Bayesian Integration of Satellite Geodetic Data with Models to Separate Land Hydrology and Surface Deformation Signals

Nooshin Mehrnegar

School of Earth and Environmental Science
Cardiff University

This dissertation is submitted for the degree of
Doctor of Philosophy

December 2021

Acknowledgements

First and foremost, I am extremely grateful to my supervisor, Prof. Ehsan Forootan, for his invaluable advice, continuous support, and patience during my PhD study. I am very grateful for all the fruitful dialogues, friendly discussions, and all the valuable insights and suggestions that without doubt help me to improve the quality of my work. His immense knowledge and plentiful experience have encouraged me in all the time of my academic research and daily life. I'm proud of, and grateful for, my time working with Ehsan. I would also like to express my sincere gratitude to my second supervision, Prof. Owen Jones, who has been my great mathematics teacher, and his understanding of mathematical formulations and comments on the Bayesian formulations considerably improved the quality of my work. I'm extremely grateful to Dr. Michael Singer, my co-supervisor, whose insightful suggestions, guidance, unwavering support and encouragement has been invaluable throughout this study.

My deep gratitude also goes to Dr. Mark Cuthbert (Cardiff University) and Prof. Philip Moore (Newcastle University) for being a referee of this thesis, and for providing valuable discussions and kind supports. Their constructive comments helped me to improve the quality of dissertation. Furthermore, my thanks go to Prof. Thomas Blenkinsop, who agreed to be the chairperson of my examination board.

I would also like to express the gratitude to A.Prof. Dr.-Ing Maike Schumacher (Aalborg University), Dr. Thomas.Jagdhuber (Augsburg University), and Prof. Paul Bates (Bristol University) for sharing insightful suggestions and detailed feedback, and for their encouraging words and thoughtful, which is hard to forget throughout my life.

I must also thank all the members in the School of Earth and Ocean Science in Cardiff University, for their kind help and support that have made my study and life in the UK a wonderful time. I very much appreciate Mrs. Joanne Poynter, our postgraduate administrator, who always did her best to answer my questions and helped me out with a lot of paper works.

I would like to acknowledge the Vice-Chancellor's International Scholarship for Research Excellence. This work would not have been possible without their financial support.

I would like to thank my friends and my lab mates for all the scientific and non-work related discussions, and the wonderful time in a great working environment, naming a few: Andrés Quichimbo, Zhibin Lei, Sophie Cox, Christopher Tully, Harry Lea, Maria Magdalena Warter, Romy Sabathier, Manny Zarate, Jaz millar, Aidan star, Eveanjelene Snee, Maximiliaan Jansen, William Smith, Niall Groome and Nurudeen Oshinlaja.

During the last year of my PhD study, I had the unique chance to spend three months as a guest researcher at Aalborg University, Denmark. I am very grateful for the kind supports of Ehsan during my research stay at his Geodesy group, at Aalborg University. Thanks to A.Prof. Peter Cederholm and A.Prof Finn Kjær Christensen for their kind supports and friendly discussions during my stay in Aalborg.

Finally, my deep and sincere gratitude goes to my family for their continuous and unparalleled love, help and support. I cannot begin to express my thanks to my husband, Majid, whose supports and encouragements during the challenges of graduate school and life is constant as tides. I am truly thankful for having him in my life. I am grateful to my sister, Mahnoosh, for always being there for me as a friend. I am forever indebted to my parents, Nezhat and Mohammad, for giving me the opportunities and experiences that have made me who I am. They selflessly encouraged me to explore new directions in life and seek my own destiny. This journey would not have been possible if not for them. I hope they would accept this work being dedicated to them.

Abstract

Reliable quantification of water mass changes (or redistribution) within the different compartments of the water cycle is important for understanding processes and feedback loops within the Earth's climate system. This information is also essential in geodesy because it changes the Earth's orientation (importance for defining reference frames) and the Earth's gravity field, which is the physical shape of the Earth and is used for defining reference datum. The Gravity Recovery and Climate Experiment (GRACE) and its Follow-On mission (GRACE-FO) provide time-variable Earth's gravity fields that contain signals related to different processes such as non-steric sea level changes, Terrestrial Water Storage Changes (TWSC), ice sheet melting, and Post Glacial Rebound (PGR). Although GRACE(-FO) data represent an accurate superposition of these anomalies, separating this integrated signal into its contributors is desirable for many hydro-climatic and geophysical applications. In this thesis, three novel Bayesian data-model fusion frameworks are developed to separate land hydrology (surface and sub-surface) and surface deformation (due to PGR) from GRACE(-FO) data. The three main frameworks of this thesis include: 1- the Dynamic Model Data Averaging (DMDA), that is formulated to merge multi-model data with GRACE(-FO) data; 2- Markov Chain Monte Carlo-Data Assimilation (MCMC-DA), as an extension of DMDA, to recursively estimate components of the TWSC, while accounting for temporal dependencies between the storage compartments; and 3- the Constrained Bayesian-Data Assimilation (ConBay-DA) to use multi-sensor data for GRACE(-FO) signal separation. DMDA is used to compare several global hydrological models and merge them with GRACE data. The groundwater and soil water storage changes are extracted within the Conterminous United States (CONUS) by implementing the MCMC-DA approach. ConBay-DA is applied, based on the hierarchical MCMC optimization, to use GRACE data and the surface uplift rates from the Global Navigation Satellite System (GNSS) stations and separate hydrological and GIA deformation components over the Great Lakes (GL) area in North America.

Table of contents

Acronyms

BMA	**B**ayesian **M**odel **A**veraging
CSR	**C**entre for **S**pace **R**esearch
C/DA	**C**alibration/ **D**ata **A**ssimilation
CM	**C**enter of **M**ass
CONUS	**C**ontinental **U**nited **S**tates
CLSM	**C**atchment **L**and **S**urface **M**odel
DA	**D**ata **A**ssimilation
DUT	**D**elft **U**niversity of **T**echnology
DMDA	**D**ynamic **M**odel **D**ata **A**veraging
ENSO	**E**l **N**iño **S**outhern **O**scillation
EnKF	**E**nsemble **K**alman **F**iltering
EO	**E**arth **O**bervation
EOP	**E**arth **O**rientation **P**arameter
ESA-CCI	**E**uropean **S**pace **A**gency-**C**limate **C**hange **I**nitiative
EWHs	**E**quivalent **W**ater **H**eights
GFZ	**G**eo**F**orschungs**Z**entrum
GHM	**G**lobal **H**ydroloical **M**odel
GIA	**G**lacial **I**sostatic **A**djustment
GL	**G**reat **L**akes
GLDAD	**G**lobal **L**land **D**lata **A**lssimilation **S**lystem
GNSS	**G**lobal **N**avigational **S**atellite **S**ystem
GPS	**G**lobal **P**ositioning **S**ystem
GRACE	**G**ravity **R**ecovery and **C**limate **E**xperiment
GRACE-FO	**G**ravity **R**ecovery and **C**limate **E**xperiment-**F**ollow-**O**n
ICA	**I**ndependent **C**omponent **A**nalysis
ICESat	**Ice**, **C**loud, and **E**levation **Sat**ellite
JPL	**J**et **P**ropulsion **L**aboratory
KBR	**K-B** (Microwave) **R**anging
LSM	**L**and **S**urface **M**odel
MCMC	**M**arkov **C**hain **M**onte **C**arlo
MCMC-DA	**M**arkov **C**hain **M**onte **C**arlo-**D**ata **A**ssimilation
MEO	**M**edium **E**arth **O**rbit
NASA	**N**ational **A**eronautics and **S**pace **A**dministration
NGL	**N**evada **G**eodetic **L**aboratory
NOAH-MP LSM	**NOAH** **M**ulti **P**arameterization **L**and **S**urface **M**odel
PC	**P**rincipal **C**omponent
PCA	**P**rincipal **C**omponent **A**nalysis
PDF	**P**robability **D**istribution **F**anction
PF	**P**article **F**ilter
PGR	**P**ost **G**lacial **R**ebound
PHDI	**P**almer **H**ydrological **D**rought **I**ndex
PS	**P**article **S**moother
RMSD	**R**oot **M**ean **S**quare of **D**ifferences
SHCs	**S**pherical **H**armonic **C**oefficients
SLR	**S**atellite **L**aser **R**anging
SNR	**S**ignal to **N**oise **R**atio
StD	**St**andard **D**eviation
TRMM	**T**ropical **R**ainfall **M**easuring **M**ission
TWSC	**T**errestrial **W**ater **S**torage **C**hanges
Total-WSC	**Total** **W**ater **S**torage **C**hanges
USA	**U**nited **S**tates
USGS	**U**nited **S**tates **G**eological **S**urvey

List of tables

List of figures

Chapter 1

Introduction

1.1 Background: Earth System Dynamics and the Global Water Cycle

The global water cycle describes the redistribution of water within the Earth system, including the hydrosphere, biosphere, atmosphere, and oceans. Water from the atmosphere reaches the continents and oceans in the form of precipitation. Once the water reaches the ground, it might be partly stored on land in the form of snow and ice, stored as surface water in lakes and wetlands, or it might infiltrate into the ground, where it is stored as soil water or groundwater for different durations of time. The continental storage of water is reduced by evapotranspiration, i.e., the sum of evaporation from soil, surface water, vegetation, and/or by river discharge and groundwater discharge, which transport water, e.g., to seas or oceans. Due to solar radiation, water in seas and oceans also evaporates and is transported back into the atmosphere, where it becomes available again in the form of precipitation as it cools and condenses, thereby completing the water cycle (see Fig. 1.1). The water cycle is also linked to energy exchanges among the atmosphere, ocean, and land, which determine the Earth's climate, and cause much of its natural climate variability. Reliable quantification of the hydrological cycle, its key fluxes and stores, and its spatiotemporal variability is required to understand numerous processes and feedback loops within the Earth's climate system. Such quantification is also crucial for many applications related to hydrometeorology, water resources, hazard assessment, and land-atmosphere interactions, including flood and drought monitoring and prediction (*Forootan et al.*, 2017, 2019; *Houborg et al.*, 2012; *Li et al.*, 2019; *Long et al.*, 2013; *Slater et al.*, 2015), assessing water resources sustainability (*Castellazzi et al.*, 2016; *Forootan et al.*, 2014c; *Scanlon et al.*, 2012a), and identifying ecohydrological links between climate and vegetation (*Singer et al.*, 2014). From the geodetic point of view, an answer to the question of how mass is distributed and redistributed within the Earth system, provides precise information to estimate temporal changes in the Earth's orientation

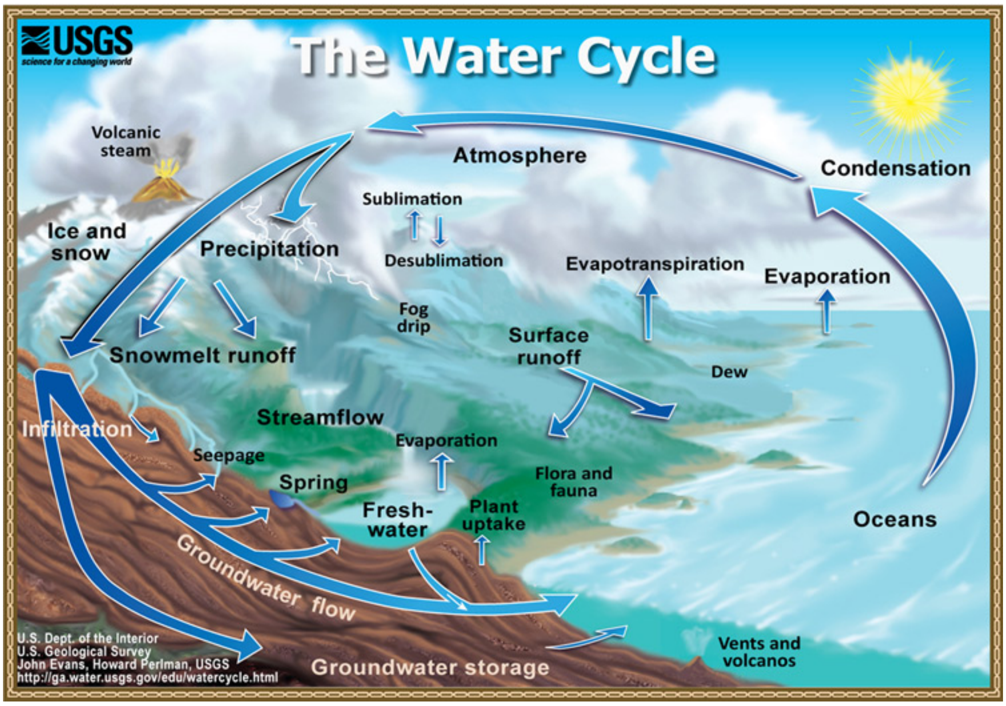

Fig. 1.1 Schematic overview of the global water cycle. This figure is taken from the official website of the United States Geological Survey (USGS; http://water.usgs.gov/edu/watercycle.html).

(importance for defining reference frames) and the Earth's gravity field, which is the physical shape of the Earth and is an important basis for geodetic applications.

Monitoring the global water cycle is difficult due to the complexity of hydrological processes and the large spatial scale, and the fact that it includes mass fluxes between the atmosphere, land, and oceans as well as between various states of water (solid, liquid, and vapour). As a result, different approaches and observational instruments have been developed and implemented to improve the understanding and quantifications of the hydrological processes within the Earth system. Hydrological models and hydrometeorological (drought/flood) monitoring systems are the primary tools to monitor the global terrestrial water cycle (*Sood and Smakhtin*, 2015). Hydrological models are typically classified as conceptual models if empirical equations and parameters are used to represent the water dynamics or as physics-based if the model equations are based on physical principles, e.g., land surface models (see, e.g., *Jaiswal et al.*, 2020). Even though the models aim for an adequate representation of the real world, uncertainties exist due to insufficient model realism, imperfect climate input data, inadequate empirical model parameters, and unrepresented feedbacks between model components.

Post Glacial Rebound (PGR), also known as isostatic rebound or crustal rebound, is referred to as surface deformation and is caused by the removal of the huge weight of ice sheets during the last glacial period. PGR and isostatic depression are phases of Glacial Isostatic Adjustment (GIA) that represent the Earth's viscoelastic response to the loading of glaciation and deglaciation of ice age cycles within the continental regions that were ice-covered during the last glacial maximum about 20,000 years ago. The solid Earth response to changes in ice loading has important implications for many geophysical processes. GIA affects the Earth's shape, gravity field, and the axis of rotation via redistribution of mantle material (*Dickey et al.*, 2002). All these changes, in turn, affect the global sea level (e.g., *Lambeck and Chappell*, 2001; *Nakada and Lambeck*, 1987). A reliable estimation of GIA is necessary to understand contemporary changes in sea level (*Peltier*, 1998). In general, geophysical forward models are the primary tools to study the GIA signal. However, large uncertainties in the present-day global GIA models exist due to the insufficient input data, e.g., uncertainty of ice loading history and the viscosity of the upper and lower mantle. This is shown by *Guo et al.* (2012), who found considerable regional differences between 14 forward model solutions, disagreeing even on the sign of vertical land motion in some areas.

In recent decades, by launching various geodetic and Earth Observation (EO) satellite missions and the availability of in-situ monitoring networks, a wealth of data has been collecting from different components of the water cycle. Traditionally, monitoring soil water and groundwater storage changes mainly relied on in-situ meteorological measurements and piezometric observations. Gauge-based measurments have also been used to estimate, e.g., precipitation, evapotranspiration, and runoff. In-situ water monitoring networks are particularly useful to understand the local hydrological characteristics of a region. For instance, in-situ water levels monitoring combined with measurements of level fluctuations over surface water bodies can be used to calculate water flow (*Council et al.*, 1999). In-situ networks, however, are typically limited to local or regional areas and to obtain a global picture of water storage variations. Today, organisations like the European Space Agency (ESA [1]) and the National Aeronautics and Space Administration (NASA [2]) provide various EO data to track changes in global water, carbon, and energy cycles (*Lindersson et al.*, 2020). For instance, altimetry satellites (e.g., Topex/Poseidon, Jason 1 and 2, and ESA's altimetry satellites) observe variations of the Earth's surface water volume changes, including oceans and seas (*Shum et al.*, 1995), as well as lakes, reservoirs, and rivers (*Berry et al.*, 2005). Modern satellite missions, e.g., the Ice, Cloud, and Land Elevation Satellite (ICESat), were designed to collect data on the topography of the Earth's ice sheets, clouds, and vegetation. Over the last two decades, precipitation has been remotely observed over tropical regions using the Tropical Rainfall Measuring Mission (TRMM, *Huffman and Bolvin*, 2015). Microwave remote sensing provides the capability to obtain soil water observation at global and regional scales (*Prigent et al.*, 2005). Global Positioning System (GPS), a constellation of Global Navigation Satellite System (GNSS), presents the displacement of the Earth's crustal deformation, including seasonal changes and long-term trend tectonic motion in the continuous GPS observation stations (*Blewitt et al.*, 2001; *Dong et al.*, 2002). GNSS observations can be used to estimate land surface deformation and the mass balance of the ice sheets, and their current contribution to the sea level rise.

[1] https://eo4sd.esa.int/category/themes/climate-resilience/
[2] https://earthobservatory.nasa.gov/

Among available satellite geodetic and remote sensing EO techniques, the Gravity Recovery And Climate Experiment (GRACE, 2002–2017) satellite mission (*Tapley et al.*, 2004a,b) and its Follow-On mission (GRACE-FO, 2018–onward) provide time-variable Earth's gravity fields that contain signals related to different processes such as non-steric sea level changes, Terrestrial Water Storage Changes (TWSC, i.e., a vertical summation of changes in water storage within plant canopies, surface water, snow, soil and groundwater), ice sheet melting, and PGR, with a spatial resolution of a few 100 km and temporal resolution of $\sim$10 days to 1 month in satellite-only solutions (*Flechtner et al.*, 2016). The ability of GRACE(-FO) satellites to detect mass changes in the surface and sub-surface, which cannot be measured by any other satellite mission, and its sensitivity to water storage changes throughout all seasons, provides a unique opportunity to extract possible intensification's of the water cycle (*Eicker et al.*, 2016; *Kusche et al.*, 2016) and for drought monitoring at global (*Forootan et al.*, 2019; *Zhao et al.*, 2017b) and regional scales (*Houborg et al.*, 2012; *Schumacher et al.*, 2018a; *Sinha et al.*, 2017; *Thomas et al.*, 2014; *Zhao et al.*, 2017a). Although GRACE(-FO) data represent an accurate superposition of water storage changes, separating this integrated signal into its contributors is desirable for many geodynamic and hydro-climatic applications.

1.2 The Quest to Improve Earth System and Geophysical Model Simulations

Geophysical and hydrological models show limited skills to perfectly describe processes within the Earth system. This is due to several assumptions and simplifications associated with the mathematical equations used in these models, and probably due to errors in their structures. In the case of hydrological models, major sources of uncertainty include the lack of accurate climate forcing, such as precipitation, temperature and solar radiation (e.g., *Döll et al.*, 2016), insufficient boundary data, and insufficient knowledge about model parameters to account for intensification of the water cycle caused by the climate change and anthropogenic modifications. Empirical model parameters are often estimated to steer the model equations, but many of these parameters cannot directly be measured, and therefore they might introduce biases to model simulations (*Vrugt et al.*, 2013). Moreover, errors are introduced due to spatial and temporal discretization, as well as due to the background information such as hydro-geology maps. For the GIA forward models, large uncertainties are caused by the lack of knowledge about past ice loading history and Earth structure (in particular, lower mantle viscosity) to estimate the lithospheric response. Besides, mantle viscosity varies laterally, which has been challenging to incorporate into GIA models (*Peltier et al.*, 2015).

In order to improve available models, one solution is to simulate more processes. However, this solution would increase model complexity and introduce new parameters that are not well considered and which render the results exceedingly difficult to interpret. Another solution is to improve model simulation quality by integrating additional observations through data-model fusion techniques, generally known as Data Assimilation (DA). This integration can improve the quality of model simulations by producing outputs that are sampled by direct observations and are closer to reality.

Previous studies indicated how satellite geodetic and EO data can improve monitoring of the water cycle and land surface deformation. For instance, *Sha et al.* (2019) suggested that merging GPS observations with a geophysical forward model can improve the estimation of GIA mass balance on a global or regional scale. Other studies indicated that GRACE TWSC in combination with remotely sensed surface soil water data can be used to improve hydrological and land surface models. (*Girotto et al.*, 2016, 2017; *Khaki et al.*, 2018c; *Schumacher et al.*, 2016, 2018a; *Tian et al.*, 2017; *Zaitchik et al.*, 2008). Regional deep crustal deformation and dynamic tectonic processes have been constrained by GPS and GRACE data in *Fu and Freymueller* (2012); *Hao et al.* (2016); *Pan et al.* (2016). The combination of GRACE, GPS, and altimetry observations is shown in an inversion framework, to provide new estimates of GIA with quantifiable uncertainties. Even though these estimations can lead to refinements of ice-load histories to be used in the forward models (*Milne et al.*, 2004; *Steffen and Kaufmann*, 2005), they are found to be highly dependent on surface density change assumptions, restricted by the low resolution of GRACE data, and limited by the spatial coverage of altimetry data in areas of high relief (*Martin-Espanol and Bamber*, 2016).

1.3 GRACE(-FO) Signal Separation

GRACE(-FO) observations represent a superposition of all mass change signals on land and within the oceans and atmosphere, with non-linear and complex interactions and with many inherent timescales. However, the separation of these integrated signals is only possible by introducing the *a priori* information on mass distribution in each compartment, so mass changes can be reasonably tied to initial state. Merging GRACE(-FO) data with models provides an alternative approach to separate GRACE(-FO) field estimates into its water storage compartments and to downscale the mission's relatively coarse spatial resolution (see e.g., *Forootan et al.*, 2014c; *Girotto et al.*, 2016). Various data-model fusion techniques have been developed in recent year to merge multi-resource data and achieve vertical separation of surface and sub-surface water storage compartments of GRACE data. These techniques are summarized into the following categories:

(a) Forward modelling techniques to evaluate different compartments of mass variations through a simple reduction process, relying on model and/or observation data for other compartments, e.g., surface water and soil water, if groundwater estimation is the target (e.g., *Feng et al.*, 2013; *Khandu et al.*, 2016; *Rodell et al.*, 2009; *Strassberg et al.*, 2009; *Tiwari et al.*, 2009). This method is relatively straight forward, but it is not necessarily the most accurate way to separate GRACE signals due to the reflection of errors in the applied models and/or observation errors on the final estimation of mass changes (see, e.g., *Forootan et al.*, 2014a). Also, the spatial and temporal resolution of the observations (from satellites or in-situ) and model outputs, as well as their signal content are not necessarily consistent (see the discussions in, e.g., *Forootan et al.*, 2014c). Most of these differences are taken into account in methods (b) and (c) that are discussed below.

(b) Statistical inversion techniques, based on statistical signal decomposition techniques, such as Principal Component Analysis (PCA, *Lorenz*, 1956) and its alternatives, e.g., Independent Component Analysis (ICA, *Forootan and Kusche*, 2012, 2013), have been used in previous studies to separate GRACE TWSC into individual water storage estimates. For example, *Schmeer et al.* (2012) used PCA to generate *a priori* information about mass changes from global ocean, atmosphere, and land hydrology models. Then, they applied the least-squares technique to use GRACE TWSC and modify these priori estimates. A statistical inversion, which works based on both PCA and ICA, is proposed in *Forootan et al.* (2014c) and *Awange et al.* (2014) to separate GRACE TWSC using auxiliary data of surface water from satellite altimetry and individual sub-surface water storage estimation from a land surface model (Global Land Data Assimilation System (GLDAS, *Rodell et al.*, 2004). This inversion harmonizes the use of all available data sets within a single least-squares framework. As a result, a more consistent mass estimation (compared to the forward modelling in (a)) for individual water storage compartments can be achieved.

(c) Data Assimilation (DA), as well as simultaneous Calibration/Data Assimilation (C/DA, *Schumacher*, 2016) have been used in recent years to merge GRACE data with hydrological model outputs or other types of observations. These techniques rely on model equations to relate water and energy fluxes with water storage changes. Therefore, unlike the inversion approach in (b), combining information from observations (e.g., GRACE TWSC) and models is performed in a physically justified way. DA or C/DA also increases physical understanding of the model and improves the model states by decreasing the simulation errors. For example, DA is used in *Girotto et al.* (2016, 2017); *Khaki et al.* (2018c,d); *Tian et al.* (2017); *Zaitchik et al.* (2008), while C/DA is applied in *Schumacher et al.* (2016, 2018a) to improve models such as GLDAS (*Rodell et al.*, 2004), World-Wide Water Resources Assessment (W3RA, *Van Dijk*, 2010), WaterGap Global Hydrological Model (WGHM, *Döll et al.*, 2003), and NOAH Multi Parameterization Land Surface Model (NOAH-MP LSM, *Niu et al.*, 2011). Most of the previous DA and C/DA applications were implemented regionally (except *Khaki et al.* (2017a, 2018a); *Van Dijk et al.* (2014)) for example, over the Mississippi River Basin (*Schumacher et al.*, 2016; *Zaitchik et al.*, 2008), Bangladesh (*Khaki et al.*, 2018c), the Middle East (*Khaki et al.*, 2018d), and Murray-Darling River Basin (*Schumacher et al.*, 2018a; *Tian et al.*, 2017). In each of these studies, multiple realisations of the model derived water storage simulations were generated by perturbing the input forcing data and or model parameters. A sequential integration approach such as the Ensemble Kalman Filtering (EnKF, *Evensen*, 1994) or its extensions were then used to merge GRACE data with (ensemble) outputs of a single model (e.g., *Khaki et al.*, 2017b; *Schumacher et al.*, 2016). The statistical information used in EnKF-DA is restricted to the covariance matrices.

Moreover, relying on the simulation of (only) one model may potentially contain errors caused by the model's imperfect structure, such as biases in the internal parameters and/or in boundary conditions. Therefore, in this thesis the formulation of DA using multiple models is addressed.

1.4 Bayesian Inference for Signal Separation

In recent years, Bayesian-based techniques have been developed to merge different observations with models and update the model outputs. For example, *Long et al.* (2017) applied the Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999) technique to average multiple GRACE TWSC products and global hydrological models to analyse spatial and temporal variability of global TWSC. However, they did not assess the update of individual surface and sub-surface water storage estimations.

Sha et al. (2019) presented a model-data synthesis method, based on Bayesian Hierarchical Modelling (BHM, see, e.g., *Banerjee et al.*, 2004), for updating a global GIA model using GPS data. The main feature of their approach is to use observations to adjust a model-based solution by modelling explicitly the discrepancy between the simulation and the true process. Their study, however, only focused on GIA and did not address the estimation of global hydrological mass changes.

It is worth mentioning here that the EnKF used for DA and C/DA can also be classified as a Bayesian-based technique because the cost function for updating the conditionality of unknown state parameters on the measurement data is formulated based on the Bayes theory (e.g., *Evensen*, 2003; *Fang et al.*, 2018; *Schumacher*, 2016).

Methods, such as Particle Filter (PF) and Particle Smoother (PS), are also Bayesian (*Särkkä*, 2013) and have already been applied in a wide range of geophysical and hydrological applications. For example, *Weerts and El Serafy* (2006) compared the capability of EnKF and PF to update a conceptual rainfall-runoff model using discharge and rainfall data. *Plaza Guingla et al.* (2013) used the standard PF to assimilate densely sampled discharge records into a conceptual rainfall-runoff model. *Bain and Crisan* (2008), however, showed that the rate of convergence of the approximate probability distribution until attainment of the true posterior is inversely proportional to the number of particles used in the filter. This means that the filter perfectly approximates the posterior distribution when the number of particles tends to infinity. However, since the computational cost of PF grows with the number of particles, choosing a specific number of particles in the design of filters is a crucial parameter for these methods. The rationale for introducing a new Bayesian data-model merging algorithm in this study is described below.

1.5 Aims and Objectives of the Thesis

This PhD aims at developing new Bayesian frameworks to extract land hydrology and surface deformation information from GRACE(-FO) data. To this end, the predicted water states derived from available models (hydrological and land surface models, as well as GIA models) and other geodetic measurements that observe a part of the water cycle (e.g., GPS measurements) are used as *a priori* information to estimate different compartments of the hydrological water states.

Merging GRACE(-FO) field estimates with the outputs of the models and other geodetic observations is a challenging problem due to the (i) spatial, spectral, and temporal resolution mismatches between these data sets, (ii) different model structures (e.g., different number of soil layers) and various assumptions to simulate water mass changes within the Earth system, (iii) the difficulty in describing the uncertainty of the models and observations in data-model fusion techniques, and (iv) the strong interactions and temporal dependencies between different water storage compartments, which are unknown and need to be considered within formulating the signal separation problem.

To address these issues, the following objectives, hypotheses and strategies for testing them are considered in this thesis.

Objective 1: To test Bayesian frameworks to separate land hydrology components from GRACE(-FO) TWSC, using (multiple) hydrological model outputs, while considering their error estimates.

Objective 2: To evaluate multiple hydrological model outputs against GRACE(-FO) TWSC within a Bayesian Framework.

Objective 3: To formulate a Bayesian framework to simultaneously estimate water storage changes and the unknown temporal dependency between them while using GRACE(-FO) data as observation.

Objective 4: To assess the performance of the extended Bayesian framework for down-scaling (vertically and horizontally) GRACE(-FO) TWSC.

Objective 5: To establish a hierarchical Bayesian optimization approach for merging multiple geodetic observations with (multiple) models.

Objective 6: To assess the performance of the proposed Bayesian optimization approach for joint estimation of land hydrology and surface deformation using GRACE and GNSS in-situ measurements.

This thesis addresses **Objective 1** and **Objective 2**, through the development of the 'Dynamic Model Data Averaging (DMDA)' approach to evaluate available multiple hydrological model outputs against GRACE(-FO) TWSC and sensibly merge them to update the global estimation of surface and sub-surface water storage changes. The main hypothesis behind this approach is that each global hydrological model has its unique characteristics and strengths in capturing different aspects of the water cycle. Therefore, relying on a single model often leads to predictions that represent some phenomena or events well at the expenses of others. Therefore, the effective combination of multiple models may provide more skilful hydrological simulations compared to a single model. The motivation to formulate the DMDA is based on its capability to deal with various observations and models with different structures.

In summary, DMDA is based on the Bayes theory, which combines the benefits of state-space merging techniques, such as Kalman filtering (*Evensen*, 1994) or Particle Filtering (PF, *Gordon et al.*, 1993), Markov Chain (MC, *Chan and Geyer*, 1994; *Kuczera and Parent*, 1998; *Metropolis et al.*, 1953), and Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999). DMDA provides time-variable weights to compute an average of hydrological model outputs, yielding the best fit to GRACE(-FO) TWSC, while considering their error estimates. These dynamic weights indicate which of the available models at a given point in time fits best to GRACE(-FO) TWSC estimates. These weights modify the estimation of water storage changes derived from individual models. Therefore, the DMDA-derived ensemble is expected to yield more skilful (realistic) hydrological simulations compared to any individual model (see similar arguments in *Duan et al.*, 2007).

A realistic synthetic example is set up to test the performance of DMDA, where the true merged values are known, and the method can be evaluated to provide the confidence that it can be applied to a real case study. To test DMDA with real data, GRACE TWSC are merged with outputs of six global hydrological and land surface models within the world's 33 largest river basins, and the results are explored and interpreted using independent validation data sets. Here, the use of DMDA is preferred over the previously introduced EnKF, PF, and PS methods due to its computational efficiency in handling large dimensional problems such as the global integration implemented in this thesis. Besides, the DMDA's time-variable weights can be used to assess the performance of hydrological models, whereas this aspect is missing in other merging techniques.

To address **Objective 3**, a novel Bayesian approach is formulated, which benefits from a combination of a forward-filtering backward-smoothing recursion (*Kitagawa*, 1987) and an efficient Markov Chain Monte Carlo (MCMC, *Geyer*, 1991) algorithm. This method 'MCMC-Data Assimilation (MCMC-DA)' is used to recursively estimate individual surface and sub-surface water storage changes, as well as temporal dependency between them, which are allowed to vary over time. The implementation is realised through a multivariate analysis using the linear 'state-space model' (*Bernstein*, 2005), while both state parameters and the error covariance matrix of the observation can vary in time. MCMC-DA is formulated in a way that the full error covariance matrix of the observed GRACE(-FO) TWSC is introduced to the state-space model as the error covariance matrix of the observation innovation, which is also known as the residual of the observation equation between GRACE(-FO) TWSC and model outputs. This matrix is then used in the forward-filtering backward-smoothing recursion, along with the observed values of the GRACE(-FO) TWSC, to update the model-derived water storage components.

The difference between MCMC-DA and DMDA is that the latter estimates the unknown state parameters using a Kalman filtering (*Kalman*, 1960) approach, while the temporal dependency between the unknown individual water storage states is controlled by a constant forgetting factor ranging from 0 to 1. The computation of this constant value can only be done empirically, and, therefore, is considered as a source of increasing uncertainties in the DMDA-derived individual water storage estimations. The central hypothesis behind formulating MCMC-DA is that the magnitude of changes in water storage components depends on the history of hydrological processes. However, there is little physical knowledge about how this dependency varies over time. Therefore, selecting a constant forgetting factor cannot be physically justified, though the value could be mathematically optimized to provide an overall best fit between the model's unknown states and GRACE(-FO) TWSC. To eliminate this drawback, an MCMC process replaces the constant forgetting factor (of DMDA) to allow a dynamic estimation of temporal dependencies among unknown state parameters (i.e., individual water storage in various compartments). As a result, more realistic estimates of individual water storage components can be expected from the introduced MCMC-DA.

To test this hypothesis, MCMC-DA is implemented to merge GRACE TWSC with water storage outputs of the W3RA water balance model (*Van Dijk*, 2010) within the Conterminous United States (CONUS), on $0.125^\circ \times 0.125^\circ$ spatial grid points for the period 2003–2017. The obtained results are then compared with those derived from the DMDA approach.

To address **Objective 4**, groundwater and soil water storage changes derived from MCMC-DA within CONUS are compared with those of the original model outputs, using independent validation data sets, while the possible relationships between the water storage changes and climatic as well as anthropogenic factors are evaluated.

To address **Objective 5**, a hierarchical Bayesian optimization approach, named 'Constrained Bayesian-Data Assimilation (ConBay-DA)', is proposed for a joint estimation of the land hydrology and PGR rates by merging GRACE(-FO) and in-situ GNSS measurements with the outputs of the hydrological and GIA models. The hypothesis behind this approach is that time-variable gravity data contain both PGR and hydrological signals. For hydrological applications, however, the effect of PGR is typically removed from GRACE(-FO) data, during the post-processing, as a linear trend based on the output of a GIA model. However, these rates might contain considerable uncertainties as shown by (*Guo et al.*, 2012; *Spada et al.*, 2011), and therefore, they negatively affect the accuracy of hydrological mass estimations.

In the ConBay-DA approach, instead of removing PGR from GRACE field estimates, the outputs of a GIA model are used as *a priori* information, along with the hydrological model outputs, to simultaneous separate land hydrology and PGR from GRACE data. It worth noting here that, in many glacial regions of the world, PGR signals are ‘contaminated’ by vertical elastic crustal deformation, which is induced by present ice mass change. Therefore, improvement in the estimation of PGR enhances the estimation of the magnitude and pattern of the elastic crustal deformation (surface deformation) within the continental regions.

ConBay-DA is formulated to estimate a multivariate state-space model between GRACE(-FO) and model outputs (both the hydrological and GIA model). However, when analysing multivariate data, e.g., in a multivariate linear model, one concerning issue is that the relationship between model parameters is unknown (or it is extremely difficult to determine). For instance, it could be expected that certain parameters have a larger effect on the dependent variables than other parameters. This issue in the application of separating TWSC and PGR from GRACE(-FO) estimates can be seen explicitly in the glacial regions, where PGR has a large effect on hydrological estimations. Moreover, PGR manifests as a trend in the relatively short era of the GRACE(-FO) mission, which needs to be introduced to the Bayesian separation framework. These theories can be transformed into the Bayesian fusion technique with specific inequality/equality constraints on the means and regression coefficients. Therefore, in-situ GNSS measurements are used in a hierarchical level to constrain the GIA part by accepting or rejecting the updated value of the GIA model suggested by GRACE(-FO) data.

Estimation of the unknown state parameters and the temporal dependency between them in ConBay-DA are realised through a combination of a forward-filtering backward-smoothing recursion (*Kitagawa*, 1987), and a Gibbs sampling (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993) algorithm. A Metropolis-Hastings (*Chib and Greenberg*, 1995) algorithm is then formulated, in a hierarchical level, to constrain the updated values of PGR rates based on in-situ GNSS data.

To address **Objective 6**, the performance of the ConBay-DA is assessed to separate land hydrology and PGR from GRACE(-FO) data within the Great Lakes (GL) region in North America. The main hypothesis to choose this region is that water mass changes due to PGR and surface deformations are the dominant un-corrected geodynamic effect in GRACE(-FO) data within the GL area. Moreover, *Schumacher et al.* (2018b) indicated that large uncertainties exist between the GIA models over the northern part of the United States. Also, extensive interactions exist between surface water and groundwater in this region due to the fact that the influence of surface water is predominant (*Sophocleous*, 2002; *Winter et al.*, 1998). The numerical implementation of MCMC-DA and ConBay-DA is compared in the GL area to demonstrate how adding GNSS measurement modifies the estimation of hydrological mass changes. It worth mentioning here that, in this thesis, the proposed Bayesian data-model fusion techniques are named ’Bayesian Data Assimilation’ and the technical differences between these two terms are ignored.

1.6 Outline of the Thesis

This PhD thesis provides three data-model fusion techniques to separate different compartments of land hydrology and surface deformation from GRACE(-FO) data, while the outputs of the model simulations and other complementary data sets are used as *a priori* information in these frameworks. Within the application of these approaches, the study is focused on the hydrology signals, while their relations to climate variability and anthropogenic factors are explored and interpreted within different regions of the world. To this end, observations and models of the terrestrial water cycle and surface deformations are described in Chapter 2. The mathematical representation of the observed gravity field changes derived from GRACE(-FO), as well as its conversion to TWSC, are provided in Section 2.1 and Section 2.2. Aspects of data post-processing and the propagation of the error information are also addressed. This is followed by the description of the second geodetic observation-data type used in this study, i.e., Global Navigation Satellite System (GNSS) (Section 2.3). A summary of the hydrological and land surface models used in different applications of this study are explained in Section 2.4. Finally, supplementary data sets, which are required to validate finding of this thesis, are summarized in Section 2.5.

The mathematical foundations of the Bayesian inference are discussed in Chapter 3. This includes an introduction to Bayes' theorem (Section 3.1) and the Gaussian Process (Section 3.2). The basic principles of the Gaussian process regression model and dynamic system modelling are presented in Sections 3.3 and 3.4, respectively. Mathematical descriptions of the state-space model, as the basis of the proposed Bayesian data-model fusion techniques in this study, are discussed in Section 3.4.2. Kalman filtering and Bayesian sampling approaches are the main tools to solve the state-space models and are used to formulate the Bayesian signal separation approach in this study. These techniques are described in detail in Sections 3.5 and 3.6, respectively. Bayesian sampling used in this study includes the MCMC approaches (Section 3.6.1), Gibbs sampling (Section 3.6.2), and Metropolis-Hastings algorithm (Section 3.6.3). Comparison between Gibbs sampling and Metropolis-Hastings is discussed in Section 3.6.4. Moreover, the convergence diagnostics for MCMC approaches are described in Section 3.6.5. This chapter is ended by presenting the concept of Bayesian Model Averaging (BMA) and its formulation in Section 3.7. BMA is then used to formulate the DMDA approach, in Chapter 4.

In Chapter 4, Dynamic Model Data Averaging (DMDA) is formulated. In Section 4.1, a dynamic state-space model is introduced for describing a linear relation between GRACE(-FO) TWSC time series and water storage changes derived from models. The Kalman filter approach, to recursively estimate the unknown state parameters and their uncertainties, is discussed in Section 4.2. In Section 4.3, BMA is formulated to estimate the time-variable weights for each of the multiple hydrological models. These weights are then used in Section 4.4 to dynamically average the model simulations and estimate the improved water storage changes that best resemble measured TWSC. Finally, the performance of DMDA is tested in a controlled synthetic simulation in Section 4.5, where the results of the Bayesian update are known by definition.

MCMC-Data Assimilation (MCMC-DA) is proposed and formulated in Chapter 5. MCMC-DA is formulated based on a multivariate analysis using a linear state-space model between GRACE(-FO) and model-derived water storage changes, as described in Section 4.1. The new formulation consists of the Gibbs sampling algorithm and the forward-filtering backward-smoothing recursion approach, which are explained in Sections 5.2 and 5.1, respectively. Finally, the estimated parameters are used to update the model-derived hydrological compartments and estimate their uncertainties, which are described in Section 5.3.

Constrained Bayesian-Data Assimilation (ConBay-DA) is formulated in Chapter 6. After a short introduction, a multivariate state-space model to define a linear relation between GRACE Total Water Storage Changes (Total-WSC) and the model outputs (both hydrological and GIA model) is explained in Section 6.1. Here, it is also shown how in-situ GNSS measurements can be used to constraint the estimation of PGR from GRACE data. To solve the multivariate state-space model with a hierarchical constraint equation, a combination of forward-filtering backward-smoothing approach, Gibbs sampling, and Metropolis-Hastings is formulated in Section 6.2. Dynamic estimations of the state parameters are then used to update model-derived water storage changes and PGR rates, along with their uncertainties in Sections 6.3.

An application of the DMDA to merge multi-model water storage simulations with GRACE TWSC is described in Chapter 7. After giving an overview of DMDA implementation (Section 7.1), an overview of GRACE TWSC and model-derived water storage changes are presented in Section 7.2 and Section 7.3, respectively. A comparison between TWSC derived from GRACE and models is provided in Section 7.4. The performance of the multiple models to predict water storage changes against GRACE data is evaluated using DMDA estimated weights in Section 7.5. The monotonic changes of the DMDA-derived surface water, soil water, and groundwater storage, in terms of long-term linear trend, are explored in Section 7.6. DMDA-derived TWSC is compared with those derived from original model outputs and GRACE data in Section 7.7. In this study, temporal correlation coefficients between model-derived storage outputs and the El Niño Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) index are used as a measure to determine whether implementing the DMDA helps to derive physically-relevant storage simulations in Section 7.8. A summary of the results and conclusion is provided in Section 7.9.

MCMC-DA is implemented in Chapter 8 to explore meso-scale (10-100 km resolution) water storage changes, mainly focus on soil water and groundwater storage, across the CONUS, with an emphasis on the use of a relatively simple water balance model. In this chapter, after presenting an introduction (Section 8.1), the importance of improving groundwater and soil water storage within CONUS is discussed in Section 8.2. The MCMC-DA groundwater and soil water storage estimates are compared with those derived from the original model outputs in Section 8.3, and the possible relationships between the storage changes and climatic and anthropogenic factors are explored and interpreted in Sections 8.5 and 8.6. Validations are made against independent measurements, i.e., in-situ USGS groundwater level observations, as well as soil water data from the European Space Agency (ESA)'s Climate Change Initiative (CCI). Evaluations using the groundwater levels are made after standardizing the available time series. To extract the influence of ENSO on groundwater and soil water storage estimates, the Independent Component Analysis (ICA, *Forootan and Kusche*, 2012, 2013; *Forootan et al.*, 2018) is applied, and the results are compared with available ENSO indices in Section 8.7. Down-scaling of GRACE TWSC observations using the MCMC-DA approach is evaluated in Section 8.8. Changes in groundwater and soil water storage within the Texas and California states, which are mostly affected by anthropogenic modifications, are evaluated and interpreted in Sections 8.9 and 8.10, respectively. A summary and conclusion is presented in Section 8.11.

An application of the ConBay-DA to test the performance of this approach, is presented in Chapter 9. An overview of the application and data used in this study are provided in Section 9.1 and Section 9.2, respectively. ConBay-DA results are shown and interpreted in Section 9.3 and Section 9.4, and a summary of the results is provided in Section 9.5.

In Chapter 10, the major finding of this PhD thesis are summarized, and an outlook for further research is provided.

Chapter 2

Data and Tools

In this chapter, after introducing the Gravity Recovery And Climate Experiment (GRACE) and its Follow On (GRACE-FO) mission (Section 2.1), the estimation of Terrestrial Water Storage Changes (TWSC) from GRACE time-variable gravity products is addressed in Section 2.2. Discussions on data preparation are explained in Sections 2.2.1, 2.2.2, 2.2.3, and 2.2.4. Spatial averaging is introduced in Section 2.2.5, and error estimation of TWSC fields in Section 2.2.6, which are necessary to understand the investigations that are performed in the application parts (Chapter 7, Chapter 8, and Chapter 9). A summary of the post-processing steps to estimate TWSC from GRACE(-FO) time-variable products is presented in Section 2.2.8. Global Navigation Satellite System (GNSS) observations are explained in Sections 2.3. In Section 2.4, a summary of the hydrological models, used in the application parts of this thesis, is provided. Finally, supplementary data sets, which are required to validate the findings of this thesis, are introduced in Section 2.5.

2.1 Gravity Recovery and Climate Experiment (GRACE) and GRACE Follow On (GRACE-FO)

GRACE is a joint satellite mission of the American National Aeronautics and Space Administration (NASA) and the German Aerospace Centre (Deutsche Zentrum für Luft- und Raumfahrt, DLR), which continuously monitor the Earth time-variable gravity field (*Tapley et al.*, 2004a,b). GRACE, as a low-low satellite-to-satellite tracking (SST) mission, was launched on 17th March 2002. Initially, the mission was targeted to cover a 5-year period, which was exceeded in 2017. The GRACE Follow-On (GRACE-FO) is a continuation of the mission on near-identical hardware, launched in May 2018, but is equipped, in addition with a high precision laser ranging system. Both missions consist of two almost-identical spacecraft in tandem formation (GRACE(-FO) twins) chasing each other in orbit, with an inter-satellite distance of about 200 km, an initial altitude of ∼500 km, and an inclination of 89.5°. Over the last 2 decades GRACE altitude decreased to ∼362 km due to atmospheric drag [1]. The distance between the two satellites is measured by a highly accurate inter-satellite K-Band (Microwave) Ranging (KBR) system in 5-second intervals with an accuracy of about 1 micron. Gravitational variations of the mass within the Earth's interior, on its surface, and in the atmosphere causes the variation in the distance between the two GRACE satellites. Variations in the Earth's gravity field occur due to rapid or slow changes, e.g., caused by mass distribution of the Earth or mass transport of water in the oceans, ice volume changes, as well as Post Glacial Rebound (PGR) and the movement of water vapour and other components in the atmosphere (*Schmidt et al.*, 2008). By this principle, GRACE(-FO) observes the integral sum of mass changes in the Earth system, but it cannot distinguish between the different sources.

Global Positioning System (GPS) receivers are installed on board of satellites to determine their precise location in the high-low mode observations (*Tapley et al.*, 2004a,b). The low-low and high-low observations are shown in Fig. 2.1. Precise orbit determination with Satellite Laser Ranging (SLR) reflectors is also used as an independent check (*Tapley et al.*, 2004a). A high precision accelerometer measures non-gravitational surface forces, dominated by atmospheric drag, which must be removed from GRACE(-FO) observations (*Tapley et al.*, 2004b).

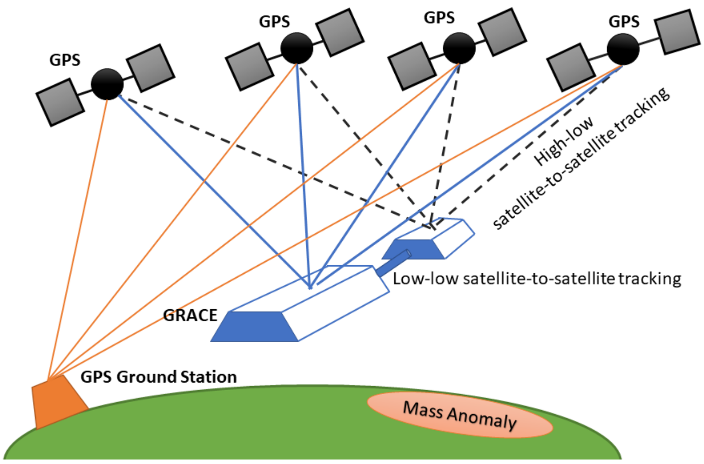

Fig. 2.1 Overview of the GRACE satellite-to-satellite tracking in the low-low and high-low modes.

2.1.1 GRACE Level 2 Products

Various approaches have been developed to process GRACE observations. The traditional processing approach that has been applied over the last 2 decades is parameterizing the Earth's gravity field using global spherical harmonics basis functions (*Wahr et al.*, 1998). Spherical harmonics have been widely

[1] https://www.csr.utexas.edu/missions/

used in satellite geodesy for several decades, based largely on the computational efficiency of the parameterization. Moreover, the satellite sensitivity is dependent on the spatial wavelength of the mass variations which is implicit in the harmonic basis function (*Wahr et al.*, 1998). Time-variable gravity field solutions, represented by the spherical harmonic expansion, are known as GRACE(-FO) level 2 data, which is computed using one month of the pre-processed along-track range (rate) data, derived from level 1B data (*Dahle et al.*, 2013).

Three official analysis centres provide level 2 products, each employing different processing techniques, background models, and assumptions: The Centre for Space Research (CSR, USA), Jet Propulsion Laboratory (JPL, USA), and the GeoForschungsZentrum (GFZ, Germany). These centres provide spherical harmonic potential coefficients (level-2 data products) from the inter-satellite distance measurements. Other research institutes also process GRACE(-FO) data, e.g., Graz University of Technology (TU Graz) in Austria [2], NASA Goddard Space Flight Centre (GSFC/NASA) in the USA [3], Space Geodesy Research Group (GRGS) in France [4], Delft University of Technology (DUT) in the Netherlands, and The Ohio State University in the USA.

One of the main applications of the GRACE(-FO) products is the estimation of seasonal hydrological signals (*Schmidt et al.*, 2008; *Tapley et al.*, 2004a,b). Recovering monthly mean gravity field solutions from the sampled data requires careful reduction of the short-term (sub-daily to monthly) variations of the atmosphere and the ocean mass changes, using a background model, since these effects may alias into longer periods (*Han et al.*, 2004). Moreover, the effect of PGR is usually not reduced, but its impact is treated by post-processing in hydrology applications. Details of post processing steps are described in Section 2.2.

GRACE time-variable level 2 gravity products are provided with a temporal resolution of one month to even one day (*Mayer-Gürr et al.*, 2018; *Ramillien et al.*, 2015), depending on the analysis technique. After post-processing GRACE(-FO) level 2 data, to retrieve reliable information about the Earth mass redistribution, TWSC can be estimated with a spatial resolution of down to a few hundred kilometres (*Schmidt et al.*, 2008). For example, destriping and filtering methods (*Swenson and Wahr*, 2006; *Wahr et al.*, 1998) have been developed with the aim of reducing the noise due to potential errors from spherical harmonic higher degrees (see Section 2.2.2). The scaling factors or other additional methods have been developed by previous studies (e.g., *Feng et al.*, 2012; *Klees et al.*, 2006; *Landerer and Swenson*, 2012; *Long et al.*, 2015; *Longuevergne et al.*, 2010) to restore the lost signal and correct leakage errors, which are mainly caused by signal truncation and filtering in post-processing (see Section 2.2.3).

[2]https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018/
[3]http://grace.gsfc.nasa.gov/
[4]http://grgs.obs-mip.fr/grace

2.1.2 GRACE Mascon Solutions

In addition to spherical harmonics, the other common basis functions to estimate mass flux from GRACE(-FO) is the mass concentration (mascon) block, which has become operational within the past couple of years, e.g., JPL mascon solutions from *Watkins et al.* (2015) and CSR mascon solutions from *Save et al.* (2016). A mascon solution approach estimates the mass anomalies at specified mass concentration blocks or grid location.

In general, there are three different approaches to compute mascon solutions, where (i) is based on an analytic expression for the mass concentration function, in which the mass variations are directly estimated using the explicit partial derivatives relating the inter-satellite range-rate measurements to the analytic mascon formulation. An example of this type can be found in *Ivins et al.* (2011) and *Watkins et al.* (2015). (ii) The second approach is provided by the group at NASA Goddard Space Flight Center (GSFC) (*Luthcke et al.*, 2006, 2013; *Rowlands et al.*, 2010; *Sabaka et al.*, 2010). Similar to (i), the mascon basis functions are directly related to the inter-satellite range-rate measurements through explicit partial derivatives, which are used in the gravity estimation. The difference is that each mascon basis function is represented by a finite truncated spherical harmonic expansion, rather than an analytical expression, such that the functional representation of each mascon has signal power outside of the mascon boundary. Finally, (iii) consists of solutions that users fit mass elements to the spherical harmonic coefficients as a form of post-processing to remove correlated errors. Therefore, these are not level1-derived mascon solutions in the sense that there is no direct relationship between the formulation of the mass elements and the inter-satellite range-rate measurements (i.e., there are no explicit partial derivatives relating the observations to the state). Examples of type (iii) include *Schrama et al.* (2014) and *Velicogna et al.* (2014).

It is known to the science community that although there are similarities in overall patterns of TWSC estimates from existing GRACE solutions, e.g., GRACE level 2 and mascon products, (*Scanlon et al.*, 2016; *Watkins et al.*, 2015), there are also some differences in their long-term trends and the seasonal magnitudes. *Scanlon et al.* (2016) indicated that time series of GRACE TWSC from mascons and spherical harmonics are highly correlated (correlation coefficients of between 0.97-0.99), and the basin average long-term trends for spherical harmonics average 15% less than mascon products. They have also discussed that differences in the long-term trends among GRACE solutions increase with decreasing basin size, which indicates that the processing approach may be more critical for small basins. The advantages of the GRACE mascon solutions relative to the traditional spherical harmonic solutions is that there is no or little requirement for post-processing, which make it much easier for non-geodesists to apply GRACE data to hydrologic problems (*Scanlon et al.*, 2016).

In this thesis, the GRACE level 2 product, i.e., the traditional spherical harmonic processing approach, is applied to estimate GRACE TWSC as observations for the applications of the proposed approaches. For this, the required formulation to convert potential coefficients to TWSC and the techniques to reduce errors are provided in what follows. The reason spherical harmonics are used in this study is to ensure that the TWSC estimates from GRACE and the hydrological model outputs, used in Bayesian data assimilation, have the same spectral content.

2.2 From Geopotential Coefficients to Terrestrial Water Storage Changes (TWSC)

The Earth gravitational potential V (J/kg) satisfies Laplace's equation outside the attracting mass, that is $\Delta V = 0$ (*Heiskanen and Moritz*, 1967). Therefore, its solution can be explained as a sum of harmonic basis functions $Y_n(\lambda, \theta)$ as

$$V(\lambda, \theta, r) = \sum_{n=0}^{\infty} \frac{1}{r^{n+1}} Y_n(\lambda, \theta), \tag{2.1}$$

here λ, θ and r are the spherical coordinate parameters denoting the geographical longitude (radian), co-latitude (radian), and distance (m) to the origin of an Earth-fixed coordinate system, respectively. Equation (2.1) is a spherical representation of the gravitational potential in the exterior of a unit sphere, where $Y_n(\lambda, \theta)$ denotes surface spherical harmonics and define a complete orthogonal system which can be expressed as

$$Y_n(\lambda, \theta) = \sum_{m=0}^{n} [\bar{P}_{nm}(cos\theta)(c_{nm}cos(m\lambda) + s_{nm}sin(m\lambda))], \tag{2.2}$$

In Eq. (2.2) the Stokes coefficients c_{nm} and s_{nm} are fully normalized potential spherical harmonic coefficients (SHCs), derived from GRACE(-FO) level 2 product, n and m are the degree and order of the SHCs respectively, and $\bar{P}_{nm}$ denotes the normalized associated Legendre functions, which can be evaluated using a stable recursion formula (*Heiskanen and Moritz*, 1967). Therefore, replacing Eq. (2.2) in Eq. (2.1), the Earth's gravitational potential V on a sphere with radius R (m) and with total mass of M (kg), can be defined as

$$V(\lambda, \theta, r) = \frac{GM}{R} \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} \bar{P}_{nm}(\cos\theta)\ [c_{nm}\cos(m\lambda) + s_{nm}\sin(m\lambda)], \tag{2.3}$$

where G $(m^3/(kg.s^2))$ denotes Newton's gravitational constant. Mass redistribution in the Earth system is time-dependent, which caused temporal changes in the Earth gravitational potential system. These temporal changes can be represented in terms of changes in the Stokes' coefficients Δc_{nm} and Δs_{nm}, which are computed by subtracting a temporal mean value of the SHCs from each month of GRACE level 2 products (i.e. $\Delta c_{nm} = c_{nm} - \bar{c}_{nm}$ and $\Delta s_{nm} = s_{nm} - \bar{s}_{nm}$, with $\bar{c}_{nm}$ and $\bar{s}_{nm}$ being the temporal means). Assuming global mass conversion, changes in the degree zero coefficient will vanish (see Section 2.2.1) and Eq. (2.3) is modified for mass changes as

$$\Delta V(\lambda,\theta,r) = \frac{GM}{R}\sum_{n=1}^{\infty}\left(\frac{R}{r}\right)^{n+1}\sum_{m=0}^{n}\bar{P}_{nm}(\cos\theta)\,[\Delta c_{nm}\cos(m\lambda)+\Delta s_{nm}\sin(m\lambda)]. \tag{2.4}$$

Wahr et al. (1998) considered a thin layer at the Earth's surface to formulate the loading effect of mass-redistribution and expressed it as surface density redistribution ($\Delta\sigma$), which is related to the temporal gravity changes as

$$\Delta\sigma(\lambda,\theta) = \frac{M}{4\pi R^2}\sum_{n=1}^{\infty}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k_n')}\bar{P}_{nm}(\cos\theta)\,[\Delta c_{nm}\cos(m\lambda)+\Delta s_{nm}\sin(m\lambda)]. \tag{2.5}$$

In Eq. (2.5), $k_n^{'}$ denotes the degree-dependent gravitational load Love numbers, which is used to represent an indirect gravitational attraction through loading and deformation of the underlying solid Earth (*Wahr et al.*, 1998). Solutions for the load Love numbers can be estimated based on Earth models, such as the spherically symmetric, non-rotating, elastic, and isotropic Earth, so-called SNREI models (see, for example, *Farrell*, 1972). Considering the average density of water, $\rho_w = 1025\frac{\mathrm{kg}}{\mathrm{m}^3}$, the density changes derived from Eq. (2.5) can be converted into the changes of Equivalent Water Heights (EWHs) as

$$\Delta E(\lambda,\theta) = \frac{M}{4\pi R^2\rho_w}\sum_{n=1}^{\infty}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k_n')}\bar{P}_{nm}(\cos\theta)\,[\Delta c_{nm}\cos(m\lambda)+\Delta s_{nm}\sin(m\lambda)]. \tag{2.6}$$

EWHs derived from the potential coefficients of GRACE(-FO) level 2 products provide TWSC (as the main input observation in this thesis), which is defined as the sum of changes in all available water storage on and below the surface of the Earth (i.e., canopy, snow, surface water body (lakes, river, ..., soil water, and groundwater storage). This estimation also contains the PGR signals, which needs to be removed from the estimates of Eq. (2.6). Corrections to be applied on the level 2 data are briefly addressed in the following.

2.2.1 Low Degree Coefficients

The spherical harmonic coefficients of lower degree can be directly related to the physical shape of the Earth. The $c_{00} = 1$ ($n = 0, m = 0$) coefficient can be interpreted as a scaling factor for $\frac{GM}{R}$ in Eq. (2.4). Due to the general assumption of mass conversion in the Earth system c_{00} does not change in time, thus $\Delta c_{00} = 0$.

Changes in degree-1 coefficients, i.e., Δc_{10}, Δc_{11}, and Δs_{11} are linked to the offset between the Earth's centre of mass (CM) and the origin of the chosen reference system (centre of figure). Temporal changes in degree-1 coefficients represent a considerable mass variation (*Chambers*, 2006). Therefore, its omission has a considerable impact on the estimation of, e.g., high-latitude mass variability and large-scale oceanic mass changes.

GRACE(-FO) level 2 products are evaluated in a reference frame that is fixed to the instantaneous CM of the Earth, ocean and its surrounding atmosphere. In this frame, the retrieval of the degree-1 coefficients of the surface loading variations cause a singularity. To recover these coefficients, one has to transform the reference frame origin of the GRACE(-FO) data by supplying auxiliary degree-1 coefficients. These coefficients, however, can be augmented by considering the geocentre motion, defined as the relative motion of the centre of figure of the Earth with respect to the CM of the Earth system. Therefore, degree-1 coefficients are usually replaced by geocentre motions estimated by analysing physical models (*Swenson et al.*, 2008) or by using the combinations of GPS, GRACE, and ocean bottom pressure observations in an inversion approach to estimate the geocentre motion (*Rietbroek*, 2014). In this thesis, the degree-1 time series are replaced by those from *Sun et al.* (2016), who used an optimization technique to combine GRACE data with an Ocean Bottom Pressure model and a Glacial Isostatic Adjustment (GIA) model for estimating geocentre motion. This data set can be downloaded from https://grace.jpl.nasa.gov/data/get-data/geocenter/.

Due to the orbital geometry and the short distance between the GRACE(-FO) satellites, the low degree spherical harmonic coefficients, especially those of degree-2 are not well determined. *Chen et al.* (2004) showed that degree-2 variations estimated from accurately measured Earth Orientation Parameters (EOP) or those obtained from the Satellite Laser Ranging (SLR) present better accuracy than those derived from GRACE(-FO). In this thesis, therefore, the time series of zonal degree-2 coefficients (C_{20}) of GRACE(-FO) level 2 are replaced by more reliable estimates of the SLR solutions following *Chen et al.* (2007). The time series can be downloaded from the Jet Propulsion Laboratory (JPL) website [5].

2.2.2 Smoothing in the Spectral Domain

GRACE(-FO) level 2 products, represented in terms of potential spherical harmonics, are strongly affected by correlated errors, which can be seen as a north-south ‘striping pattern’ in the gridded TWSC fields (*Kusche*, 2007). Correlated errors of the potential coefficients are caused by an anisotropic spatial sampling of the mission, instrument noise, and temporal aliasing from an incomplete reduction of short-term mass variations. These errors increase at higher degree and order of the coefficients (*Swenson and Wahr*, 2006).

To reduce the striping errors, smoothing (also called filtering) is applied, which can be implemented in either the spatial or spectral domain. Filtering suppresses the effect of noise in maps, where in the spatial domain in can be applied by a convolution. In the spectral domain, this can be done

[5] grace.jpl.nasa.gov

by incorporating a smoothing kernel (also called filter matrix) W that directly acts on the potential coefficients, i.e., Δc_{nm} and Δs_{nm}, to derive the smooth coefficients $\Delta c^{\mathrm{W}}_{nm}$ and $\Delta s^{\mathrm{W}}_{nm}$. As a result, the EWHs derived from Eq. (2.6) can be filtered in the spectral domain to suppress the noise and derive a smoothed field ΔE^{W} as

$$\Delta E^{\mathrm{W}}(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=1}^{\infty} \sum_{m=0}^{n} \frac{(2n+1)}{(1+k'_n)} \bar{P}_{nm}(\cos\theta)\ [\Delta c^{\mathrm{W}}_{nm} \cos(m\lambda) + \Delta s^{\mathrm{W}}_{nm} \sin(m\lambda)], \tag{2.7}$$

where

$$\begin{aligned}
\Delta c^{\mathrm{W}}_{nm} &= \sum_{n'=1}^{n_{max}} \sum_{m'=0}^{n'} w^{c\,n'm'}_{c\,nm} \Delta c_{n'm'} + w^{s\,n'm'}_{c\,nm} \Delta s_{n'm'}, \\
\Delta s^{\mathrm{W}}_{nm} &= \sum_{n'=1}^{n_{max}} \sum_{m'=0}^{n'} w^{c\,n'm'}_{s\,nm} \Delta c_{n'm'} + w^{s\,n'm'}_{s\,nm} \Delta s_{n'm'}.
\end{aligned} \tag{2.8}$$

In Eq. (2.8), w represents the filter in the spectral domain (*Han et al.*, 2005; *Kusche et al.*, 2009). The sub-indices of c and s (e.g., in $w^{c\,n'm'}_{c\,nm}$ and $w^{c\,n'm'}_{s\,nm}$) indicate that the multiplication produces filtered cosine and sine spherical harmonic coefficients, respectively (see also *Han et al.*, 2005). It should be mentioned here that, in general, the smoothing kernel can be an isotropic or anisotropic filter. Isotropic filters are only degree dependent in the spectral domain and independent of the direction in the spatial domain, e.g., the Gaussian filter introduced by *Jekeli* (1981), thus in Eq. (2.8), $w^{c\,n'm'}_{\ nm} = w^{s\,n'm'}_{\ nm} = w_n$. In contrast, anisotropic filters (decorrelated methods), e.g., those of *Klees et al.* (2008); *Kusche* (2007); *Kusche et al.* (2009); *Swenson and Wahr* (2006), are degree and order dependent in the spectral domain and location-dependent in the spatial domain. The idea behind the decorrelation or anisotropic filtering is to identify and remove error correlation in the sets of spherical harmonic coefficients (i.e., between different coefficients), either based on empirical analysis of the coefficients (*Swenson and Wahr*, 2006) or using an a priori synthetic model of the observation geometry (*Klees et al.*, 2008; *Kusche*, 2007; *Kusche et al.*, 2009). The decorrelation method proposed by *Kusche* (2007) was originally formulated based on computing and applying a filter matrix with as many rows and columns as there are spherical harmonic coefficients. *Kusche et al.* (2009) simplified the decorrelation approach of *Kusche* (2007) to an order-only convolution method (comparable to the approach in (*Swenson and Wahr*, 2006)) which still closely complied with the original, statistically optimal, and full-matrix method. These order convolution filter coefficients were provided to the scientific community for three different degrees of smoothness, which are known as DDK1, DDK2, and DDK3. Equation (2.8) can be used to implement any anisotropic filters such as the DDK filters (*Kusche et al.*, 2009). In this study we use DDK2 filter proposed by (*Kusche et al.*, 2009), which is comparable to a Gaussian filter of 340 km (*Kusche et al.*, 2009), to filter TWSC products for the application parts in Chapter 7, 8, and 9.

In practice, in Eq. (2.7), the summation over n has to be truncated at a maximum degree n_{max} according to

$$\Delta E^{W'}(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=1}^{n_{max}} \sum_{m=0}^{n} \frac{(2n+1)}{(1+k'_n)} \bar{P}_{nm}(\cos\theta)\ [\Delta c_{nm}^{W} \cos(m\lambda) + \Delta s_{nm}^{W} \sin(m\lambda)]. \tag{2.9}$$

and the error due to spectral truncated can be expressed as

(2.10)

$$delta\Delta E^{W'}(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w} \sum_{n=(n_{max}+1)}^{\infty} \sum_{m=0}^{n} \frac{(2n+1)}{(1+k'_n)} \bar{P}_{nm}(\cos\theta)\ [\Delta c_{nm}^{W} \cos(m\lambda) + \Delta s_{nm}^{W} \sin(m\lambda)].$$

(2.11)

In order to use GRACE(-FO) level 2 products in hydrological applications, n_{max} is chosen to be 60 or 90. The omission error of neglecting degrees $n > n_{max}$ can be estimated by evaluating Eq. (2.9) for $n_{max} = 60$ and $n_{max} = 90$. The differences between EWHs estimated based on $n_{max} = 60$ and $n_{max} = 90$ is estimated to be less than 5 mm on a regular grid with 1-degree spatial resolution. The omission error of higher degrees, i.e., $n > 90$, is expected to be even smaller.

2.2.3 Leakage Problem

The noise amplitude of level 2 coefficients increases at high degree and order. This means that short-wavelength spatial changes in GRACE(-FO) TWSC exhibit a high level of noise, which is usually suppressed by applying filters (*Jekeli*, 1981; *Kusche*, 2007). All filters contain a smoothing kernel that spatially averages TWSC values, and as a result, anomalies might move, for example, those inside a specified region can move outside of it, or those from outside might move into the region. This apparent movement introduces biases in the mass change estimates, which is usually referred to as spatial leakage-out and leakage-in problem (*Klees et al.*, 2006; *Swenson and Wahr*, 2002). The impact of spatial leakage on TWSC depends on the filter size, area, the amplitude of mass variations inside and outside the area of interest, and the location of the region. When using a filter with large smoothing radius, more mass signals are distorted and replaced. *Awange et al.* (2009) and *Longuevergne et al.* (2010) showed that the leakage is usually large over smaller regions or along coastlines, where land meets surface water bodies such as lakes and seas.

Since the spatial resolution of filtered GRACE(-FO) TWSC data is typically lower than that of hydrological models or other remote sensing data sets, it is necessary to account for these differences before any further analysis or comparisons can be performed. Otherwise, differences in the data sets due to the scale mismatch might be attributed to limited skills of observations or model simulation. A straightforward way to account for this inconsistency is realized by filtering each data set in the same way. This approach is usually applied for comparing GRACE(-FO) TWSC to other sources such as hydrological models in e.g., *Doell et al.* (2014); *Kusche et al.* (2009); *Schmidt et al.* (2008). Alternatively, *Klees et al.* (2006), *Longuevergne et al.* (2010), *Landerer and Swenson* (2012), and *Feng et al.* (2012) suggested to estimate a scaling factor for the TWSC to reduce the effect of spatial leakage. This can be done by dividing the mean value of the basin function before and after filtering. These factors can be either constant or change when a new field is analysed. Then, the re-scaled GRACE(-FO) TWSC can be directly compared to hydrological model outputs.

Another leakage problem, known as the 'spectral leakage' in literature, is related to the limited range of potential SHCs in GRACE(-FO) level 2 products due to the truncation at a maximum degree n_{max}. This impact restricts the spatial resolution of the GRACE(-FO) TWSC. As a result, water storage signals with spatial variability of smaller than a few hundred kilometres are not well presented in these maps. This limitation makes detection of mass anomalies over, e.g., the land-ocean boundaries, more challenging, whereas one might detect spatially propagated storage change from oceans to land or vice versa that mask each other (see, e.g., *Awange et al.*, 2009; *Chen et al.*, 2006).

2.2.4 Treating the Effect of Post Glacial Rebound (PGR)

PGR (also called isostatic rebound or crustal rebound) is known as the rise of land surface after the removal of the enormous weight of ice sheets during the last glacial period, which had caused isostatic depression. PGR and isostatic depression are phases of Glacial Isostatic Adjustment (GIA), which manifests as a trend in the relatively short era of the GRACE(-FO) mission. For hydrological applications, PGR is often removed as a linear trend based on the output of GIA models (e.g., *Sasgen et al.*, 2012; *Wahr and Zhong*, 2012). As a result, the uncertainty of present-day mass change estimated from GRACE(-FO) depends on the accuracy of GIA models. It should be mentioned that the magnitude of PGR varies over the globe. Regions such as Greenland, North America, Canada, and Scandinavia are more affected by GIA, while those over Africa and the Middle East experience negligible influence. In order to remove model-derived PGR uplift rates from GRACE(-FO) level 2 data, it is needed to expand these values (PGR uplift rates $u(\theta,\lambda)$) as a series of spherical harmonic coefficients as

$$u(\theta,\lambda) = R\sum_{n=0}^{n_{max}}\sum_{m=0}^{n} P_{nm}(cos\theta)\,(c^u_{nm}cos(m\lambda) + s^u_{nm}sin(m\lambda)) \tag{2.12}$$

where R is the mean equatorial radius, θ and λ are the co-latitude and east-longitude, , respectively, P_{nm} are the fully-normalised associated Legendre polynomials and c^u_{nm} and s^u_{nm} are the coefficients for the uplift rates, and can be obtained from Eq. (2.13) through a numerical integration approach (*Wang* *et al.*, 2006) or a least squares approach (*Sneeuw*, 1994).

$$c^u_{nm} = \frac{1}{4\pi R^2}\iint_S \frac{u(\theta,\lambda)}{R} cos(m\lambda)P_{nm}(cos\theta)R^2 d\theta d\lambda,$$
$$s^u_{nm} = \frac{1}{4\pi R^2}\iint_S \frac{u(\theta,\lambda)}{R} sin(m\lambda)P_{nm}(cos\theta)R^2 d\theta d\lambda. \tag{2.13}$$

The spherical harmonic coefficients of the uplift rates are then removed from smoothed spherical harmonic coefficients of GRACE(-FO) level 2 data in Eq. (2.9) to estimate EWHs after removing PGR rates ($\Delta E_u^{\mathrm{W}'}$). Therefore Eq. (2.9) is re-written as

$$\Delta E_u^{\mathrm{W}'}(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w}\sum_{n=1}^{n_{max}}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k'_n)}\bar{P}_{nm}(\cos\theta)\ [(\Delta c^{\mathrm{W}}_{nm} - c^u_{nm})\cos(m\lambda) + (\Delta s^{\mathrm{W}}_{nm} - s^u_{nm})\sin(m\lambda)]. \tag{2.14}$$

2.2.5 Spatial Averaging

For hydrological applications, one might need to compute area-averaged mass variations from GRACE(-FO) level 2 product for a specific area of interest, e.g., a river basin or a continent. To make computations as accurate as possible, this averaging is implemented in the spectral domain to avoid introducing extra biases caused by the resolution mismatch between the (basin) boundary and GRACE(-FO) data. The spatial average of mass changes $\Delta E_f{}^{\mathrm{W}}$ can be computed by multiplying the filtered spherical harmonic coefficients $\Delta c^{\mathrm{W}}_{nm}$ and $\Delta s^{\mathrm{W}}_{nm}$ in Eq. (2.7) with the spherical harmonic coefficients c^f_{nm} and s^f_{nm} of a basin function $f(\lambda,\theta)$ as

$$\Delta E_f{}^{\mathrm{W}}(\lambda,\theta) = \frac{M}{4\pi R^2 \rho_w A_f}\sum_{n=1}^{n_{max}}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k'_n)}\bar{P}_{nm}(\cos\theta)\ [c^f_{nm}\Delta c^{\mathrm{W}}_{nm}\cos(m\lambda) + s^f_{nm}\Delta s^{\mathrm{W}}_{nm}\sin(m\lambda)]. \tag{2.15}$$

In Eq. (2.15), R is the Earth radius, and A_f is the area of the basin function f. Spherical harmonic coefficients of the basin function f can be determined by defining a global grid with one inside the region of interest and zero outside of it, namely

$$f(\lambda,\theta) = \begin{cases} 1 & \text{Within the area of interest,} \\ 0 & \text{Outside the area of interest.} \end{cases} \tag{2.16}$$

The basin function f is defined over a sphere (with radius R) represented by a spherical harmonic expansion

$$f(\lambda,\theta) = \frac{1}{4\pi}\sum_{n=1}^{n_{\max}}\sum_{m=0}^{n}\bar{P}_{nm}(\cos\theta)\;[c^f_{nm}\cos(m\lambda)+s^f_{nm}\sin(m\lambda)]. \tag{2.17}$$

Since the basin function f is known by definition, the coefficients of c^f_{nm} and s^f_{nm} can be computed using an integral approach (*Wang et al.*, 2006) or a least-squares approach (*Sneeuw*, 1994).

2.2.6 Error Estimation of TWSC

GRACE(-FO) gravity field coefficients contain correlated errors (known as striping patterns) that can be significantly reduced by filtering the data. The Root Mean Square (RMS) of the time series of GRACE(-FO) TWSC errors without filtering reaches up to 20 cm, and after filtering (e.g., applying a Gaussian filter with 300 km half-width radius) reduces to a few centimetres (*Wahr et al.*, 2006). When the filtered GRACE(-FO) data is scaled to account for the leakage and damping impact, the residual measurement errors must also be scaled.

To estimate the error of GRACE(-FO) TWSC, the uncertainties of coefficients are propagated as the covariance matrix of EWHs using a formal least squares error propagation method. Most studies only consider the variances of the potential coefficients, which are reported in GRACE(-FO) level 2 products (e.g., level 2 data from the GFZ and JPL centres). These errors, however, represent only the variance part of the estimated variance-covariance matrix of GRACE(-FO) level 2 products. Error estimation of TWSC derived from GRACE(-FO) level 2 is more realistic when all elements of the covariance matrix are considered during the error propagation procedure.

By replacing degree-1 and degree-2 coefficients of GRACE(-FO) level 2 products with other estimations (see Section 2.2.1), one should consider the fact that the errors of new coefficients are not essentially consistent with the errors of level 2 products. To mitigate this inconsistency, one might use low degree products that are estimated in a system involving GRACE(-FO) products, e.g., estimations of the low degree coefficients in *Rietbroek et al.* (2012). Otherwise, correlations of the new low degree coefficients and the GRACE(-FO) derived higher degree coefficients must be introduced.

Moreover, within the processing of satellite gravimetry data, it is common to reduce the high-frequency signals of the Earth rotation, Earth tides, ocean tides, and high-frequency non-tidal oceanic and atmospheric mass redistribution from the level 1B measurements by using 'background' models. Otherwise, such high-frequency mass changes will be aliased into long-wavelength signals leading to misinterpretation of hydrological signals (*Flechtner et al.*, 2015). This reduction, however, is imperfect since the models are not precise enough. However, errors in these models, e.g., in the tidal and non-tidal ocean-atmosphere de-aliasing products (*Forootan et al.*, 2013, 2014a), are not included in the full variance-covariance of monthly GRACE(-FO) level 2 products. It should be mentioned here that since the model errors affect the whole spectrum of the GRACE(-FO) level 2 products, isolating them from geophysical signals is extremely difficult. Errors due to an incomplete reduction of short-term mass variations of atmospheric de-aliasing products are taken into account in this study, similar to *Forootan et al.* (2014a).

Other errors that need to be considered for uncertainty estimation include errors in the GIA models, which are used to remove GIA signal from GRACE(-FO) level 2 products within the highly effected regions, such as Greenland and North America (see Section 2.2.4). While the amplitudes of the GRACE TWSC are estimated between $\sim \pm 300$ mm in global scale, their uncertainties are approximated between 10-20 mm, depending on the region of the study, and the procedure used to estimate these values. An overview of GRACE TWSC and its uncertainty within the world's 33 largest river basins can be found in Section 7.2.

2.2.7 Filling Temporal Gaps Between GRACE and GRACE-FO

Initially, GRACE was targeted to cover a 5-year period, which was exceeded by 10 years to October 2017. The GRACE-FO mission was launched in May 2018, but suffered a failure of the main instrument processing unit between July and October 2018. This has led to approximately one year of data gap (*Li et al.*, 2020), which leads to inability to supply continuous geophysical information.

Many efforts were carried out to explore the potential of bridging the gap between the two GRACE missions with some alternative observations. However, there is no single satellite mission or proxy observation which is able to fill this gap with comparable quality (see e.g., *Cecilia Peralta et al.*, 2016; *Lück et al.*, 2018; *Rietbroek et al.*, 2014). Geodetic remote sensing observation such as SLR, GPS, and low Earth-orbiting satellites, namely the European Space Agency (ESA)'s Swarm Earth explorer mission, provide temporal gravity solutions with lower spatial resolution (*Bezděk et al.*, 2016; *Encarnacao et al.*, 2019; *Jäggi et al.*, 2016; *Lück et al.*, 2018; *Sośnica et al.*, 2015), that can be used as auxiliary observations to bridge the gap of the two GRACE missions (see e.g., *Forootan et al.*, 2020; *Lück et al.*, 2018; *Meyer et al.*, 2019).

Data-driven approaches (*Hasan and Tarhule*, 2020; *Humphrey and Gudmundsson*, 2019) are another technique to reconstruct the GRACE(-FO) TWSC by determining the relationships between TWSC and corresponding climatic and hydrological variables specifically, such as rainfall and temperature. For example, the Artificial Neural Network (ANN) was adopted to learn the relationship between TWSC and related variables (*Ahmed et al.*, 2019). *Forootan et al.* (2014b) first applied the Independent Component Analysis (ICA, *Forootan et al.*, 2012) to separate the GRACE signals into their original sources, then produced the reconstruction and derived the relations based on the autoregressive exogenous (ARX, *Ljung*, 1999). *Li et al.* (2020) applied data-driven methods for reconstructing and predicting GRACE-Like gridded TWSC using climate inputs.

Besides the above approaches, many interpolation methods, such as linear interpolation (*Zotov and Shum*, 2010), cubic spline interpolation (*Guo et al.*, 2018) and least-squares fitting (*Rangelova et al.*, 2010), were used to interpolate missing data with neighboring data. However, the performance of filling gap with these interpolation approaches is basically dependent on the lengths of time series and gaps, availability of neighboring data, and so on (*Taie Semiromi and Koch*, 2019).

For the application part of this study, we only use GRACE observations, so there is no need to fill the gap between GRACE and GRACE-FO. However, for further applications of the proposed approaches, where GRACE(-FO) will be used as the observation data set, the low-degree gravity solutions from the Swarm mission can be used in an iterative reconstruction procedure to fill the existing gap in the GRACE record and the gap between GRACE and GRACE-FO *Forootan et al.* (2020). The reconstruction approach uses Swarm data as initial values in the gapping fields, then make use of the Independent Component Analysis (ICA) to update GRACE values using the statistics existing in GRACE(-FO) and Swarm TWSC fields. This updating procedure iterated until the filling values do not change considerably from one iteration to another. These fields can be used as level 3 TWSC products covering the whole period of 2002 onwards.

2.2.8 Computational Steps for Estimating GRACE TWSC in this Study

In this study, the latest release (RL06) of the monthly GRACE level 2 product, provided by the Centre for Space Research (CSR [6]), is used to estimate TWSC from 2002 through 2017. The potential coefficients are truncated at the spherical harmonics of degree and order 90, resulting in $\sim$300 km spatial resolution at the Equator. To generate monthly TWSC fields from GRACE products, recommended corrections are applied to the GRACE spherical harmonic coefficients.

- Degree-1 coefficients are replaced by those from *Sun et al.* (2016).
- The zonal degree-2 coefficients (C_{20}) are replaced by more reliable estimates of the SLR solutions following *Chen et al.* (2007).
- When needed, surface deformation signals due to the Glacial Isostatic Adjustment (GIA) are reduced using the ICE5G-VM2 GIA model (*Wahr and Zhong*, 2012).
- Correlated errors of the potential coefficients are reduced by applying the DDK filter (DDK2 or DDK3, *Kusche et al.*, 2009). The DDK filter is preferred here to the other filtering techniques since the final smoothed solutions are generally in better agreement with the TWSC output of

[6] http://www2.csr.utexas.edu/grace/

global hydrological models (see, e.g., *Werth et al.*, 2009). It also considers correlations between potential coefficients in a more rigorous manner compared to other filter techniques such as that of *Swenson and Wahr* (2006) who only models the order-dependent correlations.

- The formulation explained in Section 2.2 that follows *Wahr et al.* (1998), is used to convert the level 2 potential coefficients to the gridded TWSC, from 2003 to 2017.
- A single scale factor as described by *Feng et al.* (2012) is used in this study to restore the signal loss due to filtering, which simultaneously minimizes the summation of leakage-in and leakage-out contributions, and to compute the basin averages within a region of interest, e.g., within the river basins.
- Uncertainties in these fields are computed by implementing a collocation error estimation (*Awange et al.*, 2016; *Ferreira et al.*, 2016) using TWSC estimates from the CSR, JPL, and GFZ level 2 data. [7]).
- Errors due to an incomplete reduction of short-term mass variations of atmospheric de-aliasing products are taken into account, similar to *Forootan et al.* (2014a).

2.3 Global Navigation Satellite System (GNSS) Station Data

GNSS refers to satellite constellations that emit signals from space which are decoded by GNSS receivers to estimate time differences between the satellite and GNSS antenna. The receivers then use these measurements to determine the vertical and horizontal components of locations. GNSS constellations include the Europe's Galileo, the United States' NAVSTAR Global Positioning System (GPS), Russia's Global'naya Navigatsionnaya Sputnikovaya Sistema (GLONASS) and China's BeiDou Navigation Satellite System.

In this study, in-situ GNSS uplift rates released by *Schumacher et al.* (2018b) are used to estimate PGR rates in combination with GRACE data in Chapter 9. This data set consists of 4072 in-situ sites over the globe (selected based on prior information from the GIA forward models to exclude tectonic signals), and for this study, 343 sites within the Great Lakes (GL) region are extracted from the original data (see Fig. 9.2).

Vertical motion from GNSS data not only contains the GIA signals in their uplift rates, but also that of other physical processes such as vertical land motion (VLM) (e.g., jumps due to earthquakes or local subsidence) and longer term changes due to natural and anthropogenic processes, tectonics, local hydrology (e.g., groundwater pumping). GNSS hardware changes can also give artificially induced jumps.

[7] https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2018/

Schumacher et al. (2018b) used the GPS data set of the Nevada Geodetic Laboratory (NGL) as the starting point for providing an observational estimate of global GIA VLM, and then used a novel fully-automatic post-processing strategy to deal with the challenges of GPS time-series analysis in general, and for GIA purposes in particular, including outlier and jump detection, atmospheric mass loading correction, elastic signal correction and filtering for stations where other sources of VLM are likely to dominate GIA.

In order to accurately account for the elastic response of the Earth's crust over Antarctica and Greenland, separate data sets are used that have been corrected for the contemporary ice mass loading impact on elastic deformation using high-resolution ice mass balance time-series (*Khan et al.*, 2016; *Martín-Español et al.*, 2016). The non-tidal oceanic and hydrological loading have a similar effect to atmospheric loading on the GPS time-series but both are less well modelled in general (*Santamaría-Gómez and Mémin*, 2015), which means that the loading computations are not as accurate. In *Schumacher et al.* (2018b) GNSS data sets, however, no explicit correction is applied for hydrological mass loading but instead they performed a spatial filtering strategy to select stations that are predominantly influenced by the long wavelength GIA signal and to exclude stations that are affected by local to regional hydrology (such as groundwater pumping).

To compare gridded PGR rates with GRACE(-FO) EWHs, PGR rates can be expanded in a series of spherical harmonics coefficients following Eqs. (2.12) and (2.13). The spherical harmonic coefficients of the uplift rates are then used in Eq. (2.6) to estimate the EWHs corresponding to the PGR rates. Figure 9.3, in Chapter 9, shows the EWHs derived from GNSS solution (*Schumacher et al.*, 2018b), and those derived from the ICE5G-VM2 GIA model (*Wahr and Zhong*, 2012) within the GL area.

2.4 Global Hydrological Models

To simulate large-scale continental and global hydrology, several hydrological models have been developed. Land surface models represent the land-atmosphere interface in climate models and numerical weather prediction and aim to represent the energy and water fluxes by implementing surface energy and water balance equations. Hydrological water balance models are used to reconstruct historical series and predict future ones. They are based on the principle of mass conservation or the continuity equation, which considers that the difference of inputs and outputs will be reflected in water storage in the catchment (*Pérez-Sánchez et al.*, 2019). The main aim of hydrological balance models is to assess inflows in a water resource system, and it is essential for appropriate analysis of its availability. These models are of conceptual structure, i.e., even though the complex physical processes are often known, the model equations are reasonably simplified due to the lack of adequate forcing data sets. In what follows, a short description of models used in this study is presented.

Monthly water storage components from six large-scale Global Hydrological Models (GHMs) including PCR-GLOBWB (*Van Beek et al.*, 2011; *Wada et al.*, 2014), SURFEX-TRIP (*Decharme et al.*, 2013), LISFLOOD (*Van Der Knijff et al.*, 2010), HBV-SIMREG (*Lindström et al.*, 1997), W3RA (*Van Dijk*, 2010), and ORCHIDEE 7 are used as priori information of water storage changes in DMDA application (Chapter 7). The output of these models are published by *Schellekens et al.* (2017), and are available at 0.5° spatial resolution covering the period of 1979–2012. These fields can be downloaded from http://earth2observe.github.io/water-resource-reanalysis-v1.

Although these models are structurally different, i.e., they use different methodologies to simulate water changes, they are driven by the same reanalysis-based forcing data set, WFDEI (WATCH Forcing Data methodology applied to ERA-Interim reanalysis *Weedon et al.*, 2014). In other words, all hydrological models that are used in this study may represent the TWSC, but their respective approaches for simulating TWSC and its corresponding storage compartments are not identical. For example, *Schellekens et al.* (2017) state that PCR-GLOBWB and SURFEX-TRIP contain all surface and sub-surface water storage components in their TWSC estimation. In contrast, TWSC derived from LISFLOOD, HBV-SIMREG, and W3RA are equal to the summation of groundwater, soil water, and snow, while that of ORCHIDEE is the summation of soil water, surface water, and snow storage components. An overview of the model outputs used in this study is provided in Table 2.1. *Schellekens et al.* (2017) also states that among all these six model, only LISFLOOD and ORCHIDEE considered human water use as an input parameter to simulate water storage changes. Comparison between these models in terms of linear trends and seasonality fitted to their water storage components, such as groundwater, soil water, and surface water storage changes, are shown in Section 7.3. To ensure

Table 2.1 Overview of models used in this study and their water storage components.

Model	Water Storage Compartments				
	Groundwater	Number of Soil layer	Surface Water	Canopy	Number of Snow layer
PCR-GLOBWB	Yes	2	Yes	Yes	1
W3RA	Yes	3	No	No	1
HBV-SIMREG	Yes	1	No	No	1
SURFEX-TRIP	Yes	14	Yes	Yes	12
LISFLOOD	Yes	2	No	No	1
ORCHIDEE	No	11	Yes	No	6

that the TWSC estimates from the GRACE level 2 data and model outputs have the same spectral content, 0.5° resolution hydrological model outputs are transformed into the spectral domain and truncated to the maximum degree and order 90 (Eq. (2.9), Section 2.2.2). Basin averages of each model components and their errors in terms of EWH are obtained from the same procedure used to process GRACE level 2 data.

The Worldwide Water Resources Assessment (W3RA, *Van Dijk*, 2010) is used in this study to merge with GRACE TWSC within CONUS for the application of MCMC-DA approach (Chapter 8). The grid-distributed W3RA water balance model was first developed in 2008 by the Commonwealth

Scientific and Industrial Research Organization (CSIRO [8]) to simulate landscape water storage in the vegetation and soil systems at $1^\circ \times 1^\circ$ spatial resolution (*Van Dijk*, 2010). For this study, the original code [9] is modified for the CONUS by using daily $0.125^\circ \times 0.125^\circ$ interpolated ERA-Interim reanalysis fields (*Dee et al.*, 2011) of precipitation, albedo, 2-meter wind, as well as minimum and maximum temperature [10] as forcing data to run the model from 1980-2017. In W3RA, each cell is modelled independently of its neighbours, but lateral mass exchanges are accounted for by implementing a routing scheme. More details on the W3RA model can be found in *Van Dijk* (2010).

For the application of MCMC-DA and ConBay-DA (Chapter 8 and 9), W3RA's monthly averaged model states (snow, surface water storage, surface soil water (top layer), shallow-rooted soil water, deep-rooted soil water storage, and groundwater storage) are used, which are known as the W3RA's water storage components, for the period January 2003 through December 2017. Our motivation in selecting W3RA is its simplicity, which makes its computational load manageable for scientific applications (see examples of W3RA's applications in, e.g., *Forootan et al.* (2019); *Khaki et al.* (2017b)), and its acceptable performance when compared with other commonly used global haydrological or land surface models *Schellekens et al.* (2017).

Model uncertainty is estimated following *Renzullo et al.* (2014) by using the perturbed meteorological forcing approach. To this end, an additive error is assumed for the short-wave radiation perturbation of 50 Wm^2, a Gaussian multiplicative error of 30% for rainfall perturbation, and a Gaussian additive error of 2 °C as the magnitude of the additive error air temperature perturbations. Estimated model uncertainty is used subsequently as the initial value of the variance/covariance matrix of the unknown state parameter (see Section 5.1.1) in the Bayesian inference, which is then updated through a forward-filtering and backward smoothing algorithm presented in Section 5.2.

2.5 Supplementary Data Sets

2.5.1 El Niño Southern Oscillation (ENSO) Index

The El Niño Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) is a large-scale inter-annual climate variability phenomenon in the Tropical Pacific Ocean, which affects the climate of many regions of the Earth due to its ability to change the global atmospheric circulation, which influences temperature and precipitation across the globe (*Forootan et al.*, 2016; *Trenberth*, 1990). The positive phase on ENSO is known as El Niño, and its opposite phase is known as La Nina. The ENSO index used in this study is derived from sea surface temperature in the Niño 3.4 region ($5^\circ N -$

[8] https://www.csiro.au/en/
[9] http://wald.anu.edu.au/challenges/water/w3-and-ozwald-hydrology-models/
[10] https://apps.ecmwf.int/datasets

$5°S, 170°E - 120°W$). Monthly ENSO index (Niño 3.4 index), which is provided by the NOAA National Centre for Environmental Information (NCEI) covering 1948 onward, is downloaded from https://www.esrl.noaa.gov/psd/data/correlation/nina34.data. This index will be used later in this study to demonstrate whether the surface and sub-surface water storage estimates in Chapter 7 and 8 are closer to reality than those from individual models.

2.5.2 In-situ USGS Groundwater Level Data

In-situ groundwater level data (GWL) can be converted to groundwater storage changes (GWS) using an effective Storage coefficient (Sc), where $GWS = Sc \times GWL$, which can be used as an independent validation data set to evaluate groundwater storage changes derived from models and/or DA results. In this thesis, we use the data provided by the US Geological Survey (USGS) groundwater network [11], which contains a record of groundwater levels between 1970-now for ∼100000 wells across the CONUS in Chapter 8. The point-wise data for 2003–2017 are downloaded and filtered to exclude measurements with large data gaps (temporal gaps > 2 years), and those without any variation, in their time series's (e.g., those time series which only contain linear and/or non-linear trends, without any other signals are excluded from data sets). Selected groundwater levels (∼38000 wells) are then temporally averaged to produce monthly time series. The stations located in each $0.125° \times 0.125°$ grid are then spatially averaged, for each time step separately, to produce a time series that is comparable to the W3RA model outputs.

The USGS groundwater network covers a range of unconfined to confined conditions that are important to consider in evaluating groundwater level records and comparing with modelled ground water storage changes. The storage coefficient (Sc), required to convert groundwater level to groundwater storage, can vary over several orders of magnitude from unconfined aquifers (e.g., between 0.02 and 0.3) to confined aquifers (known as storativity ~ 0.001) (*Freeze and Cherry*, 1979). However, there is a continuum between unconfined to confined conditions with some aquifers predominantly semi-confined, such as the California Central Valley. In systems with vertically stacked aquifers, it is often difficult to determine whether wells are screened in unconfined or confined aquifers, or both, which increases the uncertainty of estimating groundwater storage from groundwater level data. Therefore, different approaches have been introduced to approximate it regionally. For example, *Rodell et al.* (2007) applied an average value of 0.15 in the Mississippi River basin, and *Xiao et al.* (2015) used a range of 0.02 to 0.6 in the Mid-Atlantic Region of the CONUS based on the technical insights provided by USGS. However, since the *Sc* estimates are only available for some regions, for the entire CONUS, the USGS groundwater level data are standardized by subtracting their temporal mean and dividing the residuals by their standard deviations. These values are then used to evaluate standardised groundwater storage estimates derived from the W3RA and the MCMC-DA (see Section 8.5.3).

[11] https://water.usgs.gov/ogw/networks.html

Comparisons in terms of water storage changes are performed for Texas and California, where data on storage coefficients are available from previous studies, i.e., High Plain aquifer within Texas, *Gutentag et al.* (2014), and Central Valley aquifer within California, *Scanlon et al.* (2012b). The average Sc in these regions is reported to be 0.18 and 0.15, respectively.

2.5.3 ESA CCI Satellite-Derived Soil Data

The European Space Agency's Climate Change Initiative (ESA CCI) soil moisture product (*Gruber et al.*, 2019) is used in this study to validate the top layer ($<$ 10 cm) soil water storage of W3RA and those from MCMC-DA in Chapter 8. The ESA CCI soil moisture algorithm generates consistent, quality-controlled, and long-term soil moisture climate data records by harmonising and merging soil moisture retrievals from multiple satellites into (i) an active-microwave-based only (ACTIVE), (ii) a passive-microwave-based only (PASSIVE) and a (iii) combined active–passive (COMBINED) product (*Dorigo et al.*, 2017). According to (*Dorigo et al.*, 2017) and a review of existing literature, the ESA CCI product quality has steadily increased with each successive release and that the merged products generally outperform the single-sensor input products. ESA CCI soil moisture generally agrees well with the spatial and temporal patterns estimated by land surface models and observed in-situ, however, (*Dorigo et al.*, 2017) identified surface conditions (e.g., dense vegetation, organic soils) for which it still has large uncertainties (see similar discussion of this study in Section 8.6).

In this study, v04.7 released of daily ESA CCI with the spatial resolution of $0.25^\circ \times 0.25^\circ$ covering the period of 2003–2017 is downloaded from the ESA website[12]. The monthly ESA CCI soil moisture time series is computed by the temporal averaging of daily products. These values are then spatially interpolated (using linear interpolation, *Meijering*, 2002) on the same $0.125^\circ \times 0.125^\circ$ grids as in W3RA in the application part of MCMC-DA, and on $1^\circ \times 1^\circ$ grids as in W3RA in the application part of ConBay-DA. The volumetric units (m^3 m^{-3}) of the ESA CCI is converted to the vertical changes in soil water storage (in mm) using a weighted averaging of the first 2 layers (0-5 cm, 5-10 cm) of the STATSGO porosity values [13].

[12]http://www.esa-soilmoisture-cci.org
[13]http://www.soilinfo.psu.edu/

Chapter 3

An Overview of Bayesian Fusion Techniques

In this chapter the mathematical foundations of the Bayesian inference are discussed. An introduction of Bayes' theorem and the Gaussian Process are presented in Section 3.1 and Section 3.2, respectively. The basic principles of the Gaussian process regression model and dynamic system modelling are shown in Sections 3.3 and 3.4, respectively. The mathematical descriptions of the state-space model is represented in Section 3.4.2, and Kalman filtering and Bayesian sampling approaches, as the main tools to solve the state-space models in the proposed Bayesian signal separation frameworks in this PhD thesis are described in detail in Sections 3.5 and 3.6. This chapter is ended by presenting the concept of Bayesian Model Averaging (BMA) and its formulation in Section 3.7.

3.1 Introduction to Bayesian Statistics

Bayesian statistics is a class of statistical methods that provide a coherent framework for learning and problem solving under uncertainty conditions. The foundation of statistics is based on the theory of probability, which is understood as a measure of the plausibility and uncertainty of a statement. Bayesian statistics mostly involve **conditional probability**, which is the probability of an event (e.g., Θ) given another event (e.g., Y), and can be calculated using the **Bayes' theorem**. In traditional statistics, which is not founded on Bayes' theorem, the probability is only associated with random experimental results, while Bayesian statistics allow for probabilities of all statements or propositions. The advantage of Bayesian statistics in comparison to traditional statistics is that by using Bayes' theorem and estimating the probability density functions for the unknown parameters, the method of testing hypotheses or estimating their confidence regions can be readily tackled by these approaches

(*Koch*, 2007). Therefore, Bayesian methods face rapid expansion and affect many application areas, where the uncertainty is inherent in many processes involved.

3.1.1 Conditional Probability and Bayes' Theorem

Conditional probability is the probability of one event, e.g., Θ, occurring with some relationship to one or more other events, e.g., Y, and is the key concept in Bayes' theorem. One writes $\Theta|Y$ to denote the situation that Θ is true under the condition that Y is true. If Θ and Y are two events in a sample space S, then the marginal probability of events Θ and Y are denoted by $p(\Theta)$ and $p(Y)$, while the joint probability of events Θ and Y is denoted by $p(\Theta,Y)$, and the probability of Θ conditional on Y is denoted by $p(\Theta|Y)$, and is defined as

$$p(\Theta|Y) = \frac{p(\Theta,Y)}{p(Y)}, \qquad \text{where} \quad p(Y) > 0. \tag{3.1}$$

From the probability rules (*Koch*, 2007), $p(\Theta,Y)$ is the probability of both Θ and Y occurring, which is equal to $p(Y,\Theta)$. Equation (3.1) indicates that $p(\Theta,Y) = p(\Theta|Y)p(Y)$. Similarly, we can state that $p(Y,\Theta) = p(Y|\Theta)p(\Theta)$. The fact that these two expressions are equal (i.e., $p(\Theta,Y) = p(Y,\Theta)$) leads to **Bayes's theorem**, which is defined as

$$p(\Theta|Y) = \frac{p(Y|\Theta)p(\Theta)}{p(Y)}. \tag{3.2}$$

Given the hypothesis Θ (i.e., the parameters of interest), and evidence Y (i.e., observed data), Bayes' theorem states the relationship between the probability of the hypothesis before getting the evidence, i.e., $p(\Theta)$, and the probability of the hypothesis after getting the evidence, i.e., $p(\Theta|Y)$. For this reason, the marginal probability of $p(\Theta)$ is called the *prior probability* distribution of the parameter of interest, while $p(\Theta|Y)$ is called the *posterior probability* distribution. It is worth noting that Bayesian statistics typically involve using probability distributions rather than point probabilities for the quantities in the theorem. It is due to the fact that the prior probability is derived from previous samples and is not known and a fixed population quantity, which precisely determines why different sources may give various estimations of this prior probability. Therefore, our knowledge of the prior probabilities is not likely to be perfect.

From a Bayesian perspective, the prior probability distribution of a quantity in the theorem captures our prior uncertainty about its true value. The inclusion of a prior probability distribution ultimately produces a posterior probability that is also no longer a single quantity; instead, the posterior becomes also a probability distribution.

In Eq. (3.2), $p(Y|\Theta)$ is termed as the sampling distribution, which is proportional to the "likelihood function", only differing by a constant that makes it a proper density function, and sometimes written as $L(Y|\Theta) = p(Y|\Theta)$. In Bayes' theorem, $P(Y)$ is the marginal probability distribution of the observed data (or evidence), and it can be estimated as

$$p(Y) = \int p(Y|\Theta)p(\Theta)d\Theta. \tag{3.3}$$

The marginal probability distributions, sometimes called the "marginal likelihood" for the data, acts as a normalising constant to make the posterior distribution proper (see, e.g., *Raftery* (1995) for the importance of using marginal likelihood). Bayes' Theorem for probability distributions is often stated as the posterior distribution of the parameter of interest, which is proportional (denoted here as $\propto$) to the likelihood density of the observed data conditional on the parameters of interest multiplied to their initial values. This definition can be simplified as

$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}. \tag{3.4}$$

3.1.2 Prior and Posterior Probability Distributions

In all application of Bayesian methods, it must be acknowledged that both the prior and the likelihood have only been specified as a convenient approximation to the beliefs of the analyst. The prior distribution is often considered to be the most controversial element of Bayesian statistics. In principle, the prior should be found by introspection and consideration of all available information about the hypothesis before taking the evidence or data into account. However, this constitutes a non-trivial task, and several approaches have been suggested to formulate priors in practice.

The prior is usually chosen from a parametric family of distributions or mixtures of those, which often gives a reasonable compromise between an accurate representation of prior beliefs and analytical tractability. In practice, it is usually helpful to generate samples from the initial information and inspect whether the values are reasonable *a priori* information. Furthermore, the corresponding sampling distributions can be used to generate synthetic data sets whose properties should also conform with prior beliefs. Nevertheless, in any Bayesian analysis, it is recommended to examine how sensitive the posterior distribution reacts to the prior distribution changes. Comparing posteriors and priors can illustrate how informative the observed data is about the parameters. When the observed data does not contribute to reducing uncertainty about a parameter, then both distributions will be the same, expressing that the beliefs are unchanged (*Jeffreys*, 1998; *Koch*, 2007).

Prior distributions about the parameter of interests (hypothesis), which might come from a literature review or explicitly from earlier data analysis, are known as traditional *informative prior*. *Weakly informative prior* distributions do not supply any controversial information but are strong enough to pull the data away from inappropriate inferences that are consistent with the likelihood. Other prior distributions that are uniform, or nearly so, and allow the information derived from the likelihood to be interpreted probabilistically are *non-informative priors*, or maybe, in some cases, weakly informative (more information can be found in *Berger and Bernardo*, 1989; *Jeffreys*, 1998).

Selecting an appropriate prior is a key component of Bayesian modelling. With only a finite amount of data, the prior can have a very large influence on the posterior. It is important to be aware of this and understand the sensitivity of posterior inference to the choice of prior. In practice, it is common to use non-informative priors to limit this influence; when conjugate priors are chosen for tractability reasons.

In Bayesian theory, the prior is called a *conjugate prior* for the likelihood function $p(Y|\Theta)$ when the posterior distributions of the parameter of interest $p(\Theta|Y)$ have the same functional form as the prior probability distribution $p(\Theta)$. For example, the Gaussian family is conjugate to itself (or self-conjugate) concerning a Gaussian likelihood function; if the likelihood function is Gaussian, choosing a Gaussian prior over the mean will ensure that the posterior distribution is also Gaussian. In Bayesian inference, conjugate priors are useful because they reduce Bayesian updating to modify the prior distribution parameters, rather than computing integrals, and ensure that the posterior is tractable even after multiplying the likelihood by the prior. Moreover, conjugate prior allows for efficient inference algorithms because the posterior and prior share the same functional form (*Gutiérrez-Pena and Smith*, 1995).

3.2 Gaussian Process

A Gaussian process is a class of stochastic processes designed to solve probabilistic classification problems, and can be used as a prior probability distribution over countably or continuous functions in Bayesian inference. Recent advances in Bayesian frameworks have developed an elegant and principled methodology for using Gaussian processes to express uncertainty over the space of functions that fit some set of empirical data due to their flexible non-parametric nature and computational simplicity.

The distribution of a Gaussian process is defined for the infinite number of possible outputs of a function, in which the distribution over any finite number of them has a joint Gaussian distribution (*MacKay*, 1997). Gaussian distributions, known as normal distributions, are completely specified by their first and second-order statistics. In other words, a normal distribution is defined as one whose higher-order cumulants are all zero (*Cardoso*, 1999). In Eq. (3.5), $p(\Theta)$, at a particular input data Θ, is a Gaussian distributed function if

$$p(\Theta) \sim \mathcal{N}(\mu(\Theta), k(\Theta, \Theta)), \tag{3.5}$$

where $\mu(\Theta)$ and $k(\Theta,\Theta)$ are the mean and the covariance kernel functions, respectively. These two functions fully specify the Gaussian process. To mathematically show a Gaussian process, let χ be any set, and Θ^i denotes any finite subset of Θ at a particular input location, where $\Theta^i = [\Theta_n, \ldots, \Theta_{n+i}] \in \chi$, $n = 1, ..., N$. Following the above definition, $p(\Theta)$ is a Gaussian process if its values at any finite number of points, e.g., $\Theta^i, \Theta^j, \Theta^k, \dots \in \chi$, have a joint Gaussian distribution as

$$p(\Theta^i,\Theta^j,\Theta^k,\ldots) = \mathcal{N}\left(\begin{bmatrix}\mu(\Theta^i)\\ \mu(\Theta^j)\\ \mu(\Theta^k)\\ \vdots\end{bmatrix}, \begin{bmatrix}k(\Theta^i,\Theta^i) & k(\Theta^i,\Theta^j) & k(\Theta^i,\Theta^k) & \ldots\\ k(\Theta^j,\Theta^i) & k(\Theta^j,\Theta^j) & k(\Theta^j,\Theta^k) & \ldots\\ k(\Theta^k,\Theta^i) & k(\Theta^k,\Theta^j) & k(\Theta^k,\Theta^k) & \ldots\\ & & & \ddots\end{bmatrix}\right). \tag{3.6}$$

Figure 3.1 shows a graphical demonstration of a 2-D, i.e., 2 dimensional, multivariate Gaussian process defined by Eq. (3.6). Most of the structure of a Gaussian process derived from its covariance

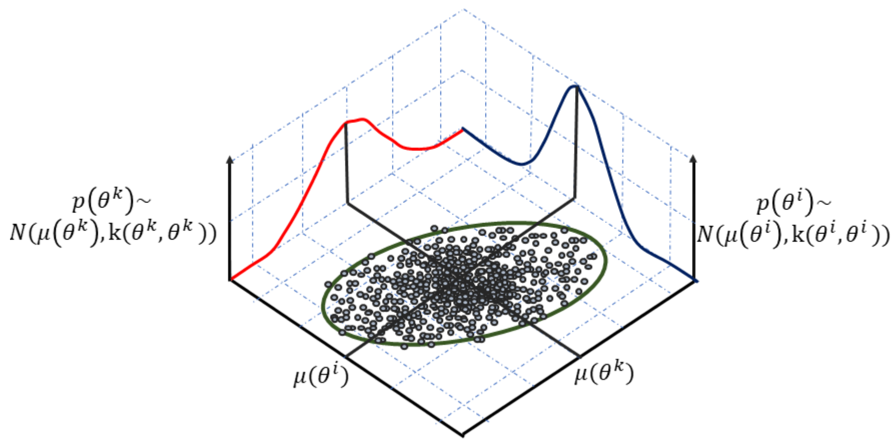

Fig. 3.1 Graphical definition of a 2-D multivariate Gaussian process, i.e., $p(\Theta^i,\Theta^k)$, defined in Eq. 3.6.

kernel function k, which describes how the values of sampled functions vary across nearby (or not-so-nearby) points. The Gaussian process covariance functions provide a very flexible and elegant methodology for establishing priors over random functions. Any positive semi-definite kernel can be used to specify the covariance function. Different covariance functions encourage different degrees of smoothness. Probably the most widely used one is the exponentiated quadratic covariance, also commonly referred to as the squared exponential, Gaussian or Radial Basis Function (RBF) kernel

$$k(\Theta^i,\Theta^j) = \sigma^2 exp(\frac{||\Theta^i - \Theta^j||^2}{\lambda^2}) \tag{3.7}$$

where i, j are the identity factors, σ^2 is a scale parameter that indicates the amount by which locality should influence the covariance, and λ is the length scale, that specifies the smoothness of the function. σ^2 and λ are the *hyperparameters* and can be selected via a maximum likelihood optimisation procedure. Having σ^2 and $\mu(\Theta)$ the general form of the normal probability density function of Θ (Eq. (3.5)) is defined as

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

3.3 Regression Models with Gaussian Process

The regression model is the simplest in which one can appreciate usefulness of Gaussian processes for time series modelling. To describe a regression model through a Gaussian process, consider a functional mapping $\mathscr{X} \to \mathscr{Y}$ from a D-dimensional input data $Z \in \mathscr{X}$, i.e., $\mathscr{X} = \mathbb{R}^D$, to a real-valued target $Y \to \mathscr{Y}$ where

$$Y = f(Z;\Theta) + \varepsilon. \tag{3.9}$$

In Eq. (3.9), f denotes a continuous function mapping $\mathscr{X} \to \mathscr{Y}$, Θ is a set of unknown parameters, known as weights to build a regression model, and ε is additive measurement error or noise. In the above formulation, a functional mapping is specified from Z to Y, but it is accepted that there may be some amount of noise involved in the mapping. Setting the noise in this manner means that for a given input Z, there is now a distribution over possible mappings to Y. In Eq. (3.9), Z can be water storage changes derived from hydrological models for a finite number of points t, i.e., $Z = [Z_1, Z_2, \ldots, Z_t]$, and Y denotes GRACE TWSC as the target values for each point, i.e., $Y = [Y_1; Y_2; \ldots; Y_t]$, while assuming zero-mean Gaussian noise, i.e., $\varepsilon \sim \mathscr{N}(0, V)$, where $V = k(\varepsilon, \varepsilon)$ is the covariance matrix of the GRACE TWSC observations. In the case of a simple linear mapping, $f(Z;\Theta)$ is defined as $f(Z,\Theta) = Z^T\Theta$, where $\Theta = [\Theta_1, \Theta_2, \ldots, \Theta_t]$ are the unknown regression parameters to define linear regression between GRACE observation and model outputs. Assuming spherical zero-mean Gaussian noise gives rise to a Gaussian likelihood over the entire data set (assuming independence) as the product of the single case likelihood over the t observations as

$$p(Y|Z,\Theta) = \mathscr{N}(Z^T\Theta, V). \tag{3.10}$$

Note that under a Gaussian noise model, an infinite number of possible mappings into Y for any given input Z is possible. Thus, the infinite number of repetitions for the above procedure will produce an infinite number of unique functions over Y. Intuitively, many of these functions are implausible to have produced the set of observed targets. When using Bayesian inference on a parametric model, an initial value is chosen on the parameter of interest $p(\Theta)$ and a posterior distribution over the parameter given the data $p(\Theta|Y)$ is estimated by combining the prior with the likelihood function $p(Y|\Theta)$ (see Eq. (3.2)).

A parametric approach to Bayesian regression consists of specifying a family of functions parameterised by a finite set of parameters, putting a prior on those parameters and performing inference. However, we can find a less restrictive and potent approach to inference on functions by directly specifying a prior over an infinite-dimensional space of functions. It contrasts with putting a prior over a finite set of parameters that implicitly specify a distribution over functions. A very useful prior over functions is the Gaussian process, which addresses how one defines a distribution over this infinite set of functions in a principled way using Bayesian probability theory. Gaussian processes take advantage of an *a priori* assumption to specify a preference over functions. This is expressed through setting a Gaussian prior over the parameter of interest, $p(\Theta_0) = \mathcal{N}(\theta_0, \Sigma_0)$, with some positive definite covariance matrix Σ_0. As the functions are defined by their parameters, specifying a distribution over the parameters is analogous to specifying a distribution over functions. Thus following Bayes' theorem (see Eq. (3.2)), a posterior distribution over the unknown parameters is defined as

$$p(\Theta|Y,Z) = \frac{p(Y|Z,\Theta)p(\Theta)}{p(Y|Z)}, \tag{3.11}$$

Following Eq. (3.3) the marginal likelihood $p(Y|Z)$ is given by integrating out the weights or parameters Θ as

$$p(Y|Z) = \int p(Y|Z,\Theta)p(\Theta)d\Theta. \tag{3.12}$$

The posterior estimation over parameters (Eq. (3.11)) can be useful in itself to e.g., update the input data Z (e.g., hydrological model outputs) with respect to the observed values Y (e.g., GRACE TWSC), but it can also be used to make much richer predictions for the targets of a given data point. The predictive distribution over targets $\hat{Y}_{t+1} = f(Z_{t+1};\Theta_{t+1})$ for a novel data point Z_{t+1} can be specified through averaging the targets given over each possible set of parameters, weighted by the posterior probability of the parameters $\Theta = [\Theta_1, \Theta_2, ..., \Theta_t]$ as

$$p(\hat{Y}_{t+1}|Z_{t+1},Z,Y) = \int p(\hat{Y}_{t+1}|Z_{t+1},\Theta)p(\Theta|Z,Y)d\Theta, \tag{3.13}$$

where the first term in the integral, i.e., $p(\hat{Y}_{t+1}|Z_{t+1},\Theta_{t+1})$, results from the Gaussian process prior linking all possible values of Θ_{t+1} and $\hat{Y}_{t+1}$ with a joint normal distribution (*Rasmussen*, 2003), and the second term, $p(\Theta|Z,Y)$, is simply the posterior of Θ from Eq. (3.11). Therefore, when the likelihood has a Gaussian distribution as Eq. (3.10), both the posterior and predictive distributions are Gaussian. For other likelihoods, one may need to resort to approximation methods (see, e.g., *Murray et al.*, 2010; *Nguyen and Bonilla*, 2014).

3.4 Dynamic System Modelling with Gaussian Process

Learning dynamical systems, also known as system identification or time series modelling, aims to create a model or improve an existent model based on measured signals. The developed/improved model can be used later to predict the system's future behaviour or explain the observed data's structure, or de-noise the original observed time series (*Lennart*, 1999). Two important families of dynamical system models are: (I) autoregressive models; (II) the state-space models. The unified description of these dynamical systems is provided in this section. Furthermore, generative models to easily identify the assumptions made by each model and the different inference/learning algorithms that have been tailored to each model are discussed in the following section.

3.4.1 AutoRegressive Models

Autoregressive models describe a time series by defining a mapping from past observations to the current observation. The term "autoregression" indicates that it is a regression of the variable against itself. An autoregressive model of order p is written as

$$Y_t = f(Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p}; \Theta_{t-1}, \Theta_{t-2}, \ldots, \Theta_{t-p}) + \varepsilon_t, \tag{3.14}$$

where $\varepsilon_t \sim \mathcal{N}(0, V_t)$ represents random noise that is independent and identically distributed across time, and in a simple linear autoregressive model, Eq. (3.14) is re-written as

$$Y_t = \Theta_1 Y_{t-1} + \Theta_2 Y_{t-2} + \cdots + \Theta_p Y_{t-p} + \varepsilon_t. \tag{3.15}$$

Changing the parameters of $\Theta_1, \Theta_2, \ldots, \Theta_p$ results in different time series patterns. Therefore, autoregressive models are remarkably flexible at handling a wide range of different time series patterns. An important characteristic of this model is that there is no measurement noise (see Section 3.4.2). Here, noise injected via ε_t influences the future trajectory of the system. Learning in this model is performed using conventional Gaussian process regression techniques. In particular, exact inference is possible if we choose a conjugate likelihood as

$$p(Y_t | Y_{t-1}, \ldots, Y_{t-p}; \Theta_{t-1}, \Theta_{t-2}, \ldots, \Theta_{t-p}) \sim \mathcal{N}(f(Y_{t-1}, \ldots, Y_{t-p}; \Theta_{t-1}, \ldots, \Theta_{t-p}), V_t). \tag{3.16}$$

Non-linear autoregressive models with external (exogenous) inputs are often known as NARX models. *Gregorcic and Lightbody* (2002) and *Kocijan et al.* (2005) presented learning of a Gaussian process-based NARX model via maximisation of the marginal likelihood, and *Girard et al.* (2002) proposed a method to propagate the predictive uncertainty in GP-based NARX models for multiple-step ahead forecasting.

3.4.2 State-Space Models

State-space models refer to a class of probabilistic graphical model (*Koller and Friedman*, 2009) that describes the probabilistic dependence between the latent (unobserved) state variables (e.g., hydrological model outputs stored in matrix Z), state parameters (e.g., Θ), and the observed measurement (e.g., GRACE TWSC in Y). The state parameters or the measurement can be either continuous or discrete. The state-space model provides a general framework for analysing deterministic and stochastic dynamical systems that are measured or observed through a stochastic process, such models have been successfully applied in engineering, statistics, computer science and economics to solve a broad range of dynamical system problems. Other terms used to describe state-space models are hidden Markov model (*Rabiner*, 1989) and latent process models. In probability theory, a Markov model is a stochastic model used to model pseudo-randomly changing systems, where it is assumed that future states depend only on the current state, not on the events that occurred before it (it is known as the Markov property). Generally, this assumption enables reasoning and computation with the model that would otherwise be intractable (*Gagniuc*, 2017).

The objective of state-space modelling is to compute the optimal estimate of the state parameters given the observed data, which can be derived as a recursive form of Bayes's rule (*Brown et al.*, 1998; *Chen et al.*, 2010b). In a general state-space formulation, let Θ_t denotes the state parameters at time t, and $Y_{1:t}$ denotes the cumulative observations up to time t, i.e., $Y_{1:t} = [Y_1, Y_2, \ldots, Y_t]$. Based on the Bayes' theorem defined in Eqs. (3.1) and (3.2), the filtering posterior probability distribution of the state parameters conditional on the sequence of observations is

$$p(\Theta_t|Y_{1:t}) = \frac{p(\Theta_t, Y_{1:t})}{p(Y_{1:t})} = \frac{p(Y_t|\Theta_t, Y_{1:t-1})p(\Theta_t|Y_{1:t-1})}{p(Y_t|Y_{1:t-1})}, \tag{3.17}$$

where the last equality of Eq. (3.17) follows the conditional independence assumption between the observations. The one-step state prediction, known as the *Chapman-Kolmogorov equation*, is defined as

$$p(\Theta_t|Y_{1:t-1}) = \int p(\Theta_t|\Theta_{t-1})p(\Theta_{t-1}|Y_{1:t-1})d\Theta_{t-1}. \tag{3.18}$$

Equations (3.17) and (3.18) provide the fundamental relations to develop state-space models, where the probability distribution (or density) $p(\Theta_t|\Theta_{t-1})$ describes a *state transition equation*, and $p(Y_t|\Theta_t, Y_{1:t-1})$ describes the *observation equation*. For an illustration purpose, considering a discrete-time multivariate linear Gaussian system, the state-space model is characterised by two linear equations:
State equation: the n-dimensional hidden state process Θ_t follows a first-order Markovian dynamics, as it only depends on the previous state at time $t-1$, and is corrupted by a (correlated or uncorrelated)

state noise process δ_t. The state equation is defined as

$$\Theta_t = B_t \Theta_{t-1} + \delta_t, \tag{3.19}$$

where B_t is the $n \times n$ state-transition matrix, also known as Markovian transition matrix, between current and prior states at time t and time $t-1$.
Observation equation: the m-dimensional measurement Y_t is subject to a linear transformation of the hidden state Θ_t, and is further corrupted by a measurement noise process ε_t

$$Y_t = Z_t \Theta_t + \varepsilon_t, \tag{3.20}$$

where Z_t is a $m \times n$ observation-transition matrix between the current observation and current state at time t. The noise processes ε_t and δ_t in the state-space model are both Gaussian with zero mean and respective covariance matrices V_t and Q, which can be shown as

$$\begin{bmatrix} \varepsilon_t \\ \delta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_t & 0 \\ 0 & Q \end{bmatrix} \right). \tag{3.21}$$

It is worth mentioning that Eq. (3.19) defines a first-order autoregressive process (see Section 3.4.1). A higher-order autoregressive structure can also be reformulated and transformed into a first-order autoregressive formulation by concatenating several state vectors into a new state vector (e.g., $\Theta_t^{new} = [\Theta_t, \Theta_{t-1}]$) (*Bernstein*, 2005).

The linear Gaussian state-space model can be extended to a broad class of dynamical Bayesian networks (*Ghahramani*, 1997) by changing one or more of the conditions about the state or measurement variables (*Chen et al.*, 2010b): **(i)** from continuous state to discrete or mixed-value state variable; **(ii)** from continuous observation to discrete or mixed observation; **(iii)** from Gaussian to non-Gaussian noise processes; **(iv)** inclusion of non-linearity in Eqs. (3.19) and (3.20). For instance, changing condition (i) may result in a discrete-time hidden Markov model or switching state-space model; changing condition (ii) or (iii) may result in a generalised state-space model with a Generalised Linear Model (GLM, *Nelder and Wedderburn*, 1972) in place of Eq. (3.20); and changing condition (iv) may result in a non-linear neural filter. In addition, a control variable can be incorporated into the state Eq. (3.19), which will result in a standard Linear Quadratic Gaussian (LQG) control system, for which the optimal solution can be derived analytically (*Bertsekas et al.*, 1995).

3.4.3 Benefits of Gaussian Process State-Space Models

In this thesis, we are interested in linear state-space models to define a dynamic relationship between observations and models within the proposed Bayesian data-model fusion techniques. In particular, we consider the case where a Gaussian process (*Rasmussen*, 2003) is responsible for modelling the underlying dynamics, which is widely known as the Gaussian process state-space model. Gaussian process state-space models are particularly appealing because they are defined generally and can accept flexible priors over the transition function. The presence of a latent state, within the state-space model, allows for a succinct representation of the dynamics in the form of a Markov chain. The state contains the information about the dynamic system (e.g., hydrological process within the Earth system), which is essential to determine future forecasts. We choose to build upon Gaussian processes for a number of reasons. First, they are non-parametric, which makes them effective in learning from small datasets. Second, we want to take advantage of the probabilistic properties of Gaussian processes. By using a Gaussian process for the latent transitions, we can get away with an approximate model and learn a distribution over functions. This allows us to account for model errors whilst quantifying uncertainty, as discussed and empirically shown by *Schneider* (1997) and *Deisenroth et al.* (2015). Consequently, the system will not become overconfident in regions of the space where data are scarce.

Moreover, a related advantage of state-space models over autoregressive ones is that to train an autoregressive model, the time series of observations Y is broken down into a set of input/output pairs, and the function that maps inputs to the outputs is obtained with regression techniques. One could use noise in the inputs regression (also known as errors invariables) to deal with observation noise. However, this would fail to exploit the fact that the particular noise realisation affecting the observation Y_t is the same when using Y_t as an input or output. When using a state-space model, the time series is not broken down into input/output pairs, and inference and estimating the unknown state parameters are performed in a way that the measurement noise is coherently taken into account as observation noise.

3.5 Kalman Filtering

A common objective of statistical inference for the state-space models is to infer the state parameters $\Theta = [\Theta_1, \Theta_2, \ldots, \Theta_t]$ (including its uncertainty) based on the time series of observations $Y = [Y_1, Y_2, \ldots, Y_t]$. In light of Eq. (3.17), the goal of the state-space analysis is to estimate the posterior probability distribution (or density) $p(\Theta|Y)$. In the special case of the linear Gaussian state-space model, the predictive posterior distribution is fully characterised by the conditional mean and conditional covariance of a Gaussian distribution. When the state equation (Eq. (3.19)) and the observation equation (Eq. (3.20)) are known, with known system matrices Z_t, B_t, V_t, Q, the optimal inference algorithm is described by a Kalman filtering (*Kalman*, 1960) (where Y_t is used for an online operation) or fixed-interval Kalman smoothing (*Bierman*, 1973) (where Y_t is used for an offline operation).

Kalman filtering (*Kalman*, 1960) is an algorithm to estimate the posterior probability distribution of the unknown state parameters in a joint probability distribution over the time series measurements and un-observed variables in each time-frame. Let us assume that within the state-space model defined by Eqs. (3.19) and (3.20), the system matrices $H_t = \{Z_t, B_t, V_t, Q\}$ are known, which are denoted here as H_t. The likelihood probability distribution (or the sample density function) associated with the state-space model for $t = 1, 2, \ldots, T$ is denoted by $p(Y_{1:T}|H_{1:T}, \Theta_{1:T})$. It should be noted here that for the rest of this thesis, $*_{1:t}$ denotes a time series for $t = 1, 2, \ldots, T$, e.g., $Y_{1:t} = [Y_1, Y_2, \ldots, Y_t]$. Considering the Bayes rule (e.g., Eq. (3.10)), the likelihood probability distribution over the entire data sets (i.e., for $t = 1, ..., T$) can be factorised as

$$\begin{aligned} p(Y_{1:T}|H_{1:T}, \Theta_{1:T}) = p(Y_1|H_1, \Theta_1)p(Y_2|Y_1, H_2, \Theta_2) \ldots p(Y_T|Y_{T-1}, H_T, \Theta_T) \\ = \Pi_{t=1}^{T} p(Y_t|Y_{t-1}, H_t, \Theta_t), \end{aligned} \tag{3.22}$$

where $Y_0 = 0$. Therefore, to construct the likelihood function $p(Y_{1:T}|H_{1:T}, \Theta_{1:T})$, we need to estimate the probability densities of $p(Y_t|Y_{t-1}, H_t, \Theta_t)$, $t = 1, 2, \ldots, T$. We can achieve this using filtering techniques, in particular when the system is linear, and errors are Gaussian. Kalman filter to solve Gaussian state-space models is a recursive procedure that involves 3 steps: (1) initialisation, (2) prediction, and (3) correction (*Kalman*, 1960). In the following, each of these steps will be discussed in greater detail. It should be noted here that in the following equations, $*_{t|s}$ denotes the prediction of the variable $*$ at time t, conditional upon information available at time s.

(1) Initialisation: The Kalman filter is initialised by deriving the best predictor of the initial state (*a priori* information for the unknown parameters when $t = 0$) $\Theta_{0|0}$, and an estimate of its covariance matrix, $\Sigma_{0|0}$.

(2) Prediction: $\Theta_{t|t-1}$ and $\Sigma_{t|t-1}$ are estimated for each time step t based on the state equation (Eq. (3.19)) as

$$\Theta_{t|t-1} = B_t \Theta_{t-1|t-1}, \tag{3.23}$$

$$\Sigma_{t|t-1} = B_t \Sigma_{t-1|t-1} B_t' + Q, \tag{3.24}$$

where Q is the covariance matrix of the state noise process δ_t (see Eq. 3.21). $\Theta_{t|t-1}$ is then used in the observation equation Eq. (3.20) to construct the forecast $Y_{t|t-1} = Z_t \Theta_{t|t-1}$. Having observation Y_t, the forecast error can be estimated as

$$u_t = Y_t - Y_{t|t-1} = Y_t - Z_t \Theta_{t|t-1} = \varepsilon_t + Z_t(\Theta_t - \Theta_{t|t-1}). \tag{3.25}$$

Because of the Gaussian errors assumption in state-space model, u_t is Gaussian, i.e.,

$$u_t \sim \mathcal{N}(0, V_t + Z_t \Sigma_{t|t-1} Z_t'). \tag{3.26}$$

According to the Eq. (3.25) $Y_t = u_t + Y_{t|t-1}$. Therefore, the probability distribution of Y_t conditional on Y_{t-1} is equal to the probability distribution of u_t which can be shown by

$$p(Y_t | Y_{t-1}, H_t, \Theta_t) = p(u_t | H_t, \Theta_t), \tag{3.27}$$

therefore, given $\Theta_{t|t-1}$ and $\Sigma_{t|t-1}$, the likelihood distribution $p(Y_t | Y_{1:t-1}, H_t, \Theta_t), t = 1, 2, \ldots, T$ can be computed from the normal density function Eq. (3.8) as

$$p(Y_t | Y_{1:t-1}, H_t, \Theta_t) = \frac{1}{\sqrt{(2\pi)^{n_y} |V_t + Z_t \Sigma_{t|t-1} Z_t'|}} exp(-\frac{u_t'(V_t + Z_t \Sigma_{t|t-1} Z_t')^{-1} u_t}{2}). \tag{3.28}$$

(3) Correction: Since the observed data Y_t is already known, the predictions $\Theta_{t|t-1}$ and $\Sigma_{t|t-1}$ can be updated (corrected) according to the *Kalman* (1960) formulation as

$$\Theta_{t|t} = \Theta_{t|t-1} + K_t(Y_t - Y_{t|t-1}) = \Theta_{t|t-1} + K_t(Y_t - Z_t \Theta_{t|t-1}), \tag{3.29}$$

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t(V_t + Z_t \Sigma_{t|t-1} Z_t') K_t', \tag{3.30}$$

where K_t is called *Kalman gain factor* and formulated as

$$K_t = \Sigma_{t|t-1} Z_t' (Z_t \Sigma_{t|t-1} Z_t' + V_t)^{-1}. \tag{3.31}$$

The intuition behind these formulae is relatively straightforward. The corrected prediction is a linear combination between the old prediction, $\Theta_{t|t-1}$, and the current prediction error $(Y_t - Y_{t|t-1})$. Given the linear form, K_t is chosen such that it minimises the prediction error variance. Kalman filter works recursively and requires only the last "best guess", rather than the entire history, of a system's state to calculate a new state. This means that the prediction step and the correction step are repeated at every time step, with the latest time step estimate, and its covariance informing the prediction used in the following iteration. The proof of these formulae is not provided here since they can be found in many literary works (see e.g., *Chui et al.*, 2017).

3.6 Bayesian Inference for State-Space Model

Bayesian inference (*Gelman et al.*, 2013) provides the opportunity to estimate the full posterior distribution of the state parameters within the state-space models and its uncertainty based on Bayesian statistics, especially when the state equation or the observation equation is unknown, i.e., the system matrices are unknown.

$$p(\Theta,X|Y) = p(\Theta|Y)p(X|Y) = \frac{p(Y|\Theta,X)p(\Theta)p(X)}{p(Y)} = \frac{p(Y|\Theta,X)p(\Theta)p(X)}{\int p(Y|\Theta,X)p(\Theta)p(X)dXd\theta}. \quad (3.32)$$

Equation (3.32) shows the joint posterior distribution ($p(\Theta,X|Y)$) of the state parameters Θ and the other unknown parameters (denoted by X) within the state-space model (Eqs. (3.19) and (3.20)) using the Bayes's rule. In this equation X can be use to denote the covariance matrix of the noise processes V_t and Q in Eq. (3.21), $p(Y|\Theta,X)$ shows the likelihood distribution, while $p(\Theta)$ and $p(X)$ denote the prior distributions for the Θ and X, respectively. The denominator of Eq. (3.32) is a normalising constant known as the partition function.

Ensemble Kalman Filter (EnKF, *Evensen*, 1994) and Particle Filtering (PF, *Gordon et al.*, 1993) or smoothing (*Doucet and Johansen*, 2009) are among popular algorithms that can be used to recursively update an estimate of the model states and produce corresponding innovation values given a sequence of observations in the state-space equation (Eqs. (3.19) and (3.20)). In theory, EnKF accomplishes this goal by linear projections, and the estimations in PF are performed through a Sequential Monte Carlo sampling. Comparing EnKF and PF, the latter includes a random element, converging to the true posterior probability function if the number of samples is vast. While the PF's strength is in its ability to account for both Gaussian and non-Gaussian error distributions, it suffers from the problem of dimensionality, which means that the sample size increases exponentially with the dimension of the state-space to achieve a certain performance. This fact precludes using PF in high-dimensional data-model fusion problems (*Bengtsson et al.*, 2008; *Daum and Huang*, 2003; *Snyder et al.*, 2008).

When accuracy is important, simulation-based (stochastic) methods, e.g., Markov Chain Monte Carlo (MCMC, *Geyer*, 2011) approaches such as Gibbs sampling (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993) offer an attractive alternative (see Sections 3.6.1 and 3.6.2) to produce a simulation sample (though not necessarily an independent one) from the (joint) posterior distributions of the unknown parameters in dynamic models, such as state-space models (see Eq. (3.32)). A simulation sample can be used to approximate almost any quantity relevant to Bayesian inference, including posterior expectations, variances, quantiles, and marginal densities. In other words, given a sample from the posterior of sufficient effective size, posterior expected values can be approximated by sample means, posterior quantiles by sample quantiles, posterior marginal densities by sample-based density estimates, and so forth.

In general, Monte Carlo-based Bayesian inference is powerful yet computationally expensive (*Geyer*, 1992, 2011; *Smith*, 2013). A trade-off between tractable computational complexity and good performance is to exploit various approximate Bayesian inference methods, such as expectation propagation (*Minka*, 2001), mean-field approximation (*Opper and Saad*, 2001) and variational approximation (*Beal et al.*, 2006; *Jordan et al.*, 1999). These techniques can also be integrated or combined to produce new methods, such as Monte Carlo Expectation Maximisation (EM) or motivational MCMC algorithms (*McLachlan and Krishnan*, 2007).

Since the approximations become more exact as more samples are used, accuracy tends to be limited only by the computational resources available. The best posterior inference is readily accomplished if an efficient method of sampling from the posterior is available. Independent sampling from the posterior is seemingly ideal since relatively few samples are required to obtain a good approximation in most cases, and the approximation error is relatively easy to characterise. However, such methods have proven difficult to implement in a general way that efficiently scales with the unknown parameters' dimension.

For example, rejection sampling (accept/reject) is efficient only if the posterior is tightly bounded by a known function proportional to an easy-to sample density (see e.g., *Casella et al.*, 2004). Finding such a function is generally difficult, and even adaptive variants struggle in high-dimensional situations. Currently, the most efficient, typically adaptable methods are using dependent sampling. Dependent sampling usually incurs a computational cost of acquiring a larger number of samples to attain a given accuracy. Still, the flexibility of these methods and their scalability to higher dimensions offset this disadvantage. During the past three decades, the category of Markov chain Monte Carlo (MCMC, *Geyer*, 2011) sampling approaches have been dramatically used in various studies. Unlike most classical methods, MCMC can often be efficiently automated, even for moderately complicated models. The underlying logic of MCMC sampling is that we can estimate any desired expectation by ergodic averages. That is, we can compute any statistic of a posterior distribution as long as we have N simulated samples from that distribution according to

$$E[f(s)]_p \approx \frac{1}{N}\Sigma_{i=1}^{N} f(s^{(i)}), \tag{3.33}$$

where p denotes the posterior distribution of interest, $E[f(s)]$ is the desired expectation, and $f(s^{(i)})$ is the i-th simulated sample from p. For example, we can estimate the mean of the unknown parameters Θ_t in the state-space Eqs. (3.19) and (3.20) by

$$E[\Theta_t]_p \approx \frac{1}{N}\Sigma_{i=1}^{N}\Theta_t^{(i)}. \tag{3.34}$$

3.6.1 Markov Chain Monte Carlo (MCMC) Sampling

MCMC is a sampling method, which follows a Bayesian inference to estimate the full posterior distribution of the parameters of interest, given a set of observations. MCMC sampling provides a class of algorithms for systematic random sampling from high-dimensional probability distributions, where each new sample depends on the existing samples, known as a Markov Chain. This property allows the algorithms to focus on the quantity that is being approximated from the distribution, even with many random variables (*Van Ravenzwaaij et al.*, 2018).

The benefit of the Monte Carlo approach is that calculating the mean of a large sample of numbers can be much more comfortable than calculating the mean directly from the normal distribution's equations. This benefit is most pronounced when random samples are easy to draw and when the distribution's equations are hard to work with in other ways. Unlike Monte Carlo sampling approaches that can draw independent samples from the distribution, Markov Chain Monte Carlo methods draw samples where the next sample is dependent on the existing sample, called a Markov Chain (*Geyer*, 2011).

The Markov chain property of MCMC (see Eq. (3.35)) is that a unique sequential process generates random samples. Each random sample is used as a stepping stone to generate the next random sample (hence the chain). It allows the algorithms to narrow in on the quantity that is being approximated from the distribution, even with a large number of random variables. A particular property of the chain is that, while each new sample depends on the one before it, new samples do not rely on any samples before the previous one (this is the "Markov" property).

MCMC is particularly useful in Bayesian inference because of the focus on posterior distributions which are often difficult to work with via analytic examination. In these cases, MCMC allows the user to approximate aspects of posterior distributions that cannot be directly calculated (e.g., random samples from the posterior, posterior means, etc.). Several Markov chain methods are available for sampling from a posterior distribution, where two important examples are the Gibbs sampler and the Metropolis-Hastings algorithm. The different MCMC algorithms differ in their performance with the speed and convergence depending on the model structure, which is briefly presented in what follows.

Markov Chain

Markov chains are an essential component of Markov chain Monte Carlo (MCMC) techniques. Under MCMC, the Markov chain is used to sample from some target distribution. A Markov chain is a stochastic process that operates sequentially (e.g., temporally), transitioning from one state to another within an allowed set of states.

$$\Theta_0 \to \Theta_1 \to \Theta_2 \to \cdots \to \Theta_t \to \dots \tag{3.35}$$

Markov chain is defined by three elements:
(I) A state-space Θ, which is a set of values that the chain is allowed to take.
(II) A transition operator $p(\Theta_t|\Theta_{t-1})$ that defines the probability of moving from state Θ_{t-1} to Θ_t.
(III) An initial condition distribution π_0 establishes the probability of being in any possible state at the initial iteration $i = 0$.
The Markov chain starts at some initial state, which is sampled from initial values, then transitions from one state to another according to the transition operator $p(\Theta_t|\Theta_{t-1})$.
A Markov chain is called *memoryless* if the next state only depends on the current state and not on any of the states previous to the current value. The mathematical definition of the Markov chain is

$$p(\Theta_t|\Theta_{t-1},\Theta_{t-2},\dots,\Theta_1) = p(\Theta_t|\Theta_{t-1}). \tag{3.36}$$

This memory-less property of a Markov chain is formally known as the Markov property in Bayesian inference. If the transition operator for a Markov chain does not change across transitions, the Markov chain is called *time-homogeneous*. A nice property of time-homogeneous Markov chains is that as the chain runs for a long time and $t \to \infty$, the chain will reach an equilibrium called the *chain's stationary distribution*.

$$p(\Theta_{t+1}|\Theta_t) = p(\Theta_t|\Theta_{t-1}). \tag{3.37}$$

If the state-space of a Markov chain takes on a finite number of distinct values, and it is time-homogeneous, then the transition operator can be defined by a matrix P, where the entries of P are

$$P_{ij} = p(\Theta_t = j|\Theta_{t-1} = i). \tag{3.38}$$

Equation (3.38) means that if the chain is currently in the i-th state, the transition operator assigns the probability of moving to the j-th state by the entries of i-th row of P (i.e., each row of P defines a conditional probability distribution on the state-space). A Markov chain can also have a continuous state-space that exists in the real numbers $\Theta \in \mathbb{R}^N$. In this case, the transition operator cannot be represented simply as a matrix but is instead some continuous function on the real numbers. Note that the continuous state-space Markov chain also has a stationary distribution. The stationary distribution of a Markov chain is important for sampling from probability distributions, a technique at the heart of the MCMC approaches.

3.6.2 Gibbs Sampling

Gibbs sampling is one of the most frequently used MCMC techniques to obtain samples from the posterior distribution. The idea in Gibbs sampling is to generate posterior samples by sweeping through each variable (or block of variables) to sample from its conditional distribution with the remaining variables fixed to their current values. For instance, consider the random variables $(X_1, X_2, \dots, X_D)$. Gibbs sampling is started by setting these variables to their initial values $X_1^{(0)}, X_2^{(0)}, \dots, X_D^{(0)}$ (often values sampled from their prior distribution $p(X_d^{(0)})$, for $d = 1, \dots, D$). At iteration (i), a Markov chain is generated by updating $X_1, X_2, \dots, X_D$ in turn by drawing from the full conditional distributions as

$$X_d^{(i)} \sim p(X_d^{(i)}|X_{-d}^{(i-1)}) \sim p(X_d^{(i)}|X_1^{(i)}, X_2^{(i)}, \dots, X_{d-1}^{(i)}, X_{d+1}^{(i-1)}, \dots, X_D^{(i-1)}), \quad d = 1, 2, \dots, D. \tag{3.39}$$

This process continues until convergence (the sample values have the same distribution as if they were sampled from the proper posterior joint distribution). Algorithm 1 details a generic Gibbs sampler.

Algorithm 1 Gibbs sampler

Initialise $X_1^{(0)}, X_2^{(0)}, X_3^{(0)}, \dots, X_D^{(0)}$

for iteration $i = 1, 2, \dots$ **do**

$X_1^{(i)} \sim p(X_1^{(i)}|X_2^{(i-1)}, X_3^{(i-1)}, \dots, X_D^{(i-1)})$

$X_2^{(i)} \sim p(X_2^{(i)}|X_1^{(i)}, X_3^{(i-1)}, \dots, X_D^{(i-1)})$

$\vdots$

$X_D^{(i)} \sim p(X_D^{(i)}|X_1^{(i)}, X_2^{(i)}, \dots, X_{D-1}^{(i)})$

In Algorithm 1, we are not directly sampling from the posterior distribution itself. Rather, we simulate samples by sweeping through all the posterior conditionals, one random variable at a time. Because we initialise the algorithm with random values, the samples simulated based on this algorithm at early iterations may not necessarily be representative of the actual posterior distribution. However, the theory of MCMC guarantees that the stationary distribution of the samples generated under Algorithm 1 is the target joint posterior that we are interested in (*Geyer*, 1992). For this reason, MCMC algorithms are typically run for a large number of iterations (in the hope that convergence to the target posterior will be achieved). Because samples from the early iterations are not from the target posterior, it is common to discard them. The discarded iterations are often referred to as the *burn-in* period (see also Section 3.6.5).

Essentially, Gibbs sampling reduces the problem of sampling $X := (X_1, X_2, ..., X_D)$ to the problem of conditionally sampling of variables X_d, for $d = 1, 2, ..., D$. Since X_d are of lower dimension (perhaps even one-dimensional), compared to X, they may be easier to sample by conventional methods. Moreover, it is often possible to choose a prior distribution such that many of the full conditionals are easy to sample. For example, when conditional priors are chosen from easily sampled families that are partially conjugate to the sampling model (see Section 3.1.2), the Gibbs sampler is easy to construct (see, e.g., *Gelman et al.*, 2013).

However, there are cases in which Gibbs sampling will be very inefficient. That is, the *mixing* of the Gibbs sampling chain might be prolonged, meaning that the algorithm may spend a long time exploring a local region with high density, and, thus, take very long to explore all areas with significant probability mass. For example, when the cross-correlation of the posterior conditional distributions between variables is high, successive samples become very highly correlated, and sample values change very slowly from one iteration to the next, resulting in chains that basically do not mix. The performance of the Gibbs sampling can sometimes be improved by modifying the algorithm. For example, the order in which the variables X_d are sampled can affect the mixing rate (e.g., *Roberts and Sahu*, 1997). Replacing some of the full conditional distributions with (partial) posterior marginals results in a partially collapsed Gibbs sampler (*Van Dyk and Park*, 2008). It may have better sampling properties, though it must be implemented carefully to preserve the stationary distribution (e.g., *Van Dyk and Jiao*, 2015).

3.6.3 Metropolis-Hastings

Metropolis-Hastings (*Chib and Greenberg*, 1995), another MCMC algorithm to simulate samples from a probability distribution of the full joint density function, relies on the (independent) proposal distributions of the variables of interest X_d, where $d = 1,2,...,D$. Algorithm 2 provides details of a generic Metropolis-Hastings algorithm where $X^{(i)} := (X_1^{(i)}, X_2^{(i)}, ..., X_D^{(i)})$.

Algorithm 2 Metropolis-Hastings algorithms

Initialise $X^{(0)}$, for $d = 1,2,\ldots,D$
for iteration $i = 1,2,...$ **do**
Propose: $X^{cand} \sim p(X^{(i)}|X^{(i-1)})$
Acceptance Probability: $\alpha(X^{cand}|X^{(i-1)}) = min\{1, \frac{p(X^{(i-1)}|X^{cand})p(X^{cand})}{p(X^{cand}|X^{(i-1)})p(X^{(i-1)})}\}$
$u \sim \text{Uniform}(u;0,1)$
if $u < \alpha$ **then**
Accept the proposal: $X^{(i)} \leftarrow X^{cand}$
else
Reject the proposal: $X^{(i)} \leftarrow X^{(i-1)}$
end if
end for

The first step is to initialise the sample value for each random variable (this value is often sampled from the variable's prior distribution). The main loop of Algorithm 2 consists of three components: (1) Generate a proposal (or a candidate) sample X^{cand} from the proposal distribution $p(X^{(i)}|X^{(i-1)})$; (2) Compute the acceptance probability via the acceptance function $\alpha(X^{cand}|X^{(i-1)})$ based upon the proposal distribution $p(X^{(i)}|X^{(i-1)})$ and the full joint density $p(X^{(i)})$; (3) Accept the candidate sample with the acceptance probability α, or reject it with probability $1-\alpha$.

Proposal Distribution

The Metropolis-Hastings algorithm starts with simulating a candidate sample X^{cand} from the proposal distribution $p(.)$. Note that samples from the proposal distribution are not accepted automatically as posterior samples. These candidate samples are accepted probabilistically based on the acceptance probability $\alpha(.)$. There are mainly two kinds of proposal distributions, symmetric and asymmetric. A proposal distribution is a symmetric distribution if

$$p(X^{(i)}|X^{(i-1)}) = p(X^{(i-1)}|X^{(i)}). \tag{3.40}$$

Clear choices of symmetric proposals include Gaussian distributions or Uniform distributions centred at the current state of the chain. For example, if we have a Gaussian proposal, then we have

$$\begin{aligned} X^{cand} &= X^{(i-1)} + \delta, \\ \delta &\sim \mathcal{N}(0, \sigma). \end{aligned} \tag{3.41}$$

Because the probability density function for normal distributions of $X^{cand} - X^{(i-1)}$ and $X^{(i-1)} - X^{cand}$ are equal to $\mathcal{N}(0, \sigma)$, this is a symmetric proposal. This proposal distribution randomly perturbs the current state of the chain, and then either accepts or rejects the perturbed value. The algorithms of this form are called the *Random-walk Metropolis algorithm* (*Chib and Greenberg*, 1995).

Random-walk Metropolis-Hastings algorithms are the most common Metropolis-Hastings algorithms. However, we may choose to (or need to) work with asymmetric proposal distributions in certain cases. For example, we may choose an inherently asymmetric proposal distribution, such as the log-normal density, which is skewed towards larger values. In other cases, we may need to work with asymmetric proposal distributions to accommodate particular constraints in our models. For example, if we wish to estimate the posterior distribution for a variance parameter, we require that our proposal does not generate values smaller than zero (*Haario et al.*, 1999).

Acceptance Function

Intuitively, the Metropolis-Hastings acceptance function is designed to strike a balance between the following two constraints:
(1) The sampler should tend to visit higher probability areas under the full joint density (this constraint is given by the ratio $\frac{p(X^{cand})}{p(X^{(i-1)})}$).
(2) The sampler should explore the space and avoid getting stuck at one site (e.g., the sampler can reverse its previous move in the space; this constraint is given by the ratio $\frac{p(X^{(i-1)}|X^{cand})}{p(X^{(i)}|X^{(i-1)})}$.
The Metropolis-Hastings acceptance function must have this particular form because this ensures that

the Metropolis-Hastings algorithm satisfies the condition of detailed balance, which guarantees that the stationary distribution of the Metropolis-Hastings algorithm is the target posterior that we are interested in (see *Geyer*, 2011; *Gilks*, 1996, , for more details).

Importantly, note that the acceptance function can be asymmetric (e.g., $\alpha(X^{(i)}|X^{(i-1)}) \neq \alpha(X^{(i-1)}|X^{(i)})$) irrespective of the proposal distribution. The acceptance function in the case of symmetric proposals, where $\alpha(X^{(i)}|X^{(i-1)}) = \alpha(X^{(i-1)}|X^{(i)})$ is driven as

$$\alpha(X^{cand}|X^{(i-1)}) = min\{1, \frac{p(X^{(i-1)}|X^{cand})p(X^{cand})}{p(X^{cand}|X^{(i-1)})p(X^{(i-1)})}\} = min\{1, \frac{p(X^{cand})}{p(X^{(i-1)})}\}, \tag{3.42}$$

where $p(X^{cand})$ and $p(X^{(i-1)})$ are the full joint density. This result is intuitive. When the proposal distribution is symmetric, the acceptance probability becomes proportional to how likely each current state $X^{(i-1)}$ and the proposal state X^{cand} are under the full joint density.

Accept/Reject a Proposal

Finally, we accept a given proposal with the acceptance probability α, which is the outcome of the acceptance function described above. The *min* operator in the acceptance function makes sure that the acceptance probability α is never larger than 1. Operationally, we draw a random number uniformly between 0 and 1, and if this value is smaller than α, we accept the proposal; otherwise, we reject it.

3.6.4 Gibbs Sampling Visa Versa Metropolis-Hastings

Gibbs sampling is a special case of Metropolis-Hastings sampling where the proposal distributions are the posterior conditionals. Recall that all proposals are accepted in Gibbs sampling, which implies that the acceptance probability is always 1. The algebra below shows that the acceptance function is equal to 1 for Gibbs sampling algorithm.

$$\begin{aligned} &\alpha(X_n^{cand}, X_{-n}^{(i-1)}|X_n^{(i-1)}, X_{-n}^{(i-1)}) \\ &= min\{1, \frac{p(X_n^{(i-1)}, X_{-n}^{(i-1)}|X_n^{cand}, X_{-n}^{(i-1)})p(X_n^{cand}, X_{-n}^{(i-1)})}{p(X_n^{cand}, X_{-n}^{(i-1)}|X_n^{(i-1)}, X_{-n}^{(i-1)})p(X_n^{(i-1)}, X_{-n}^{(i-1)})}\}, \end{aligned} \tag{3.43}$$

the proposal distributions for Gibbs sampling are the posterior conditionals distribution as

$$\begin{aligned} &p(X_n^{(i-1)}, X_{-n}^{(i-1)}|X_n^{cand}, X_{-n}^{(i-1)}) = p(X_n^{(i-1)}|X_{-n}^{(i-1)}) \\ &\qquad\qquad\text{and} \\ &p(X_n^{cand}, X_{-n}^{(i-1)}|X_n^{(i-1)}, X_{-n}^{(i-1)}) = p(X_n^{cand}|X_{-n}^{(i-1)}), \end{aligned} \tag{3.44}$$

here, the chain rule is considered, where the full joint distribution of $p(\circ,\star)$ is defined as $= p(\circ|\star)p(\star)$ (see Section 3.1.1). Therefore

$$\begin{gathered} p(X_n^{cand}, X_{-n}^{(i-1)}) = p(X_n^{cand}|X_{-n}^{(i-1)})p(X_{-n}^{(i-1)}), \\ \text{and} \\ p(X_n^{(i-1)}, X_{-n}^{(i-1)}) = p(X_n^{(i-1)}|X_{-n}^{(i-1)})p(X_{-n}^{(i-1)}). \end{gathered} \tag{3.45}$$

According to the Eqs. (3.44) and (3.45), Eq. (3.43) is rewritten as

$$\begin{aligned} &\alpha(X_n^{cand}, X_{-n}^{(i-1)}|X_n^{(i-1)}, X_{-n}^{(i-1)}) \\ &\quad = min\{1, \frac{p(X_n^{(i-1)}|X_{-n}^{(i-1)})p(X_n^{cand}|X_{-n}^{(i-1)})p(X_{-n}^{(i-1)})}{p(X_n^{cand}|X_{-n}^{(i-1)})p(X_n^{(i-1)}|X_{-n}^{(i-1)})p(X_{-n}^{(i-1)})}\} \\ &= 1. \end{aligned} \tag{3.46}$$

3.6.5 Convergence Diagnostics For MCMC

Two important issues that must be addressed while implementing MCMC are where to start and when to stop the algorithm. These two tasks are related to determining convergence of the underlying Markov chain to stationarity and convergence of Monte Carlo estimators to population quantities, respectively. It is known that under some standard conditions on the Markov chain, for any initial value, the distribution of $X_n^{(i)}$ converges to the stationary distribution (see, e.g., *Meyn and Tweedie*, 2012; *Robert and Casella*, 2013) as $i \to \infty$.

The prior distribution of X_n is not equal to the posterior distribution, known as target density π, and MCMC algorithms produce (serially) correlated samples. Therefore, the further the initial value from the posterior distribution, the longer it takes for X_n to approximate π. In particular, if the initial value is not in a high-density (π) region, the samples at the earlier iterations may not be close to the target distribution. In such cases, a common practice is to discard early realisations in the chain and start collecting samples only after the effect of the initial value has (practically) worn off. The main idea behind this method, known as burn-in, is to use samples only after the Markov chain gets sufficiently close to the stationary distribution. However, its usefulness for Monte Carlo estimation has been questioned in the MCMC community (*Geyer*, 2011). Thus, ideally, MCMC algorithms should be initialised at a high-density region, but if finding such areas is difficult, collection of Monte Carlo samples can be started only after a particular iteration M when approximately $X_n^{(M)} \sim \pi$. Once the starting value M is determined, one needs to decide when to stop the simulation. (Note that the starting value M here refers to the beginning of a collection of samples instead of the initial value of $X_n^{(0)}$ for the Markov chain, although these two values can be the same.) Often the quantities of interest regarding the target density π can be expressed as the means of certain functions defined as

$$E_\pi g \equiv \int_\chi g(\chi)\pi(\chi)d\chi, \tag{3.47}$$

where g is a real-valued function. For example, appropriate choices of g make $E_\pi g$ different measures of location, spread, and other summary features of π. Here, the support of the target density π is denoted by $\chi = \mathbb{R}^D$, for some $D \geq 1$, although it can be non-Euclidean as well. The MCMC estimator of the population mean $E_\pi g$ is the average sample $\bar{g}_{N,M}$ as

$$\bar{g}_{N,M} = \sum_{i=M+1}^{N} g(X_n^{(i)})/(N-M). \tag{3.48}$$

It is known that usually $\bar{g}_{N,M} \to E_\pi g$ as $N \to \infty$. In practice, however, MCMC users run the Markov chain for a finite N number of iterations, thus MCMC simulation should be stopped only when $\bar{g}_{N,M}$ has sufficiently converged to $E_\pi g$. The accuracy of the time average estimator $\bar{g}_{N,M}$ depends on the quality of the samples. Thus, when implementing MCMC methods, it is necessary to conclude the Markov chain convergence wisely and subsequently determine when to stop the simulation. By performing a theoretical analysis on the underlying Markov chain, an analytical upper bound on its distance to stationarity may be obtained (*Rosenthal*, 1995), which in turn can provide a rigorous method for deciding MCMC convergence and thus finding M (*Jones and Hobert*, 2001; *Jones et al.*, 2004). Similarly, using a sample size calculation based on an asymptotic distribution of the (appropriately scaled) Monte Carlo error $\bar{g}_{N,M} - E_\pi g$, an honest stopping value N can be found. In the absence of such theoretical analysis, often practical diagnostic tools are used to check the convergence of MCMC samplers and estimators, although these tools cannot determine convergence with certainty.

Since the early 1990s, with the increasing use of MCMC, much research effort have gone into developing convergence diagnostic tools. These diagnostic methods can be classified into several categories. For example, corresponding to the two types of convergence mentioned before, some of these diagnostic tools are designed to assess convergence of the Markov chain to the stationary distribution, whereas others check for convergence of the summary statistics like sample means and sample quantiles to the corresponding population quantities. The available MCMC diagnostic methods can be categorised according to other criteria as well. For example, (I) their level of theoretical foundation, if they are suitable for checking joint convergence of multiple variables; (II) whether they are based on multiple (parallel) chains or a single chain or both, (III) if they are complemented by a visualisation tool or not, (IV) if they are based on moments and quantiles or the kernel density of the observed chain, and so on. Several review articles on MCMC convergence diagnostics are available in the literature (see, e.g., *Brooks and Roberts*, 1998; *Cowles and Carlin*, 1996).

Cowles and Carlin (1996) describe 13 convergence diagnostics and summarise these according to the different criteria mentioned above. In this study, a simple graphical method suggested by *Brooks and*

Roberts (1998) and *Sinharay* (2003) is used to determine the number of iterations and to define the convergence of the sampling algorithm in the MCMC-DA, and ConBay-DA approaches, proposed in Chapters 5 and 6, respectively. Different simple graphical tools exist for convergence assessment, which examines the chain(s) of values generated for each parameter to determine if the simulation process stabilises in some sense. These tools are elementary and easy to implement and provide useful feedback about the convergence of the MCMC.

Trace Plots

Creating a trace-plot for each parameter in the model is the most popular check for convergence of an MCMC algorithm. The trace plot is a time series plot that shows the realisations of the Markov chain at each iteration against the iteration numbers. This graphical method is used to visualise how the Markov chain is moving around the state-space, that is, how well it is mixing. If the MCMC chain is stuck in some part of the state-space, or if there is a clear pattern (e.g., visible trends) in this plot, the MCMC algorithm may not have converged. It is often said that a good trace plot should look like a hairy caterpillar (Fig. 3.2). For an efficient MCMC algorithm, if the initial value is not

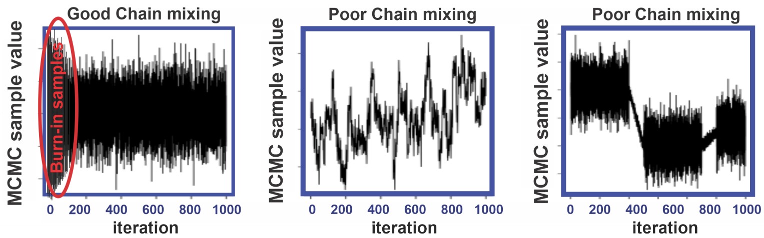

Fig. 3.2 An example of MCMC sample chain values after successive iterations. Poor chain mixing (characterized by any sort of pattern) suggests that the MCMC sampling chains may not have completely traversed all features of the posterior distribution and that more iterations are required to ensure the distribution has been accurately represented.

in the high-density region, the beginning of the trace plot shows back-to-back steps in one direction. In contrast, if the trace plot shows a similar pattern throughout, then there is no point in rejecting burn-in samples. If an MCMC algorithm consists of multiple chains, the trace plots often overrate the generated values on a typical graph for each parameter. MCMC chains not traversing the sample space, in the same way, provide evidence of lack of convergence. Some MCMC users monitor the log of posterior density (or a multiple of it) at the chain's current state. If the log of posterior density has an increasing trend, the chain has not reached the primary mode yet. If the log-density is going down, the chain probably started near a mode around, which there is little probability mass and will be more respectively part of the distribution.

Autocorrelation Functions Plots

Unlike independent and identically distributed (i.i.d) sampling, MCMC algorithms result in autocorrelated samples, which need a large number of iterations to traverse the whole sample space of the parameters. The lag-k (sample) autocorrelation is defined to be the correlation between the samples lag-k steps apart. The autocorrelation plot shows values of the lag-k autocorrelation function against increasing lag-k values. This tool is not strictly a convergence diagnostic tool but helps to interpret the trace plots and, thus, indirectly helps to assess convergence of the MCMC algorithm. For fast-mixing Markov chains, lag-k autocorrelation values drop down to (practically) zero quickly as k increases (see Fig. 3.3 first row plots). In contrast, high lag-k autocorrelation values for larger lag-k indicate a high degree of correlation and slow mixing of the Markov chain (see Fig. 3.3 second row plots). Generally, to get precise Monte Carlo estimates, Markov chains need to be run for many iterations, which takes the autocorrelation function to be practically zero. The autocorrelation plots on the right

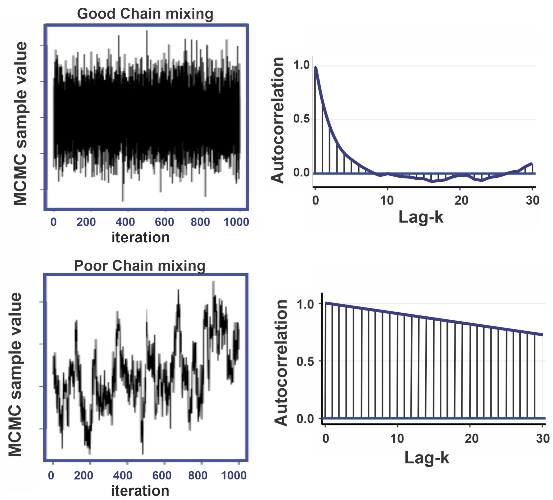

Fig. 3.3 An example of autocorrelation plots. The first graph demonstrates a sample of good chain mixing, which will converge to a stationary distribution (the posterior) quickly. The bottom graph demonstrates poor chain mixing, with slow convergence. The plots on the right show the corresponding autocorrelations. It can be seen that whereas the autocorrelation dies out for the first chain (at about the 10th lag, it remains high for poor chain mixing.

show the corresponding autocorrelations. You can see that whereas the autocorrelation dies out for the first chain pretty soon (at about the 10th lag), it remains high for the other two cases.

Running Mean Plots

Another graphical method used in practice is the running mean plot, although its use has faced criticism (*Geyer*, 2011). The running mean plot shows the mean (time average) of all Monte Carlo estimates against the iterations. Usually, running means are plotted at every k-th iteration (e.g., k=50). This line plot should stabilise to a fixed number as iteration increases, but non-convergence of the plot indicates that the simulation cannot be stopped yet. While the trace plot is used to diagnose a Markov chain's convergence to stationarity, the running mean plot is used to decide stopping times. Figure 3.4 shows an example of the running mean plot for 3 different chains. The black curve shows that the convergence is slow, while the blue curve shows a faster convergence, but still slower than the red curve. In the multivariate case, individual trace, autocorrelation, and running mean plots are generally

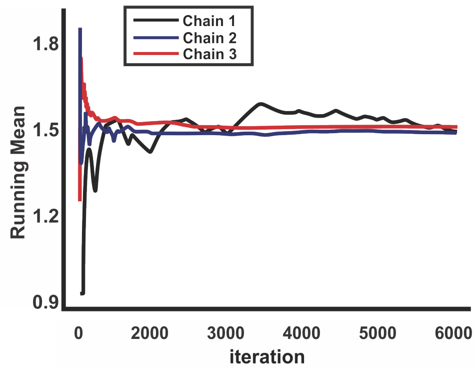

Fig. 3.4 An example time series of the running chain mean as used to check whether the chain is slowly or quickly approaching its target distribution.

made based on realisations of each marginal chain. Thus, the correlations that may be present among different components are not visualised through these plots. In multivariate settings, investigating correlation across different variables is required to check for the presence of high cross-correlation (*Cowles and Carlin*, 1996).

3.7 Bayesian Model Averaging

In practice, multiple models provide descriptions of the distributions generating the observed data Y. It is a standard statistical practice that, in such situations, a best model must be selected according to some criteria, like model fit to the observed data set, predictive capabilities or likelihood penalisations, such as information criteria. After making the selection, all inference is made, and conclusions are drawn assuming the selected model as the right model. However, there are downsides to this approach. The selection of one particular model may lead to overconfident inferences and riskier decision making since it ignores the existent model uncertainty favouring very specific distributions and assumptions on the model of choice. Therefore, modelling this source of uncertainty to select or combine multiple models appropriately is very desirable. Using Bayesian inference for this purpose has been suggested as a framework capable of achieving these goals. Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999) is an extension of the usual Bayesian inference methods, in which one not only models parameter uncertainty through the prior distribution but also model uncertainty obtaining posterior parameter and model posteriors using Bayes' theorem and therefore allowing for direct model selection, combined estimation and prediction.

Let each model is denoted by $Z^{(m)}$, where $m = 1, 2, ..., M$ is used to identify the model integrated into the observation at time t. Considering Eq. (3.11), one then obtain the posterior probability distribution of the unknown parameters $\Theta^{(m)}$, using Bayes' theorem (see Eq. (3.2)) as

$$p(\Theta^{(m)}|Y, Z^{(m)}) = \frac{p(Y|Z^{(m)}, \Theta^{(m)})p(\Theta^{(m)}|Z^{(m)})}{\int p(Y|Z^{(m)}, \Theta^{(m)})p(\Theta^{(m)}|Z^{(m)})d\Theta^{(m)}}, \tag{3.49}$$

where the integral in the denominator is calculated over the support set for each prior distribution and represents the marginal distribution of the data set for all parameter values specified in model m. This quantity is called the model's marginal likelihood (see also Eq. (3.12)) or model evidence and is denoted by

$$p(Y|Z^{(m)}) = \int p(Y|Z^{(m)}, \Theta^{(m)})p(\Theta^{(m)}|Z^{(m)})d\Theta^{(m)}. \tag{3.50}$$

BMA then adds a layer to hierarchical modelling present in Bayesian inference by assuming a prior distribution over the set of all considered models to determine the prior uncertainty over each model's capability to describe the data accurately. If there is a probability mass function over all the models with values $Z^{(m)}$ for $m = 1, 2, ..., M$, then Bayes' theorem can be used to derive posterior model probabilities given the observed data by

$$p(Z^{(m)}|Y) = \frac{p(Y|Z^{(m)})p(Z^{(m)})}{\sum_{i=1}^{M} p(Y|Z^{(i)})p(Z^{(i)})}, \tag{3.51}$$

which is a straightforward posterior model probability (*Raftery et al.*, 1997), representing the backing of each considered model by the observed data. There is also a link between these posterior model probabilities and the use of Bayes Factors. Given two models i and j, the Bayes factor of model i against model j is given by

$$BF_{i,j} = \frac{p(Z^{(i)}|Y)}{p(Z^{(j)}|Y)}, \tag{3.52}$$

thus, quantifying the relative strength of the evidence in favour of model i against that of model j. Given a baseline model, which can be arbitrarily fixed as model 1, it is clear that Eq. (3.51) can be written in terms of Bayes Factors by simply dividing by the baseline model's evidence, resulting in

$$p(Z^{(m)}|Y) = \frac{BF_{m,1}p(Z^{(m)})}{\sum_{i=1}^{M} BF_{i,1}p(Z^{(i)})}, \tag{3.53}$$

which means that one can estimate the posterior model probabilities by using Bayes Factors estimations and vice versa. These model probabilities can mainly be used for two purposes. First, the posterior probabilities can be used as straightforward model selection criteria, with the most likely model being selected. Second, consider a quantity of interest present in all models, e.g., Θ, it follows that its marginal posterior distribution across all models is given by

$$p(\Theta|Y) = \sum_{m=1}^{M} p(\Theta^{(m)}|Y,Z^{(m)})p(Z^{(m)}|Y), \tag{3.54}$$

Equation (3.54) shows the weighted average of all posterior distributions based on the posterior model probabilities estimated by Eq. (3.51). Therefore, BMA allows for a direct combination of models to obtain combined parameter estimates or predictions (*Roberts*, 1965). However, the implementation and application of BMA are not easy due to the difficulties in estimation of the model evidence, known as the marginal likelihood (Eq. (3.50)), which is non-trivial in most applications. The integral can be approximated by Monte Carlo methods or its extensions such as Importance Sampling (*Tokdar and Kass*, 2010). Namely, given a sampling weight $w(\Theta)$, defined over the same integration domain of Eq. (3.50), the evidence can be approximated by taking N random samples from the probability distribution determined by $w(\Theta)$ and computing the weighted average as

$$p(Y|Z^{(m)}) = \sum_{n=1}^{N} \frac{p(Y|\Theta_n^{(m)},Z^{(m)})p(\Theta_n^{(m)}|Z^{(m)})}{w(\Theta_n^{(m)})}, \tag{3.55}$$

The ordinary Monte Carlo approximation can be performed by using the parameter prior $p(\Theta)$ as a sampling weight (*Gelfand and Smith*, 1990). Markov Chain Monte Carlo (MCMC) methods can also be used to approximate the marginal likelihood by sampling $\Theta_1^{(m)}, \Theta_2^{(m)}, ..., \Theta_N^{(m)}$ from an MCMC chain (*Gelfand and Dey*, 1994). The estimator defined by Eq. (3.55) is guaranteed to converge to the evidence when the sample size increases. Another benefit of the MCMC comes through the use of trans-dimensional Markov Chain methods like the Reversible Jump MCMC (*Green*, 1995) or stochastic searches through the model space like variable selection through stochastic search (SSVS, *George and McCulloch*, 1997) employed in regression models. When using MCMC samples, the quality of the approximation is not guaranteed, and there are more sophisticated results ensuring its good behaviour (see *Robert and Casella* (2013) for a complete treatment).

Chapter 4

Dynamic Model Data Averaging (DMDA)

In this chapter, Dynamic Model Data Averaging (DMDA, i.e., a modified version of the Dynamic Model Averaging (DMA) approach presented by *Raftery et al.*, 2010) is formulated to merge multiple *a priori* information, derived from multi-model water storage simulations, with a set of observation, derived from GRACE(-FO) TWSC, while considering the uncertainty of all data sets (Fig. 4.1 summarises the DMDA method implemented in this study). In summary, DMDA is based on the Bayes theory and provides time-variable weights to compute an average of multiple *a priori* information, yielding the best fit to the observations. These weights can then be used to separate the components of the observation and modify the estimation of *a priori* information (e.g., surface and sub-surface water storage changes derived from hydrological models). Therefore, the DMDA water storage estimates in this study are expected to be more realistic than those of individual models. In this chapter, it will be shown that the implementation of DMDA combines the benefits of state-space merging techniques, such as the Kalman filtering (*Kalman*, 1960) or Particle Filtering (PF, *Gordon et al.*, 1993), Markov Chain (MC, *Chan and Geyer*, 1994; *Kuczera and Parent*, 1998; *Metropolis et al.*, 1953), and Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999). The proposed approach is able to deal with various observations and models with different structures, and can be applied in data assimilation applications that work with only one model, e.g., (*Girotto et al.*, 2016; *Khaki et al.*, 2017b,c; *Schumacher et al.*, 2018a), as well as in handling multiple model outputs as in *Van Dijk et al.* (2014).

In this chapter, a dynamic state-space model to define a linear relationship between a set of observation (GRACE(-FO) TWSC) and *a priori* information (hydrological model outputs) is described in Section 4.1. The Kalman filter approach to recursively estimate the unknown state parameters and their uncertainties is formulated in Section 4.2, while the Bayesian averaging of the Kalman filter estimates is presented in Section 4.3. Updating *a priori* information and estimating their uncertainties is described in Section 4.4. Finally, the performance of DMDA is tested in a controlled synthetic simulation in Section 4.5, where the results of the Bayesian update are known by definition.

4.1 Linear State-Space Model

Linear relationships between observations and *a priori* information in a dynamic system can be represented by a linear state-space model as

$$Y_t = Z_t \Theta_t + \varepsilon_t, \tag{4.1}$$

$$\Theta_{t+1} = \Theta_t + \delta_t. \tag{4.2}$$

Equation (4.1) is known as the 'observation equation', while Eq. (4.2) is known as the 'state equation' according to *Bernstein* (2005) (see also Eqs. (3.19) and (3.20)). In Eq. (4.1) the observation vector (e.g., GRACE(-FO) TWSC) for P spatial grid points at time $t = 1,2,...,T$ is shown by $Y_t = [y_1, y_2, \dots, y_P]_t$, and *a priori* information (e.g., the water storage components derived from a hydrological model) is stored in a $P \times P$ diagonal matrix Z_t, where the diagonal elements contain $[z_{p,1}, z_{p,2}, \dots, z_{p,K}]_t$ corresponding to each spatial grid point $p = 1,2,...,P$ (see also Eq. (4.3)). The number of parameters in *a priori* information is denoted by $k = 1,...,K$. For hydrological applications K shows the number of individual water storage compartments derived from hydrological model outputs, such as canopy, snow, surface water, soil water, and groundwater storage.

$$Z_t = \begin{bmatrix} [z_{1,1}, z_{1,2}, \dots, z_{1,K}] & 0 & \cdots & 0 \\ 0 & [z_{2,1}, z_{2,2}, \dots, z_{2,K}] & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & [z_{p,1}, z_{P,2}, \dots, z_{P,K}] \end{bmatrix}_t, \quad \Theta_t = \begin{bmatrix} \begin{bmatrix} \theta_{1,1} \\ \theta_{2,1} \\ \vdots \\ \theta_{K,1} \end{bmatrix} \\ \vdots \\ \begin{bmatrix} \theta_{1,P} \\ \theta_{2,P} \\ \vdots \\ \theta_{K,P} \end{bmatrix} \end{bmatrix}_t \tag{4.3}$$

In Eq. (4.1), each element of $\Theta_{t\,(P\times 1)}$ contains the unknown state parameters $[\theta_{1,p}, \theta_{2,p}, \dots, \theta_{K,p}]_t$ (see also Eq. (4.3))., which make a linear relation between the observation Y_t and *a priori* information Z_t.

In an ideal situation, where the observation (GRACE(-FO) TWSC) is equal to *a priori* information (the summation of individual water storage components of the model), the unknown state parameters Θ_t are equal to 1. However, for hydrological applications, it is expected that the model estimates contain errors due to the imperfect structure of the hydrological model and uncertainties in the input forcing data that are used to run the model simulations. Besides, GRACE(-FO) TWSC estimates contain errors as described, e.g., in *Forootan et al.* (2014a). As a result, the state parameters Θ_t are unknown, and are allowed to evolve in time according to the state equation Eq. (4.2).

In Eqs. (4.1) and (4.2), ε_t and δ_t are interpreted as the residual of the observation equation (Eq. (4.1)) and the state equation (Eq. (4.2)), respectively. The residual of the observation equation, i.e., ε_t, is assumed to be Gaussian distributed with the mean value of zero and the error covariance matrix of V_t, which can vary in time. The benefit of the Gaussian process state-space model is discussed in Section 3.4.3. The state residual δ_t is assumed to be stationary and is Gaussian distributed. It is also independent from ε_t, with the mean value of zero and its error covariance matrix Q (more details about the Gaussian process state-space model are provided in Section 3.4.2 of Chapter 3). Thus, the distribution of ε_t and δ_t can be written as

$$\begin{bmatrix} \varepsilon_t \\ \delta_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_t & 0 \\ 0 & Q \end{bmatrix}\right). \tag{4.4}$$

Uncertainty of the observation is reflected in V_t. The error covariance matrix of δ_t (Q in Eq. (4.4)) defines the temporal dependency between water storage changes at each time point to the previous time step. Equations (4.1) and (4.2) are formulated with the main assumption that there is little physical knowledge about how the defined regression model and its parameters are likely to evolve in time. In hydrological applications, since changes in water storage compartments depend on the history of hydrological processes, accounting for temporal dependency between water states is desirable. However, in practice, there might be no information about the temporal relationship between observations and *a priori* information (e.g., between GRACE(-FO) TWSC and hydrological model outputs). Therefore, Q and consequently δ_t are unknown. However, it will be shown that by introducing two parameters λ and α, which are referred to as 'Kalman forgetting factors' and 'BMA forgetting factors', respectively, one can control the temporal dependency of the DMDA solutions. These two parameters provide the opportunity to treat *a priori* information and observations of each step temporally as dependent on or independent of previous steps.

It is worth mentioning here that for linear and Gaussian-type state-space models, as presented in this chapter, the PF method will yield the same likelihood as EnKF when the number of simulations is large enough (this has been tested but the results are not shown to keep the focus of this study on presenting the DMDA). Therefore, the DMDA, which combines the benefits of the EnKF and is mathematically rigorous as the PF, is adopted in this thesis. More details of the EnKF and PF are provided in Section 3.6 of Chapter 3.

4.2 Kalman Filter to Estimate State-Space Parameters

To solve the state-space model by the Kalman filter approach, the first step is to initialize the unknown state parameters by choosing a prior value for $\Theta_{0|0}$ and its error covariance matrix $\Sigma_{0|0}$ when $t = 0$ (see Section 3.5 of Chapter 3). The numerical examples of prior values are discussed in Section 5.1.1 5. For the following equations, $Y_{1:t}$ represents the vector of observations up to the time step t, i.e., $[Y_1, ..., Y_t]$, and $\circ_{t|s}$ denotes the prediction of the variable $\circ$ at time t, conditional upon information available at time s.

In the Kalman filter solution of the state-space model, since the observation Y_t and the residual outputs ε_t are assumed to be Gaussian distributed, the state parameters $\Theta_{t|t}$ are obtained from a Gaussian posterior distribution with the mean value of $\hat{\Theta}_{t|t}$ and the variance $\hat{\Sigma}_{t|t}$. Accordingly, at time $t-1$, the posterior distribution of state parameters are assumed to be normal with the mean value of $\hat{\Theta}_{t-1|t-1}$ and the covariance $\hat{\Sigma}_{t-1|t-1}$, defined as

$$p(\Theta_{t-1}|Y_{1:t-1}) \sim N(\hat{\Theta}_{t-1|t-1}, \hat{\Sigma}_{t-1|t-1}). \tag{4.5}$$

In the next step of the Kalman filtering, the state parameter at time t, i.e., Θ_t, are predicted by introducing $\delta_t \sim \mathcal{N}(0, Q)$ to the state equation (Eq. (4.2)). The posterior probability distribution of predicated Θ_t conditional on the observed $Y_{1:t-1}$ has a normal distribution with the mean value of $\Theta_{t|t-1}$ and the covariance matrix of $\Sigma_{t|t-1}$ shown by Eq. (4.6) as

$$p(\Theta_t|Y_{1:t-1}) \sim N(\Theta_{t|t-1}, \Sigma_{t|t-1}), \tag{4.6}$$

where

$$\begin{aligned} \Theta_{t|t-1} &= \hat{\Theta}_{t-1|t-1}, \\ \Sigma_{t|t-1} &= \hat{\Sigma}_{t-1|t-1} + Q. \end{aligned} \tag{4.7}$$

In Eq. (4.7) which is derived from the state equation (Eq. (4.2)), Q is unknown (the error covariance matrix of δ_t in Eq. (4.2)). Therefore, to mathematically define a temporal dependency, $\Sigma_{t|t-1}$ in Eq. (4.6) can be replaced by

$$\Sigma_{t|t-1} = \lambda^{-1}\hat{\Sigma}_{t-1|t-1}, \tag{4.8}$$

where $0 < \lambda \leq 1$ controls the influence of previous observations on the regression value at time t and is known as 'Kalman forgetting factor' in the DMDA approach (see, e.g., *Fagin*, 1964; *Jazwinski*, 2007). Therefore, in the state-space model, the covariance matrix of the state innovation (Q) is assumed to be equal to $(\lambda^{-1} - 1)\Sigma_{t-1|t-1}$.

Hannan et al. (1989) indicated that in the recursive estimation of auto-regressive models, i.e., the state equation, the covariance of previous steps is derived as a weighted product of the current step (i.e., weighted by λ^{-1} in Eq. (4.8)). By this assumption, the effective window size of temporal dependency is estimated by $1/(1-\lambda)$. In the application of DMDA, where six global models are merged with GRACE(-FO) TWSC data, λ is chosen to be 0.95, which means that for monthly data, the effective window size is equivalent to 18 months. This value is experimentally selected as a value to minimise the Root Mean Square of Differences (RMSD) between the observed and modelled TWSC.

DMDA is formulated to update multiple sources of *a priori* information about a dynamic system, using a set of observations. Therefore, each set of *a priori* information is denoted by $Z_t^{(m)}$, for the rest of the equations in this chapter, where $m = 1, 2, \ldots, M$ is used to identify each model used in DMDA approach. Therefore, the unknown sate parameters Θ_t and its covariance matrix of Σ_t derived from each set of *a priori* information m are denoted by $\Theta_t^{(m)}$ and $\Sigma_t^{(m)}$, respectively. It is worth mentioning that available models, used as *a priori* information, might have different structures, e.g., the number of soil layers varies between models. Therefore, the length of the state vector Θ_t and the size of observation-transition matrix Z_t change from one model to another. These differences can be handled by DMDA.

To apply DMDA and update *a priori* information (e.g., water storage simulated by $m = 1, 2, \ldots, M$ different models), the parameter prediction of Eq. (4.6) is then extended as

$$p(\Theta_t^{(m)}|Y_{1:t-1}) \sim N(\Theta_{t|t-1}^{(m)}, \lambda^{-1}\hat{\Sigma}_{t-1|t-1}^{(m)}), \qquad m = 1, ..., M \tag{4.9}$$

After this prediction, the next step of the Kalman filter is to update (correct) parameters $\Theta_{t|t-1}$ and their uncertainty $\Sigma_{t|t-1}$ conditional on new observation Y_t (see also Eqs. (3.29), (3.30), and (3.31)), where the posterior distribution of the updated value is shown by

$$p(\Theta_t^{(m)}|Y_{1:t}) \sim N(\hat{\Theta}_{t|t}^{(m)}, \hat{\Sigma}_{t|t}^{(m)}), \tag{4.10}$$

where

$$\hat{\Theta}_{t|t}^{(m)} = \Theta_{t|t-1}^{(m)} + \Sigma_{t|t-1}^{(m)} Z_t^{(m)} (V_t + Z_t^{(m)} (\Sigma_{t|t-1}^{(m)} + W_t^{(m)}) Z_t^{(m)\prime})' (Y_t - Z_t^{(m)} \Theta_{t|t-1}^{(m)}), \tag{4.11}$$

$$\hat{\Sigma}^{(m)}_{t|t} = \Theta^{(m)}_{t|t-1} - \Sigma^{(m)}_{t|t-1} Z^{(m)\prime}_t (V_t + Z^{(m)}_t (\Sigma^{(m)}_{t|t-1} + W^{(m)}_t) Z^{(m)\prime}_t)^{-1} Z^{(m)}_t \Sigma^{(m)}_{t|t-1}. \tag{4.12}$$

In Eqs. (4.11) and (4.12), V_t is the covariance matrix of observation, and W_t is the covariance matrix of predictors Z_t (see Eq. (4.1)). In Eq. (4.11) and in the rest of this chapter, $\circ'$ denotes the transpose matrix of $\circ$.

4.3 Bayesian Averaging of the Kalman Filter Estimates

It is evident from Eqs. (4.11) and (4.12) that the estimation of state parameter $\hat{\Theta}_{t|t}$ is conditional on a particular model. Therefore, the DMDA solution to obtain unconditional results and update multi-model simulations involves calculating the posterior model probability $p(Z^{(m)}_t|Y_{1:t})$ as a weight for each model, which changes at each time step. In the following, it will be shown that time-variable weights need to be computed for each model m by choosing a BMA forgetting factor α in a recursive method. These weights are then used to dynamically average a priori values (the models), which leads to the best fit to observations. This justifies the term 'Dynamic' in the DMDA and makes the method different from other averaging techniques such as the Bayesian Model Averaging (BMA).

According to the Bayes' theorem defined by Eq. (3.2) (Chapter 3), the posterior model probability for each model m at time t can be estimated as

$$p(Z^{(m)}_t|Y_{1:t}) = \frac{p(Y_t|Z^{(m)}_t, Y_{1:t-1}) p(Z^{(m)}_t|Y_{1:t-1})}{\sum_{i=1}^{M} p(Y_t|Z^{(i)}_t, Y_{1:t-1}) p(Z^{(i)}_t|Y_{1:t})}, \tag{4.13}$$

where, $p(Y_t|Z^{(m)}_t, Y_{1:t-1})$ is the density of the observation, known as likelihood distribution, at time t, conditional on the *a priori* information derived from model m and the observation values up to time $t-1$, i.e., $Y_{1:t-1} = [Y_1, Y_2, ..., Y_{t-1}]$, which is estimated by a normal distribution as

$$p(Y_t|Z^{(m)}_t, Y_{1:t-1}) \sim N(Z^{(m)}_t \Theta^{(m)}_{t|t-1}, V_t + Z^{(m)}_t (\Sigma^{(m)}_{t|t-1} + W^{(m)}_t) Z^{(m)\prime}_t), \tag{4.14}$$

and, $p(Z^{(m)}_t|Y_{1:t-1})$ is the model prediction equation, which is defined by

$$p(Z^{(m)}_t|Y_{1:t-1}) = \Sigma_{i=1}^{M} p(Z^{(m)}_{t-1}|Y_{1:t-1}) a_{mi}. \tag{4.15}$$

In Eq. (4.14), $\Theta^{(m)}_{t|t-1}$ and $\Sigma^{(m)}_{t|t-1}$ are the predicted values derived from Eq. (4.7). In Eq. (4.15) $a_{mi} = P(Z^{(i)}_t|Z^{(m)}_{t-1})$ is the element of the $M \times M$ transition matrix $A(a_{mi})$ between models, which can be onerous when the number of models is large, e.g., for M models and τ time steps, the number of combinations of models will be $M^{2\tau}$. In the numerical application of DMDA in Chapter 7, the number of hydrological models is 6, therefore $M = 6$, and 122-time steps over the entire period of the study (2002–2012), which leads to 6^{244} combinations of models. To specify the transition matrix A, the implicit specification of the transition matrix is avoided using the BMA forgetting factor of $0 < \alpha < 1$, which has the same role as λ in Eq. (4.8). As a result, the model prediction equation (Eq. (4.15)) can be rewritten as

$$p(Z^{(m)}_t|Y_{1:t-1}) = \frac{p(Z^{(m)}_{t-1}|Y_{1:t-1})^{\alpha}}{\sum_{i=1}^{M} p(Z^{(i)}_{t-1}|Y_{1:t-1})^{\alpha}}. \tag{4.16}$$

The posterior model probability, or weights, for each model at time t, is estimated in a recursive solution between Eqs. (4.13), (4.14), and (4.16). This process is initialized by setting $p(Z^{(m)}_0|Y_0) = \frac{1}{M}$ for $m = 1,...,M$, and assigning an initial values for the state parameters $\Theta^{(m)}_0 \sim N(0,\Sigma^{(m)}_0)$ where $\Sigma^{(m)}_0$ =Variance (Y_t)/Variance $(Z^{(m)}_t)$. The reason for choosing this prior value is that in linear regression, a regression coefficient for a predictor Z_t is likely to be less than the standard deviation of the observations Y_t divided by the standard deviation of predictors Z_t (for more information see *Raftery*, 1993). In this study, for the application part of DMDA (Chapter 7) to merge six global hydrological models with GRACE TWSC, the optimum regression estimates are found when $0.85 < \alpha < 0.9$. By this choice, the RMSD between the DMDA TWSC and those of GRACE was found to be minimum. A BMA forgetting factor $\alpha = 0.9$ corresponds to a temporal smoothing window with the length of 36 time steps. It means that the contribution of *a priori* information at time $t-37$ into the posterior model probability of each model m at time t is negligible. The length of this smoothing window is reduced, e.g., to 8 months if $\alpha = 0.2$.

4.4 Updating *a priori* information and their Uncertainties by DMDA

The multi-model predictions of $\hat{Y}^{DMDA}_t$ is a weighted average of model-specific prediction $\hat{Y}^{(m)}_t$, using the posterior model probabilities, $p(Z^{(m)}_t|Y_{1:t})$, as its weights, i.e.,

$$\hat{Y}^{DMDA}_t = \sum_{m=1}^{M} p(Z^{(m)}_t|Y_{1:t})\hat{Y}^{(m)}_t, \tag{4.17}$$

where $\hat{Y}_t^{(m)} = Z_t^{(m)} \hat{\Theta}_{t|t}^{(m)}$. The posterior model probability for each model at time t, along with the estimated time-variable state parameter $\hat{\Theta}_{t|t}^{(m)} = [\hat{\theta}_{1,p}^{(m)}, \ldots, \hat{\theta}_{K,p}^{(m)}]_t$ from Kalman filter-type updating Eq. (4.11) are used to update the *a priori* information, which can be water storage components in hydrological application, as

$$\hat{z}_{k,p,t}^{DMDA} = \sum_{m=1}^{M} p(Z_t^{(m)}|Y_{1:t}) z_{k,p,t}^{(m)} \hat{\theta}_{k,p,t}^{(m)}, \tag{4.18}$$

where $k = 1, 2, \ldots, K$ are used to identify each of the water storage components in hydrological applications, i.e., groundwater, soil water, surface water, canopy, and snow. To update the water storage simulations of a single-model using the GRACE(-FO) TWSC and the DMDA approach, M needs to be set to 1, and the prediction step is limited to the conditional estimation of the parameter $\Theta_t^{(m)}|Z_t^{(m)}$ using Eq. (4.11).

The posterior model probability can also be used to estimate the unconditional probability distribution of state parameters $\Theta_{t|t}^{(1:M)} = (\Theta_{t|t}^{(1)}, ..., \Theta_{t|t}^{(M)})$ given by observation Y_t as shows by

$$p(\Theta_t^{(1:M)}|Y_{1:t}) = \sum_{m=1}^{M} p(\Theta_t^{(m)}|Z_t^{(m)}, Y_{1:t}) P(Z_t^{(m)}|Y_{1:t}), \tag{4.19}$$

where $p(\Theta_t^{(m)}|Z_t^{(m)}, Y_{1:t})$ shows the conditional distribution of $\Theta_{t|t}^{(m)}$, which is approximated with a normal distribution as

$$p(\Theta_t^{(m)}|Z_t^{(m)}, Y_{1:t}) \sim N(\hat{\Theta}_{t|t}^{(m)}, \hat{\Sigma}_{t|t}^{(m)}). \tag{4.20}$$

The DMDA approach can be reduced to a standard Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999) when $\alpha = \lambda = 1$ (see also Section 3.7). Then the posterior model probability of model m is given by

$$P(Z_t^{(m)}|Y_{1:t}) = \frac{p(Y_{1:t}|Z_t^{(m)})}{\sum_{i=1}^{M} p(Y_{1:t}|Z_t^{(i)})}, \tag{4.21}$$

where $p(Y_{1:t}|Z_t^{(m)})$ is the marginal likelihood, obtained by integrating the product of the likelihood, $P(Y_{1:t}|\Theta_t^{(m)}, Z_t^{(m)})$, over the parameter of interest, $P(\Theta_t^{(m)})$, according to Eq. (3.2) in Chapter 3 (see also *Hsu et al.*, 2009). Figure 4.1 summarises the work-flow of the DMDA approach.

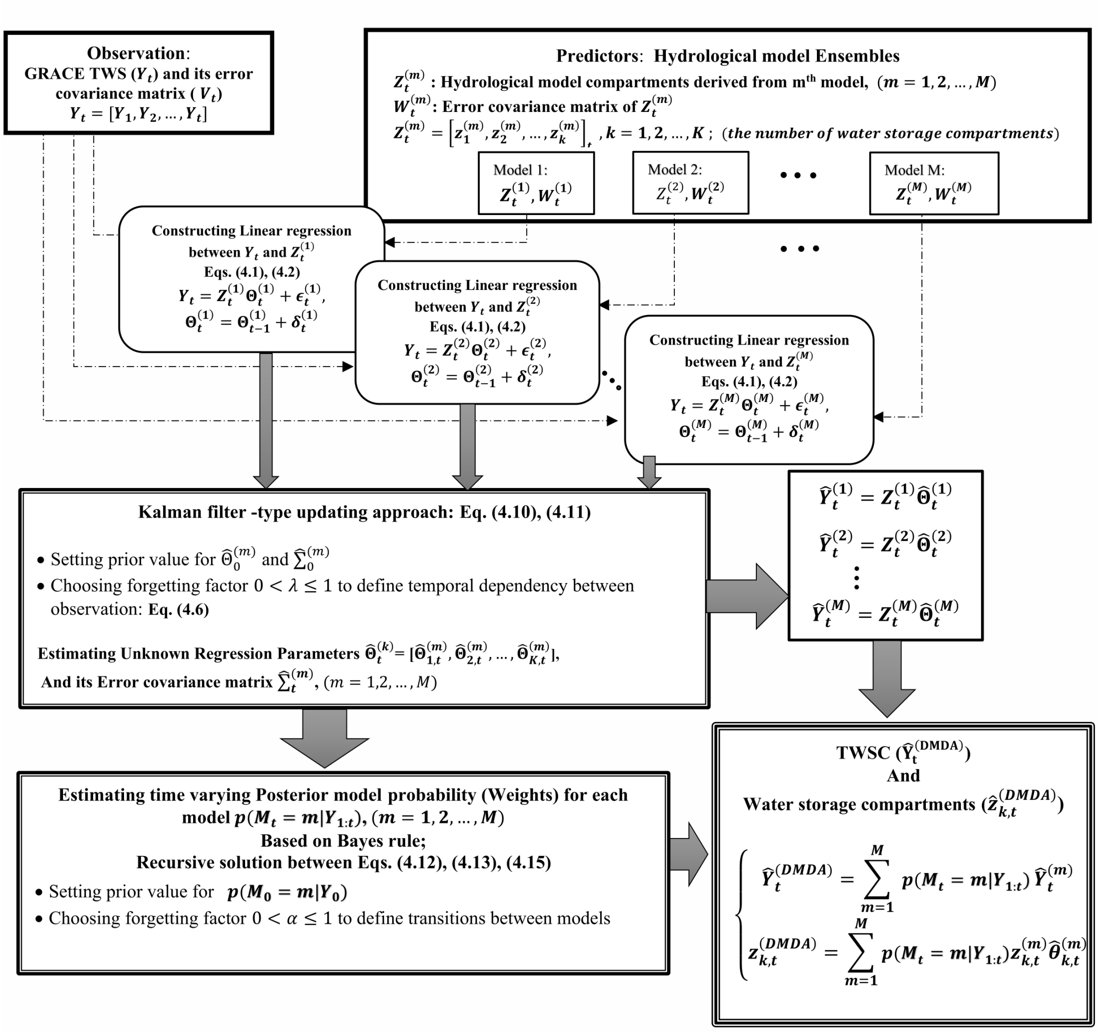

Fig. 4.1 Flowchart of the Dynamic Model Data Averaging (DMDA) method. The framework can accept an arbitrary number of models, and it can be extended to accept various types of observations.

4.5 Set up a Simulation to Test the Performance of DMDA

Before applying the DMDA method to real data, its performance is tested in a controlled synthetic example, where the results of the Bayesian update are known by definition. In the first step of this simulation, the aim is to compare DMDA and BMA in terms of updating hydrological model outputs with respect to the observations. In the second step, it will be shown that the DMDA-derived time-variable weights are the same as the expected values. To make the synthetic study simple, it

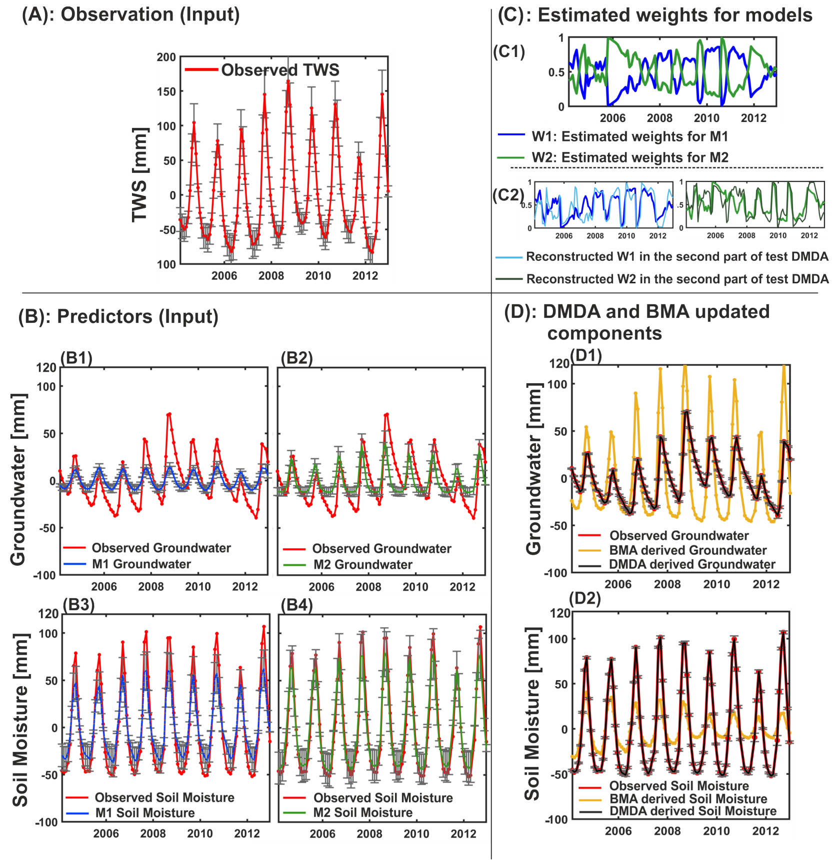

Fig. 4.2 A synthetic example, where DMDA is applied in a controlled setup, to integrate 2 hydrological models (here selected as SURFEX-TRIP (M1) and LISFLOOD (M2)) with simulated observed TWSC to separate its compartments (i.e., groundwater and soil water storage). All data sets in this simulation are related to the Niger River Basin and covering the period between 2002–2012; Figure 4.2 (A) shows TWSC simulated from PCR-GLOBWB (here standing in for observed TWSC, shown by red curves); Figure 4.2 (B1, B3) shows the time series of groundwater and soil water storage derived from model 1 (M1, the blue curves), and Fig. 4.2 (B2, B4) shows those of model 2 (M2, the green curves), which are considered as the input predictors in DMDA; The uncertainty of these data sets are shown by the grey error bars fitted to each time series. Figure 4.2 (C1) presents the time-varying weights estimated for two selected model, and Figure 4.2 (C2) shows the reconstructed of weights in the second step of this simulation. Figure 4.2 (D1) and (D2) show the updated hydrological components obtained from the DMDA and BMA method and a comparison between the obtained results and the expected values derived from simulated observation data.

is assumed that TWSC is defined as the summation of just groundwater and soil water components. By this definition, the time series of groundwater and soil water storage of two hydrological models, i.e., here selected as LISFLOOD (M_1) and SURFEX-TRIP (M_2), are introduced as predictors to the DMDA. TWSC derived from a third model, here selected to be PCR-GLOBWB, is considered as the observation (here standing in for GRACE TWSC). By this choice, after applying DMDA to merge M_1 and M_2 with simulated observed TWSC, it is expected that the updated groundwater and soil water storage estimates (DMDA results) will be fitted to those of simulated observations. Here, the results within the Niger River Basin (id:20 in Fig. 7.1) are selected, covering the period of 2002–2012. Figure 4.2 (A) shows the PCR-GLOBWB TWSC as observations (i.e., GRACE-like TWSC estimates in this study), Fig. 4.2 (B) represents the time series of groundwater and soil water derived from M_1 (B1, B3, blue curves) and M_2 (B2, B4, green curves), while the expected value of DMDA groundwater and soil water (simulated observation) are shown with the red colour curves in these figures.

The magnitude of minimum (Min), maximum (Max), and the Root Mean Squares (RMS) of the signal for all simulated data sets can be found in Table 4.1. The uncertainty of these data sets is computed following the least squares error propagation while considering the leakage error of GRACE TWSC in the Niger River Basin. It is worth mentioning that the final results of the simulation do not depend on the selection of models and the adopted simplification. The RMSD between the simulated TWSC and two selected models (reported in Table 4.1) indicates that M_2 (RMSD of $\Delta_{TWSC} = 14.1$ mm) had a better agreement with the observations compared to M_1 (RMSD of $\Delta_{TWSC} = 18.6$ mm). Figure 4.2 (C1) shows the estimated weights for the first model (W_1, Mean= 0.47) and the second model (W_2, Mean= 0.53) obtained using DMDA (Eq. (4.13)). These results show that the model which had a better agreement with observations gained higher weights.

To compare DMDA and BMA methods to average hydrological components, both of these methods are applied on simulated data sets. The final results are shown in Fig. 4.2 (D1: groundwater) and (D2: soil water). Groundwater, soil water, and consequently, TWSC derived from DMDA shows better agreement with the expected values than the BMA results. The RMS of errors for both methods is reported in Table 4.1, which indicates that although TWSC derived from BMA follow the expected value (RMS of error= 8.4 mm), the obtained individual components from this method are not close to the simulated values (RMS of errors of 20.4 mm and 18.6 mm are found for groundwater and soil water, respectively). A considerable decrease in the differences between hydrological components and the expected values of DMDA shows that the method is suitable to update multi-model water storage estimates. Details of the numerical comparisons can be found in Table 4.1.

In the second step of this simulation, the weights of the first step (W_1, W_2, Fig. 4.2 (C1)) are considered to define the true TWSC. A temporal white noise with a standard deviation of 0.02 m (equal to the standard deviation of GRACE TWSC error within the Niger River Basin) is added as error. After applying the DMDA for the second time, the reconstructed weights, using the new synthetic TWSC observations, are shown in Fig. 4.2 (C2). The correlation coefficient between W_1 and W_2 with their reconstructed values is found to be 0.73, and the RMS of the reconstruction errors is found to be 0.18.

This indicates that the DMDA's weights are close to the introduced values. These results motivate the application of DMDA to real data sets in Chapter 7.

Table 4.1 Magnitude of simulated predictors, observations, and DMDA results in a controlled synthetic simulation.

Hydrological Compartment	Model name	Min [mm]	Max [mm]	RMS [mm]
Groundwater (First model)	LISFLOOD	-10.5	16.1	7.9
Groundwater (Second model)	SURFEX-TRIP	-12.1	39.8	14.2
Groundwater (Expected value of DMDA)	PCR-GLOBWB	-39.5	70.4	24.2
Groundwater (DMDA result)	DMDA Output	-35.3	92.3	19.9
Groundwater (BMA result)	BMA Output	-46.0	130.2	43.8
Soil water (First model)	LISFLOOD	-37.4	62.2	30.8
Soil water (Second model)	SURFEX-TRIP	-45.7	79.9	41.5
Soil water (Expected value of DMDA)	PCR-GLOBWB	-52.0	107.9	48.7
Soil water (DMDA result)	DMDA Output	-58.5	113.8	51.2
Soil water (BMA result)	BMA Output	-40.8	49.6	21.0
TWSC (First model)	LISFLOOD	-46.8	75.5	37.2
TWSC (Second model)	SURFEX-TRIP	-57.6	115.2	54.6
TWSC (Expected value of DMDA results)	PCR-GLOBWB	-83.3	164.5	64.2
TWSC (DMDA result)	DMDA Output	-77.8	153.8	63.2
TWSC (BMA result)	BMA Output	-77.8	153.8	63.2
$\lvert\Delta\rvert_{\text{Groundwater}}$	$\lvert$LISFLOOD – Expected value$\rvert$	0	58.1	11.2
$\lvert\Delta\rvert_{\text{Groundwater}}$	$\lvert$SURFEX – Expected value$\rvert$	0	45.8	10.3
$\lvert\Delta\rvert_{\text{Groundwater}}$	$\lvert$DMDA – Expected value$\rvert$	0	31.2	5.3
$\lvert\Delta\rvert_{\text{Groundwater}}$	$\lvert$BMA – Expected value$\rvert$	0	87.6	20.4
$\lvert\Delta\rvert_{\text{Soil water}}$	$\lvert$LISFLOOD – Expected value$\rvert$	0	46.8	9.6
$\lvert\Delta\rvert_{\text{Soil water}}$	$\lvert$SURFEX – Expected value$\rvert$	0	29.3	5.7
$\lvert\Delta\rvert_{\text{Soil water}}$	$\lvert$DMDA – Expected value$\rvert$	0	29.2	5.2
$\lvert\Delta\rvert_{\text{Soil water}}$	$\lvert$BMA – Expected value$\rvert$	0	89.5	18.6
$\lvert\Delta\rvert_{\text{TWSC}}$	$\lvert$LISFLOOD – Expected value$\rvert$	0	94.7	18.6
$\lvert\Delta\rvert_{\text{TWSC}}$	$\lvert$SURFEX – Expected value$\rvert$	0	60.9	14.1
$\lvert\Delta\rvert_{\text{TWSC}}$	$\lvert$DMDA – Expected value$\rvert$	0	24.2	6.2
$\lvert\Delta\rvert_{\text{TWSC}}$	$\lvert$BMA – Expected value$\rvert$	0	31.4	8.4

Chapter 5

MCMC-Data Assimilation (MCMC-DA)

In this chapter, MCMC-Data Assimilation (MCMC-DA) is formulated as an extension of the Dynamic Model Data Averaging (DMDA) introduced in Chapter 4. MCMC-DA is formulated based on a linear state-space model (*Bernstein*, 2005) between observation (GRACE(-FO) TWSC) and *a priori* information (hydrological model outputs) according to the Eqs. (4.1) and (4.2), where both unknown state parameters and the error covariance matrix of the observation are allowed to vary in time. Estimation of the error covariance matrix of δ_t shown by Eq. (4.4) is the main difference between the Bayesian formulation of this study (MCMC-DA) compared to the DMDA method. In MCMC-DA, unlike DMDA, one does not need to consider a forgetting factor $0 < \lambda \leq 1$ to define temporal dependency in the Kalman filter approach (i.e., the error covariance matrix of Q in Eq. (4.7)), but instead an MCMC algorithm is applied to simultaneously estimate the unknown state parameters and the error covariance matrix Q in a recursion approach. The new formulation consists of a Markov Chain Monte Carlo (MCMC, *Geyer*, 1991; *Gilks*, 1996) algorithm, such as a Gibbs sampling algorithm (formulated in Section 5.1) and a forward-filtering backward-smoothing recursion approach (*Kitagawa*, 1987) (formulated in Section 5.2) to recursively estimate unknown state parameters as well as the temporal dependencies between them.

The posterior values of the state parameters, through MCMC-DA algorithm, are then used to update *a priori* information and to estimate their uncertainties in Section 5.3.

5.1 Gibbs Sampling Algorithm to Estimate State-Space Model

In the linear state-space model (Eqs. (4.1) and (4.2)), the conditional distribution of the parameter of interest, i.e., Θ_t and Q, is defined by the observation equation (Eq. (4.1)) and the state equation (Eq. (4.2)). To solve the linear state-space model defined by Eqs. (4.1) and (4.2), sampling techniques can be applied to generate samples from the distribution of (i) time-varying coefficients ($\Theta_{1:T} = [\Theta_1, \Theta_2, \ldots, \Theta_T]$), and (ii) the error covariance matrix of $\delta_{1:T}$, i.e., Q, conditional on the observed data ($Y_{1:T}$) and its error covariance matrix ($V_{1:T}$), and the rest of the unknown parameters, i.e., Q in (i) and $\Theta_{1:T}$ in (ii).

Gibbs sampling (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993) is one of the most common MCMC algorithms, which repeatedly generates a Markov chain of samples from the distribution of each variable, in turn, conditional on the current values of other variables and the data (see Section 3.6.2). Gibbs sampling is applicable when the joint posterior distribution of the parameter of interest is difficult to sample directly. Although, the conditional distribution of each variable is known and is easier to sample from, which is the case here.

The state-space model given by Eqs. (4.1) and (4.2) is linear, and it is assumed that the distribution of observations Y_t (GRACE(-FO) TWSC) and *a priori* information Z_t (e.g., hydrological model outputs) are Gaussian and independent from each other (though their covariances are assumed to be spatially correlated). Therefore, the conditional posterior distribution of $\Theta_{1:T}$ is a product of Gaussian Probability Density Functions (PDFs), and can be generated using the standard simulation smoother introduced by *Carter and Kohn* (1994), which showed that how we can use Gibbs sampling for Bayesian inference on a linear state-space model, with Gaussian error distribution, and with temporal variation of the state parameters Θ_t. Samples generated from the conditional posterior of Q are the product of independent Inverse-Wishart distributions (*Schuurman et al.*, 2016), which are defined on symmetric and positive definite matrices and used generally as the conjugate prior for the covariance matrix of a multivariate normal distribution in the Bayesian inference (see Section 3.1.2). The implemented Gibbs sampling for estimating the unknown parameters in linear state-space model in Eqs. (4.1) and (4.2) is summarized as follows.

Step1:
Gibbs sampling (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993) is based on the Bayes rule, which uses the knowledge derived from observations (through the likelihood function), to update prior knowledge about the parameters of interest to become posterior belief about them. Therefore, the first step of Gibbs sampling is to define initial states, or prior values, for the unknown state parameters Θ_t, where $t = 0$, and the error covariance matrix of additive innovation δ_t, i.e., $Q^{(i)}$, where i denotes the number of iteration in the Gibbs sampling (it is zero here). Details of prior values for the unknown parameters are explained in Section 5.1.1.

Step2:
Sample $\Theta_{1:T}(i)$ from the posterior PDF of $\Theta_{1:T}(i)$ conditional on the observed data ($Y_{1:T}$) and, its error covariance matrix ($V_{1:T}$) and the covariance matrix of additive innovation (residual) δ_t obtained

from the previous iteration, i.e., $Q^{(i-1)}$. This sampling is defined by:

$$p(\Theta_{1:T}^{(i)}|Y_{1:T},V_{1:T},Q^{(i-1)}) = p(\Theta_T^{(i)}|Y_{1:T},V_{1:T},Q^{(i-1)})\Pi_{t=1}^{T-1}p(\Theta_t^{(i)}|\Theta_{t+1}^{(i)},Y_{1:T},V_{1:T},Q^{(i-1)}), \quad (5.1)$$

where

$$p(\Theta_t^{(i)}|\Theta_{t+1}^{(i)},Y_{1:T},V_{1:T},Q^{(i-1)}) \sim N(\Theta_{t|t+1}^{(i)},\Sigma_{t|t+1}^{(i)}). \quad (5.2)$$

In Eq. (5.1) and for the rest of the equations in this chapter, $p(\circ|*)$ is used to denote a generic PDF of a variable (such as $\circ$) conditional on another variable (such as $*$), while $N(.)$ indicates a Gaussian PDFs, and $\Pi(.)$ is an operator to multiply PDF. Forward-filtering backward-smoothing approach, as in *Kitagawa* (1987), is used to estimate unknown state parameters $\Theta_{t|t+1}$ and their error covariance matrices $\Sigma_{t|t+1}$. Details and the corresponding equations are provided in Section 5.2.

Step3:
Generate samples from the posterior PDF of $Q^{(i)}$ conditional on the observed data $(Y_{1:T})$ and its error covariance matrix $(V_{1:T})$, and $\Theta_{1:T}^{(i)}$ that are derived from Step 2. This sampling is defined by:

$$p(Q^{(i)}|Y_{1:T},V_{1:T},\Theta_{1:T}^{(i)}) \sim IW(\bar{Q}^{(i)},\bar{v}), \quad (5.3)$$

where

$$\begin{aligned} \bar{v} &= T + v, \\ \bar{Q}^{(i)} &= Q^{(0)} + \sum_{t=1}^{T}(\Theta_t^{(i)} - \Theta_{t-1}^{(i)})(\Theta_t^{(i)} - \Theta_{t-1}^{(i)})'. \end{aligned} \quad (5.4)$$

In Eq. (5.3), $IW(.)$ denotes an Inverse-Wishart PDF, $\bar{Q}$ is the posterior scale matrix, and v is an initial value that is chosen as the degree of freedom to define the conjugate prior for Q as the product of independent Inverse-Wishart distribution (see Section 5.1.1), and $\bar{v}$ is the posterior value of the degree of freedom. In all the equations of this chapter, T denotes the total number of time steps t.

Step4:
Return to Step 2 and continue the iteration until a breaking criterion is satisfied. In this study, a simple graphical method suggested by *Brooks and Roberts* (1998) and (*Sinharay*, 2003) used to determine the number of iterations and to define the convergence of the sampling algorithm (see Section 3.6.5). This is done by creating a time-series plot for each of the parameters of interest, i.e., Θ_t in Eq. (4.1), and Q in Eq. (4.4), to view the path traversed by the chains. From the obtained results (figures not shown), it can be found that after 10000 iterations, the Gibbs sampling converged to stationary distributions. However, to increase confidence in the process, it is required to select more iterations, i.e., N=20000. The reason for this is that after initializing the sampling algorithm with a priori values for the unknown state parameters Θ_t, and the covariance of additive innovations Q, samples from

early iterations may not necessarily be representative of the actual posterior distributions. Thus, the early M=500 iterations are discarded as the 'burn-in' period.

5.1.1 Specifying Prior Values for Unknown Parameters

The first step of Gibbs sampling to solve the state-space model, given by Eqs. (4.1) and (4.2), is to define initial states, or prior values for the unknown state parameters Θ_t, where $t = 0$, and the error covariance matrix of additive innovation δ_t, i.e., $Q^{(i)}$, where i denotes the number of iteration in the Gibbs sampling (it is zero here). As discussed in Section 5.1, since the state-space equation is linear, and with assuming that the distribution of observations Y_t and Z_t to be Gaussian and independent from each other, the conditional posterior distribution of $\Theta_{1:T}$ in Eq. (5.1) is a product of their (Gaussian) PDFs. Therefore, the prior value of Θ_0 is Gaussian distributed and is defined as

$$\Theta_0 \sim N(\Theta_{0|0}, \Sigma_{0|0}), \tag{5.5}$$

where $N(.)$ represents a Gaussian (normal) distribution, and $\Theta_{0|0}$ and $\Sigma_{0|0}$ are the mean and variance of Θ_0. In GRACE(-FO) applications, $\Theta_{0|0}$ is chosen to be 1, because in theory the summation of individual water storage values Z_t must be equal to the GRACE(-FO) TWSC observation Y_t in Eq. (4.1). Uncertainty of *a priori* information (e.g., simulated individual water storage of W3RA, see Section 2.4) are stored in $\Sigma_{0|0}$, which are needed for computing updates by the Gibbs sampling in Step 2.

In Eq. (4.2), Q is the error covariance matrix of additive innovation δ_t and defines the temporal dependency between water storage states Θ_t at each time point to the previous time steps. Considering Eq. (5.5), the conjugate prior for Q can be estimated by an independent Inverse-Wishart distribution ($IW(.)$) as

$$Q^{(0)} \sim IW(S, \nu), \tag{5.6}$$

In Eq. (5.6) S is a $K \times K$ scale matrix, where K is the number of unknown state parameters, and ν is the degrees of freedom *Schuurman et al.* (2016). S is used to position the Inverse-Wishart distribution in parameter space, and $\nu > K+1$ sets the certainty about the prior information in the scale matrix. In this study, ν is set to be $K+1$, which is the minimum value that can be chosen for this parameter (*Primiceri*, 2005; *Schuurman et al.*, 2016). *Schuurman et al.* (2016) compared three Inverse-Wishart prior specifications: (I) a prior specification that is based on an identity matrix, and is often used as an uninformative prior in practice, (II) a data-based prior that uses input from maximum likelihood estimations, and (III) the default conjugate prior proposed by (*Kass and Natarajan*, 2006). Their results showed that the data-based maximum likelihood prior specification for the covariance matrix of the random parameters, based on estimates of the variances from the data, performed the best, compared to the other techniques. They also found that when the prior is specified too far from zero (e.g., Inverse-Wishart prior with S as an identity matrix), this will result in an overestimation of the variances. However, specifying the central tendencies too close to zero will result in an underestimation of the variances, as, firstly too much weight is shifted towards zero, and secondly because an element of the scale matrix set close to zero will also have a small variance. Following *Cogley* (2005); *Primiceri* (2005); *Schuurman et al.* (2016), the scale matrix S is chosen to be a constant fraction of the variance of the initial values Θ_0 as $K_Q^2\,\Sigma_{0|0}$. Therefore, Eq. (5.6) is rewritten as

$$Q^{(0)} \sim IW(K_Q^2\,\Sigma_{0|0}, K+1). \tag{5.7}$$

In Eq. (5.7), $\Sigma_{0|0}$ is the covariance matrix that is derived from the ensemble of W3RA (see Eq. (5.5)). Following *Primiceri* (2005), K_Q is chosen to be 0.01, which allows Θ_t to be a temporal variable. More details about selecting different initial values and their impact on the Inverse-Wishart estimation can be found in *Schuurman et al.* (2016). Initial values chosen for Θ_0 and $Q^{(0)}$ are then used in Step 2 and Step 3 of Gibbs sampling in Eqs. (5.2) and (5.4), respectively.

5.2 Forward-Filtering Backward-Smoothing Approach

In this section, the forward-filtering backward-smoothing approach is described, which is used in the second step of the Gibbs sampling (see Eq. (5.1)) to generate samples of $\Theta_{1:T}$ (unknown state parameters in the state-space model given by Eqs. (4.1) and (4.2)) from the PDF of $\Theta_{1:T}$ conditional on the observed data ($Y_{1:T}$) and its error covariance matrix ($V_{1:T}$), and the rest of the unknown parameters, i.e., Q. In the rest of this section, $\Theta_{t|t}$ and its variance $\Sigma_{t|t}$, as well as $\Theta_{t|t-1}$ and $\Sigma_{t|t-1}$ are defined as

$$\begin{aligned}
\Theta_{t|t} &= E(\Theta_t|Y_{1:t}, V_{1:t}, Q),\\
\Sigma_{t|t} &= Var(\Theta_t|Y_{1:t}, V_{1:t}, Q),\\
\Theta_{t|t-1} &= E(\Theta_t|Y_{1:t-1}, V_{1:t-1}, Q),\\
\Sigma_{t|t-1} &= Var(\Theta_t|Y_{1:t-1}, V_{1:t-1}, Q),
\end{aligned} \tag{5.8}$$

where t represents the (monthly) time steps between 1 and T, and $E(\circ|*)$ and $Var(\circ|*)$ denote the mean value and the variance for the normal distribution of $(\circ)$ conditional on $(*)$.

The forward-filtering backward-smoothing is a recursive approach that consists of two steps:
(1) **Forward-filtering**: where a standard Kalman filter is used following *Carter and Kohn* (1994) to recursively estimate $\Theta_{t|t}$ and $\Sigma_{t|t}$, for t=1,2,..,T, given the initial values of $\Theta_{t|t}$ and $\Sigma_{t|t}$ when t=0 ($\Theta_{0|0}$

and $\Sigma_{0|0}$, see Section 5.1.1). The Kalman filter recursion is presented in the following equations as

$$\begin{aligned}
\Theta_{t|t-1} &= \Theta_{t-1|t-1}, \\
\Sigma_{t|t-1} &= \Sigma_{t-1|t-1} + Q, \\
K_t &= \Sigma_{t|t-1} Z_t' (Z_t \Sigma_{t|t-1} Z_t' + V_t)^{-1}, \\
\Theta_{t|t} &= \Theta_{t|t-1} + K_t (Y_t - Z_t \Theta_{t|t-1}), \\
\Sigma_{t|t} &= \Sigma_{t|t-1} - K_t Z_t \Sigma_{t|t-1}.
\end{aligned} \tag{5.9}$$

The last elements of the Kalman filter recursion when t=T, i.e., $\Theta_{T|T}$ and $\Sigma_{T|T}$, are used to generate the samples of Θ_T in Eq. (5.10) as

$$\Theta_T \sim N(\Theta_{T|T}, \Sigma_{T|T}), \tag{5.10}$$

where N(.) denotes the normal distribution of Θ_T with the mean value of $\Theta_{T|T}$ and the variance of $\Sigma_{T|T}$.

(2) **Backward-smoothing**: The outputs of step (1) (i.e., $\Theta_{t|t-1}$, $\Sigma_{t|t-1}$, $\Theta_{t|t}$, and $\Sigma_{t|t}$ for t=1,2,...,T), and the generated sample of Θ_T, derived from Eq. (5.10) are then used in Eq. (5.11) to update $\Theta_{t-1|t}$ and $\Sigma_{t-1|t}$, and generate samples of Θ_{t-1}, for $t = T, T-2, ..., 1$. For a generic time t, the updating formulas of the backward recursion smoother can be written as

$$\begin{aligned}
\Theta_{t-1|t} &= \Theta_{t-1|t-1} + \Sigma_{t-1|t-1} \Sigma_{t|t-1}^{-1} (\Theta_t - \Theta_{t-1|t-1}), \\
\Sigma_{t-1|t} &= \Sigma_{t-1|t-1} - \Sigma_{t-1|t-1} \Sigma_{t|t-1}^{-1} \Sigma_{t-1|t-1}, \\
\Theta_{t-1} &\sim N(\Theta_{t-1|t}, \Sigma_{t-1|t}).
\end{aligned} \tag{5.11}$$

The backward recursion smoother is started from time T and continues until time 1. The output of the forward-filtering backward-smoother approach is the generated samples of $\Theta_{1:T}$, conditional on the observed data $(Y_{1:T})$ and its error covariance matrix ($V_{1:T}$), and the unknown covariance matrix of the additive innovation Q, which is defined as $p(\Theta_t^{(i)} | \Theta_{t+1}^{(i)}, Y_{1:T}, V_{1:T}, Q^{(i-1)})$ in Eq. (5.2). The generated samples of $\Theta_{1:T}$ are then used in the third step of Gibbs sampling to generate samples of the unknown covariance matrix of additive innovations, i.e., shown by Q in Eq. (5.3).

5.3 Updating *a priori* Information and their Uncertainties by MCMC-DA

At the end of the Gibbs sampling, $N - M$ number of generated samples for $\Theta_{1:T}$, derived from the forward-filtering backward-smoothing approach (Eq. (5.9)) is used to estimate the posterior value of unknown state parameters ($\bar{\Theta}_{1:T}$) as

$$\bar{\Theta}_t = \frac{1}{N-M}\Sigma_{i=M+1}^{N}\Theta_t^{(i)}, \tag{5.12}$$

where $\bar{\Theta}_t = [\bar{\theta}_{1,p}, \bar{\theta}_{2,p}, \dots, \bar{\theta}_{K,p}]_t$, and are used to update the priori information Z_t according to the observation equation (Eq. (4.1)) as

$$\hat{z}_{k,t,p} = z_{k,t,p}\bar{\theta}_{k,t,p}, \tag{5.13}$$

where $\hat{z}_{k,t,p}$ is the updated value of the *a priori* information $z_{k,t,p}$, and $k = 1,2,...,K$ denotes each of the *a priori* information compartments (e.g., water storage compartments).

The uncertainties of the MCMC-DA updated value of priori information ($\hat{z}_{k,t,p}$) are estimated using the variance of generated samples $\Theta_{1:T}$, i.e., $\Sigma_{1:T}$, which is derived from the forward-filtering backward-smoothing approach in Section 5.2. To this aim, the posterior value of $\Sigma_{1:T}$ is denoted by $\bar{\Sigma}_t$ which can be estimated using the following equation as

$$\bar{\Sigma}_t = \frac{1}{N-M}\Sigma_{i=M+1}^{N}\Sigma_t^{(i)}, \tag{5.14}$$

where the diagonal elements of the error covariance matrix of $\bar{\Sigma}_t$ contain $\delta\bar{\theta}^2_{k,p,t}$, corresponding to the variance of $\bar{\theta}_{K,p,t}$, which then used to estimate the uncertainties of $\hat{z}_{k,t,p}$. This is done through an error propagation procedure using the uncertainty of the *a priori* information $\delta z^2_{k,t}$ as

$$\delta\hat{z}^2_{k,p,t} = \delta\bar{\theta}^2_{k,p,t}.z^2_{k,p,t} + \delta z^2_{k,p,t}.\bar{\theta}^2_{k,p,t}, \tag{5.15}$$

where $\delta\hat{z}^2_{k,p,t}$ are the uncertainty of the MCMC-DA updated values of priori information ($\hat{z}_{k,p,t}$) for $p = 1,...,P$, $t = 1,2,...,T$ and $k = 1,2,...,K$.

MCMC-DA is implemented in Chapter 8 to merge W3RA water balance model with GRACE TWSC on a $0.125^\circ \times 0.125^\circ$ spatial grid points within the CONUS for the period of 2003-2017. Therefore, for this case study, the number of spatial grid points P is equal to 71212, and the $t = 1,2,...,168$ months covering the period of this study. Considering this scale, the application of MCMC-DA within CONUS is computationally expensive. Significantly, the convergence of the Gibbs sampling might take around 1-month using a general-purpose computer (e.g., 8-core CPU (2.9GHz) with 64GB RAM). Therefore, a paralleled implementation of the Gibbs sampling on a High-Performance Computing (HPC) system is essential to maximize the performance of MCMC-DA in large-scale applications.

Figure 5.1 summarises the work-flow of the MCMC-DA approach to update water storage components of the hydrological model using GRACE TWSC estimates.

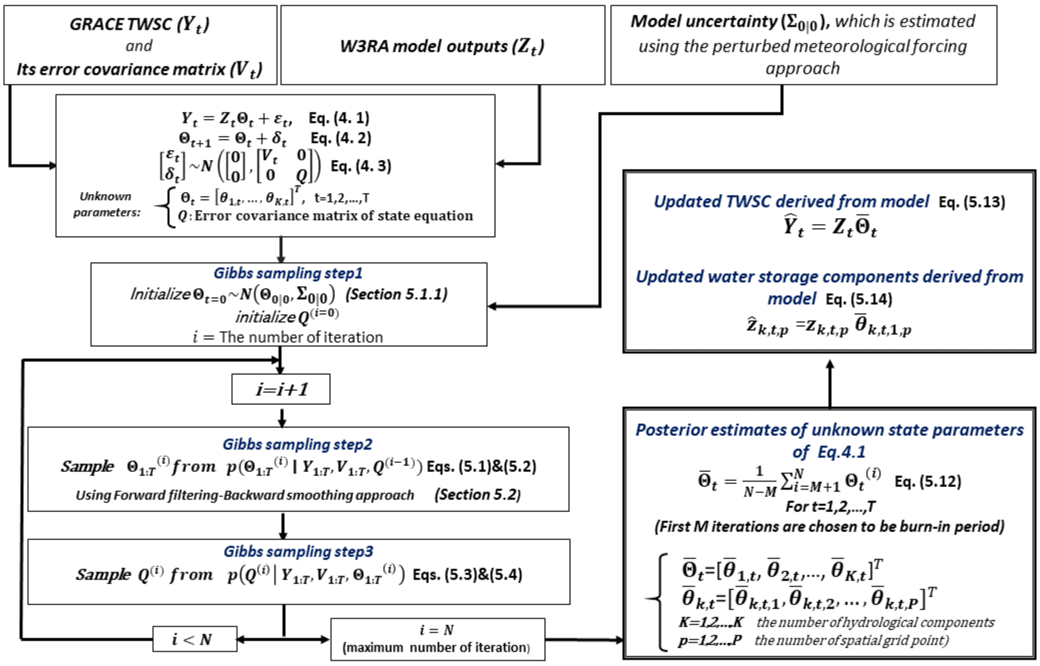

Fig. 5.1 Flowchart of the MCMC-DA method. The framework can accept an arbitrary number of models and it can be extended to accept various type of observations.

Chapter 6

Constrained Bayesian Data Assimilation (ConBay-DA)

The Constrained Bayesian Data Assimilation (ConBay-DA) approach is formulated in this chapter to merge two sets of observation with *a priori* information using a hierarchical multivariate state-space model (Section 6.1), while the state parameters and temporal dependency between them are unknown and are varying in time.

ConBay-DA benefits from a combination of forward-filtering and backward-smoothing approach (*Kitagawa*, 1987) (Section 5.2) and a Gibbs sampling algorithm (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993) (Section 6.2), which are formulated to recursively estimate time-variable state parameters and the unknown temporal dependency between them using the first set of observation. A Metropolis-Hastings algorithm (*Chib and Greenberg*, 1995) is then used in a hierarchical level to accept/reject the estimated value of a certain parameter based on the second set of observations. The posterior estimated values of the state parameters are then used to update *a priori* information and their uncertainties in Section 6.3.

ConBay-DA is formulated in this thesis with the aim of merging GRACE(-FO) field estimate and deformation rates from in-situ GNSS observation, as the first and second set of observation, respectively, for a joint estimation of the land hydrology and surface deformation signal, while hydrological and GIA model outputs are used as multiple *a priori* information within the signal separation framework. Therefore, the *a priori* information are chosen from two different sources.

6.1 Multivariate State-Space Model with Unknown State Equation

A multivariate state-space model between a set of observation and multiple *a priori* information can be represented by the observation and the state equations (*Bernstein*, 2005) as

$$Y_t = Z_t\Theta_t + X_t\beta_t + \varepsilon_t, \tag{6.1}$$

$$[\Theta_{t+1}, \beta_{t+1}] = [\Theta_t, \beta_t] + \delta_t, \tag{6.2}$$

These formulations are the extended version of Eqs. (4.1) and (4.2), where $Y_t = [y_1, y_2, \dots, y_P]_t$ represents the vector of observation for P spatial grid points, at time $t = 1,2,...,T$, while Z_t and X_t are two separate diagonal matrix to store *a priori* information from two different sources, and Θ_t and β_t are the unknown state parameters to make a relationship between the observation and *a priori* information.

For the application of ConBay-DA in this thesis, Y_t denotes the GRACE(-FO) field estimate, the diagonal elements of $Z_{t\,(P\times P)}$ contain the hydrological model outputs, and the diagonal elements of $X_{t\,(P\times P)}$ contains Equivalent Water Heights (EWHs) derived from a GIA model for the spatial grid point $p = 1,2,...,P$ at time t. Each of the diagonal elements of Z_t is a $1\times K$ vector of $[z_{1,p}, z_{2,p}, \dots, z_{K,p}]_t$, where K is the number of individual water storage components, such as snow, canopy, surface water, soil water, and groundwater storage changes. Θ_t is a $P\times 1$ vector, where each element is itself a $K\times 1$ vector containing the unknown state parameters for water storage components, $[\theta_{1,p}, \theta_{2,p}, \dots, \theta_{K,p}]_t^T$, and β_t is a $P\times 1$ vector representing the unknown state parameters for GIA signal, corresponding to the spatial grid points $p = 1,2,...,P$.

A hierarchical constraint equation is formulated here to use the second observation data sets (e.g., rates from in-situ GNSS measurements in this study) for controlling the sampling of β_t derived from Eq. (6.1) as

$$\bar{G}_t = X_t\beta_t + \gamma_t, \tag{6.3}$$

where $\bar{G}_t$ is a $P\times 1$ vector of the observation at spatial grid point $p = 1,2,...,P$, and γ_t is their corresponding measurement error.

In Eqs. (6.1), (6.2), and (6.3) ε_t, δ_t, and γ_t are additive innovations (i.e., residuals) corresponding to the observation equation, state equation, and the constraint equation, respectively. ε_t and γ_t are assumed to be Gaussian distributed with a mean value of zero and the error covariance matrices of V_t and U_t, which vary over time, and the state residual δ_t is assumed stationary Gaussian distributed and independent from ε_t and γ_t, with a mean value of zero and an error covariance matrix of Q. Thus, the distribution of the additive innovations can be written as

$$\begin{bmatrix} \varepsilon_t \\ \delta_t \\ \gamma_t \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} V_t & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & U_t \end{bmatrix} \right). \tag{6.4}$$

The uncertainty of the first set of observations (e.g., GRACE(-FO) measurements) is reflected in V_t, while the uncertainty of the second set of observation (e.g., in-situ GNSS measurements) is reflected in U_t. The error covariance matrix Q (corresponding to the state innovation δ_t) defines the temporal dependency between various compartments of the *a priori* information, which is unknown, and will be simultaneously estimated with the unknown state parameters Θ_t and β_t through the Gibbs sampling algorithm described in Section 6.2.

The ConBay-DA is formulated in the next section as a combination of a forward-filtering and backward-smoothing approach, Gibbs sampling, and Metropolis-Hastings to estimate the unknown state parameters Θ_t and β_t, and the covariance matrix Q, while the generated samples of β_t in each iteration of Gibbs sampling are not accepted automatically as posterior samples; instead they are introduced as candidate samples to the hierarchical Metropolis-Hastings to be accepted or rejected based on the GNSS measurements. These candidate samples are accepted probabilistically based on the acceptance probability α (see Section 3.6.3), which is estimated using the posterior probability distribution of β_t conditional on the second set of the observations (e.g., GNSS measurement) based on the Eq. (6.3).

6.2 ConBay-DA Formulation

The multivariate state-space model defined by Eqs. (6.1) and (6.2), provide the conditional distribution of the parameter of interest, i.e., Θ_t, β_t and Q. The Gibbs sampling algorithm is formulated here to generate samples from the posterior distribution of (i) time varying state parameters ($\Theta_{1:T} = [\Theta_1, \Theta_2, \ldots, \Theta_T]$) and ($\beta_{1:T} = [\beta_1, \beta_2, \ldots, \beta_T]$) and (ii) the error covariance matrix of $\delta_{1:T}$, i.e., Q, conditional on the first set of observation ($Y_{1:T}$) and its error covariance matrix ($V_{1:T}$), and the rest of the unknown parameters, i.e., Q in (i) and $\Theta_{1:T}$ and $\beta_{1:T}$ in (ii). However, generated samples from the posterior distribution of β_t (e.g., state-space parameters corresponding to the GIA effects in this study) in step (i) are not accepted automatically as posterior samples and will be controlled by the second set of observations $\bar{G}_t$ (e.g., in-situ GNSS measurement) based on Eq. (6.3). Before generating samples of covariance matrix Q, a Metropolis-Hastings algorithm is formulated in a hierarchical level to estimate the acceptance probabilities using the posterior distribution of candidate samples β_t conditional on the observations $\bar{G}_t$. The acceptance probabilities are then used to determine the best generated samples of β_t, which leads to decrease RMSD between the updated value of X_t (i.e., $X_t\beta_t$)

and the observations $\bar{G}_t$ according to Eq. (6.3). The accepted values of the generated samples β_t are then used to generate the posterior samples of Q in step (ii).

The state-space model given by Eqs. (6.1) and (6.2) is linear, and it is assumed that the distribution of the observations and the *a priori* information are Gaussian and independent from each other (though their covariances are assumed to be spatially correlated). Therefore, the conditional posterior distribution of $\Theta_{1:T}$ and $\beta_{1:T}$ are the products of Gaussian Probability Density Functions (PDFs), which can be generated using a standard simulation smoother introduced by *Carter and Kohn* (1994). Samples generated from the conditional posterior of Q are the product of independent Inverse-Wishart distributions (*Schuurman et al.*, 2016), which are defined on symmetric and positive definite matrices and used generally as the conjugate prior (defined in Section 3.1.2) for the covariance matrix of a multivariate normal distribution in the Bayesian inference. The mathematical formulations to estimate the posterior distribution of the parameters of interest are explained in details in what follows.

Step1:
Define initial states, or prior values, for the unknown state parameters Θ_t and β_t, where $t = 0$, and the prior value of the error covariance matrix of additive innovation δ_t, i.e., $Q^{(i)}$, where i denotes the iteration number in the Gibbs sampling (it is zero here). Details of initial values for the unknown parameters are explained in Section 5.1.1, where the prior values for β_t are determined similar to those of Θ_t.

Step2:
Sample $\Theta^{(i)}_{1:T}$, $\beta^{(i)}_{1:T}$ from the posterior PDFs of $\Theta^{(i)}_{1:T}$, $\beta^{(i)}_{1:T}$ conditional on the observation $(Y_{1:T})$, its error covariance matrix ($V_{1:T}$), and the covariance matrix of additive innovations δ_t, which is obtained from the previous iteration, i.e., $Q^{(i-1)}$. To simplify the rest of the equations, we define $\Phi_t = [\Theta_t, \beta_t]'$, where $[\circ]'$ denotes the transpose of the matrix $[\circ]$.

$$p(\Phi^{(i)}_{1:T}|Y_{1:T}, V_{1:T}, Q^{(i-1)}) = p(\Phi^{(i)}_{T}|Y_{1:T}, V_{1:T}, Q^{(i-1)})\Pi_{t=1}^{T-1} p(\Phi^{(i)}_{t}|\Phi^{(i)}_{t+1}, Y_{1:T}, V_{1:T}, Q^{(i-1)}), \tag{6.5}$$

where

$$p(\Phi^{(i)}_{t}|\Phi^{(i)}_{t+1}, Y_{1:T}, V_{1:T}, Q^{(i-1)}) \sim N(\Phi^{(i)}_{t|t+1}, \Sigma^{(i)}_{t|t+1}). \tag{6.6}$$

In Eq. (6.5) and for the rest of the equations, $p(\circ|*)$ is used to denote a generic PDF of (a variable such as $\circ$) conditional on (another variable such as $*$), while $N(.)$ indicates a Gaussian PDF, and $\Pi(.)$ is an operator to multiply PDF. A Forward-filtering backward-smoothing approach, as in *Kitagawa* (1987), is used to estimate the unknown state parameters $\Theta_{t|t+1}$ and their error covariance matrices $\Sigma_{t|t+1}$. Details and the corresponding equations are provided in Section 5.2. For the application of ConBay-DA, the parameter Θ in all the equations of the forward-filtering backward-smoothing approach of Section 5.2, must be replaced by Φ, which has already defined as $[\Theta, \beta]$. The outputs of the forward-filtering backward-smoother approach is the generated samples of $\Theta_{1:T}^{(i)}$ and $\beta_{1:T}^{(i)}$ derived from multivariate normal distribution, where each generated samples of $\Theta_t^{(i)}$ and $\beta_t^{(i)}$ can be described as a normal distribution with a mean value denoted by $\mu(*)^{(i)}$ and the covariance matrix of $\Sigma_*^{(i)}$ as

$$\begin{aligned} \Theta_t^{(i)} &\sim N\left(\mu(\Theta_t)^{(i)}, \Sigma_{\Theta_t}^{(i)}\right) \\ \beta_t^{(i)} &\sim N\left(\mu(\beta_t)^{(i)}, \Sigma_{\beta_t}^{(i)}\right) \end{aligned} \tag{6.7}$$

Step3:

Estimate the acceptance probability α_t to accept/reject generated samples of $\beta_t^{(i)}$, where $t = 1, ..., T$, based on the constraint equation of Eq. (6.3). The acceptance probability α_t are estimated for each spatial grid point $p = 1, 2, ..., P$. For each iteration (i), the acceptance functions $\alpha_t^{(j)}$ for $j = 1, 2, ..., i$ are estimated to compare the posterior distributions of the $\beta_t^{(i)}$ and $\beta_t^{(j)}$ conditional on the GNSS measurements $\bar{G}_t$ and its error covariance matrix U_t as

$$\alpha_t^{(j)} = \frac{p(\beta_t^{(j)}|\bar{G}_t, U_t)}{p(\beta_t^{(i)}|\bar{G}_t, U_t)} = \frac{p(\bar{G}_t|\beta_t^{(j)}, U_t)p(\beta_t^{(j)})}{p(\bar{G}_t|\beta_t^{(i)}, U_t)p(\beta_t^{(i)})}, \qquad j = 1, 2, ..., i. \tag{6.8}$$

where $p(\bar{G}_t|\beta_t^{(j)}, U_t)$ is the likelihood density of the second observation data set $(\bar{G}_t)$ conditional on the parameter of interest $(\beta_t^{(j)})$ and the error covariance matrix of the residual U_t, and $p(\beta_t^{(j)})$ denotes the prior distribution of $\beta_t^{(j)}$. Since the state-space model is formulated with the assumption of Gaussian distribution for all the data and model parameters, the likelihood density of the observation $\bar{G}_t$, as well as the prior distribution of $\beta_t^{(j)}$ are estimated using the normal density function as

$$p(\bar{G}_t|\beta_t^{(j)}, U_t) = \left(2\pi\Sigma_{\hat{X}_t}^{(j)}\right)^{-1} exp\left(\frac{-\left(\bar{G}_t - \hat{X}_t^{(j)}\right)^2}{2\Sigma_{\hat{X}_t}^{(j)}}\right), \qquad j = 1, 2, ..., i, \tag{6.9}$$

$$p(\beta_t^{(j)}) = \left(2\pi\Sigma_{\beta_t}^{(j)}\right)^{-1} exp\left(\frac{\left(\beta_t^{(j)} - \mu(\beta_t^{(j)})\right)^2}{2\Sigma_{\beta_t}^{(j)}}\right), \qquad j = 1, 2, ..., i, \tag{6.10}$$

where $\hat{X}_t^{(j)}$ is the updated value of X_t (GIA model output in this study) and is estimated using the mean value of generated sample $\beta_t^{(j)}$, while $\Sigma_{\hat{X}_t}^{(j)}$ denotes the uncertainty of $\hat{X}_t^{(j)}$, which is estimated using the error propagation formula as

$$\begin{aligned} hatX_t^{(j)} &= X_t.\mu(\beta_t)^{(j)}, \\ \Sigma_{\hat{X}_t}^{(j)} &= X_t\Sigma_{\beta_t}^{(j)}X_t' + U_t, \end{aligned} \tag{6.11}$$

where $j = 1,2,...,i$. It is worth noting here that $\alpha_t^{(i)} = \frac{p(\beta_t^{(i)}|\bar{G}_t,U_t)}{p(\beta_t^{(i)}|\bar{G}_t,U_t)} = 1$. At iteration (i), the generated sample of $\beta_t^{(i)}$ is accepted as $\beta_t^{accepted^{(i)}}$, when $\alpha_t^{(i)} = min\{\alpha_t^{(1)},\alpha_t^{(2)},...,\alpha_t^{(i)}\}$, otherwise $\beta_t^{accepted^{(i)}} = \beta_t^{(l)}$ if

$$\alpha_t^{(l)} = min\{\alpha_t^{(1)},\alpha_t^{(2)},...,\alpha_t^{(l-1)},\alpha_t^{(l)},\alpha_t^{(l+1)},...,\alpha_t^{(i)}\}, \qquad l = 1,2,...,i-1. \tag{6.12}$$

The minimum value of $\alpha_t^{(l)}$ indicates that the generated samples of $\beta_t^{(l)}$, and therefore, the update value of priori information X_t, i.e., $X_t\beta_t^{(l)}$, is more fitted to the observation $\bar{G}_t$ (according to the Eq. (6.3)), compared to those of $X_t\beta_t^{(i)}$. $\beta_{1:T}^{accepted^{(i)}}$ derived from this step, along with the $\Theta_{1:T}^{(i)}$ of Step 2 are then used to generate samples of the unknown covariance matrix of additive innovations δ_t, i.e., Q in Eq. (6.4), in the next step of the Gibbs sampling.

Step4:

Sample $Q^{(i)}$ from the posterior PDF of $Q^{(i)}$ conditional on the observed data $(Y_{1:T})$ and its error covariance matrix $(V_{1:T})$, and $\Phi_{1:T}^{(i)} = [\Theta_{1:T}^{(i)},\beta_{1:T}^{accepted^{(i)}}]$, where $\Theta_{1:T}^{(i)}$ are estimated in step 2, and $\beta_{1:T}^{accepted^{(i)}}$ are derived from step 3. This sampling is defined by:

$$p(Q^{(i)}|Y_{1:T},V_{1:T},\Phi_{1:T}^{(i)}) \sim IW(\bar{Q}^{(i)},\bar{\nu}), \tag{6.13}$$

where

$$\begin{aligned} \bar{\nu} &= T + \nu, \\ \bar{Q}^{(i)} &= Q^{(0)} + \sum_{t=1}^{T}(\Phi_t^{(i)} - \Phi_{t-1}^{(i)})(\Phi_t^{(i)} - \Phi_{t-1}^{(i)})'. \end{aligned} \tag{6.14}$$

In Eq. (6.13), $IW(.)$ denotes an Inverse-Wishart PDF, $\bar{Q}$ is the posterior scale matrix, and ν is an initial value that is chosen as the degree of freedom to define the conjugate prior for Q as the product of independent Inverse-Wishart distribution (see 5.1.1), and $\bar{\nu}$ is the posterior value of the degree of freedom. In all these equations, T denotes the total number of time steps t.

Step5:

Return to step 2 and continue the iteration until a breaking criterion is satisfied. In this study, the number of iteration is chosen to be $N = 20000$, such as those determined for MCMC-DA in Chapter 5, and the first $M = 500$ iterations are discarded as the 'burn-in' period.

6.3 Updating *a priori* information and their Uncertainties by ConBay-DA

At the end of the ConBay-DA algorithm, $N - M$ generated samples for $\Theta_{1:T}$ and $\beta_{1:T}$, are used to estimate the posterior value of unknown state parameters $\bar{\Theta}_{1:T}$, and $\bar{\beta}_{1:T}$ as

$$\begin{aligned}\bar{\Theta}_t &= \frac{1}{N-M}\Sigma_{i=M+1}^{N}\Theta_t^{(i)},\\ \bar{\beta}_t &= \frac{1}{N-M}\Sigma_{i=M+1}^{N}\beta_t^{accepted^{(i)}}, \qquad \text{for t=1,2,...,T}\end{aligned} \tag{6.15}$$

$\bar{\Theta}_{1:T}$, and $\bar{\beta}_{1:T}$ are then used to update the *a priori* information Z_t (e.g., hydrological model outputs) and X_t (e.g., GIA model output) as

$$\begin{aligned}\hat{Z}_t &= Z_t\bar{\Theta}_t,\\ \hat{X}_t &= X_t\bar{\beta}_t, \qquad \text{for t=1,2,...,T,}\end{aligned} \tag{6.16}$$

The uncertainties of the ConBay-DA updated signals are estimated using the variance of generated samples, i.e., $\Sigma_{\Theta_t}^{(i)}$ and $\Sigma_{\beta_t}^{(i)}$ in Eq. (6.7), which is derived from the forward-filtering backward-smoothing approach following the equations in Section 5.2. To this aim, the posterior value of $\bar{\Sigma}_{\Theta_t}$ and $\bar{\Sigma}_{\beta_t}$ for $t = 1, 2, ..., T$ are obtained as

$$\begin{aligned}\bar{\Sigma}_{\Theta_t} &= \frac{1}{N-M}\Sigma_{i=M+1}^{N}\Sigma_{\Theta_t}^{(i)},\\ \bar{\Sigma}_{\beta_t} &= \frac{1}{N-M}\Sigma_{i=M+1}^{N}\Sigma_{\beta_t}^{(i)},\end{aligned} \tag{6.17}$$

The diagonal elements of the error covariance matrix of $\bar{\Sigma}_{\Theta_t}$ contain $\delta\bar{\theta}_{k,t}^2$ corresponding to each compartment of the *a priori* information Z_t, $k = 1, .., K$, which can be used to estimate the uncertainties of $\hat{Z}_t$ through an error propagation procedure according to the Eq. (5.15).

The uncertainties of $\hat{X}_t$ derived from Eq. (6.16) is estimated similar to that of $\hat{Z}_t$ based on Eq. (5.15). The work-flow of the ConBay-DA approach, formulated in this chapter, is summarised in Fig. 6.1.

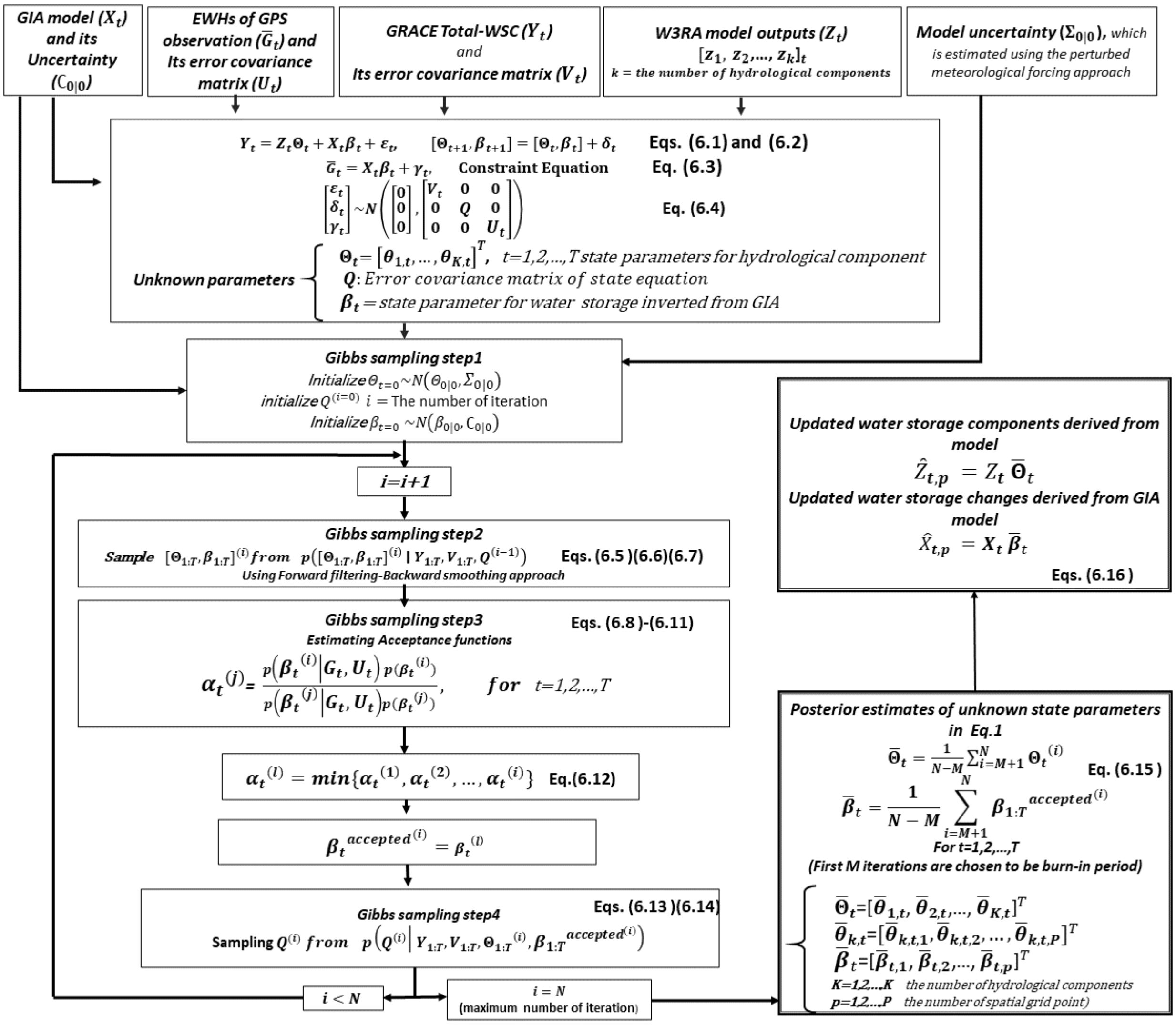

Fig. 6.1 Flowchart of the ConBay-DA method. The framework can accept an arbitrary number of models, and it can be extended to accept the various type of observations.

Chapter 7

Application of DMDA to Merge Multi-Hydrological Models with GRACE data

7.1 Introduction

The DMDA method, which is proposed in Chapter 4, is implemented here to merge multi-model water storage simulations with GRACE TWSC within the world's 33 largest river basin (Fig. 7.1) for the period of 2002–2012, during which both GRACE data and model simulations are available. To this aim, surface and sub-surface water storage simulations of the six published global hydrological and land surface models (*Schellekens et al.*, 2017) are used. These models are introduced in Section 2.4. Results of this chapter follows *Mehrnegar et al.* (2020a). A challenging problem in merging GRACE TWSC with the outputs from multiple hydrological models is related to their different spatial and temporal resolutions. To overcome the computational problem caused by the spatial and temporal mismatch, *Schumacher et al.* (2016) introduced spatial and temporal matching functions, which can avoid computational problems. In this study, the spatial/temporal operator is not implemented because both model outputs and GRACE data are set at monthly (temporal) and basin-averaged (spatial). Handling the differences in the spectral domain is described in Section 2.4.

In what follows, after presenting an overview of GRACE TWSC (Section 7.2) and model-derived water storage changes (Section 7.3) a comparison between GRACE and model-derived TWSC is presented in Section 7.4. The model outputs are compared against GRACE TWSC in Section 7.5, using the DMDA-derived temporal weights within the world's largest river basins, covering 2002–2012. The DMDA-derived updates, which are assigned to the long-term trend of surface and sub-surface water storage components, are explored and interpreted in Section 7.6. Numerous studies showed that various hydrological models are not in close agreement for many regions worldwide. The representation of long-term trends, in particular, is a common problem among hydrological model simulations (*Döll et al.*, 2014), which is the component in which we are primarily interested in this study.

In addition, the DMDA-derived TWSC is compared with those derived from original model outputs in Section 7.7. Temporal correlation coefficients between model-derived storage outputs and the ENSO index are used as a measure to determine whether implementing the DMDA helps to derive realistic storage simulations (see Section 7.8). A summary of the obtained results is provided in Section 7.9.

7.2 An Overview of GRACE TWSC

For the application of DMDA in this thesis, GRACE TWSC is estimated globally with 0.5° spatial resolution, the same as those of models outputs. Basin averages of GRACE TWSC within the world's 33 largest river basins and the leakage errors are obtained following Section 2.2.5. Figure 7.2 provides an overview of the basin averaged GRACE TWSC within the world's 33 largest river basins (Fig. 7.1), where Fig. 7.2 (A) shows the Standard Deviation (StD) of GRACE TWSC covering the period of 2002–2012. This map indicates the strength of the TWSC signal in these basins. Figure 7.2 (B) contains the Standard Deviation (StD) of the TWSC errors, which indicates the magnitude of these errors. Finally, Fig. 7.2 (C) shows the Signal to Noise Ratio (SNR) that is computed by dividing the plot (A) by (B). These maps show that the basin averaged GRACE TWSC is of acceptable accuracy and can be used with confidence in this application.

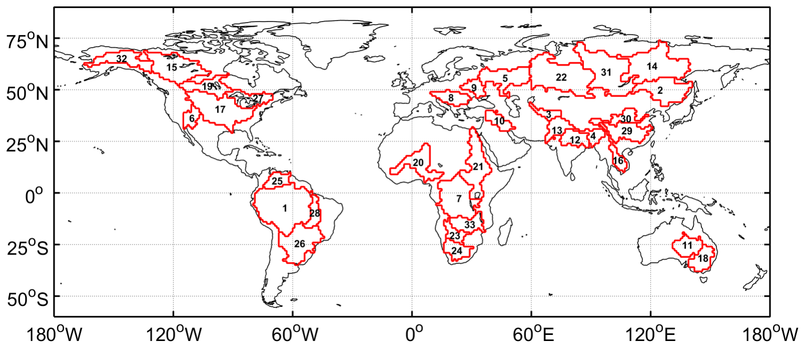

Fig. 7.1 The world's 33 largest river basins examined in this study; The numbered river basins include the following: 1: Amazon, 2: Amur, 3: Aral, 4: Brahmaputra, 5: Caspian-Volga, 6: Colorado, 7: Congo, 8: Danube, 9: Dnieper, 10: Euphrates, 11: Lake Eyre, 12: Ganges, 13: Indus, 14: Lena, 15: Mackenzie, 16: Mekong, 17: Mississippi, 18: Murray, 19: Nelson, 20: Niger, 21: Nile, 22: Ob, 23: Okavango, 24: Orange, 25: Orinoco, 26: Parana, 27: St. Lawrence, 28: Tocantins, 29: Yangtze, 30: Yellow, 31: Yenisei, 32: Yukon, 33: Zambezi

7.3 An Overview of water storage estimates from multiple hydrological model outputs

Long-term linear trends (between 2002-2012) fitted to the groundwater, soil water, and surface water storage of the six global hydrological models are shown in Fig. 7.3, Fig. 7.4 and Fig.7.5. From Fig. 7.3, it can be seen that except PCR-GLOBWB, the other 4 hydrological models, do not show considerable linear trends in groundwater storage in almost all the 33 river basins. LISFLOOD shows decreasing groundwater storage in some irrigated regions such as the Ganges, Barhmaputra, and Indus River Basins, which experienced a strong decline in rainfall over the entire period of our study (e.g., 9.0 $\pm$ 4.0 mm/decade between 1994–2014 over the Ganges and Brahmaputra River Basins *Khandu et al.*, 2016). A large difference between PCR-GLOBWB and LISFLOOD can be found in the Ganges and Indus River Basins, where PCR-GLOBWB groundwater shows positive trend and LISFLOOD groundwater shows a strong negative trend within these regions. Figure 7.4 shows that soil water

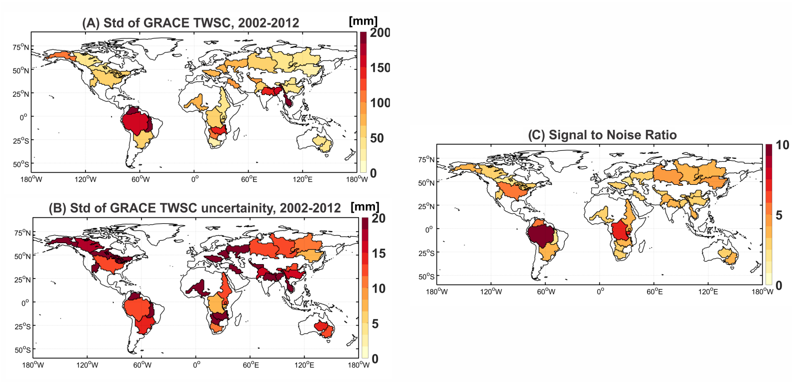

Fig. 7.2 An overview of the basin averaged GRACE TWSC for the world's 33 largest river basins. (A) Standard Deviation (StD) of the basin averaged GRACE TWSC between 2002–2012 showing the strength of its signal. (B) Standard Deviation StD of the TWSC errors. (C) Signal to Noise Ratio (SNR) computed by dividing the values in plot (A) by corresponding values of plot (B).

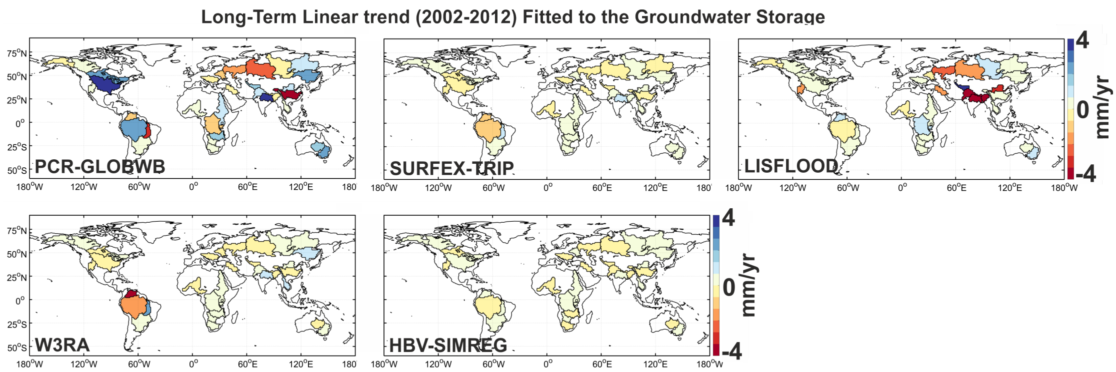

Fig. 7.3 Long-term linear trends [mm/yr] fitted to the groundwater storage of global hydrological model outputs used in this study between 2002-2012.

storage changes derived from all six hydrological models indicate an increasing trend within the Murray and lake Eyre River Basins in Australia, and a decreasing trend in the Ob and Caspian-Volga River Basins in North Asia. Among these models, ORCHIDEE and SURFEX-TRIP show the highest linear rate in the soil water changes, mostly in North Asia.

Among the six global models used in this study, only PCR-GLOBWB, SURFEX-TRIP, and ORCHIDEE models contain surface water storage estimates. Figure 7.5 indicates that PCR-GLOBWB does not show any significant trend in its surface water changes, except in the Mackenzie River Basin in the northwest of America. SURFEX-TRIP and ORCHIDEE model indicates strong linear trends in surface water changes within the basins in South America, between 2002-2012, which is a negative trend for the Amazon River Basin and a positive trend for the Parana and Tocantins. ORCHIDEE surface water is found to contain a positive trend over the Nile and Ganges River Basins between 2002-2012.

7.4 Comparison between GRACE and Model-derived TWSC

To understand how GRACE TWSC modifies model-derived individual and vertically summed water storage simulations, the mismatch between GRACE TWSC and those derived from the six hydrological models (i.e., PCR-GLOBWB, SURFEX-TRIP, LISFLOOD, HBV-SIMREG, W3RA, ORCHIDEE,

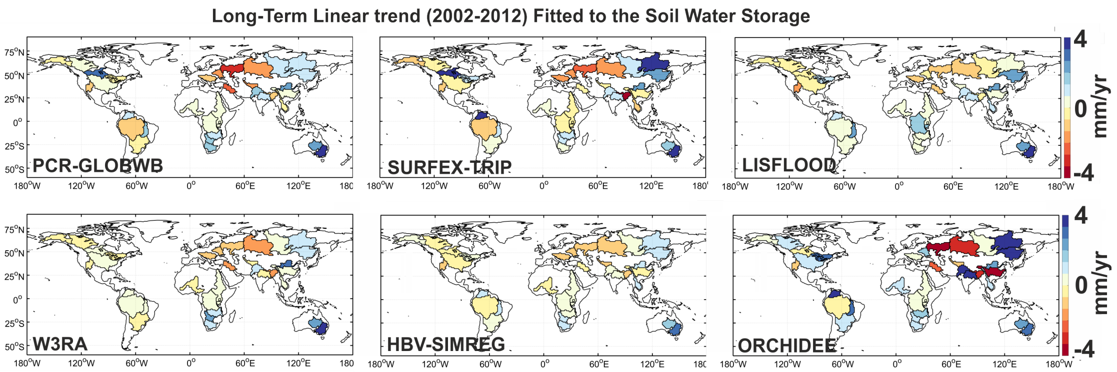

Fig. 7.4 Long-term linear trends [mm/yr] fitted to the soil water storage changes of global hydrological model outputs used in this study between 2002-2012.

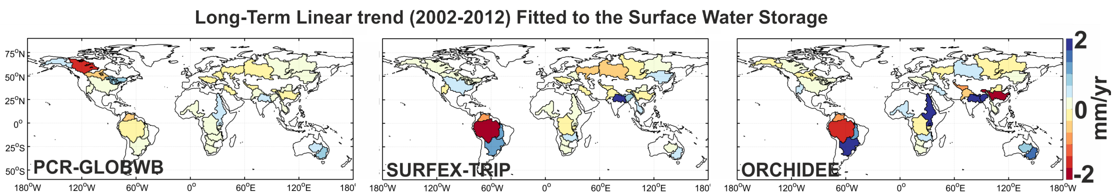

Fig. 7.5 Long-term linear trends [mm/yr] fitted to the surface water storage of six global hydrological model outputs used in this study between 2002-2012.

Table 7.1 The Root Mean Squares Differences (RMSD) between GRACE TWSC and: 6 global hydrological models, the Mean of TWSC derived from these 6 models, and DMDA TWSC, between 2002–2012.

ID	Basin Name	PCR-GLOBWB	SURFEX-TRIP	LISFLOOD	HBV-SIMREG	W3RA	ORCHIDEE	Mean Model	DMDA
1	Amazon	78.03	69.21	116.90	112.42	83.27	91.76	78.58	35.77
2	Amur	24.38	25.61	27.49	23.82	23.91	32.12	24.01	20.85
3	Aral	47.42	59.30	62.26	62.67	65.25	58.09	56.12	37.73
4	Brahmaputra	67.52	76.57	119.69	83.70	61.78	167.11	70.45	49.18
5	Caspian-Volga	22.29	44.18	40.19	61.00	50.52	32.08	38.74	20.03
6	Colorado	36.07	38.17	41.88	42.52	37.28	35.98	34.32	31.08
7	Congo	26.75	37.84	34.38	34.33	33.60	31.40	29.11	27.82
8	Danube	34.71	39.28	40.42	50.06	45.03	33.47	37.96	25.36
9	Dnieper	39.69	53.11	48.09	59.82	48.43	48.98	47.23	38.93
10	Euphrates	62.22	77.38	69.95	78.62	73.82	75.64	72.41	46.15
11	Lake Eyre	26.57	35.24	32.47	37.86	33.17	29.62	31.96	24.58
12	Ganges	87.86	97.34	90.07	92.92	80.98	98.85	74.33	40.12
13	Indus	47.73	46.56	51.95	44.56	46.64	55.25	38.58	35.80
14	Lena	30.90	21.96	41.23	29.65	32.05	36.93	29.08	17.82
15	Mackenzie	27.47	20.89	34.82	25.69	25.92	28.70	23.98	16.59
16	Mekong	70.68	52.22	127.17	103.75	75.58	58.36	70.95	50.26
17	Mississippi	17.44	32.13	39.67	42.92	37.46	29.31	29.35	17.86
18	Murray	29.13	35.86	32.44	38.47	35.65	32.56	32.65	26.16
19	Nelson	27.36	39.05	41.51	47.34	40.30	35.84	35.40	22.41
20	Niger	30.14	35.20	38.22	44.18	40.59	22.67	31.36	24.27
21	Nile	18.64	26.26	29.08	32.01	29.00	27.19	21.60	18.28
22	Ob	19.60	31.18	41.63	42.82	38.72	38.04	32.40	17.64
23	Okavango	84.51	91.20	90.68	93.86	86.11	92.52	89.11	78.26
24	Orange	19.48	22.90	20.17	25.33	22.85	23.36	21.63	20.46
25	Orinoco	104.59	86.10	142.70	116.51	79.25	79.95	80.75	50.53
26	Parana	24.26	24.93	37.27	32.09	30.48	27.42	22.60	19.33
27	St. Lawrence	39.49	58.64	66.07	66.71	54.91	38.58	50.17	27.27
28	Tocantins	92.10	82.62	134.08	120.09	92.07	76.58	87.35	59.12
29	Yangtze	33.50	25.90	34.52	26.39	21.49	48.77	27.56	19.62
30	Yellow	37.17	34.84	28.68	34.98	35.50	28.82	31.52	24.60
31	Yenisei	19.85	25.85	36.65	30.91	27.91	29.90	24.65	15.71
32	Yukon	95.58	63.80	96.54	102.20	95.93	99.00	91.34	50.93
33	Zambezi	72.32	89.30	89.38	100.50	93.29	75.78	84.22	64.07

Schellekens et al., 2017) are shown in Fig. 7.6 in terms of Root Mean Square of Differences (RMSD), without removing linear trends and seasonality. The biggest RMSD is estimated within the basins which are strongly affected by climate and human water-use such as Amazon, Orinoco and Tocantins (all corresponding to the LISFLOOD model), the Yukon basin in North America (corresponding to the HBV-SIMREG model), and some irrigated regions such as Ganges and Brahmaputra (corresponding to the ORCHIDEE model). The numerical values corresponding to Fig. 7.6 are reported in Table 7.1. The RMSD computed for each model against GRACE TWSC, after removing the linear trends and seasonality, is shown in Fig. 7.7, from which one can see that inter-annual differences exist between model outputs and GRACE in some river basins. These differences could be related to climate variability such as that of ENSO (see, e.g., *Anyah et al.*, 2018; *Chen et al.*, 2010a; *Forootan et al.*, 2016, 2019; *Hurkmans et al.*, 2009; *Ni et al.*, 2018; *Zhang et al.*, 2015). By comparing Fig. 7.6 and Fig. 7.7, it is found that the large differences between hydrological models and GRACE TWSC are mainly due to the differences in the linear trends and seasonality. GRACE represents hydrological

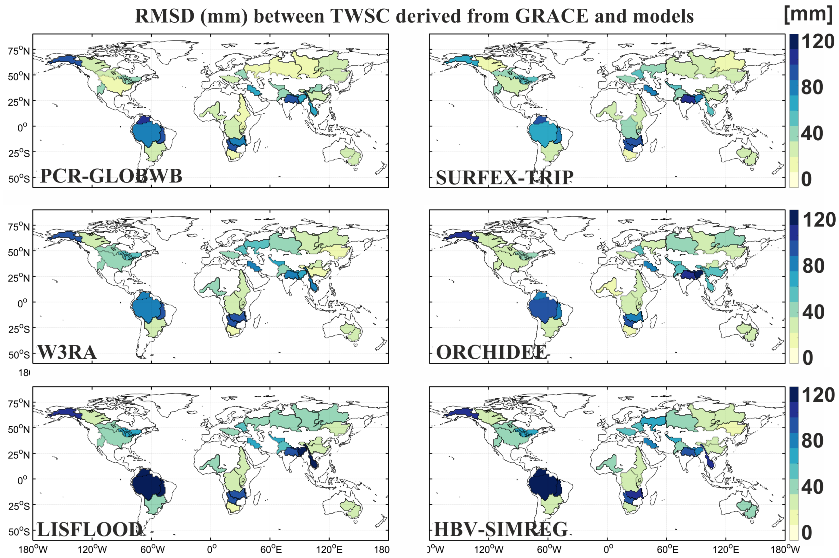

Fig. 7.6 The Root Mean Squares Differences (RMSD) between GRACE TWSC and those of derived from 6 global hydrological models, over the world's 33 largest river basins, covering the period of 2002–2012.

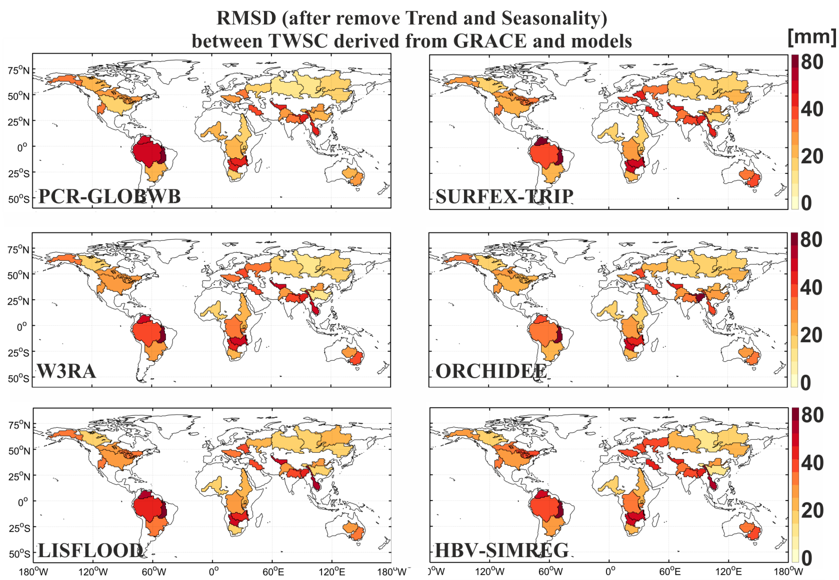

Fig. 7.7 The RMSD between GRACE TWSC and those derived from 6 hydrological models, after removing linear trends and seasonality, over the world's 33 largest river basin covering the period of 2002–2012.

variations that are composed of both climate variability and human activities, but there are certain limitations in the hydrological models to simulate these components (*Pokhrel et al.*, 2012); and that is why the proposed DMDA combination is crucial to provide more realistic water storage estimations.

7.5 DMDA Weights to Compare Global Hydrological Models

TWSC derived from DMDA is a weighted average of TWSC derived from selected models, which is obtained by estimating time-varying weights based on the Bayes rule as in Eq. (4.17). Figure 7.8 shows the estimated weights for ten basins with the largest RMSD between GRACE TWSC and those derived from individual models. Time-variable weights derived from DMDA allow us (1) to quantify the quality and compare individual water storage simulations derived from each global hydrological model against GRACE TWSC at different times, and (2) to separate GRACE TWSC in a Bayesian framework, while considering different model structures and errors within and between model simulations and GRACE data. The average of weights during 2002–2012, derived from DMDA, is considered as the basis to select the best model to simulate TWSC over 33 river basins, which are shown in the middle of Fig. 7.8.

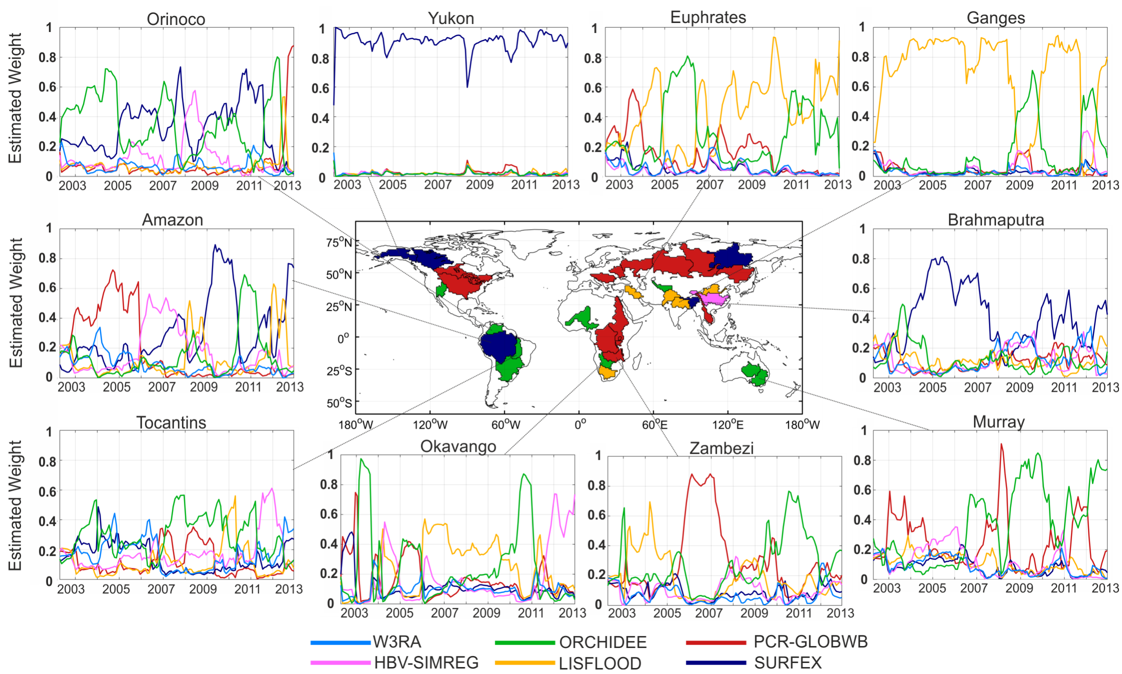

Fig. 7.8 Temporal weights derived from DMDA approach for the six initially considered models, over 10 selected river basins with the biggest RMSD computed using GRACE and models-derived TWSC. In the middle of Fig. 7.8, the most contributed models in the DMDA-derived TWSC are shown over the world's 33 largest river basins, covering the period of 2002–2012.

From the numerical results, PCR-GLOBWB is associated with the largest weights during the period of study; thus, the PCR-GLOBWB is the largest contributor to DMDA-derived TWSC in North Asia, Central Africa, and North America. The weights computed for SURFEX-TRIP are found to be larger than other models within the snow-dominated regions, such as the Yukon and Mackenzie in the north part of America and the Lena in the northeast of Asia. These results confirm the investigations by *Schellekens et al.* (2017), who compared the mentioned models against the Interactive Multi-sensor snow and Ice Mapping System (IMS, *Ramsay*, 1998). Multiple snow layers of SURFEX-TRIP helps it to better simulate snow dynamics during the cold seasons.

In Fig. 7.8 it can be seen that SURFEX-TRIP received the highest averaged weights (compared to other models) within the Amazon and Brahmaputra River Basins during 2002–2012, where surface water storage changes is the main components of TWSC (*Chen et al.*, 2009; *Khandu et al.*, 2016). In the Amazon River Basin, SURFEX-TRIP is also found to perform well between 2009–2011, during which the extreme floods (*Chen et al.*, 2009) increased the inter-annual magnitude. TWSC within the Amazon is also closely connected to the ENSO events in the tropical Pacific (*Kousky et al.*, 1984; *Ropelewski and Halpert*, 1987). Later in Section 7.8, it will be seen that surface water derived from SURFEX-TRIP is well correlated with the ENSO index compared to other models of this study. This could be another reason that the highest weights are obtained for SURFEX-TRIP between 2009–2011 within the Amazon River Basin.

Moreover, Fig. 7.8 indicates that within the river basins with considerable irrigation activities (such as the Indus, Euphrates, and Orange River Basins), the relatively highest weights are assigned to the LISFLOOD and ORCHIDEE, where both account for human water-use (*Schellekens et al.*, 2017). ORCHIDEE is also found to perform well within the Brahmaputra, Ganges, and Murray River Basins, each of which experienced a strong decline in rainfall over the entire period of this study (e.g., 9.0 ± 4.0 mm/decade between 1994–2014 over Ganges and Brahmaputra *Khandu et al.*, 2016). Specifically, ORCHIDEE contains 14 soil layers (see Table 2.1) that help it to better resolve vertical water exchange within the irrigated regions. *Mehrnegar et al.* (2020a) indicated that GRACE TWSC changes within the Murray River Basin are considerably influenced by ENSO events (see also *Forootan et al.*, 2012, 2016), and the simulated outputs of ORCHIDEE reflects these changes better than the other tested models justifying the higher weights that are assigned to this model within the DMDA procedure. Later in Section 7.7, it will be seen that after applying the DMDA, model-derived TWSC is tuned to the GRACE TWSC.

An overall interpretation of Fig. 7.6 is that those models with smaller RMSD, when compared with GRACE TWSC, receive bigger weights. However, some exceptions are found. For example, in the Ganges River Basin, LISFLOOD received the biggest weight, while the RMSD between this model and GRACE TWSC is found to be 90.07 mm, which is larger than that of PCR-GLOBWB, i.e., 87.86 mm, and W3RA, i.e., 80.98 mm. Although W3RA shows the best agreement with GRACE TWSC in terms of magnitude, it does not reproduce the trend that is evident in GRACE data. In contrast, LISFLOOD is found to be the only model that reproduced the trend, and as a result, is weighted the highest by the DMDA procedure.

7.6 Linear Trends of DMDA Water Storage Estimates

Estimated weights for the six models (presented in Section 7.5) along with the computed state parameters $\hat{\theta}_t$ (see the flowchart of Fig. 4.1) are used to compute the DMDA-derived groundwater, soil water, and surface water storage. To extract the monotonic changes of water storage changes within the 33 river basins, long-term linear trends are fitted to the DMDA results that are shown in Fig. 7.9, and the numerical values are reported in Table 7.2.

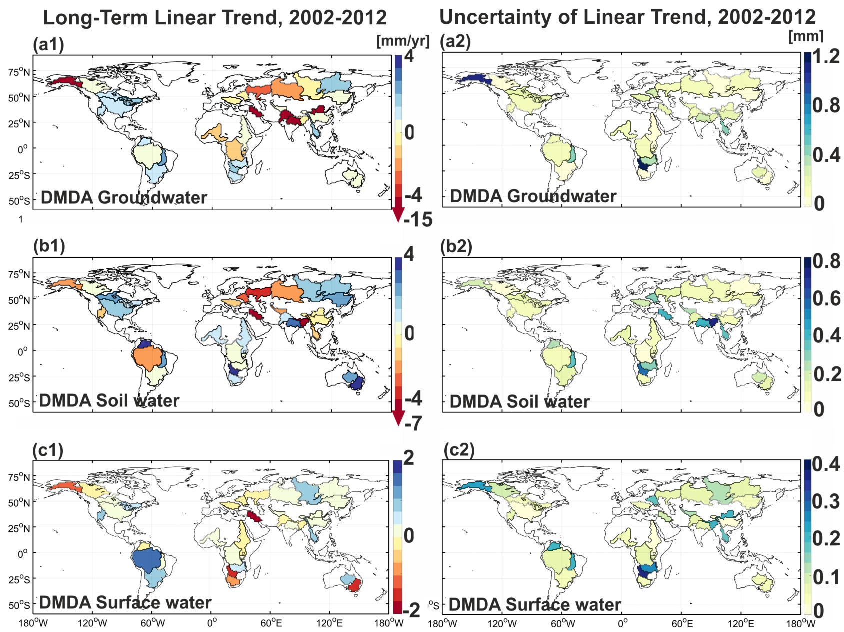

Fig. 7.9 Long-term (2002–2012) linear trends fitted to the DMDA-derived (a1) groundwater, (b1) soil water, and (c1) surface water components, expressed in mm/yr. The uncertainties of the linear trends are shown in (a2), (b2), and (c2), respectively.

Figure 7.9 (a1) and (a2) show the linear trends fitted to the DMDA-derived groundwater storage and their uncertainty. The results indicate decreasing groundwater storage in 42% of the assessed river basis (i.e., 14 of 33). The largest decreasing trends are found in basins with large-scale irrigation, such as the Ganges (-14.77 ± 0.25 mm/yr), Indus (-8.26 ± 0.16 mm/yr) and Euphrates (-5.36 ± 0.23 mm/yr). The results confirm findings by *Khandu et al.* (2016), *Forootan et al.* (2019), and *Voss et al.* (2013), respectively. The strongest increasing trends in groundwater are seen in the Tocantins basin (South America), Okavango (South Africa), and Lena (northeast Asia). However, all of these trends are not physically significant (less than 3 mm/yr). The positive trends in groundwater storage in these last two basins are associated with the heavy rainfalls, seasonal floods and the geographical location of the Okavango Delta (*McCarthy et al.*, 1998), and underground ice melting caused by global warming (*Dzhamalov et al.*, 2012), respectively. Comparison between linear trends fitted to the DMDA-derived groundwater storage and those of model outputs (Fig. 7.3) indicates that after merging GRACE TWSC with multiple hydrological model outputs, linear trends fitted to the groundwater storage changes has been modified considerably. This means that introducing GRACE data can successfully modify the anthropogenic effects, which are not well simulated by models.

Table 7.2 The amplitude of long-term linear trends [mm/yr] and their uncertainty, fitted to the DMDA-derived groundwater, soil water, and surface water storage, within the world's 33 largest river basin from 2002 to 2012.

Basin		DMDA	DMDA	DMDA
ID	**Name**	**Groundwater**	**Soil water**	**Surface water**
1	Amazon	0.17 ± 0.12	-1.92 ± 0.09	1.43 ± 0.06
2	Amur	0.46 ± 0.06	2.61 ± 0.09	0.25 ± 0.03
3	Aral	0.02 ± 0.08	-1.43 ± 0.22	0.21 ± 0.12
4	Brahmaputra	-0.44 ± 0.16	-7.00 ± 0.69	-0.13 ± 0.21
5	Caspian-Volga	-2.06 ± 0.15	-2.98 ± 0.16	-0.02 ± 0.07
6	Colorado	0.80 ± 0.11	-0.75 ± 0.09	0.82 ± 0.08
7	Congo	-0.72 ± 0.08	0.59 ± 0.03	0.06 ± 0.06
8	Danube	-0.47 ± 0.18	-0.75 ± 0.21	-0.08 ± 0.04
9	Dnieper	-0.5 ± 0.29	-2.27 ± 0.28	-0.03 ± 0.18
10	Euphrates	-5.36 ± 0.23	-5.75 ± 0.39	-2.09 ± 0.09
11	Lake Eyre	0.55 ± 0.16	2.42 ± 0.19	0.77 ± 0.04
12	Ganges	-14.77 ± 0.25	2.69 ± 0.40	0.29 ± 0.05
13	Indus	-8.26 ± 0.16	1.10 ± 0.13	-0.06 ± 0.07
14	Lena	1.74 ± 0.11	1.94 ± 0.05	0.20 ± 0.08
15	Mackenzie	0.51 ± 0.06	0.12 ± 0.05	-0.05 ± 0.10
16	Mekong	1.58 ± 0.43	-0.79 ± 0.33	0.83 ± 0.17
17	Mississippi	1.25 ± 0.09	1.36 ± 0.09	0.33 ± 0.02
18	Murray	0.06 ± 0.06	6.66 ± 0.15	-1.47 ± 0.04
19	Nelson	0.70 ± 0.18	2.45 ± 0.15	0.11 ± 0.03
20	Niger	-1.14 ± 0.15	0.75 ± 0.15	0.32 ± 0.05
21	Nile	0.45 ± 0.06	0.77 ± 0.06	-0.05 ± 0.02
22	Ob	-1.42 ± 0.08	-1.54 ± 0.06	0.05 ± 0.07
23	Okavango	1.74 ± 1.31	3.92 ± 0.55	-1.42 ± 0.37
24	Orange	1.32 ± 0.05	1.28 ± 0.06	-0.85 ± 0.05
25	Orinoco	0.87 ± 0.11	3.45 ± 0.26	-0.22 ± 0.19
26	Parana	0.68 ± 0.08	0.03 ± 0.13	1.04 ± 0.04
27	St. Lawrence	1.49 ± 0.18	1.07 ± 0.07	0.48 ± 0.05
28	Tocantins	2.41 ± 0.47	2.37 ± 0.35	0.08 ± 0.21
29	Yangtze	0.55 ± 0.23	-0.30 ± 0.09	0.20 ± 0.02
30	Yellow	-3.50 ± 0.14	-0.27 ± 0.05	0.08 ± 0.21
31	Yenisei	-0.26 ± 0.07	1.79 ± 0.06	0.75 ± 0.11
32	Yukon	-4.73 ± 1.08	-1.52 ± 0.20	-1.11 ± 0.23
33	Zambezi	1.19 ± 0.38	0.65 ± 0.31	0.35 ± 0.25

Long-term linear trends fitted to the DMDA soil water storage and their uncertainty are shown in Fig. 7.9 (b1) and (b2). The strongest increasing trends in DMDA soil water storage changes are found within the Murray (Australia), Okavango, and Orinoco (South America) River Basins with rates of 6.66 ± 0.15, 3.92 ± 0.55, and 3.45 ± 0.26 mm/yr respectively, and the largest decreasing trends are found in the Brahmaputra and Euphrates with rates of -7.00 ± 0.69 and -5.75 ± 0.39 mm/yr.

Figure 7.9 (c1) and (c2) show the linear trends and their uncertainty fitted to the surface water storage estimated through the DMDA method. Linear trends of surface water within 28 out of the 33 river basins are found to be physically insignificant (values between -2 and +2 mm/yr). The strongest negative trends are found in the Euphrates, Murray, and Okavango River Basins with rates of -2.09 $\pm$ 0.09, -1.47 $\pm$ 0.04, and -1.42 $\pm$ 0.37 mm/yr respectively. In contrast, the largest positive trends are found within the Amazon and Colorado, at the rate of 1.43 $\pm$ 0.06 and 1.04 $\pm$ 0.04 mm/yr, respectively. The heavy flood during the summer of 2008–2009 (*Chen et al.*, 2010a; *Marengo et al.*, 2011), which was considerably larger than the temporal mean, likely caused this positive trend in the Amazon River Basin. Negative trends in all three water storage compartments of the Euphrates River Basin (groundwater -5.36 $\pm$ 0.23 mm/yr, soil water -5.75 $\pm$ 0.39 mm/yr, and surface water -2.09 $\pm$ 0.09 mm/yr) can be associated with both irrigation and long-term drought as shown by *Forootan et al.* (2017).

7.7 An overview of TWSC Derived from DMDA

DMDA TWSC, as a summation of DMDA-derived water storage components (i.e., groundwater, soil water, surface water, snow, and canopy) is compared with the TWSC derived from models within the world's 33 largest river basins. Figure 7.10 shows differences between linear trends fitted to the GRACE and (left panel) DMDA TWSC, and (right panel) the mean of original model outputs. The results indicate that except for the Aral, Danube, and Nile River Basins the difference between linear trends of DMDA TWSC and GRACE data, is smaller than that of between mean of the models and GRACE data (see also Table 7.3). For example, within the irrigated river basins of Ganges and Euphrates, the differences between linear trends of GRACE TWSC and the mean of the models are estimated to be ~ -20 and ~ -18 mm/yr, respectively, while the differences between GRACE TWSC and DMA TWSC are estimated less than -5 and -9 mm/yr, respectively in these regions. The numerical values corresponding to Fig. 7.10 are given in Table 7.3.

This table also shows the differences between the linear trends fitted to GRACE TWSC and individual models. These results indicate that, as expected, in 73% of the region of the study (24 of 33 major river basins) the DMDA TWSC can reproduce that of GRACE better than individual models. In the Aral River Basin, for example, the difference between linear trend of DMDA TWSC and GRACE TWSC is estimated to be $\sim$ 2 mm/yr, while the difference between HBV-SIMRWG TWSC and GRACE TWSC is smaller than the DMDA result (-1.38 mm/yr). This is possibly because the ORCHIDEE model gains the highest weights through the DMDA approach within this region, and therefore is the most contributed model within the DMDA results (see Fig. 7.8). However, the difference between linear trend of ORCHIDEE and GRACE TWSC is 95% larger than that of HBV-SIMREG ($\sim$ 4 mm/yr). According to the Table 2.1, ORCHIDEE only simulates soil water, surface water storage, and snow. It can be postulated that, although ORCHIDEE performs well within the irrigated regions to simulate soil water and surface water storage, its lack of groundwater storage can negatively affect the DMDA results.

Table 7.3 $\Delta Trend$ [mm/yr] between GRACE TWSC and 6 hydrological models, the mean of models, and DMDA TWSC, 2002–2012.

ID	Basin Name	PCR-GLOBWB	SURFEX-TRIP	LISFLOOD	HBV-SIMREG	W3RA	ORCHIDEE	Mean Model	DMDA
1	Amazon	4.65	14.28	5.75	6.81	8.70	7.78	7.99	5.27
2	Amur	-3.47	-3.15	-2.47	-1.38	-1.12	-3.98	-2.60	-1.03
3	Aral	2.13	4.23	-1.41	-1.38	2.85	3.88	1.72	1.92
4	Brahmaputra	-14.54	-13.45	-13.14	-17.90	-14.63	-18.62	-15.38	-9.63
5	Caspian-Volga	-3.70	-5.72	-4.44	-6.85	-6.23	-3.90	-5.14	-1.93
6	Colorado	2.37	2.40	5.65	4.11	2.48	2.92	3.32	2.89
7	Congo	-1.74	-2.75	-5.17	-2.80	-3.23	-2.90	-3.10	-2.76
8	Danube	0.05	0.86	0.81	0.42	0.85	-1.61	0.23	0.95
9	Dnieper	-5.02	-5.96	-6.08	-6.38	-4.75	-8.10	-6.05	-2.30
10	Euphrates	-16.82	-18.63	-17.34	-19.16	-17.98	-17.57	-17.92	-8.83
11	Lake Eyre	0.48	2.78	2.55	3.95	2.53	1.96	2.37	1.37
12	Ganges	-24.84	-24.34	-3.36	-22.89	-20.12	-25.53	-20.18	-4.21
13	Indus	-12.04	-10.92	13.64	-10.06	-11.31	-14.26	-7.49	-3.60
14	Lena	1.91	-0.93	2.65	2.55	2.80	-2.10	1.15	-0.48
15	Mackenzie	0.39	0.38	-0.20	-1.42	-0.16	-0.56	-0.26	-0.81
16	Mekong	-6.76	-5.94	-8.02	-7.50	-8.02	-7.69	-7.32	-9.03
17	Mississippi	-3.05	0.53	1.71	1.64	1.69	0.01	0.42	0.64
18	Murray	-0.66	2.46	1.66	3.87	2.90	2.41	2.11	0.88
19	Nelson	-0.74	-1.10	3.85	5.71	3.06	2.70	2.25	-0.02
20	Niger	4.76	5.19	4.96	5.27	5.80	4.00	5.00	4.23
21	Nile	0.80	1.56	0.92	0.77	0.94	-5.67	-0.11	0.30
22	Ob	0.53	-1.56	-0.91	-2.41	-2.43	-1.91	-1.45	0.12
23	Okavango	23.77	24.97	24.92	25.09	23.60	24.63	24.50	17.92
24	Orange	0.63	2.15	1.91	2.28	1.92	1.92	1.80	0.61
25	Orinoco	17.06	19.74	13.46	16.82	21.97	15.07	17.35	8.80
26	Parana	-0.73	-3.46	-0.53	-0.85	-1.05	-6.93	-2.26	-0.78
27	St. Lawrence	1.68	3.11	3.59	3.09	4.67	1.15	2.88	1.09
28	Tocantins	5.41	1.89	2.73	1.41	2.40	-0.74	2.18	1.87
29	Yangtze	8.02	6.04	4.24	4.83	4.83	10.96	6.49	3.58
30	Yellow	-7.43	-6.42	-3.12	-4.83	-6.99	-5.86	-5.78	-1.54
31	Yenisei	4.66	4.26	4.64	5.37	5.06	4.54	4.75	2.11
32	Yukon	-23.89	-15.33	-24.78	-29.48	-23.86	-23.35	-23.45	-11.85
33	Zambezi	16.88	18.32	19.24	19.31	18.78	17.30	18.30	15.25

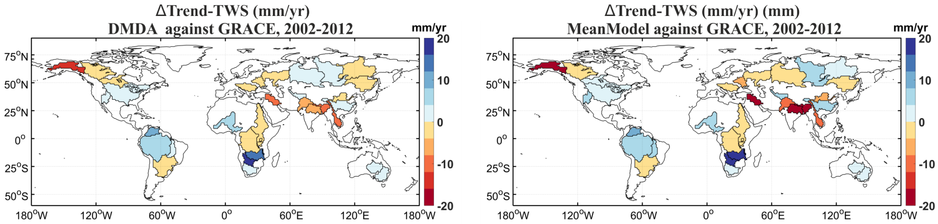

Fig. 7.10 An overview of the difference between the linear trends ($\Delta Trend$) fitted to the GRACE TWSC and: (left) the DMDA-derived TWSC, (right) the mean of 6 models over the world's 33 largest river basins covering the period of 2002–2012.

Figure 7.11 shows the RMSD between DMDA-derived TWSC and GRACE TWSC, as well as the RMSD between the mean of TWSC derived from original models and that of GRACE data. The results indicate that the magnitude of differences is considerably decreased after applying the DMDA (see a summary of the corresponding statistics in Table 7.1). Therefore, DMDA is effective in introducing missing components to model simulations, e.g., long-term linear trends and cyclic (e.g., seasonal) components, contributing to addressing the defects which were previously reported in studies such as *Scanlon et al.* (2018).

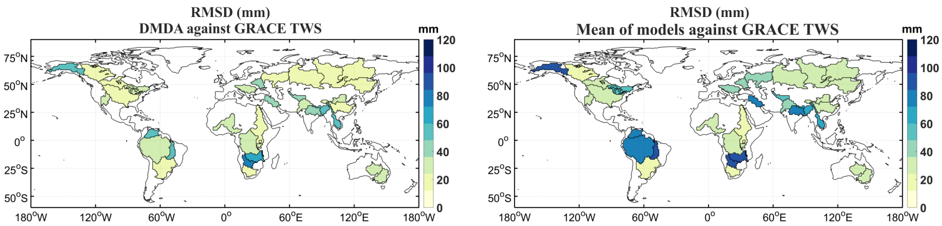

Fig. 7.11 The Root Mean Squares of Differences (RMSD) between GRACE TWSC and TWSC derived from DMDA (left), and the mean of 6 models (right) over the world's 33 largest river basins covering the period of 2002–2012.

7.8 Contribution of ENSO to the Changes of Water Storage Components

The El Niño Southern Oscillation (ENSO) is a dominant climate mode that results from ocean–atmosphere interactions over the equatorial Pacific (*Trenberth and Stepaniak*, 2001) and considerably influences precipitation and inter-annual TWSC in various regions (*Anyah et al.*, 2018; *Awange et al.*, 2014; *Chen et al.*, 2010a; *Fasullo et al.*, 2013; *Forootan et al.*, 2016, 2019; *Ni et al.*, 2018; *Zhang et al.*, 2015). Previous studies explored the influence of ENSO on GRACE TWSC and the global water balance estimates (*Eicker et al.*, 2016; *Forootan et al.*, 2018; *Phillips et al.*, 2012). Their results indicate that the ENSO teleconnection patterns are well reflected in the GRACE signal, whereas more pronounced variability is found by GRACE than the reanalysis water flux estimates shown by *Eicker* *et al.* (2016). Therefore, it is expected that by merging multi-model outputs with GRACE data, their skill in representing ENSO related water storage change will be improved.

7.8.1 Extracting the ENSO Modes from Water Storage Changes

To show this impact, the ENSO mode is extracted from water storage changes and compared with the ENSO index. The extraction is implemented here by applying the Principal Component Analysis (PCA, *Von Storch and Zwiers*, 2001, see also Appendix A.1). Before this implementation, the long-term linear trends and cyclic changes (with annual and semi-annual periods) are removed from the time series due to their dominant variance so that we can focus on inter-annual and multi-year cyclic time-scales.

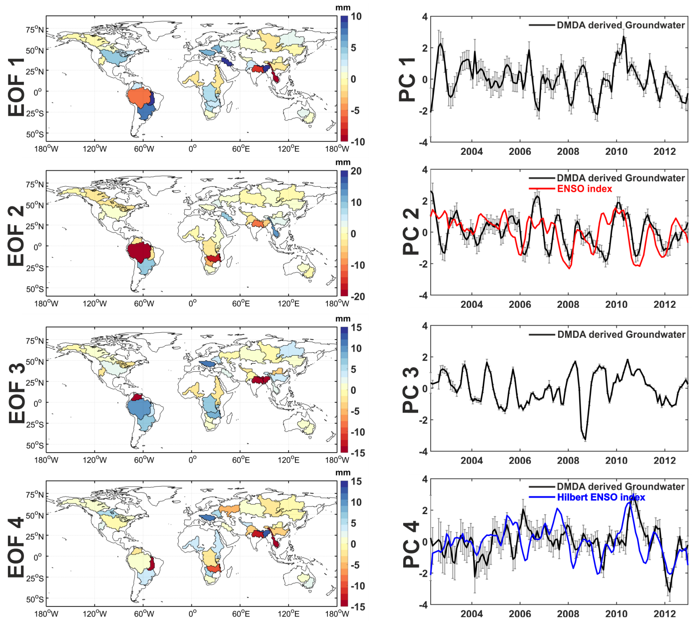

Fig. 7.12 The first four EOFs and PCs of the DMDA-derived **groundwater storage** after removing the linear trends and seasonal cycles (74% of total variance), between 2002–2012. EOF1 and PC1 explain 17% of the residual signal. EOF2 and PC2, as well as EOF4 and PC4 that respectively represent 15% and 10% of the residual signal capture the influence of ENSO on global groundwater changes. EOF3 and PC3 carry 9% of the residual signal. The uncertainty of the DMDA-derived groundwater storage, after removing linear trends and seasonal cycles is shown by the grey error bars fitted to the PC curve.

Figures 7.12, 7.13, and 7.14 show the spatial patterns (the first four dominant Empirical Orthogonal Functions, EOFs) and their corresponding temporal patterns (Principal Components, PCs) derived from the DMDA groundwater, soil water, and surface water storage estimates for the entire period of 2002–2012. The results indicate that the second and fourth modes of both groundwater and soil water, and the first and third modes of the surface water capture the influence of ENSO on global water storage changes, which are known here as the ENSO modes of water storage change. PC2 derived from the DMDA groundwater and soil water estimates, as well as PC1 derived from the surface water, are found to be in-phase with the ENSO index, while PC4 (derived from the DMDA groundwater and soil water) and PC3 (derived from the DMDA surface water) follow its out-of-phase evolution, i.e., the Hilbert transformation (*Horel*, 1984) of the ENSO index (more details can be found in *Forootan et al.*, 2018). By comparing Figs. 7.12 and 7.13, it can be seen that the magnitude of soil water

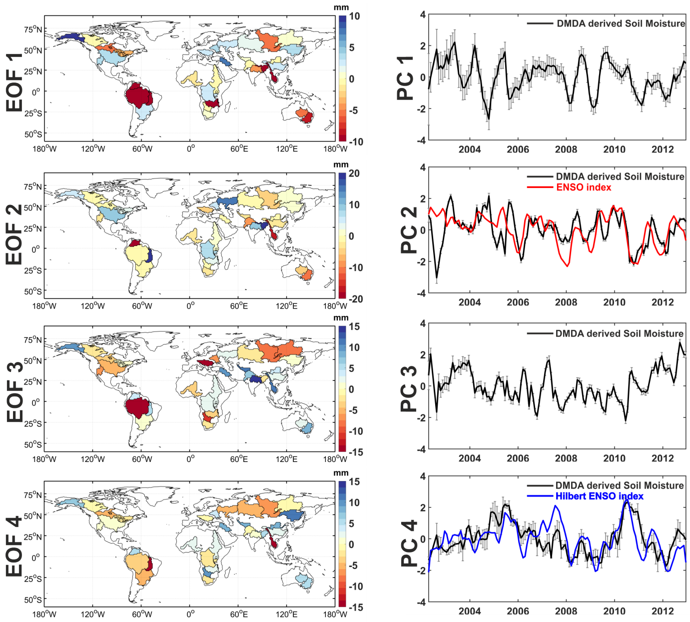

Fig. 7.13 The first four EOFs and PCs of the DMDA-derived **soil water** after removing the linear trends and seasonal cycle (74% of total variance), covering the period of 2002–2012. EOF1 and PC1 explain 19% of the residual signal. EOF2 and PC2, as well as EOF4 and PC4 respectively correspond to 15% and 10% of the residual signal and capture the influence of ENSO on global soil water storage changes. EOF3 and PC3 indicate 12% of the residual signal. The uncertainty of the DMDA-derived soil water storage, after removing linear trends and seasonal cycles is shown by the grey error bars fitted to the PC curve.

storage changes due to ENSO is higher than that of groundwater, particularly in the north of Asia and Australia. This is likely due to the stronger interactions (coupling) between rainfall (climate) and soil water (than that of rainfall and groundwater) in these regions. In some regions, e.g., Amazon and Zambezi, we can see that the effects of ENSO on groundwater storage is much bigger than that of soil water storage.

7.8.2 Correlation Coefficients of Water Storage Changes with the ENSO Index

Figure 7.15 shows temporal correlation coefficients between the ENSO index and the ENSO modes of groundwater and soil water storage derived from DMDA and the original model outputs. Maximum and minimum correlation of 0.75 and 0.53 corresponding to a maximum lag of up to 2 months are found globally between the DMDA groundwater and the ENSO index, respectively. Smaller correlation coefficients are found between those of the original models and the ENSO index. Among these models, W3RA and HBV-SIMREG indicate stronger correlations (~ 0.6 and ~ 0.4, respectively) with the ENSO index with a maximum lag of 2 months. Other models such as LISFLOOD and SURFEX-TRIP indicate notably different values. Small positive correlations are found with a maximum value of 0.3 between the original PCR-GLOBWB's groundwater and the ENSO index.

In Fig. 7.15, it can be seen that the correlation coefficients of the soil water storage derived from SURFEX-TRIP and LISFLOOD models are the highest, in comparison to those of DMDA and other model outputs, i.e., correlations of 0.6 to 0.8 within the 33 river basins examined here. PCR-GLOBWB and W3RA show a correlation coefficient of ~ 0.5, while those from HBV-SIMREG and ORCHIDEE are different from the other estimates, for example, less than 0.1 in the Niger and Nile River Basins and greater than 0.75 in North Asia. *Khaki et al.* (2018b) indicate that over the Nile River Basin, all three water storage components (i.e., groundwater, surface water, and soil water) are strongly influenced by ENSO. Therefore, the derived correlation of 0.1 in the Nile River Basin from HBV-SIMREG is likely to be unrealistic. Moreover, the high correlation coefficients of ORCHIDEE and SURFEX-TRIP with the ENSO index could be due to the fact that they have the largest number of soil layers (11 and 14, respectively) in their structure (*Schellekens et al.*, 2017), compared to the other models. Thus, these two models may simulate soil water storage better than PCR-GLOBWB and W3RA.

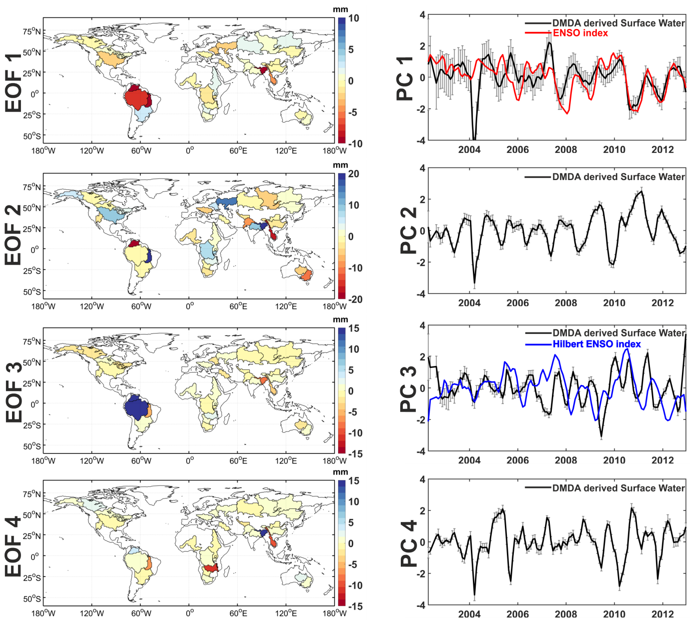

Fig. 7.14 The first four EOFs and PCs of the DMDA-derived **surface water** after removing the linear trends and seasonal cycles (54% of total variance) covering the period of 2002–2012. EOF1 and PC1 (26% of the residual signal) along with EOF3 and PC3 (13% of the residual signal) capture the influence of ENSO on global surface water changes. EOF2 and PC2, as well as EOF4 and PC4, explain to 15% and 9% of the residual signal, respectively. The uncertainty of the DMDA-derived surface water storage, after removing linear trends and seasonal cycles is shown by the grey error bars fitted to the PC curve.

Similar assessments are performed between the surface water storage changes and the ENSO index, and the results are shown in Fig. 7.16. Correlation coefficients up to 0.8 are computed from the DMDA estimates with a maximum lag of up to 2 months.

The DMDA-derived surface water storage is compared with those of PCR-GLOBWB, SURFEX-TRIP, and ORCHIDEE, which contain the surface water storage compartment. The correlation coefficients are found to be generally smaller than those of soil water and groundwater components (with a maximum of 0.5), which likely shows that the modelling of surface water needs improvement because in reality surface water in lakes and rivers within regions like East Africa shows an immediate response to ENSO (e.g., *Becker et al.*, 2010). Figure 7.16 shows that the surface water storage output of SURFEX-TRIP has the highest correlation coefficients with the ENSO index in all basins of America (values between 0.33 and 0.51) and Africa (values between 0.23 and 0.48), while ORCHIDEE shows the highest correlations (values between 0.32 and 0.58) in most parts of Asia. The correlation coefficients for PCR-GLOBWB are found to be relatively smaller, i.e., between 0.1 and 0.2 with lags of between 5-12 months.

Comparisons between the DMDA and original model outputs, in terms of correlation coefficients with ENSO index in Figs. 7.15 and 7.16, indicate that combining models with GRACE data improve the correlations with the ENSO index, and the correlation lags are considerably reduced globally. It is

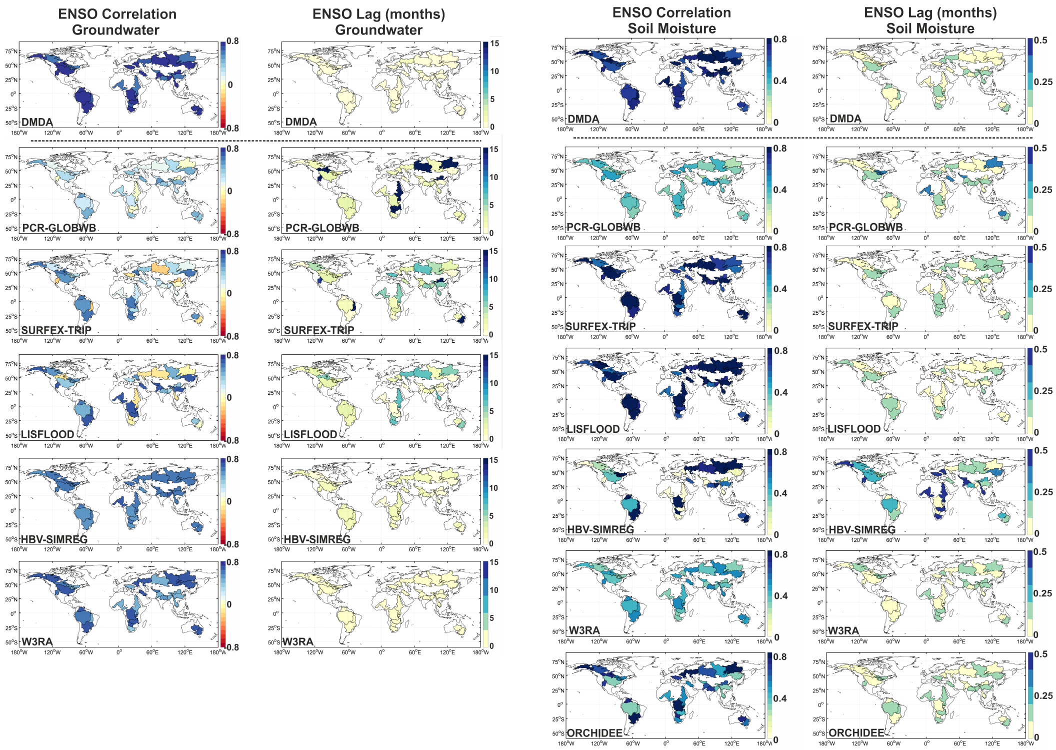

Fig. 7.15 Correlation coefficients and their lags between the ENSO index and groundwater and soil water storage derived from the DMDA method and hydrological model outputs used in this study for the period of 2002–2012.

worth mentioning that the DMDA results that are presented here are derived by setting the α value in Eq. (4.16) to 0.9. This means that 36-month temporal correlations are assumed between water storage simulations of the six models. This value guarantees an extraction of the dominant part of ENSO using two PCA modes after merging GRACE and model outputs.

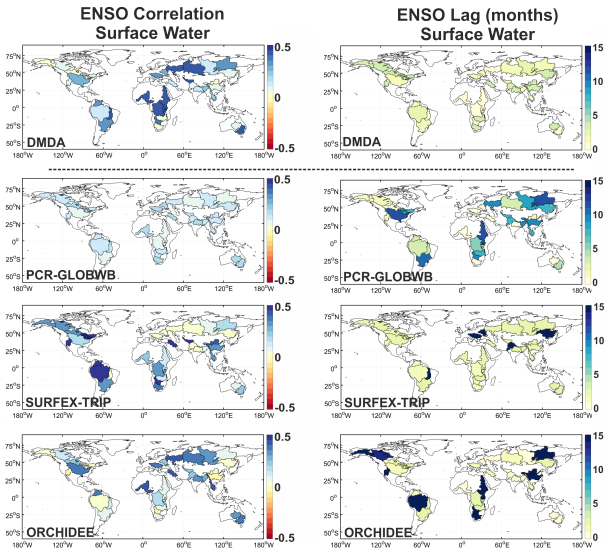

Fig. 7.16 Correlation coefficients and their lags between the ENSO index and surface water estimates derived from the DMDA method and hydrological models used in this study from 2002 to 2012.

7.9 Summary and Conclusion

In this chapter, the application of DMDA was examined (1) to compare multi-model (individual) water storage simulations against GRACE-derived Terrestrial Water Storage Changes (TWSC); and (2) to separate GRACE TWSC into its hydrological compartments within the world's 33 largest river basins.

Numerically, the DMDA method was implemented by integrating the output of six global hydrological and land surface models (*Schellekens et al.*, 2017), i.e., PCR-GLOBWB, SURFEX-TRIP, LISFLOOD, HBV-SIMREG, W3RA, ORCHIDEE, and monthly GRACE TWSC (2002–2012) within the world's 33 largest river basins, while considering the inherent uncertainties of all inputs.

DMDA provides time-variable weights (see Section 4.3) to compute an average of multiple *a priori* information, yielding the best fit to the observation. These weights can also be used to understand the behaviour of *a priori* information (which here refers to the output of hydrological models) against the observations (GRACE(-FO) TWSC) while considering their error estimates. To test this hypothesis, a realistic synthetic example was defined to evaluate the performance of DMDA (Fig. 4.2). This analysis showed that the method is able to correctly separate GRACE TWSC estimations into its individual hydrological compartments. It was also shown that the DMDA's estimation of temporal weights (for each model) is close to reality and can be used to assess the performance of models.

Temporal weights estimated for the six global hydrological and land surface models in this study were interpreted in Section 7.5 for different regions of the world. The Relationship between various climate events in different regions during the last two decades and the structure of the models were used as an evidence to confirm finding of this study. From the results, it can be concluded that DMDA time-variable weights allow us to assess each model's performance during various time steps and to explore which model can best simulate water storage changes due to various climate events.

The application of DMDA to merge six global hydrological model outputs with GRACE data showed that this approach can deal with models with different structures, and performed well to update water storage simulations with respect to GRACE data. Based on these results, we have gained confidence in this method for improving the characterization of water storage over broad regions of the globe using GRACE data. In what follows, the main conclusions and remarks of this study are summarized.

- Estimated weights (Fig. 7.8) showed that the PCR-GLOBWB model gained the largest weights; thus, it contributed the most in the DMDA-derived TWSC in North Asia, North America, and the centre of Africa. SURFEX-TRIP performed best within basins with dominant surface water storage changes, as well as in snow-dominant regions. The LISFLOOD and ORCHIDEE models were found to perform well within irrigated basins and those affected by ENSO events.
- DMDA results in Fig. 7.9 (a1) showed that considerable trends exist in groundwater storage changes within the Ganges, Indus, and Euphrates River Basins during 2002–2012. These changes were dominantly influenced by anthropogenic modifications. Trends in soil water storage changes (Fig. 7.9 (b1)) were found to be mostly related to prolonged drought events such as those in the Brahmaputra and Euphrates River Basins.
- Figures 7.15 and 7.16 showed that the ENSO mode of water storage variability in most of the world's 33 largest river basins are improved after merging GRACE TWSC with individual model outputs using the DMDA approach. DMDA assigned the largest corrections of the ENSO mode in groundwater to the Nile, Murray, Tocantins, Ob, Okavango and Orange River Basins. The highest improvement in the ENSO modes of soil water storage were found within the Nile, Niger, Zambezi, and Amur River Basins, and those of the surface water storage were found within the Nile, Niger, Congo, Tocantins, and Murray River Basin. For example, the correlation coefficient between groundwater storage and ENSO in the Murray River Basin changed from -0.2 to 0.6, and in the Nile River Basin from 0.1 to 0.4 for soil water, and from 0.3 to 0.7 for the surface water components.

Chapter 8

Application of MCMC-DA to Merge GRACE with W3RA Water Balance Model over CONUS

8.1 Introduction

The objective of this chapter is to implement the MCMC-DA approach (formulated in Chapter 5) to explore high resolution soil water and groundwater storage changes across the Conterminous United States (CONUS) (with $0.125°$ spatial resolution), with an emphasis on the use of a relatively simple water balance model. To this aim, the Worldwide Water Resources Assessment (W3RA, *Van Dijk*, 2010) model, which is simpler than the currently used NASA's Catchment Land Surface Model (CLSM, *Ducharne et al.*, 2000), is selected as the platform to be merged with GRACE TWSC, which is adapted here by defining the CONUS boundary as its domain, during 2003–2017 (see also Section 2.4). W3RA is selected both due to its computational load being manageable for scientific applications, and its reasonable performance compared to the other commonly used global hydrological or land surface models *Schellekens et al.* (2017).

GRACE TWSC and its full error covariance matrix are obtained following the computational steps in Section 2.2.8, on a $0.125° \times 0.125°$ spatial grid points within the CONUS, for the period 2003–2017.

In this chapter, it will be shown that, by formulating the rigorous Bayesian data-model integration (MCMC-DA), GRACE TWSC improves model estimates of both soil water storage and groundwater storage changes in terms of trends and seasonality, which are not well simulated by most of the available models (*Scanlon et al.*, 2019). Beyond long-term trends and seasonality, the effect of other climate processes such as that of the El Niña Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) will be explored. The results are published in *Mehrnegar et al.* (2020b).

In what follows, after providing an overview on the importance of the improved estimation of groundwater and soil water storage changes within CONUS (Section 8.2), TWSC derived from observations (GRACE) and model (W3RA) are compared in terms of their linear trends and annual cycle in Section 8.3. A comparison between DMDA (proposed in Chapter 4) and MCMC-DA to merge W3RA model outputs with GRACE observations is provided in Section 8.4. The MCMC-DA groundwater and soil water storage estimates are then compared with those of the original model outputs (Sections 8.5 and 8.6, respectively), while the possible relationships between the storage changes and climatic and anthropogenic factors are evaluated. Validations are done against independent measurements, i.e., in-situ USGS groundwater level observations, as well as soil water storage data from the European Space Agency (ESA)'s Climate Change Initiative (CCI) (introduced in Section 2.5). Evaluations using the groundwater levels are done after standardising the available time series. Within Texas and California, where reliable information is available, the equivalent groundwater storage estimates are computed from the USGS level observations, and are used to evaluate groundwater storage estimates from MCMC-DA and the original model within these states.

In order to extract the influence of ENSO on groundwater and soil water storage estimates, both Principal Component Analysis (*Von Storch and Zwiers*, 2001, PCA,) and Independent Component Analysis (ICA, *Forootan and Kusche*, 2012, 2013; *Forootan et al.*, 2018) can be applied. PCA (Appendix A.1), as the second order statistical decomposition techniques, was applied to extract the ENSO modes from DMDA water storage changes in Chapter 7. In this study, ICA (Appendix A.2), as the higher order statistical decomposition technique, is chosen to extract the ENSO modes from water storage changes, and the obtained result are compared with available ENSO indices and are interpreted in Section 8.7. Choosing ICA over PCA is due to its power to extract cyclo-stationary modes as demonstrated by *Forootan and Kusche* (2013).

Down-scaling GRACE TWSC using the outputs of the W3RA in the MCMC-DA approach is demonstrated in Section 8.8. Changes in groundwater and soil water storage within the Texas and California states, which are affected by anthropogenic modifications, are evaluated and interpreted in Sections 8.9 and 8.10, respectively. At the end of this chapter, a summary of the obtained results and a conclusion are presented in Section 8.11.

8.2 Hydrological Properties of Conterminous United States (CONUS)

Over the past decades, climate variability and change along with anthropogenic modifications and land management activities have affected water resources across the Conterminous United States (CONUS)

(*Fasullo et al.*, 2016; *Rodell et al.*, 2018). For example, changes in rainfall patterns, and, therefore, change in the annual (net-)precipitation has resulted in several extreme flood and drought events across the country (*Dong et al.*, 2011; *Leng et al.*, 2016; *Peterson et al.*, 2013; *Schubert et al.*, 2004). In addition, the population has increased with an average annual growth of 0.67% between 2010 and 2019, mostly in the south and west, and the trend is expected to continue until 2050 (*Means III et al.*, 2005; *Potter and Hoque*, 2014).

Groundwater across the CONUS accounts for almost half of the water consumed for irrigation, livestock, and drinking water (including public domestic water supply) (*Dieter*, 2018). In arid and semi-arid regions throughout much of the CONUS, groundwater is the only potential freshwater resource (*Maupin et al.*, 2017). As expected from the current hydro-climatological conditions, groundwater depletion has been reported in irrigated regions, such as the Central and Southern High Plains in Kansas and Texas and the Central Valley aquifers in California (*Famiglietti et al.*, 2011; *Scanlon et al.*, 2012a). Continuing the current negative trends in groundwater resources in these regions presents a dire threat to future crop production, natural stream-flow, groundwater-fed wetlands, saltwater intrusion, and related ecosystems (*Scanlon et al.*, 2010).

Soil water is another key variable in the water cycle that can be used as a measure of the land-atmosphere feedback (*Levine et al.*, 2016). In general, soil water conditions contribute to the natural and agricultural productivity of a region by defining the vadose zone water that is available for uptake into vegetation (*Hillel*, 1998; *Illston et al.*, 2004). In turn, water is transpired from vegetation to the atmosphere during photosynthesis, increasing low-level atmospheric moisture at various scales. Therefore, information about soil water changes has been used to predict changes in precipitation (*Brocca et al.*, 2017), climatic extremes and future climate projections (*Bolten et al.*, 2009; *Mo et al.*, 2011). Soil water information is now used for monitoring and understanding drought development (see e.g., www.drought.gov).

8.3 Comparison between GRACE and W3RA TWSC

Before concentrating on the results of MCMC-DA and their evaluation with independent data, an overview of the comparison of variance between GRACE TWSC (used as observations) and W3RA (whose individual water storage estimates are used as *a priori* information in the MCMC-DA approach) is provided in this section. Grid-based linear trends and annual amplitudes of TWSC correspond to 36% and 53% of the total variance over the CONUS, respectively (see Fig. 8.1). Groundwater and soil water storage changes of the original W3RA model and those updated by the MCMC-DA approach (Figs. 8.4 and 8.10) and their interpretation are presented in the following sections. To compute the linear trend along with annual and semi-annual cycles, a regression equation $l(t) = a + b\,t + c\,\cos(2\pi t) + d\,\sin(2\pi t) + e\,\cos(4\pi t) + f\,\sin(4\pi t) + \varepsilon(t)$ is used. Here, $l(t)$ contains

TWSC time series in each grid (as in Fig. 8.1) or individual water storage time series, and t is a time vector that varies between 2003 and 2017. To estimate the unknown coefficients (shown by $\hat{a}$, $\hat{b}$, $\hat{c}$, $\hat{d}$, $\hat{e}$, and $\hat{f}$), a least-squares approach is used while assuming $\varepsilon(t)$ to be the vector of residuals. $\hat{b}$ is considered as the linear trend (in mm/yr) and $\sqrt{\hat{c}^2 + \hat{d}^2}$ as the amplitude of the annual cycle (in mm). The non-parametric Wilcoxon-Mann-Whitney statistical test (WMW, *Fay and Proschan*, 2010) with 95% and 99% significant levels are used to evaluate the statistical significance of the estimated unknowns.

TWSC are shown in Fig. 8.1, where considerable differences are found between the original W3RA TWSC and GRACE TWSC in terms of both linear trend (compare Fig. 8.1 (A1) and Fig. 8.1 (B1)) and annual-amplitude (compare Fig. 8.1 (A2) and Fig. 8.1 (B2)) within different parts of the CONUS, where large differences in the linear trend are found in the west and south, e.g., California and Texas that are influenced by irrigation and drought (*Famiglietti et al.*, 2011; *Rodell et al.*, 2018; *Scanlon et al.*, 2012a). GRACE indicates larger trends in these regions with the mean value of -9.47 ± 1.02 mm/yr in California and -7.65 ± 1.32 mm/yr in Texas, relative to those of W3RA with the mean value of -3.15 ± 0.86 mm/yr and -3.86 ± 1.24 mm/yr within California and Texas, respectively. Within the northern CONUS (e.g., Montana, North Dakota, and South Dakota), similar results are found with higher positive trends from GRACE ($\sim 9.48 \pm 1.78$ mm/yr) than those of W3RA ($\sim 2.61 \pm 0.86$ mm/yr). More pronounced changes in GRACE TWSC within the northern CONUS are attributed to high-intensity rainfall and flooding events (e.g., the 2011's flood in the Missouri River, *Reager et al.*, 2014; *Zhang and Schilling*, 2006; *Zheng et al.*, 2014).

Large differences in the annual-amplitude of TWSC are found within the Great Lakes and in the southeast of CONUS, with the mean values of 76.23 ± 4.08 mm derived from W3RA and 34.65 ± 3.21 mm from GRACE TWSC. In contrast, the annual-amplitude of W3RA TWSC within Florida is estimated up to 96.5 ± 5.34 mm, which is much greater than those estimated by GRACE TWSC

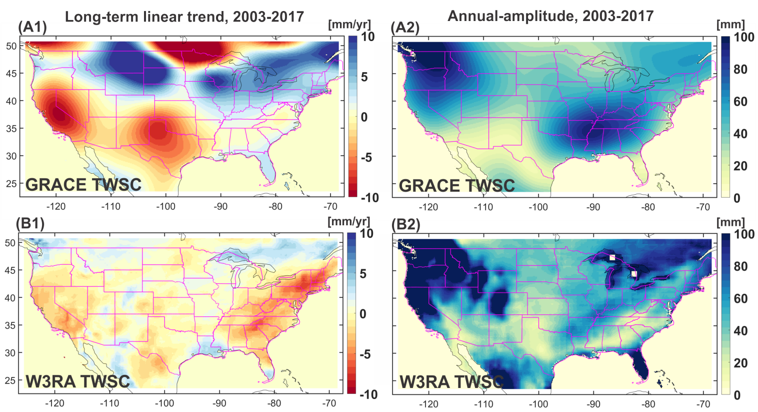

Fig. 8.1 Linear trend [mm/yr] and annual-amplitude [mm] fitted to the GRACE TWSC and the original W3RA TWSC changes within the CONUS, covering the period of 2003–2017. The plots in (A1) and (A2) correspond to GRACE TWSC, while (B1) and (B2) correspond to W3RA TWSC.

(20 ± 3.5 mm). The differences in the seasonality of TWSC, between modelled and measured, can be related to errors in the forcing data and uncertainty in model parameters to control these values (*Van Dijk et al.*, 2011; *Van Dijk*, 2010), which is the case for most available models as demonstrated by *Scanlon et al.* (2018). Comparison in Fig. 8.1 indicates that GRACE TWSC has the potential to be used for improving the W3RA's water storage simulations in terms of long-term trends and seasonality. It will also be shown that the improvement is even beyond these components, and some climate modes, such as that of ENSO is influenced after implementing MCMC-DA.

8.4 Comparison between DMDA and MCMC-DA TWSC within CONUS

The performance of the DMDA approach to merge W3RA model outputs and GRACE observation within CONUS is compared with those of MCMC-DA. This comparison is done in terms of the RMSD between modelled and measured TWSC (Fig. 8.2 (A), (B), (C)), and the phase differences for the annual amplitude of modelled TWSC with GRACE data (Fig. 8.2 (D), (E), (F)) before and after implementing DMDA and MCMC-DA. A visual representation of the median, first and third quartiles of the RMSD and phase differences of Fig. 8.2 can be seen in Fig. 8.3 (A) and (B), respectively. The numerical results indicate that the median of RMSD between modelled and measured TWSC is reduced by $\sim 50\%$, from 60 mm to 31 mm, after merging GRACE with W3RA through the DMDA approach. RMSD between MCMC-DA TWSC and GRACE data is estimated to be zero within CONUS, which indicates that MCMC-DA TWSC yields the best fit to GRACE data. Both DMDA and MCMC-DA performed well in reducing the phase differences between modelled and measured TWSC, where the median of phase differences between the annual amplitude of W3RA and GRACE TWSC is reduced from -50 deg to zero after implementing both techniques (see Fig. 8.2 (B)). Although DMDA performed well to reduce the phase differences between W3RA and GRACE TWSC in more than 70% of the CONUS, large values of phase differences (between ± 25 deg) still exist between modelled and GRACE data, mostly in the south and southeast CONUS (see Fig. 8.2 (E) and Fig. 8.2 (B)).

The DMDA soil water and groundwater storage are also validated against independent ESA CCI and USGS data (results are not shown here). The numerical results indicate that, compared to DMDA, the MCMC-DA estimates are on average 25% closer to the independent estimates. This improvement is likely gained by the dynamic estimation of temporal dependency between unknown state parameters in Eq. (4.2) that, compared to DMDA, introduces more realistic updates to the individual water storage components. This is especially true for the top soil layer, in which changes are dynamic and strongly coupled to the atmosphere rather than to groundwater change, where the hydrological memory of the region plays an important role in its evolution. In light of this assessment, the interpretations of this study are limited to those derived from MCMC-DA.

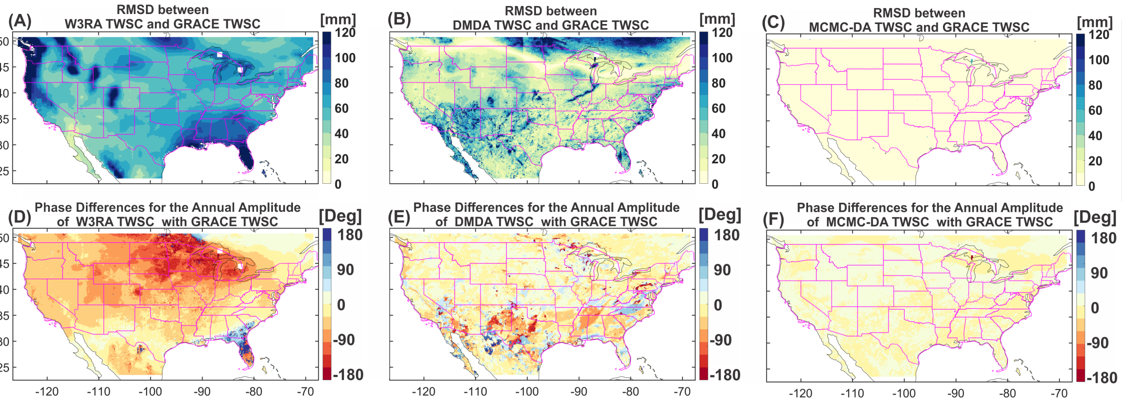

Fig. 8.2 RMSD [mm] between GRACE TWSC and (A) W3RA TWSC, (B) DMDA TWSC, and (C) MCMC-DA TWSC, and phase difference [Deg] for the annual amplitude of GRACE TWSC with (D) W3RA TWSC, (E) DMDA TWSC, (F) MCMC-DA TWSC, within CONUS, between 2003–2017. The blue horizontal line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points (blue points), which are not considered as outliers.

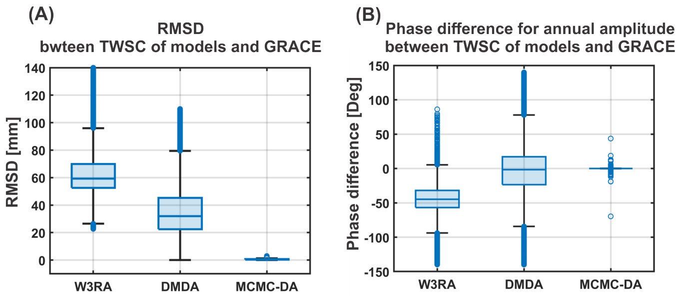

Fig. 8.3 Box plots illustrating the (A) RMSD between GRACE TWSC and modelled TWSC before and after implementing DMDA and MCMC-DA to merge W3RA with GRACE data, and (B) the phase difference for the annual amplitude of modelled TWSC with GRACE data, before and after merging W3RA with GRACE data through DMDA and MCMC-DA. This figure shows the box plots of RMSD and phase differences illustrated in Fig. 8.2.

In what follows, the performance of the MCMC-DA will be assessed in estimating individual water storage changes. The focus is to explore changes in groundwater storage, which is essential for water resources, and by integrating GRACE data into the model, it will be shown that the model's missing trends in the groundwater compartment will be introduced to the MCMC-DA result. Moreover, the soil water storage (top layer $< 10cm$) will be evaluated, which is important to understand the land-atmosphere interactions (*Brocca et al.*, 2017; *Levine et al.*, 2016). Previous Data Assimilation (DA) attempts (*Girotto et al.*, 2016) showed that a single GRACE DA might introduce unrealistic signals to the soil water storage compartments. Therefore, here, it will be examined whether such errors can be avoided using MCMC-DA, where the uncertainties and the dynamic evolution of water states are rigorously accounted for.

8.5 Groundwater Storage Changes Across the CONUS

Implementing MCMC-DA significantly modifies linear trends, as well as annual and semi-annual amplitudes of groundwater estimates. On average, changes in these components are found to be in the magnitude of 110%, 25%, 20% of the original W3RA estimates, respectively. The mean values of linear trends within the CONUS are reported in Table 8.1, which are estimated based solely on W3RA output versus those using MCMC-DA. The linear trend and annual-amplitude in

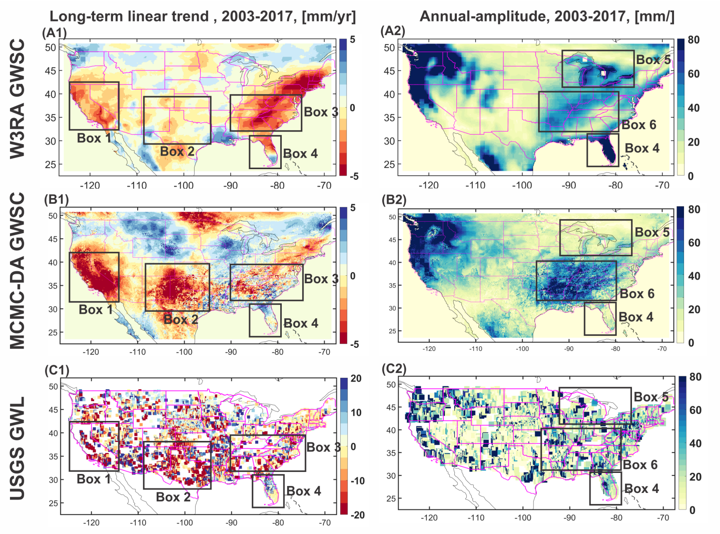

Fig. 8.4 Linear trend [mm/yr] and annual-amplitude [mm] fitted to the (A1 and A2) groundwater storage changes for the original W3RA; (B1 and B2) groundwater storage estimates from MCMC-DA; and (C1 and C2) the USGS groundwater level observations, covering the period of 2003–2017. In this figure, the terms GWSC and GWL are used to show groundwater storage changes and groundwater level, respectively. In order to enhance the visualisation, the black boxes in Fig. 8.4 (marked as Box 1, 2, 3, 4, 5, and 6) are used to show the regions with considerable groundwater changes after integrating GRACE TWSC into W3RA. The extension of these boxes is reported here as Box 1: $[112°W - 125°W, 32°N - 42°N]$, Box 2: $[95°W - 109°W, 30°N - 39°N]$, Box 3: $[75°W - 90°W, 32°N - 39°N]$, Box 4: $[79°W - 86°W, 24°N - 31°N]$, Box 5: $[75°E - 90°E, 40°N - 50°N]$, and Box 6: $[79°W - 97°W, 30°N - 40°N]$.

groundwater storage are shown in Figs. 8.4 (A1) and (A2) and Figs. 8.4 (B1) and (B2) for W3RA and MCMC-DA, respectively. Similar plots from the USGS groundwater level observations are shown as an independent comparison in Figs. 8.4 (C1) and (C2). Six black boxes are highlighted in Fig. 8.4 to show the regions with considerable changes after integrating GRACE TWSC into W3RA compared to the original model outputs. The extensions of these boxes are reported in Fig. 8.4.

8.5.1 Linear Trend of MCMC-DA Groundwater Storage

Pronounced negative trends are found using the MCMC-DA approach in the southwestern (-3.89 ± 0.34 mm/yr, Box 1, California and Nevada) and south central CONUS (-2.83 ± 0.46 mm/yr, Box 2, Texas, New Mexico, Colorado, Kansas, and Oklahoma). These values are much greater in magnitude than those from the original W3RA results (-1.76 ± 0.54 in Box 1 and -0.10 ± 0.18 mm/yr in Box 2). The USGS groundwater level data (Fig. 8.4 (C1)) also show negative trends within these regions (Box 1, -18.39 ± 1.96, and Box 2, -16.55 ± 2.04 mm/yr). Estimated linear trends of both USGS groundwater level and MCMC-DA groundwater storage are found to be statistically significant within Box 1 and Box 2 based on WMW statistical test results with a 99% significance level. As USGS observations are not converted to the storage equivalence, the long-term linear trends of standardized USGS groundwater level are used to validate standardized MCMC-DA groundwater storage and those of the original model outputs (Fig. 8.5). These data sets are standardized by subtracting their temporal mean and dividing the residuals by their standard deviations. Linear trends fitted to the standardized

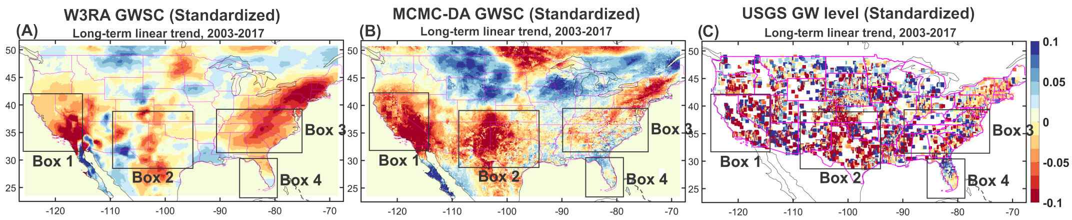

Fig. 8.5 Linear trend fitted to the standardized (A) groundwater storage changes (GWSC) from the original W3RA, (B) MCMC-DA, and (C) USGS groundwater level observations cover the period 2003–2017.

MCMC-DA groundwater storage and USGS groundwater levels in Fig. 8.5 show considerably larger values of groundwater depletion in west and south CONUS (Box1 and Box2) between 2003-2017. W3RA, however, shows groundwater depletion in these region by $\sim 50\%$ less than two other data

Table 8.1 The mean value of linear trends fitted to the W3RA and MCMC-DA groundwater storage estimates. The results (expressed in mm/yr) are computed for CONUS covering the period of 2003–2017.

State ID, Name	W3RA	MCMC-DA	State ID, Name	W3RA	MCMC-DA
1-West Virginia	-3.69±0.39	**-1.39±0.36**	**25-Ohio**	-1.54±0.28	**0.38±0.15**
2-Florida	-1.01±0.43	**1.47±0.40**	**26-Texas**	-0.11±0.34	**-2.96±0.56**
3-Illinois	0.35±0.24	**1.7±0.44**	**27-Colorado**	-0.13±0.22	**-0.59±0.13**
4-Minnesota	-0.22±0.24	**-1.01±0.30**	**28-South Carolina**	-3.04±0.45	**1.03±0.30**
5-Washington D.C.	-2.83±0.31	**-1.36±0.24**	**29-Oklahoma**	-0.22±0.31	**-2.38±0.41**
6-Rhode Island	-3.42±0.31	**0.09±0.14**	**30-Tennessee**	-1.72±0.47	**-0.42±0.58**
7-Idaho	0.08±0.3	**0.34±0.19**	**31-Wyoming**	0.39±0.22	**1.56±0.20**
8-New Hampshire	-4.30±0.39	**-1.84±0.23**	**33-North Dakota**	-0.18±0.19	**0.22±0.12**
9-North Carolina	-2.65±0.43	**0.13±0.05**	**34-Kentucky**	-2.28±0.39	**-0.73±0.23**
10-Vermont	-3.46±38	**-1.48±0.22**	**38-Maine**	-1.46±0.39	**0.31±0.21**
11-Connecticut	-4.6±0.32	**-0.77±0.16**	**39-New York**	-3.27±0.35	**-0.93±0.21**
12-Delaware	-3.14±0.31	**-1.78±0.31**	**40-Nevada**	-1.03±0.19	**-1.65±0.15**
13-New Mexico	-0.10±0.22	**-2.67±0.33**	**43-Michigan**	0.41±0.25	**0.96±0.28**
14-California	-2.63±0.32	**-4.95±0.44**	**44-Arkansas**	0.02±0.39	**-0.06±0.19**
15-New Jersey	-3.82±0.31	**-0.62±0.30**	**45-Mississippi**	0.60±0.49	**0.26±0.83**
16-Wisconsin	0.45±0.17	**1.86±0.14**	**46-Missouri**	0.34±0.48	**1.08±0.54**
17-Oregon	-0.64±0.34	**-0.63±0.26**	**47-Montana**	0.73±0.22	**1.67±0.18**
18-Nebraska	0.15±0.20	**1.40±0.21**	**48-Kansas**	-0.18±0.23	**-1.26±0.44**
19-Pennsylvania	-3.63±0.35	**-1.37±0.27**	**49-Indiana**	-0.64±0.26	**0.74±0.40**
20-Washington	1.95±0.41	**0.95±0.32**	**51-South Dakota**	0.01±0.2	**0.69±0.22**
21-Louisiana	2.55±0.48	**1.12±0.42**	**52-Massachusetts**	-4.47±0.34	**-1.04±0.25**
22-Georgia	-2.64±0.49	**0.42±0.04**	**53-Virginia**	-2.61±0.40	**-0.63±0.42**
23-Alabama	-1.63±0.54	**-0.83±0.71**	**55-Iowa**	0.66±0.20	**1.51±0.27**
24-Utah	-0.38±0.20	**-0.59±0.22**	**56-Arizona**	-0.18±0.19	**-1.21±0.26**

sets. These results indicate that the impact of the last decade's droughts that is amplified by intensive irrigation is well reflected in GRACE data (see, e.g., *Diffenbaugh et al.*, 2015; *Faunt et al.*, 2016; *Mehta et al.*, 2013; *Rodell et al.*, 2018; *Scanlon et al.*, 2012a) but was not captured by the model as the human water-use component is not included in W3RA. However, this analysis demonstrates that integrating GRACE data into this modelling framework can improve this characterization of variability and trends. Temporal correlation coefficients between the USGS observation and W3RA (CC_{W3RA}) and MCMC-DA ($CC_{MCMC-DA}$) groundwater storage are presented in Fig. 8.6 (A) and (B), respectively. These results show that the correlation coefficients between USGS observation and W3RA groundwater estimation are less than -0.5 in 70% of CONUS (including Box1 and Box2) and this value increases to over ~ 0.4 after implementing MCMC-DA. Due to the fact that on average, changes in the linear trends and annual amplitudes of MCMC-DA groundwater are in the magnitude of 110%, and 25% of those of W3RA groundwater storage, it can be said that the improvement in the correlation coefficients generally occurred due to improvements in the linear trends and annual amplitudes after merging GRACE with W3RA. In the eastern CONUS (Box 3, including South

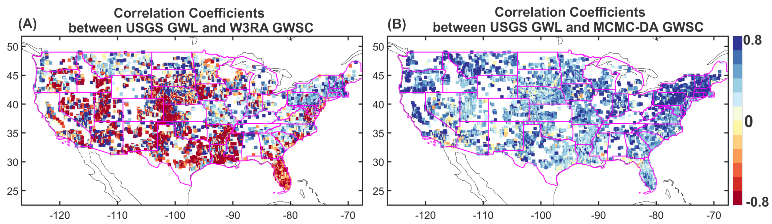

Fig. 8.6 Correlation coefficients (A) between the USGS groundwater level and W3RA groundwater storage (CC_{W3RA}); and (B) between the USGS groundwater level and MCMC-DA groundwater storage ($CC_{MCMC-DA}$). The results correspond to the period of 2003–2017.

and North Carolina, Virginia, Georgia, and Tennessee) W3RA shows a systematic negative trend in groundwater storage; however, incorporating GRACE using MCMC-DA adjusts these trends to a mixture of positive and negative values (compare Box 3 in Figs. 8.4 (A1) and (B1)) with patterns more similar to those from the USGS groundwater level data (see Box 3 in Fig. 8.4 (C1)). In Box 3 (southeast CONUS), a mixture of both positive and negative temporal correlation coefficients, (minimum -0.5 and maximum 0.2) can be seen between USGS groundwater level observation and W3RA groundwater storage changes (see Fig. 8.6 (A)). However, after merging W3RA with GRACE data the correlations are improved to be between +0.2 and +0.6 (Fig. 8.6 (B)). Analysis for Box 3 indicates that the simple groundwater output of the original W3RA has limitations in accounting for differences in physical parameters of wells in this region, and their different extraction and recharge rates are not well reflected in the model parameters. However, this dynamic complexity is introduced to the groundwater states by implementing the MCMC-DA of GRACE data.

Modification of trends and variability in groundwater storage after implementing MCMC-DA is not limited to the regions with negative trends. In Florida (Box 4), W3RA simulates negative trends (-4.9 ± 0.86 mm/yr) in the north and positive trends ($+2.4 \pm 0.53$ mm/yr) in the south, whereas both

MCMC-DA and independent USGS observations indicate trends with opposite signs (MCMC-DA trend in the north $+4.3 \pm 0.8$ mm/yr and -1.4 ± 0.4 mm/yr in the south). The positive trends in the north of the Florida state may be related to considerable inter-annual precipitation in southeastern CONUS (*Dourte et al.*, 2015) and that of the south attributed to agricultural expansion and population growth (*Renken et al.*, 2005; *Takatsuka et al.*, 2018). Moreover, temporal correlation coefficients between W3RA and USGS are significantly improved from -0.6 (on average) to 0.4 after merging with GRACE data.

8.5.2 Impact of the MCMC-DA on Seasonality of Groundwater Storage Changes

Seasonal amplitudes (i.e., only annual is shown here) in the MCMC-DA and the original W3RA groundwater storage are similar (differences $< 10\%$ in mm) in almost 75% of the grids across the CONUS. In contrast, large differences are found in Florida (Box 4) and within the Great Lakes (GL) area (Box 5), where the high annual-amplitude of W3RA is reduced by $\sim 80\%$ after implementing MCMC-DA (compare Figs. 8.4 (A2) and (B2)). Similar differences in seasonal amplitudes are apparent between GRACE TWSC and W3RA TWSC (see Fig. 8.1). Over-estimation of the seasonal amplitude in the groundwater storage component of W3RA near the GL (Box 5 in Fig. 8.4) may be caused by the model limitations in accounting for surface water changes in the region. Furthermore, in this region, uncertainties of the Glacial Isostatic Adjustment (GIA) model used to reduce the uplift load from GRACE observation are large (*Schumacher et al.*, 2018b). This reduction directly affects the estimation of long-term trends in GRACE TWSC, and consequently, it alters the individual water storage estimates after MCMC-DA. The annual and semi-annual amplitudes in

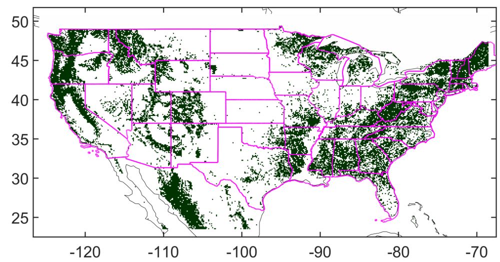

Fig. 8.7 The forest land distribution with 80% density across the CONUS, which is provided using the ESA CCI's Land Cover (LC) data (v2.0.7) http://maps.elie.ucl.ac.be/CCI. The original LC data is a consistent global LC map at 300 m spatial resolution on an annual basis from 1992 to 2015. For this study, the 2015's LC data has been extracted within the CONUS, with a spatial resolution of 12.5 km.

MCMC-DA groundwater storage in the southeastern CONUS (Box 6 in Fig. 8.4 including Georgia, Alabama, Mississippi, Tennessee, and Kentucky) are estimated to be higher than those from W3RA by more than 55% in mm. These discrepancies are attributed to the high density of vegetation in the southeastern CONUS (see Fig. 8.7), where the effect of cool temperatures on the vegetation, and the related vegetation-moisture dynamics cannot be well simulated by W3RA (*Van Dijk*, 2010).

Therefore, the seasonal patterns of groundwater storage in evergreen forests of southeastern CONUS are underestimated by the hydrological model.

8.5.3 Numerical Validation of Groundwater Estimates using USGS Data

A visual representation of the median, first and third quartiles of linear trends fitted to unit-less estimates (standardized) of groundwater storage derived from W3RA, MCMC-DA, and USGS data (see Fig. 8.5) are shown in Fig. 8.8 for 20 states across the CONUS. These states are selected within the regions where considerable modifications in estimated groundwater trends are found after implementing MCMC-DA and where USGS data are available from > 38000 wells between 2003 and 2017. The WMW statistical test indicates that the trends of MCMC-DA estimates in Fig. 8.8

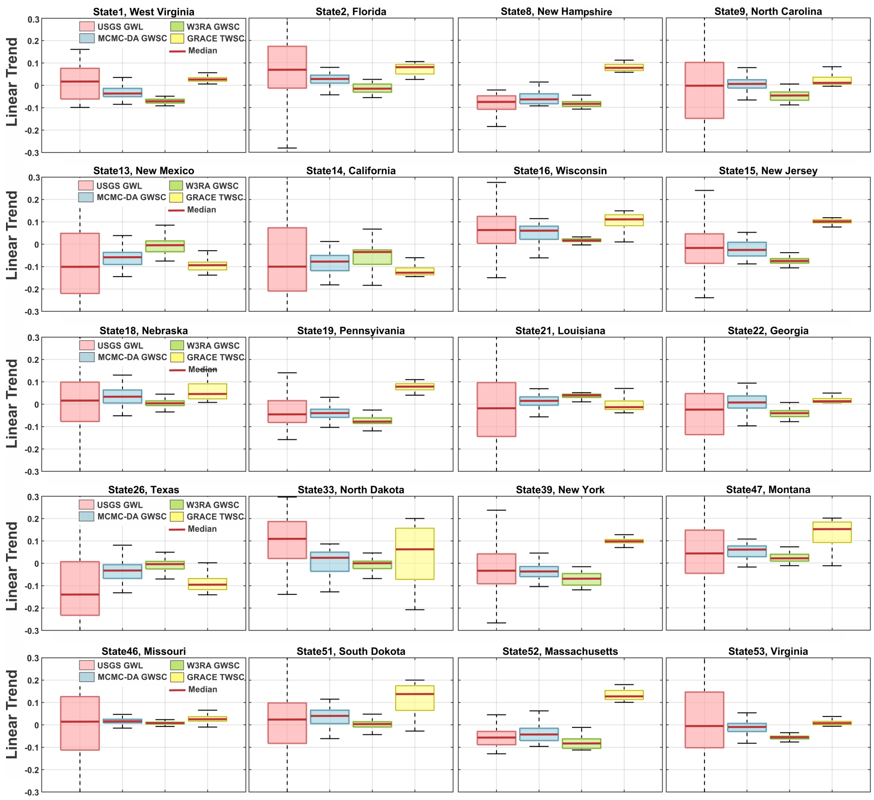

Fig. 8.8 A comparison between the linear trends fitted to the standardized (unit-less) groundwater storage time series from the original W3RA, MCMC-DA groundwater storage, USGS groundwater level observations, and standardize GRACE TWSC within different states of the CONUS, covering the period of 2003–2017. The tops and bottoms of each box are the 25th and 75th percentiles of each data sets (the first and third quartiles), and the red lines show the median of the time series. In this figure, the terms GWSC and GWL indicate groundwater storage changes and groundwater level, respectively.

are significantly closer to those of USGS compared to W3RA. Relatively significant improvements are detected in California with the median of USGS: 10% MCMC-DA: 8% and W3RA: 3% and relatively smaller changes in Missouri with the median of USGS: 1%, MCMC-DA: 1% and W3RA:

1%. In order to show further impacts of assimilating GRACE data within modelled groundwater estimates beyond those presented in Fig. 8.4, correlation coefficients are estimated between USGS groundwater level time series and those of the original W3RA and MCMC-DA, after removing trends and annual amplitudes, with the results shown in Fig. 8.9 (A) (USGS versus W3RA) and Fig. 8.9 (B) (USGS versus MCMC-DA). The greatest improvements are located within the south eastern CONUS, including Florida and Louisiana, and within central CONUS, including Texas, Utah, and Nebraska (i.e., more than 40% improvements on average in both central and south eastern CONUS).

8.6 Soil Water Storage Changes Across the CONUS

Integrating GRACE data into the model is expected to modify the vertical summation of water storage components. The portion of an update, which is assigned by MCMC-DA to each individual storage compartment, depends on the estimated $\bar{\Theta}$ in Eq. (5.13). The numerical results indicate that the estimated top ($0-10$ cm), shallow ($10-50$ cm), and deep soil layers ($50-200$ cm) account for $\sim$ 14%, 28%, and 58% of the variance in the total soil water storage, respectively. It has been found that, after implementing MCMC-DA, the magnitudes of soil water storage are changed by 72% on average. Long-term linear trends (22% of the total variance) and annual-amplitude (39% of the total variance) within the top-soil layer ($<$ 10 cm) are shown in Fig. 8.10, where those of the original model output are presented in Figs. 8.10 (A1) and (A2) and MCMC-DA results in Figs. 8.10 (B1) and (B2). Finally, the independent estimates of surface soil storage from the ESA CCI (see Section 2.5.3) are shown in Figs. 8.10 (C1) and (C2).

Linear trends in the top-soil layer of W3RA (Fig. 8.10 (A1)) are found to be small across the CONUS, i.e., less than ± 0.1 mm/yr. Comparing Fig. 8.10 (A1) with (B1) indicates that the linear trends are greatly modified after implementing MCMC-DA. Strong negative trends (~ -2 mm/yr) are found in the west (e.g., Box 1, i.e., California and Nevada) and in the south of the country (e.g., Box 2, i.e., Texas, Louisiana, Oklahoma, and New Mexico), while positive trends (up to 2 mm/yr) are found in the north of the country (e.g., Box 3, i.e., Montana, Wyoming, North and South Dakota). These

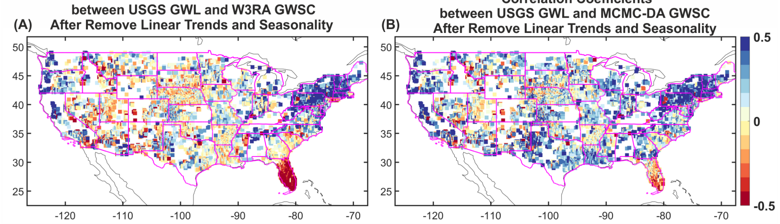

Fig. 8.9 Correlation coefficients (A) between the USGS groundwater level and W3RA groundwater storage; and (B) between the USGS groundwater level and MCMC-DA groundwater storage. Correlation coefficients are estimate after removing linear trends and seasonality from all data sets, covering 2003–2017.

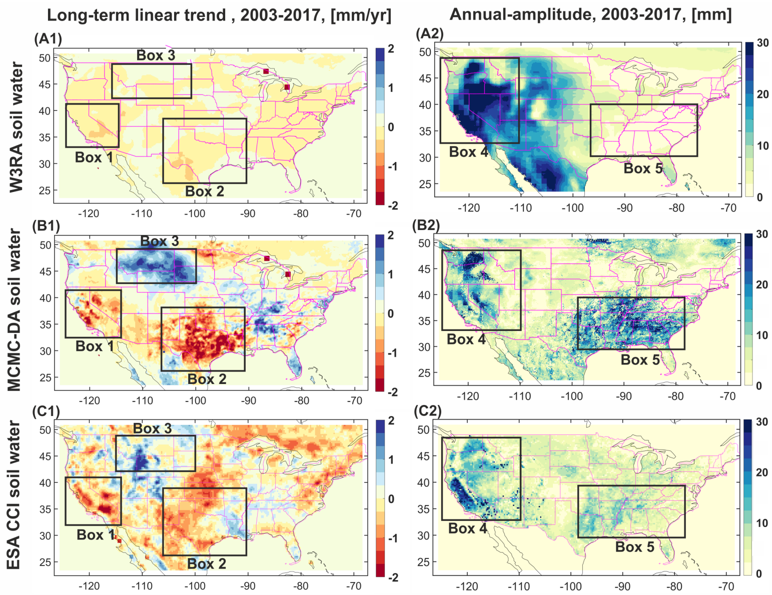

Fig. 8.10 Linear trend [mm/yr] and annual-amplitude [mm] fitted to (A1 and A2) soil water storage (top-layer, i.e., 10 cm) of the original W3RA; (B1 and B2) the updated soil water storage estimates derived from MCMC-DA; and (C1 and C2) soil water storage estimates from ESA CCI across the CONUS, covering 2003–2017. To enhance the visualisation, five boxes are shown in this figure (marked as Box 1, 2, 3, 4, and 5), where considerable changes are detected in the soil compartment after integrating GRACE TWSC into W3RA. The extension of these boxes are reported here as Box 1: $[112°W - 125°W, 32°N - 42°N]$, Box 2: $[90°W - 116°W, 26°N - 39°N]$, Box 3: $[100°W - 115°W, 42°N - 50°N]$, Box 4: $[110°W - 124°W, 32°N - 48°N]$, and Box 5: $[75°W - 90°W, 30°N - 40°N]$

estimates are consistent with those from the ESA CCI data, where the linear trends are estimated to be ~ -2, -1.7, and 1.6 mm/yr in Box 1, 2, and 3, respectively.

However, differences in estimates within some regions are found, such as the Sierra Nevada and central California (i.e., Box 1), as well as within Louisiana (east of Box 2) where ~ -2 mm/yr declining soil water storage of MCMC-DA is not detected by the ESA CCI product (i.e., -0.14 and 0.95 mm/yr within the Sierra Nevada and Louisiana, respectively). The main source of these differences is attributed to the high density of vegetation within the southeastern and northwestern CONUS (see Fig. 8.7).

Estimated negative linear trends in the MCMC-DA soil water storage changes within the Louisiana state (Fig. 8.10 (B1), Box 2) are likely associated with the rapid expansion of population and agricultural activities, which have caused the replacement of the natural forest lands with crops and modern vegetation landscapes (*Conroy et al.*, 2003; *Sun*, 2013). Removal of tree cover from forest and woodland soils may increase runoff and erosion rates, which in turn may decrease soil infiltration capacity (*Benito et al.*, 2003; *Doerr and Thomas*, 2000; *Ferreira et al.*, 2000). Furthermore, a large area of the western CONUS, in California and western Nevada, has experienced severe forest fires during the last decades, which influenced ecosystem properties, such as forest fragmentation, soil erosion rates, and sedimentation (*Agee*, 1996; *Beaty and Taylor*, 2007; *Miller et al.*, 2009). GRACE satellites can detect large-scale changes in water storage components, such as those related to massive changes in vegetation, which are successfully introduced into the MCMC-DA soil water storage estimates within these regions. ESA CCI product is, however, sensitive to the surface roughness and vegetation density parameter, which affects the quality of these data within the forest regions (*Dorigo et al.*, 2017).

W3RA soil water storage shows high seasonality in the west and northwest of the CONUS (Fig. 8.10 (A2), Box 4), where MCMC-DA decreases this value by $\sim 42\%$ on average (Fig. 8.10 (B2), Box 4). These values are found to be similar to those of ESA CCI (with differences $<\sim 5$ mm) within 76% of the region (see Fig. 8.10 (C2) Box4). In the southeastern CONUS, where considerable inter-annual precipitation is expected (*Dourte et al.*, 2015), the magnitude of the annual-amplitude in the W3RA soil water storage is estimated to be less than a few millimetres (see Fig. 8.10 (A2), Box 5). This is due to the fact that the effect of cool temperatures on vegetation and the related vegetation-moisture dynamics cannot be simulated well by W3RA (*Van Dijk*, 2010). Therefore, seasonal patterns in evergreen forests are underestimated. After implementing MCMC-DA, this value increases up to ~ 21 mm (Fig. 8.10 (B2), Box 5), which is closer to the ESA CCI value of ~ 16 mm (Fig. 8.10 (C2)).

8.6.1 Correlation Coefficients and RMSD of Soil Water With ESA CCI Soil Water Products

Temporal correlation coefficients and the Root Mean Squares of Differences (RMSD) between ESA CCI soil water storage and the top-layer soil water changes derived from W3RA and MCMC-DA are shown in Fig. 8.11, where the results indicate average improvements of 67% and 73%, respectively. Positive correlation coefficients (higher than 0.6) are found within ∼ 90% of the CONUS between the

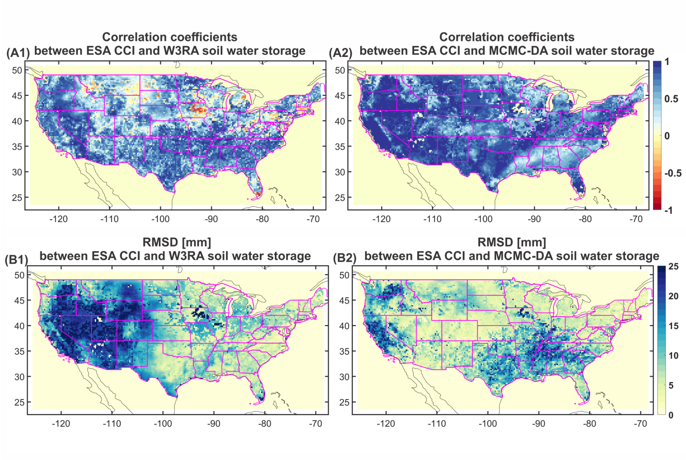

Fig. 8.11 Correlation coefficients and RMSD between the ESA CCI soil water storage and in (A1) and (B1), soil water storage estimates from the original W3RA; and in (A2) and (B2), the MCMC-DA's soil water storage. The computations use the whole period of 2003–2017.

ESA CCI and both W3RA and MCMC-DA outputs. WMW statistical test (at 95% confident level) indicates that except for the southeastern part of the CONUS (i.e., Alabama, Georgia, South Carolina, Mississippi, and Louisiana states) the values of the correlations are increased after implementing MCMC-DA. A comparison between the forest coverage across the CONUS shown by Fig. 8.7 and Figs. 8.11 (A1) and (A2) indicates that the minimum correlation coefficients (less than 0.4) are found within the forest regions, where both W3RA and ESA-CCI are relatively weak to represent the true signals. This is also reflected in the RMSD plots, where except for the southeastern CONUS the differences are decreased from ∼30 mm to ∼5 mm in the central and western parts of CONUS (see Figs. 8.11 (B1) and (B2)).

8.7 Impact of ENSO on Groundwater and Soil Water Storage Changes Across the CONUS

To demonstrate how GRACE TWSC might alter the inter- and intra-annual components of water storage changes, the dominant ENSO mode from MCMC-DA and W3RA water storage components are extracted and compared with climate indices (see, e.g., *Anyah et al.*, 2018) in terms of temporal correlation coefficients with the El Niña Southern Oscillation (ENSO). To this end, the Independent Component Analysis (ICA, *Forootan and Kusche*, 2012, 2013; *Forootan et al.*, 2018) is applied to isolate the ENSO modes from monthly W3RA and MCMC-DA groundwater and soil water storage outputs. The results are then compared in terms of amplitude and correlation coefficients with the ENSO index (El Niña 3.4 index [1]). ICA is applied to the groundwater and soil water storage

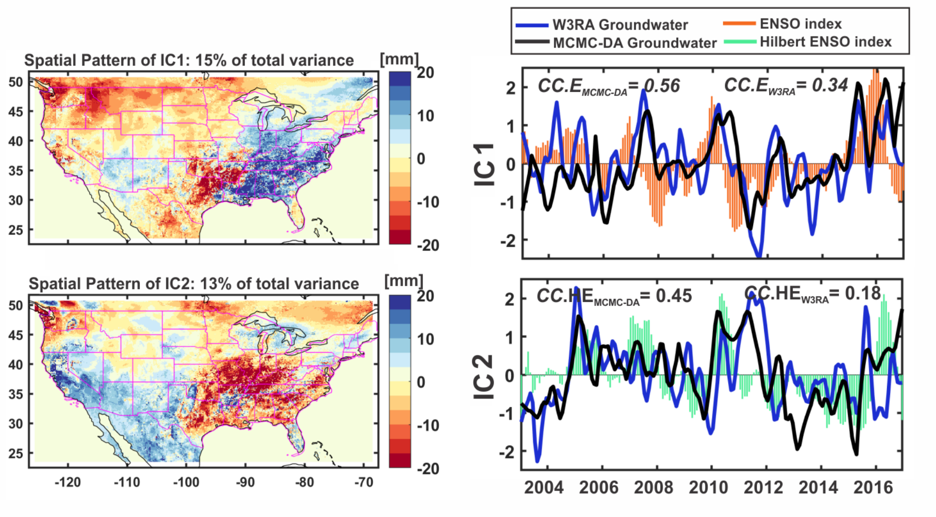

Fig. 8.12 Results of the ICA to extract the ENSO modes from groundwater storage changes after removing the linear trend and seasonal cycles (~ 59% of total variance) covering 2003–2017. The spatial anomaly map and its corresponding unit-less temporal evolution on Fig. 8.12 top represent the first independent mode. The second independent mode is shown in Fig. 8.12 bottom. $CC.E_{\text{MCMC-DA}}$ and $CC.HE_{\text{MCMC-DA}}$ mean Correlation Coefficient between ENSO index and temporal pattern of MCMC-DA groundwater storage, and between Hilbert ENSO index and temporal pattern of MCMC-DA groundwater storage respectively. The terms of $CC.E_{\text{W3RA}}$ and $CC.HE_{\text{W3RA}}$ mean Correlation Coefficient between ENSO index and temporal pattern of W3RA groundwater, and between Hilbert ENSO index and temporal pattern of W3RA groundwater, respectively.

estimates of W3RA and MCMC-DA after removing the long-term linear trend and seasonality (these components are shown in Figs. 8.4 and 8.10). ICA is applied to the water storage compartments from MCMC-DA. Then, W3RA estimates are projected onto the spatial components for comparison. In Figs. 8.12 and 8.13 the spatial anomaly modes are from the MCMC-DA estimates, and the associated temporal patterns (that are statistically independent and marked as IC1 and IC2) derived from MCMC-DA and W3RA. These are compared with the ENSO index and its Hilbert transformations. The two modes correspond to 15% and 13% of the residual groundwater, and 13% and 10% of the residual soil water storage variability, respectively. Form the obtained results, it has been found that the IC1 derived from both groundwater and soil water storage is in phase with the ENSO index, while IC2 follows its out-of-phase evaluation, i.e., the Hilbert transformation of the ENSO index. Therefore, these two modes capture the dominant influence of ENSO on groundwater and soil water storage changes across the CONUS, covering 2003–2017.

[1]https://www.esrl.noaa.gov/psd/data/correlation/nina34.data

The numerical results indicate that correlation coefficients between the ICs and the ENSO index are increased after integrating GRACE TWSC estimates into W3RA. For example, temporal correlation coefficient between IC1 of W3RA groundwater storage and the ENSO index (Fig. 8.12) is increased from 0.34 to 0.56 (65% improvement), and those of IC2 is increase from 0.18 to 0.45 (60% improvement) after merging models with GRACE data. This improvement can also be found in the soil water compartment, where the temporal correlation coefficients are increased from 0.21 to 0.64 for IC1 and from 0.11 to 0.49 for IC2 (Fig. 8.13).

In order to understand the effects of El Niña and La Niña events on soil water and groundwater storage, an independent comparison is performed by applying ICA on monthly precipitation data from ERA-Interim[2]. As expected, correlation coefficients between precipitation IC1 and the ENSO index (0.67), and between precipitation IC2 and the Hilbert ENSO index (0.43) are relatively high (Fig. 8.14).

Comparing the ENSO modes of groundwater and soil water storage, as well as the precipitation anomaly (Fig. 8.12, Fig. 8.13, and Fig. 8.14), with the ENSO index shows that La Niña events (where the ENSO index is negative) resulted in water deficits in the southeast of the CONUS including parts of Texas, Louisiana, Arkansas, Mississippi, Alabama, and Georgia (see also, *Cook et al.*,

[2]https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/

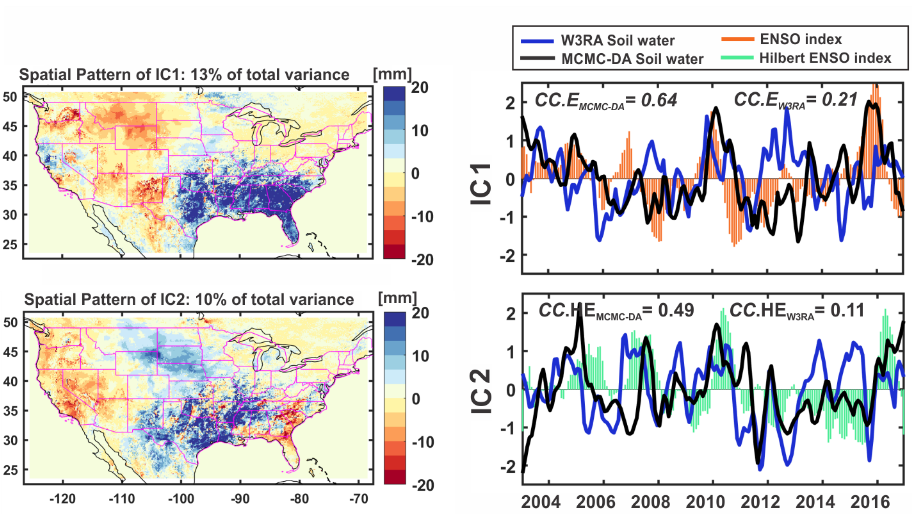

Fig. 8.13 Results of the ICA to extract the ENSO modes from MCMC-DA soil water storage changes after removing the linear trend and seasonal cycle (~ 43% of total variance) between 2003–2017. This figure shows the first two independent modes with the spatial and temporal pattern of IC1 on top and those of IC2 on the bottom. The spatial patterns correspond to the MCMC-DA. The El Niña 3.4 index and its Hilbert transformed time series are shown alongside the IC1 and IC2, respectively. $CC.E_{\text{MCMC-DA}}$ and $CC.HE_{\text{MCMC-DA}}$ mean Correlation Coefficient between ENSO index and temporal pattern of MCMC-DA soil water storage, and between Hilbert ENSO index and temporal pattern of MCMC-DA soil water storage, respectively. The terms of $CC.E_{\text{W3RA}}$ and $CC.HE_{\text{W3RA}}$ mean Correlation Coefficient between ENSO index and temporal pattern of w3RA soil water storage, and between Hilbert ENSO index and temporal pattern of W3RA soil water storage, respectively.

2007; *Manuel*, 2008; *Rippey*, 2015; *Seager et al.*, 2014). In contrast, wetter than normal conditions are detected in the northwest, e.g., the 2011's Missouri River floods (*Reager et al.*, 2014; *Zheng et al.*, 2014). Both W3RA and MCMC-DA capture the influence of La Niña events on groundwater storage changes across the CONUS, especially between 2010–2014. (Fig. 8.12). However, the ENSO mode of W3RA soil water storage shows an opposite evolution compared to that of the ENSO index (with the correlation coefficient of -0.85) during these four years. This is, however, modified after implementing the MCMC-DA approach, where a correlation coefficient of 0.72 is subsequently found between IC1 and the ENSO index (Fig. 8.13).

8.8 Down-Scaling GRACE TWSC Observation Using MCMC-DA

MCMC-DA is able to down-scale GRACE TWSC observations vertically (separating to individual storage estimates) and horizontally (i.e., improving the $\sim$ 300 km resolution to $\sim$ 12.5 km). To illustrate this gain, latitudinal (the longitude is fixed at $-100°$) and longitudinal (latitude is fixed at $40°$) profiles are shown in Fig. 8.15 left and right panels, respectively. Here, only changes in the linear trends fitted to the GRACE TWSC, W3RA and MCMC-DA groundwater and soil water estimates are shown.

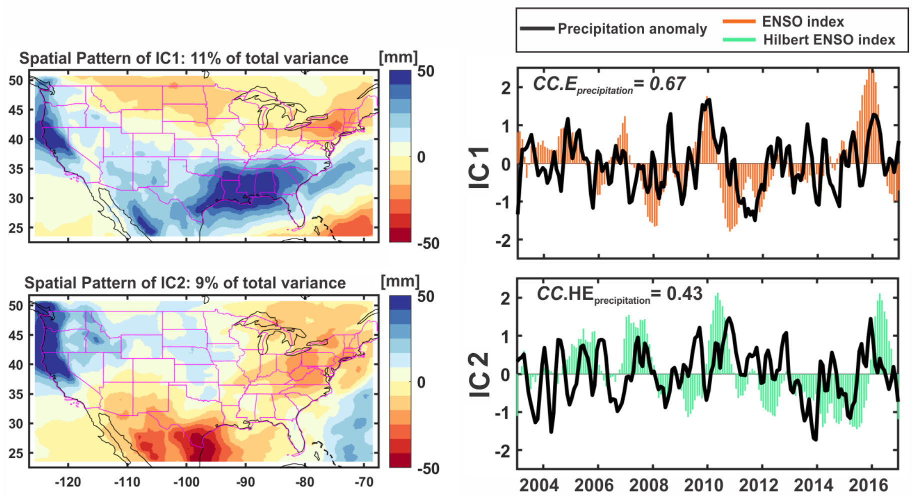

Fig. 8.14 Results of the ICA to extract the ENSO modes from precipitation anomalies within the CONUS. The linear trend and seasonal cycles ($\sim$ 53% of total variance) between 2003–2017 are removed before applying the ICA. This figure shows the first two independent modes of the residual signal derived from precipitation anomaly (i.e., corresponded to the ENSO mode), and their relation with El Niña 3.4 index (ENSO index) and its Hilbert transform, where the spatial and temporal patterns of IC1 are shown on top and those of IC2 on the bottom. The El Niña 3.4 index and its Hilbert transformed time series are shown alongside the IC1 and IC2, respectively, where $CC.E_{\text{precipitation}}$ and $CC.HE_{\text{precipitation}}$ represent correlation coefficient between ENSO index and temporal pattern of precipitation anomaly, and between Hilbert ENSO index and temporal pattern of precipitation anomaly respectively.

The GRACE signal evolves quite smoothly (see the solid black lines), but those of W3RA and MCMC-DA represent strong spatially-dependent variability. Integrating GRACE TWSC into W3RA using MCMC-DA, however, modifies the evolution of both soil and groundwater storage estimates. In the left panel (i.e., longitude profile), the larger portion of the storage update is introduced to the groundwater compartment (compare the light- and dark-blue curves), where the profile is modified towards GRACE. The soil water storage profile is only modified marginally (i.e., standard deviations of the update is 36% of groundwater). Similar magnitudes of updates to the soil water and groundwater storage are found with the standard deviations of 1.14 mm and 2.46 mm, respectively (see Fig. 8.15 right panel).

The results, therefore, indicate that the MCMC-DA considers the magnitude of storage in various compartments to modify them, and the updated estimates can be used to explore high-resolution hydrological changes across the CONUS as demonstrated in previous sections. In what follows,

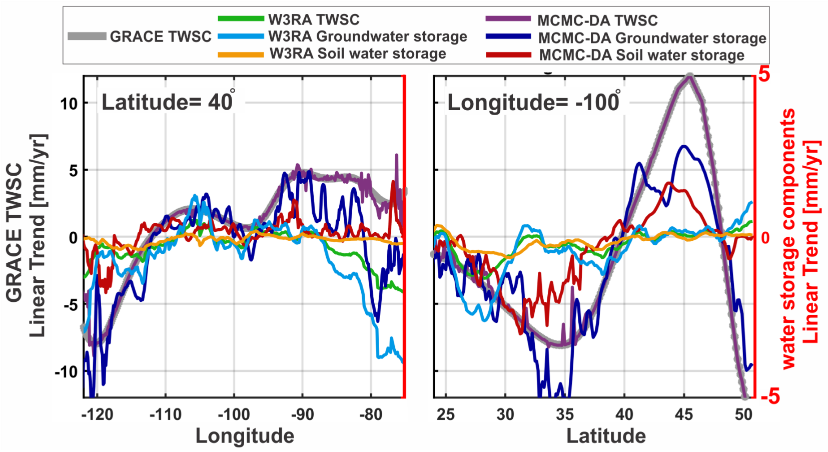

Fig. 8.15 The latitudinal (left panel, the longitude is fixed at $-100°$) and the longitudinal (right panel, the latitude is fixed at $40°$) profiles derived from the linear trends of W3RA and MCMC-DA groundwater, soil water, and TWSC, as well as TWSC derived from GRACE data. The left y-axis corresponds to the linear trend of GRACE TWSC, and the right y-axis corresponds to the linear trend of the individual water storage states from the original W3RA and MCMC-DA.

changes in the water storage components within Texas and California are explored, where considerable differences are found between the MCMC-DA and W3RA estimates (see Section 8.5 and Section 8.6). Comparisons are provided in terms of the spatial distribution of trends and spatially averaged groundwater and soil water storage changes within these states. An approximate of Storage coefficients (Sc) are available for these states from previous studies, therefore, an evaluation in terms of groundwater storage is presented. Complementary comparisons are provided with the soil storage using ESA CCI data, as well as precipitation anomalies, and the Palmer Hydrological Drought Index [3] (PHDI, *Palmer*, 1965).

[3] https://www.ncdc.noaa.gov/temp-and-precip/drought/historical-palmers/

8.9 Changes in Water Storage Components Within Texas

Texas (located in the southcentral CONUS) is the second-largest state in the country by area (after Alaska) and population (after California). According to the Office of the State Demographer and the Texas State Data Center's 2014 predictions, its population is projected to increase until 2050 (*Potter and Hoque*, 2014).

Texas experienced several drought events during the last 2 decades, and those of 2010–2014 being the strongest (Fig. 8.16 (A), *Long et al.*, 2013). The ENSO modes of the MCMC-DA groundwater (Fig. 8.12) and soil water storage (Fig. 8.13), as well as those of precipitation anomaly (8.14), indicate that La Niña events are the main climatological cause of these droughts (see also *Rippey*, 2015; *Seager et al.*, 2014). This is well reflected in the magnitude of precipitation anomalies (Fig. 8.16 (B)), where the annual precipitation is substantially reduced by 56% during the drought years of 2010–2014. This is also well reflected in the averaged GRACE TWSC and MCMC-DA TWSC in Fig. 8.16 (C), where the evolution closely follows the PHDI (with the correlation coefficient of 0.75), showing that this integration improves the representation of drought events. USGS groundwater storage within

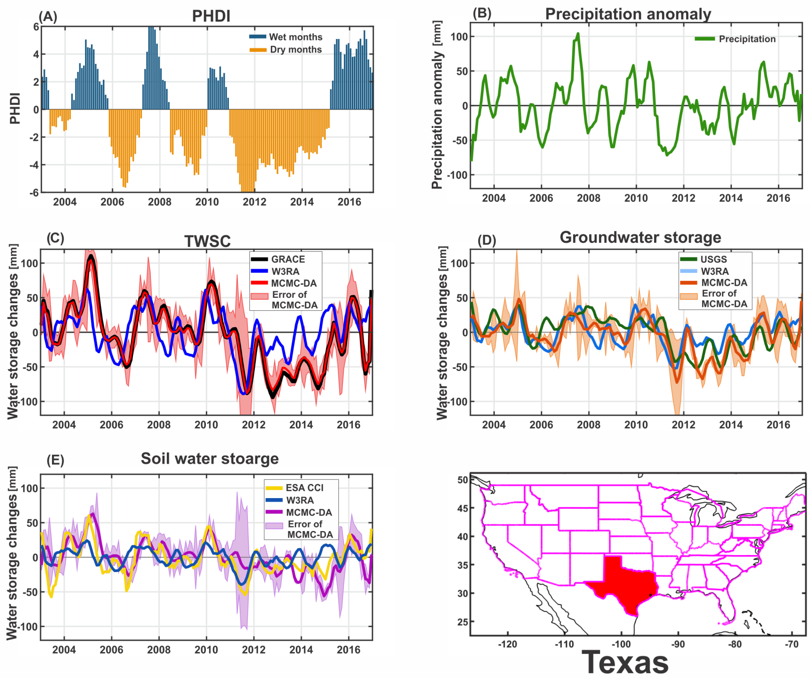

Fig. 8.16 An overview of hydro-climatological changes within Texas, where the spatially averaged time series correspond to (A) the Palmer Hydrological Drought Index (PHDI), (B) precipitation anomalies, (C) TWSC derived from GRACE, W3RA, and MCMC-DA, (D) groundwater storage derived from USGS observations, W3RA, and MCMC-DA, as well as (E) soil water changes derived from ESA CCI products, W3RA, and MCMC-DA.

Texas is obtained by averaging the observed levels and using an average storage coefficient (i.e., 0.18, see Section 2.5.2) for the conversion. The results (the dark green curve in Fig. 8.16 (D)) indicate that during the extreme drought of 2010–2014, the mean of groundwater storage in Texas markedly decreased (-36 mm) compared to the mean value of 2003–2010. Comparing groundwater storage estimates of the original W3RA and MCMC-DA with those of USGS indicates a higher agreement after integrating GRACE data (i.e., Root Mean Squares of Differences, RMSD, reduced from 60 mm to 18 mm).

Earlier in Fig. 8.10 (A), it appears that the original W3RA simulates soil water storage with a negligible linear trend within the entire CONUS. Integration of GRACE TWSC into W3RA model outputs introduces a negative linear trend to the soil compartment and modifies its cyclic components (such as seasonality). This is reflected in the averaged soil storage time series (Fig. 8.16 (E)), where the evolution of MCMC-DA is found to be closer to ESA CCI with the correlation coefficient of 0.56 compared to that of the original W3RA (i.e., 0.32). The standard deviation of storage is also modified from 16.56 mm to 26.63 mm after integrating GRACE TWSC.

8.10 Changes in Water Storage Components Within California

California is the third-largest (by area) and most densely populated state, which ranks first in the country in terms of economic activities and agricultural value. California has various types of climate from hyper-arid to polar, depending on latitude, elevation, and proximity to the coast. PHDI in Fig. 8.17 (A), precipitation anomalies in Fig. 8.17 (B), and TWSC observed by GRACE mission in Fig. 8.17 (C) indicate that California has experienced several drought events during the last decades, such as the three-year drought of 2007–2009 (*Jones*, 2010) and the five-year drought between 2012–2017 (e.g., *Diffenbaugh et al.*, 2015; *Griffin and Anchukaitis*, 2014; *Seager et al.*, 2015). The averaged GRACE TWSC estimates indicate that during 2012–2017, the state lost TWSC at the rate of -22.50 mm/yr. Large differences are found between the original W3RA and GRACE TWSC during these drought years (RMSD of 54.51 mm), indicating that in addition to the precipitation deficit (-10.45 mm/yr) anthropogenic modifications contributed strongly to the water storage decline (-12.05 mm/yr).

In Fig. 8.17 (D), the averaged groundwater storage changes within California are shown. The USGS water levels are converted to storage estimates using an average Sc of 0.15 from *Scanlon et al.* (2012b). The averaged soil water storage results are shown in Fig. 8.17 (E) with a comparison with the ESA CCI products. Comparing the curves in Fig. 8.17 (D) with Fig. 8.17 (E) indicates that 77% of the update from GRACE TWSC is introduced to the groundwater compartment. A decreasing trend of -4.95 mm/yr is derived for groundwater storage during 2003–2017 and -18.9 mm/y for the drought period of 2012–2017. These estimates are found to be close to that from USGS data, i.e., -3.46 mm/yr during 2003–2017 and -17.79 mm/yr during 2012–2017. GRACE data have mostly modified

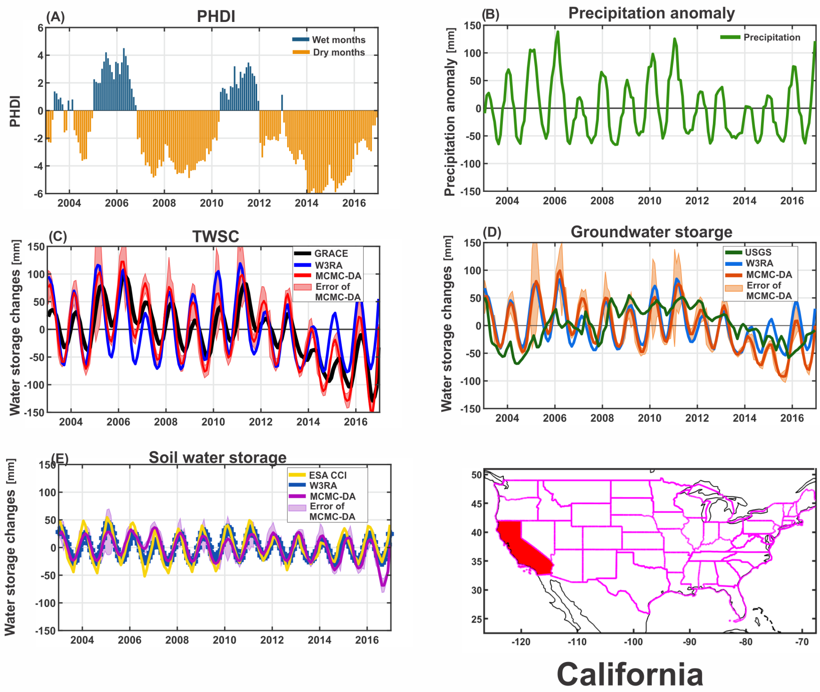

Fig. 8.17 An overview of hydro-climatological changes within California, where the spatially averaged time series correspond to (A) the Palmer Hydrological Drought Index (PHDI), (B) precipitation anomalies, (C) TWSC derived from GRACE, W3RA, and MCMC-DA, (D) groundwater storage derived from USGS observations, W3RA, and MCMC-DA, as well as (E) soil water changes derived from ESA CCI products, W3RA, and MCMC-DA.

the linear trend in the soil water storage changes (see Fig. 8.17 (E)), i.e., changing from -0.45 mm/yr to -1.66 mm/yr, which is closer to that of ESA CCI (-1.12 mm/yr).

8.11 Summary and Conclusions

In this chapter, MCMC-DA (introduced in Chapter 5) was implemented to explore high resolution ($\sim$12.5 km) groundwater and soil water storage changes across the CONUS, covering 2003–2017. This approach was tested by performing various comparisons between the original W3RA estimates and the MCMC-DA results, as well as validations against the in-situ US Geological Survey (USGS) groundwater level observations and the European Space Agency (ESA)'s Climate Change Initiative (CCI) water storage between 2003–2017.

The numerical results indicated that the MCMC-DA introduced trends, which exist in GRACE TWSC, mostly to the groundwater storage and to a less extent to the soil water storage compartments. For example, the linear trend fitted to the groundwater storage changes of the original W3RA model was changed from -0.11 ± 0.34 to -2.06 ± 0.56 mm/yr within Texas, and from -2.63 ± 0.32 to -4.95 ± 0.44 mm/yr within California. A higher similarity was found in groundwater estimation of MCMC-DA and those of USGS in the southeastern CONUS, e.g., in Florida, North and South Carolina, and Virginia states.

The linear trend in the model's original soil water storage is modified from ± 0.5 mm/yr to ± 2 mm/yr, averaged over the entire CONUS, which is closer to the independent estimations from the ESA CCI. MCMC-DA also improved the estimation of soil water storage changes in regions with high forest intensity (e.g., southeastern CONUS).

To demonstrate the impact of integrating GRACE TWSC into W3RA on the inter- and intra-annual components of water storage changes, we also investigate the storage changes associated to the El Niña Southern Oscillation (ENSO). For this, the Independents Component Analysis (ICA) is applied to isolate the ENSO modes from groundwater and soil water storage estimates. The modes derived from MCMC-DA are found to be better correlated to the Niña 3.4. ENSO index. For example, the correlation coefficients between the ENSO mode of groundwater and soil water storage and ENSO index are increased from 0.34 to 0.56 and from 0.21 to 0.64, respectively. Comparisons of ENSO modes of water storage show that MCMC-DA improve agreement between soil water and precipitation (correlation coefficient of 0.58), relative to W3RA (correlation coefficient of 0.41). This indicates that the coupling procedure between the shallow soil water storage and precipitation is stronger than that of the deep storage changes associated with the groundwater compartment.

Comparing W3RA and MCMC-DA TWSC (e.g., the red and blue curves in Fig. 8.16 (C)) indicates that integrating GRACE data into W3RA modifies the timing of water storage changes (i.e., on average it advances the phase of the W3RA TWSC time series by 2 months). Therefore, we conclude that the proposed integration of GRACE data into W3RA improved representation of slowly evolving hydrological processes, such as hydrological droughts (see e.g., *Forootan et al.*, 2019).

The results indicate that estimation of groundwater and soil water storage over the Great Lakes (GL) area (northeast CONUS) is complex. This may be due to limitations in accounting for interactions between surface water and groundwater in regions where the predominant influence is surface water (e.g., the GL area). Furthermore, over the northern part of CONUS, the GIA model uncertainties, as required to reduce the effect of PGR from GRACE TWSC estimates, are significant (*Schumacher et al.*, 2018b). This uncertainty directly affects the estimation of long-term trends in TWSC, and consequently, alters the individual water storage estimates after MCMC-DA. In the next chapter, the extended MCMC-DA, the ConBay-DA (formulated in Chapter 6), is implemented to simultaneously estimate water storage change components and surface deformation from GRACE data within the GL area, while the influence of estimating PGR uplift rates on TWSC estimation is discussed.

Chapter 9

Application of ConBay-DA for a Joint Estimation of Land Hydrology and Surface Deformation

9.1 Introduction

ConBay-DA (formulated in Chapter 6) is implemented here to merge GRACE field estimates and in-situ GNSS measurements for a joint estimation of the land hydrology and PGR uplift rates (surface deformation). The methodology can be applied globally but an illustration of its implementation is presented within the Great Lakes (GL) area, the Unites States (US), during 2003 to 2017.

GL is located in the northeast of the CONUS, where mass changes due to PGR and surface deformations are significant contributors within the GRACE(-FO) data. As *Schumacher et al.* (2018b) indicated, large uncertainties exist between the GIA models over the northern part of the US, making this region a good candidate to study the performance of ConBay-DA. *Winter et al.* (1998) and *Sophocleous* (2002) found strong interactions between surface water and groundwater storage changes within GL. Therefore, separating GRACE(-FO) signals into its compartments can be of interest to hydrological applications.

In what follows, an overview of the observations and models used here is provided in Section 9.2. The estimated PGR rates and TWSC are shown and evaluated in Sections 9.3 and 9.4, respectively, and a summary of the results can be found in Section 9.5.

9.2 An Overview of Data and Models Used in the Application of ConBay-DA

For this study, instead of removing PGR uplift rates from GRACE data during the post processing steps (as it was discussed in Section 2.2.4), the PGR rate is assumed to be unknown that can be estimated from GRACE(-FO) data through the ConBay-DA of Chapter 6. To distinguish the terminology of this chapter from previous ones, the equivalent water heights estimated from GRACE(-FO) data are named as the GRACE Total Water Storage Changes (Total-WSC), which means that they contain signals of **both** hydrology and PGR. In another words, Total-WSC estimates are obtained as a summation of terrestrial water storage changes (TWSC, as it was assumed to be solely hydrology driven in previous chapters) and the Equivalent Water Heights (EWHs) that correspond to the PGR deformation. The computations of these fields within GL ($[75°W - 92.5°W, 40°N - 50°N]$) and its full error covariance matrix follow Section 2.2.8. Grids are considered to have 1° spatial resolution, and the test is performed for the period of 2003–2017 considering the GRACE mission data only.

W3RA water balance model (*Van Dijk*, 2010) (Section 2.4) and the ICE5G-VM2 GIA model (*Wahr and Zhong*, 2012) are used as *a priori* information, while the in-situ GNSS measurements (*Schumacher et al.*, 2018b, see also Section 2.3) are used in a hierarchical level to constrain the PGR estimates from GRACE data.

Considering the resolution of GRACE, the original W3RA model outputs ($0.125° \times 0.125°$) and the ICE5G-VM2 GIA model output ($0.5° \times 0.5°$) are averaged on $1° \times 1°$ grids. Figure 9.1 shows the

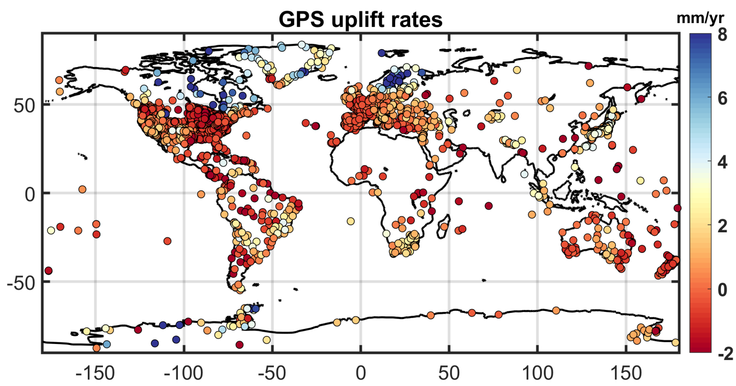

Fig. 9.1 The global PGR-related crustal uplift rates derived from in-situ GNSS measurements (*Schumacher et al.* (2018b)), between 2003–2017.

spatial distribution, as well as the PGR uplift rates of global in-situ GNSS measurement used in this study. To make the best of the in-situ GNSS measurements and match the spatial resolution of GNSS data with GRACE Total-WSC, an interpolation approach, based on the Least Squares Collocation (LSC) technique (*Moritz*, 1978), is used to obtain grid with 1-degree spatial resolution within the CONUS. The interpolated values are shown in Fig. 9.2 over the GL area, were we aim to estimate the

PGR rates and TWSC from GRACE data. The gridded uplift rates of the in-situ GNSS measurements and those of the GIA model output are converted to EWH, following the procedure explained in Section 2.2.4. Figure 9.3 shows the rate of the EWHs derived from in-situ GNSS measurements (Fig. 9.3 (A)) and those of ICE5G-VM2 GIA model (Fig. 9.3 (B)), while the differences between them are shown in Fig. 9.3 (C).

Considerable differences ($\sim \pm 14$ mm/yr) between in-situ GNSS measurements and the GIA model output can be seen in the northeast and northwest of the GL area, where the simulated PGR rates are found to be under-estimated (compared with the observations) in the northeast, and over-estimated in the northwest of GL (Fig. 9.3 (C)). In the northeast, the rate of EWHs derived from GNSS measurements are estimate around 30 mm/yr and in the northwest are estimated up to 4 mm/yr. ICE5G-VM2 GIA model, however, simulated these values to be $\sim$ 15 mm/yr.

GRACE Total-WSC within the GL area is also compared with W3RA TWSC + EWHs derived from ICE5G-VM2 GIA model in terms of long-term linear trend (Fig. 9.4 (A1) and (B1)) and annual amplitude (Fig. 9.4 (A2) and (B2)) between 2003–2017. Considerable linear trends and annual-amplitude differences are found between measured and the modelled Total-WSC. For instance, linear trends fitted to the GRACE Total-WSC in the northeast of the GL area is estimated up to $\sim$ 34 mm/yr,

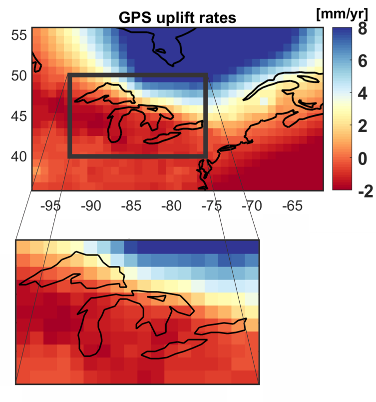

Fig. 9.2 Gridded PGR uplift rates derived from interpolated in-situ GNSS measurements (*Schumacher et al.* (2018b)) over the Great Lakes (GL) area, between 2003–2017.

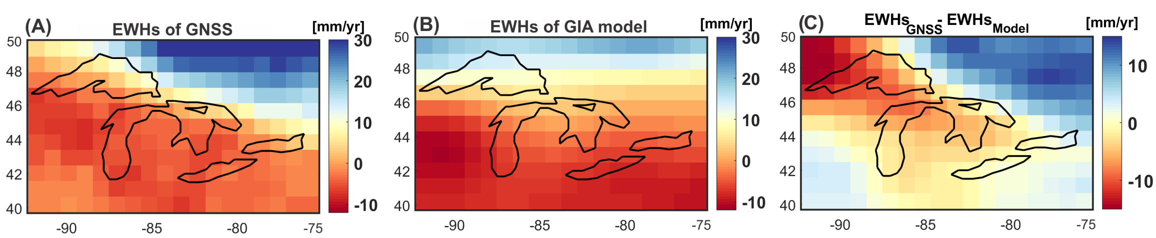

Fig. 9.3 The rate of EWHs [mm/yr] derived from (A) in-situ GNSS measurements, and (B) ICE5G-VM2 GIA model output. Differences between (A) and (B) are shown in Fig. 9.3 (C).

which is considerably larger than those derived from models (W3RA+ICE5G-VM2) with the mean value of $\sim$ 20 mm/yr. Positive trends up to 8 mm/yr are estimated for GRACE data in the northwest, while those of derived from model are estimated up to 18 mm/yr.

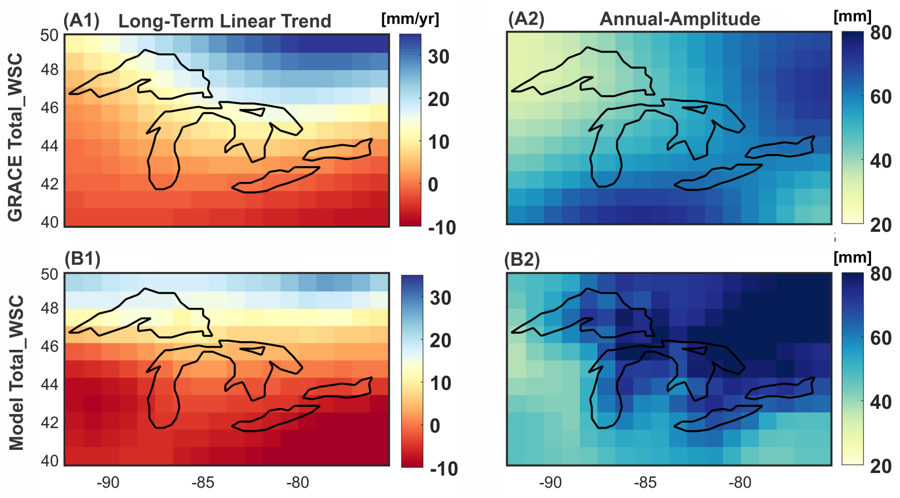

Fig. 9.4 Long-term linear trends [mm/yr] and annual-amplitudes [mm] fitted to the GRACE- and model- derived Total-WSC within the GL area, covering the period of 2003–2017. The plots in (A1) and (A2) correspond to the GRACE data, while (B1) and (B2) correspond to the summation of W3RA TWSC and ICE5G-VM2 EWHs.

Large differences are also found in terms of annual-amplitude between GRACE Total-WSC and W3RA TWSC, mostly in the northeast and southwest of GL. In these regions, the annual-amplitude of GRACE are estimated up to 70 mm, while W3RA water balance model simulates this value up to 90 mm in the northeast, and less than 45 mm in the southwest of GL. Differences in the seasonality of water storage changes between modelled (the seasonality of W3RA TWSC) and measured (GRACE), can be related to the errors in the forcing data and uncertainty in the model parameters to control these values (*Van Dijk et al.*, 2011).

9.3 PGR Uplift Rates Derived from ConBay-DA

After implementing ConBay-DA to merge GRACE observation and in-situ GNSS measurements with *a priori* information, the estimated value of PGR uplift rates, as the updated values of ICE5G-VM2 GIA model, are compared with the original model output in this section.

MCMC-DA (formulated in Chapter 5), as an unconstrained Bayesian-DA, is also implemented to separate land hydrology and PGR uplift rate from GRACE Total-WSC, and the results are compared with those of ConBay-DA. The main goal of this implementation is to assess the impact of the constraint (introduced in Eq. 6.3) on the signal separation results, and to see how in-situ GNSS measurements might affect the estimation of the PGR uplift rates and hydrological signals from GRACE data. To separate GRACE Total-WSC to its compartments, using MCMC-DA, the observation

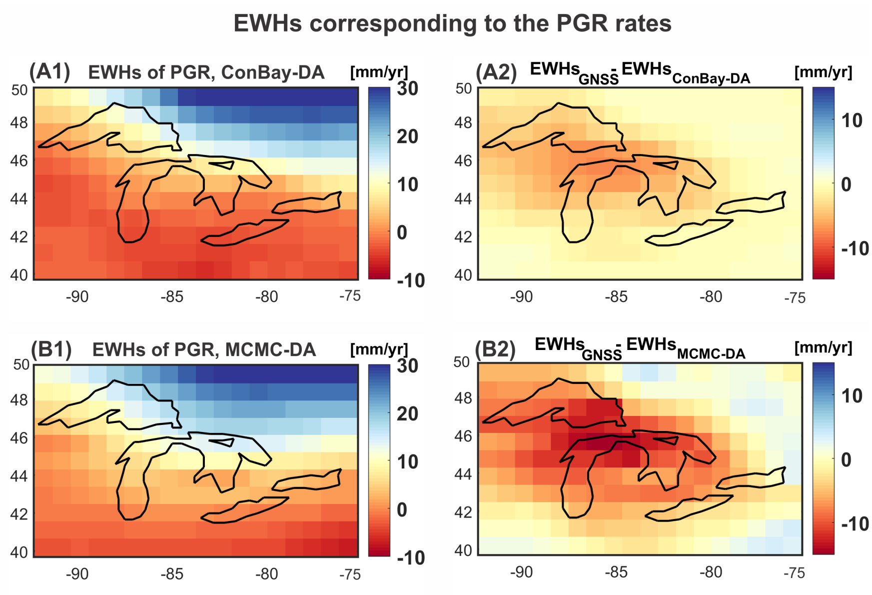

Fig. 9.5 The rate of EWHs corresponding to the PGR uplift rates [mm/yr], derived from (A1) ConBay-DA, and (B1) MCMC-DA approach. Difference between the rate of EWHs derived from in-situ GNSS and those of derived from ConBay-DA and MCMC-DA are shown in (A2) and (B2), respectively.

equation (Eq. (4.1)) and the state equation (Eq. (4.2)) of the state-space models, as the basis of formulating the MCMC-DA approach, are needed to be extended to include hydrological and GIA model outputs as *a priori* information. The extended formulation of Eq. (4.1) and Eq. (4.2) were shown by Eq. (6.1) and Eq. (6.2).

The rate of EWHs that corresponded to PGR are estimated by ConBay-DA and MCMC-DA, which are shown in Fig. 9.5 (A1) and (B1), while their differences with the EWHs of in-situ GNSS measurements (Fig. 9.3 (A1)) are shown in Fig. 9.5 (A2) and (B2), respectively. The obtained results indicate that the bias between in-situ GNSS observation and the GIA model output is considerably reduced, after implementing both MCMC-DA and ConBay-DA, by 56% and 90%, respectively. However, in the middle of GL, PGR of both techniques are over-estimated, compared to the in-situ GNSS measurements. This could be due to the inconsistencies between GRACE and GNSS signals or the fact that the hydrological model did not provide reasonable estimates of interactions between surface water changes and the other components of the water cycle.

Comparing MCMC-DA and ConBay-DA in Fig. 9.5 (A2) and (B2) shows that using in-situ GNSS measurements to constrain the estimation of PGR can reduce the effect of these uncertainties in the middle of GL, where differences between MCMC-DA and in-situ GNSS measurements are found to be 13 mm/yr, in terms of EWH magnitude, while those between ConBay-DA and in-situ measurements are estimated to be up to 4 mm. The main hypothesis to formulate the constraint equation in ConBay-

DA is that PGR manifests as a trend in the relatively short era of the GRACE(-FO) mission, which can be defined within the Bayesian signal separation framework (Eq. 6.3). To test this hypothesis, the annual-amplitudes fitted to the EWHs of PGR derived from ConBay-DA and MCMC-DA are shown in Fig. 9.6 (A) and (B), respectively. The results indicate that unwanted seasonal components are introduced to the PGR estimates of MCMC-DA. However, in ConBay-DA, the constraint equation and GNSS data removes the seasonality of the estimated PGR, i.e., the amplitude of the seasonality is of the level of GRACE noise ($\sim$ 20 mm). The reason could be due to the fact that W3RA excludes surface water storage, while GRACE detects this change within the GL area. The lack of *a priori* information about surface water storage may reflected as an uncertainty on the GRACE signal separation, which can be seen in Fig 9.6 (A).

9.4 Water Storage Changes Derived from ConBay-DA

After evaluating the performance of the ConBay-DA approach to estimate PGR uplift rates, it is important to see how merging GRACE and in-situ GNSS measurements with *a priori* information can improve the estimation of land hydrology signals compared to the original model outputs and MCMC-DA (latter pre-defines the PGR values using a model instead of co-estimating these rates with water storage compartments). Figure 9.7 (A1) and (A2) shows the long-term linear trend and annual-amplitude fitted to the Total-WSC derived from ConBay-DA covering 2003–2017, which indicates that implementing ConBay-DA, as expected, reduces biases between the modelled and GRACE Total-WSC estimates. For example, the mean value of Root Mean Square of Differences (RMSD) between measured and modelled Total-WSC is reduced from $\sim$ 68 mm to $\sim$ 15 mm within the GL area (see Fig. 9.8). Since, all the hydrological and land surface deformation signals are estimated simultaneously within the proposed Bayesian approaches (MCMC-DA and ConBay-DA), it is expected that any changes in PGR uplift rate estimation affect the estimation of hydrological signals from GRACE data. To assess this hypothesis, the long-term linear trends and the annual-amplitudes of the top-layer ($<$ 10 cm) soil water storage changes derived from MCMC-DA (unconstrained Bayesian-

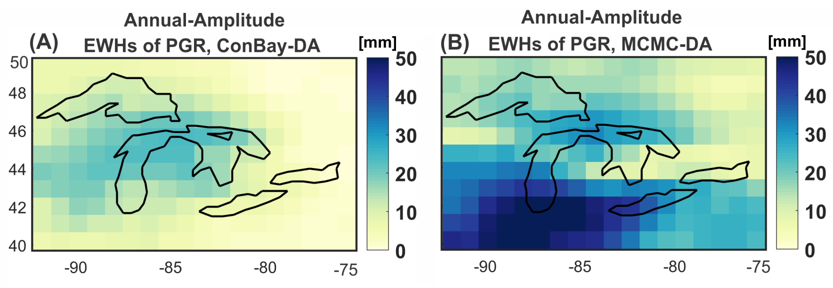

Fig. 9.6 Annual amplitudes fitted to the estimated PGR signal derived from (A) ConBay-DA, and (B) MCMC-DA, i.e., without using GNSS measurements to constrain PGR estimates.

DA) and ConBay-DA are compared in Fig. 9.9 (A1, A2) and (B1, B2), respectively. Comparing MCMC-DA and ConBay-DA results in Fig. 9.9 indicates that using in-situ GNSS measurements influences the estimation of hydrological signals. Validation against ESA CCI soil moisture products (Fig. 9.9 C1, C2), in terms of RMSD and temporal correlation coefficients (see Fig. 9.10), indicates that ConBay-DA can better estimate soil water storage from GRACE data, compared to the original model and MCMC-DA approach. For example, in south of GL, the correlation coefficients between MCMC-DA and ESA CCI soil water changes is estimated to be less than −0.4, while those of between ConBay-DA and ESA CCI product show positive values up to 0.5. A considerable improvement

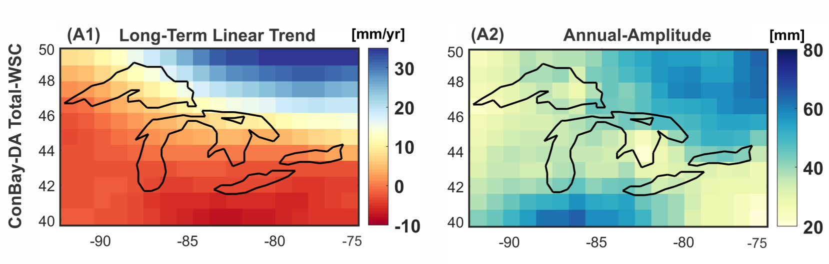

Fig. 9.7 Long-term linear trends [mm/yr] and annual-amplitudes [mm], fitted to the Total-WSC derived from ConBay-DA within the GL area, covering the period of 2003–2017.

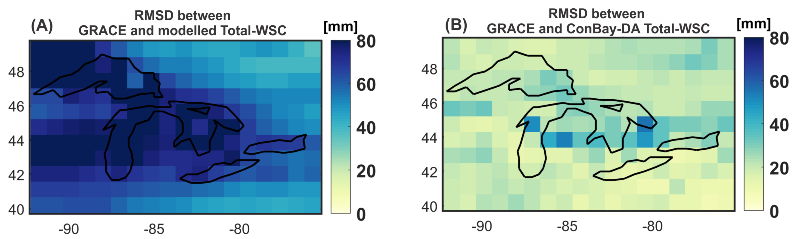

Fig. 9.8 Root Mean Square of Differences (RMSD) between (A) measured and modelled Total-WSC, and (B) measured and ConBay-DA Totoal-WSC, covering the period of 2003–2017.

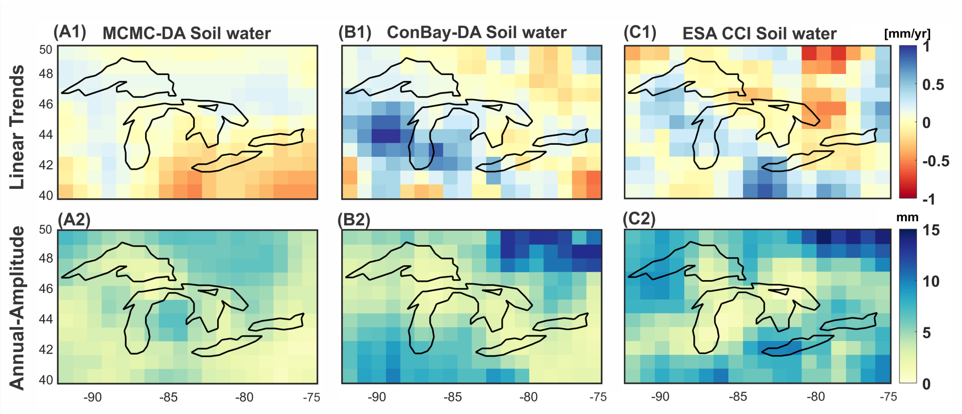

Fig. 9.9 Long-term linear trends and annual-amplitudes fitted to the soil water storage changes derived from (A1, A2) MCMC-DA, (B1, B2) ConBay-DA, (C1, C2) ESA CCI, within the GL area, covering 2003-2017.

can be seen in the northeast of the region, where the annual-amplitude of the MCMC-DA soil water storage is less than 5 mm, while after implementing ConBay-DA this value is increased up to $\sim$ 12 mm, on average, which is close to those of ESA CCI product, i.e., 13.2 mm. In the south part of GL, the annual-amplitudes of the MCMC-DA soil water storage are estimated to be less than 3 mm, while those of the ConBay-DA outputs, and the ESA CCI products, are estimated to be up to $\sim$ 8 mm. Therefore, one can say that ConBay-DA improves the estimation of soil water storage by 62%, compared to MCMC-DA, against independent ESA CCI validation data set.

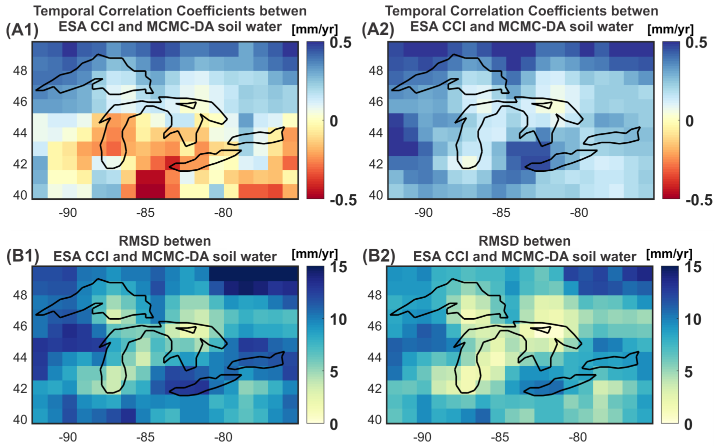

Fig. 9.10 Temporal Correlation Coefficients (A1, A2) and RMSD (B1, B2) between ESA CCI soil water storage, and soil water changes derived from MCMC-DA and ConBay-DA approach within GL are, between 2003-2017.

The reason of finding these differences between MCMC-DA and ConBay-DA is the strong contribution of PGR in the GRACE signal and the fact that these values are uncertain. Therefore, within the application of ConBay-DA and MCMC-DA, to optimise updated value of the modelled Total-WSC at each time step, with respect to GRACE data, the unknown state parameters corresponding to the GIA model output (β_t in Eq. (6.1)) gains the highest value of the posterior probability distribution compared to those of hydrological compartments (θ_t in Eq. (6.1)). Therefore, the largest updates are introduced to the GIA model output in order to reduce RMSD between GRACE and model Total-WSC. As already discussed in Section 9.3, without using in-situ GNSS measurement to constrain the estimation of PGR (as is the case in MCMC-DA), the PGR accepts a large portion of the update because statistically it will be more likely that the biases between *a priori* information and observations will be reduced. This update therefore does not care about the physical representation of PGR, which should be purely linear. As a result, the MCMC-DA estimation of PGR contains seasonal component and its hydrological estimation is negatively affected as the validation in Fig. 9.9 represents.

In order to visualise differences between the estimated parameters from MCMC-DA and ConBay-DA, a comparison of posterior values of the unknown states $\theta_t^{soilwater}$, $\theta_t^{groundwater}$, and β_t^{PGR} are shown in Fig. 9.11. The results correspond to only one spatial grid point with the latitude of 45° and the longitude of $-90°$, for 2003–2017. These parameters (and others computed for other grids) can be used in Eq. (6.16) to update soil water storage, groundwater storage, and PGR rates, respectively. These results indicate that the posterior values of β_t^{PGR} from MCMC-DA are higher than those of

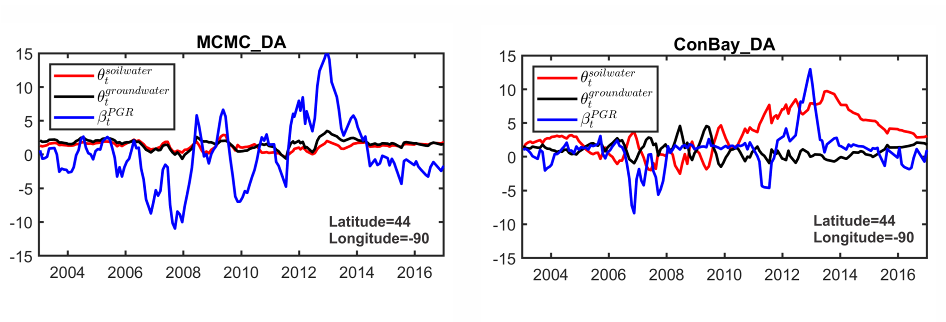

Fig. 9.11 The Posterior values of the unknown state-space parameters Θ_t and β_t corresponding to the soil water storage ($\theta_t^{soilwater}$), groundwater storage ($\theta_t^{groundwater}$), and PGR uplift rate (β_t^{PGR}), derived from MCMC-DA (without using GNSS measurements), and ConBay-DA, for the period of 2003–2017. The posterior values are shown for a spatial grid point with latitude of 45° and longitude of $-90°$.

$\theta_t^{soilwater}$, $\theta_t^{groundwater}$. Therefore, in MCMC-DA, the required values to update *a priori* information and to reduce RMSD are mostly introduced to the GIA model output, while the updated values of soil water and groundwater storage are negligible. After merging in-situ GNSS measurements with GRACE data in ConBay-DA, the posterior value of the state parameters $\theta_t^{soilwater}$ and $\theta_t^{groundwater}$ are considerably increased, compared to the MCMC-DA results, while the posterior value of β_t^{PGR} is decreased. Consequently, soil water and groundwater storage changes have gained larger updates in ConBay-DA, compared to the MCMC-DA results, which can be seen in Fig. 9.9. This shows that introducing GNSS observations in an extra step controls the updates and consequently improves the estimation.

9.5 Summary and Conclusion

ConBay-DA was implemented in this chapter to simultaneously use GRACE Total-WSC (as a summation of TWSC and EWHs of PGR uplift rate) and in-situ GNSS measurements to update *a priori* information of hydrological and surface deformation, where they were introduced using the W3RA hydrological model outputs and the ICE5G-VM2 GIA model. As a case study, the Great Lakes (GL) area and the period of 2003–2017 was chosen. The main goal was to evaluate the performance of the ConBay-DA to separate PGR uplift rates and water storage changes from GRACE data, while using GNSS measurements to constrain their estimates. To assess the impact of the constraint equation on the signal separation results, an extension to the MCMC-DA (unconstrained Bayesian-DA) was performed to separate land hydrology and surface deformation from GRACE Total-WSC. The PGR uplift rates derived from MCMC-DA and ConBay-DA were compared against in-situ GNSS measurements, and soil water storage changes derived from both techniques were validated against ESA CCI soil water storage, as an independent validation data set.

The numerical results indicated after implementing ConBay-DA that RMSD between model-derived Total-WSC and GRACE Total-WSC reduced by $\sim 70\%$, on average, and the bias between GIA model output and the in-situ GNSS observation reduced by $\sim 90\%$. MCMC-DA results showed the same improvement in the estimation of Total-WSC, but the bias between GIA model output and in-situ GNSS measurements reduced by 56% after implementing MCMC-DA. This reduction, however, was considerably smaller than those of ConBay-DA. The estimated values of PGR through MCMC-DA contained annual component (see Fig. 9.6), which is not physically realistic. ConBay-DA reduced this uncertainty by 86%, using the in-situ GNSS measurements to constrain the updated value of the ICE5G-VM2 GIA model.

Validation of top-layer soil water estimates, derived from MCMC-DA and ConBay-DA, against ESA CCI soil water storage supports the hypothesis that, using in-situ GNSS measurements in the Bayesian signal separation approach, improves the estimation of hydrological signals in terms of both linear trends and seasonality. From the results, it can be concluded that using the GNSS constraint equation in ConBay-DA defines an upper and lower boundary to update the GIA model output, which allows us to introduce more updates to the hydrological signals compared to those of MCMC-DA. Therefore, the GNSS constraint equation not only optimises the estimation of the land surface deformation, but also improves estimation of the hydrological signals, in terms of both linear trends and seasonality. A short summary of the setup to implement MCMC-DA and ConBay-DA and statistical comparisons of the results are shown in Fig. 9.12.

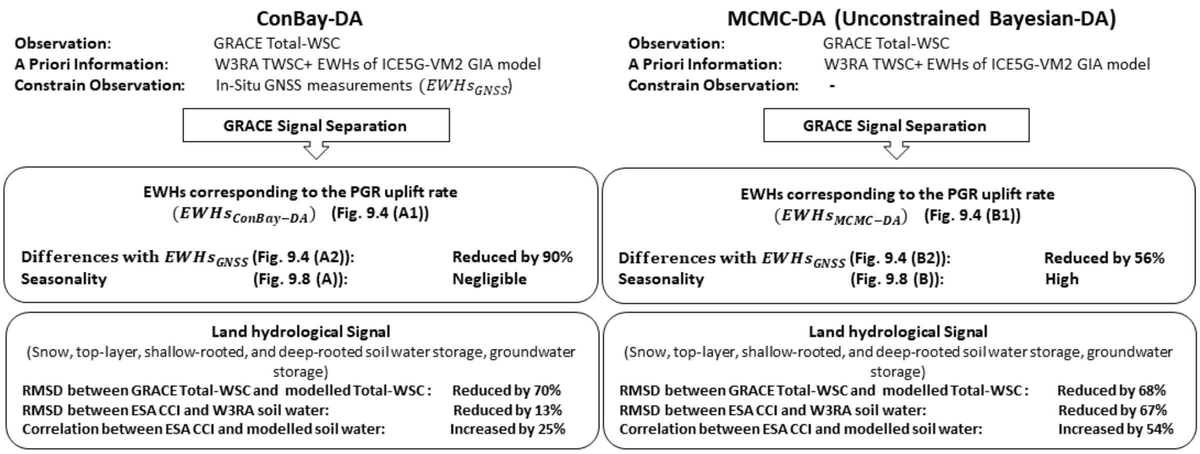

Fig. 9.12 An overview of the setup to implement MCMC-DA and ConBay-DA, and a statistical comparison between their application in G to separate TWSC and PGR from GRACE data.

Chapter 10

Conclusion and Outlook

This thesis is aimed at developing flexible data-model fusion frameworks, based on Bayesian methods, to merge multi-model states with GRACE(-FO) and other geodetic observations to: **(a)** update (modify) water states and surface deformation simulated by models, and **(b)** separate GRACE(-FO) superimposed signals into their components. To this end, in Chapter 1, the importance of using satellite geodetic techniques and models to study the Earth system, as well as the main hypothesis and objectives of the research were introduced that mainly focused on establishing Bayesian frameworks for: (1) merging multiple hydrological model outputs with GRACE(-FO) TWSC, (2) separating GRACE(-FO) TWSC and simultaneously accounting for temporal dependency between model states, and (3) a hierarchical integration of a priori information about the Earth system with multiple geodetic observations.

In Chapter 2, GRACE(-FO) data and its processing steps, as well as other geodetic observations and auxiliary models, used in this study, were introduced, while an overview of the Bayesian inference was provided in Chapter 3. As the first step toward the Bayesian framework, Dynamic Model Data Averaging (DMDA) was proposed and formulated in Chapter 4 to merge multiple *a priori* information (e.g., multiple hydrological model outputs) with a set of observation (e.g., GRACE TWSC) to address the above objective (1). In Chapter 5, the Markov Chain Monte Carlo Data Assimilation technique (MCMC-DA) was developed to update a priori information based on a set of observation by recursively estimating the unknown state parameters within a state-space model, while temporal dependency between the parameters was estimated simultaneously through a Bayesian sampling approach. This addressed objective (2). To address objective (3), a hierarchical Bayesian optimization approach, Constrained Bayesian-Data Assimilation (ConBay-DA), was formulated in Chapter 6 to merge two set of observations (e.g., GRACE and GNSS measurements) with *a priori* information, where the second set of observation is used in a hierarchical level to constrain a specific compartment of the water cycle (water storage changes due to the PGR uplift rates).

The applicability of the DMDA, MCMC-DA and ConBay-DA formulatiuons to separate GRACE signal was investigated in Chapter 7, Chapter 8, and Chapter 9 via several real case studies. In the following, the main conclusions and outlook for this research trajectory are presented.

10.1 Conclusions

10.1.1 Comparison of DMDA, MCMC-DA, and ConBay-DA approaches

DMDA, MCMC-DA, and ConBay-DA were formulated in this study based on a linear state-space model to define a relationship between observations and *a priori* information in a dynamic system, while the state parameters and the temporal dependency between them were unknown and were estimated through Bayesian approaches. The proposed approaches in this study are flexible in accounting for the full error covariance matrix of the GRACE(-FO) TWSC, and the uncertainty of the model simulations, which is allowed to vary in time.

The main difference between DMDA and two other techniques, is that in DMDA a Kalman Filter (*Kalman*, 1960) approach (Eqs. (4.11) and (4.12)) is formulated to estimate the unknown state parameters, while a forgetting factor of $0 < \lambda < 1$ is empirically estimate to define the unknown temporal dependency between the water states (see Eq. (4.8)). It is due to the fact that the magnitude of changes in water storage components depends on the history of hydrological processes, which is unknown, and there is not enough physical knowledge about how this dependency varies over time. In MCMC-DA the Kalman Filter of DMDA is replaced by a combination of forward filtering-backward smoothing approach (*Kitagawa*, 1987) and Gibbs sampling (*Gelfand and Smith*, 1990; *Smith and Roberts*, 1993), where the first builds the relationship between observations and *a priori* information, and the second accounts for unknown correlations between model states, which allows to introduce more realistic updates to the individual water states, compared to the DMDA.

This hypothesis was tested by comparing the performance of DMDA and MCMC-DA to merge W3RA model outputs with GRACE data on a $0.125° \times 0.125°$ spatial grid points within CONUS, between 2003–2017 (see Section 8.4). TWSC derived from both techniques were compared with GRACE observations in terms of RMSD and phase differences of annual amplitudes. The results indicated that although both techniques performed well to improve model simulations, MCMC-DA TWSC is more close to GRACE TWSC (RMSD and phase difference of zero) compared to the DMDA results. The finding of this study, clearly showed that estimating the unknown temporal dependency using a Bayesian sampling approach in MCMC-DA improves the estimation of the water states, compared to the empirically defined this parameters in DMDA approach.

DMDA is formulated in this study to merge a set of observation with multiple *a priori* information in two steps: (1) a Kalman Filter to solve the state-space model between a set of observations and *a priori* information. For multiple *a priori* information, the Kalman Filter is implemented to solve the state-space model between the observations and each set of *a priori* information; (2) Bayesian Model Averaging (BMA, *Hoeting et al.*, 1999) to provide a time-variable weights to average the water states derived from multiple *a priori* information of the first step, yielding the best fit to the observations, while the implicit specification of the transition matrix is avoided using the forgetting factor of $0 < \alpha < 1$.

In ConBay-DA, which is an extension of the MCMC-DA, a hierarchical Metropolis-Hastings algorithm (*Chib and Greenberg*, 1995) is combined with the forward filtering-backward smoothing approach and Gibbs sampling algorithm of MCMC-DA to constrain a specific compartments of the water cycle (e.g., water storage changes due to the PGR uplift rate) based on a second set of observation (e.g., GNSS measurement).

The formulation and presentation of MCMC-DA and ConBay-DA in Chapters 5 and 6 are based on merging the observations with only one set of *a priori* information, as the simplest case study. The formulation of this frameworks, however, can be extended to merge multiple *a priori* information with GRACE observation using the second part of DMDA, i.e., BMA in Section 4.3. Therefore, all three Bayesian techniques proposed in this study have the potential to merge and to evaluate the behaviour of multiple model outputs against GRACE(-FO) data.

The advantage of the DMDA approach, compared to the MCMC-DA is that the implementation of the DMDA is more computationally efficient compared to the two other techniques, which makes it a unique approach to integrate multiple hydrological models with GRACE(-FO) data, over a large area and for high-spatial resolution applications. As an evidence, for example, it took almost three weeks to estimate the posterior probability distributions of the water states and their uncertainties within CONUS through MCMC-DA algorithm, using a general purpose computer (MATLAB toolbox), while DMDA provided these estimations in less than 1 hour. Therefore, it can be say that, although MCMC-DA introduces more realistic updates to the water storage changes, it is computationally insufficient, compared to DMDA, to merge multiple hydrological model outputs with GRACE data in a large scale area, e.g., continental to global scale, and for a high spatial (e.g., less than 12 km) and high temporal resolution applications (e.g., daily data).

Comparison between the application of MCMC-DA and ConBay-DA to estimate land hydrology and surface deformation in Chapter 9 indicated that using a second source of observations (e.g., GNSS measurements) along with the GRACE data, can improve the estimation of both hydrological and surface deformation signals. The reason could be due to the fact that, the second set of observation provide useful information about a specific compartment of the water cycle or geophysical signal, which can optimise the performance of signal separation techniques.

Benefits of the Proposed Bayesian Frameworks Compared to Previous Studies

Application of Bayesian techniques to merge observations and models is getting more attention in geophysical and hydrological studies. From the introduced Bayesian methods to merge GRACE observations with model simulations in Chapter 4, Chapter 5, and Chapter 6, the use of Bayesian methods is quite recent in hydrological studies. In recent years, Data Assimilation (DA) and simultaneous Calibration/Data Assimilation (C/DA) have been used in various studies to merge GRACE(-FO) TWSC with hydrological model simulations or other Earth Observations (EO) data using a sequential integration approach such as the Ensemble Kalman Filtering (EnKF, *Evensen*, 1994) or its extensions.

The cost functions of DA and C/DA approaches to update model parameters, conditional on the measurement data, are formulated based on Bayes' theorem (e.g., *Evensen*, 2003; *Fang et al.*, 2018; *Schumacher*, 2016). *Schumacher* (2016) discussed that the statistical information used in EnKF-DA is restricted to the covariance matrices of the observations and models. Therefore, the observation error model and the spatial resolution of GRACE TWSC has a significant influence on C/DA results. Instead of limiting the statistical information in the data to the use of their covariances, a sampling of their Probability Density Functions (PDFs) and a Bayesian optimization approach is adopted in DMDA, MCMC-DA and ConBay-DA that results in more realistic estimations of states and their errors. Moreover, the proposed approaches are flexible for any spatial resolution of the observations and models, which does not have any influence on final results. More important, unlike the previous DA, the proposed Bayesian techniques are implemented in an offline mode, where we do not need to run the models, and we only use the outputs of the available hydrological models to merge with the observations.

Particle Filter (PF) and Particle Smoother (PS) are other types of Bayesian approaches (*Särkkä*, 2013), which have been used in some hydrological applications such as *Plaza Guingla et al.* (2013); *Weerts and El Serafy* (2006) to assimilate the observations into the models. These techniques use a set of particles (also called samples) to represent the posterior distribution of some stochastic process given noisy and/or partial observations. The positive point of these technique is that the state-space model can be non-linear and the initial state and noise distributions can take any form required (*Del Moral*, 1997). Moreover, PF techniques provide a well-established methodology for generating samples from the required distribution without requiring assumptions about the state-space model or the state distributions.

The rate of convergence of the approximate probability distribution until attainment of the true posterior in PF approaches is inversely proportional to the number of particles used in the filter (*Bain and Crisan*, 2008). This means that the filter perfectly approximates the posterior distribution when the number of particles tends to infinity. However, since the computational cost of PF grows with the number of particles, choosing a specific number of particles in the design of filters is a key parameter for these methods. Therefore, PF and PS might not be efficient for high-dimensional fusion tasks (e.g., *Bain and Crisan*, 2008; *Snyder et al.*, 2008) such as the global hydrological application presented here. The applications of the proposed Bayesian approach in this study indicated that these techniques provide the ability to deal with high-dimensional fusion tasks, and its computational load is much lower than PF and PS.

BMA has been used in many climate hydrology studies to merge (multi) models with measurements. For example, *Long et al.* (2017) applied BMA to average multiple GRACE TWSC products and hydrological models to analyse spatial and temporal variability of global TWSC. A model-data synthesis method, based on Bayesian modelling, has also been used in *Sha et al.* (2019) to update a global GIA model using GPS data. Their studies, however, did not assess the update of individual surface and sub-surface water storage estimates.

Previous data-model fusion techniques are often applied using only one set of data (e.g., GRACE, GNSS, remote sensing spoil water storage, or altimetry data alone) to merge with only one physical or geographical model with some notable exceptions (*Girotto et al.*, 2017; *Tian et al.*, 2017; *Van Dijk et al.*, 2018), or they are limited to a particular region (not global), or to a particular component of the water cycle. The proposed Bayesian approaches employed in this PhD thesis were formulated to simultaneously estimate all surface and sub-surface water storage changes, as well as surface deformation from GRACE(-FO) data. Moreover, from the application parts of DMDA, MCMC-DA and ConBay-DA in Chapters 7, 8, 9 it was shown that these approaches are flexible to implement in both regional and global scales, and for different spatial and temporal resolutions to study different aspects of the water cycle.

Application of the Proposed Bayesian Techniques to Improving Hydrological Water States and Land Surface Deformation

A fundamental question regarding the merging of GRACE(-FO) TWSC with model outputs is whether the integration of GRACE(-FO) TWSC into hydrological models improves the representation of total and individual water states. To answer this question, various case studies were considered to demonstrate how the established Bayesian framework were employed to update *a priori* model-derived estimates, using GRACE data, while rigorously accounting for uncertainties in models and measurements.

Water storage changes derived from implementing the proposed approaches in Chapters 7, 8, and 9 were interpreted and evaluated by performing various comparisons between the original model outputs and the Bayesian results, with respect to metrics such as Root Mean Square of Differences (RMSD) and temporal correlation coefficients. Validations were done against independent measurements such as in-situ US Geological Survey (USGS) groundwater level observations and the European Space Agency (ESA)'s Climate Change Initiative (CCI) soil data, introduced in Section 2.5.

Previous Data Assimilation (DA) attempts (*Girotto et al.*, 2016) showed that a single GRACE DA might introduce unrealistic signals to the soil water storage compartments. The results in this study indicated that such errors could be considerably reduced using the proposed Bayesian techniques (such as MCMC-DA), where the uncertainties and the dynamic evolution of water states are rigorously accounted for updating model outputs. For example, the MCMC-DA results showed that the soil water storage estimates were improved within the CONUS unlike the previous attempts by *Girotto et al.* (2016). We also found a stronger linear trend in MCMC-DA soil water storage across the CONUS, compared to W3RA (changing from ± 0.5 mm/yr to ± 2 mm/yr), which is closer to independent estimates from the ESA CCI. MCMC-DA also improved the estimation of soil water storage in regions with high forest intensity, where ESA CCI and hydrological models have difficulties in capturing the soil-vegetation-atmosphere continuum. Moreover, the results derived from implementing ConBay-DA within the GL area, and validation against ESA CCI soil water storage, indicated that merging EO data with models can improve the estimation of soil water storage changes, in terms of both linear trends and seasonality.

To demonstrate the impact of integrating GRACE TWSC into hydrological models on the inter- and intra-annual components of water storage changes, water storage changes associated with the El Niño Southern Oscillation (ENSO, *Barnston and Livezey*, 1987) were investigated. From this investigation, we found that the modification of water storage changes are even beyond linear trends and seasonality, since some climate modes, such as that of ENSO, are influenced after implementing the proposed Bayesian approaches.

Furthermore, it was shown that the proposed Bayesian techniques are able to improve the timing of water storage changes estimates. For example, in Section 8.4, it was shown that both MCMC-DA and DMDA considerably reduced the phase differences between modelled and measured TWSC within CONUS. The averaged time series of groundwater, soil water, and TWSC in Texas (south of CONUS) are compared with those of independent validation data sets, and it was found that MCMC-DA advanced the phase of the W3RA TWSC time series by 2 months, on average, as shown by Fig. 8.16.

The results derived from the application of the Bayesian approaches showed that the model simulations of TWSC were always improved after merging with GRACE data. This improvement was also seen for the individual water storage compartments, such as the groundwater and soil water storage. However, the quality of the GRACE signal separation to its individual compartments was not always skilful in all the regions of the case studies, and in the whole time period. A possible reason might be due to the large interactions and/or the similarity of inter-annual cycles between different water storage compartments, which are not well reflected in *a priori* information. Moreover, possible tipping points or abrupt changes caused by various phenomena (e.g., earthquakes, sudden atmospheric pressure changes, hurricanes, floods and desertification of grasslands) are difficult to simulate in models but reflected in GRACE(-FO) data. Identifying the exact source of these discrepancies and quantifying them will need further research (see e.g., *Peng et al.*, 2019).

From the application of ConBay-DA within the GL area, we found that a lack of *a priori information* on water storage (e.g., lack of surface water storage in W3RA model outputs) may increase the uncertainties of the signal separation results. For example, we found a small annual amplitude in the separated PGR signal from GRACE data, which is not physically realistic (see Fig. 9.6). It may because surface water storage is the dominant signal within the GL area, which is detected by the GRACE(-FO) mission. W3RA, however, does not contain surface water storage. The lack of *a priori* information about this compartment may be reflected as an uncertainty within the GRACE signal separation results. To reduce this uncertainty, one solution is to remove the surface water storage of the GL area from GRACE data, using altimetry measurements. The uncertainty of the altimetry measurements, however, will be reflected within the GRACE data, which will affect the final estimations of the water states. The other solution, is to use the altimetry observation as an *a priori information* for surface water storage changes. However, one of the challenging problems of using altimetry data is the spatial and temporal resolution mismatch between observation and models. Finding a spatial/temporal resolution mismatch is a subject for further research.

10.2 Outlook

Improving the Bayesian algorithm

Bayesian data-model fusion techniques in this thesis were formulated with the assumption that all the data used in this study, and their uncertainties, are Gaussian distributed. To assess the impact of different assumptions about the error distributions of the observations (whether Gaussian or non-Gaussian) on the final estimates, the Bayesian formulations of this study need to be extended by using a non-Gaussian state-space model to define the relationship between observations and models. As a result, Kalman filtering in DMDA and forward-filtering backward-smoothing recursion approach in MCMC-DA and ConBay-DA, should be replaced by an alternative sampling approach to obtain accurate approximations of the likelihood function. To understand to what extent these techniques and different observation error distributions might impact the final estimates future research and case studies must be designed.

To reduce the computational load of DMDA, instead of implementing an MCMC sampling algorithm to estimate the transition matrix between models, a BMA forgetting factor of $\alpha = 0.9$ was considered (see Section 4.3). This factor defined the temporal smoothing for predicting posterior probability values at each time step. Based on the same logic that, in MCMC-DA, we replaced the Kalman forgetting factor of DMDA with the Gibbs sampling, in future, one might apply an efficient sampling approach instead of using an empirical (α) value. This replacement might lead to better estimations of temporal weights in the DMDA approach.

The formulation of the ConBay-DA technique can be extended to merge multiple (more than two) geodetic or Earth Observations (EO) data sets, such as satellite altimetry and remote sensing soil water data, which may provide more information about different hydrological compartments. Merging various EO data with models is a challenging problem due to the different spectral/spatial and temporal resolution mismatch. Therefore, finding an optimal matching operator to solve spectral/spatial mismatch between them and evaluating their impact on final results should be assessed in future studies.

A secondary application of the proposed Bayesian techniques can also be devoted to their application for predicting (or extrapolating) water storage estimates. To achieve this purpose, however, the Bayesian formulation needs to be extended. For example, the DMDA weights can be used to identify the best model in different river basins and seasons. By analysing this information and knowing the observed TWSC by GRACE-FO, one can use a combination of different model runs (weighted by the DMDA outputs) and extrapolate the surface and sub-surface water storage estimates.

Further Applications

In this study, the GRACE level 2 product was applied to estimate GRACE TWSC as observations in the applications of the proposed approaches. It is known to the science community that although there are similarities in overall patterns of TWSC estimates from existing GRACE solutions (e.g., GRACE level 2 and mascon products, *Watkins et al.*, 2015), there are also some differences in their long-term trends and the seasonal magnitudes. These differences are more pronounced in coastal regions such as the southeastern CONUS. In general, there is no perfect solution for selecting the best gravity recovery technique for coastal regions, which is why alternative inversion techniques are being developed (*Ferreira et al.*, 2020; *Yang et al.*, 2017). The influence of using various GRACE products on DMDA, MCMC-DA, and ConBay-DA estimations will be subject to future investigations.

The error variance-covariance matrix of the GRACE field estimates is a key parameter in Bayesian signal separation approaches to estimate the posterior PDF of the unknown state parameters and the temporal dependency between them (see Eq. (4.11), Eq. (5.1), and Eq. (6.6)). The uncertainty of error-covariance matrix of GRACE data is then reflected on signal separation results. understand the impact of different approximations of the full error variance-covariance matrix of GRACE observations on Bayesian signal separation is a subject for further investigations.

MCMC-DA was implemented in Chapter 8 to merge outputs from one model, i.e., the W3RA water balance model, with GRACE TWSC. The application of MCMC-DA can be extended to merge multi-hydrological model outputs with GRACE TWSC, using the Bayesian weighting approach of DMDA. Such extensions and evaluating weights assigned to these models should be assessed in future research.

The application of the ConBay-DA to merge GRACE Total-WSC and in-situ GNSS time series in a global scale is a challenging problem due to the large spacial and temporal gaps within the time series. Therefore, defining a matching operator is necessary to establish a relationship between the observations and *a priori* information, which will be a subject to future investigations.

The Bayesian data-model fusion techniques, introduced in this thesis, have the potential to be used in different climate and hydrological applications to compare available hydrological, land-surface and climate models against observations. It can also be used to generate ensemble means from multi-model outputs such as climate projections.

References

Agee, J. K. (1996), *Fire ecology of Pacific Northwest forests*, Island press.

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

Anyah, R. O., E. Forootan, J. L. Awange, and M. Khaki (2018), Understanding linkages between global climate indices and terrestrial water storage changes over Africa using GRACE products, *Science of The Total Environment*, *635*, 1405–1416, doi:10.1016/j.scitotenv.2018.04.159.

Awange, J. L., M. A. Sharifi, O. Baur, W. Keller, W. E. Featherstone, and M. Kuhn (2009), GRACE hydrological monitoring of Australia: Current limitations and future prospects, *Journal of Spatial Science*, *54*(1), 23–36, doi:10.1080/14498596.2009.9635164.

Awange, J. L., E. Forootan, M. Kuhn, J. Kusche, and B. Heck (2014), Water storage changes and climate variability within the Nile Basin between 2002 and 2011, *Advances in Water Resources*, *73*, 1–15, doi:10.1016/j.advwatres.2014.06.010.

Awange, J. L., V. Ferreira, E. Forootan, Khandu, S. Andam-Akorful, N. Agutu, and X. He (2016), Uncertainties in remotely sensed precipitation data over Africa, *International Journal of Climatology*, *36*(1), 303–323, doi:10.1002/joc.4346.

Bain, A., and D. Crisan (2008), *Fundamentals of stochastic filtering*, vol. 60, Springer Science & Business Media.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2004), *Hierarchical modeling and analysis for spatial data*, Chapman and Hall/CRC.

Barnston, A. G., and R. E. Livezey (1987), Classification, seasonality and persistence of low-frequency atmospheric circulation patterns, *Monthly weather review*, *115*(6), 1083–1126, doi:10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2.

Beal, M. J., Z. Ghahramani, et al. (2006), Variational Bayesian learning of directed graphical models with hidden variables, *Bayesian Analysis*, *1*(4), 793–831.

Beaty, R. M., and A. H. Taylor (2007), Fire disturbance and forest structure in old-growth mixed conifer forests in the northern Sierra Nevada, California, *Journal of Vegetation Science*, *18*(6), 879–890, doi:10.1111/j.1654-1103.2007.tb02604.x.

Becker, M., W. LLovel, A. Cazenave, A. Güntner, and J.-F. Crétaux (2010), Recent hydrological behavior of the East African great lakes region inferred from GRACE, satellite altimetry and rainfall observations, *Comptes Rendus Geoscience*, *342*(3), 223–233, doi:10.1016/j.crte.2009.12.010.

Bengtsson, T., P. Bickel, B. Li, et al. (2008), Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, in *Probability and statistics: Essays in honor of David A. Freedman*, pp. 316–334, Institute of Mathematical Statistics, doi:10.1214/193940307000000518.

Benito, E., J. Santiago, E. de Blas, and M. Varela (2003), Deforestation of water- repellent soils in Galicia (NW Spain): effects on surface runoff and erosion under simulated rainfall, *Earth Surface Processes and Landforms*, *28*(2), 145–155, doi:10.1002/esp.431.

Berger, J. O., and J. M. Bernardo (1989), Estimating a product of means: Bayesian analysis with reference priors, *Journal of the American Statistical Association*, *84*(405), 200–207.

Bernstein, D. S. (2005), *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*, vol. 41, Princeton university press Princeton.

Berry, P., J. Garlick, J. Freeman, and E. Mathers (2005), Global inland water monitoring from multi-mission altimetry, *Geophysical Research Letters*, *32*(16), doi:10.1029/2005GL022814.

Bertsekas, D. P., D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas (1995), *Dynamic programming and optimal control*, vol. 1, Athena scientific Belmont, MA.

Bezděk, A., J. Sebera, J. Teixeira da Encarnação, and J. Klokočník (2016), Time-variable gravity fields derived from GPS tracking of Swarm, *Geophysical Journal International*, *205*(3), 1665–1669, doi:10.1093/gji/ggw094.

Bierman, G. J. (1973), Fixed interval smoothing with discrete measurements, *International Journal of Control*, *18*(1), 65–75, doi:10.1080/00207177308932487.

Blewitt, G., D. Lavallée, P. Clarke, and K. Nurutdinov (2001), A new global mode of Earth deformation: Seasonal cycle detected, *Science*, *294*(5550), 2342–2345, doi:10.1126/science.1065328.

Bolten, J. D., W. T. Crow, X. Zhan, T. J. Jackson, and C. A. Reynolds (2009), Evaluating the utility of remotely sensed soil moisture retrievals for operational agricultural drought monitoring, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *3*(1), 57–66, doi:10.1109/JSTARS.2009.2037163.

Brocca, L., L. Ciabatta, C. Massari, S. Camici, and A. Tarpanelli (2017), Soil moisture for hydrological applications: open questions and new opportunities, *Water*, *9*(2), 140, doi:10.3390/w9020140.

Brooks, S. P., and G. O. Roberts (1998), Assessing convergence of Markov chain Monte Carlo algorithms, *Statistics and Computing*, *8*(4), 319–335.

Brown, E. N., L. M. Frank, D. Tang, M. C. Quirk, and M. A. Wilson (1998), A statistical paradigm for neural spike train decoding applied to position prediction from ensemble firing patterns of rat hippocampal place cells, *Journal of Neuroscience*, *18*(18), 7411–7425, doi: 10.1523/JNEUROSCI.18-18-07411.1998.

Cardoso, J.-F. (1999), High-order contrasts for independent component analysis, *Neural computation*, *11*(1), 157–192, doi:10.1162/089976699300016863.

Carter, C. K., and R. Kohn (1994), On Gibbs sampling for state space models, *Biometrika*, *81*(3), 541–553, doi:10.1093/biomet/81.3.541.

Casella, G., C. P. Robert, M. T. Wells, et al. (2004), Generalized accept-reject sampling schemes, in *A Festschrift for Herman Rubin*, pp. 342–347, Institute of Mathematical Statistics.

Castellazzi, P., R. Martel, A. Rivera, J. Huang, G. Pavlic, A. I. Calderhead, E. Chaussard, J. Garfias, and J. Salas (2016), Groundwater depletion in Central Mexico: Use of GRACE and InSAR to support water resources management, *Water resources research*, *52*(8), 5985–6003, doi: 10.1002/2015WR018211.

Cecilia Peralta, F., M. James H., and J. M. Wallace (2016), Proxy representation of Arctic ocean bottom pressure variability: Bridging gaps in GRACE observations, *Geophysical Research Letter*, *43*(17), 9183–9199, doi:10.1002/2016GL070137.

Chambers, D. P. (2006), Observing seasonal steric sea level variations with GRACE and satellite altimetry, *Journal of Geophysical Research: Oceans*, *111*(C3), doi:10.1029/2005JC002914.

Chan, K. S., and C. J. Geyer (1994), Discussion: Markov chains for exploring posterior distributions, *The Annals of Statistics*, *22*(4), 1747–1758, doi:10.1214/aos/1176325754.

Chen, J., C. Wilson, B. Tapley, and J. Ries (2004), Low degree gravitational changes from GRACE: validation and interpretation, *Geophysical Research Letters*, *31*(22), doi:10.1029/2004GL021670.

Chen, J., C. Wilson, D. Blankenship, and B. Tapley (2006), Antarctic mass rates from GRACE, *Geophysical research letters*, *33*(11), doi:10.1029/2006GL026369.

Chen, J. L., C. R. Wilson, J. S. Famiglietti, and M. Rodell (2007), Attenuation effect on seasonal basin-scale water storage changes from GRACE time-variable gravity, *Journal of Geodesy*, *81*(4), 237–245, doi:10.1007/s00190-006-0104-2.

Chen, J. L., C. R. Wilson, B. D. Tapley, Z. L. Yang, and G. Y. Niu (2009), 2005 drought event in the Amazon River basin as measured by GRACE and estimated by climate models, *Journal of Geophysical Research: Solid Earth*, *114*(B5), doi:10.1029/2008JB006056.

Chen, J. L., C. R. Wilson, and B. D. Tapley (2010a), The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE, *Water Resources Research*, *46*(12), doi:10.1029/2010WR009383.

Chen, Z., R. Barbieri, and E. N. Brown (2010b), State-space modeling of neural spike train and behavioral data, in *Statistical signal processing for neuroscience and neurotechnology*, pp. 175–218, Elsevier, doi:10.1016/B978-0-12-375027-3.00006-5.

Chib, S., and E. Greenberg (1995), Understanding the metropolis-hastings algorithm, *The american statistician*, *49*(4), 327–335, doi:10.1080/00031305.1995.10476177.

Chui, C. K., G. Chen, et al. (2017), *Kalman filtering*, Springer.

Cogley, T. (2005), How fast can the new economy grow? A Bayesian analysis of the evolution of trend growth, *Journal of macroeconomics*, *27*(2), 179–207, doi:10.1016/j.jmacro.2003.11.018.

Conroy, M. J., C. R. Allen, J. T. Peterson, L. Pritchard, and C. T. Moore (2003), Landscape Change in the Southern Piedmont: Challenges, Solutions, and Uncertainty Across Scales, *Conservation Ecology*, *8*(2), 3.

Cook, E. R., R. Seager, M. A. Cane, and D. W. Stahle (2007), North American drought: Reconstructions, causes, and consequences, *Earth-Science Reviews*, *81*(1-2), 93–134, doi: 10.1016/j.earscirev.2006.12.002.

Council, N. R., et al. (1999), *Global environmental change: Research pathways for the next decade*, National Academies Press.

Cowles, M. K., and B. P. Carlin (1996), Markov chain Monte Carlo convergence diagnostics: a comparative review, *Journal of the American Statistical Association*, *91*(434), 883–904, doi: 10.1080/01621459.1996.10476956#.X-9D4ti2mUk.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H. Neumayer (2013), GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005: revised edition, January 2013, pp. 1–21.

Daum, F., and J. Huang (2003), Curse of dimensionality and particle filters, in *2003 IEEE Aerospace Conference Proceedings (Cat. No. 03TH8652)*, vol. 4, pp. 4_1979–4_1993, IEEE, doi:10.1109/AERO.2003.1235126.

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

Dee, D. P., S. M. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, 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, doi:10.1002/qj.828.

Deisenroth, M. P., D. Fox, and C. E. Rasmussen (2015), Gaussian processes for data-efficient learning in robotics and control, *IEEE transactions on pattern analysis and machine intelligence*, *37*(2), 408–423, doi:10.1109/TPAMI.2013.218.

Del Moral, P. (1997), Nonlinear filtering: Interacting particle resolution, *Comptes Rendus de l'Académie des Sciences-Series I-Mathematics*, *325*(6), 653–658, doi:10.1016/S0764-4442(97)84778-7.

Dickey, J. O., S. L. Marcus, O. de Viron, and I. Fukumori (2002), Recent earth oblateness variations: Unraveling climate and postglacial rebound effects, *Science*, *298*(5600), 1975–1977, doi:10.1126/science.1077777.

Dieter, C. A. (2018), *Water availability and use science program: Estimated use of water in the United States in 2015*, Geological Survey.

Diffenbaugh, N. S., D. L. Swain, and D. Touma (2015), Anthropogenic warming has increased drought risk in California, *Proceedings of the National Academy of Sciences*, *112*(13), 3931–3936, doi:10.1073/pnas.1422385112.

Doell, P., H. Mueller Schmied, C. Schuh, F. T. Portmann, and A. Eicker (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, doi:10.1002/2014WR015595.

Doerr, S. H., and A. D. Thomas (2000), The role of soil moisture in controlling water repellency : new evidence from forest soils in Portugal, *Journal of Hydrology*, *231*, 134–147, doi:10.1016/S0022-1694(00)00190-6.

Döll, P., F. Kaspar, and B. Lehner (2003), A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, *270*(1-2), 105–134, doi: 10.1016/S0022-1694(02)00283-4.

Döll, P., H. Mueller Schmied, C. Schuh, F. T. Portmann, and A. Eicker (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, doi:10.1002/2014WR015595.

Döll, P., H. Douville, A. Güntner, H. M. Schmied, and Y. Wada (2016), Modelling freshwater resources at the global scale: challenges and prospects, *Surveys in Geophysics*, *37*(2), 195–221, doi:10.1007%2Fs10712-015-9343-1.

Dong, D., P. Fang, Y. Bock, M. Cheng, and S. Miyazaki (2002), Anatomy of apparent seasonal variations from GPS-derived site position time series, *Journal of Geophysical Research: Solid Earth*, *107*(B4), ETG–9, doi:10.1029/2001JB000573.

Dong, X., B. Xi, A. Kennedy, Z. Feng, J. K. Entin, P. R. Houser, R. A. Schiffer, T. L'Ecuyer, W. S. Olson, K.-l. Hsu, et al. (2011), Investigation of the 2006 drought and 2007 flood extremes at the Southern Great Plains through an integrative analysis of observations, *Journal of Geophysical Research: Atmospheres*, *116*(D3), doi:10.1029/2010JD014776.

Dorigo, W., W. Wagner, C. Albergel, F. Albrecht, G. Balsamo, L. Brocca, D. Chung, M. Ertl, M. Forkel, A. Gruber, et al. (2017), ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions, *Remote Sensing of Environment*, *203*, 185–215, doi:10.1016/j.rse.2017.07.001.

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

Doucet, A., and A. M. Johansen (2009), A tutorial on particle filtering and smoothing: Fifteen years later, *Handbook of nonlinear filtering*, *12*(656-704), 3.

Dourte, D. R., C. W. Fraisse, and W.-L. Bartels (2015), Exploring changes in rainfall intensity and seasonal variability in the Southeastern US: Stakeholder engagement, observations, and adaptation, *Climate Risk Management*, *7*, 11–19, doi:10.1016/j.crm.2015.02.001.

Duan, Q., N. K. Ajami, X. Gao, and S. Sorooshian (2007), Multi-model ensemble hydrologic prediction using Bayesian model averaging, *Advances in Water Resources*, *30*(5), 1371–1386, doi:10.1016/j.advwatres.2006.11.014.

Ducharne, A., R. D. Koster, M. Suarez, M. Stieglitz, and P. Kumar (2000), A catchment-based approach to modeling land surface processes in a general circulation model. II- Parameter estimation and model demonstration, *Journal of Geophysical Research*, *105*, 24, doi:10.1029/2000JD900328.

Dzhamalov, R. G., G. N. Krichevets, and T. I. Safronova (2012), Current changes in water resources in Lena River basin, *Water Resources*, *39*(2), 147–160, doi:10.1134/S0097807812020042.

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

Encarnacao, J., P. Visser, A. Jaeggi, A. Bezdek, T. Mayer-Gürr, C. Shum, D. Arnold, E. Doornbos, M. Elmer, J. Guo, et al. (2019), Multi-approach gravity field models from swarm gps data, doi:10.13140/RG.2.2.34189.90084.

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research: Oceans*, *99*(C5), 10,143–10,162, doi:10.1029/94JC00572.

Evensen, G. (2003), The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics*, *53*(4), 343–367, doi:10.1007/s10236-003-0036-9.

Fagin, S. L. (1964), Recursive linear regression theory, optimal filter theory, and error analysis of optimal systems, in *IEEE international convention record*, vol. 12, pp. 216–245.

Famiglietti, J. S., M. Lo, S. L. Ho, J. Bethune, K. Anderson, T. H. Syed, S. C. Swenson, C. R. de Linage, and M. Rodell (2011), Satellites measure recent rates of groundwater depletion in California's Central Valley, *Geophysical Research Letters*, *38*(3), doi:10.1029/2010GL046442.

Fang, H., N. Tian, Y. Wang, M. Zhou, and M. A. Haile (2018), Nonlinear Bayesian estimation: from Kalman filtering to a broader horizon, *IEEE/CAA Journal of Automatica Sinica*, *5*(2), 401–417, doi:10.1109/JAS.2017.7510808.

Farrell, W. (1972), Deformation of the Earth by surface loads, *Reviews of Geophysics*, *10*(3), 761–797, doi:10.1029/RG010i003p00761.

Fasullo, J. T., C. Boening, F. W. Landerer, and R. S. Nerem (2013), Australia's unique influence on global sea level in 2010–2011, *Geophysical Research Letters*, *40*(16), 4368–4373, doi: 10.1002/grl.50834.

Fasullo, J. T., D. M. Lawrence, and S. C. Swenson (2016), Are GRACE-era terrestrial water trends driven by anthropogenic climate change?, *Advances in Meteorology*, *2016*, doi: 10.1155/2016/4830603.

Faunt, C. C., M. Sneed, J. Traum, and J. T. Brandt (2016), Water availability and land subsidence in the Central Valley, California, USA, *Hydrogeology Journal*, *24*(3), 675–684, doi:10.1007/s10040-015-1339-x.

Fay, M. P., and M. A. Proschan (2010), Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules, *Statistics surveys*, *4*, 1, doi: 10.1214/09-SS051.

Feng, W., L. Jean-Michel, M. Zhong, and T.-T. Hsu (2012), Terrestrial water storage changes in the Amazon basin measured by GRACE during 2002-2010, *Diqiu Wuli Xuebao*, *55*(3), 814–821, doi:10.6038/j.issn.0001-5733.2012.03.011.

Feng, W., M. Zhong, J. M. Lemoine, R. Biancale, H.-T. Hsu, and J. Xia (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, doi:10.1002/wrcr.20192.

Ferreira, A. J. D., C. Coelho, R. P. D. Walsh, R. Shakesby, A. Ceballos, and S. Doerr (2000), Hydrological implications of soil water-repellency in Eucalyptus globulus forests, north-central Portugal, *Journal of Hydrology*, *231*, 165–177, doi:10.1016/S0022-1694(00)00192-X.

Ferreira, V. G., H. D. C. Montecino, C. I. Yakubu, and B. Heck (2016), Uncertainties of the Gravity Recovery and Climate Experiment time-variable gravity-field solutions based on three-cornered hat method, *Journal of Applied Remote Sensing*, *10*(1), 1 – 20, doi:10.1117/1.JRS.10.015015.

Ferreira, V. G., B. Yong, K. Seitz, B. Heck, and T. Grombein (2020), Introducing an Improved GRACE Global Point-Mass Solution?A Case Study in Antarctica, *Remote Sensing*, *12*(19), 3197, doi:10.3390/rs12193197.

Flechtner, F., H. Dobslaw, and E. Fagiolini (2015), AOD1B Product Description Document for Product Release 05 (Rev. 4.3, April 23, 2015).

Flechtner, F., K.-H. Neumayer, C. Dahle, H. Dobslaw, E. Fagiolini, J.-C. Raimondo, and A. Güntner (2016), What can be expected from the GRACE-FO laser ranging interferometer for Earth science applications?, *Surveys in Geophysics*, *33*(2), 453?470, doi:10.1007/s10712-015-9338-y.

Forootan, E. (2014), Statistical signal decomposition techniques for analyzing time-variable satellite gravimetry data, Ph.D. thesis.

Forootan, E., and J. Kusche (2012), Separation of global time-variable gravity signals into maximally independent components, *Journal of Geodesy*, *86*(7), 477–497, doi:0.1007/s00190-011-0532-5.

Forootan, E., and J. Kusche (2013), Separation of deterministic signals using independent component analysis (ICA), *Studia Geophysica et Geodaetica*, *57*(1), 17–26, doi:10.1007/s11200-012-0718-1.

Forootan, E., J. L. Awange, J. Kusche, B. Heck, and A. Eicker (2012), Independent patterns of water mass anomalies over Australia from satellite data and models, *Remote Sensing of Environment*, *124*, 427–443, doi:10.1016/j.rse.2012.05.023.

Forootan, E., O. Didova, J. Kusche, and A. Löcher (2013), Comparisons of atmospheric data and reduction methods for the analysis of satellite gravimetry observations, *Journal of Geophysical Research: Solid Earth*, *118*(5), 2382–2396, doi:10.1002/jgrb.50160.

Forootan, E., O. Didova, M. Schumacher, J. Kusche, and B. Elsaka (2014a), Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields, over 2003–2011, *Journal of Geodesy*, *88*(5), 503–514, doi:10.1007/s00190-014-0696-x.

Forootan, E., J. Kusche, I. Loth, W.-D. Schuh, A. Eicker, J. L. Awange, L. Longuevergne, B. Diekkrüger, M. Schmidt, and C. Shum (2014b), Multivariate prediction of total water storage changes over West Africa from multi-satellite data, *Surveys in Geophysics*, *35*(4), 913–940, doi:10.1007/s10712-014-9292-0.

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

Forootan, E., J. L. Awange, M. Schumacher, R. O. Anyah, A. I. J. M. van Dijk, J. Kusche, et al. (2016), Quantifying the impacts of ENSO and IOD on rain gauge and remotely sensed precipitation products over Australia, *Remote sensing of Environment*, *172*, 50–66, doi:10.1016/j.rse.2015.10.027.

Forootan, E., A. Safari, A. Mostafaie, M. Schumacher, M. Delavar, and J. L. Awange (2017), Large-scale total water storage and water flux changes over the arid and semiarid parts of the Middle East from GRACE and reanalysis products, *Surveys in Geophysics*, *38*(3), 591–615, doi: 10.1007/s10712-016-9403-1.

Forootan, E., J. Kusche, M. Talpe, C. K. Shum, and M. Schmidt (2018), Developing a Complex Independent Component Analysis (CICA) Technique to Extract Non-stationary Patterns from Geophysical Time Series, *Surveys in Geophysics*, *39*(3), 435–465, doi:10.1007/s10712-017-9451-1.

Forootan, E., M. Khaki, M. Schumacher, V. Wulfmeyer, N. Mehrnegar, A. Van Dijk, L. Brocca, S. Farzaneh, F. Akinluyi, G. Ramillien, et al. (2019), Understanding the global hydrological droughts of 2003–2016 and their relationships with teleconnections, *Science of the Total Environment*, *650*, 2587–2604, doi:10.1016/j.scitotenv.2018.09.231.

Forootan, E., M. Schumacher, N. Mehrnegar, A. Bezděk, M. J. Talpe, S. Farzaneh, C. Zhang, Y. Zhang, and C. Shum (2020), An iterative ICA-based reconstruction method to produce consistent time-variable total water storage fields using GRACE and Swarm satellite data, *Remote Sensing*, *12*(10), 1639, doi:10.3390/rs12101639.

Freeze, R. A., and J. A. Cherry (1979), Groundwater, *Tech. rep.*

Fu, Y., and J. T. Freymueller (2012), Seasonal and long-term vertical deformation in the Nepal Himalaya constrained by GPS and GRACE measurements, *Journal of Geophysical Research: Solid Earth*, *117*(B3), doi:10.1029/2011JB008925.

Gagniuc, P. A. (2017), *Markov chains: from theory to implementation and experimentation*, John Wiley & Sons.

Gelfand, A. E., and D. K. Dey (1994), Bayesian model choice: asymptotics and exact calculations, *Journal of the Royal Statistical Society: Series B (Methodological)*, *56*(3), 501–514, doi: 10.1111/j.2517-6161.1994.tb01996.x.

Gelfand, A. E., and A. F. Smith (1990), Sampling-based approaches to calculating marginal densities, *Journal of the American statistical association*, *85*(410), 398–409.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013), *Bayesian data analysis*, CRC press.

George, E. I., and R. E. McCulloch (1997), Approaches for Bayesian variable selection, *Statistica sinica*, pp. 339–373.

Geyer, C. J. (1991), Markov chain Monte Carlo maximum likelihood.

Geyer, C. J. (1992), Practical Markov Chain Monte Carlo, *Statistical science*, pp. 473–483.

Geyer, C. J. (2011), Introduction to Markov Chain Monte Carlo, *Handbook of Markov Chain Monte Carlo*, p. 46.

Ghahramani, Z. (1997), Learning dynamic Bayesian networks, in *International School on Neural Networks, Initiated by IIASS and EMFCSC*, pp. 168–197, Springer, doi:10.1007/BFb0053999.

Gilks, W. R. (1996), Introducing Markov Chain Monte Carlo, *Markov chain Monte Carlo in practice*.

Girard, A., C. Rasmussen, J. Q. Candela, and R. Murray-Smith (2002), Gaussian process priors with uncertain inputs application to multiple-step ahead time series forecasting, *Advances in neural information processing systems*, *15*, 545–552.

Girotto, M., G. J. De Lannoy, R. H. Reichle, and M. Rodell (2016), Assimilation of gridded terrestrial water storage observations from GRACE into a land surface model, *Water Resources Research*, *52*(5), 4164–4183, doi:10.1002/2015WR018417.

Girotto, M., G. J. De Lannoy, R. H. Reichle, M. Rodell, C. Draper, S. N. Bhanja, and A. Mukherjee (2017), Benefits and pitfalls of GRACE data assimilation: A case study of terrestrial water storage depletion in India, *Geophysical research letters*, *44*(9), 4107–4115, doi:10.1002/2017GL072994.

Golub, G. H., and C. F. Van Loan (2012), *Matrix computations*, vol. 3, JHU Press.

Gordon, N. J., D. J. Salmond, and A. F. Smith (1993), Novel approach to nonlinear/non-Gaussian Bayesian state estimation, in *IEE proceedings F (radar and signal processing)*, vol. 140, pp. 107–113, IET, doi:10.1049/ip-f-2.1993.0015.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*(4), 711–732, doi:10.1093/biomet/82.4.711.

Gregorcic, G., and G. Lightbody (2002), Gaussian processes for modelling of dynamic non-linear systems, in *Proceedings of the Irish Signals and Systems Conference*, pp. 141–147.

Griffin, D., and K. J. Anchukaitis (2014), How unusual is the 2012–2014 California drought?, *Geophysical Research Letters*, *41*(24), 9017–9023, doi:10.1002/2014GL062433.

Gruber, A., T. Scanlon, R. van der Schalie, W. Wagner, and W. Dorigo (2019), Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, *Earth System Science Data*, pp. 1–37, doi:10.5194/essd-11-717-2019.

Guo, J., Z. Huang, C. Shum, and W. Van der Wal (2012), Comparisons among contemporary glacial isostatic adjustment models, *Journal of Geodynamics*, *61*, 129–137, doi:10.1016/j.jog.2012.03.011.

Guo, J., W. Li, X. Chang, G. Zhu, X. Liu, and B. Guo (2018), Terrestrial water storage changes over xinjiang extracted by combining gaussian filter and multichannel singular spectrum analysis from grace, *Geophysical Journal International*, *213*(1), 397–407, doi:10.1093/gji/ggy006.

Gutentag, E. D., F. J. Heimes, and N. C. Krothe (2014), *Geohydrology of the High Plains aquifer in parts of Colorado, Kansas, Nebraska, New Mexico, Oklahoma, South Dakota, Texas, and Wyoming*, vol. 1400-B, 63p, doi:10.3133/pp1400B.

Gutiérrez-Pena, E., and A. Smith (1995), Conjugate parameterizations for natural exponential families, *Journal of the American Statistical Association*, *90*(432), 1347–1356, doi: 10.1080/01621459.1995.10476639.

Haario, H., E. Saksman, and J. Tamminen (1999), Adaptive proposal distribution for random walk Metropolis algorithm, *Computational Statistics*, *14*(3), 375–396, doi:10.1007/s00180005002.

Han, S. C., C. Jekeli, and C. Shum (2004), Time-variable aliasing effects of ocean tides, atmosphere, and continental water mass on monthly mean GRACE gravity field, *Journal of Geophysical Research: Solid Earth*, *109*(B4), doi:10.1029/2003JB002501.

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

Hannan, E. J., A. McDougall, and D. Poskitt (1989), Recursive estimation of autoregressions, *Journal of the Royal Statistical Society. Series B (Methodological)*, *51*(2), 217–233, doi:10.1111/j.2517-6161.1989.tb01759.x.

Hao, M., J. T. Freymueller, Q. Wang, D. Cui, and S. Qin (2016), Vertical crustal movement around the southeastern Tibetan Plateau constrained by GPS and GRACE data, *Earth and Planetary Science Letters*, *437*, 1–8, doi:10.1016/j.epsl.2015.12.038.

Hasan, E., and A. Tarhule (2020), Grace: Gravity recovery and climate experiment long-term trend investigation over the nile river basin: Spatial variability drivers, *Journal of Hydrology*, *586*, 124,870, doi:10.1016/j.jhydrol.2020.124870.

Heiskanen, W. A., and H. Moritz (1967), Physical geodesy, *Bulletin Géodésique (1946-1975)*, *86*(1), 491–492.

Hillel, D. (1998), *Environmental soil physics: Fundamentals, applications, and environmental considerations*, Elsevier.

Hoeting, J. A., D. Madigan, A. E. Raftery, and C. T. Volinsky (1999), Bayesian Model Averaging: A Tutorial, *Statistical Science*, *14*(4), 382–417.

Horel, J. D. (1984), Complex principal component analysis: Theory and examples, *Journal of Applied Meteorology and Climatology*, *23*(12), 1660–1673, doi:10.1175/1520-0450(1984)023<1660:CPCATA>2.0.CO;2.

Houborg, R., M. Rodell, B. Li, R. Reichle, and B. F. Zaitchik (2012), Drought indicators based on model-assimilated Gravity Recovery and Climate Experiment (GRACE) terrestrial water storage observations, *Water Resources Research*, *48*(7), doi:10.1029/2011WR011291.

Hsu, K.-l., H. Moradkhani, and S. Sorooshian (2009), A sequential Bayesian approach for hydrologic model selection and prediction, *Water Resources Research*, *45*(12), doi:10.1029/2008WR006824.

Huffman, G. J., and D. T. Bolvin (2015), Real-time TRMM multi-satellite precipitation analysis data set documentation, *NASA Tech. Doc*, *10*.

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

Hurkmans, R., P. A. Troch, R. Uijlenhoet, P. Torfs, and M. Durcik (2009), Effects of climate variability on water storage in the Colorado River basin, *Journal of hydrometeorology*, *10*(5), 1257–1270, doi:10.1175/2009JHM1133.1.

Illston, B. G., J. B. Basara, and K. C. Crawford (2004), Seasonal to interannual variations of soil moisture measured in Oklahoma, *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *24*(15), 1883–1896, doi:10.1002/joc.1077.

Ivins, E. R., M. M. Watkins, D.-N. Yuan, R. Dietrich, G. Casassa, and A. Rülke (2011), On-land ice loss and glacial isostatic adjustment at the drake passage: 2003–2009, *Journal of Geophysical Research: Solid Earth*, *116*(B2), doi:10.1029/2010JB007607.

Jäggi, A., C. Dahle, D. Arnold, H. Bock, U. Meyer, G. Beutler, and J. Van den IJssel (2016), Swarm kinematic orbits and gravity fields from 18 months of gps data, *Advances in Space Research*, *57*(1), 218–233, doi:10.1016/j.asr.2015.10.035.

Jaiswal, R., S. Ali, and B. Bharti (2020), Comparative evaluation of conceptual and physical rainfall–runoff models, *Applied Water Science*, *10*(1), 1–14, doi:10.1007/s13201-019-1122-6.

Jazwinski, A. H. (2007), *Stochastic processes and filtering theory*, Courier Corporation.

Jeffreys, H. (1998), *The theory of probability*, OUP Oxford.

Jekeli, C. (1981), Alternative methods to smooth the Earth's gravity field.

Jones, G. L., and J. P. Hobert (2001), Honest exploration of intractable probability distributions via Markov Chain Monte Carlo, *Statistical Science*, pp. 312–334.

Jones, G. L., J. P. Hobert, et al. (2004), Sufficient burn-in for Gibbs samplers for a hierarchical random effects model, *The Annals of Statistics*, *32*(2), 784–817.

Jones, J. (2010), *California's Drought of 2007-2009: An Overview*, State of California, Natural Resources Agency, California Department of Water resources.

Jordan, M. I., Z. Ghahramani, T. S. Jaakkola, and L. K. Saul (1999), An introduction to variational methods for graphical models, *Machine learning*, *37*(2), 183–233, doi:10.1023/A:1007665907178.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *82*(1), 35–45, doi:10.1115/1.3662552.

Kass, R. E., and R. Natarajan (2006), A default conjugate prior for variance components in generalized linear mixed models (comment on article by browne and draper), *Bayesian Analysis*, *1*(3), 535–542, doi:10.1214/06-BA117B.

Khaki, M., B. Ait-El-Fquih, I. Hoteit, E. Forootan, J.L., and M. Kuhn (2017a), A two-update ensemble kalman filter for land hydrological data assimilation with an uncertain constraint, *Journal of Hydrology*, *555*, 447–462, doi:10.1016/j.jhydrol.2017.10.032.

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

Khaki, M., M. Schumacher, E. Forootan, M. Kuhn, J. L. Awange, and A. I. J. M. van Dijk (2017c), Accounting for spatial correlation errors in the assimilation of GRACE into hydrological models through localization, *Advances in Water Resources*, *108*, 99–112, doi: 10.1016/j.advwatres.2017.07.024.

Khaki, M., B. Ait-El-Fquih, I. Hoteit, E. Forootan, J. L. Awange, and M. Kuhn (2018a), Unsupervised ensemble kalman filtering with an uncertain constraint for land hydrological data assimilation, *Journal of Hydrology*, *564*, 175–190, doi:10.1016/j.jhydrol.2018.06.080.

Khaki, M., J. L. Awange, E. Forootan, and M. Kuhn (2018b), Understanding the association between climate variability and the nile's water level fluctuations and water storage changes during 1992?2016, *Science of The Total Environment*, *645*, 1509–1521, doi: 10.1016/j.scitotenv.2018.07.212.

Khaki, M., E. Forootan, M. Kuhn, J. L. Awange, F. Papa, and C. K. Shum (2018c), A study of Bangladesh's sub-surface water storages using satellite products and data assimilation scheme, *Science of the Total Environment*, *625*, 963–977, doi:10.1016/j.scitotenv.2017.12.289.

Khaki, M., E. Forootan, M. Kuhn, J. L. Awange, A. I. J. M. van Dijk, M. Schumacher, and M. A. Sharifi (2018d), Determining water storage depletion within Iran by assimilating GRACE data into the W3RA hydrological model, *Advances in Water Resources*, *114*, 1–18, doi: 10.1016/j.advwatres.2018.02.008.

Khan, S. A., I. Sasgen, M. Bevis, T. van Dam, J. L. Bamber, J. Wahr, M. Willis, K. H. Kjær, B. Wouters, V. Helm, et al. (2016), Geodetic measurements reveal similarities between posttextendashlast glacial maximum and present-day mass loss from the greenland ice sheet, *Science advances*, *2*(9), doi: 10.1126/sciadv.1600931.

Khandu, K., E. Forootan, M. Schumacher, J. L. Awange, and H. Müller-Schmied (2016), Exploring the influence of precipitation extremes and human water use on total water storage (TWS) changes in the Ganges-Brahmaputra-Meghna River Basin, *Water Resources Research*, *52*(3), 2240–2258, doi:10.1002/2015WR018113.

Kitagawa, G. (1987), Non-gaussian state?space modeling of nonstationary time series, *Journal of the American statistical association*, *82*(400), 1032–1041, doi:10.1080/01621459.1987.10478534.

Klees, R., E. Zapreeva, H. Winsemius, and H. Savenije (2006), The bias in GRACE estimates of continental water storage variations, *Hydrology and Earth System Sciences Discussions*, *3*(6), 3557–3594.

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

Koch, K.-R. (2007), *Introduction to Bayesian statistics*, Springer Science & Business Media.

Kocijan, J., A. Girard, B. Banko, and R. Murray-Smith (2005), Dynamic systems identification with Gaussian processes, *Mathematical and Computer Modelling of Dynamical Systems*, *11*(4), 411–424, doi:10.1080/13873950500068567.

Koller, D., and N. Friedman (2009), *Probabilistic graphical models: principles and techniques*, MIT press.

Kousky, V. E., M. T. Kagano, and I. F. Cavalcanti (1984), A review of the Southern Oscillation: oceanic-atmospheric circulation changes and related rainfall anomalies, *Tellus A*, *36*(5), 490–504, doi:10.1111/j.1600-0870.1984.tb00264.x.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *Journal of Hydrology*, *211*(1-4), 69–85, doi:10.1016/S0022-1694(98)00198-X.

Kusche, J. (2007), Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models, *Journal of Geodesy*, *81*(11), 733–749, doi:10.1007/s00190-007-0143-3.

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

Kusche, J., A. Eicker, E. Forootan, A. Springer, and L. Longuevergne (2016), Mapping probabilities of extreme continental water storage changes from space gravimetry, *Geophysical Research Letters*, *43*(15), 8026–8034, doi:10.1002/2016GL069538.

Lambeck, K., and J. Chappell (2001), Sea level change through the last glacial cycle, *Science*, *299*(5517), 679–686.

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

Leng, G., M. Huang, N. Voisin, X. Zhang, G. R. Asrar, and L. R. Leung (2016), Emergence of new hydrologic regimes of surface water resources in the conterminous United States under future warming, *Environmental Research Letters*, *11*(11), 114,003, doi:10.1088/1748-9326/11/11/114003.

Lennart, L. (1999), System identification: theory for the user, *PTR Prentice Hall, Upper Saddle River, NJ*, pp. 1–14.

Levine, P. A., J. T. Randerson, S. C. Swenson, and D. M. Lawrence (2016), Evaluating the strength of the land − atmosphere moisture feedback in Earth system models using satellite observations, *Hydrology and Earth System Sciences (Online)*, *20*(12), doi:10.5194/hess-20-4837-2016.

Li, B., M. Rodell, S. Kumar, H. K. Beaudoing, A. Getirana, B. F. Zaitchik, L. G. de Goncalves, C. Cossetin, S. Bhanja, A. Mukherjee, et al. (2019), Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges, *Water Resources Research*, *55*(9), 7564–7586, doi:10.1029/2018WR024618.

Li, F., J. Kusche, R. Rietbroek, Z. Wang, E. Forootan, K. Schulze, and C. Lück (2020), Comparison of data-driven techniques to reconstruct (1992–2002) and predict (2017–2018) grace-like gridded total water storage changes using climate inputs, *Water Resources Research*, *56*(5), e2019WR026,551, doi:10.1029/2019WR026551.

Lindersson, S., L. Brandimarte, J. Mård, and G. Di Baldassarre (2020), A review of freely accessible global datasets for the study of floods, droughts and their interactions with human societies, *Wiley Interdisciplinary Reviews: Water*, *7*(3), e1424, doi:10.1002/wat2.1424.

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

Ljung, L. (1999), System identification, *Wiley encyclopedia of electrical and electronics engineering*, pp. 1–19, doi:10.1002/047134608X.W1046.pub2.

Long, D., B. R. Scanlon, L. Longuevergne, A. Y. Sun, D. N. Fernando, and H. Save (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, doi:10.1002/grl.50655.

Long, D., Y. Yang, Y. Wada, Y. Hong, W. Liang, Y. Chen, B. Yong, A. Hou, J. Wei, and L. Chen (2015), Deriving scaling factors using a global hydrological model to restore GRACE total water storage changes for China's Yangtze River Basin, *Remote Sensing of Environment*, *168*, 177–193, doi:10.1016/j.rse.2015.07.003.

Long, D., Y. Pan, J. Zhou, Y. Chen, X. Hou, Y. Hong, B. R. Scanlon, and L. Longuevergne (2017), Global analysis of spatiotemporal variability in merged total water storage changes using multiple GRACE products and global hydrological models, *Remote sensing of environment*, *192*, 198–216, doi:10.1016/j.rse.2017.02.011.

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

Lorenz, E. N. (1956), Empirical orthogonal functions and statistical weather prediction.

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

Luthcke, S. B., H. J. Zwally, W. Abdalati, D. Rowlands, R. Ray, R. Nerem, F. Lemoine, J. McCarthy, and D. Chinn (2006), Recent Greenland ice mass loss by drainage system from satellite gravity observations, *Science*, *314*(5803), 1286–1289, doi:10.1126/science.1130776.

Luthcke, S. B., T. Sabaka, B. Loomis, A. Arendt, J. McCarthy, and J. Camp (2013), Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution, *Journal of Glaciology*, *59*(216), 613–631, doi:10.3189/2013JoG12J147.

MacKay, D. J. (1997), Gaussian processes-a replacement for supervised neural networks?

Manuel, J. (2008), Drought in the southeast: lessons for water management, doi:10.1289/ehp.116-a168.

Marengo, J. A., C. A. Nobre, G. Sampaio, L. F. Salazar, and L. S. Borma (2011), Climate change in the Amazon Basin: Tipping points, changes in extremes, and impacts on natural and human systems, in *Tropical rainforest responses to climatic change*, pp. 259–283, Springer, doi:10.1007/978-3-642-05383-2_9.

Martín-Español, A., A. Zammit-Mangion, P. J. Clarke, T. Flament, V. Helm, M. A. King, S. B. Luthcke, E. Petrie, F. Rémy, N. Schön, et al. (2016), Spatial and temporal antarctic ice sheet mass trends, glacio-isostatic adjustment, and surface processes from a joint inversion of satellite altimeter, gravity, and gps data, *Journal of Geophysical Research: Earth Surface*, *121*(2), 182–200.

Martin-Espanol, M. A. K. A. Z. S. B. A. P. M., Alba, and J. L. Bamber (2016), An assessment of forward and inverse GIA solutions for Antarctica, *Journal of Geophysical Research*, *121*(9), 6947–6965, doi:10.1002/2016JB013154.

Maupin, M. A., J. F. Kenny, S. S. Hutson, J. K. Lovelace, N. L. Barber, K. S. Linsey, et al. (2017), Estimated use of water in the United States in 2010.

Mayer-Gürr, T., S. Behzadpour, A. Kvas, M. Ellmer, B. Klinger, S. Strasser, and N. Zehentner (2018), Itsg-grace2018: Monthly, daily and static gravity field solutions from grace, doi: 10.5880/ICGEM.2018.003.

McCarthy, T. S., A. Bloem, and P. A. Larkin (1998), Observations on the hydrology and geohydrology of the Okavango Delta, *South African Journal of Geology/Suid-Afrikaanse Tydskrif vir Geologie*, *101*(2), 101–117.

McLachlan, G. J., and T. Krishnan (2007), *The EM algorithm and extensions*, vol. 382, John Wiley & Sons.

Means III, E. G., N. West, and R. Patrick (2005), Population growth and climate change: Will pose tough challenges for water utilities, *Journal-American Water Works Association*, *97*(8), 40–46, doi:10.1002/j.1551-8833.2005.tb07447.x.

Mehrnegar, N., O. Jones, M. B. Singer, M. Schumacher, P. Bates, and E. Forootan (2020a), Comparing global hydrological models and combining them with GRACE by dynamic model data averaging (DMDA), *Advances in Water Resources*, *138*, 103,528, doi:10.1016/j.advwatres.2020.103528.

Mehrnegar, N., O. Jones, M. B. Singer, M. Schumacher, T. Jagdhuber, B. R. Scanlon, A. Rateb, and E. Forootan (2020b), Exploring groundwater and soil water storage changes across the CONUS at 12.5 km resolution by a Bayesian integration of GRACE data into W3RA, *Science of the Total Environment*, *758*, 143,579, doi:10.1016/j.scitotenv.2020.143579.

Mehta, V. K., V. R. Haden, B. A. Joyce, D. R. Purkey, and L. E. Jackson (2013), Irrigation demand and supply, given projections of climate and land-use change, in Yolo County, California, *Agricultural water management*, *117*, 70–82, doi:10.1016/j.agwat.2012.10.021.

Meijering, E. (2002), A chronology of interpolation: from ancient astronomy to modern signal and image processing, *Proceedings of the IEEE*, *90*(3), 319–342, doi:10.1109/5.993400.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *The journal of chemical physics*, *21*(6), 1087–1092, doi:10.1063/1.1699114.

Meyer, U., K. Sosnica, D. Arnold, C. Dahle, D. Thaller, R. Dach, and A. Jäggi (2019), Slr, grace and swarm gravity field determination and combination, *Remote sensing*, *11*(8), 956, doi: 10.3390/rs11080956.

Meyn, S. P., and R. L. Tweedie (2012), *Markov chains and stochastic stability*, Springer Science & Business Media.

Miller, J. D., H. D. Safford, M. Crimmins, and A. E. Thode (2009), Quantitative evidence for increasing forest fire severity in the Sierra Nevada and southern Cascade Mountains, California and Nevada, USA, *Ecosystems*, *12*(1), 16–32, doi:10.1007/s10021-008-9201-9.

Milne, G. A., J. X. Mitrovica, H.-G. Scherneck, J. L. Davis, J. M. Johansson, H. Koivula, and M. Vermeer (2004), Continuous GPS measurements of postglacial adjustment in Fennoscandia: 2. Modeling results, *Journal of Geophysical Research: Solid Earth*, *109*(B2), doi:10.1029/2003JB002619.

Minka, T. P. (2001), A family of algorithms for approximate Bayesian inference, Ph.D. thesis, Massachusetts Institute of Technology.

Mo, K. C., L. N. Long, Y. Xia, S. Yang, J. E. Schemm, and M. Ek (2011), Drought indices based on the Climate Forecast System Reanalysis and ensemble NLDAS, *Journal of Hydrometeorology*, *12*(2), 181–205, doi:10.1175/2010JHM1310.1.

Moritz, H. (1978), Least-squares collocation, *Reviews of geophysics*, *16*(3), 421–430.

Murray, I., R. Adams, and D. MacKay (2010), Elliptical slice sampling, in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 541–548.

Nakada, M., and K. Lambeck (1987), Glacial rebound and relative sea-level variations: a new appraisal, *Geophysical Journal International*, *90*(1), 171–224, doi:10.1111/j.1365-246X.1987.tb00680.x.

Nelder, J. A., and R. W. Wedderburn (1972), Generalized linear models, *Journal of the Royal Statistical Society: Series A (General)*, *135*(3), 370–384, doi:10.2307/2344614.

Nguyen, T. V., and E. V. Bonilla (2014), Automated variational inference for Gaussian process models, in *Advances in Neural Information Processing Systems*, pp. 1404–1412.

Ni, S., J. Chen, C. R. Wilson, J. Li, X. Hu, and R. Fu (2018), Global Terrestrial Water Storage Changes and Connections to ENSO Events, *Surveys in Geophysics*, *39*(1), 1–22, doi:10.1007/s10712-017-9421-7.

Niu, G.-Y., Z.-L. Yang, K. E. Mitchell, F. Chen, M. B. Ek, M. Barlage, A. Kumar, K. Manning, D. Niyogi, E. Rosero, et al. (2011), The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, *Journal of Geophysical Research: Atmospheres*, *116*(D12), doi:10.1029/2010JD015139.

Opper, M., and D. Saad (2001), *Advanced mean field methods: Theory and practice*, MIT press.

Palmer, W. C. (1965), *Meteorological drought*, vol. 30, US Department of Commerce, Weather Bureau.

Pan, Y., W.-B. Shen, C. Hwang, C. Liao, T. Zhang, and G. Zhang (2016), Seasonal mass changes and crustal vertical deformations constrained by GPS and GRACE in Northeastern Tibet, *Sensors*, *16*(8), 1211, doi:10.3390/s16081211.

Peltier, W. R. (1998), Postglacial variations in the level of the sea: Implications for climate dynamics and solid-earth geophysics, *Reviews of geophysics*, *36*(4), 603–689, doi:10.1002/wrcr.20291.

Peltier, W. R., D. Argus, and R. Drummond (2015), Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_C (VM5a) model, *Journal of Geophysical Research: Solid Earth*, *120*(1), 450–487, doi:10.1002/2014JB011176.

Peng, X., M. Small, Y. Zhao, and J. M. Moore (2019), Detecting and predicting tipping points, *International Journal of Bifurcation and Chaos*, *29*(08), 1930,022, doi:10.1142/S0218127419300210.

Pérez-Sánchez, J., J. Senent-Aparicio, F. Segura-Méndez, D. Pulido-Velazquez, and R. Srinivasan (2019), Evaluating hydrological models for deriving water resources in peninsular spain, *Sustainability*, *11*(10), 2872, doi:10.3390/su11102872.

Peterson, T. C., R. R. Heim Jr, R. Hirsch, D. P. Kaiser, H. Brooks, N. S. Diffenbaugh, R. M. Dole, J. P. Giovannettone, K. Guirguis, T. R. Karl, et al. (2013), Monitoring and understanding changes in heat waves, cold waves, floods, and droughts in the United States: state of knowledge, *Bulletin of the American Meteorological Society*, *94*(6), 821–834, doi:10.1175/BAMS-D-12-00066.1.

Phillips, T., R. S. Nerem, B. Fox-Kemper, J. S. Famiglietti, and B. Rajagopalan (2012), The influence of ENSO on global terrestrial water storage using GRACE, *Geophysical Research Letters*, *39*(16), doi:10.1029/2012GL052495.

Plaza Guingla, D. A., R. De Keyser, G. J. De Lannoy, L. Giustarini, P. Matgen, and V. R. Pauwels (2013), Improving particle filters in rainfall-runoff models: Application of the resample-move step and the ensemble Gaussian particle filter, *Water Resources Research*, *49*(7), 4005–4021, doi:10.1002/wrcr.20291.

Pokhrel, Y., N. Hanasaki, S. Koirala, J. Cho, P. J.-F. Yeh, H. Kim, S. Kanae, and T. Oki (2012), Incorporating anthropogenic water regulation modules into a land surface model, *Journal of Hydrometeorology*, *13*(1), 255–269, doi:10.1175/JHM-D-11-013.1.

Potter, L. B., and N. Hoque (2014), Texas population projections, 2010 to 2050, *Office of the State Demographer*, *4*.

Prigent, C., F. Aires, W. B. Rossow, and A. Robock (2005), Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements, *Journal of Geophysical Research: Atmospheres*, *110*(D7), doi:10.1029/2004JD005087.

Primiceri, G. E. (2005), Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies*, *72*(3), 821–852, doi:10.1111/j.1467-937X.2005.00353.x.

Rabiner, L. R. (1989), A tutorial on hidden Markov models and selected applications in speech recognition, *Proceedings of the IEEE*, *77*(2), 257–286, doi:10.1109/5.18626.

Raftery, A. (1993), *Model selection and accounting for model uncertainty in linear regression models*, Citeseer.

Raftery, A. E. (1995), Bayesian model selection in social research, *Sociological methodology*, pp. 111–163.

Raftery, A. E., D. Madigan, and J. A. Hoeting (1997), Bayesian model averaging for linear regression models, *Journal of the American Statistical Association*, *92*(437), 179–191.

Raftery, A. E., M. Kárnỳ, and P. Ettler (2010), Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill, *Technometrics*, *52*(1), 52–66, doi:10.1198/TECH.2009.08104.

Ramillien, G., F. Frappart, S. Gratton, and X. Vasseur (2015), Sequential estimation of surface water mass changes from daily satellite gravimetry data, *Journal of Geodesy*, *89*(3), 259–282, doi:10.1007/s00190-014-0772-2.

Ramsay, B. H. (1998), The interactive multisensor snow and ice mapping system, *Hydrological Processes*, *12*(10-11), 1537–1546, doi:10.1002/(SICI)1099-1085(199808/09)12:10/11<1537::AID-HYP679>3.0.CO;2-A.

Rangelova, E., W. Van der Wal, M. Sideris, and P. Wu (2010), Spatiotemporal analysis of the grace-derived mass variations in north america by means of multi-channel singular spectrum analysis, in *Gravity, geoid and earth observation*, pp. 539–546, Springer, doi:10.1007/978-3-642-10634-7_72.

Rasmussen, C. E. (2003), Gaussian processes in machine learning, in *Summer School on Machine Learning*, pp. 63–71, Springer, doi:10.1007/978-3-540-28650-9_4.

Reager, J. T., B. F. Thomas, and J. S. Famiglietti (2014), River basin flood potential inferred using GRACE gravity observations at several months lead time, *Nature Geoscience*, *7*(8), 588, doi:10.1038/ngeo2203.

Renken, R. A., J. Dixon, J. Koehmstedt, S. Ishman, A. Lietz, R. L. Marella, P. Telis, J. Rogers, and S. Memberg (2005), Impact of anthropogenic development on coastal ground-water hydrology in southeastern Florida, 1900–2000, *US Geol Surv Circ*, *1275*, 77.

Renzullo, L. J., A. Van Dijk, J.-M. Perraud, D. Collins, B. Henderson, H. Jin, A. Smith, and D. L. McJannet (2014), Continental satellite soil moisture data assimilation improves root-zone moisture analysis for water resources assessment, *Journal of Hydrology*, *519*, 2747–2762, doi:10.1016/j.jhydrol.2014.08.008.

Rietbroek, R. (2014), Retrieval of Sea Level and Surface Loading Variations from Geodetic Observations and Model Simulations.

Rietbroek, R., M. Fritsche, S. E. Brunnabend, I. Daras, J. Kusche, J. Schröter, F. Flechtner, and R. Dietrich (2012), Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data, *Journal of Geodynamics*, *59*, 64–71, doi:10.1016/j.jog.2011.02.003.

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

Rippey, B. R. (2015), The US drought of 2012, *Weather and Climate Extremes*, *10*, 57–64, doi:10.1016/j.wace.2015.10.004.

Robert, C., and G. Casella (2013), *Monte Carlo statistical methods*, Springer Science & Business Media.

Roberts, G. O., and S. K. Sahu (1997), Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*(2), 291–317, doi:10.1111/1467-9868.00070.

Roberts, H. V. (1965), Probabilistic prediction, *Journal of the American Statistical Association*, *60*(309), 50–62, doi:10.1080/01621459.1965.10480774.

Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C. J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, et al. (2004), The global land data assimilation system, *Bulletin of the American Meteorological Society*, *85*(3), 381–394, doi:10.1175/BAMS-85-3-381.

Rodell, M., J. Chen, H. Kato, J. S. Famiglietti, J. Nigro, and C. R. Wilson (2007), Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, *Hydrogeology Journal*, *15*(1), 159–166, doi:10.1007/s10040-006-0103-7.

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

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

Ropelewski, C. F., and M. S. Halpert (1987), Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation, *Monthly weather review*, *115*(8), 1606–1626, doi:10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2.

Rosenthal, J. S. (1995), Minorization conditions and convergence rates for Markov chain Monte Carlo, *Journal of the American Statistical Association*, *90*(430), 558–566.

Rowlands, D., S. Luthcke, J. McCarthy, S. Klosko, D. Chinn, F. Lemoine, J.-P. Boy, and T. Sabaka (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), doi:10.1029/2009JB006546.

Sabaka, T., D. Rowlands, S. Luthcke, and J.-P. Boy (2010), Improving global mass flux solutions from Gravity Recovery and Climate Experiment (GRACE) through forward modeling and continuous time correlation, *Journal of Geophysical Research: Solid Earth*, *115*(B11), doi: 10.1029/2010JB007533.

Santamaría-Gómez, A., and A. Mémin (2015), Geodetic secular velocity errors due to interannual surface loading deformation, *Geophysical Journal International*, *202*(2), 763–767, doi: 10.1093/gji/ggv190.

Särkkä, S. (2013), *Bayesian filtering and smoothing*, vol. 3, Cambridge University Press, doi: 10.1201/b16018.

Sasgen, I., V. Klemann, and Z. Martinec (2012), Towards the inversion of GRACE gravity fields for present-day ice-mass changes and glacial-isostatic adjustment in North America and Greenland, *Journal of Geodynamics*, *59*, 49–63, doi:10.1016/j.jog.2012.03.004.

Save, H., S. Bettadpur, and B. D. Tapley (2016), High-resolution csr grace rl05 mascons, *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547–7569, doi:10.1002/2016JB013007.

Scanlon, B., Z. Zhang, A. Rateb, A. Sun, D. Wiese, H. Save, H. Beaudoing, M. Lo, H. Müller-Schmied, P. Döll, et al. (2019), Tracking seasonal fluctuations in land water storage using global models and GRACE satellites, *Geophysical Research Letters*, *46*(10), 5254–5264, doi: 10.1029/2018GL081836.

Scanlon, B. R., R. C. Reedy, J. B. Gates, and P. H. Gowda (2010), Impact of agroecosystems on groundwater resources in the Central High Plains, USA, *Agriculture, ecosystems & environment*, *139*(4), 700–713, doi:10.1016/j.agee.2010.10.017.

Scanlon, B. R., C. C. Faunt, L. Longuevergne, R. C. Reedy, W. M. Alley, V. L. McGuire, and P. B. McMahon (2012a), 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, doi:10.1073/pnas.1200311109.

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

Scanlon, B. R., Z. Zhang, H. Save, D. N. Wiese, F. W. Landerer, D. Long, L. Longuevergne, and J. Chen (2016), Global evaluation of new GRACE mascon products for hydrologic applications, *Water Resources Research*, *52*(12), 9412–9429, doi:10.1002/2016WR019494.

Scanlon, B. R., Z. Zhang, H. Save, A. Y. Sun, H. M. Schmied, L. P. van Beek, D. N. Wiese, Y. Wada, D. Long, R. C. Reedy, 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, doi:10.1073/pnas.1704665115.

Schellekens, J., E. Dutra, A. Martínez-de la Torre, G. Balsamo, A. van Dijk, F. S. Weiland, M. Minvielle, J.-C. Calvet, B. Decharme, S. Eisner, et al. (2017), A global water resources ensemble of hydrological models: the eartH2Observe Tier-1 dataset, *Earth System Science Data*, *9*(2), 389–413, doi:10.5194/essd-9-389-2017.

Schmeer, M., M. Schmidt, W. Bosch, and F. Seitz (2012), Separation of mass signals within GRACE monthly gravity field models by means of empirical orthogonal functions, *Journal of Geodynamics*, *59*, 124–132, doi:10.1016/j.jog.2012.03.001.

Schmidt, R., F. Flechtner, U. Meyer, K.-H. Neumayer, C. Dahle, R. König, and J. Kusche (2008), Hydrological signals observed by the GRACE satellites, *Surveys in Geophysics*, *29*(4-5), 319–334, doi:10.1007/s10712-008-9033-3.

Schneider, J. G. (1997), Exploiting model uncertainty estimates for safe dynamic control learning, in *Advances in neural information processing systems*, pp. 1047–1053.

Schrama, E. J., B. Wouters, and R. Rietbroek (2014), A mascon approach to assess ice sheet and glacier mass balances and their uncertainties from GRACE data, *Journal of Geophysical Research: Solid Earth*, *119*(7), 6048–6066, doi:10.1002/2013JB010923Citations:85SECTIONSPDFPDF.

Schubert, S. D., M. J. Suarez, P. J. Pegion, R. D. Koster, and J. T. Bacmeister (2004), Causes of long-term drought in the US Great Plains, *Journal of Climate*, *17*(3), 485–503, doi:10.1175/1520-0442(2004)017<0485:COLDIT>2.0.CO;2.

Schumacher, M. (2016), Methods for assimilating remotely-sensed water storage changes into hydrological models, Ph.D. thesis, Universitäts-und Landesbibliothek Bonn.

Schumacher, M., J. Kusche, and P. Döll (2016), A systematic impact assessment of GRACE error correlation on data assimilation in hydrological models, *Journal of Geodesy*, *90*(6), 537–559, doi:10.1007/s00190-016-0892-y.

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

Schumacher, M., M. King, J. Rougier, Z. Sha, S. A. Khan, and J. Bamber (2018b), A new global GPS data set for testing and improving modelled GIA uplift rates, *Geophysical Journal International*, *214*(3), 2164–2176, doi:10.1093/gji/ggy235.

Schuurman, N. K., R. P. Grasman, and E. L. Hamaker (2016), A comparison of inverse-wishart prior specifications for covariance matrices in multilevel autoregressive models, *Multivariate Behavioral Research*, *51*(2-3), 185–206, doi:10.1080/00273171.2015.1065398.

Seager, R., L. M. Goddard, J. A. Nakamura, N. L. Henderson, and D. E. Lee (2014), Dynamical Causes of the 2010/11 Texas_Northern Mexico Drought, *Journal of Hydrometeorology*, *15*(1), 39–68, doi:10.1175/JHM-D-13-024.1.

Seager, R., M. Hoerling, S. Schubert, H. Wang, B. Lyon, A. Kumar, J. Nakamura, and N. Henderson (2015), Causes of the 2011–14 California drought, *Journal of Climate*, *28*(18), 6997–7024, doi: 10.1175/JCLI-D-14-00860.1.

Sha, Z., J. Rougier, M. Schumacher, and J. Bamber (2019), Bayesian model–data synthesis with an application to global glacio-isostatic adjustment, *Environmetrics*, *30*(1), e2530, doi: 10.1002/env.2530.

Shum, C., J. Ries, and B. Tapley (1995), The accuracy and applications of satellite altimetry, *Geophysical Journal International*, *121*(2), 321–336, doi:10.1111/j.1365-246X.1995.tb05714.x.

Singer, M. B., C. I. Sargeant, H. Piégay, J. Riquier, R. J. Wilson, and C. M. Evans (2014), Floodplain ecohydrology: Climatic, anthropogenic, and local physical controls on partitioning of water sources to riparian trees, *Water resources research*, *50*(5), 4490–4513, doi:10.1002/2014WR015581.

Sinha, D., T. H. Syed, J. S. Famiglietti, J. T. Reager, and R. C. Thomas (2017), Characterizing drought in India using GRACE observations of terrestrial water storage deficit, *Journal of Hydrometeorology*, *18*(2), 381–396, doi:10.1175/JHM-D-16-0047.1.

Sinharay, S. (2003), Assessing convergence of the Markov chain Monte Carlo algorithms: A review, *ETS Research Report Series*, *2003*(1), i–52, doi:10.1002/j.2333-8504.2003.tb01899.x.

Slater, L. J., M. B. Singer, and J. W. Kirchner (2015), Hydrologic versus geomorphic drivers of trends in flood hazard, *Geophysical Research Letters*, *42*(2), 370–376, doi:10.1002/2014GL062482.

Smith, A. (2013), *Sequential Monte Carlo methods in practice*, Springer Science & Business Media.

Smith, A. F., and G. O. Roberts (1993), Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *Journal of the Royal Statistical Society: Series B (Methodological)*, *55*(1), 3–23.

Sneeuw, N. (1994), Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective, *Geophysical Journal International*, *118*(3), 707–716, doi: 10.1111/j.1365-246X.1994.tb03995.x.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson (2008), Obstacles to high-dimensional particle filtering, *Monthly Weather Review*, *136*(12), 4629–4640, doi:10.1175/2008MWR2529.1.

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

Sophocleous, M. (2002), Interactions between groundwater and surface water: the state of the science, *Hydrogeology journal*, *10*(1), 52–67, doi:10.1007/s10040-001-0170-8.

Sośnica, K., A. Jäggi, U. Meyer, D. Thaller, G. Beutler, D. Arnold, and R. Dach (2015), Time variable Earth's gravity field from SLR satellites, *Journal of geodesy*, *89*(10), 945–960, doi: 10.1007/s00190-015-0825-1.

Spada, G., V. R. Barletta, V. Klemann, R. Riva, Z. Martinec, P. Gasperini, B. Lund, D. Wolf, L. Vermeersen, and M. King (2011), A benchmark study for glacial isostatic adjustment codes, *Geophysical Journal International*, *185*(1), 106–132, doi:10.1111/j.1365-246X.2011.04952.x.

Steffen, H., and G. Kaufmann (2005), Glacial isostatic adjustment of Scandinavia and northwestern Europe and the radial viscosity structure of the Earth's mantle, *Geophysical Journal International*, *163*(2), 801–812, doi:10.1111/j.1365-246X.2005.02740.x.

Strassberg, G., B. R. Scanlon, and D. Chambers (2009), Evaluation of groundwater storage monitoring with the GRACE satellite: Case study of the High Plains aquifer, central United States, *Water Resources Research*, *45*(5), doi:10.1029/2008WR006892.

Sun, G. (2013), Impacts of Climate Change and Variability on Water Resources in the Southeast USA, *NCA Southeast Technical Report 204-234*, pp. 210–236, doi:10.5822/978-1-61091-509-0_10.

Sun, Y., R. Riva, and P. Ditmar (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, doi:10.1002/2016JB013073.

Swenson, S., and J. Wahr (2002), Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *Journal of Geophysical Research: Solid Earth*, *107*(B9), ETG–3, doi:10.1029/2001JB000576.

Swenson, S., and J. Wahr (2006), Post-processing removal of correlated errors in GRACE data, *Geophysical Research Letters*, *33*(8), doi:10.1029/2005gl025285.

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

Taie Semiromi, M., and M. Koch (2019), Reconstruction of groundwater levels to impute missing values using singular and multichannel spectrum analysis: application to the ardabil plain, iran, *Hydrological Sciences Journal*, *64*(14), 1711–1726, doi:10.1080/02626667.2019.1669793.

Takatsuka, Y., M. R. Niekus, J. Harrington, S. Feng, D. Watkins, A. Mirchi, H. Nguyen, and M. C. Sukop (2018), Value of irrigation water usage in South Florida agriculture, *Science of The Total Environment*, *626*, 486–496, doi:10.1016/j.scitotenv.2017.12.240.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins (2004a), GRACE measurements of mass variability in the Earth system, *Science*, *305*(5683), 503–505, doi: 10.1126/science.1099192.

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber (2004b), The gravity recovery and climate experiment: Mission overview and early results, *Geophysical Research Letters*, *31*(9), doi: 10.1029/2004GL019779.

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

Tian, S., P. Tregoning, L. J. Renzullo, A. I. J. M. van Dijk, J. P. Walker, V. R. Pauwels, and S. Allgeyer (2017), Improved water balance component estimates through joint assimilation of GRACE water storage and SMOS soil moisture retrievals, *Water Resources Research*, *53*(3), 1820–1840, doi: 10.1002/2016WR019641.

Tiwari, V., J. Wahr, and S. Swenson (2009), Dwindling groundwater resources in northern India, from satellite gravity observations, *Geophysical Research Letters*, *36*(18), doi:10.1029/2009GL039401.

Tokdar, S. T., and R. E. Kass (2010), Importance sampling: a review, *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(1), 54–60, doi:10.1002/wics.56.

Trenberth, K. E. (1990), Recent observed interdecadal climate changes in the Northern Hemisphere, *Bulletin of the American Meteorological Society*, *71*(7), 988–993, doi:10.1175/1520-0477(1990)071<0988:ROICCI>2.0.CO;2.

Trenberth, K. E., and D. P. Stepaniak (2001), Indices of el niño evolution, *Journal of climate*, *14*(8), 1697–1701, doi:10.1175/1520-0442(2001)014<1697:LIOENO>2.0.CO;2.

Van Beek, L. P. H., Y. Wada, and M. F. P. Bierkens (2011), Global monthly water stress: 1. Water balance and water availability, *Water Resources Research*, *47*(7), doi:10.1029/2010WR009791.

Van Der Knijff, J. M., J. Younis, and A. P. J. De Roo (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, doi:10.1080/13658810802549154.

Van Dijk, A., L. J. Renzullo, and M. Rodell (2011), Use of Gravity Recovery and Climate Experiment terrestrial water storage retrievals to evaluate model estimates by the Australian water resources assessment system, *Water Resources Research*, *47*(11), doi:10.1029/2011WR010714.

Van Dijk, A. I., J. Schellekens, M. Yebra, H. E. Beck, L. J. Renzullo, A. Weerts, and G. Donchyts (2018), Global 5 km resolution estimates of secondary evaporation including irrigation through satellite data assimilation, *Hydrology and Earth System Sciences*, *22*(9), 4959–4980, doi: 10.5194/hess-22-4959-2018.

Van Dijk, A. I. J. M. (2010), The Australian Water Resources Assessment System: Technical 901 Report 3, Landscape model (version 0.5) Technical Description, CSIRO, *Water for a Healthy Country National Research Flagship*.

Van Dijk, A. I. J. M., L. J. Renzullo, Y. Wada, and P. Tregoning (2014), A global water cycle reanalysis (2003–2012) merging satellite gravimetry and altimetry observations with a hydrological multi-model ensemble, *Hydrology and Earth System Sciences*, *18*(8), 2955–2973, doi:10.5194/hess-18-2955-2014.

Van Dyk, D. A., and X. Jiao (2015), Metropolis-Hastings within partially collapsed Gibbs samplers, *Journal of Computational and Graphical Statistics*, *24*(2), 301–327, doi: 10.1080/10618600.2014.930041.

Van Dyk, D. A., and T. Park (2008), Partially collapsed Gibbs samplers: Theory and methods, *Journal of the American Statistical Association*, *103*(482), 790–796, doi:10.1198/016214508000000409.

Van Ravenzwaaij, D., P. Cassey, and S. D. Brown (2018), A simple introduction to Markov Chain Monte–Carlo sampling, *Psychonomic bulletin & review*, *25*(1), 143–154, doi:10.3758/s13423-016-1015-8.

Velicogna, I., T. Sutterley, and M. Van Den Broeke (2014), Regional acceleration in ice mass loss from greenland and antarctica using grace time-variable gravity data, *Geophysical Research Letters*, *41*(22), 8130–8137, doi:10.1002/2014GL061052.

Von Storch, H., and F. W. Zwiers (2001), *Statistical analysis in climate research*, Cambridge university press.

Voss, K. A., J. S. Famiglietti, M. Lo, C. De Linage, M. Rodell, and S. C. Swenson (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, doi:10.1002/wrcr.20078.

Vrugt, J. A., C. J. ter Braak, C. G. Diks, and G. Schoups (2013), Hydrologic data assimilation using particle Markov chain Monte Carlo simulation: Theory, concepts and applications, *Advances in Water Resources*, *51*, 457–478, doi:10.1016/j.advwatres.2012.04.002.

Wada, Y., D. Wisser, and M. Bierkens (2014), Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, *Earth System Dynamics Discussions*, *5*(1), 15–40, doi:10.5194/esdd-4-355-2013.

Wahr, J., and S. Zhong (2012), Computations of the viscoelastic response of a 3-D compressible Earth to surface loading: an application to Glacial Isostatic Adjustment in Antarctica and Canada, *Geophysical Journal International*, *192*(2), 557–572, doi:10.1093/gji/ggs030.

Wahr, J., M. Molenaar, and F. Bryan (1998), Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *Journal of Geophysical Research: Solid Earth*, *103*(B12), 30,205–30,229, doi:10.1029/98JB02844.

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

Wang, H., P. Wu, and Z. Wang (2006), An approach for spherical harmonic analysis of non-smooth data, *Computers and geosciences*, *32*(10), 1654–1668, doi:10.1016/j.cageo.2006.03.004.

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

Weedon, G. P., G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo (2014), The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, *Water Resources Research*, *50*(9), 7505–7514, doi:10.1002/2014WR015638.

Weerts, A. H., and G. Y. El Serafy (2006), Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, *Water resources research*, *42*(9), doi:10.1029/2005WR004093.

Werth, S., A. Güntner, R. Schmidt, and J. Kusche (2009), Evaluation of GRACE filter tools from a hydrological perspective, *Geophysical Journal International*, *179*(3), 1499–1515, doi: 10.1111/j.1365-246X.2009.04355.x.

Winter, T. C., J. W. Harvey, O. L. Franke, and W. M. Alley (1998), *Ground water and surface water: a single resource*, vol. 1139, US geological Survey.

Xiao, R., X. He, Y. Zhang, V. G. Ferreira, and L. Chang (2015), Monitoring groundwater variations from satellite gravimetry and hydrological models: A comparison with in-situ measurements in the mid-atlantic region of the United States, *Remote Sensing*, *7*(1), 686–703, doi:10.3390/rs70100686.

Yang, F., J. Kusche, E. Forootan, and R. Rietbroek (2017), Passive-ocean radial basis function approach to improve temporal gravity recovery from GRACE observations, *Journal of Geophysical Research: Solid Earth*, *122*(8), 6875–6892, doi:10.1002/2016JB013633.

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

Zhang, Y.-K., and K. Schilling (2006), Increasing streamflow and baseflow in Mississippi River since the 1940 s: Effect of land use change, *Journal of Hydrology*, *324*(1-4), 412–422, doi: 10.1016/j.jhydrol.2005.09.033.

Zhang, Z., B. Chao, J. Chen, and C. Wilson (2015), Terrestrial water storage anomalies of Yangtze River Basin droughts observed by GRACE and connections with ENSO, *Global and Planetary Change*, *126*, 35–45, doi:10.1016/j.gloplacha.2015.01.002.

Zhao, M., I. Velicogna, and J. S. Kimball (2017a), Satellite observations of regional drought severity in the continental United States using GRACE-based terrestrial water storage changes, *Journal of Climate*, *30*(16), 6297–6308, doi:10.1175/JCLI-D-16-0458.1.

Zhao, M., I. Velicogna, and J. S. Kimball (2017b), 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, doi:10.1175/JHM-D-16-0182.1.

Zheng, H., D. Barta, and X. Zhang (2014), Lesson learned from adaptation response to Devils Lake flooding in North Dakota, USA, *Regional environmental change*, *14*(1), 185–194, doi: 10.1007/s10113-013-0474-y.

Zotov, L., and C. Shum (2010), Multichannel singular spectrum analysis of the gravity field data from grace satellites, in *AIP Conference Proceedings*, vol. 1206, pp. 473–479, American Institute of Physics, doi:10.1063/1.3292557.

Appendix A

A.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA, *Lorenz*, 1956) is a statistical decomposition technique that is widely used to extract dominant modes of multivariate data sets based on the eigenvalue decomposition of their auto-covariance matrix. The PCA approach is applied in this study follows the implementation in (*Forootan and Kusche*, 2012, chapter 3), which allows us to extract orthogonal spatial and temporal components that capture the dominant variability in GRACE TWS, hydrological model outputs, and the DMDA-derived separated water storage estimates. Therefore, PCA can be used as a tool to compare various aspects of the available data.

The PCA is applied on the available time-variable fields after removing their temporal mean. The $t \times s$ data matrix $\mathbf{Y}$ is used to store either TWS or individual water storage values, where t represents the number of time epochs and s stands for the number of grid points. In our study, $\mathbf{Y}$ contains 122 rows ($t = 122$ months) and 33 columns ($s = 33$ basins). The solution to the PCA is defined by expanding $\mathbf{Y}$ as:

$$\mathbf{Y} = \mathbf{P}\bar{\mathbf{E}}^T, \tag{A.1}$$

where $\bar{\mathbf{E}}$ contains the spatially orthogonal vectors (Empirical Othogonal Functions, i.e., EOFs), associated with temporally uncorrelated time series, known as Principal Components (PCs), and stored in the column of matrix $\mathbf{P}$. Thus, $\bar{\mathbf{E}}$ contains the unit-norm eigenvectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_m$ of the covariance matrix $\mathbf{C}_Y = E\{\mathbf{Y}^T\mathbf{Y}\}$ in its columns ($\bar{\mathbf{E}}\bar{\mathbf{E}}^T = \mathbf{I}$), which are arranged with respect to the magnitude of the corresponding singular-values $\lambda_1, \lambda_2, ..., \lambda_m$, in which $\lambda_1 > \lambda_2 > ... > \lambda_m$, and m is the maximum number of singular-values corresponding to the covariance matrix $\mathbf{C}_Y$, i.e., 33 in our implementation. Similar to equation (A.1), the data matrix $\mathbf{Y}$ can also be decomposed by the Singular

Value Decomposition (SVD) method as:

$$\mathbf{Y} = \bar{\mathbf{P}}\Lambda\bar{\mathbf{E}}^T, \tag{A.2}$$

where $\bar{\mathbf{P}}$ contains normalized PCs, i.e. $\bar{\mathbf{P}}\bar{\mathbf{P}}^T = \mathbf{I}$, and Λ is diagonal and holds the singular values λ ordered according to their magnitude. The solution to the PCA problem therefore requires the use of classic algebraic methods to find the singular-values and their corresponding singular vectors of $\mathbf{C}_Y$ (see the computational aspects in *Golub and Van Loan*, 2012). The total variance of the data matrix $\mathbf{Y}$ can be derived by summing the square of singular values, i.e., $\sum {\lambda_i}^2, i = 1, ...33$. The i'th PCA-derived orthogonal model is computed as

$$\mathbf{Y}_i = \bar{\mathbf{P}}_i\Lambda_i\bar{\mathbf{E}}_i^T, \tag{A.3}$$

which represents $100 \times {\lambda_i}^2/\sum {\lambda_i}^2, i = 1, ...33$ percent of the total variance in $\mathbf{Y}$.

A.2 Independent Component Analysis (ICA)

ICA is a higher (than second) order statistical decomposition technique, which incorporates statistical moments of available data samples in its computational procedure to derive empirical components that are statistically as independent as possible. The ICA technique that is applied here follows the formulation of *Forootan and Kusche* (2012, 2013), in which the ICA is a rotated extension of the Principal Component Analysis (PCA, *Forootan*, 2014, chapter 3). For this, an orthogonal rotation matrix $\mathbf{R}$ is computed in an optimization procedure, which can be used to rotate the PCA's orthogonal components make them as statistically as independent as possible. The ICA decomposition of the data matrix $\mathbf{X}$ can be formulated as:

$$\mathbf{X} = \mathbf{P}\mathbf{R}\mathbf{R}^T\bar{\mathbf{E}}^T, \tag{A.4}$$

where $\mathbf{PR}$ and $\mathbf{ER}$ are rotated components and always only one of them is independent. For the temporal ICA technique, columns of $\mathbf{PR}$, and for the Spatial ICA method, columns of $\mathbf{ER}$ are statistically as independent as possible (details of the Temporal and Spatial ICA techniques and the approach to estimate $\mathbf{R}$ can be found in *Forootan*, 2014, chapter 4).