Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 804

**Bramha Dutt Vishwakarma**

# Understanding and repairing the signal damage due to filtering of mass change estimates from the GRACE satellite mission

**München 2017**

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 | ISBN 978-3-7696-5216-1

Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C | Dissertationen | Heft Nr. 804

# Understanding and repairing the signal damage due to filtering of mass change estimates from the GRACE satellite mission

Von der Fakultät Luft- und Raumfahrttechnik und Geodäsie

der Universität Stuttgart

zur Erlangung des Grades

Doktors der Ingenieurwissenschaften (Dr.-Ing.)

genehmigte Dissertation

Vorgelegt von

M. Tech. Bramha Dutt Vishwakarma

Geboren am 05.12.1988 in Fatehpur, India

München 2017

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325 ISBN 978-3-7696-5216-1

**Adresse der DGK:**

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 ● D – 80 539 München

Telefon +49 – 331 – 288 1685 ● Telefax +49 – 331 – 288 1759

E-Mail post@dgk.badw.de ● http://www.dgk.badw.de

Prüfungskommission:

Vorsitzender: Prof. Dr.-Ing. Nico Sneeuw

Korreferenten: Prof. Dr.-Ing Martin Horwath (Dresden)

Tag der mündlichen Prüfung: 13.04.2017

© 2017 Bayerische Akademie der Wissenschaften, München

Alle Rechte vorbehalten. Ohne Genehmigung der Herausgeber ist es auch nicht gestattet,
die Veröffentlichung oder Teile daraus auf photomechanischem Wege (Photokopie, Mikrokopie) zu vervielfältigen

ISSN 0065-5325 ISBN 978-3-7696-5216-1

# Understanding and repairing the signal damage due to filtering of mass change estimates from the GRACE satellite mission

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

by

Bramha Dutt Vishwakarma

born in Fatehpur, India

| | |
|---|---|
| Main referee: | Prof. Dr.-Ing. Nico Sneeuw |
| Co-referee: | Prof. Dr.-Ing. Martin Horwath, Dresden |
| Date of defense: | 13 April 2017 |

Institute of Geodesy
University of Stuttgart
2017

Copyright ©Bramha Dutt Vishwakarma. All rights reserved.

## ABSTRACT

Filtering noisy observations to extract meaningful information is an old and necessary exercise for engineers and scientists. Filtering affects both the signal and the noise. While the noise is reduced to a minimum, the filtered observation is a smoothed representation of the true signal. The amount and type of smoothing required depends on the noise level. The geodetic satellite mission, GRACE provides heavily contaminated time variable gravity field of the Earth. Therefore, we have to use a strong smoothing operator before the data can be used. Over the past decade, many types of filters have been developed for treating the noisy GRACE products, which damages the signal. Therefore, along with filters, a number of methods to restore the signal damage have emerged. However, the majority of these methods use hydrological models to compute correction terms, such as leakage, bias, or scale factors, in a setup that lacks a detailed mathematical understanding. We fill this gap by studying the convolution integral on the sphere, with a motivation to revert the signal changes in filtered GRACE products.

Since the dominant time varying signal observed by GRACE comes from mass transports in the hydrosphere, we analyze the impact of filtering on catchment scale hydrology. We discuss the convolution integral in the spatial domain, which helps us to break the total impact of filtering into two parts: *leakage* and *attenuation of catchment-confined signal*, where leakage is only the contribution of signal from outside the catchment. We find that leakage changes the *amplitude* as well as the *phase* of the catchment-confined filtered signal. Previous contributions have addressed only the amplitude change due to filtering, usually with the help of a hydrological model. This practice propagates the error and the uncertainties in models to the corrected GRACE products. Therefore, we advocate avoiding models for computing correction terms.

A mathematical dissection of the convolution integral leads us to two methods for approaching the true regional average: *the method of scale* and *the method of deviation*. The method of scale uses the uniform layer approximation, while the method of deviation avoids any approximation. In a noise-free closed-loop test, we show that the method of scale is able to approach the truth, while the method of deviation gives us the true value. These methods need accurate knowledge of leakage and the deviation integral, which are estimated in a data-driven framework employing once filtered and twice filtered GRACE fields. In a closed-loop simulation environment with GRACE-type noise, we demonstrate for 32 catchments that we are able to approach the true leakage and the true deviation integral. The efficacy of data-driven method of deviation is found to be superior to three popular model dependent approaches.

After being satisfied with data-driven methods for hydrology, we intend to use them for assessing the ice mass loss in ice sheets such as Antarctica and Greenland, but we find that they fail for ice sheets. This is due to the physical difference between the spatial mass change distribution in an ice sheet and in a hydrological catchment: the former suffers from a mass

change concentrated near coast, while the later experiences a mass change throughout the catchment. Therefore, we tailor a new approximation for ice sheets giving us *the data-driven method for ice sheets*. It is tested effective in a noisy closed-loop simulation environment.

The data-driven methods are used to correct the filtered GRACE products and to analyze the total water mass loss over Aral sea, lake Urmia, lake Victoria, California, Antarctica, and Greenland. We report and compare our findings with previously reported figures. We find that the long term trend in mass change is suppressed by filtering, and overestimated by model dependent approaches.

This thesis explores the signal damage at catchment scale due to filtering of GRACE products, and develops data-driven methods to repair the signal damage. In a realistic closed-loop simulation environment, we demonstrate that the corrected signal is closer to truth. The performance decays with the catchment size, but is still better than model dependent approaches. Furthermore, the data-driven method is less accurate over arid regions (desert), however, the performance is on a par with the model dependent methods. Nevertheless, we extract our confidence from the overall performance of the data-driven methods in closed-loop environments to believe that we get superior mass change estimates from GRACE. This contribution helps us to reduce the filtering induced uncertainty in GRACE products.

# ZUSAMMENFASSUNG

Das Filtern von verrauschten Beobachtungen, um sinnvolle Informationen zu extrahieren, ist eine alte und notwendige Übung für Ingenieure und Wissenschaftler. Die Filterung wirkt sich sowohl auf das Signal als auch auf das Rauschen aus. Während das Rauschen auf ein Minimum reduziert wird, ist die gefilterte Beobachtung eine geglättete Darstellung des wahren Signals. Die Menge und Art der erforderlichen Glättung hängt vom Rauschpegel ab. Die geodätische Satellitenmission GRACE stellt stark verrauschte zeitvariable Schwerkraftmodelle der Erde zur Verfügung. Daher müssen wir einen starken Glättungsoperator verwenden, bevor die Daten verwendet werden können. In den letzten zehn Jahren wurden viele Arten von Filtern entwickelt, um die verrauschten GRACE-Produkte zu behandeln, was jedoch das Signal verändert. Daher sind zusammen mit Filtern eine Anzahl von Verfahren zur Wiederherstellung des Signals entstanden. Jedoch verwenden die meisten dieser Methoden hydrologische Modelle, um Korrekturterme – wie *Leakage*, *Bias* oder Skalierungsfaktoren – in einer Konfiguration zu berechnen, die mathematisch noch nicht im Detail verstanden wurde. Wir füllen diese Lücke, indem wir das Faltungsintegral auf der Kugel untersuchen, mit der Motivation, die Signalveränderungen in gefilterten GRACE-Produkten wiederherzustellen.

Da das von GRACE beobachtete dominante, zeitlich variable Signal aus Massentransporten in der Hydrosphäre stammt, untersuchen wir die Auswirkungen der Filterung auf die Hydrologie auf der Ebene von Einzugsgebieten. Wir diskutieren das Faltungsintegral im Ortsraum, was uns hilft, den Gesamteffekt der Filterung in zwei Teile zu spalten: ***Leakage*** und **Dämpfung des Signals innerhalb eines Einzugsgebiets**, wobei *Leakage* nur der Signalbeitrag von außerhalb des Einzugsgebietes ist. Wir erkennen, dass *Leakage* sowohl ***Amplitude*** als auch ***Phase*** des vom Einzugsgebiet begrenzten gefilterten Signals ändert. Frühere Beiträge haben nur die Amplitudenänderung durch Filtern angesprochen, meist mit Hilfe eines hydrologischen Modells. Dieser Ansatz propagiert Fehler und Unsicherheiten der Modelle zu den korrigierten GRACE-Produkten. Daher empfehlen wir, Modelle zur Berechnung von Korrekturtermen zu vermeiden.

Eine mathematische Zerlegung des Faltungsintegrals führt uns zu zwei Methoden zur Annäherung an das wahre regionale Mittel: die ***Method of Scale*** und die ***Method of Deviation***. Die *Method of Scale* verwendet die *Uniform Layer Approximation*, während die *Method of Deviation* jede Näherung vermeidet. In einem rauschfreien *Closed-Loop-Test* zeigen wir, dass die *Method of Scale* in der Lage ist, sich der Wahrheit zu nähern, während die *Method of Deviation* den wahren Wert liefert. Diese Methoden benötigen genaue Kenntnisse von *Leakage* und *Deviation-Integral*, die in einem datenbasierten Rahmen mit einmal und zweimal gefilterten GRACE-Feldern geschätzt werden. In einer *Closed-Loop-Simulation* mit GRACE-typischem Rauschen demonstrieren wir für 32 Einzugsgebiete, dass wir in der Lage sind, uns der wahren *Leakage* und dem wahren *Deviation-Integral* anzunähern. Wir stellen fest, dass die datenbasierte *Method of Deviation* den drei üblichen modellabhängigen Ansätzen überlegen ist.

Nachdem wir für die Hydrologie mit den datenbasierten Methoden zufrieden sind, wollen wir diese für die Bewertung des Eismassenverlusts von Eisschilden wie in der Antarktis und Grönland nutzen, stellen jedoch fest, dass sie scheitern. Dies ist auf den physikalischen Unterschied zwischen der räumlichen Massenverteilung in einem Eisschild und einem hydrologischen Einzugsgebiet zurückzuführen: Im ersten Fall erfolgt die Massenänderung konzentriert in der Nähe der Küste, während im zweiten Fall das gesamte Gebiet eine Änderung erfährt. Daher entwickeln wir eine neue Näherung für Eisschilde, aus welcher wir die ***Data-Driven Method for Ice Sheets*** erhalten. Wir haben dies erfolgreich in einer rauschbehafteten *Closed-Loop-Simulation* getestet.

Die datenbasierten Methoden werden verwendet, um die gefilterten GRACE-Produkte zu korrigieren und den gesamten Wassermassenverlust von Aralsee, Urmiasee und Victoriasee, sowie in Kalifornien, der Antarktis und Grönland zu analysieren. Wir berichten und vergleichen unsere Ergebnisse mit zuvor veröffentlichten Zahlen. Wir stellen fest, dass der langfristige *Trend* von Massenänderungen durch Filterung unterdrückt und durch modellabhängige Ansätze überschätzt wird.

Diese Arbeit untersucht die Signalveränderung auf der Ebene von Einzugsgebieten durch Filterung von GRACE-Produkten und entwickelt datenbasierte Methoden, um die Signalschäden zu reparieren. In einer realistischen *Closed-Loop-Simulation* zeigen wir, dass das korrigierte Signal näher an der Wahrheit liegt. Die Leistungfähigkeit fällt mit der Größe der Einzugsgebiete ab, ist aber immer noch besser als modellabhängige Ansätze. Darüber hinaus ist das datenbasierte Verfahren weniger präzise über ariden Regionen (Wüsten), jedoch ist die Leistung auf einer Stufe mit den modellabhängigen Methoden. Nichtsdestoweniger ziehen wir unser Vertrauen in die datenbasierten Methoden aus deren Gesamtleistungsfähigkeit in Closed-Loop-Umgebungen, und nehmen deshalb an, dass wir bessere Massenänderungsschätzungen von GRACE erhalten. Dieser Beitrag hilft uns, die filterungsbedingte Unsicherheit in GRACE-Produkten zu reduzieren.

# ACKNOWLEDGEMENTS

**Institute of Geodesy**, **Stuttgart**, has been like a *gurukul* to me. It has helped me grow both scientifically and intellectually. Although no amount of words can express my gratitude towards members of the Institute, still I will try my best to cage my experiences and emotions in limited vocabulary of mine.

Dear Nico, I am still amazed by your knowledge and wisdom, be it scientific or cultural. Your measure of self-righteousness, your *sense of style*, and your ability to put yourself in the shoes of the person in front of you, are a few of many attributes of yours I want to learn. You have been my strength during my PhD. You gave me the freedom to create my own path, yet ensuring that I am never lost. I wish I can imitate a small part of your intellectual aurora in my life. I am grateful to you for supervising this thesis work, for discussions in your office or around the coffee machine, and for fueling my curiosity and passion for science.

Dear Balaji, if you were not my teacher and daily-supervisor during my master thesis, I would not have been writing this now. In a foreign land, I found a friend, a helping colleague, a wonderful scientist, and a walking encyclopedia in you. The foundation of this work could not have been laid, if you were not there when I was starting my PhD. Thank you for the discussions and encouragement, which motivated me to dive into mathematics of convolution. Your patience, will to learn, selfless dedication, and your hunger for perfection is remarkable and inspiring.

I feel lucky to have met Andreas and Martin at a conference, which lead to a short visit to Dresden. I am delighted to admit that your role in the method developed in this thesis work is vital. Thank you for being such a wonderful host, for collaborating, and for adding a new dimension to my research. A special thanks to Martin for accepting to be my examiner.

I am thankful to all my colleagues for cracking jokes, for providing mental support, for scientific discussions and for joining me in both political and meteorological "going nowhere" discussions. These three years would not have been easy without you all. Special thanks goes to Anita for being so comforting and welcoming, to Ron for providing the technological help, to Matthias and Markus for winning the battle for me in translating the Abstract in German, and to the cleaning staff for maintaining healthy working environment. Dear Prof. Keller, I am grateful to you for giving a listening ear to my mathematical and scientific troubles, and for writing a positive report about my progress to DAAD.

I am thankful to Prof. Kamal Jain for his unconditional support and belief in me, which motivated me to aim higher than I could imagine. I am deeply indebted to my parents and my siblings, who always backed my decision and provided me the unconditional support. I would like to thank my all my friends, for sharing their time, advice and jokes at both the

high and low phases. Thanks goes to my cricket buddies at the university of Stuttgart and my badminton buddies at turnen und freizeit feuerbach, for helping me keep myself active and fit despite having a sitting job.

I never had sleepless nights, mental stress, or a temporary phase of depression, and I owe you big time my best friend and my partner for life. Thanks Medhavi for cheering me up, for supporting me, and for motivating me to take each day as an opportunity to improve as a person and as a researcher.

I would like to thank Ms. Ursula Habel at the international center, University of Stuttgart, for being patient and cooperative with my administrative queries. Last but not the least, I am thankful to DAAD for providing the financial support towards my PhD studies.

# CONTENTS


Abstract i

Zusammenfassung iii

Acknowledgements v

Table of symbols ix

1 Introduction 1

1.1 Historical overview 1

1.2 Mathematical background 3

1.3 Filtering: signal processing 6

1.4 Filters for GRACE 8

1.5 Motivation and objectives 11

2 Understanding the impact of filtering on the signal 13

2.1 Filtering and regional averages 13

2.2 Breaking down the process of filtering 15

2.3 Signal leakage due to filtering 16

2.4 Understanding effects of leakage 20

2.5 Impact of filtering on the catchment confined signal 22

2.6 Popular model-dependent approaches 25

2.7 Closed-loop validation 26

3 A tool for noisy fields 31

3.1 Impact of noise 31

3.2 Leakage from filtered fields 33

3.3 The deviation integral from filtered fields 38

3.4 Estimating the original leakage and the original deviation integral 39
3.5 The special case of ice sheets 50

4 Are we able to approach the truth? 53

4.1 Data, simulation environment, and validation 53
4.2 Testing in a more realistic closed-loop simulation 62
4.3 Closed-loop test of data-driven method for ice sheets 69
4.4 Validation by closing the water balance equation 69

5 Correcting the filtered grace products 73

5.1 Endorheic catchments: desiccating water bodies 74
5.2 Replenishment of lake Victoria 78
5.3 Ground water depletion: the drought in California 79
5.4 Global hydrology from grace 81
5.5 Ice mass loss in Antarctica and Greenland 81

6 Conclusions and outlook 87

6.1 The journey: aims and achievements 87
6.2 The big picture 89
6.3 Putting the pen down 90

bibliography 93

a Data and models 99

b Changing the filter 101

# TABLE OF SYMBOLS

| Symbol | Definition | first used |
|---|---|---|
| $l$ | spherical harmonic degree | (1.1) |
| $m$ | spherical harmonic order | (1.1) |
| $C_{lm}$, $S_{lm}$ | spherical harmonic coefficients | (1.1) |
| $P_{lm}$ | Legendre function of the first kind | (1.1) |
| $r$ | radial component of the Earth fixed coordinate system | (1.2) |
| $\theta$ | co-latitude | (1.1) |
| $\lambda$ | longitude | (1.1) |
| $V(\cdot)$ | gravitational potential | (1.2) |
| $a$ | mean equatorial radius of the Earth | (1.3) |
| $\rho$ | density | (1.3) |
| $\rho_{\text{avg}}$ | average density of the Earth | (1.3) |
| $\sigma$ | surface density | (1.4) |
| $k_l$ | load Love number for degree $l$ | (1.5) |
| $\rho_{\text{water}}$ | density of water | (1.7) |
| $\mathcal{F}(X)$ | Fourier transform of $X$ | (1.12) |
| $\mathcal{F}^{-1}(X)$ | inverse Fourier transform of $X$ | (1.12) |
| $\psi$ | spherical distance | (1.13) |
| $w(\psi)$ | weight as a function of spherical distance | (1.13) |
| $f(\cdot)$ | a global field | (1.14) |
| $\bar{f}(\cdot)$ | filtered field | (1.14) |
| $b(\cdot,\cdot)$ | two point filter kernel | (1.14) |
| $W_{lm}^{nk}$ | spectral weight | (1.16) |
| $R(\cdot)$ | catchment characteristic function | (2.1) |
| $A_{\text{c}}$ | area of a catchment | (2.1) |
| $f_{\text{c}}$ | regional average of the field over a catchment c | (2.1) |
| $R^*(\cdot)$ | complement of catchment characteristic function, $(1 - R(\cdot))$ | (2.5a) |
| $F(\cdot)$ | catchment confined field, $[f(\cdot) \cdot R(\cdot)]$ | (2.6b) |

| | | |
|---|---|---|
| $F^*(\cdot)$ | a field only outside the catchment, $[f(\cdot) \cdot R^*(\cdot)]$ | (2.6b) |
| $l_c$ | regional average of leakage over a catchment c | (2.11) |
| $\alpha$ | amplitude of time series | (2.14) |
| $\omega$ | angular frequency of time series | (2.14) |
| $\phi$ | phase of time series | (2.14) |
| $s$ | scale factor | (2.22) |
| $\delta F(\cdot)$ | deviation of field $F(\cdot)$ with respect to the catchment average $f_c$ | (2.29) |
| $\overline{\delta F}_c$ | deviation integral | (2.31) |
| $k$ | model derived scale factor | (2.34) |
| $l_c^m$ | model derived leakage | (2.32) |
| $b_c$ | model derived bias | (2.33) |
| $g(\cdot)$ | GRACE field | (3.1) |
| $n(\cdot)$ | noise field | (3.1) |
| $N(\cdot)$ | catchment confined noise field | (3.2) |
| $n_c$ | catchment aggregated noise | (3.5) |
| $n_c^l$ | the noise in catchment aggregated leakage | (3.5) |
| $\overline{\delta N}_c$ | the deviation integral of catchment confined noise | (3.6) |
| $\kappa(\cdot)$ | $R^*(\cdot) \cdot \bar{R}(\cdot)$ | (3.9) |
| $\gamma(\cdot)$ | $R(\cdot) \cdot \bar{R}(\cdot)$ | (3.20) |
| $p_c$ | total precipitation in a catchment | (4.1) |
| $e_c$ | total evapotranspiration from a catchment | (4.1) |
| $r_c$ | total runoff from a catchment | (4.1) |
| $\mathrm{d}f_c/\mathrm{d}t$ | rate of mass change in a catchment | (4.1) |

**general conventions**

| | |
|---|---|
| $(\cdot)$ | $(\theta, \lambda)$ |
| $\bar{X}$ | filtered $X$ |
| $X_c$ | regional average of $X$ over a catchment c |
| $\hat{X}$ | estimated value of $X$ |
| $\Delta X$ | long term mean removed (residual of) $X$ |

# 1 INTRODUCTION

## §1.1 HISTORICAL OVERVIEW

GEODESY is the science of measuring and mapping the Earth's surface [Helmert, 1880], which requires accurate measurement and understanding of three fundamental properties of the Earth: its geometric shape, its orientation in space, and its gravity field. These properties and their evolution in time can be determined by carrying out dedicated geodetic expeditions, but they are time consuming and costly. Furthermore, a good global and temporal coverage is a challenge too big to overcome. Therefore, after the launch of the first man made satellite, SPUTNIK-1 on October 4, 1957, researchers realized an opportunity to overcome this geodetic challenge, and a lot of effort was directed into studying geodesy with the help of satellites. This gave rise to a whole new discipline, known as satellite geodesy. It gained a lot of momentum at the start of the cold war because an accurate knowledge of the geodetic properties of the Earth is vital for calculating the trajectories of ballistic missiles. Therefore, in the early 1960s, obtaining superior geodetic information was one of the top priorities of the Department of Defense (DOD), the United States[*]. Their first geodetic satellite project ANNA was launched in 1962. It was named after its four contributors, the Army, the Navy, NASA and Air force. The idea was to monitor the motion of the satellite from dedicated ground stations to obtain gravity field information. NASA wanted global cooperation so that they can spread the ground stations and improve the quality of the data. However, DOD wanted the mission to be classified and this affected the quality and the glory of the first dedicated geodetic satellite mission ANNA[†]. It was followed by three NASA Geodetic Earth Orbiting Satellites (GEOS), launched in 1965, 1968, and 1975. GEOS-3 was the first satellite to carry a RADAR altimeter, which contributed towards the mapping of ocean heights. Within the next decade, a number of satellite missions were launched to explore the global sea surface heights, to determine the Earth's shape and size parameters, and to help the military navigation.

In the middle of these numerous developments, a unique idea to determine the Earth's gravity field at a spatial scale of 100 km was taking shape: the Geopotential Research Mission (GRM) [Taylor et al, 1983; Wagner, 1987]. On October 29, 1984, in the University of Maryland, Dr. Burton I. Edelson (then the Associate Administrator, NASA office of Space Science and Application) was addressing the Geopotential Research Mission Science conference as the keynote speaker. He said "*As for the GRM, I can promise you that it will*

[*] http://nsarchive.gwu.edu/

[†] http://nsarchive.gwu.edu/NSAEBB/NSAEBB509/

*receive very serious consideration in my office. It is now a new start candidate in fiscal year 1988. It is an excellent mission - from a scientific standpoint, from the point of view of practical use of the data, and from the standpoint of feasibility and technological readiness.*" Dr. Edelson was previously associated with the space program ANNA, and his words were certainly a confirmation of the importance of GRM, and a motivation to the geodetic community working for GRM. The co-chairman of the conference was Prof. William M. Kaula, who, in his conference summary, described GRM as a satellite system proposed to determine the gravitational and magnetic fields to the resolution of about 100 km. In order to achieve this, the satellite mission was designed to accurately record the range-rate between two satellites separated by 300 km, moving in a coincident near-polar orbit at an altitude of 160 km. Although a lot of effort and money was spent on GRM, it was never realized. In an interview to Rebecca Wright, Dr. Byron D. Tapley told that it was due to the first Space Shuttle disaster on January 28, 1986, that most of the satellite missions were delayed and a few got cancelled[‡]. For example, TOPEX was accepted in 1985; it was delayed and finally launched in 1992, while the ambitious GRM team was disbanded in 1986 only.

I participated in an Autumn School on 'Global Gravity Field Modeling from Satellite-to-Satellite Tracking Data', from October 4–9, 2015, held in Bad Honnef, Germany. The opening speech was delivered by Prof. Reiner Rummel, who mentioned that the geodetic community was trying hard since the early 1990s to send a GRM type mission, but finding funds was always a challenge. Then someone added climate to the name, which provided necessary financial wings to the satellite pair, and Gravity Recovery And Climate Experiment (GRACE) satellite mission was born. Furthermore, the scientific community was more interested in the time-variability of the gravity field, which required the satellite mission to fly for a longer period of time. Therefore, the altitude of GRACE satellites was increased compared to the GRM mission and an accurate accelerometer was added to measure the surface forces due to drag and radiation pressure. These changes eliminated the cost and mission life limitations associated with GRM. Finally, with a collaboration between GeoForschungZentrum (GFZ) in Potsdam, Germany, and NASA, the twin satellite mission was launched on March 17, 2002.

The GRACE satellite mission is a modified and improved version of GRM. The observations include precise measurements of the variation in inter-satellite distance, non-gravitational accelerations, spacecraft attitude, GPS coordinates, and other vital information. They constitute the level-1 data, from which one can construct gravity field solutions. There are many institutes, which have the expertise to process the level-1 information and produce the level-2 data: spherical harmonic coefficients. The level-2 data provided by most of the data centers (for example: GFZ, CSR, JPL) is at monthly time scale, which are then processed to produce maps of mass change, also known as level-3 products. Furthermore, recently another type of GRACE products have been introduced, known as mascon (mass concentration) blocks.

A mascon solution provides an estimate of the mass anomalies at a predefined grid, point, or block. They are computed from the GRACE observations with the help of constrained regularization. Currently three centers are providing mascon solutions: Jet Propulsion Laboratory (JPL) [Watkins et al, 2015], Goddard Space Flight Center (GSFC)

[‡]http://www.jsc.nasa.gov/history/oral_histories/NASA_HQ/ESS/TapleyBD/TapleyBD_1-12-10.htm

[Luthcke et al, 2006, 2013], and Center for Space Research (CSR) [Save et al, 2016]. The difference between these mascon solutions is in the procedure followed to constrain the regularization, for example Watkins et al [2015] employ explicit partial derivatives with analytical expression for mass concentration to relate the intersatellite range-rate measurements to the individual mascons. [Luthcke et al, 2013] relate the mascons to the range-rate or the range-acceleration via a spherical harmonic expansion truncated at a finite degree and order, and Save et al [2016] fit the mascons to the spherical harmonic coefficients obtained from GRACE. These mascon products do not need post-processing, and hence can be used as level 3 products. Although, these estimates of mass change claim to retain maximum signal, the signal change due to regularization is still not investigated.

## §1.2 MATHEMATICAL BACKGROUND

A square integrable analytic gravitational potential function $V(r,\theta,\lambda)$ can be represented in the frequency domain by spherical harmonic coefficients with the help of the following synthesis relation [Kaula, 1966; Chao and Gross, 1987; Wahr et al, 1998]

$$V(r,\theta,\lambda) = \frac{GM}{r}\sum_{l=0}^{\infty}\left(\frac{a}{r}\right)^{l}\sum_{m=0}^{l}\bar{P}_{lm}(\cos\theta)\,[C_{lm}\cos m\lambda + S_{lm}\sin m\lambda], \tag{1.1}$$

where $G$ is the gravitational constant, $M$ is the total mass of the Earth, $a$ is the mean radius of the Earth, $\bar{P}_{lm}$ are the fully normalized Legendre functions of the first kind, $C_{lm}, S_{lm}$ are the fully normalized spherical harmonic coefficients, $l$ is the degree and $m$ is the order. The left hand side of the equation (1.1) is the spatial domain $(r,\theta,\lambda)$, and the right hand side is the spectral domain $(l,m)$. (1.1) is similar to Fourier theory, except for the fact that the frequency decomposition involves spherical harmonic functions. We can compute the spatial domain quantity from the spectrum (spherical harmonic coefficients) and *vice versa*. The potential $V(r,\theta,\lambda)$ at point $P$ is related to the density distribution at a point $Q$ as

$$V_P = G\int\int\int_{v}\frac{\rho_Q}{r_{PQ}}\mathrm{d}v_Q, \tag{1.2}$$

where $v$ denotes the volume. $r_{PQ}$ is the distance between the points $P$ and $Q$. The total mass of the Earth is constant, but a continuous mass redistribution is taking place via natural processes. For example, water mass is continuously in motion via hydrological cycle. Therefore, if we want to observe hydrological phenomena with the help of potential, then we must focus on the change in potential over time.

The gravity potential in (1.1) is static, which means it does not represent the time variability. However, GRACE is well-known for providing the time-variable gravity information, which is represented by change in the spherical harmonic coefficients. To calculate the change in spherical harmonic coefficients we remove a long term mean of the spherical harmonic coefficients from monthly values. These residual spherical harmonic coefficients are denoted by $\Delta C_{lm}$ and $\Delta S_{lm}$. Since we are interested in mass change, the potential change reflected by the residual spherical harmonic coefficients must be represented in

terms of density change $\Delta\rho(r,\theta,\lambda)$ [Wahr et al, 1998]

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \frac{3}{4\pi a \rho_{\text{avg}}(2l+1)} \int\int\int_v \Delta\rho(r,\theta,\lambda) \left(\frac{r}{a}\right)^{l+2} \bar{P}_{lm}(\cos\theta) \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\lambda \, \mathrm{d}r, \tag{1.3}$$

where $\rho_{\text{avg}}$ is the average density of the Earth ($5517\,\text{kg/m}^3$). One big problem with the relation (1.3) is that we can not determine how the density varies with the depth within the Earth. In order to overcome this problem, we assume that most of the mass redistribution takes place within a thin layer near the Earth's surface, thus $(r/a)^{l+2}$ is approximately equal to 1 and we can replace the volume integral of $\Delta\rho$ by the surface integral of $\Delta\sigma$, and (1.3) becomes

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\}_{\text{surf mass}} = \frac{3}{4\pi a \rho_{\text{avg}}(2l+1)} \int_\Omega \Delta\sigma(\theta,\lambda) \, \bar{P}_{lm}(\cos\theta) \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \mathrm{d}\Omega. \tag{1.4}$$

Furthermore, the solid Earth responds to the mass redistribution with a deformation, which also contributes to the total gravity change. This is taken care of by the load Love numbers $k_l$ [Wahr et al, 1998]

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\}_{\text{solid Earth}} = \frac{3\,k_l}{4\pi a \rho_{\text{avg}}(2l+1)} \int_\Omega \Delta\sigma(\theta,\lambda) \, \bar{P}_{lm}(\cos\theta) \left\{ \begin{array}{c} \cos m\lambda \\ \sin m\lambda \end{array} \right\} \mathrm{d}\Omega. \tag{1.5}$$

The total change in the spherical harmonic coefficients is given by a sum of (1.4) and (1.5):

$$\left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\} = \left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\}_{\text{surf mass}} + \left\{ \begin{array}{c} \Delta C_{lm} \\ \Delta S_{lm} \end{array} \right\}_{\text{solid Earth}}.$$

The resulting synthesis equation is

$$\Delta\sigma(\theta,\lambda) = \frac{a\rho_{\text{avg}}}{3} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \bar{P}_{lm}(\cos\theta) \, \frac{2l+1}{1+k_l} \, [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda]. \tag{1.6}$$

Since the mass redistribution at monthly time scales is dominated by hydrology, the density change $\Delta\sigma(\theta,\lambda)$ is expressed in terms of Equivalent Water Height (EWH)

$$\Delta\sigma(\theta,\lambda) = \text{EWH}(\theta,\lambda) \cdot \rho_{\text{water}} = \frac{a\rho_{\text{avg}}}{3} \sum_{l,m} \bar{P}_{lm}(\cos\theta) \, \frac{2l+1}{1+k_l} \, [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda].$$

$$\text{EWH}(\theta,\lambda) = \frac{a\rho_{\text{avg}}}{3\rho_{\text{water}}} \sum_{l,m} \bar{P}_{lm}(\cos\theta) \, \frac{2l+1}{1+k_l} \, [\Delta C_{lm} \cos m\lambda + \Delta S_{lm} \sin m\lambda]. \tag{1.7}$$

The aim of the GRACE satellite mission is to obtain $\Delta C_{lm}$ and $\Delta S_{lm}$ accurately. However, it was known even before the launch that the spherical harmonic products would be noisy (cf. Figure 1.1) and the resolution would be coarse [Wahr et al, 1998]. There is a

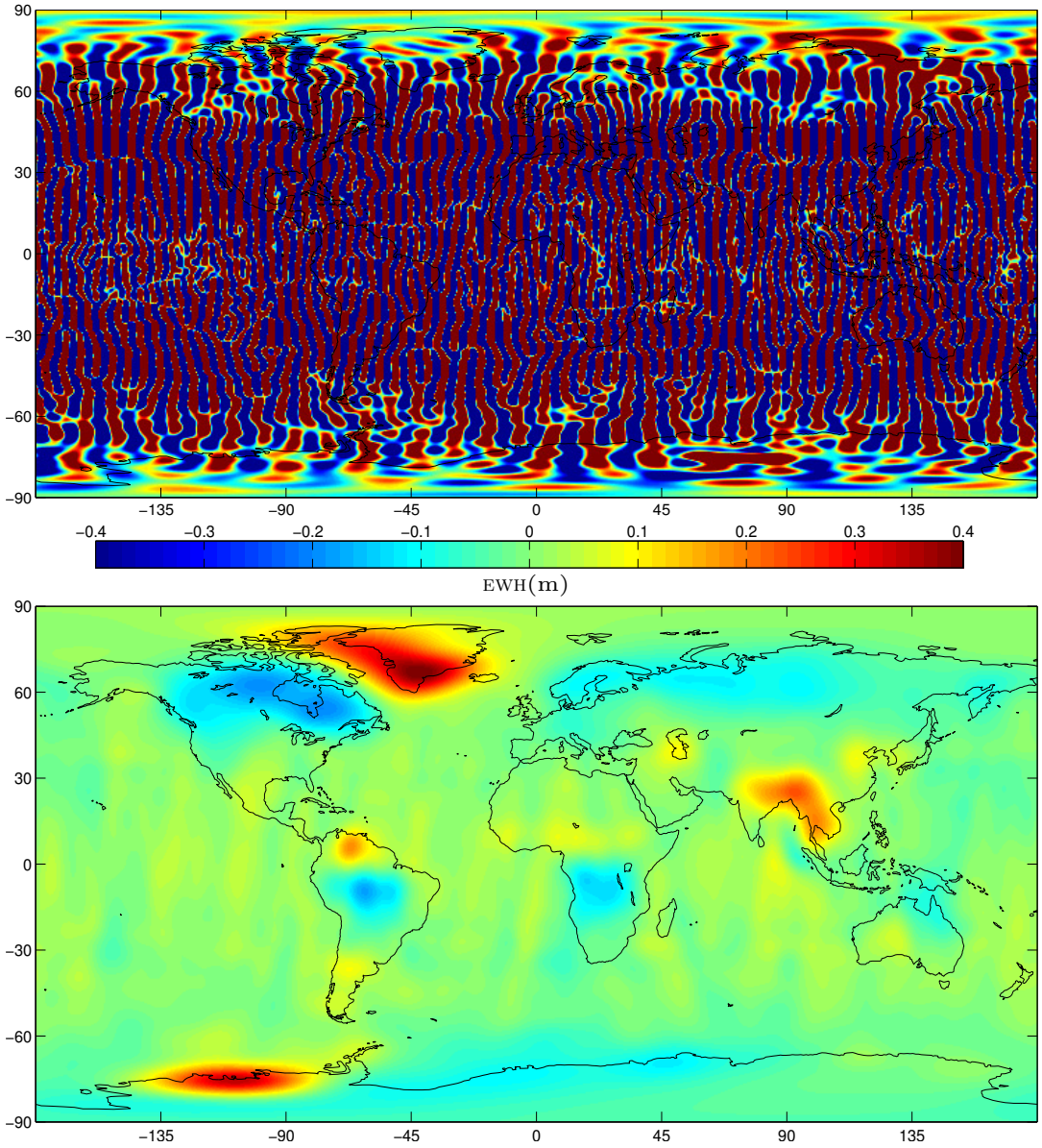


**Figure 1.1:** The monthly GRACE product for the month of October, 2004. The top figure is the noisy product and the lower figure is the filtered product. A Gaussian filter (cf. 1.15) of half width radius 500 km is used.

delicate trade-off between the resolution and noise. The spatial resolution depends on the maximum spherical harmonic degree and order that we can extract with significant signal to noise ratio (spatial scale $\approx 20000/l$ km). However, if we aim for maximum attainable spherical harmonic degree and order, then we suffer from a large amount of noise, because noise increases with spherical harmonic degree. Therefore, filtering is needed to suppress noise in the higher degree spherical harmonic coefficients [Jekeli, 1981; Wahr et al, 1998]. Jekeli [1981] studied well known filters from Fourier theory for signal processing in 2 dimensions, and adapted them for data on a sphere. Along with the benefits of filtering, we have also borrowed the side-effects. In order to understand both faces, we must attain a basic understanding of filtering from a signal processing perspective.

## §1.3 FILTERING: SIGNAL PROCESSING

A filter, in day to day life, is used to pass some component and attenuate, if not stop, the other components of a contaminated commodity (mixture). For example a coffee filter, an oil filter in an automobile, or a lens filter used in photography. Each one of them help us to extract the component that we desire. Similarly, in signal processing, a filter is used to pass certain frequencies and attenuate others.

If we look into the historical development of filters in signal processing, we find that the first documented effort towards designing an analog filter was from George Ashley Campbell in 1911. In 1930 Butterworth filters were proposed, followed by Bessel filters in 1949, and Chebyshev filters in the 1950s [Paarmann, 2001]. The process of filter design is still going on with an aim to improve on previous filters. However, taking an average of a series of measurements to increase the certainty of the information is a very old technique, which is simple and effective. The traditional method of taking average and the modern method of selecting frequencies yield similar results [Hamming, 1998], although the computation is carried out in different domains (the space or the time domain for traditional method and the spectral domain for contemporary filters).

The freedom to move from the space domain (or the time domain) to the frequency domain has increased our understanding and revolutionized the developments in signal processing. The credit goes to Jean Baptise Joseph Fourier, the great French mathematician, who developed Fourier series and the Fourier integral while investigating the propagation of heat in solid bodies. His paper *Théorie analytique de la chaleur* was initially rejected, but after 15 years, in 1882, it was accepted and is still a classic. The results from his contribution are driving the modern day science in a way, which one can not describe in a few lines. He proposed to represent an arbitrary function defined in a finite interval as a sum of sinusoids. Thus, opening the gates between the space domain (or the time domain) and the frequency domain. Since then we have been choosing a domain, where we are comfortable, and carrying out mathematical operations to extract, to process, or to modify the signal at our disposal.

To this end, we must admit that for filtering spectral domain operations have been favored over spatial domain operations. This is due to the fact that in the spatial domain filtering is performed by computing a convolution integral, while in the spectral domain filtering is a multiplication of each frequency by a weight. Thus the spectral domain

operation is simple to apply and is faster. Most importantly, it is easier to understand and express. For example, if we want a frequency to be extracted from a data, then in the filter parameters we provide the weight corresponding to that frequency as 1 and elsewhere 0. We can classify filters depending on the frequency it passes: low pass filter, band pass filter, or high pass filter. The type of filter chosen for suppressing noise, depends on the frequency spectrum of the signal and noise. For example, if the signal is dominated by low frequencies and the noise is a white Gaussian noise or a blue noise (noise power increases with frequency), then a low pass filter is preferred.

Filtering in the spatial or the time domain is written as a convolution integral between a function $f$ and filter $b$, as

$$f * b = \int_{-\infty}^{\infty} f(\tau)\, b(\tau - t)\, \mathrm{d}\tau. \tag{1.8}$$

The important properties of the convolution integral are:

*i.* It is commutative: the order of functions does not affects the result

$$f * b = b * f. \tag{1.9}$$

*ii.* It is associative: when we convolve more than two functions, then performing intermediate convolutions first will not change the result

$$(f * b) * c = f * (b * c). \tag{1.10}$$

*iii.* It is distributive: the sum of signal convolved with a function is equal to the sum of individual functions convolved separately

$$f * (b + c) = f * b + f * c. \tag{1.11}$$

*iv.* The convolution theorem: convolution in the time or the space domain is multiplication in the frequency domain

$$\mathcal{F}\left(f * b\right) = \mathcal{F}(f)\, \mathcal{F}(b). \tag{1.12}$$

In order to understand how a filter suppresses noise, we can use the fact that the operation is equivalent to computing a weighted average at each data point, where the magnitude of weights is described by the design of the filter kernel $b$. A set of observation represented by a function $f$ can be written as

$$f = \text{Signal} + \text{Noise},$$

and its average would yield

$$\text{Mean}\{f\} = \text{Mean}\{\text{Signal}\} + \text{Mean}\{\text{Noise}\}.$$

The output of filtering (the left hand side) is equal to the mean of the signal, only if the expectation or mean of noise is either equal to zero or is negligible. This is

achieved by designing an operator which minimizes the mean of noise generated by the system. Therefore, identifying the statistical properties of noise becomes an important task. In an electronic system, there are numerous sources of noise (shot noise, thermal noise, flicker noise, etc.) [Haykin and Moher, 2009]. The central limit theorem helps us with the problem of noise identification. It states that "the probability distribution of the random variable approaches a normalized Gaussian distribution in the limit, as the number of random variable approaches infinity" [Haykin and Moher, 2009]. Thus, for the most commonly studied scenarios, when independent random variables are added, their sum tends toward a normal distribution even if the original variables themselves are not normally distributed [Rice, 1995]. Therefore the expectation of noise, more often than not, is negligible.

An important point to be noted here is that, while we reduce noise to a minimum, we obtain the mean of the signal and the crude details of the signal are compromised. Thus it has been a big challenge to design a filter which retains maximum signal content while suppressing the noise to a minimum. The same is true for filters designed to extract the mass change information from the noisy GRACE products. The information content of GRACE products is compromised due to filtering. If we want to understand and restore the lost information, then we must understand the mechanism of filtering GRACE products. It is slightly different from the conventional planar filters in signal processing, because we have to work on 2-Sphere. Thankfully, the journey and the route from Euclidean space to the 2-Sphere was established in the early 1980s by Jekeli [1981].

## §1.4 FILTERS FOR GRACE

In order to perform the spatial averaging of GRACE products, we need, in its most general form, a two point filter function $b(\theta,\lambda,\theta',\lambda')$, where the weights $w$ (scalar values attained by $b$) are decided by the filter type and its parameters. The first filter proposed for GRACE was an isotropic filter, whose weights depend on the spherical distance $\psi$ between the calculation point $(\theta,\lambda)$ and the data point $(\theta',\lambda')$ [Jekeli, 1981; Wahr et al, 1998]. For a Gaussian spatial averaging the weights are described by

$$w(\psi) = \frac{\beta}{2\pi}\frac{\exp[-\beta(1-\cos\psi)]}{1-\exp^{-2\beta}}, \quad \text{where} \quad \beta = \frac{\ln(2)}{(1-\cos(r_{\text{fil}}/a))}, \tag{1.13}$$

$r_{\text{fil}}$ is known as the averaging radius of the filter, and it is defined as the distance on the Earth's surface at which $w$ has dropped to half of its value at $\psi = 0$. The spatial domain implementation of Gaussian smoothing is written as

$$\bar{f}(\theta,\lambda) = \frac{1}{4\pi}\int_{\Omega'} f(\theta',\lambda')\, b(\theta,\lambda,\theta',\lambda')\, \mathrm{d}\Omega', \tag{1.14}$$

where $\bar{f}(\theta,\lambda)$ is the filtered field, $f(\theta',\lambda')$ is the noisy field which we want to smooth, $\Omega'$ represents the domain of the surface of a unit sphere, $(\theta,\lambda)$ are co-latitude and longitude of the calculation point, $(\theta',\lambda')$ are co-latitude and longitude of the data point, and $\mathrm{d}\Omega'$ is the infinitesimal surface element $\sin\theta'\mathrm{d}\theta'\mathrm{d}\lambda'$. The corresponding spectral domain filtering

is written as [Wahr et al, 1998]

$$\overline{\Delta\sigma}(\theta,\lambda) \;=\; \frac{2a\rho_{\text{avg}}\pi}{3}\sum_{l,m} W_l \bar{P}_{lm}(\cos\theta)\,\frac{2l+1}{1+k_l}\,[\Delta C_{lm}\,\cos m\lambda \;+\; \Delta S_{lm}\,\sin m\lambda]. \tag{1.15}$$

This is similar to (1.6), except for a multiplication by the spectral weights $W_l$ defined as

$$W_l \;=\; \int_0^{\pi} w(\psi) P_l(\cos\psi)\sin\psi \mathrm{d}\psi, \quad \text{where} \quad P_l \;=\; \frac{\bar{P}_l}{\sqrt{2l+1}}.$$

For a Gaussian filter the weights decay smoothly with the degree (cf. Figure 1.2), and the filter function depends only on the degree. However, these weights can vary with location and with both the degree and order. Therefore, in general we must write [Wahr et al, 1998; Devaraju, 2015]

$$\overline{\Delta\sigma}(\theta,\lambda) \;=\; \frac{a\rho_{\text{avg}}}{12\pi}\sum_{l,m}\sum_{n,k} W_{lm}^{nk} \bar{P}_{lm}(\cos\theta)\,\frac{2l+1}{1+k_l}\,[\Delta C_{lm}\,\cos m\lambda \;+\; \Delta S_{lm}\,\sin m\lambda], \tag{1.16}$$

where $W_{lm}^{nk}$ are the spectral weights in their most general form. Depending on the combination of parameters, with which the weight varies, we can classify filters as: isotropic, anisotropic, homogeneous, and inhomogeneous [Devaraju, 2015]. The spectral form of these different type of filters is shown in Figure 1.3.

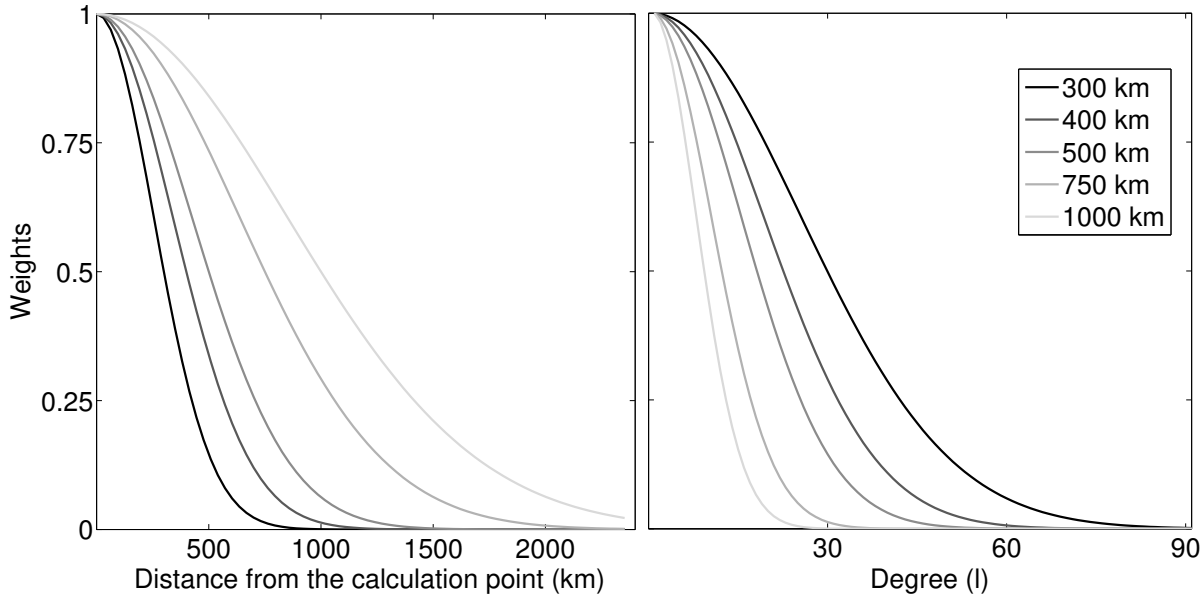


**Figure 1.2**: The weights for Gaussian filters in the spatial domain (left) and in the spectral domain (right) for different averaging radius.

The Gaussian filter is the most popular filter for GRACE post-processing because it is simple to apply, which is due to the fact that it assumes the noise to be isotropic. However, the GRACE noise is anisotropic and it varies with both degree and order [Han et al, 2005; Swenson and Wahr, 2006; Kusche, 2007; Klees et al, 2008]. The North-South linear features (*stripes*; cf. Figure 1.1) in the GRACE products is due to the fact that the spherical harmonic coefficients of certain orders are correlated [Swenson and Wahr, 2006].

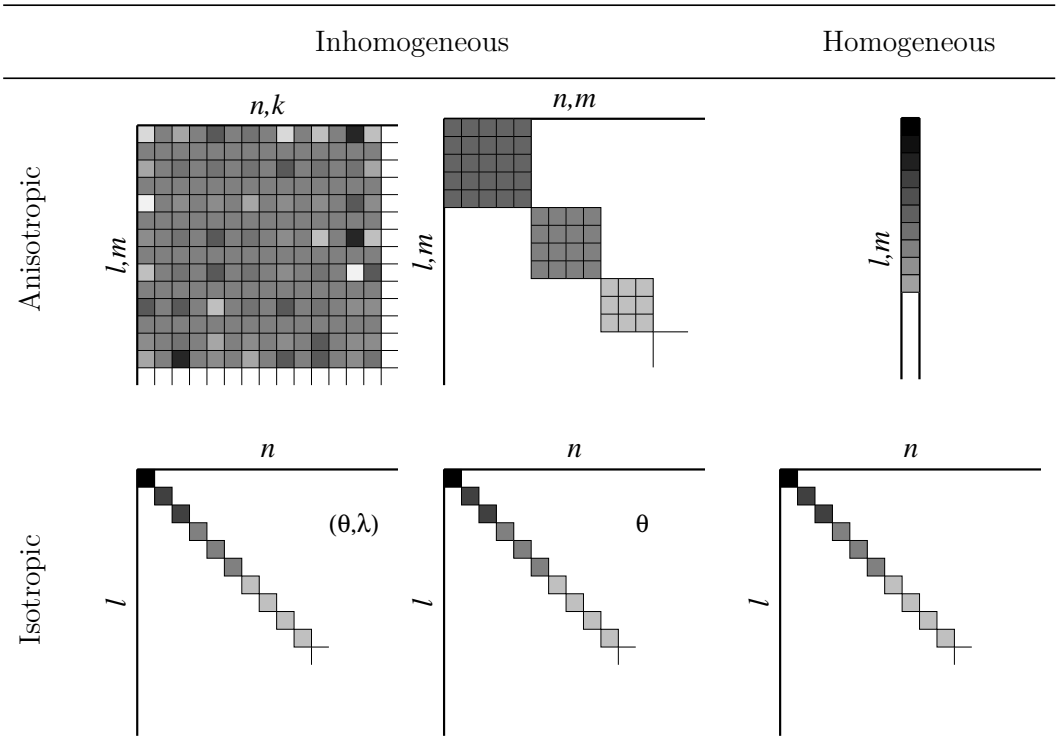


**Figure 1.3:** Different types of filters and their spectral form (*courtesy*: Balaji Devaraju). $l$ and $n$ denote degree while $m$ and $k$ denote order of the spherical harmonic coefficients. The weights of an isotropic filter depend on the spherical distance between the calculation point and the data point, while the weights of an anisotropic filter depends on both the spherical distance and the azimuth between the calculation point and the data point. The homogeneous filters are same at each point on the sphere, while the inhomogeneous filters change their shape with respect to colatitude and longitude.

Swenson and Wahr [2006] proposed a destriping filter that replaces the correlated coefficients of an order by the residuals obtained after fitting and removing a polynomial. This technique has a disadvantage of affecting North-South oriented signals, but it is efficient and widely used. Kusche [2007] proposed an anisotropic filter popularly known as DDK filter. The method involves regularization, leading to a series of decorrelation transformations enabling successive smoothing to treat the noise. The process requires a-priori signal covariance information, which is derived from a geophysical model. Zhang et al [2009] proposed a filter: fan filter, whose weights are derived from Gaussian smoothing but the averaging radius depends on the order of geopotential coefficients. This filter demonstrated an improvement in resolution along latitude. Apart from these popular filters, there were many efforts toward designing an optimal filter for GRACE. For example, Klees et al [2008] developed a filter that is unique for every month, it incorporates the variance–covariance information of noise and the full signal of the monthly solution. Duan et al [2009] proposed a moving window filter, whose width depends on the error pattern of the spherical harmonic coefficients. Devaraju [2015] developed a regularization filter that uses the variance information of the signal and noise, and drafted a set of performance metrics for evaluating filters. Today we have a garden of filters to choose from, and it raises the question which filter is the best for processing GRACE. There have been a few attempts to find the answer to this question [Werth et al, 2009], but with the ongoing developments the debate on filter supremacy is still alive.

## §1.5 MOTIVATION AND OBJECTIVES

While much effort went into designing filters for GRACE, a comprehensive effort to understand the side-effects of filtering has hardly been made. Wahr et al [1998] reported that filtering will introduce leakage between the signal from ocean and the signal from land, which will affect the amplitude of the signal. Many contributions demonstrated that filtering has a substantial impact on the GRACE signal. They refer to the change in the signal and the underlying cause by different names, e.g. bias, scaling, and leakage [Longuevergne et al, 2010; Klees et al, 2007; Landerer and Swenson, 2012; King et al, 2012]. Although a concern about signal damage due to filtering was prevalent in the community, a study on the nature of impact and its physical interpretation was missing, which is necessary to understand the information deficit in filtered GRACE products.

Klees et al [2007] reported an amplitude change of 50% to 70% in the the signal over Congo-Zambezi and nearby regions in Southern Africa. In order to restore the loss, they computed a leakage and a bias correction from a hydrological model. Longuevergne et al [2010] and Landerer and Swenson [2012] also used information derived from hydrological models to repair the damage to the signal due to filtering. These methods, using a hydrological model, raise a number of concerns. First there are numerous global and regional hydrological models and they differ from each other substantially [Werth, 2010]. Hence, choosing a model is a difficult task. Second, these models do not agree with GRACE in a consistent way [Lorenz et al, 2014]. Thus, using them will propagate their uncertainties to GRACE, which defies the whole initiative of improving filtered GRACE.

Apart from model dependent approaches, there were numerous efforts that do not use

models to correct the filtered GRACE products. Wahr et al [2007] demonstrated that for Greenland and Caspian sea (isolated regions) one can approach the true signal by using a scale factor derived from the information of regional extent and the filter. Similar scale factors improved the agreement between the altimetry data over the Mediterranean sea and the GRACE derived water height change [Fenoglio-Marc et al, 2006]. Baur et al [2009] proposed an iterative scheme to restore the signal change for Greenland without using any additional information (such as models or information from other satellite missions). However, these approaches were effective only for isolated catchments.

When I started my PhD research in October, 2013, GRACE was already more than a decade in orbit. The amount of research contributions towards post processing of GRACE products was both wide and deep. Nevertheless, the understanding towards the impact of filtering on the signal was limited. It would not be an overstatement to say that it was more philosophical than mathematical. The contribution from Klees et al [2007] was indeed the only candidate which provided a limited but better mathematical understanding of the problem. Balaji Devaraju, my colleague, was finishing his PhD thesis and he used the potential signal leakage as a performance metric for comparing filters. His research fueled my interest in developing a detailed understanding of leakage. In the following chapters we will accomplish the following:

*i.* develop the mathematical foundation for understanding the impact of filtering

*ii.* quantify the signal damage at catchment scale

*iii.* find the relation to approach the true signal

*iv.* develop a method to restore the signal loss for GRACE products

*v.* validate the findings, and discuss the limitations

In order to test the mathematical foundation or the efficacy of method, we use various global models and data provided by various institutes. An account of these models and data is provided in Appendix A.

The novelty of GRACE satellite mission has been its ability to sense mass changes beneath the surface of the Earth (Ground water storage) [Rodell et al, 2009; Tiwari et al, 2009; Shum et al, 2011], and beyond easily observable regions (ice sheets such as Antarctica and Greenland, and ocean mass) [Chambers et al, 2004; Velicogna and Wahr, 2005; Luthcke et al, 2006; Baur et al, 2009; King et al, 2012]. The qualitative supremacy of GRACE is widely acknowledged, but the quantitative aspect has been questioned from time to time. The stakeholders want a number (linear trend) to portray the rate of ice mass loss in ice-sheets, or a number that narrates the total ground water deficit in California or North-West India. But these numbers come with a large uncertainty, which exists due to GRACE errors and also due to the signal loss from filtering. The expectation of the community is that the uncertainty is reduced to a minimum. I believe that this research work is a successful contribution in this direction.

# 2 UNDERSTANDING THE IMPACT OF FILTERING ON THE SIGNAL

Although filtering in the spectral domain is popular, we prefer the spatial domain for a better visualization and for obtaining a different perspective. First we develop an understanding of convolution at point scale, then we extend it to an entire catchment, which is more relevant for GRACE observed hydrological signals. Filtering affects both the signal and the noise. In this chapter, we discuss the impact of filtering on the signal only. Therefore, the fields used are noise-free hydrological models representing the GRACE type total water storage change. The impact of convolution on the signal is studied by breaking the filtered signal into two parts, namely leakage from outside and change of signal inside. These components are then treated extensively to obtain a relation for the original signal. The discussion is mathematical, and is followed by a closed-loop test to verify the findings.

## §2.1 FILTERING AND REGIONAL AVERAGES

The relation (1.14) is used for spatial averaging. The basic idea behind filtering is embedded in this relation, where we compute a weighted average of the field over the entire globe and allot that value to the calculation point. Therefore, the filtered value $\bar{f}(\cdot)$ at $(\theta, \lambda)$ is influenced by the field values from the entire domain. However, due to weighting, the magnitude of the impact of a data point depends on its distance from the calculation point and the design of the filter kernel as well. For example, the weights for a Gaussian filter decay with spherical distance from the calculation point in an isotropic manner. Therefore, in general nearby points have a stronger impact than distant points.

The GRACE products suffer from limited spatial resolution, therefore we study the hydrological signal at catchment scale [Swenson, 2002; Velicogna and Wahr, 2006b; Longuevergne et al, 2010; Lorenz et al, 2014]. In order to understand the impact of filtering at catchment scale, it is logical to extend the concept of convolution from point scale to catchment scale, but before we do that, let us understand the mathematical definition of catchment averages.

The area integral of the field values inside a region of interest $O$, divided by the area of the region of interest gives us a scalar value $f_c$, known as the catchment average and

written as

$$f_c = \frac{1}{A_c}\int_O f(\theta,\lambda)\,\mathrm{d}\Omega \qquad \text{with the area } A_c = \int_O \mathrm{d}\Omega. \tag{2.1}$$

We can achieve the same goal by introducing a characteristic function $R(\theta,\lambda)$ instead of the region of interest $O$, which, when multiplied with the field $f(\theta,\lambda)$, gives us the field confined to the region $O$. Integrating this product over the domain $\Omega$ gives us the regional average $f_c$. The characteristic function $R(\theta,\lambda)$ is a mask (step function) with the value one inside the region of interest and zero elsewhere (cf. Figure 2.1).

$$R(\theta,\lambda) = \begin{cases} 1, & \text{inside } O \\ 0, & \text{elsewhere} \end{cases}, \tag{2.2a}$$

$$A_c = \int_\Omega R(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{2.2b}$$

Using the characteristic function in (2.1) allows us to evaluate the integral over the globe. Therefore (2.1) is equivalent to

$$f_c = \frac{1}{A_c}\int_\Omega f(\theta,\lambda)\,R(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{2.3}$$

The regional average from a filtered field (denoted by $\bar{f}_c$) can be computed in the same fashion. We just have to replace the field $f(\theta,\lambda)$ with the filtered field $\bar{f}(\theta,\lambda)$:

$$\bar{f}_c = \frac{1}{A_c}\int_\Omega \bar{f}(\theta,\lambda)\,R(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{2.4a}$$

Replacing the field $\bar{f}(\theta,\lambda)$ in (2.4a) with the convolution integral (1.14), we get

$$\bar{f}_c = \frac{1}{4\pi A_c}\int_\Omega\int_{\Omega'} R(\theta,\lambda)\,f(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'\,\mathrm{d}\Omega\,. \tag{2.4b}$$

We will now evaluate the convolution integral over the domain $\Omega$ and then change the variables from $(\theta',\lambda')$ to $(\theta,\lambda)$ to get

$$\bar{f}_c = \frac{1}{A_c}\int_\Omega f(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{2.4c}$$

In (2.4a) the field is filtered, while in (2.4c) the catchment characteristic function is filtered. Both equations yield the same result, which is a consequence of the associative property of the convolution integral. In Figure 2.1, we have shown the filtered catchment characteristic function $\bar{R}(\cdot)$, which when multiplied with field $f(\cdot)$ and integrated over the globe gives

us the regional average from filtered field ($\bar{f}_c$), the output of (2.4c). Therefore, $\bar{f}_c$ includes a weighted contribution from the points outside the catchment, while the points inside the catchments are weighted down. For a deeper understanding, let us break down the problem further.

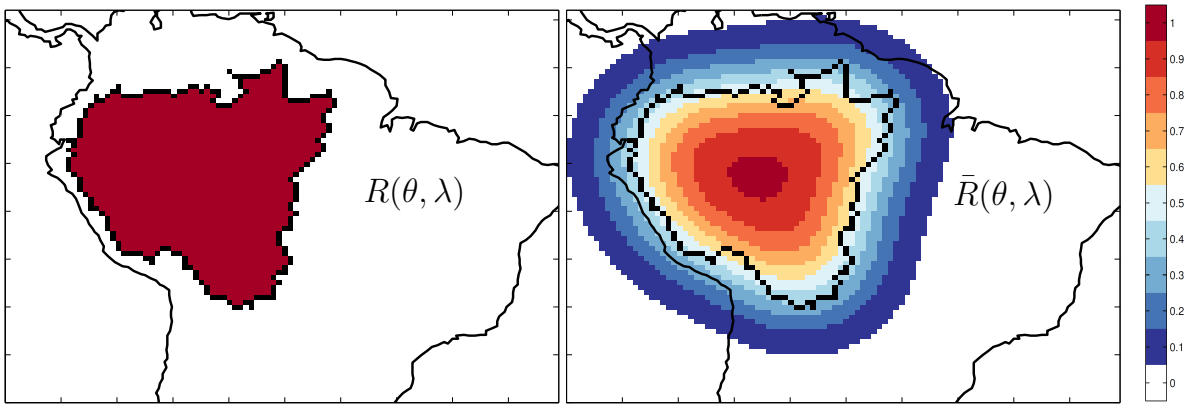


**Figure 2.1:** The catchment characteristic function $R(\cdot)$ and the filtered catchment characteristic function $\bar{R}(\cdot)$ for Amazon. A Gaussian filter of half width radius 500 km is used.

## §2.2 BREAKING DOWN THE PROCESS OF FILTERING

Breaking down the entire domain in points inside and outside the catchment will help us focus on the degradation of the signal inside the catchment and the influence of points outside separately. The catchment characteristic function $R(\cdot)$ represents the points inside, so we introduce $R^*(\cdot)$, the complement of the characteristic function $R(\cdot)$, to represent the points outside the catchment

$$R^*(\theta, \lambda) = 1 - R(\theta, \lambda) = \begin{cases} 0, & \text{inside } O \\ 1, & \text{elsewhere} \end{cases}, \tag{2.5a}$$

$$A_c^* = \int_\Omega R^*(\theta, \lambda)\, \mathrm{d}\Omega = 4\pi - A_c\,. \tag{2.5b}$$

The field $f(\cdot)$ can be written as the sum of the field inside and outside

$$f(\theta, \lambda) = f(\theta, \lambda) R(\theta, \lambda) + f(\theta, \lambda) R^*(\theta, \lambda), \tag{2.6a}$$

$$= F(\theta, \lambda) + F^*(\theta, \lambda), \tag{2.6b}$$

where the term $F(\cdot)$ contains the signal within the catchment and $F^*(\cdot)$ contains the signal outside it (cf. Figure 2.2 column 1). Although, in Figures we use a Gaussian filter for visualization, the discussion is valid for other filters too. We now insert (2.6b) into (1.14),

and perform the convolution:

$$\bar{f}(\theta,\lambda) = \frac{1}{4\pi}\int_{\Omega}\left(F(\theta',\lambda')+F^{*}(\theta',\lambda')\right)b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega', \tag{2.7}$$

$$= \bar{F}(\theta,\lambda)+\bar{F}^{*}(\theta,\lambda), \tag{2.8}$$

where the filtered field $\bar{f}(\cdot)$ contains both the signal from the catchment $\bar{F}(\cdot)$ (although attenuated), and from outside it as well $\bar{F}^{*}(\cdot)$ (cf. Figure 2.2 column 2). Confining our analysis to $O$, which is our goal, (2.8) becomes

$$\bar{f}(\theta,\lambda)R(\theta,\lambda) = \bar{F}(\theta,\lambda)R(\theta,\lambda)+\bar{F}^{*}(\theta,\lambda)R(\theta,\lambda),$$

or

$$\bar{f}_{O}(\theta,\lambda) = \bar{F}_{O}(\theta,\lambda)+l(\theta,\lambda)R(\theta,\lambda). \tag{2.9}$$

The term $\bar{f}_{O}(\cdot)$ is the filtered field over the catchment; it is the sum of the attenuated signal $\bar{F}_{O}(\cdot)$ and the leakage signal $l(\cdot)$ pulled into the region by the filter (cf. Figure 2.2 column 3). Thus, we can write leakage explicitly as follows:

$$l(\theta,\lambda) = \frac{1}{4\pi}\int_{\Omega'} f(\theta',\lambda')R^{*}(\theta',\lambda')b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'. \tag{2.10}$$

Taking the regional average of (2.9) we get

$$\bar{f}_{\mathrm{c}} = \bar{F}_{\mathrm{c}}+l_{\mathrm{c}}. \tag{2.11}$$

The filtered signal inside the region of interest is a sum of attenuated inside field $\bar{F}(\cdot)$ and a leakage field confined to the region $l(\cdot)R(\cdot)$.

The regional average of a quantity can be computed for every epoch. This information stacked in time gives us a time series, which is used to study the behaviour of that quantity over time. Since the time variable gravity information from GRACE is mostly used at time series level, we would treat the problem at hand also at time series level. The relation (2.11) is true for regional averages at an epoch and also at time series level. To this end, we seek the answers to the following questions:

*i.* how are these two components related to the true or original regional average $f_{\mathrm{c}}$?

*ii.* what is the impact on the regional average computed from the filtered field $\bar{f}_{\mathrm{c}}$?

## §2.3 SIGNAL LEAKAGE DUE TO FILTERING

We define leakage as the contribution from signal outside the region of interest, due to convolution only. Beyond this strict definition one can find different perceptions of leakage in the literature, for example, one may include Gibbs's effect into leakage [Swenson and

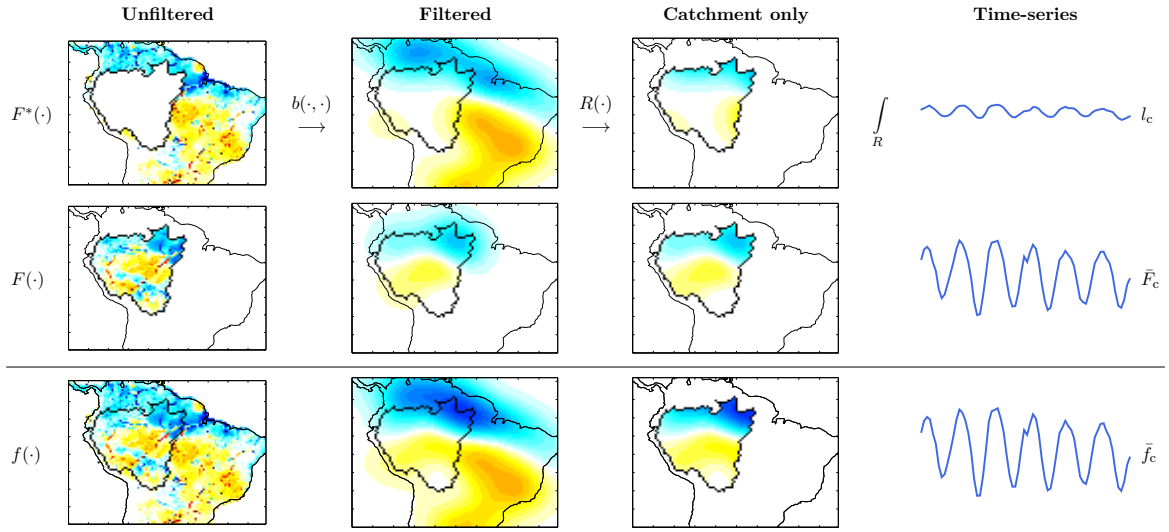


**Figure 2.2:** An illustrative example breaking down the process of filtering. The first and the second rows are spatial complements of each other. The first row shows the leakage, and second row shows the loss in catchment-confined signal. The third row is the sum of the first two rows. Column 1 is filtered to obtain column 2. The field in column 2 is global, thus we extract the field only inside the catchment in column 3. Then the catchment average at each epoch of the field in column 3 constitutes a time series, shown in column 4.

Wahr, 2006], or describe the effect of convolution by different terminology such as bias [Klees et al, 2007]. Apart from regional average specific definitions, there are also filter specific definitions [e.g., Devaraju, 2015].

The leakage field $l(\cdot)$ and its catchment average $l_c$, derived in (2.10) and (2.11) respectively, gives us the total leakage from outside the catchment. However, a catchment is typically surrounded by a number of catchments, each of them having a different hydrological behaviour. Therefore, to see if the leakage is contributed evenly by the entire region $R^*(\cdot)$ or only by certain catchments in $R^*(\cdot)$, we make a small change in the way we define our $R^*(\cdot)$. We break the region $R^*(\cdot)$ into $N$ individual catchments. Thus we can rewrite $R^*(\cdot)$ as a sum of the characteristic functions of those smaller regions as

$$R^*(\theta,\lambda) = \sum_{i=1}^{N} R_i^*(\theta,\lambda). \tag{2.12}$$

Substituting (2.12) in (2.6b), and subsequently deriving the leakage effect, we arrive at

$$\bar{f}_c = \bar{F}_c + \sum_{i=1}^{N} l_i, \tag{2.13}$$

which indicates that the total leakage received is the sum of contribution from all other catchments. To understand better, let us take the South-American catchments in Figure 2.3.

The individual leakage contribution, from only one catchment to the other is called inter-catchment leakage. In order to avoid confusion, let us call the catchment under investigation as the sink catchment, and the catchment whose impact we want to study as the source catchment. Thus, the inter-catchment leakage received by Amazon (sink catchment) only from Orinoco (source catchment), is determined by computing $l_i$ where $i$ corresponds to the Orinoco catchment. If we want to understand various aspects of inter-catchment leakage, such as dependence on catchment size, shape, and location, then we should first analyze the potential inter-catchment leakage. To obtain that, we take a homogeneous unit field only over the source catchment and filter it. Smoothing will spread the energy in the source catchment outside the borders, the portion of spread received by the sink catchment is the potential leakage field. Figure 2.4 shows the potential inter-catchment leakage among the 5 river basins in South-America for a Gaussian filter kernel of 400 km filter radius. We observe that the leakage signal in the sink catchment is strong near catchment boundaries and it decays as soon as we move away from the source catchment. A first impression from Figure 2.4 is that in comparison to a big sink catchment, relatively more area of a small sink catchment is affected by leakage. Therefore, the regional average of the leakage follows a rule of thumb, the bigger the sink catchment the smaller the leakage $l_i$. Catchments which share a longer border leak more into each other. In other words, *proximity* is an important factor when addressing expected leakage from a catchment. Apart from catchment characteristics, the filter function also plays a vital role. The bigger the filter radius the larger the leakage. Although here we have demonstrated the potential leakage with an isotropic Gaussian filter, the shape of the filter kernel (anisotropy) will have an impact on the inter-catchment leakage.

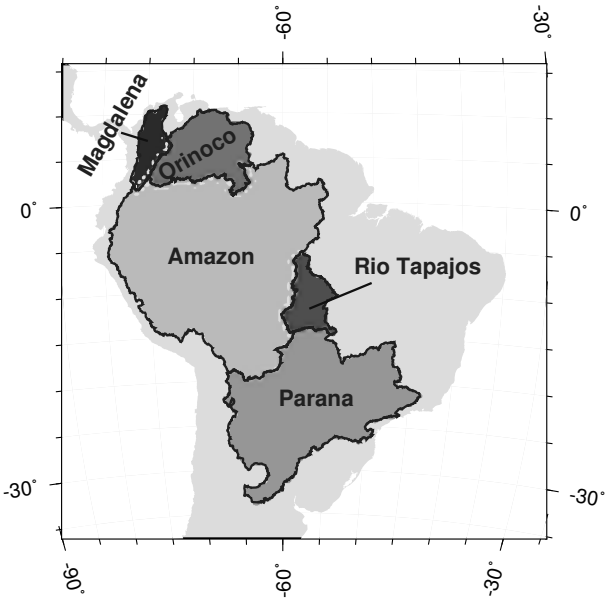


**Figure 2.3:** The South American catchments depicted in the inter-catchment leakage visualization (Figure 2.4).

Till now, we have discussed the inter-catchment leakage at an epoch under the assumption that we have a homogeneous field distribution, which helps us to understand the general behaviour of leakage. However, in reality the hydrological signal is not a uniform homogeneous layer. The field distribution plays an important role and can not be neglected. Therefore, beyond catchment characteristics and filter properties, leakage also depends on the field distribution around the sink catchment. In order to understand this better, we use a Gaussian filter kernel of 400 km half width radius, and monthly WaterGAP Global Hydrology Model (WGHM) fields [Döll et al, 2014], to compute catchment averages of the inter-catchment leakage field, attenuated field, and the WGHM field at every epoch. Then we stack them in time, to get a time series, see Figure 2.5. Indeed we can see that the amplitude of the inter-catchment leakage depends on the proximity of the sink and the source catchments. For example, Rio Tapajos receives a substantial amount of leakage

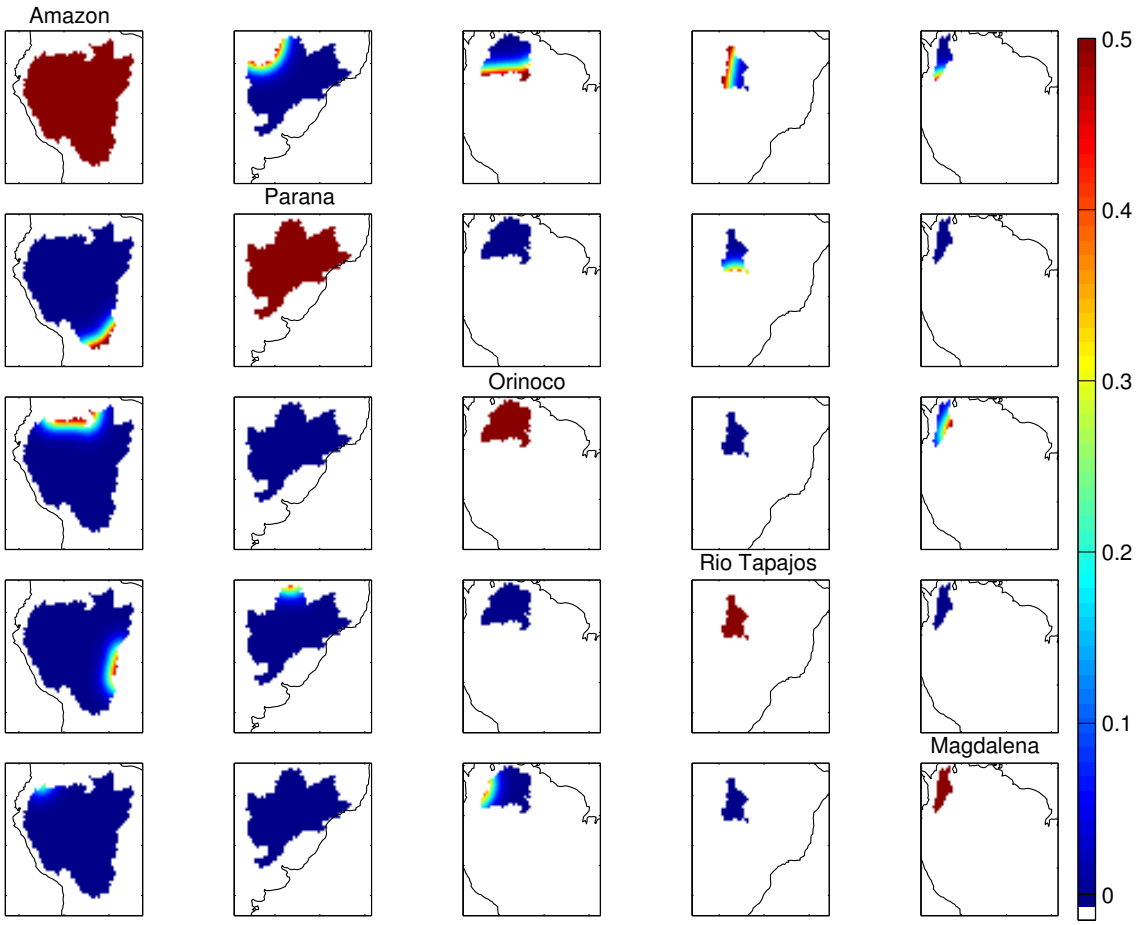


**Figure 2.4**: Visualization of potential leakage between 5 South American catchments. The maps on the diagonal represent the catchment characteristic function $R(\theta, \lambda)$. Each column shows how much leakage the catchment in red receives from other catchments. While each row shows how much the catchment in red contributes to other catchments. The colour bar represents the relative signal strength (dimensionless). Visualizing potential inter-catchment leakage helps us to obtain a preliminary idea of the relative amount of leakage to be expected.

because of its proximity to the Amazon. Apart from this the size of the sink catchment is also important, the amplitude of the leakage, relative to the attenuated signal amplitude in the sink catchment, is small if the area of the sink catchment is large, and *vice-versa* (*cf.* off-diagonal elements of the Figure 2.5).

Each catchment can have a hydrological behaviour different than the others nearby, which implies that the leakage time series from each catchment can be different. For instance, in Figure 2.3 Amazon receives a leakage which is a sum total of leakage from all other individual catchments. Since Amazon, Rio Tapajos, and Parana are in the Southern hemisphere, they have an out of phase signal with respect to Orinoco and Magdalena in the Northern hemisphere. Therefore, the leakage from Orinoco and Magdalena will be out of phase with respect to the leakage from Rio Tapajos and Parana, and *vice-versa*. The resultant leakage received by Amazon will have a different phase with respect to the signal in Amazon. The leakage $l(\cdot)$ added to the filtered confined field $\bar{F}(\cdot)$, gives us the filtered field $\bar{f}(\cdot)$. This is true for regional average at an epoch and also for the time

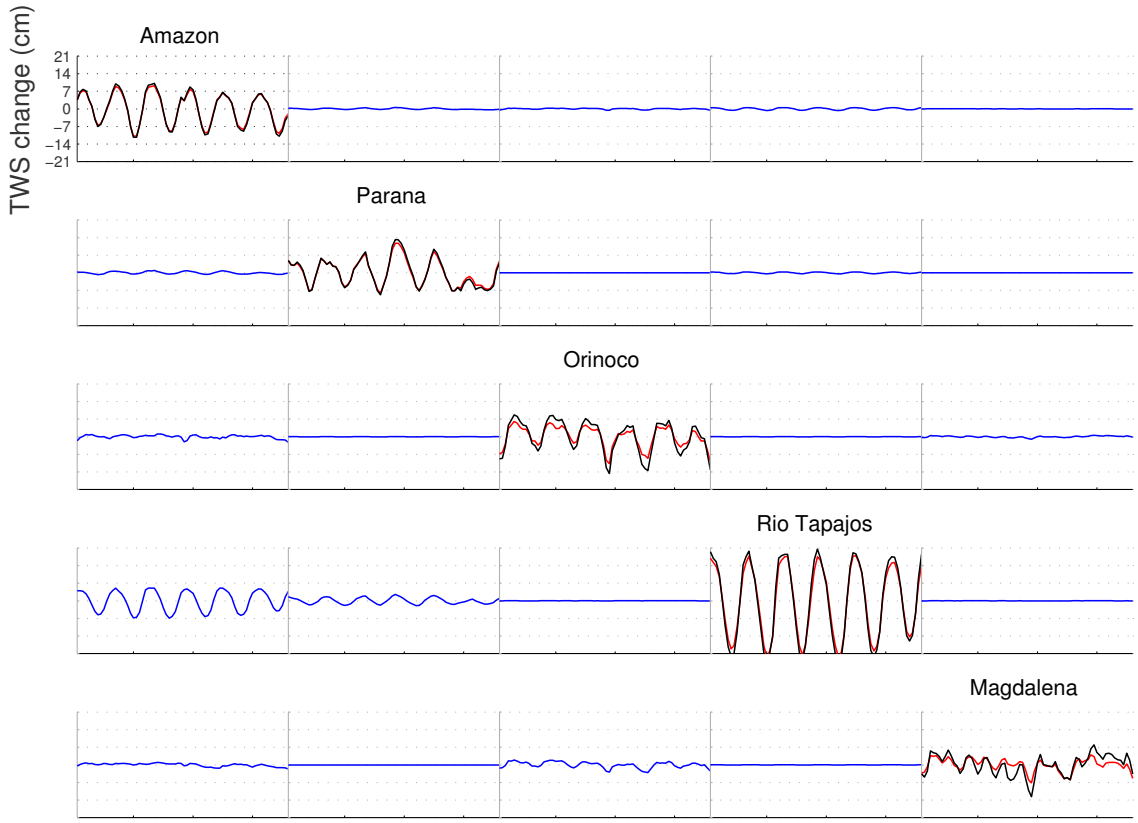


**Figure 2.5**: Inter-catchment leakage: Black represents the true time series from WGHM. Red is the time series from the filtered model and blue in off diagonal plots represents leakage. Each row shows how much leakage the catchment in diagonal receives from others.

series. Let us understand the impact of leakage on the output of filtering in more detail.

## §2.4 UNDERSTANDING EFFECTS OF LEAKAGE

A hydrological time series, typically, has a seasonal signature. Let us assume that both the leakage and the attenuated time series are sinusoids of frequency $\omega$ and phase $\phi$. The leakage time series is represented by $\alpha_l \sin(\omega t + \phi_l)$ and the time series of the attenuated signal by $\alpha_a \sin(\omega t + \phi_a)$. Let us write the time series in the last column of the third row of Figure 2.2 as $\bar{\alpha}\sin(\omega t + \bar{\phi})$, which is equivalent to sum of the leakage time series and the attenuated time series. Hence

$$\bar{\alpha}\sin(\omega t + \bar{\phi}) = \alpha_a \sin(\omega t + \phi_a) + \alpha_l \sin(\omega t + \phi_l) \tag{2.14}$$

using the identity $\sin(x+y) = \sin x \cos y + \sin y \cos x$ to expand the right hand side of (2.14), we get

$$\begin{aligned}
\bar{\alpha}\sin(\omega t + \bar{\phi}) &= \alpha_a[\sin\omega t\ \cos\phi_a + \sin\phi_a \cos\omega t] + \alpha_l[\sin\omega t\ \cos\phi_l + \sin\phi_l \cos\omega t], \\
&= \alpha_a \sin\omega t \cos\phi_a + \alpha_a \sin\phi_a\ \cos\omega t + \alpha_l \sin\omega t \cos\phi_l + \alpha_l \sin\phi_l \cos\omega t, \\
&= (\alpha_a \cos\phi_a + \alpha_l \cos\phi_l)\sin\omega t\ + (\alpha_a \sin\phi_a + \alpha_l \sin\phi_l)\cos\omega t.
\end{aligned} \tag{2.15}$$

Since the phases and the amplitudes are constants, the terms $(\alpha_a \cos\phi_a + \alpha_1 \cos\phi_1)$ and $(\alpha_a \sin\phi_a + \alpha_1 \sin\phi_1)$ are also constant. Hence, let us rewrite

$$(\alpha_a \cos\phi_a + \alpha_1 \cos\phi_1) = \bar{\alpha}\cos\bar{\phi}, \tag{2.16}$$

$$(\alpha_a \sin\phi_a + \alpha_1 \sin\phi_1) = \bar{\alpha}\sin\bar{\phi}. \tag{2.17}$$

Substituting (2.16) and (2.17) in (2.15), and again using the identity $\sin(x+y) = \sin x \cos y + \sin y \cos x$, we get the left hand side of (2.14)

$$\bar{\alpha}\sin(\omega t + \bar{\phi}) = (\alpha_a \cos\phi_a + \alpha_1 \cos\phi_1)\sin\omega t + (\alpha_a \sin\phi_a + \alpha_1 \sin\phi_1)\cos\omega t.$$

Thus the summation of two time series yields another time series with amplitude $\bar{\alpha}$ and phase $\bar{\phi}$. To find the relation for amplitude, take the square of both the sides of (2.16) and (2.17), and then add to get

$$\begin{aligned} &\alpha_a^2 \cos^2\phi_a + \alpha_a^2 \sin^2\phi_a \\ &+\alpha_1^2 \cos^2\phi_1 + \alpha_1^2 \sin^2\phi_1 \\ &+2\alpha_a\alpha_1[\cos\phi_a \cos\phi_1 + \sin\phi_a \sin\phi_1] = \bar{\alpha}^2, \end{aligned} \tag{2.18}$$

using the identity $\sin^2 x + \cos^2 x = 1$, we get the amplitude

$$\bar{\alpha} = \sqrt{{\alpha_a}^2 + {\alpha_1}^2 + 2\alpha_1\alpha_a \cos(\phi_a - \phi_1)}. \tag{2.19}$$

It is evident that the resultant amplitude is a function of the cosine of the phase difference $(\phi_a - \phi_1)$. Since the cosine function is a symmetric function, the sign of $(\phi_a - \phi_1)$ can be ignored. Moreover, the maximum value is obtained for a phase difference of $\pi$ and the minimum is for zero. The resultant amplitude can attain any value between a maximum and a minimum value, given by

$$\bar{\alpha} = \begin{cases} \alpha_a + \alpha_1, & \Delta\phi = 0 \\ \alpha_a - \alpha_1, & \Delta\phi = \pi \end{cases} \tag{2.20}$$

If we divide the equation (2.16) and (2.17), then we get a relation for the resultant phase, given by

$$\tan\bar{\phi} = \frac{\alpha_1 \sin\phi_1 + \alpha_a \sin\phi_a}{\alpha_1 \cos\phi_1 + \alpha_a \cos\phi_a}. \tag{2.21}$$

From (2.19), we can expect the catchments closer to the equator, receiving leakage dominantly from the opposite hemisphere, to suffer a large amplitude reduction. For example, Orinoco lies in the Northern Hemisphere, and it gets leakage from nearby catchments in the Southern Hemisphere, therefore the time series from filtered fields should have minimum attainable value. However, according to (2.21) the phase should not change substantially. Similarly, we can craft an understanding about the impact of filtering

on the amplitude and phase for many catchments over the globe. The important message is that both the amplitude and the phase of the signal time series change after filtering.

We recall that the filtered catchment average $\bar{f}_c$, is the sum of the filtered confined-signal $\bar{F}_c$ and the leakage signal $l_c$ brought in by the filter. In order to compute the true signal, we have to remove the leakage first and then counter the signal change inside the catchment. For that purpose, first we should understand how the signal inside the catchment is affected due to filtering, and then find a way to reverse it.

## §2.5 IMPACT OF FILTERING ON THE CATCHMENT CONFINED SIGNAL

The second row of Figure 2.2 shows that the signal inside the catchment is smothered by the filter. Since we are interested in the regional average of the filtered signal, we loose some signal inside the catchment boundary due to attenuation, which leads to an underestimation of the true regional average. In order to compensate for the loss, researchers have advocated use of a scale factor derived from the catchment characteristic function and the filter function [Velicogna and Wahr, 2006a; Wahr et al, 2007; Longuevergne et al, 2010; King et al, 2012; Vishwakarma et al, 2016].

### Method of scale

This approach helps us to mitigate the amplitude loss, but comes with its own problems. We will now discuss its genesis, underlying assumptions, and limitations. The goal is to relate the attenuated signal $\bar{F}_c$ with the true signal $f_c$, using a scale factor $s$

$$s = \frac{f_c}{\bar{F}_c}. \tag{2.22}$$

It is instructive here to write the explicit expressions for $f_c$ and $\bar{F}_c$

$$s = \frac{\int_\Omega f(\theta,\lambda)R(\theta,\lambda)\mathrm{d}\Omega}{\int_\Omega \bar{F}(\theta,\lambda)R(\theta,\lambda)\mathrm{d}\Omega}. \tag{2.23}$$

In order to evaluate (2.23) we need the true signal $f(\theta,\lambda)$, which is an unknown. However, we are interested in catchment averages only, thus we can replace $f(\theta,\lambda)$ with the catchment average $f_c$ in the numerator, and write it as

$$\frac{1}{A_c}\int_\Omega f(\theta,\lambda)R(\theta,\lambda)\mathrm{d}\Omega = \frac{1}{A_c}\int_\Omega f_c R(\theta,\lambda)\mathrm{d}\Omega. \tag{2.24}$$

The denominator can be rewritten by using the associative property, demonstrated by (2.4c)

$$\frac{1}{A_c}\int_{\Omega}\bar{F}(\theta,\lambda)R(\theta,\lambda)\mathrm{d}\Omega \;=\; \frac{1}{A_c}\int_{\Omega}F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega \;=\; \frac{1}{A_c}\int_{\Omega}f(\theta,\lambda)R(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega. \tag{2.25}$$

If the field $f(\theta,\lambda)$ is nearly uniform inside the catchment, then we can use the uniform layer approximation in (2.25). The approximation is

$$\frac{1}{A_c}\int_{\Omega}f(\theta,\lambda)R(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega \;\approx\; \frac{1}{A_c}\int_{\Omega}f_c R(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega. \tag{2.26}$$

The total water storage change observed by GRACE is dominated by the long wavelength component [Swenson, 2002]. Therefore, for most of the regions dominated by hydrology, the uniform layer approximation is a fair assumption. However, we should remember that it is still an assumption. Substituting (2.24) and (2.26) in (2.23), we get

$$s \;=\; \frac{\displaystyle\int_{\Omega} f_c\, R(\theta,\lambda)\,\mathrm{d}\Omega}{\displaystyle\int_{\Omega} f_c\, R(\theta,\lambda)\,\bar{R}(\theta,\lambda)\mathrm{d}\Omega} \;=\; \frac{\displaystyle\int_{\Omega} R(\theta,\lambda)\,\mathrm{d}\Omega}{\displaystyle\int_{\Omega} R(\theta,\lambda)\,\bar{R}(\theta,\lambda)\mathrm{d}\Omega}. \tag{2.27}$$

In (2.27) $s$ is a data independent catchment specific quantity for a given filter. With the scale factor known, we can approach the true catchment average $f_c$ from $\bar{F}_c$, i.e.

$$f_c \;=\; s\,\bar{F}_c.$$

We now replace $\bar{F}_c$ in (2.11) to get

$$\bar{f}_c \;=\; \frac{f_c}{s} + l_c \quad\Longleftrightarrow\quad f_c = s\,(\bar{f}_c - l_c). \tag{2.28}$$

As per (2.28), first we should subtract $l_c$ from $\bar{f}_c$ and then scale up to reach $f_c$. These entities are defined at an epoch. We can compute them at every epoch and stack them to get a time series. To avoid additional notation, we would refer the time series and regional averages at an epoch by the same notation, i.e. without an explicit time argument.

In (2.28), leakage can only be computed accurately, if we know the true field distribution, which is not possible for GRACE observations. Therefore, most of the research contributions determine leakage from models [Longuevergne et al, 2010; Klees et al, 2007], which brings the uncertainty in the model to (2.28). Furthermore, the scale factors are inversely related to the size of the catchment, which means that the error or uncertainty in smaller catchments is amplified more in comparison to the big catchments. Therefore, such an approach can work for big catchments with uniform total water storage change signal, but for other areas the accuracy will be low and we will not be able to close (2.28).

If we find a way to estimate leakage without using any model, and avoid uniform layer approximation to find a relation which closes (unlike (2.28)), then we can treat the

signal degradation due to filtering comprehensively. Let us take the problem at hand, one by one. Next, we will investigate further to avoid the uniform layer approximation.

**Method of averages and deviation**

Recalling (2.8), we can write the filtered field as

$$\begin{aligned}\bar{f}(\theta,\lambda) &= \frac{1}{4\pi}\int_{\Omega}\left(F(\theta',\lambda')+F^{*}(\theta',\lambda')\right)b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' \\ &= \bar{F}(\theta,\lambda)+\bar{F}^{*}(\theta,\lambda).\end{aligned}$$

Taking the regional average over a region $R(\theta,\lambda)$ we get

$$\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\bar{f}(\theta,\lambda)R(\theta,\lambda)\mathrm{d}\Omega = \frac{1}{A_{\mathrm{c}}}\int_{\Omega}F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega+\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\bar{F}^{*}(\theta,\lambda)R(\theta,\lambda)\,\mathrm{d}\Omega.$$

Our goal is to get the true regional average $f_{\mathrm{c}}$. Now if we write the catchment limited field $F(\cdot)$ as a sum of two global fields: the regional average over the catchment $f_{\mathrm{c}}$, and a deviation $\delta F(\cdot)$ from $f_{\mathrm{c}}$, i.e. $F(\cdot)=f_{\mathrm{c}}+\delta F(\cdot)$, shown in Figure 2.6, then

$$\bar{f}_{\mathrm{c}} = \frac{1}{A_{\mathrm{c}}}\int_{\Omega}(f_{\mathrm{c}}+\delta F(\theta,\lambda))\bar{R}(\theta,\lambda)\mathrm{d}\Omega+\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\bar{F}^{*}(\theta,\lambda)R(\theta,\lambda)\,\mathrm{d}\Omega. \tag{2.29}$$

The last integral on the right hand side of (2.29) is leakage $l_{\mathrm{c}}$, thus the relation becomes

$$\bar{f}_{\mathrm{c}} = \frac{1}{A_{\mathrm{c}}}\int_{\Omega}f_{\mathrm{c}}\bar{R}(\theta,\lambda)\mathrm{d}\Omega+\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\delta F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega+l_{\mathrm{c}}.$$

Since $f_{\mathrm{c}}$ is a constant we can take it out of the first integral to get

$$\bar{f}_{\mathrm{c}} = \frac{f_{\mathrm{c}}}{A_{\mathrm{c}}}\int_{\Omega}\bar{R}(\theta,\lambda)\mathrm{d}\Omega+\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\delta F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega+l_{\mathrm{c}}.$$

Now the total energy under the filter kernel is conserved, which means

$$\int_{\Omega}\bar{R}(\theta,\lambda)\mathrm{d}\Omega=\int_{\Omega}R(\theta,\lambda)\mathrm{d}\Omega=A_{\mathrm{c}}. \tag{2.30}$$

Thus, we can write

$$\bar{f}_{\mathrm{c}} = f_{\mathrm{c}}+\frac{1}{A_{\mathrm{c}}}\int_{\Omega}\delta F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega+l_{\mathrm{c}}.$$

Let us write the deviation integral $\frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega$ as $\overline{\delta F}_c$, then the true regional average is written as

$$f_c = \bar{f}_c - \overline{\delta F}_c - l_c\,. \tag{2.31}$$

Equation (2.31) has no assumptions and approximations. Therefore, the relation must close and we should be able to determine the true regional average $f_c$ from the three terms on the right hand side. This fulfills our quest for a equation that relates the true regional average $f_c$ with the regional average from filtered field $\bar{f}_c$.

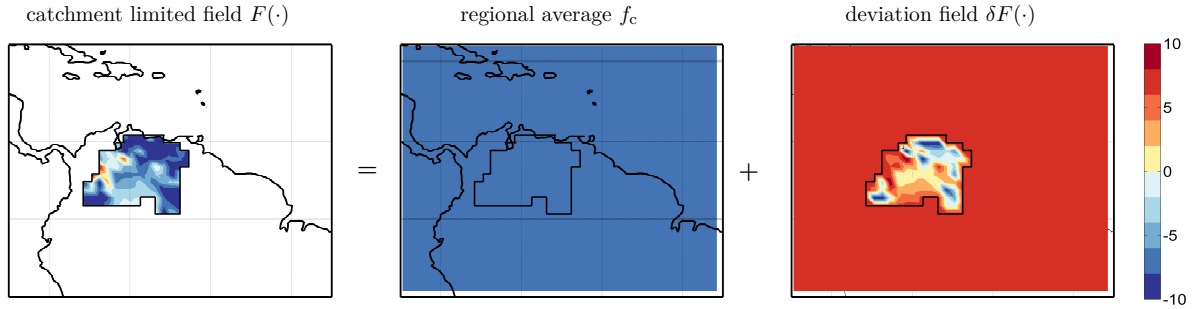


**Figure 2.6:** Breaking the catchment confined field $F(\cdot)$ inside Orinoco into two parts: a global uniform layer with the amplitude equal to the regional average $f_c$ and a deviation field $\delta F(\cdot)$. Please note that the deviation field attains a value of $-f_c$ outside the catchment.

We have established mathematical relations to approach the true regional average $f_c$, however, a numerical test of method of scale and method of averages and deviation, hereafter method of deviation, will demonstrate the accuracy of these relations. But, before we verify our findings in a closed-loop environment, let us describe a few model dependent approaches.

## §2.6 POPULAR MODEL-DEPENDENT APPROACHES

From a number of approaches to correct the filtered GRACE products, we choose the three most popular model-dependent approaches: multiplicative approach by Longuevergne et al [2010], additive approach by Klees et al [2007], and the scaling approach by Landerer and Swenson [2012]. Long et al [2015] compared these three approaches and found that the additive method performs the best. The basic formulation behind these model dependent approaches is given in Table 2.1. We can see that the multiplicative approach is equivalent to the data-driven method of scale, with the difference that the leakage in multiplicative approach is computed from a model $M(\theta,\lambda)$. The leakage from a model can be written as

$$l_c^{\mathrm{m}} = \frac{1}{A_c}\int_\Omega \frac{1}{4\pi}\int_{\Omega'} M(\theta',\lambda')R^*(\theta',\lambda')b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'\,\mathrm{d}\Omega\,. \tag{2.32}$$

| approach | relation |
|---|---|
| multiplicative | $f_c = s(\bar{f}_c - l_c^m)$ |
| additive | $f_c = \bar{f}_c - l_c^m + b_c^m$ |
| scaling | $f_c = k\,\bar{f}_c$ |

**Table 2.1:** Mathematical relations behind three popular model dependent approaches.

The leakage in additive approach is also computed with the help of model using (2.32). However, the catchment-confined signal degradation is treated using a correction term known as bias $b_c$, which is written as

$$b_c = \frac{1}{A_c}\int_{\Omega} M(\theta,\lambda)(R(\theta,\lambda) - \bar{R}(\theta,\lambda))\mathrm{d}\Omega\,. \tag{2.33}$$

The scale $k$ in scaling approach is estimated with an aim to minimize the difference between the quantity from model and from filtered model, in a least squares sense. This approach can be applied to a grid also, which gives us a map of scaling factors [Landerer and Swenson, 2012]. The concept, can be written mathematically as

$$k \;\rightarrow\; \min\{M(\theta,\lambda) - k\bar{M}(\theta,\lambda)\}\,. \tag{2.34}$$

In addition to these three model dependent approaches, a scaling approach from Wahr et al [2007] was used for many studies with GRACE [Fenoglio-Marc et al, 2006; Velicogna and Wahr, 2006b]. The scaling approach amplifies the regional averages from filtered GRACE by $s$ that is same as the scale in (2.27). Since we have proved in (2.28) that a leakage removal is essential before scaling, the approach by Wahr et al [2007] will be effective for catchments receiving negligible leakage, and with a nearly homogeneous and uniform signal.

## §2.7 CLOSED-LOOP VALIDATION

We use the monthly 1° gridded total water storage change fields from WGHM to emulate a GRACE-type signal. A description of the model is provided in Appendix A. A Gaussian filter with half width radius of 400 km is used for filtering, and we perform our computations for 12 catchments given in Table 2.2, over a span of 6 years, from January 2004 to December 2009. In order to validate the theoretical understanding obtained, we will answer the following questions:

*i.* Is it true, as we found in section 2.4, that leakage affects both the amplitude and the phase of the true time series? If yes, then by what amount is the phase of the

time series from filtered fields changed? Is it significant or not? Can scaling alone help us to recover the true time series?

*ii.* The uniform layer approximation leads us to (2.28), which has been advocated for retrieving the true regional average $f_c$. Is this relation accurate enough? What are the limitations that are brought in by the uniform layer assumption?

*iii.* We claim that (2.31) approaches the true regional average. How well does relation (2.31) closes? Is it better than (2.28)?

In order to find answers to these questions, first we compute $\bar{f}_c$, $f_c$, $l_c$, $\overline{\delta F}_c$ and $s$. We then compute $s(\bar{f}_c - l_c)$ and $\bar{f}_c - \overline{\delta F}_c - l_c$, which we refer to as the retrieved time series from the method of scale (2.28) and retrieved time series from the method of deviation (2.31) respectively. These time series are compared with the true time series $f_c$ computed from the model fields. The phase difference between two time series is computed by employing Hilbert transform [Phillips et al, 2012]. Let us suppose, we want to determine the phase difference between two time series $f_c$ and $\hat{f}_c$, then first we express one of them in terms of the other as

$$\hat{f}_c \;=\; a + b\, f_c + c\, \mathrm{Im}\{\, \mathcal{H}\,(f_c)\}.$$

The parameters $a$, $b$, and $c$ are obtained using least squares estimation. The phase difference between two time series is then given by

$$\Delta\phi \;=\; \arctan\!\left(\frac{c}{b}\right).$$

The difference between true time series and estimated time series is represented by the RMS (Root Mean Square), given by

$$\mathrm{RMS} \;=\; \sqrt{\frac{1}{m}\sum_{i=1}^{m}(f_c - \hat{f}_c)^2}. \tag{2.35}$$

In addition to that, we also compute the Nash Sutcliffe Efficiency (NSE) given by

$$\mathrm{NSE} \;=\; 1 - \frac{\sum_{i=1}^{m}(f_c - \hat{f}_c)^2}{\sum_{i=1}^{m}(f_c - \tilde{f}_c)^2}, \tag{2.36}$$

where $f_c$ represents the true value, $\hat{f}_c$ is the time series with which we compare the true time series, $\tilde{f}_c$ is the mean of the true values and $m$ is the number of epochs (months). RMS can attain any positive value, a RMS close to zero represents excellent agreement between $f_c$ and $\hat{f}_c$. NSE can attain any value between $-\infty$ and 1. A positive NSE value indicates that the time series computed is better than the mean of the true time series and a value close to 1 represents excellent agreement between $f_c$ and $\hat{f}_c$ [Nash and Sutcliffe, 1970].

Table 2.2 contains $\Delta\phi$, RMS, and NSE between true time series and a time series computed from filtered fields or its corrected version. We can see that the phase change

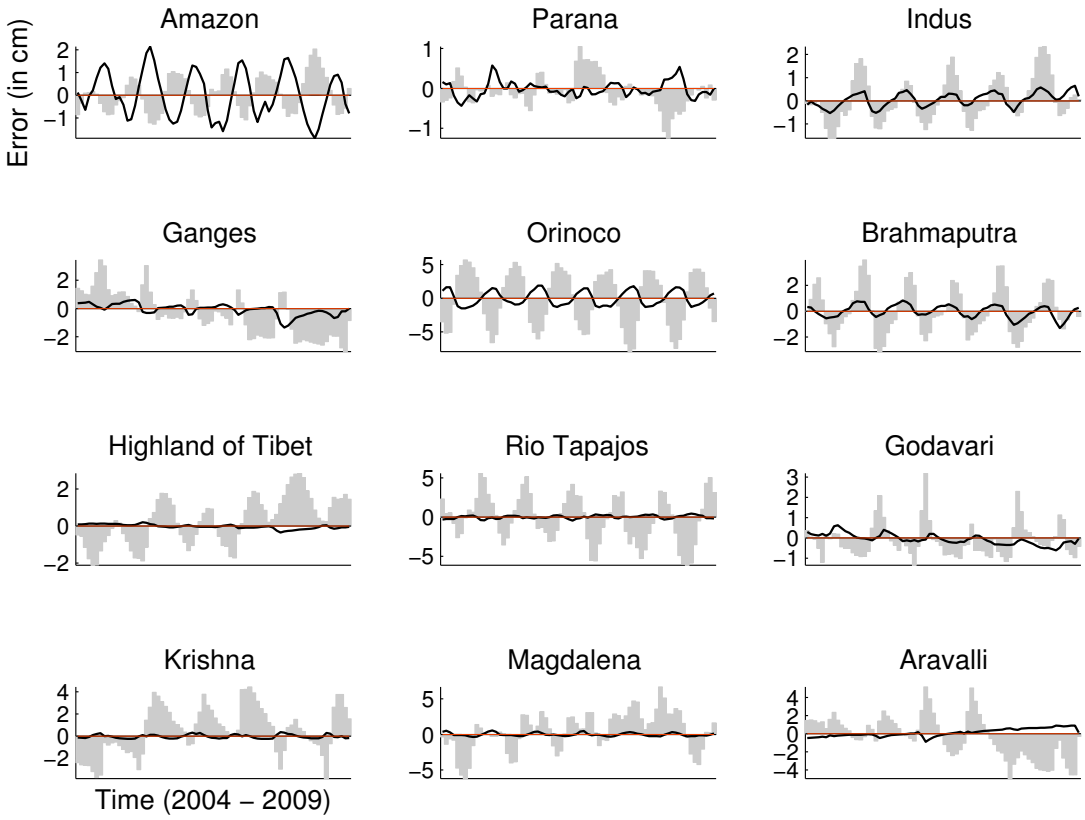


**Figure 2.7:** The total water storage time series computed from WGHM fields is the true time series in our closed-loop simulation. The difference between true time series $f_c$ and that from filtered WGHM fields $\bar{f}_c$ is denoted by the gray bars. The difference between $f_c$ and $s(\bar{f}_c - l_c)$ (2.28) is represented by the black line, and the difference between the true time series $f_c$ and that recovered by (2.31) is in red.

introduced by filtering ($\phi - \bar{\phi}$) is significant. For the Highland of Tibet it even amounts up to 22.6°(≈ 23 days). Such a huge shift in the phase means that the time series is shifted nearly by a month. The phase difference for other catchments, such as Rio Tapajos, Brahmaputra, and Krishna, is also substantial. Since there have been several attempts to recover the signal loss by scaling alone [Velicogna and Wahr, 2006a; Wahr et al, 2007; Landerer and Swenson, 2012], we also compare a scaled version of the time series from filtered fields with the true time series. The corresponding scale factors are given in the last column. We can clearly see that due to non-negligible magnitude of the phase difference, a simple scaling approach is ineffective. Furthermore, it increases the disagreement between the corrected time series and the true time series, and we can see that the time series from filtered fields are better than the scaled time series.

**Table 2.2:** Absolute phase difference, RMS (cm), and NSE between two total water storage change time series. These time series are generated from WGHM fields and their filtered version for catchments in the first column. The last column shows scale factors for a Gaussian filter with averaging radius of 400 km. Catchments are sorted according to their area. The best value is in bold.

| | Between $f_c$ and $\bar{f}_c$ | | | Between $f_c$ and $s\bar{f}_c$ | | | Between $f_c$ and $s(\bar{f}_c - l_c)$ | | | Between $f_c$ and $\bar{f}_c - \overline{\delta F}_c - l_c$ | | | $s$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Catchment | $\Delta\phi$ | RMS | NSE | $\Delta\phi$ | RMS | NSE | $\Delta\phi$ | RMS | NSE | $\Delta\phi$ | RMS | NSE | |
| Amazon | 0.6° | 0.79 | **0.99** | 0.6° | 1.86 | 0.95 | 1.9° | 1.00 | 0.99 | **0°** | **0** | **1.00** | 1.32 |
| Parana | 2.2° | 0.41 | 0.99 | 2.2° | 1.61 | 0.81 | 0.0° | 0.20 | **1.00** | **0°** | **0** | **1.00** | 1.49 |
| Indus | 0.3° | 0.96 | 0.88 | 0.3° | 2.50 | 0.20 | 1.0° | 0.30 | 0.99 | **0°** | **0** | **1.00** | 1.82 |
| Ganges | 2.6° | 1.61 | 0.95 | 2.6° | 4.46 | 0.60 | 0.5° | 0.41 | **1.00** | **0°** | **0** | **1.00** | 1.98 |
| Orinoco | 2.7° | 4.17 | 0.89 | 2.7° | 5.93 | 0.78 | 0.2° | 1.07 | 0.99 | **0°** | **0** | **1.00** | 2.09 |
| Brahmaputra | 12.8° | 1.78 | 0.89 | 12.8° | 11.40 | -0.64 | 1.0° | 0.48 | 0.99 | **0°** | **0** | **1.00** | 2.69 |
| Highland of Tibet | 22.6° | 1.33 | 0.58 | 22.6° | 2.66 | -3.60 | 0.2° | 0.12 | **1.00** | **0°** | **0** | **1.00** | 2.49 |
| Rio Tapajos | 11.3° | 3.05 | 0.95 | 11.3° | 29.74 | -4.01 | 0.6° | 0.21 | **1.00** | **0°** | **0** | **1.00** | 3.24 |
| Godavari | 2.1° | 0.85 | 0.98 | 2.1° | 12.35 | -3.35 | 1.6° | 0.27 | **1.00** | **0°** | **0** | **1.00** | 3.44 |
| Krishna | 11.9° | 2.24 | 0.84 | 11.9° | 13.35 | -4.56 | 4.0° | 0.17 | **1.00** | **0°** | **0** | **1.00** | 4.05 |
| Magdalena | 4.7° | 2.70 | 0.73 | 4.7° | 14.56 | -6.91 | 0.1° | 0.23 | **1.00** | **0°** | **0** | **1.00** | 4.06 |
| Aravalli | 5.6° | 2.36 | 0.85 | 5.6° | 15.01 | -3.95 | 1.0° | 0.43 | 0.99 | **0°** | **0** | **1.00** | 5.01 |
| Mean | 6.6° | 1.90 | 0.88 | 6.6° | 9.62 | -1.97 | 1.0° | 0.41 | 0.99 | **0°** | **0** | **1** | |

The result from the method of scale displays a substantial improvement in the phase and the amplitude. Thus, removal of leakage prior to scaling is important for the accuracy of the corrected time series. However, the relation does not close, and a small error in the phase and the amplitude is still present. The result from the method of averages and deviation demonstrates a perfect reproduction of both the phase and the amplitude. This corroborates the correctness and efficacy of the theoretical foundation laid.

We discussed earlier that due to out-of-phase leakage, Orinoco is expected to suffer a larger deviation in the amplitude. In Figure 2.7, we can indeed see that the difference between $f_c$ and $\bar{f}_c$ (the gray bars) is maximum for Orinoco, corresponding to which, in Table 2.2, the RMS between $f_c$ and $\bar{f}_c$ is also maximum.

In this chapter we studied the signal damage due to filtering, and established relations to repair it. However, its realization for GRACE products is a challenge in itself. There are two stumbling blocks in front of us:

*i.* The computation of leakage $l_c$ and $\overline{\delta F}_c$ require the knowledge of the true field, which is unknown in reality. This is the reason, why many researchers advocate

usage of models to compute leakage, bias or scale factors. However, we want to refrain from such practice, which motivates us to develop a data-driven method for determining these quantities.

*ii.* The GRACE products are contaminated with a large amount of noise. We assume that it is filtered out, but a perfect filter is fictional. Therefore, we expect to have some noise left, whose magnitude is dictated by the filter definition. However, noise should be included and treated in the general mathematical framework established here, so that we are able to demarcate the limitations pertinent to a filter.

These two aspects are treated in the next chapter, which will help us to come up with a comprehensive method to correct the filtered GRACE products.

# 3 A TOOL FOR NOISY FIELDS

In this chapter we aim to develop a comprehensive method to approach the true value of the regional average from the monthly GRACE products alone. For this purpose, first of all we introduce noise in the developed mathematical framework to identify its impact, and then craft a data-driven strategy to estimate leakage $l_c$ and the deviation integral $\overline{\delta F}_c$. The method evolves with the support from few assumptions and approximations, which are tested exhaustively to design a reliable method.

## §3.1 IMPACT OF NOISE

The relations (2.28) and (2.31) were derived for the noise free case, however, GRACE monthly products are contaminated with noise. Let us denote a GRACE field by $g(\theta,\lambda)$, which can be written as a summation of the signal $f(\theta,\lambda)$ and noise $n(\theta,\lambda)$, i.e.

$$g(\theta,\lambda) = f(\theta,\lambda) + n(\theta,\lambda). \tag{3.1}$$

The filtered GRACE field can be written as

$$\bar{g}(\theta,\lambda) = \frac{1}{4\pi}\int_{\Omega'} g(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' = \frac{1}{4\pi}\int_{\Omega'} \Big(f(\theta',\lambda') + n(\theta',\lambda')\Big)\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'.$$

This convolution integral can be written in two parts, one for the signal and the other for noise

$$\bar{g}(\theta,\lambda) = \frac{1}{4\pi}\int_{\Omega'} f(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' + \frac{1}{4\pi}\int_{\Omega'} n(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'.$$

Next, we break both the signal and noise into two spatially complementary parts like before: one inside and the other outside, to get

$$\begin{aligned}\bar{g}(\theta,\lambda) = {} & \frac{1}{4\pi}\int_{\Omega'} F(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' + \frac{1}{4\pi}\int_{\Omega'} F^*(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' \\ & + \frac{1}{4\pi}\int_{\Omega'} N(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega' + \frac{1}{4\pi}\int_{\Omega'} N^*(\theta',\lambda')\,b(\theta,\lambda,\theta',\lambda')\,\mathrm{d}\Omega'. \end{aligned}\tag{3.2}$$

The first integral on the right hand side of (3.2) gives us the attenuated catchment-confined signal, the second integral is leakage, the third integral gives us the attenuated catchment-confined noise, and the fourth integral is the noise that comes with leakage. These terms can be abbreviated and written as

$$\bar{g}(\theta,\lambda) \;=\; \bar{F}(\theta,\lambda) + l(\theta,\lambda) + \bar{N}(\theta,\lambda) + n^{l}(\theta,\lambda). \tag{3.3}$$

Taking the regional average of (3.3), we get a relation similar to (2.11), which is

$$\bar{g}_{\mathrm{c}} \;=\; \bar{F}_{\mathrm{c}} + l_{\mathrm{c}} + \bar{N}_{\mathrm{c}} + n^{l}_{\mathrm{c}}. \tag{3.4}$$

In order to counter the attenuation of catchment-confined quantity, we can follow the method of scale or the method of deviation. If we follow the uniform layer assumption for both the signal and the noise then we reach a mathematical relation similar to the relation obtained by **method of scale** (2.28)

$$\bar{g}_{\mathrm{c}} \;=\; \frac{1}{s}(f_{\mathrm{c}} + n_{\mathrm{c}}) + l_{\mathrm{c}} + n^{l}_{\mathrm{c}}. \tag{3.5}$$

On the other hand, if we write the inside-field $F(\theta,\lambda)$ and the inside-noise $N(\theta,\lambda)$ in (3.2) as a sum of the regional average and a deviation from the respective quantity, then after taking the regional average we reach a mathematical relation similar to the relation obtained by **method of deviation** (2.31). It reads as

$$\bar{g}_{\mathrm{c}} \;=\; f_{\mathrm{c}} + n_{\mathrm{c}} + \frac{1}{A_{\mathrm{c}}}\int\limits_{\Omega} \delta F(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega + \frac{1}{A_{\mathrm{c}}}\int\limits_{\Omega} \delta N(\theta,\lambda)\bar{R}(\theta,\lambda)\mathrm{d}\Omega + l_{\mathrm{c}} + n^{l}_{\mathrm{c}},$$

or

$$\bar{g}_{\mathrm{c}} \;=\; f_{\mathrm{c}} + n_{\mathrm{c}} + \overline{\delta F}_{\mathrm{c}} + \overline{\delta N}_{\mathrm{c}} + l_{\mathrm{c}} + n^{l}_{\mathrm{c}}, \tag{3.6}$$

where $\bar{g}_{\mathrm{c}}$ is the catchment average of the filtered mass change from GRACE, $n_{\mathrm{c}}$ is the aggregated noise over the catchment only, $n^{l}_{\mathrm{c}}$ is the noise term in the leakage computed from unfiltered GRACE fields, and $\overline{\delta N}_{\mathrm{c}}$ is the integral of the catchment-limited noise deviation. All other symbols have their usual meanings.

Since the noise in GRACE fields is unknown, the noise components in equation (3.5) and (3.6) can not be determined. However, if we assume that the filter dampens the noise to a large extent, then the averages of dampened noise over the region ($n_{\mathrm{c}}$ and $\overline{\delta N}_{\mathrm{c}}$) can be safely neglected. But the leakage noise $n^{l}_{\mathrm{c}}$ brought into the catchment along with the leakage signal remains a major threat to the accuracy. If we use a noise-free field to compute leakage, then we can avoid $n^{l}_{\mathrm{c}}$. This is the reason many studies advocated using a hydrological model to compute leakage $l_{\mathrm{c}}$ [Klees et al, 2007; Longuevergne et al, 2010].

In Figure 3.1 and 3.2 we demonstrate that using a model brings in error. We plot the difference between true time series from WGHM fields and the corrected time series using (2.28), but computing leakage from GLDAS or MERRA-Land model. Similarly, the error in corrected time series from (2.31) are shown in Figure 3.2, where we compute both the leakage $l_{\mathrm{c}}$ and the deviation integral $\overline{\delta F}_{\mathrm{c}}$ from a model. Even in this noise-free closed-loop

test we can clearly see that the difference in models is reflected as error. Furthermore, for certain catchments one model is better than the other, but one can not declare a single model superior and use it. This motivates us to develop a data-driven approach for estimating leakage $l_c$ and the deviation integral $\overline{\delta F}_c$.

As a first approximation, we propose to use filtered fields for computing leakage or the deviation integral. In Figures 3.1 and 3.2, we have also plotted the error in corrected time series using leakage and the deviation integral from filtered fields. It is consistently less compared to when using a model. However, the magnitude of error is not negligible. This is due to the fact that filtering affects the original field, thus the leakage time series and the deviation integral time series from them are also affected. Nevertheless, a modified information of the original field distribution is present, which helps us to perform better than models. If we can approach the original leakage and the original deviation integral, then we can improve our performance even further. In order to achieve this, we must obtain a better understanding of leakage and the deviation integral from filtered fields. Let us first analyze the leakage from filtered fields.

## §3.2 LEAKAGE FROM FILTERED FIELDS

The catchment average of the leakage from the true field, $l_c$, is explicitly given by

$$l_c = \frac{1}{A_c} \int_{\Omega} \frac{R(\theta,\lambda)}{4\pi} \int_{\Omega'} f(\theta',\lambda') R^*(\theta',\lambda') b(\theta,\lambda,\theta',\lambda') \, \mathrm{d}\Omega' \, \mathrm{d}\Omega. \tag{3.7}$$

Using the associative property of the convolution integral yields

$$l_c = \frac{1}{A_c} \int_{\Omega} f(\theta,\lambda) R^*(\theta,\lambda) \bar{R}(\theta,\lambda) \mathrm{d}\Omega \,. \tag{3.8}$$

Replacing $R^*(\theta,\lambda)\bar{R}(\theta,\lambda)$ by $\kappa(\theta,\lambda)$, we can rewrite equation (3.8) as

$$l_c = \frac{1}{A_c} \int_{\Omega} f(\theta,\lambda) \kappa(\theta,\lambda) \mathrm{d}\Omega \,. \tag{3.9}$$

Following the above proposed idea of data-driven leakage correction, let us assess what happens if the field $f(\theta,\lambda)$ is replaced with the filtered GRACE field $\bar{g}(\theta,\lambda)$:

$$\bar{l}_c = \frac{1}{A_c} \int_{\Omega} \frac{R(\theta,\lambda)}{4\pi} \int_{\Omega'} \bar{g}(\theta',\lambda') R^*(\theta',\lambda') b(\theta,\lambda,\theta',\lambda') \, \mathrm{d}\Omega' \quad \mathrm{d}\Omega \,.$$

Using the associative property of the convolution integral we get

$$\bar{l}_c = \frac{1}{A_c} \int_{\Omega'} \bar{g}(\theta',\lambda') R^*(\theta',\lambda') \bar{R}(\theta',\lambda') \, \mathrm{d}\Omega \,, \tag{3.10}$$

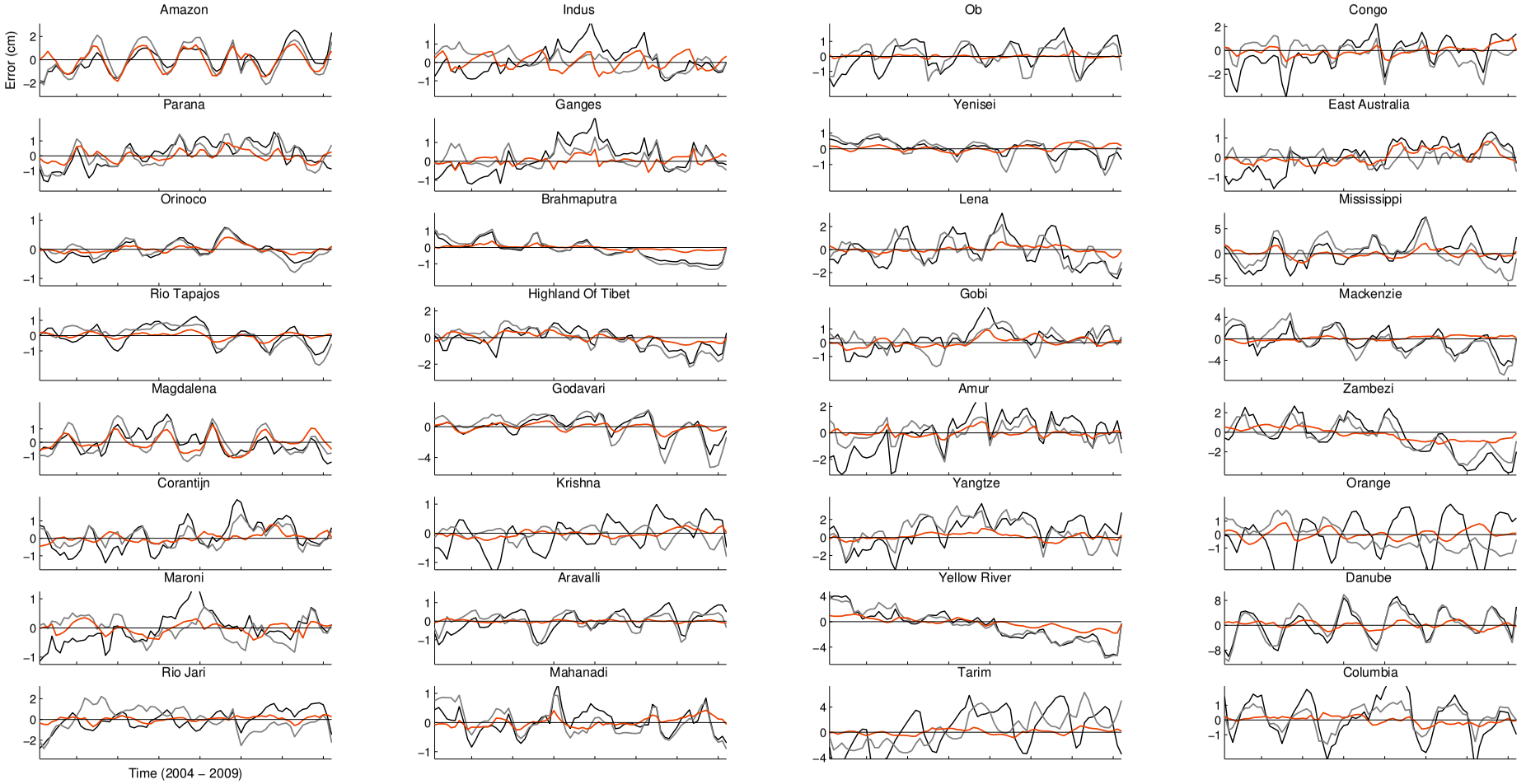


**Figure 3.1**: Error in the corrected time series from (2.28), which is computed by subtracting the original time series from corrected time series. The black line denotes the error in corrected time series using leakage from GLDAS model, and gray line denotes the error in corrected time series using leakage from MERRA-Land model. The red line denotes the error in the corrected time series using leakage from the filtered WGHM fields. A map of the catchments is given in Figure 3.10

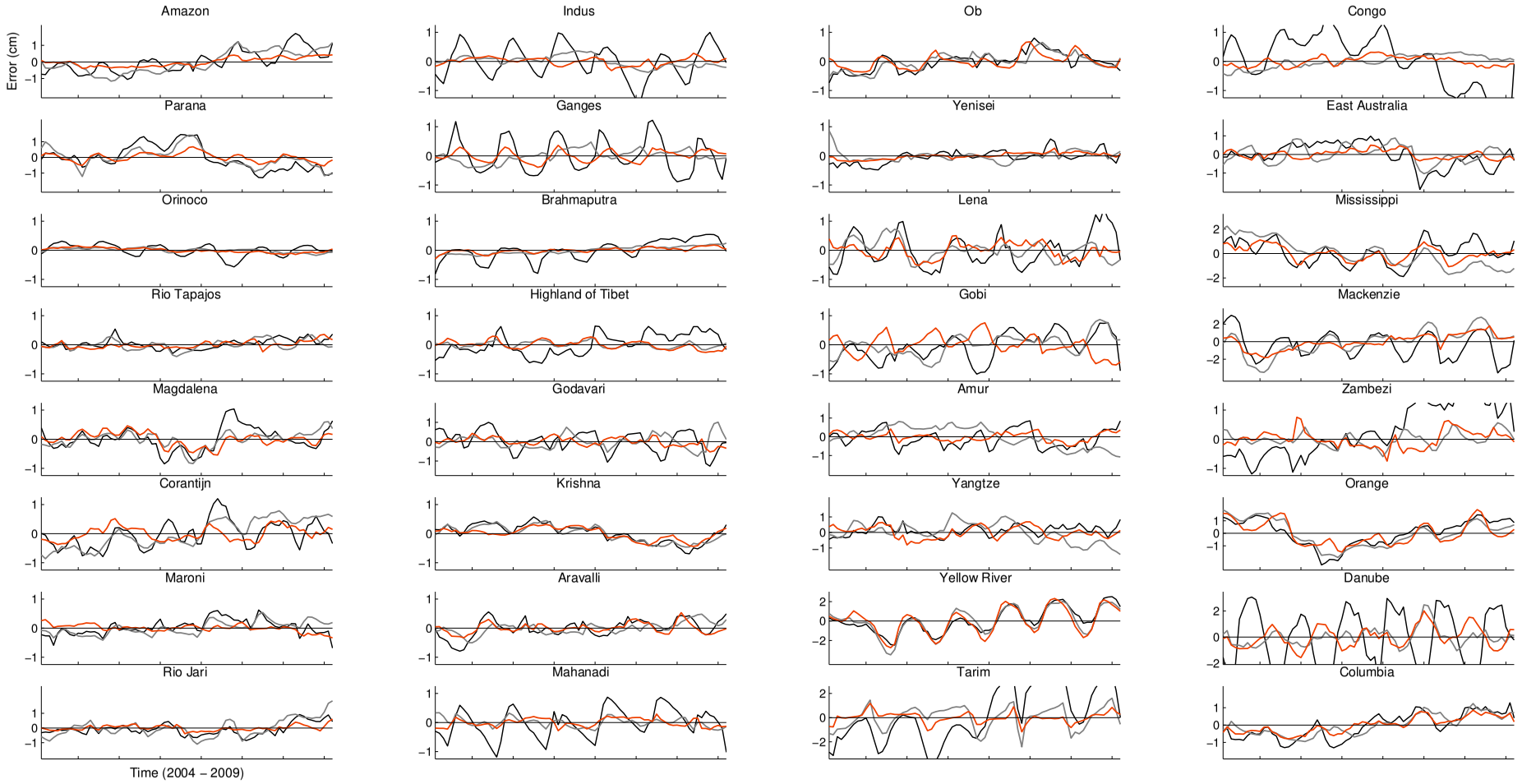


**Figure 3.2:** Error in the corrected time series from (2.31), which is computed by subtracting the original time series from corrected time series. The black line denotes the error in corrected time series using leakage and the deviation integral from GLDAS model, and gray line denotes the error in corrected time series using leakage and the deviation integral from MERRA-Land model. The red line denotes the error in the corrected time series using leakage and the deviation integral from the filtered WGHM fields. A map of the catchments is given in Figure 3.10

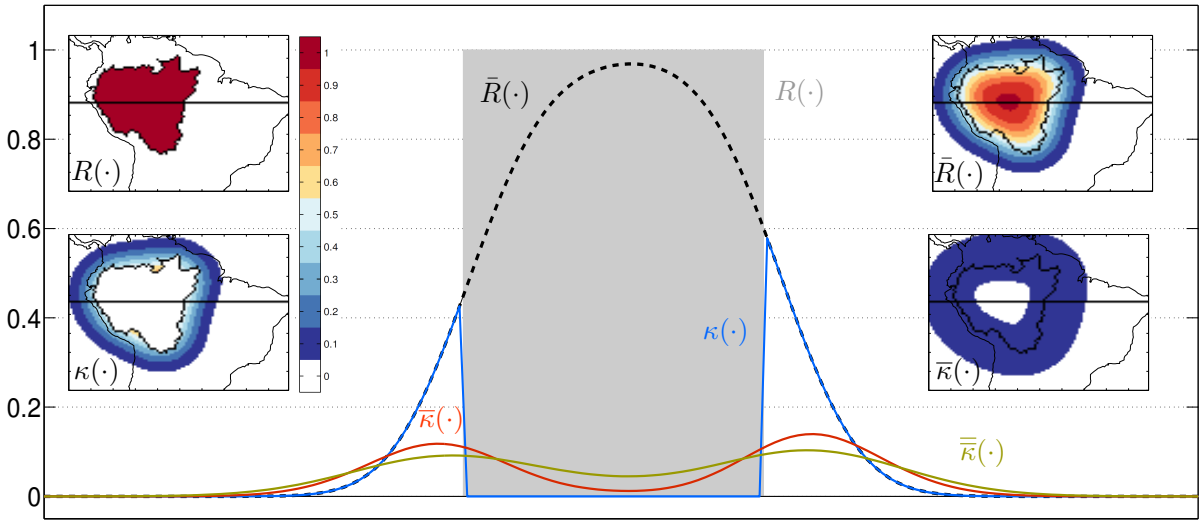


**Figure 3.3:** (a): 1-D visualisation of $R(\cdot)$, $\bar{R}(\cdot)$, $\kappa(\cdot)$, $\overline{\kappa}(\cdot)$, and $\overline{\overline{\kappa}}(\cdot)$ for Amazon. The corresponding 2-D functions are shown in subfigures. The 1 dimensional plots for various quantities are extracted along the black horizontal line dissecting the catchment.

where we now replace $\bar{g}(\theta', \lambda')$ with a convolution integral between the noisy field $g(\theta, \lambda)$ and the filter kernel $b(\theta, \lambda, \theta', \lambda')$:

$$\bar{I}_c = \frac{1}{A_c} \int_{\Omega} g(\theta, \lambda) \quad \frac{1}{4\pi} \int_{\Omega'} \left[ R^*(\theta', \lambda') \bar{R}(\theta', \lambda') \right] b(\theta, \lambda, \theta', \lambda') \mathrm{d}\Omega' \mathrm{d}\Omega \,. \tag{3.11}$$

Again replacing $[R^*(\theta, \lambda)\bar{R}(\theta, \lambda)]$ by $\kappa(\theta, \lambda)$, its filtered version by $\overline{\kappa}(\theta, \lambda)$, and $g(\theta, \lambda)$ by $[f(\theta, \lambda) + n(\theta, \lambda)]$ we can rewrite equation (3.11) as

$$\bar{I}_c = \frac{1}{A_c} \int_{\Omega} \left[ f(\theta, \lambda) + n(\theta, \lambda) \right] \overline{\kappa}(\theta, \lambda) \, \mathrm{d}\Omega, \tag{3.12a}$$

$$= \frac{1}{A_c} \int_{\Omega} f(\theta, \lambda) \, \overline{\kappa}(\theta, \lambda) \, \mathrm{d}\Omega \; + \; \frac{1}{A_c} \int_{\Omega} n(\theta, \lambda) \, \overline{\kappa}(\theta, \lambda) \, \mathrm{d}\Omega. \tag{3.12b}$$

The distribution of $\overline{\kappa}$ is shown in Figure 3.3. It decays as we go away from the borders of the catchment. In the second integral of the equation (3.12b), noise is weighted down by $\overline{\kappa}$ and averaged over the globe. Since the global aggregate of dampened noise is negligible and we can write

$$\bar{I}_c = \frac{1}{A_c} \int_{\Omega} f(\theta, \lambda) \, \overline{\kappa}(\theta, \lambda) \, \mathrm{d}\Omega \tag{3.13a}$$

$$= \frac{1}{A_c} \int_{\Omega} \bar{f}(\theta, \lambda) \, \kappa(\theta, \lambda) \, \mathrm{d}\Omega. \tag{3.13b}$$

Similarly, the catchment average of the leakage computed from twice filtered field can be written as

$$\bar{\bar{l}}_c = \frac{1}{A_c}\int_\Omega f(\theta,\lambda)\,\overline{\overline{\kappa}}(\theta,\lambda)\,\mathrm{d}\Omega \tag{3.14a}$$

$$= \frac{1}{A_c}\int_\Omega \bar{f}(\theta,\lambda)\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega \tag{3.14b}$$

$$= \frac{1}{A_c}\int_\Omega \bar{\bar{f}}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega. \tag{3.14c}$$

Using (2.8) to replace $\bar{f}(\theta,\lambda)$ in (3.13b), we get

$$\bar{l}_c = \frac{1}{A_c}\int_\Omega [\bar{F}(\theta,\lambda)+\bar{F}^*(\theta,\lambda)]\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega$$

$$= \frac{1}{A_c}\int_\Omega \bar{F}(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \bar{F}^*(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{3.15}$$

We can write the field $\bar{F}^*(\theta,\lambda)$ as the leakage field $l(\theta,\lambda)$, cf. (2.9), and use the associative property in the first integral to get

$$\bar{l}_c = \frac{1}{A_c}\int_\Omega F(\theta,\lambda)\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{3.16}$$

The catchment-confined field $F(\theta,\lambda)$ can be written as the sum of the regional average $f_c$ and the deviation $\delta F(\theta,\lambda)$ to get

$$\bar{l}_c = \frac{1}{A_c}\int_\Omega [f_c+\delta F(\theta,\lambda)]\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega\,,$$

$$= \frac{1}{A_c}\int_\Omega f_c\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega\,,$$

$$= \frac{f_c}{A_c}\int_\Omega \overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\kappa(\theta,\lambda)\,\mathrm{d}\Omega\,. \tag{3.17}$$

Similarly, using (2.8) to replace $\bar{f}(\theta,\lambda)$ in (3.14b), we get

$$\bar{\bar{l}}_c = \frac{1}{A_c}\int_\Omega [(\bar{F}(\theta,\lambda)+\bar{F}^*(\theta,\lambda)]\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega\,.$$

Following the same procedure as in (3.15) to (3.17), we obtain a relation for leakage from the fields that have been filtered twice:

$$\bar{\bar{l}}_c = \frac{f_c}{A_c}\int_\Omega \overline{\overline{\kappa}}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\overline{\overline{\kappa}}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega\,, \tag{3.18}$$

Observing (3.17) and (3.18) closely, we find that the leakage from once and twice filtered fields is a sum of damped version of the true leakage (third integral) and a weighted fraction of the field inside the region of interest (first and the second integral). This is true for an epoch and also for time series. Now let us examine the deviation integral from filtered fields and then interpret the physical meaning of these relations.

## §3.3 THE DEVIATION INTEGRAL FROM FILTERED FIELDS

The deviation integral from the filtered field can be written as

$$\frac{1}{A_c}\int_\Omega \delta\bar{F}(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega \;=\; \frac{1}{A_c}\int_\Omega [\bar{f}(\theta,\lambda)\,R(\theta,\lambda) - \bar{f}_c]\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega\,, \tag{3.19}$$

where $\delta\bar{F}(\theta,\lambda)$ is the deviation computed from the filtered field. Using (2.8) to replace $\bar{f}(\theta,\lambda)$, and writing the left hand side of (3.19) as $\overline{\delta\bar{F}}_c$, we get

$$\overline{\delta\bar{F}}_c \;=\; \frac{1}{A_c}\int_\Omega [(\bar{F}(\theta,\lambda) + \bar{F}^*(\theta,\lambda))\,R(\theta,\lambda) - \bar{f}_c]\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega\,.$$

The constant term $\bar{f}_c$ can come out of the integral to give

$$\begin{aligned}\overline{\delta\bar{F}}_c \;&=\; \frac{1}{A_c}\int_\Omega [\bar{F}(\theta,\lambda)\,R(\theta,\lambda) + \bar{F}^*(\theta,\lambda)\,R(\theta,\lambda)]\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega - \bar{f}_c \\ &=\; \frac{1}{A_c}\int_\Omega \bar{F}(\theta,\lambda)\,R(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \bar{F}^*(\theta,\lambda)\,R(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega - \bar{f}_c\,.\end{aligned}$$

Writing $R(\theta,\lambda)\,\bar{R}(\theta,\lambda)$ as $\gamma$, and using the associative property of the convolution integral in the first integral we get

$$\overline{\delta\bar{F}}_c \;=\; \frac{1}{A_c}\int_\Omega F(\theta,\lambda)\,\bar{\gamma}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \bar{F}^*(\theta,\lambda)\,\gamma\,\mathrm{d}\Omega - \bar{f}_c\,. \tag{3.20}$$

We can write the field $\bar{F}^*(\theta,\lambda)$ as the leakage field $l(\theta,\lambda)$ and the field $F(\theta,\lambda)$ as $f_c + \delta F(\theta,\lambda)$ to obtain

$$\overline{\delta\bar{F}}_c \;=\; \frac{1}{A_c}\int_\Omega [f_c + \delta F(\theta,\lambda)]\,\bar{\gamma}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\,\gamma\,\mathrm{d}\Omega - \bar{f}_c\,. \tag{3.21}$$

$$=\; \frac{1}{A_c}\int_\Omega f_c\,\bar{\gamma}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\,\bar{\gamma}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\,\gamma\,\mathrm{d}\Omega - \bar{f}_c\,. \tag{3.22}$$

Similarly, the deviation integral from the twice filtered field reads

$$\overline{\delta\bar{\bar{F}}}_c \;=\; \frac{1}{A_c}\int_\Omega f_c\,\bar{\bar{\gamma}}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\,\bar{\bar{\gamma}}\,\mathrm{d}\Omega + \frac{1}{A_c}\int_\Omega l(\theta,\lambda)\,\bar{\gamma}\,\mathrm{d}\Omega - \bar{f}_c\,. \tag{3.23}$$

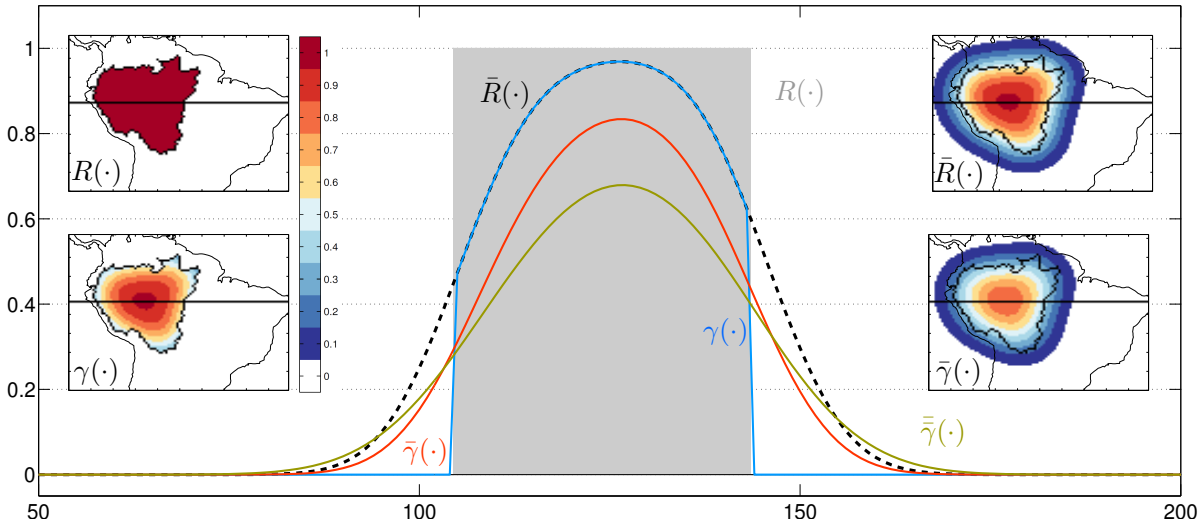


**Figure 3.4:** (a): 1-D visualisation of $R(\cdot)$, $\bar{R}(\cdot)$, $\gamma(\cdot)$, $\bar{\gamma}(\cdot)$, and $\bar{\bar{\gamma}}(\cdot)$ for Amazon. The corresponding 2-D function is shown in subfigures. The 1 dimensional plots for various quantities are extracted along the black horizontal line dissecting the catchment.

The relations (3.22) and (3.23) are similar to (3.17) and (3.18). They also contain a sum of attenuated catchment-confined quantities (deviation integral, regional average) and outside-catchment quantities (leakage). According to the discussion in section 2.4, a summation of leakage time series and attenuated catchment-confined time series affects both the amplitude and the phase of the total water storage change time series from filtered fields. Following the same line of argument, we can say that both the amplitude and the phase of leakage time series $\bar{l}_c$ and the deviation integral time series $\overline{\delta F}_c$ will be different from the corresponding original time series. To this end, we must find out how significant these amplitude and phase changes are, and how we can revert them to approach the corresponding original time series.

## §3.4 Estimating the original leakage and the original deviation integral

Mathematically, we can not determine the original leakage and the original deviation integral using (3.17), (3.18), (3.22), and (3.23). In Figure 3.5, we have shown the true leakage field and the leakage fields from once filtered and twice filtered fields. In this figure, our region of interest is Orinoco (corresponding catchment characteristic function $R(\cdot)$). We can see that the strength of the leakage field from twice filtered field is lower than the strength of the leakage from once filtered field, which has a lower signal strength compared to the true leakage field. Moreover, the leakage time series suggests that a scalar ratio will be able to relate the true leakage with the leakage from once filtered fields. Similarly, a scalar ratio will be able to relate the leakage from once filtered fields with the leakage from twice filtered fields. The same is true for the deviation integral.

Therefore, in search of any possible pattern, we plot leakage and the true deviation integral from the WGHM fields, and from filtered model fields, cf. Figure 3.6 and 3.7. We

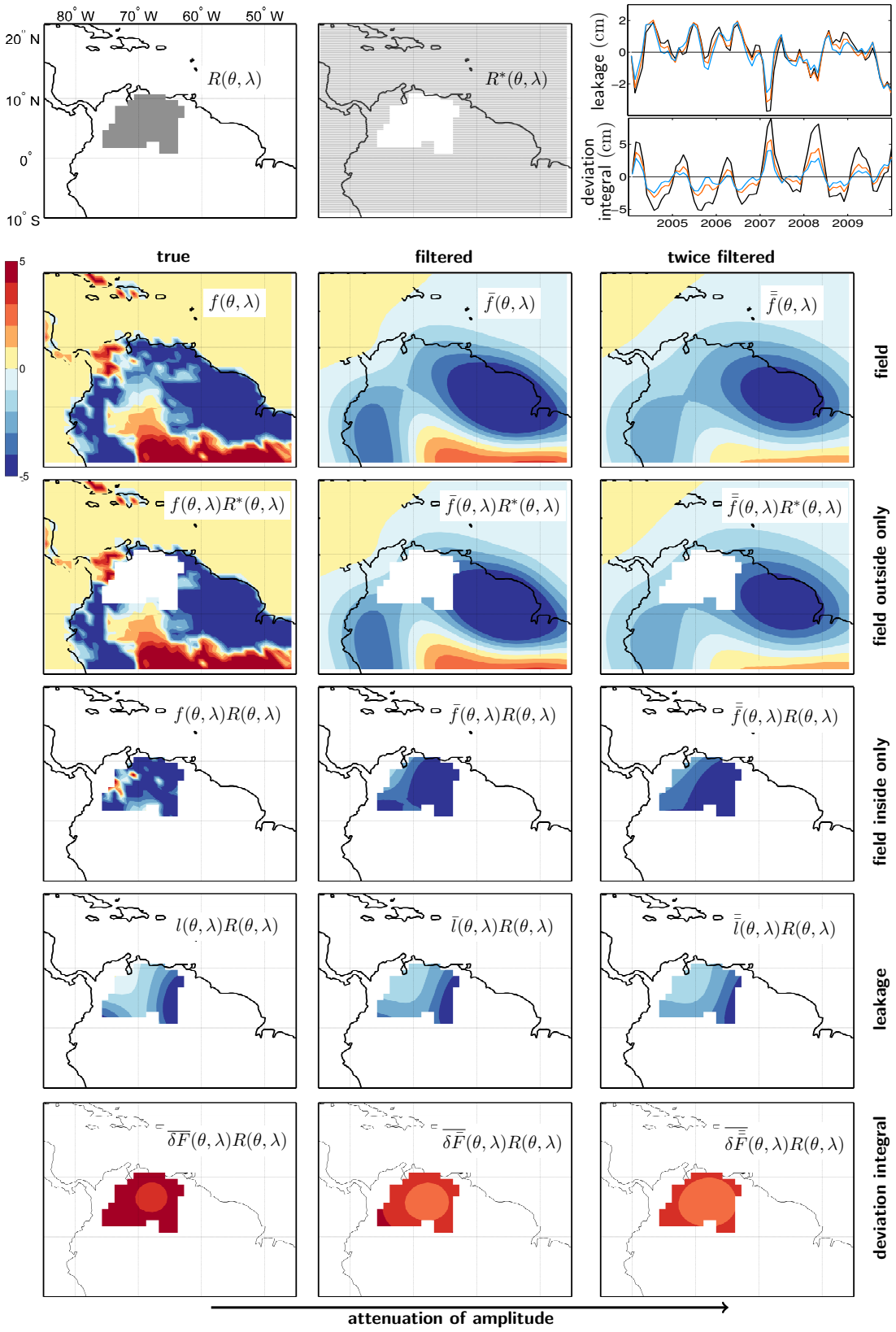


**Figure 3.5:** A visual description of true leakage, leakage from once filtered and twice filtered fields, true deviation integral, the deviation integral from once filtered and twice filtered fields. The unit of axes and colour-bars is cm.

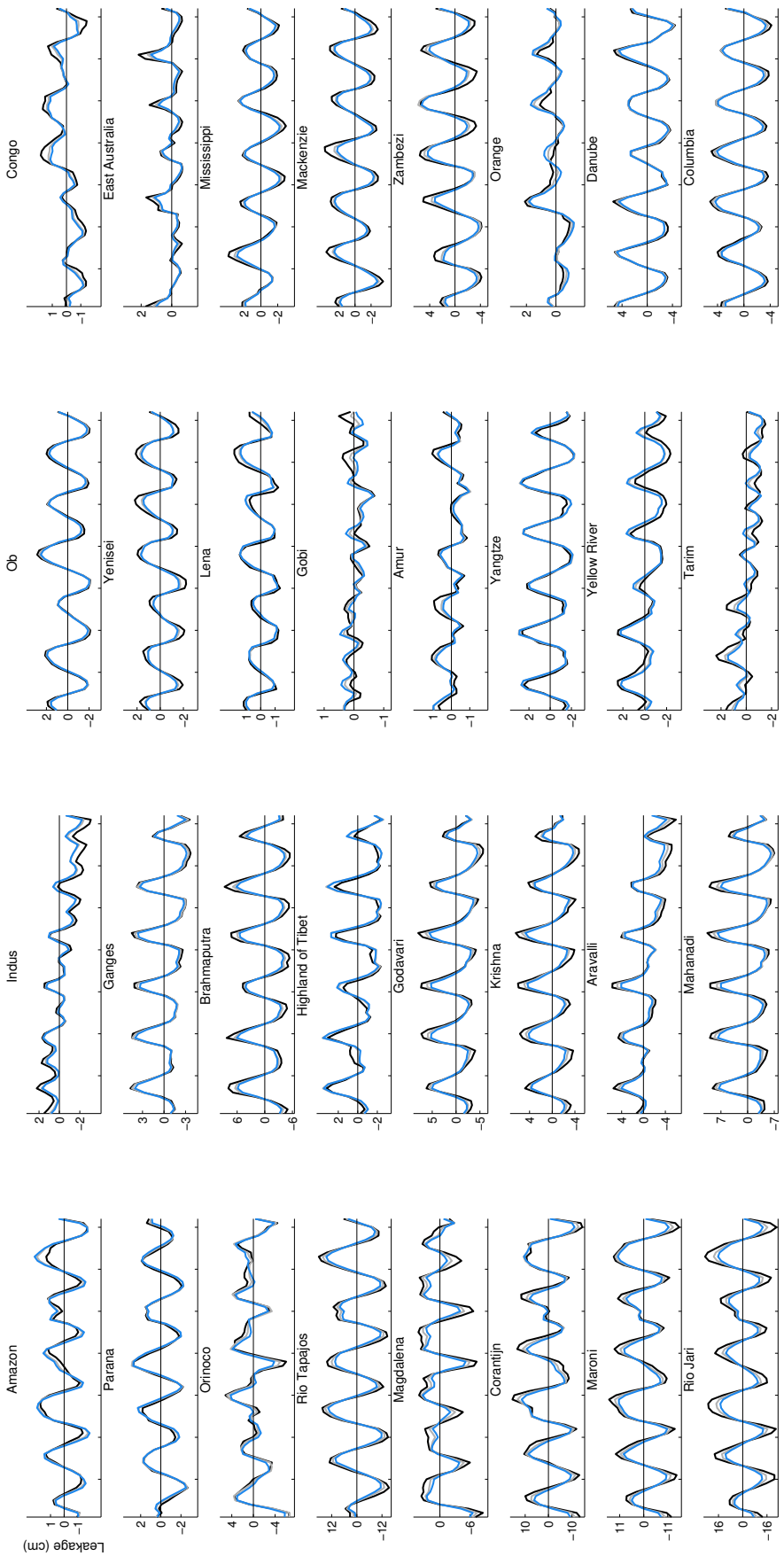


**Figure 3.6:** True leakage from WGHM model fields (black), leakage from once filtered fields (gray), and leakage from twice filtered fields (blue).

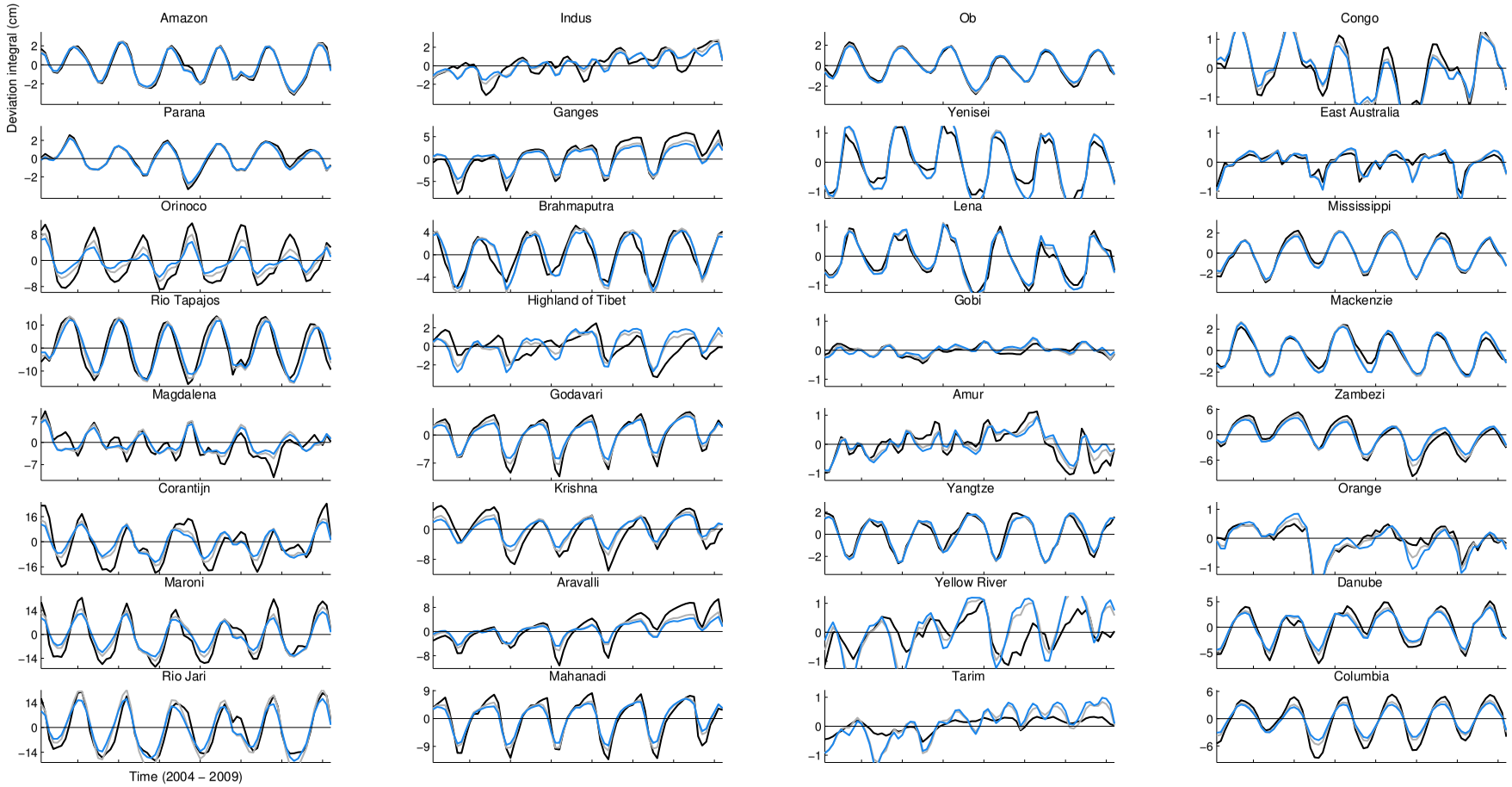


**Figure 3.7**: The deviation integral computed from WGHM model fields (black), from once filtered fields (gray), and from twice filtered fields (blue).

observe that for all the catchments, the leakage from filtered field has a lower amplitude with respect to the leakage from unfiltered field. We compute the ratio between original leakage and the leakage from filtered fields, given by

$$\frac{l_c}{\bar{l}_c} = \frac{\int_\Omega f(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega f(\theta,\lambda)\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega} = \frac{\int_\Omega f(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega \bar{f}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}. \tag{3.24}$$

Similarly, the ratio between leakage from once filtered field and leakage from twice filtered fields is

$$\frac{\bar{l}_c}{\bar{\bar{l}}_c} = \frac{\int_\Omega \bar{f}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega \bar{f}(\theta,\lambda)\,\overline{\kappa}(\theta,\lambda)\,\mathrm{d}\Omega} = \frac{\int_\Omega \bar{f}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega \bar{\bar{f}}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}. \tag{3.25}$$

Since the noise free GRACE monthly fields are unknown, the numerator in (3.24) can not be computed. However, both the numerator and the denominator in (3.25) can be computed because they require filtered fields. Now if the two ratios are related, then we can approach the true leakage $l_c$. Similarly, if the ratio between $\overline{\delta F}_c$ and $\overline{\delta \bar{F}}_c$ is related to the ratio between $\overline{\delta \bar{F}}_c$ and $\overline{\delta \bar{\bar{F}}}_c$, then we can approach the true deviation integral $\overline{\delta F}_c$.

In Table 3.1 we can observe that these ratios are approximately similar in magnitude. This is also demonstrated by the scatter plots in Figure 3.8 and 3.9. However, if we carefully examine individual catchments time series in Figure 3.6, then we see that for arid regions such as Highland of Tibet and Gobi the leakage from filtered fields deviates from original leakage in an inconsistent manner. But for other catchments, we observe a consistent amplitude dampening. Therefore, it is safe to assume that for catchments dominated by hydrological signal, these two ratios are approximately equal, and we can write

$$\frac{\int_\Omega f(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega \bar{f}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega} \approx \frac{\int_\Omega \bar{f}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega}{\int_\Omega \bar{\bar{f}}(\theta,\lambda)\,\kappa(\theta,\lambda)\,\mathrm{d}\Omega} \;\Rightarrow\; \frac{l_c}{\bar{l}_c} \approx \frac{\bar{l}_c}{\bar{\bar{l}}_c}. \tag{3.26}$$

$$\frac{\frac{1}{A_c}\int_\Omega \delta F(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega}{\frac{1}{A_c}\int_\Omega \delta \bar{F}(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega} \approx \frac{\frac{1}{A_c}\int_\Omega \delta \bar{F}(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega}{\frac{1}{A_c}\int_\Omega \delta \bar{\bar{F}}(\theta,\lambda)\,\bar{R}(\theta,\lambda)\,\mathrm{d}\Omega} \;\Rightarrow\; \frac{\overline{\delta F}_c}{\overline{\delta \bar{F}}_c} \approx \frac{\overline{\delta \bar{F}}_c}{\overline{\delta \bar{\bar{F}}}_c}. \tag{3.27}$$

The approximation in (3.26) and (3.27) is fair for a catchment surrounded by catchments with similar hydrological signal. A catchment with relatively weak hydrological activity, a

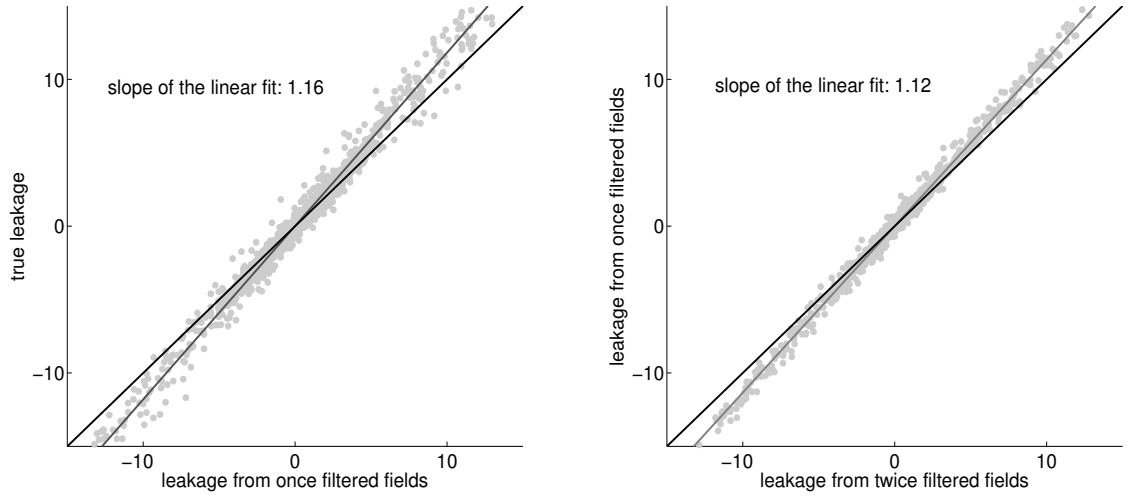


**Figure 3.8:** Scatter plot between the original leakage $l_c$ and that from once filtered fields $\bar{l}_c$ (left). Scatter plot between the leakage from once filtered $\bar{l}_c$ and twice filtered fields $\bar{\bar{l}}_c$ (right). The best fit line is in black and its slope is written in the respective plots. The unit of axes is cm.

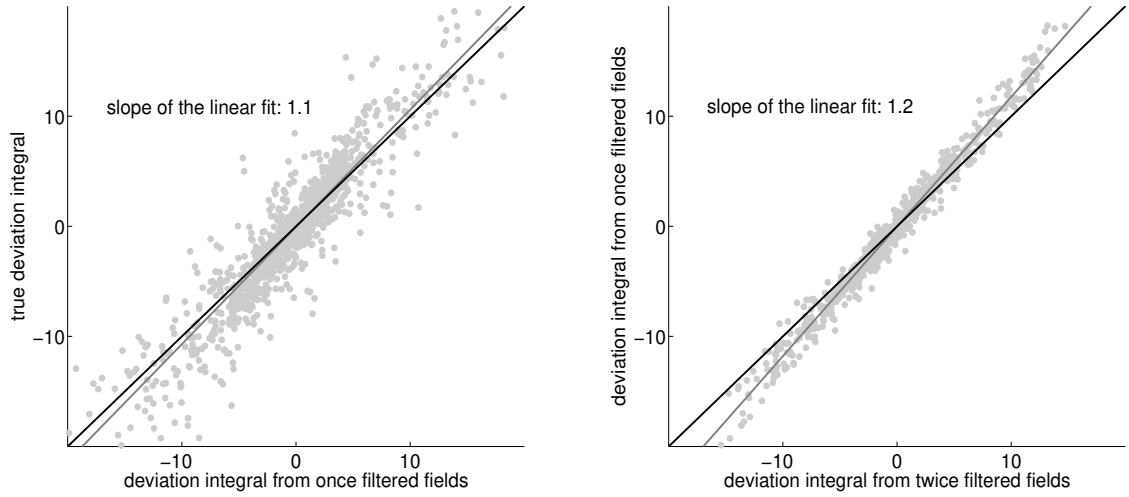


**Figure 3.9:** Scatter plot between the original deviation integral $\overline{\delta F}_c$ and that from once filtered fields $\overline{\delta \bar{F}}_c$ (left). Scatter plot between the deviation integral from once filtered $\overline{\delta \bar{F}}_c$ and twice filtered fields $\overline{\delta \bar{\bar{F}}}_c$ (right). The best fit line is in black and its slope is written in the respective plots. The unit of axes is cm.

**Table 3.1:** The phase difference, and amplitude ratio between two leakage time series, and between the deviation integral time series, for 32 catchments. These time series are generated from WGHM fields and their filtered version for catchments in the second column. A Gaussian filter with averaging radius of 500 km is employed for filtering. The ratios are determined by the method of least squares. The catchments are grouped into a set of 8 catchments separated by a vertical space. The spatial location of these catchments is shown in Figure 3.10 through the index number given in first column. Each group is sorted by area of the catchment. Catchments below the filter resolution are further separated by a horizontal line within the group.

| # | Catchment | Area ($km^2$) | $\Delta\phi$ between | | ratio between | | $\Delta\phi$ between | | ratio between | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | $l_c$ & $\bar{l}_c$ | $\bar{l}_c$ & $\bar{\bar{l}}_c$ | $l_c$ & $\bar{l}_c$ | $\bar{l}_c$ & $\bar{\bar{l}}_c$ | $\overline{\delta F}_c$ & $\overline{\delta \bar{F}}_c$ | $\overline{\delta \bar{F}}_c$ & $\overline{\delta \bar{\bar{F}}}_c$ | $\overline{\delta F}_c$ & $\overline{\delta \bar{F}}_c$ | $\overline{\delta \bar{F}}_c$ & $\overline{\delta \bar{\bar{F}}}_c$ |
| 1 | Indus | 1122836 | -2.0° | -0.7° | 1.22 | 1.07 | -2.3° | -1.5° | 0.91 | 1.08 |
| 2 | Ganges | 906200 | 1.3° | -0.1° | 1.04 | 1.12 | -4.4° | -1.2° | 1.17 | 1.16 |
| 3 | Brahmaputra | 521828 | 0.8° | -0.8° | 1.21 | 1.09 | -16.1° | -2.8° | 0.87 | 1.10 |
| 4 | High. of Tibet | 521596 | -6.0° | -1.3° | 0.90 | 0.99 | 34.7° | 4.1° | 0.81 | 0.77 |
| 5 | Godavari | 308695 | 3.5° | 1.5° | 1.11 | 1.10 | 3.2° | 3.2° | 1.19 | 1.12 |
| 6 | Krishna | 254875 | -1.9° | 2.3° | 1.22 | 1.09 | 16.6° | 2.7° | 1.13 | 1.18 |
| 7 | Aravalli | 179980 | -1.6° | -0.3° | 1.09 | 1.10 | 5.2° | 0.1° | 1.25 | 1.16 |
| 8 | Mahanadi | 123744 | 2.1° | 0.8° | 1.12 | 1.09 | 7.7° | 2.7° | 1.12 | 1.11 |
| 9 | Amazon | 4672876 | 2.6° | 2.6° | 0.85 | 0.96 | 5.1° | 1.8° | 1.01 | 1.06 |
| 10 | Parana | 2645738 | -6.1° | -3.1° | 0.94 | 1.01 | 2.3° | -1.3° | 1.02 | 1.00 |
| 11 | Orinoco | 836021 | 3.3° | 5.1° | 1.01 | 1.05 | 4.5° | 6.9° | 1.37 | 1.31 |
| 12 | Rio Tapajos | 366842 | -3.1° | -2.5° | 1.17 | 1.16 | -15.0° | -5.7° | 1.04 | 1.12 |
| 13 | Magdalena | 254778 | 1.3° | 1.9° | 1.24 | 1.20 | 6.2° | 1.6° | 0.94 | 1.22 |
| 14 | Corantijn | 67876 | 5.7° | 3.8° | 1.10 | 1.15 | 12.3° | 6.3° | 1.12 | 1.11 |
| 15 | Maroni | 61706 | 3.4° | 2.3° | 1.20 | 1.17 | 0.3° | 3.2° | 1.12 | 1.17 |
| 16 | Rio Jari | 46354 | 1.7° | 1.7° | 1.29 | 1.20 | 9.3° | 2.0° | 0.93 | 1.25 |
| 17 | Ob | 2926321 | -2.4° | -0.2° | 1.07 | 1.06 | 0.2° | -0.7° | 0.98 | 1.05 |
| 18 | Yenisei | 2454961 | 0.5° | 0.4° | 1.22 | 1.08 | -2.9° | -1.3° | 0.80 | 1.02 |
| 19 | Lena | 2417932 | -8.4° | -2.3° | 1.07 | 0.99 | 3.0° | -1.5° | 0.92 | 0.99 |
| 20 | Gobi | 2099470 | -4.1° | -0.3° | 0.92 | 0.87 | -1.9° | 1.6° | 0.94 | 0.85 |
| 21 | Amur | 1949471 | 4.0° | 0.1° | 1.06 | 1.07 | -11.2° | -7.3° | 1.07 | 0.99 |
| 22 | Yangtze | 1676801 | -3.1° | -1.7° | 0.91 | 1.01 | -4.5° | -2.6° | 0.93 | 0.98 |
| 23 | Yellow River | 902468 | -0.5° | -0.4° | 0.96 | 0.92 | 17.7° | 3.6° | 0.68 | 0.82 |
| 24 | Tarim | 884092 | -1.9° | -1.8° | 1.03 | 1.02 | -0.1° | -1.1° | 0.59 | 0.93 |
| 25 | Congo | 3615546 | 5.1° | 2.4° | 1.13 | 1.11 | -1.8° | 0.5° | 0.93 | 1.11 |
| 26 | Mississippi | 2938538 | 1.1° | 0.4° | 1.09 | 1.07 | -0.1° | -0.1° | 1.03 | 1.04 |
| 27 | East Australia | 1843833 | -3.9° | -2.0° | 1.17 | 1.09 | -5.7° | -3.1° | 0.80 | 1.00 |
| 28 | Mackenzie | 1666073 | 0.03° | -0.3° | 1.22 | 1.09 | 0.4° | -0.1° | 0.89 | 1.00 |
| 29 | Zambezi | 1115565 | -1.1° | 0.5° | 1.17 | 1.11 | 2.0° | 1.9° | 1.02 | 1.09 |
| 30 | Orange | 828475 | -3.7° | -0.2° | 0.88 | 0.89 | -4.8° | -3.0° | 0.94 | 0.89 |
| 31 | Danube | 771277 | -2.1° | -0.4° | 1.04 | 1.04 | 5.0° | 1.3° | 1.16 | 1.08 |
| 32 | Columbia | 662542 | -0.6° | 0.4° | 1.07 | 1.07 | 2.6° | 0.7° | 1.27 | 1.13 |

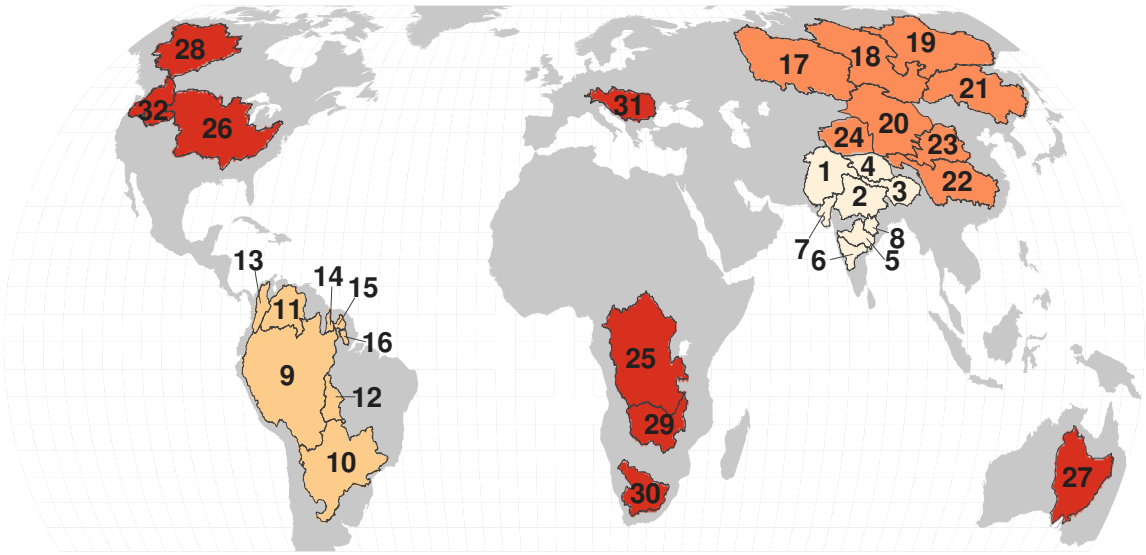


**Figure 3.10:** Spatial location of catchments investigated. The four different groups are represented by four different colours. An index number, allotted to every catchment, identifies the name and area of the catchment in table 3.1.

desert or an arid region, close to catchments with strong signal, will suffer from error with this approximation. This is due to the fact that filtering smothers the field, which affects the field distribution more, when the spatial signal contrast is stronger. For example, the annual amplitude of Highland of Tibet is below 8 cm but the annual amplitude of its neighbouring catchment Indus is approximately 25 cm. This is a case of high spatial signal contrast. Filtering it for the first time will have a larger impact on the spatial distribution than when filtering it for a second time. Hence, the deviation integral or leakage from once filtered fields will be related differently with the original value than with the value from twice filtered fields. The higher the contrast, the stronger this difference. However, one can argue, that if the ratios are not similar, then for dry catchments we must find a relation between the ratios in (3.26) and (3.27). This line of thought fails, because the leakage signal in an arid catchments, such as Highland of Tibet or Gobi, does not have a dominant annual behaviour. Over time there is no uniform relation (the fluctuation of leakage signal in time is not uniform, cf. Figure 3.11). In comparison to Godavari, the leakage time series $l_c$, $\bar{l}_c$, and $\bar{\bar{l}}_c$ for Highland of Tibet are not related by similar ratios over time.

A high spatial contrast also affects the neighbouring catchments of a dry region. The impact is small for leakage but significant for the deviation integral. Indus, Brahmaputra, Yellow river, and Tarim are in close proximity to either Highland of Tibet, or Gobi, or both. The deviation integral is directly related to the catchment-confined variations. These catchments, close to a dry catchment, are a case of high spatial signal contrast, which yields a bigger value of the deviation integral. Filtering for the first time affects the spatial signal contrast more than it does the second time. Similar is the impact on the deviation integral and, hence, on the ratios in (3.27). In Figure 3.7, we can see that for these catchments the deviation integral from models and its filtered version is not of the same nature and it also varies in time.

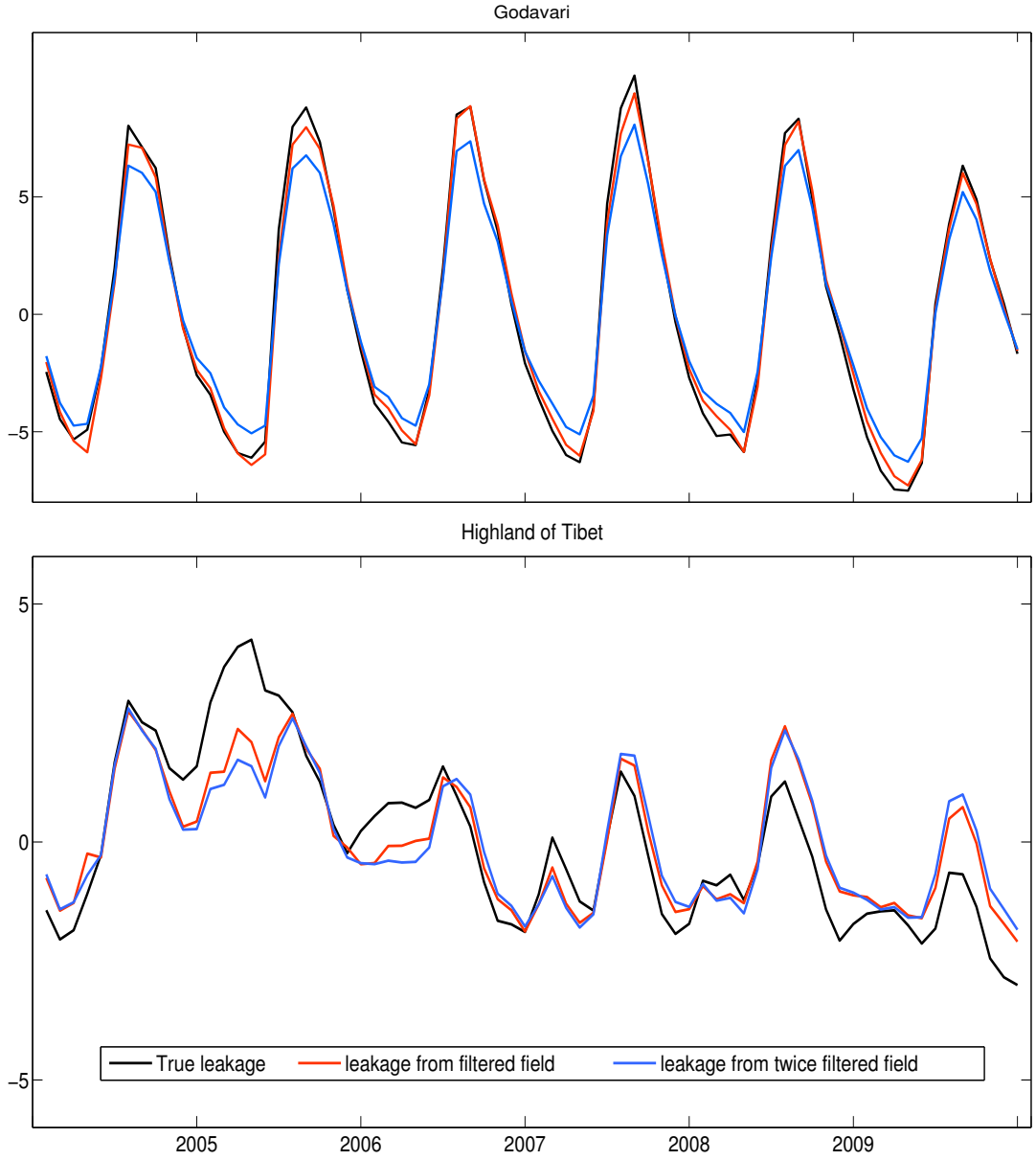


**Figure 3.11:** Leakage time series from subsequently filtered fields, and true leakage time series for Godavari and Highland of Tibet. The unit is cm.

The leakage and the deviation integral from filtered fields suffer from a phase change as well. In Table 3.1 we have also shown the phase difference between the leakage time series and the deviation integral time series computed from the model and their filtered versions. The phase difference between the leakage time series from once filtered fields and twice filtered fields is, more often than not, similar in sign and smaller than the phase difference between the original leakage and leakage from once filtered fields. The sign of phase change is different for Ganges, Brahmaputra, Krishna, and Mackenzie only. It is to be noted that the magnitude of phase change for these catchments is small. To this end, we propose that in order to correct for phase change, shift the leakage time series from once filtered fields by the phase difference between the original leakage and that from once filtered fields. This will ensure that we move towards a better estimate and do not overshoot. Similar analysis for the deviation integral shows that there is no general relation between the phase differences. Hence, we refrain from any phase corrections for the deviation integral from filtered fields.

In order to approach the true leakage time series, we propose to first calculate the leakage time series $\bar{l}_c$ and $\bar{\bar{l}}_c$, then the phase difference between them. Shift the $\bar{\bar{l}}_c$ by the phase difference towards $\bar{l}_c$, and after removing outliers calculate the mean of the ratio between shifted $\bar{\bar{l}}_c$ and $\bar{l}_c$ at each time point. Shifting of time series is performed in the frequency domain. We first Fourier transform the time series to be shifted, then multiply each frequency component by the complex exponential of the phase difference, and then transform back to the time domain. Once we have a phase difference and a mean ratio, we shift $\bar{l}_c$ by the phase difference and multiply it by the mean ratio to get $\hat{l}_c$, which should be very close to the leakage time series $l_c$. As per the relation (2.28), we remove $\hat{l}_c$ from the time series obtained from filtered field $\bar{f}_c$ and then scale up the result by $s$ to get closer to the original time series.

For the method of averages and deviations, we require an estimate of the deviation integral $\overline{\hat{\delta F}}_c$ also. It is computed by multiplying the deviation integral from filtered fields $\overline{\delta \bar{F}}_c$ with a multiplication factor computed by fitting a scale factor between $\overline{\delta \bar{F}}_c$ and $\overline{\delta \bar{\bar{F}}}_c$ using the method of least squares. The corrected time series is given by the following two equations:

$$\hat{f}_c = s(\bar{g}_c - \hat{l}_c), \quad \text{method of scale} \tag{3.28}$$

$$\hat{f}_c = \bar{g}_c - \overline{\hat{\delta F}}_c - \hat{l}_c, \quad \text{method of deviation} \tag{3.29}$$

Both methods are summarized with the help of flow charts in Figures 3.12 and 3.13. Although, GRACE has a significant contribution towards continental hydrology, it is popular for revealing the ice mass change in Antarctica and Greenland. In the previous section, we have clearly mentioned that the data-driven methods are drafted with the help of approximations that were found true for hydrological signal. If we want to apply them for ice mass change estimation, then first of all we must investigate whether these approximation are valid. If not, then which approximation suits our need.

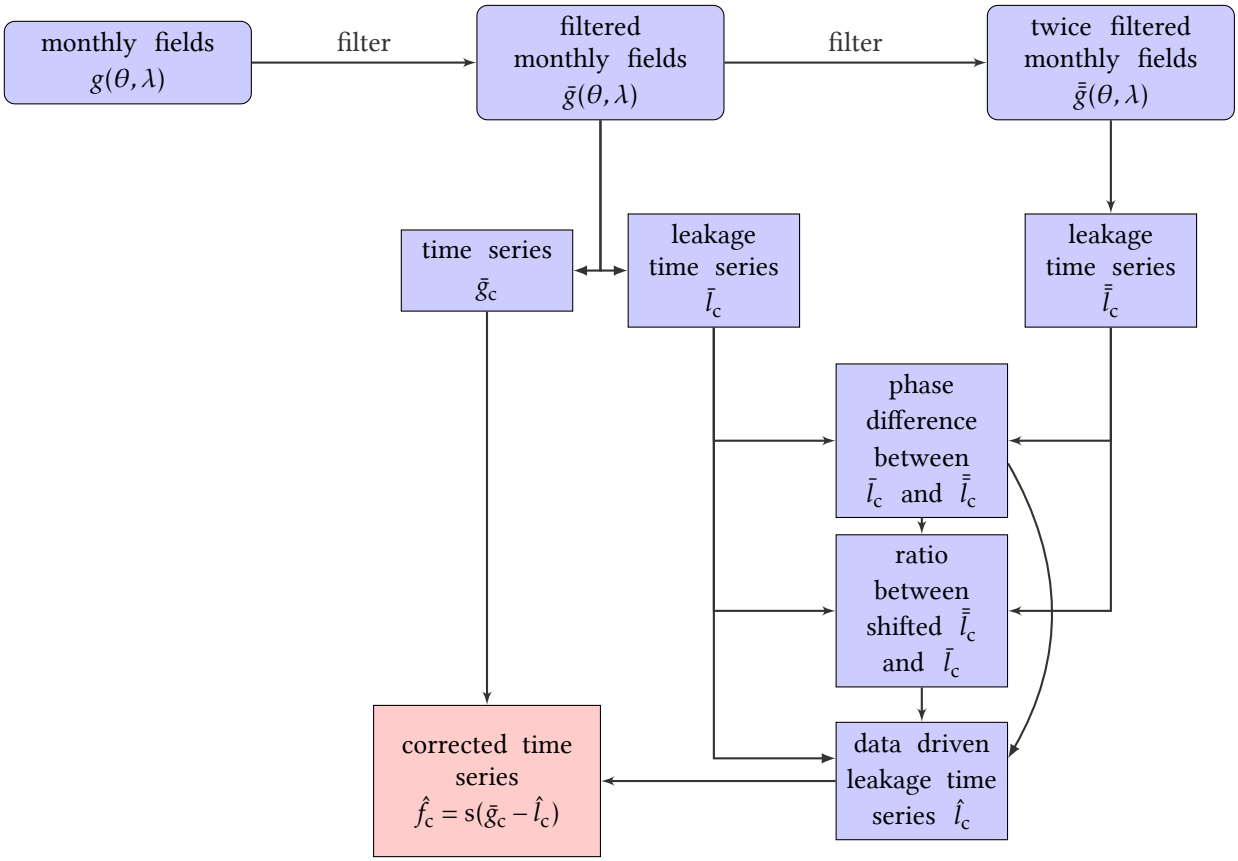


**Figure 3.12:** The flowdiagram for data-driven method of scale.

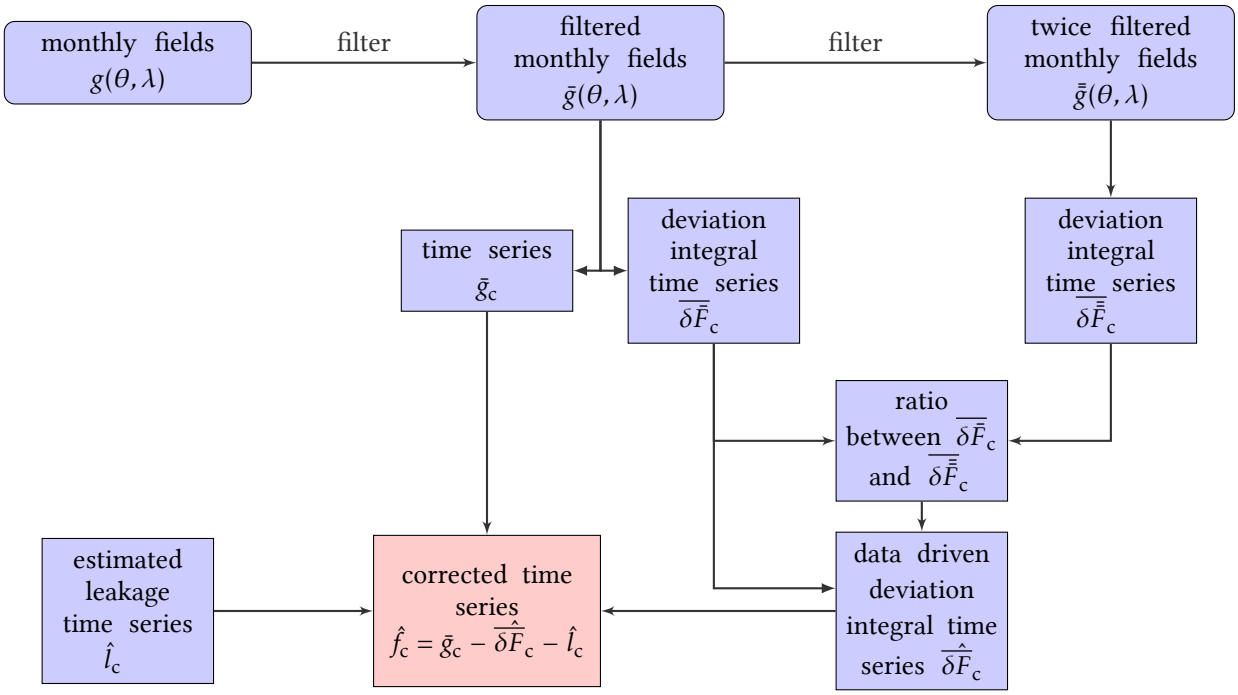


**Figure 3.13:** The flowdiagram for data-driven method of deviation.

## §3.5 THE SPECIAL CASE OF ICE SHEETS

The case of ice sheets is different in comparison to a hydrological catchment. In Antarctica or Greenland, most of the mass change is occurring near the coastal region, while in a hydrological catchment, the mass change is occurring throughout the catchment. Thus, the signal of ice mass change is much more concentrated than the signal of hydrological mass change over a catchment. These physical differences restrict us from estimating the leakage and the deviation integral from the filtered fields with the help of approximations made in (3.26) and (3.27).

In order to understand better, we consider the case of the Antarctic ice sheet. In Figure 3.14, we have shown the leakage field from a model and its filtered versions. The model comprises of GLDAS fields (hydrology on land) augmented with monthly ice mass change fields provided by Institute of Planetary Geodesy, TU Dresden. Since we have not included any ocean mass variations, we expect the leakage inside Antarctica to be zero. However, when we filter the signal in 3.14, it spreads outside the region, and leakage from this field is significant. When we filter the field once more, it spreads the signal further. Although leakage from twice filtered field is weaker than the leakage from once filtered field, it is still significantly larger than the true leakage (zero). In the data-driven methods (for hydrology), we use the leakage information from filtered fields to estimate the true leakage (cf. Figure 3.5), which is supported by the approximation given by (3.26). But, for ice sheets such as Antarctica and Greenland, such an approximation fails. Therefore, we formulate a different strategy for estimating the true signal for ice sheets.

In Figure 3.14, the leakage from the filtered field is far from the truth, because filtering spreads the signal outside the region of interest $R$, which otherwise has no or minimum signal. Filtering it once again attenuates and spreads the signal furthermore, but still the signal outside the region of interest is significantly larger than the truth. Similarly, the true deviation integral and the deviation integral from filtered fields. The deviation integral from filtered fields is much smaller than the true deviation integral because filtering spreads the energy outside the region. It is to be noted that this energy is captured as the leakage from filtered fields. Therefore it belongs to the true deviation integral, and it can be restored by adding its magnitude to the deviation integral from filtered fields. However, the leakage from filtered fields are computed by filtering an outside-only field, and this brings only a portion of the outside signal inside. Therefore, we must tackle the signal attenuation during this filtering step. We propose to achieve this by amplifying the leakage from once filtered fields with a ratio given by (3.25). Similarly, the deviation integral is also amplified to tackle the attenuation due to filtering. These ratios for amplification are computed in the same manner as they were computed for hydrological catchments. Once we obtain the amplified leakage, we must add its magnitude to the amplified deviation integral. Since the sign of leakage and the deviation integral are always opposite, changing the sign of leakage in (2.31) will fulfill our aim. Thus, the updated relation for ice sheets is written as

$$\hat{f}_{\mathrm{c}}^{\mathrm{is}} \;=\; \bar{f}_{\mathrm{c}} - \frac{1}{A_{\mathrm{c}}} \int_{\Omega} \delta F(\theta,\lambda)\, \bar{R}(\theta,\lambda)\, \mathrm{d}\Omega \;\; + l_{\mathrm{c}}\,. \tag{3.30}$$

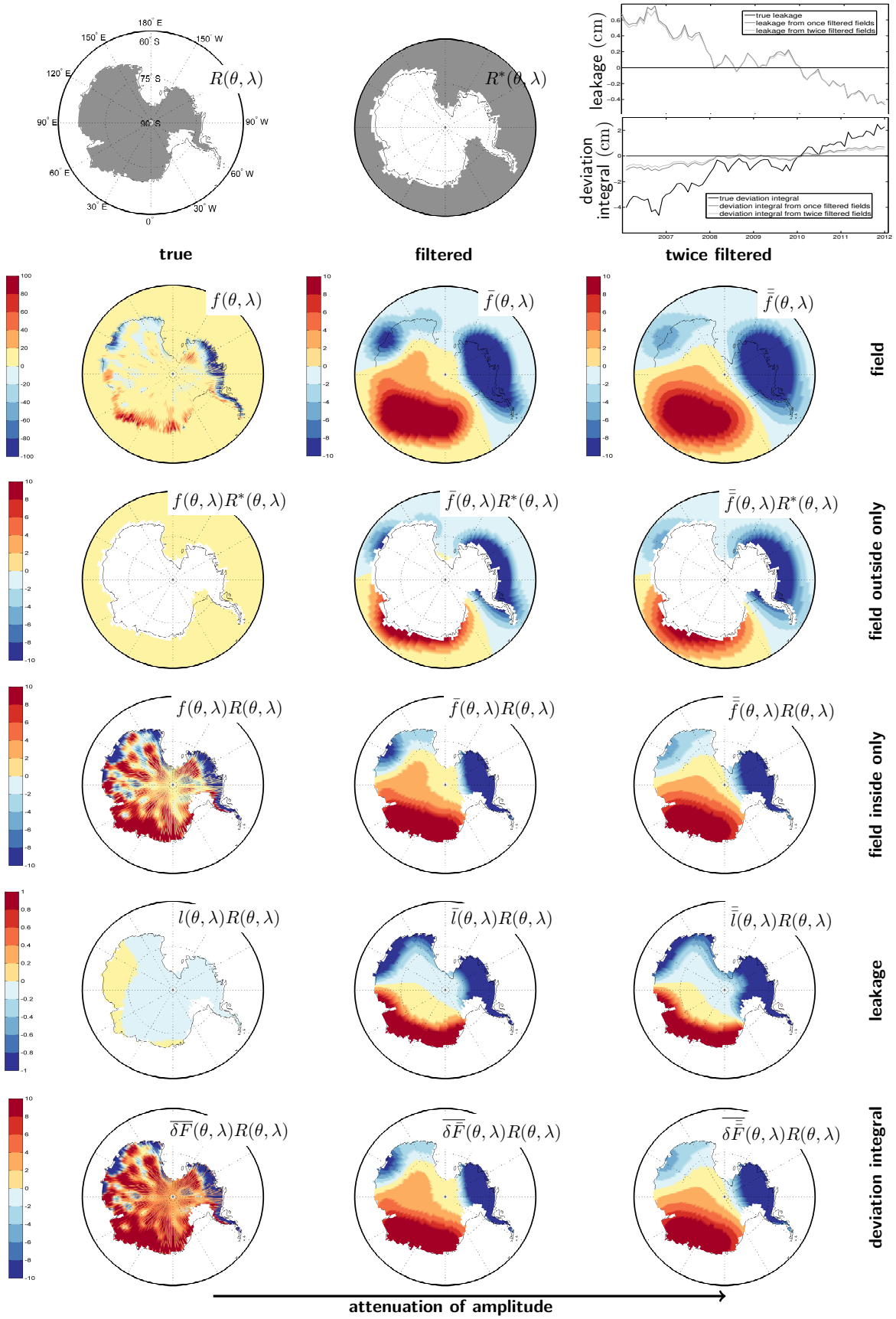


**Figure 3.14:** A visual description of true leakage, true deviation integral, leakage and deviation integral from once filtered and twice filtered fields. The unit of axes and colour-bars is cm. Please note that the scale of color bar is not same.

The data-driven methods, for hydrology and for ice sheets, are developed with the help of approximations and assumptions. The efficacy of these methods is yet to be tested.

# ARE WE ABLE TO APPROACH THE TRUTH? 4

The efficacy of the data-driven methods is investigated in this chapter. Since the GRACE observed mass change can not be validated directly, we perform a closed-loop test with realistic noisy fields. Within this closed-loop simulation environment, the results are compared with other popular model dependent approaches, which helps us to evaluate the improvement over them. Since the method is crafted with the help of a few approximations and assumptions, we validate them and study their impact to demarcate the limitations of the correction strategy proposed.

## §4.1 DATA, SIMULATION ENVIRONMENT, AND VALIDATION

We use monthly $1^\circ \times 1^\circ$ monthly GLDAS model fields [Rodell et al, 2004], from January 2004 to December 2009, as the truth. We contaminate them with noise, extracted from GRACE, to obtain realistic noisy fields. These noisy fields are then filtered and processed to obtain mass change over a region of interest, which are then corrected with three popular model dependent approaches and with both the data-driven approaches. In order to use model dependent approaches, we compute correction quantities (such as bias, leakage, and scale factors) from WGHM fields available from January 2004 to December 2009. The corrected time series are then compared with the true time series from GLDAS model fields over 32 catchments. These catchments are distributed over the globe, comprising large to very small catchments (cf. Figure 3.10 and Table 3.1). This makes the study comprehensive and brings out the impact of catchment size on the performance.

### Addition of realistic noise

In order to emulate GRACE noise in a closed-loop environment, we extract noise from Release 05, GRACE level 2 monthly products provided by GFZ [Dahle et al, 2012]. The erroneous $C_{20}$ coefficient in GRACE is replaced by $C_{20}$ from Satellite Laser Ranging mission [Cheng et al, 2013]. For extracting noise, we first filter the GRACE monthly fields with a destriping filter [Swenson and Wahr, 2006] and a Gaussian filter of half width radius of 400 km. Then these filtered fields are subtracted from their unfiltered version to get noise. The GLDAS monthly fields are available at $1^\circ$ grids, corresponding to a maximum degree of 180, while the GRACE fields are available till degree 90. We truncate the GLDAS fields at degree 90 to ensure spectral consistency. Then the monthly noise fields are added to the monthly GLDAS fields, to get noisy GLDAS fields. In Figure 4.1, we have shown

the degree variance and an arbitrary monthly field, before and after adding noise to it.

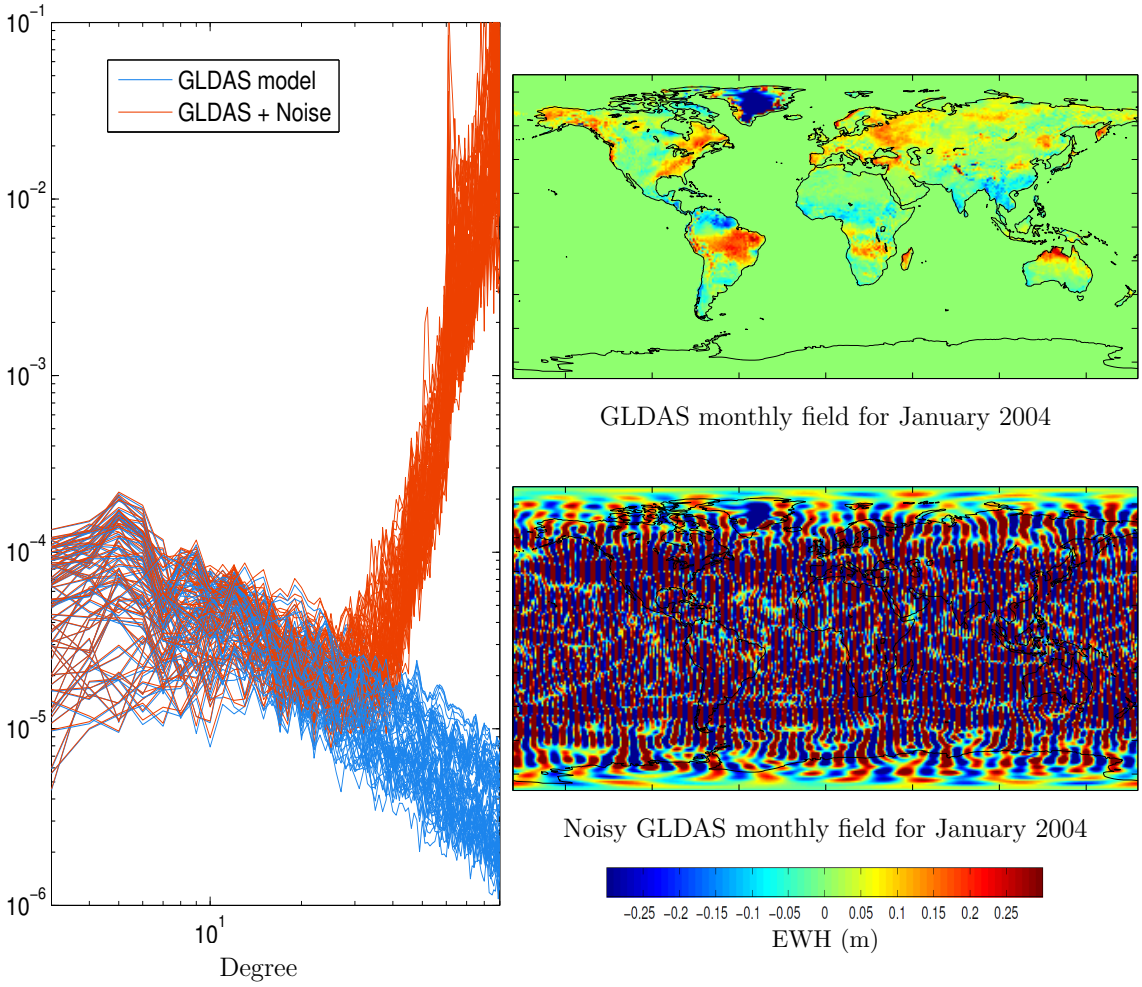


**Figure 4.1:** Degree variance in terms of EWH in meters, and spatial visualization of the GLDAS model and the noise added GLDAS model.

### Closed-loop validation

The efficiency of the data-driven approach depends on the ability of the filter to suppress noise. A good filter, which suppresses the noise to a minimum, will minimize the noise terms $n_c$ and $\overline{\delta N}_c$. However, it is also important that the filter does not suppresses the signal too much. In order to characterise whether a filter suppresses the noise enough, while retaining the maximum signal, we can use two performance metrics: processing gain and processing loss. Processing gain is the ratio between the signal-to-noise ratios of the unfiltered and the

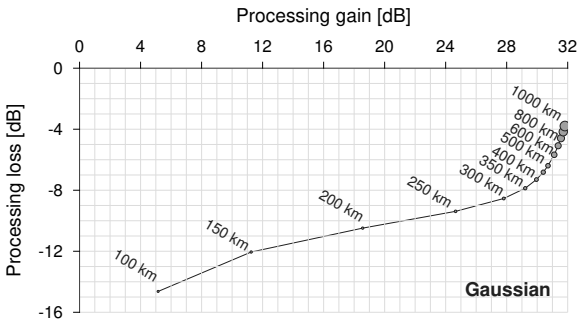


**Figure 4.2:** The processing gain and processing loss of Gaussian filters with different half width radius (*courtesy:* Balaji Devaraju).

filtered fields, while processing loss is the ratio between the difference in the signal energy before and after filtering and the signal energy before smoothing [Devaraju, 2015]. In Figure 4.2, the processing gain and processing loss of Gaussian filters with different half width radii are plotted. We can see that for a radius greater than 400 km, the processing loss increases rapidly with the radius, while the processing gain increases slowly. On the other hand, for a radius smaller than 400 km the processing gain increases rapidly with the radius, while the processing loss increases slowly. Therefore, we choose the optimum Gaussian filter of half width radius 400 km, to filter the noisy GLDAS fields [Devaraju, 2015]. Then we compute the regional average of the filtered field over a catchment for each epoch to get a time series $\bar{f}_c$. The filtered products are corrected by using three model dependent approaches: multiplicative [Longuevergne et al, 2010], additive [Klees et al, 2007], and scaling approach [Landerer and Swenson, 2012] (cf. section 2.6 and Table 2.1). In order to compute the corrected time series from these model dependent approaches, we use quantities, such as bias $b_c^m$, leakage $l_c^m$, and a scale factor $k$, computed from WGHM. We also use the data-driven method of scale [Vishwakarma et al, 2016] and the data-driven method of deviation to correct the filtered products.

In Figures 4.3 and 4.4, we plot the corrected time series from each approach and the true time series, the gap between them is filled with red colour. These fill plots are interpreted as, the less red means better performance. These figures are good for visual inspection, but a statistical assessment is vital for identifying the relative performance of these methods. Therefore, in Tables 4.1 and 4.2, RMS and NSE are provided to identify the best method. We can see that the corrected time series ($\hat{f}_c$) from the data-driven approaches perform consistently better than model dependent approaches. However, the performance of the two data-driven approaches is not similar for all the catchments. Let us discuss them one by one.

The data-driven method of scale given by equation (3.28) loses accuracy for catchments below the filter resolution [Vishwakarma et al, 2016]. This is due to the use of scale factor $s$ for amplifying the leakage removed filtered product, which also amplifies the errors in the estimated leakage. Since the scale factor $s$ increases as the size of the catchment decreases, the amplification of error in a small catchment is larger than in big catchments. For example, the scale factor for Aravalli, when using a Gaussian filter of half width radius of 400 km, is 5.01 (cf. Table 2.2). Thus, the errors in leakage are magnified 5 times. Therefore, the performance of the data-driven method of scaling decays with the size of catchment. In Table 4.1 and 4.2, we have also computed the mean of the RMS and NSE with all the 32 catchments and without the small catchments. The impact of the catchment size on the performance of the data-driven method of scaling is demonstrated by the change in the mean value after excluding the small catchments. Hence, if we assess these methods in the light of their promised efficacy, the data-driven methods are better than the model dependent methods. Apart from small catchments, arid catchments such as Highland of Tibet and Gobi also suffer from a less accurate leakage estimation.

On the other hand, the data-driven method of deviations performs consistently better than other methods. Since we are not using a scale factor $s$, there is no amplification of the errors in estimated leakage and the deviation integral. Therefore, we obtain relatively better time series for smaller catchments. However, the less accurate estimation of leakage

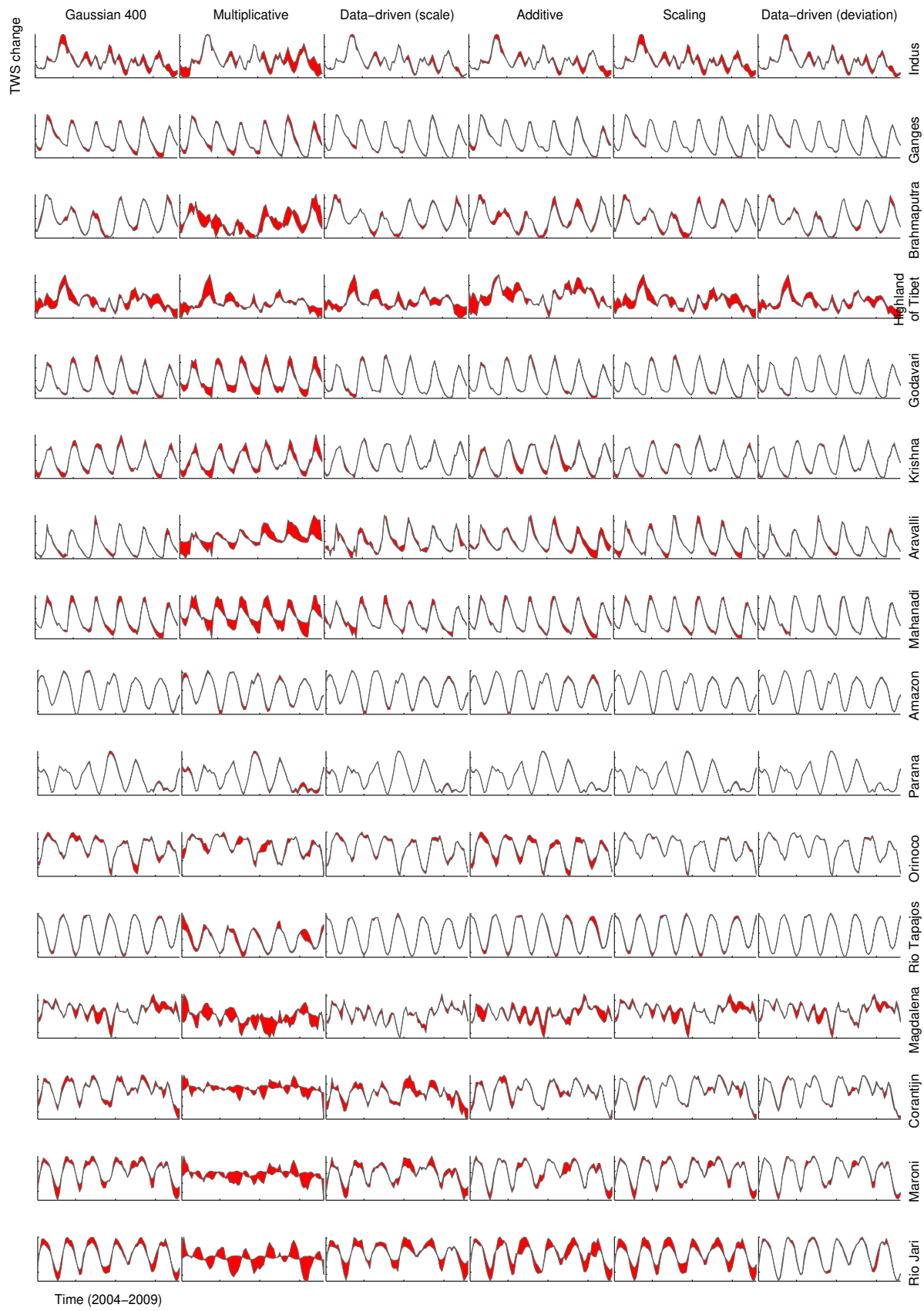


**Figure 4.3:** Fill plots comparing different approaches for catchment number 1 to 16 in Table 3.1. The ordinate of each subplot is different. Noisy GLDAS is the input field, and WGHM is used by the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

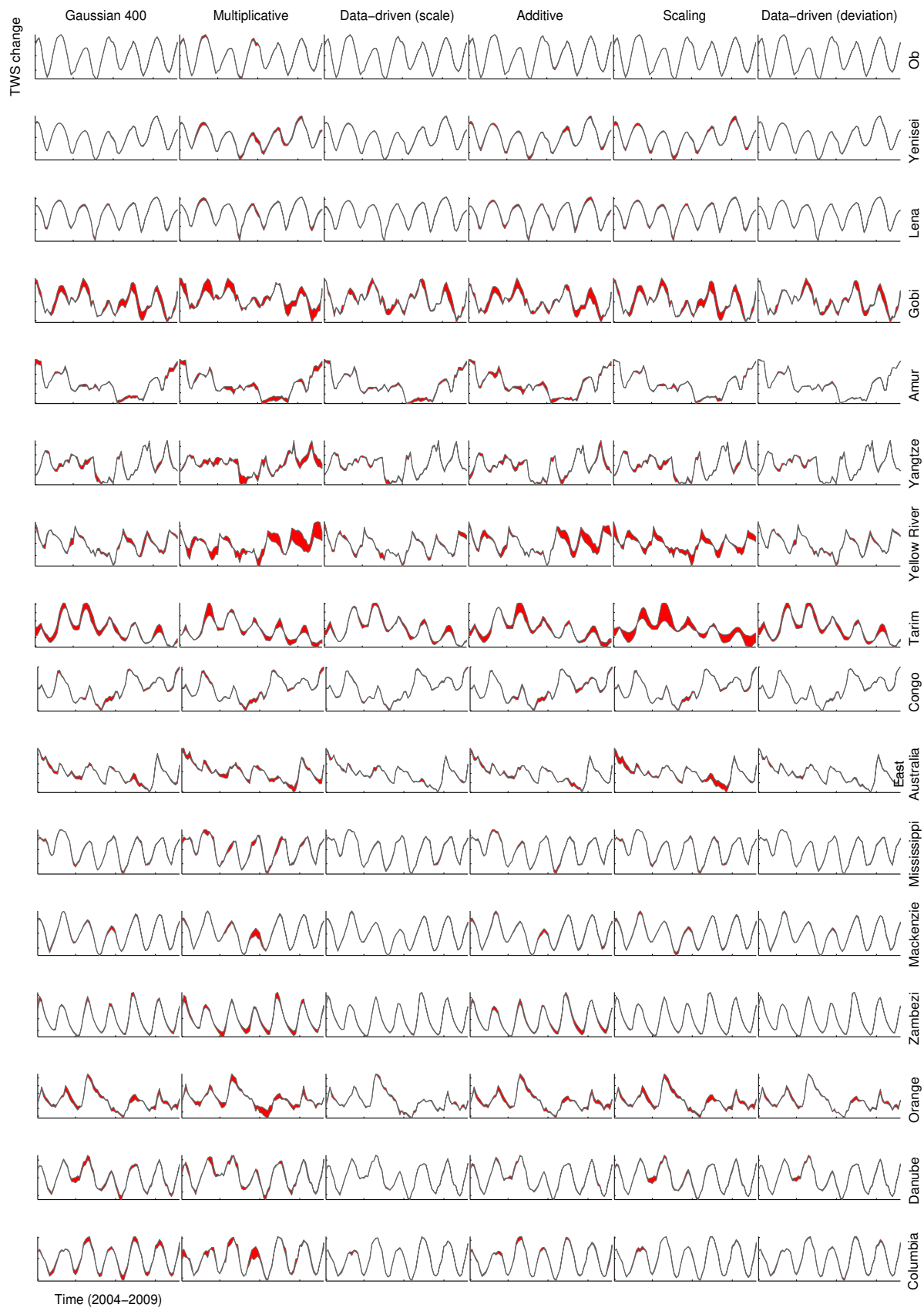


**Figure 4.4:** Fill plots comparing different approaches for catchment number 17 to 32 in Table 3.1. The ordinate of each subplot is different. Noisy GLDAS is the input field, and WGHM is used by the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

and the deviation integral for regions with high spatial signal contrast is reflected by a relatively poor performance for Highland of Tibet and Gobi. But still, it is at par with other methods. Tarim is a catchment sharing a long border with both Highland of Tibet and Gobi, while Indus, Brahmaputra, and Yangtze share borders with Highland of Tibet. As discussed in the section 3.4, the estimated deviation integral for these catchments is less accurate. Therefore, the performance in these catchments is not as good as it is for other hydrological catchments of similar size and signal strength. But overall it is superior to others.

In Tables 4.1 and 4.2, we have marked the best method and the second best method for a catchment. The data-driven method of deviation performs better than other methods for most of the catchments. In Table 4.5, we have listed the number of catchments, out of 32 catchments, for which each method is either the best or second best, and also when the method is best. We can clearly see that the data-driven methods are performing better than the model dependent approaches, and among the data-driven methods, the method of deviation is a better choice.

Although the results demonstrate the efficacy of the data-driven methods, it is important to check the validity of the assumptions and the approximations made while developing these methods. It increases the confidence in the approach and helps us to understand the reason and behaviour of the limitations.

**Validity of assumptions and approximations**

In reducing the relation (2.28) to (3.28), and (2.31) to (3.29), we assume that

*i.* the aggregated noise left after filtering, $n_c$ is approximately zero.

*ii.* the leakage term $l_c$ can be replaced by $\hat{l}_c$.

*iii.* the deviation integral term $\overline{\delta F}_c$ can be replaced by $\overline{\hat{\delta F}}_c$.

In the closed-loop simulation environment, we can monitor each component of the mathematical framework behind the data-driven approaches. This gives us the freedom to investigate the fidelity of these assumptions and approximations. We compute $n_c$, $\hat{l}_c$, $l_c$, $\overline{\delta F}_c$, and $\overline{\hat{\delta F}}_c$ with the noisy GLDAS fields.

Figures 4.5 show the box plots of the noise term $n_c$, the error in the estimated leakage ($\hat{l}_c - l_c$), and the error in the deviation integral ($\overline{\hat{\delta F}}_c - \overline{\delta F}_c$). These quantities are determined for 127 catchments, divided into three categories according to their area: white ($> 10^6\,\mathrm{km}^2$), light gray ($250\,000\,\mathrm{km}^2$-$10^6\,\mathrm{km}^2$), and dark gray ($< 250\,000\,\mathrm{km}^2$). As the area of the catchment decreases, the median shows higher fluctuation about zero, and the quartile generally increases. The whiskers and outliers are small for big catchments, but large for smaller catchments. Please note that in general the assumptions are valid because the median is close to zero for all the quantities and for all the catchments. However, the performance decays with the size of catchment due to increase in the quartiles, which is responsible for an inconsistent temporal agreement between the corrected time series and the true time series.

**Table 4.1:** RMS (in cm) of the error in the corrected time series computed in the noisy GLDAS simulation environment. Each row corresponds to a catchment and each column to a method. The smallest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments.

| | Gaussian 400 | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 1.65 | 2.26 | **0.82** | 1.13 | 1.68 | 1.15 |
| Ganges | 1.90 | 2.76 | 0.96 | 1.45 | 0.82 | **0.75** |
| Brahmaputra | 0.99 | 4.71 | 1.15 | 1.50 | 1.15 | **0.92** |
| High. of Tibet | 1.51 | 2.72 | 2.17 | 1.37 | 1.50 | **1.35** |
| Godavari | 1.97 | 7.65 | 1.48 | 1.47 | 1.24 | **0.79** |
| Krishna | 1.97 | 5.90 | 1.21 | 2.24 | 1.19 | **0.83** |
| Aravalli | 0.97 | 12.22 | 1.92 | 2.39 | 1.56 | **0.87** |
| Mahanadi | 3.41 | 23.18 | 5.19 | 2.81 | 2.47 | **1.90** |
| Amazon | 0.47 | 1.13 | 0.79 | 0.80 | 0.45 | **0.37** |
| Parana | 0.62 | 1.08 | 0.45 | 0.40 | 0.46 | **0.21** |
| Orinoco | 1.72 | 2.97 | 1.20 | 2.74 | 0.85 | **0.57** |
| Rio Tapajos | 1.93 | 8.79 | 1.49 | 2.84 | 2.08 | **1.04** |
| Magdalena | 2.40 | 10.31 | **1.68** | 3.26 | 2.43 | 1.93 |
| Corantijn | 3.65 | 39.28 | 11.44 | 3.46 | 2.50 | **2.20** |
| Maroni | 5.10 | 45.37 | 9.77 | 4.28 | 3.84 | **2.67** |
| Rio Jari | 6.05 | 126.18 | 19.26 | 9.26 | 8.42 | **3.24** |
| Ob | 0.22 | 0.42 | 0.17 | 0.28 | 0.22 | **0.11** |
| Yenisei | 0.19 | 0.85 | 0.23 | 0.55 | 0.60 | **0.16** |
| Lena | 0.33 | 0.60 | 0.24 | 0.55 | 0.42 | **0.14** |
| Gobi | 0.29 | 0.47 | 0.22 | 0.31 | 0.28 | **0.21** |
| Amur | 0.40 | 0.69 | 0.39 | 0.50 | 0.27 | **0.14** |
| Yangtze | 0.45 | 1.24 | 0.33 | 0.56 | 0.47 | **0.25** |
| Yellow River | 0.48 | 2.02 | 0.43 | 0.88 | 0.88 | **0.33** |
| Tarim | 1.13 | 1.42 | **0.71** | 1.04 | 1.88 | 0.92 |
| Congo | 0.73 | 0.90 | 0.39 | 0.77 | 0.61 | **0.31** |
| East Australia | 0.42 | 0.72 | 0.28 | 0.34 | 0.66 | **0.24** |
| Mississippi | 0.40 | 0.91 | 0.37 | 0.43 | 0.39 | **0.23** |
| Mackenzie | 0.45 | 0.76 | **0.26** | 0.51 | 0.50 | 0.31 |
| Zambezi | 1.35 | 3.44 | 0.63 | 2.22 | 0.78 | **0.46** |
| Orange | 0.76 | 1.36 | **0.40** | 0.70 | 0.79 | 0.47 |
| Danube | 1.31 | 1.82 | **0.40** | 0.73 | 1.07 | 0.67 |
| Columbia | 1.35 | 1.88 | **0.45** | 0.89 | 0.62 | 0.56 |
| **Mean** | 1.46 | 9.88 | 2.09 | 1.65 | 1.35 | **0.82** |
| **Mean*** | 1.01 | 2.58 | 0.71 | 1.13 | 0.90 | **0.57** |

**Table 4.2:** NSE values for the corrected time series computed in the noisy GLDAS simulation environment. Each row corresponds to a catchment and each column to a method. The largest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments.

| | Gaussian 400 | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 0.75 | 0.54 | **0.94** | 0.88 | 0.74 | 0.88 |
| Ganges | 0.96 | 0.91 | **0.99** | 0.98 | **0.99** | **0.99** |
| Brahmaputra | **0.96** | -0.00 | 0.94 | 0.90 | 0.94 | **0.96** |
| High. of Tibet | -0.18 | -2.64 | -1.41 | 0.05 | -0.17 | **0.06** |
| Godavari | 0.96 | 0.43 | 0.98 | 0.98 | **0.99** | **0.99** |
| Krishna | 0.94 | 0.43 | 0.98 | 0.92 | 0.98 | **0.99** |
| Aravalli | 0.95 | -4.22 | 0.82 | 0.73 | 0.88 | **0.96** |
| Mahanadi | 0.93 | -2.45 | 0.83 | 0.95 | 0.96 | **0.98** |
| Amazon | 0.99 | 0.97 | 0.98 | 0.98 | **1.00** | **1.00** |
| Parana | 0.99 | 0.96 | 0.99 | 0.99 | 0.99 | **1.00** |
| Orinoco | 0.92 | 0.77 | 0.96 | 0.80 | 0.98 | **0.99** |
| Rio Tapajos | 0.98 | 0.67 | 0.99 | 0.97 | 0.98 | **1.00** |
| Magdalena | 0.64 | -5.58 | **0.82** | 0.35 | 0.63 | 0.77 |
| Corantijn | 0.88 | -6.82 | -0.20 | 0.89 | 0.94 | **0.96** |
| Maroni | 0.84 | -7.34 | 0.43 | 0.89 | 0.91 | **0.96** |
| Rio Jari | 0.85 | -40.46 | -0.49 | 0.66 | 0.71 | **0.96** |
| Ob | **1.00** | 0.99 | **1.00** | **1.00** | **1.00** | **1.00** |
| Yenisei | **1.00** | 0.94 | **1.00** | 0.98 | 0.97 | **1.00** |
| Lena | 0.99 | 0.97 | 0.99 | 0.97 | 0.98 | **1.00** |
| Gobi | 0.70 | 0.22 | 0.83 | 0.66 | 0.73 | **0.84** |
| Amur | 0.97 | 0.92 | 0.97 | 0.96 | 0.99 | **1.00** |
| Yangtze | 0.96 | 0.65 | 0.98 | 0.93 | 0.95 | **0.99** |
| Yellow River | 0.91 | -0.45 | 0.92 | 0.70 | 0.69 | **0.96** |
| Tarim | 0.83 | 0.73 | **0.93** | 0.86 | 0.53 | 0.89 |
| Congo | 0.98 | 0.97 | 0.99 | 0.98 | 0.99 | **1.00** |
| East Australia | 0.95 | 0.87 | **0.98** | 0.97 | 0.89 | **0.98** |
| Mississippi | 0.99 | 0.95 | 0.99 | 0.99 | 0.99 | **1.00** |
| Mackenzie | 0.99 | 0.96 | **1.00** | 0.98 | 0.98 | 0.99 |
| Zambezi | 0.98 | 0.90 | **1.00** | 0.96 | 0.99 | **1.00** |
| Orange | 0.90 | 0.68 | **0.97** | 0.91 | 0.89 | 0.96 |
| Danube | 0.96 | 0.92 | **1.00** | 0.99 | 0.97 | 0.99 |
| Columbia | 0.95 | 0.91 | **0.99** | 0.98 | **0.99** | **0.99** |
| **Mean** | 0.88 | -1.61 | 0.78 | 0.86 | 0.87 | **0.93** |
| **Mean*** | 0.88 | 0.36 | 0.88 | 0.87 | 0.87 | **0.93** |

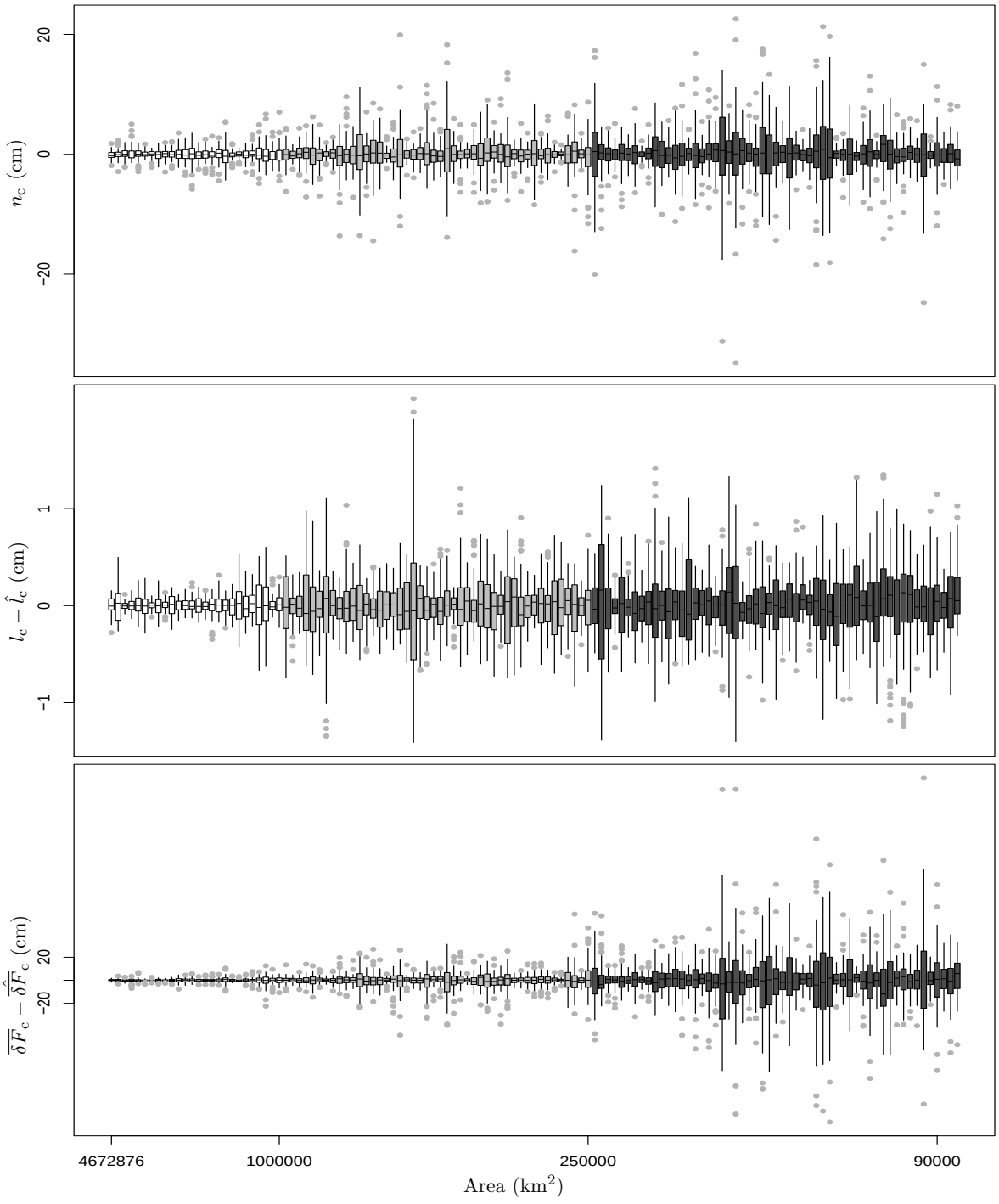


**Figure 4.5:** Box plot of the noise term $n_c$, the error in data-driven leakage, and the error in the deviation integral estimated by the data-driven approach for 127 catchments in the noisy GLDAS simulation environment. The abscissa is not linear.

The data-driven method of deviation is affected only by a sum of the error in leakage, the error in the deviation integral, and noise $n_c$, while the data-driven method of scaling is affected by an amplification of the sum of error in leakage and noise $n_c$ by a scale factor $s$. Given the behaviour of scale factor $s$, the data-driven method of scaling suffers more over a small catchment in comparison to the data-driven method of deviation. On the other hand, the error in the model dependent approaches can be attributed to the disagreement between the truth and the model amplified by a scale factor, if used. In order to compare the leakage from a model with the data-driven leakage from noisy fields, Figure 4.6 shows a scatter plot of error in leakage along with their respective histograms for 127 catchments. The scatter plot is between the error in estimated leakage $(l_c - \hat{l}_c)$, and the error in leakage from a model $(l_c - l_c^m)$. It is evident that the leakage estimated by the data-driven approach is superior to the leakage from WGHM model.

The overall impact of these assumptions and approximation can be assessed from Table 4.1, where the RMS of the total error for all the catchments is given. We can deduce that in this closed-loop simulation environment, the data-driven methods are able to approach the true time series with remarkable accuracy.

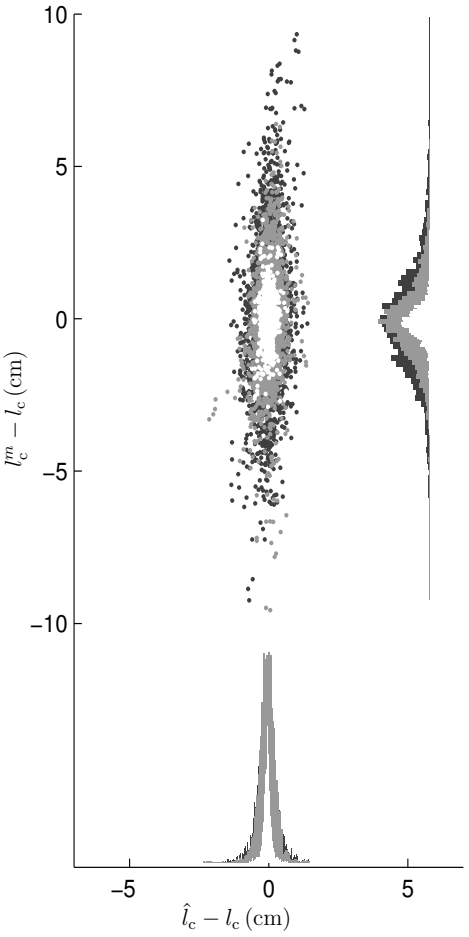


**Figure 4.6:** Histogram and scatter plot between the error in leakage from model and the error in leakage from data-driven approach.

## §4.2 TESTING IN A MORE REALISTIC CLOSED-LOOP SIMULATION

One of the novel findings of this work is the change in the phase of the catchment aggregated time series due to filtering, which is an important constituent of the data-driven approach. While estimating true leakage, we shift the leakage from filtered fields with a phase difference between leakage from twice filtered fields and from once filtered fields. The process of computing the phase change and shifting the leakage time series involves frequency domain operations (such as Fourier transform, Hilbert transform and multiplication with complex exponential). Therefore, a periodic time series with dominant annual signal is an undisputed candidate for better implementation and hence better results. However, the GRACE products come with their own problems.

First of all, GRACE level 02 monthly products suffers from uneven data acquisition (number of days of level 1 data available to generate the monthly solution) [Dahle et al, 2012]. Hence, the information content varies from one month to another. Secondly, there

are gaps in the time series (before 2004 and after 2010) and certain months are poorly observed (bad months) [Dahle et al, 2012]. Therefore, the time series exhibits a wiggly characteristic around the dominant signal corresponding to hydrology, which means that short wavelength signal is also present along with dominant annual and semi-annual signals. These problems with GRACE products are not covered in the noisy GLDAS closed-loop simulation environment. On the contrary, the simulation is rich in the signal content and there were no bad months. In order to be critical and conservative about the claims we made for the data-driven approaches, we seek a more robust examination. For that purpose, we use a GRACE type closed-loop simulation provided by GFZ [Flechtner et al, 2016]. They use the European Space Agency - Atmosphere Ocean Hydrology Ice Solid-Earth (ESA-AOHIS) models to represent the mass changes near the Earth's surface, which is used to simulate the GRACE satellite orbit. Then they introduce errors in orbits and in the range-rate observations extracted from this set-up, invert them to produce GRACE type noisy fields for a period of 5 years. A few bad months are introduced intentionally to emulate a realistic GRACE scenario. Now, we will verify our findings in this closed-loop simulation environment, hereinafter referred to as the GFZ GRACE simulation.

We follow the same procedure as followed for generating the corrected time series in noisy GLDAS simulation. We produce similar fill plots (Figure 4.7 and 4.8) and tables for RMS and NSE (Table 4.3 and 4.4) comparing the two data-driven methods and three model dependent approaches over the same 32 catchments. The WGHM fields are used to compute bias $b_c^m$, leakage $l_c^m$, and a scale factor $k$ for model dependent approaches. In this simulation also, we can observe that the data-driven method of deviation exhibit a consistently better performance in comparison to other approaches. It is to be noted that the performance of data-driven method of scale decays substantially in this more realistic closed-loop simulation compared to its performance in the noisy GLDAS closed-loop simulation. The limitations with arid regions and the size of the catchments are reflected here also. Furthermore, the scaling approach demonstrates a better performance, which may be attributed to good agreement between the background model in the simulation and WGHM. In terms of error RMS, the data-driven method of deviation is either the best or the second best in 24 catchments out of 32 catchments, and in terms of NSE it is either the best or the second best for 26 catchments out of 32, cf. Table 4.5. From the results in both the simulation environment, we can conclude that the data-driven method of deviation is able to approach the truth consistently, and it is superior to model dependent approaches.

Throughout the development of data-driven methods we have used a general two point function as our filter kernel $b(\theta, \lambda, \theta', \lambda')$. Therefore, the method itself is independent of the choice of filter, but the performance will vary from one filter to another. This is due to the fact that the residual noise $n_c$, and the accuracy of estimated leakage $\hat{l}_c$ and the deviation integral $\overline{\hat{\delta F}}_c$ is different for every filter. We have also demonstrated the data-driven methods with the destriping filter along with a Gaussian filter of 350 km radius in the GFZ GRACE closed-loop simulation environment. The fill plots and tables for RMS and NSE are given in Appendix B.

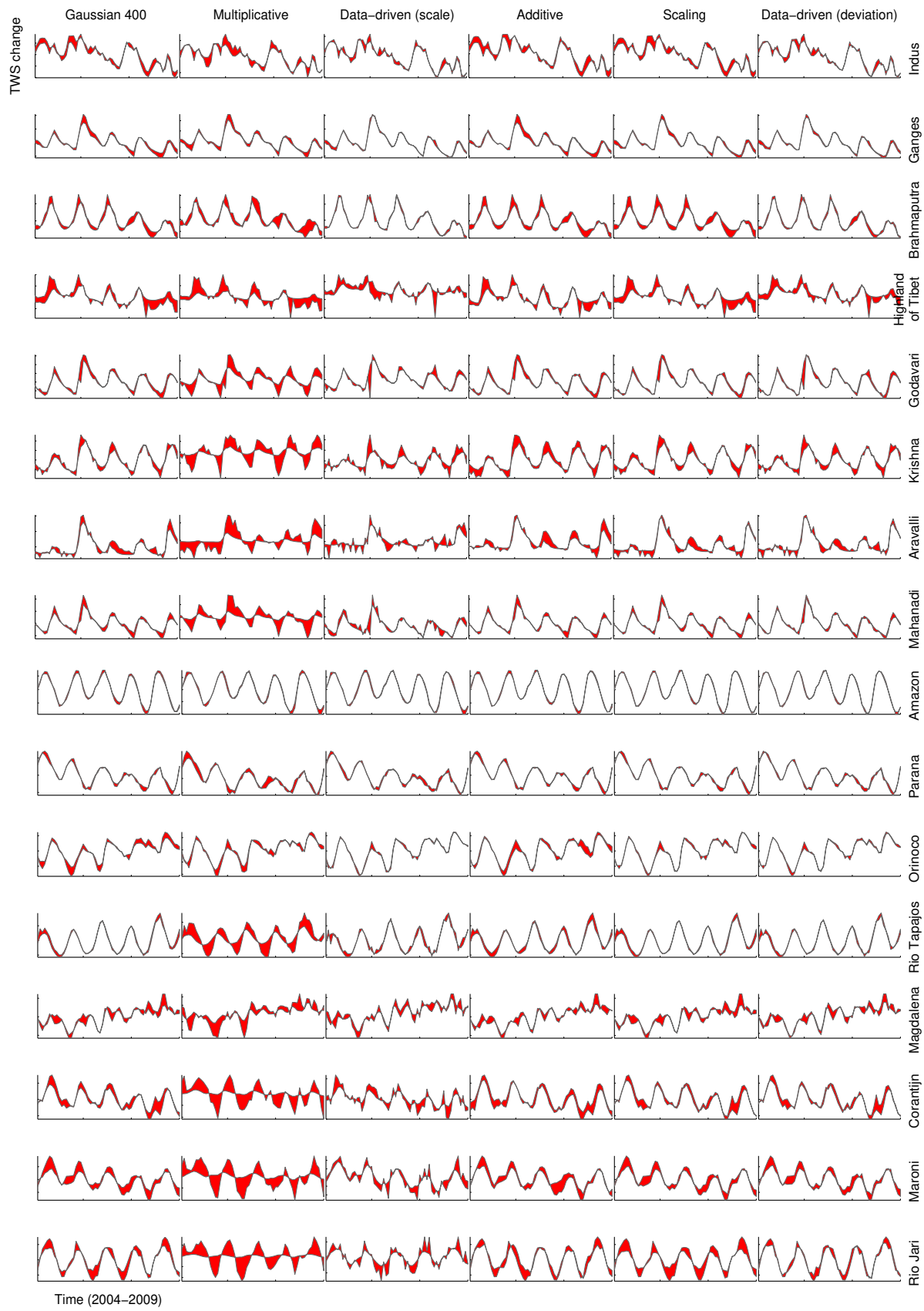


**Figure 4.7:** Fill plots comparing different approaches for catchment number 1 to 16 in Table 3.1. The ordinate of each subplot is different. The GFZ GRACE simulation fields are the input fields, and WGHM helps the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

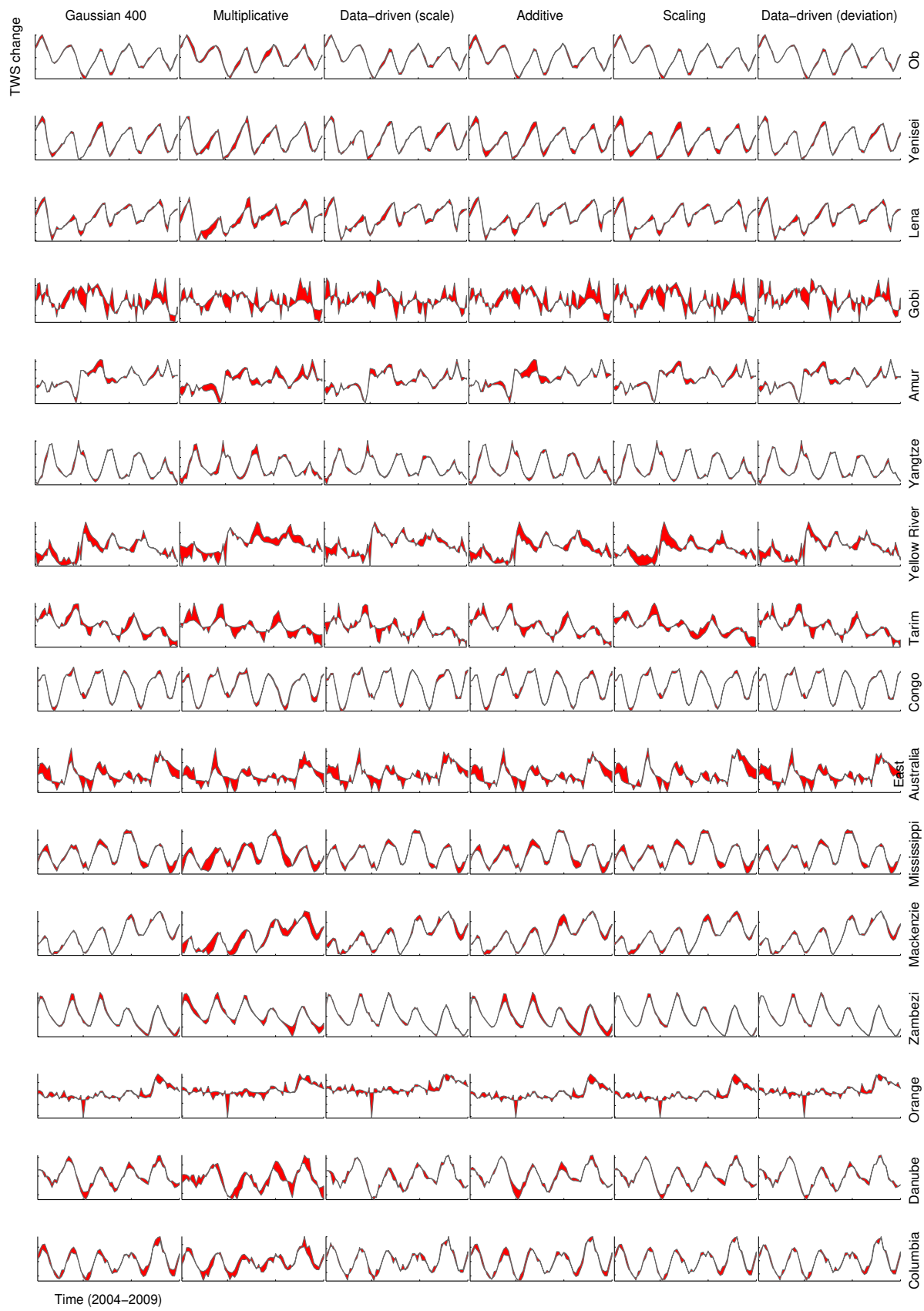


**Figure 4.8:** Fill plots comparing different approaches for catchment number 17 to 32 in Table 3.1. The ordinate of each subplot is different. The GFZ GRACE simulation fields are the input fields, and WGHM helps the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

**Table 4.3:** RMS (cm) of the error in the corrected time series computed in the GFZ GRACE simulation environment. Each row corresponds to a catchment and each column to a method. The smallest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments.

| | Gaussian 400 | Multiplicative | Scaling | Additive | Data-driven (scale) | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 1.93 | 3.14 | 1.94 | 2.00 | 1.81 | **1.60** |
| Ganges | 4.42 | 5.82 | 3.22 | 4.35 | **2.48** | 2.64 |
| Brahmaputra | 4.94 | 9.97 | 6.15 | 5.55 | **2.71** | 3.22 |
| High. of Tibet | 2.93 | 6.77 | 3.07 | **2.14** | 3.53 | 2.79 |
| Godavari | 4.55 | 18.67 | **4.06** | 4.27 | 8.13 | 4.92 |
| Krishna | **2.30** | 17.23 | 2.98 | 3.51 | 5.48 | 3.32 |
| Aravalli | **2.75** | 24.95 | 3.22 | 3.85 | 6.82 | 2.87 |
| Mahanadi | 5.93 | 61.51 | 5.08 | 5.24 | 10.31 | **4.73** |
| Amazon | 1.41 | 1.76 | **1.02** | 1.18 | 1.46 | 1.03 |
| Parana | 1.41 | 2.50 | 1.30 | 1.34 | 1.53 | **1.22** |
| Orinoco | 4.96 | 5.97 | 2.77 | 4.87 | **2.40** | 2.43 |
| Rio Tapajos | 3.66 | 27.32 | 3.65 | 4.88 | 4.92 | **3.37** |
| Magdalena | 5.18 | 17.73 | 5.20 | **4.70** | 6.47 | 5.00 |
| Corantijn | 6.77 | 89.12 | 5.94 | 6.85 | 15.88 | **5.78** |
| Maroni | 9.62 | 93.58 | 8.41 | 9.02 | 15.14 | **7.83** |
| Rio Jari | 7.54 | 201.06 | 9.46 | 7.38 | 21.75 | **5.77** |
| Ob | 0.74 | 1.44 | **0.71** | 0.93 | 0.81 | 0.78 |
| Yenisei | 0.81 | 1.28 | 1.19 | 1.10 | 0.89 | **0.75** |
| Lena | 0.68 | 1.35 | **0.70** | 0.78 | 0.79 | 0.72 |
| Gobi | 0.65 | 1.04 | **0.62** | 0.67 | 1.02 | 0.80 |
| Amur | **0.68** | 1.54 | 0.71 | 1.02 | 0.81 | **0.68** |
| Yangtze | **0.83** | 1.95 | 0.85 | 0.98 | 1.23 | 0.96 |
| Yellow River | **1.67** | 3.44 | 1.88 | 1.71 | 2.00 | 1.69 |
| Tarim | 1.30 | 3.46 | **1.06** | 1.22 | 1.77 | 1.48 |
| Congo | 1.16 | 1.79 | 1.02 | 1.26 | 1.30 | **0.96** |
| East Australia | 1.60 | 2.71 | **1.23** | 1.54 | 1.81 | 1.81 |
| Mississippi | 1.23 | 2.73 | 1.23 | 1.33 | **0.96** | 1.04 |
| Mackenzie | **0.65** | 2.60 | 0.89 | 1.10 | 1.18 | 0.84 |
| Zambezi | 2.60 | 6.12 | 1.83 | 4.06 | 1.76 | **1.62** |
| Orange | 1.83 | 3.82 | **1.80** | 1.82 | 2.99 | 2.38 |
| Danube | 1.74 | 4.90 | 1.36 | 1.90 | 1.68 | **1.34** |
| Columbia | 2.85 | 6.13 | **2.16** | 2.57 | 2.69 | 2.26 |
| **Mean** | 2.85 | 19.80 | 2.71 | 3.00 | 4.20 | **2.45** |
| **Mean*** | 2.17 | 6.04 | 2.02 | 2.33 | 2.39 | **1.91** |

**Table 4.4**: NSE values of the corrected time series computed in the GFZ GRACE simulation environment. Each row corresponds to a catchment and each column to a method. The largest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments.

| | Gaussian 400 | Multiplicative | Scaling | Additive | Data-driven (scale) | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 0.86 | 0.63 | 0.86 | 0.85 | 0.88 | **0.90** |
| Ganges | 0.89 | 0.82 | 0.94 | 0.90 | **0.97** | 0.96 |
| Brahmaputra | 0.86 | 0.41 | 0.78 | 0.82 | **0.96** | 0.94 |
| High. of Tibet | -0.72 | -8.13 | -0.89 | **0.09** | -1.49 | -0.55 |
| Godavari | 0.85 | -1.54 | **0.88** | 0.87 | 0.51 | 0.82 |
| Krishna | **0.75** | -12.32 | 0.59 | 0.43 | -0.40 | 0.49 |
| Aravalli | **0.83** | -13.03 | 0.76 | 0.66 | -0.05 | 0.81 |
| Mahanadi | 0.84 | -15.40 | 0.89 | 0.88 | 0.53 | **0.90** |
| Amazon | 0.98 | 0.97 | **0.99** | **0.99** | 0.98 | **0.99** |
| Parana | 0.95 | 0.84 | **0.96** | 0.95 | 0.94 | **0.96** |
| Orinoco | 0.88 | 0.83 | 0.96 | 0.88 | **0.97** | **0.97** |
| Rio Tapajos | 0.94 | -2.19 | 0.94 | 0.90 | 0.90 | **0.95** |
| Magdalena | 0.74 | -2.06 | 0.73 | **0.78** | 0.59 | 0.75 |
| Corantijn | 0.73 | -23.59 | 0.79 | 0.73 | -0.49 | **0.80** |
| Maroni | 0.67 | -24.39 | 0.75 | 0.71 | 0.18 | **0.78** |
| Rio Jari | 0.80 | -139.20 | 0.69 | 0.81 | -0.65 | **0.88** |
| Ob | 0.97 | 0.90 | **0.98** | 0.96 | 0.97 | 0.97 |
| Yenisei | **0.96** | 0.89 | 0.91 | 0.92 | 0.95 | **0.96** |
| Lena | **0.93** | 0.74 | **0.93** | 0.91 | 0.91 | **0.93** |
| Gobi | -0.37 | -2.44 | **-0.22** | -0.44 | -2.37 | -1.07 |
| Amur | **0.89** | 0.44 | 0.88 | 0.76 | 0.85 | **0.89** |
| Yangtze | **0.96** | 0.79 | **0.96** | 0.95 | 0.92 | 0.95 |
| Yellow River | **0.61** | -0.66 | 0.50 | 0.59 | 0.44 | 0.60 |
| Tarim | 0.60 | -1.87 | **0.73** | 0.65 | 0.25 | 0.48 |
| Congo | 0.97 | 0.94 | **0.98** | 0.97 | 0.97 | **0.98** |
| East Australia | 0.03 | -1.78 | **0.43** | 0.10 | -0.23 | -0.23 |
| Mississippi | 0.91 | 0.54 | 0.91 | 0.89 | **0.94** | 0.93 |
| Mackenzie | **0.98** | 0.63 | 0.96 | 0.93 | 0.92 | 0.96 |
| Zambezi | 0.96 | 0.76 | **0.98** | 0.89 | **0.98** | **0.98** |
| Orange | 0.58 | -0.85 | **0.59** | 0.58 | -0.14 | 0.28 |
| Danube | 0.91 | 0.26 | **0.94** | 0.89 | 0.91 | **0.94** |
| Columbia | 0.88 | 0.44 | **0.93** | 0.90 | 0.89 | 0.92 |
| **Mean** | 0.74 | -7.42 | **0.75** | 0.74 | 0.42 | 0.72 |
| **Mean*** | 0.73 | -0.81 | **0.74** | 0.73 | 0.51 | 0.70 |

**Table 4.5:** Comparison of different methods in both closed-loop simulation environments. $n_x^*$ denotes the number of catchment, out of 32, for which the method in respective column is either the best or second best in terms of the statistical measure $x$. $n_x$ denotes the same for the case when the method is best.

| | | Methods | | | | | |
|---|---|---|---|---|---|---|---|
| | | Gaussian 400 | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
| Noisy GLDAS simulation | $n^*_{RMS}$ | 4 | 0 | 17 | 3 | 9 | 31 |
| | $n_{RMS}$ | 0 | 0 | 7 | 0 | 0 | 25 |
| | $n^*_{NSE}$ | 10 | 1 | 23 | 10 | 16 | 32 |
| | $n_{NSE}$ | 3 | 0 | 12 | 1 | 5 | 26 |
| GFZ GRACE simulation | $n^*_{RMS}$ | 10 | 0 | 6 | 6 | 19 | 24 |
| | $n_{RMS}$ | 6 | 0 | 4 | 2 | 9 | 12 |
| | $n^*_{NSE}$ | 18 | 0 | 12 | 11 | 21 | 26 |
| | $n_{NSE}$ | 8 | 0 | 5 | 3 | 14 | 15 |

**Figure 4.9:** Comparison of various data-driven methods with the truth in a closed-loop simulation for Antarctica.

## §4.3 CLOSED-LOOP TEST OF DATA-DRIVEN METHOD FOR ICE SHEETS

In order to test the data-driven method for ice sheets, we seek a closed-loop environment simulating a GRACE-type product. We can not use the previous simulation setup because the GLDAS hydrological model is empty in Antarctica. Therefore, we use the ESA-CCI product as the background signal and contaminate it with the noise extracted from GRACE monthly products. The corrected time-series from data-driven methods, along with the true time series and the time series from filtered fields, are shown in Figure 4.9. We can see that the data-driven method of scale and the data-driven method of deviation are not able to approach the truth. Furthermore, the data-driven method of scale is poorer than the time series from filtered fields. It is underestimating the trend signal significantly, which proves that the uniform layer assumption is indeed inappropriate for the case of ice sheets. On the other hand, the data-driven approach for ice sheets is able to approach the truth with much better accuracy. Therefore, we expect better results with this method for ice mass loss computed from GRACE.

## §4.4 VALIDATION BY CLOSING THE WATER BALANCE EQUATION

Although we have shown that the data-driven methods, in a closed-loop simulation, are able to repair the damage due to filtering of GRACE, still a validation of corrected GRACE products is most desirable. Therefore, we follow the strategy proposed by Landerer et al [2010]; Lorenz et al [2014]; Sneeuw et al [2014], where they close the water balance equation for validating GRACE observed total water storage change over a number of catchments. A balance equation, where the sum of all the components in a system is zero, is helpful in determining a component, given all the other components are known. In a hydrological catchment, precipitation $p_c$ is equal to the sum of runoff $r_c$, actual evapotranspiration $e_c$, and the time derivative of the total water storage change ($\mathrm{d}f_c/\mathrm{d}t$). Therefore, we can write

$$p_c = r_c + e_c + \frac{\mathrm{d}f_c}{\mathrm{d}t} \quad \rightarrow \quad p_c - r_c - e_c - \frac{\mathrm{d}f_c}{\mathrm{d}t} = 0. \tag{4.1}$$

An accurate knowledge of precipitation, evapotranspiration, and runoff is essential to find the total water storage change, which is also recorded by GRACE. This gives us an opportunity to assess the accuracy of GRACE products (in an ideal case). In reality, we do not have a dense global *in-situ* measurements of these components of the hydrological balance equation. Therefore, we have to rely on poor global datasets provided by various agencies and institutes, such as Global Precipitation Climatology Centre (GPCC), Global Precipitation Climatology Project (GPCP), Climate Prediction Center (CPC), and many more [Lorenz et al, 2014]. These global datasets or models have large uncertainties that varies in space and in time. This limits the efficacy of hydrological balance equation as a tool to validate GRACE. Nevertheless, it is one of the few methods available for validating GRACE.

We use 6 different datasets for both precipitation and evapotranspiration. The grided precipitation data is obtained from CMAP, CRU, PRECL, GPCC, GPCP and Del, while the grided evapotranspiration is obtained from GLDAS, GLEAM, MOD16, FLUXNET-MTE, MERRA-Land, and

ERA-Interim (a description of these datasets is provided in Appendix A). It gives us 36 different combinations of precipitation and evapotranspiration, which along with runoff from GRDC and the time derivative of total water storage change from GRACE products provided by ITSG (either filtered or corrected), are used to close the water balance equation for 49 catchments. The balance equation is applied at catchment scale, hence we use (2.1) to compute regional average of precipitation, evapotranspiration, and total water storage change from GRACE. The runoff from GRDC is avaliable as one scalar value, for each epoch, for whole catchment. The GRACE products are numerically differentiated using three point central difference to obtain $\mathrm{d}f_c/\mathrm{d}t$. In order to maintain temporal consistency between $\mathrm{d}f_c/\mathrm{d}t$ and the time series of precipitation, evapotranspiration, and runoff, we smooth the later with a moving window suggested by Landerer et al [2010]

$$\tilde{X}_i = \frac{1}{4}X_{i-1} + \frac{1}{2}X_i + \frac{1}{4}X_{i+1}, \tag{4.2}$$

where $X_i$ is the quantity (precipitation, evapotranspiration, or runoff ) at epoch $i$. Lorenz et al [2014] demonstrated that in general, for most of the catchments and for most of the combinations, the GRACE derived runoff exhibits a good agreement with the GRDC observed runoff. However, the errors in models and datasets prevent a clear strong pattern to emerge in favour of GRACE or in favour of a combination of models.

In order to investigate whether the data-driven approach offers some improvement in the GRACE products, we seek a reduction in the misclosure of the water balance equation, when using the corrected GRACE products in comparison to when using the filtered GRACE products. These misclosures are written as:

$$\bar{\delta}_c = P_c - r_c - Eta_c - \frac{\mathrm{d}\bar{f}_c}{\mathrm{d}t} \quad \text{misclosure with filtered GRACE,} \tag{4.3}$$

$$\hat{\delta}_c = P_c - r_c - Eta_c - \frac{\mathrm{d}\hat{f}_c}{\mathrm{d}t} \quad \text{misclosure with corrected GRACE.} \tag{4.4}$$

We have shown the bias and RMS of the misclosure for each combination and for each catchment in Figures 4.10 and 4.11. We can clearly see that there is no clear pattern, which favors a combination of model or which suggests a substantial improvement after correction. There are a few catchments and a few combinations which display either improvement or degradation, but overall we can not reach a conclusion.

This exercise seems to have been a futile attempt to assess any improvement in closing the water budget with corrected GRACE products. Nevertheless, it has helped us understand that in closing the water budget, it is the uncertainty in models, which is the thorn in the flesh, not the uncertainty in GRACE. This is reflected by Figures 4.10 and 4.11, where we have also shown the bias and RMS of the correction in GRACE ($\mathrm{d}\hat{f}_c/\mathrm{d}t - \mathrm{d}\bar{f}_c/\mathrm{d}t$). It is to be noted that for most of the catchments, the correction is nearly an order of magnitude smaller than the misclosure that comes from the poor quality of the combined precipitation and evapotranspiration dataset.

We have established and proved the efficacy of the data-driven methods, in a closed-loop

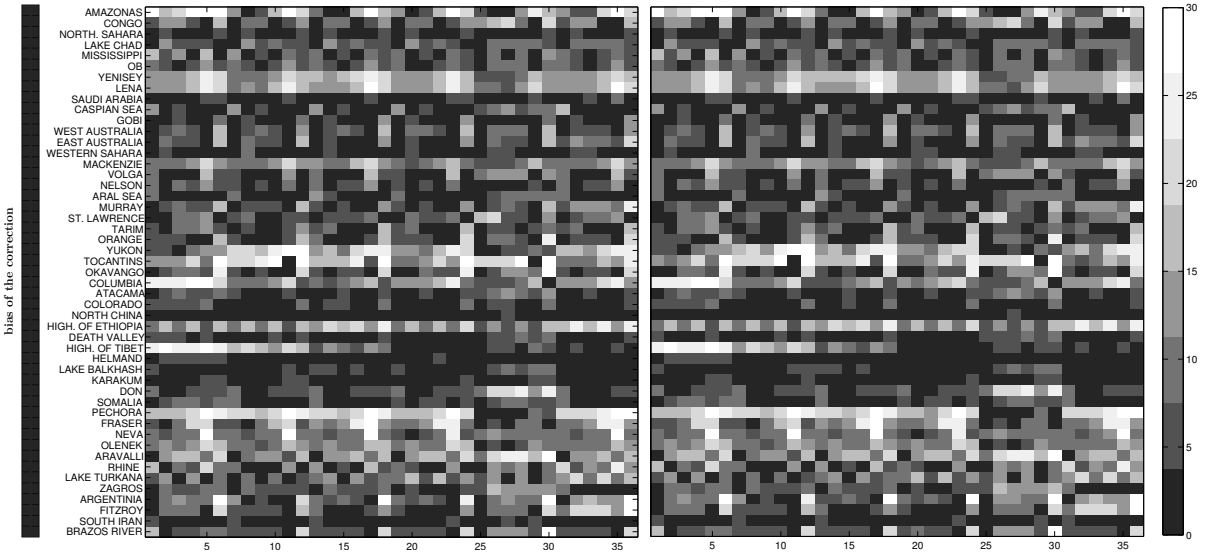


**Figure 4.10:** The bias of the misclosure with respect to zero. The left figure is the bias when we use the time series from filtered fields, and the right figure is when we use the time series corrected with data-driven method of deviation. The ordinate consists of catchments sorted by their area, and the abscissa denotes the combination of models. The bias of the correction achieved using data-driven method, for each catchment, is given by the gray bar on extreme left.

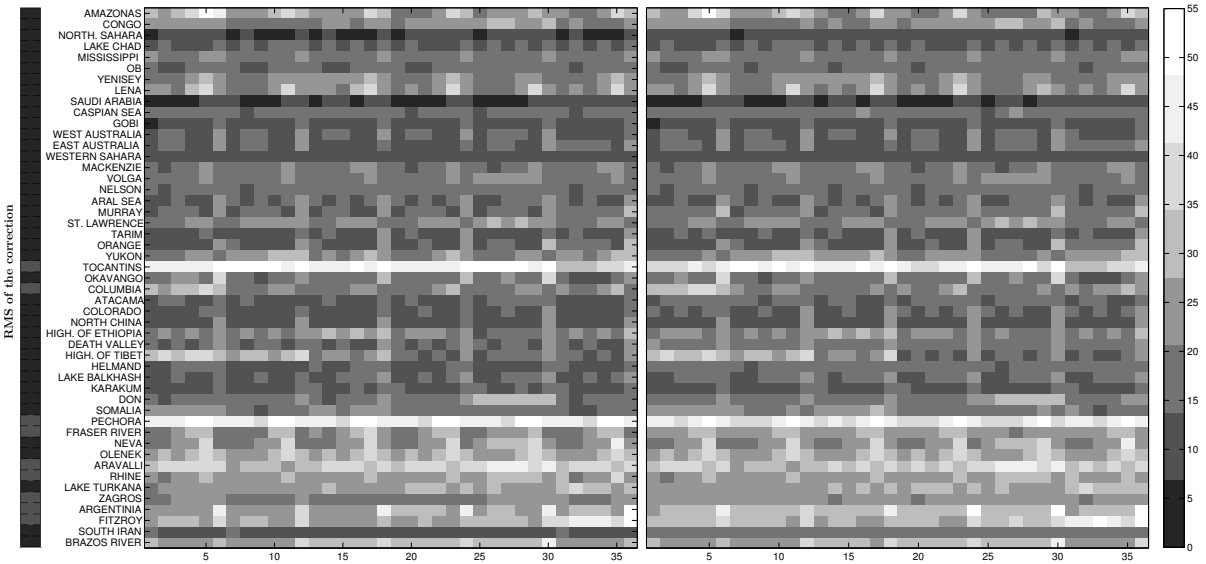


**Figure 4.11:** The RMS of the misclosure. The left figure is the RMS when we use the time series from filtered fields, and the right figure is when we use the time series corrected with data-driven method of deviation. The ordinate consists of catchments sorted by their area, and the abscissa denotes the combination of models. The RMS of the correction achieved using data-driven method, for each catchment, is given by the gray bar on extreme left.

simulation, for minimizing the signal damage due to filtering on the sphere. The interest in this work was enunciated with an aim to improve the mass change estimates obtained from GRACE. In the next chapter we will process the GRACE products and study the impact of signal improvement from data-driven approaches.

# CORRECTING THE FILTERED GRACE PRODUCTS

# 5

The changes in our planet can be monitored through various spaceborne sensors, but most of them are capable of recording the changes only on or above the Earth's surface. Therefore, the unique ability of GRACE, to record also the mass changes beneath the surface, added a new dimension to the satellite based observation of the Earth system. The total water storage change from GRACE has an everlasting impact on our understanding of the hydrosphere and cryospheric sciences, both at global and local scale. This is reflected by the usage of GRACE in determining:

- the amount of water lost by California in recent drought;
- the ground water depletion in North-West India;
- the rate of water loss in endorheic catchments such as Aral and Urmia;
- the hydrological extreme events, such as floods and droughts;
- the magnitude of ice mass loss in Greenland and Antarctica;
- and many more.

In this chapter, we choose a few regions that have shown distinguished signatures of water mass changes due to human intervention and climate change. We assess the change in signal magnitude offered by the data-driven method of deviation in comparison to the filtered GRACE time series. We start with analysis of Aral sea, Victoria lake, Urmia lake, and California. Then we report the time series for 32 hydrological catchments located in different continents and climatic zones. After discussing hydrology, we move to the ice mass loss in Antarctica and Greenland, where we apply the data-driven method for ice sheets to assess the total ice mass loss from 2003 to 2015.

We use the GRACE level 02 products provided by ITSG, Graz [Mayer-Gürr et al, 2016]. The $C_{20}$ coefficient is replaced by the $C_{20}$ coefficient from SLR [Cheng et al, 2013], and the degree 1 coeffiecients are provided by Swenson et al [2007]. We use a Gaussian filter of half width radius 400 km to filter the noisy products and then use the data driven method of deviation to restore the signal lost due to filtering. The upper bound of the uncertainty in the time series is computed by taking the standard deviation of the residual obtained after removing a long term behaviour (trend) and an annual signal from the time series [Schrama et al, 2007; Horwath et al, 2016]. To this end, the trend and the

annual signal are determined with the help of STL (Seasonal and Trend decomposition using Loess) [Cleveland et al, 1990].

STL is a procedure for decomposing a time series into trend (a long term behaviour), seasonal and remainder components. It consists of a sequence of smoothing operations, employing the loess (locally weighted regression) smoother. It has several advantages over conventional decomposition methods such as principal component analysis (PCA). STL can capture a seasonal signal with increasing amplitude, and a non-linear trend signal that represents the long term behaviour of the signal. It is a popular method for decomposing the time series in climate science and we adapt it here for time series analysis of GRACE products. In Figure 5.1, the STL decomposition of a time series is shown. We can see that the amplitude of the seasonal signal is varying with time, the trend signal exhibits the long term behaviour and the residuals contain most of the noise.

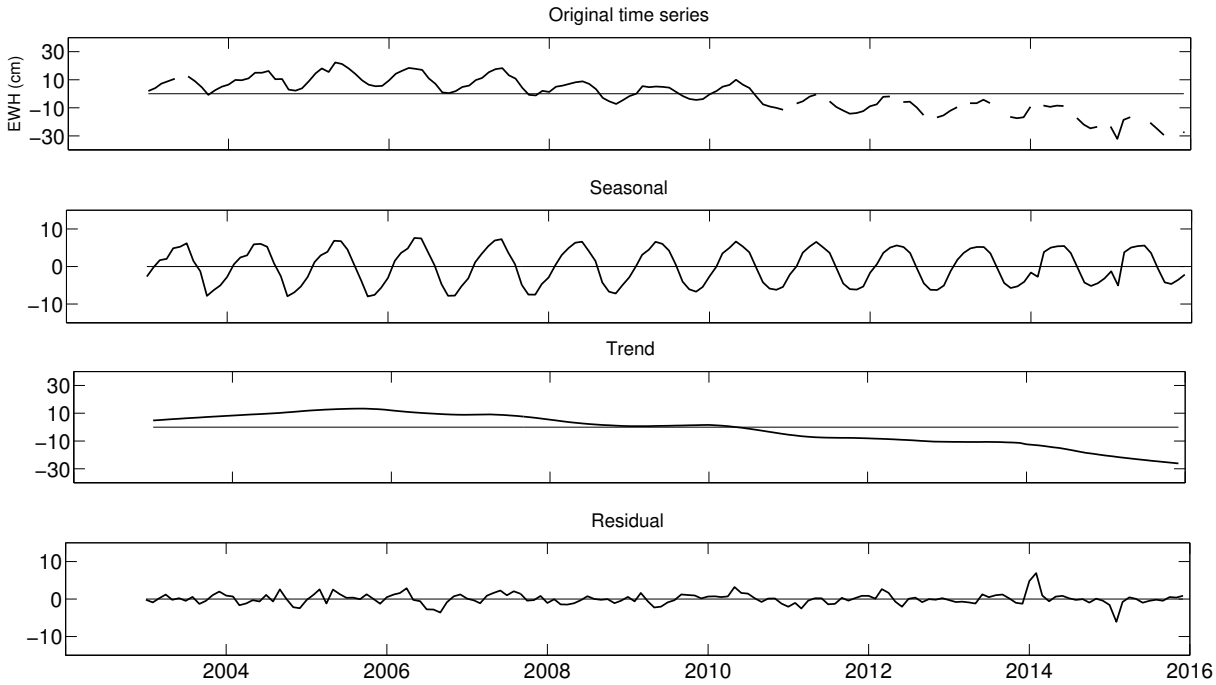


**Figure 5.1**: Decomposing a time series of EWH change over Caspian sea using STL.

We do not compute a number for indicating linear trend, instead we provide the total water volume change or the total ice mass change over a certain period of time. The total volume change is computed by multiplying the area of the catchment with the difference between the end value and the starting value of the trend component (this difference is denoted by a vertical bar in respective time series plots). We choose this convention because different definitions of trend are used by different research groups. Choosing one of them invites intensive debate [Baur, 2012], and we want to avoid any such discussion.

## §5.1 ENDORHEIC CATCHMENTS: DESICCATING WATER BODIES

An endorheic catchment is a closed drainage basin that retains water and does not allow water to flow out of it to other external bodies of water, such as rivers or oceans.

Today these catchments host big lakes, but with the water declining in many of them, we might loose a few of the lakes. Here we analyze two such water bodies: Aral sea and Urmia lake. Each one of them highlights the influence of anthropogenic activities on water resources, and the ability of GRACE to capture them.

### Aral: from sea to a desert

Aral sea was once one of the four largest lakes of the world. It is located in central Asia. Its North border is in Kazakhstan and its South border is in Uzbekistan. The two rivers Amu Darya and Syr Darya feed the sea from the South and from the North respectively. In 1960s, the Soviet government decided to divert these rivers for irrigating the desert region surrounding the Sea. Since then the water body is shrinking due to large-scale irrigation abstraction for rice and cotton fields [Singh et al, 2016]. Human intervention has decimated it to a collection of few small water bodies. In Figure 5.2, we can see the desiccation of Aral sea within a period of 3 decades. Since the majority of water was lost before the launch of GRACE satellites, we can not observe a large trend in GRACE products. However, GRACE provides an opportunity to assess the mass change in the past decade.

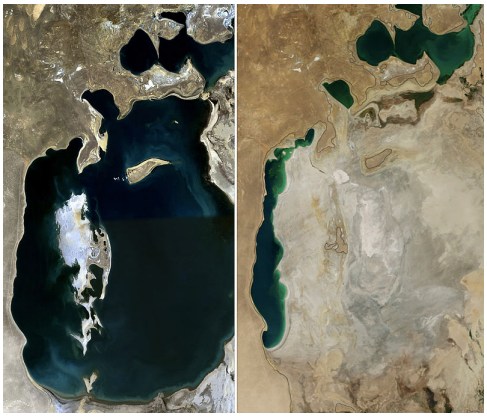

**Figure 5.2:** The water extent over Aral sea in 1989 (left) and in 2014 (right). *courtesy: wikipedia*

We have plotted the total water storage change from filtered GRACE products and the corrected GRACE products in Figure 5.3. We observe that the total volume of water loss, from filtered fields and from corrected GRACE, is $14.4 \pm 1.1\,\mathrm{km}^3$ and $19.7 \pm 1.6\,\mathrm{km}^3$ respectively. The catchment characteristic function used for computation is shown in the figure 5.3.

Singh et al [2012] demonstrated that GRACE exhibits good agreement with the water level change obtained from altimetry, but only in terms of temporal evolution. They chose a rectangular region covering Aral sea. The reason for choosing a rectangular area was to satisfy the minimum catchment area of $\approx 200000\,\mathrm{km}^2$ [Longuevergne et al, 2010]. The filtered products were then treated with the multiplicative approach to repair the signal damage due to filtering. For that purpose, a leakage computed from WGHM was removed from the filtered products followed by a scaling (scale factor $s = 2.63$). Their analysis yielded a total water volume change of $\approx 35\,\mathrm{km}^3$ for a time period of 9 years, from 2003 to 2011, which is more than double the volume obtained from the data-driven method. The reason for this discrepancy is the usage of the multiplicative method to counter impact of filtering. We have shown in Chapter 3, that the multiplicative approach amplifies both the errors in leakage (which can be large from a model) and the noise in filtered time series, by a scale factor that is big for small regions. Hence, the method has a tendency of overestimating the amplitude of total mass change.

In order to compare our findings with theirs, we also choose the same region, and

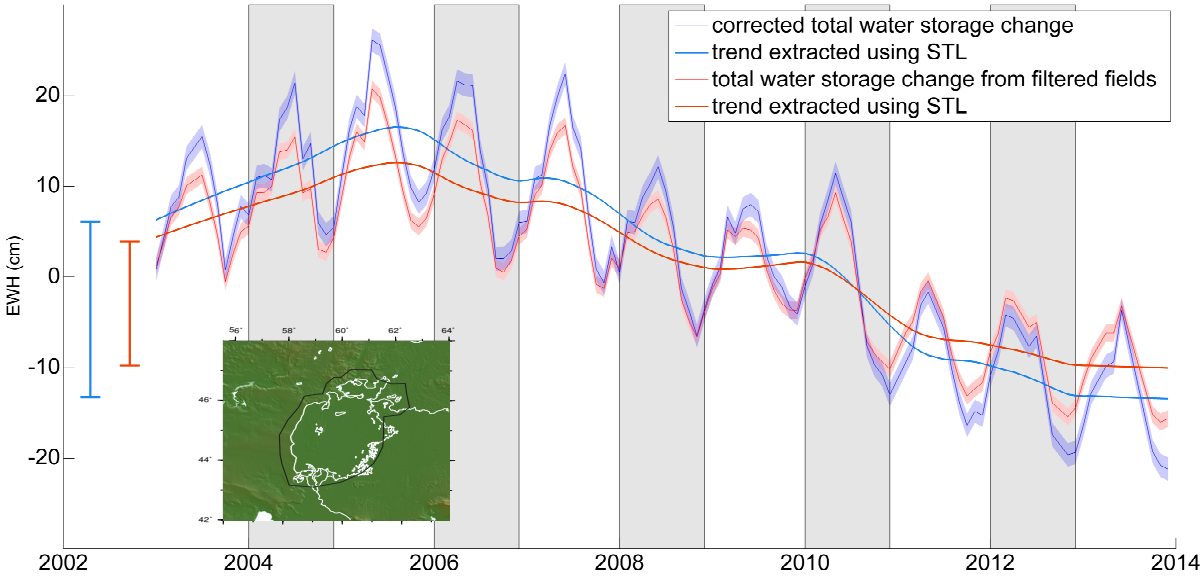


**Figure 5.3:** Total water storage change time series over Aral sea. The catchment characteristic function is the black polygon covering the Aral sea.

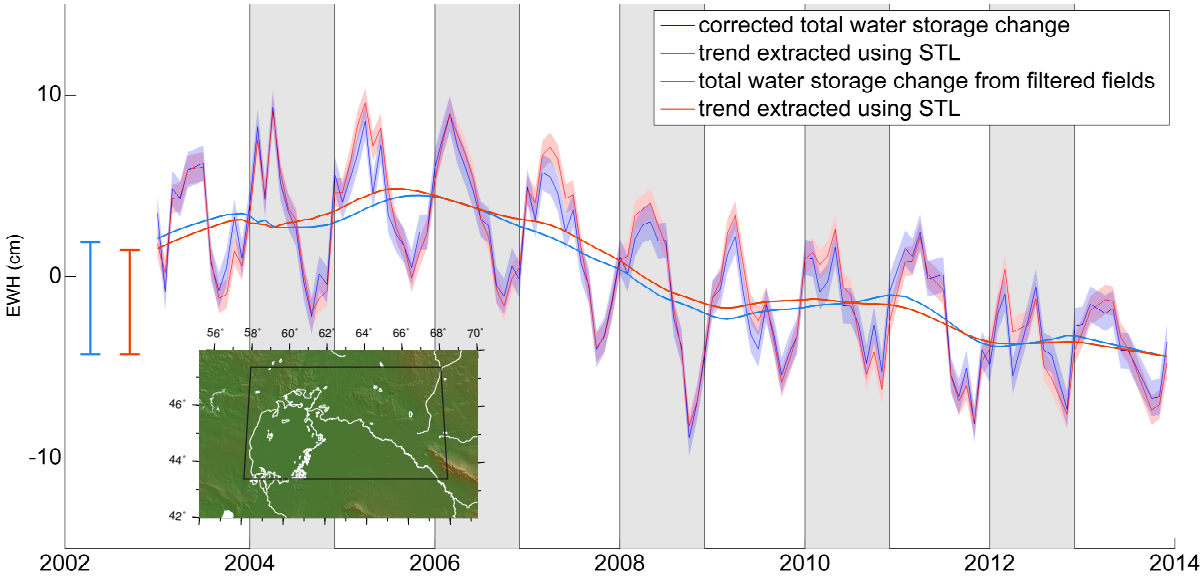


**Figure 5.4:** Total water storage change time series over a rectangular region including Aral sea. The catchment characteristic function is the black rectangle covering the Aral sea.

plot the corresponding corrected time series and time series from filtered fields (cf. Figure 5.4). Since in this region the dominant mass change signal is from Aral sea, the regional average with a bigger (rectangular) region would yield a lower value of the total water volume loss in comparison to the water volume loss from the region confined to the physical boundary of Aral sea. We observe that the volume of water loss for the rectangular region ($16.6 \pm 3.1\,\mathrm{km}^3$) is indeed less than that for the region confined to the Aral sea ($19.7 \pm 1.6\,\mathrm{km}^3$). Furthermore, in a recent contribution Singh et al [2016] used the MASCON solution from JPL. This time the total water volume loss, from 2003 to 2013, is approximately $20\,\mathrm{km}^3$, which is close to our estimate from the data-driven method of deviations. Since the MASCON solutions are shown to be superior to the GRACE spherical harmonic products [Watkins et al, 2015; Save et al, 2016], we can conclude that the data-driven approach is reliable.

### Desiccation of lake Urmia

Lake Urmia, the largest inland body of salt water in the Middle East, is now endangered. It is located in the North-West of Iran and is of great socio-economic importance. Its water level has been reported to decline since 1995 due to various climatic and anthropogenic reasons [Tourian et al, 2015]. The lake surface water extent is shrinking at an alarming rate of $220 \pm 6\,\mathrm{km}^2/\mathrm{yr}$ since 2000 (cf. Figure 5.5, [Tourian et al, 2015]). The main reason behind the rapid water loss is extraction of ground water for irrigation aggravated by climate change.

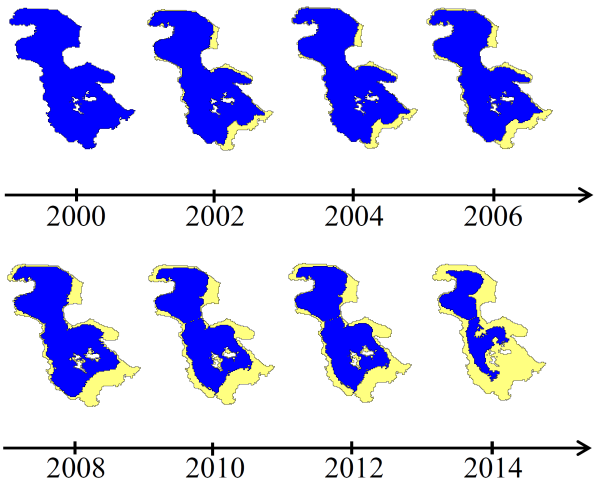


**Figure 5.5:** The desiccation of lake Urmia: snapshot of surface water extent obtained from MODIS imagery.

Although the size of this lake is below the spatial resolution of GRACE, the magnitude of mass change is large enough to register itself within the gravimetric resolution of GRACE. It is a classical example demonstrating that the ability of GRACE to sense the mass variation is not limited by the spatial resolution of approximately $200000\,\mathrm{km}^2$, as suggested by Longuevergne et al [2010]. Rather, a big mass change can be seen by GRACE irrespective of the size of the source. But after filtering, the signal attenuation and leakage makes it difficult for us to extract the information available. Hence minimizing the impact of filtering is important to observe the declining water level in the lake. For this purpose, the scaling approach from Landerer and Swenson [2012] was used by Tourian et al [2015], where the scaling factor was computed from WGHM. They reported a negative linear trend of $26.9 \pm 1.8\,\mathrm{mm/yr}$ in equivalent water height, over a period of 11 years. However, we get a trend of $21.6 \pm 0.2\,\mathrm{mm/yr}$ from the corrected time series. The difference between the two trend values can be attributed to two factors: first, they used GRACE products from GFZ, while we use ITSG products. Second, the efficacy of the model based scaling

approach is dependent on the accuracy of the model. We have demonstrated that the scaling approach performs poor for small catchments in comparison to the data-driven method of deviation (cf. Table 4.1 and Table 4.3).

Since the GRACE products from different centers should not influence the trend values by a large magnitude, we suspect and hence investigate the impact of the scaling approach. The scale factor used by Tourian et al [2015] is 1.4. If we divide their estimate of trend, $26.9 \pm 1.8$ mm/yr, by 1.4, we get $19.2 \pm 1.2$ mm/yr. The scale between the filtered product and corrected product from data-driven method is 1.2. Multiplying $19.2 \pm 1.2$ mm/yr with 1.2 gives us $23.0 \pm 1.4$ mm/yr, which is closer to the estimate from the data-driven method. Therefore, we conclude that the trend estimated in Tourian et al [2015] is most probably overestimated, and the model dependent method of scaling is to be held responsible for that. In terms of water volume, the total change between 2003 and 2013, from corrected GRACE and filtered GRACE, is $7.4 \pm 0.9\,km^3$ and $6.0 \pm 0.6\,km^3$ respectively.

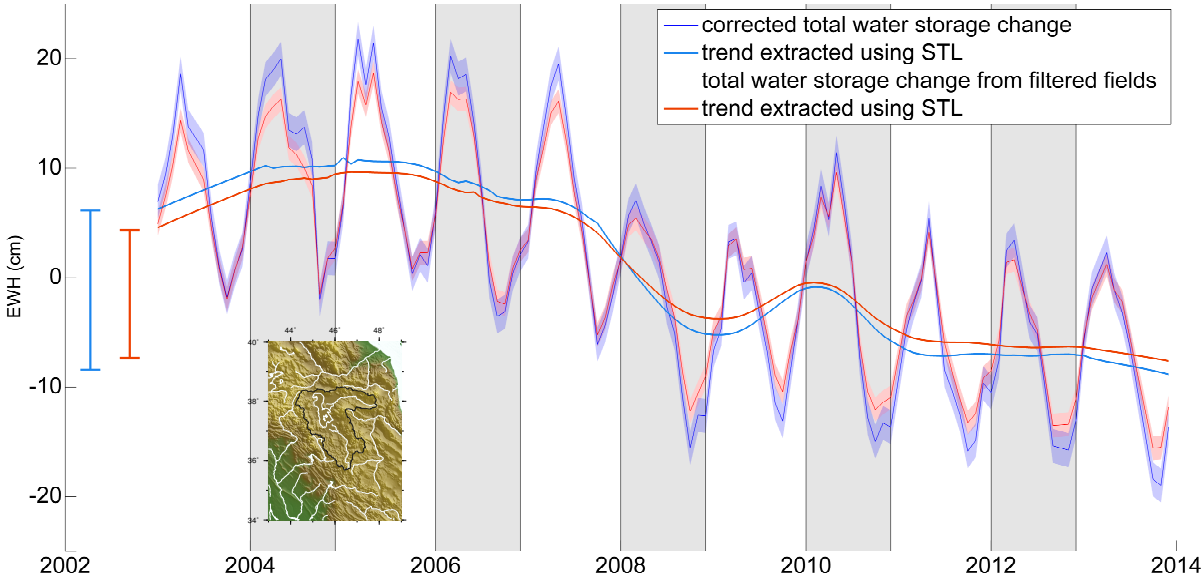


**Figure 5.6:** Total water storage change time series over Urmia lake in Iran.

## §5.2 REPLENISHMENT OF LAKE VICTORIA

Lake Victoria is situated in Africa, sharing its water with Kenya, Uganda, and Tanzania. It is the second largest fresh water lake by area, and the largest lake of Africa. It occupies an important place in the landscape of dwindling water bodies due to human intervention. The lake started loosing water after a new dam was built on the Nile in 1990s. In 2005, the lake level dropped to a level, lowest since 1951. In 2006, an independent Kenya–based hydrologic engineer reported that two dams on the Nile river, in Uganda, were the reason behind the drying Victoria lake*. These dams were diverting more than allowed volume of water for irrigation and hydroelectricity, which resulted into more outflow. Since the lake is shared between three countries, this report and the media ensured a strict action

*https://www.internationalrivers.org/resources/dams-draining-africa-s-lake-victoria-4117

and in 2006 one of the extensions of the dam was closed. This resulted into more water for lake Victoria and the water level returned to normal. The fall and rise in water mass can be seen in the GRACE time series plotted in Figure 5.7. The EWH from GRACE shows a steep fall from 2003 to early 2006 and a rise after 2006. In order to study the period of falling lake level and the period of rising lake level separately, we compute the total water mass change in two parts: from 2003 to 2005 and from 2006 to 2013. The total volume of water loss from 2003 to 2005 is $20.1 \pm 3.7\,\mathrm{km}^3$ from corrected GRACE and $13.4 \pm 2.4\,\mathrm{km}^3$ from filtered GRACE, while the water volume gained in the period between 2006 and 2013 is $25.6 \pm 3.5\,\mathrm{km}^3$ from corrected GRACE and $15.6 \pm 2.3\,\mathrm{km}^3$ from filtered GRACE. This example emphasizes that man-made problems are bigger than the climate change alone.

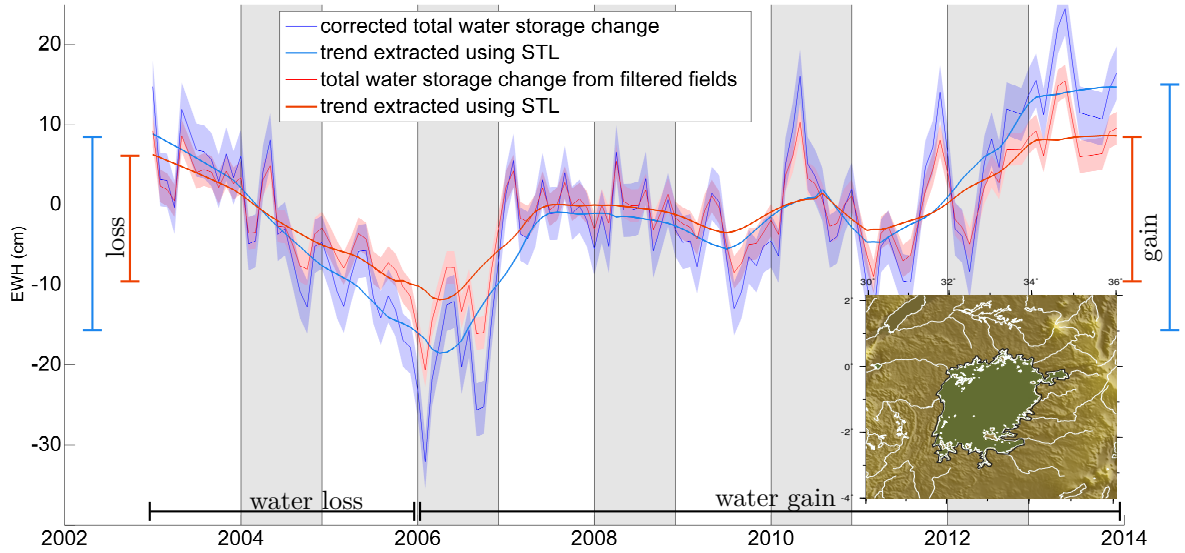


**Figure 5.7:** Total water storage change time series over Victoria lake. The catchment charateristic function used for lake Victoria follows the physical boundary of the lake shown in the figure.

## §5.3 GROUND WATER DEPLETION: THE DROUGHT IN CALIFORNIA

The ability of GRACE to monitor ground water change, was first demonstrated over North-West India [Rodell et al, 2009; Tiwari et al, 2009]. Recently, a severe drought due to overexploitation of ground water, a case similar to India, was reported in California, a state located on the Western coast of the United States. Agriculture is a significant contributor to the economy of the state and in order to keep cultivating a handsome crop, farmers are harvesting the ground water in this region for more than three decades. The source of the ground water is the Central Valley aquifer, which has dried up leading to a severe scarcity of water [Famiglietti, 2014]. Since the ground water system and the surface water system are coupled, the impact of ground water depletion is reflected in dying water reservoirs. In Figure 5.8, we can see that within three years, the water in lake Oroville has nearly disappeared.

In a report published by NASA in 2014, a mammoth 42 $km^3$ of water deficit haunts the state of California[†]. In recent years, many researchers have analyzed the problem and calculated the magnitude of heavenly bliss required to mitigate this man-made calamity [Scanlon et al, 2012; Griffin and Anchukaitis, 2014]. The numbers are growing with time, leaving us with dwindled hope. We have computed the amount of water lost by California after 2003. In Figure 5.9, we have plotted the corrected time series, the time series from filtered fields and the corresponding long term trend signal. We find that the total water lost by the state of California between 2003 and 2015, from the corrected GRACE is $84.3 \pm 7.3\,km^3$ and from filtered GRACE is $55.0 \pm 4.8\,km^3$. Over the same period and for the state of California, the water volume from filtered fields is close to the volume reported by NASA in 2014, but we have shown that filtering suppresses the trend signal. Therefore, we can safely conclude that the real amount is more, and most probably close to the volume from corrected GRACE.

**Figure 5.8**: Water level change in lake Oroville, California. The left picture is from 2011 and the right picture is from 2014. courtesy: http://sanfrancisco.cbslocal.com

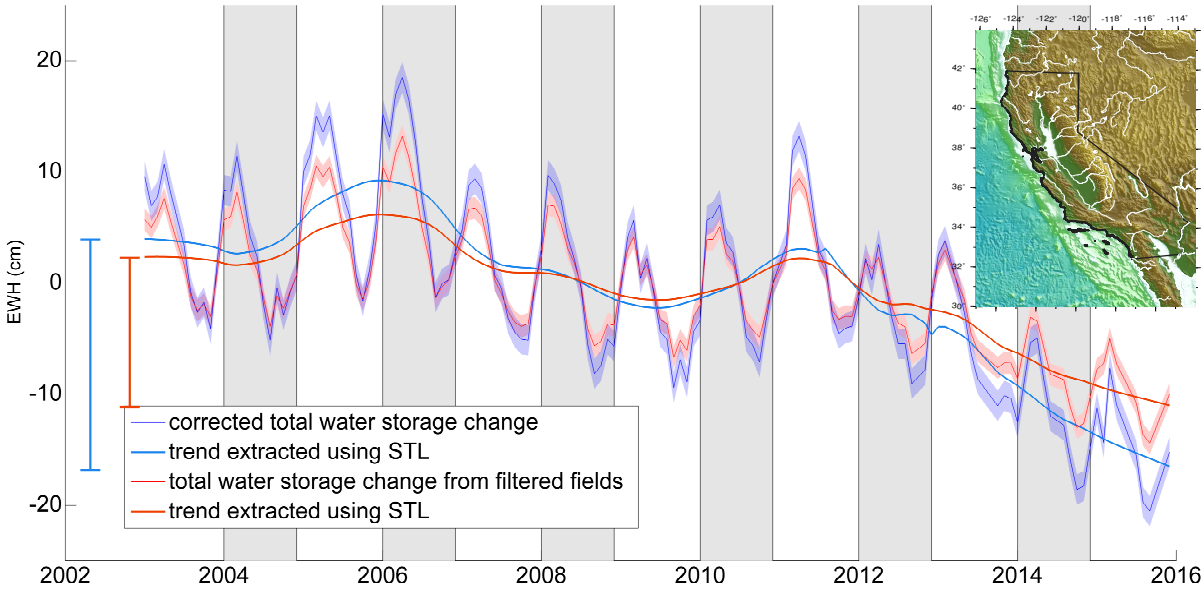


**Figure 5.9**: Total water storage change time series over California.

[†]https://www.nasa.gov/press/2014/december/nasa-analysis-11-trillion-gallons-to-replenish-california-drought-losses

## §5.4 GLOBAL HYDROLOGY FROM GRACE

We compute the total water storage change over 32 catchments listed in Table 3.1 and shown in Figure 3.10. The time series from filtered fields and from the data-driven method of deviation are shown in Figure 5.10. We can see that a substantial amount of amplitude is restored in small catchments and in Orinoco, which suffers heavily due to out of phase leakage from the Southern hemisphere. In Ganges, Indus, Godavari, and Zambezi, the linear trend in the corrected time series is more than that in the time series from filtered GRACE fields.

In order to ensure that the results obtained from processing GRACE are meaningful, we compute the Root Mean Square of the difference between the corrected time series and the time series from the filtered fields, obtained from GRACE, the Noisy-GLDAS simulation, and the GFZ GRACE simulation (cf. sections 4.1 and 4.2). RMS for these 32 catchments and their mean is given in Table 5.1. The mean of correction for Noisy-GLDAS is 1.01 cm, for GFZ GRACE it is 1.36 cm, and for GRACE it is 1.39 cm. These numbers are close to each other, which increases our confidence in the data-driven approach. Since in the closed-loop simulations the corrected time series was able to approach the truth, we can safely conclude that the increment in the GRACE signal, shown in Figure 5.10, is improvement in the signal from GRACE.

## §5.5 ICE MASS LOSS IN ANTARCTICA AND GREENLAND

The ice masses over Greenland and Antarctica are reported to be decreasing [Velicogna and Wahr, 2005; Luthcke et al, 2006; Velicogna and Wahr, 2006b; Horwath and Dietrich, 2009; Baur et al, 2009; King et al, 2012]. Consequently, the ocean mass increases with time [Chambers et al, 2004; Chambers, 2009]. However, each one of these contributions have reported a different rate of mass change (linear trend) in respective regions. The ice mass loss in Antarctica varies between $-31\,\mathrm{Gt\,yr^{-1}}$ and $-246\,\mathrm{Gt\,yr^{-1}}$ [King et al, 2012]. On the other hand, the ice mass loss trend over Greenland was reported to be $-82 \pm 28\,\mathrm{km^3\,yr^{-1}}$ in 2004, $-248 \pm 31\,\mathrm{km^3 yr^{-1}}$ in 2006 and $-258 \pm 41\,\mathrm{km^3 yr^{-1}}$ in 2013, by the same authors [Velicogna and Wahr, 2005, 2006b, 2013]. This ambiguity in the linear trend estimates arises from the differences in the data used, the time period chosen, and the choice of Glacial Isostatic Adjustment (GIA) model used in these studies. For example, the mass change trend for Antarctica and Greenland changes significantly, when we change the GIA model [King et al, 2012; Sutterley et al, 2014]. In addition, the GRACE spherical harmonic solutions from different processing centers, when processed with the same method give different trend estimates (from GFZ GRACE products: $-194 \pm 24\,\mathrm{km^3 yr^{-1}}$, from CSR GRACE products: $-242 \pm 14\,\mathrm{km^3 yr^{-1}}$, and from JPL GRACE products: $-96 \pm 23\,\mathrm{km^3 yr^{-1}}$) [Baur, 2012]. More recently mascon solutions are preferred [Arendt et al, 2013], which gives us another set of results. And finally, one can choose different filtering schemes followed by different methods to minimize signal change due to filtering, increasing the possibility of getting a different value for linear trend [Baur, 2012].

The permutation and combination of so many choices makes it difficult to identify the real ice mass loss. The impact of using one of the GRACE solution or a GIA model can

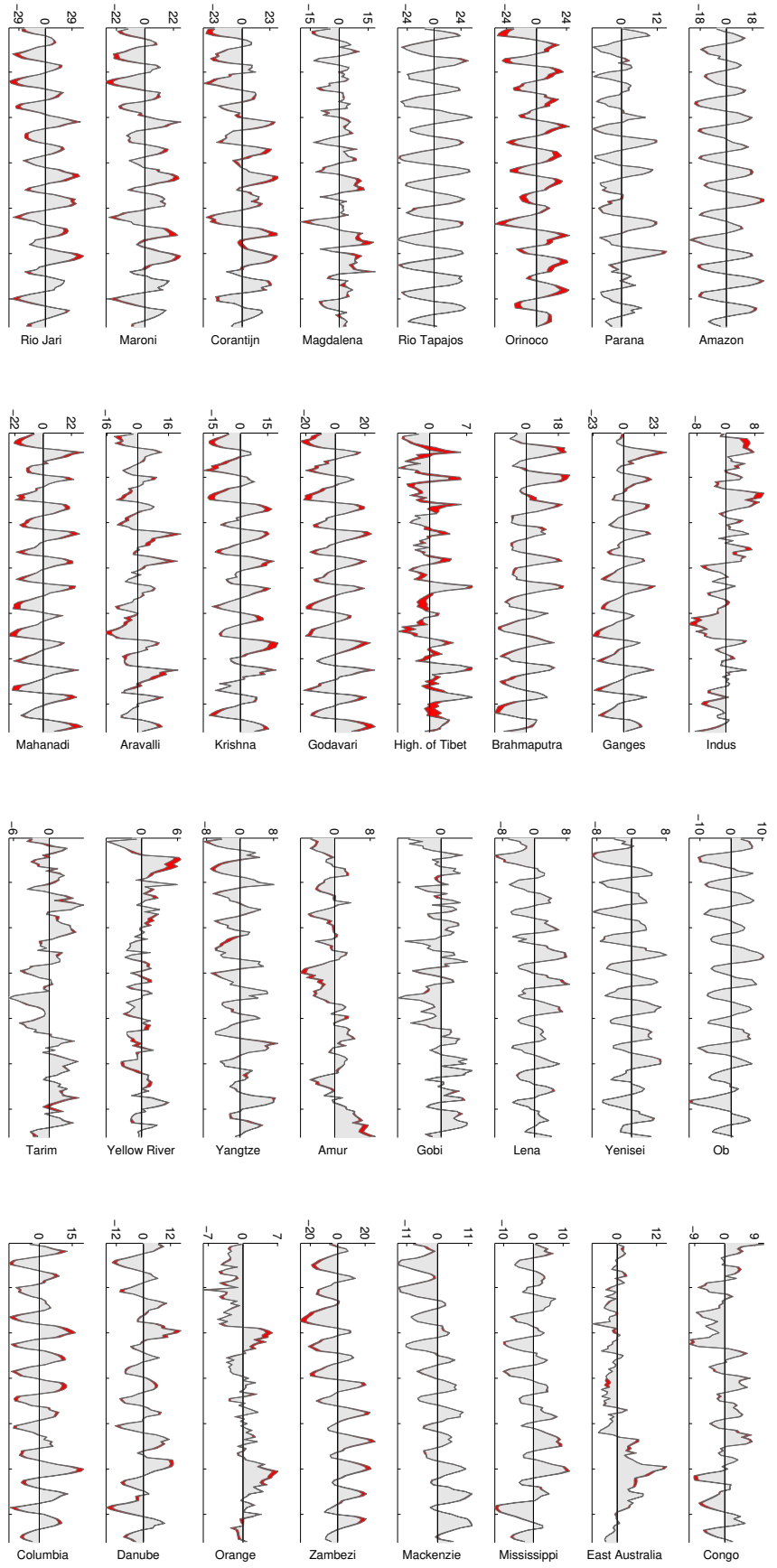


**Figure 5.10:** The time series of EWH (cm) from GRACE for 32 catchments. The increment in the signal amplitude, brought in by the data-driven method of deviation in comparison to the time series from filtered fields, is shown by the red fill. The gray fill denotes the time series from the filtered fields.

**Table 5.1:** RMS (cm) of the difference between the corrected time series and the time series from filtered fields. Each row corresponds to a catchment, while the first two columns represent the closed-loop simulation environment and the third column represents the GRACE products.

| | Noisy GLDAS | GFZ GRACE | ITSG GRACE |
|---|---|---|---|
| Amazon | 0.46 | 0.98 | 1.24 |
| Parana | 0.50 | 0.70 | 0.53 |
| Orinoco | 1.85 | 3.66 | 4.18 |
| Rio Tapajos | 1.54 | 1.36 | 1.66 |
| Magdalena | 0.81 | 2.38 | 1.64 |
| Corantijn | 2.73 | 2.51 | 2.97 |
| Maroni | 3.09 | 2.82 | 2.61 |
| Rio Jari | 3.87 | 3.24 | 3.76 |
| Indus | 0.59 | 1.08 | 0.91 |
| Ganges | 1.83 | 2.82 | 2.03 |
| Brahmaputra | 0.63 | 2.59 | 1.90 |
| High. of Tibet | 0.86 | 1.04 | 1.34 |
| Godavari | 2.04 | 2.55 | 2.72 |
| Krishna | 1.35 | 1.27 | 2.13 |
| Aravalli | 0.49 | 1.61 | 1.73 |
| Mahanadi | 2.89 | 3.05 | 3.34 |
| Ob | 0.18 | 0.32 | 0.35 |
| Yenisei | 0.16 | 0.33 | 0.28 |
| Lena | 0.24 | 0.21 | 0.31 |
| Gobi | 0.15 | 0.28 | 0.16 |
| Amur | 0.39 | 0.35 | 0.55 |
| Yangtze | 0.32 | 0.45 | 0.57 |
| Yellow River | 0.34 | 0.63 | 0.53 |
| Tarim | 0.38 | 0.52 | 0.45 |
| Congo | 0.54 | 0.78 | 0.55 |
| East Australia | 0.24 | 0.39 | 0.49 |
| Mississippi | 0.33 | 0.32 | 0.51 |
| Mackenzie | 0.24 | 0.48 | 0.46 |
| Zambezi | 1.30 | 1.57 | 1.97 |
| Orange | 0.38 | 0.88 | 0.53 |
| Danube | 0.80 | 0.97 | 1.04 |
| Columbia | 0.87 | 1.68 | 1.32 |
| **Mean** | 1.01 | 1.36 | 1.39 |

only be improved by choosing a better data or model. We prefer the widely used GIA model from Ivins and James [2005] for Antarctica and the ICE-6G model for Greenland [Argus et al, 2014; Peltier et al, 2015]. We use GRACE products form ITSG because they have have been shown to have the lowest noise level in comparison to the products from GFZ, CSR, JPL, and AIUB [Mayer-Gürr et al, 2016; Horwath et al, 2016].

In chapter 2 and chapter 3, we have demonstrated that the data-driven methods are better than model dependent approaches or a simple scaling. Therefore, we follow the data-driven method for ice-sheets to estimate the mass change signal over Antarctica and Greenland from GRACE monthly solutions. The time series of GIA corrected ice mass change from GRACE over Antarctica and over Greenland are shown in Figures 5.11 and 5.12 respectively. In Antarctica, we get a total ice mass loss of $1105.3 \pm 96.9$ Gt from filtered GRACE fields and $1460.1 \pm 134.3$ Gt from corrected GRACE fields, while in Greenland, we get a total ice mass loss of $1463.6 \pm 20.0$ Gt from filtered GRACE fields and $2750.8 \pm 37.4$ Gt from corrected GRACE fields, over a period of 13 years, from 2003 to 2015.

The ice mass estimates in Antarctica suffer from large uncertainties, which is due to the fact that the mass change signal is dominated by the post glacial rebound, which is removed with the help of models. The uncertainty in the GIA model is around 20%. Therefore, we get such large uncertainties in the final estimate. On the other hand, the contribution from post glacial rebound over Greenland is relatively small, which results into a smaller uncertainty. We conclude that filtering affects the trend signal significantly. Therefore, it is important to correct the filtered GRACE products before any assessment is made, and the data-driven methods are a better choice for the aforementioned purpose.

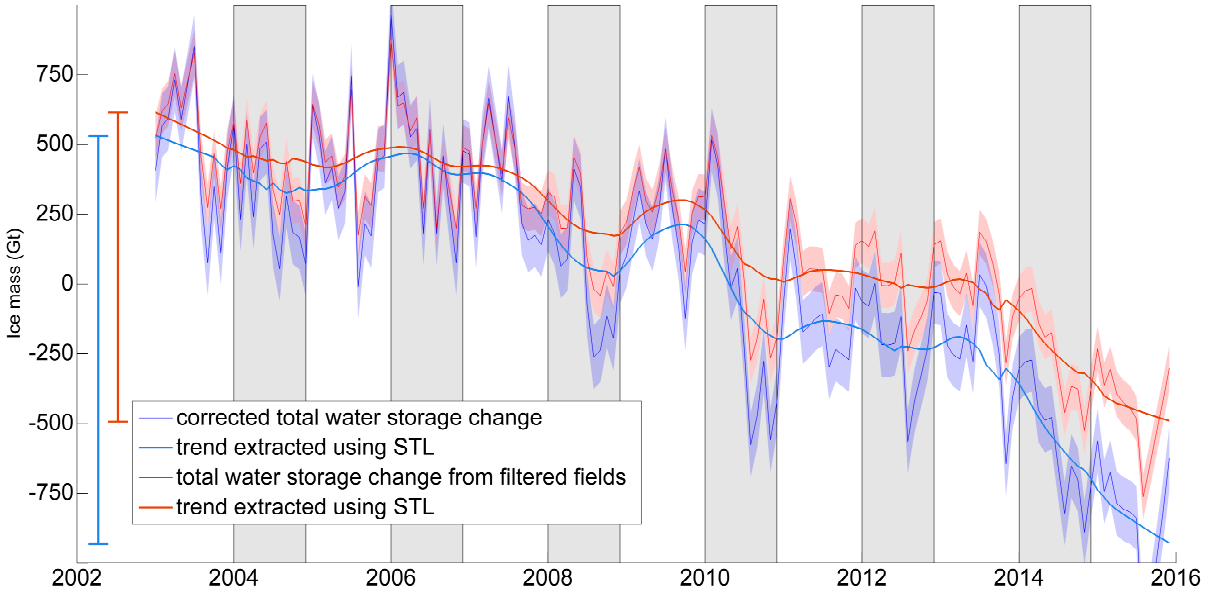


**Figure 5.11:** Total ice mass change time series over Antarctica.

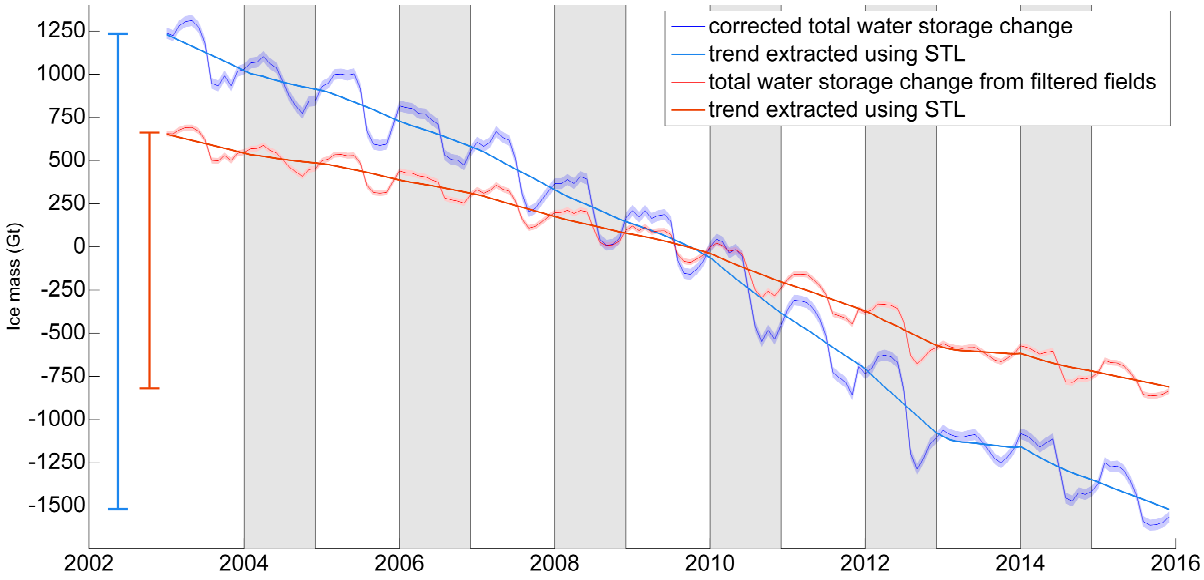


**Figure 5.12:** Total ice mass change time series over Greenland.

# 6 CONCLUSIONS AND OUTLOOK

Filtering GRACE products, correcting for the signal loss due to filtering, and then interpreting the water mass change for improving our understanding of the mass transport within the Earth system, has been a standard procedure. However, the interpretation comes with an uncertainty due to errors and the signal loss from filtering. Therefore, designing a better filter and a method to revert its impact on the signal have been hotspot for the past decade. Today we have a number of filters with different characteristics and hence a different amount of impact on the signal quality. However, a detailed mathematical understanding of the signal damage was missing, and a majority of development promoted model dependent methods to repair the signal damage. This work aimed at developing a mathematical understanding of the problem, demarcating the nature of signal loss, finding a relation to minimize the signal change, and developing a method that is not dependent on a-priori information, such as hydrological models.

## §6.1 THE JOURNEY: AIMS AND ACHIEVEMENTS

We started the discussion with defining a filter in the spatial domain, where we established that the process of convolution, along with weighting down the signal inside the catchment, invites a contribution from outside the catchment, known as leakage. Since the signal characteristic of a catchment can be different from the signal characteristic of neighbouring catchments, the leakage signal can have a different nature. When we analyzed the impact of leakage, we found that both the amplitude and the phase of the times series from filtered fields are affected, which was not reported before. The change in phase implied that only a scaling of the filtered products will not help us reach the original signal. Therefore we suggested that leakage should be removed before we tackle the weighted down catchment-confined signal. These above mentioned findings were demonstrated with the help of Table 2.2.

In order to revert the catchment-confined signal loss, we proposed two methods: *the method of scale* and *the method of deviation*. The method of scale, which uses a uniform layer assumption, requires a scale factor computed with the help of catchment characteristic function and the filter function. On the other hand, the method of deviation uses a deviation integral to rectify the signal loss. In an ideal closed-loop environment, where we can compute every component and there is no noise, we demonstrated that the method of deviation is able to recover the true signal, while the method of scale is only able

to approach the truth (cf. Figure 2.7). Nevertheless, application of these methods require knowledge of leakage and the deviation integral. Since the GRACE fields are noisy, leakage and the deviation integral from them will also be noisy. This is the reason hydrological models are employed for computing leakage, bias, scale factors, etc. These hydrological models are only as good as the data and physical modeling behind them and they tend to have large uncertainties that vary in space and time. Furthermore, choosing a model from several models is itself a topic for debate. We showed that using a model propagates the error in the model to the corrected GRACE product. Therefore, we advocate using the filtered GRACE products for estimating the leakage and the deviation integral, which provided better results in comparison to when using a model (cf. Figures 3.1 and 3.2).

We established that filtering changes the nature of the signal. Thus the leakage and the deviation integral from filtered fields are different from corresponding true values. The leakage from filtered fields suffers from an amplitude change and a phase change that is similar to the amplitude and the phase difference between the leakage from once filtered fields and twice filtered fields. On the other hand, the deviation integral demonstrated similar pattern for the amplitude but not for the phase. Hence, we proposed to shift the leakage from once filtered fields by a phase difference between the leakage from once and twice filtered fields, and then amplify the shifted leakage with an amplitude ratio between the leakage from once filtered and twice filtered fields. The deviation integral was estimated by amplifying the deviation integral from once filtered fields with an amplitude ratio between the deviation integral from once filtered and twice filtered fields. Since the leakage and the deviation integral are estimated from the GRACE fields only, we call these methods: *the data-driven method of scale* and *the data-driven method of deviation*.

The data-driven methods were tested in a closed-loop environment with GLDAS hydrological model as the background truth contaminated with GRACE-type noise. We compared the two data-driven methods with three popular model dependent approaches: multiplicative approach, additive approach and the scaling approach. We found that the data-driven methods were more accurate as well as consistent in comparison to the model dependent approaches. The methods can be arranged in order of their efficacy (from best to worst): data-driven method of deviation, data-driven method of scale, scaling approach, additive approach, and the multiplicative approach. The data-driven method of scale and the multiplicative method, given by Longuevergne et al [2010], have the same governing equation (2.28), but their performance differs because the leakage in the former method is data-driven and in the latter is from a model. Furthermore, for both methods, the accuracy declined with the size of the catchment. This is due to the dependency of the scale factor on the size of the catchment: the smaller the catchment the larger the scale factor. Hence, the error in a smaller catchment gets amplified by a larger value in comparison to that in a big catchment.

The approximation and assumptions behind the data-driven methods were tested and found true for hydrology. However, these assumptions fail for ice sheets, such as Greenland and Antarctica. This is due to the physical differences in the signal distribution: for hydrology, the mass change is throughout the catchment, while for ice sheets the mass change is near the coast and is huge in amplitude. Therefore, the method of scale, which assumes a uniform layer distribution, can not be applied. The method of deviation is

applicable, but the leakage and the deviation integral can not be determined as they were determined for hydrological catchments. Hence, we investigated and formulated another approximation, which paved the path for a data-driven method for ice sheets. The method was tested in a closed loop simulation environment and was found effective.

The aim of this contribution was to improve the total mass change estimates from GRACE. Although the validation of GRACE products is a research problem of immense magnitude, we attempted at evaluating the improvement brought in by the data-driven method of deviation, within the framework of closing the water budget. We could not conclude anything from that exercise because the correction from the data-driven method of deviation was an order of magnitude smaller than the misclosure due to errors in precipitation, evapotranspiration and runoff. However, we extract confidence from the performance of data-driven methods in a realistic GRACE-type closed-loop simulation setup provided by GFZ.

The data-driven method of deviation, which promises better signal quality from GRACE, was employed to assess the water mass loss in Aral sea, lake Victoria, lake Urmia, California, Antarctica, and Greenland. We observed that the corrected GRACE products show a higher amount of mass loss in comparison to the mass loss computed from filtered GRACE products. We also compare the results with a few published works that have used model dependent approaches for correction. We concluded that these methods have estimated a rate of mass loss with large uncertainties. Furthermore, from many research contributions we can deduce that, more often than not, anthropogenic exploitation of water resource is responsible for sustained regional droughts [Rodell et al, 2009; Tiwari et al, 2009; Famiglietti, 2014; Tourian et al, 2015; Singh et al, 2016]. If we can regulate the water use carefully and efficiently, we can ensure water safety and prosperity of the region.

The results from the data-driven method for ice sheets have been contributed to the Ice sheet Mass Balance Inter-comparison Exercise (IMBIE). The aim of this exercise is to compute a reconciled estimates of ice sheet mass balance from different methods [Shepherd et al, 2012]. We estimated a linear trend of $-99.1 \pm 40.4\,\mathrm{Gt\,yr^{-1}}$ over Antarctica and $-229.85 \pm 7.7\,\mathrm{Gt\,yr^{-1}}$ over Greenland. The previous IMBIE exercise in 2012, reported a trend of $-72 \pm 43\,\mathrm{Gt\,yr^{-1}}$ over Antarctica and $-232 \pm 23\,\mathrm{Gt\,yr^{-1}}$ over Greenland [Shepherd et al, 2012]. The analysis and inter-comparison of ice mass loss from various institutes and from various methods is yet to be done. However, through personal communication we obtained the linear trend estimated by colleagues at Institute of Planetary Geodesy, TU Dresden. They have reported $-105.4 \pm 35.3\,\mathrm{Gt\,yr^{-1}}$ for Antarctica and $-257.9 \pm 14.6\,\mathrm{Gt\,yr^{-1}}$ for Greenland. They use the same data as we have used, but they minimize noise with the help of the tailored sensitivity kernel approach [Horwath and Dietrich, 2009]. The difference between their results and our results is not significant for Antarctica. However, for Greenland the disagreement is significant, which can be attributed to the differences in the method employed to tackle noise.

## §6.2 THE BIG PICTURE

In the midst of global climate change, and under pressure of anthropogenic over-exploitation, many regions are suffering from a severe water scarcity. For example, North-West India,

California, North-West Iran, Aral sea, South Africa, and many more. In order to manage the water resource, we must know

- how much water we have lost?
- what is the rate at which the exploitation of water is increasing?
- how is it related to the rate of population growth and the agricultural growth?
- how much of the water lost is replenished by precipitation?
- and so on.

Since water is stored in a column including surface water, soil moisture, and groundwater, optical remote sensing and altimetry offer only a limited insight. Dedicated land surveys lead us to limited spatial coverage and poor temporal coverage. However, the GRACE satellite mission enables us to monitor the changes in hydrosphere with a good temporal and a decent spatial resolution. The information from the GRACE mission can be used for water resource monitoring and planning. A dedicated group of researchers, under the umbrella of European Gravity Service for Improved Emergency Management (EGSIEM) project, are already working to use GRACE, in tandem with other satellite information, to predict floods and droughts. Nevertheless, the reliability of such an experiment and their possible products depend heavily on the accuracy of GRACE products.

In the past decade, a lot of development has been made in every aspect and in every dimension of exploring the changes in the gravity field of the Earth via GRACE satellites. The amount of research findings have increased non-linearly with time. For the same region we can find a number of publications, each using a different post-processing method, providing us different numerical values to mark the severity of the problem. This work itself is guilty of providing one more post-processing strategy, but with a promise that it has a better potential in comparison to the model dependent approaches.

## §6.3 PUTTING THE PEN DOWN

Vince Lombardi said "Perfection is not attainable, but if we chase perfection we can catch excellence"[*]. Every work suffers from a few caveats and unfinished tasks. Sometimes the amount of time required is just not available, and sometimes the effort is not worth the results. Nevertheless, it is important to know the loose ends and limitations of the finished job.

This contribution has improved our understanding of the signal damage due to filtering, especially from a mathematical point of view, which has helped us to design the data-driven methods for repairing the signal damage. These methods can be applied to GRACE products filtered with any filter, but assuming that the filter suppresses the noise to a minimum. However, every filter will reduce the noise to a different amount, which makes the performance of data driven methods vary with the filter definition. Throughout this work, we have used the optimum Gaussian filter of half width radius 400 km. Nevertheless,

[*] from BrainyQuote.com

the impact of changing the filter is yet to be studied. In a preliminary analysis we found that in general the impact of changing filter is more prominent in smaller catchments. This is shown in Figure 6.1, where we have plotted the RMS of the error in corrected time series. The spread in error RMS with respect to filter definition increases as the size of the catchment decreases. Furthermore, we have discussed in chapter 4 that the accuracy of the data-driven method of deviation declines with the catchment size. However, the impact of catchment shape and size is not investigated thoroughly. Such a study might help us improve the efficacy of data-driven methods even further.

While understanding the signal damage, we encountered three different scenarios: active hydrological catchments, arid regions or deserts, and the ice sheets. We have discussed and drafted the data-driven method, with case specific approximations, for active hydrological catchments and ice sheets. However, a detailed analysis and a dedicated method for arid regions, such as Gobi, Highland of Tibet, Sahara desert, Thar desert and many more, is still missing. Furthermore a comparative analysis of data-driven method for ice sheets with other sophisticated methods, such as point mass modeling and tailored sensitivity kernel approach, is yet to be done.

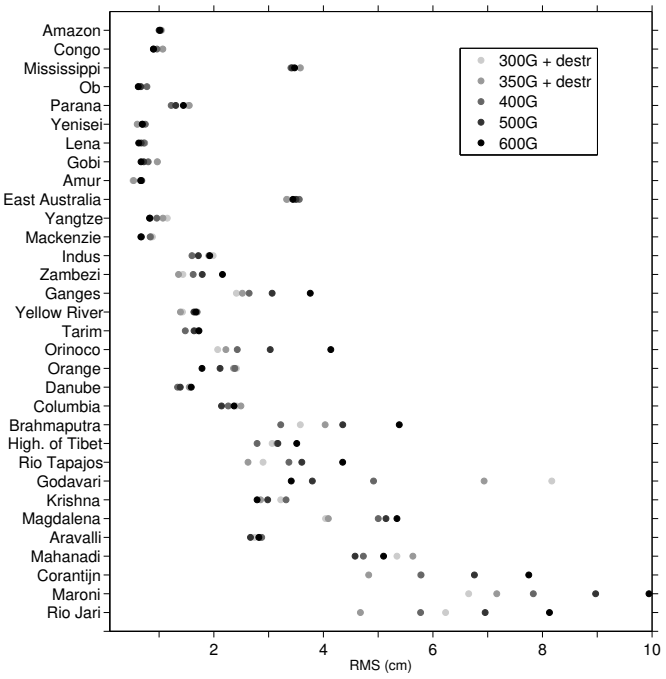


**Figure 6.1:** The RMS of error for data-driven method of deviation over 32 catchments with different filters. The catchments are sorted by their area, and we use 5 filter definitions. The filtered fields are corrected with data-driven method of deviation and it is compared to the true time series.

Within the past two years, we have seen the rise of a new type of GRACE products: mascon solutions, which do not require filtering at the user end. However, they are constructed using constrained regularization, which itself is a type of filtering. Therefore, the quality of their signal, although claimed to be better, and the damage due to regularization is yet to be tested and discussed in the community. An analysis similar to this work would provide us a better insight.

In conclusion, we believe that this contribution is a valuable addition to the toolkit for processing GRACE products to obtain better mass change estimates at catchment scale.

# BIBLIOGRAPHY


Arendt A, Luthcke S, Gardner A, O'Neel S, Hill D, Moholdt G, Abdalati W (2013) Analysis of a GRACE global mascon solution for Gulf of Alaska glaciers. Journal of Glaciology 59(217):913–924, DOI 10.3189/2013JoG12J197

Argus DF, Peltier WR, Drummond R, Moore AW (2014) The Antarctica component of postglacial rebound model ICE-6GC (VM5a) based on GPS positioning, exposure age dating of ice thicknesses, and relative sea level histories. Geophysical Journal International 198:537–563, DOI 10.1093/gji/ggu140

Baur O (2012) On the computation of mass-change trends from GRACE gravity field time-series. Journal of Geodynamics 61:120–128, DOI 10.1016/j.jog.2012.03.007

Baur O, Kuhn M, Featherstone WE (2009) GRACE-derived ice-mass variations over Greenland by accounting for leakage effects. Journal of Geophysical Research: Solid Earth 114(B6):B06,407, DOI 10.1029/2008JB006239

Chambers DP (2009) Calculating trends from GRACE in the presence of large changes in continental ice storage and ocean mass. Geophysical Journal International 176(2):415–419, DOI 10.1111/j.1365-246X.2008.04012.x

Chambers DP, Wahr J, Nerem RS (2004) Preliminary observations of global ocean mass variations with GRACE. Geophysical Research Letters 31(13):L13,310, DOI 10.1029/2004GL020461

Chao BF, Gross RS (1987) Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes. Geophysical Journal of the Royal Astronomical Society 91(3):569–596, DOI 10.1111/j.1365-246X.1987.tb01659.x

Cheng M, Tapley BD, Ries JC (2013) Deceleration in the Earth's oblateness. Journal of Geophysical Research: Solid Earth 118(2):740–747, DOI 10.1002/jgrb.50058

Cleveland R, Cleveland W, McRae J, Terpenning I (1990) STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion). Journal of Official Statistics 6:3–73

Dahle C, Flechtner F, Gruber C, König D, König R, Michalak G, Neumayer KH (2012) GFZ GRACE Level-2 Processing Standards Document for Level-2 Product Release 05. Scientific Technical Report - Data 12/02, GFZ, Potsdam, Potsdam, DOI 10.2312/GFZ.b103-12020

Devaraju B (2015) Understanding filtering on the sphere – Experiences from filtering GRACE data. PhD thesis, Universität Stuttgart, URL elib.uni-stuttgart.de/bitstream/11682/4002/1/BDevarajuPhDThesis.pdf

Döll P, Müller Schmied H, Schuh C, Portmann FT, Eicker A (2014) Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. Water Resources Research 50(7):5698–5720, DOI 10.1002/2014WR015595

Duan XJ, Guo JY, Shum CK, van der Wal W (2009) On the postprocessing removal of correlated errors in GRACE temporal gravity field solutions. Journal of Geodesy 83(11):1095–1106, DOI 10.1007/s00190-009-0327-0

Famiglietti JS (2014) The global groundwater crisis. Nature Climate Change 4(19):945–948, DOI 10.1038/nclimate2425

Fenoglio-Marc L, Kusche J, Becker M (2006) Mass variation in the Mediterranean Sea from GRACE and its validation by altimetry, steric and hydrologic fields. Geophysical Research Letters 33(19):L19,606, DOI 10.1029/2006GL026851

Flechtner F, Neumayer KH, Dahle C, Dobslaw H, Fagiolini E, Raimondo JC, Güntner A (2016) What Can be Expected from the GRACE-FO Laser Ranging Interferometer for Earth Science Applications? Surveys in Geophysics 37(2):453–470, DOI 10.1007/s10712-015-9338-y

Griffin D, Anchukaitis KJ (2014) How unusual is the 2012–2014 California drought? Geophysical Research Letters 41(24):9017–9023, DOI 10.1002/2014GL062433, 2014GL062433

Hamming RW (1998) Digital filters, third edition. Dover publications, New york

Han SC, Shum CK, Jekeli C, Kuo CY, Wilson C, Seo KW (2005) Non-isotropic filtering of GRACE temporal gravity for geophysical signal enhancement. Geophysical Journal International 163(1):18–25, DOI 10.1111/j.1365-246X.2005.02756.x

Haykin S, Moher M (2009) Communication Systems. Wiley

Helmert FR (1880) Mathematical and Physical Theories of Higher Geodesy, Part 1, Preface and the Mathematical Theories. Aeronautical Chart and Information Center (St. Louis, 1964).

Horwath M, Dietrich R (2009) Signal and error in mass change inferences from GRACE: the case of Antarctica. Geophysical Journal International 177(3):849–864, DOI 10.1111/j.1365-246X.2009.04139.x

Horwath M, Groh A, the EGSIEM team (2016) Evaluation of recent GRACE monthly solution series with an ice sheet perspective. Geophysical Research Abstracts 18:EGU2016–9728

Ivins ER, James TS (2005) Antarctic glacial isostatic adjustment: a new assessment. Antarctic Science 17:541–553, DOI 10.1017/S0954102005002968

Jekeli C (1981) Alternative methods to smooth the Earth's gravity field. Tech. Rep. 327, Department of Geodetic Science and Surveying, The Ohio State University

Kaula W (1966) Theory of Satellite Geodesy: Applications of Satellites to Geodesy. Special series of brief books covering selected topics in the Pure and applied sciences, Blaisdell Pub. Co.

King AM, Bingham JR, Moore P, Whitehouse LP, Bentley JM, Milne AG (2012) Lower satellite-gravimetry estimates of Antarctic sea-level contribution. nature 491:586–589, DOI 10.1038/nature11621

Klees R, Zapreeva EA, Winsemius HC, Savenije HHG (2007) The bias in GRACE estimates of continental water storage variations. Hydrology Earth System Sciences, 11(4) pp 1227–1241, DOI 10.5194/hess-11-1227-2007

Klees R, Revtova EA, Gunter BC, Ditmar P, Oudman E, Winsemius HC, Savenije HHG (2008) The design of an optimal filter for monthly GRACE gravity models. Geophysical Journal International 175(2):417–432, DOI 10.1111/j.1365-246X.2008.03922.x

Kusche J (2007) Approximate decorrelation and non-isotropic smoothing of time-variable GRACE-type gravity field models. Jornal of Geodesy 81(11):733–749, DOI 10.1007/s00190-007-0143-3

Landerer FW, Swenson SC (2012) Accuracy of scaled GRACE terrestrial water storage estimates. Water Resources Research 48(4):W04,531, DOI 10.1029/2011WR011453

Landerer FW, Dickey JO, Güntner A (2010) Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003-2009. Journal of Geophysical Research 115(D23):D23,115, DOI 10.1029/2010JD014584

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

Longuevergne L, Scanlon BR, Wilson CR (2010) GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. Water Resources Research 46(11):W11,517, DOI 10.1029/2009WR008564

Lorenz C, Devaraju B, Tourian MJ, Sneeuw N, Riegger J, Kunstmann H (2014) Large-scale runoff from landmasses: a global assessment of the closure of the hydrological and atmospheric water balances. Journal of Hydrometeorology 15:2111–2139, DOI 10.1175/JHM-D-13-0157.1

Luthcke SB, Zwally HJ, Abdalati W, Rowlands DD, Ray RD, Nerem RS, Lemoine FG, McCarthy JJ, Chinn DS (2006) Recent Greenland ice mass loss by drainage system from satellite gravity observations. Science 314(5803):1286–1289, DOI 10.1126/science.1130776

Luthcke SB, Sabaka T, Loomis B, Arendt A, McCarthy J, Camp J (2013) Antarctica, Greenland and Gulf of Alaska land-ice evolution from an iterated GRACE global mascon solution. Journal of Glaciology 59(216):613–631, DOI doi:10.3189/2013JoG12J147

Mayer-Gürr T, Behzadpour S, Ellmer M, Kvas A, Klinger B, Zehentner N (2016) ITSG-Grace2016 - monthly and daily gravity field solutions from GRACE. Website, GFZ Data Services, DOI 10.5880/icgem.2016.007, online available at https://www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2016/

Nash JE, Sutcliffe JV (1970) River flow forecasting through conceptual models part I—A discussion of principles. Journal of hydrology 10(3):282–290

Paarmann LD (2001) Design and Analysis of Analog Filters: A Signal Processing Perspective. Springer US, DOI 10.1007/b100752

Peltier WR, Argus DF, Drummond R (2015) Space geodesy constrains ice age terminal deglaciation: The global ICE-6GC (VM5a) model. Journal of Geophysical Research: Solid Earth 120(1):450–487, DOI 10.1002/2014JB011176, 2014JB011176

Phillips T, Nerem RS, Fox-Kemper B, Famiglietti JS, Rajagopalan B (2012) The influence of ENSO on global terrestrial water storage using GRACE. Geophysical Research Letters 39(16):L16,705, DOI 10.1029/2012GL052495

Rice J (1995) Mathematical Statistics and Data Analysis. No. v. 1 in Duxbury advanced series, Duxbury Press

Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, Bosilovich MG, Schubert SD, Takacs v, Kim GK, Bloom S, Chen J, Collins D, Conaty A, Silva Ad, Gu W, Joiner J, Koster RD, Lucchesi R, Molod A, Owens T, Pawson S, Pegion P, Redder CR, Reichle R, Robertson FR, Ruddick AG, Sienkiewicz M, Woollen J (2011) MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. Jornal of Climate 24:3624–3648, DOI 0.1175/JCLI-D-11-00015.1

Rodell M, Houser PR, Jambor U, Gottschalck J, Mitchell K, Meng CJ, Arsenault PR, Cosgrove B, Radakovich J, Bosilovich M, Entin JK, Walker JP, Lohmann D, Toll D (2004) The global land data assimilation system. Bulletin of American Meteorological Society 85(3):381–394

Rodell M, Velicogna I, Famiglietti JS (2009) Satellite-based estimates of groundwater depletion in India. nature 460(7258):999–1002, DOI 10.1038/nature08238

Save H, Bettadpur S, Tapley BD (2016) High-resolution CSR GRACE RL05 mascons. Journal of Geophysical Research: Solid Earth 121(10):7547–7569, DOI 10.1002/2016JB013007, 2016JB013007

Scanlon BR, Longuevergne L, Long D (2012) Ground referencing GRACE satellite estimates of groundwater storage changes in the California Central Valley, USA. Water Resources Research 48(4):W04,520, DOI 10.1029/2011WR011312

Schrama EJO, Wouters B, Lavallée DA (2007) Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations. Journal of Geophysical Research: Solid Earth 112(B8):B08,407, DOI 10.1029/2006JB004882

Shepherd A, Ivins ER, A G, Barletta VR, Bentley MJ, Bettadpur S, Briggs KH, Bromwich DH, Forsberg R, Galin N, Horwath M, Jacobs S, Joughin I, King MA, Lenaerts JTM, Li J, Ligtenberg SRM, Luckman A, Luthcke SB, McMillan M, Meister R, Milne G, Mouginot J, Muir A, Nicolas JP, Paden J, Payne AJ, Pritchard H, Rignot E, Rott H, Sørensen LS, Scambos TA, Scheuchl B, Schrama EJO, Smith B, Sundal AV, van Angelen JH, van de Berg WJ, van den Broeke MR, Vaughan DG, Velicogna I, Wahr J, Whitehouse PL, Wingham DJ, Yi D, Young D, Zwally HJ (2012) A reconciled estimate of ice-sheet mass balance. Science 338(6111):1183–1189, DOI 10.1126/science.1228102

Shum CK, Guo Jy, Hossain F, Duan J, Alsdorf DE, Duan Xj, Kuo Cy, Lee H, Schmidt M, Wang L (2011) Climate Change and Food Security in South Asia (Inter-annual Water Storage Changes in Asia from GRACE Data), vol 1. Springer Netherlands, DOI 10.1007/978-90-481-9516-9

Singh A, Seitz F, Schwatke C (2012) Inter-annual water storage changes in the Aral Sea from multi-mission satellite altimetry, optical remote sensing, and GRACE satellite gravimetry. Remote Sensing of Environment 123:187–195, DOI http://dx.doi.org/10.1016/j.rse.2012.01.001

Singh A, Seitz F, Eicker A, Güntner A (2016) Water budget analysis within the surrounding of prominent lakes and reservoirs from multi-sensor earth observation data and hydrological models: Case studies of the Aral Sea and lake Mead. Remote Sensing 8(11):953, DOI 10.3390/rs8110953

Sneeuw N, Lorenz C, Devaraju B, Tourian MJ, Riegger J, Kunstmann H, Bárdossy A (2014) Estimating runoff using hydro-geodetic approaches. Surveys in Geophysics 35(6):1333–1359, DOI 10.1007/s10712-014-9300-4

Sutterley TC, Velicogna I, Csatho B, van den Broeke M, Rezvan-Behbahani S, Babonis G (2014) Evaluating Greenland glacial isostatic adjustment corrections using GRACE, altimetry and surface mass balance data. Environmental Research Letters 9(1):014,004, DOI 10.1088/1748-9326/9/1/014004

Swenson S (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(B9):2193, DOI 10.1029/2001JB000576

Swenson S, Wahr J (2006) Post-processing removal of correlated errors in GRACE data. Geophysical Research Letters 33(8):L08,402, DOI 10.1029/2005GL025285

Swenson S, Chamber D, Wahr J (2007) Estimating geocenter variations from a combination of GRACE and ocean model output. Journal of Geophysical Research 113:B08,410, DOI 10.1029/2007JB005338

Taylor PT, Keating T, Kahn WD, Langel RA, Smith DE, Schnetzler CC (1983) GRM: Observing the terrestrial gravity and magnetic fields in the 1990's. Eos, Transactions American Geophysical Union 64(43):609–611, DOI 10.1029/EO064i043p00609

Tiwari VM, Wahr J, Swenson S (2009) Dwindling groundwater resources in Northern India, from satellite gravity observations. Geophysical Research Letters 36(18):L18,401, DOI 10.1029/2009GL039401

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

Velicogna I, Wahr J (2005) Greenland mass balance from GRACE. Geophysical Research Letters 32(18):L18,505, DOI 10.1029/2005GL023955, l18505

Velicogna I, Wahr J (2006a) Acceleration of Greenland ice mass loss in spring 2004. Nature 443(7109):329–331, DOI 10.1038/nature05168

Velicogna I, Wahr J (2006b) Measurements of time-variable gravity show mass loss in Antarctica. Science 311:1754–1756, DOI 10.1126/science.1123785

Velicogna I, Wahr J (2013) Time-variable gravity observations of ice sheet mass balance: Precision and limitations of the GRACE satellite data. Geophysical Research Letters 40(12):3055–3063, DOI 10.1002/grl.50527

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

Wagner CA (1987) Improved gravitational recovery from a Geopotential Research Mission Satellite pair flying en echelon. Journal of Geophysical Research: Solid Earth 92(B8):8147–8155, DOI 10.1029/JB092iB08p08147

Wahr J, Molenaar M, Bryan F (1998) Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. Journal of Geophysical Research 103(B12):30,205

Wahr J, Swenson S, Velicogna I (2007) Some hydrological and cryospheric applications of GRACE. In: Proceedings of the GRACE Science Team Meeting and DFG SPP1257 Symposium

Watkins MM, Wiese DN, Yuan DN, Boening C, Landerer FW (2015) Improved methods for observing Earth's time variable mass distribution with GRACE using spherical cap mascons. Journal of Geophysical Research: Solid Earth 120(4):2648–2671, DOI 10.1002/2014JB011547

Werth S (2010) Calibration of the global hydrological model wghm with water mass variations from grace gravity data. PhD thesis, Universität Potsdam, URL https://publishup.uni-potsdam.de/frontdoor/index/index/docId/3996

Werth S, Güntner A, Schmidt R, Kusche J (2009) Evaluation of GRACE filter tools from a hydrological perspective. Geophysical Journal International 179(3):1499–1515, DOI 10.1111/j.1365-246X.2009.04355.x

Zhang ZZ, Chao BF, Lu Y, Hsu HT (2009) An effective filtering for GRACE time-variable gravity: Fan filter. Geophysical Research Letters 36(17):L17,311, DOI 10.1029/2009GL039459

# A

# DATA AND MODELS

We describe the data and models used. Every data is global except for runoff from GRDC, which is available for a limited number of catchments.

**WGHM**: WaterGAP Global Hydrology Model, provides global $0.5° \times 0.5°$ monthly total water storage change in mm. We have used it from 2004 to 2009 [Döll et al, 2014].

**GLDAS**: Global Land Data Assimilation System, provides global $0.5° \times 0.5°$ monthly total water storage change in mm. We have used it from 2004 to 2009 [Rodell et al, 2004].
*source*: `disc.sci.gsfc.nasa.gov/uui/datasets?keywords=GLDAS`

**MERRA-Land**: Modern Era Retrospective-analysis for Research and Applications - Land, provides global $0.5° \times 0.5°$ monthly total water storage change in mm. We have used it from 2004 to 2009 [Rienecker et al, 2011].

**CMAP**: CPC Merged Analysis of Precipitation, provides global $0.5° \times 0.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `www.esrl.noaa.gov/psd/data/gridded/data.cmap.html`.

**CRU**: Climatic Research Unit, provides global $0.5° \times 0.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `badc.nerc.ac.uk/data/cru`.

**Del**: Willmott, C. J. and K. Matsuura, provide global $0.5° \times 0.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `climate.geog.udel.edu/~climate/html_pages/archive.html`.

**PRECL**: PRECipitation REConstruction over Land, provides global $0.5° \times 0.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `www.esrl.noaa.gov/psd/data/gridded/data.precl.html`.

**GPCP**: Global Precipitation Climatology Project, provides global $2.5° \times 2.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html`.

**GPCC**: Global Precipitation Climatology Center, provides global $0.5° \times 0.5°$ precipitation in mm/month. We have used it from 2004 to 2010.
*source*: `gpcc.dwd.de`.

**GLDAS**: Global Land Data Assimilation System, provides global $0.25° \times 0.25°$ evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*:`disc.sci.gsfc.nasa.gov/uui/datasets/GLDAS_NOAH025_M_V2.0/summary`.

**GLEAM**: Global Land Evaporation Amsterdam Model, provides global evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*: www.gleam.eu.

**MOD16**: MODIS Global Evapotranspiration Project, provides global evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*: www.ntsg.umt.edu/project/mod16.

**FLUXNET MTE**: FLUXNET Multi-Tree Ensemble, provides global evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*: climatedataguide.ucar.edu/climate-data/fluxnet-mte-multi-tree-ensemble.

**MERRA-Land**: Modern Era Retrospective-analysis for Research and Applications - Land, provides global evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*: gmao.gsfc.nasa.gov.

**ERA Interim**: European Centre for Medium-Range Weather Forecasts (ECMWF), provides global evapotranspiration in mm/month. We have used it from 2004 to 2010.
*source*: apps.ecmwf.int/datasets/data/interim-full-moda/levtype=sfc.

**GRDC runoff**: Global Runoff Data Centre provides the river Discharge measurements for various rivers.
*source*: www.bafg.de/GRDC.

**GRACE spherical harmonic coefficients from GFZ**: GeoForschungZentrum in Potsdam provides the level 2 GRACE data. These are dimensionless fully normalized spherical harmonic coefficients upto degree and order 90 [Dahle et al, 2012].
*source*: ftp://podaac-ftp.jpl.nasa.gov/GeodeticsGravity/grace/L2/GFZ/RL05/

**GRACE spherical harmonic coefficients from ITSG**: Institute of Geodesy in Graz, provides the level 2 GRACE data. These are dimensionless fully normalized spherical harmonic coefficients upto degree and order 90 [Mayer-Gürr et al, 2016].
*source*: www.tugraz.at/institute/ifg/downloads/gravity-field-models/itsg-grace2016

# B

## CHANGING THE FILTER

We plot the error in time series in Figures B.1 and B.2, and report the RMS and NSE in Tables B.1 and B.2. The error is the deviation from truth in the GFZ GRACE closed loop simulation environment. We use a destriping filter along with a Gaussian filter of half width radius 350 km. We can see that the data-driven method of deviation is performing better than other approaches in terms of RMS. The scaling approach with this filter improves and the performance of data-driven method of deviation decays, in comparison to their performance with the Gaussian 400 km filter.

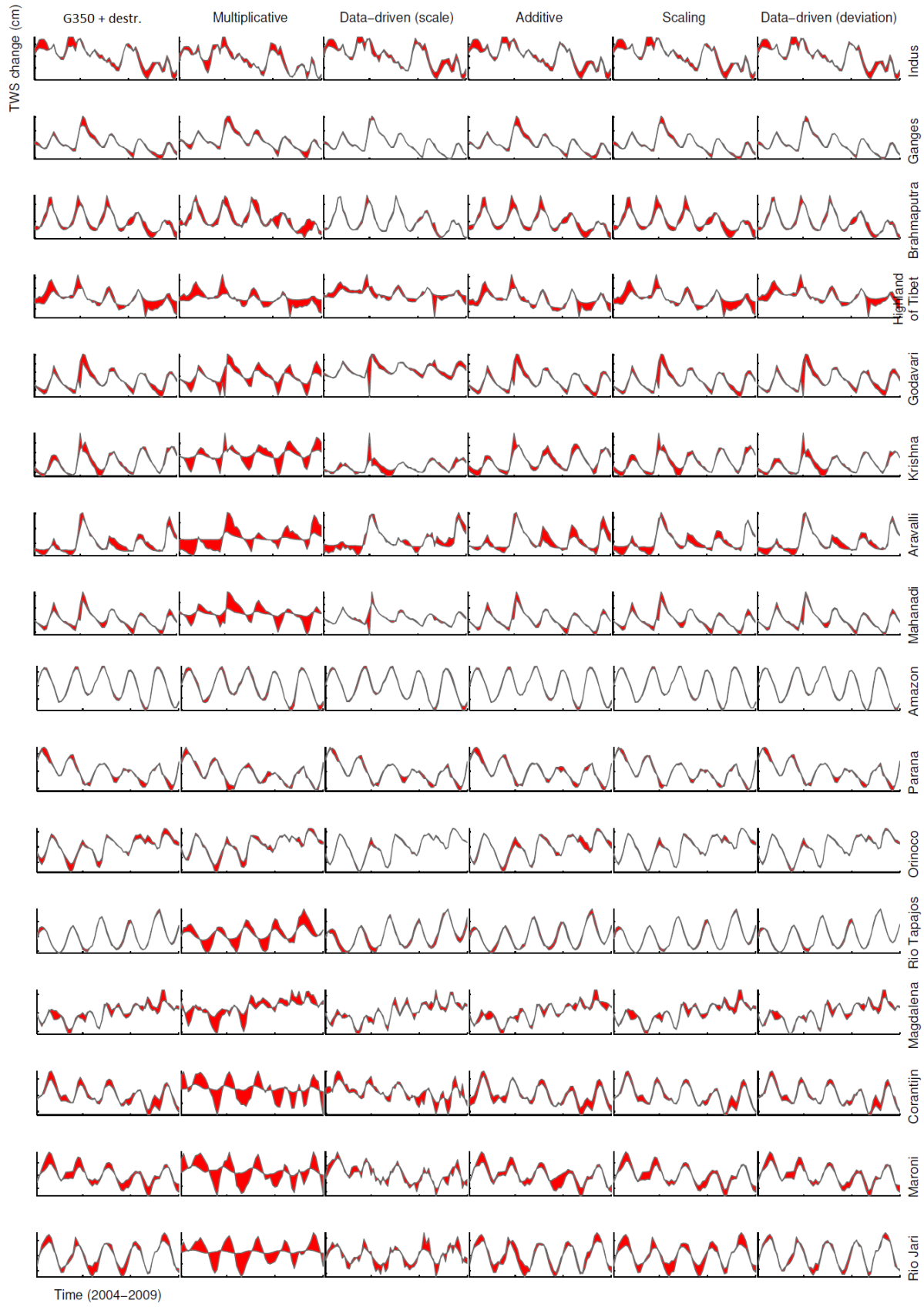


**Figure B.1:** Fill plots comparing different approaches for catchment number 1 to 16 in Table 3.1. The ordinate of each subplot is different. The GFZ GRACE simulation fields are the input fields, and WGHM helps the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

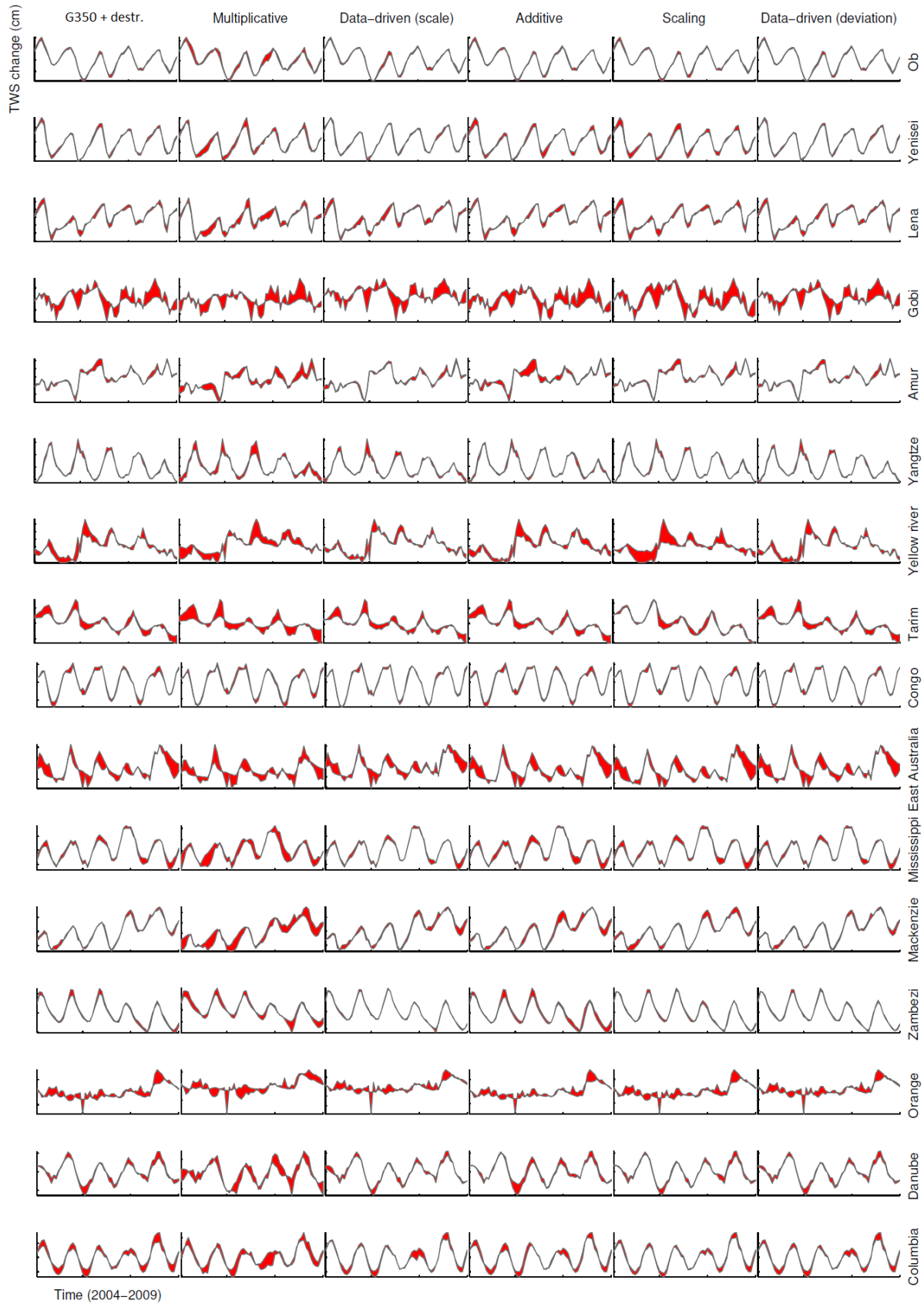


**Figure B.2:** Fill plots comparing different approaches for catchment number 17 to 32 in Table 3.1. The ordinate of each subplot is different. The GFZ GRACE simulation fields are the input fields, and WGHM helps the model dependent approaches. The amount of red is the difference between the true time series and the result from the corresponding approach. Since the ordinate for every method for a catchment is same (except for multiplicative approach), less red means better performance.

**Table B.1:** RMS (cm) of the error in the corrected time series computed in the GFZ GRACE simulation environment. Each row corresponds to a catchment and each column to a method. The smallest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments and dry catchments.

| | G350 + desrt. | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 2.11 | 2.67 | 2.14 | 1.99 | 1.94 | **1.91** |
| Ganges | 4.24 | 6.43 | 2.07 | 4.10 | 3.31 | **2.52** |
| Brahmaputra | 5.53 | 10.00 | **2.75** | 5.79 | 6.50 | 4.03 |
| High. of Tibet | 3.16 | 7.43 | 3.56 | 2.24 | **3.05** | 3.15 |
| Godavari | 6.00 | 18.40 | 11.10 | **5.37** | 5.56 | 6.93 |
| Krishna | **2.07** | 13.77 | 5.31 | 2.82 | 2.54 | 2.85 |
| Aravalli | 2.98 | 30.51 | 4.02 | 4.08 | 3.93 | **2.85** |
| Mahanadi | 6.30 | 69.29 | 11.31 | **5.41** | 6.14 | 5.63 |
| Amazon | 1.22 | 2.31 | 1.61 | 1.16 | 1.05 | **1.02** |
| Parana | 1.82 | 2.21 | 1.62 | 1.75 | **1.42** | 1.55 |
| Orinoco | 4.65 | 5.57 | **1.90** | 4.43 | 2.43 | 2.22 |
| Rio Tapajos | 2.62 | 29.00 | 4.92 | 3.84 | **2.60** | 2.62 |
| Magdalena | 4.75 | 18.10 | 4.51 | 4.53 | 4.66 | **4.09** |
| Corantijn | 6.17 | 78.73 | 14.42 | 6.61 | 4.91 | **4.83** |
| Maroni | 9.18 | 77.49 | 11.12 | 9.23 | 7.85 | **7.17** |
| Rio Jari | 6.55 | 174.46 | 17.22 | 7.53 | 8.73 | **4.68** |
| Ob | 0.64 | 1.57 | 0.73 | 0.68 | 0.68 | **0.62** |
| Yenisei | 0.68 | 1.53 | 0.66 | 1.03 | 1.08 | **0.60** |
| Lena | **0.72** | 1.19 | 0.82 | 0.78 | 0.82 | 0.74 |
| Gobi | 0.80 | 1.31 | 1.05 | 0.79 | **0.57** | 0.97 |
| Amur | 0.62 | 1.51 | 0.55 | 1.03 | 0.61 | **0.53** |
| Yangtze | 0.88 | 2.44 | 1.54 | 0.94 | **0.85** | 1.07 |
| Yellow River | 1.52 | 3.44 | 1.42 | 1.54 | 2.06 | **1.39** |
| Tarim | 1.50 | 4.65 | 1.86 | 1.54 | **0.71** | 1.73 |
| Congo | 1.45 | 1.72 | **1.06** | 1.38 | 1.17 | 1.07 |
| East Australia | 1.47 | 2.65 | 1.36 | 1.42 | **1.27** | 1.41 |
| Mississippi | 1.13 | 2.78 | **0.87** | 1.30 | 1.17 | 0.94 |
| Mackenzie | **0.77** | 3.23 | 1.11 | 1.26 | 1.02 | 0.85 |
| Zambezi | 2.42 | 6.53 | **1.22** | 3.82 | 1.66 | 1.36 |
| Orange | 2.08 | 4.32 | 2.71 | **1.88** | 2.14 | 2.36 |
| Danube | 1.89 | 4.65 | 1.53 | 2.06 | **1.50** | 1.55 |
| Columbia | 3.28 | 6.12 | 2.50 | 2.81 | **2.03** | 2.49 |
| **Mean** | 2.85 | 18.63 | 3.76 | 2.97 | 2.68 | **2.42** |
| **Mean*** | 2.24 | 6.27 | 2.31 | 2.37 | 2.03 | **1.93** |

**Table B.2:** NSE values of the corrected time series computed in the GFZ GRACE simulation environment. Each row corresponds to a catchment and each column to a method. The largest value in a row is in bold, and the next in magnitude is underlined. There are four major groups separated by space. Each group represents one region, and is sorted by the area of the catchments. Small catchments, below the filter resolution, are separated within the group by a horizontal line. **Mean*** denotes the mean value computed excluding the small catchments and dry catchments.

| | G350 + desrt. | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
|---|---|---|---|---|---|---|
| Indus | 0.83 | 0.73 | 0.83 | 0.85 | **0.86** | **0.86** |
| Ganges | 0.90 | 0.78 | **0.98** | 0.91 | 0.94 | 0.97 |
| Brahmaputra | 0.82 | 0.41 | **0.96** | 0.80 | 0.75 | 0.90 |
| High. of Tibet | -1.00 | -10.03 | -1.53 | **0.00** | -0.86 | -0.98 |
| Godavari | 0.74 | -1.47 | 0.10 | **0.79** | 0.77 | 0.65 |
| Krishna | **0.80** | -7.63 | -0.31 | 0.63 | 0.70 | 0.62 |
| Aravalli | **0.80** | -19.78 | 0.63 | 0.62 | 0.65 | 0.82 |
| Mahanadi | 0.82 | -19.75 | 0.43 | **0.87** | 0.83 | 0.86 |
| Amazon | **0.99** | 0.95 | 0.97 | **0.99** | **0.99** | **0.99** |
| Parana | 0.91 | 0.87 | 0.93 | 0.92 | **0.95** | 0.94 |
| Orinoco | 0.89 | 0.85 | **0.98** | 0.90 | 0.97 | **0.98** |
| Rio Tapajos | **0.97** | -2.57 | 0.90 | 0.94 | **0.97** | **0.97** |
| Magdalena | 0.78 | -2.21 | 0.80 | 0.80 | 0.79 | **0.84** |
| Corantijn | 0.78 | -19.17 | -0.23 | 0.74 | **0.86** | **0.86** |
| Maroni | 0.70 | -15.99 | 0.56 | 0.69 | 0.78 | **0.82** |
| Rio Jari | 0.85 | -102.03 | -0.03 | 0.80 | 0.73 | **0.92** |
| Ob | **0.98** | 0.88 | 0.97 | **0.98** | **0.98** | **0.98** |
| Yenisei | 0.97 | 0.85 | 0.97 | 0.93 | 0.92 | **0.98** |
| Lena | **0.92** | 0.79 | 0.90 | 0.91 | 0.90 | **0.92** |
| Gobi | -1.06 | -4.47 | -2.58 | -1.01 | **-0.04** | -2.03 |
| Amur | 0.91 | 0.46 | **0.93** | 0.75 | 0.91 | **0.93** |
| Yangtze | **0.96** | 0.67 | 0.87 | 0.95 | 0.96 | 0.95 |
| Yellow River | 0.68 | -0.65 | 0.72 | 0.67 | 0.41 | **0.73** |
| Tarim | 0.46 | -4.17 | 0.17 | 0.43 | **0.88** | 0.28 |
| Congo | 0.96 | 0.94 | **0.98** | 0.96 | 0.97 | **0.98** |
| East Australia | 0.18 | -1.65 | 0.30 | 0.24 | **0.39** | 0.25 |
| Mississippi | 0.92 | 0.52 | **0.95** | 0.90 | 0.92 | **0.95** |
| Mackenzie | **0.96** | 0.42 | 0.93 | 0.91 | 0.94 | **0.96** |
| Zambezi | 0.96 | 0.72 | **0.99** | 0.91 | 0.98 | **0.99** |
| Orange | 0.45 | -1.38 | 0.06 | **0.55** | 0.42 | 0.29 |
| Danube | 0.89 | 0.33 | 0.93 | 0.87 | **0.93** | **0.93** |
| Columbia | 0.84 | 0.44 | 0.91 | 0.88 | **0.94** | 0.91 |
| **Mean** | 0.70 | -6.29 | 0.50 | 0.72 | **0.75** | 0.68 |
| **Mean*** | 0.81 | -0.40 | 0.75 | 0.81 | **0.83** | **0.83** |

**Table B.3:** Comparison of different methods applied with Gaussian 350 km and the destriping filter in GFZ closed-loop simulation environment. $n_x^*$ denotes the number of catchment, out of 32, for which the method in respective column is either the best or second best in terms of the statistical measure $x$. $n_x$ denotes the same for the case when the method is best.

| | | Methods | | | | | |
|---|---|---|---|---|---|---|---|
| | | Gaussian 400 | Multiplicative | Data-driven (scale) | Additive | Scaling | Data-driven (deviation) |
| GFZ GRACE simulation | $n^*_{\mathrm{RMS}}$ | 10 | 0 | 13 | 5 | 14 | 24 |
| | $n_{\mathrm{RMS}}$ | 3 | 0 | 5 | 3 | 9 | 12 |
| | $n^*_{\mathrm{NSE}}$ | 16 | 0 | 14 | 12 | 22 | 25 |
| | $n_{\mathrm{NSE}}$ | 8 | 0 | 7 | 6 | 11 | 18 |

"...Vanishing groundwater will translate into major declines in agricultural productivity and energy production, with the potential for skyrocketing food prices and profound economic and political ramifications. Further declines in groundwater availability may well trigger more civil uprising and international violent conflict in the already water-stressed regions of the world, and new conflict in others.

From North Africa to the Middle East to South Asia, regions where it is already common to drill over 2 km to reach groundwater, it is highly likely that disappearing groundwater could act as a flashpoint for conflict. Managing the global groundwater crisis will require raising awareness of these critical issues to the level of everyday understanding. The actions outlined above are important steps in that direction. Once elected officials, environmental decision-makers and the general public truly understand the sources of water, and how they are affected by climate change, overuse and population growth, the need for action will be clear. Full appreciation of the importance of groundwater to the global water supply and security is essential for managing this global crisis, and for vastly improving management of all water resources for the generations to come."

– *J. S. Famiglietti*,
The global groundwater crisis
Nature Commentary, Volume 4, Nov 2014.