ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE ENGINEERING AND TECHNOLOGY

ANALYSIS OF GRACE L2 DATA GLOBALLY AND REGIONALLY WITH PCA

M.Sc. THESIS

Ceren ALTUNAL PODLECH

Department of Geomatics Engineering

Geomatics Engineering Programme

MAY 2015

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL OF SCIENCE
ENGINEERING AND TECHNOLOGY

ANALYSIS OF GRACE L2 DATA GLOBALLY AND REGIONALLY WITH PCA

M.Sc. THESIS

Ceren ALTUNAL PODLECH
(501111628)

Department of Geomatics Engineering

Geomatics Engineering Programme

Thesis Advisor: Assoc. Prof Dr. H. Hakan Denli

MAY 2015

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

GLOBAL VE BÖLGESEL OLARAK GRACE L2 VERİLERİNİN PCA İLE ANALİZİ

YÜKSEK LİSANS TEZİ

Ceren ALTUNAL PODLECH
(501111628)

Geomatik Mühendisliği Anabilim Dalı

Geomatik Mühendisliği Programı

Tez Danışmanı: Assoc. Prof Dr. H. Hakan Denli

MAYIS 2015

Ceren ALTUNAL PODLECH, a M.Sc. student of ITU Graduate School of Science Engineering and Technology 501111628 successfully defended the thesis entitled **"ANALYSIS OF GRACE L2 DATA GLOBALLY AND REGIONALLY WITH PCA"**, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof Dr. H. Hakan Denli** Istanbul Technical University	....
Jury Members :	**Assoc. Prof Dr. Serdar EROL** Istanbul Technical University	....
	Assoc. Prof Dr. Cüneyt AYDIN Yıldız Technical University	....
		....

Date of Submission : **4 May 2015**
Date of Defense : **May 2015**

To my familly,

FOREWORD

I am thankful to my thesis supervisor, Assoc. Prof Dr. H. Hakan Denli who advised my during Master studies. I was fortunate to benefit from his knowledge and experiences. His supportive attitude motivated me to force myself all the time to achieve a better state.

I would like to thank to Prof. Dr. Frank Flechtner, Dr.-Ing Christian Gruber and GFZ (Helmholtz-Zentrum Potsdam - Deutsches GeoForschungs Zentrum) for all their support and the supply of data to my Master thesis study.

Besides, specials thanks to my parents, my sister and my husband, Paul.

MAY 2015

Ceren ALTUNAL PODLECH
Geomatics Engineer

TABLE OF CONTENTS

Page

ABBREVIATIONS

CS2	: GFZ R05 data with 250 *km* spatial resolution
CS3	: GFZ R05 data with 350 *km* spatial resolution
CS6	: GFZ R04 data with 500 *km* spatial resolution
DLR	: Deutsches Zentrum für Luft- und Raumfahrt
EOF	: Empirical Orthogonal Function
GFZ	: Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum
GRACE	: Gravity Recovery and Climate Experiment
JPL	: Jet Propulsion Laboratories, Pasadena
L2	: Level 2
PC	: Principal Component
PCA	: Principal Component Analysis
RL04	: Release 04
RL05	: Release 05
SLR	: Satellite Laser Ranging
TWS	: Total Water Storage
UTCSR	: University of Texas at Austin, Center for Space Research
WGHM	: WaterGap Hydrological Model

LIST OF TABLES

Page

LIST OF FIGURES

Page

ANALYSIS OF GRACE L2 DATA GLOBALLY AND REGIONALLY WITH PCA

SUMMARY

The satellite gravity mission Gravity Recovery and Climate Experiment (GRACE) provides a huge amount of data which allows to quantify temporal and spatial variations of the Earth's gravity field caused by mass transport and mass distribution.

However generating and using GRACE-derived estimates of total water storage (TWS) for validation purposes is a complicated task due to the nature of the GRACE Level 2 (L2) products and their spatio-temporal resolution limitations. This results in flawed data-sets. For instance, the GRACE L2 results contain stripe shaped correlated noise, artifacts oriented from north to south, and ring shaped ones about centers of large masses. Besides, spherical harmonic coefficients are dominated by noise at high degrees. Therefore, spherical harmonic expansion has to be cut-off at an optimum maximum degree/order with keeping signal-loss as little as possible. This truncation of the input signal causes amplitude dampening and signal leakage outside of the area of interest. Also missing months of GRACE as a result of battery management causes gaps in the time-series which affect results of the analyzes.

Within this study, at first post-processing steps of GRACE data are studied to have optimum condition of our data-set. The contaminating effect of the noise is demonstrated and filtering techniques as a solution for the correlated noise are introduced. Then the performance of the Gauss filter to remove noise is assessed. Gaussian filtering removes random noise, but leaves the correlated noise intact. Following that, truncation and leakage problems for Amazon, Greenland, Mekong and Tigris/Euphrates regions are observed. A test data set, obtained by a simulation of continental hydrology, and GRACE-like example data derived from it, both supplied by Helmholtz-Zentrum Potsdam-Deutsches GeoForschungsZentrum (GFZ), are used to test truncation in the spectral domain. Since the GRACE-like data is acquired from simulated continental hydrology data, the latter serves as a reference data (true signal). Because that way the true signal is available, the GRACE-like data can be used to recover the restricted spatial resolution of GRACE data and allows making the analysis in the spectral domain without any signal loss. For this analysis the spherical harmonic synthesis in equivalent water height up to 50 and 90 degree and order is executed in the spectral domain. The results are compared with the reference signal for the regions Mekong, Amazon, Greenland, Turkey and Tigris/Euphrates. Besides, masks that minimize signal-loss are enlarged for the regions Amazon, Greenland and Tigris/Euphrates. Enlarging the region mask covers more signal, however it can not be guaranteed, that the recovered signal belongs to the region of interest.

As a second step, the statistical signal-signal decomposition technique Principle Component Analysis (PCA) is applied to the data-sets to reduce dimensionality and determine the direction of the highest variance of the data. PCA solutions of three

data-sets (CS2, CS3 and CS6) obtained by applying different post-processing techniques are compared each other. In the first mode, the results showed identical trend in each region. The CS6 data-set resulted in smoother signal than the newer releases in all modes and all regions.

The effect of gaps in GRACE time-series (due to battery management) is analyzed by adding artificial gaps to GRACE data between 2003-2010, which did not previously have any gaps. Comparing results yields that missing monthly data caused 1.54 % signal loss in Greenland, 0.71 % in Amazon and 0.22 % in Mekong. Apparently and to no surprise, the signal loss is bigger the larger the area is.

The performance of the conventional Principle Component Analysis (PCA) is investigated by applying it to a monthly series of Gauss filtered (400 *km* radius) GRACE and WaterGAP Global Hydrology Model (WGHM) data for the years 2005 to 2008. The results are compared globally and for Amazon, Orinoco, Turkey and Tigris/Euphrates regionally. Besides the signal's reconstruction is accomplished for desired variances, 80 % and 95 %, globally and regionally. The required PCA mode numbers with respect to the regions of interest are determined. As it turns out, this depends on the region. For 95 % variance globally mode 8, in Amazon mode 3 and in Orinoco mode 1 are required. WGHM and GRACE data showed the same trend for each region, however the amplitude of WGHM was lower in every case.

The PCA results of all 3 previously mentioned GRACE data-sets provided by GFZ are also validated with in-situ tide gauges over the Mekong basin between 1^{st} of July and 30^{th} of November for the years 2008, 2009, 2010 and 2011. The results depict that consistency of GRACE and in-situ data change depending on the corresponding year.

GLOBAL VE BÖLGESEL OLARAK GRACE L2 VERİLERİNİN PCA İLE ANALİZİ

ÖZET

GRACE uydu sistemi (Gravitiy Recovery and Climate Experiment) yeryuvarı sistemindeki kütle transferi ve dağılımından kaynaklanan gravite alanının zamansal ve uzaysal değişimlerinin değerlendirilmesine olanak sağlayan veriyi elde etmektedir.

GRACE verilerinden yeryuvarı sistemindeki su kaynaklarının değişimlerini elde etmek ve bu sonuçları, doğrulama amacı ile kullanmak karmaşık bir işlemdir. Bu karmaşıklık GRACE seviye 2 (Level 2, L-2) ürünlerinin doğasından, örnek vermek gerekirse, ürünlerin uzaysal ve zamansal çözünürlüklerinin kısıtlı olmasından ileri gelir. Bunun yanı sıra, lokal havza gel-git ölçmeleri ve hidroloji model verilerinin bu süreçte değerlendirilmesi de oldukça dikkat gerektirir.

GRACE sistemindeki yapısındaki bu kısıtlılıklar, L-2 ürünlerinde hata olarak ortaya çıkmaktadır. Kuzey-güney doğrultusunda şerit şeklinde ve büyük kütleler etrafında ise halkalar halinde sistematik hatalar olarak görülür.

Bunun yanı sıra, yüksek derece ve sıralarda bulunan fazla gürültü sebebi ile girdi sinyalinin işlenmesinin belirli bir derece ve sırada kesilmesinden dolayı, sinyal genliğinde sönülüm ve sinyalin ilgili bölgenin etrafına yayılması gibi bir problem açığa çıkar.

Ayrıca GRACE uydularının 2011 yılından itibaren ortaya çıkan batarya problemi sebebi ile oluşan zaman serilerindeki boşluklar analiz sonuçlarını olumsuz etkiler.

Bu tez çalışması sürecinde, bahsedilen GRACE veri setinde mevcut bulunan olumsuzlukları minimuma indirgemek amaçlanmıştır. İlk olarak veri işleme yöntemleri incelenerek verinin analiz için optimum kuşullara getirilme yolları incelendi. Gauss filitreleme tekniğini ile verideki gürültüyü yok etme performansı test edildi. Sonuçlar incelendiğinde, korelasyonlu gürültülerin bu yöntemle giderilemediği görüldü.

Sonraki aşamada, sinyalin kesintisinden ötürü oluşan sızıntı problemi Amazon, Grönland, Mekong ve Fırat/Dicle nehirleri bölgeleri için incelendi. Çalışmanın bu aşamasında GFZ (Helmholtz-Centre Potsdam-GFZ German Research Centre for Geosciences) tarafından sağlanan test verileri kullanılmıştır. Bu test verileri kıtasal hidroloji simulasyonu (gerçek sinyal) ve bu veriden elde türetilmiş yapay-GRACE verisinden oluşur.

Yapay GRACE verisi test verisinden türetildiğinden, test veri seti referans verisi kabul edilir ve iki veri seti kendi aralarında karşılaştırılabilir. Bu bize sinyal kesintisine bağlı sinyal kaybının başka etmenleri hesaba katmadan hesaplama imkanı sağlar. Dolayısı ile bu aşamada, verilerin mekansal alandaki derece/sıra limitlemesinin GRACE analiz sonuçlarına etkileri incelenebilir.

Araştırma küresel harmonic katsayı setleri maksimum derece/sıra 90 ve 50 için hesaplanmıştır. Amazon, Mekong ve Grönland bölgelerinin global ölçekten

tanımlanabilmesi için, alan sınırlarını baz alarak oluşturulmuş havza fonksiyonları (maskeler) kullanılmıştır. Sentetik GRACE verisi referans veri ile karşılaştırıldığında görülmüştür ki, sinyal kaybının etkisi, incelenen bölge alanı küçüldükce artmaktadır. Mekong bölgesi icin 90 dereceki katsayı kesintisinin % 80 sinyal kaybına yol açtığı gözlenmiştir. Aynı kesinti Amazon bölgesinde % 8.7 oranına düşmüştür.

Çalışmanın ikinci aşamasında, statistiki sinyal-sinyal ayrıştırma yöntemi Temel Bileşen Analizi (PCA) $1° \times 1°$ grid GRACE L-2 (Release-04, RL-04) 500 *km*, (Release-05, RL-04) 250 *km* ve (Release-05, RL-05) 350 *km* mekansal çözünürlüklü aylık çözümlemelerine uygulanarak, sinyaller bileşenlerine ayrılmıştır. Bu analiz kullanılan veri setlerinin en yüksek varyans ile dagıldığı dogrultuyu ve bu doğrultudaki veri gurubunu ortaya koymuştur. Ayrıca, veri setindeki gürültüyü ayrıştırmıştır. Kullanılan her üç veri de 2003-2012 periyodunu kapsamaktadır. Bu analiz bölgesel olarak gerçekleştirilmiştir. Analiz sonucunda faklı karakteristliğe sahip bu bölgeler için sonuçlar karşılaştırılıp, farklı versiyon GRACE verilerin performansları incelenmiştir.

Bölgelerin tanımlanmasında iki farklı prensibe dazanan metodlar ile gerçekleştirilmiştir. İlk olarak, Amazon, Mekong, Grönland, Türkiye ve Dicle/Fırat nehirlerini bölge verilerine sınırlarını baz alan maskeler aplike edilmiş, PCA uygulanmıştır.

Ardından çalışma sırasında tanımlanmış olan sinyal seviyelerini baz alarak tanımlanan Amazon ve Grönland maskeleri aplike edilmiştir. Her iki yaklaşımın PCA sonuçları kendi aralarında karşılaştırılmıştır. Tasarlanmış bu maskeler, sinyal değerlerinin belirli bir değerde limitlendirilmesi ile logical matrix olarak elde edilmiştir. Farklı sinyal seviyelerine göre üretilmiş maskelerle yapılar testler göstermiştir ki, küresel harmonic sayı setlerinin 50 derece/sıraya kadar geliştirip, 0.15 de limitlendirilmesi her iki bölge için de en iyi sonucu üretmekte, yakalanan sinyal oranını arttırmaktadır.

Tasarlanan bu maskelerin PCA sonuçlarına etkilerine gelince, tasarlanan maskelerin sızmış sinyali daha iyi yakalayabilmesi neticesinde ayrışan sinyallerin genliklerinde sistematik bir artış gözlenmiştir. Bu sonuç, genlik sönülümdeki iyileştirme olarak yorumlanabilir. Ayrıca kordinat bazlı maskeler manuel olarak genişletilmiş ve performansları test edilmiştir. Dicle & Fırat bölgesinde bölgenin dışından bölgenin içine sinyal sızdığı gözlenmiştir.

2011 yılından itibaren, GRACE sistemde baş gösteren batarya problemi sebebiyle, veri kaybı yasanmaktadır. Kullanılan GRACE verilerinin PCA sonuçlarında da görüldüğü üzere, zaman serilerinde aylık boşluklar bulunmaktadır. Bu boşluklar çalışma için temin edilen verilerde "0" ile doldurulmuştur. Zaman serilerindeki bu boşlukların PCA sonuclarına etkisinin boyutlarını belirlemek amacı ile, verilerin boşluk içermeyen zaman dilimini (2003-2010) kullanarak, yapay boşluklar oluşturup, bu boşluklar "0" ile doldurulmuşur.

Bu boşlukların orijinal veri ile aynı aylarda (batarya tasarufu içeren aylarda-Ocak 2009, Haziran 2009, Mayıs 2010 ve Ekim 2010) oluşturulmasına dikkat edilmiştir. Elde edilen bu yeni veriye Amazon, Mekong ve Grönland bölgeleri sınır bazlı maskeler uygulanılarak, bölgeler dahilinde PCA uygulanmıştır. Sonuçlar göstermiştir ki, bu yapay boşluklar Mekong, Amazon ve Grönland bölgeleri sırası ile % 0.22, %0.71 ve % 1.54 oranında sinyal kaybına uğramıştır. Alan küçüldükçe, sinyal kayıp oranı da azalmıştır.

GRACE sistemi yüksek doğrulukta, global ölçekte gravite degişimi veri setlerini ortaya koyar. Bu verileri kullanılarak global ve ya bölgesel olarak toplam su değişimi elde edilir. Bu veri setleri, hidroloji modellerinin kalibre edilmesi amacı ile yaygın olarak kullanılmaktadır. Bizim calışmamızın kapsamında Geleneksel Temel Bileşen Analizi (Princible Component Analysis-PCA) kullanılarak GRACE ve WaterGAP Global Hidroloji modeli (WaterGAP Global Hydrology Model) verileri global ve bölgesel olarak karşılaştırılmıştır.

400 *km* yarıçaplı Gauss filtreleme tekniği uygulanmış GRACE L-2 ve WGHM bu metodla sinyal bileşenlerine ayrılmıştır. Her iki veri de 2005-2008 periyodunu kapsamaktadır. Analiz sonuçları global ve bölgesel ölçekte karşılaştırılmıştır. Bölgesel analiz için çalışma alanı Amazon, Orinoco, Türkiye ve Dicle/Fırat havzaları seçilmiştir. Bu analiz sonuçlarından faydalanılarak, WGHM sinyalinin Global ve bölgesel olarak % 95 ve % 80 oranında orijinal veri ile uyuşacak şekilde tekar işa edilmiştir.

Bu işlem bölgesel olarak da tekrarlanmıştır. % 95 sinyal varyasyonuna ulaşmak global ölçekte analizin 8 moduna inmeyi gerektirir iken, Amazon ve Orinoco havzaları için bu sayı sırası ile 3 ve 1'e inmiştir. % 80 oranını hedeflediğimiz de ise, mod sayısı global ölçekte 3'e düşmüştür

Çalışmanın son aşamasında, GRACE verilerine referans veri olarak, GFZ tarafından temin edilmiş, Mekong deltasına ait olan yerel gel-git ölçmeleri kullanılmıştır. Temin ediler veriler 2008, 2009, 2010 ve 2011 yılları için 1 Temmuz-30 Kasım periyodunu kapsamaktadır. Sınır bazlı Mekong maskesi her üç farklı çözünürlüklü GRACE verisine uygulanmış ve sinyal PCA kullanılarak bileşenlerine ayrılmıştır. Elde edilen ilk mod sonuçları referans ölçmelerle karşılaştırılmıştır.

Sonuçlar değerlendirilir iken, hidroloji verileri ile GRACE verileri arasında yaklaşık bir aylık sistematik bir kayma olabileceği dikkate alınmıştır. Sonuçlar arasındaki en düşük korelasyon (% 0.95) GFZ RL-04 Level-2 gözlenerek, GFZ GRACE verilerindeki gelişme de ortaya konmuştur. GFZ RL-05 Level-2 verilerileri oldukça benzer sonuçlar göstererek, % 0.99 oranında ölçme verilerine uyum göstermektedir.

1. INTRODUCTION

Since 2002, the Gravity Recovery and Climate Experiment (GRACE) satellite mission monitors the Earth's gravity field. Given the extraordinary accuracy of GRACE, mass distributions induced by geophysical and hydro-climatological mechanisms above, at and below the Earth's surface can be cleared up by time-variable gravity signals of GRACE. From the absence of a global comprehensive hydrology monitoring system and the limitations in quantifying water balance components by observations and modeling, GRACE for the field of continental hydrology is ground-breaking. GRACE is an unprecedented opportunity to observe hydrological cycles, as for the first time ever, water storage changes for large continental areas are monitored with GRACE [3].

1.1 Data Description

The Level-2 data products include the static and time-variable (monthly) gravity field and related data products derived from the application of Level-2 processing at GFZ, UTCSR and JPL to the previous level data products (the Level-0, the Level-1A, the Level-1B). This level also includes the ancillary data products such as GFZ's Level-1B short-term atmosphere and ocean de-aliasing product (AOD1B)generated during the processing [4].

During this master thesis study, variety of data-sets are tested globally and regionally. These data-sets are supplied from Helmholtz-Zentrum Potsdam - Deutsches GeoForschungsZentrum (GFZ). To be able to follow the complete analysis, it is very important to understand the purpose of each data sets and in which part of the study are they used. Because of differences in data resolution, time-span and filtering technique of every single data, they will be introduced with a brief explanation of their use during analysis.

These data set can be itemized as followed:

- GRACE Level-2 data over March and September 2010 with a resolution of $1^\circ \times 1^\circ$ which is post-processed by Bonn University. the differences of these data-sets are used to demonstrate gravity function and test the Gaussian filter in chapter 2.1 for each analysis.

- A test data set, obtained by a simulation of continental hydrology on $0.5^\circ \times 0.5^\circ$, and GRACE-like example data derived from it. The spherical harmonic coefficients of the GRACE-like data are developed up to 90 degree/order. This simulation data-set is created for test purposes by GFZ, it doesn't belong to a time span. These data-sets are used for demonstration of signal loss and leakage for Tigris, Mekong, Greenland and Turkey. This analysis is also repeated with the coefficients developed up to 50 degree/order to see the effect of truncation more significantly. Analysis results and details are explained in section 4.1. Besides, defined masks are tested with using these two data-sets to assess the quality of the mask regarding to signal lost performance in section 4.1.
- Spectrally analyzed spherical harmonical coefficients that are developed up to 90 degree/order. This data-set is used to define threshold-base masks for Greenland and Amazon region in section 4.1.
- Monthly series of 400 *km* radius Gauss filtered (explained in section 2.1) GRACE and WaterGAP Global Hydrology Model (WGHM) data for the years 2005 to 2008. It is used in section 4.5 for compassion of two data-sets regionally and globally with applying PCA. For the regional study, Amazon, Orinoco, Turkey and Tigris-Euphrates rivers are chosen as region of interest. All of the inspected regions are selected from rest of the glob with using their border coordinates.
- In-situ tide gauges over the Mekong basin between 1^{st} of July and 30^{th} of November for the years 2008, 2009, 2010 and 2011 are used to compare with PCA results of Mekong with CS2, CS3 and CS6 data-sets in section 4.4.

During the study, different versions of GRACE data-sets are tested for the regional studies in section 4.2. As study area Amazon, Greenland, Mekong, Tigris, Turkey are chosen. Results for each data-sets are compared for each region. Besides, for analyzing the effect of gaps in the timeseries, these data-sets are also used in section4.3. Each version is named as it will be explained here and they will be stated in analysis as they are defined. These 3 version of GRACE data-sets can be itemized as followed:

- CS6-The GRACE Level-2 GFZ Release 04 Kusche 2 filtered and has 500 *km* spatial resolution for the span 2003 through 2012.

- CS2 and CS3 are GRACE Level-2 GFZ Release 05 data (L-2, RL-05), ("GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005" published as GFZ Scientific Technical Report which can be checked for detailed information about data [5]) for the span 2003 through 2012.
- CS3 is Kusche-3 filtered and currently has approximately 350 *km* resolution.
- CS2 is in principle unfiltered. However, it was computed with the aid of regularization, where the regularization matrix is computed with a Kalman filter and it has been driven by the hydrology model WGHM. For CS2, effective resolution is the bandwidth of the spherical harmonic expansion (degree 80) which corresponds to 250 *km* spatial resolution. This data can not be referenced currently, it is about to be published during the next months. To have more detail information about this data, please contact to Dr.-Ing. Christian Gruber (gruber@gfz-potsdam.de).

Finally, to reference to the start point of my master thesis, the summer school which was held at the Institute of Geodesy and Geoinformation (IGG), Bonn University in 2011 was inspiring for the study. All material can be access via this link for free of charge (ftp://www.dgfi.badw.de/pub/SPP1257/compact/).

1.2 Purpose of Thesis

There exist several regions all across the globe having critical situations regarding the Total Water Storage (TWS) change. Therefore GRACE is a very powerful tool to obverse and evaluate these changes in any region of choice, even if this TWS data is inaccessible in other ways. However evaluating GRACE data can be challenging. These measurements contain correlated noise originating from the orbit of the satellites and random noise, mostly at the higher degree and order of spherical harmonics. Within the study Gaussian filtering is introduced as a solution to noise in the data-set. To minimize the noise at higher degrees/order an optimum maximum degree/order is obtained and signal loss regarding to this truncation is calculated. This can be observed as signal leakage in GRACE data around the area of interest. Therefore it is necessary to define adjusted region masks for regional analyzes as done in section 4.1.
Differently post-processing approaches can modify the analyzes results. Hence the impact of the choice of input data to PCA results are investigated in the regions Turkey, Tigris/Euphrates, Amazon, Greenland and Mekong over 120 months.

Unfortunately there is gaps in GRACE measurements (due to battery management). To investigate the impact of gaps in time-series data to PCA results, artificial gaps are injected to the years between 2003 to 2010, which previously did not have any gaps, and the results are compared.
To test the validity of PCA results, they are compared to in-situ measurement data for Mekong area.
As GRACE provides a record of variations in total terrestrial water storage (defined as all of the snow, surface water, soil moisture, and groundwater) across the globe [6], it can be used to refine and/or verify hydrological models to improve their accuracy. Thus PCA results from GRACE data and WaterGAP Hydrology Model are compared in global and regional scale, whereas regional refers to the areas Amazon, Orinoco, Turkey and Tigris/Euphrates.

2. GRACE OBSERVATIONS of MASS CHANGES

In this chapter, firstly, the concept of the Gravity Recovery And Climate Experiment (GRACE) mission is introduced. Proper estimation of total water storage (TWS) fields from GRACE time-variable gravity products is addressed (Section 2.1).

In section 2.1, the processing steps, to estimate TWS changes from GRACE time-variable products, is presented in section 2.1.2. Therefore, discussions on data preparation (subsections 2.1.1 and 2.1.2) are necessary to understand the investigations that are performed in the application part (chapter 4).

The GRACE (gravity recovery and climate experiment) mission is a joint project between the National Aeronautics and Space Administration (NASA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The mission has been proposed in 1996 jointly by the University of Texas at Austin, Center for Space Research (UTCSR), the German Research Centre for Geosciences (GFZ) and the Jet Propulsion Laboratories (JPL) in Pasadena. GRACE was selected in 1997 as second mission in NASA's Earth System Science Pathfinder (ESSP) program. As an innovation, the Principal Investigator Prof. Byron Tapley (UTCSR) and his team are ultimately responsible for developing the flight mission hardware from selection to a launch-ready condition, for accomplishing the scientific objectives and delivering the proposed measurements to the broader Earth science community and general public as expediently as possible. Project management and systems engineering activities are carried out by JPL [7].

(a) View of twin satellites in space (b) View of twin satellites in space center

Figure 2.1: Satellite system of GRACE mission

The primary science objective of the GRACE mission is to measure the Earth's gravity field and it's time variability with unprecedented accuracy. The increase in accuracy has been achieved by utilizing two satellites following each other on the same orbital track. To consider precise altitude and non-gravitational forces both satellites are equipped with star cameras and accelerometers. The position and velocity of the satellites is measured using onboard GPS antenna and (for validation purposes) Satellite Laser Ranging (SLR) retro-reflectors [7].

Additionally, the twin satellites are interconnected by a K-band microwave link to measure the exact separation distance and its rate of change to an accuracy of better than 0.1 μm/s. These measurements are directly coupled to the seasonal and sub-seasonal variations in the continental hydrological cycle, to ice mass loss in the large glacier systems in Greenland or Antarctica, to long wavelength ocean circulation processes or to the transport of ocean heat to Earth's poles. Additionally, GRACE provides the time variability of the Earth's overall physical shape, the geoid. Consequently, since its launch in March 2002 this fundamental data set has enabled dramatic improvements of seasonal and inter-annual climate change estimates [7].

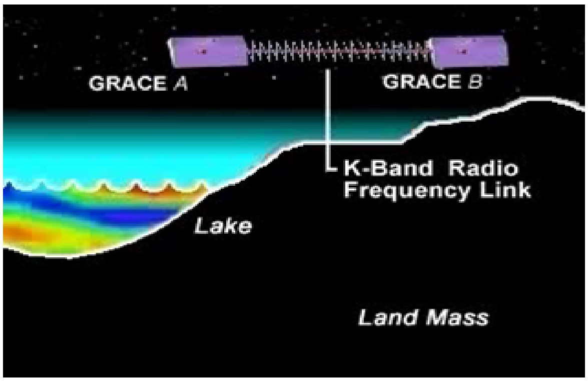


Figure 2.2: Measurement principle of GRACE

The secondary science objective of the GRACE mission is to obtain about 150 very precise globally distributed vertical temperature and humidity profiles of the atmosphere per day using the GPS radio occupation technique [7].

Because of solar activity, instrument conditions and battery health, the lifetime of the mission is restricted. Today, it has been already 12 years. Since many years strong request by the community to launch a GRACE Follow-on mission as fast as possible to extend the mass flux time series and to keep the gap as short as possible. GRACE-FO is again built under US-German partnership (NASA/GFZ MOU) and entered Phase-B in September 2012 with a target launch date in August 2017 [8].

2.1 Estimate TWS changes from Geopotential Coefficients

At product level 2, GRACE data are condensed to monthly sets of fully normalized spherical harmonic coefficients. These coefficients are the outcome of a rather complex data processing scheme. For scientific analysis, users of GRACE data will have to perform a number of basic operations on the GRACE coefficients: transform dimensionless geopotential coefficients into gridded geoid heights or maps of surface density expressed through equivalent water heights, filtering to reduce noise, which is explained below, and average such maps over the surface of some hydrological catchment area or ocean basin [9].

In this part of the study, level 2 GRACE data, the coefficients of the monthly solution, for the time slot 2008-2010 March and 2008-2010 September are used. Firstly, difference between the coefficient of the monthly solutions over 6 months is derived. Next, the result is filtered with using Gaussian smoothing method. Filtering method is applied multiple time for different half-width parameter value. Water Storage Change is calculated from non-filtered and filtered data, and the results are compared and evaluated.

Using the Gauss filtered monthly solutions difference which covers 6 months (500 km half-width), total water storage and gravity disturbances over are produced as gridded data on the Earth's surface.

2.1.1 Filtering

One of the problems that users of the GRACE level-2 products face is the presence of increasing correlated noise stripes at higher frequencies [10]. To filter out the spurious north-south stripes in the GRACE TWS products different techniques base on empirical, statistical, inverse, or regional approaches are in use [11].

Smoothing operators can be applied in either the spatial or spectral domain in order to suppress the effect of noise in maps and area averages.

2.1.1.1 Gaussian Filter

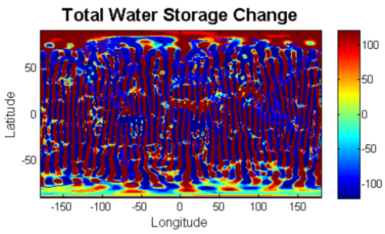
The simple isotropic Gaussian filer proposed earlier by Jekeli in 1981 and It was used in pre-launch studies. Later on, it was applied to the real data to demonstrate the abilities of GRACE to retrieve time variations of water mass anomalies over land [11].

Since than, at global and basin scales, Gaussian smoothing have been commonly used to suppress the noise which dominates high-degree and high-order Stokes coefficients in GRACE fields. Gaussian smoothing gives equal weight to all orders of Stokes coefficients at each degree, equivalent to convolution with a circular Gaussian shaped filter. Major drawbacks of of this filter are choice of smoothing radius and inability to distinguish between geophysical signals and noise [11].

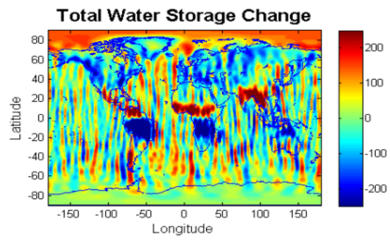
Gaussian smoothing is effective when proper radius of the smoothing in case the spatial distribution of noise is isotropic. This radius corresponds to the distance at which the weight drops to half its peak value at the shortest wavelength. Since the operation corresponds to convolution over the Earth's surface with a circularly symmetric function, which is not isotropic, GRACE noise does not fully meet this assumption [12].

In the previous study, geographical distribution of for 3, 6 and 12 months periods, like April/May 2002 minus April/May 2003 differences in GRACE gravity field solutions and WGHM continental hydrology model predictions are smoothed with Gauss type filter. It is pointed that, Gauss type filter with 750 *km* half wavelength (as a rule of thumb to a spherical harmonic degree of l=27) results nearly no high frequency signal remained in the filtered coefficients [13].

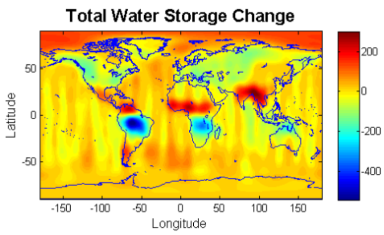
To suppress the errors contributed by higher degree spherical harmonic coefficients in GRACE solution, Gaussian filter is exerted [13]. The difference between coefficients over 2010 March-September used and total water storage change calculated according to common approximation. Coefficients differences filtered with Gauss filter averaging radius 250 *km*, 500 *km* and 1000 *km* and total water storage is derived for each. The results are illustrated in figure 2.3.



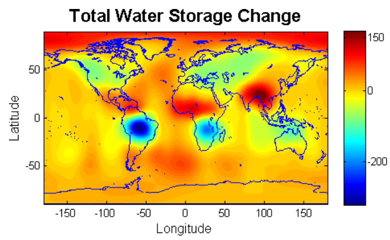
(a) Unfiltered water storage change (cm), (0 km half-width) over 2010 March-September.



(b) Gaussian filtered water storage change (cm) with 250 km half-width over 2010 March-September.



(c) Gaussian filtered water storage change (cm) with 500 km half-width over 2010 March-September.



(d) Gaussian filtered water storage change (cm) with 1000 km half-width over 2010 March-September.

Figure 2.3: Effect of Gaussian filtering and role of smoothing radius in water storage change which is calculated from GRACE spherical harmonics coefficients differences are demonstrated. It should be noted that Figure (a), Figure (b) are scaled 10 times greater, in mm unit, than Figure (c) and Figure (d), in *cm* unit, and all result computed up to maximum degree 60.

Figure 2.3 (a), 2.3 (b), 2.3 (c) and 2.3 (d) illustrate unsmoothed and with Gaussian filters of different half-widths GRACE-derived maps of monthly anomaly of total water storage (TWS) change. The results exhibit that, increasing the filtering radius causes the amplitude of the stripes to decrease. However, despite of the supplied improvement, filtered data still appear to have unphysical anomalies, dominated by stripes [12].

Two main limitations of Gaussian smoothing can be summarized:

- As effective radius increases, there is increased leakage error associated with a limited range of spherical harmonics.
- It assigns isotropic weights in the spatial domain respectively only degree-dependent weights in spherical harmonics domain.

Since Gaussian filter is degree dependent (isotropic), it is inadequate to remove non-isotropic noises. Remainders results as longitudinal stripes mostly due to coefficients related to the polar orbit ground tracks with a concentration at high orders. Besides, some studies show that these smoothing can affect water storage estimates.

For instance, a previous study, which is done by J. L. Chen et al. in 2006, demonstrates that Gaussian smoothing significantly affects basin-scale water storage estimates, even for the largest river basins. For example, with 1000 *km* Gaussian smoothing, seasonal magnitudes of GRACE water storage estimates in the Amazon and Mississippi are reduced to about 35 %, and the phase of the annual change is also affected [14].

2.1.2 Spherical Harmonics Synthesis and Gravity Functions

The Earth's global gravity field is commonly described in terms of the shape of the geoid and it is usual to expand the geoid undulations, N, as a sum of spherical harmonics [9].

$$N(\theta,\lambda) = a\sum_{l=0}^{\infty}\sum_{m=0}^{l}\overline{P}_{lm}(\cos(\theta))(C_{lm}\cos(m\lambda)+S_{lm}\sin(m\lambda)) \qquad \textbf{(2.1)}$$

where a is the radius of the Earth, θ and λ are colatitude and east longitude, C_{lm} and S_{lm} are dimensionless coefficients and the P_{lm} are normalized associated Legendre functions for every degree(l) and every order (m) [9].

$$\overline{P}_{lm} = \sqrt{(2-\delta_{m0})(2l+1)\frac{(l-m)!}{(l+m)!}}\cdot\frac{(1-x^2)^{\frac{m}{2}}}{2^l l!}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l \qquad \textbf{(2.2)}$$

A typical GRACE geoid model consists of numerical values for the C_{lm} and S_{lm} variables complete to degree and order of about 60-90 every few weeks [9].

Suppose there is a time-dependent change in the geoid ΔN. One could imagine ΔN as representing either the change in N from one time to another, or as the difference between N at one time and a time average of N, or as some other representation of a changing N. This change in N can be represented in terms of changes ΔC_{lm} and ΔS_{lm} in the spherical harmonics geoid coefficients changes of Legendre function ΔP_{lm} as

$$\Delta N(\theta,\lambda) = a\sum_{l=0}^{\infty}\sum_{m=0}^{l}\Delta\overline{P}_{lm}(\cos(\theta))(\Delta C_{lm}\cos(m\lambda)+\Delta S_{lm}\sin(m\lambda)). \qquad \textbf{(2.3)}$$

$\Delta\rho(r,\theta,\lambda)$ be the density redistribution causing this geoid change. For the Spherical harmonics coefficients it can be shown like:

$$\begin{Bmatrix}\Delta C_{lm}\\ \Delta S_{lm}\end{Bmatrix} = \frac{3}{4\pi\alpha\rho_{ave}(2l+1)}\int\Delta\rho(r,\theta,\lambda)\overline{P}_{lm}(\cos(\theta)) \times\left(\frac{r}{a}\right)^{l+2}\begin{Bmatrix}\cos(m\lambda)\\ \sin(m\lambda)\end{Bmatrix}\sin\theta\; d\theta\; d\lambda\; dr \qquad \textbf{(2.4)}$$

where ρ_{ave} is average density of the Earth(=5517 kg/m^3) [15].

Suppose $\Delta\rho$ is concentrated in a thin layer of thickness H at the Earth's surface. This thickness must be thick enough to include those portions of the atmosphere, oceans, ice caps, and below-ground water storage with significant mass fluctuations. Thus H is mostly determined by the thickness of the atmosphere and is of the order of 10-15 *km* [15].

The change in the surface density, $\Delta\sigma$, is the radial integral of $\Delta\rho$ through this layer:

$$\Delta\sigma(\theta,\lambda) = \int_{thinlayer} \Delta\rho(r,\theta,\lambda)dr \qquad \textbf{(2.5)}$$

The GRACE errors for large values of l are very large. It is unlikely to recover useful time variable geoid coefficients for $l >\approx 100$. In fact, most of the recoverable time-dependent gravity signal will be concentrated at degrees well below 80 or so. Hence, the sum over (l,m) in geoid change can be truncated to degrees $l < l_{max}$, where, at most, $l_{max} \approx 100$. Assume H is thin enough that $(l_{max}+2)H/a << 1$. Then $(r/a)^{l+2} \approx 1$, and so the equation (**2.5**) reduces to following equation [15]:

$$\begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix}_{surf\ mass} = \frac{3}{4\pi\alpha\rho_{ave}(2l+1)} \int \Delta\sigma(\theta,\lambda)\overline{P}_{lm}(\cos(\theta)) \times \begin{Bmatrix} \cos(m\lambda) \\ \sin(m\lambda) \end{Bmatrix} \sin\theta\ d\theta\ d\lambda. \qquad \textbf{(2.6)}$$

Equation (**2.6**) describes the contribution to the geoid from the direct gravitational attraction of the surface mass. That surface mass also loads and deforms the underlying solid Earth, which causes an additional geoid contribution [15]:

$$\begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix}_{solid\ E} = \frac{3k_l}{4\pi\alpha\rho_{ave}(2l+1)} \int \Delta\sigma(\theta,\lambda)\overline{P}_{lm}(\cos(\theta)) \times \begin{Bmatrix} \cos(m\lambda) \\ \sin(m\lambda) \end{Bmatrix} \sin\theta\ d\theta\ d\lambda \qquad \textbf{(2.7)}$$

where k_l is the load Love number of degree l. The total geoid changes is the sum of (**2.6**) and (**2.7**) [15]:

$$\begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix} = \begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix}_{surf\ mass} + \begin{Bmatrix} \Delta C_{lm} \\ \Delta S_{lm} \end{Bmatrix}_{solid\ E} \qquad \textbf{(2.8)}$$

These results for ΔC_{lm} and ΔS_{lm} can be written in a more compact form with expanding $\Delta\sigma$ as [15]:

$$\Delta\sigma(\theta,\lambda) = a\rho_w \sum_{l=0}^{\infty}\sum_{m=0}^{l} \overline{P}_{lm}(\cos\theta)(\Delta\widehat{C}_{lm}\cos(m\lambda) + \Delta\widehat{S}_{lm}\sin(m\lambda)) \qquad \textbf{(2.9)}$$

where ρ_w is density of the water. Including this to the above equations $\Delta\widehat{C}_{lm}$ and $\Delta\widehat{S}_{lm}$ become dimensionless. $\Delta\sigma/\rho_w$ is the change in surface mass expressed in equivalent water thickness. As $\overline{P}_{lm}$ variables are normalized so that

$$\int_0^{\pi} \overline{P}_{lm}^2(\cos\theta)\sin\theta\, d\theta = 2(2-\delta_{m,0}), \tag{2.10}$$

it can be concluded from (**2.9**) that [15]:

$$\left\{\begin{array}{c}\Delta\widehat{C}_{lm}\\ \Delta\widehat{S}_{lm}\end{array}\right\} = \frac{1}{4\pi a\rho_w}\int_0^{2\pi} d\lambda \int_0^{\pi} \sin\theta\, d\theta \times\Delta\sigma(\theta,\lambda)\,\overline{P}_{lm}(\cos\theta)\left\{\begin{array}{c}\cos(m\lambda)\\ \sin(m\lambda)\end{array}\right\}. \tag{2.11}$$

By introducing (**2.6**) and (**2.7**) in (**2.8**) and using (**2.11**) a simple relation between ΔC_{lm}, ΔS_{lm}, $\Delta\widehat{C}_{lm}$ and $\Delta\widehat{S}_{lm}$ can be obtained:

$$\left\{\begin{array}{c}\Delta C_{lm}\\ \Delta S_{lm}\end{array}\right\} = \frac{3\rho_w + 1 + k_l}{\rho_{ave} 2l+1}\left\{\begin{array}{c}\Delta\widehat{C}_{lm}\\ \Delta\widehat{S}_{lm}\end{array}\right\} \tag{2.12}$$

. Or, conversely,

$$\left\{\begin{array}{c}\Delta\widehat{C}_{lm}\\ \Delta\widehat{S}_{lm}\end{array}\right\} = \frac{\rho_{ave}}{3\rho_w}\frac{2l+1}{1+k_l}\left\{\begin{array}{c}\Delta C_{lm}\\ \Delta S_{lm}\end{array}\right\}. \tag{2.13}$$

Introducing(**2.13**) in (**2.9**) gives

$$\Delta\sigma(\theta,\lambda) = \frac{a\rho_{ave}}{3}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\overline{P}_{lm}(\cos\theta)\frac{2l+1}{1+k_l}\times(\Delta C_{lm}\cos(m\lambda)+\Delta S_{lm}\sin(m\lambda)) \tag{2.14}$$

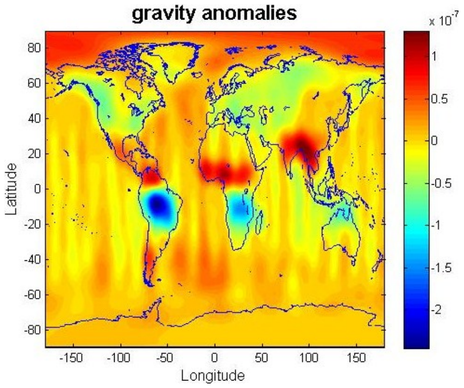
which can be used to find the change in surface mass density from changes ΔC_{lm} and ΔS_{lm} in the geoid coefficients [15].

Similarly, using equation (**2.12**) gives

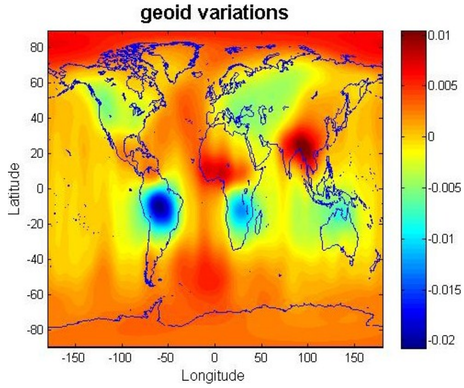
$$\Delta N(\theta,\lambda) = \frac{3a\rho_w}{\rho_{ave}}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\overline{P}_{lm}(\cos\theta)\frac{1+k_l}{2l+1}\times(\Delta\widehat{C}_{lm}\cos(m\lambda)+\Delta\widehat{S}_{lm}\sin(m\lambda)) \tag{2.15}$$

which together with (**2.11**) gives the change in the geoid from knowledge of the change in surface mass density [15].

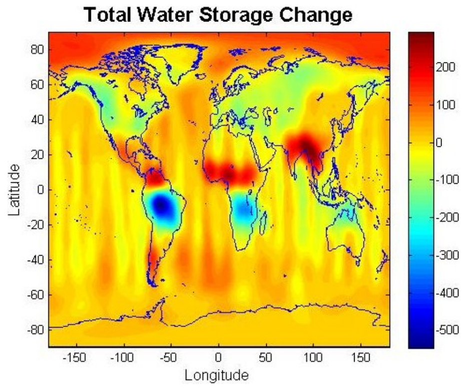
The level-2 GRACE products are generally given as gravitational potential in terms of spherical harmonics. In this part of study, different filtered gravity field functions, which are gravity anomalies, geoid variations and total water storage, are computed from 2010-2008 March and 2010-2008 September GRACE level-2 data as gridded on the Earth's surface.



(a) Gravity anomalies ($MGal$)



(b) Gravity variations ($MGal$)



(c) Total water storage change (cm)

Figure 2.4: Variance for each function during the period, between March and September, calculated from GRACE level-2 data. The data is filtered with Gaussian filter with spherical harmonics degrees of 30 and 500 *km* half wavelengths.

As seen in the Figure 2.4, difference of two data gives changes during six months. With this difference, the regions which shows normally semiannually variation in total water storage both on land and oceans are visible. The mean density of the Earth is used in this study as $q_e = 5540 kg/m^3$ and the $k_l^{'}$ the elastic gravity load Love numbers. As J. Wahr et al. in 1998 explained; $k_l^{'}$ can be computed with using the Preliminary Reference Earth Model (PREM), values until degree 200 are calculated in their study and located in a table, for higher degrees, the values can be interpolated [15].

3. PRINCIPAL COMPONENT ANALYSIS

Products of geodetic observing systems (GRACE, altimeter) and geophysical modeling are most often represented in form of time series of spatial maps (total water storage, sea level anomalies, and others). The user of these products will often find spatial patterns dominating the variability within these maps. Identifying these patterns can aid in physical interpretation, comparison of different data sets, and removing unnecessary small-scale signals or noise. Eigenspace techniques as the principal component method are among the most popular analysis techniques supporting these objectives [16].

In this part of the study Principal Component (PCA) will be briefly introduced. In chapter 4, local and global analysis with using GRACE and WGHM will be carried out.

3.1 Theory

Sampling spatio-temporal fields can lead to huge amounts of data. It is now a challenging task to reduce the dimensionality of the data vector and to identify the most important patterns explaining the variability of the system. The Empirical Orthogonal Functions (EOF) technique, also called Principal Component Analysis (PCA), has become one of the most widely used methods. PCA reduces a data set containing a large number of variables to a data set containing fewer new variables. These new variables are linear combinations of the original ones, and these linear combinations are chosen to represent the maximum possible fraction of the variability contained in the original data [16].

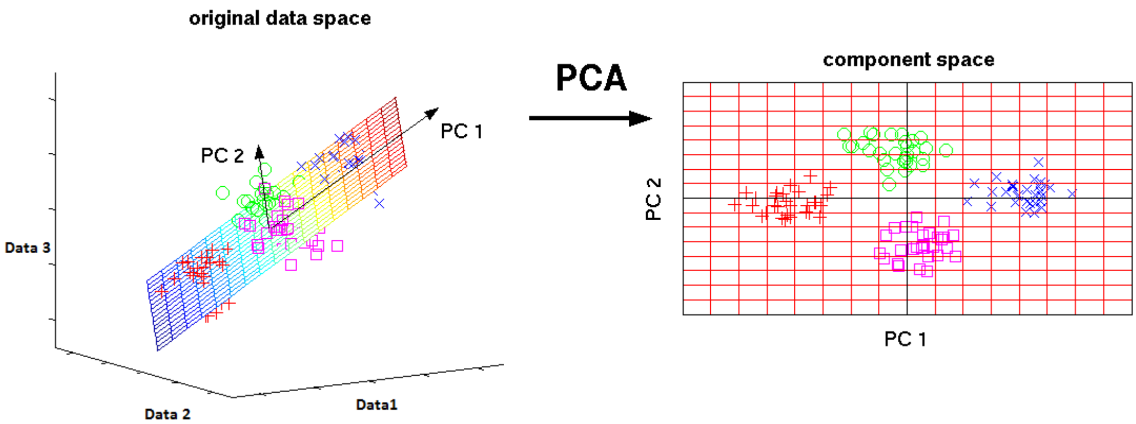


Figure 3.1: Understanding of PCA [1].

PCA has been used extensively to extract individual dominant modes, which are connected with low data variability, and reducing the number of data efficiently as demonstrated in the figure 3.1. The physical interoperability of the obtained pattern is, however, a point of discussion as the obtained modes are by definition orthogonal in space and time and this is not necessarily so in reality. Observed or modelled data in n locations are expressed as n x 1 data vector y, given for p time epochs t_i [17],

$$y_i = \begin{pmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{n,i} \end{pmatrix} \quad i = 1 \ldots p \tag{3.1}$$

For the following, the y vector will be assumed centered, so the time average per node $\frac{1}{p}\sum_{i=1}^{p} y_{j,i}$ is already reduced from the observations $y_{j,i}$. Another way to look at the (**3.1**) is decomposition of the data vector $y_i = Iy_i$ according to the individual locations. The base vectors u_j are independent of time, orthogonal, normalized with respect to the standard scalar product $(a,b) = a^T b$ [17].

$$y_i = y_{1,i}\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + y_{2,i}\begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + y_{n,i}\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} = y_{1,i}u_1 + y_{2,i}u_2 + \ldots + y_{n,i}u_n \tag{3.2}$$

All y_i can take part in one matrix with assuming there is no gap in the time series;

$$Y = (y_1, y_2, \ldots, y_n) = \begin{pmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,p} \\ y_{2,1} & y_{2,p} & \cdots & y_{2,p} \\ \vdots & \vdots & & \vdots \\ y_{n,1} & y_{n,p} & \cdots & y_{n,p} \end{pmatrix}. \tag{3.3}$$

Its rows contain the time series per location, whereas its columns contain the entire data from all locations per time epoch. The first principal components (Y_1) is given by the linear combination of the variables $X_1, X_2, ..., X_p$.

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \ldots + a_{1p}X_p \tag{3.4}$$

or, in matrix notation

$$Y_1 = a_1^T X \tag{3.5}$$

The first principle component is calculated such that it accounts for the greatest possible variance in the data set. Weights are calculated with constraint that their sum of squares is 1.

$$a_{11}^2 + a_{12}^2 + \ldots + a_{1p}^p = 1 \tag{3.6}$$

The second principle component is calculated in the same way, with the condition that it is uncorrelated with (perpendicular to) the first principle component and that it expresses the next highest variance.

$$Y_2 = a_{21}X_1 + a_{22}X_2 + \ldots + a_{2p}X_p \tag{3.7}$$

This continues until a total of p principle components have been calculated, equal to the original number of variables. At this point, the sum of the variances of the principle components will equal the sum of the variances of all of the individual variables. Collectively, all of these transformations of the original variables to the principal components leads to [17]:

$$Y = AX. \tag{3.8}$$

The rows of matrix A are called the eigenvectors of the variance-covariance matrix of the original data. The elements of an eigenvector are the weights a_{ij}, and are also known as loadings. The elements in the diagonal of the variance covariance matrix of the principal components, are known as the eigenvalues. Eigenvalues are the variance explained by each principal component, and to repeat, are constrained to decrease monotonically from the first principal component to the last [17]. The positions of each observation in this new coordinate system of principal components are called scores and are calculated as linear combinations of the original variables and the weights a_{ij}.

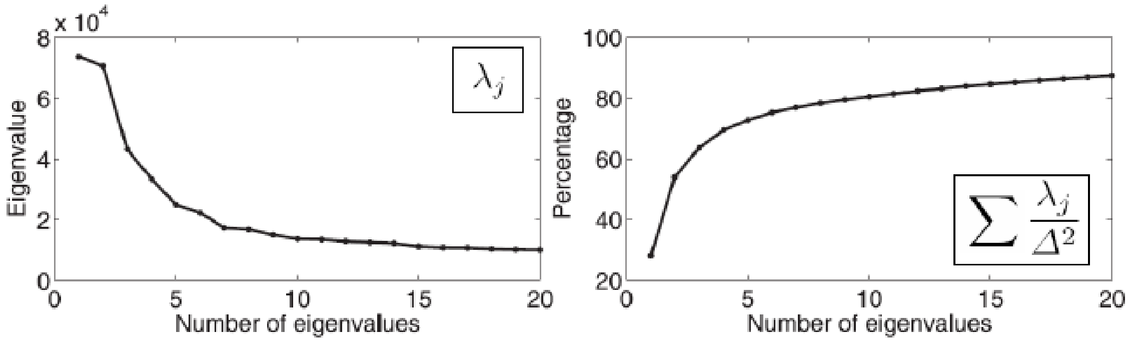


Figure 3.2: Relationship between eigenvalues and variance.

For example, the score for the r^{th} sample on the k^{th} principal component is calculated as [17],

$$Y_{kr} = a_{k1}x_{k1} + a_{k2}x_{k2} + \ldots + a_{kp}x_{kp}. \qquad \textbf{(3.9)}$$

The correlations of the original variables with the principal components for variable X_i and principal component Y_j can be calculated as followed [17]:

$$r_{ij} = \sqrt{a_{ij}^2 Var/s_{ii}}. \qquad \textbf{(3.10)}$$

3.1.1 Number of Modes

The question of interest is: How many PCs should be investigated and which mode is enough to describe the data sufficiently? There are several approaches about it;

- One common criteria is to ignore principal components at the point at which the next PC offers little increase in the total variance explained [17].
- A second criteria is to include all those PCs up to a predetermined total percent variance explained, such as 90 % [17].
- A third standard is to ignore components whose variance explained is less than 1, when a correlation matrix is used or less than the average variance explained when a covariance matrix is used, with the idea being that such a PC offers less than one variable's worth of information [17].
- A fourth standard is to ignore the last PCs whose variance explained is all roughly equal to each other [17].

4. APPLICATION of PCA for ANALYZING GRACE DATA

In this chapter, application part of the study will be carried out. It consist of 4 main parts.

Mask definition and observation of signal loss and leakage Comparison of PCA of different GFZ release GRACE data regionally examined in section 4.2. CS2, CS3 and CS6 data-sets are used to analyze Amazon, Greenland, Mekong, Turkey and Tigris/Euphrates rivers. Analysis effects of the gaps to PCA results investigated in section 4.3 for Amazon, Greenland and Mekong regions. PCA results of CS2, CS3 and CS6 are compared with in-situ data over in section 4.4. PCA results of WGHM and GRACE data compared globally and regionally. Amazon, Greenland, Orinoco, Turkey and Tigris/Euphrates rivers are chosen as study are in section 4.5. For deep understanding of the data-sets which are used during the analyses, please check section 1.1.

4.1 Area Averaging

Goal of this part of the thesis is, with considering the signal loss as result of the truncation in the spherical harmonics synthesis at maximum degree/order in the spectral domain, examining ways of defining basin function for extracting regional mass anomalies from GRACE gravity coefficients and use these basin functions for regional water storage estimates. The basin function is defined with following two different approaches for Greenland, Mekong and Amazon regions. Firstly, basin function according to basin shape is constructed for the region of interest. Each basin function is applied to both hydrological and GRACE-like data. GRACE-like data spherical harmonic coefficient are estimated up to 90 and 50 degree and order. Signal loss due to this truncation is observed according to the area average difference of two data-sets. Basin functions are scaled and signal lost over the area calculated again. After as the second approach, regarding to signal amplitude Greenland and Amazon masks are defined. This masks will be qualified with GRACE data in the chapter 4.2.

According to first approach, surface mass change estimate for a space-limited region of interest R with area R_0 requires a basin function h, which is defined as 1 inside R and 0 outside the basin, i.e.,

$$h(X) = \begin{cases} 1 & \text{if X} \in R \\ 0 & \text{if X} \in \Omega - R \end{cases} \quad \textbf{(4.1)}$$

, where X is position on the earth surface $X = (\theta, \varphi)$, and Ω represents the entire Earth surface. If L_{max} (estimated maximum degree and order for stokes coefficients) approaches to infinity, the basin function approaches its optimum shape, but for finite L_{max} it becomes only an approximation to the exact h(X) [15].

The GRACE water storage estimate $\hat{\hat{S}}_0$ over the region of interest of area R_0 is expressed as [15]:

$$\hat{\hat{S}}_0 = \frac{1}{R_0} \int \hat{\hat{S}} h d\Omega \quad \textbf{(4.2)}$$

.

A basin function can be described either on the Earth's surface or in terms of SH coefficients $h(X) \rightarrow (h^c_{l,m}$ and $h^s_{l,m})$; therefore, the estimate can also be written as a linear combination of the GRACE Stokes coefficients

$$\hat{\hat{S}}_0 = \frac{4\pi a^3 \rho_e}{3R_0} \sum_{L_{max}}^{l=0} \sum_{l}^{m=0} \frac{2l+1}{1+k_l'} (C_{l,m} h^C_{l,m} + S_{l,m} h^S_{l,m}) \quad \textbf{(4.3)}$$

, where ρ_e represents the Earth's mean density. The weighting factor $\frac{2l+1}{1+k_l'}$ accounts for Earth's response to surface mass.

Gaussian filter, which is examined in previous chapter, however not capable of isolating a specific region. In order to use GRACE to estimate regional changes in surface mass, a different approach should be followed which extract regional mass anomalies from GRACE coefficients [18]. Since water storage estimates can be express as a set of linear weights applied to the Stokes coefficients, these estimates can also be developed with the goal of both maximizing resolution within the region of interest and filtering high-degree and high-order Stokes coefficients to suppress noise [19].

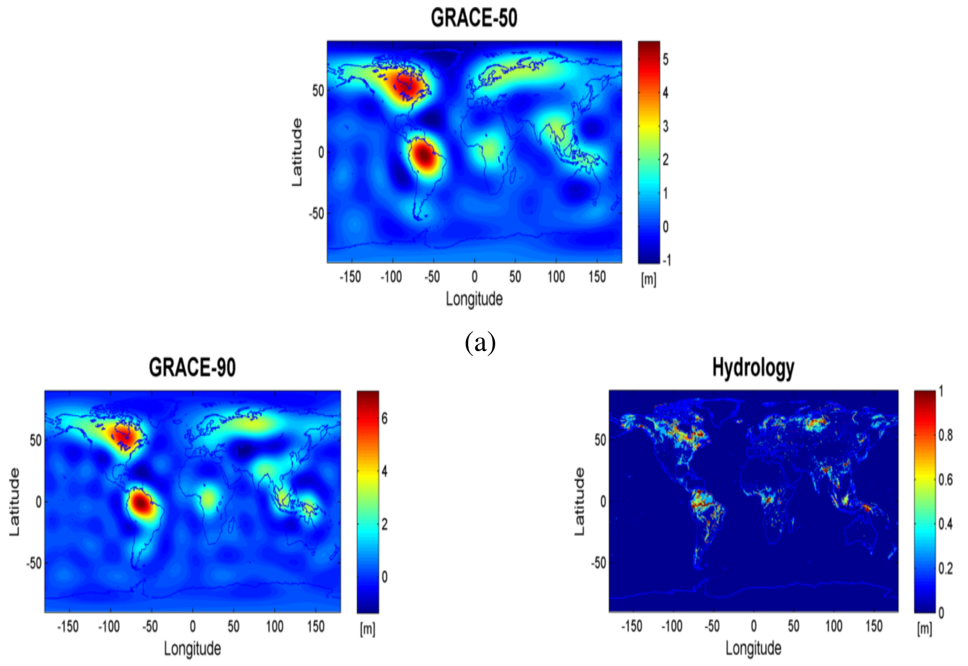
The effective basin function (EBF) $\hat{h}$ is defined as the spatial function on the surface of the sphere, and its corresponding SH weights are applied to GRACE Stokes coefficients developed. $\hat{h}$ describes the ability of GRACE to determine an average water storage change within the basin of interest [19].

Because GRACE estimates Stokes Coefficients until L_{max} components of surface mass variability corresponding to $l > L_{max}$ will be missing from calculation of the basin average. The contribution of high l Stokes coefficients to large basin averages may be disregarded, since they are quite small. However, as the basin size decreases, their contribution becomes more significant. In addition to this truncation error, GRACE will also contain satellite measurement errors. Truncating the sum at $L_{max} < 100$ will minimize these measurements errors but, on the other hand, it increases truncation error [18].

Since GRACE has limited spatial resolution, the data resolution corresponds to a spectral resolution of 360 degree and order of the spherical harmonic expansion, can not be recovered by GRACE products. Bias and leakage effects which occur due to this limitation, depends on actual water storage variations (both interior and exterior to the basin) and must be generally estimated from hydrological model [19]. To enable the creation of the desired result in the spatial domain, a test data set that simulates continental hydrology (figure 4.1c) and GRACE-like data derived from this data set was created by GFZ. This permits us to make a comparison of both data-set in the spectral domain. In order to analyze the truncation effect using the GRACE like data, the spherical harmonic synthesis in equivalent water heights is estimated up to 50 and 90 degree and order (figure 4.1a and 4.1b). Differences between the results of the synthetic GRACE product and its reference data (the continental hydrology simulation data set) are compared each other for the Amazon, Greenland and Mekong regions. Region of interest defined with using basin function which base on region borders as our first approach. Variance according to the maximum degree is observed for each of these areas.

The border base masks of Tigris-Euprates rivers system, Amazon, Greenland, Mekong regions are tested with continental hydrology (true data) and the GRACE-like data to observe leakage according to truncation. These masks are enlarged according to nearest-neighbor interpolation method and the signal lost examination is repeated. Analysis results are collected in the table 4.1.

For the Tigris-Euphrates masks, it is determined that, GRACE like data has negative signal lost. This can be interpreted that this region is occupied by leakage of surroundings. As also seen in the table 4.1 Greenland can have more than 4% signal lost

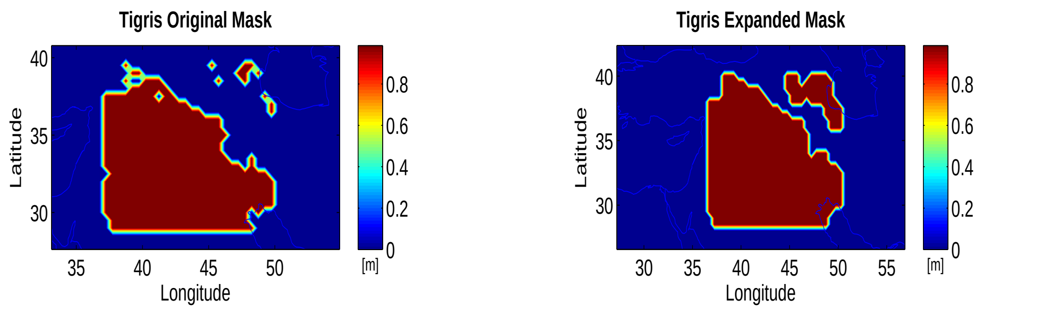


(a) (b) (c)

Figure 4.1: Synthetic GRACE product derived from simulated continental hydrological data in equivalent water height to a maximum degree of 90 (a), maximum degree of 50 (b) and the test data that simulates continental hydrology (c).

due to truncation. This amount is more significant for higher lower degree/order. As Greenland, Mekong and Amazon masks also increase signal coverage at high maximum degree/order. However, signal lost can not be recovered as efficiently as it is desired with the expanded mask.

Failure to include absence of basin coefficients for L_{max} (ϑ^c_{lm} and ϑ^s_{lm}) results an inaccurate presentation of basin shape. The ringing which occur near the boundaries of the basin mask, called the Gibbs phenomenon, is due to this deficiency. Ringing



(a) Original Tigris & Euphrates mask.

(b) Expanded Tigris & Euphrates mask

Figure 4.2: Border base Tigris & Euphrates and extended version of the mask.

Table 4.1: Signal coverage of masks over and expanded mask

Mask	Maximum degree	Signal loss
Tigris Euphrates	90	-1.0478
Tigris Euphrates	50	-0.5901
Expanded Tigris	90	-0.5901
Expanded Tigris	50	-0.5901
Amazon	90	0.0903
Amazon	50	0.1493
Expanded Amazon	90	0.1172
Expanded Amazon	50	0.1695
Mekong	90	0.8177
Mekong	50	0.8179
Expanded Mekong	90	0.8259
Expanded Mekong	50	0.8263
Greenland	90	1.9856
Greenland	50	4.7067
Expanded Greenland	90	1.8351
Expanded Greenland	50	4.3607

increases as L_{max} (l_{trnc}) decreases, resulting in a basin average more of the region outside of the basin [18].

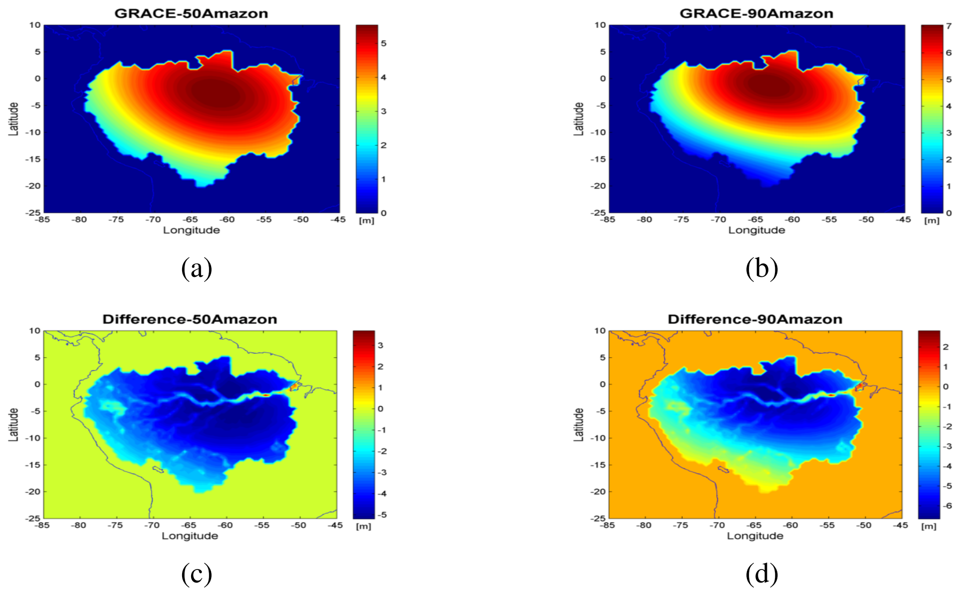


Figure 4.3: Total water storage derived from GRACE like data (until maximum degree 90 and 50) delimited with the Amazon border mask and differences between continental hydrology and GRACE-like data in equivalent water height.

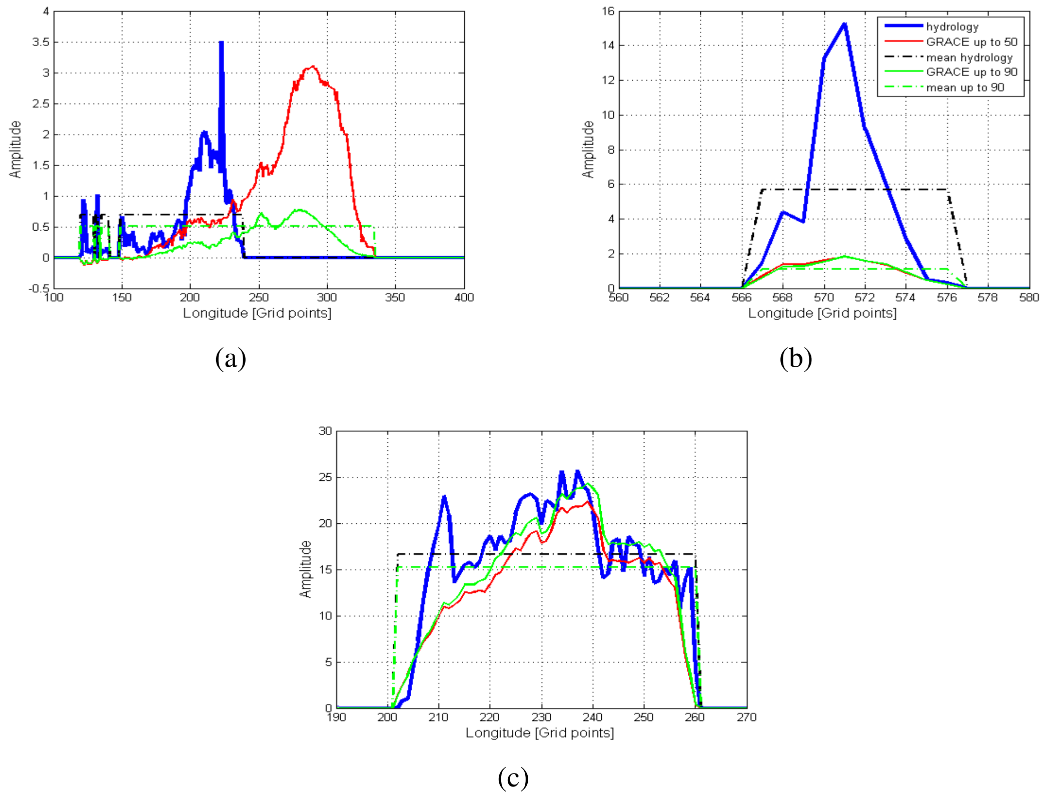


Figure 4.6: Demonstration of signal lost due to truncation. Figure (a) for Greenland, figure (b) for Amazon and figure (c) for Mekong region compares hydrology (blue), Grace like data truncated at 90 (red) and 50 (green) degree and order. Also mean hydrology (black) and (dash-dot line green) emphasizes lost signal amount.

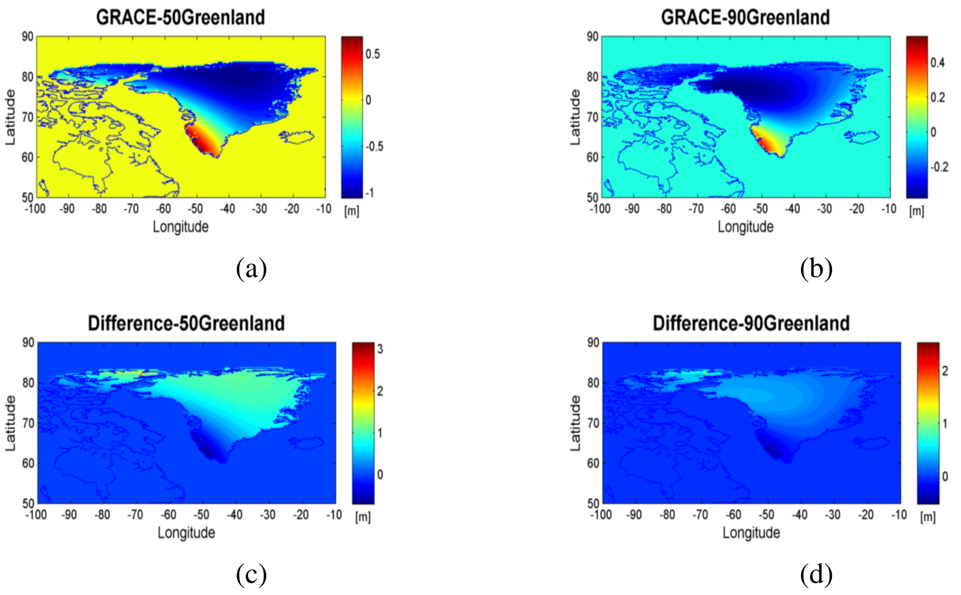
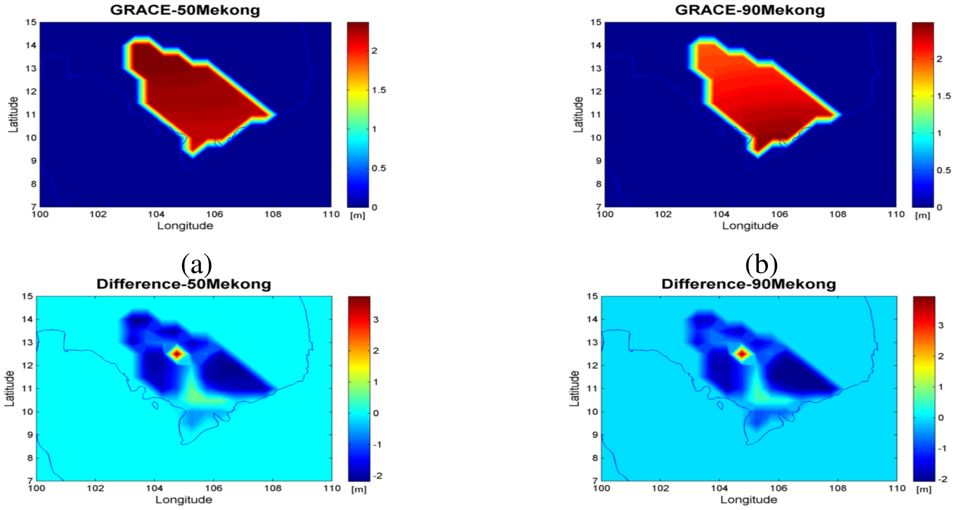


Figure 4.4: Total water storage derived from GRACE like data (until maximum degree 90 and 50) delimited with the Greenland border mask and differences between continental hydrology and GRACE-like data in equivalent water height.

The results indicate that, the TWS from GRACE-like data up to 90 degree/order (maximum degree) match with the hydrological data better than TWS from GRACE-like data until 50 degree in each region. The improvement is more distinctive at the region borders, where the leakage error is dominant. This point demonstrates how the leakage is altered according to the truncated order/degree. Additionally, it is observed that signal lost is most significant for the smaller regions as seen in Mekong. Signal lost between hydrology and GRACE-like which is calculated up to 90 degree and order is determined 80.6%. Nonetheless, this value is nearly same for $L_{max} = 50$. For Amazon 8.7% and 15.8% of the initial signal missing for $L_{max} = 90$ and $L_{max} = 50$ respectively. Signal lost for each region illustrated in figure 4.6.

A direct basin comparison requires signal omission due to bandwidth limitation of the solution (degree/order maximum degree 50-90). This leads in general to an augmentation of the investigated region. But if the region of interest neighbors with geophysical signal of opposite sign, such an increase would again reduce the signal by leakage-in effects from the neighboring geophysical phenomenon. Thus, a trade-off for an optimal mask has to be found that ideally separates both influences. Such a mask can be applied in the spatial and spectral domain equally [20], [21], [22].



(a) (b) (c) (d)

Figure 4.5: Total water storage derived from GRACE like data (until maximum degree 90 and 50) delimited with the Mekong border mask and differences between continental hydrology and GRACE-like data in equivalent water height.

In order to define basin functions as second approach, the supplied data-set which consists of spectrally analysed coefficients for variable masks is used. Spherical harmonic synthesis in equivalent water level is computed with up to 50 and 90 degree and order. To identify the regions covering the signal wavelength, a threshold is set, resetting all float mask values above it to 1 and every value below to 0. It is evident that lower degree/order requires larger masks due to the increased leakage effect.

Since filtering and truncation modify the basin shape, the mean value of the EBF is no longer unity (which leads to bias estimate), and EBF is not zero outside of the basin, making the estimates sensitive to water storage changes outside the basin of interest (leakage). Bias and leakage time series can be calculated numerically considering the EBF $\hat{h}$: $\bar{S_0} = \hat{\bar{S_0}} + bias - leak$ where

$$bias = \frac{1}{R_0} \int_R S_0 (h - \hat{h}) d\Omega \tag{4.4}$$

$$leak = \frac{1}{R_0} \int_{\Omega - R} S_{leak} d\Omega \tag{4.5}$$

. S_0 is true stored water variation and S_0 is average stored water over R. S_{leak} is true stored water variation outside the basin, and $\hat{\bar{S_0}}$ is the bias estimate before leakage and bias corrections [23]. The resulting masks according to signal amplitude for Amazon and Greenland are illustrated in figure 4.7.

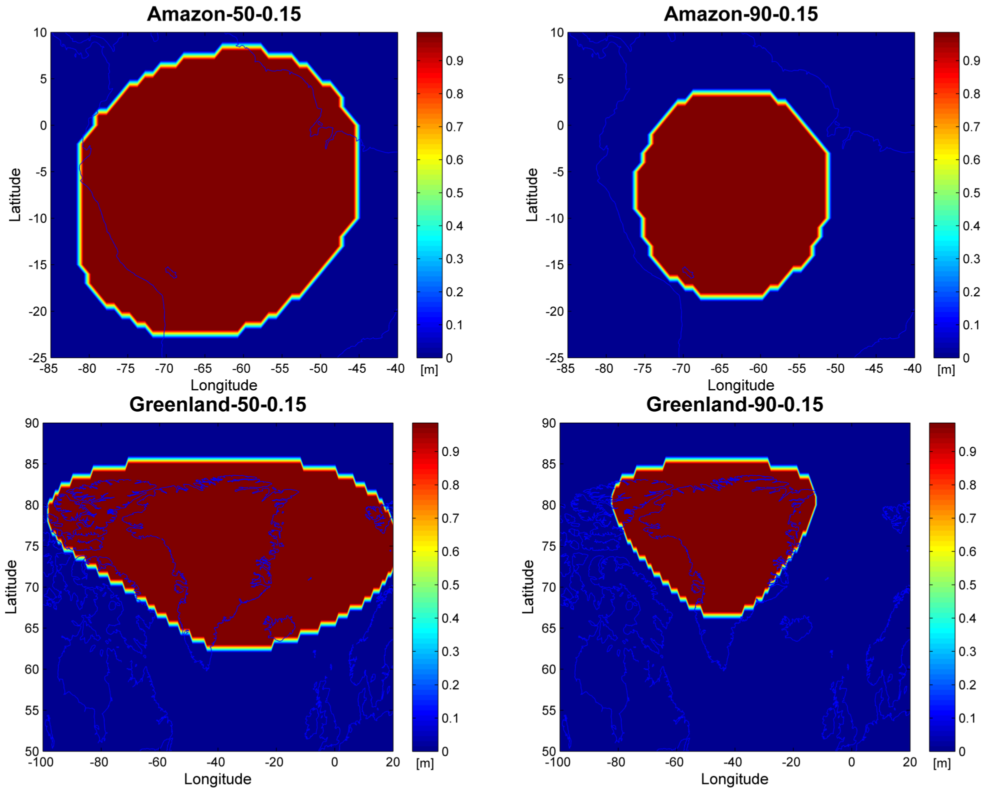


Figure 4.7: Defined masks for Amazon and Greenland which ore developed up to 50 and 90 degree and order. The threshold chosen is 0.15.

As mentioned in previously, the low spatial resolution of GRACE solutions causes leakage where signals spread out into nearby regions. Figure 4.7 illustrate the leakage effect unmistakably. As a consequence of signal damping, basin shapes are modified in each data set. Since truncation error is more significant for L_{max}=50 than L_{max}=90 modification is also higher. Performance of these masks will be tested with different GRACE data in the following chapter.

4.2 Comparison of Different Version of GRACE Data Locally with Using PCA

In this part of the study, different GFZ GRACE L-2 release data sets are tested regionally. As analysis method, PCA (explained in the 3) used. The improvement between RL04 to RL05 assesed.

According to the results of *Bonin and Chambers'* study, the "block layout" produces larger errors. Trying to divide into basins base on best estimates of signal covariance rather then easily definable rectangles can reduce the errors [24]. It is a good argument to use masks for regional analysis. However, it is also crucial if the mask covers the signal or not. In this part of the study, the Amazon, Greenland, Mekong, Turkey, Tigris&

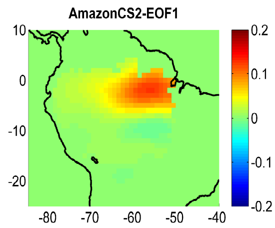
Euphrates regions are examined with using the time series CS6, CS3 and CS2 covering the time-span 2003 till 2012. Moreover, localization is carried out with border based masks for Amazon, Greenland, Mekong and the threshold-defined masks for Amazon and Greenland areas. Afterwards PCA is executed regionally, variations at the results with respect to the resolution of the GRACE data and mask type are displayed. For more information about the data, please check 1.1.

4.2.1 Amazon

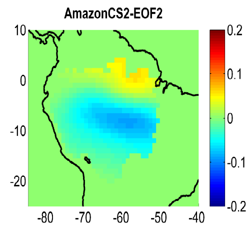
With tributaries extending from the vast savannas to its north and south, the Amazon River runs almost 4,000 *miles* (1 *mile* equals 1.6 *km*) across northern South America from the highland biomes in the foothills of the Andes Mountains to the Atlantic Ocean. It carries twenty percent of all river water discharged into Earth's oceans-ten times the volume of the Mississippi River. From December to May each year, torrential rains and snow melt from the Andes increase the main river channel's depth 30-45 *feet* (9 to 14 *m*), and water backs up in tributaries and inundates forest several miles from the main channel. In the central Amazon Basin alone, the flood waters can cover an area up to 97,000 mi^2 (about 250,000 km^2). The river and the flooded forests then come together as a giant, slow-moving swamp [25].

In this part of the study, data from GRACE will reveal how the water flow in the Amazon basin changes with respect to the seasons and which Asian aquifers are replenished by the monsoons. Later on, as a result of the Principle Component Analysis (PCA), behavior of total water storage change over the region visualized in the first three dominant modes. Performance of the PCA is tested with normalized, centered GRACE data with three different release and different resolutions for 120 month covers from 2003 to 2012. The results are compared.

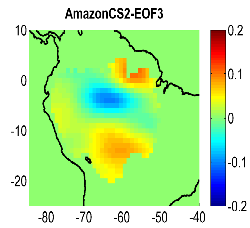
The study area is selected from the entire globe with a defined basin function, which is defined in chapter 4.1. Defined basin border mask, according to first approach, is applied to GFZ's CS2, CS3 and CS6 solutions and used for analyzing. To give insight into data and masking, September 2010 solution for each data-set are visualized in figure 4.8, before and after applying the mask.



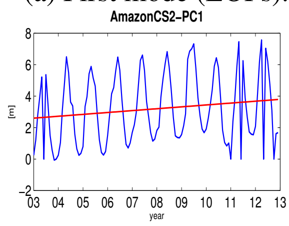
(a) First mode (EOFs).



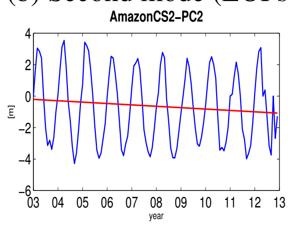
(b) Second mode (EOFs).



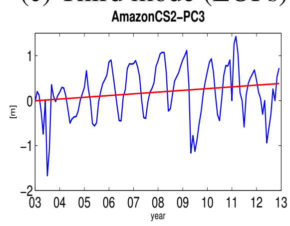
(c) Third mode (EOFs).



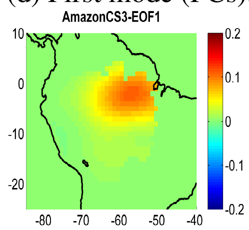
(d) First mode (PCs).



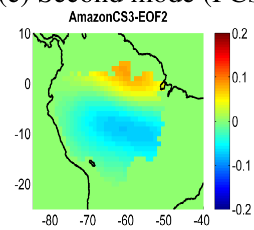
(e) Second mode (PCs).



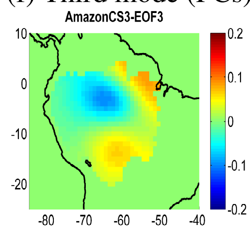
(f) Third mode (PCs).



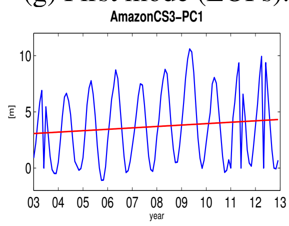
(g) First mode (EOFs).



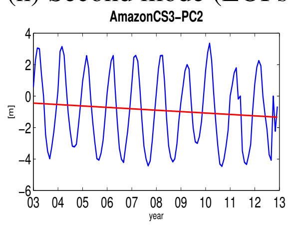
(h) Second mode (EOFs).



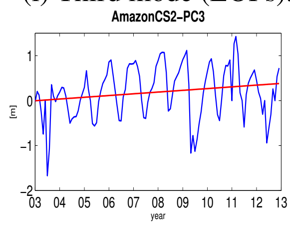
(i) Third mode (EOFs).



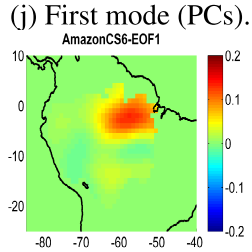
(j) First mode (PCs).



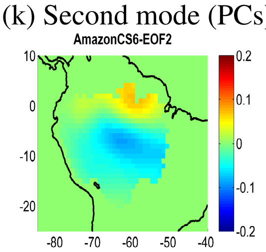
(k) Second mode (PCs).



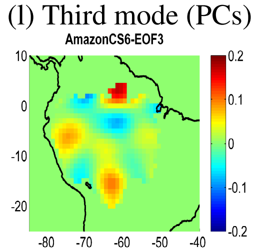
(l) Third mode (PCs).



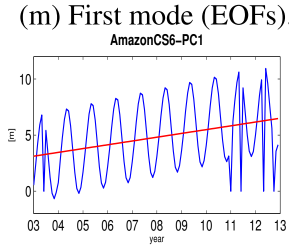
(m) First mode (EOFs).



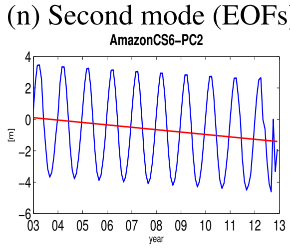
(n) Second mode (EOFs).



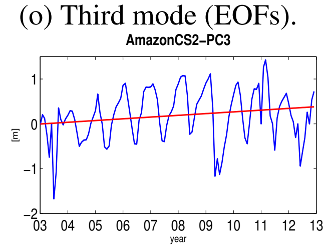
(o) Third mode (EOFs).



(p) First mode (PCs).



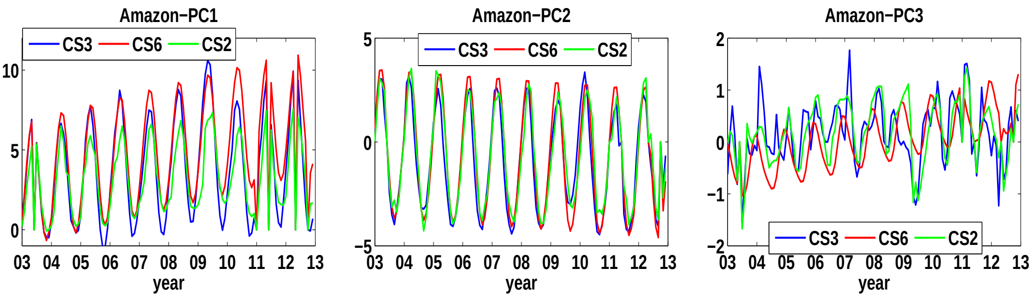
(q) Second mode (PCs).



(r) Third mode (PCs).

Figure 4.8: The spatial patterns (EOFs) and the temporal evolution (PCs) of CS2 (a, b, c, d, e, f) CS3 (g, h, i, j, k, l) and CS6 (m, n, o, q, r) data-set, normalized and centered for the years 2003-2012 with Amazon mask, first three dominant modes left to right.

As demonstrated in the figure 4.8, masking localizes the analysis. It eliminates any mass distribution data from the irrelevant regions. This process is repeated to all months of the GRACE time series to restrict the information to the constrained Amazon region. PCA is applied afterward.



(a) First mode. (b) Second mode. (c) Third mode.

Figure 4.9: Comparison of Principal components for the first three PCA modes. The graphs belong to CS3 (blue, 350 *km* spatial res.), CS2 (green, 250 *km* spatial res.) and CS6 (red, 500 *km* spatial res.) data with the Amazon border based mask applied.

It is seen that at the first mode for every data type a semi-annual behaviour manifests. Besides, the first mode is dominated by the effects of missing months in latter years. However, even though the CS6 effective data has an increasing trend at the third mode, this trend is not visible in CS2 and CS3 data, which might be an indication of the improvement of release 05 over the previous one.

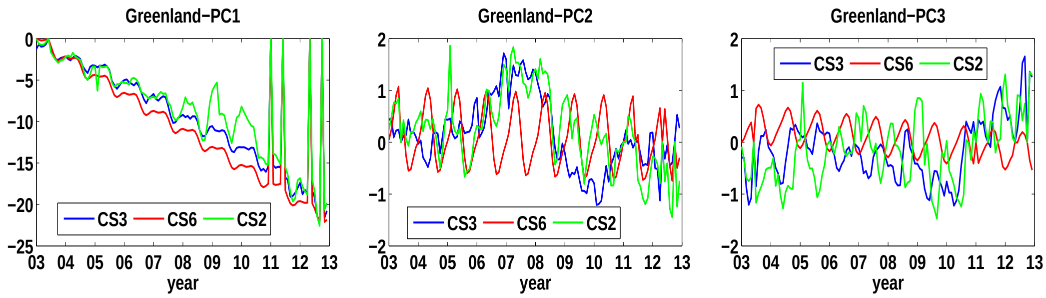
4.2.2 Greenland

The size of Greenland extends about 2,670 *km* from north to south and more than 1,050 *km* from east to west at its widest point. Two-thirds of the island lies within the Arctic Circle, and the island's northern extremity extends to within less than 800 *km* of the North Pole. Greenland is separated from Canada's Ellesmere Island to the north by only 26 *km*. The nearest European country is Iceland, lying about 320 *km* across the Denmark Strait to the southeast. Greenland's deeply indented coastline is 39,330 *km* long, a distance roughly equivalent to the Earth's circumference at the Equator. Greenland's major physical feature is its massive ice sheet, which is second only to Antarctica's in size. The Greenland Ice Sheet has an average thickness of 1,500 metres, reaches a maximum thickness of about 3,000 *m*, and covers with more than 1,800,000 km^2 over four-fifths of Greenland's total land area [26].

While scientists have been aware that climate change has caused Greenland's ice sheet to thin for decades, melting in the sheet's northeast region has been discovered for the first time, USA Today reports. Ice loss in the northeast region of Greenland has nearly tripled since 2003, with the ice sheet retreating 12.4 *miles* over the past decade and losing 10 billion tons of ice per year from April 2003 to April 2012, according to co-author Michael Bevis, a professor of earth sciences at Ohio State University [24].

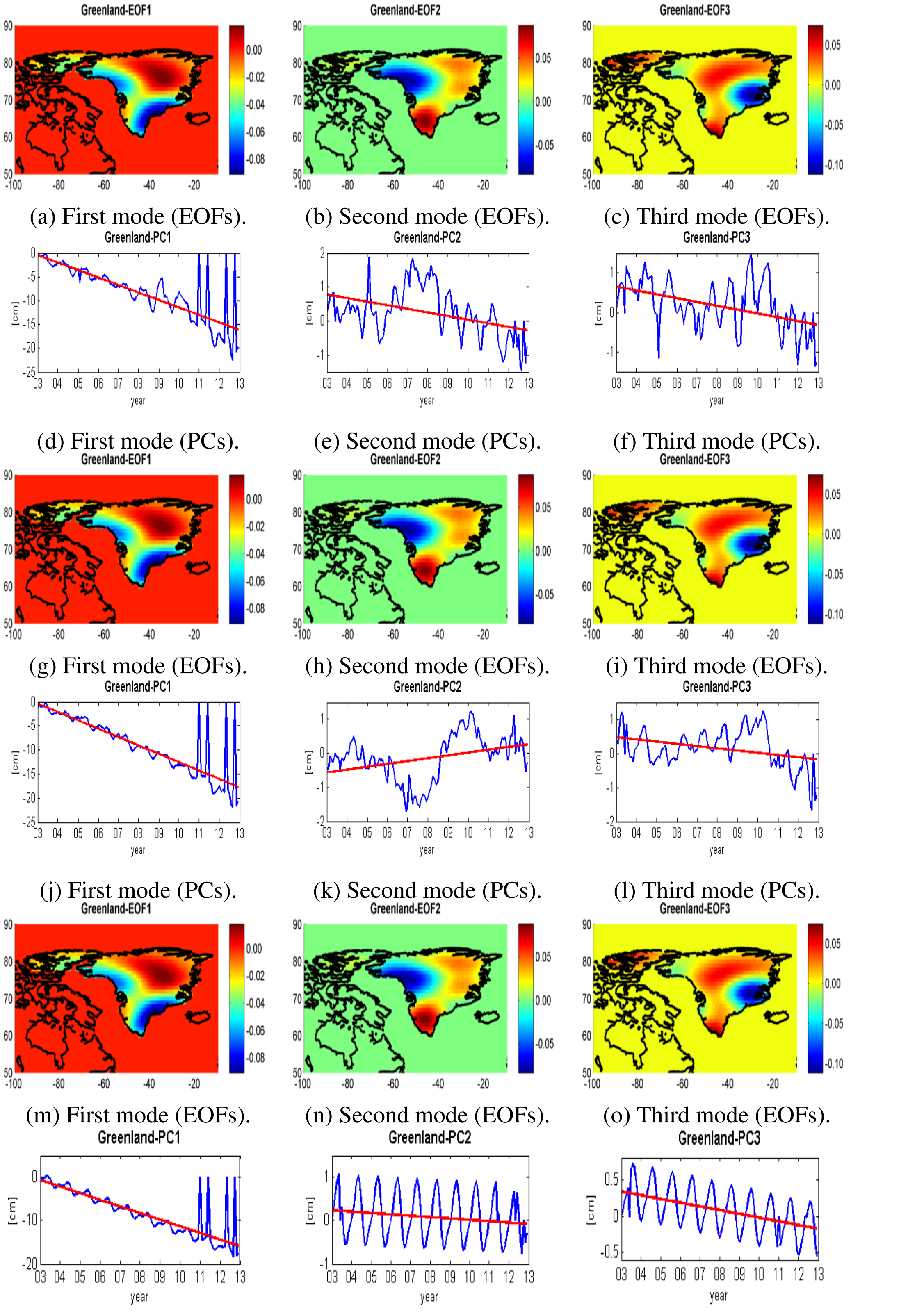
To get an overview of the great Greenland meltdown, gravity distribution provided by GRACE for each 30-day period has been used and this ice loss for each year, mainly from southern Greenland was determined in previous studies [4]. GRACE is also providing clues as to how the situation varies from year to year and enables to examine melting trends.

The border based Greenland mask is applied the CS2, CS3 and CS6 GRACE solutions from 2003 to 2013. Water storage change for Greenland is calculated and PCA applied. Results are compared and the effect of spatial resolution to the PCA results are analyzed.



(a) First mode. (b) Second mode. (c) Third mode.

Figure 4.10: Comparison of Principal components for the first three PCA modes. The graphs belong to CS3 (blue, 350 *km* spatial res.), CS2 (green, 250 *km* spatial res.) and CS6 (red, 500 *km* spatial res.) data with the Greenland border based mask applied.



(a) First mode (EOFs). (b) Second mode (EOFs). (c) Third mode (EOFs).

(d) First mode (PCs). (e) Second mode (PCs). (f) Third mode (PCs).

(g) First mode (EOFs). (h) Second mode (EOFs). (i) Third mode (EOFs).

(j) First mode (PCs). (k) Second mode (PCs). (l) Third mode (PCs).

(m) First mode (EOFs). (n) Second mode (EOFs). (o) Third mode (EOFs).

(p) First mode (PCs). (q) Second mode (PCs). (r) Third mode (PCs).

Figure 4.11: The spatial patterns (EOFs) (1st 3th and 5th row) and the temporal evolution (PCs) of CS2 (a, b, c, d, e, f), CS3 (g, h, i, j, k, l) and CS6 (m, n, o, q, r) data-set, normalized and centered for the years 2003-2012 with Greenland mask, first three dominant modes left to right.

As seen in the first mode of PCA for each data-set, there is a big mass of ice loss in Greenland.In the first mode, CS2 shows higher noise peaks starting from 2007. CS3 and CS6 have increasing trend up to 2007 and a decrease trend as following in second component. CS6 data show a permanent decrease trend instead. As a result, CS3 and CS2 show relatively similar results.

4.2.3 Test of threshold-defined masks

In this part of the investigation, the defined masks which are created in the section 4.1 according to the second mask creation approach 0.15 threshold are tested with CS2, CS3 and CS6 data (for more information about the data-set see section 1.1). Improvements at covered signal percentage with defined mask are demonstrated.

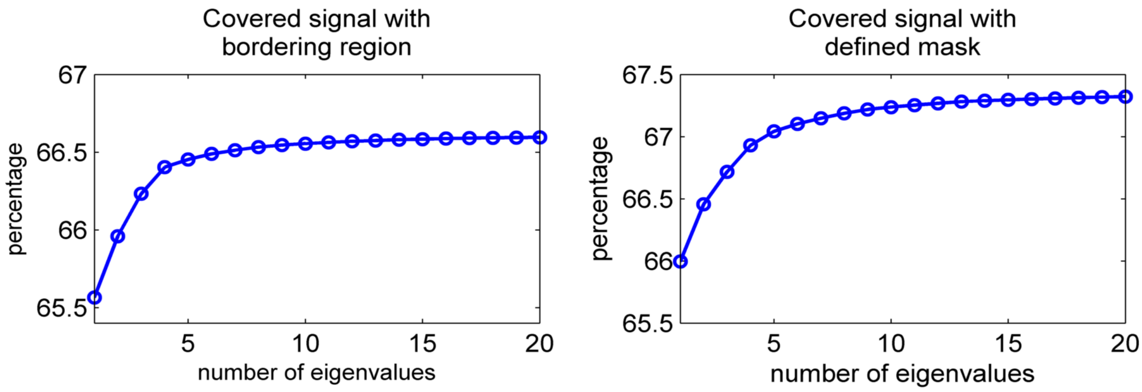


Figure 4.12: Covered signal percentages of CS2 data for the Greenland area. Addressing to the Greenland region is accomplished with a region border based mask (left) and a threshold-defined mask. It is defined by setting 0.15 threshold to the GRACE derived spherical harmonic coefficients (developed up to 90 degree and order by GFZ) transformed in equivalent water height up to 50 maximum degree (right).

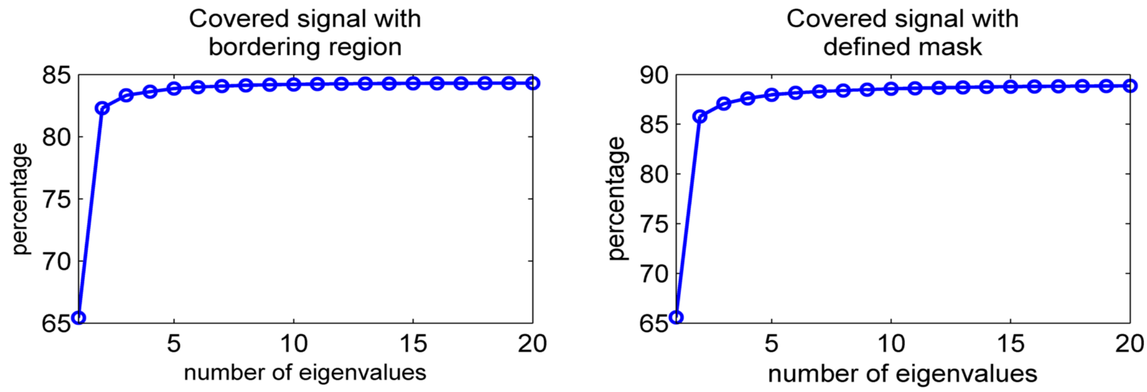


Figure 4.13: Covered signal percentages of CS3 data for the Amazon area. Addressing to the Amazon region is accomplished with a region border based mask (left) and a threshold-defined mask. It is defined by setting 0.15 threshold to the GRACE derived spherical harmonic coefficients (developed up to 90 degree and order by GFZ) transformed in equivalent water height up to 50 maximum degree (right).

Covered signal percentages of CS2, CS3 and CS6 data for the Amazon and Greenland regions. For both regions, masks are acquired from the GRACE derived spherical harmonic coefficients (developed up to 90 degree and order by GFZ) transformed in equivalent water height up to 50 maximum and 90. Increase in percentage achieved by threshold-masks is highlighted in red in table 4.14. The results demonstrate that for the Amazon region, it is possible to improve PCA results with using threshold base masks up to 4.5%.

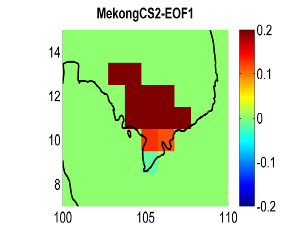
		bordered **RL05-250**	masked **RL05-250**	bordered **RL04-500**	masked **RL04-500**	bordered **RL05-350**	masked **RL05-350**	Diff 250	Diff 500	Diff 350
Maximum deg=**50**	first mode	65.38	65.5	65.35	65.47	65.42	65.57	0.12	0.12	0.15
Amazon	second mode	88.96	90.28	79.03	81.74	82.3	85.78	1.32	2.71	3.48
Th=**0.15**	third mode	90.54	92.08	79.58	82.74	83.32	87.06	1.54	3.16	3.74
	20th mode	92.3	94.37	80	83.4	84.3	88.85	**2.07**	**3.4**	**4.55**
	first mode	65.69	66.14	65.55	65.95	65.69	66.14	0.45	0.4	-0.26
Greenland	second mode	65.97	66.43	65.68	66.06	65.97	66.43	0.46	0.38	-0.09
	third mode	66.17	66.61	65.72	66.11	66.17	66.61	0.44	0.39	0.06
	20th mode	66.36	67.32	65.74	66.15	66.36	66.92	**0.96**	**0.41**	**0.21**
Maximum deg=**90**	first mode	65.38	65.36	65.35	65.33	65.42	65.41	-0.02	-0.02	-0.01
Amazon	second mode	88.96	90.75	79.03	79.05	82.3	83.34	1.79	0.02	1.04
Th=**0.15**	third mode	90.54	92.45	79.58	80.51	83.32	84.43	1.91	0.93	1.11
	20th mode	92.03	93.92	80	80.95	84.3	85.27	**1.89**	**0.95**	**0.97**
	first mode	65.56	65.62	65.55	65.66	65.69	65.79	0.06	0.11	0.1
Greenland	second mode	65.96	66.15	65.58	65.58	65.97	66.05	0.19	0	0.08
	third mode	66.23	66.05	65.72	65.76	66.17	66.24	-0.18	0.04	0.07
	20th mode	66.6	67.4	65.74	65.83	66.36	66.51	**0.8**	**0.09**	0.15

Figure 4.14: Covered signal percentages of CS2, CS3 and CS6, over Amazon, Greenland.

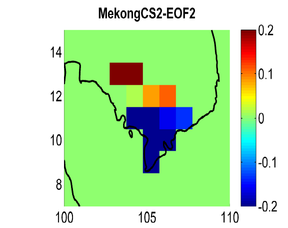
4.2.4 Mekong

The Mekong River basin can be divided into two parts: The Upper Basin located in Tibet, Burma, and China (where the river is called the Lancang Jiang), and the Lower Mekong Basin located in Lao PDR, Thailand, Vietnam and Cambodia. The climate of the Mekong Basin is dominated by the Monsoon, which generates wet and dry seasons of more or less equal length. The Southwest monsoon generates the wet season which usually lasts from May until late September or early October. Tropical cyclones occur over much of the area during August and September and even October (in the delta). The Northeast Monsoon brings lower temperatures from China and causes dry weather in the Lower Mekong Basin from late October until April [27].

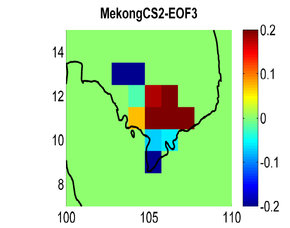
The Mekong basin is analyzed with using CS2, CS3 and CS6 GRACE data. This data is localized to the Mekong basin with applying the Mekong masks which is created according to the area borders. The results of PCA decomposition for CS2, CS3 and CS6 for the first three components are visualized in the figure 4.15 respectively.



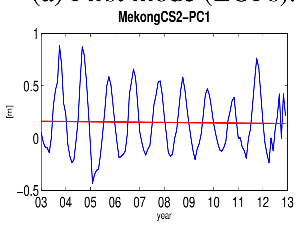
(a) First mode (EOFs).



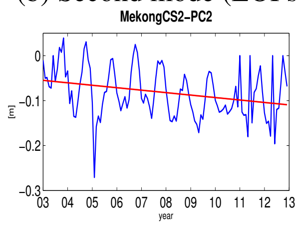
(b) Second mode (EOFs).



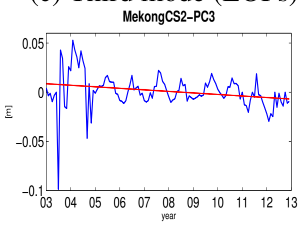
(c) Third mode (EOFs).



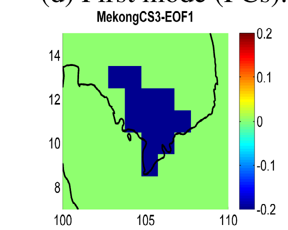
(d) First mode (PCs).



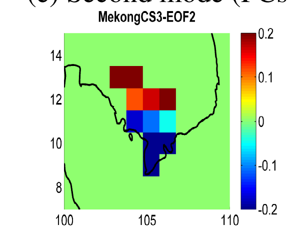
(e) Second mode (PCs).



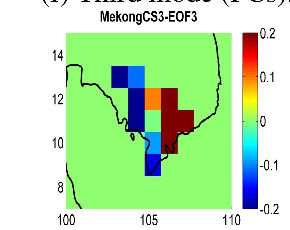
(f) Third mode (PCs).



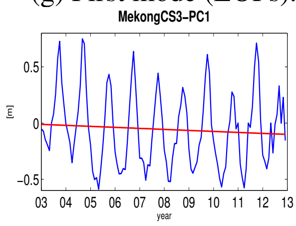
(g) First mode (EOFs).



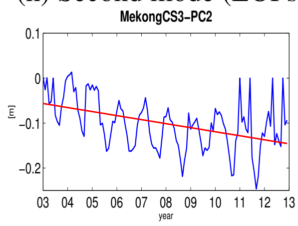
(h) Second mode (EOFs).



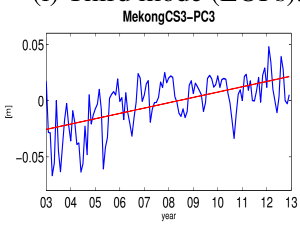
(i) Third mode (EOFs).



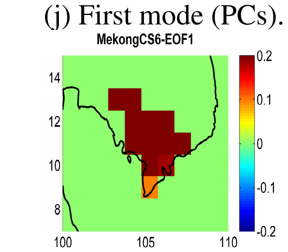
(j) First mode (PCs).



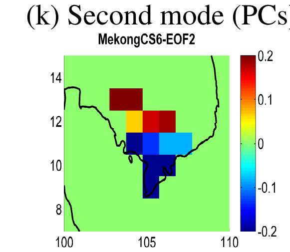
(k) Second mode (PCs).



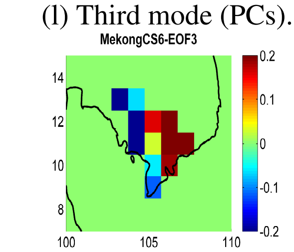
(l) Third mode (PCs).



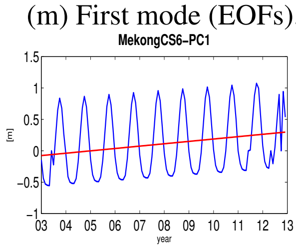
(m) First mode (EOFs).



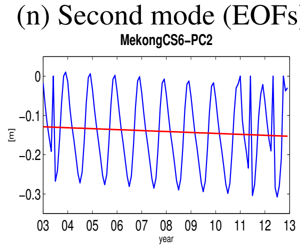
(n) Second mode (EOFs).



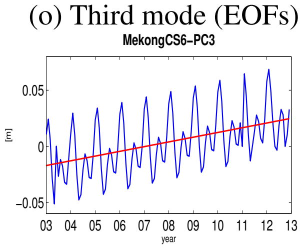
(o) Third mode (EOFs).



(p) First mode (PCs).

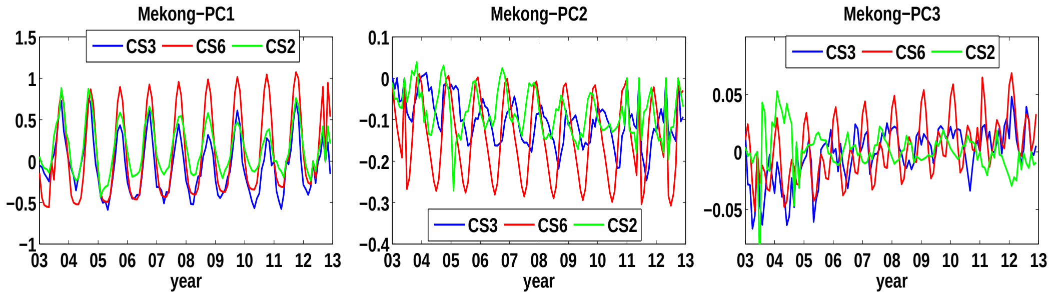


(q) Second mode (PCs).



(r) Third mode (PCs).

Figure 4.15: The spatial patterns (EOFs) (1st 3th and 5th row) and the temporal evolution (PCs) of CS2 (a, b, c, d, e, f) CS3 (g, h, i, j, k, l) and CS6 (m, n, o, q, r) data-set, normalized and centered for the years 2003-2012 with Mekong mask, first three dominant modes left to right.



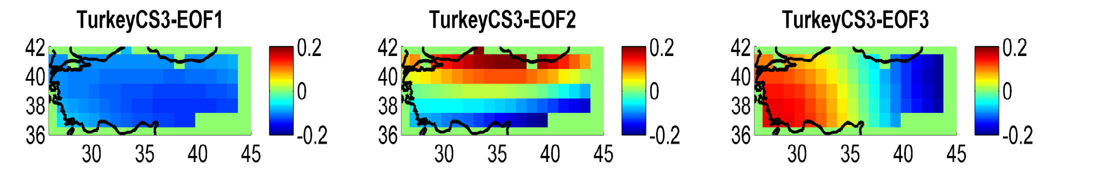
(a) First mode. (b) Second mode. (c) Third mode.

Figure 4.16: Comparison of Principal components for the first three PCA modes. The graphs belong to CS3 (blue, 350 *km* spatial res.), CS2 (green, 250 *km* spatial res.) and CS6 (red, 500 *km* spatial res.) data with the Mekong border based mask applied.

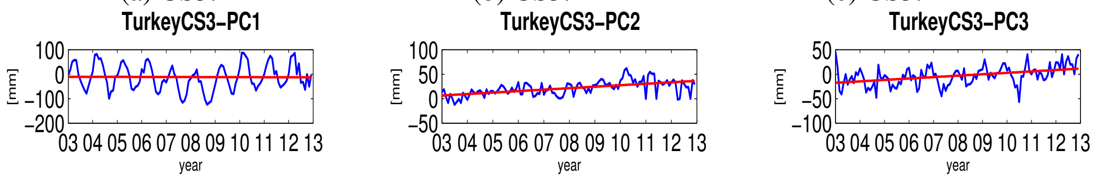
Analysis result show that CS2, CS3 and CS6 data-sets show similar trend in the first three components. However CS6(red), has higher amplitude respect to CS2 and CS3.

4.2.5 Turkey

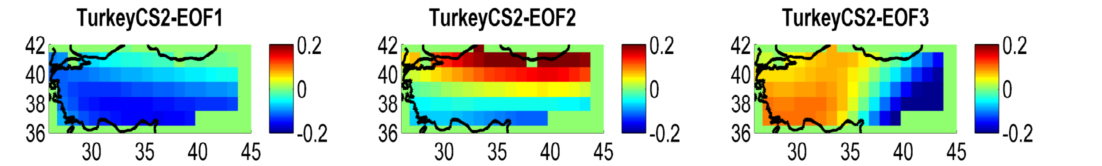
Turkey is one of the most water rich countries of the Mediterranean. The enormous population increase from 28 million (in the 1960's) to 68 million (in 2000) caused decrease of the available water resources from around 4000 m^3 to 1500 m^3 per capita/year today. Water demand in Turkey approximately has doubled in the second half of the last century [28].



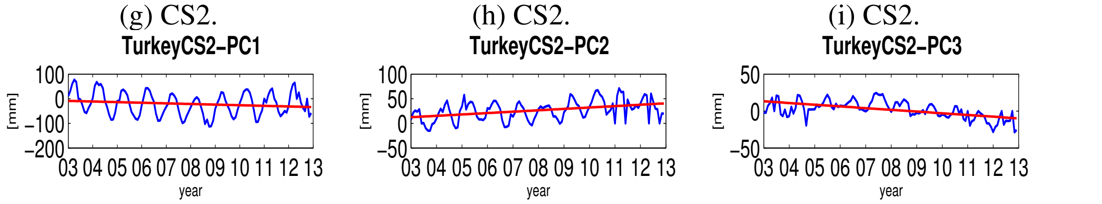
(a) CS3. (b) CS3. (c) CS3.



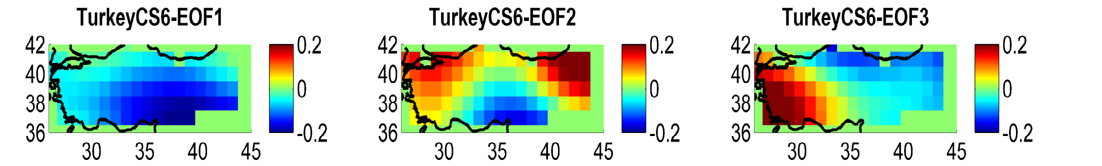
(d) CS3. (e) CS3. (f) CS3



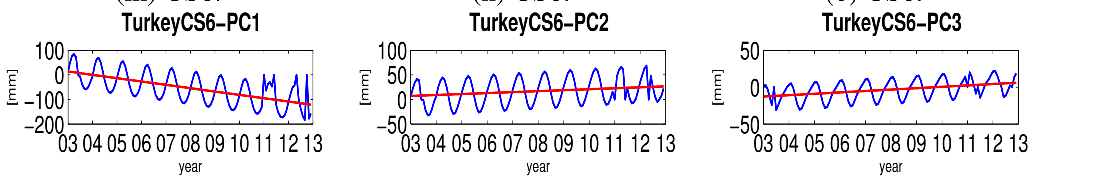
(g) CS2. (h) CS2. (i) CS2.



(j) CS2. (k) CS2. (l) CS2.



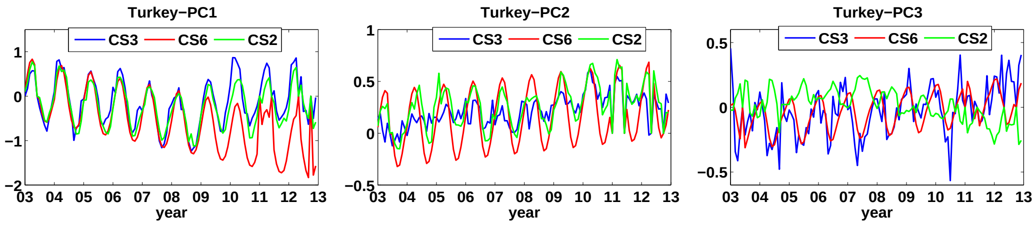
(m) CS6. (n) CS6. (o) CS6.



(p) CS6. (q) CS6. (r) CS6.

Figure 4.17: Comparison of CS2, CS3, CS6 (2003-2010) with utilizing PCA for the study region Turkey.

As seen in the figure 4.17 CS2, CS3 and CS6 data (for more information please see section 1.1) is analyzed with PCA method for Turkey and results are compared eachouter. The decreasing trend is the most significant at the output of the CS6 data in the first mode (subfigure p). As also seen spatial representation of the data (subfigure m), TWS change in the South Anatolia region contribute strongly to this deduction. In the results of PCA second modes, slight increase of TWS can be observed in the Black sea area. Evaluation of the third mode output is hard regarding to small random changes in the results.



(a) 1^{st} mode. (b) 2^{nd} mode. (c) 3^{th} mode.

Figure 4.18: Comparison of temporal changes of CS2, CS3, CS6 (2003-2010) with utilizing PCA for the study region Turkey.

4.2.6 Tigris & Euphrates

The Euphrates arises from the eastern highlands of Turkey, between Lake Van and the Black Sea, and it consists of two major tributaries, the Murat and the Karasu and enters the Syrian territory at Karkamis, downstream from the Turkish town of Birecik.

It joins to its major tributaries, the Balik and Khabur, which also originate in Turkey, and flows southeast across the Syrian plateaus before entering the Iraqi territory near Qusaybah.

The Euphrates river is 3 000 *km* long, run through the countries Turkey (1 230 *km*), the Syrian Arab Republic (710 *km*), and Iraq (1 060 *km*). Besides, 62% of the catchment area that produces inputs into the river is situated in Turkey and 38% in the Syrian Arab Republic.

28%, 17%, 40%, 15%t and just 0.03% of the lies of the Euphrates basin locates in Turkey, in Syrian Arab Republic, in Iraq, in Saudi Arabia, and in Jordan respectively. The Saudi Arabian stretch of the Euphrates dries in summer; there are no perennial rivers.

It is estimated that Turkey contributes 89% and the Syrian Arab Republic 11% of the annual flow. The remaining riparian countries contribute to the annual flow with very little amount of water [29].

The Tigris, like The Euphrates originates in eastern Turkey, flows through the country until the border city of Turkey, Cizre. From Cizre, it passes across the border of Turkey and the Syrian Arab Republic and occupies a minor area and then crosses into Iraq at Faysh Khabur.

The river is 1 850 *km* long in total, with 400 *km* in Turkey, 32 *km* on the border between Turkey and the Syrian Arab Republic and 1 418 *km* in Iraq.

12%, 0.2%, 54%,34% of the lies of the Tigris basin locates in Turkey, in the Syrian Arab Republic, in Iraq and in the Islamic Republic of Iran respectively.

Turkey provides 51%, Iraq 39%, and the Islamic Republic of Iran 10% of the annual water volume of the Tigris, but because of geographic and climatic conditions of the latter, the water cannot be used for agriculture or hydropower.

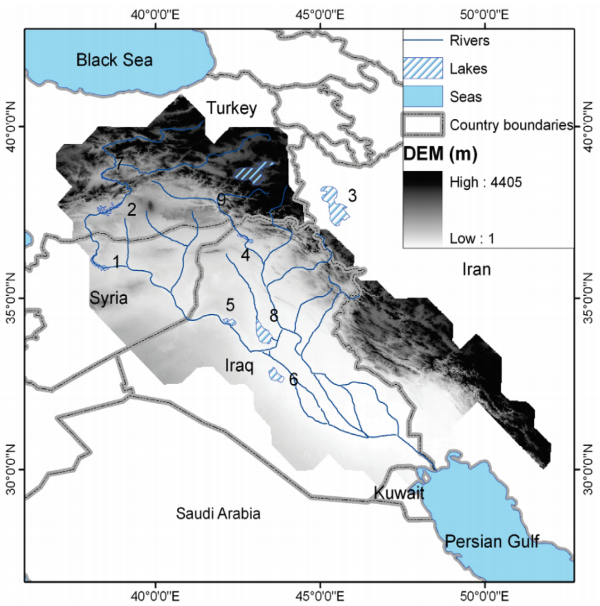


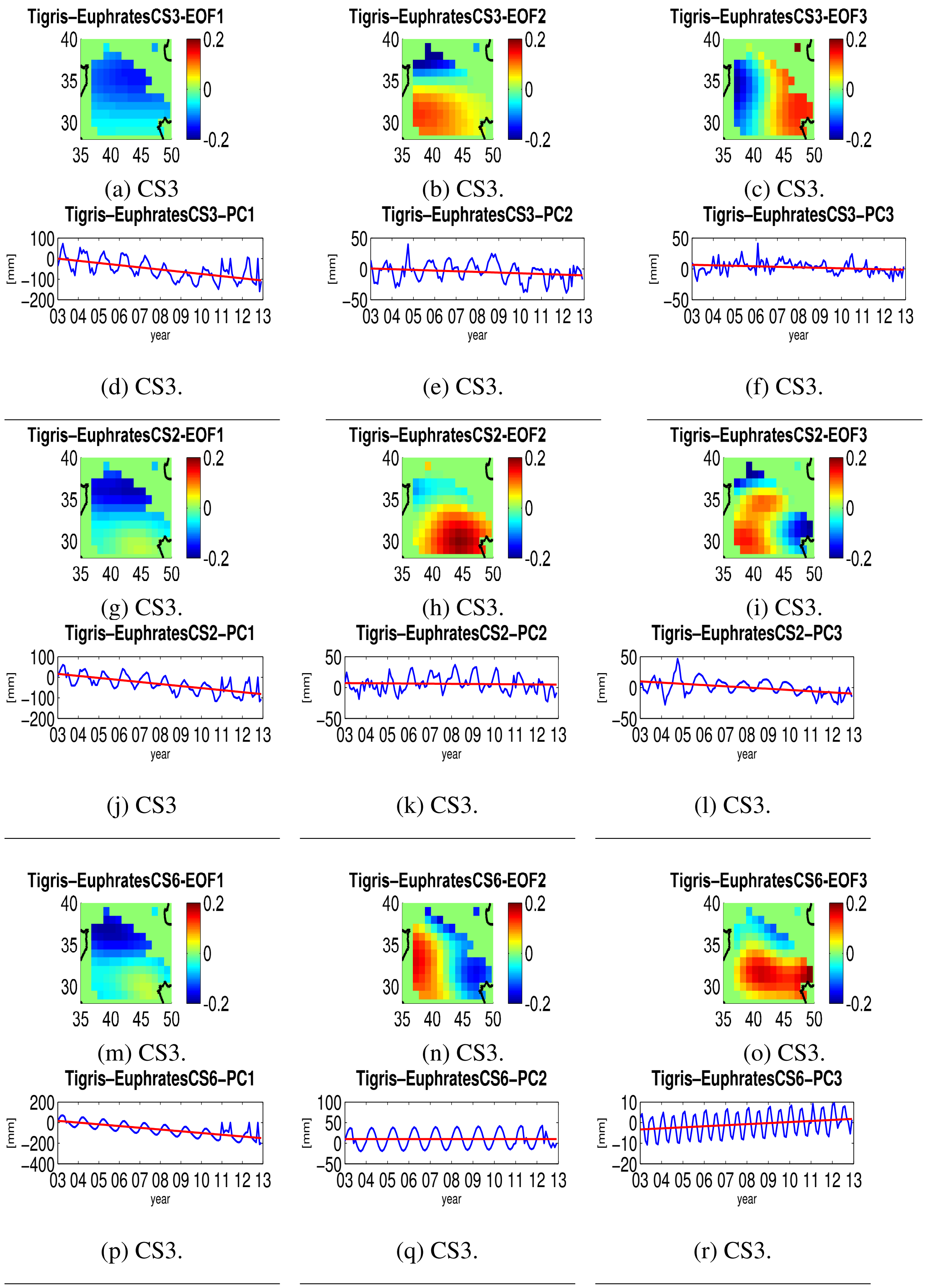
Figure 4.19: Tigris-Euphrates basin visualization [2].

Inadequacy of water supply in the Middle East and conflicts to share the existing few resources within the region is a well-known issue. And the drought which started in 2007 has made the situation even more critical and led to a dire situation. Fields lay fallow, wetland ecosystems were destroyed, and hundreds of farmers migrated to urban centers in search of employment. Water management of the Tigris-Euphrates river was every time difficult to manage because of strategical location of the river system. The paucity of hydrologic data for the region have been a big challenge. As a consequence despite its importance, there have been few basin-wide hydrological studies using observational data for the Tigris-Euphrates Basin in recent years [6].

Although previous studies in the region showed that publicly available observations of stream-flow, precipitation, and evaporation data are sparse to nonexistent, and if available, data sets are often incomplete. Besides existing classified, government-controlled data stands in need the permission and cooperation of the respective governments to access to these data. Similarly groundwater information is also constrained, with limited or no data related to water table height or annual groundwater extraction available publicly. These data scarcity and inaccessibility result in an incomplete understanding of water availability and use in this area of the Middle East. GRACE stands in the breach so to say and provides an accessible record of variations in total terrestrial water storage (defined as all of the snow, surface water, soil moisture, and groundwater) across the globe data with sufficient resolution and accuracy to benefit water management [6].

The study which examines the most of the Tigris River Basin, the upper and middle section of Euphrates River Basin, and western Iran as study area to analyze the regional trends present in the global GRACE data set from January 2003 to December 2009. Monthly groundwater storage variations as the residual of the water storage balance, after subtracting snow water equivalent, surface water, and soil moisture storage from those of total water storage observed by GRACE. Three land surface models of NASA Global Land Data Assimilation System (GLDAS) for precipitation, evapotranspiration, streamflow, soil moisture, and snow water equivalent are used as additional information. The soil moisture and snow water equivalent were combined with satellite altimetry measurements of lake and reservoir heights. The study result is addressed that, the results show an alarming rate of decrease in total water storage of approximately 27.26±0.6 mm y^{-1} equivalent water height, equal to a volume of 143.6km^3 during the course of the study period [6].

Objective of this part of the study to evaluate TWS derived from GRACE data which is post-processed with 3 different techniques. Principal Component Analysis (PCA) is use to isolate spatial and temporal signals [30]. The first three component of the PCA examine in contrast. We used 120 months, from January 2003 to December 2012, of GRACE-derived variations in TWS for the analysis. GRACE data set represented in terms of centimeters of equivalent water height at a spatial resolution of 1° x 1° and a temporal resolution of a month with some data gaps. These GRACE data-sets are named CS2, CS3 and CS6 as abbreviation within the study. Each data is a set of monthly solutions with 2003-2012 (120 months) temporal resolution and on global grid with resolution 1° x 1°. **CS2** represent the GRACE data version **release 05** of GFZ during the study, and it has **250 km** spatial resolution. **CS3** is the current version **release 05** of GFZ during the study and it has **350km** spatial resolution. **CS6** stands for the the time-series for the previous version **release 04** of GFZ during the study and it has **500km** spatial resolution. For detailed information about the data-set please check section 1.1.



(a) CS3 (b) CS3. (c) CS3.

(d) CS3. (e) CS3. (f) CS3.

(g) CS3. (h) CS3. (i) CS3.

(j) CS3 (k) CS3. (l) CS3.

(m) CS3. (n) CS3. (o) CS3.

(p) CS3. (q) CS3. (r) CS3.

Figure 4.20: Comparison of CS2, CS3, CS6 (2003-2010) with utilizing PCA for the study region Tigris & Euprates.

The first component of EOFs shows negative trend in all data-sets. As a common point, decreasing trend is most dominant at the north part of the basin. However, the data-set differs slightly to localize this trend.

Basically the basin can be divided in 3 different regions according to level of decrease in the first mode outcome. The maximum amount decrease (-0.1 to -0.2) is located at the north part of the basin and south-west part of the basin is spotted as middle effected area as a common result.

CS6 and CS2 data-set agree to mark the south-east part of the basin as less effected area (≈0) over the basin. However, the same region shows the middle decreasing trend (0 to-0.1) in the result of the CS3.

Effect of the the drought which started at 2007 is visible in all 3 data-set's PC results. Also the data-gaps within the results is contaminating the PC results starting 2011. PCA results of CS6 are relatively smooth compared to CS2, CS3. Sevan lake which located in Armenia is seen in north-east of the basin with a dominating trend.

4.3 Analyzing Effects of the Gaps in GRACE Data

Existing missing values in the data is also a concern in the application of decomposition techniques for extracting information from geophysical time series. Hence, this issue must be considered when applying the introduced PCA. 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 [31].

Generation of a good gravity field solution requires accumulation of satellite-to-satellite tracking data for about 1 month. It has to be taken into account that there are several missing consecutive weeks due to 'battery management' of the GRACE satellites starting in 2011, occurring approximately every 6 months as shown in the Table 4.2 [32] which interrupts the continuity of the time series.

Table 4.2: Missing months of GRACE data.

Month	Year	Day
January	2011	01-31
November	2011	321-345
April	2012	110-121
May	2012	122-152
October	2012	275-305
March	2013	60-90

The PCA problem in the presence of missing values may at first sight seem as easy as standard PCA, there are some important distinctions. Because of them, the situation actually resembles in many ways nonlinear modeling problems. Firstly, there is no analytical solution available for PCA with missing values because even the estimation of the data covariance matrix is nontrivial. Secondly, the optimized cost function typically has local minima, which makes finding of the global optimum more difficult. Regularization is often required in probabilistic approaches because standard PCA approaches can easily lead to over-fitting. And algorithms may have a heavy computational load especially in large-scale problems [33].

A typical method for handling missing values is so-called imputation method. In its simplest form, this means in PCA replacing the missing values in the data by the mean value of that component. A more advanced imputation method uses suitably designed iterations for replacing the missing values. The data vectors completed in this way are then used to re-estimate the sample covariance matrix in, and this iteration is continued until it converges. One other method for handling missing values in real-world data sets is to look for such a data vector that is as similar as possible to the data vector to be handled which has missing value(s), and then replace the missing values with the respective components of that most similar another data vector. A natural method is to seek for such a data vector by computing the Euclidean distances between the data vector to be handled and the other data vectors in the data set, and then select the nearest neighbor that has minimum distance. Another possibility is to compute the angles between the data vectors, and replace the missing component from the respective one of the vector that has the smallest angle to the data vector to be handled [33].

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. This approach, however, does not necessarily lead to a semi-positive defined auto-covariance matrix and consequently its corresponding EOFs are over-estimated [31].

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 re-evaluating the field and estimating new EOFs until the procedure convergences [31].

In our study, we didn't interpolate missing months and the amount of effect is omitted. The missing months are left with filled "0" as it is provided. However, in this part of the study, the affects of these gaps in the GRACE time series over the regional PCA result will be analyzed. For this purpose, part of the CS3 time series which doesn't have missing months are set apart and used as reference data. To simulate the effect of the missing months, some zero filled months are injected to the CS3 data-set. Zero filled months are chosen as January 2009, June 2009, May 2010 and October 2010 to get close the impact as real data (see in table 4.2). Afterward, Amazon, Greenland and Mekong masks are applied to simulated and reference data. PCA is applied to both results. First three mode of principle components, eigenvalues and variance for Amazon , Mekong and Greenland are compared.

It is determined that, missing months caused 1.54 %, 0.71 % and 0.22 % signal loss in Greenland, Amazon and Mekong areas relatively. It means that the larger area is, the more effectively the gaps influence the result.

4.4 Compression of In-situ Data and PCA Results

In this part of the study, PCA is applied to monthly GRACE satellite solutions to separate useful geophysical signals. The results are then validated with in-situ measurements of the Mekong basin. GRACE observations are used to estimate variations over the Mekong basin at an inter-annual scale from July to November between 2008 and 2011.

The annual variations from GRACE are compared to the annual TWS of the Mekong basin which is the sum of in-situ observations (surface water and ground water). In-situ observations are considered as the reference in the following analysis. Centered and normalized in-situ daily observations over the period from July to November for each year are illustrated in figure 4.21.

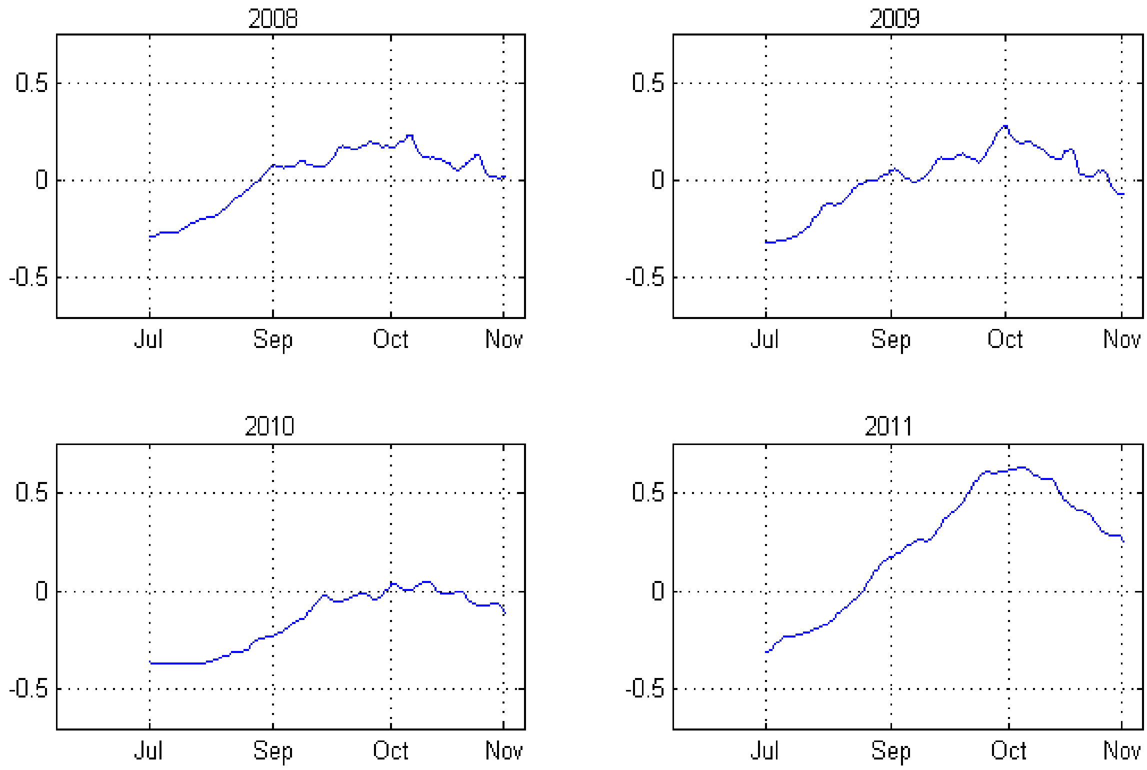


Figure 4.21: Normalized and centered in-situ data of Mekong delta covering from 01. July to 30. November annual time period for year 2008 (left top), 2009 (right top), 2010 (left bottom) and 2011 (right bottom).

To center the attention on Mekong among the global solutions, the Mekong basin region border mask is applied to CS3 GRACE solutions from 2008 to 2011. After these integral observations are separated and time coefficients (Principle components, PCS) are calculated. The most dominant mode for the time period between July and November (5 months) for each year is calculated. Results are compared with in-situ data and visualized in figure 4.22.

Principle component analysis (PCA) first mode of CS3 and in-situ monthly means exhibit very similar temporal patterns in each year. The correlation coefficients between two time series in all years surpass 0.95.

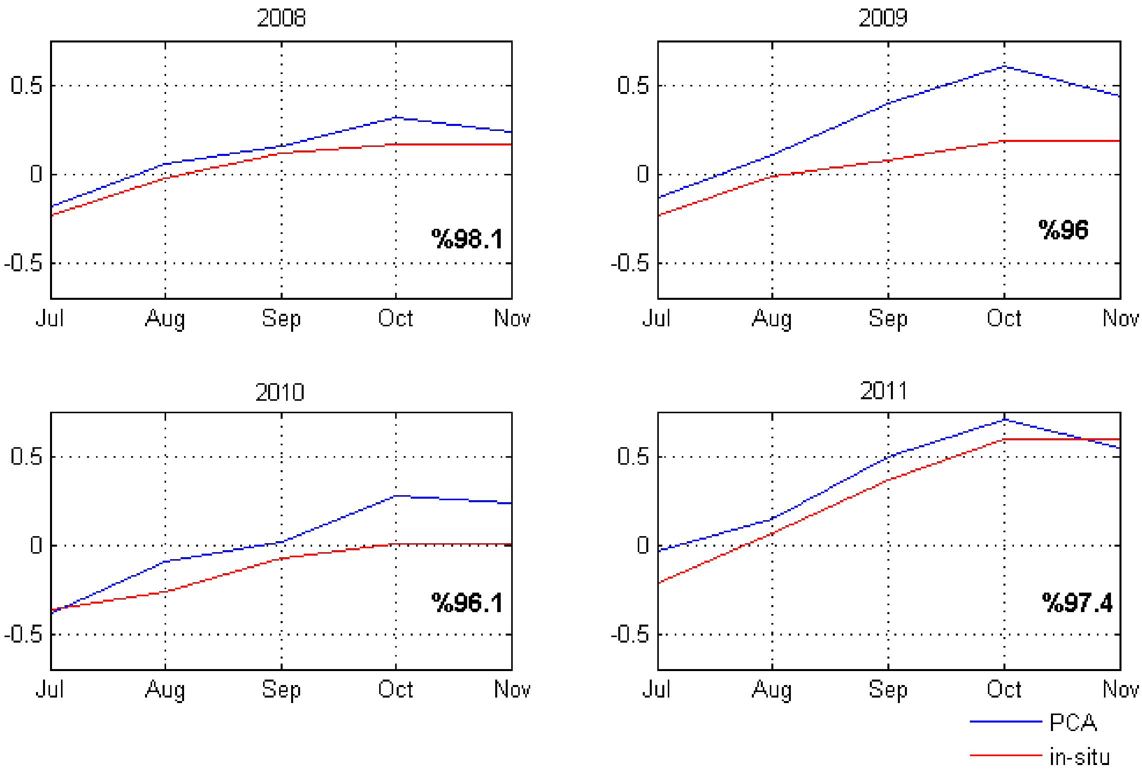


Figure 4.22: Most dominant mode of the time coefficients (principal components) is derived from CS3 GRACE solution (blue) and monthly average of in-situ data (red) of Mekong delta covering from 01.July to 30.November annual time period for year 2008 (left top), 2009 (right top), 2010 (left bottom) and 2011 (right bottom).

The results demonstrate that, GRACE data and in-situ data show a good agreement with high correlations, like 98.1 %, 96 %, 96.1 % and 97 % for 2008, 2009, 2010 and 2011 years respectively. However, despite the high correlations, for the hydrology there is a clear phase shift for the annual signal with GRACE observing the extrema about 1 month than in-situ data (similar shift is also observed in the previous study [34].). The

Table 4.3: Correlation coefficients between in-situ and GRACE data with different resolutions.

Resolution of GRACE data	2008	2009	2010	2011
CS2 250 *km*	0.9706	0.9651	0.9705	0.9923
CS6 350 *km*	0.9812	0.9597	0.9616	0.9746
CS3 500 *km*	0.9706	0.9651	0.9705	0.9923

in-situ and GRACE data comparison is repeated with the border based Mekong mask applied to CS2 and CS6 GRACE data to test the correlation accordingly.

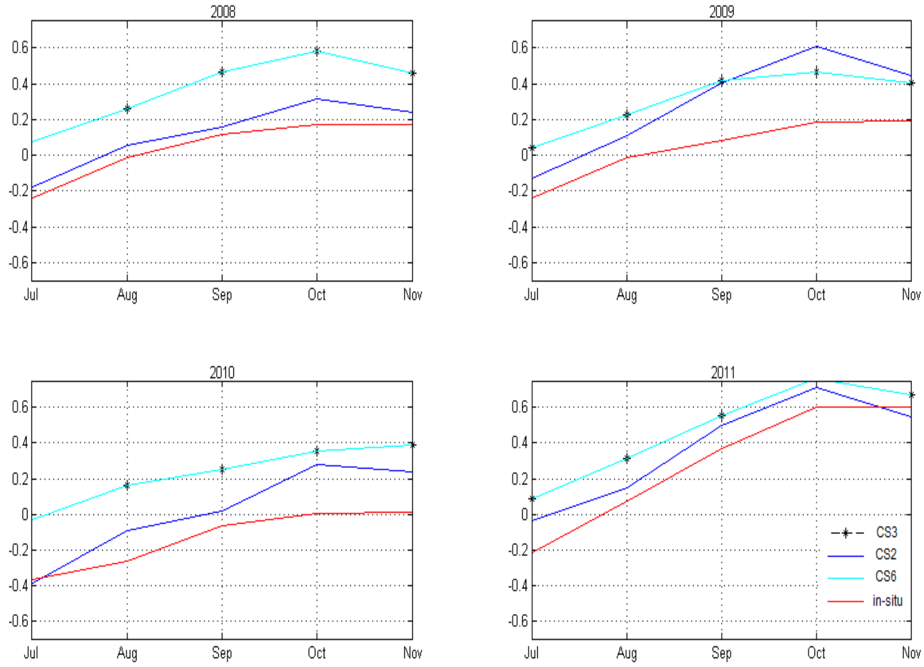


Figure 4.23: Comparison of most dominant modes of the time coefficients (principal components) are derived from CS3 (black), CS6 (turquoise), CS2 (blue) GRACE solutions and monthly average in-situ data (red) of the Mekong delta covering from 01. July to 30. November annual time period for year 2008 (left top), 2009 (right top), 2010 (left bottom) and 2011 (right bottom).

Eventually, the highest correlation occurs in year 2011 between CS3, CS2 and in-situ data with 0.9923. CS3 and CS2 have same correlations in every year. The Lowest correlation revels between CS6 data and in-situ data in year 2009 with 0.9597. The results depict that consistency of GRACE and in-situ data change respect to corresponding year. However, resolution of GRACE doesn't account for significant effect to the correlation. However it is evident that the GFZ release 05 has implemented some improvement over release 04.

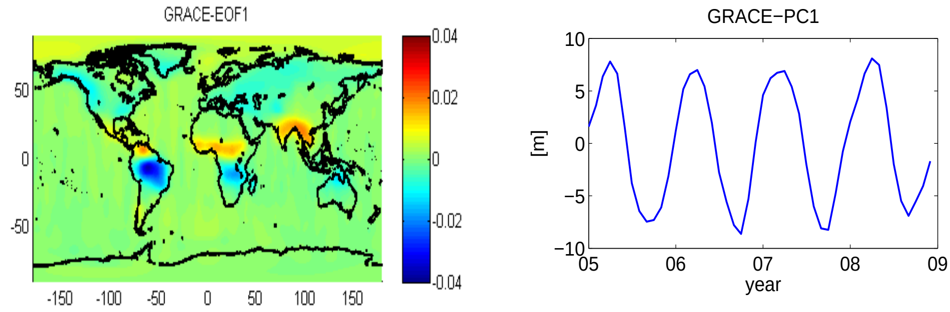
4.5 Comparison of WGHM and GRACE Data with Using PCA

As soon as sufficiently long GRACE solutions were released by the formal official providers (CSR, JPL, GFZ), various regional studies for validating these products were made. Comparison with global hydrology model outputs and surface measurements revealed acceptable agreements between GRACE-derived and model changes of continental water storage versus time, especially at seasonal scale [11]. This WaterGAP model, has been developed at the Center for Environmental Systems Research at the

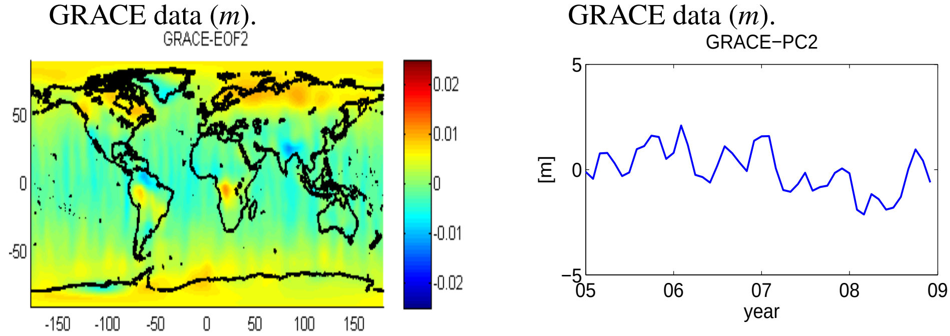
University of Kassel in Germany in cooperation with the National Institute of Public Health and the Environment of the Netherlands. The aim of the model is to provide a basis to compare and assess current water resources and water use in different parts of the world, and to provide an integrated long-term perspective of the impacts of global change on the water sector [35].

The Water GAP global Hydrology Model (WGHM) has been explicitly designed to simulate river discharge for water resources assessments. WGHM simulates the water balance on $0.5^\circ \times 0.5^\circ$ grid by using simplified conceptual algorithms to represent the hydrological process. Climate forcing is by monthly data, being decomposed into daily values corresponding to the modeling time step. Water storage in the snow pack, rooted soil zone, ground water, on vegetation surfaces, and as surface storage in rivers, lakes, and wetlands is accounted for. The model has been tuned to match observed mean annual river discharge in 724 river basins worldwide, covering 50% of the global land area (excluding Greenland and Antarctica) [36].

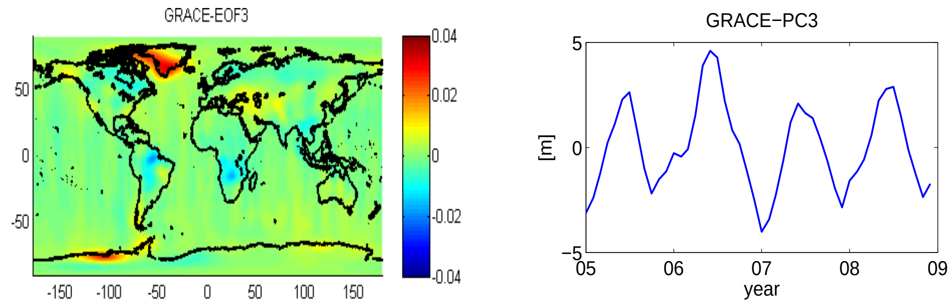
The data which are used during following analysis contain a monthly time series of filtered gridded values (Gauss filter with 400 *km* radius, years 2005-2008) on a $1^\circ \times 1^\circ$ geographical grid stored in the data matrix Y.
Next, the spatial patterns (EOFs) and the temporal evolution (PCs) are both calculated from GRACE and WaterGAP Hydrology Model (WGHM) data set . The temporal evolution for the first three modes, are visualized.



(a) The EOFs of first mode from the GRACE data (m).
(b) The PCs of the first mode from the GRACE data (m).



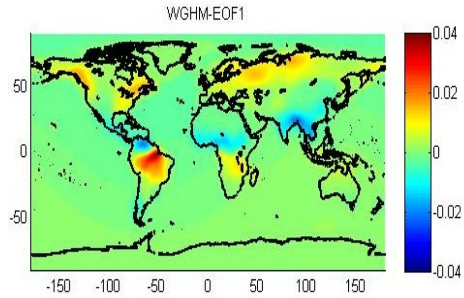
(c) The EOFs of the second mode from the GRACE data (m).
(d) The PCs of the second mode from the GRACE data (m).



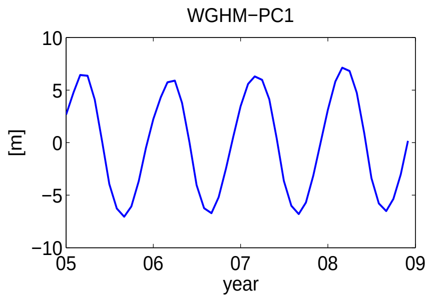
(e) The EOFs of the third mode from the GRACE data (m).
(f) The PCs of the third mode from the GRACE data (m).

Figure 4.24: The spatial patterns (EOFs) and the temporal evolution (PCs) of GRACE data of the 2005-2008 years for three modes.

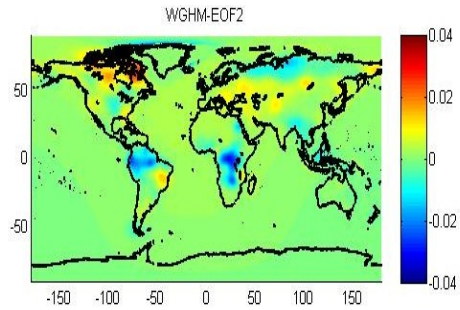
As seen in the Figure 4.24, the first component of GRACE (a), (b) data clearly determines the large water mass anomalies which are observed in the tropical regions like Amazon, the Ganges and Mekong basins and this annual signals act sinusoidally. The second and third components, as shown in the Figure (c), (d), (e) and (f), derive less dominant behavior (like semi-annual signal) and trends, linear increase and decrease relatively, can be noticed. Besides, in consequence of the rainfall in Congo (2007) the dramatic increase at PCs of second mode for the year 2007 is obtained.



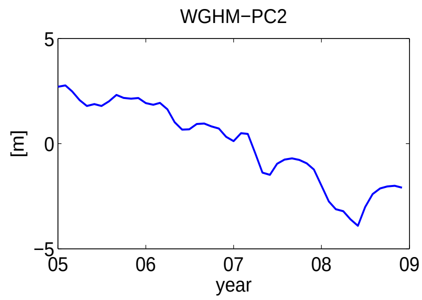
(a) The EOFs of the first mode from the WGHM data (m).



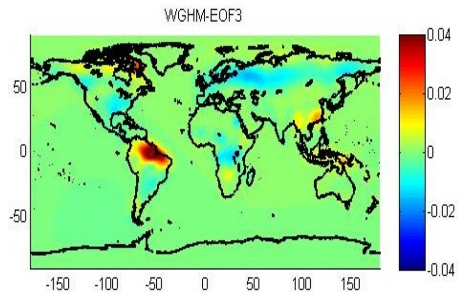
(b) The PCs of the second mode from the WGHM data (m).



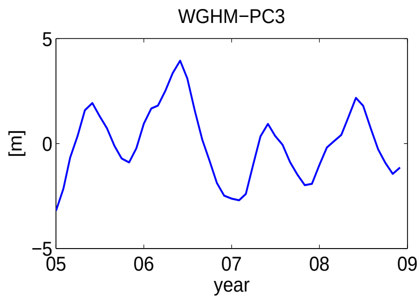
(c) The EOFs of the second mode from the WGHM data (m).



(d) The PCs of the second mode from the WGHM data (m).



(e) The EOFs of the third mode from the WGHM data(m).



(f) The PCs of the third mode from the WGHM data (m).

Figure 4.25: The spatial patterns (EOFs) and the temporal evolution (PCs) of WGHM data of the 2005-2008 years for the first three dominant modes.

As seen in the Figure 4.25, the first component of the WaterGAP Hydrology Model (WGHM) data detects the annual signal like the first component of GRACE data. However the second and third component of the data don't have that much similarity with the second and third mode of the GRACE data. For instance, the second mode of GRACE data have an increasing trend but second component of the WGHM has a deceasing trend. This kind of discrepancies are expected when a model and real measurements are compared. In such positions, GRACE real data may be used for calibration of water models.

4.5.1 Reconstruction

A decision on how many modes are actually required to reconstruct data sufficiently may be based upon statistical criteria. The relation between truncated mode and reconstruction variance will be described fundamentally.

Let $k < n$ be the number of retained modes (those associated with the largest singular value), the reconstruction of data matrix is:

$$Y \cong Y_k = \overline{P_k}\Lambda_k E_k^t \tag{4.6}$$

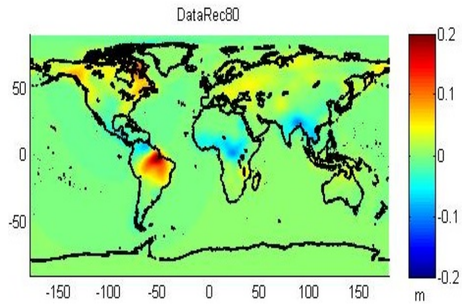
$$EE_k^t = I \tag{4.7}$$

$$\overline{P_k P_k}^t = I \tag{4.8}$$

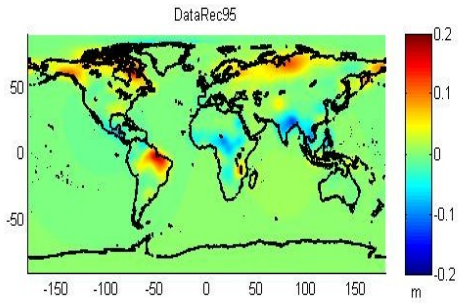
where E contains the eigenvectors of Y normalized to unit length in its columns, $\overline{P}$ contains normalized PCs and Λ_k is diagonal and holds the singular values of the data matrix ordered according to magnitude [37]. The fraction of the signal variability reconstructed by k modes according to

$$var_k = \frac{\sum_{j=1}^{\overline{k}} \lambda_j}{\Delta^2} \text{ with total variance } \Delta^2 = \sum_{j=1}^{\overline{p}} \lambda_j. \tag{4.9}$$

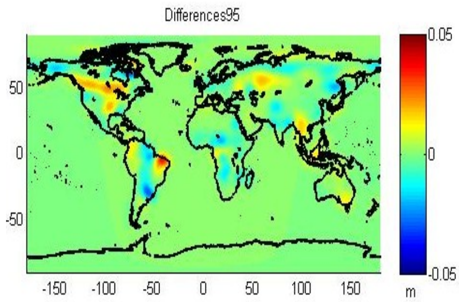
As result of the global signal construction showing specific variance of WGHM, it is determined that, 8 EOFs and 3 EOFs are necessary to reconstruct signal from the global WGHM time series with 95 % and 80 % variance respectively.



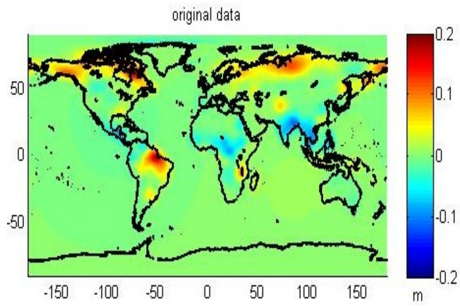
(a) Reconstructed signal with 80 % variance (m).



(b) Reconstructed signal with 95 % variance (m).



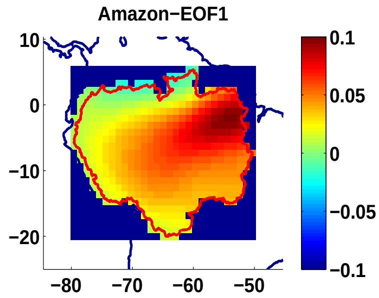
(c) Difference between original and 95 % reconstructed signal (m).

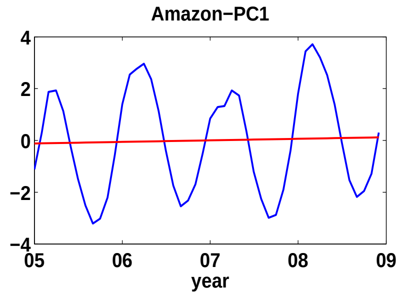


(d) Original data (m).

Figure 4.26: Difference between reconstructed and original signal with 95 % and 80 % variance in m.

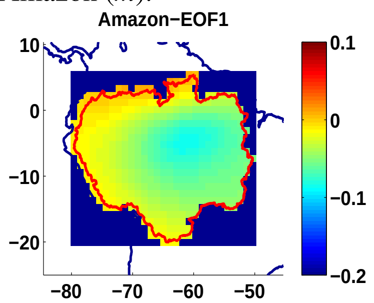
Figure 4.26 illustrates truncation at 8^{th} mode (95 % reconstructed signal) and 3^{rd} mode (80 % reconstructed signal) alter $\pm 5cm$. After it, the global WGHM time series is used and the data of Amazon and Orinoco basins are cut out of the entire data with using coordinates of the regions within polygon function. Water storage change for each region between 2005-2008 is calculated and performance of PCA is tested locally and as a result, the first EOF and PCs are visualized.

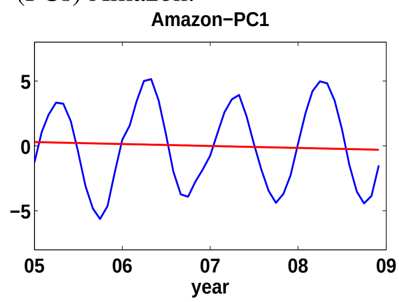




(a) The First temporal evolution (PCs) in Amazon (m).

(b) The First mode, temporal evolution (PCs) Amazon.

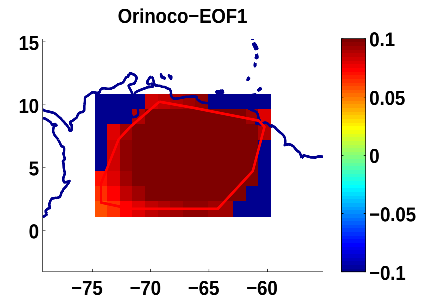


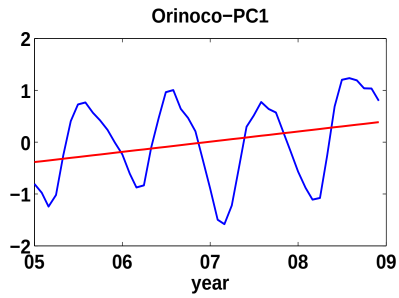


(c) The First temporal evolution (PCs) in Amazon (m).

(d) The First mode, temporal evolution (PCs) Amazon.

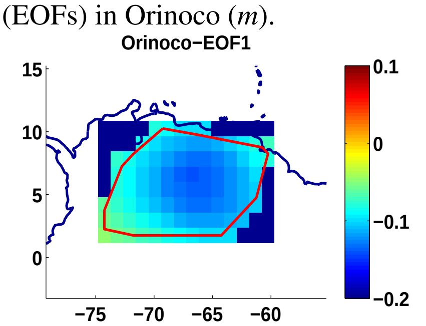
Figure 4.27: Principle component analysis for regional solutions, first mode of Amazon.

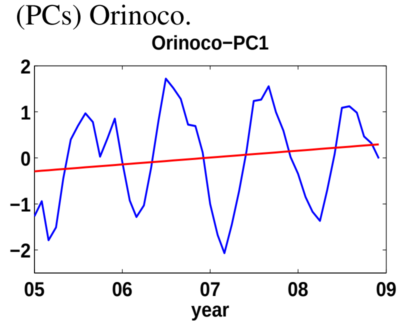




(a) The First mode, spatial evolution (EOFs) in Orinoco (m).

(b) The First mode, temporal evolution (PCs) Orinoco.





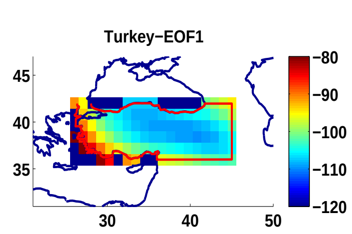
(c) The First mode, spatial evolution (EOFs) in Orinoco (m).

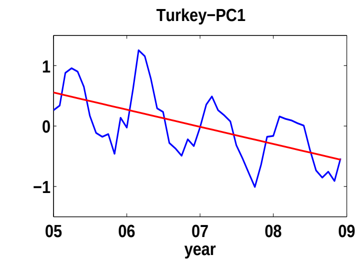
(d) The First mode, temporal evolution (PCs) Orinoco.

Figure 4.28: Principle component analysis for regional solutions, first mode of Orinoco.

A decision on how many modes are actually required to reconstruct the Amazon basin signal with 95 % is made and it turns out that 3 modes are adequate to achieve the desired variance. Moreover, it can be seen that, to reconstruct signal from the global WGHM time series with 95 % of the signal just mode 1 is sufficient for the Orinoco basin. In conclusion, the regional results are compared with the global results which are examined in subsection 4.5.1 and it is revealed that, high signal variability can be achieved with a lower number of modes in local reconstructions than in global ones. The previous study examined linear trends of large-scale TWS and groundwater storage variations in Turkey during the period from March 2003 to March 2009. The satellite-based estimate of TWS and groundwater revealed significant decreasing trends in central Anatolia reaching a rate of 4 cm/year in the southern part of central Anatolia. Results of the study also addressed that decreasing trends of TWS observed by GRACE in the southern part of central Anatolia were largely explained by the decreasing trends of groundwater variations which were confirmed by the limited available well groundwater level data in the region [38].

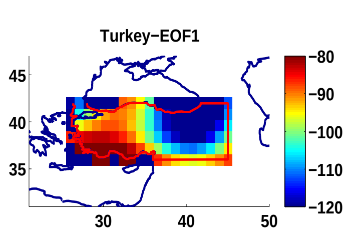
The goal of our study to examine the most dominating trend of TWS derived from GRACE and compare the results with WGHM. PCA results of these two different data-sets are calculated. Spatial and temporal variation over the area is visualized.

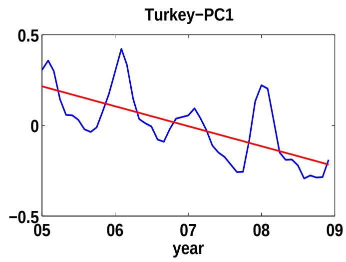




(a) The First mode, spatial evolution (EOFs) of Turkey, GRACE data (*mm*).

(b) The First mode, temporal evolution (PCs) of Turkey, GRACE data.

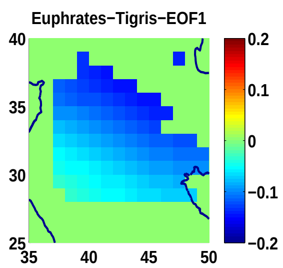


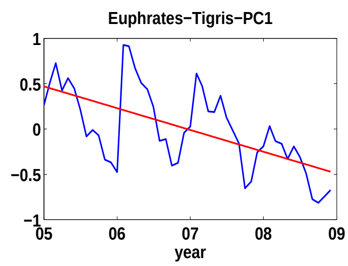


(c) The First mode, spatial evolution (EOFs) of Turkey, WGHM (*mm*).

(d) The First mode, temporal evolution (PCs) of Turkey, WGHM.

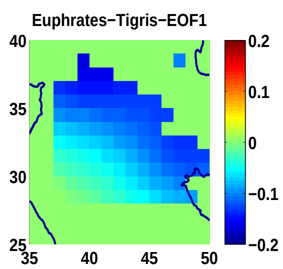
Figure 4.29: Principle component analysis of GRACE and WGHM data (2005-2008) for region Turkey. Comparison of first mode results.

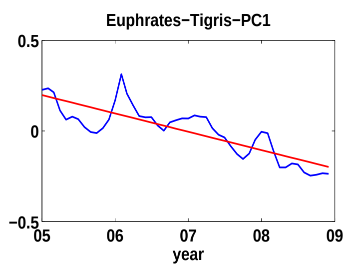




(a) The First mode, spatial evolution (EOFs) of Tigris & Euphrates river system with GRACE data (*m*).

(b) The First mode, temporal evolution (PCs) of Tigris & Euphrates river system with GRACE data.





(c) The First mode, spatial evolution (PCs) of Tigris & Euphrates river system with WGHM (*m*).

(d) The First mode, temporal evolution (PCs) of Tigris & Euphrates river system with WGHM.

Figure 4.30: Principle component analysis of GRACE and WGHM data (2005-2008) for Tigris & Euphrates river system. Comparison of first mode results.

A dominant trend at the south Anatolia region of Turkey due to part of the Tigris-Euphrates river system observed. As a second step, whole river system is examined as an extra study area. More detailed information about the Tigris-Euphrates river system can be found in subsection . The region Tigris& Euphrates shows a negative trend in the both data-sets as seen in the figure 4.30. Small variation between two data-set can be seen in the south-west part of the basin. In both results, it can be noticed that Sevan lake had deceasing as well as the rest of the basin. When the trend of the region is examined, effect of the reservoirs should be also considered. A previous study points out that, impact of nonuniform distribution of water storage distribution on GRACE TWS changes as basin-wide averages should be considered. GRACE can estimate basin-wide average over or under up to a factor ≈ 2 depending of reservoir location and area extent. Even though reservoirs mostly cover less than 1% of the basin area, their spatial extent may be unsolved by GRACE and the reservoir storage may dominate some changes in the basin. The reservoir storage accounts for 10% in the Tigris-Euphrates since the reservoirs are used to mitigate droughts and buffer against climate extremes. As a result of that, high influence on interannual timescales can occur. It is also stated that TWS storage during 2007-2009 drought in the Tigris-Euphrates basin measured by GRACE was $\approx 90km^3$. Despite this impact on GRACE reached $45km^3$ 50% of GRACE trend, the volume of actual reservoir storage which is determined from satellite altimetry data was limited to 27 km^3. As it is demonstrated, during the evaluation of the GRACE especially when separating groundwater storage from GRACE total water storage (TWS) changes [2].

5. CONCLUSIONS

The performance analysis of the Gaussian filtering shows that increasing the half-width of the filter decreases the amplitude of the stripes which is present in the dataset. Nonetheless, since it is an isotropic filter, it is incapable to remove non-isotropic noises (see section 2.1).

The truncation degree/order of spherical harmonic synthesis in spectral domain analysis results clearly show that the GRACE-like data corresponds with the reference signal better at degree/order 90 than 50. Improvements in the region borders demonstrate the recovered leaked signal. Tigris/Euphrates region is dominated by leakage of its surroundings.

For an upgraded regional analysis, special threshold masks are defined for Amazon and Greenland in Chapter 4.1. These masks are applied to the GRACE data for quality assessment.

A threshold of 0.15 and up to 50 degree and order yields best results for each of the regions. Expanding masks according to the nearest-neighbor interpolation method did not work well with Amazon, Mekong and Tigris/Euphrates regions, but Greenland area indicated increasing signal coverage.

Applying PCA GRACE data-sets CS2, CS3 and CS6 regionally with using border based masks for Amazon, Greenland, Mekong, Turkey and Tigris/Euphrates and calculated masks for the regions except for Mekong indicates that the generated masks via threshold has been increasing the covered signal percentage for each data set (see Table 4.14).

The most significant improvement of sıgnal lost is accomplished with the Amazon mask for the CS3 data-set where 4.55 % more signal is caught. Also with the Greenland mask, the CS2 signal coverage is enhanced by 0.96 %. Although, the PCA results show that, border based and designed masks have graphed same characteristic in first three modes.

However the amplitude of the signal for each mode has systematically increased with threshold based masks. This result could be evaluated as recovery of the dampening effect.

The effect of gaps in the time series to the PCA results is tested regionally for Amazon, Mekong and Greenland by applying border based masks to the CS3 time series. Artificial gaps are injected into the time series between 2003-2010 for the months January 2009, June 2009, May 2010 and October 2010.

It is determined that, missing months caused 1.54 %, 0.71 % and 0.22 % signal loss in Greenland, Amazon and Mekong respectively. It seems that the larger area is, the more significant is the effect.

PCA results of GRACE data are validated with in-situ measurements of the Mekong basin between 01^{st} July and 30^{th} of November for the years 2008, 2009, 2010 and 2011. Comparing results it can be derived that the consistency of GRACE with in-situ data changes with respect to corresponding year.

Eventually, the highest correlation occurs at year 2011 between CS3 data and CS2 data and in-situ data with 0.9923. These data have same correlations in years 2009 and 2010 as well. The lowest correlation reveals between CS6 data and in-situ data in year 2009 with 0.9597. This result also demonstrates the improvement of GRACE release 05 over GRACE release 04.

The PCA method has been used to extract modes of variability from GRACE and model based TWS time series. Performance analysis of PCA in comparison to WGHM data has shown, that the results of the first mode are substantially parallel.

The amplitudes of WGHM data are lower than those of PCA because WGHM does not include all water resources over the regions. Using these results, WGHM signals are reconstructed as inverse operator of PCA with specific variances globally and regionally (for Amazon, Orinoco, Turkey and Tigris/Euphrates) in chapter 3.

To regenerate signal with 95 % variability, global results require up to mode 8. However, this variance is achieved with 3 modes for Amazon and mode 1 for Orinoco. In case the expected variability is diminished to 80 %, global reconstruction is acquired with mode 3.

Subsequently, global and regional results are compared each other. It is demonstrated that reconstruction of local regions needs lower number of PCA modes than global reconstruction.

In conclusion, the study depicts that PCA is a useful tool for interpreting GRACE and other geophysical data and model outputs, with extracting behaviors of signals and expressing them in modes. By the help of PCA, quality assessment of masks, effects of gaps in time series, comparison of GRACE data with WGHM and in-situ tide gauges are fulfilled in an effective way.

REFERENCES

[1] **Scholz, M.**, 2015, PCA - Principal Component Analysis, http://www.nlpca.org/pca_principal_component_analysis.html, accessed: 2015-03-05.

[2] **Longuevergne, L.**, **Wilson, C.R.**, **Scanlon, B.R. and Crétaux, J.F.**, 2013. GRACE water storage estimates for the Middle East and other regions with significant reservoir and lake storage, *Hydrology and Earth System Sciences*, **17(12)**, 4817–4830, http://www.hydrol-earth-syst-sci.net/17/4817/2013/.

[3] **Güntner, A.**, 2010. Continental Water Storage Variations from GRACE Time-Variable Gravity Data, **F.M. Flechtner**, **T. Gruber**, **A. Güntner**, **M. Mandea**, **M. Rothacher**, **T. Schöne and J. Wickert**, editors, System Earth via Geodetic-Geophysical Space Techniques, Advanced Technologies in Earth Sciences, Springer Berlin Heidelberg, pp.369–375, http://dx.doi.org/10.1007/978-3-642-10228-8_31.

[4] **Flechtner, F.**, 2010. Improved GRACE Level-1 and Level-2 Products and Their Validation by Ocean Bottom Pressure, **F.M. Flechtner**, **T. Gruber**, **A. Güntner**, **M. Mandea**, **M. Rothacher**, **T. Schöne and J. Wickert**, editors, System Earth via Geodetic-Geophysical Space Techniques, Advanced Technologies in Earth Sciences, Springer Berlin Heidelberg, pp.95–104, http://dx.doi.org/10.1007/978-3-642-10228-8_8.

[5] **Dahle, C.**, **Flechtner, F.**, **Gruber, C.**, **Koenig, D.**, **Koenig, R.**, **Michalak, G. and Neumayer, K.H.**, 2013, GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 0005 : revised edition, January 2013, http://dx.doi.org/10.2312/GFZ.b103-1202-25.

[6] **Voss, K.A.**, **Famiglietti, J.S.**, **Lo, M.**, **Linage, C.**, **Rodell, M. and Swenson, S.C.**, 2013. Groundwater depletion in the Middle East from GRACE with implications for transboundary water management in the Tigris-Euphrates-Western Iran region, *Water Resources Research*, **49**, 904–914.

[7] **Hofmann-Wellenhof, B. and Moritz, H.**, 2006. Physical Geodesy, Springer.

[8] **Geomonitoring, G.S..G. and Field, G.**, 2013, Gravity Recovery and Climate Experiment-Follow-On (GRACE-FO) Mission, http://www.gfz-potsdam.de/grace/, accessed: 2014-01-05.

[9] **Kusche, J.**, **Eicker, A. and Forootan, E.**, 2011, Analyis Tools for GRACE and Related Data Sets Theoretical Basis, Summerschool Global Hydrological Cycle, lecture Notes.

[10] **Kusche, J.**, **Schmidt, R.**, **Petrovic, S. and Rietbroek, R.**, 2009, Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model.

[11] **Frédéric, F. and Guillaume, R.**, 2012, Contribution of GRACE Satellite Gravimetry in Global and Regional Hydrology, and in Ice Sheets Mass Balance, Water Resources Management and Modeling.

[12] **Sean Swenson& John Wahr**, 2006. Post-processing removal of correlated errors in GRACE data, *Geophysical research letters*, **33**.

[13] **Schmidt, R.**, **Schwintzer, P.**, **Flechtner, F.**, **Reigber, C.**, **Güntner, A.**, **Döll, P.**, **Ramillien, G.**, **Cazenave, A.**, **Petrovic, S.**, **Jochmann, H. and Wünsch, J.**, 2006. GRACE observations of changes in continental water storage, *Global and Planetary Change*, **50**, 112–126.

[14] **J. L. Chen, C. R. Wilson, J.S.F..M.R.**, 2007. Attenuation effect on seasonal basin-scale water storage changes from GRACE time-variable gravity, *Journal of Geodesy*, **81**, 237–245.

[15] **Molenaar, J.W..M.**, 1998. Time variability of the Earth's gravity field:Hydrological and oceanic effects and their possible detection using GRACE, *Journal of geophysical research*, **103**, 30,205–30,229.

[16] **Wilks, D.**, 2006. Statistical Methods in the Atmospheric Sciences, volume 91 of *International Geophysics Series*, Elsevier, Department of Earth and Atmospheric Sciences Cornell University, second edition.

[17] **Holand, S.M.**, 2008. Principle components analysis (PCA), Technical ReportGA 30602-2501, Department of Geology, University of Georgia.

[18] **Swenson, S. and Wahr, J.**, 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *Journal of Geophysical Research*, **107**.

[19] **Laurent Longuevergne, Bridget R. Scanlon, C.R.W.**, 2010. GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA, *Water Resources Research*, **46**, 105–134.

[20] **Velicogna, I. and Wahr, J.**, 2013. Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the GRACE satellite data, *Geophysical Research Letters*, **40**, 3055–3063.

[21] **Baur, O.**, **Kuhn, M. and Featherstone, W.E.**, 2009. GRACE-derived ice-mass variations over Greenland by accounting for leakage effects, *Journal of Geophysical Research (Solid Earth)*, **114(B13)**, 6407–+.

[22] **Horwath, M. and Dietrich, R.**, 2006. Errors of regional mass variations inferred from GRACE monthly solutions, *Geophysical Research Letters*, **33**, 7502–+.

[23] **Klees, R.**, **Liu, X.**, **Wittwer, T.**, **Gunter, B.**, **Revtova, E.**, **Tenzer, R.**, **Ditmar, P.**, **Winsemius, H. and Savenije, H.**, 2008. A Comparison of Global and Regional GRACE Models for Land Hydrology, *Surveys in Geophysics*, **29(4-5)**, 335–359.

[24] **Bonin, J. and Chambers, D.**, 2013. Uncertainty estimates of a GRACE inversion modelling technique over Greenland using a simulation, *Geophysical Journal International*, **194**, 212–229.

[25] **Lindsey, R.**, 2002, Introduction to the Large-scale Biosphere-Atmosphere Experiment in Amazonia, http://earthobservatory.nasa.gov/Features/LBA/, accessed: 2014-01-05.

[26] **Rasmussen2013, R.O.**, 2013, Greenland, http://www.britannica.com/EBchecked/topic/245261/Greenland#toc2834.

[27] **Tom Cochrane, Mauricio Arias, T.P..T.D.D.**, 2010, Mekong Basin, http://mekongriver.info/mekong-basin, accessed: 2014-01-05.

[28] **unknown**, 2015, Fresh water resources Turkey Present situation in Turkey, http://www.climateadaptation.eu/turkey/fresh-water-resources/, accessed: 2015-03-05.

[29] **AQUASTAT website, F. and of the United Nations (FAO), A.O.**, 2009, Euphrates-Tigris Basin, http://www.fao.org/nr/water/aquastat/basins/euphrates-tigris/index.stm, accessed: 2015-03-05.

[30] **Banerjee, C. and Nagesh Kumar, D.**, 2014. Identification of prominent spatio-temporal signals in GRACE derived terrestrial water storage for India, *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **XL-8**, 333–338.

[31] **Forootan, E.**, 2014. Statistical Signal Decomposition Techniques for Analyzing Time-Variable Satellite Gravimetry Data, Ph.D. thesis, Landwirtschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

[32] **Qu, Z.**, 2014, GRACE missing months, http://grace.jpl.nasa.gov/data/GraceMonths/, accessed: 2014-12-05.

[33] **Karhunen, J.**, 2011, Robust PCA methods for complete and missing data, Aalto University School of Science1, Dept. of Information and Computer Science, accessed: 2014-02-01.

[34] **Roberts, S. and Jackson, J.**, 1991. Active normal faulting in central Greece: an overview, *Geological Society Special Publication*, **56(1)**, 125–142.

[35] **unknown**, An Overview of the WaterGAP model, http://www.usf.uni-kassel.de/watclim/pdf/watergap_model.pdf, accessed: 2014-05-05.

[36] **Petra Döll, Frank Kaspar, B.L.**, 2003. A global hydrological model for deriving water availability indicators: model tuning and validation, *Journal of Hydrology*, **270**, 105–134.

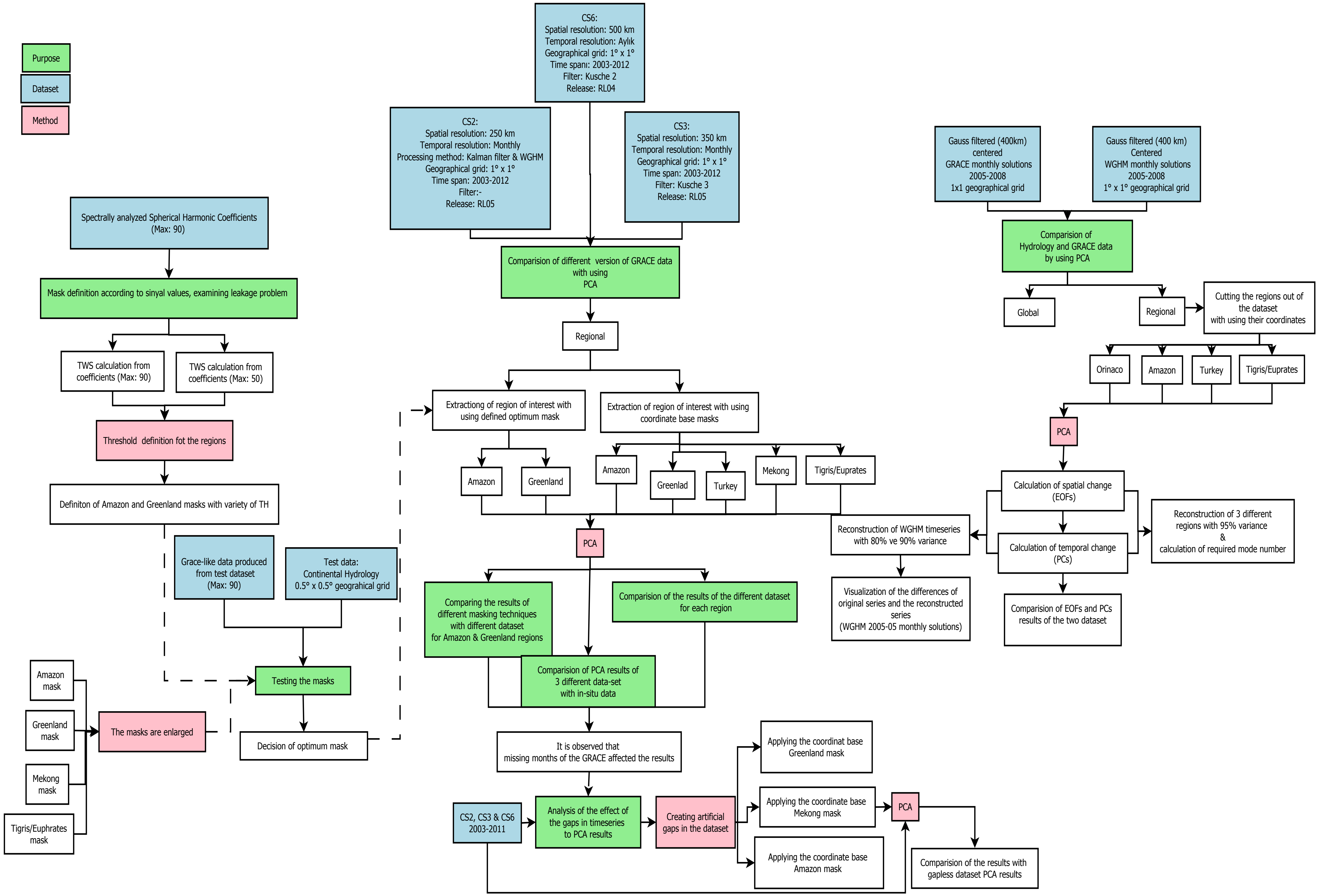
[37] **Foroostan, E. and Kusche, J.**, 2012. Separation of global time-variable gravity signals into maximally independent components, *Journal of Geodesy*, **86**, 477–497.

[38] **Lenk, O.**, 2013. Satellite based estimates of terrestrial water storage variations in Turkey, *Journal of Geodynamics*, **67(0)**, 106 – 110, http://www.sciencedirect.com/science/article/pii/S0264370712000713.

APPENDICES

APPENDIX A.1 : Outline of the Thesis

APPENDIX A.1



Purpose
Dataset
Method
CS6:
Spatial resolution: 500 km
Temporal resolution: Aylık
Geographical grid: 1° x 1°
Time spanı: 2003-2012
Filter: Kusche 2
Release: RL04
CS2:
Spatial resolution: 250 km
Temporal resolution: Monthly
Processing method: Kalman filter & WGHM
Geographical grid: 1° x 1°
Time span: 2003-2012
Filter:-
Release: RL05
CS3:
Spatial resolution: 350 km
Temporal resolution: Monthly
Geographical grid: 1° x 1°
Time span: 2003-2012
Filter: Kusche 3
Release: RL05
Gauss filtered (400km) centered GRACE monthly solutions 2005-2008 1x1 geographical grid
Gauss filtered (400 km) Centered WGHM monthly solutions 2005-2008 1° x 1° geographical grid
Spectrally analyzed Spherical Harmonic Coefficients (Max: 90)
Comparision of different version of GRACE data with using PCA
Comparision of Hydrology and GRACE data by using PCA
Mask definition according to sinyal values, examining leakage problem
Global
Regional
Cutting the regions out of the dataset with using their coordinates
Regional
TWS calculation from coefficients (Max: 90)
TWS calculation from coefficients (Max: 50)
Orinaco
Amazon
Turkey
Tigris/Euprates
Extractiong of region of interest with using defined optimum mask
Extraction of region of interest with using coordinate base masks
Threshold definition fot the regions
PCA
Amazon
Greenland
Amazon
Greenlad
Turkey
Mekong
Tigris/Euprates
Calculation of spatial change (EOFs)
Definiton of Amazon and Greenland masks with variety of TH
Reconstruction of 3 different regions with 95% variance & calculation of required mode number
PCA
Reconstruction of WGHM timeseries with 80% ve 90% variance
Calculation of temporal change (PCs)
Grace-like data produced from test dataset (Max: 90)
Test data: Continental Hydrology 0.5° x 0.5° geographical grid
Comparing the results of different masking techniques with different dataset for Amazon & Greenland regions
Comparision of the results of the different dataset for each region
Visualization of the differences of original series and the reconstructed series (WGHM 2005-05 monthly solutions)
Comparision of EOFs and PCs results of the two dataset
Amazon mask
Testing the masks
Comparision of PCA results of 3 different data-set with in-situ data
Greenland mask
The masks are enlarged
Decision of optimum mask
It is observed that missing months of the GRACE affected the results
Applying the coordinat base Greenland mask
Mekong mask
Applying the coordinate base Mekong mask
PCA
CS2, CS3 & CS6 2003-2011
Analysis of the effect of the gaps in timeseries to PCA results
Creating artificial gaps in the dataset
Tigris/Euphrates mask
Applying the coordinate base Amazon mask
Comparision of the results with gapless dataset PCA results

PHOTO

CURRICULUM VITAE

Name Surname: Ceren ALTUNAL PODLECH

Place and Date of Birth: Cankaya 04.09.1989

Adress: Wilhelmshavener Straße 14 10551 Berlin Germany

E-Mail: ceren-altunal@hotmail.com

B.Sc.: Geomatics Engineering Istanbul Technical University