Analyzing and characterizing spaceborne observation of water storage variation

past, present, future

A thesis accepted by the Faculty of Aerospace Engineering and Geodesy of the University of Stuttgart in fulfilment of the requirements for the degree of Doctor of Engineering Sciences (Dr.-Ing.)

by

Peyman Saemian

born in Chalus, Iran

Main referee: Prof. Dr.-Ing. Nico Sneeuw
Co-referee: Prof. Dr. Amir AghaKouchak

Date of defence: 20 October 2023

Institute of Geodesy
University of Stuttgart
2024

Contents

Acknowledgement

Completing a PhD thesis is a long and challenging journey, and I am grateful to have had the support and guidance of many people along the way. First and foremost, I would like to express my sincere appreciation to my supervisor, Prof. Nico Sneeuw, for his invaluable advice, encouragement, and patience throughout my research. His exceptional leadership not only made this thesis possible but also transformed the challenges into stepping stones for academic growth, creating an indelible mark on my scholarly pursuits.

I am also grateful to my co-referee, Prof. Amir Aghakouchak, whose incisive comments and invaluable suggestions have significantly elevated the quality of this thesis. His expertise and discerning insights not only enriched the scholarly content but also highlighted the collaborative spirit that fuels academic excellence.

I am immensely grateful to Dr. Mohammad J. Tourian for his consistent daily supervision, his unwavering support, and his dedicated mentorship, which have been pivotal to the success of my PhD journey. He has been a true friend, a steadfast colleague, and an exceptional scientist whose unwavering support not only shaped the trajectory of my research but also elevated my academic journey into a collaborative and enriching experience.

To my family, including my parents Nader Saemian and Shahrbanoo Ahangari, and my brother Saman Saemian, I express my deepest gratitude. Their boundless love, encouragement, and unwavering belief in me have been the bedrock of my success, and for this, I am eternally grateful. Their unwavering support has not only shaped my academic journey but has also enriched my personal growth in ways that words cannot fully capture.

I would like to express my heartfelt thanks to all the participants who took part in this study, including my colleagues at the Institute of Geodesy. Without their willingness to share their experiences, this research would not have been possible. Thank you to all who have been a part of my journey, and who have helped me reach this significant milestone.

Last but not least, my heartfelt appreciation also goes to my wife, Simineh Ghorbani, who has been my constant source of strength and inspiration. Her unwavering love, support, and encouragement have been the driving force behind my success, and I am immensely grateful for her presence in my life. I cherish every moment we've shared on this journey, and her unwavering belief in me has made all the difference. I could not have accomplished this without her unwavering support and commitment.

Peyman Saemian
Stuttgart, October 2023

Abstract

Water storage is an indispensable constituent of the intricate water cycle, as it governs the availability and distribution of this precious resource. Any alteration in the water storage can trigger a cascade of consequences, affecting not only our agricultural practices but also the well-being of various ecosystems and the occurrence of natural hazards. Therefore, it is essential to monitor and manage the water storage levels prudently to ensure a sustainable future for our planet. Despite significant advancements in ground-based measurements and modeling techniques, accurately measuring water storage variation remained a major challenge for a long time. Since 2002, the Gravity Recovery and Climate Experiment (GRACE) and its successor GRACE Follow-On (GRACE-FO) satellites have revolutionized our understanding of the Earth's water cycle. By detecting variations in the Earth's gravity field caused by changes in water distribution, these satellites can precisely measure changes in total water storage (TWS) across the entire globe, providing a truly comprehensive view of the world's water resources. This information has proved invaluable for understanding how water resources are changing over time, and for developing strategies to manage these resources sustainably. However, GRACE and GRACE-FO are subject to various challenges that must be addressed in order to enhance the efficacy of our exploitation of GRACE observations for scientific and practical purposes. This thesis aims to address some of the challenges faced by GRACE and GRACE-FO.

Since the inception of the GRACE mission, scholars have commonly extracted mass changes from observations by approximating the Earth's gravity field utilizing mathematical functions termed spherical harmonics. Various institutions have already processed GRACE(-FO) data, known as level-2 data in the GRACE community, considering the constraints, approaches, and models that have been utilized. However, this processed data necessitates post-processing to be used for several applications, such as hydrology and climate research. In this thesis, we evaluate various methods of processing GRACE(-FO) level-2 data and assess the spatio-temporal effect of the post-processing steps. Furthermore, we aim to compare the consistency between GRACE and its successor mission, GRACE-FO, in terms of data quality and measurement accuracy. By analyzing and comparing the data from these two missions, we can identify any potential discrepancies or differences and establish the level of confidence in the accuracy and reliability of the GRACE-FO measurements. Finally, we will compare the processed level-3 products with the level-3 products that are presently accessible online.

The relatively short record of the GRACE measurements, compared to other satellite missions and observational records, can limit some studies that require long-term data. This short record makes it challenging to separate long-term signals from short-term variability and validate the data with ground-based measurements or other satellite missions. To address this limitation, this thesis expands the temporal coverage of GRACE(-FO) observations using global hydrological, atmospheric, and reanalysis models. First, we assess these models in estimating the TWS variation at a global scale. We compare the performance of various methods including data-driven and machine learning approaches in incorporating models and reconstruct GRACE TWS change. The results are also validated against Satellite Laser Ranging (SLR) observations over the pre-GRACE period. This thesis develops a hindcasted GRACE, which provides a better understanding of the changes in the Earth's water storage on a longer time scale.

The GRACE satellite mission detects changes in the overall water storage in a specific region but cannot distinguish between the different compartments of TWS, such as surface water, ground-water, and soil moisture. Understanding these individual components is crucial for managing water resources and addressing the effects of droughts and floods. This study aims to integrate various data sources to improve our understanding of water storage variations at the continental to basin scale, including water fluxes, lake water level, and lake storage change data. Additionally, the study demonstrates the importance of combining GRACE(-FO) observations with other measurements, such as piezometric wells and rain-gauges, to understand the water scarcity predicament in Iran and other regions facing similar challenges.

The GRACE satellite mission provides valuable insights into the Earth's system. However, the GRACE product has a level of uncertainty due to several error sources. While the mission has taken measures to minimize these uncertainties, researchers need to account for them when analyzing the data and communicate them when reporting findings. This thesis proposes a probabilistic approach to incorporate the Total Water Storage Anomaly (TWSA) data from GRACE(-FO). By accounting for the uncertainty in the TWSA data, this approach can provide a more comprehensive understanding of drought conditions, which is essential for decision makers managing water resources and responding to drought events.

Zusammenfassung

Der Wasserspeicher ist ein essentieller Bestandteil des komplexen Wasserkreislaufs, da sie die Verfügbarkeit und Verteilung dieser wertvollen Ressource steuert. Jede Änderung des Wasserspeichers kann eine Reihe von Konsequenzen auslösen, die nicht nur unsere Landwirtschaft, sondern auch die Intaktheit verschiedener Ökosysteme und das Auftreten von Naturgefahren beeinflussen. Daher ist es von entscheidender Bedeutung, die Wasserstandsmessung sorgfältig zu überwachen und zu verwalten, um eine nachhaltige Zukunft für unseren Planeten zu sicherstellen. Trotz bedeutender Fortschritte bei bodenbasierten Messungen und Modellierungstechniken blieb die genaue Messung von Wasserstandsänderungen lange Zeit eine große Herausforderung. Seit 2002 haben die Satelliten der Gravity Recovery and Climate Experiment (GRACE) und ihr Nachfolger GRACE Follow-On (GRACE-FO) unsere Kenntnisse über den Wasserkreislauf der Erde umgestellt. Durch die Feststellung von Variationen im Gravitationsfeld der Erde, die durch Änderungen in der Wasserverteilung verursacht werden, können diese Satelliten Änderungen im gesamten Wasserspeicher (TWS) auf der Erde präzise messen und damit eine umfassende Sicht auf die weltweiten Wasserressourcen bieten. Diese Informationen sind essentiell, um zu verstehen, wie sich Wasserressourcen im Laufe der Zeit verändern und um Strategien zu entwickeln diese Ressourcen nachhaltig zu verwalten. Allerdings sind GRACE und GRACE-FO verschiedenen Herausforderungen ausgesetzt, die angegangen werden müssen, um die Nutzung von GRACE-Beobachtungen für wissenschaftliche und praktische Zwecke zu verbessern. Diese Arbeit hat das Ziel, einige der Herausforderungen zu behandeln, die bei GRACE und GRACE-FO auftreten.

Seit Beginn der GRACE-Mission verwenden Wissenschaftler zum Herleiten der Massenänderungen (meist) eine Darstellung des Gravitationsfeldes mit Hilfe von so genannten Kugelfunktionen. Verschiedene Institutionen haben bereits GRACE(-FO)-Daten, in der GRACE Community als Level-2-Daten bekannt, unter Berücksichtigung der verwendeten Einschränkungen, Ansätze und Modelle verarbeitet. Diese verarbeiteten Daten erfordern jedoch eine Nachbearbeitung, damit sie für Anwendungen, wie beispielsweise die Hydrologie und Klimaforschung, genutzt werden können. In dieser Arbeit bewerten wir verschiedene Methoden zur Verarbeitung von GRACE(-FO) Level-2-Daten und untersuchen den Effekt der Nachbearbeitung auf das räumliche und das zeitliche Verhalten der Daten. Darüber hinaus möchten wir der Konsistenz zwischen GRACE und seiner Nachfolgemission, GRACE-FO, hinsichtlich der Datenqualität und Messgenauigkeit vergleichen. Durch die Analyse und den Vergleich der Daten aus diesen beiden Missionen können potenzielle Unstimmigkeiten oder Unterschiede identifiziert sowie das Maß an Genauigkeit und Zuverlässigkeit der GRACE-FO-Messungen festgestellt werden. Schließlich werden die neu gewonnenen Level-3-Produkte mit den Level-3-Produkten verglichen, die derzeit online zugänglich sind.

Der im Vergleich zu anderen Satellitenmissionen und Beobachtungsreichen kurze Zeitraum der GRACE-Messungen kann einige Studien einschränken, die langfristige Daten erfordern. Dieser kurze Zeitraum erschwert die Trennung von langfristigen Signalen und kurzfristiger Variabilität sowie die Validierung der Daten mit bodengestützten Messungen oder anderen Satellitenmissionen. Um diese Einschränkung zu bewältigen, nutzt diese Arbeit globale hydrologische, atmosphärische und Reanalyse Modelle, um die zeitliche Abdeckung der GRACE(-FO)-Beobachtungen zu erweitern. Zunächst bewerten wir diese Modelle in der Schätzung der

TWS-Variation im globalen Maßstab. Wir vergleichen die Leistung verschiedener Methoden, einschließlich datengesteuerter und maschineller Lernansätze, um Modelle zu integrieren und die Veränderungen der GRACE-TWS zu rekonstruieren. Die Ergebnisse werden auch über den Zeitraum vor GRACE gegenüber Satelliten-Laserentfernungsmessungen validiert. Diese Arbeit entwickelt einen rückwirkend erzeugten GRACE, die ein besseres Verständnis der Veränderungen des Erdwasserspeichers auf einer längeren Zeitskala ermöglicht.

Die GRACE-Satellitenmission erfasst Veränderungen des gesamten Wasserspeichers in einer bestimmten Region, kann jedoch nicht zwischen den verschiedenen Kompartimenten des TWS, wie Oberflächenwasser, Grundwasser und Bodenfeuchte unterscheiden. Das Verständnis dieser einzelnen Komponenten ist entscheidend für die Bewirtschaftung von Wasserressourcen und die Bewältigung der Auswirkungen von Dürren und Überschwemmungen. Diese Studie zielt darauf ab, verschiedene Datenquellen zu integrieren, um unser Verständnis von Wasserspeichervariationen auf kontinentaler bis hin zur Beckenskala zu verbessern, einschließlich Wasserflüssen, Wasserstand von Seen und Veränderungen des Seevolumens. Zusätzlich zeigt die Studie die Bedeutung des Zusammenhangs von GRACE(-FO)-Beobachtungen mit anderen Messungen wie Piezometern und Niederschlagsmessern, um das Problem der Wasserknappheit im Iran und anderen Regionen mit ähnlichen Herausforderungen zu verstehen.

Die GRACE-Satellitenmission bietet wertvolle Einblicke in das Erdsystem. Das GRACE-Produkt ist jedoch aufgrund verschiedener Fehlerquellen mit einem gewissen Grad an Unsicherheit behaftet. Obwohl die Mission Maßnahmen ergriffen hat, um diese Unsicherheiten zu minimieren, müssen Forscher sie berücksichtigen, wenn sie die Daten analysieren und sie bei der Berichterstattung übermitteln. Diese Arbeit schlägt einen probabilistischen Ansatz vor, um die Daten der Total Water Storage Anomalie (TWSA) von GRACE(-FO) in die Bewertung von Dürregefahren zu integrieren. Durch Berücksichtigung der Unsicherheit in den TWSA-Daten, kann dieser Ansatz ein umfassenderes Verständnis von Dürrebedingungen vermitteln, was für Entscheidungsträger bei der Bewältigung von Wasserknappheit und Dürreereignissen unerlässlich ist.

List of Abbreviations

ACC	SuperSTAR Accelerometers		**DLR**	German Aerospace Center
AI	Aridity Index		**DSI**	Drought Severity Index
AIUB	Astronomical Institute University Bern		**DT**	Decision Tree
			EC	Eddy Covariance
CES	Coarse Earth Sun Sensor		**ECMWF**	European Centre for Medium-Range Weather Forecasts
CF	Center of Figure			
CHAMP	CHAllenging Minisatellite Payload			
			EM	Ensemble Mean
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data		**ENKF**	Ensemble Kalman Filter
			ENSO	El Niño-Southern Oscillation
CMI	Crop Moisture Index		**EOF**	Empirical Orthogonal Function
CNES	Centre National d'Etudes Spatiales		**ENVISAT**	Environmental Satellite
			ERS	European Remote Sensing
CM	Center of Mass		**ESG**	Earth System Grid
COST-G	International Combination Service for Time-variable Gravity Fields		**EUMETSAT**	Exploitation of Meteorological Satellites
			EPA	U.S. Environmental Protection Agency
C3S	Copernicus Climate Change Service			
			ERS	European Remote Sensing
CRU	Climatic Research Unit		**EWH**	Equivalent Water Height
CSA	Canadian Space Agency			
CSR	University of Texas Centre for Space Research		**FGGE**	First GARP Global Experiment
			FluxCom	FLUXNET COMprise
CV	Coefficient of Variation		**GIA**	Glacial Isostatic Adjustment
DAHITI	Database for Hydrological Time Series of Inland Waters		**GMAO**	Global Modeling and Assimilation Office
DDK	Denoising and Decorrelation Kernel		**GOCE**	Gravity field and steady-state Ocean Circulation Explorer
DFGI	Deutsches Geodätisches Forschungsinstitut		**GPCC**	Global Precipitation Climatology

	Centre	**GSWD**	Global Surface Water Dataset
GPM	Global Precipitation Measurement	**GREALM**	Global Reservoirs and Lakes Monitor
GPCP	Global Precipitation Climatology Project	**GRRATS**	Global River Radar Altimetry Time Series
GPM IMERG	Global Precipitation Measurement Integrated Multi-satellitE Retrievals for Global Precipitation Measurement	**GSFC**	Goddard Space Flight Center
		GSIM	Global Streamflow Indices and Metadata Archive
		GSWD	Global Surface Water Dataset
GREALM	Global Reservoirs and Lakes Monitor	**GW**	Groundwater
GRACE	Gravity Recovery and Climate Experiment	**GWLA**	GroundWater Level Anomaly
		ICGEM	International Center for Global Earth Models
GRACE-FO	GRACE Follow-On		
GRDC	Global Runoff Data Centre	**ICA**	Independent Component Analysis
GRRATS	Global River Radar Altimetry Time Series	**ICESat**	Ice, Cloud, and Land Elevation Satellite
GRWM	GroundWater Level Monitoring	**ICGEM**	International Center for Global Earth Models
GSFC	Goddard Space Flight Center		
GSIM	Global Streamflow Indices and Metadata Archive	**IGG**	Institute of Geodesy and Geophysics
GSWD	Global Surface Water Dataset	**IERS**	International Earth Rotation and Reference Systems
G-RUN	Global RUNoff	**IRIMO**	Iran's Meteorological Organization
GSWP	Global Soil Wetness Project		
GFZ	German Research Center for Geoscience	**IWRM**	Iran Water Resources Management Company
GNSS	Global Navigation Satellite System	**ITSG**	Institute of Geodesy at Graz University of Technology
GPR	Gaussian Process Regression	**JAXA**	Japan Aerospace Exploration Agency
GPS	Global Positioning System		
G-RUN ENSEMBLE	Global RUNoff ENSEMBLE	**JMA**	Japan Meteorological Agency
		JPL	Jet Propulsion Laboratory
GSIM	Global Streamflow Indices and Metadata Archive	**KBR**	K-band Ranging System
		KGE	Kling-Gupta Efficiency

LRI	Laser-Ranging Interferometers		**PML**	Penman-Monteith-Leuning Evapotranspiration
LST	Land Surface Temperature		**PREC/C**	Precipitation Reconstruction over Land
LSTM	Long Short-Term Memory		**PSDI**	Probabilistic Storage-based Drought Index
LUH	Leibniz Universität Hannover			
MBE	Mean Bias Error		**Q**	Runoff
ML	Machine Learning		**RAI**	Rainfall Anomaly Index
MLR	Multiple Linear Regression		**RF**	Random Forest
MODIS	Moderate Resolution Imaging Spectroradiometer		**RBF**	Radial Basis Function
MSE	Mean Squared Error		**RMS**	Root Mean Squares
MSWEP	Multi-Source Weighted-Ensemble Precipitation		**RMSE**	Root Mean Square Error
			SDI	Storage-based Drought Index
MTA	Mass Trim Assembly		**SDS**	Science Data System
NASA	U.S. National Aeronautics and Space Administration		**SEBAL**	Surface Energy Balance Algorithm for Land
NCEI	National Centers for Environmental Information		**SM**	Soil Moisture
NCEP	National Centers for Environmental Prediction		**SMAC**	Soil Moisture Accounting model
			SMS	Soil Moisture Storage
NDVI	Normalized Difference Vegetation Index		**SPEI**	Standardized Precipitation Evapotranspiration Index
NNLS	None Negative Least Squares		**SPI**	Standardized Precipitation Index
NOAA	U.S. National Oceanic and Atmospheric Administration		**SAR**	Synthetic Aperture Radar
			SARIMAX	Seasonal AutoRegressive Integrated Moving Average with exogenous variables
NSE	Nash-Sutcliffe model efficiency coefficient			
OLTC	Open-Loop Tracking Command		**SMDI**	Soil Moisture Deficit Index
P	Precipitation		**SMAP**	Soil Moisture Active Passive
PDSI	Palmer Drought Severity Index		**SNHT**	Standard Normal Homogeneity Test
P-LSH	Process-based Land Surface ET/Heat		**SNR**	Signal to Noise Ratio
PM	Penman-Monteith method		**SPI**	Standardized Precipitation Index

SPEI   Standardized Precipitation
                Evapotranspiration Index

SPI   Standardized Precipitation Index

SPEI   Standardized Precipitation
                Evapotranspiration Index

SVM   Support Vector Machine

STL   Seasonal and Trend
                decomposition using Loess

SURFEX-TRIP SURFace EXternalisée -
                TRansferts d'Interception et de
                Particules

SWE   Snow Water Equivalent

SWS   Surface Water Storage

SWSI   Surface Water Supply Index

SWAT   Soil and Water Assessment Tool

SWBM   Simple Water Balance Model

SWJTU   Southwest Jiao-Tong University

SWPU   Southwest Petroleum University

SWOT   Surface Water and Ocean
                Topography

TIROS   Television Infrared Observation

                Satellite

TSDI   Total Storage Deficit Index

TU   Technical University

TWS   Total Water Storage

TWSC   Terrestrial Water Storage Change

UKSA   UK Space Agency

UNSD   United Nations Statistics
                Division

USGS   United States Geological Survey

USO   Ultra Stable Oscillator

VCI   Vegetation Condition Index

W3RA   Wide Water Resources
                Assessment

WHU   Wuhan University

WCI   Water Cycle Integrator

WSDI   Water Storage Deficit Index

XISM SSTC   Xi'an Research Institute of
                Surveying and Mapping (XISM)
                Space Star Technology

1

Introduction

1.1. Introduction

The Earth system refers to the interconnected and interrelated processes and components. It includes the physical components of the planet, such as the land, water, and atmosphere, as well as the biological and geological processes that shape it over time. The Earth system can be divided into several subsystems, each of which contributes to the overall functioning of the planet and interact to form a cohesive and interdependent system (Figure 1.1). These subsystems include:

- the geosphere, which encompasses the solid Earth, including the crust, mantle, and core. It exercises a key influence in shaping the Earth's surface through processes such as erosion and tectonics.

- the biosphere, which encompasses all living things on the planet, including plants, animals, and microorganisms, as well as the ecosystems they form. The biosphere plays a crucial role in maintaining the overall health and well-being of the planet.

- the atmosphere, which is the layer of gases surrounding the Earth, composed primarily of nitrogen and oxygen. It plays a vital role in regulating the Earth's temperature, weather patterns, and atmospheric circulation.

- the cryosphere, which includes all the frozen water on Earth, such as ice caps, glaciers, and permafrost. It also plays a decisive part in regulating the Earth's climate, as ice reflects a large amount of incoming solar radiation and thus helps to cool the planet.

- the hydrosphere, which refers to all the water on Earth, including oceans, rivers, lakes, and groundwater. It serves a crucial function in the Earth's climate, as water has a high heat capacity and is able to store and transport heat.
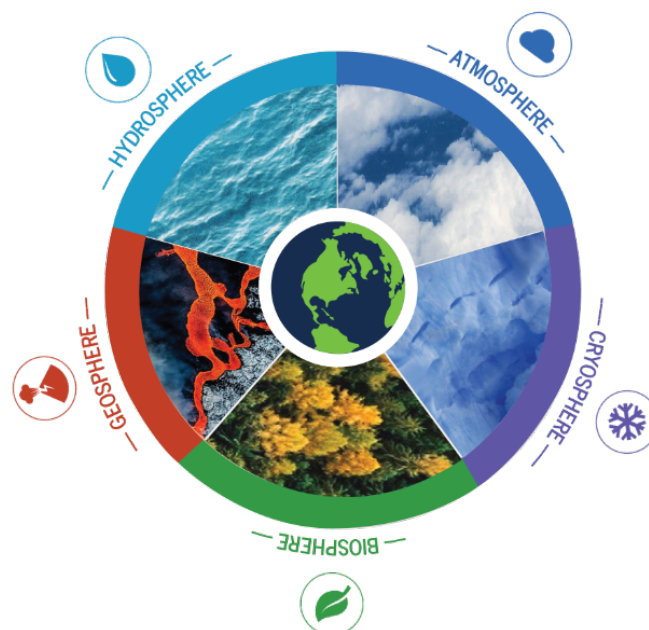

Figure 1.1: The components of the Earth as a system. Courtesy: U.S. National Aeronautics and Space Administration (NASA).

These subsystems interact with one another in complex ways. For example, the emission of greenhouse gases from human activities can alter the composition of the atmosphere and cause global warming, which in turn can lead to melting of the cryosphere and rising sea levels, affecting the geosphere and hydrosphere (Desonie, 2008; Kundzewicz, 2008; Rockström et al., 2009). The Earth system also plays a critical role in regulating the planet's climate and maintaining the conditions necessary for life to thrive. Therefore, understanding the Earth system is crucial for understanding how our planet works, and how human activities can impact it. One of the most important components of the Earth system is the water cycle. Water is constantly moving through the atmosphere, hydrosphere, cryosphere, geosphere, and biosphere, in a process known as the water cycle. In what follows, a brief introduction of the water cycle and its importance in regulating the Earth's climate is presented.

1.2. Water Cycle

The hydrologic cycle, also known as the water cycle, is an ongoing movement of water in the Earth-atmosphere system (Figure 1.3). While the overall volume of water remains constant, it is constantly distributed through various processes such as evaporation, transpiration, sublimation, condensation, precipitation, infiltration, surface runoff, and subsurface flow. The sun's energy heats water at the Earth's surface, causing it to evaporate and rise into the atmosphere as water vapor. In addition to evaporation, plants also release water vapor into the atmosphere through transpiration. A portion of water vapor in the atmosphere also forms directly from the phase change of ice to vapor, known as sublimation. The water vapor released through evaporation, transpiration, and sublimation is less dense than the primary components of the atmosphere, such as nitrogen (N_2) and oxygen (O_2), and is lifted upward by buoyancy. At high altitudes, the water vapor condenses into liquid droplets due to the lower temperature and pressure, forming clouds. These droplets may merge, growing larger, and eventually fall to the ground as hail, rain, sleet, or snow.

The droplets that fall on the Earth's surface can take various paths, such as infiltrating into the soil or permeating through rock, flowing over the ground as surface runoff, falling back into the ocean as rain, or accumulating in ice sheets, glaciers, and snowpacks. Water that infiltrates into the soil increases its moisture, and some may even reach deep enough to recharge aquifers, where it can be stored for long periods. This groundwater may also emerge as freshwater springs or flow into oceans or rivers. Some surface runoff flows into rivers and ultimately discharges into oceans, lakes, man-made reservoirs, and wetlands. Figure 1.2 illustrates the distribution of these processes on both impervious and natural surfaces.

Pools that store water play an essential role in the water cycle, as they act as reservoirs that hold water until it can be released into the environment through various processes. These pools include a variety of water bodies, such as lakes, oceans, and permafrost, as well as the moisture present in the atmosphere. Lakes and oceans are some of the most significant water pools, with oceans containing about 97 % of the world's water. Permafrost, a layer of soil that remains frozen for at least two consecutive years, also serves as a vital water storage pool. Water stored in solid form in glaciers and snowpacks is a critical component of the water cycle, providing a steady source of freshwater to downstream ecosystems and human populations. Moisture in the atmosphere, in the form of water vapor, is another important pool that plays a crucial role in the water cycle, as it forms clouds and precipitation, which ultimately replenish the Earth's water supply. Soil moisture and groundwater are also essential pools in the water cycle, playing a critical role in providing water to plants and animals, as well as replenishing surface water sources. These pools are interconnected, and the movement of water between them is critical for the functioning of the water cycle and sustaining life on our planet.
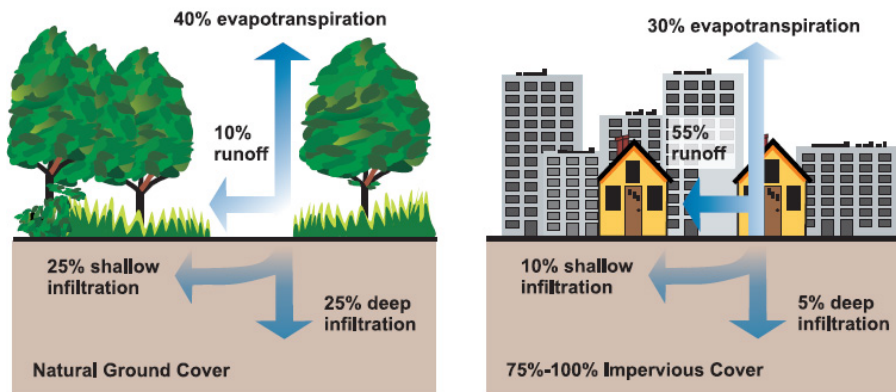
Figure 1.2: Relationship between impervious surfaces and natural surface. Courtesy: U.S. Environmental Protection Agency (EPA).

Fluxes circulate water between storing sections. The water travels between the atmosphere and the land surface via precipitation, evaporation, and transpiration. Across the Earth's surface, snowmelt, runoff, and streamflow are the main drivers the water transpiration and moves to the ground via infiltration and groundwater recharge. Through these processes, water is constantly moving from one storage section to another, helping to regulate the Earth's climate and maintain the conditions necessary for life to thrive. Without water fluxes, the water cycle would not function, and there would be serious consequences for life on Earth.

The water cycle and climate change are closely linked. Climate change is altering the water cycle and its components. As the Earth's temperature increases, it causes more evaporation, which leads to more precipitation in some regions, and less precipitation in others. The altered water cycle can cause changes in the timing, intensity, and distribution of precipitation, which can lead to changes in the amount of water stored in various reservoirs such as surface and subsurface reservoirs, glaciers, ice caps, and the ocean (Levizzani & Cattani, 2019; Wu et al., 2013; Yang et al., 2014). These changes in water storage can have significant impacts on water resources, agriculture, and ecosystems, as well as on natural hazards such as floods and droughts (Kundzewicz, 2008; Yoon et al., 2015). Climate change also affects the amount of water vapor in the atmosphere (Schneider et al., 2010; Voigt & Shaw, 2015), which can lead to changes in the amount of water stored in clouds and precipitation. Additionally, changes in temperature and precipitation can cause changes in the timing and amount of snowmelt, which can have significant impacts on water resources, agriculture, and ecosystems (Adam et al., 2009; Qin et al., 2020; Yang et al., 2022).

Climate change can also affect evapotranspiration, which is the combination of evaporation and plant transpiration, which is an important component of the water cycle (Abtew et al., 2013; Tabari & Talaee, 2014). As the temperature increases, it causes more evaporation, which leads to more evapotranspiration, which can lead to changes in water resources, agriculture, and ecosystems. It is worth noting that, while the water cycle is a global phenomenon, the impacts of climate change on the water cycle are not uniform around the world. Climate change can cause more extreme weather events such as heat waves, heavy precipitation, and prolonged droughts, which can lead to water scarcity in some regions and water excess in others.
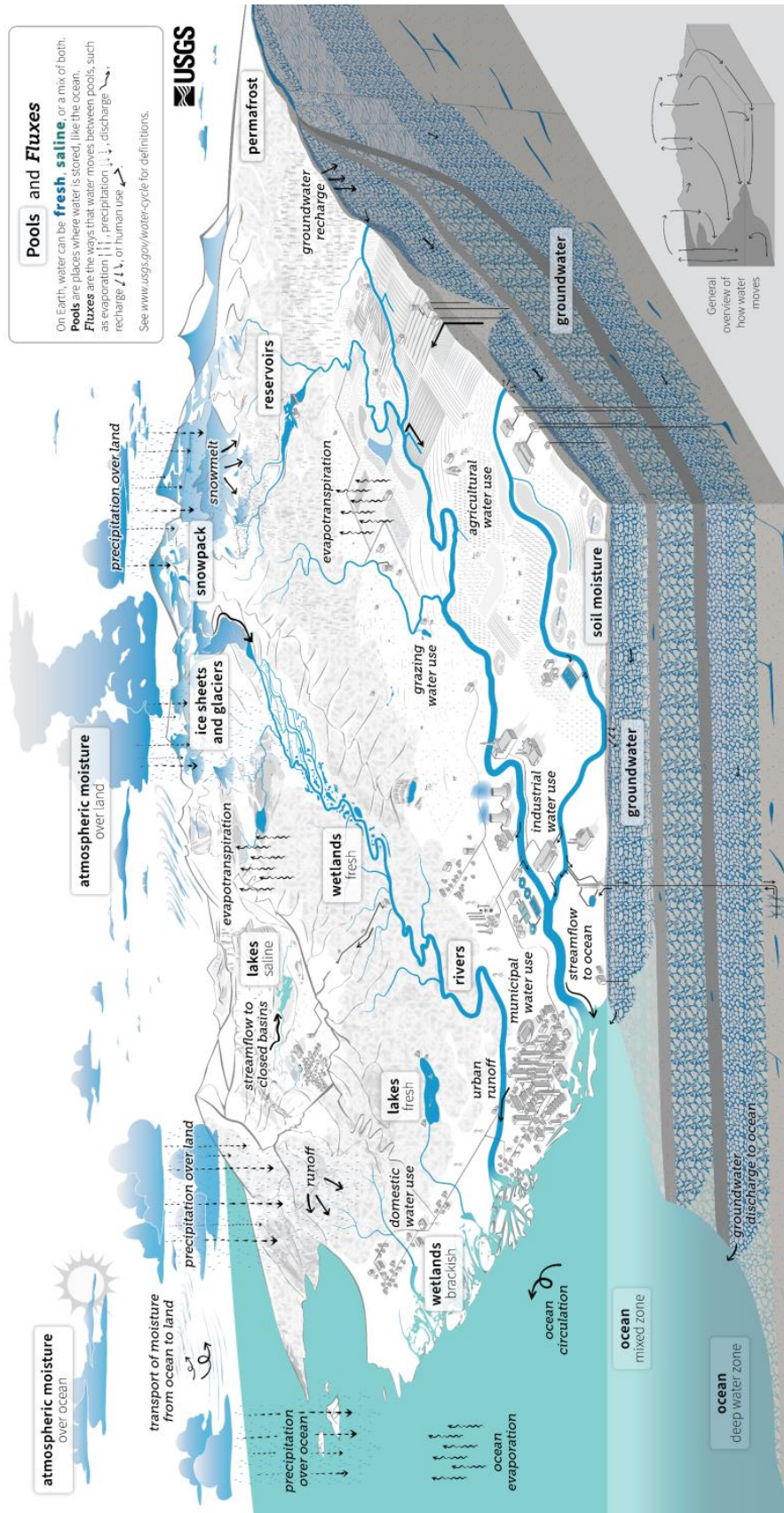
1

Figure 1.3: The water cycle diagram. Courtesy: United States Geological Survey (USGS).

1.3. Total Water Storage

The continental Total Water Storage (TWS) is of crucial importance for the global hydrological cycle, significantly impacting climatic variability, sea level budget, and water resource availability for human life (e.g., Dieng et al., 2017; Jensen et al., 2019; Pellet et al., 2020; Syed et al., 2008; Zhang et al., 2016). The Steering Committee of the Global Climate Observing System (GCOS) has recently recognized TWS as a new Essential Climate Variable (ECV). TWS encompasses all above and below surface water storages, including canopy water, rivers and lakes, soil moisture, groundwater, snow, and ice. Soil water storage is a crucial factor in water and energy fluxes over the land's surface and, alongside precipitation, contributes to hydrological extremes such as droughts and floods. Groundwater storage, or water stored in the saturated zone of soil layers, comprises a significant percentage (about 30%) of total freshwater on Earth and supports agriculture, industry, and domestic needs. Surface water, including water in wetlands, floodplains, lakes, rivers, and man-made reservoirs, supports global agriculture and energy production, particularly in wet tropics and sub-polar regions, while influencing hydro-meteorological and biogeochemical processes (Landerer & Swenson, 2012; Zhao et al., 2021).

TWS is the total of all potential water reserves (Humphrey et al., 2023):

$$\text{TWS} = \text{GW} + \text{SM} + \text{SWE} + \text{SW} + \text{LI} + \text{BW}, \tag{1.1}$$

where GW is the groundwater, SM is the soil moisture, SWE is the snow water equivalent, SW is the surface water, LI is the land ice, and BW is the biomass water.
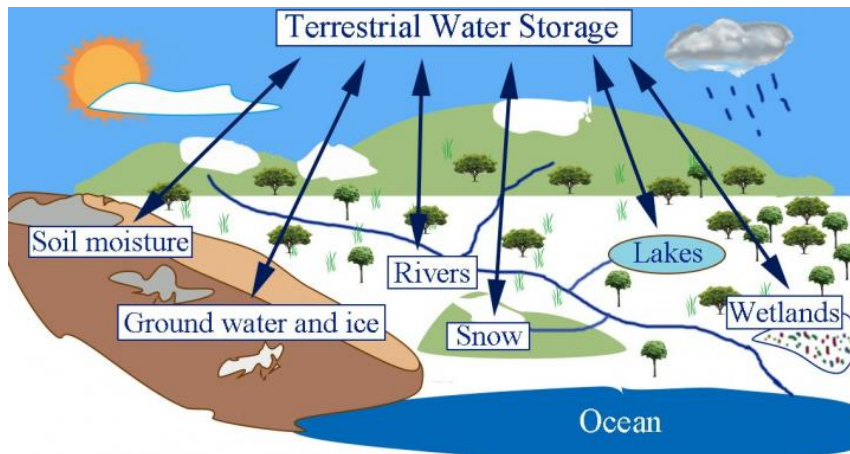
Figure 1.4: The components of the TWS. Source:www.eurekalert.org.

The water balance equation can also be used to estimate the change in TWS by calculating the difference between inputs (precipitation, surface water inflows) and outputs (evapotranspiration, surface water outflows, groundwater recharge) over a specific time period. A generalized equation for a region (like basin or sub-basin) can be written as:

$$\int_{\text{specific time period}} P - \text{ET} - Q \, dt = \Delta \text{TWS} \tag{1.2}$$

where P is Precipitation, ET is Evapotranspiration (ET), and Q is the runoff.

To determine the change in TWS, or ΔTWS, we must measure or estimate the different components of the water balance equation. The precision of our estimate will rely on the quality and accuracy of the data used in the calculations. It is important to note that the water balance equation mentioned above is a general one, and depending on the scale and location of the area being studied, different models may use different formulations and inputs.

1.4. Monitoring Total Water Storage

Monitoring the TWS is of paramount importance for understanding the Earth's climate system and its impact on water resources. Accurate and reliable monitoring of TWS is essential for water resource management, drought and flood early warning, and prediction of future changes in the water cycle (Pellet et al., 2020; Syed et al., 2008). Furthermore, it provides valuable information for agriculture, industry, and domestic needs by providing insights into the availability of freshwater resources (Landerer & Swenson, 2012; Zhao et al., 2021). Such information can be used to improve irrigation and water management strategies, as well as to identify areas that are at risk of water scarcity. Moreover, it facilitates a deeper comprehension of the variability and distribution of water resources, and the correlation between surface and subsurface water storage.

Several methods exist to accurately gauge TWS at global and regional levels:

- Ground-based measurements: This approach uses in-situ measurement of precipitation, evapotranspiration, runoff, and groundwater storage to estimate TWS. These data can be collected from a variety of sources, including weather stations, stream gauges, and wells.

- Hydrological models: This approach uses numerical models to simulate the water cycle and estimate TWS. These models can be driven by both satellite-based remote sensing data and in-situ data, and can be used to estimate TWS at both global and regional scales.

- Space-based observation: This approach uses satellite-borne sensors to measure the gravity field of the Earth, which can be used to infer changes in TWS.

- Data assimilation: This approach combines satellite-based remote sensing data, in-situ data, and hydrological models to estimate TWS. Such a approach allows for the integration of different types of data to improve the accuracy of TWS estimates.

In the following, a succinct overview is presented for each of the previously mentioned methods.

1.4.1. Ground-based measurements approach

The Ground-based measurements approach is a method used to estimate TWS that relies on measurements collected from various in-situ sensors and stations. These measurements include precipitation, evapotranspiration, runoff, soil moisture, and groundwater storage. The data can be collected from a variety of sources, including weather stations, stream gauges, and wells. The history of this approach goes back to the early 20th century, when the first systematic attempts to measure precipitation and streamflow were made. Since then, the network of in-situ data collection has expanded significantly, with the installation of weather stations, stream gauges, and other types of sensors around the world (Robock et al., 2000). With the increasing availability of these data, the ground-based measurements approach has become an important tool for understanding the water cycle and estimating TWS (Vorosmarty et al., 2000).

The advantage of this approach is that it provides high-resolution and accurate data, which

can be used to estimate TWS at a local or regional scale. However, it also has some limitations such as the data availability is limited by the spatial distribution of the measurements and the sensors, and it can be affected by the measurement errors and bias. The development of new and more advanced sensors and measurement devices has led to an increase in the accuracy and resolution of in-situ data. For example, the use of automatic weather stations and high-frequency stream gauges has improved the measurement of precipitation and stream-flow, respectively (Al Sawaf & Kawanisi, 2019; De Vos et al., 2017; Nitu & Wong, 2010; Nsabagwa et al., 2019). Besides, efforts have been made to standardize the measurement methods, instruments, and data formats, allowing for easier data sharing and comparison across different locations and time periods.

Despite the enhancement both in terms of hardware and data management, the number of precipitation and stream-flow gauges has decreased over the last two decades in some regions and countries due to financial constraints, lack of maintenance and obsolescence of the equipment (Strangeways, 2006; Sun et al., 2018; Tourian et al., 2013, e.g.,). Figure 1.5 show the decrease in the number of precipitation stations used by the Global Precipitation Climatology Centre (GPCC) dataset since 1980s after a long-term increase from 1890. Figure 1.6 highlights the decline of in-situ streamflow observation globally, especially after 1980s. Such a continuos decrease in the gourd-based network has led to a significant challenge for the hydrological research and the management of water resources. Moreover it affects the accuracy and reliability of the data, and it may also lead to a gap in the long-term data record that is essential for understanding the variability and change in the hydrological cycle.
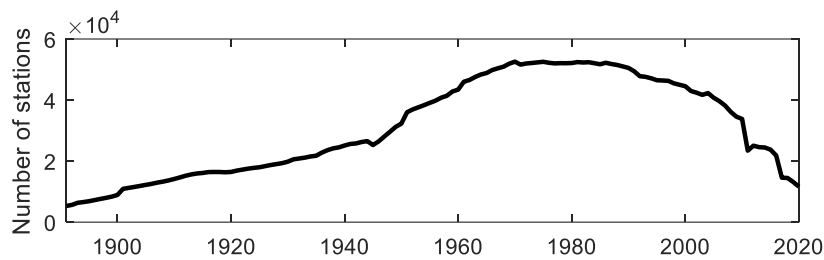
Figure 1.5: The number of monthly precipitation data points in the different databases as a function of time across the period covered by the Global Precipitation Climatology Centre (GPCC) dataset.
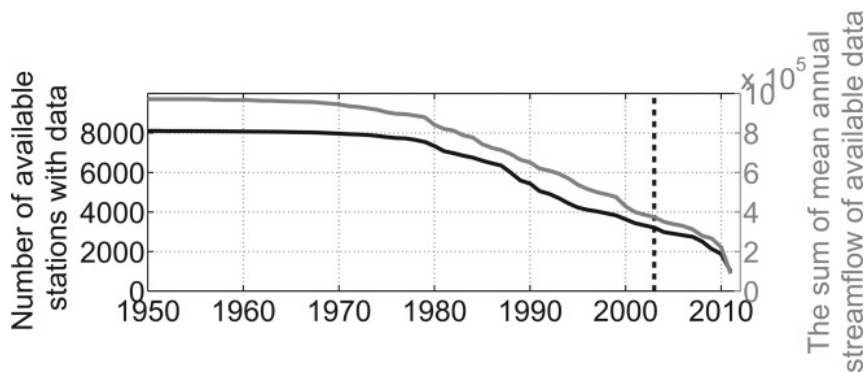
Figure 1.6: The number of Global Runoff Data Centre (GRDC) database stations over time together with the total mean annual streamflow from the stations with available data. Out of the 8424 global stations, only 3323 gauging stations remain in use following the launch of ENVISAT and GRACE, as indicated by the dashed line. Courtesy: Tourian et al. (2013).

1.4.2. Hydrological models

A hydrological model is a mathematical representation of the water cycle, used to estimate TWS and other hydrological variables, such as precipitation, evapotranspiration, and runoff. These models can be classified into different categories based on their complexity, spatial and temporal resolution, and the data inputs required.

- Conceptual models: These models are based on simple physical or empirical relationships and are used to estimate TWS at a regional or catchment scale. They are easy to use and require only a limited amount of input data. Examples of conceptual models include the Simple Water Balance Model (SWBM), the Budyko model (Budyko, 1974), and the Soil Moisture Accounting model (SMAC) (Brutsaert, 1975, 2023).

- Semi-distributed models: These models divide a catchment into multiple sub-catchments and use distributed parameters to represent the spatial variability of the hydrological processes. They are more complex than conceptual models and require more input data, such as land use, soil properties, and topography. Examples of semi-distributed models include the Soil and Water Assessment Tool (SWAT) and the Hydrological Simulation Program - FORTRAN (HSPF) (Arnold et al., 2012)

- Distributed models: These models represent the spatial variability of the hydrological processes at a high resolution, using a grid or mesh of cells. They are the most complex and data-intensive models, requiring large amounts of input data, such as digital elevation models, land use maps, and meteorological data. Examples of distributed models include the Catchment Land Surface Model (CLSM) and the Community Land Model (CLM) (Dai et al., 2003)

- Stochastic models: Stochastic models are a type of hydrological models that incorporate randomness or probability into their calculations. These models are used to simulate and predict the occurrence of natural phenomena, such as precipitation, evaporation, and runoff, which are inherently uncertain. Stochastic models are useful in situations where there is a high degree of uncertainty in the input data, and they can be used to estimate the range of possible outcomes and their likelihoods. These models can also be used to estimate model parameters that are uncertain, such as precipitation and evaporation rates.

In addition to these three main categories, there are also other types of hydrological models, such as hybrid models, which combine elements of different types of models, and data-driven models, which use statistical or machine learning techniques to estimate TWS.

Hydrological models have the advantage of providing detailed information on the water cycle, including the distribution and variability of TWS, and can be used to make predictions of future changes in the water cycle. Despite their ability to provide detailed information on the water cycle, hydrological models present several limitations in accurately estimating TWS. These limitations include:

- Complexity: Hydrological models can be complex, requiring a large amount of input data and computational resources. This can make it difficult to use these models in regions where data is scarce or of poor quality.

- Uncertainty: Hydrological models are based on a set of assumptions and parametrizations, which can introduce uncertainty into the TWS estimates. The accuracy of the model's predictions is dependent on the quality and availability of input data, the model's
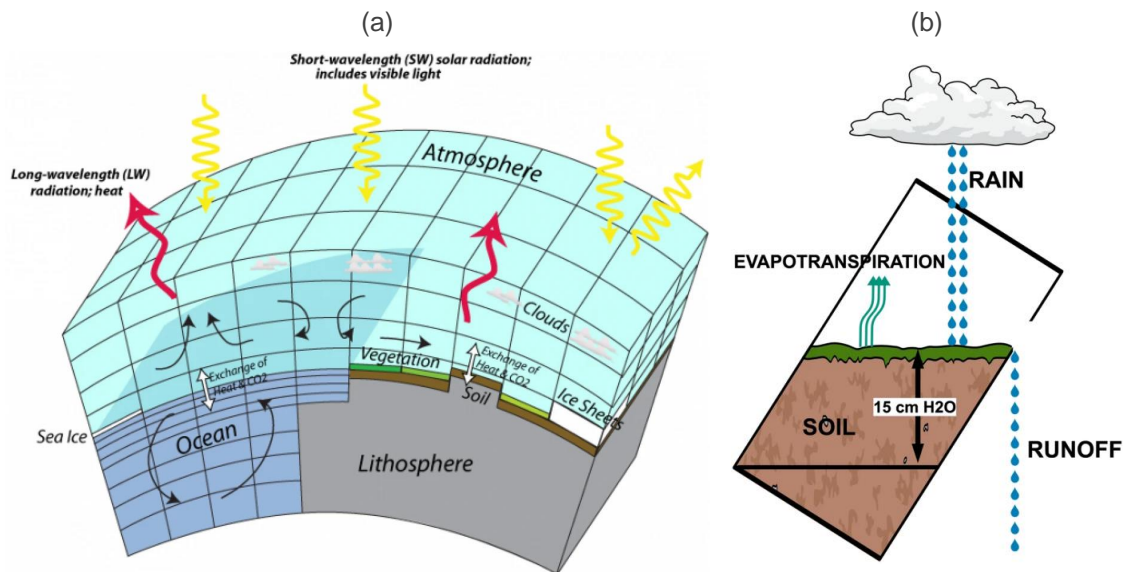
Figure 1.7: (a) A modified version of the schematic structure of a General Circulation Model, as outlined by Vargas Godoy et al. (2021). (b) A diagram that illustrates the Budyko bucket model, which was developed by Manabe (1969). It depicts a single layer of soil that acts as a reservoir with a maximum capacity of 15 cm for field water. The rate of evaporation from the soil is directly proportional to the amount of water remaining in the reservoir.

structure and parametrization, and the skill of the modeler (Sherwood et al., 2020).

- Lack of flexibility: Hydrological models are designed to simulate specific hydrological processes and may not be able to account for changes in the water cycle due to climate change or human activities.

- Lack of validation: Many hydrological models are not validated against independent data, making it difficult to assess their accuracy and reliability.

- Data requirements: Hydrological models require a large amount of input data, including meteorological, land use, and topographical data, which can be difficult to obtain in some regions.

1.4.3. Space-based TWS change observation

Satellites are widely used to measure various water fluxes such as precipitation, evapotranspiration, and runoff, which are important components of the water cycle. These measurements can be used to estimate the water balance, which is the balance of water inflows and outflows on a global scale. The earliest attempts to estimate the water cycle using satellites were focused on measuring precipitation. In the 1970s, the first geostationary weather satellites, such as the U.S. National Oceanic and Atmospheric Administration's (NOAA) Television Infrared Observation Satellite (TIROS), were launched (Schwalb, 1978). These satellites provided measurements of cloud cover and precipitation, which were used to estimate precipitation rates. In the following decades, new satellite missions were launched that improved the accuracy of precipitation estimates. For example, the launch of the NASA's Tropical Rainfall Measuring Mission (TRMM) in 1997 (Simpson et al., 1988), provided measurements of precipitation over the tropics and subtropics, which are regions that are particularly challenging to measure precipitation (e.g., AghaKouchak et al., 2009; Maggioni et al., 2016). In the early 2000s, the Global Precipitation Measurement (GPM) mission was developed as a joint mission by NASA and the Japan Aerospace Exploration Agency (JAXA), to provide global coverage of precipitation (Skofronick-

Jackson et al., 2017; Tapiador et al., 2012) (Figure 1.8). The GPM mission provided more accurate measurements of precipitation, by combining data from multiple instruments and satellites.
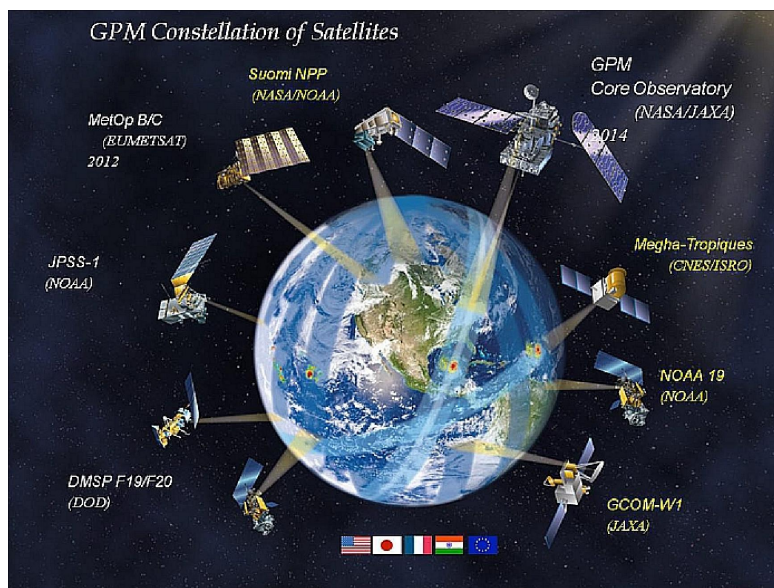

Figure 1.8: GPM (Global Precipitation Measurement) Mission constellation satellites. Source: NASA.

In the 1980s and 1990s, the focus of estimating water balance fluxes shifted to measuring evaporation. The earliest attempts to estimate evapotranspiration from satellite observations focused on using measurements of surface temperature and vegetation cover. For example, the Normalized Difference Vegetation Index (NDVI) was developed, which uses measurements of visible and infrared radiation to estimate vegetation cover (Myneni et al., 1995; Running, 1990). NDVI values are then related to evapotranspiration through empirical relationships (e.g., Nemani & Running, 1989; Rossato et al., 2005). However, these early methods had limited accuracy and were affected by various sources of error. In the late 1990s, new satellite missions were launched that provided measurements of surface wind speed and direction, which improved the accuracy of evapotranspiration estimates. For example, the National Aeronautics and Space Administration's (NASA) Scatterometer (NSCAT) was launched in 1996, and the European Space Agency's (ESA) Advanced Scatterometer (ASCAT) in 1999. These missions provided measurements of surface wind speed and direction, which can be used to estimate evaporation rates through incorporating the satellite data in the land surface models (Evans, 1993; Yagci & Yilmaz, 2021).

Since then, new satellite missions have been launched that have improved the ability to estimate evapotranspiration. For example, the European Space Agency's Sentinel-1, Sentinel-2 and Sentinel-3 missions provide information on vegetation cover, soil moisture, temperature, and surface wind speed and direction, which can be used to estimate evapotranspiration (e.g., Guzinski et al., 2020; Vanino et al., 2018). NASA's Soil Moisture Active Passive (SMAP) mission, provide measurements of soil moisture, which is important for understanding evapotranspiration (e.g., Purdy et al., 2018; Walker et al., 2019). Empirical or theoretical models, such as the Penman-Monteith method (PM) (Monteith, 1965) and the Surface Energy Balance Algorithm for Land (SEBAL) (Bastiaanssen, Menenti, et al., 1998; Bastiaanssen, Pelgrum, et al., 1998), have been developed to convert satellite imageries observations into ET estimates. In recent years, the use of remote sensing data and machine learning techniques has been developed to estimate evapotranspiration, this method combines multi-sensor data, such as NDVI, Land

Surface Temperature (LST), and surface wind speed, as well as meteorological data to estimate evapotranspiration (e.g., Chia et al., 2020; Dou & Yang, 2018; Granata, 2019). The use of satellite imagery provides a cost-effective and efficient means of estimating ET over large areas, as it eliminates the need for ground-based measurements, which are both time-consuming and labor-intensive.
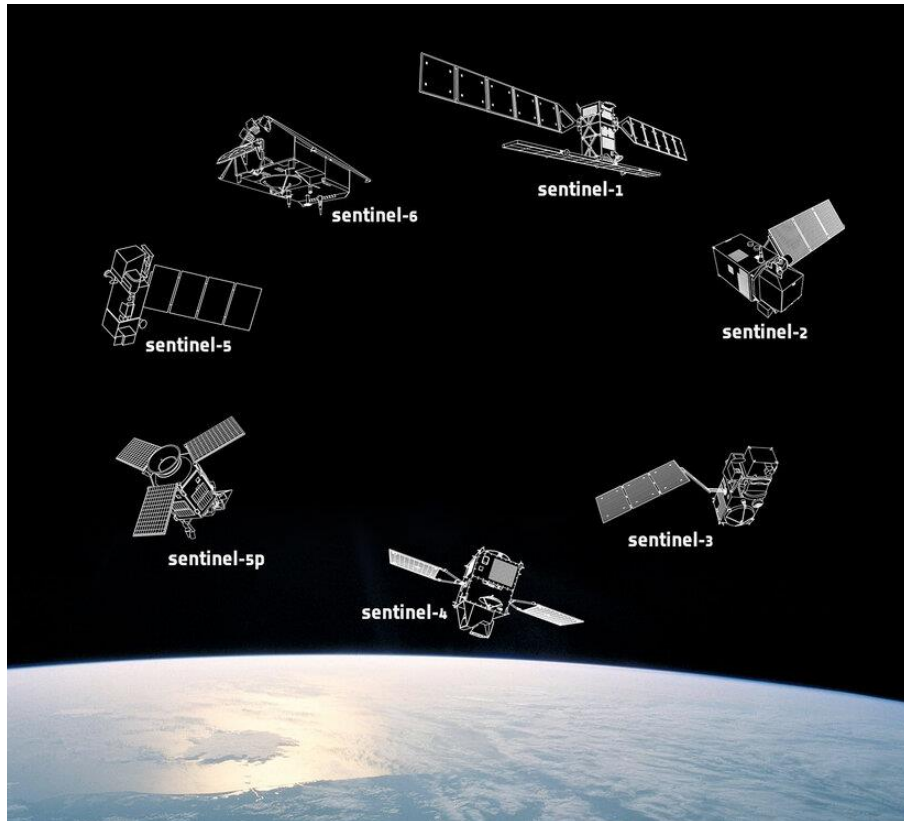

Figure 1.9: The Sentinel missions. Source: the European Space Agency (ESA).

The last term of the water balance fluxes that became observable by satellite is the runoff. The history of using satellite observations for estimating runoff, which is the amount of water that flows over land and into rivers, lakes and oceans, is relatively recent. The earliest attempts to estimate runoff from satellite observations focused on using measurements of surface water bodies, such as lakes and rivers. For example, the European Space Agency's (ESA) ERS-1 and ERS-2 satellites, launched in 1991 and 1995, respectively, provided measurements of surface water bodies, which were used to estimate the storage of water in these surface reservoirs (e.g., Nagler & Rott, 1997; Nagler et al., 2000). However, the spatial resolution of these early measurements was relatively low, and the estimates of runoff were affected by various sources of error. In the early 2000s, new satellite missions were launched that provided measurements of surface water bodies, such as the European Space Agency's Sentinel-1 and Sentinel-2, which improved the accuracy of runoff estimates (Brombacher et al., 2020; Tarpanelli et al., 2022). These missions provide high-resolution Synthetic Aperture Radar (SAR) images, which can be used to estimate the storage of water in these surface reservoirs.

In recent years, new satellite missions have been launched that have improved the ability to estimate runoff. For example, the Sentinel-6 Michael Freilich satellite, launched in November 2020. This satellite is part of the European Space Agency's Copernicus program and is a collaboration between ESA, NASA, European Organisation for the Exploitation of Meteorological

Satellites (EUMETSAT), and the the French space agency i.e., Centre National d'Etudes Spatiales (CNES). Although the main mission of Sentinel-6 is to measure the height of the ocean surface with high accuracy and precision, it will also be able to measure the height of river and lake surfaces, providing data on the water level changes, which can be used to estimate the runoff from these sources as well. Last but not least is the NASA's Surface Water and Ocean Topography (SWOT) mission, was launched on November 21, 2021, which will provide measurements of surface water bodies, including oceans, lakes, and rivers, to estimate the storage of water in these surface reservoirs and the amount of water that is discharged as runoff (Tourian et al., 2021).
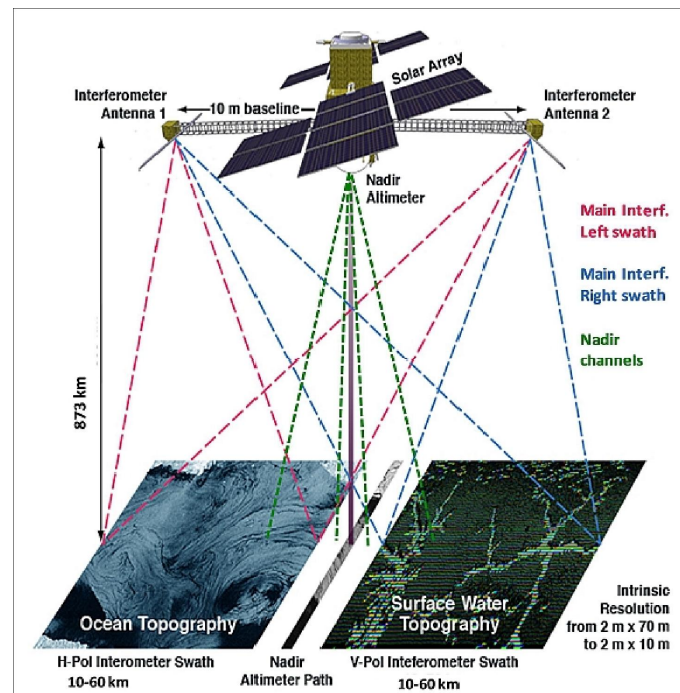

Figure 1.10: The SWOT (Surface Water Ocean Topography) measurement concept. Source: NASA, CNES

Other than the attempts to estimate the water balance fluxes using spaceborne observations, several methods have been developed to quantify the TWS state variables such as surface water storage or soil moisture:

- Measuring surface water bodies: Satellites such as Sentinel-1 and Sentinel-2, Landsat, Moderate Resolution Imaging Spectroradiometer (MODIS), and SWOT mission, provide measurements of surface water bodies, such as lakes and rivers. These measurements can be used to estimate the storage of water in these surface reservoirs (e.g., Duan & Bastiaanssen, 2013; Papa & Frappart, 2021; Tortini et al., 2020).

- Measuring the water level: Satellites altimetry missions such as ENVISAT, Sentinel-3A, Sentinel-3B, and CryoSat-2 provide measurements of the water level for the inland water bodies, which can be used to estimate total water storage over rivers and lakes (e.g., Ni et al., 2017; Song et al., 2013; Zhang et al., 2006).

- Measuring ice elevation: Satellites such as NASA's Ice, Cloud, and Land Elevation Satellite (ICESat-2) can be used to measure the elevation of ice sheets, glaciers, and ice caps. These measurements can be used to estimate the amount of water stored in these ice reservoirs (e.g., Fricker et al., 2021; Kwok et al., 2019; Markus et al., 2017).

- Measuring soil moisture: Satellites such as NASA's Soil Moisture Active Passive (SMAP) mission, provide measurements of the amount of water in the top few centimeters of soil (e.g., Chan et al., 2016; Entekhabi et al., 2010; Zhang & Zhou, 2016).

The above-mentioned spaceborn measurements, either estimating the TWS state components such as soil moisture or the water balance fluxes would not deliver a holistic view of the TWS change. One other approach that can solve this problem would be the estimation of the temporal gravity field which can be then inferred as the TWS change. The static gravity field which refers to the Earth's gravity field as it exists at a given point in time, without taking into account any changes that may occur over time. However, the temporal gravity field refers to the variations in the Earth's gravity field that occur over time, due to changes in the distribution of mass on the Earth's surface. These changes in mass distribution can be caused by various factors, such as changes in the amount of water stored in surface and subsurface reservoirs, changes in the amount of ice stored in glaciers and ice sheets, and changes in the amount of water stored in the oceans. The temporal gravity field is typically measured using satellite gravimetry, which is the measurement of the Earth's gravity field from space, and it allows monitoring changes in TWS over time.

The use of the Earth's gravity field for estimating TWS dates back to the early 20th century, when scientists first began to use gravity measurements to study the Earth's crust and subsurface. However, it wasn't until the advent of satellite technology in the late 20th century that the use of the Earth's gravity field for estimating TWS became possible. The CHAllenging Minisatellite Payload (CHAMP), launched in 2000 by the German Aerospace Center, was the first satellite mission to provide precise measurements of the Earth's magnetic and gravity field. CHAMP's measurements were used to study the Earth's magnetic field and its interaction with the solar wind, as well as the gravity field's variations due to the Earth's internal structure and mass redistribution in the atmosphere and oceans. The Gravity field and steady-state Ocean Circulation Explorer (GOCE),launched in 2009 by the European Space Agency, provided the highest accuracy and spatial resolution gravity field measurements to date. It was designed to measure the gravity field signals from Earth's crust and mantle, ocean circulation, and ice mass changes. GOCE's data was crucial in understanding the Earth's geoid and improving our knowledge of the planet's interior structure.

The Gravity Recovery and Climate Experiment (GRACE), which was a joint mission between NASA and the German Aerospace Center (DLR) launched in 2002. GRACE provided measurements of the Earth's gravity field, which were used to estimate changes in TWS, mainly in surface and subsurface reservoirs such as groundwater, lakes, and rivers. The successor of the GRACE mission, GRACE Follow-On, was launched in 2018. While GOCE and CHAMP have contributed significantly to our understanding of the Earth's gravity field, they differ from GRACE and its follow-on mission, GRACE Follow-On (GRACE-FO), in several ways. GOCE was primarily designed to measure the static part of the gravity field, while GRACE and GRACE-FO were designed to measure the time-varying part of the gravity field due to mass redistributions. Similarly, CHAMP was not specifically designed to measure the time-varying part of the gravity field related to TWS changes.

The use of the Earth's gravity field for estimating TWS has become an essential tool for monitoring and understanding TWS changes and its variability. The method has been used to estimate TWS changes in various regions of the world, such as the Amazon Basin, the Arctic, the Himalayas, and the major rivers basins in the world. It has also been used to estimate TWS changes due to climate change, such as the melting of glaciers and ice sheets, and changes in groundwater storage. Other than the GRACE satellite, no other satellite provides a holistic view

of TWS change. Other satellite missions, such as those that measure surface water bodies, sea surface height, ice elevation, soil moisture, and salinity, use different measurement principles to estimate TWS. These measurements are sensitive to specific components of the water cycle such as surface water, soil moisture, ocean water and ice. Moreover, subsurface water storage such as groundwater and deep soil moisture is a crucial component of the TWS. The satellites other GRACE fail to observe it and it is challenging to estimate it through other means than satellite gravimetry.
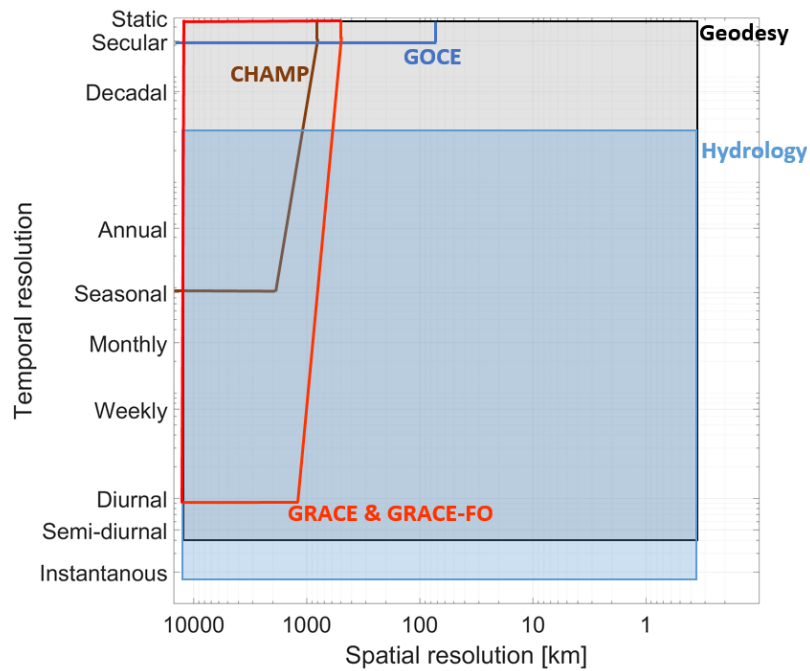

Figure 1.11: Comparison of the spatio-temporal resolution of the satellite gravimetry missions including GOCE, CHAMP, and GRACE(-FO). The region of interest for geodesy and hydrology is shown with the background colors. Courtesy: Mohammad J. Tourian.

1.5. Motivation and Objectives

GRACE and GRACE-FO provide unique insights into the Earth system and have been used for various applications:

- Understanding the water cycle: TWS is an essential component of the water cycle, and GRACE provides a unique way to measure changes in TWS over time and across the globe. GRACE data has been used to estimate TWS changes in different regions of the world and is shown that such a variation can be attributed to both natural and human-induced factors such as climate variability and human activities such as irrigation and dam building (e.g., Forootan et al., 2014; Hosseini-Moghari et al., 2020; Joodaki et al., 2014; Rodell et al., 2018; Tourian et al., 2015; Voss et al., 2013). Understanding these changes can provide important information for comprehending the water cycle and its variability, which is crucial for water resource management and climate change studies.

- Groundwater monitoring: GRACE is particularly useful for monitoring changes in subsurface water storage, such as groundwater (e.g., Chen et al., 2016; Richey et al., 2015; Thomas & Famiglietti, 2019; Thomas et al., 2017). For example, Scanlon et al. (2012) used GRACE data to estimate changes in groundwater storage in the United States and found that groundwater depletion is a significant problem in the western and southern parts of the country. Similarly, (Rodell et al., 2009) found similar signature of the groundwater depletion in GRACE signal over India. Groundwater is an important source of water for many regions of the world, and changes in groundwater storage can have significant impacts on water resources, agriculture, and ecosystems.

- Climate change studies: GRACE can also be used to measure changes in TWS due to climate change, such as the melting of glaciers and ice sheets, and changes in soil moisture (Tapley et al., 2019). The first direct measurement of ice mass change was made possible by GRACE, including clear signal of ice mass loss in Greenland and Antarctica (Velicogna & Wahr, 2005, 2006). Besides, GRACE allowed for a direct correlation to be made between inter-annual fluctuations in ice sheet mass and global variability in atmospheric circulation patterns, in addition to long-term trends (Hanna et al., 2014). Moreover, GRACE is regularly employed in conjunction with ocean hydrographic profiles from Argo to investigate the worldwide sea-level budget, which improves our comprehension of how contributions are shifting with time (Wouters, van de Wal, et al., 2018). Understanding these changes can provide important information for understanding the impacts of climate change on water resources and ecosystems.

- Natural hazards monitoring: GRACE data can provide valuable information for understanding the dynamics of natural hazards and for improving hazard assessments and risk management. GRACE data have been used to study volcanic activity and coastal land subsidence. Moreover, studies have shown that GRACE data can be used to detect changes in surface deformation caused by earthquakes (Chen et al., 2007; Wang et al., 2012), volcanic eruptions (Li et al., 2023), and ground water withdrawal (Castellazzi et al., 2016; Liu et al., 2014; Zheng et al., 2018). GRACE data is vital for monitoring and managing droughts and floods, enabling scientists to pinpoint areas at risk of water stress and track changes in water availability, vital for effective water allocation. The comprehensive datasets provided by GRACE allow policymakers to make informed decisions on water policy and management, reducing the negative impacts of natural disasters such as droughts and floods.

Despite its invaluable impacts, GRACE(-FO) faces a number of challenges that must be addressed in order to maintain the accuracy and reliability of its measurements. This thesis is motivated to address some of these challenges, including:

The GRACE(-FO) post-processing approaches and consistency evaluation

Since the early days of the GRACE mission, researchers have commonly extracted mass changes from observations by approximating the Earth's gravity field using mathematical functions called spherical harmonics. Considering the constraints, approaches, and models that have been used, several institutes have already developed their own processed GRACE(-FO) data. This processed data, known as level-2 data in the GRACE community, requires post-processing to be used for various applications, including hydrology and climate research (Swenson & Wahr, 2006). For example, the GRACE Level-2 solutions contain a significant amount of noise and errors, which can reduce the accuracy of gravity field estimates (Devaraju, 2015). To address this issue, various filters have been developed and can be applied during the post-processing stage (e.g., Jekeli, 1981; Klees et al., 2008; Kusche et al., 2009; Yi & Sneeuw, 2022). The choice of filter can depend on the specific application and the noise characteristics in the data (Devaraju & Sneeuw, 2017). Consequently, post-processed data, or so-called level-3 products, of the GRACE(-FO) mission are not unique and depend on the post-processing steps chosen.

In this thesis, we evaluate different methods of processing GRACE(-FO) level-2 data and assess the spatio-temporal effect of the post-processing steps. In addition, we aim to compare the consistency between GRACE and its successor mission, GRACE-FO, in terms of the data quality and measurement accuracy. By analyzing and comparing the data from these two missions, we can identify any potential discrepancies or differences and establish the level of confidence in the accuracy and reliability of the GRACE-FO measurements. Furthermore, the processed level-3 products will be compared to the level-3 products that are currently accessible online.

Expanding the short record of GRACE(-FO) observations

The GRACE mission, launched in 2002, was designed to last for 5 years, but it was extended multiple times and it operated until October 2017. However, even with the multiple extensions, the record of GRACE measurements is relatively short compared to other satellite missions and observational records. This short record can be a limitation for some studies that require long-term data, such as studying long-term trends in sea level rise, ice sheet mass loss, or groundwater depletion (e.g., Reager & Famiglietti, 2009; Scanlon et al., 2015; Tapley et al., 2019; Thomas et al., 2014). The short record also makes it more difficult to separate the long-term signals from the short-term variability, which can affect the accuracy of the gravity field solutions (Eicker et al., 2016). Additionally, the short record can make it more challenging to compare the GRACE data with other observational records, such as ground-based measurements or other satellite missions, which can limit the ability of researchers to validate the GRACE data and gain a more complete understanding of the Earth's gravity field.

This thesis endeavors to address the limited temporal scope of GRACE observations by leveraging the capabilities of global hydrological, atmospheric, and reanalysis models. To attain this objective we employ various methods and compare our results with the ensemble mean and ensemble weighted mean of the datasets, which serves as the naive method. Additionally, we evaluate the performance of the methods against Satellite Laser Ranging (SLR) observations. The hindcasted GRACE, GRACEH, not only expands the temporal coverage of GRACE(-FO) observations but also provides a better understanding of the changes in the Earth's water storage on a longer time scale.

GRACE's limitations and opportunities in differentiating water storage components

The GRACE satellite mission can only detect changes in the overall water storage (TWS) in a specific region and is unable to differentiate between variations in different compartments of TWS, such as variations in surface water, groundwater, or soil moisture (Rodell et al., 2009; Wada et al., 2013; Yeh et al., 2006). For instance, when a region undergoes an increase in total water storage, it could be a result of an increase in surface water, groundwater, or soil moisture. Without supplementary data or measurements, it is impossible to determine which component of water storage is changing. This poses a significant constraint when attempting to comprehend the movement of water throughout the Earth's system, as variations in different compartments of water storage can have distinct impacts on water resources and the environment. For example, an increase in surface water could indicate a flood event, while an increase in groundwater could signify a sustainable increase in water resources. Understanding the individual components of water storage is crucial for managing water resources and addressing the effects of droughts and floods (Landerer & Swenson, 2012; Reichle & Koster, 2005; Sheffield et al., 2012).

The other objective of this thesis is to build upon the near two-decade record of TWS change observations from GRACE, with the aim of improving our understanding of water storage variations at the continental to basin scale. To achieve this objective, we integrate a myriad of water fluxes, lake water level, and lake storage change data and measurements. Additionally, we demonstrate how the integration of GRACE(-FO) observations in conjunction with peizometric wells and rain-gauges measurements, prove instrumental in comprehending the water scarcity predicament in Iran. This case study highlights the alarming negative trends in the Middle East region and serves as a prime example of how scrutinizing GRACE(-FO) observations can furnish vital insights for managing water resources in regions facing similar challenges.

Accounting GRACE(-FO)'s uncertainty in drought characterization

The GRACE satellite mission, like any other scientific measurement, has its own level of uncertainty. While the mission has been successful in providing valuable insights into the Earth's system, the accuracy and precision of the data obtained by GRACE are not perfect. There are several sources of uncertainty associated with the GRACE data, including measurement error, instrument noise, and uncertainties in the models and algorithms used to process the data (Tapley et al., 2004; Tapley et al., 2019; Wahr et al., 1998). The GRACE mission has taken various measures to minimize these uncertainties, such as calibrating the instruments and using multiple models to process the data (Devaraju, 2015; Tourian, 2013). However, it is still important for researchers to account for these uncertainties when analyzing the GRACE data, as they can affect the accuracy and reliability of the results. It is also essential to communicate the uncertainty associated with the GRACE data when reporting findings based on the mission's observations.

This thesis aims to address the uncertainty that exists in the GRACE(-FO) TWSA data, and proposes a probabilistic approach to utilize this data in characterizing storage-based drought. Traditional drought indices such as the Standardized Precipitation Index (SPI) or the Palmer Drought Severity Index (PDSI) do not take into account the availability of water in storage, and may lead to an incomplete understanding of drought conditions. By incorporating the TWSA data and accounting for its uncertainty, the proposed approach can provide a more comprehensive and informative assessment of drought risk. This will be particularly beneficial for decision makers who rely on accurate and timely information for managing water resources and responding to drought events.

1.6. Outline of the thesis

In line with the aforementioned motivations and objectives, the following outline has been devised:

- **Chapter 2** provides an overview of the GRACE and GRACE-FO missions, including an introduction to the different levels of GRACE(-FO) data. The post-processing of GRACE level-2 products is then explained step-by-step, with an assessment of the spatio-temporal analysis of the consequences of different choices in the post-processing steps on the final result at a global scale. Additionally, the chapter examines the coherence and consistency of the accuracy between GRACE and its successor GRACE-FO. Finally, the result from the post-processing scheme offered in this study is compared to the level-3 products that are already available publicly, providing valuable insights into the performance of the proposed scheme.

- **Chapter 3** aims to extend the relatively short record of GRACE observations back to 1980. The chapter highlights the potential pitfalls of relying solely on GRACE(-FO) observations for climate-related assessments and hydrological extreme characterizations. To overcome these limitations, a range of global models are presented, followed by an analysis of their strengths and weaknesses. Different hindcasting methods, including the Gaussian Process Regression (GPR), are then introduced and their results are compared to assess over selected basins and globally.

- **Chapter 4** analyzes TWS variation from 2003 to 2021, examining TWS variations and separating them into their components, namely precipitation, evapotranspiration, and runoff across major river basins and sub-continents. Additionally, further investigations are conducted into the water storage variation in Iran, utilizing auxiliary data such as precipitation gauges and piezometric groundwater wells to augment our understanding of the data.

- **Chapter 5** provides an introduction on drought in general and proposes a probabilistic approach to utilize GRACE(-FO) TWSA data in characterizing storage-based drought, which can provide a more comprehensive assessment of drought risk for decision makers. The aim is to contribute to the development of more effective monitoring and risk management strategies for water resources.

- **Chapter 6** succinctly encapsulates the accomplishments of this thesis, draws insightful conclusions, and sheds light on future prospects.

Bibliography

Abtew, W., Melesse, A., Abtew, W., & Melesse, A. (2013). Climate change and evapotranspiration. *Evaporation and evapotranspiration: Measurements and estimations*, 197–202. https://doi.org/10.1007/978-94-007-4737-1_13

Adam, J. C., Hamlet, A. F., & Lettenmaier, D. P. (2009). Implications of global climate change for snowmelt hydrology in the twenty-first century. *Hydrological Processes: An International Journal, 23*(7), 962–972. https://doi.org/10.1002/hyp.7201

AghaKouchak, A., Nasrollahi, N., & Habib, E. (2009). Accounting for uncertainties of the TRMM satellite estimates. *Remote sensing, 1*(3), 606–619. https://doi.org/10.3390/rs1030606

Al Sawaf, M. B., & Kawanisi, K. (2019). Novel high-frequency acoustic monitoring of streamflow-turbidity dynamics in a gravel-bed river during artificial dam flush. *Catena, 172*, 738–752. https://doi.org/10.1016/j.catena.2018.09.033

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R., Van Griensven, A., Van Liew, M. W., et al. (2012). SWAT: Model use, calibration, and validation. *Transactions of the ASABE, 55*(4), 1491–1508. https://doi.org/10.13031/2013.42256

Bastiaanssen, W. G., Menenti, M., Feddes, R., & Holtslag, A. (1998). A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *Journal of hydrology, 212*, 198–212. https://doi.org/10.1016/S0022-1694(98)00253-4

Bastiaanssen, W. G., Pelgrum, H., Wang, J., Ma, Y., Moreno, J., Roerink, G., & Van der Wal, T. (1998). A remote sensing surface energy balance algorithm for land (SEBAL).: Part 2: Validation. *Journal of hydrology, 212*, 213–229. https://doi.org/10.1016/S0022-1694(98)00254-6

Brombacher, J., Reiche, J., Dijksma, R., & Teuling, A. J. (2020). Near-daily discharge estimation in high latitudes from Sentinel-1 and 2: A case study for the Icelandic Þjórsá river. *Remote sensing of Environment, 241*, 111684. https://doi.org/10.1016/j.rse.2020.111684

Brutsaert, W. (1975). A theory for local evaporation (or heat transfer) from rough and smooth surfaces at ground level. *Water resources research, 11*(4), 543–550. https://doi.org/10.1029/WR011i004p00543

Brutsaert, W. (2023). *Hydrology*. Cambridge university press. https://doi.org/10.1017/9781316471562

Budyko, M. I. (1974). *Climate and Life*. Academic press.

Castellazzi, P., Martel, R., Galloway, D. L., Longuevergne, L., & Rivera, A. (2016). Assessing groundwater depletion and dynamics using GRACE and InSAR: Potential and limitations. *Groundwater, 54*(6), 768–780. https://doi.org/10.1111/gwat.12453

Chan, S. K., Bindlish, R., O'Neill, P. E., Njoku, E., Jackson, T., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., et al. (2016). Assessment of the SMAP passive soil moisture product. *IEEE Transactions on Geoscience and Remote Sensing, 54*(8), 4994–5007. https://doi.org/10.1109/TGRS.2016.2561938

Chen, J. L., Wilson, C. R., Tapley, B. D., & Grand, S. (2007). GRACE detects coseismic and post-seismic deformation from the Sumatra-Andaman earthquake. *Geophysical Research Letters, 34*(13). https://doi.org/10.1029/2007GL030356

Chen, J., Famigliett, J. S., Scanlon, B. R., & Rodell, M. (2016). Groundwater storage changes: present status from GRACE observations. In *Remote sensing and water resources* (pp. 207–227). Springer.

Chia, M. Y., Huang, Y. F., Koo, C. H., & Fung, K. F. (2020). Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy*, *10*(1), 101. https://doi.org/10.3390/agronomy10010101

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. (2003). The common land model. *Bulletin of the American Meteorological Society*, *84*(8), 1013–1024. https://doi.org/doi.org/10.1175/BAMS-84-8-1013

De Vos, L., Leijnse, H., Overeem, A., & Uijlenhoet, R. (2017). The potential of urban rainfall monitoring with crowdsourced automatic weather stations in Amsterdam. *Hydrology and Earth System Sciences*, *21*(2), 765–777. https://doi.org/10.5194/hess-21-765-2017

Desonie, D. (2008). *Climate: Causes and effects of climate change*. Infobase Publishing.

Devaraju, B. (2015). *Understanding filtering on the sphere: Experiences from filtering GRACE data*. https://doi.org/10.18419/opus-3985

Devaraju, B., & Sneeuw, N. (2017). The polar form of the spherical harmonic spectrum: implications for filtering GRACE data. *Journal of Geodesy*, *91*(12), 1475–1489. https://doi.org/10.1007/s00190-017-1037-7

Dieng, H.-B., Cazenave, A., Meyssignac, B., & Ablain, M. (2017). New estimate of the current rate of sea level rise from a sea level budget approach. *Geophysical Research Letters*, *44*(8), 3744–3751. https://doi.org/10.1002/2017GL073308

Dou, X., & Yang, Y. (2018). Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. *Computers and Electronics in Agriculture*, *148*, 95–106. https://doi.org/10.1016/j.compag.2018.03.010

Duan, Z., & Bastiaanssen, W. (2013). Estimating water volume variations in lakes and reservoirs from four operational satellite altimetry databases and satellite imagery data. *Remote Sensing of Environment*, *134*, 403–416. https://doi.org/10.1016/j.rse.2013.03.010

Eicker, A., Forootan, E., Springer, A., Longuevergne, L., & Kusche, J. (2016). Does GRACE see the terrestrial water cycle "intensifying"? *Journal of Geophysical Research: Atmospheres*, *121*(2), 733–745. https://doi.org/10.1002/2015JD023808

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., Entin, J. K., Goodman, S. D., Jackson, T. J., Johnson, J., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Evans, D. L. (1993). Earth Remote Sensing at the Jet Propulsion Laboratory.

Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M., Awange, J., Schmidt, M., Omondi, P., & Famiglietti, J. (2014). Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sensing of Environment*, *140*, 580–595. https://doi.org/10.1016/j.rse.2013.09.025

Fricker, H. A., Arndt, P., Brunt, K. M., Datta, R. T., Fair, Z., Jasinski, M. F., Kingslake, J., Magruder, L. A., Moussavi, M., Pope, A., et al. (2021). ICESat-2 meltwater depth estimates: application to surface melt on Amery Ice Shelf, East Antarctica. *Geophysical Research Letters*, *48*(8), e2020GL090550. https://doi.org/10.1029/2020GL090550

Granata, F. (2019). Evapotranspiration evaluation models based on machine learning algorithms—A comparative study. *Agricultural Water Management*, *217*, 303–315. https://doi.org/10.1016/j.compag.2018.03.010

Guzinski, R., Nieto, H., Sandholt, I., & Karamitilios, G. (2020). Modelling high-resolution actual evapotranspiration through Sentinel-2 and Sentinel-3 data fusion. *Remote Sensing*, *12*(9), 1433. https://doi.org/10.3390/rs12091433

Hanna, E., Fettweis, X., Mernild, S. H., Cappelen, J., Ribergaard, M. H., Shuman, C. A., Steffen, K., Wood, L., & Mote, T. L. (2014). Atmospheric and oceanic climate forcing of the ex-

ceptional Greenland ice sheet surface melt in summer 2012. *International Journal of Climatology*, *34*(4), 1022–1037. https://doi.org/10.1002/joc.3743

Hosseini-Moghari, S.-M., Araghinejad, S., Ebrahimi, K., Tang, Q., & AghaKouchak, A. (2020). Using GRACE satellite observations for separating meteorological variability from anthropogenic impacts on water availability. *Scientific Reports*, *10*(1), 15098. https://doi.org/10.1038/s41598-020-71837-7

Humphrey, V., Rodell, M., & Eicker, A. (2023). Using Satellite-Based Terrestrial Water Storage Data: A Review. *Surveys in Geophysics*, 1–29. https://doi.org/10.1007/s10712-022-09754-9

Jekeli, C. (1981). *Alternative methods to smooth the Earth's gravity field* (tech. rep.). https://ntrs.nasa.gov/citations/19820014947

Jensen, L., Eicker, A., Dobslaw, H., Stacke, T., & Humphrey, V. (2019). Long-term wetting and drying trends in land water storage derived from GRACE and CMIP5 models. *Journal of Geophysical Research: Atmospheres*, *124*(17-18), 9808–9823. https://doi.org/10.1029/2018JD029989

Joodaki, G., Wahr, J., & Swenson, S. (2014). Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research*, *50*(3), 2679–2692. https://doi.org/10.1002/2013wr014633

Klees, R., Revtova, E., Gunter, B., Ditmar, P., Oudman, E., Winsemius, H., & Savenije, H. (2008). The design of an optimal filter for monthly GRACE gravity models. *Geophysical Journal International*, *175*(2), 417–432. https://doi.org/10.1111/j.1365-246X.2008.03922.x

Kundzewicz, Z. W. (2008). Climate change impacts on the hydrological cycle. *Ecohydrology & Hydrobiology*, *8*(2-4), 195–203. https://doi.org/10.2478/v10104-009-0015-y

Kusche, J., Schmidt, R., Petrovic, S., & Rietbroek, R. (2009). Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *Journal of geodesy*, *83*(10), 903–913. https://doi.org/10.1007/s00190-009-0308-3

Kwok, R., Cunningham, G., Markus, T., Hancock, D., Morison, J., Palm, S., Farrell, S., Ivanoff, A., & Wimert, J. (2019). ATLAS/ICESat-2 L3A sea ice height, version 1. *Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center*, *20*.

Landerer, F. W., & Swenson, S. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research*, *48*(4). https://doi.org/10.1029/2011WR011453

Levizzani, V., & Cattani, E. (2019). Satellite remote sensing of precipitation and the terrestrial water cycle in a changing climate. *Remote sensing*, *11*(19), 2301. https://doi.org/10.3390/rs11192301

Li, R., Lei, J., Kusche, J., Dang, T., Huang, F., Luan, X., Zhang, S.-R., Yan, M., Yang, Z., Liu, F., et al. (2023). Large-Scale Disturbances in the Upper Thermosphere Induced by the 2022 Tonga Volcanic Eruption. *Geophysical Research Letters*, *50*(3), e2022GL102265. https://doi.org/10.1029/2022GL102265

Liu, R., Li, J., Fok, H. S., Shum, C., & Li, Z. (2014). Earth surface deformation in the north China plain detected by joint analysis of GRACE and GPS data. *Sensors*, *14*(10), 19861–19876. https://doi.org/10.3390/s141019861

Maggioni, V., Meyers, P. C., & Robinson, M. D. (2016). A review of merged high-resolution satellite precipitation product accuracy during the Tropical Rainfall Measuring Mission (TRMM) era. *Journal of Hydrometeorology*, *17*(4), 1101–1117. https://doi.org/10.1175/JHM-D-15-0190.1

Manabe, S. (1969). Climate and the ocean circulation: I. The atmospheric circulation and the hydrology of the Earth's surface. *Monthly weather review*, *97*(11), 739–774. https://doi.org/10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., et al. (2017). The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): science requirements, concept, and implementation. *Remote sensing of environment, 190*, 260–273. https://doi.org/10.1016/j.rse.2016.12.029

Monteith, J. L. (1965). Evaporation and environment. *Symposia of the society for experimental biology, 19*, 205–234.

Myneni, R. B., Hall, F. G., Sellers, P. J., & Marshak, A. L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and remote Sensing, 33*(2), 481–486. https://doi.org/10.1109/TGRS.1995.8746029

Nagler, T., & Rott, H. (1997). The application of ERS-1 SAR for snowmelt runoff modelling. *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences, 242*, 119–128.

Nagler, T., Rott, H., & Glendinning, G. (2000). Snowmelt runoff modelling by means of Radarsat and ERS SAR. *Canadian journal of remote sensing, 26*(6), 512–520. https://doi.org/10.1080/07038992.2000.10874792

Nemani, R. R., & Running, S. W. (1989). Estimation of regional surface resistance to evapotranspiration from NDVI and thermal-IR AVHRR data. *Journal of Applied Meteorology and Climatology, 28*(4), 276–284. https://doi.org/10.1175/1520-0450(1989)028<0276:EORSRT>2.0.CO;2

Ni, S., Chen, J., Wilson, C. R., & Hu, X. (2017). Long-term water storage changes of Lake Volta from GRACE and satellite altimetry and connections with regional climate. *Remote Sensing, 9*(8), 842. https://doi.org/10.3390/rs9080842

Nitu, R., & Wong, K. (2010). Measurement of solid precipitation at automatic weather stations, challenges and opportunities. *Meteorological Service of Canada, 4905*.

Nsabagwa, M., Byamukama, M., Kondela, E., & Otim, J. S. (2019). Towards a robust and affordable Automatic Weather Station. *Development Engineering, 4*, 100040. https://doi.org/10.1016/j.deveng.2018.100040

Papa, F., & Frappart, F. (2021). Surface water storage in rivers and wetlands derived from satellite observations: A review of current advances and future opportunities for hydrological sciences. *Remote Sensing, 13*(20), 4162. https://doi.org/10.3390/rs13204162

Pellet, V., Aires, F., Papa, F., Munier, S., & Decharme, B. (2020). Long-term total water storage change from a Satellite Water Cycle reconstruction over large southern Asian basins. *Hydrology and Earth System Sciences, 24*(6), 3033–3055. https://doi.org/10.5194/hess-24-3033-2020

Purdy, A. J., Fisher, J. B., Goulden, M. L., Colliander, A., Halverson, G., Tu, K., & Famiglietti, J. S. (2018). SMAP soil moisture improves global evapotranspiration. *Remote Sensing of Environment, 219*, 1–14. https://doi.org/10.1016/j.rse.2018.09.023

Qin, Y., Abatzoglou, J. T., Siebert, S., Huning, L. S., AghaKouchak, A., Mankin, J. S., Hong, C., Tong, D., Davis, S. J., & Mueller, N. D. (2020). Agricultural risks from changing snowmelt. *Nature Climate Change, 10*(5), 459–465. https://doi.org/10.1038/s41558-020-0746-8

Reager, J. T., & Famiglietti, J. S. (2009). Global terrestrial water storage capacity and flood potential using GRACE. *Geophysical research letters, 36*(23). https://doi.org/10.1029/2009GL040826

Reichle, R. H., & Koster, R. D. (2005). Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. *Geophysical Research Letters, 32*(2). https://doi.org/10.1029/2004GL021700

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., & Rodell, M. (2015). Quantifying renewable groundwater stress with GRACE. *Water resources research, 51*(7), 5217–5238. https://doi.org/10.1002/2015WR017349

Robock, A., Vinnikov, K. Y., Srinivasan, G., Entin, J. K., Hollinger, S. E., Speranskaya, N. A., Liu, S., & Namkhai, A. (2000). The global soil moisture data bank. *Bulletin of the American Meteorological Society*, *81*(6), 1281–1300. https://doi.org/10.1175/1520-0477(2000) 081<1281:TGSMDB>2.3.CO;2

Rockström, J., Steffen, W., Noone, K., Persson, Å., Chapin III, F. S., Lambin, E., Lenton, T. M., Scheffer, M., Folke, C., Schellnhuber, H. J., et al. (2009). Planetary boundaries: Exploring the safe operating space for humanity. *Ecology and society*, *14*(2). https://www.jstor.org/stable/26268316

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J., Beaudoing, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature*, *557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Rodell, M., Velicogna, I., & Famiglietti, J. S. (2009). Satellite-based estimates of groundwater depletion in India. *Nature*, *460*(7258), 999–1002. https://doi.org/10.1038/nature08238

Rossato, L., Alvala, R. C., Ferreira, N. J., & Tomasella, J. (2005). Evapotranspiration estimation in the Brazil using NDVI data. *Remote Sensing for Agriculture, Ecosystems, and Hydrology VII*, *5976*, 377–385. https://doi.org/10.1117/12.626793

Running, S. W. (1990). Estimating terrestrial primary productivity by combining remote sensing and ecosystem simulation. *Remote sensing of biosphere functioning*, 65–86. https://doi.org/10.1007/978-1-4612-3302-2_4

Scanlon, B. R., Longuevergne, L., & Long, D. (2012). Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. *Water Resources Research*, *48*(4). https://doi.org/10.1029/2011WR011312

Scanlon, B. R., Zhang, Z., Reedy, R. C., Pool, D. R., Save, H., Long, D., Chen, J., Wolock, D. M., Conway, B. D., & Winester, D. (2015). Hydrologic implications of GRACE satellite data in the Colorado River Basin. *Water Resources Research*, *51*(12), 9891–9903. https://doi.org/10.1002/2015WR018090

Schneider, T., O'Gorman, P. A., & Levine, X. J. (2010). Water vapor and the dynamics of climate changes. *Reviews of Geophysics*, *48*(3). https://doi.org/10.1029/2009RG000302

Schwalb, A. (1978). The Tiros-N/NOAA AG Satellite Series. *NASA STI/Recon Technical Report N*, *79*, 12135.

Sheffield, J., Wood, E. F., & Roderick, M. L. (2012). Little change in global drought over the past 60 years. *Nature*, *491*(7424), 435–438. https://doi.org/10.1038/nature11575

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019RG000678

Simpson, J., Adler, R. F., & North, G. R. (1988). A proposed tropical rainfall measuring mission (TRMM) satellite. *Bulletin of the American meteorological Society*, *69*(3), 278–295. https://doi.org/10.1175/1520-0477(1988)069<0278:APTRMM>2.0.CO;2

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, E. F., Kirschbaum, D. B., Kakar, R., Braun, S. A., Huffman, G. J., Iguchi, T., et al. (2017). The Global Precipitation Measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*, *98*(8), 1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1

Song, C., Huang, B., & Ke, L. (2013). Modeling and analysis of lake water storage changes on the Tibetan Plateau using multi-mission satellite data. *Remote Sensing of Environment*, *135*, 25–35. https://doi.org/10.1016/j.rse.2013.03.013

Strangeways, I. (2006). *Precipitation: Theory, measurement and distribution*. Cambridge University Press.

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics*, *56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data. *Geophysical research letters*, *33*(8). https://doi.org/10.1029/2005GL025285

Syed, T., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research*, *44*(2). https://doi.org/10.1029/2006WR005779

Tabari, H., & Talaee, P. H. (2014). Sensitivity of evapotranspiration to climatic change in different climates. *Global and Planetary Change*, *115*, 16–23. https://doi.org/10.1016/j.gloplacha.2014.01.006

Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., Garcıéa-Ortega, E., Machado, L. A., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., et al. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, *104*, 70–97. https://doi.org/10.1016/j.atmosres.2011.10.021

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters*, *31*(9). https://doi.org/10.1029/2004GL019920

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. (2019). Contributions of GRACE to understanding climate change. *Nature climate change*, *9*(5), 358–369. https://doi.org/10.1038/s41558-019-0456-2

Tarpanelli, A., Mondini, A. C., & Camici, S. (2022). Effectiveness of Sentinel-1 and Sentinel-2 for flood detection assessment in Europe. *Natural Hazards and Earth System Sciences*, *22*(8), 2473–2489. https://doi.org/10.5194/nhess-22-2473-2022

Thomas, A. C., Reager, J. T., Famiglietti, J. S., & Rodell, M. (2014). A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters*, *41*(5), 1537–1545. https://doi.org/10.1002/2014GL059323

Thomas, B. F., & Famiglietti, J. S. (2019). Identifying climate-induced groundwater depletion in GRACE observations. *Scientific reports*, *9*(1), 4124. https://doi.org/10.1038/s41598-019-40155-y

Thomas, B. F., Famiglietti, J. S., Landerer, F. W., Wiese, D. N., Molotch, N. P., & Argus, D. F. (2017). GRACE groundwater drought index: Evaluation of California Central Valley groundwater drought. *Remote Sensing of Environment*, *198*, 384–392. https://doi.org/10.1016/j.rse.2017.06.026

Tortini, R., Noujdina, N., Yeo, S., Ricko, M., Birkett, C. M., Khandelwal, A., Kumar, V., Marlier, M. E., & Lettenmaier, D. P. (2020). Satellite-based remote sensing data set of global surface water storage change from 1992 to 2018. *Earth System Science Data*, *12*(2), 1141–1151. https://doi.org/10.5194/essd-12-1141-2020

Tourian, M., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., & Sneeuw, N. (2015). A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran. *Remote Sensing of Environment*, *156*, 349–360. https://doi.org/10.1016/j.rse.2014.10.006

Tourian, M. J., Sneeuw, N., & Bárdossy, A. (2013). A quantile function approach to discharge estimation from satellite altimetry (ENVISAT). *Water Resources Research*, *49*(7), 4174–4186. https://doi.org/10.1002/wrcr.20348

Tourian, M. J. (2013). *Application of spaceborne geodetic sensors for hydrology*. https://doi.org/10.18419/opus-3929

Tourian, M. J., Elmi, O., Shafaghi, Y., Behnia, S., Saemian, P., Schlesinger, R., & Sneeuw, N. (2021). HydroSat: a repository of global water cycle products from spaceborne geodetic

sensors. *Earth System Science Data Discussions*, 1–42. https://doi.org/10.5194/essd-2021-174

Vanino, S., Nino, P., De Michele, C., Bolognesi, S. F., D'Urso, G., Di Bene, C., Pennelli, B., Vuolo, F., Farina, R., Pulighe, G., et al. (2018). Capability of Sentinel-2 data for estimating maximum evapotranspiration and irrigation requirements for tomato crop in Central Italy. *Remote Sensing of Environment*, *215*, 452–470. https://doi.org/10.1016/j.rse.2018.06.035

Vargas Godoy, M. R., Markonis, Y., Hanel, M., Kyselỳ, J., & Papalexiou, S. M. (2021). The global water cycle budget: A chronological review. *Surveys in Geophysics, 42*(5), 1075–1107. https://doi.org/10.1007/s10712-021-09652-6

Velicogna, I., & Wahr, J. (2005). Greenland mass balance from GRACE. *Geophysical research letters, 32*(18). https://doi.org/10.1029/2005GL023955

Velicogna, I., & Wahr, J. (2006). Measurements of time-variable gravity show mass loss in Antarctica. *science, 311*(5768), 1754–1756. https://doi.org/10.1126/science.1123785

Voigt, A., & Shaw, T. A. (2015). Circulation response to warming shaped by radiative changes of clouds and water vapour. *Nature Geoscience, 8*(2), 102–106. https://doi.org/10.1038/ngeo2345

Vorosmarty, C. J., Green, P., Salisbury, J., & Lammers, R. B. (2000). Global water resources: Vulnerability from climate change and population growth. *science, 289*(5477), 284–288. https://doi.org/10.1126/science.289.5477.284

Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., & Swenson, S. C. (2013). 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*(2), 904–914. https://doi.org/10.1002/wrcr.20078

Wada, Y., Van Beek, L. P., Wanders, N., & Bierkens, M. F. (2013). Human water consumption intensifies hydrological drought worldwide. *Environmental Research Letters, 8*(3), 034036. https://doi.org/10.1088/1748-9326/8/3/034036

Wahr, J., Molenaar, M., & Bryan, F. (1998). 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*(B12), 30205–30229. https://doi.org/10.1029/98JB02844

Walker, E., Garcıéa, G. A., Venturini, V., & Carrasco, A. (2019). Regional evapotranspiration estimates using the relative soil moisture ratio derived from SMAP products. *Agricultural Water Management, 216*, 254–263. https://doi.org/10.1016/j.agwat.2019.02.009

Wang, L., Shum, C., Simons, F. J., Tapley, B., & Dai, C. (2012). Coseismic and postseismic deformation of the 2011 Tohoku-Oki earthquake constrained by GRACE gravimetry. *Geophysical Research Letters, 39*(7). https://doi.org/10.1029/2012GL051104

Wouters, B., van de Wal, R., et al. (2018). Global sea-level budget 1993–present. *Earth System Science Data, 10*(3), 1551–1590. https://doi.org/10.3929/ethz-b-000287786

Wu, P., Christidis, N., & Stott, P. (2013). Anthropogenic impact on Earth's hydrological cycle. *Nature Climate Change, 3*(9), 807–810. https://doi.org/10.1038/nclimate1932

Yagci, A. L., & Yilmaz, M. T. (2021). Mapping and monitoring of soil moisture, evapotranspiration, and agricultural drought. *Agro-geoinformatics: Theory and Practice*, 299–320. https://doi.org/10.1007/978-3-030-66387-2_15

Yang, K., Wu, H., Qin, J., Lin, C., Tang, W., & Chen, Y. (2014). Recent climate changes over the Tibetan Plateau and their impacts on energy and water cycle: A review. *Global and Planetary Change, 112*, 79–91. https://doi.org/10.1016/j.gloplacha.2013.12.001

Yang, Y., Chen, R., Liu, G., Liu, Z., & Wang, X. (2022). Trends and variability in snowmelt in China under climate change. *Hydrology and Earth System Sciences, 26*(2), 305–329. https://doi.org/10.5194/hess-26-305-2022

Yeh, P. J.-F., Swenson, S. C., Famiglietti, J. S., & Rodell, M. (2006). Remote sensing of ground-water storage changes in Illinois using the Gravity Recovery and Climate Experiment (GRACE). *Water Resources Research*, *42*(12). https://doi.org/10.1029/2006WR005374

Yi, S., & Sneeuw, N. (2022). A novel spatial filter to reduce north–south striping noise in GRACE spherical harmonic coefficients. *Journal of Geodesy*, *96*(4), 1–17. https://doi.org/10.1007/s00190-022-01614-z

Yoon, J.-H., Wang, S. S., Gillies, R. R., Kravitz, B., Hipps, L., & Rasch, P. J. (2015). Increasing water cycle extremes in California and in relation to ENSO cycle under global warming. *Nature Communications*, *6*(1), 8657. https://doi.org/10.1038/ncomms9657

Zhang, D., Zhang, Q., Werner, A. D., & Liu, X. (2016). GRACE-based hydrological drought evaluation of the Yangtze River Basin, China. *Journal of Hydrometeorology*, *17*(3), 811–828. https://doi.org/10.1175/JHM-D-15-0084.1

Zhang, D., & Zhou, G. (2016). Estimation of soil moisture from optical and thermal remote sensing: A review. *Sensors*, *16*(8), 1308. https://doi.org/10.3390/s16081308

Zhang, J., Xu, K., Yang, Y., Qi, L., Hayashi, S., & Watanabe, M. (2006). Measuring water storage fluctuations in Lake Dongting, China, by Topex/Poseidon satellite altimetry. *Environmental Monitoring and Assessment*, *115*, 23–37. https://doi.org/10.1007/s10661-006-5233-9

Zhao, M., Zhang, J., Velicogna, I., Liang, C., Li, Z., et al. (2021). Ecological restoration impact on total terrestrial water storage. *Nature Sustainability*, *4*(1), 56–62. https://doi.org/10.1038/s41893-020-00600-7

Zheng, M., Deng, K., Fan, H., & Du, S. (2018). Monitoring and analysis of surface deformation in mining area based on InSAR and GRACE. *Remote Sensing*, *10*(9), 1392. https://doi.org/10.3390/rs10091392

2

Spaceborne Monitoring of Water Storage Variation

2.1. GRACE(-FO) Satellite Mission

The Gravity Recovery and Climate Experiment (GRACE) was a joint mission of the National Aeronautics and Space Administration (NASA) and the German Aerospace Center (DLR) (Tapley et al., 2004a) (Figure 2.1(a)). On March 17th, 2002, the twin satellites were launched on a Rockot launch vehicle from the Eurockot's launch facility at the Plesetsk cosmodrome in Russia. GRACE consisted of two identical spacecrafts in tandem named GRACE-A and GRACE-B, or Tom and Jerry, flying 220 km apart from each other (Nelson, 2017; Wahr et al., 1998). The trailing satellite passes over a specific region approximately 28 s after the leading satellite. GRACE satellites fly in a co-planar, nearly polar orbit (89° inclination) to ensure near-global coverage of the Earth (Kirschner et al., 2001; Visser, 2005). The mission was designed in low orbit to increase the spatial resolution of the final gravity field anomalies up to a few hundred kilometers (Ray & Luthcke, 2006; Schrama & Visser, 2007). Due to atmospheric drag, the initial altitude of 500 km decreased to 330 km by mid-2017. One revolution took about 94 minutes which resulted in 15.5 revolutions per day. The Earth's dense spatial coverage can be achieved after one month of continuous observation (Figure 2.2). The complete orbital configuration can be found at http://www.csr.utexas.edu/grace/.

The GRACE-FO (GRACE Follow-On) mission, launched on 22 May 2018, is a continuation of the original GRACE mission. The satellite mission launched aboard a SpaceX Falcon 9 rocket from Vandenberg AFB, California. GRACE-FO builds on the success of the original GRACE mission by using more advanced technologies like the Laser-Ranging Interferometers (LRI) and is designed to continue providing high-precision measurements of Earth's gravity field (Figure 2.1(b)). The interferometers use a laser beam, which is split into two paths and then recombined, to measure the distance between the two satellites to within a fraction of a micrometer. This allows for much more precise measurements of the distance between the satellites, and therefore more accurate measurements of changes in the Earth's gravity field. Additionally, the laser ranging system operates at a much higher frequency than the microwave system, which results in a reduction of signal noise and a more stable measurement. Therefore, GRACE-FO's measurements are expected to be at least as accurate as those of the original GRACE mission. Additionally, the satellites feature a new set of navigation and attitude determination instruments, such as a Global Positioning System (GPS) receiver and a fiber-optic gyroscope. GRACE-FO has a design life of 5 years.

(a)   (b)

Figure 2.1: (a) Illustration of the twin GRACE satellites (credit: NASA); (b) The GRACE-FO satellites. The inter-satellite range is measured using lasers instead of microwaves (credit: Max-Planck institute for gravitational physics)

Figure 2.2: The path that the GRACE satellite mission follows is shown in three different time periods: one day (left panel), one week (middle panel), and one month (right panel). (credit: Szabó and Marjańska, 2020)

GRACE satellites were equipped with the following instruments (Figure 2.3) (for more details, see https://earth.esa.int/eogateway/missions/grace) :

- **K-band Ranging System (KBR)**: provides precise (within 10 μm) measurements of the distance change between the two satellites needed to estimate fluctuations in gravity.

- **Ultra Stable Oscillator (USO)**: provides a stable frequency generation for the K-band ranging system.

- **SuperSTAR Accelerometers (ACC)**: precisely measures the non-gravitational accelerations acting on the satellites.

- **Star Camera Assembly (SCA)**: precisely determines the two satellite's orientation by tracking them relative to the position of the stars.

- **Coarse Earth Sun Sensor (CES)**: provides omnidirectional, reliable, and robust, but fairly coarse, Earth and Sun tracking. Used during initial acquisition and whenever GRACE operates in safe mode.

- **Center of Mass Trim Assembly (MTA)**: precisely measures the offset between the satellite's centre of mass and the ACC proof-mass and adjusts center of mass as needed during the flight.

- **BlackJack GPS Receiver and Instrument Processing Unit (GPS)**: provides digital signal processing; measures the distance change relative to the GPS satellite constellation.

In an idealized case (i.e., no atmospheric drag, no solar radiation pressure, etc.), the change in distance between two co-orbiting satellites is essentially proportional to the difference in the Earth's gravitational potential at the respective locations (Wolff, 1969). Variations in gravity exist due to mass changes in the hydrosphere, atmosphere, biosphere, oceans, and mass variations inside the Earth. Initially, the GRACE mission was designed for a 5-year lifespan (Tapley et al., 2004b). Since 2011, battery issues have led to missing solutions approximately every six months. GRACE was operational until June 2017, providing 15 years of records. GRACE-FO now carries on the observation strategy of the GRACE mission while testing a new laser ranging system, which is more precise than the K-band instrument (Sheard et al., 2012; Tapley et al., 2019).
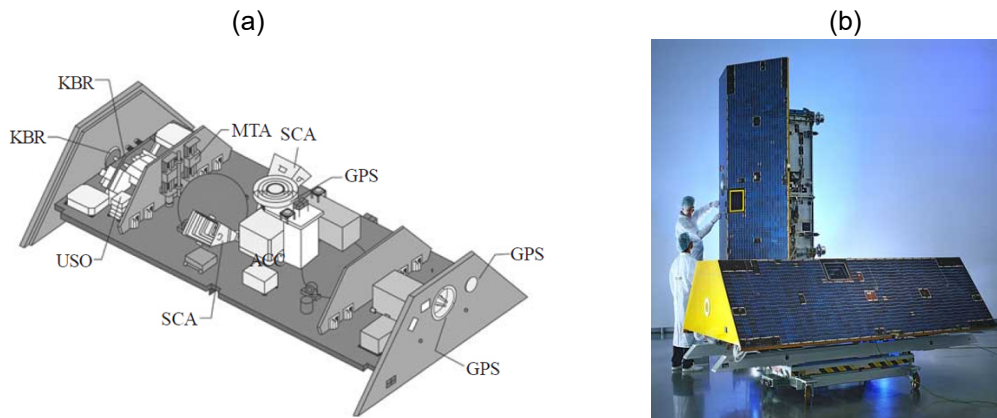
Figure 2.3: (a) Interior view of the GRACE instruments (Credit: NASA). (b) The twin GRACE satellites (Credit: Astrium).

2.1.1. Measurement Principle

The GRACE measurement principle is based on three fundamental techniques: precise orbit determination, inter-satellite ranging, and accelerometry. The GRACE orbit tracking primarily uses the Global Positioning System (GPS). Additionally, the orbit is tracked from the ground stations using the Satellite Laser Ranging (SLR) observation technique. For these purposes, a laser retro-reflector is mounted on the nadir panel of each satellite. The SLR observations are then used for validation of the GPS-based orbits. The precise orbit determination is performed on the ground by applying the so-called kinematic, dynamic, or reduced dynamic approach (Jäggi et al., 2007; Kang et al., 2006).

It is not only the gravitational attraction of the Earth but also the gravitational attraction of the Sun, the Moon, and other celestial bodies, which perturb the satellite's orbit. All these forces have to be precisely modeled and removed from the observations. Along with the gravitational forces, non-gravitational forces such as air drag, solar radiation pressure, and Earth albedo act on the satellite vehicle. Each satellite was equipped with two simultaneously operating star cameras, which provided orientation towards the stars i.e., external orientation in an inertial frame. Knowing the location (GPS) can then turn this information into an Earth-fixed or Satellite-fixed orientation. In order to obtain the orbit perturbations caused solely by the gravitational attraction of the Earth's mass, these non-gravitational forces are sensed by ultra-sensitive accelerometers located at the center of mass of each satellite and subsequently reduced from the original observations as well. The key instrument of the GRACE satellites was a dual one-way KBR system, which measured the inter-satellite range and its time derivative. After processing, biased ranges, range rates, and range accelerations are available at 5-second sampling.

2.2. Data Pre-Processing

Before the onboard data is used for gravity field investigations, several correction and filtering steps are required. Generally, the following background models are considered (e.g., Dahle et al., 2013; Watkins & Yuan, 2012):

- **N-Body Perturbations**: the satellites' orbits are influenced by gravitational forces from third bodies, i.e., the sun, the moon, and the planets. These direct and indirect accelerations acting on the GRACE satellites are computed from planetary ephemerides and eventually removed from the observations (Agnew, 2010).

- **Solid Earth Tides**: tidal forces induce deformations of the solid Earth, which can be modeled using Love numbers (Wahr et al., 1998). Subsequently, the influence on the Earth's gravitational potential is computed.

- **Ocean Tides**: oceanic mass variations due to tidal forces are considered by combining models of the different tidal constituents.

- **Pole Tides**: changes in the Earth's rotation result in centrifugal force changes. The influence on the gravitational potential is computed from the polar motion and (i) an Earth model for the contribution of the solid Earth and (ii) an equilibrium model for the oceanic contribution (Desai, 2002).

- **De-aliasing**: high-frequency (< 30 days) non-tidal variations of atmosphere and ocean are modeled and removed using the Atmosphere and Ocean De-aliasing Level- 1B (AOD1B) products (Dobslaw et al., 2017; Flechtner et al., 2014). By this means, the aliasing of high frequencies into monthly gravity field solutions is reduced. It should be mentioned that errors in this model can introduce biases to estimates of mass change from GRACE(-FO). AOD1B Product is subject to several sources of error that can affect its accuracy and reliability. Such errors including error in the modeling of the atmosphere and ocean dynamics and errors in the estimation of the AOD1B coefficients can become aliased in the GRACE solutions, meaning that they can appear as signals in the estimated changes in the Earth's gravity field.

- **Non-gravitational Forces**: non-gravitational accelerations, i.e., atmospheric drag and radiation pressure from Sun and Earth, were measured by the satellites' accelerometers. A few-parameter model relating linear acceleration observations to non-gravitational forces is set up. During gravity field estimation, biases and scale factors of the accelerometer measurements are co-estimated.

- **Relativistic Correction**: instead of a fully consistent relativistic formulation of the equation of motion, general relativistic corrections are applied as described in Chapter 10 of the International Earth Rotation and Reference Systems (IERS) 2010 conventions (Petit & Luzum, 2010).

2.3. Data levels

The GRACE(-FO) raw telemetry data called **Level-0** is processed by the Science Data System (SDS) and stored in binary encoded measurements. At the next level, the binary data is transformed into engineering units through editing, quality control, and sensor calibration to yield **Level-1A** data. The **Level-1B** data will then be derived from the **Level-1A** data after being down-sampled, time-tagged, and filtered. The **Level-1A** and **Level-1B** data are generated by two of the SDS centers, namely the German Research Center for Geoscience (GFZ) and the Jet Propulsion Laboratory (JPL). The GRACE(-FO) **Level-2** products, which hold the monthly and static gravity field models in terms of Spherical Harmonics (SHs) coefficients, are computed from various gravity field solutions. Three official centers calculate temporal global gravity field models using the classical variation approach (or dynamic approach) and provide SHs solutions, namely the University of Texas Centre for Space Research (CSR), GFZ, and JPL. Besides the official centers, several other centers have obtained their own solutions using different approaches and background models. Table 2.1 lists the solutions, including their time span and maximum degree/order. Temporal gravity field models developed by different institutions and agencies can be downloaded from the International Center for Global Earth Models (ICGEM) http://icgem.gfz-potsdam.de/series (last access: 1 June 2022). Please note that

the COST-G product is a combination of AIUB-RL02, GFZ-RL06, GRGS-RL04 (unconstrained solution), ITSG-GRACE2018, and CSR-RL06.

GRACE(-FO) **Level-3** products, i.e., gridded geopotential functionals (e.g. water equivalent mass change), can be generated from the Level-2 solutions. To this end, additional steps known as post-processing must be conducted. These steps are later introduced in the section 2.4. Within the past several years, an alternative approach for processing GRACE level-1 has been proposed, which considers parameterizing the gravitational field with regional mass concentration functions (mascons) (**watkins2015mascon**). Three mascons solutions (RL06) are currently available: the CSR mascon solutions (Save, 2020; Save et al., 2016), the NASA Goddard Space Flight Center (GSFC) (Rowlands et al., 2010; Watkins et al., 2015) that are calculated through expansion of spherical harmonics and the JPL mascon solution which is acquired from analytical partial derivatives (Wiese et al., 2016).

Table 2.1: List of GRACE(-FO) Level-2 solutions.

Center	Solution(s)	max d/o	time span (GRACE)	Approach
Solutions that include GRACE and GRACE-FO				
CSR	CSR RL06	60, 96	200204–201706	dynamic
	CSR RL06 (GFO)	60, 96	201806–present	
GFZ	GFZ RL06	60, 97	200204–201706	dynamic
	CSR RL06 (GFO)	60, 96	201806–present	
JPL	JPL RL06	60, 98	200204–201706	dynamic
	CSR RL06 (GFO)	60, 96	201806–present	
ITSG	ITSG-Grace2018	60, 96, 120	200204–201706	dynamic
	ITSG-Grace_op	60, 96, 120	201806–present	
LUH	LUH-Grace2018	80*	200301–201603	dynamic
	LUH-GRACE-FO-2020	80*	201806–present	
COST-G**	Grace	90	200204–201706	celestial mechanics
	Grace-FO	90	201806–present	
AIUB	AIUB-RL02	90*	200302–201403	celestial mechanics
	AIUB-GRACE-FO_op	90*	201806–present	
CNES	CNES_GRGS_RL05	90	200209–present	dynamic
Solutions that include only GRACE				
Tongji	Tongji-Grace2018	96	200204–201608	modified short arc
HUST	HUST-Grace2020	60, 90	200301–201607	dynamic
IGG	IGG-RL01	60	200204–201607	dynamic
SWJTU	SWJTU-GRACE-RL01	60	200303–201110	dynamic
SWPU	SWPU-GRACE2021	96	200204–201705	dynamic
WHU	WHU RL01	60, 90, 120	200204–201607	dynamic
XISM&SSTC	GRACE01	60	200204-201603	dynamic

* The maximum d/o might change at some epochs.

The SDS provides quality-controlled data from Level-0 (KBR range data) to Level-3 (grids). Moreover, Level-3 solutions from the mascons approach and expressed in terms of TWSA can be accessed from CSR, JPL, and NASA-GSFC. Other than Level-3 data, several centers help to visualize GRACE TWSA. The JPL and GSFC mascons, for instance, can be visualized using the Mascon Visualization Tool from the University of Colorado Boulder, and the basin-wise variability of TWS can be obtained from the Gravity over basins Information Service (GravIS) website. Furthermore, several data browsers allow the interactive retrieval of GRACE and GRACE-

FO data, including the one developed within the International Center for Global Earth Models (ICGEM) project, the GRACE Plotter, and NASA data Analysis Tool. Table 2.2 lists the above-mentioned centers and products.

Table 2.2: List of centers which provide Level-3 TWSA from GRACE and GRACE-FO.

Product	Sensor(s)	Source/Reference
Level-3 datasets		
JPL	GRACE	Landerer (2020d); Landerer and Swenson (2012)
	GRACE-FO	Landerer (2020e)
	GRACE/GRACE-FO (mascons)	Landerer et al. (2020); Wiese et al. (2018); Watkins et al. (2015)
CSR	GRACE	Landerer (2020a)
	GRACE-FO	Landerer (2020b)
	GRACE/GRACE-FO (mascons)	Save (2020); Save et al. (2016)
GFZ	GRACE	Landerer (2020c)
	GRACE-FO	Landerer (2020b)
	GRACE/GRACE-FO	Boergens et al. (2020)
GSFC	GRACE/GRACE-FO (mascons)	Loomis et al. (2019)
Visualization centers		
CU Boulder	GRACE/GRACE-FO	https://ccar.colorado.edu/grace
ICGEM	GRACE/GRACE-FO	http://icgem.gfz-potsdam.de/home
The GRACE Plotter	GRACE/GRACE-FO	http://thegraceplotter.com
NASA	GRACE/GRACE-FO	https://grace.jpl.nasa.gov/data-analysis-tool
GFZ	GRACE/GRACE-FO	http://gravis.gfz-potsdam.de/land
HydroSat	GRACE/GRACE-FO	http://hydrosat.gis.uni-stuttgart.de

2.4. Data post-processing

Each monthly solution of GRACE(-FO) contains the full hydrological, cryospheric, and Glacial Isostatic Adjustment (GIA) signal in the form of fully normalized SH coefficients after removing the contributions from other phenomena like tides (ocean, solid earth, and atmospheric), atmospheric and non-tidal oceanic mass changes. To accurately obtain the Total Water Storage Anomaly (TWSA), which represents the water equivalent mass change, several corrections still need to be applied to the GRACE(-FO) solutions, known as post-processing steps. Figure 2.4 illustrates the scheme of the data post-processing from GRACE(-FO) Level-2 to Level-3. The post-processing steps are color-coded in four colors: gray : formal uncertainties and the final TWSA, blue : steps that are well-established in the community, orange : steps that depend on the user choice, and green : steps that are not common but improve the final results. In what follows, each of the post-processing steps is introduced. For the orange and green steps, a comparison between possible approaches is discussed.
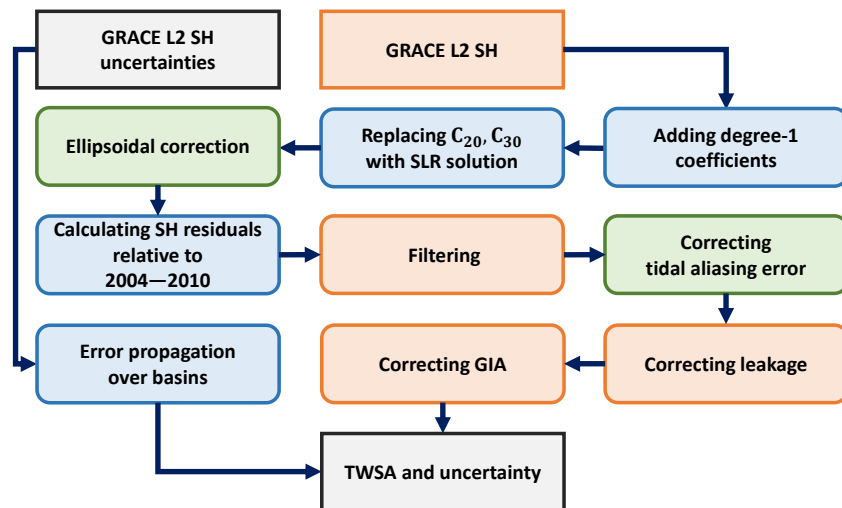

Figure 2.4: Flowchart of obtaining Total Water Storage Anomaly (TWSA) and its corresponding uncertainty from GRACE(-FO) Level-2 solutions.

Most comparisons have been conducted over the 405 major global river basins. The boundaries of the basins are obtained from the Global River Discharge Center (GRDC) (https://www.bafg.de/GRDC). Figure 2.5 illustrates the global distribution of the major basins while highlighting the 27 largest and most important river basins worldwide, categorized and listed based on their continent. Based on the area of the basins, we have categorized them into four major groups namely small to medium (area $< 10^4 \, \text{km}^2$), large (area within 10^4–$10^5 \, \text{km}^2$), very large (area within 10^5–$10^6 \, \text{km}^2$), and mega (area $> 10^6 \, \text{km}^2$) (Table 2.3). Based on such criteria, $\sim 10\%$ of the basins are small to medium sizes (e.g., Daule, Mira, and Geba), $\sim 60\%$ are large (e.g., Tana, Alabama, and Po), $\sim 23\%$ are very large (e.g., Ganges, Colorado, Highland of Tibet), and $\sim 7\%$ are mega (e.g., Greenland, Antarctica, Amur, Nelson, Lena).

Table 2.3: Summary of the area of the four major groups of the basins in terms of the basins' size.

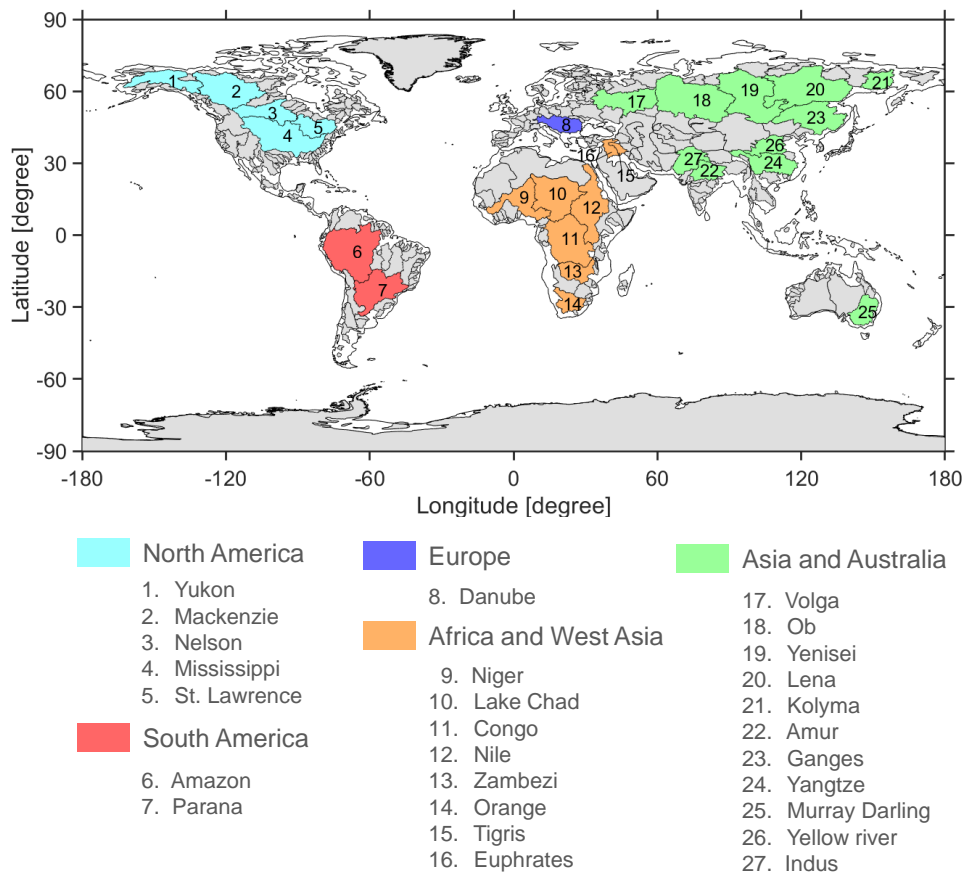|	# of basins	total area [$10^6 \, \text{km}^2$]
small/medium	40	0.25
large	246	9.8
very large	91	31.1
mega	28	72.4

Figure 2.5: Global distribution of the major river basins with gray background. The 27 largest and most important river basins are highlighted, listed, and color coded based on their continent.

These basins represent the global range of climate, land cover, and extent of irrigation. Characterization of climate for the basins is based on the mean annual aridity index (AI, the mean of annual precipitation over annual potential evapotranspiration), including:

- humid (AI > 0.65),

- dry sub-humid (AI ≤ 0.65, and AI > 0.5),

- semi-arid (AI ≤ 0.5 and > 0.2),

- arid (AI ≤ 0.2 and > 0.05),

- hyper-arid (AI ≤ 0.05).

This study grouped arid and hyper-arid into one group, arid to hyper-arid (Figure 2.6 a). Based on AI classification, 60 % of the river basins are categorized as humid, ~ 10 % as sub-humid, 22 % as semi-arid, and ~ 8 % as arid to hyper-arid). In order to understand the effect of latitude on the comparison of the GRACE(-FO) post-processing steps, four dominant latitude zones have been considered namely Polar (from 66.5 ° north (resp. south) to the North (resp. South) Poles), Temperate (40 °–66.5 ° north and south), Sub-tropical (23.5–40 ° north and south), and Tropical (-23.5–23.5 °) (Figure 2.6 b). Based on the latitude zones, the majority of the river basins lie in the tropical zones (~ 85 %), a balance between sub-tropical and temperate (each ~ 6.5 %), and only (~ 2 %) are located in the polar zones including Greenland and Antarctica.

The Root Mean Squares (RMS) of the GRACE(-FO) TWSA represent the strength of signal in each basin. The magnitude can be seen as a measure to characterize basins. In this thesis, we have divided the RMS of TWSA into four categories, namely <30 mm (~ 25 %), 30–60 mm (~ 40 %), 60–90 mm (~ 20 %), and > 90 mm (~ 15 %) (Figure 2.6 c).
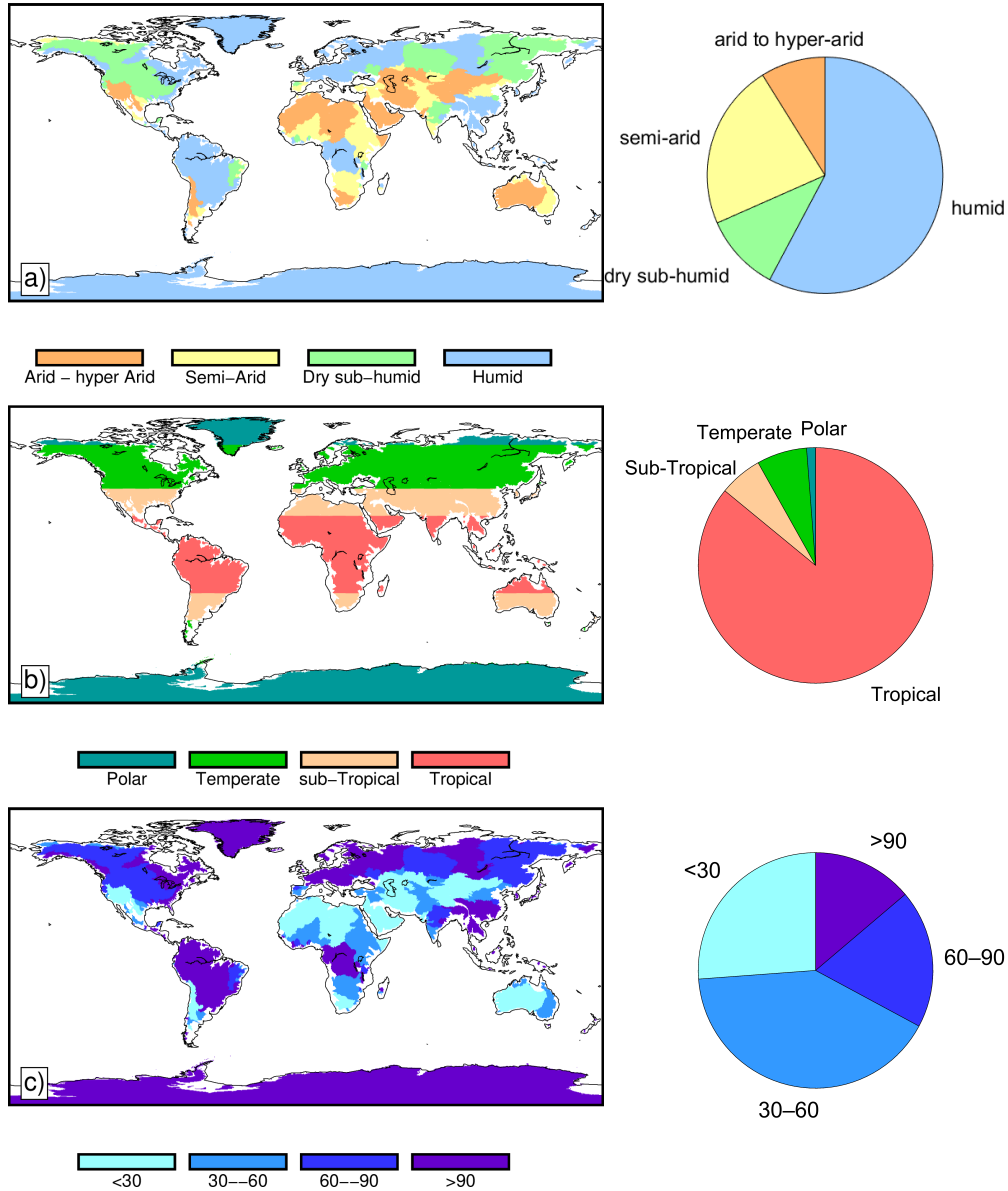

Figure 2.6: Characteristics of Global River Basins (a) aridity index (AI), (b) Latitude Zones, and (c) Root Mean Squared (RMS) of the TWSA signal from GRACE(-FO). The pie charts in the right panel show the percentage of the basins that lies in each category.

2.4.1. Total Water Storage Anomalies from spherical harmonics

GRACE(-FO) measures the Earth's gravity field commonly illustrated in terms of the shape of geoid: the equipotential surface corresponding to mean sea level over the oceans. The Earth's gravitational field is described by the geopotential V, using the following synthesis relation (Heiskanen & Moritz, 1967):

$$V(r,\theta,\lambda) = \frac{GM}{R}\left\{\sum_{l=0}^{\infty}\sum_{m=0}^{l}\left(\frac{R}{r}\right)^{l+1}P_{lm}(\cos\theta)\,[C_{lm}\cos(m\lambda)+S_{lm}\sin(m\lambda)]\right\}, \qquad (2.1)$$

where G is the Newton's gravitational constant, M is the total mass of the Earth, R represent the mean equatorial radius of the Earth, r is the radius, θ and λ are the co-latitude and longitude respectively. P_{lm} are the fully normalized Legendre polynomials of degree l and order m while C_{lm} and C_{lm} are the fully normalized dimensionless spherical harmonic coefficients.

Beyond the static gravity potential (Equation 2.1), GRACE(-FO) provides a valuable insight into the time-variable gravity information represented by change in the spherical harmonic coefficients. This variability can be seen as either the change in V from one time to the other, or the difference with respect to a time average of V. We considered the later interpretation and removed the long-term mean of spherical harmonic coefficients from monthly values. These residual spherical harmonic coefficients are denoted by ΔC_{lm} and ΔS_{lm} can be related to the density redistribution ($\Delta\rho(r,\theta,\lambda)$) occurring on and in the Earth:

$$\begin{Bmatrix}\Delta C_{lm}\\\Delta S_{lm}\end{Bmatrix} = \frac{3}{4\pi R\rho_{\text{ave}}(2l+1)}\int\int\int_{v}\Delta\rho(r,\theta,\lambda)P_{lm}(\cos\theta)\times\left(\frac{r}{R}\right)^{l+2}\begin{Bmatrix}\cos(m\lambda)\\\sin(m\lambda)\end{Bmatrix}\sin\theta\,d\theta\,d\lambda\,dr,$$
$$(2.2)$$

Where ρ_{ave} is the average density of the Earth (= 5517 kg/m^3). We assume that most of the mass redistribution takes place within a thin layer of thickness H, including those portion of the atmosphere, oceans, ice caps, and below groundwater storage with significant mass fluctuations. Moreover, we pretend that this thin layer consists of water. This is a meaningful assumption in hydrology , but must be used with care in ice sheet and and in GIA/seismo analysis. When H is thin enough (i.e. $(l_{\max}+2)H/R\ll 1$) then $\left(\frac{r}{R}\right)^{l+2}$ will be approximately equal to 1 and by replacing the volume integral $\Delta\rho$ by the surface integral $\Delta\sigma$, Equation 2.2 becomes:

$$\begin{Bmatrix}\Delta C_{lm}\\\Delta S_{lm}\end{Bmatrix} = \frac{3}{4\pi R\rho_{\text{ave}}(2l+1)}\int_{\Omega}\Delta\sigma(\theta,\lambda)P_{lm}(\cos\theta)\times\begin{Bmatrix}\cos(m\lambda)\\\sin(m\lambda)\end{Bmatrix}d\Omega. \qquad (2.3)$$

The surface mass density change ($\Delta\sigma$) can be expressed in terms of *Equivalent Water Height* (EWH). EWH or the TWSA is obtained as follow (Wahr et al., 1998):

$$\text{TWSA} = \text{EWH}(\theta,\lambda) = \frac{R\rho_{\text{ave}}}{3\rho_{\text{w}}}\sum_{l=0}^{\infty}\frac{2l+1}{1+k_l}\sum_{m=0}^{l}P_{lm}(\cos\theta)[\Delta C_{lm}\cos(m\lambda)+\Delta S_{lm}\sin(m\lambda)] \qquad (2.4)$$

2

2.4.2. GRACE(-FO) Level-2 solutions quality assessment

The GRACE(-FO) gravity field solutions depend on the dynamic process models, methods, assumptions, and background force models (ocean and atmospheric models together with ocean dealiasing product). Although most solutions have comparable signal amplitude, their performance in reducing noise should be evaluated, especially at higher degrees. To this end, we have compared the Level-2 solutions in the spectral domain using geoid degree RMS relative to the state-of-the-art static model EIGEN6C4. Figure 2.7 shows the comparison for three months, namely July 2003, December 2008, and April 2011. In all selected epochs in Figure 2.7, various solutions agree well up to degree 30. Such a performance can be expected since the gravity field signals dominate for lower degrees (~ below 30) (Chen et al., 2018; Meyer et al., 2016). ITSG-Grace2018 perform slighly better than other solutions. Such a performance has already been reported by Chen et al. (2019), but within a smaller group of solutions. In contrast, considering the coefficients up to degree 30, LUH and AIUB showed the highest difference among all the solutions. Among official centers, CSR demonstrates slightly better performance while JPL and GFZ lie within the same range of RMS error.

The performance of the CNES solution is similar to other solutions up to degree (order) 30. However, unlike other solutions, the degree RMS in the spherical harmonic coefficients of the CNES solutions continues to decrease after degree 30 with an increases after degree 75. The overall performance of the solutions outperforms others. Such an outstanding performance has been achieved via regularization. Unlike a simple Cholesky inversion, in the CNES solutions, the normal matrices are first diagonalized, ordered by decreasing order of the eigenvalues, and only the best-defined sets of linear combinations of the spherical harmonics are solved. Moreover, to obtain an accurate and consistent estimation of the very low degree of the gravity field, the data from GRACE, including the SCA, ACC, KBR, and GPS receiver, are combined with the data from 5 SLR satellites are also used (Lageos, Lageos-2, Starlette, Stella, and Ajisai). Therefore, in the process of the CNES SHs, we do not apply the corrections for the very low degree (see subsection 2.4.3 and subsection 2.4.4 for more details). For more details about the CNES solutions, please see https://grace.obs-mip.fr/.

To evaluate the stochastics of the gravity field solutions, known as formal errors, we have considered a ratio between the average formal errors and the empirical errors following the approach suggested by Kvas et al. (2019). For the average formal errors, we have computed the mean of the reported variance of the spherical harmonics coefficients for monthly solutions from January 2005 to December 2010, which is assumed to hold a homogeneous data quality. In order to estimate the empirical errors, we compute the standard deviation of the coefficients after removing the mean, linear trend, and annual and semi-annual signal. The ratio between these two quantities for all solutions is shown in Figure 2.8. We have excluded solutions that do not provide formal errors, namely IGG-China, LUH, and SWJTU. The optimal value is one, and values below one indicate an underestimation, while values bigger than one show overestimation. ITSG-Grace2018 shows more realistic values among all solutions while formal centers, SWPU, WHU, and AIUB are mostly overestimating, especially for d/o higher than 30. The peculiar pattern in the CNES product is related to the regularization that has been applied to obtain the gravity field from the Level-1 dataset.

It is important to assess the variation in the final TWSA estimation from different GRACE(-FO) Level-2 solutions. To this end, for each of the solutions, we have applied the following steps the necessary steps shown in Figure 2.4 including adding degree-1, replacing C_{20} and C_{30} coefficients, calculating the SH residuals and then filtering using Gaussian with radius 400 km and de-striping approach proposed by Swenson and Wahr (2006). Then, we have aggregated TWSA over the major river basins. Figure 2.9 compares magnitudes and spatial patterns of all
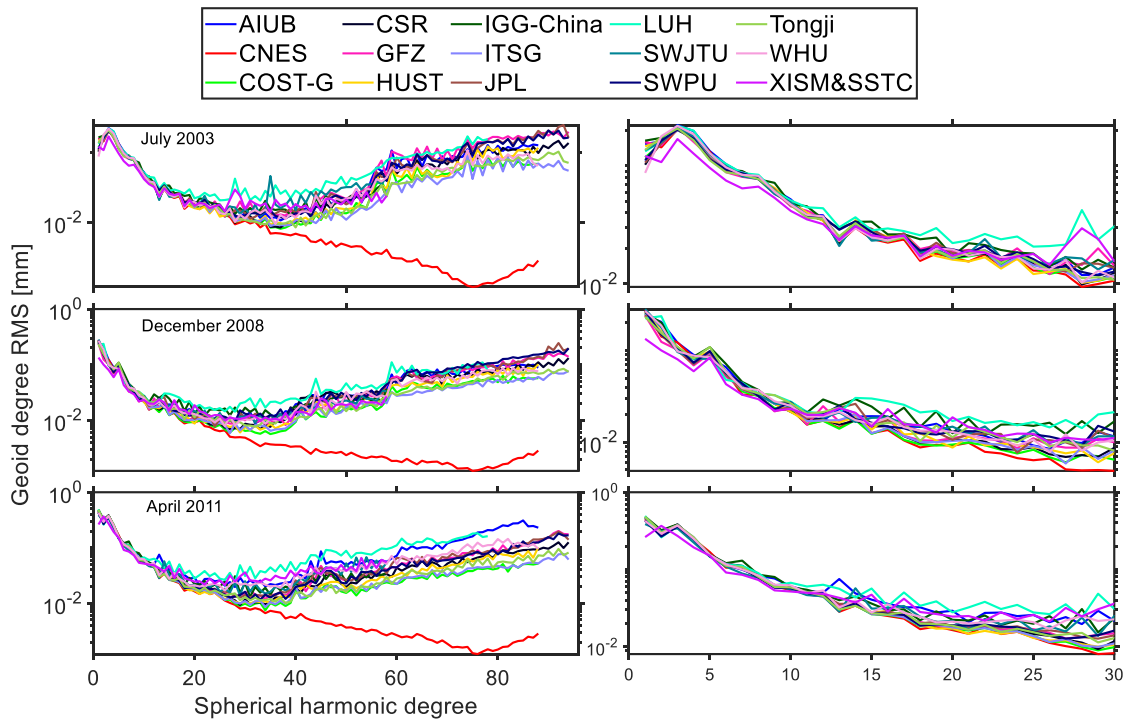
Figure 2.7: Geoid degree RMS of various gravity field solutions with respect to EIGEN6C4 for three epochs.

GRACE(-FO) Level-2 solutions in terms of the coefficient of variation or CV (defined as the ratio between the standard deviation and the mean) of the RMS and linear trend of TWSA within 2003–2011. The full set of solutions would provide a full set of RMS for each basin which provides $CV_{RMS} = \sigma_{RMS}/\mu_{RMS}$. The same can be applied for the linear trends. Moreover, the CV of the RMS and the linear trend are compared based on different categories of climate, latitude zones, basin area, and the RMS of the GRACE TWSA signal (Figure 2.10).

Based on the results, on average, the CV of the RMS is about 0.07 while 70 % of the basins shows CV≤0.10. Significant disparities in RMS were found in arid to hyper-arid areas, e.g., western and eastern Sahara, Gobi, and North China. The lowest range of CV is observed over sub-humid regions, e.g., Ganges, Lena, Nelson, and Amur. In terms of latitude zones, the maximum variation of RMS is shown in the sub-tropical regions, while tropical areas have the minimum disparities. Moreover, based on the results of the boxplots in Figure 2.9 generally, the CV of the RMS from various solutions decreases as the size of the basin increase and has the same range of variation among each category of basins area. Finally, the disparities in RMS are considerable for basins with large (or small) RMS of GRACE signals (RMS<30 or RMS>90), e.g., SanJuan, West Australia, Zambezi, Rio Parnaiba, and Greenland while it is at lowest for basins with RMS within 30–60, e.g., Po, Kolyma, Alabama, and Niger.

In terms of trends, the discrepancies are higher among different solutions. The CV of the trend lies below 0.3 in more than 60 % of the basins. However, 20 % of the river basins have CV>0.8 and even 15 % show CV>1. Relatively low dispersion in trend values was found in sub-humid regions, e.g., Central Amazon, Mississippi, and Central Africa. In contrast, low disparities in the linear trend (e.g., CV<0.4) were found in semi-arid areas, e.g., the southern U.S. and Mexico, eastern Australia, Middle East, Northwest China, and Mongolia. Basins in the sub-tropical region would be affected the least by the choice of solutions regarding trend values. In contrast, the sensitivity of basins located in polar regions to the choice of GRACE solution for trend estimation is significant. The size of the basin also plays a crucial role, with larger basins such as
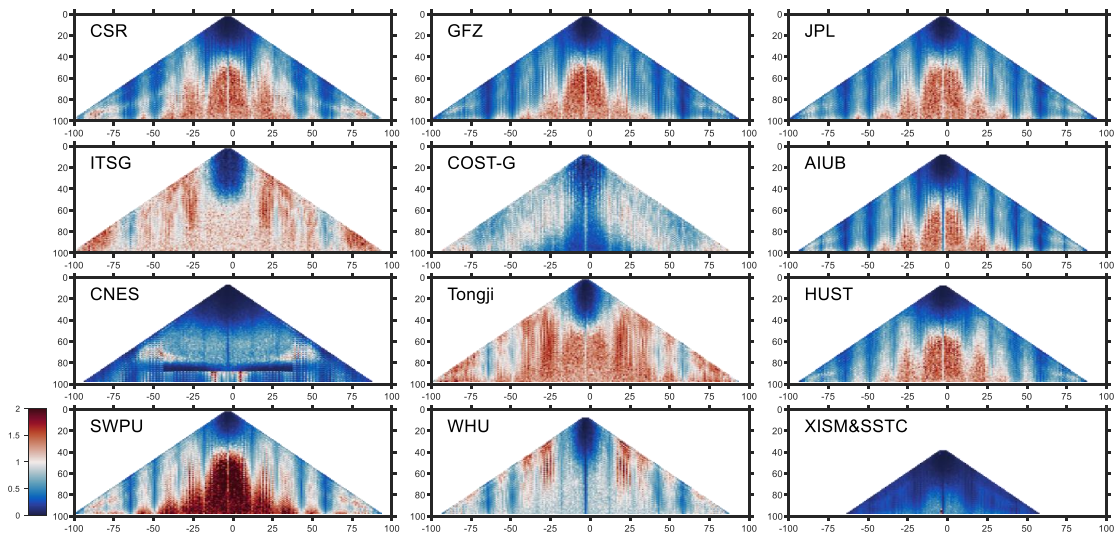
Figure 2.8: The coefficient-wise ratio of average formal errors and empirical errors of the GRACE solutions following the approach suggested by Kvas et al. (2019). The optimal value is one, and values below one indicate an underestimation, while values bigger than one show overestimation. The solutions that do not provide formal errors, namely IGG-China, LUH, and SWJTU are excluded.

Western Dvina, Po, Karun, Savannah, and Alabama exhibiting relatively minor differences. On the other hand, small to medium basins exhibit the largest discrepancies. In terms of the RMS, a clear message from the boxplot is that the disparities in the trend values among the solutions decrease significantly with the increase of the GRACE signal RMS.
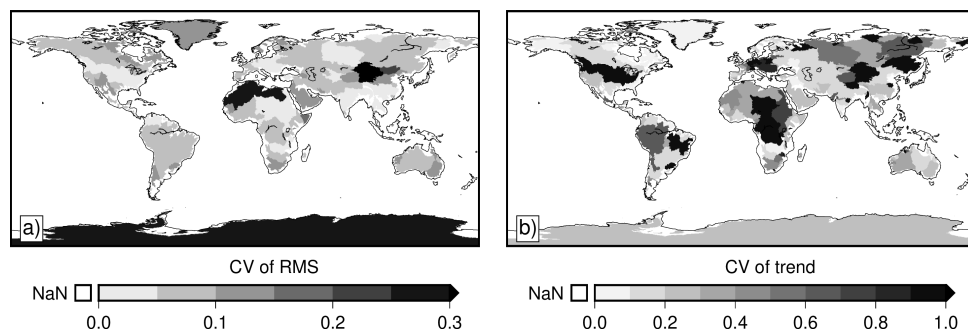

Figure 2.9: Basin-wise CVs of monthly TWSA derived from various GRACE(-FO) Level-2 products across the global land surface in terms of, (a) RMS and (b) the linear trend.

Figure 2.11 illustrates the TWSA retrieved from various GRACE Level-2 solutions over selected river basins. For the post-processing, the degree-1 coefficients are added from the estimation by Sun et al. (2016) and the C_{20} and C_{30} coefficients are replaced with SLR estimations. Moreover, the noisy fields of the residuals are filtered using the de-striping filter (Swenson & Wahr, 2006) and a Gaussian filter with radius 400 km for all solutions. It should be noted that the GIA is not removed from the TWSA field during the post-processing. Among the selected basins, the highest discrepancies between the solution are observed over the Murray Darling basin in Australia, Yangtze in China, and Niger and Nile in Africa. Such significant discrepancies can be related to the difference between the background models. The HUST solution shows outlier values in February 2015 over several basins, namely Death Valley, Danube, Tigris, and Murray Darling.
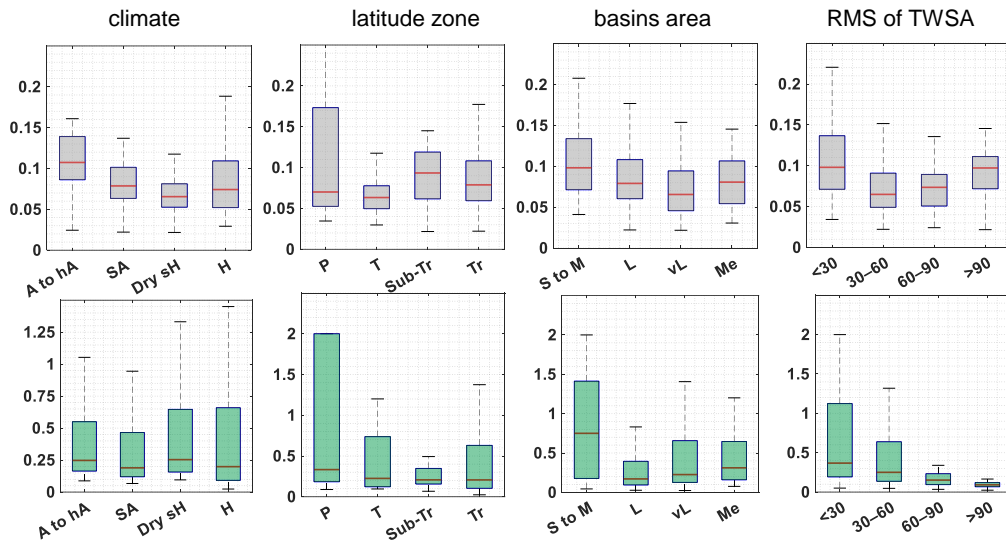
Figure 2.10: Boxplot of the RMS and trend values from various GRACE(-FO) Level-2 products categorized in terms of climate, latitude zone, basin area, and RMS of the GRACE TWSA signal.
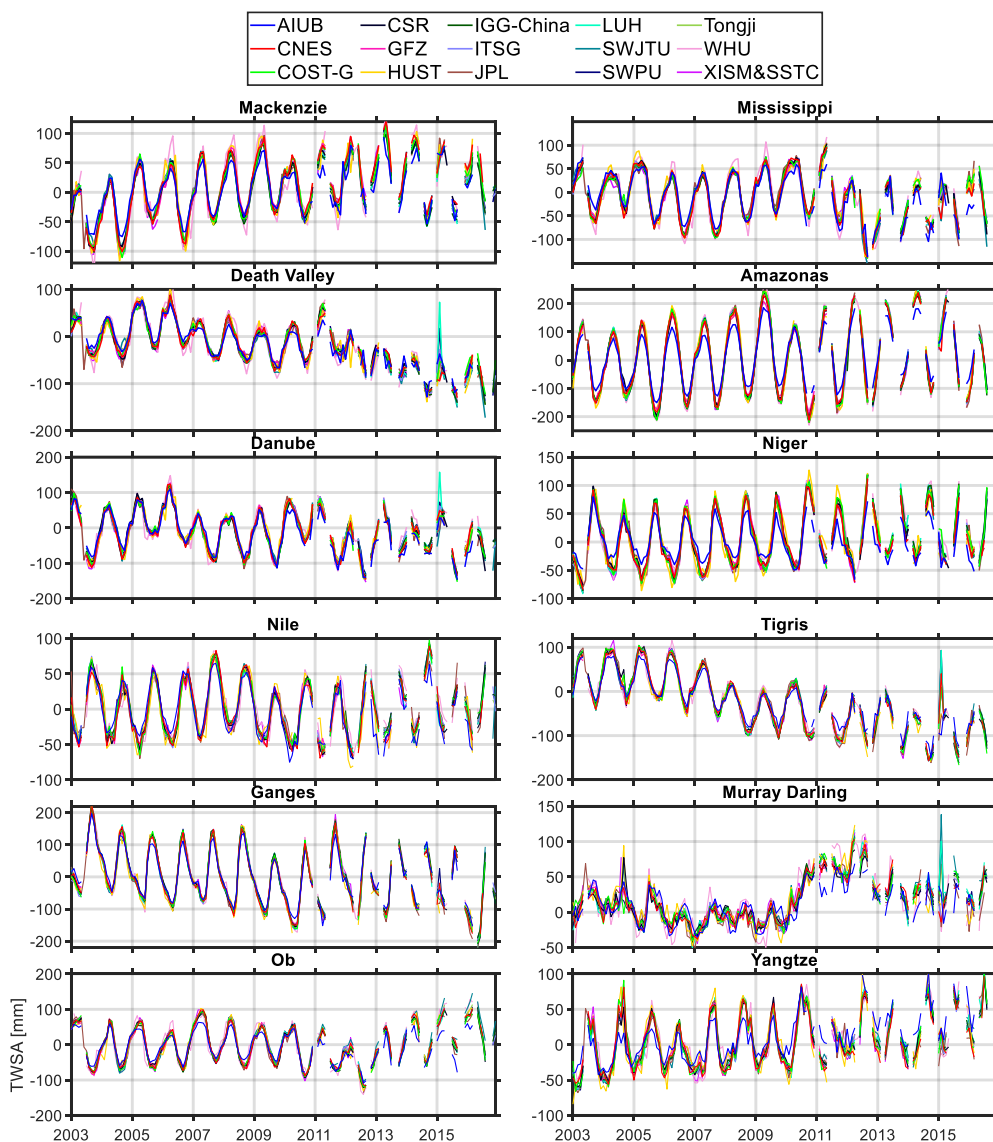

Figure 2.11: GRACE TWSA from various GRACE Level-2 products over selected basins. The noisy fields of the SHs residuals are filtered using de-striping (Swenson & Wahr, 2006) and Gaussian filter with a radius of 400 km. The degree-1 coefficients are added, and C_{20} and C_{30} are replaced with SLR estimations. The GIA trend is not removed from the filtered TWSA fields.

2.4.3. Adding Degree-1 Coefficients

The spherical harmonic coefficients of degree 1 (x: $C_{1,1}$, y: $S_{1,1}$, z: $C_{1,0}$) represent the geocentric motion of the Earth. The geocenter moves with respect to the Center of Figure (CF) of the solid Earth. The origin of the reference frame for the gravity field solutions and also the twin satellites is the Center of Mass (CM). Hence, GRACE is blind to degree 1 signal and therefore cannot measure it. The lack of degree 1 information has the potential to have a significant effect on the recovery of high-latitude mass variability and large-scale inter-basin ocean mass exchange (Swenson et al., 2008). Therefore, degree 1 coefficients have to be added to the GRACE gravity solutions using independent external information.

Different approaches exist for modeling the degree1 coefficients. The geocentric motion can be determined using GPS observations which lack measurements in the oceans and remote areas (Sun et al., 2016). The other method is to use SLR measurements using the distances along the axis of rotation. In another approach, Swenson et al. (2008) and later on Sun et al. (2016) used a numerical ocean model to estimate the degree 1 time-variable gravity coefficients, which are more consistent with the so-called sea-level-equation and the patterns of sea-level rise. In this thesis, the time series of degree-1 and corresponding error information from Sun et al. (2016) is used. The degree 1 coefficients in terms of (fully normalized) coefficients of the geopotential estimated based on (Sun et al., 2016) are available as the GRACE Technical Note #13 in https://podaac.jpl.nasa.gov/gravity/grace-documentation website and is shown in Figure 2.12.
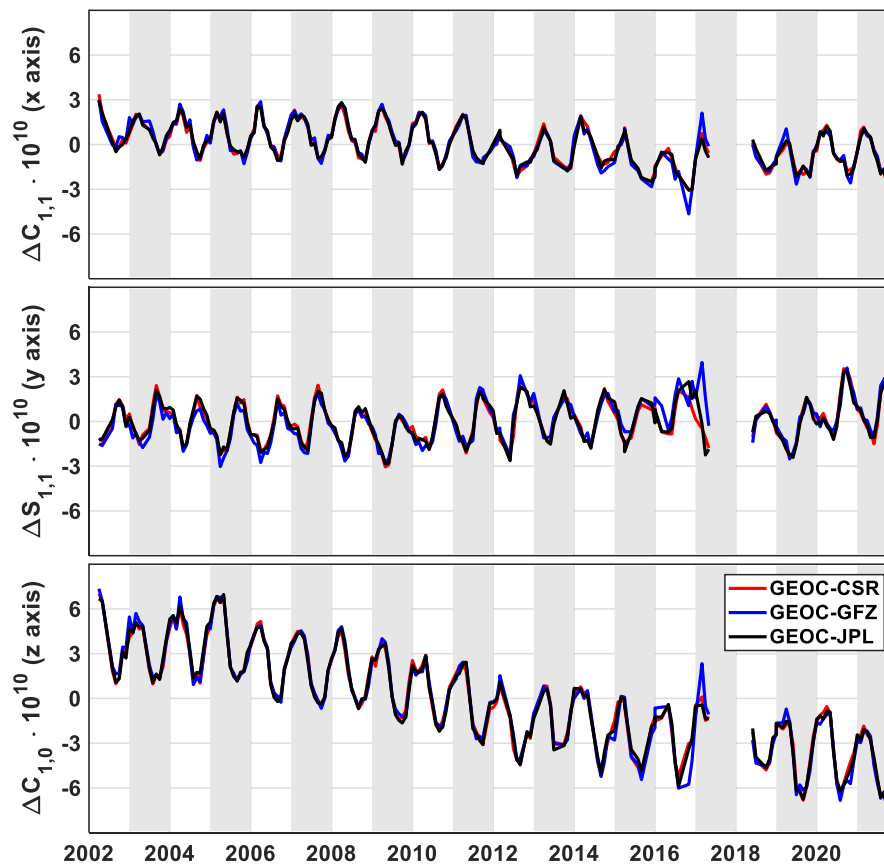

Figure 2.12: Time series of the degree-1 coefficients from SLR estimated by Sun et al. (2016).

2.4.4. Replacing Degree-2 Coefficients

The spherical harmonics of degree 2 represent the Earth's moment of inertia. In particular, the zonal coefficient, $C_{2,0}$, represents the flattening of the Earth, also known as Earth's dynamic oblateness. Satellite observations since 1979 show that $C_{2,0}$ has decreased constantly (Cheng et al., 2013). Two main phenomena contribute to the variation of $C_{2,0}$:

1 GIA following the last ice age

2 Climate related mass exchange between the tropics (23.5°N to 23.5°S latitudes) and the *extratropical* areas

GRACE should be able to reflect the effects of ice mass loss on $C_{2,0}$. However, the GRACE-derived $C_{2,0}$ is not accurate due to its polar orbit and because it is affected by large tidal aliases (Liu, 2019). The SLR tracking from geodetic satellites provides a unique estimation of C_{20}. The SLR-derived $C_{2,0}$ is derived from 5 satellites, namely LAGEOS-1 and 2, Stella, Starlette, and AJISAI. The coefficients can be accessed from the GRACE Technical Note #14 in https://podaac.jpl.nasa.gov/gravity/grace-documentation website. Figure 2.13 illustrates the time series of the $C_{2,0}$ and the $C_{3,0}$ coefficients from ITSG-Grace2018 and the release 6 of GFZ, SLR, and JPL compared with the SLR estimation. More discrepancies exist between the GRACE observations and the SLR values in estimating the degree-2 coefficients than degree-3. The GRACE solutions show similar variation compared with the SLR estimation in estimating the $C_{3,0}$. However, the RMS of the degree-3 coefficients from various solutions shows higher RMS for the GRACE-FO period
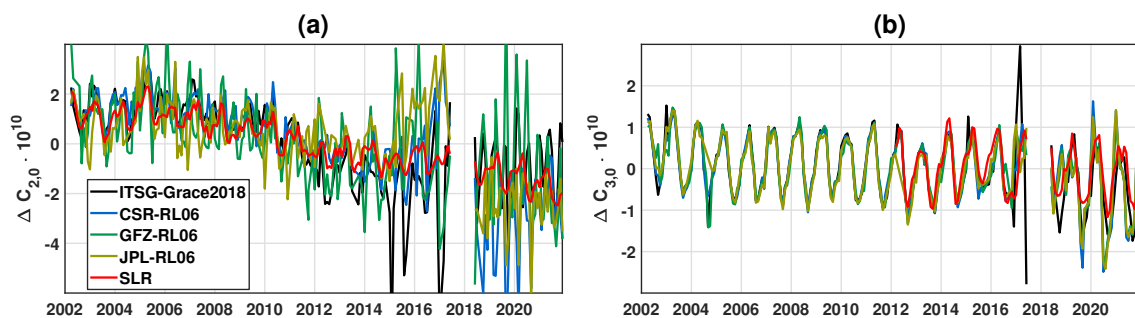

Figure 2.13: Time series of the $C_{2,0}$ and the $C_{3,0}$ coefficients from ITSG-Grace2018 and the release 6 of GFZ, SLR, and JPL compared with the SLR estimation of the five satellites.

2.4.5. Ellipsoidal Correction

It is common to use a spherical approximation to synthesize the GRACE(-FO) solutions. The true form of the Earth resembles more an ellipsoid than a sphere, with the radius of the Earth at the north (or south) pole being ∼ 22 km shorter than that at the equator. Therefore, the mentioned approximation would cause noticeable bias, especially in high-latitude regions. Li et al. (2017) showed that the differences (underestimations) can be 4.3 % over Greenland to 6.6 % for Svalbard Islands. In this thesis, each solution is corrected from spherical to ellipsoidal coefficients following the iterative method proposed by Li et al. (2017). The spherical approximation leads to underestimation, and removing the underestimation is called the ellipsoidal correction.

Figure 2.14 compares the global field of TWSA with and without ellipsoidal correction for January 2004. The analysis reveals significant differences in the Northern Antarctic Peninsula, Greenland's periphery, and the East Canadian Arctic Archipelago regions. To determine the

effect of ellipsoidal correction on the RMS and trend of TWSA at the basin scale, we have calculated the relative RMS difference $(\text{RMS}(\text{TWSA}_{\text{without}}) - \text{RMS}(\text{TWSA}_{\text{with}}))/\text{RMS}(\text{TWSA}_{\text{without}})$ for the three official centers, namely CSR, GFZ, and JPL. We also performed the same calculation for the linear trends, and the results are presented in Figure 2.15. The relative difference in estimating RMS ellipsoidal correction are mostly positive (correction increases the RMS values) and increase by latitude which can be as high as 6 %. In terms of the effects on the trend values, the correction mostly leads to an increase in trend values (0–15%). However, in some regions such as central Australia, mid and southern South America, and central China, the correction results in a decrease in trend values by up to 10 %.

To investigate the spatial distribution of the significance of the ellipsoidal correction, the relative RMS and trend values are plotted within specific categories of climate, latitude zones, basin area, and the RMS of the GRACE TWSA Figure 2.16. The results from the climate categories show that sub-humid to humid regions are more affected by the ellipsoidal correction in terms of RMS and trend values. As expected, the effect of the ellipsoidal correction increases by latitude and is at its highest over Polar regions. Moreover, the basin size does not play a significant role in the effect of corrections for estimating the RMS and trend values but shows a higher range of variations over small to large basins than very large or mega basins. Although basins with higher RMS of the TWSA signal show a bigger difference regarding ellipsoidal corrections, this might be related to the fact that many of such regions are located at higher latitudes; the boxplots should not be over-interpreted.

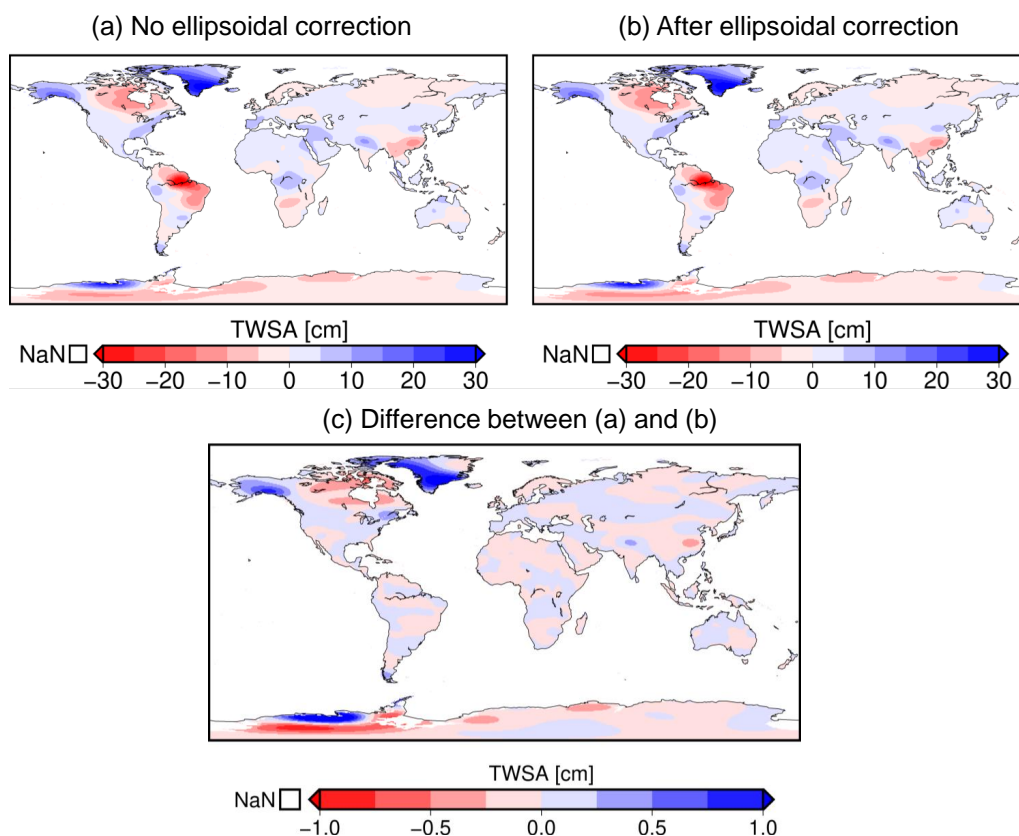Figure 2.14: The global gridded TWSA values with ellipsoidal correction and without for January 2004 (a and b). (c) the differences obtained by subtracting (b) from (a). Note the difference in color bar scales. For all plots the residual of the CSR Release 06 solution with respect to 2004–2010 is filtered by a Gaussian filter with filter radius 400 km after applying de-striping filter proposed by Swenson and Wahr (2006).
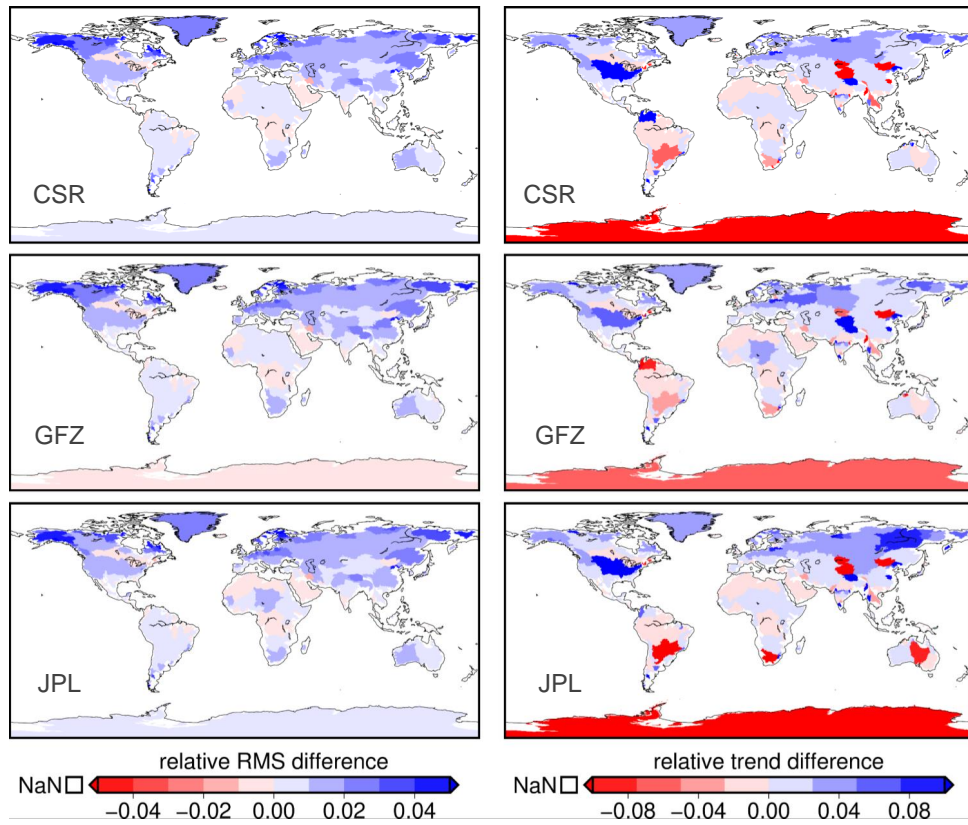
Figure 2.15: Basin-wise distribution of the relative RMS and trend values of the effect of ellipsoidal corrections for the three official GRACE(-FO) solutions.
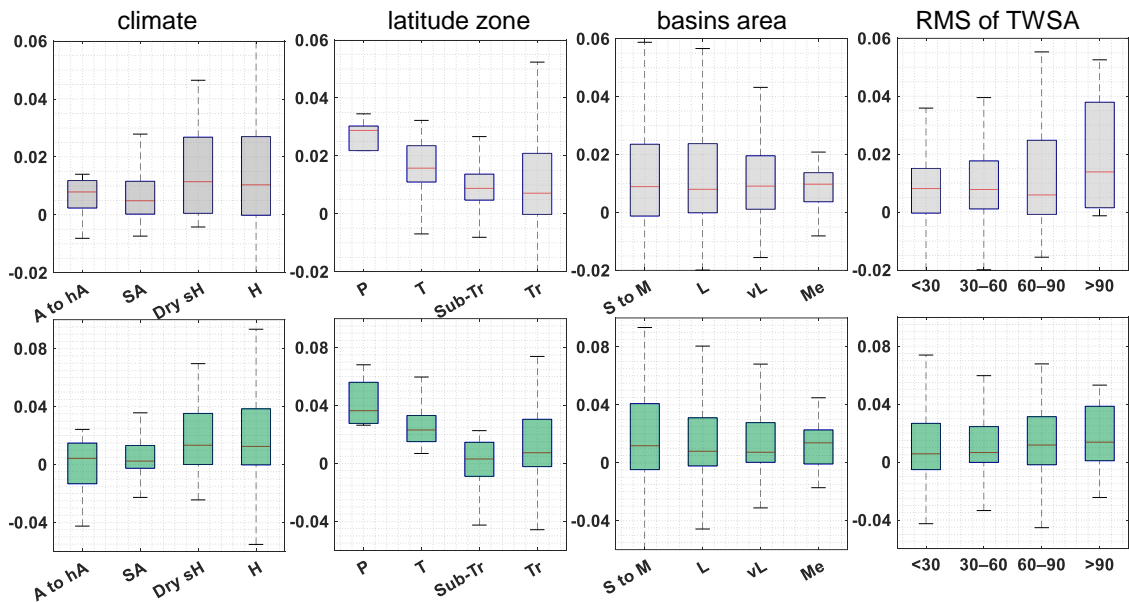
Figure 2.16: Boxplot of the relative difference in RMS and trend values of the TWSA over major river global basins before and after ellipsoidal correction in terms of climate, latitude zone, basin area, and RMS of the GRACE TWSA signal. The CSR RL06 GRACE(-FO) Level-2 products is used as the input and the noisy fields of the residuals are filtered using de-striping filter (Swenson & Wahr, 2006) and Gaussian with radius 400 km.

2

2.4.6. Removing the Remaining Tidal Aliasing Errors

Undersampling the mass motions from GRACE's orbital configuration results in aliasing (Seo et al., 2008). Due to imperfect tidal models, GRACE SHs are contaminated by residual tidal aliasing error, a primary and a secondary one (Liu & Sneeuw, 2021; Tourian, 2013). To avoid mixing the real seasonal variation with the tidal error such a remaining aliasing error should be corrected. Therefore, within the post-processing of GRACE(-FO) Level-2 solutions, we eliminate the primary and secondary tidal aliasing errors of the main tidal constituents, S1, S2, P1, K1, K2, M2, O2, O1, and Q1 from GRACE monthly solutions using a least-squares Fourier analysis proposed by Tourian (2013). To this end, a parametric model of $y = Ax + e$ can be defined:

$$\Delta C_{lm}(t) = a_1 t + \sum_{i}^{N} b_i \cos(\omega_i t) + c_i \sin(\omega_i t), \tag{2.5}$$

$$\Delta S_{lm}(t) = a_2 t + \sum_{i}^{N} d_i \cos(\omega_i t) + e_i \sin(\omega_i t), \tag{2.6}$$

where ΔC_{lm} and ΔS_{lm} are the spherical harmonics residuals, a_1 and a_2 are the trend coefficients, N is the number of tidal constituents, ω_i is the tidal aliasing angular frequency and the b_i, c_i, d_i, e_i are the unknown coefficients of the model. The amplitude of each tidal aliasing error can be calculated by:

$$A_i = \sqrt{b_i^2 + c_i^2} \tag{2.7}$$

Based on the analysis conducted by Tourian (2013), removing tidal aliasing error from GRACE monthly solutions improves the solutions towards being more consistent with the hydrometeorological water storage change. Moreover, the maximum tidal aliasing error occurs over Greenland, where the investigation into its ice mass loss plays an important role in climatic studies. Besides, the impact of removing the error on trend is not significant according to the given uncertainty in GRACE data (Table 2.4). The low impact of tidal aliasing error on the estimated trend is better represented in the frequency domain, where the low frequencies belonging to the trend are not affected after removing the error.

2.4.7. Filtering Noise

The GRACE(-FO) monthly gravity field solutions suffer from correlated noise (cf. Figure 2.17), resulting from measurement errors, orbit geometry, sensitivity of the observable, and undersampling. Due to the near-polar orbit, GRACE(-FO) measurements are sensitive north-south, while the sensitivity is weaker in the east-west direction. Besides, temporal aliasing caused by imperfect background models and the sampling frequency would result in an error in the solutions. Moreover, the GRACE(-FO) level 2 coefficients are contaminated by the instrument noise resulting from the orbit errors, star cameras, accelerometer observations, and the noise in the K-band ranging system. The aforementioned errors manifest themselves as distinctive North-South striping patterns (Swenson & Wahr, 2006; Wahr et al., 2006). To investigate any physical signal, such noise must be removed (or suppressed) from the GRACE(-FO) solutions. There is a trade-off between the resolution and noise, i.e., increasing the degree of spherical harmonics coefficients would lead to higher resolution but also greater noise. Conversely, decreasing the degree of coefficients would result in lower resolution but less noise. Therefore, a straightforward way to lower the noise would be to truncate the spherical harmonic series around degrees 20–30 (cf. Figure 2.7). However, the monthly results from higher degrees include a complex noise for applications with high spatial resolution demand, such as in hydrology.

Table 2.4: The estimated linear trend from GRACE for Jan. 2005 to December 2011 for the 19 drainage basins of Greenland (Tourian, 2013).

Drainage basin	Area [10^3 km^2]	Linear Trend [Gt/yr]	
		before	after
A1	130	-7.8	7.7
A2	63	-5.6	-5.5
A3	46	-5.0	-4.9
A4	17	-4.2	-4.1
B1	273	-11.2	-11.1
B2	50	-8.0	-7.9
C1	148	-9.8	-9.6
C2	35	-9.8	-9.6
C3	72	-12.2	-12.0
D1	64	-11.5	-11.3
D2	47	-13.0	-12.8
D3	33	-11.7	-11.5
E1	49	-11.0	-10.9
F1	48	-11.6	-11.5
F2	135	-13.9	-13.7
G1	95	-10.8	-10.7
G2	129	-12.6	-12.4
H1	240	-16.1	-15.8
H2	34	-16.3	-16.1
Total	1718	-202	-199

Several filters have been developed to reduce the noise in the GRACE(-FO) solutions. Wahr et al. (1998) utilized the low-pass isotropic Gaussian smoothing operator designed by Jekeli (1981) which was further developed as an anisotropic version in Han et al. (2005) and Zhang et al. (2009). Other examples of the spectral filters include the Wiener filter (Klees et al., 2008; Sasgen et al., 2006), the denoising and decorrelation kernel (DDK) filter (Kusche et al., 2009), and the regularization filter (Devaraju & Sneeuw, 2017). Such filters use the error variance-covariance matrix but also require a priori signal covariance information (Devaraju, 2015; Duan et al., 2009). Another category of filters attempts to improve the signal-to-noise ratio by removing the correlation in the spectral domain (stripes in the spatial domain). Swenson and Wahr (2006) proposed such a de-striping filter by smoothing the spherical harmonics coefficients for a particular order m with a polynomial of order p using least squares estimation. Following Swenson and Wahr (2006) several filters have been developed which investigate such a pattern in the unconstrained solutions (Chambers, 2006; Chen et al., 2007; Duan et al., 2009). These filters are mostly followed by a spatial filter like a Gaussian filter to reduce the remaining noise. Another category of filters benefits from the temporal information in the GRACE(-FO) data including the Empirical Orthogonal Function (EOF) filter (Schrama et al., 2007; Wouters & Schrama, 2007), the Independent Component Analysis (ICA) filter (Frappart et al., 2011), the stochastic filter (Wang et al., 2016), the Singular Spectrum Analysis (SSA) filter (Prevost et al., 2019). Recently Yi and Sneeuw (2022) introduced a new filter named Spatial SSA (SSAS), which mainly follows the strategy suggested by Swenson and Wahr (2006) while applying SSA on the spatial domain to suppress the residual noise. Figure 2.17 illustrates the global field of TWSA for April 2006 retrieved from the release 06 of the CSR solution and applying commonly used filters together with the noisy field.
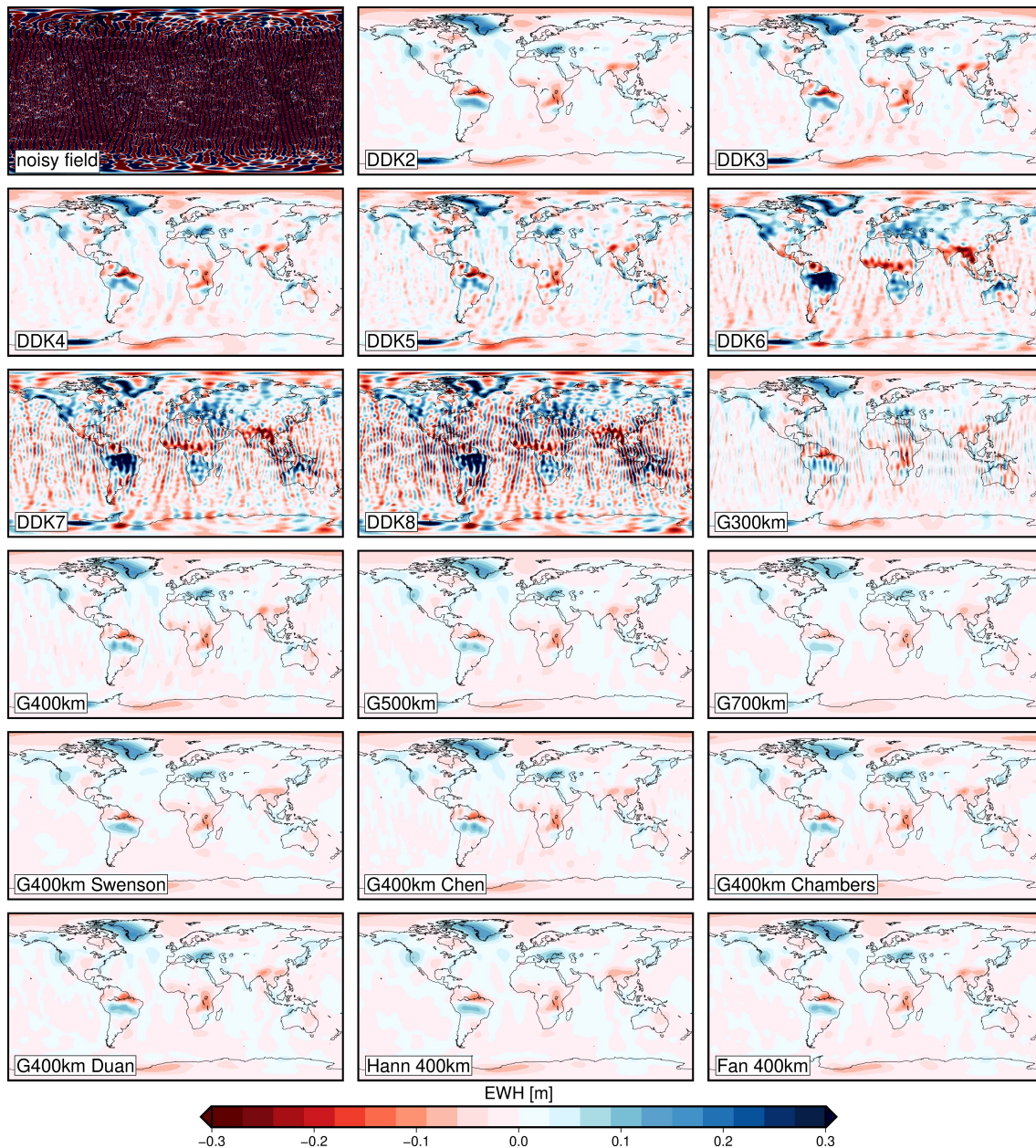
2

Figure 2.17: The Global field of the TWSA retrieved from the noisy data of GRACE and also after applying various commonly used filters for April 2006. The CSR RL06 has been used as the solution for the plots.

To assess the impact of filter choice on the estimation of TWSA from GRACE(-FO) Level-2 data, we conducted a comparative analysis of the RMS and linear trend of TWSA obtained using various filters, as shown in Figure 2.18. We also examined the CV of the TWSA signal in different climate categories, latitude zones, basin areas, and RMS values using boxplots, as illustrated in Figure 2.19. Our analysis of the TWSA signal across various basins and regions suggests that the choice of filter significantly affects the RMS of the TWSA signal in arid to semi-arid regions. Specifically, we observed that the effect of filter choice on the RMS of the TWSA signal varied depending on the latitude zone of the basin. While this effect was minimal in polar and tropical regions, it was significant in tropical basins of Africa, such as Niger, Nile, Congo, and Chad. Our findings also indicate that the choice of filter did not have a significant impact on trend values, although the CV of trends was generally higher in semi-arid regions. Notably, we found that the effect of basin size on the RMS or linear trend estimation of the TWSA signal was substantial,

as it significantly reduced the discrepancies among different filters.

Additionally, we observed that the disparities between various filters decreased as the magnitude of the TWSA signal's RMS increased. This suggests that the choice of filter may have a greater impact on the analysis of TWSA signals in regions where the signal's variability is low. Furthermore, our analysis revealed that the estimation of trends exhibited substantial differences in polar and tropical regions. This finding suggests that researchers should exercise caution when interpreting TWSA trends in these regions and carefully consider the choice of filter. In conclusion, our study highlights the importance of carefully selecting a filter when analyzing TWSA signals, particularly in arid to semi-arid regions and tropical basins in Africa. Additionally, our findings suggest that basin size and the magnitude of the TWSA signal's RMS should be taken into consideration when selecting a filter. These insights can help researchers to better understand and interpret TWSA data, ultimately improving our understanding of global water resources.
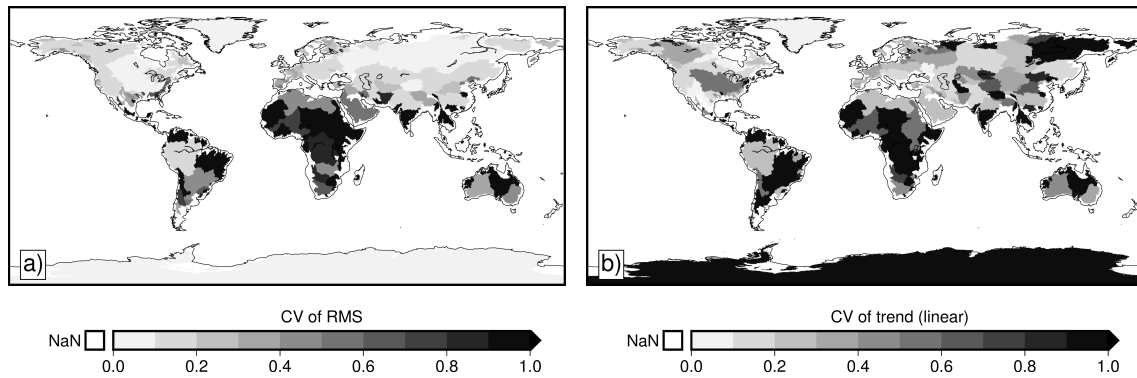
Figure 2.18: Basin-wise CVs of monthly TWSA derived from various filters and CSR RL06 Level-2 product across the global land surface in terms of, (a) RMS and (b) linear trend.
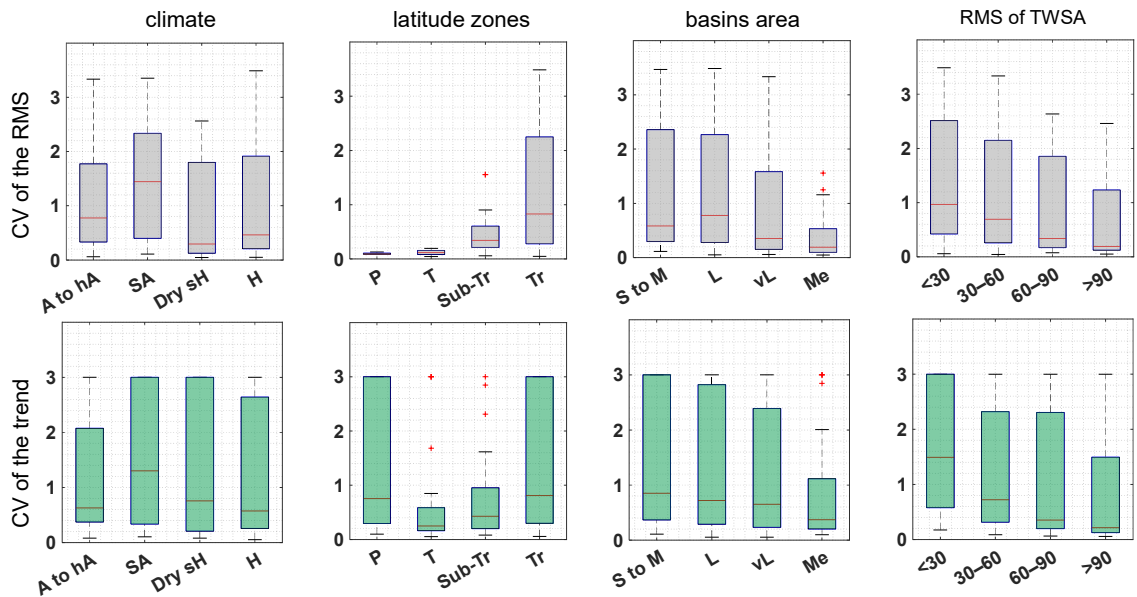

Figure 2.19: Boxplot of the relative difference in RMS and trend values of the TWSA over major river basins from various filters categorized in terms of climate, latitude zone, basin area, and RMS of the GRACE TWSA signal. The CSR RL06 GRACE(-FO) Level-2 products are used as the input.

2.4.8. Correcting Leakage

Filtering, in the context of GRACE data analysis, can be viewed as a spectral multiplication operation that selectively modifies the spectral content of the signal. However, it is important to note that leakage, which occurs during the spectral analysis or transformation, is primarily a consequence of spatial convolution operations. Correcting leaked signal is an inevitable step, especially in coastal regions. Various approaches have been proposed to restore the GRACE TWSA signal at pixel and basin scale including:

- (1) the scaling factor approach proposed by Landerer and Swenson (2012) which suggest a scaling factor by least squares fit between the filtered and unfiltered field of a Land Surface Model (LSM),

- (2) the additive correction approach by Klees et al. (2007) that add the difference between the unfiltered field of TWSA from a Land Surface Model (LSM) and the filtered field (same filter as applied to the GRACE noisy field) and excluding the signal outside of the area of interest (basin or a grid cell),

- (3) the multiplicative correction approach proposed by Longuevergne et al. (2010) which adjusts the filtered basin function's amplitude to that of the unfiltered basin function and multiplies the difference between the filtered GRACE signal and the leakage,

- (4) the data-driven approach proposed by Vishwakarma et al. (2016) which follows the idea of the multiplicative approach while the leakage inside and outside of the basin is reduced using a catchment mask and a filter kernel,

- (5) the forward modeling algorithm proposed by Chen et al. (2015) which reduces the leaked signal via an iterative process that makes incremental changes to a model until the difference between the filtered model and filtered GRACE is minimal.

The above-mentioned models have been extensively compared at basin and grid-level (Khaki et al., 2018; Long et al., 2015; Vishwakarma et al., 2017). In this thesis, we deliberately exclude further analysis and use the forward modeling approach for two main reasons. First, the forward modeling has shown superior performance over the first three methods (Long et al., 2016) and second, the data-driven, despite its great performance over multiple comparisons (Vishwakarma et al., 2017), can only be applied over a specific region (basin) and not grids.

2.4.9. Glacial Isostatic Adjustment

The Earth's crust is slowly recovering from the ice load of the past ice ages and sea-level change. During the ice ages, the continental and polar ice sheets and mountains glaciers expanded due to the significant reduction in Earth's temperature. The most recent global deglaciation even (started 21 000 years ago and was completed about 6 000 years ago) has unloaded a huge amount of mass from land to the ocean, which is a major contributor to the sea level variations that has been observed recently.

The Earth's geopotential and geoid change includes the GIA signal. Therefore, to compute the present surface mass change (i.e., trends in water content of hydrologic basins, or ocean bottom pressure, or ice sheet mass) the GRACE(-FO) data must be corrected for GIA (e.g., Geruo et al., 2013; Peltier, 2009). It is noteworthy that a GIA surface mass equivalent is not a physical mass that exists at the surface; rather, it is a hypothetical equivalent mass that would produce the same geopotential effect if observed from space. Several GIA models exist to account for the GIA. Geruo et al., 2013 simulated the viscoelastic response of a compressible Earth to the

surface load. Peltier (2009) developed the ICE6G-D model by computing the solid Earth viscoelastic response to forcing applied by surface mass redistributions during waxing and waning of Pleistocene glaciations using more data. Geruo et al. (2013) used a Bayesian hierarchical modeling approach to combine GPS measurements with a new geophysical GIA model based on the ICE-6G_C (VM5a) deglaciation history (the Wahr Model in the Figure 2.20). More recently, Caron et al. (2018) have adapted new GIA modeling and inversion approaches that enable a physical uncertainty quantification of GIA model errors and uncertainty ranges across a range of realistic GIA model parameters. Recently Sun and Riva (2020) have developed a global empirical model of present-day GIA, solely based on GRACE data and on geoid fingerprints of mass redistribution. The global representation of the GIA annual trend values from the four datasets mentioned above is shown in Figure 2.20.
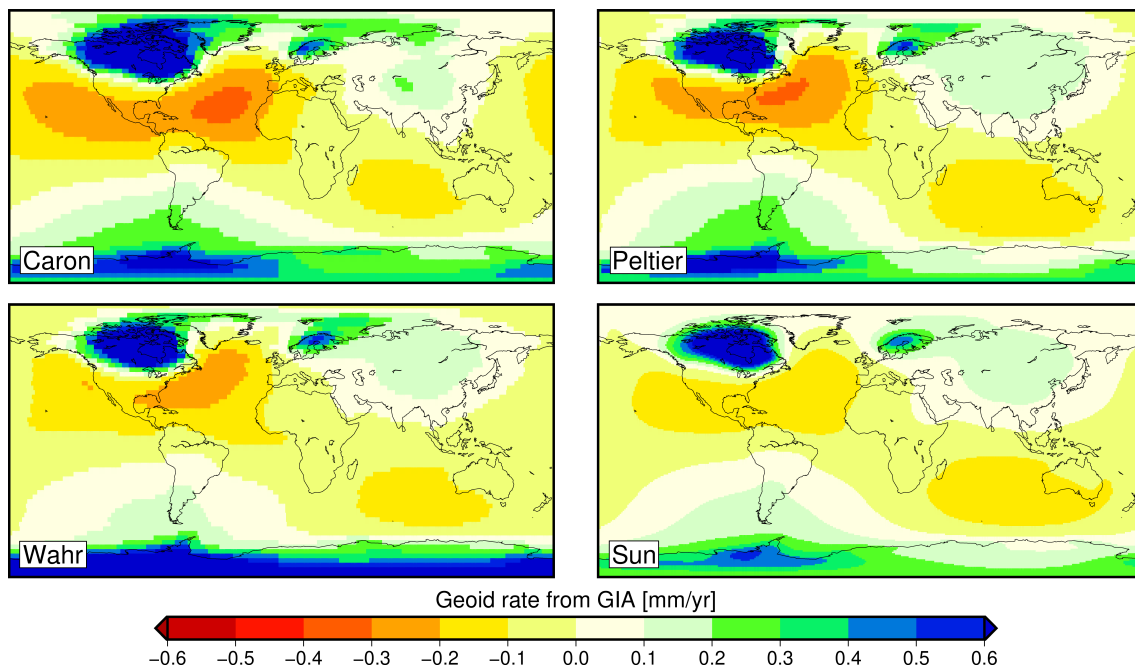

Figure 2.20: Contemporary geoid rates (in mm/yr) from GIA as predicted by the four GIA models.

2.4.10. Filling gaps

During the GRACE mission, there were some months when the data from the satellites were not available due to various issues such as equipment malfunctions or maintenance activities. These gaps in data availability can affect the continuity and accuracy of the gravity field solutions. The GRACE Follow-On mission, which was launched in 2018, is a continuation of the GRACE mission and aims to continue the measurements of changes in the Earth's gravity field. However, there is a gap in the data between the end of the original GRACE mission in 2017 and the beginning of the GRACE Follow-On mission in 2018 (Figure 2.21). This gap in data can affect the continuity of the time-series of gravity field solutions. Moreover, gaps in data availability can also make it difficult to study long-term trends and changes in the Earth's gravity field, which is one of the main goals of the GRACE mission.

Several methods have been proposed to address this issue. One approach is to use data from other satellite missions, such as Swarm satellites and SLR, to reconstruct TWSA and reproduce TWSA variations at large scales (Löcher & Kusche, 2021; Richter et al., 2021). However, the spatial resolution of these datasets is not comparable to GRACE and GRACE-FO datasets, leading to large uncertainty for some small catchments and high-latitudes (Li et al., 2021). EOF analy-
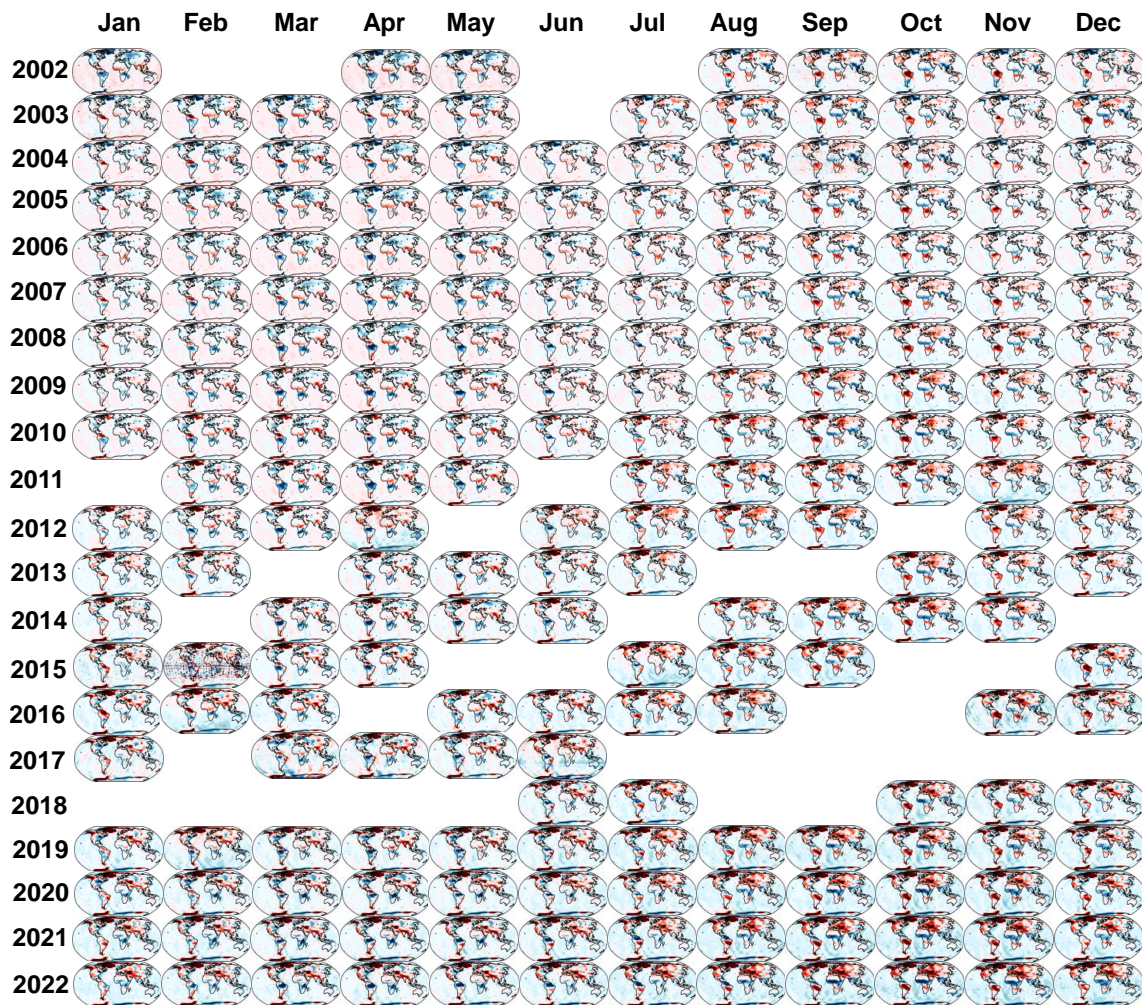
Figure 2.21: The global monthly variation of the TWS observed by GRACE mission. To estimate the TWS change, the unconstrained solution from the Institute of Geodesy of the Graz University of Technology (ITSG-Grace2018) is used.

sis is a statistical technique that can be used to extract dominant patterns of variability from a dataset. It has been used to fill the temporal gap in GRACE data by estimating the missing data based on the correlation between the dominant patterns of variability in GRACE and other datasets, such as the Global Land Data Assimilation System (GLDAS) dataset (Lenczuk et al., 2022; Sun et al., 2021; Zhang et al., 2022).

Another approach involves using hydroclimatic factors as predictors and reconstructing GRACE-derived TWSA at the gridded resolution. For example, Humphrey et al. (2017) and Humphrey and Gudmundsson (2019) used precipitation and temperature as explanatory variables and reconstructed the de-seasonalized and de-trended TWSA using a statistical model. Sun et al. (2019) employed a deep neural network with precipitation, temperature, and soil moisture in the GLDAS model as explanatory variables for the prediction of TWSA. Li et al. (2021) reconstructed global TWSA for the gap period using various statistical methods. Moreover, regional TWSA reconstructions have been accomplished for India (Kumar et al., 2022), Africa (Ahmed et al., 2019), and China (Chu et al., 2023). Despite the success of these methods, their accuracy and reliability still depend on various factors, such as the quality of the input data, the selection of predictor variables, and the statistical methods used for reconstruction.

Yi and Sneeuw (2021) proposed a non-parametric and data-adaptive approach to fill the missing epochs in the monthly gravity product of the GRACE and GRACE-FO missions. The approach is composed of Singular Spectrum Analysis gap-filling technique, cross-validation, and spectral testing for significant components to produce reasonable gap-filling results in the form of Spherical Harmonic (SH) coefficients. The quality of the gap-filling product is evaluated through comparison with a surface mass balance based estimate in Greenland, Swarm gravity solutions and a climate-driven water storage model, all of which confirm the good performance of the results. The authors demonstrate that the proposed method effectively suppresses noise and can reduce noise in the oceans without sacrificing resolutions on land. The method is applicable to smoothed gridded observations and provides a ready-to-use gap-filling product in the form of SH coefficients with proper error estimates. In our study, we have used this method to fill the gap in GRACE and GRACE-FO data.

2.5. GRACE vs GRACE-FO

One way to assess the accuracy of GRACE TWSA observations is to compare them with altimetric water level time series over large lakes. Considering the spatial resolution of GRACE, we have selected several lakes, including the Caspian Sea, Superior, Victoria, Huron, Michigan, Tanganyika, Malawi, Erie, and Turkana (Figure 2.22 and Table 2.5). To conduct the evaluation, we have used the solutions that incorporate GRACE-FO level-2 data and all three mascons products: JPL, CSR, and GSFC. To process the spherical harmonic coefficients, we followed the flowchart previously outlined in Figure 2.4, using a Gaussian filter with a 400 km radius and the destriping method proposed by Swenson and Wahr (2006). Additionally, to correct leakage in the SH coefficients, we employed the data-driven approach as suggested by Vishwakarma et al. (2017). In addition, we also computed the ensemble mean of the mascons products and the ensemble mean of the processed solutions. By calculating the ensemble mean, we can mitigate the influence of the specific solution or product chosen and obtain a more robust estimate of the TWSA. This approach can improve the accuracy of the evaluation and provide more reliable results.
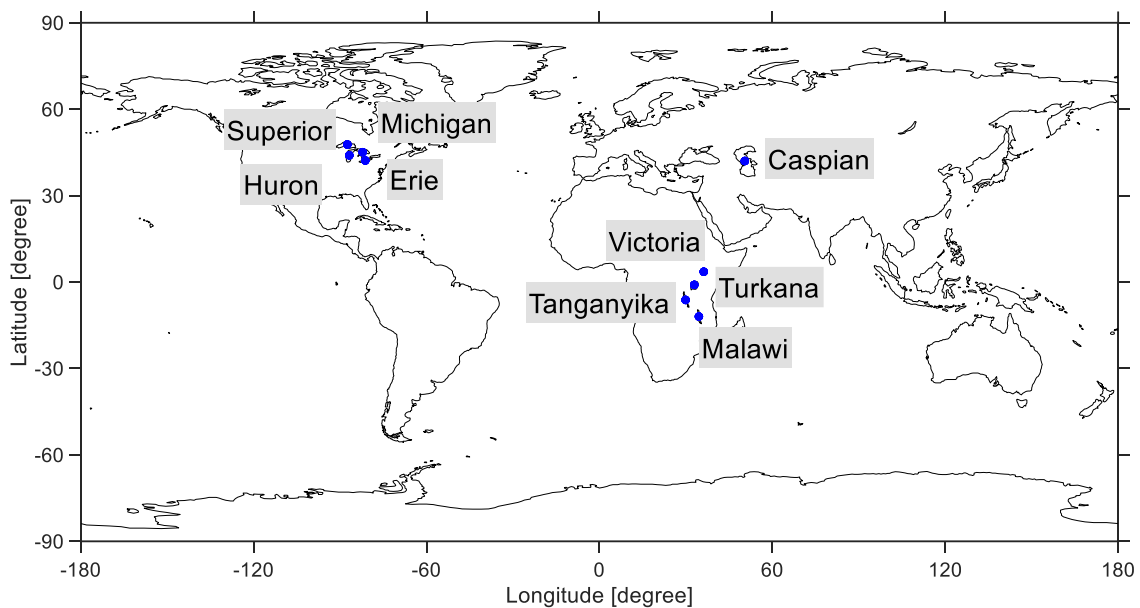

Figure 2.22: Distribution of the selected lakes.

Figure 2.23 displays the time series of GRACE(-FO) data in comparison to the altimetric water level data for a selection of lakes. The entire time period under consideration is divided into four segments: the mission duration (2002-2020), the GRACE-FO period (2018-2020), the pre-GRACE-FO period (2002-2010), and an intermediate period (2011-2017). The GRACE period is further subdivided into two periods, the first from 2002 to 2010, during which the satellites were in their most stable period, and the second from 2011 to 2017, which experienced some gaps and battery problems in the last few months. The envelope surrounding the GRACE(-FO) time series represents the 1-σ uncertainty, which is estimated from various calculations based on the mascon products of SHs solutions.

In Table 2.6, the correlation coefficients between GRACE TWSA observations and altimetric water level time series over the selected lakes for the four periods mentioned above are presented. The correlation coefficients were computed for two types of GRACE solutions: mascons and processed SH coefficients. The results reveal that the accuracy of the TWSA observations varies across different lakes, time periods, and types of GRACE solutions. Some lakes, such as Superior and Turkana, consistently exhibit high correlation coefficients across all time periods and GRACE solutions, while others, such as Huron and Erie, show lower correlations.

The correlation coefficients between GRACE and altimetric water level time series are generally higher than those between GRACE-FO and the same time series across all lakes and time periods. For example, the correlation coefficient between GRACE and the altimetric water level time series for Superior is consistently higher than that for GRACE-FO for all time periods and types of GRACE solutions. However, some lakes, such as Malawi and Turkana, display very high correlation coefficients between GRACE-FO and the altimetric water level time series, even higher than those between GRACE and the same time series. Overall, while there are variations across different lakes, time periods, and types of GRACE solutions, the general trend suggests that GRACE is more accurate than GRACE-FO. It is important to note that the results should not be over-interpreted due to the coarse spatial resolution of GRACE, the use of only correlation coefficients, and the limited scope of lakes considered in the assessment.

Table 2.5: Summary of the selected lakes with their surface area and type.

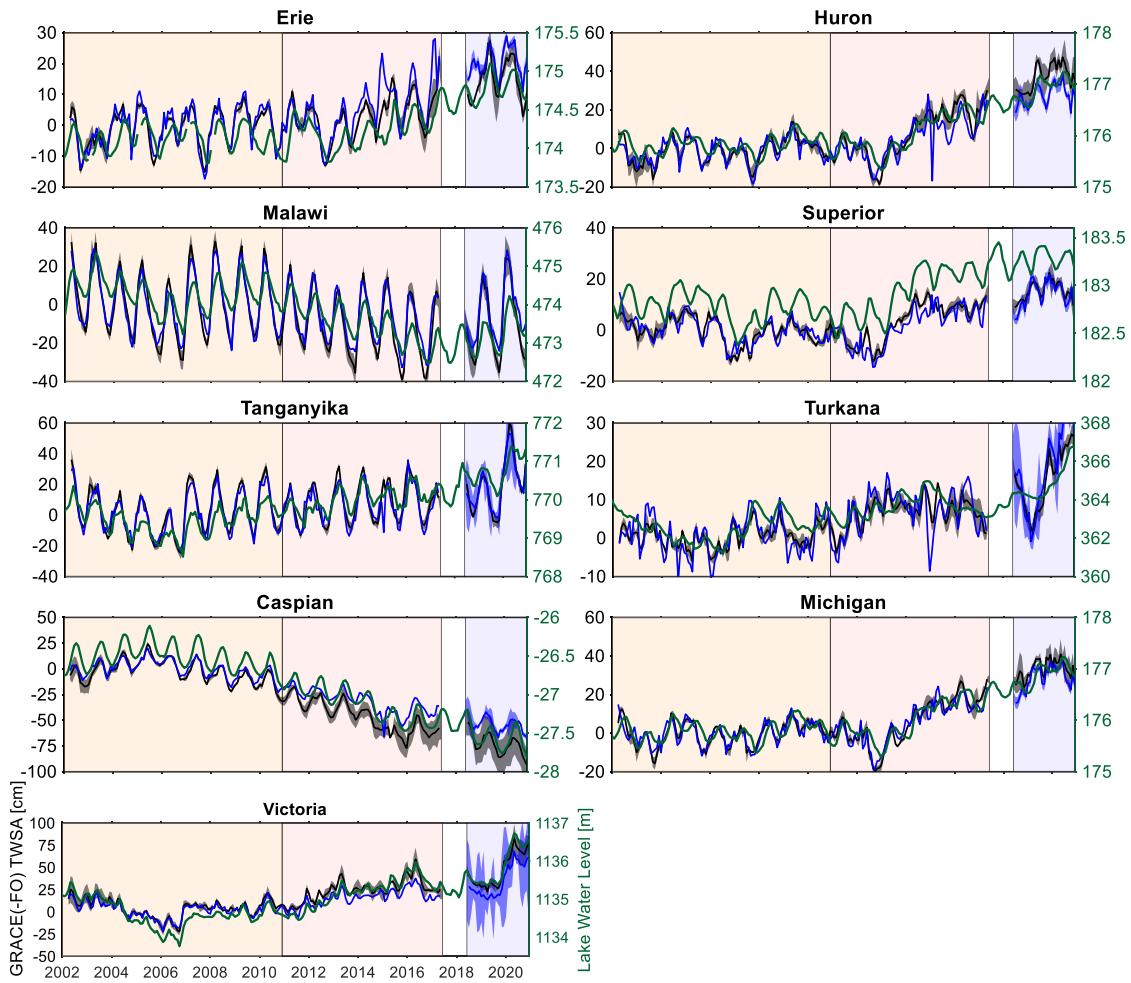| Lake Name	Surface Area [km^2]	Type
Caspian Sea	371 000	Saline
Superior	82 100	Freshwater
Victoria	68 870	Freshwater
Huron	59 600	Freshwater
Michigan	58 000	Freshwater
Tanganyika	32 600	Freshwater
Malawi	29 500	Freshwater
Erie	25 700	Freshwater
Turkana	6 405	Saline

Figure 2.23: Time series of TWSA from GRACE and GRACE-FO from the ensemble mean of the mascons solutions (the black line) and the ensemble mean of the processed SH coefficients from various solutions (the blue line) over selected lakes. The standard deviation of the ensemble means are shown as an envelope. The time series of the water level from satellite altimetry are shown in green. The time period is divided into three periods namely from April 2002 to December 2010, from Januaray 2011 to June 2017, and from June 2018 to the end of 2021.

Table 2.6: Summary of the GRACE(-FO) performance (correlation coefficient) over four periods from two approach, ensemble mean of all possible scenarios of the SHs approach and the ensemble weighted mean of the mascons products.

Lake Name	GRACE(-FO) (2002–2020)		GRACE (2002–2010)		GRACE (2011–2017)		GRACE-FO (2018–2020)	
	mascons	SHs	mascons	SHs	mascons	SHs	mascons	SHs
Erie	0.81	0.79	0.55	0.50	0.72	0.59	0.66	0.54
Huron	0.94	0.86	0.58	0.35	0.91	0.74	0.79	0.57
Malawi	0.73	0.67	0.81	0.79	0.71	0.66	0.82	0.85
Superior	0.83	0.71	0.59	0.33	0.82	0.69	0.00	-0.09
Tanganyika	0.75	0.76	0.76	0.75	0.70	0.68	0.80	0.77
Turkana	0.86	0.76	0.64	0.52	0.58	0.18	0.83	0.82
Caspian	0.99	0.97	0.91	0.84	0.96	0.89	0.89	0.79
Michigan	0.93	0.89	0.58	0.45	0.90	0.83	0.56	0.48
Victoria	0.97	0.96	0.91	0.86	0.90	0.91	0.98	0.98

2.6. Ensemble of the scenarios

No universal and unique procedure is accepted by the Earth's science community in post-processing GRACE(-FO) data. Therefore, after delivering an analysis of the outcomes from various choices in the post-processing steps in the previous sections, we offer an ensemble of the most common scenarios in this section. To this end, we have chosen centers with both GRACE and GRACE-FO solution to be consistent during the whole time period of the satellite missions. These solutions enclose releasing 6 of the official centers, namely CSR, GFZ, and JPL. Besides, we have campuses the latest version of the solutions from ITSG, COST-G, and CNES considering their superior performance which has been discussed in subsection 2.4.2. The steps in blue and green have been considered for all scenarios (see Figure 2.4).

To filter the noisy field of the GRACE(-FO), we have incorporated four different approaches, including a Gaussian filter with three different radii together with the de-striping methods described in subsection 2.4.7. Moreover, we have encompassed two anisotropic filters, namely Hann and Fan filters, each with three radius choices. The DDK2, DDK3, and DDK4 are included as widely used filters among the spatial filters, and finally, the recent filter SSAS together with DDK6 proposed by Yi and Sneeuw (2022). It should be noted that since the noise level in the CNES solutions has already been decreased via regularization, we have only considered the Gaussian, Hann, Fann with radius 300 km and DDK4 among all the possible filter choices. To correct the leakage, we have included the data-driven approach proposed by Vishwakarma et al., 2017. All four GIA models have been considered in the processing resulting in different scenarios to account for the GIA. GRACE and GRACE-FO solutions include 11-month gap between the two missions. Moreover, in total 35 months gap exists in the GRACE Level 2 solutions. To fill these gaps, we have employed the gap-filling technique proposed by Yi and Sneeuw (2021) that benefits from the SSA. The above-mentioned post-processing would end up in about 600 level 3 products from each of the level-2 solutions. Let's call them SHs products. The ensemble mean of the outcomes is calculated together with the standard deviation representing the uncertainty.
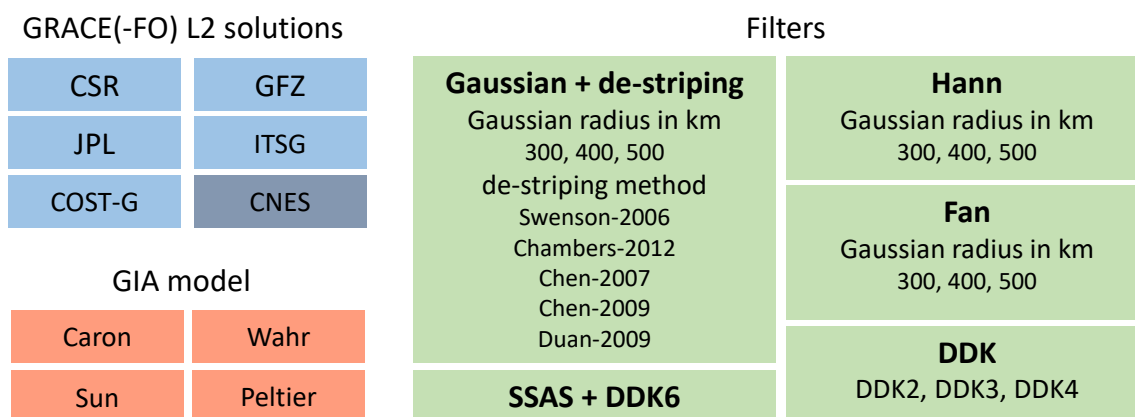
Figure 2.24: Schematic of the different choices for different processing steps of the GRACE(-FO) solutions.

The average coefficient of variation (CV) has been calculated for various scenarios across major river basins. Figure 2.25 presents the results, displaying the boxplot for each category: climate, latitude zone, basin area, and signal strength (RMS of the TWS signal). The findings demonstrate substantial differences in the coefficient of variation (CV) of TWSA across all categories.

One interesting finding is that the arid to hyper-arid climate category had the highest mean CV

value (0.39), indicating a high level of sensitivity to the processing choices. This suggests that the accuracy of TWSA estimates in these regions may be particularly affected by the processing choices, which could have important implications for water resource management in these areas.

In terms of latitude zone, the results showed that the polar zone had the highest mean CV value (1.11), which is consistent with the higher variability in temperature and precipitation patterns in these regions. This suggests that accurate TWSA estimates in polar regions may be more challenging than in other zones, and that the processing choices can have a significant impact on the accuracy of the estimates.

Another interesting finding is that the mean CV values were relatively consistent across different area categories, ranging from 0.25 to 0.39. This suggests that the processing choices have a similar level of impact on TWSA estimates across different basin sizes. However, it is important to note that the choice of basin size can still affect the TWSA estimates themselves, even if the processing choices have a similar level of impact across different basin sizes.

Finally, the results showed that the mean CV values generally increased with increasing signal strength, which indicates that the processing choices have a greater impact on TWSA estimates when the signal is stronger. This finding highlights the importance of carefully considering the processing choices for GRACE Level 2 data when analyzing TWSA time series, particularly in regions with a strong signal.

Overall, these results suggest that the processing choices for GRACE Level 2 data can have a significant impact on the accuracy of TWSA estimates, particularly in regions with certain climate characteristics, latitudes, and signal strengths. By carefully considering these factors, researchers can optimize their processing choices and improve the accuracy of TWSA estimates, which has important implications for water resource management and hydrological modeling.
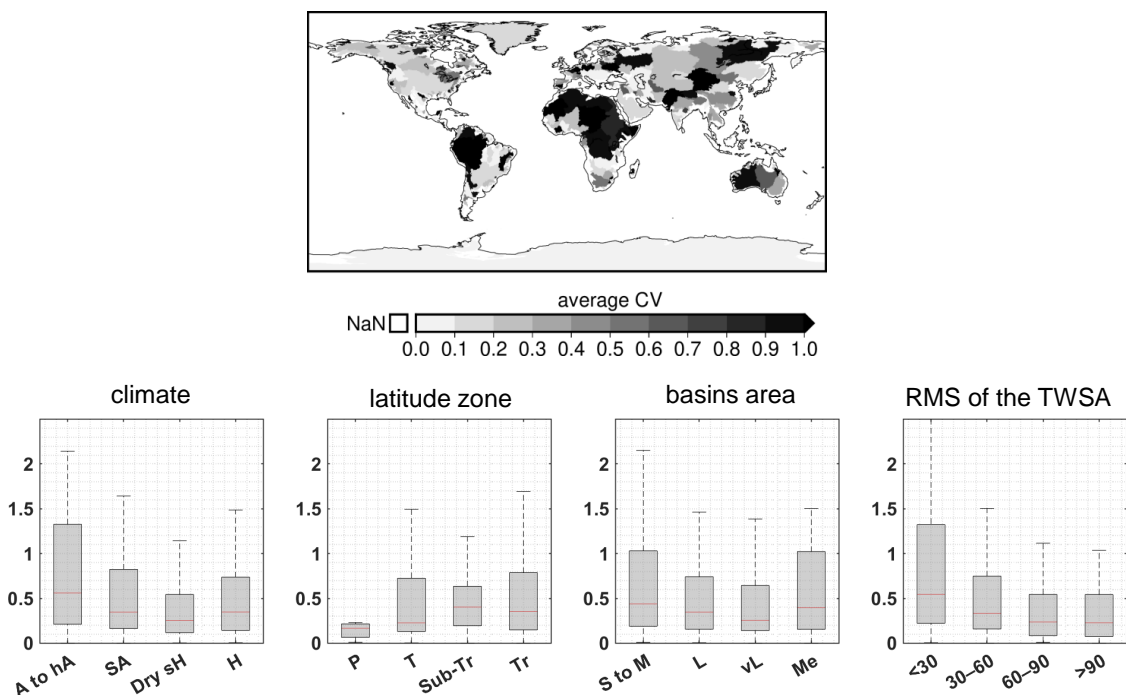
Figure 2.25: Top:Basin-wise average CV from various scenarios. Bottom: Boxplot of the average CV from 600 scenarios over major river basins categorized in terms of climate, latitude zone, basin area, and RMS of the GRACE TWSA signal.

Figure 2.26 illustrates the time series of GRACE(-FO)-TWSA using the ensemble of all the scenarios compared with three mascons solutions, namely CSR mascon v2, JPL mascon v2, and GSFC mascon v2 over selected basins. The ensemble approach (GRACE-Merged) is aligned with all three mascons solutions, while the mascons approach has a different procedure for processing the data.

Figure 2.26: Time series of the ensemble products together with the three mascons products namely CSR mascon v2, JPL mascon v2, and GSFC mascon v2 over selected basins.

Bibliography

Agnew, D. (2010). 6 Earth Tides. *Treatise on Geophysics, Volume 3: Geodesy*, 163.

Ahmed, M., Sultan, M., Elbayoumi, T., & Tissot, P. (2019). Forecasting GRACE data over the African watersheds using artificial neural networks. *Remote Sensing, 11*(15), 1769. https://doi.org/10.3390/rs11151769

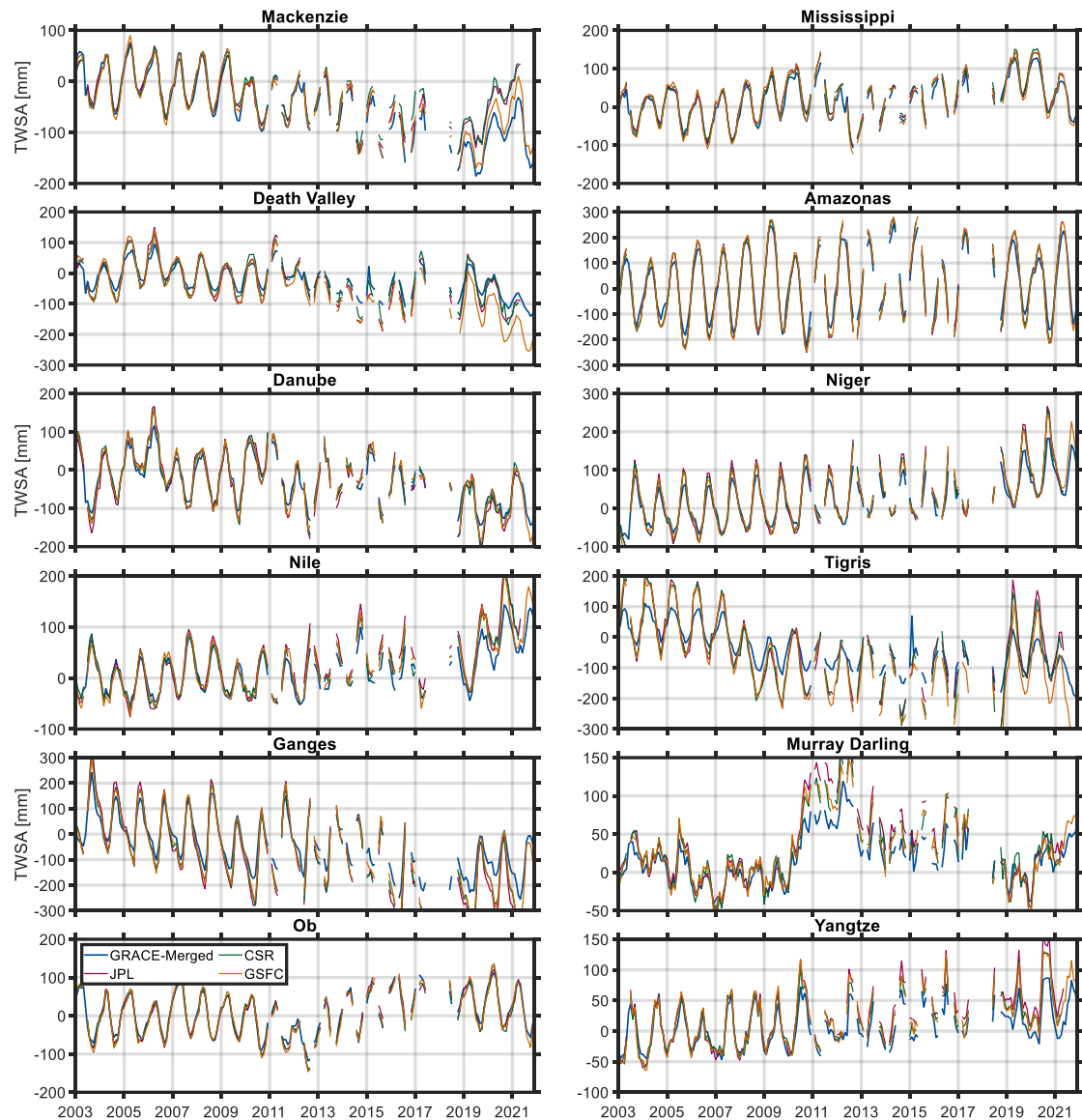Boergens, E., Dobslaw, H., Dill, R., Thomas, M., Dahle, C., Murböck, M., & Flechtner, F. (2020). Modelling spatial covariances for terrestrial water storage variations verified with synthetic GRACE-FO data. *GEM-International Journal on Geomathematics, 11*(1), 1–25. https://doi.org/10.1007/s13137-020-00160-0

Caron, L., Ivins, E., Larour, E., Adhikari, S., Nilsson, J., & Blewitt, G. (2018). GIA model statistics for GRACE hydrology, cryosphere, and ocean science. *Geophysical Research Letters, 45*(5), 2203–2212. https://doi.org/10.1002/2017GL076644

Chambers, D. P. (2006). Evaluation of new GRACE time-variable gravity data over the ocean. *Geophysical Research Letters, 33*(17). https://doi.org/10.1029/2006GL027296

Chen, J. L., Wilson, C. R., Tapley, B. D., & Grand, S. (2007). GRACE detects coseismic and postseismic deformation from the Sumatra-Andaman earthquake. *Geophysical Research Letters, 34*(13). https://doi.org/10.1029/2007GL030356

Chen, Q., Shen, Y., Chen, W., Francis, O., Zhang, X., Chen, Q., Li, W., & Chen, T. (2019). An optimized short-arc approach: Methodology and application to develop refined time series of Tongji-Grace2018 GRACE monthly solutions. *Journal of Geophysical Research: Solid Earth, 124*(6), 6010–6038. https://doi.org/10.1029/2018JB016596

Chen, Q., Shen, Y., Francis, O., Chen, W., Zhang, X., & Hsu, H. (2018). Tongji-Grace02s and Tongji-Grace02k: high-precision static GRACE-only global Earth's gravity field models derived by refined data processing strategies. *Journal of Geophysical Research: Solid Earth, 123*(7), 6111–6137. https://doi.org/10.1029/2018JB015641

Chen, Wilson, C., Li, J., & Zhang, Z. (2015). Reducing leakage error in GRACE-observed longterm ice mass change: a case study in West Antarctica. *Journal of Geodesy, 89*(9), 925–940. https://doi.org/10.1007/s00190-015-0824-2

Cheng, M., Tapley, B. D., & Ries, J. C. (2013). Deceleration in the Earth's oblateness. *Journal of Geophysical Research: Solid Earth, 118*(2), 740–747. https://doi.org/10.1002/jgrb.50058,2013

Chu, J., Su, X., Jiang, T., Qi, J., Zhang, G., & Wu, H. (2023). Filling the gap between GRACE and GRACE-FO data using a model integrating variational mode decomposition and long short-term memory: a case study of Northwest China. *Environmental Earth Sciences, 82*(1), 38. https://doi.org/10.1007/s12665-022-10716-y

Dahle, C., Flechtner, F., Gruber, C., König, D., König, R., Michalak, G., & Neumayer, K.-H. (2013). GFZ GRACE Level-2 processing standards document for level-2 product release 0005: revised edition, January 2013. https://doi.org/10.2312/GFZ.b103-1202-25

Desai, S. D. (2002). Observing the pole tide with satellite altimetry. *Journal of Geophysical Research: Oceans, 107*(C11), 7–1. https://doi.org/10.1029/2001JC001224

Devaraju, B. (2015). *Understanding filtering on the sphere: Experiences from filtering GRACE data*. https://doi.org/10.18419/opus-3985

Devaraju, B., & Sneeuw, N. (2017). The polar form of the spherical harmonic spectrum: implications for filtering GRACE data. *Journal of Geodesy, 91*(12), 1475–1489. https://doi.org/10.1007/s00190-017-1037-7

Dobslaw, H., Bergmann-Wolf, I., Dill, R., Poropat, L., & Flechtner, F. (2017). Product description document for AOD1B release 06. *GFZ German Research Centre for Geosciences Department, 1.*

Duan, X., Guo, J., Shum, C., & Van Der Wal, W. (2009). On the postprocessing removal of correlated errors in GRACE temporal gravity field solutions. *Journal of Geodesy, 83*(11), 1095–1106. https://doi.org/10.1007/s00190-009-0327-0

Flechtner, F., Dobslaw, H., & Fagiolini, E. (2014). AOD1B product description document for product release 05 (Rev. 4.4, December 14, 2015). *Technical Note.*

Frappart, F., Ramillien, G., Leblanc, M., Tweed, S. O., Bonnet, M.-P., & Maisongrande, P. (2011). An independent component analysis filtering approach for estimating continental hydrology in the GRACE gravity data. *Remote Sensing of Environment, 115*(1), 187–204. https://doi.org/10.1016/j.rse.2010.08.017

Geruo, A., Wahr, J., & Zhong, S. (2013). Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada. *Geophysical Journal International, 192*(2), 557–572. https://doi.org/10.1093/gji/ggs030

Han, S.-C., Shum, C., Jekeli, C., Kuo, C.-Y., Wilson, C., & Seo, K.-W. (2005). Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. *Geophysical Journal International, 163*(1), 18–25. https://doi.org/10.1111/j.1365-246X.2005.02756.x

Heiskanen, W. A., & Moritz, H. (1967). Physical geodesy(Book on physical geodesy covering potential theory, gravity fields, gravimetric and astrogeodetic methods, statistical analysis, etc).

Humphrey, V., & Gudmundsson, L. (2019). GRACE-REC: a reconstruction of climate-driven water storage changes over the last century. *Earth System Science Data, 11*(3), 1153–1170. https://doi.org/10.5194/essd-11-1153-2019

Humphrey, V., Gudmundsson, L., & Seneviratne, S. I. (2017). A global reconstruction of climate-driven subdecadal water storage variability. *Geophysical Research Letters, 44*(5), 2300–2309. https://doi.org/10.1002/2017GL072564

Jäggi, A., Hugentobler, U., Bock, H., & Beutler, G. (2007). Precise orbit determination for GRACE using undifferenced or doubly differenced GPS data. *Advances in Space Research, 39*(10), 1612–1619. https://doi.org/10.1016/j.asr.2007.03.012

Jekeli, C. (1981). *Alternative methods to smooth the Earth's gravity field* (tech. rep.). https://ntrs.nasa.gov/citations/19820014947

Kang, Z., Tapley, B., Bettadpur, S., Ries, J., & Nagel, P. (2006). Precise orbit determination for GRACE using accelerometer data. *Advances in Space Research, 38*(9), 2131–2136. https://doi.org/10.1016/j.asr.2006.02.021

Khaki, M., Forootan, E., Kuhn, M., Awange, J., Longuevergne, L., & Wada, Y. (2018). Efficient basin scale filtering of GRACE satellite products. *Remote sensing of environment, 204,* 76–93. https://doi.org/10.1016/j.rse.2017.10.040

Kirschner, M., Montenbruck, O., & Bettadpur, S. (2001). Flight dynamics aspects of the GRACE formation flying. *2nd International Workshop on Satellite Constellations and Formation Flying,* 19–20.

Klees, R., Revtova, E., Gunter, B., Ditmar, P., Oudman, E., Winsemius, H., & Savenije, H. (2008). The design of an optimal filter for monthly GRACE gravity models. *Geophysical Journal International, 175*(2), 417–432. https://doi.org/10.1111/j.1365-246X.2008.03922.x

Klees, R., Zapreeva, E., Winsemius, H., & Savenije, H. (2007). The bias in GRACE estimates of continental water storage variations. *Hydrology and Earth System Sciences, 11*(4), 1227–1241. https://doi.org/10.5194/hess-11-1227-2007

Kumar, K. S., Sridhar, V., Varaprasad, B. J. S., & Chinnapa Reddy, K. (2022). Bridging the Data Gap between the GRACE Missions and Assessment of Groundwater Storage Variations for Telangana State, India. *Water, 14*(23), 3852. https://doi.org/10.3390/w14233852

Kusche, J., Schmidt, R., Petrovic, S., & Rietbroek, R. (2009). Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *Journal of geodesy, 83*(10), 903–913. https://doi.org/10.1007/s00190-009-0308-3

Kvas, A., Behzadpour, S., Ellmer, M., Klinger, B., Strasser, S., Zehentner, N., & Mayer-Gürr, T. (2019). ITSG-Grace2018: Overview and evaluation of a new GRACE-only gravity field time series. *Journal of Geophysical Research: Solid Earth, 124*(8), 9332–9344. https://doi.org/10.1029/2019JB017415

Landerer, F. (2020a). CSR TELLUS GRACE Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly Release 6.0 version 03 [Last accessed: 1 February 2021]. https://doi.org/10.5067/TELND-3AC63

Landerer, F. (2020b). CSR TELLUS GRACE-FO Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly Release 6.0 version 03 [Last accessed: 1 February 2021]. https://doi.org/10.5067/GFLND-3AC63

Landerer, F. (2020c). GFZ TELLUS GRACE Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly Release 6.0 version 03 [Last accessed: 1 February 2021]. https://doi.org/10.5067/TELND-3AG63

Landerer, F. (2020d). JPL TELLUS GRACE Level-3 Monthly Land Water Equivalent Thickness Surface Mass Anomaly Release 6.0 version 03 [Last accessed: 1 February 2021]. https://doi.org/10.5067/TELND-3AJ63

Landerer, F. (2020e). JPL TELLUS GRACE-FO Level-3 Monthly Land Water-Equivalent-Thickness Surface Mass Anomaly Release 6.0 version 03 [Last accessed: 1 February 2021]. https://doi.org/10.5067/GFLND-3AJ63

Landerer, F., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., Bettadpur, S. V., Byun, S. H., Dahle, C., Dobslaw, H., et al. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters, 47*(12), e2020GL088306. https://doi.org/10.1029/2020GL088306

Landerer, F., & Swenson, S. (2012). Accuracy of scaled GRACE terrestrial water storage estimates. *Water resources research, 48*(4). https://doi.org/10.1029/2011WR011453

Lenczuk, A., Weigelt, M., Kosek, W., & Mikocki, J. (2022). Autoregressive Reconstruction of Total Water Storage within GRACE and GRACE Follow-On Gap Period. *Energies, 15*(13), 4827. https://doi.org/10.3390/en15134827

Li, F., Kusche, J., Chao, N., Wang, Z., & Löcher, A. (2021). Long-Term (1979-Present) total water storage anomalies over the global land derived by reconstructing GRACE data. *Geophysical Research Letters, 48*(8), e2021GL093492. https://doi.org/10.1029/2021GL093492

Li, J., Chen, J., Li, Z., Wang, S.-Y., & Hu, X. (2017). Ellipsoidal correction in GRACE surface mass change estimation. *Journal of Geophysical Research: Solid Earth, 122*(11), 9437–9460. https://doi.org/10.1002/2017JB014033

Liu, W. (2019). Understanding ocean tide aliasing in satellite gravimetry. https://doi.org/10.18419/opus-10485

Liu, W., & Sneeuw, N. (2021). Aliasing of ocean tides in satellite gravimetry: a two-step mechanism. *Journal of Geodesy, 95*(12), 1–12. https://doi.org/10.1007/s00190-021-01586-6

Löcher, A., & Kusche, J. (2021). A hybrid approach for recovering high-resolution temporal gravity fields from satellite laser ranging. *Journal of Geodesy, 95*(1), 6. https://doi.org/10.1007/s00190-020-01460-x

2

Long, D., Chen, X., Scanlon, B. R., Wada, Y., Hong, Y., Singh, V. P., Chen, Y., Wang, C., Han, Z., & Yang, W. (2016). Have GRACE satellites overestimated groundwater depletion in the Northwest India Aquifer? *Scientific reports*, *6*(1), 1–11. https://doi.org/10.1038/srep24398

Long, D., Longuevergne, L., & Scanlon, B. R. (2015). Global analysis of approaches for deriving total water storage changes from GRACE satellites. *Water Resources Research*, *51*(4), 2574–2594. https://doi.org/10.1002/2014WR016853

Longuevergne, L., Scanlon, B. R., & Wilson, C. R. (2010). GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research*, *46*(11). https://doi.org/10.1029/2009WR008564

Loomis, B., Luthcke, S., & Sabaka, T. (2019). Regularization and error characterization of GRACE mascons. *Journal of geodesy*, *93*(9), 1381–1398. https://doi.org/10.1007/s00190-019-01252-y

Meyer, U., Jäggi, A., Jean, Y., & Beutler, G. (2016). AIUB-RL02: An improved time-series of monthly gravity fields from GRACE data. *Geophysical Journal International*, *205*(2), 1196–1207. https://doi.org/10.1093/gji/ggw081

Nelson, J. (2017). Prolific Earth Gravity Satellites End Science Mission.

Peltier, W. (2009). Closure of the budget of global sea level rise over the GRACE era: the importance and magnitudes of the required corrections for global glacial isostatic adjustment. *Quaternary Science Reviews*, *28*(17-18), 1658–1674. https://doi.org/10.1016/j.quascirev.2009.04.004

Petit, G., & Luzum, B. (2010). *IERS conventions (2010)* (tech. rep.). Bureau International des Poids et mesures sevres (france).

Prevost, P., Chanard, K., Fleitout, L., Calais, E., Walwer, D., van Dam, T., & Ghil, M. (2019). Data-adaptive spatio-temporal filtering of GRACE data. *Geophysical Journal International*, *219*(3), 2034–2055. https://doi.org/10.1093/gji/ggz409

Ray, R., & Luthcke, S. (2006). Tide model errors and GRACE gravimetry: towards a more realistic assessment. *Geophysical Journal International*, *167*(3), 1055–1059. https://doi.org/10.1111/j.1365-246X.2006.03229.x

Richter, H. M. P., Lück, C., Klos, A., Sideris, M. G., Rangelova, E., & Kusche, J. (2021). Reconstructing GRACE-type time-variable gravity from the Swarm satellites. *Scientific reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-020-80752-w

Rowlands, D., Luthcke, S., McCarthy, J., Klosko, S., Chinn, D., Lemoine, F., Boy, J.-P., & Sabaka, T. (2010). Global mass flux solutions from GRACE: A comparison of parameter estimation strategies—Mass concentrations versus Stokes coefficients. *Journal of Geophysical Research: Solid Earth*, *115*(B1). https://doi.org/10.1029/2009JB006546

Sasgen, I., Martinec, Z., & Fleming, K. (2006). Wiener optimal filtering of GRACE data. *Studia Geophysica et Geodaetica*, *50*(4), 499–508. https://doi.org/10.1007/s11200-006-0031-y

Save, H. (2020). CSR GRACE and GRACE-FO RL06 Mascon Solutions v02. 2020. https://doi.org/10.15781/cgq9-nh24

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547–7569. https://doi.org/10.1002/2016JB013007

Schrama, E., & Visser, P. (2007). Accuracy assessment of the monthly GRACE geoids based upon a simulation. *Journal of Geodesy*, *81*(1), 67–80. https://doi.org/0.1007/s00190-006-0085-1

Schrama, E. J., Wouters, B., & Lavallée, D. A. (2007). Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations. *Journal of Geophysical Research: Solid Earth*, *112*(B8). https://doi.org/10.1029/2006JB004882

Seo, K., Wilson, C., Han, S., & Waliser, D. (2008). Gravity Recovery and Climate Experiment (GRACE) alias error from ocean tides. *Journal of Geophysical Research: Solid Earth, 113*(B3). https://doi.org/10.1029/2006JB004747

Sheard, B., Heinzel, G., Danzmann, K., Shaddock, D., Klipstein, W., & Folkner, W. (2012). Inter-satellite laser ranging instrument for the GRACE follow-on mission. *Journal of Geodesy, 86*(12), 1083–1095. https://doi.org/10.1007/s00190-012-0566-3

Sun, A., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? *Water Resources Research, 55*(2), 1179–1195. https://doi.org/10.1029/2018WR023333

Sun, A. Y., Scanlon, B. R., Save, H., & Rateb, A. (2021). Reconstruction of GRACE total water storage through automated machine learning. *Water Resources Research, 57*(2), e2020WR028666. https://doi.org/10.1029/2020WR028666

Sun, Y., Riva, R., & Ditmar, P. (2016). 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*(11), 8352–8370. https://doi.org/10.1002/2016JB013073

Sun, Y., & Riva, R. E. (2020). A global semi-empirical glacial isostatic adjustment (GIA) model based on Gravity Recovery and Climate Experiment (GRACE) data. *Earth System Dynamics, 11*(1), 129–137. https://doi.org/10.5194/esd-11-129-2020

Swenson, S., Chambers, D., & Wahr, J. (2008). Estimating geocenter variations from a combination of GRACE and ocean model output. *Journal of Geophysical Research: Solid Earth, 113*(B8). https://doi.org/10.1029/2007JB005338

Swenson, S., & Wahr, J. (2006). Post-processing removal of correlated errors in GRACE data. *Geophysical research letters, 33*(8). https://doi.org/10.1029/2005GL025285

Szabó, V., & Marjańska, D. (2020). Accuracy analysis of gravity field changes from GRACE RL06 and RL05 data compared to in situ gravimetric measurements in the context of choosing optimal filtering type. *Artificial Satellites, 55*(3), 100–117. https://doi.org/10.2478/arsa-2020-0008

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004a). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters, 31*(9). https://doi.org/10.1029/2004GL019920

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004b). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters, 31*(9). https://doi.org/10.1029/2004GL019920

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. (2019). Contributions of GRACE to understanding climate change. *Nature climate change, 9*(5), 358–369. https://doi.org/10.1038/s41558-019-0456-2

Tourian, M. J. (2013). *Application of spaceborne geodetic sensors for hydrology.* https://doi.org/10.18419/opus-3929

Vishwakarma, B. D., Devaraju, B., & Sneeuw, N. (2016). Minimizing the effects of filtering on catchment scale GRACE solutions. *Water Resources Research, 52*(8), 5868–5890. https://doi.org/10.1002/2016WR018960

Vishwakarma, B. D., Horwath, M., Devaraju, B., Groh, A., & Sneeuw, N. (2017). A data-driven approach for repairing the hydrological catchment signal damage due to filtering of GRACE products. *Water Resources Research, 53*(11), 9824–9844. https://doi.org/10.1002/2017WR021150

Visser, P. (2005). Low-low satellite-to-satellite tracking: a comparison between analytical linear orbit perturbation theory and numerical integration. *Journal of Geodesy, 79*(1), 160–166. https://doi.org/10.1007/s00190-005-0455-0

Wahr, J., Molenaar, M., & Bryan, F. (1998). 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*(B12), 30205–30229. https://doi.org/10.1029/98JB02844

Wahr, J., Swenson, S., & Velicogna, I. (2006). Accuracy of GRACE mass estimates. *Geophysical Research Letters, 33*(6). https://doi.org/10.1029/2005GL025305

Wang, L., Davis, J. L., Hill, E. M., & Tamisiea, M. E. (2016). Stochastic filtering for determining gravity variations for decade-long time series of GRACE gravity. *Journal of Geophysical Research: Solid Earth, 121*(4), 2915–2931. https://doi.org/10.1002/2015JB012650

Watkins, M. M., Wiese, D. N., Yuan, D.-N., Boening, C., & Landerer, F. W. (2015). Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. *Journal of Geophysical Research: Solid Earth, 120*(4), 2648–2671. https://doi.org/10.1002/2014JB011547

Watkins, M. M., & Yuan, D.-N. (2012). JPL Level-2 processing standards document for Level-2 product release 05. *Report No. GRACE*, 327–744.

Wiese, D., Landerer, F. W., & Watkins, M. M. (2016). Quantifying and reducing leakage errors in the JPL RL05M GRACE mascon solution. *Water Resources Research, 52*(9), 7490–7502. https://doi.org/10.1002/2016WR019344

Wiese, D., Yuan, D., Boening, C., Landerer, F., & Watkins, M. (2018). JPL GRACE mascon ocean, ice, and hydrology equivalent water height release 06 coastal resolution improvement (CRI) filtered version 1.0 [Last accessed: 1 February 2021]. *DAAC: Pasadena, CA, USA*. https://doi.org/10.5067/TEMSC-3MJC6

Wolff, M. (1969). Direct measurements of the Earth's gravitational potential using a satellite pair. *Journal of Geophysical Research, 74*(22), 5295–5300. https://doi.org/10.1029/JB074i022p05295

Wouters, B., & Schrama, E. J. (2007). Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering of spherical harmonics. *Geophysical Research Letters, 34*(23). https://doi.org/10.1029/2007GL032098

Yi, S., & Sneeuw, N. (2021). Filling the data gaps within GRACE missions using singular spectrum analysis. *Journal of Geophysical Research: Solid Earth, 126*(5), e2020JB021227. https://doi.org/10.1029/2020JB021227

Yi, S., & Sneeuw, N. (2022). A novel spatial filter to reduce north–south striping noise in GRACE spherical harmonic coefficients. *Journal of Geodesy, 96*(4), 1–17. https://doi.org/10.1007/s00190-022-01614-z

Zhang, X., Li, J., Dong, Q., Wang, Z., Zhang, H., & Liu, X. (2022). Bridging the gap between GRACE and GRACE-FO using a hydrological model. *Science of The Total Environment, 822*, 153659. https://doi.org/10.1016/j.scitotenv.2022.153659

Zhang, Z.-Z., Chao, B., Lu, Y., & Hsu, H.-T. (2009). An effective filtering for GRACE time-variable gravity: Fan filter. *Geophysical Research Letters, 36*(17). https://doi.org/10.1029/2009GL039459

3

Hindcasting GRACE(-FO)

3.1. Introduction

The Terrestrial Water Storage (TWS) dynamics represent the available freshwater on land and play a critical role in understanding the Earth's coupled climate system (e.g., Frappart & Ramillien, 2018; McCabe et al., 2017; Richey et al., 2015; Tapley et al., 2019). The GRACE and GRACE-FO missions have measured the TWS change with unprecedented accuracy for more than 20 years (Tapley et al., 2004). Using GRACE(-FO) we can investigate the footprint of the natural ecosystems and human interventions on the spatio-temporal variability of the TWS at the regional to global scales. While these measurements are valuable for identifying short-term variations in water storage, such as seasonal and interannual variations, they are insufficient for detecting long-term trends in water storage, which may be important for various applications including testing climate model simulations (Goosse, 2015; Jensen et al., 2019; Trenberth & Trenberth, 1992, e.g.,), understanding past extreme hydro-climate events (drought or flood) (Zhang et al., 2016; Zhao et al., 2018), and constructing the sea level budget (Horwath et al., 2022; Wouters, van de Wal, et al., 2018). Understanding and monitoring these changes in water storage are particularly crucial in regions where water availability is a major concern, such as arid and semi-arid regions, where long-term changes in water storage can have significant impacts on water resource management and agricultural productivity (Van Beek et al., 2011).

The climatology of TWS refers to the long-term monthly average of TWS, which characterizes the baseline variation across a region. To appraise the impact of observation records on the TWS climatology, we computed the extended-term mean of TWS anomalies (TWSA) for three distinct periods. The first period spans from 2003 to 2016, during the GRACE era, utilizing GRACE observations. The second period ranges from 1980 to 2002, prior to the availability of GRACE data, employing an ensemble mean of models, as described in section 3.3. The third period consists of a merged dataset, encompassing 1980 to 2016, which includes the GRACE period and the pre-GRACE period. Figure 3.1 presents the outcomes for twelve selected basins and their corresponding level of uncertainty. We calculated the uncertainty for each month by determining the standard deviation of TWSA values within the given time period. Notably, we observed wider envelopes for 1980–2002 since we employed a merged model version. While some basins, such as the Danube, show no change in climatology between the GRACE era and a longer record back to 1980, selecting only the GRACE record would significantly underestimate TWS in some basins (e.g., Murray) and overestimate it in others (e.g., Amazon). This underscores the importance of considering an adequate length of record for estimating the TWS climatology of a region.

To obtain a global perspective, we assessed the differences between the climatology calculated using a longer-term record (1980–2012) and the GRACE data record (2003–2012) for 403 major river basins worldwide, except for Greenland and Antarctica. Although the result from the longer-term record is not evaluated yet, we assume the longer period includes more of the climate variation and hence can be considered as the true baseline. Figure 3.2 displays the climatology difference for each month. Generally, May, June, July, and December exhibit the highest differences (positive or negative) in climatology, while April and October show the lowest variations. The maximum underestimation is observed in December and January, whereas the maximum overestimation occurs in June and July when compared to the long-term reference record. The number of basins with underestimated (overestimated) climatology ranges from approximately 130 (190) basins in January (July) to about 210 (265) basins in July (January). The Tana, Bravo, and Zagros basins exhibit the maximum overestimation, while the Atrato, Svarta, and South Iran basins exhibit the maximum underestimation. These findings highlight the necessity of employing a record of observations spanning over 30 years to establish a reliable climatology.
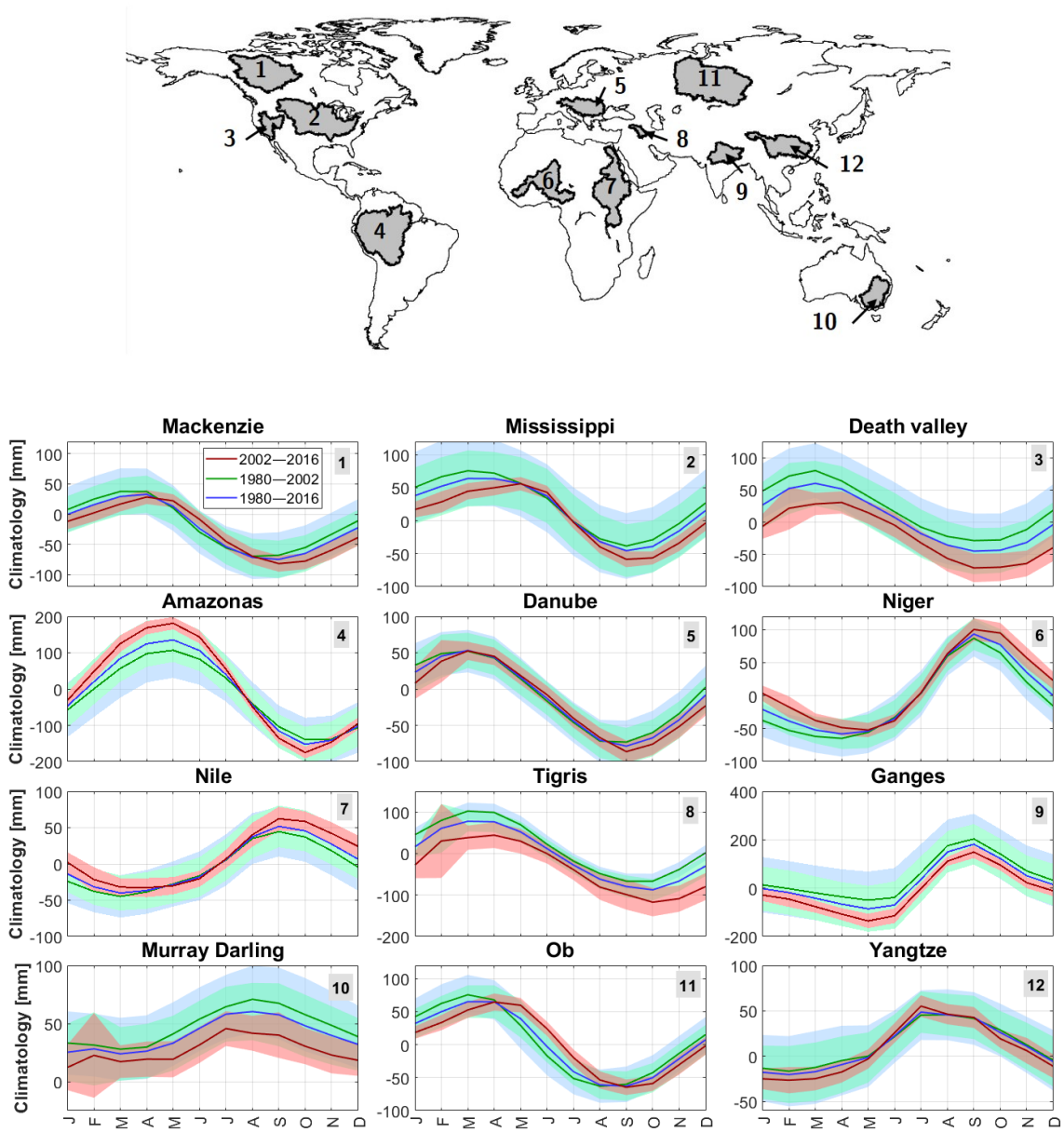
Figure 3.1: Long-term monthly mean (climatology) of TWSA over selected basins within three periods namely 1980–2002 (pre-GRACE period), 2002–2016 (GRACE period), and 1980–2016. The envelopes represent the climatology uncertainty. The global distribution of the selected basins is shows in the top panel.
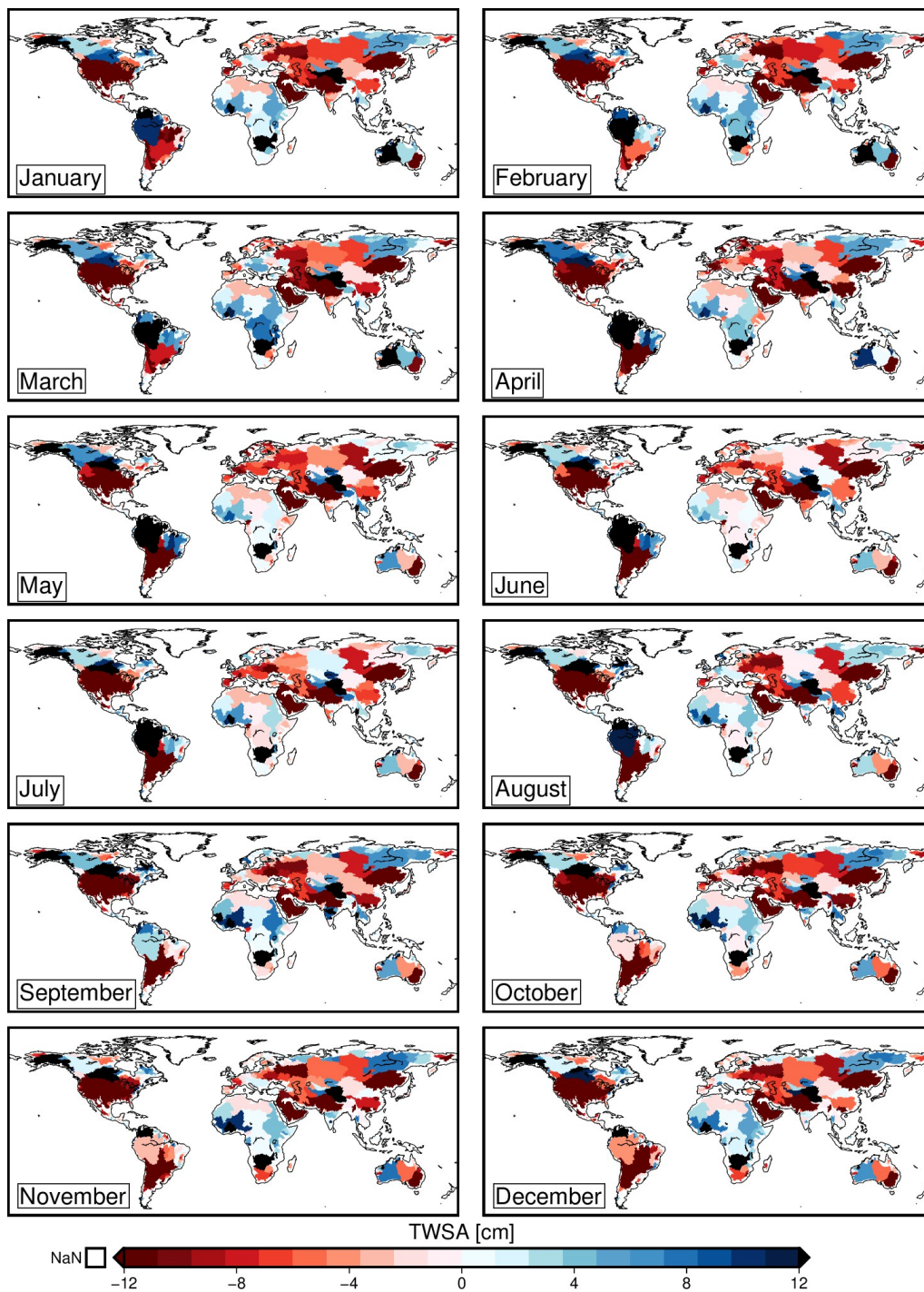
Figure 3.2: Distribution of the difference of the TWSA climatology within the restricted GRACE time period (2002–2012) and 1980–2012 for the 403 major river basins. The long-term TWSA dataset is derived from the ensemble of the models as described in subsection section 3.3.

3.2. Brief review on the hindcasting methods

There are several approaches to estimate TWSA for the pre GRACE period. The first option is to use in-situ observations of groundwater or surface water levels. Such an approach is not an effective one due to several reasons. Firstly, the spatial coverage of in-situ measurements is limited and unevenly distributed across the globe, making it difficult to capture the large-scale changes in water storage that GRACE is capable of detecting. Moreover, in-situ measurements are often limited to specific depths or locations, which may not be representative of the entire basin or region, leading to biased estimates. Secondly, in-situ measurements are subject to errors, uncertainties, and biases due to various factors such as instrument drift, measurement noise, data gaps, and inconsistencies in sampling frequency. These errors can accumulate over time and affect the accuracy and reliability of the reconstructed data, particularly for long-term analysis. Thirdly, in-situ measurements are often influenced by local processes such as groundwater pumping, irrigation, and reservoir management, which can obscure the signal of large-scale hydrological processes that GRACE aims to capture. These local effects can vary spatially and temporally and can be challenging to separate from the signal of interest. Overall, while in-situ measurements are valuable for validating and calibrating remote sensing data, they are not a substitute for GRACE data for reconstructing global water storage changes over time.

Several space-borne sensors are employed to estimate TWSA, including Swarm satellites (de Teixeira da Encarnação et al., 2019; Lück et al., 2018; Olsen et al., 2013; Richter et al., 2021, e.g.,), Global Navigation Satellite System (GNSS) inversion (Rietbroek et al., 2014), and SLR (Li et al., 2021; Talpe et al., 2017). While space-borne sensors such as Swarm and GNSS inversion offer global coverage, their products have lower spatial resolution and accuracy than GRACE, making them insufficient to fully reconstruct GRACE TWSA. Besides, while the GRACE mission provided a continuous record of TWSA changes over a period of more than 15 years, other spaceborne sensors may have limited temporal coverage due to their shorter operational lifetimes. Moreover, each spaceborne sensor measures TWSA using a different principle and technology, which can lead to different sources of errors and uncertainties. As a result, combining data from different sensors may introduce additional sources of error that are difficult to quantify and correct for. Finally, processing and analyzing data from different spaceborne sensors can be challenging due to differences in data format, resolution, and accuracy. Integrating these data sources requires careful calibration, quality control, and validation, which can be time-consuming and resource-intensive.

Several studies have assessed approaches to reconstruct TWSA prior to the GRACE era (i.e., before 2002). Humphrey et al. (2017) developed a statistical model based on precipitation and temperature to reconstruct climate-driven TWSA. The study aims to reconstruct TWSA changes over the last century, which is a significant period for understanding the variability of water resources on a global scale. The accuracy of the reconstructed data depends on the accuracy of the input data used to drive the model. Additionally, the water balance model used in their study does not take into account changes in groundwater storage, which can be a significant component of TWSA changes in some regions. Similarly, Liu et al. (2022) reconstructed the Climate-driven Water Storage Anomalies (CWSAs) using precipitation and temperature data as input. The model significantly outperformed LSMs and produced similar results to machine learning-based methods. However, limitations were observed in some basins with widespread frozen ground and glaciers or huge lakes. While these studies show promise in reconstructing long-term TWSA trends, they are limited by the accuracy of the input data used to drive the models. Additionally, the water balance models used in these studies do not account for changes in groundwater storage, which can be a significant component of TWSA changes in

high-intensity human-intervention basins. Future studies could incorporate additional data sources, such as remote sensing data or groundwater monitoring networks, to improve the accuracy of reconstructed TWSA.

Many studies have improved the performance of the physically-based hydrological models, especially in region with considerable groundwater extraction, by assimilating GRACE observations (Eicker et al., 2014; Khaki et al., 2017; Kumar et al., 2016; Schumacher et al., 2018). Zaitchik et al. (2008) assimilated GRACE data into a hydrological model to improve estimates of groundwater recharge in the Mississippi River basin. The assimilation of GRACE data improved the accuracy of the model, particularly during the drought period, and provided valuable information on the spatial patterns of groundwater recharge. For example, The Global Land Data Assimilation System (GLDAS) 2.21 DA model is the latest version of the GLDAS product, which was released in 2021. The DA in the name refers to the Data Assimilation component of the model. In this version, GLDAS uses the Community Land Model (CLM) version 4.5 as the land surface model and assimilates various observational data sources into the model using the Ensemble Kalman Filter (EnKF) algorithm. GLDAS 2.21 DA model provides several variables, including soil moisture, temperature, and snow water equivalent, at a spatial resolution of 0.25 ° and at a temporal resolution of three hours. The model has a global coverage and provides data from January 2000 to the present. The GLDAS 2.21 DA model is useful for a range of applications, such as drought monitoring, flood forecasting, and climate modeling. The data is publicly available and can be accessed through the NASA Earthdata website or the GLDAS website.

Some recent studies have employed Machine Learning (ML) techniques to reconstruct GRACE TWSA and in particular to hindcast TWSA back in time. These methods have shown promising results in capturing the spatiotemporal patterns of TWSA and reducing the uncertainties associated with the GRACE data. One approach is to use Convolutional Neural Network (CNN), a type of deep learning algorithm that extracts relevant features from input data using convolutional filters. For example, (Sun et al., 2019), proposed a TWSA-CNN model that utilized a 3D CNN architecture to learn the temporal and spatial features of TWSA. The model achieved good performance in reproducing TWSA with a correlation coefficient of 0.94 between the reconstructed and observed TWSA.

Another machine learning technique that has gained popularity is Long Short-Term Memory (LSTM), a type of recurrent neural network that can model the temporal dependencies of sequential data. (Wang et al., 2021) applied an LSTM-based model to reconstruct TWSA over the period of 2002–2016, achieving a high correlation coefficient of 0.98 between the reconstructed and observed TWSA. Another method that combines machine learning and spatial analysis is the random forest algorithm with a spatially moving window structure. (Jing et al., 2020) used this method to reconstruct TWSA over the period of GRACE and showed good performance in capturing the spatial heterogeneity of TWSA. In addition, some studies have utilized a combination of machine learning methods to improve the accuracy of TWSA reconstruction. For example, (Sun et al., 2020) proposed a TWSA-MLs model that integrated Deep Neural Network (DNN), Multiple Linear Regression (MLR), and Seasonal AutoRegressive Integrated Moving Average with exogenous variables (SARIMAX) methods. The model showed significant improvement over individual methods in capturing the temporal and spatial patterns of TWSA.

3.3. Global TWSA from models

3.3.1. Overview of the global models

The total amount of water on Earth is fixed while it continuously circulates between the oceans, the atmosphere, cryosphere, and the land in the so-called water cycle. We count on global models to predict the effects of human intervention and the environment on water resources (Döll et al., 2016; Scanlon et al., 2018). Models, from a simple box model to a recent sophisticated deep learning model, have been designed to enhance our understanding and acuity of the Earth's water system that occurs as an exchange between the terrestrial biosphere and atmosphere. In general, three different groups of models have been developed, namely Land Surface Models (LSMs), Global Hydrological Models (GHMs), and global atmospheric reanalysis models.

LSMs, as a part of climate models, simulate processes happening at the Earth's surface between the land surface and atmosphere. Such processes include the coupled fluxes of water, energy, and carbon and incorporate the direct and indirect human interventions and ecological dynamics (Fisher & Koven, 2020). Land surface models have expanded through time from their simple biophysical configurations to including various units such as soil carbon cycling, surface hydrological processes, soil moisture dynamics, and phosphorus cycling (Abramowitz et al., 2008; Lawrence et al., 2019; Wiltshire et al., 2020; Yokohata et al., 2019). These expansions have been driven by the needs of various users and the availability of satellite data and projects such as FLUXNET (an international network of Eddy Covariance (EC) sites) (Stöckli et al., 2008; Williams et al., 2009). Deardorff (1978) performed one of the first LSM assessments for a wheat crop in the United Kingdom, utilizing fluxes derived from a few meteorological data over two summer days. Since then, several models have been developed, including on-layer or big-leaf models, also known as Penman-Monteith type models, two-layer models, and multi-layer models (Figure 3.3).
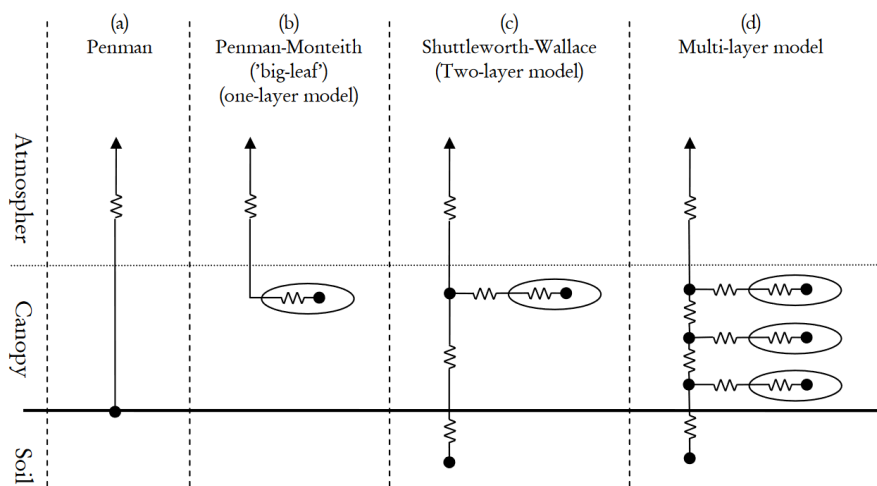
Figure 3.3: Different type of LSMs from (a) Penman to (d) Multi-layer model (credit: Overgaard et al. (2006))

In response to global water scarcity problems, GHMs are developed using empirical water budget approaches rather than a complex physical basis in LSMs (Döll et al., 2016; Scanlon et al., 2018). Such models describe the spatio-temporal distribution of the water resources and water stress under the impact of both human and natural mechanisms. Global hydrological modeling is used to predict water fluxes in the Earth's terrestrial regions, such as evapotranspiration, river discharge, groundwater recharge, and water storage (or merely water storage

fluctuations) in various compartments, such as soil, groundwater, and surface water bodies (Müller Schmied et al., 2021; Sood & Smakhtin, 2015; Sutanudjaja et al., 2018a). Although actual GHMs have few calibratable parameters, the recent advanced models are becoming more complex, provide invaluable spatio-temporal estimates of global water resources, and help to analyze possible projections/scenarios of changes in those estimates (Sood & Smakhtin, 2015). A continuous improvement of modeling in the GHMs aims to reduce uncertainty, incorporating fine-scale satellite observations, calibrating/assimilating models, and downscaling in time and space (Hoch et al., 2022; Schewe et al., 2019).

An atmospheric reanalysis model systematically monitors and forecasts the climate state by combining models with observations. The data assimilation scheme and model(s) are unchanged ("frozen") during the process, which enables a dynamically consistent estimate of the states on the global scale (Dee et al., 2016). The reanalysis models contain a numerical description of the climate, including surface parameters such as rainfall, soil moisture content, and sea surface temperature over a long period that can extend back several decades (Cucchi et al., 2020; Muñoz-Sabater et al., 2021). Several groups have developed global atmospheric reanalysis models, including the recent products of the JRA-55 produced by the Japan Meteorological Agency (JMA) (Kobayashi et al., 2015), the MERRA-2 reanalysis from the NASA Global Modeling and Assimilation Office (GMAO) (Gelaro et al., 2017), and the Climate Forecast System Reanalysis (CFSR) produced by the National Centers for Environmental Prediction (NCEP) (Saha et al., 2006; Saha et al., 2014) (for a comprehensive list of the reanalysis products please see Dee et al. (2016)). Besides the mentioned models, the European Centre for Medium-Range Weather Forecasts (ECMWF) has a long history with reanalysis products starting from 1979 with the First GARP Global Experiment (FGGE) project (Bengtsson et al., 1982) and ERA-15 in the 1990s (Gibson, 1997), followed by ERA-40 from 2001 to 2003 (Uppala et al., 2005), ERA-Interim from 2006 to 2019, and finally the fifth generation, which is called ERA5 in 2019 (Hersbach et al., 2020). A successive product aims for higher spatio-temporal resolution, incorporating more sophisticated data assimilation techniques and delivering consistent estimates of the climate states using the atmosphere, land surface, ocean sea ice, and carbon cycle (Hersbach et al., 2020).

3.3.2. Global models used in this study

In this study, we have employed in total 13 state-of-the-art datasets of Global Hydrological Models (GHMs), Land Surface Models (LSMs), and atmospheric reanalysis models (Table 3.1). Nine multi-decadal global water resources datasets were obtained from the eartH2Observe Water Cycle Integrator (WCI; ftp://wci.earth2observe.eu (last access: 31 May 2021), including PCR-GLOBWB, SURFEX-TRIP, HBV-SIMREG, HTESSEL-CaMa, JULES, LISFLOOD, ORCHIDEE, SWBM, and W3RA. The output of these datasets are available at 0.5° spatial resolution over the period 1979–2012. Beside datasets from eartH2Observe, we have included the Community Land Model Version 5 (CLM5) with two standard forcing datasets, namely the Global Soil Wetness Project forcing data set (GSWP3) and CRUNCEP (the combination of the Climate Research Unit (CRU) and the National Centers for Environmental Prediction (NCEP)). The CLM5 datasets are at 0.5° spatial resolution covering the period 1901–2014 (for more detail about the CLM5 model, please see Lawrence et al., 2019). The CLM5 products are accesible via Earth System Grid (ESG) (Oleson et al., 2019). We have also included the latest version of the WaterGAP Global Hydrology Model (WaterGAP v2.2d) (Müller Schmied et al., 2021), covering the period 1901–2016 and at 0.5° spatial resolution. The outputs of the WaterGap v2.d are available at (https://doi.pangaea.de/10.1594/PANGAEA.918447). Finally, we have included the fifth generation ECMWF atmospheric reanalysis of the global climate (ERA5) at 0.25° spatial resolution which provide data from 1979 to present. The data is downloaded from the Copernicus Climate Change Service (C3S) at ECMWF (https://cds.climate.copernicus.eu)(last access: 30 May

2021). TWSA from models carries a higher spatial resolution and therefore values with higher frequency. To set the same spectral content in models comparing to GRACE TWSA, we have transferred the models outputs into the spectral domain and truncated the SHs to the maximum degree and order 96. Finally we recovered the TWSA fields from the truncated SHs.

Table 3.1: Summary of global models used in this study. GHM: Global Hydrological Model; LSM: Land Surface Model; ReA: Reanalysis Model.

	Model	Time Period	Data Provider	Reference
GHM	WGHM	1901–2016	Goethe University Frankfurt	Müller Schmied et al., 2021
	PCRGLOB-WB	1979–2012	Utrecht University (UU)	Sutanudjaja et al., 2018b; Wada et al., 2014
	HBV-SIMREG	1979–2012	Joint Research Centre (JRC)	Lindström et al., 1997
	LISFLOOD	1979–2012	Joint Research Centre (JRC)	Van Der Knijff et al., 2010
	W3RA	1979–2012	CSIRO**	Van Dijk, 2010
	SWBM	1979–2012	Simple Water Balance Model	Koster and Mahanama, 2012; Orth and Seneviratne, 2013
LSM	CLM5	1940–2014	The Earth System Grid (ESG) at NCAR	Lawrence et al., 2019
	HTESSEL	1979–2012	ECMWF	Balsamo et al., 2015
	JULES	1979–2012	Centre for Ecology and Hydrology (CEH)	Best et al., 2011; Clark et al., 2011
	ORCHIDEE	1979–2012	French National Centre for Scientific Research	Polcher et al., 2011
	SURFEX-TRIP	1979–2012	Meteo France	Decharme et al., 2013
ReA	ERA5	1979–2016	ECMWF*	Hersbach et al., 2020

* ECMWF: European Centre for Medium-Range Weather Forecasts

** CSIRO: Commonwealth Scientific and Industrial Research Organisation

3.3.3. Assessment of the global models

To assess the consistency between models, we have computed the temporal RMS and the annual trend from the models mentioned above within 1980–2012 for the 403 major river basins (Figure 3.5). The envelope represents the range (minimum and maximum) of values across different models, and basins are grouped in different continents or sub-continents separated by the white and gray background color. The basins within each sub-continent are sorted according to the latitude of their center. In general, Figure 3.5 indicates high variation within models in estimating the two principal components of the TWSA time series. The highest discrepancies in RMS are observed in North America, north of South America, Australia, and southern Asia. The same pattern can be observed in the annual trend estimation while considerable variations over North America and North Europe. Figure 3.5 highlights the need for considering a combination selected models rather than a single model for estimating TWSA over a different part of the world.

In order to assess the performance of the above-mentioned models over different climate categories, we have computed the correlation coefficient and the Nash–Sutcliffe model efficiency coefficient (NSE) between each model with GRACE from 2003 to 2012 over the 403 major river basins. The boxplots in Figure 3.6 illustrate the correlations in four different climate categories (please see section 2.4 for details about the categories). On average, the highest correlations (> 0.8) are obtained from JULES, followed by CLM5-GSWP3 and HTESSEL, while the lowest correlations (< 0.75) are observed from HBV-SIMREG, followed by SURFEX-TRIP and W3RA. In general, the correlation increases by the humidity so that arid to hyper-arid regions show the lowest correlations between models and GRACE, while the highest correlations are achieved over sub-humid to humid regions.

In terms of NSE, generally HTESSEL outperforms others ($\overline{\text{NSE}} = 0.53$) followed by JULES ($\overline{\text{NSE}} = 0.50$) and ORCHIDEE ($\overline{\text{NSE}} = 0.43$). The highest variation of NSE among basins is observed over the arid to hyper-arid regions among the CLM5 datasets, ERA5, and SWBM. For the arid to hyper-arid category, most of the models have negative NSE values, indicating poor model
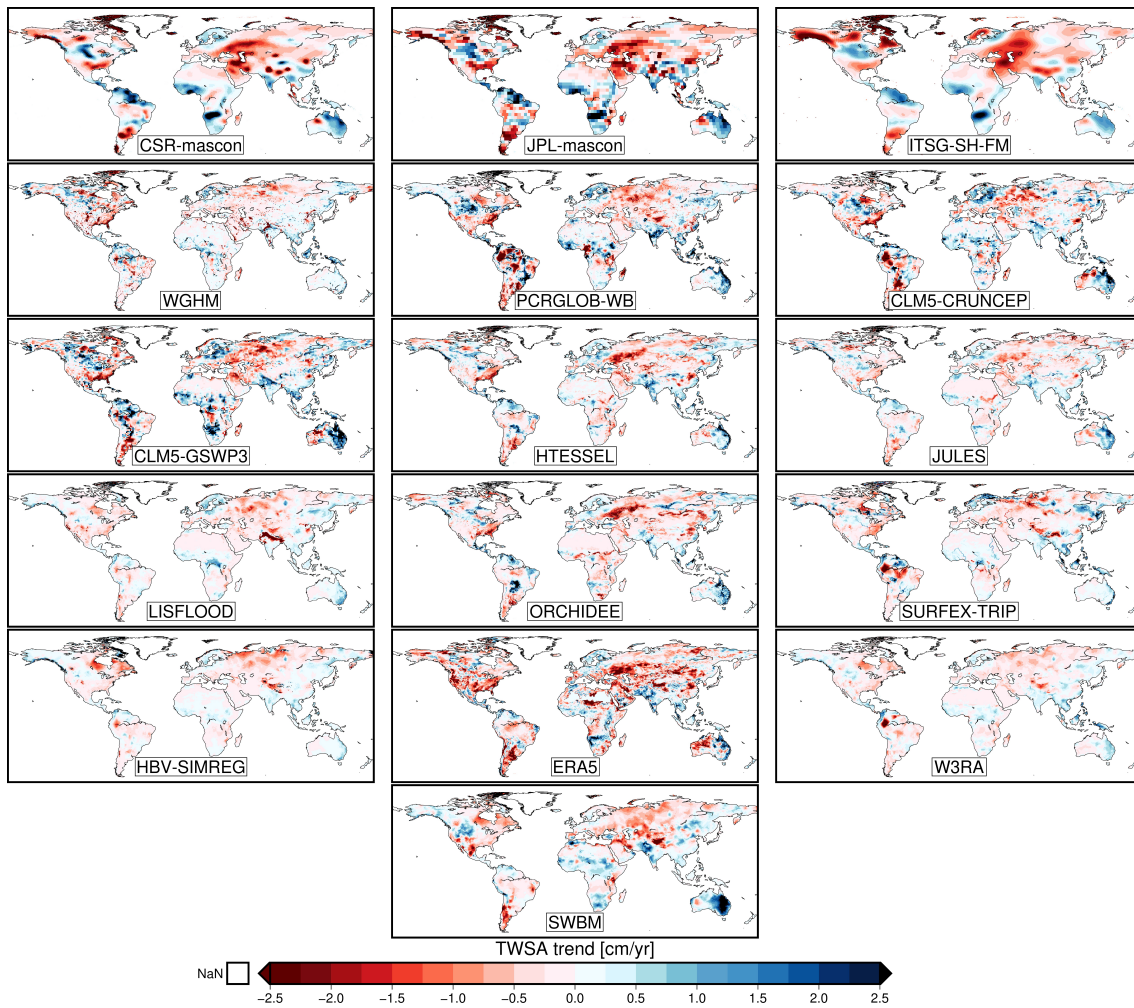
Figure 3.4: Annual trend value of the TWSA over the global land (except for Greenland and Antarctica) from GRACE (first row) and various models. The trend values are obtained during the period from 2003 to 2012.

performance. The only models that show positive NSE values are HTESSEL, ORCHIDEE, and PCRGLOBWB. In the semi-arid category, all models have positive NSE values, indicating better performance compared to the arid to hyper-arid category. The highest NSE value is observed for the model JULES. In the Dry sub-humid category, all models have positive NSE values, and the highest NSE value is observed for the model ERA5. Finally, in the humid category, most models have positive NSE values, with the highest value observed for the model JULES. Overall, the results suggest that hydrological models perform better in regions with higher moisture availability, such as Dry sub-humid and humid climates. This is expected, as the availability of water is a critical factor in hydrological processes, and accurate representation of water resources is necessary for accurate hydrological modeling. The results from Figure 3.6 and Figure 3.6 also support the conclusion mentioned earlier in this section regarding the importance of model combination, as different models perform differently in different climatic regions, and no single model performs consistently well across all regions.
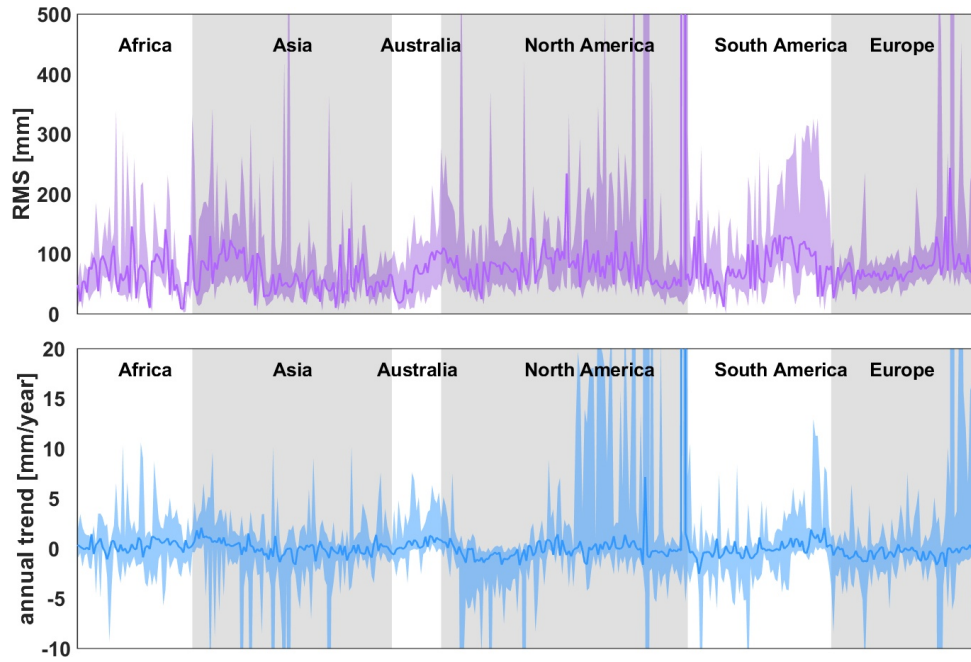
Figure 3.5: The Root Mean Square (RMS) and the annual trend from the models (hydrological, atmospheric, and land surface) for the global major river basins, except Greenland and Antarctica, within 1980–2012. The envelope shows the range of the variations, and the line with solid color represents the ensemble median of the models. Each continent or subcontinent has been separated using the gray/white background shade, and within each continent or subcontinent, the basins are sorted by latitude of the basins' center point in ascending order.

Figure 3.6: Boxplot of the correlation coefficient between the models and GRACE over major basins within 2003–2012 for different categories of climate regions.

Figure 3.7: Boxplot of the NSE between the models and GRACE within 2003–2012 for different categories of climate regions.

3.4. Methodology

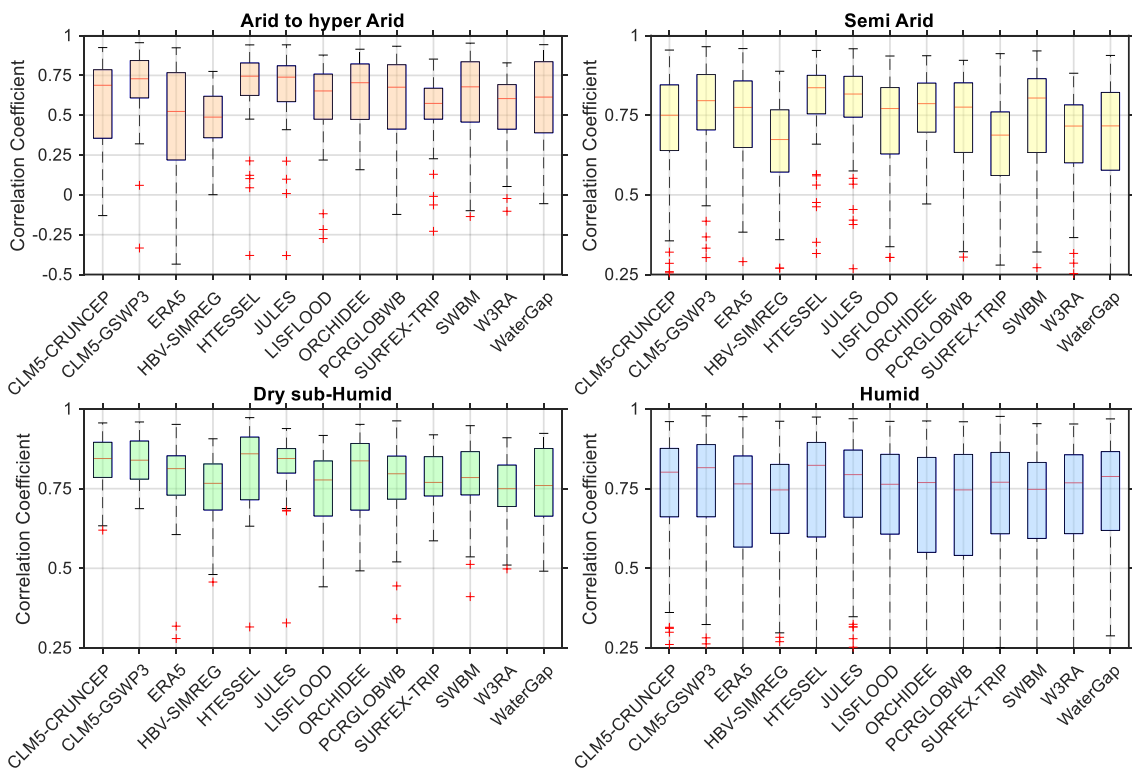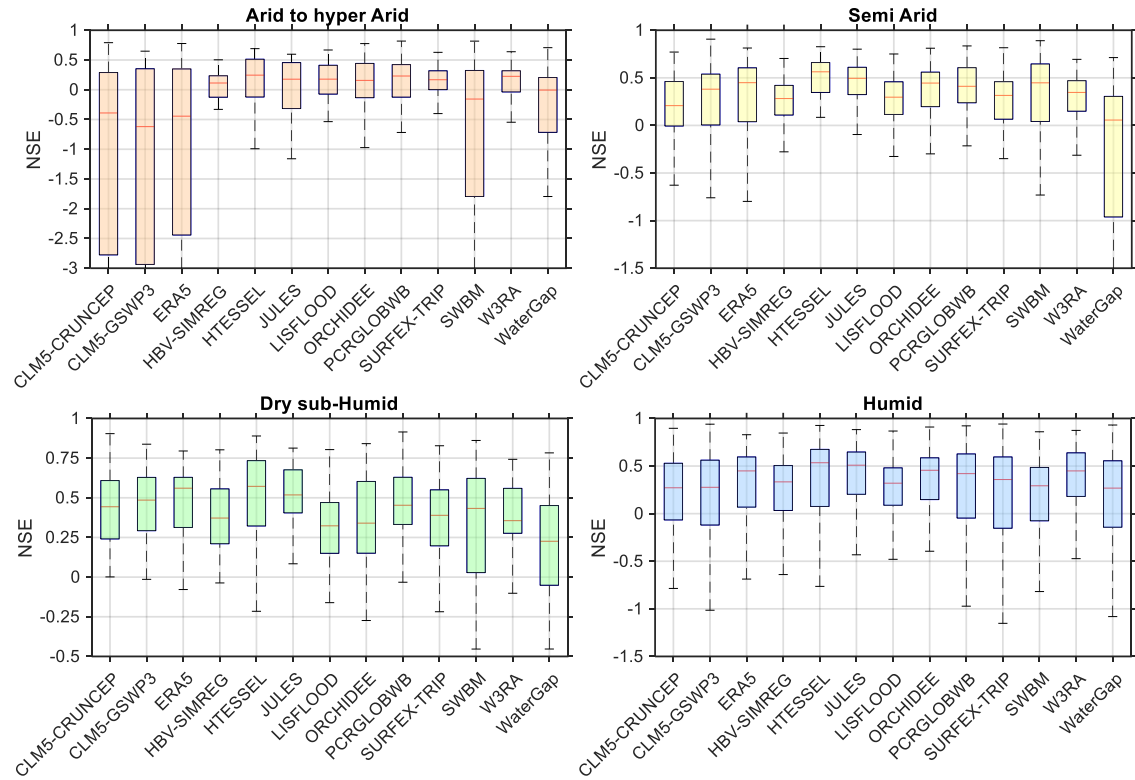As it is shows in subsection 3.3.3, each model incorporates unique assumptions, algorithms, or parameterizations, which can lead to distinct perspectives on the underlying processes driving TWSA variations. By considering a broader set of models, we increase the likelihood of capturing a wider range of TWSA patterns and potential sources of uncertainty. To combine the models and estimate TWSA before GRACE era, we have calculated the ensemble mean (EM) and the ensemble weighted mean (EWM) of the models. In addition to the ensemble mean and ensemble weighted mean, we have evaluated six alternative approaches for hindcasting GRACE TWSA using a combination of models. These approaches were considered to explore potential variations and improvements beyond the ensemble-based methods. By investigating a diverse range of modeling techniques, we aimed to capture a comprehensive understanding of the complex dynamics involved in the GRACE TWSA hindcasting process. The six methods include:

- Multivariate Linear Regression (MLR)

- Non-Negative Least Squares (NNLS)

- Support Vector Machine (SVM)

- Decision Tree (DT)

- Random Forest (RF)

- Gaussian Process Regression (GPR)

Multivariate Linear Regression (MLS) and Non-Negative Least Squares Regression (NNLS) are the extensions of the classical Least Squares Regression (LSR) method. They differ in their approach to addressing the limitations of LSR and have specific advantages and limitations. MLS extends the LSR method to multiple response variables, allowing for a more comprehensive analysis of the relationships between multiple predictor and response variables. NNLS, on the other hand, constrains the regression coefficients to be non-negative, which can result in a more interpretable and easily understood model.

We employed a variety of machine learning techniques to analyze the relationships between our predictor and response variables. We used Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Gaussian Process Regression (GPR). SVM is a supervised learning algorithm that constructs a hyperplane or a set of hyperplanes in a high-dimensional space to separate the different classes in our dataset. DT and RF are also supervised learning algorithms that create decision trees or a forest of decision trees to classify or regress our data. Gaussian Process Regression (GPR) is a Bayesian non-parametric method that can also improve the accuracy and precision of regression parameter estimates. Compared to MLS and NNLS, GPR has the advantage of being able to capture nonlinear relationships between the predictor and response variables. However, GPR can be more computationally intensive and require the selection of a kernel function to model the covariance between the observations. In what follows, each of these approaches are introduced.

3.4.1. The ensemble (weighted) mean

To enhance our comprehension and forecast of terrestrial water storage changes, we can utilize GRACE data to hindcast TWSA data by integrating multiple models. One simple strategy to incorporate these models is to employ an ensemble mean of the models, which functions as a

3

benchmark for evaluating outcomes from other techniques. This ensemble mean is achieved by averaging the output of each model, which helps to alleviate the impact of individual model biases, errors, and uncertainties. Besides the ensemble mean, one can integrate models using the weighted mean. In addition to the ensemble mean, models can also be integrated using a weighted mean. Since the performance of the models varies over different regions and climate (see subsection 3.3.3), a combination of models using different weights might have a better performance. To accomplish this, we calculated the Root Mean Squared (RMS) of the GRACE signal from 2003–2012 and compared it with the RMS from the models during the same period. The inverse of the absolute difference between the RMS from the models and GRACE were then used as their corresponding weights. The RMS is a measure of the overall agreement between the GRACE observations and the models. Using this measure allows you to identify models that closely match the observed GRACE data, indicating their potential accuracy in representing the TWSA variations. Besides, by considering the differences between the RMS values, we are effectively taking into account potential biases in the models. Models with systematic biases that consistently deviate from the GRACE signal will have larger differences in their RMS values.

$$\text{TWSA}_{\text{EWM}} = \frac{\sum_{i=1}^{n} w_i \cdot \text{TWSA}_i}{\sum_{i=1}^{n} w_i} \tag{3.1}$$

$$w_i = |\text{RMS}_{\text{GRACE}} - \text{RMS}_{\text{Model}}|^{-1} \tag{3.2}$$

where n is the number of models, TWSA_i represents the TWSA from model i, and w_i represents the weight associated with model i.

Figure 3.8 displays the time series of GRACE, the ensemble mean of the models (EM), and the ensemble weighted mean (EWM) of the models across selected basins. The disparities between GRACE and the ensemble models are negligible in certain basins such as Danube, Nile, and Ob. Conversely, the ensemble models depicts considerable differences with the GRACE observations in the RMS of the signal in basins like Amazon, or the trend in basins such as Tigris, Mackenzie, and Mississippi. In the results section, we compare the performance of such an ensemble mean over the GRACE time period with other approaches. To compare results from different model categories, it is essential to calculate the ensemble mean for each category. This is crucial because each category, namely Global Hydrological Models (GHMs), Land Surface Models (LSMs), and Reanalysis models, operates under distinct assumptions and uses different input data, resulting in diverse output. Therefore, calculating the ensemble mean for each category of models provides a more precise and robust estimate of the TWSA. This approach also helps to determine the strengths and weaknesses of each category of models accurately.

We compare the ensemble mean (EM) and the ensemble weighted mean (EWM) of three groups of models - Global Hydrological Models (GHM), Land Surface Models (LSM), and Reanalysis (ReA) - with GRACE over the major river basin within 2003-2012. We used three metrics to evaluate the models including Kling-Gupta Efficiency (KGE), Normalized Root Mean Squared Error (NRMSE), and Mean Bias Error (MBE). The results are presented in Table 3.2 that provides a comparison between EM and EWM for each of the three model groups across all basins and four climate categories - arid to hyper-arid, semi-arid, Dry sub-humid, and humid. It appears that the EWM performs slightly better than EM in terms of KGE and NRMSE for all three model groups. However, for MBE, the performance of EWM is mixed. In the case of GHM and ReA, the EWM performs slightly worse than EM, while for LSM, EWM performs slightly better than EM. When considering the results by climate category, it appears that the performance of the models varies depending on the climate category. For arid to hyper-arid, EWM performs better than EM for all three model groups across all three metrics. However, for semi-arid and Dry

sub-humid, the performance of EWM is mixed, and for humid, EM performs better than EWM for all three model groups across all three metrics. The results suggest that using an ensemble weighted mean can improve model performance in some cases, but the results may vary depending on the model group and climate category.

Table 3.2: Performance comparison of ensemble mean (EM) and ensemble weighted mean (EWM) of different types of models (Global Hydrological Models (GHM), Land surface Models (LSM) and Reanalysis Models (ReA)) in terms of three metrics (KGE, NRMSE, and MBE) against GRACE TWSA over major river basins. The performance evaluation is conducted separately for different climatic categories including arid, semi-arid, dry sub-humid, and humid. The shaded cells represent the top three performance within each column for each metric and category while the darker the color, the better the performance.

|	Metric	EM-GHM	EM-LSM	EM-ReA	EWM-GHM	EWM-LSM	EWM-ReA
all basins	KGE	−11.34	−5.44	−9.21	−10.33	−5.11	−10.12
	NRMSE	0.65	0.62	0.65	0.63	0.58	0.65
	MBE	1.78	1.25	−0.70	1.47	1.00	−1.01
arid	KGE	−2.39	−1.29	−0.75	−0.88	−0.70	−0.75
	NRMSE	0.61	0.61	0.59	0.62	0.50	0.58
	MBE	1.30	4.30	0.20	1.67	1.85	−0.90
semi-arid	KGE	−29.00	−13.86	−27.43	−32.29	−11.73	−29.85
	NRMSE	0.62	0.61	0.64	0.62	0.58	0.64
	MBE	3.49	1.67	0.81	2.71	0.99	0.45
dry sub-humid	KGE	−14.33	−7.80	−12.25	−12.75	−6.26	−14.68
	NRMSE	0.44	0.44	0.47	0.42	0.38	0.51
	MBE	10.01	4.64	6.48	9.28	5.18	8.03
humid	KGE	−5.18	−2.30	−2.73	−2.63	−2.96	−2.91
	NRMSE	0.70	0.66	0.70	0.68	0.62	0.70
	MBE	−0.35	−0.02	−2.78	−0.50	0.10	−3.28

3
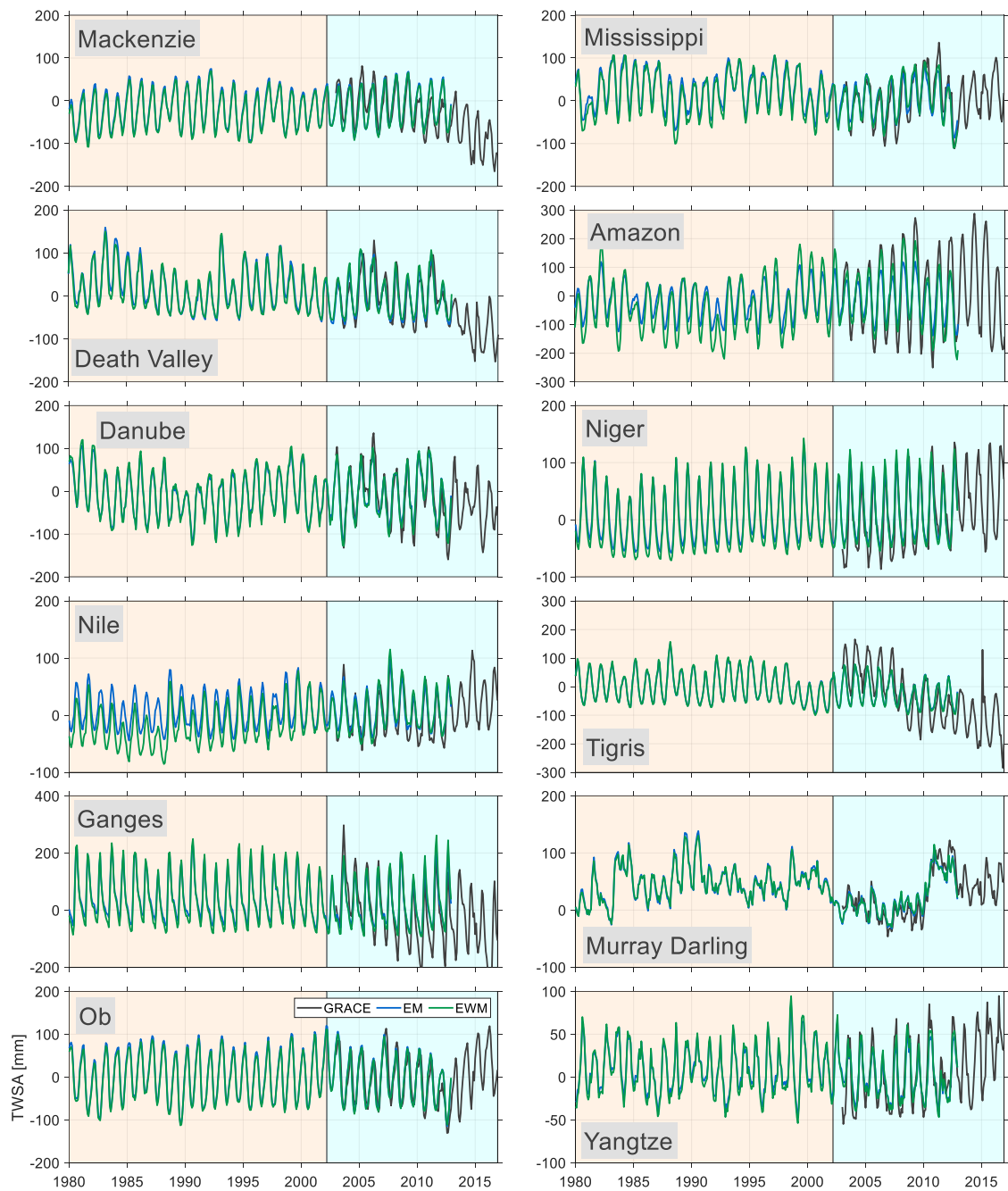
Figure 3.8: Time series of TWSA obtained from GRACE compared with the one estimated from the models' ensemble mean (EM) and ensemble weighted mean (EWM) over selected basins.

3.4.2. Multivariate Linear Regression

Multivariate Linear Regression (MLR) is a statistical method used for estimating the parameters of a linear regression model with multiple independent variables. The history of MLS dates back to the early 1800s, when it was first proposed by Carl Friedrich Gauss for the analysis of astronomical data. Later, it was further developed by several statisticians, including Francis Galton and Karl Pearson. The basic idea behind MLS is to find the coefficients that minimize the sum of squared errors between the predicted and actual values of the dependent variable. The formula for MLS is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{3.3}$$

Where, $\mathbf{y}$ is the vector of dependent variable values, $\mathbf{X}$ is the matrix of independent variable values, $\boldsymbol{\beta}$ is the vector of coefficients to be estimated, and $\boldsymbol{\epsilon}$ is the vector of errors, which are assumed to be normally distributed with mean zero and constant variance.

MLS has several advantages, including its ability to handle multiple independent variables and to model complex relationships between variables. It also provides estimates of the coefficients and their standard errors, which can be used to test hypotheses and construct confidence intervals. However, MLS assumes that the errors are normally distributed and have constant variance, which may not always be true in practice. Additionally, it can be sensitive to outliers and multicollinearity among the independent variables.

3.4.3. Non-Negative Least Squares Regression

Non-Negative Least Squares (NNLS) regression is a method used for solving linear regression problems where the coefficients are constrained to be non-negative. The basic idea behind NNLS is to find the coefficients that minimize the sum of squared errors subject to the non-negativity constraint. The history of NNLS dates back to the early 1970s, when it was first proposed by Lawson and Hanson as a method for solving linear regression problems with non-negative coefficients. Since then, NNLS has been widely used in various fields, including biology, chemistry, and economics. The formula for NNLS is as follows:

$$\min_{\boldsymbol{\beta} \geq 0} \left| \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right|_2^2 \tag{3.4}$$

Where, $\mathbf{y}$ is the vector of dependent variable values, $\mathbf{X}$ is the matrix of independent variable values, $\boldsymbol{\beta}$ is the vector of coefficients to be estimated, subject to the constraint that $\boldsymbol{\beta} \geq 0$, $|\cdot|_2^2$ denotes the Euclidean norm.

NNLS has several advantages, including its ability to handle problems where the coefficients must be positive, such as in signal processing and image analysis. It also provides a unique solution and can be efficiently solved using iterative algorithms. However, NNLS assumes that the relationship between the dependent and independent variables is linear, which may not always be true in practice. Additionally, the non-negativity constraint may result in a biased estimation of the coefficients.

3.4.4. Support Vector Machine

Support Vector Machines (SVMs) are a popular class of supervised learning algorithms that can be used for classification and regression tasks. In the context of regression, the goal of SVM is

to find a function $f(x)$ that approximates the relationship between input features x and output y, given a set of training examples $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$.

The SVM algorithm for regression involves finding a hyperplane in a high-dimensional feature space that has the maximum margin from the training data. This hyperplane can be represented as:

$$f(x) = w^T \phi(x) + b \tag{3.5}$$

where w is the weight vector, $\phi(x)$ is the feature mapping function that transforms the input x into a higher-dimensional space, and b is the bias term. The feature mapping function is usually determined by the choice of a kernel function, which defines the similarity between two input data points in the high-dimensional space.

The kernel function plays a crucial role in the SVM algorithm, as it allows the algorithm to find nonlinear relationships between the input features and output. The most commonly used kernel functions for SVM regression are:

1. Linear kernel: $K(x, x') = x^T x'$

2. Polynomial kernel: $K(x, x') = (x^T x' + c)^d$

3. Radial Basis Function (RBF) kernel: $K(x, x') = \exp(-\gamma ||x - x'||^2)$

where c, d, and γ are hyperparameters that control the shape and complexity of the decision boundary.

To find the optimal weight vector w and bias term b, the SVM algorithm minimizes a regularized objective function that balances the margin and the loss incurred by violating the margin. The objective function can be written as:

$$\min_{w,b} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \mathcal{L}_\epsilon (y_i, f(x_i)) \tag{3.6}$$

where $||w||^2$ is the L2-norm of the weight vector, $\mathcal{L}_\epsilon$ is the epsilon-insensitive loss function that penalizes the deviations of predicted values from true values by at least epsilon (ϵ), and C is the regularization parameter that controls the trade-off between fitting the training data and preventing overfitting.

3.4.5. Random Forest

Random Forest is a powerful machine learning algorithm that can be used for both classification and regression tasks. It belongs to the family of ensemble learning algorithms that combine multiple decision trees to form a single model. The Random Forest algorithm can be used for regression tasks by averaging the predictions of multiple decision trees.

The algorithm works by first creating a set of decision trees using random subsets of the training data and random subsets of the features. During the training phase, each tree is grown using a portion of the training data, selected randomly with replacement (bootstrap sample). Then, at each node of each tree, a subset of the features is randomly selected and the best split is made using one of several criteria, such as Mean Squared Error (MSE) or Mean Absolute Error (MAE).

The prediction of the Random Forest model for a new input x is obtained by averaging the predictions of all the trees in the forest. For regression tasks, the predicted value $\hat{y}$ is the average of the predicted values of all the trees:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} f_b(x), \tag{3.7}$$

where B is the number of trees in the forest, f_b is the prediction of the b-th tree, and $\hat{y}$ is the final prediction.

The Random Forest algorithm can also be used to calculate the importance of each feature in the dataset. Feature importance is a measure of how much each feature contributes to the prediction of the model. The feature importance score I_j for feature j is calculated by averaging the reduction in the impurity measure (such as MSE or MAE) over all the trees in the forest that use feature j for splitting:

$$I_j = \frac{1}{B} \sum_{b=1}^{B} \sum_{i \in N_b(j)} w_b p_b(i)(y_i - \bar{y}Nb(j))^2, \tag{3.8}$$

where $N_b(j)$ is the set of nodes that use feature j for splitting in the b-th tree, w_b is the weight of the b-th tree, $p_b(i)$ is the proportion of samples that reach node i in the b-th tree, y_i is the true label of the i-th sample, and $\bar{y}Nb(j)$ is the mean of the labels of the samples in node $N_b(j)$.

The Random Forest algorithm can be further enhanced by using different types of splits or by combining it with other algorithms. For example, Random Forest with Extra Trees (Extremely Randomized Trees) is a variant that uses extremely randomized splits instead of optimal splits, resulting in faster training and lower variance at the cost of higher bias. Additionally, using Random Forest as a feature selection method in combination with other algorithms such as SVM or Gradient Boosted Trees can lead to improved performance.

3.4.6. Decision Tree

Decision Trees (DT) are a non-parametric supervised learning method used for both regression and classification problems. In regression problems, DT is used to approximate the mapping function between input features and target values. The idea behind DT is to recursively split the input space into smaller regions using a set of rules based on the input features until the resulting regions become homogeneous with respect to the target variable.

The splitting of the input space is performed using a tree-like structure where the internal nodes represent the splitting rules based on the input features, and the terminal nodes, also called leaves, represent the resulting regions with the corresponding target values. DT can be trained using a top-down greedy approach called recursive binary splitting. The idea is to select the best feature and the corresponding split value that best separates the target variable in the input space.

The quality of the split is evaluated using a criterion that measures the impurity of the resulting regions. One commonly used impurity measure is the mean squared error (MSE). Once a split is performed, the process is recursively applied to each resulting region until a stopping criterion is met, such as a maximum depth of the tree or a minimum number of samples required to split a node.

One limitation of DT is that it tends to overfit the training data, resulting in poor generalization performance on unseen data. One way to address this issue is to use ensemble methods such as Random Forest (RF) that combine multiple DT models to improve the predictive performance.

In summary, the main steps involved in training a DT model for regression are:

1. Split the input space using a set of rules based on the input features.

2. Evaluate the quality of the split using an impurity measure, such as the mean squared error.

3. Recursively apply the splitting process to each resulting region until a stopping criterion is met.

4. Use the resulting regions as leaves with the corresponding mean target values as predictions.

3.4.7. Gaussian Process Regression

Gaussian Process Regression (GPR) is a non-parametric Bayesian regression method used for modeling and predicting complex nonlinear relationships between variables. The history of GPR dates back to the early 1950s, when it was first introduced by Andrey Kolmogorov as a tool for modeling random processes. Later, it was further developed by several researchers, including David MacKay and Carl Rasmussen, who introduced the concept of Bayesian inference in GPR. Today, GPR is widely used in various fields, including engineering, finance, and biology. The basic idea behind GPR is to model the target variable as a Gaussian process, which is a collection of random variables, any finite number of which have a joint Gaussian distribution.

GPR has several advantages, including its ability to model complex nonlinear relationships and to provide a probabilistic estimate of the predictions, which can be used to construct confidence intervals. It also has the ability to handle missing data and to incorporate prior knowledge into the model. However, GPR can be computationally expensive and may suffer from overfitting if the covariance function is not properly specified. The formula for GPR is as follows:

$$y = f(\mathbf{x}) + \epsilon \tag{3.9}$$

Where, y is the target variable to be predicted, $f(\mathbf{x})$ is the latent function, which is modeled as a Gaussian process, $\mathbf{x}$ is the vector of input variables, ϵ is the random error term, which is assumed to be normally distributed with mean zero and constant variance. The latent function values are assumed to follow a multivariate Gaussian distribution with a mean of zero and a covariance matrix $\mathbf{K}$, where $\mathbf{K}_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function evaluated at inputs $\mathbf{x}_i$ and $\mathbf{x}_j$. The noise values are also assumed to follow a multivariate Gaussian distribution with a mean of zero and a covariance matrix $\sigma_n^2 \mathbf{I}$, where σ_n^2 is the noise variance and $\mathbf{I}$ is the identity matrix. The formula for the prediction of the target variable at a new input $\mathbf{x}_*$ is:

$$\hat{y}* = \mathbf{k}*^\top \mathbf{K}^{-1} \mathbf{y} \tag{3.10}$$

where $\mathbf{k}*$ is the vector of kernel values between $\mathbf{x}*$ and each training input.

In matrix form, the main formula for GPR can be written as:

$$\begin{bmatrix} \mathbf{y} \; \mathbf{f}* \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{K}* \; \mathbf{K}*^\top & \mathbf{k}** \end{bmatrix} \right) \tag{3.11}$$

where $\mathbf{f}*$ is the vector of latent function values at the new input, $\mathbf{K}*$ is the vector of kernel values between the training inputs and the new input, and $\mathbf{k}_{**}$ is the kernel value at the new input. The formula for the prediction of the target variable at a new input can be written as:

$$\hat{y}* = \mathbf{k}*^\top (\mathbf{K} + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} \tag{3.12}$$

where the inverse of the covariance matrix is used to compute the weights of the linear combination of training targets, as determined by the kernel similarity between the training inputs and the new input. The kernel function is seen in the calculation of the covariance matrix $\mathbf{K}$ and the kernel vectors $\mathbf{K}*$ and $\mathbf{k}**$. The choice of kernel or covariance function plays a crucial role in modeling the relationship between the input variables and the target variable. The kernel specifies the similarity between any two input variables, and thus determines the smoothness and shape of the underlying function that is being modeled by the Gaussian process. A kernel function must be positive semi-definite to ensure that the resulting covariance matrix is also positive semi-definite. There are several commonly used kernel functions in GPR, including:

1. Radial Basis Function (RBF) kernel: This is the most commonly used kernel function in GPR, and it assumes that the function being modeled is smooth and continuous. The RBF kernel has a single length scale hyperparameter, which controls the degree of smoothness of the function.

2. Matérn kernel: The Matérn kernel is a more flexible kernel that can model functions with varying degrees of smoothness. It has two hyperparameters: the length scale and the smoothness parameter.

3. Periodic kernel: The periodic kernel is used when the function being modeled has a repeating pattern. It has a single hyperparameter, which controls the period of the function.

4. Linear kernel: The linear kernel assumes that the function being modeled is linear. It has no hyperparameters.

GPR involves estimating the hyperparameters of the Gaussian process, which include the length scale and the amplitude of the covariance function. The covariance function determines the similarity between the input variables and is typically chosen based on prior knowledge of the problem.

3.4.8. Selecting the Kernel and Hyperparameters

The selection of the kernel function and hyperparameters is contingent upon the nature of the modeled problem and the available prior knowledge. Certain kernels may be better suited for specific types of data, for instance, the periodic kernel for data exhibiting periodic patterns. Gaussian Process Regression (GPR) and Support Vector Machine (SVM) are two powerful algorithms used for regression problems in Machine Learning. One of the key factors that contribute to their effectiveness is the use of a kernel function. In GPR, the kernel function plays a crucial role in defining the covariance structure of the prior distribution over the function values. It determines the similarity between any two points in the input space, and hence how

strongly their output values are correlated. The choice of kernel function reflects the assumptions made about the underlying function, such as smoothness or periodicity. Popular choices of kernel functions in GPR include the RBF kernel and the Matérn kernel.

Similarly, in SVM, the kernel function serves as a mapping between the input space and a higher dimensional feature space, where the linearly non-separable problem can be solved. The kernel function measures the similarity between any two points in the input space, and the SVM algorithm uses this similarity to find the optimal separating hyperplane. By choosing an appropriate kernel function, the SVM algorithm can capture complex nonlinear relationships between the input features and the output variable. Common kernel functions used in SVM include the RBF kernel, polynomial kernel, and sigmoid kernel.

In this study, we have incorporated the most prevalent kernels for GPR. We have considered a variety of kernels, ten in total, such as exponential, squared-exponential, Matern32, Matern52, rational quadratic, ard-exponential, ard-squared-exponential, ard-Matern32, ard-Matern52, and ard-rational quadratic. We have also employed four different kernels for SVM including Gaussian, Radial Basis Function (RBF), linear, and polynomial kernel. For each kernel, the hyperparameters that yield optimal performance during the training period are selected. For each river basin, we have computed ten reconstructions of TWSA resulting from the kernels, utilizing the best hyperparameter choices.

Hyperparameters are parameters that are set before the model training process, unlike model parameters, which are learned during the training process. Other than kenel type, examples of hyperparameters include learning rate, regularization strength, number of hidden layers. The values of hyperparameters significantly impact the model's performance, and the optimal values vary depending on the specific problem and dataset. In general, the hyperparameter tuning process involves selecting a range of values for each hyperparameter and evaluating the model's performance for each combination of hyperparameters.

Common approached for hyperparameter tuning are:

- Grid search: a grid of hyperparameters is defined, and the model is trained and evaluated for each combination of hyperparameters in the grid.

- Random search: hyperparameters are randomly sampled from a distribution, and the model is trained and evaluated for each combination of hyperparameters.

- Maximum likelihood estimation (MLE): a commonly used method for hyperparameter tuning in machine learning algorithms. MLE is a statistical method that seeks to find the values of hyperparameters that maximize the likelihood of the observed data given the model. In other words, MLE estimates the values of hyperparameters that make the model most likely to have generated the observed data.

- Bayesian optimization: a probabilistic model is used to model the performance of the model as a function of the hyperparameters. The probabilistic model is then used to select the next set of hyperparameters to evaluate, based on the expected improvement in the model's performance.

In this study we have employed the Bayesian optimization approach to tune the hyperparameters for each algorithm. Bayesian optimization is a powerful method for tuning hyperparameters in machine learning algorithms because it combines the benefits of both random and grid search methods. Unlike grid search, Bayesian optimization does not require a predefined grid of hyperparameters to explore. Instead, it models the performance of the model as a function of

the hyperparameters and uses this model to select the next set of hyperparameters to evaluate. This approach allows the algorithm to focus on promising areas of the hyperparameter space and quickly eliminate unpromising regions. Moreover, Bayesian optimization provides a probabilistic model of the model's performance that can be used to estimate the uncertainty of the performance estimate, which is valuable for making decisions about which hyperparameters to use in the final model.

3.5. Results

3.5.1. Training period sensitivity

The duration of the training period is a critical aspect to consider when training and assessing a statistical model. Generally, an increase in the length of the training data is likely to lead to improved model performance, as it enables the model to capture the patterns and interrelationships in the data more accurately and glean more information. However, an excessive amount of training data may result in overfitting, where the model becomes too specialized to the training data and fails to generalize to new data. Therefore, it is essential to strike a balance between using sufficient training data to achieve good performance and avoiding overfitting. The optimal length of the training data depends on various factors, including model intricacy, the amount and quality of available data, and the specific problem under consideration. Conducting experiments with different training durations can assist in determining the ideal length for a particular model and problem.

We assessed the performance of the six different methods for reconstructing TWSA over major river basins using varying training periods ranging from two to eight years. Figure 3.9 displays the performance of these six methods, indicating that the mean NRMSE for all methods does not vary significantly as the training period decreases. Among the methods, the multiple linear regression (MLR) and Gaussian process regression (GPR) methods displayed the lowest mean NRMSE across all training periods. In contrast, the support vector machine (SVM) and decision tree (DT) methods exhibited higher mean NRMSE, indicating a lower accuracy in TWSA estimation, particularly for shorter training periods. In general, it appears that the SVM and DT models are more sensitive to the length of the training period, while the MLR and GPR models are less so.

We conducted a similar analysis by grouping basins into four climate categories, revealing that the methods' performance varies across different climatic regions and training period lengths. We found that the performance of all models generally improves as the training period lengthens for all climate categories. However, the rate of improvement varies for each model and climate category. For the arid to hyper-arid category, the SVM and DT models appear to be the most sensitive to the length of the training period, with their performance improving significantly as the training period lengthens. For the semi-arid category, the SVM model is again the most sensitive, followed by the DT model. For the dry sub-humid category, the SVM model shows the most considerable improvement with an increase in training period length, followed by the DT model. For the humid category, the SVM and DT models again appear to be the most sensitive to training period length. It is also noteworthy that the MLR model displays the least sensitivity to training period length for all climate categories. The GPR model also exhibits low sensitivity, especially for the humid and dry sub-humid categories.

3.5.2. Cross-Validation analysis

Cross-validation is a crucial technique in machine learning and data analysis, which plays a vital role in assessing model accuracy and robustness. In K-fold cross-validation, the data is
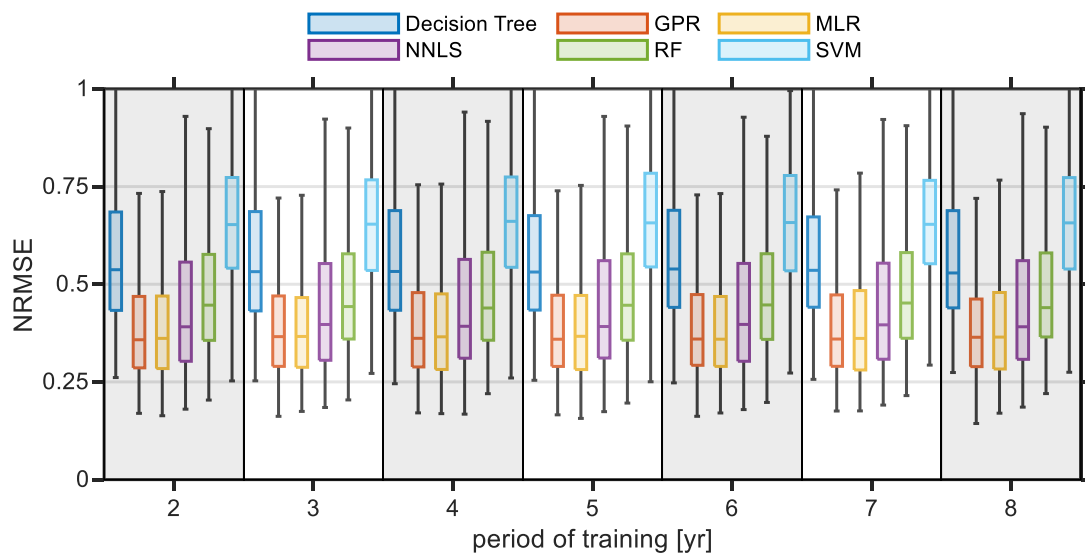
3

Figure 3.9: Boxplot comparing the NRMSE of six algorithms, including Decision Tree, Gaussian Process Regression (GPR), Multivariate Linear Regression (MLR), Non-negative Least Squares (NNLS), Random Forest (RF), and Support Vector Machine (SVM), for different training periods ranging from 2 to 8 years.

partitioned into K equally sized subsets or "folds". The model is trained on $K - 1$ of these folds and tested on the remaining fold. This process is repeated K times, with each fold being used exactly once as the testing data. The results from each iteration are then averaged to estimate the model's overall performance. K-fold cross-validation provides a more accurate estimate of the model's performance compared to simple partitioning of the data into a single training and testing set since it uses all the data for both training and testing. Moreover, K-fold cross-validation can also help identify whether the model is overfitting or underfitting the data. If the training error is much lower than the testing error, it suggests overfitting, which may result in poor generalization to new data. Conversely, if both the training error and testing error are high, it indicates underfitting, meaning that the model is unable to capture the underlying patterns in the data.

For a time series dataset like TWSA, where the order of the data points matters, it is important to use a variation of cross-validation called time series cross-validation. Time series cross-validation takes into account the temporal aspect of the data and ensures that the model is evaluated on data that it has not seen before in time. One common approach to time series cross-validation is to use a sliding window approach. In this approach, you split the time series data into multiple segments or folds, where each fold consists of a certain number of consecutive time points. You can then use each fold as the test set and the remaining data as the training set. In this study, GPR was used to estimate TWSA using various models as features, with GRACE data serving as the target variable. The number of folds to use in time series cross-validation depends on the length of the time series and the size of the segments you want to use. To evaluate the performance of the models, cross-validation was applied using K-folds equal to 10. The study utilized a time series data spanning 120 months from 2003 to 2012, which provided a sufficient duration for robust model evaluation.

The performance of the various models on the training and testing sets for the major river basins is presented in Figure 3.10 and Figure 3.11. The results demonstrate that most of the models exhibit good performance on both the training and testing sets, indicating that they are able to capture the underlying patterns in the data and generate accurate predictions on

new data. However, there are some discrepancies in the performance of the Decision Trees and SVM models between the training and testing sets, suggesting potential overfitting or underfitting issues. To address these issues, we employed L2 regularization, also known as weight decay, which is a technique that adds a penalty term to the loss function of the model to constrain the weights of the model during training (Cortes et al., 2012). By adding this penalty term, we encourage the model to produce simpler and smoother solutions that generalize better to new data, thereby reducing the potential for overfitting or underfitting. L2 regularization is particularly effective in controlling overfitting by shrinking the weights towards zero, which reduces their influence on the predictions and makes the model more robust.
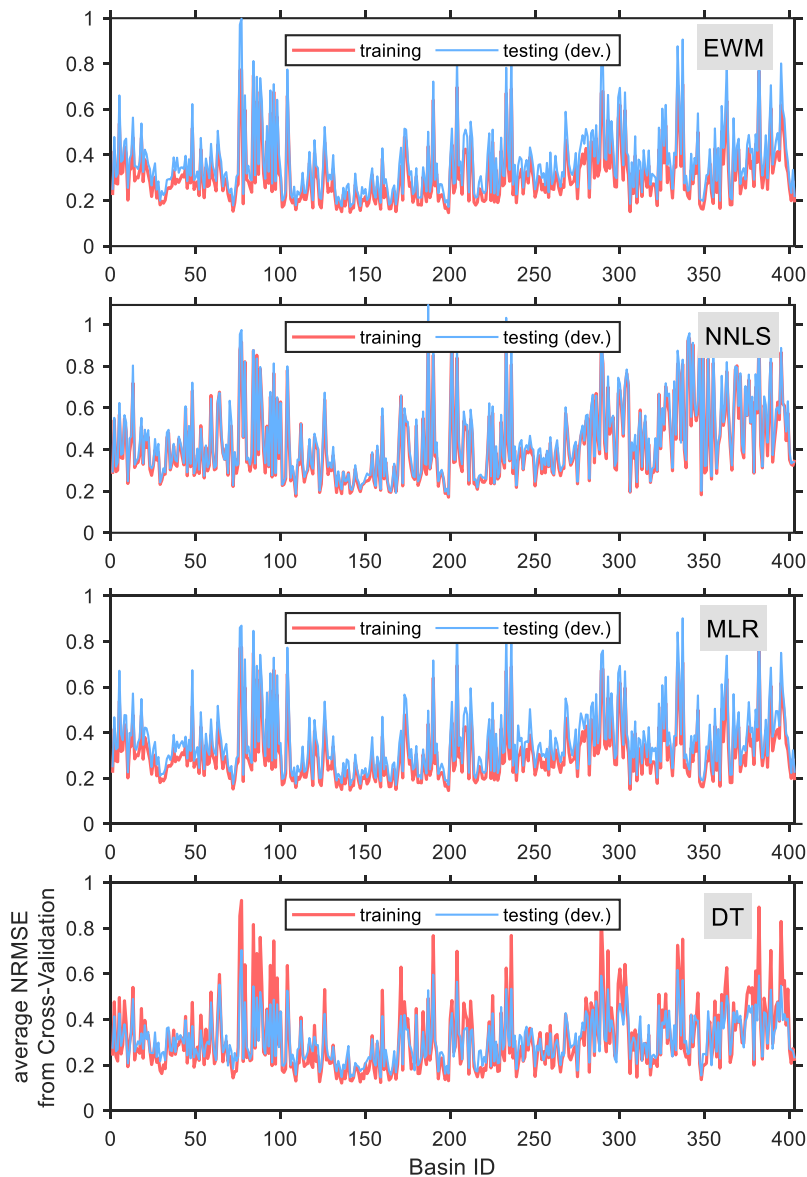

Figure 3.10: The figure shows the average NRMSE obtained from *K*-fold cross-validation for the training and testing sets of Ensemble Weighted Mean (EWM), Non-Negative Least Squares (NNLS), Multivariate Linear Regression (MLR), and Decision Tree (DT) over the major river basins.
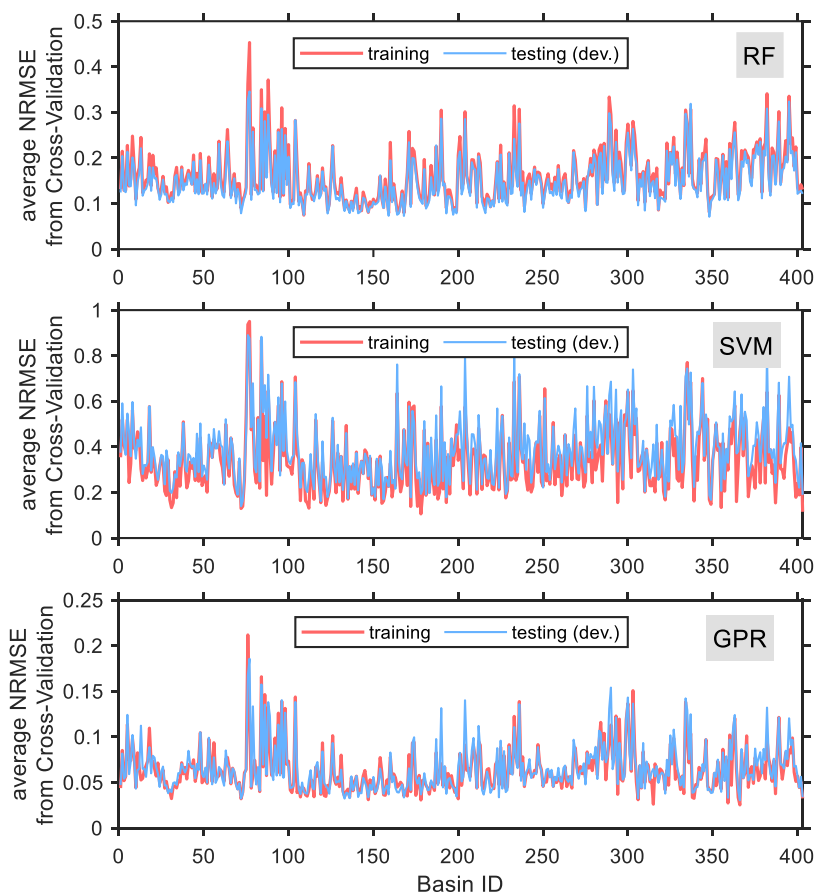
Figure 3.11: The figure shows the average NRMSE obtained from K-fold cross-validation for the training and testing sets of Random Forest (RF), Support Vector Machine (SVM), and Gaussian Process Regression (GPR) over the major river basins.

3.5.3. Assessing the performance of kernels

To derive a quantitative measure of the Kernels performance variability, we have computed the Coefficient of Variation (CV) of TWSA estimates from all kernels at each time step. Figure 3.12 portrays the mean coefficient of variation from various kernels, both from GPR and SVM, across major river basins during the period spanning April 2002 to December 2012. Overall, the results suggest that GPR and SVM are relatively insensitive to the choice of kernels, as evidenced by the similar CV values across different kernels. This suggests that the choice of kernel may not be critical for achieving accurate TWSA estimates using these methods. However, the sensitivity of GPR and SVM to kernel choice does appear to vary somewhat across different climate categories. For example, in arid to hyper-arid regions, both GPR and SVM exhibit relatively low CV values across all kernel types, suggesting that these methods are relatively robust in these regions. In contrast, in semi-arid regions, both GPR and SVM exhibit higher CV values, suggesting that the choice of kernel may have a greater impact on TWSA estimates in these regions. There also appears to be some differences in the sensitivity of GPR and SVM to kernel choice. For example, in humid regions, GPR exhibits a higher mean CV value compared to SVM, suggesting that GPR may be more sensitive to the choice of kernel in these regions.

3.5.4. Comparison with SLR

Satellite Laser Ranging (SLR) is a powerful geodetic technique that can be used to measure the precise range between a satellite and ground-based receivers using laser beams. SLR observa-
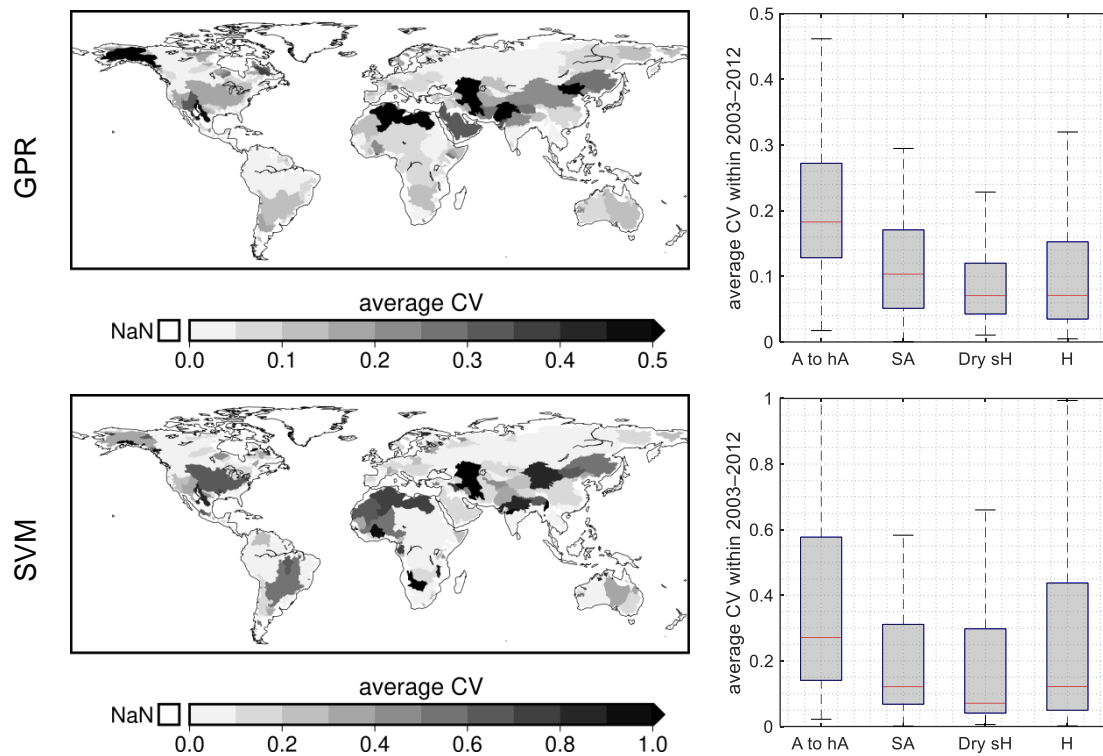
Figure 3.12: Performance comparison of GPR and SVM using various kernels and optimized hyperparameters. The averaged Coefficient of Variation (CV) is shown for each kernel over major river basins. Additionally, boxplots of the average CVs among the four climate categories are shown for each method.

tions can be used to infer changes in the Earth's gravitational field, which are primarily caused by the redistribution of water, ice, and atmospheric mass. SLR provides a valuable method to estimate temporal gravity fields prior to the GRACE mission. While SLR solutions contain more noise in higher degrees compared to GRACE solutions, they allow for high-resolution estimates of temporal gravity fields in the pre-GRACE period. This approach can be used to validate the performance of models that estimate TWSA before the launch of GRACE. Löcher and Kusche (2021) developed a new hybrid approach that combines SLR observations with the leading GRACE empirical orthogonal functions to produce high-resolution temporal gravity fields for the period between 1992 and 2020. This approach allows for high-resolution esti- mates of temporal gravity fields with up to 60° resolution, enabling us to estimate TWSA vari- ations in large basins, such as the Amazon, with high accuracy. The data can be downloaded from https://www.apmg.uni-bonn.de/daten-und-modelle/slr. We followed the approach uti- lized by Li et al. (2021) and compared the results from SLR first with GRACE. To this end, we found the *reliable* grid cells that was introduced in Li et al. (2021) and also reliable basins by comparing the time series of the TWSA from SLR with the GRACE observation (Figure 3.13(a) and Figure 3.13(b)). Considering the criterian of strong correlation, in our case larger than 0.8, we selected 112 basins out of 403 major river basins (Figure 3.13(c)).

The boxplot of correlation values (Figure 3.14) over the reliable basins shows that the mod- els have varying performance across different regions, with some regions showing consistently high correlation values and others showing more variability. All the models are able to cap- ture the variability in the observed TWSA, as evidenced by the moderate to high correlation coefficients obtained for both time periods. The MLR, NNLS, DT, RF, SVM, and GPR models all perform similarly, with average correlation coefficients ranging from 0.80 to 0.86, indicating their effectiveness in capturing the dynamics of TWSA. Interestingly, the ensemble mean and

ensemble weighted mean methods perform similarly to the individual models, suggesting that introducing weights may not necessarily lead to better performance. Additionally, it is noteworthy that the GRACE satellite mission, which directly measures changes in TWSA, outperforms all the models, with a correlation coefficient of 0.87 for the GRACE time period.
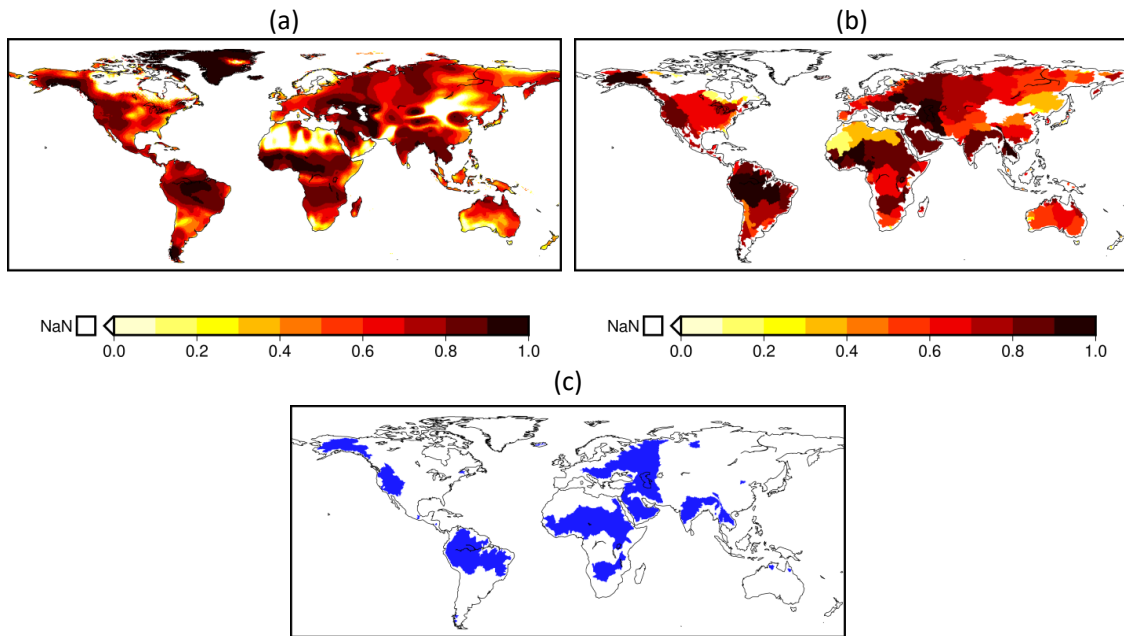

Figure 3.13: (a) Pixel-wise and (b) basinwise correlations between Satellite Laser Ranging (SLR) TWSA and GRACE(-FO) TWSA for the period from April 2002 to the end of 2021. (c) Basins with reliable correlation coefficient ($r \geq 0.8$) are highlighted in blue.
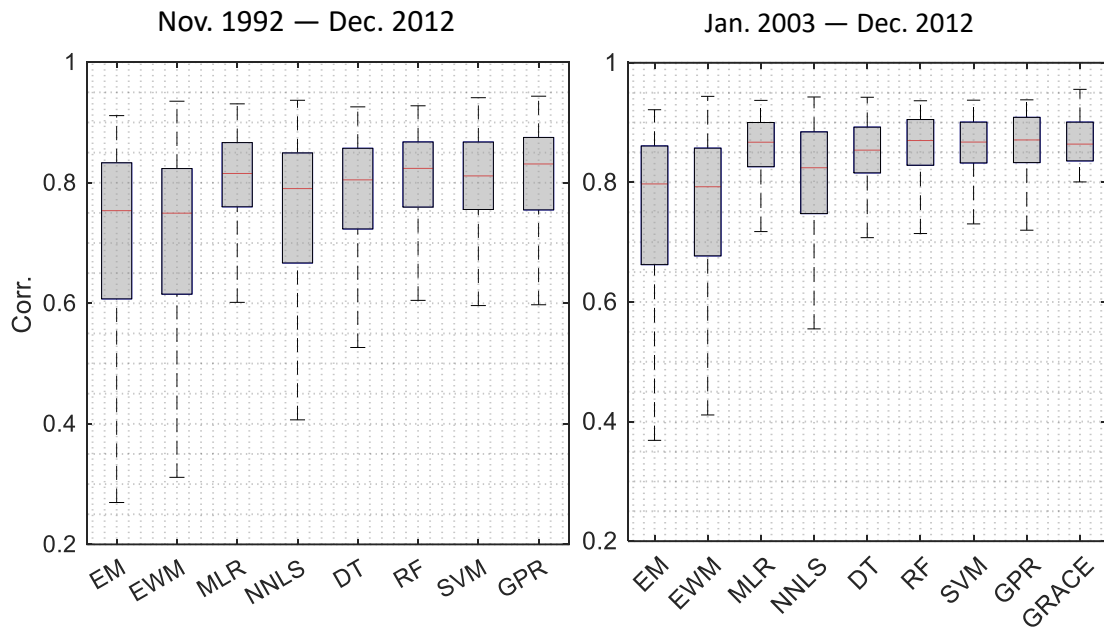

Figure 3.14: Boxplots of correlation coefficients between SLR TWSA and the results of various models over selected (reliable) river basins. The left plot shows the correlation between Nov. 1992 to Dec. 2012, while the right plot shows the correlation between Jan. 2003 to Dec. 2012, with the boxplot for the comparison between GRACE and SLR also included.

3.5.5. Comparison with GRACE

We compared the performance of various methods against GRACE in reconstructing TWSA over major river basins from 2003-2012. To assess the efficacy of each method, we utilized three metrics: KGE, NRMSE, and Mean Bias Error (MBE). KGE provides an overview of overall performance, while NRMSE quantifies the difference between the GRACE TWSA, taken as truth, and the reconstructed TWSA. Finally, MBE indicates the overall direction of the error, whether overestimation or underestimation. The results for KGE, NRMSE, and MBE over the major river basins, excluding Greenland and Antarctica, are depicted in Figures 3.15, 3.16, and 3.17, respectively.

Table 3.3 presents the results of comparing the performance of different methods with GRACE for reconstructing TWSA across all basins and specific basin types, namely arid to hyper-arid, semi-arid, dry sub-humid, and humid. The table shows the KGE, NRMSE, and MBE values for eight different methods, namely EM, EWM, MLR, NNLS, DT, RF, SVM, and GPR. We found that MLR and GPR showed the best overall performance in KGE and NRMSE, respectively. However, GPR consistently shows the best performance in most of the basins, except for arid to hyper-arid regions where MLR showed better performance in KGE. SVM and DT has poor performance in most of the basins.

The results of the comparison show that the performance of the different methods varies depending on the climate conditions of the region being analyzed. In the arid to hyper-arid regions, the MLR and NNLS methods show comparable results with the GPR method, while in semi-arid regions, the MLR and NNLS methods perform poorly. In the humid regions, the GPR and DT methods perform the best, while the MLR and NNLS methods perform poorly. It is also interesting to note that the Ensemble Mean (EM) method, which is the simplest and most naive method used as a baseline, performs poorly in all climate regions compared to the other methods. This suggests that the more advanced and sophisticated methods used in this study are necessary to accurately reconstruct TWSA in different regions and climates. Overall, the results suggest that the choice of method for reconstructing TWSA should be made with consideration of the specific climate conditions of the region being analyzed. The MLR method may be a good choice for dry sub-humid regions, while the GPR method may be a good choice for most regions.

3

Table 3.3: Performance comparison of various approaches for hindcasting GRACE in terms of three metrics (KGE, NRMSE, and MBE) against GRACE TWSA over major river basins. The performance evaluation is conducted separately for different climatic categories including arid, semi-arid, dry sub-humid, and humid. The shaded cells represent the top three performance within each column for each metric and category while the darker the color, the better the performance.

|	Metric	EM	EWM	MLR	NNLS	DT	RF	SVM	GPR
all basins	KGE	−6.46	−5.45	0.92	−2.64	0.90	0.38	−0.07	0.94
	NRMSE	0.60	0.58	0.31	0.43	0.32	0.15	0.31	0.19
	MBE	1.03	0.79	0.00	−0.14	0.00	−0.08	0.19	0.00
arid	KGE	−2.48	0.06	0.93	0.46	0.93	0.89	0.83	0.94
	NRMSE	0.58	0.50	0.30	0.38	0.29	0.15	0.28	0.20
	MBE	3.21	1.00	0.00	0.20	0.00	−0.11	0.27	0.00
semi-arid	KGE	−17.67	−14.26	0.91	−6.55	0.89	−0.24	−0.94	0.92
	NRMSE	0.60	0.57	0.32	0.42	0.32	0.16	0.31	0.21
	MBE	1.82	1.06	0.00	−1.12	0.00	−0.09	0.57	0.00
dry sub-humid	KGE	−9.35	−7.19	0.96	−2.98	0.96	0.48	−0.97	0.97
	NRMSE	0.42	0.40	0.21	0.28	0.21	0.11	0.28	0.15
	MBE	5.75	6.29	0.00	2.90	0.00	0.11	0.14	0.00
humid	KGE	−2.10	−2.50	0.91	−1.51	0.89	0.53	0.31	0.94
	NRMSE	0.65	0.62	0.32	0.46	0.34	0.16	0.32	0.19
	MBE	−0.49	−0.36	0.00	−0.37	0.00	−0.10	0.03	0.00
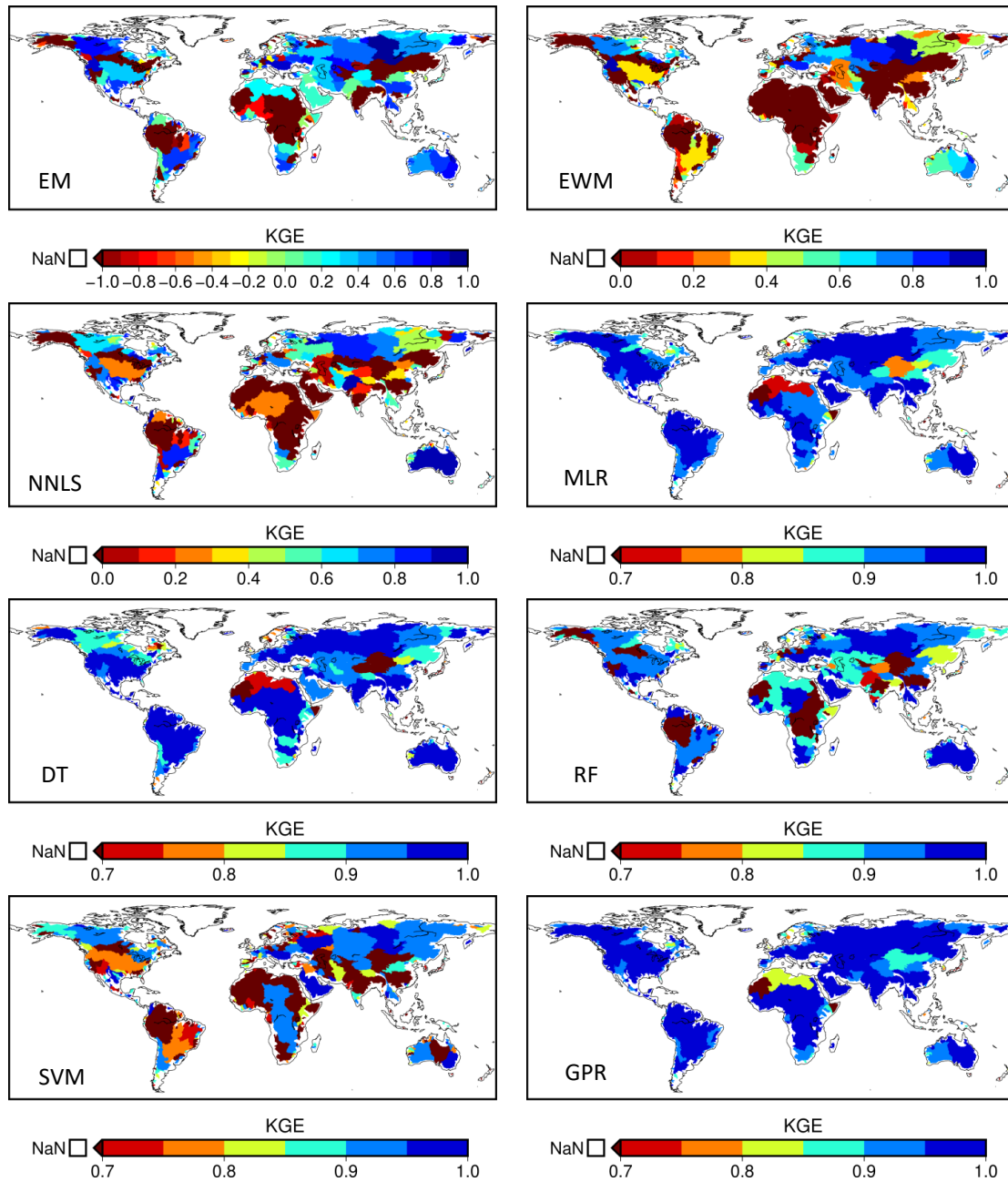
Figure 3.15: Global distribution of Kling-Gupta Efficiency (KGE) values for major river basins (excluding Greenland and Antarctica) obtained from the comparison between the reconstructed TWSA from various methods and GRACE during 2003–2012. Higher KGE values indicate better agreement between the reconstructed TWSA and GRACE measurements.
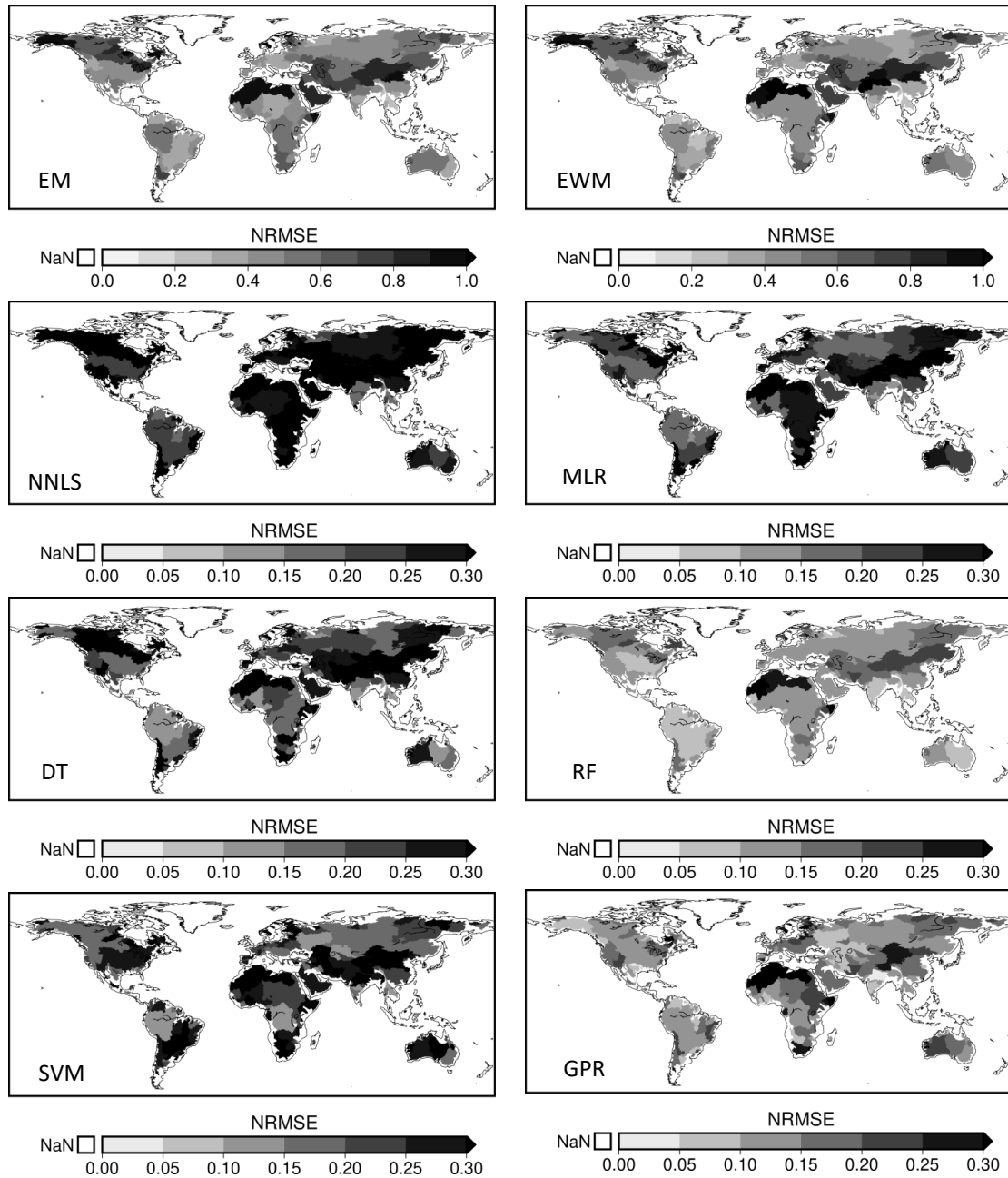
Figure 3.16: Global distribution of NRMSE values for major river basins (excluding Greenland and Antarctica) obtained from the comparison between the reconstructed TWSA from various methods and GRACE during 2003–2012. Lower NRMSE values indicate better agreement between the reconstructed TWSA and GRACE measurements.
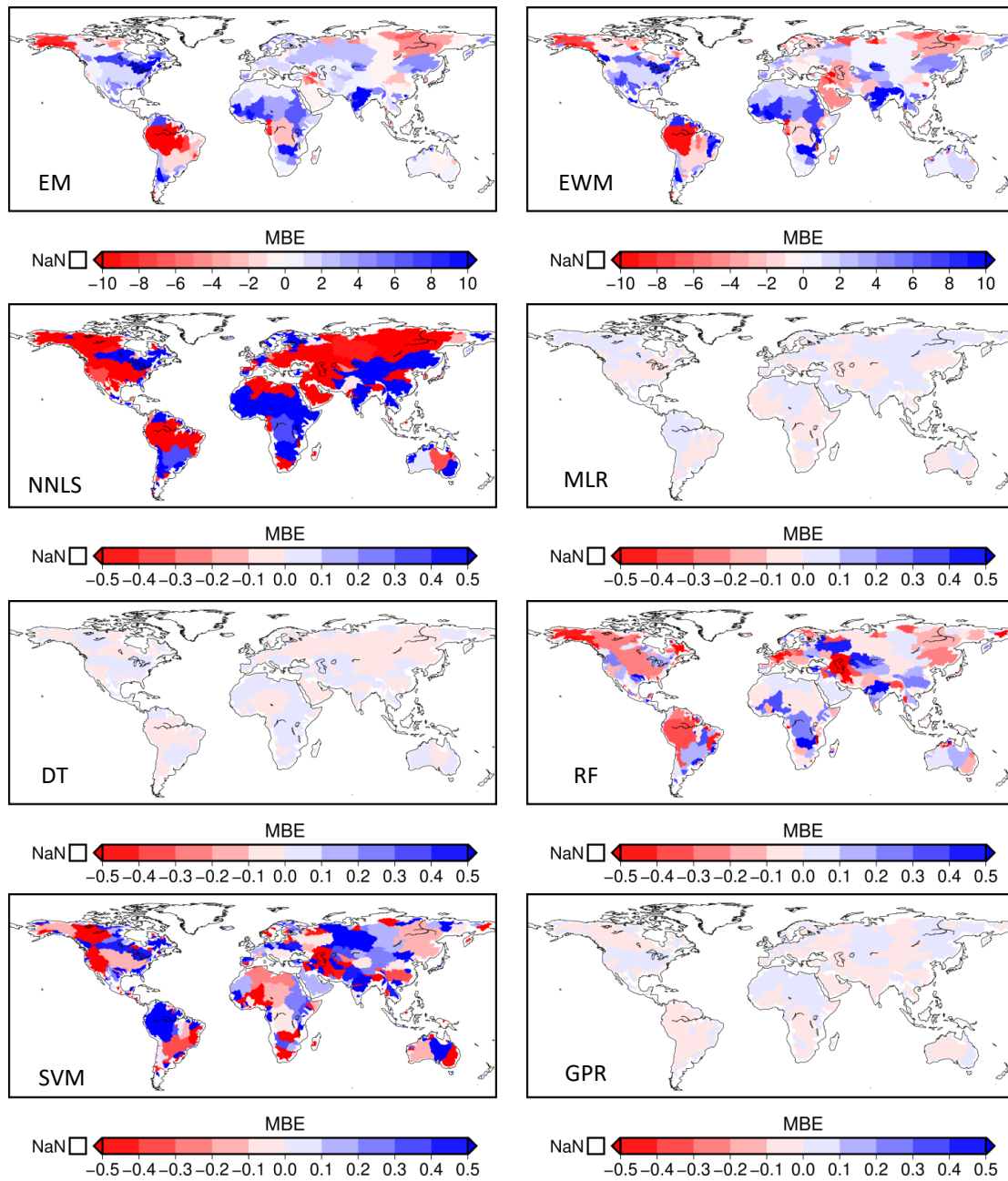
Figure 3.17: Global distribution of Mean Bias Error (MBE) values for major river basins (excluding Greenland and Antarctica) obtained from the comparison between the reconstructed TWSA from various methods and GRACE during 2003–2012. Positive values shows over estimation and negative values indicate underestimation.

Bibliography

Abramowitz, G., Leuning, R., Clark, M., & Pitman, A. (2008). Evaluating the performance of land surface models. *Journal of Climate*, *21*(21), 5468–5481. https://doi.org/10.1175/2008JCLI2378.1

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., et al. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389–407. https://doi.org/10.5194/hess-19-389-2015

Bengtsson, L., et al. (1982). FGGE 4-dimensional data assimilation at ECMWF.

Best, M., Pryor, M., Clark, D., Rooney, G., Essery, R., Ménard, C., Edwards, J., Hendry, M., Porson, A., Gedney, N., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 1: energy and water fluxes. *Geoscientific Model Development*, *4*(3), 677–699. https://doi.org/10.5194/gmd-4-677-2011

Clark, D., Mercado, L., Sitch, S., Jones, C., Gedney, N., Best, M., Pryor, M., Rooney, G., Essery, R., Blyth, E., et al. (2011). The Joint UK Land Environment Simulator (JULES), model description–Part 2: carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, *4*(3), 701–722. https://doi.org/10.5194/gmd-4-701-2011

Cortes, C., Mohri, M., & Rostamizadeh, A. (2012). L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*. https://doi.org/10.48550/arXiv.1205.2653

Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., Hersbach, H., & Buontempo, C. (2020). WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, *12*(3), 2097–2120. https://doi.org/10.5194/essd-12-2097-2020

Deardorff, J. W. (1978). Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation. *Journal of Geophysical Research: Oceans*, *83*(C4), 1889–1903. https://doi.org/10.1029/JC083iC04p01889

Decharme, B., Martin, E., & Faroux, S. (2013). Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *Journal of Geophysical Research: Atmospheres*, *118*(14), 7819–7834. https://doi.org/10.1002/jgrd.50631

Dee, D., Fasullo, J., Shea, D., Walsh, J., & NCAR, S. (2016). The climate data guide: Atmospheric reanalysis: Overview & comparison tables. *National Center for Atmospheric Research, Boulder, CO). Available at https://climatedataguide. ucar. edu/climatedata/atmospheric-reanalysis-overview-comparison-tables. Accessed June*, *1*, 2017.

de Teixeira da Encarnação, J., Visser, P., Arnold, D., Bezdek, A., Doornbos, E., Ellmer, M., Guo, J., van den IJssel, J., Iorfida, E., Jäggi, A., et al. (2019). Multi-approach gravity field models from Swarm GPS data. *Earth Syst. Sci. Data Discuss*. https://doi.org/10.5194/essd-2019-158

Döll, P., Douville, H., Güntner, A., Müller Schmied, H., & Wada, Y. (2016). Modelling freshwater resources at the global scale: Challenges and prospects. *Surveys in Geophysics*, *37*(2), 195–221. https://doi.org/10.1007/s10712-015-9343-1

Eicker, A., Schumacher, M., Kusche, J., Döll, P., & Schmied, H. M. (2014). Calibration/data assimilation approach for integrating GRACE data into the WaterGAP Global Hydrology Model (WGHM) using an ensemble Kalman filter: First results. *Surveys in Geophysics*, *35*(6), 1285–1309. https://doi.org/10.1007/s10712-014-9309-8

Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. *Journal of Advances in Modeling Earth Systems, 12*(4), e2018MS001453. https://doi.org/10.1029/2018MS001453

Frappart, F., & Ramillien, G. (2018). Monitoring groundwater storage changes using the Gravity Recovery and Climate Experiment (GRACE) satellite mission: A review. *Remote Sensing, 10*(6), 829. https://doi.org/10.3390/rs10060829

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of climate, 30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Gibson, J. (1997). *ECMWF Re-Analysis Project report series: ERA description* (Vol. 1). European Centre for Medium-Range Weather Forecasts. https://doi.org/10.1175/1520-0477-63.1.29.

Goosse, H. (2015). *Climate system dynamics and modeling*. Cambridge University Press.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hoch, J. M., Sutanudjaja, E. H., Wanders, N., van Beek, R., & Bierkens, M. F. (2022). Hyper-resolution PCR-GLOBWB: opportunities and challenges of refining model spatial resolution to 1 km over the European continent. *EGUsphere*, 1–29. https://doi.org/10.5194/egusphere-2022-111

Horwath, M., Gutknecht, B. D., Cazenave, A., Palanisamy, H. K., Marti, F., Paul, F., Le Bris, R., Hogg, A. E., Otosaka, I., Shepherd, A., et al. (2022). Global sea-level budget and ocean-mass budget, with a focus on advanced data products and uncertainty characterisation. *Earth System Science Data, 14*(2), 411–447. https://doi.org/10.5194/essd-14-411-2022

Humphrey, V., Gudmundsson, L., & Seneviratne, S. I. (2017). A global reconstruction of climate-driven subdecadal water storage variability. *Geophysical Research Letters, 44*(5), 2300–2309. https://doi.org/10.1002/2017GL072564

Jensen, L., Eicker, A., Dobslaw, H., Stacke, T., & Humphrey, V. (2019). Long-term wetting and drying trends in land water storage derived from GRACE and CMIP5 models. *Journal of Geophysical Research: Atmospheres, 124*(17-18), 9808–9823. https://doi.org/10.1029/2018JD029989

Jing, W., Zhang, P., Zhao, X., Yang, Y., Jiang, H., Xu, J., Yang, J., & Li, Y. (2020). Extending GRACE terrestrial water storage anomalies by combining the random forest regression and a spatially moving window structure. *Journal of Hydrology, 590*, 125239. https://doi.org/10.1016/j.jhydrol.2020.125239

Khaki, M., Hoteit, I., Kuhn, M., Awange, J., Forootan, E., Van Dijk, A. I., Schumacher, M., & Pattiaratchi, C. (2017). Assessing sequential data assimilation techniques for integrating GRACE data into a hydrological model. *Advances in Water Resources, 107*, 301–316. https://doi.org/10.1016/j.advwatres.2017.07.001

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II, 93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Koster, R. D., & Mahanama, S. P. (2012). Land surface controls on hydroclimatic means and variability. *Journal of Hydrometeorology, 13*(5), 1604–1620. https://doi.org/10.1175/JHM-D-12-050.1

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. (2016). Assimilation of gridded GRACE terrestrial water storage estimates in the North American Land Data Assimilation System. *Journal of Hydrometeorology, 17*(7), 1951–1972. https://doi.org/10.1175/JHM-D-15-0157.1

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. (2019). The Community Land Model version 5: Description of new features, benchmarking, and impact of forcing uncertainty. *Journal of Advances in Modeling Earth Systems, 11*(12), 4245–4287. https://doi.org/10.1029/2018MS001583

Li, F., Kusche, J., Chao, N., Wang, Z., & Löcher, A. (2021). Long-Term (1979-Present) total water storage anomalies over the global land derived by reconstructing GRACE data. *Geophysical Research Letters, 48*(8), e2021GL093492. https://doi.org/10.1029/2021GL093492

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology, 201*(1-4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Liu, B., Zou, X., Yi, S., Sneeuw, N., Li, J., & Cai, J. (2022). Reconstructing GRACE-like time series of high mountain glacier mass anomalies. *Remote Sensing of Environment, 280*, 113177. https://doi.org/10.1016/j.rse.2022.113177

Löcher, A., & Kusche, J. (2021). A hybrid approach for recovering high-resolution temporal gravity fields from satellite laser ranging. *Journal of Geodesy, 95*(1), 6. https://doi.org/10.1007/s00190-020-01460-x

Lück, C., Kusche, J., Rietbroek, R., & Löcher, A. (2018). Time-variable gravity fields and ocean mass change from 37 months of kinematic Swarm orbits. *Solid Earth, 9*(2), 323–339. https://doi.org/10.5194/se-9-323-2018

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E., Franz, T. E., et al. (2017). The future of Earth observation in hydrology. *Hydrology and earth system sciences, 21*(7), 3879–3914.

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. (2021). The global water resources and use model WaterGAP v2. 2d: Model description and evaluation. *Geoscientific Model Development, 14*(2), 1037–1079. https://doi.org/10.5194/gmd-14-1037-2021

Muñoz-Sabater, J., Dutra, E., Agustı́-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data, 13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Oleson, K., Lawrence, D., Lombardozzi, D., & Wieder, D. (2019). CLM land-only release [Accessed: 2021-05-30]. https://doi.org/https://doi.org/10.5065/d6154fwh

Olsen, N., Friis-Christensen, E., Floberghagen, R., Alken, P., Beggan, C. D., Chulliat, A., Doornbos, E., Da Encarnação, J. T., Hamilton, B., Hulot, G., et al. (2013). The Swarm satellite constellation application and research facility (SCARF) and Swarm data products. *Earth, Planets and Space, 65*(11), 1189–1200. https://doi.org/10.5047/eps.2013.07.001

Orth, R., & Seneviratne, S. I. (2013). Predictability of soil moisture and streamflow on subseasonal timescales: A case study. *Journal of Geophysical Research: Atmospheres, 118*(19), 10–963. https://doi.org/10.1002/jgrd.50846,2013

Overgaard, J., Rosbjerg, D., & Butts, M. (2006). Land-surface modelling in hydrological perspective– a review. *Biogeosciences, 3*(2), 229–241. https://doi.org/10.5194/bg-3-229-2006

Polcher, J., Bertrand, N., Biemans, H., Clark, D. B., Floerke, M., Gedney, N., Gerten, D., Stacke, T., van Vliet, M., & Voss, F. (2011). Improvements in hydrological processes in general hydrological models and land surface models within WATCH.

Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., & Rodell, M. (2015). Quantifying renewable groundwater stress with GRACE. *Water resources research*, *51*(7), 5217–5238. https://doi.org/10.1002/2015WR017349

Richter, H. M. P., Lück, C., Klos, A., Sideris, M. G., Rangelova, E., & Kusche, J. (2021). Reconstructing GRACE-type time-variable gravity from the Swarm satellites. *Scientific reports*, *11*(1), 1–14. https://doi.org/10.1038/s41598-020-80752-w

Rietbroek, R., Fritsche, M., Dahle, C., Brunnabend, S.-E., Behnisch, M., Kusche, J., Flechtner, F., Schröter, J., & Dietrich, R. (2014). Can GPS-derived surface loading bridge a GRACE mission gap? *Surveys in Geophysics*, *35*(6), 1267–1283. https://doi.org/10.1007/s10712-013-9276-5

Saha, S., Nadiga, S., Thiaw, C., Wang, J., Wang, W., Zhang, Q., Van den Dool, H., Pan, H.-L., Moorthi, S., Behringer, D., et al. (2006). The NCEP climate forecast system. *Journal of Climate*, *19*(15), 3483–3517. https://doi.org/10.1175/JCLI3812.1

Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., Behringer, D., Hou, Y.-T., Chuang, H.-y., Iredell, M., et al. (2014). The NCEP climate forecast system version 2. *Journal of climate*, *27*(6), 2185–2208. https://doi.org/10.1175/JCLI-D-12-00823.1

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. (2018). Global models underestimate large decadal declining and rising water storage trends relative to GRACE satellite data. *Proceedings of the National Academy of Sciences*, *115*(6), E1080–E1089. https://doi.org/10.1073/pnas.1704665115

Schewe, J., Gosling, S. N., Reyer, C., Zhao, F., Ciais, P., Elliott, J., Francois, L., Huber, V., Lotze, H. K., Seneviratne, S. I., et al. (2019). State-of-the-art global models underestimate impacts from climate extremes. *Nature communications*, *10*(1), 1–14. https://doi.org/10.1038/s41467-019-08745-6

Schumacher, M., Forootan, E., van Dijk, A. I., Schmied, H. M., Crosbie, R. S., Kusche, J., & Döll, P. (2018). Improving drought simulations within the Murray-Darling Basin by combined calibration/assimilation of GRACE data into the WaterGAP Global Hydrology Model. *Remote Sensing of Environment*, *204*, 212–228. https://doi.org/10.1016/j.rse.2017.10.029

Sood, A., & Smakhtin, V. (2015). Global hydrological models: A review. *Hydrological Sciences Journal*, *60*(4), 549–565. https://doi.org/10.1080/02626667.2014.950580

Stöckli, R., Lawrence, D., Niu, G.-Y., Oleson, K., Thornton, P. E., Yang, Z.-L., Bonan, G., Denning, A., & Running, S. W. (2008). Use of FLUXNET in the Community Land Model development. *Journal of Geophysical Research: Biogeosciences*, *113*(G1). https://doi.org/10.1029/2007JG000562

Sun, A., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: can we learn from mismatch? *Water Resources Research*, *55*(2), 1179–1195. https://doi.org/10.1029/2018WR023333

Sun, Long, D., Yang, W., Li, X., & Pan, Y. (2020). Reconstruction of GRACE data on changes in total water storage over the global land surface and 60 basins. *Water Resources Research*, *56*(4), e2019WR026250. https://doi.org/10.1029/2019WR026250

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., Van Der Ent, R. J., De Graaf, I. E., Hoch, J. M., De Jong, K., et al. (2018a). PCR-GLOBWB 2: a 5 arcmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Sutanudjaja, E. H., Van Beek, R., Wanders, N., Wada, Y., Bosmans, J. H., Drost, N., Van Der Ent, R. J., De Graaf, I. E., Hoch, J. M., De Jong, K., et al. (2018b). PCR-GLOBWB 2: a 5 ar-

cmin global hydrological and water resources model. *Geoscientific Model Development*, *11*(6), 2429–2453. https://doi.org/10.5194/gmd-11-2429-2018

Talpe, M. J., Nerem, R. S., Forootan, E., Schmidt, M., Lemoine, F. G., Enderlin, E. M., & Landerer, F. W. (2017). Ice mass change in Greenland and Antarctica between 1993 and 2013 from satellite gravity measurements. *Journal of Geodesy*, *91*(11), 1283–1298. https://doi.org/10.1007/s00190-017-1025-y

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical research letters*, *31*(9). https://doi.org/10.1029/2004GL019920

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. (2019). Contributions of GRACE to understanding climate change. *Nature climate change*, *9*(5), 358–369. https://doi.org/10.1038/s41558-019-0456-2

Trenberth, K. E., & Trenberth, K. E. (1992). *Climate system modeling*. Cambridge University Press.

Uppala, S. M., Kållberg, P., Simmons, A. J., Andrae, U., Bechtold, V. D. C., Fiorino, M., Gibson, J., Haseler, J., Hernandez, A., Kelly, G., et al. (2005). The ERA-40 re-analysis. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(612), 2961–3012. https://doi.org/10.1256/qj.04.176

Van Beek, L., Wada, Y., & Bierkens, M. F. (2011). Global monthly water stress: 1. Water balance and water availability. *Water Resources Research*, *47*(7). https://doi.org/10.1029/2010WR009792

Van Der Knijff, J., Younis, J., & De Roo, A. (2010). LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science*, *24*(2), 189–212. https://doi.org/10.1080/13658810802549154

Van Dijk, A. (2010). The Australian Water Resources Assessment System Technical Report 3. Landscape Model (version 0.5) Technical Description. https://doi.org/10.4225/08/5852dd9bb578c

Wada, Y., Wisser, D., & Bierkens, M. F. (2014). Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources. *Earth System Dynamics*, *5*(1), 15–40. https://doi.org/10.5194/esd-5-15-2014

Wang, F., Chen, Y., Li, Z., Fang, G., Li, Y., Wang, X., Zhang, X., & Kayumba, P. M. (2021). Developing a Long Short-Term Memory (LSTM)-Based Model for Reconstructing Terrestrial Water Storage Variations from 1982 to 2016 in the Tarim River Basin, Northwest China. *Remote Sensing*, *13*(5), 889. https://doi.org/10.3390/rs13050889

Williams, M., Richardson, A. D., Reichstein, M., Stoy, P. C., Peylin, P., Verbeeck, H., Carvalhais, N., Jung, M., Hollinger, D. Y., Kattge, J., et al. (2009). Improving land surface models with FLUXNET data. *Biogeosciences*, *6*(7), 1341–1359. https://doi.org/10.5194/bg-6-1341-2009

Wiltshire, A. J., Duran Rojas, M. C., Edwards, J. M., Gedney, N., Harper, A. B., Hartley, A. J., Hendry, M. A., Robertson, E., & Smout-Day, K. (2020). JULES-GL7: the Global Land configuration of the Joint UK Land Environment Simulator version 7.0 and 7.2. *Geoscientific Model Development*, *13*(2), 483–505. https://doi.org/10.5194/gmd-13-483-2020

Wouters, B., van de Wal, R., et al. (2018). Global sea-level budget 1993–present. *Earth System Science Data*, *10*(3), 1551–1590. https://doi.org/10.3929/ethz-b-000287786

Yokohata, T., Kinoshita, T., Sakurai, G., Pokhrel, Y., Ito, A., Okada, M., Satoh, Y., Kato, E., Nitta, T., Fujimori, S., et al. (2019). MIROC-INTEG1: A global bio-geochemical land surface

model with human water management, crop growth, and land-use change. *Geoscientific Model Development Discussions, 2019*, 1–57.

Zaitchik, B. F., Rodell, M., & Reichle, R. H. (2008). Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *Journal of Hydrometeorology, 9*(3), 535–548. https://doi.org/10.1175/2007JHM951.1

Zhang, D., Zhang, Q., Werner, A. D., & Liu, X. (2016). GRACE-based hydrological drought evaluation of the Yangtze River Basin, China. *Journal of Hydrometeorology, 17*(3), 811–828. https://doi.org/10.1175/JHM-D-15-0084.1

Zhao, C., Huang, Y., Li, Z., & Chen, M. (2018). Drought monitoring of southwestern China using insufficient GRACE data for the long-term mean reference frame under global change. *Journal of Climate, 31*(17), 6897–6911. https://doi.org/10.1175/JCLI-D-17-0869.1

3

4

Analysis of Water Storage Variation

4.1. Introduction

The terrestrial water cycle consists of groundwater, surface water, soil moisture, canopy water, snow, and ice. The sum of the mentioned compartments is defined as the Total (Terrestrial) Water Storage (TWS). Despite its tendency to remain range-bound, the TWS has shown some significant negative trends (water loss), especially in regions with over-exploitation of the groundwater (Castellazzi et al., 2016; Joodaki et al., 2014; Voss et al., 2013) and where the ice sheets and glaciers are melting as a consequence of climate change (Khan et al., 2010; Velicogna et al., 2020). Moreover, positive trends (water gain) are also observed over vast regions of Africa and the Amazon while progressing from dry to wet periods (Huang et al., 2015; Seyoum, 2018; Syed et al., 2008). Accurate accounting of the TWS gain (or loss) is essential for sustainable water resource management, assessment of the water scarcity, ensuring the food supply cycle, and energy generation (Rodell et al., 2018).

GRACE(-FO) missions have delivered an unprecedented monthly measurement of the TWS anomaly at the continental and regional scales over the last two decades. GRACE(-FO) observations have been employed in several studies to quantify the TWS change together with its uncertainty at a global to regional scale (e.g., Joodaki et al., 2014; Pokhrel et al., 2021; Reager & Famiglietti, 2013; Rodell et al., 2018; Wang et al., 2018). Figure 4.1 shows the long-term trend of TWSA observed from GRACE(-FO) from 2003 to the end of 2021. To calculate the linear trend, we have used ordinary least squares regression and the NASA Goddard Space Flight Center (GSFC) mascon solution. Significant water storage loss has occurred over Greenland, the Middle East, Central Asia, the Indian sub-continent, the east and south of South America, western Canada, and states like California and Texas in the US. In contrast, regions like the Amazon, large parts of Africa (except for North Africa), and the Great Lakes basin have gained considerable water.
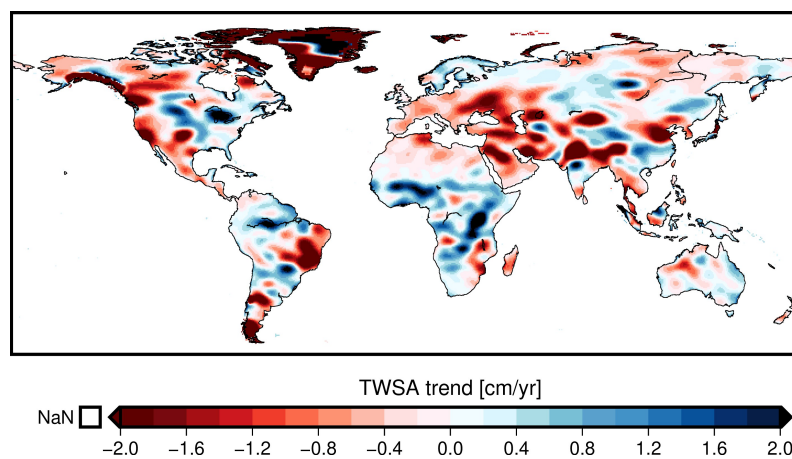

Figure 4.1: Annual trend of TWSA from 2003 to 2021. The mascon solution by the the NASA Goddard Space Flight Center (GSFC) has been used https://earth.gsfc.nasa.gov/geo/data/grace-mascons.

In this chapter we analyze the total water storage variation. According to the Merriam-Webster dictionary, *Analysis* is "a detailed examination of anything complex to understand its nature or to determine its essential features". Moreover, as its second meaning, *Analysis* is the act of separating a whole into its components. Following the mentioned definitions, i.e., to analyze the water storage variation, this chapter quantifies the TWS change from 2003 to the end of 2021. Moreover, we have analyzed the water storage change's three components: precipitation, evapotranspiration, and runoff. All the analyses are presented over the main sub-continents and the major river basins. Finally, further investigations are presented over Iran and its major river basins using auxiliary data like precipitation gauges and piezometric groundwater wells.

4.2. Study areas

In this chapter, we analyzed GRACE data in three different study areas, ranging from subcontinents to major river basins and finally over Iran, to demonstrate the versatility of using this data at different regional scales. By considering different regional scales, we were able to gain a better understanding of the spatial and temporal variability of water resources, and how they are impacted by factors such as climate change and human activities. This information is crucial for making informed decisions regarding water management and resource allocation, particularly in regions where water scarcity is a pressing issue.

4

4.2.1. Sub-continents

We have divided the global land into 16 sub-continental regions, excluding Greenland, Antarctica, Southeastern Asia, and Northern Asia (Figure 4.2). The boundaries are taken from the United Nations geo-scheme, devised by the United Nations Statistics Division (UNSD). We have deliberately excluded Antarctica, Greenland, Siberia, and the east and north of Russia as their TWSA signal includes significant contributions from phenomena like the permafrost or extensive glacier melting, which is beyond the scope of our study. Moreover, we have excluded southwestern Asia due to the considerable contamination of leaked signals from the ocean.
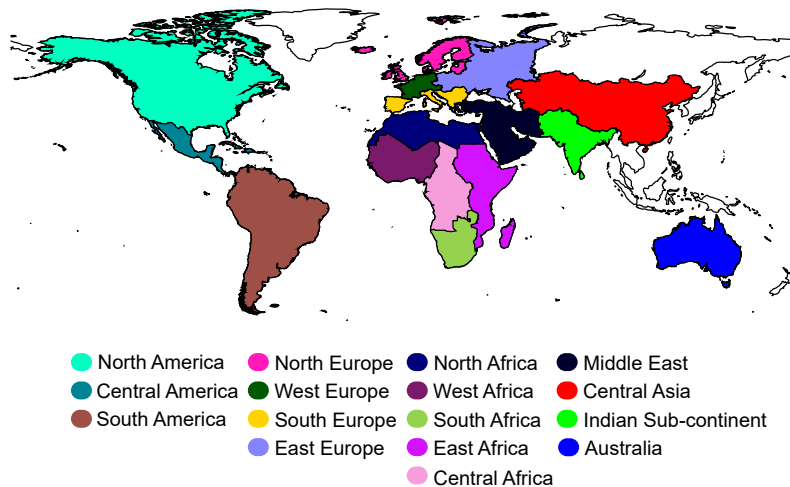

Figure 4.2: Global distribution of the sub-continents selected in this study.

4.2.2. Major river basins

To investigate the TWS loss (gain) at the basin scale, 405 major global river basins are selected. The distribution of the basins is shown in Figure 4.3. We use the categories introduced in section 2.4 to analyze the results, including the basins' size, climate, latitude zone, and the signal's RMS. For the characteristics of the basins in terms of the mentioned categories, please see Figure 2.6.
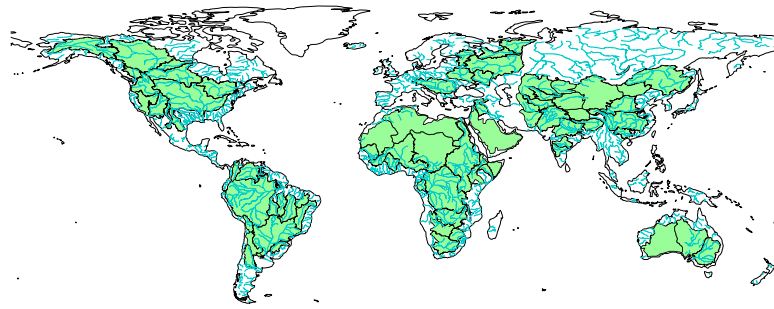
Figure 4.3: Global distribution of the selected basins.

4.2.3. Iran

Iran has an area of about 1.7 million km^2 and is located in the south-west of Asia (Figure 4.4). The country's main water bodies include the world's largest (by area) inland water body called the Caspian Sea in the north, the Persian Gulf and the Sea of Oman in the south, and Lake Urmia in the northwest. Two large mountain ranges cover 60 % of the area: the Alborz chain running from the northwest to the northeast along the southern edge of the Caspian Sea and the Zagros range, which runs from the northwest southward to the shores of the Persian Gulf. The central part of the country is covered by two large deserts: Dasht-e Kavir (77 600 km^2) and the Lut Desert (Dasht-e Lut) (51 800 km^2), which are the world's 24th and 25th-largest deserts. Iran is divided into six main water basins, which are subdivided into 30 major river basins. The characteristics of the river basins are listed in Table 4.1. The first digit in the basin's ID represents the number of 6 major basins in Iran, and the second indicates the sub-basins.
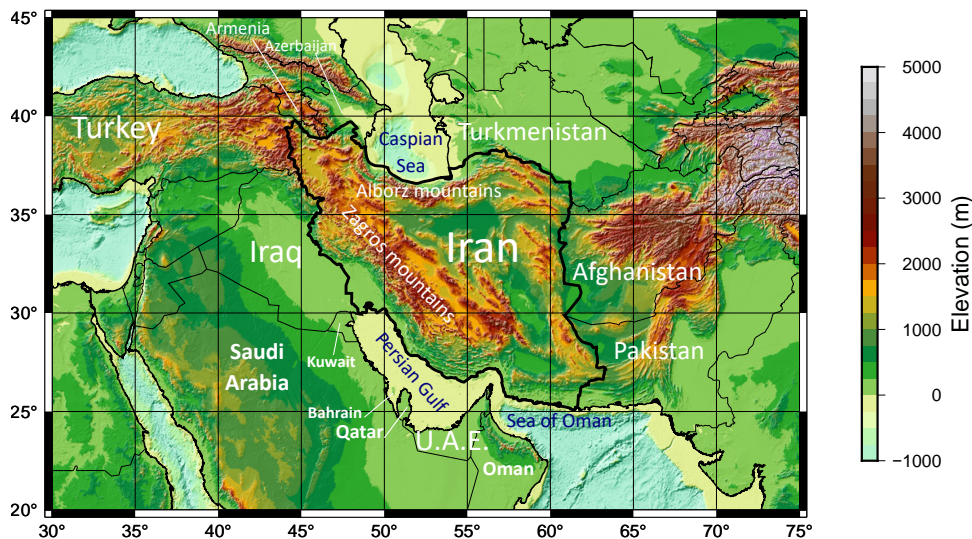
Figure 4.4: Elevation map of the Middle East including Iran.

In terms of climate, Iran is located in the subtropical high-pressure belt of the Earth. However, the variety of topographic regions, with heights varying from 25 m to 5600 m, has led to a wide range of climates across the country. Most of the country is arid (65 %) to semi-arid (20 %) with sweltering summers in the central and southern coastal regions. In contrast, only 15 % is humid, mainly in regions close to the Caspian Sea and partly in areas close to the Persian Gulf and Sea of Oman (Madani, 2014) (Figure 4.5 (a)). Using the data from Global Precipitation Climatology Centre (GPCC), the long-term (1960–2016) mean annual precipitation for the entire

country is around 225 mm (~ 370 km^3), while precipitation can be as low as 50 mm in deserts and exceed 1500 mm in the northern side of the Alborz Mountain range and the coastal areas of the Caspian Sea (Figure 4.5 (b)). In terms of annual precipitation, Iran ranked 158 among 189 countries over 1960–2016 period using GPCC as the reference dataset. Around 30 % of the total precipitation falls in the form of snow (Mousavi, 2005). However, this share seems to be declining over the last decade (Araghi & Mousavi-Baygi, 2020). About 70 % of the precipitation is lost through evaporation (Lehane, 2014).
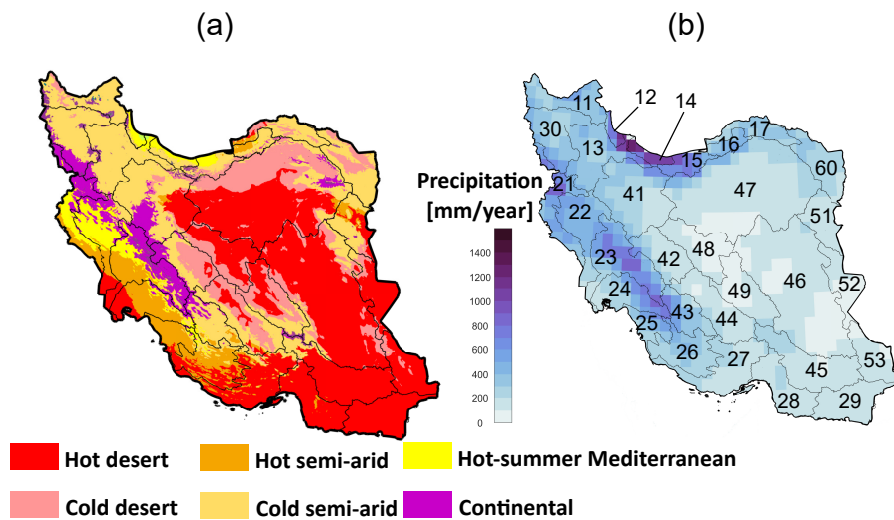

Figure 4.5: (a) Köppen-Geiger climate classification map (1980–2016) based on calculations by Beck et al. (2018), and (b) Spatial distribution of mean annual precipitation using long-term (1980–2016) precipitation data from GPCC. Borders and numbers inside the country represent the 30 major river basins in Iran, listed in Table 4.1.

Table 4.1: Iran's major river basins and their areas. Mean annual precipitation and discharge were calculated for each basin from ERA5 dataset for the period from 1983 to 2019.

ID	name	area $[10^3 \text{ km}^2]$	mean annual precipitation [cm/yr]	mean annual discharge [cm/yr]
11	Aras	41	16.5	6.5
12	Talesh	7	51.5	50.9
13	Sefidrood	60	26.5	5.5
14	Haraz-Sefidrood	11	27.0	15.5
15	Haraz	19	31.0	18.5
16	Gharesoo	13	15.5	3.5
17	Atrak River	27	17.0	1.5
21	West-border	39	32.5	9.0
22	Karkheh	50	30.5	7.0
23	Karun	64	40.5	23.0
24	Jarahi and Zohreh	38	23.5	9.5
25	Helle	20	21.0	3.5
26	Mand	44	18.5	1.5
27	Mehran-Kal	58	13.5	1.0
28	Bandar Abbas	41	9.5	0.5
29	South Baluchestan	44	10.5	0.2
30	Lake Urmia	53	23.0	0.0
41	Salt Lake	91	19.5	2.0
42	Gavkhuni	40	13.0	0.4
43	Tashk	29	20.5	2.0
44	Abarghoo-Sirjan	54	11.0	0.2
45	Hamun-Jazmurian	64	10.5	0.5
46	Lut Desert	195	9.5	0.3
47	Central Desert	224	13.5	0.8
48	Siahkooh	47	6.0	0.3
49	Saghand	48	11.5	0.8
51	Khaf	32	12.5	0.8
52	Hamun Hirmand	3.2	4.0	0.2
53	Hamun Mashkil	3.3	8.0	0.2
60	Ghareghoom	44	20.5	3.0

4.3. Data

4.3.1. Water balance fluxes

As expressed in chapter 1, the total water storage change can be determined from the terrestrial water balance equation:

$$\frac{\text{dTWSA}}{\text{d}t} = P - ET - R, \tag{4.1}$$

where $\frac{\text{dTWSA}}{\text{d}t}$ is the derivative of GRACE-TWSA and P, ET, and R are precipitation, evapotranspiration, and runoff, respectively. Due to the lack of one superior dataset which can deliver water balance fluxes with high accuracy on a global scale, we have determined the water fluxes using an ensemble mean of selected datasets. To this end, we have first compared the aggregated precipitation from all datasets over the Amazon basin using the GPCC version 2020 as the reference. GPCC is acquired from more than 85,000 stations worldwide, and several studies have already used GPCC as a reference to compare precipitation products globally and regionally (e.g., Becker et al., 2013; Sun et al., 2018). For evapotranspiration, we have used the Penman-Monteith-Leuning Evapotranspiration Version 2 (PML-V2) dataset as the reference, which has already shown the overall best global and regional performance among available gridded datasets (e.g., Chao et al., 2021; Elnashar et al., 2021; Zhang et al., 2019). Finally, we included datasets with various approaches, including hydrological, reanalysis, and atmospheric models for runoff. The datasets are compared with the global runoff reconstruction product, named Global RUNoff ENSEMBLE (G-RUN ENSEMBLE), derived by Ghiggi et al., 2021 from a machine learning approach and 21 different atmospheric forcing datasets and streamflow observations from the Global Streamflow Indices and Metadata Archive (GSIM).

For each flux, the datasets are assessed with respect to the reference dataset via three commonly used metrics, namely the Correlation Coefficient (CC), Root Mean Square Error (RMSE), and the Kling-Gupta Efficiency (KGE). For the full description of these metrics, please see the information provided in Appendix section B. We have selected datasets with KGE>0.5 (KGE>0.3 for runoff) and RMSE< 10 % of the mean annual. The final estimation for each flux was obtained from the ensemble mean of the selected datasets at each pixel. It should be noted that for all the fluxes, the ensemble of the selected datasets holds the spatial resolution of 0.5°. Daily data were aggregated to monthly values, and when necessary, data were interpolated to 0.5° resolution using the nearest-neighbor approach. Such an interpolation would not affect the ranking of datasets (Saemian et al., 2021).

4.3.2. Surface water extent

Surface water storage is a vital variable in the TWS change over a vast region of the world. With the advent of satellite imagery missions, we have started to monitor the surface water extent of the global water bodies, including lakes, rivers, and wetlands. Observations from missions like the Moderate Resolution Imaging Spectroradiometer (MODIS), Landsats, and Sentinel-2 have been used in several studies to extract water bodies' extent (Donchyts et al., 2016; Khandelwal et al., 2017; Klein et al., 2017; Pekel et al., 2016; Schwatke et al., 2020; Zhang & Gao, 2016). Four datasets are available which provide the time series of the lakes surface area, including Hydroweb (Crétaux et al., 2011), Database for Hydrological Time Series of Inland Waters (DAHITI) (Schwatke et al., 2019), the Bluedot observatory, and HydroSat (Tourian et al., 2021). Table 4.3 lists these datasets, including their links. In this study, we have employed the surface water area from HydroSat (Tourian et al., 2021). The dataset is developed based on the

Table 4.2: Summary of the datasets used in this study to compute the ensemble mean of the water balance fluxes.

	Dataset	Method/Source(s)	Spatial Resolution	Time period	Reference
Precipitation	AgCFSR	G, R	0.25° × 0.25°	1980–2010	Ruane et al., 2015
	AgMERRA	G, S, R	0.25° × 0.25°	1980–2010	Ruane et al., 2015
	CHIRPS	G, S, R	0.05° × 0.05°	1981–present	Funk et al., 2015
	CRUv4.04	G	0.5° × 0.5°	1901–2019	Harris et al., 2020
	GPCCv2020	G	0.25° × 0.25°	1982–2019	Schneider et al., 2020
	GPCPv2.3	G, S	2.5° × 2.5°	1979–present	**adler2018GPCPv2.3**; Adler et al., 2003
	GPCP 1DD	G, S	1.0° × 1.0°	1996–present	Huffman et al., 2001
	GPM IMERG v6 Final	G, S	0.1° × 0.1°	2000–present	Huffman et al., 2019
	MSWEP v2.8	G, S, R	0.1° × 0.1°	1979–2020	Beck et al., 2019
	PERSIANN-CDR	G, S	0.25° × 0.25°	1983–present	Ashouri et al., 2015
	PREC/L	G	0.5° × 0.5°	1948–present	Chen et al., 2002
	TRMM-3B42-adj	G, S	0.25° × 0.25°	1998–2019	Huffman et al., 2007
	UDELv5.01	G	0.5° × 0.5°	1900–2017	Willmott and Matsuura, 1995
ET	ERA5	R	31 km	1979–present	Hersbach et al., 2020
	FluxCom	R	31 km	1979–present	Jung et al., 2019
	P-LSH	R	31 km	1979–present	Zhang et al., 2010
	PML-v2	R	31 km	1979–present	Zhang et al., 2019
Runoff	G-RUN	R	31 km	1979–present	Ghiggi et al., 2021
	SURFEX-TRIP	R	31 km	1979–present	Lindström et al., 1997
	W3RA	R	31 km	1979–present	Van Dijk, 2010

Global Surface Water Dataset (GSWD) and has improved GSWD in terms of its temporal gaps and also biases (see Figure 8 and section 3 of (Tourian et al., 2021) for more details). HydroSat contains time series of 4980 lakes and reservoirs worldwide (Figure 4.6).

Table 4.3: List of sources for providing time series of surface water extent from satellite imagery (adapted from (Tourian et al., 2021))

| Product	operated by	source	Remark
Hydroweb	CNES	http://hydroweb.theia-land.fr	available for lakes
DAHITI	Deutsches Geodätisches Forschungsinstitut (DGFI)	https://dahiti.dgfi.tum.de	available for lakes
HydroSat	Insititute of Geodesy University of Stuttgart	http://hydrosat.gis.uni-stuttgart.de	available over rivers and lakes
Bluedot observatory	Copernicus, European commission ESA, USGS, Amazon Web Services	https://blue-dot-observatory.com	available for lakes and reservoirs
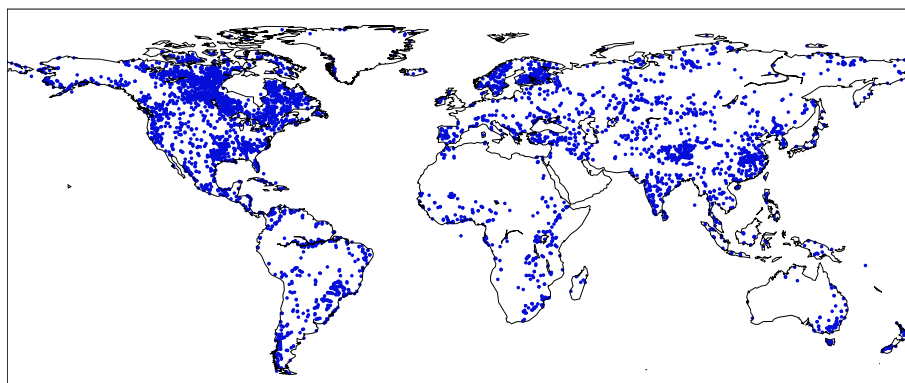
Figure 4.6: Distribution of the lakes and reservoirs with surface water extent time series used in this study.

4.3.3. Satellite altimetric water level

Monitoring the water level over the inland water bodies, including lakes, reservoirs, rivers, and wetlands, is crucial for various applications, including water resource management, natural hazard monitoring, and hydrological extreme events mitigation. Despite its importance, gauge observations are sparse and, in many cases, lack continuous measurement. Satellite altimetry, initially intended to track sea level budgets, has shown promise as a virtual lake and river gauge (e.g., Alsdorf et al., 2007; Papa et al., 2010). TOPEX/Poseidon and Envisat are the first satellite altimetry missions that used Open-Loop Tracking Command (OLTC), especially for inland water bodies monitoring. Later, the Delay Doppler technique or the concept of Synthetic ApertureRadar (SAR) is used in the Sentinel-3A and Sentinel-3B missions. Recently, the Surface Water and Ocean Topography (SWOT) is launch in December 2022, which has improved the spatio-temporal resolution of the altimetric water level for both the inland water bodies and the Ocean. The mission is a collaboration between NASA and CNES, the French space agency, in partnership with the Canadian Space Agency (CSA) and the UK Space Agency (UKSA).

The centers provide satellite altimetry data on three main levels. The level 1 data includes raw telemetry from instruments on the satellites, which are timed and located. Such data needs to be corrected in terms of instrument errors, atmospheric errors, geophysical corrections (solid Earth, ocean and pole tides, etc.), and error from the signal perturbations caused by surface reflection. The corrected data is called level 2 data. Such data can be processed using various algorithms and adjustments and is released in terms of water level, called level 3 data. Hydroweb was the first website that delivered the altimetric water level time series. After Hydroweb, several centers have released their own level-3 data for the inland water bodies. Table 4.4 lists the repositories that provide altimetric water time series. In this study, we have combined all the available level-3 products with continuous measurement from 2003 to 2020.

Table 4.4: List of the providers for the altimetric water level time series, adapted from Tourian et al. (2021).

Product	operated by	source
Hydroweb	CNES	http://hydroweb.theia-land.fr
River& Lake	ESA	http://altimetry.esa.int/riverlake
DAHITI	Deutsches Geodätisches Forschungsinstitut (DGFI), TU Munich	https://dahiti.dgfi.tum.de Schwatke et al., 2015
HydroSat	Insititute of Geodesy University of Stuttgart	http://hydrosat.gis.uni-stuttgart.de
G-REALM	United States Department of Agriculture	https://ipad.fas.usda.gov/cropexplorer/global_reservoir
GRRATS	The Ohio State University	https://doi.org/10.5067/PSGRA-SA2V1 Coss et al., 2020
AltEx	USAID and NASA	https://altex.servirglobal.net/ Markert et al., 2019
C3S LWL	CLS on behalf of Copernicus and European Commission	https://doi.org/10.24381/cds.5714c668
Water Level on VITO	Copernicus Global Land Operations CNES, LEGOS, and CLS	https://land.copernicus.eu/global/products/wl

4.3.4. Data over Iran

Precipitation data (In-situ)

A dense network of 2850 precipitation stations has been collected from Iran's Meteorological Organization (IRIMO) and Iran Water Resources Management Company (IWRM), covering 1983–2013. Figure 4.7 illustrates the distribution of gauges throughout Iran. Continuous in-situ observations up to the end of 2019 are available only for a limited number of rain gauges (less than 400), which does not provide a desirably dense measurement network. Therefore, we evaluate the performance of 10 gridded precipitation datasets that include observations from 1983 to 2013 over Iran's basins (Table 4.5). At each basin, we assess the performance of these models using the in-situ observations and select the best group of them (see subsection 4.5.2 for more details). The ensemble mean of the selected datasets at each basin is then used for calculating the long-term monthly mean (1983–2002) and precipitation anomaly over the study period (2003–2019).
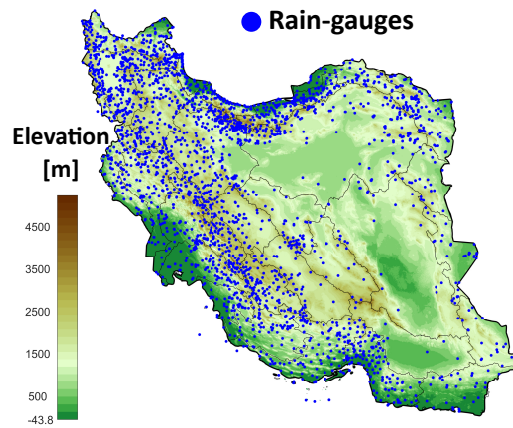

Figure 4.7: Distribution of the rain-gauges over Iran.

To compute the reference in-situ dataset, we first perform quality control and test the homogeneity for all the stations by four tests: the Buishand range test (Buishand, 1982), the Von Neumann ratio test (Von Neumann, 1941), the Standard Normal Homogeneity Test (SNHT) (Alexandersson, 1986) and the Pettit test (Pettit, 1979). In order to identify inconsistencies, we have applied the double mass curve test (Searcy & Hardison, 1960). The tests mentioned above are explained briefly in Appendix section A . Based on the aforementioned tests, we have excluded 19 stations from our assessments, because of sudden changes in precipitation patterns, such as abrupt increases or decreases or inconsistencies or shifts in the double mass curve. The remaining stations do not show data outages within 1983–2013. During this period, at each month, we average gauge values for each $0.5° \times 0.5°$ grid cells. In order to reduce the uncertainty from gauge measurements, only grid cells containing at least three gauges are considered in the estimation (Adler et al., 2003; Xue et al., 2013). Please see the distribution of cells with and without data in Figure 4.8. Two basins, 47 and 46, suffer from sparsity of gauged cells. These spatial gaps are located at two large deserts: Dashte-e-Kavir and the Lut Desert, where we expect limited precipitation variation. Therefore, the gauged cells of these two basins can be generalized to the whole basin.
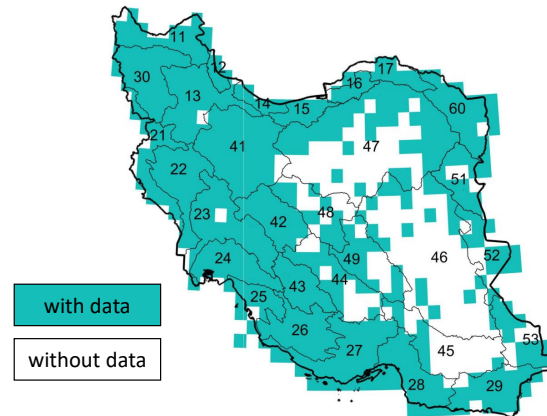
Figure 4.8: Distribution of gauged and ungauged grid cells in the in-situ dataset.

Precipitation data (globally gridded datasets)

In this thesis, we have analyzed 10 gridded precipitation datasets over Iran and its 30 major basins. The datasets are classified into three classes: gauge-based, satellite-based, and reanalysis products (see Table 4.5). All gridded precipitation datasets are re-sampled to $0.5° × 0.5°$ to be consistent with the gridded data from ground data.

Table 4.5: Summary of global precipitation datasets used for estimating the precipitation over Iran. Abbreviations in the data source(s) defined as: G, gauge; S, satellite; and R, reanalysis.

Dataset	Class	Resolution		Coverage		Reference
		Spatial	Temporal	Spatial	Temporal	
Gauge-Based Products						
PRECL	G	$0.5° × 0.5°$	1 mo	Global land	1948–2019	(Chen et al., 2002)
CPC	G	$0.5° × 0.5°$	1 d	Global land	1979–2019	(Chen et al., 2008)
Satellite-Based Products						
GPCP	G, S	$2.5° × 2.5°$	1 mo	Global	1979–2019	(Adler et al., 2003)
CMAP	G, S	$2.5° × 2.5°$	1 mo	Global	1979–2019	(Xie et al., 2003)
PERSIANN-CDR	G, S	$0.25° × 0.25°$	3,6 h /1 d	60°S–60°N	1983–2019	(Ashouri et al., 2015)
CHIRPS	G, S, R	$0.05° × 0.05°$	1d	50°S–50°N	1981–2019	(Funk et al., 2015)
Reanalysis Products						
ERA5	R	$0.25° × 0.25°$	6 h/ 1 mo	Global	1979–2019	(Dee et al., 2011)
NCEP 1	R	$2.5° × 2.5°$	6 h/1 d/ 1 mo	Global	1948–2019	(Kalnay et al., 1996)
NCEP 2	R	$1.875° × 1.875°$	6 h/1 d/ 1 mo	Global	1979–2019	(Kanamitsu et al., 2002)
MERRA-2	R	$0.5° × 0.67°$	1 d	Global	1979–2019	(Rienecker et al., 2011)

Groundwater level data

This thesis utilizes groundwater observations from 13 879 piezometric wells from IWRM (http://wrs.wrm.ir/amar, last access 20 April 2020), with the most updated available data up to June 2017 (Figure 4.9). The wells selected for analysis cover the period from 2003 to 2016, with less than 12 months of gaps, while data after 2016 was not available at the time of writing. Quality control was applied to eliminate outliers and biases in each group of wells within their corresponding aquifer, and spline interpolation was used to fill gaps in the data. The study's rigorous data processing methodology, including quality control, gap filling, and standardization, enhances the reliability and accuracy of the resulting time series of mean groundwater level, enabling more informed decisions regarding groundwater resource management in the region.

Figure 4.10 illustrates the data process to obtain the time series of Mean Groundwater Level

Anomaly (GWLA). To estimate GWLA at each basin, the time series at each well is standardized using its mean and standard deviation. The standardization enables us to merge wells that carry different properties (Tourian et al., 2015). Then, in each aquifer, we calculate the ensemble mean of all standardized levels and multiply the obtained time series by the average of the well's standard deviations before standardization. Finally, to determine the time series of mean groundwater level, at each basin we compute the ensemble mean of the aquifers' time series weighted by the aquifers' area. The standard deviation of the groundwater levels at each month is considered as the uncertainty of the final time series for each basin. Figure 4.11 illustrates the validity of the Gaussian distribution for the errors at each epoch of the GWLA. Kurtosis is a metric for how *tailed* the probability distribution is. The kurtosis of the GWLA residuals is 3.2, indicating that it is slightly higher than three. However, it can still be considered within the range of a normal distribution.
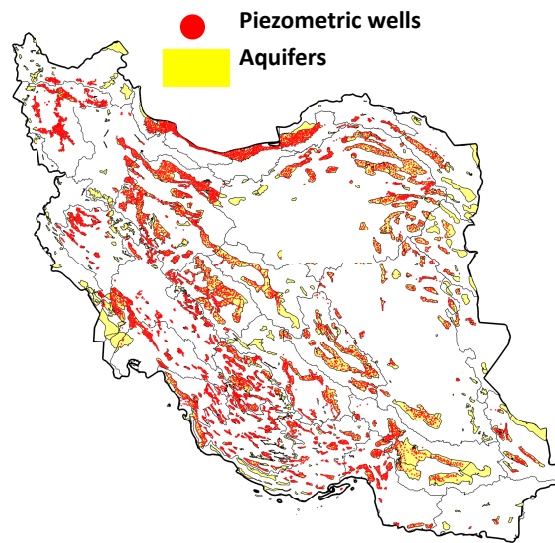

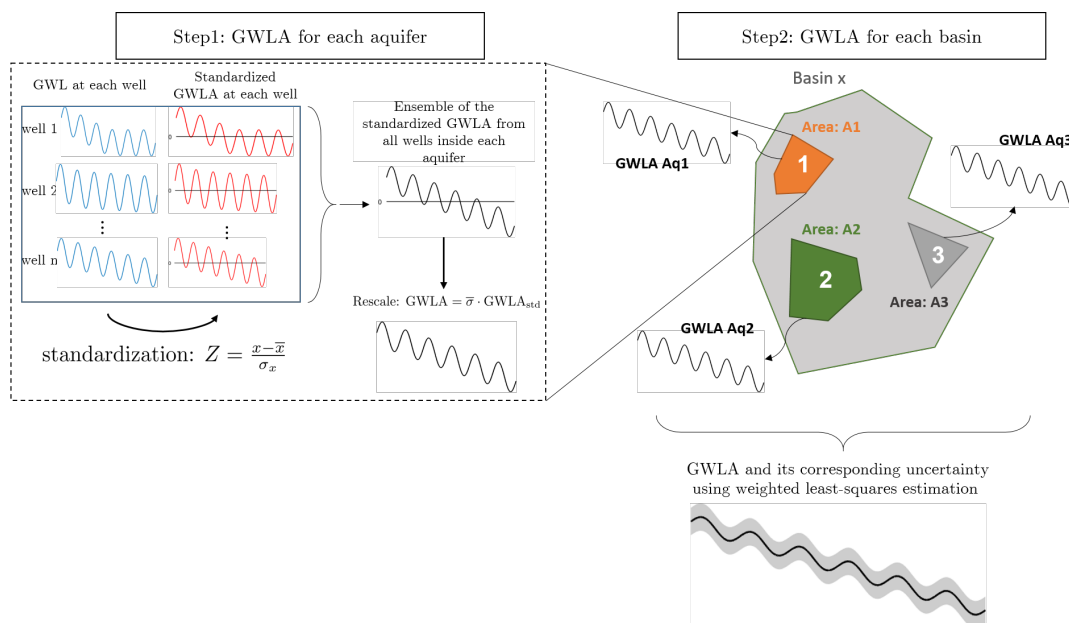Figure 4.9: Spatial distribution of piezometric wells and the aquifers in Iran.

Figure 4.10: Schematic of the main two steps to estimate GWLA for each basin.
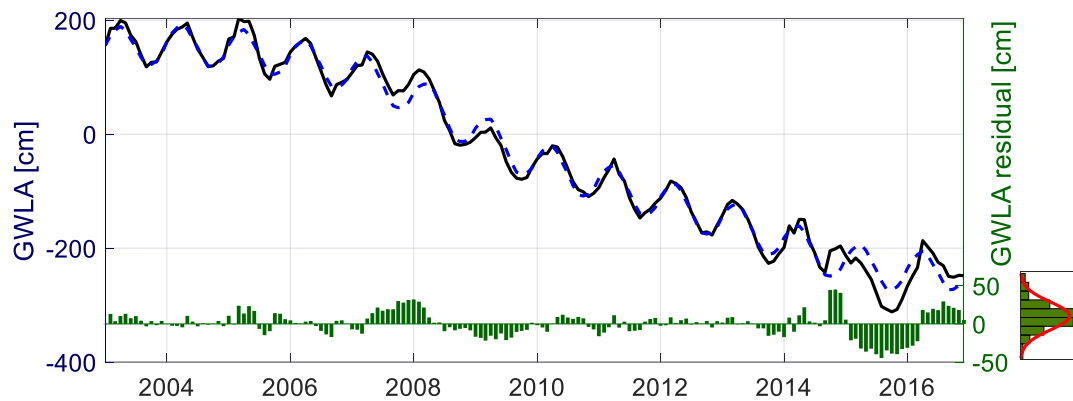
Figure 4.11: (a) GWLA time series together with the estimated combination of nonlinear trend and seasonality signal over Iran within 2003–2016. The residuals, GWLA signal after removing trend and seasonal variation, are shown as green bars. (b) Histogram of the GWLA residuals together with the fitted distribution.

4.4. Validation of the datasets

In this chapter, we employed a range of evaluation metrics to thoroughly assess the performance of the models. These metrics encompassed the correlation coefficient, Kling-Gupta Efficiency (KGE), its two constituent parameters (γ and β), Normalized Root Mean Square Error (NRMSE), and Nash-Sutcliffe Efficiency (NSE). For a comprehensive understanding of each metric and its calculation, please refer to the detailed information provided in Appendix section B.

4.4.1. Validation of TWSA from fluxes

In order to assess the performance of the fluxes estimations, we have compared the left side (derivative of GRACE-TWSA) with the right side (fluxes) of the Equation 4.1 over the study regions.

Over sub-continents

Figure 4.12 and Figure 4.13 show the result of this comparison over the sub-continents. To quantify the accuracy of the water balance fluxes, we have employed four evaluation metrics namely Correlation, NSE, and NRMSE. Summary of the evaluation metrics is shown in Table 4.6. Generally, the ensemble mean of flux datasets agrees well ($r \geq 0.56$) with the GRACE TWSA derivation. The correlation coefficient varies between 0.56 to 0.95, while the highest correlation is obtained over the Indian-sub-continent ($r = 0.95$), followed by West Africa ($r = 0.92$) and East Europe ($r = 0.92$) and the lowest values are obtained over Central Asia and North Africa with $r = 0.56$. In terms of NSE, the performances vary between 0.17 and 0.90, with the highest NSE over the Indian sub-continent (NSE = 0.90), followed by West Africa (NSE = 0.85). The lowest NSE is observed over East Africa ($r = 0.04$), followed by Central Africa ($r = 0.17$) and Central Asia ($r = 0.18$). Time series of the fluxes balance ($P - ET - R$) with respect to the derivative of GRACE-TWSA over the sub-continents are shown in Figure 4.12 and Figure 4.12.

Table 4.6: Evaluation of the fluxes with respect to the derivative of GRACE-TWSA as the reference over the sub-continents.

| Region	r	NSE	NRMSE
Australia	0.82	0.35	0.14
Central Asia	0.56	0.18	0.16
Indian sub-continent	0.95	0.90	0.08
Middle East	0.83	0.67	0.12
North Europe	0.74	0.33	0.23
South Europe	0.91	0.57	0.14
West Europe	0.89	0.47	0.16
East Europe	0.92	0.79	0.12
South America	0.91	0.65	0.17
North America	0.91	0.81	0.12
Central America	0.87	0.57	0.15
North Africa	0.56	0.31	0.12
South Africa	0.91	0.81	0.11
Central Africa	0.75	0.17	0.18
East Africa	0.65	0.04	0.18
West Africa	0.92	0.85	0.10

4
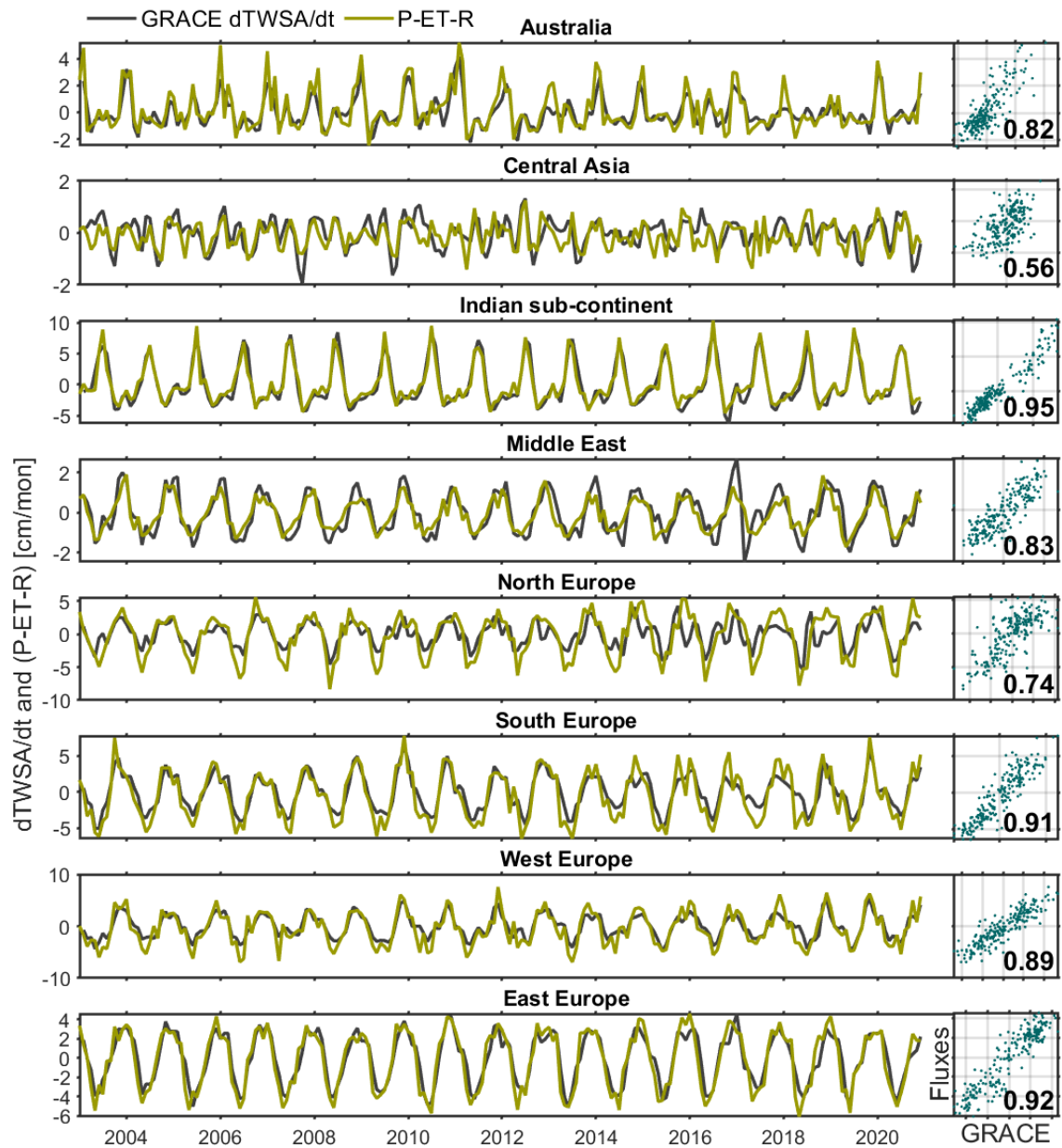
Figure 4.12: Comparison of the derivative of GRACE-TWSA ($dTWSA/dt$) with the balance of the water fluxes ($P - ET - R$) over sub-continents. The scatter plot and the correlation coefficient for each region are shown on the right side of the time series.
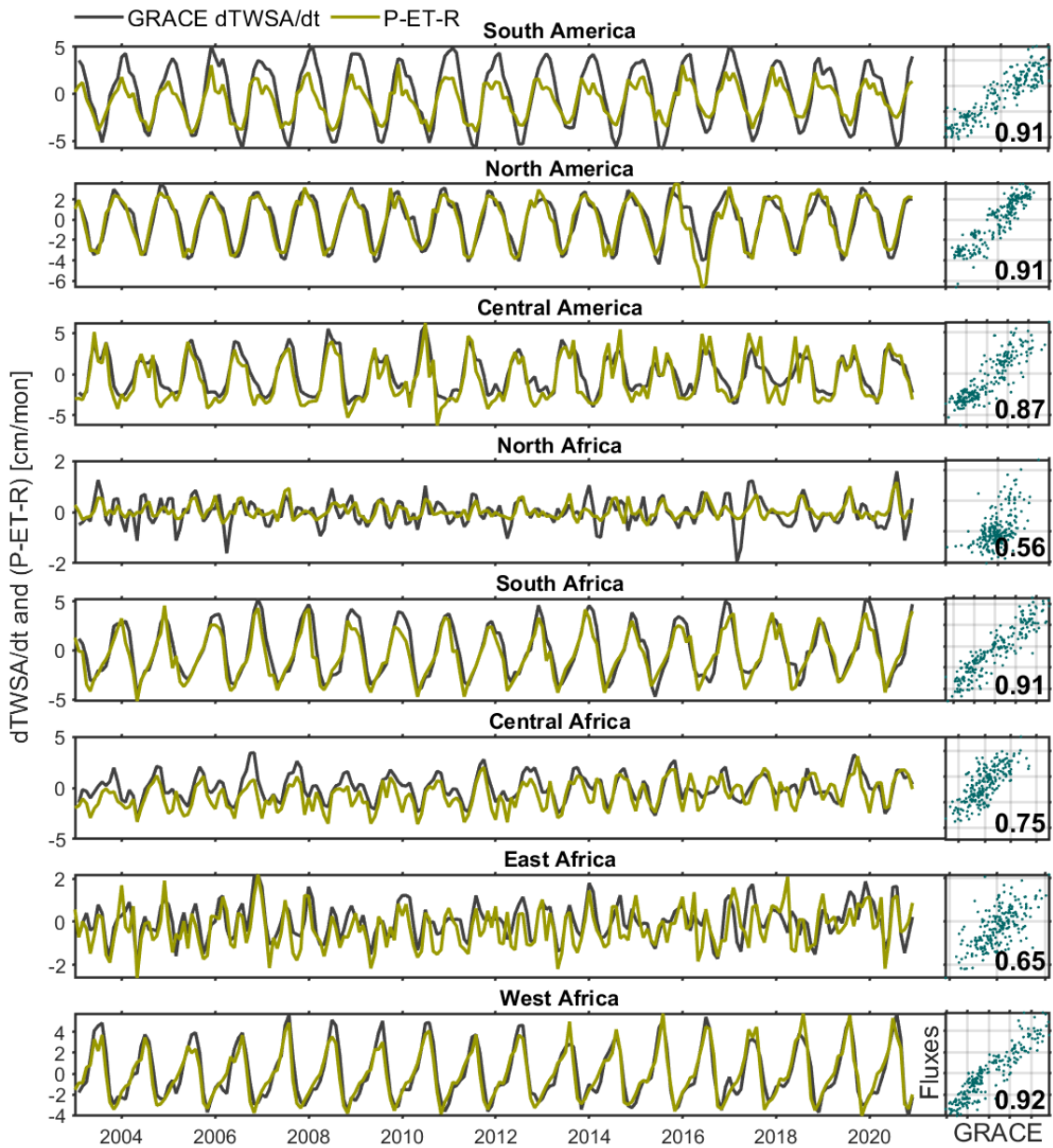
Figure 4.13: Figure 4.12 continued.

Over major river basins:

The results of the evaluation using three metrics are presented in Figure 4.14. These metrics reveal that the performance of the flux datasets varies across the basins. On average, the flux estimations correlate well with the derivative of GRACE(-FO), with 85 % of the basins exhibiting a correlation coefficient greater than 0.6. The Northern Sahara basin displays the lowest correlation coefficient of 0.15, while the Columbia basin in the US displays the highest correlation coefficient of 0.95. In terms of correlation, the datasets perform best over Europe ($\overline{r} = 0.87$), followed by South America ($\overline{r} = 0.75$). The worst performance is observed, on average, over Australia.

Regarding the Nash-Sutcliffe Efficiency (NSE), the performance of the datasets ranges from as low as -2.2 in the St. Lawrence basin to as high as 0.88 in the Ganges basin. On average, the NSE over the selected basins is 0.74, with 72 % of the basins showing an NSE greater than 0.1, and over 75 % of them having an NSE greater than 0.5. The best performance, on average, is observed over the basins in Europe (NSE = 0.65). However, the datasets fail to represent the dynamics of storage change in the three major basins of Australia. One possible reason for this low performance could be the strong effect of the El Niño-Southern Oscillation (ENSO), which significantly affects the RMS and trend of the flux values over Australia, resulting in severe droughts and heavy flood events.

The NRMSE yields similar results to NSE across the basins. In the Ganges basin, the NRMSE is less than 10 %, while in the St. Lawrence basin, it is 30 %. More than half of the basins have NRMSE values between 12–16 %, which is acceptable given the ensemble mean of the datasets. However, some basins exhibit low accuracy in estimating water storage change, as evidenced by negative NSE and high NRMSE values. It is important to note that such results should not be over-interpreted for monthly evaluations. Furthermore, it is worth mentioning that the water fluxes are presented as total estimates over the study period in the results section, rather than as monthly or annual comparisons.

Figure 4.15 illustrates the boxplot of the metrics in Figure 4.14 over different climate categories. The correlation between the fluxes balance and the derivative of GRACE TWSA is generally high across all climate types, but it is the highest in the arid to hyper-arid regions (0.81 to 0.94), and lowest in the humid regions (-0.02 to 0.94). This suggests that the relationship between the two variables is stronger in arid regions compared to humid regions. A lower NRMSE indicates better model performance. The NRMSE is generally lowest in dry sub-humid regions (0.37 to 0.68), and highest in arid to hyper-arid regions (0.38 to 4.70). This suggests that the accuracy of the model is better in regions with moderate precipitation.

The boxplot in Figure 4.15 depicts the metrics from Figure 4.14 across diverse climate categories. Overall, the correlation between the fluxes balance and the derivative of GRACE TWSA is robust across all climate types. However, it is most pronounced in arid to hyper-arid regions, ranging from 0.81 to 0.94, while being the lowest in humid regions, ranging from -0.02 to 0.94. This observation implies a more robust relationship between the two variables in arid regions in comparison to humid regions. Furthermore, a lower NRMSE value implies better model performance. It was found that the NRMSE is generally lowest in dry sub-humid regions, ranging from 0.37 to 0.68, and highest in arid to hyper-arid regions, ranging from 0.38 to 4.70. This finding suggests that the model's accuracy is superior in regions with moderate precipitation.

The bias (β) gauges the disparity between predicted and observed values, and lower bias signifies superior model performance. Notably, the bias appears to be minimal in dry sub-humid regions (0.67 to 3.51), and at its peak in semi-arid regions (2.84 to 1070.37), indicating that the

model has a tendency to over-predict flux balance in semi-arid regions. Conversely, the variability ratio (γ) measures the ratio of the predicted values' standard deviation to that of the observed values, and a lower variability ratio suggests better model performance. The variability ratio is generally lower in dry sub-humid regions (0.02 to 0.49) and higher in humid regions (0.01 to 179.89). The model, therefore, tends to overestimate the variability of flux balance in humid regions.

Overall, these findings indicate that the model's accuracy is higher in arid to hyper-arid regions than in humid areas. The model performs better in regions with moderate precipitation, lower variability, higher correlation, and lower bias. Notably, the model's error is primarily attributable to variability ratio and bias, rather than correlation.

4
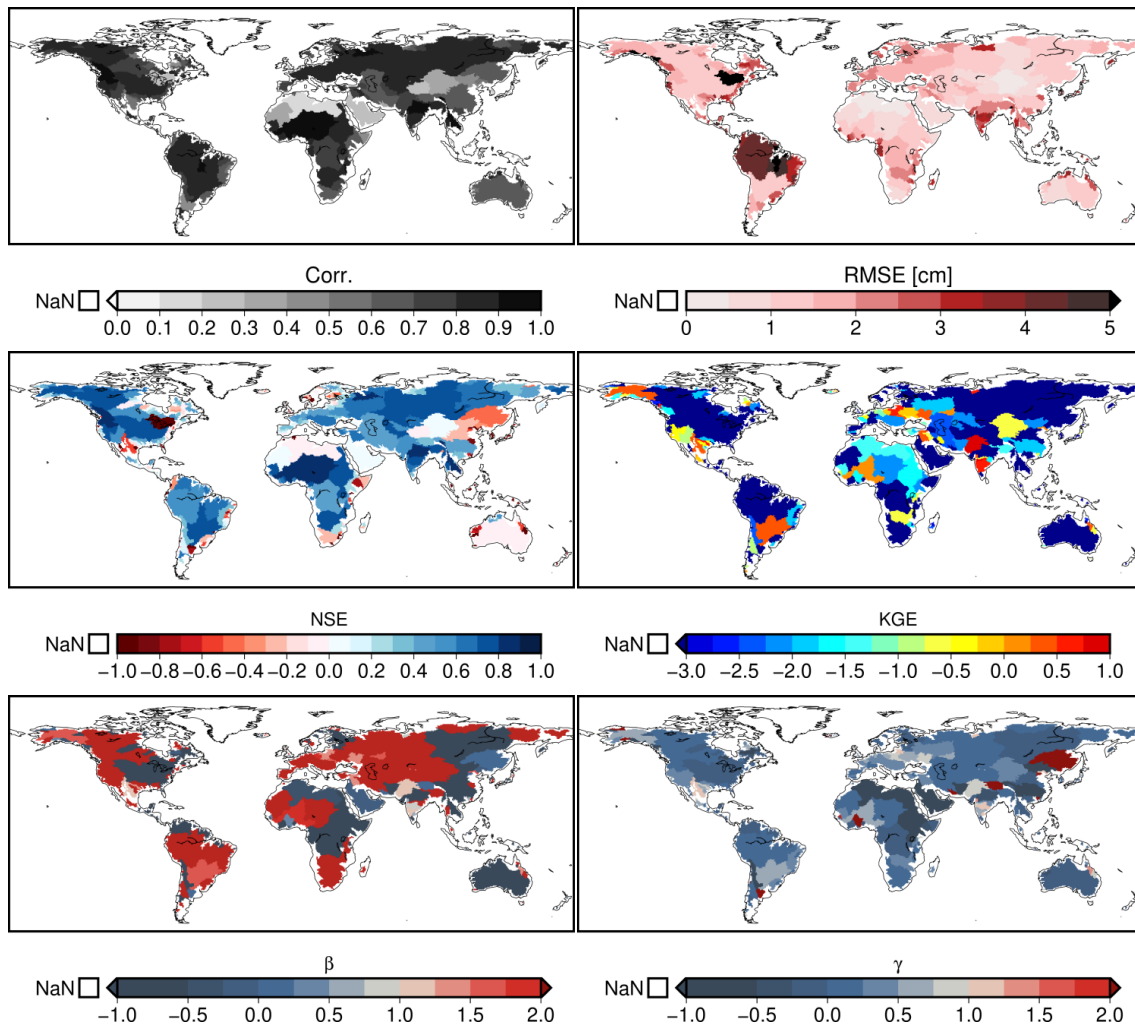
Figure 4.14: Evaluation of the fluxes with respect to the derivative of GRACE-TWSA as the reference over the major river basins. The figure includes Pearson's correlation (r), Root Mean Squared Error (RMSE), Nash-Sutcliffe Efficiency (NSE), Kling-Gupta efficiency (KGE), bias (β), and variability ratio (γ).

Figure 4.15: Boxplot of the evaluation metrics presented in Figure 4.14 over the four categories of climate. Only the Normalized RMSE is used instead of RMSE to make the comparison easier.

4.4.2. Validation of the surface water extent

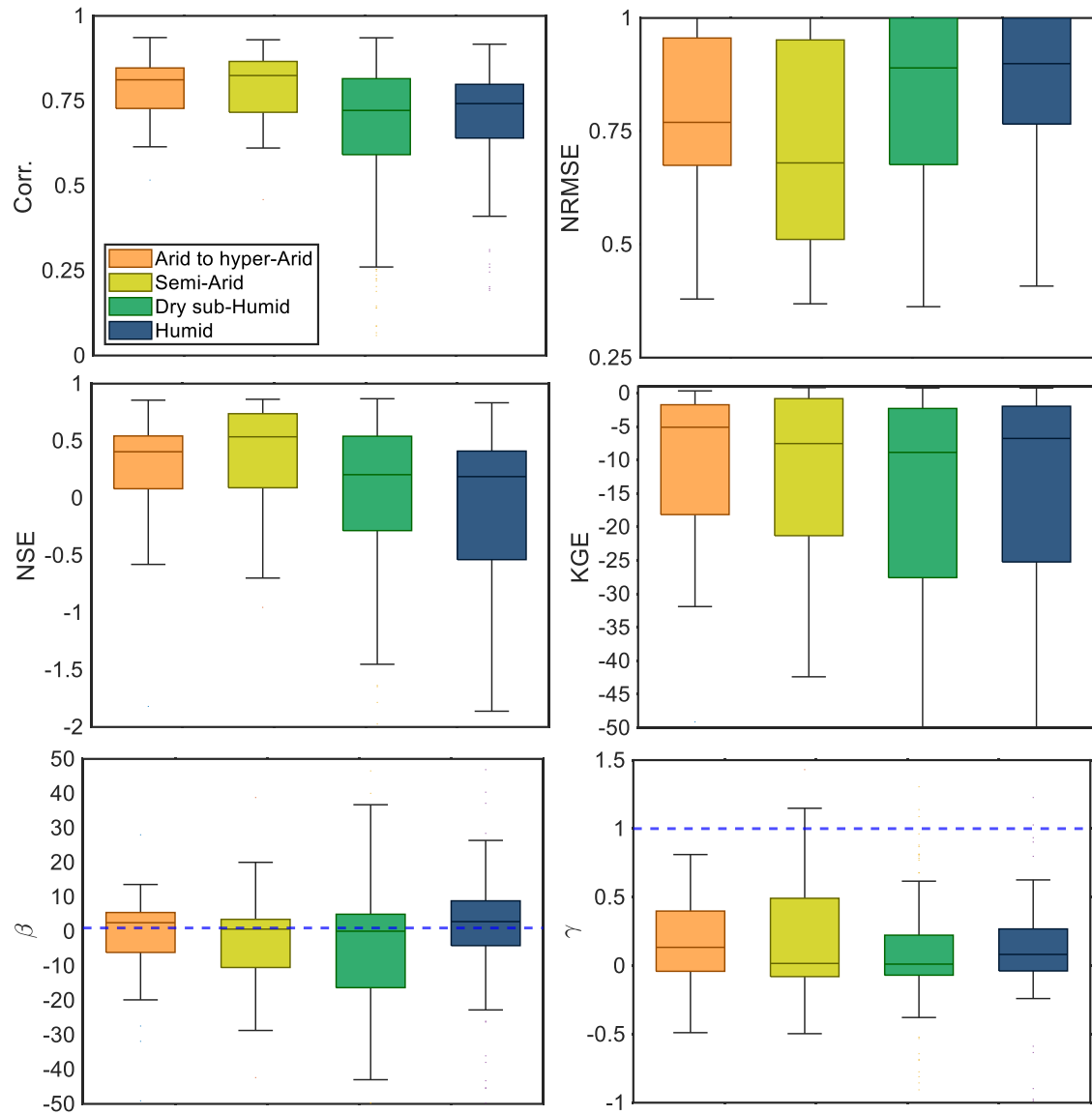To validate the time series of the surface water extent for each lake or reservoir, the correlation coefficient between the monthly surface area from satellite imagery with water level from monthly satellite altimetry is calculated. Figure 4.16 illustrates the correlations worldwide. A high correlation (0.5 or higher) is obtained for most lakes and reservoirs (>70 %). The algorithm used to extract the lake extent performed well, as evidenced by the strong association between the lake's area and the altimetric water level (Figure 4.16). Although the high correlation between the surface water area and the altimetric water level indicates a satisfactory performance of the algorithm in estimating the surface water extent, the low correlation can be the result of various reasons, including (1) high regulation over reservoirs and regulated lakes which significantly affect the natural dynamic of the area and level relationship, (2) the satellite altimetry may overpass the part of the lake that is disconnected to the lake which can not represent the lake (reservoir) variation.
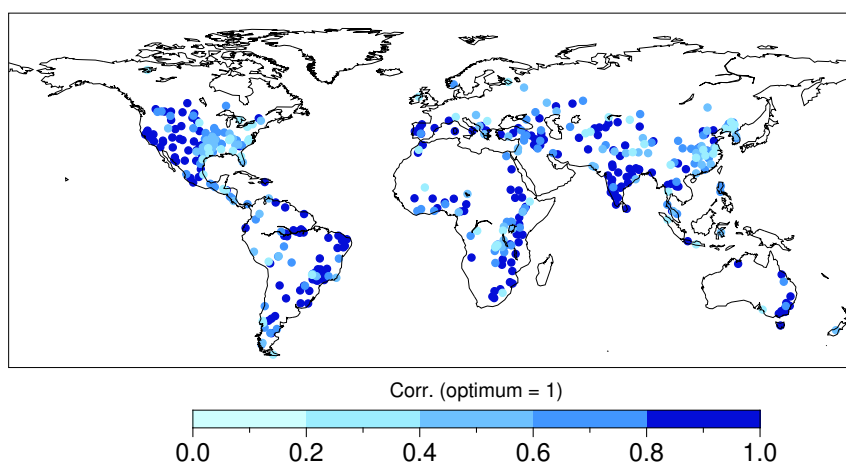

Figure 4.16: Global distribution of the correlation coefficient between the surface water extent from satellite imagery and the altimetric water level.

4.5. Methodology

4.5.1. Trend analysis

A linear trend does not necessarily represent the long-term behavior of TWSA time series over many regions, because it may fail to capture seasonal and cyclic patterns. Moreover, time series data are often autocorrelated, meaning that each observation is correlated with previous observations. This autocorrelation can result in changes in the trend over time that are not linear and cannot be adequately represented by a linear trend. Furthermore, the long-term behavior of a time series may exhibit non-monotonic changes, where the trend is not consistently increasing or decreasing over time. A non-linear trend may better capture these complex patterns and provide a more accurate representation of the long-term behavior of the time series.

There are several methods that can be used to extract the non-linear trend in a time series. Here are some of the most commonly used methods:

- Non-linear regression: Non-linear regression is a statistical method that involves fitting a non-linear equation to the data. Non-linear regression can be used to model a wide range of non-linear trends, and it allows for the inclusion of additional predictors and covariates in the model.

- Polynomial regression: Polynomial regression involves fitting a polynomial equation to the data. Polynomial regression can be used to model a wide range of non-linear trends, but it is limited by the degree of the polynomial that can be used.

- Smoothing methods: Smoothing methods involve averaging the data over a moving window or a fixed interval. Smoothing methods can be used to estimate the non-linear trend by removing the short-term fluctuations in the data.

- Wavelet analysis: Wavelet analysis involves decomposing the time series into a set of wavelet coefficients. Wavelet analysis can be used to identify the non-linear trend and other features of the data at different scales and frequencies.

- Machine learning methods: Machine learning methods, such as neural networks and support vector machines, can be used to model non-linear trends. Machine learning methods can be particularly useful when the data is high-dimensional and complex.

- Seasonal and Trend decomposition using Loess (STL): STL is another data-driven method that involves decomposing the time series into three components: the seasonal component, the trend component, and the remainder component. The trend component is estimated using a non-parametric smoothing method called Locally Weighted Scatterplot Smoothing (Loess). STL can be used to extract non-linear trends from time series data with complex seasonal patterns, and it is particularly useful for data with multiple seasonal cycles.

- Singular Spectrum Analysis (SSA): SSA is a data-driven method that involves decomposing the time series into a set of orthogonal components called empirical orthogonal functions (EOFs). The first EOF represents the trend component, while the remaining EOFs represent the noise and cyclical components. SSA can be used to extract non-linear trends from noisy and complex time series data, and it is particularly useful when the data contains strong cyclical or seasonal patterns.

Singular Spectrum Analysis (SSA) is a powerful and versatile method for analyzing time series data (Blewitt & Lavallée, 2002). It has several advantages over other methods for extracting non-linear trends, including:

1. Data-driven approach: SSA is a data-driven method that does not require a priori assumptions about the shape or form of the trend. Instead, it uses the data itself to identify the underlying patterns and structures in the time series.

2. Adaptive decomposition: SSA decomposes the time series into a set of orthogonal components called empirical orthogonal functions (EOFs), which are adaptive to the data and capture the most important patterns and structures. This allows for a flexible and efficient decomposition of the time series, even when it is noisy and complex.

3. Separation of components: SSA separates the time series into its different components, including the trend, cyclical, and noise. This allows for a more detailed and interpretable analysis of the underlying patterns and structures in the data.

4. Outlier detection: SSA can be used to detect outliers and anomalies in the time series, which can be useful for identifying data quality issues or unusual events.

5. Parameter selection: SSA allows for the selection of important parameters, such as the window length and the number of EOFs, using cross-validation or other statistical crite-

ria. This ensures that the analysis is robust and reliable.

Compared to other methods discussed above, such as non-linear regression, polynomial regression, smoothing methods, and STL, SSA has some unique advantages. Non-linear regression and polynomial regression can be sensitive to the choice of model and the initial conditions, and may not be able to capture complex and non-linear patterns in the data. Smoothing methods, such as moving averages and exponential smoothing, can remove important information from the data and may not be able to adapt to changes in the trend over time. STL can be useful for extracting non-linear trends with complex seasonal patterns, but it may not be able to handle data with multiple seasonal cycles or strong noise

Figure 4.17 (a) presents the time series of the TWSA over Australia from 2003 to 2021, with its uncertainty shown as the gray envelope. We obtain the long-term non-linear signal using the SSA approach with a 2-year window. The 2-year window results in the minimum linear trend left in the seasonal signal after removing the SSA non-linear trend. However, the SSA method does not deliver uncertainty for the estimated non-linear trend. Therefore, to obtain a realistic uncertainty, we perturb storage and groundwater time series according to their stochastic information using a Monte-Carlo simulation (Metropolis & Ulam, 1949; Mooney, 1997). Monte Carlo simulation is a statistical method that can be used to assess the uncertainty in a model or an analysis by generating multiple realizations of the data. In the context of time series analysis, Monte Carlo simulation can be used to deal with the uncertainty in the data and estimate the non-linear trend using SSA.

To use Monte Carlo simulation with SSA, one would generate multiple simulations of the time series data by adding noise to the original data. To perturb, we assume a normal distribution for each monthly value, centered at the signal with the estimated uncertainties as its standard deviation. We simulate 10 000 realizations of the time series to acquire a comprehensive representation of all possible realizations (Figure 4.17 (b)). Each simulation would be analyzed using SSA to extract the non-linear trend component (Figure 4.17 (b)). The resulting trend estimates from the multiple simulations can be used to assess the uncertainty in the trend estimate and calculate confidence intervals.

We estimate the total water loss or gain within the study period from each of the non-linear realizations by subtracting the last value from the first value (Figure 4.17 (c)). The total water storage loss from all realizations forms a normal distribution according to the Monte-Carlo-Simulation assumption (see the bottom left of Figure 4.17 (c)). Finally, the mean of all 10 000 realizations would be the representative total water storage loss or gain, and their standard deviation indicates uncertainty.

4.5.2. Precipitation analysis over Iran

We quantify relative gain or deficit in precipitation within the past 17 years (2003–2019). Gain (deficit) is considered to be the precipitation higher (lower) than a reference, which is defined as the long-term monthly mean from 1983 to 2002. To this end, we first analyze the datasets to nominate the most reliable precipitation data of each basin. As error measure at each basin, we subtracted the monthly gauged precipitation throughout 1983–2013 from the corresponding values in the gridded precipitation datasets. It should be noted that to calculate the monthly values from precipitation datasets, at each basin we have included only the grid cells with measurements in the in-situ dataset (colored cells in Figure 4.8). To evaluate the results, we do not rely on any standard metric like Nash-Sutcliffe, Root Mean Square of Errors (RMSE) or bias as they provide only a summary rather than a full view of the error. Instead, we look at the whole distribution of the error which allows us to evaluate errors in all quantiles. The errors contain
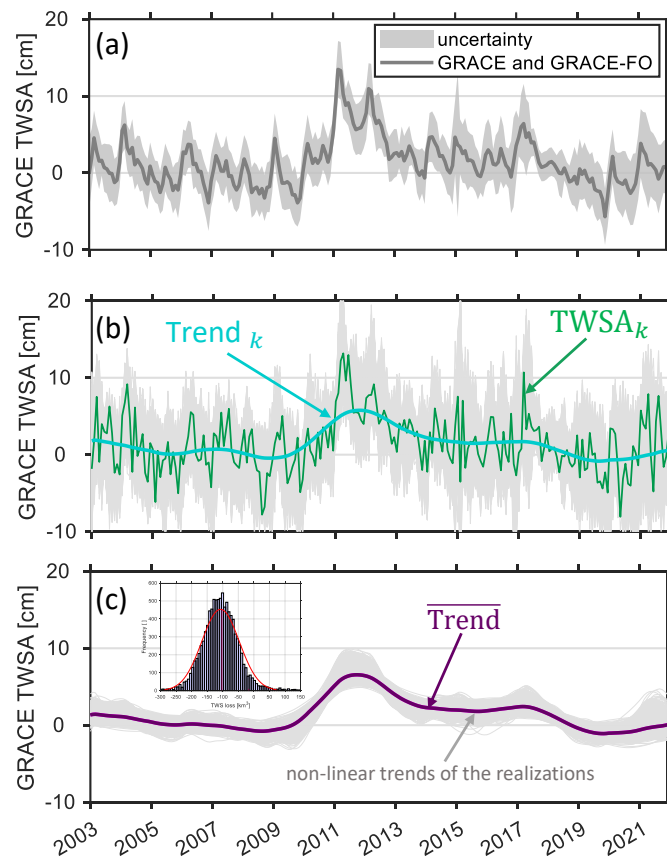
Figure 4.17: (a) GRACE TWSA time series including uncertainty (gray envelope). (b) Simulating 10 000 time series in a Monte Carlo simulation approach for which the gray envelope represents all realizations. The green line shows the simulated signal for the kth realization (TWS_k) and the blueish line its corresponding non-linear trend (Trend_k). (c) The non-linear trend of the 10 000 realizations using SSA (gray envelope) together with the mean of all non-linear trends ($\overline{\text{Trend}}$). For each non-linear trend, we calculated the total water storage loss by subtracting the last value from the first. The histogram of all 10 000 TWS loss which forms a normal density function fit in red, is shown in the left corner of the plot (c). This normal distribution's mean and three standard deviations are the estimated total water storage loss and its corresponding uncertainty.

both negative and positive values and have typically a non-zero bias. The Cumulative Distribution Function (CDF) plot can not be directly used as a metric. To benefit from the CDF as a metric, one should fold the error over a specific number. One choice can be zero (absolute value of the error) assuming the dataset as bias-free which might not hold in our case. We fold the error over its median (i.e., $|e - \textbf{Median}(e)|$) and then build its CDF. Using this method we can detect the bias in the dataset that will affect our final ranking of the datasets.

Figure 4.18 (b) shows the CDF of 10 precipitation datasets over the Lake Urmia basin. They perform similarly, with 55–70 % of the centered errors below 10 mm. At each basin, we select the precipitation datasets with an error at the 90th percentile less than 15 % of the mean annual precipitation of the basin, with the mean taken from GPCC (see Figure 4.18 (b)). The selected datasets are shown in Figure 4.18 (a). For each basin, we calculate the time series of monthly precipitation for the period 1983–2019 from the selected datasets and the corresponding error at 90 % CDF as their weight using least squares adjustment.

We quantify water input stability at each basin by the precipitation anomaly $\delta P(t_{y,m})$ for the period of 2003–2019 with respect to its long-term monthly mean (1983–2002):
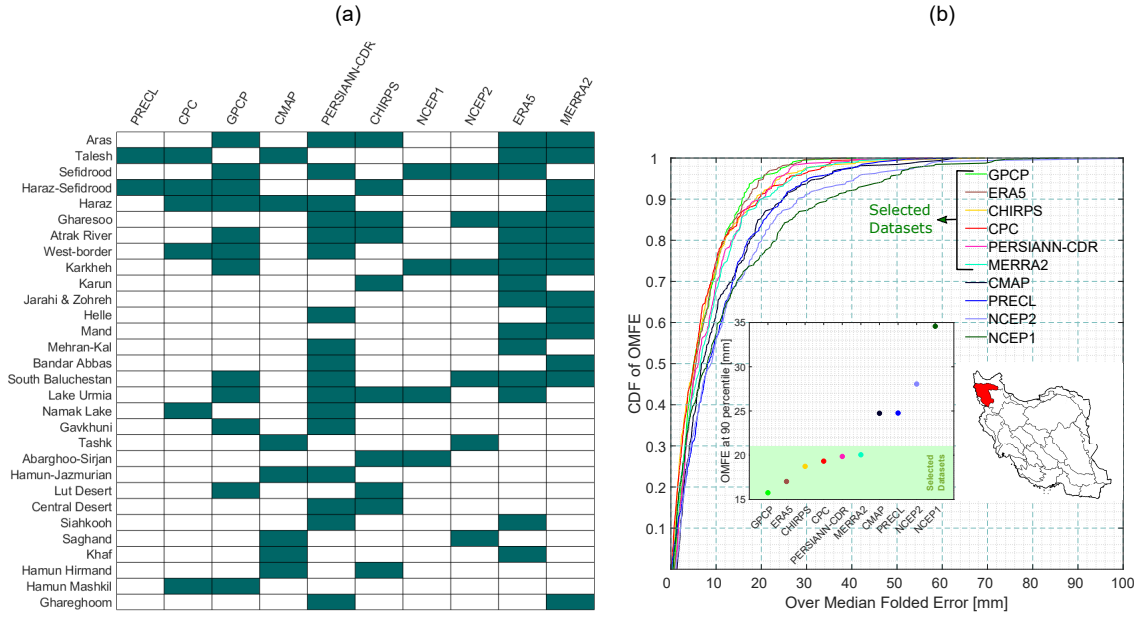
Figure 4.18: (a) Filled rectangles at each row represent the selected precipitation datasets at each basin of Iran. (b) CDF of the centered error for all datasets in the Lake Urmia basin and the selected datasets including their error at CDF = 0.9.

$$\delta P(t_{y,m}) = P(t_{y,m}) - \frac{1}{20} \sum_{y=1983}^{2002} P(t_{y,m}) , \tag{4.2}$$

where P is the precipitation, $t_{y,m}$ represents the month (m) and year (y). We then obtain the long-term mean annual anomaly by:

$$\overline{\delta P} = \frac{1}{17} \sum_{y=2003}^{2019} \sum_{m=1}^{12} \delta P(t_{y,m}) . \tag{4.3}$$

4.6. Results

4.6.1. Sub-continental analysis

In accordance with the methodology outlined in subsection 4.5.1, this study estimates the non-linear trend and corresponding uncertainty of GRACE-TWSA for sub-continents during the period spanning from 2003 to 2021. Figure 4.19 presents the TWS loss or gain, as well as its corresponding uncertainty, relative to the 2003 state. It is pertinent to note that the initial state, 2003, is arbitrary, and we are solely investigating the relative change in TWS. With the exception of sub-Saharan Africa, all sub-continents examined in this study have experienced a decline in water storage compared to their status in 2003. In certain regions, such as Central Asia, the Indian sub-continent, the Middle East, Europe, and North America, water storage levels remained below the 2003 state for the majority of the 2003–2021 time-frame (exceeding 70 %) (Table 4.7). North America registered the largest TWS loss of $2302 \pm 321\,\text{km}^3$ at a rate of $-121 \pm 17\,\text{km}^3/\text{year}$. Despite a three-year increase in TWS from 2003 to 2006, a controversial negative trend emerged in 2006, resulting in this significant water loss. Following North America, the Middle East, Central Asia, and Eastern Europe have experienced losses of $813 \pm 70\,\text{km}^3$, $590 \pm 99\,\text{km}^3$, and $570 \pm 48\,\text{km}^3$, respectively. In all the regions mentioned above, the onset of the negative trend can be observed around 2006–2008.

The overall TWS loss for Europe is estimated to be $890 \pm 57\,\text{km}^3$, with the vast majority (~90 %) occurring in the North and West of the continent. Between 2003 and 2014, the South and West of Europe experienced fluctuating TWS levels, with intermittent periods of water gain and loss. However, both regions exhibited a significant negative trend in TWS from 2015 to 2020, leading to TWS losses of $108 \pm 12\,\text{km}^3$ and $52 \pm 11\,\text{km}^3$, respectively (Figure 4.19). In contrast, West Europe experienced a noteworthy positive trend starting from 2019, which resulted in a TWS increase of $37 \pm 12\,\text{km}^3$.

Australia has experienced a significant decrease in its TWS, amounting to $100 \pm 32\,\text{km}^3$ at a rate of $-5.2 \pm 1.7\,\text{km}^3/\text{year}$ from 2003 until the conclusion of 2021. The evolution of TWS in Australia is characterised by six primary phases. During the initial phase, which spanned from 2003 until 2008, the region lost $152 \pm 37\,\text{km}^3$. This period coincided with one of the most severe droughts in recent Australian history known as the *Millennium Drought*. Subsequently, the region experienced a series of floods from 2008 until 2012, leading to a TWS gain of $547 \pm 50\,\text{km}^3$ and commencing the second phase. The third phase began in 2012 with a substantial negative trend that ceased in 2014, causing a TWS loss of $234 \pm 35\,\text{km}^3$. Following this remarkable loss of water in a mere two years, the region remained in equilibrium regarding water storage fluctuations for four years (2012–2016). However, a notable negative trend ensued, resulting in the loss of $261 \pm 20\,\text{km}^3$ from TWS. The region showed signs of TWS recovery starting from early 2019, yet the recovery remains insufficient to offset the colossal water loss from 2016 until 2018.

Africa as a continent has observed a net gain in water, except for the North region. The total TWS of the continent has increased by $2380 \pm 165\,\text{km}^3$. Except for North Africa, the sub-continents of Africa have mostly maintained a higher TWS (>70 %) than their 2003 baseline during the period from 2003 to 2021. South, Central, and East Africa experienced a negative trend from 2003 to 2006, resulting in TWS losses of $301 \pm 88\,\text{km}^3$, $120 \pm 27\,\text{km}^3$, and $87 \pm 22\,\text{km}^3$, respectively. In contrast, North Africa showed a decline in TWS, with a loss of $150 \pm 30\,\text{km}^3$ from 2003 to 2021, despite a weak positive trend ($30 \pm 12\,\text{km}^3$) from 2003 to 2006.

Central Asia, the Indian sub-continent, and the Middle East have experienced a significant drop in TWS from 2003 to 2021. During this period, these regions were mostly (>84 %) below their 2003 TWS levels. The Middle East recorded the highest TWS loss of $-813 \pm 70\,\text{km}^3$. This enor-

mous loss in a region with an arid to semi-arid climate began in late 2007 and has shown no sign of recovery as of 2021. In Central Asia, despite a weak positive trend that increased TWS by $70 \pm 15\,\text{km}^3$ within 2003–2005, the region has lost $-590 \pm 99\,\text{km}^3$ of its TWS over the last two decades. The Indian sub-continent experienced a considerable TWS loss from 2008 to the end of 2019, resulting in a loss of $-527 \pm 29\,\text{km}^3$. The region has exhibited a weak positive trend from early 2020 onward, which partially recovered $63 \pm 13\,\text{km}^3$ (~11 %) of its TWS.

The Table 4.7 provides estimates of the relative gain (or deficit) in water fluxes including precipitation, evapotranspiration, and runoff 16 sub-continents of the world. Such a gain or deficit is calculated by subtracting the total amount of fluxes from 1984 to 2002 from the total amount of 2003 to 2021. Additionally, the table presents the relative TWS for each sub-continent, with respect to the TWS in 2003. The estimates of the water fluxes show that Australia experienced no significant changes in the water balance during the observed period. In contrast, Central Asia had the highest relative positive change in precipitation, evapotranspiration, and runoff, with a respective increase of 6, 7, and 2 $10^3\,\text{km}^3$ in which the balance result in water loss over the region. The Indian sub-continent experienced a positive relative change in precipitation and evapotranspiration, but no change in runoff. The TWS, however, showed a negative relative change of -419 ± 37. In the Middle East, the relative change in precipitation was negative, while evapotranspiration and runoff remained unchanged. This negative trend in precipitation has led to a significant decrease in TWS, with a negative value of -813 ± 70.

During the study period, the North European sub-continent experienced an increase in precipitation, although there was a decrease in runoff. Despite this seemingly positive balance, there was actually a decline in TWS during this time. This discrepancy may be attributed to other factors that affect water storage, such as changes in groundwater pumping, land-use, and soil moisture. On the other hand, the South and West European sub-continents did not show significant changes in the water fluxes balance, which is consistent with the TWS estimations from GRACE data. Similarly, East Europe was estimated to have the same balance, but the result was not consistent with the relative TWS change observed. This inconsistency may be explained by the contribution of permafrost in the region, which is not fully captured in global datasets used to estimate the water fluxes. This highlights the importance of accounting for local factors when studying changes in water resources.

The flux balances of North and Central America indicates no significant change in water fluxes. Although the GRACE observation supports the findings over Central America, a significant discrepancy is observed between the relative change in water fluxes and TWS. This discrepancy can be attributed to the contribution of ice-sheet loss, which is observable by GRACE and can be inferred by an increase in evapotranspiration. In contrast, South America displays a positive balance, mainly due to the decrease in runoff. This decline is consistent with the decrease in relative precipitation over the sub-continent.

The African sub-continents exhibited mixed results during the study period. North Africa showed a small negative change in TWS relative to the state in 2003, which was also supported by the results from the water fluxes balance that showed no significant change. It is worth noting that much of North Africa is covered by the Sahara Desert, where TWS variations are minimal. This small variation decreases the Signal to Noise Ratio (SNR) and increases the possibility that the negative trend observed in TWS is partly due to measurement errors. Conversely, the relative TWS increased in the other sub-continents of Africa, which can be explained by a decrease in runoff, especially in Central Africa.

Table 4.7: Summary of total water fluxes and changes in TWS across selected sub-continents. The relative values for the fluxes are derived by subtracting the total amount between 1984 and 2002 from the total amount between 2003 and 2021. Regarding TWS, the relative status indicates the level in 2021 compared to the level in 2003.

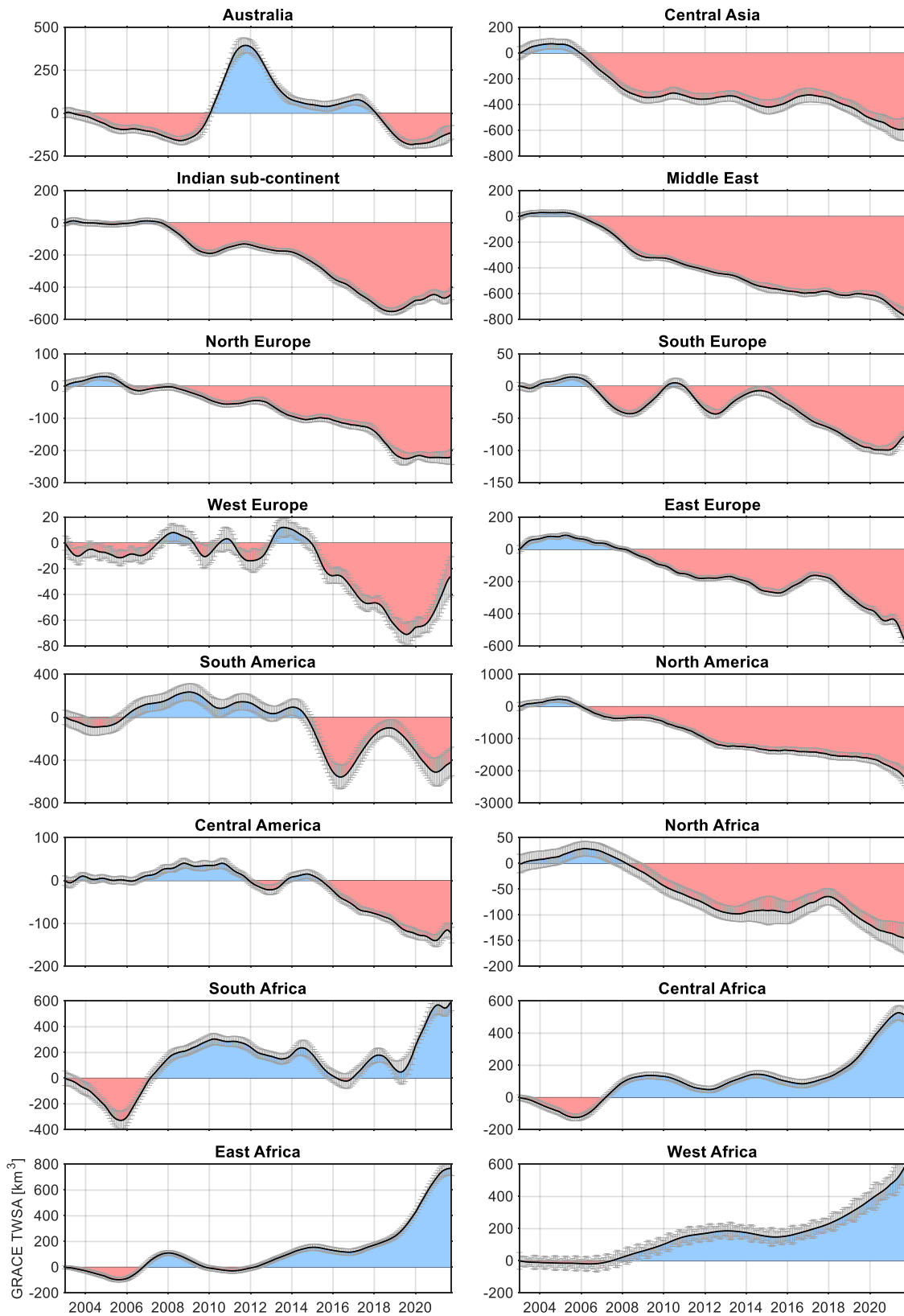| Region	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3]	rel. TWS [km^3]
Australia	7.7	1	0	0	-100 ± 32
Central Asia	14.9	6	7	2	-590 ± 99
Indian sub-continent	5.1	5	5	0	-419 ± 37
Middle East	6.9	-1	0	0	-813 ± 70
North Europe	1.8	1	0	-2	-215 ± 27
South Europe	1.7	1	1	0	-79 ± 8
West Europe	1.2	-1	0	-1	-26 ± 14
East Europe	5.3	1	2	-1	-570 ± 48
South America	17.8	-4	2	-5	-370 ± 130
North America	19.3	2	3	-1	-2302 ± 321
Central America	2.7	1	2	-1	-151 ± 25
North Africa	7.7	1	1	0	-150 ± 30
South Africa	9.8	2	3	-2	613 ± 86
Central Africa	6.6	1	3	-6	488 ± 47
East Africa	7.7	0	1	-1	737 ± 52
West Africa	6.0	4	4	-1	690 ± 120

Figure 4.19: Inter-annual trend of TWSA estimated from MC-SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over sub-continents. The gain period is shown by the blue color, and the loss period by red. The standard deviation of the trend estimation is shown with gray error bars, representing the uncertainty of the trend.

Surface Water (SW) constitutes a major part of the TWS change as a component of the total water storage. Therefore, to understand the TWS loss and gain estimated from GRACE(-FO), we have compared the GRACE-TWSA with the SW time series obtained from satellite imagery. Figure 4.21 compares the mean annual TWSA observed by GRACE(-FO) with the time series of the SW during 2003–2020 over the selected sub-continents. It should be noted that we have not included the change of the water volume in the river system. To measure the level of agreement, we have determined the correlation between the SW and the annual mean of the GRACE-TWSA for the whole period of study over the selected sub-continents (Table 4.8). The time series of the TWSA agrees well (r>0.7) with the time series of SW over Australia, Central Asia, East Europe, South America, South Africa, Central Africa, and East Africa. The highest correlation is observed in Australia (r=0.9) followed by Central Asia (r=0.86) and East Africa (r=0.85).

TWSA and SW show a negative correlation over the Indian sub-continent, North Europe, West Europe, North America, and North Africa. Such a negative correlation can be explained by the fact that the surface water over such regions carries an insignificant role in the TWS variations. For example, TWS dynamics are mainly governed by snow and glaciers melting in North Europe and North America and groundwater extraction is the main factor of the TWS change over the Indian sub-continent. The TWS over North Africa primarily represents the variations of the Sahara Desert, where the surface water at its highest range is as low as 16 km^2. In West Europe, the total variation of the surface water within the study period was insignificant (<1 % of the total surface water). It is noteworthy that for the lakes and reservoirs, surface water variations represent a portion of the total surface water variations as the water change over the river systems can also contribute significantly, especially in regions with large river networks like Amazon in south America and Nile in Africa.

The surface water from lakes and reservoirs can be divided into two main categories: regulated and non-regulated. The former includes reservoirs and regulated lakes, while the latter contains only natural lakes. The time series of the regulated and non-regulated surface water is shown for each of the selected sub-continent within the inside plot (Figure 4.21). Figure 4.20 shows the average ratio of non-regulated versus regulated with respect to the total surface water over the selected regions. The higher the ratio of the regulated, the more possibility for surface water resource management, especially during flood events. Therefore, the surface water in regions like Australia, the Indian sub-continent, East Europe, North Africa, and West Africa has the potential to be managed via reservoirs and regulated lakes. Such a potential is limited in Central Asia, North Europe, West Europe, South Africa, Central Africa, and East Africa.

Table 4.8: Correlation between the annual mean of the TWSA and SW over the selected sub-continents.

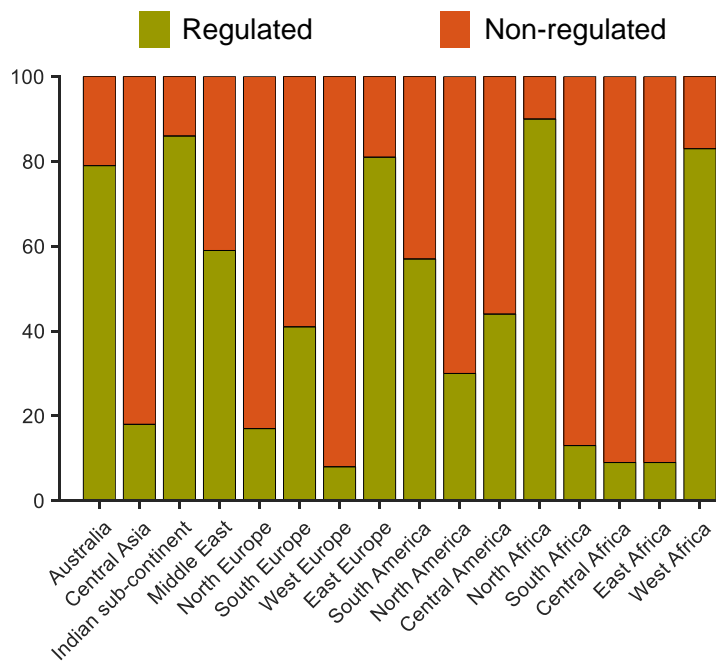| Region	r
Australia	0.90
Central Asia	0.86
Indian sub-continent	−0.28
Middle East	0.38
North Europe	−0.78
South Europe	0.35
West Europe	−0.30
East Europe	0.59
North America	0.79
Central America	−0.59
South America	0.12
North Africa	−0.40
South Africa	0.69
Central Africa	0.67
East Africa	0.85
West Africa	0.47

Figure 4.20: Average volume percentage of the regulated (reservoirs and regulated lakes) versus non-regulated (natural lakes) over the sub-continents from 2003 to 2020.

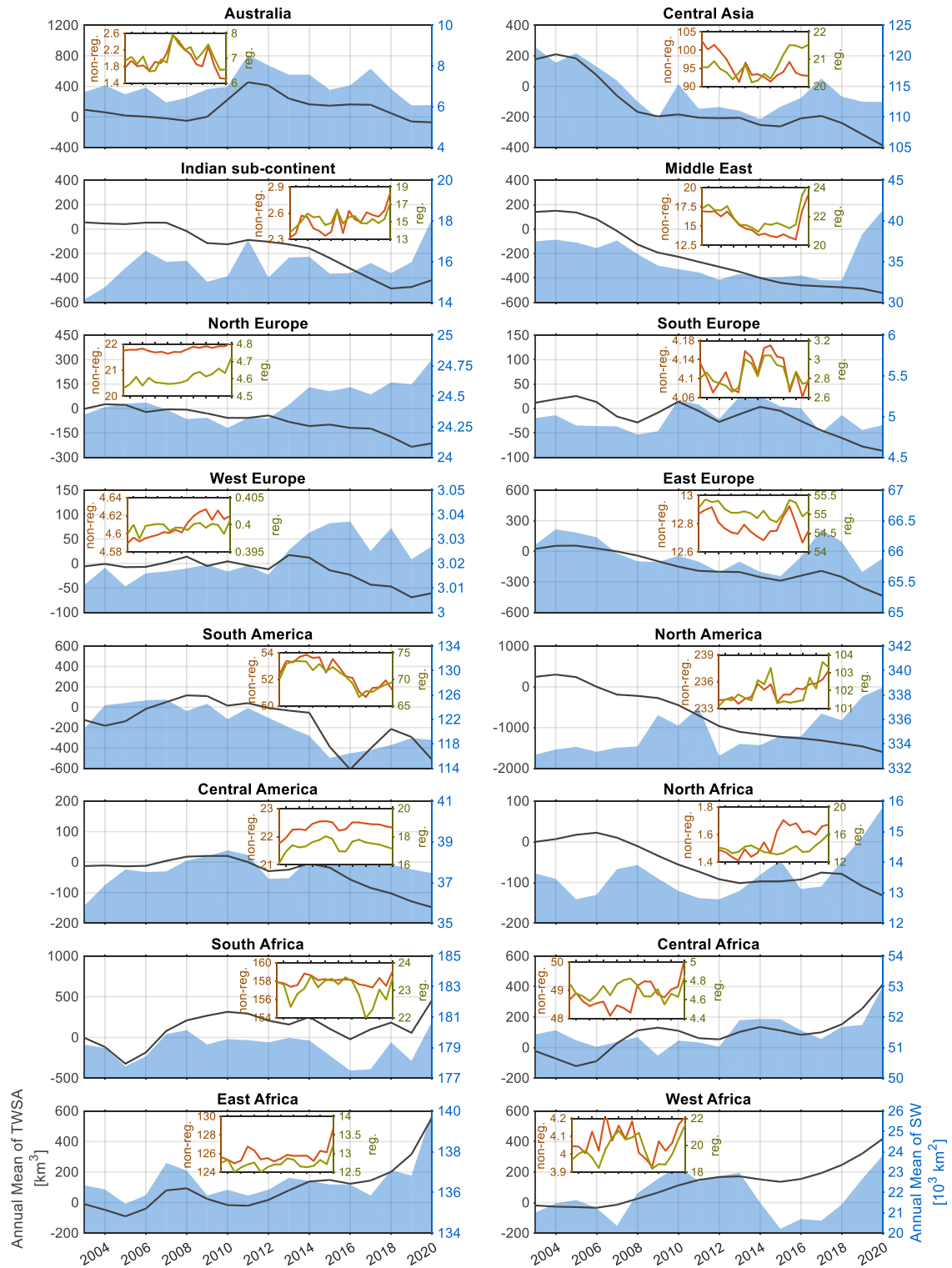Figure 4.21: Annual mean of GRACE-TWSA from 2003 to 2021 over sub-continents versus the annual mean of the surface water from lakes and reservoirs. The time series of the regulated surface water (reservoirs and regulated lakes) and the time series of the non-regulated lakes for each region is shown inside the plots. The x axis in the inside plot is time and share the same range as the main plot.

4.6.2. Analysis over major river basin

Africa

Thirteen major river basins are selected over Africa with area larger than 0.2×10^6 km^2. The distribution of the selected basins over Africa is shown in Figure 4.22. The area of the basins varies from 0.23×10^6 km^2 (Shebelle) to 3.62×10^6 km^2 (Congo). Figure 4.23 shows the TWS loss or gain obtained from the GRACE(-FO) observations with respect to the state in 2003 over the basins. Except for two basins, Limpopo and Northern Sahara, all other basins are gauged at a higher level with respect to their stand in 2003. The maximum gain is observed over Congo (422 $\pm$ 27 km^3) followed by Nile (399 $\pm$ 20 km^3) and Niger (289 $\pm$ 13 km^3). The basins like the Nile, Niger, Senegal, Volta, Shebelle, Congo, and Lake Chad have an overall continuous and gradual water gain. In contrast, basins like Ogooue, Zambezi, Orange, and Western Sahara despite an overall gaining status in 2021 has varied between higher and lower level within 2003–2021.

Only two major basins in Africa gained water with respect to their level in 2003, namely northern Sahara and Limpopo. Since 2003 the northern Sahara has witnessed a strong negative trend (-10 ± 0.8 km^3/yr), except for a weak positive trend from 2003 to 2005 (2.9 ± 0.8 km^3/yr), and has lost 289 ± 13 km^3 of its TWS. On the other hand, the TWS of the Limpopo basin has oscillated between positive and negative trends over the last two decades. Despite positive trend periods since 2003, the basin was only gauged positive from early 2009 to the end of 2012. The same periodic behavior is observed over the adjacent basins, namely Orange and Okavango. However, those basins were mostly leveled above their status in 2003.

1. Northern Sahara
2. Western Sahara
3. Senegal
4. Volta
5. Niger
6. Lake Chad
7. Nile
8. Ogooue
9. Congo
10. Highland of Ethiobia
11. Shebelle
12. Somalia
13. Zambezi
14. Ocavango
15. Limpopo
16. Orange

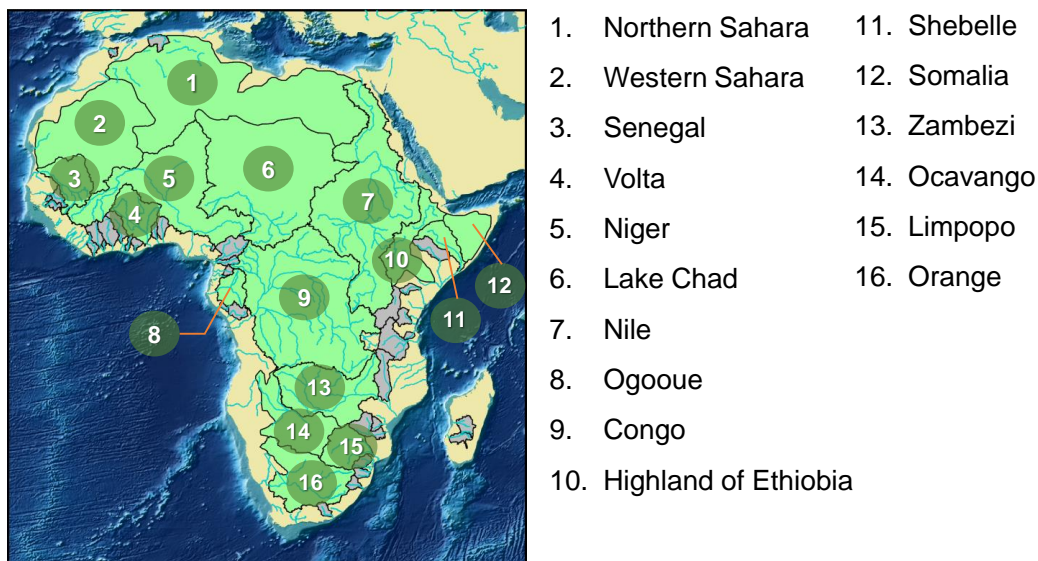Figure 4.22: Geographic location of the selected basins over Africa. The gray regions are among the 405 major river basins defined by the GRDC but with area < 200 000 km^2.
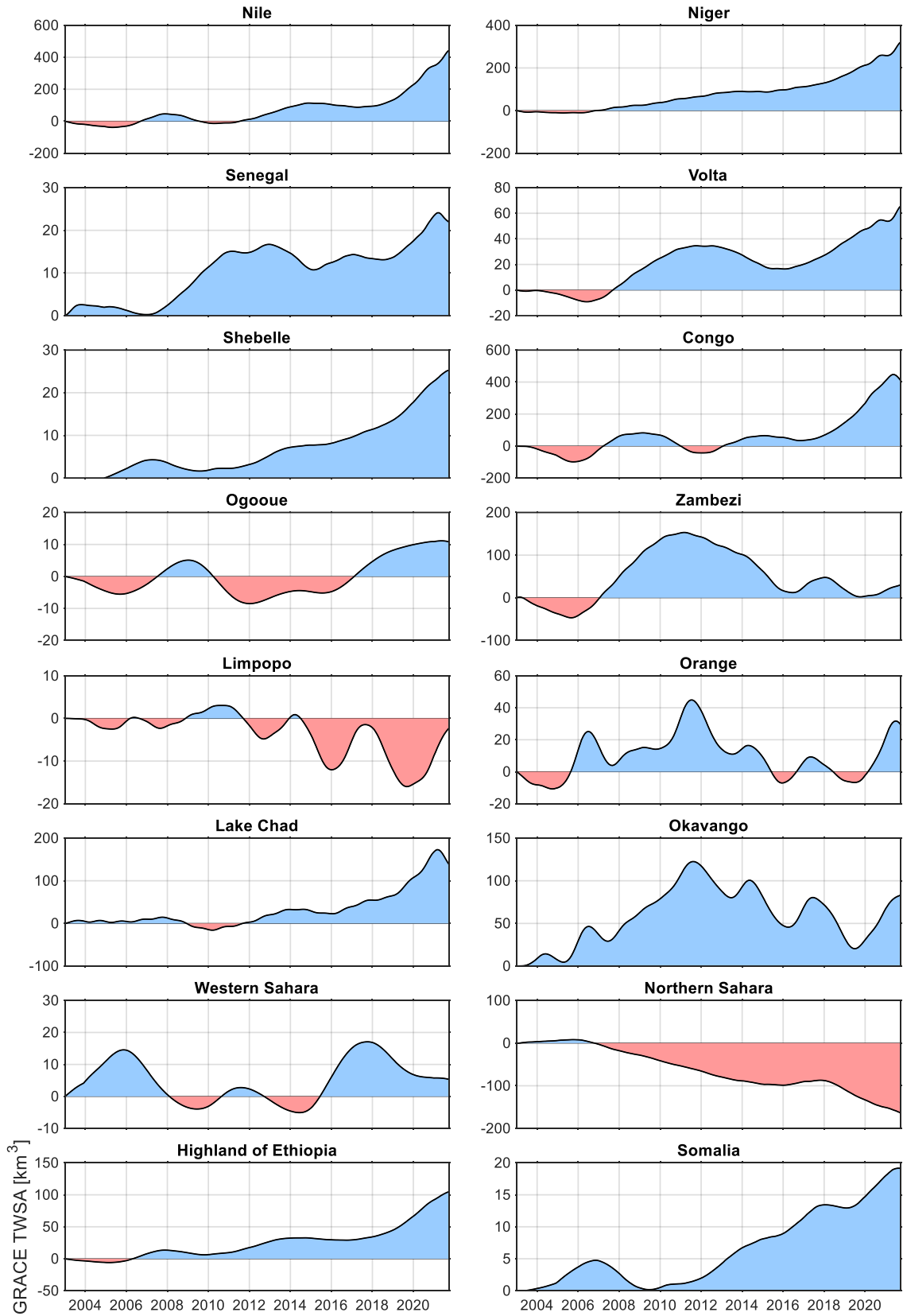
Figure 4.23: Inter-annual TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area ≥ 200 000 km²) of Africa. The gain period is shown in blue, and the loss period in red.

Table 4.9 presents the TWS gain (loss) since 2003 as observed by GRACE(-FO) within the period of 2003–2021, along with the relative water fluxes estimated for the same period. The relative water fluxes were obtained with respect to the 1984–2002. The relative increase (decrease) within the study period with respect to the long-term mean is indicated in parentheses. The Congo basin experienced the greatest decline in precipitation (94×10^3 km^3), followed by the Nile (35×10^3 km^3) and Niger (24×10^3 km^3) basins. Total precipitation over the study period remained relatively unchanged in most basins, except for the Nile, Niger, Senegal, Zambezi, Lake Chad, and Okavango. The only region with less precipitation compared to the long-term was the Ogooue basin. In terms of evapotranspiration, most basins experienced an increase of 1×10^3 km^3 over the study period, with the largest increase occurring in the Niger basin. In some regions, such an increase compensated for the rise in total precipitation, as observed in the Nile, Niger, Zambezi, and Okavango. The Congo basin exhibited the highest total ET among the selected African basins (73×10^3 km^3), followed by the Nile (29×10^3 km^3) and Niger (18×10^3 km^3) basins. Consistent with the results of the precipitation analysis, most basins (eleven basins) had evapotranspiration less than 8×10^3 km^3.

The primary discharge from the significant basins in Africa is predominantly from the Congo basin (28×10^3 km^3), trailed by the Nile (7×10^3 km^3) and the Niger (5×10^3 km^3) basins. A few of the basins (six basins) are dischargeless (DL), including Shebelle, Limpopo, Orange, Western and Northern Sahara, and Somalia. The overall discharge from the Congo basin has decreased significantly (-4×10^3 km^3). The total discharge from the major basins of Africa has slightly decreased (-2×10^3 km^3) in comparison to the 1984–2002 period. This decline happened despite an increase in the total precipitation (7×10^3 km^3), which can be partially explained by the increase in the overall ET of 11×10^3 km^3.

Table 4.9: Summary of total water fluxes and changes in TWS across the major basins of Africa. The relative values for the fluxes are derived by subtracting the total amount between 1984 and 2002 from the total amount between 2003 and 2021. Regarding TWS, the relative status indicates the level in 2021 compared to the level in 2003.

Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3]	rel. TWS [km^3]
Nile	3.05	1	1	1	399 ± 20
Niger	2.10	2	2	0	289 ± 13
Senegal	0.42	1	0	1	23 ± 2
Volta	0.40	0	1	0	59 ± 3
Shebelle	0.23	0	0	0	24 ± 1
Congo	3.62	0	1	−4	422 ± 27
Ogooue	0.21	−1	0	−1	9 ± 3
Zambezi	1.12	1	1	0	42 ± 14
Limpopo	0.34	0	0	0	-4 ± 3
Orange	0.83	0	0	0	28 ± 8
Lake Chad	3.18	2	1	0	158 ± 23
Okavango	0.74	1	1	1	79 ± 9
Western Sahara	1.81	0	1	0	8 ± 6
Northern Sahara	3.47	0	1	0	-159 ± 12
Highland of Ethiopia	0.57	0	1	0	100 ± 3
Somalia	0.37	0	0	0	19 ± 1

Australia

The basins exhibit a wide variation in their areas, ranging from 1.2×10^6 km^2 for Murray Darling to 2.04×10^6 km^2 for West Australia. Furthermore, the figure 4.25 illustrates the trend in TWS over these basins since 2003. The TWS data reveals that Murray Darling basin experienced a significant increase in TWS, with a net gain of 31 ± 6 km^3 compared to its status in 2003. On the other hand, East and West Australia basins have undergone a loss in TWS of around -24 ± 11 km^3 and -63 ± 11 km^3, respectively (4.10).

Although the final TWS levels in the basins varied at the end of the study period, the dynamic changes in TWS across the major basins in Australia exhibited a consistent pattern. TWS levels in all the basins have been decreasing since 2003, with declines ending in 2007, 2009, and 2010 in the Murray Darling (-10 ± 1 km^3/yr), East Australia -4.5 ± 0.7 km^3/yr, and West Australia -16.5 ± 0.8 km^3/yr, respectively. Subsequently, there was an increase in TWS levels, which ended in early 2012, end of 2010, and early 2012 over the Murray Darling (33.7 ± 2 km^3/yr), East Australia 75.7 ± 7.5 km^3/yr, and West Australia 22.9 ± 6.6 km^3/yr, respectively. Following this gaining period, the East and West Australia basins started to lose water, which continued until the end of 2021, resulting in a negative trend of -12.6 ± 0.8 km^3/yr and -4.9 ± 0.9 km^3/yr, respectively. Similarly, the Murray Darling experienced a negative trend from 2013 to 2018 (-5.5 ± 1 km^3/yr). However, within the last two years of the study period (2019–2021), there has been a continuous positive trend of TWS recovery (28.8 ± 1.5 km^3/yr), and the basin has returned to its late 2016 status.

The table in Table 4.10 provides a summary of the estimated total water fluxes and volumetric TWS gain or loss since 2003. The fluxes are consistent across the three basins. Total precipitation and evapotranspiration over the basins are estimated to be within the same range of 8–11×10^3 km^3. Regarding runoff, Murray Darling and West Australia align with their estimates from the pre-GRACE era (1984–2002). However, the runoff from East Australia has increased by 1000 km^3.

In terms of water flux balance with respect to the same period (1983–2002), there has been a slight decrease in Murray Darling by -1×10^3 km^3, an increase in West Australia by the same amount, and no change in East Australia. The GRACE(-FO) observations indicate water loss over West Australia, but this finding is not aligned with the water flux balance over Murray Darling and East Australia. Such discrepancies can be explained by the limited accuracy of the flux estimates in capturing the TWS change, which has been previously discussed in Figure 4.4.1.

1. West Australia
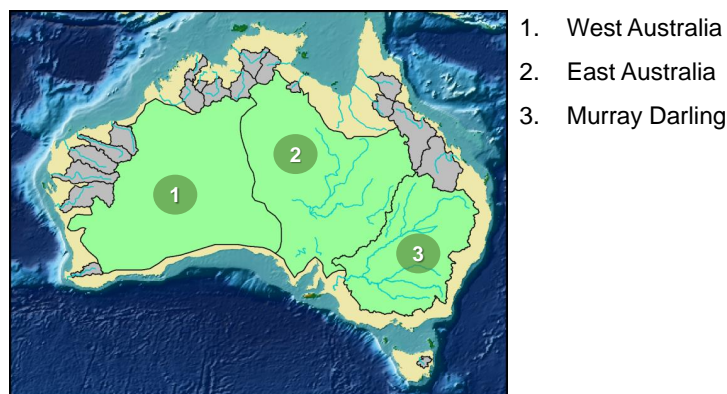2. East Australia
3. Murray Darling

Figure 4.24: Geographic location of the selected basins over Australia. The gray regions are among the 405 major river basins defined by the GRDC but with area $< 200\,000$ km^2.
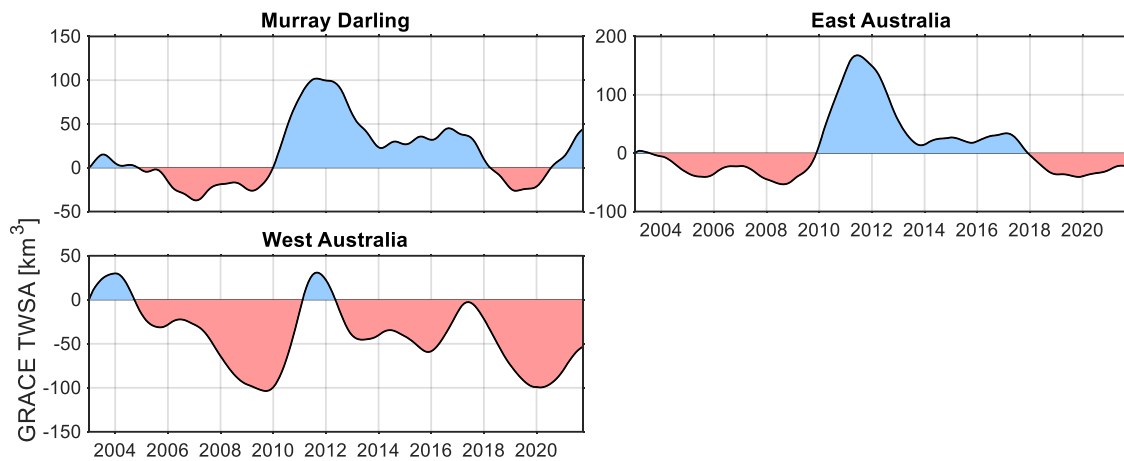
Figure 4.25: Inter-annual trend of TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area ≥ 200 000 km^2) of Australia. The gain period is shown by the blue color, and the loss period by red.

Table 4.10: Summary of total water fluxes and changes in TWS across the major basins of Australia. The relative values for the fluxes are derived by subtracting the total amount between 1984 and 2002 from the total amount between 2003 and 2021. Regarding TWS, the relative status indicates the level in 2021 compared to the level in 2003.

Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3	rel. TWS
Murray Darling	1.02	−1	0	0	31 ± 6
East Australia	1.84	1	0	1	−24 ± 11
West Australia	2.04	0	1	0	−63 ± 11

North America

The spatial distribution of the eleven major basins across North America is illustrated in Figure 4.26. The size of these basins ranges from 0.3×10^6 km^2 in Churchill basin to 2.94×10^6 km^2 in Mississippi. The corresponding temporal trends of TWS gain or loss since 2003 over these selected basins are presented in Figure 4.27. Among these basins, only two, namely St. Lawrence (137 ± 19 km^3) and Mississippi (132 ± 38 km^3), have exhibited a net gain in TWS since 2003. On the other hand, the Yukon basin shows the maximum water loss of -199 ± 10 km^3, followed by Mackenzie (-105 ± 17 km^3) and Death Valley (-74 ± 3 km^3).

The relative TWS changes over Colorado, Bravo, Death Valley, Churchill, and Fraser can be segregated into two distinct phases. These basins have experienced a net gain in water in the first phase, primarily from 2003 to 2011. However, all basins have witnessed a significant negative trend in TWS in the second phase, resulting in a considerable decline in the TWS. The Nelson basin has recorded a consistently higher TWS compared to its status in 2003, while the TWS in the Yukon and Mackenzie basins has shown a decline compared to the TWS level in 2003. It is noteworthy that the observed discrepancies between the TWS loss/gain and water fluxes over the basins can be attributed to the limitations of flux estimates in predicting TWS variations, as discussed in Figure 4.4.1.

1.	Yukon	5.	Death Valley	9.	Nelson
2.	Mackenzie	6.	Colorado	10.	Mississippi
3.	Fraser	7.	Bravo	11.	St. Lawrence
4.	Columbia	8.	Churchill		

Figure 4.26: Geographic location of the selected basins over North America. The gray regions are among the 405 major river basins defined by the GRDC but with area $< 200\,000\,\text{km}^2$.

Table 4.11 presents a summary of the estimated total water fluxes and volumetric TWS gain (or loss) since 2003 for the major basins of North America. Mississippi received the maximum precipitation (42×10^3 km^3) followed by St. Lawrence (16×10^3 km^3), Mackenzie (12×10^3 km^3), and Nelson (10×10^3 km^3). In contrast, Colorado, Death Valley, Churchill, and Fraser received the minimum amount of precipitation (3×10^3 km^3). Most basins have gained the same amount of water from precipitation as in the previous period (1984–2002). However, the total precipitation has relatively increased ($+1 \times 10^3$ km^3) in three basins since 1984–2002, while the Colorado river basin has experienced a 25 % reduction in precipitation. Notably, basins with high precipitation values also have the highest ET rates.

In terms of evapotranspiration, more than half of the basins have remained unchanged compared to the previous period (1984–2002). However, three basins, namely St. Lawrence, Colorado, and Death Valley, have shown a relatively lower ET rate (-1×10^3 km^3) over the last two decades. Concerning runoff, most basins discharged the same amount of water from 2003 to 2020 as they did from 1984–2002. However, Columbia has experienced a relative total runoff of 33 % less than the long-term average, while four basins, namely Colorado, Bravo, Death Valley, and Churchill, have shown discharge-less behavior. The considerable contribution of ice sheet loss explains the 66 % increase in runoff above the long-term average in some basins. Finally, the balance in the total water fluxes ($P - ET - R$) indicates significant water storage loss in the St. Lawrence basin. On the other hand, the water storage in Colorado, Mackenzie, Churchill, Nelson, and Fraser basins has remained unchanged compared to the previous period (1984–2002).
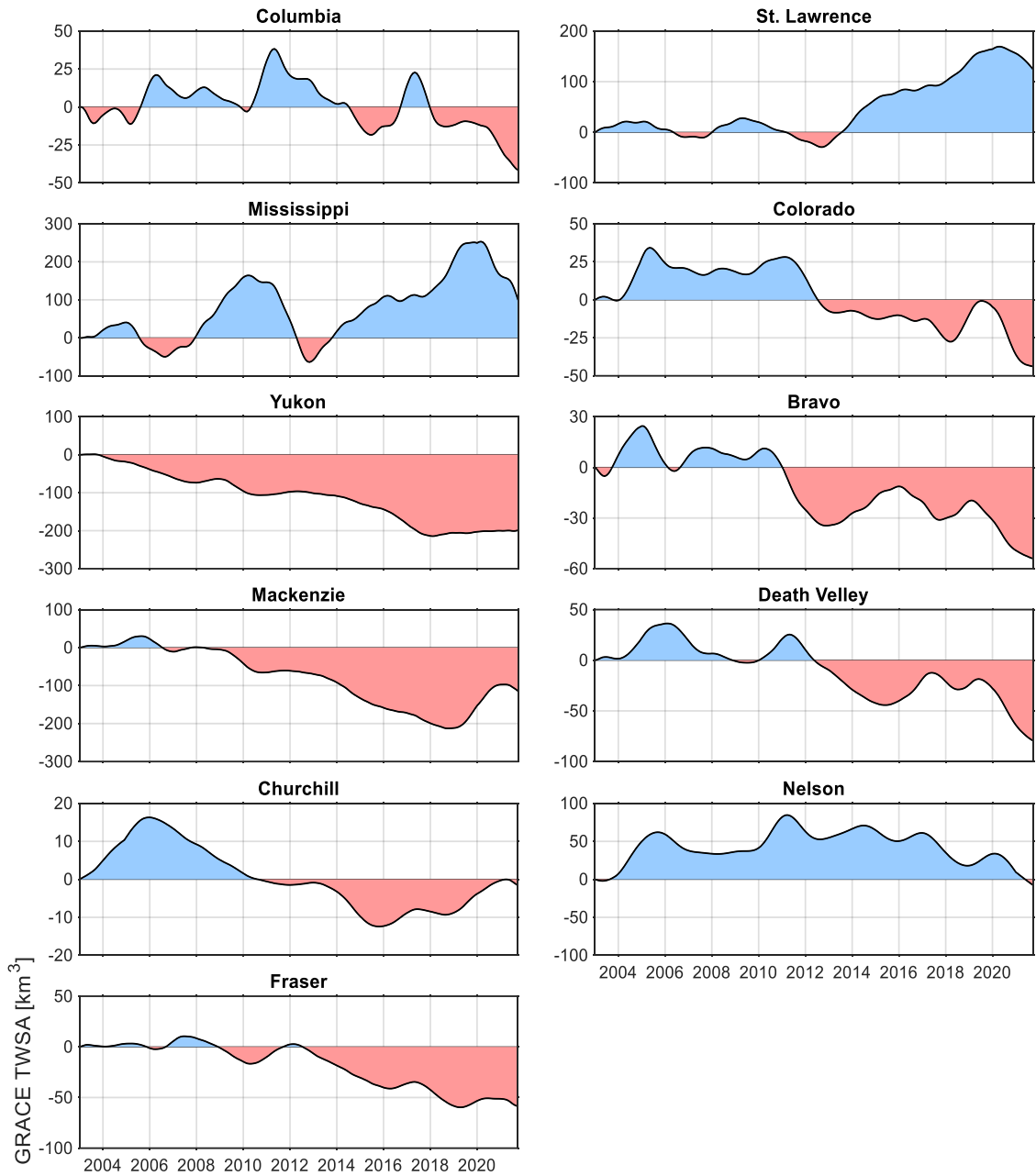
Figure 4.27: Inter-annual trend of TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area ≥ 200 000 km^2) of North America. The gain period is shown by the blue color, and the loss period by red.

Table 4.11: Summary of total water fluxes and changes in TWS across the major basins of North America. The relative values for the fluxes are derived by subtracting the total amount between 1984 and 2002 from the total amount between 2003 and 2021. Regarding TWS, the relative status indicates the level in 2021 compared to the level in 2003.

Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3	rel. TWS
Columbia	0.66	0	0	−1	−36 ± 6
St. Lawrence	0.94	1	−1	4	137 ± 19
Mississippi	2.94	1	1	0	132 ± 38
Colorado	0.64	−1	−1	0	−42 ± 3
Yukon	0.82	1	0	0	−199 ± 10
Bravo	0.51	0	1	0	−52 ± 2
Mackenzie	1.67	0	0	0	−105 ± 17
Death Valley	0.55	0	−1	0	−74 ± 3
Churchill	0.30	0	0	0	−1 ± 3
Nelson	1.13	0	0	0	−29 ± 49
Fraser	0.23	0	0	0	−55 ± 2

South America

Figure 4.28 illustrates the distribution of the selected major river basins over South America. The area of the selected basins varies from 0.24×10^6 km^2 in Uruguay to 4.67×10^6 km^2 in Amazonas. Figure 4.29 shows the TWS loss or gain since 2003 over the selected basins. Since 2003 most of the basins in Asia have leveled below their status in 2003. The maximum water loss is observed over the Parana basin (-184 ± 59 km^3) followed by Sao Francisco (-117 ± 4 km^3). Three basins indicate an overall gain in TWS, including Magdalena, Orinoco, and Amazonas. While some basins like Colorado (Argentinia) and the Atacama mostly leveled below their status in 2003, others like Amazonas, Rio Tapajos, and Parana have been oscillating between water loss and water gain over the study period. Five basins have experienced almost two phases in their TWS change. Some of these basins, including Xingu, Tocantins, Sao Francisco, and Rio Parnaiba, have witnessed first the positive level for almost half of the study period while started to lose water for the second half. The Uruguay basin, in contrast, has to experience phases contrariwise over the study period.

The estimation of the total water fluxes since 2003 over the selected major basins of South America is shown in Table 4.12. The total precipitation varies from as low as in Colorado to as high as in Amazonas. Considering all the basins, the same amount of precipitation has fallen over the study period compared with 1984–2002. A slight increase of 2 % is observed over Amazonas while two basins, namely Tocantins and Parana, have observed slightly (~ 3.5 %) less water with respect to the 1984–2002 period. The Uruguay basins have received considerable (~ 15 %) less water from precipitation over the last two decades.

In terms of ET, on average 70 % of the precipitation has been evaporated from the basins ranging from 54 % in Orinoco to 100 % in Colorado (Argentinia). The total ET has slightly increased (2×10^3 km^3) over the whole region of South America while most of the basins (~ 75 %) remained unchanged with respect to 1984–2002. Only one basins, Colorado, is dischargeless while the runoff over the basins varies from 1×10^3 km^3 in Atacama to 88×10^3 km^3 in Amazonas. Other than the Amazonas basin, two other basins namely Orinoco (17×10^3 km^3) and the Parana (14×10^3 km^3) also have discharged significant share of their water gain from precipitation. The total discharge from the selected basins in South America shows no significant

change with respect to 1984–2002. However, a slight increase (~ 4 %) and decrease (~ 1 %) is observed over Amazonas and Orinoco, respectively.
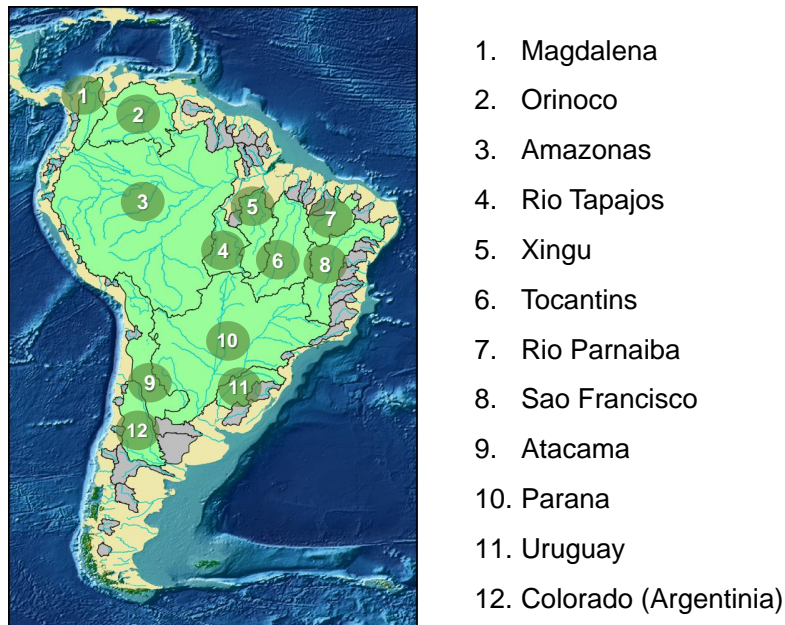

Figure 4.28: Geographic location of the selected basins over South America. The gray regions are among the 405 major river basins defined by the GRDC but with area $< 200\,000\,\text{km}^2$.

Table 4.12: Summary of the total water fluxes and the TWS gain(loss) since 2003 over the major basins of South America.

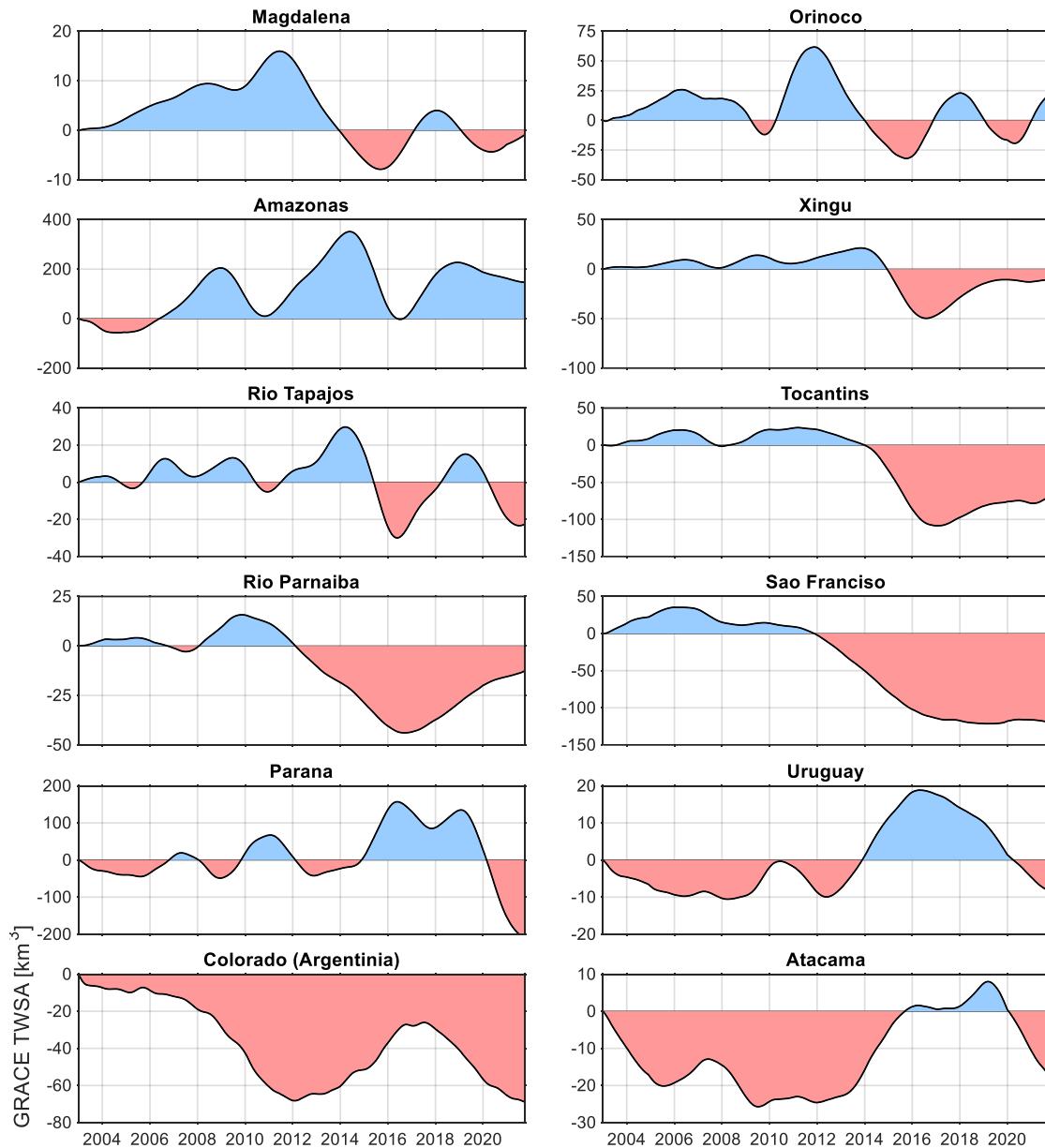| Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3]	rel. TWS [km^3]
Magdalena	0.25	0	0	1	1 ± 5
Orinoco	0.84	0	1	−2	14 ± 23
Amazonas	4.67	4	3	3	152 ± 38
Xingu	0.45	0	0	0	-11 ± 3
Rio Tapajos	0.37	0	0	−1	-19 ± 4
Tocantins	0.75	−1	0	−1	-74 ± 5
Rio Parnaiba	0.30	0	0	0	-14 ± 2
Sao Francisco	0.63	0	−1	0	-117 ± 4
Parana	2.65	−2	−1	−1	-184 ± 59
Uruguay	0.24	−1	0	0	-7 ± 1
Colorado	0.39	0	0	0	-68 ± 3
Atacama	0.65	0	0	0	-15 ± 3

Figure 4.29: Inter-annual trend of TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area $\geq$ 200 000 km^2) of South America. The gain period is shown by the blue color, and the loss period by red.

Asia

The distribution of the selected river basins over Asia is illustrated in Figure 4.30. The area of the basins varies between 0.25×10^6 km^2 in Krishna and 2.40×10^6 km^2 in Saudi Arabia. Figure 4.31 and Figure 4.32 show the TWS loss or gain since 2003 over the selected basins. Since 2003 most of the basins in Asia have lost their TWS. The maximum water loss is observed over the Saudi Arabia basin (-256 ± 9 km^3) followed by Indus (-228 ± 9 km^3) and Ganges (-219 ± 8 km^3). In contrast, six river basins, mostly in China, have gained water since 2003, including Yangtze, Xi Jiang, Godavari, Krishna, Highland of Tibet, and Amur. These basins have gained in total 366 ± 19 km^3 since 2003 in which most of this gain ($\sim 80\%$) is occurred over two major basins namely Yangtze (113 ± 13 km^3) and Amur (181 ± 14 km^3).

The temporal change in the relative TWS over the selected major basin of Asia can be categorized into four groups. In the first group, the basins have been leveled in water loss since 2003 for most of the study period. Such basins include Indus, Brahmaputra, Ganges, and Saudi Arabia. The Syr Darya, Amu Darya, Yellow River, Euphrates, Tarim, North China, Gobi, Helmand, Karakum, and the Aral Sea are among the second groups has observed a short gaining period followed by a significant negative trend which resulted in a considerable TWS over these basins. The length of the gaining period varies over these basins from one to two years in the basins like Yellow River, Euphrates, and North China to even nine years for Karakum. In the third group, which includes basins like Amur, Yangtze, Xi Jiang, and Highland of Tibet, the basins mostly remained above their status in 2003. Finally, the last group belongs to the basins that have observed a considerable fluctuation in their TWS since 2003, including Lake Balkhash, Godavari, and Krishna.
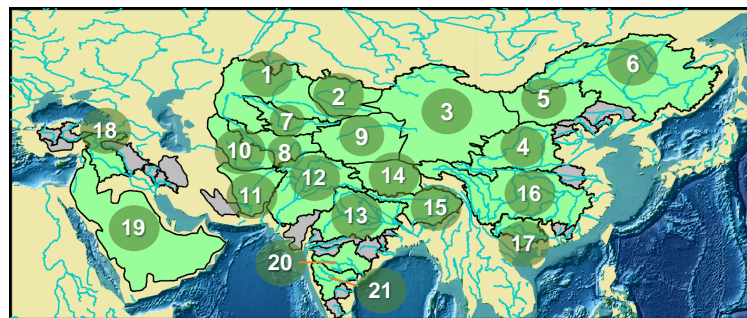
1.	Aral Sea	8.	Amu Darya	15.	Brahmaputra
2.	Lake Balkhash	9.	Lake Balkhash	16.	Yangtze
3.	Gobi	10.	Tarim	17.	Xi Jiang
4.	Yellow River	11.	Karakum	18.	Euphrates
5.	North China	12.	Helmand	19.	Saudi Arabia
6.	Amur	13.	Indus	20.	Godavari
7.	Syr Darya	14.	Highland of Tibet	21.	Krishna

Figure 4.30: Geographic location of the selected basins over Asia. The gray regions are among the 405 major river basins defined by the GRDC but with area $< 200\,000\,\text{km}^2$.

4

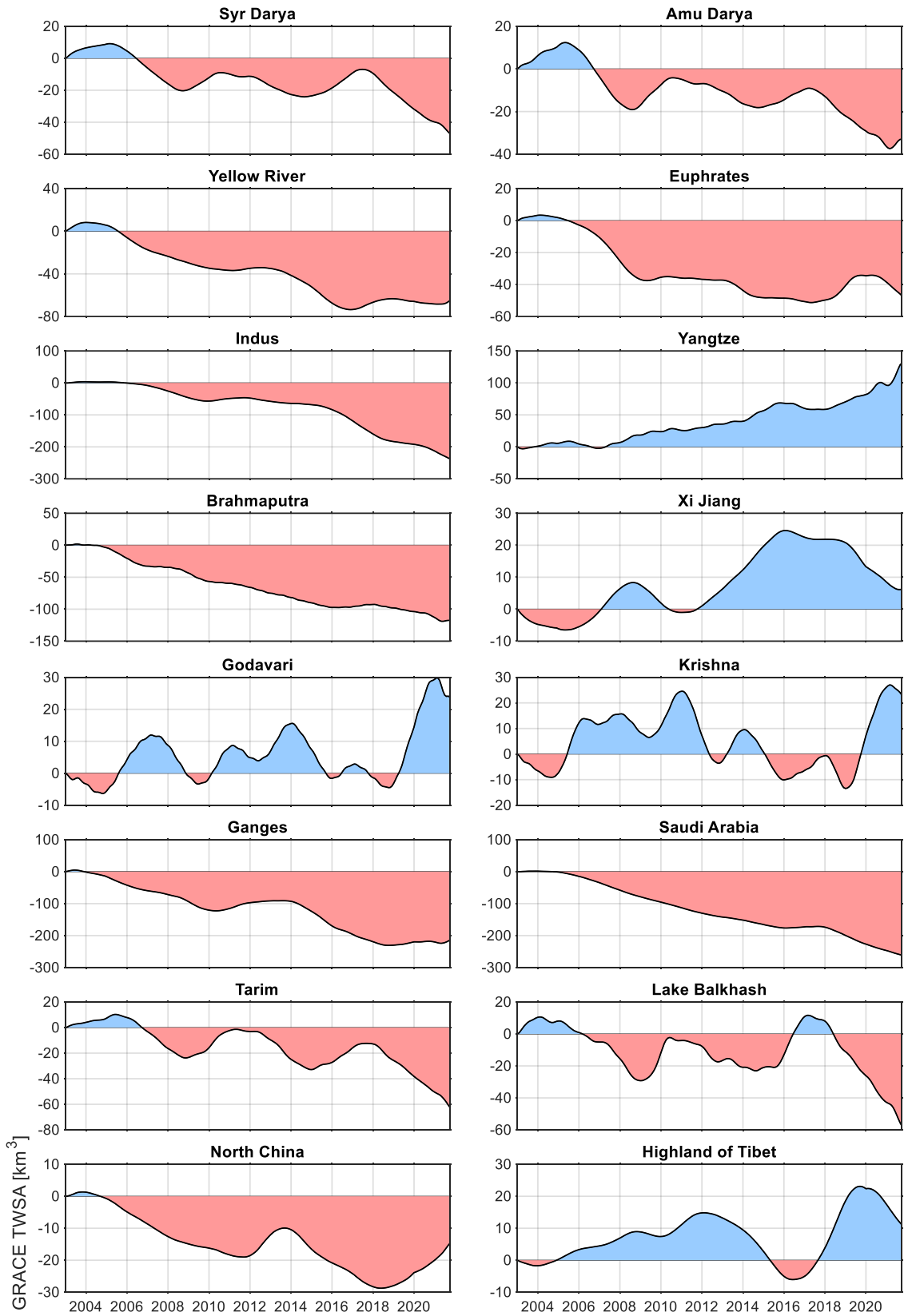

Figure 4.31: Inter-annual trend of TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area ≥ 200 000 km^2) of Asia. The gain period is shown by the blue color, and the loss period by red.
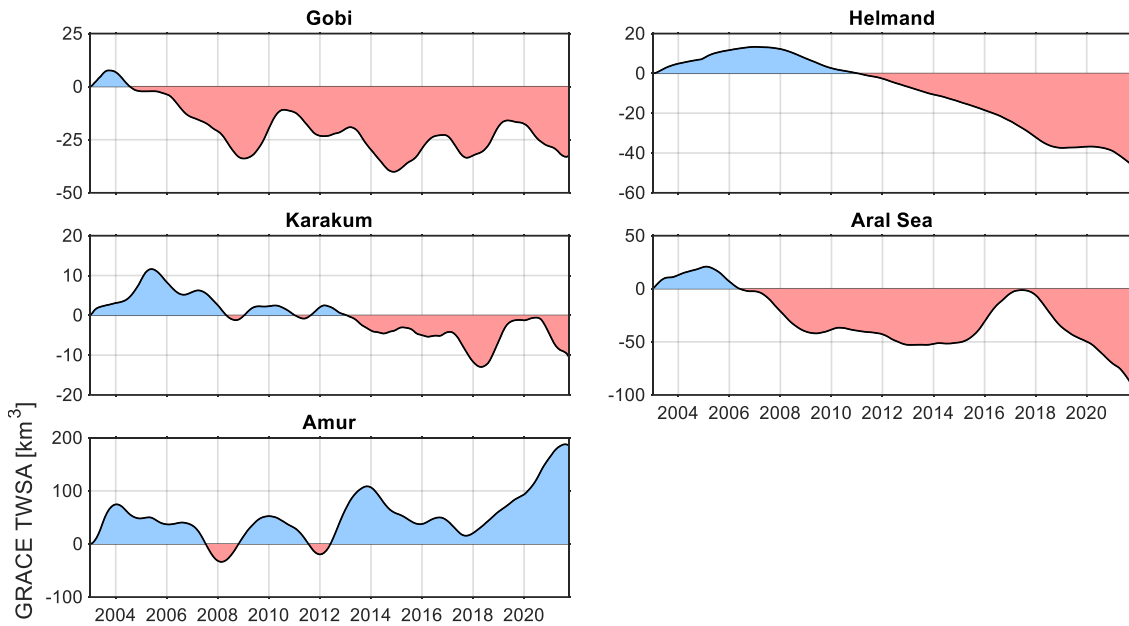
Figure 4.32: Figure 4.31 continued.

To put the observation of the GRACE-TWSA into perspective, we have estimated the total amount of the water balance fluxes over the selected major basins of Asia (Table 4.13). The precipitation varies from as high as 32×10^3 km^3 over Yangtze to as low as 1×10^3 km^3 over Helmand. About 60 % of the basins have received less than 5×10^3 km^3. In contrast, four basins, namely Yangtze, Ganges, Brahmaputra, and Amur, have received more than 50 % of the total precipitation over Asia. The relative precipitation with respect to 1984–2002 shows an increase of 0.5 %. Only six major basins have observed this gain in their input from precipitation, namely Yellow River, Indus, Lake Balkhash, Gobi, Karakum, and Amur.

In terms of ET, on average 80 % of the precipitation over the basins in Asia has been evaporated, ranging from ~ 65 % over Highland of Tibet to 100 % over nine basins like the Syr Darya and Aras Sea. The relative ET shows an increase in eight basins, each only 1×10^3 km^3. In some basins like Indus, Lake Balkhash, and Amur, the increase in the ET has compensated for the slight increase in the precipitation, which results in the same recharge ($P - ET$) over these basins with respect to the 1984–2002 period.

A great number of the basins (~ 50 %) are dischargless (DL) in Asia. Moreover, five basins also discharge a small amount of water (1×10^3 km^3) including Syr Darya, Amu Darya, Yellow River, Krishna, and Highland of Tibet. In contrast, a considerable ratio of the precipitation has been discharged from Yangtze (~ 40 %), Brahmaputra (~ 60 %), Ganges (~ 30 %), and Ganges (~ 20 %). With respect to the period of 1984–2002, the basins of Asia have mostly remained unchanged in terms of runoff. Only a slight increase ($+1 \times 10^3$ km^3) and decrease (-1×10^3 km^3) is observed over Yangtze and Ganges river basins, respectively.

The balance in the water fluxes represents the water storage change in a basin. Based on the results from the fluxes, most of the basins show no change (zero water balance) in their water storage. In three basins, namely Syr Darya, Amu Darya, and the Brahmaputra, the result from the water balance is aligned with the observation of the GRACE. However, the Yangtze and Amur river basins show a negative balance over the study period, while the GRACE observation reveals a significant gain in the basin. Such discrepancies can be explained by the low performance of the fluxes datasets over Amur while the emergence of the new lakes and

restoration of the others in the Yangtze river basins (e.g., Huang et al., 2019; Li et al., 2020; Yang & Lu, 2014).

Table 4.13: Summary of the total water fluxes and the TWS gain(loss) since 2003 over the major basins of Asia.

Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3]	rel. TWS [km^3]
Syr Darya	0.35	0	0	0	-44 ± 2
Amu Darya	0.43	0	0	0	-35 ± 2
Yellow River	0.90	1	0	0	-67 ± 4
Euphrates	0.33	0	1	0	-44 ± 5
Indus	1.12	1	1	0	-228 ± 9
Yangtze	1.68	0	1	1	113 ± 13
Brahmaputra	0.52	0	1	0	-117 ± 7
Xi Jiang	0.30	0	0	0	7 ± 2
Godavari	0.31	0	0	0	26 ± 3
Krishna	0.25	0	0	0	25 ± 2
Ganges	0.91	0	1	-1	-219 ± 8
Saudi Arabia	2.40	0	0	0	-256 ± 9
Tarim	0.88	0	0	0	-57 ± 7
Lake Balkhash	0.48	1	1	0	-50 ± 4
North China	0.62	0	0	0	-17 ± 3
Highland of Tibet	0.52	0	0	0	14 ± 2
Gobi	2.10	1	0	0	-31 ± 11
Helmand	0.52	0	0	0	-43 ± 2
Karakum	0.44	1	0	0	-9 ± 6
Aral Sea	1.07	0	1	0	-81 ± 8
Amur	1.95	1	1	0	181 ± 14

Europe

The distribution of the seven major basins over Europe is shown in Figure 4.33. The area of the basins varies from 0.23×10^6 km^2 in Neva to 1.35×10^6 km^2 in Volga. Figure 4.34 shows the TWS loss or gain since 2003 over the selected basins. Four basins have leveled below their status in 2003 ranging from -57 ± 5 km^3 in Danube to -67 ± 6 km^3 in Dniepr. In contrast, three basins in the northeastern of Europe namely Pechora ($+2 \pm 2$ km^3), North Dvina ($+8 \pm 1$ km^3), and Neva ($+9 \pm 1$ km^3) has gain water since 2003. The Danube, Dniepr, and Don basins have experienced the same evolution. In these basins, a short-lived positive trend is observed at the beginning of the study period, followed by a significant negative trend (5.6–7 ± 1 km^3/yr). The Pechora basin has observed periodic gain and loss over the last two decades. The TWS in the Volga basin has remained negative since 2003 after a weak positive trend in 2004. The basin has suffered from two significant negative trends, namely -23.8 ± 1.8 km^3/yr during 2008–2011 and -11.25 ± 1.6 km^3 from 2017 to 2021. The former was recovered partially by the positive trend of $+18 \pm 1.4$ km^3 within 2012–2016.

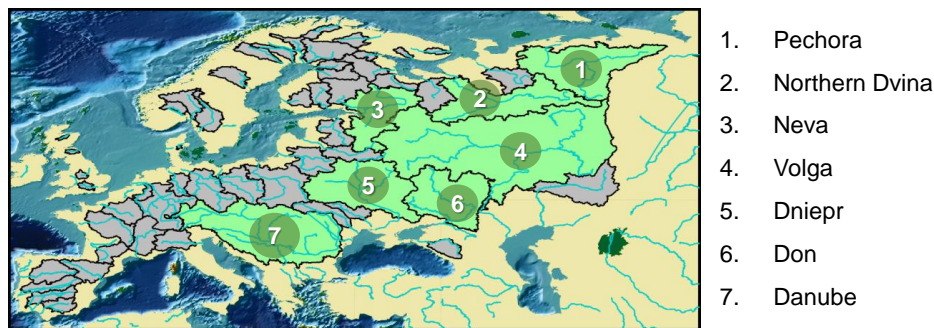|	
1.	Pechora
2.	Northern Dvina
3.	Neva
4.	Volga
5.	Dniepr
6.	Don
7.	Danube

Figure 4.33: Geographic location of the selected basins over Europe. The gray regions are among the 405 major river basins defined by the GRDC but with area < 200 000 km^2.

Table 4.14 summarizes the total amount of water fluxes from 2003 over the major river basins of Europe. The Volga basin receive the maximum amount of precipitation with 15×10^3 km^3 followed by Danube (11×10^3 km^3). Three basins, namely Pechora, Neva, and Don, are gaining the minimum precipitation (3×10^3 km^3) among Europe's major basins. Most of the basins show no change in the total precipitation water with respect to the long-term (1984–2002). Only a small increase (1×10^3 km^3) and small decrease (-1×10^3 km^3) is observed over Danube and Don, respectively.

In terms of ET, the majority of the precipitation (on average $\sim 75\,\%$) has evaporated from the basins. The whole precipitation water has been evaporated over the Don basin while this share is 65 % over Pechora and Neva. The amount of ET over the basins follows the long-term (1984–2002) with the exception of $+1 \times 10^3$ km^3 over the Northern Dvina. The runoff over the basins varies from zero over Don to ($+4 \times 10^3$ km^3) over the Volga. Except for the Volga with a decrease of 20 % of its discharge, all the other basins runoff remained unchanged with respect to the period 1984–2002. The water fluxes balance ($P - ET - R$) shows no storage change over most basins. Only two basins, namely the Pechora and Northern Dvina, have lost water (-1×10^3 km^3) in terms of the fluxes.
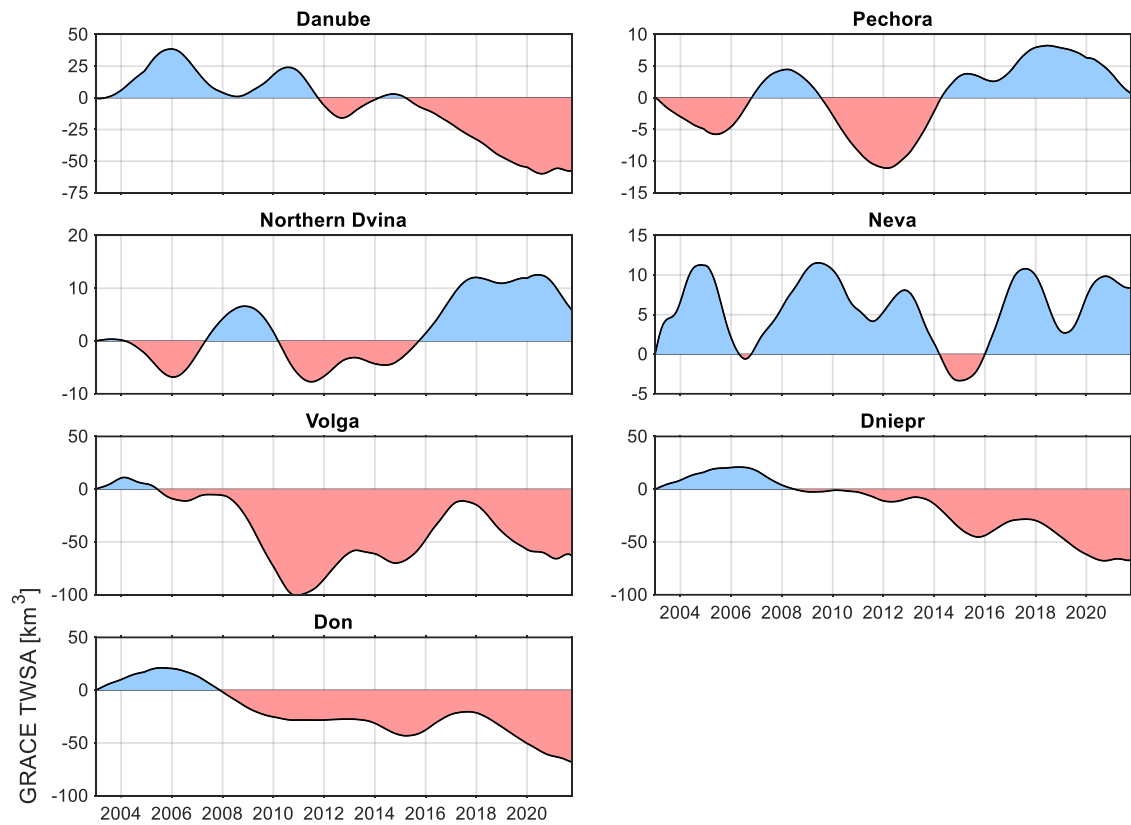
Figure 4.34: Inter-annual trend of TWSA estimated from SSA using the GRACE and GRACE-FO observations from 2003 to 2021 over the major basins (Area ≥ 200 000 km^2) of Europe. The gain period is shown by the blue color, and the loss period by red.

Table 4.14: Summary of the total water fluxes and the TWS gain(loss) since 2003 over the major basins of Europe.

Basin	Area [10^6 km^2]	rel. Total P [10^3 km^3]	rel. Total ET [10^3 km^3]	rel. Total R [10^3 km^3]	rel. TWS [km^3]
Danube	0.77	1	0	0	-57 ± 5
Pechora	0.30	0	0	0	2 ± 2
Northern Dvina	0.33	0	1	0	8 ± 1
Neva	0.23	0	0	0	9 ± 1
Volga	1.35	0	0	-1	-64 ± 5
Dniepr	0.46	0	0	0	-67 ± 6
Don	0.38	-1	0	0	-65 ± 2

4.6.3. Analysis over Iran

Like many other Middle Eastern countries, Iran has been enduring severe water shortage over the last two decades (Dubreuil et al., 2013; Madani, 2014; Michel, 2017; Zehtabian et al., 2010). Drying rivers, lakes and wetlands (Alborzi et al., 2018; Arsanjani et al., 2015; Jones et al., 2015; Madani et al., 2016; Saemian et al., 2020; Tourian et al., 2015), deforestation, desertification, soil erosion, and sand and dust storms are among the visible products of reduced surface water in this nation (Amiraslani & Dragovich, 2011; Khalyani & Mayer, 2013; Khormali et al., 2009; Mardi et al., 2018). Besides surface water loss, groundwater levels have fallen significantly over the past decades (Ashraf et al., 2017; Döll et al., 2014; Gleeson et al., 2012; Marc et al., 2012;

Voss et al., 2013), leading to problems such as land subsidence and the emergence of sinkholes across the country (Emadodin et al., 2012; Motagh et al., 2008). Using point data from wells, the total groundwater depletion in Iran is estimated at around 75 km^3 within 2002–2015 (Ashraf et al., 2021; Noori et al., 2021). The observed changes in water availability have been attributed to the compounding effects of human activities and climatic variability and change (Ashraf et al., 2019). With more than 85 % of the country's area having an arid or semi-arid climate, any significant change in water availability can result in substantial environmental and socio-economic impacts (Madani, 2014), turning water scarcity into a national security threat.

The observed depletion of water resources is the product of frequent meteorological droughts (natural variability), climatic changes, and human activities (increased water use and withdrawal). However, many studies consider anthropogenic activities including population growth (increasing demand), inefficient agricultural water use, and unsustainable water resources management as the primary cause of Iran's water bankruptcy (e.g., Madani, 2014; Madani et al., 2016; Maghrebi et al., 2020; Mirnezami et al., 2018; Mirzaei et al., 2019; Nabavi, 2018; Panahi et al., 2020). The quantitative knowledge on Iran's water bankruptcy problem at the national scale is still very limited due to lack of conclusive ground data. Filling this knowledge gap and developing a deep understanding of the level of water loss across Iran will be an essential step toward addressing this significant national problem.

Estimation of Total Water Storage (TWS), defined as the sum of all storage components such as surface water, soil moisture, snow water, and groundwater, provides valuable insight into the water resources availability at the regional scale (Riegger & Tourian, 2014; Syed et al., 2005; Tourian et al., 2018). Conventional ground-based measurements of TWS components, e.g., change in Surface Water Storage (SWS), Soil Moisture Storage (SMS), and GroundWater Storage (GWS) are being done at local scales. However, ground-based measurements are often associated with data inconsistencies, spatial, temporal and physical data gaps (e.g. unknown storage coefficients), and instrumental and human errors (Forootan et al., 2014; Lorenz et al., 2015; Rodell et al., 2007). Evaluation of TWS can also be done using Land Surface Models (LSMs) or hydrological models. The performance of these models, however, varies in different parts of the world which can make them unreliable for water management and decision-making purposes, especially during extreme events like droughts or floods (e.g., Felfelani et al., 2017; Long et al., 2014; Long et al., 2013).

GRACE-TWS observations have been previously employed for water resources monitoring in Iran over the last two decades. Voss et al. (2013) reported an alarming loss at a rate of 2.7 cm/yr in the north-central Middle East, including the Tigris and Euphrates River Basins and western Iran, from 2003 to 2009. A similar negative trend was also reported in Joodaki et al. (2014) over western Iran and eastern Iraq from 2003 to 2012. By subtracting contributions from soil moisture, snow, canopy storage, and river storage, groundwater depletion was found to represent 60 % of the total volume of water lost (Joodaki et al., 2014; Voss et al., 2013). Merging the Global Land Data Assimilation System (GLDAS) model and satellite altimetry data as a prior data with GRACE-TWS, Forootan et al. (2014) estimated a negative average trend of 1.5 cm/yr over central and northwestern Iran during the 2005–2011 period. A recent study by Rahimzadegan and Entezari (2019) has shown that the trend values obtained from GRACE after removing soil moisture from GLDAS correlated well with the observed groundwater level variations in 4 sub-watersheds in Iran.

The aforementioned studies have assessed water loss in Iran over a short time period and did not include the GRACE-FO observations. Moreover, satellite gravimetric data have not been investigated together with groundwater level observations gauged via a country-wide dense

network of piezometric wells. We provides the first estimate of total water storage loss together with the uncertainty in Iran over the last two decades. The analysis incorporates the measurements of TWS from GRACE and GRACE-FO, precipitation from a dense network of rain gauges together with globally gridded datasets, and also groundwater level from piezometric wells. The results of our analysis are reported for the entire country and, for the first time, over its 30 major river basins. Our analysis shows significant total water storage losses over the whole country. Recently, the massive flood events in early 2019 took place in a large area of Iran (Yadollahie, 2019), which brought a considerable amount of water to the system. This study, for the first time, quantifies the contribution of these events to the TWS retrieval of Iran. Furthermore, using a dense network of piezometric wells, we quantify groundwater depletion as one of the leading representative indicators of anthropogenic activities and water bankruptcy.

Results:

Figure 4.35 (a) presents the time series of GRACE TWSA over Iran from 2003 to 2019. The time evolution of the total water storage loss relative to the initial epoch is shown in Figure 4.35 (b). The total water storage exhibited a weak positive trend from 2003 to 2005, which brought $23 \, km^3$ water to the system. The short-lived positive trend in TWS was followed by a continuous negative trend from 2006 to 2016, triggered by the drought in 2007, during which Iran has lost $256 \, km^3$ water, corresponding to a rate of $-23.25 \, km^3/yr$. The negative trend ended in early 2017 and TWSA showed a weak positive trend from 2017 to the end of 2019. Within 2017–2019, TWS gained $31 \, km^3$ at a rate of $+10.3 \, km^3/yr$. Overall since 2003 Iran has lost $211 \pm 34 \, km^3$ of its total water storage at an average rate of $-12 \pm 2 \, km^3/yr$. This water amount corresponds to about half of Lake Erie's volume (EPA), one of the Great Lakes in North America.

Our findings are consistent with those reported by previous studies focusing on different time periods (e.g., Ashraf et al., 2021; Forootan et al., 2014; Joodaki et al., 2014; Panahi et al., 2020; Voss et al., 2013). The total water consumption in Iran is estimated at $96 \, km^3$ (Khoosefi, 2018). Hence, our analysis shows that Iran has lost more than twice of its annual water consumption over the last two decades. Most of the water has been lost during 2008–2016, triggered by one of the two most severe droughts of the last 50 years in the Middle East.

Moreover, Least Squares Spectral Analysis (LSSA) method (Wells et al., 1985) revealed that the annual amplitude amounted to 7.4 cm over the study period (2003–2019), which represents $122 \, km^3$ of water. The annual TWS variation has not remained constant within the last two decades and has followed three main phases, i.e., from 2003 to 2007 with 8.5 cm, from 2008 to 2015 with 7.4 cm, and finally from 2016 to 2019 with 9 cm. The smaller annual variation throughout 2008–2015 is mainly driven by a continuous decrease in the precipitation (cf. Figure 4.36 and Figure 4.35). Given a normal situation without drought or flood periods, the annual TWS variation seems to vary within 8.5–9 cm (140–$148 \, km^3$).

The basin-wise distribution of the total water storage loss rates and relative precipitation gain or deficit in Iran over the last 17 years (2003–2019) are shown in Figure 4.37 (a) and (b), respectively. Figure 4.37 (a) presents the result of both rates of total water storage loss in terms of water height (colors) and volumetric (circular disks). All major basins suffer from a significant water loss within 2003–2019, varying between -0.07 and $-1.8 \, km^3/yr$. The maximum total water storage loss rate has occurred in the Central Desert (47) at a rate of $-1.8 \pm 0.26 \, km^3/yr$, followed by Salt Lake (41), Lut Desert (46), and Sefidrood (13), each at a rate of more than $-1 \, km^3/yr$. Ghareghoom (60) in the southeast of Iran has experienced the minimum water loss ($-0.07 \pm 0.05 \, km^3/yr$). Due to the inevitable post-processing process described in section 2.4 and the coarse spatial resolution of GRACE (e.g., Longuevergne et al., 2010; Lorenz et al., 2014; Rowlands et al., 2005; Vishwakarma et al., 2018), the final results carry uncertainties and errors.
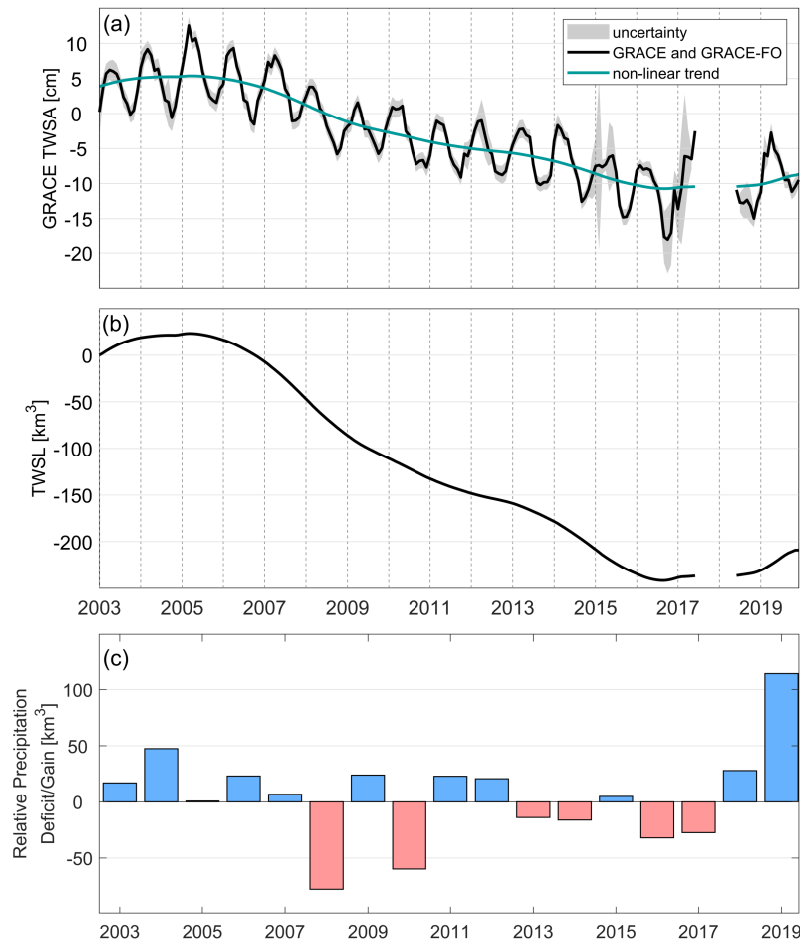
Figure 4.35: (a) TWSA derived from GRACE and GRACE-FO over Iran; The error envelope represents the uncertainty of the processed data. The non-linear trend is achieved using SSA algorithm with 24 month window. (b) Time evolution of the accumulated total water storage loss since 2003. (c) Inter-annual variation of the relative losing water status (red) or gaining water status (blue) from precipitation over Iran with respect to the long-term mean (1983–2002).

Therefore, GRACE observations should not be over-interpreted, especially in small catchments like Talesh (12) or Haraz-Sefidrood (14). It is noteworthy that due to the coarse spatial resolution of GRACE, the results over smaller basins come with inherent uncertainty. Moreover, the results in the Caspian Sea coastal region are prone to leakage error due to the strong negative trend in the signal of the Caspian Sea.

Lake Urmia in northwestern Iran is one of the largest lakes in the Middle East. This hypersaline lake has shrunk drastically over the last two decades to less than 30 % of its original surface area and it lost more than 30 % of its water volume from 1995 to 2015. Based on the results presented in Figure 4.40 and Figure 4.41, the basin has lost water at a rate of $-0.7 \pm 0.05 \, \mathrm{km^3/yr}$, one of the top 6 river basins in losing water storage in Iran over the study period (Figure 4.37 (a)). Although the overall deficit from precipitation over the last two decades is mild, the basin has suffered from a long period of persistent water loss from 2007 onwards (Figure 4.38). The expansion of agricultural land (Khazaei et al., 2019), together with water loss from precipitation, put pressure on this basin's water resources and encouraged increased groundwater use.

It is noteworthy that for the same period, the water input reflects a different picture Figure 4.37 (b). Relative to climatology, the water resource system has gained water from precipitation at a total rate of $+10.28 \pm 0.03 \, \mathrm{km^3/yr}$ in 17 basins (54 % of the area), while in the 13 other
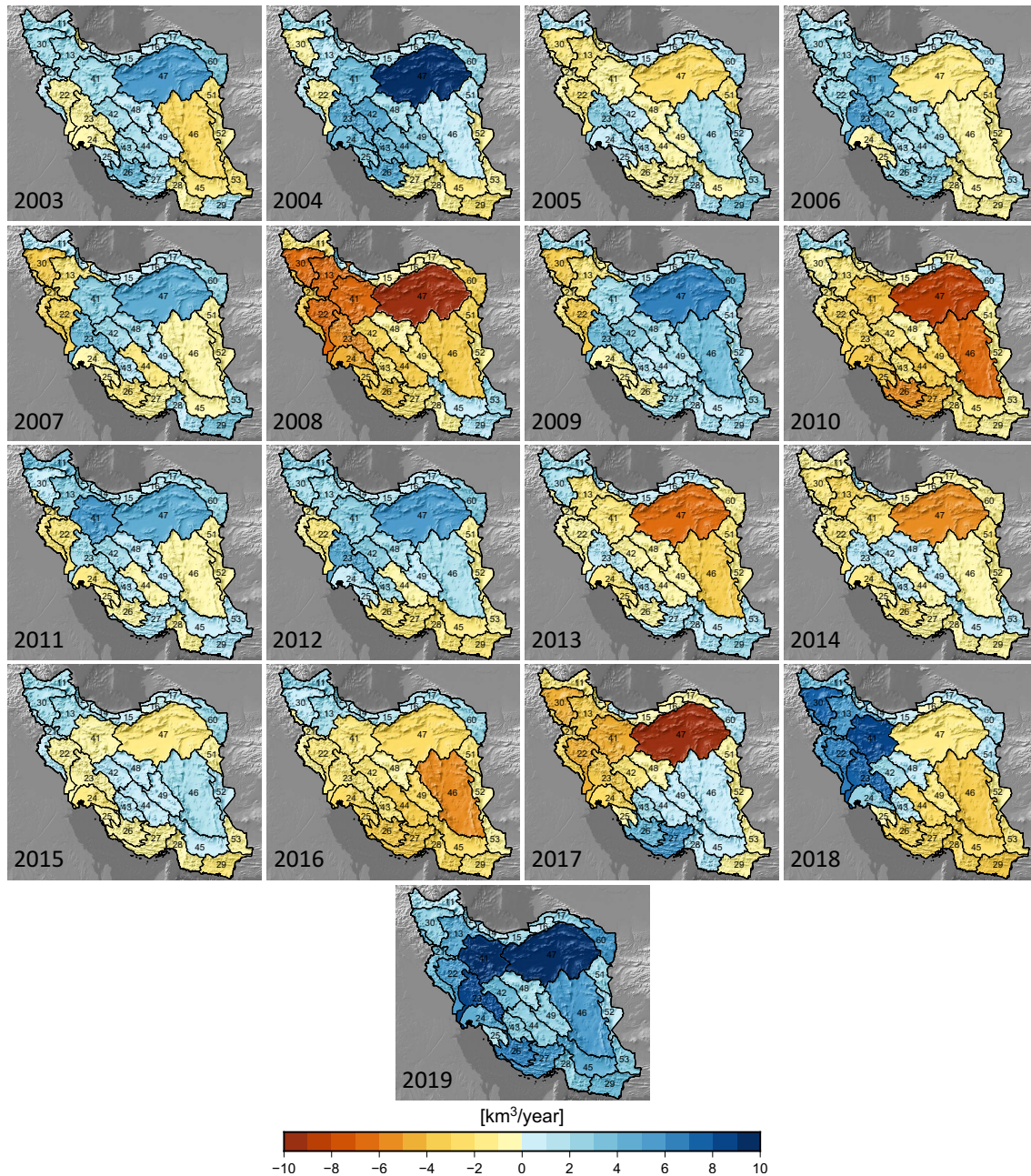
Figure 4.36: Basin-wise gain or deficit from precipitation at each year from 2003 to 2019 with respect to the long-term mean (1983–2002).

basins (46 % of the area) the total rate is $-5.42 \pm 0.03\,\mathrm{km}^3/\mathrm{yr}$. Overall, the whole country has gained water from precipitation at the rate of $+4.86 \pm 0.02\,\mathrm{km}^3/\mathrm{yr}$. The Karun basin (23) has experienced the maximum amount of gain at a rate of $+1.87 \pm 0.25\,\mathrm{km}^3/\mathrm{yr}$ while the adjacent Karkheh basin (22) shows the maximum deficit at a rate of $-1 \pm 0.11\,\mathrm{km}^3/\mathrm{yr}$. These results are mirrored by in-situ observations shown in (Figure 4.37 (d)). We calculated the gain or loss at 380 gauge stations, gauged from 1983 to the end of 2019 with less than one-year data outages within 2003–2019. Gauge results match very well the gridded data sets (Figure 4.37 (b)). The poor density of stations in the middle and eastern parts of the country is due to two vast deserts, namely the Lut Desert (46) and Central Desert (47) in the middle of the Iranian plateau.

Besides the TWS gain or loss, it is crucial to scrutinize how the water status evolved within 2003–
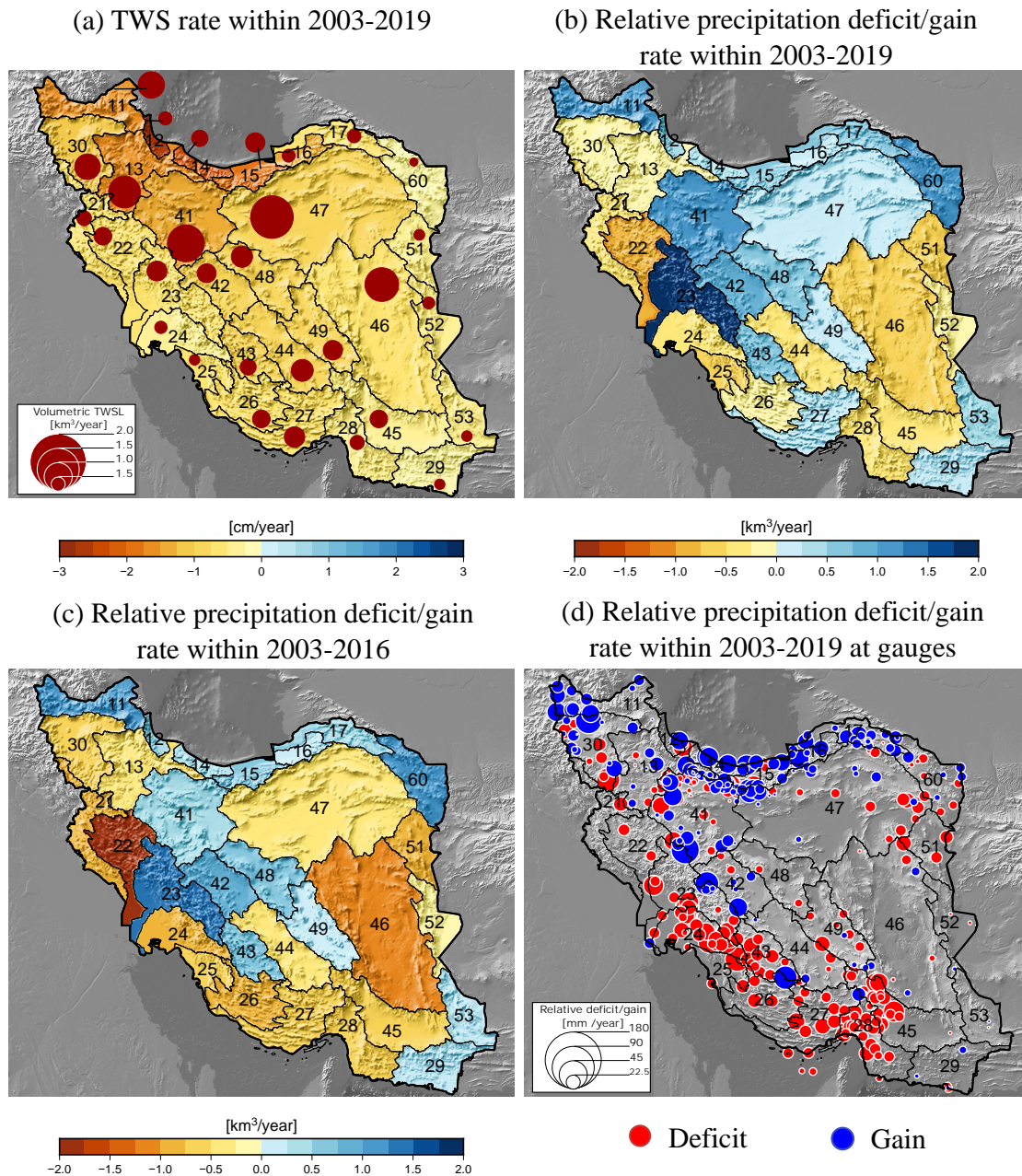
Figure 4.37: (a) Basin-wise distribution of TWS rate per unit area from 2003 to 2019. The absolute TWS rates are shown in km^3 per year using circular disks. (b) Relative deficit or gain in precipitation calculated from the selected precipitation datasets for the period 2003–2019. (c) same as (b) but for the period 2003–2016. (d) Relative deficit or gain in precipitation calculated from in-situ gauges for the period 2003–2019. The area of the disks at each station represents the magnitude of gain or deficit in mm/year.

2019. The drought event around 2007–2008 and the heavy rainfall in 2019 are two significant events that had a notable influence on the evolution (Figure 4.38). Eleven basins with an area of 40 % of the country have remained in deficit after the 2007 drought including Karkheh (22), Lake Urmia (30), and Helle (25). In general, 25 % of the whole country (6 basins) including Lut Desert (46), Karkheh (22), and Hamun Hirmand (52) have never experienced a gain in the relative precipitation during 2003–2019.

Almost all basins underwent heavy rainfall in early 2019 (Figure 4.36). Considering the long-term mean for the period 1983–2002 as the reference, the whole country gained around 115 ±
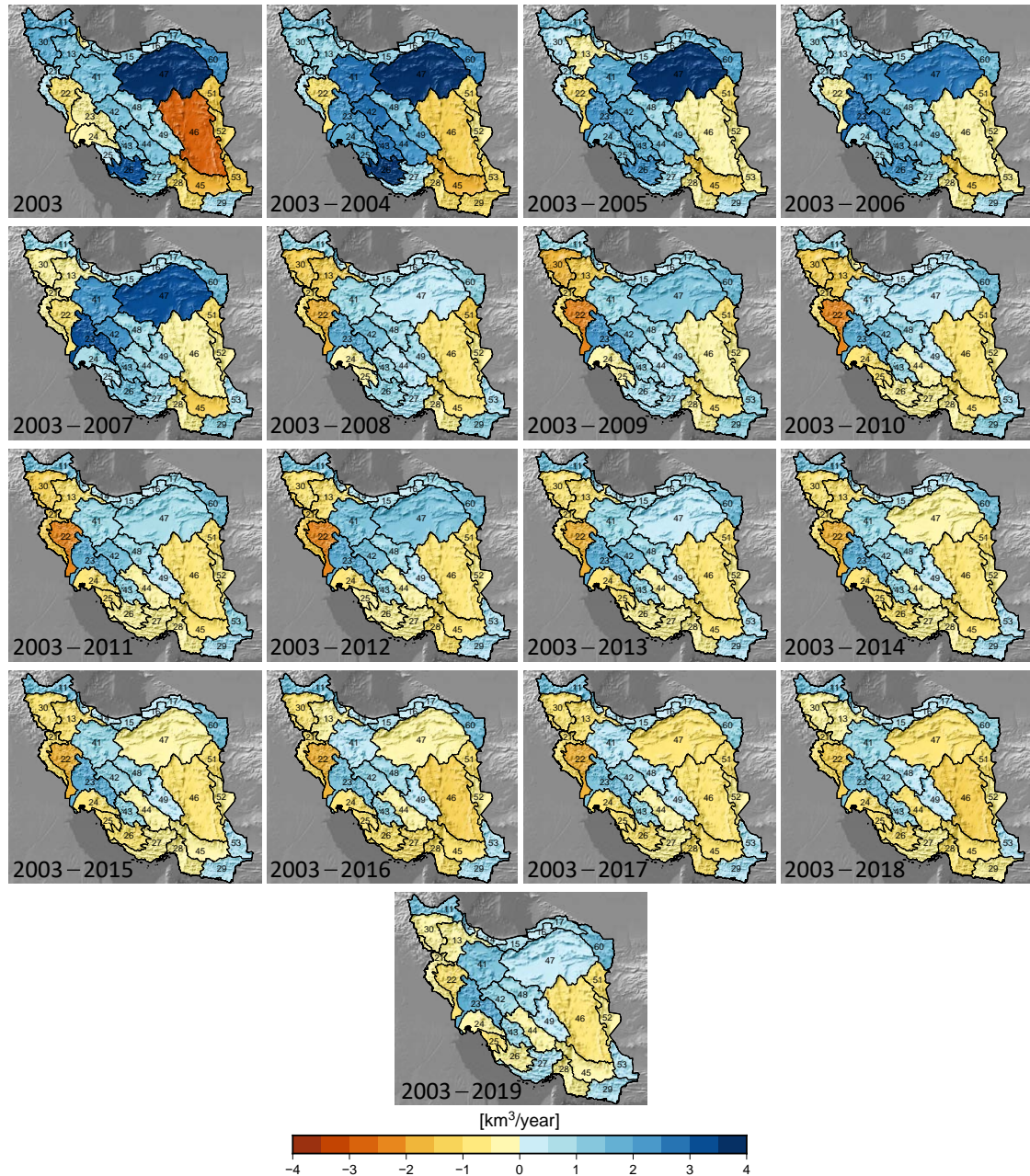
Figure 4.38: Basin-wise time evolution of gain or deficit in precipitation. At each epoch, the time period shown in the left corner is used to calculated gain or deficit with respect to the long-term mean (1983–2002).

$0.8\,km^3$ water in 2019. The Central Desert (47) received the maximum gain with $1.04 \pm 0.05\,km^3$ followed by the Salt Lake (41), the Mand (26), the Karun (23), and the Lut basin (46), all gaining more than $0.4\,km^3$ water. Comparing Figure 4.37 (b) and (c), such a heavy rainfall led to a change in the gain-deficit pattern in three basins (20 % of the total area) namely Mehran-Kal (27), South Baluchestan (29), and the Central Desert (47). Moreover, significant changes in gaining or losing water status are observed in Karkheh (22), Karun (23), and Lut Desert (46) (cf. the last two sub-figures in Figure 4.38).

To properly interpret the water loss from satellite gravimetry, we need to look at the mean annual precipitation gain or deficit from 2003 to 2019 relative to their mean annual precipitation, determined within the reference period 1983–2002. Figure 4.39 (a) shows gaining or losing wa-

ter status in percentage representing alarming rainfall deficit in basins Hamun Hirmand (52), Khaf (51), and Helle (25) by more than −10 %. In the northern and central major basins, we observe a general gain by more than 10 % of their mean annual precipitation. The percentage values of the total water loss with respect to the corresponding amplitude of the TWSA signal, trend-to-variability ratio, gives a sense how relevant the trend is with respect to the natural variability (Lehmann et al., 2015; Rahmstorf & Coumou, 2011). Considering the Root Mean Squared (RMS) of the TWSA as the amplitude, the trend-to-variability varies between −8 and −14 % per year (Figure 4.39 (b)).
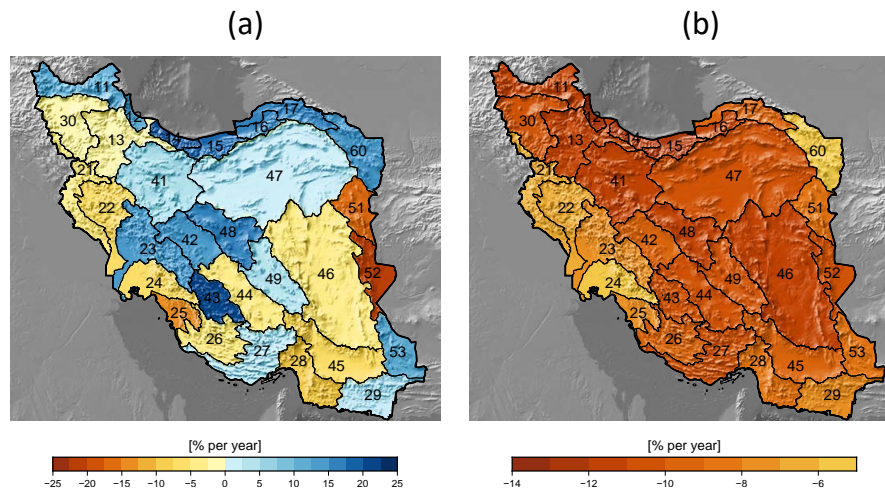

Figure 4.39: (a) The percentage of the relative precipitation gain or deficit over 2003–2019 with respect to 1983–2002. (b) The trend-to-variability ratio of mean annual TWS loss with respect to the RMS of the TWSA signal.

All the obtained results are associated with uncertainties, which we have tried to address in this study. Figure 4.40 (a) illustrates the rate of water loss together with their uncertainty. Despite their size the Lut Desert (46) and Central Desert (47) show high uncertainty, due to the weak signal which does not go beyond the GRACE noise level. The basin-wise gain or loss from precipitation and their corresponding uncertainty are shown in Figure 4.40 (b). The gain or loss of five basins, namely Sefidrood (13), Mehran-Kal (27), South Baluchestan (29), Lake Urmia (30), and Central Desert (47), is undecided as their error bar contains the zero level.

It is remarkable that basins like Aras (11), Karun (23), Salt Lake (41), and Ghareghoom (60) have gained more than 1 km^3 water while GRACE senses a negative trend. It should be noted that the GRACE estimates for small basins, like Aras (11), Talesh (12), Haraz-Sefidrood (14), Haraz (15), Gharesoo (16), and Atrak River (17) are prone to leakage. Such added uncertainty, however, will marginally affect the estimates of the trend. Since the water stored as surface water, soil moisture, canopy water and snow equivalent water is negligible in an arid to semi-arid climate of Iran (Abou Zaki et al., 2019; Van Camp et al., 2010), the negative trend in GRACE and simultaneous precipitation gain can only be explained by increased groundwater extraction. Agriculture consumes more than 90 % of Iran's water and heavily relies on groundwater for irrigation (Madani, 2014). The number of wells in the time period 1971-2013 has dramatically increased from just over 47 000 to nearly 789 000 (Noor, 2017). Moreover, reliance on groundwater has increased steadily during droughts.

To quantify the drop rate of the mean groundwater level, we analyzed groundwater level data from the piezometric stations (Figure 4.5 (b)). Figure 4.41 represents the relative annual loss or gain rate of the groundwater level of the major basins in Iran, including their uncertainties. Given the unknown storage coefficients of aquifers, results are reported in cm as an ab-
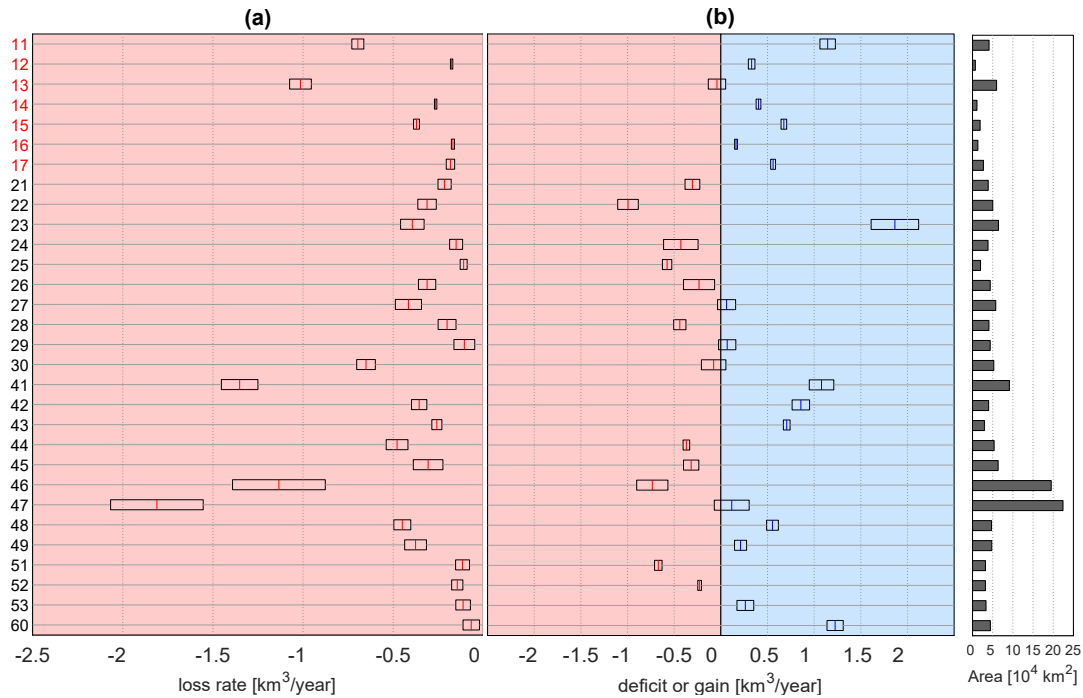
Figure 4.40: (a) Loss rate of TWSA at each basin together with its corresponding uncertainty shown by rectangle. (b) Annual deficit or gain rate from precipitation at each basin together with its corresponding uncertainty shown with rectangle. Basins with red ID are prone to GRACE leakage due to the strong negative trend of the Caspian Sea. The results of these basins (TWSA results) should not be over-interpreted.

solute volume quantification is not possible. The heterogeneous behavior of different aquifers in each basin resulted in large uncertainty values compared to those from GRACE or precipitation. Groundwater level has been dropped in most main river basins of Iran, except for the basins near the Caspian Sea, namely Talesh (12), Haraz-Sefiidrood (14), Haraz (15), Gharesoo (16), and South Baluchestan (29) (see Figure 4.42 (b)). The maximum loss has occurred in the Mehran-Kal (27) at the rate of -47.7 ± 6 cm/yr, followed by Saghand (49), Abarghoo-Sirjan (44), Ghareghoom (60), and Salt Lake (41). Based on the piezometric well observations, more than 90 % of the aquifers show negative trends within the study period while in 35 % of them, the groundwater level has dropped more than 40 cm/yr. Figure 4.42 (a) depict the TWS loss rate using the GRACE observation within 2003–2016 which is the same period of GWLA data (Figure 4.42 (b)). The annual rate of the mean groundwater level drop in this study is consistent with a recent estimation by Noori et al. (2021).

Figure 4.43 (a) represents the mean GroundWater Level Anomaly (GWLA) time series for entire Iran, together with the TWSA from GRACE, over the period from 2003 to the end of 2016. We observe that the GWL is highly correlated with TWSA ($r = 0.97$). The mean GWLA in Iran has dropped dramatically during the last two decades, triggered by the drought in 2007. Considering the period from 2003 to 2016, the mean groundwater shows a significant negative trend of -28 ± 1.4 cm/yr while this rate was -8.1 ± 3.3 cm/yr and -25.3 ± 1.9 cm/yr before and after 2008, respectively. The high correlation and the same trend behavior between TWSA and GWLA highlight the notable contribution of groundwater depletion in Iran's total water storage loss observed by GRACE.

The green and red background in Figure 4.43 (a) refer to the two distinct patterns observed in the scatter plot of the TWSA versus groundwater level Figure 4.43 (b). Except for the humid regions (15 % of the country) with considerable contribution from soil moisture anomaly (Rah-
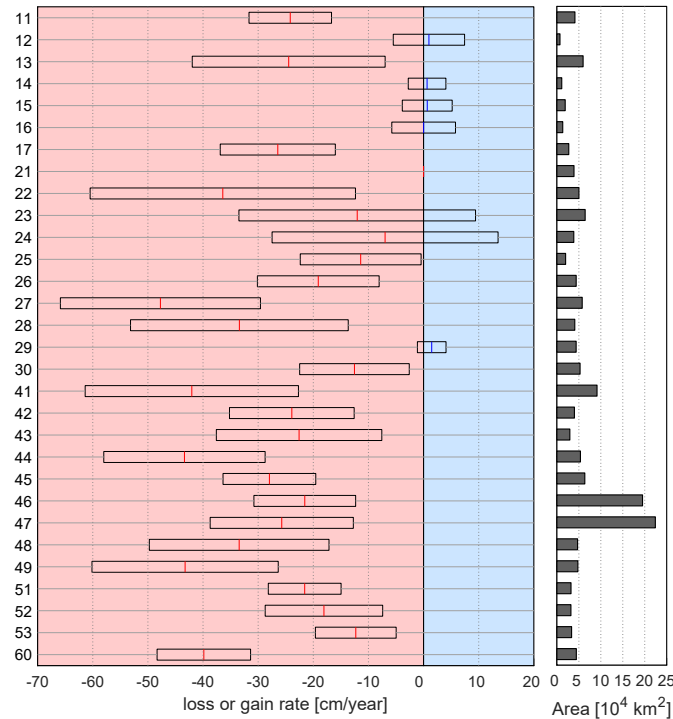
Figure 4.41: Loss or gain rate of groundwater level at each basin together with its corresponding uncertainty shown with the rectangles.
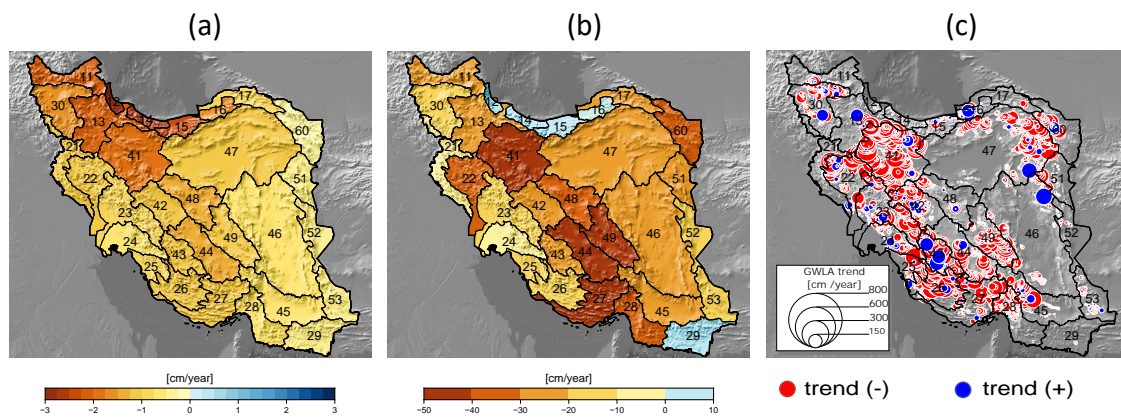

Figure 4.42: (a) Basin-wise distribution of TWS loss rate using GRACE observation within 2003–2016. (b) Basin-wise distribution groundwater level rate from the piezometric well observations within 2003–2016.

mani et al., 2016) and the variation in Lake Urmia in the northwest (Ashraf et al., 2019; Tourian et al., 2015), the GWS anomaly is the dominant compartment of the TWSA in most of Iran. Therefore, the scatter points' slope implicitly reveals the storativity or the storage coefficient since it maps the water table to the volume quantity. For an unconfined aquifer, which most aquifers in Iran are, the storage coefficient is approximately equal to the specific yield. Two distinct slopes, the green line from 2003 to 2007 0.072 and the red one from 2008 to 2016, 0.04, implicitly indicates that the acceleration in groundwater loss in last years brought the groundwater to a deeper level with a different soil structure (Figure 4.43 (b)). Another likely reason can be that groundwater extraction in many regions of Iran has become more challenging and expensive as the groundwater drops. Also, below a certain level, there may not be even much water left for extraction (at least for a period of time in each year).

Using the piezometric wells observations, two recent studies, namely Ashraf et al. (2021) and Noori et al. (2021), estimated Iran's groundwater storage depletion within 2002–2015 as 74 km^3 and 75 km^3, respectively. Based on our analysis, Iran has lost 241 km^3 of its total water storage within 2003–2015. Considering total water storage as the sum of groundwater, surface water, soil moisture, and snow water, we can derive an estimate of 166 km^3 for the water loss from the surface, soil, and snow in Iran within 2003 to 2015. The shrinkage of the Lake Urmia (11 km^3) and the reported d(r)ying up lakes like Lake Hamoon and Lake Bakhtegan, narrowing of permanent rivers width, and the disappearance of seasonal rivers are only a few shreds of evidence in agreement with this estimate.
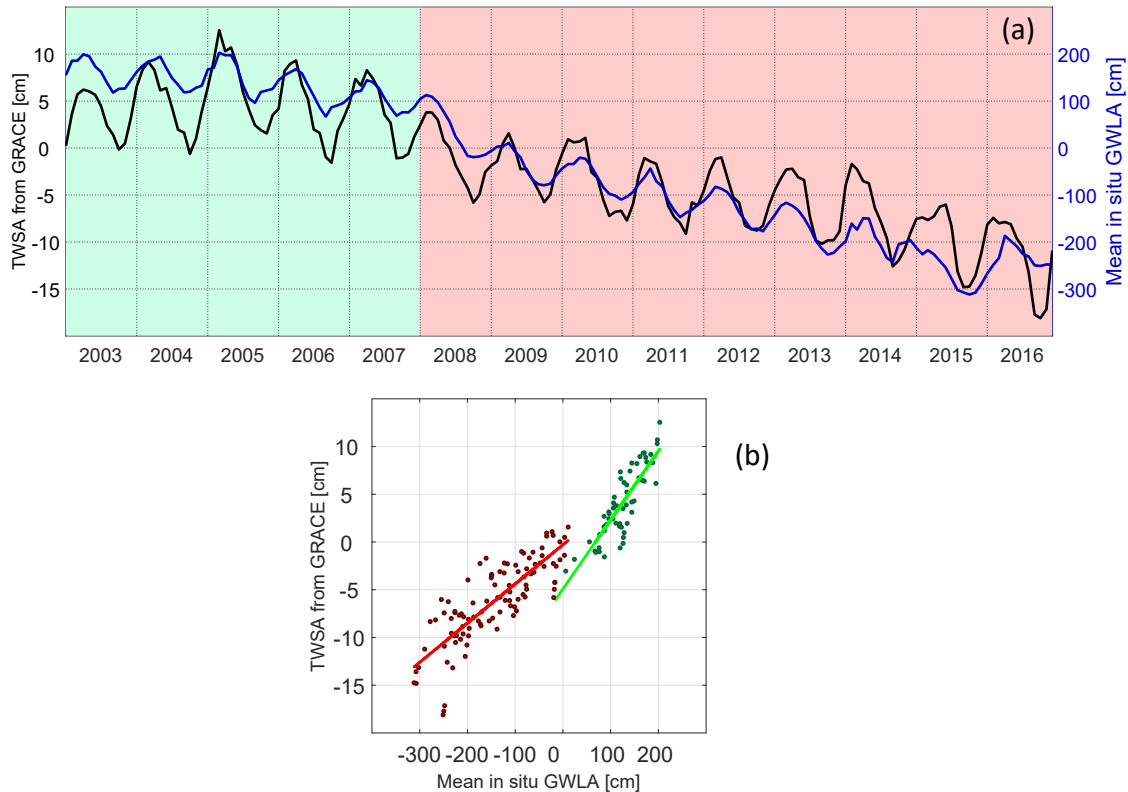

Figure 4.43: (a) TWSA derived from GRACE over Iran together with the mean groundwater level from piezometric wells network. (b) Scatter plot of TWSA with respect to mean groundwater level including two distinct slopes, color-coded in green (2003–2007) and in red (2008–2016).

We highlight that the piezometric wells' observations are limited mainly to the groundwater level in aquifers. These aquifers are located in the plains, excluding groundwater flow in mountainous terrains. In case the measurements in mountainous regions were available, we would expect a higher rate of groundwater withdrawal due to the bedrocks' steeper slope. However, since the main groundwater level fluctuations occur in irrigated plains, the contribution of groundwater depletion in the mountainous regions in the estimation of total groundwater decline should be negligible. Considering the potential role of the surface-groundwater interaction, it should be noted that, the accelerated increase in the water withdrawals over the recent years would result in the decrease of the groundwater contribution to the baseflow. This decline has already been observed in Lake Urmia basin by Vaheddoost and Aksoy, 2018 and needs to be further investigated over other basins, especially mountainous regions.

In an arid to a semi-arid region like Iran, groundwater is a precious resource functioning as the backbone of irrigated farming. The alarming accelerated rate of groundwater withdrawal after the 2007 drought continued till the end of 2016. Maghrebi et al. (2020) assessed Iran's agri-

cultural activities within 1981–2013 using agricultural area/production data from Iran's Ministry of Agriculture Jihad, Ministry of Energy, and Ministry of Roads and Urban Development. Our results from piezometric wells confirm the finding from Maghrebi et al. (2020) that the groundwater over-exploitation for agriculture and consequently irrigated agricultural production increased despite declining water availability during 2003–2016. Since the rate of recharge is slower than the pumping, many aquifers would be in danger of being depleted, and their water content never be recovered like aquifers in the agriculturally active regions of the world such High Plain (Scanlon et al., 2012) and Central Valley aquifers (Famiglietti et al., 2011) in the United States or the North China Plain (Feng et al., 2013). The observed decrease of groundwater extraction in many parts of Iran is attributable to the inefficiency of pumping (due to high groundwater depth or low water quality) and not necessarily the result of groundwater conservation and monitoring efforts across the country (Ashraf et al., 2017; Ashraf et al., 2021; Madani, 2014; Noori et al., 2021; Tourian et al., 2015).

Bibliography

Abou Zaki, N., Torabi Haghighi, A., M. Rossi, P., Tourian, M. J., & Kløve, B. (2019). Monitoring groundwater storage depletion using gravity recovery and climate experiment (GRACE) data in Bakhtegan Catchment, Iran. *Water*, *11*(7), 1456. https://doi.org/10.3390/w11071456

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2

Alborzi, A., Mirchi, A., Moftakhari, H., Mallakpour, I., Alian, S., Nazemi, A., Hassanzadeh, E., Mazdiyasni, O., Ashraf, S., Madani, K., et al. (2018). Climate-informed environmental inflows to revive a drying lake facing meteorological and anthropogenic droughts. *Environmental Research Letters*, *13*(8), 084010. https://doi.org/10.1088/1748-9326/aad246

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of climatology*, *6*(6), 661–675. https://doi.org/10.1002/joc.3370060607

Alsdorf, D., Rodriguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics*, *45*(RG2002). https://doi.org/10.1029/2006RG000197

Amiraslani, F., & Dragovich, D. (2011). Combating desertification in Iran over the last 50 years: An overview of changing approaches. *Journal of Environmental Management*, *92*(1), 1–13. https://doi.org/10.1016/j.jenvman.2010.08.012

Araghi, A., & Mousavi-Baygi, M. (2020). Variability in snowfall/total precipitation-day ratio in Iran. *Theoretical and Applied Climatology*, *140*(1), 547–558. https://doi.org/10.1007/s00704-020-03101-x

Arsanjani, T. J., Javidan, R., Nazemosadat, M. J., Arsanjani, J. J., & Vaz, E. (2015). Spatiotemporal monitoring of Bakhtegan Lake's areal fluctuations and an exploration of its future status by applying a cellular automata model. *Computers & Geosciences*, *78*, 37–43. https://doi.org/10.1016/j.cageo.2015.02.004

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., Nelson, B. R., & Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society*, *96*(1), 69–83. https://doi.org/10.1175/bams-d-13-00068.1

Ashraf, S., AghaKouchak, A., Alizadeh, A., Baygi, M. M., Moftakhari, H. R., Mirchi, A., Anjileli, H., & Madani, K. (2017). Quantifying anthropogenic stress on groundwater resources. *Scientific reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-12877-4

Ashraf, S., AghaKouchak, A., Nazemi, A., Mirchi, A., Sadegh, M., Moftakhari, H. R., Hassanzadeh, E., Miao, C.-Y., Madani, K., Baygi, M. M., et al. (2019). Compounding effects of human activities and climatic changes on surface water availability in Iran. *Climatic change*, *152*(3-4), 379–391. https://doi.org/10.1007/s10584-018-2336-6

Ashraf, S., Nazemi, A., & AghaKouchak, A. (2021). Anthropogenic Drought Dominates Groundwater Depletion in Iran. *Scientific Reports*, *11*. https://doi.org/10.1038/s41598-021-88522-y

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I., McVicar, T. R., & Adler, R. F. (2019). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/bams-d-17-0138.1

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific data*, *5*, 180214. https://doi.org/10.1038/s41597-020-00616-w

Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., & Ziese, M. (2013). A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth System Science Data*, *5*(1), 71–99. https://doi.org/10.5194/essd-5-71-2013

Blewitt, G., & Lavallée, D. (2002). Effect of annual signals on geodetic velocity. *Journal of Geophysical Research: Solid Earth*, *107*(B7), ETG–9. https://doi.org/10.1029/2001jb000570

Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, *58*(1-2), 11–27. https://doi.org/10.1016/0022-1694(82)90066-X

Castellazzi, P., Martel, R., Galloway, D. L., Longuevergne, L., & Rivera, A. (2016). Assessing groundwater depletion and dynamics using GRACE and InSAR: Potential and limitations. *Groundwater*, *54*(6), 768–780. https://doi.org/10.1111/gwat.12453

Chao, L., Zhang, K., Wang, J., Feng, J., & Zhang, M. (2021). A comprehensive evaluation of five evapotranspiration datasets based on ground and grace satellite observations: Implications for improvement of evapotranspiration retrieval algorithm. *Remote Sensing*, *13*(12), 2414. https://doi.org/10.3390/rs13122414

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres*, *113*(D4). https://doi.org/10.1029/2007jd009132

Chen, M., Xie, P., Janowiak, J. E., & Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, *3*(3), 249–266. https://doi.org/10.1175/1525-7541(2002)003<0249:glpaym>2.0.co;2

Coss, S., Durand, M., Yi, Y., Jia, Y., Guo, Q., Tuozzolo, S., Shum, C., Allen, G. H., Calmant, S., & Pavelsky, T. (2020). Global River Radar Altimetry Time Series (GRRATS): new river elevation earth science data records for the hydrologic community. *Earth System Science Data*, *12*(1), 137–150. https://doi.org/10.5194/essd-12-137-2020

Crétaux, J.-F., Jelinski, W., Calmant, S., Kouraev, A., Vuglinski, V., Bergé-Nguyen, M., Gennero, M.-C., Nino, F., Rio, R. A. D., Cazenave, A., & Maisongrande, P. (2011). SOLS: A lake database to monitor in the Near Real Time water level and storage variations from remote sensing data. *Advances in Space Research*, *47*(9), 1497–1507. https://doi.org/10.1016/j.asr.2011.01.004

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, *137*(656), 553–597. https://doi.org/10.1002/qj.722

Döll, P., Mueller Schmied, H., Schuh, C., Portmann, F. T., & Eicker, A. (2014). Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, *50*(7), 5698–5720. https://doi.org/10.1002/2014wr015595

Donchyts, G., Baart, F., Winsemius, H., Gorelick, N., Kwadijk, J., & Van De Giesen, N. (2016). Earth's surface water change over the past 30 years. *Nature Climate Change*, *6*(9), 810–813. https://doi.org/10.1038/nclimate3111

Dubreuil, A., Assoumou, E., Bouckaert, S., Selosse, S., Maïzi, N., et al. (2013). Water modeling in an energy optimization framework–The water-scarce Middle East context. *Applied energy, 101*, 268–279. https://doi.org/10.1016/j.apenergy.2012.06.032

Elnashar, A., Wang, L., Wu, B., Zhu, W., & Zeng, H. (2021). Synthesis of global actual evapotranspiration from 1982 to 2019. *Earth System Science Data, 13*(2), 447–480. https://doi.org/10.5194/essd-13-447-2021

Emadodin, I., Narita, D., & Bork, H. R. (2012). Soil degradation and agricultural sustainability: an overview from Iran. *Environment, Development and Sustainability, 14*(5), 611–625. https://doi.org/10.1007/s10668-012-9351-y

Famiglietti, J. S., Lo, M., Ho, S. L., Bethune, J., Anderson, K., Syed, T. H., Swenson, S. C., de Linage, C. R., & Rodell, M. (2011). Satellites measure recent rates of groundwater depletion in California's Central Valley. *Geophysical Research Letters, 38*(3). https://doi.org/10.1029/2010gl046442

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

Feng, W., Zhong, M., Lemoine, J.-M., Biancale, R., Hsu, H.-T., & Xia, J. (2013). Evaluation of groundwater depletion in North China using the Gravity Recovery and Climate Experiment (GRACE) data and ground-based measurements. *Water Resources Research, 49*(4), 2110–2118. https://doi.org/10.1002/wrcr.20192

Forootan, E., Rietbroek, R., Kusche, J., Sharifi, M., Awange, J., Schmidt, M., Omondi, P., & Famiglietti, J. (2014). Separation of large scale water storage patterns over Iran using GRACE, altimetry and hydrological data. *Remote Sensing of Environment, 140*, 580–595. https://doi.org/10.1016/j.rse.2013.09.025

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data, 2*, 150066. https://doi.org/10.1038/sdata.2015.66

Ghiggi, G., Humphrey, V., Seneviratne, S., & Gudmundsson, L. (2021). G-RUN ENSEMBLE: A Multi-Forcing Observation-Based Global Runoff Reanalysis. *Water Resources Research, 57*(5), e2020WR028787. https://doi.org/10.1029/2020WR028787

Gleeson, T., Wada, Y., Bierkens, M. F., & Van Beek, L. P. (2012). Water balance of global aquifers revealed by groundwater footprint. *Nature, 488*(7410), 197–200. https://doi.org/10.1038/nature11295

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data, 7*(1), 1–18. https://doi.org/10.1038/s41597-020-0453-3

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Huang, J., Zhang, Y., Arhonditsis, G. B., Gao, J., Chen, Q., Wu, N., Dong, F., & Shi, W. (2019). How successful are the restoration efforts of China's lakes and reservoirs? *Environment international, 123*, 96–103. https://doi.org/10.1016/j.envint.2018.11.048

Huang, Y., Salama, M. S., Krol, M. S., Su, Z., Hoekstra, A. Y., Zeng, Y., & Zhou, Y. (2015). Estimation of human-induced changes in terrestrial water storage through integration of GRACE satellite detection and hydrological modeling: A case study of the Y angtze R iver basin. *Water Resources Research, 51*(10), 8494–8516. https://doi.org/10.1002/2015WR016923

Huffman, G. J., Adler, R. F., Morrissey, M. M., Bolvin, D. T., Curtis, S., Joyce, R., McGavock, B., & Susskind, J. (2001). Global precipitation at one-degree daily resolution from multisatellite observations. *Journal of hydrometeorology*, *2*(1), 36–50. https://doi.org/10.1175/1525-7541(2001)002<0036:gpaodd>2.0.co;2

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/jhm560.1

Huffman, G., Stocker, E., Bolvin, D., Nelkin, E., & Tan, J. (2019). GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06. https://doi.org/10.5067/GPM/IMERGDF/DAY/06

Jones, M. D., Djamali, M., Holmes, J., Weeks, L., Leng, M. J., Lashkari, A., Alamdari, K., Noorollahi, D., Thomas, L., & Metcalfe, S. E. (2015). Human impact on the hydroenvironment of Lake Parishan, SW Iran, through the late-Holocene. *The Holocene*, *25*(10), 1651–1661. https://doi.org/10.1177/0959683615594242

Joodaki, G., Wahr, J., & Swenson, S. (2014). Estimating the human contribution to groundwater depletion in the Middle East, from GRACE data, land surface models, and well observations. *Water Resources Research*, *50*(3), 2679–2692. https://doi.org/10.1002/2013wr014633

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., & Reichstein, M. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific data*, *6*(1), 1–14. https://doi.org/10.1038/s41597-019-0076-8

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society*, *77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., & Potter, G. (2002). NCEP–DOE AMPI-II reanalysis (r-2). *Bulletin of the American Meteorological Society*, *83*(11), 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631

Khalyani, A. H., & Mayer, A. L. (2013). Spatial and temporal deforestation dynamics of Zagros forests (Iran) from 1972 to 2009. *Landscape and urban planning*, *117*, 1–12. https://doi.org/10.1016/j.landurbplan.2013.04.014

Khan, S. A., Wahr, J., Bevis, M., Velicogna, I., & Kendrick, E. (2010). Spread of ice mass loss into northwest Greenland observed by GRACE and GPS. *Geophysical Research Letters*, *37*(6). https://doi.org/10.1029/2010GL042460

Khandelwal, A., Karpatne, A., Marlier, M. E., Kim, J., Lettenmaier, D. P., & Kumar, V. (2017). An approach for global monitoring of surface water extent variations in reservoirs using MODIS data. *Remote sensing of Environment*, *202*, 113–128. https://doi.org/10.1016/j.rse.2017.05.039

Khazaei, B., Khatami, S., Alemohammad, S. H., Rashidi, L., Wu, C., Madani, K., Kalantari, Z., Destouni, G., & Aghakouchak, A. (2019). Climatic or regionally induced by humans? Tracing hydro-climatic and land-use changes to better understand the Lake Urmia tragedy. *Journal of hydrology*, *569*, 203–217. https://doi.org/10.1016/j.jhydrol.2018.12.004

Khoosefi, M. E. (2018). *Water crisis and feasibility study of connecting the northern and southern water bodies of the country, 7th Edition, Plan and Budget Organization of Iran,(In Persian)*. Plan and Budget Organization of Iran.

Khormali, F., Ajami, M., Ayoubi, S., Srinivasarao, C., & Wani, S. (2009). Role of deforestation and hillslope position on soil quality attributes of loess-derived soils in Golestan province, Iran. *Agriculture, ecosystems & environment*, *134*(3-4), 178–189. https://doi.org/10.1016/j.agee.2009.06.017

4

Klein, I., Gessner, U., Dietz, A. J., & Kuenzer, C. (2017). Global WaterPack–A 250 m resolution dataset revealing the daily dynamics of global inland water bodies. *Remote sensing of environment, 198*, 345–362. https://doi.org/10.1016/j.rse.2017.06.045

Lehane, S. (2014). The iranian water crisis. *Strategic Analysis Paper, Future Directions International International Pty Ltd.: Perth, Australia, 11.* https://apo.org.au/node/38233

Lehmann, J., Coumou, D., & Frieler, K. (2015). Increased record-breaking precipitation events under global warming. *Climatic Change, 132*(4), 501–515. https://doi.org/10.1007/s10584-015-1434-y

Li, P., Li, H., Chen, F., & Cai, X. (2020). Monitoring long-term lake level variations in middle and lower Yangtze basin over 2002–2017 through integration of multiple satellite altimetry datasets. *Remote Sensing, 12*(9), 1448. https://doi.org/10.3390/rs12091448

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology, 201*(1-4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Long, D., Longuevergne, L., & Scanlon, B. R. (2014). Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research, 50*(2), 1131–1151. https://doi.org/10.1002/2013wr014581

Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N., & Save, H. (2013). GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophysical Research Letters, 40*(13), 3395–3401. https://doi.org/10.1002/grl.50655

Longuevergne, L., Scanlon, B. R., & Wilson, C. R. (2010). GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research, 46*(11). https://doi.org/10.1029/2009wr008564

Lorenz, C., Kunstmann, H., Devaraju, B., Tourian, M. J., Sneeuw, N., & Riegger, J. (2014). Large-scale runoff from landmasses: a global assessment of the closure of the hydrological and atmospheric water balances. *Journal of Hydrometeorology, 15*(6), 2111–2139. https://doi.org/10.1175/jhm-d-13-0157.1

Lorenz, C., Tourian, M. J., Devaraju, B., Sneeuw, N., & Kunstmann, H. (2015). Basin-scale runoff prediction: An Ensemble Kalman Filter framework based on global hydrometeorological data sets. *Water Resources Research, 51*(10). https://doi.org/10.1002/2014WR016794

Madani, K. (2014). Water management in Iran: What is causing the looming crisis? *Journal of environmental studies and sciences, 4*(4), 315–328. https://doi.org/10.1007/s13412-014-0182-z

Madani, K., AghaKouchak, A., & Mirchi, A. (2016). Iran's socio-economic drought: challenges of a water-bankrupt nation. *Iranian studies, 49*(6), 997–1016. https://doi.org/10.1080/00210862.2016.1259286

Maghrebi, M., Noori, R., Bhattarai, R., Yaseen, Z. M., Tang, Q., Al-Ansari, N., Mehr, A. D., Karbassi, A., Omidvar, J., Farnoush, H., et al. (2020). Iran's Agriculture in the Anthropocene. *Earth's Future*, e2020EF001547. https://doi.org/10.1029/2020ef001547

Marc, V., Radfar, M., Martens, K., & Walraevens, K. (2012). Analysis of the groundwater resource decline in an intramountain aquifer system in Central Iran. *Geologica belgica*. https://popups.uliege.be/1374-8505/index.php?id=3644

Mardi, A. H., Khaghani, A., MacDonald, A. B., Nguyen, P., Karimi, N., Heidary, P., Karimi, N., Saemian, P., Sehatkashani, S., Tajrishy, M., et al. (2018). The Lake Urmia environmental disaster in Iran: A look at aerosol pollution. *Science of The Total Environment, 633*, 42–49. https://doi.org/10.1016/j.scitotenv.2018.03.148

Markert, K. N., Pulla, S. T., Lee, H., Markert, A. M., Anderson, E. R., Okeowo, M. A., & Limaye, A. S. (2019). AltEx: An open source web application and toolkit for accessing and ex-

4

ploring altimetry datasets. *Environmental Modelling & Software, 117,* 164–175. https: //doi.org/10.1016/j.envsoft.2019.03.021

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American statistical association, 44*(247), 335–341. https://doi.org/10.1080/01621459.1949.10483310

Michel, D. (2017). Iran's impending water crisis. *Water, Security and US Foreign Policy. Routledge, USA,* 168–188. https://doi.org/10.4324/9781315168272-10

Mirnezami, S. J., Bagheri, A., & Maleki, A. (2018). Inaction of society on the drawdown of groundwater resources: a case study of Rafsanjan plain in Iran. *Water Alternatives, 11*(3), 725–748.

Mirzaei, A., Saghafian, B., Mirchi, A., & Madani, K. (2019). The groundwater–energy–food nexus in Iran's agricultural sector: Implications for water security. *Water, 11*(9), 1835. https: //doi.org/10.3390/w11091835

Mooney, C. Z. (1997). *Monte Carlo simulation* (Vol. 116). Sage publications.

Motagh, M., Walter, T. R., Sharifi, M. A., Fielding, E., Schenk, A., Anderssohn, J., & Zschau, J. (2008). Land subsidence in Iran caused by widespread water reservoir overexploitation. *Geophysical Research Letters, 35*(16). https://doi.org/10.1029/2008gl033814

Mousavi, S.-F. (2005). Agricultural drought management in Iran. *Water conservation, reuse, and recycling: Proceedings of an Iranian-American workshop,* 106–113.

Nabavi, E. (2018). Failed policies, falling aquifers: Unpacking groundwater overabstraction in Iran. *Water Alternatives, 11*(3), 699.

Noor, H. (2017). Analysis of groundwater resources utilization and their current condition in Iran. *Iranian Journal of Rainwater Catchment Systems, 5*(2), 29–38. http://jircsa.ir/article-1-183-en.html

Noori, R., Maghrebi, M., Mirchi, A., Tang, Q., Bhattarai, R., Sadegh, M., Noury, M., Haghighi, A. T., Kløve, B., & Madani, K. (2021). Anthropogenic depletion of Iran's aquifers. *Proceedings of the National Academy of Sciences, 118*(25). https://doi.org/10.1073/pnas.2024221118

Panahi, D. M., Kalantari, Z., Ghajarnia, N., Seifollahi-Aghmiuni, S., & Destouni, G. (2020). Variability and change in the hydro-climate and water resources of Iran over a recent 30-year period. *Scientific Reports, 10*(1), 1–9. https://doi.org/10.1038/s41598-020-64089-y

Papa, F., Durand, F., Rossow, W. B., Rahman, A., & Bala, S. K. (2010). Satellite altimeter-derived monthly discharge of the Ganga-Brahmaputra river and its seasonal to interannual variations from 1993 to 2008. *Journal of Geophysical Research, 115,* C12013. https://doi.org/10.1029/2009JC006075

Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature, 540*(7633), 418–422. https://doi.org/10.1038/nature20584

Pettit, A. (1979). Anon-parametric approach to the change-point detection. *Appl. Stat, 28,* 126–135.

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., et al. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change, 11*(3), 226–233. https://doi.org/10.1038/s41558-020-00972-w

Rahimzadegan, M., & Entezari, S. A. (2019). Performance of the Gravity Recovery and Climate Experiment (GRACE) method in monitoring groundwater-level changes in local-scale study regions within Iran. *Hydrogeology Journal, 27*(7), 2497–2509. https://doi.org/10.1007/s10040-019-02007-x

Rahmani, A., Golian, S., & Brocca, L. (2016). Multiyear monitoring of soil moisture over Iran through satellite and reanalysis soil moisture products. *International journal of applied*

earth observation and geoinformation, 48, 85–95. https://doi.org/10.1016/j.jag.2015.06.009

Rahmstorf, S., & Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences, 108*(44), 17905–17909. https://doi.org/10.1073/pnas.1101766108

Reager, J., & Famiglietti, J. S. (2013). Characteristic mega-basin water storage behavior using GRACE. *Water resources research, 49*(6), 3314–3329. https://doi.org/10.1002/wrcr.20264

Riegger, J., & Tourian, M. J. (2014). Characterization of runoff-storage relationships by satellite gravimetry and remote sensing. *Water Resources Research, 50*(4), 3444–3466. https://doi.org/10.1002/2013wr013847

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., et al. (2011). MERRA: NASA's modern-era retrospective analysis for research and applications. *Journal of climate, 24*(14), 3624–3648. https://doi.org/10.1175/JCLI-D-11-00015.1

Rodell, M., Chen, J., Kato, H., Famiglietti, J. S., Nigro, J., & Wilson, C. R. (2007). Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE. *Hydrogeology Journal, 15*(1), 159–166. https://doi.org/10.1029/1999wr900141

Rodell, M., Famiglietti, J. S., Wiese, D. N., Reager, J., Beaudoing, H. K., Landerer, F. W., & Lo, M.-H. (2018). Emerging trends in global freshwater availability. *Nature, 557*(7707), 651–659. https://doi.org/10.1038/s41586-018-0123-1

Rowlands, D. D., Luthcke, S., Klosko, S., Lemoine, F. G., Chinn, D., McCarthy, J., Cox, C., & Anderson, O. (2005). Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements. *Geophysical Research Letters, 32*(4). https://doi.org/10.1029/2004gl021908

Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology, 200*, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016

Saemian, P., Elmi, O., Vishwakarma, B., Tourian, M., & Sneeuw, N. (2020). Analyzing the Lake Urmia restoration progress using ground-based and spaceborne observations. *Science of The Total Environment*, 139857. https://doi.org/10.1016/j.scitotenv.2020.139857

Saemian, P., Hosseini-Moghari, S.-M., Fatehi, I., Shoarinezhad, V., Modiri, E., Tourian, M. J., Tang, Q., Nowak, W., Bárdossy, A., & Sneeuw, N. (2021). Comprehensive evaluation of precipitation datasets over Iran. *Journal of Hydrology*, 127054. https://doi.org/10.1016/j.jhydrol.2021.127054

Scanlon, B. R., Faunt, C. C., Longuevergne, L., Reedy, R. C., Alley, W. M., McGuire, V. L., & McMahon, P. B. (2012). Groundwater depletion and sustainability of irrigation in the US High Plains and Central Valley. *Proceedings of the national academy of sciences, 109*(24), 9320–9325. https://doi.org/10.1073/pnas.1200311109

Schneider, U., Becker, A., Finger, P., Rustemeier, E., & Ziese, M. (2020). GPCC full data monthly product version 2020 at 0.25°: monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. *Global Precipitation Climatology Centre at Deutscher Wetterdienst: Offenbach, Germany*. https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025

Schwatke, C., Dettmering, D., Bosch, W., & Seitz, F. (2015). DAHITI – an innovative approach for estimating water level time series over inland waters using multi-mission satellite altimetry. *Hydrology and Earth System Sciences, 19*(10), 4345–4364. https://doi.org/10.5194/hess-19-4345-2015

Schwatke, C., Dettmering, D., & Seitz, F. (2020). Volume Variations of Small Inland Water Bodies from a Combination of Satellite Altimetry and Optical Imagery. *Remote Sensing, 12*(10), 1606. https://doi.org/10.3390/rs12101606

Schwatke, C., Scherer, D., & Dettmering, D. (2019). Automated Extraction of Consistent Time-Variable Water Surfaces of Lakes and Reservoirs Based on Landsat and Sentinel-2. *Remote Sensing, 11*(9), 1010. https://doi.org/10.3390/rs11091010

Searcy, J. K., & Hardison, C. H. (1960). *Double-mass curves*. US Government Printing Office.

Seyoum, W. M. (2018). Characterizing water storage trends and regional climate influence using GRACE observation and satellite altimetry data in the Upper Blue Nile River Basin. *Journal of Hydrology, 566*, 274–284. https://doi.org/10.1016/j.jhydrol.2018.09.025

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics, 56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Syed, T., Famiglietti, J. S., Rodell, M., Chen, J., & Wilson, C. R. (2008). Analysis of terrestrial water storage changes from GRACE and GLDAS. *Water Resources Research, 44*(2). https://doi.org/10.1029/2006WR005779

Syed, T., Famiglietti, J., Chen, J., Rodell, M., Seneviratne, S. I., Viterbo, P., & Wilson, C. R. (2005). Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance. *Geophysical Research Letters, 32*(24). https://doi.org/10.1029/2005gl024851

Tourian, M., Elmi, O., Chen, Q., Devaraju, B., Roohi, S., & Sneeuw, N. (2015). A spaceborne multisensor approach to monitor the desiccation of Lake Urmia in Iran. *Remote Sensing of Environment, 156*, 349–360. https://doi.org/10.1016/j.rse.2014.10.006

Tourian, M. J., Elmi, O., Shafaghi, Y., Behnia, S., Saemian, P., Schlesinger, R., & Sneeuw, N. (2021). HydroSat: a repository of global water cycle products from spaceborne geodetic sensors. *Earth System Science Data Discussions*, 1–42. https://doi.org/10.5194/essd-2021-174

Tourian, M. J., Reager, J. T., & Sneeuw, N. (2018). The total drainable water storage of the Amazon River Basin: A first estimate using GRACE. *Water Resources Research, 54*(5), 3290–3312. https://doi.org/10.1029/2017WR021674

Vaheddoost, B., & Aksoy, H. (2018). Interaction of groundwater with Lake Urmia in Iran. *Hydrological processes, 32*(21), 3283–3295. https://doi.org/10.1002/hyp.13263

Van Camp, M., Radfar, M., & Walraevens, K. (2010). Assessment of groundwater storage depletion by overexploitation using simple indicators in an irrigated closed aquifer basin in Iran. *Agricultural Water Management, 97*(11), 1876–1886. https://doi.org/10.1016/j.agwat.2010.02.006

Van Dijk, A. (2010). The Australian Water Resources Assessment System Technical Report 3. Landscape Model (version 0.5) Technical Description. https://doi.org/10.4225/08/5852dd9bb578c

Velicogna, I., Mohajerani, Y., Landerer, F., Mouginot, J., Noel, B., Rignot, E., Sutterley, T., van den Broeke, M., van Wessem, M., & Wiese, D. (2020). Continuity of ice sheet mass loss in Greenland and Antarctica from the GRACE and GRACE Follow-On missions. *Geophysical Research Letters, 47*(8), e2020GL087291. https://doi.org/10.1029/2020GL087291

Vishwakarma, B. D., Devaraju, B., & Sneeuw, N. (2018). What is the spatial resolution of GRACE satellite products for hydrology? *Remote Sensing, 10*(6), 852. https://doi.org/10.3390/rs10060852

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics, 12*(4), 367–395. https://www.jstor.org/stable/2235951

Voss, K. A., Famiglietti, J. S., Lo, M., De Linage, C., Rodell, M., & Swenson, S. C. (2013). 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*(2), 904–914. https://doi.org/10.1002/wrcr.20078

Wang, J., Song, C., Reager, J. T., Yao, F., Famiglietti, J. S., Sheng, Y., MacDonald, G. M., Brun, F., Schmied, H. M., Marston, R. A., et al. (2018). Recent global decline in endorheic basin water storages. *Nature geoscience, 11*(12), 926–932. https://doi.org/10.1038/s41561-018-0265-7

Wells, D. E., Vanicek, P., & Pagiatakis, S. D. (1985). Least squares spectral analysis revisited.

Willmott, C. J., & Matsuura, K. (1995). Smart interpolation of annually averaged air temperature in the United States. *Journal of Applied Meteorology, 34*(12), 2577–2586. https://doi.org/10.1175/1520-0450(1995)034<2577:sioaaa>2.0.co;2

Xie, P., Janowiak, J. E., Arkin, P. A., Adler, R., Gruber, A., Ferraro, R., Huffman, G. J., & Curtis, S. (2003). GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *Journal of Climate, 16*(13), 2197–2214. https://doi.org/10.1175/2769.1

Xue, X., Hong, Y., Limaye, A. S., Gourley, J. J., Huffman, G. J., Khan, S. I., Dorji, C., & Chen, S. (2013). Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu Basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *Journal of Hydrology, 499*, 91–99. https://doi.org/10.1016/j.jhydrol.2013.06.042

Yadollahie, M. (2019). The flood in Iran: a consequence of the global warming? *The international journal of occupational and environmental medicine, 10*(2), 54. https://doi.org/10.15171/ijoem.2019.1681

Yang, X., & Lu, X. (2014). Drastic change in China's lakes and reservoirs over the past decades. *Scientific reports, 4*(1), 1–10. https://doi.org/10.1038/srep06041

Zehtabian, G., Khosravi, H., & Ghodsi, M. (2010). High demand in a land of water scarcity: Iran. In *Water and sustainability in arid regions* (pp. 75–86). Springer. https://doi.org/10.1007/978-90-481-2776-4_5

Zhang & Gao. (2016). A novel algorithm for monitoring reservoirs under all-weather conditions at a high temporal resolution through passive microwave remote sensing. *Geophysical Research Letters, 43*(15), 8052–8059. https://doi.org/10.1002/2016GL069560

Zhang, K., Kimball, J. S., Nemani, R. R., & Running, S. W. (2010). A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research, 46*(9). https://doi.org/10.1029/2009WR008800

Zhang, Y., Kong, D., Gan, R., Chiew, F. H., McVicar, T. R., Zhang, Q., & Yang, Y. (2019). Coupled estimation of 500 m and 8-day resolution global evapotranspiration and gross primary production in 2002–2017. *Remote Sensing of Environment, 222*, 165–182. https://doi.org/10.1016/j.rse.2018.12.031

4

5

Probabilistic storage-based drought characterization

5.1. Introduction

In the earth sciences, any natural phenomenon that poses a threat to human life, infrastructure, or the environment is classified as a natural hazard (e.g., McPhillips et al., 2018; Pelling, 2003). Scientists in this field are dedicated to mapping, characterizing, and modeling natural hazards such as extreme atmospheric, geophysical, and hydrological events. The primary objective is to comprehend the magnitude, underlying factors, and scope of each hazard, which can then be utilized for risk management and future prediction (Garcin et al., 2008; Peng & Wang, 2015). Extreme weather conditions including heat waves, cold waves, and tropical cyclones can lead to extreme hydrological events, i.e., drought and flood. A flood event occurs when the flow of a river or stream quickly and unexpectedly overtops its banks, inundating the surrounding dry land. In contrast, droughts occur more gradually and affect wider areas. By examining the occurrence, frequency, and impacts of these extreme events, scientists can improve their understanding of natural hazards and develop strategies to mitigate their effects.

Understanding the temporal and spatial scale of drought is crucial for predicting and mitigating their impacts on water resources and human activities. Temporal scale refers to the duration and frequency of drought events. Droughts can last from a few weeks to several years, depending on the underlying climate conditions and the available water resources. They can also occur in different frequencies, ranging from rare, extreme events to more frequent and moderate droughts. Spatial scale refers to the extent and intensity of drought events. Droughts can occur at different spatial scales, from local, regional to continental, affecting different hydrological systems such as groundwater, surface water, and soil moisture.

Storage-based drought refers to the depletion of water resources in storage systems, such as groundwater, reservoirs, and lakes, resulting in a shortage of water supply for human activities and ecosystems. Understanding the dynamics of storage-based drought is essential for effective water management and allocation, especially in regions where water resources are limited and vulnerable to climate variability and change. Storage-based drought events can have significant impacts on the water supply for human activities, including domestic, industrial, and agricultural use. In regions where agriculture is a significant contributor to the local economy, storage-based drought events can result in crop failure, food insecurity, and economic losses. In addition, storage-based drought events can have detrimental effects on natural ecosystems, leading to the depletion of water resources and the loss of biodiversity.

To manage and mitigate the impacts of storage-based drought events, it is essential to understand their dynamics. This requires accurate and reliable information on the frequency, intensity, and duration of storage-based drought events, as well as their underlying causes and drivers. In this chapter, we aim to develop a probabilistic approach to characterize storage-based drought events based on historical data and climate projections. We will first provide a definition of drought and its impacts on water resource systems and human activities. Then, we will briefly mention some historical and recent drought events. We describe the methodology for characterizing storage-based drought events based on a probabilistic approach and the GRACE-H (hindcast GRACE, developed in chapter 3). Finally, we will present the results of our study, which will include comparison of the new drought index with the deterministic version of it. The outcome of this research will provide valuable information for water managers and decision-makers to develop effective strategies for mitigating the impacts of storage-based drought events.

5.1.1. Defining drought and its impacts

According to Wilhite (2000), drought is one of the most complex, but least understood among all natural hazards. Even though this was mentioned in Wilhite's book about 18 years ago, it is still true for several reasons. First, a worldwide accepted definition of drought is still missing. Although numerous definitions have been proposed so far, the literature lacks one precise definition that fulfils both the scientific and policymaker's expectations. The difficulty of defining drought comes from the different points of view from several sectors. Moreover, in many cases it is a hard task to distinguish between the role of nature and humans. Therefore, it is difficult to identify its key characteristics such as duration, severity, and spatial extent. Second, drought is known as a creeping phenomenon. It may stay in a region for years after the termination, which makes the determination of the onset and the end of drought quite challenging. Third, in contrast to floods, hurricanes, earthquakes, and tornadoes the effects of drought seldom result in structural damage (Mishra & Singh, 2010). Finally, drought affects a wide geographical area, in most cases wider than any other natural hazard, making it one of the most significant and far-reaching environmental challenges facing humanity today. It can affect entire regions, countries, or even continents, making it a complex and challenging phenomenon to monitor and manage. For example, In 2015, a severe drought in Brazil affected the entire country, including its largest city, São Paulo, and led to water rationing for millions of people. The other example is the recent drought in California, which lasted from 2012 to 2016, affected the entire state and was one of the worst in its history. The drought led to water shortages, wildfires, and economic losses in the agricultural sector.

Drought is an extreme event with a far-reaching web of economical, social, and environmental impacts that can occur in all climatic regimes (Mishra & Singh, 2010; Svoboda et al., 2002; Wilhite et al., 2007). With its wide range of effects, drought has impacted more people worldwide than other natural disasters over the past four decades (FAO, 2021; Wilhite, 2000). Many countries have faced considerable economic loss due to the drought events. For example, based on the NOAA's National Centers for Environmental Information (NCEI) report (https://www.ncei.noaa.gov/access/billions/), within the last four decades there have been approximately 29 significant droughts in the United States, costing the country at least $291 billion (more than $10 billion for each event). In Europe, Naumann et al. (2021) estimated the annual cost of droughts (especially in the southern and western parts) to be up to €9 billion, which could rise to more than €65 billion without climate action. Apart from the costs, climate change and unsustainable water management have raised the frequency and severity of drought events globally over the last two decades, which is expected to increase even more for the future (e.g., Coumou & Rahmstorf, 2012; Donat et al., 2016; Hisdal et al., 2001; Li et al., 2021; Teuling, 2018; Yu et al., 2014; Zhao et al., 2020).

5.1.2. Types of drought

Drought is traditionally classified according to three physically-based perspectives (Mishra & Singh, 2010):

- Meteorological: lack of precipitation

- Agricultural : deficit in soil moisture and vegetation response

- Hydrological: deficit in the runoff, streamflow, or groundwater storage

Moreover, a new class of drought called socio-economic drought (Mehran et al., 2015), also termed as anthropogenic drought (AghaKouchak et al., 2021), has been proposed recently. It is mainly related to the overuse or abuse of water due to human activities that can lead to water

stress in an area. In this chapter, we specifically focus on storage-based drought as a sub-class of hydrological drought.

The four perspectives of drought are different in timing, impact, and recovery rate with a rapid sense of dryness in meteorological drought, an intermediate response in agricultural and long-term dryness indicative of hydrological drought (Figure 5.1). It can be concluded from the figure that drought designations among sectors may or may not coincide in space and time (Wilhite, 2000). This means that several weeks of dryness may trigger an agricultural drought but it may not be sensed by groundwater or have little effect on streamflow. Moreover, the same duration of dryness may cause a severe drought in one region but may be classified as just abnormally dry in one another.
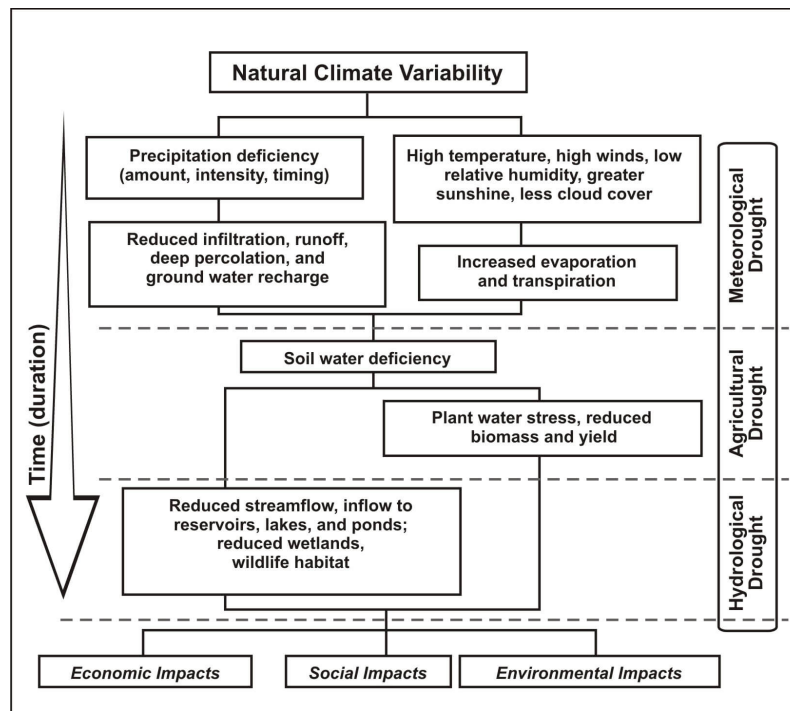

Figure 5.1: Sequence of drought occurrence and impacts for commonly accepted drought types (Wilhite, 2000)

5.1.3. Historical and recent occurrences of drought

Drought has been a recurring phenomenon throughout the history of human civilization, and its impacts have been observed globally. The historical and recent occurrences of drought can provide valuable insights into the patterns and trends of drought, as well as the societal and environmental impacts associated with this natural hazard. In this section, we provide a few example of the historical and recent occurrences of drought, with a focus on some notable events and their impacts.

Droughts have been documented throughout human history, and many ancient civilizations have experienced devastating drought events. One notable example is the Akkadian Empire, which was located in Mesopotamia (modern-day Iraq) and existed from 2334 BCE to 2154 BCE (Kerr, 1998). The empire collapsed during a prolonged drought, which was documented in cuneiform tablets from the time (Marshall, 2022). These tablets describe crop failures, famine, and societal collapse, and suggest that the drought lasted for several years. The Akkadian drought is thought to have been caused by a combination of natural climate variability and human activities, such as deforestation and overgrazing.

Another well-known historical drought event occurred in the 14th century in what is now modern-day Mexico. The region was home to the Aztec Empire, which flourished from the 14th to the 16th centuries. However, the empire was devastated by a series of droughts that occurred between 1350 and 1450 CE. These droughts caused crop failures, famine, and disease, and contributed to the eventual collapse of the empire. Researchers have suggested that the droughts may have been caused by changes in the Atlantic Multidecadal Oscillation, which is a natural climate pattern that affects sea surface temperatures in the Atlantic Ocean. The Aztec droughts serve as a reminder of the devastating impacts that climate variability can have on human societies, and underscore the importance of drought monitoring and early warning systems.
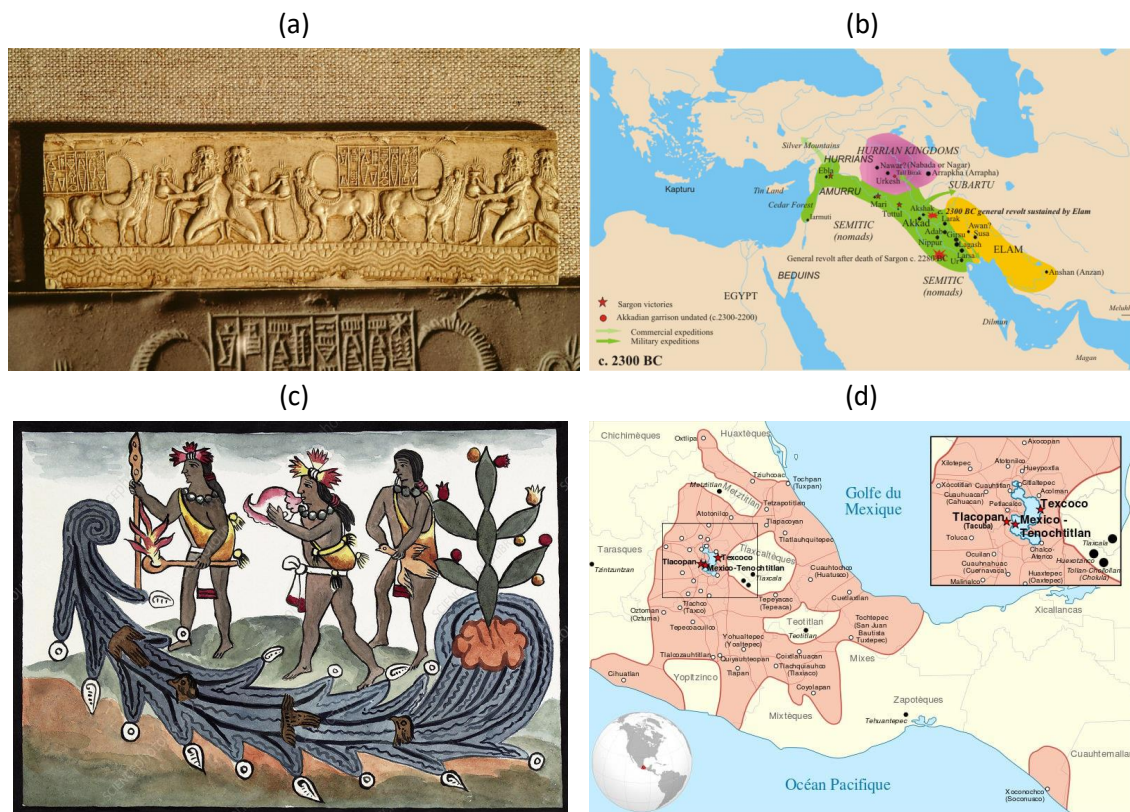
(a)   (b)

(c)   (d)

Figure 5.2: (a) Representation of a seal of an Akkadian king from circa 2200 BC, during the period of a severe drought that affected the empire. The seal depicts a god holding a vessel from which water flows, possibly symbolizing the hope for rain during the dry spell. Credit: The Art Archive/Shutterstock (b) Map of the Akkadian Empire (Illustration). Credit: World History Encyclopedia (c) Aztec drought rituals, 16th century (Courtesy: Library of congress/science photo library). The illustration depicts priests wearing jade necklaces conducting these ceremonies. To ward off drought, the priests cast decapitated birds into a stream. One priest carries a staff and a copal incense burner while another blows a conch shell. A bird is being killed by the third priest in the picture. The upper right of the illustration features a flowering cactus, which represents Tenochtitlan, now known as Mexico City. (d) A map indicating the maximum extent of the Aztec civilization. Credit: World History Encyclopedia

Regarding more recent events, one of the most severe droughts in recorded history occurred in the Sahel region of Africa in the late 1960s and early 1970s, which caused widespread famine and displacement of millions of people. The drought was attributed to a combination of natural climate variability and human factors, such as deforestation and overgrazing. The severity of this event highlighted the vulnerability of the region to drought and the need for improved drought preparedness and response measures. The effects of the Sahel drought were not limited to Africa, as it had ripple effects across the globe. The decline in food production caused by the drought led to a significant increase in food prices, which impacted consumers worldwide.

Additionally, the displacement of millions of people resulted in the spread of disease and social unrest, exacerbating the humanitarian crisis. This event serves as a reminder of the complex interplay between natural and human factors in drought occurrence and its impacts.

From the late 1990s to the late 2000s, Australia experienced a severe and prolonged drought that became known as the Millennium Drought. The drought had a significant impact on agriculture, water resources, and biodiversity, and led to major policy changes and investments in water management. The drought was caused by a combination of factors, including a prolonged period of below-average rainfall, increased evaporation rates due to higher temperatures, and the over-extraction of water from the Murray-Darling Basin, which is a major river system that supports agriculture, urban centers, and environmental ecosystems in the region. During the Millennium Drought, many rivers, lakes, and wetlands in the region dried up, and water levels in reservoirs reached critically low levels. This resulted in severe water shortages for both urban and rural communities, and many farmers were forced to sell or cull their livestock and crops, leading to significant economic losses. The impacts of the Millennium Drought were also felt beyond Australia, as it affected global food prices due to reduced agricultural productivity, particularly in the wheat and rice industries. The Millennium Drought ended in 2010, when heavy rainfall replenished the water levels in many of the region's dams and reservoirs, although the impacts of the drought are still being felt in some parts of the region.

The Syrian drought was a prolonged period of below-average rainfall that occurred in the Eastern Mediterranean region, including Syria, from 2006 to 2011. The drought was considered one of the most severe in the region's history and was exacerbated by anthropogenic factors such as overuse of water resources, mismanagement of land, and unsustainable agricultural practices. The drought had significant impacts on Syria's agriculture and economy, leading to the displacement of hundreds of thousands of people, particularly in rural areas. Many farmers were forced to abandon their crops and livestock due to a lack of water, causing a ripple effect on food prices, unemployment, and poverty. The drought also exacerbated political tensions and social unrest, contributing to the Syrian Civil War, which began in 2011. The Syrian drought highlighted the vulnerability of countries in the Eastern Mediterranean region to droughts and the need for effective water management and adaptation strategies to cope with the impacts of climate change.

The United States has experienced several severe drought events throughout its history. One of the most notable droughts is the Dust Bowl, which occurred in the 1930s in the central and southern Great Plains region of the country. The combination of prolonged drought, poor land management practices, and high winds led to massive dust storms that devastated agriculture and caused widespread ecological damage. The Dust Bowl had significant economic and social impacts, including widespread poverty, migration, and changes in agricultural practices and government policies. The event led to the establishment of soil conservation programs and other measures to prevent future ecological disasters. Another notable drought in the United States occurred in California from 2011 to 2017. The drought, which was considered the worst in the state's history, was caused by a combination of factors, including a lack of precipitation and record-high temperatures. The drought led to significant impacts on the state's agriculture, energy, and water resources. It also had social and economic impacts, including the loss of jobs and the migration of people to other states. The California drought highlighted the need for more effective water management practices and the development of drought-resistant crops and technologies. It also spurred efforts to increase public awareness about water conservation and the impacts of climate change on water resources.

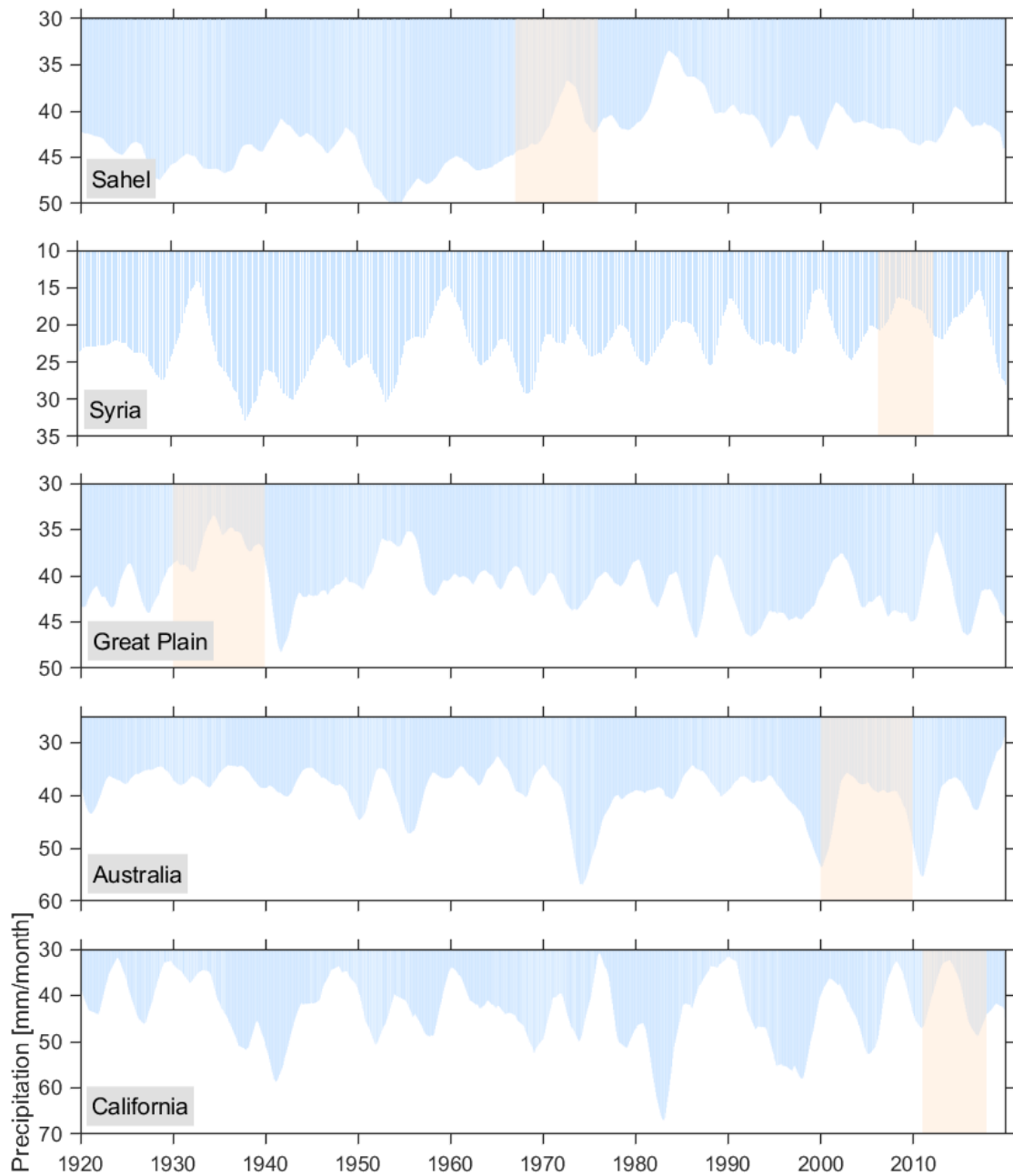Figure 5.3: Visualizing the impact of severe droughts in selected regions through a long-term record of precipitation using GPCC (Global Precipitation Climatology Centre) version 2020 data from 1920 to 2019. Each region's recorded drought period is highlighted in orange for better understanding. The GPCC data can be downloaded from https://opendata.dwd.de/climate_environment/GPCC/html/download_gate.html

5.1.4. Drought monitoring and early warning systems

Traditionally, our responses to drought events have been conducted in a very reactive, post-event fashion, which has been called *crisis management.* A crisis is an event that occurs at a specific time that is usually unforeseen, public in nature, and has the potential to cause great harm to different sectors. In contrast, a risk is an activity or event that has the potential to harm those sectors. The negative impacts of droughts can be mitigated using risk management strategies rather than crisis management (Wilhite, 2000). Whereas crisis management is concerned with responding to, managing, and recovering from an unforeseen event, *risk management* is concerned with identifying, assessing, and mitigating activities or events that could harm society. Risk management in the drought concept includes preparedness planning, mitigation, monitoring, and early warning, and also prediction to reduce the impacts of drought.

Such a proactive response requires drought monitoring, including early warning and forecasting systems, at national to local levels (Wilhite et al., 2007). Drought early warning systems are able to trigger activities that help communities cope with the effects of drought (Mishra & Singh, 2011). In order to improve drought monitoring and assist decision-makers, drought indices are developed which summarize drought complexity into a single number and characterize it into onset, intensity, frequency, and duration (Ahmadalipour et al., 2017; Wilhite, 2000; Zargar et al., 2011). Such indices describe drought using single or multiple climatic and hydrometeorological variables data such as precipitation, streamflow, evapotranspiration, temperature, and snowpack (Svoboda, Fuchs, et al., 2016).

Scientists have developed such indices using multiple factors including drought nature and characteristics and the impacts considered. Manifold drought indices have been developed (more than 150, Niemeyer et al. (2008)). The nature of drought indices reflects different events and conditions; they can reflect climate dryness anomalies (mainly based on precipitation) or correspond to delayed agricultural and hydrological impacts such as soil moisture loss or lowered reservoir levels. In addition, the categorization of drought indices can also be based on the data and technology used. For example, a considerable number of indices use remote-sensing imaging to detect vegetation health as an indicator of drought. Several drought indices have been developed so far, each related to one or two different classes of drought (i.e., meteorological, agricultural, hydrological, and socio-economical droughts). Zargar et al. (2011) reviewed 74 operational and proposed drought indices and describes research directions. In each of these indices either the deficiency in water for each component of the TWS or the impact of drought (i.e., change in the vegetation quality and quantity) have been considered to develop the index and each index has its own strengths and weaknesses. Some of the more common and practical indices include (see e.g., Saemian, 2021 or Zargar et al. (2011) for more comprehensive list).

Table 5.1: A list of some common drought indices.

Acronym	full name	source
PDSI	Palmer Drought Severity Index	Palmer (1965)
RAI	Rainfall Anomaly Index	Van Rooy (1965)
CMI	Crop Moisture Index	Palmer (1968)
BMDI	Bhalme and Mooly Drought Index	Bhalme and Mooley, 1980
SWSI	Surface Water Supply Index	Shafer, 1982
SPI	Standardized Precipitation Index	McKee (1995) and McKee et al. (1993)
SMDI	Soil Moisture Deficit Index	Hollinger et al., 1993
VCI	Vegetation Condition Index	Liu and Kogan (1996)
SPEI	Standardized Precipitation Evapotranspiration Index	Vicente-Serrano et al. (2010)

5.1.5. Brief literature review on storage-based drought index

A comprehensive characterization of drought extension requires observation of the TWS, i.e., the water on and beneath the land surface, including snow, surface water, soil moisture, and groundwater storage (Zhao et al., 2017). The GRACE-derived TWS estimates have been used to develop indices to examine drought at regional to global scales. The identification of droughts using GRACE was first proposed by Seitz et al. (2008), whose work made a significant early contribution to the field. Then, Yirdaw et al. (2008) developed Total Storage Deficit Index (TSDI) based on Palmer Drought Severity Index (PDSI, Palmer (1965)) and Soil Moisture Deficit Index (SMDI, Narasimhan and Srinivasan (2005)), for characterizing 2002/2003 Canadian Prairie droughts. Many studies have used TSDI or a modified version for characterizing in regional scales over Canada (Agboma et al., 2009), the Greater Horn of Africa (Awange et al., 2016), China (Cao et al., 2015), the Tigris-Euphrates Basin (Chao et al., 2018), and over Markazi basin in Iran (Hosseini-Moghari et al., 2019).

Afterward, Chen et al. (2010) utilized TWSA from GRACE to identify drought events in the La Plata basin, in Argentina. Houborg et al. (2012) developed GRACE-based drought indicators and enhanced its spatial resolution using data assimilation. Long et al. (2013) and Wang et al. (2014) investigated the correlation between the depletion in GRACE TWSA time series and the drought events in Texas and the Haihe river basin HRB in Northern China, respectively. In another attempt, Thomas et al. (2014) presented a GRACE-based drought characterization framework using TWSA from a GRACE-based monthly mean and evaluated the result of drought severity over Amazon, Zambezi, Texas, and the southeastern United States. Yi and Wen (2016) developed the GRACE-based Hydrological Drought Index (GHDI) for characterizing drought in the continental United States between 2003 and 2012, keeping the concept of PDSI. Sinha et al. (2017) expanded the method presented by Thomas et al. (2014) and devised a new drought index called the Water Storage Deficit Index (WSDI) to characterize drought in India. Recently some more sophisticated drought indices using GRACE TWS anomaly have been introduced (e.g., A long-term standardized GRACE reconstructed TWSA index (SGRTI) by Zhong et al. (2023), and GRACE-DSI by Zhao et al. (2017)).

Despite their potential for monitoring and assessing the TWS drought at regional to global scales, the indices mentioned above follow a deterministic approach while ignoring the fact that drought characterization using GRACE observations comes with its own uncertainties. Such uncertainties exist due to the inherent uncertainty in the GRACE data, the various postprocessing approaches of GRACE data, and different options for de-aliasing products. In this chapter, we aim to incorporate the uncertainty of GRACE TWSA in the drought characterization.

Table 5.2: A list for some of GRACE based drought indices.

Acronym	full name	source
TSDI	Total Storage Deficit Index	Yirdaw et al. (2008)
GHDI	GRACE-based Hydrological Drought Index	Yi and Wen (2016)
WSDI	Water Storage Deficit Index	Sinha et al. (2017)
GRACE-DSI	GRACE-Drought Severity Index	Zhao et al. (2017)
SGRTI	Standardized GRACE Reconstructed TWSA Index	Zhong et al. (2023)

5.2. Data

5.2.1. GRACE TWSA

Three mascon products are generated by different processing centers for the GRACE and GRACE-FO satellite missions: the Goddard Space Flight Center (GSFC), the Jet Propulsion Laboratory (JPL), and the German Research Centre for Geosciences (GFZ). Each of these products has its own strengths and weaknesses that should be considered when selecting a dataset for a particular study. The GSFC mascon product has a resolution of 1°, which allows for detailed investigations of regional-scale gravity field variations. However, it may have larger errors in the higher harmonics due to limitations in the spherical harmonic inversion method used in its processing. The JPL mascon product has a coarser resolution of 3°, but has a more accurate representation of the higher harmonics due to a regularized least squares method used in its processing. However, it may have lower accuracy in certain regions due to variations in the gravity field signal-to-noise ratio. The GSFC mascon product has a similar spatial resolution to the JPL product, but also includes a higher-resolution (0.5°) dataset for Antarctica. It uses a combination of spherical harmonic and mascon modeling to improve the accuracy of its gravity field estimates, but may have larger errors at high latitudes due to limitations in the mascon representation of the gravity field.

For this study, to ensure reproducibility of our results, we utilized the mascon product generated by the Goddard Space Flight Center (GSFC) of NASA. This dataset has been widely used in the geodesy and Earth science communities to investigate a range of phenomena, including hydrology, glaciology, and solid Earth dynamics. The GSFC mascon product can be downloaded from https://earth.gsfc.nasa.gov/geo/data/grace-mascons. We used the latest version of the dataset available at the time of our analysis, which covers the period from August 2002 to the present. The dataset includes monthly gravity field solutions with grid size of 0.5°.

5.3. Methodology

To properly characterize storage-based drought using GRACE TWSA observations, certain decisions must be made regarding data treatment. These decisions include how to handle trends in the data, whether to normalize the data, whether to include time integration, whether to use a GRACE-only approach or an augmented dataset, and finally how to incorporate uncertainty in the TWSA estimation. In the following sections, we discuss each of these steps and provide a justification for our choices.

5.3.1. Handling trends

Several studies have suggested that before investigating drought indices using the GRACE TWSA time series, detrending is necessary (e.g., Hosseini-Moghari et al., 2020; Khorrami & Gunduz, 2021; Liu et al., 2020). Liu et al. (2020), for instance, have demonstrated that without detrending TWSA time series drought severity can be overestimated over some basins in China after 2013. While the soil moisture data suggests that the drought ceased in September 2014, their GRACE indices (GRACE-DSI) show a continuous drought condition. In contrast to the aforementioned studies, we deliberately retain the trend in the time series. Our reasoning is that the trend reflects long-term changes in climate, such as temperature increases or precipitation pattern alterations, which can affect the frequency and severity of droughts. Eliminating the trend would essentially omit these long-term changes from the analysis, providing an incomplete understanding of the hydrological system.

To demonstrate the impact of detrending, we calculated the TWSA time series in two real cases using the SSA approach with a 24-month window to remove the trend in the data. The two cases, the Tigris basin in the Middle East with a negative trend and the Niger basin in Africa with a positive trend, are presented in Figure 5.4 and Figure 5.5, respectively. In each case, we compared the results from two scenarios: one without detrending, denoted by the solid line in (c) and (d) and labelled as (a), and one with detrending, shown as the dashed line in (c) and (e) and labelled as (b).

The Tigris basin experienced a prolonged period of water loss, particularly after 2007, which is apparent in the red area in Figure 5.4(d). Detrending the data resulted in higher values for the climatology (the long-term monthly mean, see Equation 5.1) compared to the non-detrended data, as shown in Figure 5.4(c), and caused oscillations between wet and dry years, as seen in Figure 5.4(e). On the other hand, the Niger basin exhibited a positive trend mainly after 2010, resulting in wetter years in the basin, as depicted in Figure 5.5(d). Although detrending did not significantly alter the climatology, as illustrated in Figure 5.5(c), it did reveal dry years after 2010, which is inconsistent with actual conditions.
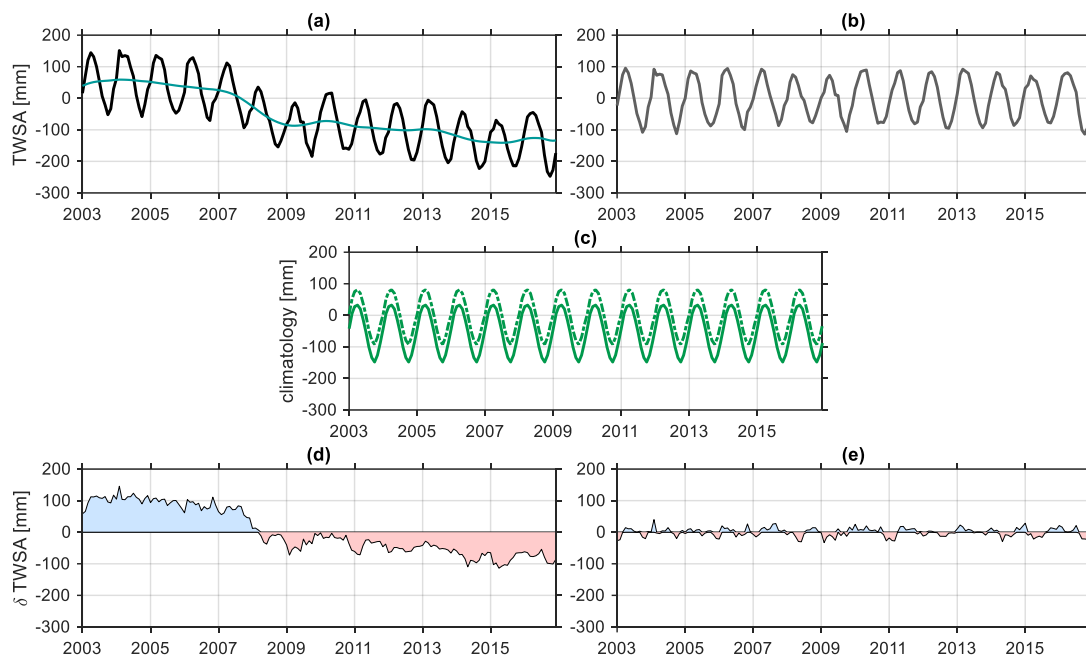

Figure 5.4: This figure presents a comprehensive analysis of TWSA for the Tigris river basin in the Middle East, using data from GRACE satellite mission. (a) shows the time series of TWSA from GRACE, along with its inter-annual variations which are extracted using the Singular Spectrum Analysis (SSA) approach with a 24-month window. (b) displays the TWSA after removing the inter-annual variations, highlighting the long-term trends. (c) illustrates the climatology of TWSA, which represents the long-term monthly mean. The solid and dashed lines represent the climatology obtained from (a) and (b), respectively. (d) and (e) show the TWSA residuals, obtained by subtracting the corresponding climatology from panels (a) and (b), respectively. These residual plots reveal the short-term fluctuations in TWSA that are not captured by the climatology.

5.3.2. Normalizing

Normalizing refers to the transformation of a variable to follow a standard normal distribution (i.e., mean of 0 and a standard deviation of 1). This process is commonly used to eliminate the effects of differing units or scales in the data, allowing for easier comparison and analysis across variables. In the context of developing a drought index from GRACE TWSA, normalization can help to account for differences in the magnitude of TWSA across different regions or time periods, and can improve the comparability of drought conditions across regions. One of the ad-
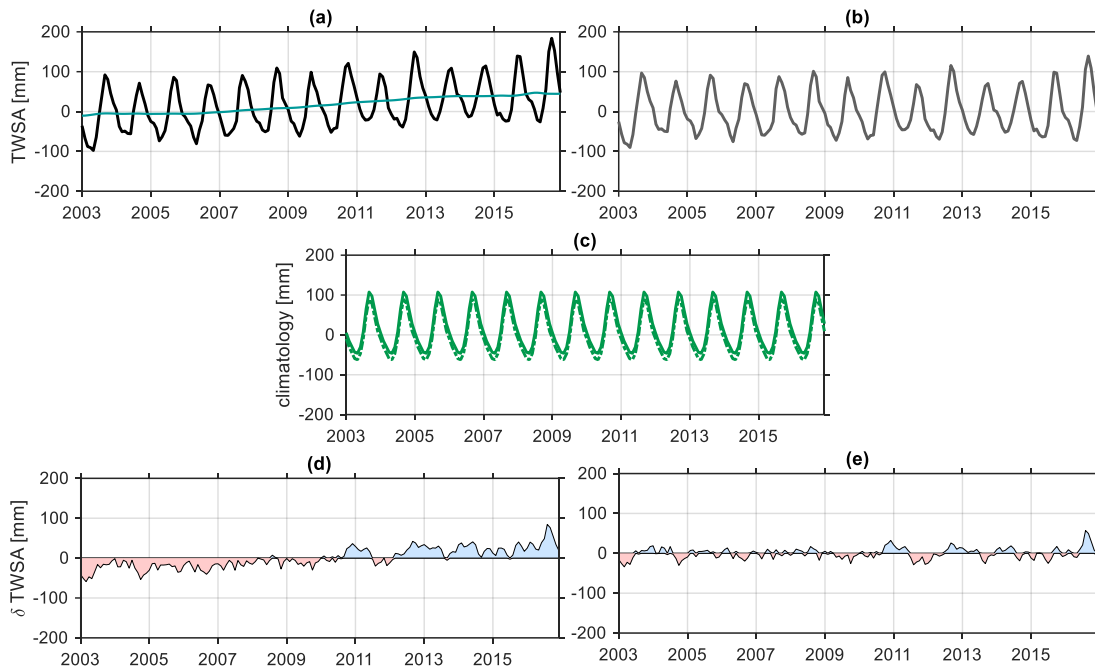
Figure 5.5: Same as Figure 5.4 but for Niger river basin in West Africa, flowing through 10 countries: Guinea, Mali, Niger, Benin, Burkina Faso, Cote d'Ivoire, Ghana, Togo, Cameroon, and Nigeria.

vantages of normalization is that it can make it easier to compare the severity of drought across regions that have different hydrological regimes and climate conditions. For example, regions with higher average TWSA values may still experience severe drought conditions if their TWSA levels drop below a certain threshold, and normalization can help to identify these conditions more accurately. Normalization can also help to reduce the impact of outliers or extreme values in the data, which can skew the results of drought analysis.

However, there are also some potential disadvantages to normalization. Normalization can obscure the actual magnitude of the TWSA values, and can make it harder to interpret the results of drought analysis in terms of actual water volumes. In addition, the choice of normalization method can affect the results of the analysis, and different normalization methods may be more or less appropriate depending on the characteristics of the data being analyzed. There are several approaches that can be used to normalize the drought index for TWSA including Z-score normalization (standardization) and min-max scaling. Standardization involves subtracting the mean TWSA value from each observation and then dividing by the standard deviation of TWSA. Min-max scaling involves scaling the TWSA values to a range between 0 and 1, by subtracting the minimum value from each observation and then dividing by the range of the entire time series.

n this study, instead of applying normalization techniques to the time series of TWSA observations, we choose to calculate the TWSA deviation from the normal condition (δTWSA). This deviation represents the difference between the observed TWSA values and the expected or average TWSA under normal conditions. By calculating this deviation, we can assess the extent to which the current TWSA values deviate from what is considered normal, providing valuable information about the severity or anomalous nature of the drought conditions. This approach allows us to analyze the TWSA data in relation to its baseline or expected state without altering the original values through normalization.

5.3.3. GRACE-only approach or with an augmented dataset

Drought is a prolonged water deficiency. To estimate such a relative TWS deficiency, we need a reference. It is common to compute the long-term monthly average as the reference or *normal* condition in a particular region. The long-term monthly mean, also known as the *climatology*, is calculated by:

$$\overline{\text{TWSA}}[t_m] = \frac{1}{N} \sum_{y=y_1}^{y_N} \text{TWSA}[t_{y,m}] \tag{5.1}$$

$$\delta\text{TWSA}[t_{y,m}] = \text{TWSA}[t_{y,m}] - \overline{\text{TWSA}}[t_m] \tag{5.2}$$

where $\overline{\text{TWSA}}[t_m]$ stands for the climatology, y is the year and can vary from y_1 to y_N, and $m = 1,2,3,\ldots,12$ is the month of a year. To quantify the TWS deficiencies, we subtract the climatology from the GRACE TWSA time series. We call the result δTWSA, in which negative values represent the water storage deficits.

TWSA is subject to significant variation over short time scales due to factors such as precipitation and evapotranspiration. Obtaining accurate long-term monthly means from shorter time series can be challenging. However, using long period data sets allows us to average out the effects of short-term variability, resulting in a more robust estimate of the long-term average conditions. This is particularly important for studies aimed at characterizing storage-based drought over time, as reliable baseline estimates of long-term monthly means are required. Calculating the climatology over at least 30 years, preferably 60 years, is standard practice, as this time frame provides a sufficiently long and representative sample of the Earth's climate variability. The climate system is complex and dynamic, exhibiting natural variations in behavior over time scales ranging from years to centuries. By calculating the climatology over 30 years or more, we can obtain a robust estimate of the long-term average conditions. This estimate can be used as a reference point for understanding how the climate is changing over time. The 30-year time frame is long enough to capture significant climate events and phenomena that influence long-term climate averages, such as El Niño-Southern Oscillation (ENSO) and the North Atlantic Oscillation (NAO). A shorter time frame can result in biased estimates of long-term averages, as it may not capture the full range of natural variability in the climate system. Therefore, a time frame of 30 years is the minimum requirement for calculating the climatology, with longer time frames providing even more robust estimates of the long-term average conditions. Various options can be considered to have an augmented longer time period of TWSA (more than 30 years) that has already been discussed in chapter 3. For this study, we used the hindcasted GRACE TWSA, which was developed and discussed in chapter 3 and covers the period from 1980 to 2012, to obtain a long-term monthly mean from a longer time period.

5.3.4. Time integration

Time integration involves aggregating the GRACE TWSA data over a given period of time, which can be done using different methods such as running averages, moving sums, or cumulative sums. One advantage of time integration is that it can help to smooth out short-term variability in the TWSA data, making it easier to identify longer-term trends and patterns in the data. This can be particularly useful for characterizing drought, which often involves a slow depletion of water resources over an extended period of time.

In our study, we chose to use a 3-month running average to smooth the TWSA deviation:

$$\delta\text{TWSA}[t_{y,m}]_{\textbf{smoothed}} = (1/3) \cdot \sum_{i=t_{y,m-1}}^{i=t_{y,m+1}} \delta\text{TWSA}[i] \tag{5.3}$$

Such an average reduces the effects of short-term fluctuations due to precipitation and other factors. By using a running average, we were able to capture the overall drought or wetness signal more accurately, which is important for characterizing storage-based drought over time. One disadvantage of time integration is that it introduces a lag in the data, which means that changes in TWSA may not be immediately apparent. For example, if there is a sudden increase in precipitation, it may take several months for this to be reflected in the time-integrated TWSA data. Additionally, the choice of the time window used for time integration can also have an impact on the results obtained, as different time scales may be more appropriate for characterizing different types of drought or wetness events.

5.3.5. Handling uncertainty

GRACE observations inherently contain uncertainties arising from various factors such as the data itself, post-processing approaches, and options for de-aliasing products. Additionally, GRACE-H, utilized for long-term TWSA estimation, includes uncertainty for each observation, thereby introducing uncertainty to the calculated climatology. Assuming that the error in TWSA and the error in $\overline{\text{TWSA}}$ are uncorrelated, error propagation provides:

$$\sigma^2_{\overline{\text{TWSA}[t_m]}} = \frac{1}{N} \sum_{y=y_1}^{y_N} \text{TWSA}[t_{y,m}] \tag{5.4}$$

$$\sigma_{\delta\text{TWSA}[t_{y,m}]} = \left(\sigma^2_{\text{TWSA}[t_{y,m}]} + \sigma^2_{\overline{\text{TWSA}[t_m]}} \right)^{1/2} \tag{5.5}$$

where N is the number of years in the long-term dataset. Figure 5.6 shows the δTWSA for the Death Valley basin in the US after applying the above-mentioned steps.

To account for the uncertainty, we need to utilize a probabilistic method that incorporates δTWSA and its corresponding uncertainty. One common way to generate random samples from a normal distribution is using the Box-Muller transform, which is a widely used method for generating independent and identically distributed normal random variables. The Box-Muller transform works by transforming pairs of uniformly distributed random variables into pairs of normal random variables, using trigonometric functions and logarithms. To apply the Box-Muller transform to a time series, we first need to estimate the mean and standard deviation of the data, which can be obtained from the time series itself or from a subset of the data. Once we have estimated the mean and standard deviation of the data, we can use the Box-Muller transform to generate random samples from the normal distribution with the same mean and standard deviation. Specifically, we can follow these steps:

1. Generate pairs of uniformly distributed random variables U_1 and U_2, with values between 0 and 1.

2. Compute the variables $R = \sqrt{-2 \cdot \log(U_1)}$ and $\theta = 2 \cdot \pi \cdot U_2$.

3. Compute the normal-distributed samples $Z_1 = R \cdot \cos(\theta)$ and $Z_2 = R \cdot \sin(\theta)$.

4. Scale the samples by the desired mean and standard deviation, by computing $Z = \text{mean} + \text{std} \cdot Z$, where mean and std are the estimated mean and standard deviation of the data.
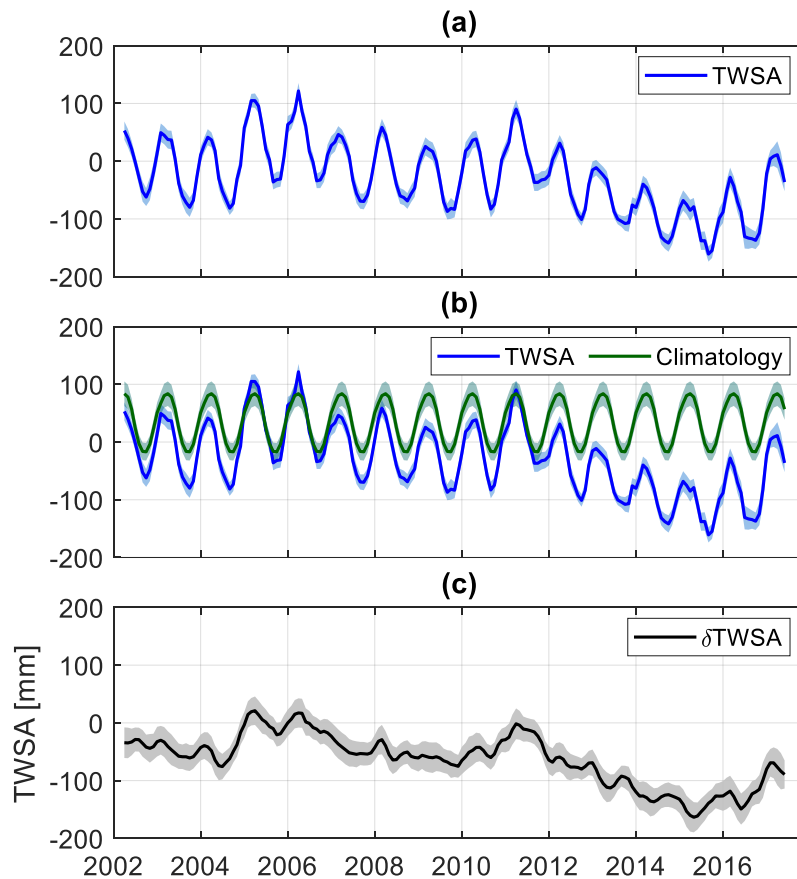
Figure 5.6: (a) Time series of the long-term TWSA together with the uncertainty (light blue envelope). (b) Same as (a) together with the long-term climatology (1980–2012) and its uncertainty (light green envelope). (c) δTWSA using the equation **??** together with its uncertainty (gray envelope). Here the results are shown for the Death Valley basin in the US.

To perturb the realizations of the time series, we can use Monte Carlo Simulation (Metropolis & Ulam, 1949; Mooney, 1997), which involves generating multiple sets of random samples and computing the corresponding statistics for each set. By using Monte Carlo Simulation with the Box-Muller transform, we can generate a large number of possible realizations of the time series, which can be used to estimate the uncertainty associated with the data, and to explore the range of possible outcomes under different scenarios or assumptions. Figure 5.7 shows the results over Death Valley in California, US. The density of the realizations is high around the mean signal and decays following a Gaussian distribution.

To accurately characterize drought, it is essential to quantify the quantile values associated with each of the drought categories. This can be achieved by using the Cumulative Distribution Function (CDF) to project all the realizations to their corresponding quantile values. In accordance with the U.S. Drought Monitor (USDM) (see Table 5.3), we use five drought categories: D0 (abnormally dry, representing percentile values in the range of 20–30%), D1 (moderate drought, representing percentile values in the range of 10–20%), D2 (severe drought, representing percentile values in the range of 5–10%), D3 (extreme drought, representing percentile values in the range of 2–5%), and D4 (exceptional drought, representing percentile values less than 2%).

Each month, we analyze the cloud of points based on the cumulative distribution functions (CDFs) (Figure 5.8) and observe that it falls within the range of δTWSA$\pm 3\cdot\sigma_{\delta\text{TWSA}}$. To determine
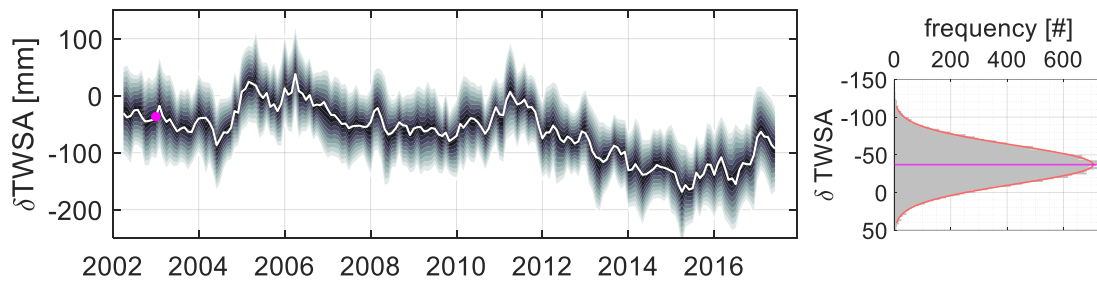
Figure 5.7: The TWSA deficit (δTWSA) together with its 10 000 realizations, calculated using Monte Carlo simulation. Here the results are shown for the Death Valley basin in the US. The distribution of realizations for January 2003 is presented in the right plot. The colored dot on the right plot shows the location of the January 2003 on the time series.

the probability of each drought category, we calculate the ratio of the number of points in each category to the total number of points in all categories. For decision-making purposes, we can select the category with the highest probability as the final drought category for that month. The flowchart of the probabilistic approach proposed in this thesis is shown in Figure 5.9.

Table 5.3: USDM classification for drought characterization (Svoboda et al., 2002)

Category	Description	Color	Percentile range
D0	Abnormal		21–30
D1	Moderate		11–20
D2	Severe		6–10
D3	Extreme		3–5
D4	Exceptional		0–2
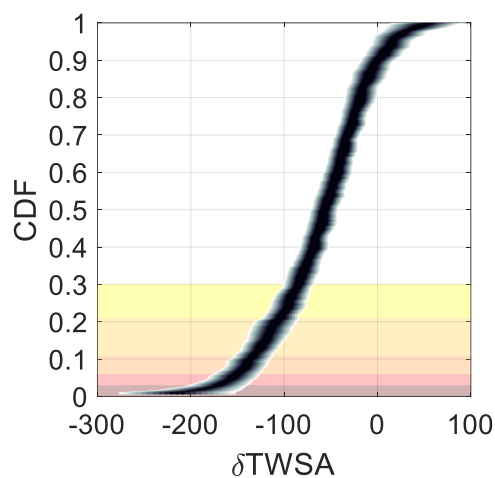
Figure 5.8: The cumulative distribution functions (CDF) of δTWSA realizations. The five categories of drought that has been shown in Table 5.3 are illustrated with their corresponding color.
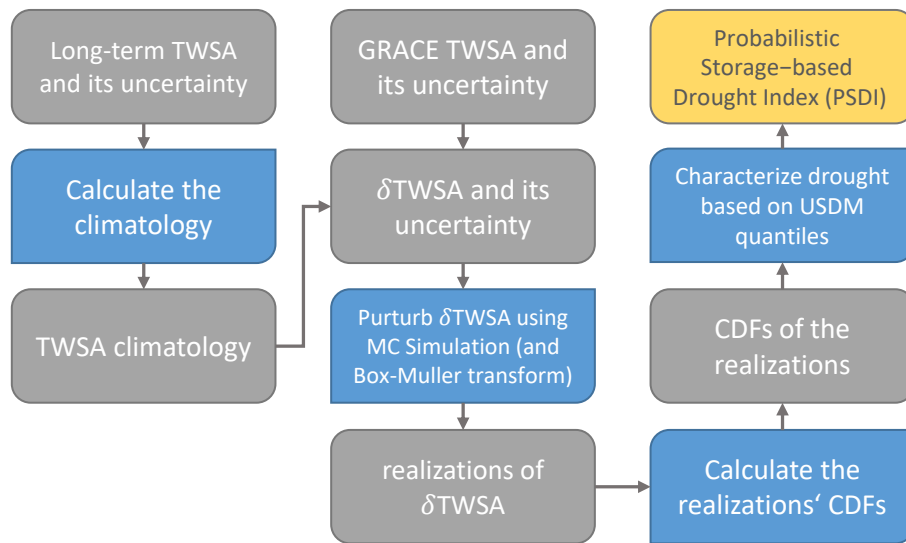
Figure 5.9: Flowchart of the PSDI approach proposed in this study.

5.4. Results and Analysis

5.4.1. Results over selected basins

Figure 5.10 provides a visual comparison between two approaches for characterizing drought: probabilistic (PSDI) and deterministic (SDI). This comparison is conducted over several selected basins. The distribution of the basins is shown in the top panel of the Figure 5.10. For each basin, the drought categories, ranging from the status of no drought to exceptional drought (D4), are displayed in columns. The probability assigned to each category at every time step is depicted using gray-scale. The deterministic perspective is illustrated with red boxes, allowing for an easy and intuitive comparison of the two approaches.

The PSDI approach offers a more nuanced understanding of drought conditions compared to the SDI approach. This is because PSDI captures the uncertainty associated with drought severity, while the SDI approach may oversimplify the classification of drought conditions. Although the SDI categorization is often the most probable category according to the PSDI, the neighboring categories may also have significant probabilities. This tendency becomes more pronounced as the intensity of the drought increases. This can be attributed to the lower slope of the CDF curve over more severe droughts and the wider range of quantile values.

5

Figure 5.10: Top: The global distribution of the selected basins. Bottom: The SDI together with its probability (PSDI) for selected basins. Red boxes show the drought category from the deterministic approach. The basins are shown in two groups considering the time period with more frequency of drought, first row within 2015–2016 and the second row within 2006–2017. The "-" represent no drought or normal state of the water storage.
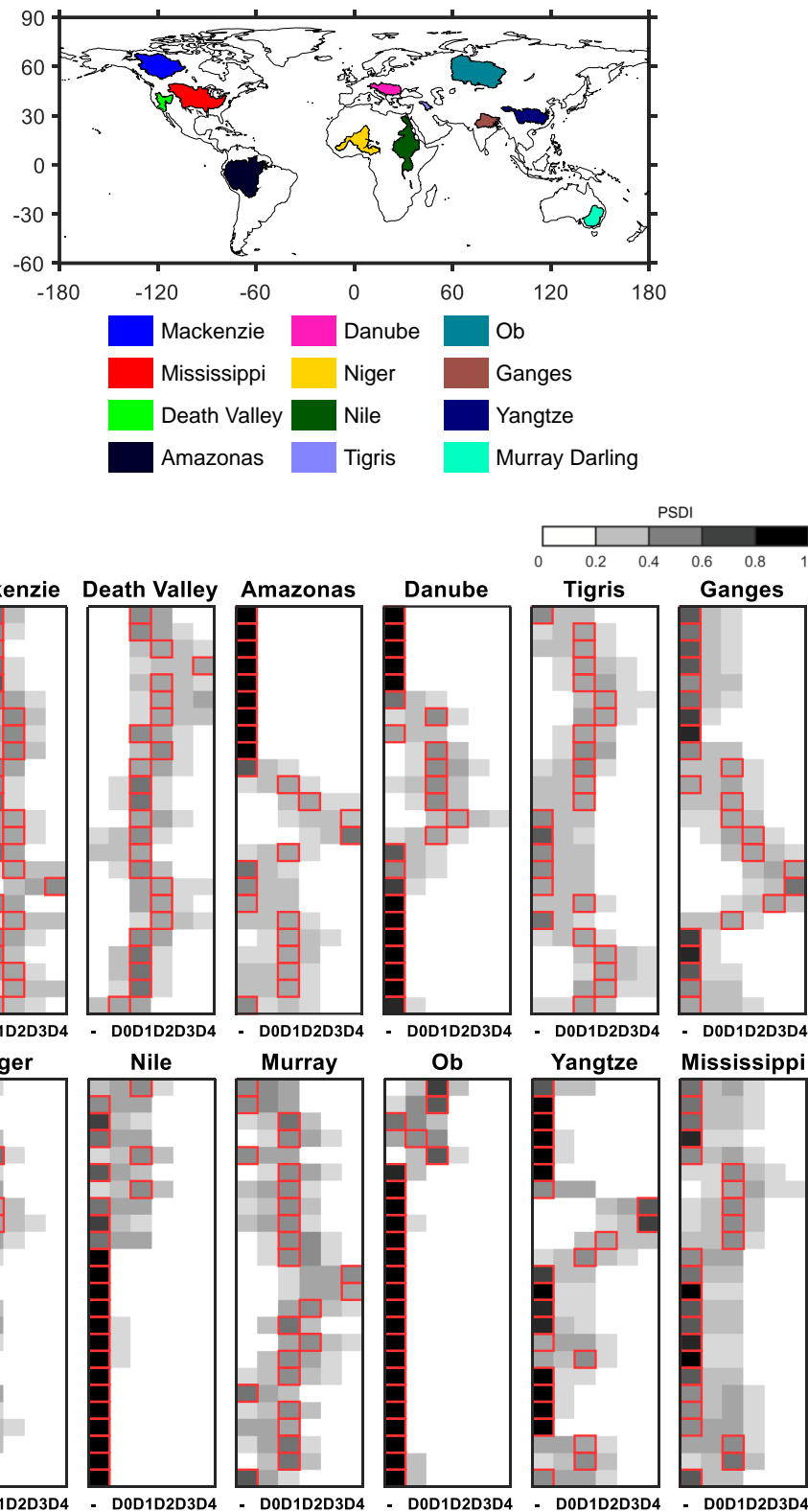
5.4.2. Performance of the PSDI during extreme hydrologic events

To evaluate the accuracy of the PSDI, we compared its results with several documented extreme hydrologic events that occurred between 2002 and 2016. These events include the moderate to exceptional drought that affected the United States in 2012 (Ault et al., 2013; Boyer et al., 2013), southern Europe (Oikonomou et al., 2020; Spinoni et al., 2015), and many regions of the Middle East in 2007–2008 (Barlow et al., 2016). Southern Africa also experienced severe to exceptional droughts from 2005 to early 2006 (Nicholson, 2014), while central Argentina and Paraguay were affected by droughts throughout 2009 (Guha-Sapir et al., 2016). Additionally, Australia suffered from the worst drought recorded since European settlement in the 2000s, with a peak in 2006 that affected many regions from the south to the east, including the agricultural lands of the Murray-Darling basin (Heberger, 2012; Van Dijk et al., 2013). Overall, the PSDI exhibited high performance in characterizing droughts in the selected hydrologic events (see Figure 5.11).
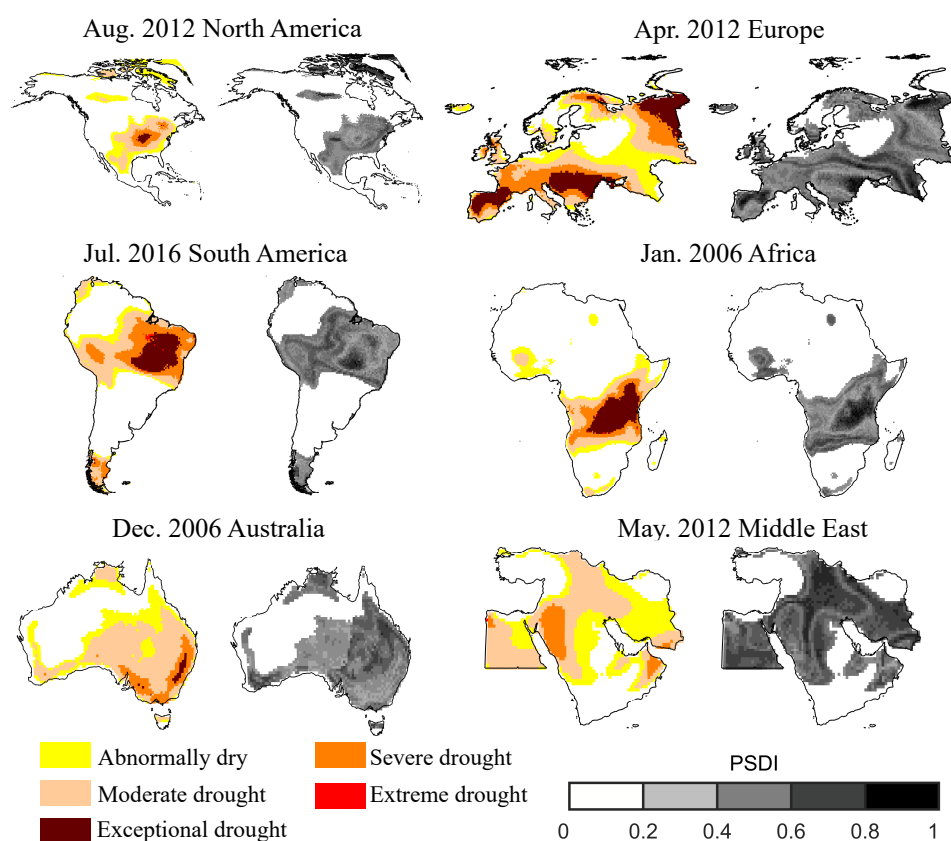
Figure 5.11: SDI performance during some reported drought events. The maps of the probability associated with the SDI, PSDI, are shown in grayscale.

5.5. Discussion

The GRACE satellite mission has provided highly precise quantitative estimates of Total Water Storage Anomaly (TWSA). However, due to the pre- and post-processing of the GRACE observations, these estimates are accompanied by uncertainty. Most centers that provide GRACE level-2 products include formal errors in the instruments and observation geometry, expressed as spherical harmonic coefficients. Additionally, mascon products provide estimates of TWSA uncertainty, represented as spherical harmonics or global grids. It is important to note that errors in the level-2 products and mascon approaches vary among different centers, which utilize diverse processing methods and background models. Figure 5.12 demonstrates a comparison

of two widely used mascon datasets, JPL RL06-v02 and GSFC RL06-v02. 5.12(a) presents a comparison of the time series of global land-averaged TWSA uncertainty between 2003 and 2016. The time series reveals a sharp peak in 2015, followed by a positive trend attributable to the battery failure (Bandikova et al., 2019; Mayer-Gürr et al., 2018; Save et al., 2016).

Moreover, Figure 5.12 (b–c) presents the average TWSA uncertainty from the aforementioned products and the current study between 2003 and 2016. The highest uncertainty values occur in Greenland, the Amazonas, the Indian sub-continent, and the northwest of Canada. A spatio-temporal comparison outlined in (Save et al., 2016) demonstrates that error estimates for GRACE TWSA observations vary depending on the dataset or approach utilized. Additionally, the impact of such discrepancies is not negligible in the probabilistic drought characterization provided in this study for the aforementioned regions.
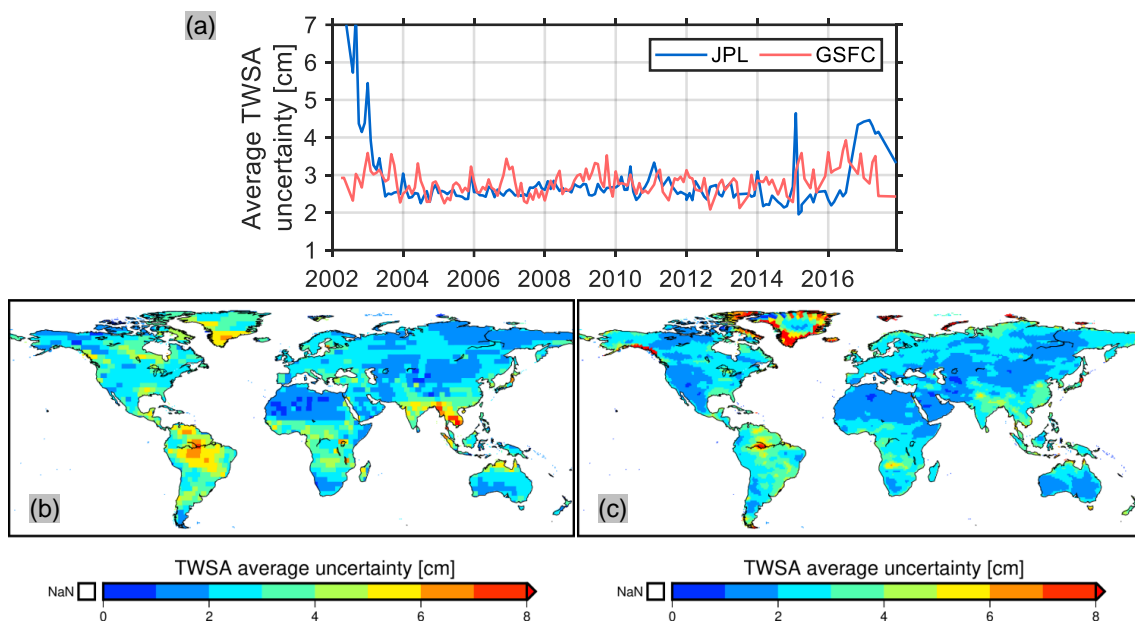
Figure 5.12: (a) Time series of the global averaged TWSA uncertainty from two mascons datasets, namely, JPL RL06-v03 and GSFC RL06-v02. Global distribution of the averaged TWSA uncertainty from April 2002 to June 2017 from (b) GSFC and JPL RL06-v03 (c).

GRACE(-FO) provides a comprehensive perspective on changes in terrestrial water storage, which encompasses various components such as canopy water, soil moisture, surface water (e.g., rivers and lakes), groundwater, and snow equivalent water. However, a GRACE-based drought indicator should be used in conjunction with proxy observations, such as water balance fluxes and commonly used drought indices (e.g., Palmer Drought Severity Index, Standardized Precipitation Index, and Standardized Precipitation Evapotranspiration Index), to create a more realistic drought monitoring and mitigation system. Despite this limitation, GRACE(-FO) observations offer a unique ability to measure deep water and reflect the combined impacts of human activities and climate change. GRACE(-FO) products provide global solutions at monthly time intervals, with typical latency of 1-2 months, which is not optimal for an early warning system. Some studies have proposed processing methods to improve temporal sampling to 10 days or even daily measurements with a latency of several days (e.g., Jäggi et al., 2019; Mayer-Gürr et al., 2018). However, the daily GRACE products have low spatial resolution (up to spherical harmonics degree and order of about 40), which limits their application for regional drought analysis, particularly in coastal regions that are highly affected by leaked signal effects Eicker et al., 2020.

Bibliography

Agboma, C., Yirdaw, S., & Snelgrove, K. (2009). Intercomparison of the total storage deficit index (TSDI) over two Canadian Prairie catchments. *Journal of hydrology, 374*(3-4), 351–359. https://doi.org/10.1016/j.jhydrol.2009.06.034

AghaKouchak, A., Mirchi, A., Madani, K., Di Baldassarre, G., Nazemi, A., Alborzi, A., Anjileli, H., Azarderakhsh, M., Chiang, F., Hassanzadeh, E., et al. (2021). Anthropogenic drought: Definition, challenges, and opportunities. https://doi.org/10.1029/2019RG000683

Ahmadalipour, A., Moradkhani, H., Yan, H., & Zarekarizi, M. (2017). Remote sensing of drought: vegetation, soil moisture, and data assimilation. In *Remote sensing of hydrological extremes* (pp. 121–149). Springer. https://doi.org/10.1007/978-3-319-43744-6_7

Ault, T., Henebry, G., De Beurs, K., Schwartz, M., Betancourt, J. L., & Moore, D. (2013). The false spring of 2012, earliest in North American record. *Eos, Transactions American Geophysical Union, 94*(20), 181–182. https://doi.org/10.1002/2013EO200001

Awange, J., Schumacher, M., Forootan, E., Heck, B., et al. (2016). Exploring hydro-meteorological drought patterns over the Greater Horn of Africa (1979–2014) using remote sensing and reanalysis products. *Advances in water resources, 94*, 45–59. https://doi.org/10.1016/j.advwatres.2016.04.005

Bandikova, T., McCullough, C., Kruizinga, G. L., Save, H., & Christophe, B. (2019). Grace accelerometer data transplant. *Advances in Space Research, 64*(3), 623–644. https://doi.org/10.1016/j.asr.2019.05.021

Barlow, M., Zaitchik, B., Paz, S., Black, E., Evans, J., & Hoell, A. (2016). A review of drought in the Middle East and southwest Asia. *Journal of climate, 29*(23), 8547–8574. https://doi.org/10.1175/JCLI-D-13-00692.1

Bhalme, H. N., & Mooley, D. A. (1980). Large-scale droughts/floods and monsoon circulation. *Monthly Weather Review, 108*(8), 1197–1211. https://doi.org/10.1175/1520-0493(1980)108<1197:LSDAMC>2.0.CO;2

Boyer, J., Byrne, P., Cassman, K., Cooper, M., Delmer, D., Greene, T., Gruis, F., Habben, J., Hausmann, N., Kenny, N., et al. (2013). The US drought of 2012 in perspective: a call to action. *Global Food Security, 2*(3), 139–143. https://doi.org/10.1016/j.gfs.2013.08.002

Cao, Y., Nan, Z., & Cheng, G. (2015). GRACE gravity satellite observations of terrestrial water storage changes for drought characterization in the arid land of northwestern China. *Remote Sensing, 7*(1), 1021–1047. https://doi.org/10.3390/rs70101021

Chao, N., Luo, Z., Wang, Z., & Jin, T. (2018). Retrieving Groundwater Depletion and Drought in the Tigris-Euphrates Basin Between 2003 and 2015. *Groundwater, 56*(5), 770–782. https://doi.org/10.1111/gwat.12611

Chen, J., Wilson, C., Tapley, B., Longuevergne, L., Yang, Z., & Scanlon, B. (2010). Recent La Plata basin drought conditions observed by satellite gravimetry. *Journal of Geophysical Research: Atmospheres, 115*(D22). https://doi.org/10.1029/2010JD014689

Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nature climate change, 2*(7), 491. https://doi.org/10.1038/nclimate1452

Donat, M. G., Lowry, A. L., Alexander, L. V., O'Gorman, P. A., & Maher, N. (2016). More extreme precipitation in the world's dry and wet regions. *Nature Climate Change, 6*(5), 508. https://doi.org/10.1038/nclimate2941

Eicker, A., Jensen, L., Wöhnke, V., Dobslaw, H., Kvas, A., Mayer-Gürr, T., & Dill, R. (2020). Daily GRACE satellite data evaluate short-term hydro-meteorological fluxes from global atmospheric reanalyses. *Scientific reports, 10*(1), 1–10. https://doi.org/10.1038/s41598-020-61166-0

FAO. (2021). *Drought* [(accessed September 16, 2021)].

Garcin, M., Desprats, J.-F., Fontaine, M., Pedreros, R., Attanayake, N., Fernando, S., Siriwardana, C., De Silva, U., & Poisson, B. (2008). Integrated approach for coastal hazards and risks in Sri Lanka. *Natural Hazards and Earth System Sciences, 8*(3), 577–586. https://doi.org/10.5194/nhess-8-577-2008

Guha-Sapir, D., Below, R., & Hoyois, P. (2016). EM-DAT: the CRED/OFDA international disaster database.

Heberger, M. (2012). Australia's millennium drought: Impacts and responses. In *The world's water* (pp. 97–125). Springer. https://doi.org/10.5822/978-1-59726-228-6_5

Hisdal, H., Stahl, K., Tallaksen, L. M., & Demuth, S. (2001). Have streamflow droughts in Europe become more severe or frequent? *International Journal of Climatology, 21*(3), 317–333. https://doi.org/10.1002/joc.619

Hollinger, S., Isard, S., & Welford, M. (1993). A new soil moisture drought index for predicting crop yields. *Preprints, Eighth Conference on Applied Climatology*, 187–190.

Hosseini-Moghari, S.-M., Araghinejad, S., Ebrahimi, K., Tang, Q., & AghaKouchak, A. (2020). Using GRACE satellite observations for separating meteorological variability from anthropogenic impacts on water availability. *Scientific Reports, 10*(1), 15098. https://doi.org/10.1038/s41598-020-71837-7

Hosseini-Moghari, S.-M., Araghinejad, S., Ebrahimi, K., & Tourian, M. J. (2019). Introducing modified total storage deficit index (MTSDI) for drought monitoring using GRACE observations. *Ecological indicators, 101*, 465–475. https://doi.org/10.1016/j.ecolind.2019.01.002

Houborg, R., Rodell, M., Li, B., Reichle, R., & Zaitchik, B. F. (2012). Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations. *Water Resources Research, 48*(7). https://doi.org/10.1029/2011WR011291

Jäggi, A., Weigelt, M., Flechtner, F., Güntner, A., Mayer-Gürr, T., Martinis, S., Bruinsma, S., Flury, J., Bourgogne, S., Steffen, H., et al. (2019). European gravity service for improved emergency management (EGSIEM)—from concept to implementation. *Geophysical journal international, 218*(3), 1572–1590. https://doi.org/10.1093/gji/ggz238

Kerr, R. A. (1998). Sea-floor dust shows drought felled Akkadian Empire. *Science, 279*(5349), 325–326. https://doi.org/10.1126/science.279.5349.325

Khorrami, B., & Gunduz, O. (2021). An enhanced water storage deficit index (EWSDI) for drought detection using GRACE gravity estimates. *Journal of Hydrology, 603*, 126812. https://doi.org/10.1016/j.jhydrol.2021.126812

Li, W., Pan, R., Jiang, Z., Chen, Y., Li, L., Luo, J.-J., Zhai, P., Shen, Y., & Yu, J. (2021). Future changes in the frequency of extreme droughts over China based on two large ensemble simulations. *Journal of Climate, 34*(14), 6023–6035. https://doi.org/10.1175/JCLI-D-20-0656.1

Liu, W., & Kogan, F. (1996). Monitoring regional drought using the vegetation condition index. *International Journal of Remote Sensing, 17*(14), 2761–2782. https://doi.org/10.1080/01431169608949106

Liu, X., Feng, X., Ciais, P., Fu, B., Hu, B., & Sun, Z. (2020). GRACE satellite-based drought index indicating increased impact of drought over major basins in China during 2002–2017. *Agricultural and Forest Meteorology, 291*, 108057. https://doi.org/10.1016/j.agrformet.2020.108057

Long, D., Scanlon, B. R., Longuevergne, L., Sun, A. Y., Fernando, D. N., & Save, H. (2013). GRACE satellite monitoring of large depletion in water storage in response to the 2011 drought in Texas. *Geophysical Research Letters*, *40*(13), 3395–3401. https://doi.org/10.1002/grl.50655

Marshall, M. (2022). Did a mega drought topple empires 4,200 years ago? *Nature*, *601*(7894), 498–501. https://doi.org/10.1038/d41586-022-00157-9

Mayer-Gürr, T., Behzadpour, S., Ellmer, M., Kvas, A., Klinger, B., Strasser, S., & Zehentner, N. (2018). ITSG-Grace2018–monthly, daily and static gravity field solutions from GRACE, GFZ dataservices. https://doi.org/10.5880/ICGEM.2018.003

McKee, T. B. (1995). Drought monitoring with multiple time scales. *Proceedings of 9th Conference on Applied Climatology, Boston, 1995*.

McKee, T. B., Doesken, N. J., Kleist, J., et al. (1993). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, *17*(22), 179–183.

McPhillips, L. E., Chang, H., Chester, M. V., Depietri, Y., Friedman, E., Grimm, N. B., Kominoski, J. S., McPhearson, T., Méndez-Lázaro, P., Rosi, E. J., et al. (2018). Defining extreme events: A cross-disciplinary review. *Earth's Future*, *6*(3), 441–455. https://doi.org/10.1002/2017EF000686

Mehran, A., Mazdiyasni, O., & AghaKouchak, A. (2015). A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand. *Journal of Geophysical Research: Atmospheres*, *120*(15), 7520–7533. https://doi.org/10.1002/2015JD023147

Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American statistical association*, *44*(247), 335–341. https://doi.org/10.1080/01621459.1949.10483310

Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. *Journal of hydrology*, *391*(1-2), 202–216. https://doi.org/10.1016/j.jhydrol.2010.07.012

Mishra, A. K., & Singh, V. P. (2011). Drought modeling–A review. *Journal of Hydrology*, *403*(1-2), 157–175. https://doi.org/10.1016/j.jhydrol.2011.03.049

Mooney, C. Z. (1997). *Monte Carlo simulation* (Vol. 116). Sage publications.

Narasimhan, B., & Srinivasan, R. (2005). Development and evaluation of Soil Moisture Deficit Index (SMDI) and Evapotranspiration Deficit Index (ETDI) for agricultural drought monitoring. *Agricultural and Forest Meteorology*, *133*(1-4), 69–88. https://doi.org/10.1016/j.agrformet.2005.07.012

Naumann, G., Cammalleri, C., Mentaschi, L., & Feyen, L. (2021). Increased economic drought impacts in Europe with anthropogenic warming. *Nature Climate Change*, *11*(6), 485–491. https://doi.org/10.1038/s41558-021-01044-3

Nicholson, S. E. (2014). A detailed look at the recent drought situation in the Greater Horn of Africa. *Journal of Arid Environments*, *103*, 71–79. https://doi.org/10.1016/j.jaridenv.2013.12.003

Niemeyer, S., et al. (2008). New drought indices. *Options Méditerranéennes. Série A: Séminaires Méditerranéens*, *80*, 267–274.

Oikonomou, P. D., Karavitis, C. A., Tsesmelis, D. E., Kolokytha, E., & Maia, R. (2020). Drought characteristics assessment in Europe over the Past 50 Years. *Water Resources Management*, *34*(15), 4757–4772. https://doi.org/10.1007/s11269-020-02688-0

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30). US Department of Commerce, Weather Bureau.

Palmer, W. C. (1968). Keeping track of crop moisture conditions, nationwide: The new crop moisture index. https://doi.org/10.1080/00431672.1968.9932814

Pelling, M. (2003). *Natural disaster and development in a globalizing world*. Routledge.

5

Peng, S.-H., & Wang, K. (2015). Risk evaluation of geological hazards of mountainous tourist area: a case study of Mengshan, China. *Natural Hazards, 78*(1), 517–529. https://doi.org/10.1007/s11069-015-1724-8

Saemian, P. (2021). *Characterizing storage-based drought using satellite gravimetry* (Master's thesis). https://doi.org/10.18419/opus-11734

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution CSR GRACE RL05 mascons. *Journal of Geophysical Research: Solid Earth, 121*(10), 7547–7569. https://doi.org/10.1002/2016JB013007

Seitz, F., Schmidt, M., & Shum, C. (2008). Signals of extreme weather conditions in Central Europe in GRACE 4-D hydrological mass variations. *Earth and Planetary Science Letters, 268*(1-2), 165–170. https://doi.org/10.1016/j.epsl.2008.01.001

Shafer, B. (1982). Developemnet of a surface water supply index (SWSI) to assess the severity of drought conditions in snowpack runoff areas. *Proceedings of the 50th Annual Western Snow Conference, Colorado State University, Fort Collins, 1982.*

Sinha, D., Syed, T. H., Famiglietti, J. S., Reager, J. T., & Thomas, R. C. (2017). Characterizing drought in India using GRACE observations of terrestrial water storage deficit. *Journal of Hydrometeorology, 18*(2), 381–396. https://doi.org/10.1175/JHM-D-16-0047.1

Spinoni, J., Naumann, G., Vogt, J. V., & Barbosa, P. (2015). The biggest drought events in Europe from 1950 to 2012. *Journal of Hydrology: Regional Studies, 3*, 509–524. https://doi.org/10.1016/j.ejrh.2015.01.001

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., Rippey, B., Tinker, R., Palecki, M., Stooksbury, D., et al. (2002). The drought monitor. *Bulletin of the American Meteorological Society, 83*(8), 1181–1190. https://doi.org/10.1175/1520-0477-83.8.1181

Svoboda, M. D., Fuchs, B. A., et al. (2016). *Handbook of drought indicators and indices.* World Meteorological Organization Geneva, Switzerland.

Teuling, A. J. (2018). A hot future for European droughts. *Nature Climate Change, 8*(5), 364–365. https://doi.org/10.1038/s41558-018-0154-5

Thomas, A. C., Reager, J. T., Famiglietti, J. S., & Rodell, M. (2014). A GRACE-based water storage deficit approach for hydrological drought characterization. *Geophysical Research Letters, 41*(5), 1537–1545. https://doi.org/10.1002/2014GL059323

Van Dijk, A. I., Beck, H. E., Crosbie, R. S., de Jeu, R. A., Liu, Y. Y., Podger, G. M., Timbal, B., & Viney, N. R. (2013). The Millennium Drought in southeast Australia (2001–2009): Natural and human causes and implications for water resources, ecosystems, economy, and society. *Water Resources Research, 49*(2), 1040–1057. https://doi.org/10.1002/wrcr.20123

Van Rooy, M. (1965). A rainfall anomaly index independent of time and space. *Notos, 14*(43), 6.

Vicente-Serrano, S. M., Beguerıéa, S., & López-Moreno, J. I. (2010). A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of climate, 23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wang, J., Jiang, D., Huang, Y., & Wang, H. (2014). Drought analysis of the Haihe River Basin based on GRACE terrestrial water storage. *The Scientific World Journal, 2014.* https://doi.org/10.1155/2014/578372

Wilhite, D. A. (2000). Drought as a natural hazard: Concepts and definitions.

Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water resources management, 21*(5), 763–774. https://doi.org/10.1007/s11269-006-9076-5

Yi, H., & Wen, L. (2016). Satellite gravity measurement monitoring terrestrial water storage change and drought in the continental United States. *Scientific reports, 6*, 19909. https://doi.org/10.1038/srep19909

Yirdaw, S. Z., Snelgrove, K. R., & Agboma, C. O. (2008). GRACE satellite observations of terrestrial moisture changes for drought characterization in the Canadian Prairie. *Journal of Hydrology*, *356*(1-2), 84–92. https://doi.org/10.1016/j.jhydrol.2008.04.004

Yu, M., Li, Q., Hayes, M. J., Svoboda, M. D., & Heim, R. R. (2014). Are droughts becoming more frequent or severe in China based on the standardized precipitation evapotranspiration index: 1951–2010? *International Journal of Climatology*, *34*(3), 545–558. https://doi.org/10.1002/joc.3701

Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews*, *19*(NA), 333–349. https://doi.org/doi.org/10.1139/a11-013

Zhao, C., Brissette, F., Chen, J., & Martel, J.-L. (2020). Frequency change of future extreme summer meteorological and hydrological droughts over North America. *Journal of Hydrology*, *584*, 124316. https://doi.org/10.1016/j.jhydrol.2019.124316

Zhao, M., Velicogna, I., & Kimball, J. S. (2017). 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*(8), 2117–2129. https://doi.org/10.1175/JHM-D-16-0182.1

Zhong, Y., Hu, E., Wu, Y., An, Q., Wang, C., Bai, H., & Gao, W. (2023). Reconstructing a long-term water storage-based drought index in the yangtze river basin. *Science of The Total Environment*, *883*, 163403.

5

6

Conclusion and outlook

6.1. Conclusion

GRACE and its successor GRACE-FO are highly successful missions with numerous applications. They provide valuable insights into the Earth system and have been used for a variety of purposes. For instance, they have helped to understand the water cycle, monitor changes in groundwater, study climate change, and monitor natural hazards. GRACE has been instrumental in measuring ice mass changes, which is crucial for understanding the impact of climate change on water resources and ecosystems. Moreover, GRACE has enabled scientists to pinpoint areas at risk of water stress and track changes in water availability, which is vital for effective water allocation and management. Overall, GRACE and GRACE-FO are essential tools for studying the Earth's water and understanding its dynamics. Despite such a success story, GRACE(-FO) faces a number of challenges that must be addressed. This thesis aimed to address some of these challenges, including:

1. the GRACE(-FO) post-processing approaches and consistency evaluation

2. expanding the short record of GRACE(-FO) observations

3. analyzing long-term changes in water storage

4. accounting GRACE(-FO)'s uncertainty in drought characterization

(1) The GRACE(-FO) post-processing approaches and consistency:

In the second chapter, we considered one of the challenges of using GRACE data, namely the various post-processing approaches. The chapter provides a brief overview of the main steps involved in obtaining TWSA from Level-2 products of GRACE and GRACE-FO, and for each step, we examined the various data/methods that can be chosen and assessed the impact of these choices on the final TWSA. Among the various gravity field solutions, ITSG-Grace2018 outperforms other solutions with clear noise reduction even at the highest degrees (> 80) for higher degrees. LUH and AIUB showed the highest noise level among all the solutions. Among official centers, CSR demonstrates slightly better performance while JPL and GFZ lie within the same range of error. The evaluation of the stochastic of the gravity field solutions shows that ITSG-Grace2018 delivers more realistic values than all other solutions, while centers like SWPU, WHU, and AIUB are mostly overestimating, especially for d/o higher than 30.

The evaluation of various post-processing methods for obtaining TWSA from GRACE data shows that ITSG-Grace2018 outperforms other solutions in terms of noise reduction and realistic stochastic values. The choice of GRACE solution can lead to significant disparities in RMS and trends, with larger basins exhibiting relatively minor differences and smaller to medium basins exhibiting the largest discrepancies. The ellipsoidal correction mostly leads to an increase in both RMS and trend values, but there are some regions where it results in a decrease in trend values. The choice of filter significantly affects the RMS of the TWSA signal in arid to semi-arid regions, but has little impact on trend values. Lastly, the correlation coefficients between GRACE and altimetric water level time series are generally slightly higher for mascons solutions than for processed spherical harmonics, but the results should be interpreted with caution due to the coarse spatial resolution of GRACE and the limited scope of lakes considered.

(2) Expanding the short record of GRACE(-FO) observations:

The GRACE mission's record is relatively short compared to other satellite missions, thereby restricting the scope of studies requiring long-term data. Chapter3 aimed to address this issue by employing global models to expand the temporal coverage of GRACE observations. The

third chapter introduces the models briefly and evaluates them against GRACE TWSA observations. The ensemble mean (EM) and the ensemble weighted mean (EWM) of the models are computed over major river basins as the base approach. In addition, six methods, including Multivariate Linear Regression (MLR), Non-Negative Least Squares (NNLS), Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Gaussian Process Regression (GPR), were introduced and executed to extend TWSA before the GRACE era. The results were compared with GRACE within the GRACE period (i.e., Apr 2002 to Dec 2012) and with high-resolution SLR TWSA outside the GRACE period (i.e., 1992–2002).

Based on the results, all six methods improved the base approaches (i.e., EM and EWM), highlighting the need for advanced and sophisticated methods to accurately reconstruct TWSA in different regions and climates. GPR and MLR outperformed the others, while SVM and DT had poor performance in most basins. The choice of method for reconstructing TWSA should consider the specific climate conditions of the region under analysis. The MLR method may be a suitable option for dry sub-humid regions, while the GPR method may be a good choice for most regions. This chapter offers a reliable approach for reconstructing long-term total water storage anomaly fields before the GRACE period (i.e., before 2002) which can be used for studies that necessitate long-term observations of TWSA like drought monitoring or calculating the long-term climatology.

(3) Analyzing long-term changes in water storage:

The fourth chapter has focused on improving our understanding of water storage variations at the continental to basin scale, using GRACE(-FO) observations integrated with other water flux measurements. The case study of water scarcity in Iran has demonstrated the importance of scrutinizing GRACE(-FO) observations for managing water resources in regions facing similar challenges. The study revealed unsustainable water management in Iran, and there is a need for effective and efficient water management policies in the country.

Moreover, the chapter analyzed the non-linear trend and corresponding uncertainty of GRACE-TWSA for sub-continents during the period spanning from 2003 to 2021. The results showed that all sub-continents examined in this study have experienced a decline in water storage compared to their status in 2003, with North America registering the largest TWS loss. The Middle East, Central Asia, and Eastern Europe have also experienced significant losses, and Europe has shown a negative trend in TWS from 2015 to 2020. The study provides crucial insights into the decline in global water storage and highlights the need for sustainable management of water resources to address the issue.

(4) Accounting GRACE(-FO)'s uncertainty in drought characterization:

The GRACE satellite mission has uncertainty in its data due to measurement error, instrument noise, and uncertainties in models and algorithms. While the mission has taken measures to minimize these uncertainties, researchers should account for them when analyzing the data. Chapter five proposed a probabilistic approach to utilize GRACE data in characterizing storage-based drought, incorporating uncertainty using the Box-Muller transform and Monte Carlo Simulation. The proposed approach can provide a more comprehensive and informative assessment of drought risk and is beneficial for decision-makers managing water resources. The thesis discusses decisions regarding data treatment, including time integration, and justifies the choices made. The new index developed is called Probabilistic Storage-based Drought Index (PSDI), and it is consistent with reported hydrological droughts. Comparing PSDI to conventional method (SDI) results shows the importance of incorporating uncertainty in drought characterization for realistic risk management.

6.2. Perspective for future research

Apart from the significant achievements highlighted in the previous chapters of this thesis, several challenges still persist, creating avenues for further investigations:

Improving the spatio-temporal resolution of GRACE(-FO): GRACE and GRACE-FO have relatively low temporal resolution, which limits their ability to resolve rapid changes in the water cycle. Besides, the spatial resolution of GRACE and GRACE-FO is also limited, which can make it difficult to accurately measure changes in small-scale hydrological processes, such as groundwater storage. To address these limitations, the Next Generation of Satellite Gravity Missions (NGGM) has been proposed by the European Space Agency (ESA). Additionally, a collaborative effort between ESA and NASA, known as the Mapping and Geodesy Imaging Constellation (MAGIC), offers a double pair contribution to overcome these challenges. These projects aim to provide a much higher spatial (100 km) and temporal resolution (3 days) of the Earth's gravity field compared to the current missions. The future mission will use advanced laser interferometry and microwave technologies to improve the measurement precision and reduce the measurement noise. The satellites are expected to be launched in the mid-2020s and will provide significant improvements in our understanding of the Earth's gravity field and its variations.

Another approach to improving the spatio-temporal resolution of GRACE(-FO) is to use Data Assimilation techniques that combine the satellite gravity data with other types of data, such as satellite altimetry, in-situ observations, and hydrological models. Moreover, Machine Learning (ML) and Deep Learning (DL) methods have also been proposed as potential alternatives to the traditional methods of processing the satellite gravity data. ML and DL models can learn complex patterns and relationships from large volumes of data and can provide accurate and fast predictions of the gravity field. However, a major challenge with these approaches is the validation of the results with independent measurements, as GRACE(-FO) have been so far the main dataset that can be considered as the truth. One possible approach is to use vertical motion data from Global Navigation Satellite System (GNSS) stations, which can serve as a proxy for the gravity field changes. GNSS stations measure the position of the ground with respect to a fixed reference frame, and any changes in the gravity field can cause vertical motions of the ground surface. However, such a validation also presents challenges such as the fact that the vertical motion data from GNSS stations are affected by various factors, such as atmospheric delays, tectonic motion, and post-seismic deformation, which need to be carefully accounted for in the validation process.

Correction for non-hydrological signals: The signals measured by GRACE and GRACE-FO can be contaminated by other factors which must be corrected for in order to obtain accurate measurements of changes in the water cycle. For instance, post-seismic deformation caused by earthquakes can persist for years, and such prolonged deformation can influence the GRACE data. Additionally, the Earth's gravity field is substantially altered by the phenomenon of ocean loading. This can lead to erroneous interpretations of changes in water storage, causing significant measurement errors. Therefore, it is crucial to model these external effects with precision and remove them from the GRACE and GRACE-FO data to obtain highly accurate measurements of the water cycle.

The corrections for external factors can be challenging, and their accuracy can affect the quality of the final results. For example, post-seismic deformation correction necessitates a meticulous modeling of the Earth's mantle's viscoelastic response to the surface loading induced by earthquakes, a complex and challenging process. Similarly, ocean loading correction requires precise modeling of the intricate dynamics of oceans and their interaction with the solid Earth,

which can be influenced by various factors such as tides, winds, and currents. Consequently, developing enhanced modeling techniques and assimilating supplementary data sources, like GPS measurements and satellite altimetry, can aid in precisely correcting for external factors and boosting the accuracy of the GRACE and GRACE-FO data.

Separation of different water compartments: Distinguishing the contributions of different water compartments, such as soil moisture, surface water, and groundwater, from GRACE and GRACE-FO data is a complex task. This is because the data only reveals information on total water storage changes. To resolve this, hydrological models and land surface models are utilized to account for the different characteristics of each compartment. Accurate separation of surface water and groundwater is vital for water resource management, drought monitoring, and groundwater depletion assessment. Evapotranspiration (ET) is another crucial component of the terrestrial water balance, but estimating it using GRACE and other satellite data remains uncertain.

Additionally, the estimation of groundwater storage from GRACE data is still a major challenge, as it requires accurate modeling of the complex hydrological processes that control the movement and storage of groundwater in the subsurface. While GRACE can detect changes in total water storage, it cannot directly distinguish between different compartments. Therefore, there is a need to develop new techniques and improve the accuracy of existing methods to better estimate ET and groundwater storage from GRACE data. Further investigations are needed to accurately separate different water compartments using GRACE and GRACE-FO data. Advanced modeling techniques, complemented by hydrological and land surface models, are required to better understand the complex hydrological processes that control water movement and storage. These advancements will assist in water resource management, drought monitoring, and groundwater depletion assessment.

6

Appendices

A. In-situ (gauge) test

- Buishand's Range Test: The Buishand range test is employed for change-point detection of a normal variate. Let X denote a normal random variate, then the adjusted partial sum (S_k) can be defined as:

$$S_k = \sum_{i=1}^{k} (x_i - \overline{x})$$

In random series, the samples are distributed on both sides of the mean. Therefore, a homogeneous series would hold $S_k = 0$. The re-scaled adjusted range (R) is computed as follows to detect the significance of shift:

$$R = \frac{\max(S_k) - \min(S_k)}{\overline{x}}$$

For detecting possible change points, the value of $R/\sqrt{N}$ (N is the total number of observations or samples) is compared with the critical values proposed by Buishand, 1982.

- Von Neumann Ratio Test: The Von Neumann ratio test is a parametric test, first introduced by Von Neumann (1941). The test statistic is defined as the ratio of the mean square successive difference to the variance:

$$N = \frac{\sum_{i=1}^{n-1} (x_i - x_{i-1})^2}{\sum_{i=1}^{n} (x_i - \overline{x})^2}$$

The series of samples are homogeneous according to the test if the expected value of N ($\mathbb{E}[N]$) is equal to 2. An expected value lower than 2 represents in-homogeneity in the series otherwise the test implies a rapid variation in the mean.

- Standard Normal Homogeneity Test (SNHT): The parametric test SNHT was first proposed by Alexandersson (1986) to detect homogeneity of a gauge using the variation of the neighboring gauges. The test statistic (T_k) compares the mean of first k observations with the mean of the remaining $(n-k)$ observations with n data points. By examining the differences between these means, the SNHT helps detect any potential shifts or inhomogeneities in the gauge data. The test statistics are computed as follows:

$$T_k = k Z_1^2 - (n-k) Z_2^2$$

$$Z_1 = \frac{1}{k} \sum_{i=1}^{k} \frac{x_i - \overline{x}}{\sigma_x}$$

$$Z_2 = \frac{1}{n-k} \sum_{i=k+1}^{n} \frac{x_i - \overline{x}}{\sigma_x}$$

where $\overline{x}$ and σ_x are the mean and standard deviation of the series and where Z_1 and Z_2 are the mean values before and after the shift. The Null hypothesis is that all values in the normalized series (Z_1 and Z_2) are normally distributed with a mean value equal to zero and standard deviation equal to one ($Z \sim \mathcal{N}(0,1)$). The null hypothesis can be rejected if the test metric (T_k) attains the maximum value and is above the selected significance level. The significant level depends on the length of the series. In this study, the critical metric values for 10%, 5%, and 2.5% levels were used based on the suggestions by Alexandersson (1986). The neighboring gauges are crucial to the SNHT as they provide reference data for comparison. By considering the differences between means before and after a suspected shift, the test assesses the potential inhomogeneity of the gauge under investigation relative to its neighboring gauges. This approach allows for the identification of significant changes in the gauge data that may be indicative of shifts in measurement practices, changes in the local environment, or other factors affecting the data's homogeneity.

- Pettitt's Test: The Pettitt's test is a non-parametric test developed by Pettit (1979) to identify abrupt changes in climatic variables. Based on this test, the time series has a change point at time t when the distribution function before and after time t is different. To apply the test, one should calculate the test statistics U_t and K as follows:

$$U_t = \sum_{i=1}^{t} \sum_{j=t+1}^{n} \text{sign}(x_t - x_j)$$

$$K = \max |U_t|$$

To check the null hypothesis, the confidence level (ρ) for the sample length (n) can be calculated as:

$$\rho = \exp\left(\frac{-K}{n^2 + n^3}\right)$$

When ρ is smaller than critical values (given by Pettit, 1979), the null hypothesis (the series is independent and has identically distributed random quantities) is rejected.

- The Double Mass Curve test: The double mass curve test proposed by Searcy and Hardison (1960) has been used to check and adjust the inconsistencies in many hydro-climate variables including precipitation. The theory behind the test is simple and is based on a comparison between the graph of accumulation of one quantity versus the other quantity within the same time period. The null hypothesis is that the plot holds a straight line while a break in the in the double mass curve represents a change in the relation between the variables caused by observation methods or data processing (Searcy & Hardison, 1960).

B. Goodness of fit evaluation metrics

There are several common metrics used to evaluate the goodness-of-fit between a model and observations. These metrics assess different aspects of the fit, such as the overall fit, the vari-

ability of the residuals, and the ability of the model to predict new data. In this thesis, we have utilized some of the most common metrics including:

Mean Bias Error (MBE): that quantifies the average difference between the observed values and the predicted values. It provides insight into the systematic tendency of the predictions to overestimate or underestimate the true values. MBE is a signed score, meaning that positive values indicate overestimation while negative values indicate underestimation. The MBE is calculated by summing up the differences between observed and predicted values and dividing by the total number of observations, as shown in the following formula:

$$\text{MBE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i) \tag{6.1}$$

where y_i represents the true value, $\hat{y}_i$ represents the predicted value, and n represents the total number of observations in the dataset.

The MBE provides valuable information about the overall bias or tendency of the predictions. A MBE close to zero indicates that, on average, the predictions are unbiased and have a balanced tendency to overestimate or underestimate. Positive MBE values indicate a systematic overestimation, while negative MBE values indicate a systematic underestimation. Analyzing the MBE helps understand the overall performance and bias of the prediction model.

Root Mean Squared Error (RMSE): which is the square root of the average of the squared differences between the observed values and the predicted values. RMSE is a negatively-oriented score, i.e., lower values show better result. The RMSE neutralizes the effect of negative pairwise discrepancies between estimated and observed values. This means both negative and positive errors contribute to the overall error. Moreover, RMSE gives more importance to big errors by giving higher weights to them, which is desirable when emphasizing extremes. The formula for RMSE is:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{6.2}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and n is the total number of observations in the dataset.

Normalized Root Mean Squared Error (NRMSE): The Normalized Root Mean Squared Error (NRMSE) is a variation of the RMSE that provides a normalized measure of the prediction error. It is particularly useful when comparing models or datasets with different scales or units. The NRMSE is calculated as the ratio of the RMSE to the range of the observed values, which scales the error to a relative measure.

The formula for NRMSE is given by:

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}}}{\max(y) - \min(y)} \tag{6.3}$$

where y_i represents the true value, $\hat{y}_i$ represents the predicted value, n is the total number of observations in the dataset, and $\max(y)$ and $\min(y)$ represent the maximum and minimum observed values, respectively.

Similar to RMSE, the NRMSE is also a negatively-oriented score, where lower values indicate better predictive performance. By normalizing the RMSE with the range of the observed values, the NRMSE allows for meaningful comparisons across different datasets or models. It takes into account the relative magnitude of the errors in relation to the variability of the observed values, providing a standardized measure of prediction accuracy.

Pearson correlation coefficient: also known as the Pearson's r, is a statistical measure developed by Karl Pearson after an idea introduced by Francis Galton in the 1880s. It describes the linear relationship between two variables and is widely used in many fields, including statistics, economics, social sciences, and engineering. The value of r ranges from -1 to 1, where -1 indicates a perfect negative correlation, 0 indicates no correlation, and 1 indicates a perfect positive correlation. A positive correlation means that as one variable increases, the other variable also increases. A negative correlation means that as one variable increases, the other variable decreases. The Pearson correlation coefficient has diverse applications, including trend identification, prediction, and hypothesis testing. Nonetheless, it is imperative to recognize that correlation does not necessarily connote causation, and extraneous variables may potentially affect the association between the two variables under examination.

The formula for the Pearson correlation coefficient is:

$$r = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}_i)}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}}_i)^2}} \tag{6.4}$$

where x_i and y_i are the true and predicted values, respectively, and $\bar{x}$ and $\bar{y}$ are the means of the true and predicted values, respectively. Here n is the total number of observations.

Nash-Sutcliffe Efficiency (NSE): is a widely used statistical metric for evaluating the accuracy of models that predict the behavior of hydrological systems (Nash & Sutcliffe, 1970). NSE ranges from $-\infty$ to 1, where a value of 1 indicates perfect agreement between the observed and simulated values, and a value of 0 or less indicates that the model performs no better than the mean of the observed values. It measures the relative magnitude of the residual variance (i.e., the difference between the observed and simulated values) compared to the variance of the observed values. This makes it a robust metric for evaluating the accuracy of hydrological models, especially when the observed values have large temporal and/or spatial variability. However, NSE has some weaknesses. First, it assumes that the variance of the observed values is the best possible estimate of the total variation in the system, which may not always be the case. Second, it tends to favor models that are biased towards the mean of the observed values, as this can reduce the residual variance and increase the NSE. Third, NSE may not capture important aspects of the system behavior, such as the shape of the hydrograph or the timing of peak flows. The formula for NSE is:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum i = 1^n (y_i - \bar{y})^2} \tag{6.5}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the true values, and n is the total number of time steps.

Kling-Gupta Efficiency (KGE): is a statistical metric used to evaluate the accuracy of hydrological models. It was developed by Gupta et al. (2009) as an improvement over the Nash-Sutcliffe

Efficiency (NSE) by taking into account three key aspects of the model performance: correlation, variability, and bias. In this study, we have calculated modified KGE scores using the estimation from in-situ observations and datasets estimation (Kling et al., 2012). The strength of KGE is that it considers three important aspects of model performance: correlation, variability, and bias. This makes it a more comprehensive metric than NSE, which only considers variability. KGE has been shown to be a useful metric for evaluating the performance of hydrological models in a variety of applications (Moriasi et al., 2007).

However, KGE also has some weaknesses. First, it can be sensitive to outliers in the data, which can affect the correlation component of the metric. Second, the choice of scaling factors can have a significant impact on the KGE value, and there is no consensus on the optimal values for these factors. Finally, like NSE, KGE assumes that the variance and mean of the observed values are the best possible estimates of the total variation and central tendency of the system, respectively.

$$KGE = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2} \tag{6.6}$$

where r is the Pearson correlation coefficient (optimum value = 1), β is bias (optimum value = 1) and γ represents the variability ratio (optimum value = 1) as follow:

$$\beta = \frac{\mu_e}{\mu_o} \quad \text{and} \quad \gamma = \frac{\sigma_e/\mu_e}{\sigma_e/\mu_o} \tag{6.7}$$

where μ and σ are the mean and standard deviations of the time series, respectively. The subscripts e and o indicate estimation and observation, respectively.

Normalized Root Mean Square Error (NRMSE): the square root of the second moment of the differences between the estimation and the reference data. It is normalized by its standard deviation:

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}}{\sigma(y)} \tag{6.8}$$

where σ is the standard deviation of the observed values y. The lower the NRMSE, the better the estimation. The optimum value of NRMSE is zero (RMSE = 0) and it occurs in case of the perfect estimation with respect to the reference data.

C. Water balance fluxes

Precipitation

Precipitation, the process of water droplets falling from the atmosphere to the Earth's surface, is a crucial component of the Earth's hydrological cycle. Precipitation datasets play a critical role in various fields, including hydrology, agriculture, climate science, and water resource management. There are several types of precipitation that can occur and is determined by the temperature and moisture content of the atmosphere. These types include:

- Rain: liquid precipitation that falls from the sky and reaches the ground.

- Snow: precipitation that falls as ice crystals and reaches the ground as snowflakes.

- Sleet: a mixture of rain and snow or ice pellets that falls to the ground.

- Freezing rain: liquid rain that freezes upon contact with surfaces that are at or below freezing temperatures.

- Hail: precipitation that falls as balls or irregular lumps of ice.

- Graupel: soft, rounded pellets of snow that form when supercooled water droplets freeze onto falling snowflakes.

Three primary categories of precipitation datasets exist: gauge-based datasets, satellite-based datasets, and reanalysis datasets. Gauge-based precipitation datasets utilize ground-based rain gauges to measure the amount of precipitation at specific locations. These datasets provide high accuracy and resolution but are limited in spatial coverage due to the limited number of gauges. Satellite-based precipitation datasets, on the other hand, use remote sensing instruments to estimate precipitation over a larger spatial extent, albeit at lower resolution. These datasets employ various techniques, such as passive microwave and infrared sensors, to estimate precipitation rates from satellite observations. Reanalysis datasets are another type of precipitation dataset that combines observations from various sources, including gauge and satellite-based datasets, with numerical models to provide consistent and continuous estimates of precipitation over a long time period. Reanalysis datasets are valuable for studying long-term climate variability and changes, as well as for assessing the reliability of other precipitation datasets.

Despite the advances in precipitation datasets, there is no single dataset that performs well everywhere due to the complex nature of precipitation and the diverse topographical and meteorological conditions across the globe. Several studies have shown that different datasets can have significant variations in their precipitation estimates, particularly in regions with complex terrain, convective rainfall, or mixed precipitation types (e.g., AghaKouchak et al., 2011; Behrangi et al., 2011; Prakash et al., 2016; Saemian et al., 2021; Sun et al., 2018). For instance, gauge-based datasets may suffer from spatial and temporal heterogeneity due to the limited coverage and density of rain gauges, while satellite-based datasets may encounter difficulties in accurately detecting light rainfall or distinguishing between different precipitation types (Derin & Yilmaz, 2014; Kidd et al., 2017). Furthermore, the accuracy of precipitation datasets can be affected by the type and quality of input data, calibration methods, and interpolation techniques (Ruelland et al., 2008; Sorooshian et al., 2011; Sun et al., 2018).

In this thesis, we have used precipitation for various applications. Table C.1 lists the selected precipitation datasets and their main properties.

Table C.1: Summary of global precipitation data sets. Abbreviations in the data source(s) defined as: G, gauge; S, satellite; and R, reanalysis.

Dataset	Data source(s)	Resolution		Coverage		Reference
		Spatial	Temporal	Spatial	Temporal	
CRUv4.04	G	0.5° × 0.5°	1 mo	Global land	1901–2019	Harris et al., 2020
GPCCv2020	G	0.5° × 0.5°	1 mo, 1d	Global land	1982–2016	Schneider et al., 2020
PREC/L	G	0.5° × 0.5°	1 mo	Global land	1948–present	Chen et al., 2002
UDELv5.01	G	0.5° × 0.5°	1 mo	Global land	1900–2017	Willmott and Matsuura, 1995
CPC	G	0.5° × 0.5°	1 d	Global land	1979–present	Chen et al., 2008; Xie et al., 2007
GPCPv2.3	G, S	2.5° × 2.5°	1 mo	Global	1979–present	**adler2018GPCPv2.3**; Adler et al., 2003
CMAP	G, S	2.5° × 2.5° 0.5° × 0.5°	1 mo	Global	1979–present	Xie et al., 2003
NCEP 1	R	2.5° × 2.5°	6 h/1 d/ 1 mo	Global	1948–present	Kalnay et al., 1996
NCEP 2	R	1.875° × 1.875°	6 h/1 d/ 1 mo	Global	1979–present	Kanamitsu et al., 2002
ERA-Interim	R	79 km	6 h/ 1 mo	Global	1979–2019	Dee et al., 2011
ERA5	R	31 km	1 h/ 1 mo	Global	1979–present	Hersbach et al., 2020
20CRv3	R	2° × 2°	6 h/1 d/ 1 mo	Global	1836–2015	Compo et al., 2011
PGFv3	G, R	0.25° × 0.25°	3 h/1 d	Global	1948–2016	Sheffield et al., 2006
WFDEI-CRU	G, R	0.5° × 0.5°	3 h	Global	1979–2016	Weedon et al., 2014
WFDEI-GPCC	G, R	0.5° × 0.5°	3 h	Global	1979–2016	Weedon et al., 2014
AgCFSR	G, R	0.25° × 0.25°	1 d	Global	1980–2010	Ruane et al., 2015
AgMERRA	G, S, R	0.25° × 0.25°	1 d	Global	1980–2010	Ruane et al., 2015
MERRA Land	G, S, R	0.5° × 0.67°	1 h/1 d/ 1 mo	Global land	1980–2016	Reichle et al., 2011
MSWEP v2.2	G, S, R	0.1° × 0.1°	3 h/1 d	Global	1979–2017	Beck et al., 2019

Evapotranspiration

Evapotranspiration is the process of water transfer from the earth's surface to the atmosphere through the combination of evaporation and transpiration. Evaporation occurs when water changes from its liquid state to its vapor state, while transpiration is the process by which water vapor is released into the atmosphere through plant leaves. Evapotranspiration is a critical component of the water cycle, affecting both the water supply and the energy balance of the planet. It plays a crucial role in agriculture, water resources management, hydrology, and climate science. Accurate estimation of evapotranspiration is therefore necessary for various applications such as irrigation scheduling, crop growth monitoring, and drought monitoring.

Several global datasets provide estimates of evapotranspiration based on either mathematical models or satellite data. ET can be mapped based on ground-based measurements, such as lysimeters or eddy covariance systems. However, such a estimation can be affected by several factors that make them challenging and costly to obtain including the limited spatial coverage and the errors result from interpolation methods. Model-based and remote sensing-based datasets are the two most commonly used types of ET datasets, particularly for global applications. Table C.1 lists the selected evapotranspiration datasets and their main properties.

Runoff

Runoff refers to the portion of precipitation that flows over land surfaces and eventually reaches streams, rivers, lakes, and oceans. Accurate estimation of runoff is essential for various applications such as water resource management, flood forecasting, and climate modeling. There are several global datasets that provide estimates of runoff, which can be classified into two main types: (1) model-based and (2) remote sensing-based. Model-based datasets use mathematical models that simulate the physical processes of runoff, such as the Variable Infiltration Capacity (VIC) model and the Soil and Water Assessment Tool (SWAT) model. Remote sensing-based datasets use satellite data to estimate runoff based on various parameters, such as precipitation, land cover, and topography. In this thesis, we have used runoff for various applications. Table C.3 lists the selected runoff datasets used in this thesis.

Table C.2: Summary of the evapotranspiration datasets.

Data set	Resolution		Coverage		Reference
	Spatial	Temporal	Spatial	Temporal	
GLEAM3.6a	0.25° × 0.25°	1 mo	Global	1980–2021	Martens et al., 2017; Miralles et al., 2011
GLEAM3.6b	0.25° × 0.25°	1 mo	Global	2003–2021	Martens et al., 2017; Miralles et al., 2011
SSEBop	0.05° × 0.05°	1 d	50°S–50°N	1981–present	Funk et al., 2015
JRA-55	60 km	1 mo	Global	1958–present	Ebita et al., 2011; Kobayashi et al., 2015
WECANN	1° × 1°	1 mo	60°S–90°N	2007–2015	Alemohammad et al., 2017
PMLv2	0.05° × 0.05°	1d	60°S–90°N	2002–2019	Zhang et al., 2016
FLDAS-NOAH	0.1° × 0.1°	1 mo	60°S–90°N	1982–present	McNally et al., 2018
ERA5	31 km	1 h/ 1 mo	Global	1979–present	Hersbach et al., 2020
MERRA-Land	0.5° × 0.5°	1 mo	Global	1980–2016	Reichle et al., 2011
GLDAS2.0-CLSM	0.25° × 0.25°	1 d	Global	1948–2014	Li et al., 2019; Rodell et al., 2004
GLDAS2.1-CLSM-DA1	0.25° × 0.25°	1 d	Global	2003–present	Rodell et al., 2004
GLDAS2.0-NOAH	0.25° × 0.25°	1 mo	Global	1948–2014	Rodell et al., 2004
GLDAS2.1-NOAH	0.25° × 0.25°	1 mo	Global	2000–present	Rodell et al., 2004
GLDAS2.0-VIC	1° × 1°	1 mo	Global	1948–2014	Rodell et al., 2004
GLDAS2.1-VIC	1° × 1°	1 mo	Global	2000–present	Rodell et al., 2004
FluxCom	0.5° × 0.5°	1 mo	60°S–90°N	2001–2013	Jung et al., 2019
P-LSH	0.083° × 0.083°	1 mo	Global	1982–2013	Zhang et al., 2009; Zhang et al., 2010
TerraClimate	0.0417° × 0.0417°	1 mo	Global	1958–2015	Abatzoglou et al., 2018

Table C.3: Summary of global runoff data sets.

Dataset	Spatial Resolution	Time period	Reference
ERA-Interim	79 km	1979–2019	Dee et al., 2011
ERA5	31 km	1979–present	Hersbach et al., 2020
ERA-land	0.1° × 0.1°	1950–present	Muñoz-Sabater et al., 2021
GLDAS2.0-CLSM	0.25° × 0.25°	1948–2014	Li et al., 2019; Rodell et al., 2004
GLDAS2.1-CLSM-DA1	0.25° × 0.25°	2003–present	Rodell et al., 2004
GLDAS2.0-NOAH	0.25° × 0.25°	1948–2014	Rodell et al., 2004
GLDAS2.1-NOAH	0.25° × 0.25°	2000–present	Rodell et al., 2004
GLDAS2.0-VIC	1° × 1°	1948–2014	Rodell et al., 2004
GLDAS2.1-VIC	1° × 1°	2000–present	Rodell et al., 2004
GDFC	0.25° × 0.25°	1950–2016	He et al., 2020
HBV-SIMREG	0.5° × 0.5°	1979–2012	Lindström et al., 1997
HTESSEL	0.25° × 0.25°	1980–2014	Balsamo et al., 2015
LISFLOOD	0.25° × 0.25°	1980–2014	Van Der Knijff et al., 2010
ORCHIDEE	0.25° × 0.25°	1980–2014	Polcher et al., 2011
SURFEX-TRIP	0.5° × 0.5°	1979–2012	Decharme et al., 2013
SWBM	0.5° × 0.5°	1979–2012	Koster and Mahanama, 2012; Orth and Seneviratne, 2013
W3RA	0.5° × 0.5°	1979–2012	Van Dijk, 2010
GloFAS	0.1° × 0.1°	1980–2018	Alfieri et al., 2020
G-Run	0.5° × 0.5°	1980–2018	Ghiggi et al., 2021

Bibliography

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data*, *5*(1), 1–12. https://doi.org/10.1038/sdata.2017.191

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., et al. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:tvgpcp>2.0.co;2

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research: Atmospheres*, *116*(D2). https://doi.org/10.1029/2010JD014741

Alemohammad, S. H., Fang, B., Konings, A. G., Aires, F., Green, J. K., Kolassa, J., Miralles, D., Prigent, C., & Gentine, P. (2017). Water, Energy, and Carbon with Artificial Neural Networks (WECANN): a statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence. *Biogeosciences*, *14*(18), 4101–4124. https://doi.org/10.5194/bg-14-4101-2017

Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of climatology*, *6*(6), 661–675. https://doi.org/10.1002/joc.3370060607

Alfieri, L., Lorini, V., Hirpa, F. A., Harrigan, S., Zsoter, E., Prudhomme, C., & Salamon, P. (2020). A global streamflow reanalysis for 1980–2018. *Journal of Hydrology X, 6*, 100049. https://doi.org/10.1016/j.hydroa.2019.100049

Balsamo, G., Albergel, C., Beljaars, A., Boussetta, S., Brun, E., Cloke, H., Dee, D., Dutra, E., Muñoz-Sabater, J., Pappenberger, F., et al. (2015). ERA-Interim/Land: a global land surface reanalysis data set. *Hydrology and Earth System Sciences*, *19*(1), 389–407. https://doi.org/10.5194/hess-19-389-2015

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., van Dijk, A. I., McVicar, T. R., & Adler, R. F. (2019). MSWEP V2 global 3-hourly 0.1 precipitation: methodology and quantitative assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/bams-d-17-0138.1

Behrangi, A., Khakbaz, B., Jaw, T. C., AghaKouchak, A., Hsu, K., & Sorooshian, S. (2011). Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology*, *397*(3-4), 225–237. https://doi.org/10.1016/j.jhydrol.2010.11.043

Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, *58*(1-2), 11–27. https://doi.org/10.1016/0022-1694(82)90066-X

Chen, M., Shi, W., Xie, P., Silva, V. B., Kousky, V. E., Wayne Higgins, R., & Janowiak, J. E. (2008). Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research: Atmospheres*, *113*(D4). https://doi.org/10.1029/2007jd009132

Chen, M., Xie, P., Janowiak, J. E., & Arkin, P. A. (2002). Global land precipitation: A 50-yr monthly analysis based on gauge observations. *Journal of Hydrometeorology*, *3*(3), 249–266. https://doi.org/10.1175/1525-7541(2002)003<0249:glpaym>2.0.co;2

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., et al. (2011). The twentieth century reanalysis

project. *Quarterly Journal of the Royal Meteorological Society, 137*(654), 1–28. https://doi.org/10.1002/qj.776

Decharme, B., Martin, E., & Faroux, S. (2013). Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *Journal of Geophysical Research: Atmospheres, 118*(14), 7819–7834. https://doi.org/10.1002/jgrd.50631

Dee, D. P., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society, 137*(656), 553–597. https://doi.org/10.1002/qj.722

Derin, Y., & Yilmaz, K. K. (2014). Evaluation of multiple satellite-based precipitation products over complex topography. *Journal of Hydrometeorology, 15*(4), 1498–1516. https://doi.org/10.1175/JHM-D-13-0191.1

Ebita, A., Kobayashi, S., Ota, Y., Moriya, M., Kumabe, R., Onogi, K., Harada, Y., Yasui, S., Miyaoka, K., Takahashi, K., et al. (2011). The Japanese 55-year reanalysis "JRA-55": an interim report. *Sola, 7*, 149–152. https://doi.org/10.2151/sola.2011-038

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data, 2*, 150066. https://doi.org/10.1038/sdata.2015.66

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2021). G-run ensemble: A multi-forcing observation-based global runoff reanalysis. *Water Resources Research, 57*(5), e2020WR028787. https://doi.org/10.1029/2020WR028787

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology, 377*(1-2), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Harris, I., Osborn, T. J., Jones, P., & Lister, D. (2020). Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Scientific data, 7*(1), 1–18. https://doi.org/10.1038/s41597-020-0453-3

He, X., Pan, M., Wei, Z., Wood, E. F., & Sheffield, J. (2020). A global drought and flood catalogue from 1950 to 2016. *Bulletin of the American Meteorological Society, 101*(5), E508–E535. https://doi.org/10.1175/BAMS-D-18-0269.1

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society, 146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Jung, M., Koirala, S., Weber, U., Ichii, K., Gans, F., Camps-Valls, G., Papale, D., Schwalm, C., Tramontana, G., & Reichstein, M. (2019). The FLUXCOM ensemble of global land-atmosphere energy fluxes. *Scientific data, 6*(1), 1–14. https://doi.org/10.1038/s41597-019-0076-8

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American meteorological Society, 77*(3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang, S.-K., Hnilo, J., Fiorino, M., & Potter, G. (2002). NCEP–DOE AMPI-II reanalysis (r-2). *Bulletin of the American Meteorological Society, 83*(11), 1631–1644. https://doi.org/10.1175/BAMS-83-11-1631

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society, 98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology, 424,* 264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al. (2015). The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II, 93*(1), 5–48. https://doi.org/10.2151/jmsj.2015-001

Koster, R. D., & Mahanama, S. P. (2012). Land surface controls on hydroclimatic means and variability. *Journal of Hydrometeorology, 13*(5), 1604–1620. https://doi.org/10.1175/JHM-D-12-050.1

Li, B., Rodell, M., Sheffield, J., Wood, E., & Sutanudjaja, E. (2019). Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models. *Scientific reports, 9*(1), 1–13. https://doi.org/10.1038/s41598-019-47219-z

Lindström, G., Johansson, B., Persson, M., Gardelin, M., & Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of hydrology, 201*(1-4), 272–288. https://doi.org/10.1016/S0022-1694(97)00041-3

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A., Fernández-Prieto, D., Beck, H. E., Dorigo, W. A., & Verhoest, N. E. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development, 10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017,2017

McNally, A., et al. (2018). FLDAS noah land surface model L4 global monthly 0.1× 0.1 degree (MERRA-2 and CHIRPS). *Atmos. Compos. Water Energy Cycles Clim. Var.*

Miralles, D. G., Holmes, T., De Jeu, R., Gash, J., Meesters, A., & Dolman, A. (2011). Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences, 15*(2), 453–469. https://doi.org/10.5194/hess-15-453-2011,2011

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE, 50*(3), 885–900.

Muñoz-Sabater, J., Dutra, E., Agustié-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., Boussetta, S., Choulga, M., Harrigan, S., Hersbach, H., et al. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data, 13*(9), 4349–4383. https://doi.org/10.5194/essd-13-4349-2021

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of hydrology, 10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Orth, R., & Seneviratne, S. I. (2013). Predictability of soil moisture and streamflow on subseasonal timescales: A case study. *Journal of Geophysical Research: Atmospheres, 118*(19), 10–963. https://doi.org/10.1002/jgrd.50846,2013

Pettit, A. (1979). Anon-parametric approach to the change-point detection. *Appl. Stat, 28,* 126–135.

Polcher, J., Bertrand, N., Biemans, H., Clark, D. B., Floerke, M., Gedney, N., Gerten, D., Stacke, T., van Vliet, M., & Voss, F. (2011). Improvements in hydrological processes in general hydrological models and land surface models within WATCH.

Prakash, S., Mitra, A. K., Pai, D., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources, 88,* 1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Reichle, R. H., Koster, R. D., De Lannoy, G. J., Forman, B. A., Liu, Q., Mahanama, S. P., & Touré, A. (2011). Assessment and enhancement of MERRA land surface hydrology estimates. *Journal of climate, 24*(24), 6322–6338. https://doi.org/10.1175/jcli-d-10-05033.1

Rodell, M., Houser, P., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological society, 85*(3), 381–394. https://doi.org/10.1175/BAMS-85-3-381

Ruane, A. C., Goldberg, R., & Chryssanthacopoulos, J. (2015). Climate forcing datasets for agricultural modeling: Merged products for gap-filling and historical climate series estimation. *Agricultural and Forest Meteorology, 200*, 233–248. https://doi.org/10.1016/j.agrformet.2014.09.016

Ruelland, D., Ardoin-Bardin, S., Billen, G., & Servat, E. (2008). Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa. *Journal of Hydrology, 361*(1-2), 96–117. https://doi.org/10.1016/j.jhydrol.2008.07.049

Saemian, P., Hosseini-Moghari, S.-M., Fatehi, I., Shoarinezhad, V., Modiri, E., Tourian, M. J., Tang, Q., Nowak, W., Bárdossy, A., & Sneeuw, N. (2021). Comprehensive evaluation of precipitation datasets over Iran. *Journal of Hydrology*, 127054. https://doi.org/10.1016/j.jhydrol.2021.127054

Schneider, U., Becker, A., Finger, P., Rustemeier, E., & Ziese, M. (2020). GPCC full data monthly product version 2020 at 0.25°: monthly land-surface precipitation from rain-gauges built on GTS-based and historical data. *Global Precipitation Climatology Centre at Deutscher Wetterdienst: Offenbach, Germany*. https://doi.org/10.5676/DWD_GPCC/FD_M_V2020_025

Searcy, J. K., & Hardison, C. H. (1960). *Double-mass curves*. US Government Printing Office.

Sheffield, J., Goteti, G., & Wood, E. F. (2006). Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of climate, 19*(13), 3088–3111. https://doi.org/10.1175/jcli3790.1

Sorooshian, S., AghaKouchak, A., Arkin, P, Eylander, J., Foufoula-Georgiou, E., Harmon, R., Hendrickx, J. M., Imam, B., Kuligowski, R., Skahill, B., et al. (2011). Advanced concepts on remote sensing of precipitation at multiple scales. *Bulletin of the American Meteorological Society, 92*(10), 1353–1357. http://www.jstor.org/stable/26218594

Sun, Q., Miao, C., Duan, Q., Ashouri, H., Sorooshian, S., & Hsu, K.-L. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Reviews of Geophysics, 56*(1), 79–107. https://doi.org/10.1002/2017RG000574

Van Der Knijff, J., Younis, J., & De Roo, A. (2010). LISFLOOD: a GIS-based distributed model for river basin scale water balance and flood simulation. *International Journal of Geographical Information Science, 24*(2), 189–212. https://doi.org/10.1080/13658810802549154

Van Dijk, A. (2010). The Australian Water Resources Assessment System Technical Report 3. Landscape Model (version 0.5) Technical Description. https://doi.org/10.4225/08/5852dd9bb578c

Von Neumann, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *The Annals of Mathematical Statistics, 12*(4), 367–395. https://www.jstor.org/stable/2235951

Weedon, G. P, Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., & Viterbo, P. (2014). The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research, 50*(9), 7505–7514. https://doi.org/10.1002/2014wr015638

Willmott, C. J., & Matsuura, K. (1995). Smart interpolation of annually averaged air temperature in the United States. *Journal of Applied Meteorology, 34*(12), 2577–2586. https://doi.org/10.1175/1520-0450(1995)034<2577:sioaaa>2.0.co;2

6

Xie, P., Chen, M., Yang, S., Yatagai, A., Hayasaka, T., Fukushima, Y., & Liu, C. (2007). A gauge-based analysis of daily precipitation over East Asia. *Journal of Hydrometeorology, 8*(3), 607–626. https://doi.org/10.1175/JHM583.1

Xie, P., Janowiak, J. E., Arkin, P. A., Adler, R., Gruber, A., Ferraro, R., Huffman, G. J., & Curtis, S. (2003). GPCP pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *Journal of Climate, 16*(13), 2197–2214. https://doi.org/10.1175/2769.1

Zhang, K., Kimball, J. S., Mu, Q., Jones, L. A., Goetz, S. J., & Running, S. W. (2009). Satellite based analysis of northern ET trends and associated changes in the regional water balance from 1983 to 2005. *Journal of Hydrology, 379*(1-2), 92–110. https://doi.org/10.1016/j.jhydrol.2009.09.047

Zhang, K., Kimball, J. S., Nemani, R. R., & Running, S. W. (2010). A continuous satellite-derived global record of land surface evapotranspiration from 1983 to 2006. *Water Resources Research, 46*(9). https://doi.org/10.1029/2009WR008800

Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y. Y., et al. (2016). Multi-decadal trends in global terrestrial evapotranspiration and its components. *Scientific reports, 6*(1), 1–12. https://doi.org/10.1038/srep19124

6