DGK Deutsche Geodätische Kommission
der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 763

Ehsan Forootan

Statistical Signal Decomposition Techniques for Analyzing Time-Variable Satellite Gravimetry Data

München 2016

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5175-1

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 45, Bonn 2015

DGK Deutsche Geodätische Kommission
der Bayerischen Akademie der Wissenschaften

Reihe C Dissertationen Heft Nr. 763

Statistical Signal Decomposition Techniques for Analyzing Time-Variable Satellite Gravimetry Data

Inaugural-Dissertation zur
Erlangung des Grades
Doktor-Ingenieur (Dr.-Ing.)
der Hohen Landwirtschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms Universität
zu Bonn

vorgelegt am 19.09.2014 von

Ehsan Forootan

aus
Ghaemshahr, Iran

München 2016

Verlag der Bayerischen Akademie der Wissenschaften
in Kommission beim Verlag C. H. Beck

ISSN 0065-5325 ISBN 978-3-7696-5175-1

Diese Arbeit ist gleichzeitig veröffentlicht in:
Schriftenreihe des Instituts für Geodäsie und Geoinformation
der Rheinischen Friedrich-Wilhelms Universität Bonn
ISSN 1864-1113, Nr. 45, Bonn 2015

Adresse der Deutschen Geodätischen Kommission:

Deutsche Geodätische Kommission
Alfons-Goppel-Straße 11 ● D – 80 539 München
Telefon +49 – 89 – 23 031 1113 ● Telefax +49 – 89 – 23 031 - 1283 / - 1100
e-mail hornik@dgfi.badw.de ● http://www.dgk.badw.de

Diese Publikation ist als pdf-Dokument veröffentlicht im Internet unter den Adressen /
This volume is published in the internet

<http://dgk.badw.de> / <http://hss.ulb.uni-bonn.de/2014/3766/3766.htm>

Prüfungskommission

Referent: Dr.-Ing. Jürgen Kusche
Korreferenten: Prof. Dr. techn. Wolf-Dieter Schuh
Prof. Dr.-Ing. Michael Schmidt

Tag der mündlichen Prüfung: 09.09.2014

ISSN 0065-5325

ISBN 978-3-7696-5175-1

Acknowledgment

I am grateful to my promoter Prof. Dr.-Ing. Jürgen Kusche who offered me the chance to be involved in his APMG (Astronomical, Physical, and Mathematical Geodesy) research group. During almost five years of my study at Bonn University, I was fortunate to benefit from Jürgen's inestimable advises and opinions through my research. Our dialog and discussions were always very open, and his continuous remarks and valuable suggestions helped me to gain a broad range of skills and knowledge about signal processing. His deep insights about geodesy helped me to improve the results of this work and several joint publications.

My deep gratitude goes to Prof. Dr.-techn. Wolf-Dieter Schuh (University of Bonn) and Prof. Dr.-Ing. Michael Schmidt (Deutsches Geodätisches Forschungsinstitut, DGFI) for their evaluations of my thesis. Their constructive comments helped me to improve the quality of this dissertation.

I have been privileged to have the chance to know and benefit from several scientists. Especially, I would like to thank the help of Prof. Dr.-Ing. Joseph Awange (University of Curtin) whose contributions in the success of my research are undeniable. I would also like to thank Prof. Dr. C.K. Shum (Ohio State University) for being always interested in discussing my research topic, and giving me the chance to meet his group at Columbus, Ohio.

I would like to acknowledge the financial supports, awarded by the German Research Foundation (DFG, die Deutsche Forschungsgemeinschaft) under: the Bayesian Methods in Geodetic Earth System Research (BAYES-G) project.

I would like to thank Mrs. Catharina Schnier, our APMG group secretary, who always did her best to answer any questions and helped me out with a lot of paper works.

I will never forget my first travel to the beautiful city of Amsterdam, where I had the chance to meet Prof. Remi Lanne and Prof. Willem Bouten (University of Amsterdam) and worked with them for slightly more than half a year. Here, I would like to express my gratitude to Remi and Willem whose honest help changed my life.

During my graduate studies, I met many nice colleagues, and will remember all the fun times I shared with them, naming a few: Michael Kemp, Jamie McLaren (University of Amsterdam), Jan Martin Brockmann, Basem Elsaka, Ina Loth, Roelof Rietbroek, Lutz Roese-Koerner, Judith Schall, Maike Schumacher, Akbar Shabanloui, Bernd Uebbing (University of Bonn), and Olga Didova (University of Delft). My special thanks would go to Roelof for many invaluable technical insights and generous comments on different stages of my work. I would also like to thank Jan Martin who always helped me to fix my computer issues. I thank Olga for her great help in computation of atmospheric de-aliasing products, as well as Bernd for sharing his altimetry knowledge with me. I owe particular thanks to Maike for being a wonderful friend, her several proof readings of this dissertation, and providing the German summary of this thesis.

I would like to thank my two brothers Afshin and Omid for their kind supports and encouragements. I may never be able to fully appreciate the love and support of my parents Azar and Fereidoun. I learned from them to be honest and responsible for my duties. I hope they would accept this work being dedicated to them.

Ehsan Forootan

Bonn, Germany
July, 2014

Statistical Signal Decomposition Techniques for Analyzing Time-Variable Satellite Gravimetry Data

Abstract

The time-variable gravity fields from the Gravity Recovery and Climate Experiment (GRACE) satellite mission provide valuable information about total water storage variations on a global scale. This quantity is difficult to observe with in-situ measurements but important for understanding regional energy balance, as well as for agricultural, and water resource management. In order to utilize GRACE time-variable level 2 products for studying global mass transport, there are two major problems that users face: 1) the presence of correlated noise in the level 2 potential spherical harmonic coefficients that increases with harmonic degree and causes 'striping' in the spatial domain, and 2) the fact that different physical signals are overlaid and difficult to separate from each other. These problems are termed the 'signal-noise' separation problem and the 'signal-signal' separation problem.

In this thesis, statistical decomposition methods are investigated to perform signal-noise and signal-signal separation using the time series of total water storage changes derived from satellite gravimetry products. In particular, the focus lies on the mathematical foundation of the second order statistical decomposition approach such as the principal component analysis (PCA), and its ordinary extensions, as well as the higher order statistical decomposition method of independent component analysis (ICA). The mathematical relationships between second and higher order statistical signal decomposition techniques are discussed. Uncertainties introduced in the extracted patterns, e.g., due to the limited time span of observations in computing auto-covariance matrices and higher order moment tensors, are addressed. The ICA approach is extended to the Complex ICA technique, which allows extraction of patterns that vary in space and time. Simulations of GRACE-like total water storage time series are used to assess the performance of the introduced statistical approaches. The ICA approach is applied to reduce the spatial leakage over the Australian continent, and to partition total water storage changes into terrestrial and groundwater storage changes over the Middle East. A new statistical approach is also introduced to forecast total water storage changes over West Africa, where it exhibits strong atmosphere-land-ocean interactions.

Statistische Methoden der Signaltrennung zur Analyse zeitabhängiger Satellitengravimetriedaten

Zusammenfassung

Zeitabhängige Gravitationsfelder der Gravity Recovery And Climate Experiment (GRACE) Satellitenmission liefern wertvolle Informationen über Gesamtwasserspeicheränderungen auf der Erde mit globaler Abdeckung. Diese Größe ist einerseits schwierig mit Hilfe lokaler Messungen zu beobachten, andererseits aber wichtig für das Verständnis regionaler Energiebilanzen und ebenfalls für Agrarwirtschaft und Gewässerbewirtschaftung. Um die zeitabhängigen Level 2 Produkte der GRACE-Mission zur Untersuchung globaler Massentransporte zu verwenden, muss der Nutzer sich mit zwei wichtigen Problemen auseinandersetzen: 1) mit dem Vorhandensein korrelierten Rauschens in den sphärisch harmonischen Potentialkoeffizienten, das mit zunehmendem harmonischen Grad ansteigt und ein 'Streifenmuster' im Ortsbereich verursacht und 2) mit der Tatsache, dass sich verschiedene physikalische Signale überlagern und schwer voneinander zu trennen sind. Diese Problematiken werden als 'Signal-Rausch' Trennungsproblem und 'Signal-Signal' Trennungsproblem bezeichnet.

In der vorliegenden Arbeit werden statistische Methoden der Signaltrennung untersucht, um 'Signal-Rausch' und 'Signal-Signal' Trennung unter Verwendung von Zeitreihen terrestrischer Wasserspeicheränderungen, die aus Produkten der Satellitengravimetrie abgeleitet werden, durchzuführen. Der Fokus liegt besonders auf der mathematischen Grundlage statistischer Ansätze zweiter Ordnung zur Zerlegung des Signals, wie der Principle Component Analysis (PCA) und seiner gebräuchlichen Erweiterungen, sowie der statistischen Methode höherer Ordnung zur Signaltrennung, bekannt als Independent Component Analysis (ICA). Die mathematischen Zusammenhänge der statistischen Methoden zweiter und höherer Ordnung zur Signalzerlegung werden erörtert. Unsicherheiten, die z.B. durch die Verwendung begrenzter Beobachtungszeiträume bei der Berechnung der Auto-Kovarianzmatrizen und Momente/Tensoren höherer Ordnung entstehen, werden behandelt. Der ICA Ansatz wird weiterentwickelt zur Komplexen ICA Methode, die das Extrahieren räumlich und zeitlich veränderlicher Muster ermöglicht. Simulierte Zeitreihen des GRACE Gesamtwasserspeichers werden erzeugt, um das Ergebnis der vorgestellten statistischen Methoden zu bewerten. Die ICA Methode wird zur Reduktion des (räumlichen) Leakage Effektes über Australien und zur Trennung der Gesamtwasserspeicheränderungen in Änderungen des terrestrischen und des Grundwasserspeichers im Mittleren Osten verwendet. Außerdem wird ein neuer statistischer Ansatz eingeführt, der Gesamtwasserspeicheränderungen über Westafrika, einer Region, die starke Interaktionen der Atmosphäre, des Landes und des Ozeans aufweist, prognostiziert.

Contents

1. Introduction

1.1 Background

The Earth System is composed of the geosphere and biosphere, from which the geosphere consists of the lithosphere, hydrosphere, cryosphere, and atmosphere. The biosphere is a global sum of all ecosystems, integrating all living beings and their relationships. Most of the mass transport through the Earth System is associated with the water (in its various forms) that moves through the system. The part of water that exists on the surface of the Earth is called surface water, that of its depth is called groundwater, while water vapor is termed atmospheric water. The global water cycle, also known as the hydrological cycle, describes the continuous movement of water within the Earth and its atmosphere.

Water from the atmosphere reaches the surface of the Earth as precipitation. Once the water reaches the ground, it might be partly stored as water in lakes, wetlands or oceans, or as snow or ice. Water might also percolate through the soil to become groundwater. A part of collected water that remains on the surface of the Earth is called runoff, which flows into the surface water bodies. Evaporation from soil and surface water, or evaporation and transpiration from vegetation (evapotranspiration) return water back to the atmosphere. Water in the atmosphere is available in the form of precipitation as it cools and condenses, thereby completing the water cycle. Processed that are connected to this cycle is illustrated in 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.

The largest fluxes over most of the globe include moisture transport in and out of the atmosphere, i.e. precipitation and evapotranspiration. Runoff is also a flux, which can be considered as the major source of available water for human use. Water stored in the subsurface, especially as soil moisture, controls vegetation moisture stress, and influences evapotranspiration. Soil moisture content affects storm runoff production, too (NRC, 1999). Groundwater and its variations are important for human use. Its interactions with soil moisture and runoff, depending on hydrological and climatic conditions of the region, considerably influence the regional water cycle. Assessing the water fluxes and interactions between different compartments within the water cycle is important, not only for understanding the Earth's hydrological mechanisms, but also for framing environmental policies by decision-makers (e.g., NRC, 1999, Ilk et al., 2005).

The global water cycle is difficult to observe due to its large spatial scale, and the fact that it includes changes in location (atmosphere, land, and oceans) and various states of water (solid, liquid, and vapor). As a result, different observational instruments and methodologies have been developed and implemented to enhance its monitoring. Traditionally, most estimates of, e.g., precipitation, evapotranspiration, and runoff, have been gauge-based. The same is true for monitoring soil moisture and groundwater storage changes, which mainly rely on in-situ meteorological measurements and piezometric observations. In-situ water monitoring networks are very useful to explore the local hydrological characteristics of a region. For instance, measurements of water levels in a network of wells in combination with measurements of level fluctuations over surface water bodies can be used to calculate water flow (NRC, 1999). In-situ networks, however, are normally limited to local or regional areas, from which obtaining a global picture of water storage variations is extremely difficult (NRC, 2002).

To support water monitoring tasks, various satellite missions have been launched, each observing a particular part of the water cycle. Among all, since three decades, altimetry satellites (e.g.,

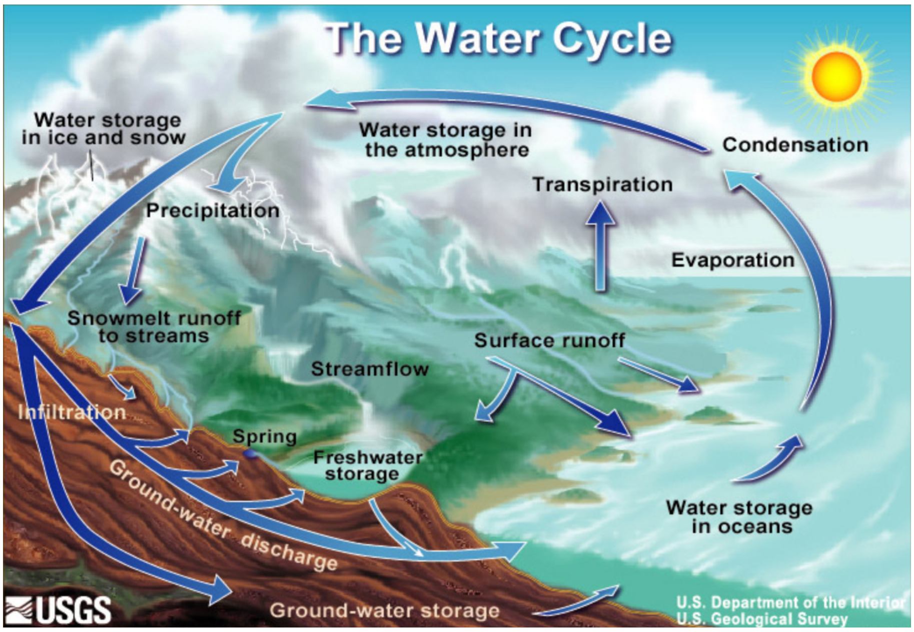


Figure 1.1: Illustration of the water cycle (source: the official website of the United States Geological Survey, USGS, http://water.usgs.gov/edu/watercycle.html).

Topex/Poseidon, Jason 1 and 2, and the European Space Agency (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. It has been more than one and a half decade that precipitation is remotely observed over the tropical regions by means of the Tropical Rainfall Measuring Mission (TRMM) (Huffman and Bolvin, 2012). Microwave remote sensing provides the capability to obtain observations of soil moisture at global and regional scales (Owe et al., 1999).

Since March 2002, with the launch of the Gravity Recovery And Climate Experiment (GRACE) mission, the global monitoring of time-variable (daily to monthly) gravity measurements, and subsequently global water storage changes has become possible (Tapley et al., 2004a,b, Kusche et al., 2012, Famiglietti and Rodell, 2013). GRACE measurements provide hydrologists the unique opportunity to close the terrestrial water budget by providing a quantitative estimate of total integrated water mass change over time, known as, total water storage (TWS) fields. TWS changes from GRACE in combination with complementary data, for instance, terrestrial water storage derived from Land Surface Models (LSM)s, and surface water estimates derived, e.g., from altimetry and tide gauge data sets can be used to estimate groundwater storage changes or other compartments of the water cycle (see examples in Kusche et al., 2012).

Besides the in-situ and satellite observations, hydrological modeling, including the development of conceptual and land surface models (e.g., Döll et al., 2003, Rodell et al., 2004), is another common approach for estimating the compartments of the water cycle. Nevertheless, models might experience limited skills in their water storage simulations, including effects on seasonal to long-term time-scales, due to various reasons. For instance, land surface models usually miss a part of the water balance, i.e. the groundwater compartment. Physical models might suffer from simplification of the physical processes introduced as mathematical equations to the model, or they might depend on boundary conditions, on data availability, and also on model calibration (e.g., Döll et al., 2003, Rodell et al., 2004, Güntner, 2008). Therefore, models and observations must be applied in a complementary way to observe the water cycle.

1.2 Motivation

In order to utilize GRACE observations for studying mass variations, there are two major problems that users of GRACE time-variable products face. These are:

1. The presence of correlated noise in the gravitational spherical harmonic coefficients that increases with harmonic degree and causes ‘striping’ in the spatial domain.
2. The fact that different physical signals are overlaid and difficult to separate from each other in the GRACE data.

These problems are termed the ‘signal-noise’ separation problem and the ‘signal-signal’ separation problem, and are addressed in the sections below.

1.2.1 Signal-Noise Separation Problem

The GRACE mission consists of two identical spacecrafts following each other in tandem in a single orbital plane. They are equipped with a K-band ranging system (KBR) to provide the inter-satellite range, as well as its derivatives with respect to time. In addition, both satellites carry GPS receivers, as well as accelerometers to remove non-gravitational forces prior to gravity modeling. Within the current processing of GRACE gravity field products, the observed inter-satellite ranges are linked to the time-variable gravity field. High-frequency mass redistribution is either removed, or taken into account within the parameter estimation process (Dobslaw et al., 2013). For instance, high-frequency non-tidal oceanic and atmospheric mass changes are reduced by using GRACE Atmosphere and Ocean De-aliasing level 1B (GRACE-AOD1B) products as ‘background’ models (Flechtner, 2007a,b). These reductions are necessary otherwise the high-frequency mass changes will alias into long wavelength signals, which deteriorate the accuracy of mass transport models (Flechtner, 2007a, Flechtner et al., 2010).

GRACE-derived time-variable gravity solutions, however, display anisotropic errors due to the geometry of the constellation in connection with limitations in the current gravity field analysis strategies (Sneeuw et al., 2004, Schrama et al., 2007, Elsaka, 2010) and temporal aliasing caused by incomplete reduction of short term mass variations by background models (Flechtner et al., 2010, Forootan et al., 2013, 2014a). Therefore, post-processing or filtering (signal-noise separation) is required to remove the correlated noise in the products (Jekeli, 1981, Swenson and Wahr, 2006, Kusche, 2007, Klees et al., 2008, Kusche et al., 2009). Most of the filtering approaches include an averaging operator in their process. As a result of smoothing (spatial sampling during the averaging process), mass anomalies are replaced and might mask each other. Such spatial leakage prevents GRACE data to be used to their full information extent, and subsequently, a leakage reduction approach must be used to counteract the attenuation of the filtering procedure and thereby deriving unbiased mass estimations (e.g., Fenoglio-Marc et al., 2006, Longuevergne et al., 2010).

1.2.2 Signal-Signal Separation Problem

GRACE-derived TWS changes contain information about vertically integrated mass redistribution, which is caused by different phenomena, including variability of land water storage and atmospheric moisture, non-steric sea level change, and ice melting. As a result,

> (i) GRACE observations represent a superposition of all mass change signals on land, and within the oceans and atmosphere. Separation of these integrated signals is only possible by introducing prior information on mass distribution in each compartment.
>
> (ii) GRACE-derived TWS changes contain non-linear and complex interactions (caused by different compartments) with many inherent time-scales.

Identifying the most dominant mass variation patterns in each compartment might help in understanding the role of natural and anthropogenic factors on the global water storage changes. Various methodologies have been proposed to estimate and evaluate mass variations, corresponding to different parts of the water cycle. In order to separate GRACE-derived mass signals, one can use physical models to account for a part of the observed mass signals and remove it from GRACE observations. The direct removal of mass variations from GRACE has been interpreted here as the 'forward or direct modeling' approach. This interpretation is due to the fact that one should rely on a priori information of mass variations derived from models while removing them from GRACE observations. An example is the reduction of high-frequency mass changes before computing the averaged time-variable gravity fields, discussed earlier. One should note that in the direct modeling approach, the modeling errors, with magnitudes greater than what the GRACE satellites are able to detect, would be reflected in the final GRACE-derived mass products (see Chapter 2 for more details). A similar problem also occurs when one uses models to reduce long-term variability from time-variable GRACE products. Guo et al. (2012), for instance, indicated that large uncertainties exist in the Glacial Isostatic Adjustment (GIA) models, which are usually used to account for the linear trend caused by viscoelastic response of the Earth to the loading of glaciation and deglaciation of ice age cycles. Removing imperfect GIA models from GRACE products might introduce errors in trend of mass estimations over the regions that exhibit strong post glacial displacements such as North America (Spada et al., 2011, Sasgen et al., 2012).

Different combinations of observations and model outputs have also been used to perform signal separation. Over the oceans, for example, monthly gravity field solutions were converted into estimates of the fluctuating ocean bottom pressure (OBP), which is the sum of atmospheric and oceanic mass variations. GRACE products were then validated against in-situ OBP observations (Kanzow et al., 2005, Rietbroek et al., 2006) or OBP derived from altimetry observations after removing the steric sea level contributions derived from salinity and temperature observations (Chambers, 2006). In contrast, Beek et al. (2012) used an improved regional hydrological model to correct estimates of seawater mass derived from GRACE and to account for freshwater inflow into surface water bodies of the Mediterranean and Black Seas. The accuracy of mass estimation was evaluated by comparing the results to an independent estimate derived from satellite altimetry with steric correction from regional oceanographic models.

For partitioning GRACE-derived TWS signals, over land, into soil moisture, groundwater, and surface water storage, most of the previous studies acquired their a priori information from, for example, altimetry observations to account for the surface water storage changes (Swenson and Wahr, 2007, Becker et al., 2010) and hydrological models for terrestrial water storage changes (Syed et al., 2005, Rodell et al., 2007, van Dijk et al., 2011, Awange et al., 2013). Subsequently, GRACE-TWS changes have been compared with (or reduced by) altimetry- and/or model-derived water storage changes. The accuracy of the estimation in such approaches might be limited since, for instance, altimetry observations contain relatively large errors over inland waters (Birkett, 1995, Kouraev et al., 2011) and hydrological models might show limited skills (Grippa et al., 2011, van Dijk, 2011).

Another approach to deal with the separation problem is to improve hydrological models through assimilation and/or calibration (Werth and Güntner, 2010, Houborg et al., 2012, Schumacher, 2012, van Dijk et al., 2013). These approaches represent their own strengths by allowing a quantitative understanding of interactions between different compartments of water storage. The computation load of these approach is, however, high.

In summary, separating GRACE signals into their physically meaningful sources (information on individual processes) addresses a challenging signal separation problem, which will be the focus of this contribution.

1.3 Application of Statistics in Signal Separation

Traditionally, in order to extract information from observations using parametric methods, it is assumed that each time series consists of at least three parts: a trend (defined as long-term evolution of the series), periodic components (typical seasonal cycles) and a random part, i.e. noise. Then, the sought-for parameters are approximated using, e.g., a least squares adjustment (Chatfield, 1989). The long-term variability derived from GRACE and geophysical observations, however, is perfectly linear in time. The periodical components of global mass changes cannot not be necessarily explained by sinusoids (Schmidt et al., 2008b). Therefore, one might use alternative statistical methods to extract data-adjusted components from observed signals, instead of a simple parametric decomposition of time series into a trend map and maps of certain periodic signals.

Several principles have been developed in statistics to extract linear and/or non-linear parameterizations of random variables. The algorithms that are used to find such parameterizations can be categorized according to the statistical information used in their extraction (decomposition) procedures. The methods that find the representations using only the information contained in the auto-covariance or auto-correlation matrices, built on the observations, are called ‘second order’ techniques. In other words, the second order methods assume that the statistical moments up to degree two adequately represent the probability distribution of observations (see details in Chapter 3). ‘Higher order’ statistical methods, however, go one step further than the second order approaches by incorporating higher than two statistical moments in their decomposition procedure. The latter is used when the probability distribution of the observations is non-Gaussian. In this case, more statistical moments are needed to represent the underlying distribution of the observations (see details in Chapter 4).

Eigenspace techniques, such as the Principal Component Analysis (PCA) also called Empirical Orthogonal Function (EOF) method, are among the most popular second order analysis techniques that are used to extract dominant patterns from time series of geophysical data (Preisendorfer, 1988). PCA was used, for instance, in Schrama et al. (2007) to identify the leading part of the surface mass signal from gridded GRACE-TWS changes and GPS height variations and, thus, to suppress what was identified as noise. In Chambers and Willis (2008), the PCA method was used as a tool for comparing the dominant part of GRACE-derived OBP changes with those OBP from ocean models. In such cases, a combination of leading PCA modes was interpreted as signal and those modes associated with low-amplitude variability were assumed to be noise (see also Wouters and Schrama, 2007).

PCA has also been used to identify geophysical patterns from observations. de Viron et al. (2006), for instance, correlated individual dominant modes of TWS changes from GRACE to the El Niño Southern Oscillation pattern (ENSO) derived from the Southern Oscillation Index (SOI). In a regional study, Rieser et al. (2010) used PCA to derive the main modes of spatial and temporal water storage variability over the Australian continent and compared each of them with rainfall anomalies. Schmeer et al. (2012) implemented an inversion approach in which the orthogonal components of atmospheric, oceanic and hydrological models, found by PCA, were used to separate the time series of GRACE-TWS changes into their land, ocean, and atmosphere compartments. Rietbroek (2014) also used PCA to extract sets of orthogonal base-functions from ocean and land water storage changes. The orthogonal components were then fitted, through a least squares inversion, to GRACE and altimetry observations, thereby time-variable surface loading changes were estimated.

In altimetric applications, several studies (e.g., Fenoglio-Marc, 2001) have implemented PCA to derive the trend and dominant patterns of sea level on both global and regional scales. Omondi et al. (2012, 2013b) used PCA to extract the decadal patterns of rainfall variations over East Africa and related them to the dominant patterns of sea surface temperature over the tropical oceans. In these applications, each of the PCA-derived component was treated individually or used as base-functions for further analysis.

Several studies, however, indicated that the objective of the PCA decomposition, i.e. to maximize the variance explained by each component while the components are orthogonal, leads to a clustering of different physical modes within a single extracted 'mathematical' mode and results in artificial features. This problem is called the 'mixing' problem in this thesis (see also e.g., Hyvärinen, 1999a, Aires et al., 2002, Hannachi et al., 2009, Forootan and Kusche, 2012). Examples on the typical mixing behavior of the PCA decomposition can be found in the application chapter of this thesis (Section 5.1). Therefore, depending on the problem, the ability of PCA to extract interpretable components may be limited (Jolliffe, 2003). To overcome some of these limitations, alternative approaches for PCA have been developed. For instance, rotated PCA/EOF techniques, simply called here Rotated EOF (REOF), that rotate PCA-derived components to obtain 'simpler' structures and thereby improving the interpretation of decomposition results (see, e.g., Richman, 1986).

PCA and its rotated extension REOF, however, use only simultaneous information between the existing time series and ignore the lagged information (lagged correlations) between them (Preisendorfer, 1988). Therefore, the assumption behind the PCA and REOF techniques is that the existing time series represent a 'stationary' process, i.e. mean and variance of the samples do not change over time. In fact, geophysical time series represent dynamical processes, which also include several cyclic phenomena. As a result, incorporating any lagged information, based on the methods such as the Extended Empirical Orthogonal Function (EEOF) (Weare and Nasstrom, 1982) and the Complex Empirical Orthogonal Function (CEOF) (Rasmusson et al., 1981), seems to be useful. The EEOF approach is also called Multi-Channel Singular Spectrum Analysis (MSSA) (Broomhead and King, 1986a,b). Rangelova et al. (2010) applied MSSA to GRACE-TWS changes in North America. Their investigation showed that using a lag-covariance matrix within the MSSA approach successfully extracts a long-term linear trend and annual variability of TWS changes. CEOF was implemented by García-García et al. (2011) to extract the spreading characteristics of the annual TWS changes over Australia. Applications of decomposition methods that works based lagged information (such as MSSA), however, require special cares, which will be discussed in this thesis.

Higher order statistical approaches, such as the Independent Component Analysis (ICA) (Cardoso and Souloumiac, 1993, Comon, 1994a, Hyvärinen, 1999a), can be considered as an alternative to the second order methods (e.g., PCA) to be used as an exploratory tool to decompose time series of geophysical observations. This is motivated from the fact that the probability density function of geophysical time series, such as GRACE-derived TWS changes, very often show non-Gaussian behavior (Aires et al., 2002, Forootan and Kusche, 2012). Therefore, more statistical information is included in the time series, which can be used for signal decomposition. Furthermore, studies, for example Richman (1986) and Jolliffe (2003), showed that PCA might represent limited skills in extracting physically meaningful patterns from geophysical observations. For these cases, extensions of the PCA approach, e.g., via incorporating higher order statistical moments in the decomposition procedure, might help in interpreting the decomposition (Aires et al., 2002, Hannachi et al., 2009). Examples of using ICA in order to decompose GRACE-TWS changes can be found e.g., in Frappart et al. (2011a,b), Forootan and Kusche (2012), and Boergens et al. (2014).

Statistical approaches can also be used to relate water storage variations to the indicators of climate variability and change such as precipitation, evapotranspiration, as well as large scale ocean-atmosphere interactions, such as ENSO and the Indian Ocean Dipole (IOD) phenomena (von Storch and Zwiers, 1999). For instance, Reager and Famiglietti (2013) presented an experimental forecasting approach that relates water storage changes to precipitation forcing and then generalize the relation based on large-scale basin characteristics. An alternative way to implement statistical TWS forecast is realized by applying decomposition methods, such as the PCA or ICA techniques, to extract the dominant uncorrelated (in case of PCA) or independent (in case of ICA) patterns from time series of TWS changes, rainfall, or sea surface temperature (SST). After constructing mathematical relationships between the dominant pattern of TWS changes, assumed as predictand, and its indicators, for example the dominant patterns of rainfall and SST, the relationship can be used to forecast or reconstruct TWS changes over a region of interest (see, e.g., Forootan et al., 2014b). The statistical forecasting approach, therefore, has the potential to be used for bridging the present GRACE data gaps, for instance those of 2004 where GRACE had an unfavorable orbit geometry, as well as a limited gap between satellite gravimetry missions such as GRACE and its GRACE follow-on mission. A similar concept has also been used, e.g., by the USA's National Oceanic and Atmospheric Administration (NOAA) for predicting climatic parameters (http://www.cpc.ncep.noaa.gov/).

1.4 Aims and Objectives of the Present Study

The application of statistics in extracting information from time-variable gravity field observations is methodologically rather complicated. This is due to the fact that all the available data sets, including those of, e.g., GRACE-TWS changes and the other complementary data that observe a part of the water cycle such as those from altimetry, are interrelated both in space and time. Although these spatio-temporal relations offer a great chance of reconstructing the state of the water cycle, for statistical analysis, it imposes a limitation since the sampled data sets are usually limited to the observation time span and subsequently a robust identification of the 'true' patterns from such dependent data sets is difficult. In addition, without having exact knowledge about the 'true' patterns, building scientific measures for evaluating the significance of the extracted information and testing them is rather complicated (see also von Storch and

Navarra, 1999). Therefore, this research is motivated by a main research question, which is:

'how can statistical signal separation methods be used to explore information from GRACE integrated observations?'

Important questions to be investigated in the context of *statistical signal separation methods* are:

1. Is the second order statistical information (e.g., from PCA) sufficient for decomposing spatio-temporal data sets such as GRACE time-variable products?
2. Several extension methods exist to enhance the interpretation of the PCA-derived components. Which strategy should be used to incorporate higher order statistical information in the decomposition procedure?
3. How are the second order and higher order statistical signal decomposition techniques related? How do they differ?
4. Estimation of the statistical independence, within the ICA approach, can be done within a variety of criteria. Which statistical criterion is suitable to be used for decomposing GRACE time-variable products?
5. Which statistical decomposition method is better suited to extract patterns that vary in space and time?
6. Is the application of statistical methods to decompose deterministic signals justified?

In order to perform signal-noise and signal-signal separation using the time series of satellite gravimetry products and complementary data, in this thesis, we will particularly address the mathematical foundation of the second and higher order statistical decomposition techniques. This includes the statistical assumptions behind each technique, as well as their applications in separating TWS changes.

In the existing literature, the second order statistical methods are usually formulated based on the PCA approach, which is usually performed using the Singular Value Decomposition (SVD) technique. Applying SVD to a matrix (containing random variables), one can factorize it into the left- and right- orthogonal singular vectors (Preisendorfer, 1988, von Storch and Zwiers, 1999). Formulations of the ordinary extensions of PCA techniques (such as REOF) also follow that of PCA (see, e.g., Jolliffe, 2003, Hannachi et al., 2009). The higher order statistical techniques are, however, discussed in the context of the 'Information Theory' or the 'Blind Source Separation (BSS)' techniques (Hyvärinen, 1999a, Stone, 2004). The BSS problem is, therefore, formulated by considering sampled random variables (observations) as a 'mixture' of some 'source' components. The desired sources are then sought-for by assuming a statistical property between them such as their mutual independence, thereby the method is called ICA (Cardoso and Souloumiac, 1993, Hyvärinen, 1999a). Another original contribution of this study is that, here, we add to the existing literature (e.g., Comon, 1994a, Aires et al., 2002, Hannachi et al., 2009, Forootan and Kusche, 2012) by describing the mathematical connections between the two families of second and higher order statistical decomposition methods. We will also argue that in which applications the second order approaches (such as PCA) may yield similar results to ICA, and in which specific situations the ICA-derived components are more representative than those based on PCA.

Forootan and Kusche (2012) applied ICA to identify independent patterns from time series of GRACE-TWS changes. The independence criterion in Forootan and Kusche (2012) was the diagonalization of the fourth-order cumulant tensor as in Cardoso and Souloumiac (1993). Forootan and Kusche (2013) also showed that such criterion is indeed adequate to separate any linear mixtures of a linear trend and sinusoidal cycles, assuming an infinite length for time series. Estimation of the statistical independence, however, can be done within a variety of criteria. In this study, we extend the discussions in Forootan and Kusche (2012) by relating the two common family of ICA criteria: including those based on the higher order statistical cumulants (e.g., Cardoso and Souloumiac, 1993, Moreau, 2001, Blaschke and Wiskott, 2004) and those based on entropy (Hyvärinen, 1999a,b, Hyvärinen and Oja, 2000).

For the first time, a complex extension of the ICA technique ('Complex ICA') is introduced in this thesis that can be used to extract non-stationary patterns (the patterns that change in both space and time) from geophysical time series. Identifying such patterns is important for many research topics such as climate and hydrological studies. This is due to the fact that the water cycle includes several dynamical processes that represent space- and time-variable signatures in the observations. To extract such patterns, the application of a Hilbert transformation (Rasmusson et al., 1981) is proposed to include information about the rate of change of the original data sets in the decomposition procedure. The ICA formulation based on the diagonalization of the fourth-order cumulant tensor (Forootan and Kusche, 2012, 2013) is then extended to a complex case. Finally, mathematical formulations to compute space- and time-variable components, derived from the Complex ICA method, are presented.

Application of statistical decomposition approaches requires an estimation of auto-covariance matrices and/or higher order cumulant tensors. Uncertainties exist in the estimation of these statistical properties, for example, due to the limited time span of observations (about 10 years of GRACE data). This has been addressed in this study by reviewing previous literature (e.g., Preisendorfer et al., 1981, Preisendorfer, 1988, chapter 5). A numerical Bootstrap approach (Efron, 1979) is presented to estimate the uncertainty of the decomposition results.

In order to assess the performance of the introduced statistical approaches, when they are applied to separate GRACE-TWS time series, various application examples are also investigated and discussed in Chapter 5. The applications of this thesis are formulated in such a way that the following objectives are covered:

1. To assess the performance of the statistical decomposition methods on realistically simulated time series, when the 'true' underlying patterns are known by definition. Here, after reviewing the simulated investigations in Forootan and Kusche (2012) and Forootan et al. (2012), we extend them by studying the performance of the introduced statistical decomposition techniques to identify the ENSO pattern in the time series of simulated global TWS changes.

2. To evaluate the functionality of each decomposition method in order to extract interpretable dominant patterns from GRACE time-variable observations. Therefore, the results in Forootan and Kusche (2012) are extended by applying ICA to the last release of GRACE products, covering the period of January 2003 to June 2013. Uncertainties of the extracted independent patterns, due to sampling errors and the imperfect reduction of atmospheric de-aliasing products (as in Forootan et al., 2014a), are evaluated. In this application, statistical decomposition methods are applied to identify the ENSO pattern in the time series of real GRACE-TWS changes.

3. To examine the applicability of the introduced statistical decomposition methods to reduce the spatial leakage, caused by limited spatial resolution of GRACE-TWS fields and the applied filtering methods. Similar to Forootan et al. (2012), the application of ICA to separate ocean- and land- water storage variations over a region including the Australian continent is discussed. Forootan et al. (2014c) recently adopted the inversion method introduced in Schmeer et al. (2012) and Rietbroek (2014) to reduce the leakage problem, and at the same time, to partition GRACE-TWS time series into land and surface water storage changes over a region including Iran. In this study, we extend this application by considering the latest release of GRACE products and studying a broader region that contains the Middle East region. Our results provide maps of groundwater and soil moisture water storage changes over the Middle East, while preserving the spatial resolution of GRACE products. This feature has been missed in the previous studies that address water storage variations over this region (see e.g., Voss et al., 2013).

4. To provide a statistical approach for forecasting and reconstructing TWS changes over a period when they are not observed. Similar to Forootan et al. (2014b), an application of ICA to reduce the high-dimensionality of a statistical TWS-forecast model is presented. The approach is then applied to forecast two years of TWS changes over West Africa.

1.5 Outline of the Thesis

The main goal of this study is to discuss how statistical signal separation methods can be used to extract useful information from GRACE satellite observations. As an application, the study is mainly focused on the hydrological signals and their relation to climate variability. To this end, a post-processing of the GRACE-derived time-variable gravity, to provide time series of TWS changes, is presented in Chapter 2. The results from Chapter 2 will be used in the applications of this thesis.

The mathematical foundations of the second and higher order decomposition approaches are respectively, discussed in the Chapters 3 and 4. In Chapter 3, a variety of second order statistical decomposition methods, including the Principal Component Analysis/Empirical Orthogonal Function (PCA/EOF) and its extensions Rotated EOF (REOF), Extended EOF (EEOF)/Multi-Channel Singular Spectrum Analysis (MSSA) and Complex EOF (CEOF) are discussed (see, e.g., Preisendorfer, 1988). The mathematical assumptions behind each decomposition approach and the ability of each to extract information from the available TWS time series are also demonstrated.

In Chapter 4, after reviewing the previous studies, the fundamental problem of source separation in the context of BSS is introduced in Section 4.1. The statistical concept of independence is defined in Section 4.2. In order to understand the similarities and differences of the second and higher order statistical approaches, the relationships between ICA and the second order methods of Chapter 3 are discussed in Section 4.3. In Section 4.4, the ICA filtering of GRACE-TWS products, (as in Frappart et al., 2011a,b), is discussed. In fact, estimation of the ICA can be done within variety of independent criteria. Therefore, in Section 4.5, some selected ICA criteria (based on higher order cumulants and entropy) and their mathematical relationships are discussed. Then an extension of the ICA approach, through the Complex ICA approach, is described to deal with the non-stationary behavior of geophysical time series. In Section 4.6, the application of ICA to separate a mixture of deterministic signals, including a linear trend and sinusoidal cycles, is mathematically investigated. Finally in Section 4.8, an approach to estimate the uncertainty of independent components is provided.

Chapter 5 contains five application sections. To complete the mathematical discussion of the Chapters 3 and 4, the performance of the commonly used decomposition methods (such as PCA and ICA), when they are applied to separate simulated GRACE-like TWS changes, is illustrated in Section 5.1. In Section 5.2, the dominant independent patterns of global GRACE-TWS changes are extracted and their errors due to the data noise, sampling errors, and imperfect atmospheric de-aliasing are investigated.

An application of the ICA approach, to reduce the leakage problem over the Australian continent and a large part of the Middle East, are respectively discussed in the Sections 5.3 and 5.4. The motivation for selecting the Australian continent and the Middle East arise from the challenging spatial leakage signal from surface water in and around these regions. The climatic conditions of the selected regions are, however, quite different. On the one hand, the Australian continent is strongly influenced by the large-scale ocean-atmosphere interactions such as ENSO and IOD (Risbey et al., 2009). On the other hand, the Middle East is located in an arid and semi-arid climatic area, and exhibits a decline of water storage (Voss et al., 2013). In Section 5.3, therefore, the ICA derived patterns of TWS changes over the Australian continent are also related to the ENSO and IOD climate indices. In Section 5.4, a statistical approach to optimally partition GRACE-TWS changes into soil moisture, surface, and groundwater storage changes is introduced. Such partitioning is necessary for the Middle East to understand the reasons behind the loss of TWS.

In Section 5.5, a new statistical, data-driven approach is described, which can be used to forecast TWS changes from GRACE, while using rainfall data, and sea surface temperature as indicators. The proposed method, therefore, capitalizes on the availability of remotely sensed observations for forecasting monthly TWS changes. West Africa is selected to test the application of the proposed method since it exhibits strong ocean-land-atmosphere interactions (Douville et al., 2007). Another motivation for selecting West Africa as a case study stems from the fact that the existing hydrological models in the region represent considerable uncertainties on the regional water balance due to missing data (Schuol and Abbaspour, 2006). A satellite-based forecasting method, therefore, might enhance large-scale TWS monitoring of the region. The major contributions of this thesis are summarized in Chapter 6, where remaining problems are reported and recommendations for further research are made.

2. GRACE Observations of Mass Changes

In this chapter, after introducing the concept of the Gravity Recovery And Climate Experiment (GRACE) mission, first, a proper estimation of total water storage (TWS) fields from GRACE time-variable gravity products is addressed (Section 2.1). In Section 2.2, the possible impacts of high-frequency mass redistribution on the final estimation of TWS from GRACE time-variable products is described. A summary of the processing steps, to estimate TWS changes from GRACE time-variable products, is presented in Section 2.3. Therefore, discussions on data preparation (Sections 2.1, 2.2, and 2.3) are necessary to understand the investigations that are performed in the application part (Chapter 5).

GRACE is a joint United States (National Aeronautics and Space Administration, NASA) and German (Deutsche Zentrum für Luft- und Raumfahrt, DLR) space mission, which provides estimates of variations in the gravity field arising from mass movements within the Earth system. GRACE was launched on 17th March 2002, and consists of two almost-identical spacecraft in the same orbit separated by roughly 220 km, having an initial altitude of $\sim$ 500 km, and an inclination of 89.5°. The altitude has been decreasing due to atmospheric drag and is $\sim$ 437 km at present. Initially, the mission was targeted to cover a 5-year period, which has long been exceeded. The GRACE Follow-On (GRACE-FO) mission, scheduled for launch in 2017, adopts the GRACE design of two satellites flying in one orbital plane but are also equipped with a modern (high precision) laser ranging system.

The GRACE satellites operate in what is termed as satellite-to-satellite tracking in the low-low mode. The distance between the satellites is measured using a K-Band (microwave) Ranging (KBR) system every five seconds. At the same time, the Global Positioning System (GPS) is employed in the high-low mode to determine their precise location (Tapley et al., 2004a,b). The low-low and high-low observations are exemplified in Fig. 2.1. Precise orbit determination with Satellite Laser Ranging (SLR) reflectors is also used as an independent check (Tapley et al., 2004a). The effects of non-gravitational surface forces are removed from along track observations before gravity modeling using measurements from on-board accelerometers (Tapley et al., 2004b).

The variation in the distance between the two GRACE satellites is largely attributed to gravitational variation of the mass within the Earth's interior, on its surface, and in the atmosphere. These variations in gravity may be due to rapid or slow changes, e.g., caused by mass distribution of the Earth or mass transport of water in the oceans, movement of water vapor and other components in the atmosphere (Schmidt et al., 2008a). Thus, they become visible in the GRACE level 0 and level 1 data, where level 0 consists of raw telemetry data reception that are collected by DLR, and level 1, including level 1A and level 1B products, are known as the processed level 0 data (for details see, e.g., Case et al., 2002).

Over any typical 30-day span (the nominal data accumulation interval of GRACE gravity field solutions) of the non-repeating orbit configurations, a dense ground track is obtained (see e.g., Elsaka, 2010, page 88). Therefore, usually one month of the pre-processed GRACE along-track range (rate) data, derived from GRACE level 1B data, are used to compute time-variable gravity field solutions that are known as GRACE level 2 data (Flechtner, 2007b). In fact, estimating the long-term and seasonal hydrological signals is one of the major applications of GRACE level 2 products (Tapley et al., 2004a,b, Schmidt et al., 2008a). Therefore, recovering monthly mean

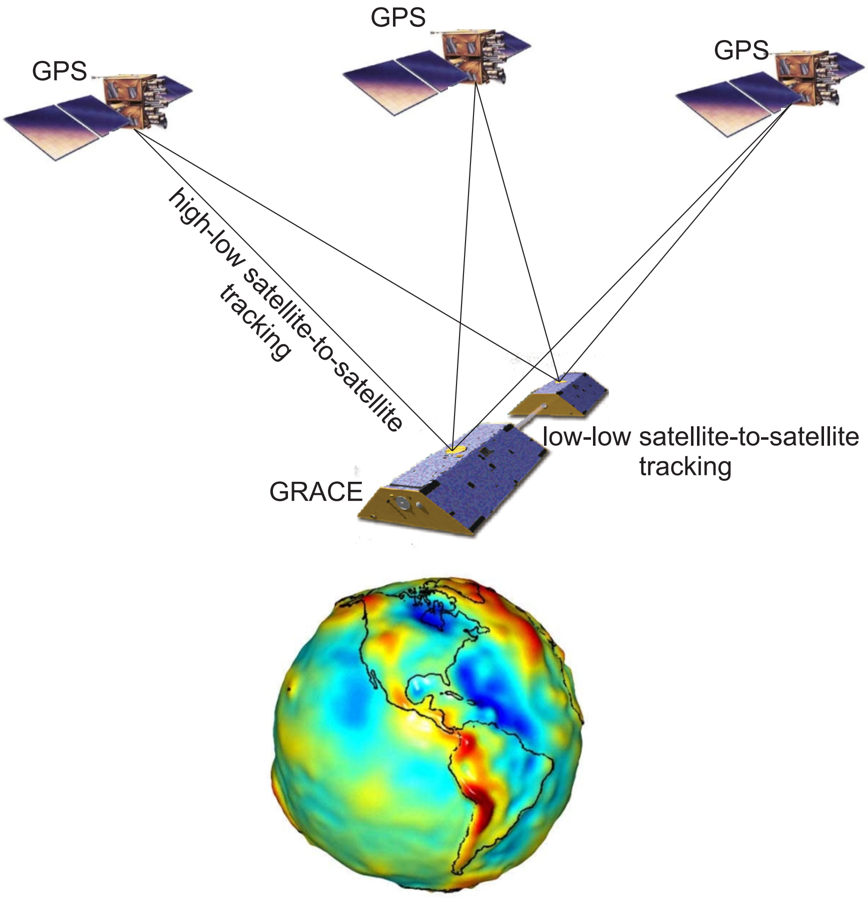


Figure 2.1: Overview of the GRACE satellite-to-satellite tracking in the low-low and high-low modes.

gravity field solutions from the sampled data requires a careful reduction of the short-term (subdaily to monthly) variations of the atmosphere and the ocean mass changes, since these effects may alias into longer periods (Han et al., 2004). This procedure is discussed in Section 2.2.

GRACE level 2 products are provided by a number of institutions, each employing different processing techniques, background models, and assumptions. Three official centers: Center for Space Research (CSR), Jet Propulsion Laboratory (JPL) in the USA, and the GeoForschungsZentrum (GFZ), Potsdam in Germany are responsible for providing the GRACE monthly level 2 solutions. Yet, GRACE data are also processed by other groups, e.g., Bonn University in Germany that provided the GRACE-ITG2010 monthly solutions (Mayer-Gürr et al., 2010a), as well as GSFC/NASA in the USA (http://grace.gsfc.nasa.gov/), GRGS in France (http://grgs.obs-mip.fr/grace), and DUT in the Netherlands.

2.1 From Geopotential Coefficients to Total Water Storage (TWS) Changes

Each GRACE level 2 time-variable gravity field product consists of a set of fully normalized potential spherical harmonic coefficients (SHCs) c_{nm} and s_{nm}, where n is the degree of the SHC and m its order. Using GRACE level 2 products, the Earth's gravitational potential V can be described by

$$V(\lambda,\theta,r) = \frac{GM}{R}\sum_{n=0}^{n_{\max}}\left(\frac{R}{r}\right)^{n+1}\sum_{m=0}^{n}\bar{P}_{nm}(\cos\theta)\,[c_{nm}\cos(m\,\lambda)+s_{nm}\sin(m\,\lambda)], \tag{2.1}$$

where λ, θ, and r are the spherical coordinates, R the Earth's radius, M the total mass of the Earth, G Newton's gravitational constant, and $\bar{P}_{nm}$ are the normalized associated Legendre functions (Heiskanen and Moritz, 1967).

Wahr et al. (1998) assumed a thin layer at the surface of the Earth to formulate the density redistribution ($\Delta\sigma$) related to time-dependent changes of the gravity field as

$$\Delta\sigma(\lambda,\theta) = \frac{M}{4\pi R^2}\sum_{n=1}^{n_{\max}}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k_n^{'})}\bar{P}_{nm}(\cos\theta)\,[\Delta c_{nm}\cos(m\,\lambda)+\Delta s_{nm}\sin(m\,\lambda)], \tag{2.2}$$

where $k_n^{'}$ represents the degree-dependent gravitational load Love numbers (see for example, Farrell, 1972), and Δc_{nm} and Δs_{nm} are the temporal changes of the potential SHCs, which are computed by reducing 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).

The density values of Eq. (2.2) can be converted to changes in equivalent water height (EWH) using

$$\Delta E(\lambda,\theta) = \frac{M}{4\pi R^2\rho_w}\sum_{n=1}^{n_{\max}}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k_n^{'})}\bar{P}_{nm}(\cos\theta)\,[\Delta c_{nm}\cos(m\,\lambda)+\Delta s_{nm}\sin(m\,\lambda)]. \tag{2.3}$$

In Eq. (2.3), $\rho_w = 1025\frac{\mathrm{kg}}{\mathrm{m}^3}$ is the average density of water. EWH derived from the potential coefficients of GRACE level 2 products provides 'total water storage (TWS)' changes. Therefore, TWS (the main parameter of interest in this thesis) is defined as the sum of all available water storage on and below the surface of the Earth. Using GRACE level 2 products to estimate ocean and land mass changes, through the formulations mentioned earlier, requires a number of nontrivial issues to be taken into account, which are briefly addressed in the following.

2.1.1 Low Degree Coefficients

Degree 1 Coefficients

GRACE level 2 products are computed in a reference frame that is fixed to the center of mass (CM) of the Earth, ocean and its surrounding atmosphere. In the CM frame, the retrieval of the degree 1 coefficients of the surface loading variations cause a singularity. In order to recover these, one has to transform the reference frame origin of the GRACE data by supplying auxiliary degree 1 coefficients. Temporal changes in degree 1 coefficients represent a considerable mass variations (Chambers, 2006). Therefore, its omission might have a significant impact on the recovery of, e.g., high-latitude mass variability and large scale oceanic mass exchange. GRACE degree 1 coefficients, however, can be augmented by considering the geocenter motion, defined as the relative motion of the center of figure of the Earth with respect to the CM of the Earth system. This can be done e.g., by using the results of Rietbroek et al. (2009, 2012b) who employed the observations of GPS, GRACE, and ocean bottom pressure in an inversion approach in order to estimate the geocenter motion. To compute degree 1 coefficients, in this thesis, time series of the geocenter motion were downloaded from the website of the Astronomical, Physical, and Mathematical Geodesy (APMG) group, Bonn University (http://www.igg.uni-bonn.de/apmg/index.php?id=geozentrum). Desired degree 1 coefficients were estimated following the formulation in Kusche and Schrama (2005).

Degree 2 Coefficients

Due to the orbital geometry and the short distance between the GRACE 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 Parameter (EOP), or those obtained from satellite laser ranging present better accuracy than those derived from GRACE. In this thesis, therefore, the time series of degree 2 coefficients (Δc_{2m} and Δs_{2m} in Eq. (2.3)), which are obtained from the analysis of SLR data (Cheng and Tapley, 2004) are used to replace those of GRACE-derived degree 2 coefficients. The time series can be downloaded from the Jet Propulsion Laboratory (JPL) website (grace.jpl.nasa.gov).

By replacing degree 1 and degree 2 coefficients of GRACE level 2 products by other estimations, 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 products, e.g., estimations of the low degree coefficients in Rietbroek et al. (2012b). Otherwise, correlations of the new low degree coefficients and the GRACE-derived higher degree coefficients must be introduced.

2.1.2 Smoothing

The GRACE-derived SHCs at higher degrees are strongly affected by correlated noise, which manifests itself as a 'striping pattern' in grids of TWS anomalies. Thus, smoothing the TWS results, or averaging over specific regions, needs to be applied to reduce the noise. This issue is usually discussed in the context of filtering approaches (see e.g., Jekeli, 1981, Swenson and Wahr, 2006, Kusche, 2007, Klees et al., 2008, Kusche et al., 2009). Smoothing operators can be applied in either spatial or spectral domain in order to suppress the effect of noise in maps and area averages. In general, a smoothed version of an arbitrary field $\Delta E^{\mathrm{W}}(\lambda,\theta)$ can be derived by convolving the original field E against a smoothing kernel W (see e.g., Kusche et al., 2011). As a result, the EWH field of Eq. (2.3) can be filtered in the spectral domain to suppress the noise and derive a smoothed field ΔE^{W} as

$$\begin{aligned}
\Delta E^{\mathrm{W}}(\lambda,\theta) &= \frac{M}{4\pi R^2 \rho_w} \sum_{n=1}^{n_{max}} \sum_{m=0}^{n} \frac{(2n+1)}{(1+k'_n)} \bar{P}_{nm}(\cos\theta) \times \\
&\Big\{ \underbrace{\Big[\sum_{n'=1}^{n_{max}} \sum_{m'=0}^{n'} w_{c\,nm}^{c\,n'm'} \Delta c_{n'm'} + w_{c\,nm}^{s\,n'm'} \Delta s_{n'm'} \Big]}_{\Delta c^{\mathrm{W}}_{nm}} \cos(m\,\lambda) + \\
&\underbrace{\Big[\sum_{n'=1}^{n_{max}} \sum_{m'=0}^{n'} w_{s\,nm}^{c\,n'm'} \Delta c_{n'm'} + w_{s\,nm}^{s\,n'm'} \Delta s_{n'm'} \Big]}_{\Delta s^{\mathrm{W}}_{nm}} \sin(m\,\lambda) \Big\}. \qquad (2.4)
\end{aligned}$$

In Eq. (2.4), w represents the filter in the spectral domain (Han et al., 2005, Kusche et al., 2009). The subindices of c and s (e.g., in $w_{c\,nm}^{c\,n'm'}$ and $w_{s\,nm}^{c\,n'm'}$) indicate that the multiplication produces filtered cosine and sine spherical harmonic coefficients, respectively (see also Han et al., 2005). As a result, $\Delta c^{\mathrm{W}}_{nm}$ and $\Delta s^{\mathrm{W}}_{nm}$ represent the smoothed coefficients. It should be mentioned here that, in general, a filter can be isotropic i.e., for instance, the Gaussian filter introduced in Jekeli (1981) or anisotropic, for example, those of Swenson and Wahr (2006), Kusche (2007), and Klees et al. (2008). When applying an isotropic Gaussian filter, the smoothing kernel is degree-dependent only, thus in Eq. (2.4), $w^{c}{}_{nm}^{n'm'} = w^{s}{}_{nm}^{n'm'} = w_n$. Equation (2.4) can be used to implement any anisotropic filters such as the DDK filter (Kusche et al., 2009), which has been applied in this study to filter TWS products in Chapter 5.

2.1.3 Area Averaging

To use GRACE level 2 data in hydrological applications, one might need to compute area-averaged mass variations for a specific area of interest, e.g., a river basin or a continent. Assume a basin function $f(\lambda,\theta)$ that is a global grid that contains 1 inside the region of interest and zero outside of it. The basin function f can be defined over a sphere (with radius R) represented by a spherical harmonic expansion

$$f(\lambda,\theta) = \frac{1}{4\pi}\sum_{n=1}^{n_{\max}}\sum_{m=0}^{n}\bar{P}_{nm}(\cos\theta)\,[c^f_{nm}\cos(m\,\lambda) + s^f_{nm}\sin(m\,\lambda)]. \tag{2.5}$$

Since the basin function f is known by definition, the coefficients of c^f_{nm} and s^f_{nm} in Eq. (2.5) can be derived using an integral approach (Wang et al., 2006) or a least squares approach (Sneeuw, 1994). Considering the filtering in Eq. (2.4) and the basin function of Eq. (2.5), one can estimate the average of TWS over the region f by

$$\begin{aligned}\Delta E_f{}^{\mathrm{W}}(\lambda,\theta) = \frac{M}{4\pi R^2\rho_w A_f}\sum_{n=1}^{n_{max}}\sum_{m=0}^{n}\frac{(2n+1)}{(1+k'_n)}\bar{P}_{nm}(\cos\theta)\times\\ [c^f_{nm}\Delta c^{\mathrm{W}}_{nm}\cos(m\,\lambda) + s^f_{nm}\Delta s^{\mathrm{W}}_{nm}\sin(m\,\lambda)].\end{aligned} \tag{2.6}$$

In Eq. (2.6), R is the Earth radius, and A_f is the area of the basin function f.

2.1.4 Leakage Problem

As shown in Eq. (2.3), GRACE-TWS changes are derived from time-variable potential coefficients. The limited range of potential SHCs in GRACE level 2 products (degree and order 60 in CSR and 120 in JPL and GFZ products) restricts the spatial resolution of the GRACE-derived TWS fields. As a result, water storage signals with spatial variability of smaller than a few hundred kilometers are not present in GRACE-derived TWS maps. This limitation makes detection of mass anomalies over, e.g., the land-ocean boundaries even more difficult, whereas one might detect spatially propagated storage change signals from oceans to land or vice versa that mask each other (see, e.g., Chen et al., 2006, Awange et al., 2009). This impact is know as the ‘spectral leakage’ in the literature.

Another issue in GRACE level 2 products is the noise contamination of potential SHCs that increases with increasing degree and order. In other words, short-wavelength spatial changes in GRACE-derived TWS products exhibit a high level of noise, which is usually dealt with by applying filters (Jekeli, 1981, Kusche, 2007). As a result of filtering in Eq. (2.6), some mass anomalies appear to be shifted. This apparent movement introduces biases in the mass change estimations, which is usually referred to as the 'spatial leakage' problem (Swenson and Wahr, 2002, Klees et al., 2006). Depending on the mass distribution, the bias might lead to under-estimation or over-estimation of the storage changes. Fenoglio-Marc et al. (2006, 2012), Awange et al. (2009), and Longuevergne et al. (2010) showed that the leakage is usually large over the regions, where land meets surface water bodies such as lakes and seas, and also for small basins. The leakage effect depends on the filter size, basin area and the amplitude of mass variations inside and outside the area of interest. In the case of using a filter with a large smoothing radius, more mass signals are distorted and replaced.

Due to the fact that the spatial resolution of filtered GRACE-TWS 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. 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-TWS estimations to other sources such as hydrological models (Schmidt et al., 2008a, Kusche et al., 2009, Forootan et al., 2012). An alternative approach is to account for the leakage problem by computing a damping factor, which is derived by dividing the mean value of the basin function before the filtering and its mean value after that (see also, e.g., Longuevergne et al., 2010, Landerer and Swenson, 2012). Klees et al. (2006) introduced a time variable damping factor, which can be alternatively used to account for the spatial leakage. In Sections 5.3 and 5.4, an application of statistical decomposition techniques to reduce a combination of the spectral and spatial leakage problem is discussed.

2.1.5 Glacial Isostatic Adjustment

Besides water storage variations, GRACE level 2 data also contains signals associated with the Glacial Isostatic Adjustment (GIA) of continental regions that were ice covered during the last glacial maximum about 20,000 years ago. GIA represents the Earth's viscoelastic response to the loading of glaciation and deglaciation of ice age cycles, which manifests as a trend at the relatively short era of the GRACE mission. For hydrological applications, the GIA effect is normally removed as a linear trend based on the output of GIA models (e.g., Sasgen et al., 2012). As a result, GRACE-derived present-day mass change signals are GIA model dependent. Large differences among available GIA models, can be due to the assumptions on spatial distribution of ice and history of melting, as well as viscosity differences in the Earth models, and the numerical implementation of model features such as temporal evolution of shorelines and paleo coastlines (Spada et al., 2011). Guo et al. (2012), for instance, compared 14 GIA models and showed that significant different sea level changes or land water storage changes may be obtained when different GIA models are applied to correct altimetry or GRACE data. It should be mentioned that the contribution of GIA to GRACE signal is not homogeneous 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.

2.1.6 Error Estimation

As mentioned earlier, the original GRACE gravity field coefficients contain correlated errors that can be significantly reduced by filtering the data. The root mean square (RMS) of the time series of GRACE-TWS 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 centimeter (Wahr et al., 2006). When the filtered GRACE data is scaled to account for the leakage and damping impact, the residual measurement errors must also be scaled. The rescaling might increase the errors by a factor, reaching up to ~ 3 (Landerer and Swenson, 2012). During the computation of GRACE-derive mass estimation errors, the following items should be considered:

- Most studies only consider the variances of the potential coefficients, which are reported in GRACE level 2 products (e.g., level 2 data from the GFZ and JPL centers). These errors, however, represent only the variance part of the estimated variance-covariance matrix of GRACE level 2 products. Error estimation of TWS changes derived from GRACE level 2 is more realistic when all elements of the covariance matrix are considered during the error propagation procedure.
- Even the full variance-covariance of monthly GRACE level 2 products do not account for the errors introduced by imperfect reduction of ‘background’ models. Such errors might affect the accuracy of mass estimation using level 2 products. This issue is discussed in Sections 2.2 and 5.2.
- For regions, such as Greenland and North America that are highly influenced by GIA, errors in the GIA models should be considered along with the other error sources for uncertainty estimation.

2.2 High-frequency Atmospheric and Oceanic Mass Redistribution

Within the processing of satellite gravimetry data, it is common to reduce the high-frequency signals of the Earth rotation, Earth tides, ocean tides, as well as high-frequency non-tidal oceanic and atmospheric mass redistribution from the level 1B measurements by using ‘background’ models (Flechtner, 2007a,b). Otherwise, such high-frequency mass changes will be aliased into long wavelength signals leading to misinterpretation of hydrological signals (Flechtner, 2007a,b). GFZ Potsdam routinely produces non-tidal high-frequency atmospheric and oceanic mass variation products called GRACE Atmosphere and Ocean De-aliasing level 1B (GRACE-AOD1B) products (Flechtner, 2007b).

Reducing tidal and non-tidal high-frequency mass variations from GRACE level 1B relies on a priori information of mass variations derived from models (e.g., tidal or de-aliasing ocean and atmosphere models). This reduction, however, is imperfect since the models are not accurate enough, as will be addressed following this section. It should be mentioned here that since the model errors affect the whole spectrum of the GRACE level 2 products, isolating them from geophysical signals is extremely difficult.

Tidal models are used to remove the tidal effects from GRACE measurements, since they are not well sampled by GRACE observations. Ray et al. (2003) stated that the ocean tide models are still imperfect, while the errors are mostly due to the uncertainty in radar altimetry and tide gauge observations specifically over polar and some coastal areas. Subsequently, tidal errors will appear in monthly level 2 data at alias periods (Han et al., 2004, Ray and Luthcke, 2006). Examples include the solar semi-diurnal (S2), the lunar diurnal (K1), and the lunar semi-diurnal (K2) tides, which respectively will appear as a sinusoidal variation at about 161 days, 7.5 and 3.7 years in GRACE time series (Knudsen, 2003, Ray et al., 2003, Chen et al., 2009).

GRACE-AOD1B consists of sets of 6-hourly SHCs up to degree and order 100 (Flechtner, 2007a,b) corresponding to non-tidal atmospheric and oceanic mass variations. The ocean part of the latest (RL05) version of GRACE-AOD1B is based on an improved baroclinic Ocean Model for Circulation and Tides (OMCT) (Flechtner et al., 2010, Dobslaw et al., 2013). Flechtner et al. (2013) indicated a notable noise reduction in GRACE level 2 (RL05) data, in which the better quality of the oceanic part of GRACE-AOD1B is a moderate factor (less than 10% of improvement is due to the new de-aliasing products). The atmospheric part of the RL05 is however the same as RL04 version of the data. GRACE-AOD1B can be downloaded from the Information System and Data Center (ISDC), GeoForschungsZentrum (GFZ), Potsdam (http://isdc.gfz-potsdam.de/index.php). The data are described in details at http://www.gfz-potsdam.de/AOD1B.

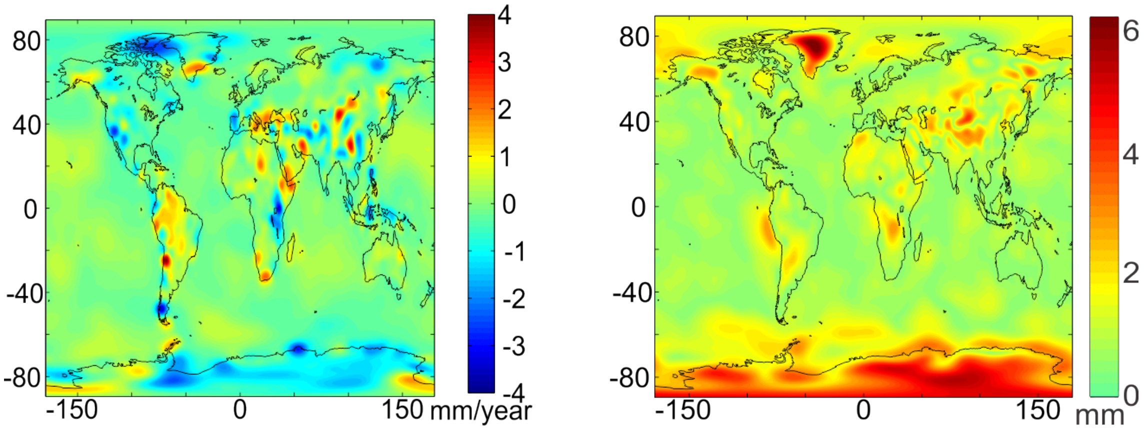
The atmospheric part of GRACE-AOD1B is computed by converting surface pressure, geopotential, temperature, and specific humidity fields, from the European Centre for Medium-Range Weather Forecasts operational analysis (ECMWFop) to potential coefficients (Flechtner, 2007a). This conversion has been realized using a three-dimensional (3D) integration approach including various approximations, e.g., in Boy and Chao (2005) and Flechtner (2007a).

Forootan et al. (2013) assessed the possible improvements within the computation of the atmospheric part of the de-aliasing products. Their modifications included: i) geometrical modification including ellipsoidal radius $r_e(\theta)$ instead of a constant radius, incorporating geoid heights from the ITG-GRACE2010s static solution instead of using the surface geopotential from ECMWF, as well as using a more accurate transformation for computing radial coordinates; ii) physical modification including the more accurate latitude- and altitude-dependent gravity acceleration formula by Heiskanen and Moritz (1967) within the vertical integration instead of a simple linear approximation of the latitude-dependent gravity acceleration, and finally iii) numerical improvements to compute more accurate atmospheric de-aliasing SHCs (Forootan et al., 2013). The modified 3D integral approach (ITG3D) that considers all above improvements represent a noticeable impact on the final results of monthly GRACE level 2 products. The possible impacts are illustrated in terms of the linear trend, as well as the annual and semi-annual amplitudes of TWS changes in Fig. 2.2.

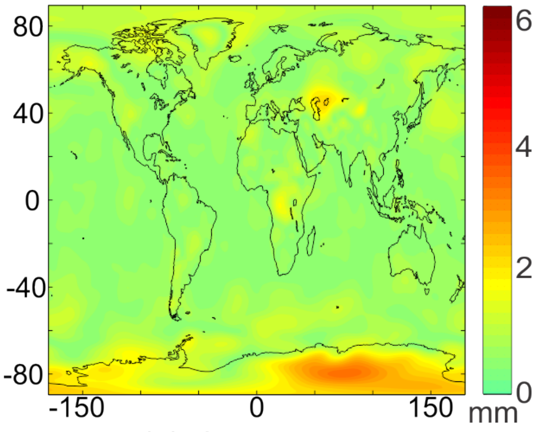
Considering the input atmospheric fields of the de-aliasing products, in fact, two large jumps in the GRACE-AOD1B atmospheric data can be identified, which occurred in January-to-February of the years 2006 and 2010, as a result of the level and resolution changes of ECMWFop fields. The existence of these jumps causes an error in GRACE level 2-derived mass estimations (Duan et al., 2012) and, as a result, should be considered for studies that address GRACE-TWS changes. To asses the impact of using a more long-term consistent atmospheric de-aliasing product on GRACE-derived mass estimations, therefore, a new set of 6-hourly total-atmospheric mass fields 'ITG3D-ERA-Interim' covering the period of January 2003 to July 2011 was computed by Forootan et al. (2014a). The new set has been improved in the atmospheric part of GRACE-AOD1B in terms of (a) the computational process and (b) the input parameters. In (a), the ITG3D integral approach (Forootan et al., 2013) is used to compute the 6-hourly products. For (b), the ERA-Interim reanalysis atmospheric fields are used as the input of the 3D integration. The jumps of 2006 and 2010 do not exist in ERA-Interim time series, thus, they are more consistent on long time-scales. ITG3D-ERA-Interim is available via the website of the Astronomical, Physical, and Mathematical Geodesy (APMG) group, Bonn University (http://www.igg.uni-bonn.de/apmg/index .php?id=itg3d_erainterim).

Figure 2.2 shows the linear trend, annual, and semi-annual amplitudes of the differences between the total atmospheric mass derived from ITG3D-ERA-Interim and GRACE-AOD1B, over January 2003 to July 2009. These results are shown after application of a Gaussian filter (Jekeli,

1981) of 300 km (half-width). The largest magnitudes for the linear rates of the differences (Fig. 2.2 (a)) are concentrated over steep orography (e.g., over Central Asia and the west part of South America), and also over Greenland and Antarctica. The cause of these differences is mainly due to the jump of 2006 in GRACE-AOD1B. Figure 2.2 (b) presents the amplitude of the annual differences between the two atmospheric mass estimations derived from ITG3D-ERA-Interim and GRACE-AOD1B. The magnitude of the differences reaches up to 6 mm in EWH over Antarctica and Greenland. Finally, the amplitudes of the semi-annual differences are shown in Fig. 2.2 (c), which are almost half of the annual differences. The same computations were repeated for the period of January 2003 to July 2011, the maximum amplitude of the differences did not change significantly over most regions of the globe (results are not shown here). For more details, we refer to Forootan et al. (2014a). The impact of errors in the atmospheric de-aliasing products on GRACE-derived mass estimation has been considered for the statistical signal separation of Section 5.2.



(a) Linear Trend (b) Annual



(c) Semi-annual

Figure 2.2: Atmospheric mass differences (in EWH) derived from the ITG3D-ERA-Interim and GRACE-AOD1B (based on ECMWFop), over January 2003 to July 2009. The differences are filtered using a Gaussian filter with a half-width radius of 300 km. Figure 2.2 (a) shows the slopes of the linear trend, fitted to monthly averaged differences. Figure 2.2 (b) represents the amplitude of the annual differences and that of Fig. 2.2 (c) corresponds to the amplitude of the semi-annual differences. The results can be considered as mass uncertainties, which are caused by imperfect reduction of atmospheric de-aliasing products (for more details see Forootan et al., 2014a).

2.3 Computational Steps for Estimating TWS Changes

To estimate TWS changes for most of the applications in this thesis or to compare them with the TWS output of hydrological models, the following steps are usually performed:

- The GRACE level 2 data are derived in terms of fully normalized geopotential SHCs. The products are downloaded from the website of official processing centers (GFZ, JPL, and CSR) or the website of APMG group at Bonn University (GRACE ITG2010).
- The gravity solutions are augmented by the degree 1 term from Rietbroek et al. (2009) in order to include the variation of the Earth's center of mass with respect to a crust-fixed reference system. This transformation is recommended for those studies that address mass changes in the oceans and high latitude TWS changes.
- The zonal degree 2 coefficients are replaced by the SLR results, downloaded from the JPL website (http://grace.jpl.nasa.gov).
- GRACE level 2 SHCs at higher degrees are affected by correlated noise and are, therefore, smoothed by applying the DDK anisotropic filter (DDK2 or DDK3, Kusche, 2007, 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 TWS output of global hydrological models (see, e.g., Werth et al., 2009).
- Filtered level 2 solutions are then used to generate the global TWS changes according to the approach presented by Wahr et al. (1998).
- Errors of GRACE level 2 coefficients are propagated to grids of TWS changes considering full error covariance matrix provided in ITG2010 products at Bonn university (http://www.igg.uni-bonn.de/apmg/index.php?id=gravitationsfeldmodelle). A bootstrap approach is used to generate errors from the estimated error fields when applying the introduced statistical decomposition methods of Chapters 3 and 4.
- Errors due to incomplete reduction of short-term mass variations of atmospheric de-aliasing products are taken into account similar to Forootan et al. (2014a).
- Similar to the GRACE products above, the DDK filter (DDK2 or DDK3, Kusche et al., 2009) is applied to the TWS output of hydrological models in order to preserve exactly the same spectral content as with the filtered GRACE products.
- The GIA corrections are not considered in this study, since the case studies include those areas that are not significantly influenced by GIA.

3. Second Order Statistical Signal Decomposition

In this chapter, the mathematical foundation of the second order statistical decomposition methods are described, and their applications to reduce the high-dimensionality of geophysical data sets are addressed. Therefore, in Section 3.1, the Principal Component Analysis/Empirical Orthogonal Function (PCA/EOF) method is described. This includes: its central idea (Section 3.1.1), mathematical derivation (Section 3.1.2), and its application to data transformation (Section 3.1.3). To improve the interpretation of PCA result, the Rotated Empirical Orthogonal Function (REOF) method is discussed in Section 3.2. The Extended Empirical Orthogonal Function (EEOF) and the Complex Empirical Orthogonal Function (CEOF) methods are respectively introduced in Sections 3.3 and 3.4 to deal with time series that contain non-stationary patterns. This is of interest to this study since time series of GRACE-derived total water storage (TWS) changes are sampled from the global water cycle. As a result, they represent dynamical characteristics, that are seen as trends and oscillations in the time series. In Table 3.1, the second order methods, discussed in this chapter, are summarized. Finally, the significance and uncertainty estimations of the methods are presented in Sections 3.5 and 3.6, respectively.

Method	Assumption	Details
PCA/EOF Section 3.1	observations are stationary	Auto-covariance is used for decomposition
Rotated EOF Section 3.2	observations are stationary	PCA components are rotated to derive simpler patterns
Extended EOF Section 3.3	observations are non-stationary	Lagged auto-covariance is used for decomposition
Complex EOF Section 3.4	observations are non-stationary	A complexified field is used to compute auto-covariance matrix and its decomposition

Table 3.1: Summary of the second order statistical signal separation methods discussed in this chapter.

3.1 Principal Component Analysis/Empirical Orthogonal Function (PCA/EOF)

Time series of geophysical data, such as maps of TWS changes, sea surface height (SSH), sea surface temperature (SST), and rainfall, contain significant temporal and spatial correlations. To use these observables for hydrological applications, it is helpful to find a few spatio-temporal patterns dominating the variability within the data sets. Identification of these patterns can aid in physical interpretation of the available data, comparison of different data sets, and removing irrelevant small-scale signals or noise (Preisendorfer, 1988, von Storch and Navarra, 1999).

3.1.1 Central Idea of PCA

To illustrate the main idea of PCA, assume two random variables, x_1 and x_2, with a Gaussian distribution as illustrated in Fig. 3.1. PCA aims at finding optimal orthogonal directions, known as eigenvectors, along which the observation values are maximally distributed. These directions are shown by the black lines in Fig. 3.1, while one eigenvector points along the axis of the largest variability (i.e. from bottom left to top right), and the other is orthogonal to it.

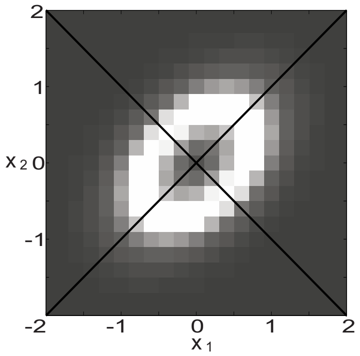


Figure 3.1: Illustration of the PCA result applied to a bi-variate Gaussian distribution. Black lines show the orthogonal directions (eigenvectors) on which the joint distribution is maximally distributed.

3.1.2 PCA from Eigenvalue Decomposition

In this section, a brief description of PCA, applied to multivariate observations is presented. One can consider the observed time series, which can be grid maps of, e.g., TWS changes, after removing their temporal mean, stored in data vectors $\mathbf{x}_i$, containing n time epochs corresponding to the i'th observation (or the i'th grid element)

$$\mathbf{x}_i = \begin{pmatrix} x_{1,i} \\ \vdots \\ x_{n,i} \end{pmatrix} \qquad i = 1, \dots, p. \tag{3.1}$$

Therefore, the centered data (i.e. $\frac{1}{n}\sum_{t=1}^{n} x_{t,i} = 0, i = 1, \dots, p$) can be stored in the $n \times p$ data matrix $\mathbf{X}$, where n represents the time epochs, and p indicates the number of grid points. The data matrix $\mathbf{X}$ is defined as

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p) = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,p} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \cdots & x_{n,p} \end{pmatrix}. \tag{3.2}$$

It should be mentioned here that the temporally centered data of Eq. (3.2) is actually derived from the original data of $\mathbf{X}'$, which is ordered in the same way as Eq. (3.2), or

$$\mathbf{X} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)\mathbf{X}' = \mathbf{H}\mathbf{X}', \tag{3.3}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{1}_n = [1, \dots, 1]^T_{n\times 1}$, and $\mathbf{H} = \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T\right)$ is a centering operator of order n. Note also that the data matrix could as well be assembled from stacking

p n-dimensional row-vectors, where each row represents all observations for a particular time epoch. Such formulation is provided in Kusche et al. (2011).

To keep the notation simple, from now on and unless otherwise stated, the dash-matrix $\mathbf{X}'$ will be dropped and $\mathbf{X}$ (Eq. (3.2)), which simply denotes the anomaly data (centered data) matrix will be used. Therefore, the entries x in Eq. (3.2) generate the 'observation space', considering

$$\mathbf{x}_i = x_{1,i}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}_{n\times1} + x_{2,i}\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}_{n\times1} + \cdots + x_{n,i}\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}_{n\times1}, i = 1,\ldots,p. \tag{3.4}$$

$$\Leftrightarrow\ \mathbf{x}_i = x_{1,i}\mathbf{p}_1 + x_{2,i}\mathbf{p}_2 + \cdots + x_{n,i}\mathbf{p}_n,$$

and

$$\tilde{\mathbf{x}}_m = x_{m,1}\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}^T_{p\times1} + x_{m,2}\begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}^T_{p\times1} + \cdots + x_{m,p}\begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}^T_{p\times1}, m = 1,\ldots,n. \tag{3.5}$$

$$\Leftrightarrow\ \tilde{\mathbf{x}}_m = x_{m,1}{\mathbf{e}_1}^T + x_{m,2}{\mathbf{e}_2}^T + \cdots + x_{m,p}{\mathbf{e}_p}^T,$$

where $\tilde{\mathbf{x}}$ represents one row of the data matrix $\mathbf{X}$. The entries of the base-functions $\mathbf{p}_i$ in Eq. (3.4) refer to time epochs. In contrast, the elements of the base-functions $\mathbf{e}_m$ in Eq. (3.5) represent spatial patterns. Both $\mathbf{p}_i$ and $\mathbf{e}_m$ are orthogonal directions and normalized with respect to the standard scalar product, i.e. $(\mathbf{p}_k, \mathbf{p}_l) = {\mathbf{p}_k}^T\mathbf{p}_l$. Using Eqs. (3.4) and (3.5), one is able to span the observation space with regard to the orthogonal base-functions of $\mathbf{p}_i$ and $\mathbf{e}_m$ (Koch, 1999, page 11 to 13).

Statistical decomposition methods, which will be discussed in Chapters 3 and 4 also search for alternative directions (base-functions) in the observation space (see e.g., Fig. 3.1). The goal of these approaches is to find directions in the data that aid in revealing more relevant information hidden in observations (Preisendorfer, 1988, von Storch and Navarra, 1999).

Once the anomaly data matrix was defined in Eq. (3.2), the sample auto-covariance matrix can be written as

$$\mathbf{C} = \frac{1}{n}\mathbf{X}^T\mathbf{X}, \tag{3.6}$$

which contains the covariances between the time series of any pair of grid points. The extended form of Eq. (3.6) is written as

$$\mathbf{C} = \frac{1}{n}\begin{pmatrix}\sum_{t=1}^n {x_{t,1}}^2 & \sum_{t=1}^n x_{t,1}x_{t,2} & \cdots & \sum_{t=1}^n x_{t,1}x_{t,p}\\ \sum_{t=1}^n x_{t,2}x_{t,1} & \sum_{t=1}^n {x_{t,2}}^2 & \cdots & \sum_{t=1}^n x_{t,2}x_{t,p}\\ \vdots & \vdots & \ddots & \vdots\\ \sum_{t=1}^n x_{t,p}x_{t,1} & \sum_{t=1}^n x_{t,p}x_{t,2} & \cdots & \sum_{t=1}^n {x_{t,p}}^2\end{pmatrix}. \tag{3.7}$$

It should be mentioned that the auto-covariance matrix of $\mathbf{X}$ (Eq. (3.6)) can also be computed as

$$\tilde{\mathbf{C}} = \frac{1}{p}\mathbf{X}\mathbf{X}^T, \tag{3.8}$$

which in contrast to Eq. (3.6), contains the spatial variance and covariances of the data viewed as a function of position. The two auto-covariance matrices of $\mathbf{C}$ in Eq. (3.7) and $\tilde{\mathbf{C}}$ in Eq. (3.8) share the same set of eigenvalues since both are derived based on the same data matrix $\mathbf{X}$. However, Eq. (3.6) is considered in the reminder of this section.

The aim of PCA is to find directions such as $\mathbf{e} = (e_1, e_2, \ldots, e_p)^T$ in the data $\mathbf{X}$, while the projection of $\mathbf{X}$ on $\mathbf{e}$ (i.e. $\mathbf{Xe}$) has the maximum variability (variance). The vector $\mathbf{e}$ is an eigenvector of $\mathbf{C}$ and called Empirical Orthogonal Function (EOF). The variance of the projection, $\mathbf{Xe}$, can be written as

$$\mathrm{var}(\mathbf{Xe}) = \frac{1}{n}(\mathbf{Xe})^T\mathbf{Xe} = \frac{1}{n}(\mathbf{e}^T\mathbf{X}^T)\mathbf{Xe}, \tag{3.9}$$

while inserting Eq. (3.6) in Eq. (3.9) leads to

$$\mathrm{var}(\mathbf{Xe}) = \mathbf{e}^T\mathbf{C}\ \mathbf{e}. \tag{3.10}$$

Equation (3.10) indicates that $\mathbf{e}$ is unitary since it factorizes the variance of the auto-covariance matrix $\mathbf{C}$. Therefore, finding a projection that maximizes the projected variance in Eq. (3.10) is equivalent to the solution of PCA that is

$$\max_{\mathbf{e}}(\mathbf{e}^T\mathbf{C}\ \mathbf{e}), \tag{3.11}$$

where $\mathbf{e}^T\mathbf{e} = 1$. Equation (3.11) can simply be solved by an eigenvalue decomposition of the auto-covariance matrix as

$$\mathbf{C}\ \mathbf{e} = \lambda^2\ \mathbf{e}, \tag{3.12}$$

where λ^2 is an eigenvalue of $\mathbf{C}$ (Jolliffe, 1986). It should be mentioned here that, since the auto-covariance matrix $\mathbf{C}$ is always symmetric and positive definite/semidefinite, Eq. (3.12) is always diagonalizable. The maximum number of positive eigenvalues that one can compute, is the rank of $\mathbf{C}$, i.e. $\leq \min(n, p)$. Here, we assume that the columns of matrix $\mathbf{X}$ are distinguishable. As a result of this assumption, the rank of the auto-covariance matrix is $\min(n, p)$. Considering the fact that the PCA-derived eigenvectors are computed based only up to second order statistical information (considering the mean and the information contained in the auto-covariance matrix), the PCA method is categorized as a 'second order' statistical decomposition approach (Hyvärinen, 1999a).

Stacking all the possible directions of $\mathbf{e}$ in the columns of matrix $\mathbf{E}$, Eq. (3.12) becomes

$$\mathbf{C}\ \mathbf{E} = \mathbf{E}\ \Lambda^2, \tag{3.13}$$

where Λ^2 is a diagonal matrix that stores all computed eigenvalues and $\mathbf{E}^T\mathbf{E} = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. For an auto-covariance matrix $\mathbf{C}$ with the full rank of n, the dimension of Λ^2 is $n \times n$, and that of $\mathbf{E}$ is $p \times n$. The variance of $\mathbf{X}$ is equal to the trace of $\mathbf{E}^T\ \mathbf{C}\ \mathbf{E} = \mathbf{E}^T\mathbf{E}\Lambda^2 = \mathbf{I}\ \Lambda^2$, thus, one can derive the total variance (energy) of the data matrix $\mathbf{X}$ as

$$\mathrm{var}(\mathbf{X}) = \sum_{i=1}^{\min(n,p)} {\lambda_i}^2, \tag{3.14}$$

or

$$\mathrm{var}(\mathbf{X}) = \mathrm{trace}(\Lambda^2), \tag{3.15}$$

where trace(.) is an operator that computes the sum of the diagonal elements (Koch, 1999, page 44). Therefore, the i'th eigenvalue (${\lambda_i}^2$) corresponding to the i'th EOF ($\mathbf{e}_i$) gives a measure of the explained variance by $\mathbf{e}_i$, which in percentage is

$$\eta_i\% = \frac{100{\lambda_i}^2}{\mathrm{trace}(\Lambda^2)}\%. \tag{3.16}$$

In Eq. (3.16), η_i is the total variance fraction that is represented by $\mathbf{e}_i$ (for details see e.g., Kusche et al., 2011).

The projection of the data $\mathbf{X}$ on the i'th EOF ($\mathbf{e}_i$), i.e. $\mathbf{p}_i = \mathbf{X}\mathbf{e}_i$, gives its corresponding temporal evolution that is called Principal Component (PC) or

$$\mathbf{p}_i(t) = \sum_{s=1}^{p} x(t,s)\ \mathbf{e}_i(s). \tag{3.17}$$

Considering Eqs. (3.11) and (3.17), one can conclude that PCA expands $\mathbf{X}$ in terms of a new set of spatially orthogonal vectors (EOFs) associated with temporally uncorrelated time series known as Principal Components (PCs) (Jolliffe, 1986, Preisendorfer, 1988). The PCA decomposition is written as

$$\mathbf{X} = \mathbf{P}\mathbf{E}^T, \tag{3.18}$$

where the orthogonal components are arranged with respect to the magnitude of the eigenvalues, and $\mathbf{P}$ stores the PCs in its columns ($\mathbf{P}^T\mathbf{P} = \mathbf{I}\boldsymbol{\Lambda}^2$). In matrix form, the PCs are the projection of the original data $\mathbf{X}$ on the orthogonal base-functions $\mathbf{E}$, which means

$$\mathbf{P} = \mathbf{X}\mathbf{E}. \tag{3.19}$$

$\mathbf{P}$ in Eq. (3.19) contains the values of Eq. (3.17) in its columns.

Similar to Eq. (3.18), the data matrix can also be decomposed by the Singular Value Decomposition (SVD) method as (Preisendorfer, 1988)

$$\mathbf{X} = \bar{\mathbf{P}}\boldsymbol{\Lambda}\mathbf{E}^T, \tag{3.20}$$

where $\bar{\mathbf{P}}$ contains normalized PCs, i.e. $\bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I}$, and $\boldsymbol{\Lambda}$ is diagonal and holds the singular values λ ordered according to their magnitude. From Eq. (3.20), it follows that

$$\begin{aligned}
\mathbf{X}^T\mathbf{X} &= \mathbf{E}\boldsymbol{\Lambda}\bar{\mathbf{P}}^T\bar{\mathbf{P}}\boldsymbol{\Lambda}\mathbf{E}^T,\ \bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I},\\
\mathbf{X}^T\mathbf{X} &= \mathbf{E}\boldsymbol{\Lambda}^2\mathbf{E}^T,\\
\mathbf{X}^T\mathbf{X}\mathbf{E} &= \mathbf{E}\boldsymbol{\Lambda}^2\mathbf{E}^T\mathbf{E},\ \mathbf{E}^T\mathbf{E} = \mathbf{I},\\
\mathbf{X}^T\mathbf{X}\mathbf{E} &= \mathbf{E}\boldsymbol{\Lambda}^2.
\end{aligned} \tag{3.21}$$

Similarly, one can also write that

$$\begin{aligned}
\mathbf{X}\mathbf{X}^T &= \bar{\mathbf{P}}\boldsymbol{\Lambda}\mathbf{E}^T\mathbf{E}\boldsymbol{\Lambda}\bar{\mathbf{P}}^T,\ \mathbf{E}^T\mathbf{E} = \mathbf{I},\\
\mathbf{X}\mathbf{X}^T &= \bar{\mathbf{P}}\boldsymbol{\Lambda}^2\bar{\mathbf{P}}^T,\\
\bar{\mathbf{P}}^T\mathbf{X}\mathbf{X}^T &= \bar{\mathbf{P}}^T\bar{\mathbf{P}}\boldsymbol{\Lambda}^2\bar{\mathbf{P}}^T,\ \bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I},\\
\bar{\mathbf{P}}^T\mathbf{X}\mathbf{X}^T &= \boldsymbol{\Lambda}^2\bar{\mathbf{P}}^T,\\
\mathbf{X}\mathbf{X}^T\bar{\mathbf{P}} &= \bar{\mathbf{P}}\boldsymbol{\Lambda}^2.
\end{aligned} \tag{3.22}$$

Equations (3.21) and (3.22) demonstrate that the columns of $\mathbf{E}$ and the columns of $\bar{\mathbf{P}}$ are the eigenvectors of $\mathbf{X}^T\mathbf{X}$ and $\mathbf{X}\mathbf{X}^T$, respectively.

One benefit of the PCA decomposition, that can be derived from SVD (Preisendorfer, 1988), is that it allows to reduce the dimension of spatio-temporal data sets. Let $j < \min(n,p)$ be the number of retained modes (those associated with the largest singular values), the reconstruction of the data matrix is

$$\mathbf{X} \simeq \mathbf{X}_j = \bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{E}_j^T, \tag{3.23}$$

with the $n \times j$, $j \times j$, and $p \times j$ matrices $\bar{\mathbf{P}}_j$, $\mathbf{\Lambda}_j$, and $\mathbf{E}_j$, and ${\mathbf{E}_j}^T\mathbf{E}_j = \mathbf{I}_j$, $\bar{\mathbf{P}}_j^T\bar{\mathbf{P}}_j = \mathbf{I}_j$. Each $\bar{\mathbf{P}}_i\mathbf{\Lambda}_i\mathbf{E}_i^T, i \in \{1, \ldots, \min(n,p)\}$, is called a 'mode' of variability in $\mathbf{X}$ (Preisendorfer, 1988). The variance percentage that each mode represents is computed in Eq. (3.16). Equation (3.23) is used afterwards for dimension reduction using the second order approach of PCA. A discussion on the selection of j can be found in Section 3.5 (see details in Preisendorfer, 1988, chapter 5).

3.1.3 PCA as a Data Whitening Method

'Whitening' or 'sphering' is a statistical concept, which can be expressed as a linear transformation of time series to from a set of uncorrelated time series with variances equal to unity. This transformation has been used in other techniques, e.g., rotated PCA (Section 3.2), Blind Source Separation (BSS, Hyvärinen, 1999a,b) and ICA (Chapter 4) as a first step to enhance the computations (Cardoso, 1992, 1999).

Consider Eq. (3.20), which was introduced to decompose the data matrix $\mathbf{X}$ into orthonormal components of $\bar{\mathbf{P}}$ and $\mathbf{E}$. By definition, these components can be interpreted as a whitened version of the original time series. For instance, the temporally white components can be derived from

$$\bar{\mathbf{P}} = \mathbf{X}\mathbf{E}\mathbf{\Lambda}^{-1}, \tag{3.24}$$

where $\bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I}$.

To illustrate the transformation in Eq. (3.24), two annual and semi-annual sinusoidal signals of $\mathbf{a}(t)$ and $\mathbf{b}(t)$ are considered, where t represents 200 sampling epochs in month. Thus, by definition we consider that $\mathbf{a}(t) = 100\,\sin(\frac{2\pi t}{12}) + n_1(t)$ and $\mathbf{b}(t) = 50\,\sin(\frac{2\pi t}{6}) + n_2(t)$, with $n_1(t)$ and $n_2(t)$ are random noise. The graph on the left side of Fig. 3.2 shows the two signals plotted against each other. The signals are not centered and the variance on the two axes are not the same.

Then, Eq. (3.24) is applied on the data matrix $\mathbf{X}$, including $\mathbf{a}$ and $\mathbf{b}$ in its columns, to derive the transformed results of $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$, which are shown against each other on the right side of Fig. 3.2. The variance of the projections on both axes are now equal to one, and the correlation between $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$ is zero. This transformation is useful since after its application, all odd statistical moments of $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$, as well as the covariances of $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$ are zero (Cardoso, 1999, Comon, 1994a). After applying the whitening transformation, the joint distribution of $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$ is sphered (see the right side of Fig. 3.2). This symmetrical property of the joint distribution will enhance the computations of higher order statistics, as will be shown in Chapter 4.

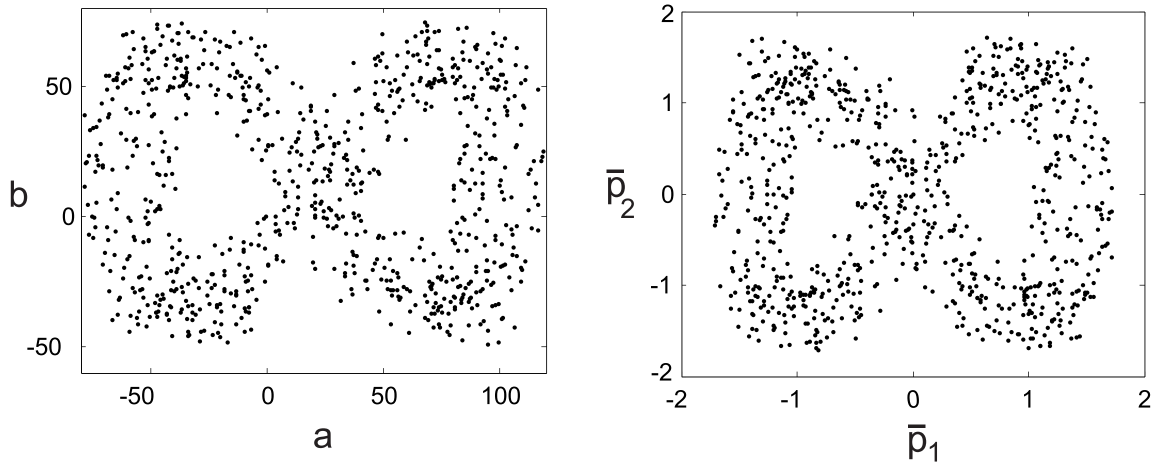


Figure 3.2: The performance of PCA as a whitening procedure. On the left, two original signals of $\mathbf{a}(t) = 100\sin(\frac{2\pi t}{12}) + n_1(t)$ and $\mathbf{b}(t) = 50\sin(\frac{2\pi t}{6}) + n_2(t)$, $t = 1, \ldots, 200$, are plotted against each other. On the right, the whitened temporal components of $\bar{\mathbf{p}}_1$ and $\bar{\mathbf{p}}_2$, derived from Eq. (3.24), are plotted against each other.

3.2 Rotated EOF

Rotated PCA/EOF techniques (REOF) are considered as an extension of the PCA decomposition to improve the interpretation of the components (e.g., Richman, 1986). REOF techniques can be derived by including a rotation matrix that rotates both EOFs and PCs in Eq. (3.23), without changing the left-hand side. The rotation can be orthogonal or oblique, and is usually applied to the leading components (Hannachi et al., 2007). To define a suitable rotation a criterion such as the 'simplicity' of the rotated modes has usually been applied (Kaiser, 1958, Richman, 1986, 1987, Jolliffe, 1987, 1995). Let $\mathbf{R}$ denote an orthogonal rotation matrix, i.e. $\mathbf{R}\mathbf{R}^T = \mathbf{I}$ (Koch, 1999, pages 43 and 65). Assuming that after an initial PCA, one decides that j data modes shall be retained, then $\mathbf{X}_j$ in Eq. (3.23) can be cast into

$$\mathbf{X}_j = \mathbf{P}_j\,\mathbf{R}_j\,\mathbf{R}_j^T\,\mathbf{E}_j^T = \mathbf{B}_j\mathbf{U}_j^T, \tag{3.25}$$

where $\mathbf{P}_j = \bar{\mathbf{P}}_j\mathbf{\Lambda}_j$, $\mathbf{U}_j = \mathbf{E}_j\mathbf{R}_j$ contains the j rotated EOFs (REOFs), and $\mathbf{B}_j = \mathbf{X}\mathbf{U}_j$ represents the rotated PCs (RPCs). Several analytical criteria for the 'simplicity' property exist in which an optimum rotation matrix $\mathbf{R}_j$ can be estimated (e.g., Browne, 2001). However, there are also difficulties regarding the implementation of most of them. For instance,

- finding a suitable number of retained EOFs or PCs to be rotated is difficult since, as we shall see e.g., later in this section and Chapter 4, the rotation criteria are usually sensitive to the number of retained EOFs or PCs (see discussions in Section 3.5).
- selecting the type of rotation, i.e. orthogonal or oblique is difficult. This is due to the fact that there is usually no 'true' pattern for justifying the performance of the decomposition.
- selecting a proper simplicity criteria is difficult due to existence of numerous criteria.

All these issues make the REOF method quite controversial (Richman, 1986, 1987, Jolliffe, 1987, 1995, Hannachi et al., 2007). Despite such drawbacks, Jolliffe (1987, 1989) and others recommended the use of rotation in order to enhance the interpretation of the PCA-derived components.

In this thesis, only those REOF techniques that are based on orthogonal rotations are discussed. Considering $\mathbf{R}$ as an orthogonal rotation matrix, $\mathbf{U}_j = \mathbf{E}_j\mathbf{R}_j$ contains the j rotated and still orthogonal REOFs, and $\mathbf{B}_j = \mathbf{X}\mathbf{U}_j$ represents the corresponding expansion coefficients of the data RPCs. Hannachi et al. (2007) found the orthogonal rotations being computationally more efficient than oblique rotations, due to matrix inversion in the latter. It should be mentioned here that, even in the case of orthogonal rotation in Eq. (3.25), the RPCs lose the property of being uncorrelated.

In general, the rotation in Eq. (3.25) aims at finding a new basis such that either the spatial patterns or the temporal expansion coefficients appear as simple as possible (e.g., Richman, 1986, Hannachi et al., 2007), with either more localized temporal or spatial structures. The actual choice of $\mathbf{R}$ is made by solving an optimization problem as

$$\max_{\mathbf{R}_j} f(\mathbf{E}_j\mathbf{R}_j), \tag{3.26}$$

which is written with respect to the spatial patterns $\mathbf{E}_j$. The functional $f(.)$ in Eq. (3.26) represents the rotation criterion. In Eq. (3.26), $\mathbf{ER}$ can be replaced by $\bar{\mathbf{P}}\mathbf{R}$ to derive simpler temporal evolutions.

In the following, some most commonly used REOF criteria are discussed. It should be mentioned here that selecting normalized or non-normalized base-functions (e.g., $\mathbf{P}$ or $\bar{\mathbf{P}}$) prior to rotation will also lead to different results. A discussion about the effect of normalization and orthogonality in REOF is provided in Mestas-Nuñez (2000). In this study, all the rotation criteria, such as the REOF in this section and that of the ICA method in Chapter 4, are based on the pre-whitened components of Section 3.1.3. Therefore, the choice of normalization is not discussed here.

VARIMAX Criterion

One of the commonly used rotation criteria to derive simpler patterns is VARIMAX, which was introduced in Kaiser (1958) as

$$f(\mathbf{R}_j) = \sum_{q=1}^{j}\left[\sum_{s=1}^{p} u_{sq}^4 - \frac{1}{p}\left(\sum_{s=1}^{p} u_{sq}^2\right)^2\right], \tag{3.27}$$

where $u_{s,q}$, $s = 1,\ldots,p$ and $q = 1,\ldots,j$, are the elements of the matrix $\mathbf{U}_j$. Assume x is a random variable. Thus, the variance of its squared elements can be estimated from $\mathrm{var}(x^2) = E(x^4) - (E(x^2))^2$, where var(.) is the variance and $E(.)$ is the expectation operator. Considering this definition, the term in squared brackets of Eq. (3.27) represents an approximation to the variance of the squared elements of the rotated EOFs ($\mathbf{U}_j = \mathbf{E}_j\mathbf{R}_j$). In other words, for the analysis of spatio-temporal data, such as grids of GRACE-derived mass changes, application of VARIMAX REOF comes down to choosing the rotation such that it maximizes the spatial variability of the REOFs. This variability will be small if the root mean squares (RMS) of the EOFs will be nearly equally distributed (in space), and it will be large if for each pattern a few regions dominate with large RMS (approaching to ±1 since the EOFs are normalized). Therefore, VARIMAX REOF maximizes the 'contrast' of the orthogonal base-functions.

As mentioned before, the procedure to derive simplified RPCs is similar to the above with the difference that one selects the first j normalized PCs ($\bar{\mathbf{P}}_j$) to be rotated and to be implemented in the criterion of Eq. (3.27) instead of the EOFs. Since the length of temporal components is n, p in Eq. (3.27) should be replace by n for deriving simplified RPCs.

QUARTIMAX Criterion

QUARTIMAX is another criterion, which seeks to maximize the variance of the orthogonal patterns (Kaiser, 1958) as

$$f(\mathbf{R}_j) = \frac{1}{pj}\sum_{s=1}^{p}\sum_{q=1}^{j}\left[{u_{sq}}^2 - \frac{1}{pj}\sum_{s=1}^{p}\sum_{q=1}^{j}{u_{sq}}^2\right]^2. \tag{3.28}$$

In Eq. (3.28), $\mathbf{U}_j^T\mathbf{U}_j = \mathbf{I}_j$, thus, the sum of the squared elements of $\mathbf{U}_j$ is constant. Therefore, Eq. (3.28), can be simplified as

$$f(\mathbf{R}_j) = \frac{1}{pj}\left[\sum_{s=1}^{p}\sum_{q=1}^{j}{u_{sq}}^4\right]. \tag{3.29}$$

VARIMAX is in general preferred to QUARTIMAX because it is slightly less sensitive to changes in the number of variables (e.g., Richman, 1986, Jolliffe et al., 2002). Hannachi et al. (2007) showed that the difference in practice is not significant. When using a rotation extension, one should note the following:

- By rotating the first j components, the rotated components jointly account for the same total variability as the j unrotated components in PCA. However, each of the rotated components no longer individually represents a maximal amount of variation. After rotation, the variance is usually spread more evenly between individual components than before. This aspect might enhance the interpretation of each rotated component. However, one might lose information on the most dominant individual sources of variation in the data (see, e.g., Richman, 1986).
- It is recommended to divide the first j PCA components into subsets such that variances of components within subsets are nearly equal. Each subset is then rotated separately within the rotation criterion (Jolliffe, 1989).

3.3 Extended EOF/Multi-Channel Singular Spectrum Analysis

PCA and its rotated extensions, discussed in the previous sections, maximize the variance of orthogonal components using the information provided by the auto-covariance matrix (Eq. (3.6)). Thus, the assumption behind the conventional PCA is that the space-time field of $\mathbf{X}$ contains stationary patterns in the sense that they are not evolving. When two time series $\mathbf{x}_t$ and $\mathbf{y}_t$, $t = 1, \dots, n$ are stationary, their statistics, such as mean and variance, do not change over time and do not follow any trends. As a result, their auto-covariance with any permutation (denoted by π here) yields the same result, i.e.

$$\mathrm{cov}(\mathbf{x}_t, \mathbf{y}_t) = \mathrm{cov}(\mathbf{x}_{\pi(t)}, \mathbf{y}_{\pi(t)}). \tag{3.30}$$

Time series of climate data, such as GRACE-TWS changes, usually contain periodic patterns such as annual and semi-annual components. It should be mentioned here that referring to the 'periodic' term does not necessarily mean that their frequency always stays the same (Schmidt et al., 2008b). Therefore, incorporating lagged correlations into the decomposition procedure might help in separating periodic components. This can be done based on the Extended Empirical Orthogonal Function (EEOF) method (Weare and Nasstrom, 1982). Broomhead and King (1986a,b) developed an approach similar to EEOF to deal with low order chaotic systems, and called the method Singular System (Spectrum) Analysis (SSA). An extended version of SSA that deals with multivariate time series is called Multi-Channel Singular Spectrum Analysis (MSSA) in Broomhead and King (1986a,b). EEOF and MSSA methods are mathematically equivalent.

In the EEOF approach, the data matrix of Eq. (3.2) at time t, $t = 1, \ldots, n$, is extended to include lagged temporal information as

$$\mathcal{X} = \begin{pmatrix} x_{1,1} & \cdots & x_{1,p} & x_{2,1} & \cdots & x_{2,p} & \cdots & x_{M,1} & \cdots & x_{M,p} \\ x_{2,1} & \cdots & x_{2,p} & x_{3,1} & \cdots & x_{3,p} & \cdots & x_{M+1,1} & \cdots & x_{M+1,p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \cdots & \ddots & \vdots \\ x_{n-M+1,1} & \cdots & x_{n-M+1,p} & x_{n-M+2,1} & \cdots & x_{n-M+2,p} & \cdots & x_{n,1} & \cdots & x_{n,p} \end{pmatrix}, \tag{3.31}$$

where M is the temporal lag. Therefore, the new data matrix $\mathcal{X}$ can be seen as a systematic resample of the original data matrix $\mathbf{X}$. In Eq. 3.31, $\mathcal{X}$ contains $d = Mp$ columns instead of p columns in $\mathbf{X}$. Considering Eq. (3.31), the temporally-lagged auto-covariance matrix can be written as

$$\mathcal{C} = \frac{1}{n - M + 1} \mathcal{X}^T \mathcal{X}. \tag{3.32}$$

One could also compute the lagged auto-covariance based on the grid points. An example can be found in Plaut and Vautard (1994), who formulated the EEOF based on the spatially-lagged covariance matrix. In our case, however, since the number of grid points is usually much bigger than the number of time steps (see the applications in Chapter 5), formulating EEOF based on the grid points can result in unnecessarily large auto-covariance matrices, which can be computationally intensive to diagonalize. Therefore, the formulation of Eq. (3.32) will be considered here.

Similar to Eq. (3.20), $\mathcal{X}$ in Eq. (3.31) can be expanded as

$$\mathcal{X} = \bar{\mathcal{P}} \mathbf{\Theta} \mathcal{E}^T, \tag{3.33}$$

where the $d \times d$ matrix $\mathcal{E}$ is made up by the right singular vectors of $\mathcal{X}$ and contains the EEOFs with the length of $d = Mp$ in its columns. The diagonal matrix $\mathbf{\Theta}$ contains the singular values $\theta_1, \ldots, \theta_d$ of $\mathcal{X}$, and $\bar{\mathcal{P}} = (\bar{\mathcal{P}}_1, \ldots, \bar{\mathcal{P}}_d)$ represents components of the left singular vectors or extended PCs, where $\bar{\mathcal{P}}_i = (\bar{p}_i(1), \ldots, \bar{p}_i(n - M + 1))^T$.

Similar to PCA, the EEOF method can also be used for transforming the original data set or for dimension reduction. For reconstruction of the original data $\mathbf{X}$, using EEOFs, one needs to consider that EEOFs also contain time lagged components. As a result, the reconstruction using $\bar{\mathcal{P}}$ and $\mathcal{E}$ yields the extended matrix $\mathcal{X}$ in Eq. (3.33), instead of the data matrix $\mathbf{X}$. To derive spatial base-functions similar to those of PCA, one possibility is to compute the average of the additional elements as

$$\mathbf{E}_i = \left(\frac{1}{M} \sum_{l=1}^{M} e_{(i+l-1),1}, \ldots, \frac{1}{M} \sum_{l=1}^{M} e_{(i+l-1),p} \right)^T. \tag{3.34}$$

The corresponding temporal evolution is derived by projecting the original data $\mathbf{X}$ onto the base-functions found from Eq. (3.34), using $\mathbf{P}_i = \mathbf{X}\mathbf{E}_i$ (similar to Eq. (3.19)). The formulation offered by Eq. (3.34) can be referred as a 'simple' average of the EEOF decomposition, since it considers only the additional multiple elements of the spatial eigenvectors (EEOFs). For reconstructing the original data, an alternative averaging is introduced e.g., in Golyandina et al. (2001) and Hannachi et al. (2007), which takes advantage of all the multiple values in $\mathcal{X}$. Considering the j retained components, the reconstruction $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,p})$ at time t is written as

$$\mathbf{x}_t^T = \begin{cases} \frac{1}{t}\sum_{l=1}^{t}\sum_{k=1}^{j}\theta_k p_k(t-j+1)\mathcal{E}_k & ; \quad for\ 1 \leq t \leq M-1 \\ \frac{1}{M}\sum_{l=1}^{M}\sum_{k=1}^{j}\theta_k p_k(t-j+1)\mathcal{E}_k & ; \quad for\ M \leq t \leq n-M+1 \\ \frac{1}{n-t+1}\sum_{l=t-n+M}^{M}\sum_{k=1}^{j}\theta_k p_k(t-j+1)\mathcal{E}_k & ; \quad for\ n-M+2 \leq t \leq n, \end{cases} \tag{3.35}$$

where $\mathcal{E}_k$ is the k'th column of $\mathcal{E}$. The EEOF method can be efficient in detecting periodic components (Rangelova et al., 2010). However, the decomposition results of EEOF depend on the proper selection of the lag window M. For instance, Monahan et al. (1999) stated that when the data contains a strong standing wave, choosing a correlation lag M close to the quarter of the wavelength (of the standing wave) leads to a sample lagged auto-correlation of zero or close to zero. As a result, even wrong decomposition results might be derived. The EEOF approach is not used for the applications of this thesis.

3.4 Complex EOF

Complex Empirical Orthogonal Function (CEOF) method is an alternative to the PCA technique, which similar to EEOF/MSSA offers the possibility to extract non-stationary patterns (i.e. patterns that change in space and time). Before describing the CEOF method, the performance of PCA in capturing a propagating structure is described below. To demonstrate this point, consider a data set consisting of five unit-amplitude annual sinusoids with the length of 100 months. This length is selected to be similar to the temporal coverage of a nominal satellite gravimetry mission, e.g., GRACE. For the first experiment, it is assumed that all the five time series are exactly the same and their variance is one. After applying the SVD of Eq. (3.20), as expected, the first eigenvalue is found to be one (for mode 1), but zero for all the other modes. In other words, the data set contains only one mode that represents its total variability. Results of the first experiment are not shown here.

For the second experiment, the first time series stays the same as before, but the second is shifted forward by one month. The third one by another one month, totally two-months, and the fourth and fifth time series are manipulated in the same manner. Figure 3.3 (top) shows the time series of the second experiment. The data matrix $\mathbf{X}$ was constructed by containing the five time series of Fig. 3.3 (top) in its columns. The PCA decomposition of the shifted time series is shown in the graph on the middle of Fig. 3.3. As a result of the phase shift, PCA extracts two standing orthogonal components, which according to Fig. 3.3 (bottom), indicates the maximum correlation when the lag is three months. The two extracted components express the maximum variance of the shifted time series (stored in $\mathbf{X}$). It should be mentioned here that the derived lag of three months is caused by the introduced phase shifts and does not depend on the sampling length.
Using the Complex EOF approach, the simulated signals of Fig. 3.3 (top) can be extracted in one complex mode. The real and imaginary parts of the first dominant component are shown on top and middle of Fig. 3.4, respectively. The relationship between the phases of the extracted complex components is shown in the bottom graph. As a result, one can see that, using the complex EOF, the introduced phase differences between the five simulated time series is well captured, whereas

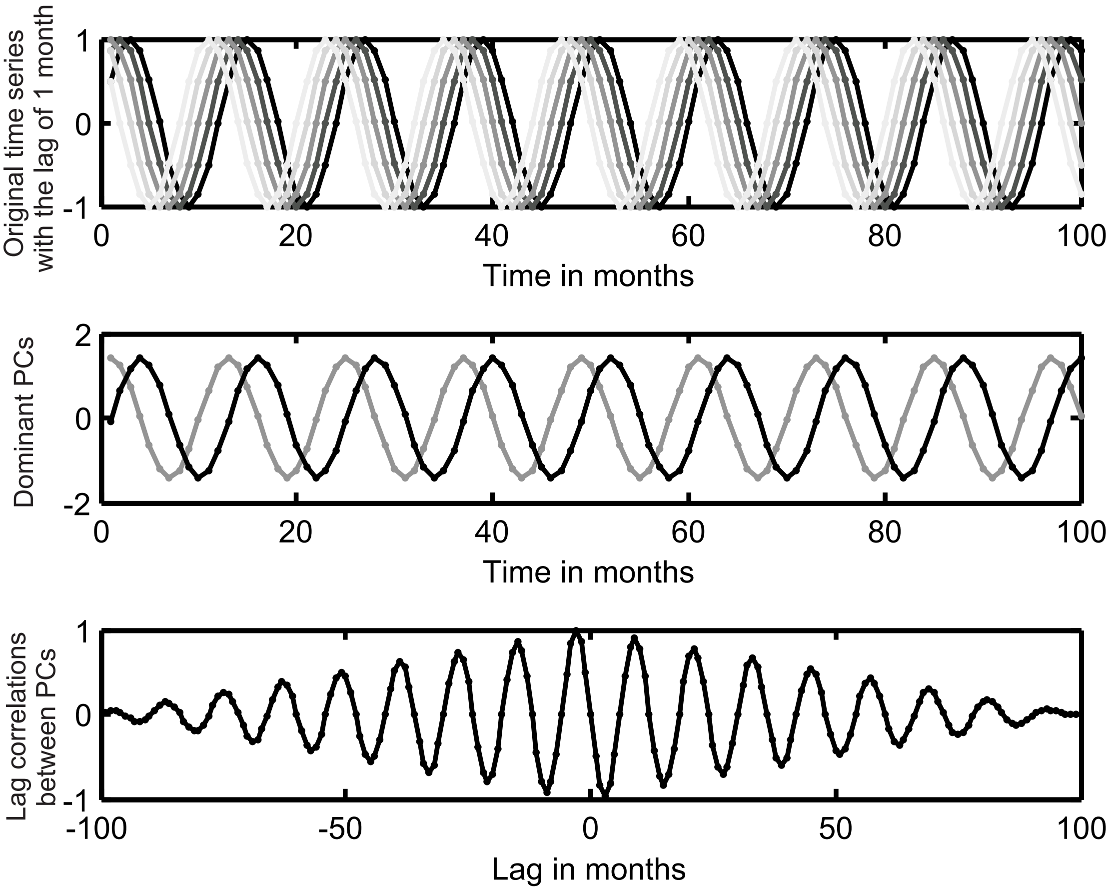


Figure 3.3: Performance of PCA in decomposing propagating waves. The graph on top shows five annual signals, with a lag of one month with respect to each other. The middle graph shows the first two dominant PCA-derived orthogonal components (PCs). The lag-correlations between the two PCs in the middle graph (PC1 and PC2) are shown in the bottom graph.

a temporal phase of 30° (that corresponds to one month delay) is found between the time series (columns of $\mathbf{X}$). Horel (1984) provided various examples that represent the performance of PCA and CEOF for decomposing propagating signals, similar to the example in Fig. 3.3. In Section 5.1, the performance of CEOF to decompose GRACE-like simulated TWS changes is demonstrated. There are, in principle, two ways to perform CEOF in time domain, known as the 'conventional' CEOF (C-CEOF) and the 'Hilbert' CEOF (H-CEOF), which will be discussed in the following.

Conventional-Complex EOF (C-CEOF)

Conventional-Complex EOF (C-CEOF) is usually applied to a complex field obtained from a pair of associated variables. Examples of its applications can be found e.g., in von Storch and Zwiers (1999). Yet, C-CEOF can be applied to a single field of $\mathbf{X}$. Before applying C-CEOF to $\mathbf{X}$, it has to be 'complexified' as

$$\mathbf{Y} = \mathbf{X}_t + i\,\mathbf{X}_{t+\tau}, \tag{3.36}$$

where τ is an arbitrary lag time. In Eq. (3.36), the entry of the new data matrix at each grid point s_l, $l = 1, \dots, p$ at time t is given by $y_{t,l} = x_{t,l} + i\ x_{t+\tau,l}$. The complex auto-covariance matrix is defined by

$$\mathbf{C} = \frac{1}{n}\mathbf{Y}^{*T}\mathbf{Y}, \tag{3.37}$$

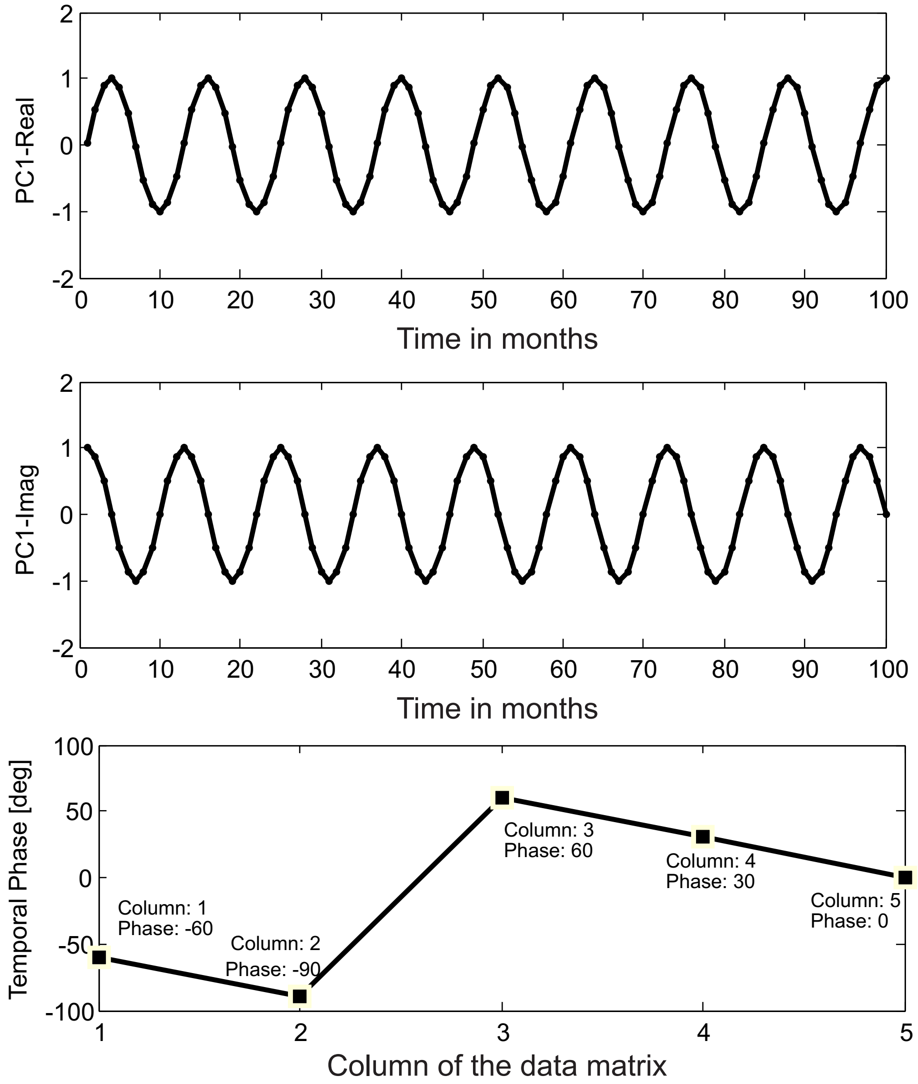


Figure 3.4: Performance of Complex EOF for decomposing the example of Fig. 3.3. The graphs on top and middle respectively show the real and imaginary part of the first dominant component, which corresponds to an annual cycle. The bottom graph shows the phase (in degree) that corresponds to each column.

where (*) is the complex conjugate operator. The auto-covariance matrix in Eq. (3.37) is also written as $\mathbf{C} = \frac{1}{n}\mathbf{Y}^H\mathbf{Y}$, where H is known as the Hermitian transpose (e.g., Horel, 1984). The extended form of $\mathbf{C}$ is written as $\mathbf{C} = \frac{1}{n}\big(\mathbf{X}^T\mathbf{X} + \mathbf{X}_{t+\tau}^T\mathbf{X}_{t+\tau} + i(\mathbf{X}^T\mathbf{X}_{t+\tau} - \mathbf{X}_{t+\tau}^T\mathbf{X})\big)$.

Similar to EEOF/MSSA, Eq. (3.36) uses temporally lagged information, but this time to define the complex field $\mathbf{Y}$. Using the eigenvalue decomposition of Eq. (3.37), one can compute complex EOFs and PCs, which respectively represent the dominant variability of observations and their spatial and temporal phase propagations. As was seen in Eq. (3.36), the entries of $\mathbf{Y}$ depend on the choice of τ, which reflects the temporal characteristic of the propagating features. This field is also used in the C-CEOF approach to compute the auto-covariance matrix, as well as its eigenvalue decomposition (decomposing Eq. (3.37) using Eq. (3.13)). Therefore, a successful C-CEOF decomposition strongly depends on a proper choice of τ. For example, if the observed time series are monthly sampled and exhibit a strong annual variability, a lag of $\sim \tau = 12$ would be a proper choice to extract such periodic pattern. This selection, however, requires a prior knowledge about the variability of the observed time series $\mathbf{X}$. To avoid this difficulty, the Hilbert transform has been introduced, which is described below.

Hilbert-Complex EOF (H-CEOF)

The Hilbert-Complex EOF (H-CEOF), described here, is equivalent to the one introduced by Rasmusson et al. (1981), which unlike the C-CEOF works based on phase shift in the frequency domain, introduced by the Hilbert transformation. Assume the columns of the data matrix $\mathbf{X}$ (shown by $\mathbf{x}_t$) to be a scalar field with a discrete Fourier representation of

$$\mathbf{x}_t = \sum_{\omega_k} \mathbf{a}(\omega_k)\cos(\omega_k t) + \mathbf{b}(\omega_k)\sin(\omega_k t),\ t = 1, \dots, n, \tag{3.38}$$

where $\mathbf{a}(\omega_k)$ and $\mathbf{b}(\omega_k)$ are vector Fourier coefficients at frequency ω_k (with accepting values between $-\pi$ and π that are selected according to the sampling rate). Thus, values of ω_k have to satisfy the Nyquist frequency rule and its equivalent wavelength should not exceed the length of time series (Chatfield, 1989, chapter 7). The Hilbert transform ($\mathcal{H}$) of $\mathbf{x}_t$ is given by

$$\mathcal{H}(\mathbf{x}_t) = \sum_{\omega_k} \mathbf{b}(\omega_k)\cos(\omega_k t) - \mathbf{a}(\omega_k)\sin(\omega_k t),\ t = 1, \dots, n. \tag{3.39}$$

Each complexified time series $\mathbf{y}_t$ is derived from multiplying Eq. (3.39) by $i = \sqrt{-1}$, and adding it to Eq. (3.38), which yields

$$\mathbf{y}_t = \mathbf{x}_t + i\,\mathcal{H}(\mathbf{x}_t). \tag{3.40}$$

The Hilbert transform of Eq. (3.39) is related to the first time derivative of Eq. (3.38). Therefore, the complex field of Eq. (3.40) contains information on the time series as well as their rate of change. The auto-covariance matrix, required for performing the H-CEOF decomposition, can be computed from Eq. (3.37), or equivalently as

$$\mathbf{C} = \frac{1}{n}\big(\mathbf{X}^T\mathbf{X} + \mathcal{H}(\mathbf{X})^T\mathcal{H}(\mathbf{X}) + i(\mathbf{X}^T\mathcal{H}(\mathbf{X}) - \mathcal{H}(\mathbf{X})^T\mathbf{X})\big), \tag{3.41}$$

which contains information on the cross-spectral values, averaged over all frequencies ($-\pi < w_k < \pi$). Therefore, its decomposition leads to an average depiction of the propagating disturbances present in the original data matrix $\mathbf{X}$. It should be mentioned here that, in case, a priori knowledge on the spectral frequency range of a certain pattern exists, then it is better to accordingly filter the original data and its Hilbert transform before performing the CEOF analysis. This can be done by applying a band-limited filter (centered around the known frequency) to the data and its Hilbert transformation. Such pre-filtering will enhance extraction of the pattern of interest (Horel, 1984).

The auto-covariance matrix in Eq. (3.41) is Hermitian, i.e. $\mathbf{C}$ is equal to its own conjugate transpose, and is therefore diagonalizable. After applying Eq. (3.13) to the matrix $\mathbf{C}$, a matrix that contains complex eigenvectors $\mathbf{E}$, and $\bar{\mathbf{P}}$ that stores complex temporal patterns, as well as a set of real singular values (stored in the main diagonal of $\mathbf{\Lambda}$) can be computed. The complex field ($\mathbf{Y}$) can be written as $\mathbf{Y} = \bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T = \mathbf{P}\mathbf{E}^T$, $\mathbf{P} = \bar{\mathbf{P}}\mathbf{\Lambda}$. Since the columns of $\mathbf{E}$ and $\bar{\mathbf{P}}$ contain complex values, the following transformations need to be applied. The spatial amplitude is derived as

$$\mathbf{u}_l = \left(\mathbf{e}_l \bullet \mathbf{e}_l^*\right)^{1/2}, \tag{3.42}$$

where $l \in \{1, \ldots, \min(n,p)\}$ (n and p are dimensions of $\mathbf{X}$) and '$\bullet$' is a vector valued operator that provides an element-by-element multiplication product of the columns. The spatial phase values are computed from

$$\boldsymbol{\Omega}_l = \arctan\left[\frac{\mathrm{Im}(\mathbf{e}_l)}{\mathrm{Re}(\mathbf{e}_l)}\right], \tag{3.43}$$

where 'arctan(.)' in Eq. (3.42) is an operator that performs inverse tangent ($\tan^{-1}$), Im(.) and Re(.) respectively extract the imaginary and real parts of the complex entries, and the fraction has to be implemented element-by-element, thus $\boldsymbol{\Omega}_l$ represents a vector. Similarly, the temporal amplitude and phase values are respectively derived from

$$\mathbf{z}_l = \left(\mathbf{p}_l \bullet \mathbf{p}_l^*\right)^{1/2}, \tag{3.44}$$

and

$$\boldsymbol{\Phi}_l = \arctan\left[\frac{\mathrm{Im}(\mathbf{p}_l)}{\mathrm{Re}(\mathbf{p}_l)}\right]. \tag{3.45}$$

Equation (3.42) represents spatial amplitude of the l'th dominant mode, and Eq. (3.43) indicates its corresponding spatial phase propagation. Similarly, the temporal amplitude of the l'th dominant mode is derived in Eq. (3.44), and Eq. (3.45) provides information about the corresponding temporal phase. For data matrices that include time series with various scales of variability, the spatial phase plot can be very difficult to interpret. Similarly, the interpretation of temporal phase would be difficult when the sampled time series consist of several cyclic patterns with the same range of amplitudes.

Each mode of variability, derived from the H-CEOF decomposition can be estimated in a similar manner to the PCA decomposition, i.e. $\mathrm{Re}\big(\mathbf{p}_l\, \mathbf{e}_l^T\big)$, $l < \min(n,p)$. The original matrix $\mathbf{X}(n \times p)$ can be reconstructed using the CEOF approach as

$$\mathbf{X}_j = \mathrm{Re}\big(\mathbf{P}_j \mathbf{E}_j^T\big), \tag{3.46}$$

where $(p \times j)$ $\mathbf{E}_j$ and $(n \times j)$ $\mathbf{P}_j$ store the first j dominant complex components, derived by singular value decomposition of Eq. (3.41). Since $\mathbf{P}_j$ and $\mathbf{E}_j$ contain complex entries, their multiplication product will also contain complex values. As a result, the Re(.) operator has been used in Eq. (3.46) to extract the real part of $\mathbf{P}_j\mathbf{E}_j^T$, which is considered as a reconstruction of the original data matrix $\mathbf{X}$. In order to extract non-stationary patterns from GRACE-TWS changes (Sections 5.1.3 and 5.2), the H-CEOF technique is preferred over the C-CEOF and MSSA methods. This is motivated from the fact that H-CEOF incorporates information about the non-stationary variability of the observations without requiring any priory assumptions about the periodical characteristics of them. In Section 4.5.3, the H-CEOF decomposition is extended to the ICA technique.

3.5 Statistical Tests for the Significance of Modes

PCA and its extensions, described in the previous sections, provide a tool to extract dominant patterns from spatio-temporal fields. The first $1, \ldots, j$ dominant modes (e.g., in Eqs. (3.23), (3.25), (3.35), and (3.46)) can be used to reduce the large number of dimensions in the original data matrix $\mathbf{X}$, since j is usually considerably smaller than n and p (dimension of $\mathbf{X}$). For instance in the PCA case, one could compare Eq. (3.20) with Eq. (3.23) assuming that EOFs and PCs are ordered with respect to the magnitude of the variance they represent. The data matrix **X** $(n \times p)$ therefore can be written as

$$\mathbf{X} = \bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{E}_j^T + \bar{\mathbf{P}}_{\min(n,p)-j} \mathbf{\Lambda}_{\min(n,p)-j} \mathbf{E}^T_{\min(n,p)-j},\ j < \min(n,p), \tag{3.47}$$

where $\bar{\mathbf{P}}_j$ $(n \times j)$, $\mathbf{\Lambda}_j$ is diagonal $(j \times j)$, and $\mathbf{E}_j$ has p rows and j columns. On this basis, the first j orthogonal components represent the dominant variance of **X** and are counted as the 'signal' of interest and the other $\min(n,p) - j$ components represent the 'noise'. Therefore, a proper j should be defined, using a rigorous decision rule, to choose a subset of EOFs and PCs that confidentially approximate the original data **X** (Jolliffe, 1989). The selection of j also has an impact on the results of the rotated PCA methods, e.g., those of Section 3.2, as well as on the ICA method in Chapter 4.

Preisendorfer (1988) categorizes the decision rules into the 'dominant-variance', 'time-history', and 'space-map' approaches. The dominant-variance approach provides a tool to select the j largest eigenvalues and their corresponding EOFs and PCs that retain a dominant portion of the total energy in **X**. This selection approach is, therefore, follows the concept in Eq. (3.47) and is more relevant to the applications that are presented in this thesis. Using the time-history rules, one usually examines the noise behavior of temporal components, therefore, those temporal components that are normally distributed will not be used for reconstructing the data set **X**. In Section 3.5.2, normality tests are introduced, which can be used as the time-history rule to decide whether a temporal component follows a normal distribution or not. The space-map rules use the same concept as the time-history rules, but with the statistical testing of spatial components.

In this section, first some selected dominant-variance rules are introduced, which will be used later in Chapters 4 and 5. Some possible tests will also be discussed to address the time-history rules. The space-map rules are not covered in this thesis since justifying the spatial behavior of the time-variable TWS changes (the main parameter of interest in this thesis) is still unknown. This means that it is extremely difficult to design a statistical test to decide whether an extracted spatial pattern represent a possible variability in TWS or not. More details about decision rules can be found in (Preisendorfer, 1988, chapter 5).

3.5.1 Dominant-variance Rules

The dominant-variance rules originated from the idea that the existing observations (stored in the data matrix **X**) are produced by a physical process. For instance, TWS observations derived from satellite gravimetry missions are related to the water cycle, which is caused by various physical processes. The gathered observations are also usually contaminated with noise, whereas the noise can, for example, originate from the measuring instrument, and/or is caused by environmental factors. By applying Eq. (3.12) to the auto-covariance matrix of **X**, its eigenvalue spectrum can be computed. In this thesis, the eigenvalue spectrum refers to the plot of either singular values or eigenvalues when they are ordered with respect to their magnitude. Since **X** consists of observations that contain a superposition of signals, i.e. here the physical process and noise, one should detect a break in spectrum, where the eigenvalues of the process (related to the signal of interest) leave off and those of pure noise begin.

In order to illustrate this, simplified TWS time series were simulated by extracting annual and semi-annual cycles from $1° \times 1°$ monthly products of the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) covering the period of January 2003 to December 2012. Extraction of the cycles was done using a least squares adjustment. A random noise with the magnitude of 3 cm was also added to the time series. The RMS of the simulated water storage time series are shown in Fig. 3.5 (left). The first 10 singular values (computed using Eq. (3.12)) are shown in Fig. 3.5 (right). The first two singular values correspond to the simulated annual cycle, while the singular values corresponding to the semi-annual cycle are the third and fourth values. This can be seen by plotting the corresponding EOFs and PCs, which are however not shown here. The break, discussed above, is seen in the fifth singular value, which means that the rest of the spectrum corresponds to the simulated noise. A break in the spectrum may be detected visually if the discontinuity is large enough (similar to that of Fig. 3.5 (right)). Examples might be found in which the eigenvalue spectrum does not contain apparent breaks. For instance, when the magnitude of noise is close to that of signal. One can also use an objective statistical procedure to detect the 'break'. Following this section, this is discussed in the context of 'selection rules'.

3.5.1.1 Simple Rules

Three common rules built along the magnitude of estimated eigenvalues are described in the following.

Scree Plot

A Scree Plot (or eigenvalue spectrum in this thesis) refers to a plot of singular values against their order (Cattell, 1966). Smaller singular values, representing random variations, tend to lie along a straight line. Therefore, those singular values above the straight line are most likely related to the signal. One can decide on the number of significant modes by visually detecting the break, i.e. located in the fifth order of Fig. 3.5 (right). The interpretation of Scree Plot is, however, sometimes complicated due to either lack of any obvious break in the plot or the possibility of existing multiple breaks. For instance, the correct order of the break in Fig. 3.5 (right) is five. However, another break also exists in its second order, which in a real case, might be misleading.

Kaiser-Guttman

This rule is based on the idea that the retained singular values should be greater than their mean. With this assumption, the selected components most likely represent the dominant variance of the original data. Although the rationale for excluding less dominant components makes sense, problems exist regarding the singular values that are only slightly greater or smaller than the average value. Therefore, the Bootstrap resampling technique (Efron, 1979) was applied, e.g., in Lambert et al. (1990) to compute an uncertainty bound for singular values. As a result, the determination of significant components could be applied with respect to a confidence limit. The Kaiser-Guttmann rule was applied to the example in Fig. 3.5 (right). Uncertainty of the mean value was computed following Lambert et al. (1990). As a result, only the first three singular values were found to be significant. Thus, the results indicate that the approach was unsuccessful in detecting the correct number of significant modes of simulated TWS time series in Fig. 3.5.

Variance Portion

Singular values computed from the PCA/EOF approach (or equivalently from SVD) represent the amount of information captured by orthogonal components. It has been shown in Eq. (3.16) that each eigenvalue represents a fraction of total variance of the data. Therefore, the first j dominant components can be selected in a way that the following variance criterion is satisfied

$$\frac{\sum_{i=1}^{j} \lambda_i^2}{\mathrm{trace}(\Lambda^2)} > \varepsilon, \tag{3.48}$$

where ε is an arbitrary variance fraction. Variance percentages of the first 10 dominant PCA components of Fig 3.5 are shown in Fig 3.6. Thus, to retain all the significant components of the simulation example, ε in Eq. (3.48) must be selected to be ~ 0.935. By selecting smaller ε, some useful information, contained in the observations, will be lost. For instance, by selecting $\varepsilon = 0.9$, only the first two singular values corresponding to the annual cycle will be selected. Thus, the semi-annual variability of $\mathbf{X}$ in Fig. 3.5 will be ignored in Eq. (3.47) during data reconstruction.

3.5.1.2 Considering Sampling Errors

In fact, computation of singular values in Eq. (3.12) involves sampling errors due to limited time span of observations, as well as the presence of noise in the data itself. Both errors were not considered for the rules described above (Section 3.5.1.1). Following this section, after discussing the sampling errors, a statistical test based on Monte Carlo sampling will be introduced. More examples of selection rules are discussed in Preisendorfer et al. (1981).

On Perturbation Theory

Methods for estimating sampling errors are usually based on the perception that the data $\mathbf{X}$ represent independent realizations or samples of a random field with unknown stochastic moments. From these realizations, one approximates the true auto-covariance $\mathbf{C}$ only up to an error that depends on the number of data realizations (North et al., 1982). Thus, following Eq. (3.13), the diagonal matrix Λ^2 containing the true eigenvalues can be derived after decomposing the true auto-covariance matrix $\mathbf{C}$ as

$$\mathbf{\Lambda}^2 = \mathbf{E}^T \mathbf{C} \mathbf{E}. \tag{3.49}$$

In reality, neither $\mathbf{\Lambda}^2$ nor $\mathbf{E}$ of Eq. (3.49) are precisely known, since the sampled auto-covariance matrix is of finite length. Therefore, applying Eq. (3.49) on the finite samples (denoted by $\hat{}$), one obtains

$$\hat{\mathbf{L}}^2 = \mathbf{E}^T \hat{\mathbf{C}} \mathbf{E}. \tag{3.50}$$

In Eq. (3.50), the eigenvalues of $\hat{\mathbf{L}}^2$ (that correspond to $\hat{\mathbf{C}}$) are expressed in the basis of the true eigenvectors $\mathbf{E}$. Thus, the matrix $\hat{\mathbf{L}}^2$ is no longer exactly diagonal, but it has the same eigenvalues as $\hat{\mathbf{C}}$.

To better understand the perturbation concept, one can assign a geometrical interpretation to eigenvectors. All EOFs (columns of $\mathbf{E}$, i.e. $\mathbf{e_i}$, $i = 1, \ldots, n$, with n being the rank of $\mathbf{C}$) are the true directions (orthogonal coordinates) that span the observation space $\mathbf{X}$. Each EOF ($\hat{\mathbf{e}}_\mathbf{i}$) of a finite sampled data set is located at an n-dimensional angle α from its true direction $\mathbf{e}$. The sampling error of $\hat{\mathbf{e}}$ can be expressed as a linear combination of all the other true EOFs $\mathbf{e}_i$, since columns of $\mathbf{E}$ span the observation space. Estimating the sampling error, therefore, is equivalent to the solution of a perturbation problem (Wilkinson, 1965). Thus, $\hat{\mathbf{L}}^2$ can be interpreted as a perturbed form of the true eigenvalues $\mathbf{\Lambda}$ ($\hat{\mathbf{L}}^2 = \mathbf{\Lambda}^2 + \epsilon\, \mathbf{N}$). When the number of samples is large enough, i.e. $\epsilon = 1/\sqrt{n} \ll 1$, with n being the length of time series, Bretherton et al. (1999) showed that the elements of $\mathbf{N}$ can be expressed in terms of the true eigenvalues λ_i^2, and uncorrelated perturbations $\omega_{i,k}$. The distribution of $\omega_{i,k}$ is unit random normal ($E(\omega) = 0, \mathrm{var}(\omega) = 1$) since each perturbation has been generated by a sum of n independent variables (caused by the n-dimensional angle between $\mathbf{e}$ and $\hat{\mathbf{e}}$). Thus, by the central limit theorem, the sum of all the differences, when n is large enough, must be approximately normal distributed. Its unit variance is due to the scaling of the auto-covariance matrix. According to Bretherton et al. (1999), $\mathbf{N}$ is written as

$$\mathbf{N}_{i,k} = \begin{cases} \sqrt{2}\lambda_i^2\omega_{i,k} & i = k \\ \lambda_i\lambda_k\omega_{i,k} & i \neq k. \end{cases} \tag{3.51}$$

Wilkinson (1965) showed that the first-order perturbation to the i^{th} sampled eigenvector ($\hat{\mathbf{e}}_i$) is derived from a summation of all errors in other (n-1) eigenvectors as

$$\hat{\mathbf{e}}_{i1} = \sum_{i\neq k} \frac{\mathbf{N}_{ik}\mathbf{e}_k}{\lambda_i^2 - \lambda_k^2}. \tag{3.52}$$

The perturbed eigenvectors can be derived by adding the estimated error of Eq. (3.52) to the true eigenvector as:

$$\hat{\mathbf{e}}_i = \mathbf{e}_i + \hat{\mathbf{e}}_{i1} = \mathbf{e}_i + \sum_{i\neq k} \epsilon\omega_{i,k}\frac{\lambda_i\ \lambda_k}{\lambda_i^2 - \lambda_k^2}\mathbf{e}_k, \tag{3.53}$$

where ϵ is the perturbation parameter. The first-order estimation of the perturbed eigenvalue is derived as

$$\hat{\lambda}_i^2 \simeq \lambda_i^2(1 + \sqrt{2}\epsilon\omega_{i,k}), \tag{3.54}$$

where $\hat{\lambda}_i^2$ is the i^{th} diagonal element of $\hat{\mathbf{L}}$ (Bretherton et al., 1999).

North's Rule

North et al. (1982) also derived an explicit form for the first-order errors in the eigenvalues of a sampled auto-covariance matrix. Considering the eigenvalues and eigenvectors of the sampled auto-covariance matrix as being stochastic, North et al. (1982) assessed the sampling error of two numerically close ('neighboring') eigenvalues (e.g., λ_i^2 and λ_k^2, $i > k$), and estimated it as

$$\delta\lambda_i^2 = \sqrt{\frac{2}{n^*}}\lambda_i^2, \tag{3.55}$$

where n^* is the number of independent observations in the sample, also known as the effective sample size (Trenberth, 1984). An estimation of n^* for the 95% level of confidence is given by Thiiébaux and Zwiers (1984) as

$$n^* = n\big(1 + 2\sum_{k=1}^{n-1}(1 - \frac{k}{n})\rho(k)\big), \tag{3.56}$$

where $\rho(k)$ is the auto-correlation value between samples. The sampling error of eigenvectors is derived as

$$\delta\mathbf{e}_i = \frac{\delta\lambda_i^2}{\lambda_k^2 - \lambda_i^2}\mathbf{e}_k, \tag{3.57}$$

where λ_k^2 and λ_i^2 are two numerically close eigenvalues, and $\mathbf{e}_k$ is the corresponding eigenvector of λ_k^2. The antecedent in Eq. (3.57) is derived from Eq. (3.55). One can see that if the sampling error of Eq. (3.55) is comparable to the difference of λ_i^2 and λ_k^2, then $\frac{\delta\lambda_i^2}{\lambda_k^2-\lambda_i^2}$ becomes one. As a result, the typical error of the corresponding EOF (Eq. (3.57)) will be of the size of $\mathbf{e}_k$ itself. One might disregard such modes when reconstructing the data. This means that one should define j in a way that

$$\delta\lambda_i^2 < |\lambda_i^2 - \lambda_k^2| = \min_{i\neq k}|\lambda_i^2 - \lambda_k^2|. \tag{3.58}$$

It should be mentioned here that Eq. (3.54) is equivalent to the criterion of Eq. (3.55) in North et al. (1982).

Rules Based on Monte Carlo

As clarified above, the investigation of sampling error, based on the degeneracy of the auto-covariance matrix spectrum aims at deriving a measure of uncertainty for each eigenvalue. This is quite difficult to get since such estimation requires several prior assumptions on the distribution and the length of observations. In practice, there exist two ways to compute the uncertainty of the eigenvalues. The first one is based on asymptotic results as was seen in the North et al. (1982)'s rule of thumb (Eq. (3.58)). The other alternative is using Monte Carlo simulations to assess uncertainty on the spectrum of the covariance matrix (e.g., Preisendorfer et al., 1981).

In this thesis, finding the number of statistically significant components follows an approach similar to the Rule 'N' in (Preisendorfer et al., 1981, pages 199 to 205), that involves a Monte Carlo approach to simulate sampling data Ψ from a random distribution $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$, with $\mathbf{\Sigma}$ containing the column variances of $\mathbf{X}_{n\times p}$. Thus, var($\boldsymbol{\Psi}$)= diag(σ_i^2)$\mathbf{I}_p$, where var(.) is a variance operator, diag$(\sigma_i), i = 1, \ldots, p$ stores the column-wise variances of $\mathbf{X}_{n\times p}$ in its diagonal elements, and $\mathbf{I}_p$ is an identity matrix of dimension p. The null hypothesis is that $\mathbf{X}$ is drawn from such a distribution. Therefore, in Rule N, the assumption is that the original time series consist of noise, which are uncorrelated from each grid point to another. To apply this rule, a number of realizations of $\mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$ are generated and the eigenvalues computed and placed in decreasing order. Then, the 5'th and 95'th percentile are taken as the significant boundaries corresponding to 5% and 95% of confidence level, respectively. Eigenvalues from the actual data sets that fall above the derived confidence boundaries are deemed as unlikely to result from a data set consisting of only noise. Therefore, they correspond to statistically significant components, which can be used to reconstruct the signal using Eq. (3.23).

The Monte Carlo approach, with 100 realization, was applied to the simulation example of Fig. 3.5, and the corresponding results of 5% and 95% confidence level were plotted in Fig. 3.5 (right). As a result, the computed significance levels could be used to correctly identify the 4 significant eigenvalues. More examples regarding the application of the Monte Carlo rule are presented in Chapter 5.

3.5.2 Time-history Tests

Time-history tests can be used for quantifying the temporal behavior of the components derived from the methods in the current section. The introduced approaches below include some selected normality tests . The motivation of using the normality tests is due to the fact that the temporally normal components might be most likely related to noise rather than signal. Therefore, such components should not be included in the reconstruction of the original data set in Eq. (3.23).

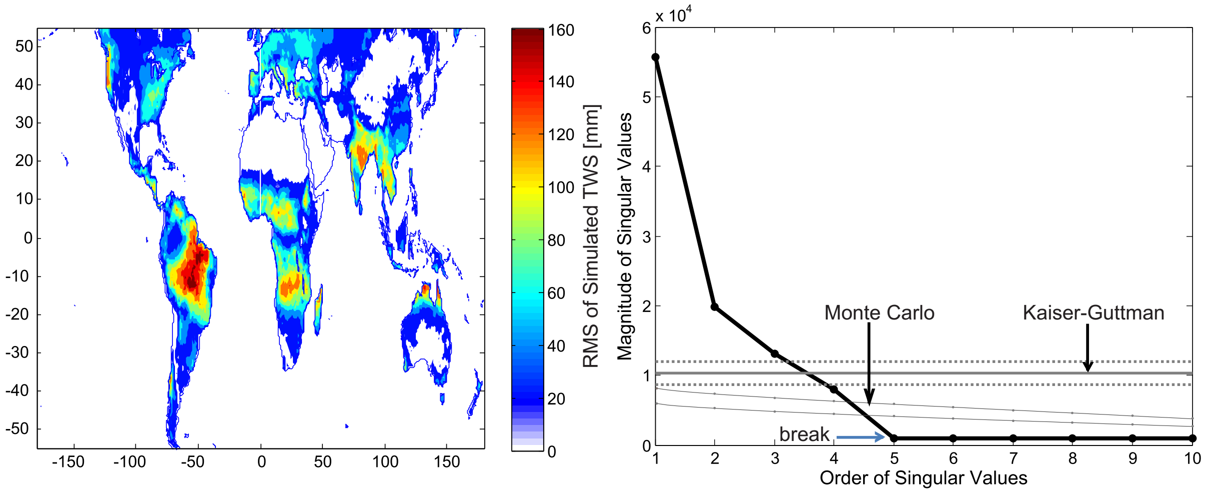


Figure 3.5: Eigenvalue spectrum of simulated TWS time series. Figure 3.5 (left) represents the RMS of the annual and semi-annual cycles of TWS, derived from monthly GLDAS/Noah hydrological models (Rodell et al., 2004) covering the period of January 2003 to December 2012. Before computing the RMS, a random noise with the magnitude of 3 cm was also added to the extracted cycles. Figure 3.5 (left) shows the first 10 dominant singular values (computed using Eq. (3.12)) against their orders. A 'break' of the spectrum is detected in the fifth order. The result of the Kaiser-Guttman rule (Section 3.5.1.1) and its uncertainty (dashed lines), as well as the confidence levels of the Monte Carlo rule (solid gray lines) are also shown in the plot.

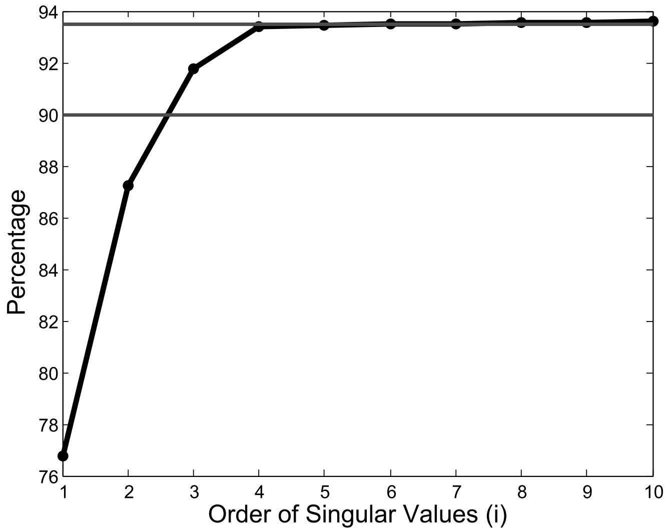


Figure 3.6: Variance percentages corresponding to the singular values that are shown in Fig. 3.5 (right). X-axis represents the order of singular values (i in Eq. (3.48)) and y-axis indicates their corresponding variance percentages. The horizontal lines indicate ε of 0.9 and 0.935 in Eq. (3.48). The results show that for retaining the simulated signal of Fig. 3.5 (left), one needs to select ε being ~ 0.93.

Kolmogorov-Smirnov (KS) Test

The Kolmogorov-Smirnov (KS) test is a common goodness of fit approach (Smirnov, 1948), which compares a sample with a reference probability distribution (one-sample KS). KS can also be applied to compare two samples (two-sample KS). The KS test is frequently used in climate research (von Storch and Zwiers, 1999) to examine the temporal behavior of samples. In Wouters and Schrama (2007), KS was used to test the noise behaviors of the principal components derived from time series of GRACE level 2 coefficients. Basically, the test quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference distribution, or between the empirical distribution functions of two samples. The null hypothesis of KS is that the samples are drawn from the same distribution. Assume that the function $F(t)$ is a given cumulative distribution and $S(t)$ is the empirical cumulative distribution of the sample, KS is written as

$$D = \sup_{t} |F(t) - S(t)|, \tag{3.59}$$

where $\sup_{t}(.)$ is the supremum of the set of differences and t is the time argument. One can assume here that $F(t)$ is a normal distribution with known mean and standard deviation. In Eq. (3.59), $S(t)$ has to be derived by estimating the mean and standard deviation of the sample. If D exceeds the $1 - \alpha$ quantile as given by the KS test statistics, then the null hypothesis is rejected, which means that the distribution of the sample is not equal to the hypothesized distribution. The KS test is, therefore, appropriate when the parameters of the hypothesized distribution are well known. In practice, however, the statistic requires a relatively large number of data points to properly test the null hypothesis in the KS test.

Alternative Normality Tests

To improve the performance of the KS test, Lilliefors (1967) used the same statistic test as the KS test but applied adjusted critical values for the test hypothesis. A table of critical values for the Lilliefors can be obtained by Monte Carlo approximation as in Conover (1999). The values of the critical values in the Lilliefors table are smaller than those of the KS, since the test also considers the sampling error of the hypothesized distribution. The Lilliefors test is, however, very time consuming if one uses the Monte Carlo approach to generate the exact critical values. Therefore, analytical approximations were introduced, e.g., in Stephens (1974) to estimate the required critical values.

There are also several alternative tests for normality, including: the Jarque-Bera, Shapiro-Wilk, Anderson-Darling, Cramer-von Mises, Pearson chi-square, and Shapiro-Francio tests (Thode, 2002). The alternative tests, specially that of Shapiro-Wilk, represent better skills to decide about the normal behavior of samples with limited number n (for example $n < 100$).

3.6 Error Estimation of Reconstruction

A proper approach to estimate the uncertainty of an experiment can be realized by repeating it under relevant conditions. Therefore, using a resampling approach, one essentially attempts to simulate this time-consuming process through numerical experiments. Popular methods are the Jackknife and the Bootstrap (Efron, 1979). As a result of resampling, the determination of significant components could be done with respect to a confidence limit. It is important to note that the Bootstrap can be performed in two modes, namely non-parametric and parametric, which are also known as unconditional and conditional, respectively. In the non-parametric mode, one draws rows of the data matrix $\mathbf{X}$ with replacement, whereas in the parametric mode, one constructs a (parametric) model and resamples the residuals instead. The term ‘conditional’, thus, refers to the residuals that are calculated conditionally to a predefined model. The Jackknife approach is similar to the non-parametric Bootstrap (Efron, 1979).

Once the proper number of retaining modes (j in Eq. (3.23)) was defined from the methods in Section 3.5.1 or with discarding noisy components as in Section 3.5.2, to estimate the uncertainties of the eigenvalues themselves, one can randomly select a subsample of the data sets and apply PCA, then select another subsample and repeat this operation several times. This yields various realizations of the eigenvalues of the original data matrix, from which the uncertainty can be estimated (see e.g., uncertainty results in Chapter 5).

It is worth mentioning here that existing missing values in the data is also a concern in the application of decomposition techniques to extract information from geophysical time series. Thus, this issue must be considered when applying the introduced techniques of this chapter and the ICA technique of Chapter 4. In general, missing data can be caused due to misfunctions of instruments, or due to the application of a processing step such as an outliers detection procedure. A possible method for dealing with missing data is to decompose the auto-covariance matrix that is calculated using only the available data (considering zero for missing values, von Storch and Zwiers, 1999). This approach, however, does not necessarily lead to a semi-positive defined auto-covariance matrix and consequently its corresponding EOFs are over-estimated.

An alternative approach to deal with missing points can be realized by filling the gaps in the data sets and determining the structure and number of significant EOFs. Basically, the dominant EOFs represent a large portion of variance in the data and should not be influenced by local changes in the values of a few points. Thus, one could calculate the EOFs based on the data matrix in which a first rough estimate of the missing values are used to fill the gaps (e.g., using an interpolation technique to provide a first guess of the missing value). Once the dominant EOFs and their amplitudes are estimated, they can serve to calculate the value of the field at the missing points. Then, this process can be repeated by reevaluating the field and estimating new EOFs until the procedure convergences (Beckers and Rixen, 2003).

4. Higher Order Statistical Signal Decomposition

Various second order methods including PCA/EOF, REOF, MSSA/EEOF, and CEOF, which can be used to decompose time series of observations were discussed in Chapter 3. From a statistical point of view, however, all the mentioned methods use up to the second order statistical information, contained in the auto-covariance or correlation matrices that are built based on the anomaly time series (mean removed), during the decomposition procedure. Therefore, they disregard a large part of information provided in higher order moments of the probability distribution function (PDF) of the observed signal (Jolliffe, 1986). Generally, there is no higher than two statistical information to be explored whenever the PDF of the observed variables is Gaussian (Hyvärinen, 1999a), but in geophysical time series (e.g., GRACE-derived mass changes), more often than not, the PDF is non-Gaussian (Aires et al., 2002, Beven, 2001, Forootan and Kusche, 2012, Westra et al., 2007). Therefore, covariances or correlations will not be sufficient as a measure of statistical dependence between signals found in non-Gaussian data.

In order to incorporate more information from the PDF underlying the data, Cardoso and Souloumiac (1993) and Hyvärinen (1999a,b) suggested to involve higher order statistical moments in the decomposition procedure, leading to what is otherwise known as Independent Component Analysis (ICA). ICA has been considered as an extension of the second order methods (e.g., PCA) to be used as an exploratory tool for climate data analysis. Application of ICA is motivated from the fact that the objective of PCA, i.e. to maximize the variance explained by each component in succession, usually clusters different physical modes within a single extracted 'mathematical' mode and results in artificial features. This problem is called the 'mixing' problem (see e.g., Hyvärinen, 1999a). Other motivation to use ICA for signal decomposition originated from studies, e.g., Richman (1986) who indicated that EOFs highly correspond to the shape of the data domain, owing to the orthogonality assumption built-in to the decomposition procedure. This is clearly a relevant issue if one is interested in identifying modes of variability in GRACE-derived total water storage (TWS) time series, e.g., for some areas of limited extension like a catchment or ocean basin (see Forootan et al. (2012, 2014c), as well as Sections 5.4 and 5.5).

An application of ICA to filter GRACE-TWS products was shown in Frappart et al. (2011a,b) who combined Gaussian-filtered GRACE solutions from three different analysis centers in their ICA implementation, with the assumption that each one of these fields contained independent information. Their results indicated that the TWS signals extracted by their ICA-filtering approach represented better correspondence to independent observations, compared to TWS derived from GRACE solutions that were filtered using a Gaussian filter (Jekeli, 1981). The ICA approach for filtering GRACE products, as in Frappart et al. (2011a,b), is discussed in Section 4.4.

In contrast to Frappart et al. (2011a,b), Forootan and Kusche (2012) proposed an ICA approach to identify statistically independent patterns from time series of GRACE-TWS changes. The main idea behind the implementation of ICA in this study is that independent physical processes will generate statistically independent source signals that are superimposed in the GRACE observations. Therefore, decomposing the GRACE observations into maximally statistically independent components should lead to base-functions that are physically at least more representative than others based on orthogonality only. These base-functions might be better suited for applications that look at each separated component individually, such as those aiming at separating GRACE and other data into signals from different compartments of the Earth system (e.g., Schmeer et al., 2012, Rietbroek et al., 2012a). The implemented ICA approach in this thesis, therefore, follows that of Forootan and Kusche (2012), which unlike Frappart et al. (2011a,b) does not rely on using multiple GRACE solutions as input.

The ICA approach of this thesis was applied in Forootan et al. (2012, 2014c) and Awange et al. (2014) to reduce a combination of the spectral and spatial leakage problem in GRACE-TWS products over Australia, Iran and the Nile Basin, respectively. Omondi et al. (2013a) applied ICA to extract the main independent behavior of TWS changes over the Greater Horn of Africa and related them to the climate condition of the region. Recently, Boergens et al. (2014) applied ICA to extract the post-glacial rebound trend in North America and the dominant variability of TWS in Africa. Applications of ICA for forecasting TWS changes is addressed in Forootan et al. (2014b).

To understand the mathematical fundamentals of ICA and its algebraic relation to the methods of Chapter 3, this chapter is organized as follows: the idea of higher order statistical signal separation is addressed in Section 4.1. Statistical independence is defined in Section 4.2, and the mathematical formulation of the ICA approach via rotation of de-correlated components (derived from the approaches in Chapter 3) is reported in Section 4.3. In Section 4.4, the ICA filtering of GRACE-TWS products (as in Frappart et al., 2011a,b) is discussed. Various approaches to approximate independence are introduced in Section 4.5. The ICA criterion of the diagonalization of the fourth-order cumulant tensor is presented in Section 4.5.1. Alternative criteria based on the diagonalization of different orders of the cumulant tensor are discussed in Section 4.5.2. In Section 4.5.3, a complex extension of the ICA method is introduced. The ICA technique based on the entropy estimation is presented in Section 4.5.4. Separation of deterministic signals, i.e. a linear trend and sinusoidal signals, using ICA is addressed in Section 4.6. In Section 4.7, a summary of the algebraic properties of the statistical decomposition techniques is provided, and finally in Section 4.8, an approach to estimate the uncertainty of independent components is treated.

4.1 Blind Source Separation

Blind source separation (BSS) is a signal processing technique, aiming at recovering unobserved patterns or 'sources' from observations that are measured by an array of sensors, known as 'mixture' signals (Hyvärinen and Oja, 2000). Such source patterns can be exploited from the observed mixtures by a statistical assumption such as the assumption of mutual independence between the sources (Cardoso, 1992). For instance, long-term observations such as that of continuous geodetic and geophysical measurements contain information about geophysical and climatic phenomena that are superimposed in the observed time series. Therefore, it is desirable to separate the observed time series (mixtures) and find a suitable representation (individual sources) for reasons of computational simplicity or enhancing the interpretation of observations. In the BSS approach, the representation is often sought-for as a linear transformation of the original data (mixture) and the mutual independence between source signals.

In order to separate multiple unknown sources from the observations, a typical approach is to utilize many sensors. Each sensor $\mathbf{x}_i$, $i = 1, \dots, n$ contains a combined information about the sources $\mathbf{s}_j$, $j = 1, \dots, m, m < n$. One needs to study the variation of the signal in each of the sensors and identify and separate the unknown sources. This situation occurs in many different applications, such as brain mapping, voice recognition, image enhancement (Stone, 2004), and climate studies (Aires et al., 2002, Hannachi et al., 2009). Assuming $\mathbf{X}$ is the data matrix that contains continuous multi-dimensional measurements or random variables, i.e. $\mathbf{X}$ contains $\mathbf{x}_i$ in its columns, the signal separation problem can be formulated as finding a function f that transforms the observations $\mathbf{X}$ into the sources $\mathbf{S}$ (containing $\mathbf{s}_j$), defined by

$$\mathbf{S} = f(\mathbf{X}). \tag{4.1}$$

In most applications, the relationship in Eq. (4.1) is assumed to be linear, thus it becomes

$$\mathbf{S} = \mathbf{W}\mathbf{X}, \tag{4.2}$$

where $\mathbf{W}$ is the so-called 'de-mixing matrix' to be determined, $\mathbf{S}$ and $\mathbf{X}$ are shown by capital letters to indicate that both mixture and source might be multivariate. Considering Eq. 4.2, $\mathbf{S}$ can also be seen as a linear transformation of $\mathbf{X}$ under a particular statistical assumption (such as independence). The motivation for preferring the linear transformation of Eq. (4.2) over Eq. (4.1) is mainly due to its computational simplicity. Some studies have already used non-linear methods for the separation purpose, most of them, however, do not provide significant improvements in skills over linear methods (see e.g., Goddard et al., 2001). It should be mentioned here that the adjective 'blind' in BSS is used to emphasize the fact that for the recovery of Eq. (4.2) no a priori knowledge is available of either the de-mixing matrix $\mathbf{W}$ or the sources $\mathbf{S}$. Thus, the source signals are estimated only by a statistical assumption between them.

Several principles can be used to define a suitable linear transformation to extract source signals from observed mixtures. These also include the transformations (decomposition approaches) that were introduced in Chapter 3. To implement all these transformations, one needs to define a statistical criterion, in which the de-mixing matrix $\mathbf{W}$ will be estimated. In Table 4.1 (Section 4.5.3), various transformations are presented, which can be used to extract patterns with specific statistical properties. For instance, using the singular value decomposition (SVD) technique (Eq. (3.20)), the data matrix $\mathbf{X}$ is transformed as $\mathbf{X} = \bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T$. Thus, the source signals in SVD derived as

$$\begin{aligned} \mathbf{X} &= \bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T, \\ \mathbf{X}\mathbf{E} &= \bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T\mathbf{E},\ \mathbf{E}^T\mathbf{E} = \mathbf{I}, \\ \mathbf{X}\mathbf{E}\mathbf{\Lambda}^{-1} &= \bar{\mathbf{P}}, \\ \mathbf{S} &= \bar{\mathbf{P}} = \mathbf{X}\mathbf{E}\mathbf{\Lambda}^{-1}, \end{aligned} \tag{4.3}$$

and

$$\begin{aligned} \mathbf{X} &= \bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T, \\ \bar{\mathbf{P}}^T\mathbf{X} &= \bar{\mathbf{P}}^T\bar{\mathbf{P}}\mathbf{\Lambda}\mathbf{E}^T, \bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I}, \\ \mathbf{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X} &= \mathbf{E}^T, \bar{\mathbf{P}}^T\bar{\mathbf{P}} = \mathbf{I}, \\ \tilde{\mathbf{S}} &= \mathbf{E} = \mathbf{X}^T\bar{\mathbf{P}}\mathbf{\Lambda}^{-1}. \end{aligned} \tag{4.4}$$

Therefore, considering Eq. (4.2), the de-mixing matrix $\mathbf{W}$ of the SVD transformation that can be used to estimate temporally de-correlated sources ($\mathbf{S}$) is derived by $\mathbf{\Lambda}^{-1}\mathbf{E}^T$, while considering $\mathbf{X}^T$ instead of the original data matrix $\mathbf{X}$. The de-mixing matrix $\mathbf{W}$ of $\mathbf{\Lambda}^{-1}\bar{\mathbf{P}}^T$ must be used to estimate spatially de-correlated sources ($\tilde{\mathbf{S}}$). In this chapter, the statistical method of ICA will be presented, which allows to extract source signals that are mutually statistically independent.

4.2 Statistical Independence

Statistical independence is a mathematical term that is defined based on the probability density function (PDF) as

$$p(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_j) = \prod_{i=1}^{j} p(\mathbf{s}_i). \tag{4.5}$$

In Eq. (4.5), $p(\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_j)$ denotes the joint PDF and $p(\mathbf{s}_i)$ denotes the marginal PDF of each source. Thus, random variables are independent if and only if their joint distribution can be factorized to the product of their marginal distributions (Hyvärinen, 1999a). Let $\mathbf{s}$ be a discrete random variable taking values of $s_1, s_2, \ldots, s_n$ with probabilities of $p_1, p_2, \ldots, p_n$, respectively. Then, the expected value of this random variable $E(\mathbf{s})$ is derived as

$$E(\mathbf{s}) = \sum_{m=1}^{n} s_m \, p_m, \tag{4.6}$$

where $E(.)$ is the expectation operator. From Eqs. (4.5) and (4.6), one can conclude that independence always implies orthogonality, i.e. for any two independent variables $\mathbf{s}_i$ and $\mathbf{s}_k$, one can write $E(\mathbf{s}_i\mathbf{s}_k) = E(\mathbf{s}_i)E(\mathbf{s}_k)$. This is due to the fact that the factorization of the joint PDF in Eq. (4.5) guarantees the factorization of the joint empirical covariance matrix. Obviously, the orthogonality does not always imply independence in Eq. (4.5) since it includes only up to second order statistical information (see an example in Hyvärinen, 1999a).

Statistical properties of a random variable such as x (without distinguishing in notation between random variable $\mathbf{x}$ or its realizations x) can be described by its moments or, more conveniently, by its cumulants that are denoted in this thesis as $C(x)$ or $K(x)$. For instance, it has been shown that the moments of random variables and the relationships between their distributions can be derived by the moment generating functions (Koch, 1999, page 106). The cumulants κ_n of a random variable x are defined via the cumulant-generating function $g(t)$, which is the logarithm of the moment-generating function $g(t) = \log[E(e^{tx})] = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}$. Therefore, the cumulants κ_n can be obtained by n times differentiating the expansion of $g(t)$ and evaluating the result at zero, or $\kappa_n = \frac{\partial^n}{\partial t^n} g(t)|_{t=0}$. Ferreira et al. (1997) showed that the cumulants κ_n appear similarly to the Taylor coefficients of the logarithmized moment function. To perform statistical signal separation, working with cumulants is advantageous over using moments. This is due to the fact that, e.g., the cumulant of the sum of two statistically independent random variables s_1 and s_2 can be written as the sum of the cumulant of each, which can be seen as $g_{s_1+s_2}(t) = \log[E(e^{t(s_1+s_2)})] = \log[E(e^{ts_1})E(e^{ts_2})] = \log[E(e^{ts_1})] + \log[E(e^{ts_2})] = g_{s_1}(t) + g_{s_2}(t)$. This property simplifies the estimation of independence criteria, discussed in Section 4.5.

Joint cumulant of several random variables $x_1, x_2, \ldots, x_n$ is similarly defined by the above cumulant generating function. For example, $K(x_1, x_2, x_3) = C(x_1, x_2, x_3) = E(x_1x_2x_3) - E(x_1x_2)E(x_3) - E(x_1x_3)E(x_2) - E(x_2x_3)E(x_1) + 2E(x_1)E(x_2)E(x_3)$. Considering random variables x with zero mean, the sample cumulants up to order four (Cardoso, 1999) are derived as

$$\begin{aligned} C(x_i) &= 0 \\ C(x_i, x_j) &= E(x_ix_j) \\ C(x_i, x_j, x_k) &= E(x_ix_jx_k) \\ C(x_i, x_j, x_k, x_l) &= E(x_ix_jx_kx_l) \\ &- E(x_ix_j)E(x_kx_l) \; - E(x_ix_k)E(x_jx_l) - E(x_ix_l)E(x_jx_k). \end{aligned} \tag{4.7}$$

4.3 Independent Component Analysis (ICA)

ICA was introduced for the first time in the context of the BSS techniques (Hyvärinen, 1999a) in the form of Eq. (4.2) or equivalently as

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \tag{4.8}$$

where $\mathbf{A}$ is a mixing matrix to be determined, while $\mathbf{W}$ in Eq. (4.2) can be identified with the pseudo-inverse matrix of $\mathbf{A}$ in Eq. (4.8), i.e. $\mathbf{W} = \mathbf{A}^{\dagger}$. Typically, in ICA algorithms, $\mathbf{W}$ or $\mathbf{A}$ is sought such that $\mathbf{S}$ contains rows that are mutually as statistically independent as possible (Cardoso, 1992, Cardoso and Souloumiac, 1993). Basically, as was shown in Section 4.2, when variables are independent, the values of each variable provide no information about the values of the other. By contrast, even though some variables are uncorrelated, the value of one variable can still provide information (higher than two statistical relationship) about the values of the other variables (Hyvärinen, 1999a). ICA, therefore, seeks a set of statistically independent source signals $\mathbf{S}$ amongst a set of observed mixtures $\mathbf{X}$. This can be realized by maximizing a measure of the joint entropy of the extracted signals (Hyvärinen and Oja, 2000, Stone, 2004), or minimizing the high-order statistical dependence between the signals (e.g., Cardoso and Souloumiac, 1993). In practice, iterative methods are used to maximize or minimize such given cost functions (see Section 4.5).

However, in order to estimate source signals $\mathbf{S}$ from observation $\mathbf{X}$, using an ICA transformation (Eq. (4.2) or (4.8)), the following assumptions are made:

1. The number of observed linear mixtures must be equal or larger than the number of independent sources, otherwise Eq. (4.8) would be under-determined, and $\mathbf{S}$ cannot be estimated.
2. All the independent components (sources $\mathbf{s}_i$) should be non-Gaussian. In case that only one of the independent components is Gaussian, the ICA model can still be estimated (Cardoso, 1999). More details will be provided in Section 4.5.

Several studies, e.g., Comon (1994a), Aires et al. (2002), and Forootan and Kusche (2012) stated that a simple way to estimate $\mathbf{W}$ in Eq. (4.2) is to first de-correlate the data, as discussed in Section 3.1.3, and then seek a proper rotation to make the components as independent as possible. Mansour and Jutten (1995) showed that the whitening (de-correlation) step improves the performance of signal separation by avoiding the occurrence of several local maxima in the solutions of independence criteria.

Following Comon (1994a), Aires et al. (2002), and Forootan and Kusche (2012), the ICA decomposition is derived in this thesis as a rotated extension of the PCA or SVD transformation. As was stated in Section 3.1.1, the main benefit of PCA or SVD is that it provides a few orthogonal modes that represent the dominant part of the observations. Thus, if the data matrix $\mathbf{X}$ can be fairly well represented by $\mathbf{X}_j = \bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{E}_j^T$, where $j < \min(n, p)$ is the number of retained modes (associated with the largest singular values). The ICA decomposition, via the rotation of orthogonal modes, proceeds as follows

$$\mathbf{X} \simeq \mathbf{X}_j = \underbrace{\bar{\mathbf{P}}_j \mathbf{\Lambda}_j \mathbf{R}_j}_{\mathbf{A}_{n\times j}} \underbrace{\mathbf{R}_j^T \mathbf{E}_j^T}_{\tilde{\mathbf{S}}_{j\times p}}, \tag{4.9}$$

where $\mathbf{X}$ and $\mathbf{X}_j$ are $n \times p$ data matrices. The $n \times j$, $j \times j$, and $p \times j$ matrices $\bar{\mathbf{P}}_j$, $\mathbf{\Lambda}_j$, and $\mathbf{E}_j$ are derived from SVD. To derive the ICA modes, an optimum $j \times j$ rotation matrix $\mathbf{R}_j$

has to be defined that rotates either $\bar{\mathbf{P}}_j$ or $\mathbf{E}_j$ while at the same time making their columns as statistically independent as possible. Orthogonal rotations as $\mathbf{R}_j{\mathbf{R}_j}^T = \mathbf{I}_j$ are used in this thesis to derive independent components. As it will be shown later, selecting $\mathbf{R}_j$ as orthogonal is computationally beneficial since the associated mixing (or de-mixing) matrix in Eq. (4.8) (or Eq. (4.2)) remains orthogonal. Defining a proper rotation matrix requires optimization of a measure of independence, which is addressed in Section 4.5. The statistically independent source signals from Eq. (4.9) are estimated as

$$\tilde{\mathbf{S}}_{j\times p} = \mathbf{R}_j^T \mathbf{\Lambda}_j^{-1} \bar{\mathbf{P}}_j^T \mathbf{X}_j = \mathbf{W}\mathbf{X}_j, \tag{4.10}$$

in which rows of $\tilde{\mathbf{S}}$ are obtained spatially as independent as possible and the decomposition method called 'Spatial ICA (SICA)'. The mixing matrix $\mathbf{A}_{j\times n}$ of SICA is derived by pseudo-inversing of $\mathbf{W}_{j\times n}$, i.e. $\mathbf{A}=\mathbf{W}^{\dagger}$, where † is the pseudo-inverse operator. In the SICA decomposition, the rotated PCs are still orthogonal.

The transpose of Eq. (4.9) provides $\mathbf{X}^T \simeq \mathbf{X}_j^T = \underbrace{\mathbf{E}_j \mathbf{\Lambda}_j \mathbf{R}_j}_{\mathbf{A}_{p\times j}} \underbrace{\mathbf{R}_j^T \bar{\mathbf{P}}_j^T}_{\mathbf{S}_{j\times n}}$. Accordingly, the source signals are derived as

$$\mathbf{S}_{\mathrm{j\times n}} = \mathbf{R}_j^T \mathbf{\Lambda}_j^{-1} \mathbf{E}_j^T \mathbf{X}_j^T = \mathbf{W}\mathbf{X}_j^T, \tag{4.11}$$

in which rows of $\mathbf{S}$ are temporally as independent as possible. The transformation in Eq. (4.11) is called 'Temporal ICA (TICA)'. The mixing matrix $\mathbf{A}_{j\times p}$ of TICA is derived by pseudo-inversing of $\mathbf{W}_{j\times p}$, i.e. $\mathbf{A}=\mathbf{W}^{\dagger}$. In the TICA decomposition, the rotated EOFs are still orthogonal.

In the language of statistics, the two options of Eqs. (4.10) and (4.11) allow us to interpret the original data $\mathbf{X}$ as a mixture of either independent spatial sequences or temporal chains. When applied to GRACE-TWS, both ways have their own advantages, see Chapter 5 for discussion.

In order to illustrate what happens within the whitening and rotation steps, assume a mixture of an annual signal $s_1(t) = 5\,\sin(2\pi t) + n_1(t)$ and a linear trend $s_2(t) = t + n_2(t)$, where t represents monthly epochs within the period of 2002 to 2013.5, n_1 and n_2 are temporal random noise. Also consider that s_1 and s_2 are linearly mixed using $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, as $\mathbf{X} = \mathbf{A}\mathbf{S}$, where rows of $\mathbf{X}$ store the mixtures $x_1(t)$ and $x_2(t)$ and rows of $\mathbf{S}$ contain the source signals $s_1(t)$ and $s_2(t)$.

Figure 4.1 (left) shows the plot of x_1 against x_2. The result of applying SVD on the mixture is shown by gray dots in Fig. 4.1 (right). As a result of whitening, the joint distribution is normalized and sphered. Figure 4.1 (right) shows two source signals, derived from the Temporal ICA transformation (implemented following Forootan and Kusche, 2012), that are plotted against each other (black dots). As a result of rotation in Eq. (4.11), the sphered distribution is rotated in a way that the joint distribution is as independent as possible (compare the black and gray dots in Fig. 4.1 (right)).

Figure 4.2 shows the performance of the SVD and ICA techniques in extracting the source signals $s_1(t)$ and $s_2(t)$ from the simulated linear mixtures $x_1(t)$ and $x_2(t)$. Figure 4.2 (left) shows the source signals derived from SVD. The results of SVD indicate that the linear component is still contaminated by the annual source. The results of ICA shown in Fig. 4.2 (right) indicate that the introduced annual and linear sources are fairly well extracted.

In this section, the algebraical connections between the transformation defined in the context of BSS and the second order transformations of Chapter 3 were established. In the following, after providing some remarks on the application of ICA for noise reduction, the estimation of rotation matrix to derive the independent sources of Eqs. (4.10) and (4.11) will be discussed.

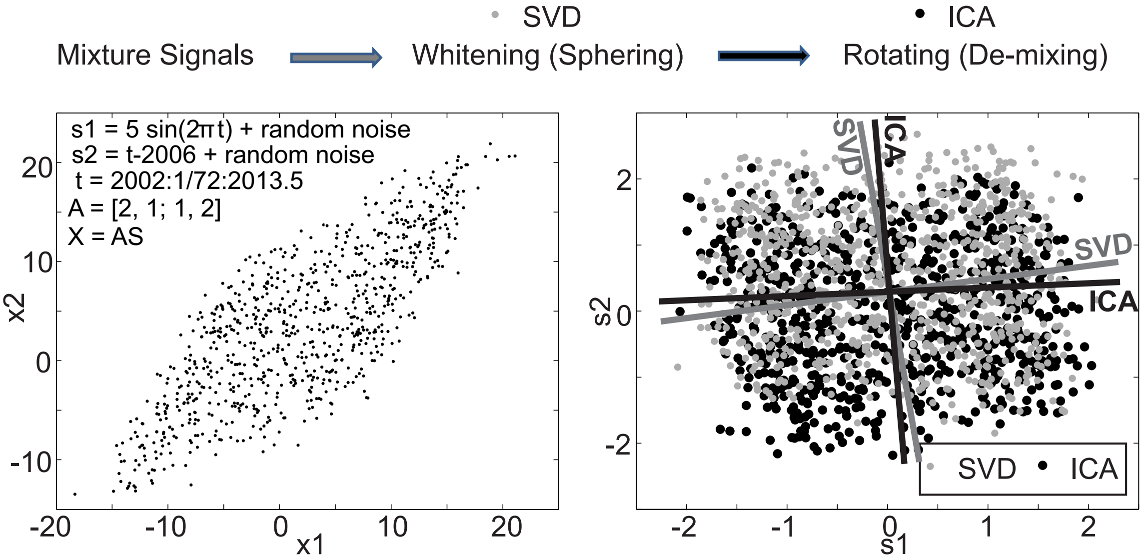


Figure 4.1: Impact of whitening and rotation transformations on a mixture of an annual signal and a linear trend in the presence of noise, i.e. $s_1(t) = 5\,\sin(2\pi t) + n_1(t)$ and $s_2 = t + n_2(t)$. Figure 4.1 (left) shows the two mixtures ($x_1(t) = 2s_1(t) + s_2(t)$ and $x_2(t) = s_1(t) + 2s_2(t)$) plotted against each other. Figure 4.1 (right) corresponds to the left, but shows a plot of the source signals (against each other) after applying the SVD (gray dots) and ICA (black dots) transformations.

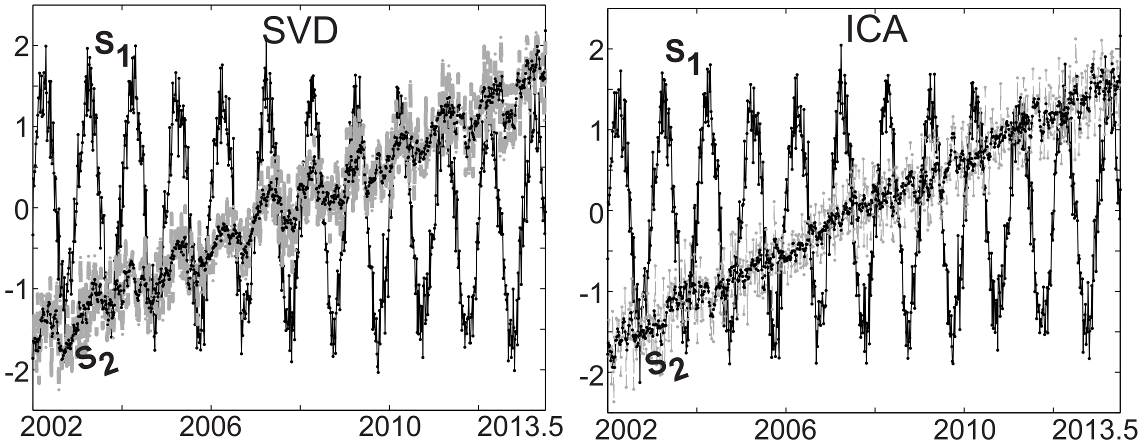


Figure 4.2: A comparison between the performance of SVD and ICA for recovering the source signals in Fig. 4.1. Figure 4.2 (left) shows source signals ($s_1(t)$ and $s_2(t)$) of the SVD approach and those sources of the ICA decomposition are shown in Fig. 4.2 (right). In order to better visualize the performance of separations, the linear components are smoothed and shown with a darker color, along with the original component.

4.4 A Remark on the Application of ICA for Noise Reduction in GRACE Products

Before discussing different criteria to approximate independence, in this section, the application of ICA for filtering GRACE-TWS products, that has been introduced in Frappart et al. (2011a,b) as the 'ICA-filtering', is discussed. This discussion is necessary since the assumption behind the application of ICA in this thesis and the way of its implementation are different from that of Frappart et al. (2011a,b).

In Frappart et al. (2011a), GRACE level 2 solutions from their official providers (CSR, GFZ, and JPL) were first pre-filtered using the Gaussian filter (Jekeli, 1981) with half-width radius of 300, 400, and 500 km. The Gaussian filter was applied to reduce the striping noise level in TWS products. The pre-filtered solutions (corresponding to one of the selected smoothing radii) were then converted to global $1^\circ \times 1^\circ$ TWS maps, using the approach in Wahr et al. (1998). For each month (with available solutions from the three providers), a data matrix $\mathbf{Y}_{3\times 64800}$ was created that contained the Gaussian-filtered TWS values of CSR, GFZ, and JPL solutions in its rows. Finally, the ICA technique was used in its BSS definition (Eq. (4.8)) to decompose the data matrix $\mathbf{Y}$ into a linear mixture of a 'true' TWS signal and noise. In Frappart et al. (2011a,b), the ICA decomposition is read as

$$\mathbf{Y} = \mathbf{A}\mathbf{S}, \tag{4.12}$$

where the 3×3 matrix $\mathbf{A}$ is a mixing matrix whose computation is discussed in Section 4.5. The source matrix $\mathbf{S}$ can be computed from $\mathbf{S} = \mathbf{A}^{-1}\mathbf{Y}$ or similar to Eq. (4.2), $\mathbf{S} = \mathbf{W}\mathbf{Y}$, thus $\mathbf{A}^{-1} = \mathbf{W}$. Rows of $\mathbf{W}$ are ordered with respect to the variance they represent. Since, the variance of TWS is dominant in the Gaussian filtered TWS products, one can expect that the product of the first row of $\mathbf{W}$ (shown by $\mathbf{w}_1$) and $\mathbf{Y}$ is likely associated to TWS and the products of the other two rows ($\mathbf{w}_2$ and $\mathbf{w}_3$) and $\mathbf{Y}$ are likely related to noise. Thus, a combination of the three Gaussian filtered TWS products of CSR, GFZ and JPL, stored in the rows of $\mathbf{Y}$, has been interpreted as the 'ICA-filtered solution', i.e. shown here by $\mathbf{Y}_F$

$$\mathbf{Y}_F = \mathbf{w}_1\mathbf{Y}, \tag{4.13}$$

where $\mathbf{w}_1 = [w_{1,1}\ w_{1,2}\ w_{1,3}]$ is the first row of the de-mixing matrix $\mathbf{W}$. From Eq. (4.13), it is clear that each entry of $\mathbf{Y}_F$ ($y_F(1,j), j = 1, \ldots, 64800$) is a weighted value of the three solutions in $\mathbf{Y}$, i.e. $y_F(1,j) = \sum_{i=1}^{3} w(1,i)y(i,j), j = 1, \ldots, 64800$, where $y(i,j)$ are the entries of $\mathbf{Y}$. The noise component $\mathbf{Y}_N$ has been computed as

$$\mathbf{Y}_N = \mathbf{w}_2\mathbf{Y} + \mathbf{w}_3\mathbf{Y}, \tag{4.14}$$

where $\mathbf{w}_2$ and $\mathbf{w}_3$ are the second and third rows of the de-mixing matrix $\mathbf{W}$ ($\mathbf{w}_2 = [w_{2,1}\ w_{2,2}\ w_{2,3}]$ and $\mathbf{w}_3 = [w_{3,1}\ w_{3,2}\ w_{3,3}]$).

Frappart et al. (2011a,b) claim that the ICA-filtered $\mathbf{Y}_F$ products represent less striping pattern, when compared to the Gaussian-filtered products and those products that are filtered by an anisotropic filter (e.g., Swenson and Wahr, 2006). This statement might be true since $\mathbf{Y}_F$ in Eq. (4.13) is derived as a weighted combination of the three pre-filtered TWS products. As a result of such averaging, the noise level has been decreased in the ICA-filtered solution.

The application of Eqs. (4.13) and (4.14), however, might be considered with cautions due to the following reasons: In Eq. (4.12), it has been assumed that the TWS and noise patterns of the three official products are the same and their differences are only in their strength, which is reflected in the estimated $\mathbf{A}^{-1}$ or $\mathbf{W}$ matrix. Obviously, this cannot be true since the three providers use different approaches for recovery of the time-variable gravity solutions. The inconsistency between different products might be due to several technical issues, such as the difference between spatial resolution of the gravity recovery (the highest degree and order of the level 2 solutions), methods for constraining the normal equation (required for estimating level 2 products from level 1B products) (see e.g., Mayer-Gürr, 2008), and the difference in the background models. Macrander et al. (2010), for instance, indicated large differences between mass estimations derived from different GRACE solutions over the oceans. Although, the applied Gaussian filter in the pre-filtering step damp the noise level of high-order potential coefficients in level 2 products, there is no guarantee that the noise and signal patterns of the three products can be treated in the same way, as it has been considered in Frappart et al. (2011a). As a result, it has been concluded here that the applied averaging in Eqs. (4.13) and (4.14) to respectively estimate the ICA-filtered product is questionable and might not correctly consider the uncertainty and signal structure of the three official products.

To modify the filtering formulation of Eq. (4.12), one has to consider that each monthly TWS field from CSR, GFZ, and JPL (each stored in a 1×64800 vector $\mathbf{y}_i, i = 1, 2, 3$) consists of two parts, that are signal $\mathbf{s}_i$ and noise $\mathbf{n}_i$ ($\mathbf{y}_i = \mathbf{s}_i + \mathbf{n}_i$). One should also consider that in general the signal and noise parts of the three products are not the same, i.e. $\mathbf{s}_1 \neq \mathbf{s}_2 \neq \mathbf{s}_3$ and $\mathbf{n}_1 \neq \mathbf{n}_2 \neq \mathbf{n}_3$ (since different background models are reduced and the products have been differently processed). This yields an under-determined system equation that cannot be solved with the proposed formulations of Eq. (4.12). One might solve this equation with an assumption on the noise or signal behaviors of the three products, whose discussion is out of the scope of this thesis.

Unlike Frappart et al. (2011a,b), the ICA approach in this thesis has been applied to a time series of TWS products derived from one specific data provider at each time that ICA is performed. Therefore, following the same concept in Chapter 3, the $n \times p$ data matrix $\mathbf{X}$ contains n epochs (months) observations of TWS (e.g., for $1° \times 1°$ global maps, $p = 64800$). Thus, $\mathbf{X}$ contains the data from either CSR, GFZ, or JPL, or any other data centers. Then, Eq. (4.10) or (4.11) is applied to respectively extract spatially or temporally independent patterns that are superimposed in the repeated TWS observations of $\mathbf{X}$. It will be shown in Chapter 5 that this decomposition is beneficial for several signal-signal and signal-noise separation applications.

4.5 On Independence Criteria

Different criteria exist, which can be used to measure independence and equivalently the demixing matrix $\mathbf{W}$ in Eq. (4.2) (see examples in Comon, 1994a,b, Hyvärinen, 1997, Cardoso, 1999). As already mentioned in Section 4.2, in order to extract independent components, $\mathbf{W}$ should be defined in a way that the joint PDF of source signals is factorized to a product of individual variables (Hyvärinen, 1999a). Thus, similar to the introduced techniques of Chapter 3, ICA is also a data-driven method that can be used to explore information from multivariate observations. It is worth mentioning here that the data-driven methods do not assume any parametric model for observations as it is usual, e.g., in the general linear models (Koch, 1999, chapter 3). By contrast, ICA relies on the signal type that is specified with its statistical structure. Thus, one can conclude that the models implicit in data-driven methods, such ICA, are generic because they attempt to extract patterns with specific statistical characteristics, rather than a best fit to a specific model signal (Stone, 2004).

A possible way to perform such factorization can be realized by defining the higher order comulants of the observed mixture, then, finding a transformation to diagonalize it (Cardoso, 1999). Another family of ICA resulted from the central limit theorem (Stone, 2004, chapter 5), which states that the distribution of a sum of independent random variables tends toward a Gaussian distribution. Therefore, maximizing the non-Gaussian property ('non-Gaussianity') of the source signals ($\mathbf{WX}$ in Eq. (4.2)) provides the desired independent components (Hyvärinen and Oja, 2000). In the following, two families of ICA criteria based on higher order cumulants (Section 4.5.1) and entropy (Section 4.5.4) are described. The focus of this thesis is, however, on the ICA algorithm based on the fourth-order cumulants that is described in the following.

4.5.1 ICA by Joint Diagonalization of the Fourth-Order Cumulant Tensor

A common measure of non-Gaussianity is the kurtosis, which is a simple form of the fourth-order cumulant of a random variable x (see Eq. (4.7)). In general, the fourth-order cumulant of a random variable x can be written as

$$\kappa_4(x) = \mu_4 - 4\mu_3\mu_1 - 3{\mu_2}^2 + 12\mu_2{\mu_1}^2 - 6{\mu_1}^4, \tag{4.15}$$

where $\mu_n = E[(x - E(x))^n]$ is the n'th statistical moment of x (Comon, 1994a). When the random variable x is symmetrically distributed and centered, odd moments of Eq. (4.15) become zero (i.e. $\mu_j = 0$, for j odd). The second-order cumulant equals to the variance $\mu_2 = \kappa_2 = E(x^2)$, and the fourth-order cumulant equals to the kurtosis $\text{kurt}(x) = \kappa_4(x) = \mu_4 - 3{\mu_2}^2 = E(x^4) - 3E^2(x^2)$. For a unit variance random variable x that is temporally centered, the kurtosis is also derived as $E(x^4)/E^2(x^2) - 3$. Random variables with zero kurtosis are Gaussian distributed. Negative kurtosis indicates sub-Gaussian distribution (e.g., periodic signals), while positive kurtosis corresponds to signals with super-Gaussian distribution (peaky signals). Forootan and Kusche (2012), for instance, examined the kurtosis of GRACE-TWS time series for the period of 2003 to 2011 and showed that more than half of the time series exhibit an absolute value of kurtosis bigger than 0.5, which proves their non-Gaussian distribution. This justifies the application of ICA for separating time-variable TWS observations.

In general, those ICA algorithms that extract source signals based on the information in the kurtosis or the fourth-order cumulant are computationally efficient. This is due to the linear property of cumulants, which indicates that if random variables are statistically independent, their cumulants can be factorized (Cardoso, 1999). This can be written as

$$\kappa_4\left(\sum_{i=1}^{p} a_i s_i\right) = \sum_{i=1}^{p} {a_i}^4 \kappa_4(s_i), \tag{4.16}$$

where $a_i, (i = 1, 2, ..., p)$ are the entries of the mixing matrix $\mathbf{A}$, and $s_i, (i = 1, 2, ..., p)$ are elements of the corresponding rows from the source signal $\mathbf{S}$. The relation in Eq. (4.16), therefore, has been used to design independent criteria based on the fourth-order cumulants (e.g., Cardoso and Souloumiac, 1993). In the multivariate case, with centered, symmetrically distributed n-dimensional random vectors $\mathbf{x}_i$, the second-order multivariate cumulant corresponds to the auto-covariance matrix $C(\mathbf{x}_i, \mathbf{x}_j) = E(\mathbf{x}_i\mathbf{x}_j)$ and the fourth-order cumulant corresponds to

$$\begin{aligned} C(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) &= E(\mathbf{x}_i\mathbf{x}_j\mathbf{x}_k\mathbf{x}_l) \\ &- E(\mathbf{x}_i\mathbf{x}_j)E(\mathbf{x}_k\mathbf{x}_l) - E(\mathbf{x}_i\mathbf{x}_k)E(\mathbf{x}_j\mathbf{x}_l) - E(\mathbf{x}_i\mathbf{x}_l)E(\mathbf{x}_j\mathbf{x}_k). \end{aligned} \tag{4.17}$$

Later in this section, we will show that to perform ICA, Eq. (4.17) is built based on the whitened random variables, thus, our assumption on the symmetric joint distribution of the random variables is justified.

In particular, under the assumption of statistical independence of $\mathbf{x}$, their joint PDF decouples into a product of individual PDFs. Thus, replacing the random variables $\mathbf{x}$ in Eq. (4.17) by independent variables $\mathbf{s}$, one can see that

$$C(\mathbf{s}_i, \mathbf{s}_j, \mathbf{s}_k, \mathbf{s}_l) = \kappa_4(\mathbf{s}_i)\delta_{ijkl}, \tag{4.18}$$

where δ_{ijkl} is Kronecker's function (Cardoso, 1999), defined by

$$\delta_{ijkl} = \begin{cases} 1 & \text{if } i = j = k = l \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, ICA decomposition aims at finding a linear transformation of the random variables $\mathbf{x}$ whose fourth-order cumulant attains the form of Eq. (4.18).

A fourth-order cumulant is a tensorial quantity. In practice, it is more handy to work with matrices and, therefore, one introduces cumulant matrices $\mathbf{Q}(\mathbf{M})$ (e.g., Cardoso, 1999) as a '2D-contraction' of a four-dimensional cumulant tensor with an arbitrary $n \times n$ matrix $\mathbf{M} = (m_{ij})$ (with zeros everywhere except 1 at index (i, j)), or $\mathbf{Q}(\mathbf{M}) = (q_{ij})$ with

$$q_{ij} = \sum_{k=1}^{n} \sum_{l=1}^{n} C(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l) m_{kl} \ . \tag{4.19}$$

Under the mentioned assumptions on the random vector (a random $n \times 1$ vector $\mathbf{x}$), one can define the associated cumulant matrix $\mathbf{Q}(\mathbf{M})$ as (Cardoso, 1999)

$$\mathbf{Q}(\mathbf{M}) = E\left((\mathbf{x}^T\mathbf{M}\mathbf{x})(\mathbf{x}\mathbf{x}^T)\right) - \mathbf{\Sigma}\ \mathrm{trace}(\mathbf{M}\mathbf{\Sigma}) - \mathbf{\Sigma}(\mathbf{M} + \mathbf{M}^T)\mathbf{\Sigma}, \tag{4.20}$$

with $\mathbf{\Sigma}$ being the auto-covariance matrix of $\mathbf{x}$. The fourth-order cumulant of independent variable $\mathbf{s}_i$, shown by $\mathbf{Q}(\mathbf{s}_i)$ can simply be derived from Eq. (4.17) as

$$C(\mathbf{s}_i^4) = E(\mathbf{s}_i^4) - 3E^2(\mathbf{s}_i^2). \tag{4.21}$$

It has been mentioned that all statistical information of a Gaussian variable is summarized in the first and second moments and that high-order cumulants are all zero. This makes the independence concept equivalent with orthogonality (uncorrelatedness) for Gaussian signals. Therefore, one can interpret the fourth-order cumulant matrix in Eq. (4.20) as a generalized form of the auto-covariance matrix, which contains all statistical information of the available signals up to order four.

The random variable $\mathbf{x}$ in Eq. (4.20) can be replaced by the source components of Eq. (4.10) or (4.11). As a result, the solution of ICA is sought by finding a suitable rotation $\mathbf{R}_j$ that makes the fourth-order cumulant matrix of Eq. (4.20) as diagonal as possible. Following Eq. (4.18), such diagonalization would guarantee the independence of the desired sources.

Basically, the joint diagonalization problem can be solved in different ways. Cardoso and Souloumiac (1993), for instance, solved the problem using a Jacobi eigenvalue algorithm. Consider a maximal set of cumulant matrices Eq. (4.20) for either rotating EOFs or rotating PCs. The cumulant tensor $\mathbf{Q}$ in Eq. (4.20) contains n^4 entries, i.e. the number of fourth-order cross-cumulants. Then, $\mathbf{R}_j$ is found as the minimizer of the squared off-diagonal cumulant entries

$$f(\mathbf{R}_j) = \sum_{t \neq s}^{j} f_{ts}^2 \qquad \mathbf{F} = \sum_{m=1}^{n^2} \mathbf{R}_j^T \mathbf{Q}(\mathbf{M}_m) \mathbf{R}_j, \tag{4.22}$$

where t and s represent the row and column of each entry. After solving the joint diagonalization optimization of Eq. (4.22) to find the $j \times j$ orthogonal rotation matrix $\mathbf{R}_j$ ($\mathbf{R}_j\mathbf{R}_j^T = \mathbf{I}_j$), one can replace the rotation matrix in Eq. (4.10) or Eq. (4.11) to identify most independent sources. Note that j in Eq. (4.22) represents the number of dominant modes. In Eqs. (4.17), (4.18), and (4.19), however, j refers to an arbitrary variable. The ICA algorithm based on the diagonalization of the fourth-order cumulants is summarized in Algorithm 1. In this section, stage 1, 2, and 3 of the ICA algorithm were described (see Section 4.5.1.2).

4.5.1.1 Joint Diagonalization (JD)

To search for $\mathbf{R}$ that minimizes Eq. (4.22), the cumulant tensor $\mathbf{Q}$ in Eq. (4.20) has been written as a set of n^2 symmetric matrices

$$\mathbf{Q} = \{\mathbf{Q}_1, \dots, \mathbf{Q}_k\}, \tag{4.23}$$

where $k = n^2$. Joint diagonalization (JD) implies minimizing the squared sum of the off-diagonal elements of the following contrast function

$$f(\hat{\mathbf{R}}) = \sum_{m=1}^{k} \|\hat{\mathbf{V}}_m^T \mathbf{Q}_m \hat{\mathbf{V}}_m - \mathrm{diag}(\hat{\mathbf{V}}_m^T \mathbf{Q}_m \hat{\mathbf{V}}_m)\|^2, \tag{4.24}$$

where the operator diag(.) produces a matrix with all off-diagonal elements being set to zero, and $\|.\|$ represents the Frobenius norm or the Hilbert-Schmidt norm. The estimated rotation matrix $\hat{\mathbf{R}}$ in Eq. (4.24) is defined as a product of k rotation matrices $\hat{\mathbf{V}}_m$ ($\hat{\mathbf{R}} = \prod_{m=1}^{k} \hat{\mathbf{V}}_m$), where each $\hat{\mathbf{V}}_m$ diagonalizes a cumulant matrix ($\mathbf{Q}_m$). The product $\hat{\mathbf{R}}$ is, therefore, the joint diagonalizer of Eq. (4.23) or equivalently Eq. (4.22) (e.g., Févotte and Theis, 2007).

A sufficient criterion for the existence of the solution $\hat{\mathbf{R}}$ is that all $\mathbf{Q}_m$ in Eq. (4.23) ‘commute’, i.e., they are simultaneously triangularizable (e.g., Cardoso and Souloumiac, 1995, Févotte and Theis, 2007). In reality, however, the cumulants matrices in Eq. (4.23) are estimated based on length-limited samples, thus, they only approximately share the same eigenstructure and do not fully commute (Févotte and Theis, 2007). As a result, the exact solution of the joint diagonalization in Eq. (4.23) does not perfectly diagonalize the fourth-order cumulant matrix of Eq. (4.22) or Eq. (4.23). Therefore, those source components that are defined following this computation are statistically ‘as independent as possible’ rather than ‘absolutely’ independent (e.g., Cardoso and Souloumiac, 1993).

Algorithms for performing joint diagonalization include, for instance, gradient descent (e.g., Afsari and Krishnaprasad, 2004), Jacobi-like iterative by planar rotation (e.g., Cardoso and Souloumiac, 1993, 1995), and subspace fitting (e.g., van der Veen, 2001). In order to perform joint diagonalization, a Jacobi-like technique is used in this study, which was originally proposed in Cardoso and Souloumiac (1993, 1995) in the form of an ICA algorithm known as the Joint Approximate Diagonalization of Eigenmatrices (JADE). The technique solves the optimization of Eq. (4.24) using plane rotations under unitary constraints of $\mathbf{V}_m$.

In general, a Jacobi rotation matrix of $\mathbf{V}_m(\theta, t, s)$, $t < s$ is defined by the identity matrix except for the entries at positions t and s, i.e. $v_{tt} = v_{ss} = \cos(\theta)$, and $v_{ts} = -v_{st} = \sin(\theta)$. Considering Eq. (4.24), the Jacobi rotation only affects the elements in rows and columns t and s of $\mathbf{Q}_m$ in each rotation (sweep). To find an optimum solution for the contrast function in Eq. (4.24), one has to minimize the following equation

$$f = \sum_{m=1}^{k} \tilde{q}_{ssm}^2 + \tilde{q}_{ttm}^2, \tag{4.25}$$

where $\tilde{q}_{ssm} = (v_{ss}\ v_{st})\begin{pmatrix} q_{ssm} & q_{stm} \\ q_{tsm} & q_{ttm} \end{pmatrix}(v_{ss}\ v_{st})^T$ and $\tilde{q}_{ttm} = (v_{ts}\ v_{ss})\begin{pmatrix} q_{ssm} & q_{stm} \\ q_{tsm} & q_{ttm} \end{pmatrix}(v_{ts}\ v_{ss})^T$. The subindices stm denote that the entry belongs to the row s and the column t of the m'th matrix in Eq. (4.23). Cardoso and Souloumiac (1995) and Févotte and Theis (2007) show that the quadrant form of Eq. (4.25) can be written as

$$f = \mathbf{w}^T \mathbf{G} \mathbf{w}, \tag{4.26}$$

where $\mathbf{w} = (\cos(2\theta)\ \sin(2\theta))$ and $\mathbf{G} = 1/4(q_{ssm} - q_{ttm}\ q_{stm} - q_{tsm})^T(q_{ssm} - q_{ttm}\ q_{stm} - q_{tsm})$. Therefore, the rotation angle θ that optimally rotates only one cumulant matrix, follows from the dominant eigenvector $\mathbf{w}$ of $\mathbf{G}$. To jointly diagonalize the other cumulant matrices, the iteration should be repeated over different choices of s and t, $1 \leq s < t \leq n$ (Cardoso and Souloumiac, 1995).

To stop the algorithm from iterations, one needs to introduce a criterion. Cardoso (1999) selected $\theta_{\min}$ in a way that rotations by angles smaller than $\theta_{\min} = 10^{-2}/\sqrt{T}$, with T being the length of time series, is not possible. It should be mentioned here that the described joint diagonalization procedure can be extended to the case of complex random variables. The general formulations of the optimization problem in Eq. (4.24) remains the same. To compute the optimal Jacobi rotation, however, one needs to estimate the dominant eigenvectors of a real 3×3 matrix $\mathbf{G}$ instead of the 2×2 matrix $\mathbf{G}$ in Eq. (4.26). The formulation of joint diagonalization of complex variables has been discussed in (Cardoso and Souloumiac, 1993, 1995).

4.5.1.2 Spatial and Temporal ICA Algorithms

In this section, Algorithm 1 is provided to estimate the components of the Spatial ICA and Temporal ICA techniques (shown respectively by SICA and TICA), which were respectively introduced as Eqs. (4.10) and (4.11). In the algorithm, the subindices ‘Spatial’ and ‘Temporal’ are used to refer to the components that are spatially and temporally orthogonal, respectively. An optimum rotation matrix in Algorithm 1 is found by joint diagonalization of the fourth-order cumulant tensor described in the previous section.

The presented ICA algorithm starts by estimating a de-correlation transformation computed based on the observation matrix $\mathbf{X}$. In stage 1 of the algorithm, two choices of the whitening matrix $\hat{\mathbf{W}}_{\text{Spatial}} = \mathbf{\Lambda}^{-1}\bar{\mathbf{P}}^T$ and $\hat{\mathbf{W}}_{\text{Temporal}} = \mathbf{\Lambda}^{-1}\mathbf{E}^T$ are provided, where the first choice ($\hat{\mathbf{W}}_{\text{Spatial}}$) will lead to a set of spatially independent sources (Eq. (4.10)) and the second choice ($\hat{\mathbf{W}}_{\text{Temporal}}$) provides patterns that are temporally independent (Eq. (4.11)). The whitening transformation is applied using the SVD technique (see Eqs. (4.3) and (4.4)).

In stage 2, one must choose between the two whitening matrices to extract either spatial or temporal orthogonal patterns from $\mathbf{X}$, i.e. $\mathbf{Z}_{\text{Spatial}} = \hat{\mathbf{W}}_{\text{Spatial}}\mathbf{X}$ or $\mathbf{Z}_{\text{Temporal}} = \hat{\mathbf{W}}_{\text{Temporal}}\mathbf{X}^T$. Since the estimated $\mathbf{Z}_{\text{Spatial/Temporal}}$ are centered and uncorrelated, one can use the formulation of Eq. (4.20) to estimate their corresponding fourth-order cumulant tensor, which is performed in stage 3.

In stage 4, the JD algorithm of Section 4.5.1.1 is applied to define an optimum rotation $\mathbf{R}$ that makes the estimated cumulant tensor (of stage 3) as diagonal as possible. A default value of $\xi = 10^{-2}/\sqrt{(T)}$, with T being the length of time series, has been selected to prevent the algorithm from iterations without convergence.

Using the estimated rotation matrix $\mathbf{R}$, in stage 5, the de-mixing matrix of SICA or TICA, shown by $\hat{\mathbf{W}}_{\text{SICA/TICA}}$, are computed. Finally in stage 6, the desired spatially or temporally independent components ($\hat{\mathbf{S}}_{\text{SICA/TICA}}$) are derived as $\hat{\mathbf{S}}_{\text{SICA}} = \hat{\mathbf{W}}_{\text{SICA}}\mathbf{X}$ or $\hat{\mathbf{S}}_{\text{TICA}} = \hat{\mathbf{W}}_{\text{TICA}}\mathbf{X}^T$, while $\hat{\mathbf{S}}_{\text{SICA}}$ is an estimation of $\tilde{\mathbf{S}}$ in Eq. (4.10) and $\hat{\mathbf{S}}_{\text{TICA}}$ estimates $\mathbf{S}$ in Eq. (4.11). The mixing matrix Eq. (4.8) is computed by $\hat{\mathbf{A}}_{\text{SICA/TICA}} = \hat{\mathbf{W}}^{\dagger}_{\text{SICA/TICA}}$. This procedure is reflected in Algorithm 1. In Section 4.5.3.1, this algorithm will be extended to a complex case.

Algorithm 1: ICA algorithm based on diagonalization of the fourth-order cumulants

```
Data: $\mathbf{X}$
Result: source signals $\mathbf{S}$, and mixing matrix $\mathbf{A}$
Default: $\xi = 10^{-2}/\sqrt{(T)}$, $T$ is the length of time series;
1: Estimate a whitening matrix $\hat{\mathbf{W}}$ via eigenvalue decomposition of the sample auto-covariance matrix $\hat{\mathbf{W}}_{\text{Spatial}} = \mathbf{\Lambda}^{-1}\bar{\mathbf{P}}^T$ or $\hat{\mathbf{W}}_{\text{Temporal}} = \mathbf{\Lambda}^{-1}\mathbf{E}^T$ ;
2: Compute de-correlated components by setting $\mathbf{Z}_{\text{Spatial}} = \hat{\mathbf{W}}_{\text{Spatial}}\mathbf{X}$ or $\mathbf{Z}_{\text{Temporal}} = \hat{\mathbf{W}}_{\text{Temporal}}\mathbf{X}^T$ ;
3: Estimate the cumulant tensor of $\mathbf{Q_Z}$ by inserting the entries of $\mathbf{Z}$ in Eq. (4.20) and order it in Eq. (4.22);
4: Find a rotation matrix $\hat{\mathbf{R}} = \prod_i \hat{\mathbf{V}}_i$ as the minimizer of the squared off-diagonal cumulant entries of Eq. (4.22);
while Sweep each cumulant matrix of Eq. (4.23) in Eq. (4.24) and derive $\hat{\mathbf{V}}_i(\theta)$ do
    if off-diagonal elements are zero then
        Select another pair;
    else
        $\theta$ is smaller than $\theta_{min} = \xi$;
        Select another pair;
    end
end
5: Estimate the mixing matrix of $\hat{\mathbf{W}}_{\text{SICA}} = \hat{\mathbf{R}}^T\hat{\mathbf{W}}_{\text{Spatial}}$ or $\hat{\mathbf{W}}_{\text{TICA}} = \hat{\mathbf{R}}^T\hat{\mathbf{W}}_{\text{Temporal}}$ ;
6: Estimate the source signals of $\hat{\mathbf{S}}_{\text{SICA}}$ as $\hat{\mathbf{W}}_{\text{SICA}}\mathbf{X}$ or $\hat{\mathbf{S}}_{\text{TICA}}$ as $\hat{\mathbf{W}}_{\text{TICA}}\mathbf{X}^T$, and $\hat{\mathbf{A}}_{\text{SICA/TICA}} = \hat{\mathbf{W}}^{\dagger}_{\text{SICA/TICA}}$, where $\dagger$ can be inverse or pseudo-inverse operator.
```

4.5.2 Alternative ICA Criteria Based on Joint Diagonalization of Cumulants

The ICA algorithm, described in Section 4.5.1, can be summarized as follows:

1. De-correlate (whiten) the observations using PCA or SVD;
2. Calculate the fourth-order cumulant tensor using Eq. (4.20) and build the objective criterion Eq. (4.22) or Eq. (4.24);
3. Use the joint diagonalization approach to find an optimum rotation matrix,
4. Compute independent sources using Eq. (4.10) or Eq. (4.11).

In fact, the formulation in Eq. (4.17) has been derived for symmetrically distributed n-dimensional random vectors. Such assumption is valid since the observed time series have been de-correlated by applying the PCA technique in the first step of the ICA algorithm. As was illustrated in Figs. 3.2 and 4.1, as a result of whitening, the joint distribution of the PCA-derived components is expected to be symmetric. This assumption might, however, be arguable for the applications in which one could be sure that the source signals are significantly non-symmetric and the level of noise is considerably low. Therefore, for these special cases, one possibility is to use the third-order cumulants to extract independent components, instead of Eq. (4.22). The third-order cumulant has the advantage that it offers the possibility to account for observations with non-symmetric, skew distributions (Mendel, 1991). The third-order cumulant, corresponding to multi-variate time series, is written as

$$C(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) = E(\mathbf{x}_i\mathbf{x}_j\mathbf{x}_k). \tag{4.27}$$

In step 2 of the above algorithm, instead of using Eq. (4.20) to derive the objective function in Eq. (4.22), one can use Eq. (4.27) and estimate as statistically independent as possible source signals with respect to the third-order statistical cumulants.

This idea is further extended by requiring joint diagonality of different order cumulants as proposed by the Extended JADE algorithm in Moreau (2001). Blaschke and Wiskott (2004) proposed an ICA algorithm 'CuBICA' that simultaneously takes the third- and fourth-order cumulant tensors into account. Their results indicated that in those cases where most of the sources were skew-symmetrically distributed, using the third-order cumulant tensor or adding it as an extra information along with fourth-order cumulant tensor helps the performance of ICA algorithm.

When periodicity is strong in the observed mixtures, another source assumption can be made to improve the signal separation algorithms. This can be realized by replacing step 1 of the ICA algorithm by the EEOF/MSSA approach of Section 3.3. Recovering source signals from lagged auto-covariance and lagged cumulants are addressed in the context of the FOBIUM algorithm (Fourth-Order Blind Identification of Underdetermined Mixtures of Sources, Ferréol et al., 2005). Applying EEOF/MSSA, however, requires a suitable selection of the lag value as was mentioned in Section 3.3. In this thesis, a complex extension of the ICA algorithm is introduced to deal with these cases, which is addressed in the following section.

4.5.3 Complex ICA by Joint Diagonalization of the Fourth-Order Cumulant Tensor

It was shown in Section 3.4 that a complex extension of the PCA technique (CEOF) can improve the identification of non-stationary patterns. In this section, for the first time, the same concept has been transferred to the ICA technique by joint diagonalizing the fourth-order cumulants of a generated complex input data. We expect that this extension, called 'Complex ICA (CICA)', is useful to extract patterns that vary in space and time.

Similar to the Hilbert-Complex EOF (H-CEOF) method, first, a complex data matrix $\mathbf{Y}$ is generated by considering the data matrix $\mathbf{X}$ as the real part of $\mathbf{Y}$, and the Hilbert transform of Eq. (3.40) as its imaginary part ($\mathbf{Y} = \mathbf{X} + i\ \mathcal{H}(\mathbf{X})\,, i = \sqrt{-1}$). In this way, the new data matrix $\mathbf{Y}$ contains information about the observed time series ($\mathbf{X}$), as well as the rate of change of $\mathbf{X}$ (introduced by $\mathcal{H}(\mathbf{X})$). The complex orthogonal components of $\mathbf{Y}$ are computed by SVD (Eq. (3.23)) of the auto-covariance matrix (Eq. (3.41)). This yields a set of orthonormal complex matrices $\mathbf{E}$ and $\mathbf{P}$, as well as a set of real singular values $\mathbf{\Lambda}$. Similar to the ICA algorithm in the real case, the complex $\mathbf{E}$ and $\mathbf{P}$ should be used in Eq. (4.20) to generate the fourth-order cumulants based on the complex values as

$$\begin{aligned} C(\mathbf{x}_i, \mathbf{x}_j^*, \mathbf{x}_k, \mathbf{x}_l^*) &= E(\mathbf{x}_i\mathbf{x}_j^*\mathbf{x}_k\mathbf{x}_l^*) \\ &- E(\mathbf{x}_i\mathbf{x}_j^*)E(\mathbf{x}_k\mathbf{x}_l^*) - E(\mathbf{x}_i\mathbf{x}_k)E(\mathbf{x}_j^*\mathbf{x}_l^*) - E(\mathbf{x}_i\mathbf{x}_l^*)E(\mathbf{x}_j^*\mathbf{x}_k)\ , \end{aligned} \tag{4.28}$$

where $E(.)$ is the expectation operator, * represents the complex conjugate, and $\mathbf{x}$ represents a column of $\mathbf{E}$ or $\mathbf{P}$. Accordingly, the fourth-order cumulant tensor (Eq. (4.20)) is extended to the complex case, using the same arbitrary matrix $\mathbf{M}$ as

$$\mathbf{Q}(\mathbf{M}) = E\left((\mathbf{x}^H\mathbf{M}\mathbf{x})(\mathbf{x}\mathbf{x}^H)\right) - \mathbf{C}\ \mathrm{trace}(\mathbf{M}\mathbf{C}) - \mathbf{C}(\mathbf{M} + \mathbf{M}^T)\mathbf{C}, \tag{4.29}$$

with $\mathbf{C}$ being derived from Eq. (3.41) and H represents the Hermitian transpose (see Section 3.4). The required rotation matrix for diagonalizing (Eq. (4.29)) can be computed using, e.g,, the joint diagonalization (JD) approach described in Cardoso and Souloumiac (1995).

Similar to the real case (Eq. (4.10)), the Spatial ICA for the complex extension (S-CICA method) is derived from

$$\tilde{\mathbf{S}} = \mathbf{R}_j^H \mathbf{\Lambda}_j^{-1} \bar{\mathbf{P}}_j^H \mathbf{Y}_j = \mathbf{W}\mathbf{Y}_j, \tag{4.30}$$

in which the rows of $\tilde{\mathbf{S}}$ (containing complex entries) are spatially mutually as independent as possible. The mixing matrix $\mathbf{A}$ is estimated as $\mathbf{A} = \mathbf{W}^\dagger = (\mathbf{R}_j^H \mathbf{\Lambda}_j^{-1} \bar{\mathbf{P}}_j^H)^\dagger$, where † is the pseudo-inverse operator. The rotated complex PCs of S-CICA are still orthogonal (see the property of SICA in Section 4.3).

Source signals in the Temporal Complex ICA method (T-CICA) are computed from

$$\mathbf{S} = \mathbf{R}_j^H \mathbf{\Lambda}_j^{-1} \mathbf{E}_j^H \mathbf{Y}_j^H = \mathbf{W}\mathbf{Y}_j^H, \tag{4.31}$$

in which the rows of $\mathbf{S}$ are temporally mutually as independent as possible. Similar to Eq. (4.30), entries of $\mathbf{S}$ are complex values. The mixing matrix is estimated as $\mathbf{A} = \mathbf{W}^\dagger = (\mathbf{R}_j^H \mathbf{\Lambda}_j^{-1} \mathbf{E}_j^H)^\dagger$. Similar to the TICA formulation in Section 4.3, the rotated complex EOFs are still orthogonal.

The entries of the source and mixing/de-mixing components of both S-CICA and T-CICA are complex numbers. Therefore, one needs to compute spatial and temporal amplitudes and their associated phase patterns. For example, the spatial amplitude of the T-CICA case is derived as $\mathbf{w}_i \bullet \mathbf{w}_i^*$, where $\mathbf{w}_i$ is the i'th row of the de-mixing matrix, $i \in \{1, \ldots, \min(n,p)\}$, and '$\bullet$' provides an element-by-element multiplication product of the rows. The phase values are computed from $\mathbf{\Theta}_i = \arctan\left[\frac{\mathrm{Im}(\mathbf{w}_i)}{\mathrm{Re}(\mathbf{w}_i)}\right]$, where Im(.) and Re(.) respectively extract the imaginary and real parts.

The temporal amplitude and phase values are respectively derived from $\mathbf{s}_i \bullet \mathbf{s}_i^*$, and $\mathbf{\Phi}_i = \arctan\left[\frac{\mathrm{Im}(\mathbf{s}_i)}{\mathrm{Re}(\mathbf{s}_i)}\right]$, where $\mathbf{s}_i$ is the i'th row of the source matrix $\mathbf{S}$ in Eq. (4.31). Table 4.1 summarizes the computation of the source and de-mixing matrices of the introduced signals separation methods in Chapters 3 and 4.

4.5.3.1 Spatial and Temporal Complex ICA Algorithms

In this section, the ICA algorithm of the real case (Algorithm 1) is extended to a complex case in the form of Algorithm 2. Using Algorithm 2, one can extract complex independent components from a complex data matrix $\mathbf{Y}$.

Similar to Section 4.5.1.2, there are two possibilities for extracting independent patterns, from which the one that provides spatially independent complex patterns is called Spatial Complex ICA (S-CICA). The alternative technique whose outputs contain temporally independent complex patterns is called Temporal Complex ICA (T-CICA).

An optimum rotation matrix in Algorithm 2 is found by joint diagonalization of the fourth-order cumulant tensor (that contains complex entries in this case). The complex field $\mathbf{Y}$ is generated by adding the Hilbert transform of Eq. (3.40) as the imaginary part to the observed data matrix $\mathbf{X}$ ($\mathbf{Y} = \mathbf{X} + i\,\mathcal{H}(\mathbf{X})$). This is reflected in the stage 1 of Algorithm 2. In stage 2, $\mathbf{Y}$ is de-correlated using the SVD decomposition, which provides the complex components of $\bar{\mathbf{P}}$ and $\mathbf{E}$. In stage 3, one must select one of the transformations $\mathbf{Z}_{\text{Spatial}} = \hat{\mathbf{W}}_{\text{Spatial}}\mathbf{Y}$ or $\mathbf{Z}_{\text{Temporal}} = \hat{\mathbf{W}}_{\text{Temporal}}\mathbf{Y}^H$. In stage 4, the fourth-order cumulant tensor $\mathbf{Q}$ is built (using Eq. (4.20)) while considering the entries of $\mathbf{Z}$ as its inputs, and in stage 5, an orthogonal joint diagonalizer $\hat{\mathbf{R}}$ is computed.

The required de-mixing matrix of $\hat{\mathbf{W}}_{\text{S-CICA}}$ or $\hat{\mathbf{W}}_{\text{T-CICA}}$ to extract spatially or temporally independent components ($\hat{\mathbf{S}}_{\text{S-CICA}}$ or $\hat{\mathbf{S}}_{\text{T-CICA}}$) is computed in stage 6, using the orthogonal rotation $\hat{\mathbf{R}}$. The desired source patterns are the outputs of Algorithm 2 that are defined in stage 7. In Algorithm 2, $\hat{\mathbf{S}}_{\text{S-CICA}}$ and $\hat{\mathbf{S}}_{\text{T-CICA}}$ approximate the source signals in Eqs. (4.30) and (4.31), respectively. The mixing matrix is defined as $\hat{\mathbf{A}}_{\text{SICA/TICA}} = \hat{\mathbf{W}}^\dagger_{\text{SICA/TICA}}$.

Algorithm 2: Complex ICA algorithm based on diagonalization of the fourth-order cumulants

```
Data: $\mathbf{X}$
Result: complex source signals $\mathbf{S}$, and complex mixing matrix $\mathbf{A}$
Default: $\xi = 10^{-2}/\sqrt{(T)}$, $T$ is the length of time series;
1: Add the Hilbert transformation of $\mathbf{X}$ to the original data and derive the complex data $\mathbf{Y}$ (Eq. (3.40));
2: Estimate a whitening matrix $\hat{\mathbf{W}}$ via eigenvalue decomposition of Eq. (3.37) $\hat{\mathbf{W}}_{\text{Spatial}} = \mathbf{\Lambda}^{-1}\bar{\mathbf{P}}^H$ or $\hat{\mathbf{W}}_{\text{Temporal}} = \mathbf{\Lambda}^{-1}\mathbf{E}^H$ ;
3: Compute de-correlated components by setting $\mathbf{Z}_{\text{Spatial}} = \hat{\mathbf{W}}_{\text{Spatial}}\mathbf{Y}$ or $\mathbf{Z}_{\text{Temporal}} = \hat{\mathbf{W}}_{\text{Temporal}}\mathbf{Y}^H$ ;
4: Estimate the cumulant matrix $\mathbf{Q}_{\mathbf{Z}}$ using the entries of $\mathbf{Z}$ Eq. (4.29) and order it in Eq. (4.22);
5: Find a rotation matrix $\hat{\mathbf{R}} = \prod_i \hat{\mathbf{V}}_i$ as the minimizer of the squared off-diagonal cumulant entries of Eq. (4.22);
while Sweep each cumulant matrix of Eq. (4.23) in Eq. (4.24) and derive $\hat{\mathbf{V}}_i(\theta)$ do
    if off-diagonal elements are zero then
        Select another pair;
    else
        $\theta$ is smaller than $\theta_{min} = \xi$;
        Select another pair;
    end
end
6: Estimate the mixing matrix of $\hat{\mathbf{W}}_{\text{S-CICA}} = \hat{\mathbf{R}}^H\hat{\mathbf{W}}_{\text{Spatial}}$ or $\hat{\mathbf{W}}_{\text{T-CICA}} = \hat{\mathbf{R}}^H\hat{\mathbf{W}}_{\text{Temporal}}$ ;
7: Estimate the source signals of $\hat{\mathbf{S}}_{\text{S-CICA}}$ as $\hat{\mathbf{W}}_{\text{S-CICA}}\mathbf{X}$ or $\hat{\mathbf{S}}_{\text{T-CICA}}$ as $\hat{\mathbf{W}}_{\text{T-CICA}}\mathbf{X}^H$, and $\hat{\mathbf{A}}_{\text{SICA/TICA}} = \hat{\mathbf{W}}^{\dagger}_{\text{SICA/TICA}}$, where $\dagger$ can be inverse or pseudo-inverse operator.
```

4.5.4 ICA Based on Entropy

For sake of completeness, this section briefly introduces another approach for measuring non-Gaussianity, which is the basic idea of the FastICA algorithm in Hyvärinen (1999b). The alternative approach is derived by the definition of entropy and negentropy in Hyvärinen and Oja (2000). The entropy H of a continuous valued random vector x with density function $f(x)$ is defined as (Hyvärinen and Oja, 2000)

$$H(x) = -\int f(x) log f(x) dx. \tag{4.32}$$

Among all random variables of equal variances, the quantity $H(x)$ of a variable with Gaussian distribution is larger than the entropy of non-Gaussian variables (Koch, 1990, page 17). To obtain a measure of non-Gaussianity that is zero for a Gaussian variable and always positive for non-Gaussian variables (e.g., Hyvärinen and Oja, 2000) defined negentropy $N(x)$ as follows

$$N(x) = H(x_G) - H(x), \tag{4.33}$$

where x_G is a Gaussian variable with same mean and variance as x. The advantage of using Eq. (4.33) as a measure of non-Gaussianity is that: it is not limited to a certain statistical moment. In practice, however, measuring Eq. (4.33) is computationally very difficult, since one needs to estimate the probability density function $f(x)$ to derive $H(x)$ from Eq. (4.32). Hence, the negentropy of Eq. (4.33) is commonly approximated by higher order cumulants as

$$N(x) \simeq \frac{1}{12}(E(x^3))^2 - \frac{1}{48}(\kappa_4(x))^2, \tag{4.34}$$

where x is assumed to be zero mean and unit variance (Hyvärinen and Oja, 2000). For symmetrically distributed random variables x, the approximation in Eq. (4.34) is equivalent to that of the fourth-order cumulants criterion (Eq. (4.22)). Therefore, theoretically, the results of both entropy- and cumulant-based ICA algorithms should be similar, while considering the approximation of Eq. (4.34). Hyvärinen and Oja (2000), however, concluded that the validity of Eq. (4.34) might be rather limited. In particular, when the length of observations is short, the kurtosis in Eq. (4.34) is usually estimated with a bias, and subsequently the approximation will not perform well.

To avoid the problems encountered with the approximation of Eq. (4.34), Hyvärinen (1999b) developed a new approximation, which replaces the cumulants in Eq. (4.34) with a non-quadratic function $G(.)$ as

$$N(x) \approx |E(G(x)) - E(G(v))|^2, \tag{4.35}$$

where v is a Gaussian variable with zero mean and unit variance. The approximation (Eq. (4.35)) is basically a generalization of the moments that was used in Eq. (4.34). Hyvärinen (1999b) selected the function $G(.)$ in a way that it does not grow too fast, e.g., $G(x)$ =log(cosh(x)), or $G(x) = -e^{-x^2/2}$. The FastICA algorithm (Hyvärinen, 1999b), therefore, is based on a fixed-point iteration scheme for finding a transformation of the observed random variables $\mathbf{w}^T\mathbf{x}$, considering x as the entries of the random vector $\mathbf{x}$, that maximizes the non-Gaussianity in Eq. (4.35). Consider the fact that the variance of $\mathbf{w}^T\mathbf{x}$ must be constrained to unity; for whitened data, this is equivalent to constraining the norm of $\mathbf{w}$ to be unity. FastICA derives an optimum $\mathbf{w}$ by solving a Newton iteration (see details in Hyvärinen, 1999b).

4.6 Separation of Deterministic Signals

It was shown in Section 4.5 that the mathematical foundation of ICA is rooted in the theory of random signals. In this section, the application of ICA to separate deterministic signals is discussed. In this context, Kirimoto et al. (2011) provided the mathematical justification of ICA, based on the diagonalization of the fourth-order cumulants, to separate deterministic sinusoidal signals, when the length of sinusoids is infinite. In many geophysical and geodetic applications, however, understanding long-term trend in the presence of periodical components of an observed phenomenon is desirable. In this section, similar to Forootan and Kusche (2013), the proof in Kirimoto et al. (2011) is extended with mathematically showing that the same ICA algorithm will indeed perfectly separate an unknown mixture of a linear trend and sinusoids in the data, considering the length of data sets is infinite.

4.6.1 Setup of the Proof

Consider p centered observations ($\mathbf{x}_m, m = 1, 2, ..., p$) taken at n time epochs that are stored in columns of a matrix $\mathbf{X}_{n\times p}$. From Eq. (4.8), $\mathbf{X} = \mathbf{AS}$, the mixing matrix $\mathbf{A}$ and the unknown sources $\mathbf{S}$ are derived by rotation, while assuming the source signals $\mathbf{S}$ are to be estimated as mutually independent as possible (Cardoso, 1999). It is worth mentioning here that for deterministic signals, neither 'independence' and 'orthogonality' nor statistical quantities such as 'moments' can be defined, since their PDFs are not mathematically determined. In practice, however, one treats stochastic and deterministic signals equivalently, since they both have been observed (sampled). The statistical definitions, therefore, are usually based on the ergodicity hypothesis applied to the observed time series.

Get the fourth-order cumulant κ_4 as in Eq. (4.15), without distinguishing in notation between random variable X or S and their realizations x or s. The n'th statistical moment is denoted by $\mu_n = E[(x-\overline{x})^n], \overline{x} = E(x)$. If p random variables $s_m, (m = 1, 2, ..., p)$ are statistically independent, Cardoso and Souloumiac (1993) showed that Eq. (4.16) holds, thus, $\kappa_4\left(\sum\limits_{i=1}^{p} a_i s_i\right) = \sum\limits_{i=1}^{p} {a_i}^4 \kappa_4(s_i)$.

Kirimoto et al. (2011) showed analytically that when **S** includes a set of deterministic sinusoidal signals, maximizing the diagonality of fourth-order cumulants separates the signals. This means that Eq. (4.16) holds for a summation of sinusoids (s_i) with distinguishable frequencies. Their simulations showed that the ICA, based on the joint diagonalization of the fourth-order cumulant tensor, is successful to separate mixtures of sinusoidal signals with frequencies even closer than those the fast Fourier transform (FFT) resolution.

For the mathematical proof, besides assuming ergodicity, they replaced the expectation E, which should be used in Eq. (4.15) for computing μ_1, μ_2 and μ_3, by infinite-time averaging $E(.) = \lim_{n\to\infty} \frac{1}{n}\sum\limits_{l=1}^{n}(.)$. In the following, with the same assumptions, but without distinguishing between $E(.)$ and the time average in notation, the mathematical proof for separation of a linear trend and sinusoidal signals is explained.

4.6.2 Separation of Sinusoidal Signals and a Linear Trend

In Kirimoto et al. (2011), the observations **X** consist of a finite mixture of sinusoidal signals. Therefore, one only needs to show that when a trend is added to the observations in Eq. (4.8), the ICA according to Eq. (4.16) is still able to separate them.

For simplicity, the mixture is considered as $x = a_1 s_1 + a_2 s_2$, where x, s_1 and s_2 are time-dependent ($x = x(t)$, $s_1 = s_1(t)$ and $s_2 = s_2(t)$ with $t_l, l = 1, 2, ..., n$ being time steps). Here, s_1 represents a linear trend ($s_1 = b\ t$) with b being a constant value, and s_2 denotes a deterministic sinusoidal signal, e.g., $s_2 = \exp|i(2\pi f t + \varphi)|$, where $i = \sqrt{-1}$, and f and φ are an arbitrary frequency and phase-shift, respectively. Finally, a_1 and a_2 are constant values that scale the amplitude of the signals. Here, we should show that $\kappa_4(x) = \kappa_4(a_1 s_1 + a_2 s_2) = {a_1}^4 \kappa_4(s_1) + {a_2}^4 \kappa_4(s_2)$.

Without loss of generality, the temporal mean from the signals are reduced and they are scaled by their Euclidean norm, which is common in ICA algorithms (e.g., Cardoso and Souloumiac, 1993). Therefore, the mixture can be written as

$$x = a_1\left(\frac{s_1 - \overline{s}_1}{|s_1 - \overline{s}_1|}\right) + a_2\left(\frac{s_2 - \overline{s}_2}{|s_2 - \overline{s}_2|}\right). \tag{4.36}$$

In Eq. (4.36), $|.|$ represents the Euclidean norm and the temporal average is abbreviated by overline $\overline{.} = E(.) = \lim_{n\to\infty} \frac{1}{n}\sum\limits_{l=1}^{n}(.)$. Thus, $\overline{s_1 - \overline{s}_1} = \overline{s_2 - \overline{s}_2} = 0$ and $|\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}| = |\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}| = 1$. When temporal means are zero, all the odd moments disappear (e.g., $\mu_1 = \mu_3 = 0$) and Eq. (4.15) is summarized as

$$\kappa_4(x) = \mu_4(x) - 3(\mu_2(x))^2 = E\left(x^4\right) - 3\left(E(x^2)\right)^2. \tag{4.37}$$

In this equation,

$$\begin{aligned}
\mu_4(x) = E\left(x^4\right) = E&\left[\left(a_1\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)+a_2\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right)^4\right] \\
= E&\left[{a_1}^4\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^4+{a_2}^4\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^4\right. \\
&+4{a_1}^3a_2\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^3\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right) \\
&+6{a_1}^2{a_2}^2\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^2\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^2 \\
&\left.+4a_1{a_2}^3\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^3\right]
\end{aligned} \tag{4.38}$$

and

$$\begin{aligned}
\mu_2(x) = E\left(x^2\right) = E&\left[a_1\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)+a_2\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right]^2 \\
= E&\left[{a_1}^2\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^2+{a_2}^2\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^2\right. \\
&\left.+2a_1a_2\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right].
\end{aligned} \tag{4.39}$$

Since the evaluation of Eq. (4.37) with respect to s_2 is discussed in Kirimoto et al. (2011), here the focus is only on separation of a trend and a sinusoid. In case, s_1 is a linear trend (or any other function that is 'antisymmetric' with respect to its mean value), it is easy to see that for every positive value of $s_1 - \overline{s}_1$, there is a negative value. Thus, all the terms with $E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^k\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^k\right], k = 1, 3$ in Eq. (4.38) and $E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right]$ in Eq. (4.39) become zero. Therefore, Eq. (4.38) is simplified as

$$\begin{aligned}
\mu_4(x) = {a_1}^4\ E&\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^4\right]+{a_2}^4\ E\left[\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^4\right] \\
&+6{a_1}^2{a_2}^2\ E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^2\right]\ E\left[\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^2\right],
\end{aligned} \tag{4.40}$$

while Eq. (4.39) becomes

$$\mu_2(x) = {a_1}^2\ E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^2\right]+{a_2}^2\ E\left[\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^2\right]. \tag{4.41}$$

When the trend source s_1 is not antisymmetric with respect to its mean value, those terms of $E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)^3\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right]$, $E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)^3\right]$ in Eq. (4.38) and $E\left[\left(\frac{s_1-\overline{s}_1}{|s_1-\overline{s}_1|}\right)\left(\frac{s_2-\overline{s}_2}{|s_2-\overline{s}_2|}\right)\right]$ in Eq. (4.39) are related to the cross-covariance values of the trend function s_1 and the sinusoid s_2.

Therefore, they are eventually zero, when the averaging time extends to infinite $t \rightarrow \infty$. Equation (4.37), then, becomes

$$\begin{aligned} \kappa_4(x) = {a_1}^4\ E\left[\left(\frac{s_1-\bar{s}_1}{|s_1-\bar{s}_1|}\right)^4\right] + {a_2}^4\ E\left[\left(\frac{s_2-\bar{s}_2}{|s_2-\bar{s}_2|}\right)^4\right] \\ -3{a_1}^4\left(E\left[\left(\frac{s_1-\bar{s}_1}{|s_1-\bar{s}_1|}\right)^2\right]\right)^2 - 3{a_2}^4\left(E\left[\left(\frac{s_2-\bar{s}_2}{|s_2-\bar{s}_2|}\right)^2\right]\right)^2. \end{aligned} \tag{4.42}$$

Computing the fourth-order cumulant κ_4 for each source signal separately shows that

$$\begin{aligned} \kappa_4(s_1) = \mu_4(s_1) - 3(\mu_2(s_1))^2 = \\ E\left[\left(\frac{s_1-\bar{s}_1}{|s_1-\bar{s}_1|}\right)^4\right] - 3\left(E\left[\left(\frac{s_1-\bar{s}_1}{|s_1-\bar{s}_1|}\right)^2\right]\right)^2, \end{aligned} \tag{4.43}$$

and

$$\begin{aligned} \kappa_4(s_2) = \mu_4(s_2) - 3(\mu_2(s_2))^2 = \\ E\left[\left(\frac{s_2-\bar{s}_2}{|s_2-\bar{s}_2|}\right)^4\right] - 3\left(E\left[\left(\frac{s_2-\bar{s}_2}{|s_2-\bar{s}_2|}\right)^2\right]\right)^2. \end{aligned} \tag{4.44}$$

By substituting the left side of Eq. (4.16) with the right side of Eq. (4.42) and inserting Eq. (4.43) and Eq. (4.44) on its right side, it is easy to see that Eq. (4.16) holds ($\kappa_4(x) = {a_1}^4\kappa_4(s_1) + {a_2}^4\kappa_4(s_2)$). Therefore, using the fourth-order cumulant, it is analytically shown that a linear trend can be separated from any arbitrary deterministic sinusoids.

4.6.3 Numerical Illustration

It was assumed in the analytical proof that the expectation operator $E(.)$ is replaced by temporal averaging, while the length of the time series were assumed infinite. In practice, however, the time span of time series is usually limited (e.g., restricted to the active period of a satellite mission or observation period). A simulation is designed here to examine the performance of the ICA technique in separation of a length-limited mixture of a linear trend and two sinusoidal annual and semi-annual signals as the source signals ($\mathbf{S}$) of interest. In fact, extracting linear trend (also called linear rate) and seasonal components is quite common in geodetic studies such as detecting the dominant hydrological cycles from GRACE-TWS time series or long-term sea level variations from altimetry missions. Therefore, selecting these two source signals might be of interest of various geodetic studies. An arbitrary random mixing matrix $\mathbf{A}$ was chosen to aggregate the mixed signal $\mathbf{X} = \mathbf{AS}$. Random Gaussian noise with a standard deviation equal to 20% of the source signal was also added.

The set-up and the results of the ICA algorithm with respect to the sample length of 50 and 100 are shown in the parts A and B of Fig. 4.3, respectively. The results illustrate that the theoretical performance of ICA is successfully transferred to practice with separating the introduced short length time series. On the other hand, improved performance of ICA was found when the length of samples was increased (compare the bottom graph of Fig. 4.3 (A) to the bottom graph of Fig. 4.3 (B)).

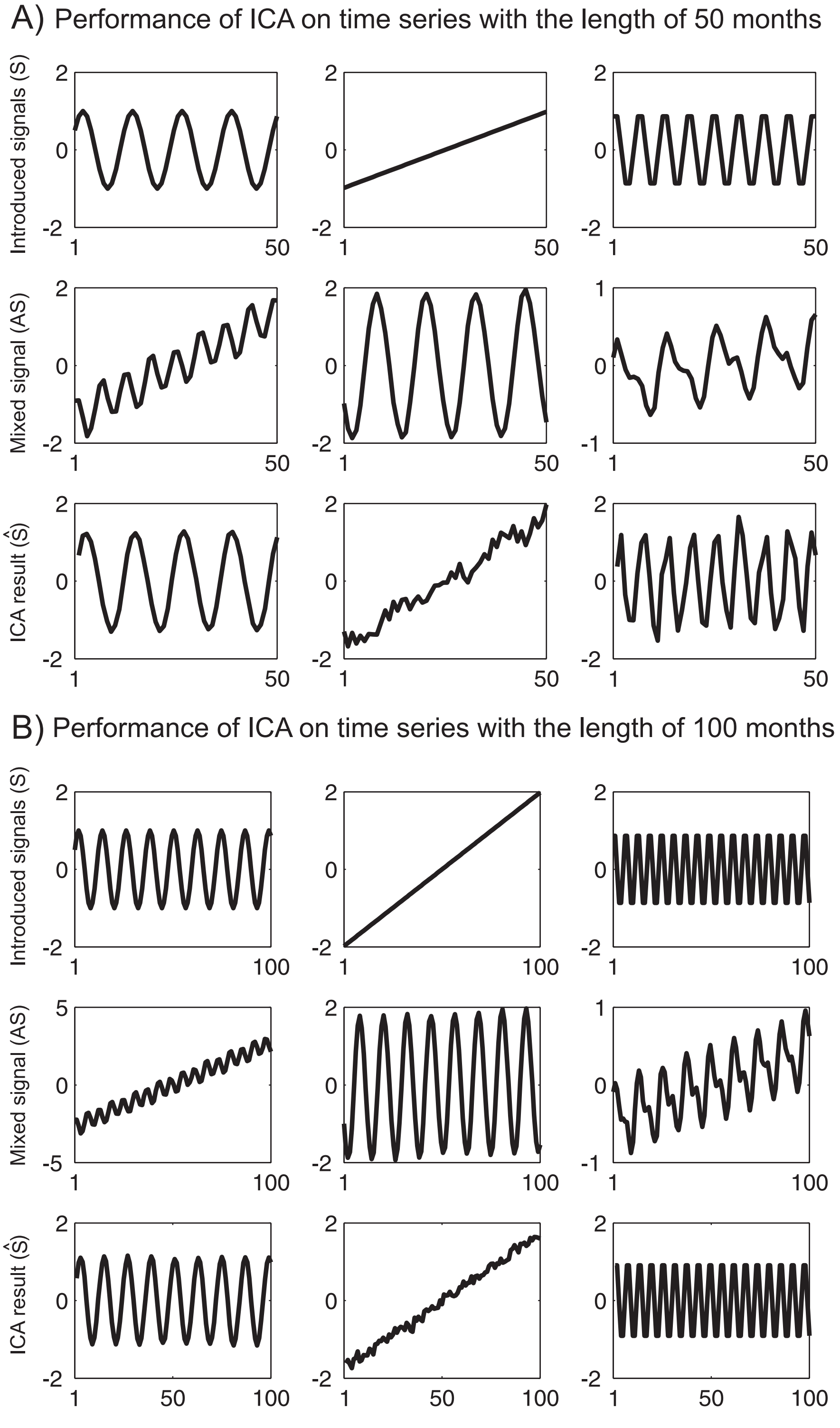


Figure 4.3: Decomposition of a mixture of a linear trend, annual and semi-annual signals in the presence of a random noise using the ICA method. Figure 4.3 (A-top): introduced samples ('sources') with 50 time steps; (A-middle): mixed signals **AS** ('observed data') and (A-bottom): ICA results $\hat{\mathbf{S}}$. Figure 4.3 (B-top): introduced samples with 100 time steps; (B-middle) mixed signals and (B-bottom) ICA results. The example is taken from Forootan and Kusche (2013).

4.7 Summary of the Introduced Statistical Decomposition Techniques

In Chapter 3 and the current chapter, various transformations were introduced, which can be used to transform the observation matrix $\mathbf{X}$ and derive components with desired statistical properties. As mentioned earlier, in the literature, for example (Preisendorfer, 1988, von Storch and Navarra, 1999), the formulation of the second order techniques usually followed the notation in SVD as $\mathbf{X} = \bar{\mathbf{P}}\boldsymbol{\Lambda}\mathbf{E}^T$. This notation was also followed by the ordinary extensions of the SVD technique such as REOF, where we derived $\mathbf{X} = \bar{\mathbf{P}}\boldsymbol{\Lambda}\mathbf{R}\mathbf{R}^T\mathbf{E}^T$. The formulations of EEOF and CEOF techniques followed that of SVD, see Sections 3.3 and 3.4 for more details. In this chapter, ICA was introduced as its BSS definition, where the observation matrix $\mathbf{X}$ is seen as a mixture of unknown sources that are mutually statistically independent ($\mathbf{X} = \mathbf{A}\mathbf{S}$). To estimate $\mathbf{S}$ from $\mathbf{X}$, one needs to compute a de-mixing matrix $\mathbf{W}$ or its pseudo-inverse $\mathbf{A}$ ($\mathbf{A} = \mathbf{W}^\dagger$). In Table 4.1, the relationships between $\mathbf{W}$ and the matrix notations of Chapter 3 are summarized. In Chapter 5, it will be shown how each of the introduced transformations can be used to extract information from multivariate time series. It is worth mentioning here that the SVD, REOF, and ICA transformation are formulated based on the stationary assumption, which means that we assume the auto-covariance matrix or the cumulants tensor does not change through the time. The MSSA and Conventional CEOF techniques incorporate lagged information in the decomposition procedure to deal with the non-stationary behavior of time series. The success of these techniques strongly depends on a proper selection of the lag information. Therefore, these techniques are not considered for the applications of this thesis. The Hilbert CEOF and Complex ICA techniques incorporate information about the rates of change of observations in the decomposition procedure. Both approaches, therefore, have the potential to deal with non-stationary time series and are considered for the applications of Chapter 5.

Method	Source $\mathbf{S}$	De-mixing Matrix $\mathbf{W}$	Details
PCA/SVD Section 3.1.3	$\tilde{\mathbf{S}} = \mathbf{E}^T = \boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X}$ $\mathbf{S} = \bar{\mathbf{P}}^T = \boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T$	$\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X} = \mathbf{W}\mathbf{X}$ $\boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T = \mathbf{W}\mathbf{X}^T$	$\bar{\mathbf{P}}$ and $\mathbf{E}$ are derived by eigenvalue decomposition of the auto-covariance matrix
Rotated EOF (Spatial) Rotated EOF (Temporal) Section 3.2	$\tilde{\mathbf{S}} = \mathbf{R}^T\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X}$ $\mathbf{S} = \mathbf{R}^T\boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T$	$\mathbf{R}^T\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X} = \mathbf{W}\mathbf{X}$ $\mathbf{R}^T\boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T = \mathbf{W}\mathbf{X}^T$	$\mathbf{R}$ is defined by solving a statistical criterion e.g., VARIMAX
Extended EOF (MSSA) Section 3.3	$\tilde{\mathbf{S}} = \mathcal{E}^T = \boldsymbol{\Theta}^{-1}\bar{\mathcal{P}}^T\mathcal{X}$ $\mathbf{S} = \bar{\mathcal{P}}^T = \boldsymbol{\Theta}^{-1}\mathcal{E}^T\mathcal{X}^T$	$\boldsymbol{\Theta}^{-1}\bar{\mathcal{P}}^T\mathcal{X} = \mathcal{W}\mathcal{X}$ $\boldsymbol{\Theta}^{-1}\mathcal{E}^T\mathcal{X}^T = \mathcal{W}\mathcal{X}^T$	$\mathcal{P}$ and $\mathcal{E}$ are derived by eigenvalue decomposition of the lagged auto-covariance matrix
Complex EOF Section 3.4	$\tilde{\mathbf{S}} = \mathbf{E}^T = \boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^H\mathbf{Y}$ $\mathbf{S} = \bar{\mathbf{P}}^H = \boldsymbol{\Lambda}^{-1}\mathbf{E}^H\mathbf{Y}^H$	$\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^H\mathbf{Y} = \mathbf{W}\mathbf{Y}$ $\boldsymbol{\Lambda}^{-1}\mathbf{E}^H\mathbf{Y}^H = \mathbf{W}\mathbf{Y}^H$	$\bar{\mathbf{P}}$ and $\mathbf{E}$ are derived by eigenvalue decomposition of the auto-covariance matrix of the complex data matrix $\mathbf{Y}$
Spatial ICA (SICA) Temporal ICA (TICA) Sections 4.3 and 4.5	$\tilde{\mathbf{S}} = \mathbf{R}^T\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X}$ $\mathbf{S} = \mathbf{R}^T\boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T$	$\mathbf{R}^T\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^T\mathbf{X} = \mathbf{W}\mathbf{X}$ $\mathbf{R}^T\boldsymbol{\Lambda}^{-1}\mathbf{E}^T\mathbf{X}^T = \mathbf{W}\mathbf{X}^T$	$\mathbf{R}$ is defined by joint diagonalization of the fourth-order cumulants
Spatial Complex ICA (S-CICA) Temporal Complex ICA (T-CICA) Section 4.5.3	$\tilde{\mathbf{S}} = \mathbf{R}^H\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^H\mathbf{Y}$ $\mathbf{S} = \mathbf{R}^H\boldsymbol{\Lambda}^{-1}\mathbf{E}^H\mathbf{Y}^H$	$\mathbf{R}^H\boldsymbol{\Lambda}^{-1}\bar{\mathbf{P}}^H\mathbf{Y} = \mathbf{W}\mathbf{Y}$ $\mathbf{R}^H\boldsymbol{\Lambda}^{-1}\mathbf{E}^H\mathbf{Y}^H = \mathbf{W}\mathbf{Y}^H$	$\mathbf{R}$ is defined by joint diagonalization of the fourth-order cumulants, and $\mathbf{Y}$ is generated by the Hilbert transform

Table 4.1: Summary of matrix relations regarding to the statistical signal separation methods introduced in Chapters 3 and 4. The data matrix $\mathbf{X}_{n\times p}$ represents p time series with the length of n. The 'source' and 'de-mixing' matrices are derived as $\tilde{\mathbf{S}} = \mathbf{W}\mathbf{X}$ or $\mathbf{S} = \mathbf{W}\mathbf{X}^T$. The mixing matrix $\mathbf{A}$ can be derived as $\mathbf{A} = \mathbf{W}^\dagger$.

4.8 Uncertainty Computation of the ICA Decomposition

After applying the ICA decomposition using one of the methods described in this chapter, one needs to estimate the uncertainty of extracted independent sources. In other words, one should compute error bars for the extracted components. Estimated uncertainties can be due to several reason such as; existing noise in observations, limited length of the observations (sample size), or inadequate modeling. Regarding the latter, it is meant that the algorithm might not be adequate for separation of variables with particular distributions. Similar to what was proposed in Section 3.6, a numerical approach to estimate the uncertainties of the leading independent components can be derived based on the Bootstrap resampling. Section 4.1 indicated that, in order to derive statistically independent modes, one can rotate whitened components derived from, for example, the PCA method. Uncertainties of the whitened components will be available from the error estimation approach in Section 3.6. Subsequently, in order to estimate the uncertainty of the independent modes, one can generate m number of realizations of $\mathbf{X}$, that are reconstructed by inserting $\bar{\mathbf{P}}_j$ and $\mathbf{E}_j$ along with m realizations of their errors in Eq. (3.23). The number of realizations m should be large enough to justify using normal theory for estimating the standard error. Applying Eq. (4.10) or (4.11) to the m realizations of $\mathbf{X}$, m realizations of the source signals $\mathbf{S}$ will be computed, that allows estimation of the uncertainties.

5. Applications of Statistical Signal Separation Techniques for Analyzing GRACE-TWS

In this chapter, some examples are presented to show how the introduced statistical signal separation approaches can be used to explore information from GRACE integrated total water storage (TWS) observations. Therefore, in each section of this chapter, first, a review of the problem and the research question is introduced. Then, the data and method are described. Finally, the results of signal separation are presented and interpreted.

In Section 5.1, the performance of the commonly used decomposition methods when they are applied to simulated GRACE-TWS changes is discussed. The application of different statistical methods to decompose real GRACE-TWS changes is presented in Section 5.2. The aim of this section is to extract the dominant independent patterns of TWS products and evaluate their errors. Application of ICA for reducing the spectral and spatial leakage problems over Australia, as well as the relationship between the ICA-extracted regional patterns of TWS changes and major climate variability of the region are discussed in Section 5.3. A new statistical approach for separation of large-scale GRACE-TWS patterns into terrestrial and surface water storage changes, as well as groundwater storage changes is discussed in Section 5.4. Finally, an ICA-based statistical forecasting approach to predict TWS changes is introduced in Section 5.5.

5.1 Decomposition of GRACE-like Simulated Total Water Storage Changes

In order to evaluate the performance of the statistical signal separation approaches described in the previous chapters, simulations of GRACE-TWS changes are generated. In each simulation, the true spatial and temporal solutions to the decomposition problem are known. Thus, one can compare the components extracted by the statistical methods with the introduced patterns and evaluate the success of each method. Here, the result of signal decomposition is accepted as 'successful', when the extracted patterns are fully comparable with the introduced TWS anomalies without generating any artificial patterns in each of the extracted modes. Otherwise, the performance of the method is referred to as 'unsuccessful'. In Section 5.1.1, the three common decomposition methods of PCA, VARIMAX-REOF and ICA (SICA and TICA) are compared. A simulated case on the extraction of localized TWS anomalies is shown in Section 5.1.2. In Section 5.1.3, the performance of the TICA, CEOF, and T-CICA methods in identifying simulated propagating patterns are compared.

5.1.1 Performance Analysis of the PCA, REOF and ICA

Similar to the study by Forootan and Kusche (2012), the simulation presented here is introduced to investigate the performance of the three most popular methods of PCA (Section 3.1.3), REOF methods (Section 3.2), and ICA (Section 4.5.1) in a controlled situation when the 'observed' signals are derived from a super-positioning process and corrupted by artificial GRACE-type correlated noise. To define reasonable spatial and temporal patterns for the simulation, 1° × 1° monthly water storage output of the Global Land Data Assimilation System (GLDAS, Rodell et al., 2004) covering the period January 2003 to August 2009 (comparable to those of GRACE-ITG2010 products) are used.

Since the long-term linear trend and annual cycle are the dominant variabilies in GRACE-TWS time series (see e.g., Wouters and Schrama, 2007, Schmidt et al., 2008a,b), for simulating the 'true' GRACE-like TWS anomalies, a superposition of a linear trend and annual cycle (with a period of 365.25-days) over South America and the annual cycle only over Africa are extracted from GLDAS-TWS using a least squares fit. This set-up should render the separation task even easier for the PCA method since the spatial and temporal differences are being exaggerated with comparison to realistic hydrological data sets. Specifically, no spatial overlaps exist between the anomalies over South America and Africa. The temporal variability is also restricted to two very distinguishable components. The set-up of the simulation is shown in Fig. 5.1.

To simulate realistic spatially correlated noise, the covariance matrix corresponding to March of 2006 from monthly ITG2010 solutions (Mayer-Gürr et al., 2010b) was used. Using Cholesky decomposition, the covariance matrix was split into an upper triangular and its conjugate transpose matrix (Koch, 1999, page 30). Then, multiplying each entry of the upper triangular matrix with a unit random vector, GRACE-type realizations of monthly errors were generated. It should be mentioned here that the length of random vectors should be equal to the length of simulated temporal patterns ($n = 79$). The generated errors in this way are spatially correlated but temporally uncorrelated. To reduce the amplitude of the noise to be comparable with the real filtered GRACE-TWS data, each of the monthly noise realizations was smoothed using a Gaussian filter (Jekeli, 1981) with 500 km half-width radius. Then, maps of simulated noise for South America and Africa are added to the monthly snapshots. Typical noise patterns are shown in Fig. 5.1 (bottom). The variance of the simulated noise was ∼ 30% of the variance of the signal in Fig. 5.1 (top).

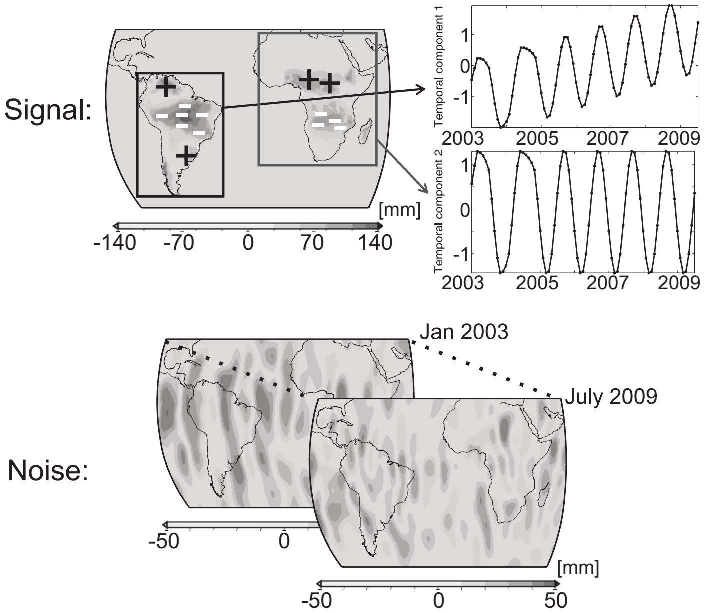


Figure 5.1: The simulation covers the period between January 2003 and July 2009: Fig. 5.1 (top) shows the introduced spatial and temporal patterns. One can reconstruct the simulated dataset by multiplying the spatial anomalies by their corresponding temporal evolutions. The data matrix **X** contains 79 rows and 12604 columns. Figure 5.1 (bottom) presents the colored noise, simulated using the covariance matrix of GRACE-ITG2010 for the same period January 2003 to July 2009 (see also Forootan and Kusche, 2012).

Results of the PCA decomposition are shown in the first row of Fig. 5.2. They clearly illustrate that the PCA method was unsuccessful in separating the predefined anomalies in its two modes, since the linear trend appears in both modes. The reason for this behavior is that the first eigenvector is oriented in the direction of maximum variance. As a result, the first component contains both linear and annual signals and the amplitude of the first mode is over-fitted for the African part since the spatial pattern has the same amplitude as what was simulated, but the corresponding temporal component shows more variability. Continuing the decomposition procedure, the following eigenvectors turn out orthogonal to the first one adding remaining signal to South America and creating an opposite anomaly over Africa. Consequently, in both of the PCA components the linear trend and annual signals prevails (see also Table 5.1).

The VARIMAX criterion orthogonally rotates the PCA-derived components to obtain a simpler structure, that in this case yields another mixture of the defined linear and annual patterns. Results of VARIMAX rotation with either rotating PCs or EOFs are respectively illustrated in the second and third row of Fig. 5.2. Rotating the EOFs shows a better performance than rotating PCs, since its first mode shows no trend over Africa and the artificial linear trend of the second mode (over Africa) was obviously smaller than the results derived from VARIMAX with rotating PCs. It should be mentioned here that changing the number of components to be rotated within the REOF method did not enhance separation for this case. In the 4th row of the figure, the results of TICA Eq. (4.11) are shown in which the first pattern on the left is related to the annual signal over South America and Africa, and the second pattern (on the right) shows the spatial distribution of the linear trend over South America. Their corresponding temporal evolution is illustrated in the second and fourth columns of Fig. 5.2, showing that the

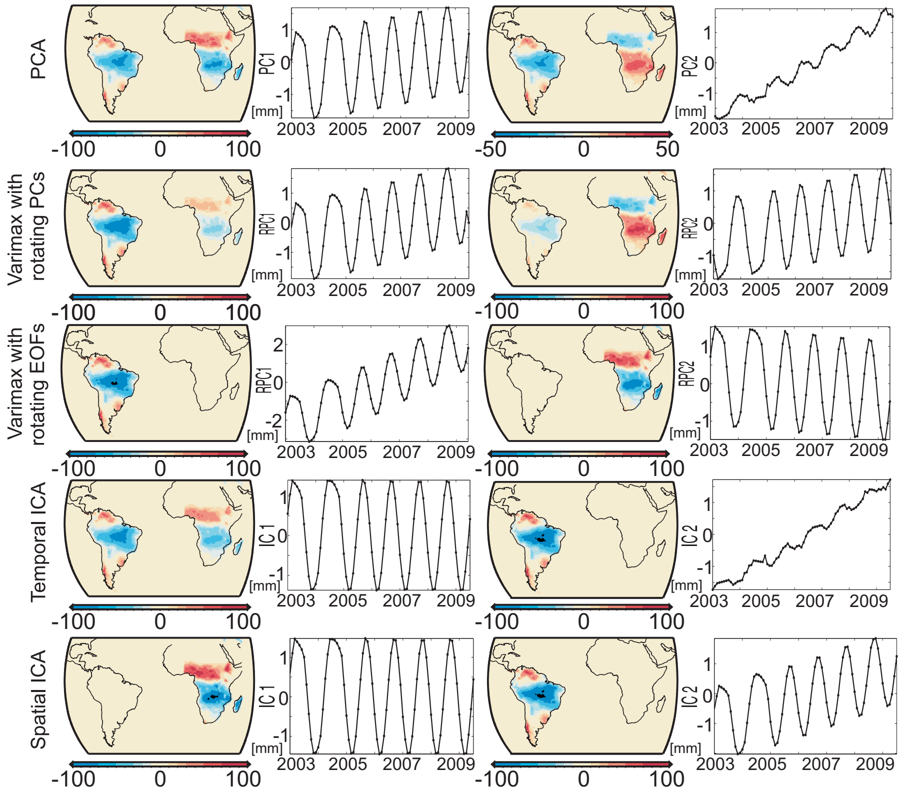


Figure 5.2: Decomposition of the simulated patterns. In this figure, the performance of PCA, VARIMAX REOF and ICA methods is compared. (First row) the derived patterns from implementing PCA, (Second row) results of VARIMAX REOF from rotating PCs, (Third row) VARIMAX REOF from rotating EOFs, (Fourth row) TICA, (Fifth row) SICA. These results are taken from Forootan and Kusche (2012).

annual cycle is separated from the linear trend. The linear component (IC2) is still contaminated by a low-amplitude annual signal. This likely happens due to the small length of the simulated time series, which does not permit the computation of the fourth order cumulants correctly. This statement is scrutinized using a longer simulation (120 months of GLDAS), and the problem indeed vanishes. A similar improvement was also seen in the separation of deterministic signals as it was exemplified in Fig. 4.3 and Forootan and Kusche (2013).

The performance of SICA (Eq. (4.10)) is also assessed here, and results are shown in the last row of Fig. 5.2. At least with the presented simple simulation set-up, SICA was able to completely separate the predefined TWS patterns. Within the simulation, it is shown that from the three methods, only ICA successfully extracts the components of a simple super-position mixture, without generating artificial patterns in its independent modes. For more details, see a summary of the decomposition results in Table 5.1, where the root mean squares (RMS) of the extracted annual and linear patterns are presented. Keeping in mind that a simulation cannot claim the general validity of a mathematical proof, this result nevertheless encourages the use of this method on the real time-variable TWS data set from GRACE, which is addressed in Section 5.2.

	South America	Africa
True Pattern	Linear Trend + Annual Cycle	Annual Cycle
PCA	Mode1: $\sim$ 80% of the annual-RMS $\sim$ 90% of the linear trend-RMS Mode2: $\sim$ 20% of the annual-RMS $\sim$ 10% of the linear trend-RMS	Mode1: $\sim$ 115% of the annual-RMS artificial linear trend $\sim$ 5 mm/year Mode2: $\sim$ 15% of the annual-RMS artificial linear trend $\sim$ 5 mm/year
VARIMAX with rotating PCs	Mode1: $\sim$ 75% of the annual-RMS $\sim$ 95% of the linear trend-RMS Mode2: $\sim$ 25% of the annual-RMS $\sim$ 5% of the linear trend-RMS	Mode1: $\sim$ 120% of the annual-RMS artificial linear trend $\sim$ 8 mm/year Mode2: $\sim$ 20% of the annual-RMS artificial linear trend $\sim$ 8 mm/year
VARIMAX with rotating EOFs	Mode1: $\sim$ 100% of the annual-RMS $\sim$ 100% of the linear trend-RMS Mode2: 0% of the annual-RMS 0% of the linear trend-RMS	Mode1: 0% of the annual-RMS no artificial linear trend Mode2: $\sim$ 100% of the annual-RMS artificial linear trend $\sim$ 3 mm/year
Temporal ICA (TICA)	Mode1: $\sim$ 100% of the annual-RMS 0% of the linear trend-RMS Mode2: $\sim$ 8% of the annual-RMS $\sim$ 100% of the linear trend-RMS	Mode1: $\sim$ 100% of the annual-RMS no artificial linear trend Mode2: 0% of the annual-RMS no artificial linear trend
Spatial ICA (SICA)	Mode1: 0% of the annual-RMS 0% of the linear trend-RMS Mode2: $\sim$ 100% of the annual-RMS $\sim$ 100% of the linear trend-RMS	Mode1: $\sim$ 100% of the annual-RMS no artificial linear trend Mode2: 0% of the annual-RMS no artificial linear trend

Table 5.1: A summary of the decomposition results derived in Section 5.1.1.

5.1.2 Performance Analysis of PCA and ICA for a Regional Case

In this section, the performance of ICA and PCA are compared for extracting mass anomalies that are spatially close to each other (see also Forootan et al., 2012). Four TWS anomaly concentrations are considered over the north, west, east and southeast of Australia. We simplify the simulation by considering that the regions have no overlap with each other. Then, we assume that the northern part exhibits only an annual signal while the other parts exhibit a superposition of a low amplitude annual signal and a linear trend (Fig. 5.3). Synthetic data is then generated by scaling the known spatial anomalies with their corresponding temporal components (see Fig. 5.3). In the previous section, one could see that the simulated noise does not appear in the decomposed results. Thus, in this simulation, no disturbance noise is considered

Now, both the PCA and SICA methods are employed to separate the signals in Fig. 5.3. The results are presented in Fig. 5.4, which again indicates that the output of the PCA method contains a mixture of both linear trend and annual signals over all regions, while SICA succeeds in extracting TWS anomalies that were introduced. The mixing behavior detected in PCA happens because the method maximizes the variance explained by each individual orthogonal component. Therefore, PCA clusters the linear and annual anomalies in both modes (see Fig. 5.4, PCA results). The SICA method, however, uses the statistical information contained in the fourth order cumulants to rotate the PCA components so that they are as statistically (spatially) independent as possible. As a result, the introduced anomalies (Fig. 5.3) that have different probability density characteristics (since they correspond to the trend and annual cycle) are recovered in two different modes (see Fig. 5.4, ICA results).

Therefore, one can conclude that when spatially distinguishable TWS anomalies exist in the data, SICA is able to extract them fairly well. This means that the overlap between the storage anomalies has to be distinguishable. In Section 5.3, SICA is applied to separate water storage anomalies from real GRACE-TWS time series over Australia. Application of TICA is, however, more interesting when the goal of the analysis is to extract patterns with different temporal behavior (time-scale), for instance, extracting the El Niño pattern from time series of GRACE-TWS products. An example can be found in Section 5.2.

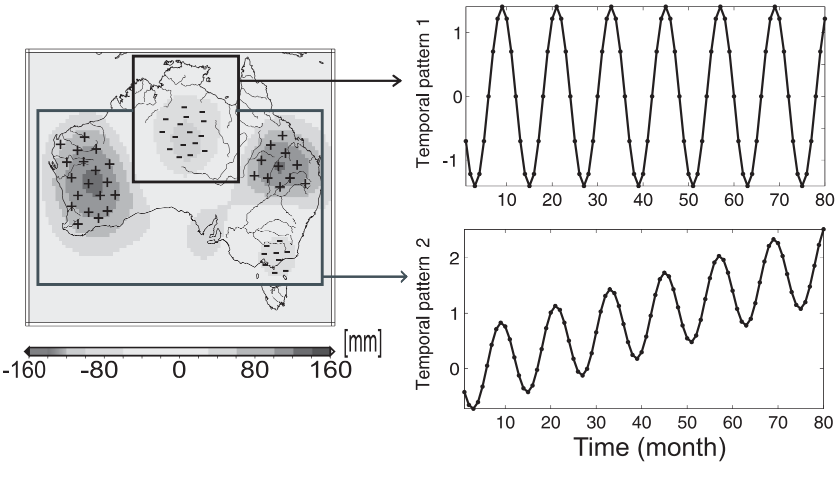


Figure 5.3: A synthetic example; TWS in the north of Australia is simulated to exhibit only an annual signal while the west, east and southeast contain a superposition of a weaker annual signal and a linear trend. In order to reconstruct the synthetic data set, one should multiply the spatial patterns (left) by the temporal components (right). This example was presented by Forootan et al. (2012).

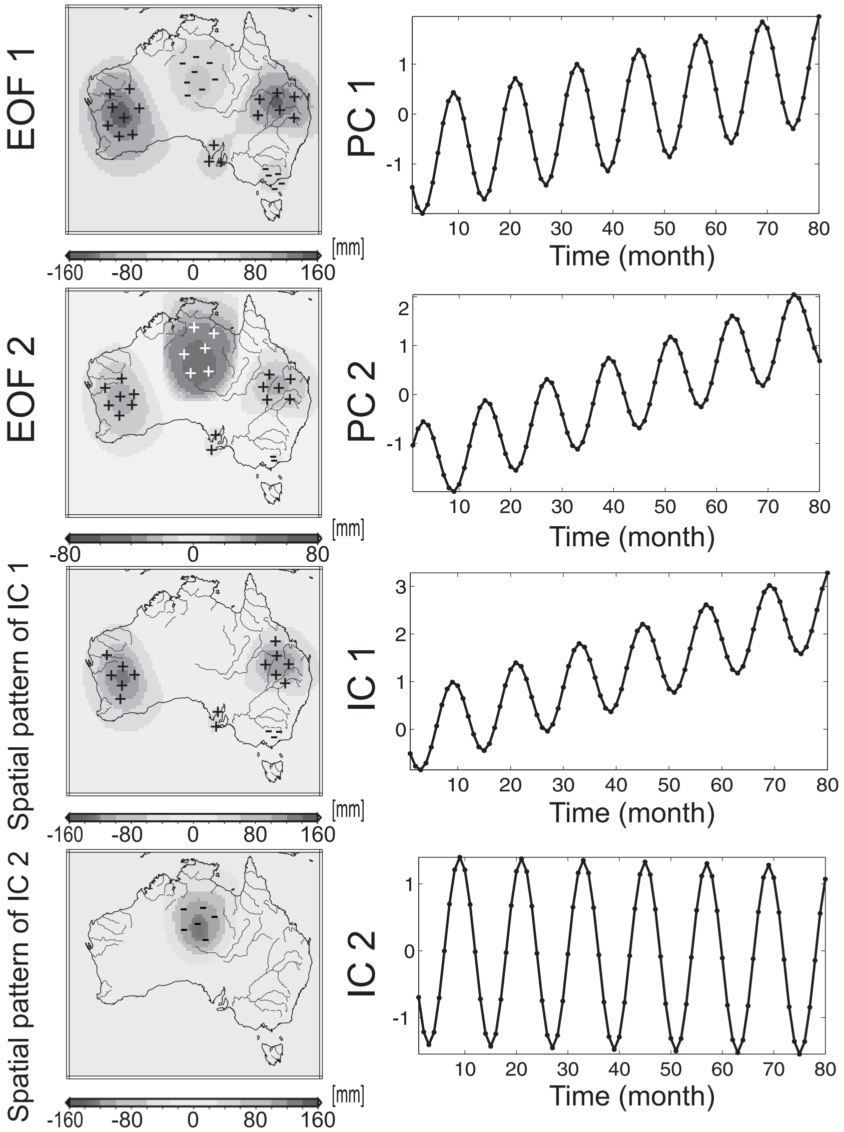


Figure 5.4: Separation of the simulated hydrological signals of Fig. 5.3 using PCA and ICA methods. The first two rows are related to the PCA results while the last 2 rows are related to the SICA results (see also Forootan et al., 2012).

5.1.3 Performance Analysis for a Case with Propagating Signals

In Sections 5.1.1 and 5.1.2, the benefit of using higher order statistics to decompose water storage time series was demonstrated. In both simulations, however, the introduced water storage patterns were stationary (e.g., the linear trend) or cyclostationary (e.g., the sine and cosine components). In this section, a non-cyclostationary case is simulated to assess the performance of the introduced separation methods.

An approximation of the El Niño Southern Oscillation (ENSO) pattern in TWS changes is selected here as our non-cyclostationary case. Our motivation for this selection is twofold: (i) ENSO represents an important impact on global water storage changes (e.g., de Viron et al., 2006), therefore, being able to isolate its fingerprints in water storage time series is important for studies that address the trend of global water storage variations or those studies that address the relationships between water storage changes and climate variability (Enfield and Mestas-Nuñez, 1999). (ii) Most of previous studies, for example Rangelova et al. (2010) and García-García et al.

(2011), demonstrated the application of extended statistical methods, e.g., EEOF (Section 3.3) and CEOF (Section 3.4), for extracting cyclic patterns. This means that they investigated the propagation of the sinus-like patterns or their modulations. Considering the temporal pattern of ENSO, one can see the complex behavior of the phenomenon that includes different periods (e.g., ~ 5 and ~ 11 years), as well as its time-varying amplitude (extreme strength during some periods, e.g., that of 2010 and less active ENSO during 2003).

We assume that the temporal variability of ENSO is known, for instance, from the Southern Oscillation Index (SOI), provided by the Australian Bureau of Meteorology (http://www.bom.gov.au/climate/enso/). The solid line in Fig. 5.5 represents the SOI. The difference between SOI and its values after introducing 3 months lag (3 months temporal shift) is shown by the dashed line in Fig. 5.5. The amplitude of the normalized SOI was co-estimated in a least squares fit along with annual and semi-annual components from the $1^\circ \times 1^\circ$ monthly GLDAS-derived water storage values (Rodell et al., 2004) covering the period of January 2002 to January 2013. The amplitudes are shown in Fig. 5.6 (left).

In order to simulate the propagation, water storage of the equatorial regions are considered to immediately respond to ENSO. The other regions of the globe, however, respond with some temporal delay. The delay has been simulated as a simple lateral phase delay that is shown in Fig. 5.6 (right). Multiplying the amplitudes of Fig. 5.6 (left) by the temporal pattern of SOI in Fig. 5.5, while considering the phase lags of Fig. 5.6 (right), one can simulate a propagated ENSO pattern. Thus, the simulated values of TWS changes $x(l,p,t)$ at latitude l, longitude p, and time t are derived as $x(l,p,t) = a(l,p)\ s(t + \phi(l,p))$, where $a(l,p)$ are the grid values in Fig. 5.6 (left), $\varphi(l,p)$ are simulated delay values in Fig. 5.6 (right), and $s(t, \phi(l,p))$ are derived from the SOI pattern in Fig. 5.5. It is worth mentioning that the simulated pattern is much simpler than what happens in reality. For instance, the spatial propagation of the signal is not considered here. This simplicity, however, helps to better interpret the performance of the introduced decomposition methods.

From the methods that were applied in Section 5.1.1, we chose TICA to decompose the simulated data due to its better performance in extracting patterns that exhibit different temporal behaviors (see Fig. 5.2). The blue curve in Fig. 5.7 shows the temporal pattern of the first dominant independent component, which is located between the SOI at zero lag and three-months lag. The second dominant pattern of TICA, shown in Fig. 5.8, also represents a part of SOI, which was not captured by IC1. This can be clearly seen by comparing the IC2 pattern and that of SOI shown by the dashed yellow and red lines (see, e.g., the years 2004 and 2010). This clustering behavior happens due to the introduced phase lag. For reasons of brevity, in this section, we only discuss the extracted temporal components.

The CEOF approach (Section 3.4) and T-CICA (Section 4.5.3) were also applied to decompose the simulated data. The results show that both methods capture the SOI pattern fairly well, i.e. the amplitude derived from Eq. (3.44) is similar to the SOI pattern in Fig. 5.5. We do not show the extracted amplitudes derived from CEOF and T-CICA. To evaluate the performance of the methods, Fig. 5.9 (top) shows the temporal phase of the SOI with respect to the 1, 2 and 3 months lags. To estimate phase patterns for SOI, $\omega_i(t) = \arctan\left[\frac{s(t)}{s(t+\varphi_i))}\right]$, with t being time, three times were evaluated. In this equation, $s(t)$ is the SOI pattern in Fig. 5.5, φ_i $(i = 1,2,3)$ are lags of 1, 2 and 3 months, while $\omega_i(t)$ are their corresponding temporal phase that are shown in Fig. 5.9 (top). In Fig. 5.9 (bottom), the temporal phase derived from CEOF is shown by the dashed line, and that of T-CICA is presented by the black solid line. As a result, it appears that the T-CICA-derived temporal phase is more similar to the simulated values (compare them to the graph in Fig. 5.9 (top)). Therefore, the results of this simulation indicate that when the

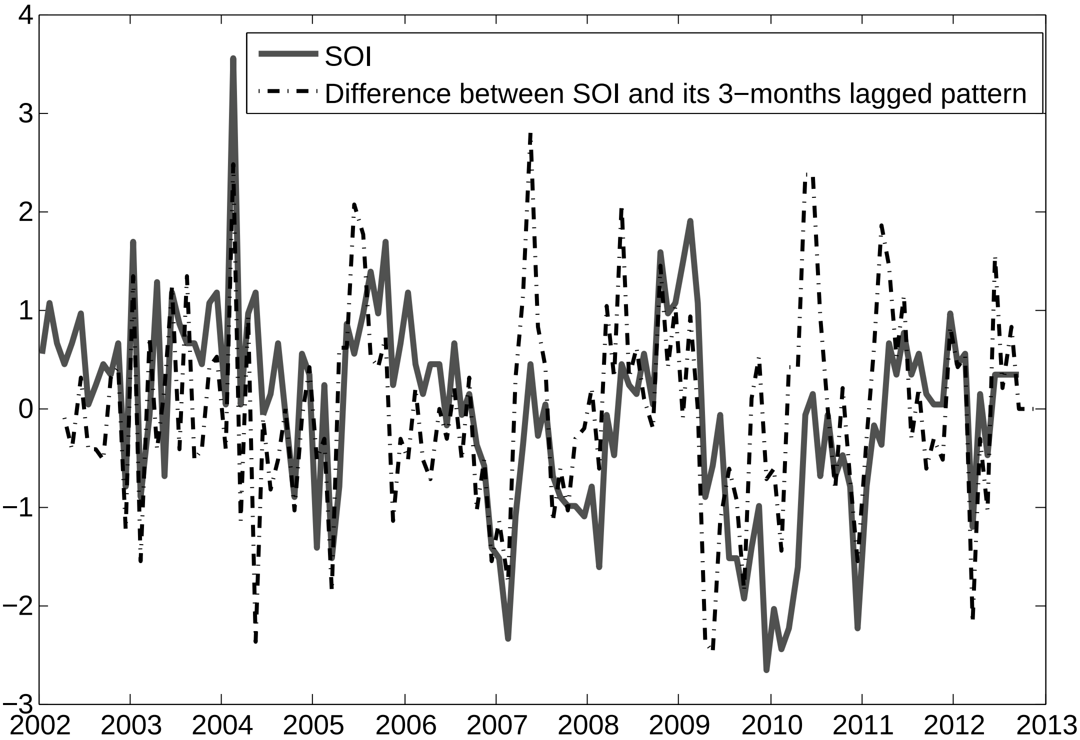


Figure 5.5: The solid line represents the normalized pattern of the ENSO index (Southern Oscillation Index, SOI) derived from the Australian Bureau of Meteorology. The dashed line presents the difference between SOI and its pattern after 3 months temporal shift.

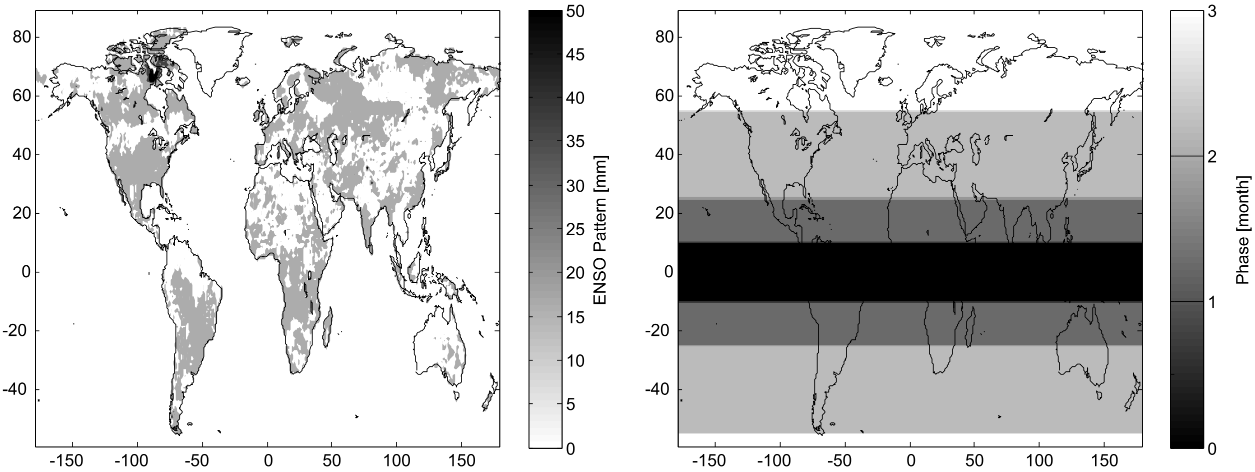


Figure 5.6: Simulated amplitude and phase of ENSO in global TWS changes. Figure 5.6 (left) shows the amplitude of ENSO derived by fitting the normalized SOI pattern along with the annual and semi-annual cycles to the GLDAS-derived water storage changes (Rodell et al., 2004) over the period of January 2002 to July 2013. Figure 5.6 (right) presents the assumed phase delay between the occurrence of the ENSO phenomenon and global water storage changes. The selected phase pattern is considerable simpler than its complex behavior in reality. This selection, however, helps to better interpret the performance of the applied statistical decomposition methods. The temporal pattern of SOI is shown in Fig. 5.5.

propagating pattern is not fundamentally cyclic, adding fourth-order statistical information into the analysis improves the performance of signal separation.

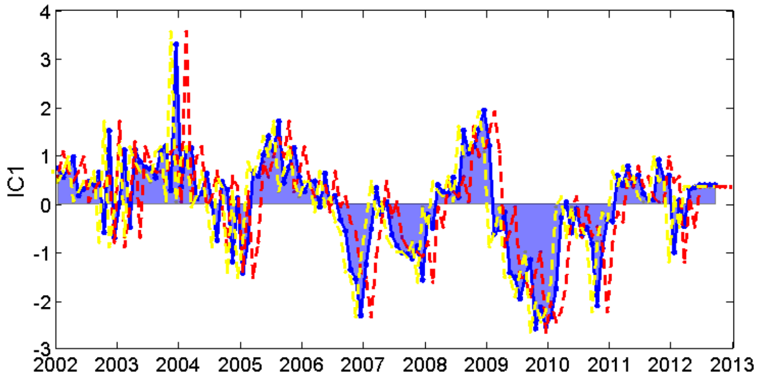


Figure 5.7: The blue curve shows the first dominant independent component extracted by applying TICA (Eq. (4.11)) to the simulated propagating ENSO. The yellow curve represents the SOI pattern, and the red curve is the SOI pattern with 3 months phase delay.

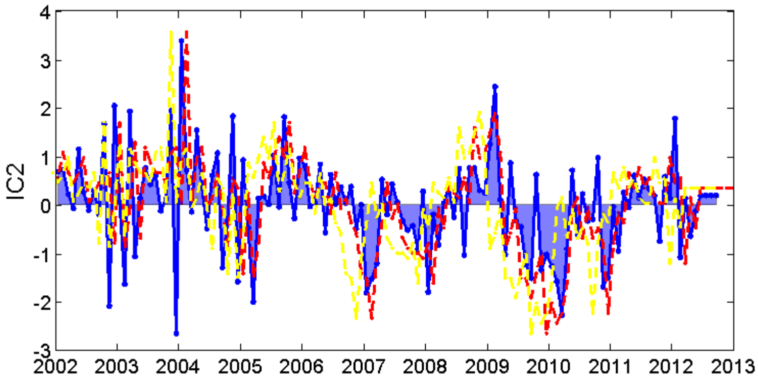


Figure 5.8: The blue curve shows the second dominant independent component extracted by applying TICA (Eq. (4.11)) to the simulated propagating ENSO. The yellow curve represents the SOI pattern, and the red curve is the SOI pattern with 3 months phase delay.

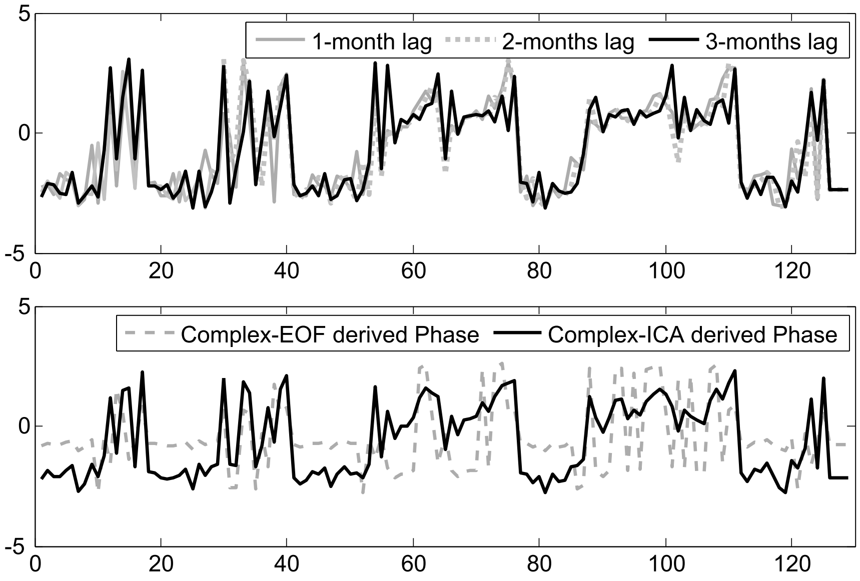


Figure 5.9: Comparison between simulated ENSO phase and its extractions using the CEOF and T-CICA techniques. Figure 5.9 (top) presents the temporal phase of SOI with respect to the 1, 2 and 3 months lags. Figure 5.9 (bottom) presents the temporal phase derived by applying the CEOF and T-CICA. For both graphs, the horizontal axis represents months from January 2002 and the vertical axis represents unit-less arc-tangent values.

5.2 Independent Patterns of Global Total Water Storage Changes

In Forootan and Kusche (2012), PCA, VARIMAX REOF, TICA, and SICA were evaluated in extracting large-scale TWS patterns. The major findings of which can be summarized as follows:

The geometrical properties of PCA can be very useful since the covariance matrix of any subset of retained PCs is always diagonal. PCA also captures the dominant part of the variance in the data set when the components are ordered with respect to the descending magnitude of singular values. From an interpretation point of view, however, when the extracted components are treated individually, the PCA method can also be misleading since it combines many of the actually separate signals into its retained components. This was evident when Forootan and Kusche (2012) applied PCA to extract dominant patterns of the simulations and also the real GRACE-TWS changes, where most of the PCs turned out to contain apparently a ‘mixture’ of signals. In theory, the PCA results may be enhanced to some extent by VARIMAX rotation. Application of VARIMAX REOF on GRACE-TWS changes, however, did not significantly improve the interpretation of the results. The application of the TICA and SICA techniques indicated that the estimated independent components contained less ‘mixing’ of signals of different time-scales.

In this section, first, TICA is applied to GRACE-TWS changes (derived from GFZ RL05) covering the period of January 2003 to June 2013. The goal is to extract the main independent patterns of TWS changes and evaluate their uncertainties due to sampling errors as discussed in Section 4.8 and background errors as discussed in Section 2.2. After removing the dominant trends and the annual cycle, the T-CICA method of Section 4.5.3 is applied with the aim to extract those water storage changes that are due to the El Niño Southern Oscillation (ENSO) phenomenon.

5.2.1 Temporally Independent Patterns from Global GRACE-TWS Changes

TICA (Eq. (4.11)) was applied to the DDK2-filtered (see http://icgem.gfz-potsdam.de/ICGEM/TimeSeries.html and Kusche et al., 2009) global TWS changes. Spatial patterns of the four dominant independent modes (in mm) are shown in Fig. 5.10, and their corresponding unit-less temporal patterns are presented in Fig. 5.11. The error-bars shown in Fig. 5.11 account for two sources of errors; the one shown by the black color is derived by applying the resampling method of Section 4.8, while considering the sampling errors and nominal errors of GRACE monthly solutions; and the error shown by the gray color derived by projecting the DDK2-filtered absolute differences between ITG3D-ERA-Interim products and the atmospheric part of GRACE-AOD1B on the spatial patterns of Fig. 5.10.

The first independent mode (spatial pattern of IC1 in Fig. 5.10 and IC1 in Fig. 5.11) captures the dominant linear trend that has been detected over polar regions such as Greenland, Alaska and Antarctica. IC1 thus represents a considerable ice-mass loss over the polar regions during the period of study (cf. Velicogna and Wahr, 2005). Some smaller mass decrease is also detected in the first independent mode such as the one over west of Australia (see also Forootan et al., 2012). In contrast, signals of moderate mass increase have been detected over the northern part of South America, as well as crustal uplift in the northern part of the Canadian shield (Rangelova and Sideris, 2008).

The annual cycle is extracted by the second and third independent modes (IC2 and IC3 and their corresponding spatial patterns) with different phase. Computing the cross-correlation of IC2 and IC3 shows that these two components are uncorrelated at time lag zero, and the maximum correlation of 0.95 is reached for a three months lag. The annual signal is clearly evident in

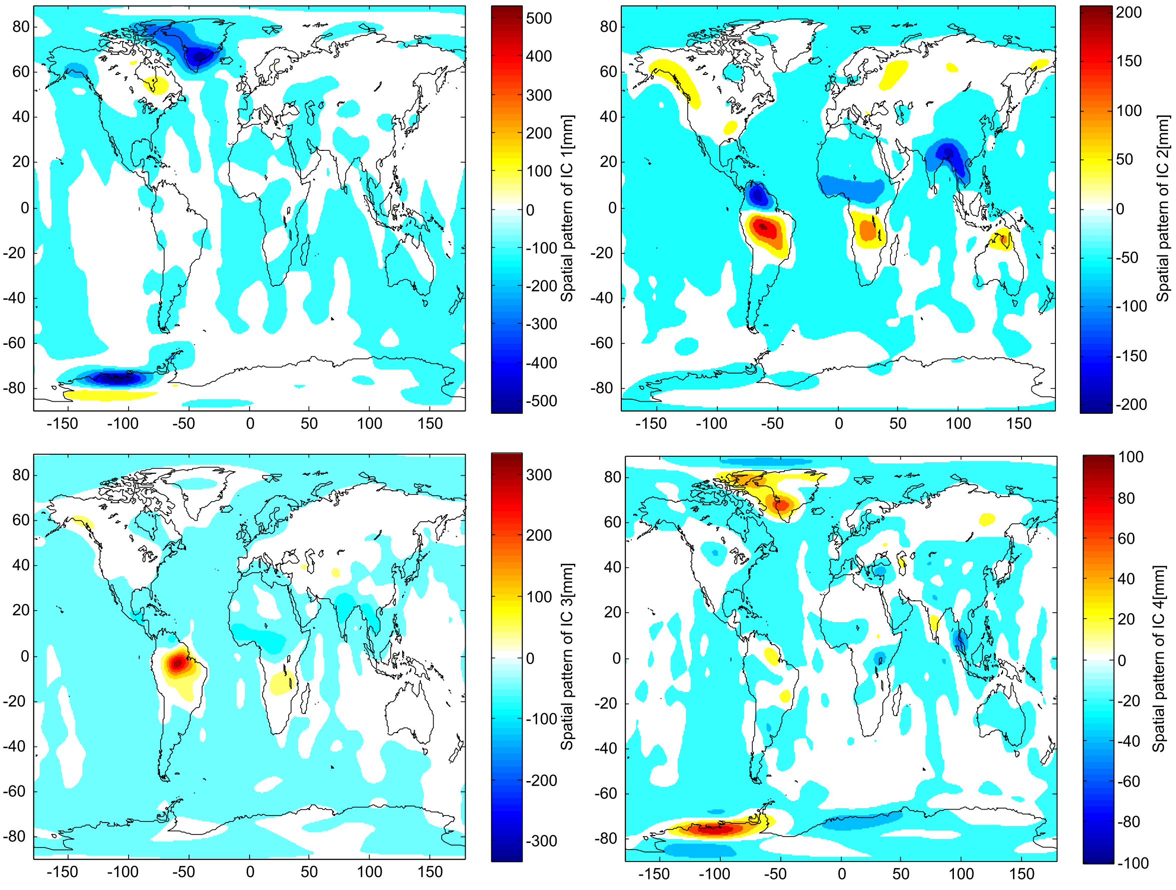


Figure 5.10: The first four leading spatial anomaly maps derived from applying TICA to GRACE-TWS changes (from GFZ RL05). The derived maps are ordered with respect to the magnitude of total variance that they represent. The corresponding temporal patterns and their associated errors are shown in Fig. 5.11.

the tropical regions, such as South America, Africa, South Asia, and northern Australia. Yet, weaker annual signals were also detected over mid and higher latitude regions such as Alaska, Siberia and northern and central Europe, which might to be related to variations in snow and ice coverage. IC4 represents a trend similar to an acceleration curve. The corresponding spatial pattern (spatial pattern of IC4 in Fig. 5.10) shows that the acceleration is mainly concentrated over Antarctic and Greenland, Lake Victoria, and Thailand.

Considering the magnitude of propagated errors, for the trend component (IC1), the atmosphere error (gray lines) is dominant (IC1 in Fig. 5.11). For the other components, those of GRACE coefficients and the sampling error (black lines) are the dominant source of uncertainty (IC2, IC and IC4 in Fig. 5.11). The dominant behavior of the atmospheric error for IC1 is most likely due to the erroneous behavior of the atmospheric de-aliasing products over the polar regions (Forootan et al., 2014a).

After removing the first four dominant independent modes (Figs. 5.10 and 5.11) from the original TWS time series, TICA was applied to the residuals. The first two dominant modes of the residuals (labeled as IC5 and IC6) are shown in Fig. 5.12, which both represent a pattern similar to ENSO. For comparison, the Southern Oscillation Index (SOI) is shown along with the temporal patterns (see Fig. 5.12). As was shown in Section 5.1.3, since ENSO has a propagating nature,

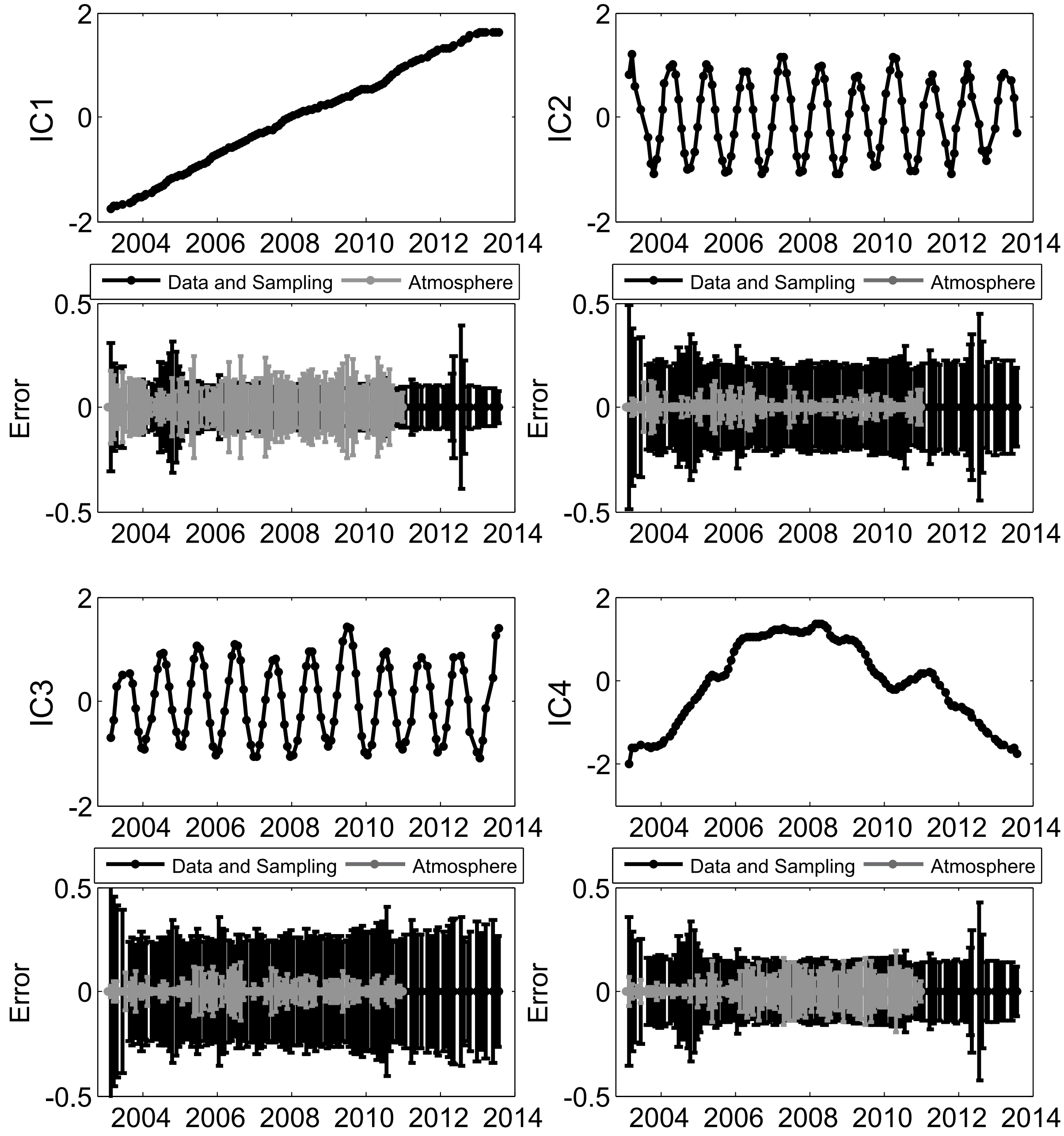


Figure 5.11: The corresponding temporal patterns of Fig. 5.10. For each independent component, two sets of error-bars are computed. The black lines represent the error in the DDK2 filtered GRACE-TWS data and the sampling error of the ICA approach (Section 4.8). The gray error-bars are derived by projecting the DDK2-filtered absolute differences of monthly atmospheric de-aliasing products from ITG3D-ERA-Interim and the atmospheric part of GRACE-AOD1B RL05 (Section 2.2). The length of atmospheric errors is limited to the period of 2003 to July 2011, since ITG3D-ERA-Interim was only available over this period.

the ordinary TICA method fails to extract this ENSO-type signal in one specific mode. Rather, the dominant part of the pattern is clustered into two modes. Errors of the extracted patterns are shown in Fig. 5.12, from which comparable amplitudes were found for both sampling and background errors.

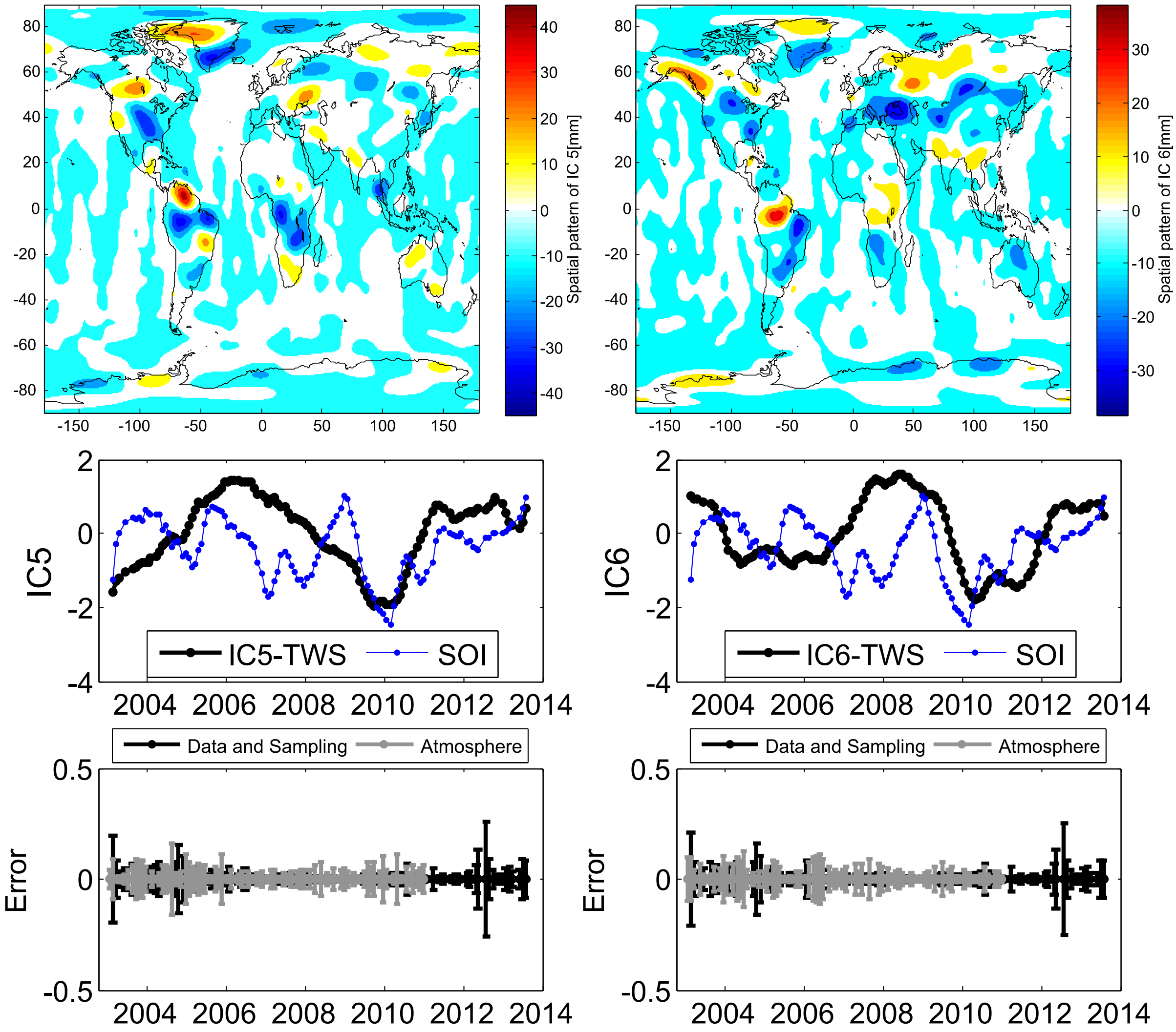


Figure 5.12: Two TICA-derived independent modes, which can be related to the ENSO phenomenon. For comparison, the ENSO index (Southern Oscillation Index, SOI) is plotted along with the temporal patterns. The error-bars are computed similar to those of Fig. 5.11.

5.2.2 Extracting the El Niño Pattern from Global GRACE-TWS Changes

To mitigate the clustering behavior seen in Fig. 5.12, the CEOF and T-CICA techniques, introduced in Sections 3.4 and 4.5.3, were applied to the residuals of TWS changes and the first four independent modes, from which only the results of T-CICA are shown in Figs. 5.13 and 5.14. The results from applying CEOF turned out similar to those from T-CICA with substantial differences in the phase patterns. The temporal amplitude of the first dominant component derived from T-CICA is quite similar to the SOI pattern, see Fig. 5.13 (top). Figure 5.13 (bottom) shows the temporal phase. In 2009, the phase pattern is quite peaky, which coincides with the extreme activity of ENSO. However, interpreting this behavior needs more research. The spatial amplitude of the first complex mode is shown in Fig. 5.14 (left), from which a strong influence of ENSO is detected over the Amazon Basin. Relatively strong anomalies are also found over a large region Africa (see Fig. 5.14 (laft)). Figure 5.14 (right) shows the spatial phase extension, which represents the response of TWS changes to the ENSO phenomenon.

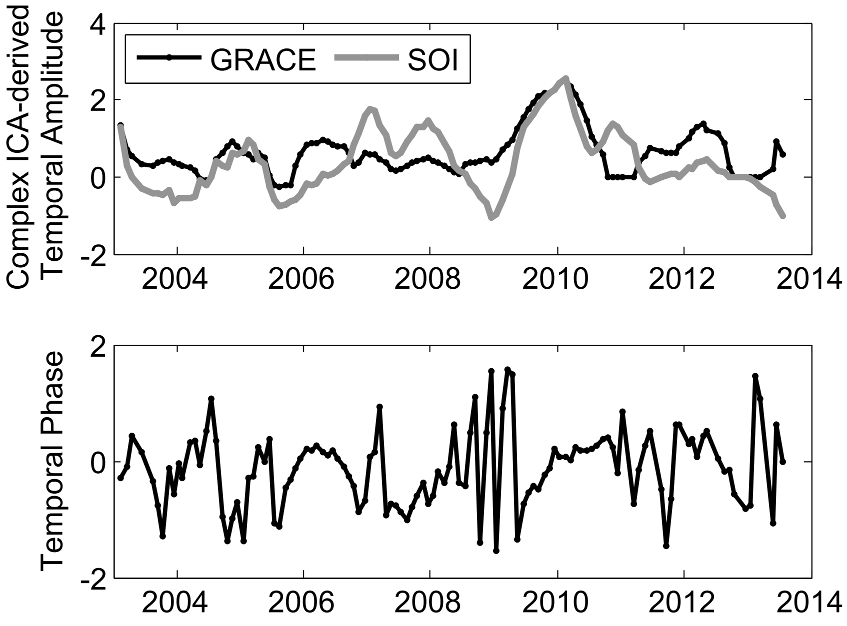


Figure 5.13: Temporal pattern of ENSO derived by applying T-CICA to GRACE-TWS time series after removing the trend and annual components (Figs. 5.10 and 5.11). Figure 5.13 (top) shows the temporal amplitude derived from the first complex component. The presented amplitude is scaled by its standard deviation making it unit-less. Figure 5.13 (bottom) represents the temporal phase (in radian) derived by applying Eq. (3.45) to the real and imaginary temporal patterns, corresponding to the amplitude in the upper graph. The related spatial pattern is shown in Fig. 5.14.

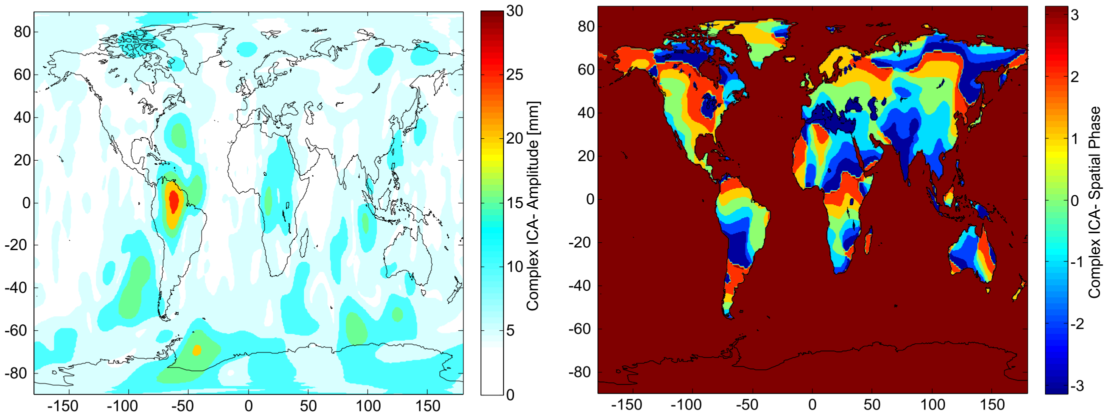


Figure 5.14: Spatial pattern of ENSO derived by applying T-CICA to GRACE-TWS time series after removing the trend and annual components (Figs. 5.10 and 5.11). Figure 5.14 (left) shows the spatial amplitude derived from the first complex component. Figure 5.14 (right) shows the spatial phase (in radian) derived by applying Eq. (3.43) to the real and imaginary spatial patterns. The corresponding temporal pattern is shown in Fig. 5.13.

5.2.3 Summary and Discussion of the Global Results

In this section, first, the TICA method was applied to the latest version of the global GRACE-TWS changes (from GFZ RL05), covering January 2003 to July 2013. The results showed the practical benefits of ICA, which does not only de-correlate the data set but goes one step further by finding components that are mutually independent. As a result, the dominant trends and the annual cycle were well separated. The sampling errors and the possible influence of the atmospheric de-aliasing products were also evaluated. The uncertainty computed for the extracted linear trend shows a considerable level over the polar regions, which is usually ignored in studies that address mass variations over these regions. This includes studies based on GRACE level 2 products and also those who used regional approaches.

When the dominant patterns, hidden in the data, exhibit a propagating nature similar to that of ENSO in global GRACE-TWS changes, almost none of the ordinary methods such as PCA, its rotation extensions e.g., VARIMAX EOF, or the ordinary ICA, are capable of extracting it in one specific mode. To deal with this issue, the CEOF and T-CICA were applied to the global TWS changes after removing the dominant modes derived from TICA. As a result, both CEOF and T-CICA extracted complex modes, from which the first mode was highly similar to the ENSO pattern (Figs. 5.13 and 5.14). This finding might suggest the suitability of using the complex extensions to deal with propagating signals.

5.3 Independent Patterns of Total Water Storage Changes over Australia

Water availability in Australia is highly variable from year to year, with various parts of the continent (e.g., the southern and eastern regions) having suffered from severe droughts conditions during the last decade (e.g., Ummenhofer et al., 2011). The long-term and inter-annual climate variability over Australia is affected by ocean-atmospheric phenomena such as ENSO (Risbey et al., 2009) and the Indian Ocean Dipole (IOD) teleconnections (Cai et al., 2011).

Monthly GRACE-TWS products have been used to estimate the annual and inter-annual water variations over Australia (e.g., in Ellett et al., 2006, Leblanc et al., 2009). Regarding the application of GRACE to monitor Australian water storage changes, Awange et al. (2009) pointed out that (i) much of Australia contain a relatively small TWS signal, which is very difficult to detect using the current GRACE system and processing strategies, and (ii) the effect of considerable spatial and spectral leakage from the surrounding oceans masks the GRACE-derived TWS changes over land. It is worth mentioning here that the impact of leakage is not the same for different parts of the continent. For instance, Tregoning et al. (2008) found a notable land hydrological signal over the Gulf of Carpentaria, while Brown and Tregoning (2010) reported that the hydrological variability over the Murray-Darling Basin does not need to account for ocean leakage (see also van Dijk et al., 2011).

With respect to (i), Awange et al. (2011) evaluated a regional solution computed with the mass concentration 'mascon' method. Their study indicated that the mascon products slightly improved the identification of TWS over the Australian continent. The focus of this section is on mitigating spatial leakages (mentioned as point (ii)). Therefore, the SICA method (Eq. (4.10)) is applied to decompose the Australian GRACE-TWS changes (from GFZ RL04) over the period from October 2002 to May 2011. For comparison, TWS outputs from the WaterGAP (Water-Global Assessment and Prognosis) global hydrology model (Döll et al., 2003), and from the Australian Water Resources Assessment (AWRA) system (van Dijk et al., 2011) are used. After decomposing GRACE-TWS anomalies, those independent modes that are concentrated over land are used to reconstruct the continental TWS changes. Our motivation for selecting SICA for performing the decomposition is that, here, detection of spatially localized anomalies is of interest. Therefore, following the simulation in Section 5.1.3, it makes sense to use a decomposition that considers the data set as a spatial sequence. The complex extension was not considered since, unless the spatial components are localized, the clustering behavior does not affect the final result of reconstruction.

This section further studies the links between the derived independent patterns of GRACE-TWS changes and climate indices e.g., ENSO and IOD. This investigation is vital to identify those areas of the Australian continent that are controlled by teleconnections and those that are not. Note that the data sets used for this section are the same as those in Forootan et al. (2012). For brevity, therefore, details of the data sets and the performed processing strategy are not discussed here. This section starts by a presentation of the numerical results and their interpretation in Section 5.3.1, and then followed by a discussion in Section 5.3.2.

5.3.1 Numerical Results over Australia

An overview of the signal RMS of the three main data sets (GRACE-TWS, AWRA-TWS, and WGHM-TWS) is shown in Fig. 5.15 in order to compare the signal strength over different regions of Australia. The RMS indicates that the main water storage signal is detected over northern Australia, where GRACE and AWRA showed stronger variability than WHGM. This can be particularly seen over the southern and western Australia (except for a smaller region of the southwest coast), where WGHM-derived TWS variability is considerably weaker. The TWS variability over southern Australia as simulated by AWRA is also weaker than that seen by GRACE.

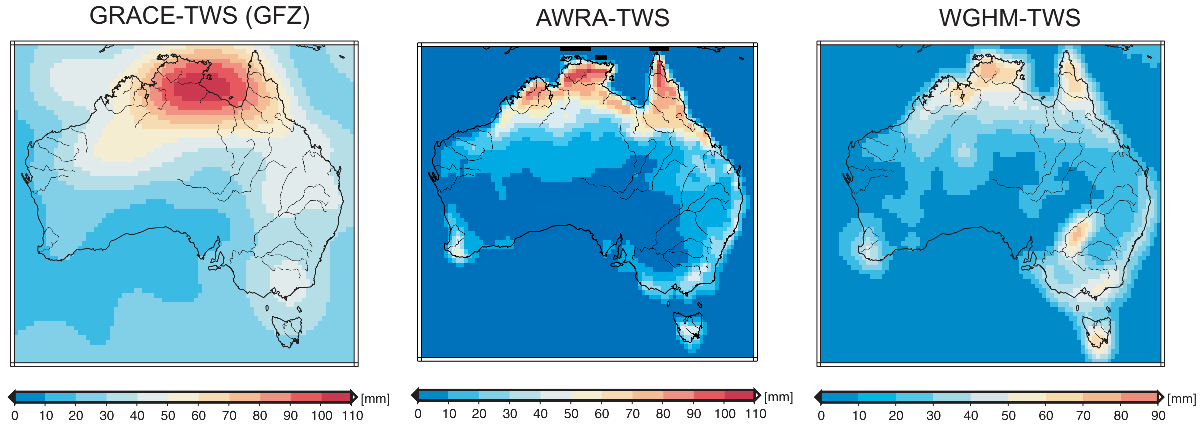


Figure 5.15: Comparing the signal variability of the three main data sources used in this study after smoothing using the Kusche et al. (2009) DDK2 filter; GRACE GFZ-data (left), AWRA-data (middle), and WGHM-data (right). Grid values represent the RMS of data sets. Details of data processing are reported by Forootan et al. (2012).

Spatial ICA for Leakage Reduction

Implementing SICA to the data sets starts by applying PCA as the first step to each data set before rotating them towards independence (Section 4.3). Considering the computed eigenvalues (not shown here), one can see that reconstructing more than 90% of the total variability of GRACE-TWS changes required selecting the first 11 PCA components. The remaining 10% of the variance in data sets was assumed to be noise. After applying SICA (Eq. (4.10)), Fig. 5.16 shows the spatial independent patterns of the GRACE-TWS data. The corresponding temporal evolutions after applying a 12-months moving average filter are shown in Fig. 5.17. Filtering is applied to enhance the comparisons with the climate indices.

To compare GRACE-TWS changes to those of hydrological models, anomalies over the continent should be separated from the surrounding oceans. This is done by selecting the independent modes of $1, 4, 5, 6, 8$ and 11 (Fig. 5.16 and Fig. 5.17), that are located over land, and reconstructing land-TWS changes using Eq. (4.8). The RMS of the reconstructed results is shown in Fig. 5.18 (A). Comparing Fig. 5.18 (A) to Fig. 5.15 (top-left), it is clear that implementing SICA has considerably isolated the signals from the surrounding oceans (e.g., no anomalies over the Gulf of Carpentaria and the eastern oceans can be seen). This shows that the application of SICA mitigated the spectral leakage over this area.

Figure 5.18 (B) shows the differences between the linear rate computed from the ICA-reconstructed GRACE-TWS and AWRA-TWS changes, covering the period of 2003 to 2011. To compute the rate, a linear trend along with annual and semi-annual sinusoids are fitted to the time series of TWS changes derived from different grids. Compared to AWRA, GRACE estimated a stronger drying trend in northwest Australia, as well as, a stronger mass gain in east and northeast Australia. These results confirmed findings by van Dijk et al. (2011), but they associated the differences to the unexplained trend in AWRA. Finally, the GRACE-TWS variability separated by SICA over the continent were temporally correlated with the time series of WGHM and AWRA from 2003 to 2010 and 2003 to 2011, respectively. The correlation results showed a high agreement between the two hydrological models and the SICA-separated results of GRACE over the continent, see Fig. 5.18 (C and D). This shows that the spatial leakage has also been decreased after application of SICA, since the reconstructed products represent higher correlations with hydrological models. A detailed comparison between GRACE-TWS changes and the hydrological models (e.g., in terms of trend and seasonal components) is presented in Forootan et al. (2012).

Independent Modes of TWS Changes over Australia

From Fig. 5.16, it was seen that the first three spatially independent patterns of GRACE-TWS changes were concentrated over the northern part of Australia, where their corresponding ICs show an annual cycle (12-months moving average filtered patterns are shown in Fig. 5.17). Particularly, IC1 isolates the annual signal over north Australia, IC2 represents the oceanic mass over the north of the Gulf of Carpentaria (Tregoning et al., 2008), and IC3 mainly shows annual mass change over the Timor Sea. The computed long-term linear rates for IC1 and IC2 between October 2002 and January 2011 showed that these regions gained mass at rates of 6 ± 2 $mm/year$ and 5 ± 2 $mm/year$, respectively. Vinogradova et al. (2011) pointed out that the variability detected over the Gulf of Carpentaria was related to the self-attraction forces that are not well removed from GRACE level 2 products. The computed linear rate for IC3 (Timor Sea) was not found statistically significant, showing an almost steady mass balance.

The fourth independent mode (IC4) from GRACE-TWS isolates a long-term mass loss along with an annual cycle over the northwestern Australia. A linear rate of TWS changes from October 2002 to May 2011 shows a loss of -19.2 ± 2 $mm/year$ in the region. IC5 depicts an increasing rate over the eastern and northeastern parts of Australia due to the 2010-2011 wet conditions. The linear rate from October 2002 to May 2011 shows a gain of 12 ± 2 $mm/year$. This rate for the period of October 2002 to January 2010 (the period before the 2011 floods) was 7.4 ± 3.6 $mm/year$. The linear rate of IC6 can be split into three sections; one from 2002 to the last months of 2005 that shows a mass gain of 8.2 ± 4 $mm/year$, then a decline in mass storage with a rate of -45.2 ± 8 *mm/year* is detected until the starting months of 2007, and finally a mass gain at a rate of 12 ± 4 *mm/year* to May 2011.

As explained in Section 2.2, during the GRACE processing procedure, high-frequency mass variations, e.g., caused by oceanic tides, are reduced as they cannot be resolved by monthly gravity field solutions. However, current ocean tide models are not accurate enough to fully reduce the tidal signal in GRACE level 2 products (Knudsen, 2003, Ray et al., 2003, Chen et al., 2009). Any imperfect modeled tidal mass variations occur in the monthly gravity field solutions at alias periods. One well-recognized example is the S2 semi-diurnal tide, which is mapped onto a 161-day period, and thus does not cancel out in the monthly solutions (Ray and Luthcke, 2006, Chen et al., 2009). IC7 of GRACE separates the S2 aliasing effect over the ocean located in northwest of Australia. This pattern was previously reported by Melachroinos et al. (2009) who fitted a predetermined cyclic signal (with a period of 161 days) to the GRACE time series. Therefore, one of the contributions of this study is to explore such a pattern as an independent component without using any predefined deterministic model. It is worth mentioning here that, by using PCA, this pattern cannot be extracted as an individual mode (Forootan et al., 2012). Fitting a sinusoidal function to IC7 showed a period of 161.4 days, agreeing with the theoretical derivation of Ray and Ponte (2003), and matching the observations of Melachroinos et al. (2009).

The water loss in west Australia is concentrated in IC8 (see Fig. 5.16). This pattern is also extended towards the ocean. This shows that the performance of ICA in separating the relatively low amplitude signals is poor. A rate of -8.2 ± 3 *mm/year* from March 2002 to December 2009 was derived from IC8. The computed linear rate of GRACE-TWS changes in the area during the years 2010 to 2011 was 6 ± 4 *mm/year*, which was due to an increased precipitation in this period. IC8 also shows an opposite TWS gain over central Australia corresponding to 7 ± 2 *mm/year* from March 2002 to December 2009.

IC9 and IC10 show mainly inter-annual mass fluctuations over the ocean in the northeastern and southwestern parts of Australia (with a frequency of 99.7 and 121 days, respectively). These reported frequencies are derived by fitting sinusoidal cycles. For deriving more reliable results together with their associated uncertainty, one might use advanced methods (e.g., in Schmidt et al., 2008b). Finally IC11 localizes the TWS anomalies over southeast Australia. The linear rates of TWS change in these regions were not found as statistically significant.

Effects of ENSO and IOD on Australia

Since the derived GRACE-ICs are spatially independent, it is possible to investigate the water variability of each component individually. This gives the unique opportunity to study the links between climate teleconnections (i.e. ENSO and IOD) and the derived GRACE-ICs. To this end, the Southern Oscillation Index (SOI) is used for ENSO and the Dipole Mode Index (DMI) for IOD. The long-period temporal correlations at 95% level of confidence between 12-month moving average smoothed ICs of GRACE-TWS and the indices (SOI and -DMI) were computed. Correlation analysis was done for the periods of October 2002 to 2011, as well as 2006 to 2011. We selected the period 2006 to 2011, besides the long-term period, for computing the correlations because of the high influence of teleconnections on the Australian TWS changes (see, e.g., van Dijk et al., 2011). Significant correlation values are reported in Fig. 5.17. It should be mentioned here that SOI and -DMI were selected for correlation analysis, as well as the 12-month filter for smoothing to make the results comparable with previous studies such as García-García et al. (2011).

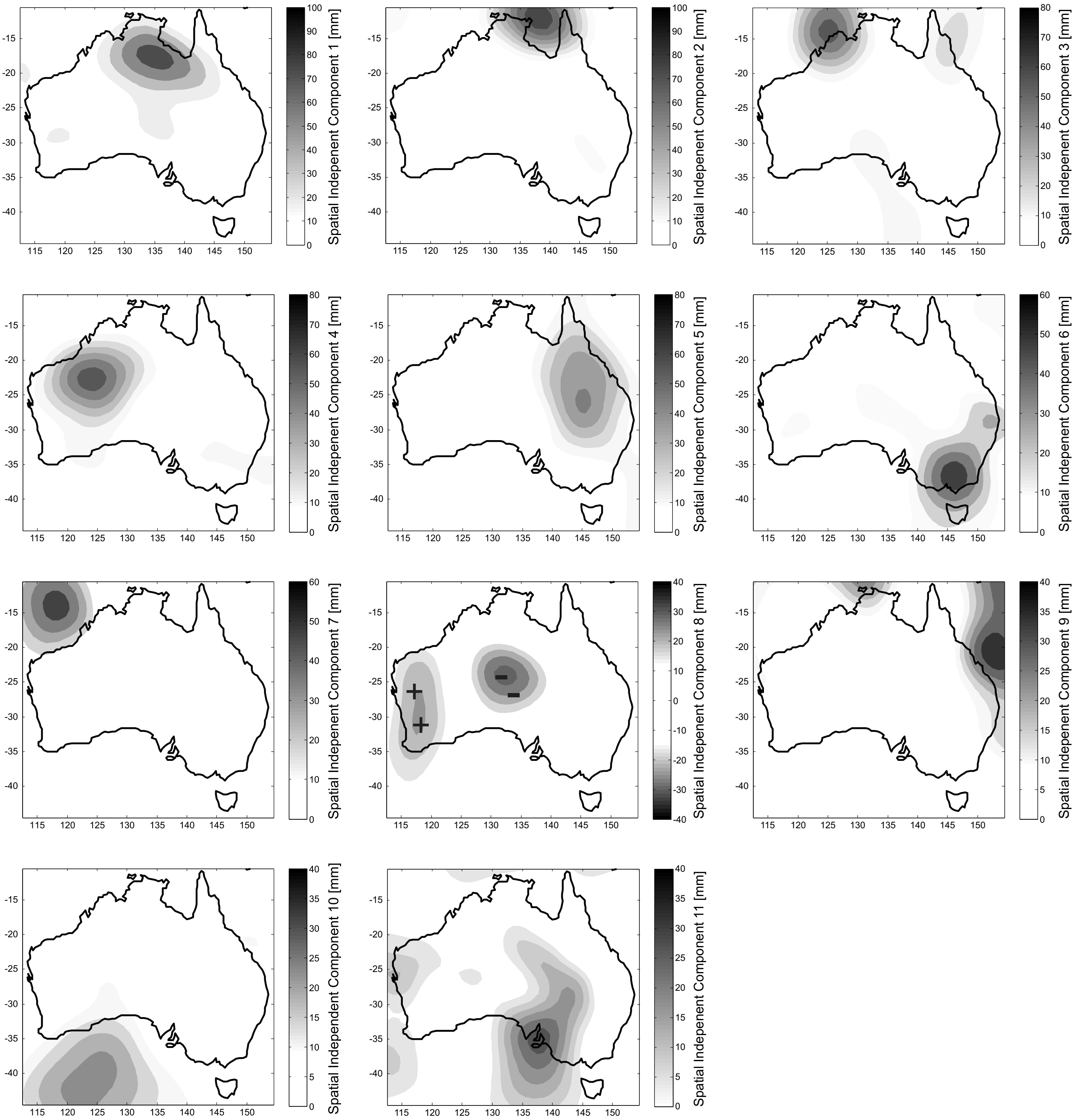
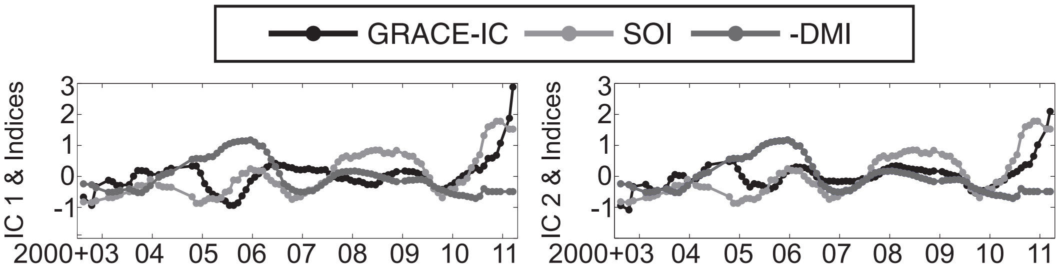


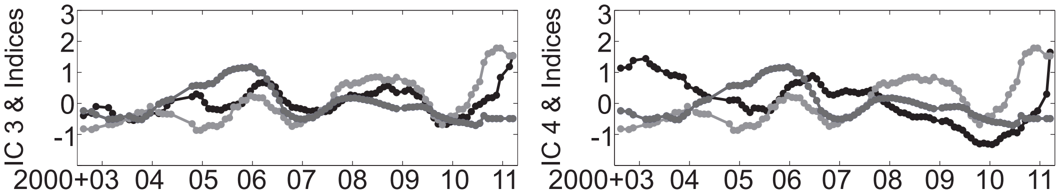
Figure 5.16: Results of the SICA method, where Eq. (4.10) was applied to the GRACE-derived TWS changes over Australia. The spatial patterns are anomalies related to the GRACE GFZ data, which are scaled using the standard deviation of their corresponding temporal evolutions (shown in Fig. 5.17). The results are ordered according to the signal strength they represent (see also Forootan et al., 2012). All presented spatial components include positive values except for that of IC8, in which the positive and negative anomalies are marked by '+' and '-' signs.

The correlation results indicate a strong stable influence of ENSO for the period 2006 to 2011 on IC1, IC2, and IC3 (with correlations of 0.57, 0.76, and 0.64, respectively). Computing the long-period correlation coefficients (October 2002 to May 2011) between the first three ICs and SOI also showed significant correlations of 0.51, 0.71, and 0.61, respectively. These correlations confirm the effect of the tropical ocean-atmosphere variability associated with ENSO rainfall in

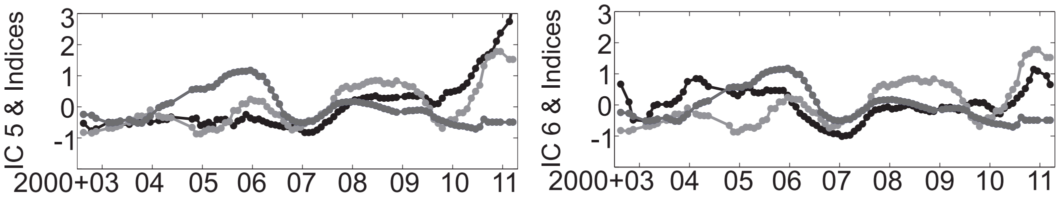


Correlation between IC 1 and ENSO, for 2006-2011: 0.57
Correlation between IC 1 and IOD, for 2006-2011: 0.35

Correlation between IC 2 and ENSO, for 2006-2011: 0.76
Correlation between IC 2 and IOD, for 2006-2011: 0.26

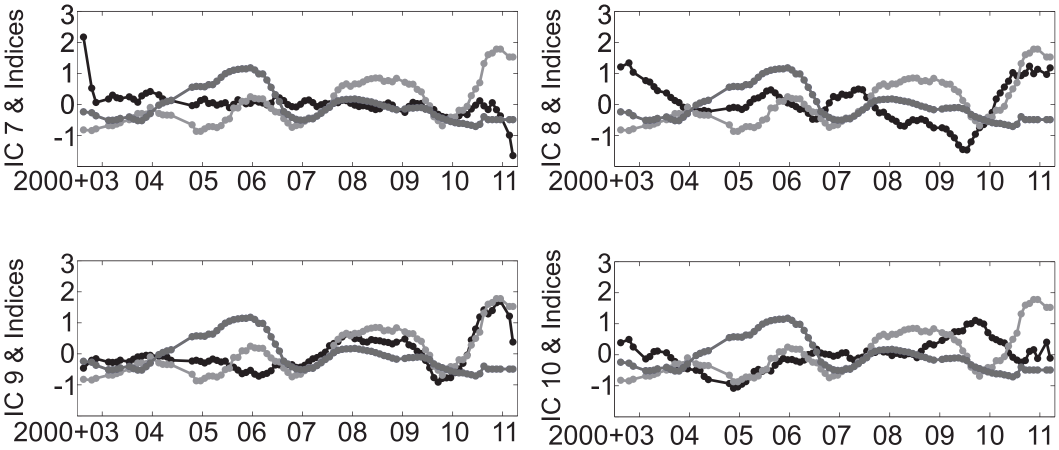


Correlation between IC 3 and ENSO, for 2006-2011: 0.64
Correlation between IC 1 and IOD, for 2006-2011: 0.32



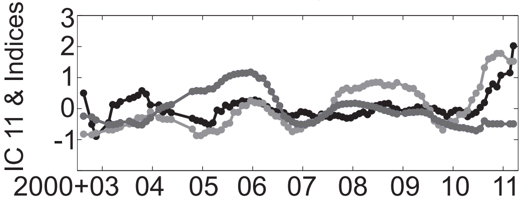
Correlation between IC 5 and ENSO, for 2002.8-2011: 0.79
Correlation between IC 5 and IOD, for 2002.8-2011: -0.36

Correlation between IC 6 and ENSO, for 2006-2011: 0.51
Correlation between IC 6 and IOD, for 2006-2011: 0.82



Correlation between IC 9 and ENSO, for 2002.8-2011: 0.81
Correlation between IC 9 and IOD, for 2002.8-2011: -0.31

Correlation between IC10and ENSO, for 2006-2011: -
Correlation between IC10and IOD, for 2006-2011: 0.53



Correlation between IC 11 and ENSO, for 2002.8-2011: 0.44
Correlation between IC 11 and IOD, for 2002.8-2011: 0.54

Figure 5.17: Temporal relationships between climate indices (ENSO and IOD) and TWS evolutions over Australia. In each graph, GRACE-derived ICs, SOI and -DMI indices are filtered using a 12-months moving average filter. The correlations are computed at 95% confidence level (see also Forootan et al., 2012). Insignificant corrlations are not reported here.

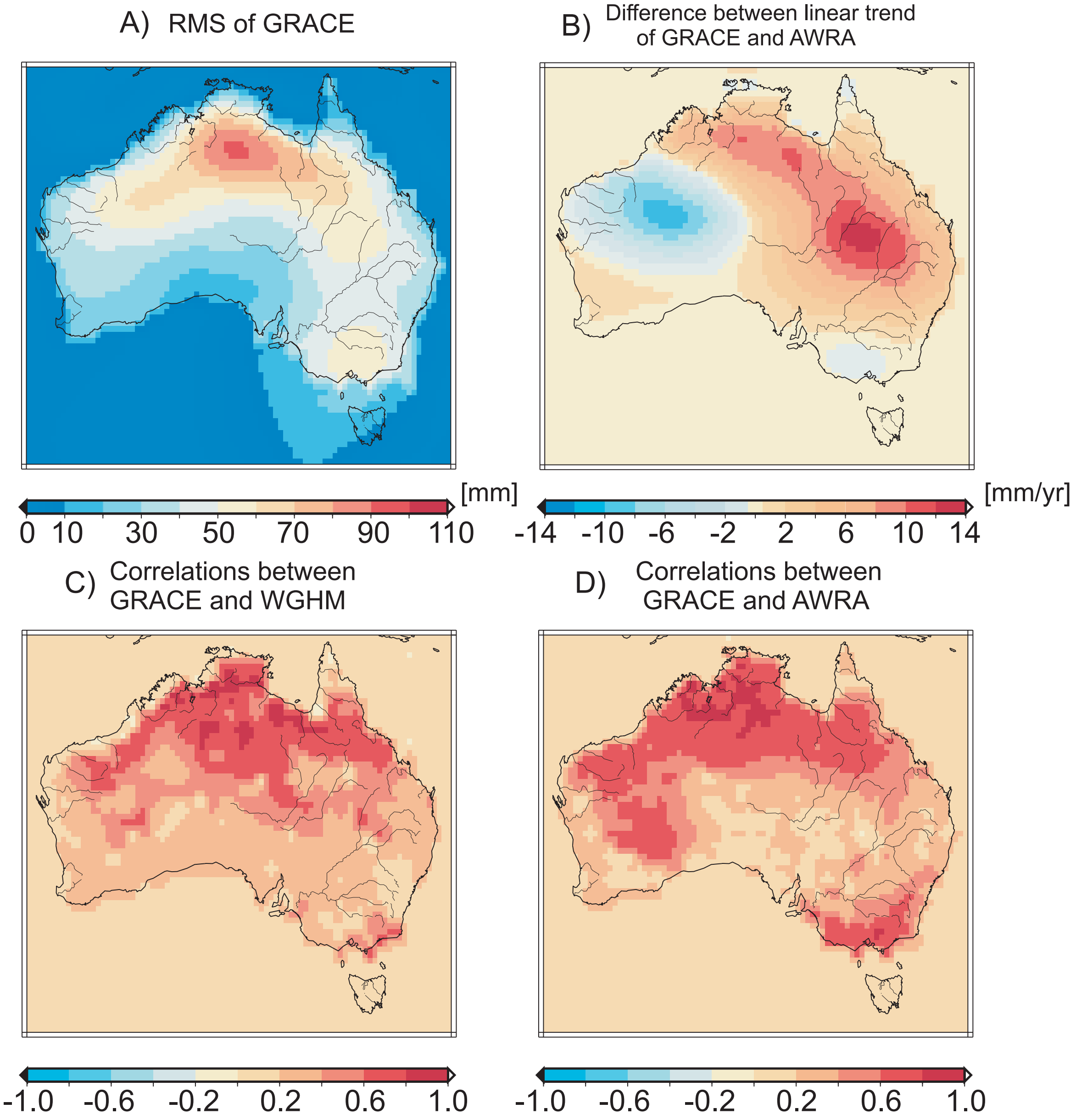


Figure 5.18: Reconstruction of GRACE-TWS variations over the Australian continent using SICA. (A) RMS of the reconstructed GRACE-TWS derived by multiplying the independent components of 1, 4, 5, 6, 8 and 10 of Fig. 5.16 to their corresponding temporal components in Fig. 5.17. (B) The difference between the linear trend computed from the ICA reconstructed time series of GRACE-TWS over the continent and the linear trend of AWRA, covering the period 2003 to 2011 (C) Temporal correlations between the ICA reconstructed time series of GRACE-TWS and WGHM-TWS changes for the period of 2003 to 2010. (D) Temporal correlations between the ICA reconstructed time series of GRACE-TWS and AWRA-TWS changes for the period of 2003 to 2011 (see also Forootan et al., 2012).

the northern regions, where IC2 and IC3 (indicating the mass variability over the ocean in the north of Australia) have stronger correlations than the northern land signal (IC1). García-García et al. (2011) reported similar results for the northern region using the Complex EOF technique.

Computed correlations between -DMI and the first three ICs were high in some years, e.g., 2006 and 2009. Their long-period correlation values for October 2002 to May 2011, however, were 0.35, 0.26, and 0.32, respectively. The correlations decreased for the period of October 2002 to May 2011. This might show that the effect of IOD on TWS variations over the northern regions is relatively weak compared to that of ENSO.

Correlations of IC4 and IC8 with SOI show the contribution of ENSO to the short-period mass gain between 2010 and 2011. Their long-period correlations for October 2002 to May 2011, however, were not found to be statistically significant. This statement applies as well to the long-period correlation of IC4 and IC8 with -DMI.

The effect of ENSO is also evident in IC5 (i.e. TWS variations in the eastern and northeastern regions) and IC9 (which concentrates over the ocean in east Australia) with a significant correlation of 0.79 and 0.81 respectively with SOI for October 2002 to May 2011. For the same period, the correlation between IC5 and -DMI was -0.36, while the correlation value between IC9 and -DMI was -0.31. These values indicate the effect of IOD on long-term variations of TWS over the eastern and northeastern regions is smaller than that of ENSO.

A significant influence of IOD from middle of 2005 to 2009 over the southeastern and southern parts of Australia is shown in IC6 and IC11. The computed correlation between IC6 and -DMI for the period of 2006 to 2011 was 0.82. This correlation, for the same period, between SOI and IC6 was 0.51. The correlation value between IC11 and -DMI for October 2002 to May 2011 was 0.54, and this value for IC11 and SOI was 0.44. Greater correlations of IC6 and IC11 with -DMI show the stronger influence of IOD on the southeastern and southern parts of the continent.

Correlation between IC10 and -DMI, for the period 2006 to 2011 was identified as 0.53, while in this period, no significant correlation with ENSO was found. Note that -DMI is selected for computing the correlations since negative IOD indicates increase in water budget and positive IOD indicates decrease in water budget. Since the spatial anomalies are positive (as they are in Fig. 5.16, the 6th and 11th patterns), -DMI should follow the pattern of the derived ICs (see Fig. 5.17).

5.3.2 Summary and Discussion of the Results over Australia

In this section, the application of ICA for extracting the large-scale TWS patterns over Australia was examined. The results indicate that; SICA has the capability to isolate the effects of spatial and spectral leakage from the surrounding oceans that mask potential terrestrial hydrological signals detectable by GRACE in a regional case such as Australia. This was unachievable using mascon or PCA and its ordinary extensions (Awange et al., 2011, Forootan et al., 2012). The result of this section can be viewed as a practical extension of the simulation case in Section 5.1.3, which showed that SICA method can successfully separate mass anomalies with distinguishable spatial patterns, even though the temporal characteristics are quite similar.

Some hidden physical processes such as the S2 tidal-aliasing along with other model deficiencies were also detected and localized from GRACE-derived time-variable TWS data without fitting any pre-defined deterministic model, which has been done in previous studies.

Studying the correlation between the ICA-localized TWS changes with the ENSO and IOD phenomena revealed the influences of the climatic teleconnections on each individual statistically independent hydrological region. As a result, a strong link between SOI and the TWS variations in the northern regions and the relation of the IOD to the eastern and southeastern Australia was established. ICA, thus, provides an alternative tool of analyzing the relationship between hydrological changes and climate variability.

5.4 Statistical Partitioning of Total Water Storage Changes

In this section, a statistical approach is provided to partition GRACE-derived total water storage (TWS) changes into surface water storage, terrestrial water storage (integration of soil moisture layers), and groundwater storage changes. The original idea of this approach has been published in Forootan et al. (2014c). Here, we extend the investigations of Forootan et al. (2014c) by (i) considering a broad region of the Middle East, and (ii) updating the GRACE data to DDK2-filtered (Kusche et al., 2009) GFZ RL05 products, covering the period of January 2003 to July 2013. As a result, maps of groundwater and terrestrial water storage changes are extracted and briefly interpreted. After introducing the research objective in below, in Section 5.4.1, the method of signal separation is introduced. This is followed by a discussion about the separated results in Section 5.4.2, and in Section 5.4.3, the application is concluded.

To account for the leakage problem (a combination of the spectral and spatial leakage problems), most of the previous studies focused on basin-wide approaches (e.g., Fenoglio-Marc et al., 2006, 2012, Longuevergne et al., 2010, Awange et al., 2013, Jensen et al., 2013). However, an approach that allows the retrieval of spatially varying water storage (WS) changes might be useful for several hydrological and climate applications. This capability is a feature that is usually lost when one applies basin-wide averaging methods. For partitioning GRACE-TWS changes, most of the previous studies use altimetry observations to account for the surface WS changes (e.g., Swenson and Wahr, 2007, Becker et al., 2010) and hydrological models for terrestrial (soil moisture) WS changes (e.g., Rodell et al., 2007, van Dijk, 2011, van Dijk et al., 2011). Subsequently, GRACE-TWS signals are compared or reduced with altimetry and/or model derived WS values. The accuracy of the estimation in such approaches might be limited since, for instance, altimetry observations contain considerable errors over inland waters (e.g., Birkett, 1995, Kouraev et al., 2011, Sharifi et al., 2013, Khaki et al., 2014, Uebbing et al., 2014) and hydrological models might show limited skills (e.g., Grippa et al., 2011, van Dijk et al., 2011).

Here, instead of removing those surface and terrestrial WS (respectively considering altimetry and hydrological models as the truth) from GRACE-TWS maps, they are used as a priori information to introduce the spatial patterns of surface and terrestrial WS changes. Then, GRACE-TWS signals are separated by adjusting the derived spatial patterns to TWS anomalies. Rietbroek (2014) applied a similar concept, known as a ‘fingerprint’ inversion approach, to estimate time-variable surface loading changes from a combination of GRACE and altimetry observations. The PCA-derived spatial components of WS changes in the oceans and land were used in Rietbroek (2014) as predefined base-functions. Surface loading changes were then computed by implementing an inversion that fits the predefined spatial patterns to GRACE and altimetry products.

In order to demonstrate this procedure, TWS data within a rectangular box that includes a large part of the Middle East region is extracted from each monthly GRACE-TWS map. As mentioned before, the main source of TWS variability, within each map, is made of the contribution of the terrestrial and surface WS changes. In this case, the surface water variations are mainly caused by water reservoirs within the selected box, e.g., the Caspian Sea, Persian and Oman Gulfs, Aral Sea, Black Sea, Red Sea as well as other small lakes. After fitting the terrestrial and surface WS changes to GRACE-derived TWS products, groundwater storage changes will be computed as a residual of TWS changes and the fitted values.

The strong seasonal mass fluctuations in the Caspian Sea will cause a time variable change in the geoid. On very short time scales (typically days), water bodies will adapt themselves to this force, similar to the tidal response of the ocean. This implies that the sea level in the Gulfs and the Black Sea are (indirectly) influenced by the variations in the Caspian Sea. This effect is known as the self-consistent sea level response (see also Farrell and Clark, 1976). This effect should be taken into account otherwise it may be misinterpreted as mass anomalies caused by, e.g., terrestrial water storage changes. Forootan et al. (2014c), however, showed that the effect of self-gravitational forces, caused by the Caspian Sea and Black Sea, are negligible over the Middle East region.

The TICA method (Section 4.5.1, Eq. (4.11)) is used to identify statistically independent patterns from (a) monthly terrestrial WS outputs of the GLDAS model (Rodell et al., 2004) over the selected rectangular box; (b) Surface WS changes derived from altimetry observations of Jason1&2 missions over the surface water bodies of the selected box. Thus, the derived independent patterns in (a) and (b) were taken as known spatial patterns (base-functions) in a Least Squares Adjustment (LSA) procedure to separate GRACE-TWS changes. This procedure gives the opportunity to make the best use of all available data sets in a LSA framework. The concept of the approach is summarized in Fig. 5.19.

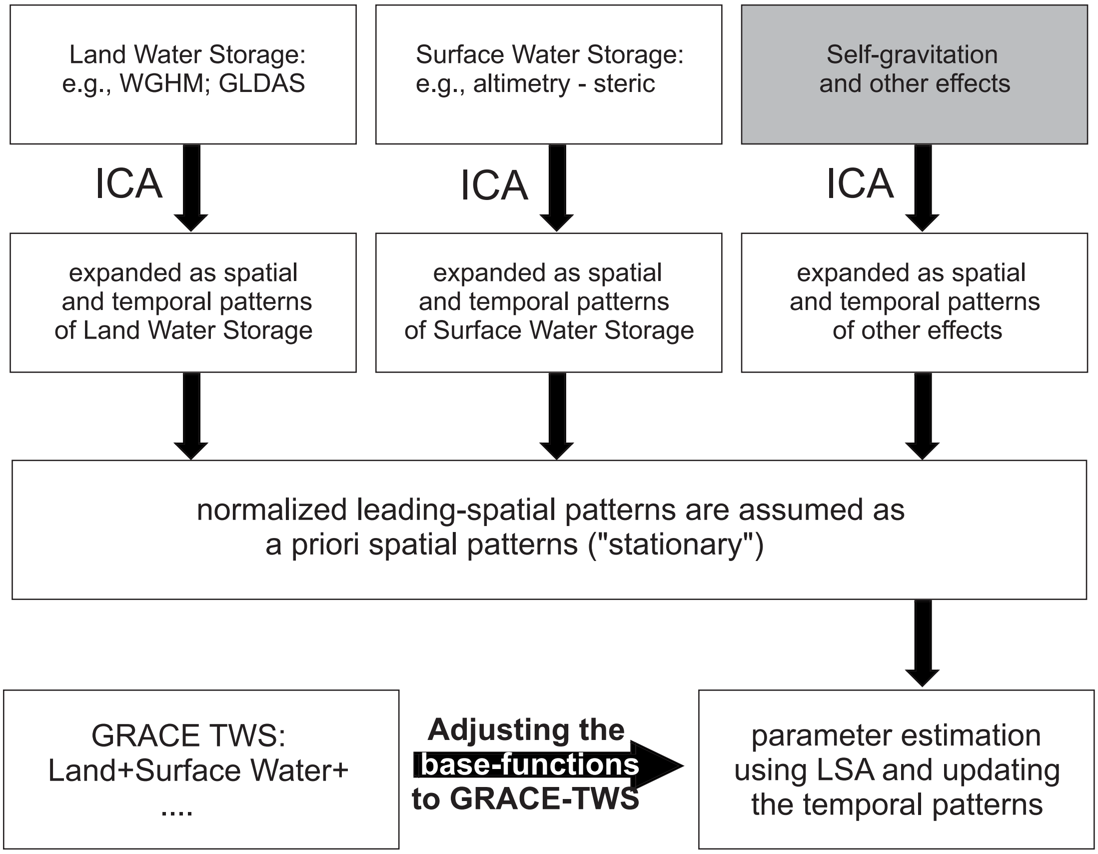


Figure 5.19: Overview of the statistical signal separation approach. The method uses the ICA-derived spatial patterns of terrestrial and surface water storage changes as a priori information. Then they are fitted to the GRACE-TWS products, in a Least Squares Adjustment (LSA) procedure, to derive GRACE-adjusted temporal evolutions.

5.4.1 Methodology of Statistical TWS Partitioning

Monthly GRACE-TWS changes, when the ocean and atmospheric mass variations are removed using de-aliasing products, reflect a combined effect of mass variations due to terrestrial water storage changes (H), and surface WS changes of seas, lakes and reservoirs (R). Assuming that GRACE-TWS fields are stored in a matrix $\mathbf{T} = \mathbf{T}(t, s)$, where t is the time, and s stands for spatial coordinate (grid points). TWS changes $\mathbf{T}$ can be factorized into spatial and temporal patterns as

$$\mathbf{T} = \mathbf{C}_H \mathbf{B}_H^T + \mathbf{C}_R \mathbf{B}_R^T, \tag{5.1}$$

where $\mathbf{C}_{H/R} = \mathbf{C}_{H/R}(t)$ and $\mathbf{B}_{H/R} = \mathbf{B}_{H/R}(s)$ are respectively the temporal and spatial patterns (base-functions). In Eq. (5.1), H and R as subindices to show the base-functions that are computed from terrestrial WS (H) and surface WS (R). Once either of $\mathbf{C}_{H/R}(t)$ or $\mathbf{A}_{H/R}(s)$ is determined in Eq. (5.1), the other component can be computed by solving a LSA. Schmeer et al. (2012) used a similar approach for separating global GRACE-TWS integral into its atmospheric, hydrologic, and oceanic contributors. To derive the required spatial base-functions in Eq. (5.1), the TICA method is applied here to decompose the centered (temporal mean removed) time series of $\mathbf{H}$ and $\mathbf{R}$ as

$$\mathbf{H}= \bar{\mathbf{P}}_H \hat{\mathbf{R}}_H \hat{\mathbf{R}}_H^T \mathbf{E}_H^T = \mathbf{C}_H {\mathbf{B}_H}^T, \tag{5.2}$$

and

$$\mathbf{R}= \bar{\mathbf{P}}_R \hat{\mathbf{R}}_R \hat{\mathbf{R}}_R^T \mathbf{E}_R^T = \mathbf{C}_R {\mathbf{B}_R}^T. \tag{5.3}$$

Similar to Section 4.3, Eq. (4.11), $\mathbf{C}_{H/R} = \bar{\mathbf{P}}_{H/R} \hat{\mathbf{R}}_{H/R}$ contains statistically mutually independent temporal components. Since $\hat{\mathbf{R}}$ is an orthogonal rotation matrix, $\mathbf{B}_{H/R} = \bar{\mathbf{E}}_{H/R} \hat{\mathbf{R}}_{H/R}$ stores the corresponding spatial maps (of $\mathbf{C}_{H/R}$) that are still orthogonal. Estimated $\mathbf{B}_{H/R}$, therefore, will be used in Eq. (5.1) as known spatial patterns and a new temporal expansions of $\hat{\mathbf{C}}_{H/R}$ will be computed using in a LSA procedure (Koch, 1999, page 57),

$$[\hat{\mathbf{C}}_H \ \hat{\mathbf{C}}_R]^T = \Big[[\mathbf{B}_H \ \mathbf{B}_R]^T [\mathbf{B}_H \ \mathbf{B}_R] \Big]^{-1} [\mathbf{B}_H \ \mathbf{B}_R]^T \ \mathbf{T}^T. \tag{5.4}$$

In Eq. (5.4), $\hat{\mathbf{C}}_{H/R}$ contains adjusted temporal components over land and surface waters, and $\mathbf{T}$ contains GRACE-TWS observations. Then, $\hat{\mathbf{C}}_H$ and $\hat{\mathbf{C}}_R$ can be respectively replaced in Eqs. (5.2) and (5.3) to reconstruct terrestrial WS changes over land and surface WS changes.

It is worth mentioning here that the separation of TWS changes in Eq. (5.4) to the land and surface water storage changes (H and R) can also be done based on the EOFs, i.e. $\mathbf{E}_H$ and $\mathbf{E}_R$ in Eqs. (5.2) and (5.3). Therefore, instead of Eq. (5.4), one can alternatively formulate the separation as

$$[\hat{\mathbf{P}}_H \ \hat{\mathbf{P}}_R]^T = \Big[[\mathbf{E}_H \ \mathbf{E}_R]^T [\mathbf{E}_H \ \mathbf{E}_R] \Big]^{-1} [\mathbf{E}_H \ \mathbf{E}_R]^T \ \mathbf{T}^T. \tag{5.5}$$

In theory, $\hat{\mathbf{C}}_{H/R}$ (in Eq. (5.4)) and their corresponding spatial patterns $\mathbf{B}_{H/R}$ should expand the same space as $\hat{\mathbf{P}}_{H/R}$ (in Eq. (5.5)) and their associated $\mathbf{E}_{H/R}$. In practice, however, the results of the reconstruction might be different since $\hat{\mathbf{C}}_{H/R}$ and $\hat{\mathbf{P}}_{H/R}$ are derived by fitting the spatial base-functions to the TWS changes that are not exactly equal to the superposition of the two storage compartments. We believe that those components of Eq. (5.4) are better suited to perform the separation since they are usually physically better interpretable.

5.4.2 Numerical Results over the Middle East Region

The RMS of GRACE-TWS changes, shown in Fig. 5.20 (top-left), clearly demonstrates a combination of the spectral and spatial leakage problem. For instance, a part of storage changes over the Caspian Sea obviously leaked into surrounding terrestrial signals. The linear rate of TWS changes is also computed and shown in Fig. 5.20 (top-right). To estimate the linear rates, the annual and semi-annual periodic changes of storage changes are taken into the account while implementing a least squares fit. Since most parts of the region exhibit a negative linear trend, the positive value of the colorbar in Fig. 5.20 (top-right) is set to white. The RMS of GLDAS-derived WS changes is also shown in Fig. 5.20 (bottom).

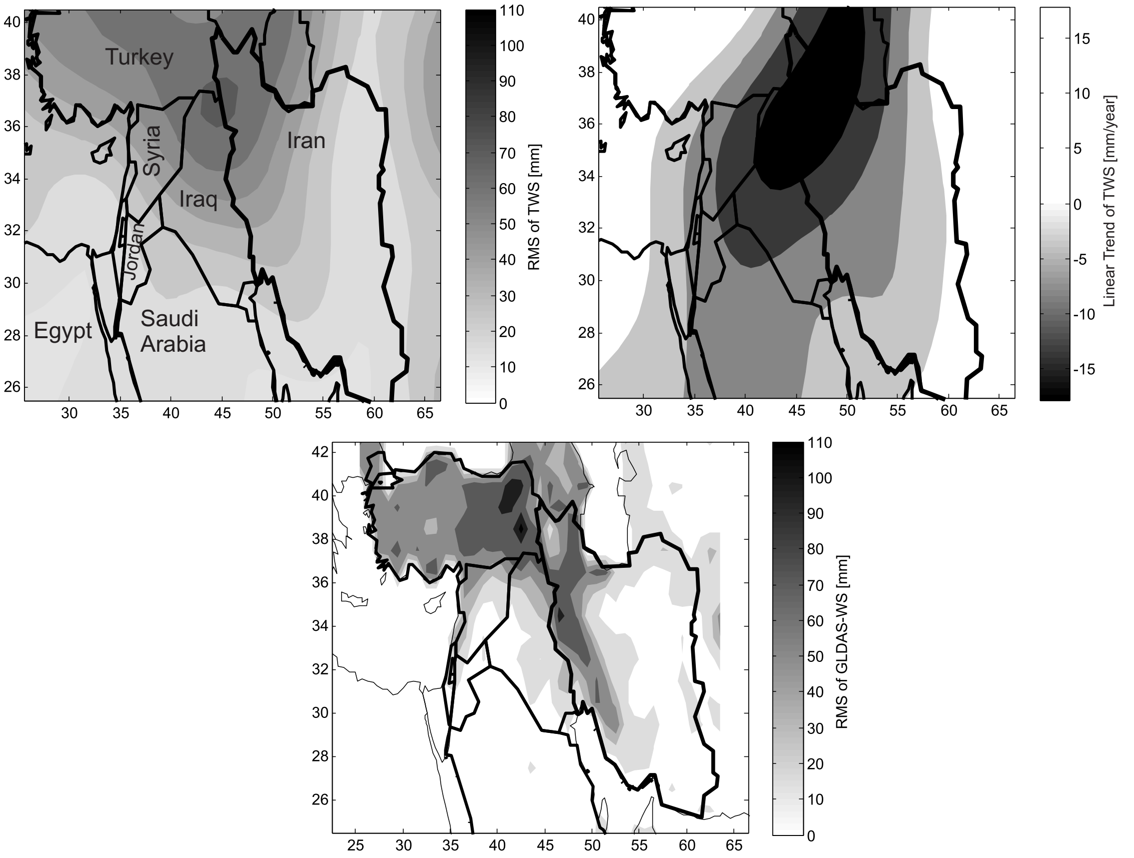


Figure 5.20: Water storage strength over the Middle East derived from GRACE and GLDAS. Figure 5.20 (top-left) shows the signal strength (RMS) of GRACE-TWS changes after smoothing using Kusche et al. (2009)'s DDK2 filter. The data covers the period of January 2003 to July 2013. Figure 5.20 (top-right) shows the estimated linear rates of TWS changes for the same period. Figure 5.20 (bottom) shows the RMS of GLDAS-derived terrestrial water storage changes.

In order to separate GRACE-TWS changes, independent modes of WS changes from altimetry and GLDAS outputs were extracted using Eqs. (5.2) and (5.3) (for brevity, the results are not shown here). The estimated spatial patterns of $\mathbf{B}_H$ and $\mathbf{B}_R$, corresponding to 90% of the total variance, were postulated as known patterns in Eq. (5.4). Using Eq. (5.4), the temporal patterns of surface and terrestrial WS changes were adjusted to GRACE observations. The adjusted temporal values $\hat{\mathbf{C}}_H$ and $\hat{\mathbf{C}}_R$ along with their corresponding spatial patterns ($\mathbf{B}_H$ and $\mathbf{B}_R$) were used to reconstruct the GRACE-fitted values of the terrestrial WS $\hat{\mathbf{H}} = \hat{\mathbf{C}}_H {\mathbf{B}_H}^T$ and surface WS $\hat{\mathbf{R}} = \hat{\mathbf{C}}_R {\mathbf{B}_R}^T$ changes.

Figure 5.21 (left) shows the linear rates of the adjusted terrestrial WS changes over the period of January 2003 to July 2013. The results show a decreasing pattern over a region around the west border of Iran and also over the east part of Turkey. The adjusted terrestrial water storage changes, then, were removed from GRACE-TWS changes over the region to derive groundwater values. Figure 5.21 (right) shows the linear rates of groundwater changes over the period of study. By comparing the left and right graphs, it is easy to see that the pattern of groundwater changes is different from that of terrestrial water storage. From our results, a vast region including the north-western and central parts of Iran, as well as an extended region of the Tigris/Euphrates River Basin (Turkey-Iraq), and the northern part of Jordan and Saudi Arabia exhibit a loss in groundwater storage. The decreasing patterns might be caused by the recent drier climatic condition over the whole area. Population increase and economic growth might also have spurred higher demands for the limited water resources (Voss et al., 2013).

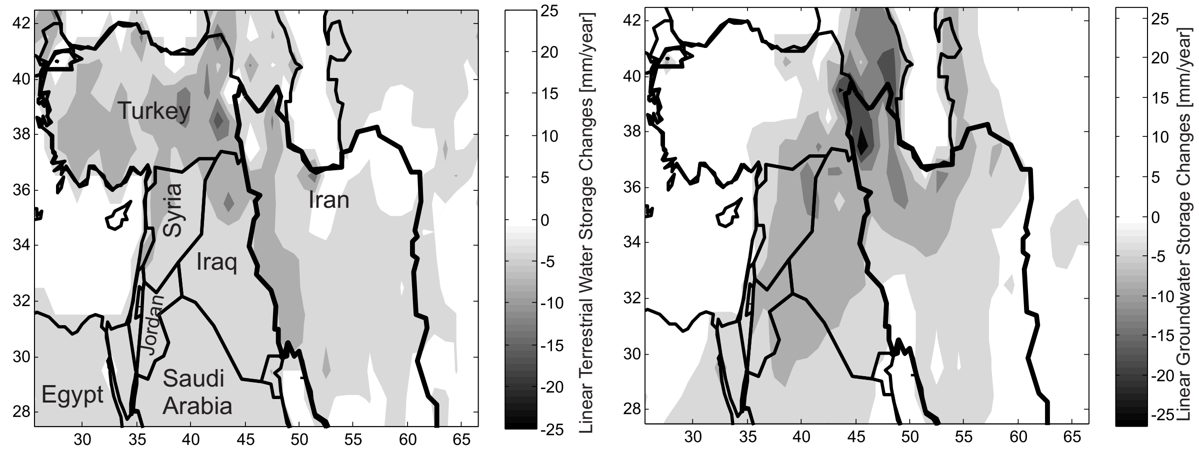


Figure 5.21: Linear rates of terrestrial water storage and groundwater storage changes over the Middle East during the period of January 2003 to July 2013. Figure 5.21 (left) corresponds to the GRACE-adjusted terrestrial water storage changes. Figure 5.21 (right) shows the results of groundwater changes over the same period. Note that to improve the visibility of the negative trend, the positive values are set to white.

5.4.3 Summary and Discussion of the Statistical Partitioning Results

In this section, we extended the study of Forootan et al. (2014c) by investigating large-scale total and terrestrial water storage changes, as well as groundwater patterns over a large part of the Middle East, where the decline of water storage is becoming a serious issue. For instance a groundwater decrease at a rate of $\sim$ 1-2 cm/year was found over the northwest of Iran (see Fig. 5.21 (right)). The decreasing pattern is also found over the trans-boundary river basins (such as Tigris/Euphrates River Basin) and aquifers (e.g., along the border of Iran and Iraq). The extracted patterns are important since the spatial variability of the data sets are kept and, thus, may be used to study natural and man-made impacts on the regional climate.

The core of the presented separation procedure lies in the ICA-decomposition of the GLDAS and altimetry outputs. Such decompositions will contain errors as a result of the short length of observations, as well as due to the errors of observations themselves. Those errors can be estimated similar to the approach presented in Section 5.2 and be introduced in the least squares procedure of Section 5.4.1. To keep the length of this thesis restricted, the results of such implementation are not presented.

5.5 Multivariate Forecasting of Total Water Storage Changes

Understanding hydrological conditions and, in particular, being able to simulate or forecast them over timescales of several months is vital for scientists and policy makers. The term 'forecast' in this study refers to estimation of a quantity of total water storage (TWS), for the period that TWS has not been observed, using its indicators. As a case study, the region of West Africa is selected, in which hydrological models usually perform poorly due to data scarcity (Schuol and Abbaspour, 2006). Besides, over West Africa, interactions between land-ocean surfaces and atmosphere are strong (see e.g., Douville et al., 2007). Therefore, it might be reasonable to use the large-scale ocean-atmospheric patterns, that are observed by remote sensing satellites, as indicators of TWS changes.

This idea has been considered, here, to design a new statistical, data-driven approach to forecast West Africa's TWS changes from gravity data obtained from GRACE, rainfall data from the Tropical Rainfall Measuring Mission (TRMM), and sea surface temperature (SST) data over the Atlantic, Pacific, and Indian Oceans. Major teleconnections within these data sets were identified using TICA (Section 4.1) and linked via low-degree autoregressive models (Section 5.5.2) to build a predictive framework.

In what follows, in Section 5.5.1, we introduce the data used for the forecast. In Section 5.5.2, the proposed forecasting method is described. The main results presented and discussed in Section 5.5.3, and in Section 5.5.4, conclusions from the study are drawn.

5.5.1 Required Data for Forecasting TWS

The GRACE-TWS data used in this section is derived from the monthly DDK2-filtered (Kusche et al., 2009) GRACE data (GFZ RL04) covering August 2002 to May 2011. The missing solutions of January 2003, 2004, May 2003, and December 2008 were not interpolated. For comparisons, monthly DDK2 filtered TWS anomalies from WGHM (Döll et al., 2003) covering the years 2003 to 2010 were used. The RMS of GRACE-TWS changes showed a strong mass anomaly over Lake Volta, which is likely due to level changes of the reservoir (RMS is not shown here). This anomaly was removed from GRACE fields before applying ICA. For details of the performed preprocessing, we refer to Forootan et al. (2014b).

SST data consisted of 1° × 1° Reynolds data from 2002 to 2012 (http://www.esrl.noaa.gov/psd/data/ gridded/data.ncep.oisst.v244.html). Similar to Omondi et al. (2012), the SST data over three major ocean basins including the Atlantic Ocean box (-66° to 13° E and -20° to 31° N), the Pacific Ocean box (159° to 275° E and -30° to 19° N) and the Indian Ocean box (34° to 114° E and -50° to 1° N) are extracted and subjected separately to further analysis. For rainfall, Version 7 of the TRMM-3B42 products (Huffman and Bolvin, 2012) covering 2002 to 2012 (http://mirador.gsfc.nasa.gov/) was used.

5.5.2 Methodology of Statistical Forecasting

ICA

On the ground of numerical efficiency, it is recommended to use a dimension reduction method before constructing the mathematical relationship between predictors and predictands. This is done here by applying TICA (Section 4.1) to the available data sets. In the following, the 'TICA' is abbreviated to 'ICA'. We assume that GRACE-TWS fields (in mm), after removing their temporal mean, are stored in a matrix $\mathbf{X}_{\text{TWS}} = \mathbf{X}_{\text{TWS}}(t, s)$, where t is the time, and s stands for spatial coordinate (grid points). Applying ICA (Forootan and Kusche, 2012), $\mathbf{X}_{\text{TWS}}$ are decomposed into spatial and temporal components as

$$\mathbf{X}_{\text{TWS}} = \mathbf{Y}_j \mathbf{V}_j^T, \tag{5.6}$$

where $\mathbf{Y}_j$ stores the j dominant unit-less temporally independent components of TWS changes, and $\mathbf{V}_j$ contains the corresponding spatial maps in its columns. Similarly, the centered maps of rainfall over West Africa $\mathbf{X}_{\text{Rainfall}}$ and SST over the major oceans $\mathbf{X}_{\text{SST}}$ are separately decomposed using the ICA decomposition as

$$\mathbf{X}_{\text{SST or Rainfall}} = \mathbf{U}_{j'} \mathbf{B}_{j'}^T, \tag{5.7}$$

where $\mathbf{U}_{j'}$ stores the j' dominant unit-less temporally independent components of SST or rainfall changes, and $\mathbf{B}'_j$ contains their corresponding spatial maps in its columns. Different indices j and j' in Eqs. (5.6) and (5.7) are used to indicate that the number of retained modes from different data sets are not necessarily the same. The index j' can also be different for the selected ocean basins or the rainfall data. Selecting a proper subset (j or j') was addressed in Section 3.5.1.2, while error estimation of the independent modes was addressed in Section 4.8. Applying Eq. (5.6) to different length of GRACE-TWS data showed that the spatial patterns of $\mathbf{V}_j$ is relatively stable. This means that, for instance, the spatial patterns ($\mathbf{V}_j$) derived from 10 years of the data do not differ significantly from those derived from 8 or 12 years of data. Therefore, for building the forecasting model, one can link the ICs of the predictor data sets (all columns of $\mathbf{U}_{j'}$ derived from SSTs and TRMM-rainfall) to individual ICs of the predictand (each column of $\mathbf{Y}_j$ in Eq. (5.6)). Finally, $\mathbf{V}_j$ of TWS will be used to reconstruct the forecasting maps as $\hat{\mathbf{X}}_{\text{TWS}} = \hat{\mathbf{Y}}_j \mathbf{V}_j^T$, where $\hat{\mathbf{Y}}_j$ is the output of the forecasting model.

Autoregressive Model with Exogenous Variables (ARX)

To forecast ICs of TWS changes, an Autoregressive Model with Exogenous Variables (ARX) process is used, which describes the relationship between the current and previous values of an output and the values of inputs. In this case, the ARX model is formulated as a multiple-inputs (ICs of SST and rainfall all together) and single-output (each IC of TWS) model (Ljung, 1987) by

$$y(t) + \sum_{i=1}^{n_a} a_i y(t-i) = \sum_{q=1}^{m} \sum_{l=1}^{n_b} b_{q,l} u_q(t - k_q - (l-1)) + \xi(t), \tag{5.8}$$

where y represents a particular independent temporal evolution of TWS, i.e. $y(t)$, $t = 1, \ldots, n$, represent a column of $\mathbf{Y}_j$ in Eq. (5.6). In Eq. (5.8), n_a is the order of the ARX model with respect to the predictand, $u_q(t)$, $q = 1, \cdots, m$, and $t = 1, \ldots, n$, are ICs of SSTs and rainfall from $\mathbf{U}_{j'}$ in Eq. (5.7), while m is the number of predictors. The order of the ARX model with respect to the predictors is n_b, and k_q denotes the number of time-steps before the q'th input (predictor) affects the output y, known as the dead time of the system. Finally, ξ is a white-noise disturbance value. The coefficients of the ARX models a_i, $i = 1, \cdots, n_a$, and $b_{q,l}$, $q = 1, \cdots, m$, and $l = 1, \cdots, n_b$, have to be derived in the 'simulation' step, using both predictand and predictors (Ljung, 1987). Once the coefficients are computed, in the forecasting step, only the predictors (ICs of SST and rainfall) are used to estimate the values of TWS changes after the simulation period.

In the literature simulation or 'training step' (see e.g., Ljung, 1987) is performed under the assumption that the output and inputs up to the time $t = t_n - 1$ are known. Furthermore, the outputs and exogenous values on the right hand side of Eq. (5.8) are not stochastic. To avoid negative indices, one might consider the observations $\mathbf{y}(t) = [y(t), y(t-1), \cdots, y(c)]^T$, where $c = \max(n_a, n_b) + \max(k_q) + 1$. Equation (5.8) is expanded as

$$\mathbf{y} = \begin{bmatrix} -y(t-1) & \cdots & -y(t-n_a) & u_q(t-k_q) & \cdots & u_q(t-k_q-n_b+1) \\ -y(t-2) & \cdots & -y(t-n_a-1) & u_q(t-k_q-1) & \cdots & u_q(t-k_q-n_b) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -y(c-1) & \cdots & -y(c-n_a) & u_q(c-k_q) & \cdots & u_q(c-k_q-n_b+1) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_{q,1} \\ \vdots \\ b_{q,n_b} \end{bmatrix} + \Xi(t), \tag{5.9}$$

$q = 1, \cdots, m$ and $\Xi(t) = [\xi(t), \xi(t-1), \cdots, \xi(c)]^T$. Equation (5.9) can be re-written compactly as

$$\mathbf{y}(t) = \mathbf{\Phi}(t)\mathbf{\Theta} + \Xi(t), \tag{5.10}$$

leading to the least squares estimation of the ARX system coefficients as

$$\hat{\mathbf{\Theta}} = \left(\mathbf{\Phi}(t)^T \mathbf{\Phi}(t)\right)^{-1} \mathbf{\Phi}(t)^T \, \mathbf{y}(t). \tag{5.11}$$

The quality of the fit (η) can be assessed by computing the signal-to-noise ratio as

$$\eta = 1 - \frac{\mathbf{y}(t)^T \mathbf{\Phi}(t) \hat{\mathbf{\Theta}}}{\mathbf{y}(t)^T \mathbf{y}(t)}. \tag{5.12}$$

The residual of fitting the ARX model ($\hat{\Xi}(t) = [\hat{\xi}(t), \hat{\xi}(t-1), \cdots, \hat{\xi}(c)]^T$) are estimated as

$$\hat{\Xi}(t) = \mathbf{y}(t) - \mathbf{\Phi}(t)\, \hat{\mathbf{\Theta}}. \tag{5.13}$$

In the forecasting step, based on $\hat{\mathbf{\Theta}} = \begin{bmatrix} \hat{a}_1 \, \ldots \, \hat{a}_{n_a} \, \hat{b}_{q,1} \, \hat{b}_{q,2} \, \cdots \, \hat{b}_{q,n_b} \end{bmatrix}^T$, when the inputs $u_q(t)$ are known, one can predict the output $\hat{y}(t_n)$ at time t_n using

$$\hat{y}(t_n) = -\sum_{i=1}^{n_a} \hat{a}_i y(t_n - i) + \sum_{q=1}^{m} \sum_{l=1}^{n_b} \hat{b}_{q,l} u_q(t_n - k_q - (l-1)). \tag{5.14}$$

To estimate the uncertainty of the ARX model in Eq. (5.8), we used Monte Carlo sampling, in which several realizations of the ICs (described in Section 4.8) were numerically generated. By inserting them into Eq. (5.8) and fitting ARX models, an error assessment of the fitted model up to the time t_n was performed. For error estimation of the forecast (the ARX value at time $t_n + 1$ and later), however, one should compute an accumulated error since there is no observed value for the output y at time $t_n + 1$ and later.

5.5.3 Numerical Results over West Africa

Statistical Decomposition Results

The two leading ICs of GRACE-TWS changes and the corresponding ICs of SST and rainfall changes are summarized in Figs. 5.22 and 5.23. For brevity, from those ICA results applied on

SST and rainfall data, only some selected temporal patterns are shown in Fig. 5.23. The first independent mode of GRACE-TWS represents the annual water variability over West Africa, where a damping of the signal magnitude can be seen in the year 2005 (temporal IC1 of GRACE-TWS). From the spatial pattern of IC1 GRACE-TWS, it is visible that the dominant concentration of annual variability appears over the tropic and coastal regions. IC2 of GRACE-TWS contains inter-annual variations of TWS along with periodic components with $\sim$ 3 and 5 years period. Nicholson (2000) also found a similar period in rainfall variations over West Africa. For comparisons, WGHM-TWS was projected on the spatial independent patterns of GRACE-TWS. The derived temporal amplitudes of WGHM-TWS are shown along with the ICs of GRACE-TWS in Fig. 5.22. The results showed a good agreement between the annual variability detected by GRACE solutions and WGHM-TWS. For the inter-annual time-scale, however, a large difference appears between what GRACE-TWS represents over West Africa and what WGHM simulates (see IC2 of TWS in Fig. 5.22). Further research will need to address the cause of these differences.

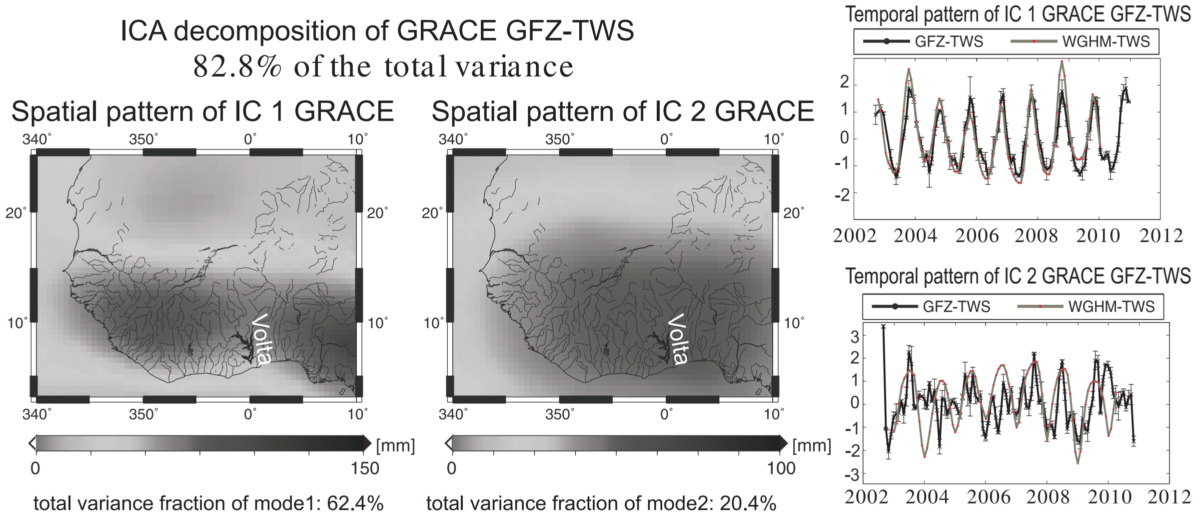
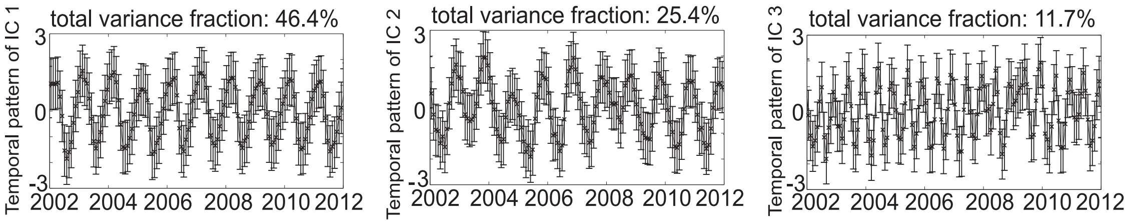


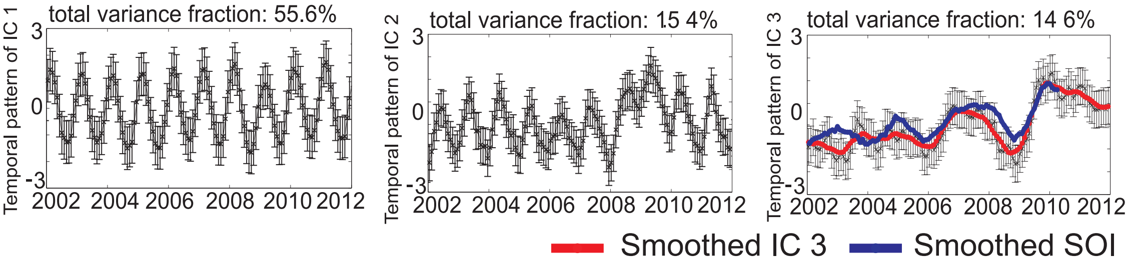
Figure 5.22: Temporal ICA decomposition of GRACE-TWS changes over West Africa (counted as predictands). For comparisons, WGHM-TWS changes are projected on the spatial patterns of IC1 and IC2. The results are presented along with temporal ICs of GRACE-TWS. The variance fraction of each independent mode is also presented.

ICA applied to SST changes over the three oceanic basins shows that the first two independent modes of surface temperature are related to the annual variability of SST (IC1 and IC2 in Fig. 5.23 (a,b, and c)). Over the Atlantic, for instance, IC1 and IC2 are related to the annual dipole structure, which are also very much correlated with IC1 of GRACE-TWS. The correlation between IC1-Atlantic SST and IC1 GRACE-TWS is 0.81 with a lag of 0 month, and between IC2-Atlantic SST and IC1 GRACE-TWS, it is 0.89 with a lag of 2 months. A damping of the annual amplitude in the year 2005 is seen for IC2-Atlantic SST, whereas a similar damp is also seen in IC1 of GRACE-TWS. This result confirms that the recent annual variability of TWS over West Africa is related to the Atlantic ocean-atmospheric interactions, reflected in the SST data (Mohino et al., 2011). IC3 of SST changes over the Atlantic and Indian Oceans represent semi-annual variability (not shown here), while IC3-Pacific SST represents the ENSO pattern: the IC3-Pacific SST was compared with the monthly ENSO pattern (shown by SOI) provided by the Australian Bureau of Meteorology. The results show a high correlation of 0.84 confirming that the pattern is physically meaningful (IC3 in Fig. 5.23 (b)). A significant correlation of 0.68

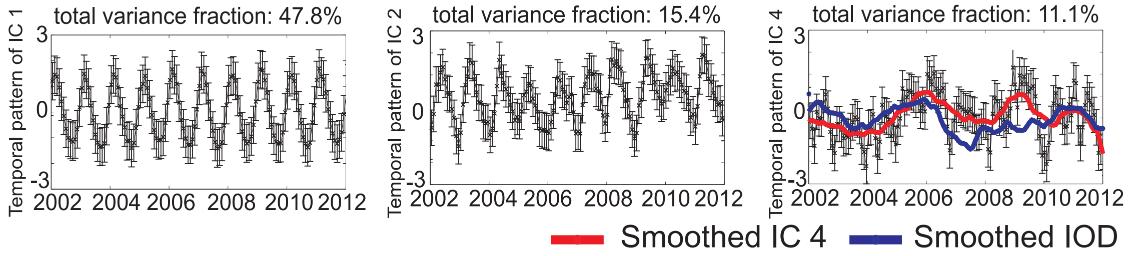
(a) ICA of SST-Atlantic Ocean



(b) ICA of SST-Pacific Ocean



(c) ICA of SST-Indian Ocean



(d) ICA of rainfall-West Africa

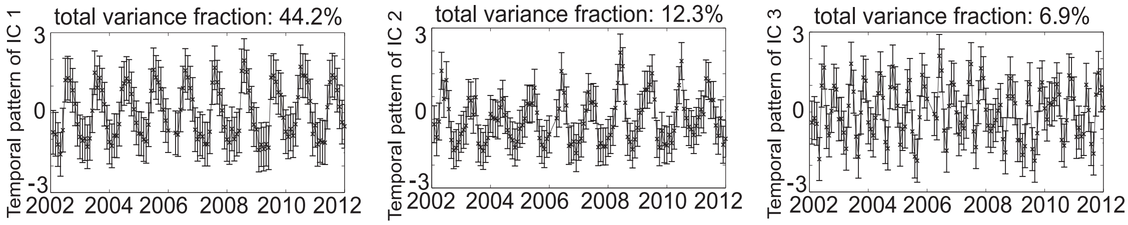


Figure 5.23: An overview of selected temporally independent patterns derived by applying the Temporal ICA method to the predictors (i.e. SST changes over the (a) Atlantic, (b) Pacific and (c) Indian Ocean basins, as well as (d) rainfall changes of West Africa). The variance fraction, that each of the ICs represents, is also presented above each graph. Uncertainties are derived from using the approach in Section 4.8 and shown with the error-bars (see the details in Forootan et al., 2014b).

was also found between IC2 of TWS and SOI, revealing a relationship between West African TWS changes and ENSO.

IC4-Indian SST (IC4 in Fig. 5.23 (c)) follows the Indian Ocean Dipole (IOD) pattern (Cai et al., 2011). Comparing the extracted IC4-Indian SST to the IOD index derived from the Japan Agency for Marine-Earth Science and Technology (http://www.jamstec.go.jp/frcgc/research/d1/iod/HTML/Dipole%20Mode%20Index.html) shows a correlation of 0.73. It should be mentioned here that modes derived by ICA show higher correlations with the ENSO and IOD indices, derived from independent data, rather than those derived by PCA (results based on PCA are not shown here).

Finally, four independent modes were extracted from TRMM-rainfall data, from which IC1-TRMM and IC2-TRMM were related to the annual rainfall variability with 3 months phase differences (Fig. 5.23 (d)). In the year 2005, a damping of the signal magnitude can be seen in IC2-TRMM, which is less pronounced than that of IC2-Atlantic SST. IC3-TRMM and IC4-TRMM represent the semi-annual rainfall variations. A lag of 2 months was found between the ICs of rainfall and those of TWS. The complete results of the Temporal ICA decomposition of SST and rainfall data are presented in Forootan et al. (2014b).

Statistical Forecasting Results

To compute the best-fitting ARX model, we inserted the first 72 months of each individual IC of GRACE-TWS (IC1 and IC2 of GRACE-TWS in Fig. 5.22) and the first 72 months of all temporal ICs of SST and rainfall (12 ICs of Fig. 5.23 and 3 other ICs that are not shown here) in Eq. (5.8). Before performing the training step, the values of January 2003, 2004, May 2003 and December 2008 were excluded from the inputs (ICs of SSTs and rainfall) to synchronize them with those ICs of GRACE-TWS. The optimum n_a and n_b were found experimentally by varying them between one to three and for k_q between zero to three, then running the ARX simulation step (Eq. (5.8)). Higher orders for n_a and n_b were not considered in order to have the model as simple as possible. For k_q, previous studies found a delay of up to three months between SST-rainfall and TWS changes. The coefficients for each ARX model $\hat{\mathbf{\Theta}}$ were computed using Eq. (5.11).

The numerical results, from the simulation of both IC1 and IC2 of GRACE-TWS, show that an ARX model with n_a=1 and n_b=3 provides the best fit; then, the residuals also pass the normality test. The fit criterion, here, is the RMS of differences between the simulated TWS values from the ARX process and the ICs of GRACE-TWS (Eq. (5.12)). Two sets of k_q corresponding to the simulations of the IC1 and IC2 of GRACE-TWS are reported in Table 5.2. The simulation results (Fig. 5.24) show that the ARX models provide a fit of 93% and 83% for simulating IC1 and IC2 of GRACE-TWS, respectively. The simulation fit of ARX corresponding to IC2 is, however, lower than that of IC1 since its temporal pattern appears much more complicated than the annual pattern in IC1. Therefore, it might have not been fully captured by the predictors (see Fig. 5.24).

Having ARX models derived for IC1 and IC2 of GRACE-TWS, the predictor values after the 72'th month alone were used to forecast the ICs of GRACE-TWS. The values of the forecast were derived from Eq. (5.14), and their uncertainties evaluated using the Monte Carlo approach (see also Forootan et al., 2014b). Results of the forecast along with their uncertainties are shown in Fig. 5.24 (a and b). The fit of the forecast for IC1 GRACE-TWS, when compared to the observed GRACE-TWS values, after one year was 79%, while after two years it was reduced to 62%. As Fig. 5.24 (a) also shows, after two years, the magnitude of uncertainty is quite large. This indicates that the proposed approach is reliable for forecasts of up to about two years. Figure 5.24 (b) shows that the fit of the forecast for IC2 of GRACE-TWS after one year is reduced to 67%. After two years, a fit of 56% was found. Comparing projected values of WGHM-TWS (gray lines in Fig. 5.22) with the ICs of GRACE-TWS (black lines), during the first year of forecast, we found a fit of 78% for the annual and a reduced fit of 53% for the inter-annual pattern. This result indicates that the TWS output of the proposed statistical method is closer to the observed GRACE-TWS changes over West Africa, when compared to hydrological modeling.

	IC1 SST Atlantic Ocean	IC2 SST Atlantic Ocean	IC3 SST Atlantic Ocean	IC4 SST Atlantic Ocean	IC1 SST Pacific Ocean	IC2 SST Pacific Ocean	IC3 SST Pacific Ocean	IC1 SST Indian Ocean	IC2 SST Indian Ocean	IC3 SST Indian Ocean	IC4 SST Indian Ocean	IC1 TRMM West Africa	IC2 TRMM West Africa	IC3 TRMM West Africa	IC4 TRMM West Africa
k_q related to IC1 of GRACE-TWS	1	0	2	1	0	3	1	1	0	1	0	0	3	3	1
k_q related to IC2 of GRACE-TWS	1	0	1	1	0	0	1	1	0	1	0	1	3	1	3

Table 5.2: Time delays k_q derived from simulation of IC1-GRACE and IC2-GRACE. The values are in month and denote the number of time-steps before each predictor (ICs of SSTs and TRMM-rainfall) affect the output (each individual ICs of GRACE).

In order to assess the robustness of the performed forecast with respect to the training period, a backward simulation and forecast were performed in Forootan et al. (2014b). The results showed a fit similar to the forward forecast. For reason of brevity, the backward results are not presented here.

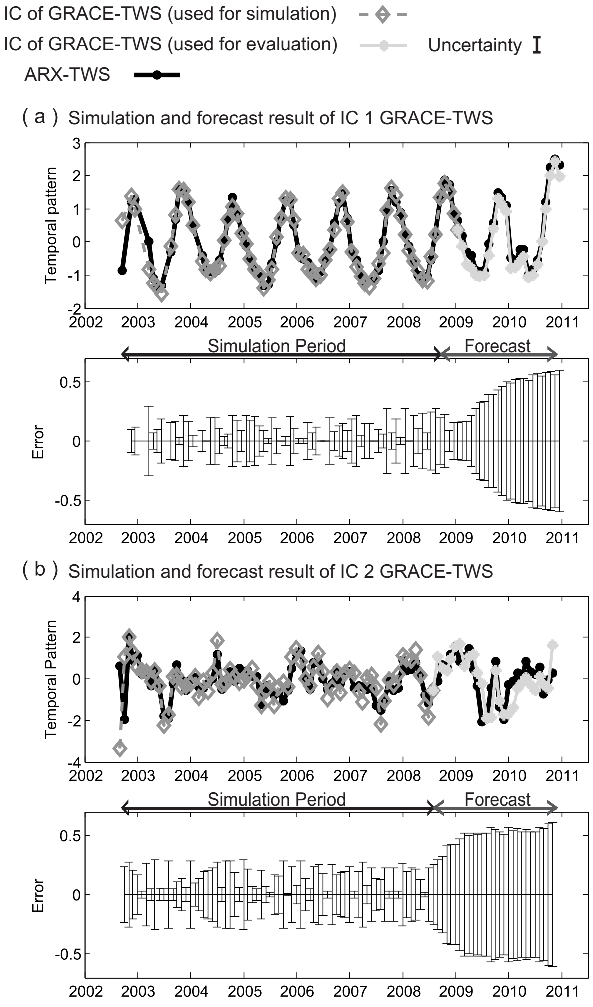


Figure 5.24: Results of simulations and forecasts of IC1 and IC2 of GRACE-TWS (Fig. 5.22), using the ARX models (shown by black-lines). Figure 5.24 (a-top) represents the results for IC1 of GRACE-TWS while using the ICs of SST and TRMM-rainfall as indicators. Figure 5.24 (a-bottom) shows the uncertainty of the forecast on top. Figure 5.24 (b-top) represents the same results as (a) but corresponding to IC2 of GRACE-TWS. Figure 5.24 (b-bottom) indicates the uncertainty of the forecast on top. For simulation, the first 72 months of TWS are used (shown in dark-gray). TWS values after the 72'th month are then used for evaluating the forecasts (shown in light-gray). The temporal patterns in this figure are unit-less.

5.5.4 Summary and Discussion of the Forecast Results

In this section, a new statistical multivariate seasonal forecasting approach was introduced, which can be used to forecast TWS changes over West Africa. It is of interest to develop statistical approaches rather than physical models because the simplified physical equations in the latter prevents an accurate estimation of TWS changes (see also Grippa et al., 2011). The proposed ICA/ARX approach does not directly simulate the complex physical process of ocean-land-atmosphere over West Africa, but instead, it statistically learns the relationships between the main physical processes of the region (such as teleconnections and the soil-precipitation feedback) and uses it to predict TWS changes. The successful implementation of the ICA/ARX approach relies on the proper selection of TWS indicators and avoiding over-parametrization of the model, data quality, as well as a learning phase, which has to reflect the relationships between predictors and predictands. To investigate this issue, we performed a numerical validation, which showed that the seasonal forecast of TWS is close to TWS that is actually measured by GRACE. We also applied a numerical uncertainty assessment to account for the sampling errors and data noise. Our results showed that the TWS forecast of ICA/ARX is valid to forecast TWS in a period of up to about two years.

It is worth mentioning here that since the proposed method is trained on the GRACE products, it provides relatively coarse resolution TWS maps. The approach also assumes that the spatial pattern of TWS changes remains stationary within the two years of the forecast. Before applying this approach to other regions, one should analyze whether this assumption holds for different time frames. In fact, for West Africa, the large-scale interactions between ocean and atmosphere are the major climate driver. Subsequently, selecting SST as the indicator of TWS seems to be logical. For implementing this approach to other study areas, one might revisit the meaningfulness of the indicators. Another issue is that the training of the ARX model was performed based on six years of data. Since SST and rainfall are available for a longer period (e.g., for TRMM, after 1998), one could use TWS outputs of models for the time before October 2002 and extend the training period. Addressing the impact of such extension in terms of the quality of the ARX coefficients and the consistency of the model-derived TWS changes with those of GRACE is, however, out of scope of this section. Regarding the numerical results, one might conclude that the presented statistical method could be helpful for filling the current data gaps of the GRACE products and a possible gap period between GRACE and its follow-on mission at least over West Africa.

6. Conclusion and Outlook

This research was motivated by a main research question:
'how can statistical signal separation methods be used to explore information from GRACE integrated observations?'.
To answer this question, a number of detailed questions were defined that mainly focused on the mathematical formulations of statistical decomposition techniques. After introducing GRACE data and its processing steps to derive accurate total water storage (TWS) values in Chapter 2, a variety of second order statistical decomposition methods such as the Principal Component Analysis (PCA) also called Empirical Orthogonal Function (EOF), and its extensions were introduced in Chapter 3. The assumptions within each method were addressed and it was mathematically investigated how these methods can be used to reduce the high dimensionality of the available data sets. In Chapter 4, the Independent Component Analysis (ICA) approach was introduced as a higher order statistical technique to extract information underlined in the GRACE-derived time-variable TWS signals. In Chapter 5, the applicability of the introduced statistical approaches to decompose time series of TWS changes were investigated via simulations and several real case studies. In the following, the main conclusions and outlooks are presented.

6.1 Conclusion

Relationships between the Second and Higher Order Statistical Methods

From the introduced statistical decomposition methods in Chapters 3 and 4 the use of ICA is quite recent in geophysical studies, thus, the number of research papers and dissertations are limited. Most of the available references looked at the ICA problem in the Blind Source Separation (BSS) context. Therefore, connecting the decomposition formulations, usually used for the second order methods in Chapter 3, and the 'source' and 'mixing/de-mixing' components that are introduced in the BSS context (Chapter 4) was not properly investigated prior to this study. Therefore, we followed Comon (1994a), Aires et al. (2002), Hannachi et al. (2009), and Forootan and Kusche (2012), by introducing the ICA approach as a rotation extension of PCA (see Section 4.1). Then, the algebraic relationships between the second order and higher order statistical signal decomposition techniques were introduced. The concept of source and mixing/de-mixing was defined with respect to both second and higher order statistical information.

Incorporation of the higher-order statistical moments in the decomposition procedure originated from the Rotated EOF techniques (REOF) (Richman, 1986, Hannachi et al., 2007). In Section 3.2, we showed that the common criteria of VARIMAX and QUARTIMAX (Kaiser, 1958) are related to the variance of the squared rotated components, indicating that higher-order statistical moments can be useful to enhance interpretation of the decomposition results. In Chapter 4, we discussed the benefit of using the independence criterion as a stronger statistical assumption (compared to orthogonality) within the separation process. We found that the ICA approach considerably mitigates the clustering behaviors that usually occur after application of the second order statistical decomposition techniques.

Estimation of the Statistical Independence

Among the ICA criteria, we found those based on the diagonalization of higher order cumulants being more straight forward and computationally efficient to be used for decomposing the non-

Gaussian time-variable gravimetry data than those ICA methods based on entropy (Section 4.5.4). As for the joint diagonalization of the fourth-order cumulant tensor (Section 4.5.1), used in this study, we mathematically proved that the ICA algorithm perfectly separated an unknown mixture of a linear trend and sinusoids, under the assumption that the length of data sets is infinite (Forootan and Kusche, 2013).

Identifying non-stationary patterns in TWS time series is important to better understand dynamical changes within the water cycle. Therefore, we applied a Hilbert transformation to include information about the rate of change of the original data sets in the decomposition procedure. Then, the ICA algorithm, based on the diagonalization of the fourth-order cumulant tensor (Forootan and Kusche, 2012, 2013), was extended to a complex case (Section 4.5.3). The method was accordingly called 'Complex ICA'. Finally, required mathematical formulations to compute non-stationary spatial and temporal components (extracted by Complex ICA) were presented.

Various approaches were also introduced to test the significance of the components. To apply most of these methods, however, one has to assume that observations are randomly sampled from a certain population. In this thesis, we chose a resampling approach that seems to be appropriate since it needs fewer prior assumptions about the distribution of random variables (observations). Therefore, the sampling problem, caused by the limited time span of data sets, is better taken into account by numerically repeating the decomposition procedure.

Applications of the Second and Higher Statistical Techniques

From an application point of view, we found that the decomposition results derived from each of the introduced methods are more representative than a simple parametric decomposition of TWS series into a trend map and maps of annual, semiannual and certain aliasing contributions would show. The latter method always relies on a simple a priori model of temporal behavior, which disregards inter-annual or episodic features as 'residuals'. In contrast, PCA and ICA approaches always contain the full temporal information, and smoothing or compression of information happens by disregarding spatial patterns of apparently less significance.

The geometrical property of PCA was found to be very useful since the covariance matrix of any subset of retained PCs is always diagonal. PCA also captures the dominant part of the variance in the data set when the components are ordered with respect to the descending magnitude of singular values. From an interpretation point of view when the components are treated individually, however, the PCA method can also be misleading since it combines many of the actually separate signals into its retained components (see Section 5.1.1). In those cases, alternative approaches (such as ICA) are more reasonable.

Among the introduced methods, what makes the ICA approach interesting to be applied for decomposing non-Gaussian geophysical observation, was the hypothesis in Chapter 4 that states: independent physical processes most likely generate statistically independent source signals that are superimposed in e.g., the GRACE time-variable observations. Therefore, decomposing them into maximally statistically independent components leads to base-functions that are physically at least more representative than others based on orthogonality only (Forootan and Kusche, 2012). For applications that look at each separated components individually (Fenoglio-Marc, 2001, Rieser et al., 2010), or those aiming at separating GRACE and other data into signals from different compartments (Schmeer et al., 2012, Rietbroek, 2014), therefore, the independent components derived from the ICA decomposition seem to be better suited than PCA or other ordinary extensions. In Section 5.1, within a variety of simulations with known input signals and artificially generated noise, one could see that the ICA-derived components were much closer to the optimal decomposition when compared to PCA and its VARIMAX-rotated extension. Although theoretically one cannot know what the true decomposition of TWS into its source signals is, these results provided confidence that the ICA method is a useful tool to separate complex signals.

An application of ICA to explore real GRACE time-variable products was shown in Chapter 5. Other examples can also be found in, e.g., Omondi et al. (2013a) who applied ICA to extract the main independent behavior of water storage changes over the Greater Horn of Africa and related them to the climate condition of the region. Applications of ICA to extract information from GRACE-derived TWS products over Australia, Iran, and the Nile Basin were shown by Forootan et al. (2012, 2014c) and Awange et al. (2014), respectively. Forootan et al. (2014b) proposed a forecasting framework to combine the ICA and autoregressive methods and predict West Africa's TWS changes based on remotely sensed rainfall changes of the region and sea surface temperature changes over the tropical oceans (see also Section 5.5).

Considering orthogonal rotation matrices to formulate Spatial ICA and Temporal ICA in Section 4.3, we demonstrated that always one set of ICA components remains orthogonal (i.e. temporal components in Spatial ICA and spatial components in Temporal ICA) while the other set being as statistically independent as possible. This property was found practically beneficial since in some applications. For example in Forootan et al. (2014c), Temporal ICA was applied to extract better interpretable independent temporal patterns of terrestrial and surface water storage changes. Then, their corresponding orthogonal spatial patterns were used as base-functions in a Least Squares Adjustment (LSA) procedure to partition GRACE-TWS changes (see also Section 5.4). Obviously, retaining the orthogonality of the spatial components (in Temporal ICA) mitigate numerical problems while performing the least squares inversion.

In a recent study, Boergens et al. (2014) applied the Spatio-temporal ICA (StICA) and the Temporal ICA techniques, which were formulated based on the entropy criterion (Hyvärinen, 1999b) to identify patterns of gravity changes over North America and the African continent. Their results indicated a slightly better separation performance of StICA compared to those of Temporal ICA. The StICA technique is slightly different from the Spatial ICA and Temporal ICA introduced in this thesis. StICA searches for the patterns in the data set that contain small dependences in space and time. Therefore, the StICA-derived patterns are not strictly independent with compared to the ICs derived from the introduced Spatial ICA and Temporal ICA techniques. In practice, however, the results of StICA might be easier to interpret since in reality there exist small dependences between different spatial, as well as between different temporal source signals. For example, one can consider that two regions exhibit similar TWS changes but with slightly different time latencies. Thus, their temporal changes would be statistically correlated. An application of the Temporal ICA to separate TWS changes over these two areas results to a clustered behavior as it was shown in the simulation study of Section 5.1.3. One might argue that in such cases a trade off between the mutual independence of the spatial and temporal patterns (as provided by StICA) likely mitigates the clustered behavior. However, it is worth mentioning that after application of StICA, both of the StICA-derived spatial and temporal components are not anymore orthogonal. Furthermore, StICA maximizes the independence of sources over space and time, without necessarily producing independence in either space or time. Therefore, in case of comparable outcomes, we recommend the use of either Temporal or Spatial ICA due to the mentioned computational and statistical benefits.

To extract patterns with propagating nature, the Hilbert Complex EOF (Section 3.4) and Complex ICA (Section 4.5.3) algorithms are preferred over the decomposition methods that work based on lagged information such as the Extended EOF (EEOF) approach (Section 3.3) or Conventional Complex EOF. The application of the Hilbert transform does not require the predefinition of a temporal lag that is required, e.g., in the EEOF method. This was illustrated in Sections 5.1.3 and 5.2, where Complex ICA was respectively used to indicate the spatial and temporal amplitude of a simulated as well as real GRACE-derived El Niño pattern and its phase propagations.

6.2 Outlook

Improvements in the ICA Algorithm

The uncertainty estimation technique, introduced in this thesis, has the potential to be improved by sampling techniques that generate random numbers while considering the second and/or higher order statistical relationships between components (e.g., Preisendorfer et al., 1981, Preisendorfer, 1988). Another approach would be realized from a bootstrapping approach in a Bayesian framework, which creates new data sets through reweighting the initial data (Rubin, 1981). A common problem also occurs in the joint diagonalization of the fourth-order cumulant tensor, where the entries of the tensor might be biased due to the limited time span of observations. This will be investigated in a Bayesian framework by simulating samples of the common eigenstructure of the tensor and estimate these uncertainties.

The ICA criterion, based on joint diagonalization, has been generalized by, e.g., Moreau (2001) to include a variety of higher order cumulants. Such extension, however, should be applied to GRACE time series with care since adding more cumulants requires the computation of more statistical moments from the length-limited time series, which itself introduces uncertainty. A rigorous investigation of such extensions will be addressed in future research.

The joint diagonalization problem was solved in this thesis using a Jacobi eigenvalue algorithm (Cardoso and Souloumiac, 1993). The computational load of this approach is extensive when the number of modes is too big. Thus, a tuned optimization algorithm such as a faster gradient decent technique will be considered in future.

Future Applications

The Temporal and Complex ICA techniques were applied to extract the El Niño pattern from time series of GRACE-TWS changes. The El Niño pattern, however, might be different in the compartments of TWS (such as soil moisture and groundwater) or in the fluxes (such as precipitation and evapotranspiration) because of their different response mechanisms to the phenomenon. A consistent extraction of the El Niño pattern from TWS changes, storage compartments, and fluxes is an ongoing research.

The possibility of using ICA to reduce a combination of the spectral and spatial leakage was shown in the regional simulation case (Section 5.1.2). Our results in Section 5.3 indicated that the Spatial ICA method can successfully separate mass anomalies with a distinct spatial patterns, even though the temporal characteristics represent quite a similar behavior. However, when the positions of spatial anomalies are considerably close to each other this application might be limited. Defining a suitable measure to indicate which types of anomalies can be separated by applying this approach needs further research.

In Section 5.4, a statistical approach was introduced to partition GRACE-TWS changes into terrestrial and surface water storage changes, as well as groundwater storage changes (Forootan et al., 2014c). Uncertainty estimation of such partitioning will be addressed in future studies. The results of statistical partitioning will be compared to the studies that directly account for the storage contribution of different compartments (Longuevergne et al., 2010, Landerer and Swenson, 2012), as well as the outputs of hydrological models after their assimilation and calibration with GRACE products (Schumacher, 2012).

A framework via combination of ICA and low-degree Autoregressive Model with Exogenous Variables (ARX) models was introduced in Section 5.5 to statistically forecast TWS changes (Forootan et al., 2014b). Forecast skills of the proposed approach might be improved by applying the Complex ICA approach. Relating complex components in the ARX models needs further research.

Statistical signal decomposition approaches, introduced in this study, have the potential to be used for downscaling of GRACE-derived TWS fields by identifying empirical links between large-scale patterns of TWS changes (predictors) and local model-derived storage variations (the predictand). Such extensions will be assessed in future researches.

Acronyms

APMG	**A**stronomical **P**hysical and **M**athematical **G**eodesy Group, University of Bonn
ARX	**A**utoregressive **M**odel with **E**xogenous Variables
AWRA	**A**ustralian **W**ater **R**esources **A**ssessment
AOD	**A**tmospheric-**O**ceanic **D**e-aliasing
BSS	**B**lind **S**ource **S**eparation
C-CEOF	**C**onventional-**C**omplex **E**mpirical **O**rthogonal **F**unction
CEOF	**C**omplex **E**mpirical **O**rthogonal **F**unction
CICA	**C**omplex **I**ndependent **C**omponent **A**nalysis
CM	**C**enter of **M**ass
CSR	**C**enter for **S**pace **R**esearch, University of Texas at Austin
DLR	**D**eutscheszentrum für **L**uft- und **R**aumfahrt (German Aerospace Center)
DUT	**D**elft **U**niversity of **T**echnology
ECMWF	**E**uropean **C**enter for **M**edium-range **W**eather **F**orecast
EEOF	**E**xtended **E**mpirical **O**rthogonal **F**unction
ENSO	**E**l Niño **S**outhern **O**scillation
EOF	**E**mpirical **O**rthogonal **F**unction
EWH	**E**quivalent **W**ater **H**eight
FastICA	**FastI**ndependent **C**omponent **A**nalysis
FFT	**Fast** **F**ourier **T**ransform
GFZ	**G**eo**F**orschungs**Z**entrum Potsdam
GIA	**G**lacial **I**sostatic **A**djustment
GLDAS	**G**lobal **L**and **D**ata **A**ssimilation **S**ystem Hydrology Model
GNSS	**G**lobal **N**avigation **S**atellite **S**ystems
GPS	**G**lobal **P**ositioning **S**ystem
GRACE	**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment
GRACE-FO	**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment-**F**ollow **O**n
GRGS	**G**roupe de **R**echerches de **G**eodesie **S**patiale
GSFC	**G**oddard **S**pace **F**light **C**enter
H-CEOF	**H**ilbert-**C**omplex **E**mpirical **O**rthogonal **F**unction
IC	**I**ndependent **C**omponent
ICA	**I**ndependent **C**omponent **A**nalysis
IOD	**I**ndian **O**cean **D**ipole
ICESat	**I**ce **C**loud and **L**and **E**levation **Sat**ellite
ITG	**I**nstitute of **T**heoretical **G**eodesy, University of Bonn
JADE	**J**oint **A**pproximate **D**iagonalization of **E**igenmatrices
JD	**J**oint **D**iagonalization
JPL	**J**et **P**ropulsion **L**aboratory
KBR	**K**-**B**and **R**anging system
KS	**K**olmogorov-**S**mirnov Test
LSA	**L**east **S**quares **A**djustment
LSM	**L**and **S**urface **M**odel
MSSA	**M**ulti-**C**hannel **S**ingular **S**pectrum **A**nalysis

NASA	**N**ational **A**eronautics and **S**pace **A**dministration
OMCT	**O**cean **M**odel for **C**irculation and **T**ides
PC	**P**rincipal **C**omponent
PCA	**P**rincipal **C**omponent **A**nalysis
PDF	**P**robability **D**istribution **F**unction
REOF	**R**otated **E**mpirical **O**rthogonal **F**unction
RMS	**R**oot **M**ean **S**quare
SHC	**S**pherical **H**armonic **C**oefficient
S-CICA	**S**patial **C**omplex **I**ndependent **C**omponent **A**nalysis
SICA	**S**patial **I**ndependent **C**omponent **A**nalysis
SLR	**S**atellite **L**aser **R**anging
SOI	**S**outhern **O**scillation **I**ndex
StICA	**S**patio-temporal **I**ndependent **C**omponent **A**nalysis
SSA	**S**ingular **S**ystem(**S**pectrum) **A**nalysis
SSH	**S**ea **S**urface **H**eight
SST	**S**ea **S**urface **T**emperature
SVD	**S**ingular **V**alue(**V**ector) **D**ecomposition
T-CICA	**T**emporal **C**omplex **I**ndependent **C**omponent **A**nalysis
TICA	**T**emporal **I**ndependent **C**omponent **A**nalysis
TRMM	**T**ropical **R**ainfall **M**easuring **M**ission
TWS	**T**otal **W**ater **S**torage
USGS	**U**nited **S**tates **G**eological **S**urvey
WGHM	**W**aterGAP **G**lobal **H**ydrology **M**odel
WS	**W**ater **S**torage

List of Figures

List of Tables

References

Afsari, B., and P.S. Krishnaprasad (2004). *Some gradient based joint diagonalization methods for ICA*. In: *Independent Component Analysis and Blind Signal Separation*, Vol. 3195 Series *Lecture Notes in Computer Science*, 437–444. Springer.

Aires, F., W.B. Rossow and A. Chedin (2002). *Rotation of EOFs by the independent component analysis: toward a solution of the mixing problem in the decomposition of geophysical time series*. Journal of the Atmospheric Sciences, 59:111–123.

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, 54:23–36.

Awange, J.L., K.M. Fleming, M. Kuhn, W.E. Featherstone, B. Heck and I. Anjasmara (2011). *On the suitability of the* $4^\circ \times 4^\circ$ *GRACE mascon solutions for remote sensing Australian hydrology*. Remote Sensing of Environment, 115 (3):864–875.

Awange, J.L., E. Forootan, J. Kusche, J.B.K. Kiema, P.A. Omondi, B. Heck, K. Fleming, S.O. Ohanya and R.M. Gonçalves (2013). *Understanding the decline of water storage across the Ramser-Lake Naivasha using satellite-based methods*. Advances in Water Resources, 60:7–23.

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 to 2011*. Advances in Water Resources, in-press.

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.

Beckers, J.M., and M. Rixen (2003). *EOF calculations and data filling from incomplete oceanographic datasets*. Journal of Atmospheric and Oceanic Technology, 20:1839–1856.

Beek, T., L. Menzel, R. Rietbroek, L. Fenoglio-Marc, S. Grayek, M. Becker, J. Kusche and E.V. Stanev (2012). *Modeling the water resources of the Black and Mediterranean Sea river basins and their impact on regional mass changes*. Journal of Geodynamics, 59-60:157–167.

Berry, P.A.M., J.D. Garlick, J.A. Freeman and E.L. Mathers (2005). *Global inland water monitoring from multi-mission altimetry*. Geophysical Research Letters, 32 (16):L16401.

Beven, K. (2001). *How far can we go in distributed hydrological modelling?* Hydrology and Earth System Sciences, 5 (1):1–12.

Birkett, C. (1995). *The global remote sensing of lakes, wetlands and rivers for hydrological and climate research*. In: *Geoscience and Remote Sensing Symposium, 1995. IGARSS '95. 'Quantitative Remote Sensing for Science and Applications', International*, Vol. 3, 1979–1981.

Blaschke, T., and L. Wiskott (2004). *CuBICA: independent component analysis by simultaneous third- and fourth-order cumulant diagonalization*. Signal Processing, IEEE Transactions on, 52 (5):1250–1256.

Boergens, E., E. Rangelova, M. G. Sideris and J. Kusche (2014). *Assessment of the capabilities of the temporal and spatio-temporal ICA method for geophysical signal separation in GRACE data.* Journal of Geophysical Research: Solid Earth, 119:4429–4447.

Boy, J-P., and B.F. Chao (2005). *Precise evaluation of atmospheric loading effects on Earth's time-variable gravity field.* Journal of Geophysical Research: Solid Earth, 110:B08412.

Bretherton, C.S., M. Widmann, V.P. Dymnikov, J.M. Wallace and I. Bladé (1999). *The effective number of spatial degrees of freedom of a time-varying field.* Journal of Climate, 12:1990–2009.

Broomhead, D.S., and G.P. King (1986a). *Extracting qualitative dynamics from experimental data.* Physica D Nonlinear Phenomena, 20 (2-3):217–236.

Broomhead, D.S., and G.P. King (1986b). *On the qualitative analysis of experimental dynamical systems.* Nonlinear Phenomena and Chaos, 113–144.

Brown, N.J., and P. Tregoning (2010). *Quantifying GRACE data contamination effects on hydrological analysis in the Murray-Darling Basin, southeast Australia.* Australian Journal of Earth Sciences, 57 (3):329–335.

Browne, M.W. (2001). *An overview of analytic rotation in exploratory factor analysis.* Multivariate Behavioral Research, 36 (1):111–150.

Cai, W., P. van Rensch, T. Cowan and H.H. Hendon (2011). *Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall.* Journal of Climate, 24: 3910–3923.

Cardoso, J-F. (1992). *Fourth-order cumulant structure forcing: application to blind array processing.* In: *Statistical Signal and Array Processing, 1992. Conference Proceedings., IEEE Sixth SP Workshop on*, 136–139.

Cardoso, J-F. (1999). *High-order contrasts for independent component analysis.* Neural Computatation, 11:157–192.

Cardoso, J-F., and A. Souloumiac (1993). *Blind beamforming for non-Gaussian signals.* IEEE Proceedings, 140:362–370.

Cardoso, J-F., and A. Souloumiac (1995). *Jacobi angles for simultaneous diagonalization.* SIAM, Journal of Mathematical Analysis and Applications, 17:161–164.

Case, K., G. Kruizinga and S. Wu (2002). *GRACE level 1B data product user handbook.* Technichal Report, JPL Publication.

Cattell, R.B. (1966). *The scree test for the number of factors.* Multivariate Behavioral Research, 1:245–276.

Chambers, D.P. (2006). *Observing seasonal steric sea level variations with GRACE and satellite altimetry.* Journal of Geophysical Research: Oceans, 111:C03010.

Chambers, D.P., and J.K. Willis (2008). *Analysis of large-scale ocean bottom pressure variability in the North Pacific.* Journal of Geophysical Research: Oceans, 113:C11003.

Chatfield, C. (1989). *The analysis of time series: an introduction.* Chapman and Hall/CRC.

Chen, J.L., C. Wilson, B.D. Tapley and J.C. Ries (2004). *Low degree gravitational changes from GRACE: Validation and interpretation.* Geophysical Research Letters, 31:L22607.

Chen, J.L., C. Wilson, , D.D. Blankenship and B.D. Tapley (2006). *Antarctic mass rates from GRACE.* Geophysical Research Letters, 33:L11502.

Chen, J.L., C.R. Wilson and K-W. Seo (2009). *S2 tide aliasing in GRACE time-variable gravity solutions.* Journal of Geodesy, 83 (7):679–687.

Cheng, M.K., and B.D. Tapley (2004). *Variations in the Earth's oblateness during the past 28 years.* Journal of Geophysical Research: Solid Earth, 109:B09402.

Comon, P. (1994a). *Independent component analysis, A new concept?* Signal Processing, 36 (3): 287–314.

Comon, P. (1994b). *Tensor diagonalization, a useful tool in signal processing.* In: *IFAC Symposium on System Identification*, 77–82. IFAC-SYSID.

Conover, W.J. (1999). *Practical nonparametric statistics.* Wiley, New York.

de Viron, O., I. Panet and M. Diament (2006). *Extracting low frequency climate signal from GRACE data.* eEarth, 1 (1):9–14.

Dobslaw, H., F. Flechtner, I. Bergmann-Wolf, C. Dahle, R. Dill, S. Esselborn, I. Sasgen and M. Thomas (2013). *Simulating high-frequency atmosphere-ocean mass variability for dealiasing of satellite gravity observations: AOD1B RL05.* Journal of Geophysical Research: Oceans, 118(7):3704–3711.

Döll, P., F. Kaspar and B. Lehner (2003). *A global hydrological model for deriving water availability indicators: model tuning and validation.* Journal of Hydrology, 207:105–134.

Douville, H., S. Conil, S. Tyteca and A. Voldoire (2007). *Soil moisture memory and West African monsoon predictability: artefact or reality?* Climate Dynamics, 28 (7-8):723–742.

Duan, J., C.K. Shum, J. Guo and Z. Huang (2012). *Uncovered spurious jumps in the GRACE atmospheric de-aliasing data: potential contamination of GRACE observed mass change.* Geophysical Journal International, 191 (1):83–87.

Efron, B. (1979). *Bootstrap methods: another look at the Jackknife.* The Annals of Statistics, 7: 1–26.

Ellett, K.M., J.P. Walker, A.W. Western and M. Rodell (2006). *A framework for assessing the potential of remote sensed gravity to provide new insight on the hydrology of the Murray-Darling Basin.* Australian Journal of Water Resources, 10 (2):89–101.

Elsaka, B. (2010). *Simulated satellite formation flights for detecting the temporal variations of the Earth's gravity field.* PhD Thesis, University of Bonn, Germany.

Enfield, D.B., and A.M. Mestas-Nuñez (1999). *Global modes of ENSO and non-ENSO SST variability and their associations with climate.* In: Diaz, H.F., and V. Markgraf (Hrsg.), *El Niño and the Southern Oscillation: Multiscale Variability and Its Impacts on Natural Ecosystems and Society*, 89–112. Cambridge University Press.

Famiglietti, J.S., and M. Rodell (2013). *Water in the balance.* Science, 340(6138):1300–1301.

Farrell, W.E. (1972). *Deformation of the Earth by surface loads.* Reviews of Geophysics, 10 (3): 761–797.

Farrell, W.E., and J.A. Clark (1976). *On postglacial sea level.* Geophysical Journal of the Royal Astronomical Society, 46 (3):647–667.

Fenoglio-Marc, L. (2001). *Analysis and representation of regional sea-level variability from altimetry and atmospheric-oceanic data.* Geophysical Journal International, 145 (1):1–18.

Fenoglio-Marc, L., J. Kusche and M. Becker (2006). *Mass variation in the Mediterranean Sea from GRACE and its validation by altimetry, steric and hydrologic fields.* Geophysical Research Letters, 33 (19):L19606.

Fenoglio-Marc, L., R. Rietbroek, S. Grayek, M. Becker, J. Kusche and E. Stanev (2012). *Water mass variation in the Mediterranean and Black Seas.* Journal of Geodynamics, 59-60:168–182.

Ferreira, P., J. Magueijo and J. Silk (1997). *Cumulants as non-Gaussian qualifiers.* Physical Review, 56:4592–4603.

Ferréol, A., L. Albera and P. Chevalier (2005). *Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM).* Signal Processing, IEEE Transactions on, 53 (5): 1640–1653.

Févotte, C., and F.J. Theis (2007). *Orthonormal approximate joint block-diagonalization.* Technical report, Groupe Audio, Acoustique et Ondes, Telecom Paris.

Flechtner, F. (2007a). *AOD1B product description document for product releases 01 to 04.* Technichal Report, GeoForschungsZentrum Potsdam.

Flechtner, F. (2007b). *GFZ Level-2 processing standards document for level-2 product release 0004.* Technichal Report, GeoForschungsZentrum Potsdam.

Flechtner, F., M. Thomas and H. Dobslaw (2010). *Improved Non-tidal Atmospheric and Oceanic De-aliasing for GRACE and SLR Satellites.* Advanced Technologies in Earth Sciences, 131–142.

Flechtner, F., C. Dahle, C. Gruber, R. König, G. Michalak and K-H. Neumayer (2013). *The GFZ RL05 GRACE gravity field model time series.* Poster EGU2013-2993, Geophysical Research Abstracts, Vol. 15, EGU2013-2993, EGU General Assembly 2013, Vienna, Austria., 2013.

Forootan, E., and J. Kusche (2012). *Separation of global time-variable gravity signals into maximally independent components.* Journal of Geodesy, 86 (7):477–497.

Forootan, E., and J. Kusche (2013). *Separation of deterministic signals using independent component analysis (ICA).* Studia Geophysica et Geodaetica, 57 (1):17–26.

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.

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.

Forootan, E., O. Didova, M. Schumacher, J. Kusche and B. Elsaka (2014a). *Comparisons of atmospheric mass variations derived from ECMWF reanalysis and operational fields.* Journal of Geodesy, 88:503–514.

Forootan, E., J. Kusche, I. Loth, W-D. Schuh, A. Eicker, J.L. Awange, L. Longuevergne, B. Diekkrüger, M. Schmidt and C.K. Shum (2014b). *Multivariate prediction of total water storage changes over West Africa from multi-satellite data.* Surveys in Geophysics, 35 (4): 913–940.

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.

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

Frappart, F., G. Ramillien, P. Maisongrande and M-P. Bonnet (2011b). *Denoising satellite gravity signals by independent component analysis.* Geoscience and Remote Sensing Letters, IEEE, 7 (3):421–425.

García-García, D., C.C. Ummenhofer and V. Zlotnicki (2011). *Australian water mass variations from GRACE data linked to Indo-Pacific climate variability.* Remote Sensing of Environment, 115 (9):2175–2183.

Goddard, L., S.J. Mason, S.E. Zebiak, C.F. Ropelewski, R. Basher and M.A. Cane (2001). *Current approaches to seasonal to interannual climate predictions.* International Journal of Climatology, 21 (9):1111–1152.

Golyandina, N., V. Nekrutkin and A. Zhigljavsky (2001). *Analysis of time series structure - SSA and related techniques.* Chapman & Hall.

Grippa, M., L. Kergoat, F. Frappart, Q. Araud, A. Boone, P. de Rosnay, J-M. Lemoine, S. Gascoin, G. Balsamo, C. Ottlé, B. Decharme, S. Saux-Picart and G. Ramillien (2011). *Water storage variability over West Africa estimated by Gravity Recovery and Climate Experiment (GRACE) and land surface models.* Water Resources Research, 47:W05549.

Güntner, A. (2008). *Improvement of global hydrological models using GRACE data.* Surveys in Geophysics, 29 (4-5):375–397.

Guo, J.Y., Z.W. Huang, C.K. Shum and W. van der Wal (2012). *Comparisons among contemporary glacial isostatic adjustment models.* Journal of Geodynamics, 61:129–137.

Han, S-C., C. Jekeli and C.K. 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:B04403.

Han, S-C., C.K. 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.

Hannachi, A., I.T. Jolliffe and D.B. Stephenson (2007). *Empirical orthogonal functions and related techniques in atmospheric science: a review.* International Journal of Climatology, 27 (9):1119–1152.

Hannachi, A., S. Unkel, N.T. Trendafilov and I.T. Jolliffe (2009). *Independent component analysis of climate data: A new look at EOF rotation.* Journal of Climate, 22:2797–2812.

Heiskanen, W.A., and H. Moritz (1967). *Physical geodesy.* Freeman, San Francisco.

Horel, J. D. (1984). *Complex principal component analysis: theory and examples.* Journal of Applied Meteorology, 23 (12):1660–1673.

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):W07525.

Huffman, G., and D. Bolvin (2012). *TRMM and other data precipitation data set documentation.* Technichal Report, Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center and Science Systems and Applications, Inc.

Hyvärinen, A. (1997). *A family of fixed-point algorithms for independent component analysis.* In: *Acoustics, Speech, and Signal Processing, 1997. ICASSP-97., 1997 IEEE International Conference on*, Vol. 5, 3917–3920.

Hyvärinen, A. (1999a). *Survey on independent component analysis.* Neural Computing Surveys, 2:94–128.

Hyvärinen, A. (1999b). *Fast and robust fixed-point algorithms for independent component analysis.* Neural Networks, IEEE Transactions on, 10 (3):626–634.

Hyvärinen, A., and E. Oja (2000). *Independent component analysis: algorithms and applications.* Neural Networks, IEEE Transactions on, 13:411–430.

Ilk, K-H., J. Flury, R. Rummel, P. Schwintzer, W. Bosch, C. Haas, J. Schröter, D. Stammer, W. Zahel, H. Miller, R. Dietrich, P. Huybrechts, H. Schmeling, D. Wolf, J. Riegger, A. Bardossy and A. Güntner (2005). *Mass transport and mass distribution in the Earth system - Contribution of the new generation of satellite gravity and altimetry missions to geosciences.* `http://www.massentransporte.de/fileadmin/Dokumente/programmschrift-Ed2.pdf`.

Jekeli, C. (1981). *Alternative methods to smooth the Earth's gravity field.* Reports of the Department of Geodetic Science, 327, Ohio State University, Columbus, Ohio.

Jensen, L., R. Rietbroek and J. Kusche (2013). *Land water contribution to sea level from GRACE and Jason-1measurements.* Journal of Geophysical Research: Oceans, 118 (1):212–226.

Jolliffe, I.T. (1986). *Principal component analysis.* Springer, New York.

Jolliffe, I.T. (1987). *Rotation of principal components: some comments.* Journal of Climatology, 7:507–510.

Jolliffe, I.T. (1989). *Rotation of ill-defined principal components.* Journal of the Royal Statistical Society, 38:139–147.

Jolliffe, I.T. (1995). *Rotation of principal components: choice of normalization constraints.* Journal of Applied Statistics, 22:29–35.

Jolliffe, I.T. (2003). *A cautionary note on artificial examples of EOFs.* Journal of Climate, 16: 1084–1086.

Jolliffe, I.T., M. Uddin and S.K. Vines (2002). *Simplified EOFs-three alternatives to retain.* Climate Research, 20:271–279.

Kaiser, H.F. (1958). *The varimax criterion for analytic rotation in factor analysis.* Psychometrika, 23 (3):187–200.

Kanzow, T., F. Flechtner, A. Chave, R. Schmidt, P. Schwintzer and U. Send (2005). *Seasonal variation of ocean bottom pressure derived from Gravity Recovery and Climate Experiment (GRACE): local validation and global patterns.* Journal of Geophysical Research: Oceans, 110: C09001.

Khaki, M., E. Forootan and M.A. Sharifi (2014). *Waveform retracking over the Caspian Sea.* International Journal of Remote Sensing, in-press.

Kirimoto, T., T. Amishima and A. Okamura (2011). *Separation of mixtures of complex sinusoidal signals with independent component analysis.* IEICE Transactions on Communications, E94.B (1):215–221.

Klees, R., E.A. Zapreeva, H.C. Winsemius and H.H.G. Savenije (2006). *The bias in GRACE estimates of continental water storage variations.* Hydrology and Earth System Sciences Discussions, 3 (6):3557–3594.

Klees, R., A. Revtova, B.C. Gunter, P. Ditmar, E. Oudman, H.C. Winsemius and H.H.G. Savenije (2008). *The design of an optimal filter for monthly GRACE gravity models.* Geophysical Journal International, 175:417–432.

Knudsen, P. (2003). *Ocean tides in GRACE monthly averaged gravity fields.* Space Science Reviews, 108 (1-2):261–270.

Koch, K.R. (1990). *Bayesian inference with geodetic applications.* Springer, Berlin.

Koch, K.R. (1999). *Parameter estimation and hypothesis testing in linear models.* Springer-second edition, New York.

Kouraev, A.V., J.-F. Crétaux, S.A. Lebedev, A.G. Kostianoy, A.I. Ginzburg, N.A. Sheremet, R. Mamedov, E.A. Zakharova, L. Roblou, F. Lyard, S. Calmant and M. Bergé-Nguyen (2011). *Satellite altimetry applications in the Caspian Sea.* In: Vignudelli, S., A.G. Kostianoy, P. Cipollini and J. Benveniste (Hrsg.), *Coastal Altimetry*, 331–366. Springer Berlin Heidelberg.

Kusche, J. (2007). *Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models.* Journal of Geodesy, 81(11):733–749.

Kusche, J., and E.J.O. Schrama (2005). *Surface mass redistribution inversion from global GPS deformation and Gravity Recovery and Climate Experiment (GRACE) gravity data.* Journal of Geophysical Research: Solid Earth, 110(9):1–13.

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.

Kusche, J., A. Eicker and E. Forootan (2011). *Analysis tools for GRACE and related data sets, theoretical basis.* Technichal Report. `http://www.igcp565.org/workshops/Johannesburg_2011/kusche_LectureNotes_analysistools.pdf`.

Kusche, J., V. Klemann and W. Bosch (2012). *Mass distribution and mass transport in the Earth system.* Journal of Geodynamics, 59-60:1–8.

Lambert, Z.V., A.R. Wildt and R.M. Durand (1990). *Assessing sampling variation relative to number-of-factors criteria.* Educational and Psychological Measurement, 50 (1):33–48.

Landerer, F. W., and S. Swenson (2012). *Accuracy of scaled GRACE terrestrial water storage estimates.* Water Resources Research, 48:W04531.

Leblanc, M.J., P. Tregoning, G. Ramillien, S.O. Tweed and A. Fakes (2009). *Basin-scale, integrated observations of the early 21st century multiyear drought in southeast Australia.* Water Resources Research, 45 (4):W04408.

Lilliefors, H. (1967). *On the Kolmogorov-Smirnov test for normality with mean and variance unknown.* Journal of the American Statistical Association, 62:399–402.

Ljung, L. (1987). *System identification - Theory for the use.* Prentice Hall, Englewood Cliffs, N.J.

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:W11517.

Macrander, A., C. Böning, O. Boebel and J. Schröter (2010). *Validation of GRACE gravity fields by in-situ data of ocean bottom pressure.* In: Flechtner, F., T. Gruber, A. Güntner, M. Mandea, M. Rothacher, T. Schöne and J. W. Berlin (Hrsg.), *System Earth via Geodetic-Geophysical Space Techniques*, 169–185. Springer Berlin Heidelberg.

Mansour, A., and C. Jutten (1995). *Fourth-order criteria for blind sources separation.* Signal Processing, IEEE Transactions on, 43 (8):2022–2025.

Mayer-Gürr, T. (2008). *Gravitationsfeldbestimmung aus der Analyse kurzer Bahnbögen am Beispiel der Satellitenmissionen CHAMP und GRACE.* PhD Thesis, Universität Bonn, Deutschland.

Mayer-Gürr, T., A. Eicker, E. Kurtenbach and K-H. Ilk (2010). *ITG-GRACE: Global static and temporal gravity field models from GRACE data.* In: *System Earth via Geodetic-Geophysical Space Techniques*, Advanced Technologies in Earth Sciences, 159–168. Springer Berlin Heidelberg.

Mayer-Gürr, T., E. Kurtenbach and A. Eicker (2010). *ITG-Grace2010 gravity field model.* `http://www.igg.uni-bonn.de/apmg/index.php?id=itg-grace2010`.

Melachroinos, S.A., J-M. Lemoine, P. Tregoning and R. Biancale (2009). *Quantifying aliased S2 tidal errors from multiple space-geodesy techniques, GPS and GRACE, over North West Australia.* Journal of Geodesy, 83 (10):915–923.

Mendel, J.M. (1991). *Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications.* Proceedings of the IEEE, 79 (3):278–305.

Mestas-Nuñez, A.M. (2000). *Orthogonality properties of rotated empirical modes.* International Journal of Climatology, 20 (12):1509–1516.

Mohino, E., B. Rodríguez-Fonseca, C.R. Mechoso, S. Gervois, P. Ruti and F. Chauvin (2011). *Impacts of the tropical Pacific/Indian Oceans on the seasonal cycle of the West African monsoon.* Journal of Climate, 24:3878–3891.

Monahan, A.H., F.T. Tangang and W.W. Hsieh (1999). *A potential problem with extended EOF analysis of standing wave fields.* Atmosphere-Ocean, 37 (3):241–254.

Moreau, E. (2001). *A generalization of joint-diagonalization criteria for source separation.* Signal Processing, IEEE Transactions on, 49 (3):530–541.

Nicholson, S.E. (2000). *The nature of rainfall variability over Africa on time scales of decades to millenia.* Global and Planetary Change, 26 (1-3):137–158.

North, G.R., T.L. Bell and F.J. Moeng (1982). *Sampling errors in the estimation of empirical orthogonal functions.* Monthly Weather Review, 110:699–706.

NRC (1999). *Global environmental change: research pathways for the next decade.* National Research Council (NRC), Washington DC: The National Academies Press.

NRC (2002). *Review of USGCRP plan for a new science initiative on the global water cycle.* National Research Council (NRC), Washington DC: The National Academies Press.

Omondi, P., J.L. Awange, L.A. Ogallo, R.A. Okoola and E. Forootan (2012). *Decadal rainfall variability modes in observed rainfall records over East Africa and their relations to historical sea surface temperature changes.* Journal of Hydrology, 464-465:140–156.

Omondi, P., J.L. Awange, E. Forootan and Coauthers (2013a). *Changes in temperature and precipitation extremes over the Greater Horn of Africa region from 1961 to 2010.* International Journal of Climatology, 34:1262–1277.

Omondi, P., J.L. Awange, L.A. Ogallo, J. Ininda and E. Forootan (2013b). *The influence of low frequency sea surface temperature modes on delineated decadal rainfall zones in Eastern Africa region.* Advances in Water Resources, 54:161–180.

Owe, M., A.A. Van de Griend, R. deJeu, J.J. deVries, E. Seyhan and E.T. Engman (1999). *Estimating soil moisture from satellite microwave observations: Past and ongoing projects, and relevance to GCIP.* Journal of Geophysical Research: Atmosphere, 104(D16):19735–19742.

Plaut, G., and R. Vautard (1994). *Spells of low-frequency oscillations and weather regimes in the northern hemisphere.* Journal of the Atmospheric Sciences, 51:210–236.

Preisendorfer, R. (1988). *Principal component analysis in meteorology and oceanography.* Elsevier: Amsterdam.

Preisendorfer, R., F.W. Zwiers and T.P. Barnett (1981). *Foundations of principal component selection rules.* `http://gate1.baaqmd.gov/pdf/0543_Foundations_Principal_Component_Selection_Rules_University_California_San_Diego_1981.pdf`.

Rangelova, E., and M.G. Sideris (2008). *Contributions of terrestrial and GRACE data to the study of the secular geoid changes in North America.* Journal of Geodynamics, 46 (3-5):131–143.

Rangelova, E., W. van der Wal, M.G. Sideris and P. Wu (2010). *Spatiotemporal analysis of the GRACE-derived mass variations in North America by means of multi-channel singular spectrum analysis.* International Association of Geodesy Symposia, 135 (7):539–546.

Rasmusson, E.M., P.A. Arkin, W-Y. Chen and J.B. Jalickee (1981). *Biennial variations in surface temperature over the United States as revealed by singular decomposition.* Monthly Weather Review, 109:587–598.

Ray, R.D., and S.B. Luthcke (2006). *Tide model errors and GRACE gravimetry: towards a more realistic assessment.* Geophysical Journal International, 167 (3):1055–1059.

Ray, R.D., and R.M. Ponte (2003). *Barometric tides from ECMWF operational analyses.* Annales Geophysicae, 21 (8):1897–1910.

Ray, R.D., D.D. Rowlands and G.D. Egbert (2003). *Tidal models in a new era of satellite gravimetry.* Space Science Reviews, 108 (1-2):271–282.

Reager, J.T., and J.S. Famiglietti (2013). *Characteristic mega-basin water storage behavior using GRACE.* Water Resources Research, 49:3314–3329.

Richman, M.B. (1986). *Rotation of principal components.* Journal of Climatology, 6 (3):293–335.

Richman, M.B. (1987). *Rotation of principal components: A reply.* Journal of Climatology, 7 (5): 511–520.

Rieser, D., M. Kuhn, R. Pail, I.M. Anjasmara and J.L. Awange (2010). *Relation between GRACE-derived surface mass variations and precipitation over Australia.* Australian Journal of Earth Sciences, 57 (7)(7):887–900.

Rietbroek, R. (2014). *Retrieval of sea level and surface loading variations from geodetic observations and model simulations: an integration approach.* PhD Thesis, University of Bonn, Germany.

Rietbroek, R., P. LeGrand, B. Wouters, J.-M. Lemoine, G. Ramillien and C.W. Hughes (2006). *Comparison of in situ bottom pressure data with GRACE gravimetry in the Crozet-Kerguelen region.* Geophysical Research Letters, 33:L21601.

Rietbroek, R., S.E. Brunnabend, C. Dahle, F. Flechtner, J. Kusche, J. Schröter and R. Timmermann (2009). *Changes in total ocean mass derived from GRACE, GPS, and ocean modeling with weekly resolution.* Journal of Geophysical Research: Oceans, 114:C11004.

Rietbroek, R., S-E. Brunnabend, J. Kusche and J. Schröter (2012a). *Resolving sea level contributions by identifying fingerprints in time-variable gravity and altimetry.* Journal of Geodynamics, 59-60:72–81.

Rietbroek, R., M. Fritsche, S-E. Brunnabend, I. Daras, J. Kusche, J. Schröter, F. Flechtner and R. Dietrich (2012b). *Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data.* Journal of Geodynamics, 59-60:64–1.

Risbey, J.S., M.J. Pook, P.C. McIntosh, M.C. Wheeler and H.H. Hendon (2009). *On the remote drivers of rainfall variability in Australia.* Monthly Weather Review, 137:3233–3253.

Rodell, M., P.R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J.K. Entin, J.P. Walker, D. Lohmann and D. Toll (2004). *The global land data assimilation system.* American Meteorological Society, 381–394.

Rodell, M., J. Chen, H. Kato, J.S. Famigliette, J. Nigro and C.R. Wilson (2007). *Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE.* Hydrogeology Journal, 15:159–166.

Rubin, D.B. (1981). *The Bayesian bootstrap.* Annals of Statistics, 9:130–134.

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, 59-60:49–63.

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, 59-60:124–132.

Schmidt, R., F. Flechtner, U. Meyer, K.H. Neumayer, C. Dahle, R. König and J. Kusche (2008a). *Hydrological signals observed by the GRACE satellites.* Surveys in Geophysics, 29 (4-5):319–334.

Schmidt, R., S. Petrovic, A. Güntner, F. Barthelmes, J. Wünsch and J. Kusche (2008b). *Periodic components of water storage changes from GRACE and global hydrology models.* Journal of Geophysical Research: Solid Earth, 113:B08419.

Schrama, E.J.O., B. Wouters and D.A. Lavallée (2007). *Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations.* Journal of Geophysical Research: Solid Earth, 112:B08407.

Schumacher, M. (2012). *Assimilation of GRACE data into a global hydrological model using an ensemble Kalman filter.* Master Thesis, University of Bonn, Germany.

Schuol, J., and K.C. Abbaspour (2006). *Calibration and uncertainty issues of a hydrological model (SWAT) applied to West Africa.* Advances in Geosciences, 9:137–143.

Sharifi, M.A., E. Forootan, M. Nikkhoo, J.L. Awange and M. Najafi-Alamdari (2013). *A point-wise least squares spectral analysis (LSSA) of the Caspian Sea level fluctuations, using TOPEX/Poseidon and Jason-1 observations.* Advances in Space Research, 51 (5):858–873.

Shum, C.K., J.C. Ries and B.D. Tapley (1995). *The accuracy and applications of satellite altimetry.* Geophysical Journal International, 121 (2):321–336.

Smirnov, N.V. (1948). *Tables for estimating the goodness of fit of empirical distributions.* Annals of Mathematical Statistics, 19:279–281.

Sneeuw, N. (1994). *Global spherical harmonic analysis by least-squares and numerical quadrature methods in historical perspective.* Geophysical Journal International, 118 (3):707–716.

Sneeuw, N., J. Flury and R. Rummel (2004). *Science requirements on future missions and simulated mission scenarios.* Earth Moon and Planets, 94 (1-2):113–142.

Spada, G., V.R. Barletta, V. Klemann, R.E.M. Riva, Z. Martinec, P. Gasperini, B. Lund, D. Wolf, L.L.A. Vermeersen and M.A. King (2011). *A benchmark study for glacial isostatic adjustment codes.* Geophysical Journal International, 185:106–132.

Stephens, M.A. (1974). *EDF statistics for goodness of fit and some comparisons.* Journal of the American Statistical Association, 69 (347):730–737.

Stone, J.V. (2004). *Independent component analysis: a tutorial introduction.* MIT Press, London.

Swenson, S., and J.M. 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):2193.

Swenson, S., and J.M. Wahr (2006). *Post-processing removal of correlated errors in GRACE data.* Geophysical Research Letters, 33(8).

Swenson, S., and J.M. Wahr (2007). *Multi-sensor analysis of water storage variations of the Caspian Sea.* Geophysical Research Letters, 34:L16401.

Syed, T.H., J.S. Famigliette, J. Chen, M. Rodell, S.I. Seneviratne, P. Viterbo and C.R. Wilson (2005). *Total basin discharge for the Amazon and Mississippi River basin from GRACE and a land-atmosphere water balance.* Geophysical Research Letters, 32:L24404.

Tapley, B. D., S. Bettadpur, M. Watkins and C. Reigber (2004a). *The gravity recovery and climate experiment: mission overview and early results.* Geophysical Research Letters, 31:L09607.

Tapley, B.D., S. Bettadpur, J.C. Ries, P.F. Thompson and M.M. Watkins (2004b). *GRACE measurements of mass variability in the Earth system.* Science, 305:503–505.

Thiiébaux, H.J., and F.W. Zwiers (1984). *The interpretation and estimation of effective sample sizes.* Journal of Climate and Applied Meteorology, 23:800–811.

Thode, H.C. (2002). *Testing for normality.* CRC Press.

Tregoning, P., K. Lambeck and G. Ramillien (2008). *GRACE estimates of sea surface height anomalies in the Gulf of Carpentaria, Australia.* Earth and Planetary Science Letters, 271 (1-4):241–244.

Trenberth, K.E. (1984). *Some effects of finite sample size and persistence in meteorological statistics. part I: autocorrelations.* Monthly Weather Review, 112:2359–2368.

Uebbing, B., J. Kusche and E. Forootan (2014). *Waveform retracking for improving Topex/Poseidon, Jason-1 and -2 altimetry results over African lakes.* Geoscience and Remote Sensing, IEEE Transactions On, accepted.

Ummenhofer, C.C., A. Gupta, P.R. Briggs and Coauthors (2011). *Indian and Pacific Ocean influences on Southeast Australian drought and soil moisture.* Journal of Climate, 24:1313–1336.

van der Veen, A-J. (2001). *Joint diagonalization via subspace fitting techniques.* In: *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, Vol. 5, 2773–2776.

van Dijk, A.I.J.M. (2011). *Model-data fusion: using observations to understand and reduce uncertainty in hydrological models.* In: *19th International Congress on Modelling and Simulation, Perth, Australia, 12-16 December 2011.*

van Dijk, A.I.J.M., 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:W11524.

van Dijk, A.I.J.M., L.J. Renzullo, Y. Wada and P. Tregoning (2013). *A global water cycle reanalysis (2013-2012) reconciling satellite gravimetry and altimetry observations with a hydrological model ensemble.* Hydrol. Earth Syst. Sci. Discuss., 10:15475–15523.

Velicogna, I., and J.M. Wahr (2005). *Greenland mass balance from GRACE.* Geophysical Research Letters, 32:L18505.

Vinogradova, N.T., R.M. Ponte, M.E. Tamisiea, K.J. Quinn, E.M. Hill and J.L. Davis (2011). *Self-attraction and loading effects on ocean mass redistribution at monthly and longer time scales.* Journal of Geophysical Research: Oceans, 116:C08041.

von Storch, H., and A. Navarra (1999). *Analysis of climate variability.* Springer, New York.

von Storch, H., and F. Zwiers (1999). *Statistical analysis in climate research.* Cambridge University Press, Cambridge.

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.

Wahr, J.M., 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, 108(B12):30205–30229.

Wahr, J.M., S. Swenson and I. Velicogna (2006). *Accuracy of GRACE mass estimates.* Geophysical Research Letters, 33 (6):L06401.

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.

Weare, B.C., and J.N. Nasstrom (1982). *Examples of extended empirical orthogonal function analyses.* Monthly Weather Review, 110:784–812.

Werth, S., and A. Güntner (2010). *Calibration analysis for water storage variability of the global hydrological model WGHM.* Hydrology and Earth System Science, 14:59–78.

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.

Westra, S., C. Brown, U. Lall and A. Sharma (2007). *Modeling multivariable hydrological series: Principal component analysis or independent component analysis?* Water Resources Research, 43 (6):W06429.

Wilkinson, J. (1965). *The algebraic eigenvalue problem.* Oxford University Press.

Wouters, B., and E.J.O. Schrama (2007). *Improved accuracy of GRACE gravity solutions through empirical orthogonal function filtering of spherical harmonics.* Geophysical Research Letters, 34 (23):L23711.