction, the small range is considered for further computation. Thus, the chance that the filter has more balanced error ranges is higher.

A second possibility for adapting the uncertainty of model states was presented in Sec. 5.3 with inflation factors. Applying an inflation factor can help avoid a strong decrease of the ensemble width throughout the assimilation, stabilize the run, and optimize the performance by inflating the ensemble of the prediction step to spread up again. The effect of applying inflation on the assimilation in PDAF is controlled via the forgetting factor (inverse squared inflation factor, Sec. 5.3) for which now various options are tested. It should be reminded that the test for inflation now is combined with the localization via the ESTKF (exponential weight, radius of 700 km) as for a possible new release it is not useful to test inflation completely separated from a new filter and localization.

Fig. 6.11 (right) shows the global mean standard deviation of TWSA from assimilation runs for a forgetting factor of 1, 0.95, 0.9, 0.85, and 0.8 and compares them. The smaller the forgetting factor (larger inflation factor), the stronger the standard deviation increases towards the end of the time series. This strong decrease is a hint, that the assimilation might get unstable over time when the model receives less weight than observations compared to an assimilation run where inflation does not play a role. The assimilation seems to be numerically more stable as soon as the observations exhibit a smaller ensemble spread than the model predictions. This effect was also observed in the past, when working on the stabilization of the assimilation framework (Sec. 5.3), thus, I hypothesize here that the found destabilization of the assimilation framework is very case-specific and might show these findings because of the detailed model structures for WaterGAP. The destabilization of the model can be avoided by choosing a forgetting factor of 0.95 or higher. Thus, for the final global assimilation framework, the forgetting factor 0.95 is chosen, which is equal to an inflation factor of about 1.11.

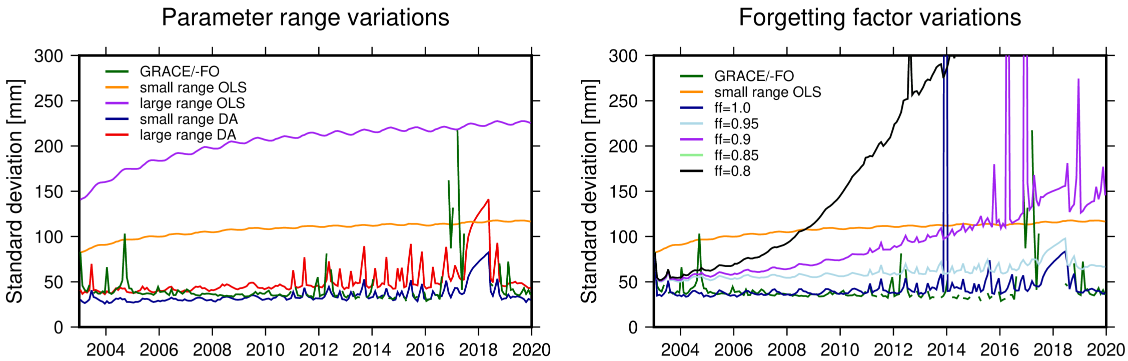


Figure 6.11: Global mean standard deviation of TWSA [mm] from the open-loop simulation (OLS) and the assimilation (DA), either for using a small or a large parameter range (left) or for varying forgetting factors (right) as well as the GRACE/-FO standard deviation ($3\sigma$).

### 6.2.3 Final release choices for GLWS3.0 and comparison to GLWS2.0

This section now summarizes the findings of refining the data assimilation setup. For the GRACE/-FO processing, it was found that restoring the signal affected by leakage is an important step, the lake and reservoir relocation and the rescaling both positively affect the data assimilation. To incorporate leakage at coastlines, rescaling is preferred over the lake and reservoir relocation. Incorporating a correction for earthquakes did not significantly impact the assimilation. No negative aspects were observed when removing earthquakes and with a view to future global assimilation that includes finer GRACE/-FO spatial grids, the correction is included in the processing. A change in the filter algorithm from EnKF to ESTKF indicated only minor differences but is necessary for applying localization and including more than only the diagonal entries of the observation variance-covariance matrix. The localization radius of 700 km with exponential weights showed to be a good choice for a stable global assimilation framework, as well as the forgetting factors of 0.95 or higher, in fact, a forgetting factor of 0.95 was chosen, which corresponds to an inflation factor of approximately 1.11.

To provide an indication of which of the discussed new processing steps has the largest effect on the assimilation output of TWSA, Fig. 6.12 shows the percentage change in RMS for the previously mentioned GRACE/-FO and assimilation tuning options for the assimilation. The RMS change is either calculated using a global spatial average (blue) of RMS for land TWSA from assimilation, or the average RMS computed from 140 basin averages (orange). The change in RMS is computed with respect to the standard options within GLWS release 2 (Tab. 6.2, left). The lake and reservoir relocation and a change in inflation factor (together with localization) are responsible for the largest percentage change in RMS globally with about 8-10%. The lowest RMS change on the global scale can be found when exchanging the filter algorithm from EnKF to ESTKF (below 1%) or correcting for earthquakes (about 2%).

For the lake and reservoir relocation the change in RMS is only slightly lower when calculating the RMS for basin averaged TWSA. This is expected since the relocation is spatially limited to surface waters that belong to basins. It might surprise that the lake and reservoir correction leads to a stronger change in RMS than the rescaling. Here, I hypothesize that adding back water to the TWSA from relocation via altimetry has a larger impact than rescaling amplitudes via models since the models generally are found to underestimate trends and amplitudes compared to GRACE/-FO observations (Scanlon *et al.*, 2018). In addition, the rescaling factors are limited to an upper factor of 3. Larger rescaling factors are not allowed to avoid overestimation, however, in some cases, this could also limit restoring realistic values. In the future, one could implement tests with varying upper limits. For this work, the rescaling is still the choice of removing the

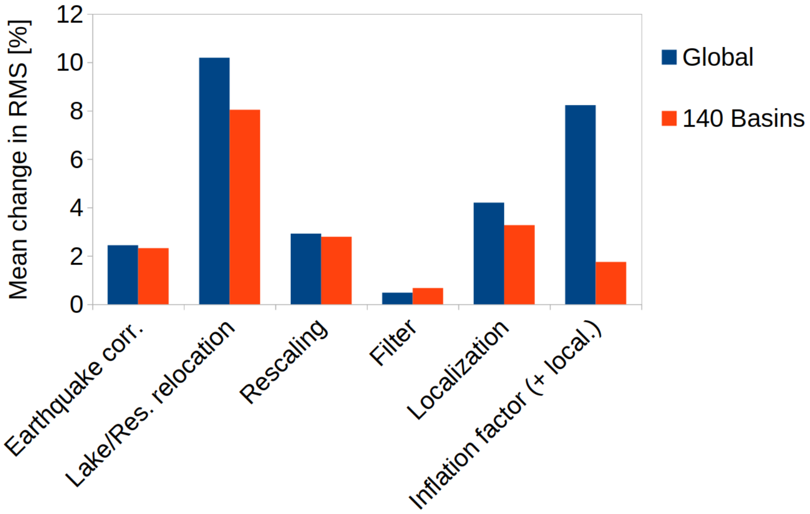


Figure 6.12: Percentage change in RMS [%] for using different settings: Removing earthquake signals, lake and reservoir relocation, rescaling, changing the filter algorithm from EnKF to ESTKF, localization with a radius of 700 km and an exponential weight, and setting the inflation factor (forgetting factor = 0.95). The change in RMS is either computed as the spatial average for the global land gridded RMS of TWSA or as the average of 140 basin averages of RMS of TWSA.

leakage signal because it is a global correction that includes coastlines. Furthermore, it is striking that the inflation factor (and localization) shows a much lower mean percentage change for the RMS averaged for basins as compared to the global RMS. From these results, it can be understood that the inflation has a strong impact on arid regions.

Due to the above-summarized reasons, and the findings from the GRACE/-FO settings (Sec. 2), the final new release of the assimilation framework – provided as Global Land Water Storage release 3 (RL003) – includes the settings as shown in Tab. 6.2.

## 6.3 Properties of the global land water storage (GLWS)

In the previous sections, the data assimilation was tuned with different options to find the most suitable data assimilation framework for the application of a global drought monitoring system. To analyze the performance of GLWS release 3, the signatures are analyzed on the spatial and spectral domain compared to model simulation and observation (Sec. 6.3.1), water cycle dynamics are explored in depth together with other climate variables (Sec. 6.3.2) and finally, a validation against surface water, groundwater and GNSS data is implemented (Sec. 6.4). For ease of use, GLWS release 3 is only denoted as GLWS in the following. For the comparison, the TWSA from GRACE/-FO observations have similar options as compared to what is assimilated for release 3 (Tab. 6.2) but is for spatial better comparison provided on the 0.5° grid. In addition, the classical standard model run of WaterGAP is used for the comparison, which means that the model is run with no perturbations for forcing data and parameters and does not require external initial states.

Table 6.2: Settings for the GLWS releases 2 and 3. The abbreviations r, w, and ff are the localization radius, localization weight, and forgetting factor, respectively.

| | **GLWS** | |
|---|---:|---:|
| **Option** | **RL002** | **RL003** |
| SHC degree/order | 96 | 96 |
| Lower degree repl. | degree 1, $c_{20}$ | degree1, $c_{20}$, $c_{30}$ |
| Ref. period | 2003 – 2016 | 2003 – 2019 |
| GRACE/-FO filter | DDK3 | DDK3 |
| GIA model | ICE5G-D | ICE6G-D |
| Earthquake correction | No | Yes |
| Lake/Reservoir relocation | No | No |
| Rescaling | No | Yes |
| Assimilation filter | EnKF | LESTKF |
| Localization | No | r=700 km, w=1 |
| Inflation | No | $ff$=0.95 |

### 6.3.1 Dominant signal patterns of hydrological signatures

**Signatures in the spatial domain**

This section analyzes the dominant signal patterns in the GLWS TWSA as well as surface water (sum of lakes, wetlands, reservoirs, and rivers), soil moisture, and groundwater as these are the main contributors to TWSA. Nonetheless, it should be mentioned that the snow and canopy storage are also contained in the TWSA from assimilation but only cover a small part of the TWSA. Figure 6.13 shows the TWSA linear trends (left column) as extracted from multi-linear regression (Sec. 6.1.2) and TWSA variability via RMS (right column) derived from WaterGAP model simulations (top), GRACE/-FO observations (middle top) and GLWS assimilation (middle bottom) of the new release 3. In addition, the previous release 2 of GLWS (bottom) is shown for comparison but the analysis will mainly concentrate on the three afore-mentioned data sets. The found patterns for the new release are very similar to the findings presented in Gerdener *et al.* (2023a) for release 2: The observations show much more intense trends than the model simulations. For example, in North East Brazil, the trends reach values of -30 mm per year or higher, whereas linear trends in WaterGAP are much lower. The underestimation of WaterGAP linear trends – but also other global hydrological models – compared to GRACE/-FO observation was also identified in Scanlon *et al.* (2018). In contrast, the simulated TWSA have a higher spatial resolution than the observed TWSA. The assimilation is now a synthesis of the model and observation and shows stronger trends than the model simulation but with a finer spatial resolution as for the observations. Linear trends in TWSA worldwide can be related to different hydrological processes or anthropogenic influences. All three TWSA data sets in this thesis agree with the sign of the GRACE/-FO trends as summarized in Rodell *et al.* (2018) for several regions, e.g., for negative trends in North East Brazil related to drought (Getirana, 2016), Middle East related to groundwater depletion (Joodaki *et al.*, 2014) or Alaska due to glaciers retreating (Luthcke *et al.*, 2013). In addition, agreements of positive trends are found for the Amazon basin due to a recovery from a dry period (Frappart *et al.*, 2012) or South East China related to the Three Gorges dam and other reservoir filling (Wang *et al.*, 2011).

However, there are also regions, where WaterGAP does not at all or only partly capture trends that can be observed with GRACE/-FO and GLWS. One example is groundwater extraction in India (Rodell *et al.*, 2009), which is very prominent in the observations over large regions in Northern India and Bangladesh. WaterGAP can capture the strong negative trends in the surrounding area of New-Dehli, which can also be observed in the model simulation but, in

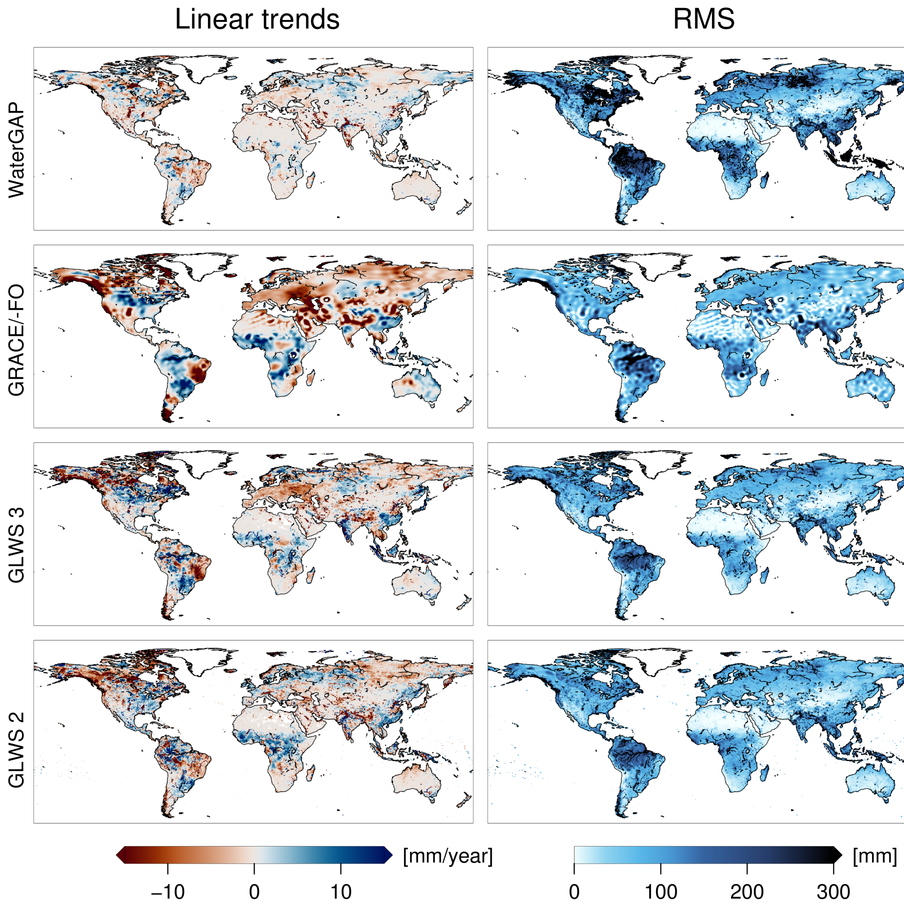


Figure 6.13: Linear trends [mm/year] (left) and RMS [mm] (right) of TWSA derived from WaterGAP model simulations (top), GRACE/-FO observations (middle top), GLWS release 3 assimilation (middle bottom), and GLWS release 2 assimilation. Based on Gerdener *et al.* (2023a).

addition, GLWS shows negative trends for the Ganges-Brahmaputra basin. GLWS thus adds more detail to the spatial extent of the negative trends and, as a hypothesis, more detail to the extent of groundwater depletion in this case. GLWS also seems to better capture wetting after dry periods compared to WaterGAP. Positive linear trends in Eastern Australia and the Okavango basin coincide with findings from Rodell *et al.* (2018), which are not visible in WaterGAP.

Comparing the variability of the three data sets, I find the highest RMS values of about 300 mm or higher for WaterGAP. GRACE/-FO has much lower variability, probably resulting from the filtering applied during the observation processing. Again, the assimilation inherits the finer spatial resolution of the model simulation, while increasing the RMS variability compared to the observations. The highest RMS values for GLWS are logically found in humid regions with large surface water, for example, the Amazon River basin, the Okavango River basin, the Great Lakes, or the Ganges-Brahmaputra delta. Thus, for the trends and variability of TWSA, the assimilation represents a smooth transition between model simulation and observation.

As the assimilation enables vertical disaggregation of the TWSA into the storage compartments, Fig. 6.14 shows linear trends (left column), annual amplitudes (central column), and annual phases (right column) of the four hydrological variables TWSA, soil moisture, surface water, and groundwater. As expected, most of the linear trends already discussed for TWSA from GLWS also appear in the groundwater trends. In other words, the groundwater trends are a much more dominant component of the TWSA trends than the surface water and soil moisture trends. This is also the case for GLWS release 2 (Fig. A.3).

As expected, linear surface water trends are most intense for large parts of surface water locations. For example, strong negative trends are found for Alaska and the Ob River, and positive trends for the Tibetian plateau, African Great Lakes, the Niger basin, and the Yenisei basin. Mixed signatures of strong positive and negative trends are found for the Amazon basin and large parts of Northern America. Kraemer *et al.* (2020) studied trends for lakes via altimetry and identified positive trends for lakes for the Tibetian plateau, the African Great Lakes, and North America lakes, which is in line with the surface water trends from GLWS.

The soil moisture trends from assimilation are discussed next and relate to the root depth in WaterGAP (globally varying). Positive soil moisture trends are found for South East China and in northern latitudes e.g., Northern Europe and East Canada, whereas negative trends are found for Central Indonesia. The Amazon basin shows, similar to the surface water trends, mixed linear trends for soil moisture. Other studies on soil moisture trends found overlapping trend directions, for example, a decrease for Central Indonesia and an increase for Northern Europe for soil moisture content was found from CMIP6 models or the GLDAS model (Gu *et al.*, 2019; Qin *et al.*, 2023) (for a soil depth of 0-200 cm depth).

For East Canada, studies using hydrological models such as GLDAS models usually find a decline in soil moisture opposite to the findings in this thesis (e.g., Deng *et al.*, 2020; Qin *et al.*, 2023). In contrast, the ESA-Climate Change Initiative (CCI) soil moisture (approximately 0-5 cm) used in Guevara *et al.* (2021) and also shown in Fig. A.13 indicates a positive trend, hence the models might not perfectly present the reality. Trends across the equator are not available for the ESA-CCI product used in that study, which impedes further comparison. Other possible explanations for the trend mismatch with model simulations in East Canada are that the ESA-CCI soil moisture is typically poorer in densely vegetated areas or that GRACE/-FO TWSA, as explained in Sec. 2.3.2, requires a correction for the GIA. The correction fields are derived from models and thus do not perfectly present the reality. I hypothesize here that an imperfect GIA simulation might be responsible for the strong positive trends and the found pattern has no direct physical background for soil moisture. I here refer to a more detailed comparison of soil moisture from ESA-CCI to soil moisture from GLWS in Sec. 6.4.4.

At this point it should be concluded that a standard comparison across studies for soil moisture linear trends is difficult since (1) soil moisture from models and observations can relate to very different depths of the soil, (2) process presentation of soil moisture is strongly different across models, (3) periods for trend detection vary, and (4) some observational products do not provide spatially consistent data because data gaps are contained.

When comparing the annual amplitudes of the four hydrological variables from assimilation, groundwater does not dominate the TWSA much more than soil moisture as is the case for the linear trends. Soil moisture and groundwater indicate strong annual amplitudes in humid regions especially around the equator. In some regions, the soil moisture annual amplitudes are even higher than the groundwater annual amplitudes, for example, for the Amazon basin, for South East Asia, Indonesia, and Europe. A possible explanation is that these humid climate regions are covered by forest and thus direct evaporation from soil is limited. At the same time, precipitation in these climate regimes can get very high. In contrast, for most parts of West and Central Africa, the groundwater amplitudes are higher than the soil moisture amplitudes. The regions

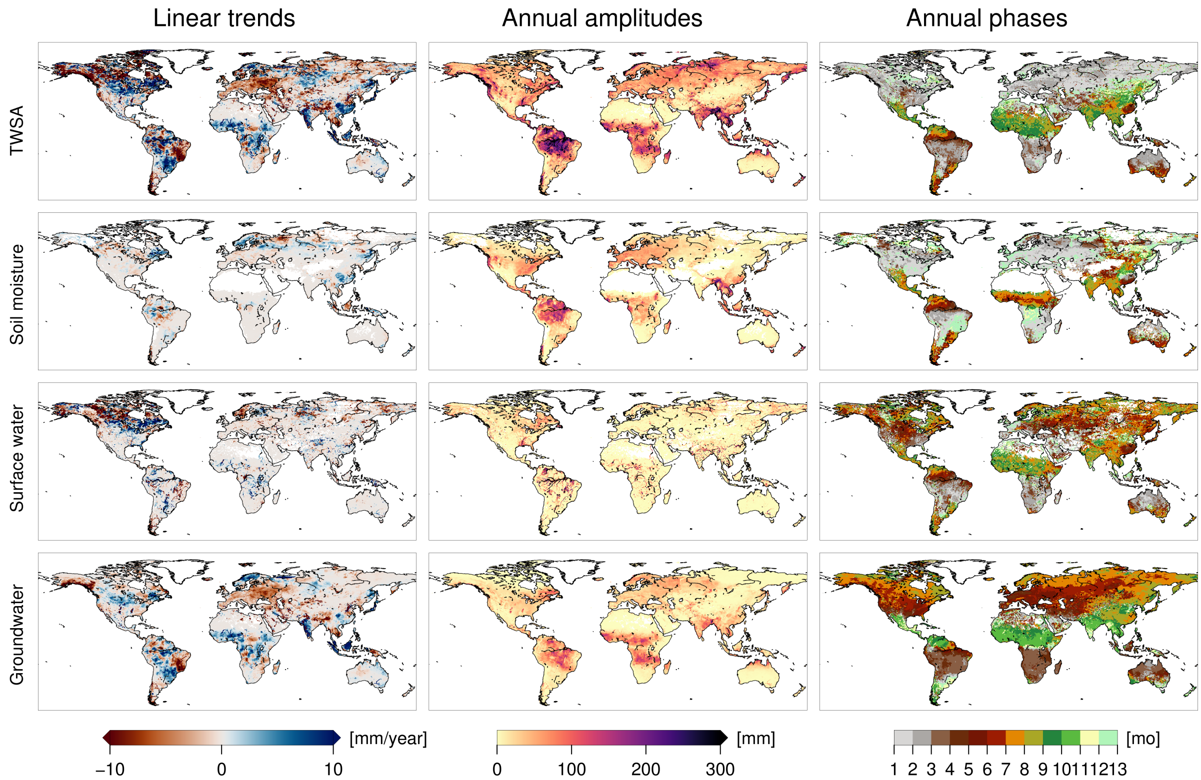


Figure 6.14: Linear trends [mm/year] (left), annual amplitudes [mm] (center) and annual phases [months] (right) of TWSA (top), soil moisture (middle top), surface waters (middle bottom), and groundwater (bottom) derived from GLWS.

are covered by grassland, savanna, and shrubland, and evaporation from soil occurs (Fig. 4.4).

The corresponding annual phases of the four water variables TWSA, soil moisture, surface water, and groundwater indicate the timing of the annual peak maximum during a year for the twelve calendar months. For example, a value of six means that the annual maximum is peaking in June. A general aspect that can be extracted is that there is a shift between the water storages over a few months. For example, in Southern Africa, soil moisture amplitude seems to be highest in December/January, whereas the surface water peaks in February or March, and after that the groundwater peaks in March/April. To get a better picture of seasonal and non-seasonal residual signal peaks in water storages and the timing between the peaks of different hydrological variables of the water cycle, a detailed analysis is made in a subsequent section (Sec. 6.3.2).

The new release of GLWS was built from several new options (Sec. 6.2), which led to different changes in TWS compared to release 2. After analyzing the signatures of the new release, I add an indication of how much trends, annual amplitudes, and semi-annual amplitudes have changed (in percent) from release 2 to release 3 and how the change propagates from TWSA to the storage compartments surface water, soil moisture, and groundwater all shown in Tab. 6.3. The numbers are determined by computing the global mean of the corresponding signature. For TWSA and groundwater, the global mean trend, annual amplitudes, and semi-annual amplitudes do not differ much between GLWS releases 2 and 3. When comparing surface water from the two releases, I do not find large differences in the trends and semi-annual amplitudes as well but the annual amplitudes amplitudes differ by about 20 mm, respectively. The same can be observed for the annual amplitudes of the soil moisture, which are about 20 mm lower for release 3 than for release 2. The aim of GLWS is to find a more stable version than release 2, and all strong

Table 6.3: Global mean linear trends [mm/year], annual amplitudes [mm] and semi-annual amplitudes [mm] for water storages of GLWS releases 2 and 3. The trends and amplitudes are estimated from 2003 to 2019 using multi-linear regression.

| **Signal component** | **TWSA** | | **Soil** | | **Surface** | | **Ground** | |
|---|---|---|---|---|---|---|---|---|
| | RL2 | RL3 | RL2 | RL3 | RL2 | RL3 | RL2 | RL3 |
| Linear trends [mm/year] | −1.50 | −1.32 | −0.73 | −0.16 | 0.14 | 0.02 | −0.42 | −1.11 |
| Annual amplitudes [mm] | 64.16 | 66.31 | 42.48 | 23.32 | 49.12 | 29.77 | 18.84 | 23.11 |
| Semi-annual ampl. [mm] | 20.09 | 17.44 | 21.07 | 8.49 | 15.81 | 10.33 | 5.45 | 5.36 |

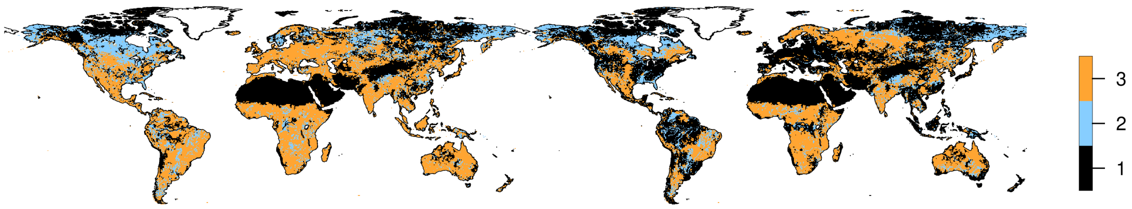


Figure 6.15: Most dominant storage compartment of TWSA for the linear trend (left) or annual amplitude (right): Choices are soil moisture (1), surface water (2), and groundwater (3) derived from GLWS assimilation.

amplitude differences I mentioned before indicate a lower amplitude for the new release. This coincides with the typical experience that outliers typically occur in the surface water and it can be hypothesized that some large outliers of release 2 in surface waters and soil moisture were reduced now in the new version.

To quantify the most dominant signature for the storages per grid, the results from multi-linear regression are further categorized into the most dominant storage for each grid cell and afterwards compared to results from signal decomposition techniques as PCA and ICA (Sec. 6.1.2). Fig. 6.15 shows which storage has the largest absolute trend (left) and amplitude (right) derived from multi-linear regression. As already concluded, the groundwater linear trends are much more intense than surface water or soil moisture trends for 50.00% of the global land surface. This is not the case for higher latitudes. For example, in Canada and Alaska surface water trends and soil moisture trends are more intense than groundwater trends. In total, soil moisture trends dominate for 31.60% and surface water trends for 18.40% of the global land area. In contrast, for the annual amplitudes, the soil moisture is most dominant. Soil moisture and groundwater annual amplitudes dominate large parts of the global land surface at 50.04% and 35.16%, respectively. Especially in dry climates groundwater is the prevalent storage for annual amplitudes in water storage, whereas in humid regions soil moisture often dominates, for example for tropical regions.

Performing signal decomposition via PCA for TWSA from assimilation, the first mode is the mode with the most dominant pattern with 33.98%, the second mode presents 13.23% and the third mode 7.78%. In total, the three most dominant modes of TWSA from GLWS represent 54.98% of the full TWSA. Fig. 6.16 shows the spatial EOFs and temporal PCs (Eq. 6.16) of the first three dominant modes. The first mode confirms the findings for TWSA that were found using multi-linear regression, as PC1 clearly shows linear trends contained in the data. In 83.48% of the global land area, the PCA trends match in signs with the trends extracted from multi-linear regression. The second mode (EOF2 and PC2) shows a clear annual signal. The spatial extent of the annual amplitudes extracted from PCA also overlaps with the spatial extent identified when computing annual amplitudes from multi-linear regression. The third mode (EOF3, PC3) also partly indicates a signature that was not discussed before. In general,

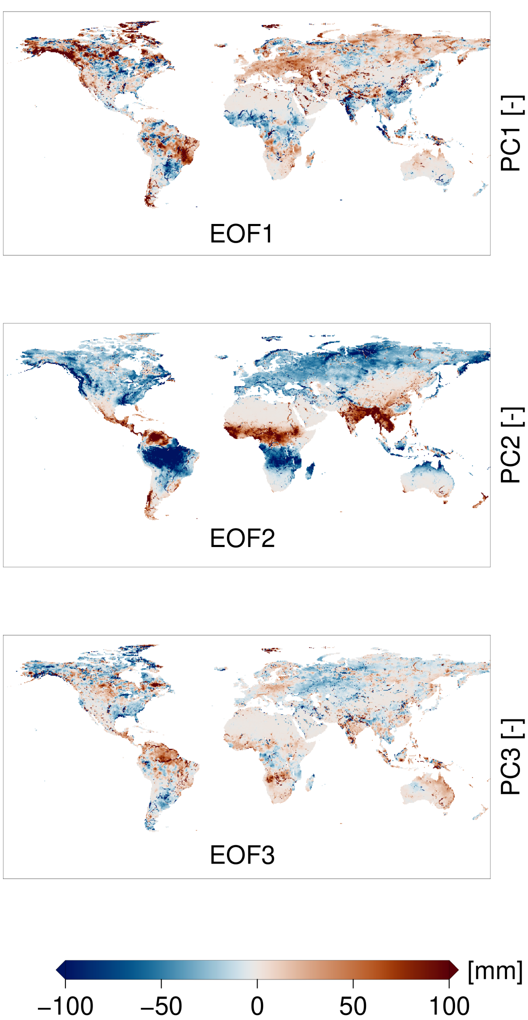


Figure 6.16: First three dominant PCA modes (EOFs [mm] and PCs [-]) of GLWS derived TWSA.

the PC3 starts with positive values, increasing until approximately 2013 and decreasing again after 2013. Forootan (2014) found a similar dominant pattern in global observed TWSA and related it to an acceleration dominant for Greenland, Antarctica, Thailand, and Lake Victoria. Greenland and Antarctica are excluded in GLWS, however, the spatial extent of this mode is much larger than for Forootan (2014). Possibly, accelerations worldwide are more dominant, because the length of the time series is five years longer in this work but it could also be that the seasonal signature in this component partly hides the spatial extent of accelerations. This seasonal signature in the third mode has a cycle of approximately 24 months, i.e., 2 years, as was found from Fast Fourier transform (e.g., Nussbaumer & Nussbaumer, 1982) that was used to compute the power spectral density of the PCs to identify periods (Fig. A.4). Probably, the two-year cycle in the PC3 results from the overlapping of seasonal variation with a sub-cycle of the two years, e.g. a yearly or a half-yearly cycle. When referring to Fast Fourier transforms, this phenomenon is known as harmonics that exist in the power spectrum, which are integer multiples of the fundamental power frequency (e.g., Ingale, 2014).

The variance and assignment to dominant signatures of the first three dominant modes of surface water, soil moisture and groundwater (EOFs and PCs attached in Figures A.5, A.6,

Table 6.4: Percentage variance and assignments of the first three PCA and ICA modes of soil moisture, surface water, and groundwater from GLWS. The assignments to a certain signature are abbreviated with "T" for linear trends, "A" for constant accelerations, and "S" for annual seasonality. It should be noted that the the ICA requires a predefined number of modes as input (here 3), thus the variance percentage adds up to 100.

| **Method** | | **PCA** | | | **ICA** | | |
|---|---|---|---|---|---|---|---|
| **Storage** | | **Soil** | **Surface** | **Ground** | **Soil** | **Surface** | **Ground** |
| **Variance** [%] | **Mode 1** | 36.31 | 26.46 | 60.28 | 46.17 | 44.01 | 60.24 |
| | **Mode 2** | 11.12 | 11.61 | 10.67 | 43.46 | 24.39 | 30.92 |
| | **Mode 3** | 5.46 | 9.66 | 5.73 | 10.37 | 19.06 | 8.84 |
| **Sum** | | 52.89 | 47.73 | 76.68 | 100 | 100 | 100 |
| **Assignment** | **Mode 1** | S | T | T | S | T/A | T/A |
| | **Mode 2** | S | S | A | S | S | A/S |
| | **Mode 3** | T | - | S | T | - | S |

and A.7) are jointly summarized with ICA dominant modes in Tab. 6.4. It should be noted that the ICA algorithm requires a predefined number of modes as input (here three modes), thus the variance percentage adds up to 100. For the PCA the number of modes is set to the number of time steps and thus the first three modes do not add up to 100% as for the ICA. Thus, PCA and ICA variance here are not directly comparable, nonetheless, the numbers provide information on which signature is most dominant in the storages depending on the method.

For groundwater, the first three modes represent similar signatures as is the case for TWSA with a total variance of 76.68%. The three most dominant modes of surface water only have a total variance of 47.73% with the first mode that represents trends covering most of it with 26.46%. Soil moisture does not have a strong trend component (mode 3 with only 5.46%), as was also discovered with multi-linear regression. Instead, the most dominant signature in soil moisture is an annual amplitude that is present in the first mode with 36.31% and in the second mode with 11.12%. The third mode of surface water was not clearly assigned as it presents a two-year cycle. Again, this might result from the overlapping of annual or semi-annual variations.

From other literature, it was found that ICA better separates hydrological signatures in the data than PCA as the components are aimed to be statistically independent (Sec. 6.1.2), which is why it was introduced in this work. However, for this global application on data assimilation outputs, a clear separation of the signatures was not possible, even when the number of input modes for the algorithm changed. For surface water and groundwater, two of the three most dominant modes are a mixture of two signals, for example, the second independent mode of groundwater was assigned to show annual amplitudes and acceleration. However, there is a tendency of the first three ICA modes to show similar modes as PCA but not completely separated.

#### Signatures in the spectral domain

As GRACE/-FO is a gravity mission, the signatures of GLWS are also analyzed with respect to the spectral domain to identify relevant features of the data set for the geodetic community. First of all, the gridded TWSA is transformed to Spherical Harmonic Coefficientss (SHCs) in the spectral domain via spherical harmonic analysis. It is very important to remember that TWSA are given on a global spatial grid on the land area without Greenland and Antarctica and therefore the final GRACE/-FO SHCs cannot be expected as similar to coefficients from the official processing centers.

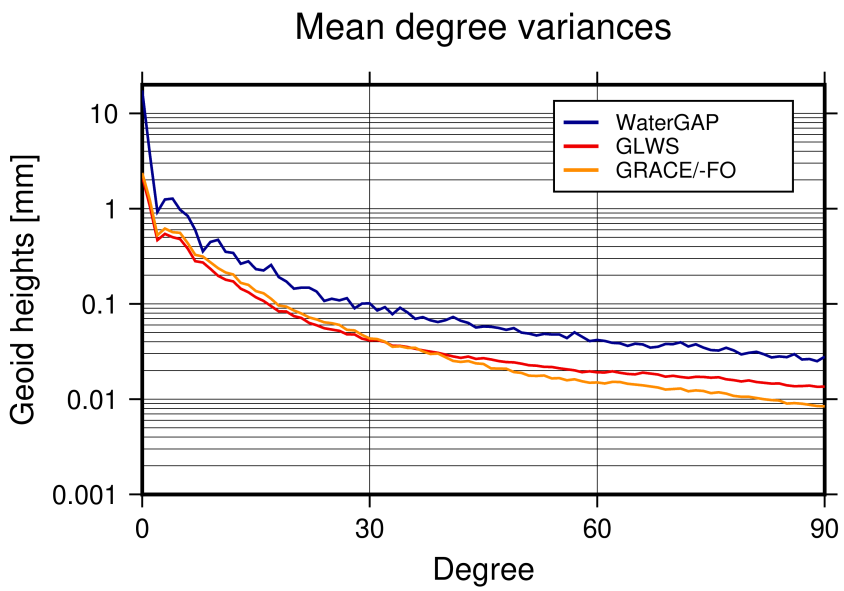


Figure 6.17: Temporal mean signal degree variances [mm geoid heights] from 2003 to 2019 for GLWS assimilation (red), WaterGAP model simulation (blue), and GRACE/-FO observations (orange). Based on Gerdener *et al.* (2023a).

Fig. 6.17 shows the signal degree variances (Eq. 2.17) up to degree 90 temporally averaged and given in mm geoid heights (Eq. 2.24). Degree variance is a measure of how much power each spherical harmonic degree has within the data. For lower degrees, the mean degree variances are very close to the GRACE/-FO degree variance. Thus, especially the long-term signal of GRACE/-FO transplants to the GLWS. The WaterGAP model has a higher degree variance for the lower degrees as compared to GLWS and GRACE/-FO. For degrees higher than approximately 30, the GLWS starts to derive equal or higher degree variances than GRACE/-FO. With increasing degrees, the GLWS degree variances get closer to the WaterGAP degree variances but never become equal. These degrees are associated with medium to short-term signatures or noise in the data, thus, especially for medium to short-term signals the assimilation inherits more from model simulations. In fact, GLWS represents a smooth transition between observations and model simulations for the degree variances.

The correlation of SHCs from assimilation with model simulations (left) or with observations (center) is shown for each of the coefficients in Fig. 6.18. In addition, the differences (right) between these two correlation maps are shown to emphasize where the correlations resemble each other and where they differ. The lowest degrees are visualized at the bottom of the figure and increase up to degree 90 at the top and the lowest orders are found in the center of the figure and increase to order 90 to the left and right outer parts. The left-hand side of the triangle in the figure visualizes the s-coefficients, whereas the center and the right-hand side show c-coefficients per degree and order. The correlations for the SHC from assimilation with the observations indicate a high agreement for degrees and orders of 30 or lower. For degrees higher than 30, the correlations are more noisy, which means positive and negative but also zero correlations are observed. In contrast, the correlations between SHCs of the assimilation with the model simulation are generally positive for more coefficients than for GRACE/-FO. A look at the difference between the correlations reveals more detail. Positive differences indicate that the correlations between assimilation and observation are higher than the correlations of assimilation and model simulation, blue indicates the reverse. Although correlations between GLWS and

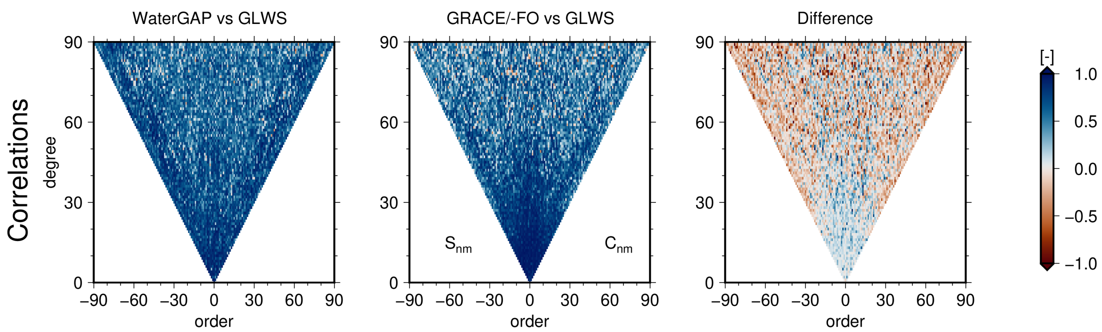


Figure 6.18: Correlations [-] between the SHCs from GLWS assimilation with WaterGAP simulations (left), with GRACE/-FO observations (center) and the difference between both (right).

WaterGAP seem to be high for degrees lower than 30, the correlations of GRACE/-FO are even higher. Another interesting aspect for degrees higher than 30 is that the differences between the correlations are stronger negative for the upper orders as compared to lower orders. It means that especially for larger orders, the assimilation is closer to the model simulations as compared to the observations.

Fig. 6.19 shows the correlation between the SHCs from GLWS with WaterGAP (blue) or GRACE/-FO (orange) either averaged per degree or per order. Up to degree 35, the observed SHCs per degree has a higher correlation to the assimilation as compared to model simulations. An exception is found for the very low degrees zero and one. As presented in Sec. 2.3.1 especially these degrees are imprecisely measured by GRACE/-FO and were replaced by other solutions. Since correlation is a metric that is sensitive to differences in amplitude between two time series but not sensitive to biases, the findings could be a result of dominant amplitudes in the lower degrees. This does not necessarily mean that GLWS is, in general, closer to WaterGAP for the degrees below 2. Considering the SHCs that are averaged per order, the assimilation-observation correlations are higher up to approximately order 20. After order 20, the model simulations have a higher correlation to the assimilation as compared to the observations.

To provide more details about specific coefficients and to highlight the advantages of assimilating GRACE/-FO TWSA into WaterGAP for the spherical-temporal domain, Fig. 6.20 shows the $c_{20}$ and $c_{30}$ coefficients per month. The coefficients from assimilation are temporally consistent, whereas the observations have well-known missing months and a gap between the two gravity missions. GLWS $c_{20}$ inherits much of the GRACE/-FO signature here as well. Compared to model simulations, the linear trend in the assimilation-derived $c_{20}$ coefficient is stronger negative. In fact, WaterGAP has a linear trend of $8.48 \cdot 10^{-13}$ and the GLWS trend is with $-4.19 \cdot 10^{-12}$ closer to the GRACE/-FO trend of $-8.97 \cdot 10^{-12}$. Similarly, the assimilation shows a smooth transition for the annual amplitude, being in between the observations and the model simulation. For $c_{30}$, the differences between model simulation, assimilation, and observation are not as intense as for the $c_{20}$, however, the amplitudes from assimilation and observation are slightly larger with $7.28 \cdot 10^{-11}$ and $7.73 \cdot 10^{-11}$ respectively as compared to the model simulations with $6.16 \cdot 10^{-11}$.

**Uncertainty quantification**

For all outputs of the data assimilation, not only the ensemble median is derived but the information of the spread of the 32 ensemble members is offered. This information can be used to provide an uncertainty quantification via computing a standard deviation per grid, time step, and

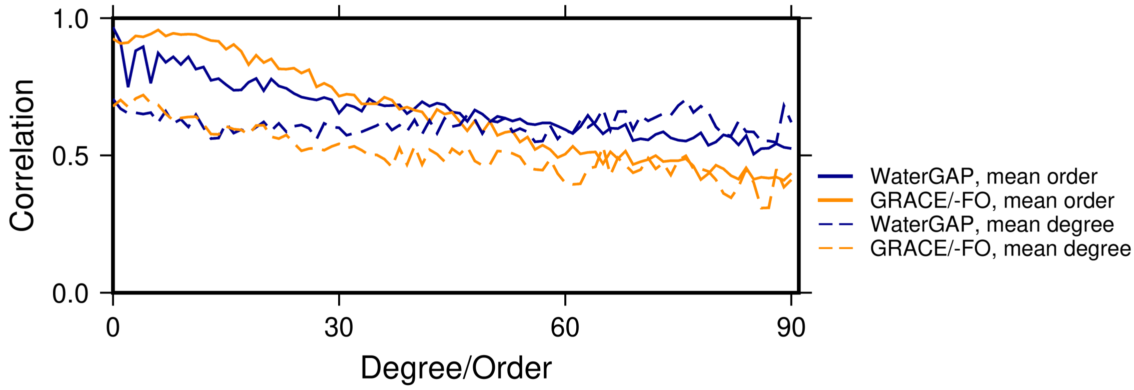


Figure 6.19: Correlations [-] between SHCs from GLWS with WaterGAP (blue) or with GRACE/-FO (orange), both averaged per degree (dashed line) or order (continuous line).

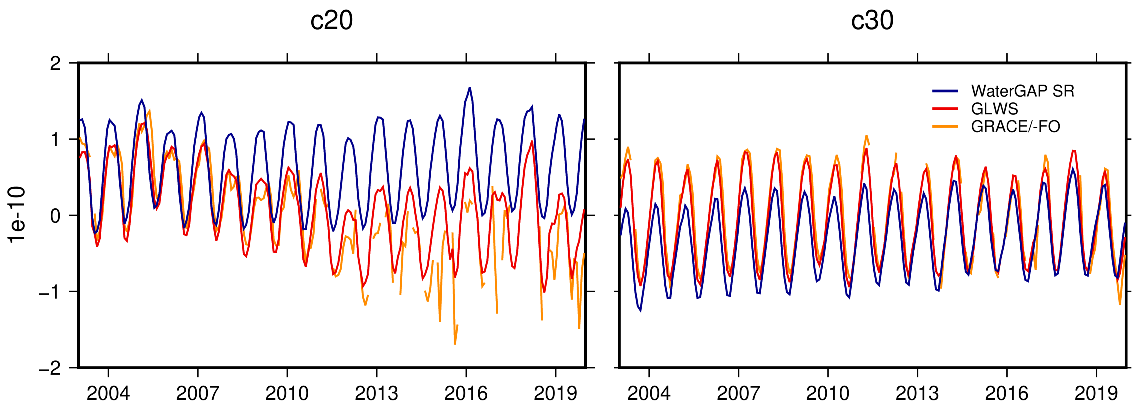


Figure 6.20: Spherical Harmonic Coefficients (SHC) $c_{20}$ (left) and $c_{30}$ (right) [-] from 2003 to 2019 for GLWS assimilation (red), WaterGAP simulations (blue), and GRACE/-FO observations (orange).

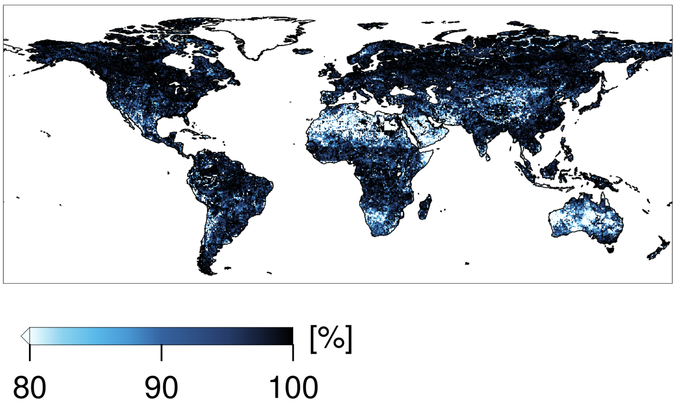


Figure 6.21: Percentage [%] of months of TWSA from GLWS assimilation that showed a significant standard deviation. The significance was tested using a z-test with a 5% significance level.

storage from the ensemble. To show whether the standard deviations are representative of the ensemble, for each grid cell a z-test is executed (e.g., Casella & Berger, 2002). The z-test assumes Gaussianity and tests if a given data vector is from a normal distribution with a certain mean and standard deviation at a confidence level of 95%. Fig. 6.21 shows the percentage of months of assimilation derived TWSA, where the standard deviation is significantly representative for the ensemble spread. The continents South America, North America, and Europe show a high percentage of significant months with 90% or higher. The percentages for Asia are also very high for most regions, but for the Himalayas, the percentage can fall below 80% only for some grids. For Australia and Africa, large areas can be identified that have an insignificant standard deviation for 20% of the months or more, for example, Northern Africa, a region in Southern Africa, and West/Central Australia. These areas are deserts and are the most dry and hot climates on Earth, thus, the results can be explained as the modeling in those areas is especially challenging and the observations experience a low signal-to-noise ratio.

The standard deviations estimated from the ensemble spread per month and grid can then be used and further propagated for specific applications, for example, for deriving standard deviations for linear trends as shown for TWSA, surface water, soil moisture, and groundwater in Fig. 6.22. An important point is that the standard deviation can get much larger than the colorbar indicates with 1 mm per year, but is limited to this value to better compare spatial structures across the four hydrological variables. In general, the standard deviations of all variables show large standard deviations in those regions that were identified for intense trends and amplitudes before. To give one example, the annual amplitudes of soil moisture are most dominant for the equator regions, the same regions are also striking for the standard deviations of soil moisture. As expected, the TWSA linear trends have the highest uncertainty compared to soil moisture, surface water, and groundwater as it is the vertical disaggregation of them. The grids at the coastline have a very large uncertainty (exemplarily shown with zoom for TWSA standard deviations of linear trends for West Africa in Fig. A.8). The reason is that during the assimilation these grid cells do not directly receive an update because the 4° GRACE/-FO grids do not cover all of the coastline cells and thus, the ensemble spread does not get smaller as is the case for cells that directly receive an update. Nonetheless, it should be mentioned that the grid cells still receive updates indirectly via the routing of water through the river.

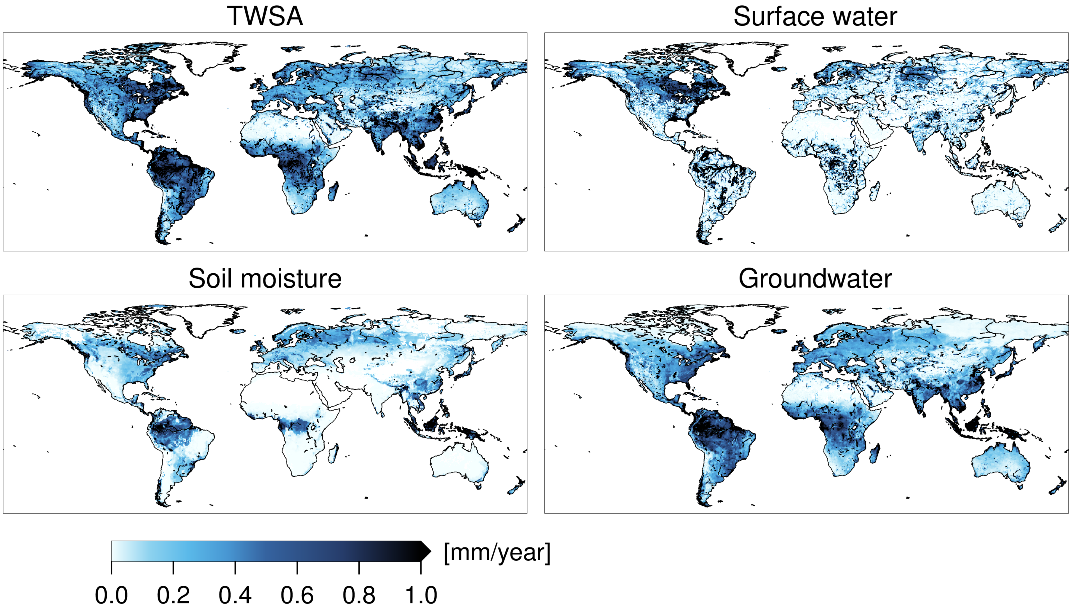


Figure 6.22: Standard deviations of the linear trends [mm per year] of TWSA (top left), surface waters (top right), soil moisture (bottom left), and groundwater (bottom right) derived from GLWS assimilation. Based on Gerdener *et al.* (2023a).

### 6.3.2 Water cycle dynamics

In this part, water cycle dynamics are investigated in detail because it is important to build a realistic picture of the response dynamics of fluxes, water storages, and vegetation. The timing of seasonal peaks and non-seasonal events in precipitation, water storages, and vegetation is analyzed and related to each other to better understand dynamic responses in the global water system, with hindsight to drought, especially for the non-seasonal part of the data. However, it must be mentioned here, that peak-to-peak time is an indicator for how long water needs on average to transfer from flux to storage or from storage to vegetation. As the propagation of single water particles is a complex process, the time can not be understood as the time that a single water particle requires.

For analyzing precipitation – water storage –vegetation dynamics, I expand the study of Gerdener *et al.* (2022) to the global scale and distinguish between seasonal and residual signal components. First, I analyze the duration between peaking times of precipitation and water storages and subsequently for the peaking times from water storages to vegetation with the help of the statistical tools as presented in 6.1. To prepare useful information for the suggested drought warning system that is implemented in Chap. 7, the focus of this section is on non-seasonal signatures in the data. Lags between the data set are determined as median per biome cluster (Sec. 6.1.3). The precipitation data set that is used for this study is the W5E5 precipitation data, the same as for the WaterGAP forcing data set (Sec. 4.6). To incorporate vegetation, I additionally shortly introduce remote-sensed vegetation proxies and the chosen data set. A moving average filter over time of nine months is applied to each data set for the computations in this section to reduce possible temporal noise effects and derive stable outputs.

#### Precipitation - water storage response

The dynamics for the propagation of precipitation to water storage behave differently for the seasonal signal and the non-seasonal signal. Thus, at first, annual phases of precipitation are analyzed with respect to the annual phases of water storage from assimilation or from model simulations in Fig. 6.23. Annual phases are computed per grid for precipitation and the storages separately via multi-linear regression (Sec. 6.1.2). Afterwards, relative phases between precipitation and each storage are estimated per grid following Eq. 6.10, and finally, the relative phases are spatially averaged via computing the median for each biome cluster in order to find spatial coherent properties. The left panel in Fig. 6.23 shows the relative annual phases between precipitation and the water storages as modeled, whereas the right panel shows the same for assimilation.

In general, only a few clusters show larger differences between WaterGAP and GLWS. For example, it takes more than six months in Northern Europe and North East Asia for the annual precipitation to refill the modeled annual surface water in WaterGAP, whereas it takes between two and five months for the assimilation. Considering all 73 biome clusters, for 42% of the clusters the duration between precipitation peaks to groundwater peaks increases for GLWS as compared to the model simulation, whereas for 30% of clusters, the duration decreases, and for 28% no change is observed. In conjunction, it takes longer from peak precipitation to peak annual soil moisture or surface water in the assimilation as compared to the model simulation for 23% or 26% of the clusters. It can also be found, that the assimilation has the strongest effect on relative phases for groundwater, because only 25% of the clusters show the same relative phases with model simulation, whereas for soil moisture and surface water, the relative phases for the clusters stay equal for more than 50%. I now divide the analysis into two parts: (1) results shown for latitudes below 35°, this means approximately all southern of Europe, and (2) results shown for latitudes higher than 35°, which includes Europe, Canada, Alaska, and Northern Asia.

For latitudes below 35°, the duration from annual water peaks of precipitation to water storage from assimilation is shortest for the soil moisture with an immediate response or 1-month delay. Annual surface water shows a delay of about one to two months and the longest delay is observed for groundwater with about three months. As expected, the precipitation recharges soil moisture nearly immediately. The delay towards surface water and groundwater is also expected since on average, water in surface bodies is routed, water seeps to the soil before replenishing groundwater and both thus takes time.

For latitudes above 35°, the findings are very different and do not confirm the above-mentioned sequence. Seasonal soil moisture indicates a response time towards seasonal precipitation of about five months, the results for groundwater show a shorter response time for most clusters and surface water shows basins with a very short response time but also with response times of 5 months or higher toward seasonal precipitation. As the annual cycle is twelve months, the approach has one main important point to make. Especially for soil moisture and surface water negative feedbacks might occur due to evaporation at the surface. This means that the maximum precipitation peaks do not always lead to a change in soil moisture a few months later but a change in soil moisture can also lead to a change in precipitation (e.g., Yang *et al.*, 2018) after intense evapotranspiration. However, this case is not included in the approach of relative phases because negative phases are excluded due to the periodicity of annual phases. The reason can be explained with Eq. 6.10. If we now assume that annual precipitation indeed is peaking in December and soil moisture is peaking 2 months later in February, the approach would derive a negative relative phase. Thus, relative phases that are estimated over the turn of the year could be interpreted incorrectly, for example, as a precipitation peak that is following a soil moisture peak.

Thus, the approach enables one to relate significant signatures in seasonal variations of data sets but according to the mentioned limitations and to the purpose of drought application within this thesis, the further analysis of precipitation peaks refreshing water storage focuses on the

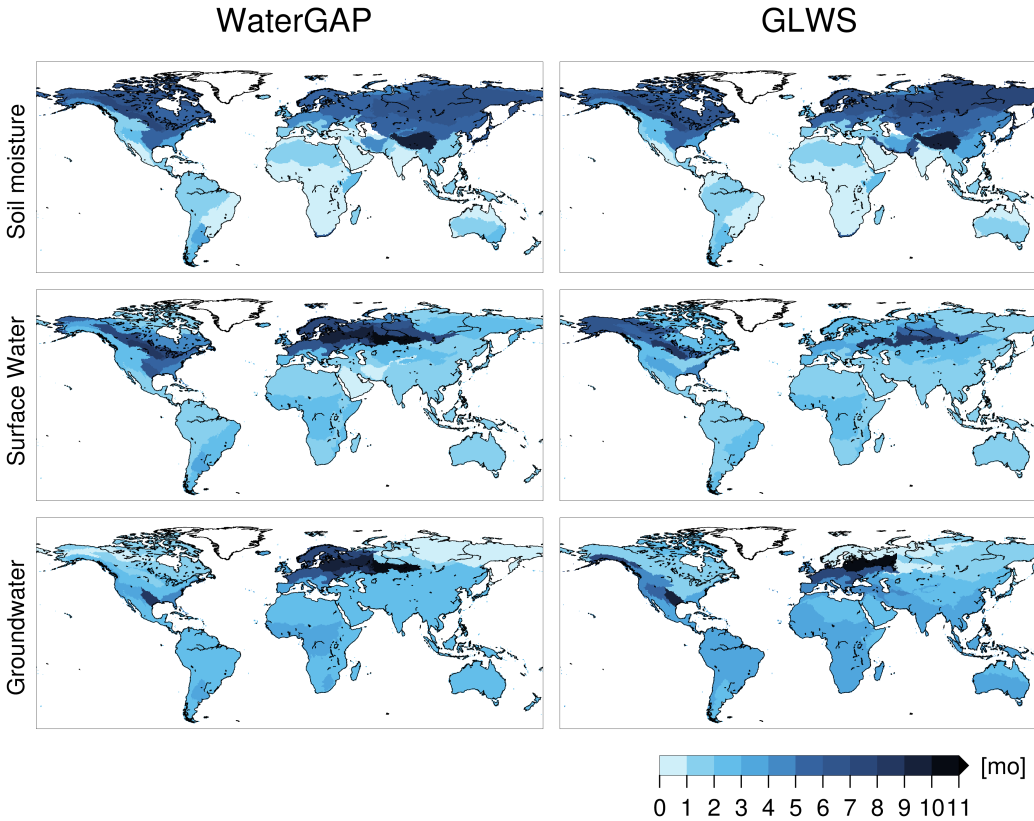


Figure 6.23: Annual phases [months] of precipitation relative to soil moisture (top), surface water (middle), or groundwater (bottom) either derived from WaterGAP model simulation (left) or GLWS assimilation (right). The annual relative phases are estimated using multi-linear regression from 2003 to 2019.

non-seasonal part of the data. When using the GLWS release 2 instead of release 3, the relative phases are very similar to the here analyzed results (Fig. A.10) leading to the same conclusions.

To derive a residual signal that contains non-seasonal variations, the seasonal, linear trend, and constant bias signals are separated from the data via multi-linear regression (Eq. 6.12). Fig. 6.24 shows the temporal lags between non-seasonal precipitation and water storages estimated via cross-correlation analysis after removing seasonality, trends, and biases. The peaking time of non-seasonal precipitation immediately leads to a peak in non-seasonal soil moisture for most of the global land area with a time delay of less than two months, especially for latitudes lower than 45°. Surface water responds at approximately the same time or one to two months later than the soil moisture. The largest time delay is found for non-seasonal precipitation peaks propagating to groundwater peaks. It takes three or more months worldwide for the precipitation-groundwater interaction. The lags can get even higher than six months, which is much longer than the duration it took from the seasonal signal. Thus, for non-seasonal events like droughts, it can be expected that a starting deficit in precipitation is observed three months later at the earliest in groundwater.

Relations between non-seasonal precipitation and water storages can also be specified for the current month. Computing an auto-regressive process as described in Eq. 6.1.2 enables providing a measure of how much the current-month storage is depending on the previous-month storage ($c$-coefficients, Fig. 6.25, left) or on same-month precipitation ($d$-coefficients, Fig. 6.25, right). The closer the $c$-coefficients are to one, the stronger the water storage of the current month depends

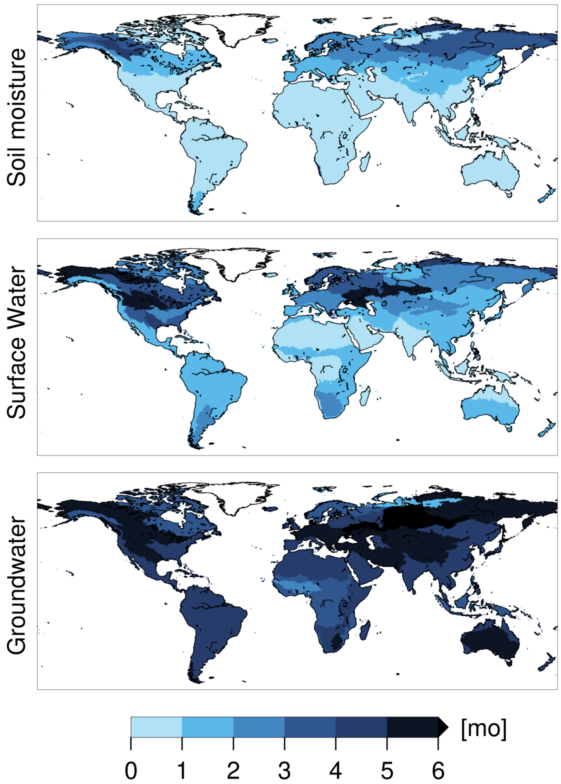


Figure 6.24: Lags [months] of precipitation relative to assimilation-derived soil moisture (top), surface water (middle), or groundwater (bottom) for the non-seasonal residual part of the data. The lags are estimated using cross-correlation analysis from 2003 to 2019.

on the previous-month storage. The more the $d$-coefficients depart from zero, the stronger the storage depends on same-month precipitation. Non-seasonal soil moisture is generally strongly dependent on same-month non-seasonal precipitation for many clusters of the world (values larger than 0.8). Dry regions, for example North Africa, show a decrease in the auto-regressive coefficients compared to wetter regions. For Southeast Asia, West Africa, Europe, North America, South America, and Central Africa, a strong precipitation-soil moisture interaction is found for the de-seasoned signal. At the same time, the dependency of soil moisture on previous-month soil moisture is high for northern latitudes, whereas it is generally lower for latitudes below 30 °. In contrast, surface water does not show a large dependency on previous-month precipitation worldwide. This means – as expected – that in case of non-seasonal events like droughts, a fast response towards precipitation deficits is most apparent for soil moisture.

For non-seasonal groundwater, a high dependency on previous-month de-seasoned groundwater is found. Compared to de-seasoned soil moisture, the response of groundwater to same-month precipitation is lower. This is again an expected result as it takes time for the precipitation to seep through the soil. The surface water response of the current to the previous month is found in between the results for soil moisture and groundwater and thus completes the picture that from surface to sub-surface, the dependency on previous months for the respective water storage is increasing. Nonetheless, surprisingly the response of surface water to same-month precipitation is much lower than for groundwater. This results from the fact that cells with rivers or lakes containing large water masses show a much stronger dependency on same-month precipitation as compared to cells that contain smaller water masses in lakes, rivers, or wetlands. The median then leads to small values because the cells with smaller water masses cover a much larger area than the cells with larger water masses. This regional difference is not found when analyzing the dependency of current-month surface water to the previous-month, for the relative phases shown before, and also not for the lags shown before. Thus, I hypothesize that in this specific case, the process model results for surface water together with median computation per cluster do not show reasonable results.

**Water storage - vegetation response**

Next, the timing between water storages peaks to peaks in vegetation growth is analyzed. For observing vegetation, a wide range of remotely sensed vegetation metrics exist. Here, I aim to use a global data set, which has a similar or higher spatial resolution than the assimilation (0.5°). The techniques to observe vegetation from space differ from mission to mission but most of the satellites use imaging spectrometry, for example, the Sentinel-3 OLCI launched by ESA and the MODIS missions launched by NASA. Common vegetation indicators provided by such types of missions are the Normalized Difference Vegetation Index (NDVI), Actual Evapotranspiration (AET), and Leaf Area Index (LAI). The Copernicus Global Land Service is a platform that provides, among others, vegetation data sets from different satellite missions and with different spatial and temporal resolutions as the NDVI and LAI. AET is not available on that platform.

Ukasha *et al.* (2022) provided a detailed comparison of NDVI and LAI and how the two metrics correlate to other hydro-climatological variables for two river basins in California dominated by grass and shrubland. They found that both indicators are suitable for correlation analyses as they are correlated with the hydro-climatic variables but differ in the peaking time of the seasonal cycle. In case only one index is used, Ukasha *et al.* (2022) suggest using the LAI because it might have a better indication of vegetation response to drought and climate variations. The suggestion for LAI might change for other vegetation regimes, but in addition, NDVI only captures the horizontal projection of vegetation cover, while LAI also captures the leaf growth below the most upper canopy layer (e.g., Gitelson *et al.*, 2003). Due to these reasons, the LAI is considered

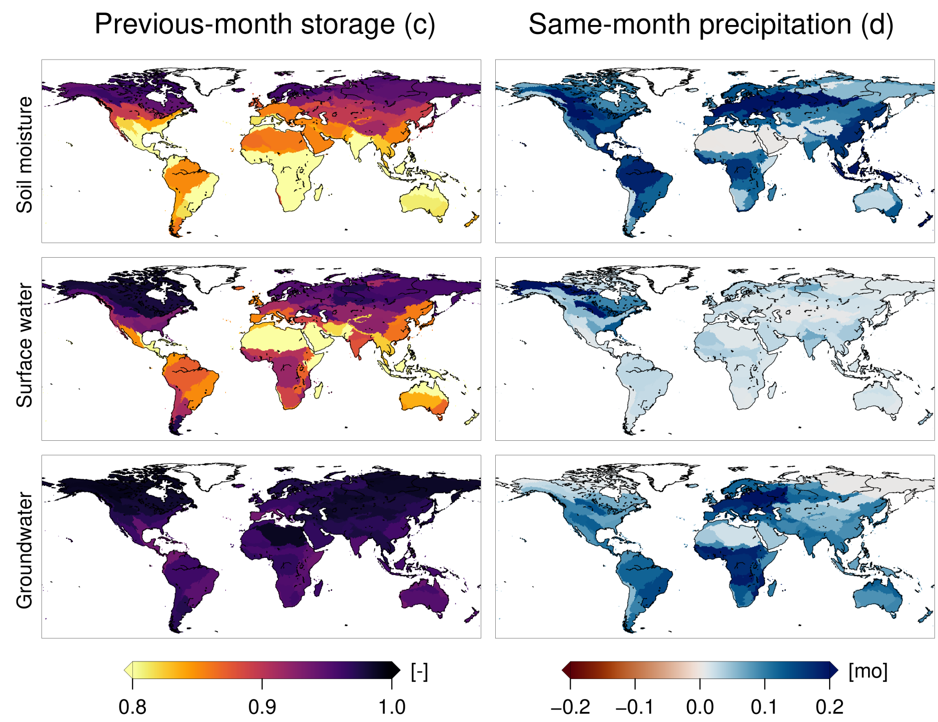


Figure 6.25: Auto-regressive process coefficients (c and d, Eq. 6.25) of order 1 for water storage of the current month to previous-month storage [-] (left) or of precipitation relative to soil moisture (top), surface water (middle), or groundwater (bottom) for the same month [months] (right). The coefficients are estimated using an auto-regressive process model from 2003 to 2019 for the non-seasonal part of the data. The closer the $c$-coefficients are to one, the stronger the water storage of the current month depends on the previous-month storage. The more the $d$-coefficients depart from zero, the stronger the storage depends on same-month precipitation.

Table 6.5: Percentage of clusters per water storage of the GLWS assimilation that experience an annual relative phase towards LAI that is longer, shorter, or equally long as the WaterGAP model simulation or that experiences invalid values (not a number, nan) because of invalid LAI values.

| **Storage** | **Percentage of clusters** [%] | | | |
|---|---|---|---|---|
| | **longer** | **shorter** | **equal** | **nan** |
| Soil moisture | 6.85 | 16.44 | 64.38 | 12.33 |
| Surface water | 19.18 | 17.81 | 50.68 | 12.33 |
| Groundwater | 34.25 | 21.92 | 31.51 | 12.33 |

within this work.

LAI is a biophysical variable suitable as a measure for vegetation growth and thus also crop growth. It is defined as the leaf area per unit horizontal ground area. As it relates two areas to each other, it is a unitless measure ($m^2/m^2$). In this thesis, I consider the LAI from the data sets SPOT/VEGETATION and PROBA-V on 1 km spatial resolution version 2 (copyright CNES and distribution by VITO), which can be downloaded from the Copernicus Global Land Service. The data set is spatially aggregated to the GLWS resolution of 0.5° and temporally averaged to transform the 10-day data to monthly values. Some grid cells for the LAI experience a bias, which is visible in a jump in the time series after 2013 and a different magnitude of amplitudes because the data are based on two satellite missions. Therefore, a simple method for minimizing the biases is applied. The temporal mean and annual amplitudes are computed for 2003 to 2013 and for 2014 to 2019 separately and used to adapt the 2014 to 2019 time series by removing the bias and applying a multiplication factor.

The previous section analyzed the timing of water peaks for the seasonal and non-seasonal patterns in the precipitation-storage dynamics. As it was decided to concentrate on the non-seasonal signature of the data, I provide a short summary (Tab. 6.5) on the comparison between the model simulations and the assimilation for timing between seasonal water storage and vegetation peaks as measure with the vegetation metric LAI via relative amplitudes estimated for each biome cluster (Sec. 6.1.3). The computation of relative amplitudes is similarly done as for the precipitation-water storage dynamics (Eq. 6.10). Considering the dynamic soil moisture or surface water seasonal peaks contributing to seasonal vegetation growth, no large differences between the model simulation and GLWS are observed and more than half of the clusters show equal relative amplitudes. For seasonal groundwater peaks contributing to seasonal LAI growth, only 31.51% of the clusters show an equal relative phase for assimilation and model simulation. In the assimilation, more clusters experience a longer relative amplitude of the groundwater-LAI dynamics as compared to the model simulation. Spatial details and the same analysis applied for the GLWS release 2 can be found in the attachment (Fig. A.9 and A.11), which leads to similar findings. It should be noted, that for the assimilation-derived storages the percentages of relative phases shorter, longer, or equal to model simulations do not add up to 100%, due to some clusters that experience invalid values. The invalid values mainly occur in the north due to the properties of missions used for the LAI.

Let us now analyze the water-storage-vegetation dynamics for the non-seasonal residual signal. Again, the seasonal signals, linear trends, and constant biases are removed from the data sets to derive a residual signal containing the non-seasonal pattern. Fig. 6.26 shows the lags between the deseasoned water storages from assimilation and the deseasoned LAI estimated from cross-correlation analysis. Positive lags indicate that a non-seasonal event in water storage is followed by a non-seasonal event in vegetation, thus vegetation growth or decrease is observed after a storage change. Negative lags indicate that a non-seasonal event in storage is found after a non-seasonal event in vegetation. As a reminder: The lags between non-seasonal precipitation and water storages showed positive values (Sec. 6.3.2), which indicates that first there is a event in precipitation, which is then followed by a event in the storages. For the lags estimated between the storages and the LAI this behavior is similar for soil moisture considering deseasoned signals: a soil moisture peak is followed by a peak in LAI. The same behavior can only be observed for a very few clusters for surface water and groundwater. Surface water shows many clusters with a lag of zero towards LAI, which means both variables are peaking at the same time. The groundwater lags to LAI are (slightly) negative for many clusters, which means that the peak in LAI is followed by a peak in groundwater. Thus LAI does not clearly respond to groundwater for many clusters but rather to soil moisture and surface water. Therefore, one could assume that a non-seasonal even with increasing soil moisture values often faster leads to vegetation growth than it refreshes the groundwater storage. A second reason for the negative lags and lags around zero might result from anthropogenic influences. Farmers might irrigate their fields in response to dry vegetation conditions and the water for irrigation is taken from surface water or groundwater (Fig. 4.2).

Only for a few humid regions, the lags between groundwater storage and LAI are positive, which means that a change in non-seasonal groundwater could be an indicator of non-seasonal vegetation growth here. In view of the prototype drought monitoring system that is set up in this thesis, soil moisture seems to be the better proxy for vegetation growth, while surface water and groundwater can better be integrated into systems that focus on other impacted economic sectors, e.g. water supply. Nonetheless, as a few clusters show zero or positive lags of the groundwater-LAI dynamic, groundwater still can positively support drought warning systems built on soil moisture and vegetation.

To identify relevant relations between the vegetation of the current month to previous-month vegetation ($e$-coefficients) and compare it to the dependency of water storage and LAI for the same month ($f$-coefficients), Fig. 6.27 shows the auto-regressive process coefficients for order one (Sec. 6.1.2). This time, the process model seems to reflect reasonable dependencies of same-month storages towards vegetation because soil moisture shows many clusters that lead to more vegetation growth as groundwater. The groundwater-vegetation interaction of the same month seems to be nearly zero, whereas the surface water leads to more vegetation growth than groundwater but less than soil moisture, and thus is found in between both storages. However, it is striking that the values are generally low, whereas the relation between vegetation of the current month to previous-month vegetation is very close to one, which means a high dependency of same-month vegetation to the previous month exists independently of if the process model for vegetation is estimated jointly with soil moisture, surface water, or groundwater.

In summary, the results indicate that soil moisture is the main driver for vegetation growth but in some regions also surface water and groundwater contribute to the growth with time delays, especially for the seasonal part of the signals. In the case of using the residual part of the data, many clusters were found that showed an immediate response of vegetation growth to changes in surface water or – reversely – vegetation growth followed by changes in groundwater. Both findings might result from anthropogenic water use (Fig. 4.2).

## 6.4 Validation of the global assimilation framework

To validate the outputs of GLWS, independent in-situ data of different observation types are used. TWSA can be transformed to vertical displacement time series, therefore I include GNSS vertical displacements for the validation of all water compartments (Sec. 6.4.1). The surface

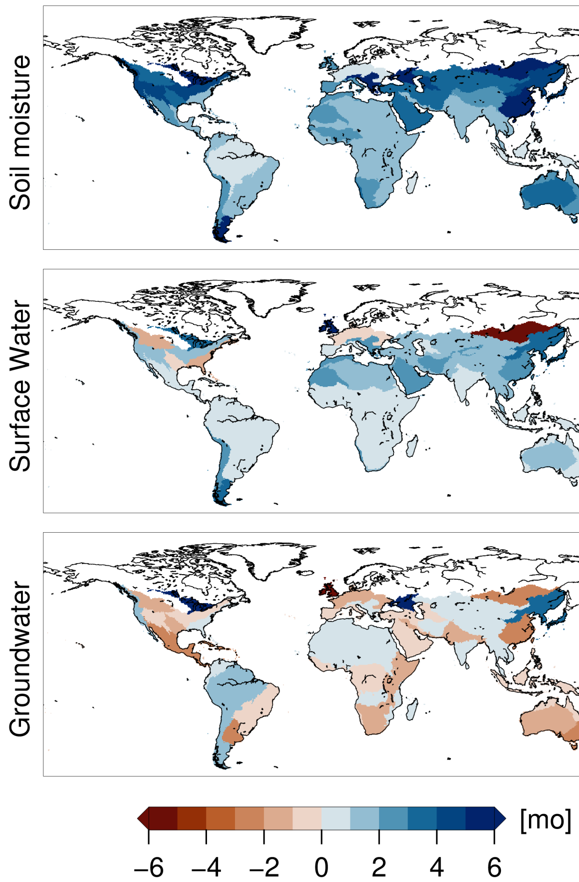


Figure 6.26: Lags [months] of soil moisture (top), surface water (middle), or groundwater (bottom) from assimilation relative to LAI. The lags are estimated using cross-correlation analysis from 2003 to 2019 for the non-seasonal residual part of the data. Positive lags mean that the LAI is following the water storage and negative lags indicate the reverse.

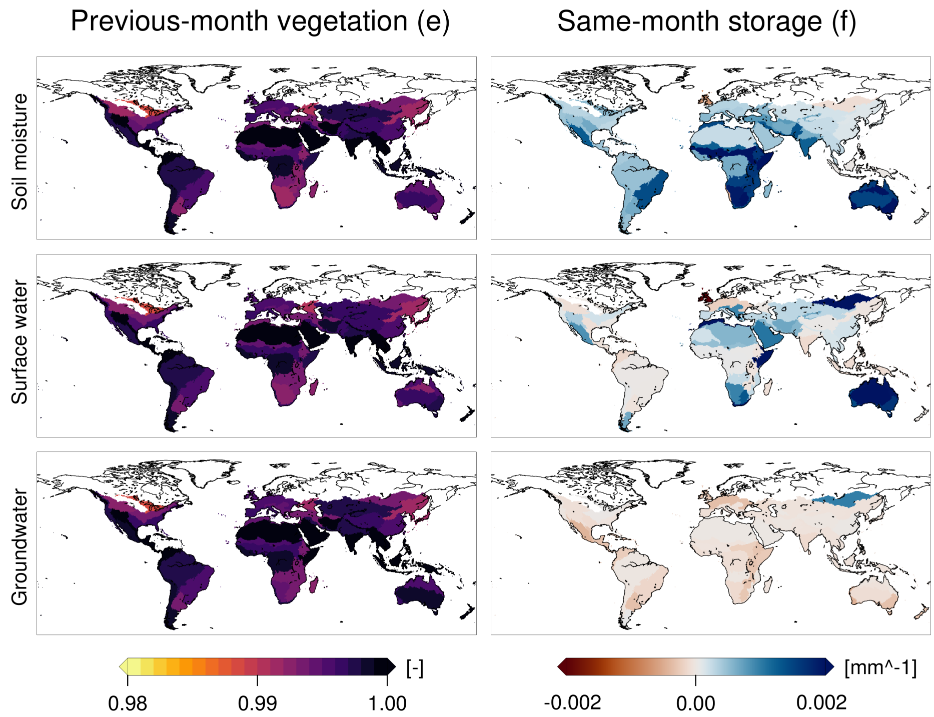


Figure 6.27: Auto-regressive process coefficients (e and f, Eq. 6.26) of order 1 for vegetation observed with LAI of the current month to previous-month vegetation [-] (left) or of vegetation relative to soil moisture (top), surface water (middle), or groundwater (bottom) for the same month [$mm^{-1}$] (right). The coefficients are estimated using an auto-regressive process model from 2003 to 2019 for the residual part of the data. The closer the $c$-coefficients are to one, the stronger the vegetation of the current month depends on the previous-month vegetation. The more the $d$-coefficients depart from zero, the stronger the vegetation depends on same-month storage.

water from assimilation is evaluated against the relocation data set RECOG-LR (Sec. 2.3.2) as this data set includes independent lake and reservoir storage information from remote sensing. Groundwater is compared against available borehole data in France and East Brazil. To evaluate the performance of soil moisture, GLWS is compared against the ESA-CCI global soil moisture product (6.4.4).

### 6.4.1 GNSS

The GLWS data sets release 2 was validated against GNSS vertical displacement time series from over 1030 stations worldwide in Gerdener *et al.* (2023a). I here now take the same validation data of vertical displacements and use it for validation of the new release 3. The station data is downloaded from the International GNSS Service (Repro 3, Rebischung, 2021) and further processed by the Military University of Warsaw in Poland (personal communication with Anna Klos). The processing includes outlier removal with the Interquartile Range rule, manual jumps removal from comparing various GNSS time series, and the removal of atmospheric and oceanic loading effects via predictions from the GFZ Earth System Modeling group.

Following Eq. 2.25, the TWSA from assimilation, satellite observation, and model simulation are transformed to vertical displacements. Fig. 6.28 shows histograms of how many of the 1032 GNSS vertical displacements have a certain correlation (top row) or RMSD (bottom row) to either the GLWS assimilation (blue), the WaterGAP model simulation (orange), or the GRACE/-FO observation (grey) vertical displacements. This analysis is applied for the detrended signal of vertical displacements (left), the seasonal signal as estimated via multi-linear regression (center), or the non-seasonal residual signal (right). Generally, more GNSS stations have a higher correlation to the assimilation as compared to the model simulations when considering the detrended signal of the data. In fact, for 510 stations the correlation of GLWS with GNSS is larger than 0.5, whereas for the correlation between WaterGAP and GNSS 486 stations have a correlation of 0.5 or higher. This is partly a result of the slight improvements in correlation for the non-seasonal part of the GLWS signal. Moreover, for the non-seasonal part of the data, the correlations of the observations with GNSS are lower as compared to the model simulation and assimilation. Thus, assimilation seems to derive improved correlations to GNSS as a combination of both, model simulations and observations for this part of the data. For the seasonal signal, the observations, the simulation, and the assimilation show a high number ($> 650$) of stations with a correlation higher than 0.5 towards GNSS with 754 stations, 683 stations, and 736 stations, respectively. Thus, for the seasonal signal, the assimilation seems to be slightly closer to the GNSS data as compared to the model simulation but not as close as the observation. When analyzing the RMSD, the same conclusion can be drawn. For example, the spatially median RMSD between vertical displacements from GLWS and GNSS is lowest for the non-seasonal residual signal when compared to the RMSD of vertical displacements between WaterGAP model or GRACE/-FO observations and GNSS.

So in summary, the new release of GLWS shows improvements compared to the GRACE/-FO data and model simulations for the non-seasonal part of the data and improvements compared to WaterGAP simulation when considering the seasonal signal. In contrast, for the release 2 of GLWS, a correlation of higher than 0.5 towards GNSS is found for 754 stations when considering the seasonal signal and for 174 stations when considering the non-seasonal residual signal (Fig. A.12). This means that the correlation towards GNSS improved from release 2 to release 3 mainly for the non-seasonal signal.

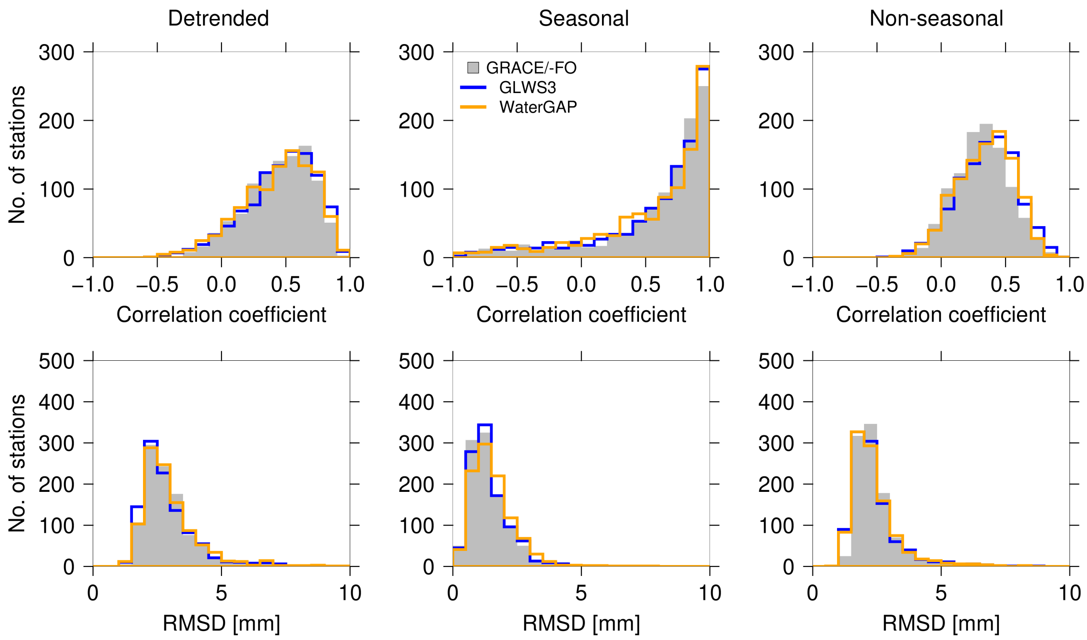


Figure 6.28: Correlations (top) and RMSD (bottom) of vertical displacements from GNSS station worldwide with vertical displacements from GRACE/-FO observations (grey), GLWS assimilation (blue), and WaterGAP simulations (orange) estimated for 2003 to 2019. Additionally, the correlations and RMSD are either shown for the detrended signal (left) or for the seasonal (center) and the non-seasonal residual part (right) part of the data.

### 6.4.2 Groundwater wells

The validation of groundwater storage presents many challenges. The most obvious one is that in-situ/borehole data are often provided as water levels (GL), whereas assimilated or simulated groundwater is given as water storage or water storage anomalies (GWA). To combine both, a specific yield factor ($S_y$) is required that transforms the groundwater levels to anomalies via

$$\mathrm{GWA} = \mathrm{S}_y \cdot \mathrm{GL}. \tag{6.32}$$

The specific yield has a major influence on comparing storage and well data. It is typically derived from pumping tests that are executed for the corresponding well (e.g., Ramsahoye & Lang, 1961) or estimated from literature research via comparison of rock types and tables with the region-specific soil characteristics.

For this thesis, well data are available for East Brazil and France. East Brazil well levels are downloaded from the Geological Survey of Brazil (CPRM – Companhia de Pesquisa de Recursos Minerais[2]). The specific yield factors for East Brazil are taken from Li *et al.* (2019) and used to compute groundwater storage anomalies per site. The data available for France is provided by the HafenCity University of Hamburg in the context of the GlobalCDA project[3] (personal communication). Due to its use within the project, groundwater head variation sites are area-averaged onto the 0.5° WaterGAP model grid and thus provided for six French basins (the Adour, Garonne, Loire, Rhone, Seine, and Vilaine basins) on the same spatial resolution as used within this thesis. As the number of station data for East Brazil is very low with only 61 sites, the model simulations and assimilation grid cells are interpolated to the stations instead. Similar to other studies, Li *et al.* (2019) choose only unconfined aquifers because the groundwater level or

[2] `https://www.sgb.gov.br/` (last accessed 19.05.2024)
[3] `http://globalcda.de/` (last accessed 23.03.2024)

Table 6.6: Spatially median correlation [-], NSE [-], and KGE [-] for East Brazil and for six basins in France for comparing groundwater anomalies from in-situ observations with the GLWS assimilation, or with the WaterGAP model simulations. The metrics are estimated for the period 2003 to 2016 and are either shown for the full signal or for the seasonal and non-seasonal part of the groundwater.

| | | GLWS | | | WaterGAP | | |
|---|---|---|---|---|---|---|---|
| | **Signal** | **Corr.** | **NSE** | **KGE** | **Corr.** | **NSE** | **KGE** |
| East Brazil | Full | 0.54 | -3.32 | -0.44 | 0.43 | -26.76 | -3.50 |
| | Seasonal | 0.81 | 0.28 | 0.33 | 0.81 | 0.29 | 0.26 |
| | Non-seasonal | 0.48 | 0.05 | 0.12 | 0.39 | 0.08 | -0.01 |
| France | Full | -0.17 | -4.35 | -0.53 | -0.03 | -21.53 | -2.60 |
| | Seasonal | -0.45 | -0.44 | -0.66 | 0.22 | 0.00 | -0.16 |
| | Non-seasonal | -0.08 | -0.26 | -0.28 | -0.04 | -0.03 | -0.33 |

Table 6.7: Spatially median linear trends [mm/yr], annual amplitudes [mm] and semi-annual amplitudes [mm] for East Brazil and for six basins in France for comparing groundwater from in-situ stations with groundwater from GLWS assimilation or WaterGAP model simulations. The metrics are estimated for the period 2003 to 2016.

| | **Hydrological signature** | **GLWS** | **WaterGAP** | **In-situ** |
|---|---|---|---|---|
| East Brazil | Linear trend [mm/yr] | 2.38 | 0.21 | 0.88 |
| | Annual amplitude [mm] | 33.70 | 19.18 | 29.39 |
| | Semi-annual amplitude [mm] | 2.97 | 2.48 | 5.84 |
| France | Linear trend [mm/yr] | −0.44 | 0.01 | 1.13 |
| | Annual amplitude [mm] | 22.74 | 9.70 | 40.92 |
| | Semi-annual amplitude [mm] | 4.44 | 2.56 | 6.91 |

storage data should contain a seasonal signal, which is not the case for confined aquifers because they can be fully saturated. I follow this approach here.

Table 6.6 summarizes spatially median metrics for East Brazil and the six basins in France with correlations, NSE, and KGE for the comparison of the in-situ groundwater anomalies with the GLWS, or WaterGAP groundwater. For East Brazil, correlations between GLWS and in-situ groundwater are higher than for WaterGAP. The largest improvement is found in the non-seasonal signal (also supported by the KGE) of the data but overall, the correlations are medium-high with values between 0.48 and 0.39. Thus, the assimilation of GRACE/-FO TWSA shows improvements in the groundwater in East Brazil. Unexpectedly, France does not show such improvements. The correlations between WaterGAP and the in-situ groundwater are slightly better than for GLWS, however, all correlations are generally very low or even negative and do not indicate a strong correlation to the in-situ groundwater.

When extracting linear trends, annual amplitudes, and semi-annual amplitudes of the three data sets (Tab. 6.7) – again as spatial median metrics for East Brazil and the French basins – an improved seasonality is found. For example, after assimilation, the groundwater storage spatially median annual amplitude is – with 33.70 mm – closer to the in-situ annual amplitude of 29.39 mm as compared to the model simulation with an amplitude of 19.18 mm. Thus, the assimilation indeed improves parts of the signal, but the correlations could not show this effect probably because there exists a temporal delay between the data sets. For the NSE and the KGE, low values of 0.32 or lower are found, which could also result from a temporal lag.

At this stage, it can already be summarized, that the assimilation has a positive impact on groundwater for East Brazil, especially for improving the seasonality. However, the data in France showed deterioration after assimilation, and linear trends of GLWS differ much more from in-situ data as compared to the WaterGAP simulations for both areas. It is possible that the assimilation indeed deteriorates linear trends very locally, but this is not expected in general, as the groundwater trends for GLWS match with findings of Rodell *et al.* (2018) as described in Sec. 6.3.1. Further reasons could be responsible for the mismatch:

1. **Length of time series:** Trend estimation from short time series is less representative than for long periods. As the overlapping period of in-situ groundwater for East Brazil with model simulations and assimilation outputs is a maximum of 110 months, this might not be representative. The length of groundwater in-situ measurements for France is about 168 months, which has better temporal coverage.

2. **Location characteristics:** A small study region can lead to a less representative comparison, as is the case for France. The total area of France is much smaller than East Brazil. Furthermore, France is directly located at the coast, and, as shown in Sec. 2.3.2, GLWS has larger uncertainties at the coast due to leakage.

3. **Specific yield factor:** Uncertainties in the estimation of the specific yield factor might bias the comparison. It can be expected that metrics like the linear trend or amplitude for the in-situ measurements might increase or decrease in size (trends would not change the sign).

4. **Specific sensitivities of metrics:** The comparison of seasonal correlations and extraction of amplitudes shows that different metrics can analyze the same part of the data but lead to different conclusions.

Most of these aspects cannot be largely influenced by the user, because the availability of temporally and spatially consistent data in the right format is time-consuming or sometimes even restricted, especially for in-situ data. Since the choice of metric is up to the user, the analysis is now expanded for a better understanding of the results for France. Tab. 6.8 shows the correlation and cross-correlation between spatial median in-situ groundwater storage and the assimilation or the model simulation for six different hydrological basins in France: the Adour, Garonne, Loire, Rhone, Seine, and Vilaine basins. It can clearly be extracted, that the cross-correlations are higher than the correlations. Considering the full groundwater signal, the assimilation has a stronger cross-correlation to the in-situ measurements than the model simulations in the Garonne, Loire, Rhone, and Vilaine basins. This finding confirms the assumption that time lags bias the improvements in signal magnitude and course. The choice of metrics is therefore very important to separately uncover several improvements or deterioration of the assimilation compared to the validation data. In conclusion, the assimilation also has a positive impact on groundwater in some of the sub-basins of France but was either hidden by a temporal delay in combination with the choice of metric or the spatial averaging that is applied to derive one final metric for the whole of France from the six basins.

All in all, it can be summarized that improvements in the groundwater from assimilation could be confirmed with in-situ data, especially for East Brazil. Still, further in-depth analysis should be taken up in future research with a global more consistent coverage of stations and temporally longer in-situ measurements with various statistical metrics to understand improvements and limitations of the assimilation more precisely.

Table 6.8: Correlation and cross-correlation [-] for comparing spatial median groundwater anomalies from in-situ observations with the GLWS assimilation, or with the WaterGAP model simulations for six french basins: The Adour, Garonne, Loire, Rhone, Seine, and Vilaine basins. The metrics are estimated for the period 2003 to 2016, and are either shown for the full signal or for the seasonal and non-seasonal part of the groundwater.

| | | **GLWS** | | **WaterGAP** | |
|---|---|---|---|---|---|
| | **Signal** | **Corr.** | **X-Corr.** | **Corr.** | **X-Corr.** |
| Adour | Full | -0.26 | 0.00 | -0.25 | -0.23 |
| | Seasonal | -0.32 | 0.87 | -0.02 | 0.95 |
| | Non-seasonal | -0.12 | 0.04 | -0.20 | -0.14 |
| Garonne | Full | -0.22 | 0.46 | 0.03 | 0.08 |
| | Seasonal | -0.31 | 0.95 | 0.55 | 0.89 |
| | Non-seasonal | -0.15 | 0.04 | -0.12 | -0.07 |
| Loire | Full | -0.18 | 0.12 | 0.06 | 0.10 |
| | Seasonal | -0.31 | 0.94 | 0.76 | 0.95 |
| | Non-seasonal | -0.09 | 0.06 | 0.32 | 0.44 |
| Rhone | Full | -0.10 | 0.45 | -0.03 | 0.02 |
| | Seasonal | -0.28 | 0.90 | 0.49 | 0.91 |
| | Non-seasonal | 0.18 | 0.37 | 0.15 | 0.15 |
| Seine | Full | -0.31 | 0.42 | 0.27 | 0.46 |
| | Seasonal | -0.41 | 0.95 | 0.44 | 0.92 |
| | Non-seasonal | -0.32 | 0.17 | 0.07 | 0.10 |
| Vilaine | Full | -0.50 | 0.44 | 0.18 | 0.34 |
| | Seasonal | -0.85 | 0.94 | 0.44 | 0.82 |
| | Non-seasonal | -0.29 | 0.22 | -0.10 | 0.17 |

Table 6.9: Global median metrics for linear trends [mm/yr], annual amplitudes [mm] and semi-annual amplitudes [mm] for lakes and reservoirs from RECOG-LR for comparing surface water from remote sensing with surface water from GLWS assimilation or WaterGAP model simulations. The metrics are estimated for the period 2003 to 2016.

| **Signal component** | **GLWS** | **WaterGAP** | **RECOG-LR** |
|---|---|---|---|
| Linear trend [mm/yr] | −0.04 | 0.00 | −0.14 |
| Annual amplitude [mm] | 7.79 | 6.52 | 35.05 |
| Semi-annual amplitude [mm] | 2.75 | 2.07 | 6.89 |

### 6.4.3 Surface water

As the lake and reservoir relocation data set (Sec. 2.3.2) is produced by converting independent altimetry and remote sensing lake levels to water storage anomalies and was not included in the final GLWS release, it presents an opportunity now for having an independent evaluation data set. To make the assimilation and model simulations comparable to RECOG-LR, only lakes and reservoirs are accounted for because the relocation data set does not contain rivers and wetlands. Cells that are zero in RECOG-LR are removed from all data sets. In addition, cells are removed that do not cover the time frame from 2003 to 2016 to provide as long time series as possible. Before analyzing the results, it should be mentioned that the surface water storage from RECOG-LR is a data set that is produced by merging observations of remote sensing with altimetry and underlies a static water body area. A comparison against RECOG-LR can indicate improvement but also contains many processing steps and plans for improvements on the validation data set side. Thus, the results should be taken with care.

Tab. 6.9 shows spatially median signatures of linear trends, annual amplitudes, and semi-annual amplitudes for the GLWS assimilation, the WaterGAP model simulations, and the surface water storage from RECOG-LR. An improvement is found for the linear trends from assimilation as compared to the model simulations with trends of 1.16 mm per year for RECOG-LR, 7.05 mm per year for GLWS, and 13.82 mm per year for WaterGAP. In addition, the assimilation is closer to RECOG-LR as compared to the model when considering the annual and semi-annual amplitudes. When computing a histogram of linear trends differences between RECOG-LR and the assimilation or RECOG-LR and the model simulations (Fig. 6.29), no clear tendency is found if the model simulations or the assimilation linear trends compare better to RECOG-LR. Furthermore, correlations and cross-correlations indicate that the model simulation is only very slightly closer to the relocation data set as compared to the assimilation. For example, the cross-correlation for the seasonal surface water of RECOG-LR with the assimilation is 0.85, whereas the cross-correlation towards the model simulation is 0.88.

In fact, with the current assimilation setup, neither clear deterioration nor improvement is found, which is probably a result of groundwater being the largest component of TWSA and surface water and soil moisture being relatively small compared to groundwater.

### 6.4.4 Soil moisture

In this section, the soil moisture from assimilation and simulation is compared to the global soil moisture product provided by the ESA. The soil moisture product is part of the ESA Climate Change Initiative (CCI) that aims to produce an updated soil moisture product every year[4]. In fact, the combined active and passive soil moisture version v08.1 (Dorigo *et al.*, 2017; Gruber *et al.*, 2019; Preimesberger *et al.*, 2020; Dorigo *et al.*, 2023) is used which provides

[4]https://climate.esa.int/en/projects/soil-moisture/ (last accessed 23.08.2024)

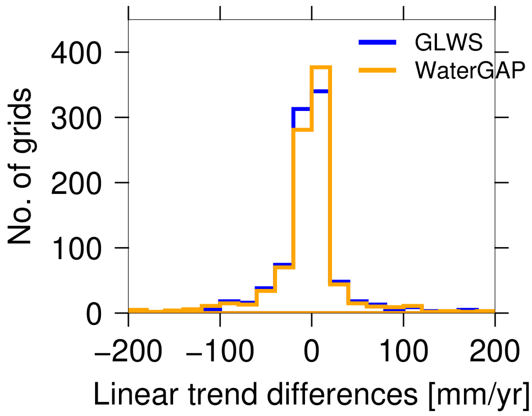


Figure 6.29: Linear trend differences [mm/year] between surface water from RECOG-LR with GLWS assimilation, or WaterGAP simulation estimated for 2003 to 2019.

soil moisture for approximately the first five centimeters of the soil. The data set is developed by combining multiple single sensor active and passive microwave soil moisture products (e.g., Dorigo *et al.*, 2023). The combined daily soil moisture data set is available from 1978 to 2022, here the time series is shortened to the period used in this thesis, i.e. 2003 to 2019. The daily values are averaged to monthly values and the grid cells are aggregated from their native 0.25° spatial resolution to the 0.5° spatial resolution of GLWS and WaterGAP.

A direct comparison of GLWS or WaterGAP soil moisture storage to the ESA-CCI soil moisture with the same units requires the transformation of soil moisture storage given in millimeters to volumetric soil moisture given in $m^3$ per $m^3$. Another possibility to perform a comparison is by normalizing all soil moisture products in advance of the comparison. This is commonly done in the soil moisture community (e.g., Koster *et al.*, 2009; Escorihuela & Quintana-Seguí, 2016) because soil moisture from models, remote sensing, and in-situ stations are often provided in different units. Thus, the assimilation, model simulation, and ESA-CCI soil moisture is now normalized by removing the temporal mean and dividing by the temporal standard deviation computed from all months per grid cell. Nonetheless, by applying the normalization, the data sets cannot be compared in magnitude but metrics such as correlations are still applicable because they do not change due to the normalization.

Fig. 6.30 shows the difference between the correlations of normalized soil moisture from GLWS with ESA-CCI and the correlations of normalized soil moisture from WaterGAP with ESA-CCI. If the differences are positive, this means that soil moisture from assimilation is closer to the ESA-CCI soil moisture as compared to the model simulation. If the differences are negative, the soil moisture from the simulation is closer to the ESA-CCI soil moisture. Positive correlation differences are, for example, found in Southern Africa, Northern Australia, West U.S. south of the U.S. Great lakes, and western China. In these regions, the assimilation correlates better to the ESA-CCI soil moisture as compared to the simulations. In contrast, the soil moisture simulations correlate better to the ESA-CCI compared to the assimilation in northern latitudes, for example, in Scandinavia and in the north of Siberia. However, although it is possible to identify regions where the assimilation shows improved or decreased correlations towards the validation data set compared to the simulation, the assimilation and model simulation are very close to each other. For example, for 70% of the land grid cells, the RMSD between the assimilation and model simulation is lower than 2 mm. Again, I assume that the assimilation majorly improves the groundwater storage instead of surface water and soil moisture, because groundwater is the

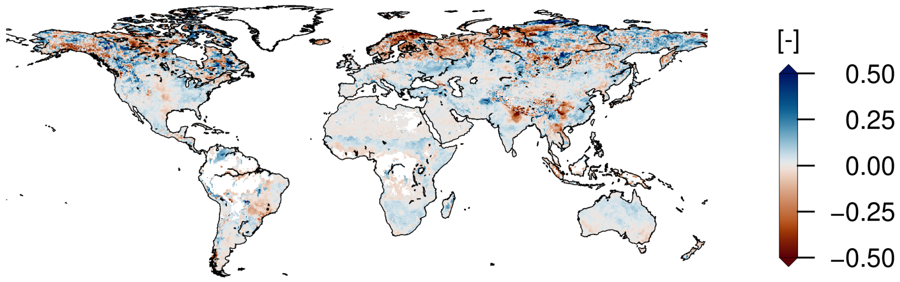


Figure 6.30: Difference between the correlations of normalized soil moisture from ESA-CCI with GLWS and ESA-CCI with WaterGAP. Positive correlation differences indicate that GLWS fits better to ESA-CCI than WaterGAP. Negative correlation differences indicate that WaterGAP fits better to ESA-CCI than GLWS. The correlation differences are unitless.

largest contributor to TWSA.

Since the soil moisture comparison is based on normalization, the correlation provides an indication of improvements/deteriorations for soil moisture from assimilation versus the simulation. However, as mentioned before, the two data sets differ in their definitions of soil moisture depth. As explained in Sec. 4.4, WaterGAP defined soil moisture within the effective root zone that can be up to 4 m deep depending on the land cover, whereas the ESA-CCI product refers approximately to the first 5 cm of the soil (e.g., Dorigo *et al.*, 2023). In addition, the ESA-CCI soil moisture can have limitations in densely vegetated and snow-covered areas. In summary, the comparison showed that on average the assimilation and simulation both fit equally well to the ESA-CCI soil moisture. In the future, a more expanded evaluation should be considered that performs the transformation of the data sets to the same units to also compare the magnitude. Furthermore, one should use approaches for referring the data within different depths of soil to each other and includiung more in-depth regional analyses while considering region-specific limitations of the ESA-CCI soil moisture.

# Chapter 7

# Drought monitoring with GRACE/-FO and assimilation outputs

This chapter provides a detailed picture of the envisaged application of GRACE/-FO satellite observations and assimilation results for drought monitoring. Direct validation data for droughts is difficult to derive because there is no unique definition of drought and common drought databases quantify only the impact of drought on some aspect (e.g., human life). Within a synthetic study, existing GRACE drought indices and their performance are tested towards signals that are typically contained in the GRACE satellite data, for example, linear trends. After deciding on a final index for the global application, the results for monitoring droughts with GRACE/-FO and outputs from assimilating GRACE/-FO into WaterGAP are presented. After that, hazard risk maps are computed and a prototype warning system is set up incorporating precipitation, water storages, and vegetation to observe drought conditions in different fluxes and storage compartments. The chapter closes with an evaluation of the results of monitoring drought events with two newly developed prototype warning systems.

## 7.1 Reliability of GRACE/-FO indices

A huge problem in studying the performance of drought monitoring is that no direct validation data based on observations exists for drought and no universal method for defining drought indices. A few databases record drought events like EM-DAT database[1], the National Drought Mitigation Center[2], the U.S. Drought Monitor[3], etc. by using economic and ecologic damage values. But as there exists no explicit objective measurement of drought and there is no universal consensus about the determination of drought events open questions for the performance and sensitivities of the drought indices that were presented in Sec. 3.4.3 remain:

- How to validate the drought indices?
- How are trends biasing drought indices?
- How is the GRACE/-FO specific spatial and temporal noise influencing the drought monitoring?
- Is there a preferred method for deriving index values as a spatial average?

[1] https://www.emdat.be/ (last accessed 22.05.2025)
[2] https://drought.unl.edu/ (last accessed 22.05.2024)
[3] https://droughtmonitor.unl.edu/ (last accessed 22.05.2024)

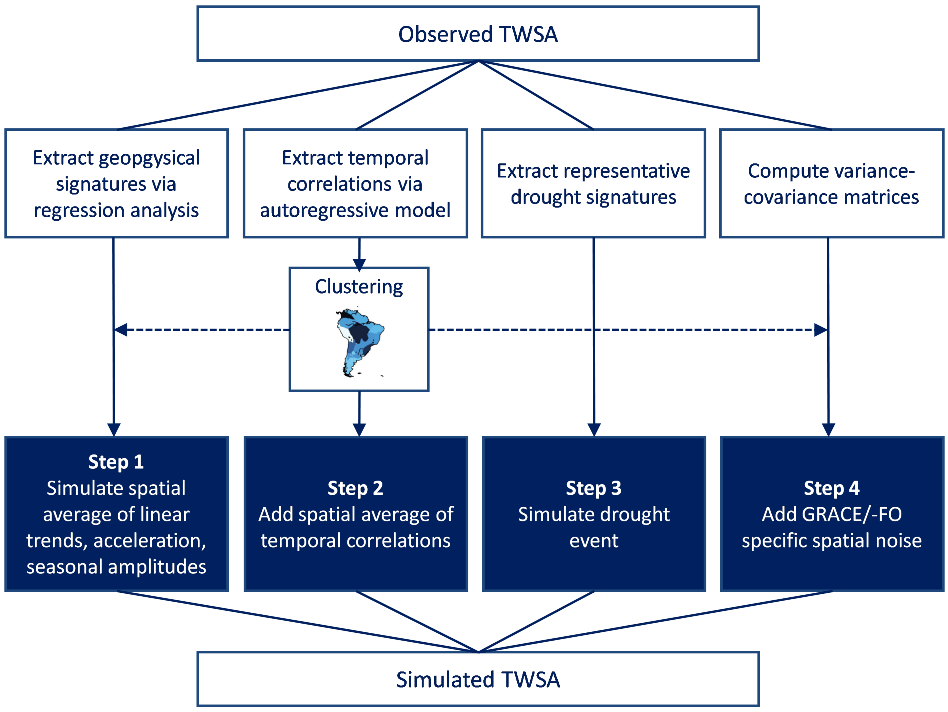


Figure 7.1: Schematic illustration of the synthetically generated TWSA that includes the simulation of trends, seasonality, spatial and temporal noise, and a simulated drought event.

- What is the effect of accumulating or differencing storage information before index computation as, for example, typically done for SPI?
- Is the time series length suitable for drought monitoring?

To answer these questions and provide a measure for validation, a synthetic study is set up, aiming to simulate synthetic TWSA time series with GRACE/-FO specific characteristics and one drought event. Different drought indices are computed that use the synthetic time series as input to investigate and get an understanding of how trends, noise, and other signals contained in the GRACE/-FO data bias the drought detection. This section is based on Gerdener *et al.* (2020a) but slightly adapts the development of the synthetic TWSA and introduces new indices.

### 7.1.1 Synthetic study setup

In this part, I extract relevant geophysical signatures and noise from real GRACE/-FO TWSA to simulate synthetic TWSA time series including a drought event for the drought index evaluation. With the help of this synthetic data, one can identify whether the application of drought indices correctly monitors the drought or if the application changes important aspects of the drought, such as the timing or duration. It would also be possible, to use TWSA from model simulations as synthetic truth for this study but then the effect of GRACE/-FO specific spatial and temporal noise on the drought detection could not be investigated. Some details of the simulation are shown in Fig. 7.1 and are now further elaborated.

In the beginning, observed TWSA from the satellites on a global 0.5° grid from 2003 to 2019 are analyzed with respect to dominant temporal correlations in the TWSA via an auto-regressive process model that relates TWSA of the current month $t$ at grid $j$ to multiple previous months as

$$x_{j,t} = \psi_{j,1} x_{j,t-1} + \psi_{j,2} x_{j,t-2} + ... + \psi_{j,q} x_{j,t-q}, \tag{7.1}$$

where $q$ is the order of the process model and $\psi$ are the auto-regressive process coefficients. Similar to in Gerdener *et al.* (2020a), I found that most global land cells experience a dominant temporal correlation concerning the previous-month storage, which more precisely is described by an order one auto-regressive process model. Thus, an order one model is set up and the process coefficients $\psi$ are derived per grid cell and used as a basis for finding spatial coherent regions with similar spatiotemporal properties via EM-clustering (Sec. 6.1.3). The final clusters are shown in Fig. 7.2.

Next, the real GRACE/-FO TWSA are analyzed with respect to existing geophysical signals in the data. Via multi-linear regression (Sec. 6.1.2) linear trends, accelerations, annual signal, and a semi-annual signal are extracted per grid cell. The signatures are spatially averaged for the identified clusters. One synthetic time series per cluster is created by applying the spatial average signature coefficients to the time (Eq. 6.7). The estimated spatial averaging process coefficient of the auto-regressive model from the real data is used to apply an order one auto-regressive model to the synthetic data to add temporal noise for the basin averages per cluster (cmp. Gerdener *et al.*, 2020a:Appendix A). Simulating a drought event is difficult as the onset of drought and water deficit varies from region to region and from drought event to drought event. Gerdener *et al.* (2020a) showed that for droughts in the Amazon basin, Texas, and Europe deficits from -23 mm to -350 mm are found with a duration of three to ten months. To simulate a drought event that is clearly visible in the data, I follow the approach of Gerdener *et al.* (2020a). A synthetic drought event is added at the beginning of the synthetic TWSA, starting in January 2005 and finishing in September 2005 with a constant water deficit of -100 mm for the Southern Brazil and -50 mm for the Southern Africa cluster. Until here, the synthetic TWSA time series results in a single time series per cluster that represents spatial average geophysical signatures, spatial average temporal noise, and a synthetic drought event.

As a final step for the simulation, GRACE/-FO specific spatial error characteristics shall be included. As the GRACE/-FO spatial correlations are very mission-specific and grid-dependent, the aim is to simulate the final TWSA time series for each 0.5° grid within a cluster. In the first step, the single time series of length $n$ per cluster (produced from signature averages before) is reproduced $m$ times yielding in a matrix $\mathbf{T}_{\mathrm{m \times n}}$. For each time step, a gridded full variance-covariance matrix is computed for the chosen grid within each cluster from the real GRACE/-FO satellite data (Sec. 2.4) and decomposed by Cholesky decomposition ($\mathbf{\Sigma} = \mathbf{R}^T\mathbf{R}$). By multiplying the decomposed matrix $\mathbf{R}_{\mathrm{m \times m}}$ with a random noise vector $w_{\mathrm{m \times 1}}$, a vector $\epsilon_{\mathrm{m \times 1}}$ is yielded that presents the spatial noise characteristics of GRACE/-FO. By gathering the spatial noise vectors for each time step, the matrix $\mathbf{V}_{\mathrm{m \times n}}$ is computed and can be added to the simulation in matrix $\mathbf{T}$. Here, the procedure is applied for two clusters (orange clusters in Fig. 7.2) to represent different characteristics of GRACE/-FO. The first cluster is located in Southern Brazil and has a significant trend and annual amplitudes representing a humid vegetation region. The second cluster represents an arid regime in Southern Africa and thus shows a strong spatial and temporal noise and no significant trend and seasonality. One could also choose multiple clusters that cover more different hydrological regimes than those two but to limit the size of the study the number of clusters is limited for showcasing.

The final simulated and spatially average time series are shown in Fig. 7.3, indicating a significant negative trend and seasonal variations for the Southern Brazil cluster and a strong noise for the Southern Africa cluster. The drought event (grey background) in 2005 is clearly visible. The

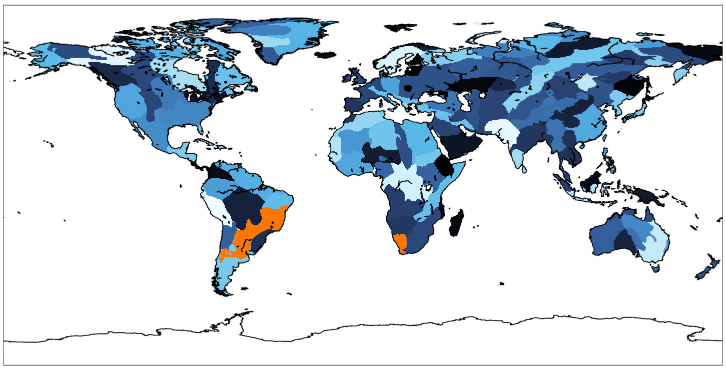

Figure 7.2: Clusters derived from EM-clustering of the auto-regressive process coefficients order one from real GRACE/-FO TWSA. The two clusters used for simulation are marked in orange.

following performance study of indices is based on the indices presented in Sec. 3.4.3: The DSI, DI, TSDI, and GGDI that use the synthetic time series as input.

### 7.1.2 Results of the synthetic study

#### Masking effect of trends, accelerations and seasonality

In this section, it is investigated how geophysical signals that are measured by the GRACE/-FO mission (hydrology, signals from the nearby ocean, melting of glaciers, GIA, etc.), can mask drought signals and consequently bias the detection of droughts. In fact, the analysis refers to systematic signals like linear trends, constant accelerations, and seasonal variations, which are simulated in step one of the simulation framework (Fig. 7.1). Each index might propagate these systematic signals differently. Figure 7.4 shows the severity classes of each month from 2003 to 2019 identified with the four indices presented in Sec. 3.4.3 (and their adaption with TWSA accumulations and differences): the DSI (top left), DI (top right), TSDI (bottom left), and GGDI (bottom right). The indices were computed with TWSA spatially averaged for the Southern Brazil cluster from 2003 to 2019. Let us first concentrate on the original indices without accumulation or differences of TWSA before index computation, which is found in row zero on the y-axis of Fig. 7.4 as the index does not use accumulation or differences.

All four indices show drought events in 2005 and from about 2014 to 2019. The detected drought in 2005 matches with the simulated drought event, thus the indices are able to detect the drought but the duration and severity vary. The intensity of the drought is severe or less intense for all four indices but the DI shows the lowest severity. Nonetheless, the identified severity of all indices is much too low given that we know that an intense deficit of -100 mm is simulated. The explanation for the fact that indices point to a less severe drought is the same why a drought event from 2014 to the end is monitored. A linear trend and acceleration in the synthetic time series lead to strong negative values towards the end of the time series and bias the drought monitoring. This would not have been the case when the drought event would also occur towards the end of the time series (Fig. B.1). In this case, the drought severity class might be overestimated. A possibility for avoiding such biases from linear trends or constant accelerations would be the removal of trends and accelerations from the time series in advance of index computation. However, distinguishing if a trend is contained in the data or if the trend is caused by a drought event at the very beginning or end of the time series is not always possible and the trend removal might also not perfectly work as droughts at the beginning or end of the time series might be seen as trends.

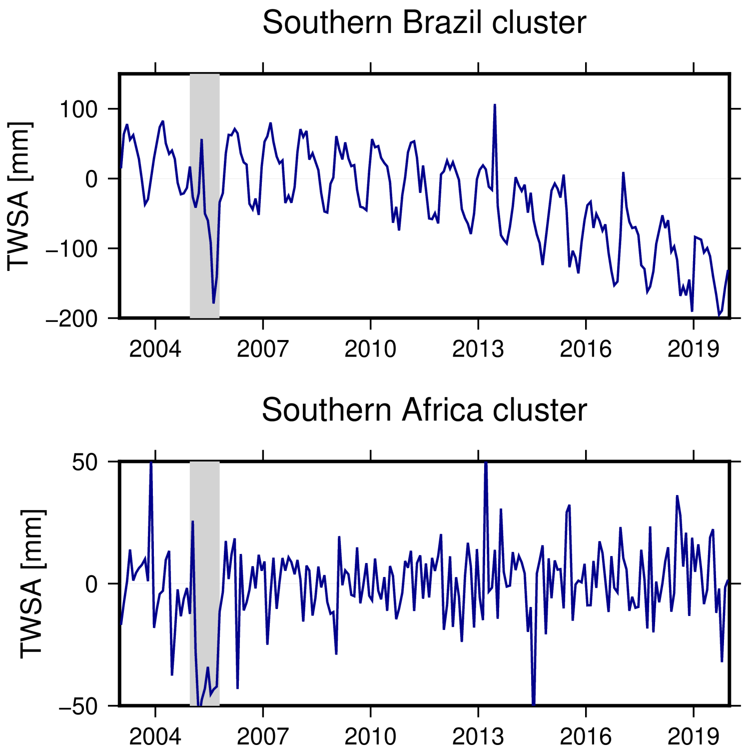


Figure 7.3: Synthetic TWSA time series spatially averaged for the Southern Brazil (top) and the Southern Africa cluster (bottom). The TWSA are given in mm. The grey shade marks the period of simulated drought.

In contrast to the steady decrease or increase in water storage that affects the indices, periodic events like an annual signal do not propagate to the drought indices: no periodic events are visible in the index time series. This is expected, since each index uses a method to remove seasonality via, for example, computing climatology.

Summing up, linear trends and accelerations propagate through the indices, seasonality is removed per index methodology and a trend removal can be tested but still carefully analyzing droughts at the beginning and end of a time series should be undertaken.

**Effect of accumulation and difference periods in index definition on geophysical signatures and temporal noise**

In Sec. 3.2, I presented the possibility of accumulation and differencing of TWSA in advance of index computation. The accumulation or differencing is applied for each of the four drought indices considering time scales of two to 12 months (Fig. 7.4). The upper part of the four figures always refers to the indices based on accumulation. The longer the accumulation period, the more seem the drought indices to be temporally smoothed, which applies to each index but in general, the DI is more noisy than DSI, TSDI, and GGDI. The smoothing has the effect that the drought severity for the 2005 drought is more dampened and temporally shifts in time as compared to the original indices but it has also the positive effect that the temporal GRACE/-FO noise is minimized. Thus, the accumulation leads to a more robust drought detection if the time scales of the accumulation are short, for long time scales the smoothing is too strong to identify intense severity and correct timing.

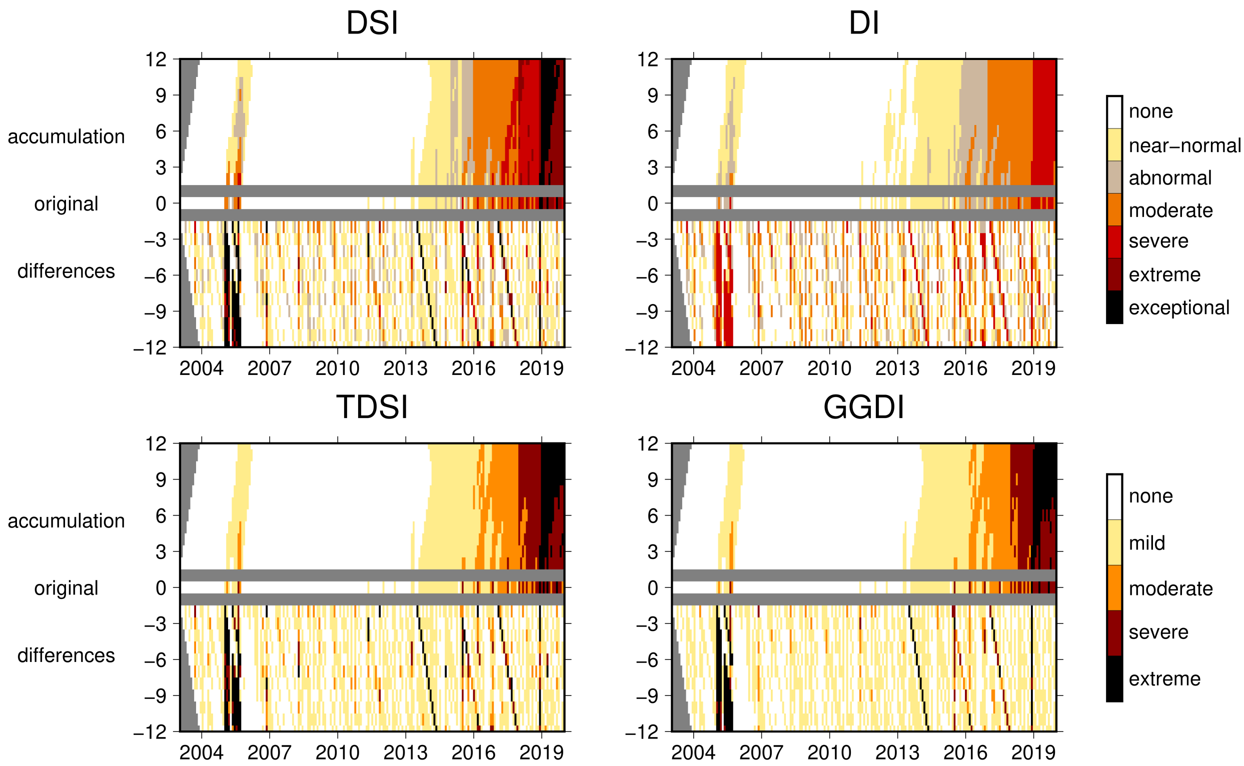


Figure 7.4: Synthetic TWSA-DSI, -DI, -TSDI, and -GGDI and their variants with accumulation or differences computed for spatially averaged TWSA for the Southern Brazil cluster. The y-axis describes which variant of the respective index is used: Zero is the original index, positive values describe the indices based on accumulation period from two to 12 months (e.g., DSIA12) and negative values describe the indices based on a differencing period (e.g., DSID12). The index severity classes are unitless.

In contrast, the use of differencing periods – these are shown in the lower part of each index figure – encounters different problems compared to the use of accumulation periods. The indices that are based on differences of TWSA show an increase in temporal noise as compared to the original indices and the index with accumulation. Thus, the indices that use differencing are less robust against the temporal noise compared to the original indices and indices that use accumulation. The use of differences leads the indices to identify shorter time period of intense drought. The short intense drought event is always followed by a period where no drought is identified. In fact, this non-drought period is very wet (not shown). The interpretation can be followed by referencing to Eq. 3.3 that relates precipitation to water storage differences. Under the given assumptions, deriving differences of water storage is related to as accumulated precipitation. When a precipitation deficit leads to a hydrological drought event, a strong wet period is required to recover the water storage from its deficit. For long time scales of the differencing, the length of drought duration increases and better reflects the real length of the simulated drought period.

No coherent drought period is identified for the period after 2013 as is the case for the original indices and the indices that use accumulation, which means that the trend and acceleration do not propagate to the indices that use differences. Although the differences lead to an increased sensitivity against temporal noise, it is not significantly affected by linear trend or acceleration biases.

**Effect of spatially correlated GRACE/-FO errors**

The GRACE/-FO specific spatial noise patterns applied to the synthetic TWSA time series are not observed by other satellite missions or observations. In contrast, geophysical signals like hydrology can be observed by other observations. To study the effect of spatially correlated noise, indices are spatially averaged within the two clusters via two methods:

1. Compute the spatial average of gridded TWSA and computing the index for that one time series.

2. Compute drought indices for each of the TWSA grids within the cluster and average the final gridded indices to one time series.

Due to the fact that the indices are non-linear, the two resulting averages from the two methods are different. The first method is denoted as "**averaged TWSA**" and the second method is denoted as "**averaged index**" in the following.

Fig. 7.5 shows the results from the two averaging methods for the DSI and for the Southern Brazilian cluster (top) as well as the Southern Africa cluster (bottom). For the Brazilian cluster, the DSI from averaged TWSA differs only very slightly from the average DSI. The findings can be explained from the synthetic TWSA from that cluster (Fig. 7.3). The TWSA time series shows a strong seasonality and trend, whereas the noise in the time series is smaller in magnitude. This means that the signal-to-noise ratio is high and the noise is not most prominent in the time series. In addition, the ratio of drought deficit to noise is also high, as the drought was defined to have a deficit of 100 mm. In consequence, the noise is also not propagating dominantly to the indices, independent of which the averaging method is chosen. Considering the Southern Africa cluster, the signal-to-noise ratio is much lower, the drought deficit is smaller at 50 mm, and the spatial and temporal correlations are very prominent. The strong noise is propagating to the indices within this cluster leading to noticeable differences for the two averaging methods. The DSI from averaged TWSA has generally elevated index values per month (positive as well as negative). During the synthetic drought period in 2005, the difference between the averaging method leads to differences in the identified severity of the drought. Whereas the DSI of averaged

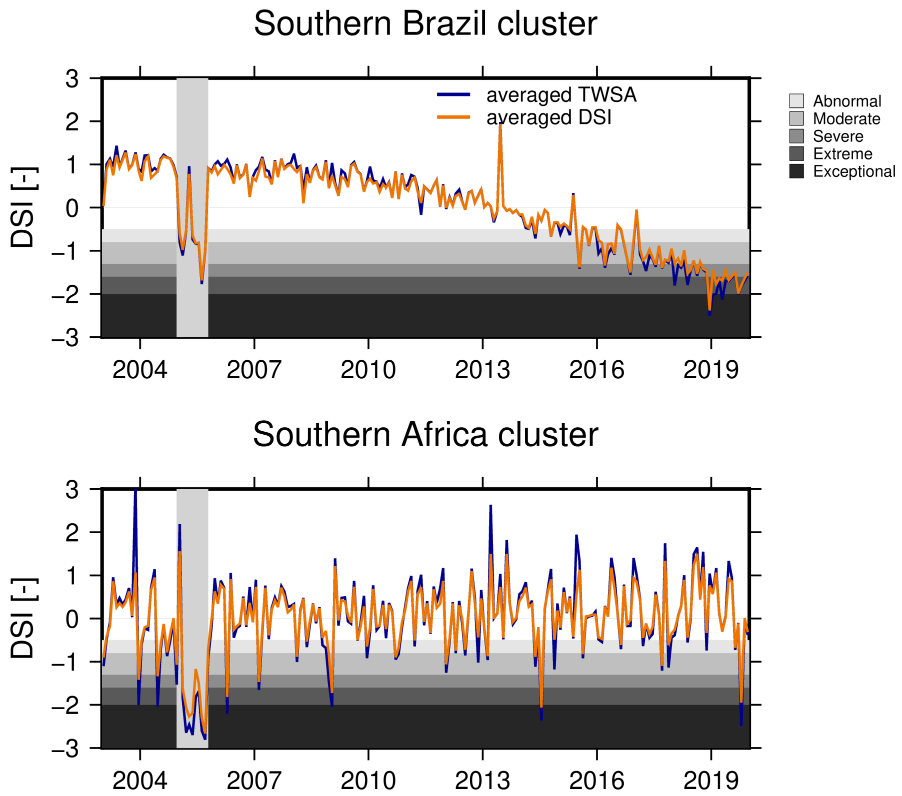


Figure 7.5: Synthetic TWSA-DSI for Southern Brazil and the Southern Africa clusters computed from two different averaging methods: either from (blue) spatially averaged TWSA or from (orange) spatially averaging the gridded DSI. The index values are unitless. The vertical grey shade marks the period of simulated drought, whereas the horizontal grey shades mark the severity classes from abnormal (lightest grey shade) to exceptional drought (darkest grey shade).

TWSA identifies the drought months as extreme and exceptional, the averaged DSI identifies the drought as moderate to exceptional.

To conclude, as soon as the TWSA time series have a low signal-to-noise ratio, the different averaging methods return different severity classifications of the same drought. If the noise is not very prominent in the data, the choice of averaging method is insignificant. To avoid an underestimation of the drought for regions where the signal is affected disproportionally by noise, it is recommended to use averaging method one, which is computing the index from averaged TWSA. The same conclusions are found for the DI, TSDI, and GGDI (results not shown).

#### Length of time series

According to recommendations, drought indices should cover about 30 years of data to produce a realistic distribution of the data (Hoylman *et al.*, 2022). This is especially difficult for remotely sensed data (Yihdego *et al.*, 2019) because some missions, among others also GRACE/-FO, do not have a long record of data or have observational gaps (Sec. 2). Therefore, data assimilation can also help to prolong the available time series by filling gaps and extending time frames. In this work, the length of the time series from assimilation is 17 years and the simulated TWSA is set up with the same length.

From the previous analyses, all indices had at least one month of the whole time period within one of the severity classes that are defined per index, except for the DI or the DI with accumulated TWSA (DIA). The methodology of DI is using empirical CDFs to compute percentile ranks. With a time series of 17 years, the lowest percentile rank has an occurrence probability of 5.9% showing that the length is not suitable to detect exceptional drought. To monitor exceptional drought with DI, and thus derive an occurrence probability of equal or lower than 2%, at least 50 years of observations are required to build an empirical CDF that includes all severity classes. This fact that occurrence probabilities of lower than 2% can only be determined with 50 years underlines that the DI output is not fully reliable for short time series. Since the DSI, TSDI, and GGDI work with standardization, all severity classes are included in the monitoring. As mentioned earlier, a bias correction could be used to develop an empirical DI for longer time scales. Then, the indices are adapted towards the underlying distribution of the respective data set used for bias correction, which requires that one trusts the data set.

### 7.1.3 Index choice

In Gerdener *et al.* (2020a) and as also discussed in the previous section, it was found that all indices have their individual advantages and disadvantages. There exists no clear "best" index for all applications but the choice of the final index strongly depends on the current application and one should be aware of how input signals contained in the data affect drought detection (Sec. 7.1.2) to better interpret the results.

A noisy index – as is the case for all indices that use TWSA differences – hides drought events for the risk map computation because a drought might be interrupted through the noise more easily. Such indices might be a good choice when comparing TWSA indices directly to precipitation indices. The indices that use accumulation are "smoothed" indices and are better suitable for monitoring consecutive drought periods but with a slight delay the intensity can be decreased. The original indices without accumulation or differences in the computation did not change the timing of the drought and were able to detect drought events with high severity although more temporal noise is propagating from the TWSA to the indices. For a drought warning system that tracks subsequent drought events in precipitation through storages and affecting vegetation, the timing of the drought should be determined correctly as it is one of the most important features required for the system. I thus suggest to stick to the original index versions without accumulation or differences. Nonetheless, one should not forget that the original indices of DSI, DI, TSDI, and GGDI are not able to suppress linear trend and acceleration signatures that bias the drought detection. As the incorrect removal of such trends and accelerations also could inhibit the correct monitoring of a drought, no clear recommendation can be given. The analysis of droughts at the beginning or end of a time series should then be evaluated carefully.

Most of the findings are transferable to other indices that use standardization methods and or aggregation/differences and are also transferable to other storages, for example, groundwater or soil moisture. Linear trends and accelerations would show biases in a similar way, the accumulation and differencing would smooth or increase noise and change the severity and duration of drought and a prolonged time series should better capture the seasonality signatures and underlying distributions. The only two aspects that could change are the temporal and the spatial noise. The temporal and spatial noise might be less or stronger as compared to the TWSA but would not significantly change the performance of the index and it would not change the choice of the averaging method.

It remains the question of which of the four drought indices should be used. As the length of the time series of the current study setup is significantly biasing the DI, a choice should consider only the three other indices, although prolonging the time series in the future should be aimed for as more data improve the precision of the underlying distributions, which will be feasible with the new gravity missions. Only very minor differences between DSI, TSDI, and GGDI were observed. The main difference was how the temporal and spatial noise propagates through the indices but this did not lead to significantly different results for monitoring of the 2005 synthetic drought. Consequently, all three indices are suitable for application in this thesis, but I concentrate on the DSI because the standardization per calendar month is most straight forward in my opinion as compared to the final standardization using all months of the time series as is the case for TSDI and GGDI.

## 7.2 Monthly monitoring of hydrological drought

In this section, drought is analyzed as retrospectively monitored with the DSI (Sec. 3.4.3) for monthly storages mainly from GLWS assimilation but also in comparison with GRACE/-FO observations, and WaterGAP model simulations. First, I show how drought manifested in TWSA globally and analyze drought with and without the use of a trend removal in the TWSA before computation of the drought indices. Then, the monthly drought monitoring is investigated on the regional domain for the Congo basin and the Ganges basin and extended with the vertically disaggregated outputs of GLWS to distinguish between drought in the soil moisture, surface water, and groundwater. Finally, an opportunity is shown how to derive uncertainty estimates for the drought indices from the assimilation.

### 7.2.1 Global drought monitoring

To provide a broad global picture of drought monitoring in the retrospective with TWSA-DSI, the year 2010 is exemplarily chosen to showcase where on the global land mass extreme or exceptional drought has been identified. The TWSA is derived from assimilation and for each of the the monthly fields, a global index map can be derived. Fig. 7.6 shows the number of months that experienced at least extreme drought in 2010 per grid cell of the TWSA-DSI from assimilation. The year 2010 is chosen as it is a year in the middle of the time period to minimize a possible bias introduced by trends or its incorrect removal on drought detection. Several regions are found to experience extreme or exceptional drought for multiple months in 2010. For example, in Southern Asia, about five months are detected to have been at least extremely dry, or in the La Plata basin, up to six months are detected. Further regions that show striking drought patterns are the Orinoco basin, the Horn of Africa, the regions surrounding Lake Chad, the Lena basin, and the Yenisei basin (Russia). The EM-DAT database confirms these regions with recorded drought months in 2010 for some countries as follows:

- **Horn of Africa:** Somalia, Ethiopia, Djibuti, South Sudan
- **Lake Chad and surrounding regions:** Chad, Niger
- **Northern of South America:** Guyana, Barbados, Grenada, Saint Lucia, Trinidad and Tobago, Saint Vicent and the Grenadines, Brazil (Amazonas province)
- **Volta basin:** Russia (Multiple provinces in Eastern Europe)
- **Southern Asia:** China (eastern provinces), Thailand

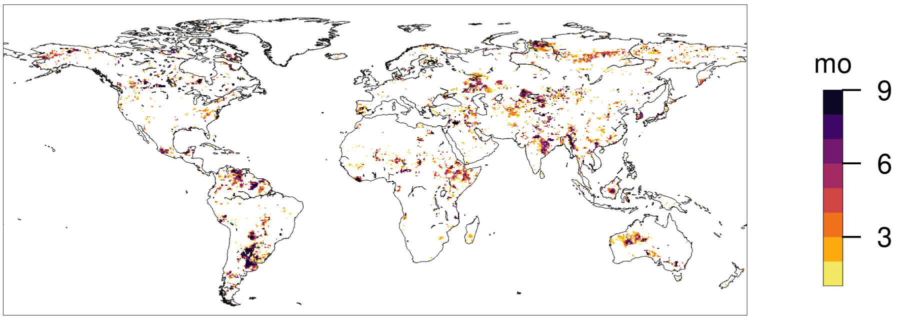


Figure 7.6: Number of months with extreme or exceptional drought in 2010 using TWSA-DSI from GLWS assimilation.

The Yenisei basin and the La Plata basin are not confirmed by the EM-DAT database. Instead, a drought in the La Plata basin is recorded in 2009 (Argentina, Paraguay) due to strong El Niño. As El Niño leads to strong rainfall deficits, it can explain the temporal delay towards drought in TWSA a few months later. No corresponding relation is found for the Yenisei basin. As a hypothesis, the Yenisei basin is not recorded in the EM-DAT database as it is mainly covered by tundra with cold temperatures, and although water storage is indicating a strong deficit no human or economic losses occur, which are the base criteria for recording events for the EM-DAT database. Further events in 2010 recorded by the database cannot be completely confirmed with GLWS drought monitoring, which might be due to the fact that only very small areas are showing droughts in the assimilation or it might be related to a temporal shift between the record in the EM-DAT database and GLWS monitoring: Mali, Mozambique, Zimbabwe, Burundi, Mauritania, and Bolivia.

As shown for the synthetic drought study, the influence of linear trends or a corresponding trend removal on the drought detection with DSI is largest for the droughts in the early years of the period of GLWS or for the droughts in the later years. To discuss the differences that might occur and provide a workaround with trend biases during drought monitoring, Fig. 7.7 shows, similarly to the previous figure, the number of months that are at least extremely dry monitored with DSI based on TWSA from assimilation but for the years 2003 (top left) and 2019 (bottom left). The right side shows the same but this time TWSA were detrended per grid before computing the DSI. For 2003, few regions show drought independent of a detrending or not, which are, among others, the Orinoco basin, Senegal, and Finland. For 2019, droughts are identified with both methodologies for Mexico, Central and Eastern Europe, and Thailand, but with varying spatial extent and duration. The direct comparison of the DSI with and without trend removal helps to identify drought events independent of trend biases.

Strong differences in the results with and without trend removal are found in 2003 for, for example, Eastern Brazil, West Africa, and Europe. Without further analysis and evaluation, no conclusions can be drawn. One should compare the detected drought events with independent drought databases and look at possible trends in the TWSA time series to uncover the direction of the trend and its overlap with the drought. For example, for East Brazil Fig. 6.13 showed strong negative trends and the EM-DAT database does not record a drought event in 2019 in Brazil. In this case, the DSI using detrended TWSA is probably better suitable to monitor droughts in Brazil as compared to the version with trends included. However, the removal should always be explicitly tested for each region and no overall advice is given for or against a trend removal. The good news is that the longer the time series, the more stable a trend is estimated and the drought can be separated from the trend. So in the future, the complexity through trend biases should decrease.

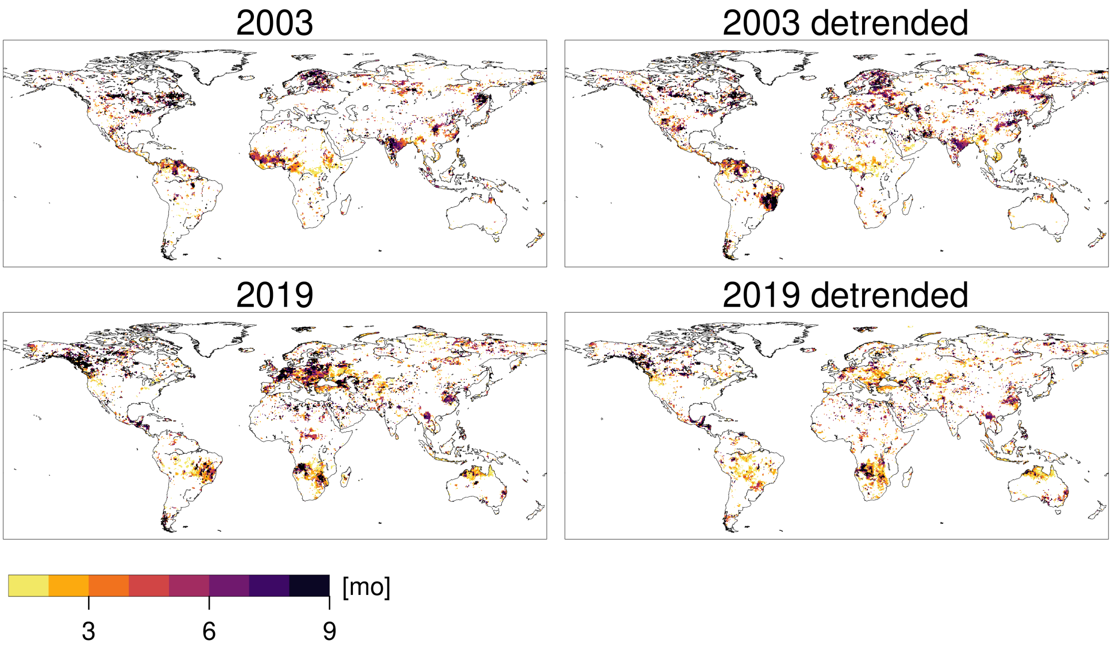


Figure 7.7: Number of months with extreme or exceptional drought in 2003 or 2019 using TWSA-DSI from assimilation without (left) or with (right) trend removal.

### 7.2.2 Regional drought monitoring

**Drought in total water storage**

For a detailed analysis of how drought monitoring with TWSA from the assimilation performs compared to observations and model simulation, the TWSA-DSI is now analyzed for the Congo basin and the Ganges basin. These basins are chosen as they are study regions for the Global-CDA[4] project and cover two different hydrological regimes but the in-detail analysis could be applied to any other basin.

Specifically, Congo is the largest African basin located in Central Africa. The Congo river originates in the southeast of the basin in the highlands and routes to the Atlantic Ocean in the West of the Congo basin. The basin experiences an annual rainfall of about 1500 to 2000 mm (e.g., CSC, 2013) with two rainfall seasons. The first rainfall season is from March to May and the second season is from September to November. Most of the land area within the basin is covered by rainforest or savanna. The largest lake within the basin is Lake Tanganyika in the East and the largest dam is the Grand Inga Dam in the West, which is close to the capital Kinshasa. The second basin, the Ganges basin, is a major river basin on the Asian continent routing through India and Bangladesh. The Ganges river originates in the western Himalayas and routes eastwards and southwards to the Bay of Bengal. Due to the summer monsoon, most of the mean annual rain of the basin of about 1094 mm (e.g., Khan *et al.*, 2015) falls from July to October. The monsoon leads to regular flooding events almost every year. There is also a winter monsoon but it is weaker than the summer monsoon. The vegetation cover of the land area within the basin is diverse and ranges from croplands for most parts of India, tropical forests, to tundra for the Himalayas in the north. One of the major dams is the Theri dam which is located in the North West of the basin. The outlines of the Congo and the Ganges basins can be extracted from Fig. 6.1 (left).

[4] http://globalcda.de/ (last accessed 16.05.2024)

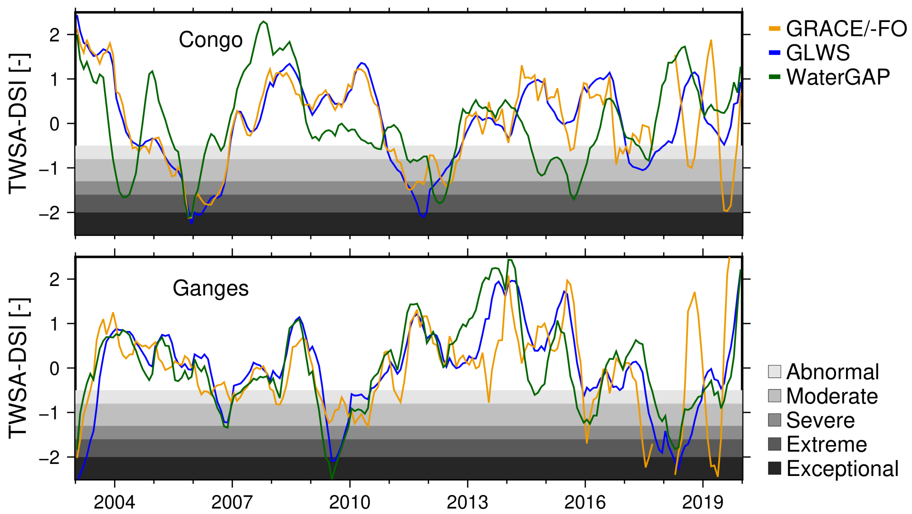


Figure 7.8: DSI computed by using spatially averaged TWSA for the Congo basin and Ganges basin. Green represents the TWSA-DSI from the WaterGAP model simulation, yellow from the GRACE/-FO observation, and blue from the GLWS assimilation. The index values are unitless. The horizontal grey shades mark the severity classes from abnormal (lightest grey shade) to exceptional drought (darkest grey shade).

Fig. 7.8 shows the TWSA-DSI for the GRACE/-FO observations (orange), the WaterGAP model simulations (green), and the GLWS assimilation (blue) computed from basin-averaged TWSA for the two basins. The TWSA is detrended before index calculation, as strong linear trends are found for the two basins (wetting trend for the Congo basin and drying trend for the Ganges basin). For the Congo basin, the TWSA-DSI from assimilation detects drought at the end of 2005 and in 2006 as mostly extreme or in some months as exceptional, whereas at the end of 2011 and the beginning of 2012 mainly severe or extreme drought is monitored. The TWSA-DSI from model simulations indicates only a few months of drought during both periods and in addition drought at the beginning of 2004 and in 2015. Especially at the beginning of the full period, the DSI from GLWS is closer to GRACE/-FO as compared to WaterGAP and towards the end of the period, e.g. in 2019, the reverse is found, which is not surprising as there is a gap between the two satellite missions in 2017/2018.

For the Ganges basin, the TWSA-DSI from assimilation shows extreme and severe drought events in 2009 and 2018, respectively. The DSI from assimilation is much closer to the model simulations as compared to the observations and as compared to the Congo basin. A possible explanation is that the basin has a complex topography with mountains in the north and the coast in the south. As described in Sec. 2.3.2, the GRACE/-FO mission is prone to leakage error, which is very dominant in coastal regions. Consequently, during assimilation, the update might be closer towards the model than to GRACE/-FO. Nonetheless, the DSI that uses TWSA from assimilation is closer to the observations as compared to the model for a few months, for example, at the beginning of 2013. Hence, the smooth transition of GLWS between GRACE/-FO and WaterGAP that was found when analyzing the properties in Sec. 6.3 is also valid for the application for drought monitoring.

The two drought periods in the Congo basin in 2005/2006 and 2012 discovered with the TWSA-DSI from GLWS assimilation are confirmed by literature, for example, Elameen *et al.* (2023) identified droughts longer than one year in 2005/2006 and 2011/2012 in Congo basin and the EM-DAT database records drought in 2005 and 2012 for Angola and Cameroon, and drought in 2005 in Zambia, all countries that are part of the Congo basin. In agreement with the findings from assimilation and GRACE/-FO but in disagreement with the model simulation, the database does not record a drought event for the beginning of 2004. The two droughts identified in the Ganges basin are also coinciding with literature. The EM-DAT database records a drought in Bangladesh and India in 2009 and a drought in India in 2018. The 2009 drought can be confirmed by Dharpure *et al.* (2022) and was related to a deficit in the summer monsoon.

The spatial extent of the 2005 drought in the Congo basin is shown with monthly TWSA-DSI maps (detrended TWSA) derived from GLWS assimilation and shown in Fig. 7.9. At the beginning of 2005, only a few cells experience moderate or severe drought. A striking pattern is found for the southeast of the Congo basin emerging towards the centre of the Congo basin. During nearly all months in 2005, this region experienced drought or dryness, but the patterns do not overlap with coherent vegetation regimes, river patterns, etc., and thus it might result from specific precipitation characteristics. Towards the end of 2005, the number of cells experiencing drought increases significantly, and also the severity of drought intensifies. In December 2005, more than half of the cells are occupied with moderate or more severe drought. Most of the south of the Congo basin experiences severe, extreme, or exceptional drought. For example, exceptional drought stretches from the south towards the inner of the basin along the tributary Sankuru, where also a large city is located: Mbuji-Mayi with a population of two million. Coherently, the south is underlying mixed forest vegetation (compare Sorí *et al.* (2017)), which obviously is more prone to drought for this year as the dominant vegetation regime of the northern part, the evergreen broadleaf forest.

**Vertical disaggregation of drought**

The assimilation of GRACE/-FO TWSA into the global WaterGAP model enables not only spatial downscaling but also vertical disaggregation of the TWSA into the model compartments. Hence, the temporal developments of a drought through the water compartments can be understood, i.e. temporally subsequent droughts in different water storage compartments. Fig. 7.10 shows on the left-hand side the percentage area of the Congo basin that is experiencing one of the drought severity classes either for the DSI based on soil moisture (top), surface water (middle), or groundwater (bottom), all derived from assimilation. On the right-hand side, the DSI of the corresponding storage is shown as a spatial map for December 2005 for analyzing the spatial patterns. The storage information was detrended as in the previous section.

The soil moisture DSI shows a very noisy temporal pattern. The 2005/2006 and 2011/2012 droughts that were identified with the TWSA-DSI and the EM-DAT database cannot be clearly detected. Instead, the DSI from soil moisture monitors show most cells affected by drought in 2003. The DSI from surface water and groundwater shows a drought event in 2005/2006. Especially for groundwater, the area covered by severe, extreme, or exceptional drought is about 50%. For the surface water DSI, a maximum of 40% of the area is affected by drought. It seems as if the droughts in 2005/2006 and 2011/2012 had no serious impacts on soil moisture but on surface water and groundwater. Possibly, the seasonal soil water content is relatively stable from year to year thus strong temporal noise is prevalent in the drought indices, and extreme events do not lead to a massive decrease in soil moisture. Fig. 6.14 supports this finding, no strong trends are prevalent in the Congo basin for soil moisture from GLWS, the same is found for the ESA-CCI soil moisture (Fig. A.13).

The spatial extent of the drought in December 2005 in the water storages is shown on the right side of Fig. 7.10. As expected, the drought is mainly dominant in the surface water and groundwater storage and, in accordance with the TWSA-DSI, strong drought patterns are found in the south along the Sukuru tributary with strong drought severity. The difference between the results for

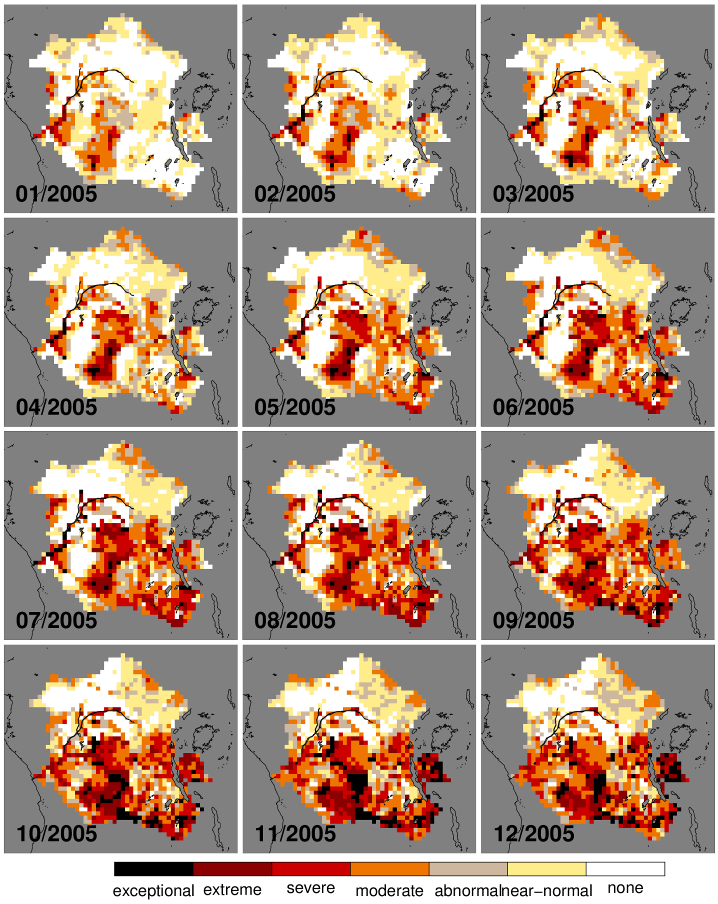


Figure 7.9: Spatial maps of the TWSA-DSI severity classes for the Congo basin computed from GLWS assimilation for each month in 2005.

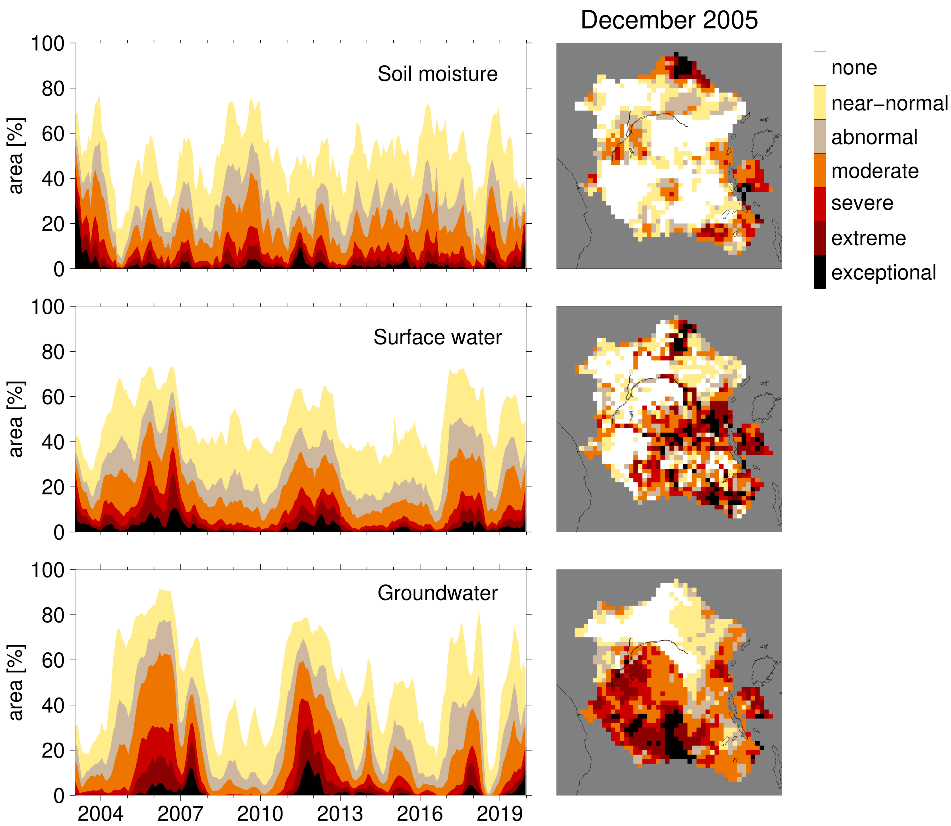


Figure 7.10: Percentage area of DSI severity classes for each month from 2003 to 2019 (left) and spatial maps for December 2005 (right) for using soil moisture (top), surface water (middle), and groundwater (bottom) from GLWS assimilation for the Congo basin.

surface and groundwater is that the surface water drought maps exhibit much more spatial noise and obviously more dominant along the rivers as compared to the groundwater drought. Although the soil moisture storage does not show a striking drought pattern in the temporal domain for the end of 2005, the spatial domain reveals coherent regions of moderate to extreme drought in the southeast, nonetheless with a much smaller spatial extent than for surface water and groundwater. All in all, this case study highlights the opportunities of drought monitoring with assimilation output as it enables new insights into the spatial extent of drought and temporally subsequent drought events in different water compartments, which could not be made with GRACE/-FO TWSA only.

#### Uncertainty estimation

As described in Chap. 5, the assimilation uses a filter algorithm based on an ensemble to represent model and observation uncertainty and this enables deriving an ensemble of state updates. Until now, all results for the drought monitoring with assimilation outputs are derived by computing the ensemble median from the derived $n = 32$ ensemble members. Here, I use the ensemble members instead of the ensemble median to compute a drought index for each member. With the help of the ensemble of indices, an uncertainty quantification of the specific drought index in a specific month can be provided. Fig. 7.11 gives an example of uncertainty quantification for drought monitored with the TWSA-DSI from assimilation. It shows the percentage of ensemble members that experienced at least extreme drought in the Congo basin in December 2005. The uncertainty quantification could also be provided for the soil moisture, surface water, and groundwater compartments or for other months and regions but here is limited to the current setup to complement the regional study.

A high percentage of the ensemble members show drought in TWSA for the south (Fig. 7.11. Two coherent regions are shown. The first region is prevalent from the south towards the interior of the Congo basin. The same region was discussed in the previous section and it is located along the Sankuru tributary. The second region that shows strikingly high percentages around the Lake Tanganyika, the second largest lake in Africa and part of the African Great Lakes. This region is also showing up in previous results (Sec. 7.2.2) but was not found to show similar intensity as the Sukuru tributary. In contrast, the area between the two regions shows much lower percentages of DSI ensemble members that point towards extreme or higher drought (20% or lower). In consequence, for both regions, the occurrence of drought in December 2005 can be estimated with low uncertainty as the agreement of ensemble members is overall high.

## 7.3 Drought monitoring system

In the following, I show how outputs from GRACE/-FO assimilation can be used to distinguish between economic sectors affected by drought and could be used for designing monitoring systems with two different objectives in mind: (1) Determine drought hazard risk at the current state and (2) set up a framework for a prototype of a warning alert system which could be used for near-real-time in future. In preparation for the hazard risk computation and warning alerts system, consecutive periods of droughts are extracted from the monthly indices presented previously (Sec. 7.2.1). Then, the details of the system setup are presented first and the results for drought hazard risk and the warning alerts are presented at the end of this section.

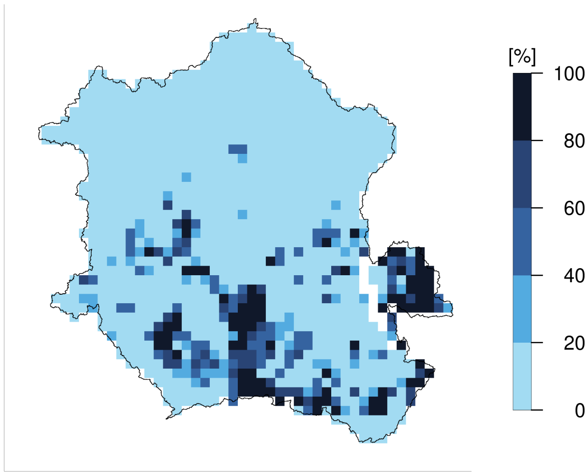


Figure 7.11: Percentage of ensemble members from GLWS assimilation per cell where the TWSA-DSI shows extreme or exceptional drought in the Congo basin in December 2005.

### 7.3.1 Prototype warning system and risk assessment setup

This section describes how the prototype drought hazard risk assessment and warning alert system, which will make use of GRACE/-FO assimilated into WaterGAP, are exactly set up. Under optimal conditions, a warning alert system (or monitoring system) for drought would consist of a near-real-time application that is regularly updated (Sec. 3.4). This means that a setup of a drought monitoring system should be based on, among other aspects, near-real data providers, automatic data downloads and workflows, sufficient storage capacities, security measures, and quality checks, etc. As not all of these aspects can be guaranteed for this thesis, I aim to set up the drought monitoring system by using the retrospective. Nonetheless, the retrospective system will be denoted as a monitoring system to simplify the terminology and also underline its potential use in the future for the near-real-time. In addition, drought risk hazard assessments typically do use all available data at the time of computation. This means, risk assessments provide a current state estimate of drought risk using the retrospective (e.g., Carrão *et al.*, 2016; Meza *et al.*, 2020) and can also use near-real-time data if available.

A drought monitoring system and drought hazard risk assessment should refer to a specific sector that is affected (Meza *et al.*, 2020; Herbert & Döll, 2023). Common sectors are the sectors for health, energy, economy, waterborne transport, forestry, agriculture, and water supply. Limitations in public water supply and agriculture are those sectors that are most directly connected to a shortage of water storage leading to a shortage in food and drinking water. Following Meza *et al.* (2020), the agricultural sector can further be separated to distinguish between irrigated and rainfed agriculture. I here exemplarily show the use of assimilation outputs for drought hazard risk assessment and suggested warning alert systems for the three sectors of rainfed agriculture, irrigated agriculture, and water supply but it should be noted that the use of assimilation products might be considered for other system setups as well.

The next question is which variables to choose to monitor drought or apply drought hazard risk assessments for the specific sectors appropriately. Rainfed agriculture refers to the direct availability of water for crops. As plants derive water via their roots from the soil, soil moisture is the preferred hydrological variable in this study for the drought hazard risk assessment for rainfed agriculture. For setting up a prototype drought monitoring system for rainfed agriculture, it should be noted the impact of drought is apparent in reduced crops, which can be seen as a part of vegetation. First, the precipitation falls on the soil and then leads to the growth of vegetation for most biomes (Sec. 6.3.2). This setup perfectly matches building a warning alert system for rainfed agricultural drought. Building a monitoring system and computing drought risk based on soil moisture is commonly done, for example, with the European Drought Observatory (Sec. 3.6), but including products of assimilation is not widely integrated, especially on a global scale.

For irrigated agriculture, the plants receive water anthropogenically through irrigation. Water for irrigation is either received from groundwater wells or surface water bodies such as rivers, lakes, etc. Similarly, the water for water supply is also extracted from surface water bodies and groundwater wells or springs. For this reason, the assessment of drought risk for irrigated agricultural sectors and water supply is determined both from surface water and groundwater in this thesis. There exist regional groundwater risk assessments in the literature, for example, for the Pingtung Plain in Taiwan (Lin *et al.*, 2021) by using groundwater stations or for the Kansai river basin (Bera *et al.*, 2024) using machine learning products. However, to my knowledge, including groundwater globally – especially from assimilation – in the computation of drought hazard risk has not been applied so far and is important as global groundwater additionally presents a major source for solving water requirements and thus presents a very new approach.

A monitoring system or drought hazard risk assessment for irrigated agriculture would require a different setup as compared to rainfed agriculture. Sec. 6.3.2 revealed that for many vegetation regimes, the vegetation responds mainly to soil moisture instead of surface water and groundwater, which was hypothesized to result from soil moisture being for most areas the main moisture source for vegetation or anthropogenic effects like irrigation in response to dry vegetation conditions. In addition, the soil moisture storage in the model (and thus in GLWS) refers to the root zone with up to four meters depth (Sec. 4.4) instead of a few centimeters of surface soil moisture only. Thus, with the observed results, the vegetation would rather be a temporal proxy for drought in surface or groundwater as compared to surface water and groundwater being a proxy for vegetation. Due to these reasons, a monitoring system that can be used for irrigated agriculture builds upon precipitation, soil moisture, and the combination of surface- and groundwater.

At the same time, a drought monitoring system for water supply would probably refer to the measured impacts where the water is supplied, e.g., for drinking water or water for the industry. This means, for example, statistics about drinking water shortages affecting the population or industrial damage would be required. In addition, one would also need to monitor only the amount of surface- and groundwater used for drinking water or the industry. At this point, the reader should be reminded that the thesis is focusing on drought as measured by GRACE and GLWS, and therefore no anthropogenic impact variable, for example, the number of people that are exposed to drought hazard, is included for the prototype monitoring system for water supply. Since the irrigated agriculture and the water supply systems both require surface water and groundwater for their monitoring, the precipitation – soil moisture – surface-/groundwater monitoring system can be used for both sectors.

As an interim summary, two drought hazard risk assessments will be applied. The first one will focus on soil moisture risk and the second one will focus on surface-/groundwater risk. In addition, two prototype warning alert systems will be developed. The first one is using precipitation, soil moisture, and vegetation, and the second suggested system will use precipitation, soil moisture, and surface-/groundwater. Since much more information (e.g., statistics about exposure and vulnerability per sector) is required for setting up a complete risk assessment or warning alert system for the mentioned sectors of rainfed agriculture, irrigated agriculture, and water supply, I will refer to the assessments and prototype systems as soil moisture risk assessment, surface-/groundwater risk assessment, soil moisture warning alert system, and surface-/groundwater

Table 7.1: Considered thresholds, minimum duration, and maximum pause for extracting consecutive drought events from drought indices. The values are depending on the specific application.

| **Condition** | **Risk determination** | **Warning system** |
|---|---|---|
| Minimum threshold | -1.3 | -1.0 |
| Minimum duration | 9 | 2 |
| Maximum pause | 2 | 2 |

warning alert system.

For the drought hazard risk determination and for the warning alert system, specific requirements need to be formulated in order to allow to extract consecutive periods of drought from the monthly drought indices. The start of a drought should be determined with a minimum duration of consecutive months with a certain drought severity. Typically, for meteorological drought indices as the SPI the threshold of -1 is chosen, which includes the severity classes of mild, severe, and extreme drought for that specific index (McKee *et al.*, 1993) and a minimum duration of two months is chosen (e.g., Meza *et al.*, 2020). The indices can be affected by noise in the observations; one positive month might interrupt the drought, although it continues in subsequent months. To avoid a too-fast interruption of a drought (e.g., from temporal noise), a condition is introduced that stipulates that the end of a drought is determined when two consecutive months above the threshold exist. All these options that define consecutive drought events can be adapted depending on the application.

As the extraction of consecutive drought events changes depending on whether the warning alert system or the risk assessment is used, the options chosen for this thesis are summarized in Tab. 7.1. The risk analysis concentrates on long-term droughts, which is why a strong drought severity of at least extreme or exceptional drought is chosen as the minimum threshold for the water storages. This means the long-term drought events are extracted from soil moisture indices or from indices that include the combination of surface water and groundwater. A minimum duration of nine months is chosen to isolate the long-term events. The drought ends when the corresponding index value is two months above the threshold. The two final drought hazard risk assessments – one for soil moisture and one for surface-/groundwater) yields from the determined consecutive drought events collected per grid. Following Meza *et al.* (2020), the mean normalized severity is multiplied by the frequency of drought events to determine drought hazard risk per assessment:

$$R_{hazard} = \bar{S} \cdot f \tag{7.2}$$

So, the severity of each consecutive drought is estimated by averaging the accumulation of the monthly index magnitudes for each drought event. To provide a risk map comparable across grid cells, the final risk values are standardized to the range zero to one, where zero means the lowest risk and one is the highest risk.

The prototype warning alert systems intend to detect a drought as early as possible, thus, a weaker drought intensity is chosen with a minimum threshold of -1 and a minimum drought duration of two months to extract consecutive drought events. Again, the drought ends when the corresponding index value is two months above the threshold. The soil moisture and surface-/groundwater systems additionally require drought information from precipitation and vegetation, realized with the SPI and normalized LAI (as with DSI for storages). For both variables, an index threshold of -1, a minimum duration of two months, and an interruption of two months are specified to define the start and end of a drought following McKee *et al.* (1993) and Meza *et al.* (2020), respectively. Fig. 7.12 illustrates the simplified version of warning levels for the rainfed agricultural system

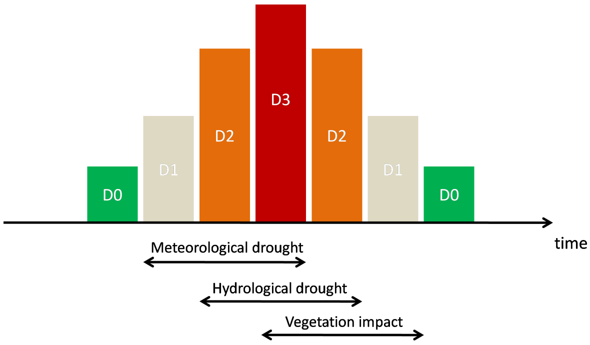


Figure 7.12: Illustration of warning alert levels for a precipitation-soil moisture-vegetation drought warning system.

and how different warning levels could be selected to relate to meteorological and hydrological drought as well as the impact on vegetation. Level D0 indicates no drought and no current warning alert. As soon as the meteorological drought index detects a consecutive drought event, level D1 is reached, level D2 starts with the beginning of a hydrological drought, and level D3 when the impact is noticeable in the vegetation. Compared to the drought monitors from the European Commission Joint Research Centre (EDO and GDO, Sec. 3.6), D1 can be interpreted as a "watch" phase, D2 as a "warning" phase and D3 as an "alert" phase (e.g., Sepulcre-Canto *et al.*, 2012). After the highest alert phase D3, the drought warning levels are stepwise reduced again to return to normal with no warning given. In a similar way, the second system built upon precipitation, soil moisture, and the combination of surface water and groundwater is set up.

### 7.3.2 Consecutive periods of drought and drought hazard risk

In this section, consecutive drought events are extracted from water storage variables as derived from assimilation and used to compute drought hazard risk maps. The detection of consecutive drought events is here shown for TWSA first instead of showing results for soil moisture and the combination of surface water and groundwater to compare the results from assimilation to the GRACE/-FO observations and show up chances and challenges for both.

To extract droughts monitored with the TWSA-DSI from assimilation, a minimum duration of nine months is chosen below a threshold of -1.3 (Tab. 7.1). The drought is interrupted if two consecutive months with zero or positive index values occur. These settings are strict compared to, e.g., the common settings for defining consecutive drought events with SPI (McKee *et al.*, 1993), who define the start of a drought as soon as the index falls below zero. In addition, the method in McKee *et al.* (1993) does not require a minimum number of months to define a drought. The reason for the strict choices of thresholds and minimum durations used here is that the focus of this section shall be set on long-term events of at least severe drought severity and with that the analysis aims to extract the longest events in water storage instead of, for example, the droughts with the strongest deficit.

Fig. 7.13 shows the duration, deficit, and the timing of the drought with the longest duration per grid cell either extracted from the TWSA-DSI (detrended) from GRACE/-FO observations (left) or from the GLWS assimilation (right). The TWSA are detrended for this analysis as the linear trends bias the drought detection results in the way that the longest droughts are only identified at the beginning or end of the time period. The droughts with a strong deficit for the observations that the EM-DAT database or literature can confirm are, for example, the droughts in East Brazil and India around the years 2003/2004 (e.g., Sinha *et al.*, 2017; Hulsman *et al.*, 2021), the drought in the Zambezi basin in 2006 (e.g., Thomas *et al.*, 2014; Hulsman *et al.*, 2021), drought in the Amazon basin, Argentina, Congo basin around 2009/2010 (e.g., Frappart *et al.*, 2013; Abelen *et al.*, 2015; Elameen *et al.*, 2023), and drought in the Great Lakes around 2012/2013.

The same events are found when repeating the analysis with the TWSA-DSI from assimilation. Much more spatial detail is given due to the ability of the assimilation system to spatially downscale the satellite observations to the model's resolution. However, it should be noted that assimilation does not apply a simple statistical downscaling but considers spatial correlations and correlations between the water storages. A good example of showcasing is the results for East Brazil. The area that is prone to drought for GRACE/-FO shows a larger extent as found with GLWS. Interestingly, the increase in spatial resolution also leads to higher noise, for example, in the north of Africa but this can be explained: Finding consecutive drought periods requires the definition of temporal limitations and drought in neighboring cells can vary for a few months. Hard limits of a minimum requirement of drought duration of nine months then might introduce spatial noise. As an interim summary, the usage of GRACE/-FO TWSA as well as the TWSA from assimilation for identifying consecutive drought events and the choice between the two depends on the application and its requirements for spatial resolution.

A further basin where a long-term drought is evident in the GRACE/-FO and GLWS data is the Yenisei basin. In connection with GRACE/-FO no clear drought event was found in the literature. In addition, the EM-DAT database does not record an event and other literature does – to my knowledge – not report a drought event in 2012 in the Yenisei basin. This could be a hint that the drought detection failed in this case due to the trend removal but could also indicate an under-representation of the North Asian basins in the literature. Also, GLWS shows some regions affected by long-term drought which are not shown for GRACE/-FO, for example, a drought in 2009/2010 in Sierra Leone. Indeed, Sierra Leone experienced severe rainfall deficits in 2009 (Fayiah *et al.*, 2022) that could be responsible for the shown drought in assimilation-derived TWSA. As Sierra Leone is a very small country with a large coastline, the satellite observations might miss the drought due to the mentioned leakage effect (Sec. 2.3.2) or the coarse resolution. GLWS inherits in this case the WaterGAP properties. The smooth transition of assimilation between observation and model simulation is already frequently discussed in this thesis and will not further be elaborated on.

The same approach of extracting consecutive drought events is subsequently applied to the chosen variables for the drought hazard risk computation. As explained in Sec. 7.3.1, the drought hazard risk is computed from soil moisture (related to rainfed agriculture) and from adding the storages of surface water and groundwater (related to irrigated agriculture or water supply). For the surface-/ groundwater combination a detrending showed more realistic results than keeping trends in the data, so the same approach as for TWSA is applied. For soil moisture, a more complex situation is given. The soil storage in WaterGAP has a lower boundary and cannot be negative, whereas the lakes and groundwater can be negative because they are defined as relative storage in WaterGAP. Although a trend removal for soil moisture might be reasonable in some study regions, an imperfect detrending can also introduce misleading trends. For example, when soil moisture is close to zero in arid regions, and has stronger noise towards the end than at the beginning of the time series, a trend removal would introduce artificial trends. These artifacts are stronger impacting results as compared to if trends are kept in the data, therefore, a trend removal is not applied for the soil moisture. Apart from the trend removal, the same threshold and minimum duration from the TWSA application are also applied to soil moisture and the surface-/groundwater combination.

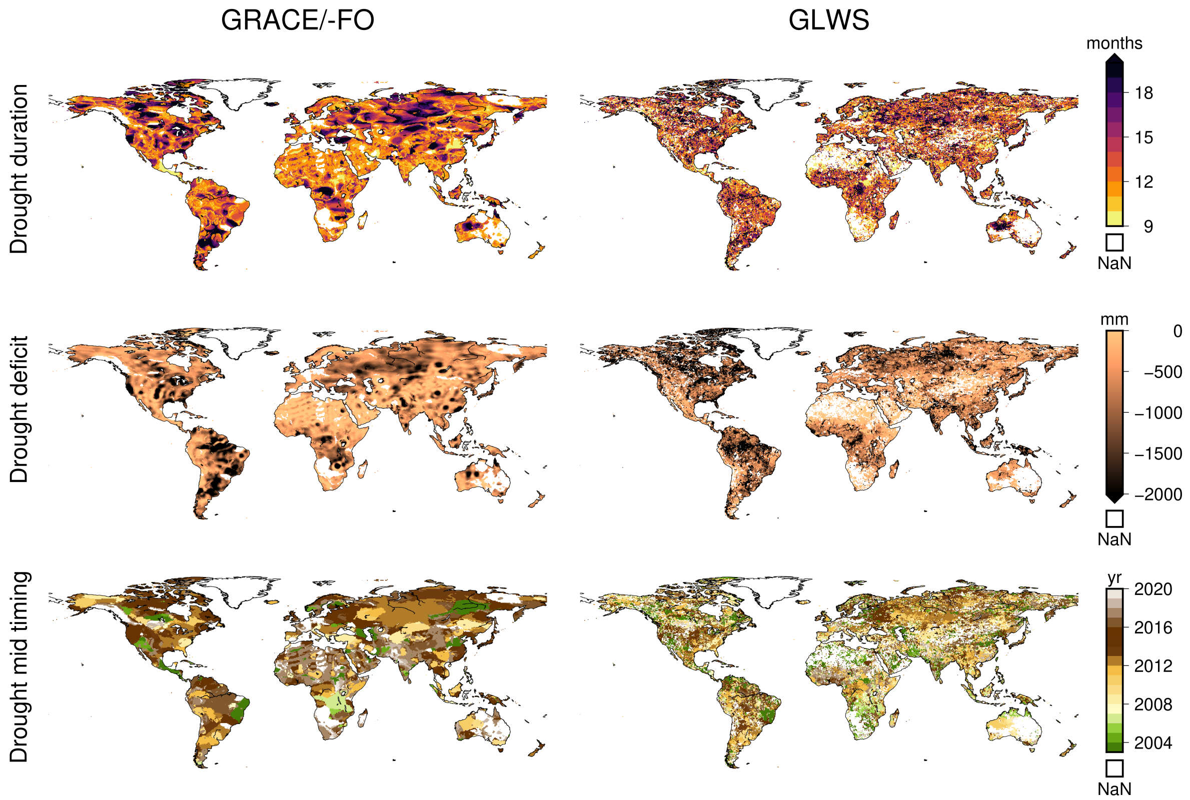


Figure 7.13: Duration (top), deficit (middle), and timing (bottom) of the longest consecutive drought per grid estimated from TWSA-DSI for GRACE/-FO observations (left) or GLWS assimilation (right). The duration is given in months, the deficit is given in mm and the mid timing is given in years.

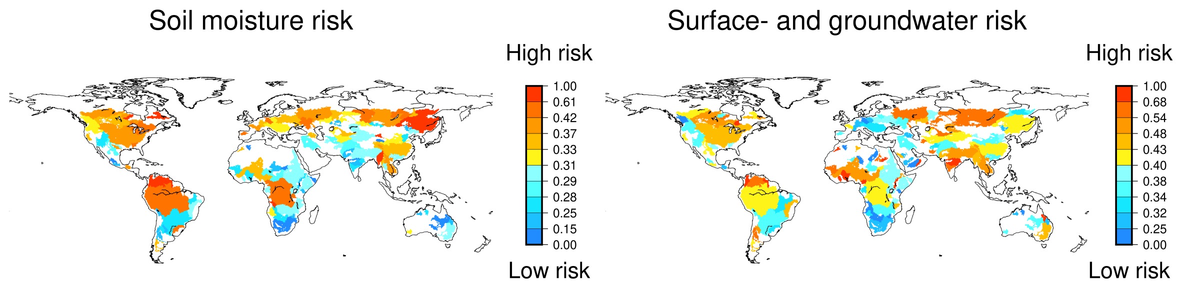


Figure 7.14: Drought hazard risk for long-term (9 months) risk for soil moisture and for the combination of surface and groundwater from GLWS assimilation averaged per basin. The drought risk values are standardized and therefore unitless. The colorbar ticks are chosen according to the decimal quantiles of the respective drought hazard risk assessment.

Fig. 7.14 shows the drought hazard risk either for soil moisture (left) or surface and groundwater (right). The drought risk is shown as basin average to highlight watersheds at risk but could also be evaluated for other regional boundaries, for example, the biomes presented in this thesis or national borders. The basins are chosen as some of the biomes clusters are very large and cover much larger areas than the basin polygons and the national borders are anthropogenic outlines apart from meteorological, hydrological, or vegetation features, which is not the focus of this thesis. However, as national borders for the risk assessment with GLWS might become relevant for mitigation tasks of regional or country-wide governments, the results for the risk assessment with national borders is attached (Fig. B.2). In addition to drought hazard risk assessment with basin polygons in Fig. 7.14, the 15 basins with the highest risk are gathered in Tab. 7.2 for each of the two hazard risk assessments.

The highest risk for long-term drought equal to or longer than nine months in soil moisture for the 17-year period (2003-2019) is found for river basins in North America, which are the La Grande basin, the Churchill basin, and the Severn basin. Thus, river basins in Hudson Bay seem to be at a high risk of soil moisture drought. Multiple basins from Asia are found to be at a high risk as well, for example, the Shule basin, Aldan basin, and Maya basin. Additionally, the Upper Lena basin and the Amur basin also experience a high risk of soil moisture drought. A third striking region with a number of river basins that experience a high risk of drought for soil moisture is in the north of South America with the Orinoco, Essequibo, and Magdalena basins. Similarly, a high drought risk is also found in the north of South America when considering groundwater and surface water. The other regions do not show up prominent in the 15 basins with the highest risk. In contrast, Africa and Asia – more precisely India – are the regions where a high surface/-groundwater risk is found. This is underlined by the fact that the top four basins for the highest surface-/groundwater risk are found in Africa for the Sanaga basin, Bandama basin, Lake Toshka, and Volta basin. An important point to mention is that a high risk of surface-/groundwater drought risk is also found for basins that temporally carry water as the Wadi Draa. In general, the risk values for surface- and groundwater are higher compared to the risk values of soil moisture. As a normalization from between zero and one was applied to compute the risk, it should be mentioned that a direct comparison of the soil moisture risk and surface-/groundwater risk is not reasonable.

The approach for drought hazard risk determination in this thesis differs from other risk assessments as long-term droughts of nine months with at least severe severity are extracted from the underlying soil moisture and the combination surface-/groundwater. In contrast, Meza *et al.* (2020) use evapotranspiration and streamflow for the computation of drought hazard, extract drought with a minimum duration of two months and do not provide information about

Table 7.2: Top 15 river basins for the longest drought events either for the soil moisture drought hazard risk or the surface- and groundwater drought hazard risk from GLWS assimilation. The countries on rank one have the highest risk and with increasing rank, the risk is decreasing.

| | **Soil moisture drought risk** | | **Surface- and groundwater drought risk** | |
|---|---|---|---|---|
| **Rank** | **River basin** | **Continent** | **River basin** | **Continent** |
| 1 | La Grande | North America | Sanaga | Africa |
| 2 | Douro | Europe | Bandama | Africa |
| 3 | Churchill (Atlantic) | North America | Lake Toshka | Africa |
| 4 | Severn (Can.) | North America | Volta | Africa |
| 5 | Essequibo | South America | Narmada | Asia |
| 6 | Maya | Asia | Mahanandi | Asia |
| 7 | Shule | Asia | Turkana | Africa |
| 8 | Orinoco | South America | Magdalena | South America |
| 9 | Magdalena | South America | Essequibo | South America |
| 10 | Aldan | Asia | Burdekin | Australia |
| 11 | Irrawaddy | Asia | Godovari | Asia |
| 12 | Cuanza | Africa | Dhofar | Asia |
| 13 | Amur | Asia | Moose | North America |
| 14 | Upper Lena/Vitim | Asia | Wadi Draa | Africa |
| 15 | Sanaga | Africa | Colorado (Arg.) | South America |

the extracted severity and Carrão *et al.* (2016) use precipitation for the computation of drought hazard and partitions the drought and non-drought months via threshold from Fisher-Jenks algorithm. For example, Meza *et al.* (2020) identifies high drought hazard risk for the agriculture (rainfed and irrigated agriculture joint) in arid regions mainly located in, among others, North and South Africa, Australia, Mongolia, Russia, the USA, and Canada. Despite the differences in methodology, some of the findings agree, for example, high drought risk in Canada, Russia, and Mongolia. In more detail, a high surface-/groundwater risk and a low soil moisture risk is found for Burkina Faso and Côte d'Ivoire when the drought assessment is computed per country instead of basins (Fig. 7.15). This agrees with findings from Meza *et al.* (2020) (Fig. 5), who found a high risk of irrigated agriculture and a low risk of rainfed agriculture there. However, there are also disagreements, for example, for the drought risk assessments for South Africa in general or Australia. As these two examples are arid regions and the experience showed that such arid regions have a low signal-to-noise ratio, the strong restrictions of long-term drought might exclude these regions from our analysis. But overall, the risk computation based on water storage information of GRACE/-FO assimilation presents a new opportunity to determine regions that are experiencing high risk and the approach could be adapted in the future to include for example short-term droughts or choose a different minimum severity of drought event depending on the applications.

### 7.3.3 Warning alerts

Finally, the two emulated prototypes of a warning alert system – one for drought hazard in rainfed agriculture and one for irrigated agriculture and water supply – based on GRACE/-FO assimilation were introduced in Sec. 7.3.1 and are now presented. In fact, the first system includes drought warning alerts estimated from drought indices for precipitation, soil moisture, and vegetation (with LAI), and for the second system precipitation, soil moisture, and the combination of surface water and groundwater is used. The determination of drought events includes an index threshold of -1, minimum duration of two months, and maximum pause of a drought

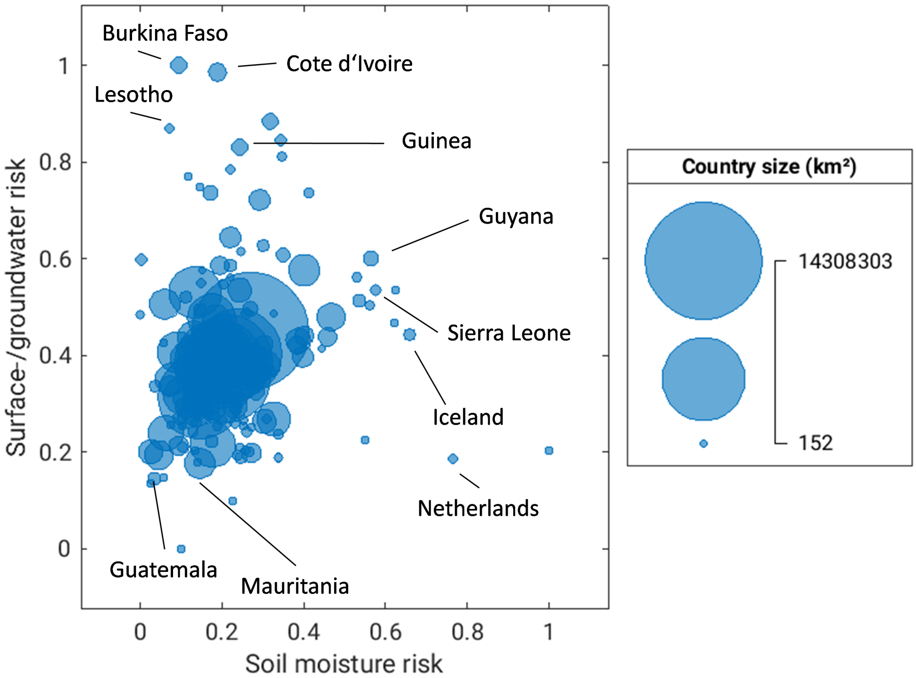


Figure 7.15: Drought hazard risk for long-term (9 months) risk for soil moisture versus the risk of the combination of surface and groundwater from GLWS assimilation averaged per country. The drought risk values are standardized and therefore unitless. The country size is indicated by the size of the bubbles [$km^2$].

of two months (as described in Tab. 7.1) for each of the variable of the two warning systems. Considering the results from the monthly drought monitoring (Sec. 7.2) and the risk analysis (Sec. 7.3.2) it was figured out that a drought detection with surface water and groundwater might be biased by linear trends and these are removed, whereas soil moisture performs for some basins better if there is no trend removal because of no possible negative values for that storage. Similar to soil moisture, precipitation and LAI are per definition zero or positive, therefore, a trend removal is only applied for the surface-/groundwater.

Fig. 7.16 gives an overview of the emulated warning alerts from 2003 to 2019 within the 140 drainage basins. Basins number 1 to 34 are African basins, 35 to 69 are Asian basins, 70 to 77 are Australian basins, 78 to 81 are Canadian basins, 82 to 106 are European basins, 107 to 125 are North American basins and 126 to 140 are South American basins (Sec. 6.1.3). On basin average, the highest warning alert "alert" is executed approximately one time per basin and system during the period 2003 to 2019. The categorization into continents/countries helps to find overlaps of droughts in space and in time. Multiple basins in South America show longer watch-warning phases, and some basins even experience the alert mode in 2016, which can be confirmed by, for example, Erfanian *et al.* (2017) who found that this drought was unprecedented in severity. In Africa, the years 2010 and around 2016 also show many basins with the highest warning level of the system, which is also known in the literature, for example for the 2010 drought at the Horn of Africa and the Sahel zone (e.g., Dutra *et al.*, 2013; Swift & Saulle, 2015). Multiple basins in Europe show longer warnings around the years 2003 and 2018. It is well-known that a drought existed in Europe (e.g., Andersen *et al.*, 2005; Boergens *et al.*, 2020), however, soil moisture is not detrended for this application. The GRACE/-FO observations that are assimilated experience high uncertainty in Northern America and Europe due to the ice ages that still affect gravity nowadays (Sec. 2.3.2), and although the effect is removed via geophysical models, a possible (smaller) influence might still be included. Thus, especially drought events at the beginning and the end of the period should be handled with caution (Sec. 7.1.2. As not-detrended soil moisture is also included in the second system (Fig. 7.16, right), the careful analysis of early and very late events is important.

In general, the second system including surface- and groundwater shows overlapping temporally coherent drought events, e.g., in 2010 in Asia, and in 2016 in South America but with a different temporal course of warning levels per basin. In some cases, the surface-/groundwater system detects temporal coherency that is not found in the soil moisture system. For example, around 2010, many basins in Asia experience a high level of warning, which is confirmed by the EM-DAT database for e.g., India. In addition, alerts are found for basins across Australia, nonetheless, Australia still experiences drought, for example, in multiple basins around 2005 to 2008 (e.g., Wang *et al.*, 2021) but the comparison of the two systems indicates that Australia is more prone to drought in soil moisture than in surface- and groundwater. For surface water, this was expected because the extent of surface water bodies in Australia is small compared to, for example, North America (Fig. 4.3). Thus, the comparison of the two systems is a further possibility to identify which storage per basin is more prone to drought hazards.

The findings indicate, that a drought warning system cannot only be used for the determination of temporally subsequent drought events but also relate this temporal relation to the spatial propagation of drought through multiple basins across one continent. Therefore, a closer look into warning levels of a choice of basins for Asia is exemplarily given.

Fig. 7.17 shows the warning alerts for the precipitation – soil moisture – groundwater system for the five river basins Ganges, Godavari, Mahanandi, Mekong, and Salween from 2003 to 2019 (top) and the spatial extent of the warning levels for South East Asian river basins (bottom) from January to December 2010. The location of the five basins and their designation is additionally presented in Fig. 7.18. All five basins experienced one of the warning levels in 2010 but with different durations, intensities, and timing. For example, the Ganges basin experienced the longest duration of the alert level of about five months in 2009 accompanied by watch or warning levels before and after. Similarly, the Mekong basin shows watch, warning, and alert but with a temporal delay. The alert is executed in 2010 instead of 2009. A positive example of a successful watch-warning-alert sequence is the Godavari basin, which starts with the watch mode and turns into warning mode before the alert is imminent. After that, the alert is reduced to warning first and watch mode next before the drought event is finished. A look at the spatial maps reveals the spatial extent of the droughts for each of the basins. In March 2009, the Ganges and Godavari basin starts to experience a drought warning or watch mode. In mid-2009, the mode increases for the two basins, while other basins more eastern experience a watch phase. For multiple months, the alert status is activated in the Ganges basin, and the alert or warning status in the Godavari basin until October. After October, the two basins return to watch conditions whereas more eastern basins as the Mekong and Salween basins turn into the alert status. Together with the drought timing identified above, the drought in 2009 and 2010 propagates from west to east in time. This comprehensive picture of drought spatial and temporal relation would not be feasible by only considering single climate variables, which highlights the benefit of the suggested warning alert system.

In summary, the development of the two prototype warning systems provides multiple opportunities. When extending the systems to an operational framework, stakeholders would have the opportunity to make improved management plans depending on the warning level and affected system and the comparison of basins or other regional aggregated information per choice enables the identification of spatiotemporal relations in a simplified warning environment

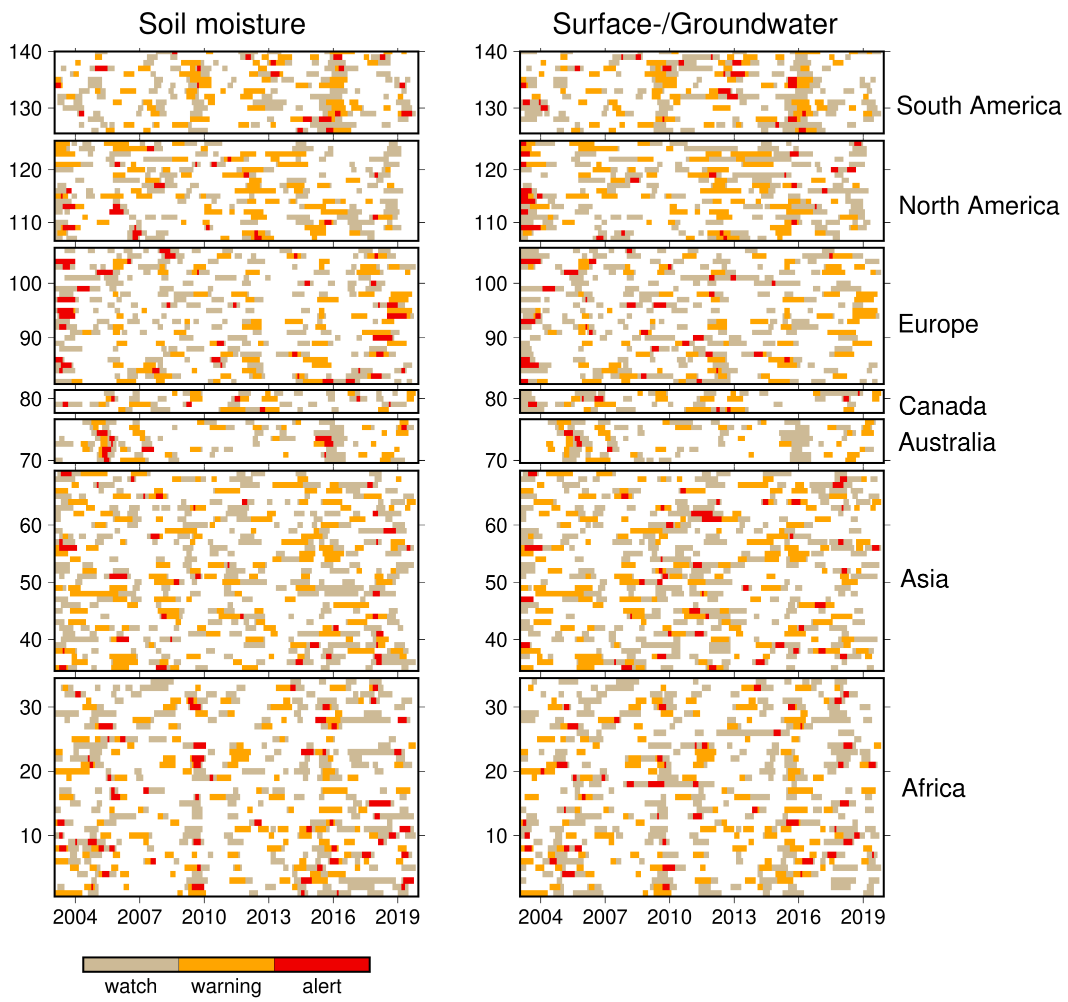


Figure 7.16: Warning alerts for the two prototype warning alert systems SPI, soil moisture DSI, and LAI-DSI or SPI, soil moisture DSI, and surface water/groundwater DSI for each of the 140 basins plotted across time. The storage information is derived from GLWS assimilation.

instead of judging single drought indices. Nonetheless, it is important to keep in mind that the two suggested systems also have challenges that need to be tackled in the future, such as biases in drought detection which may be caused by, for example, trends in climate variables. In addition, confirming spatial and temporal relations of drought events is very difficult as no external validation is available that find spatiotemporal relations of drought.

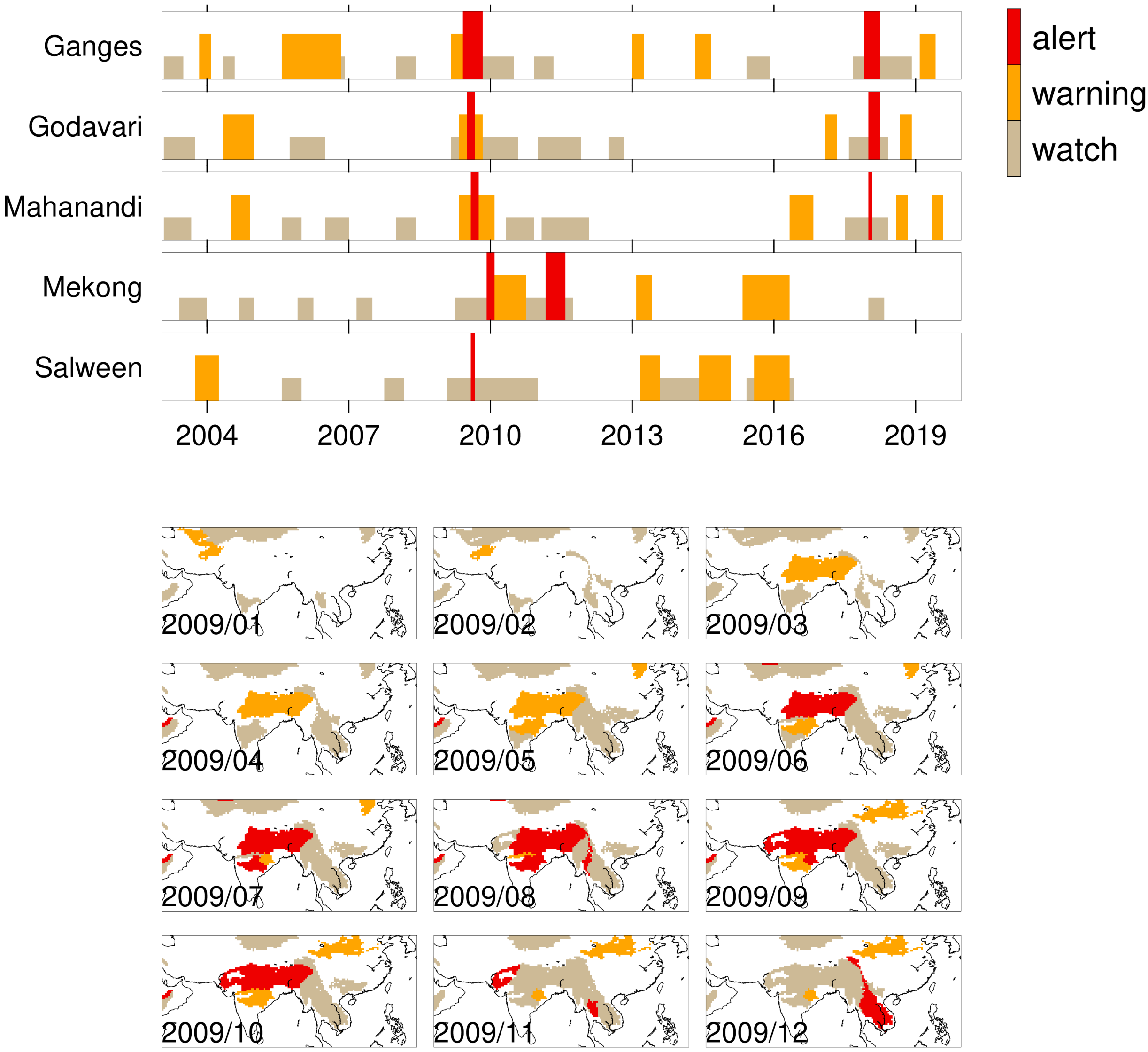


Figure 7.17: Warning alerts for the prototype warning alert system with the precipitation-based SPI, the soil moisture DSI, and the surface-/groundwater DSI either shown from 2003 to 2019 for the five river basins Ganges, Godavari, Mahanandi, Mekong, and Salween (top) or as spatial maps for January to December in 2009 for South East Asia (bottom). The surface water and groundwater is derived from GLWS assimilation.

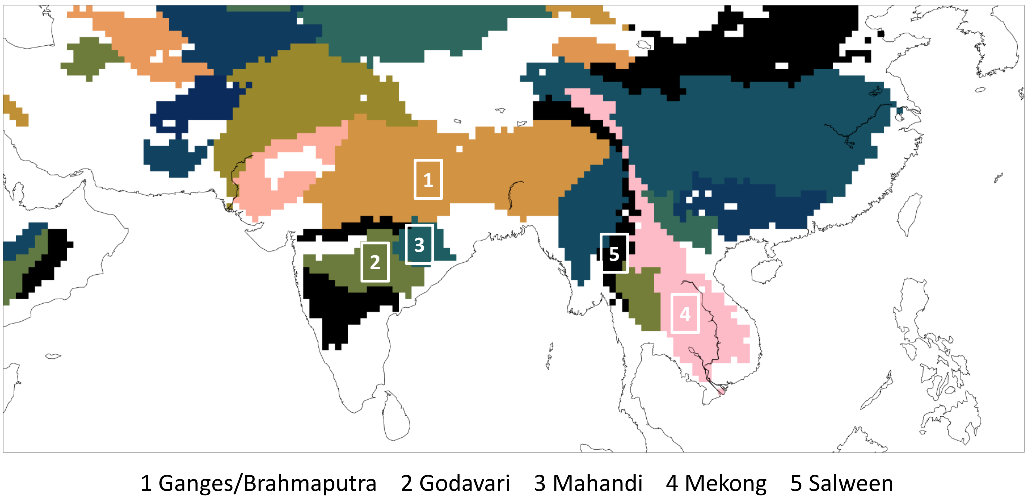


Figure 7.18: Location and designation of the basins Ganges, Godavari, Mahandi, Mekong, and Salween.

# Chapter 8

# Conclusion and outlook

## 8.1 Conclusion

The thesis had three major objectives. The first objective was the development of a global data assimilation system that assimilates GRACE and GRACE/-FO (GRACE/-FO) satellite observations of Total Water Storage Anomalies (TWSA) into the global hydrology model Water Global Assessment and Prognosis (WaterGAP). The TWSA from the assimilation are more realistic than from model simulations only and are spatially downscaled compared to GRACE/-FO. In addition, the observations can be vertically disaggregated to the storage compartments via assimilation. The second objective was analyzing the assimilation outputs according to signatures, estimating the timing between precipitation peak refreshing water storage and leading to vegetation growth, and validating the assimilation with independent data. The third objective was developing a framework for GRACE/-FO observations and the assimilation outputs to monitor droughts in water storage on the surface but also in subsurface water storages, compute drought risk, and develop a prototype of a warning alert system that monitors drought by considering the precipitation-storage- vegetation dynamics of the hydrological water cycle.

#### Development of a global data assimilation framework

To develop a stable assimilation system on the global scale, tuning options for the GRACE/-FO processing and the assimilation setup were precisely tested to find an optimal configuration for the drought monitoring framework. Non-standard corrections for the GRACE/-FO processing were tested, where I looked specifically into a correction for earthquake signatures in the TWSA time series and correcting signal loss due to filtering via lake and reservoir relocation and rescaling. While the earthquake correction significantly reduces co-seismic and postseismic signals in the GRACE/-FO TWSA on a fine spatial resolution of 0.5° and the 4° input grid for the assimilation, it did not significantly impact the assimilation output itself. On the other hand, the lake and reservoir relocation led to a percentage change in global mean RMS of about 10% and the rescaling of about 3%. As the rescaling applies to all spatial grids instead of only lakes and reservoirs, the rescaling was the preferred choice for signal restoration in this thesis but as the RECOG-LR data set might be updated in the future with more data, and data for smaller lakes and reservoirs, it should be considered for new future setups of assimilation.

I transformed a regional data assimilation framework into a global one by including the Parallel Data Assimilation Framework (Nerger & Hiller, 2013) for improving run-time and storage efficiency. I contributed to the restart functionality of WaterGAP to stabilize assimilation in online mode and implemented that existing model water storage limits are also applied after the monthly update from assimilation. The comparison of two different filter algorithms for assimilation showed no significant impact on the TWSA. The representation of model errors was found to be very sensitive to parameter range and forgetting factor, in fact, the parameters were perturbed using the “small” parameter range, and a forgetting factor of 0.95 was used. A special focus was laid on the representation of full error variance-covariance matrices for the TWSA observations and the consequences of including them in the assimilation framework. To my knowledge, for the first time, a global data assimilation framework was set up that considers error correlations between spatial grid points to account for the GRACE/-FO specific mission constellation. Only a few groups worldwide have set up a global GRACE/-FO data assimilation system but none include real correlations derived from propagating errors from formal covariance matrices from the Level 1-2 processing. As spurious long-range correlation did lead to filter divergence, a workaround with localization for various techniques was extensively tested to include as much model and observation error information as possible. The results showed that domain localization with an exponential weight and a localization radius of $r = 700$ km appears as better suited as compared to covariance localization as it preserves a stable assimilation run. Choosing a spatial resolution of the input observations finer than 4° led to instabilities in the assimilation.

The code for the final assimilation framework was frozen with the mentioned options and a production run provided as the Global Land Water Storage (GLWS) data set release 3.

**Signatures, timing and evaluation of water storages from assimilation**

The analysis of the GLWS in the spatial and spectral domain revealed a smooth transition of the TWSA from assimilation between the GRACE/-FO observations and WaterGAP model simulations. For example, linear trends of TWSA or lower degree spherical harmonic coefficients of model simulations were nudged towards the observations. The partitioning of TWSA trends and annual amplitudes into major contributors of the single compartments of surface water, groundwater, and soil moisture were shown. Nonetheless, it should be noted that the snow and canopy storage are also contained in the TWSA from the assimilation but present a very small part of the TWSA as compared to the other comaprtments. The results revealed that the groundwater linear trends are for 50% of the global land are the strongest contributor to TWSA trends but for northern latitudes soil moisture trends have an increasing relevance. An example for the spectral domain is that the lower degree coefficients of GLWS had a closer variability to the GRACE/-FO observations as compared to the model simulations. A hypothesis test on TWSA from assimilation showed that for most of the global land area cells and months the provided standard deviation is significant, only in very dry climate regions like the Sahara, the significance was found lower.

The precipitation – water storage – vegetation dynamics were analyzed in this thesis in two main parts: (1) The response time of water storages to precipitation peaks and (2) the response time of vegetation to water storage peaks. I used precipitation from a combination of the WATCH Forcing Data methodology and ERA5 reanalysis data sets, surface water, soil moisture, and groundwater from GRACE/-FO assimilation or WaterGAP model simulations, and the LAI from Copernicus data sets SPOT/VEGETATION and PROBA-V. Considering the seasonal evolution of the climate variables, the timing of precipitation peaks leading to water storage peaks identified the shortest response for precipitation with soil moisture, an intermediate response of about one to two months for surface water, and the longest response for groundwater of three to six months depending on the vegetation regime. Similar conclusions were also found when considering the non-seasonal behavior but the duration of precipitation for recharging the groundwater storage took longer with the non-seasonal residual signal as compared to the seasonal signal. Especially interesting was that only for the soil moisture storage, I found that soil moisture peaks led to vegetation growth for most of the considered regions when considering the non-seasonal resiudal part of the data. In contrast, nearly no temporal delay between non-seasonal events in surface water and vegetation was identified in many regions, and groundwater non-seasonal events were often observed as delayed with respect to the non-seasonal vegetation events. The findings explained that vegetation is for many regions on the global land mainly driven by soil moisture. An explanation for the surface water and groundwater peaks that were observed temporally delayed to vegetation peaks was hypothesized to be a response of irrigation to missing vegetation growth. For the use of the water storages from assimilation in the suggested drought monitoring system, these results highlight that soil moisture is well-suited as an early indication that vegetation health might decrease. Nonetheless, a warning alert systems for rainfed agriculture should make use of both storage and vegetation measurements. The estimation of groundwater and surface water changes from remote sensing (GRACE/-FO data assimilation) can contribute to develop warning alert system for rainfed agriculture but are better suited for other sectors such as water supply.

A validation against independent observations was applied against GNSS observations, groundwater well data, surface water from altimetry and remote sensing (RECOG-LR), and soil moisture from a combined product by the ESA-CCI. The validation of TWSA via comparing vertical displacements from GNSS confirmed improvements in the assimilation compared to the model simulation via a correlation analysis, especially for the non-seasonal part of the data. The groundwater validation showed that GLWS has improvements compared to model simulations when considering well data collected over East Brazil and all signals (seasonal versus non-seasonal) but deteriorations when considering spatial median groundwater metrics for six basins in France. The reason was found in temporal lags between the validated data sets and validation data sets as well as in the spatial aggregation methods. A more detailed analysis per basin indicated that for most of the six basins, the assimilation indeed is closer to the in-situ groundwater than the model simulation or performs equally well when a time delay is neglected. Only insignificant changes between surface water or soil moisture from model simulation or assimilation were found, leading to the conclusion that assimilation majorly improves groundwater storage as it is the largest contributor to TWSA. However, it should be noted that the validation of the groundwater, surface water, and soil moisture compartments is generally comlicated as, for example, conversion methods to compare simulated storage to in-situ measurements can be imprecise. Therefore, a future validation should concentrate on the extension of the validation (e.g, more stations) and improve comparability of the different data sets.

#### Drought monitoring using GRACE/-FO and GRACE/-FO assimilation

Ideally, a validation would also be possible for the monitoring of drought with outputs from GRACE/-FO or assimilation but, the validation of droughts is very difficult as the concept of drought is not uniformly defined. Therefore, a synthetic study was constructed to simulate TWSA time series with GRACE/-FO specific properties that contained a synthetic drought event. The advantage of the synthetic study is that the detection of the drought event can be tested for various drought indices. The results showed that trend, accelerations, and noise contained in the synthetic TWSA strongly bias the drought detection, thus, results should be analyzed carefully according to the signatures. Furthermore, the accumulations or temporal differencing of TWSA before computing the indices lead to changes in identified drought severity, timing, and duration. Under specific circumstances, each index has advantages or disadvantages but overall the Drought Severity Index (DSI) showed the most consistent drought monitoring results. The final decision was made to use the DSI for all the following drought applications.

The GRACE/-FO observations and outputs of the data assimilation framework were used to showcase the opportunity of the satellite data for their usage for drought monitoring. The monitoring of monthly drought severity with the DSI applied to monthly TWSA maps revealed several drought hotspots worldwide in 2010 that were confirmed by other literature, for example, drought for the Horn of Africa. A short test highlighted how sensitive drought detection at the beginning and end of the time series is toward the removal of the temporal mean. Subsequently, results for the Congo showcased the extent of the 2005 drought. The results showed that the drought was mainly apparent in the south of the basin and most dominant in the surface and groundwater storage but less in the soil moisture storage. To provide a measure of uncertainty, the percentage of ensemble members experiencing a certain drought severity can be estimated for each storage. The uncertainty for extreme or exceptional drought in TWSA in the Congo basin was highest along the Sukuru tributary and around Lake Tanganyika.

Finally, I showed how one could use assimilation products for risk computation and for setting up a prototype of a warning alert system under the consideration of different impacted sectors. As the identification of droughts within the proposed warning alert system requires consecutive drought events, these are extracted from the drought indices of the respective climate variables by introducing thresholds for minimum drought intensity and duration. The setup yielded two systems, one that can be used for, for example, rainfed agriculture by focusing on soil moisture and vegetation growth, and one that can be used for, for example, irrigated agriculture and water supply by integrating the combination of surface water and groundwater. The soil moisture drought risk was identified as highest for basins in North America around the Hudson Bay, whereas drought hazard risk for surface- and groundwater was highest for many basins in Africa. In a final analysis, the warning alert algorithms were designed to trigger a watch, a warning, and an alert phase. For the first system, the precipitation-based SPI, the soil moisture-DSI, and the LAI-DSI were used for the implementation. For the second system, the LAI-DSI was exchanged by the surface-/groundwater-DSI. The systems enabled the identification of temporally subsequent drought events within the water cycle occurring in multiple basins on the same continent and also identified spatiotemporal relations of drought, for example, for a drought in 2009/2010 in South East Asia.

All in all, the monthly drought monitoring, risk computation, and prototype warning alert system presented an overall picture of the great opportunity for using multiple water storages from GRACE/-FO assimilation for drought detection.

## 8.2 Outlook

The thesis was developed to find a strategy for global GRACE/-FO assimilation and its use for monitoring drought. Further improvements are still possible and new opportunities in the future enable the expansion of each part presented in this work.

A main goal for the future is to enable the global GRACE/-FO data assimilation framework used in this thesis to produce reliable output when providing the input GRACE/-FO observations and full variance-covariance matrices on a spatial grid finer than 4°. One possibility is adapting the assimilation framework in a way that GRACE/-FO observations including full error information can be assimilated without encountering numerical issues that lead to unrealistic state updates. It is expected that under these conditions, the lake and reservoir removal approach would derive more realistic results. In case the leakage effect is additionally reduced, the removal of earthquake correction would also be more significant for the assimilation. Another possibility for improving the spatial resolution of the observations is given with future satellite missions that are already confirmed or planned. With, for example, the MAGIC mission planned by ESA and NASA[1] consisting of two satellite pairs, not only better spatial and temporal resolution is expected as

[1] https://www.eoportal.org/satellite-missions/magic#eop-quick-facts-section (last accessed 15.03.2024)

compared to GRACE/-FO, but the launch will also extend the length of the existing observation time series. Further future missions could include new and better sensors for measuring gravity changes. Currently, the assimilation is performed by assimilating monthly fields of TWSA. As future missions also might provide sub-monthly fields and the WaterGAP model can run daily, the update step could be performed more often. In addition to the improved spatiotemporal resolution of the observations, one could integrate the current global assimilation framework in models that use spatial resolutions below 0.5° for deriving water storages on a scale finer than the WaterGAP model.

A few studies further developed assimilation systems to not only perform data assimilation but also calibrate model parameters simultaneously. On a regional scale, some successful applications already exist (e.g., Schumacher *et al.*, 2016) but to my knowledge, no global calibration/data assimilation was performed. By including parameter calibration, the hope is to derive a more realistic prediction of water storages in general and during periods where no observations are available. Further techniques that might be used in the future to fill gaps between the gravity missions is deriving TWSA from other satellite missions, for example, from Satellite Laser Ranging or predicted fields with deep learning/neural networks. In addition, the data sets could be provided years before the launch of GRACE/-FO or, in the case of deep learning, could be provided as forecasts. The extension and gap-filling of storage data sets could be, among other applications, helpful for the monitoring of extreme events such as droughts and floods because the indices used for the monitoring benefit from long time series.

The single compartments from assimilation could be further improved. Changes in the assimilation setup (e.g. increased spatial observation resolution) hopefully improve the simulation of groundwater, surface water, and soil moisture. In recent years, the assimilation of multiple observation data sets gained more attention. GRACE/-FO TWSA was, for example, jointly assimilated with soil moisture (e.g., Tian *et al.*, 2017; Girotto *et al.*, 2019) or discharge (Schulze *et al.*, 2024). The approaches present a new opportunity to improve multiple water compartments or fluxes at the same time, but existing literature also shows that it might introduce trade-offs. As this topic is still under development, it is a very interesting field for future research. Furthermore, the validation of the water storages should be extended to further regions all over the world, for example, the groundwater validation. However, spatial expansion is often difficult because it requires often missing hydrogeophysical information to transform the groundwater level measurements into groundwater storage. Future research could concentrate on improving the availability of hydrogeophysical information.

Currently, the thesis incorporates soil moisture, surface water, and groundwater as these three water storages present the dominant contributor to the TWSA for most regions of the world. As the world is prone to climate change that especially affects the melting of glaciers etc., analyzing snow from assimilation (e.g. in mountain regions) would complement the analysis of trends, amplitudes, response time, etc., and possibly also be relevant for drought detection. In addition, the simulations of hydrological models (and thus the assimilation) could be further improved by integrating or coupling glacier models into the hydrological models.

The monitoring of drought with indices always relies on comparing the current state to the defined "normal", which is here often defined as seasonality. Herbert & Döll (2023) encountered that removing the seasonality to derive drought indices could not always perfectly reflect the "normal" situation of water storage or flux in some regions, i.e. the habituation of the population. This is for example the case in regions where the population is better habituated to the inter-annual variability. Together with the conclusions that linear trends, accelerations, and noise bias drought detection, an automatic workaround could be set up that includes the removal of certain signatures and choosing “normal” according to the habituation of the population living in the study region to improve the performance of the indices.

The last but most important point for monitoring systems is to enable near real-time application. The strongest limiting factor at the moment is that the GRACE/-FO data are operationally available for near-real-time applications with a delay of approximately two to three months. If future gravity missions like MAGIC would have a shorter delay of, for example, one to two weeks, near-real-time TWSA could be provided for and used to set up an operational global assimilation framework and both could be implemented in early warning systems for drought. Depending on for which sector the operational warning alerts shall be implemented, the current system setup could be adapted, for example, by unifying the two presented systems to systems that present warning levels for subsequent drought events for the precipitation - soil moisture - vegetation - surface water - groundwater dynamic.

# Appendix A

# Additional results assimilation

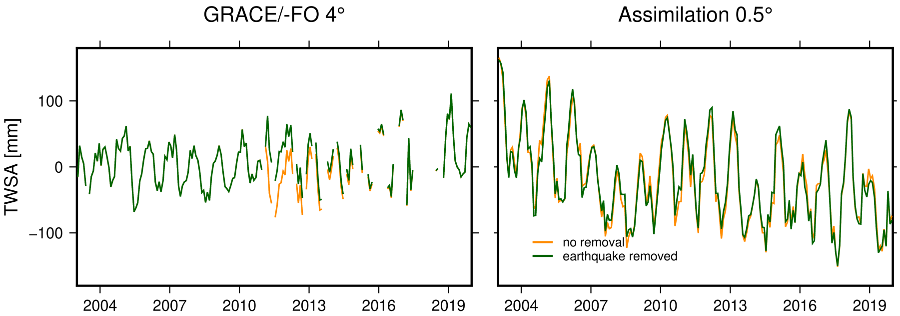


Figure A.1: TWSA [mm] with and without earthquake correction derived from a 4° GRACE/-FO grid or from the 0.5° GLWS assimilation spatially averaged for Japan.

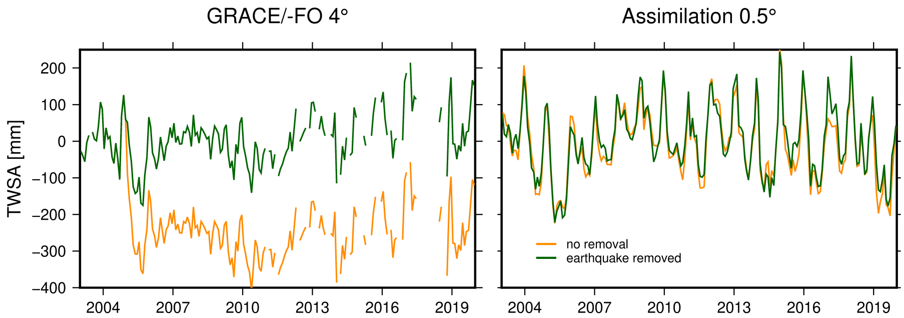


Figure A.2: TWSA derived from a 4° GRACE/-FO grid or from the 0.5° GLWS assimilation spatially averaged for West Malaysia. All is derived from applying rescaling to the GRACE/-FO TWSA.

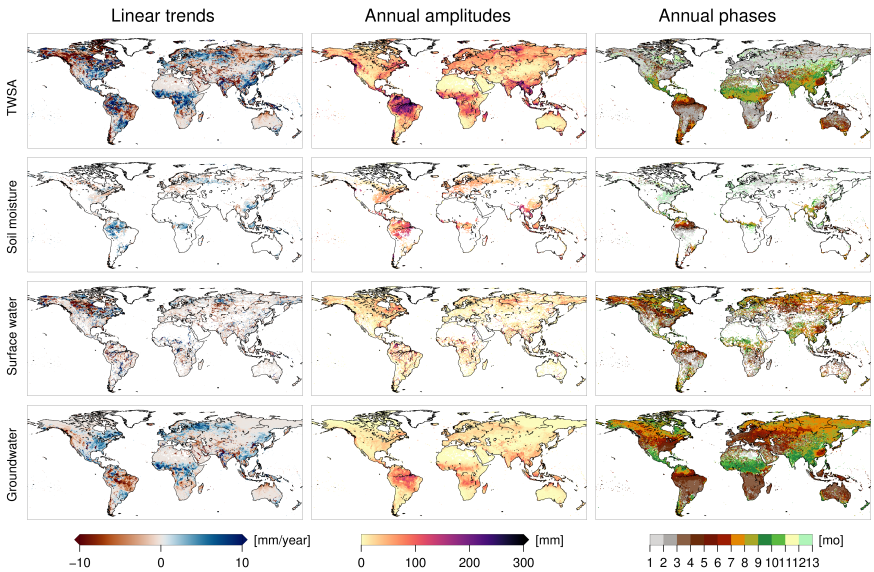


Figure A.3: Linear trends [mm/year] (left), annual amplitudes [mm] (center) and annual phases [months] (right) of TWSA (top), soil moisture (middle top), surface waters (middle bottom), and groundwater (bottom) derived from GLWS release 2.

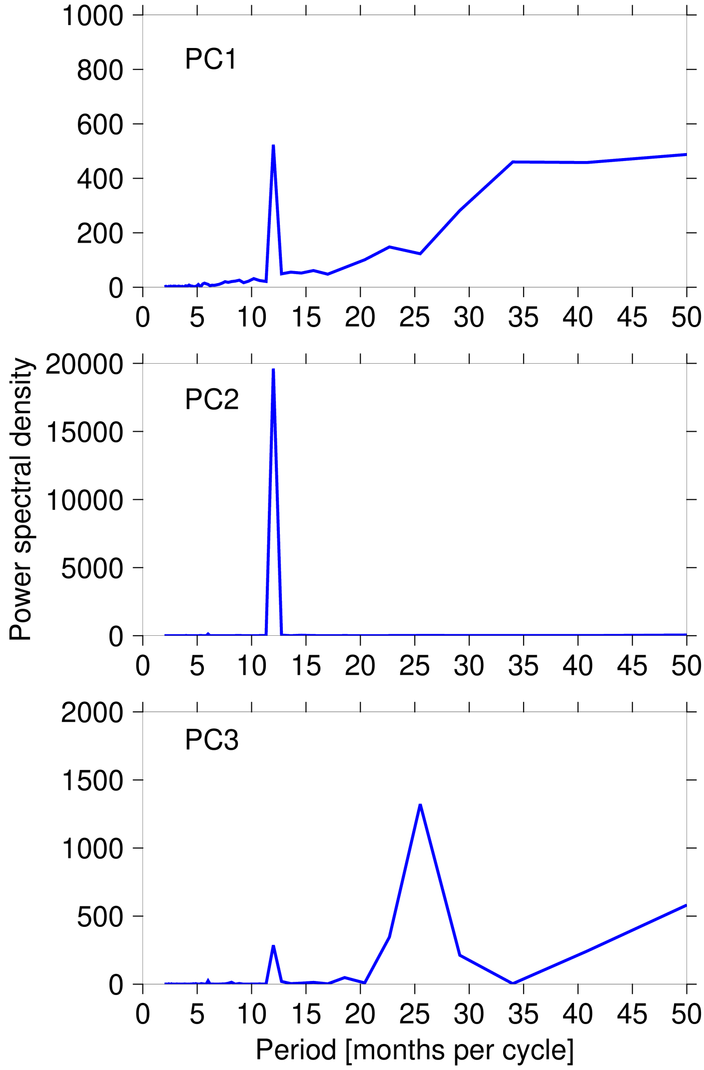


Figure A.4: Power spectral density of first three dominant PCs [-]) of GLWS derived TWSA.

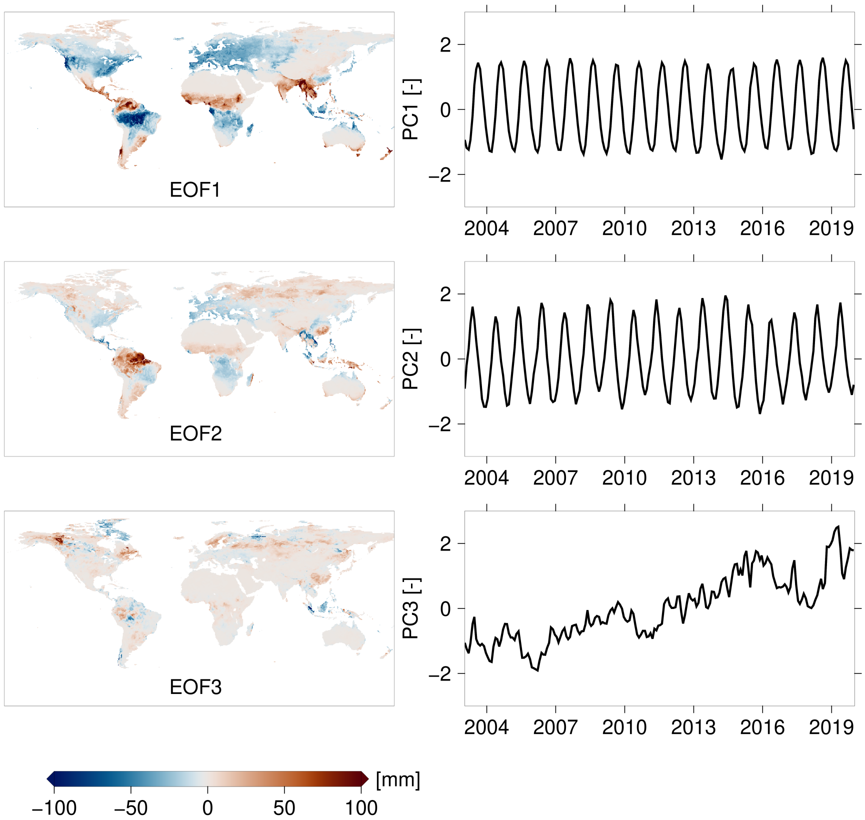


Figure A.5: First three dominant PCA modes (EOFs and PCs) of GLWS assimilation-derived soil moisture. The EOFs are given in mm and the PCs are unitless.

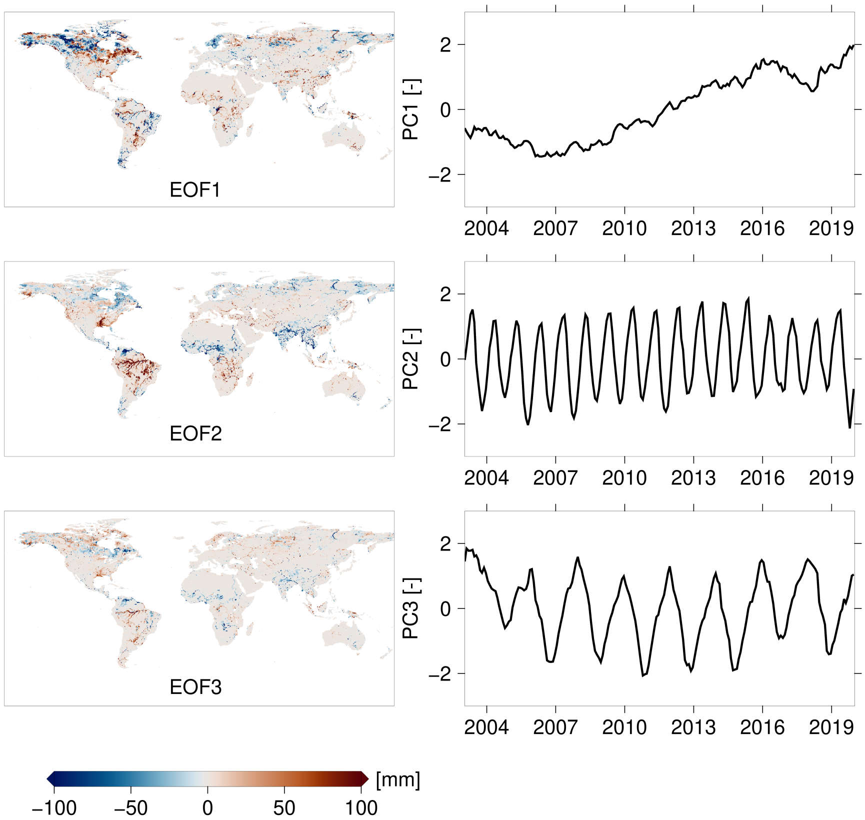


Figure A.6: First three dominant PCA modes (EOFs and PCs) of GLWS assimilation-derived surface water. The EOFs are given in mm and the PCs are unitless.

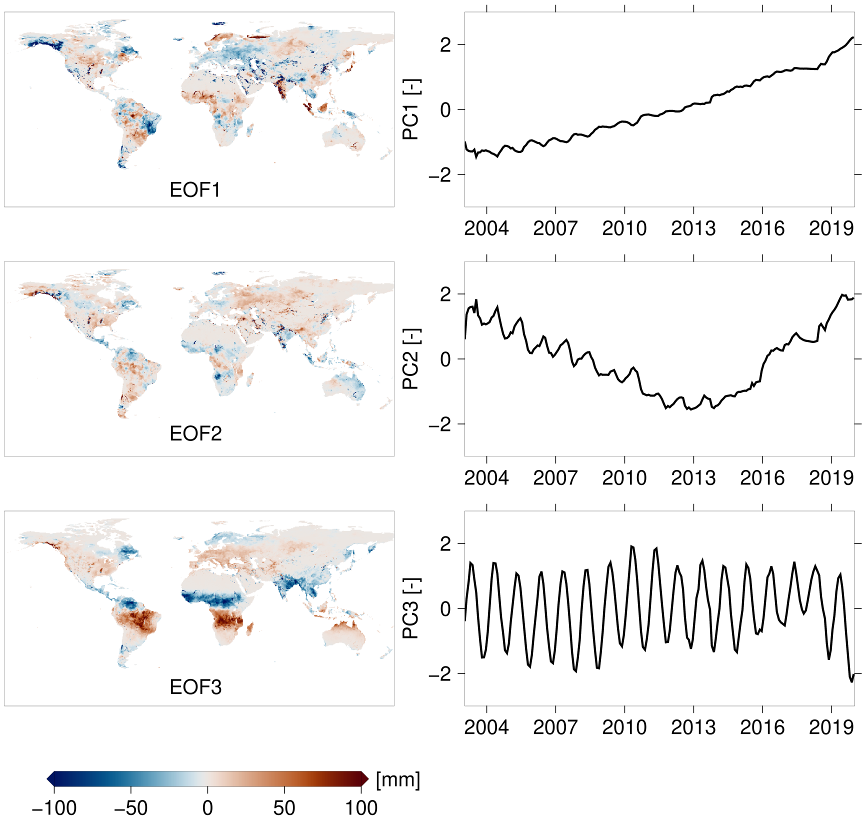


Figure A.7: First three dominant PCA modes (EOFs and PCs) of GLWS assimilation-derived groundwater. The EOFs are given in mm and the PCs are unitless.

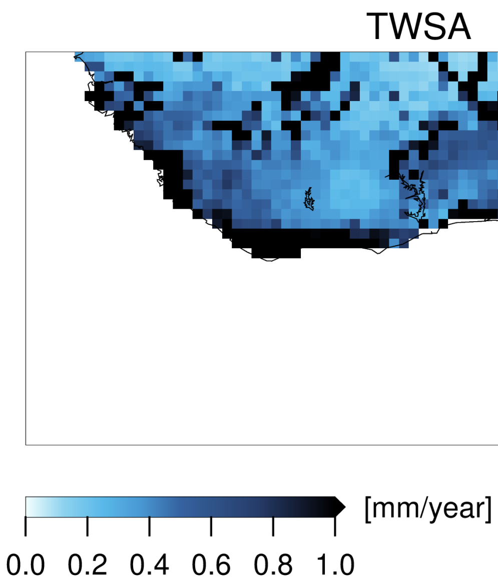


Figure A.8: Standard deviations of the TWSA linear trends [mm per year] derived from GLWS assimilation. This is the same figure as Fig. 6.22 but zoomed in to West Africa to showcase large uncertainties at the coast. Based on Gerdener *et al.* (2023a).

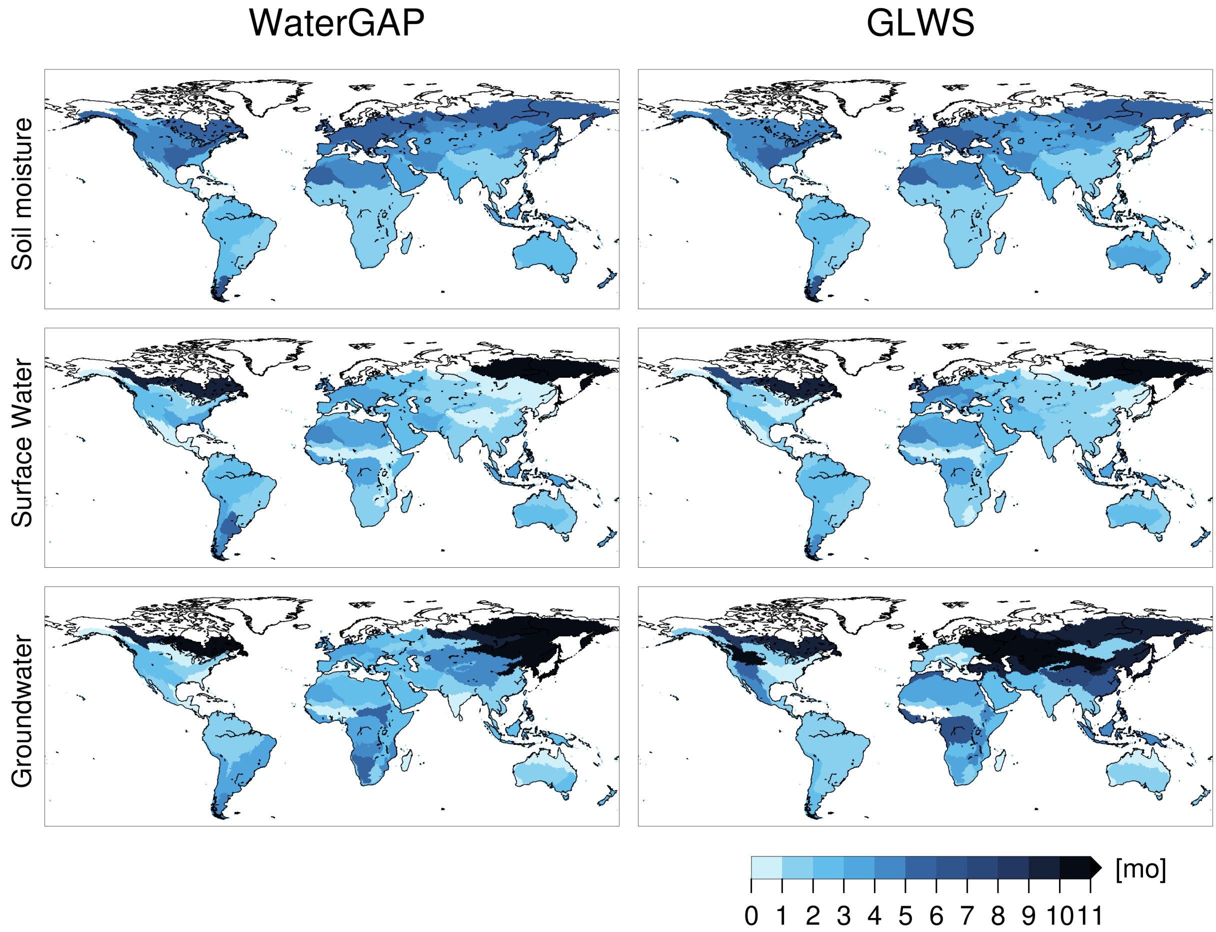


Figure A.9: Annual phases [months] of soil moisture (top), surface water (middle), or groundwater (bottom) relative to LAI either derived from WaterGAP model simulation (left) or GLWS release 3 assimilation (right). The relative phases are estimated using multi-linear regression from 2003 to 2019.

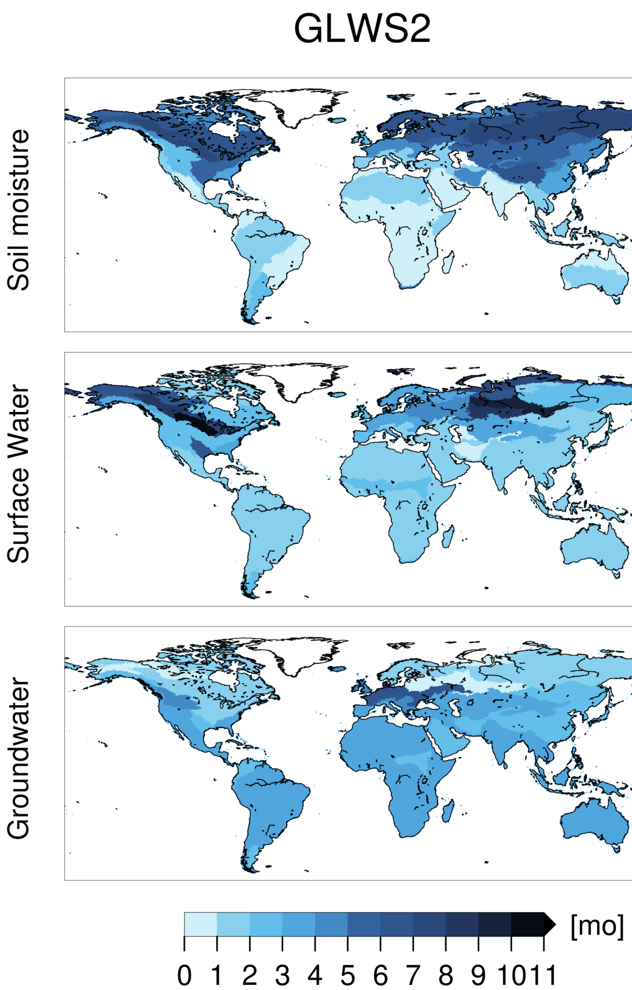


Figure A.10: Annual phases [months] of precipitation relative to soil moisture (top), surface water (middle), or groundwater (bottom) derived GLWS release 2 assimilation. The annual relative phases are estimated using multi-linear regression from 2003 to 2019.

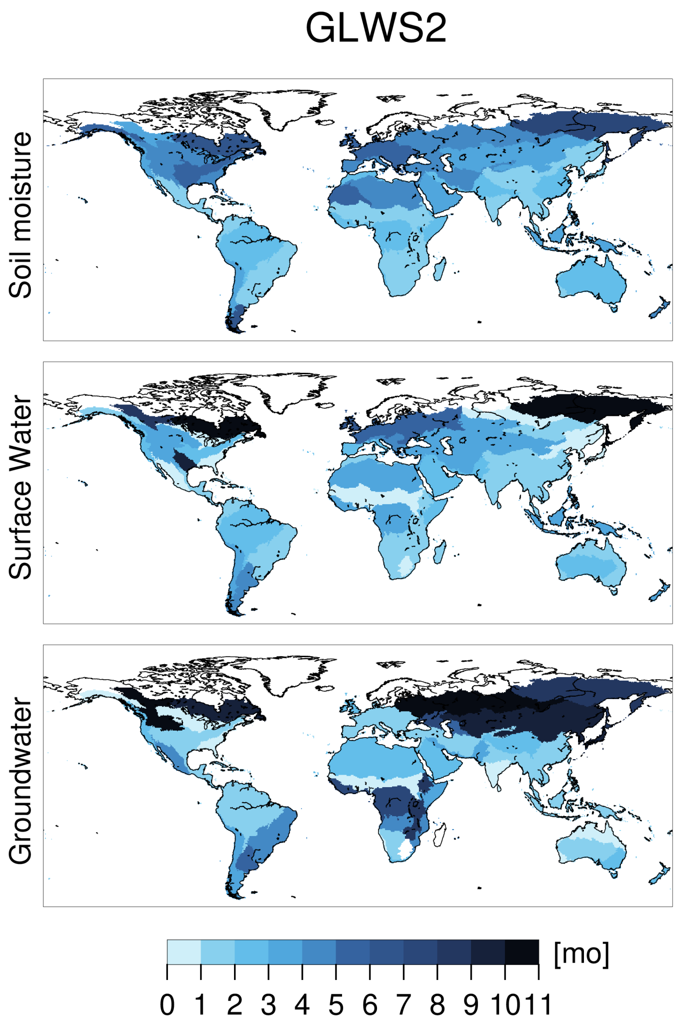


Figure A.11: Annual phases [months] of soil moisture (top), surface water (middle), or groundwater (bottom) relative to LAI derived from GLWS release 2 assimilation. The relative phases are estimated using multi-linear regression from 2003 to 2019.

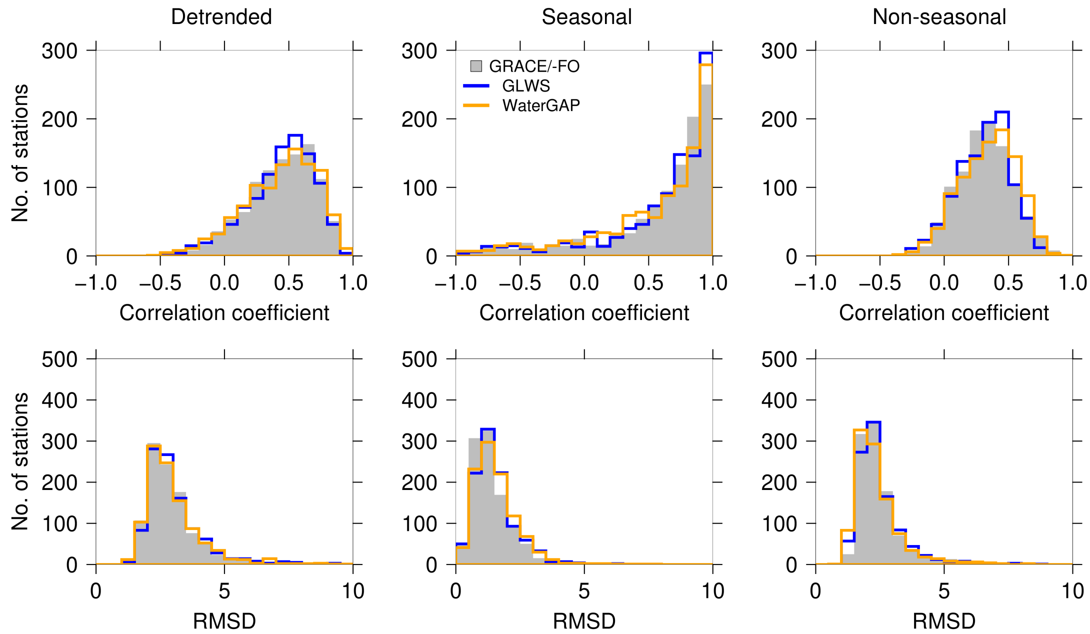


Figure A.12: Correlations (top) and RMSD (bottom) of vertical displacements from GNSS station worldwide with vertical displacements from GRACE/-FO observations (grey), GLWS release 2 assimilation (blue), and WaterGAP simulations (orange) estimated for 2003 to 2019. Additionally, the correlations and RMSD are either shown for the detrended signal (left) or for the seasonal (center) and the non-seasonal part (right) part of the data.

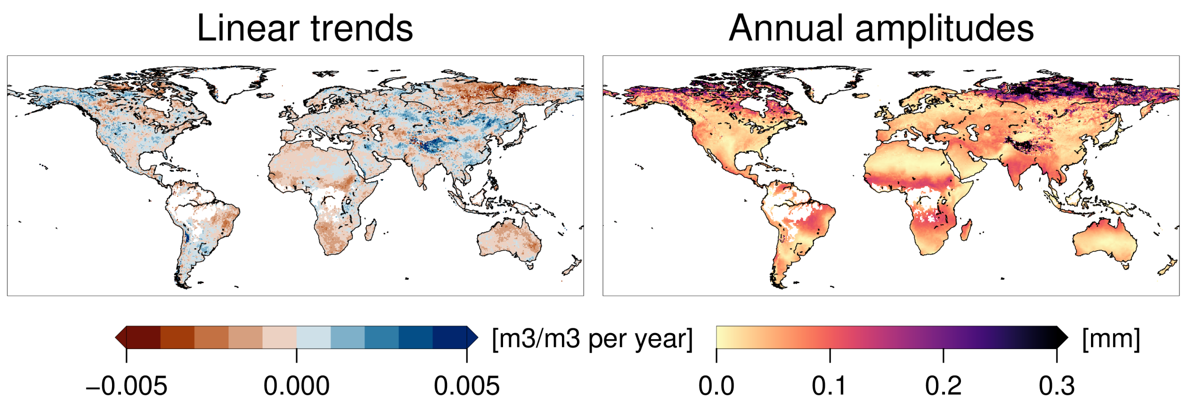


Figure A.13: Linear trends [$m^3/m^3$/year] (left) and annual amplitudes [$m^3/m^3$] (right) of soil moisture derived from ESA-CCI.

Appendix B

# Additional results drought monitoring

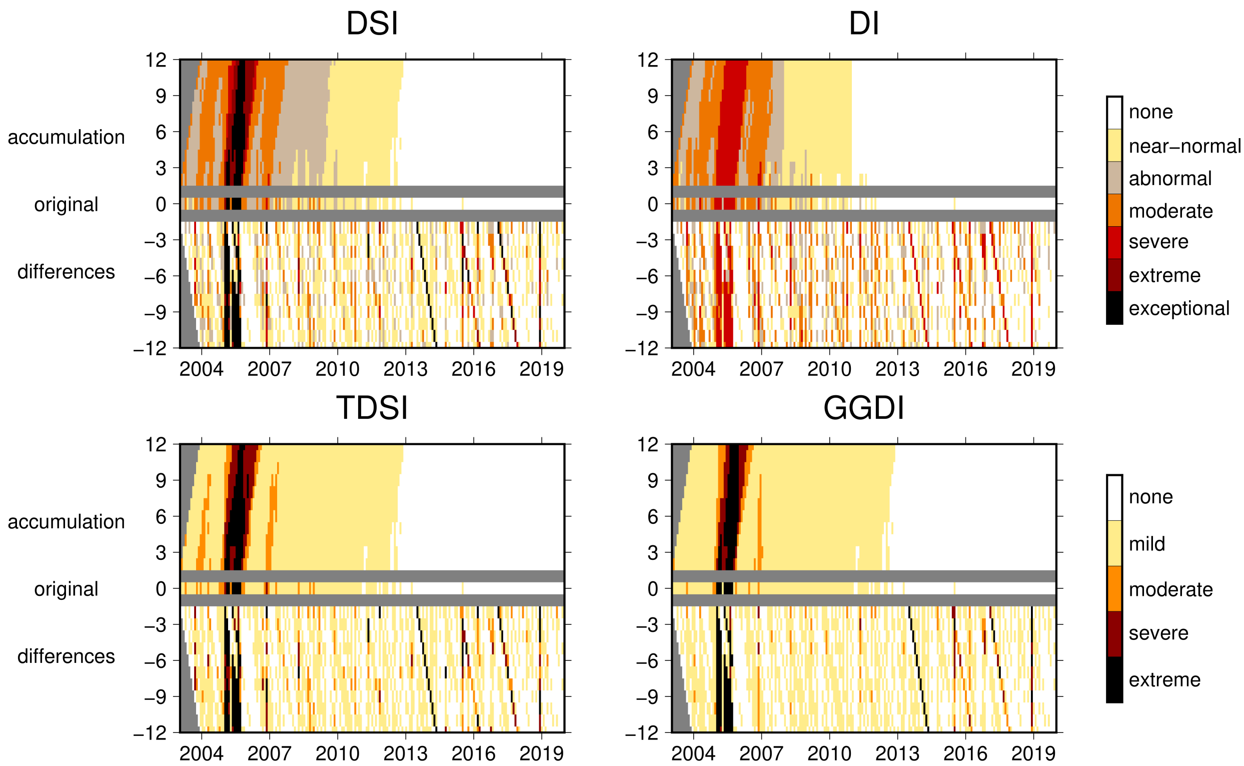


Figure B.1: Synthetic TWSA-DSI, -DI, -TSDI, and -GGDI and their variants with accumulation or differences computed for spatially averaged TWSA for for the Southern Brazil cluster but in contrast to Fig. 7.4, the acceleration and trend are simulated to be positive instead of negative. The y-axis describes which variant of the respective index is used: Zero is the original index, positive values describe the indices based on an accumulation period from two to 12 months (e.g., DSIA12) and negative values describe the indices based on a differencing period (e.g., DSID12). The index values are unitless.

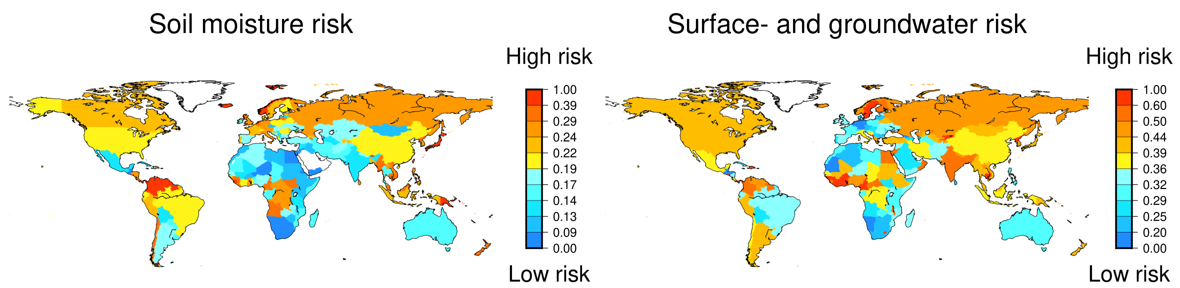


Figure B.2: Drought hazard risk for long-term (9 months) risk for soil moisture and for the combination of surface and groundwater from GLWS assimilation averaged per nation. The drought risk values are standardized and therefore unitless. The colorbar ticks are chosen according to the decimal quantiles of the respective drought hazard risk assessment.

# Appendix C

# Glossary

# Appendix D

# Acronyms

| | |
|---|---|
| **AET** | Actual Evapotranspiration |
| **CCI** | Climate Change Initiative |
| **CDF** | Cumulative Distribution Function |
| **CLSM** | Catchment Land Surface Model |
| **CSR** | Center for Space Research |
| **DI** | Drought Indicator |
| **DLR** | German Aerospace Center |
| **DSI** | Drought Severity Index |
| **EDO** | European Drought Observatory |
| **EOF** | Empirical Orthogonal Function |
| **EnKF** | Ensemble Kalman Filter |
| **ESA** | European Space Agency |
| **ESTKF** | Error Subspace Transform Kalman Filter |
| **EM** | Expectation-Maximization |
| **GCOS** | Global Climate Observing System |
| **GDO** | Global Drought Observatory |
| **GFZ** | GeoForschungsZentrum |
| **GHM** | Global Hydrology Model |
| **GSFC** | Goddard Space Flight Center |
| **GGDI** | GRACE Groundwater Drought Index |
| **GIA** | Glacial Isostatic Adjustment |
| **GLWS** | Global Land Water Storage |
| **GLDAS** | Global Land Data Assimilation System |
| **GGMN** | Global Groundwater Monitoring Network |
| **GNSS** | Global Navigation Satellite System |
| **GRACE** | Gravity Recovery And Climate Experiment |
| **GRACE-FO** | GRACE Follow-On |
| **GRACE/-FO** | GRACE and GRACE/-FO |
| **GWSWUSE** | Groundwater and Surface Water USE |
| **IGRAC** | International Groundwater Resources Assessment Centre |
| **ICA** | Independent Component Analysis |
| **JPL** | Jet Propulsion Laboratory |
| **KGE** | Kling-Gupta Efficiency |
| **PC** | Principal Component |
| **PCA** | Principal Component Analysis |
| **LAI** | Leaf Area Index |
| **LSM** | Land Surface Model |
| **MAGIC** | Mass-Change and Geosciences International Constellation |

| | |
|---|---|
| **mascon** | Mass Concentration block |
| **NSE** | Nash-Sutcliffe efficiency |
| **NASA** | National Aeronautics and Space Administration |
| **NDVI** | Normalized Difference Vegetation Index |
| **PDAF** | Parallel Data Assimilation Framework |
| **PDF** | Probability Distribution Function |
| **PDSI** | Palmer Drought Severity Index |
| **RECOG-LR** | REgional COrrections for GRACE for Lakes and Reservoirs |
| **RMS** | Root Mean Square |
| **RMSD** | Root Mean Square Difference |
| **SDG** | Sustainable Development Goals |
| **SHC** | Spherical Harmonic Coefficients |
| **SMAP** | Soil Moisture Active Passive |
| **SMDI** | Soil Moisture Deficit Index |
| **SPEI** | Standardized Precipitation Evapotranspiration Index |
| **SPI** | Standardized Precipitation Index |
| **SSI** | Standardized Soil Moisture Index |
| **SWOT** | Surface Water and Ocean Topography |
| **TRMM** | Tropical Rainfall Measuring Mission |
| **TSDI** | Total Storage Deficit Index |
| **TWS** | Total Water Storage |
| **TWSA** | Total Water Storage Anomalies |
| **W3RA** | World-Wide Water Resources Assessment system |
| **WMO** | World Meteorological Organization |
| **WaterGAP** | Water Global Assessment and Prognosis |
| **WGHM** | WaterGAP Global Hydrology Model |

# List of Figures

1.1 Illustration of drought propagation from precipitation to water storages . . . . . 4

2.1 GRACE/-FO mission concept . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
2.2 GRACE/-FO data levels . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
2.3 GRACE/-FO TWSA in May 2010 filtered with DDK1, DDK3 and DDK5 . . . . 21
2.4 Range of linear trends of GIA models . . . . . . . . . . . . . . . . . . . . . . . . 22
2.5 Illustration of lake leakage . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 23
2.6 Linear trends for the lake and reservoir removal and relocation of RECOG-LR . 25
2.7 GRACE/-FO TWSA with and without earthquake correction in West Malaysia and Japan on 0.5° . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 26

4.1 Model structure of WaterGAP . . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
4.2 Simulated linear trends of surface water and groundwater abstraction . . . . . . 47
4.3 Fractions of WGHM surface water bodies . . . . . . . . . . . . . . . . . . . . . . 49
4.4 WGHM land cover classes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 52

5.1 PDAF offline and online mode . . . . . . . . . . . . . . . . . . . . . . . . . . . . 61
5.2 Mean condition numbers for GRACE/-FO variance-covariance matrices . . . . . 63
5.3 Illustration of localization radius and weight . . . . . . . . . . . . . . . . . . . . 69

6.1 Basin outlines and biomes . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78
6.2 Linear TWSA trends with and without lake and reservoir relocation from GRACE/-FO and assimilation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 80
6.3 Rescaling factors from six different hydrological models . . . . . . . . . . . . . . 81
6.4 Differeneces between rescaled and not rescaled TWSA from GRACE/-FO or assimilation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 83
6.5 TWSA with and without earthquake correction in West Malaysia for 4° GRACE/-FO and 0.5° assimilation . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
6.6 Linear trend differences for TWSA with and without earthquake correction . . . 84
6.7 TWSA trend differences for EnKF versus ESTKF assimilation . . . . . . . . . . 86
6.8 Diagonal and full GRACE variance-covariance matrix . . . . . . . . . . . . . . . 86
6.9 Linear trends derived from assimilation with EnKF and full observation error matrix 87
6.10 Global mean standard deviation for varying localization settings with ESTKF . . 88
6.11 Global mean standard deviation for small and large parameter ranges and for varying forgetting factors . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 90
6.12 Change in RMS for multiple settings of the assimilation . . . . . . . . . . . . . . 91
6.13 Linear trends and RMS of TWSA from GRACE/-FO, WaterGAP, and GLWS . . 93
6.14 Linear trends, annual amplitudes, and annual phases of TWSA, soil moisture, surface water, and groundwater from GLWS assimilation . . . . . . . . . . . . . . 95
6.15 Dominant storage for linear trends and annual amplitudes of GLWS assimilation 96
6.16 First three dominant PCA modes of TWSA from GLWS assimilation . . . . . . . 97
6.17 Temporal mean degree variance for WaterGAP, GRACE/-FO and GLWS . . . . 99
6.18 Correlations of SHCs from GLWS with WaterGAP or GRACE/-FO . . . . . . . 100

6.19 Correlations between SHCs of GLWS with WaterGAP or with GRACE/-FO averaged per degree and order . . . 101
6.20 Spherical harmonic coefficients $c_{20}$ and $c_{30}$ for GLWS, WaterGAP, and GRACE/-FO 101
6.21 Percentage of months of GLWS TWSA with significant standard deviation . . . . 102
6.22 Standard deviations of linear trends in water storages from GLWS assimilation . 103
6.23 Annual phases of precipitation relative to storages either for model simulations or assimilation . . . 105
6.24 Lags of precipitation relative to soil moisture, surface water, or groundwater for the residual part of the storage data . . . 106
6.25 Auto-regressive process coefficients for the water storages to previous-month storage or to same-month precipitation for the non-seasonal part of the data . . . 108
6.26 Lags of surface water, soil moisture or groundwater from assimilation relative to LAI 111
6.27 Auto-regressive process coefficients for vegetation to previous-month vegetation or to same-month storage . . . 112
6.28 Correlations and RMSD of vertical displacements from GNSS with GRACE/-FO, GLWS, and WaterGAP . . . 114
6.29 Linear trend differences of surface water from RECOG-LR with GLWS, and WaterGAP . . . 119
6.30 Difference between the correlation of normalized soil moisture from ESA-CCI with GLWS and ESA-CCI with WaterGAP . . . 120

7.1 Illustration of the synthetically generated TWSA . . . 122
7.2 Cluster derived from EM-clustering of auto-regressive process coefficients from real GRACE/-FO TWSA . . . 124
7.3 Synthetic TWSA for the Southern Brazil and Southern Africa clusters . . . 125
7.4 Synthetic TWSA-DSI, -DI, -TDSI, and -GGDI and their variants with accumulation or differences computed for the Southern Brazil cluster . . . 126
7.5 Synthetic TWSA-DSI for two different spatial averaging methods for Southern Brazil and Southern Africa clusters . . . 128
7.6 Number of months with extreme or exceptional drought in 2010 using TWSA-DSI from assimilation . . . 131
7.7 Number of months with extreme or exceptional drought in 2003 or 2019 using TWSA-DSI from assimilation with or without trend removal . . . 132
7.8 DSI computed by using spatially averaged TWSA from assimilation for the Congo basin and Ganges basin . . . 133
7.9 Spatial maps of the TWSA-DSI severity classes for the Congo basin computed from assimilation for each month in 2005 . . . 135
7.10 Percentage area of DSI severity classes and spatial maps for using soil moisture, surface water, and groundwater from assimilation for the Congo basin . . . 136
7.11 Percentage of ensemble members per cell where the TWSA-DSI shows extreme or exceptional drought in the Congo basin in December 2005 . . . 138
7.12 Illustration of warning alert levels for a precipitation-soil moisture-vegetation drought warning system . . . 141
7.13 Duration, deficit, and timing of the longest consecutive drought per grid estimated from TWSA-DSI for observations or assimilation . . . 143
7.14 Drought hazard risk for long-term (9 months) risk for soil moisture and the combination of surface and groundwater from assimilation averaged per basin . . 144
7.15 Drought hazard risk for long-term (9 months) risk of soil moisture versus the combination of surface and groundwater from assimilation . . . 146
7.16 Warning alerts for the two prototype warning alert systems SPI, soil moisture DSI, and LAI or SPI, soil moisture DSI, and surface water/groundwater DSI for each of the 140 basins plotted across time . . . 148

7.17 Warning alerts for the precipitation-soil moisture-surface-/groundwater prototype warning alert system in South East Asia . . . 150
7.18 Location and designation of the basins Ganges, Godavari, Mahandi, Mekong, and Salween . . . 151

A.1 TWSA with and without earthquake correction in Japan for 4° GRACE/-FO and 0.5° assimilation . . . i
A.2 TWSA with and without earthquake correction and rescaling in West Malaysia for 4° GRACE/-FO and 0.5° assimilation . . . ii
A.3 Linear trends, annual amplitudes, and annual phases of TWSA, soil moisture, surface water, and groundwater from GLWS release 2 assimilation . . . ii
A.4 Power spectral density of first three dominant PCs of TWSA from GLWS assimilation iii
A.5 First three dominant PCA modes of soil moisture from GLWS . . . iv
A.6 First three dominant PCA modes of surface water from GLWS . . . v
A.7 First three dominant PCA modes of groundwater from GLWS . . . vi
A.8 Standard deviations of TWSA linear trends from GLWS assimilation zoomed into West Africa . . . vi
A.9 Annual phases of soil moisture, surface water, or groundwater relative to LAI either for model simulations or GLWS release 3 assimilation . . . vii
A.10 Annual phases of precipitation relative to storages for assimilation from GLWS release 2 . . . viii
A.11 Annual phases of soil moisture, surface water, or groundwater relative to LAI for assimilation from GLWS release 2 . . . ix
A.12 Correlations and RMSD of vertical displacements from GNSS with GRACE/-FO, GLWS release 2, and WaterGAP . . . x
A.13 Linear trends and annual amplitudes of soil moisture from ESA-CCI . . . x

B.1 Synthetic TWSA-DSI, -DI, -TDSI, and -GGDI and their variants with accumulation or differences computed for the Southern Brazil clusters but with positive trend and acceleration . . . xii
B.2 Drought risk for long-term (9 months) risk for soil moisture and the combination of surface and groundwater from assimilation averaged per nation . . . xii

# List of Tables

2.1 Percentage of months where either CSR or GFZ TWSA fit into the error band of the ITSG TWSA . . . 28

3.1 Drought severity level of the DSI . . . 39
3.2 Drought severity level of the DI . . . 39
3.3 Drought severity level of the TDSI . . . 40

4.1 WaterGAP water storage limits . . . 50

5.1 Global mean RMSD between GRACE/-FO and WaterGAP TWSA for different reference periods . . . 61
5.2 Model parameters from WaterGAP, ranges for perturbation, and distributions . . . 65

6.1 Settings for GLWS releases 2 and options that are step-wise tested . . . 79
6.2 Settings for GLWS releases 2 and 3 . . . 92
6.3 Global mean hydrological signatures of GLWS releases 2 and 3 . . . 96
6.4 Percentage variance and signature assignment of the first three PCA and ICA modes for soil moisture, surface water, and groundwater . . . 98
6.5 Percentage of clusters per water storage of the assimilation that experience an annual relative phase towards LAI that is longer, shorter, or equally long as the model simulation. . . . 109
6.6 Spatially median metrics in East Brazil and six basins in France for comparing groundwater anomalies from in-situ observations with the assimilation, or with the model simulations . . . 115
6.7 Spatially median trends and amplitudes in East Brazil and six basins in France for comparing groundwater anomalies from in-situ observations with the assimilation, or with the model simulations . . . 115
6.8 Correlation and cross-correlation for in-situ groundwater with assimilated or modeled groundwater in six French basins . . . 117
6.9 Global median linear trends and amplitudes for lakes and reservoirs for comparing surface water from remote-sensed observations with the assimilation, or with the model simulations . . . 118

7.1 Considered thresholds, minimum duration, and maximum pause for extracting consecutive drought events from drought indices. . . . 140
7.2 Top 15 river basins for the longest drought events either for the soil moisture drought hazard risk or the surface- and groundwater drought hazard risk from assimilation . . . 145

# Bibliography

1. A, G., Wahr, J. & Zhong, S. Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada. *Geophysical Journal International* **192,** 557–572. ISSN: 0956-540X. doi:10.1093/gji/ggs030 (2012).
2. Abaza, M., Garneau, C. & Anctil, F. Comparison of sequential and variational streamflow assimilation techniques for short-term hydrological forecasting. *Journal of Hydrologic Engineering* **20,** 04014042. doi:10.1061/(ASCE)HE.1943-5584.0001013 (2015).
3. Abelen, S., Seitz, F., Abarca-del-Rio, R. & Güntner, A. Droughts and floods in the La Plata Basin in soil Moisture data and GRACE. *Remote Sensing* **7,** 7324–7349. ISSN: 2072-4292. doi:10.3390/rs70607324 (2015).
4. Akaike, H. in *Selected Papers of Hirotugu Akaike* 199–213 (Springer New York, New York, 1998). ISBN: 978-1-4612-1694-0. doi:10.1007/978-1-4612-1694-0_15.
5. Alley, W. M. The Palmer Drought Severity Index: limitations and assumptions. *Journal of Climate and Applied Meteorology* **23,** 1100–1109. ISSN: 07333021, 21635366. http://www.jstor.org/stable/26181407 (1984).
6. Andersen, O. B., Seneviratne, S. I., Hinderer, J. & Viterbo, P. GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave. *Geophysical Research Letters* **32.** doi:10.1029/2005GL023574 (2005).
7. Anderson, J. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic meteorology and oceanography* **61,** 72–83. doi:10.1111/j.1600-0870.2008.00361.x (2009).
8. Anderson, J. L. An ensemble adjustment Kalman Filter for data assimilation. *Monthly Weather Review* **129,** 2884–2903. doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2 (2001).
9. Azmi, M., Rüdiger, C. & Walker, J. P. A data fusion-based drought index. *Water Resources Research* **52,** 2222–2239. doi:10.1002/2015WR017834 (2016).
10. Barker, L. J., Hannaford, J., Chiverton, A. & Svensson, C. From meteorological to hydrological drought using standardised indicators. *Hydrology and Earth System Sciences* **20,** 2483–2505. doi:10.5194/hess-20-2483-2016 (2016).
11. Baur, O., Kuhn, M. & Featherstone, W. GRACE-derived ice-mass variations over Greenland by accounting for leakage effects. *Journal of Geophysical Research: Solid Earth* **114.** doi:10.1029/2008JB006239 (2009).
12. Bazrafshan, O., Zamani, H. & Shekari, M. A copula-based index for drought analysis in arid and semi-arid regions of Iran. *Natural Resource Modeling* **33.** doi:10.1111/nrm.12237 (2020).

13. Bera, A., Baranval, N. K., Kumar, R. & Pal, S. K. Groundwater drought risk assessment in the semi-arid Kansai river basin, West Bengal, India using SWAT and machine learning models. *Groundwater for Sustainable Development* **26,** 101254. doi:10.1016/j.gsd.2024.101254 (2024).

14. Bergman, K., Sabol, P. & Miskus, D. *Experimental indices for monitoring global drought conditions* in *Proceedings of the 13th annual climate diagnostics workshop, Cambridge, MA, USA* **31** (1988), 190–197.

15. Bhuiyan, C. *Various drought indices for monitoring drought condition in Aravalli terrain of India* in *Proceedings of the XXth ISPRS Congress, Istanbul, Turkey* **2004** (2004), 12–23. https://www.isprs.org/proceedings/xxxv/congress/comm7/papers/243.pdf.

16. Bijaber, N., El Hadani, D., Saidi, M., Svoboda, M. D., Wardlow, B. D., Hain, C. R., Poulsen, C. C., Yessef, M. & Rochdi, A. Developing a remotely sensed drought monitoring indicator for Morocco. *Geosciences* **8,** 55. doi:10.3390/geosciences8020055 (2018).

17. Bishop, C. H., Etherton, B. J. & Majumdar, S. J. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Monthly weather review* **129,** 420–436. doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2 (2001).

18. Boergens, E., Dobslaw, H. & Dill, R. *GFZ GravIS RL06 continental water storage anomalies* 2019. doi:10.5880/GFZ.GRAVIS_06_L3_TWS.

19. Boergens, E., Güntner, A., Dobslaw, H. & Dahle, C. Quantifying the Central European droughts in 2018 and 2019 with GRACE Follow-On. *Geophysical Research Letters* **47,** e2020GL087285. doi:10.1029/2020GL087285 (2020).

20. Bonaccorso, B., Bordi, I., Cancelliere, A., Rossi, G. & Sutera, A. Spatial variability of drought: an analysis of the SPI in Sicily. *Water resources management* **17,** 273–296. doi:10.1023/A:1024716530289 (2003).

21. Broerse, D. *Megathrust earthquakes: Study of fault slip and stress relaxation using satellite gravity observations* PhD thesis (Delft University of Technology, 2014). doi:10.4233/uuid:3e46f5b1-1887-4c7c-9d5e-9a6a56126ebf.

22. Buchhorn, M. Copernicus Global Land Service: Global biome cluster layer for the 100m global land cover processing line. *Zenodo.* doi:10.5281/zenodo.5848609 (2022).

23. Byun, H.-R. & Wilhite, D. A. Daily quantification of drought severity and duration. *Journal of Climate* **5,** 2747–2756. doi:10.1175/1520-0442(1999)012<2747:DQODSA>2.0.CO;2 (1996).

24. Cardona, O. D., Van Aalst, M. K., Birkmann, J., Fordham, M., Mc Gregor, G., Rosa, P., Pulwarty, R. S., Schipper, E. L. F., Sinh, B. T., Décamps, H., *et al.* in *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change* 65–108 (Cambridge University Press, 2012).

25. Cardoso, J.-F. & Souloumiac, A. *Blind beamforming for non-Gaussian signals* in *IEE proceedings F (radar and signal processing)* **140** (1993), 362–370. doi:10.1049/ip-f-2.1993.0054.

26. Caron, L., Ivins, E., Larour, E., Adhikari, S., Nilsson, J. & Blewitt, G. GIA model statistics for GRACE hydrology, cryosphere, and ocean science. *Geophysical Research Letters* **45,** 2203–2212. doi:10.1002/2017GL076644 (2018).

27. Carrão, H., Naumann, G. & Barbosa, P. Mapping global patterns of drought risk: An empirical framework based on sub-national estimates of hazard, exposure and vulnerability. *Global Environmental Change* **39,** 108–124. doi:10.1016/j.gloenvcha.2016.04.012 (2016).

28. Carrão, H., Singleton, A., Naumann, G., Barbosa, P. & Vogt, J. V. An optimized system for the classification of meteorological drought intensity with applications in drought frequency analysis. *Journal of Applied Meteorology and Climatology* **53,** 1943–1960. doi:10.1175/JAMC-D-13-0167.1 (2014).

29. Case, K., Kruizinga, G. & Wu, S.-C. *GRACE Level 1B Data Product User Handbook* tech. rep. (Jet Propulsion Laboratory, California Institute of Technology, 2010).

30. Casella, G. & Berger, R. L. *Statistical Inference* 199–213. ISBN: 0-534-24312-6 (Duxbury/Thomson Learning, 2002).

31. Cerón, W. L., Molina-Carpio, J., Ayes Rivera, I., Andreoli, R. V., Kayano, M. T. & Canchala, T. A principal component analysis approach to assess CHIRPS precipitation dataset for the study of climate variability of the La Plata Basin, Southern South America. *Natural Hazards* **103,** 767–783. doi:10.1007/s11069-020-04011-x (2020).

32. Chen, J., Wilson, C., Blankenship, D. & Tapley, B. Antarctic mass rates from GRACE. *Geophysical research letters* **33.** doi:10.1029/2006GL026369 (2006).

33. Chen, J., Wilson, C., Famiglietti, J. & Rodell, M. Spatial sensitivity of the Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations. *Journal of Geophysical Research: Solid Earth* **110.** doi:10.1029/2004JB003536 (2005).

34. Chen, M. *EM Algorithm for Gaussian Mixture Model (EM GMM)* program. 2018. https://www.mathworks.com/matlabcentral/fileexchange/26184-em-algorithm-for-gaussian-mixture-model-em-gmm.

35. Cheng, M., Ries, J. C. & Tapley, B. D. Variations of the Earth's figure axis from satellite laser ranging and GRACE. *Journal of Geophysical Research: Solid Earth* **116,** 148–227. doi:10.1029/2010JB000850 (2011).

36. Cheng, M., Tapley, B. D. & Ries, J. C. Deceleration in the Earth's oblateness. *Journal of Geophysical Research: Solid Earth* **118,** 740–747. doi:10.1002/jgrb.50058 (2013).

37. Cobbing, J. Groundwater and the discourse of shortage in Sub-Saharan Africa. *Hydrogeology Journal* **28,** 1143–1154. doi:10.1007/s10040-020-02147-5 (2020).

38. Colombo, O. L. *Numerical methods for harmonic analysis on the sphere* (Department of Geodetic Science, The Ohio State University, 1981).

39. CSC. *Climate Change Scenarios for the Congo Basin* tech. rep. (Climate Service Centre Report No. 11, Hamburg, Germany, 2013). https://www.climate-service-center.de/imperia/md/content/csc/csc-report11_optimized.pdf.

40. Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H. & Buontempo, C. WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data* **12,** 2097–2120. doi:10.5194/essd-12-2097-2020 (2020).

41. Dai, Y., Zeng, X., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G., *et al.* The common land model. *Bulletin of the American Meteorological Society* **84,** 1013–1024. doi:10.1175/BAMS-84-8-1013 (2003).

42. Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., *et al.* The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society* **137,** 553–597. doi:10.1002/qj.828 (2011).

43. Deggim, S., Eicker, A., Schawohl, L., Ellenbeck, L., Dettmering, D., Schwatke, C., Mayr, S. & Klein, I. *RECOG-LR RL01: Correcting GRACE total water storage estimates for global lakes and reservoirs* data set. 2020. doi:10.1594/PANGAEA.921851.

44. Deggim, S., Eicker, A., Schawohl, L., Gerdener, H., Schulze, K., Engels, O., Kusche, J., Saraswati, A. T., van Dam, T., Ellenbeck, L., *et al.* RECOG RL01: correcting GRACE total water storage estimates for global lakes/reservoirs and earthquakes. *Earth System Science Data* **13,** 2227–2244. doi:10.5194/essd-13-2227-2021 (2021).

45. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)* **39,** 1–22. doi:10.1111/j.2517-6161.1977.tb01600.x (1977).

46. Deng, Y., Wang, S., Bai, X., Luo, G., Wu, L., Cao, Y., Li, H., Li, C., Yang, Y., Hu, Z., *et al.* Variation trend of global soil moisture and its cause analysis. *Ecological Indicators* **110,** 105939. doi:10.1016/j.ecolind.2019.105939 (2020).

47. Dharpure, J. K., Goswami, A., Patel, A., Kulkarni, A. V. & Meloth, T. Drought characterization using the Combined Terrestrial Evapotranspiration Index over the Indus, Ganga and Brahmaputra river basins. *Geocarto International* **37,** 1059–1083. doi:10.1080/10106049.2020.1756462 (2022).

48. Dibi-Anoh, P. A., Koné, M., Gerdener, H., Kusche, J. & N'Da, C. K. Hydrometeorological extreme events in West Africa: droughts. *Surveys in Geophysics* **44,** 173–195. doi:10.1007/s10712-022-09748-7 (2023).

49. Dobslaw, H., Dill, R., Bagge, M., Klemann, V., Boergens, E., Thomas, M., Dahle, C. & Flechtner, F. Gravitationally consistent mean barystatic sea level rise from leakage-corrected monthly GRACE data. *Journal of Geophysical Research: Solid Earth* **125.** doi:10.1029/2020JB020923 (2020).

50. Döll, P., Hasan, H. M. M., Schulze, K., Gerdener, H., Börger, L., Shadkam, S., Ackermann, S., Hosseini-Moghari, S.-M., Müller Schmied, H., Güntner, A. & Kusche, J. Leveraging multi-variable observations to reduce and quantify the output uncertainty of a global hydrological model: evaluation of three ensemble-based approaches for the Mississippi River basin. *Hydrology and Earth System Sciences* **28,** 2259–2295. doi:10.5194/hess-28-2259-2024 (2024).

51. Döll, P., Douville, H., Güntner, A., Müller Schmied, H. & Wada, Y. Modelling freshwater resources at the global scale: challenges and prospects. *Surveys in Geophysics* **37,** 195–221. doi:10.1007/s10712-015-9343-1 (2016).

52. Döll, P. & Fiedler, K. Global-scale modeling of groundwater recharge. *Hydrology and Earth System Sciences* **12,** 863–885. doi:10.5194/hess-12-863-2008 (2008).

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

54. Döll, P., Müller Schmied, H., Schuh, C., Portmann, F. T. & Eicker, A. Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research* **50,** 5698–5720. doi:10.1002/2014WR015595 (2014).

55. Döll, P. & Siebert, S. Global modeling of irrigation water requirements. *Water resources research* **38,** 8.1–8.10. doi:10.1029/2001WR000355 (2002).

56. Dorigo, W., Preimesberger, W., Hahn, S., Van der Schalie, R., De Jeu, R., Rodriguez-Fernandez, N., Hirschi, M., Stradiotti, P., Frederikse, T., Gruber, A. & Madelon, R. ESA soil moisture Climate Change Initiative (Soil_Moisture_cci): Version 08.1 data collection. *NERC EDS Centre for Environmental Data Analysis.* http://catalogue.ceda.ac.uk/uuid/ff890589c21f4033803aa550f52c980c/ (2023).

57. Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., *et al.* 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 (2017).

58. Dracup, J. A., Lee, K. S. & Paulson Jr, E. G. On the definition of droughts. *Water resources research* **16,** 297–302. doi:10.1029/WR016i002p00297 (1980).

59. Dutra, E., Magnusson, L., Wetterhall, F., Cloke, H. L., Balsamo, G., Boussetta, S. & Pappenberger, F. The 2010–2011 drought in the Horn of Africa in ECMWF reanalysis and seasonal forecast products. *International Journal of Climatology* **33,** 1720–1729. doi:10.1002/joc.3545 (2013).

60. Ehrendorfer, M. A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift* **16,** 795–818. doi:10.1127/0941-2948/2007/0256 (2007).

61. Eicker, A., Forootan, E., Springer, A., Longuevergne, L. & Kusche, J. Does GRACE see the terrestrial water cycle "intensifying"? *Journal of Geophysical Research: Atmospheres* **121,** 733–745. doi:10.1002/2015JD023808 (2016).

62. Eicker, A., Schumacher, M., Kusche, J., Döll, P. & Schmied, H. M. Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an Ensemble Kalman filter: First results. *Surveys in Geophysics* **35,** 1285–1309. doi:10.1007/s10712-014-9309-8 (2014).

63. Einarsson, I., Hoechner, A., Wang, R. & Kusche, J. Gravity changes due to the Sumatra-Andaman and Nias earthquakes as detected by the GRACE satellites: a reexamination. *Geophysical Journal International* **183,** 733–747. doi:10.1111/j.1365-246X.2010.04756.x (2010).

64. Einarsson, I. *Sensitivity analysis for future gravity satellite missions* PhD thesis (Deutsches GeoForschungsZentrum GFZ Potsdam, 2011), 160. doi:10.2312/GFZ.b103-11107.

65. Ek, M., Mitchell, K., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. & Tarpley, J. Implementation of Noah land-surface model advances in the NCEP operational mesoscale Eta model. *J Geophys Res* **108.** doi:10.1029/2002JD003296 (2003).

66. Elameen, A. M., Jin, S. & Olago, D. Identification of drought events in major basins of Africa from GRACE total water storage and modeled products. *Photogrammetric Engineering & Remote Sensing* **89,** 221–232. doi:10.14358/PERS.22-00092R2 (2023).

67. Elsaka, B., Forootan, E. & Alothman, A. Improving the recovery of monthly regional water storage using one year simulated observations of two pairs of GRACE-type satellite gravimetry constellation. *Journal of Applied Geophysics* **109,** 195–209. ISSN: 0926-9851. doi:10.1016/j.jappgeo.2014.07.026. https://www.sciencedirect.com/science/article/pii/S0926985114002389 (2014).

68. Erfanian, A., Wang, G. & Fomenko, L. Unprecedented drought over tropical South America in 2016: significantly under-predicted by tropical SST. *Scientific reports* **7,** 1–11. doi:10.1038/s41598-017-05373-2 (2017).

69. Escorihuela, M. J. & Quintana-Seguí, P. Comparison of remote sensing and simulated soil moisture datasets in Mediterranean landscapes. *Remote Sensing of Environment* **180.** Special Issue: ESA's Soil Moisture and Ocean Salinity Mission - Achievements and Applications, 99–114. ISSN: 0034-4257. doi:https://doi.org/10.1016/j.rse.2016.02.046. https://www.sciencedirect.com/science/article/pii/S0034425716300748 (2016).

70. Evensen, G. Open boundary conditions for the Extended Kalman filter with a quasi-geostrophic ocean model. *Journal of Geophysical Research: Oceans* **98,** 16529–16546. doi:10.1029/93JC01365 (1993).

71. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* **99,** 10143–10162. doi:10.1029/94JC00572 (1994).

72. Evensen, G. The Ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean dynamics* **53,** 343–367. doi:10.1007/s10236-003-0036-9 (2003).

73. Evensen, G. *et al. Data assimilation: the Ensemble Kalman filter* doi:10.1007/978-3-642-03711-5 (Springer, 2009).

74. Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K., Syed, T. H., Swenson, S. C., de Linage, C. R. & Rodell, M. Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters* **38.** doi:10.1029/2010GL046442 (2011).

75. Fayiah, M., Tulcan, R. X. S., Bockarie, R., Saccoh, S., Singh, S. & Fayiah, M. S. Sierra Leone rainfall and drought dynamics from 1990-2020: a holistic perspective. *World Journal of Environmental Biosciences* **11,** 34–42. doi:10.51847/8Wz28ID8Mn (2022).

76. Felsberg, A., De Lannoy, G. J., Girotto, M., Poesen, J., Reichle, R. H. & Stanley, T. Global soil water estimates as landslide predictor: the effectiveness of SMOS, SMAP, and GRACE observations, land surface simulations, and data assimilation. *Journal of Hydrometeorology* **22,** 1065–1084. doi:10.1175/JHM-D-20-0228.1 (2021).

77. Fengler, M. J., Freeden, W., Kohlhaas, A., Michel, V. & Peters, T. Wavelet modeling of regional and temporal variations of the Earth's gravitational potential observed by GRACE. *Journal of Geodesy* **81,** 5–15. doi:10.1007/s00190-006-0040-1 (2007).

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

79. Food and Agriculture Organization of the United Nations. *2017: the impact of disasters and crises on agriculture and food security.* tech. rep. (2021), 182.

80. Food and Agriculture Organization of the United Nations. *The state of food and agriculture 2021: making agrifood systems more resilient to shocks and stresses* 2021. doi:10.4060/cb4476en.

81. Forootan, E. *Statistical signal decomposition techniques for analyzing time-variable satellite gravimetry data* PhD thesis (Universität Bonn, Institut für Geodäsie und Geoinformation, 2014). doi:20.500.11811/1452.

82. Forootan, E., Awange, J., Schumacher, M., Anyah, R., Van Dijk, A., Kusche, J., *et al.* Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over Australia. *Remote sensing of Environment* **172,** 50–66. doi:10.1016/j.rse.2015.10.027 (2016).

83. Forootan, E. & Kusche, J. Separation of global time-variable gravity signals into maximally independent components. *Journal of Geodesy* **86,** 477–497. doi:10.1007/s00190-011-0532-5 (2012).

84. Forootan, E., Mehrnegar, N., Schumacher, M., Schiettekatte, L. A. R., Jagdhuber, T., Farzaneh, S., van Dijk, A. I., Shamsudduha, M. & Shum, C. Global groundwater droughts are more severe than they appear in hydrological models: An investigation through a Bayesian merging of GRACE and GRACE-FO data with a water balance model. *Science of the Total Environment* **912,** 169476. doi:10.1016/j.scitotenv.2023.169476 (2024).

85. Frappart, F., Papa, F., da Silva, J. S., Ramillien, G., Prigent, C., Seyler, F. & Calmant, S. Surface freshwater storage and dynamics in the Amazon basin during the 2005 exceptional drought. *Environmental Research Letters* **7,** 044010. doi:10.1088/1748-9326/7/4/044010 (2012).

86. Frappart, F., Papa, F., Güntner, A., Tomasella, J., Pfeffer, J., Ramillien, G., Emilio, T., Schietti, J., Seoane, L., da Silva Carvalho, J., *et al.* The spatio-temporal variability of groundwater storage in the Amazon River Basin. *Advances in Water Resources* **124,** 41–52. doi:10.1016/j.advwatres.2018.12.005 (2019).

87. Frappart, F., Ramillien, G. & Ronchail, J. Changes in terrestrial water storage versus rainfall and discharges in the Amazon basin. *International Journal of Climatology* **33,** 3029–3046. doi:10.1002/joc.3647 (2013).

88. Gardner, A. S., Moholdt, G., Cogley, J. G., Wouters, B., Arendt, A. A., Wahr, J., Berthier, E., Hock, R., Pfeffer, W. T., Kaser, G., *et al.* A reconciled estimate of glacier contributions to sea level rise: 2003 to 2009. *science* **340,** 852–857. doi:10.1126/science.1234532 (2013).

89. Gaye, C. B. & Tindimugaya, C. Review: Challenges and opportunities for sustainable groundwater management in Africa. *Hydrogeology Journal* **27,** 1099–1110. doi:10.1007/s10040-018-1892-1 (2019).

90. Gerdener, H., Engels, O. & Kusche, J. A framework for deriving drought indicators from the Gravity Recovery and Climate Experiment (GRACE). *Hydrology and Earth System Sciences* **24,** 227–248. doi:10.5194/hess-24-227-2020 (2020).

91. Gerdener, H., Kusche, J., Schulze, K., Döll, P. & Klos, A. The global land water storage data set release 2 (GLWS2. 0) derived via assimilating GRACE and GRACE-FO data into a global hydrological model. *Journal of Geodesy* **97.** doi:10.1007/s00190-023-01763-9 (2023).

92. Gerdener, H., Kusche, J., Schulze, K., Ghazaryan, G. & Dubovyk, O. Revising precipitation–water storages–vegetation signatures with GRACE-based data assimilation. *Journal of Hydrology* **612,** 128096. doi:10.1016/j.jhydrol.2022.128096 (2022).

93. Gerdener, H., Schulze, K., Engels, O. & Kusche, J. *RECOG-EQ RL01 Earthquake correction for CSR, GFZ and ITSG solutions of GRACE level 3 total water storage anomalies from 2003-01 to 2016-12 correcting for the Sumatra-Andaman (2004) and Tohoku (2011) earthquakes* data set. 2020. doi:10.1594/PANGAEA.921923.

94. Gerdener, H., Schulze, K. & Kusche, J. *GLWS 2.0: A global product that provides total water storage anomalies, groundwater, soil moisture and surface water with a spatial resolution of 0.5° from 2003 to 2019* data set. 2023. doi:10.1594/PANGAEA.954742.

95. Getirana, A. Extreme water deficit in Brazil detected from space. *Journal of Hydrometeorology* **17,** 591–599. doi:10.1175/JHM-D-15-0096.1 (2016).

96. Girotto, M., De Lannoy, G. J., Reichle, R. H. & Rodell, M. Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model. *Water Resources Research* **52,** 4164–4183. doi:10.1002/2015WR018417 (2016).

97. Girotto, M., De Lannoy, G. J., Reichle, R. H., Rodell, M., Draper, C., Bhanja, S. N. & Mukherjee, A. Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India. *Geophysical research letters* **44,** 4107–4115. doi:10.1002/2017GL072994 (2017).

98. Girotto, M., Reichle, R. H., Rodell, M., Liu, Q., Mahanama, S. & De Lannoy, G. J. Multi-sensor assimilation of SMOS brightness temperature and GRACE terrestrial water storage observations for soil moisture and shallow groundwater estimation. *Remote Sensing of Environment* **227,** 12–27. doi:10.1016/j.rse.2019.04.001 (2019).

99. Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G. & Leavitt, B. Remote estimation of leaf area index and green leaf biomass in maize canopies. *Geophysical research letters* **30.** doi:10.1029/2002GL016450 (2003).

100. Glanville, K., Sheldon, F., Butler, D. & Capon, S. Effects and significance of groundwater for vegetation: A systematic review. *Science of the Total Environment* **875,** 162577. doi:10.1016/j.scitotenv.2023.162577 (2023).

101. Gleick, P. H. *The World's Water: The Biennial Report on Freshwater Resources. Volume 8* ISBN: 978-1-61091-483-3. doi:10.5822/978-1-61091-483-3 (Island press Washington, DC, 2014).

102. Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W. & Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth System Science Data* **11,** 717–739. doi:10.5194/essd-11-717-2019 (2019).

103. Gu, X., Zhang, Q., Li, J., Singh, V. P., Liu, J., Sun, P., He, C. & Wu, J. Intensification and expansion of soil moisture drying in warm season over Eurasia under global warming. *Journal of Geophysical Research: Atmospheres* **124,** 3765–3782. doi:10.1029/2018JD029776 (2019).

104. Guevara, M., Taufer, M. & Vargas, R. Gap-free global annual soil moisture: 15 km grids for 1991–2018. *Earth System Science Data Discussions* **13,** 1711–1735. doi:10.5194/essd-13-1711-2021 (2021).

105. Güntner, A. Improvement of global hydrological models using GRACE data. *Surveys in geophysics* **29,** 375–397. doi:10.1007/s10712-008-9038-y (2008).

106. Güntner, A., Gerdener, H., Boergens, E., Kusche, J., Kollet, S., Dobslaw, H., Hartick, C., Sharifi, E. & Flechtner, F. Veränderungen der Wasserspeicherung in Deutschland seit 2002 aus Beobachtungen der Satellitengravimetrie. *Hydrologie und Wasserbewirtschaftung* **67,** 74–89. doi:10.5675/HyWa_2023.2_1 (2023).

107. Gupta, H. V. & Kling, H. On typical range, sensitivity, and normalization of Mean Squared Error and Nash-Sutcliffe Efficiency type metrics. *Water Resources Research* **47.** doi:10.1029/2011WR010962 (2011).

108. Gusyev, M., Hasegawa, A., Magome, J., Kuribayashi, D., Sawano, H. & Lee, S. *Drought assessment in the pampanga river basin, the Philippines. Part 1: A Role of dam Infrastructure in Historical Droughts* in *Proceedings of the 21st International Congress on Modelling and Simulation (MODSIM 2015), Broadbeach, Queensland, Australia* (2015). doi:0.13140/RG.2.1.4931.4321.

109. Hagemann, S., Chen, C., Clark, D. B., Folwell, S., Gosling, S. N., Haddeland, I., Hanasaki, N., Heinke, J., Ludwig, F., Voss, F., *et al.* Climate change impact on available water resources obtained using multiple global climate and hydrology models. *Earth System Dynamics* **4,** 129–144. doi:10.5194/esd-4-129-2013 (2013).

110. Hagenlocher, M., Meza, I., Anderson, C. C., Min, A., Renaud, F. G., Walz, Y., Siebert, S. & Sebesvari, Z. Drought vulnerability and risk assessments: state of the art, persistent gaps, and research agenda. *Environmental Research Letters* **14,** 083002. doi:10.1088/1748-9326/ab225d (2019).

111. Hamill, T. M., Whitaker, J. S. & Snyder, C. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review* **129,** 2776–2790. doi:10.1175/1520-0493(2001)129%3C2776:DDFOBE%3E2.0.CO;2 (2001).

112. Hao, Z. & AghaKouchak, A. A nonparametric multivariate multi-index drought monitoring framework. *Journal of Hydrometeorology* **15,** 89–101. doi:10.1175/JHM-D-12-0160.1 (2014).

113. Hayes, M. J., Svoboda, M. D., Wardlow, B. D., Anderson, M. C. & Kogan, F. Drought monitoring: Historical and current perspectives. *Drought Mitigation Center Faculty Publications.* http://digitalcommons.unl.edu/droughtfacpub/94 (2012).

114. Heim Jr, R. R. A review of twentieth-century drought indices used in the United States. *Bulletin of the American Meteorological Society* **83,** 1149–1166. doi:10.1175/1520-0477-83.8.1149 (2002).

115. Herbert, C. & Döll, P. Analyzing the informative value of alternative hazard indicators for monitoring drought hazard for human water supply and river ecosystems at the global scale. *Natural Hazards and Earth System Sciences* **23,** 2111–2131. doi:10.5194/nhess-23-2111-2023 (2023).

116. Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146,** 1999–2049. doi:10.1002/qj.3803 (2020).

117. Hirabayashi, Y., Kanae, S., Motoya, K., Masuda, K. & Döll, P. A 59-year (1948-2006) global near-surface meteorological data set for land surface models. Part I: Development of daily forcing and assessment of precipitation intensity. *Hydrological Research Letters* **2,** 36–40. doi:10.3178/hrl.2.36 (2008).

118. Hisdal, H., Tallaksen, L. M., Gauster, T., Bloomfield, J. P., Parry, S., Prudhomme, C. & Wanders, N. in *Hydrological Drought (Second Edition)* (eds Tallaksen, L. M. & van Lanen, H. A.) Second Edition, 157–231 (Elsevier, 2024). ISBN: 978-0-12-819082-1. https://doi.org/10.1016/B978-0-12-819082-1.00006-0.

119. Hofmann-Wellenhof, B. & Moritz, H. *Physical geodesy* doi:10.1007/978-3-211-33545-1 (Springer Science & Business Media, 2006).

120. Horvath, A., Murböck, M., Pail, R. & Horwath, M. Decorrelation of GRACE time variable gravity field solutions using full covariance information. *Geosciences* **8,** 323. doi:10.3390/geosciences8090323 (2018).

121. Hosseini-Moghari, S.-M., Araghinejad, S., Tourian, M. J., Ebrahimi, K. & Döll, P. Quantifying the impacts of human water use and climate variations on recent drying of Lake Urmia basin: the value of different sets of spaceborne and in situ data for calibrating a global hydrological model. *Hydrology and Earth System Sciences* **24,** 1939–1956. doi:10.5194/hess-24-1939-2020 (2020).

122. Houborg, R., Rodell, M., Li, B., Reichle, R. & Zaitchik, B. F. Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research* **48.** doi:10.1029/2011WR011291 (2012).

123. Houtekamer, P. L. & Mitchell, H. L. Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review* **126,** 796–811. doi:10.1175/1520-0493(1998)126%3C0796:DAUAEK%3E2.0.CO;2 (1998).

124. Hoylman, Z. H., Bocinsky, R. K. & Jencso, K. G. Drought assessment has been outpaced by climate change: empirical arguments for a paradigm shift. *Nature communications* **13,** 2715. doi:10.1038/s41467-022-30316-5 (2022).

125. Hulsman, P., Savenije, H. H. & Hrachowitz, M. Satellite-based drought analysis in the Zambezi River Basin: Was the 2019 drought the most extreme in several decades as locally perceived? *Journal of Hydrology: Regional Studies* **34,** 100789. doi:10.1016/j.ejrh.2021.100789 (2021).

126. Humphrey, V. & Gudmundsson, L. GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data* **11,** 1153–1170. doi:10.5194/essd-11-1153-2019 (2019).

127. Hyvärinen, A. Survey on Independent Component Analysis. English. *Neural Computing Surveys* **2,** 94–128. https://www.cs.helsinki.fi/u/ahyvarin/papers/NCS99.pdf (1999).

128. Ilk, K. H. *Potentialtheorie* doi:10.1007/978-3-662-62361-9 (Springer, 2021).

129. Ingale, R. Harmonic analysis using FFT and STFT. *International Journal of Signal Processing, Image Processing and Pattern Recognition* **7,** 345–362. doi:0.14257/ijsip.2014.7.4.33 (2014).

130. IPCC. *Climate change 2014 impacts, adaptation, and vulnerability. Part A: global and sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change ed V R Barros et al* tech. rep. (Cambridge University Press, Cambridge, United Kindom and New York, NY, USA, 2014). https://www.ipcc.ch/report/ar5/wg2/.

131. IPCC. *Climate change 2022: impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* tech. rep. (Cambridge University Press, Cambridge, United Kindom and New York, NY, USA, 2022). doi:10.1017/9781009325844. https://www.ipcc.ch/report/ar6/wg2/.

132. Jacob, T., Wahr, J., Pfeffer, W. T. & Swenson, S. Recent contributions of glaciers and ice caps to sea level rise. *Nature* **482,** 514–518. doi:10.1038/nature10847 (2012).

133. Jekeli, C. *Alternative methods to smooth the Earth's gravity field* tech. rep. (1981). https://ntrs.nasa.gov/citations/19820014947.

134. Jensen, L., Gerdener, H., Eicker, A., Kusche, J. & Fiedler, S. Observations indicate regionally misleading wetting and drying trends in CMIP6. *npj Climate and Atmospheric Science, in review* (2024).

135. Joodaki, G., Wahr, J. & Swenson, S. Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research* **50,** 2679–2692. doi:10.1002/2013WR014633 (2014).

136. Kalman, R. E. & Bucy, R. S. New results in linear filtering and prediction theory. *Journal of Basic Engeneering* **83.** doi:10.1115/1.3658902 (1961).

137. Kalnay, E. *Atmospheric modeling, data assimilation and predictability* doi:10.1017/CBO9780511802270 (Cambridge university press, 2003).

138. Keyantash, J. & Dracup, J. A. The quantification of drought: an evaluation of drought indices. *Bulletin of the American Meteorological Society* **83,** 1167–1180. doi:10.1175/1520-0477-83.8.1167 (2002).

139. Khaki, M., Forootan, E., Kuhn, M., Awange, J., Papa, F. & Shum, C. A study of Bangladesh's sub-surface water storages using satellite products and data assimilation scheme. *Science of the Total Environment* **625,** 963–977. doi:10.1016/j.scitotenv.2017.12.289 (2018).

140. Khaki, M., Han, S.-C., Ghobadi-Far, K., Yeo, I.-Y. & Tangdamrongsub, N. Assimilation of GRACE Follow-On inter-satellite laser ranging measurements into Land Surface Models. *Water Resources Research* **59,** e2022WR032432. doi:10.1029/2022WR032432 (2023).

141. Khan, A. A., Pant, N. C., Goswami, A., Lal, R. & Joshi, R. *Critical evaluation and assessment of average annual precipitation in the Indus, the Ganges and the Brahmaputra basins, Northern India* in *Dynamics of climate change and water resources of Northwestern Himalaya* (2015), 67–84. doi:10.1007/978-3-319-13743-8_7.

142. Kim, H. *Global soil wetness project phase 3 atmospheric boundary conditions (experiment 1)* 2017.

143. Klees, R., Revtova, E., Gunter, B., Ditmar, P., Oudman, E., Winsemius, H. & Savenije, H. The design of an optimal filter for monthly GRACE gravity models. *Geophysical Journal International* **175,** 417–432. doi:10.1111/j.1365-246X.2008.03922.x (2008).

144. Klees, R., Zapreeva, E., Winsemius, H. & Savenije, H. The bias in GRACE estimates of continental water storage variations. *Hydrology and Earth System Sciences* **11,** 1227–1241. doi:10.5194/hess-11-1227-2007 (2007).

145. Koch, K.-R. & Schönfeld, P. Parameter estimation and hypothesis testing in linear models. *Statistical Papers* **30,** 212–212 (1989).

146. Kornfeld, R. P., Arnold, B. W., Gross, M. A., Dahya, N. T., Klipstein, W. M., Gath, P. F. & Bettadpur, S. GRACE-FO: the gravity recovery and climate experiment follow-on mission. *Journal of spacecraft and rockets* **56,** 931–951. doi:10.2514/1.A34326 (2019).

147. Koster, R. D. & Suarez, M. J. Modeling the land surface boundary in climate models as a composite of independent vegetation stands. *Journal of Geophysical Research: Atmospheres* **97,** 2697–2715. doi:10.1029/91JD01696 (1992).

148. Koster, R. D., Suarez, M. J., Ducharne, A., Stieglitz, M. & Kumar, P. A catchment-based approach to modeling land surface processes in a general circulation model: 1. Model structure. *Journal of Geophysical Research: Atmospheres* **105,** 24809–24822. doi:10.1029/2000JD900327 (2000).

149. Koster, R. D., Guo, Z., Yang, R., Dirmeyer, P. A., Mitchell, K. & Puma, M. J. On the Nature of Soil Moisture in Land Surface Models. *Journal of Climate* **22,** 4322–4335. doi:10.1175/2009JCLI2832.1. https://journals.ametsoc.org/view/journals/clim/22/16/2009jcli2832.1.xml (2009).

150. Kraemer, B. M., Seimon, A., Adrian, R. & McIntyre, P. B. Worldwide lake level trends and responses to background climate variation. *Hydrology and Earth System Sciences* **24,** 2593–2608. doi:10.5194/hess-24-2593-2020 (2020).

151. Kumar, S. V., Zaitchik, B. F., Peters-Lidard, C. D., Rodell, M., Reichle, R., Li, B., Jasinski, M., Mocko, D., Getirana, A., De Lannoy, G., *et al.* Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology* **17,** 1951–1972. doi:10.1175/JHM-D-15-0157.1 (2016).

152. Kusche, J. & Schrama, E. Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data. *Journal of geophysical research: solid earth* **110.** doi:10.1029/2004JB003556 (2005).

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

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

155. Kvas, A., Behzadpour, S., Ellmer, M., Klinger, B., Strasser, S., Zehentner, N. & Mayer-Gürr, T. ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series. *Journal of Geophysical Research: Solid Earth* **124,** 9332–9344. doi:10.1029/2019JB017415 (2019).

156. Landerer, F. W. & Swenson, S. Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research* **48.** doi:10.1029/2011WR011453 (2012).

157. Lange, S., Mengel, M., Treu, S. & Büchner, M. *ISIMIP3a atmospheric climate input data* version 1.0. 2022. doi:10.48364/ISIMIP.982724.

158. Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., *et al.* The Community Land Model version 5: description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems* **11,** 4245–4287. doi:10.1029/2018MS001583 (2019).

159. Le Dimet, F.-X., Castaings, W., Ngnepieba, P. & Vieux, B. Data assimilation in hydrology: variational approach. *Data assimilation for atmospheric, oceanic and hydrologic applications,* 367–405. doi:10.1007/978-3-540-71056-1_20 (2009).

160. Leblanc, M. J., Tregoning, P., Ramillien, G., Tweed, S. O. & Fakes, A. Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia. *Water resources research* **45.** doi:10.1029/2008WR007333 (2009).

161. Lehner, B., Verdin, K. & Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos, Transactions American Geophysical Union* **89,** 93–94. doi:10.1029/2008EO100001 (2008).

162. Li, B., Rodell, M., Kumar, S., Beaudoing, H. K., Getirana, A., Zaitchik, B. F., de Goncalves, L. G., Cossetin, C., Bhanja, S., Mukherjee, A., *et al.* Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges. *Water Resources Research* **55,** 7564–7586. doi:10.1029/2018WR024618 (2019).

163. Li, B., Rodell, M., Zaitchik, B. F., Reichle, R. H., Koster, R. D. & van Dam, T. M. Assimilation of GRACE terrestrial water storage into a land surface model: Evaluation and potential value for drought monitoring in western and central Europe. *Journal of Hydrology* **446-447,** 103–115. ISSN: 0022-1694. doi:https://doi.org/10.1016/j.jhydrol.2012.04.035. https://www.sciencedirect.com/science/article/pii/S0022169412003228 (2012).

164. Li, F., Kusche, J., Chao, N., Wang, Z. & Löcher, A. Long-Term (1979-Present) total water storage anomalies over the global land derived by reconstructing GRACE data. *Geophysical Research Letters* **48,** e2021GL093492. doi:10.1029/2021GL093492 (2021).

165. Li, F., Kusche, J., Sneeuw, N., Siebert, S., Gerdener, H., Wang, Z., Chao, N., Chen, G. & Tian, K. Forecasting Next Year's Global Land Water Storage using GRACE Data. *Geophysical Research Letters, under review* (2024).

166. Li, R., Tsunekawa, A. & Tsubo, M. Index-based assessment of agricultural drought in a semi-arid region of Inner Mongolia, China. *Journal of Arid Land* **6,** 3–15. doi:10.1007/s40333-013-0193-8 (2014).

167. Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research: Atmospheres* **99,** 14415–14428. doi:10.1029/94JD00483 (1994).

168. Lin, Y.-C., Kuo, E.-D. & Chi, W.-J. Analysis of meteorological drought resilience and risk assessment of groundwater using signal analysis method. *Water Resources Management* **35,** 179–197. doi:10.1007/s11269-020-02718-x (2021).

169. Lloyd-Hughes, B. The impracticality of a universal drought definition. *Theoretical and applied climatology* **117,** 607–611. doi:10.1007/s00704-013-1025-7 (2014).

170. Long, D., Longuevergne, L. & Scanlon, B. R. Global analysis of approaches for deriving total water storage changes from GRACE satellites. *Water Resources Research* **51,** 2574–2594. doi:10.1002/2014WR016853 (2015).

171. Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N. & Save, H. GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophysical Research Letters* **40,** 3395–3401. doi:10.1002/grl.50655 (2013).

172. Long, D., Shen, Y., Sun, A., Hong, Y., Longuevergne, L., Yang, Y., Li, B. & Chen, L. Drought and flood monitoring for a large karst plateau in Southwest China using extended GRACE data. *Remote Sensing of Environment* **155,** 145–160. doi:10.1016/j.rse.2014.08.006 (2014).

173. Loomis, B., Luthcke, S. & Sabaka, T. Regularization and error characterization of GRACE mascons. *Journal of geodesy* **93,** 1381–1398. doi:10.1007/s00190-019-01252-y (2019).

174. Loomis, B. D., Rachlin, K. E., Wiese, D. N., Landerer, F. W. & Luthcke, S. B. Replacing GRACE/GRACE-FO with satellite laser ranging: Impacts on Antarctic Ice Sheet mass change. *Geophysical Research Letters* **47,** e2019GL085488. doi:10.1029/2019GL085488 (2020).

175. Loukas, A. & Vasiliades, L. Probabilistic analysis of drought spatiotemporal characteristics inThessaly region, Greece. *Natural Hazards and Earth System Sciences* **4,** 719–731. doi:10.5194/nhess-4-719-2004 (2004).

176. Luthcke, S. B., Sabaka, T., Loomis, B., Arendt, A., McCarthy, J. & Camp, J. Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. *Journal of Glaciology* **59,** 613–631. doi:10.3189/2013JoG12J147 (2013).

177. Margat, J. & Van der Gun, J. *Groundwater around the world: a geographic synopsis* doi:10.1201/b13977 (Crc Press, 2013).

178. Mayer-Gürr, T., Behzadpur, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S. & Zehentner, N. *ITSG-Grace2018 - monthly, daily and static gravity field Solutions from GRACE* data set. 2018. doi:10.5880/ICGEM.2018.003.

179. McGlade, J., Bankoff, G., Abrahams, J., Cooper-Knock, S., Cotecchia, F., Desanker, P., Erian, W., Gencer, E., Gibson, L., Girgin, S., *et al. Global assessment report on disaster risk reduction 2019* tech. rep. (United Nations Office for Disaster Risk Reduction, Geneva, Switzerland, 2019). doi:10.1007/978-3-642-03711-5.

180. McKee, T. B., Doesken, N. J., Kleist, J., *et al. The relationship of drought frequency and duration to time scales* in *Proceedings of the 8th Conference on Applied Climatology. American Meteorological Society* **17** (1993), 179–183.

181. Meigh, J., McKenzie, A. & Sene, K. A grid-based approach to water scarcity estimates for eastern and southern Africa. *Water Resources Management* **13,** 85–115. doi:10.1023/A:1008025703712 (1999).

182. Mekonnen, M. & Hoekstra, A. Y. National water footprint accounts: the green, blue and grey water footprint of production and consumption. Volume 1: main report. *Value of Water Research Report Series No. 48, UNESCO-IHE, Delft, the Netherlands.* https://www.waterfootprint.org/resources/Report50-NationalWaterFootprints-Vol1.pdf (2011).

183. Meza, I., Rezaei, E. E., Siebert, S., Ghazaryan, G., Nouri, H., Dubovyk, O., Gerdener, H., Herbert, C., Kusche, J., Popat, E., *et al.* Drought risk for agricultural systems in South Africa: Drivers, spatial patterns, and implications for drought risk management. *Science of the Total Environment* **799,** 149505. doi:10.1016/j.scitotenv.2021.149505 (2021).

184. Meza, I., Siebert, S., Döll, P., Kusche, J., Herbert, C., Eyshi Rezaei, E., Nouri, H., Gerdener, H., Popat, E., Frischen, J., *et al.* Global-scale drought risk assessment for agricultural systems. *Natural Hazards and Earth System Sciences* **20,** 695–712. doi:10.5194/nhess-20-695-2020 (2020).

185. Mishra, A. K. & Singh, V. P. A review of drought concepts. *Journal of hydrology* **391,** 202–216. doi:10.1016/j.jhydrol.2010.07.012 (2010).

186. Modarres, R. Streamflow drought time series forecasting. *Stochastic Environmental Research and Risk Assessment* **21,** 223–233. doi:10.1007/s00477-006-0058-1 (2007).

187. Mohammad, A. H., Jung, H. C., Odeh, T., Bhuiyan, C. & Hussein, H. Understanding the impact of droughts in the Yarmouk Basin, Jordan: monitoring droughts through meteorological and hydrological drought indices. *Arabian Journal of Geosciences* **11.** doi:10.1007/s12517-018-3433-6 (2018).

188. Müller Schmied, H., Trautmann, T., Ackermann, S., Cáceres, D., Flörke, M., Gerdener, H., Kynast, E., Peiris, T. A., Schiebener, L., Schumacher, M. & Döll, P. The global water resources and use model WaterGAP v2.2e: description and evaluation of modifications and new features. *Geoscientific Model Development Discussions* **2023,** 1–46. doi:10.5194/gmd-2023-213. https://gmd.copernicus.org/preprints/gmd-2023-213/ (2023).

189. Müller Schmied, H., Cáceres, D., Eisner, S., Flörke, M., Herbert, C., Niemann, C., Peiris, T. A., Popat, E., Portmann, F. T., Reinecke, R., *et al.* The global water resources and use model WaterGAP v2. 2d: Model description and evaluation. *Geoscientific Model Development* **14,** 1037–1079. doi:10.5194/gmd-14-1037-2021 (2021).

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

191. Narasimhan, B. & Srinivasan, R. Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agricultural and forest meteorology* **133,** 69–88. doi:10.1016/j.agrformet.2005.07.012 (2005).

192. Nerger, L. & Hiller, W. Software for ensemble-based data assimilation systemsImplementation strategies and scalability. *Computers & Geosciences* **55,** 110–118. doi:10.1016/j.cageo.2012.03.026 (2013).

193. Nerger, L., Hiller, W. & Schröter, J. A comparison of error subspace Kalman filters. *Tellus A: Dynamic Meteorology and Oceanography* **57,** 715–735. doi:10.3402/tellusa.v57i5.14732 (2005).

194. Nerger, L., Janjić, T., Schröter, J. & Hiller, W. A unification of ensemble square root Kalman filters. *Monthly Weather Review* **140,** 2335–2345. doi:10.1175/MWR-D-11-00102.1 (2012).

195. Nerger, L., Tang, Q. & Mu, L. Efficient ensemble data assimilation for coupled models with the Parallel Data Assimilation Framework: example of AWI-CM (AWI-CM-PDAF 1.0). *Geoscientific Model Development* **13,** 4305–4321. doi:10.5194/gmd-13-4305-2020 (2020).

196. Ni, S., Chen, J., Wilson, C. R., Li, J., Hu, X. & Fu, R. Global terrestrial water storage changes and connections to ENSO events. *Surveys in Geophysics* **39,** 1–22. doi:10.1007/s10712-017-9421-7 (2018).

197. Nie, N., Zhang, W., Chen, H. & Guo, H. A global hydrological drought index dataset based on gravity recovery and climate experiment (GRACE) data. *Water Resources Management* **32,** 1275–1290. doi:10.1007/s11269-017-1869-1 (2018).

198. Niemeier, W. *Ausgleichungsrechnung: Statistische Auswertemethoden* doi:10.1515/9783110206784 (Walter de Gruyter, 2008).

199. Nussbaumer, H. J. & Nussbaumer, H. J. *The fast Fourier transform* ISBN: 978-3-540-11825-1. doi:10.1007/978-3-642-81897-4_4 (Springer, 1982).

200. Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. & Yorke, J. A. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A: Dynamic Meteorology and Oceanography* **56,** 415–428. doi:10.3402/tellusa.v56i5.14462 (2004).

201. Palmer, W. C. *Meteorological drought* (US Department of Commerce, Weather Bureau, 1965).

202. Palmer, W. C. Keeping track of crop moisture conditions, nationwide: the new crop moisture index. *Weatherwise* **21,** 1156–161. doi:10.1080/00431672.1968.9932814 (1968).

203. Panet, I., Mikhailov, V., Diament, M., Pollitz, F., King, G., De Viron, O., Holschneider, M., Biancale, R. & Lemoine, J.-M. Coseismic and post-seismic signatures of the Sumatra 2004 December and 2005 March earthquakes in GRACE satellite gravity. *Geophysical Journal International* **171,** 177–190. doi:10.1111/j.1365-246X.2007.03525.x (2007).

204. Peduzzi, P., Dao, H., Herold, C. & Mouton, F. Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. *Natural hazards and earth system sciences* **9,** 1149–1159. doi:10.5194/nhess-9-1149-2009 (2009).

205. Peltier, W. R. Global glacial isostasy and the surface of the ice-age Earth: the ICE-5G (VM2) model and GRACE. *Annual Review of Earth and Planetary Sciences.* **32,** 111–149. doi:10.1146/annurev.earth.32.082503.144359 (2004).

206. Peltier, W. R., Argus, D. F. & Drummond, R. Comment on "An assessment of the ICE-6G_C (VM5a) glacial isostatic adjustment model" by Purcell et al. *Journal of Geophysical Research: Solid Earth* **123,** 2019–2028. doi:10.1002/2016JB013844 (2018).

207. Pham, D. T. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly weather review* **129,** 1194–1207. doi:10.1175/1520-0493(2001)129%3C1194:SMFSDA%3E2.0.CO;2 (2001).

208. Phillips, T., Nerem, R., Fox-Kemper, B., Famiglietti, J. & Rajagopalan, B. The influence of ENSO on global terrestrial water storage using GRACE. *Geophysical Research Letters* **39.** doi:10.1029/2012GL052495 (2012).

209. Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., *et al.* Global terrestrial water storage and drought severity under climate change. *Nature Climate Change* **11,** 226–233. doi:10.1038/s41558-020-00972-w (2021).

210. Portmann, F. *Global irrigation in the 20th century: extension of the WaterGAP Global Irrigation Model (GIM) with the spatially explicit Historical Irrigation Data set (HID)* workingpaper. 2017.

211. Preimesberger, W., Scanlon, T., Su, C.-H., Gruber, A. & Dorigo, W. Homogenization of structural breaks in the global ESA CCI soil moisture multisatellite climate data record. *IEEE Transactions on Geoscience and Remote Sensing* **59,** 2845–2862. doi:10.1109/TGRS.2020.3012896 (2020).

212. Preisendorfer, R. Principal Component Analysis in meteorology and oceanography. *Elsevier Sci. Publ.* **17,** 425 (1988).

213. Qin, T., Feng, J., Zhang, X., Li, C., Fan, J., Zhang, C., Dong, B., Wang, H. & Yan, D. Continued decline of global soil moisture content, with obvious soil stratification and regional difference. *Science of The Total Environment* **864,** 160982. doi:10.1016/j.scitotenv.2022.160982 (2023).

214. Ramsahoye, L. & Lang, S. M. *A simple method for determining specific yield from pumping tests* (US Government Printing Office, 1961).

215. Reager, J. T., Thomas, A. C., Sproles, E. A., Rodell, M., Beaudoing, H. K., Li, B. & Famiglietti, J. S. Assimilation of GRACE terrestrial water storage observations into a land surface model for the assessment of regional flood potential. *Remote Sensing* **7,** 14663–14679. doi:10.3390/rs71114663 (2015).

216. Rebischung, P. *Terrestrial frame solutions from the IGS third reprocessing* in *EGU General Assembly Conference Abstracts* (2021), EGU21–2144.

217. Redner, R. A. & Walker, H. F. *Mixture densities, maximum likelihood and the EM Algorithm* 1984. doi:10.1137/1026034.

218. Reinecke, R., Foglia, L., Mehl, S., Trautmann, T., Cáceres, D. & Döll, P. Challenges in developing a global gradient-based groundwater model (G3M v1. 0) for the integration into a global hydrological model. *Geoscientific Model Development* **12.** doi:10.5194/gmd-12-2401-2019 (2019).

219. Rissanen, J. Modeling by shortest data description. *Automatica* **14,** 465–471. doi:10.1016/0005-1098(78)90005-5 (1978).

220. Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J., Beaudoing, H. K., Landerer, F. W. & Lo, M.-H. Emerging trends in global freshwater availability. *Nature* **557,** 651–659. doi:10.1038/s41586-018-0123-1 (2018).

221. Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., *et al.* The global land data assimilation system. *Bulletin of the American Meteorological society* **85,** 381–394. doi:10.1175/BAMS-85-3-381 (2004).

222. Rodell, M., Velicogna, I. & Famiglietti, J. S. Satellite-based estimates of groundwater depletion in India. *Nature* **460,** 999–1002. doi:10.1038/nature08238 (2009).

223. Roser, M. Employment in Agriculture. *Our World in Data.* https://ourworldindata.org/employment-in-agriculture (2023).

224. Saha, T. R., Shrestha, P. K., Rakovec, O., Thober, S. & Samaniego, L. A drought monitoring tool for South Asia. *Environmental Research Letters* **16,** 054014. doi:10.1088/1748-9326/abf525 (2021).

225. Satoh, Y., Yoshimura, K., Pokhrel, Y., Kim, H., Shiogama, H., Yokohata, T., Hanasaki, N., Wada, Y., Burek, P., Byers, E., *et al.* The timing of unprecedented hydrological drought under climate change. *Nature Communications* **13,** 3287. doi:10.1038/s41467-022-30729-2 (2022).

226. Scanlon, B. R., Zhang, Z., Save, H., Sun, A. Y., Müller Schmied, H., Van Beek, L. P., Wiese, D. N., Wada, Y., Long, D., Reedy, R. C., *et al.* Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences* **115,** E1080–E1089. doi:10.1073/pnas.1704665115 (2018).

227. Schulze, K., Kusche, J., Gerdener, H., Döll, P. & Müller Schmied, H. Benefits and Pitfalls of GRACE and Streamflow Assimilation for Improving the Streamflow Simulations of the WaterGAP Global Hydrology Model. *Journal of Advances in Modeling Earth Systems, under review* (2024).

228. Schumacher, D. L., Keune, J., Dirmeyer, P. & Miralles, D. G. Drought self-propagation in drylands due to land–atmosphere feedbacks. *Nature Geoscience* **15,** 262–268. doi:10.1038/s41561-022-00912-7 (2022).

229. Schumacher, M. *Methods for assimilating remotely-sensed water storage changes into hydrological models* PhD thesis (Universität Bonn, Institut für Geodäsie und Geoinformation, 2016). doi:20.500.11811/6630.

230. Schumacher, M., Forootan, E., van Dijk, A. I., Schmied, H. M., Crosbie, R. S., Kusche, J. & Döll, P. 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 (2018).

231. Schumacher, M., Kusche, J. & Döll, P. A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models. *Journal of Geodesy* **90,** 537–559. doi:10.1007/s00190-016-0892-y (2016).

232. Schwarz, M., Landmann, T., Cornish, N., Wetzel, K.-F., Siebert, S. & Franke, J. A spatially transferable drought hazard and drought risk modeling approach based on remote sensing data. *Remote Sensing* **12,** 237. doi:10.3390/rs12020237 (2020).

233. Sepulcre-Canto, G., Horion, S., Singleton, A., Carrao, H. & Vogt, J. Development of a Combined Drought Indicator to detect agricultural drought in Europe. *Natural Hazards and Earth System Sciences* **12,** 3519–3531. doi:10.5194/nhess-12-3519-2012 (2012).

234. Sheffield, J. & Wood, E. F. *Drought: past problems and future scenarios* doi:10.4324/9781849775250 (Routledge, 2011).

235. Shukla, S. & Wood, A. W. Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical research letters* **35.** doi:10.1029/2007GL032487 (2008).

236. Sinha, D., Syed, T. H., Famiglietti, J. S., Reager, J. T. & Thomas, R. C. Characterizing drought in India using GRACE observations of terrestrial water storage deficit. *Journal of Hydrometeorology* **18,** 381–396. doi:10.1175/JHM-D-16-0047.1 (2017).

237. Sivakumar, M. V., Motha, R., Wilhite, D. & Wood, D. *Agricultural drought indices. Proceedings of an expert meeting: 2-4 June, 2010, Murcia* doi:10.1007/978-3-642-03711-5 (World Meteorological Organization Geneva, Switzerland, 2011).

238. Sivakumar, M. V., Stefanski, R., Bazza, M., Zelaya, S., Wilhite, D. & Magalhaes, A. R. High level meeting on national drought policy: summary and major outcomes. *Weather and climate Extremes* **3,** 126–132. doi:10.1016/j.wace.2014.03.007 (2014).

239. Smith, M. D., Wilkins, K. D., Holdrege, M. C., Wilfahrt, P., Collins, S. L., Knapp, A. K., Sala, O. E., Dukes, J. S., Phillips, R. P., Yahdjian, L., *et al.* Extreme drought impacts have been underestimated in grasslands and shrublands globally. *Proceedings of the National Academy of Sciences* **121,** e2309881120. doi:10.1073/pnas.2309881120 (2024).

240. Sood, A. & Smakhtin, V. Global hydrological models: a review. *Hydrological Sciences Journal* **60,** 549–565. doi:10.1080/02626667.2014.950580 (2015).

241. Sorí, R., Nieto, R., Vicente-Serrano, S. M., Drumond, A. & Gimeno, L. A Lagrangian perspective of the hydrological cycle in the Congo River basin. *Earth System Dynamics* **8,** 653–675. doi:10.5194/esd-8-653-2017 (2017).

242. Spinoni, J., Barbosa, P., Bucchignani, E., Cassano, J., Cavazos, T., Christensen, J. H., Christensen, O. B., Coppola, E., Evans, J., Geyer, B., *et al.* Future global meteorological drought hot spots: a study based on CORDEX data. *Journal of Climate* **33,** 3635–3661. doi:10.1175/JCLI-D-19-0084.1 (2020).

243. Springer, A. *A water storage reanalysis over the European continent: assimilation of GRACE data into a high-resolution hydrological model and validation* PhD thesis (Universität Bonn, Institut für Geodäsie und Geoinformation, 2019). doi:20.500.11811/7987.

244. Sternberg, W. J. & Smith, T. L. *The theory of potential and spherical harmonics* (University of Toronto Press, 1944).

245. Sun, Y., Riva, R. & Ditmar, P. Optimizing estimates of annual variations and trends in geocenter motion and J2 from a combination of GRACE data and geophysical models. *Journal of Geophysical Research: Solid Earth* **121,** 8352–8370. doi:10.1002/2016JB013073 (2016).

246. Swenson, S., Chambers, D. & Wahr, J. Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research: Solid Earth* **113.** doi:10.1029/2007JB005338 (2008).

247. Swenson, S. & Wahr, J. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity. *Journal of Geophysical Research: Solid Earth* **107,** ETG 3-1 –ETG 3-13. doi:10.1029/2001JB000576 (2002).

248. Swenson, S. & Wahr, J. Post-processing removal of correlated errors in GRACE data. *Geophysical research letters* **33.** doi:10.1029/2005GL025285 (2006).

249. Swift, L. & Saulle, J. Dealing with food crises in the Sahel. *Humanitarian Assistance in West Africa and Beyond,* 59. doi:10.1007/978-3-642-03711-5 (2015).

250. Tadikamalla, P. R. A look at the Burr and related distributions. *International Statistical Review/Revue Internationale de Statistique,* 337–344. doi:10.2307/1402945 (1980).

251. Tallaksen, L. M. & Van Lanen, H. A. *Hydrological drought: processes and estimation methods for streamflow and groundwater* doi:10.1016/C2017-0-03464-X (Elsevier, 2004).

252. Tangdamrongsub, N., Ditmar, P., Steele-Dunne, S., Gunter, B. & Sutanudjaja, E. Assessing total water storage and identifying flood events over Tonlé Sap basin in Cambodia using GRACE and MODIS satellite observations combined with hydrological models. *Remote Sensing of Environment* **181,** 162–173. doi:10.1016/j.rse.2016.03.030 (2016).

253. Tangdamrongsub, N., Steele-Dunne, S. C., Gunter, B. C., Ditmar, P. G., Sutanudjaja, E. H., Sun, Y., Xia, T. & Wang, Z. 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,** 2053–2074. doi:10.5194/hess-21-2053-2017 (2017).

254. Tapley, B. D., Bettadpur, S., Watkins, M. & Reigber, C. The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters* **31.** doi:10.1029/2004GL019920 (2004).

255. Tapley, B. D., Watkins, M. M., Flechtner, F., Reigber, C., Bettadpur, S., Rodell, M., Sasgen, I., Famiglietti, J. S., Landerer, F. W., Chambers, D. P., *et al.* Contributions of GRACE to understanding climate change. *Nature climate change* **9,** 358–369. doi:10.1038/s41558-019-0456-2 (2019).

256. Telesca, L., Lovallo, M., Lopez-Moreno, I. & Vicente-Serrano, S. Investigation of scaling properties in monthly streamflow and Standardized Streamflow Index (SSI) time series in the Ebro basin (Spain). *Physica A: Statistical Mechanics and its Applications* **391,** 1662–1678. doi:10.1016/j.physa.2011.10.023 (2012).

257. Thom, H. C. A note on the gamma distribution. *Monthly weather review* **86,** 117–122. doi:10.1175/1520-0493(1958)086<0117:ANOTGD>2.0.CO;2 (1958).

258. Thomas, A. C., Reager, J. T., Famiglietti, J. S. & Rodell, M. A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters* **41,** 1537–1545. doi:10.1002/2014GL059323 (2014).

259. Thomas, B. F., Famiglietti, J. S., Landerer, F. W., Wiese, D. N., Molotch, N. P. & Argus, D. F. GRACE groundwater drought index: Evaluation of California Central Valley groundwater drought. *Remote Sensing of Environment* **198,** 384–392. doi:10.1016/j.rse.2017.06.026 (2017).

260. Thornthwaite, C. W. An approach toward a rational classification of climate. *Geographical review* **38,** 55–94. doi:10.2307/210739 (1948).

261. Tian, S., Tregoning, P., Renzullo, L. J., Van Dijk, A. I., Walker, J. P., Pauwels, V. R. & Allgeyer, S. Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals. *Water Resources Research* **53,** 1820–1840. doi:10.1002/2016WR019641 (2017).

262. Trautmann, T., Koirala, S., Guentner, A., Kim, H. & Jung, M. Calibrating global hydrological models with GRACE TWS: does river storage matter? *Environmental Research Communications* **5,** 081005. doi:10.1088/2515-7620/acece5 (2023).

263. Tweed, S., Leblanc, M. & Cartwright, I. Groundwater–surface water interaction and the impact of a multi-year drought on lakes conditions in South-East Australia. *Journal of Hydrology* **379,** 41–53. ISSN: 0022-1694. doi:https://doi.org/10.1016/j.jhydrol.2009.09.043. https://www.sciencedirect.com/science/article/pii/S0022169409006179 (2009).

264. Ukasha, M., Ramirez, J. A. & Niemann, J. D. Temporal variations of NDVI and LAI and interactions with hydroclimatic variables in a large and agro-ecologically diverse region. *Journal of Geophysical Research: Biogeosciences* **127,** e2021JG006395. doi:10.1029/2021JG006395 (2022).

265. United Nations. *The sustainable development goals report 2022* tech. rep. (United Nations Publications, 2022). https://unstats.un.org/sdgs/report/2022/.

266. Van Loon, A. F. Hydrological drought explained. *Wiley Interdisciplinary Reviews: Water* **2,** 359–392. doi:10.1002/wat2.1085 (2015).

267. Van Loon, A. F., Kumar, R. & Mishra, V. Testing the use of standardised indices and GRACE satellite data to estimate the European 2015 groundwater drought in near-real time. *Hydrology and Earth System Sciences* **21,** 1947–1971. doi:10.5194/hess-21-1947-2017 (2017).

268. Van Loon, A. F., Tijdeman, E., Wanders, N., Van Lanen, H. J., Teuling, A. & Uijlenhoet, R. How climate seasonality modifies drought duration and deficit. *Journal of Geophysical Research: Atmospheres* **119,** 4640–4656. doi:10.1002/2013JD020383 (2014).

269. Van Loon, A. F. & Van Lanen, H. A. A process-based typology of hydrological drought. *Hydrology and Earth System Sciences* **16,** 1915–1946. doi:10.5194/hess-16-1915-2012 (2012).

270. Van Rooy, M. *A rainfall anomally index independent of time and space, notos* in (Weather Bureau South Africa, 1965). https://api.semanticscholar.org/CorpusID:130892564.

271. Van Vliet, M. T., Yearsley, J. R., Ludwig, F., Vögele, S., Lettenmaier, D. P. & Kabat, P. Vulnerability of US and European electricity supply to climate change. *Nature Climate Change* **2,** 676–681. doi:10.1038/nclimate1546 (2012).

272. Velicogna, I. & Wahr, J. Greenland mass balance from GRACE. *Geophysical research letters* **32.** doi:10.1029/2005GL023955 (2005).

273. Velicogna, I. & Wahr, J. Measurements of time-variable gravity show mass loss in Antarctica. *science* **311,** 1754–1756. doi:10.1126/science.1123785 (2006).

274. Vetra-Carvalho, S., Van Leeuwen, P. J., Nerger, L., Barth, A., Altaf, M. U., Brasseur, P., Kirchgessner, P. & Beckers, J.-M. State-of-the-art stochastic data assimilation methods for high-dimensional non-Gaussian problems. *Tellus A: Dynamic Meteorology and Oceanography* **70,** 1–43. doi:10.1080/16000870.2018.1445364 (2018).

275. Vicente-Serrano, S. M., Beguería, S., López-Moreno, J. I., Angulo, M. & El Kenawy, A. A new global 0.5 gridded dataset (1901–2006) of a multiscalar drought index: comparison with current drought index datasets based on the Palmer Drought Severity Index. *Journal of Hydrometeorology* **11,** 1033–1043. doi:10.1175/2010JHM1224.1 (2010).

276. Vicente-Serrano, S. M., López-Moreno, J. I., Gimeno, L., Nieto, R., Morán-Tejeda, E., Lorenzo-Lacruz, J., Beguería, S. & Azorin-Molina, C. A multiscalar global evaluation of the impact of ENSO on droughts. *Journal of Geophysical Research: Atmospheres* **116.** doi:10.1029/2011JD016039 (2011).

277. Vishwakarma, B. D., Horwath, M., Devaraju, B., Groh, A. & Sneeuw, N. A data-driven approach for repairing the hydrological catchment signal damage due to filtering of GRACE products. *Water Resources Research* **53,** 9824–9844. doi:10.1002/2017WR021150 (2017).

278. Vishwakarma, B. D., Zhang, J. & Sneeuw, N. Downscaling GRACE total water storage change using partial least squares regression. *Scientific data* **8.** doi:10.1038/s41597-021-00862-6 (2021).

279. Vörösmarty, C. J., Moore III, B., Grace, A. L., Gildea, M. P., Melillo, J. M., Peterson, B. J., Rastetter, E. B. & Steudler, P. A. Continental scale models of water balance and fluvial transport: An application to South America. *Global biogeochemical cycles* **3,** 241–265. doi:10.1029/GB003i003p00241 (1989).

280. Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M. & Swenson, S. C. Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region. *Water resources research* **49,** 904–914. doi:10.1002/wrcr.20078 (2013).

281. Wahr, J. Time-variable gravity from satellites. *Treatise on geophysics* **3,** 213–237. doi:10.1016/B978-044452748-6.00176-0 (2007).

282. Wahr, J., Molenaar, M. & Bryan, F. Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research: Solid Earth* **103,** 30205–30229. doi:10.1029/98JB02844 (1998).

283. Wahr, J., Swenson, S. & Velicogna, I. Accuracy of GRACE mass estimates. *Geophysical Research Letters* **33.** doi:10.1029/2005GL025305 (2006).

284. Wanders, N., van Lanen, H. A. & van Loon, A. F. *Indicators for drought characterization on a global scale* tech. rep. (Wageningen Universiteit, 2010). https://edepot.wur.nl/160049.

285. Wang, H., Xiang, L., Jia, L., Jiang, L., Wang, Z., Hu, B. & Gao, P. Load Love numbers and Green's functions for elastic Earth models PREM, iasp91, ak135, and modified models with refined crustal structure from Crust 2.0. *Computers & Geosciences* **49,** 190–199. doi:10.1016/j.cageo.2012.06.022 (2012).

286. Wang, W., Shen, Y., Wang, F. & Li, W. Two severe prolonged hydrological droughts analysis over Mainland Australia using GRACE satellite data. *Remote sensing* **13,** 1432. doi:10.3390/rs13081432 (2021).

287. Wang, X., de Linage, C., Famiglietti, J. & Zender, C. S. Gravity Recovery and Climate Experiment (GRACE) detection of water storage changes in the Three Gorges Reservoir of China and comparison with in situ measurements. *Water Resources Research* **47.** doi:10.1029/2011WR010534 (2011).

288. Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C. & Landerer, F. W. Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth* **120,** 2648–2671. doi:10.1002/2014JB011547 (2015).

289. Weedon, G., Gomes, S., Viterbo, P., Shuttleworth, W., Blyth, E., Österle, H., Adam, J., Bellouin, N., Boucher, O. & Best, M. Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century. *Journal of Hydrometeorology* **12,** 823–848. doi:10.1175/2011JHM1369.1 (2011).

290. Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J. & Viterbo, P. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research* **50,** 7505–7514. doi:10.1002/2014WR015638 (2014).

291. Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S. & Lee, X. Revisiting the contribution of transpiration to global terrestrial evapotranspiration. *Geophysical Research Letters* **44,** 2792–2801. doi:10.1002/2016GL072235 (2017).

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

293. Wiese, D., Nerem, R. & Lemoine, F. Design considerations for a dedicated gravity recovery satellite mission consisting of two pairs of satellites. *Journal of Geodesy* **86,** 81–98. doi:10.1007/s00190-011-0493-8 (2012).

294. Wikle, C. K. & Berliner, L. M. A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena* **230,** 1–16. doi:10.1016/j.physd.2006.09.017 (2007).

295. Wilhite, D. A. *Preparing for drought: A guidebook for developing countries* https://wedocs.unep.org/20.500.11822/30153 (Diane Publishing, 1994).

296. Wilhite, D. A. & Glantz, M. H. Understanding: the drought phenomenon: the role of definitions. *Water international* **10,** 111–120. doi:10.1080/02508068508686328 (1985).

297. World Meteorological Organization. *The status of the global climate observing system 2021: The GCOS status report* (eds Zemp, M., Eggleston, S., Míguez, B. M., Oakley, T., Rea, A., Robbez, M. & Tassone, C.) 2021. https://gcos.wmo.int/en/publications/gcos-status-report-2021.

298. World Meteorological Organization & Global Water Partnership. *Handbook of drought indicators and indices* (Integrated Drought Management Programme (IDMP), Integrated Drought Management Tools and Guidelines Series 2. Geneva, 2016).

299. World Meteorological Organization, United Nations Environment Programme, International Science Council & Intergovernmental Oceanographic Commission of the United Nations Educational, Scientific and Cultural Organization. *GCOS 2022 implementation plan* (eds Zemp, M., Chao, Q., Han Dolman, A. J., Herold, M., Krug, T., Speich, S., Suda, K., Thorne, P. & Yu, W.) 2022. https://library.wmo.int/idurl/4/58104.

300. Xia, Y., Mocko, D., Huang, M., Li, B., Rodell, M., Mitchell, K. E., Cai, X. & Ek, M. B. Comparison and assessment of three advanced land surface models in simulating terrestrial water storage components over the United States. *Journal of Hydrometeorology* **18,** 625–649. doi:10.1175/JHM-D-16-0112.1 (2017).

301. Yang, L., Sun, G., Zhi, L. & Zhao, J. Negative soil moisture-precipitation feedback in dry and wet regions. *Scientific reports* **8,** 4026. doi:10.1038/s41598-018-22394-7 (2018).

302. Yi, H. & Wen, L. Satellite gravity measurement monitoring terrestrial water storage change and drought in the continental United States. *Scientific reports* **6,** 1–9. doi:10.1038/srep19909 (2016).

303. Yihdego, Y., Vaheddoost, B. & Al-Weshah, R. A. Drought indices and indicators revisited. *Arabian Journal of Geosciences* **12,** 1–12. doi:10.1007/s12517-019-4237-z (2019).

304. Yirdaw, S. Z., Snelgrove, K. R. & Agboma, C. O. GRACE satellite observations of terrestrial moisture changes for drought characterization in the Canadian Prairie. *Journal of Hydrology* **356,** 84–92. doi:10.1016/j.jhydrol.2008.04.004 (2008).

305. Zaitchik, B. F., Rodell, M. & Olivera, F. Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resources Research* **46.** doi:10.1029/2009WR007811 (2010).

306. Zhao, M., Geruo, A., Velicogna, I. & Kimball, J. S. A global gridded dataset of GRACE drought severity index for 2002–14: comparison with PDSI and SPEI and a case study of the Australia millennium drought. *Journal of Hydrometeorology* **18,** 2117–2129. doi:10.1175/JHM-D-16-0182.1 (2017).