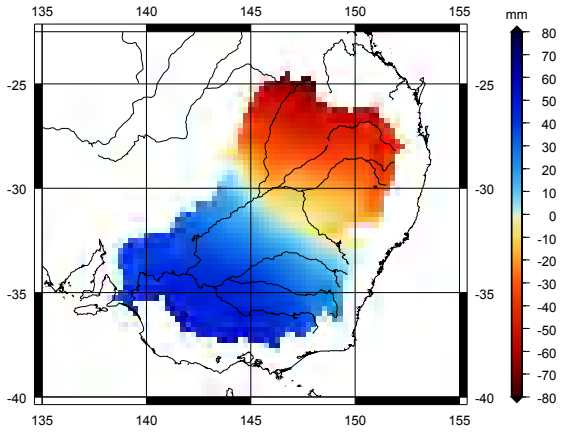
Universität Stuttgart
Geodätisches Institut

Curtin University of Technology
Western Australian Centre for Geodesy
& The Institute of Geoscience Research

Assessing hydrological changes in the Murray-Darling Basin



Diplomarbeit im Studiengang

Geodäsie und Geoinformatik

an der Universität Stuttgart

Stefanie Schmid

Perth, Juli 2010

Betreuer: Prof. Dr. sc. techn. Wolfgang Keller
Universität Stuttgart

Dr.-Ing. Michael Kuhn
Curtin University of Technology

Erklärung der Urheberschaft

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher in gleicher oder ähnlicher Form in keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Ort, Datum Unterschrift

Acknowledgements

I would like to thank several people who are involved in the evolution of this thesis:

- Prof. Dr. sc. techn. Wolfgang Keller, my supervisor at University of Stuttgart, who proposed me to do my thesis at Curtin University of Technology in Australia. Furthermore, I want to thank for his expertise and support whenever I had a question.
- Dr.-Ing. Michael Kuhn, my supervisor at Curtin University of Technology, for his help, guidance, and support. Many discussions helped me to broaden my view and knowledge. Moreover, I want to thank for providing the routines for Principal Component Analysis. He also helped me with the search for available data.
- Dipl.-Ing. Daniel Rieser for providing the Matlab routines for the Multiple Linear Regression Analysis.
- Dr.-Ing. Oliver Baur, for the provision and help with the routines for mass estimate calculation.
- M.Sc. Ira Anjasmara for her help with the handling of the routines for Principal Component Analysis and GMT.
- Murray-Darling Basin Authority, New South Wales, and Queensland for providing in-situ hydrological data (water level and river flow)

Abstract

In this diploma thesis GRACE-derived time-variable gravity, TRMM precipitation, and in-situ hydrological observations (water level, river flow) between May 2003 and April 2009 (6 years) are used to assess hydrological changes in the Murray-Darling Basin (MDB). The investigations in terms of Multiple Linear Regression Analysis (MLRA) and Principal Component Analysis (PCA) show distinct differences between the northern and southern climatic regions of the MDB. While the north is dominated by precipitation occurring within the early part of each year, precipitation appears in the south half a year later. Furthermore, a severe drought is detected, which appeared in 2006/07. The analysis of seasonal variations reveals that gravity changes are preceded by precipitation by about one month in the north of the MDB, while the south shows the inverse behaviour. The phase shift of gravity changes between the north and the south suggests a mass transport taking about 5 months to cross the MDB from the north to the south, which is confirmed by the in-situ hydrological observations.

Key words: Time-variably gravity, GRACE, Murray-Darling Basin, hydrology, mass transport, Principal Component Analysis, Multiple Linear Regression Analysis, Correlation Analysis

Zusammenfassung

Um hydrologische Änderungen im Murray-Darling Becken (MDB) aus den von GRACE abgeleiteten zeitlich variablen Schweredaten abschätzen zu können, werden diese Daten im Zeitraum von Mai 2003 bis April 2009 (6 Jahre) zusammen mit den Niederschlagsdaten der TRMM-Mission und den hydrologischen Beobachtungen vor Ort (Wasserstand, Abfluss) verwendet. Untersuchungen mittels Multipler Linearer Regressionsanalyse und Hauptkomponentenzerlegung zeigen deutliche Unterschiede zwischen den nördlichen und südlichen Klimaregionen des MDBs. Während der Norden durch Niederschläge dominiert ist, welche Anfang jedes Jahres auftreten, erscheinen Niederschläge im Süden ein halbes Jahr später. Desweiteren wird eine strenge Dürre detektiert, welche in den Jahren 2006/07 auftrat. Die Analyse jahreszeitlicher Variationen zeigt, dass Niederschläge im Norden ein Monat vor Schwereänderungen auftreten, während der Süden das gegenteilige Verhalten aufzeigt. Die Phasenverschiebung der Schwereänderungen zwischen dem Norden und Süden weist auf einen Massentransport mit einer Dauer von ca. 5 Monaten vom Norden zum Süden hin. Dieser Massentransport wird durch die hydrologischen Beobachtungen vor Ort bestätigt.

Schlüsselwörter: Zeitlich variable Schwere, GRACE, Murray-Darling Becken, Hydrologie, Massentransport, Hauptkomponentenzerlegung, Multiple Lineare Regressionsanalyse, Korrelationsanalyse

Contents

List of Figures

List of Tables

1 Introduction

> *"... while the world's fresh-water resources are massive, they are not only unevenly distributed in both space and time but also have a finite limit in terms of practical utilisation. ... improving knowledge of the globe's water resources is indispensable for the well-being of mankind and for the protection of the environment. Reliable information on the state and the trends of water resources is a prerequisite for sound decisions on their sustainable management."*

The above quotation from the publication 'Water Resources Assessment - Progress in the Implementation of the Mar del Plata Action Plan and Strategy of the 1990s' published by the WMO/UNESCO, succintly states the importance of water monitoring. This is of particular importance for Australia with its predominantly arid climate. The knowledge of Australia's water resources over time is crucial for water management. Australia's climate varies from a tropical climate in the north, to the temperate climate in both the south-west and south-east. When viewed as a whole, Australia is a very arid country, with 80% of the continent having less than 600 millimetres of rain per year and 50% of that has even less than 300 millimetres per year. Approximately 40% of Australia's landmass is covered by sand dunes (e.g. Wikipedia (2010a)). Therefore, the management of water resources is a big, but necessary challenge for the future. To ensure the continuity of Australia's drinking water supplies, monitoring of hydrological changes is very important. In this thesis, the hydrological changes in the Murray Darling Basin (MDB) located in the south-east of Australia are examined. The MDB is Australia's largest river basin as well as being one of the world's major river systems. Throughout recent history, the MDB has already overcome several periods of drought (e.g. Nicholls (2004)). Dry conditions, along with reduced water availability for irrigators persist throughout the Murray-Darling Basin. The MDB is vital to Australia, in that it is home to large-scale agricultural activity.

Earlier studies, undertaken throughout the whole of Australia, have revealed some interesting relationships between time variable gravity observations from the Gravity Recovery and Climate Experiment (GRACE) satellite mission, and precipitation data in the MDB. These relationships suggest that the MDB is governed by

two distinctively different regimes in the northern and southern part (e.g. Rieser (2008)). While the north is dominated by precipitation at the beginning of a year, precipitation occurs in the southern part of the MDB approximately half a year later. For example a comparison between both data sets suggests a phase lag of approximately 1 to 2 months in the northern part, and a negative phase lag of 1 to 2 months in the southern part, whereby in the north precipitation occurs 1 month before the detection of gravity changes and the inverse behaviour happens in the southern part. Another study, which used GRACE data together with in-situ hydrological data, showed the reduction of water resources caused by a multiyear drought (e.g. Leblanc et al. (2009)). In this study a mean value of the annual changes of groundwater storage for the whole Murray-Darling Basin was determined.

To study the general behaviour of hydrological changes in the MDB, this study uses time-variable gravity observations (expressed as surface mass changes) from the GRACE-satellite mission, along with precipitation data from the Tropical Rainfall Measurement Mission (TRMM). Furthermore, in-situ hydrological observations (water level and river flow) are used. Since its launch in 2002 the GRACE satellite mission delivers maps of the Earth's spatio-temporal gravity field with a spatial and temporal resolution of approximately 400-600 km and 10-30 days, respectively. These gravity changes are directly linked to mass changes on or close to the Earth's surface. GRACE delivers the integrated gravity measurement of all mass changes. Because of this, a separation (mostly in the vertical) of different sources of mass changes cannot be performed by GRACE. The TRMM mission is designed to detect and study tropical rainfall. The satellite observations are available on a grid with a 0.25 degree resolution covering the world in a latitude band between 50° S and 50° N. These measurements from space (TRMM and GRACE) are compared with in-situ hydrological observations. These observations are water level and river flow, which are distributed over the whole MDB. While the water level represents the height of the water in metres, river flow describes the mass of water, which flows through the river at the gauging station with respect to a given time interval (e.g. day).

The above mentioned differing hydrological regimes in the north and south of the MDB should be verified, in order that a key question be answered, being: "Is it possible to detect the water mass movements from the north to the south of the MDB with the aid of GRACE and hydrological data?"

The three different data types are analysed in terms of temporal and spatial changes over the last six years (e.g. May 2003 - April 2009) using both the Multiple Linear Regression Analysis (MLRA) and the Principle Component Analysis (PCA). Finally, Correlation Analysis is applied to examine how good the three different data types fit together.

First basic background knowledge is given in chapter 2. Techniques for the determination of the Earth's gravity field are shown. This chapter ends with the presentation of the three data types used within this thesis. Chapter 3 provides the methodology, with which the data types are examined. In Chapter 4 and 5 results are presented as well as the key question answered, whether monitoring of mass transport within the MDB is possible. Chapter 6 ends with a summary of all results.

2 Basic background knowledge

The first section of this chapter gives a review about the Earth's gravity field with its spatial and temporal variations. Furthermore, methods for the determination of the gravity field are briefly described. Section 2.2 presents three satelllite missions, which map the Earth's gravity field. This chapter ends with the presentation of the measurement techniques of the data types used within this thesis, being the GRACE and TRMM missions, as well as in-situ hydrological observations, namely water level and river flow.

2.1 The Earth's gravity field

The Earth's surface and its interior undergo several dynamic processes. Some of these processes are:

- continental drift,
- seismic,
- volcanos,
- changes in the hydrology, atmosphere,
- ocean circulations,
- etc.

These processes produce continuing mass redistribution both within the Earth and on its surface, being one of the reasons for the non-uniformity of the Earth's density. Consequently, the Earth's gravity field is irregular, e.g. spatially and temporally dependent. Often the geometrical shape of the Earth is described in a first approximation by a geometrical model such as the surface of a sphere or an ellipsoid. An example of this is the reference surface of GPS measurements, the WGS84, which is an ellipsoid of revolution. Thus, the non-uniform gravity field of the Earth requires another reference surface, the so-called geoid, which describes the physical figure of the Earth. The geoid can be imagined as a theoretical surface that coincides to mean sea level (e.g. equipotential surface) and distends at the same level under the continents. More precisely, the geoid is by definition an equipotential surface, being

a surface with constant gravitational potential.
The gravitational potential can be described by an infinite spherical harmonic series as (e.g. Heiskanen/Moritz (1967)):

$$V(r,\theta,\lambda) = \frac{GM}{R}\sum_{l=0}^{L_{max}}\left(\frac{R}{r}\right)^{l+1}\sum_{m=0}^{l}\overline{P}_{lm}(cos\theta)\,(\overline{c}_{lm}cosm\lambda + \overline{s}_{lm}sinm\lambda) \qquad (2.1)$$

where

λ, θ, r	... polar spherical coordinates
GM	... geocentric constant
R	... major semi-axis of a reference ellipsoid
l, m	... degree, order
L_{max}	... maximal spectral resolution
$\overline{P}_{lm}(\cos\theta)$	... 4π-normalized Legendre functions of the first kind
$\overline{c}_{lm}, \overline{s}_{lm}$	... 4π-normalized dimensionless gravity field spherical harmonic coefficients

As seen in equation 2.1 the sum is truncated at a maximum degree L_{max}. Of course a geoid does not conform with a geometrical model, such as the surface of an ellipsoid or a sphere. Deviations with respect to the surface of a reference ellipsoid are described by the geoid undulation. Globally, the geoid undulation can range between ±100 m. According to Wahr et al. (1998), the Earth's gravity field can be described in terms of the shape of the geoid, e.g. geoid undulations. Applying the well known equation of Bruns to equation 2.1 leads to the geoid undulation N also expressed by a spherical harmonic series

$$N(\theta,\lambda) = R\sum_{l=0}^{L_{max}}\sum_{m=0}^{l}\overline{P}_{lm}(cos\theta)\,(\overline{c}_{lm}cosm\lambda + \overline{s}_{lm}sinm\lambda)\,. \qquad (2.2)$$

The spherical harmonic coefficients $\overline{c}_{lm}$, $\overline{s}_{lm}$ are amongst the output data of satellite missions such as CHAMP, GRACE, and GOCE, which map the Earth's gravity field.

Imagine a time-dependent change in the Earth's gravity field causing a geoid change ΔN. The change ΔN can be expressed as the difference from the instanteous geoid undulation from a time average of N. Therefore, the change ΔN is described in terms of changes of the spherical harmonic coefficients Δc_{lm} and Δs_{lm}, respectively, as given by:

$$\Delta N(\theta,\lambda) = R\sum_{l=0}^{L_{max}}\sum_{m=0}^{l}\overline{P}_{lm}(cos\theta)\,(\Delta\overline{c}_{lm}cosm\lambda + \Delta\overline{s}_{lm}sinm\lambda) \qquad (2.3)$$

whereas
$\Delta\overline{c}_{lm} = \overline{c}_{lm} - \overline{c}_{lm}^{mean}$
$\Delta\overline{s}_{lm} = \overline{s}_{lm} - \overline{s}_{lm}^{mean}$
and $\overline{c}_{lm}^{mean}$ and $\overline{s}_{lm}^{mean}$ are the averaged coefficients over a specified time period.

Let $\Delta\rho(r, \theta, \lambda)$ be the density change causing the instantaneous change in the Earth's gravity field. To determine the mass changes caused by the three dimensional density alteration, a two dimensional approximation is applied. According to Wahr et al. (1998), it is assumed that $\Delta\rho$ is accumulated in a thin layer of height H surrounding the Earth surface. The change in surface mass density $\Delta\sigma$ is then defined by the radial integral over the thickness of this layer:

$$\Delta\sigma(\theta, \varphi) = \int_{thin\ layer} \Delta\rho(r, \theta, \lambda)dr \tag{2.4}$$

It can be shown that equation 2.4 can be expressed by means of the spherical harmonic coefficients (e.g. Wahr et al. (1998)):

$$\Delta\sigma(\theta, \lambda) = \frac{R\rho_{ave}}{3} \sum_{l=0}^{L_{max}} \frac{2l+1}{1+k_l} \sum_{m=0}^{l} \overline{P}_{lm}(cos\theta)\,(\Delta\overline{c}_{lm}cosm\lambda + \Delta\overline{s}_{lm}sinm\lambda) \tag{2.5}$$

with
ρ_{ave} ... average density of the Earth ($= 5517\frac{kg}{m^3}$)
k_l ... degree dependent load Love numbers of an elastic Earth model

Finally, the simple relation

$$\Delta\nu(\theta, \lambda) = \frac{\Delta\sigma(\theta, \lambda)}{\rho_w} \tag{2.6}$$

leads to Equivalent Water Thickness (EWT) values. The denominator of equation 2.6 ρ_w indicates the average density of water ($= 1000\frac{kg}{m^3}$). It is common to express surface mass changes in terms of Equivalent Water Thickness values acknowledging that hydrological changes are the most common sources.

2.2 Determining the Earth's gravity field from space

Various gravity measurement methods exist to determine the Earth's gravity field. By making use of three different satellite missions CHAMP, GRACE and GOCE, which apply three different techniques, it is possible to infer the Earth's gravity field in a high resolution on an almost global scale. In the following the three different techniques employed by these missions, namely Satellite-to-Satellite Tracking in high-low mode (SST-hl), Satellite-to-Satellite Tracking in low-low mode (SST-ll), and Satellite Gravity Gradiometry (SGG) are briefly described. Each of these missions delivers the spherical harmonic coefficients, which, according to section 2.1, can be used to describe / model the Earth's gravity field.

2.2.1 Satellite-to-Satellite Tracking in high-low mode (SST-hl)

The satellite mission CHAllenging Minisatellite Payload (CHAMP) (see figure 2.1) is realized through the Satellite-to-Satellite Tracking in high-low mode technique (e.g. Flechtner (2010)). Launched on 15 July 2000, the CHAMP satellite is the first satellite, which maps amongst other parameters the Earth's gravity field on a global scale. The satellite has a GPS receiver on board, which receives code and carrier phase measurements of the higher flying GPS. These measurements allow a precise determination of its orbit. Further satellite-laser ranging measurements to ground stations afford an individual control of the orbit determination. An accelerometer onboard of the CHAMP satellite determines the non-gravitational accelerations such as air drag, solar radiation-, and Earth radiation pressure, so that these accelerations can be removed from the measurements. The knowledge of the current orbit allows the determination of orbit accelerations caused by spatio-temporal changes of the Earth's gravity field. The integration of the measured acceleration leads to the gravitational geopotential.

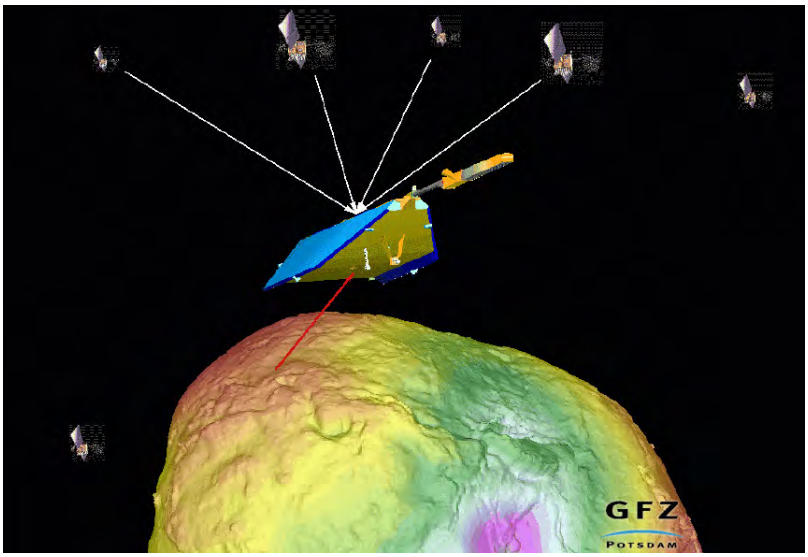


Figure 2.1: The CHAMP satellite [©GFZ Potsdam]

2.2.2 Satellite-to-Satellite Tracking in low-low mode (SST-ll)

The concept of Satellite-to-Satellite Tracking in low-low mode (SST-ll) realized in the Gravity Recovery and Climate Experiment (GRACE) mission, delivers a more accurate determination of the Earth's gravity field (e.g. Flechtner (2010)). The range and range rate, respectively, between two co-orbiting GRACE satellites (see figure 2.2) in a low and near-polar orbit are measured. So instead of absolute disturbing accelerations such as used in the CHAMP mission, differences of disturbing accelerations are used to derivate the Earth's changing gravity field. More details of the GRACE mission are presented in section 2.3.1.

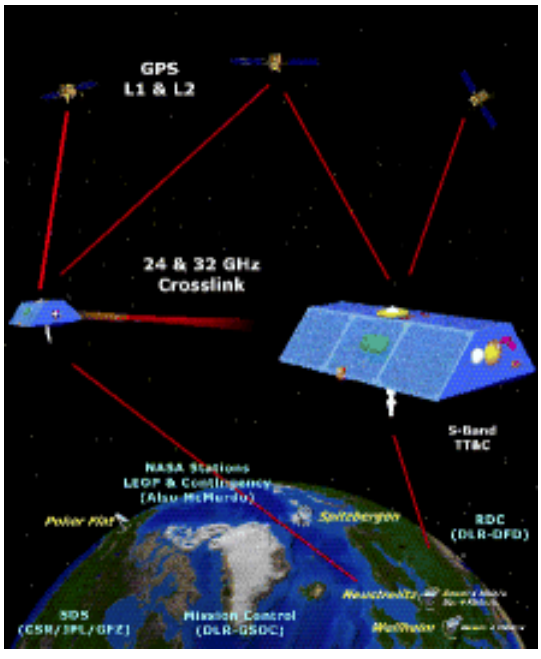


Figure 2.2: The GRACE satellites [©GFZ Potsdam]

2.2.3 Satellite Gravity Gradiometry (SGG)

The Gravity field and steady-state Ocean Circulation Explorer (GOCE) was launched on 17 March, 2009. The gradiometer, the key instrument of that mission, consists of six accelerometers (e.g. ESA (2010)). Due to the very small baselines between the accelerometers, differential disturbing accelerations are measured. By using this method and a lower orbit as compared to GRACE, the short wavelength parts of the Earth's gravity field can be determined. Like the two other missions, the GOCE satellite (see figure 2.3) is equipped with GPS receivers for precise orbit determination.

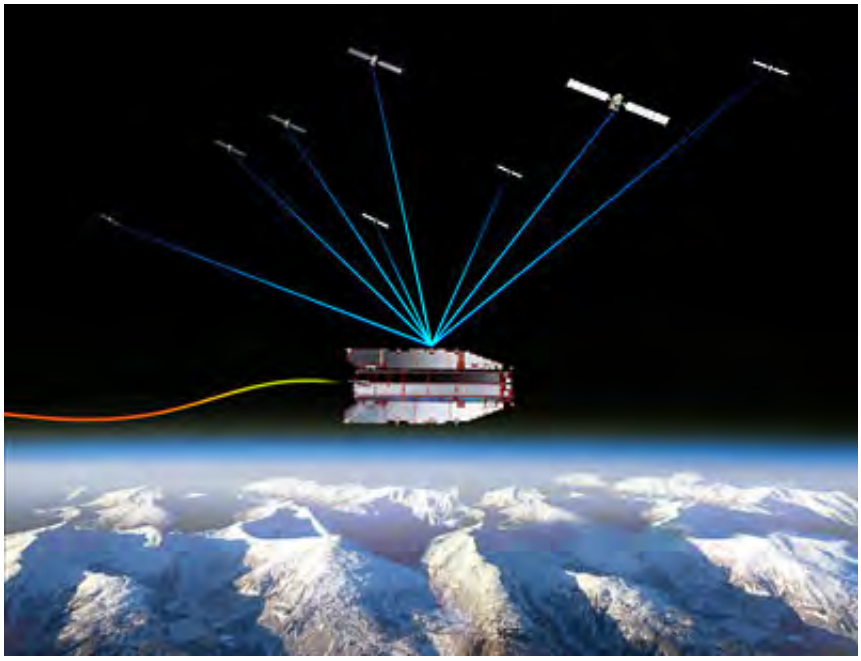


Figure 2.3: The GOCE satellite [©ESA]

2.3 Measurement techniques for the data types used within this thesis

2.3.1 Gravity Recovery and Climate Experiment

The Gravity Recovery and Climate Experiment (GRACE) is a joint mission between the National Aeronautics and Space Administration (NASA) and the Deutsches Zentrum für Luft- und Raumfahrt (DLR). The GRACE project was proposed in 1996 by the University of Texas at Austin, Center for Space Research (UTCSR), the GeoForschungsZentrum Potsdam (GFZ), the Jet Propulsion Laboratories (JPL), Space Systems/Loral (SSL), the Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), and Astrium GmbH (e.g. Flechtner (2010)). The primary goal of the GRACE mission is to map the Earth's spatio-temporal gravity field with a spatial and temporal resolution of approximately 400-600 km and 10-30 days, respectively. Mass movements in the Earth's oceans, atmosphere, and land surfaces cause a time dependent non-uniform Earth's gravity field.
GRACE delivers important information about the distribution and mass movements on and close to the Earth's surface. The following mass changes can be inferred from the data provided by GRACE (e.g. CSR (2010)):

- oceanic changes caused by surface and deep currents,
- land hydrological changes: runoff and ground water storage on land masses,
- cryospheric changes: substitutions between ice sheets or glaciers and the oceans,
- solid Earth changes: variations of mass within the Earth, and
- atmospheric changes: accumulation of a better atmosphere profile

The accuracy of these measurements is orders higher than previous measurements from ground-based observations or remote-sensing spacecrafts.
On March 17, 2002 the GRACE satellites were launched from Plesetsk Cosmodrome, Russia on a Rockot launch vehicle. The two identical co-planar spacecrafts are situated in a low and near-polar orbit initally at approximately 500 km altitude. The crafts move along their near circular orbit with an along-track separation of about 220 km. In figure 2.4 the groundtracks of two revolutions of a GRACE satellite is displayed.

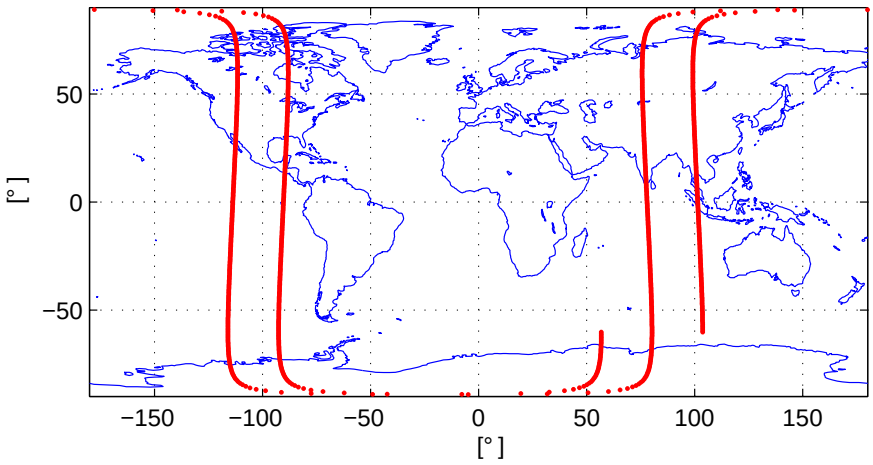


Figure 2.4: Groundtracks of a GRACE satellite for two consecutive revolutions

Changes in the spatio-temporal dependent gravity field of the Earth influence the trajectories of the two satellites. An alteration in the particular trajectory of the satellites causes a change of the satellite-to-satellite range. This change is measured by the so called K-Band ranging system (see section 2.3.1.1). The relative range and range-rate is proportional to the integrated differences of the gravity accelerations, which are apprehended from both satellites at their differing locations. In order to cover the whole Earth, the inclination of the orbit was chosen to be 89.5° (near polar).

2.3.1.1 The satellite payload

Each of the twin-satellites (see figure 2.5) has the following instruments on board (e.g. Flechtner (2010)):

- K-band Ranging System (KBR),
- Accelerometer (ACC),
- GPS Space Receiver (GPS),
- Laser Retro-Reflector (LRR),
- Star Camera Assembly (SCA),
- Coarse Earth and Sun Sensor (CES),
- Ultra Stable Oscillator (USO), and
- Centre of Mass Trim Assembly (CMT)

K-band Ranging System (KBR):
The main instrument of GRACE is the K-band ranging system (KBR). The KBR measures the one-way range between the satellites, with an accuracy of about 1 micrometer per second. The only difference between the KBRs on each of the two satellites, is that they are shifted by 500 KHz, so that no cross-talking between the received and transmitted signals can appear. The dual microwave signals at 24 GHz and 32 GHz, called K- and Ka-Band, respectively, are transmitted and received from a single horn antenna. No ionospheric corrections have to be applied to the measurements because carrier phase signals on two frequencies are used. An ultra-stable oscillator provides the frequency reference. Up- and down-converting between the frequency levels happens through a microwave assembly. Finally, the K-band carrier phase signals are then sampled and digital signal processed by an instrument processing unit (IPU).

Accelerometer (ACC):
All non-gravitational accelerations on the GRACE satellite such as air drag, solar radiation-, and Earth radiation pressure have to be removed from the measurements. This is performed by the SuperSTAR accelerometer. The accelerometer is based on a proof-mass located in a cage, the current position of which is measured by capacitive sensors. Whenever the proof-mass is in motion, electrostatic forces provided by the capacitive sensors constrain the proof-mass to stay statically in the centre of the cage. Due to the known proof-mass and the applied forces on the proof-mass, the three dimensional acceleration vector can be inferred at the desired point in time. To achieve a high accuracy, the proof-mass must be located precisely at the centre of gravity of the GRACE satellite. Any offsets between the centre of gravity and the proof-mass, are measured and rectified by a Centre of Mass Trim Assembly (CMT). The resolution of the ACC amounts $10^{-10} m/s^2$.

GPS Space Receiver (GPS):
The GPS TurboRogue Space Receiver assembly has the followings tasks:

- Precise Orbit Determination (POD) (cm-accuracy),
- coarse Positioning for real-time use,
- time Tagging of all data, and
- atmospheric and ionospheric profiling.

Satellite-to-satellite tracking between the GRACE satellites and the GPS satellites is applied. Two omnidirectional POD-antennas serve for a centimetre accurate

orbit determination. Another antenna, namely an aft-pointing helix antenna, is responsible for providing both atmospheric and ionospheric profiling.

Laser Retro-Reflector (LRR):
The Laser Retro-Reflector (LRR) is composed of four prisms, which reflect short laser pulses transmitted by Laser ground stations. The deduced range from the laser measurements is used for an independent POD control. The LRR also serves for a calibration of the GPS Space Receivers, which are on board of each GRACE satellite, and for further technological experiments.

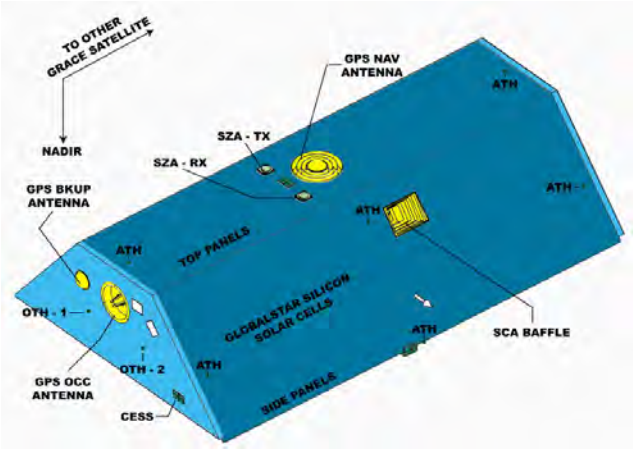
Star Camera Assembly (SCA):
For a correct interpretation and accurate orientation of the satellite relative to the position of the stars, the Star Camera Assembly (SCA) in a celestial reference frame is utilised. Two simultaneous DTU star cameras measure the attitude with an accuracy of 0.3 mrad or better.

Coarse Earth and Sun Sensor (CES):
Consisting of six sensor heads, the Coarse Earth and Sun Sensor (CES) provides the course state of the sun and the Earth, which is used for initial acquisition.

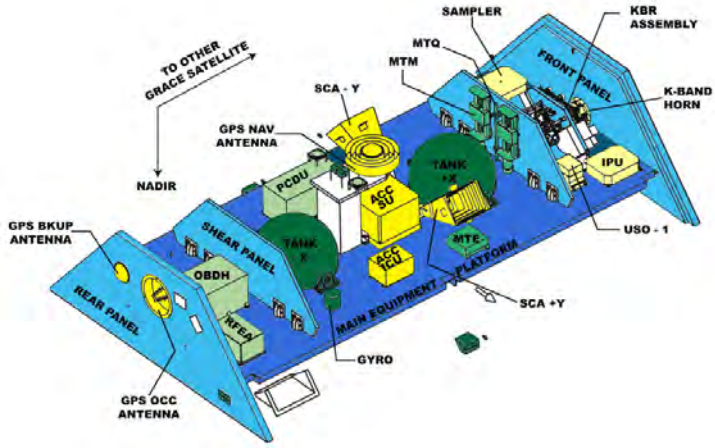
Ultra-Stable Oscillator (USO):
The Ultra Stable Oscillator (USO) provides the frequency generation for the K-band ranging system.

Centre of Mass Trim Assembly (CMT):
As described above, the CMT precisely measures the offset between the satellite's centre of mass and the centre of the proof-mass in the direction of three independent axes. If it is needed, the centre of mass is adjusted.

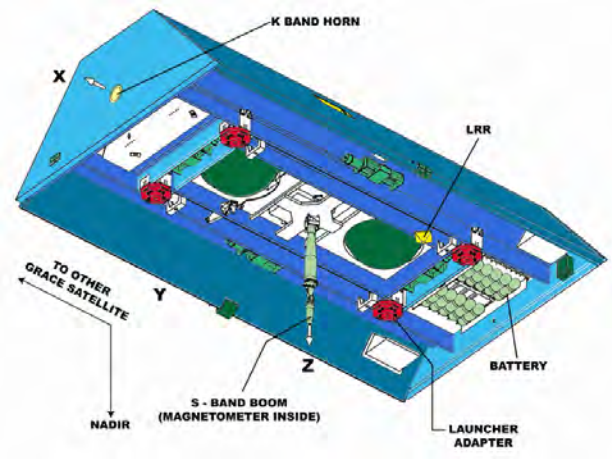
Figure 2.5 shows a GRACE satellite with its instruments onboard.



(a) top view



(b) internal view



(c) bottom view

Figure 2.5: GRACE satellite with its instruments [©GFZ Potsdam]

2.3.1.2 GRACE products

The team of Science Data System (SDS) consisting of the University of Texas at Austin, Center for Space Research (UTCSR), GeoforschungsZentrum Potsdam (GFZ), and the Jet Propulsion Laboratories (JPL) provides and processes the products from GRACE.

There are three levels of products available (e.g. Flechtner (2010)):

- Level-0: The Raw Data Center (RDC) of the Mission Operation System (MOS) located in Neustrelitz/Germany receives, collects and decommutates the raw data.
- Level-1: These pre-processed, time-tagged and normal pointed data include the K-band ranging, accelerometer, star camera, and GPS data of both satellites. Furthermore, the preliminary orbits of both GRACE satellites are generated. The level-1 products are available at JPL's Physical Oceanography Distributed Active Data Center (PODAAC) and at GFZ's Integrated System Data Center (ISDC).
- Level-2: These data contain the short-term (30 days) and mean gravity field, which are derived from calibrated and validated GRACE level-1 data. Data sets about temperature, and pressure fields, ocean bottom pressure and hydrological data are also provided in order to eliminate time variabilities in gravity field solutions. Moreover, the precise orbits of both GRACE satellites are generated. All level-2 products should be available at PODAAC and ISDC 60 days after data taking.

In each processing an a priori best-known gravity potential model is updated with the estimates by means of a least-square approach. The a priori gravity potential model is part of a background model, which contains also mathematical models, the proper parameters, and a priori gravity models for ocean- and polar tides, etc. This, together with numerical techniques, the background model allows a prediction of the satellite-to-satellite ranges and range rates, respectively. The differences between the predicted and measured values lead to the residuals. The residuals are then minimized with a least-square approach and the "old" background model is updated, which produces the new gravity information derived from GRACE.
More details of the GRACE mission and its products are found in Tapley et al. (2004) and Schmidt et al. (2008).

2.3.1.3 Details of the data in use

For this thesis level-2 data processed at the GFZ is used. The data set contains the spherical harmonic coefficients up to degree and order 120 for a global gravity field solution in a monthly resolution. These coefficients can be interpreted as the change in mass distribution at a particular time. In this thesis, the spherical harmonic coefficients from May 2003 until April 2009 (6 years) are used. Unfortunately, two months (June 2003 and January 2004) are completely missing in the data set.

2.3.2 The Tropical Rainfall Measuring Mission (TRMM)

On November 27 1997, the satellite of the Tropical Rainfall Measuring Mission (TRMM) was launched by means of a H-II rocket from the Tanegashima Space Center in Tanegashima, Japan (e.g. Wikipedia (2010b), Jaxa (2010)). The aim of this joint venture between NASA, the Japan Aerospace Exploration Agency (JAXA), and the Communciation Research Laboratory (CRL), is to detect and study tropical rainfall both over the land surface as well as over oceans. Forming 2/3 of the Earth's total rainfall, the tropical and subtropical precipitation plays an important role for climate processes on the whole globe. TRMM makes a contribution to understand the climate system better and also to predict floods and other catastrophes having to do with water and precipitation. Together with in-situ on ground observations, the TRMM mission allocates the satellite observations on a grid with a 0.25 degree in a latitude band between 50° S and 50° N. The orbit altitude was first approximately 350 km with an inclination of 35 degrees. One revolution takes 90 minutes, i.e. 16 orbits per day. The lower the orbit altitude, the better is the spatial resolution. In 2001, the altitude was changed to 402.5 km.

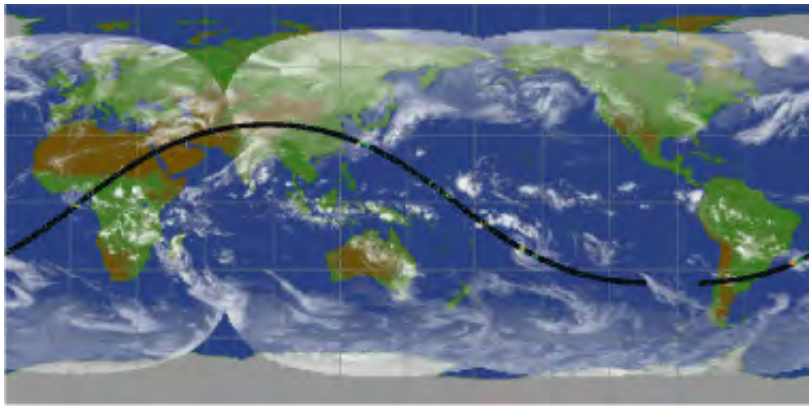


Figure 2.6: Groundtrack of one revolution of the TRMM satellite [©JAXA]

Five instruments aboard provide the data, namely the Precipitation Radar (PR), the TRMM Microwave Imager (TMI), the Visible and Infrared Scanner (VIRS), the Clouds and the Earth's Radiant Energy System (CERES), and the Lightning Imaging Sensor (LIS).

Precipitation Radar (PR):
The PR is one of the main instruments of the TRMM mission and provides three dimensional profiles of rainfall up to a height of approximately 20 kilometres. These profiles contain information about the intensity and distribution of the rain, the rain type, the storm depth, and the height melting the snow into rain. This is done by emitting microwaves from the PR, which return back after transmitted from the rain. Furthermore, the PR detects the rain by means of a close electronic scan, which uses phased array antenna technique. Rain rates can be detected up to 0.7 mm/hour. On the ground the horizontal resolution is about five kilometres along track and the swath width accounts for 220 km at the equator.

TRMM Microwave Imager (TMI):
The PR and the TMI form the key instruments of the TRMM missions. The TMI is a multi-channel / dual-polarized microwave radiometer measuring the radiation energy emitted from the Earth. The TMI delivering a swath width of 760 km has the ability to detect water vapor, cloud water and rainfall intensity in the atmosphere.

Visible and Infrared Scanner (VIRS):
As the name of this instrument suggests the VIRS cross-tracking radiometer provides measurements of the radiance from visible to infrared light. Separated in visible and infrared areas precipitation can be better detected. The swath width amounts to 720 km at the equator.

Clouds and the Earth's Radiant Energy System (CERES):
The fourth instrument, CERES, is an optical staring telescope and filtering imaging system, which detects the atmospheric radiation energy and the atmosphere at the Earth's surface. Besides the radiation energy CERES also provides some cloud characteristics such as cloud-amount, altitude-, and thickness.

Lightning Imaging Sensor (LIS):
This sensor measures lightning. A point on the Earth is monitored for 80 seconds. This time is enough to calculate the flashing rate whereby it is known if a storm is increasing or decreasing.

TRMM provides a number of products in different levels. In this thesis a monthly precipitation product processed in the so-called algorithm 3B-43 is used. The satellite data is combined with on-ground observations. The precipitation data is provided in a monthly resolution and in a 0.25° x 0.25° spatial resolution ranging between 50 degrees south and 50 degrees north latitude.

2.3.3 In-situ hydrological observations

2.3.3.1 Water level

The height of the water surface at a location along a stream or river is called water level, stream stage or gauge height. There are several techniques, which exist that can be used to measure water level (e.g USGS (2010)). When carrying out gauge height meausurements, regard should be taken to the height datum. Two approaches by which this can be undertaken are described below.

- **Stilling well**
  A stilling well is installed on the foreshore of a river (see figure 2.7). Through underwater pipes, water from the river is transmitted into the stilling well. According to the principle of communication vessels, the water level in the stilling well is at the same height like as that of the river. An acoustic, float or optic sensor measures then the height (water level) in the well.

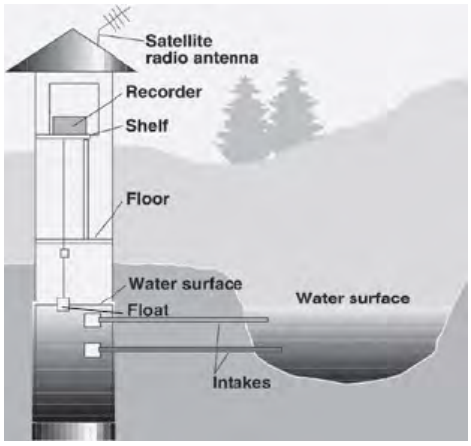


Figure 2.7: Stilling well [©USGS]

- **Pressure measurements**
  A tube with a small flow of gas may be placed at the surface of the river bed. The water applies pressure to the tube, which causes bubbles to form at a specific place under water. This pressure is in accordance with the hydrostatic pressure. The measured pressure is proportional to the water height at the location where the bubbles reach the surface.

2.3.3.2 River flow

River flow or discharge describes the mass of water flowing through a river or stream in a certain time unit, which is calculated by multiplying the area of water in a channel cross section with its average velocity (e.g USGS (2010)).

$$River flow = Cross\ section \cdot velocity \tag{2.7}$$

There are several methods, which can be used to measure the velocity and the cross section. Two such techniques are described below.

- **Current-meter method**
  The cross section is divided into several vertical sections (see figure 2.8). In each vertical section the width and depth is measured, which makes it possible to determine the vertical section area. The widths can be gauged by use of a cable or steel tape, and the depth may be measured by an implement such as a wading rod. Moreover, an average velocity is measured by a so-called current meter. The principle is based on counting the number of rotations around a vertical axis of a wheel consisting of metal cups. This is performed by transmitting an electronic signal with each rotation, which is directly linked to the water velocity. For every subsection the river flow is determined by equation 2.7. The total river flow is determined by summing all discharges from the vertical section areas.

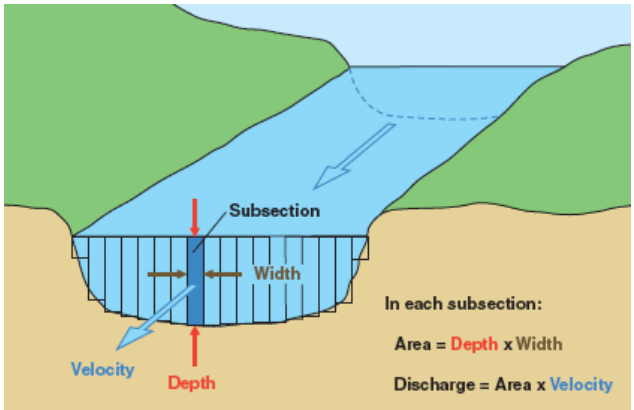


Figure 2.8: Current-meter method [©USGS]

- **Acoustic Doppler Current Profiler**

  The Acoustic Doppler Current Profiler makes use of the Doppler Effect to determine water velocity. A sound pulse is sent into the water, which is reflected on the river bed. The change of the frequency is measured, which allows for the determination of the water velocity. Furthermore, acoustic signals are sent to the bottom of the river to determine the river depth. This is done by measuring the time between the emitted and received signal. To determine the discharge, the Acoustic Doppler Currrent Profiler is fixed on a boat. Measurements are done continuously on the boat, which is guided through the river. Therewith a profile through the river is obtained. With the different depths, widths and the water velocities, the river flow can be determined according to equation 2.7.

3 Methodology

To carry out a meaningful analysis of the data, some theoretical basics are necessary. Firstly, the technique of Multiple Linear Regression Analysis is introduced in section 3.1. This technique is used for the search of a linear model fitting "best" to a time series. In section 3.2 the Principal Component Analysis is described in more detail. The goal of the Principal Component Analysis is to find another representation of the original data in a reduced Euclidean space in order to eliminate noise as much as possible and to keep only the signal in the best case. This chapter concludes with an explanation of cross correlation analysis, which provides the correlation and phase shift between two signals.

3.1 Multiple Linear Regression Analysis (MLRA)

3.1.1 Fitting a model to data

The goal of a multiple linear regression model is to find a relationship between one dependent and one or more independent variables with the following functional model (e.g Brook/Arnold (1985), Johnson/Wichern (2007)):

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_u x_{ui} + \varepsilon_i \tag{3.1}$$

In equation 3.1 $Y_i \quad (\forall i = 1, .., n)$ denotes the n dependent variables, namely n observations. The vector of all observations $\underline{Y}$ is also called the response. $\beta_j \quad (\forall j = 1, ..., u)$ stands for the predictor variables or regressor parameters, e.g. u unknown parameters. The vector $\underline{\varepsilon}$ describes the error, which arises from measurement errors and not modelled effects. The error $\underline{\varepsilon}$ is a stochastic variable.
The following characteristics are assumed for the error:

1. $E(\underline{\varepsilon}) = 0$
2. $Var(\varepsilon_i) = \sigma^2$
3. $Cov(\varepsilon_j, \varepsilon_k) = 0$

where E stands for the expectation value, Var for the variance, and Cov for the covariance.
This means the ε_i are normally distributed.

Equation 3.1 can also be expressed in vector-matrix notation:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1u} \\ 1 & x_{21} & x_{22} & \dots & x_{2u} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nu} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_u \end{pmatrix} + \begin{pmatrix} \varepsilon_0 \\ \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \tag{3.2}$$

$$\Leftrightarrow$$

$$\underline{Y} = \underline{A}\,\underline{x} + \underline{\varepsilon}$$

With the assumption of no existing correlation between the n observations Y_i, which all have the same variance, the well-known adjustment of observation equations using least-square is applied to determine the unknown parameters β_j, e.g. the sum of square residuals is minimized.

$$\sum_{i=1}^{n} \varepsilon_i^2 = \underline{\varepsilon}^T \underline{\varepsilon} = (\underline{Y} - \underline{A}\,\underline{x})^T (\underline{Y} - \underline{A}\,\underline{x}) = min \tag{3.3}$$

The adjusted regressors are then obtained by:

$$\underline{\widehat{x}} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \vdots \\ \widehat{\beta}_u \end{pmatrix} = (\underline{A}^T \underline{A})^{-1} \underline{A}^T \underline{Y} \tag{3.4}$$

The standard deviations $\sigma_{\widehat{\beta}_j}$ of the adjusted parameters $\widehat{\beta}_j$ are obtained from the covariance matrix

$$Q_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0 (\underline{A}^T \underline{A})^{-1} = \frac{\underline{v}^T \underline{v}}{n-u} (\underline{A}^T \underline{A})^{-1} \tag{3.5}$$

with

$$\underline{v} = \underline{Y} - \underline{\widehat{Y}} = \underline{Y} - \underline{A}\,\underline{\widehat{x}} \ . \tag{3.6}$$

The standard deviations are then defined as the square roots of the diagonal elements of the covariance matrix given by:

$$\sigma_{\widehat{\beta}_j} = \sqrt{Q_{\widehat{x}\widehat{x}}} \tag{3.7}$$

To study the statistical relationship between the adjusted regression parameters the linear correlation coefficient can be calculated based on the values of the covariance matrix. It is defined by

$$r_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \tag{3.8}$$

with

$$r_{ij} \in [-1, 1]\,. \tag{3.9}$$

The correlation coefficient can reach a value between -1 and 1. A value of zero denotes that there is no relationship between the parameters $\widehat{\beta}_i$ and $\widehat{\beta}_j$, whereas a value of 1 indicates a perfect increasing linear relation (e.g. correlation) and a value of -1 a perfect decreasing linear relation (e.g. anti-correlation).

3.1.2 'Goodness of fit' of the chosen model

To find out how many regression parameters should be chosen for the linear regression model, a significance test for each single adjusted parameter must be applied. The corresponding null hypothesis is given by:

$$H_0 : \ \widehat{\beta}_j = 0 \quad \forall j = 0, ..., u \tag{3.10}$$

The particular null hypothesis H_0 for each of the regression parameters (equation 3.10) means that the examined regression parameter does not affect the response significantly. If the null hypothesis is true for a certain $\widehat{\beta}_j$, this parameter should be eliminated from the original regression model. The test statistic

$$t_j = \frac{\widehat{\beta}_j}{\sigma_{\widehat{\beta}_j}} \sim t_{n-u} \tag{3.11}$$

follows a Student's t distribution with the quantile t_{n-u}. The equations above only show the significance of a single parameter. A further test can be applied to test whether the complete model fits the data. To determine whether the model fits the data, the coefficient of determination R^2 is introduced. The coefficient describes the proportion of the variation, which appears by the chosen statistical model. R^2 can also be seen as a measure about the 'goodness of fit', i.e. how well the output can be predicted. The coefficient of determination R^2 is defined by

$$R^2 = \frac{SSR}{SST} \tag{3.12}$$

where SSR and SST denotes the regression sum of squares and total sum of squares, respectively, given by:

$$SSR = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 \tag{3.13}$$

$$SST = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 \tag{3.14}$$

with
$\overline{Y}$ denoting the mean value of all Y_i.

The total sum of squares SST can also be expressed by the sum of SSR and SSE given by:

$$SST = SSR + SSE \tag{3.15}$$

whereas

$$SSE = \sum_{i=1}^{n} \varepsilon_i^2 \tag{3.16}$$

SSE indicates the residual sum of squares. The coefficient of determination can reach values between 0 and 1. A value of 0.5 indicates that half of the variation cannot be described by the chosen model. The larger the value for R^2, the better the model fits to the data. A best fit would indicate a value of 1. However, the coefficient of determination also has a disadvantage. According to Brook/Arnold (1985) the size of R^2 depends on the sample size and the number of parameters to be estimated. For this reason an adjusted value for R^2 is introduced, which disposes of this disadvantage. This adjusted coefficient of determination is defined by

$$\tilde{R}^2 = \left(R^2 - \frac{u}{n-1} \right) \frac{n-1}{n-u-1} \,. \tag{3.17}$$

The adjusted coefficient of determination $\tilde{R}^2$ has also a small disadvantage. If the value of R^2 is close to zero, the adjusted coefficient of determination can even become negative, which cannot be interpreted.
To calculate the significance of the model with its regressors, a statistical Fisher-test for R^2 or $\tilde{R}^2$ can be applied. With this test the ratio of the sum of squares and of the residual sum of the square is determined.

The corresponding null hypothesis is given by

$$H_0: \ R^2 = 0 \qquad or \qquad H_0: \ \tilde{R}^2 = 0 \tag{3.18}$$

with the test variable

$$F = \frac{SSR}{SSE}\frac{n-u-1}{u} = \frac{R^2}{1-R^2}\frac{n-u-1}{u} \sim F_{u,n-u-1} \tag{3.19}$$

and the quantile $F_{u,n-u-1}$.
If the test variable F is smaller than the quantile $F_{u,n-u-1}$, then the null hypothesis is true. This means R^2 or $\tilde{R}^2$ is not significant different from zero. Consequently, the chosen model with its regressors is not adequate in this case . A different model should be chosen. Conversely, if the test variable F is bigger than the quantile $F_{u,n-u-1}$, then the model is feasible. However, the statistical Fisher test doesn't testify how well the different regressors fit to the model. It is possible that one regressor leads to a large sum of squares, while other parameters even could be neglected. Consequently, the t-test for every single parameter is also applied, which is already described at the beginning of this section.

Depending on the chosen model it does not always make sense to apply a t-test on every single variable. For example if the chosen model is of the form

$$Y = c_0 + c_1 t + a_1 cos(\omega t) + b_1 sin(\omega t) \tag{3.20}$$

with $\omega = \frac{2\pi}{T}$ and $T = 12$ months
the parameters a_1 and b_1 should not be separated. The variables a_1 and b_1 combined form the parameters of the annual part of the model. Both parameters a_1 and b_1 can be tested in a significance test (t-test see above) separately, however it makes no sense. It is better to test phase Φ and amplitude A, which are functions of a_1 and b_1. The standard deviations of the phase and the amplitude can be calculated by means of error propagation. Possible correlations between the parameters a_1 and b_1 are neglected here. Phase Φ and amplitude A with its corresponding standard deviations σ_Φ and σ_A, respectively, are given by equations 3.21 to 3.24:

phase

$$\Phi = arctan\left(-\frac{b_1}{a_1}\right) \tag{3.21}$$

$$\sigma_\Phi = \sqrt{\left(\frac{b_1}{a_1^2+b_1^2}\right)^2 \sigma_{a_1}^2 + \left(\frac{-a_1}{a_1^2+b_1^2}\right)^2 \sigma_{b_1}^2} \tag{3.22}$$

amplitude

$$A = \sqrt{a_1^2+b_1^2} \tag{3.23}$$

$$\sigma_A = \sqrt{\left(\frac{a_1}{\sqrt{a_1^2+b_1^2}}\right)^2 \sigma_{a_1}^2 + \left(\frac{b_1}{\sqrt{a_1^2+b_1^2}}\right)^2 \sigma_{b_1}^2} \tag{3.24}$$

Furthermore, the t-tests for every single parameter are not independent. Possible correlations between the parameters are not considered. A possible approach to consider correlations is the technique of reduced models. The method of reduced models is based on the comparison of the sum of squares of the residuals from the full model, which is a suitable parameterization of the data, with the sum of squares of a reduced model. A reduced model contains less regressors. If in the example above, equation 3.20 described the full model, a reduced model could be

$$Y_{reduced} = c_0 + c_1 t \ . \tag{3.25}$$

In this case the influence of the annual part of the model is investigated. This means a hypothesis test can be applied to a set of parameters. In the example above the null hypothesis would be $H_0 : a_1 = b_1 = 0$.
The test variable for the F-test is then defined by:

$$F = \frac{SSE_{reduced} - SSE_{full}}{SSE_{full}} \frac{n-u}{r} \sim F_{r,n-u} \tag{3.26}$$

The quantile is defined by $F_{r,n-u}$. The factor r indicates the number of omitted parameters, e.g. in the example r would be 2. If the null hypothesis is true, the reduced model is suitable for the representation of the data and the full model evocates a not significant improvement.

3.2 Principal Component Analysis (PCA)

3.2.1 Motivation of Principal Component Analysis

To illustrate the aim of a Principal Component Analysis (PCA) a simple two dimensional example is given (Krzanowski (2007)). Imagine the heights and weights from a number of people are measured. After mean-centring the residuals of heights (x_1-axis) are plotted against the residuals of weights (x_2-axis) (see figure 3.1).

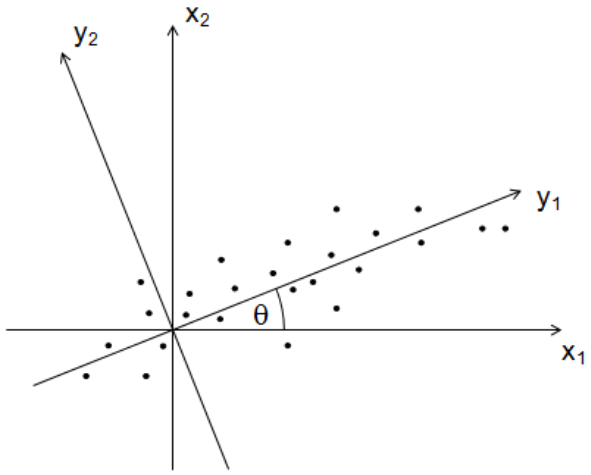


Figure 3.1: A two-dimensional example

The coordinate system is now rotated around the angle θ so that the new axis y_1 is located along the largest variances. The second axis y_2 is of course chosen to be orthogonal to the y_1-axis. The data set is now related to the new coordinate system $(y_1 - y_2)$ by a simple coordinate transformation.

$$y_1 = x_1 \cdot cos\theta + x_2 \cdot sin\theta \tag{3.27}$$

$$y_2 = -x_1 \cdot sin\theta + x_2 \cdot cos\theta \tag{3.28}$$

By projecting the data set along the y_1-axis a simpler representation in only one dimension is achieved, which is a suitable approximation. Hence, a reduction of one dimension by this simple representation of the data along the y_1-axis is gained. In most cases the data set is much more complex, i.e. more dimensions, and in the first instance, the structures of the signal are not obvious and may be hidden. Therefore, the goal of the PCA is to reduce a data set by a coordinate transformation to a few new basis vectors in such a way that the signal is revealed, while the noise is filtered out. PCA is based on the assumption that large variances belong to the signal, which represents dominant structures and in contrast small variances illustrate noise. In the following subsection a brief description of PCA for arbitrary

complex data is given.

3.2.2 A brief description of Principal Component Analysis

A n x p data matrix Z, whose elements are already mean-centred, is formed. The values are given for locations $x = 1, ..., p$ and for the time $t = 1, ..., n$. The matrix Z with the dimension n x p has the following form.

$$Z = \begin{bmatrix} z(1,1) & z(1,2) & \dots & z(1,p) \\ z(2,1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ z(n,1) & \dots & \dots & z(n,p) \end{bmatrix} \tag{3.29}$$

with

$z(t,x) = \tilde{z}(t,x) - \overline{z}$ ... mean centred value

$\tilde{z}(t,x)$ ... original value

$\overline{z}$ ... mean value

Each row in Z stands for a specific time t_i for all locations $x = 1, ..., p$, while each column represents one location x_j for all dates $t = 1, ..., n$.
In the next step a so-called scatter matrix S is formed. This matrix S is defined by:

$$S = Z^T \cdot Z \tag{3.30}$$

The scatter matrix S is related to the covarince matrix Σ by:

$$\Sigma = \frac{1}{n-1} S \tag{3.31}$$

This scatter matrix S is now investigated to find the directions (compare with y_1 in the two-dimensional introductive example) along the variances of the scatter matrix appear with a maximum. This demand leads to the so-called eigen-system equation:

$$SE = E\Lambda \tag{3.32}$$

with

Λ ... diagonal matrix containing all eigenvalues, and

E ... containing all corresponding eigenvectors

The eigenvectors, called empirical orthogonal functions (EOFs), are the new basis vectors in a p dimensional Euclidean vector space, which are uncorrelated and orthogonal among each other. The data set is now transformed to the new basis

vectors (EOFs). Their corresponding coordinates a_j are called principal components (PCs) (analysis step), which are given in matrix notation by

$$A = ZE \tag{3.33}$$

and in scalar notation by

$$a_j(t) = \sum_{x=1}^{p} z(t,x)e_j(x) \qquad t = 1, ..., n \ . \tag{3.34}$$

Sometimes the EOFs are called amplitudes and the PCs loadings. The i-th PC and i-th EOF form together the i-th mode of the PCA. Conversely, the original matrix Z is recovered by multiplying equation 3.33 with E^T from the right hand side (synthesis step).

$$Z = AE^T = \sum_{j=1}^{p} a_j e_j^T \tag{3.35}$$

In the following two important properties of the principal components are represented.

- **Property 1**:
  The principal component series $a_j(t)$ $(t = 1, ..., n)$ are pair wise uncorrelated. This can be shown by

$$A^T A = (ZE)^T ZE = E^T(Z^T Z)E = E^T(SE) = E^T(E\Lambda) = \Lambda \, . \tag{3.36}$$

- **Property 2**:
  The total scatter of the data set is given by the sum of all eigenvalues λ_i (e.g. Preisendorfer (1988)).

$$\sum_{x=1}^{p}\sum_{t=1}^{n} z^2(t,x) = \sum_{i=1}^{p}\sum_{t=1}^{n} a_i^2(t) = \sum_{i=1}^{p} \lambda_i \tag{3.37}$$

As the magnitudes of the eigenvalues λ_i are measures of the relative importance (see equation 3.39) of the corresponding PCA mode, they are ordered in the matrix Λ in decreasing order ($\lambda_1 > \lambda_2 > ... > \lambda_p$).
According to Krzanowski (2007) the p-dimensional geometrical model formed from the sample can be considered to be the 'true' picture of the data. To receive a more clearly arranged representation of the complex data the data points are projected into r subspaces ($r << p$), which are defined by the r principal components and empirical orthogonal functions, respectively. As a result of this, the matrix Z

can be approximated by the synthesis formula

$$Z_{PCA} = \sum_{j=1}^{r} a_j e_j^T \approx Z \tag{3.38}$$

with $r << p$.

It is common to calculate the proportion of the cumulative sum of the first r eigenvalues of the total variance, which are taken into account.

$$P_r = 100 \cdot \frac{\sum_{i=1}^{r} \lambda_i}{trace(\Lambda)} \tag{3.39}$$

For a suitable representation of the data the quotient P_r should at least account for 75-80% of the total variance. One graphical approach is to plot the magnitude of all eigenvalues in a so-called scree diagram. An example of a scree diagram is illustrated in figure 3.2.

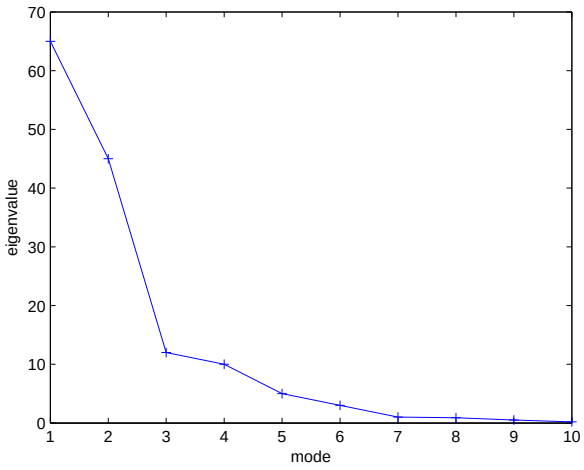


Figure 3.2: Scree diagram

A sharp decline can be recognized in the first three modes, while in the further modes the gradient is much smaller. In this case probably the first three modes containing the signal would be chosen and all other modes would account for noise. It is common to choose the 'elbow' of the curve as a breakup between signal and noise. If in a scree diagram no significant change in the gradient appears, it is an indication that the data has a larger variation from the mean, or includes more noise, respectively.
Further details on the PCA can be found in e.g. Preisendorfer (1988) and Jolliffe (2002).

3.3 Cross correlation

Cross correlation is a statistical tool to analyse if two signals are correlated and to calculate the phase difference between the two signals.
The cross correlation function between two signals is defined by (e.g. TU Freiberg (2010)):

$$K(\tau) = \int_{-\infty}^{+\infty} x(t)y(t+\tau)dt \tag{3.40}$$

If, for example, two identical signals are time shifted by -5τ , equation 3.40 would have a maximum value at -5τ (see figure 3.3).

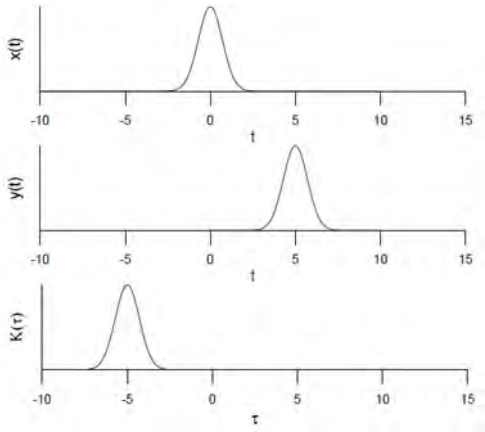


Figure 3.3: Cross correlation function between two signals x(t) and y(t)

However, in practice there are only discrete signals. Therefore, a transformation of the integral into a sum must be applied to equation 3.40. For two discrete signals given at dates $t_0+\Delta t,\ t_0+2\Delta t,\ ...,\ t_0+N\Delta t$ the cross correlation function becomes:

$$corr(k) = \frac{\sum_{i=1}^{N} x[i] \cdot y[i+k]}{\sqrt{\sum_{i=1}^{N}(x[i])^2 \cdot \sum_{i=1}^{N}(y[i+k])^2}} \tag{3.41}$$

where

$\sqrt{\sum_{i=1}^{N}(x[i])^2 \cdot \sum_{i=1}^{N}(y[i+k])^2}$ ... normalisation term

A value of 1 denotes that the two discrete signals are equal at a special time, whereas a value of -1 indicates that the two signals are in phase opposition. No correlation between the two signals accounts for a value of 0. A value of exactly 1 or -1 very rarely appears. Therefore, the maximum value of the cross correlation function is determined, in order to calculate the relation between two signals as well as the phase difference between them.

4 Results

After the representation of the methodology, the investigation area, the Murray-Darling Basin, is briefly described and the three different data types are investigated by means of MLRA and PCA.

4.1 Geographical setting - the Murray-Darling Basin (MDB)

The MDB (e.g. Prasad/Khan (2002)) is located in the south-east of Australia, which spans part of the states of New South Wales (57%), Queensland(25%), Victoria (12%), South Australia (6%), and the Australian Capital Territory (less than 1%) (see figure 4.1). With a length of approximately 3400 kilometres in a north eastern direction, and covering approximately 14% of Australia's total area, the MDB is the world's fifteenth biggest basin in terms of length and the twenty first largest in terms of area. This corresponds to an area of 1,061,469 km^2. The MDB consists of 23 river valleys and several wetlands. The three longest rivers in Australia, being the Darling River (2,740 km), the Murray River (2,530 km) and the Murrumbidgee River (1,690 km), all belong to the MDB. The name of the MDB is derived from the two longest rivers, which form part of it.

Figure 4.1: The Murray-Darling Basin

Most of the MDB is low lying (below 200 metres above sea level). Therefore, many rivers are due to low gradients very slow flowing. Around 40% of all Australia's farms are located in the MDB. These farms produce wool, cotton, wheat, sheep, cattle, dairy produce, rice, oil-seed, wine, fruit and vegetables. Forming more than one third of Australia's gross value of agricultural production, the MDB is the most important agricultural area in Australia. Many farmers use irrigation causing an augment of the salinity. They extract 70% of all water used in Australia. The irrigation system, introduced at the end of the nineteenth century, has caused significant problems for the MDB in terms of the proper use of its water resources. The MDB plays an important role in maintaining biodiversity, with at least 35 species of birds, 16 species of mammals and more than 35 fish species living in the area. For the fish species the MDB is a huge interconnecting river-network.

4.1.1 Climate, rainfall, run-off and water storage in the MDB

In comparison to other areas of Australia, the MDB has mostly a dry climate. Due to the size of the MDB, there are different climate conditions throughout, ranging from the cool and humid eastern uplands with its rainforests, the temperate south-east, the subtropical north-east, to the hot, dry semi-arid and arid lands of the west. Although the MDB is a very large catchment area, the surface water runoff is very small. However, compared to some areas of Australia the MDB receives comparatively more rainfall. While the south-east and east receive the most rainfall (up to 1000 mm), the western and north-western areas receive much less rainfall (less than 300 mm) (see figure 4.2). Furthermore, the seasonal appearance of precipitation is different in the north and the south of the MDB. The north is dominated by precipitation at the beginning of each year, whereas the south receives most of the precipitation approximately half a year later (e.g. in the second half of the year). The central part of the MDB defines a transition zone between both regimes. As will be seen throughout all analysis in this thesis the climatic north-south division plays an important role in the interpretation of results.

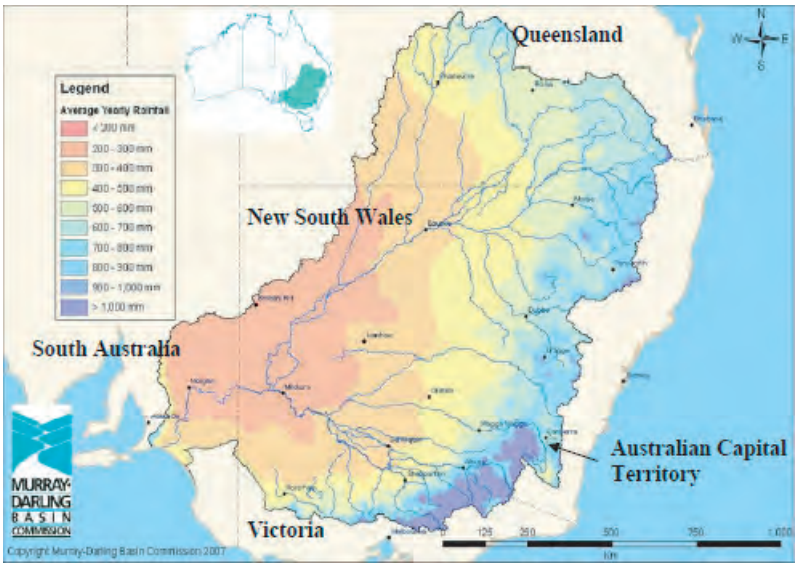


Figure 4.2: Rainfall distribution in the Murray Darling Basin [©Murray-Darling Basin Commission]

4.1.2 Problems in the MDB

Due to the changing climate, such as the drought conditions in the last decade many animal species are now endangered and some have already become extinct. Many rivers now have unusual low water flow in the austral winter and spring. The extraction of large amounts of water from the rivers has caused a decrease in the annual and seasonal variability. Due to the extensive clearing of native vegetation, more and more rainfall leaches the ground up to the groundwater. In turn augmenting groundwater levels cause increasing salinity.

4.2 Analysis of GRACE-derived surface mass changes

4.2.1 Preprocessing of original GRACE observations

Before the start of the investigation of Equivalent Water Thickness values derived from GRACE, some preprocessing steps have been carried out. In the first step the spherical harmonic coefficients (output of the level-2 GRACE product) are reduced by their individual long-term mean values ($\overline{c}_{lm}^{mean}$, $\overline{s}_{lm}^{mean}$) of the period May 2003 - April 2009 in order to receive residual time-variable coefficients (see equation 2.3). Some error sources cause a limitation of the spatial and temporal sampling of GRACE (e.g. Han et al. (2004)). It is not possible to differentiate between different error sources, as the GRACE observations contain the total detected mass variations on the Earth. Sources of these errors may include errors in the background model, limitations in the GRACE instruments, uneven distribution of the

groundtrack, system-noise and orbital errors, etc. For example about 10 days are needed for the GRACE mission to cover the whole Earth. Consequently, filter techniques are required to remove a significant part of the noise. A map of surface mass anomalies determined from unsmoothed GRACE observations, shows longitudinal stripes. The stripes denote a high degree of spatial correlation in the GRACE errors (e.g. Swenson/Wahr (2006)). It is assumed that the correlated errors are based on the along-track measurements of the K-Band instrument. Swenson/Wahr (2006) developed a filter technique to remove these correlated errors. It can be verified that certain spherical harmonic coefficients are correlated. Whereas even and odd coefficients are not correlated among each other, coefficients of even or odd parity are correlated among each other. Furthermore, the correlation errors for the coefficients of even or odd parity are different for the same order, which requires a separate examination of all coeffficients of even and odd degree for one particular order. However, not all spherical harmonic coefficients of even or odd parity are correlated. Not until order $m = 8$ correlations are apparent. For every order a particular polynomial is fitted separated for all odd degrees and for all even degrees, respectively. Subtracting the fitted coefficients from the original coefficients results in smoothed coefficients, and visible stripes are considerably removed. This technique is effective on spherical harmonic coefficients with short wavelengths, i.e. coefficients with a high degree contain noise, since the formal error degree is an accumulation of all formal errors δc_{lm} and δs_{lm}. In addition to removing correlated errors, the use of a Gaussian filter may yield even better results by further reducing high frequency errors. The Gaussian filter determines spatial averages of the surface mass density by

$$\overline{\Delta\sigma}(\theta,\lambda) = \int sin\theta' d\theta' d\lambda' \Delta\sigma(\theta',\lambda') W(\theta,\lambda,\theta',\lambda') \tag{4.1}$$

where $W(\theta,\lambda,\theta',\lambda')$ denotes an average function, which compensates the effects of poorly known, short-wavelength spherical harmonic coefficients (e.g. Wahr et al. (1998)). The average function $W(\theta,\lambda,\theta',\lambda')$ itself only depends on the spherical distance α between two points (θ,λ) and (θ',λ'), indirectly defined through

$$cos(\alpha) = cos(\theta)cos(\theta') + sin(\theta)sin(\theta')cos(\lambda - \lambda') \; . \tag{4.2}$$

After some manipulations on equation 2.5 together with the average function the averaged surface mass density is obtained by

$$\overline{\Delta\sigma}(\theta,\lambda) = \frac{2R\rho_{ave}\pi}{3} \sum_{l=0}^{L_{max}} \frac{2l+1}{1+k_l} \sum_{m=0}^{l} \overline{P}_{lm}(cos\theta) W_l \left(\Delta\overline{c}_{lm} cosm\lambda + \Delta\overline{s}_{lm} sinm\lambda\right) \; . \tag{4.3}$$

Dividing this equation again with the average water density ρ_w $(= 1000\frac{kg}{m^3})$ leads to the Averaged Equivalent Water Thickness (AEWT) $\overline{\Delta\nu}(\theta, \lambda)$.

In spatial representation the average function is defined by

$$W(\alpha) = \frac{b}{2\pi}\frac{e^{-b(1-cos\alpha)}}{1-e^{-2b}} \tag{4.4}$$

with

$$b = \frac{ln2}{1-cos(r_{\frac{1}{2}}/R)} \tag{4.5}$$

where R is the radius of the sphere, and $r_{\frac{1}{2}}$ being the distance on the Earth's surface at which $W(\alpha)$ has dropped to 1/2 of its value at $W(\alpha = 0)$. Therefore, $r_{\frac{1}{2}}$ is called the averaging or smoothing radius. With the choice of α the level of smoothing can be devised.
In the spectral domain, the filter coefficients W_l for a particular degree l are determined by a recursion formula:

$$\begin{aligned} W_0 &= \frac{1}{2\pi} \\ W_1 &= \frac{1}{2\pi}\left[\frac{1+e^{-2b}}{1-e^{-2b}} - \frac{1}{b}\right] \\ W_{l+1} &= -\frac{2l+1}{b}W_l + W_{l-1} \end{aligned} \tag{4.6}$$

As seen in equation 4.6, W_l only depends on the degree. The filter is called isotropic, since the filter coefficients do not depend on the spatial location.
In figure 4.3, the filter coefficients W_l for three different smoothing radii $r_{\frac{1}{2}} = 300\ km$, $r_{\frac{1}{2}} = 400\ km$, and $r_{\frac{1}{2}} = 500\ km$ are illustrated.

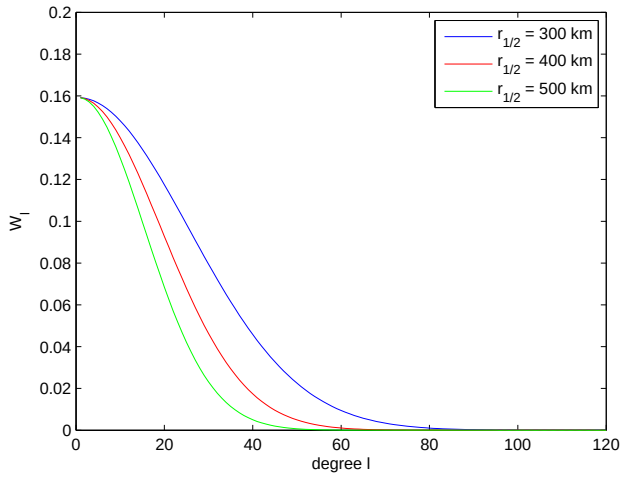


Figure 4.3: Gaussian averaging function in terms of spherical harmonic coefficients W_l

It can be clearly seen that the higher the chosen smoothing radius, the greater the effect of the smoothing. Spherical harmonic coefficients of a degree higher than approximately 50 are largely dampened for a smoothing radius of $r_{\frac{1}{2}} = 500\ km$, whereas for a smoothing radius of 300 km coefficients are not dampened until a degree of approximately 80. In return, maps of averaged surface mass changes determined with a smoothing radius of $r_{\frac{1}{2}} = 300\ km$ contain more noise (e.g. higher frequency signals).

It is not straight forward to choose the correct smoothing radius. In order to make a decision for the MDB, the Averaged Equivalent Water Thickness values with the smoothing radii 300 km, 350 km, 400 km, 450 km, and 500 km were determined for the period from May 2003 - April 2009. Before the application of the Gaussian smoothing, the filter for removal of correlated errors (Swenson/Wahr (2006)) was applied. It was performed with the software package FROG developed by Dr.-Ing Oliver Baur (Universität Stuttgart, Germany). In figure 4.4 the Averaged Equivalent Water Thickness (AEWT) values of September 2005 determined with a smoothing radius of 300 km and 500 km, respectively, are shown for the MDB.

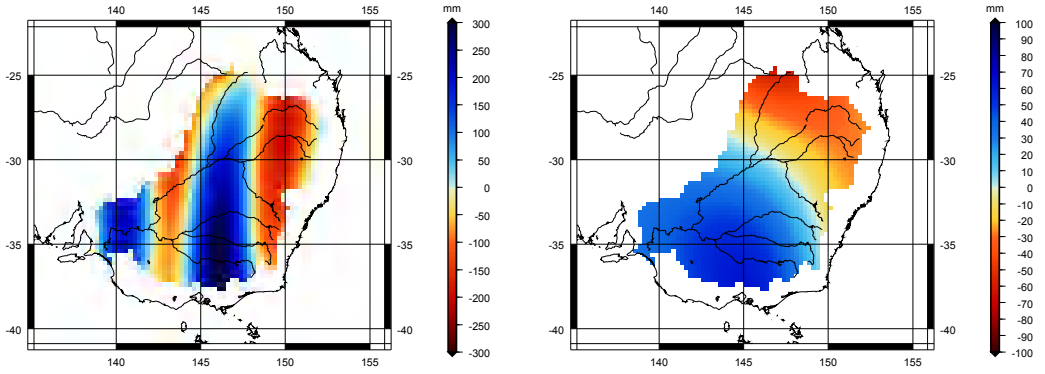


Figure 4.4: AEWT for the Murray-Darling Basin (left: $r_{\frac{1}{2}}$ =300 km - right: $r_{\frac{1}{2}}$ =500 km) for September 2005

There is a big difference between the AEWT values determined with a smoothing radius of 300 km and those determined with a radius of 500 km. In the left picture of figure 4.4 "artificial stripes" appear. These stripes start to vanish with a smoothing radius of 400 km. The best visual result of the five examined smoothing radii (300 km, 350 km, 400 km, 450 km, 500 km) is achieved with a smoothing radius of 500 km (right picture of figure 4.4), where the stripes disappear completely. The two different areas, being red in the north blue in the south coincide with the two different climatic zones, which cover the MDB (see section 4.1.1). While in the north, precipitation is heavy at the beginning of a year, whilst precipitation in the south is delayed by about half a year. As a result of the foregoing, the decision was made to use a smoothing radius of 500 km for the MDB, which corresponds to a resolution of approximately 5 degree at the equator ($\approx$ 500 km x 500 km

at the equator). Furthermore, a smoothing radius of 500 km is quite commonly used for such applications. The selection of a smoothing radius of 500 km is also a compromise in that it removes the striping effects, but still allows to study sub-basin changes.

4.2.2 Investigation by means of Multiple Linear Regression Analysis

With the aid of the methodology of Multiple Linear Regression Analysis described in section 3.1, the monthly determined Averaged Equivalent Water Thickness values for the MDB are investigated. For the application of the MLRA Matlab routines developed by Dipl.-Ing. Daniel Rieser (Technische Universiät Graz, Austria) were used. Rieser (2008) found a regression model, which is suitable for the whole of Australia, and showed that this model is statistically significant in nearly all parts of Australia. This model has the following form:

$$M_1(t) = c_0 + c_1 t + a_1 cos\omega t + b_1 sin\omega t \tag{4.7}$$

The parameters c_0 and c_1 describe an offset and trend, respectively.
The term

$$K(t) = a_1 cos\omega t + b_1 sin\omega t \tag{4.8}$$

with

$$\omega = \frac{2\pi}{T}$$

represents the annual part of the signal. Therefore, the period T is chosen to be T=12 month.
Applying the well-known least-squares adjustment technique delivers the adjusted parameters c_0, c_1, a_1, and b_1 for every grid point in the MDB. In the following, the original time series and the fitted model (left picture: red and blue line, respectively) together with its parameters and standard deviations (right picture: red and blue bar, respectively) are shown for three grid points representative for the northern, the central, and the southern part of the MDB (figures 4.5, 4.6, and 4.7).

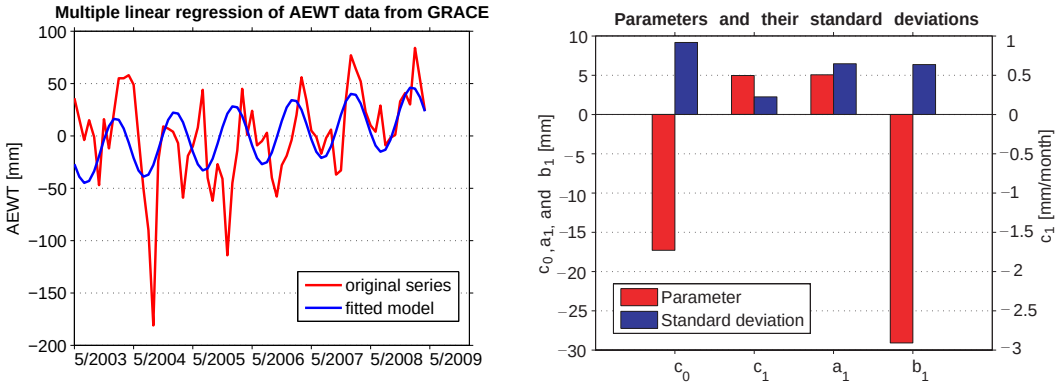


Figure 4.5: Multiple linear regression of AEWT values at $\lambda = 146.875°$, $\varphi = -24.625°$ (close to Warrego River)

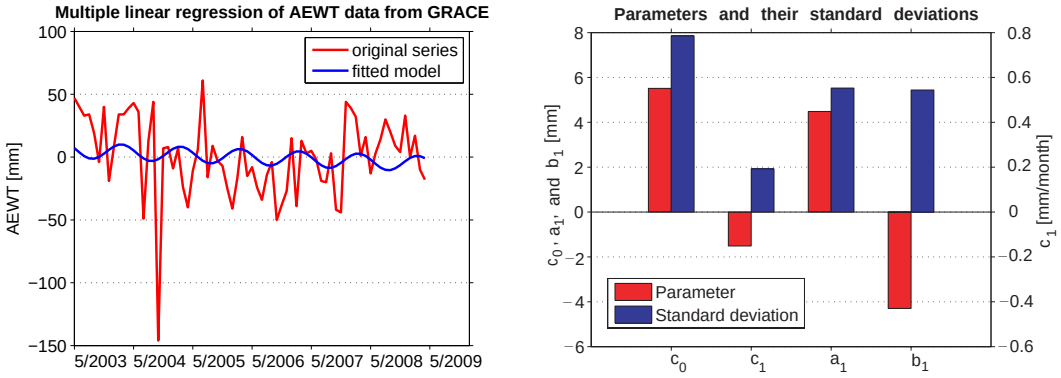


Figure 4.6: Multiple linear regression of AEWT values at $\lambda = 144.125°$, $\varphi = -29.375°$ (close to Darling River)

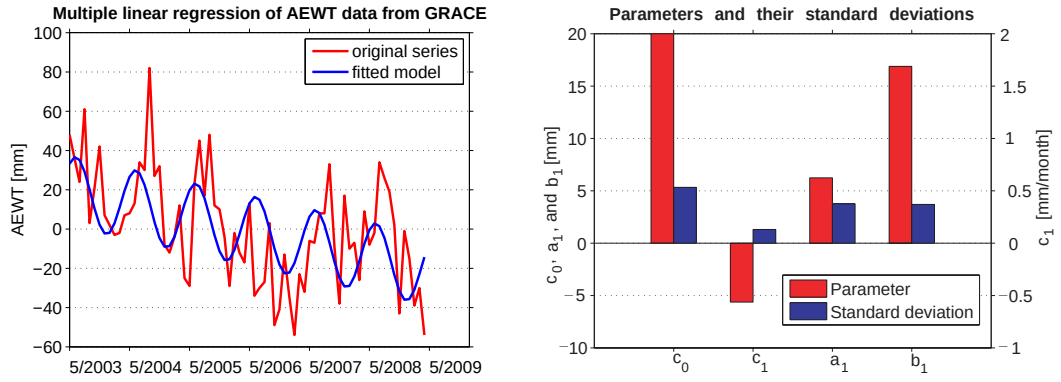


Figure 4.7: Multiple linear regression of AEWT values at $\lambda = 141.875°$, $\varphi = -35.625°$ (close to Murray River)

It can be seen that the model fits quite well for the grid point in the northern and the southern part of the MDB. In both cases, an annual signal and a linear trend can be recognized in the original time series. However, some outliers exist, which do not fit neatly within the chosen parameter model. While in figure 4.5

(which is located in the north of the MDB) the biggest amplitude appears in the early months of each year, in figure 4.7 (which is located in the south of the MDB) the maximal amplitudes occur approximately half a year later in each year. The absolute minimum values of the original time series in figure 4.7 are attributed to a long drought period, which peaked in the years 2006/2007. The peak of the drought was caused by an 18 month decile in rain from May 2006 until October 2007 (e.g. Campbell (2008)). During this period, the lowest rainfall on record in the upper Murray catchment was recorded.

It is remarkable that the trend parameter in figures 4.5 - 4.7, in comparison with the other parameters, is very small, and therefore, c_1 plays a minor role in the parameter model. In contrast, when one observes the left picture in figure 4.6, it appears as though the model doesn't fit to the original time series. This fact is affirmed by comparing the values of the parameters with the values for the standard deviations, which show that the values of the standard deviations are larger than the parameter values. Furthermore, the AEWT values do not show a clear periodic (e.g. annual) behaviour. These early results confirm that the central part of the MDB is a transition zone (see section 4.1.1) between the northern and the southern part of the MDB.

The model above is applied to the time series in every grid point over the whole MDB. The next step is to investigate the significance of the four parameters for all grid points. A Student t-test, as explained in chapter 3.1.2, is applied to all four parameters of the model (see equation 3.11). Table 4.1 provides an overview of the results of the applied t-test for the MDB. The theoretical quantile acounts for 1.997 (particular null hypothesis $H_0 : j = 0$ with $j = c_0,\ c_1,\ a_1$, and b_1, probability 97.5% (two-sided test)).

parameter j	min $\lvert t_j \rvert$	max $\lvert t_j \rvert$	# significant	# not significant
c_0	0.0090	4.4949	834 (51.6%)	782 (48.4%)
c_1	0.0032	5.1840	893 (55.3%)	723 (44.7%)
a_1	0.00002	4.1816	247 (15.2%)	1369 (84.7%)
b_1	0.0093	4.8324	1082 (67.0%)	534 (33.0%)

Table 4.1: Significance tests for the four different parameters of the MLRA model applied to the AEWT values

With the exception of the parameter a_1 of the annual signal, all parameters are significant with a probability of 97.5% in more than half of all 1616 grid points in the MDB. Only 15.2% of all locations show a significant behaviour in terms of the cosine-fraction of the annual signal (a_1 magnitude), which is a very low value in comparison to all other parameters. Selecting May 2003 as start of the time series (e.g. $t = 0$ for May 2003) the time series follows a sine wave in the south of the MDB and a sine wave shifted by six months in the north of the MDB (see

alternating sign of the parameter b_1 in figure 4.5 and figure 4.7). This is consistent with a maximum early in the year for the north and six months later in the year for the south. Therefore, the low value of 15.2% of the cosine part of the annual signal (a_1) is not surprising.
Since the parameters of the annual part of the model belong together and should not be investigated separately, another approach to test the significance of the annual part should be applied. Furthermore, the Student t-test is based on the assumption that the parameters c_0, c_1, a_1, and b_1 are uncorrelated between each other. This is of course not true, as can be seen in figure 4.8, illustrating the correlation coefficients for the grid point $\lambda = 146.875°$, $\varphi = -24.625°$ in the northern part of the MDB. For example there is a strong negative correlation between the offset und trend parameters c_0 and c_1, respectively, while the parameters a_1 and b_1, describing the annual part, have a rather weak correlation.

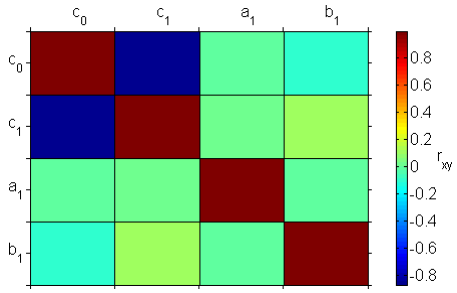


Figure 4.8: Correlation coefficients between model parameter at $\lambda = 146.875°$, $\varphi = -24.625°$ (GRACE)

In order to account for possible correlation among the different parameters, the technique of reduced models explained in chapter 3.1.2 can be applied. With this approach, a significance test can also be applied to determine whether the addition of a_1 and b_1 makes a significant contribution (e.g. better fit to the observations) to the model containing only an offset and a trend. The same significance can also be used to determine whether the addition of further parameters will yield improved results. Therefore, it will provide some indication which of the examined models is statistically "best" suited to represent the original observations.

The following cases as given below are investigated:

1. $M_0 \rightarrow M_1$:
   reduced model: $M_0 = c_0 + c_1 t$
   $\longrightarrow$ full model: $M_1 = c_0 + c_1 t + a_1 cos(\omega t) + b_1 sin(\omega t)$
   null hypothesis: $a_1 = b_1 = 0$ (annual signal)

2. $M_1 \rightarrow M_2$:
   reduced model: $M_1 = c_0 + c_1 t + a_1 cos(\omega t) + b_1 sin(\omega t)$
   $\longrightarrow$ full model: $M_2 = c_0 + c_1 t + ... + a_2 cos(2\omega t) + b_2 sin(2\omega t)$
   null hypothesis: $a_2 = b_2 = 0$ (semi-annual signal)

3. $M_2 \rightarrow M_3$:
   reduced model: $M_2 = c_0 + c_1 t + ... + a_2 cos(2\omega t) + b_2 sin(2\omega t)$
   $\longrightarrow$ full model: $M_3 = c_0 + c_1 t + ... + a_3 cos(3\omega t) + b_3 sin(3\omega t)$
   null hypothesis: $a_3 = b_3 = 0$ (1/3 - annual signal)

4. $M_3 \rightarrow M_4$:
   reduced model: $M_3 = c_0 + c_1 t + ... + a_3 cos(3\omega t) + b_3 sin(3\omega t)$
   $\longrightarrow$ full model: $M_4 = c_0 + c_1 t + ... + a_4 cos(4\omega t) + b_4 sin(4\omega t)$
   null hypothesis: $a_4 = b_4 = 0$ (1/4 - annual signal)

The one-sided F-test from equation 3.26 is now applied. The results are demonstrated in table 4.2.

case	$F_{0.95,2,n_{full}-u_{full}}$	# F significant	# F not significant
1 ($M_0 \rightarrow M_1$)	3.1359	1349 (83.5%)	267 (16.5%)
2 ($M_1 \rightarrow M_2$)	3.1404	1404 (86.9%)	212 (13.1%)
3 ($M_2 \rightarrow M_3$)	3.1453	1398 (86.5%)	218 (13.5%)
4 ($M_3 \rightarrow M_4$)	3.1504	1385 (85.7%)	231 (14.3%)

Table 4.2: Significance tests for the comparison of different multiple linear regression models (GRACE)

All tests show that for more than 80% of all grid points covering most of the MDB the inclusion of an annual signal is significant. Even if the best case appears for expanding from M_1 to M_2, the chosen model M_1 is kept, as with an additional semi-annual part an augmentation of only 3% is achieved, which is not overly different when using only M_1. Furthermore, a physical interpretation of an additional semi-annual part and all further periodic parts in the model is difficult.
The coefficient of determination R^2, and the adjusted coefficient of determination $\tilde{R}^2$ introduced in section 3.1.2, can be interpreted as a statistical measure of the 'goodness of fit' of a particular model. Both coefficients were determined for each grid point (see equation 3.12 and equation 3.17, respectively). The minimum and maximum for both coefficients are shown in table 4.3.

	minimum	maximum
R^2	0.0265	0.4213
$\tilde{R}^2$	-0.0334	0.3857

Table 4.3: Coefficient of determination R^2 and $\tilde{R}^2$ (GRACE)

Indeed the values are very low. A rule of thumb (Brook/Arnold (1985)) denotes that the coefficient of determination should be greater than 0.5 to have much confidence in the model. However, as this rule of thumb is not based on a statistical test, it is not applied in this thesis with table 4.3 showing that it might not be suited for the MDB. In order to investigate the low values of table 4.3, the significance of the chosen model (see equation 3.19 in section 3.1.2) is determined, and is illustrated in figure 4.9 (null hypothesis: $\tilde{R}^2 = 0$, theoretical quantile $F_{u,n-u-1} = F_{4,65} = 2.51$ with a probability of 95% (one-sided test)). It is observed that especially in the central part of the MDB, where values of the test variable are close to zero, the model M_1 is not suitable. The model is significant along the Murray River, Murrumbidgee River, and the Lachlan River (encompassing the south of the region), where the highest values of F appear, and for a small band in the north. Altogether, 66% of all grid points are significant. Therefore, the transition zone (see section 4.1.1) in the central part of the MDB between an annual signal in the north and an annual signal in the south finds additional support.

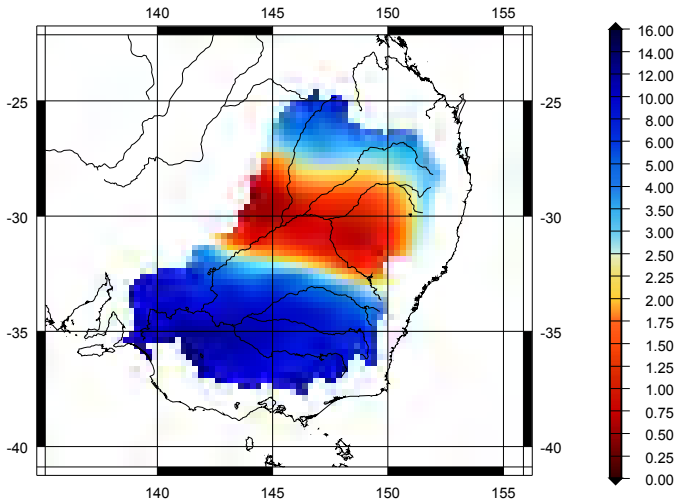
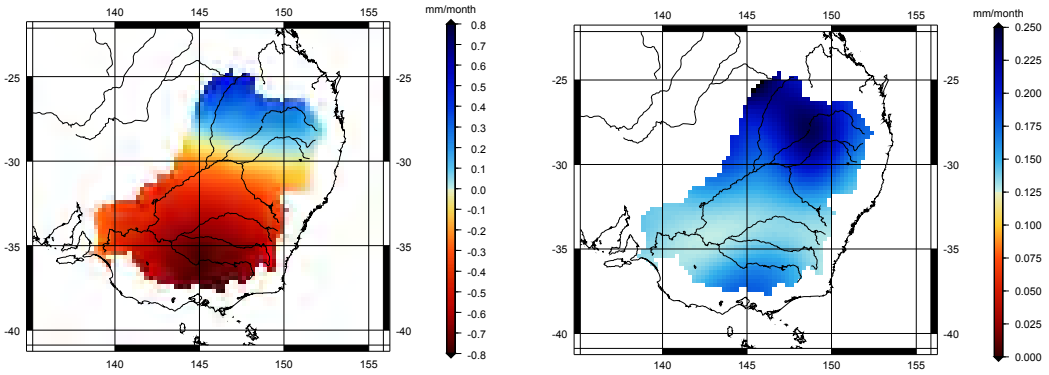


Figure 4.9: Significance test for the chosen model M_1 (GRACE)
(red colours mean not significant, blue colours mean significant)

Finally, the trend and annual part of the model are investigated separately. The trend parameters, together with its standard deviations for the whole MDB, are determined (see figure 4.10). By comparing the trend parameters in the north and in the south, an inverse behaviour is visible. In the north, maximum values of approximately 0.5 mm/month can be detected, while in the south the maximum values account for about -0.8 mm/month. The negative values in the south indicate a mass loss, probably caused by a long drought period. This leads to the supposition that the north was not as badly affected by the drought, indicating mass gain over the period examined. The standard deviation values are relative high especially in the north of the MDB in comparison with its corresponding trend values seen in the figures 4.5, 4.6, and 4.7.



(a) trend of AEWT values (b) standard deviation of trend

Figure 4.10: Trend together with its standard deviation for AEWT values

The results of the applied t-test for the trend parameters (see table 4.1) in the MDB are illustrated in figure 4.11. A significant trend is present for the complete southern part, whereas a trend in the northern part is not significant apart from the very northern part.

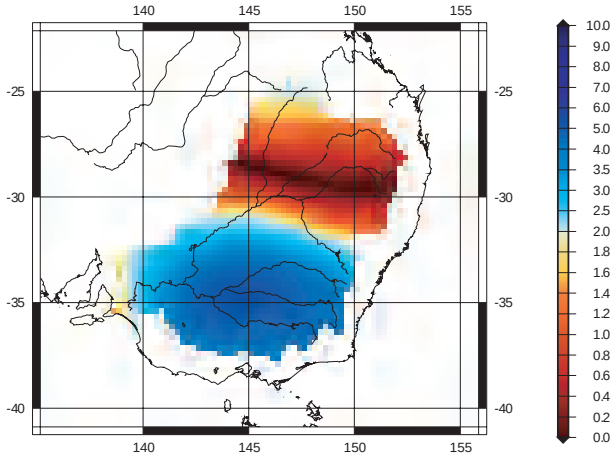


Figure 4.11: Significance test for the trend parameter (GRACE)
(red colours mean not significant, blue colours mean significant)

In addition to the trend parameter, the annual part of the signal is analysed, which consists of a linear combination between the sine and cosine waves of M_1. The technique of reduced models (see table 4.2) indicates that an additional annual part to the model M_0 (containing only offset and trend) is necessary. The null hypothesis $a_1 = b_1 = 0$ was rejected for 84% of all grid points. In figure 4.12 the F-test (theoretical quantile $F_{2,66} = 1.997$ with a probability of 95% (one-sided test)) for the annual part of the signal is illustrated. In the central part of the MDB the test variable is smaller than the theoretical quantile, which indicates that the null hypothesis $a_1 = b_1 = 0$ is true. This means an annual signal cannot be found in the central part of the MDB. This confirms again that this area is a transition zone (see section 4.1.1) between the northern and southern part of the MDB.

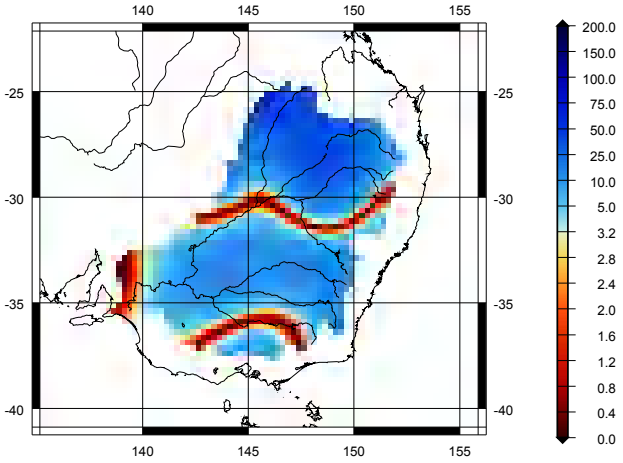
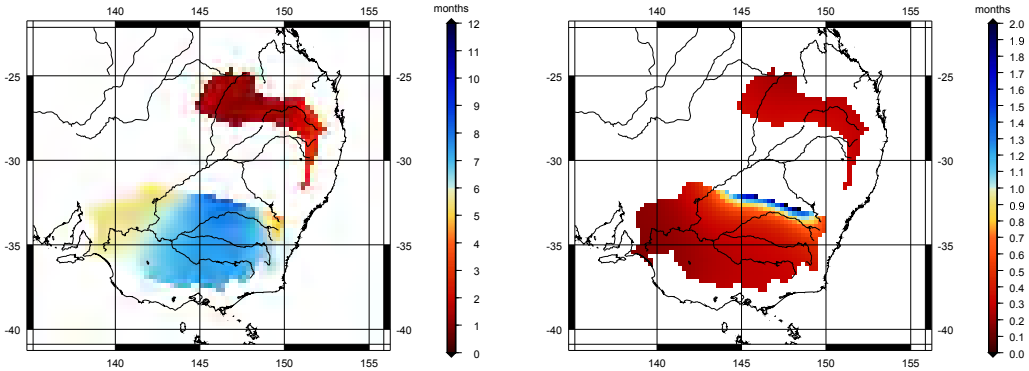


Figure 4.12: Significance test for the annual signal (GRACE)
(red colours mean not significant, blue colours mean significant)

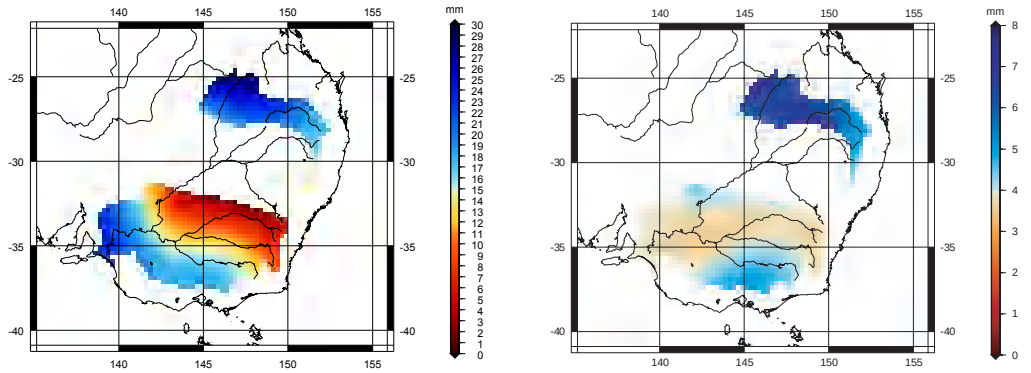
The annual signal can be characterised by the parameters phase and amplitude. Both phase and amplitude with their corresponding standard deviations are determined by equations 3.21 to 3.24 in section 3.1.2. Figure 4.13 illustrates the determined phase for all 1616 grid points denoted in the unit months, along with the corresponding standard deviation for each grid point. The phase denotes here the month in which the maximum of the annual signal with respect to the beginning of a year appears (e.g. start of a cycle of a cosine signal). Since the surface mass changes are monthly mean values, the phase is determined with reference to the middle of January. Note the phase with its corresponding standard deviation is only plotted for grid points, where the M_1 model is significant.



(a) phase of annual signal (b) standard deviation of phase

Figure 4.13: Phase of the annual signal together with its standard deviation (GRACE)
(phase is defined with respect to mid-January)

A phase difference between the northern and southern part of approximately 5 months can be identified. In the northern part the mean phase accounts for approximately 1.5 months, while the mean phase in the south is approximately 6.5 months. Therefore, 1.5 months corresponds with the beginning of March and 6.5 months corresponds with the beginning of August. This fact coincides with the seasonal occurrence of precipitation in the north (beginning of the year) and south (middle of the year). The accuracy of the phase is generally within half a month as would be expected with monthly mean data. More details of the phase analysis are represented in chapter 5, where amongst others the phases of the different data types are compared.
The amplitude and the corresponding standard deviation are also represented for significant grid points only (see figure 4.14).



(a) amplitude of annual signal (b) standard deviation of amplitude

Figure 4.14: Amplitude of the annual signal together with its standard deviation (GRACE)

The largest amplitudes are reached in the northern and in the southern part of the MDB along the Murray River (up to 30 mm). In contrast the amplitudes along the Murrumbidgee River, Lachlan River (also both in the southern part of the MDB), and along parts of the Darling River (northern to southern part of the MDB) are only half of the magnitude.

4.2.3 Investigation by means of Principle Component Analysis

Before the application of the PCA, the Averaged Equivalent Water Thickness values must be preprocessed as described in chapter 3.2.2. First the residuals of the AEWT values with respect to the particular mean value are determined. To do this for each of the 1616 grid points the mean value for the time period from May 2003 until April 2009 inclusive are calculated. After this, for every grid point the particular mean value is subtracted, which leads to the residuals $z(t, x)$. The parameters t and x, respectively, stand for the month (time) and grid point (location), respectively. In the next step the residuals are sorted into the matrix Z (see equation 3.29 in section 3.2.2), e.g. Z has the following form.

$$Z = \begin{bmatrix} z(1,1) & z(1,2) & \dots & z(1,1616) \\ z(2,1) & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ z(70,1) & \dots & \dots & z(70,1616) \end{bmatrix} \tag{4.9}$$

With the formulae from section 3.2.2 the PCA can now be carried out, which is performed with FORTRAN 77 software scripts provided by Dr.-Ing. Michael Kuhn. In order to decide how many modes should be taken into account, figure 4.15 represents the relative amount of overall variability for a particular mode (cf. equation 3.39 in section 3.2.2), together with its mode number. The first 10 modes are illustrated in figure 4.15.

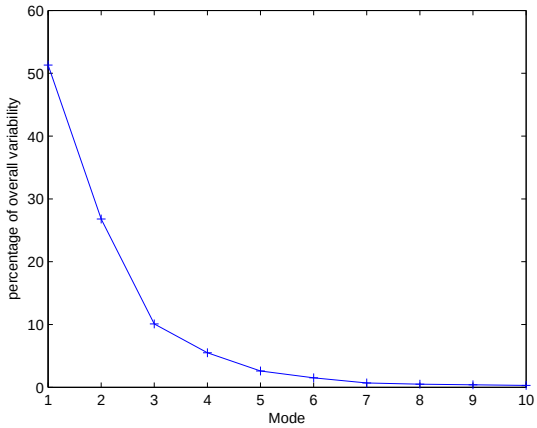
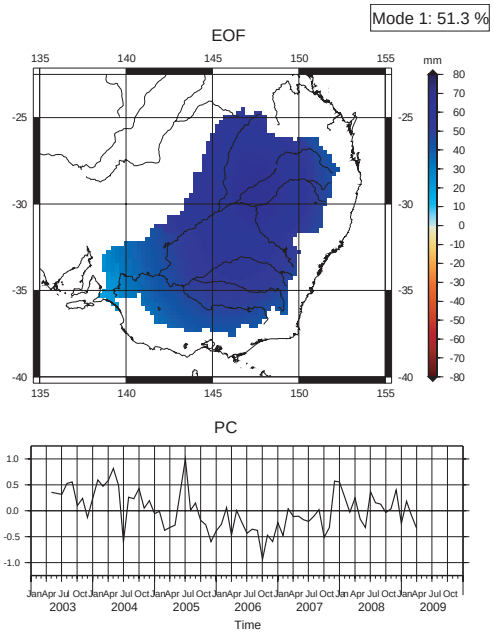
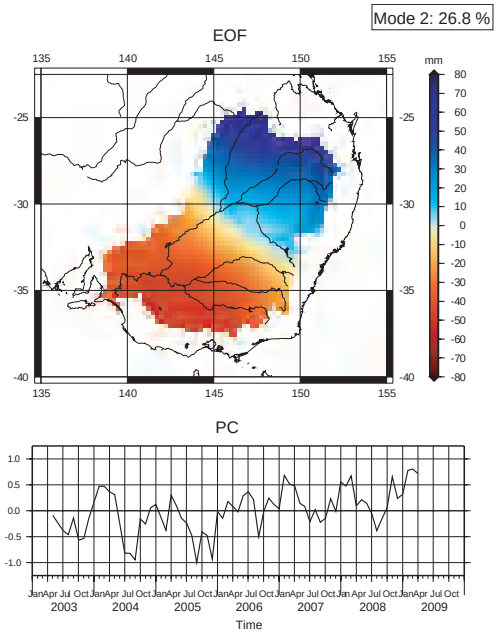


Figure 4.15: Percentage of total variability (GRACE)

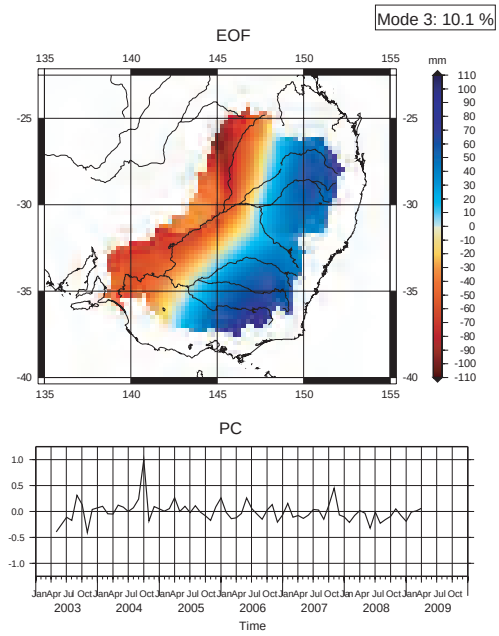
It can be seen that the first four modes combined account for approximately 95% per cent of the signal's variability. The largest gradient appears until mode 3 or mode 4. Therefore, it can be assumed that at least the first three modes and maybe also the fourth mode represent almost the total signal variability. The EOF's with their particular PC's for the first four modes are illustrated in figure 4.16.



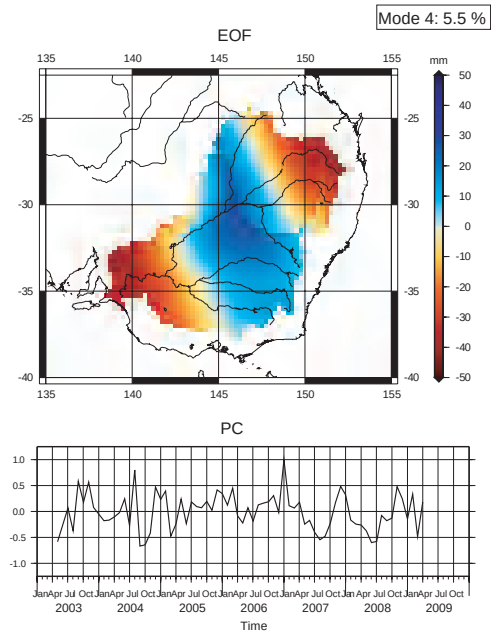
(a) Mode 1



(b) Mode 2



(c) Mode 3



(d) Mode 4

Figure 4.16: PCA modes 1-4 of AEWT values

Just over half (51.3%) of the variation of GRACE-derived surface mass changes is represented by mode 1. It is remarkable that no clear annual signal can be recognized in the first mode of the PC time series, which is quite unusual. The EOF map, describing the spatial distribution of the magnitude of the PC, shows an almost uniform distribution for the whole area with a mean value of the EOF pattern of approximately +50 mm. Besides a missing annual signal in the PC time series of mode 1, it is noticeable that between October 2005 and November/December 2007 the PC time series only shows negative values which, together with the positive values of the PC time series, indicates a mass deficit for the whole MDB during this time period. The largest amount of mass deficit appears between July 2006 and January 2007. Multiplying these PC values with the mean value of 50 mm from the EOF yields a mass loss between 25 and 50 mm over the whole MDB. The reason for that long period of mass loss is a long term drought (e.g. Campbell (2008)). This drought took place between May 2006 and October 2007. This time period coincides nearly exactly with the PC time series, because during this time, the absolute minimum values appear in the PC time series (absolute minimum: October 2006). In the beginning of 2008, an increase in mass can be detected. Consequently, instead of an annual signal the first mode is mainly dominated by the drought.

Mode 2 represents almost 27% of the total surface mass variation. In contrast to mode 1, an annual signal can now be detected in the PC time series with the maxima at the beginning of each year. Furthermore, the MDB is now separated in two zones, namely the north and the south. The EOF pattern has a positive sign in the north, while the south is negative, whereby the north has larger values. This means the PC time series is dominated by the northern part of the MDB where mass gain (through precipitation) occurs at the beginning of each year. During this time, the southern part is largely dry, which is visible by the mass deficit. The transition zone (see section 4.1.1) between mass deficit and mass surplus is exactly in the central part of the MDB with values of close to zero, e.g. neither mass deficit nor mass surplus. Consequently, the phase difference of half a year between the north and the south detected by the phase in figure 4.13 (MLRA in section 4.2.2) can be affirmed by the second mode. Mode 1 and mode 2 combined account for more than 78% of the total variability.

For the third mode it is hard to find a clear interpretation. This mode likely describes mainly noise, since the PC time series oscillates around zero. However, an anomaly with a positive peak appears for October 2004. In contrast to mode 2 the spatial pattern is divided into two zones in the east and in the west. There might be a connection with the location of the different rivers in the MDB. For example, the negative values are only concentrated along the Darling River, which connects the north with the south. A remarkable mass deficit would only appear along the

Darling river for the anomaly occuring in October 2004. Note however that this mode only represents about 10% of the total variability and is described according to the appearance of the PC time series mainly by noise.
As with the third mode, mode 4 does not readily give rise to a conclusive physical explanation. Beginning in 2007, an annual signal can be recognized in the PC time series. It can be also assumed that mode 4 describes largely noise, since less than 6% of the total variation is represented by this mode. The EOF map is divided into three zones. High amplitudes appear in a stripe going from north to south.
In summary mode 1 and mode 2 are the most important ones. Mode 1 describes a uniform behaviour for the whole MDB and is mostly dominated by the drought. An annual signal is first detectable in the second mode, where a phase difference of approximately half a year between the north and south of the MDB appears.

4.3 Analysis of precipitation observations from TRMM

Prior to the application of Multiple Linear Regression Analysis and Principal Component Analysis of the precipitation observations this data is also filtered, as with the GRACE-derived surface mass changes, by applying a Gaussian filter (smoothing radius 500 km). This is necessary for a direct comparison with the GRACE-derived surface mass changes.

4.3.1 Investigation by means of Multiple Linear Regression Analysis

The investigation starts with the same procedure used for the GRACE-derived surface mass changes. The model approach is in turn the model M_1 (see equation 4.7) consisting of an offset, trend and annual part as used for the GRACE-derived surface mass changes.
In figure 4.17, 4.18, and 4.19 the original time series and the fitted models are again demonstrated for three grid points, representing the northern, the central, and the southern part of the MDB.

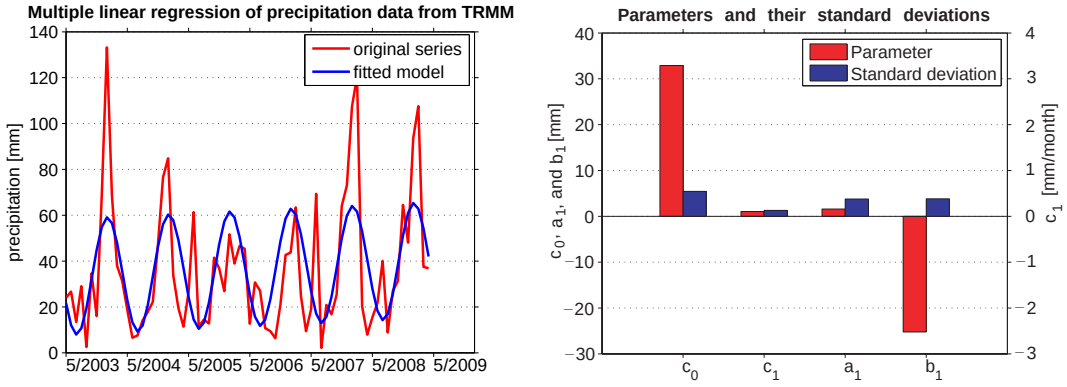


Figure 4.17: Multiple linear regression of precipitation observations at $\lambda = 146.875°$, $\varphi = -24.625°$ (close to Warrego River)

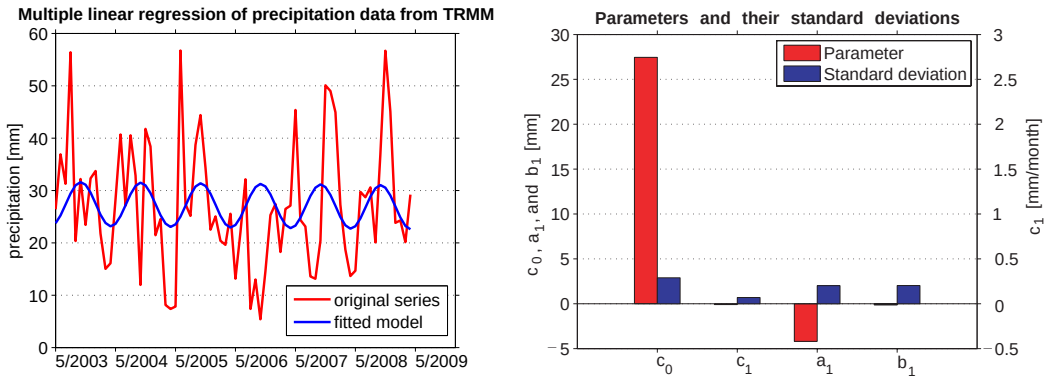


Figure 4.18: Multiple linear regression of precipitation observations at $\lambda = 142.375°$, $\varphi = -32.125°$ (close to Darling River)

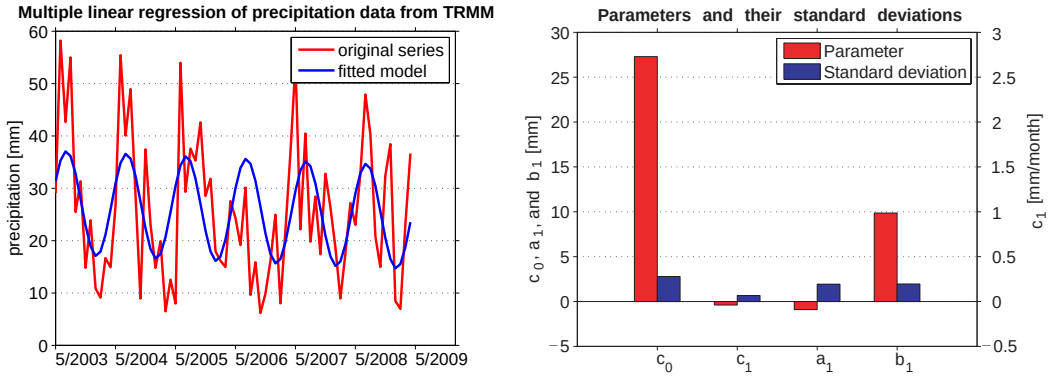


Figure 4.19: Multiple linear regression of precipitation observations at $\lambda = 139.125°$, $\varphi = -35.625°$ (close to Murray River)

Comparing the six illustrations above, it is immediately apparent that the model doesn't fit in figure 4.18, representing the central part of the MDB, which already was ascertained for the GRACE-derived surface mass changes. When reviewing the original time series, parts of an annual signal cannot be identified. Therefore, a

suitable model with offset, trend parameter and an annual part cannot be found for the central part of the MDB being a climatic transition zone (see section 4.1.1). The fitted model in the north and south fits quite well to the data with the exception of a few peaks. As with the GRACE-derived surface mass changes (see figure 4.5, 4.6, and 4.7) the trend values are very low. Offset and b_1-Parameter are, with the exception of figure 4.18, considerabley larger than their corresponding standard deviations. Therefore, both parameters are likely to be significant. The time series of these three grid points (precipitation observations) shows (as with the results of the analysis of GRACE-derived surface mass changes) that annual parts can only be found in the north and the south, whereas no annual signal can be found in the central part of the MDB. In order to confirm this fact some hypothesis tests are also applied. A Student t-test is accomplished for each of the four parameters. The theoretical quantile value accounts for 1.996 (probability 97.5% (two-sided test)). The results for all 1616 grid points are represented in table 4.4.

parameter j	min $\|t_j\|$	max $\|t_j\|$	# significant	# not significant
c_0	6.0419	11.1347	1616 (100.0%)	0 (0.0%)
c_1	0.0030	1.2492	0 (0.0%)	1616 (100.0%)
a_1	0.0020	2.1253	121 (7.5%)	1495 (92.5%)
b_1	0.0007	6.6131	1121 (69.4%)	495 (30.6%)

Table 4.4: Significance tests for the four different parameters of the MLRA model applied to the precipitation observations

The t-test clearly shows that the trend parameter is not significant for all 1616 grid points. Therefore, a trend should not be included in the model. However, to make a fair comparison with the MLRA results of GRACE-derived surface mass changes, the same model should be used and as a consequence the trend is not excluded from the model. Furthermore, only about 8% of all grid points show a significant a_1 parameter. This is again obvious, since figures 4.17 and 4.19 show as for the GRACE-derived surface mass changes (see figures 4.5 and 4.7) a sine wave in the southern part of the MDB and a sine wave shifted by 6 months in the northern part of the MDB.

To study the influence of the annual part of the signal represented by the parameter a_1 and a_2 again the method of reduced models is applied. As mentioned in section 4.2.2, the one-sided F-test for the technique of reduced models (see equation 3.26) shows the influence of additional parameters to the model. Table 4.5 shows the results of the applied test.

case	$F_{0.95,2,n_{full}-u_{full}}$	# F significant	# F not significant
1 ($M_0 \rightarrow M_1$)	3.1317	1335 (82.6%)	281 (17.4%)
2 ($M_1 \rightarrow M_2$)	3.1359	1383 (85.6%)	233 (14.4%)
3 ($M_2 \rightarrow M_3$)	3.1404	1399 (86.6%)	217 (13.4%)
4 ($M_3 \rightarrow M_4$)	3.1453	1383 (85.6%)	233 (14.4%)

Table 4.5: Significance tests for the comparison of different multiple linear regression models (TRMM)

When compared to the GRACE-derived surface mass changes the tests show nearly the same results. For more than 80% the null hypothesis for the four different cases must be declined (cf. table 4.5). There is no appreciable improvement for an expansion from M_1 to M_2 and to all further models. For this reason the M_1 model (see equation 4.7) is again suitable for the investigations. The parameters R^2 and $\tilde{R}^2$ are computed for each grid point, which describe the goodness of fit of the model. The minimum and maximum value are given in table 4.6.

	minimum	maximum
R^2	0.0083	0.4070
$\tilde{R}^2$	-0.0509	0.3716

Table 4.6: Coefficient of determination R^2 and $\tilde{R}^2$ (TRMM)

It also appears that both coefficients have rather low values. Once again the rule of thumb (Brook/Arnold (1985)) is not fulfilled ($R^2 > 0.5$) for any of the grid elements. Therefore, it is also better to determine additionally the significance of the model M_1 by applying an F statistical test. The theoretical quantile is $F_{4,n-u-1} = F_{4,67} = 2.5087$ (probability 95% (one-sided test)). The results of the hypothesis test are illustrated in figure 4.20.

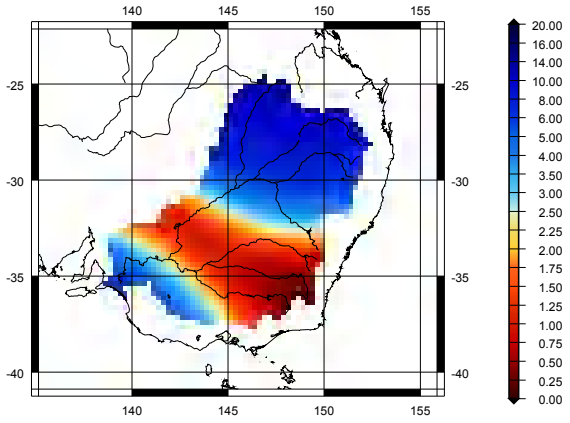


Figure 4.20: Significance test for the model M_1 (TRMM)
(red colours mean not significant, blue colours mean significant)

This again demonstrates that the M_1 model is least significant in the central part of the MDB. Significant regions (covering 60% of the MDB) form a small band in the south and a larger band in the north. Thus, this test provides further support

for the proposition that the central part of the MDB is a climatic transition zone (see section 4.1.1).
Considering the annual part of the model, the results of the F-test (reduced models) for the annual part of the signal from table 4.5 are illustrated in figure 4.21.

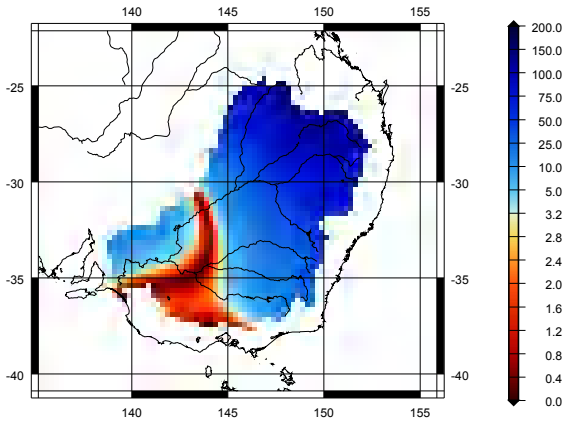
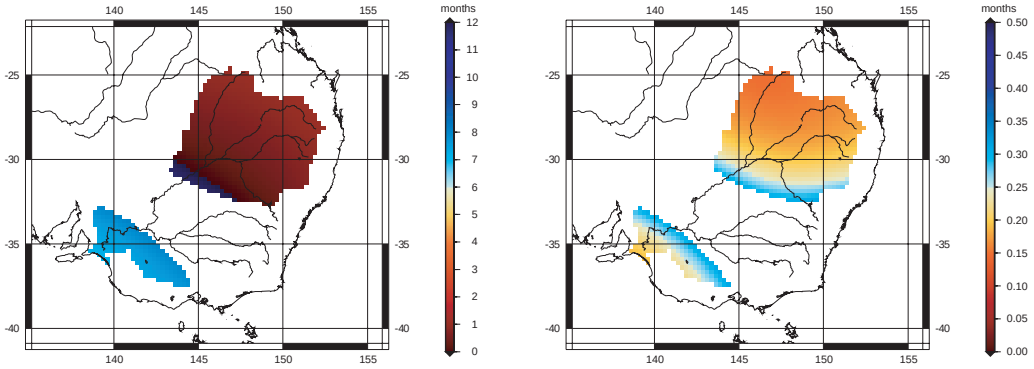


Figure 4.21: Significance test for the annual signal (TRMM)
(red colours mean not significant, blue colours mean significant)

For the whole northern part of the MDB, an annual part in the model is necessary, whereas the central part and parts of the south of the MDB aren't significant.
Phase and amplitude, which describe the parameter of the annual part of the model, are computed. The phase and its standard deviation (see equations 3.21 and 3.23 in section 3.1.2) are determined for all grid points (see figure 4.22).

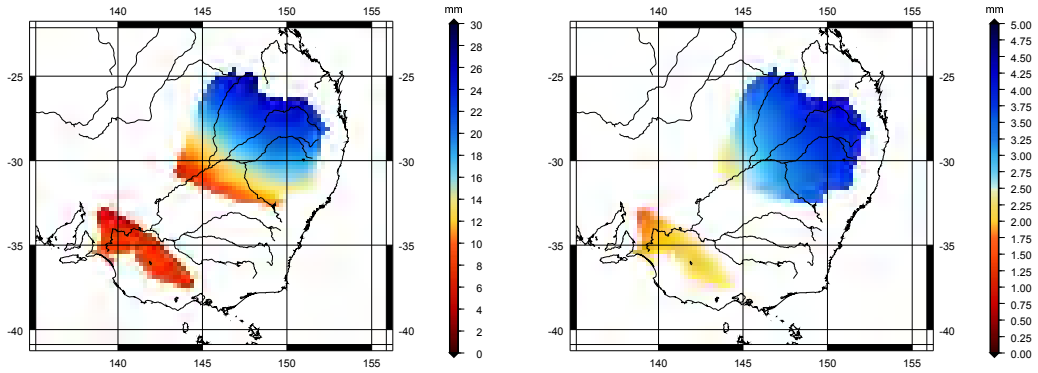


(a) phase of annual signal (b) standard deviation of phase

Figure 4.22: Phase of the annual signal together with its standard deviation (TRMM)
(phase is defined with respect to mid-January)

Again two different phases are detectable for the north and the south. The maximum of the annual signal indicates with respect to mid-January again in which months of the year most of precipitation occurs. In the north a maximum of the annual signal appears in the beginning of February, whereas the south has a maximum in the beginning of September. The magnitude of the standard deviations is the same throughout (approximately 0.2 - 0.4 months).

Figure 4.23 illustrates the amplitude and the corresponding standard deviation (see equations 3.22 and 3.24 in section 3.1.2) in all 1616 grid points.



(a) amplitude of annual signal (b) standard deviation of amplitude

Figure 4.23: Amplitude of the annual signal together with its standard deviation (TRMM)

Compared to the south the magnitude of precipitation in the north is approximately four times greater (south: $\approx$ 3-8 mm, north: $\approx$ 20-25 mm). The standard deviations throughout the MDB vary between 2 and 4 mm.

4.3.2 Investigation by means of Principal Component Analysis

As with the PCA for the GRACE-derived surface mass changes, mean-centred precipitation observations (mean taken over the whole time period (e.g. 6 years) considered) are also determined, before the Principal Component Analysis can be applied. After that the mean-centred values are ordered in the matrix Z (see equation 4.9) and the Principal Component Analysis is applied. However, in contrast to the GRACE-derived surface mass changes, where two months (June 2003 and January 2004) are completely missing in the data, the Z matrix for the precipitation observations has 72 rows (72 months). Figure 4.24 illustrates the relative amount of the overall variability for the first 10 modes, together with their mode numbers.

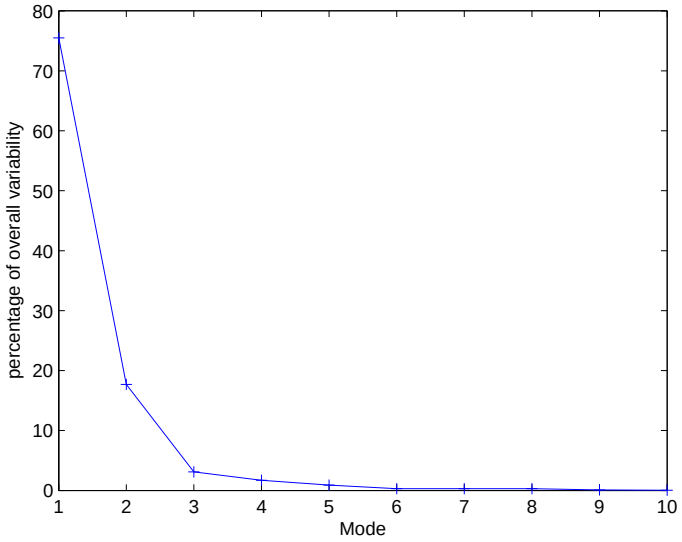
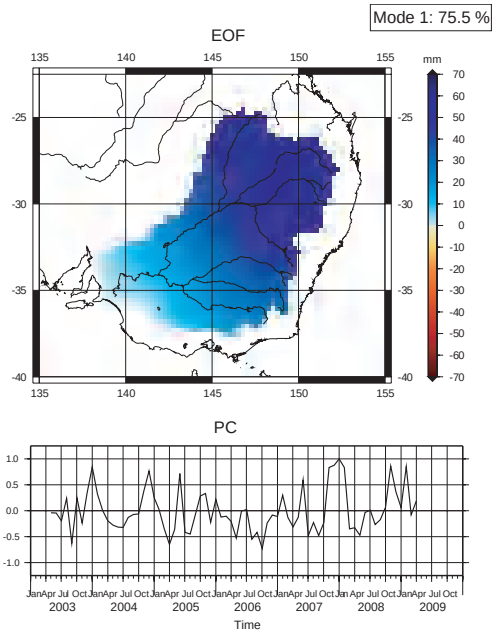
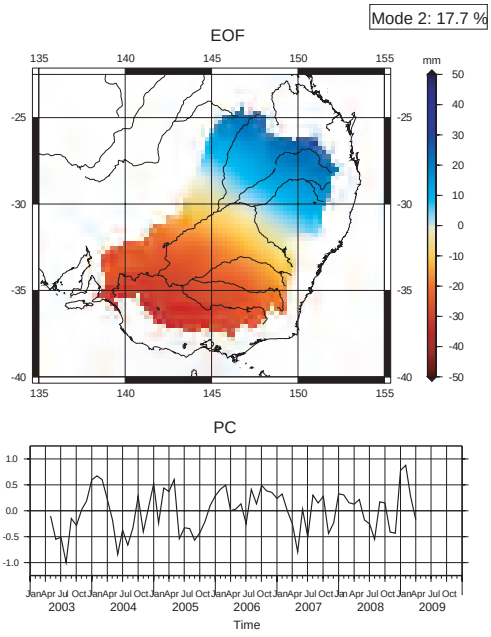


Figure 4.24: Percentage of total variability (TRMM)

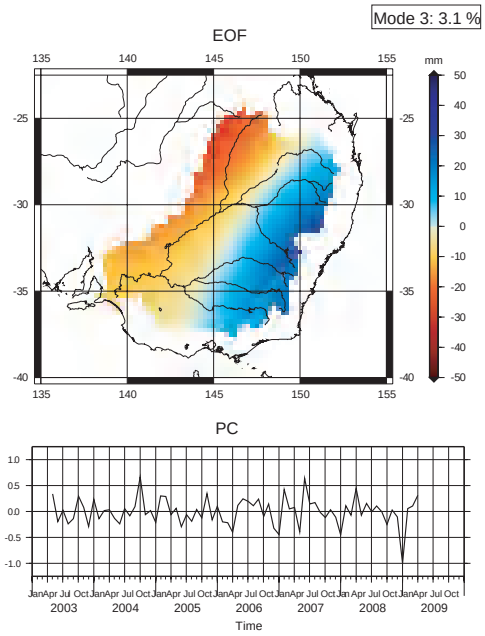
The first mode covers almost 76% of the total variability. Compared to the GRACE-derived surface mass changes (cf. section 4.2.3), here the first mode covers 25% more of the overall variation. A big decline between the first and second mode is visible in figure 4.24 (mode 2 18%). At mode 3 the gradient becomes much lower, which means that further modes likely contain more noise. Since the different modes between precipitation observations and GRACE-derived surface mass changes will be compared, the fourth mode of the precipitation observations is also investigated, even if this mode only covers 1.7% of the variability of the original precipitation observations. With the first four modes the original precipitation observations can almost completely be described (98% of variability). Figure 4.25 shows the EOF maps and PC time series for the first four modes.



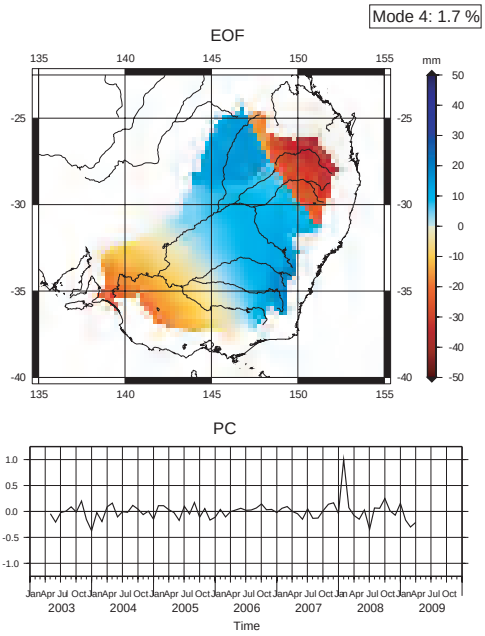
(a) Mode 1



(b) Mode 2



(c) Mode 3



(d) Mode 4

Figure 4.25: PCA modes 1-4 of precipitation observations from TRMM

When comparing these results to that obtained from the GRACE-derived surface mass changes, nearly the same spatial pattern with the same magnitudes in the EOF map of the first mode can be identified, which indicates that the whole MDB follows the same temporal behaviour. Only the south-east has slightly lower values (light blue). Once again, an annual signal cannot be found in the first mode. Remarkably, the first mode takes 76% of the overall variability into account, which is nearly 25% more than for the GRACE-derived surface mass changes. Consequently, the drought period is more evident in terms of the precipitation observations than in terms of the GRACE-derived surface mass changes. Comparing the PC time series of the first mode with the results from the GRACE-derived surface mass changes the same peaks are found. For example, in October 2006, both data sets have an absolute minimum, indicating the peak of the drought.

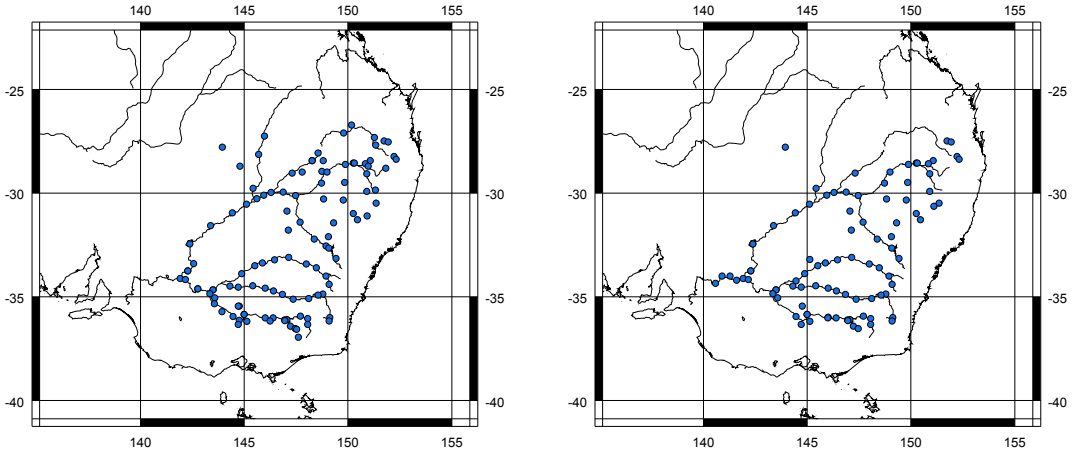
In contrast to mode 1, an annual signal in the PC time series can be recognized in mode 2, which represents 18% of the total precipitation variation. In turn, the EOF map of the MDB is divided into two different regimes. While the northern part of the MDB has a positive sign, the south is negative. The PC time series shows the maxima at the beginning of each year, which coincides with the periods in which precipitation occurs in the north. In the central part of the MDB a transition zone (see section 4.1.1) can be identified, where values close to zero are present. With the application of the PCA to the precipitation observations, the detected phase difference between the two regimes of approximately half a year (MLRA) can also be verified. The minima of the PC time series occur in the middle of each year. Together with the negative values of the EOF map (red) the maxima (positive sign) for all grid points in the south are left, indicating that precipitation occurs mainly half a year later than in the northern part of the MDB.

The third and fourth mode of the precipitation observations together represent much less of the overall variability than for the GRACE-derived surface mass changes. The third mode of the TRMM observations shows small oscillations around zero in the PC time series. While these oscillations are (when compared to the GRACE-derived surface mass changes) larger, the contribution to the overall variability has to kept in mind. In the month of October 2004, an anomaly appears in the PC time series of the GRACE-derived surface mass changes, which can also be found in the PC time series of the precipitation observations. However, this anomaly is less distinct here. A larger negative anomaly appears in January 2009. In the EOF pattern there are two distinct zones separating the east from the west. The amplitudes of the EOF pattern are roughly half as large as the corresponding amplitudes for the GRACE-derived surface mass changes in the third mode (EOF pattern). Again, finding an explanation for the two different regimes is difficult. As mentioned in relation to the GRACE-derived surface mass changes, there may be a connection with the locations of the rivers, such as the Darling River, which connects the south-east

of the MDB with the north-east. The PC time series for the fourth mode shows a nearly constant value of zero with the exception of an occurring anomaly in February 2008. But note this mode only accounts for 1.7% of the overall variability. The same pattern as for the GRACE-derived surface mass changes appears in the fourth mode. But again note the two modes 3 and 4 together represent only 5% of the overall variability, whereas the first and second mode combined represent more than 90% of the variability. Due to the small percentage and the oscillation of the PC time series around zero, it can be assumed that mode 3 and 4, respectively, are comprised mainly of noise.

4.4 Analysis of in-situ hydrological observations

For the whole MDB in-situ hydrological observations in terms of water level and river flow are available. The measurements along the Murray river were received from the Murray-Darling Basin Authority, which is responsible for the water management of the MDB. The remaining data was acquired from the states New South Wales and Queensland. Figure 4.26 illustrates the spatial distribution in the MDB of both water level and river flow observations.



(a) water level (b) river flow

Figure 4.26: Spatial distribution of in-situ hydrological observations

The rivers of the MDB are quite well covered by water level and river flow observations. For 112 locations along the different rivers, water level observations are available and for 92 locations river flow observations are available. Since water level and river flow observations are very strong correlated, which will be shown in chapter 5.2, only the results in terms of the MLRA and the PCA for the water level observations are shown here. Hence, the water level observations, for which compared to the river flow observations more observations are available, represent the in-situ hydrological observations for this study. For the sake of completeness, all

results for river flow observations are provided in the appendix A. In contrast to the GRACE-derived surface mass changes and the precipitation observations observed from the TRMM mission, no filter techniques are applied to the in-situ hydrological observations, since the data is not uniformly distributed, e.g. not available on a grid. However, both data sets are available on a daily resolution. Therefore, monthly mean values were determined for both, water level and river flow observations by averaging.

4.4.1 Investigation by means of Multiple Linear Regression Analysis

For the analysis of water level observations in terms of MLRA, the same model approach M_1 (see equation 4.7) is applied. Again the three figures 4.27, 4.28, and 4.29 depict a graphical representation of how well the chosen model M_1 fits to the original time series. Again, three points are considered, representing the northern, the central, and the southern part of the MDB.

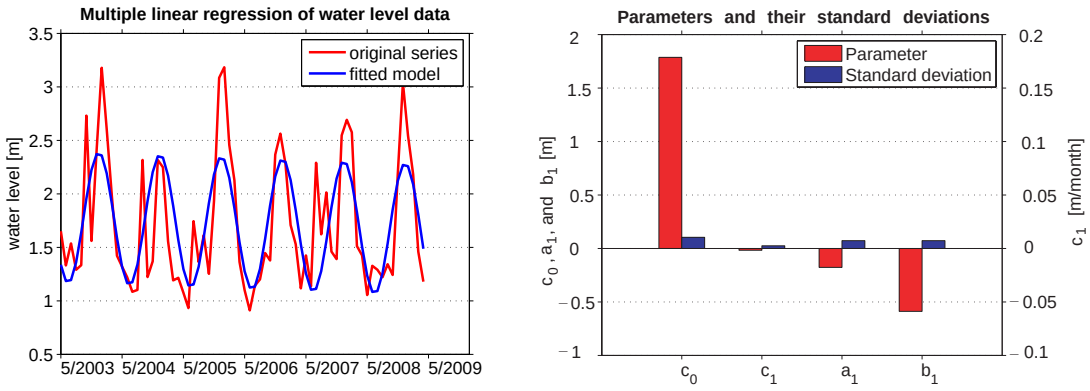


Figure 4.27: Multiple linear regression of water level observations at $\lambda = 149.874°$, $\varphi = -28.609°$ (Macintyre River)

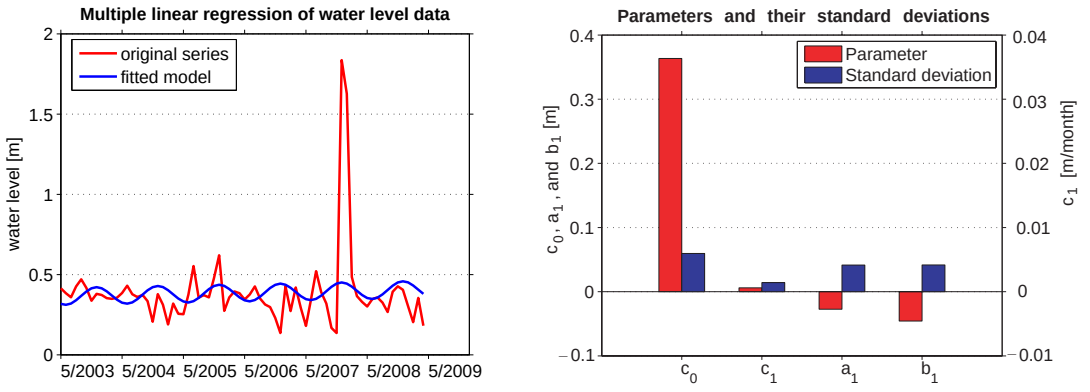


Figure 4.28: Multiple linear regression of water level observations at $\lambda = 147.052°$, $\varphi = -30.857°$ (Macquarie River)

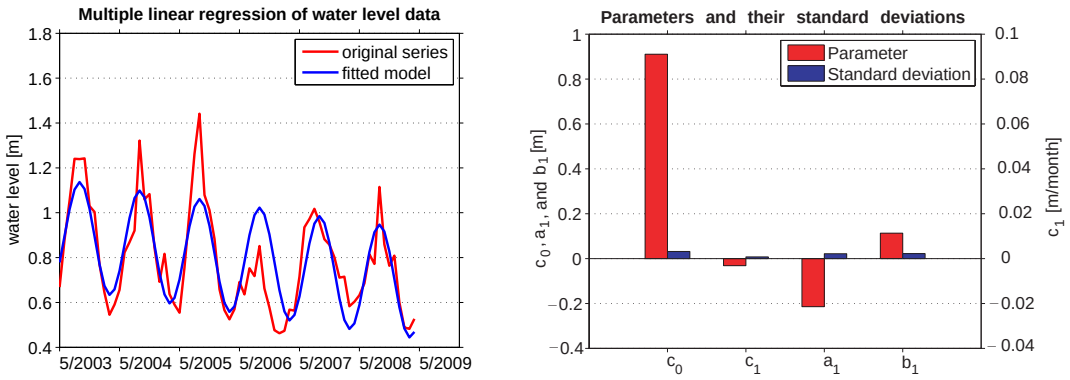


Figure 4.29: Multiple linear regression of water level observations at $\lambda = 147.606°$, $\varphi = -36.946°$ (Murray River)

Compared to the three figures of GRACE-derived surface mass changes (figure 4.5, 4.6, and 4.7) and precipitation observations from TRMM (figure 4.17, 4.18, and 4.19) the same results are obtained. However, the figures 4.17 and 4.19 (representing the northern and southern part of the MDB) show an even better fit to the original time series. Figure 4.28, representing the central part of the MDB, shows that the chosen model does not fit the original time series. This is not surprising, since in the original time series neither a trend nor an annual signal is identifiable. Another confirmation is found by looking at the parameters and their standard deviations (right picture of figure 4.28). The values of the standard deviations of the a_1 and b_1 parameters are nearly as large as their corresponding parameters. Consequently, the analysis of water level observations in terms of MLRA is another indication that the central part of the MDB is a climatic transition zone (see section 4.1.1). A large peak (see figure 4.28), appearing in early 2008, shows the end of the drought. In contrast, annual signals can be found in the north and the south of the MDB. All three illustrations show a very large value for the offset parameter c_0 compared to its corresponding standard deviation. Furthermore, the magnitudes of the trend parameter are again small. Following the same procedure as before it is studied in the next step whether the chosen model is suitable. Once again, with a significance test separately applied for the four regression parameters c_0, c_1, a_1, and b_1 is started. The results for the applied Student t-test (probability 97.5% (two-sided test), theoretical quantile: 1.996) for all 112 locations are represented in table 4.7.

parameter j	min $\lvert t_j \rvert$	max $\lvert t_j \rvert$	# significant	# not significant
c_0	1.230	1989.054	110 (98.2%)	2 (1.8%)
c_1	0.043	11.170	49 (43.8%)	63 (56.3%)
a_1	0.008	9.762	55 (49.1%)	57 (50.9%)
b_1	0.026	9.772	78 (69.6%)	34 (30.4%)

Table 4.7: Significance tests for the four different parameters of the MLRA model applied to the water level observations

The offset parameter c_0 as a part of the model is significant for nearly all of the 112 locations. The b_1 parameter is significant in almost 70% of the measurement locations, whereas a trend and the a_1 parameter are significant in about half of all locations. When compared to the results of the respective significance tests for the GRACE-derived surface mass changes and the precipitation observations, the results show that the model M_1 fits better to water level observations. This is shown by the proportion of locations where the M_1 model is significant.
For the sake of completeness, the results of the F-test for the investigation of the annual parts of a signal (probability 95% (one-sided test)) are given in table 4.8 (technique of reduced models). However, it is clear from the results of the t-test in table 4.7, that an annual part should be included.

case	$F_{0.95,2,n_{full}-u_{full}}$	# F significant	# F not significant
1 ($M_0 \rightarrow M_1$)	3.1317	111 (99.1%)	1 (0.9%)
2 ($M_1 \rightarrow M_2$)	3.1359	109 (97.3%)	3 (2.7%)
3 ($M_2 \rightarrow M_3$)	3.1404	107 (95.5%)	5 (4.5%)
4 ($M_3 \rightarrow M_4$)	3.1453	106 (94.6%)	6 (5.4%)

Table 4.8: Significance tests for the comparison of different multiple linear regression models (water level)

Table 4.8 shows that for nearly all water level observation locations (99%) an expansion of the M_0 model containing only offset and trend parameters to the model M_1 is suitable. This includes an annual part, in addition to the offset and trend parameter. The results for all further expansions to the models M_2, M_3, and M_4, show that even fewer points can be considered significant. Therefore, the model M_1 is again very suitable to model water level observations. Furthermore, it is an indication that water level observations are more periodic than GRACE-derived surface mass changes or precipitation observations, as the proportion of significant locations is much higher.
To study the 'goodness of fit' of the whole model M_1, the coefficient of determination R^2 and the adjusted coefficient of determination $\tilde{R}^2$, respectively, are determined. The minima and maxima of both coefficients are given in table 4.9.

	minimum	maximum
R^2	0.0278	0.6850
$\tilde{R}^2$	-0.0303	0.6662

Table 4.9: Coefficient of determination R^2 and $\tilde{R}^2$ (water level)

The minimum and maximum values of both coefficients are now much higher in comparison to the coefficients of determination of GRACE-derived surface mass changes and precipitation observations. Consequently, for some measurement points (19 of 112 points) the rule of thumb from Brook/Arnold (1985) ($R^2 > 0.5$) is fulfilled.

However, it is once again better to apply an one-sided F-test (null hypothesis: $\tilde{R}^2 = 0$, probability 95%) to determine the significance of the whole model. The theoretical quantile accounts for 2.509. The results of the Fisher test are illustrated in figure 4.30.

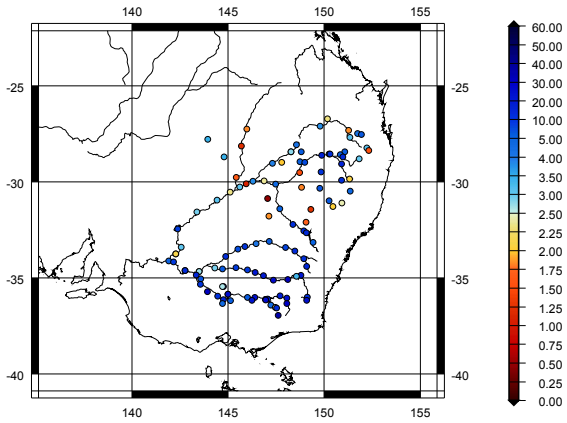
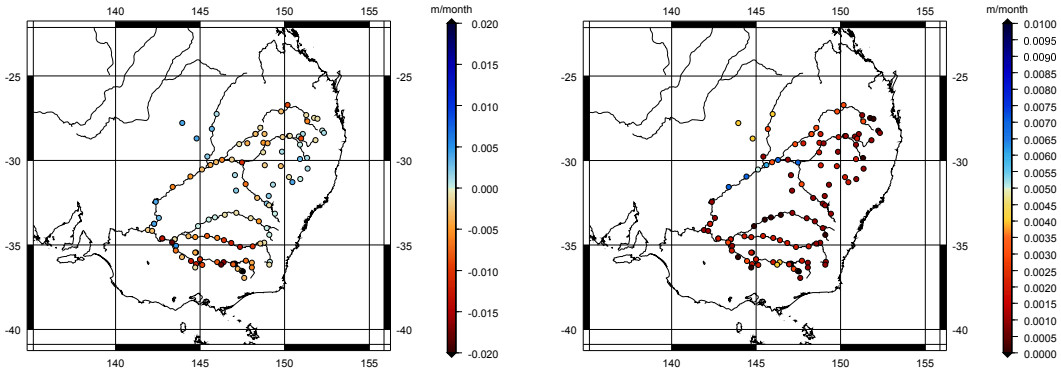


Figure 4.30: Significance test for the M_1 model (water level)
(red colours mean not significant, blue colours mean significant)

90 locations show the M_1 model as significant. This corresponds to 80% of all 112 points. Only in the central part of the MDB there are points where the water level observations show that the M_1 model is not suitable. Consequently, the Fisher test also confirms that the central part of the MDB is a transition zone (see section 4.1.1) between the north and the south of the MDB.

In the last part of this subsection, the trend parameter and the annual part of the M_1 model are examined in more detail. Figure 4.31 illustrates the trend parameter values together with their standard deviations for all 112 observation points.



(a) trend (b) standard deviation of trend

Figure 4.31: Trend together with its standard deviation for water level observations

The trend values and their corresponding standard deviation values are very low (cm/month). A negative trend can be found predominantly along the Murray River and Murrumbidgee River (in the south of the MDB). Along the Lachlan River (also in the south of the MDB) the trend values are close to zero. In the north of the

MDB, some trend parameter values have a small positive value. This is consistent with the results of the GRACE-derived surface mass changes and precipitation observations.

The results of the Student t-test for the trend parameter c_1 (all 112 points) are illustrated in figure 4.32.

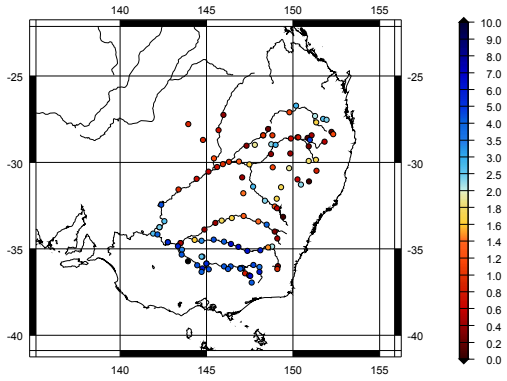
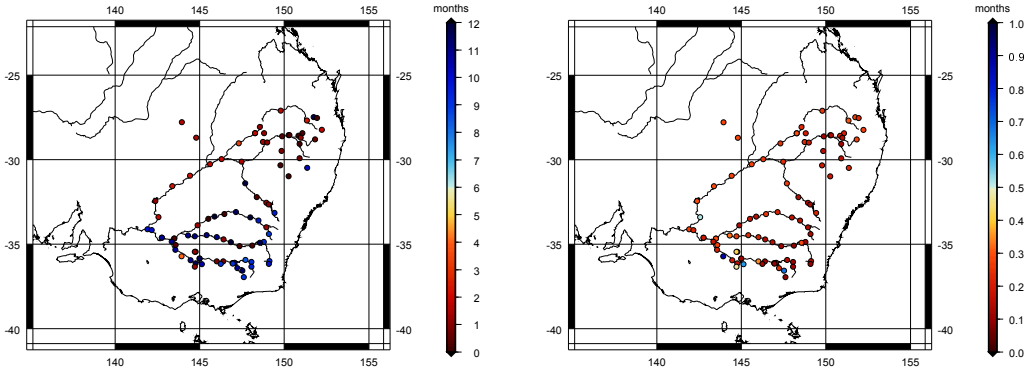


Figure 4.32: Significance test for the trend parameter c_1 (water level)
(red colours mean not significant, blue colours mean significant)

About half (49 of 112) of all points show a significant trend parameter. When reviewing figure 4.32 the south (Murray River and Murrumdbidgee River) is significant, whereas the north shows many non-significant trend parameters.

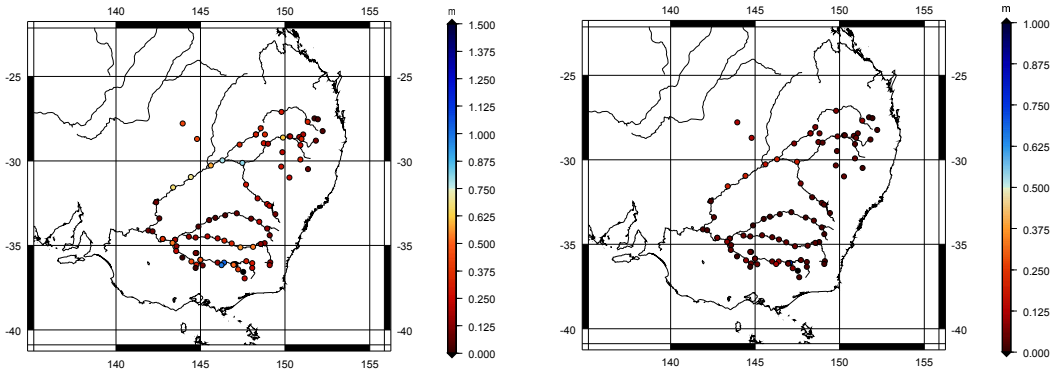
The annual part of the model is in turn described by the parameters a_1 and b_1. It is not necessary to illustrate the results of the F-test for the expansion from the model M_0 to M_1, as it will not provide further information (only one point along the Murray River is not significant). Phase and amplitude characterise the annual part of the M_1 model. Both are determined by equations 3.21 and 3.23 in section 3.1.2. The corresponding standard deviations are obtained via equations 3.22 and 3.24 in section 3.1.2. In figure 4.33 the phases with their corresponding standard deviations are illustrated but only for points where the model M_1 is significant (90 of 112 points).



(a) phase (b) standard deviation of phase

Figure 4.33: Phase together with its standard deviation for water level observations
(phase is defined with respect to mid-January)

For the northern part (Darling River and confluents) a mean value for the phase of approximately 0.6 months is determined, which corresponds exactly with the phase for the north of the precipitation observations. Consequently, water level and precipitation observations have a high relationship in the north. In contrast, in the south parts of the MDB - along the Murray River, Murrumbidgee River, and Lachlan River - the mean phase accounts for approximately 10.7 months. This value clearly differs from the mean phases in the south of the precipitation observations. For the precipitation observations, a mean value of approximately 7.6 months was determined. A reason for the deviation of three months might be human operations. For example, a number of dams are located along the Murray River. Additonally, water is extracted mostly along the southern rivers for irrigation. But this assumption is refuted by the phase values in the south of the GRACE-derived surface mass changes. A difference of four months to the water level observations appears. Since the GRACE mission detects all mass changes, not only changes caused by precipitation like the TRMM mission, mass changes due to human operations along the rivers should be visible in the detected surface mass changes from the GRACE mission. Therefore, the reason for the large deviations of three and four months, respectively, remains unclear at this point. However, a possible explaination is given in section 5.4. The mean value of the standard deviation accounts for 0.2 months. Next the amplitude of the annual part of the M_1 model is analysed. Figure 4.34 represents the amplitudes with their corresponding standard deviations. These are also shown only at significant points of the M_1 model.



(a) amplitude (b) standard deviation of amplitude

Figure 4.34: Amplitude together with its standard deviation for water level observations

Almost the entire MDB shows the same values of the amplitude with an average of approximately 0.4 m. For the standard deviation of the amplitude values, all points show nearly the same value as well. The mean value of the standard deviation of the amplitude accounts for 7 mm, which is approximately 6 times smaller than the mean value of the amplitude. Therefore, both, phase and amplitude, were determined with a relative high accuracy.

4.4.2 Investigation by means of Principal Component Analysis

First, mean-centred values (mean taken over the 6-year time period) for the water level observations are calculated. The mean-centred values are ordered once again in the matrix Z (see equation 3.29) and the PCA is applied. In figure 4.35 the relative amount of the overall variability for a particular mode, together with its mode number is illustrated.

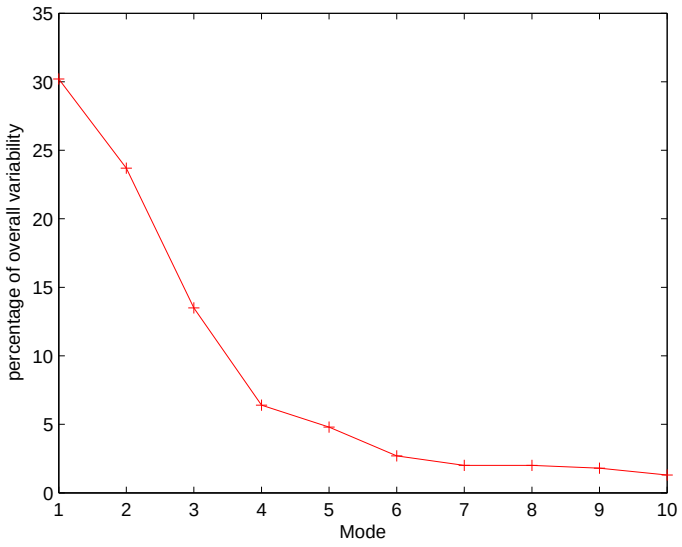
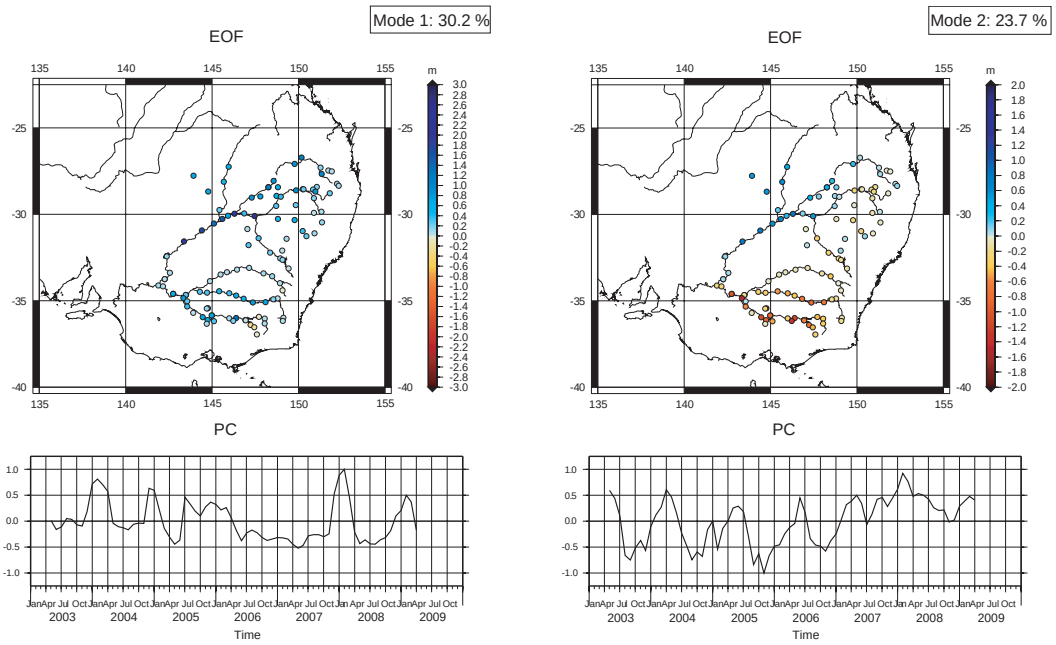
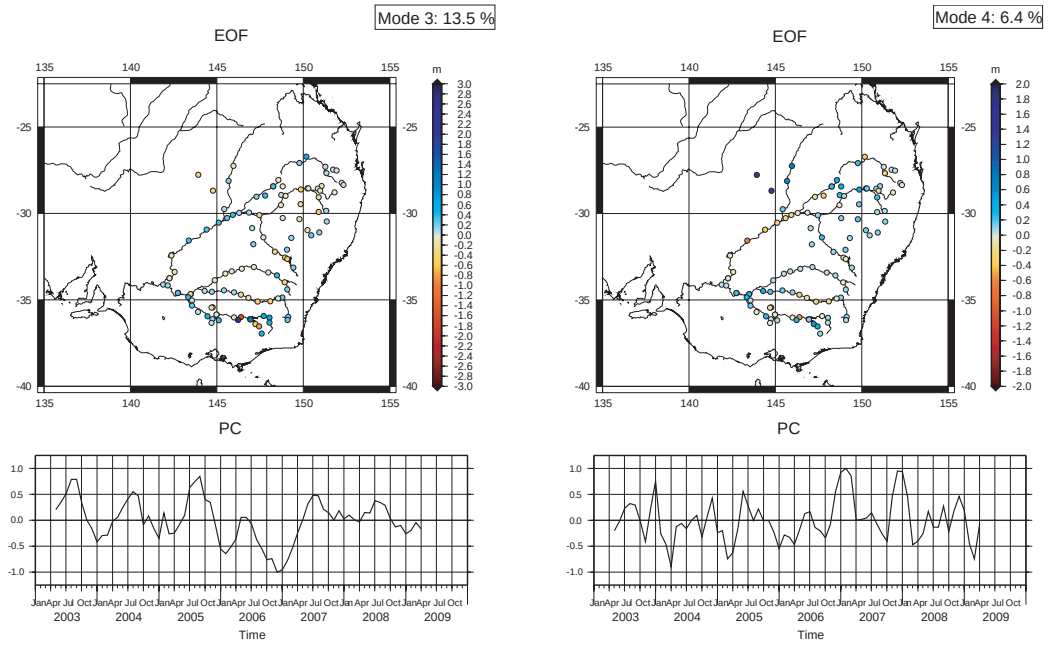


Figure 4.35: Percentage of overall variability (water level)

The first mode covers only 30% of the overall variability of the water level observations. This percentage differs only little from the contribution of the second mode, which accounts for approximately 24%. Consequently, the gradient between the first mode and second mode is relatively flat, when compared to the figures 4.15 and 4.24 (GRACE and TRMM). The 'elbow' of the curve can be found at mode 4, where the gradient becomes much lower. Therefore, it can be assumed that all further modes represent mainly noise. The first four modes combined represent nearly 75% of the overall variability. This means that with the first four modes, 75% of the original variablity of the signal can be described, which is a much lower value in comparison to the first four modes from GRACE-derived surface mass changes and precipitation (94% and 98%, respectively). The first four modes are illustrated in figure 4.36.



(a) Mode 1 (b) Mode 2



(c) Mode 3 (d) Mode 4

Figure 4.36: PCA modes 1-4 of water level observations

In contrast to mode 1 of the GRACE-derived surface mass changes and precipitation observations, the PC time series of the first mode of water level observations shows a different behaviour. Parts of an annual signal in the PC time series can be recognized with the annual peaks always appearing at the beginning of a year (January/ February). The annual signal is interrupted by the 2006/07 drought. During the period from the middle of 2006 until beginning of 2008 no annual signal appears. Note that as described above, the drought period ranged from May 2006 until October 2007. Consequently, the drought can be confirmed by mode 1. In February 2008 a large peak appears in the PC time series, which can also be found in mode 1 of the precipitation observations. This large peak in February 2008 indicates that after the end of the drought, unusually high water level were present, which were likely the result from unusually high rainfall (see mode 1 of precipitation observations in figure 4.25 (section 4.3.2)). This water likely came from the north, along the Darling River, since the EOF map shows larger values along that river. This can be an indication that in the north of the MDB the drought was not as severe as in the south in terms of available water. The EOF map of mode 1 shows an almost uniform pattern, which coincides with the EOF map of the GRACE-derived surface mass changes and precipitation observations. However, here mode 1 accounts only for 30% of the overall variability.

Mode 2 has a similar contribution (24%) to the overall variability as mode 1. Two regimes can also be recognized in the EOF map. While the northern part of the Darling River has measurement values, which are positive (blue) ranging from approximately 0.2 m to 1.0 m, the southern part of the MDB including the Murray River and Murrumbidgee River has negative values (red) ranging from approximately -0.2 to -1.8 m. The Lachlan River and the Macquarie River (located in the central part of the MDB) have values close to zero, indicating the transition zone between the north and the south (see section 4.1.1). For the locality of the different rivers in the MDB refer to figure 4.1 at the beginning of this chapter. In the PC time series of mode 2 an annual signal can also be identified, which is mainly dominated by the southern part of the MDB (the largest values in the EOF map appear along the southern rivers). The minima of the annual signal appear in the middle of each year. Together with the negative values in the south again maxima are obtained. Maxima values in the middle of the year coincide with the occurrence of precipitation in the south. A water mass deficit for the south of the MDB can then be recognized from February 2007 onwards. During this period, the north of the MDB has positive values, indicating that mainly the south was affected by the drought. Mode 1 and mode 2 together represent 54% of the overall variability. This value corresponds approximately with mode 1 of the GRACE-derived surface mass changes. Consequently, drought and annual signal are not separately found in the first two modes. Moreover, they are mixed in both mode 1 and mode 2 of the water

level observations. In contrast, mode 1 of GRACE-derived surface mass changes and precipitation showed mostly the drought, whereas the second mode illustrated more annual parts of a signal.

The third mode accounts for 14% of the overall variability of the signal. All points in the EOF map show very small values (light colours). Values of zero are found along the Lachlan river, which is located in the central part of the MDB. This might be another indication that the central part of the MDB is a climatic transition zone (see section 4.1.1). Two points (dark red and black) along the Murray River attract attention. Their colours differ clearly from their surrounding points. The explaination for that difference is very simple, in that the points are located at two dams. Therefore, the water levels are much larger than others, which is reflected in the EOF pattern. When analysing the PC time series the drought can immediately recognized. From July 2006 until May 2007 negative values appear. After November 2007 there is neither a water mass surplus nor a water mass deficit observable, which show the zero values of the PC time series. This situation does not improve until April 2008. The minimum appears in December 2006. Note the corresponding minimum appeared in October 2006 for both GRACE-derived surface mass changes and precipitation observations. From the beginning of the time series until the beginning of the drought (May 2006) an annual signal can be recognized, whereas the amplitude in May 2006 is already much lower (close to zero) than the years before. Consequently, the signatures of the drought and annual signals are not only found in mode 1 and mode 2, but also in mode 3.

Finally, mode 4 describes 6.4% of the total variability. With the exception of the points located along the Darling River and a few points located along the Murrumbidgee River, all points show nearly the same amplitudes (positiv (light blue)) in the EOF map. There are parts of an annual signal visible in the PC time series with mostly maxima at the beginning of each year. But some maxima values are also found in the middle of the year. This makes an accurate interpretation very difficult. In mode 1 and also in mode 4 of the precipitation observations, a peak was found in February 2008. This peak appears in the fourth mode of the water level observations 1-2 months earlier.

Altogether the PCA of water level observations shows a difference to the PCA of GRACE-derived surface mass changes and precipitation data. While in the first mode of GRACE-derived surface mass changes and precipitation data the drought is mainly found, the signature of the drought is in the PCA of water level observations divided into the first three modes. Furthermore, the annual part can be found almost in every mode, whilst it is mainly found in the second mode of the GRACE-derived surface mass changes and precipitation observations.

5 Monitoring mass transport

The aim of this chapter is to answer the key question, being whether it is possible to monitor mass transport in the MDB based on GRACE-derived surface mass changes and precipitation observations. In order to answer this question the different data sets are compared. Furthermore, some results, which were derived in chapter 4, are briefly summarized.

5.1 Surface mass changes vs. precipitation

The first study focuses on GRACE-derived surface mass changes and precipitation and tries to establish whether there are similarities and differences. To determine the relationship between both data sets, linear correlation coefficients are assessed. This is obtained by assessing the correlation coefficient between the time series of GRACE-derived surface mass changes and precipitation for every grid point. The results for the correlation coefficients are illustrated in figure 5.1.

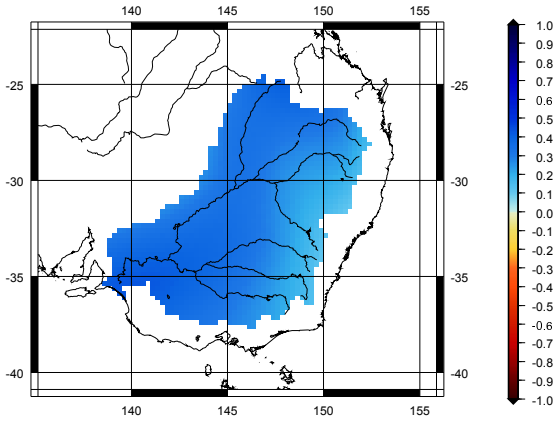


Figure 5.1: Correlation coefficients of GRACE-derived surface mass changes and precipitation

It can be recognized that for the whole MDB very similar correlation coefficients (around 0.3) are present. Only in the east the correlation coefficients have slightly smaller values. The mean value accounts for 0.3, which indicates a rather low correlation. The largest correlation coefficients appear in the south-west of the MDB, where they reach values of about 0.4. Consequently, it can be assumed that mass

changes are not only caused by the occurrence of precipitation. There must be some other processes involved such as mass transport from one region to another, which cause mass changes without precipitation.

GRACE observes the sum of all mass changes, and is not able to differentiate between the different processes, which might have caused the mass change. Furthermore, a significance test for the correlation coefficients $r_i \ \forall i = 1, ..., 1616$ of each grid point is applied in order to confirm whether a relationship can be significantly detected or not. The null hypothesis $H_0 : r = 0$ means that no significant relationship between the two time series can be ascertained. The test variable follows a Student t-test and is determined by

$$t_r = r\sqrt{\frac{n-2}{1-r^2}}, \tag{5.1}$$

where the parameter n denotes the number of observations. In this case, it is the number of months of the 6 year time period. Since June 2003 and January 2004 are missing in the GRACE observations, the number of observations is 70 instead of 72. The two-sided test is performed with a probability of 97.5%. The theoretical quantile accounts for $T_{n-2} = T_{68} = 1.996$. The results of the applied Student t-test are illustrated in figure 5.2.

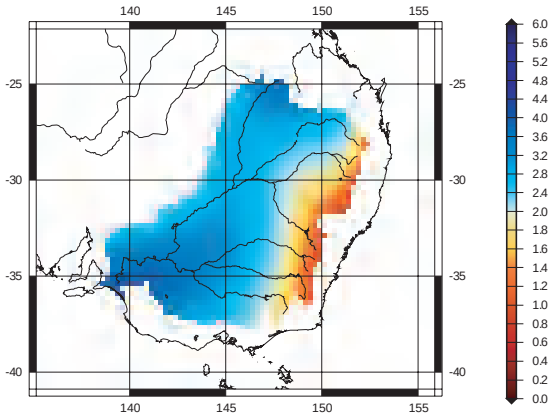


Figure 5.2: Significance test for correlation coefficients of GRACE-derived surface mass changes and precipitation (red colours mean not significant, blue colours mean significant)

A small band in the east shows that the correlation coefficients are not significant, which coincides with the smaller correlation coefficients in figure 5.1. All other parts of the MDB show significant correlation coefficients. Note the test variable is not much larger than the theoretical quantile, as the values of the correlation coefficients are rather low.

Next, it will be examined whether the results of MLRA show similar or different results for the GRACE-derived surface mass changes and for the precipitation observations. The MLRA results from the GRACE-derived surface mass changes showed

that a significant trend could be found for just over half of all 1616 grid points, whereas a trend was not significant for the precipitation observations. For the GRACE-derived surface mass changes the trend had values of maximal 0.5 mm/month in the north, while in the south values reached up to -0.8 mm/month suggesting that the south was loosing mass whereas the north was gaining mass over the time period considered (see figure 4.10 in section 4.2.2). Therefore, it seems that the south was more affected by the drought than the north.
Generally, precipitation occurs over a shorter time period than surface mass changes. Surface mass changes caused by precipitation cannot detect the precipitation at once, since the water from precipitation evaporates and runs off into rivers and lakes. A small portion of precipitation accumulates in the region. Therefore, the period of detecting surface mass changes caused by for example precipitation is longer, as part of the water stays longer in the region than the time period of precipitation. Consequently, the determined phase from MLRA between GRACE-derived surface mass changes and precipitation are different. In figure 5.3 the phase differences between GRACE and TRMM ($\Phi_{GRACE} - \Phi_{TRMM}$) are illustrated. Note the phase difference is only plotted for grid points where the model M_1 (see equation 4.7 in section 4.2.2) is significant for both, GRACE-derived surface mass changes and precipitation.

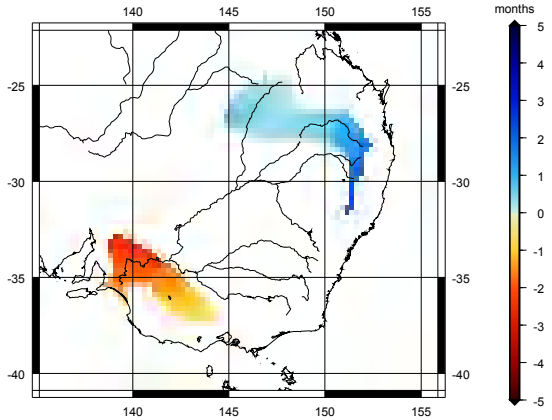


Figure 5.3: Significant phase difference between GRACE-derived surface mass changes and precipitation

Since the model M_1 is significant mainly in the southern part of the MDB for GRACE-derived surface mass changes, and is mainly significant in the northern part of the MDB for the precipitation, the area, which is not considered significant for both data types, is rather large. Nevertheless, in the north, precipitation appears approximately one month before GRACE detects mass changes. In the southern part of the MDB an inverse behaviour is visible. The GRACE-derived surface mass changes occur one month before precipitation is detected by the TRMM mission. The explanation for the north of the MDB is very simple, and is merely that a pre-

cipitation event has occured. It then takes about one month until the precipitation and water mass can be detected by the GRACE mission. For the north and for the south, mean values for the phases were determined both for GRACE-derived surface mass changes and precipitation. The mean phases and differences are represented in table 5.1.

phase [months]	GRACE	TRMM	difference [months]
north	1.6	0.7	0.9
south	6.6	7.6	-1.0

Table 5.1: Mean phases for the north and south (GRACE - TRMM)
(phase is defined with respect to mid-January)

Since the phase is defined here with respect to a cosine curve with its maximum at the beginning of each year, cosine curves for the north and for the south are plotted in figure 5.4 in order to compare the detected phases between the TRMM and GRACE mission.

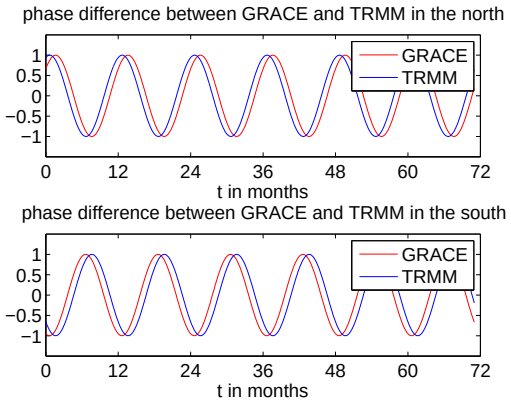


Figure 5.4: Phases illustrated in a cosine curve for the north and the south of the MDB
(GRACE - TRMM)

This illustration again shows a positive phase difference in the north and a negative phase difference in the south between both data types. For the south it seems unreasonable that mass is detected just before precipitation. But what is the reason for the negative phase difference in the south? The MDB is mainly influenced by the precipitation events in the north, which take place at the beginning of each year. The remaining parts of the water mass, which does not evaporates or soak into the ground, run-off along the different rivers to the south, especially along the Darling River and its northern confluents, which largely flow from the north-east to the south-west. Precipitation in the south mainly occurs in the middle of each year and can be seen at the phase value of precipitation in the south (7.6 months). However, the detected mass changes by GRACE appear one month earlier ($\Phi^{south}_{GRACE} = 6.6$ months). Therefore, it may be assumed that the run-off from the north takes approximately five months as indicated by the phase values in the

north and the south from the GRACE-derived surface mass changes ($\Phi^{north}_{GRACE} = 1.6$ months and $\Phi^{south}_{GRACE} = 6.6$ months, respectively) to pass the MDB from the north to the south. In order to find a verification or a confutation of this assumption, the in-situ hydrological observations are analysed in terms of cross correlation in section 5.4 later. By means of cross correlations the time the water mass needs to flow from the north to south will be determined.

The amplitudes of the annual signal (MLRA) of the GRACE-derived surface mass changes and precipitation show large similarities. While in the north amplitudes are found up to 30 mm, the amplitudes in the south accounts for almost half of the values from the north (see figure 4.14 in section 4.2.2 and figure 4.23 in section 4.3.1).

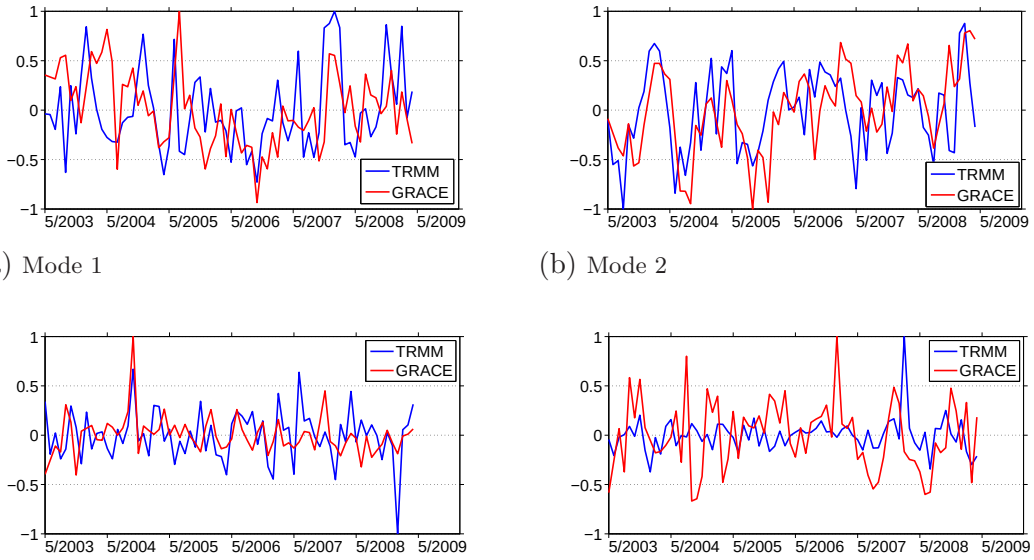
Now the similarities and differences of the results from PCA are analysed. Table 5.2 shows again the percentage of overall variability for the first four modes.

	percentage TRMM	percentage GRACE
Mode 1	75.5%	51.3%
Mode 2	17.7%	26.8%
Mode 3	3.1%	10.1%
Mode 4	1.7%	5.5%
$\sum$	98.0%	93.7%

Table 5.2: Percentage of overall variability for GRACE and TRMM

The cumulative sum of the percentage of overall variability for all four modes is for both, GRACE-derived surface mass changes and precipitation, nearly the same, with the results from the TRMM mission showing approximately 4% more of the percentage of total variability of the signal. It is noticeable that the percentage of the first mode of the precipitation observations is 24% higher than the value for the GRACE-derived surface mass changes. As shown in section 4.3.2 this indicates that the TRMM mission detects more signal from the drought period than the GRACE mission. As described in section 4.2.3 and 4.3.2, the first mode shows a long period of drought, both through precipitation observations and the GRACE-derived surface mass changes. For both no annual signal was found in the first mode. Furthermore, the EOF pattern shows a uniform pattern for the whole MDB, which indicates that the whole area was affected by the drought period (see mode 1 in figure 4.16 in section 4.2.3 and figure 4.25 in section 4.3.2). In contrast, the second mode shows for both an annual signal in the PC time series. A phase difference of half a year for the two different regimes in the north and the south was discovered. The PC time series is mainly dominated by the northern part of the MDB (see mode 2 in figure 4.16 in section 4.2.3 and figure 4.25 in section 4.3.2). While the first two modes describe more than 93% of the signal variability for the TRMM observations, the percentage for the GRACE-derived surface mass changes is approximately 15%

lower. All further modes describe more or less noise and/or higher frequency variations in both cases. Therefore, it can be assumed that the GRACE-derived surface mass changes contain more noise and/or higher frequency variations than does the precipitation observations. In figure 5.5, the PC time series for the first four modes are plotted.



(a) Mode 1 (b) Mode 2

(c) Mode 3 (d) Mode 4

Figure 5.5: Comparison of PC time series of precipitation observations and GRACE-derived surface mass changes (mode 1 - mode 4)

At first view, the PC time series from GRACE and TRMM show the same curves. However, it is noticeable that the amplitudes of the PC time series of the GRACE-derived surface mass changes are often higher than those of the precipitation observations. This together with almost the same amplitudes in the EOF map for both data sets (but only for mode 1, 2 and 4) denotes that GRACE detects often more mass changes, which are not only caused by precipitation. For example the absolute minimum in the first mode (October 2006) is larger in the GRACE-derived surface mass changes. But note this is only the case for the modes 1, 2 and 4, since in mode 3 the amplitudes of the EOF map are different. Furthermore, it must be considered that the percentage of overall variability of GRACE-derived surface mass changes and precipitation is different for these modes. In the second mode, which includes the annual signal, a phase difference between GRACE and TRMM is detectable. As described above, both PC time series of mode 2 (TRMM and GRACE) are dominated by the northern part of the MDB. The PC time series of GRACE-derived surface mass changes lags behind approximately 1-2 months. This fact coincides with the results of MLRA where a phase difference of one month between the precipitation and GRACE-derived surface mass changes was detected. Both show an anomaly in October 2004 in mode 3. A further anomaly is visible in mode 4 for both data types. However, the anomaly in the GRACE-derived surface

mass changes appears approximately 3 months earlier (GRACE: November 2007, TRMM: February 2008). Consequently, this mode may be dominated by the south. Note this is only a very disputable speculation, since the percentages of variability are very low for the modes 4. Furthermore, the mode 4 of TRMM precipitation shows mainly noise with the exception of that anomaly.

After the first look at the PC time series the similarities and differences between the particular modes of the GRACE-derived surface mass changes and precipitation observations are investigated on a more scientific level below. Cross correlation is applied, which was explained in chapter 3.3. The correlation coefficient r as well as the phase difference for all four modes are determined. Furthermore, a significance test is applied to decide whether the calculated correlation coefficient is significantly different from zero. The theoretical quantile accounts for $T_{n-2} = T_{68} = 1.996$. The two-sided hypothesis test (equation 5.1) is performed with a probability of 97.5%. In table 5.3 the results are given.

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.217	0.457	0.261	0.058
Test variable t	0.479	4.243	2.233	0.479
Test decision	not significant	significant	significant	not significant
Phase difference $\Delta\Phi$	-1 month	-1 month	+ 5 months	+1 month

Table 5.3: Correlation coefficients together with the phase difference for the PC times series of the first four modes (GRACE - TRMM)

With the exception of mode 2 all other modes show a relative low correlation coefficient. The reason for the low coefficient in mode 1 might be found in the percentage of the total variability, since the precipitation observations account for 24% more variability compared to the mode 1 of GRACE-derived surface mass changes. Parts of the variability contained in TRMM mode 1 (75.5%) might be split amongst mode 1 and 2 for GRACE-derived surface mass changes, since the cumulative percentage of mode 1 and 2 accounts for 78.1%. The significance test shows that the correlation coefficient of 0.217 for the first mode is not significant. Note the values of the phase differences are related to the precipitation observations (e.g. $\Phi_{TRMM} - \Phi_{GRACE}$). For example a negative value means that the PC time series of GRACE lags behind the PC time series of TRMM and a positive value indicates the inverse case. The phase difference -1 months for mode 1 shows that GRACE lags one month behind TRMM. But note the test showed that the correlation coefficient is not significant. Therefore, the so derived phase difference should be dealt with care.

The significant correlation coefficients of mode 2 accounts for 0.46. The almost similar percentage of overall variability for GRACE and TRMM confirms the higher correlation coefficient compared to mode 1. In the preceding chapter it was ascertained that the PC time series for mode 2 is dominated by the northern part of the

MDB for both precipitation and GRACE-derived surface mass changes. The phase difference of -1 month denotes that the GRACE-derived surface mass changes lag one month behind the precipitation observations. This is exactly the same result, which was received from the phases from MLRA (see beginning of this chapter). Consequently, mode 2 of PCA coincides with the MLRA results for GRACE-derived surface changes and precipitation observations.
The third mode also shows a significant correlation coefficient of 0.261. But note the test variable has nearly the same magnitude as the theoretical quantile. Furthermore, the PC time series especially of GRACE-derived surface mass changes shows oscillating values around zero. Indeed the values of the precipitation observations have a little bit larger oscillation around zero. But together with the low percentage for both GRACE and TRMM it must be assumed that this mode represents mainly noise. Therefore, the determined phase difference of +5 months is a more or less random value and should be treated with care.
Finally, the fourth mode shows the lowest correlation coefficient of 0.06, which is of course not significant. The phase difference delivers a value of +1 months, indicating that TRMM lags one month behind GRACE. Therefore, the very disputable speculation from above can be "confirmed" that the PC time series of the fourth mode is dominated by the southern part of the MDB. But again since the correlation coefficient is not significant and the contribution of the fourth mode is very low for both data types (see table 5.2) it also can be assumed that the fourth mode represents mainly noise. Therefore, the phase difference of one month must be dealt with care as well.
Besides the PC time series the correlation coefficients of the EOF pattern were also determined. Hence, the spatial distribution of the different modes is compared. Furthermore, a significance test for the correlation coefficients is applied. The test variable accounts for $T_{n-2} = T_{1614} = 1.961$ (probability 97.5% (two-sided test). The results are represented in table 5.4.

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.772	0.989	0.884	0.825
Test variable t	48.805	266.656	75.802	58.651
Test decision	significant	significant	significant	significant

Table 5.4: Correlation coefficients for the EOF map of the first four modes (GRACE - TRMM)

The correlation coefficients are very high, which can be confirmed by visual inspection of the EOF maps (see figure 4.16 in section 4.2.3 and figure 4.25 in section 4.3.2). The EOF map for a particular mode for GRACE and TRMM show the same pattern. As can be expected all correlation coefficients are significant. Mode 2 even shows a correlation coefficient of almost 1. Consequently, a much stronger

relationship for the spatial distribution of the different modes than for the temporal behaviour exist. Knowing this information the magnitudes of the PC time series, but only for modes 1, 2 and 4, can indeed be compared between GRACE-derived surface mass changes and precipitation.
Altogether the second mode representing the annual signal shows the largest accordance between TRMM and GRACE. This can be confirmed by nearly the same values of percentages for GRACE-derived surface mass changes and precipitation observations.

5.2 Precipitation vs. in-situ hydrological observations

Before studying the relationships between precipitation and in-situ hydrological observations the relationship between the in-situ hydrological observations of water level and river flow is scrutinized. In section 4.4 it was established that the water level and river flow have a very strong relationship. Therefore, only the results of the water level observations were presented. To prove this relationship between both data types Correlation Analysis between water level and river flow observations is also applied. Since both measurement types are not always taken at the same locations, a grid by five degree elements is formed for the MDB (see figure 5.6). The MDB area covers the grid elements A2, A3, B1, B2, B3, C1, C2, and C3. A mean value of all observations points (water level, river flow), which are within one grid element, is determined for every month of the time period considered. After that, correlation coefficients are calculated for the time series of every grid element. The division in such a grid by five degree elements is illustrated in figure 5.6.

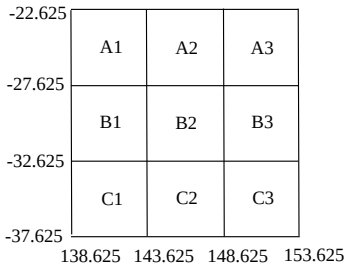


Figure 5.6: Grid by 5 degree elements for Correlation Analysis

The correlation coefficients are illustrated in figure 5.7. Additionally, on the right picture of figure 5.7 the correlation coefficient for each grid element is stated.

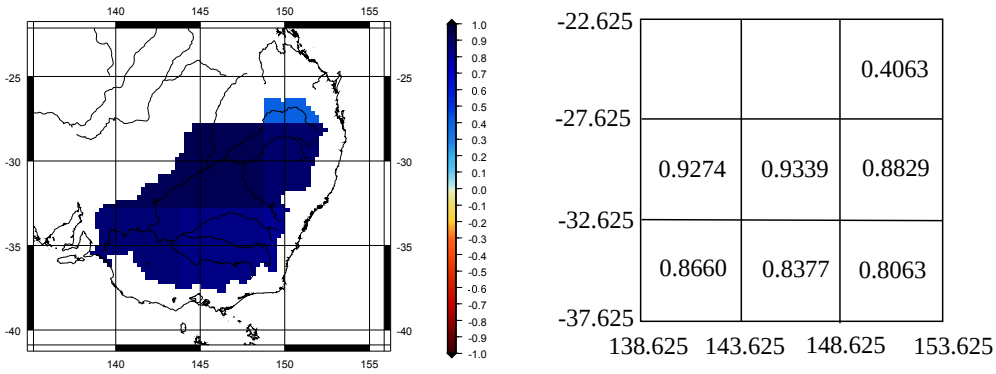


Figure 5.7: Correlation coefficients for water level and river flow observations

Unfortunately, for the grid element A2 no river flow observations are available. Therefore, no Correlation Analysis can be performed for this grid element. All grid elements show a very high correlation coefficient. Only grid element A3 is an exception, where the correlation coefficient is relative low compared to the other grid elements. The reason for that is very simple. In this grid element only two river flow observations are located. All other grid elements show a value of 0.8 and higher. In the central part of the MDB the correlation coefficients are even higher, at approxiamtely 0.9. A significance test reveals of course that all correlation coefficients (even for the grid element A3) are significant. The representation of the results of a hypothesis test is not considered here, since the significance of the correlation coefficients is obvious by looking at figure 5.7. Consequently, the water level observations suffice as a representation for the in-situ hydrological observations.

After this short intermezzo the similarities and differences between the precipitation and in-situ hydrological observations in the form of water level observations are now analysed. As already mentioned the precipitation observations were filtered with a smoothing radius of 500 km, which corresponds to a resolution of approximately five degree at the equator. Therefore, a Correlation Analysis between both data sets for a grid of five degree by five degree resolution is reasonable. Consequently, mean values of the precipitation and water level observations for all eight grid elements (A2, A3, B1, B2, B3, C1, C2, and C3) are determined. As with the water level and river flow observations, a correlation coefficient (for each of the eight grid elements) between the time series of precipitation and water level is determined. Furthermore, a two-sided significance test is applied in order to determine whether the coefficients are significantly different from zero or not. The theoretical quantile accounts for $T_{72-2} = T_{70} = 1.994$ with a probability of 97.5%. The correlation coefficients together with the results of the significance test are illustrated in figures 5.8 and 5.9.

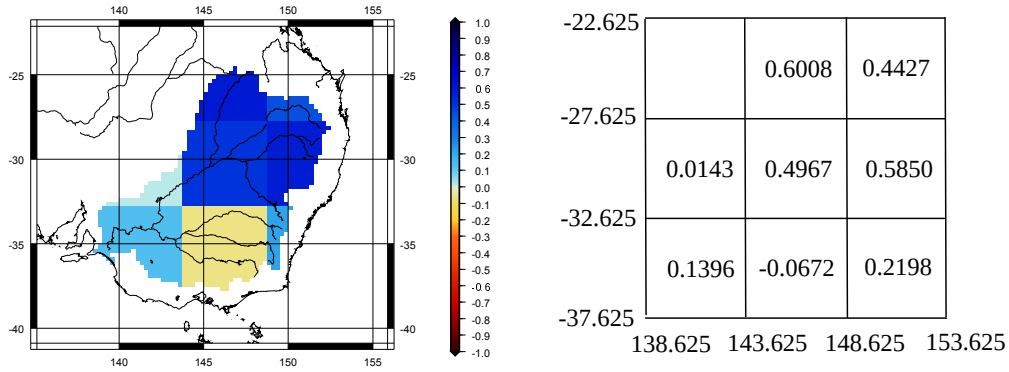


Figure 5.8: Correlation coefficients for precipitation and water level observations

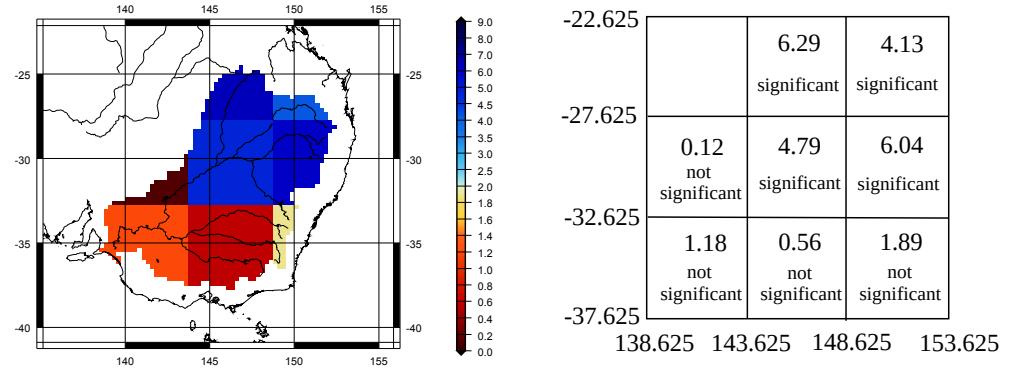


Figure 5.9: Significance test for correlation coefficients of precipitation and water level observations (red colours mean not significant, blue colours mean significant)

Figure 5.8 shows that the highest correlation coefficients between precipitation and water level observations are found in the central and the northern part of the MDB. Here the correlation coefficients vary between 0.4 and 0.6. The southern rivers, namely Murray River, Murrumbidgee River, and Lachlan River (for the locations of the rivers see figure 4.1), which are located in the grid elements C1, C2, and C3 show low correlation coefficients between 0 and 0.25. A possible explanation for these low values could be human operations along the southern rivers. As described above a few dams are found in the southern part of the MDB. In addition to precipitation, the southern rivers receive also run-off water from the northern rivers. Since the TRMM mission only detects precipitation, this could be an explanation why the correlation coefficients in the north are much larger compared to the south, which may also be affected by mass transport. Similarly, a possible explanation for the low correlation coefficient in grid element B1 (southern part of the Darling River) could be that the water level is composed of the run-off water mass from the upper Darling River and of precipitation in the south. Furthermore, the relatively low correlation coefficients show that there are probably other processes involved besides precipitation (for example mass transport), which cause water mass changes. In the following the results from MLRA and PCA are analysed. As already descri-

bed above, a significant trend parameter could not be found for the precipitation product (see section 4.3.1). For the southern rivers (Murray River and Murrumbidgee River), a negative trend (a few centimetres/month) was exhibited for water level observations. The northern parts showed values just above zero, indicating that the northern part of the MDB was less affected by the drought than the south (see figure 4.31 in section 4.4.1).

Both, precipitation and water level observations, showed two different regimes (north and south), where different phases were determined by means of MLRA (see figure 4.22 in section 4.3.1 and figure 4.33 in section 4.4.1). To compare the different phases a mean value for the north and the south both for precipitation and for water level observations is determined. The mean values are provided in table 5.5. Again cosine curves for the north and for the south are plotted (see figure 5.10) in order to compare the detected phases between the precipitation and water level observations.

phase [months]	TRMM	water level	difference [months]
north	0.7	0.6	0.1
south	7.6	10.7	-3.1

Table 5.5: Mean phases for the north and south of the MDB (TRMM - water level) (phase is defined with respect to mid-January)

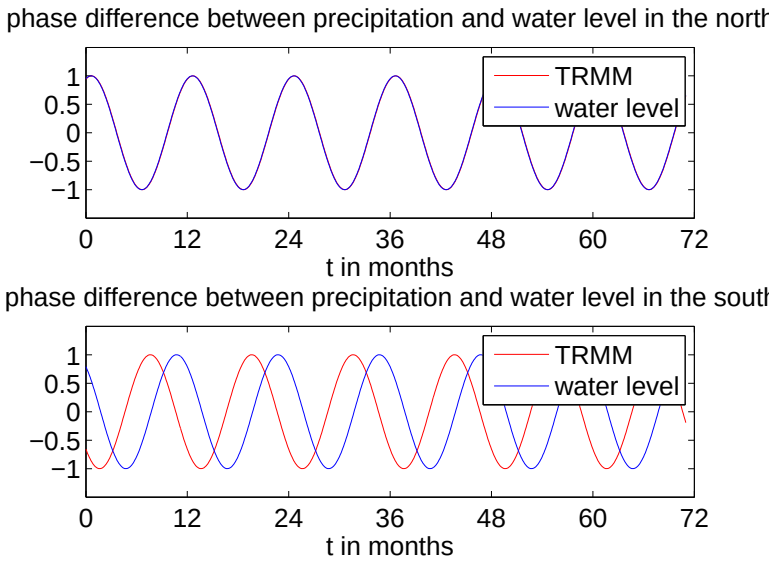


Figure 5.10: Phases from precipitation and water level illustrated in a cosine curve for the north and the south of the MDB

In the north of the MDB both precipitation and water level observations show almost exactly the same phase with the maximum at the beginning of each year, whereas in the south the maximum in the cosine curve of the water level observations appears in comparison with the maximum in the cosine curve of the precipitation

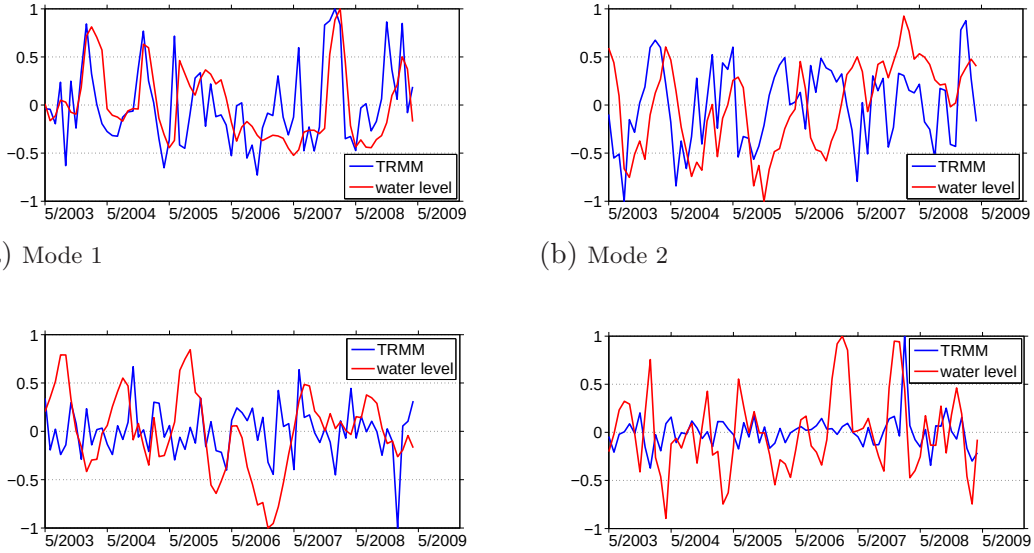
observations approximately three months later. The deviation of three months from the water level observations might be attributable to human operations such as water regulations by dams and water extraction for irrigation along the southern rivers (e.g. holding water longer in the region).
The annual amplitudes (from MLRA) of both data types show big differences. While the amplitudes of the precipitation observations in the south show only half of their values of the north (up to 30 mm) (see figure 4.23 in section 4.3.1), the amplitudes of the water level observations indicate nearly the same values (mean value approximately 0.4 m) for the whole MDB (see figure 4.34 in section 4.4.1).
In the last part of this section the results from PCA for both data types are compared. Table 5.6 shows again the percentage of overall variability for the first four modes.

	percentage TRMM	percentage water level
Mode 1	75.5%	30.2%
Mode 2	17.7%	23.7%
Mode 3	3.1%	13.5%
Mode 4	1.7%	6.4%
$\sum$	98.0%	73.8%

Table 5.6: Percentage of overall variability for precipitation and water level observations

The percentages show a big difference. The amount of the cumulative percentage of all four modes of the water level observations is as large as the percentage of the first mode of the precipitation observations. Furthermore, the first and second mode of the water level observations are almost of the same amount. As described in section 4.4.2, the drought and annual signal is divided into the first three modes of the water level observations, while drought and annual signal are separated in the first and second mode for the precipitation observations. Therefore, it can be expected that a strong relationship between the modes of precipitation and water level observations cannot be found. To confirm or to disapprove this assumption the PC time series of the first four modes from the precipitation and from the water level observations are now analysed. In figure 5.11 the PC time series are illustrated for mode 1 to mode 4. At first view the biggest accordance between precipitation and water level observations can be found in mode 1. In this mode the PC time series show the same peaks with nearly the same amplitudes. Only the period of the drought (May 2006 - October 2007) shows different amplitudes. While the precipitation observations and also the GRACE-derived surface mass changes observations have the deepest peak in October 2006, the deepest peak in the PC time series of the water level observations is found as recently as May 2007. In January/February 2008 both PC time series show the largest value, indicating that half a year after the end of the drought more water than usual was available. In the second mode

only the first two years (May 2003 - May 2005) show a very good accordance. After that both time series vary considerably. The reason for that is probably the division of drought and annual signals in the first three modes of the water level observations, while drought and annual signal are separately represented for precipitation observations in mode 1 and mode 2. The third mode also shows differences. The precipitation observations have much lower amplitudes than the water level observations, indicating that the third mode contains much more noise than for the water level observations. This can be confirmed by the small percentage of mode 3 for the precipitation observations (3.1%), whereas the percentage of mode 3 for the water level observations accounts for 13.5% and is nearly five times larger. The lowest value for the water level observations appears in December 2006, which shows that the drought is still evident in mode 3. Note the lowest peak was found in mode 1 in October 2006 for both GRACE-derived surface mass changes and precipitation observations. While the fourth mode of precipitation observations oscillates around zero (with the exception of an anomaly in February 2008), the amplitudes of the water level observations are much larger. The same anomaly is also found in the water level observations, thus given reason for not being an artefact.



(a) Mode 1 (b) Mode 2

(c) Mode 3 (d) Mode 4

Figure 5.11: Comparison of PC time series of precipitation and water level observations (mode 1 - mode 4)

Next the correlation coefficients r and phase differences (PC time series) between the particular modes from precipitation and water level observations with the use of cross correlation are determined. Again a Student t-test is applied to examine the significance of the correlation coefficients. The test variable accounts for $T_{n-2} = T_{70} = 1.994$. The two-sided test (equation 5.1) is performed with a

probability of 97.5%. In table 5.7 the results are given.

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.434	0.130	0.088	0.149
Test variable t	4.035	1.093	0.741	1.260
Test decision	significant	not significant	not significant	not significant
Phase difference $\Delta\Phi$	-1 month	-3 months	+ 0 months	+1 month

Table 5.7: Correlation coefficients together with the phase differences for the PC times series of the first four modes (TRMM - water level)

As already seen in figure 5.11 the largest accordance is present for the first mode with a correlation coefficient of 0.434. All other modes show a very low correlation with correlation coefficients close to zero. Consequently, only the correlation coefficient of the PC time series for the first mode is significantly different from zero. Besides the correlation coefficients for the PC time series, the phase difference between the PC time series of both data sets were determined for each mode. Note the phase difference is taken with respect to the PC time series of the precipitation observations (e.g. $\Phi_{TRMM} - \Phi_{water\ level}$). As for the comparison between the GRACE-derived surface mass changes and precipitation observations, a negative value means that the PC time series of water level lags behind the PC time series of TRMM and a positive value indicates the inverse case. Consequently, the PC time series of the water level observations lags one month behind the PC time series of the precipitation observations (mode 1). All other modes are not significant, thus the following three modes are not examined in detail. It was shown that a high correlation between the original time series of water level and river flow observations exists. However, a Correlation Analysis of the PC time series in the appendix A.3 shows a low correlation between river flow and water level observations. A possible reason could be obtained by examining their contribution to the overall variability (see Appendix A.2 and figure 4.36 in section 4.4.2). The percentage of mode 1 from river flow observations is nearly as large as the percentages of mode 1 and mode 2 from water level observations. This would indicate that mode 1 of river flow observations contains the signal of mode 1 and mode 2 of water level observations.
The similarities and differences between the EOF maps are now analysed (see figure 4.25 in section 4.3.2 and figure 4.36 in section 4.4.2). The whole MDB shows a uniform pattern for both precipitation and water level observations. In the second mode the EOF map of the water level observations consists of mainly negative amplitudes for the southern rivers (red), while especially along the Darling River positive values (blue) are obtained. In the central part of the MDB (Lachlan River and Macquarie River (for the locations of the rivers see also figure 4.1) the amplitudes are close to zero, indicating once again that the central part of the MDB is a transition zone between the north and the south (see section 4.1.1). This pattern

is confirmed by the EOF map of mode 2 from the precipitation observations, where positive amplitudes in the north and negative amplitudes in the south can be found. The EOF map of mode 3 as well as the EOF map of mode 4 of both data types look different. The reason for that might be found in the relative contribution to the overall variability (see table 5.6). Mode 3 and all further modes of the precipitation observations contain mainly noise and/or higher frequency variations, whereas mode 3 and mode 4 of the water level observations still contain parts of the signal. Further determination of correlation coefficients for the EOF maps is not performed here.

5.3 Surface mass changes vs. in-situ hydrological observations

Finally, the similarities and differences between GRACE-derived surface mass changes und in-situ hydrological observations are investigated. First, the correlation coefficients for the by five degree grid for both data types are once again determined. Again a two-sided Student t-test (probability 97.5%) shows whether the correlation coefficients are significant or not. The theoretical quantile accounts for $T_{70-2} = T_{68} = 1.996$. The results are illustrated in figures 5.12 and 5.13. A band going through the central part of the MDB from the north to the south shows the largest correlation coefficients ranging from 0.3859 to 0.5363. These values have approximately the same magnitude as the correlation coefficients between the precipitation and water level observations. In fact, it could be expected that the correlation coefficients between GRACE-derived surface mass changes and water level observations are larger, since the GRACE mission detects mass changes from all involved processes and not only precipitation like the TRMM mission. At this point it is not clear what the reason for such similiar values of the correlation coefficients is. Significant correlation coefficients are located at the described band going from north to south.

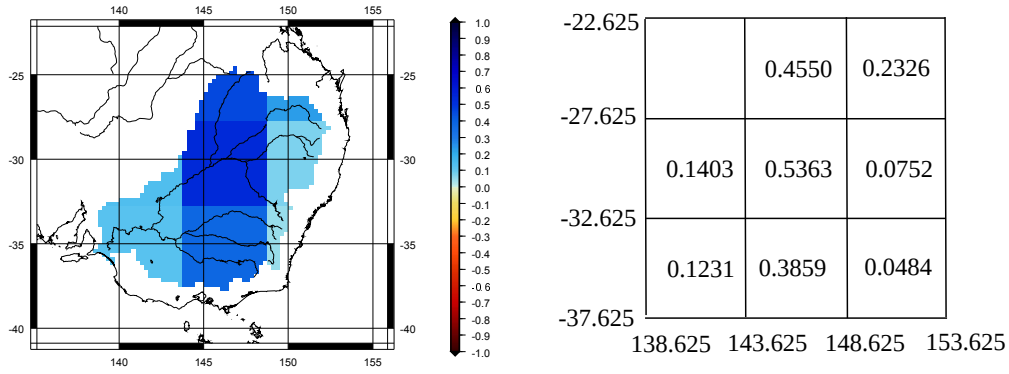


Figure 5.12: Correlation coefficients for GRACE-derived surface mass changes and water level observations

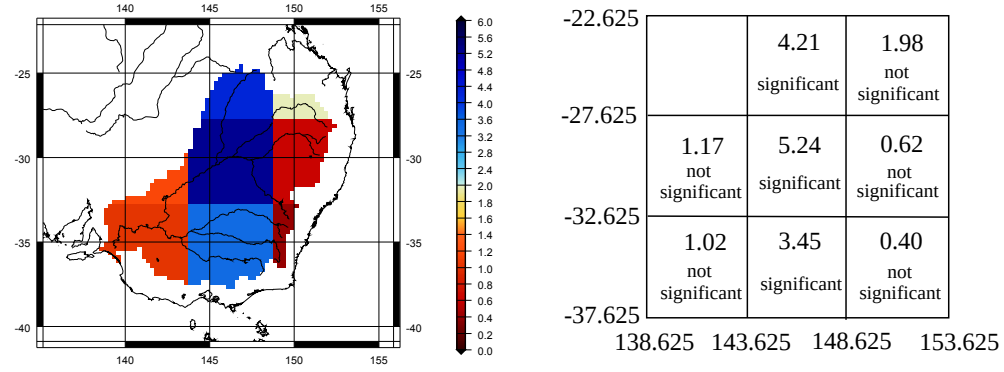


Figure 5.13: Significance test for correlation coefficients of GRACE-derived surface mass changes and water level observations (red colours mean not significant, blue colours mean significant)

Next the results from MLRA from both data types are analysed. For both GRACE-derived surface mass changes and water level observations a significant trend could be found for approximately half of all grid points (see figure 4.10 in section 4.2.2 and figure 4.31 in section 4.4.1). The trend in the north showed amplitudes just above zero, while in the south small negative values are present, indicating that the south was more affected by the drought than the northern part of the MDB.
The MLRA results showed for both data types that two regimes (north and south) exist, which can be seen at the phase values. For the north and the south the mean values of the phases are determined and are represented in table 5.8. These mean phases are once again illustrated in a cosine curve (see figure 5.14).

phase [months]	GRACE	water level	difference [months]
north	1.6	0.6	1.0
south	6.6	10.7	-4.1

Table 5.8: Mean phases for the north and south of the MDB (GRACE - water level) (phase is defined with respect to mid-January)

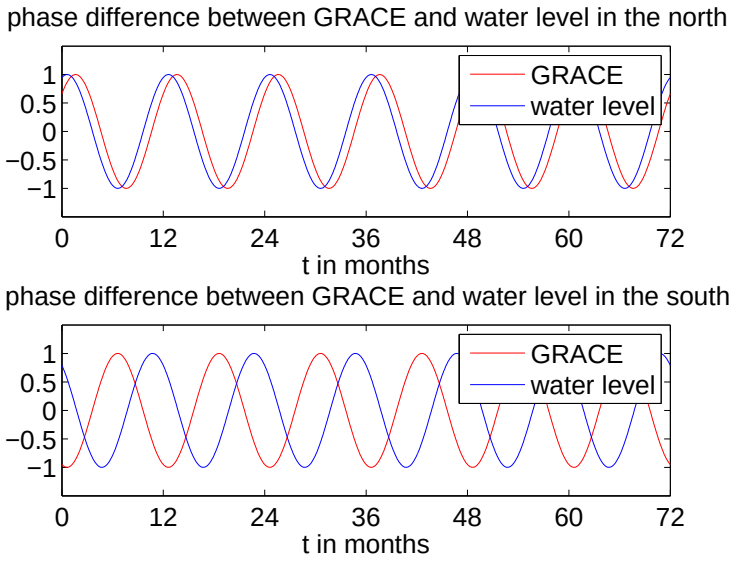


Figure 5.14: Phases from GRACE-derived surface mass changes and water level illustrated in a cosine curve for the north and the south of the MDB

Figure 5.14 illustrates that the water level observations lag one month behind the GRACE-derived surface mass changes in the north. In section 5.2 it was asserted that the mean phase of the water level observations is in the north exactly in accordance with the mean phase of the precipitation observations. In the south a phase difference of four months between the GRACE-derived surface mass changes and water level observations is present. As described in section 5.2 it is difficult to find an interpretation for this big phase difference. Once again a possible explanation for the three months deviation of precipitation and water level observations could be human operations (water regulations by dams and water extraction for irrigation), thus keeping water masses longer in the region. But since the GRACE mission detects all mass changes such as dam regulations, etc. there should be no big phase difference between GRACE-derived surface mass changes and water level observations. The reason for the four months deviation is unclear at this point. However, a possible explaination is given in section 5.4.

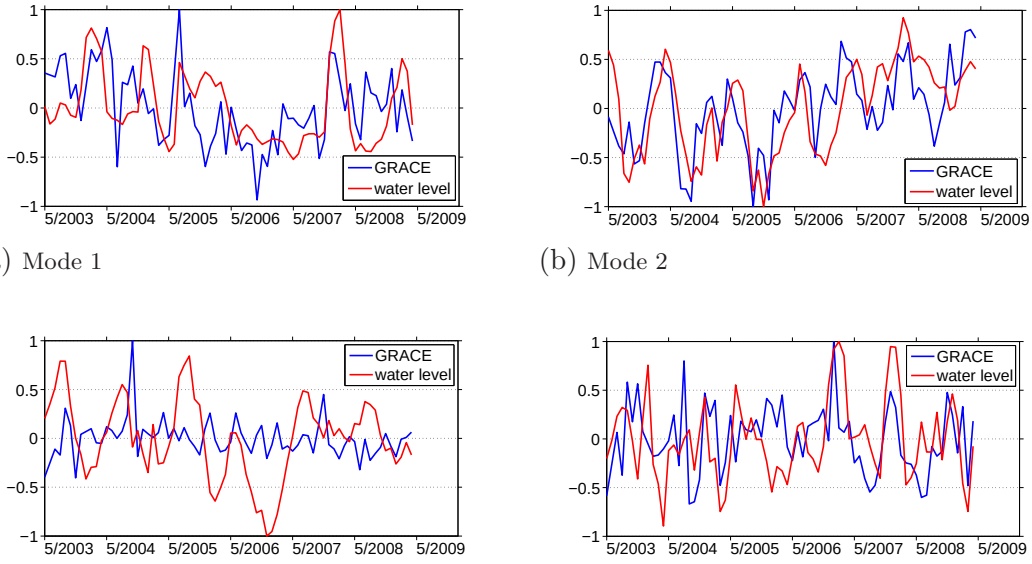
Comparing the annual amplitudes (from MLRA) of both data types show big differences (see figure 4.14 in section 4.2.2 and figure 4.34 in section 4.4.1). While the amplitudes of the precipitation observations in the south shows only half of the values of the north (up to 30 mm), the amplitudes of the water level observations are the same nearly everywhere (approximately 0.4 m).

Finally, the similarites and differences between the PCA results of GRACE-derived surface mass changes and water level observations are analysed. Table 5.9 illustrates again the percentage of overall variability for the first four modes.

	percentage GRACE	percentage water level
Mode 1	51.3%	30.2%
Mode 2	26.8%	23.7%
Mode 3	10.1%	13.5%
Mode 4	5.5%	6.4%
$\sum$	93.7%	73.8%

Table 5.9: Percentage of overall variability for GRACE-derived surface mass changes and water level observations

A big difference between the percentages of GRACE-derived surface mass changes and water level observations can be recognized. The cumulative percentage sum of the first two modes of GRACE-derived surface mass changes has approximately the same magnitude than the cumulative percentage sum of the first four modes of water level observations. As described above, the drought and annual signal are found separately in the first two modes of the GRACE-derived surface mass changes, while they are divided into the first three modes of the water level observations. This leads to the assumption that a big accordance between the modes of the GRACE-derived surface mass changes and water level observations cannot be found. To study this assumption the PC time series are now analysed. In figure 5.15 the PC time series are illustrated for the first four modes.



(a) Mode 1 (b) Mode 2

(c) Mode 3 (d) Mode 4

Figure 5.15: Comparison of PC time series of GRACE-derived surface mass changes and water level observations (mode 1 - mode 4)

Optically, the PC time series of GRACE-derived surface mass changes and water level observations fits much better than the PC time series of precipitation and water level observations. This result was expected since the GRACE mission detects all mass changes and not only changes due to precipitation like the TRMM mission. Especially, the first, second and fourth mode fit quite well. To deter-

mine the statistical relationships between the PC time series of both data types Correlation Analysis is applied. The correlation coefficients and the results of the Student t-test are presented in table 5.10. The theoretical quantile accounts for $T_{n-2} = T_{68} = 1.996$ (probability 97.5% (two-sided test)). The results show (as already noticed in figure 5.15) that significant correlation coefficients are found for mode 1, mode 2, and mode 4. For mode 3 a correlation coefficient close to zero was determined which was expected, since the PC time series of GRACE-derived surface mass changes oscillates very low around zero with the exception of one anomaly, demonstrating that this PC time series describes mainly noise. The largest correlation coefficient is found for mode 2 accounting for 0.63. Both time series show an annual signal in this mode. The phase of -1 months determined with respect to the GRACE-derived surface mass changes (e.g. $\Phi_{GRACE} - \Phi_{water\ level}$) denotes that the time series of the water level observations lags one month behind the GRACE-time series. The phase difference of 0 months for mode 1 denotes that the drought was detected by both data types at the same time. No attention should be paid to the 2 months phase difference for mode 3, since the correlation coefficient is almost zero and of course not significant.

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.408	0.629	-0.011	0.291
Test variable t	3.689	6.674	0.092	2.512
Test decision	significant	significant	not significant	significant
Phase difference $\Delta\Phi$	0 month	-1 months	+ 2 months	+0 month

Table 5.10: Correlation coefficients together with the phase difference for the PC times series of the first four modes (GRACE - water level)

Finally, the results of the EOF maps for the different modes are briefly discussed (see figure 4.16 in section 4.2.3 and figure 4.36 in section 4.4.2). The whole MDB shows a uniform pattern for both GRACE-derived surface mass changes and water level observations (mode 1). Both EOF maps for the mode 2 show negative amplitudes for the south of the MDB and positive values in the north, whereas the corresponding maps for mode 3 and mode 4 look different. Overall, the PCA results of the water level observations show that they are in a larger accordance with those of the GRACE-derived surface mass changes. This is expected, since, as explained, GRACE detects the sum of all mass changes and not only mass changes caused by precipitation like TRMM. Therefore, water level / river flow observations can be taken as a better representation of overall mass changes, as the in-situ hydrological observations also account for mass transport, while precipitation not. Further determination of correlation coefficients for the EOF map is not performed as well.

5.4 Monitoring mass transport

In section 5.1 it was asserted that the time difference between the north and the south for GRACE-derived surface mass changes accounts for approximately five months. In the south surface mass changes are detected by the GRACE mission one month earlier than precipitation by the TRMM mission, while the contrary in the north with a positive phase difference of one month was ascertained. It will be dealt now with the key question whether it is possible to monitor mass transport from the north to the south in the MDB by the GRACE mission that could explain the above described behaviour. The five month phase difference between the north and the south of the GRACE-derived surface mass changes together with the negative phase difference between GRACE and TRMM in the south suggests that the stream flow takes about five months to reach the south. In order to either confirm or refute this assumption the water level observations are now analysed. With the use of cross correlation it is possible to determine the phase difference between two time series (e.g. time it takes for water to flow from one location to the other). Therefore, the phase differences for all locations, where water level observations along the Darling River are available are determined. The results are represented in figure 5.16. The phase differences are determined with respect to the point P (northernmost point).

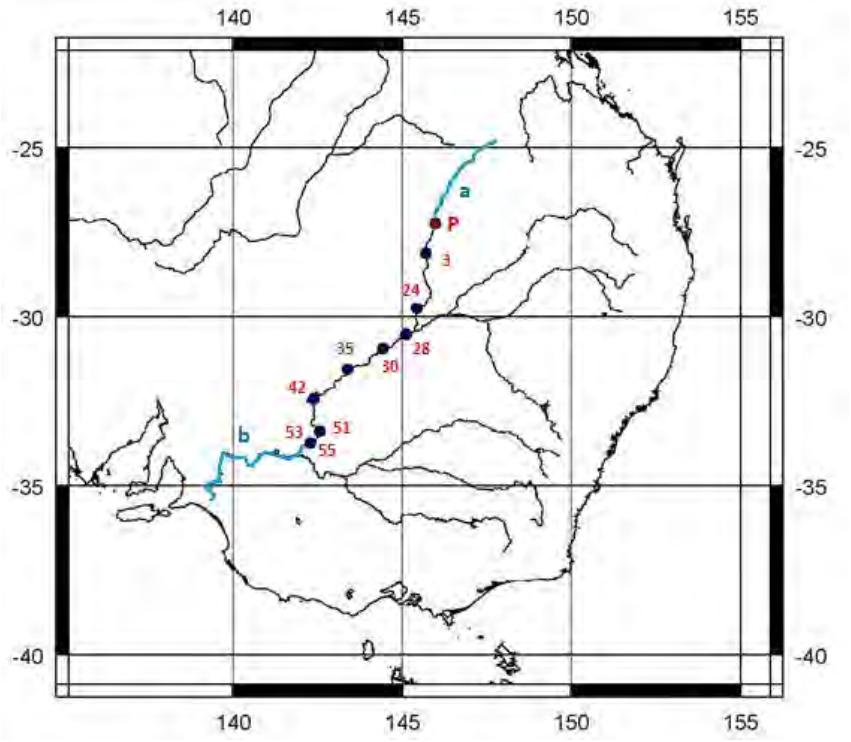


Figure 5.16: Duration of stream flow from the north to the south derived from water level observations

The red numbers at the points along the Darling River denote the number of days it takes for water to flow from the point P in the north to the location of a particular point. For example, the duration from the point P to the point in the southernmost point takes 55 days. Unfortunately no observations are available for the blue lines marked with a and b, respectively. Therefore, the time for the distance (blue line) marked with a and b, respectively, can only be estimated by means of extrapolation. However, it is difficult to perform extrapolation, since the gradients for the north and the south are not known by the author. Furthermore, as seen in figure 5.16, the augmentation of the time duration from the north to the south doesn't augment uniformly. Despite this problem, for the blue line in the north (marked with a) a duration of 10-15 days and for the southern part (marked with b) a duration of 20-30 days is estimated, assuming the flow time only depends on the approximated distance. Note these values are only estimates and their values are not assured by measurements. Together with the 55 days a total duration of 85-100 days is achieved. Of course, in reality, it can be more or less than 100 days. Consequently, a rounding would lead to a total duration of almost four months. Note the GRACE mission detected a five month phase difference between the north and the south (see figure 5.17). Consequently, it is a strong indication that the GRACE mission has the ability to monitor mass transport in the MDB and the key question is now answered. Of course it is not a definitive proof, since extrapolation was applied in order to find the time duration for the missing measurement points. Cross correlation was also applied to the other rivers, but since the northern rivers, Balonne River + Condamine River and Barwon River + Macintyre River (for the location of the rivers see figure 4.1) have many smaller confluents, the signals for the cross Correlation Analysis are disturbed and leads to inconclusive results.
Finally, the above results are summarized in figure 5.17.

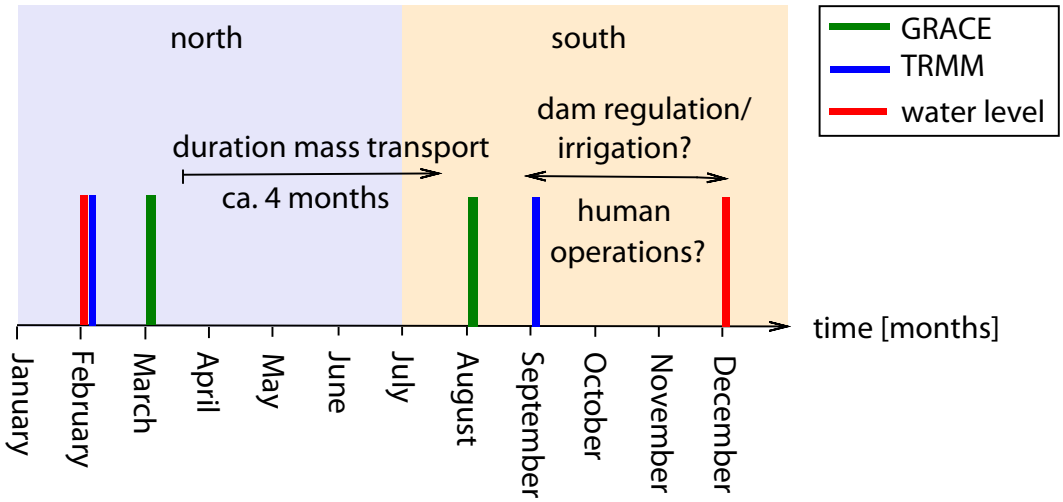


Figure 5.17: Mean phases for the MDB (GRACE - TRMM - water level)

In the north, where the phases of precipitation and water level observations coincide nearly exactly, the precipitation event occurs one month before the detection of surface mass changes by the GRACE mission. The phase difference of five months between the north and the south detected by the GRACE mission together with the negative phase difference of precipitation and surface mass changes in the south indicates a mass transport from the north to the south. This detected mass transport is confirmed by the water level observations (see demonstration above). Consequently, GRACE is able to monitor sub-basin surface mass changes - e.g. north vs. south. In summary, the mass transport of about four months can provide a reasonable explaination of the rather peculiar phase difference of one month between GRACE-derived surface mass changes and precipitation observations.

The reason for the three and four months phase difference, respectively, between precipitation - water level and GRACE-derived surface mass changes - water level, respectively, in the south suggests human operations like for example dam regulations and water extraction used for irrigation (e.g. water is longer in a region). But this assumption cannot be confirmed by the GRACE-surface mass changes, which also should detect mass changes caused by human operations. A possible explaination could be that, in terms of GRACE-derive surface mass changes, the south is more dominated by the mass transport from the north rather than from precipitation in the south. On the other hand water level observations are probably dominated by human operations (e.g. holding water longer in the region). This could explain the phase difference of four months between GRACE-derived surface mass changes and water level observations. However, this is a pure speculation and requires further analysis, which is left for further studies.

6 Conclusions and outlook

Hydrological changes within the MDB were examined in this diploma thesis. Firstly, basic background knowledge was provided in chapter 2. The modelling of the Earth's gravity field from space was discussed as well as the satellite missions, CHAMP, GRACE, and GOCE, which have the ability to detect gravity changes on the Earth, have been explained. Three different data types - GRACE-derived surface mass changes, precipitation from the TRMM mission, and hydrological in-situ observations such as water level and river flow - were used for the hydrological investigation of the MDB. These data sources were also discussed in chapter 2. Methods used for the analysis of hydrological changes, Multiple Linear Regression Analysis, Principal Component Analysis, and Correlation Analysis were explained in chapter 3. The goal of MLRA is to find a linear model fitting to a time series, whereas PCA is used to reduce the dimension of complex data in order to extract dominant structures and reduction of noise. Correlation Analysis is a statistical tool to analyse if two data series are correlated. In chapter 4 these techniques were applied to the three different data types. In chapter 5 the results of the different data types were compared by the use of correlation and cross correlation. In this chapter 6 conclusions and outlook are presented.

Before the performance of MLRA and PCA, the GRACE-derived surface mass changes were filtered in the space domain in order to reduce high-frequency noise. To make a fair comparison, the precipitation observations were filtered in the same way, using a Gaussian smoothing with a smoothing radius of 500 km. Furthermore, for the comparison the in-situ hydrological observations have been averaged on a 5° x 5° grid.

Based on statistical tests a suitable model to describe all three data types could be found, consisting of a linear combination of an offset, trend and annual signal. Even if the precipitation observations showed that a trend is not significant for the whole MDB, the model was not changed for the analysis of precipitation observations. A change of the model for one of the three data types compromises the comparison between them. For the in-situ hydrological observations, water level and river flow observations were available for the whole MDB. A Correlation Analysis of water level and river flow observations revealed correlation coefficients of close to one over the whole MDB. While the original analysis was performed for both, water level and

river flow observations, results for only the water level observations are presented, thus water level observations are taken here as in-situ hydrological observations. For completeness, river flow results are presented in the appendix A without further discussion.

The results showed that the above mentioned model was significant in most parts of the MDB for all three data types. It turned out that mostly the central part of the MDB was not significant, indicating that this area is a transition zone between different climatic zones in the north and south.

The trend analysis showed that the amplitude of the trend was slightly negative in the southern part of the MDB, whereas in the northern part of the MDB the trend was slightly positive. This is consistent with a drought taking place in the period from May 2006 until October 2007, with the south mostly affected, which explains the different values of the trend for the north and the south.

The analysis of the annual part of the model also confirmed the results of an earlier study (e.g. Rieser (2008)) that the MDB is divided into two different regimes - the north and the south. For example the determined phases showed that in the north in average precipitation occurs one month before the GRACE mission detects surface mass changes. Furthermore, the phase analysis of water level observations and precipitation observations showed almost identical phases in the north. This makes sense, since surface mass changes caused by precipitation cannot detect the precipitation immediately, since the water from the precipitation evaporates, runs off in rivers and lakes, and a small portion of precipitation accumulates in the region. The phase analysis showed that precipitation mainly occurs at the beginning of each year in the north, while precipitation in the south takes place about half a year later again indicating two different climatic zones. For the southern part of the MDB it turned out that in average precipitation lags approximately one month behind surface mass changes. However, this is reasonable taken into account the mass transport from north to south, which takes approximately four months as confirmed by the in-situ hydrological observations in section 5.4. This is rather close to the five months phase difference detected by surface mass changes between north and south. Therefore, the mass transport can be taken as reason why the surface mass changes are observed one month before the precipitation time series in the south. This also explains the one month larger detected phase difference for precipitation observations between the north and the south. By means of cross correlation the duration of mass transport from the north to the south was determined for the in-situ hydrological observations. Thereby the phase difference with respect to the water level observation point in northernmost part of the MDB was determined. The phase difference denotes the duration of the time for the water to flow from the north to the south. Unfortunately, water level observations along some river arms in the north and the south were missing. Consequently, extrapolation was also ap-

plied to determine the phase difference for these missing parts. Thereby, the above mentioned 4-months duration of mass transport from the north to the south was determined. Consequently, the key question of this diploma thesis could be answered, being whether it is possible to monitor water mass transport in the MDB. Since the performance of cross correlation and extrapolation revealed a value of up to four months, the assumption that the GRACE mission is able to detect water mass transport, can be confirmed. Of course, the analysis by means of cross correlation and extrapolation is not a definitive proof, however it is a strong indication that mass transport can be monitored by GRACE within the MDB. Furthermore, the two determined time spans for the stream flow from the north and the south are not exactly in accordance with one another (water level: up to 4 months, GRACE: 5 months). Consequently, GRACE is able to monitor sub-basin surface mass changes - e.g. north vs. south.

The phase analysis also showed that in the south water level observations have a phase difference of three months to the precipitation observations, and a phase difference of even four months to the GRACE-derived surface mass changes. For the derivation to the precipitation observations a possible explanation could be human operations such as water regulations by dams and water extraction for irrigation along the southern rivers. But coincidently, this assumption is refused by the phase of the GRACE-derived surface mass changes. Since the GRACE mission is able to detect all mass changes, including extraction of water and dam regulation the phases of surface mass changes and water level observations should be in accordance, which is not the case. A possible explaination for the big phase difference could be that GRACE-derived surface mass changes are more dominated by the mass transport from the north rather than from precipitation in the south, whereas water level observations are supposable dominated by human operations (e.g. holding water longer in the region). However, this is only a pure speculation and requires further analysis, which is left for further study.

The amplitudes of the annual signal showed that in the south of the MDB the magnitudes are half of the size of the amplitudes in the north. Only the in-situ hydrological changes showed that almost the same amplitudes were obtained everywhere in the MDB. The small amplitudes in the south confirm the assumption that the south of the MDB was more affected by the long drought period than the northern part of the MDB.

The PCA showed similar results especially for the precipitation and GRACE-derived surface mass changes. Both showed the long drought period in the first mode of the PC time series, whereas an annual signal was not found in the first mode. The EOF map represented an uniform pattern over the whole MDB. The annual signal was then detected in the second mode of the PC time series. Furthermore, the EOF map showed two distinct regimes in the north and the south of the MDB, whereas

in the central part of the MDB a climatic transition zone was confirmed. All further modes of precipitation and GRACE-derived surface mass changes contain more or less noise and/or higher frequency variations. The cumulative percentage of overall variability of the signal (mode 1 and mode 2) accounts for 93% (TRMM) and 78% (GRACE), respectively. Consequently, it can be assumed that the GRACE-derived surface mass changes contain more noise and/or higher frequency variations than does the precipitation observations. In contrast the in-situ hydrological observations showed that the drought and annual signal could not be found separately in two different modes. Instead, parts of the drought and the annual signal are found from mode 1 to mode 3. This fact is also reflected in the percentage of overall variability for the first four modes, which are all much lower compared to precipitation and GRACE-derived surface mass changes. In the first two modes a similar pattern in the EOF map compared to GRACE-derived surface mass changes and precipitation could be found. A Correlation Analysis of the EOF map for the GRACE-derived surface mass changes and precipitation observations showed a very high correlation. In contrast, only the second mode for the PC time series showed a larger correlation coefficient. Furthermore, the Correlation Analysis for the PC time series demonstrated that the relationships between GRACE-derived surface mass changes and in-situ hydrological changes are higher than between precipitation and in-situ hydrological changes. This is, of course, expected since the GRACE mission detects not only mass changes caused by precipitation, but also mass changes caused by other processes, in contrast to the TRMM mission, which detects only those changes caused by precipitation.

Finally, a Correlation Analysis for the time series was performed between the three data types. It turned out that the largest correlation coefficients were found in the north and the central part of the MDB. Actually, it would be expected that the correlation between GRACE-derived surface mass changes and in-situ hydrological observations is higher than the correlation between precipitation and in-situ hydrological observations. But this is only the case for the PC time series. Consequently, it can be assumed that some noise remains in the time series, even after the application of Gaussian filtering in the spatial domain. This noise is then filtered out by means of PCA. Therefore, a further approach could be the application of a filter technique in the time domain to reduce more noise/higher frequency variation. This, however, is left for further study. A possible approach could be to use PCA also as a filter here, e.g. a filtered signal is re-constructed by taking only the first few modes of PCA, with which the analysis is then performed.

Bibliography

Brook, R. J./Arnold, G. C. (1985): Applied regression analysis and experimental design, Marcel Dekker, inc., New York.

Campbell, B. (2008): Unprecedented Drought - Climate Change and the Murray-Darling Basin, Murray-Darling Basin Commission, http://www.sfu.ca/act/program/documents/Bruce_Campbell.ppt, last access: 4.6.10.

CSR (2010): GRACE Gravity Recovery and Climate Experiment, http://www.csr.utexas.edu/grace/overview.html, last access: 4.6.10.

ESA (2010): GOCE, http://www.esa.int/SPECIALS/GOCE, last access: 4.6.10.

Flechtner, F. (2010): GFZ's Gravity Recovery And Climate Experiment, GeoForschungsZentrum Potsdam, http://op.gfz-potsdam.de/grace/index_GRACE.html, last access: 4.6.10.

Han, S.-C./Jekeli, C./Shum, C. K. (2004): Time-variable aliasing effects of ocean tides, atmosphere, and continental water mass on monthly mean GRACE gravity field, in: Journal of Geophysical Research, 109.

Heiskanen, W. A./Moritz, H. (1967): Physical Geodesy, W. H. Freeman and Company, San Francisco.

Jaxa (2010): TRMM Tropical Rainfall Measuring Mission, http://www.eorc.jaxa.jp/TRMM/index_e.htm, last access: 4.6.10.

Johnson, R. A./Wichern, D. W. (2007): Applied Multivariate Statistical Analysis, Pearson Prentice Hall, New York.

Jolliffe, I. T. (2002): Principle Component Analysis, Springer.

Krzanowski, W. J. (2007): Principles of multivariate analysis, Oxford Univ. Press.

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

Nicholls, N. (2004): The Changing Nature of Australian Droughts, in: Climatic Change, 63, p. 323-336.

Prasad, A./Khan, S. (2002): Murray-Darling Basin Dialogue on Water and Climate, presentation at River Symposium, Brisbane 6 September 2002, http://www.waterandclimate.org/dialogue/Basin/Murray-Darling/..., last access: 4.6.10.

Preisendorfer, R. W. (1988): Principal Component Analysis in Meteorology and Oceanography, Elsevier, Washington.

Rieser, D. (2008): Comparison of GRACE-derived monthly Surface Mass Variations with Rainfall Data in Australia, master thesis, Graz University of Technology.

Schmidt, R./Petrovic, S./Guentner, A./Barthelmes, F./Wuensch, J./Kusche, J. (2008): Periodic components of water storage changes from GRACE and global hydrology models, in: Journal of Geophysical Research, 113.

Swenson, S./Wahr, J. (2006): Post-processing removal of correlated errors in GRACE data, in: Geophysical Research Letters, 33.

Tapley, B. D./Bettadpur, S./Ries, J. C./Thompson, P. F./Watkins, M. M. (2004): GRACE Measurements of Mass Variability in the Earth System, in: science, 305(5683), p. 503-505.

TU Freiberg (2010): Die Kreuzkorrelation, TU Freiberg, http://tu-freiberg.de/fakult4/imfd/lehre/fluid/MT/Kreuzkorrelation.pdf, last access: 4.6.10.

USGS (2010): Water Science for Schools, http://ga.water.usgs.gov/edu/measureflow.html, last access: 4.6.10.

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, in: Journal of Geophysical Research, 103, p. 30,205 - 30,229.

Wikipedia (2010a): Climate of Australia, Wikipedia the free encyclopedia, http://en.wikipedia.org/wiki/Climate_of_Australia, last access: 4.6.10.

Wikipedia (2010b): Tropical Rainfall Measuring Mission, Wikipedia the free encyclopedia, http://en.wikipedia.org/wiki/Tropical_Rainfall_Measuring_Mission, last access: 4.6.10.

Appendix A

Illustrations of the river flow results

In the following all results of the analysis of river flow observations are presented without further discussion. Due to the high correlation between water level and river flow observations all conclusions made for the water level observations equally hold for the river flow data.

A.1 Results from Multiple Linear Regression Analysis

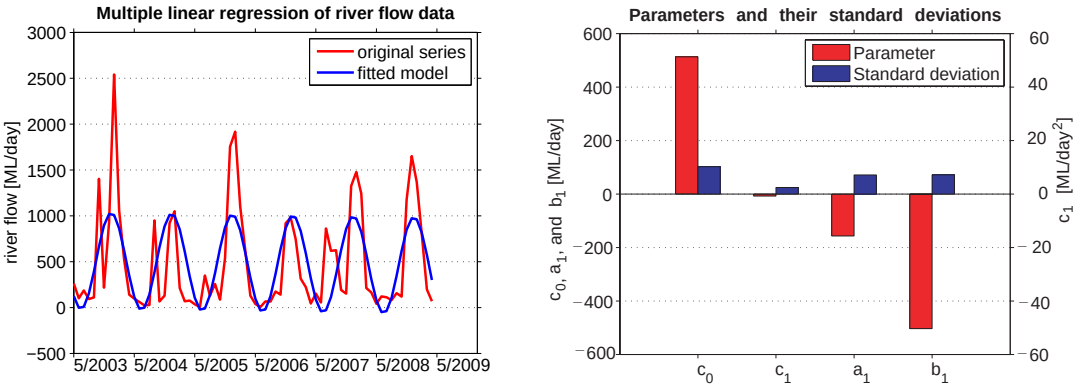


Figure A.1: Multiple linear regression of river flow observations at $\lambda = 149.874°$, $\varphi = -28.609°$ (Macintyre River)

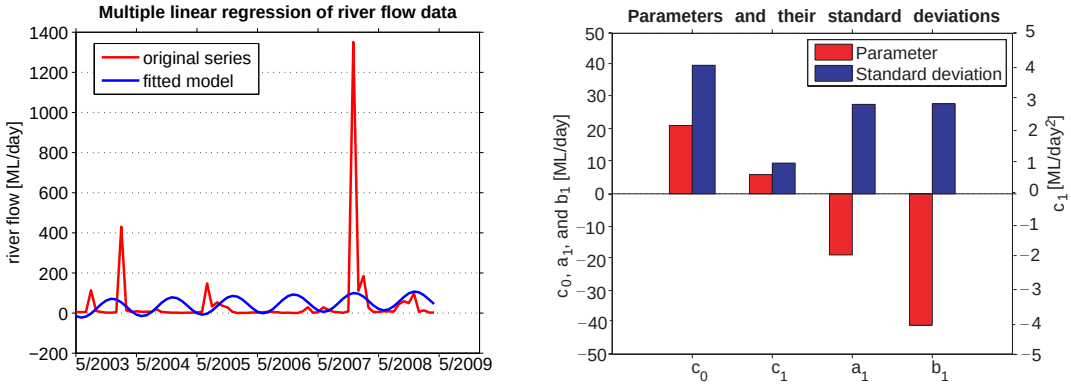


Figure A.2: Multiple linear regression of river flow observations at $\lambda = 149.310°$, $\varphi = -31.420°$ (Namoi River)

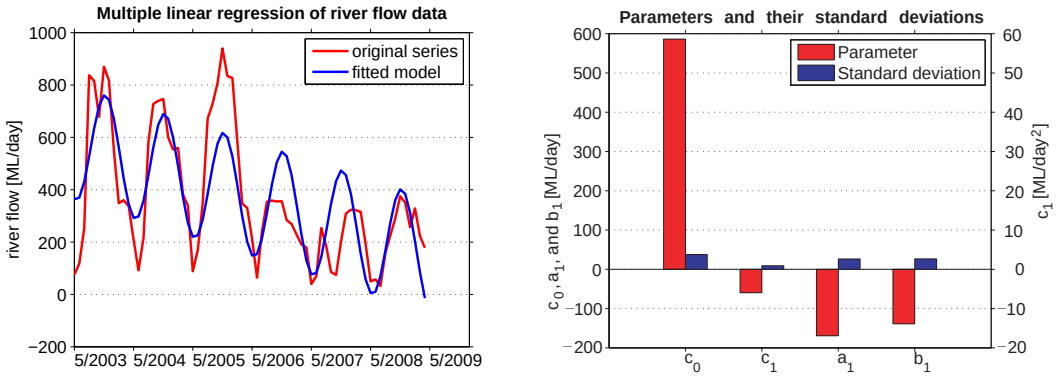


Figure A.3: Multiple linear regression of river flow observations at $\lambda = 144.988°$, $\varphi = -35.849°$ (Murray River)

Significance tests

1. Two-sided significance test for the different parameters:
   quantile T=1.996, probability 97.5%

coefficient j	min $\lvert t_j \rvert$	max $\lvert t_j \rvert$	# significant	# not significant
c_0	0.056	19.755	77 (83.7%)	15 (16.3%)
c_1	0.076	11.161	36 (39.1%)	56 (60.9%)
a_1	0.140	7.972	40 (43.5%)	52 (56.5%)
b_1	0.005	8.874	63 (68.5%)	29 (31.5%)

Table A.1: Significance tests for the four different parameters of the MLRA model applied to the river flow observations

2. One-sided significance test for different MLRA models: probability 95%

case	$F_{0.95,2,n_{full}-u_{full}}$	# F significant	# F not significant
1 ($M_0 \rightarrow M_1$)	3.1317	90 (97.8%)	2 (2.2%)
2 ($M_1 \rightarrow M_2$)	3.1359	87 (94.6%)	5 (5.4%)
3 ($M_2 \rightarrow M_3$)	3.1404	89 (96.7%)	3 (3.3%)
4 ($M_3 \rightarrow M_4$)	3.1453	90 (97.8%)	2 (2.2%)

Table A.2: Significance tests for the comparison of different multiple linear regression models (river flow)

3. Significance test for the M_1 model

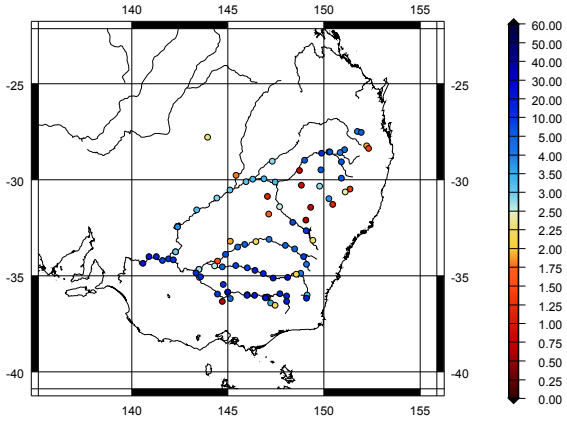
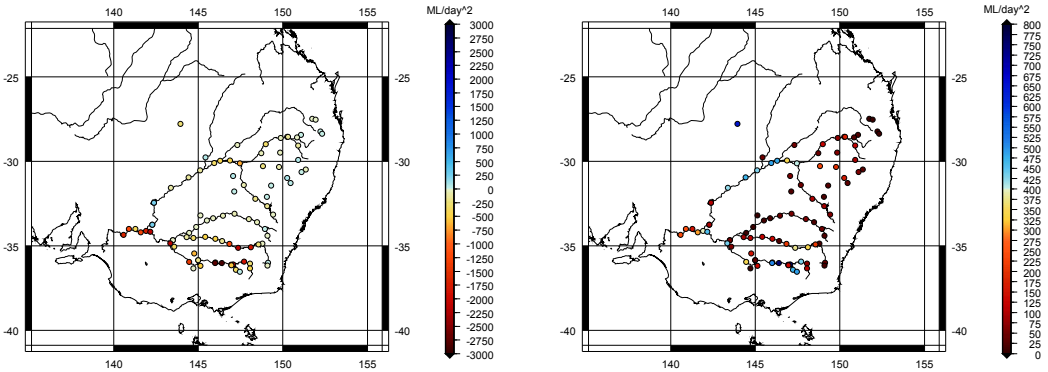


Figure A.4: One-sided significance test for the M_1 model (river flow), probability 95%, theoretical quantile: 2.5087 (red colours mean not significant, blue colours mean significant)

Analysis of trend



(a) trend (b) standard deviation of trend

Figure A.5: Trend together with its standard deviation for river flow observations

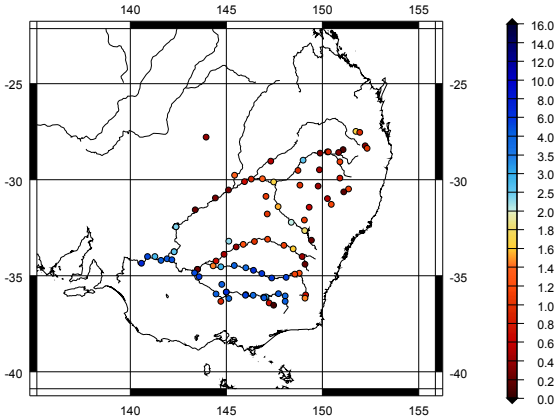
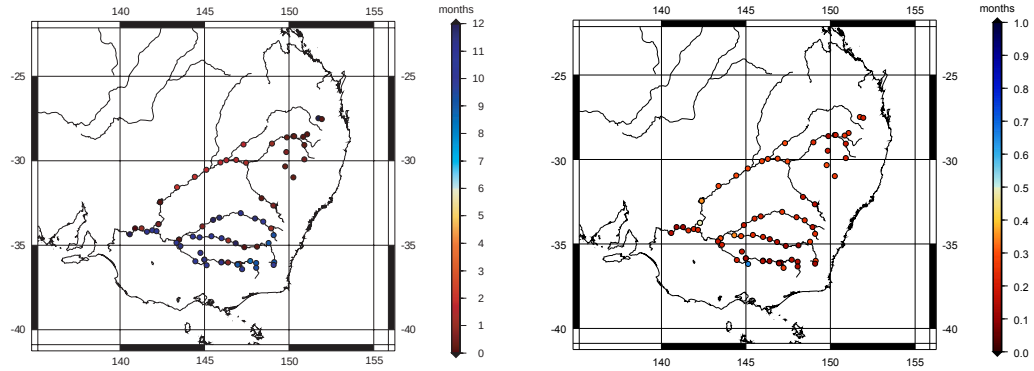


Figure A.6: Two-sided significance test for the trend parameter c_1 (river flow), probability 97.5%, theoretical quantile: 1.996 (red colours mean not significant, blue colours mean significant)

Analysis of the annual part of the model M_1

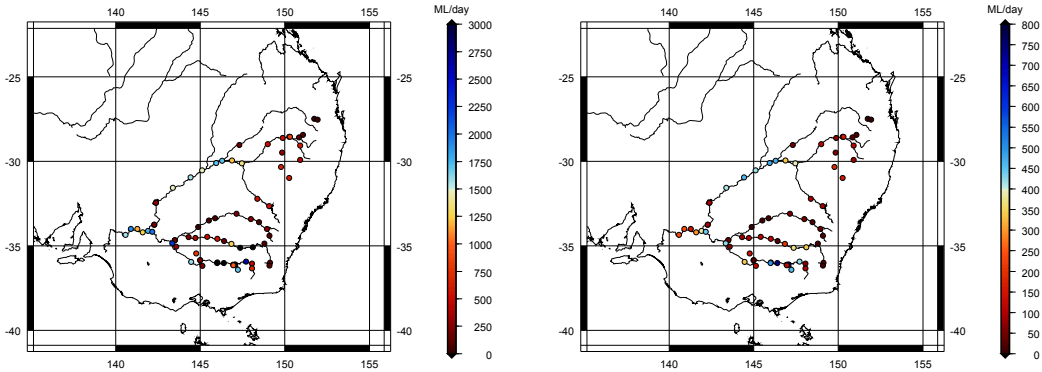
1. Phase at significant points of the model M_1



(a) phase (b) standard deviation of phase

Figure A.7: Phase together with its standard deviation for river flow observations (phase is defined with respect to mid-January)

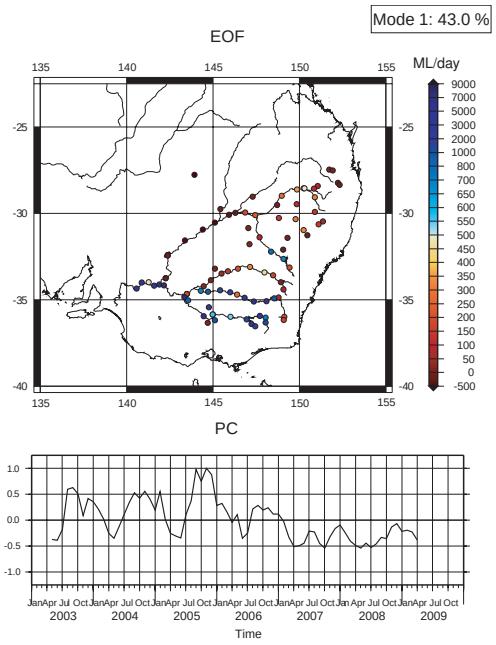
2. Amplitude at significant points of the model M_1



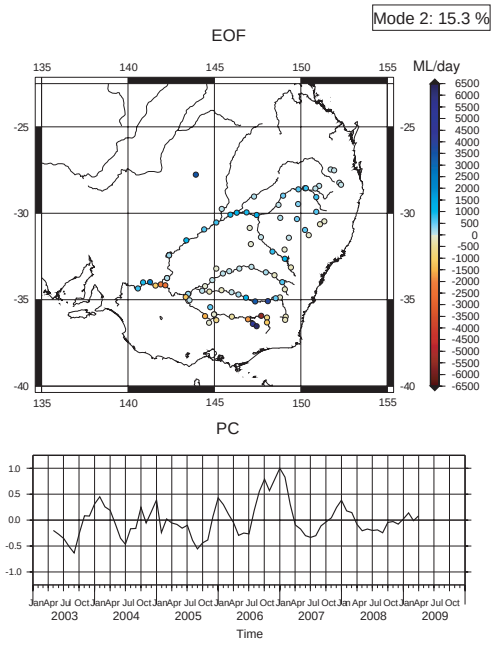
(a) amplitude (b) standard deviation of amplitude

Figure A.8: Amplitude together with its standard deviation for river flow observations

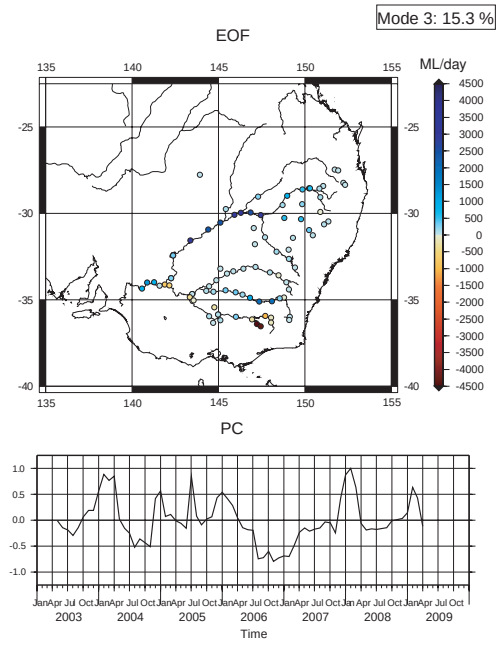
A.2 Results from Principal Component Analysis



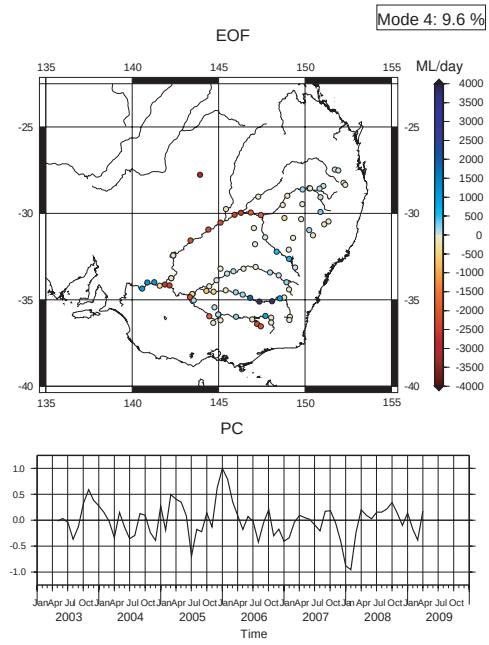
(a) Mode 1



(b) Mode 2



(c) Mode 3



(d) Mode 4

Figure A.9: PCA modes 1-4 of river flow observations

A.3 Results from Correlation Analysis

Correlation Analysis between river flow and the other data types

1. **river flow vs. surface mass changes**

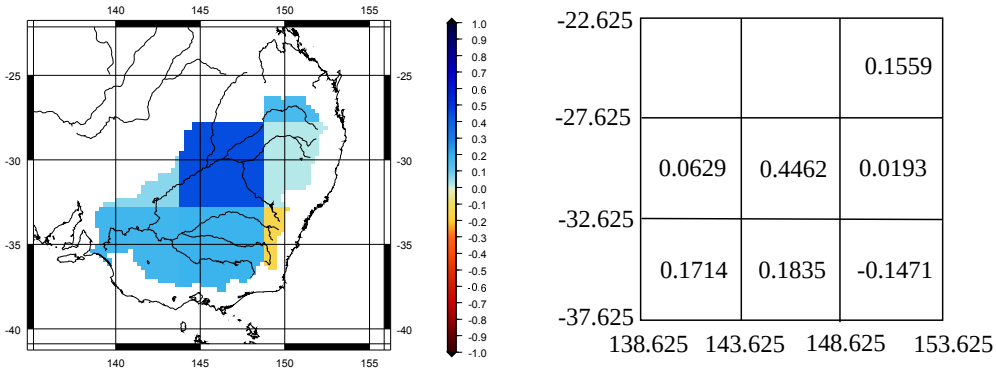


Figure A.10: Correlation coefficients for river flow observations and surface mass changes

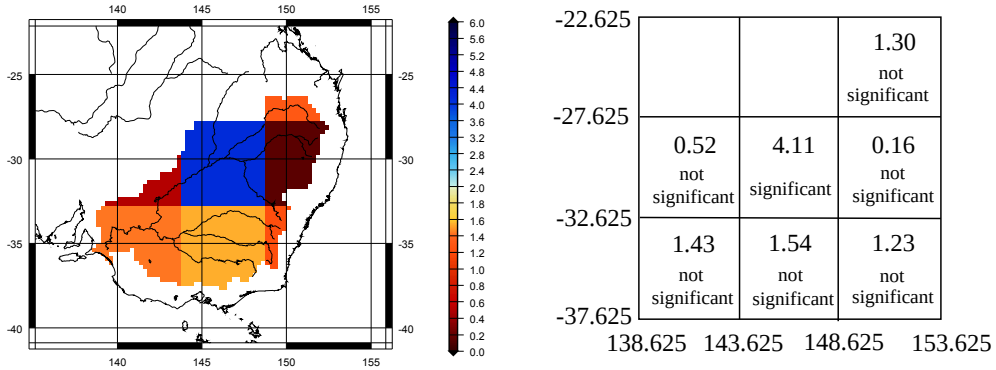


Figure A.11: Two-sided significance test for correlation coefficients of river flow observations and surface mass changes, probability: 97.5%, theoretical quantile: 1.996 (red colours mean not significant, blue colours mean significant)

2. **river flow vs. precipitation**

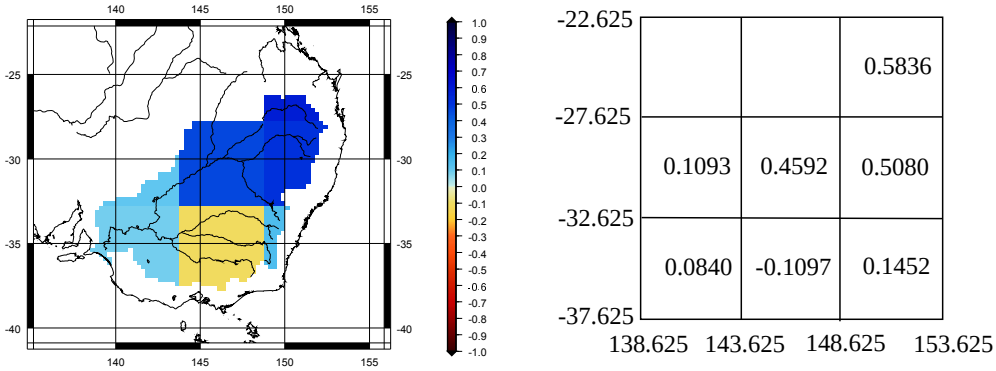


Figure A.12: Correlation coefficients for river flow and precipitation observations

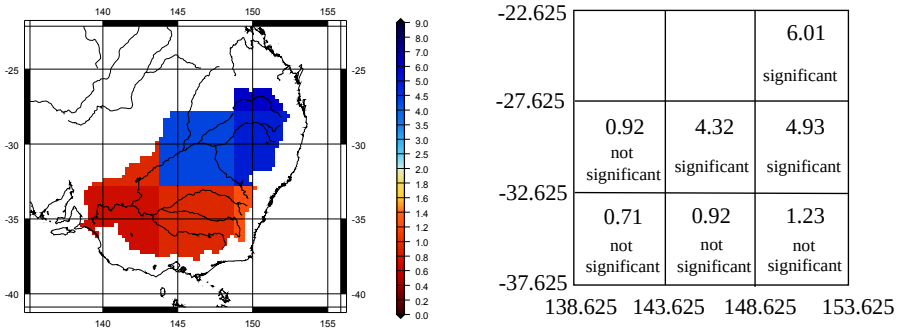


Figure A.13: Two-sided significance test for correlation coefficients of river flow and precipitation observations, probability: 97.5%, theoretical quantile: 1.994 (red colours mean not significant, blue colours mean significant)

Correlation Analysis of PC time series between the different data types

Note all phase differences are determined with respect to the river flow data.

1. **river flow vs. water level observations**

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.376	0.013	0.095	-0.429
Test variable t	3.396	0.112	0.797	3.973
Test decision	significant	not significant	not significant	significant
Phase shift $\Delta\Phi$	-2 months	-2 months	-6 months	+4 months

Table A.3: Correlation coefficients together with the phase difference for the PC times series of the first four modes (river flow - water level), two-sided significance test: probability: 97.5%, theoretical quantile: 1.994

2. **river flow vs. precipitation observations**

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	0.063	0.607	-0.154	-0.302
Test variable t	0.530	6.385	1.307	2.654
Test decision	not significant	significant	not significant	significant
Phase shift $\Delta\Phi$	-2 months	0 months	-4 months	+2 months

Table A.4: Correlation coefficients together with the phase difference for the PC times series of the first four modes (river flow - precipitation), two-sided significance test: probability: 97.5%, theoretical quantile: 1.994

3. **river flow vs. surface mass changes**

	Mode 1	Mode 2	Mode 3	Mode 4
Correlation coeff. r	-0.025	0.402	-0.105	-0.039
Test variable t	0.204	3.624	0.867	0.323
Test decision	not significant	significant	not significant	not significant
Phase shift $\Delta\Phi$	-6 months	-1 months	+4 months	+5 months

Table A.5: Correlation coefficients together with the phase difference for the PC times series of the first four modes (river flow - surface mass changes), two-sided significance test: probability: 97.5%, theoretical quantile: 1.996