Institute of Geoecology

Calibration of the global hydrological model WGHM with water mass variations from GRACE gravity data

A dissertation submitted to the
Faculty of Mathematics and Natural Sciences at the University of Potsdam, Germany
for the degree of Doctor of Natural Sciences (Dr. rer. nat.) in Hydrology by

Susanna Werth

Submitted: August 20, 2009
Defended: February 17, 2010
Published: March, 2010

Referees

Prof. Bruno Merz, University of Potsdam, Institute for Geoecology
Prof. Hubert Savenije* TU Delft, Civil Engineering and Geosciences
Prof. Reinhardt Dietrich, TU Dresden, Institute of Planetary Geodesy

Published online at the
Institutional Repository of the University of Potsdam:
URL http://opus.kobv.de/ubp/volltexte/2010/4173/
URN urn:nbn:de:kobv:517-opus-41738
http://nbn-resolving.org/urn:nbn:de:kobv:517-opus-41738

Author's declaration

I prepared this dissertation without illegal assistance. The work is original except where indicated by special reference in the text and no part of the dissertation has been submitted for any other degree. This dissertation has not been presented to any other University for examination, neither in Germany nor in another country.

Susanna Werth
Potsdam, August 2009

Contents

Abstract

Since the start-up of the GRACE (Gravity Recovery And Climate Experiment) mission in 2002 time dependent global maps of the Earth's gravity field are available to study geophysical and climatologically-driven mass redistributions on the Earth's surface. In particular, GRACE observations of total water storage changes (TWSV) provide a comprehensive data set for analysing the water cycle on large scales. Therefore they are invaluable for validation and calibration of large-scale hydrological models as the WaterGAP Global Hydrology Model (WGHM) which simulates the continental water cycle including its most important components, such as soil, snow, canopy, surface- and groundwater. Hitherto, WGHM exhibits significant differences to GRACE, especially for the seasonal amplitude of TWSV. The need for a validation of hydrological models is further highlighted by large differences between several global models, e.g. WGHM, the Global Land Data Assimilation System (GLDAS) and the Land Dynamics model (LaD).

For this purpose, GRACE links geodetic and hydrological research aspects. This link demands the development of adequate data integration methods on both sides, forming the main objectives of this work. They include the derivation of accurate GRACE-based water storage changes, the development of strategies to integrate GRACE data into a global hydrological model as well as a calibration method, followed by the re-calibration of WGHM in order to analyse process and model responses. To achieve these aims, GRACE filter tools for the derivation of regionally averaged TWSV were evaluated for specific river basins. Here, a decorrelation filter using GRACE orbits for its design is most efficient among the tested methods. Consistency in data and equal spatial resolution between observed and simulated TWSV were realised by the inclusion of all most important hydrological processes and an equal filtering of both data sets. Appropriate calibration parameters were derived by a WGHM sensitivity analysis against TWSV. Finally, a multi-objective calibration framework was developed to constrain model predictions by both river discharge and GRACE TWSV, realised with a respective evolutionary method, the ϵ-Non-dominated-Sorting-Genetic-Algorithm-II (ϵ-NSGAII).

Model calibration was done for the 28 largest river basins worldwide and for most of them improved simulation results were achieved with regard to both objectives. From the multi-objective approach more reliable and consistent simulations of TWSV within the continental water cycle were gained and possible model structure errors or mis-modelled processes for specific river basins detected. For tropical regions as such, the seasonal amplitude of water mass variations has increased. The findings lead to an improved understanding of hydrological processes and their representation in the global model. Finally, the robustness of the results is analysed with respect to GRACE and runoff measurement errors. As a main conclusion obtained from the results, not only soil water and snow storage but also groundwater and surface water storage have to be included in the comparison of the modelled and GRACE-derived total water budged data. Regarding model calibration, the regional varying distribution of parameter sensitivity suggests to tune only parameter of important processes within each region. Furthermore, observations of

single storage components beside runoff are necessary to improve signal amplitudes and timing of simulated TWSV as well as to evaluate them with higher accuracy.

The results of this work highlight the valuable nature of GRACE data when merged into large-scale hydrological modelling and depict methods to improve large-scale hydrological models.

German Abstract

Seit dem Start der GRACE-Mission (Gravity Recovery And Climate Experiment) im Jahr 2002 sind globale Daten von zeitlichen Veränderungen des Erdschwerefeldes verfügbar, mit deren Hilfe sich geophysikalische und klimatologische Massenumverteilungen auf der Erdoberfläche studieren lassen. Insbesondere die von GRACE beobachteten Variationen des gesamten Wasserspeichers (TWSV) bieten erstmals einen globalen Datensatz für die Analyse des Wasserkreislaufes auf großen räumlichen Skalen. Sie sind außerordentlich wertvoll für die Validierung und Kalibrierung von großskaligen hydrologischen Modellen, wie das "WaterGAP Global Hydrology Model" (WGHM), welches kontinentale Wasserspeicher, einschließlich der wichtigsten Komponenten (Boden, Schnee, Interzeption, Oberflächen- und Grundwasser), simuliert. Bisher weist WGHM insbesondere in der saisonalen Amplitude der TWSV gegenüber GRACE signifikante Differenzen auf. Sehr große Unterschiede zwischen hydrologischen Modellen, z.B. dem WGHM, dem "Global Land Data Assimilation System" (GLDAS) und dem "Land Dynamics model" (LaD) betonen die Notwendigkeit hydrologische Modelle zu validieren.

Zu diesem Zweck verbindet GRACE die Wissenschaftsbereiche der Geodäsie und der Hydrologie. Diese Verknüpfung verlangt von beiden Seiten die Entwicklung geeigneter Methoden zur Datenintegration, welche die Hauptaufgaben dieser Arbeit darstellen. Dabei handelt sich es insbesondere um die Ableitung von genauen GRACE-basierten TWSV und um die Strategie-Entwicklung zur Integration von GRACE Daten in ein hydrologisches Modell sowie zur Kalibrierung von WGHM. Das abschließende Ziel ist die Rekalibrierung von WGHM, mit der Motivation Prozess- und Modellverhalten zu analysieren. Um diese Ziele zu erreichen, wurden in der vorliegenden Arbeit für bestimmte Flusseinzugsgebiete verschiedene GRACE-Filter evaluiert, die zur Ableitung von räumlich gemittelten TWSV dienen. Als effizienteste unter den getesteten Methoden erwies sich ein Dekorrelationsfilter, für dessen Design GRACE-Orbits angewendet werden. Die Konsistenz zwischen den zu vergleichenden Daten und deren räumlicher Auflösung wurde durch den Einschluss aller wichtigen hydrologischen Prozesse sowie eine äquivalente Filterung beider Datensätze realisiert. Durch eine Sensitivitätsanalyse des Modells auf TWSV wurden geeignete Kalibrierparameter bestimmt. Abschließend konnte ein multi-kriterieller Kalibrierrahmen entwickelt werden, der eine entsprechende evolutionäre Kalibriermethode ϵ-Non-dominated-Sorting-Genetic-Algorithm-II (ϵ-NSGAII) anwendet, um hydrologische Modellsimulationen an gemessene Abflusszeitreihen und an GRACE-abgeleitete TWSV anzupassen.

Die Modellkalibierung wurde weltweit für die 28 größten Flusseinzugsgebiete durchgeführt. In den meisten Fällen ergab sich eine Simulationsverbesserung gegenüber beiden Kalibrierkriterien. Mit Hilfe des multi-kriteriellen Ansatzes wurden verlässlichere und konsistentere Simulationen von TWSV erreicht sowie mögliche Modell-Strukturfehler oder unkorrekt modellierte Prozesse aufgedeckt. Für tropische Regionen ergeben sich z.B. größere saisonale Amplituden in den Wassermassenvariationen. Die Ergebnisse führen zu einem verbesserten Verständnis hydrologischer Prozesse und helfen bei der Optimierung des globalen Modells. Zum Schluss konnte

die Robustheit der Ergebnisse gegenüber Fehlern in GRACE- und Abflussmessungen erfolgreich getestet werden. Nach den wichtigsten Schlussfolgerungen, die aus den Ergebnissen abgeleitet werden konnten, sind nicht nur Bodenfeuchte- und Schneespeicher sondern auch Grundwasser- und Oberflächenwasserspeicher in Vergleiche des gesamten Wasserbudgets zwischen simulierten und GRACE detektierten Analysen einzubeziehen. Für die Modellkalibrierung sind regional variierende Parametersensitivitäten zu beachten, z.B. in dem für ein bestimmtes Einzugsgebiet nur Parameter bedeutender Prozesse kalibriert werden. Weiterhin sind neben Abflussmessungen zusätzlich Beobachtungen von Einzelspeicherkomponenten notwendig, um die Signalstärke von simulierten TWSV sowie deren zeitliche Anpassung zu verbessern und mit größerer Genauigkeit zu evaluieren.

Die Ergebnisse dieser Arbeit zeigen wie wertvoll GRACE Daten für die großskalige hydrologische Modellierung sind und eröffnen eine Methode zu Verbesserung selbiger.

List of Publications

The present work constitutes a cumulative dissertation. It includes three articles of internationally accredited scientific journals, with one being published, one accepted and one under review by the time the dissertation was submitted to the University of Potsdam. The publication of all three article could be completed, by the time of the defense of the present thesis:

List of Figures

List of Tables

Abbreviations

A3HM	**A**verage of **3** global **H**ydrological **M**odels
CC	**C**orrelation **C**oefficient
CHAMP	**CHA**llenging **M**inisatellite **P**ayload
CSR	**C**enter for **S**pace **R**esearch, Austin, USA
ECMWF	**E**uropean **C**entre for **M**edium-Range **W**eather **F**orecasts
GDAS	**G**lobal **D**ata **A**ssimilation **S**ystem
GFZ	Deutsches **G**eo**F**orschungs**Z**entrum, Helmoltz-Zentrum Potsdam (German Research Centre for Geosciences, Helmholtz Centre Potsdam)
GLDAS	**G**lobal **L**and **D**ata **A**sssimilation **S**ystem
GLDAS-NOAH	**GLDAS** **N**ational Centers for Environmental Prediction / **O**regon State University / **A**ir Force / **H**ydrologic Research Lab Model
GPCC	**G**lobal **P**repitation **C**limatology **C**entre
GRACE	**G**ravity **R**ecovery **A**nd **C**limate **E**xperiment
GRDC	**G**lobal **R**unoff **D**ata **C**entre
JPL	**J**et **P**ropulsion **L**aboratory, Pasadena, USA
LaD	**La**nd **D**ynamic model
LEO	**L**ow **E**arth **O**rbiting Satellite
MCWH	**M**aximum **C**anopy **W**ater **H**eight
MPET	**M**aximum **P**otential **E**vapotranspiration
NCEP	**N**ational **C**enters for **E**nvironmental **P**rediction
NSC	**N**ash-**S**utcliffe-**C**oefficient
ORCHIDEE	**OR**ganising **C**arbon and **H**ydrology **I**n **D**ynamic **E**cosyst**E**ms
PT	**P**riestley-**T**aylor
RL	**R**ea**L**ease
RMS	**R**oot **M**ean **S**quare
RMSE	**R**oot **M**ean **S**quare **E**rror
SH	**S**pherical **H**armonics
SST	**S**atellite to **S**atellite **T**racking
SST-ll	**S**atellite to **S**atellite **T**racking im **l**ow-**l**ow Modus
TWS	**T**otal **W**ater **S**torage
TWSC	**T**otal **W**ater **S**torage **C**hange
TWSV	**T**otal **W**ater **S**torage **V**ariations
WGHM	**W**aterGAP **G**lobal **H**ydrological **M**odel
SMOS	**S**oil **M**oisture and **O**cean **S**alinity

1 Introduction

1.1 Interfaces between geodesy and hydrology

Geodesy comprises the survey of the Earth's figure. Herein, an important research objective is the determination of the gravity field, which provides a natural height system and influences all geodetic measurements. It derives from a superimposition of centrifugal as well as gravitational forces and therefore, it depends on mass distributions on and below the Earth's surface. For many years, the potential field was considered as a geometrical and static figure. Since the end of the 20th century, satellite missions enable a global survey of the Earth including is gravity field. The continuously increasing measurement accuracy directed the focus to dynamical behaviours. Temporal variations in the gravity field are caused by geophysical or climatic induced mass transfers above and below the Earth's surface. Within geosphere, oceans, atmosphere, hydrosphere, cryosphere and biosphere system typical substances change their location and distribution inside and between the individual subsystems. By the reflection of these mass transfers in changes of the planets gravitational field, they affect the location of satellite orbits in space. In return, the geodetic survey of these orbits enables the observation of mass transfers in the Earth-system (see Sect. 1.1.3). Developed as such, the Gravity Recovery And Climate Experiment (GRACE, see Sect. 1.1.2) provides monthly up to 10-day measurements of the gravity field with global coverage and regional resolution. From the view point of geodesy, the gravitational measurements have to be reduced by short scale mass transfers in oceans, atmosphere, cryosphere and hydrosphere, to separate them from geodynamics or long-term changes in the geosphere.

Research fields that belong to the "residual" signals profit from the possibility to detect respective mass variations on the Earth's surface. As one of these fields, in hydrology it is an important aim to understand the mechanisms of water transports on the continents (see Sect. 1.1.1). Water is a vital resource, but its quantity on the globe is critical. Approximately 3% of the total water available on the planet is fresh water (Baumgartner & Liebscher, 1990). A large proportion of that is bounded as ice in polar regions. Due to the fast increase of human population and climate warming, water availability decreases in many regions. Knowledge of the development of fresh water resources is of significant interest. On the other hand, extreme precipitation or runoff events on the continents threaten the life of humans and animals. Hence, knowledge about the possible enhancement of hazard due to climate change are relevant for the protection of civilised regions. In order to achieve this knowledge, the observation of water discharge and storage processes is necessary to provide reliable predictions for future water availability and hazardous risks. In hydrology, research is undertaken on different scales, like small catchments of a few hundred meter up to the scale of large river basins. But a global observation and understanding of the water cycle is necessary to gain a broad overview on the hydrological system, to link it to other subsystems like oceans or atmosphere and to couple it with climate studies. Widely applied

and useful tools depict large-scale hydrological models, as the conceptual WaterGAP Global Hydrological Model (WGHM, see Sect. 1.2). But the application of such models also increases the need of global hydrological observations to validate model simulations or calibration.

The link between Hydrology and Geodesy is the gravity attraction of water mass, that causes a change of the gravity field (see Fig. 1.3) when the water cycles through the Earth's subsystems (see Fig. 1.1). Geodetically recovered water mass variations by GRACE deserve as an input for hydrological process studies. Furthermore, they are applicable as model input and to discover simulation uncertainties, that reflect a lack of our knowledge about the water cycle mechanisms. In order to take these challenges, the aim of this work is the integration of water mass variations from global GRACE gravity fields into the global hydrological model WGHM (see Sect. 1.4), by a sophisticated model optimisation method (see Sect. 1.3). A main difficulty depicts the analysis and the combination of data from both research fields, with e.g. different spatial resolution. The latter is realised in a multi-objective re-calibration of WGHM model parameter with river discharge data and GRACE-based mass variations (see Sect. 1.5).

1.1.1 The global water cycle

Water on the Earth is not fixed to certain storages of the planet, but it circulates between oceans, atmosphere and the continental surface as a medium of energy fluxes. The global water cycle (see Fig. 1.1) is a complex system on different scales, and it is a main force of life on our planet (Jones, 1997; Baumgartner & Liebscher, 1990) and it is closely linked to the global climate.

The available energy for water transport on the Earth depends on the radiation intensity of the sun and the absorption capacity of the ground. Depending on energy fluxes, water changes its phases between ice, liquid and vapour. Hence, water evaporates from the oceans and from land sites, to precipitate as rain and snow back to oceans and land. On the continents, the liquid may be temporarily stored in soil, plants, lakes, wetlands or groundwater storages. Ice and snow accumulates in glaciated or cold regions and during melt seasons liquid water transfers to rivers. Not evaporated or stored water on the continents fluxes back to the oceans, which occurs

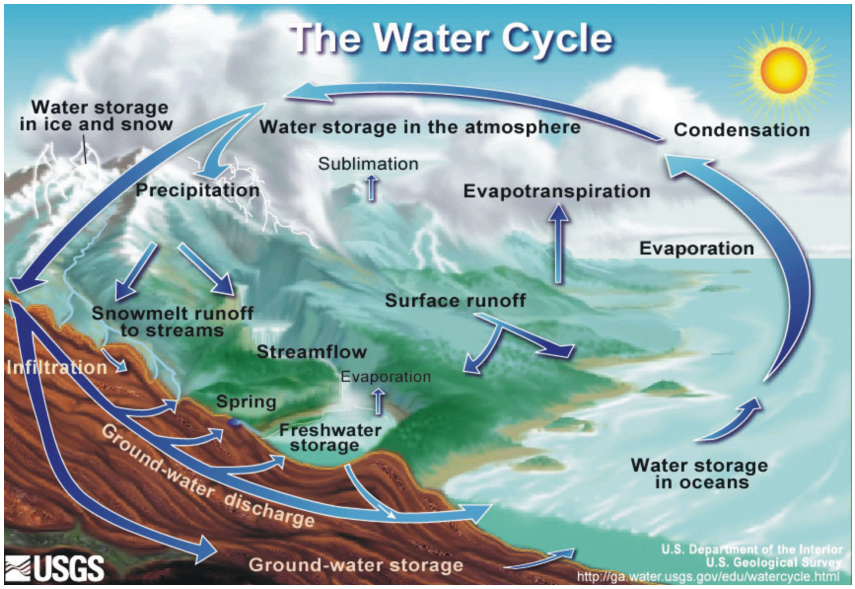


Figure 1.1: The global water cycle.

above the surface along rivers or below the surface through soil and groundwater aquifers. The individual processes of water transports depend on climatic conditions and the constitution of the ground, like vegetation coverage, soil characteristics, elevation, disposition or roughness and they exhibit large differences in their temporal behaviour. The continental water cycle includes hydrological water transports on the land sites, and its dynamics determine the availability of fresh water for human consumption on the globe. Especially, soil moisture represents a key parameter of the hydrological cycle, since it links the energy cycle of evapotranspiration with the subsurface processes of infiltration and solute transports (Ilk et al., 2005).

The importance of fresh water for human makes the continental water cycle and its variations in particular an important subject of research studies. A fundament in its survey is the hypothesis of mass conservation. For any time interval, it asserts that the mass volume entering a defined space body is equal to the volume exiting plus the storage changing in the space body, no matter of the space size and the interval length (Baumgartner & Liebscher, 1990). Following that principle, the dynamic water budged for the drainage basin of a river is described by

$$\frac{\delta S}{\delta t} = \frac{\delta P}{\delta t} - \frac{\delta ET}{\delta t} - \frac{\delta R}{\delta t}, \tag{1.1}$$

with the change of water storage S, precipitation P, evapotranspiration ET and total discharge R for an infinitesimal time element δt. The static approach

$$\Delta S = P - ET - R \tag{1.2}$$

describes the total water storage change TWSC (ΔS) of the basin for a specific time interval Δt. On the global scale, P and R are accessible by various ground or space based measurements. ET may be determined from satellite observations followed by physical modelling of the energy balance (Sheffield et al., 2009). For a long time only ΔS (or S) was not directly accessible by measurements at large scales, since it accumulates from all existing water storages on and below the surface between canopy and the deepest groundwater storage. Therefore, the total water storage change is simulated on large scales by hydrological models, that imitate the continental water cycle with different strategies (see a comparison of three global models in Chapter 2). The only independent data set to validate these models is available from satellite gravity missions, which indirectly detect water storage variations by their mass change induced effect on the gravitational field of the Earth.

1.1.2 Satellite gravimetry by GRACE

Gravitation is the attraction between mass bodies. As every large mass body in space, the Earth features a measurable gravitational field, that forces the Moon as well as artificial satellites staying on an orbit around the Earth. The gravitational potential V at a location (x, y, z) in the field (Heiskanen & Moritz, 1967; Torge, 2003)

$$V(x, y, z) = G \cdot \iiint_{Earth} \frac{dm}{l} \tag{1.3}$$

is an integrated effect of all mass elements $dm = \rho dV$. These elements may be described by their density ρ and volume element dV and they are of distance l to the location (x, y, z). Consequently, the gravitational field is induced by the density distribution of the planet. Changes in the density, e.g., due to mass transfers in or between the Earth's subsystems (see Fig. 1.3) lead to changes in the field. In satellite gravimetry, the inverse effect of mass changes on satellite orbits is used to measure the global gravity field.

In satellite geodesy, a spherical harmonic (SH) expansion of the time dependent gravitational potential V is applied, which can be written outside the Earth's surface ($r \geq R$) as (Torge, 2003; Schmidt et al., 2008b):

$$V(r, \lambda, \theta, t) = \frac{GM}{r} \sum_{n=0}^{\infty} \sum_{m=0}^{n} \left(\frac{a}{r}\right)^n \tilde{P}_{nm}(\cos\theta) \left[C_{nm}(t) \cos m\lambda + S_{nm}(t) \sin m\lambda\right], \tag{1.4}$$

where a is the semi major axis, G the gravitational constant and M the total mass of the Earth. The colatitude θ, the longitude λ and the radius r are geographical coordinates that belong to an Earth-fixed reference frame. $\tilde{P}_{nm}(cos\theta)$ denotes the fully normalised associated Legendre-polynomials, with the integers n and m being degree and order of the spherical harmonic expansion. The dimensionless spherical harmonic coefficients C_{nm} and S_{nm} are the gravitational parameters which describe the mass distribution of the Earth. Their lower terms can be interpreted as physical properties of the Earth (Schmidt et al., 2008b). For example, the C_{00}-term defines the total mass and C_{20} the flattening of the planet.

In satellite gravimetry, the SH-coefficients are determined by the analysis of satellite orbits. Due to the decreasing sensitivity to spatial variations in the gravitational signal with increasing distance from the Earth, so called low-earth orbiting (LEO) satellite missions as the Gravity Recovery And Climate Experiment (GRACE) are launched to an altitude of about 500 km are. GRACE is the direct successor of the first GFZ-1 and CHAMP (CHAllenging Minisatellite Payload) gravity missions. Started in 2002, it consists of two satellites that chase each other in about 220 km distance on the same polar orbit with an inclination of 89,5°. The mission configuration (see Fig. 1.2) enables low-low satellite to satellite tracking (SST-ll), which is realised by accurate (μm-level) quasi instantaneous distance measurements with a microwave K-band instrument between the two LEO's (Schmidt et al., 2008b). The reference of the measurement to an Earth fixed frame is realised by on-board GPS receivers, that absolutely track the satellite orbits with mm-resolution. The orientation of both space-crafts is undertaken with observations of a mrad-accurate two star camera assembly (Reigber et al., 2005). Non-gravitational forces on the satellites, e.g. from atmospherical friction, are determined in three directions with a capacitive accelerometer, that is located in the mass center of each satellite. Furthermore, a laser retro-reflector is installed at the bottom of each satellite to enable laser range measurements for calibration of GPS and K-Band data (Schmidt et al., 2008b). Though, the expected lifetime of the satellites was 5 years, optimistic estimations expect to receive data for several more years, which depends on the survival of the individual measurement instruments as well as on remaining fuel resources necessary to prevent a critical orbit-height decline.

The analysis of the raw data is mainly undertaken by three scientific teams at the German Research Center for Geosciences (GFZ) in Germany, the Center for Space Research (CSR) in Texas, USA and the Jet Propulsion Laboratory (JPL) in California, USA. Here, gravitational

field and other unknown parameters of the measurement system are derived in a least-squares adjustment of the satellite equations that base on the Newtonian equations of motion. Data of about one month are necessary to achieve a spatial resolution of maximal 400 km, which corresponds up to degree and order 50 of the SH (Schmidt et al., 2008b). Consequently, GRACE provides information of mass change induced variations in the gravitational field averaged for one month. These variations can be expressed by monthly averaged changes of the gravitational parameters relative to a reference, e.g. a mean field, by:

$$\overline{V}(r,\lambda,\theta,\Delta t) - V_{ref}(r,\lambda,\theta) = \overline{\Delta V}(r,\lambda,\theta,\Delta t) = f\left(\overline{\Delta C}_{nm}(\Delta t), \overline{\Delta S}_{nm}(\Delta t)\right), \tag{1.5}$$

if $\Delta t = 1$ Month. Only at the cost of spatial accuracy, some processing centers started to provide gravity fields with 10-days or weekly temporal resolution (e.g. Lemoine et al., 2007).

Followed from the formulations above, the GRACE signal integrates from all mass variations on the Earth's surface. Fig. 1.3 provides an overview of GRACE's spatial and temporal sensitivity towards mass changes of the Earth. The separation of the gravitational signal from one of these systems is only possible with further information on mass variations of the other system. Therefore, known proportions of the gravitational influences are a-priori estimated by geophysical models and reduced from the GRACE SH coefficients during the gravity field computation. These known effects are due to gravitational attraction of solar system bodies, the luni-solar lunar tides of the Earth system (solid Earth, oceans and atmosphere) including loading and deformation effects, monthly as well as sub-monthly mass redistributions in the Earth's subsystems of oceans and atmosphere as well as effects from variations in the Earth's rotation (for details see Schmidt et al., 2008b).

The previously unrivaled accuracy of the GRACE mission for the determination of the gravitational field enables the detection of further signals, that remain unmodelled in the GRACE gravity fields. This concerns mass changes of continental hydrology, post glacial rebound, the cryosphere as well as seismic deformation of the solid Earth from earthquakes. But a draw-

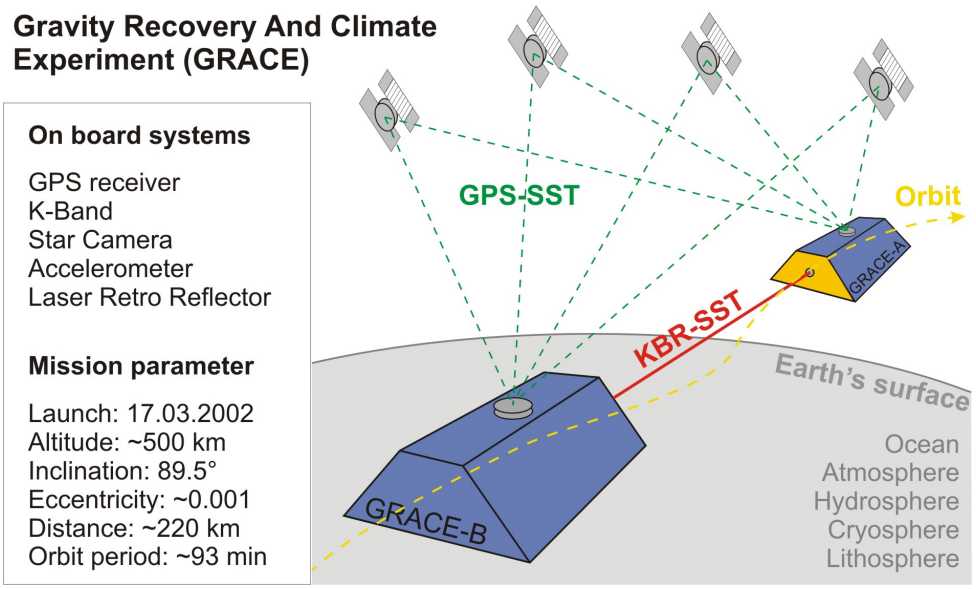


Figure 1.2: Configuration principle of the GRACE satellite mission. After (Schmidt et al., 2008b).

back for the mass change recovery (see next section) from theses signals are uncertainties in the GRACE data that constitute from their limited spatial resolution as well as errors in the GRACE instrument data (e.g. accelerometer), parametrisation deficiencies or errors in the applied geophysical models (Schmidt et al., 2008b). These uncertainties generate spurious gravity signals that indicate correlated errors in the models, i.e. their SH coefficients (Kusche, 2007) and that become visible by characteristic north-south-stripe artifacts in maps of GRACE based gravity fields. Since errors of SH coefficients increase with increasing degree (i.e. increasing spatial resolution) filtering methods are introduced, that enable a smooth down-weighting of SH coefficients with increasing degree (Swenson & Wahr, 2002). The weighting factors are represented in the SH domain by filter coefficients $w_{nm} = [0; 1]$ that are directly applied to the SH coefficients:

$$\left\{ \begin{array}{c} C^w_{nm} \\ S^w_{nm} \end{array} \right\} = \left\{ \begin{array}{c} w^c_{nm} \cdot C_{nm} \\ w^s_{nm} \cdot S_{nm} \end{array} \right\}. \tag{1.6}$$

In the literature, several isotropic, non-isotropic or de-correlating filter methods are available and user of GRACE data may decide for a method that generates an optimal signal-to-noise relation in the smoothed data. This is is described in Chapter 2.

1.1.3 Hydrological prospects from satellite gravimetry

The mass volume of the seasonal water cycle depicts the largest of the unmodelled contributions in the GRACE gravity field (Tapley et al., 2004a). Furthermore, temporal scales and spatial occurrence of the unmodelled signals are sufficiently dissimilar, which is important to split the integrative gravity signal into mass variations from hydrology and other subsystems of the Earth.

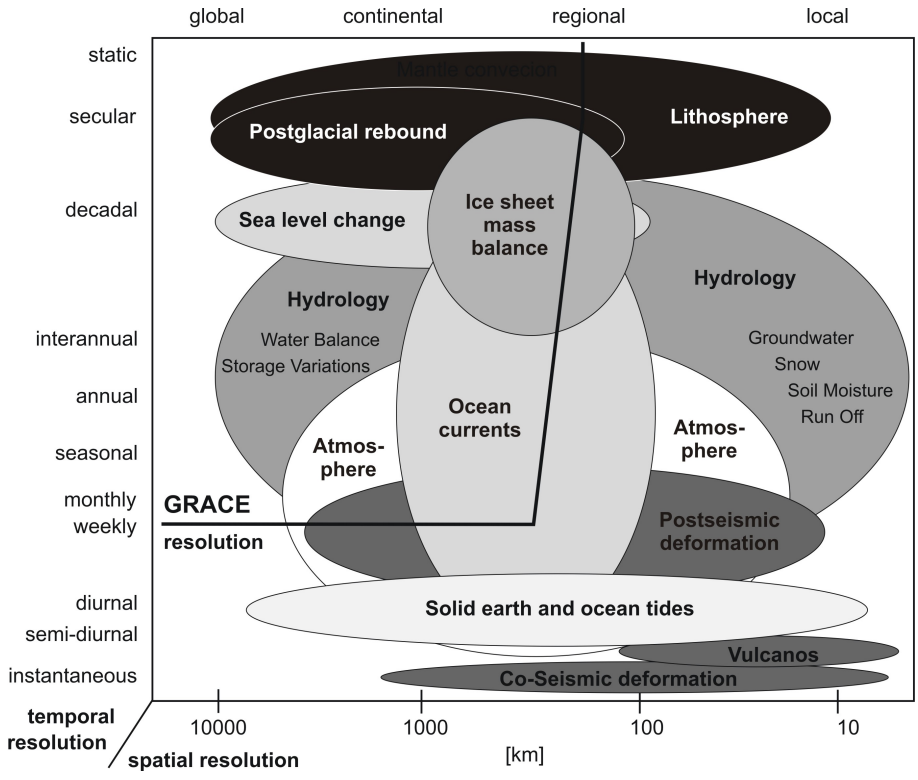


Figure 1.3: Temporal and spatial scales of geoid variations Ilk et al. (2005)

For example, ice mass changes occur mainly in the polar regions, while hydrology is of main interest for the inhabited continents. Solid earth processes are either instantaneous (seismic deformation) or secular (post glacial rebound), and therefore, they can be separated from the non-secular proportions of the hydrological signal. A separation of hydrological trends from post glacial rebound would be necessary for e.g., Scandinavia in Europe and for the North American Great Lake region. But long term changes in hydrology are not of subject in the present study. Therefore, GRACE depicts the first large-scale monitoring system of continental water storage variations ΔS (Ilk et al., 2005), that enables the closure of the continental water cycle (e.g., Sheffield et al., 2009) and the study of hydrological mass transports on the continents (see an overview from Ramillien et al., 2008b) with unrivaled accuracy. GRACE-based estimates of ΔS are a useful tool to validate and calibrate large-scale hydrological models (Güntner, 2009). Therewith, the combination of satellite gravimetry and hydrological modelling shall lead to an improved understanding of the Earth's water cycle in order to better analyse and predict developments and consequences of the changing climate onto freshwater resources or hazardous potentials for different regions.

Numerous existing large-scale or global hydrological models (see Dirmeyer et al., 2006; Widen-Nilsson et al., 2007; Liu et al., 2007, 2009; Milly & Shmakin, 2002a; Rodell et al., 2004b) differ in terms of spatial and temporal resolution, model strategies, number and type of parameters or input data, but they may be classified into two main types. On the one hand, land surface models simulate energy and water fluxes between the Earth's surface and the atmosphere by physically-based heat and mass balance equations. State-variables of these models are usually tuned by a direct integration of measurements as parameter or system states into the model, e.g. with Kalman-filtering. A land surface model widely used within GRACE studies is the Global Land Data Assimilation System (GLDAS) (Syed et al., 2008). On the other hand, water balance models more completely represent the water cycle based on more conceptual equations and they mainly developed to simulate the streamflow of a river basin (Güntner, 2009). Here, parameter tuning is done by model calibration, which denotes the selection of model parameter values by evaluating the simulation performance via a model output objective against observations. In contrary to data assimilation, the system is tuned by determining model parameter values during a pre-defined time interval, and the resulting parameter set may be used for subsequent independent model runs. An example for a land surface model with calibrated parameters is the Land Dynamics model (LaD) (Milly & Shmakin, 2002a). Comparative studies (see Chapter 2) have shown large differences between existing large-scale hydrological models of various types. GRACE data will help to understand these discrepancies to reduce errors in the models and therewith increase our knowledge about the water cycle mechanisms.

In order to derive monthly water mass variations from GRACE, surface density variations are recovered from the SH coefficients by (Wahr et al., 1998; Swenson & Wahr, 2002):

$$\overline{\Delta\sigma}(\theta,\lambda,\Delta t) = a\sum_{n=0}^{\infty}\sum_{m=0}^{n} K_n \tilde{P}_{nm}(cos\theta)\left[\overline{\Delta C}^{w}_{nm}(\Delta t)\cos m\lambda + \overline{\Delta S}^{w}_{nm}(\Delta t)\sin m\lambda\right] \tag{1.7}$$

$$\text{with } K_n = \frac{\rho_E}{3}\frac{(2n+1)}{(1+k_n)}. \tag{1.8}$$

Where the average density of the Earth ρ_E is 5517 kg/m^3 and the load love numbers k_n describe the elasticity of the Earth (Farrell, 1972). $\overline{\Delta C}^w_{nm}(\Delta t)$ and $\overline{\Delta S}^w_{nm}(\Delta t)$ represent the filtered gravitational parameter as difference to a mean field and averaged for $\Delta t = 1$ Month. $\overline{\Delta\sigma}$ may be expressed as mass equivalent to a water column [1mm w.eq. = 1kg/m^2].

The measurements errors and the lack of small scale information decreases for regional averages of GRACE data (Swenson & Wahr, 2002). Therefore, a more reliable estimate of water mass variations is given by a regionally averaged surface mass density. The average may be obtained in the frequency domain by the following formula (Swenson & Wahr, 2002), applied throughout this work:

$$\overline{\Delta\sigma}_{region}(\Delta t) = \frac{a}{\Omega_A}\sum_{n=0}^{\infty}\sum_{m=0}^{n} K_n \left(\vartheta^c_{nm}\overline{\Delta C}^w_{nm}(\Delta t) + \vartheta^s_{nm}\overline{\Delta S}^w_{nm}(\Delta t)\right). \tag{1.9}$$

The coefficients ϑ^c_{nm} and ϑ^s_{nm} describe the spatial characteristics of the regional averaging function $\vartheta(\theta, \lambda)$, which is 1 inside and 0 outside the region of interest. Ω_A represents the angular area of the investigated region (i.e., the area value of the region on a unit sphere).

$\overline{\Delta\sigma}_{region}$ is equal to the monthly TWSC ΔS of a river basin (see 1.2), if the regional function describes the shape of that basin. But the user has to keep in mind the limited spatial resolution of GRACE. Practically, the ineluctable smoothing of GRACE coefficients in Eq. 1.9 applies also to the coefficients of the regional averaging function. The incomplete information on small scales results in a fuzzy or non discrete realisation of the averaging function, which may be larger than 0 outside and smaller than 1 inside the basin. No matter if the spatial average is undertaken in the frequency (as in Eq. 1.9) or in the spatial domain, the limited spatial accuracy results in a leakage error in GRACE-based mass estimates. The error accumulates from the influence of signals outside and the non-unity weighting of the signal inside the region of interest and prevents a clear spatial separation of mass variations. This cutback of GRACE data has to be considered for any hydrological application of the satellite observations (see also Chapter 2).

1.2 The WaterGAP Global Hydrological Model (WGHM)

The WaterGAP Global Hydrological Model (WGHM) was developed by Kaspar (2004) and Döll et al. (2003) as hydrological component for the water use model WaterGAP (Alcamo et al., 2003). The original application of WGHM is to provide information about water availability on large-scales and the determination of water stress within river basins (Alcamo et al., 2003). Therefore, the heuristic water balance model simulates the most important components of the continental water cycle, which constitute to the total water storage change of WGHM:

$$\Delta S_{WGHM} = \Delta S_{c(canopy)} + \Delta S_{sn(snow)} + \Delta S_{sw(surface)} + \Delta S_{g(groundwater)} + \Delta S_{s(soil)} \tag{1.10}$$

$$\Delta S_{sw(surface)} = \Delta S_{l(lakes)} + \Delta S_{wl(wetlands)} + \Delta S_{r(river)} \tag{1.11}$$

A daily water balance is calculated for the land fraction of each 0.5°-grid cell. A vertical balance simulates interception by canopy, snow accumulation and water throughfall to soil. In the lateral water balance groundwater transport and surface runoff are computed (see a routing-scheme in Fig. 1.4). Each grid cell of WGHM is located inside the river network of the global drainage-

direction map DDM 30 (Döll & Lehner, 2002) and provided with information about land cover, soil properties, hydrogeology as well as reservoirs, lakes and wetlands from different data sets (see Fig. 1.4, Hunger & Döll, 2008; Kaspar, 2004). The model is forced by monthly 0.5°-gridded climate data for precipitation, temperature, cloudiness and number of wet days. Daily values for rainfall are distributed synthetically as a two-state, first order Markov-chain (for details see Kaspar, 2004). The conceptual formulations of WGHM represent physical processes of the water cycle in a simplified manner. Therefore, WGHM parameters are not measurable but clearly set to different processes along the water path. The most important and relevant parameter of WGHM 2.1f (Hunger & Döll, 2008) have been considered for calibration of the present study. They are listed in Chapter 4, Table 4.1 together with their standard values and uncertainty ranges. Below, a brief description of the model equations is given. Parameter abbreviations in brackets refer to Table 4.1.

Atmospherical influences The actual evapotranspiration E_c describes the real water transfer from the surface to the atmosphere. But in WGHM, the potential evapotranspiration E_p has to be calculated ahead. E_p constitutes from evaporation and transpiration. The first is the physically drive transfer of water between the uncovered ground and the atmosphere. The latter describes the water transfer due to variable physiological characteristics of the growth (evaporation from

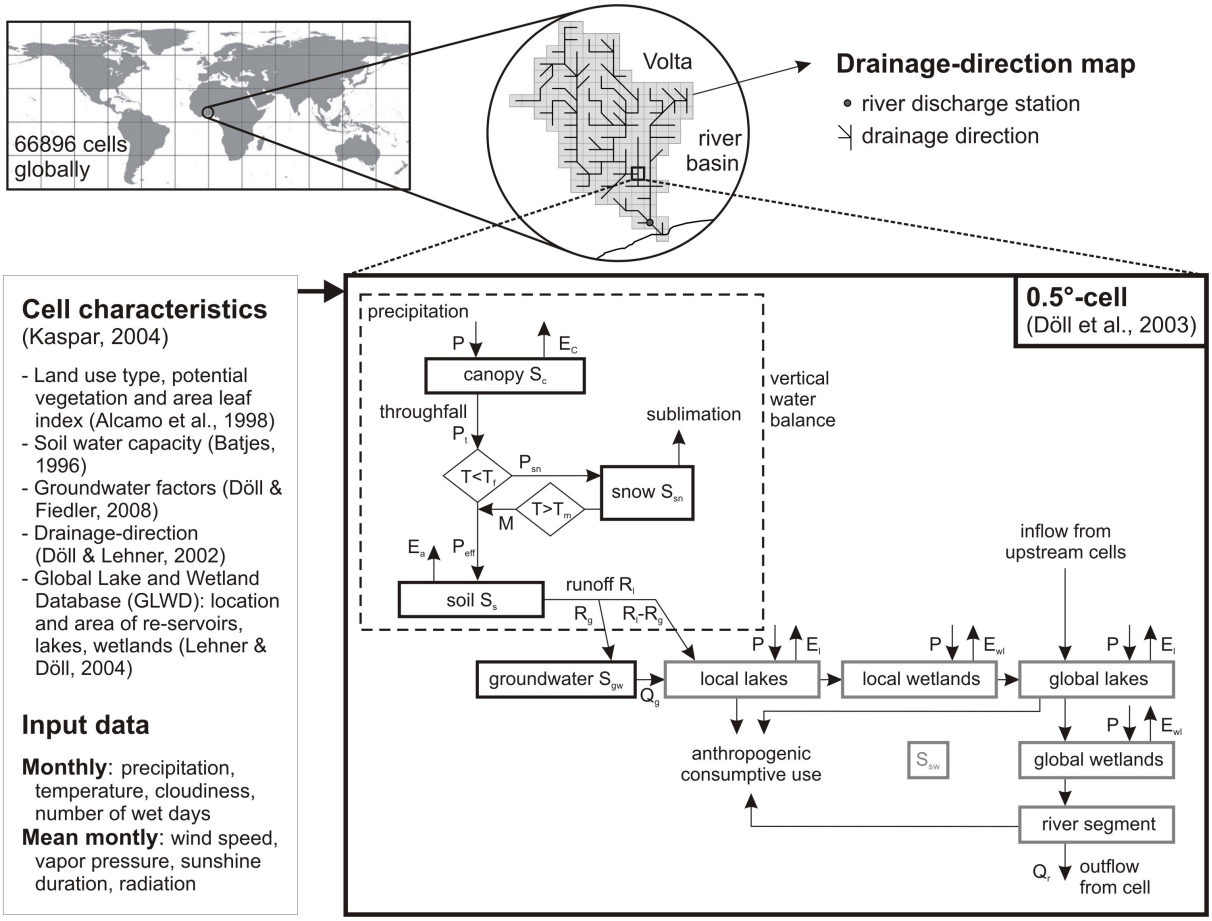


Figure 1.4: Scheme of the WaterGAP Global Hydrological Model (WGHM, Döll et al., 2003; Kaspar, 2004; Hunger & Döll, 2008) and its input and forcing data.

plants). From open water bodies, only potential evaporation occurs, but for all other areas transpiration has to be taken into account.

The potential evapotranspiration E_p isolates the influence of atmospherical conditions onto the evapotranspiration processes from the growth characteristics and in many hydrological models it provides a basis to derive E_c. In WGHM, E_p quantifies the amount of water, that may evaporate for the given atmospheric conditions from an idealised area-unit covered by grass and with an albedo of 0.23. The computation of E_p is realised by the Priestley-Taylor equation (Priestley & Taylor, 1972):

$$E_p = \alpha_{PT} \frac{\Delta}{\Delta + \gamma_{psy}} (R_n - G), \tag{1.12}$$

where Δ is the grade of the saturation vapour pressure that depends on the actual temperature and which is applied according to Shuttleworth (1993). γ_{psy} is the psychrometer constant determined by atmospheric conditions (Kaspar, 2004). The soil heat flow G is neglected, as recommended by Shuttleworth (1993). The Priestley-Taylor coefficient α_{PT} is an empirical parameter. It differs between humid (parameter ER-5 in Table 4.1) and wet (ER-6) areas and in the original WGHM its values are taken from Shuttleworth (1993). R_n represents the net radiation, its formulations belong to the Priestley-Taylor approach and they are explained in detail by Kaspar (2004). Exemplarily, R_n is determined by temperature T, sunshine hours n or the surface albedo α as well as by parameters for radiation proportion a_s (ER-1) of the global radiation that reaches the Earth's surface and the radiation correction for cloudiness a_c. The latter is differentiated for humid (ER-2) and arid areas: $R_n = f(a_s, a_{c,humid}, a_{c,arid}, T, n, \alpha)$.

Canopy water balance Subsequent to the determination of atmospherical influences on evaporation processes, WGHM begins the simulation of the water cycle itself. The contact of water with the ground occurs in the vegetation. Here, the surface of the plants constitutes a canopy storage S_c. Its maximum capacity is $S_{c,max}$ in WGHM. Consequently, rain P_t either falls through the canopy storage, if $Sc \geq Sc_{c,max}$. Or the precipitated water stays inside the canopy, if $Sc < Sc_{c,max}$, then $P_t = 0$. Subsequently, evaporation occurs from the interception water, which is computed by (Deardorff, 1978) in WGHM:

$$E_c = E_p \left(\frac{S_c}{S_{c,max}} \right)^{\beta} \tag{1.13}$$

In case, the canopy storage is not filled completely, the canopy evaporation exponent β determines the leaf enclosure by water representing an area proportion. From this proportion the potential evaporation is taken. The maximum canopy storage is computed from the maximum canopy water height m_c (IN-1) and the leaf area index:

$$S_{c,max} = m_c \cdot LAI. \tag{1.14}$$

The leaf area index depends on two further parameters, the biomass multiplier (IN-3) and the specific leaf area multiplier (IN-2). The leaf mass is derived from the input data sets (see Fig. 1.4). Here, a minimal and a maximal LAI value is computed to enable a differentiation of growth and non-growth seasons depending on temperature and precipitation (Kaspar, 2004).

Snow accumulation Water, that falls through interception stores to the snow storage S_{sn}, if the actual temperature T is below the snow freeze temperature T_f (SN-1). Sublimation from snow is computed similar to the potential evaporation (Kaspar, 2004) but with a snow albedo (ER-3). The snow storage remains filled until the actual temperature rises above the snow melt temperature T_m (SN-2) and snow melting initiates:

$$M = K_s \cdot (T - T_m) \tag{1.15}$$

Herein, K_s is the degree-day factor (SN-3), that determines the proportion of snow melt to the difference of T and T_m. To consider the spatial heterogeneity of elevation and temperature within a 0.5°cell, the WGHM snow algorithm is interpolated on a 30" subgrid resolution (Schulze & Döll, 2004). A temperature gradient parameter (SN4) determines the decrease of the actual temperature with increasing sub-cell height.

Soil water balance The remaining effective precipitation $P_{eff} = P_t - P_{sn} + M$ enters the soil water balance, that is modelled as one layer in WGHM and where processes of evaporation E_a and runoff from the landside R_l occur:

$$\frac{dS_s}{dt} = P_{eff} - R_l - E_a \tag{1.16}$$

The actual evaporation from the soil depends on a parameter for maximum potential evapotranspiration $E_{p,max}$ (ER-7), the soil water content S_s in the effective root zone and the total available soil water capacity $S_{s,max}$:

$$E_a = min\left((E_p - E_c), (E_{p,max} - E_c)\frac{S_s}{S_{s,max}}\right). \tag{1.17}$$

By this, the actual evaporation is proportional to the saturation of the soil ($S_s/S_{s,max}$), but it can not be larger than the potential evapotranspiration reduced by canopy evapotranspiration. The latter substraction is necessary to sustain the energy balance. The maximum soil storage is derived thought the land-cover specific root depth and the water capacity of the belonging root zone: $S_{s,max} = m_{droot} \cdot d_{root} \cdot C_s$. These parameter are taken from input data sets (see Fig. 1.4). For model calibrations, the root depth is calibrated by the multiplier m_{droot} (SL-1), which is 1 in a normal simulation.

The second soil process, runoff from landside, depends on effective precipitation as well as on the saturation of the soil. According to Bergström (1995):

$$R_l = P_{eff}\left(\frac{S_s}{S_{s,max}}\right)^{\gamma \cdot m_\gamma} \tag{1.18}$$

Runoff strongly depends on the runoff coefficient ($0.3 \leq \gamma \leq 3$). In presence of a rain event, $\gamma < 1$ leads to a fast increase of runoff, even if the soil storage is empty. In contrast precipitated water will first saturate the soil and slowly lead to an increased runoff, for $\gamma > 1$. m_γ (SW-1) is a neutral factor of 1 in the original WGHM version (see its function in Sect. 1.5). The proportion of effective precipitation that is not discharged via surface runoff directly enters the soil storage S_s. Here, an exception is made for cropped areas (given by the land cover data set), where soil

infiltration is set to zero and the water is directly diverted to open water bodies. Exchange of soil and groundwater (capillary rise) is not taken into account, because the lack of information about groundwater tables on the global scale (Döll et al., 2003).

Groundwater transport The groundwater model was developed by Döll et al. (2002). Its recharge volume R_g is proportional to the landside runoff (see also Fig. 1.4) but may not exceed the maximal groundwater recharge, characteristic for the soil texture of the present grid-cell:

$$R_g = min(R_{g,max}, R_l \cdot f_g), \tag{1.19}$$

where a comprehensive groundwater factor $0 \leq f_g \leq 1$ derives from $f_g = f_s f_t f_a f_{pg}$. The individual groundwater factors are slope-, texture-, aquifer- and permafrost/glacier-related, respectively. The factors as well as $R_{g,max}$ are taken from the input data sets (see Fig. 1.4). The groundwater outflow Q_g derives from the groundwater storage S_g and the outflow coefficient k_g (GW-1):

$$Q_g = k_g \cdot S_g \tag{1.20}$$

and it is assumed to discharge to the surface water bodies (see Fig. 1.4) because information of groundwater flow paths are not available at the global scale (Döll et al., 2003).

Surface water transport Additional to the inflow from groundwater discharge, lakes and wetlands are filled by landside runoff (reduced for groundwater recharge), the inflow from upstream cells and direct precipitation on the water body area. The water is transported from local lakes to wetlands, further to global lakes and wetlands and finally the river segment (see Fig. 1.4). Local and global water bodies are simulated equally, but their differentiations refers to the extension of global surface water bodies over more than one grid-cell, which are computed comprehensively. Local lakes and wetlands only get inflow from the cell of their location. Man-made reservoirs are treated as natural-lakes in the applied WGHM version (Hunger & Döll, 2008). Location and area of open water bodies are taken from the input data sets (see Fig. 1.4).

The evaporation of lakes and wetlands is equal to potential evaporation reduced by a factor that considers the decreasing surface area available for evaporation from open water bodies with decreasing water level: $E_{l,wl} = E_p(\alpha_{sw}) \cdot r_{l/wl}$. The potential evaporation E_p (Eqn. 1.12) is computed for an idealised open water body but with a specific open water albedo $\alpha_{sw} = 0.08$ (ER-4), in the original model. The reduction factor

$$r_{l/wl} = 1 - \left(\frac{\mid S_{l/wl} - S_{l/wl,max} \mid}{f_{l/wl} \cdot S_{l/wl,max}} \right)^{3.32} \tag{1.21}$$

was introduced by Hunger & Döll (2008) and leads to increased simulation accuracy of lake level dynamics, because it prevents lake water levels below $-S_{l/wl,max}$ and wetland water levels below 0, when for lakes $f_l = 2$ and for wetlands $f_{wl} = 2$ is applied.

The surface water outflow of lakes and wetlands is proportional to the relation of the actual active storage $S_{l/wl}$ and the maximum surface water storage $S_{l/wl,max}$:

$$Q_{l/wl} = k_{l/wl} \cdot S_{l/wl} \left(\frac{S_{l/wl}}{S_{l/wl,max}} \right)^x \tag{1.22}$$

that is regulated by the outflow coefficient parameter $k_{l/wl}$ (SW-5). The value of outflow exponent $x = 1.5$ for lakes leads to a faster outflow compared to wetlands with $x = 2.5$. The maximum surface water storage is a product from the area of the open water body F and the maximum active storage depth $h_{l/wl,max}$: $S_{l/wl,max} = F \cdot h_{l/wl,max}$. In the original model, the latter parameter values $h_{l,max} = 5$m for lakes (SW-3) and $h_{l,max} = 2$m for wetlands (SW-4).

The outflow of lakes and wetlands enters the river segment, from where water can be transported to the next cell. The outflow from the river storage S_r is given by

$$Q_r = \frac{v_r}{s} \cdot S_r \tag{1.23}$$

with the river velocity v_r (SW-2) and the distance s between two neighbouring grid-cells. The flow direction is provided by the drainage-direction map (see Fig. 1.4). A further calculation step of WGHM accounts the water loss that occurs by human consumption for irrigation or water supply. This water demand is derived from the Global Water Use Model of WaterGAP (Alcamo et al., 2003) and reduced in respective order from the available storages of river, global and local lake within a grid cell. The reduction is done until the water demand is satisfied or memorized for the next day (see details in Döll et al., 2003).

Original model calibration The runoff coefficient γ (Eqn. 1.18) represents the only calibration parameter of the original WGHM (Döll et al., 2003; Kaspar, 2004). It is determined in a global calibration (the latest version by Hunger & Döll, 2008), which is done consecutively downstream for each available measurement station in a river basin, so that the longterm-mean of simulated gauge levels from 30 years fit optimal to the measurements. γ becomes equal for all grid-cells of the sub-basin area upstream to a measurement stations until the next available station or the river source in the river network. The worldwide 1240 applied measurement stations cover about 70% of active drainage areas. For the remaining sites, γ is regionalised. Adjacently, runoff correction factors (a station and an area-based factor) are computed to improve the model performance. A detailed explanation of the original model calibration is given by Kaspar (2004) and a global validation of WGHM simulations by Hunger & Döll (2008).

In a superimposed modus, the calibration results for γ from the original model provide a basis for the calibration work of the present study. But, because of the limited resolution of GRACE data sets, calibration is done at once for a complete river basin. The sub-basin variability of river discharge regulation from the original model calibration is kept by the introduction of the multiplier m_γ (parameter SW-1, see Eq. 1.18), that is valid for the whole river to tune the complete γ-set of the basin. The physical limitations for the parameter space of the runoff coefficient γ (see Kaspar, 2004) are kept as constraints ($0.3 \leq (\gamma \cdot m_\gamma) \leq 3$). The runoff correction factors are not applied for the present study, to sustain a closed water cycle.

1.3 Model optimisation

1.3.1 Computer models and calibration methods

Models represent a simplified copy of the real world. There exists no optimal model (Beven & Binley, 1992), because the purpose of the creator of the model determines its composition. Models are designed to learn how the mechanisms of real processes act together (Fig. 1.5a). Herein, measured phenomenon consider as input of mathematic models, which imitate the reaction of the real world. Model parameters and model structure determine the reaction characteristics of the model system. The quality of the simulations and therefore our understanding of reality can be controlled by a comparison of the model output with respective measurements. Model parameter are derived from this comparison by inversion techniques, e.g. from calibration. Further, feedback may be given to model formulations. This principle of modelling an parameter estimation holds for any research field. A specific difficulty in hydrology is that the number of measurements is usually smaller than the number of parameters. Furthermore, hydrological models are usually non-linear. Therefore, analytic inversion techniques are not eligible for parameter estimations.

In hydrology, parameters are derived with calibration methods, which denote a trial-and-error process of parameter variation and model simulation, until a sufficient simulation accuracy is achieved. For this purpose, a various number of techniques exists. Characteristics of specific automatic techniques, that depict an important criteria for the selection of a technique are:

- Ability of the algorithm to find a global optimum
- Accuracy of the optimum solution
- Computation efficiency, i.e. convergence speed
- Applicability of the method settings for different calibration problems

But in the past years, the problem of parameter equifinality, that arises from similar simulation accuracies for different parameter sets, led to the development of further method characteristics (Gupta et al., 2005):

- Consideration of calibration uncertainties (e.g., Thiemann et al., 2001; Beven & Freer, 2001; Beven & Binley, 1992)
- Ability of multi-criterial calibration (e.g., Yapo et al., 1998; Duan, 2003; Vrugt et al., 2003a)
- Calibration of model structure (e.g., Jakeman & Hornberger, 1993; Wagener et al., 2003; Clark et al., 2008)

To decide for a calibration method the user can select one of the three following main types (Duan, 2003), depending on the specific calibration requirements and available resources:

- Passive search: Methods with gridded or random sampling of the parameter space are highly adaptive but computationally intensive and they inhere slow convergence to find a global optima. Their convergence is proportional to the grid size or the density of the random samples, respectively.

- Controlled random search: Random methods may be speeded up by sophisticated sampling, that learn from previous model evaluations. Examples are gradient-based methods, a search with greater density in promising regions, simulated annealing methods or genetic algorithms. Combined with a multi-start search these methods can inhere fast convergence speeds with medium computational demands.

- Stochastic methods. A probabilistic optimisation is derived with stochastic distributions of parameters. Such methods demand high computational efforts and inhere a low adaptability. They are highly complex but sophisticated, since they consider uncertainties of the calibration. Examples are bayesian or generalised likelihood methods.

Independent from the type of the calibration method, a decision about a good or bad parameter set has to be made. Therefore, objective (i.e. mathematical) functions are applied, that measure the fit of simulations to observations. Examples for such functions are the root mean squared error (RMSE), the correlation coefficient (CC) or the Nash-Sutcliffe-coefficient (NSC, see Chapter 2), which is widely used in hydrological applications.

1.3.2 Main aspects of multi-objective evolutionary algorithms

The application of multi-objective methods is ineluctable, if different aspects of the simulation output should be captured (e.g., peak-flow and base-flow of hydrographs) or several measurements have to be integrated into the process of parameter selection. Such multi-criterial values can only be captured together by the evaluation of model simulations with more than one objective function.

Hence, for a multi-objective or multi-criterial calibration, the selection of "good" parameter sets becomes a multi-dimensional problem. Therefore, the theory of Pareto-optimality (Vincent & Grantham, 1981) was introduced to such hydrological calibration problems by (Yapo et al., 1998). The parameters are partitioned into "good" solutions, i.e. Pareto-solutions and "bad" solutions (see Fig. 1.5b). The Pareto solutions are located along a Pareto-frontier towards the

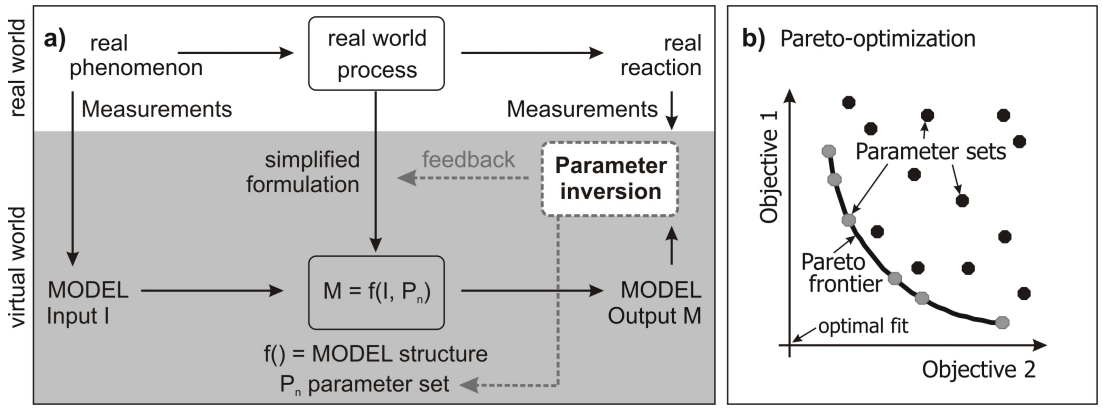


Figure 1.5: a) Principles of modelling and calibration. b) Principle of Pareto-optimisation for a multi-objective calibration problem. The Pareto-frontier is given by the non-dominated Pareto-solutions (gray dots) which are separated from all dominated parameter sets (black dots) of a two-dimensional optimisation problem within the objective space.

optimal fit and they are "non-dominated" (gray dots) over all other "dominated" solutions (black dots). A ranking between Pareto solutions is not possible without additional information or restrictions, because parameter sets that provide a good fit for one criteria, may lead to large errors concerning another. When moving along this frontier, one may increase the fit for one criteria but decrease the fit for at least one other criteria.

As advantages of multi-objective calibration methods, they illustrate the trade-off between opposing objectives, they inhere a smaller likelihood of unbalanced model performance and the calibration results are not too sensitive to a specific measure (Gupta et al., 2003). Therefore, such methods reduce the number of uncertain variables and provide more reliable calibration results.

The higher order of multi-objective optimisation problems, leaded to an agglomerate application of genetic algorithms in hydrology, which feature high computational efficiency and strong convergence abilities (Yapo et al., 1998; Gupta et al., 1998; Reed et al., 2003; Kollat & Reed, 2006). The basic principle of evolutionary methods is the transfer of biological courses like mutation, crossover and selection into the mathematics of model optimisation. Parameter sets are not only varied on a random basis to generate new samples. They are also combined with each other (crossover). Or only one parameter of the whole set is changed (mutation). After model evaluation the best parameters are kept (selection). Herein, elitist methods keep parameter sets, that provide a good fit in an memory archive. This depicts one of three main strategies, that can guaranty the convergence of evolutionary multi-objective algorithms towards the real Pareto-frontier. A second strategy is the usage of an efficient fitness assignment like the determination of non-dominated parameter sets, to separate the Pareto-solutions from the "bad" solutions. And thirdly, the preservation of diversity in the solutions is realised by statistical density estimations of the parameter sets, which prevents that the algorithm traps in on area of the objective space. A successfully validated and applied multi-objective genetic algorithm, that applies an elitist strategy with an dynamic archive, an efficient non-dominating sorting scheme and that eliminates sharing parameter to preserve diversity is the ϵ-Non-dominated-Sorting-Genetic-Algorithm-II (ϵ-NSGA-II Kollat & Reed, 2006; Tang et al., 2006). Herein, the ϵ-dominance archiving enables the user to specify the precision of the objective-function quantification.

1.4 Research objectives

The main motivation to calibrate hydrological models with GRACE gravity data, was due to differences between GRACE-recovered estimates of water mass variations and respective simulations of global hydrological models (Schmidt et al., 2006). The latter often exhibits smaller seasonal amplitudes of mass variations in the water cycle, than the observation-derived values by GRACE, which is specifically true for WGHM simulations (see Introduction of Capter 3 and 4; Schmidt et al., 2008a; Güntner, 2009).

Furthermore, differences between several models are large (see more details in Chapter 2; Güntner, 2009). None of the models fits globally best to GRACE, when looking at correlations of time series for water mass variations of different river basins (see Table 1.1). This situation complicates the explanation of differences between GRACE and specific models as well as the determination of possible model structure errors by inter-model comparisons. Consequently, and

Table 1.1: Correlation correspondence of basin averaged monthly time series of total water storage from the global models WGHM, GLDAS, LaD with GRACE data (2003-2006) with 500 km Gaussian smoothing (Werth & Güntner, 2008).

	Correlation with GRACE		
Basin	WGHM	GLDAS	LaD
Amazon	0.93	0.95	0.84
Amur	0.40	0.35	0.42
Columbia	0.89	0.76	0.86
Danube	0.83	0.80	0.79
Ganges	0.98	0.96	0.94
Huang He	0.53	0.79	0.59
Indus	0.44	0.33	0.67
Lena	0.82	0.59	0.80
Mackenzie	0.90	0.46	0.94
Mekong	0.88	0.90	0.89
Mississippi	0.83	0.86	0.81
Murray	0.59	0.34	0.62
Nelson	0.74	0.59	0.66
Niger	0.95	0.96	0.85
Nile	0.87	0.89	0.84
Ob	0.92	0.56	0.91
Orange	0.56	0.74	0.53
Orinoco	0.97	0.92	0.94
Parana	0.86	0.92	0.81
St. Lawrence	0.88	0.59	0.83
Tocantins	0.96	0.96	0.84
Volga	0.89	0.58	0.80
Volta	0.88	0.95	0.77
Yangtze	0.96	0.77	0.96
Yenisei	0.88	0.47	0.89
Yukon	0.91	0.22	0.88
Zaire (Congo)	0.75	0.58	0.65
Zambezi	0.90	0.94	0.71
Global correlation	0.54	0.49	0.52

what is the basic innovation of the present work, GRACE-based estimates of water storage variations shall be applied to re-calibrate WGHM, in order to achieve model simulations consistent to the satellite observations.

For the region of a specific river basin, WGHM simulations and GRACE data can be compared by:

$$\Delta S_{WGHM} + \epsilon_{struc} + \epsilon_{in} + \epsilon_{leak} = \Delta S_{GRACE} + \epsilon_{meas} + \epsilon_{leak}, \tag{1.24}$$

where ΔS_{WGHM} can be derived from Eq. (1.10) and ΔS_{GRACE} from Eq. (1.9). For GRACE, measurement (ϵ_{meas}) and leakage errors (ϵ_{leak}) have to be considered. The hydrological simulations include model structure errors ϵ_{struc} that derive from an incomplete or mis-modelled representation of the water cycle as well as uncalibrated model parameter. Furthermore, model input errors represent measurement errors of input and calibration data like precipitation, temperature, runoff or TWSV. Errors in input data may also cause errors in calibrated model parameters. The hydrological modeler has no influence on the size of ϵ_{in} but he can determine its influence on the model parameter and the model output uncertainty. For a comparison of TWSV data sets, model simulations have to be filtered to the same spacial resolution as GRACE. Therefore, WGHM simulations are altered by leakage errors as well. If both data sets inhere the

same leakage error, it simply constraints the spatial resolution of the comparison and may be neglected. A general similarity of leakage errors is likely, since the spatial distribution of seasonal hydrological signals is similar for both data sets (see Chapter 2), though this may differ between the geographical regions.

The intention of this study is the reduction of ϵ_{struc}. For regions where ϵ_{meas} is smaller than $\epsilon_{struc} + \epsilon_{in}$ (and ϵ_{in} is smaller than ϵ_{struc}), GRACE data are applicable to decrease ϵ_{struc} and therefore, improve WGHM simulations of total water storage change. This condition may be given for large river basins with sufficient GRACE data accuracy (see Sect. 1.1.2). For such basins, the calibration results of WGHM by a comparison of GRACE data and model simulations may help to understand possible reasons for mis-modelled processes or structural errors of WGHM. Hence, for the integration of GRACE TWSV into the global hydrological model WGHM, three main research questions arise:

Research questions

- How can GRACE based estimates of TWSV be integrated in a global hydrological model?
- Do GRACE data help to improve large-scale TWSV simulations and if yes, for which regions?
- What can we learn from the results for global hydrological model development?

In response to these questions and in order to achieve a better accuracy for ΔS_{WGHM}, the main objectives that follow for this study are:

Research Objectives

- Derivation of GRACE-based storage change estimates for hydrological applications with best accuracy (Chapter 2)
- Development of a strategy to integrate GRACE data into WGHM (Chapter 3)
- Development of a calibration approach for storage change (Chapter 3)
- Global re-calibration of WGHM by the developed calibration approach (Chapter 4)
- Process analysis of calibration results and detection of possible WGHM model structure errors (Chapter 4)

1.5 Data and computational challenges

The calibration approach for the integration of GRACE data into WGHM is explained in Chapter 3 (see especially Fig. 4.2). To meet the above explained main objectives by the developed calibration approach, specific challenges had to be beared within the present study. The computational challenges and respective consequences are explained in detail by the paragraphs below.

Relative nature of GRACE data GRACE data provide an basin-averaged integrated estimate of water storages changes. Hence, GRACE does not provide data of absolute water storages in a river basin. In a model calibration, solutions for equal signal variations but different absolute values of water storage in the river basin are not distinguishable from GRACE data only. Therefore, absolute measurements as from river discharge should be used for the calibration as well. A multi-objective calibration method enables the combination of several measurements of different nature and increases parameter accuracy (e.g., Vrugt et al., 2003a; Gupta et al., 2005). By such an approach, GRACE data with regional resolution may be combined with station based discharge measurements.

Limited spatial resolution of GRACE data Because of the limited spatial resolution of GRACE, the accuracy of mass variations is only sufficient for large river basins with diameter of several hundred kilometer. Therefore, the 28 largest and most important river basins are selected for a calibration of WGHM (as listed in Table 1.1). Since storage variations of GRACE have to be averaged for whole regions, e.g. a river basin, the model calibration is done river basin wise. River discharge measurements are taken from the last available station of each river basin.

The limited spatial resolution of GRACE demands the application of filtering techniques and causes the GRACE leakage error. While filtering decreases the GRACE measurement errors, it increases the leakage error in the GRACE data. A minimal sum of both is desirable. Therefore, an estimation of optimal filter techniques, that may be different for various river basins, is undertaken for the 28 river basins (in Chapter 2). Furthermore, to ensure equal spatial resolution in the modelled data, a-priori to the comparison of modelled and GRACE-based storage variations, WGHM simulations are smoothed with equal techniques as the GRACE data. For a complete filtering, the signal within the surrounding region of the calibration basin has to be loaded from the original model version and the simulated mass variations are transferred to a SH representation, by new integrated software into the source code of WGHM.

Lack of alternative measurements On the global scale GRACE data and model simulations are the only purely measured estimations for water mass variations on the continents. Consequently, the data accuracy improvements are limited to an iterative approach. Combined atmospherical-terrestrial water balances may provide an GRACE-alternative estimate of TWSV (e.g., Hirschi et al., 2006; Seneviratne et al., 2004). But such water balance studies depend on atmospherical reanalyses and runoff measurements. Therefore, they represent modelled not observed data and due to the usage of runoff they are not independent to hydrological model simulations. The lack of alternative measurements in combination with the limited spatial resolution of GRACE, makes each of the two data set a validation tool for the other. Despite the need of input data for model calibration, GRACE filter methods have to be validated by hydrological model estimations,

because the filtering of GRACE data has a huge effect on the amplitudes of derived time series for mass variations. The estimation of GRACE filter techniques is done by the fit of filtered GRACE data to a filtered mean field of three widely used global hydrological models, including WGHM (see Chapter 2).

Increase model sensitivity to all processes of the water cycle The original model is calibrated against river discharge for one runoff parameter. Since total storage variations inheres water mass variations from other storage compartments besides river water, model parameter from these storages should be calibrated as well. The calibration of further model parameter is only reasonable, because of the usage of additional measurements as calibration data (Savenije, 2009), hence the application of a multi-objective instead of a GRACE-only calibration. In an a-priori model sensitivity analysis for each river basin (Chapter 3 and 4), the most sensitive parameter against storage variations and river discharge are determined. The six to eight most sensitive of the overall 26 parameter are then optimised in the multi-objective calibration. A reduction of the calibration parameter also decreases the number of demanded evaluation runs.

High computational costs WGHM model runs effort high computational costs, as visible from the overview of the evaluation time for different run types in Table 1.2. These temporal statistics limit the number of model evaluations for a calibration of WGHM to 1000 until 2000. Furthermore, the optimisation problem of WGHM for six to eight parameters is highly non-linear and the multi-objective optimisation of WGHM with GRACE and river discharge demands a decision making in two dimensions. To minimise the computational demands, for the calibration of WGHM with GRACE data an efficient genetic and multi-objective calibration algorithm ϵ-NSGA-II is applied, that is able to diverge fast to global optima. The method was coupled with the WGHM software for a multi-objective calibration.

Table 1.2: Computational costs of WGHM model runs for the period [1]) 1992-2007 and [2]) 2002-2007, inclusive 3 initial years as well as spherical harmonic analysis for smoothing and GRACE data comparison for 2003-2007. The employed hardware cluster holds an AMD Opteron CPU of 284 GHz and a RAM of 1 GB.

run type	evaluation time
complete global run[1]	6h, 40min
1 x Amazon[2]	19min
1 x Volta[2]	17min
1200 x Amazon[2]	15d, 22h
1200 x Volta[2]	14d, 3h
1200 x 28 basins[2]	parallel ca. 15d

1.6 Content overview

In preface to the following chapters, an overview of the structure and the content of the present work is provided in Fig. 1.6. The work is divided into overall five chapters. Chapter 1 provides the introductorily theoretical and technical background for the following three main Chapters 2-4. The three following chapters depict the main body of this work. These comprise the derivation of regionally averaged water mass variations from GRACE gravity data with

best possible accuracy (Chapter 2), the development of a calibration technique that enables the integration of GRACE data into the global hydrological model WGHM (Chapter 3) and the global calibration analysis of WGHM (Chapter 4). Each of these parts contributes a piece to the outlined main objectives (Sect. 1.4) and an improved data set of total water storage variations. Chapter 5 summarises the complete work and provides an overall conclusion.

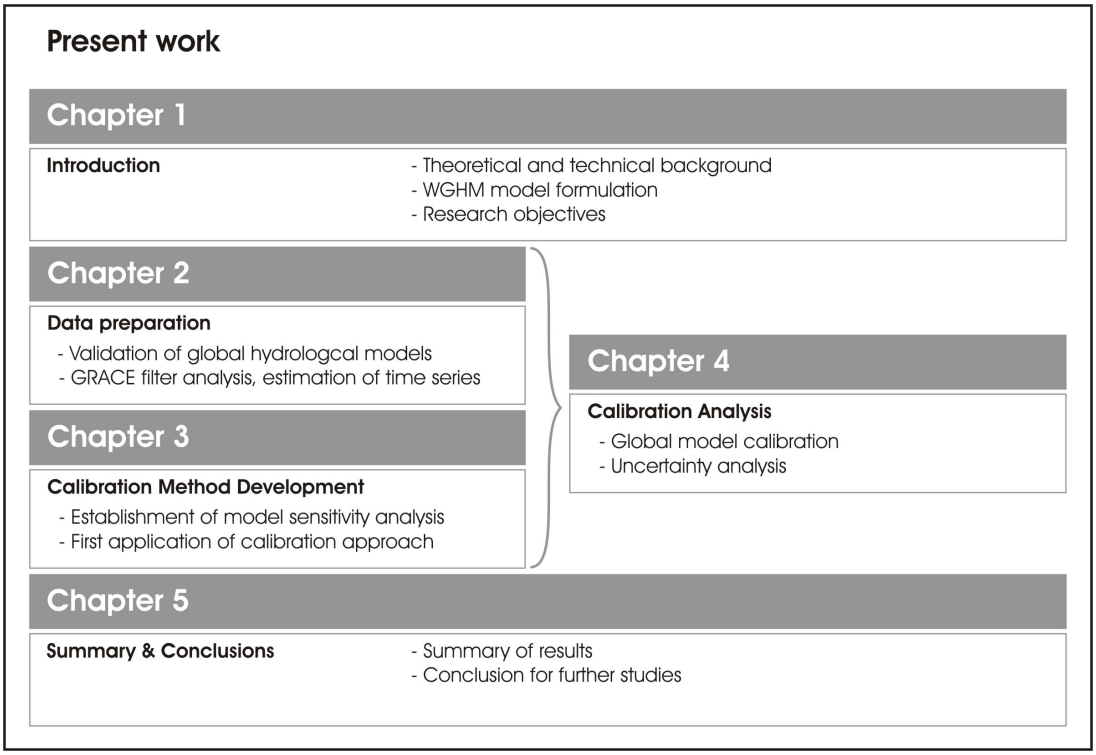


Figure 1.6: Content and structure of the chapters.

2 Data Preparation: "Evaluation of GRACE filter tools from a hydrological perspective"*

Abstract

Approximately seven years of time-variable gravity data from the satellite mission GRACE (Gravity Recovery And Climate Experiment) are available to quantify present-day mass variations on and near the Earth's surface. Mass variations caused by the continental water cycle are the dominant signal component after subtracting contributions from atmosphere and oceans. This makes hydrology a primary area of application of GRACE data. To derive water storage variations at the scale of large river basins, appropriate filter techniques have to be applied to GRACE gravity fields given in a global spherical harmonic representation. A desirable filter technique minimises both GRACE data error and signal leakage across the border of the region of interest. This study evaluates the performance of six widely used filter methods (isotropic filters, anisotropic filters and decorrelation methods) and their parameter values to derive regionally averaged water mass variations from GRACE data. To this end, filtered time series from GRACE for 22 of the world's largest river basins were compared to continental water mass variations from a multi-model ensemble mean of three global hydrological models (WGHM, GLDAS and LaD). Filter-induced biases for seasonal amplitudes and phases of water storage variations, as well as satellite and leakage error budgets, were quantified for each river basin and explained in terms of storage variations in and around the basin. The optimum filter types and filter parameters were identified for each basin. The best results were provided by a decorrelation method that uses GRACE orbits for the filter design. Our ranking between all filter types and parameters depended on the geographical location, shape and signal characteristics of the specific river basin. Based on a multi-criterial evaluation of satellite and leakage error, as well as an error assessment of the hydrological data, the filter selection and parameter optimisation results were shown to be reliable for 17 river basins. The results serve as a guideline for the optimal filtering of GRACE global spherical harmonic coefficients for hydrological applications.

*Werth, S., Güntner, A., Schmidt, R., Kusche, J. (2009b), Geophysical Journal International (accepted).

2.1 Introduction

Global monthly gravity field solutions from the US-German satellite gravity mission GRACE (Gravity Recovery And Climate Experiment) trace mass redistributions close to the Earth's surface (Reigber et al., 2005; Tapley et al., 2004b). Assumptions as defined by Chao (2005) and Wahr et al. (1998) enable the transformation of gravity variations into time series of global maps of surface mass anomalies. Due to their integrative nature, global coverage and previously unrivalled accuracy, GRACE-derived mass variations give insight into processes within the Earth's subsystems. This knowledge helps to improve understanding and modelling of geophysical mass transfers within the Earth's system. For example, recent studies have considered estimations of ice mass loss within polar regions (Chen et al., 2008; Wouters et al., 2008), observations of oceanic circulations (Dobslaw & Thomas, 2007), components of the continental water cycle (Boronina & Ramillien, 2008; Swenson et al., 2008; Niu et al., 2007a) and interactions between these subsystems (Chambers et al., 2007; Ramillien et al., 2008a).

Water mass variations within the continental hydrological cycle are a major signal recovered from the GRACE gravity data after removal of signals from tides, atmosphere and oceans. Numerous studies show an overall good agreement between variations of total continental water storage (TWS) from global hydrological models and from GRACE, especially for large river basins (for a recent overview see Güntner, 2009). Ramillien et al. (2005) and Schmidt et al. (2006) compared output from the WaterGAP Global Hydrology Model (WGHM) and the Land Dynamics model (LaD) to GRACE data and found a good general correspondence. Syed et al. (2008) confirmed this agreement for the Global Land Data Assimilation System (GLDAS). Schmidt et al. (2008c) found similar dominant seasonal and inter-annual TWS periods for GRACE and simulated data of GLDAS, LaD and WGHM for the Amazon, the Ganges and the Mississippi river basins. However, the degree of agreement between GRACE and hydrological models clearly varies with the region and river basin of interest (e.g., Ramillien et al., 2005). Except for a few regions (e.g., Swenson et al., 2006), there is a lack of independent TWS observation data at large spatial scales that are consistent with GRACE. Thus, in spite of the uncertainties inherent in hydrological models, simulation data are currently the only way to evaluate TWS variations from GRACE for large areas.

A method of deriving time series of total regionally-averaged mass variations from global GRACE gravity fields represented as coefficients of spherical harmonics (SH) has been suggested by Wahr et al. (1998). GRACE measurement and processing errors, which are often referred to as satellite errors, mostly distort SH coefficients of high resolution. One way to reduce noise in the monthly solutions is to constrain (or regularise) the coefficients (e.g., towards a mean field) in the course of GRACE data processing (e.g., Save et al., 2008; Watkins et al., 2008; Lemoine et al., 2007). But the need for regularisation characteristics vary widely between different scientific applications of GRACE data (Kusche, 2007). As a result, unconstrained solutions are mostly published by the processing centres, making the application of a post-processing filter technique indispensable. Filtering aims at the suppression of noisy high-resolution coefficients of the gravity field, i.e., smoothing the original data to a lower spatial resolution. Furthermore, decorrelation techniques can be applied to remove striping artefacts of GRACE gravity data, which can be interpreted as realisation of anisotropically correlated noise in the coefficients (Swenson & Wahr,

2006). In this paper, both smoothing and decorrelation techniques are subsumed under the term filtering.

As a drawback, filtering implies the leakage of signals from outside the region of interest into the resulting time series as well as the non-unity weighting of the signal variability inside the region of interest (Swenson & Wahr, 2002; Klees et al., 2007). Hereafter, both effects will be referred to as leakage. According to Swenson & Wahr (2002), leakage varies between different filter types. Klees et al. (2007) listed three simplified cases of leakage scenarios: a) The signal (TWS anomaly) outside the area of interest is of the same sign as the signal inside, b) The signal outside is zero, and c) The signal outside is of a different sign than the signal inside. Using a Gaussian smoother of different filter widths, Klees et al. (2007) concluded that the first case would lead to the lowest total leakage error and the third case to the highest leakage error. Since the signal intensities outside and inside of a river basin vary widely between different regions due to varying hydrological characteristics, the leakage error also depends on the region of interest.

Consequently, in order to select an appropriate filter method, the user has to balance between remaining satellite errors and the spatial resolution (i.e., leakage error), and has to find an optimal balance, specifically, for each river basin they intend to analyse. Filter types developed so far differ in their assumptions on signal-noise properties of the true GRACE-derived mass variations. Some studies evaluate specific filter types. For example, Swenson & Wahr (2002) developed two anisotropic methods and compared them with an isotropic Gaussian filter by evaluation of signal leakage. Han et al. (2005) showed that anisotropic smoothing is necessary to consider the degree and order dependence of GRACE coefficient errors. Seo et al. (2006) described error reductions within GRACE water-mass variations when using a time-dependent noise-minimising filter instead of the Gaussian method. Schrama et al. (2007) determined the radius of the Gaussian smoother with an empirical orthogonal function (EOF) analysis and by comparing it with GPS load measurements. A recent decorrelating filter method and an overview of several GRACE filter techniques was given by Kusche (2007). In order to compensate for the effect of amplitude damping by filtering, Velicogna & Wahr (2006) introduced a scaling factor to recover the full hydrological signal in time series of TWS variations. Similarly, Chen et al. (2007a) used a scale factor to readjust amplitude damping effects caused by the Gaussian filtering of GRACE data relative to TWS from GLDAS simulations.

An optimised spatial resolution of GRACE data by use of an adequate filter algorithm is especially crucial for hydrological studies, where a separation of water-mass variations of different river basins is of high interest for water balance studies. The transport of water masses can be concentrated to small regions like the river network and its inundation areas with low signal correlations to other mass transport processes. In addition, the hydrological signal of interest is composed of mass variations in several water storage compartments of the continental water cycle (such as snow, surface water or groundwater), which differ in their modes of temporal variability or spatial correlation lengths (Güntner et al., 2007b). Hence, particular hydrological features have to be considered when selecting appropriate GRACE filter techniques with small leakage and satellite errors for applications in continental hydrology. Nevertheless, a comprehensive evaluation of various filters from this perspective is missing in the literature so far. In particular, the following questions arise: (1) Which filter is optimal for which scale, location or shape of a river basin of interest? (2) Which filter is superior for which regional signal properties, i.e.,

for different sources of water mass variations in continental hydrology? (3) What are the filter properties in terms of TWS amplitude falsification and phase shifts? (4) Which filter removes striping artefacts sufficiently well? In this study, we address the first three questions from the perspective of hydrological applications such as water balance analysis or hydrological modelling. To this end, we evaluate GRACE filter methods by simulating data from global hydrological models, which at present provide the only alternative data set of TWS variations for large areas.

2.2 Methods and data

To evaluate different filter types, time series of continental water storage variations from GRACE (Sect. 2.2.2) were evaluated using the three global hydrological models WGHM, GLDAS and LaD (Sect. 2.2.3). In the absence of alternative observation data at the relevant scale, hydrological model data were considered the most realistic information on continental water mass variations. An analysis of differences between the models was undertaken to uncover their uncertainties (Sect. 2.3.1). To reduce model-specific errors in the evaluation data set, an ensemble mean of the three models was used for the filter analysis. It was assumed that the reduced GRACE signal used in this study is governed by hydrological processes, and that the GRACE data are corrupted by satellite errors but not by other geophysical processes. The different filter methods (Sect. 2.2.1) were applied to compute time series of water storage variations for selected large river basins, after converting the hydrological fields into a spherical harmonic representation. To assure consistency, GRACE and model data were filtered in the same way. The similarity of measured and modelled TWS time series was evaluated by a correspondence criterion, which is described in Sect. 2.2.4. Computations were repeated for each filter method with varying filter parameters. The optimal filter method of deriving water mass variations was selected for each river basin from the maximum of the correspondence criteria, which is expected when the total error (leakage and satellite error) in the filtered GRACE data is minimal.

2.2.1 Filter methods

In this study, six post-processing filter methods for derivation of regionally averaged water mass variations from GRACE's global gravity field solutions were evaluated. The smoothing of the gravity field can be interpreted by a weighted spatial averaging for a region of every point on the globe in order to reduce noise that disturbs the signal components on higher spatial scales. A short description of the isotropic (degree dependent) filters, the anisotropic (degree and order dependent) filters and the two anisotropic decorrelation methods used in our study, is given below. For details on the filter methods, the reader should refer to the respective original publications. For each filter method, the parameters that define the degree of smoothing strength are explained below (see a list in Table 2.1).

(I) The widely used isotropic Gaussian filter was proposed by Jekeli (1981) as a way of smoothing out the Earth's gravity fields. Its weighting function is derived from the Gaussian probability density function, which has its highest weight in the centre and diverges to zero with increasing distance from the kernel. The form parameter of the symmetric bell-shaped weighting function may be expressed as filter width r_g (eqn. 59 in Jekeli, 1981). r_g represents the radius at which the filter weighting function declines to 50% of its maximum value, and it is used to tune this

degree-dependent smoothing method. (II) Another filtering method was developed by Swenson & Wahr (2002) and applied with degree- and order-dependency for this analysis. The idea behind this filter design is to apply less smoothing to GRACE coefficients with relatively small errors that are relevant to a signal within the region of interest. Hence, a spherical harmonic representation of the basin function is used to compute the filter weights. No direct assumption about the signal is introduced. The user may tune this method by deciding for a total maximum satellite error of basin average Δ_{max} (eqn. 45 in Swenson & Wahr, 2002). To approximate this a-priori fixed maximum satellite error, the filter weights are computed iteratively from a propagation of the smoothed GRACE coefficient errors to the basin average.

(III) Another degree- and order-dependent technique by Swenson & Wahr (2002) minimises the sum of GRACE satellite error and signal leakage. The satellite error is propagated from the GRACE coefficient errors. Signal leakage is estimated by an exponential signal model, which is parameterised by the auto-correlation length G_l and standard deviation σ_0 of the expected geophysical signal (eqn. 41 in Swenson & Wahr, 2002). (IV) Seo et al. (2006) proposed a time-dynamic filter that optimises the signal-to-noise ratio of each GRACE coefficient individually. We applied the method B_4 of their study, which uses the GRACE SH coefficients themselves as a signal estimate. Seo et al. (2006) derived a monthly filter version from the monthly GRACE coefficient errors. For the present study, a static filter was computed from the variance of the monthly coefficient errors. These variances were modified with a dimensionless error factor f, as a means of tuning the filter's degree of smoothing.

(V) Swenson & Wahr (2006) published an empirical decorrelation method that has to be followed by a subsequent application of one of the filter methods explained above. To reduce the correlation between coefficients of the same order but increasing degrees, they fit and remove a quadratic polynomial in a moving window from the coefficients, and they do so separately for even and odd degrees. The moving window is cantered at the coefficient to be filtered. No details on the window size are provided by Swenson & Wahr (2006); therefore, its design orients on Press et al. (1992) for the present study. The size of the window has to be decreased (e.g., with a Gaussian function) for increasing degrees in order to avoid too much signal damping. Consequently, one has to define four parameters for the decorrelation process: the initial and the final window size, w_a and w_e, as well as the degree of the first and last coefficient to be filtered, n_a and n_e. For computations represented below, $n_a = 2$ and $w_e = 3$ were fixed. Discrete variations for $w_a = [10, 20, 30, 40, 50]$ and $n_e = [10, 20, 30]$ were tested. Thereafter, a global filter optimisation described by Chen et al. (2006b), who proposed to maximise the ratio of the spatial signal root mean square (RMS) for ocean versus land, was applied. The three optimised versions of filter V were concluded from a combination of $n_e = [30]$ with $w_a = [10, 20, 30]$, which were used for further investigations.

(VI) Another decorrelation method, by Kusche (2007), makes use of the GRACE orbital geometry and can be interpreted as an anisotropic filter. This method imitates the regularisation of GRACE data processing, using a-priori diagonal signal and dense error covariance matrices. The latter are derived synthetically from GRACE orbits. The filter's degree of smoothing may be tuned by a regularisation parameter $a = 10^x$ of the signal covariance matrix (eqn. 22 in Kusche, 2007). Three filter versions with $x = 12$, $x = 13$ and $x = 14$ were applied in this study.

Table 2.1: Overview of the tested filter methods I-VI.

Method	Variable parameter	Reference
I	r_g	Jekeli (1981)
II	Δ_{max}	Swenson & Wahr (2002)
III	G_l, σ_0	Swenson & Wahr (2002)
IV	f	Seo et al. (2006)
V	w_a, n_e	Swenson & Wahr (2006)
VI	x	Kusche (2007)

2.2.2 GRACE data

Monthly basin-averaged surface-mass variations were derived from GRACE-only global gravity field-model time series generated at GFZ German Research Centre for Geosciences (GRACE Level-2 products, version GFZ-RL04, Schmidt et al., 2008b). These data were obtained from the GFZ Information System and Data Center for a period ranging from 02/2003 until 07/2007 (excluding unavailable months 06/2003 and 01/2004) up to degree and order 120. They consist of unconstrained gravity fields (Flechtner, 2007). Effects of the atmosphere and oceans are removed at the GRACE data centre by applying the appropriate model data. For this study, water mass variations are derived relative to a mean field for the years 2003-2006, and trends were removed from the time series. Coefficients from degree 2 were used, and degree 1 coefficients were set to zero. This is adequate because degree-1 coefficients are also excluded from the hydrological model data used for comparison. The accuracy of GRACE gravity fields varies in time and space. Schmidt et al. (2008b) quantified the global average error of derived water mass variations to 13-15 mm of a water mass equivalent column (w.eq.) for a circular area with a radius of 800 km. For filter parameterisations, estimates of GRACE error covariances were taken from calibrated coefficient errors, which are published together with GFZ-RL04 fields.

2.2.3 Hydrological data

Continental water-storage data provided by three global hydrological models were used for the analyses.

The WaterGAP Global Hydrology Model (WGHM, Döll et al., 2003) is a conceptual global model that simulates the continental water cycles, excluding the regions of Antarctica and Greenland. Modelled water storages include interception, soil water, snow, groundwater and surface water. For this study, data sets were available from 01/2003 until 12/2007 from the most recent version of the model (Hunger & Döll, 2008). WGHM was forced by monthly climate data from ECMWF (European Centre for Medium-Range Weather Forecasts) and precipitation data from GPCC (Global Precipitation Climatology Centre). Output were of 0.5° resolution and were calibrated by tuning a runoff coefficient parameter against observed river runoff at 1,235 discharge stations worldwide. Water storage simulated with WGHM has recently been analysed at the global scale by Güntner et al. (2007b).

The Global Land Data Assimilation System (GLDAS, Rodell et al., 2004b) may incorporate a variety of land-surface models. For this study, the 'National Centers for Environmental Prediction / Oregon State University / Air Force / Hydrologic Research Lab Model' (NOAH, Ek et al., 2003)

was used. GLDAS was forced by precipitation data from NRL (U.S. Naval Research Laboratory) as well as a number of atmospheric conditions from different sources, such as ECMWF and GDAS (Global Data Assimilation System, Rodell et al., 2004b). Model tuning was realised by assimilation of skin temperature observations from the Television Infrared Observation Satellite (TIROS, Rodell et al., 2004b). GLDAS-NOAH represented simulations for snow, canopy and soil water storages covering the period from 03/2000 until 04/2008 on a 0.25°-grid between latitude 60°S and 90°N. The Land Dynamics (LaD) model was developed as a land-surface model by Milly & Shmakin (2002a) to simulate global water and energy balances with ISLSCP (International Satellite Land Surface Climatology Project) data for radiation, precipitation, surface pressure, temperature, humidity and wind speed. The variability of soil, groundwater and snow storages was modelled over all continents, excluding Antarctica and Greenland, with a spatial resolution of 1°. The model was tuned by an adjustment of seven parameters of land properties, e.g., surface albedo, thermal conductivity or surface roughness length (Milly & Shmakin, 2002b). Validation of the model output was undertaken by observation-based discharge measurements for large river basins (Milly & Shmakin, 2002a). For this study, the LaD model version, LadWorld-Gascoyne, was available from 01/1980 until 07/2007.

Model strategies, tuning concepts and input data vary widely between the three models used here. GLDAS and LaD were developed as land-surface models with physically based model equations that describe both water and energy fluxes. The sub-grid variability of hydrological processes within these models is either ignored (LaD) or captured by additional parameters or functions (GLDAS). In contrast, WGHM is a water-balance model with conceptual equations that are a simplified representation of water transport processes on large scales. In a station-based calibration, WGHM parameters that are not directly observable are varied until a sufficient agreement of modelled and observed river discharge is achieved. Similarly, LaD is calibrated by river discharge applying spatially distributed parameters. In contrast, data assimilation in GLDAS denotes the direct integration of spatially distributed satellite measurements as parameter or system states into the model by Kalman-filtering.
In addition, it has to be noted that the three models represent different water storage compartments on the continents. While soil water and snow storage changes are simulated by all models, only WGHM simulates the water transport and storage in surface water bodies and only LaD includes an ice component. Moreover, GLDAS-NOAH does not include groundwater in its model structure. Due to the small variability of canopy interception water, its absence in LaD can be neglected.
Errors in input data, model structure and parameters propagate to errors in the model output. Due to the different concepts and data used by GLDAS, LaD and WGHM, their errors are expected to be of different spatial and temporal characteristics, which are analysed by differences in TWS in Sect. 2.3.1.

To reduce uncertainties caused by specific errors of individual models, multi-model ensembles and, in particular, the ensemble mean, are often used in hydrology, oceanography and atmospheric sciences as a more robust estimate of the system state or of forecast fields (e.g., Hagedorn et al., 2005; Tebaldi & Knutti, 2007; Regonda et al., 2006). In this study, comparisons of GRACE with simulated hydrological data was undertaken with a multi-model mean of WGHM, GLDAS and LaD, hereafter named as Average of three global Hydrological Models (A3HM). In order

to compute the A3HM, global fields of total continental water storage were calculated for each hydrological model by adding up all simulated storage compartments. The TWS data of each model were averaged to monthly means and re-gridded to a common 0.5° resolution. Then, for each month, the mean of the three model data sets was calculated to give the A3HM monthly time series. Antarctica and Greenland were excluded from the analysis. See Table 2.2 for global signal intensity of A3HM compared to the other models. A3HM data were transformed into time series of spherical harmonic coefficient sets, up to degree and order 150. To ensure consistency of A3HM and GRACE data, monthly basin-average TWS variations around the mean were computed with the same filter methods as the GRACE data. The common period of analysis in this study was 02/2003-07/2007.

For the regional analysis, the 22 biggest river basins worldwide, with catchment areas greater than 730000 km^2,were selected (Fig. 2.1). As example basins of different climate zones, the Amazon, the Indus, the Nile and the Ob river basins were analysed in more detail.

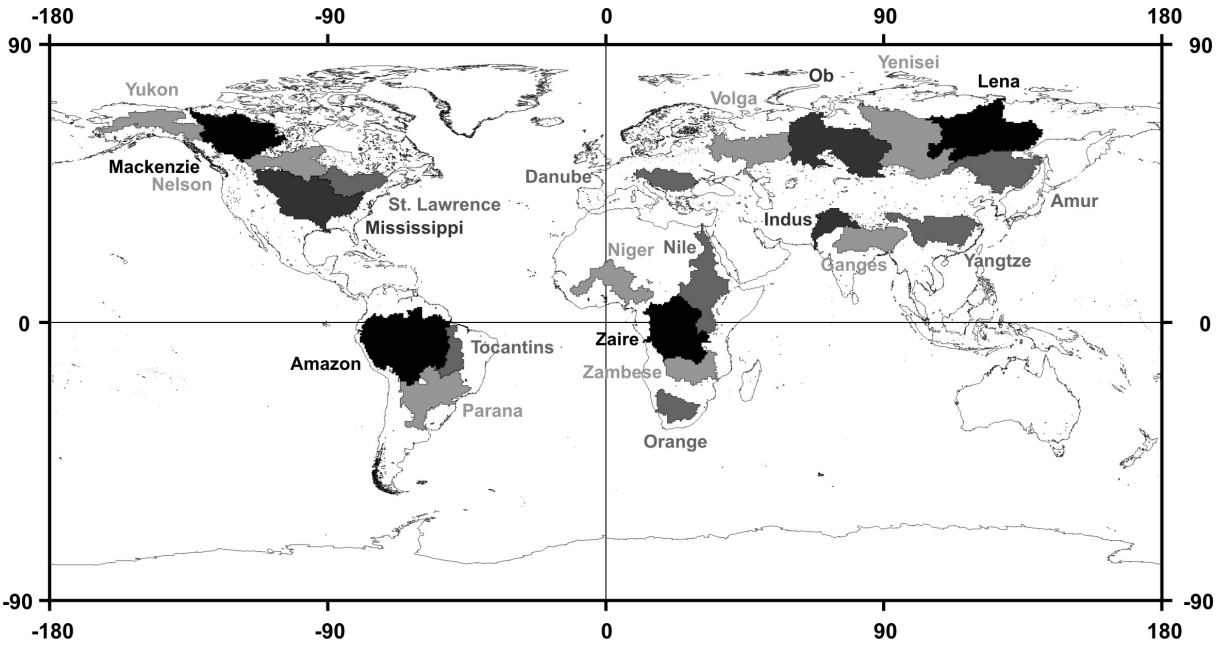


Figure 2.1: The 22 largest river basins worldwide (with an area greater than 730000 km^2).

2.2.4 Correspondence criteria

In hydrology, the Nash-Sutcliffe coefficient (NSC, Nash & Sutcliffe, 1970) is a widely used parameter to measure the performance of simulated time series against observations. The coefficient is defined by the sum of squared differences between predicted (P) and observed (O) values, normalised by the sum of squared deviations of the observations to their mean, during the period of interest with n time steps:

$$NSC = 1 - \sum_{i=0}^{n}(O_i - P_i)^2 \cdot [\sum_{i=0}^{n}(O_i - \bar{O})^2]^{-1}, \tag{2.1}$$

where $\bar{O}$ is the mean of the observations over the examined period. NSC ranges from 1.0 (indicating perfect fit) down to $-\infty$. A value lower than zero denotes that the model is worse than

if $\bar{O}$ was used as a predictor. Therefore, results with values < 0 were discarded in this study. NSC not only evaluates consistency in phase, like the correlation coefficient (CC), but also in amplitude and absolute level of simulated versus observed time series. This is demonstrated in Fig. 2.2 by comparing two sine waves that only differ either in phase (x-axis in Fig. 2.2a) or in amplitude (x-axis in Fig. 2.2b). In this study, NSC was used as a correspondence criterion to evaluate several filter techniques by comparing measured (filtered GRACE data) and simulated (filtered modelled data) time series.

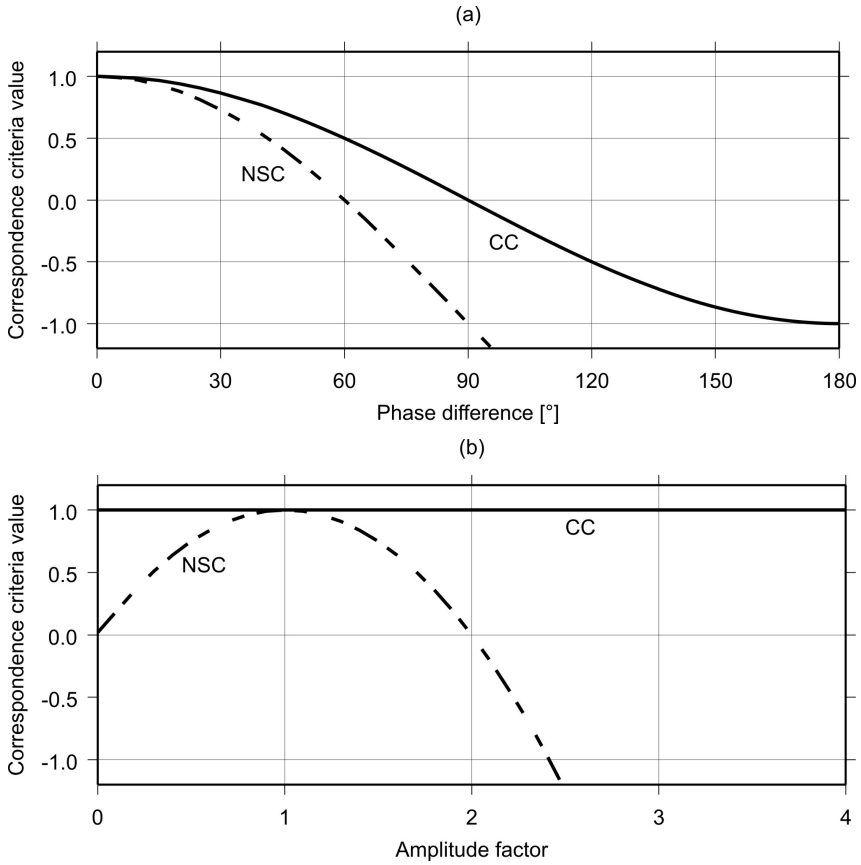


Figure 2.2: Nash-Sutcliffe coefficient (NSC) versus correlation coefficient (CC) for two sine waves that differ in (a) phase or (b) amplitude.

For stronger smoothing, the amplitudes of seasonal TWS variations usually are reduced more strongly, due to an increasing leakage effect. At the same time, the satellite error would decrease in GRACE time series, while it is zero for any filter parameter in the hydrological time series. If the modelled hydrological data comprehend no simulation error, and if they represent the only remaining seasonal signal in GRACE, the leakage error would be the same in both time series, and they would become more similar to each other for stronger smoothing. This may misleadingly cause higher NSC values for higher filter parameter values. Therefore, a measure of the leakage effect was introduced by weighting NSC with an attenuation factor w, which accounts for strong signal attenuation due to filtering.Hence, w was computed from the summed squared difference ($\epsilon^* = \sum_{i=0}^{n}(P_i^* - P_i)^2$) between the filtered ($P$) and unfiltered ($P^*$) time series of simulated hydrological data normalised by the squared sum of the unfiltered time series ($\sigma^* = \sum_{i=0}^{n} P_i^{*2}$).

Finally, we get

$$wNSC = \left(1 - \frac{\epsilon^*}{\sigma^*}\right) * NSC. \tag{2.2}$$

To evaluate the reliability of the results obtained with $wNSC$, an alternative correspondence measure, the *index of agreement* (see Willmot, 1984) was used. This measure also evaluates phase and amplitude differences between modelled and observed time series.

2.3 Results and Discussion

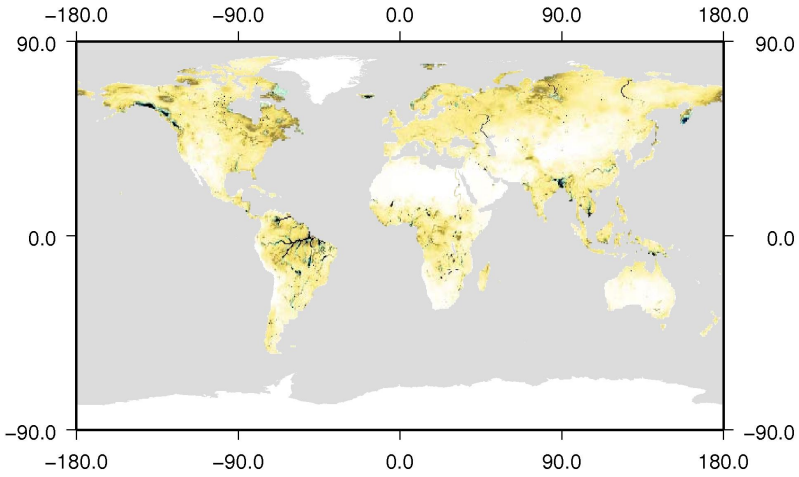
2.3.1 Uncertainties of the hydrological model data

Since GRACE provides the only large-scale observation data of continental water storage change, global hydrological models provide the only means of evaluating GRACE methods for the estimation of TWS variability. In this respect, errors and differences in hydrological models need to be carefully considered. In Sect. 2.2.3, it was shown that model structure, forcing data and strategies for parameter tuning, differ considerably between WGHM, GLDAS and LaD. Maps of TWS variability for the hydrological models (Fig. 2.3) expose the consequences of these different concepts.

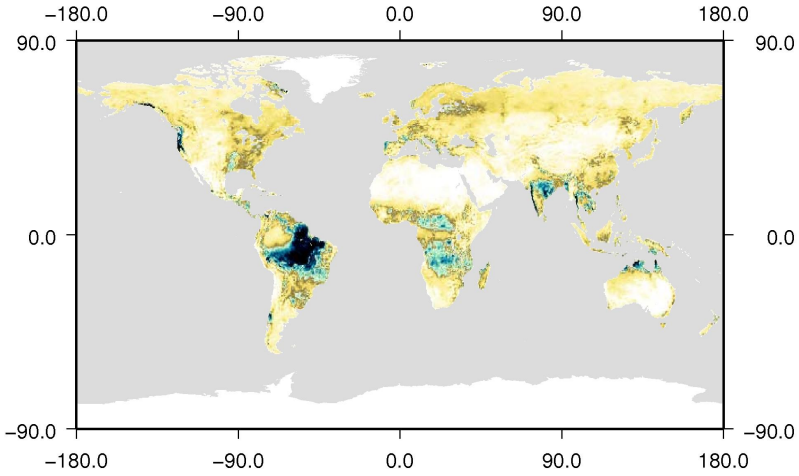
The spatial distribution of the TWS variability in WGHM tends to exhibit linear patterns, reflecting the presence of the surface water storage compartment in the model, including rivers and their inundation areas (see Fig. 2.3a). In contrast, TWS variability from GLDAS and LaD is more gradually distributed in space (Fig. 2.3b and c) in line with larger correlation lengths of soil-water storage (and groundwater for LaD), which dominates TWS in these models. Furthermore, GLDAS amplitudes of TWS variations are larger than those of the other two models. Thus, for the 0.5°-cell-wise RMS-differences between the models (Fig. 2.4), the largest differences occur for GLDAS versus WGHM or LaD, whereas WGHM and LaD are more similar to each other. The differences in the simulated TWS variability between the models may amount to 300 mm w.eq., which is close to the signal magnitude itself. In the difference maps of Fig. 2.4, the linear patterns caused by surface water storage in WGHM not present in the other models, are obvious again. The largest differences occur within the river basins of the Amazon, Congo, Ganges, Mekong, Yukon, St. Lawrence and Ob rivers. Thus, a main difference in TWS variability between the models can be attributed to the fact that different storage compartments with different spatial characteristics are represented in the models.

However, the cell comparisons between models, as shown in Fig. 2.4, may be misleading if basin-average water storage variations and water balances are of interest. This is the case when considering the lower resolution GRACE data. Relative differences between models decrease on the river basin or global scale, e.g., after computing basin averages or reducing the resolution of TWS data by applying a GRACE-filter method. Relative model differences of global (latitude) weighted RMS of TWS reduce after a Gaussian filtering of 500 km, when compared to unfiltered data (Table 2.2). Nevertheless, much smaller differences in signal magnitudes between WGHM and LaD, than of both models relative to GLDAS, remain even after global averaging (Table 2.2).

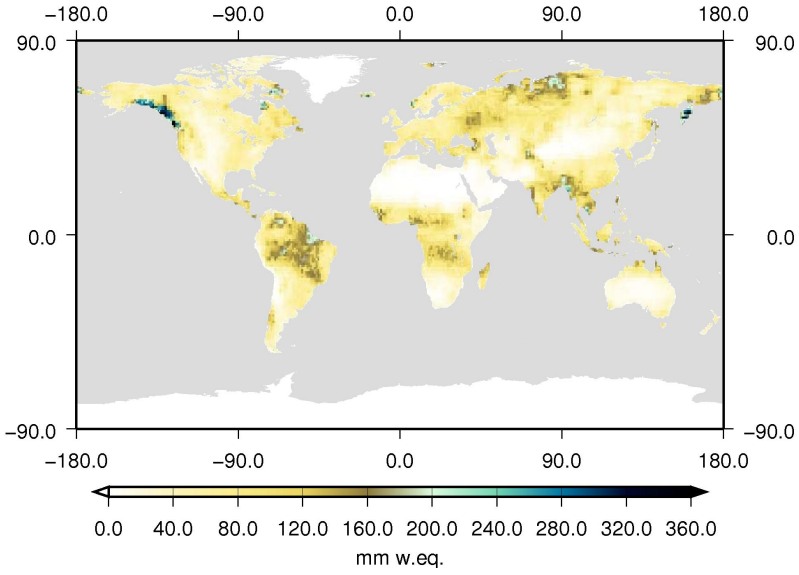
In contrast, temporal correlations of TWS time series are very high between the hydrological models (Fig. 2.5). WGHM and LaD are nearly perfectly correlated on all land areas (Fig. 2.5b), except for a small region in the Himalayas and some linear river courses (e.g. Lena river).



(a) WGHM



(b) GLDAS



(c) LaD

Figure 2.3: RMS of monthly variability of TWS from (a) WGHM, (b) GLDAS and (c) LaD during 2003-2006 (unfiltered).

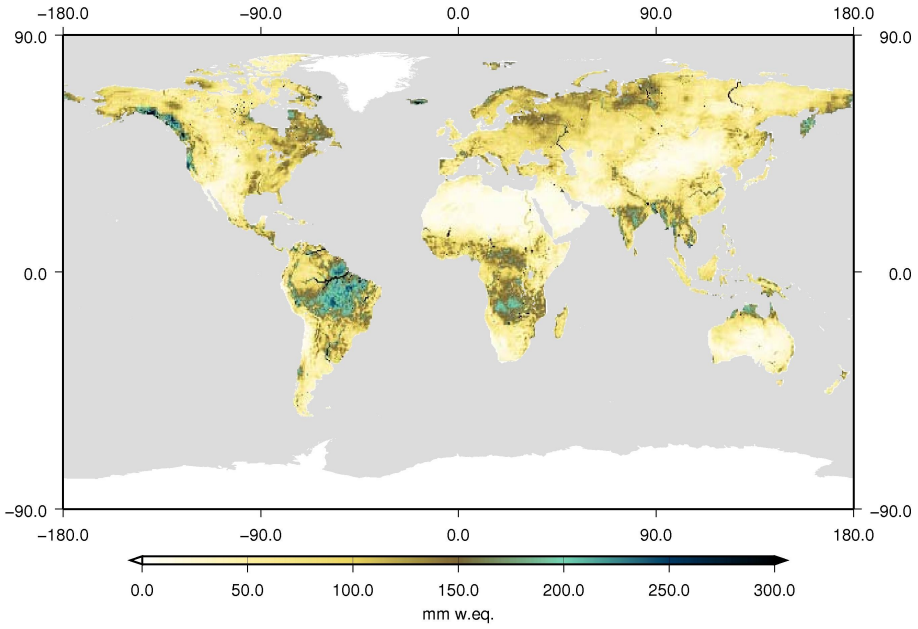


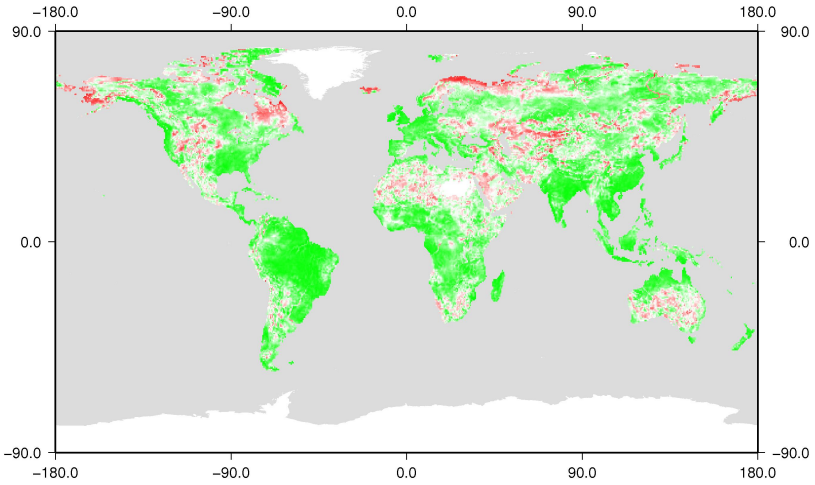
Figure 2.4: RMS of monthly differences of TWS variability between WGHM, GLDAS and LaD during 2003-2006 (unfiltered).

The first deviation may be due to differences in the snow algorithms and the latter due to the absence of surface water routing in LaD. This process causes longer residence times and, thus, delayed storage depletion within river basins for WGHM. But correlation maps for WGHM versus GLDAS (Fig. 2.5a) and WGHM versus LaD (Fig. 2.5c) indicate good temporal correlations for the major river basins (e.g., Amazon, Zaire, Ganges, Mississippi, or large parts of Ob, Yenisei and Lena) despite large differences in amplitudes of TWS variations as shown in Fig. 2.4. Dry areas, such as North Africa, central North America, central Australia and central Asia, are not well correlated in time between the models, but the TWS change signals are very small (compare to Fig. 2.3) or negligible in these areas. Low correlations for regions with large TWS variability only appear in small areas of Scandinavia, East-Siberia and the northeast of North America.

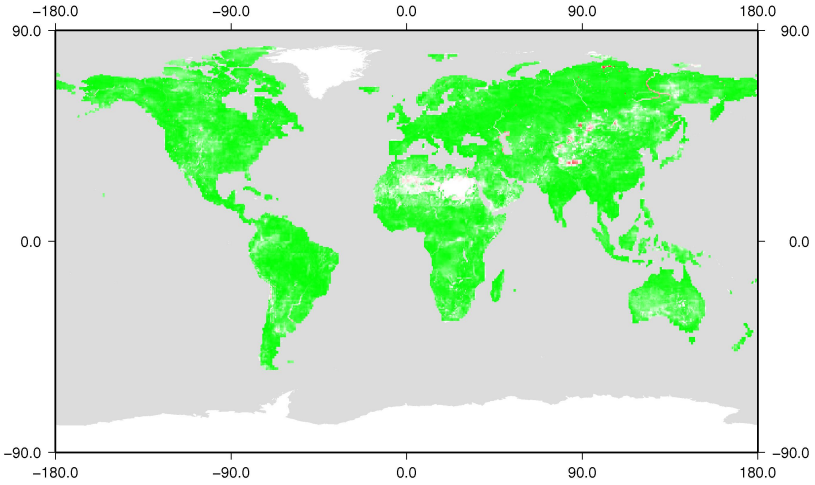
To conclude, differences of TWS variations between the three global hydrological models are quite large when evaluated at the grid scale. These differences are mainly due to different model structures in terms of water storage components represented in each of the models. In previous

Table 2.2: Global weighted-RMS of TWS variations (in mm of a water mass equivalent column) for the global hydrological models WGHM (W), GLDAS (G), LaD (L) and the multi-model mean A3HM (col. 2-5) are derived from unfiltered data sets after application of a Gaussian filter with 500 km half-length. Columns 6-8 show relative differences of wRMS values between the hydrological models.

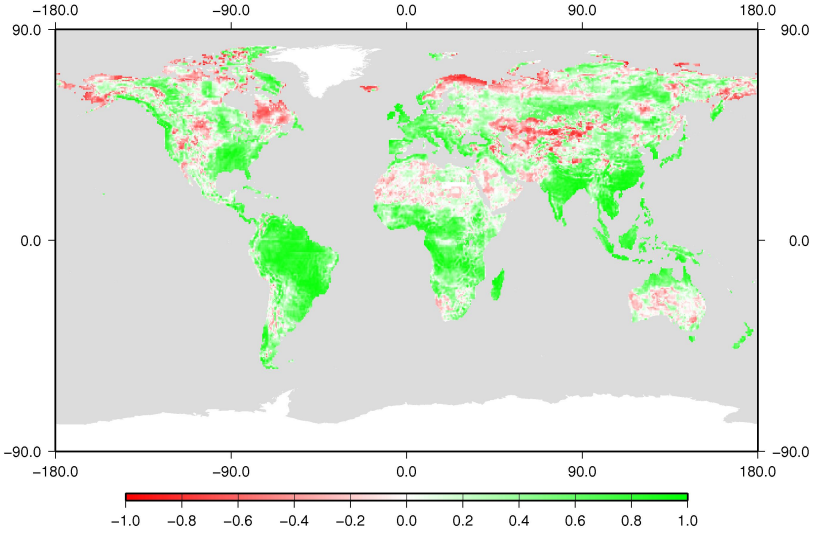
	wRMS [mm]				relative difference		
	W	G	L	A3HM	(G-W)/G	(G-L)/G	(L-W)/L
unfiltered	60.8	97.8	62.4	64.4	0.38	0.36	0.03
gaussian, 500km	15.5	20.7	16.0	16.3	0.25	0.23	0.03



(a) WGHM vs. GLDAS



(b) WGHM vs. LaD



(c) GLDAS vs. LaD

Figure 2.5: Correlation of monthly TWS between WGHM, GLDAS and LaD during 2003-2006.

studies (e.g., compare Güntner, 2009; Werth & Güntner, 2008), no model was shown to be most consistent relative to GRACE-derived TWS variations. Nevertheless, global hydrological models represent the most comprehensive and state-of-the-art data on continental water-cycle processes on large scales. Therefore, they are the only data source to evaluate GRACE-derived estimations of TWS variability. Relative differences between the models reduce on the scale of river basins and are relevant for comparisons to GRACE data. Furthermore, the models show a good temporal agreement, especially within regions of large TWS variations. For this study, the model mean A3HM provided a compromise between the three independent model realisations of different concepts. A3HM averaged out particular model errors due to individual model structures and input data sets. Only a few systematic errors that may prevail in all input data sets, such as those due to the generally small number of precipitation stations in specific regions (e.g., parts of Africa, South America or Central Asia), cannot be reduced in this way. In evaluating GRACE filter methods, A3HM currently provides the most adequate estimation of water storage variations on the continents.

2.3.2 Filter evaluation

2.3.2.1 Filter effects on seasonal amplitude and phase

Different filter methods cause different GRACE error reduction and leakage effects when applied to different river basins. To understand reasons for such differences, filtered time series with non-decorrelating filter methods and different filter parameters are shown in Fig. 2.6-2.8 (a-d) for GRACE (top) and A3HM (bottom) derived TWS variations. Examples are given for three river basins (Amazon, Indus, Ob) to illustrate the effects of different climate zones with diverse hydrology and different regimes of TWS variations. The Amazon exhibits a strong signal that dominates northern South America (Fig. 2.6, dotted time series). The signal of a surrounding area of the Amazon basin (defined by a latitudinal and longitudinal buffer of 8° around the catchment boundaries) exhibits a much smaller signal with a slight phase shift (triangles). The application of filters with a weaker smoothing strength (blue coloured time series) generates erroneous time series in terms of GRACE and nearly undamped time series in terms of A3HM. Stronger smoothing (pink coloured time series) leads to higher TWS amplitude attenuation due to the small signal in the surrounding areas. The amplitude damping is stronger for filter I compared to filter IV and most prominent for filter II with small parameter values, as well as for filter III, with very small signal variance parameter values. Phase shifts of the surrounding areas are too small to have a noticeable influence on the seasonal phase of the filtered Amazon time series.

In contrast to the Amazon basin, higher differences between the model-based and GRACE-derived TWS data occur for the Indus river basin. Also, the smaller size of this basin leads to more erroneous GRACE time series for weaker smoothing, and differences between the filter methods become more evident (Fig. 2.7). The Indus basin is influenced by a strong signal in surrounding regions with opposite seasonal phase. For example, the closely located Ganges River has a strong signal caused by the Indian summer monsoon, whereas Indus water storage variations are more influenced by snow accumulation and melt. Furthermore, the eastern desert in the Indus basin exhibits low TWS variability. Thus, a strong leakage effect of the surrounding

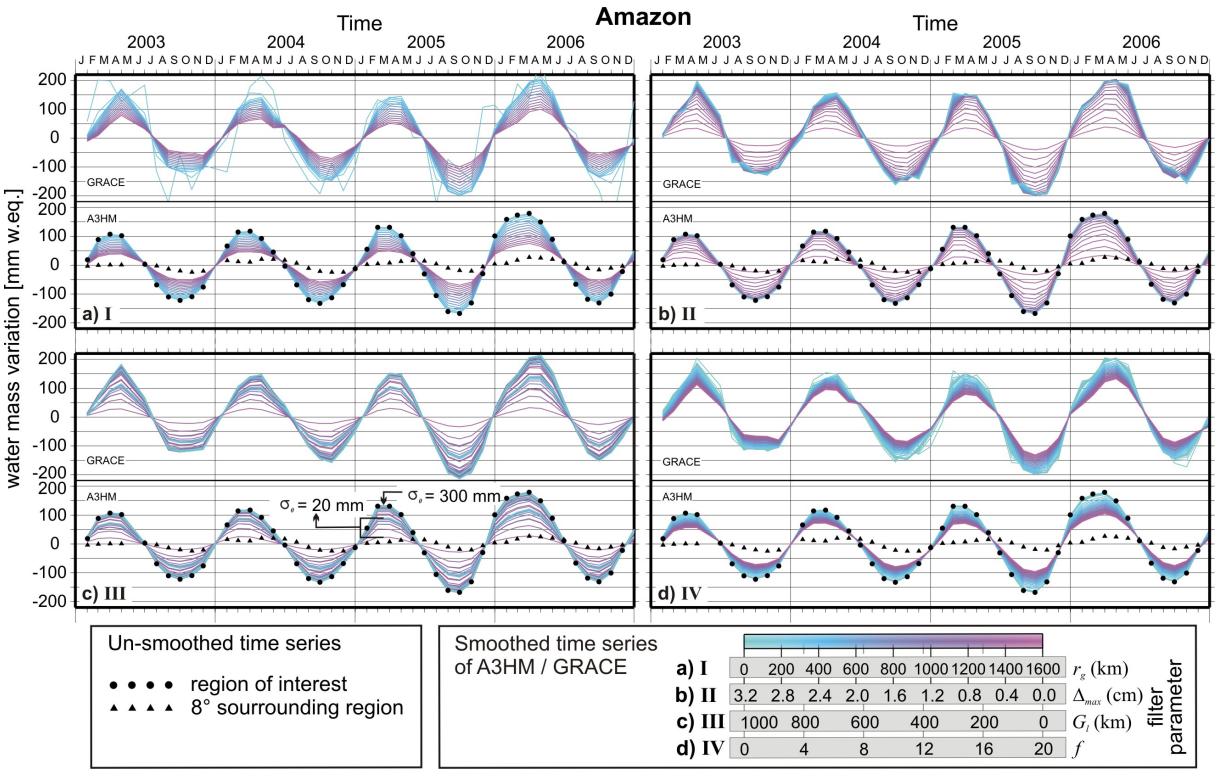


Figure 2.6: Time series of TWS variations for the Amazon river basin after applying the non-decorrelating filter methods: (a) I, (b) II, (c) III and (d) IV. For different values of the filter parameter, the graphs colour gradually changes from blue (weak smoothing) to pink (strong smoothing). From the two parameters of filter III, G_l is colour-coded and σ_0 graphs, with maximal and minimal values, are exemplarily indicated in sub-Figure (c). See further explanations in the main text.

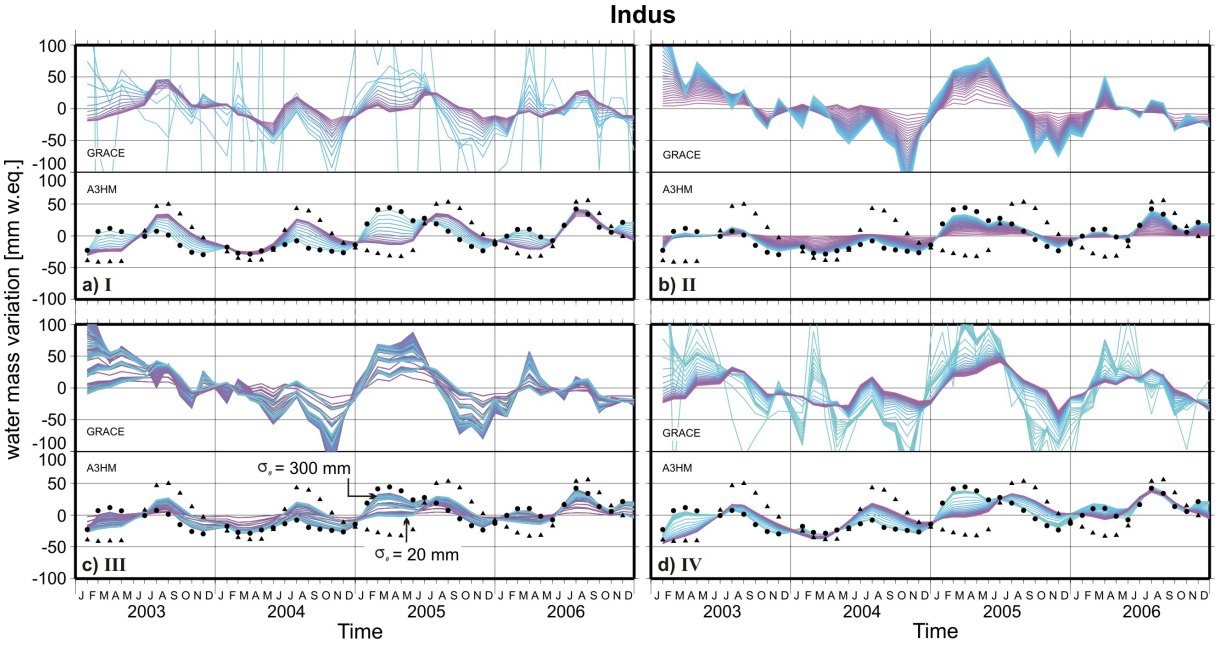


Figure 2.7: Same as Figure 2.6 but for the Indus basin.

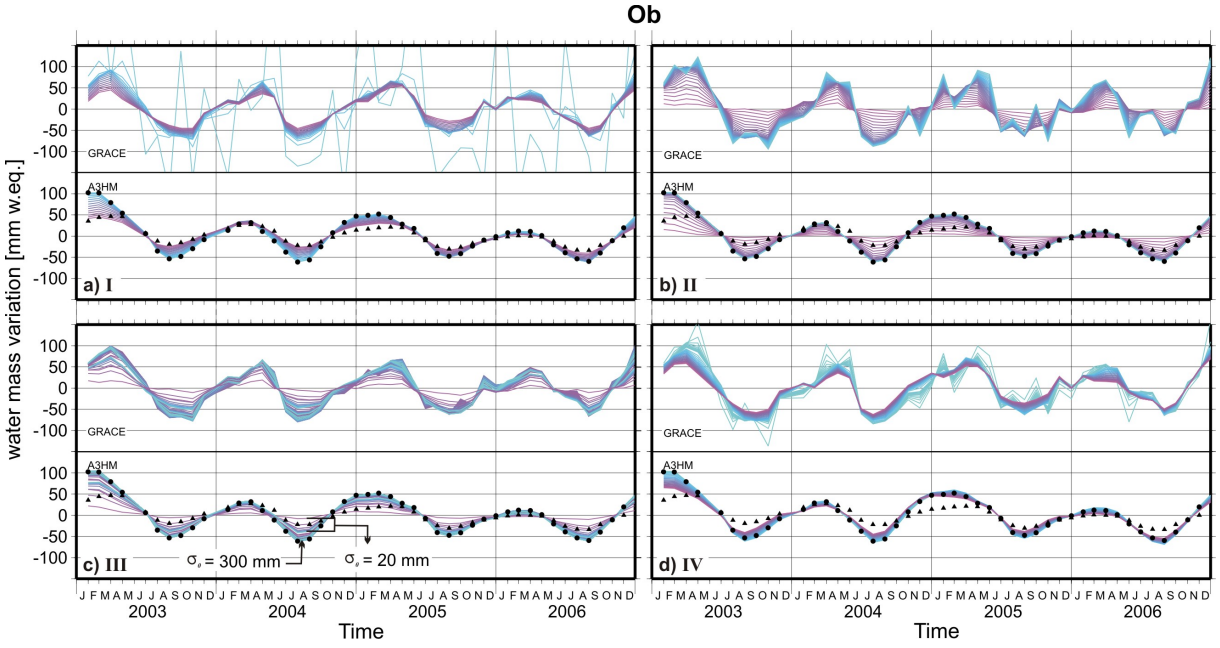


Figure 2.8: Same as Figure 2.6 but for the Ob basin.

areas causes strong phase shifts towards the surrounding signal when using filter IV and I, with strong smoothing. This occurs with filter III, as well, with signal variance parameters (50 or 20 mm w.eq.) that are too small. In contrast, for filters II and III (the latter with signal variance parameters greater than 100), only amplitude damping can be observed.

The Ob basin (Fig. 2.8) is surrounded by regions with equal phase and similar amplitude (see Fig. 2.3). Hence, signal leakage is less dominant. The similar hydrological signal characteristics of surrounding river basins (e.g. Volga, Yenisei) balance signal truncation inside the Ob basin. Therefore, filters I and IV cause, overall, very little amplitude damping. Filter IV even exhibits slightly increased amplitudes for some parameters compared to the unfiltered A3HM signal. On the other hand, amplitude damping of filters II and III becomes strong for very small parameters of II and small signal variance parameters of III. For the Ob basin, phase shifts are a negligible filter effect. Compared to the Amazon, the more erroneous GRACE time series for lower smoothing are explainable by other factors, such as the more complex shape, or smaller size, of the Ob river basin.

These three examples show that phase shifts and amplitude attenuation of TWS time series differ between the river basins depending on the applied filter methods, the signal properties inside and outside the basin, and the basin size or shape. If phase shifts between the time series outside and inside a basin are of negligible size, the leakage scenarios established by Klees et al. (2007) are comprehensible. For example, the small signal of the Amazon's surrounding region biases the TWS time series of the Amazon basin more strongly than does the time series of the Ob basin when influenced by a signal of similar size outside the Ob basin. But when a marked phase shift between the signal outside and inside the basin is present, the three leakage scenarios vary between months because the ratio of the signal outside the basin to the signal inside the basin varies. Therefore, phase shifts occur in the filtered time series of such regions. This applies especially to high (strong) parameters of filters I and IV (smoothing), as shown for the Indus

basin. In such cases, the application of amplitude scaling or bias correction factors as proposed by Velicogna & Wahr (2006) and Klees et al. (2007) will not recover the hydrological signal after filtering.

A summary of seasonal phases and amplitudes for the 22 river basins and their surrounding areas is given in Table 2.3. Phase shifts and amplitude differences as an effect of filtering are shown in Table 2.4. To make them comparable, the different filter methods were parameterised in such a way that they generate the same RMS of monthly satellite error (propagated from the coefficient errors) as a Gaussian smoother of 500 km radius. If phase shifts of the surrounding region are large, and if amplitudes are of similar magnitude as the signal inside the basin, an impact of the phase of the surrounding signal is clearly visible in the filtered time series (e.g., Amur, Indus). A much smaller signal amplitude in the surrounding region compared to the basin itself results in strong amplitude damping (e.g., Ganges, Tocantins, Zambezi). Both effects are simultaneously visible for a few basins (e.g. Amur, Indus, Parana). Also, the size and sign of both effects vary between the filter methods. Some basins (e.g., Amur, Lena, Nelson, Nile) exhibit phase shifts of different signs. For other basins, the size of amplitude damping differs largely between filter methods of equal satellite error reduction (e.g., Danube, Ganges, St. Lawrence, Tocantins). Due to a weak annual signal of the Orange basin (see seasonal amplitude of Orange in Table 2.3), leakage tends to increase the annual amplitude for this basin. Probable reasons for the different filter effects will be given in the next section.

Table 2.3: Seasonal amplitude (A, col. 2) and phase (Φ, col. 3) of TWS variations for the 22 river basins, derived from the ensemble model mean A3HM. Respectively, seasonal amplitude difference (ΔA, col. 4) and phase shift ($\Delta\Phi$, col. 5) are computed for an 8° surrounding region.

	Basin		Surr. region	
Basin	A	Φ	ΔA	$\Delta\Phi$
	[mm]	[day]	[mm]	[day]
Amazon	136	-23	-119	-32
Amur	10	154	-4	-54
Danube	47	-1	-15	+13
Ganges	125	-182	-75	-3
Indus	14	-65	+27	-126
Lena	26	36	-17	-3
Mackenzie	34	9	-8	-2
Mississippi	26	5	-16	+43
Nelson	23	26	+8	-16
Niger	77	165	-59	+19
Nile	37	-182	-34	-32
Ob	50	-2	-26	-6
Orange	1	-40	+18	+21
Parana	76	-5	-20	-7
St. Lawrence	85	3	-49	+12
Tocantins	230	-15	-129	-17
Volga	73	-3	-41	+1
Yangtze	39	-146	+9	-31
Yenisei	32	9	-16	-9
Yukon	43	9	-17	+1
Zaire	23	8	-19	+180
Zambezi	103	-19	-60	+7

Table 2.4: Filter-induced bias of the seasonal amplitude (ΔA, col. 3-8) and phase (ΔΦ, col. 9-14) for 22 river basins and six filter methods (I-VI). Parameters for each filter were set to give the same propagated satellite error (ϵ_{sat} in col. 2) as a Gaussian smoother of 500 km radius. Results were computed from A3HM data. See Table 2.3 for seasonal amplitude and phases of the un-smoothed signal.

	ϵ_{sat}	ΔA [mm]						ΔΦ [day]					
Basin	[mm]	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Amazon	10	-18	-2	-6	-18	-47	+1	-2	+0	+0	-1	-2	+0
Amur	11	-3	+0	-4	-3	-9	-1	-29	+7	+8	-40	-64	-7
Danube	15	-3	-15	-1	-1	-17	-2	+5	+1	+4	+2	+9	+3
Ganges	14	-27	-19	-22	-43	-49	-6	-3	-1	-2	+1	-3	-2
Indus	18	-2	-7	-7	+2	+7	-5	-94	-41	-67	-91	-117	-29
Lena	10	-3	-1	-7	+6	-8	+0	+1	+1	+6	-9	-5	+3
Mackenzie	12	+5	-1	-3	-9	-1	+2	+0	-1	-1	+4	+1	+1
Mississippi	10	-1	+0	-1	-7	-7	+0	+2	-1	+1	+17	+1	-1
Nelson	12	+1	-4	+1	-5	-4	+1	-3	+2	+1	+6	-14	+4
Niger	14	-8	-9	-15	+2	-23	+2	+1	-1	-2	-1	+2	+0
Nile	15	-4	-4	-1	-17	-9	+0	-4	-7	-4	+8	-5	-5
Ob	10	-3	-1	-7	-2	-24	+1	+0	+0	-1	+2	+5	+0
Orange	19	+3	+7	+5	+4	+5	-1	+21	-167	-171	+11	+20	-119
Parana	14	-10	-31	-29	-16	-20	-11	+2	+9	+8	+7	+3	+4
St. Lawrence	14	-17	-16	-16	-47	-42	-2	+6	-1	+4	+4	+6	+1
Tocantins	22	-42	-145	-27	-45	-89	-22	-4	+2	-3	-4	-6	-4
Volga	12	-12	-7	-9	-10	-33	-2	+1	+0	+2	+4	+4	+1
Yangtze	13	+2	-4	-1	-11	-5	+0	-8	+2	-3	+1	-9	+0
Yenisei	10	+0	+0	+1	+7	-15	+2	-1	+0	-2	-2	-2	-1
Yukon	13	+4	+3	-7	+6	-8	+17	-1	-3	-3	-3	-1	-2
Zaire	13	-1	-1	+0	+4	-8	+1	-5	-2	-3	-19	-5	-5
Zambezi	16	-17	-22	-25	-21	-38	-2	+2	+2	+2	+1	+2	+1

2.3.2.2 Correspondence of GRACE to hydrology data

The different filter methods and smoothing rates were evaluated with the $wNSC$ correspondence criteria (explained in Sect. 2.2.4) against A3HM data (Table 2.5). Example results for the Amazon and Ob basins are shown in detail in Fig. 2.9.Applying the Gaussian filter for the Amazon basin, averaged time series are very sensitive to damping when evaluated by $wNSC$. A radius of 300 km results in the highest wNSC value, i.e., the best correspondence of GRACE and A3HM data. This radius is similar to the results of Schrama et al. (2007), who selected a globally optimal Gaussian filter radius of 275 km by comparison with GPS load measurements. The $wNSC$ correspondence rapidly decreases for smaller and higher radii than 300 km. The degree-only dependency of filter method I does not take into account differences in accuracy for coefficients of equal degree but different order. Therefore, method I may either filter coefficients with an acceptable signal-to-noise ratio too strongly or may not sufficiently filter coefficients containing large errors. Thus, basin average values are either affected by signal leakage from surrounding areas or by large errors.The Amazon basin is located close to the ocean (which inheres a signal close to zero) both to its east and to its west. To its north it borders on the equator (where a shifted seasonality of water storage occurs further north). Therefore, basin averages for the Amazon are quite sensitive to leaking signals or amplitude damping (as shown above) for large filter radii. This causes high parameter sensitivity of filter I.

Filter II and IV also exhibit a distinct sensitivity to filter parameters. In the case of filter II,

the correspondence to hydrological model data is poor for maximum admitted satellite errors smaller than about 5 mm in both river basins. In this case, parameter values that are too low (i.e., strong smoothing) increase signal leakage (i.e., amplitude damping). Smoothing with filter II, for admitted maximum satellite errors in the range of 8 to 15 mm, performs well for both basins. Maximum $wNSC$ values of II are slightly higher than optimum results of I for the Amazon, and are somewhat lower than optimum results of I for the Ob basin. These results follow from the design of filter II, which preferably preserves coefficients that contain important signals of the examined basin and, thus, reduces signal leakage. Furthermore, the anisotropic

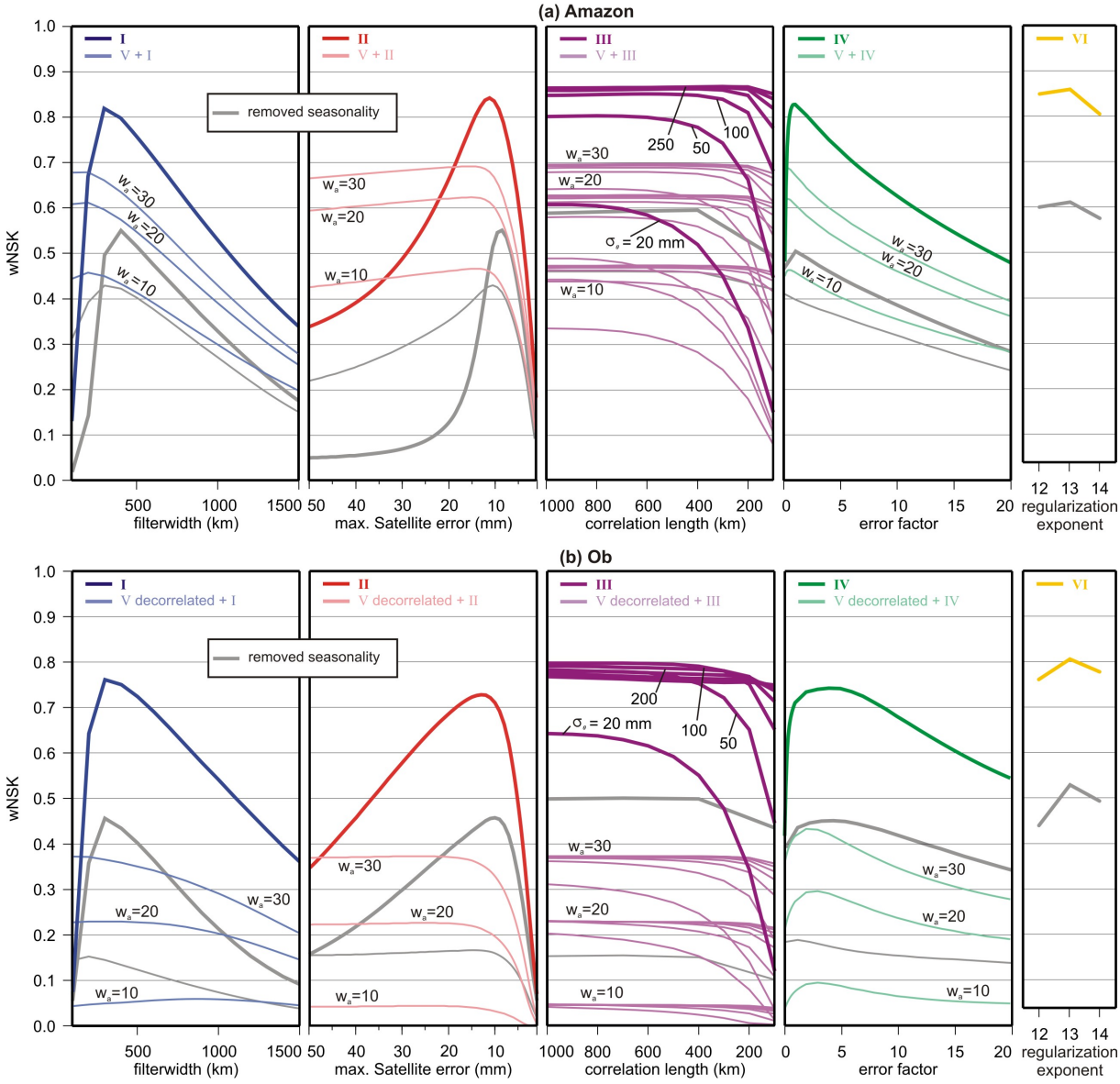


Figure 2.9: Weighted NSC ($wNSC$) performance of different filter types (I-VI) and grades of smoothing strengths for: (a) the Amazon and (b) the Ob river basins. Bold lines in blue: Gaussian (I), red: basin optimised (II), purple: signal model optimised (III), green: SNR optimised (IV), yellow: decorrelation VI. Light coloured lines: decorrelation V (additionally applied to I-IV) with $n_e = 30$ for all displayed graphs. Grey lines: seasonality removed before computation of $wNSC$ for all filters. See a filter description in Sect. 2.2.1.

design of II (compared to the isotropic Gaussian) distinguishes between the orders of coefficients with the same degree, which enables a finer adjustment of the filter weights. The filter II design is of particular benefit for the Amazon basin, where signal separation from the surrounding areas is important for preventing amplitude damping, as described above. This strategy of filter II is of less advantage for the Ob basin, where leakage of I is compensated by similar signals in surrounding areas (as described above). This also explains the lower filter parameter sensitivity of II for the Ob basin.

The optimal error factor of the anisotropic filter IV is $f = 1$ for Amazon and $f = 4$ for Ob. This implies that the correlated GRACE coefficient errors are properly estimated within the Amazon basin, whereas they are regionally underestimated within the Ob region. Optimal error factors for IV differ between the river basins because the quality of GRACE coefficient error assessment varies regionally (see also Horwath & Dietrich, 2006). The optimised filter IV is nearly as good as each of the other optimised filter types for the Amazon. Again, due to similarly large signal characteristics around the Ob basin, signal damping by leakage is small when using filter IV. The positive leakage for some smoothing rates of that filter increases TWS amplitudes, as shown in the previous section. Therefore, larger parameter values for IV hardly damp regionally averaged time series and only reduces errors. This also leads to a smaller parameter sensitivity of that filter for the Ob basin than for the Amazon.

By contrast, the anisotropic filter type III is comparatively insensitive in terms of $wNSC$ values to parameter variations. Correlation lengths greater than 300 km, and standard deviations greater than 100 mm, provide $wNSC$ values that differ less than 0.04 in both river basins. Sensitivity of the standard deviation parameter is higher than that of the correlation length parameter. This confirms Swenson & Wahr (2002), that an exponential signal model is a good approximation for estimating the leakage error, and that it does not strongly depend on the exact estimation of its parameter values, as long as σ_0 and G_l are not too small. Compared to the other filters, III provides the highest $wNSC$ results for the Amazon and the second highest for the Ob basin.

For the Amazon basin, decorrelation by V (thin coloured lines in Fig. 2.9) does not improve the correspondence between filtered GRACE and hydrological model time series for any of the four filter types discussed above. This follows from the low efficiency of V in equatorial regions. Signals of these regions are dominant in near-sectorial coefficients (with similar degree and order), which are corrected incompletely by that method (Swenson & Wahr, 2006). Decorrelation with filter V does not give better results than the four non-decorrelating methods for the Ob basin. Outside the equatorial region, improvement by decorrelation filter V only occurs for the Lena, Orange, Mississippi, Parana and St. Lawrence basins.

The alternative decorrelation method VI for $a = 10^{12}$ gives a $wNSC$ value close to the filter III optima for the Amazon (black dashed line in Fig. 2.9). Results of VI for $a = 10^{13}$ are superior to all filter methods for the Ob basin. A low parameter sensitivity of VI is visible for both river basins.

Schaefli & Gupta (2007) showed that the NSC is very sensitive to seasonality. Since seasonality is the most dominant signal in most river basins, $wNSC$ was re-computed after removing the seasonal signal from the time series. These results are shown by the grey graphs in Fig. 2.9 for all filter types, respectively (including the optimal filter V with $w_a = 30$). The $wNSC$ values

become smaller due to relatively high errors in the small non-seasonal water storage signal of GRACE and the models. But the highest $wNSC$ is likely to occur for similar filter parameter values when compared to results that include the seasonal signal. This shows that the optimum filter technique for a specific river basin is more a function of filter properties in combination with the geographical characteristics of the region of interest, than a function of the selected time period or of TWS temporal dynamics. Thus, the results obtained here can be expected to be of broad relevance for hydrological studies.

A summary of filter comparison for all 22 basins is given in Table 2.5, with the highest $wNSC$ value and the corresponding filter parameter value for each filter type. The optimal filter type for each basin is indicated by bold numbers. Among the basins, different filter methods with different filter parameter values appear to be optimal. GRACE basin-average TWS time series with optimum filtering generally have a high correspondence to A3HM, except for Amur, Indus, Nelson, St. Lawrence and Orange. For the latter five basins, the error map (Fig. 2.4) shows large differences of TWS variability between the hydrological models for the St. Lawrence basin only; but temporal correlations between the models are poor for large areas inside all of the five basins (Fig. 2.5). Hence, large uncertainties in the hydrological model data may lead to uncertain results for the filter optimisation for these basins.For filter I, the highest $wNSC$ occurs for radii from 300 km to 400 km for nearly all river basins. This indicates that the spatial resolution of GRACE-derived TWS variations is mostly better than 500 km. The optimised maximum satellite error of filter II tends to be larger than the 10 mm water equivalent. This illustrates the limitations in accuracy of GRACE TWS estimates due to GRACE measurement errors. Furthermore, optimal

Table 2.5: Weighted NSC ($wNSC$) evaluation of GRACE filter types with A3HM data: highest $wNSC$ for each filter type and corresponding filter parameter values in brackets. Bold $wNSC$ values indicate the overall optimal filter method for each basin.

	weighted Nash-Sutcliffe-Coefficient (wNSC)					
	I	II	III	IV	V	VI
Basin	($r_g[km]$)	($\Delta_{max}[mm]$)	($\sigma_0[mm]$, $G_l[km]$)	(f)	(w_a)	(x)
Amazon	0.82 (300)	0.84 (11)	**0.87** (250,300)	0.83 (1)	0.70 (30),II	**0.86** (13)
Amur	0.27 (300)	**0.35** (25)	0.31 (300,100)	0.26 (2)	0.15 (30),I	0.21 (13)
Danube	0.63 (300)	0.69 (27)	0.70 (250,1000)	0.66 (0.6)	0.46 (30),II	**0.75** (12)
Ganges	0.81 (300)	0.81 (17)	0.88 (300,500)	0.77 (1)	0.76 (30),II	**0.91** (12)
Indus	0.15 (400)	0.29 (21)	**0.33** (200,1000)	0.25 (2)	0.11 (30),III	**0.32** (13)
Lena	**0.49** (300)	0.42 (13)	**0.49** (300,1000)	**0.49** (2)	**0.50** (20),IV	**0.49** (12)
Mackenzie	0.60 (400)	0.59 (13)	**0.65** (150,200)	0.60 (1)	0.42 (30),II	0.60 (12)
Mississippi	0.61 (400)	0.59 (13)	0.64 (150,1000)	0.54 (1)	0.60 (30),I	**0.66** (12)
Nelson	0.31 (500)	0.29 (30)	0.33 (200,1000)	0.22 (0.4)	0.30 (30),II	**0.51** (12)
Niger	0.85 (300)	**0.88** (23)	**0.89** (200,200)	0.86 (0.7)	0.78 (30),IV	**0.89** (12)
Nile	0.56 (400)	0.58 (14)	**0.61** (150,900)	0.43 (0.5)	0.57 (30),II	0.59 (13)
Ob	0.76 (300)	0.73 (13)	**0.80** (100,900)	0.74 (4)	0.43 (30),IV	**0.81** (13)
Orange	0.29 (600)	0.09 (41)	0.32 (20,1000)	0.17 (7)	**0.38** (20),I	0.28 (14)
Parana	0.67 (500)	0.48 (16)	0.69 (200,1000)	0.63 (2)	0.58 (30),II	**0.73** (12)
St. Lawrence	**0.37** (200)	0.15 (20)	0.24 (20,1000)	0.24 (0.1)	0.22 (30),I	0.21 (14)
Tocantins	0.78 (400)	0.78 (34)	**0.85** (300,900)	0.80 (0.7)	0.69 (30),II	**0.85** (12)
Volga	0.70 (300)	0.66 (15)	0.75 (100,900)	0.70 (1)	0.50 (30),II	**0.78** (13)
Yangtze	0.74 (400)	0.71 (17)	0.79 (300,700)	0.69 (2)	0.62 (30),III	**0.82** (12)
Yenisei	0.60 (400)	0.57 (12)	0.63 (50,500)	0.63 (16)	0.42 (30),IV	**0.66** (14)
Yukon	0.50 (300)	**0.59** (16)	**0.59** (150,100)	0.52 (1)	0.24 (30),IV	0.57 (12)
Zaire	0.41 (400)	0.47 (12)	0.49 (100,300)	0.47 (2)	0.41 (30),III	**0.51** (13)
Zambezi	0.75 (300)	**0.81** (27)	**0.81** (300,200)	**0.82** (0.7)	0.64 (30),III	**0.82** (12)

correlation lengths of filter III vary considerably between the river basins. Lower correlation lengths may be due to the importance of surface water storage concentrated in a small spatial domain, as pointed out by Güntner et al. (2007b). Filter V provides optimal filter results for only two river basins (Lena, Orange). Parameter optimisation of V is not straightforward, because the four filter parameters (w_a, w_e, n_a and n_e) may have to be adjusted individually for each basin in addition to the parameter of the subsequently applied filter method. Method VI provides the highest $wNSC$ for sixteen river basins, and its performance is also in the same range as the best alternate filter methods for the remaining basins. The anisotropic decorrelation method of VI seems to efficiently preserve the hydrological signal while reducing GRACE satellite errors. In addition, VI exhibits low filter parameter sensitivity. This supports the method's strategy of deriving an error covariance matrix from satellite orbits in order to decorrelate the coefficients in the filter process.Finally, the computations were repeated for the Index of Agreement (Willmot, 1984). This measure of correspondence between GRACE and hydrological model data generally confirms the results as provided above (not shown).

For all filter methods, a final estimation of biases of the seasonal amplitude and phase in the TWS time series after application of the optimised filter, is provided in Table 2.6, based on A3HM data. For both amplitudes and phases, biases are reduced for many river basins in comparison to Table 2.4, where a standard Gaussian filter, or filter of equivalent smoothing strength, were applied. This indicates a successful optimisation of the filter type and parameter. Large phase shifts remain for the Indus and Orange basin only. For most of the other basins, filters III and

Table 2.6: Filter-induced bias of the seasonal amplitude (ΔA, col. 2-7) and phase ($\Delta\Phi$, col. 8-13) for 22 river basins and the six optimised filter methods (I-VI) as listed in Table 2.5. Results were computed from A3HM data. See Table 2.3 for seasonal amplitude and phases of the un-smoothed signal.

	ΔA [mm]						$\Delta\Phi$ [day]					
Basin	I	II	III	IV	V	VI	I	II	III	IV	V	VI
Amazon	-7	-4	+0	-4	-27	-1	-1	+0	-1	+0	+0	+0
Amur	-2	+0	+0	+0	-8	-1	-11	+3	+5	-12	+23	-7
Danube	-2	-5	-2	-4	-10	-3	+3	+0	+2	+2	+10	+2
Ganges	-13	-12	-5	-13	-25	-4	-1	+0	-1	+0	-3	-1
Indus	-5	-5	-4	-4	+0	-5	-74	-27	-24	-26	-105	-29
Lena	-1	-2	+0	+5	-3	+0	+2	+2	+2	-4	-11	+3
Mackenzie	+5	-2	+1	+1	+2	+2	+0	-1	+0	-1	-2	+1
Mississippi	-1	-1	+0	-1	-3	+0	+1	-1	+0	+3	-1	+0
Nelson	+1	-1	+0	-2	-9	-1	-3	+0	+2	+1	-9	+4
Niger	-2	-2	+0	+5	-9	+2	+1	+0	+0	-1	+1	+0
Nile	-3	-3	-1	-6	-5	-1	-3	-5	-3	+3	-1	-1
Ob	+0	-2	+0	+3	-18	+1	+0	+0	+0	+0	+6	+0
Orange	+4	+0	+1	+4	+4	+2	+21	+11	-181	+12	+18	-138
Parana	-10	-14	-8	-12	-20	-6	+2	+4	+3	+5	+6	+3
St. Lawrence	-4	-6	-37	+0	-42	-14	+2	+0	+4	+0	+6	+3
Tocantins	-31	-21	-10	-16	-55	-7	-3	-2	-2	-1	-4	-3
Volga	-6	-7	-3	-2	-27	-2	+1	+0	+2	+1	+5	+1
Yangtze	+1	-2	+0	-7	-6	+0	-5	+1	+0	+0	-2	+0
Yenisei	+1	-1	+0	+3	-9	+2	-1	-1	-2	+0	-2	-2
Yukon	+10	+4	+6	+8	+1	+13	-1	-3	-3	-1	-1	-2
Zaire	-1	-1	+1	+6	-4	+1	-5	-1	-3	-18	-6	-5
Zambezi	-6	-1	+0	-1	-21	+2	+1	+1	+1	+0	+1	+1

VI had the smallest seasonal phase shifts and amplitude damping.

2.3.2.3 Multi-criterial error analysis

For an alternative evaluation of the filter methods with hydrological data, satellite and leakage errors of the different filters were evaluated in a multi-criterial way (Fig. 2.10). The leakage error was derived as an RMS of differences between filtered and unfiltered A3HM time series in monthly TWS variations. The satellite error was derived as an RMS from a monthly propagation of the calibrated coefficient errors into the basin averages. The total error is given by the squared sum of both error components. Hence, the point closest to the origin in Fig. 2.10 provides the smallest total error and indicates the optimum filter type according to the error budget.

The error budgets of the Amazon (Fig. 2.10a) and the Ob (Fig. 2.10d) basin show a well defined ranking between the filter methods. The decorrelation method VI is superior in reducing the total error. The second best error budget is provided by filter III. Furthermore, the more complex the shape (e.g., Nile) or the smaller the size of a river basin (e.g., Indus), the larger is the total error and the smoother are the error graphs in Fig. 2.10. For the Indus and the Nile basin, leakage for low smoothing rates amounts to several millimetres. Also, for these critical basins' characteristics, filters VI and III, respectively, provide the filter versions with the best error budgets, though versions of method II are located close to the optimum as well. The method V exhibits high leakage errors in the error plots of all basins, which explains the $wNSC$ results from above. Hence, method V is not a generally efficient filter approach for deriving basin-averaged TWS variations from GRACE gravity fields.

In summary, the order of the filters in terms of their total error budget in Fig. 2.9 closely matches the filter type ranking by the $wNSC$-evaluation (Table 2.5). For most cases, the decorrelation method VI provides the best error budget. For a similar satellite error reduction in the GRACE data, the leakage error of VI is much smaller compared to the other filter methods. This explains the small seasonal amplitude damping and phase shifts for this method in many river basins (Table 2.4). A list of TWS satellite, leakage and total error for the 22 river basins after application of the optimised decorrelation method VI, is provided in Table 2.7. The comparison of these total error values with estimations of seasonal TWS amplitudes from A3HM (Table 2.3) indicates that the estimation of GRACE-derived seasonal water mass variations is not reliable for the Amur, Indus, Nelson and Orange basins, as the error exceeds the signal magnitudes. This coincides with the small correspondence of GRACE and A3HM-derived time series of TWS variations for these river basins in Table 2.5.

Besides method III in the Indus basin, the best filter methods found for each river basin in the previous section, by the $wNSC$-evaluation (black circles in Fig. 2.10), are very close to the minimum satellite and leakage error budget. This result confirms the broader validity of the optimum filter selection procedure.

2.3.2.4 Sensitivity to errors in amplitude of the hydrological data

In Sect. 2.3.1 it was shown that the differences in TWS variations between the hydrological models consist of amplitude differences rather than phase shifts. Consequently, the influence of amplitude errors in the hydrological data on the $wNSC$-evaluation of filter parameter and

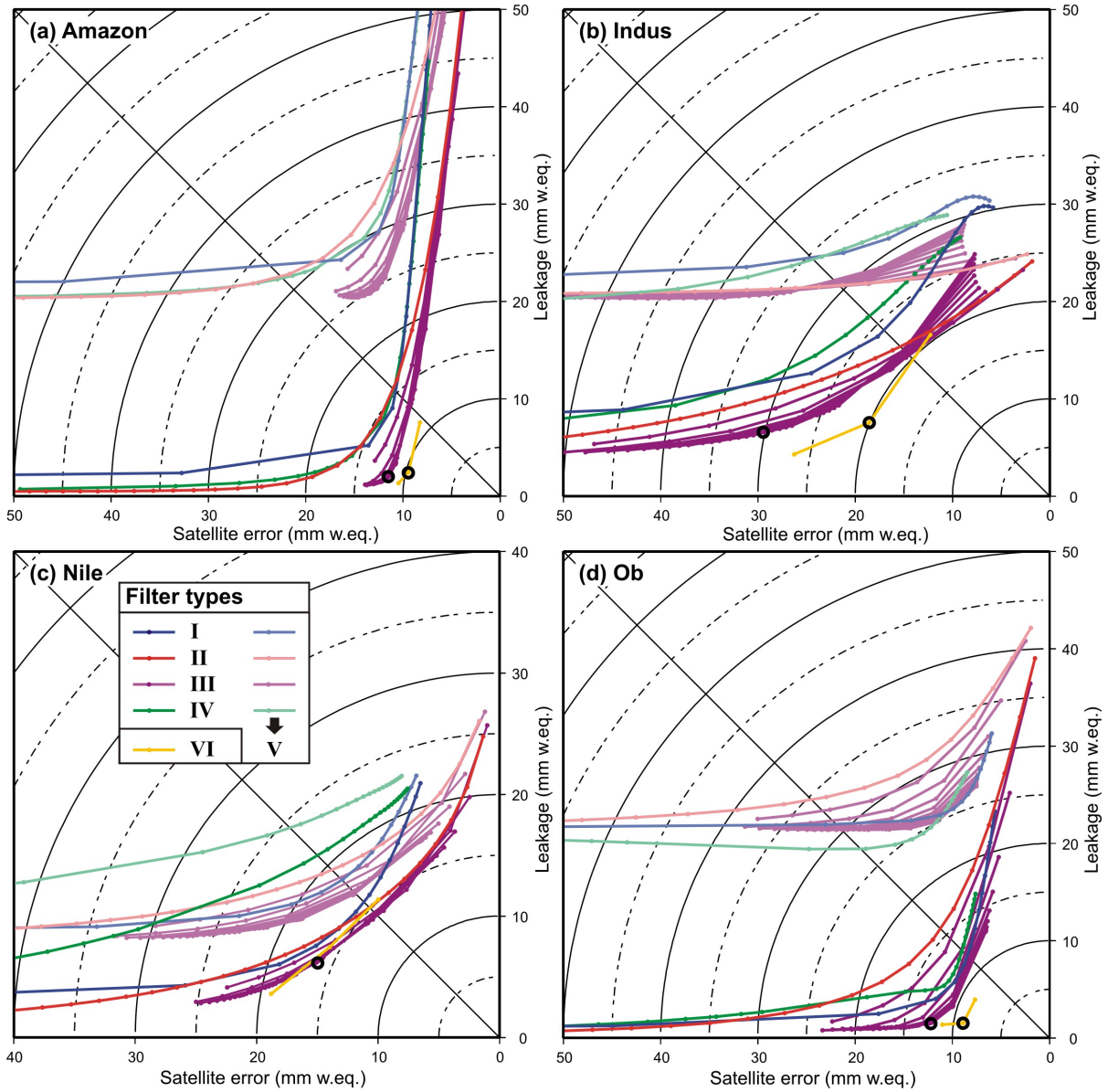


Figure 2.10: Error budget as hydrological leakage error (RMS differences of filtered and unfiltered A3HM time series) versus GRACE satellite error (RMS of propagated monthly coefficient errors) of TWS variations of : (a) the Amazon, (b) Indus, (c) Nile and (d) Ob river basins. Applied filter methods are shown by lines in blue: Gaussian (I), red: basin optimised (II), purple: signal model optimised (III), green: SNR optimised (IV). Light coloured lines: decorrelation filter V. Yellow line: decorrelation filter VI. Black circles indicate the individual optimised filter parameter from Table 2.5 and the bold black circles indicate the respective optimal filter method. See a filter description in Sect. 2.2.1.

Table 2.7: 22 basin individual standard deviation of monthly satellite and leakage error for decorrelation VI optimised filtering (see Table 2.5 for respective parameter).

Basin	ϵ_{sat} [mm]	ϵ_{leak} [mm]	ϵ_{ges} [mm]
Amazon	9.4	2.4	9.7
Amur	11.1	2.6	11.4
Danube	21.6	4.5	22.1
Ganges	17.1	4.2	17.6
Indus	18.6	7.6	20.1
Lena	10.0	1.3	10.1
Mackenzie	12.3	2.2	12.5
Mississippi	11.0	1.6	11.1
Nelson	16.2	5.6	17.1
Niger	18.3	2.0	18.4
Nile	14.0	7.5	15.9
Ob	8.9	1.5	9.0
Orange	12.9	4.7	13.7
Parana	16.3	5.8	17.3
St. Lawrence	10.3	16.7	19.6
Tocantins	32.0	12.3	34.3
Volga	11.2	2.6	11.5
Yangtze	14.9	1.4	15.0
Yenisei	8.0	2.8	8.5
Yukon	16.3	11.6	20.0
Zaire	11.9	5.3	13.0
Zambezi	20.7	3.6	21.0

methods has to be estimated. Therefore, a second $wNSC$-evaluation is undertaken in this section. Ahead of filtering and $wNSC$ evaluation, the monthly A3HM grid data are multiplied by a factor of 1.5. This factor is estimated as an average maximum difference between A3HM and GRACE TWS amplitudes. Subsequently, the $wNSC$-evaluation is repeated. Normalised differences of the optimised filter parameters relative to the ones optimised with the original A3HM data (Table 2.5) are shown in Fig. 2.11. The results for the Nelson and St. Lawrence basins are excluded because $wNSC$ values below zero were obtained and, therefore, no optimised parameter values could be achieved. Differences for the other basins mainly occur for parameters of filter III, which exhibits a low sensitivity for its filter parameter concerning filter performance (see Fig. 2.9). Parameter selection of VI also shows differences (Amur, Ob, Orange, Zaire), but here as well, the sensitivity of filter performance is low (see Sect. 2.3.2.2). Parameter differences of I, II and IV are either zero or are of expected evaluation uncertainties of one or two parameter step sizes (100 km, 2 mm and 1, respectively). Hence, except for the Nelson and St. Lawrence basins, these results prove that a possible error in the amplitude of the hydrological data would have small effects on the selection of optimal filter parameters by the $wNSC$-evaluation. This robustness of filter parameters is mainly due to the identical filtering of both data sets (GRACE and hydrological data) for the $wNSC$ evaluation, in combination with an accounting of leakage errors by the weighting factor w in $wNSC$. This approach prevents a simple fitting of GRACE to hydrological amplitudes, since amplitude damping affects both data sets.

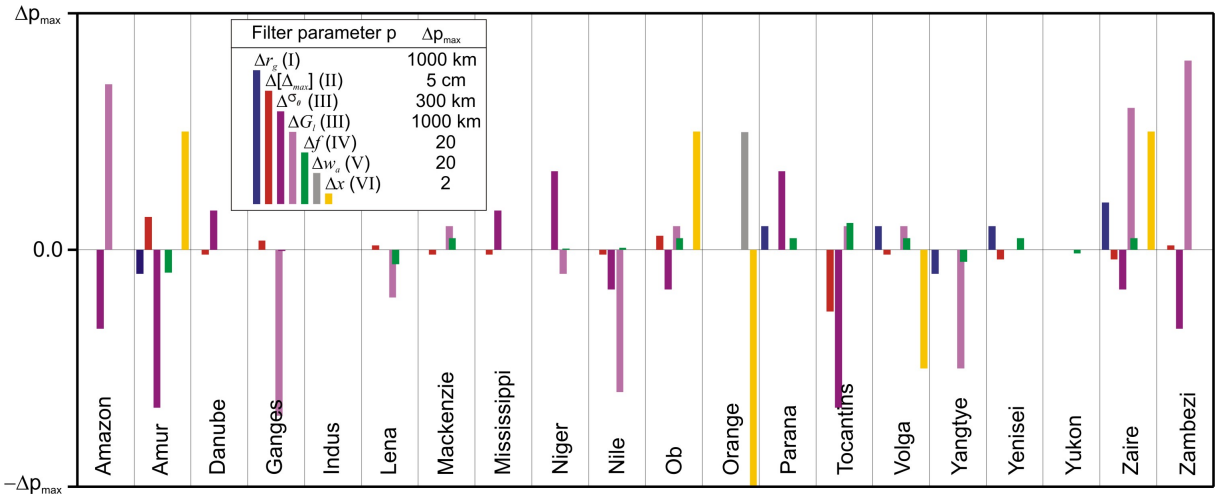


Figure 2.11: Parameter differences for Gaussian (I, blue), basin optimised (II, red), signal model optimised (III, light and dark purple), SNR optimised (IV, green), decorrelation V (grey) and decorrelation VI (black) filter methods for 20 river basins. Parameter differences are derived from $wNSC$-evaluations with A3HM versus a modified A3HM version of one-and-a-half times increased signal amplitudes: Δp = p(A3HM) - p(1.5*A3HM). In the graphic, differences are normalised by the maximum parameter values chosen in this study (see legend).

2.4 Conclusions

The results show that filter types and their corresponding parameters have to be selected carefully in order to derive basin-averaged time series of water storage variations from GRACE spherical harmonic data. The different smoothing effects of the different filter methods lead to varying balances of satellite and leakage errors in each river basin. We could determine the individual best filter methods for deriving basin-averaged water mass variations for the 22 largest river basins worldwide. When being evaluated by global hydrology, optimal parameters of the individual filter types vary for different basin sizes, shapes, and locations, as well as for signal type and intensity. Filter type VI provides generally good results. The differences of signal characteristics, like seasonal amplitude and phase, inside and outside a region of interest, highly influences the efficiency of a filter method. If phase shifts due to signals outside the river basin affect the TWS estimation, a bias or amplitude correction by a scale factor will not adequately recover the signal. Instead, a previous selection of an optimal filter type is expected to allow for a best possible bias correction. Additionally, for many filter types, the selection of an optimal parameter for the specific location and shape of the basin or process is necessary.

The decorrelation method VI was be the most efficient approach for the set of river basins analysed in this study. Only for basins of generally poor agreement between GRACE and hydrological data (Amur, Orange and St. Lawrence), was there a considerably higher correspondence provided by other filter methods. The usage of GRACE orbit-configurations to design a synthetic error covariance matrix sufficiently reduces the satellite error while preserving most of the hydrological signal for most of the river basins - even if they exhibit a small size or complex signal characteristics (e.g., Danube). To conclude, the general and global adaptability with moderate parameter sensitivity makes method VI the most reliable of the six analysed filter tools.

The isotropic Gaussian filter technique (I), or filter methods with little information on total error characteristics (II, IV, V), does not sufficiently address globally varying conditions for the extraction of basin-averaged TWS variations. Strong smoothing filter versions of method II and III tend to have more pronounced amplitude damping, while filters I and IV generally lead to phase shifts in the time series. Thus, the particular parameter values for these filter methods must be chosen carefully. It was shown that filter I gives acceptable results if the signal around the basin exhibits equal characteristics (Ob, Lena). If the river basin is characterised by a strong signal of TWS variability and is of large size and circuit shape, like the Amazon, leakage effects may be small if the parameter of filter I is chosen carefully. But if the signal around the basin is of different characteristics because of such factors as the vicinity of oceans (Ganges, Yukon), deserts (Nile, Indus) or smaller signals in surrounding regions (Danube), leakage may reach high values for filter I. Furthermore, the leakage effect may be time-dependent in cases in which surrounding areas are characterised by a different seasonal water storage regime. In this case, method I is inappropriate. The principle of method II is only advantageous when its parameter values are optimised and the river basin of interest exhibits a complex shape or small size (e.g., Niger, Yukon and Zambezi). Method III can efficiently deal with similar (e.g., Ob) or different (e.g., Nile) signal characteristics outside the area of interest, as well as with small or complex basin shapes (Niger, Zambezi). It provides good filter results for half of the river basins due to its efficient leakage estimation with an exponential signal model. Because of high leakage effects, the decorrelation method V only provides satisfactory filter results in some basins (Orange, Lena). However, when using method V for decorrelation additional smoothing, and therefore parameter optimisation, is necessary. A non-practical basin-based (instead of global) optimisation of the decorrelation parameter, in addition to the parameter of the superimposed filter method, may lead to improved results for that method. The conclusions above have been supported by reduction of amplitude and phase differences, total error budget maps and an amplitude sensitivity test.

It should be remembered that the results are derived from comparisons with model-based hydrological data, which might contain structural errors within specific basins. Such errors are caused by mis-modelled or missing processes, within any model, erroneous mode-forcing data or parameters. However, global hydrological models provide the only source of alternative TWS data sets for evaluating GRACE data. The ensemble mean A3HM was used as a compromise between three widely used hydrological models (GLDAS, LaD and WGHM) in order to evaluate GRACE filter parameters with the best possible accuracy, at present. Furthermore, the applied models mainly differ in amplitudes of TWS and it is shown that an amplitude error has a small effect on the filter evaluation. Only for the Indus, Amur, St. Lawrence, Orange and Nelson, may filter parameter type and selection be unreliable. A rather low correspondence between modelled and GRACE-derived data sets is due to temporal uncertainty of simulated TWS variability for these basins. The results in terms of optimum filters were shown to be robust, both for seasonal and non-annual TWS dynamics, in river basins. Nevertheless, for other applications, such as those with a focus on spatial patterns or secular trends, another prioritisation of filter methods may be more appropriate.

Hence, the discussion of adequate filter methods will likely continue as long as there is no breakthrough in accuracy for GRACE or GRACE-Follow on gravity field models. Filter types

and parameters, as derived in this study, are particularly useful for GRACE data analysis within hydrological applications, such as monitoring of water mass exchange on the continent, studies of inter-annual variability in the hydrosphere, or using reliable water storage data as input for assimilation into large-scale hydrological models. The results obtained here can also be used as a guideline for filter selection for areas that were not specifically considered in this study.

Acknowledgements

We wish to thank P.C.D. Milly, M. Rodell, J. Alcamo and P. Döll for providing the LaD, GLDAS and WGHM model data, respectively. Thanks to M. Scheinert for his GravTools software. The German Ministry of Education and Research (BMBF) supported these investigations within the geo-scientific R+D program GEOTECHNOLOGIEN „Erfassung des Systems Erde aus dem Weltraum". We also thank two anonymous reviewers for their constructive comments, which helped to improve the manuscript.

3 Calibration Method Development: "Integration of GRACE mass variations into a global hydrological model"*

Abstract

Time-variable gravity data of the GRACE (Gravity Recovery And Climate Experiment) satellite mission provide global information on temporal variations of continental water storage. In this study, we incorporate GRACE data for the first time directly into the tuning process of a global hydrological model to improve simulations of the continental water cycle. For the WaterGAP Global Hydrology Model (WGHM), we adopt a multi-objective calibration framework to constrain model predictions by both measured river discharge and water storage variations from GRACE and illustrate it on the example of three large river basins: Amazon, Mississippi and Congo. The approach leads to improved simulation results with regard to both objectives. In case of monthly total water storage variations we obtained a RMSE reduction of about 25 mm for the Amazon, 6 mm for the Mississippi and 1 mm for the Congo river basin. The results highlight the valuable nature of GRACE data when merged into large-scale hydrological modelling. Furthermore, they reveal the utility of the multi-objective calibration framework for the integration of remote sensing data into hydrological models.

*Werth, S., Güntner, A., Petrovic, S., Schmidt, R. (2009a), Earth and Planetary Science Letters, 270(1-2), 166-173.

3.1 Introduction

By mapping time variations of the Earth's gravity field with the Gravity Recovery and Climate Experiment satellite mission (GRACE) since its launch in 2002, an unprecedented global data set of mass variations close to the Earth surface became available (Tapley et al., 2004b). After removal of mass variations due to tides and non-tidal atmospheric and oceanic transport processes, the time-variable gravity data mainly represent water mass variations in continental hydrology, i.e., total water storage change (TWSC) on the continents (see a recent review by Schmidt et al., 2008b). In specific regions, also mass variation from post glacial rebound (Tamisiea et al., 2007) and seismic activities (Chen et al., 2007b) could be revealed from the GRACE data.

For the field of hydrology, the past six years of GRACE operation contributed to a significantly improved understanding of the spatio-temporal patterns of water storage variations on the continents because no comprehensive TWSC data were available before at large spatial scales due to the absence of adequate monitoring systems (Lettenmaier & Famiglietti, 2006). Thus, the GRACE TWSC data give new insights into the Earth's water cycle including the contribution of TWSC to sea level variations (Ramillien et al., 2008a), the impact of climate variability or extremes on water storage (e.g. Andersen et al., 2005; Seitz et al., 2008), or melting of glaciers and ice caps (e.g. Chen et al., 2006a; Luthcke et al., 2006). Numerous regional or river basin studies analysed GRACE TWSC from seasonal to inter-annual time scales (see a recent review by Schmidt et al., 2008b). Others solved the water balance using TSWC from GRACE for other hydrological components such as evapotranspiration (Rodell et al., 2004a; Ramillien et al., 2006) or runoff (Syed et al., 2007), or separated individual storage compartments such as groundwater (Rodell et al., 2006; Strassberg et al., 2007) or snow (Frappart et al., 2006; Niu et al., 2007a).

Besides observation data, hydrological simulation models are an indispensable tool to assess the impact of environmental change on the continental water cycle and the particular processes mentioned above. Thus, in turn, they are a prerequisite for implementing measures of sustainable management of water-related issues in future. At continental to global scales, hydrological models are an integral part of atmospheric circulation models where they represent the land surface processes for climate and weather prediction simulations, see Dirmeyer et al. (2006) for an overview on land surface models and their comparison. In addition, water balance models are used to represent the full water cycle in river basins for purposes such as stream flow forecasting and water resources assessment (for a recent overview on global water balance models see Widen-Nilsson et al., 2007). However, these large-scale hydrological models are known to suffer from uncertainties in terms of model structure, parameter values and climate forcing data. As a consequence, simulation results for hydrological state variables and water fluxes on the continents vary considerably between models (e.g. Dirmeyer et al., 2006). While river discharge has for a long time been the only observable to validate and calibrate global water balance models (Hunger & Döll, 2008), considerable model uncertainties remain for other components of the water cycle, e.g., water storage, evapotranspiration or groundwater recharge due to the lack of adequate observation data.

In this context, GRACE provides a unique data set to evaluate and improve the simulation of TWSC on large scales and therewith to uncover shortcomings in model designs and parameters. Numerous studies compared GRACE-derived TWSC data with simulation results of hydrological

models and concluded with a recommendation to use GRACE data as a model constraint (see a recent overview by Güntner, 2009). First attempts have been made to modify large-scale hydrological models and to evaluate the modifications with GRACE observations (Niu & Yang, 2006; Ngo-Duc et al., 2007) and very recently, Zaitchik et al. (2008) assimilated GRACE TWSC into a land surface model for the Mississippi river basin. A global integration of GRACE data with hydrological models to improve model performance by calibration has not been reported so far.

This motivated the present study to incorporate for the first time GRACE data into the tuning process of a global hydrological model (Sect. 3.2.1). For this purpose, a multi-objective calibration scheme has been developed (see Sect. 3.2.2). Calibration denotes the selection of model parameter values by evaluating the simulation performance via a model output objective against observations. In contrary to data assimilation, the system is tuned by determining model parameter values during a pre-defined time interval, and the resulting parameter set may be used for subsequent independent model runs. Multi-objective calibration denotes that more than one model output objectives are taken into consideration. In this study, two different types of measured data are used to constrain parameter sets (Sect. 3.2.3). Improvements for the simulation of TWSC are analysed (in Sect. 3.3) and the value of calibration procedure using GRACE data towards enhanced predictions of the continental water cycle is outlined (Sect. 3.4).

3.2 Methods and Data

3.2.1 Global Hydrological Model

The WaterGAP Global Hydrology Model (WGHM) is a conceptual water balance model which simulates the continental water cycle including the most important water storage components, i.e., interception, soil water, snow, groundwater and surface water. The major hydrological processes are simplified by conceptual formulations. WGHM has a 0.5°x0.5° spatial resolution and a daily computation time step. Information on land surface characteristics such as the spatial distribution of vegetation, soil types, land use, groundwater and surface water bodies is given in the model from global data sets. For details on model equations and their parameters see Döll et al. (2003). The model has widely been used to analyse continental water storage change (Güntner et al., 2007b). In comparisons with GRACE TWSC, a general agreement of seasonal and other periodic characteristics of TWSC was found at the global scale, but amplitudes and phases in the model showed significant differences (larger than GRACE errors) in particular river basins (Ramillien et al., 2005; Schmidt et al., 2006, 2008c).

In this study, WGHM is driven by climate data (temperature, cloudiness and number of rain days per month) of the European Centre for Medium-Range Weather Forecast (ECMWF) and monthly precipitation data of the Global Precipitation Climatology Centre (GPCC). Precipitation is disaggregated to a daily resolution with the given number of rain days per month. The climate input data are available from 01/1992 until 12/2007 for this study. Antarctica and Greenland were excluded from the simulations.

We used the most recent WGHM version as described by Hunger & Döll (2008), who calibrated (i.e. tuned) the model against observed mean annual river runoff at 1235 discharge stations

worldwide, by varying one runoff generation parameter. This model version is called the original version in the following. Overall, the model includes 26 process parameters. Their values in the original model as well as parameter ranges for the calibration are based on literature and qualitative reasoning (Kaspar, 2004), see Table 3.2 for the parameters calibrated in this study. Thereof, the parameter root depth is based on the global land cover distribution and can be calibrated by a multiplicative factor. The Priestley-Taylor coefficient is used in the corresponding approach to quantify potential evapotranspiration. The radiative fraction of the extraterrestrial radiation that reaches the Earth's surface is determined by cloud cover data and the radiation proportion parameter. The variability of snow melt temperature is due to different elevation and vegetation cover of different regions. A more detailed description of the model parameters is provided by Döll et al. (2003).

3.2.2 Calibration Technique

Combining both the present station-based accuracy of WGHM in terms of river discharge and the integrative nature of the GRACE data with global coverage, improved simulation results were expected from a multi-objective calibration approach. Calibration in the sense used here denotes an iterative method of testing different parameter values and selecting the best parameter sets based on performance criteria that evaluate simulation results against observation data. Calibration methods differ in their strategies to select parameter sets for each iteration from the given parameter space. Furthermore, multi-objective calibration denotes the selection of parameter values through evaluating model performance against more than one objective. In this study, these objectives are based on two observation data sets: river discharge and periodic TWSC (see Sect. 3.2.3); hence, it is a two dimensional problem. Instead of a single optimum parameter set, such an approach will lead to a Pareto set of optimal solutions (Gupta et al., 1998). Each Pareto optimum of this set is an optimal solution from a multi-objective point of view in the sense that no other solution exists that provides a better simulation performance for both model output objectives. Hence, when moving from one Pareto solution to another, simulation performance increases for one objective while it decreases for the other objective. Without additional information it is not possible to undertake a ranking among the Pareto solutions. The trade-off (i.e. the spread) between the Pareto solutions reflects the minimum parameter uncertainty (Vrugt et al., 2003a) caused by errors in the input and the measured data as well as by model structure.

The calibration of a number of model parameters against more than one objective depicts a highly non-linear optimisation problem and requires a global optimisation method. Furthermore, only stochastic methods like a multi-start simulated annealing or an evolutionary algorithm assure a feasible computing time for the calibration of the global hydrological model WGHM. Therefore, to handle the complexity of a multi-objective and multi-parameter calibration problem as well as the computational demands we select the ϵ-Non-dominated-Sorting-Genetic-Algorithm-II (ϵ-NSGAII) (Kollat & Reed, 2006), which ranks among the most effective and efficient multi-objective optimisation methods (Tang et al., 2006). This global optimisation algorithm solves multi-objective problems using the concept of evolutionary parameter variation (mutation, crossover and selection). It is an elitist algorithm with a Pareto ranking routine. Furthermore, as an extension of NSGAII (Deb et al., 2000) by the concept of ϵ-dominance, it

allows to specify the accuracy to be fulfilled by each objective. For this study, we parameterise its operators as proposed by Kollat & Reed (2006). Furthermore, we use a population size of $N = 8$ and an ϵ-resolution of 0.05 for both objectives and stop the optimisation after 400 iterations.

The calibration of WGHM is exemplarily done for the Amazon, the Mississippi and the Congo river basins in this study. These basins were selected because of their large size of over three million km^2. The period 01/2003-12/2006 was used for WGHM calibration.

Güntner et al. (2007b) showed that WGHM parameter sensitivity for TWSC simulations varies considerably between the river basins. This inter-basin variability of parameter sensitivity can be explained by differences of the climatic conditions (represented in the model by the climate input data and parameters steering evaporation or snow melt processes, for instance) and of the land surface properties (represented by, e.g., vegetation or soil parameters) between the river basins. This results in different water flow and storage characteristics in the basins. In particular, different storage components dominate the individual river basin response, e.g., snow storage in higher latitude areas or surface water storage in some tropical areas with large inundation zones. Thus, also the sensitivity of model parameters used to govern these individual dominant storage processes varies between the river basins. Consequently, ahead of the calibration work, a sensitivity study was undertaken by a Latin Hypercube sampling for 2000 parameter sets and by an analysis scheme going back to Hornberger & Spear (1981), who selected sensitive parameters based on their ability to provide behavioural model simulations. For each river basin, we selected the six most sensitive parameters for calibration against TWSC and river discharge (see row (e) and row (f) of Table 3.1). Parameter values and ranges are documented in Table 3.2.

For the Amazon basin, three of these parameters concern the process of surface water transport, because of the high water volume during an important flood season. In contrast, evaporation is most important in the tropical Congo river basin with a distinct dry season. A diverse set of important processes (e.g. snow, evaporation and surface water) provides the most sensitive parameter of the Mississippi river basin, due to its location in three different climate regions (cold in the north, subtropical in the southeast and dry in the southwest).

The evaluation of model performance for each iteration is effected by the following four steps: 1) Model simulation of monthly global TWSC fields and river discharge with the current parameter set. 2) Application of a GRACE-equivalent filter procedure, which comprises the conversion of WGHM TWSC fields into the frequency domain, i.e. spherical harmonic coefficients, followed by Gaussian smoothing (Jekeli, 1981) and the computation of basin averages of TWSC according to Wahr et al. (1998). 3) Fitting amplitudes and phases of significant periods which were determined from GRACE data (see Sect. 3.2.3.2) to the simulated basin averages of TWSC and reconstruction of a basin-average time series of TWSC from these periods. 4) Evaluation of each calibration objective (discharge and TWSC) by computation of the Nash-Sutcliffe-efficiency coefficient (NSC) (Nash & Sutcliffe, 1970) as a criterion of agreement between modelled and measured time-series.

NSC is a simulation performance measure that normalises the squared difference of a predicted (P) to an observed (O) time series by the variance of the observed values with n time steps:

$$NSC = 1 - \frac{\sum_{i=0}^{n}(O_i - P_i)^2}{\sum_{i=0}^{n}(O_i - \bar{O})^2}, \tag{3.1}$$

where $\bar{O}$ is the mean of the observations over the examined period. NSC evaluates both phase and amplitude agreement between two time series. It ranges from $-\infty$ to 1 (optimal fit), with a value of 0 indicating a simulated time series that performs as well as a model being equal to the mean of the observable. Therefore Pareto solutions are restricted to NSC values greater than 0.

3.2.3 Calibration data

3.2.3.1 River basin discharge: Objective 1

River discharge data of Amazon, Mississippi and Congo from the most downstream gauging station were used (Table 3.1). We computed monthly mean values for the calibration period. For the Congo river where no up-to-date measurements were available, we assigned the monthly mean discharge of earlier observations to the calibration period.

3.2.3.2 GRACE TWSC: Objective 2

Reconstructed significant periodic parts of basin-averaged TWSC resulting from the investigation presented in Schmidt et al. (2008c) are used as calibration input for this study. These data are chosen, because errors in the GRACE original data and the difficulty to separate the errors from real signals mark the greatest challenge for application of satellite gravity solutions.

Schmidt et al. (2008c) developed a technique to extract significant water storage change information from GRACE data by three steps: 1) Identification of the dominant spatio-temporal patterns in mass variations derived from GRACE observations through a principal component analysis (applied at the scale of the river basins to grids previously filtered by a Gaussian smoothing with a 500 km averaging radius), 2) Identification of significant periods of TWSC contained in the principal components without fixing a priori the period lengths, and 3) Reconstruction of (error-reduced) basin-average time series of TWSC from the significant periods.

As a basis, monthly GRACE-only time series of global gravity fields generated as spherical harmonic expansions up to degree and order 120 at the GFZ German Research Center for Geosciences (GRACE Level-2 products, version GFZ-RL04, Schmidt et al., 2008b) for the time period from 02/2003 until 12/2006 (excluding unavailable months 06/2003 and 01/2004) were used. The noise contained in the spherical harmonics increases with the degree of the expansion terms, and the noise/signal ratio reaches unacceptably high values in higher-degree terms. In the space domain this noise becomes visible in the form of the typical meridional-oriented spurious gravity signals ("stripes") (e.g. Swenson & Wahr, 2006; Schmidt et al., 2008b). Hence, a spatial filtering is mandatory when computing water storage variations from GRACE gravity field models in order to reduce these errors. For the present study a widely used Gaussian smoothing (Jekeli, 1981) with an averaging radius of 500 km was applied. Mass variations (TWSC) were derived relative to a mean field (i.e. in the form of mass anomalies) for the considered data period applying the procedure presented by Swenson & Wahr (2002).

Since the effects of the atmospheric and the oceanic circulations were previously removed in the course of the gravity field recovery from the raw GRACE data by applying appropriate geophysical models (Flechtner, 2007), the major part of the signal contained in the derived grids of mass anomalies can be attributed to hydrological variations. Due to the rather short time period covered by the available GRACE data, the long-term trends determined both from the

Table 3.1: (a) Re-calibrated river basins with (b) corresponding area and (c) discharge station. (d) Discharge source and time series for computation of monthly means. (e) Number of WGHM parameters from different processes (S: Soil, SW: Surface water, GW: groundwater, ER: Evaporation and Radiation, SN: Snow, IN: Interception) derived from a sensitivity study against TWSC and river discharge. The underlined process includes the most sensitive parameter. (f) Calibration parameter in corresponding order to row (e) (MCWH: maximum canopy water height, PT: Priestley-Taylor). (g) Significant GRACE derived TWSC periods P_n of basin averages with associated amplitudes A_n and phases ϕ_n, with t_0 =01.01.2005. (h) Cumulative proportion of the significant periods in the full GRACE signal variability.

(a)	Amazon	Mississippi	Congo
(b)	5.9 Mio km^2	3.0 Mio km^2	3.6 Mio km^2
(c)	Obidos	Tarbert Landing	Kinshasa
	1.9°S, 55.5°E	31.6°N, 91.5°W	4.3°S, 15.3°W
(d)	ORE HYBAM	US ACE	GRDC
	2003-2006	2003-2006	1903-1983
(e)	3 <u>SW</u>, 1 GW, 1 S,	1 <u>SW</u>, 1 S, 2 ER	2 <u>ER</u>, 1 S, 1 GW,
	1 IN	1 SN, 1 IN	2 SW
(f)	runoff coefficients	runoff coefficient	radiation proportion
	river velocity	root depth	PT coefficient
	wetland depth	radiation proportion	rooting depth
	GW baseflow coeff.	PT coefficient	GW baseflow coeff.
	rooting depth	snow melt temperature	wetland depth
	MCWH	MCWH	SW baseflow coeff.
(g)	P_1=0.9833 a	P_1=0.9826 a	P_1=0.9881 a
	A_1=146 mm	A_1=33 mm	A_1=30 mm
	ϕ_1=3.82 mon	ϕ_1=2.99 mon	ϕ_1=1.82 mon
	P_2=2.5297 a	P_2=2.4824 a	P_2=0.5022 a
	A_2=22 mm	A_2=22 mm	A_2=15 mm
	ϕ_2=19.29 mon	ϕ_2=29.59 mon	ϕ_2=4.75 mon
(h)	99%	75%	73%

Table 3.2: Calibration parameter values and their ranges for the calibration work (GW: groundwater, MCWH: maximum canopy water height, PT: Priestley-Taylor, SW: surface water).

Parameter	Standard value and unit	Minimum	Maximum
GW baseflow coefficient	0.01 / day	0.006	0.1
MCWH	0.3 mm	0.1	1.4
PT coefficient	1.26	0.885	1.65
radiation proportion	0.25	0.08	0.54
river velocity	1 m/s	0.05	2.0
root depth mult.	1	0.5	2.0
runoff coefficient mult.	1	0.5	2.0
snow melt temperature	0°C	-3.75	3.75
SW baseflow coefficient	0.01 / day	0.001	0.1
wetland depth	2 m	1.0	5.0

hydrology model WGHM and from the GRACE gravity fields should be regarded as less reliable than the periodic components resulting from the same data. Therefore, as the last preparatory step, the data used in this study have been de-trended.

Subsequently, the three-step strategy for the detection of significant periodic components, depicted at the beginning of this section, was realised, see (Schmidt et al., 2008c) for more details. It is important to note, that the period search was not a-priori constrained to seasonal or other postulated variations. For all three river basins, considered in this study, two periods resulted to be significant with respect to their signal proportion and an uncertainty study. Corresponding amplitudes and phases used for the calibration are given in Table 3.1, row (g). TWSC of all three basins exhibit a seasonal period. A second period of inter-annual scale (about 2.5 years) occurs for the Amazon as well as the Mississippi and of semi-annual scale for the Congo river basin. The cumulative variability of the reconstructed periodic components dominates the integral GRACE signal (see Table 3.1 row (h) for percentage proportion).

Error estimations of GRACE data differ between several studies. For example, using a Gaussian smoothing with an averaging radius of 750 km Wahr et al. (2006) derived latitude-dependant errors of GRACE mass estimates ranging from 8 mm near the poles up to 25-27 mm at low latitudes, when expressed in water column equivalents. This results in a global area-weighted mean of 21 mm. Schmidt et al. (2007) gave for a 500 km Gaussian filtering a global error estimate of 24-30 mm water column. According to Schmidt et al. (2008b) the accuracy of the GFZ-RL04 used in this study is approximately two times better than the accuracy of the earlier releases used in both cited studies. However, it should be taken into account that errors may be higher for particular regions and months, and are also influenced by leakage errors after forming basin-average values.

3.3 Results and Discussion

The multi-objective calibration of WGHM with GRACE TWSC and river discharge led to improved simulation results in all three river basins (Figure 3.1). Each Pareto solution (on the red line) is superior to the original model version (green dot) with regard to both objectives.

Best results were obtained for the Amazon basin. NSC performances better than 0.95 with respect to both objectives were achieved for the Pareto solution closest to the optimum (hereafter referred as the selected Pareto-optimum, blue dot in Figure 3.1a). The amplitude of periodic terms of TWSC increased markedly in the Pareto solutions when compared to the original model (Figure 3.2a). Since the narrow uncertainty band given by the Pareto set of solutions does not include the original model time series, the significance of model improvement is substantiated. Although the amplitudes of basin-average TWSC were slightly overestimated by the selected Pareto solution in 2003 and 2006, its root mean square error (RMSE) of the complete (but de-trended) TWSC signal was reduced by 50% compared to the original model version (Table 3.3). The reduction of RMSE for discharge was even greater, since a phase shift of discharge seasonality could be corrected by the multi-criteria calibration (see Figure 3.3a). A main reason for the model improvements in the Amazon basin could be attributed to longer residence times of surface water in rivers and floodplains as expressed by lower values for the flow velocity parameter in the Pareto solutions.

Table 3.3: RMSE of simulated versus detected hydrological states for the calibration period 01/2003-12/2006: monthly mean river discharge (col. 2-3) and de-trended TWSC signal with non-periodic components (col. 4-5). RMSE is given for the original WGHM (original) and for the selected Pareto solution of the re-calibrated (re-cal.) WGHM version. For the validation period 01/2007-12/2007 RMSE of TWSC is given in col. 6-7.

	Discharge [kg^3/month]		TWSC [mm]		TWSC [mm] (2007)	
Basin	original	re-cal.	original	re-cal.	original	re-cal.
Amazon	126.6	28.1	49.6	24.7	64.0	34.1
Mississippi	17.2	4.0	21.8	16.1	18.9	13.4
Congo	22.0	13.5	24.7	23.6	25.4	30.5

Also in the Mississippi basin a very good fit to observations with NSC performances of about 0.9 for both objectives were obtained for the selected Pareto-optimum (Figure 3.1b). Although the results for river discharge are more uncertain than for TWSC, the improvement compared to the original WGHM is greater for discharge than for TWSC. This is reflected by the reduction of the RMSE of the monthly mean discharge of about 80%, respectively 13 km^3/month (Table 3.3) for the selected Pareto-optimum. The clear improvement of monthly discharge simulations is also due to the fact that the original model was calibrated for mean annual values and did not take into account the seasonal distribution of discharge as in the present scheme. Therefore, the overestimated peaks of monthly discharge during spring in the standard model version could be corrected for all Pareto solutions (see Figure 3.3b). The reconstructed calibrated time series of water storage variations shows a slightly shifted phase and an amplitude which is closer to the GRACE time series (Figure 3.2b). The RMSE of the full de-trended time series of TWSC was improved about 6 mm compared to the original model version (Table 3.3). This improvement was most likely caused by changes of two model parameters. An increased effective root zone increases the soil storage capacity and an increased snow melt temperature smooths the previously overestimated runoff peaks.

Calibration for the Congo basin resulted in a much wider trade-off between both objectives (note the different scaling of both axes in Figure 3.1c). The performance of the Pareto solutions varies between 0.0 and 0.8 for discharge and between 0.7 and 0.9 for TWSC (Figure 3.1c). This trade-off resulted in a wider uncertainty band for the calibrated TWSC periods of the Pareto solutions (Figure 3.2c). Nevertheless, a small phase shift of TWSC periods was achieved for all Pareto solutions. The RMSE of the full TWSC signal for the selected Pareto-optimum was improved by about 1 mm (Table 3.3). All other Pareto solutions provide greater RMSE reductions, since they show a higher simulation performance for the significant periods of TWSC, as the selected Pareto-optimum. For discharge, there were slight improvements in the monthly regime (Figure 3.3c), as indicated by higher peaks during the turns of the year (from October till January) for the re-calibrated hydrograph of the selected Pareto-optimum. While the RMSE for discharge could clearly be decreased by the calibration procedure, the NSC value for the selected Pareto-optimum of 0.76 still indicates only moderate correspondence of simulated and observed river discharge. Though, the rather discontinuous course of the Pareto frontier may imply that a higher number of function evaluations would give better calibration results. These limitations

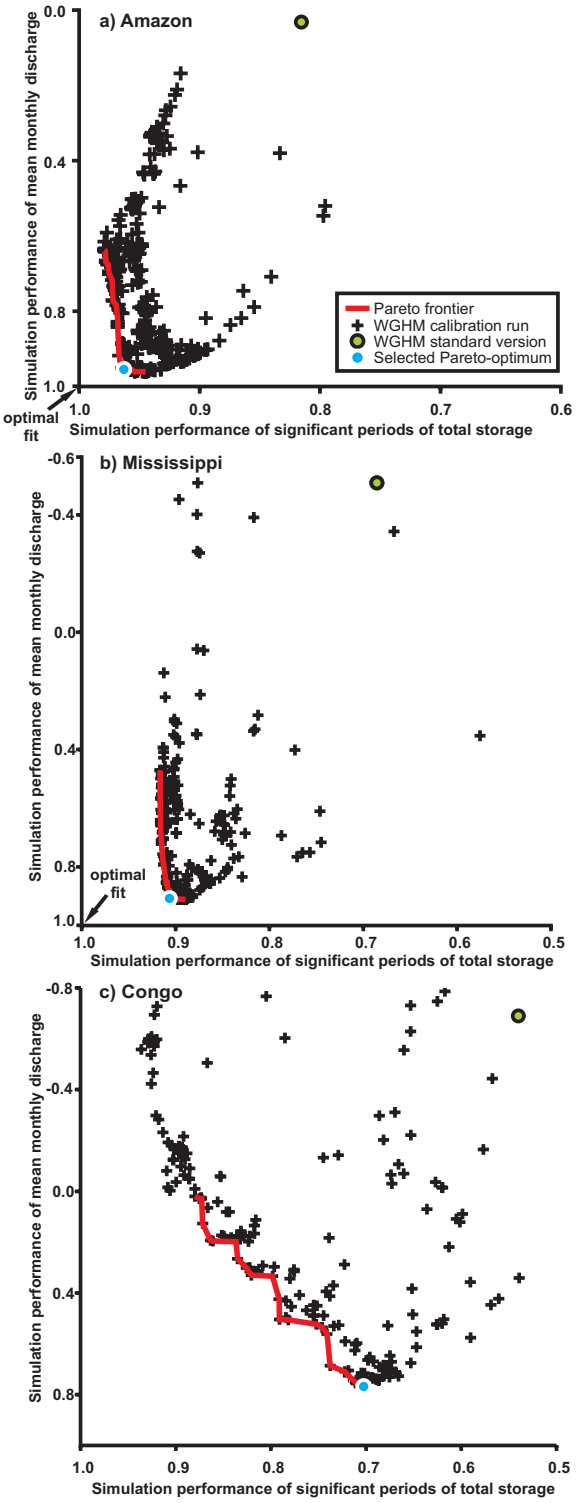


Figure 3.1: Calibration results in terms of NSC indicating the simulation performance for a) the Amazon, b) the Mississippi and c) the Congo river basin.

in achieving better discharge and TWSC simulations as well as the wider uncertainty in the calibration of the Congo basin are likely due to the lack of river runoff measurements during the calibration period and complicate the assignment of improved processes for the Congo basin. The particular characteristics of the rainfall distribution in each year will cause substantial deviations from the mean hydrograph that was used for model evaluation in this basin (Figure

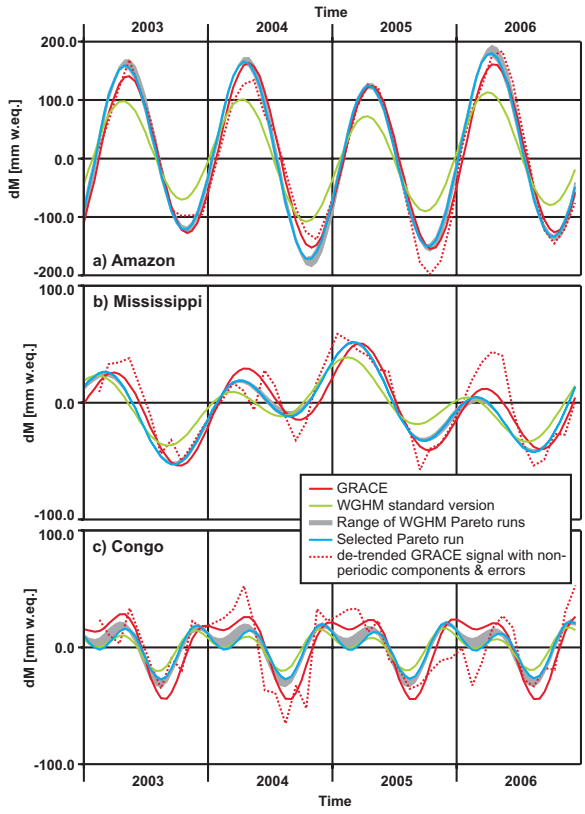


Figure 3.2: Calibration results in terms of time series of TWSC from reconstructed periodic terms for a) the Amazon, b) the Mississippi and c) the Congo river basin.

3.3c). This may also point out errors in the model structure, the model input data, or in the parameter space allowed for calibration in the Congo basin and is subject to further studies. Introduction of further observables to the multi-objective calibration scheme could further reduce the resulting equifinality of parameter sets as expressed by the dense Pareto-Frontier shown for Amazon and Mississippi. In particular, parameter values of storage processes that are represented by these additional observations could be more effectively constrained. For example, surface water storage derived from satellite altimetry and imagery can provide such data sets for an individual storage compartment (Papa et al., 2008). Though, the success will be limited as long as the observables contain high errors (e.g. groundwater, Döll & Fiedler, 2008) or the approach demands sophisticated model modifications to make model state variables match the observables (as for remotely sensed surface soil moisture).

A validation of the calibrated model was performed for de-trended GRACE signals including non-periodic components and errors from January until December 2007 (see Figure 3.4). For this year, a simulation run was realized with WGHM using the parameter values that were calibrated for the period 2003-2006. For the Amazon and the Mississippi river basins, simulation results were markedly better for the validation period, when they are compared to the results of the standard model in terms of amplitude, phase and RMSE values. This improvement is similar to what was achieved in the calibration period (see Table 3.3). This corroborates the model improvement of TWSC that could be achieved by the multi-criterial calibration for these basins. For the Congo river basin, however, the RMSE value increased, indicating that the model performs somewhat worse with the re-calibrated parameter set in the validation period.

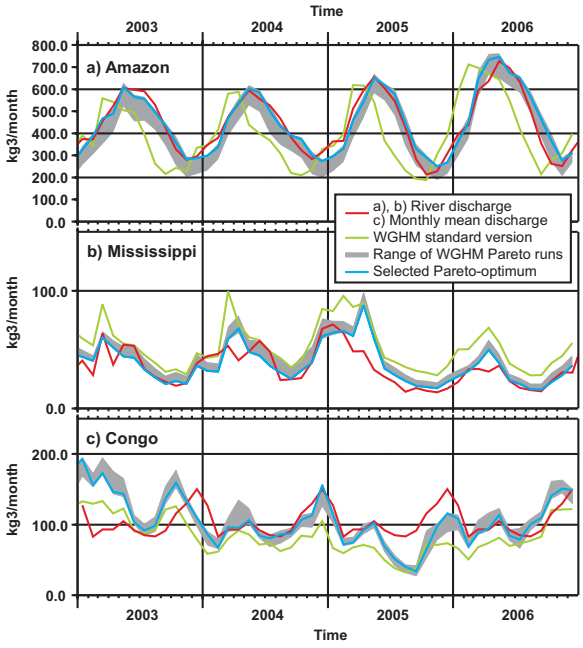


Figure 3.3: Calibration results in terms of measured time series of monthly river discharge for a) the Amazon, b) the Mississippi and of mean monthly river discharge for c) the Congo river basin. See Table 3.1 for detailed sources of discharge measurements.

This confirms the above results that improvements by calibration are difficult to achieve with the present model set up and data availability for this river basin. For further studies it should also be taken into consideration that it might be justified to reduce the weight assigned to the river discharge data during calibration in the Congo basin due to their high uncertainties. This may enable the selection of Pareto optima with higher TWSC-simulation performance (see Figure 3.1c).

3.4 Conclusions

The first multi-objective calibration of the global hydrology model WGHM with TWSC data from GRACE and monthly mean river discharge was successfully carried out. By this approach, phase and amplitude differences of periodic water storage variations between GRACE and WGHM could be significantly reduced as compared to earlier versions of WGHM. We could show that the direct integration of GRACE data into the calibration process of WGHM leads to a clear improvement of simulated monthly TWSC signals on a scale of large river basins. At the same time, a better simulation of river discharge could be achieved. This highlights the particular value of multi-objective process analyses. If two observables are considered within the calibration approach, the trade-off in model performance of different hydrological variables is taken into account. Finally, this allows for an improved representation of the water balance as a whole.

It should be pointed out that the calibration approach adopted in this study followed two principles that can be seen as a prerequisite for the successful integration of GRACE water storage data into large-scale hydrological models (Güntner, 2009). First, GRACE and WGHM model data were treated exactly in the same way before comparison and parameter adjustment, i.e., the same methods of filtering and basin-averaging were applied to both data sets. This

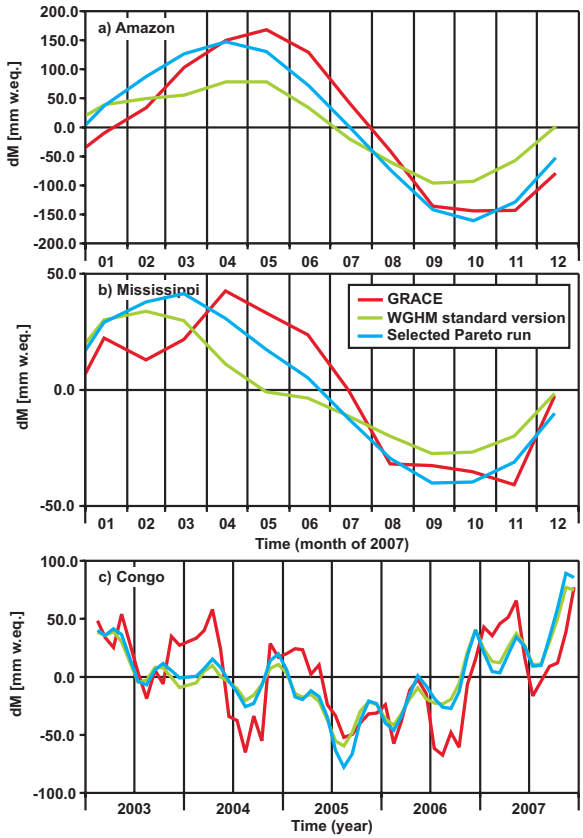


Figure 3.4: Validation of de-trended TWSC by model simulations during the period 01-12/2007 (which was not used for calibration) for a) the Amazon, b) the Mississippi and c) the Congo river basins.

excludes the risk of poor comparability of the time series if unfiltered model data are compared to filtered GRACE data which may include filter-induced biases. Secondly, with WGHM a hydrological model was used that represents all relevant water storage compartments in the analysed river basins, including surface water storage. Thus, it is assured that water storage calibrated in the model is consistent with the observation variable, i.e., the integrative nature of GRACE-based TWSC.

A better process understanding in global hydrology is necessary to provide more reliable estimates of changes in the continental water cycle, which constitutes an important input for climate studies or water resources management. In order to get a closer view into the reasons why the model differs from the real world, more accurate input data and improved calibration settings should be applied. The former can be achieved by using up-to-date river discharge data (i.e. for the Congo basin) and better GRACE filter methods. For the latter, technically more extensive model calibrations in terms of the size of parameter set population and of function evaluation are necessary to shift the Pareto frontier towards an even better model performance. Also, the analysis of a posteriori model states and parameter sets will help to uncover potential errors in model structure or input data. In this way, an improved understanding of continental water storage processes may finally be achieved by a stepwise modification of the modelling concept (Fenicia et al., 2008). Especially for regions like the Congo river basin with a very inaccurate or lacking coverage of terrestrial data, the usage of GRACE data is most proliferous concerning model

improvement. Longer GRACE time series and the continuing error reduction within GRACE gravity recovery are likely to reduce the uncertainty of GRACE TWSC recovery and therefore the data assimilation into global hydrology modelling in further studies. Additionally, the presented approach is promising for the integration of alternative data sets from remote sensing, such as soil moisture, snow cover or surface water volumes into hydrological models. Furthermore, the methods considered here to achieve consistency of model variables and GRACE observations in terms of, e.g., data filtering and the selection of dominant signals, may similarly apply to other areas of Earth system modelling where GRACE data are to be used as a model constraint, such as for processes of the cryosphere or the Earth's interior.

Acknowledgements

The German Ministry of Education and Research (BMBF) supported these investigations within the geoscientific R+D programme GEOTECHNOLOGIEN "Erfassung des Systems Erde aus dem Weltraum" under grant 03F0424A. Thanks to M. Scheinert for providing the GravTools software.

4 Calibration Analysis: "Calibration analysis for water storage estimations of the global hydrological model WGHM"*

Abstract

This study contributes to an improved global simulation of continental water storage variations by calibrating the WaterGAP Global Hydrology Model (WGHM) for 28 of the largest river basins worldwide. Five years (01/2003-12/2007) of satellite-based estimates of total water storage changes from the GRACE mission are combined with river discharge data in a multi-objective calibration framework of the most sensitive WGHM model parameters. The uncertainty and significance of the calibration results is analysed with respect to errors in the observation data. An independent simulation period (01/2008-12/2008) is used for validation. The contribution of single storage compartments to the total water budget before and after calibration is analysed in detail. A multi-objective improvement of the model states is obtained for most of the river basins, with mean error reductions up to 110 km^3/month for discharge and up to 24 mm of a water mass equivalent column for total water storage changes, as for the Amazon basin. Errors in phase and signal variability of seasonal water mass changes are reduced. The calibration is shown to primarily affect soil water storage in most river basins. The variability of groundwater storage variations is reduced at the global scale after calibration. Structural model errors are identified from a small contribution of surface water storage including wetlands in river basins with large inundation areas, such as the Amazon or the Mississippi. The results demonstrate the value of GRACE data and the multi-objective calibration approach for improvements of large-scale hydrological simulations, as they constitute a starting-point for improvements of model structure. The integration of complimentary observation data to further constrain the simulation of single storage compartments is encouraged.

*Werth, S and Güntner, A. (2009), Hydrology and Earth System Sciences - Discussions, 6, 4813-4861.

4.1 Introduction

In the face of global climate change, forecasts about water shortage accumulate for many regions and water shortage becomes an increasing social-humanitarian problem. Global hydrological models are indispensable to track consequences of the alternating climate and to study the dynamics of water resources distribution. For a reliable monitoring of the stability and dynamical behaviour of the water cycle, changes in the water budget (change in total water storage $\Delta TWS = P - E - R$) of specific regions like large river basins play a key-role. To simulate the water cycle, hydrological models are forced by e.g., precipitation (P) and different climatic conditions, to estimate flow and storage of water on the continents and its charge to other Earth's subsystems like atmosphere and oceans by processes of evaporation (E) and runoff (R), respectively. A consistent representation of the continental water cycle and its components are a major issue for hydrological modelling. Only recently, however, variations of TWS have become a key variable in evaluating large-scale models (Güntner, 2009).

Several large-scale or global hydrological models exist (see Dirmeyer et al., 2006; Widen-Nilsson et al., 2007; Liu et al., 2007, 2009; Milly & Shmakin, 2002a; Rodell et al., 2004b), but estimates of variations in the total water storage (TWS) differ largely between them. Werth et al. (2009b) compared global TWS variations (TWSV) of the conceptual WaterGAP Global Hydrological Model (WGHM) with two physically based land surface models (the Global Land Data Assimilation System, GLDAS and the Land Dynamics model, LaD) and discovered differences in the magnitude of the signal itself between the three models, though temporal correlations are high. Reasons are different input data and modelling strategies for representing storage and flow processes at the coarse scale. Also, there is still a lack of knowledge about the regional importance and characteristics of individual storage processes. For example, surface water storage or deeper groundwater are absent or inattentively treated in many land surface models (Güntner, 2009; Niu et al., 2007b).

Syed et al. (2008) assessed TWS variability of GLDAS on the global scale being too small and concluded that the absence of groundwater and surface water or uncertain snow parameterisations were possible reasons for model errors. For the land surface model ORCHIDEE, TWS amplitudes and phases could be improved by introducing a cumulative surface water and groundwater reservoir that allowed for a longer residence time of water in the river basins (Ngo-Duc et al., 2007). Recent regional studies focus on modelling of groundwater storage with land surface models (e.g., Gulden et al., 2007; Lo et al., 2008; Kollet & Maxwell, 2008) but groundwater is still absent in several large-scale or global models. Although the global model WGHM simulates the most important storages compartments, including surface water and groundwater, simulation accuracy of the conceptual model was originally low for river discharge in snow dominated and semi-arid regions. Here, difficulties in the representation of evaporation or snow accumulation appeared (Döll et al., 2003). In response, Hunger & Döll (2008) and Schulze & Döll (2004) improved model equations for both processes. For TWS, however, WGHM still tended to underestimate seasonal TWS variations and phase shifts appeared (Schmidt et al., 2008c, 2006). Güntner et al. (2007b) found a regional varying sensitivity of WGHM parameters. Since only one parameter of the original model has globally been calibrated so far, this calls for an extension towards a regional calibration with respect to dominant processes of a river basin.

Theoretical studies propagate an iterative working process of model prediction, model analysis and process understanding (e.g., Fenicia et al., 2008; Savenije, 2009). An evaluation of model predictions should be undertaken by comparisons of simulated states of the water cycle to real-world observations. Model behaviour during tuning processes like data assimilation (e.g., Houtekamer & Mitchell, 1998; Reichle et al., 2002) or model calibration (e.g., Duan et al., 2003; Gupta et al., 2005) provides information on process behaviour and structural model deficits. But, the learning process is especially difficult on the global scale and limited to iterative steps, primarily because of the lack of adequate model forcing and validation data with global coverage and acceptable resolution and accuracy.

In this respect, the Gravity Recovery And Climate Experiment (GRACE) is of extraordinary benefit for large-scale hydrological studies. With global coverage, monthly gravity observations from this twin-satellite-mission are transferable to the variability of water stored on and below the Earth's surface with a resolution of a few hundred kilometres (e.g., Tapley et al., 2004b; Wahr et al., 2004). After removal of atmospheric and oceanic gravity effects, GRACE observations enable temporarily reliable studies of different hydrological processes (like snow and ice, groundwater, soil, surface, as done by Wouters et al., 2008; Niu et al., 2007b; Swenson et al., 2008; Papa et al., 2008, respectively) that include different climatic conditions and extreme events for many regions (e.g., Zeng et al., 2008; Seitz et al., 2008) or the water balance itself (Sheffield et al., 2009). Since the first GRACE record became available, large progress has been made in order to improve GRACE data accuracy and, thus, the reliability of water mass variations from GRACE. These include studies on dealiasing (Han et al., 2004), error estimates (Horwath & Dietrich, 2006), development of filter (Swenson & Wahr, 2002) and decorrelation techniques (Kusche, 2007) as well as filter optimisation (Werth et al., 2009b). Consequently, GRACE depicts a valuable tool for validation and calibration of large-scale hydrological models (Schmidt et al., 2008b; Güntner, 2009; Lettenmaier & Famiglietti, 2006). Application of GRACE data for large-scale hydrological modelling started out with validation of simulated water storage variations for large river basins or with global coverage (e.g., Ngo-Duc et al., 2007; Syed et al., 2008; Güntner, 2009). More recently, promising further steps were made towards the integration of GRACE data into model development and model tuning for particular regions, e.g., the Amazon or Mississippi basin (e.g., Zaitchik et al., 2008; Werth et al., 2009a; Lo et al., 2010). As a subsequent step that makes full use of the global coverage of GRACE, a world-wide integration of TWS variations towards an improved simulation of continental TWSV as a whole would be desirable. But many combinations of simulated single storage compartments may lead to a good fit for the integrative GRACE TWS variations with only coarse resolution. Hence, to obtain additional model constraints, higher parameter accuracy (Yapo et al., 1998; Vrugt et al., 2003a; Gupta et al., 2005) and to reduce parameter equifinality (Beven & Binley, 1992), the combination with other system states, like river discharge, in a multi-objective method is promising. In addition, using GRACE-based TWSV and river discharge is of particular interest for water balance analyses as both are integrated measures of the hydrological dynamics in a river basin.

In this context, this study makes a step forward in the iterative learning process of large-scale hydrological modelling towards improved global simulation of the continental water cycle and its storage compartments by a multi-objective calibration (Sect. 4.2.2) of the global model WGHM (Sect. 4.2.1) against river discharge and GRACE-based estimations (Sect. 4.2.3) for 28 of the

largest and most important river basins world wide (Sect. 4.3.1).

4.2 Methods and Data

4.2.1 Global Hydrological Model

The WaterGAP Global Hydrology Model (WGHM, Döll et al., 2003) simulates the continental water cycle by conceptual formulations of the most important hydrological processes. WGHM was originally developed by Döll et al. (2003) for water availability studies at the continental scale (Alcamo et al., 2003, e.g.,). But since the model provides estimates of water masses, it may serve for for hydrological analyses of water storage and its global dynamics (Güntner et al., 2007b) as well as for individual storage compartments, such as groundwater recharge (Döll & Fiedler, 2008) or storage of surface water bodies (Papa et al., 2008). WGHM was numerously applied for comparison of continental water storage variability to GRACE-based water mass variations (Schmidt et al., 2006, 2008c).

The conceptual model equations of WGHM are described in detail by Döll et al. (2003), Kaspar (2004) and Hunger & Döll (2008). In general, if water precipitates as rain it is passed through the storages of interception, surface water (including rivers, reservoirs, lakes and wetlands), soil and groundwater, reduced for evapotranspiration losses. In case of precipitation falling as snow, it accumulates as snow storage and follows the above liquid water cycle after melting. Additionally, human water consumption is considered (Döll et al., 2003). Accumulation of ice or permafrost is not accounted for in WGHM (Hunger & Döll, 2008). The model is computed on a daily time step and cell-wise with a 0.5°spatial resolution, excluding Antarctica and Greenland, hence, 66896 grid cells world wide. The water passes from cell to cell according to a global drainage direction map (Döll & Lehner, 2002) until it reaches a coastal cell, where it discharges to the oceans. The simulations of the hydrological cycle are supplied by cell-based information on properties of soil, land cover, hydrogeology as well as on locations of reservoirs, lakes and wetlands (Döll et al., 2003).

A very recent version of WGHM as described by Hunger & Döll (2008) with updates for the input data for surface water bodies and human water consumption, an improved snow algorithm and a more realistic formulation of evaporation of lakes and wetlands was used in this study. To allow model runs for the GRACE period (2002 - to date), the model was forced by climate data (temperature, cloudiness and number of rain days per month) from the operational forecasts of the European Centre for Medium-Range Weather Forecasts (ECMWF). Monthly precipitation input from the Global Precipitation Climatology Centre (GPCC) was used. Precipitation data were corrected for precipitation measurement errors according to Legates & Willmott (1990) following Fiedler & Döll (2007). This model set up formed the reference of the present study and is hereafter called the *original model version*.

Döll et al. (2003) and Hunger & Döll (2008) tuned the original WGHM against long-term river discharge by a runoff coefficient parameter, which determines the fraction of effective precipitation that translates into runoff, depending on the saturation of soil water (Eq. 3, Döll et al., 2003). Both studies noted that calibrating this parameter only was not sufficient for some areas to get acceptable simulation results for river discharge because, for instance, the water

balance of lakes and wetlands is not influenced by this calibration approach, and because of other mis-modelled processes. Therefore, this study intends to calibrate WGHM parameters of all important process formulations besides runoff within a river basin (see Sect. 4.2.2.1). We consider calibrated parameter values as effective values that account for non-resolvable features in a large-scale model such as sub-scale variability, input data errors, model structure errors or simplifications in model equations.

WGHM consists of 36 model parameters. They are explained in detail in the publications of the original model versions while an overview of the 21 relevant WGHM parameters for this study is given below and in Table 4.1. The admitted parameter ranges for calibration were based on literature data and qualitative reasoning (Kaspar, 2004).

The soil storage capacity depends on the soil type and the land cover and is regulated by the *root depth* parameter. This parameter is calibrated as a multiplicative factor (SL-1), i.e., the particular value for soil storage capacity based on the soil and land cover data in each model cell is multiplied by the value of SL-1 (here in the range of 0.5 to 2, see Table 4.1). Groundwater storage and outflow is governed by the *groundwater baseflow coefficient* (GW-1).

Surface water transport may on the one hand be calibrated by the *river velocity* (SW-2). On the other hand, the *surface water flow coefficient* (SW-5) as well as the maximum range of water levels in lakes (*lake depth*, SW-3) and wetlands (*wetlands depth*, SW-4) determine storage rates of surface water bodies and are possible calibration parameter for surface water transport processes. Furthermore, the *runoff coefficient* parameter, which was tuned against river discharge for the original model versions, is calibrated as a multiplier (SW-1) in this study.

The potential evapotranspiration is computed in WGHM by the approach of Priestley & Taylor (1972) (PT). The equation is adjusted by the *PT-coefficient* that differentiates between humid (average relative humidity of 60% or more, ER-5) and arid regions (average relative humidity less than 60%, ER-6). The net radiation required as input for the PT-approach is computed according Shuttleworth (1993) (see Döll et al., 2003). Herein, the *radiation proportion* parameter (ER-1) is used to determine the radiation fraction of the extraterrestrial radiation that reaches the Earth's surface. The radiation fraction may be reduced by cloud cover following a *radiation correction* parameter (ER-2). The actual evaporation of open water can be calibrated by the *open water albedo* (ER-4) and sublimation of snow by the *snow albedo* (ER-3). Land surface evapotranspiration is limited by the *maximum potential evapotranspiration* (*MPET*, ER-7) parameter (see Döll et al., 2003).

Interception storage capacity depends on three parameters: The *maximum canopy water height* (*MCWH*, IN-1) as well as a *specific leaf area multiplier* (IN-2) and a *biomass multiplier* (IN-3).

The rates of snow melt and accumulation depend on land cover and elevation. Snow melt is computed in WGHM by a degree-day approach. The *degree-day factor* depends on the land cover type. It is calibrated in this study by a multiplicative factor (SN-3). Sub-grid variability of elevation within a 0.5 degree model cell is represented in WGHM (100 sub-units per 0.5°-cell) and elevation effects are accounted for by a *temperature gradient* (SN-4). Additional effects on snow storage processes can be adjusted by a cell-averaged *snow freeze temperature* (SN-1) and *snow melt temperature* (SN-2).

Table 4.1: Detail information on the calibration parameter (col. 1; MCWH: maximum canopy water height, MPET: maximum potential evapotranspiration, PT: Priestley-Taylor) is provided by belonging processes and numbering (col. 2; SL: Soil, GW: groundwater, SW: Surface water, ER: Evaporation and Radiation, SN: Snow, IN: Interception), original WGHM value (col. 3), minimum and maximum value (col. 4 and 5). Literature references to model parameter and according equation numbers are provided in col. 6 and 7, English references are preferred.

Parameter	Abbrev. number	Original value & unit	Min. value	Max. value	Literature reference	(Eq.) or page
Root depth mult.	SL-1	1	0.5	2	Kaspar (2004)	(2.26)
GW baseflow coefficient	GW-1	0.01/day	0.006	0.1	Döll et al. (2003)	(5)
Runoff coefficient mult.	SW-1	1	0.5	2	Döll et al. (2003)	(3)
River velocity	SW-2	1 m/s	0.05	2	Kaspar (2004)	(2.38)
Lake depth	SW-3	5 m	1	20	Döll et al. (2003)	(6)
Wetland depth	SW-4	2 m	1	5	Döll et al. (2003)	(6)
SW baseflow coefficient	SW-5	0.01/day	0.001	0.1	Döll et al. (2003)	(6)
Radiation proportion	ER-1	0.25	0.08	0.54	Kaspar (2004)	(2.11)
Radiation correction	ER-2	1.0	0.7	1.3	Kaspar (2004)	(2.13)
Albedo snow	ER-3	0.4	0.3	0.9	Kaspar (2004)	p. 19
Albedo open water	ER-4	0.08	0.03	0.5	Kaspar (2004)	p. 15
PT coeff. (humid areas)	ER-5	1.26	0.885	1.65	Kaspar (2004)	(2.4)
PT coeff. (arid areas)	ER-6	1.74	1.365	2.115	Kaspar (2004)	(2.4)
MPET	ER-7	10 mm/day	6.25	13.75	Döll et al. (2003)	(2)
MCWH	IN-1	0.3 mm	0.1	1.4	Döll et al. (2003)	(1)
Specific leaf area mult.	IN-2	1	-0.2	2.2	Kaspar (2004)	(2.19)
Biomass mult.	IN-3	1	0.25	1.75	Kaspar (2004)	(2.19)
Snow freeze temperature	SN-1	0°C	-1	3	Kaspar (2004)	(2.22)
Snow melt temperature	SN-2	0°C	-3.75	3.75	Güntner et al. (2007b)	(2)
Degree day factor	SN-3	1	0.5	2	Güntner et al. (2007b)	(2)
Temperature gradient	SN-4	0.006°C/m	0.004	0.01	Hunger & Döll (2008)	p. 845

4.2.2 Calibration technique

4.2.2.1 Calibration regions and parameter sensitivity

Due to the limited resolution of GRACE data, the 28 largest and most important river basin worldwide were selected for this study (Fig. 4.1). Except for Volta in western Africa, all basins are larger than 600000 km^2 in size (see Table 4.2). WGHM calibration is carried out separately for each basin.

Güntner et al. (2007b) showed that WGHM parameter sensitivity for water storage variations varied between the river basins. This inter-basin variability is due to different climatic conditions as well as land surface properties and, thus, varying relevance of different storage processes. Consequently, for each region, only the sensitive parameters should be calibrated in order to reduce computational costs and to simplify the interpretation of the calibration results. A sensitivity analysis (SA) against TWSV and river discharge was undertaken (see also Werth et al., 2009a) following the SA approach of Hornberger & Spear (1981). The parameter sensitivity was analysed by a Latin Hyper-cube sampling for 2000 parameter sets for all 28 river basins. Applied parameter ranges are given in Table 4.1. The resulting six to eight most sensitive parameters for TWSV and river discharge (Table 4.3) were used for the regional calibration of each river basin and non-sensitive parameters were set to their original values (Table 4.1, col. 3).

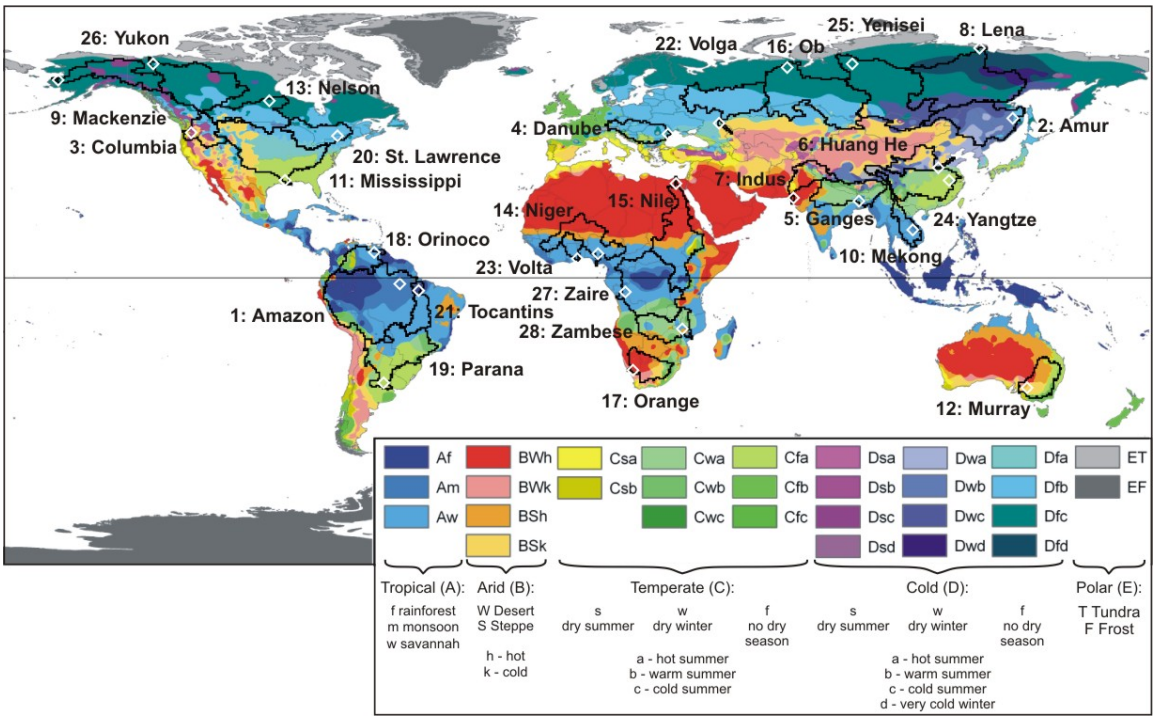


Figure 4.1: The 28 largest and most important river basins worldwide (black polygons) with underlying Köppen-Geiger climate zones (for 1951-2000, by Peel et al., 2007) and gauging stations (white diamonds) of each basin used for calibration of river discharge. See Table 4.2 for station names.

The results of the SA confirmed that the subset of sensitive parameters varied considerably between the river basins. While snow parameters are not sensitive in tropical basins, parameters that control surface water transport appeared as particularly sensitive in basins with important flood plains, such as the Amazon. A broader range of sensitive parameters resulted, for instance, in the Indus river basin which is, on the one hand, dominated by snow storage in the northern mountain area and, on the other hand, high evaporation rates in desert region of the lower Indus. Hence, sensitive parameters belong to these two processes and, e.g., soil water parameters are comparatively less important in the Indus basin. As an example of a river basin that stretches among three different climate regions (cold in the north, subtropical in the southeast and dry in the southwest), important parameters for the Mississippi cover a variety of processes (soil, snow, evaporation, interception and surface water).

4.2.2.2 Multi-objective calibration approach

The multi-objective calibration approach of WGHM was explained in detail by Werth et al. (2009a). Fig. 4.2 and the description below gives an overview. The calibration was done for all 28 river basins in an automated framework for the period 01/2003-12/2007.

Calibration is a widely used optimisation technique in hydrological modelling. In an iterative process, different parameter values are tested for their ability to generate model system states that fit well to observations. The best parameter set provides the lowest simulation error or the highest simulation performance expressed by an objective value. Several functions to measure the objective value are possible, like the normalised root mean square error or the correlation

Table 4.2: Details of the 28 calibrated river basins (col. 1-3) and calibration data (col. 4-6) used. Col. 4: River runoff station, col. 5: Source of discharge data (1: GRDC, 2: USACE, 3: ORE-HYBAM) and period runoff data applied for calibration, col. 7: applied GRACE filter method and belonging filter parameter (I: isotropic filter of Swenson & Wahr (2002) for an a-priori given maximum error of basin average Δ_{max}; II: Swenson & Wahr (2002) computed by by the auto-correlation length G_l and standard deviation σ_0 of an exponential signal model; III: decorrelation method by Kusche (2007) with the power x of the regularisation factor $a = 10^x$ of the signal covariance matrix.)

No.	Basin	Basin area [Mio km^2]	Discharge data source & period	Discharge station	Filter parameter I) a II) Δ_{max},III) σ_0/G_l
B1	Amazon	5.96	3: 2003-2007	Obidos	III: 250/300
B2	Amur	1.87	1: 1975-2004	Bogorodskoye	II: 2.5
B3	Columbia	0.67	1: 1977-2006	Dalles	I: 13
B4	Danube	0.80	1: 1973-2002	Ceatal Izmail	I: 12
B5	Ganges	1.59	1: 1973-2002	Farakka	I: 12
B6	Huang He	0.80	1: 1973-2002	Huayuankou	I: 13
B7	Indus	0.85	1: 1950-1979	Kotri	III: 200/1000
B8	Lena	2.45	1: 1973-2002	Stolb	I: 12
B9	Mackenzie	1.70	2: 2003-2007	Arctic Red River	III: 150/200
B10	Mekong	0.80	1: 1980-1991	Kompong Cham	I: 12
B11	Mississippi	3.24	1: 2003-2007	Tarbert Landing	I: 12
B12	Murray	1.06	1: 1965-1984	Lock 9	III: 150/900
B13	Nelson	1.20	1: 1976-2005	Kelsey	I: 12
B14	Niger	1.80	1: 1977-2006	Lokoja	I: 12
B15	Nile	2.91	1: 1973-1984	El Ekhsase	III: 150/900
B16	Ob	2.70	2: 2003-2007	Salekhard	I: 13
B17	Orange	0.96	1: 1972-2001	Vioolsdrif	III: 20/1000
B18	Orinoco	0.97	1: 1960-1989	Tunente Angostura	II: 4.1
B19	Parana	2.58	1: 1965-1994	Timbues	I: 12
B20	St. Lawrence	1.05	1: 1976-2005	Cornwall	III: 200/1000
B21	Tocantins	0.88	1: 1978-1999	Tucurui	I: 12
B22	Volga	1.39	1: 1973-2002	Volgograd	I: 13
B23	Volta	0.41	1: 1955-1984	Senchi	I: 13
B24	Yangtze	1.93	1: 1975-2004	Datong	I: 12
B25	Yenisei	2.54	2: 2003-2007	Igarka	I: 14
B26	Yukon	0.83	1: 1977-2006	Pilot Stn.	III: 150/100
B27	Congo (Zaire)	3.72	1: 1954-1983	Kinshasa	I: 13
B28	Zambezi	1.39	1: 1976-1979	Matundo-Cais	I: 12

coefficient. Within this study, the Nash-Sutcliffe-efficiency coefficient (NSC, Nash & Sutcliffe, 1970) is applied. NSC is a simulation performance measure that normalises the squared difference of a predicted to an observed time series by the sum of squared deviations of the observations to their mean during the period of interest. It ranges from $-\infty$ to 1 (optimal fit), with a value of 0 indicating a simulated time series that performs as well as a model being equal to the mean of the observable. NSC is applied here because it measures errors in phase, amplitude and mean of a simulated time series at the same time.

Within a multi-objective calibration, more than one observation is applied to evaluate the model simulations, which makes the selection of the best parameter set less trivial. Due to errors in the model structure and the input data (Vrugt et al., 2003a), the approach will no longer provide one single optimal parameter set, but lead to a Pareto set of optimal solutions (Gupta et al., 1998). Each Pareto solution provides a better simulation performance than any other Pareto solutions for at least one of the objectives (but not all objectives). Without additional

Table 4.3: Most sensitive and calibrated parameter for the 28 river basins. See Tab. 4.2 for complete basin names and parameter description.

B1	B2	B3	B4	B5	B6	B7	B8	B9	B10
SL-1	SL-1	SL-1	SL-1	SL-1	SW-1	SW-1	SL-1	SW-1	SL-1
GW-1	SW-1	SW-1	SW-1	GW-1	SW-5	SW-2	SW-1	SW-3	GW-1
SW-1	ER-3	ER-1	SW-3	SW-3	ER-3	SW-5	ER-1	SW-5	SW-1
SW-2	ER-1	ER-3	ER-1	SW-4	IN-1	ER-1	ER-3	ER-3	SW-2
SW-4	IN-1	SN-2	ER-5	ER-1	IN-2	ER-3	ER-5	ER-4	SW-4
IN-1	SN-2	SN-4	SN-2	SN-2	IN-3	ER-5	IN-1	SN-1	IN-1
						SN-2	IN-2	SN-2	IN-2
						SN-4	SN-2	SN-3	IN-3

B11	B12	B13	PB14	B15	B16	B17	B18	B19	B20
SL-1	GW-1	SW-1	SL-1	SL-1	SW-1	GW-1	SL-1	SL-1	SL-1
SW-1	SW-1	SW-5	GW-1	GW-1	SW-2	SW-1	GW-1	GW-1	SW-1
ER-1	SW-5	ER-1	SW-2	SW-2	SW-5	ER-7	SW-2	SW-1	SW-3
ER-5	ER-2	ER-3	SW-4	SW-3	ER-2	IN-1	SW-5	SW-3	ER-4
IN-1	ER-5	ER-4	SW-3	SW-4	ER-3	IN-2	ER-1	SW-4	ER-5
SN-2	ER-6	ER-5	ER-1	ER-1	SN-1	IN-3	IN-2	SW-5	IN-1
	IN-1	SN-2	IN-2	ER-3	SN-2			ER-1	IN-2
	IN-2	SN-3	IN-1	SN-2	SN-3			ER-5	SN-2

B21	B22	B23	B24	B25	B26	B27	B28
SL-1	SL-1	SL-1	SL-1	ER-1	SL-1	SL-1	SL-1
GW-1	GW-1	GW-1	SW-2	ER-3	SW-1	GW-1	ER-6
SW-2	SW-2	SW-1	ER-1	ER-5	SW-4	SW-4	SW-1
SW-3	SW-3	SW-2	ER-3	SN-1	ER-1	SW-5	SW-3
SW-4	ER-1	SW-3	ER-5	SN-2	ER-3	ER-1	ER-1
ER-1	SN-2	ER-1	SN-2	SN-3	SN-2	ER-5	ER-7
ER-4		IN-1					IN-1
IN-2		IN-2					IN-2

information on the observations or a defined priority of simulation accuracy, the Pareto solutions are equal. In this study, river discharge and TWSV were applied for the calibration of WGHM and a balanced improvement of simulation performance for both objectives was intended. Therefore, the solution closest to the optimum of the objective values (here a value of NSC=1 for both objectives) was selected as the best parameter set and used for further analyses.

For parameter variation, ranking and archiving the calibration algorithm ϵ-Non-dominated Sorting Genetic Algorithm-II (ϵ-NSGAII, Kollat & Reed, 2006) was used. The multi-start scheme and the evolutionary strategy of the algorithm (mutation, crossover and selection) enable a global optimisation of the parameter values and are able to solve highly non-linear optimisation problems. The algorithm is one of the most efficient and effective multi-objective optimisation methods used in hydrological modelling (Tang et al., 2006). These features enable a multi-objective calibration for more than one parameter of the non-linear and computational expensive WGHM model system. ϵ-NSGAII operators were set to values proposed by Kollat & Reed (2006) and a population size of N=12, an ϵ-resolution of 0.05 for both objectives and a generation size of 100 (hence, a maximum of 1200 model evaluations) were used.

In contrast to Werth et al. (2009a) who applied significant signal periods within the GRACE data for their calibration, a calibration against full time series of GRACE TWSV was undertaken in the present study (see data Sect. 4.2.3). During the calibration of WGHM, TWSV simulations

were filtered in the same way as GRACE data (see Sect. 4.2.3.2 and Table 4.2) to ensure equal resolution and a consistent comparison of both data sets.

4.2.3 Calibration data

4.2.3.1 Discharge data

River discharge data of the most downstream gauging station of each river basin were used (Table 4.2, col. 4 and Fig. 4.1). Data were obtained from the Arctic Regional Integrated Hydrological Monitoring System for the Pan-Arctic Land Mass (ArcticRIMS, http://rims.unh.edu), the Environmental Research Observatory for geodynamical, hydrological and biogeochemical control of erosion/alteration and material transport in the Amazon (ORE HYBAM, http://www.ore-hybam.org) and the Global Runoff Data Center (GRDC, grdc.bafg.de).

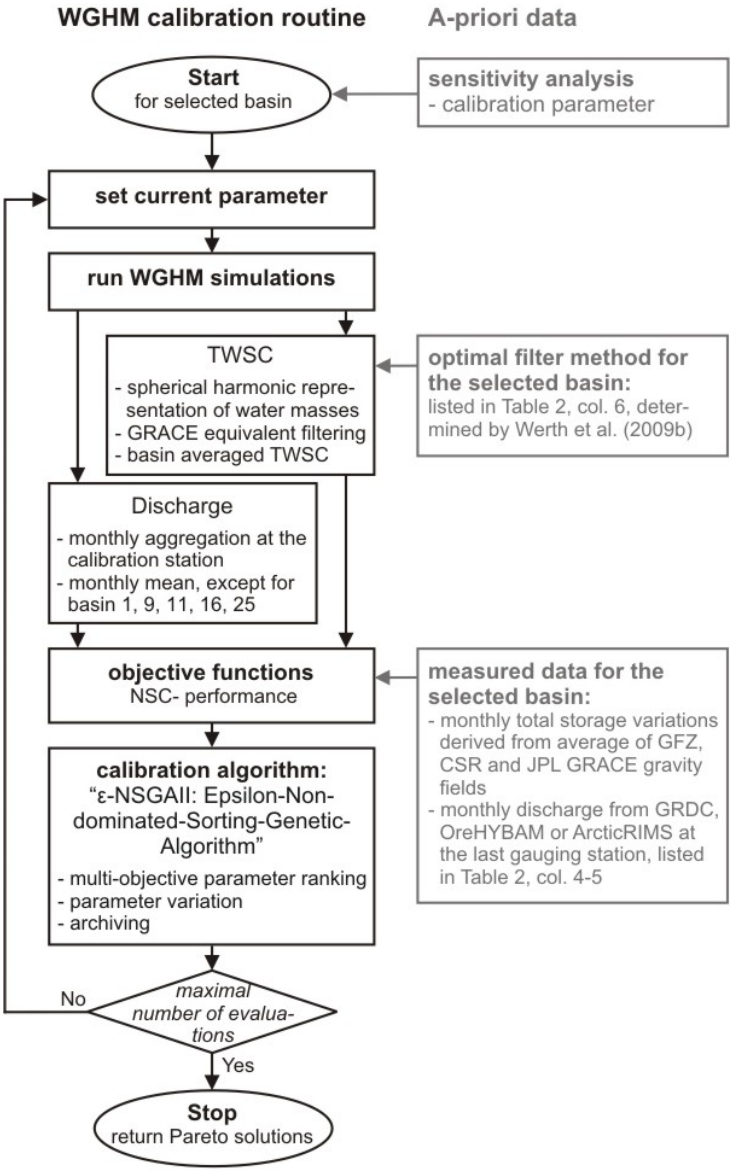


Figure 4.2: Concept scheme of multi-objective WGHM calibration for a specific river basin and with input from Werth et al. (2009b) for applied GRACE filter methods.

For the Amazon, Mississippi, Mackenzie, Ob and Yenisei monthly discharge data were available for the GRACE operation period. For all other basins, up-to-date measurements were not available and mean monthly river discharge (for Jan-Dec) was computed from the most recent period of available data (maximum period of 30 years, see Table 4.2).

Errors of discharge measurements depend on the individual measurement methods and channel cross sections are likely to vary for the individual stations and time periods. Unfortunately, no details are provided from the data centres on the accuracy of discharge measurements. Therefore, the error of discharge measurements was set to a conservative value of 20% for the uncertainty analysis of the calibration results.

4.2.3.2 GRACE data

The greatest challenge in the application of GRACE-based TWSV is marked by the difficulty of separating error from signal as well as separating signal from the region of interest and its neighbouring regions. The spatial resolution of the GRACE data is limited due to the decreasing sensitivity of the satellites to mass variations with smaller geographical extent. Simulation data of atmospheric and oceanic circulation models are applied to de-alias the gravity fields from sub-monthly circulation effects in both systems. Errors in these de-aliasing data and satellite measurement errors increase the noise in spherical harmonic coefficients particularly for higher degrees of the expansion terms, i.e., higher spatial resolution (e.g., Schmidt et al., 2008b). The error budget is also influenced by signal leakage errors from surrounding areas. As a consequence, the application of filter methods is indispensable to reduce noise in the GRACE data. Nevertheless, the magnitude of errors varies between particular regions and months. Therefore, the user has to decide on an adequate filter method as well as for filter parameter settings to balance and minimise GRACE measurement errors and leakage errors. Filtering in turn may change the final signal properties. Werth et al. (2009b) showed that filter induced amplitude damping and phase shifts in time series of basin-averaged TWSV differs between regions because of varying signal characteristics inside and outside of the river basin and basin shape. Hence, the selection of an optimum filter method is a function of the river basins. For the present study, the optimal filter methods (and parameter values) of Werth et al. (2009b) were applied for smoothing of GRACE and hydrological data in 22 river basins. For the remaining six basins (Columbia, Huang He, Mekong, Murray, Orinoco and Volta) optimum filter settings were derived by repeating the method of Werth et al. (2009b) (see Table 4.2 for a summary of applied filter methods).

GRACE derived time series of TWSV from different processing centres show significant differences (as for the Lena basin in Fig. 4.3). These differences are due to different processing strategies, background models or processing software (Schmidt et al., 2008b) and reflect uncertainties in the GRACE data. Consequently, an average of GRACE gravity fields (Level-2 products, most recent version RL04) from three processing centres was used (Flechtner, 2009): the German Research Center for Geosciences (GFZ, until degree 120), the Center for Space Research (CSR, until degree 60) and the Jet Propulsion Laboratory (JPL until degree 120). The three sets of coefficients were averaged from degree 2 to 60 for each month in the period from 02/2003 until 12/2008, excluding 06/2003 and 01/2004 due to missing data from GFZ. For GFZ, regularised solutions for 07-10/2004 and 12/2006 were applied.

GRACE errors were estimated from the error coefficients of the individual data sets published by the processing centres, i.e., correlated errors as provided by GFZ and CSR (Schmidt et al., 2008b; Wahr et al., 2006). As correlated errors were not available for JPL gravity fields, the confidence interval of JPL coefficient errors is increased to 99% by assuming a normal distribution. This results in a multiplication by ≈2.6 of the formal coefficient errors. The final error estimates for the averaged coefficients from the three processing centres amounts to:

$$\epsilon_{knm}^{avefield} = \sqrt{\epsilon_{knm}^{GFZ^2} + \epsilon_{knm}^{CSR^2} + \epsilon_{knm}^{JPL^2}}, k = [0,1], n = [2,60], m = [0,60]. \quad (4.1)$$

Errors in the coefficients are propagated to the basin averages of water storage for each river basin. See Fig. 4.3 for an example of basin-averaged TWSV derived from the three gravity solutions, the average solution and associated errors.

4.2.4 Uncertainty estimation due to observational errors

The uncertainty of the calibration results due to errors in the calibration data is estimated for each river basin by the following procedure: 1) Selection of the calibration run with the Pareto solution closest to the optimum (see an example for the Lena river basin in Fig. 4.4). 2) Propagation of GRACE coefficient errors to basin-averaged estimates of TWSV as well as determination of the 20% discharge error. 3) Generation of 5000 normally distributed samples within the estimated error ranges for the monthly data points of GRACE-based TWSV and monthly river discharge, respectively. The sample size was tested ahead and selected to provide stable statistical results. 4) Estimation of both objective functions (NSC) for each sample against simulated time series of the selected optimal solution, respectively for TWSV and discharge. 5) Determination of the NSC standard deviations for both objectives as the semiaxis for an error ellipse around the selected optimal solution. And 6) Selection of all calibration runs within the error ellipse (see Fig. 4.4 for the Lena basin).

The described approach determines all Pareto solutions around the selected optimum and non-Pareto solutions close to the Pareto frontier, which cannot be evaluated to provide a better fit to the observations than the selected Pareto solution if the error range of the observations is considered. The selected cluster of calibration solutions represents the total uncertainty of the calibration results in view of the observation errors.

4.3 Results and Discussion

4.3.1 Calibration results

Detailed results for Lena basin (Fig. 4.4) show a typical objective function response that was found after calibration of most river basins. A clear trade-off exists between both objective functions for TWSV and mean monthly discharge. The best solutions for the single objectives are located at the end of the Pareto frontier (crossed dots). Best results for a single objective, however, give an undesirable decrease in the accuracy for the other objective. The selected Pareto optimum (large gray dot) provides a balanced improvement between both objectives. The multi-objective calibration approach also decreases equifinality of the parameter sets, since unac-

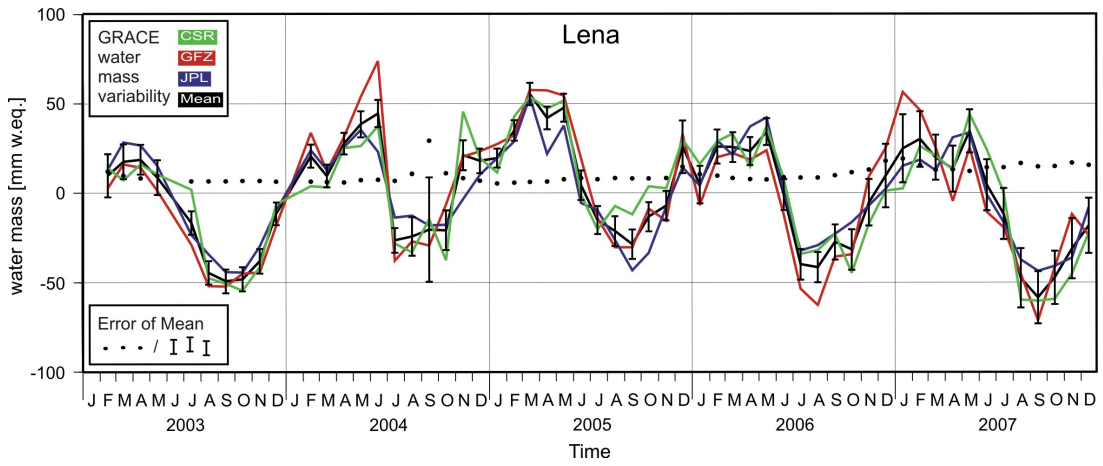


Figure 4.3: Basin-averaged time series of TWS variations from GRACE for the Lena river basin from the processing centres CSR (green), GFZ (red) and JPL (blue) and the averaged field (black) with propagated coefficient errors (black dots and error bars).

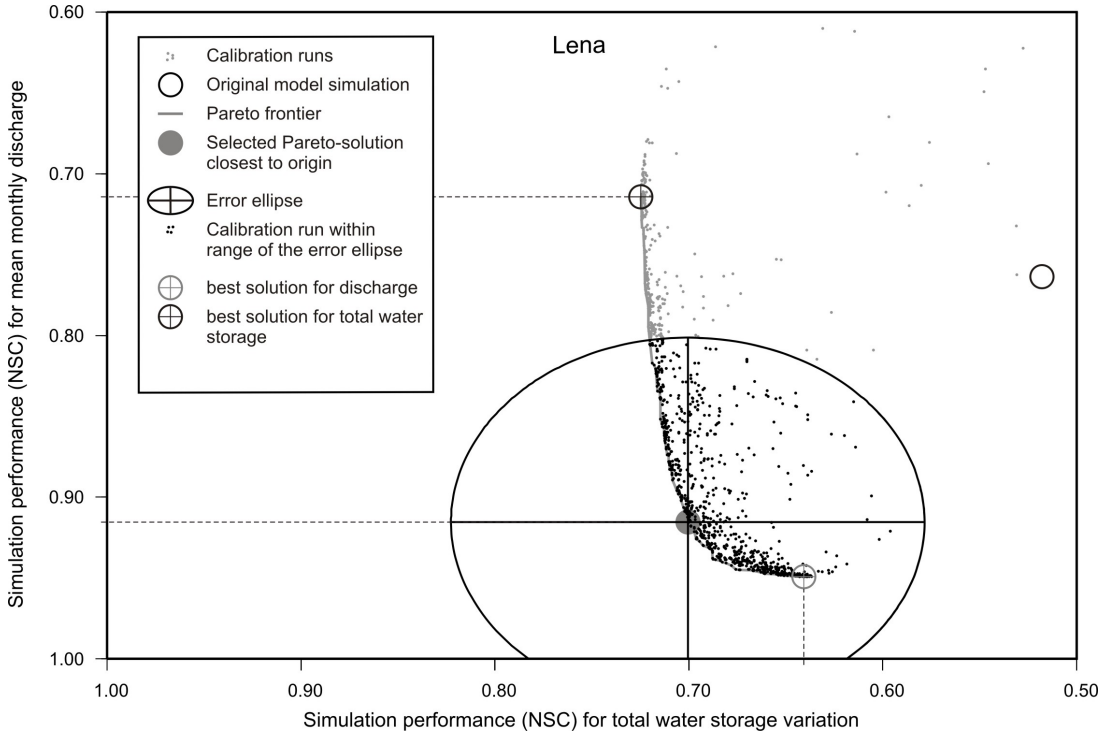


Figure 4.4: Calibration results for the Lena river basin in terms of objective function values. Each point (gray and black) represents one model run. The Pareto optimal solutions form a frontier (gray solid line) towards the optimal model fit (lower left corner). The Pareto solution closest to the optimum (gray large dot) is selected as the optimal solution of the calibration providing a balanced improvement for both objectives and it is used for further studies. Best solutions for each single objective are located at the end of the Pareto frontier (crossed large dots). From errors of the measured calibration data, an uncertainty range for both objectives is indicated by an error ellipse around the selected Pareto solution. The solutions lying in that range (black small dots) show a significant improvement of the calibrated model compared to the original model simulation (plain black circle).

ceptable parameter sets for any of the objectives are excluded by the multi-objective evaluation scheme. A more pronounced equifinality for simulating total water storage variations originates from the character of total water storage data. Since GRACE provides no absolute values but only variations of water masses, the same storage variations may be simulated by different model representations with different absolute amounts of water stored in the river basin. This is not the case for river discharge where both absolute values and variations are given by the observation data. Hence, a smaller number of model realisations provides good objective values for evaluation by discharge than by TWSV. The large ellipse around the selected Pareto optimum represents its uncertainty caused by measurement errors in the calibration data. Variations of parameter values or model output for model realisations within this range are not significant for the assumed observation data errors. Nevertheless, a significant improvement was achieved for both objective values relative to the original model for the example of the Lena basin.

An overview of the calibration results for all river basins is given in terms of relative root mean squared error (RMSE, Fig. 4.5). The relative RMSE was computed from the RMSE of time series of mean monthly discharge (circles) and TWSV (squares) against root mean squared (RMS) values of the respective measurements. Absolute values of signal RMS and model RMSE are presented in Table 4.4. Uncertainty ranges due to observational errors were transferred to RMSE and relative RMSE values and they are indicated by error bars in (Fig. 4.5). A comparison of the results for the calibrated model (black symbols) and the original model (gray symbols) indicate a successful calibration with significant improvements for both objectives for most of the basins. The highest relative improvement of TWSV simulations are provided (and respective RMSE improvements as height of a water column) for the Amazon (ca. 24 mm), Danube (7 mm), Lena (4 mm), Mekong (13 mm), Mississippi (8 mm), Volga (13 mm) and Zambezi (15 mm). Mean monthly discharge simulations improved in particular for the Amazon (with 10 km^3/month decrease in RMSE), Danube (3 km^3/month), Niger (14 km^3/month), Tocantins (10 km^3/month) and Volga (18 km^3/month). For Huang He, Indus and Mekong, improvements were achieved for TWSV simulations only. For the first two of these basins, discharge accuracy is of the same level for the calibrated compared to the original model and the accuracy decreased slightly for Mekong. But the discharge simulations of all three basins are within the measurement error bands. Nelson, Orange, Yukon and Congo (Zaire) exhibit an improvement of discharge simulations while TWSV simulations are of the same performance as for the original model.

With the selected optimum parameter sets, WGHM simulations were repeated between 01/2008-12/2008 beyond the calibration period for validation against GRACE-based TWSV. Table 4.4 shows that the improvement relative to the original model is similar to the calibration period for most of the river basins. For example, RMSE differences to the original model are promising for the Amazon (31 mm), the Lena (10 mm), Mackenzie (10 mm), Mekong (14 mm), St. Lawrence (19 mm) or Zambezi (25 mm). For Murray, Nelson, Orange and Yenisei only a slight improvement for TWSV simulation is achieved in the validation period. A larger RMSE than for the original model was found for Ganges, Huang He, Indus, Orinoco, Nelson, Orange and Congo. This corresponds to the calibration failure of the latter three basins mentioned above.

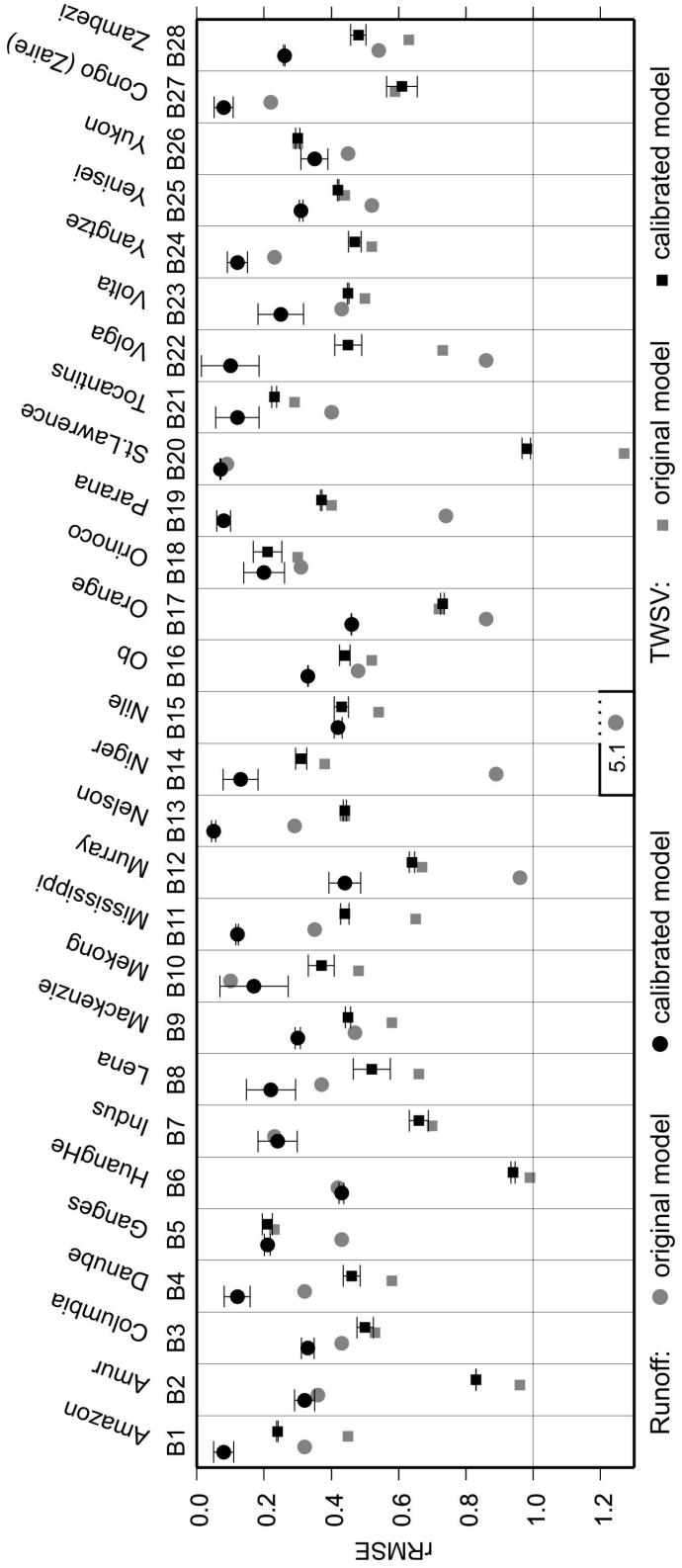


Figure 4.5: Simulation performance for the 28 calibrated river basins in terms of relative root mean squared error (rRMSE) for river discharge (circles) and TWSV (squares) of the original (gray) and the calibrated model version (black). See Table 4.4 for absolute values. Error bars are derived from GRACE and discharge measurement errors as described in Sect. 4.2.4.

4.3.2 Simulation of seasonal TWSV

The effects of model calibration on seasonal amplitudes and phases of TWSV are given in Fig. 4.6. For the most basins, the amplitude was shifted towards the GRACE observations. The strongest improvements for the seasonal amplitude are achieved, e.g., for Amazon, Mackenzie, Niger, Orinoco and Zambezi. For some basins, reduced seasonal phase differences between GRACE and WGHM could be achieved by calibration (e.g., Amazon, Mississippi, Ob and Congo). Only phases could be corrected for Columbia, Danube, Lena, Nelson, Parana and Yenisei. No success for the calibration results again for Huang He in case of the seasonal signal. For Amur and Orange phases differ strongly between GRACE and WGHM, but TWS does not exhibit a distinct seasonal signal in both basins (not shown).

4.3.3 Parameter values and single storage compartments

A detailed analysis of parameter changes (Fig. 4.7) and their effects on single storage compartments (Fig. 4.8-4.9) is provided below for the example of seven river basins of different continents, climatic conditions and calibration success. Storage in lakes, floodplains and wetlands (denoted surface water) is analysed separately from water in the river channel (denoted river storage) in the following sections.

Amazon. The better representation of TWSV simulations for the tropical Amazon after multi-criterial calibration is mainly due to a lower river flow velocity (SW-2) in the calibrated model version as well as a larger runoff coefficient (SW-1). The adjustment of both parameters is stable against calibration uncertainty from observation errors (Fig. 7a). The parameter changes cause a longer-lasting storage of more water in the river network which leads to larger and delayed seasonal amplitudes of TWS in line with GRACE observations (Fig. 4.8a). Also, interannual variations of TWS such as a heavy drought experienced in the Amazon in 2005 (Zeng et al., 2008) are better represented with the calibrated model (Fig. 4.8a). A slightly increased soil water storage is due to the larger rooting depth (SL-1) in the re-calibrated model. But the rooting depth parameter is highly uncertain and it is not significant relative to the original model as can be seen from the wide spread of parameter values for the Pareto solutions in Fig. 4.7a. The larger value of the parameter wetland depth (SW-4) has nearly no effect on the storage variability in lakes and wetlands in spite of the large importance of wetlands and floodplains for water storage in the Amazon (e.g., Papa et al., 2008). Surface water storage is mainly attributed to river channel storage in WGHM (Fig. 4.8a) although the large inundation areas are taken into account as model input. This may indicate structural model errors in representing surface water exchange processes between floodplains and the channel due to the conceptual model formulations and the cell-based simulation of surface water bodies in WGHM.

Mississippi. The Mississippi basin is located in different climate zones ranging from cold to temperate (Fig. 4.1) and therefore it shows a more complex contribution of the individual storage compartments (see Fig. 4.8b) than the Amazon. The most important change in TWSV after model calibration is due to a larger soil storage variability and a longer storage persistence in the early summer, caused by a deeper rooting depth (SL-1). Secondly, a higher snow melt temperature (SN-2) causes an increased snow peak and later melting by one month. The changes for snow and soil storage are supported by a lower radiation proportion absorbed by the surface,

which leads to higher snow accumulation as well as a delayed snow melt. These parameter changes for the Mississippi compared to the original model are reliable considering calibration uncertainty (Fig. 4.7d). An earlier seasonal peak of simulated TWS compared to GRACE data (see Fig. 4.8b) can possibly be attributed to underestimated groundwater storage that are typically characterized by a later seasonal phase compared to near-surface storage. In fact, studies of (Rodell et al., 2006) and Zaitchik et al. (2008) indicate a higher groundwater volume than represented by WGHM. A change for groundwater was prevented by the missing sensitivity of groundwater parameter for WGHM (B11 in Table 4.3), which may be due to the overlap with soil storage variations. The groundwater parameters should therefore be included in further calibration studies.

Lena. For the Lena basin, the seasonality of river water storage exhibits an opposite phase to total storage which is dominated by snow storage variations. This makes a fit of the overall small TWSV amplitude (below 50 mm w.eq. in average) more difficult than for the two previous basins. Model improvements by calibration for this cold, high-latitude basin (Fig. 4.1) mainly are of temporal nature. The phase of TWSV could be corrected (see also Fig. 4.6) based on changes of water accumulation in snow, river and soil (Fig. 4.8c). Due to a larger snow melt temperature (SN-2), snow accumulation lasts nearly one month longer while snow melt finally

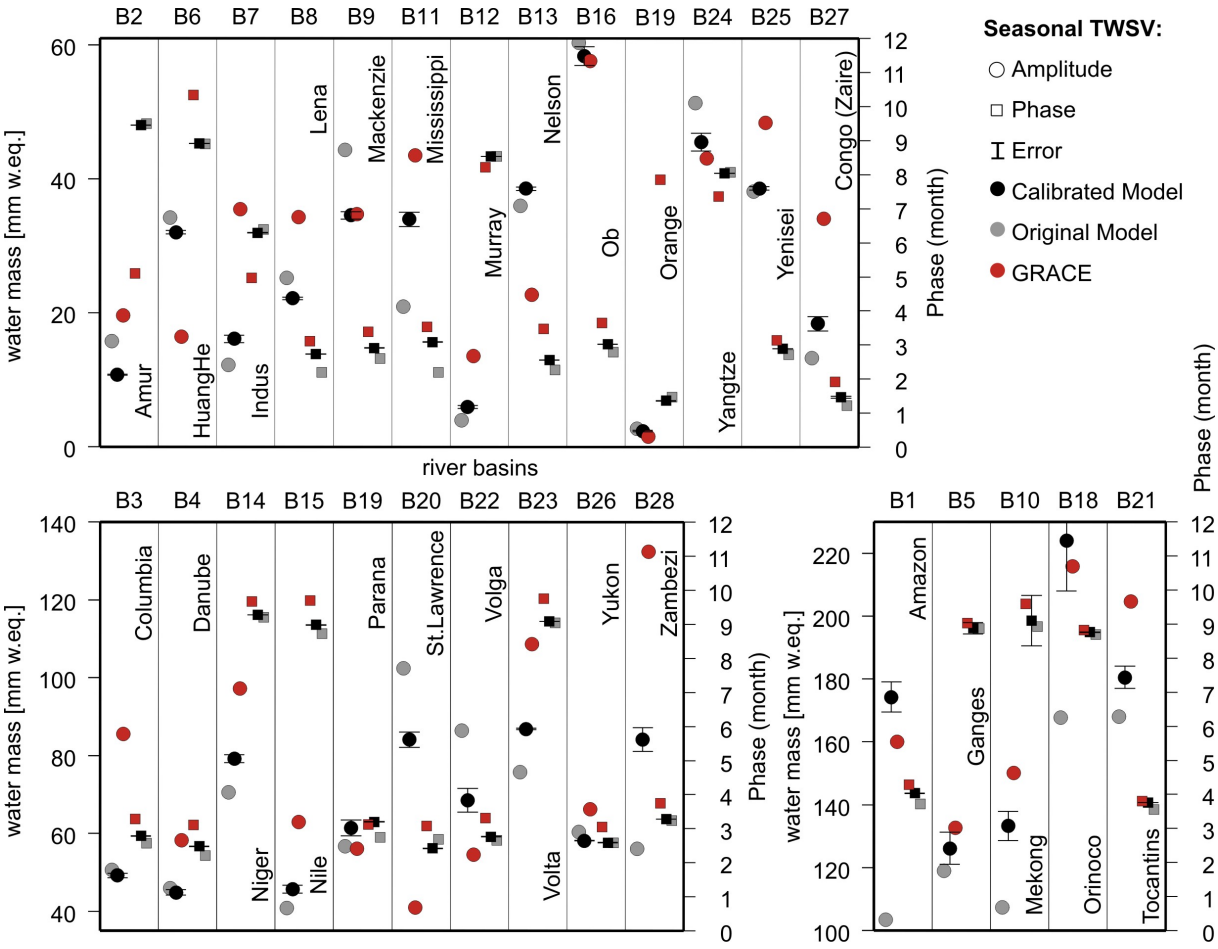


Figure 4.6: Results for seasonal amplitude (circles) and phase (squares) of TWSV for the original (gray) and the calibrated model version (black) compared to GRACE (red). Error bars of TWSV amplitudes are derived from GRACE and discharge measurement errors as described in Sect. 4.2.4.

occurs later but more rapidly in April and May. The larger snow albedo (ER-3) decreases snow sublimation and supports the slightly larger variability of snow storage. In line with the later and faster snow melt in spring, water storage dynamics in river network change accordingly. A larger and later monthly runoff peak also corresponds to the river discharge measurements and is better represented by the calibrated model (see embedded graph in Fig. 4.8c). Changes in soil storage dynamics due to calibration are of minor importance in the Lena basin but, in general, are characterised by slightly larger seasonal variations with a later phase commensurate to the snow dynamics but also to overall lower evapotranspiration rates caused by smaller radiation proportion (ER-1) and PT-coefficient (ER-5) parameters.

Danube. As for Lena, mainly a phase correction of TWSV was achieved by calibration (Fig. 4.6) for the cold and partly temperate (Fig. 4.1) Danube basin. This resulted in a smaller RMSE of TWSV time series (Fig. 4.5). While the seasonal amplitude was not changed, a better fit of extreme events like heat waves or floods as observed by Andersen et al. (2005); Seitz et al. (2008) are visible in the time series for autumn of 2003, 2005 and 2006, as well as for the water mass maxima in 2004 and 2006 (see Fig. 4.8). In the calibrated model, snow is melting faster due to a higher snow melt temperature, hence reducing the snow storage volume. The released water is mainly stored in the soil of which the storage capacity was increased by a larger root depth parameter after calibration. Also river water is reallocated to the soil where it can remain for longer periods during the spring season than in the quickly draining river network. The smaller river discharge in spring is confirmed by observations (not shown here, due to limited space), hence, a smaller RMSE for mean monthly discharge (Fig. 4.5). Groundwater storage variations slightly decreased and delayed in the Danube basin.

Zambezi. Increased storage variations in the hot-temperate and partly dry Zambezi basin (Fig. 4.1) are due to larger soil, groundwater and surface water storage amplitudes (Fig. 4.9a). The largely corrected seasonal variability of TWSV (Fig. 4.6) in the calibrated model originates mainly from less evapotranspiration of surface and soil water as controlled by a smaller PT-coefficient (ER-6) and a smaller maximum potential evapotranspiration (ER-7). As WGHM contains only one soil layer, it may be exhausted too quickly by evapotranspiration in the dry Zambezi region instead of being stored in deeper soil layers. This is supported by the increased groundwater volume, that confirms the high relevance of water exchange with deeper soil zones for Zambezi basin (see also Winsemius et al., 2006a). Surface water volume changes in wetlands increase after calibration and cause longer residence times of water in the Zambezi basin. The importance of this storage mechanism in the Zambezi basin was also found by Winsemius et al. (2006a).

Nelson. The seasonality of snow and groundwater storage exhibits a marked anti-phase in the Nelson basin according to the WGHM simulation results (Fig. 4.9b). This decreases model sensitivity for TWS variations and makes an effective calibration of the individual storage components difficult, since many combinations of different snow and groundwater states can lead to an equally good fit of simulated to GRACE-based TWSV. In addition, the required smoothing of GRACE data has a huge effect on overall water storage dynamics for this basin (Fig. 4.9). Major seasonal signals are smoothed out, but remaining TWSV time series correspond reasonably well between GRACE and WGHM. Comparatively small changes occur by model re-calibration relative to the original model.

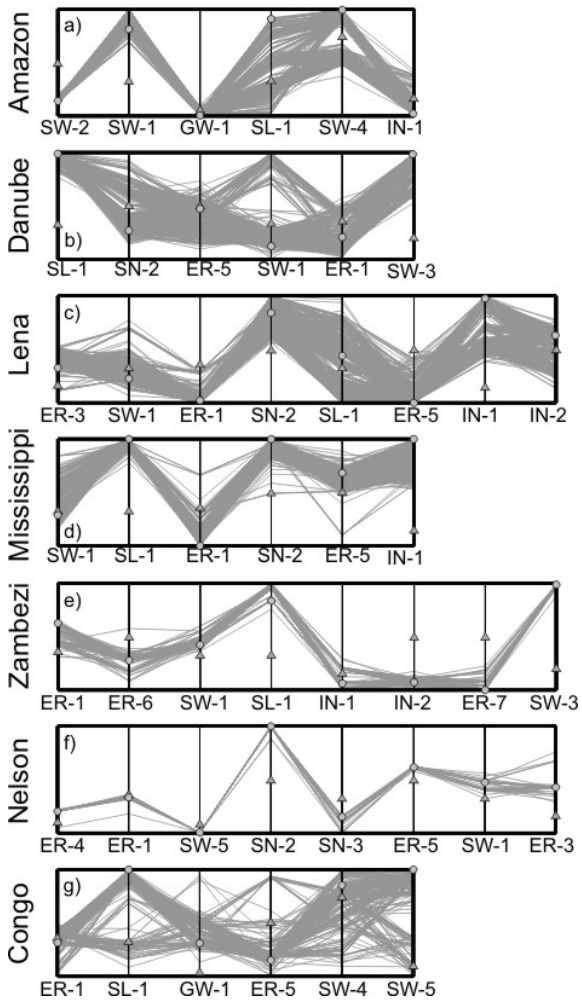


Figure 4.7: Normalised parameter for exemplary river basins. Parameter sets are shown for the selected optimum (black solid line), the original model version (gray dashed line) and all calibration runs within the uncertainty range (gray solid lines) due to observational errors.

Congo. TWSV in the Congo (Zaire) basin is dominated by inter-annual patterns such as a water loss between 2003 and 2005 as described before by Crowley et al. (2006). But as assumed by these authors, the loss is not of secular nature and the storage is filled up again during 2006 and 2007 (Fig. 4.9c). Though the calibrated WGHM exhibits an improved simulation for seasonal amplitude and phase of the Congo basin (Fig. 4.6), the simulated inter-annual variability of basin-average TWS is still different from GRACE, e.g. a too large negative anomaly in 2005. Also, RMSE values did not improved after calibration (Table 4.4). The inter-annual variations in TWS mainly derive from soil and groundwater storage (Fig. 4.9c). For the calibrated model, a larger seasonal variability in soil storage causes a slightly delayed phase of storage variability. This delay appears to be compensated by a negative phase shift in groundwater. As a result, the faster outflow of the groundwater (due to a larger outflow coefficient GW-1) causes a smaller groundwater volume and decreases the inter-annual variation of groundwater storage in the calibrated model.

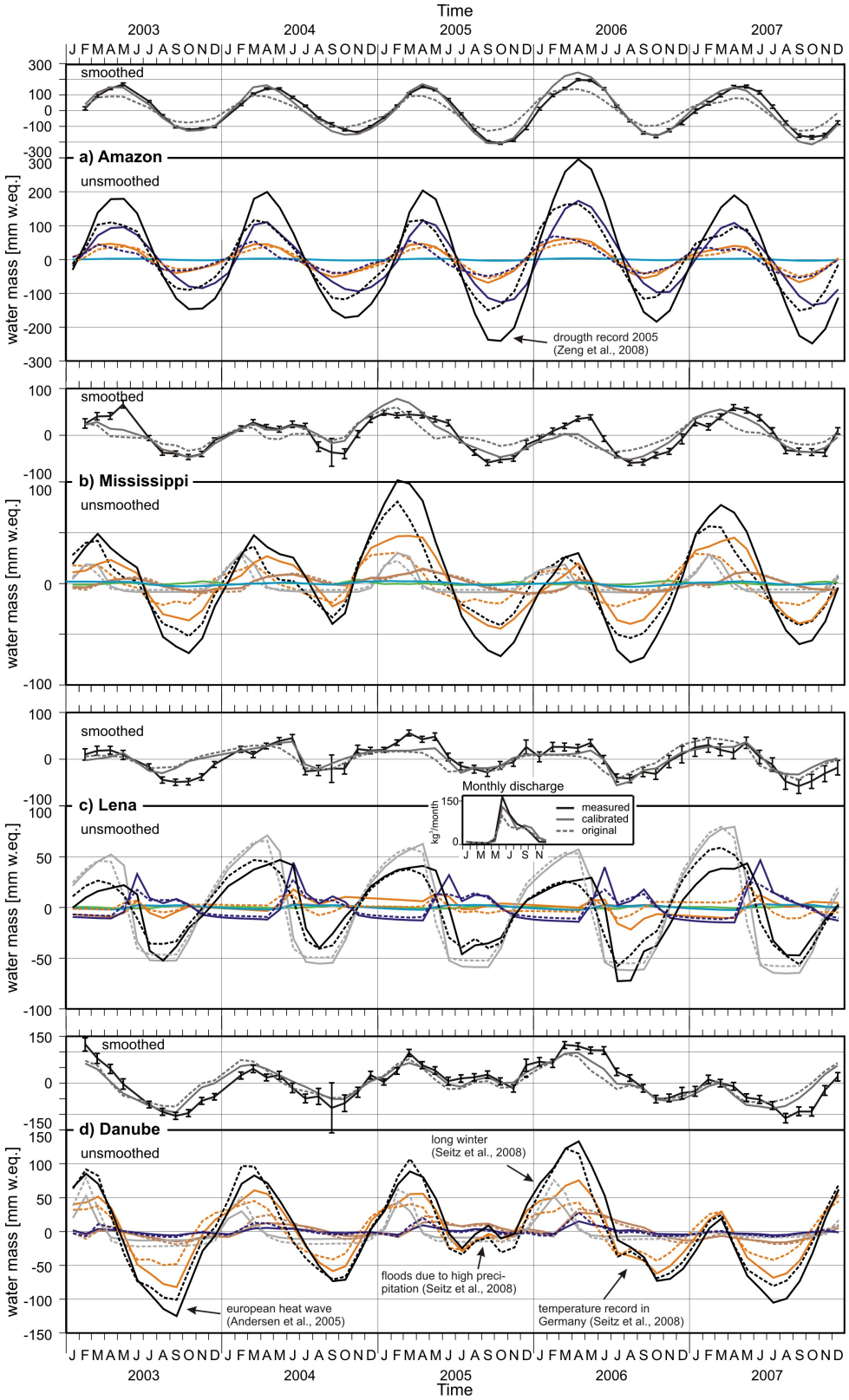


Figure 4.8: Basin-averaged time series of single storage compartments from the calibrated and the original model version (unsmoothed, below) as well as smoothed total storage from both model versions and GRACE (smoothed, above) for a) the Amazon, b) the Mississippi, c) the Lena and d) the Danube basin. See Fig. 4.8 for legend.

Three of the four basins (Nelson, Orange, Congo) with an unsuccessful calibration for TWS exhibit strong inter-annual variations (see Fig. 4.9b,c for Nelson and Congo). The inter-annual variations are visible in GRACE as well, but the short period of five years used here may impede the effective calibration of inter-annual changes in total storage variability and its components. Furthermore, for Congo, Nelson and Orange a large trade-off occurs for the Pareto solutions between simulation performance of river discharge simulation and TWS (not shown). Therefore, calibration difficulties within these basins may also be due to the use of mean monthly discharge values, which neglect inter-annual variations during the calibration period. As a further drawback for Congo, available discharge data are from the period 1954-1983 for this basin.

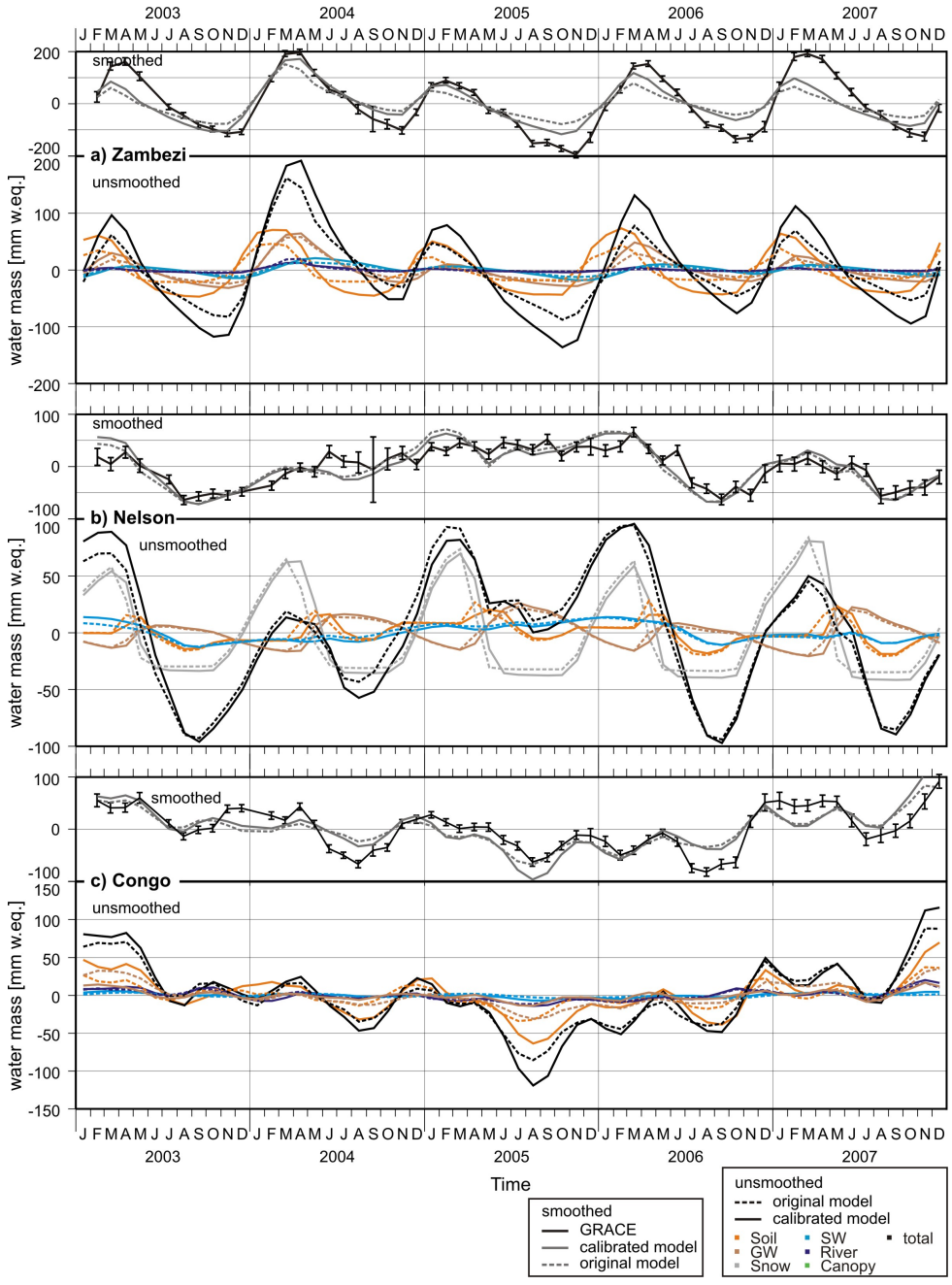


Figure 4.9: Same as Fig. 4.8 but for a) the Zambezi, b) the Nelson and c) the Congo basin.

Table 4.4: Root mean squared signal of observed river discharge (σ_{Dis}^{meas}, col. 2) and standard deviation of GRACE TWSV (σ_{TWSV}^{meas}, col. 5) compared to the root mean squared error (RMSE) of the calibrated (ϵ^{cal}, col. 3 and 6) and the original (ϵ^{org}, col. 4 and 7) model against respective observation data for all 28 river basins. Col. 8 provides differences of RMSE values of TWSV from the calibrated and the original model for the validation period (01/2008-12/2008). Here, negative values indicate an improved simulation compared of the calibrated compared to the original model.

Basin No.	σ_{Dis}^{meas} [km^3/mth]	ϵ_{Dis}^{cal} [km^3/mth]	ϵ_{Dis}^{org} [km^3/mth]	σ_{TWSV}^{meas} [mm]	ϵ_{TWSV}^{cal} [mm]	ϵ_{TWSV}^{org} [mm]	$\Delta\epsilon_{TWSV}^{2008}(cal-org)$ [mm]
B1	471	39	149	118	29	53	-31
B2	31	10	11	30	25	29	-2
B3	13	4	6	65	33	35	-3
B4	17	2	6	61	28	36	-7
B5	44	9	19	103	21	24	2
B6	4	2	2	26	25	26	0.4
B7	11	3	3	40	26	28	2
B8	64	14	24	31	16	20	-10
B9	29	9	14	34	15	20	-10
B10	34	6	3	113	42	54	-14
B11	42	5	15	41	18	26	-12
B12	0.8	0.4	0.8	24	15	16	-0.1
B13	5.3	0.3	2	46	20	20	-1
B14	18	2	16	76	23	29	-5
B15	3	1	17	50	22	27	-6
B16	43	15	21	46	20	23	-5
B17	0.7	0.3	0.6	12	9	9	-0.1
B18	98	20	31	168	36	50	18
B19	45	4	33	49	17	20	-7
B20	20	1	2	39	38	50	-19
B21	36	4	14	157	36	46	-10
B22	23	2	20	48	22	35	-7
B23	3	0.9	1	84	38	42	-7
B24	80	10	18	36	17	19	-3
B25	73	23	38	37	16	16	-0.4
B26	21	8	10	65	20	20	-5
B27	112	9	25	41	25	24	-4
B28	9	2	5	107	52	67	-25

The water mass variation of the Orange basin, which also exhibit inter-annual variations (not shown), are smaller than 12 mm of a water column (see Table 4.4 and Fig. 4.6) and for some months below GRACE data accuracy. While inter-annual variations are not relevant for the Yukon basin, similar to Nelson, a clear anti-phase between snow and groundwater storage as well as soil storage causes a small model sensitivity for TWS variations.

4.3.4 Global analysis

A global analysis of simulated TWSV for the calibrated model (see spatial distribution in Fig. 4.10 and variability of basin averages in Table 4.5) shows that its variability increased for the most river basins compared to the original model. On the global average (last row of Table 4.5), TWS variability increased by 7 mm w.eq., which is mainly due to a larger variations of soil, river and surface water storage. Most variability is gained within the tropical and temperate regions, like the Amazon (total 60 mm for the basin average), Congo (9 mm), Niger (14 mm), Mekong

Table 4.5: Variability of unfiltered and basin-averaged continental TWSV simulations from the calibrated WGHM version for total storage and single compartments: $\sigma^{cal}(storage)$ (TS: total storage, SL: soil, GW: groundwater, SN: snow, R: river, SW: surface water, C: canopy). Every other line provides deviations of storage variability to the original model: $\Delta\sigma_{storage} = \sigma^{cal}(storage) - \sigma^{org}(storage)$.

Basin	$\sigma^{cal}(TS)$	$\Delta\sigma_{TS}$	$\sigma^{cal}(SL)$	$\Delta\sigma_{SL}$	$\sigma^{cal}(GW)$	$\Delta\sigma_{GW}$	$\sigma^{cal}(SN)$	$\Delta\sigma_{SN}$	$\sigma^{cal}(R)$	$\Delta\sigma_{R}$	$\sigma^{cal}(SW)$	$\Delta\sigma_{SW}$	$\sigma^{cal}(C)$	$\Delta\sigma_{C}$
B1	150	+60	37	+8	25.7	+0.4	0.1	+0.0	82.9	+49.6	2.1	+0.8	0.0	+0.0
B2	20	+3	10	+3	7.8	-0.4	21.0	+1.7	5.0	+0.2	1.4	+0.2	0.5	+0.3
B3	55	-1	12	-7	4.4	-0.6	40.2	+6.1	3.5	+0.1	1.7	-0.2	0.2	+0.0
B4	64	+4	43	13	10.6	-2.2	16.0	-9.4	4.1	-3.0	1.4	+0.9	0.4	+0.0
B5	90	+7	17	-8	21.0	-5.6	1.8	+0.3	20.1	+3.6	10.6	+4.2	0.1	+0.1
B6	18	-2	9	+1	5.9	-1.4	0.3	+0.0	2.7	-0.6	0.3	-0.2	0.1	+0.1
B7	28	+4	7	+1	4.8	+0.7	24.6	+3.4	6.4	+0.2	1.0	+0.4	0.0	+0.0
B8	32	+0	8	+2	1.7	+0.2	47.9	+2.8	15.2	+4.7	1.8	+0.1	0.8	+0.5
B9	44	-8	7	-1	7.8	+0.5	50.7	+0.1	6.9	+3.5	1.5	-1.3	0.3	+0.0
B10	129	+36	54	+22	33.5	+0.6	0.3	+0.0	37.1	+11.2	3.1	+0.7	0.1	+0.1
B11	48	+14	27	+11	6.3	-0.4	12.1	+2.6	3.4	-0.6	1.9	+0.0	1.1	+0.9
B12	17	+3	9	+1	2.1	-0.6	0.0	+0.0	0.6	+0.4	2.4	+0.7	0.0	-0.1
B13	57	+2	10	-1	12.2	+1.2	39.8	+2.4	0.5	-0.2	7.7	+0.5	0.2	+0.0
B14	58	+14	26	+12	14.8	-4.6	0.0	+0.0	11.1	+4.2	3.2	-0.2	2.6	+2.6
B15	35	+2	21	+8	1.6	-5.3	0.0	+0.0	9.7	+1.2	3.5	+0.3	0.0	+0.0
B16	61	+0	14	-2	14.6	+2.2	67.5	+9.3	5.3	+0.4	1.5	-0.7	0.3	+0.0
B17	6	-1	3	-1	2.3	-0.3	0.0	+0.0	0.5	-0.3	0.4	-0.1	0.3	+0.3
B18	169	+51	57	+18	35.8	+1.7	0.0	+0.0	54.6	+26.0	7.6	+1.3	0.1	+0.0
B19	59	+1	22	+6	19.3	-0.8	0.0	+0.0	5.1	-10.3	5.6	+2.4	0.1	+0.0
B20	78	-20	15	-5	9.8	-4.9	43.4	-21.4	1.2	-1.4	9.2	-0.3	0.6	+0.3
B21	145	+18	39	+17	43.0	+0.8	0.0	+0.0	16.0	-9.4	13.1	+2.3	0.3	+0.3
B22	68	-16	33	+8	11.9	-2.6	56.0	-15.1	11.5	-5.4	1.7	+0.6	0.3	+0.0
B23	80	+26	49	+28	19.2	+1.9	0.0	+0.0	2.5	-0.4	4.6	-1.8	1.0	+1.0
B24	30	-5	4	-2	13.0	+1.2	1.2	+0.5	12.3	-3.4	0.7	+0.0	0.3	+0.0
B25	41	+2	9	+1	6.8	+1.3	56.0	+7.5	9.1	+3.9	1.9	+0.5	0.3	+0.0
B26	52	-3	7	+0	3.4	+0.3	57.6	+0.0	8.8	+1.1	2.9	+0.5	0.2	+0.0
B27	47	+9	26	+12	6.3	-8.8	0.0	+0.0	7.3	+1.2	2.8	+0.6	0.0	+0.0
B28	80	+26	41	+20	23.3	+5.3	0.0	+0.0	3.6	-0.7	9.1	+2.6	0.0	+0.0
global	66	+7	24	+3	15.4	-0.6	20.5	+0.2	9.4	+3.3	13.2	+2.7	0.2	+0.1

(35 mm) as well as for the Mississippi (14 mm). A spatial redistribution between sub-regions for some of these basins is visible in Fig. 4.10, e.g., Ganges and Parana. A smaller total water budget appears only for basins in cold regions like Mackenzie, St. Lawrence, Volga or Yangtze (Table 4.5). Some further cold regions like Lena or Ob exhibit an unchanged water budget. This comparison shows that TWS variability in the original WGHM was mainly underestimated in tropical and temperate regions but overestimated in cold regions, similar to the seasonal components (Fig. 4.6).

For the individual basins and storages, largest differences to the original model occur within soil storage, mainly for tropical and temperate regions like Mekong, Mississippi, Orinoco, Volta or Zambezi, which is visible by area distributed TWSV differences to the original model in the lower Fig. 4.10 and reflected in the basin-averages (Table 4.5). Soil has the highest seasonal capacity to store water and contributes most to the gravity signal discovered by GRACE that is usually dominated by seasonal features. Linear structures in the spatial distribution TWSV differences to the original model are mainly due to changes in river storage, being the second most contributor

to changes for the basin-averages (Table 4.5). Very large increase of river water volumes occur in rainy tropical regions of Amazon, Mekong and Orinoco, where a slow discharge in the river network causes a longer maintenance of river water in the basin (see analysis for Amazon in Sect. 4.3.3 above). In contrast, a decrease of river water volume is visible for temperate and dry regions. Snow storage increases for regions in cold climate zones e.g., Columbia, Ob, Yenisei). But it decreases for cold climates with a warm summer (St. Lawrence, Volga, Danube). In these transition zones, less snow precipitation may be due to global climate warming, that is relevant for the calibration period but not incorporated in the calibration of the original model.

Simulated groundwater storage changes decreased on the global scale. A large decrease of groundwater variations occurred for regions with a distinct dry season (Ganges, Niger, Nile) and for some cold regions (St. Lawrence, Volga). Groundwater seasonality is usually delayed compared to soil and surface storage, because groundwater recharge and runoff are temporarily filtered by soil transfer processes. As seen from seasonal phase shift between GRACE and WGHM, water often drains too quickly from river basins compared to GRACE even for the calibrated model version. This may be explained by a too small groundwater recharge and volume in WGHM (e.g. Zambezi or Mississippi). Also the sensitivity of the model to changes in groundwater storage may be superimposed by the soil storage with a different seasonal phase. Therefore, future calibrations against GRACE data should include groundwater timing and volume parameters for each river basin.

4.4 Conclusions

This study demonstrates that a multi-objective calibration with TWS variations from the GRACE satellite mission and river discharge enables a world-wide improved simulation of changes in the continental water cycle and its compartments. The presented strategy for improving simulations of continental water storage includes the following key points: 1) Inclusion of the most important storage compartments (soil, snow, canopy, rivers, surface water and groundwater) in the simulation of continental water storage for a comparison with satellite observations. 2) Multi-objective calibration by absolute values of river discharge and relative values of TWS variations. 3) Basin-specific calibration of dominant processes, hence of the most sensitive model parameters. 4) Assuring consistency of observables and model state variables (equal spatial scale) by identical smoothing of GRACE and model data, as well as the application of most optimal filter method for each river basin. 5) Consideration of measurements errors in an uncertainty analysis of the calibration results. 6) Analysis of calibration results to reveal model structural errors and to broaden the knowledge about hydrological processes on large-scales.

The multi-objective calibration of WGHM led to higher simulation accuracy for TWS variations and river discharge for most of the 28 calibrated river basins. Seasonal amplitudes and phases of the water budget for most river basins were improved. A global comparison showed that TWS variability was mostly increased for tropical regions. The highest proportion of the increase occurred for soil storage. An analysis of single storage compartments for seven river basins from different continents and diverse climatic regions revealed reasonable changes within single storage compartments of the calibrated model that contributed to a better representation of TWS variability. Herein, the deviation of the calibrated parameter sets to the original model

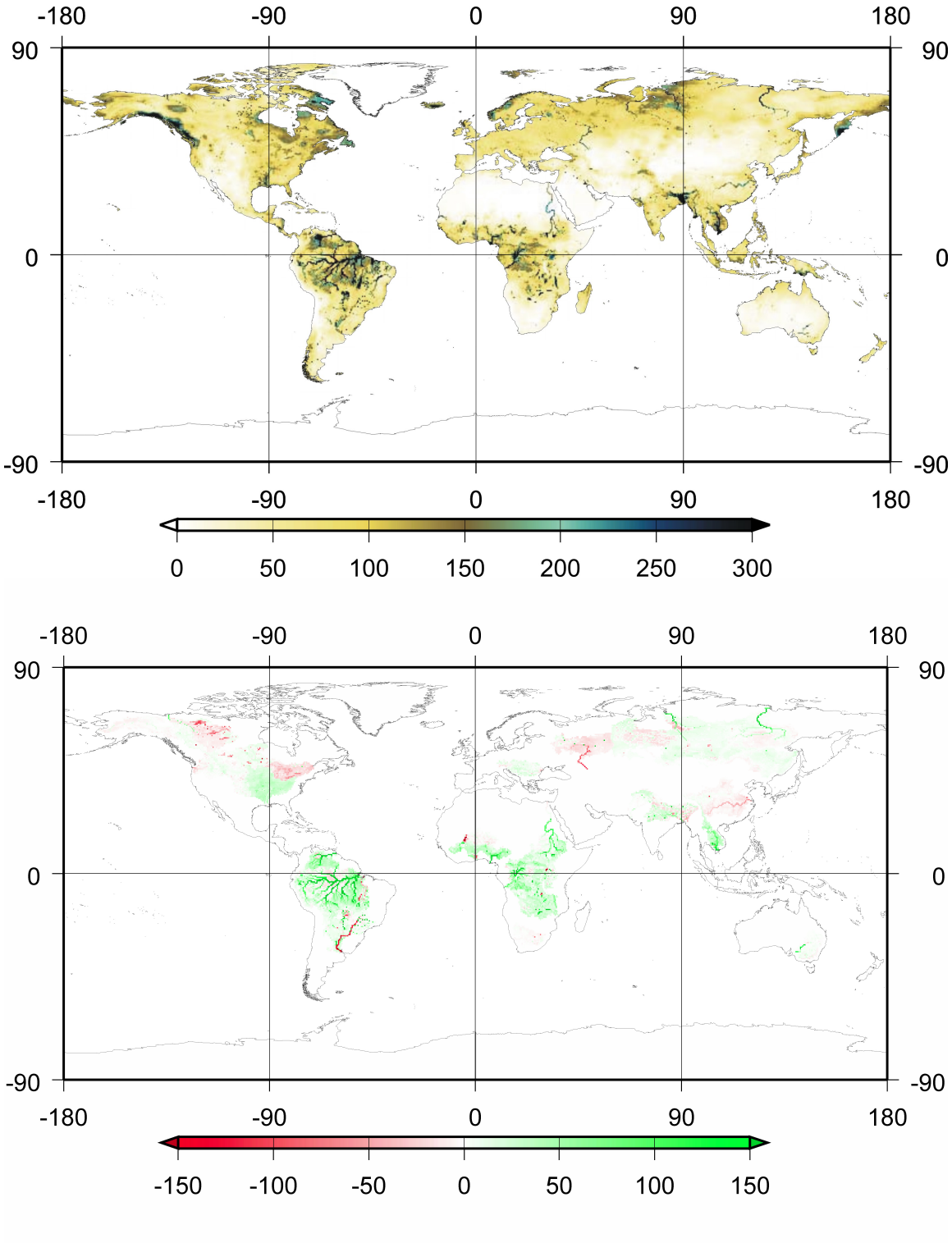


Figure 4.10: Global distribution of total storage variability of the calibrated WGHM (above) and its deviations to the original model version (below). Negative values below indicate decreased and positive values increased variability. Units are in mm of water column.

version and their uncertainty due to measurement errors provide an insight into the control and reliability of the individual process simulations.

For some basins, possible model structural errors are uncovered by the calibration, e.g., too small wetland volumes in the Amazon and the Mississippi basin. For some basins, errors or limits of the calibration data restrict the calibration success. An update of global river discharge data sets to the GRACE mission period is an urgent need for further progress. As another strategy for the calibration of basins with strong inter-annual variations and scarce discharge data availability, smaller weights could be given to mean monthly discharge data in the calibration process. However, the model representation of TWS variations inheres more parameter equifinality than river discharge due to the lack of absolute values and the integrative nature and limited spatial resolution of GRACE TWS variations. Consequently, GRACE data alone are not adequate for calibrating water storage state variables in large-scale hydrological models.

Calibration difficulties are also due to the complexity of interaction between single storage components and to the inability to separate these storages with the integrative TWSV data. Many different single storage combinations can lead to similar variations in the total water budget of a river basin. The decrease of model sensitivity for TWS or its components is catalysed if clear anti-phases occur between storage variations in individual compartments. Since groundwater seasonality deviates mostly from the other storages, it plays an important role in the timing of TWSV in a river basin. A parameter sensitivity analysis for such basins should be undertaken carefully and for future studies, an increased attention should be given to groundwater storage in the calibration process.

The improvement of large-scale hydrological models and the validation of GRACE water mass estimates is an iterative process. Model structure errors may complicate the calibration of WGHM with GRACE TWSV. But also limited spatial resolution or regional varying accuracy (e.g., Winsemius et al., 2006b) as well as different smoothing effects between GRACE and modelled data may affect the calibration performance. Therefore, GRACE uncertainties are still an important object of research. Furthermore, due to the general data scarcity of hydrological observations at the global scale, newly developed observation systems like GRACE in turn depend on global model estimations for validation and error reduction. This will complicate the independence of model re-calibrations and again it limits the application of GRACE for hydrology (and vice versa) in the sense of an iterative learning process.

As consequence to the named difficulties, it is desirable to include further hydrological observations into the calibration and validation process. Only satellite data are in the run for large-scale hydrological modelling and the rise of satellite observation systems with global coverage are promising. Groundwater observations are not available with global coverage and this storage is not directly accessible by space techniques. But satellite observations of snow storage (e.g., by MODIS, Parajka & Blöschl, 2008), surface water (Papa et al., 2008) or soil moisture from the future satellite missions such as SMOS and SMAP are applicable for tuning or validation of large scale-hydrological models with more than two objectives.

Due to the large diversity of processes in different regions of manifold climatic conditions, global hydrological modelling is a challenging ambition. The present study expands experiences on representing hydrological processes on the global scale with a particular emphasis on water storage dynamics. The continuation of similar studies is motivated by the steadily improved

accuracy of GRACE solutions and the future prospect of a GRACE-follow on mission. Longer time series of gravity data will in particular allow focusing on hydrological extremes, inter-annual variations and secular trends in both observations and modelling capabilities.

Acknowledgements

We acknowledge J. Alcamo and P. Döll for providing the WGHM model code. The German Ministry of Education and Research (BMBF) supported these investigations within the geoscientific R+D programme GEOTECHNOLOGIEN „Erfassung des Systems Erde aus dem Weltraum". The authors also want to thank Ross Woods productive discussions on the calibration analysis methods.

5 Final Summary and Conclusions

The main aim of this work was to improve global simulations of total water storage variations (TWSV) from the hydrological model WGHM by applying monthly TWSV estimates from GRACE gravity fields and therewith, to learn about the reasons for simulation discrepancies to GRACE, in particular about model errors. To achieve this aim, a multi-objective calibration approach for the integration of GRACE into WGHM was developed and applied. Several questions arisen for such a novel data combination and not all of them may be answered, but they shall motivate our attempts for the multi-data integration of remotely sensed hydrological observations into large-scale hydrological modelling in order to improve our understanding of the Earth's water cycle as a whole.

5.1 Main strategies and results

The undertaken studies are a combination of a-priori model and data analyses, model re-calibration as well as a-posteriori data and uncertainty analyses. Herein, the study period was reclined to the GRACE mission period from 2003 until 2007 (Chapter 3) and 2008 (Chapter 4), respective to the state of the studies. Nevertheless, applied discharge data partly go back to 1950, depending on the most recent available data sets (see Table 4.2). In the following sections, the components of the developed approach and the main results of the re-calibration are summarised.

5.1.1 Iterative concept

A challenge in the application of GRACE data is the separation of signal from one region to another as well as signal from noise (Schmidt et al., 2008b). The limited spatial resolution of GRACE is reflected by increasing errors in GRACE coefficients of higher spatial resolution. Therefore, the coefficients have to be smoothed before transferring them to mass variations. Herein, a critical point depicts the selection of an adequate filter method and a respective parameter value. The discussions of proper filter strategies (just some main examples are Jekeli, 1981; Swenson & Wahr, 2002; Han et al., 2005; Sasgen et al., 2006; Chen et al., 2006b; Swenson & Wahr, 2006; Seo et al., 2006; Kusche, 2007; Schrama et al., 2007; Davis et al., 2008; Klees et al., 2008; Kusche et al., 2009) and of restoring strategies (e.g., Swenson & Wahr, 2007; Chen et al., 2007a; Klees et al., 2007) are likely to prolong until there is no break through regarding the GRACE gravity field accuracy.

A validation of GRACE filter methods and respective parameter demands independent global data sets of mass variations. For continental hydrology, only global models provide such data and they are widely used to validate filter methods (e.g., Chen et al., 2005; Swenson & Wahr, 2006; Chen et al., 2007b; Kusche et al., 2009; Klees et al., 2008). On the other hand, differences

between global hydrological models are as large as the signal they feature, which is due to different modelling strategies and input data (see Chapter 2). In return, this underlines the need of GRACE-based hydrological mass variations to improve large-scale hydrological modelling concepts (Schmidt et al., 2006; Niu & Yang, 2006; Ngo-Duc et al., 2007; Syed et al., 2008; Schmidt et al., 2008c; Zaitchik et al., 2008). Consequently, at this stage and until further global observations become available, which cover the main components of the global water cycle, the improvement of TWSV simulations and the validation of GRACE data analysis tools is an iterative procedure. This constraint is reflected by the present work, where a validation of GRACE filter methods was undertaken with hydrological model simulations.

5.1.2 GRACE filter evaluation

The strategy and the strength of a GRACE filter, that is applied to suppress coefficients, strongly influences the spatial resolution and the total error in GRACE data (Chapter 2). The latter is an accumulation of the satellite error and the filter induced leakage (e.g., Klees et al., 2007). In this study, the respective methods and their parameters were evaluated (see Chapter 2) by a comparison of six of the most common non-decorrelation and decorrelation filter methods (Jekeli, 1981; Swenson & Wahr, 2002; Seo et al., 2006; Swenson & Wahr, 2006; Kusche, 2007) and for different values of the filter parameters. This comparison was undertaken by a measure of correspondence between basin averaged TWSV derived from equally filtered GRACE data and model simulations (Table 2.5). The latter are computed from an average of three global hydrological models GLDAS, LaD and WGHM, which are widely applied for comparisons with GRACE. It is expected, that the total hydrological simulation error is reduced when averaging three independent models, since all models rely on different strategies and input data (see Chapter 2).

A comparison of the filter methods shows that smoothing strategies and strengths but also basin properties and signal characteristics inside and outside a river basin influence the filter's performance and efficiency. Different filter and parameter settings introduce different amplitude and phase distortion of the TWSV time series of different river basins (Table 2.4). Two of the filter (method II and III) more likely produce amplitude damping, while two other methods (I and IV) can lead to strong phase shifts in the time series of TWSV (Fig. 2.6-2.8). If smoothing induced phase shifts are of relevance, amplitude restoring factors are not applicable alone to recover the signal amplitude. Optimal filter methods and parameter values for individual basins were determined to be used in hydrological applications, i.e. WGHM model calibration. These optimised methods and filter parameter provide reduced differences in amplitude and phase as well as a better total error budget compared to not optimised filter methods (see Table 2.4 and 2.6). It is shown that the integration of a-priori information into the filter design like a signal model or orbit-configurations is benefiting and leads to a better error budged, than e.g., the standard Gaussian filter method (Fig. 2.10). One of the decorrelation methods (Kusche, 2007), which uses orbit-configurations, was exposed to be optimal for most of the river basins (Table 2.5). Exception occurs only for Amur, Orange and St. Lawrence, where agreement between model simulation and GRACE data is generally low. Additionally, the results for Nelson and Indus are not reliable due to low general correspondence for all filter methods. The filter evaluation results were supported by estimation and graphical comparison of the total error budget for all filter methods and admitted parameter values. Finally, a sensitivity test towards amplitude

errors in hydrological simulations confirmed the reliability of the filter evaluation. The filter evaluation results are an important and valuable input for the application of GRACE data in WGHM re-calibration studies.

5.1.3 Data consistency

GRACE data and hydrological model simulations have been filtered equally. Because GRACE is not sensitive to small scale mass variations, no clear separation between signal of different regions is possible with any analysis strategy or without alternative data sets. Since such data sets are not available, the resolution of hydrological simulations has to be down-scaled. A consistency to the resolution of GRACE data may be achieved by filtering the simulated data. The remarkable differences between filtered and unfiltered time series of basin-averaged TWSV are visible for some example river basins in Fig. 4.8 and 4.9. They have a significant effect on the detection of signal amplitudes, e.g. in particular of extreme events. Some studies neglect the necessity to equally filter both data sets (e.g., Zaitchik et al., 2008; Rodell et al., 2004a). But as it is reasonable only to compare data with equal temporal resolution, similar spatial resolution should be ensured as well before the comparison of modelled and GRACE based TWSV (as done by Chen et al., 2009; Klees et al., 2008; Schmidt et al., 2008c).

Another type of data consistency refers to the inclusion of relevant hydrological process. WGHM is an appropriate global hydrological model to be calibrated by GRACE data, since it simulates the main components of the continental water cycle, including soil, canopy, snow, surface water storage (see Chapter 1). Especially surface water (rivers, lakes and wetlands) is often neglected in global hydrological models (Syed et al., 2008; Ngo-Duc et al., 2007; Milly & Shmakin, 2002a, see also Chapter 2), though a representation of all relevant components of the continental water cycle is important to preserve consistency with the integrative GRACE observations. Furthermore, a closed water balance is guaranteed if all important storages are simulated. The complete closure of the water cycle is a basic demand to be able to increase our understanding of the water cycle as a whole.

5.1.4 Parameter sensitivity

The global hydrology model WGHM is originally calibrated against one river runoff parameter, because it bases on river discharge measurements only. The application of GRACE data for the re-calibration of WGHM enables the tuning of parameter from other processes besides river runoff, since GRACE data include variations of all water storage components. This makes the regulation of the surface water balance possible, which is not affected by the runoff parameter (Döll et al., 2003). Furthermore, the empirical or literature-based values of overall 36 model parameters may be verified. But since a calibration of all parameter is technical not reasonable, the six to eight most sensitive parameter towards river discharge and TWSV were selected by a sensitivity analysis with a Latin Hypercube sampling for 2000 model evaluations (see Chapter 3 and 4). Here, the number of six to eight is a compromise between the optimisation of all parameters with high sensitivity towards model output and the reduction of parameter equifinality as well as computational effort. The latter would rise fast with the number of optimisation parameters. The results of the sensitivity analysis are reasonable from a hydrological point of view,

since they reflect the importance of specific processes within a river basin, e.g., they allocate a high importance of surface and soil water to the Amazon region or a large importance of snow parameter to the Mackenzie region (see Table 4.3). The elimination of insensitive parameters from the calibration list not only reduces computation time but also simplifies the interpretation of the re-calibration results for WGHM.

5.1.5 Calibration technique

For the monitoring of the continental water cycle on large scales GRACE depicts a very valuable tool providing access to a previously immeasurable variable, the total water mass variation of a specific region. Nevertheless, due to the integrative and relative nature of GRACE data it is necessary to use further measurements during the calibration which ensure a correct water level within the simulated water cycle. Furthermore, a new calibration of WGHM was ought to predicate on the station-based accuracy of the original model version evaluated by river discharge measurements. Therefore, a key strategy for the re-calibration of WGHM is the multi-objective decision making against two sets of measurements, being GRACE TWSV and river discharge measurement.

The application of multi-objective methods in hydrological model calibration started with (Gupta et al., 1998), as an answer to parameter equifinality (Beven & Binley, 1992). The multi-objective approach provides more reliable results for the system optimisation, because any single observations is contaminated by errors and it would reflect only one aspect of the system behaviour (Vrugt et al., 2003a; Wagener et al., 2003; Fenicia et al., 2007; Yapo et al., 1998; Gupta et al., 1998; Duan, 2003). Furthermore, hydrological model optimisation is often an over-parameterised optimisation problem. So, the integration of additional measurements increases the predictability of the system (Beven, 2001) and a multi-objectively calibrated parameter set will lead to model simulations that are consistent to different system states. The second answer in hydrological sciences to equifinality was the consideration of parameter uncertainty. In this work, parameter uncertainty is respected by a a-posteriori uncertainty analysis for GRACE and discharge measurement errors (Chapter 4). The application of respective optimisation methods (e.g., GLUE, BaRE or SCEM-UA by Beven & Binley, 1992; Thiemann et al., 2001; Vrugt et al., 2003b, respectively), recently combined with multi-objective methods (e.g., MOSCEM-UA by Vrugt et al., 2003a) are not applicable for WGHM re-calibration at the moment, because they demand a very large number of model evaluations (minimum between 10.000 and 100.000) to enable statistical analyses.

The gain from a multi-objective calibration strategy is best shown for the example of the Congo river basin (see Fig. 3.1), where a large trade-off between both calibration objectives occurs. A single-objective calibration for Congo would lead to good estimates for either TWSV or river discharge but at the same time result in very inaccurate simulations for the second variable. As mentioned above, this problem is of high relevance, when using GRACE data. A large number of model versions may reflect the relative variations of TWS but different absolute water levels in the river basin. Consequently, a larger equifinality towards GRACE TWSV (as for Amazon and Mississippi in Fig. 3.1 and for Lena in Fig. 4.4) is exhibited. Here, a multi-objective calibration which additionally constrains absolute river discharge measurements decreases parameter equifinality and restricts good TWSV solutions to model versions consistent

to both components of the continental water cycle. For the multi-objective parameter ranking and sampling the efficient genetic algorithm ϵ-NSGAII (Tang et al., 2006; Kollat & Reed, 2006; Deb et al., 2000) was linked to the WGHM software.

To accommodate the limited spatial resolution of GRACE data, the re-calibration of WGHM was done basin-wise for basin averaged TWSV and for river discharge at the last discharge station of the basin towards the ocean. The sub-basin heterogeneity of river water flow of the original model version was kept by the calibration of a runoff coefficient factor for the whole river basin (see Chapter 2).

The algorithm was first applied for the Amazon, the Mississippi and the Congo (Zaire) river basins (Chapter 3) with a limited number of 400 model evaluations. The river basins were selected for for varying signal intensity and to test the calibration method on different geographical locations. An aim for the test runs was to minimise the noise in GRACE data at best possible, since a standardly used Gaussian filter had to be applied at that time. Therefore, the calibration was done against significant TWSV periods in the GRACE signal, that are verifiable by hydrological model simulations, though they may exhibit different size and timing (Schmidt et al., 2008c). From the calibration an improved simulation of TWSV signal periods as well as its complete signal could be achieved for all three river basins, parallel to an increased simulation accuracy for river discharge. This successful calibration encouraged the application of the developed approach for further river basins world wide.

5.1.6 Global model calibration

Due to the limited spatial resolution of GRACE data, their application is limited to large river basins. Though, the size of GRACE observation error is site dependent (Winsemius et al., 2006b; Horwath & Dietrich, 2009), satellite errors may mask out the signal of mass variations at very small basins. Therefore, the final calibration (Chapter 4) was done for the 28 largest and most important river basins worldwide (see the basin list in Table 4.2).

For each basin most sensitive parameter (Table 4.3) are calibrated with the developed multi-objective approach after 1200 model runs (Chapter 3). Time series of GRACE TWSV were smoothed by a-priori determined optimal filter methods (Chapter 2). Herein, most updated monthly GRACE RL04 solutions were taken from the data centres of GFZ, CSR and JPL. Differences between the solutions originate from different processing strategies of the data centres (Klees et al., 2008). Averages of these three different fields were used as a best guess of the monthly GRACE mission's gravity fields. Calibrated errors of the three data sets where propagated to an error estimation of the average fields (Eq. 4.1). River discharge data were taken from different data centers at the last available discharge station of a river basin (see Table 4.2). Due to missing error estimations from the runoff data centers, the error of discharge measurements was conservatively set to 20%. This agrees quite well with a very resent and first general estimation of discharge measurements by Di Baldassarre & Montanari (2009), who estimated the error to about 25% for the Po River. Consequently, the reliability of the calibration towards GRACE and river discharge errors was analysed in an a-posteriori uncertainty analysis (see Sect. 4.2.4, Fig. 4.5 and 4.6).

Strongest accuracy increase in hydrological simulations was achieved for the Amazon basin, where RMSE reduced about 24 mm for TWSV and 110 km^3/month for river discharge. Accuracy

increase is best for the Amazon, because the strong signal of water mass variations routing through the river basin more significantly contrasts to measurement errors. Though, smaller than for the Amazon a significant improved simulation performance regarding both objectives could be achieved for other basins as well (see Fig. 4.5). Only for Huang He, Indus and Mekong improvements are limited to TWSV, and for Nelson, Orange, Yukon and Congo to river discharge respectively. Seasonal amplitude and phase of TWSV improved similarly, except for basins lacking a distinct seasonal signal in the GRACE data like Amur and Orange (see Fig. 4.6). For these two basins, GRACE data accuracy is not sufficient.

A validation of the calibration results was done for a temporal period 01/2008-12/2008 following after the calibration time frame. Regarding river basins, the validation results correlate with calibration success.

5.1.7 Simulations of TWSV and its components

After the re-calibration, simulations of TWSV and its components are analysed in detail for seven river basins (Sect. 4.3.3) and for the global scale (Sect. 4.3.4). In total, water storage variations show a global increased variation of 7 mm, which mainly results from a larger variation in tropical and temperate regions (e.g., Amazon, Mekong, Niger, Figure 4.10, Table 4.5). In contrast, TWSV was overestimated by the original model for most of the cold regions (e.g., for Mackenzie or St. Lawrence).

Analysing differences between single storage compartments of the calibrated and the original model version on the global scale, largest changes occur within soil, river and surface water storage (see last row of Table 4.5). On the basin scale, it becomes clear that these changes mainly derive from changes in tropical regions (whole Table 4.5). But also increasing snow storage is exhibited for cold regions. Decrease in water storage variations of single WGHM compartments occurs for river water in temperate and dry regions as well as for snow in cold regions with a warm winter. In these transition zones, global climate warming during the GRACE period may have an effect, because the original model calibration belongs to earlier periods (Hunger & Döll, 2008). The variation in groundwater was decreasing globally on the basin scale. This loss in variability mainly derives from regions with a distinct dry region or some cold river basins.

It is also interesting to have a look on TWSV of the re-calibrated single storage compartments of WGHM. In average, soil depicts the highest variable capacity to store water on the continents and therefore, it inheres the largest part of changes in the GRACE gravity field (last row of Table 4.5). Variations in snow storage may reach very high values for cold regions and average to the second largest storage variations on the global scale. Together, rivers and water of lakes and wetlands (surface water) are of similar size as snow storage. Otherwise, storage variations of lakes and wetlands alone are the fourth largest contributor to the variable gravity signal, together with variations in the groundwater. Compared to other storages and to signal accuracy canopy storage variations are negligible on river basin and on global scale. But because vegetation has an important influence on evaporation, the simulation of canopy storage is necessary for a complete represenation of the continental water cycle.

Time series of simulated basin-averages of the single storage components, exhibit a complex interaction (some examples are shown in Fig. 4.8 and 4.9). Single storage variations are not directly evaluated during calibration, but only through their proportion in TWSV. Instead, a

brief loot at single storage variations and parameter changes was undertaken to investigate the reasonability of the calibration results (Sect. 4.3.3). Phase shifts between single storages is visible in the time series, that are strongest for groundwater or snow in cold regions. Furthermore, a too low surface water variability was found for Amazon and Mississippi, which may be caused by surface water exchange of river and surface water due to the conceptual WGHM formulation on an aggregated cell-basis. It is likely, that also the absence of reservoir information in the applied WGHM version introduces an error to the timing of time series of surface water storage.

5.1.8 Difficulties and disappointments

The limitation of the WGHM calibration with GRACE became apparent by the analysis of time series of single storage compartments (Fig. 4.8 and 4.9). Total water storage depicts a sum of all components and it is not possible to separate GRACE time series without further information. Other studies separated the components from GRACE mass variations using modelled hydrological data (Swenson & Wahr, 2009; Rodell et al., 2006; Ramillien et al., 2006; Frappart et al., 2006; Niu et al., 2007a). But such a separation is illegitimate, when GRACE data are applied for hydrological modelling itself, as in this study. Hence, during model calibration a combination of the strongly seasonal single storage variations could sum up to similar TWSV time series for many model versions, i.e. many combinations of different storage compartments. Consequently, the calibration is rather insensitive to the simulation of single storages. Furthermore, it is known from independent model simulations, that groundwater storage exhibits the strongest phase shifts to other components of TWSV and therefore, it plays an important role in the timing of TWSV (Güntner et al., 2007a). For WGHM, larger groundwater variations would lead to a better timing of seasonal TWSV. But Fig. 2.2 showed that NSC is more sensitive to amplitude differences than to phase differences of the seasonal TWSV. This may have reduced the sensitivity of the groundwater factor parameter, which is not among the six to eight most sensitive parameters for any basin. On the other hand, the groundwater baseflow coefficient regulates the groundwater phase and its sensitivity might have been overestimated. These limitations could be responsible for the reduced groundwater variation after the calibration, which is not reasonable regarding previous studies, e.g. for the Mississippi basin (due to Rodell et al., 2006; Zaitchik et al., 2008).

Further difficulties occurred for river basins with strong inter-annual variations. These variations are not present in mean monthly river discharge, applied for the multi-objective calibration. This points out the high need of more up to date river discharge measurements on the global scale.

It might be expected to apply real alternative data sets to validate the WGHM re-calibration. But due to data scarcity of hydrological observations on global scales and due to the difficulties in the application of remotely sensed surface soil moisture into global hydrological models (e.g., Basara & Crawford, 2000; Wilson et al., 2003; Choi et al., 2007) such data were not available for the present study. Therefore, a validation of the calibration results was limited to a second time period instead of alternative measurements.

A further limitation for the re-calibration respects to the long evaluation time of WGHM, reducing the possible number of model evaluations (see Table 1.2). To achieve a secure global optima, usually many thousands of model evaluations are calculated. For WGHM, a compromise between model evaluation time and security of the calibration results had to be taken.

The selected number of evaluations follows findings of Tang et al. (2006) regarding convergence efficiency of ϵ-NSGAII.

5.1.9 Open questions

The present study concentrates on TWSV variations close to the seasonal time scale. An evaluation of simulated trends in hydrological variables would be another important insight into the model performance for evaluating modelling strategies. Steffen et al. (2008) excluded secular trends in GRACE data from geophysical signals like glacial isostatic rebound and compared them to hydrological models. He argued that hydrological models do not well represent trends in continental hydrology and that they have to be improved before their secular components are applicable for further studies. This issue is still an open question in hydrological research. But the knowledge gained by the present study serves as a guideline for respective model analyses.

Climate warming has a significant effect on water storages of ice (e.g., glaciers) and permafrost (e.g., Haeberli & Beniston, 1998; Lawrence & Slater, 2005). These storages are not included in WGHM (see Sect. 1.2). On the one hand, increased melting may introduce errors in simulated discharge variations by WGHM compared to GRACE or to models that include these storage components. On the other hand, effects of climate change may explain some differences of water storage variations between the re-calibrated and the original model version, especially for transition climate zones (see Chapter 4). As the influence of climate change may increase together with the extension of GRACE time series, it would be interesting to quantify this effect specifically for WGHM to evaluate its representation of the hydrological cycle or to separate effects of climate change.

Another open question that appears at this point and in the face of global climate change, is the transferability of the re-calibrated WGHM parameter to the past and to the future. Due to the IPCC report (Bates et al., 2008), the global climate change causes very different hydrological system states. But at what stage of the system change are we and how long will it maintain? Climate modelers will help to answer this questions. But changes in the distribution of climate zones, as we know them today (e.g., Peel et al., 2007), may limit the calibrated model states to certain periods. Dynamic behaviour of certain parameters or e.g., land cover types may have to be introduced into global hydrological modelling, especially for long-term studies.

The calibration with GRACE was be done for large river basins only. For small basins, resolution of GRACE is insufficient. It has still to be tested, whether a calibration of a group of several river basins clustered to a region of sufficient size would work out. To answer this open question would exceed the time limit for this work. But it would help to realise a complete integration of GRACE into global models as well as the application of GRACE for continents like Europe, where rivers are relatively small but very important for e.g., hazard assessment or water demand.

5.2 Integration of GRACE data into WGHM

The main research questions of this study, that were posed in the introducing Chapter 1 shall be answered by the following sections.

5.2.1 How can GRACE be integrated into global hydrological modelling?

The most important characteristics of GRACE data integration into the global hydrological model WGHM can be summarised to the following.

I. Separation of GRACE signal from error by a basin specific filter evaluation towards an optimal error budged in GRACE-based time series of TWSV: Smoothing of GRACE gravity fields is indispensable and the available filter methods have different effects on different regions. Filter specific parameter especially of non-decorrelation methods should be optimised to reduce the satellite error in GRACE data but achieve the best possible spatial resolution, hence, an optimal separation of signal and noise in GRACE data. Not optimised filter methods may cause insufficient GRACE data accuracy.

II. A-priori model analysis: Process analysis and parameter calibration based on GRACE data integrations may be simplified by a-priori considerations of regional characteristics. An a-priori model sensitivity analysis may eliminate insensitive parameter against the calibration variables for specific river basins. Therewith it not only reduces computation time but also simplifies interpretation of the results.

III. Consistency of compared data sets: a) First, this respects to the simulation of all important water storage components when they are compared to total water mass variations from GRACE. Outside polar regions, these depict storages of soil, snow, rivers, canopy, surface and groundwater. b) Secondly, consistency of data sets respects to an equal spatial resolution among the compared data sets. This demands equal filtering of GRACE and modelled data.

IV. Calibration for large river basins: The limited spatial resolution of GRACE restricts their applicability to regions with a diameter of a few hundred kilometer. Therefore, only simulations for large river basins or for a region clustering several basins may be calibrated with the satellite data. The minimum size of the region of interest depends on its shape, location and the real regional distribution of GRACE errors.

V. Multi-criterial calibration: The continental water cycle includes a complex composition of many water storages and transport processes. A single type of observation represents only one sight or aspect of that cycle and may bias the calibration of a hydrological model. The integration of multiple data sets increases the stability of the results and enables a more consistent representation of the continental water cycle.

VI. Uncertainty estimation of the calibration results: Errors in calibration data are propagating towards errors in parameter and errors in model output. This uncertainty has to be taken into account and to be quantified for an estimation of the reliability of the results.

Two strategies were applied concerning the issue II.b within this study. The global calibration (Chapter 4) based on optimised filter methods and respective parameter (Chapter 2). For the test calibration, significant signal periods were extracted (Schmidt et al., 2008c) to separate signal from error (Chapter 4). Though, the second method demands a larger computational effort,

amplitudes and phases of these significant periods showed to be applicable for an adjustment of hydrological simulations, as well.

The listed strategies provide a guideline for similar studies, that compass the integration of GRACE data into large-scale hydrological models in order to solve the water balance of a river basin. Components of the method may be of relevance for studies on further Earth subsystems as well, e.g. periodic signals from oceans or secular processes of the solid earth.

5.2.2 Were simulations of TWSV improved?

The re-calibration of WGHM for three basins in test modus (Chapter 3), as well as globally for 28 of the largest river basins (Chapter 4) proofed the benefit of GRACE satellite data for large-scale hydrological modelling. Improved time series of TWSV were achieved worldwide and for most of the river basins parallel to a better representation of mean monthly river discharge. Most effects on single storage compartments appeared to be reasonable and provided an input for further improvements of the model structure, hence, of TWSV simulations. The re-calibration was only limited for river basins that exhibit large errors or scarcity in the calibration data. These results confident the further usage of GRACE data for large-scale hydrological modelling.

5.2.3 What can we learn from the results for global hydrological model development?

The sensitivity analysis of all available model parameter and the re-calibration showed two important aspects, that have to be regarded for global hydrological simulations. Except for the calibration parameter, all parameters are of the same values worldwide in the original model version, though some are global factors multiplied to land properties (like the rooting depth) or other characteristics of the individual land-cells. In fact, it is the aim of global hydrological modelling to find an algorithm that is worldwide valid. But different processes are of highest importance within different river basins, hence, different parameter need to be calibrated. This was shown by the sensitivity analysis. Furthermore, in different regions other parameter values improve simulations of water cycle components, which was shown by the re-calibration results. For example, the river velocity was reduced in the Amazon basin, leading to a slower outflow from the watershed and better timing of TWSV time series. For many other basins, river velocity was either modelled sufficiently accurate by the original model version or not that significant as other processes (e.g., the Lena basin).

In general, the most important parameter changes were significant compared to the original model version and respective calibration data errors. This enables the application of the calibrated parameters within further studies.

From the re-calibration, further consequences follow for subsurface water and groundwater storages. On global average, groundwater variations are the fourth largest proportion of TWSV, that includes more than ten percent of mass transports within the water cycle. If river water is included, subsurface water amounts to the second largest proportion of TWSV from WGHM. Both storages are of large importance for a complete representation of the water cycle and therefore should be included in any water balance study.

Due to the conceptional approach of WGHM, the distribution of water between rivers, lakes, wetlands and reservoirs may be erroneous (as it was shown for the Amazon and the Mississippi riverss). A revision of the subsurface algorithms may be necessary, if a separation of the individual subsurface components is of interest. Furthermore, the model sensitivity of WGHM towards TWSV appeared to be critical for groundwater factor parameters, which may be due to inaccuracies for WGHM simulations of this storage as well.

5.3 Outlook for future research

5.3.1 Consequences from the experiments

The correct modelling of single storages has to be evaluated by further measurements beside GRACE data or river discharge measurements. Such validations are necessary for a further accuracy improvement of global hydrological simulations. Herein, the general data scarcity and limited data accuracy on the global scale is a significant drawback, which has to be overcome to enable the complete simulation, understanding and the closing of the terrestrial water cycle with sufficient accuracy (Sheffield et al., 2009).

Additional data from satellite and ground observations are desirable to fill the data gap. Another chance is the development of new techniques to assimilate available data into hydrological models for which it has been difficult until now. Some regional examples are given by Papa et al. (2008) for surface water extends, Parajka & Blöschl (2008) for snow coverage or river discharge speed by Smith & Pavelsky (2008). These techniques together with others should be made applicable for global studies.

5.3.2 Suggestions

To integrate GRACE data into the simulations of smaller than the 28 selected watersheds, a calibration of basin clusters would be possible. These basins should belong to similar climate regions or exhibit similarities in important characteristics of the water cycle. This is necessary to enable the calibration of clustered parameter and to minimise the additional number of calibration parameter.

The sensitivity of groundwater factors against TWSV could be increased by an inclusion of seasonal phases in the performance measure, in order to give them a larger weight in the evaluation process. Herein, further objective values or an aggregation-based combination of several aspects of the time series into one objective function are applicable.

In future, global hydrological model adjustments should be undertaken more independently from GRACE data, to be able to continue with the iterative concept of GRACE data evaluation and improved modelling of global TWSV. For example, GRACE data integration into models can provide an estimation which specific processes are mis-modelled (as done for groundwater and surface water). Afterwards, improved simulation of TWSV could base on improved model structure instead of GRACE-based parameter calibration.

5.3.3 Prospects

Interim updates of the WGHM algorithm undertaken parallel to this study enable more accurate analysis for future studies. These updates include a new reservoir algorithm (Döll et al., 2009), which is important for a correct timing of surface water storage changes due to the anthropogenic regulation of water outflows from lakes. Furthermore a technically improved model algorithm significantly reduces the model evaluation time (Verzano, 2009). This enables more comprehensive calibration runs in the future, which is a basis for a greater reliability of model and parameter uncertainty analyses.

GRACE time series are enlarging month by month. The mission is performing very well, though its expected lifetime is already exceeded. Also, a GRACE follow-on missions is likely to come (Ries & Bettadpur, 2008). Enlarged time series of TWSV increase the probability of extreme events available in the data (as already available for Amazon and Australia from Chen et al., 2009; Leblanc et al., 2009, respectively). This further challenges the evaluation of large-scale hydrological models. The determination of interannual variations and trends in TWSV as well as increased variability of TWSV due to climate change will be more reliable from longer time series of GRACE satellite measurements.

GRACE errors have been under strong investigation (e.g., Han et al., 2004; Horwath & Dietrich, 2006; Winsemius et al., 2006b; Kusche, 2007; Klees et al., 2007). Significant error decreases were already achieved during the last years (e.g., de-aliasing, de-correlation, smoothing Schmidt et al., 2008b). This process is likely to continue by ongoing investigations of instrumental and processing errors (Ries & Bettadpur, 2008). Smaller errors in the GRACE data would increase the accuracy of GRACE data integrations as well as the applicability of these data on smaller scales and for smaller signal components.

As on the global scale, the same link of gravity changes and mass variations is given for terrestrial gravity measurements and local hydrological variations, (Neumeyer et al., 2006). Promisingly, the observation of the single storage compartments is even easier and faster to realise on local scales (Creutzfeldt et al., 2008) and current local studies investigate on the integration of gravity data into hydrological models (Christiansen et al., 2008). Besides difference in scale, the calibration technique of the present study can be valuable for local data combinations, as well. In return, global studies may profit from experimental results on the small scales, in cases they are transferable to the globe. Furthermore, a first combination of terrestrial gravity measurements, e.g., from superconducting gravimeters and global hydrological simulations by Wziontek et al. (2009) are promising for global hydrology.

Lastly, planed and soon started measurement systems, e.g., the European Space Agency's satellite mission Soil Moisture and Ocean Salinity (SMOS) will provide additional global observations of the continental water cycle. The rise of such missions trends towards a permanent observation of the Earth's subsystems, including hydrology.

References

Alcamo, J., Leemans, R., & Kreileman, E., 1998. *Global change scenarios of the 21st century. Results from the IMAGE 2.1 model*, Pergamon, Oxford.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T., & Siebert, S., 2003. Development and testing of the WaterGAP 2 global model of water use and availability, *Hydrolog. Sci. J.*, **48**(3), 317–338.

Andersen, O. B., Seneviratne, S. I., Hinderer, J., & Viterbo, P., 2005. GRACE-derived terrestrial water storage depletion associated with the 2003 European heat wave, *Geophys. Res. Lett.*, **32**, L18405.

Basara, J. B. & Crawford, T. M., 2000. Improved installation procedures for deep-layer soil moisture measurements, *J. Atmos. Ocean Tech.*, **17**(6), 879–884.

Bates, B., Kundzewicz, Z., Wu, S., & Palutikof, J. P., 2008. *Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change*, IPCC Secretariat, Geneva, pp. 210.

Batjes, N. H., 1996. Development of a world data set of soil water retention properties using pedotransfer rules., *Geoderma*, **71**(1-2), 31–52.

Baumgartner, A. & Liebscher, H. J., 1990. *Allgemeine Hydrologie*, Gebrüder Borntraeger, Berlin Stuttgart.

Bergström, S., 1995. The HBV model, in *Computer Models of Watershed Hydrology*, pp. 443–476, Water Resources Publications, Colorado.

Beven, K. & Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, **6**(3), 279–298.

Beven, K. & Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, **249**(1-4), 11–29.

Beven, K. J., 2001. *Rainfall-Runoff Modelling. The Primer*, John Wiley & Sons LTD, Chichester.

Boronina, A. & Ramillien, G., 2008. Application of AVHRR imagery and GRACE measurements for calculation of actual evapotranspiration over the Quaternary aquifer (Lake Chad basin) and validation of groundwater models, *J. Hydrol.*, **348**(1-2), 98–109.

Chambers, D. P., Tamisiea, M. E., Nerem, R. S., & Ries, J. C., 2007. Effects of ice melting on GRACE observations of ocean mass trends, *Geophys. Res. Lett.*, **34**, L05610.

Chao, B. F., 2005. On inversion for mass distribution from global (time-variable) gravity field, *J. Geodyn.*, **39**(3), 223–230.

Chen, J., Wilson, C., Famiglietti, J., & Rodell, M., 2007a. Attenuation effect on seasonal basin-scale water storage changes from GRACE time-variable gravity, *J. Geodesy*, **81**(4), 237–245.

Chen, J. L., Wilson, C. R., Famiglietti, J. S., & Rodell, M., 2005. Spatial sensitivity of the Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations, *J. Geophys. Res.*, **110**, B08408.

Chen, J. L., Tapley, B. D., & Wilson, C. R., 2006a. Alaskan mountain glacial melting observed by satellite gravimetry, *Earth Planet. Sc. Lett.*, **248**(1-2), 368–378.

Chen, J. L., Wilson, C. R., & Seo, K. W., 2006b. Optimized smoothing of Gravity Recovery and Climate Experiment (GRACE) time-variable gravity observations, *J. Geophys. Res.*, **111**, B06408.

Chen, J. L., Wilson, C. R., Tapley, B. D., & Grand, S., 2007b. GRACE detects coseismic and postseismic deformation from the Sumatra-Andaman earthquake, *Geophys. Res. Lett.*, **34**, L13302.

Chen, J. L., Wilson, C. R., Tapley, B. D., Blankenship, D., & Young, D., 2008. Antarctic regional ice loss rates from GRACE, *Earth Planet. Sc. Lett.*, **266**(1-2), 140–148.

Chen, J. L., Wilson, C. R., Tapley, B. D., Yang, Z. L., & Niu, G. Y., 2009. 2005 drought event in the Amazon River basin as measured by GRACE and estimated by climate models, *J. Geophys. Res.*, **114**(B13), 5404.

Choi, M., Jacobs, J. M., & Cosh, M. H., 2007. Scaled spatial variability of soil moisture fields, *Geophys. Res. Lett.*, **34**, L01401.

Christiansen, L., Leiriao, S., He, X., Andersen, O. B., & Bauer-Gottwein, P., 2008. Spatial and temporal gravity data used for hydrological model calibration: Field study of a recharge event in the Okavango Delta, Botswana, *Trans. Am. Geophys. U.*, **89**(53).

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., & Hay, L. E., 2008. Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, **44**, W00B02.

Creutzfeldt, B., Güntner, A., Klügel, T., & Wziontek, H., 2008. Simulating the influence of water storage changes on the superconducting gravimeter of the geodetic observatory Wettzell, Germany, *Society of Exploration Geophysics*, **73**, 6.

Crowley, J. W., Mitrovica, J. X., Richard, C. B., Tamisiea, M. E., & Davis, J. L., 2006. Land water storage within the Congo Basin inferred from GRACE satellite gravity data, *Geophys. Res. Lett.*, **33**, L19402.

Davis, J. L., Tamisiea, M. E., Elosegui, P., Mitrovica, J. X., & Hill, E. M., 2008. A statistical filtering approach for gravity recovery and climate experiment (grace) gravity data, *J. Geophys. Res.*, **113**, B04410.

Deardorff, J. W., 1978. Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation, *J. Geophys. Res.*, **83**(C4).

Deb, K., Agrawal, S., Pratap, A., & Meyarivan, T., 2000. A fast elitist non-dominated sorting genetic algorithm for mulit-objective optimization: NSGA-II, Tech. rep., Kanpur Genetic Algorithms Laboratory.

Di Baldassarre, G. & Montanari, A., 2009. Uncertainty in river discharge observations: A quantitative analysis, *Hydrol. Earth Syst. Sc.*, **13**(6), 913–921.

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., & Hanasaki, N., 2006. GSWP-2: Multimodel analysis and implications for our perception of the land surface, *B. Am. Meteorol. Soc.*, **87**(10), 1381–1397.

Dobslaw, H. & Thomas, M., 2007. Simulation and observation of global ocean mass anomalies, *J. Geophys. Res.*, **112**, C05040.

Döll, P. & Fiedler, K., 2008. Global-scale modeling of groundwater recharge, *Hydrol. Earth Syst. Sc.*, **12**, 863–885.

Döll, P. & Lehner, B., 2002. Validation of a new global 30-min drainage direction map, *J. Hydrol.*, **258**(1-4), 214–231.

Döll, P., Lehner, B., & Kaspar, F., 2002. Global modeling of groundwater recharge, in *Proceedings of Third International Conference on Water Resources and the Environment Research, Vol. 1* , pp. 27–31, Technical University of Dresden, Germany.

Döll, P., Kaspar, F., & Lehner, B., 2003. A global hydrological model for deriving water availability indicators: Model tuning and validation, *J. Hydrol.*, **270**(1-2), 105–134.

Döll, P., Fiedler, K., & Zhang, J., 2009. Global-scale analysis of river flow alterations due to water withdrawals and reservoirs, *Hydrol. Earth Syst. Sc. Disc.*, **6**, 4773–4812.

Duan, Q., 2003. Global optimization for watershed model calibration, in *Calibration of watershed models*, American Geophysical Union, Washington, D.C.

Duan, Q., Gupta, H. V., Sorooshian, S., Rousseau, A. N., & Turcotte, R., 2003. *Calibration of watershed models*, American Geophysical Union, Washington, D.C.

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., & Tarpley, J. D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model, *J. Geophys. Res.*, **108**(D22), 8851.

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

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L., 2007. A comparison of alternative multiobjective calibration strategies for hydrological modeling, *Water Resour. Res.*, **43**, W03434.

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L., 2008. Understanding catchment behavior through stepwise model concept improvement, *Water Resour. Res.*, **44**, W01402.

Fiedler, K. & Döll, P., 2007. Global modelling of continental water storage changes - sensitivity to different climate data sets, *Adv. Goesci.*, **11**, 63–68.

Flechtner, F., 2007. GFZ Level-2 processing standards document for Level-2 product Release 0004, Rev. 1.0, GRACE 327-743 (GR-GFZ-STD-001), Deutsches GeoForschungsZentrum (GFZ), Potsdam, Germany.

Flechtner, F., 2009. GRACE science data system monthly report, feb. 2009, Tech. rep., isdc.gfz-potsdam.de.

Frappart, F., Ramillien, G., Biancamaria, S., Mognard, N. M., & Cazenave, A., 2006. Evolution of high-latitude snow mass derived from the GRACE gravimetry mission (2002-2004), *Geophys. Res. Lett.*, **33**, L02501.

Gulden, L. E., Rosero, E., Yang, Z.-L., Rodell, M., Jackson, C. S., Niu, G.-Y., Yeh, P. J. F., & Famiglietti, J., 2007. Improving land-surface model hydrology: Is an explicit aquifer model better than a deeper soil profile?, *Geophys. Res. Lett.*, **34**, L09402.

Güntner, A., 2009. Improvement of global hydrological models using GRACE data, *Surv. Geophys.*, **29**(4-5), 375–397.

Güntner, A., Schmidt, R., & Döll, P., 2007a. Supporting large-scale hydrogeological monitoring and modelling by time-variable gravity data, *Hydrogeol. J.*, **15**(1), 167–170.

Güntner, A., Stuck, J., Werth, S., Döll, P., Verzano, K., & Merz, B., 2007b. A global analysis of temporal and spatial variations in continental water storage, *Water Resour. Res.*, **43**, W05416.

Gupta, H. V., Sorooshian, S., & Yapo, P. O., 1998. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, **34**(4), 751–763.

Gupta, H. V., Sorooshian, S., Hogue, T. S., & Boyle, D. P., 2003. Advances in automatic calibration of watershed models, in *Calibration of watershed models*, American Geophysical Union, Washington, D.C.

Gupta, H. V., Beven, K. J., & Wagener, T., 2005. Model calibration and uncertainty estimation, in *Encyclopedia of Hydrological Sciences, 3, 131*, Wiley, Chichester.

Haeberli, W. & Beniston, M., 1998. Climate change and its impacts on glaciers and permafrost in the Alps, *Research for Mountain Area Development: Europe*, **27**(4), 258–265.

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N., 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting - I. Basic concept, *Tellus A*, **57**(3), 219–233.

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, *J. Geophys. Res.*, **109**, B04403.

Han, S. C., Shum, C. K., Jekeli, C., Kuo, C. Y., Wilson, C., & Seo, K. W., 2005. Non-isotropic filtering of GRACE temporal gravity for geohysical signal enhancement, *Geophys. J. Int.*, **163**(1), 18–25.

Heiskanen, W. & Moritz, H., 1967. *Physical Geodesy*, W.H. Freeman and Co., San Francisco, CA/USA.

Hirschi, M., Seneviratne, S. I., & Schär, C., 2006. Seasonal variations in terrestrial water storage for major midlatitude river basins, *J. Hydrometeorol.*, **7**(1), 39–60.

Hornberger, G. M. & Spear, R. C., 1981. An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, **12**(1), 7–18.

Horwath, M. & Dietrich, R., 2006. Errors of regional mass variations inferred from GRACE monthly solutions, *Geophys. Res. Lett.*, **33**, L07502.

Horwath, M. & Dietrich, R., 2009. Signal and error in mass change inferences from GRACE: the case of Antarctica, *Geophys. J. Int.*, **177**(3), 849–864.

Houtekamer, P. L. & Mitchell, H. L., 1998. Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, **126**(3), 796–811.

Hunger, M. & Döll, P., 2008. Value of river discharge data for global-scale hydrological modeling, *Hydrol. Earth Syst. Sc.*, **12**(3), 841–861.

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

Jakeman, A. & Hornberger, G., 1993. How much complexity is warranted in a rainfall-runoff model?, *Water Resour. Res.*, **29**, 2637–2649.

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

Jones, J. A. A., 1997. *Global hydrology. Processes, resources and environmental management*, Longman, Harlow.

Kaspar, F., 2004. *Entwicklung und Unsicherheitsanalyse eines globalen hydrologischen Modells.*, Ph.D. thesis, Universität Kassel.

Klees, R., Zapreeva, E. A., Winsemius, H. C., & Savenije, H. H. G., 2007. The bias in GRACE estimates of continental water storage variations, *Hydrol. Earth Syst. Sc.*, **11**(4), 1227–1241.

Klees, R., Liu, X., Wittwer, T., Gunter, B., Revtova, E., Tenzer, R., Ditmar, P., Winsemius, H., & Savenije, H., 2008. A comparison of global and regional GRACE models for land hydrology, *Surv. Geophys.*, **29**(4-5), 335–359.

Kollat, J. B. & Reed, P. M., 2006. Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design, *Adv. Water Resour.*, **29**(6), 792–807.

Kollet, S. J. & Maxwell, R. M., 2008. Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model, *Water Resour. Res.*, **44**, W02402.

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

Kusche, J., Schmidt, R., Petrovic, S., & Rietbroek, R., 2009. Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model, *J. Geodesy*, (Online first, doi: 10.1007/s00190-009-0308-3).

Lawrence, D. M. & Slater, A. G., 2005. A projection of severe near-surface permafrost degradation during the 21st century, *Geophys. Res. Lett.*, **32**, L24401.

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, *Water Resour. Res.*, **45**, W04408.

Legates, D. R. & Willmott, C. J., 1990. Mean seasonal and spatial variability in gauge-corrected global precipitation, *Int. J. Climatol.*, **10**(2), 111–127.

Lehner, B. & Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands, *J. Hydrol.*, **296**(1-4), 1–22.

Lemoine, J.-M., Bruinsma, S., Loyer, S., Biancale, R., Marty, J.-C., Perosanz, F., & Balmino, G., 2007. Temporal gravity field models inferred from GRACE data, *Adv. Space Res.*, **39**(10), 1620–1629.

Lettenmaier, D. P. & Famiglietti, J. S., 2006. Hydrology: Water from on high, *Nature*, **444**(7119), 562–563.

Liu, J., Williams, J. R., Zehnder, A. J., & Yang, H., 2007. GEPIC - modelling wheat yield and crop water productivity with high resolution on a global scales, *Agr. Syst.*, **94**(2), 478–493.

Liu, J., Zehnder, A. J. B., & Yang, H., 2009. Global consumptive water use for crop production: The importance of green water and virtual water, *Water Resour. Res.*, **45**, W05428.

Lo, M.-H., Yeh, P. J.-F., & Famiglietti, J., 2008. Constraining water table depth simulations in a land surface model using estimated baseflow, *Adv. Water Resour.*, **31**(12), 1552–1564.

Lo, M.-H., Famiglietti, J. S., Yeh, P. J.-F., & Syed, T. H., 2010. Improving parameter estimation and water table depth simulation in a land surface model using GRACE water storage and estimated baseflow data, *Water Resour. Res.*, (in press).

Luthcke, S. B., Zwally, H. J., Abdalati, W., Rowlands, D. D., Ray, R. D., Nerem, R. S., Lemoine, F. G., McCarthy, J. J., & Chinn, D. S., 2006. Recent Greenland ice mass loss by drainage system from satellite gravity observations, *Science*, **314**(5803), 1286–1289.

Milly, P. C. & Shmakin, A. B., 2002a. Global modeling of land water and energy balances. Part I: The Land Dynamics (LaD) model, *J. Hydrometeorol.*, **3**(3), 283–299.

Milly, P. C. & Shmakin, A. B., 2002b. Global modeling of land water and energy balances. Part II: Land-characteristic contributions to spatial variability, *J. Hydrometeorol.*, **3**(3), 301–310.

Nash, J. E. & Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part 1 - A discussion of principles, *J. Hydrol.*, **10**(3), 282–290.

Neumeyer, J., Barthelmes, F., Dierks, O., Flechtner, F., Harnisch, M., Harnisch, G., Hinderer, J., Imanishi, Y., Kroner, C., Meurers, B., Petrovic, S., Reigber, C., Schmidt, R., Schwintzer, P., Sun, H. P., & Virtanen, H., 2006. Combination of temoral gravity variations resulting from superconducting gravimeter (SG) recordings, GRACE satellite observations and global hydrology models, *J. Geodesy*, **79**(10-11), 573–585.

Ngo-Duc, T., Laval, K., Ramillien, G., Polcher, J., & Cazenave, A., 2007. Validation of the land water storage simulated by Organising Carbon and Hydrology in Dynamic Ecosystems (ORCHIDEE) with Gravity Recovery and Climate Experiment (GRACE) data, *Water Resour. Res.*, **43**, W04427.

Niu, G. Y. & Yang, Z. L., 2006. Assessing a land surface model's improvements with GRACE estimates., *Geophys. Res. Lett.*, **33**, L07401.

Niu, G. Y., Seo, K. W., Yang, Z. L., Wilson, C., Su, H., Chen, J., & Rodell, M., 2007a. Retrieving snow mass from GRACE terrestrial water storage change with a land surface model, *Geophys. Res. Lett.*, **34**, L15704.

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H., 2007b. Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data, *J. Geophys. Res.*, **112**, D07103.

Papa, F., Güntner, A., Frappart, F., Prigent, C., & Rossow, W. B., 2008. Variations of surface water extent and water storage in large river basins: A comparison of different global data sources, *Geophys. Res. Lett.*, **35**, L11401.

Parajka, J. & Blöschl, G., 2008. The value of MODIS snow cover data in validating and calibrating conceptual hydrologic models, *J. Hydrol.*, **358**(3-4), 240–258.

Peel, M. C., Finlayson, B. L., & McMahon, T. A., 2007. Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sc.*, **11**, 1633–1644.

Press, W. H., Teukolsky, S. T., Vetterling, W. T., & Flannery, B. P., 1992. *Numerical recipes in FORTRAN: The art of scientific computing (2nd ed.)*, Cambridge University Press.

Priestley, C. H. & Taylor, R. J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters, *Mon. Weather Rev.*, **100**(2), 81–92.

Ramillien, G., Frappart, F., Cazenave, A., & Güntner, A., 2005. Time variations of land water storage from an inversion of 2 years of GRACE geoids, *Earth Planet. Sc. Lett.*, **235**(1-2), 283–301.

Ramillien, G., Frappart, F., Güntner, A., Ngo-Duc, T., Cazenave, A., & Laval, K., 2006. Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry, *Water Resour. Res.*, **42**, W10403.

Ramillien, G., Bouhours, S., Lombard, A., Cazenave, A., Flechtner, F., & Schmidt, R., 2008a. Land water storage contribution to sea level from GRACE geoid data over 2003-2006, *Global Planet. Change*, **60**(3-4), 381–392.

Ramillien, G., Famiglietti, J., & Wahr, J., 2008b. Detection of continental hydrology and glaciology signals from GRACE: A review, *Surv. Geophys.*, **29**(4-5), 361–374.

Reed, P., Minsker, B. S., & Goldberg, D. E., 2003. Simplifying multiobjective optimization: An automated design methodology for the nondominated sorted genetic algorithm-II, *Water Resour. Res.*, **39**(7), 1196.

Regonda, S. K., Rajagopalan, B., Clark, M., & Zagona, E., 2006. A multimodel ensemble forecast framework: Application to spring seasonal flows in the Gunnison River basin, *Water Resour. Res.*, **42**, W09404.

Reichle, R. H., McLaughlin, D. B., & Entekhabi, D., 2002. Hydrologic data assimilation with the ensemble Kalman filter, *Mon. Weather Rev.*, **130**(1), 103–114.

Reigber, C., Schmidt, R., Flechtner, F., König, R., Meyer, U., Neumayer, K. H., Schwintzer, P., & Zhu, S. Y., 2005. An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S, *J. Geodyn.*, **39**(1), 1–10.

Ries, J. & Bettadpur, S., 2008. Proceeding summary of the GRACE Science Team Meeting, CSR-GR-08-01, San Francisco, December 12-13, 2008. Center for Space Research, University of Texas at Austin.

Rodell, M., Famiglietti, J. S., Chen, J., Seneviratne, S. I., Viterbo, P., Holl, S., & Wilson, C. R., 2004a. Basin scale estimates of evapotranspiration using GRACE and other observations, *Geophys. Res. Lett.*, **31**, L20504.

Rodell, M., Houser, P. R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C. J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, J. K. E. M., Walker, J. P., Lohmann, D., & Toll, D., 2004b. The Global Land Data Assimilation System, *B. Am. Meteorol. Soc.*, **85**(3), 381–394.

Rodell, M., Chen, J., Kato, H., Famiglietti, J., Nigro, J., & Wilson, C., 2006. Estimating groundwater storage changes in the Mississippi River basin (USA) using GRACE, *Hydrogeol. J.*, **15**(1), 159–166.

Sasgen, I., Martinec, Z., & Fleming, K., 2006. Wiener optimal filtering of GRACE data, *Stud. Geophys. Geod.*, **50**(4), 499–808.

Save, H., Bettadpur, S., & Tapley, B. D., 2008. The use of regularization for global GRACE solutions, in *Proceeding of the GRACE Science Team Meeting, San Francisco*, CSR-GR-08-01, Center for Space Research, The University of Texas at Austin, USA.

Savenije, H. H. G., 2009. HESS Opinions "The art of hydrology", *Hydrol. Earth Syst. Sc.*, **13**(2), 157–161.

Schaefli, B. & Gupta, H. V., 2007. Do Nash values have value?, *Hydrol. Process.*, **21**(15), 2075–2080.

Schmidt, M., Seitz, F., & Shum, C. K., 2008a. Regional four-dimensional hydrological mass variations from GRACE, atmospheric flux convergence, and river gauge data, *J. Geophys. Res.*, **113**, B10402.

Schmidt, R., Schwintzer, P., Flechtner, F., Reigber, C., Güntner, A., Döll, P., Ramillien, G., Cazenave, A., Petrovic, S., Jochmann, H., & Wünsch, J., 2006. GRACE observations of changes in continental water storage, *Global Planet. Change*, **50**(1-2), 112–126.

Schmidt, R., Flechtner, F., Güntner, A., König, R., Meyer, U., Neumayer, K. H., Reigber, C., Rothacher, M., Petrovic, S., & Zhu, S. Y., 2007. GRACE time-variable gravity accuracy assessment, in *IAG Symposium Dynamic Planet. Monitoring and Understanding a Dynamic Planet with Geodetic and Oceanographic Tools, International Association of Geodesy Symposia, Vol. 130, Cairns/Australia, 22.-26. August 2005, pp. 237-243*, pp. 237–243, Springer, Berlin Heidelberg New York.

Schmidt, R., Flechtner, F., Meyer, U., Neumayer, K. H., Dahle, C., König, R., & Kusche, J., 2008b. Hydrological signals observed by the GRACE satellites, *Surv. Geophys.*, **29**(4-5), 319–334.

Schmidt, R., Petrovic, S., Güntner, A., Barthelmes, F., Wünsch, J., & Kusche, J., 2008c. Periodic components of water storage changes from GRACE and global hydrology models, *J. Geophys. Res.*, **113**, B08419.

Schrama, E. J. O., Wouters, B., & Lavalle, D., 2007. Signal and noise in Gravity Recovery and Climate Experiment (GRACE) observed surface mass variations, *J. Geophys. Res.*, **112**, B08407.

Schulze, K. & Döll, P., 2004. Neue Ansätze zur Modellierung von Schneeakkumulation und -schmelze im globalen Wassermodell WaterGAP, in *7th Workshop for large-scale modeling in Hydrology. Munic, November 2003*, pp. 145–154, Kassel University Press, Kassel.

Seitz, F., Schmidt, M., & Shum, C. K., 2008. Signals of extreme weather conditions in Central Europe in GRACE 4-D hydrological mass variations, *Earth Planet. Sc. Lett.*, **268**(1-2), 165–170.

Seneviratne, S. I., Viterbo, P., Lüthi, D., & Schär, C., 2004. Inferring changes in terrestrial water storage using ERA-40 reanalysis data: The Mississippi river basin, *J. Climate*, **17**(11), 2039–2057.

Seo, K. W., Wilson, C. R., Famiglietti, J. S., Chen, J. L., & Rodell, M., 2006. Terrestrial water mass load changes from Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.*, **42**, W05417.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., & McCabe, M. F., 2009. Closing the terrestrial water budget from satellite remote sensing, *Geophys. Res. Lett.*, **36**, L07403.

Shuttleworth, W. J., 1993. Evaporation, in *Handbook of hydrology*, pp. 4.1–4.53, ed. Maidment, D., McGraw-Hill Inc.

Smith, L. C. & Pavelsky, T. M., 2008. Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena river, siberia, *Water Resour. Res.*, **44**, W03427.

Steffen, H., Denker, H., & Müller, J., 2008. Glacial isostatic adjustment in Fennoscandia from GRACE data and comparison with geodynamical models, *J. Geodyn.*, **46**(3-5), 155–164.

Strassberg, G., Scanlon, B. R., & Rodell, M., 2007. Comparison of seasonal terrestrial water storage variations from GRACE with groundwater-level measurements from the High Plains Aquifer (USA), *Geophys. Res. Lett.*, **34**, L14402.

Swenson, S. & Wahr, J., 2002. Methods for inferring regional surface-mass anomalies from Gravity Recovery and Climate Experiment (GRACE) measurements of time-variable gravity, *J. Geophys. Res.*, **107**(B9), 2193.

Swenson, S. & Wahr, J., 2006. Post-processing removal of correlated errors in GRACE data, *Geophys. Res. Lett.*, **33**, L08402.

Swenson, S. & Wahr, J., 2007. Multi-sensor analysis of water storage variations of the Caspian Sea, *Geophys. Res. Lett.*, **34**, L16401.

Swenson, S. & Wahr, J., 2009. Monitoring the water balance of lake Victoria, East Africa, from space, *J. Hydrol.*, **370**(1-4), 163–176.

Swenson, S., Yeh, P. J. F., Wahr, J., & Famiglietti, J. S., 2006. A comparison of terrestrial water storage variations from GRACE with in situ measurements from Illinois, *Geophys. Res. Lett.*, **33**, L16401.

Swenson, S., Famiglietti, J., Basara, J., & Wahr, J., 2008. Estimating profile soil moisture and groundwater variations using GRACE and Oklahoma Mesonet soil moisture data, *Water Resour. Res.*, **44**, W01413.

Syed, T. H., Famiglietti, J. S., Zlotnicki, V., & Rodell, M., 2007. Contemporary estimates of pan-arctic freshwater discharge from GRACE and reanalysis, *Geophys. Res. Lett.*, **34**, L19404.

Syed, T. H., J. S., F., Rodell, M., Chen, J., & Wilson, C. R., 2008. Analysis of terrestrial water storage changes from GRACE and GLDAS, *Water Resour. Res.*, **44**, W02433.

Tamisiea, M. E., Mitrovica, J. X., & Davis, J. L., 2007. GRACE gravity data constrain ancient ice geometries and continental dynamics over laurentia, *Science*, **316**(5826), 881–883.

Tang, Y., Reed, P., & Wagener, T., 2006. How effective and efficient are multiobjective evolutionary algorithms at hydrologic model calibration?, *Hydrol. Earth Syst. Sc.*, **10**, 289–307.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M., 2004a. GRACE measurements of mass variability in the earth system, *Science*, **305**(5683), 503–505.

Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C., 2004b. The gravity recovery and climate experiment: Mission overview and early results, *Geophys. Res. Lett.*, **31**, L09607.

Tebaldi, C. & Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections, *Philos. T. R. Soc. A*, **365**(1857), 2053–2075.

Thiemann, M., Trosset, M., Gupta, H. V., & Sorooshian, S., 2001. Bayesian recursive parameter estimation for hydrologic models, *Water Resour. Res.*, **37**(10), 2521–2535.

Torge, W., 2003. *Geodäsie*, de Gruyter, Berlin.

Velicogna, I. & Wahr, J., 2006. Measurements of time-variable gravity show mass loss in Antarctica, *Science*, **311**(5768), 1754–1756.

Verzano, K., 2009. *Climate change impacts on flood related hydrological processes: Further develpoment and application of a global scale hydrological model*, Ph.D. thesis, Kassel University, Germany.

Vincent, T. & Grantham, W., 1981. *Optimality in parametric systems*, John Wiley & Sons LTD, New York.

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., & Sorooshian, S., 2003a. Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, **39**(8), 1214.

Vrugt, J. A., Gupta, H. V., Bouten, W., & Sorooshian, S., 2003b. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, **39**(8), 1201.

Wagener, T., Wheater, H., & Gupta, H., 2003. Identification and evaluation of watershed models, in *Calibration of watershed models*, pp. 29–47, American Geophysical Union, Washington, D.C.

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, *J. Geophys. Res.*, **108**(B12), 30,205–30,229.

Wahr, J., Swenson, S., Zlotnicki, V., & Velicogna, I., 2004. Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, **31**, L11501.

Wahr, J., Swenson, S., & Velicogna, I., 2006. Accuracy of GRACE mass estimates, *Geophys. Res. Lett.*, **33**, L06401.

Watkins, M. M., Yuan, D. N., Kuang, D., Bertiger, W., Byun, S., Lu, W., & Kuizinga, G. L., 2008. JPL L-2 GRACE solutions: Harmonics, mascons, iteration and constaints, in *Proceeding of the GRACE Science Team Meeting, San Francisco*, CSR-GR-08-01, Center for Space Research, The University of Texas at Austin, USA.

Werth, S. & Güntner, A., 2008. Intercomparison of global hydrological models in terms of water storage, in *AGU Fall Meeting, December 2008, San Francisco, USA*.

Werth, S. & Güntner, A., 2009. Calibration analysis for water storage variability of the global hydrological model WGHM, *Hydrol. Earth Syst. Sc. Disc.*, **6**, 4813–4861.

Werth, S., Güntner, A., Petrovic, S., & Schmidt, R., 2009a. Integration of GRACE mass variations into a global hydrological model, *Earth Planet. Sc. Lett.*, **277**(1-2), 166–173.

Werth, S., Güntner, A., Schmidt, R., & Kusche, J., 2009b. Evaluation of GRACE filter tools from a hydrological perspective, *Geophys. J. Int.*, **179**(3), 1499–1515.

Widen-Nilsson, E., Halldin, S., & Xu, C.-y., 2007. Global water-balance modelling with WASMOD-M: Parameter estimation and regionalisation, *J. Hydrol.*, **340**(1-2), 105–118.

Willmot, C. J., 1984. On the evaluation of model performance in physical geography, in *Spatial Statistics and Models*, pp. 443–460, D. Reidel, Dordrecht, Netherlands.

Wilson, D. J., Western, A. W., Grayson, R. B., Berg, A. A., Lear, M. S., Rodell, M., Famiglietti, J. S., Woods, R. A., & McMahon, T. A., 2003. Spatial distribution of soil moisture over 6 and 30 cm depth, Mahurangi river catchment, New Zealand, *J. Hydrol.*, **276**(1-4), 254–274.

Winsemius, H. C., Savenije, H. H. G., Gerrits, A. M. J., Zapreeva, E. A., & Klees, R., 2006a. Comparison of two model approaches in the Zambezi River basin with regard to model reliability and identifiability, *Hydrol. Earth Syst. Sc.*, **10**(3), 339–352.

Winsemius, H. C., Savenije, H. H. G., van de Giesen, N. C., van den Hurk, B. J. J. M., Zapreeva, E. A., & Klees, R., 2006b. Assessment of Gravity Recovery and Climate Experiment (GRACE) temporal signature over the upper Zambezi, *Water Resour. Res.*, **42**, W12201.

Wouters, B., Chambers, D., & Schrama, E. J., 2008. GRACE observes small-scale mass loss in Greenland, *Geophys. Res. Lett.*, **35**, L20501.

Wziontek, H., Wilmes, H., Wolf, P., Werth, S., & Güntner, A., 2009. Time series from superconducting gravimeters and water storage variations from the global hydrology model WGHM, *J. Geodyn.*, (submitted), 166–177.

Yapo, P. O., Gupta, H. V., & Sorooshian, S., 1998. Mulit-objective global optimization for hydrologic models, *J. Hydrol.*, **204**(1-4), 83–97.

Zaitchik, B. F., Rodell, M., & Reichle, R. H., 2008. Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin, *J. Hydrometeorol.*, **9**(3), 535–548.

Zeng, N., Yoon, J.-H., Marengo, J. A., Subramaniam, A., Nobre, C. A., Mariotti, A., & Neelin, J. D., 2008. Causes and impacts of the 2005 Amazon drought, *Environ. Res. Lett.*, **3**, 014002.

Acknowledgments

At this point, I wish to thank all colleagues, friends and my family who supported me in preparing this PhD-thesis, no matter whether it was of professional or private nature.

First of all, I thank Andreas Güntner for a very valuable supervision and guidance through the labyrinth of my doctoral studies. He managed to urge me to intensive work but also to provide me with sufficient scope for development. I want to thank Professor Bruno Merz for his provision of my PhD-position at GFZ and his guidance as professor of my thesis. Thank you, Professor Hubert Savenije, Professor Reinhard Dietrich and Professor Jakob Flury for your willingness to review this dissertation.

I like to mention procreative teamwork with my colleges in the Project TIVAGAM and my coauthors of the published work: Svetozar Petrovic, Roland Schmidt, Professor Jürgen Kusche, Professor Petra Döll and Kristina Fiedler.

My parents, my sister and Manoo: thank you for resisting against dissertation-related bad moods and for supporting my staying power. In special Manoo, thank you for very supportive discussions and your correction work.

I thankfully remember all colleges from the GFZ Section 5.4 for numerous support, no matter technical, experiential or practical know-how. Here, in special memory is Heiko Thoss for having time at all hours for any problem. Steffi Uhlemann, Benjamin Creutzfeldt and Florian Elmer for discussions, support and share of our PhD-fortune. Furthermore, I like to name very valuable experiences I made together with Steffi Uhlemann, Anastasia Galkin and Philipp Kuhn as PhD-representatives at GFZ, that payed out for organising and finishing my dissertation-project at last.

Last but not least, I thankfully memorise time with very important friends for being my flatmates, travel or discussion partners that helped me to sweat out the longsome years as a Ph.D.-student.