KUMULATIVE DISSERTATION ZUR ERLANGUNG DES GRADES
DOKTOR DER NATURWISSENSCHAFTEN (DR. RER. NAT)

Closure of the Water Cycle: Global Analysis and Hydrometeorological Data Assimilation

DIPL.-ING. CHRISTOF LORENZ

Universität Augsburg

Fakultät für Angewandte Informatik

Institut für Geographie

AUGUST 2016

First Reviewer: Prof. Dr. Harald Kunstmann, University of Augsburg
Second Reviewer: Prof. Dr. Peter Fiener, University of Augsburg
Third Reviewer: Prof. Dr. Nico Sneeuw, University of Stuttgart

Date of Oral Exam: 03.08.2016

Abstract

In the discussion about the global climate change, one central topic are the projected changes in the water cycle. It is predicted that there will be an increase of extreme hydrometeorological events like heavy precipitation or droughts. It is obvious that this *intensification* of the hydrological cycle will have a significant impact on the society. Such predictions require reliable and consistent datasets for the major hydrological variables precipitation, evapotranspiration, runoff, and water storage changes. Today, there are various data sources for each of these variables. While some of these datasets are still based on *in situ* measurements, there are alternative data sources available, which are often derived from satellite-based measurements. The advantages of these observations are obviously the homogeneous spatial and temporal resolutions on the global scale. Besides such satellite-based products, state-of-the-art hydrological and hydrometeorological models and reanalyses also provide consistent long-term estimates for the major hydrological variables.

In order to evaluate the past, present, and future state of the climate system, it is mandatory that there is both temporal and spatial consistency between these data sources. Otherwise, the mismatch between the different water cycle variables cause *imbalances* in the empirical evaluation of the hydrological cycle, which, in the end, hinder the analysis of extreme events or variations on climatic time scales. It is thus of major importance to investigate the strengths and weaknesses of the data sources for precipitation, evapotranspiration, runoff, and water storage changes, but also the level of consistency between different water cycle variables.

This doctoral thesis, which comprises of four articles, shall therefore serve as a comprehensive overview over the current status of our knowledge about and our data basis for the large-scale water cycle. Therefore, various data sources for the four major water cycle variables are compared and evaluated on different temporal and spatial scales. These sources comprise gridded observations (GPCC, GPCP, CRU, DEL, CPC), atmospheric reanalysis models (ERA-Interim, MERRA, CFSR), partially model-based datasets (GLEAM, MOD16, FLUXNET MTE), land surface models (GLDAS, MERRA Land), satellite-derived water storage changes from GRACE, and *in situ* runoff observations from GRDC,

The study reveals serious shortcomings in the empirical evaluation of the large-scale water cycle. On the global scale, significant differences can be identified when comparing the model estimates from the three reanalyses against gridded observations of precipitation and temperature. However, differences with similar magnitudes can also be observed between the applied observation-based datasets. A catchment-scale analysis over 96 catchments of different sizes and climatic conditions worldwide confirms that these differences occur on both the global- and the catchment-scale. In the context of the gridded precipitation observations, this can be (at least partly) explained with a significant decrease of rain gauges worldwide. Looking at the spatial distribution of the gauges reveals large data gaps e.g. in the Tropics or the African continent, which leads to a high level of uncertainty in these regions.

The shortcomings in the data sources for each of the four water cycle variables are confirmed by an analysis of the global- and basin-scale water budgets. Due to their ability to simulate the whole climate system, the three reanalysis models allow in principle a consistent evaluation of the global water cycle. However, it is shown that there are significant imbalances and numerical artifacts in their oceanic and continental water budgets, which obviously hinder the use of such model estimates for e.g. extreme value or climate trend studies. On the basin-scale, the evaluation of the water budgets from different combinations of widely used data sources for precipitation, evapotranspiration, runoff, and water storage changes reveal imbalances of more than 25 % of the mean annual runoff over most of the 96 study regions. Even if some data combinations allow a reasonable closure of the water budgets over certain catchments, it is not possible to identify a single best dataset which performs consistently on the global scale. That being said, the significant decrease in the number of stream gauges worldwide further aggravates a continuous analysis of the basin-scale water cycle.

The study therefore presents an approach, with which basin-scale time series of precipitation, evapotranspiration, runoff, and water storage changes can be predicted or corrected. The method is based on an Ensemble Kalman Filter framework, where all required input parameter are derived from an ensemble of hydrological and hydrometeorological datasets. In order to evaluate the performance of the proposed framework, the filter is used for predicting runoff over 16 catchments. A comparison with observed runoff shows correlations larger than 0.5, relative errors lower than ±20 %, and NSE-values larger than 0.5 for most of the study regions.

Overall, the study shows that our current datasets for the major water cycle variables have to be used with care. The large imbalances and inconsistencies in the water budgets on both the global- and basin-scale deny the direct use of such estimates for e.g. climate trend studies or the analysis of extreme events. Thus, in order to use our current datasets for studying the projected changes in the global water cycle, a careful analysis and data correction has to be performed. It is further stressed that, despite the promising performance of certain alternative methods like e.g. the presented EnKF-approach, there is still an urgent need for *in situ* observations for precipitation and runoff. Otherwise, it will become more difficult in the near future to perform water budget studies or climate analyses, but also to validate hydrological or hydrometeorological models.

Zusammenfassung

Die Änderungen des globalen Wasserkreislaufs sind ein zentrales Thema in der Debatte über den globalen Klimawandel. Es wird angenommen, dass die Häufigkeit von beispielsweise Starkniederschlägen oder Dürren zunehmen wird. Die Auswirkungen einer solchen *Intensivierung* des Wasserkreislaufs auf die Gesellschaft liegen auf der Hand. Um jedoch entsprechende Aussagen über solche Änderungen treffen zu können, werden zuverlässige und konsistente Datenquellen für Niederschlag, Verdunstung, Abfluss sowie Wasserspeicheränderungen benötigt. Heutzutage werden für jede einzelne dieser Variablen verschiedenste Datenquellen angeboten. Wärend einige dieser Produkte auf *in situ*-Messungen basieren, gibt es alternative Datensätze, welche oftmals von satellitengestützten Beobachtungen abgeleitet werden. Der Vorteil hierbei ist die (auf globaler Skala) homogene rämliche und zeitliche Auflösung. Neben solchen Satelliten-basierten Produkten liefern moderne hydrologische und hydrometeorologische Modelle und Reanalysen ebenfalls konsistente Schätzungen der wichtigsten hydrologischen Variablen.

Um sowohl die Vergangenheit, als auch Gegenwart und Zukunft unseres Klimasystems zu beschreiben, müssen die Datenquellen sowohl zeitlich, als auch räumlich, konsistent sein. Andernfalls treten bei der empirischen Beschreibung des Wasserkreislaufs Ungleichgewichte auf, welche durch die Diskrepanzen der verschieden Variablen hervorgerufen werden. Diese Ungleichgewichte erschweren die Analyse von extremen Ereignissen oder Variationen über klimatische Zeitskalen. Es ist daher äußerst wichtig, die verschiedenen Stärken und Schwächen der einzelnen Datenquellen für Niederschlag, Verdunstung, Abfluss und Wasserspeicheränderungen, aber auch die Gemeinsamkeiten der verschiedenen Wasserkreislaufvariablen zu untersuchen und zu verstehen.

Diese Arbeit, welche aus vier Artikeln besteht, soll daher einen umfassenden Überblick über den aktuellen Kenntnisstand und die Datengrundlagen für den großskaligen Wasserkreislauf bieten. Dafür werden verschiedene Datenquellen für die Wasserkreislaufsvariablen auf verschiedenen zeitlichen und räumlichen Skalen miteinander

verglichen und evaluiert. Die untersuchten Daten basieren auf gerasterten Beobachtungen (GPCC, GPCP, CRU, DEL, CPC), atmosphärischen Reanalysemodellen (ERA-Interim, MERRA, CFSR), teilweise modellbasierten Ansätzen (GLEAM, MOD16, FLUXNET MTE), Land-Oberflächenmodellen (GLDAS, MERRA Land), Satelliten-basierten Wasserspeicheränderungen von GRACE sowie Abflussmessungen des GRDC.

In dieser Arbeit werden erhebliche Mängel in der empirischen Auswertung des großskaligen Wasserkreislaufs aufgezeigt. Auf globaler Skala lassen sich große Unterschiede zwischen den Modellergebnissen der drei Reanalysen und gerasterten Niederschlags- und Temperaturbeobachtungen feststellen. Die Unterschiede zwischen den verschiedenen beobachtungsbasierten Datensäte jedoch erreichen eine ähnliche Größenordnung. Ein Vergleich für insgesamt 96 Einzugsgebiete mit unterschiedlicher Größe und klimatischen Bedingungen zeigt deutlich, dass diese Unsicherheiten sowohl auf globaler, als auch auf Einzugsgebietsskala auftreten. Bei den gerasterten Niederschlagsbeobachtungen kann dies zumindest teilweise mit einem deutlichen Rückgang der Messstationen weltweit erklärt werden. Betrachtet man die räumliche Verteilung der Messstationen, lassen sich große Datenlücken z.B. über den Tropen oder dem gesamten Afrikanischen Kontinent erkennen. Diese führen natürlich zu erheblichen Unsicherheitsspannen in den entsprechenden Regionen.

Die Defizite in den Datenquellen für die vier Wasserkreislaufsvariablen werden noch offensichtlicher, wenn man die Schließung der Wasserbilanz auf globaler Skala sowie für einzelne Einzugsgebiete untersucht. Eine globale Auswertung des Wasserkreislaufs wäre prinzipiell mit den drei untersuchten Reanalysemodellen möglich, da sie das gesamte Klimasystem simulieren. Allerdings zeigen sich deutliche Ungleichgewichte und numerische Artefakte in den ozeanischen und kontinentalen Wasserbilanzen, was beispielsweise Extremwertanalysen oder Trendstudien erheblich erschwert. Auf der Einzugsgebietssakala können die Wasserbilanzen durch verschiedene Kombinationen der Datenquellen für Niederschlag, Verdunstung, Abfluss und Wasserspeicheränderungen untersucht werden. Dabei lassen sich Ungleichgewichte von mehr als 25 % des mittleren Jahresabflusses für die Mehrzahl der 96 untersuchten Einzugsgebiete feststellen. Auch wenn einige Kombinationen eine annähernde Schließung der Wasserbilanz erlauben, kann hierbei kein einzelner Datensatz ermittelt werden,

welcher in allen Einzugsgebieten konsistente Ergebnisse liefert. Da zusätzlich ein erheblicher Rückgang in der Anzahl der Flusspegel weltweit zu erkennen ist, wird die kontinuierliche Untersuchung des Wasserkreislaufs auf Einzugsgebietsebene immer schwieriger.

In dieser Arbeit wird daher ein Ansatz präsentiert, welcher zur Prädiktion oder Korrektur von Zeitreihen für Niederschlag, Verdunstung, Abfluss oder Wasserspeicheränderungen verwendet werden kann. Die Methode basiert auf einem Ensemble Kalman Filter, wobei sämtliche benötigten Eingabeparameter aus einem Ensemble von hydrologischen und hydrometeorologischen Datensätzen bestimmt werden. Um die Leistungsfähigkeit des Ansatzes zu überprüfen, werden mit Hilfe des Filters Abflusszeitreihen für 16 Einzugsgebiete prädiziert. Der Vergleich mit Abflussbeobachtungen liefert Korrelationen größer 0.5, relative Fehler kleiner als ±20 % sowie NSE-Werte größer als 0.5 für die meisten untersuchten Gebiete.

Insgesamt zeigt die Studie, dass unsere derzeitigen Datensätze für die vier Wasserkreislaufsvariablen nur mit Vorsicht benutzt werden dürfen. Die großen Ungleichgewichte und Unstimmigkeiten in den Wasserbilanzen sowohl auf globaler, als auch auf Einzugsgebietsskala erschweren die Analyse von beispielsweise extremen Ereignissen oder Klimatrends. Zur Untersuchung der Änderungen im globalen Wasserkreislauf muss daher zunächst eine sorgfältige Überprüfung und Korrektur der Datengrundlage erfolgen. Es muss ebenfalls darauf hingewiesen werden, dass trotz der vielversprechenden Ergebnisse des EnKF-Ansatzes nach wie vor ein dringender Bedarf an *in situ*-Beobachtungen für Niederschlag und Abfluss herrscht. Andernfalls wird es in Zukunft erheblich schwieriger, aussagekräftige Wasserhaushaltsstudien oder Klimaanalysen, aber auch Validierungen von hydrologischen und hydrometeorologischen Modellen durchzuführen.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Harald Kunstmann for all his support, patience, kindness, and the possibility to work in this beautiful environment of Garmisch-Partenkirchen. I would also like to thank my Geodesy-supervisor Prof. Dr. Nico Sneeuw, with whom I work together since my Diploma Thesis. My sincere thanks are due to both supervisors for all the fruitful discussions, comments, guidelines, and scientific input which (most of the time) made things a little clearer for me.

I would further like to thank my Geodesy-colleagues Balaji Devaraju, who once was the supervisor of my Diploma Thesis and became a true friend and colleague, and Mohammad J. Tourian for all those many interesting, inspiring, and amicable discussions. This work would not have been possible without their help and support.

I also deeply appreciate the great community in our working group at the Institute of Meteorology and Climate Research. Special thanks are due to my colleagues Michael Warscher, Christian Chwala, Sven Wagner, Jakob Garvelmann, Ingo Völksch, Patrick Laux, Felix Keis, Benjamin Fersch, and the rest of the Regional Climate and Hydrology Team at the IMK-IFU for all the interesting conversations and activities both at and (maybe even more important) off the institute.

Finally, I would like to thank my whole family, but especially my wife Kristine. She always had an open ear and listened to the dozens of issues which I encountered during my thesis. Even if she could not always find a solution to my problems, she nevertheless encouraged me to continue with my work and gave me a lot of motivation and confidence during the last few years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The observed and projected changes in the freshwater availability due to climate change is often expressed by the common and prominent phrase *The rich get richer and the poor get poorer* [e.g., Trenberth, 2011]. Besides the global long-term mean amounts and trends of the four major water cycle variables precipitation, evapotranspiration, runoff, and water storage changes, also changes in both the temporal and spatial distribution are of major concern. These changes are often summarized as the *intensification of the global water cycle* [e.g., Huntington, 2006], which was copiously discussed in IPCC's Fourth Assessment [IPCC, 2007] and IPCC's dedicated report [Field et al., 2012]. In general, it is assumed that there will be an increase in the frequency of extreme hydrometeorological events like heavy precipitation or severe droughts.

The impacts of changes in the water cycle are of high social relevance. In IPCC's fifth assessment [IPCC, 2014], it is concluded that *the fraction of global population experiencing water scarcity and the fraction affected by major floods increase with the level of warming in the 21st century*. Furthermore, they state that *climate change over the 21st century is projected to reduce renewable surface water and groundwater resources significantly in most dry subtropical regions, ..., while water resources are projected to increase at high latitudes*.

Carpenter et al. [2011] estimate that only 0.26 % of liquid freshwater on Earth is in lakes, reservoirs, and rivers. Thus, the freshwater that is directly accessible to human use is only a minor portion of the hydrosphere, which also includes the oceans or the vast ice shields, for example, over Greenland or polar and high-latitude regions. With that background, it is obvious that even small changes in the hydrological cycle can have an alarming impact on the availability of freshwater resources.

Postel et al. [1996] summarize all the terrestrial freshwater compartments, that are accessible to human use, under the term *accessible runoff*. They give early estimates that until 2025, accessible runoff can be increased by about 10 %, while the population is projected to increase by more than 45 %. This clearly indicates that we will face a declining water availability per capita, primarily due to a growing world population. The increase of water requirements was also estimated by Falkenmark and Rockström [2004], who state that each additional person requires 1,300 m^3 of freshwater per capita per year. As a summary, Vörösmarty et al. [2010] estimate that 80 % of the word's population is exposed to high levels of threat to water security and that 65 % of the total continental discharge is classified as moderately to highly threatened.

It is therefore evident that one of the major challenges of the modern hydrological and hydrometeorological communities is a better understanding, modeling, and prediction of the hydrological cycle. This is also identified by Rodell et al. [2015], who state that *in order to identify change, one must first establish the present condition*. This led to several comprehensive initiatives whose aim is the detailed analysis of the major hydrological processes. One of the core projects of the World Climate Research Programme (WCRP) is the Global Energy and Water Exchanges Project [GEWEX; Chahine, 1992a,b], which was established in the late 1980s. Within this project, various comprehensive studies from leading geo-scientific institutes have been realized (e.g., the Baltic Sea Experiment [BALTEX; Raschke et al., 1998] or the African Monsoon Multidisciplinary Analysis Project [AMMA; Redelsperger et al., 2006]) and several widely used datasets have been developed and published (e.g., the global precipitation datasets from the Global Precipitation Climatology Centre [GPCC; Schneider et al., 2008] or the Global Precipitation Climatology Project [GPCP; Adler et al., 2003]). A similar project was founded by the National Aeronautics and Space Administration NASA. The central challenge of the NASA Energy and Water cycle

Study (NEWS) is *to document and enable improved, observationally-based, predictions of energy and water cycle consequences of Earth system variability and change* [NSIT, 2007].

Even if the quantitative description of the water cycle components goes back till the ancient times (e.g., streamflow observations at the river Nile), the start of the satellite era in the 1960s can be considered as a quantum leap in the observation, understanding, and description of the contributing hydrological variables. The main reason is that hydrological processes can now be observed globally with homogeneous spatial and temporal resolutions, independent of, for example, the distribution of *in situ* stations. Today, all components of the terrestrial and atmospheric water cycle can be derived from satellite measurements. Precipitation data can be obtained from the Tropical Rainfall Measuring Mission [TRMM; Kummerow et al., 2000] or the Global Precipitation Measurement [GPM; Smith et al., 2007], water storage changes from the Gravity Recovery and Climate Experiment [GRACE; Schmidt et al., 2008], runoff from satellite altimetry and synthetic aperture radar [Alsdorf et al., 2007], soil moisture from the Soil Moisture and Ocean Salinity [SMOS; Kerr et al., 2010] or the Soil Moisture Active Passive (SMAP) satellites [Entekhabi et al., 2010], evapotranspiration from the Moderate-resolution Imaging Spectroradiometer [MODIS; Mu et al., 2007, 2011], and a host of variables concerning the hydrological cycle from *Aqua* [Parkinson, 2003]. This led to several comprehensive studies of the water cycle using satellite data [e.g., Huntington, 2006, Schlosser and Houser, 2007, Sheffield et al., 2009].

There have also been significant improvements in the performance and reliability of state-of-the-art hydrological and hydrometeorological models. The much higher performance of current computers allows to simulate the climate system in unprecedented spatial and temporal resolutions. Improvements in the understanding of the interactions between the different climate variables also allowed for a much more realistic mathematical description of the climate dynamics in the models. Furthermore, since several decades, the assimilation of observation data into models received more and more attention. It is assumed that this allows to provide far more realistic estimates compared to free model runs. An increasing number of studies from various subjects thus made use of the so-called reanalysis models, where millions of observations are assimilated into complex hydrological or hydrometeorological models in order to provide the most realistic and consistent estimates of climatic variables. In terms of large-scale water balance studies, it has therefore been normal practice to resort to sophisticated hydrological and atmospheric reanalysis models or global datasets based on terrestrial observations [e.g., Trenberth et al., 2007].

1.2 Motivation and Problem Description

However, there is one fundamental question when we use all these datasets for studying the climate system: Are these data sources reliable enough for drawing meaningful conclusions on the past, present, and future state of the climate system? A first guess about the reliability can be obtained when looking at the past and present distribution of rainfall gauges or the distribution of gauged and ungauged river basins in Figure 1.1. Obviously, there are large areas where not a single gauge is located and the number of both streamflow and rainfall gauges significantly decreases during the period 1980 to 2009. If we therefore use current estimates of precipitation minus evapotranspiration from various observation-based datasets for performing a simple evaluation of the trend[1] in the availability of freshwater resources [e.g., Oki and Shinjiro, 2006, Parish et al., 2012], one ends up with highly inconsistent results (see Figure 1.2, top). Moreover, using models from the Coupled Model Intercomparison Project Phase 5 [CMIP5; Taylor et al., 2012] for such an evaluation results in completely different trend estimates compared to the observation-based datasets (see Figure 1.2, bottom). Over most regions of the Earth, we are therefore not able to give a clear statement about the future development of the water availability when using state-of-the-art data sources for the most important water cycle variables. With respect to the intensification of the water cycle, Huntington [2006] report that empirical evidence to date does not consistently support an increase in the frequency or intensity of extreme hydrometeorological events.

---

[1] It should be noted that this work does not focus on trend studies. The figure should just illustrate the discrepancies between our current observation- and model-based data sources for the major hydrological variables.

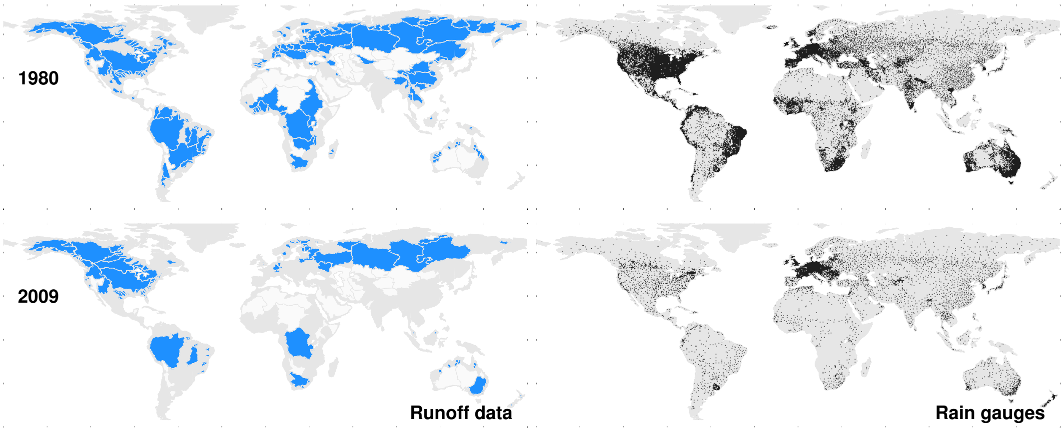

FIGURE 1.1: From Lorenz et al. [2014]: The maps show the decline in the number of runoff (left) and rain (right) gauges. The catchments whose runoff measurements are publicly available (blue), that do not discharge into the oceans or other open water bodies (white) and those that are not gauged or whose runoff measurements are not publicly available (gray) are shown on the left column. In the right column the individual dots represent one or more rain gauges in a $0.5^\circ \times 0.5^\circ$ gridcell.

There are various reasons for the these discrepancies and each water cycle variable suffers from different and individual limitations. It is known that, for example, the quality of observation based global precipitation data depends significantly on the number and distribution of *in situ* stations [Rudolf and Schneider, 2005, Schneider et al., 2008]. Even if there are dense precipitation observation networks in developed countries, our understanding of the distribution of precipitation in data-sparse regions relies on sporadic rain gauges and information gathered by spaceborne sensors [Bytheway and Kummerow, 2013]. However, satellite-based precipitation estimates suffer from non-negligible biases and random errors associated with inadequate sampling, algorithm errors, and the indirect nature of the physical relationship between precipitation and the observations [Xie and Arkin, 1996].

For a long time, evapotranspiration could not directly be observed on large scales—that is, one had to rely solely on models. Although there are several attempts to provide global evapotranspiration data [e.g., Jung et al., 2009, Miralles et al., 2011b, Mu et al., 2007, Salvucci and Gentine, 2013], these products nevertheless involve a significant portion of modeling. On the other hand, there is still a lack of a global benchmark for evapotranspiration [Mueller et al., 2013] in order to validate these datasets on larger spatial scales and dense global coverage of point measurements, for

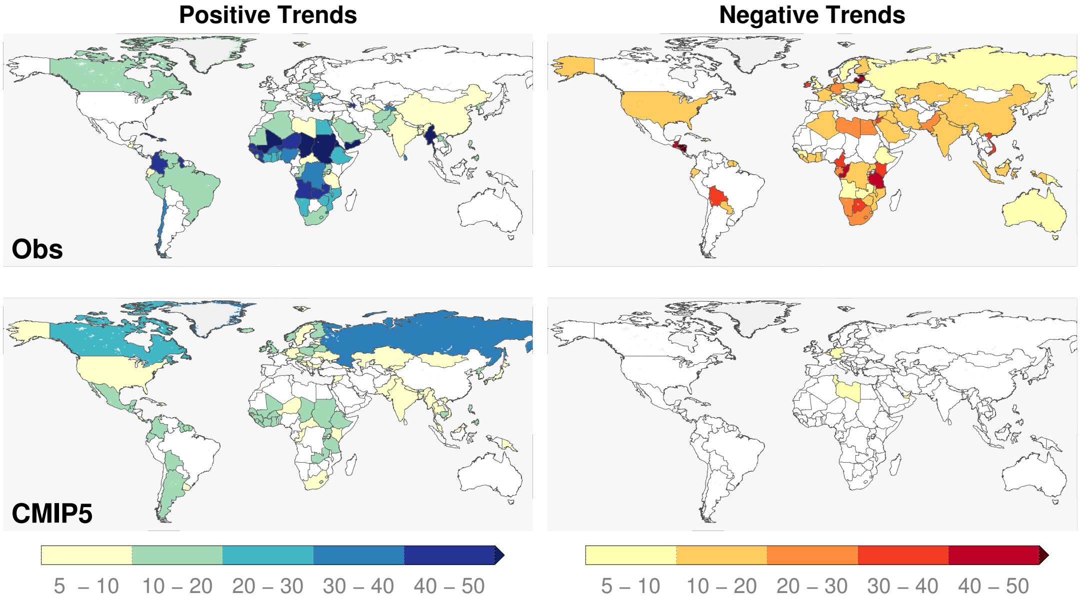

FIGURE 1.2: Percent of 20 observation-based datasets (top) and 28 CMIP5 models (bottom), which show a significant positive (left) and negative (right) trend in the annual sums of precipitation minus evapotranspiration. On longer time-scales, $P-ET$ serves as an approximation of the amount of renewable freshwater resources. The trends have been derived from the period 1980–2005.

example, from the FLUXNET project [Baldocchi et al., 2001], is not feasible [Mueller et al., 2011].

Until the launch of GRACE, observing large-scale water storage variations was not possible. But even after more than 12 years of experience with GRACE data, the spatial resolution is still too low for many basin-scale hydrological applications [e.g., Landerer and Swenson, 2012]. It is widely assumed that the horizontal resolution of GRACE is $\approx$ 400 km [e.g., Rodell and Famiglietti, 1999, Wahr et al., 1998] which limits its applications to basins with an area larger than 400,000 km^2. Furthermore, validation of water storage variations from GRACE is rather difficult as it is still the only sensor of it's kind and comparing *in situ* measurements to data from GRACE is, at least, questionable due to its coarse resolution.

Finally, even if runoff is still the most accurately observed variable of the basin-scale water cycle [Fekete et al., 2012], many catchments are ungauged [Blöschl et al., 2013] and the number of gauges is steadily decreasing [Fekete et al., 2012, Milzow et al., 2011, Shiklomanov et al., 2002, Sivapalan et al., 2003]. Dai and Trenberth [2002] report that the annual runoff rate over the unmonitored areas is comparable to that over the monitored areas. As discharge is one of the most important sources of freshwater, there is thus an urgent need for alternative methods in order to estimate runoff over ungauged or sparsely gauged regions. However, even if Alsdorf et al. [2007] provide a comprehensive overview of methods based on satellite data for deriving runoff, they also report of large uncertainties in such estimates.

Obviously, there are still significant shortcomings in our current knowledge and the representativeness of each single variable in the terrestrial water cycle, which, in the end, complicate reliable water budget analyses. It is therefore mandatory to compare and analyze our present data sources in order to better understand the similarities, but, maybe even more important, also the uncertainties with which we have to cope. Such comparisons have to be conducted on various spatial and temporal scales for obtaining a reasonable evaluation of our present hydrological and hydrometeorological data sources.

1.3 Overall Aim and Structure of the Study

This work shall thus serve as a comprehensive state-of-the-art analysis of our current knowledge of the global hydrological cycle. The main research questions behind this thesis are

1. How well do we know the basin- to global-scale water cycle?
2. How reliable are our current state-of-the-art hydrological and hydrometeorological datasets?
3. Are our current hydrological and hydrometeorological data sources able to close the water budget?
4. Can we use our current database for deriving reliable predictions of the major hydrological processes?

In order to answer these questions, the thesis comprises four peer-reviewed articles and is structured as follows:

Chapter 2: The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis In the first article, three widely used atmospheric reanalysis models, namely ERA-Interim, MERRA, and CFSR, are compared and evaluated with independent observation based datasets for precipitation, temperature, and atmospheric water vapor content. The analysis is carried out on global, hemispheric, and continental scales during the period 1989 to 2006. The article also discusses the spatial and temporal distribution of *in situ* stations, which are used for generating the global precipitation datasets. It further comprises a spatial comparison, where the long-term mean patterns of precipitation and temperature are compared against precipitation data from GPCC, CRU, and CPC, and temperature data from CRU. The temporal variations are evaluated by comparing annual and monthly time series of precipitation globally, over the Northern and Southern Hemisphere, the tropics, North America, Europe, Asia, South America, Africa, and Australia. It is further evaluated how well the three reanalysis models close the large-scale water balance. Therefore, long-term mean area aggregated precipitation, evapotranspiration, runoff, and moisture flux divergences are used as inputs in different water balance equations and the remaining imbalances are discussed in detail. Finally, estimates of the total atmospheric water vapor from the three reanalyses are compared. For an evaluation of the precipitable water over the oceans, the modeled estimates are compared against the Hamburg Ocean Atmosphere Parameters and fluxes from Satellite data (HOAPS).

As global atmospheric reanalyses seek to assimilate a huge number of observations in order to represent the *true* state of the climate system as good as possible, they are used in a very wide range of applications. Some of these are, for example, Climate change [e.g., Frauenfeld, O. W. and Zhang, T. and Serreze, 2005, Santer et al., 2004, Trenberth, 2007] or Water budget studies [Trenberth and Fasullo, 2013, Trenberth et al., 2011, 2007]. Besides such investigations, reanalysis data is mandatory as forcing data for regional downscaling [Fersch et al., 2012, Qian et al., 2006]. The quality and reliability of such reanalysis models is thus of major importance for large parts of the geo-scientific community. The article *The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis* thus serves as a general introduction as it discusses very basic questions about the closure of the global water cycle of three of the most widely used datasets.

Chapter 3: Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances In the second article, a closer look at the catchment-scale water cycle is provided. According to Fekete et al. [2012], runoff is still the most accurately observed variable in the large-scale hydrological cycle. Therefore, the article discusses the performance of terrestrial and atmospheric-terrestrial water-balance-based methods to estimate monthly runoff. The study employs different data sources for precipitation, evapotranspiration, water storage changes, and moisture flux divergences from gridded observations (GPCC, GPCP, CPC, CRU, and DEL), atmospheric reanalysis models (ERA-Interim, CFSR, MERRA), partially-model-based datasets (GLEAM, MOD16, FLUXNET MTE) and GRACE. The datasets for the different hydrological variables are then combined in the terrestrial and atmospheric-terrestrial water balance equation, which finally allows the estimation of catchment-scale runoff. The ensemble of runoff estimates comprises, in total, 90 members. This ensemble is evaluated against monthly runoff observations from the Global Runoff Data Center (GRDC) over 96 catchments of different size and climatic conditions, worldwide. The analysis further evaluates the budget-based estimates against runoff estimates from different land surface models (GLDAS, MERRA Land) and a simple predictor based on the ratio between precipitation and runoff (P-R ratio). The performance of the estimates is analyzed in terms of the relative bias, correlation and Nash-Sutcliffe Efficiency with respect to the monthly runoff observations.

As a matter of fact, the differences between the estimates from the hydrological and hydrometeorological water-balance-based methods and the runoff observations are a direct measure of the water budget imbalance. Thus, the article is the logical continuation of the first publication as it discusses the water budget closure on catchment scales using a wide variety of state-of-the-art hydrological and hydrometeorological datasets. Furthermore, in the first article, an evaluation of the gridded observation-based datasets was not possible due to the lack of reliable reference data. Comparing a large ensemble of runoff estimates to observed data, as it is presented in the second publication, now gives the opportunity to analyze how all the combinations using a single dataset (e.g., precipitation from GPCC) perform. This finally allows to evaluate the uncertainty of each single water cycle variables over different catchments and the performance of each single input dataset.

Chapter 4: Estimating Runoff Using Hydro-Geodetic Approaches The third article is a more method-oriented paper, which focuses on five different approaches for estimating catchment-scale runoff. It serves as a companion article to the second paper as it puts the performance of the runoff estimates from the terrestrial- and atmospheric-terrestrial water budget and the *P-R* ratio into comparison with satellite-altimetry derived estimates and estimates from the relationship between runoff and water storage. The evaluation is carried out over six catchments (Amazon, Mississippi, Ob, Yenisei, Niger, and Danube). The analyzed methods can be differentiated into global runoff estimation methods, which are (in principle) able to provide runoff globally, and catchment-specific methods, which require at least an appropriate period of *in situ* runoff observations.

The main motivation for the second and third article was the necessity for alternative runoff estimation methods as the number of (traditionally) gauged catchments is significantly decreasing. It is thus evaluated if such estimates are able to provide realistic runoff values as these are mandatory for reasonable water budget studies.

Chapter 5: Basin-scale Runoff Prediction: An Ensemble Kalman Filter Framework based on Global Hydrometeorological Datasets The fourth article aims at combining all previously studied datasets and derived estimates in a mathematically consistent way. The proposed algorithm finally allows to provide reliable and, even more important, consistent estimates of precipitation, evapotranspiration, runoff, and water storage changes. Furthermore, it is described how water budget closure between the different data sources can be enforced and how such constraints affect the final estimates of the water cycle variables. The algorithm is based on an Ensemble Kalman Filter framework, where the different state variables (i.e., the water cycle variables) are predicted using a least squares prediction approach. This allows to exploit the joint temporal and spatial auto- and cross-covariance structures of precipitation, evapotranspiration, water storage changes, and runoff.

The article thus serves as a round-up of all previous studies as it considers all the advantages and disadvantages of our present hydrological and hydrometeorological data sources. Even if the algorithm theoretically allows to predict every single variable of the terrestrial water cycle, the article focuses on the prediction of runoff as the estimates can be directly validated against *in situ* observations from the GRDC. Therefore, the study focuses on 29 large catchments of different climate regions, with which runoff is predicted and evaluated for a subset of 16 river basins.

Chapter 2

Article I: The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis

Lorenz, C. and H. Kunstmann, 2012: The Hydrological Cycle in Three State-of-The-Art Reanalyses: Intercomparison and Performance Analysis, *Journal of Hydrometeorology*, **13**, 1397–1420, doi: 10.1175/JHM-D-11-088.1[1]

Abstract

The three state-of-the-art global atmospheric reanalysis models—namely, ECMWF Interim Re-Analysis (ERA-Interim), Modern-Era Retrospective Analysis for Research and Applications (MERRA; NASA), and Climate Forecast System Reanalysis (CFSR; NCEP)—are analyzed and compared with independent observations in the period between 1989 and 2006. Comparison of precipitation and temperature estimates from the three models with gridded observations reveals large differences between the reanalyses and also of the observation datasets. A major source of uncertainty in the observations is the spatial distribution and change of the number of gauges over time. In South America, active measuring stations were reduced from 4267 to 390.

---

[1] ©American Meteorological Society. Used with permission

The quality of precipitation estimates from the reanalyses strongly depends on the geographic location, as there are significant differences especially in tropical regions. The closure of the water cycle in the three reanalyses is analyzed by estimating long-term mean values for precipitation, evapotranspiration, surface runoff, and moisture flux divergence. Major shortcomings in the moisture budgets of the datasets are mainly due to inconsistencies of the net precipitation minus evaporation and evapotranspiration, respectively, $(P-E)$ estimates over the oceans and landmasses. This imbalance largely originates from the assimilation of radiance sounding data from the NOAA-15 satellite, which results in an unrealistic increase of oceanic $P-E$ in the MERRA and CFSR budgets. Overall, ERA-Interim shows both a comparatively reasonable closure of the terrestrial and atmospheric water balance and a reasonable agreement with the observation datasets. The limited performance of the three state-of-the-art reanalyses in reproducing the hydrological cycle, however, puts the use of these models for climate trend analyses and long-term water budget studies into question.

Introduction

Global and regional atmospheric retrospective analysis models (*reanalyses*) play a crucial role in today's hydrological and hydrometeorological research. These global atmospheric reanalyses aim at assimilating a large amount of historical observation data to provide a physically consistent basis for the most important hydrological, hydrometeorological, and atmospheric quantities. To bring these various observations into a consistent scheme, computation of the reanalysis models is performed via state-of-the-art data assimilation methods (3D-Var or 4D-Var) that constrain the observations with physically reasonable time evolution and budget equations. These reanalyses can be used to analyze the global climate system, atmosphere, and land surface processes on large to continental scales and to understand exchange processes between these different regimes. Global atmospheric reanalyses also are often used as forcing data for regional hydrological or hydrometeorological simulations, such as numerical weather predictions and regional climate simulations. Three of the most widely used reanalyses are the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim), the Modern-Era Retrospective Analysis for Research and Applications (MERRA) from the National Aeronautics and Space Administration (NASA), and the Climate Forecast System Reanalysis (CFSR) from the National Centers for Environmental Prediction (NCEP).

Reanalyses represent an approximation of the real world. Because of the changing amount of assimilated observational data, different data assimilation methods, and different model equations and assumptions, results of reanalysis models deviate significantly, even if they should be similar in principle. Therefore, it is necessary to validate these global atmospheric models with observational datasets.

Such comparisons were made by, for example, Janowiak et al. [1997], Poccard et al. [2000], or Higgins et al. [2010], with rainfall estimates from the CFSR and its predecessor, the NCEP/NCAR reanalysis, being validated against precipitation observations. Bosilovich et al. [2008] compared precipitation from the 40-yr ECMWF Re-Analysis (ERA-40), the two older NCEP reanalyses (which are often referred to as NR1 and NR2), and the Japanese 25-yr reanalysis (JRA-25) with data from the Global Precipitation Climatology Project (GPCP) and the widely used Climate Prediction Center Merged Analysis of Precipitation (CMAP) on both the continents and the oceans. In Hagemann et al. [2005], different quantities contributing to the global hydrological cycle of ERA-Interim's predecessor ERA-40 were analyzed in detail, while Chido and Haimberger [2009] or Mueller et al. [2010] investigated the closure of water and energy budgets in the ERA-Interim reanalysis. A more detailed comparison is given in, for example, Trenberth et al. [2007], where estimates of the most important quantities of the global water cycle are presented. On regional scales, Yeh and Famiglietti [2008] concentrated on the estimation of evapotranspiration. Considerations relating to the hydrological cycle over the United States were presented by Roads et al. [1994]. Seneviratne et al. [2004] analyzed the water budget closure over the Mississippi basin and presented estimates of monthly water storage variations based on water vapor flux convergence, atmospheric water vapor content, and river runoff. Similar work was performed by Betts et al. [1999, 2003, 2005, 2009], who analyzed energy and mass budgets of ERA-40 and ERA-Interim over

Continent	Area [$10^6\,km^2$]
North America	19.3
Europe	5.7
Asia	37.4
South America	17.8
Africa	30.0
Australia	7.7

FIGURE 2.1: Land-sea mask used for computing the spatial averages over North America, South America, Europe, Africa, Asia, and Australia. The table shows the areas of the regions considered within each continent.

several river basins (especially in North America). An assessment of the applicability of the ERA-40 model for the detection of climate trends was made by Bengtsson [2004].

In this study, the three state-of-the-art reanalyses ERA-Interim, MERRA, and CFSR are compared. The reanalyses are evaluated by comparing quantities—such as precipitation, temperature, and atmospheric water vapor—with observational datasets from the Global Precipitation Climatology Centre (GPCC), the GPCP, the Climate Prediction Center (CPC), the Climate Research Unit (CRU), the University of Delaware (DEL), and the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data (HOAPS). Differences in the total amount, spatial variability, and distribution of gauges of the gridded rainfall observations are analyzed in order to estimate the uncertainties incorporated in these datasets. Special emphasis is devoted to the comparison of precipitation estimates from the reanalyses because of their importance in the hydrological cycle.

In addition, the closure of the water budgets in the three reanalyses are analyzed and it will be estimated how well the transport processes between the oceans and the continents as well as moisture exchange between the land surface and the atmosphere are balanced. For this purpose, long-term mean values of precipitation, evapotranspiration, surface runoff, and atmospheric moisture flux divergences are computed. As evapotranspiration and surface runoff are the dominating quantities of moisture transport from the surface back into the atmosphere and oceans, respectively, the estimates from the reanalyses are used to investigate how well the water budgets in the models are closed.

Data and Methods

Reanalysis Data

For comparison, three different global atmospheric retrospective analyses are used—namely, ERA-Interim from ECMWF [Berrisford et al., 2009, Simmons et al., 2006], MERRA from the NASA Goddard Space Flight Center [GSFC; Rienecker et al., 2011], and CFSR from the National Oceanic and Atmospheric Administration (NOAA) National Centers for Environmental Prediction [NCEP; Saha et al., 2010]. The two latter reanalyses cover the satellite period from 1979 to the present, while ERA-Interim was intended to cover the period from 1989 to the present to provide a bridge between ECMWF's previous reanalysis ERA-40 [Uppala et al., 2005] and a forthcoming next-generation reanalysis. Recently, the ERA-Interim archive was extended to cover the years between 1979 and 1989 as well. The CFSR dataset succeeds the widely used NCEP-NCAR reanalysis [Kalnay et al., 1996]. The novelties of this reanalysis are the coupling to the ocean during the generation of the 6-h guess field, an interactive sea ice model, and the assimilation of satellite radiances for the entire period [Saha et al., 2010]. Furthermore, the analysis system used in CFSR for the atmosphere, the Gridpoint Statistical Interpolation (GSI) scheme, is nearly the same as the one used by MERRA at NASA GSFC. The MERRA atmosphere-only reanalysis is being conducted over the same years with nearly the same input data [Saha et al., 2010]. However, observation processing, model equations, and the main scope of the reanalyses differ significantly. The resulting differences in modeled variables thus reveal uncertainty ranges of present-day reanalysis models (see Table 2.1 for further details of these datasets).

TABLE 2.1: Summary of the three reanalyses.

	ERA-Interim	MERRA	CFSR
Hor. resolution	T255 ($\sim$ 78 km)	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	T382 ($\sim$ 38 km)
Vert. levels	60	72	64
Top level	0.1 hPa	0.01 hPa	0.26 hPa
Period	1979–present	1979–present	1979–2009
Output times	6h, daily, monthly	1h, 6h, daily, monthly	1h, 6h, monthly
Assimilation scheme	4D-Var	3D-Var	3D-Var

According to Kalnay et al. [1996] or Kistler et al. [2001], gridded variables from reanalyses can be separated into three classes, which vary by the influence of assimilated observations on the variable. The type A variables (e.g., upper-air temperatures or horizontal winds) are strongly influenced by the observations, and are thus assumed to be the most reliable variables. Type B variables (e.g., surface and 2-m temperatures) are influenced by both the observations and the model while type C variables (e.g., precipitation or surface runoff) are derived solely from the model.

Gridded Observation Data

In order to validate the three different reanalyses, we compare precipitation and temperature estimates from the reanalyses with gridded observations from GPCC [Rudolf and Schneider, 2005], GPCP [Adler et al., 2003], CRU [Mitchell and Jones, 2005], the Unified Gauge-based Analysis of Global Daily Precipitation from the Climate Prediction Center CPC [Chen et al., 2008], and DEL [Matsuura and Willmot, 2009]. For validation of the atmospheric water vapor over the oceans, data from the HOAPS product [Andersson et al., 2010] is used, which is based on satellite observations from the *Special Sensor Microwave Imager* SSM/I on satellites of the Defense Meteorological Satellites Program and provides reliable estimates of oceanic precipitation, evaporation, and other atmospheric variables.

The continental precipitation and temperature datasets contain at least daily (CPC) or monthly (GPCC, CRU, DEL) means at a spatial resolution of 0.5° × 0.5° for the whole world (see Table 2.2 for further details of the gridded observation products).

In principle, the different datasets should provide similar precipitation and temperature values. Differences of global fields must be considered as uncertainty ranges, which can be expected when using such datasets for validation purposes. To generate gridded observations from *in situ* measurements, the different data centers apply similar interpolation algorithms and may therefore exhibit similar biases (particularly in areas with complex terrain).

Two main error sources lead to uncertainties in precipitation observations. The sampling error, which is due to the irregular distribution of gauges, has a magnitude of about ±7 %–40 % of the true area-mean precipitation [Schneider et al., 2008]. Rudolf and Rubel [2005] report that sampling errors between 15 % and 100 % can be expected for sparsely gauged regions (less than 3 gauges per 2.5° × 2.5° grid cell). The second error is due to the under-catch of precipitation gauges which results from wind-field deformation above the gauge orifice, losses from wetting on internal walls of the collector and in the container, and losses due to evaporation from the container [Rudolf and Rubel, 2005]. The gauge under-catch error might be large especially during winter in the high-latitude regions or over mountain ranges, as there will be a high amount of solid precipitation. This leads to an underestimation of the true precipitation of up to 50 %. Since 2007, GPCC has been providing event-based correction factors [Fuchs et al., 2001, Schneider et al., 2008] to account for the systematic gauge under-catch error. Before 2007, the corrections consisted of monthly climatologies as proposed in Legates and Willmott [1990], which are still applied to the GPCP precipitation product. The original GPCC full data product used for this study does not include such corrections (Andreas Becker 2011, personal communication).

In the course of this study, the GPCC precipitation product was updated from version 4.0 to 5.0. Even though the new dataset is based on a denser station network, the differences of area-averaged values or long-term mean fields are not significant (not shown here). Therefore, the GPCC-v4.0 product was used for reference observations in this study, but the differences in the distribution and total number of gauges between version 4.0 and 5.0 are discussed briefly (see section 3a).

TABLE 2.2: Summary of the observation datasets containing precipitation P, near surface temperature $T2$, and the atmospheric water vapor content W.

	Variables	Hor. resolution	Period	Output times	Version number
GPCC	P	0.5° × 0.5°	1901–2009	monthly	4.0
GPCP	P	2.5° × 2.5°	1979–2009	monthly	2.1
CRU	$P, T2$	0.5° × 0.5°	1901–2009	monthly	3.0
CPC	P	0.5° × 0.5°	1979–present	daily	1.0
DEL	$P, T2$	0.5° × 0.5°	1900–2008	monthly	2.01
HOAPS	P, E, W	0.5° × 0.5°	1987–2005	daily, 5-daily, monthly	3

Area-Averaging of Gridded Data

For the validation of the reanalyses' rainfall estimates with the observation datasets, all fields were remapped to the resolution of the GPCC dataset (i.e., 0.5°×0.5°) using a first-order conservative interpolation [Jones, 1999]. From these fields, area-weighted averages were computed over different regions using the continental mask shown in Figure 2.1. As GPCC only contains gauge-based observations, the oceans or the poles were not considered for comparison of the precipitation fields. Consequently, the global and hemispheric averages do only represent the rainfall over land. For investigating the water budget closure, a correct differentiation between the processes over land and the oceans is crucial. We did not perform any additional interpolation for this analysis, but used the fields in the models' native resolutions. The area-averaged values over the continents and oceans were calculated using the land-sea masks from the three reanalyses. For the evaluation of the oceanic water cycle components, we used a dynamic land-sea mask, as the satellite observations from HOAPS are available over ice-free ocean only.

The Global Water Balance

The terrestrial large-scale water balance (mm month^{-1}) can be written as

$$\frac{dS}{dt} = P - E - R, \tag{2.1}$$

where dS/dt is the change in the terrestrial water storage, P is precipitation, E is evapotranspiration, and R is total runoff—that is, the sum of surface runoff and subsurface runoff [Peixoto and Oort, 1992, Willmott, 1985]. According to Yeh and Famiglietti [2008], the change in the total terrestrial water storage depends on its surface component, the soil moisture, and the groundwater components. The total soil water depends largely on the characteristics of the land surface model in the reanalyses. ERA-Interim and CFSR provide soil moisture values divided into multiple levels. MERRA does not consider a multi-level soil model for computing the interaction between the land surface and the atmosphere. dS/dt can thus be computed as a residual term like in, for example, Roads et al. [1994]. When analyzing water budgets in atmospheric reanalysis models, it is convenient to further consider a term for the analysis increment—that is, the increment that is due to the forcing of the models towards the observations. The terrestrial water balance equation is then modified to

$$\frac{dS}{dt} = P - E - R + RES_s', \tag{2.2}$$

where RES_s' is accounting for surface residual water forcing. As further proposed in Roads et al. [2002] and Szeto et al. [2008], the water storage tendency term is combined with residual forcing—that is, $RES_s = RES_s' - dS/dt$. According to Kleidon and Schymanski [2008] or Seneviratne et al. [2004], it can be assumed that for climatic time scales the total soil water content remains constant and its tendency can be neglected. This assumption results in the simplified terrestrial water balance equation of

$$\overline{R} \approx \overline{P} - \overline{E}, \tag{2.3}$$

where overbars indicate averaging over a climatic time scale. The imbalance of this equation provides information on the magnitude of the errors introduced by surface water forcing RES_s'.

Apart from the continental water budget, exchange of water between the continents and oceans is balanced as well. Over multi-year averages, the global water budget should be closed—that is, the convergence $(P - E)$ of moisture over land should equal the divergence $(E - P)$ of moisture over the ocean [Hagemann et al., 2005];

that is,

$$\overline{P}_{\text{land}} - \overline{E}_{\text{land}} = \overline{E}_{\text{ocean}} - \overline{P}_{\text{ocean}} . \tag{2.4}$$

In general, this value must be positive over the continents due to a surplus of precipitation, while there is more evaporation over the oceans.

These terrestrial budgets can be linked to the atmosphere by the atmospheric water balance equation

$$\frac{dW}{dt} + \mathbf{\nabla} \cdot \mathbf{Q} = E - P , \tag{2.5}$$

where E and P are actual gridpoint evapotranspiration and precipitation at the surface [Roads et al., 1994]. W denotes the total column water content in the atmosphere and $\mathbf{\nabla} \cdot \mathbf{Q}$ is the net balance of moisture flux (i.e. moisture flux divergence), which is defined as

$$\mathbf{\nabla} \cdot \mathbf{Q} \equiv \frac{1}{g} \mathbf{\nabla} \cdot \int_{p=0}^{p_{sfc}} \boldsymbol{v}_h(p)\, q(p)\, dp \tag{2.6}$$

with air pressure p (Pa), the gravitational acceleration g (m s^{-2}), the horizontal wind vector $\boldsymbol{v}_h$ (m s^{-1}), and the specific humidity q (kg kg^{-1}). When computed from reanalyses, moisture flux divergences are based on *type A* and *type B* variables only, while precipitation and evaporation are both *type C* variables. Again, it is convenient to add a term accounting for the analysis increment to the atmospheric water balance equation:

$$\frac{dW}{dt} + \mathbf{\nabla} \cdot \mathbf{Q} = E - P + RES'_w . \tag{2.7}$$

The atmospheric tendency term dW/dt can be combined with the residual forcing; that is, $RES_w = RES'_w - dW/dt$ [Roads et al., 2002, Szeto et al., 2008]. On annual or longer time scales, the variations of the atmospheric water storage W are often assumed to be negligible [Peixoto and Oort, 1992]. For monthly time scales, this assumption does not hold, however. The vertically integrated moisture flux divergences are directly linked to the vertical exchange terms of the terrestrial water balance:

$$\mathbf{\nabla} \cdot \mathbf{Q} = \overline{E} - \overline{P} . \tag{2.8}$$

As in case of the surface water balance, misclosure of the equation when using long-term averages is an estimate of the magnitude of the analysis increment of the atmospheric water forcing RES_w'. Equation (2.1) can be combined with (2.5) to obtain another linked balance equation:

$$\frac{dS}{dt} = -\frac{dW}{dt} - \nabla \cdot \mathbf{Q} - R. \tag{2.9}$$

As the atmospheric and terrestrial tendency terms can be assumed to be negligible over longer time scales, the atmospheric net input of moisture in a certain area must be balanced by a terrestrial net outflow at the surface; that is, $\overline{R} \approx -\nabla \cdot \mathbf{Q}$ [Roads et al., 1994]. The imbalance of this equation is a rough estimate of the total atmospheric and surface water analysis increments.

Computation of Spatial Correlations

To analyze the agreement of spatial patterns between two datasets, spatial correlations are computed for further analysis. This yields information about the extent to which certain events (e.g., large-scale rainfall) agree in terms of location, dimension, and magnitude when using various datasets. We compute the spatial correlations between two datasets x and y according to

$$\mathrm{corr}_\chi = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_\chi - 1}\sum_{i=1}^{n_\chi}\frac{\left(x_{i,t,\chi} - \overline{x}_{t,\chi}\right)\left(y_{i,t,\chi} - \overline{y}_{t,\chi}\right)}{\sigma_{x,t,\chi}\sigma_{y,t,\chi}}, \tag{2.10}$$

where n is the number of gridpoints of a given area χ, $x_{i,t,\chi}$ and $y_{i,t,\chi}$ are the actual gridpoint values at the time t, the overbar denotes the spatial mean value of the area, and $\sigma_{x,t,\chi}$ and $\sigma_{y,t,\chi}$ are the standard deviations of the two datasets x and y of the area χ at the time t. T is the number of time steps contributing to a temporal subset like, for example, all Januaries of the considered time series or all months of a specific year.

Apart from time series of spatial correlations, we use the Taylor diagrams [Taylor, 2001] to analyze the level of agreement of rainfall patterns from different data sources. In this case, the standard deviation (the radial distance of a data point from the origin) is a measure of the intensity and variability of the patterns, while

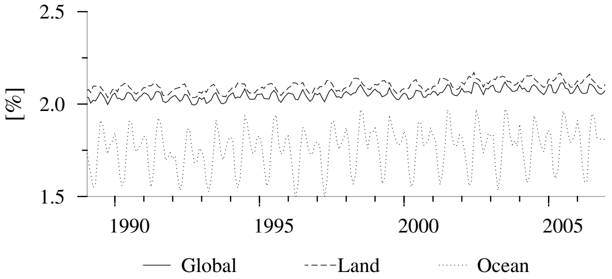

FIGURE 2.2: Relative difference between CFSR evaporation with and without considering changes in the near-surface temperature. The impact is generally higher over the continents, while over the oceans, the amplitude of the intra-annual variations is larger.

the correlations (the angle between the x-axis and a data point) reflect how well the analyzed datasets reproduce the rainfall patterns from a reference dataset. The root-mean-square difference (radial distance between the reference data point and another data point) is a measure of the average pixel-wise differences between two datasets and computed from the standard deviations and the correlations.

Computation of CFSR Evapotranspiration

In contrast to ERA-Interim and MERRA, CFSR does not provide fields of total evapotranspiration. Therefore, these fields are computed from latent heat flux, which is given in energy flux form (i.e., in units of W m^{-2}). The transformation into units of mm (i.e., mass flux form) was performed via

$$E = \frac{\lambda E}{L_e}, \tag{2.11}$$

where E is evapotranspiration (mm), λE is the latent heat flux (W m^{-2}), and L_e is the latent heat of evaporation (J kg^{-1}), which can be approximated through

$$L_e \approx 2.501 \cdot 10^6 - 2370 \cdot T_c \tag{2.12}$$

with T_c being the near-surface temperature in degrees Celsius [e.g., Jacobson, 2007]. The latent heat flux fields from CFSR include both evaporative flux from liquid and snow sublimation from snow surface. Equation (2.11) must consequently be corrected for sublimation:

$$E = \frac{\lambda E - \lambda S}{L_e} + \frac{\lambda S}{L_s}, \tag{2.13}$$

where λS is sublimation and L_s is the latent heat of sublimation which is the sum of the latent heats of evaporation and melting [e.g., Jacobson, 2007]:

$$L_s \approx 2.501 \cdot 10^6 + 3.35 \cdot 10^5 - T_c \cdot (340 + 10.46 \cdot T_c) \, . \tag{2.14}$$

According to CFSR, the influence of the temperature can be neglected (R. Yang 2011, personal communication). Our computations support this assumption. On monthly time scales, the temperature causes an increase of the continental evaporation of about 2.1 % with a maximum during the summer months (Figure 2.2). Over the oceans, the impact is smaller (about 1.7 %), but has a strong semi-annual signal with its maxima in the summer and winter months. By considering the magnitudes of other major water cycle components, the functional dependencies of L_e and L_c on temperatures are found to exert little impact on the calculations of evaporation and evapotranspiration, respectively, from latent heat flux. Thus, the corrections for temperature can be neglected on these spatial and temporal scales. For small-scale studies and especially during summer months, however, the influence might be significant and should be taken into account.

Results

Distribution of Gauges in the Observation Data Sets

To validate the global atmospheric reanalyses, rainfall observations, interpolated to a regular grid as described in, for example, Chen et al. [2008] and Rudolf and Schneider [2005], are used. As a matter of fact, the quality of these gridded precipitation fields depends primarily on the number of active gauges and their spatial distribution. The interpretation of interpolated gridded observations in regions with a few gauges only or a disadvantageous spatial distribution of such observation stations remains open.

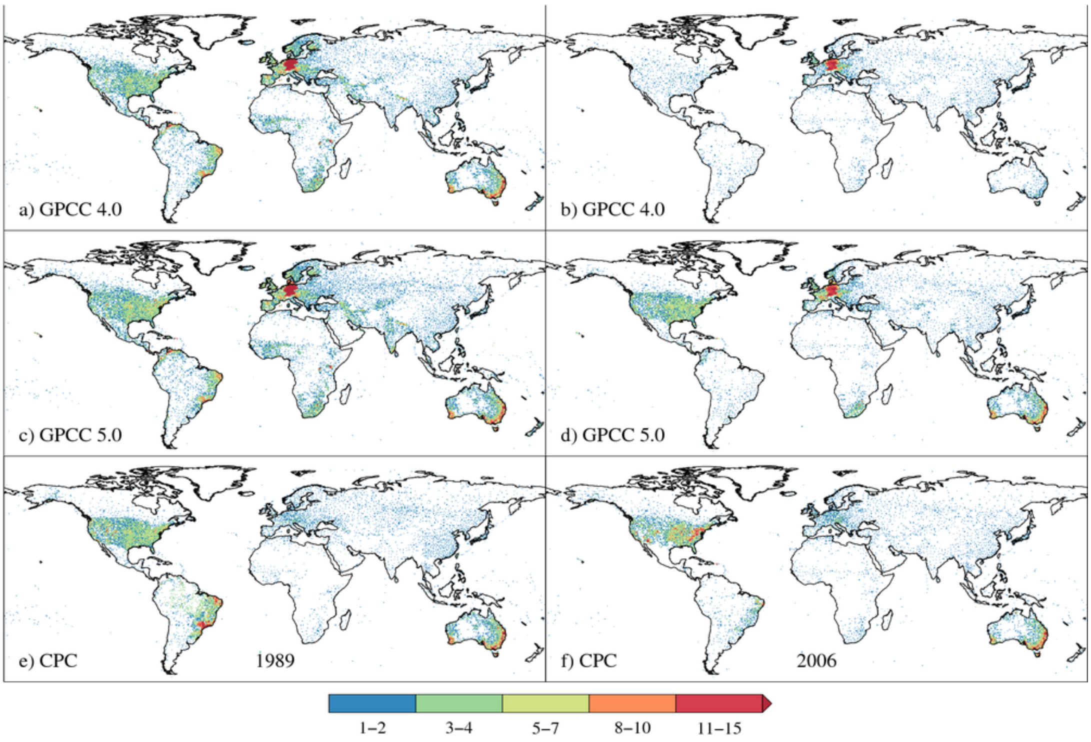

FIGURE 2.3: Number of monitoring stations per 0.5° × 0.5° grid cell in (a),(c),(e) January 1989 and (b),(d),(f) December 2006 for the (a),(b) GPCC-v4.0; (c),(d) GPCC-v5.0; and (e),(f) CPC datasets. A good spatial coverage with observation stations can be observed over North America (GPCC-v5.0, CPC) and Europe (GPCC-v4.0, GPCC-v5.0), while the number of gauges over North America is significantly reduced in GPCC-v4.0. Over most of the tropical regions like the Congo or Amazon basin, high-latitude regions, and large parts of Asia, the three datasets use a maximum of 1–2 gauges per grid cell, whereas some areas are completely ungauged.

Figure 2.3 shows the number of gauges per grid cell at the beginning and the end of the considered time series for GPCC-v4.0, GPCC-v5.0, and CPC. In 1989 (Figures 2.3a,c,e), a dense network of observation stations existed over North America, central Europe, coastal regions of Australia, and the eastern part of Brazil. The GPCC products also exhibit a good spatial coverage of South Africa, while only few gauges are located in the rain-laden regions of tropical Africa, South America, and Southeast Asia and large parts of the subtropics, Eurasia, and high-latitude regions. Depending on the geographic location of these ungauged regions, an interpolation might introduce large uncertainties. This is particularly true if the complex cycle of tropical precipitation or the high spatial variability of rainfall over mountain ranges is considered. Figures 2.3b,d,f show the amount of gauges per grid cell in December 2006. Spatial coverage with observation stations has changed drastically, especially

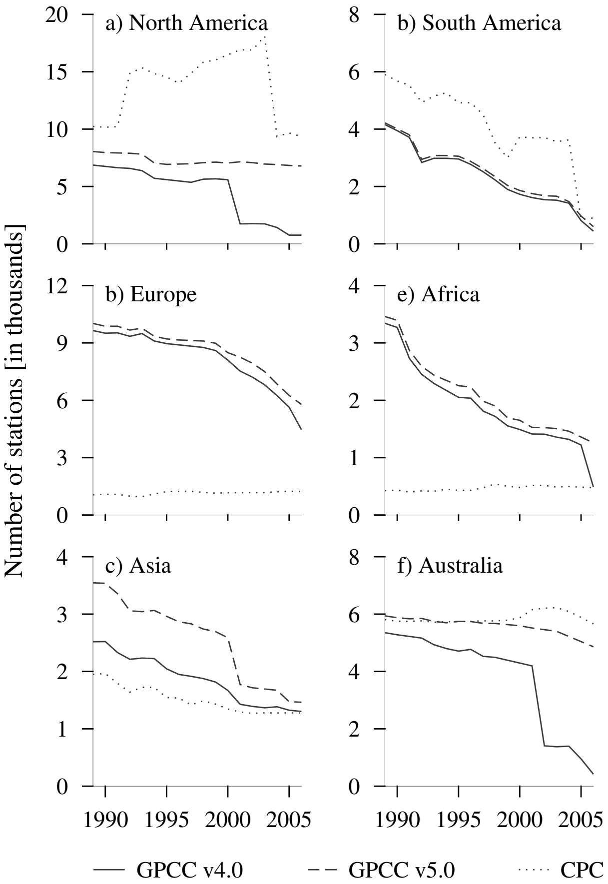

FIGURE 2.4: Mean annual number of gauges used in the precipitation observations from GPCC-v4.0, GPCC-v5.0, and CPC. Version 4.0 shows a significant drop in the number of gauges between 2000 and 2001 over North America and Australia, while version 5.0 of the GPCC product is based on a nearly constant number of observation stations during the complete time series. Over South America, Europe, and Africa, the update from v4.0 to v5.0 results in little improvement only, as both versions show a nearly constant decline over time. Over Asia, GPCC-v5.0 is using about 1,000 gauges more than version 4.0 until 2000. The CPC product is based on about 1,000 gauges over Europe and 500 gauges over whole Africa, while more than 14,000 gauges are used to generate the gridded precipitation observations over North America between 1991 and 2003.

for the GPCC-v4.0 data in North America, South America, and Africa. Large parts of the tropics and deserts remain completely ungauged over hundreds of kilometers in both GPCC and CPC datasets. The update from version 4.0 to 5.0 of the GPCC product significantly improved spatial coverage of North America and Australia, while there is only little improvement over South America, central Africa or large parts of Eurasia. As is obvious from Figure 2.4, the number of gauges decreased significantly for all three observation datasets over most of the regions. At the end of the period studied, only 1314 (CPC), 390 (GPCC-v4.0), and 555 (GPCC-v5.0) gauges remain, which are used to compute the precipitation fields over South America. Although the decrease in active gauges is not that significant over the Asian continent, comparison of the numbers of gauges over Europe and Asia again illustrates the very sparse distribution of gauges in the latter regions. In contrast to this, the CPC dataset is based on about $10,000$ gauges over North America in the beginning and the end of the time period, while in between, the number of gauges increases up to $\sim 17,000$ in 2003. This shows that there are certain regions where the gridded GPCC and CPC products are based on a dense network of gauges and, hence, provide a scientifically sound basis to validate modeled precipitation fields. On the other hand, the reliability of the observation datasets remains questionable especially over the tropics, deserts, mountain ranges, and large parts of the Asian continent due to the decreasing number of active gauges and their sparse spatial distribution.

Precipitation

Long-term Mean Annual Precipitation

The long-term mean annual rainfalls (Figure 2.5) obtained from the different reanalyses are in general agreement with the observational references when looking at the spatial precipitation patterns. The large-scale rain-laden regions of the tropics in South and Central America, central Africa, and Southeast Asia show precipitation rates of up to $11\,\mathrm{mm\ day}^{-1}$ in all products. These moist regions are clearly separated from the large subtropic desert regions with very limited precipitation. In addition, good agreement in the precipitation patterns can be found, for example, in Australia and over the moist regions in the southeastern part of North America and the drier Great Plains. Large mountain ranges, especially the Andes, the Alps, and the Himalayas, can be identified because of their wet conditions compared to the surrounding regions. Except for MERRA, all datasets show a maximum in annual precipitation at the headwaters of the Amazon River, which extends along the course of the river down to the Atlantic. In the MERRA dataset, this maximum is shifted eastward. In the regions between southern Brazil and the southern foothills of the

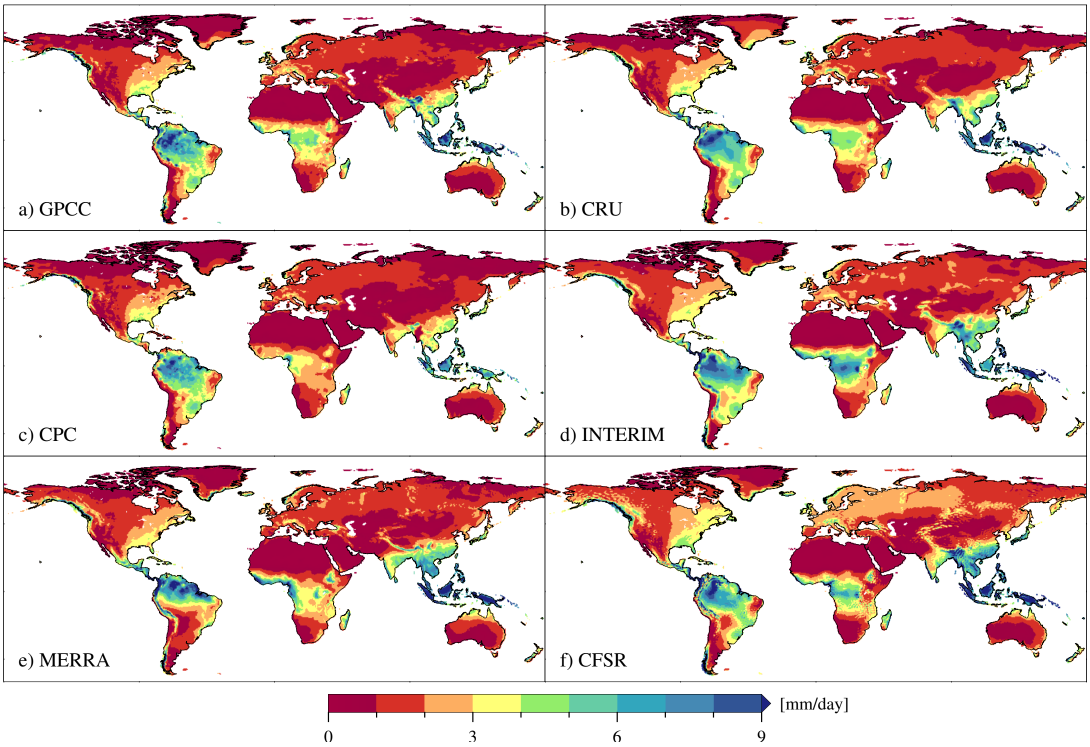

FIGURE 2.5: Long-term mean annual precipitation between 1989–2006 in [mm day^{-1}]. The three observation datasets GPCC (a), CRU (b), and CPC (c) are in good agreement over most of the regions, even if CPC assumes less rainfall over Central Africa. The precipitation estimates from the three reanalyses ERA-Interim (d), MERRA (e), and CFSR (f) show similar large-scale patterns, while significant differences exist in the spatial distribution and the amount of rainfall on smaller scales.

Andes, GPCC, CRU, CPC, ERA-Interim, and CFSR show a mean precipitation rate of about 4–5 mm day^{-1}, contrary to a distinguished dry region with less than 2 mm day^{-1} predicted by MERRA.

In general, the highest differences in spatial variability and the amount of rainfall can be found over tropical South America, central Africa, Southeast Asia, and the large mountain ranges of the Andes and the Himalayas. These differences cause large-scale deviation patterns, which can reach magnitudes of up to ±4 mm day^{-1} (Figure 2.6). Even when focusing on the ensemble of the observation datasets GPCC, GPCP, CRU, DEL, and CPC, differences of up to 3 mm day^{-1} result in central Africa (Figure 2.7a), which are likely introduced by uncertainties in the observations due to the sparse distribution of gauges in these regions. In the three reanalyses, the differences in spatial variability and the amount of precipitation are even larger compared to

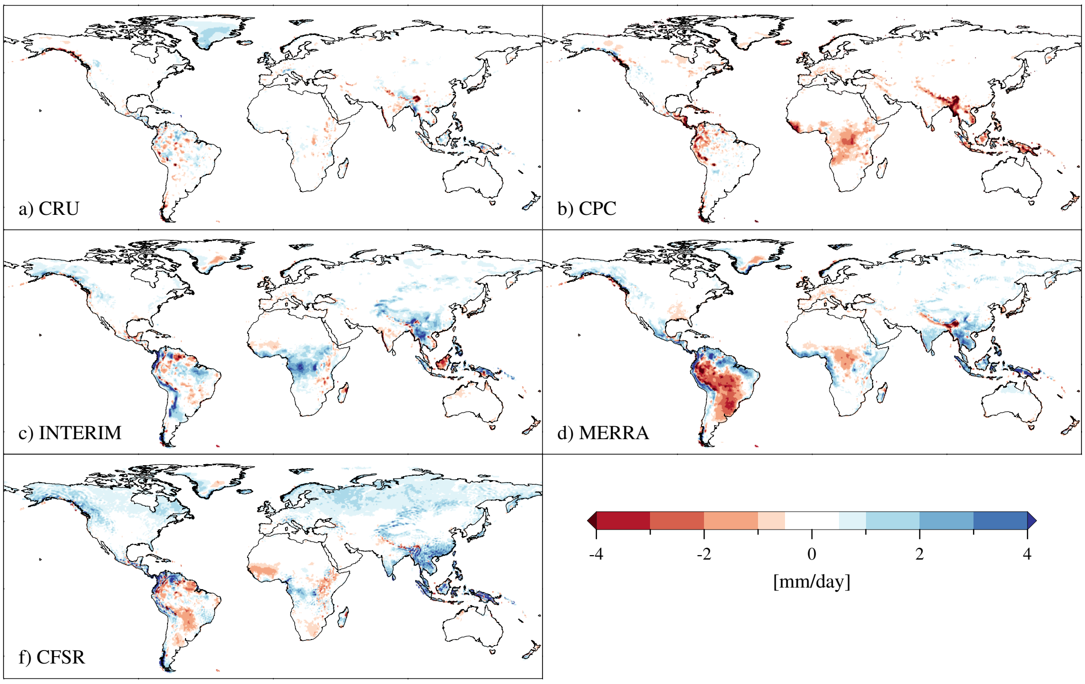

FIGURE 2.6: Absolute differences of the mean annual precipitation from 1989–2006 between GPCC and CRU (a), CPC (b), ERA-Interim (c), MERRA (d), and CFSR (e) in [mm day^{-1}]. CRU shows a good agreement with GPCC. CPC is drier over the Congo Basin, the Himalayas, and the northern part of the Andes. The largest differences between GPCC and the three reanalyses can be observed over the tropics and the mountain ranges. CFSR also has a wet bias over mid-to-high northern latitudes, while MERRA shows a dry pattern which extends over large parts of South America.

the observations. The mid- to high-latitude rainfall estimates by CFSR appear to be significantly biased, as there are deviations of up to 2 mm day^{-1} (Figure 2.6e). Higgins et al. [2010] investigated the reliability of CFSR precipitation over North America and concluded that parts of this bias can be explained by an overactive diurnal cycle in the atmospheric component of CFSR. The observation datasets are based on a dense network of gauges and show only small deviations in these regions. It can thus be assumed that there are some inaccuracies in the CFSR estimates.

A significant discrepancy exists between the precipitation patterns from GPCC, ERA-Interim, and MERRA in South America and central Africa (Figures 2.6c,d). These differences were also noted in Trenberth et al. [2011]. ERA-Interim overestimates rainfall over the Andes and central Africa up to 2.5 mm day^{-1}, while MERRA shows a large-scale underestimation in central South America and central Africa and an overestimation over coastal regions. It is well known that tropical precipitation in the MERRA reanalysis over South America has its shortcomings. Therefore, a corrected dataset for land hydrology will be released in the near future [Reichle et al., 2011]. CFSR indicates conditions that are too moist in the center and underestimates rainfall east of the Congo basin along the course of the Nile (Figure 2.6e). Poccard et al. [2000] and Sylla et al. [2010] discuss that rainfall simulations in these regions is a very complex task and might lead to large discrepancies. As the largest part of precipitable water in central Africa arises from evapotranspirating water in the tropical rain forests [Van der Ent et al., 2010], these deviations might be due to shortcomings in the models' land–atmosphere interactions in this complex environment. On the other hand, the number of active gauges (Figure 2.3d) shows that their spatial density decreased significantly during the period considered. This means that uncertainty is up to $\pm 3\,\mathrm{mm\ day^{-1}}$ in these regions because of the variability of the ensemble of observations (Figure 2.7a). Only the ERA-Interim precipitation exceeds the uncertainty given by the observations over a large area. The other two reanalyses are within the bounds given by the observations and are therefore assumed to be more realistic.

Precipitation over the Andes is generally overestimated in the reanalyses, while all datasets show less rainfall over the Himalayas than to GPCC. This might be due to the impact of orography on convective events caused by the differences in resolution and the description of the underlying terrain model. On the other hand, the high spatial variability of precipitation in mountain ranges aggravates reliable areawide observations. Because of the sparse distribution of gauges and the errors caused by the undercatch of solid precipitation, the quality of interpolated rainfall values from GPCC, CPC, and CRU remains questionable in these regions.

Time Evolution of Global, Hemispheric, and Tropic Precipitation

The global and Northern Hemispheric correlations and differences (Figures 2.8a,b and 2.9a,b) of the four observation datasets and the ERA-Interim and MERRA reanalyses are relatively constant over time. Although the number of gauges used for generating the observation datasets and the amount of observations assimilated in the reanalyses

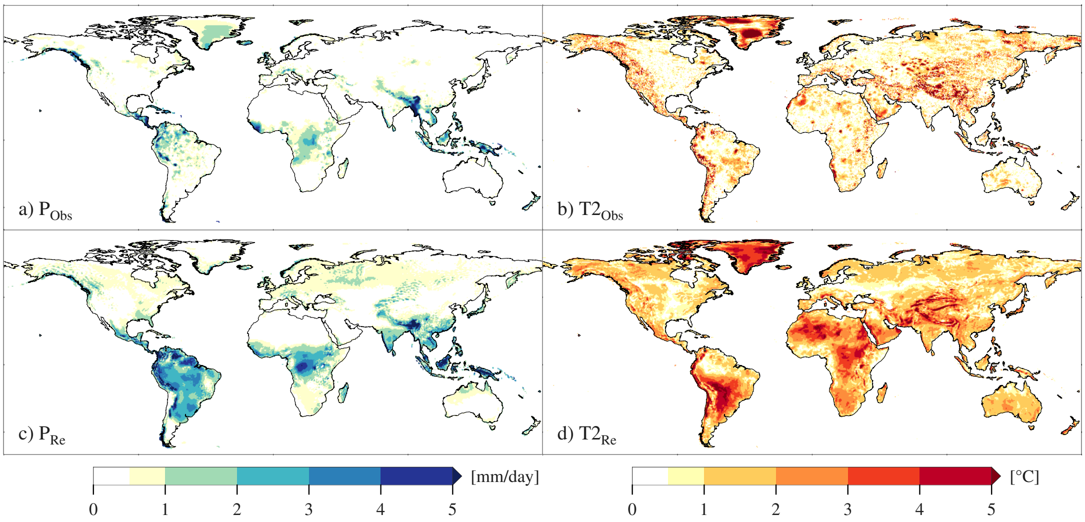

FIGURE 2.7: (a),(c) Variability of rainfall P [mm day^{-1}] and (b),(d) range of temperature $T2$ [°C] of (a),(b) the ensemble of gridded observations and (c),(d) the ensemble of reanalyses. The ensemble of rainfall observations is generated from GPCC, GPCP, DEL, CRU, and CPC, while the temperature range is based on DEL and CRU. The reanalysis ensemble consists of ERA-Interim, MERRA, and CFSR for both precipitation and temperature. The reanalyses generally produce a larger variability especially over the tropics and the whole of South America. Over the Congo basin, however, the precipitation variability of the observation ensemble reaches values of up to 3 mm day^{-1}. The temperature range from the three reanalyses shows the largest values over South America, the Congo Basin, the Sahara, and Greenland, where differences of more than 5 °C can be observed.

changed significantly between 1989 and 2006, there is only a minor impact on the agreement with GPCC on these scales. Over the Southern Hemisphere and the tropics, the spatial correlations (Figures 2.8c,d) exhibit a wider range between the datasets. It is difficult to determine, however, whether this range is due to the reduction of gauges or changes in the assimilated observations.

The CFSR series show a general wet bias over the Northern Hemisphere and a drop in the global continental rainfall toward GPCC from 1998 (Figures 2.9a,b). Interestingly, the differences between CFSR and GPCC increase again after 2000 over both the tropical regions between 15 °N and 15 °S and the Southern Hemisphere (Figures 2.9c,d), while the bias over the Northern Hemisphere is slightly reduced. The decline between 1998 and 2000 is also evident from the ERA-Interim rainfall over the tropics and the Southern Hemisphere but not in MERRA. During these years, all three reanalyses and the CRU observations show a sudden increase in the

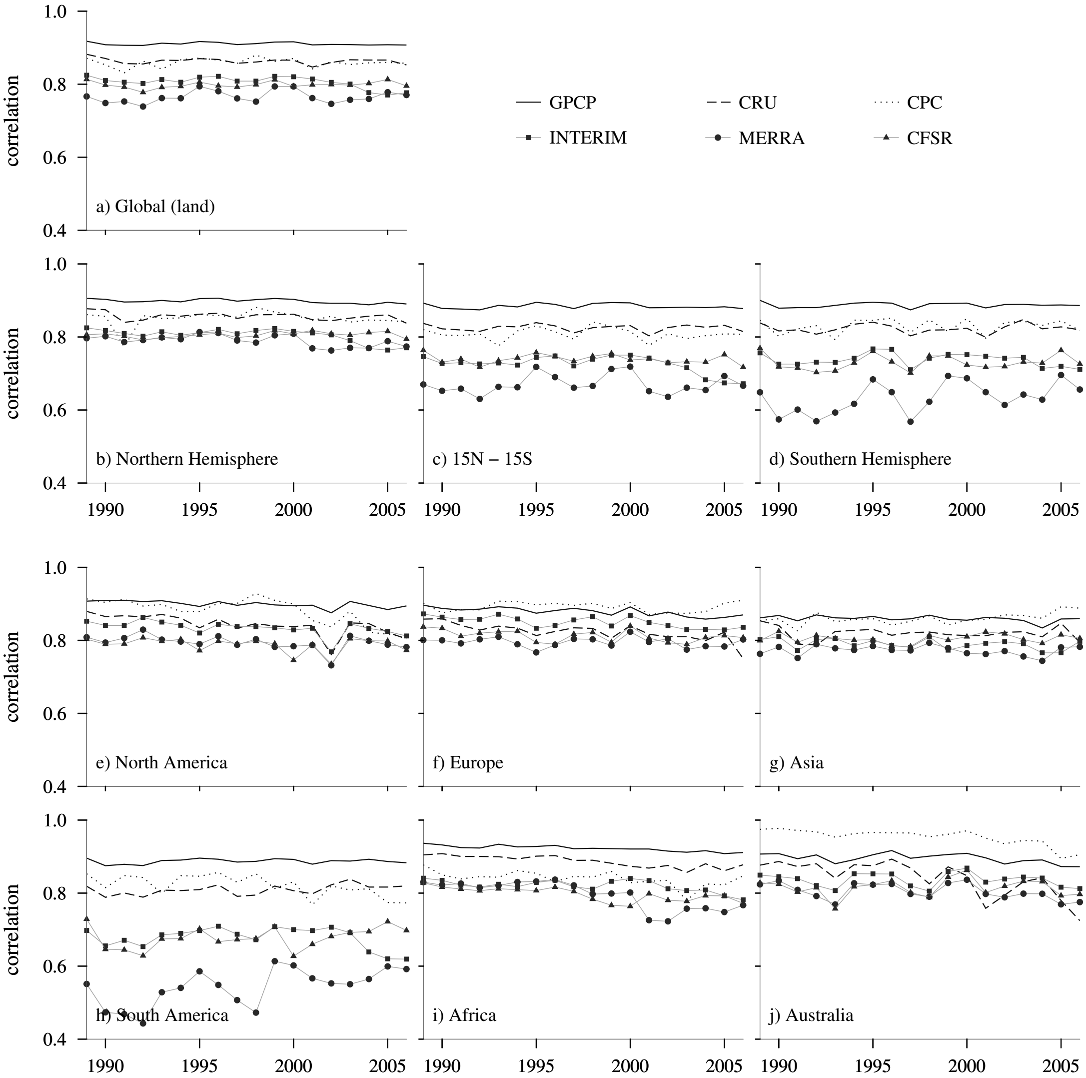

FIGURE 2.8: Area-averaged spatial correlations in mean annual precipitation of GPCP, CRU, CPC, ERA-Interim, MERRA, and CFSR in relation to GPCC. Values close to 1 indicate that the precipitation patterns from the respective dataset are in good agreement with the spatial distribution of rainfall from GPCC. In most of the regions, all datasets reproduce the spatial rainfall patterns from GPCC with a correlation coefficient > 0.7 between 1989 and 2006. The largest deviations can be observed over South America, where especially MERRA shows correlation coefficients < 0.6 (until 1998) and ∼ 0.6 (from 1998). Compared to the reanalyses, the agreement between GPCC and the observation datasets is generally better. Over most of the regions, ERA-Interim shows the highest correlation coefficients in relation to GPCC.

near-surface temperature (not shown), which might be related to the gaps in the precipitation estimates.

The CPC observations show a dry bias of about $-0.3\,\mathrm{mm\,day^{-1}}$ over the Northern and $-0.4\,\mathrm{mm\,day^{-1}}$ over the Southern Hemisphere in relation to GPCC (Figures 2.9b,d). In general, precipitation over the Southern Hemisphere and the tropics has a higher

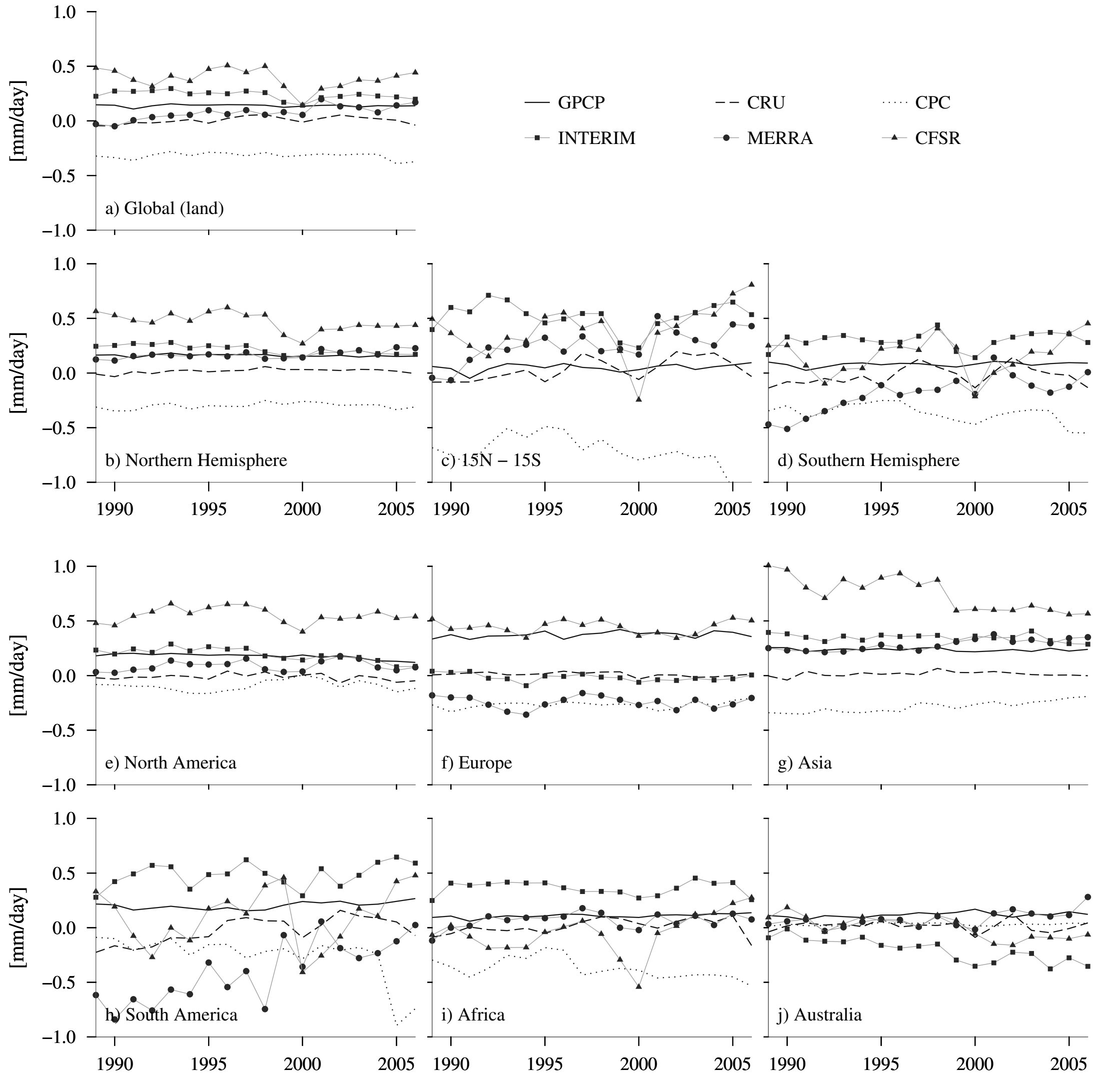

FIGURE 2.9: Area-averaged differences in the mean annual precipitation of GPCP, CRU, CPC, ERA-Interim, MERRA, and CFSR in relation to GPCC in [mm day^{-1}]. On a global scale, MERRA agrees best with GPCC, but shows significant deviations especially over the southern hemisphere (likely due to the large differences over South America). Good agreement between GPCC and MERRA can also be observed over North America, while over Europe, ERA-Interim performs best. Over Australia, MERRA and CFSR are of comparable agreement with the observations, but the differences between GPCC and ERA-Interim increase towards the end of the time series. In general, the observations from CPC show large deviations from GPCC over South America and Africa due to CPC's drier conditions in these regions. GPCP has a slight wet bias especially over the northern hemispheric regions (North America, Europe, Asia), which might be due to corrections for the gauge under-catch error.

variability in spatial correlations and deviations relating to GPCC. In contrast to the reanalyses, the observation datasets are in good agreement, even if the deviations from GPCC are larger because of the higher precipitation rates in these regions. Spatial variability of rainfall observations over the Southern Hemisphere correlates with a correlation coefficient of at least 0.8 and, hence, is a reliable basis for validating the reanalyses on large spatial scales. It can also be noticed that the reanalyses' spatial correlations over the Southern Hemisphere are dominated by the variations between 15 °N and 15 °S. Especially after 1995, the MERRA dataset depicts a quasi-periodic signal in the spatial correlations over the Southern Hemisphere and the tropics (Figures 2.8c,d).

The global intra-annual differences between the CFSR rainfall estimates and the GPCC observations have an annual cycle with maximal deviations in the period from March to June (Figure 2.10a). ERA-Interim tends to slightly overestimate the GPCC rainfall over the Northern Hemisphere with largest deviations of about 0.3 mm/day occurring from March to May. The tropical and Southern Hemispheric ERA-Interim precipitation rates are higher throughout the year with a distinct peak during the period from September to December where deviations from GPCC of up to 0.75 mm/day can be observed. Thus, on the global scale, ERA-Interim assumes slightly higher precipitation rates than GPCC, with the largest differences occurring in the periods from March to May and from September to December—that is, in the Northern and Southern Hemispheric spring months (Figures 2.10a–d). MERRA does not exhibit a clear annual cycle over the Northern or Southern Hemisphere; deviations in the tropics are maximal during the period from November to April. The intra-annual spatial correlations between MERRA and GPCC show a clear annual cycle especially over the Southern Hemisphere, which is mainly dominated by variations between 15 °N and 15 °S (Figures 2.11c,d). CFSR and ERA-Interim show a similar annual cycle with a generally higher correlation except for the period from September to November over the Southern Hemisphere, where ERA-Interim agrees better with GPCC than CFSR. Over the Northern Hemisphere, the reanalyses are in good agreement with an average correlation coefficient of about 0.8 (Figure 2.11b).

The Taylor diagrams in Figure 2.12 show that on a global scale, all three reanalyses and the observations reproduce the spatial rainfall patterns from GPCC with a correlation coefficient of > 0.7. The statistics over the Northern or Southern Hemisphere and the tropics indicate that the level of agreement between GPCC and the other datasets decreases when the area of interest is reduced. It is also evident that CFSR

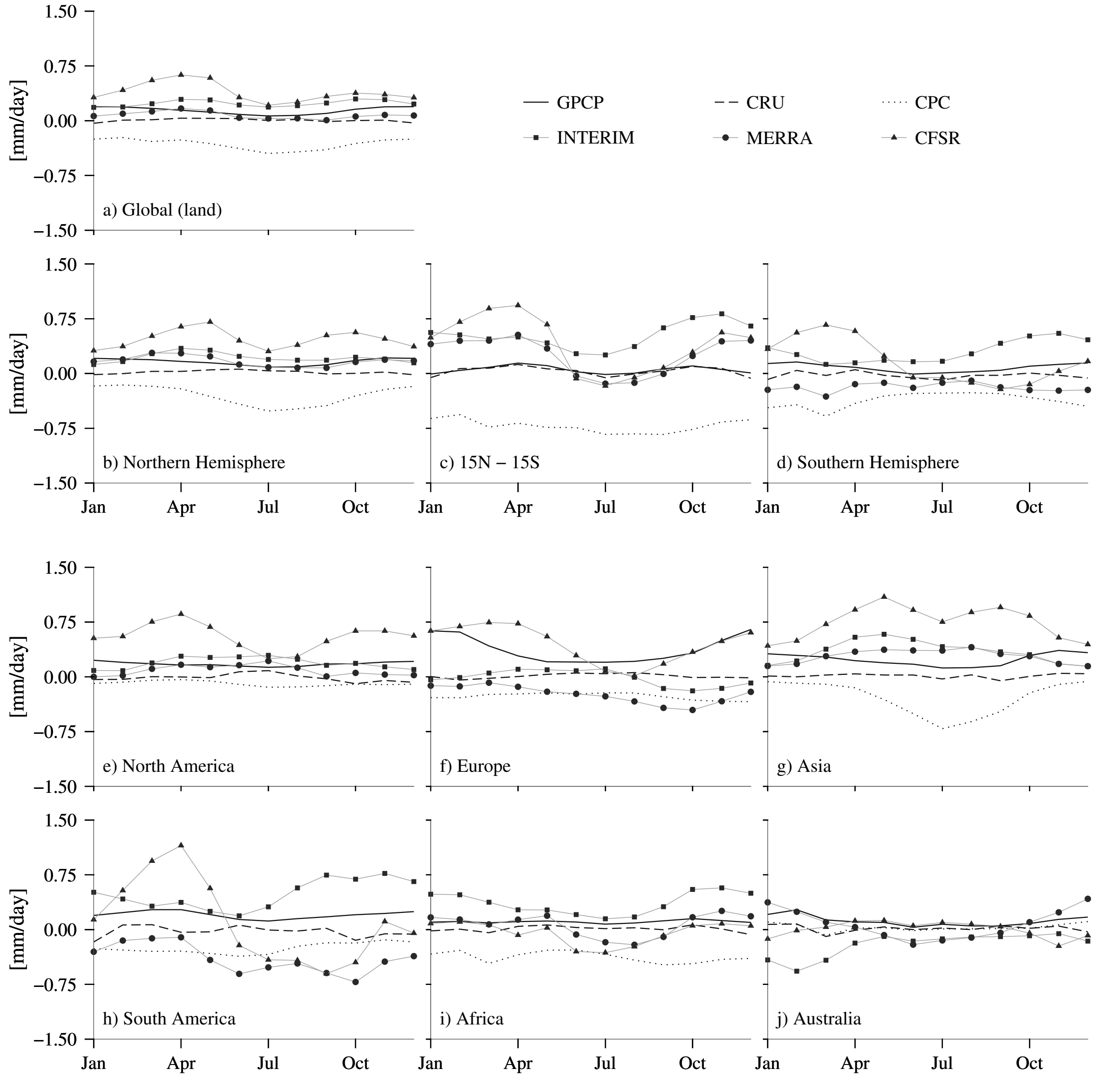

FIGURE 2.10: Long-term (17 years) averaged differences of monthly precipitation of GPCP, CRU, CPC, ERA-Interim, MERRA, and CFSR with respect to GPCC in [mm day^{-1}]. Over most of the regions, the differences between CFSR and GPCC show an intra-annual cycle which is obvious in South America and to a lesser extent over North America, Europe, and Asia. Both ERA-Interim and MERRA show a good agreement with GPCC over North America, Europe, and Asia, while over South America, ERA-Interim and MERRA assume too moist and too dry conditions, respectively, during an intra-annual cycle. A significant dry bias between GPCC and ERA-Interim can be observed over Australia during the period from January to March, whereas the other datasets are in good agreement with GPCC during the same period.

predicts a too-high spatial variability compared to GPCC during the summer months of the Northern and Southern Hemisphere, which is indicated by higher RMSD values. MERRA agrees best with GPCC during the boreal summer. Over the Southern Hemisphere and the tropics, MERRA shows the lowest correlation coefficients (< 0.6 for some years) of the three reanalyses. The performance of the models in reproducing

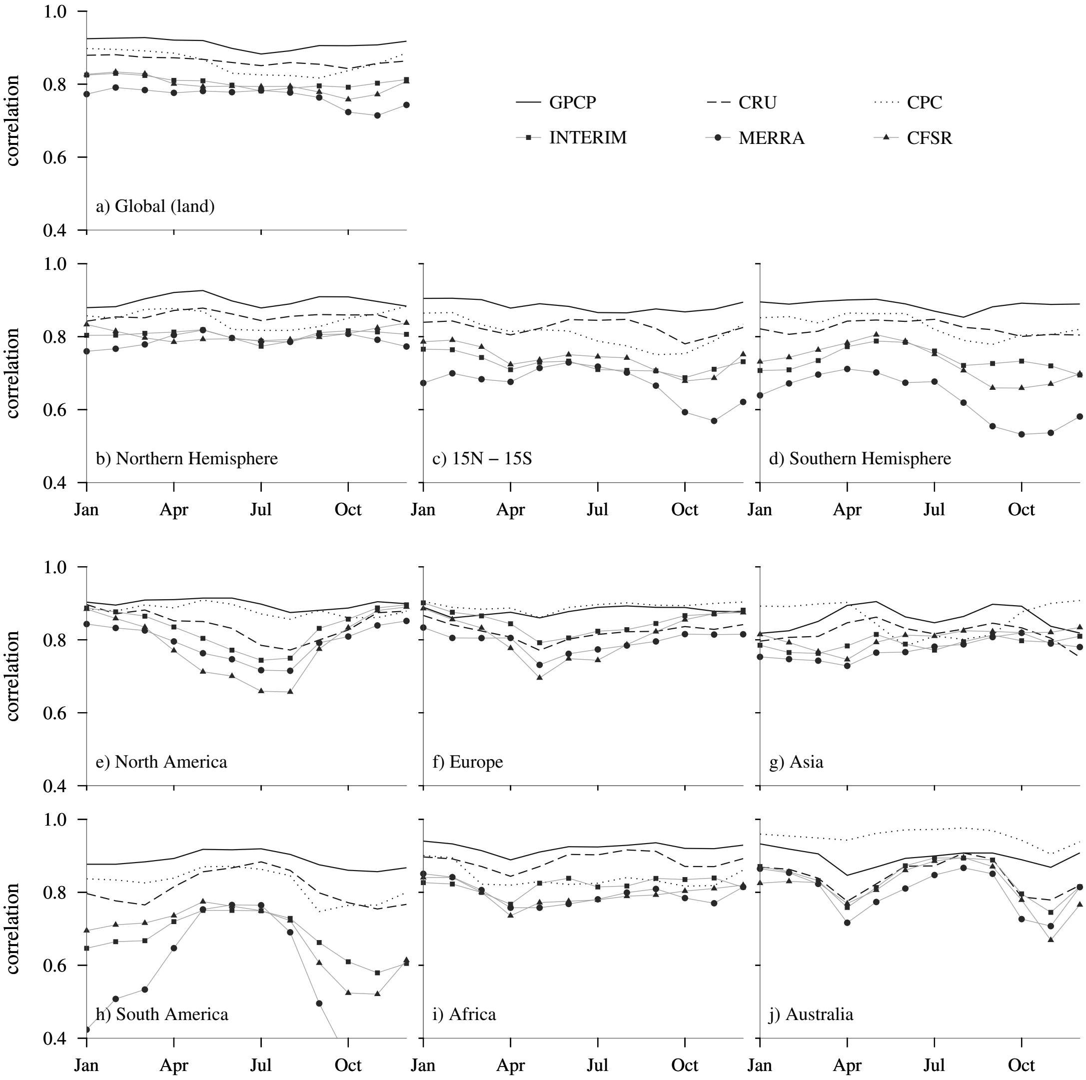

FIGURE 2.11: Area-averaged spatial correlations of the long-term (17 years) mean monthly rainfall of GPCP, CRU, CPC, ERA-Interim, MERRA, and CFSR with GPCC. Values close to 1 indicate that the respective dataset is in good agreement with the spatial distribution of rainfall from the GPCC product. Globally, the three reanalyses and observations reproduce the variations in the intra-annual rainfall patterns with a correlation coefficient > 0.7 and > 0.8, respectively. A significant intra-annual cycle can be observed over North America and Europe, which has its lowest values during the period from July to August (N. America). Over South America, MERRA shows correlation coefficients < 0.5 during from September to February. A similar intra-annual cycle, but less pronounced, can be observed for ERA-Interim and CFSR with its minimum between October and November. Over Australia, a significant drop in the spatial correlations can be observed in April and the period from October to November. As this drop is evident in both the reanalyses and the observations, there might be some shortcomings in the GPCC precipitation patterns during these months.

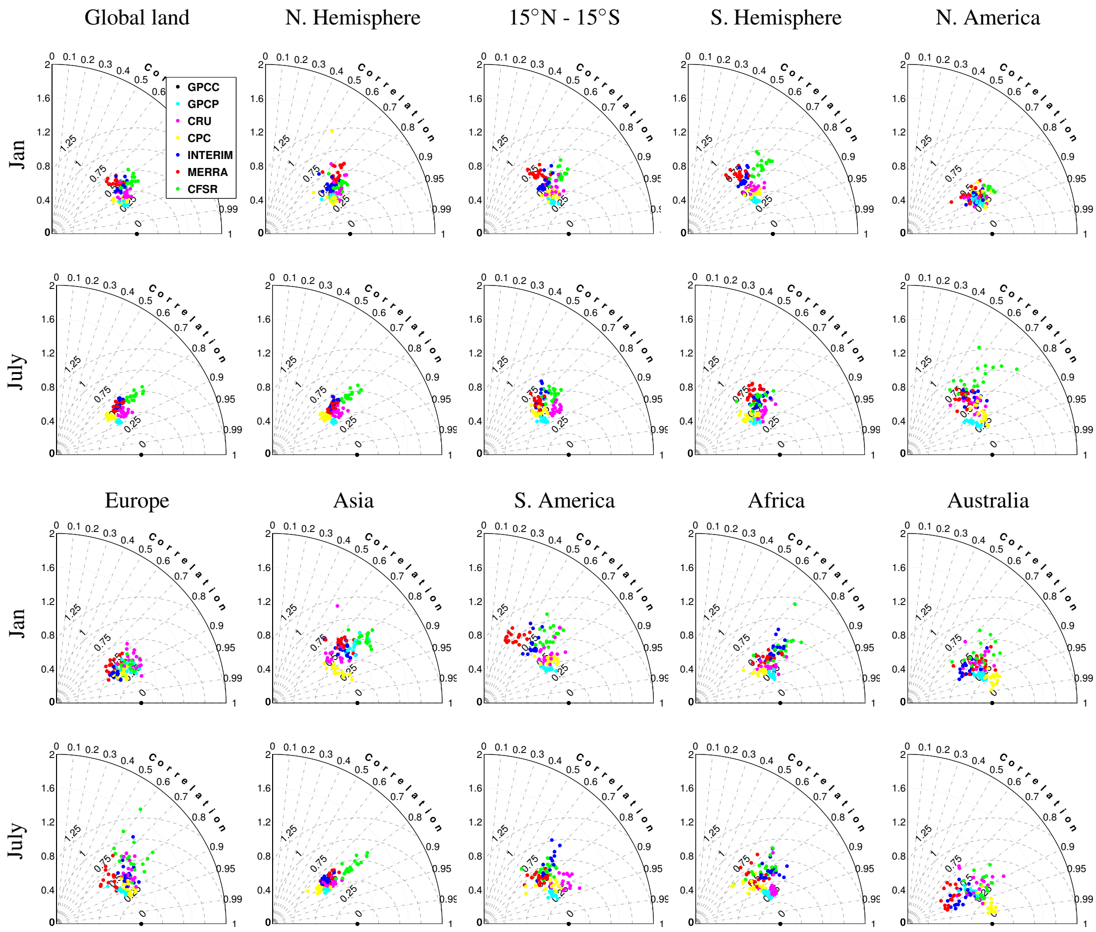

FIGURE 2.12: Taylor plots of spatial statistics of the mean monthly precipitation in January and July for GPCP, CRU, CPC, ERA-Interim, MERRA, and CFSR with respect to GPCC; each data point in a plot displays the correlation as the angle between the x-axis and the data point, the standard deviation (normalized) as the y-coordinate, and the root-mean-square difference (normalized) as the radial distance of one month of a specific year with respect to GPCC.

the GPCC rainfall patterns changes significantly depending on the region and time (month) but even from year to year. This is also true for the observation datasets although the other gridded rainfall observations on these scales agree better with GPCC than the reanalyses.

Time Evolution of Continental-Scale Precipitation

Over South America, the correlations of ERA-Interim and CFSR are in good agreement with an average correlation coefficient of about 0.7, while MERRA predicts completely different rainfall patterns, resulting in a low spatial correlation coefficient of ~ 0.5

(Figure 2.8h). As regards the intra-annual spatial variability (Figure 2.11h), the lowest correlations of MERRA are found between October and January. On the other hand, the differences in total precipitation between MERRA and CFSR (Figure 2.10h) show a reduced annual cycle compared to the correlations. This indicates that intra-annual variations in the amount of precipitation are in good agreement with the observations, while there are major differences in spatial variability. The Taylor diagrams (Figure 2.12) confirm the problems of the MERRA dataset in this respect, which cannot be explained by outliers exclusively. MERRA's annual mean correlations and deviations (Figures 2.8h and 2.9h) converge toward GPCC and the other datasets over South America, leading to better precipitation estimates at the end of the time series. The time when the MERRA precipitation estimates improve coincides with the assimilation of observations from the Advanced Microwave Sounding Unit (AMSU) on the NOAA-15 satellite. This assimilation is performed only over the oceans, but Bosilovich et al. [2011] note that such satellite epoch changes might indirectly affect the MERRA water balances over land through altered moisture.

Over North America, Europe, and Asia, significant wet biases in the mean annual and intra-annual CFSR precipitation are found (Figures 2.9e,f,g and 2.10e,f,g). While spatial correlations decrease during the period from May to August over North America, they are in good agreement with the other reanalyses (Figures 2.8e,f,g and 2.11e,f,g). Over Europe, the ERA-Interim reanalysis matches well with the GPCC observations with an average spatial correlation coefficient > 0.8 and a deviation of less than 0.1 mm day^{-1} on interannual and intra-annual time scales. The Taylor diagrams representing July over North America and Europe reveal that some data points predict a correlation coefficient < 0.7, which is likely due to an increase of convective precipitation. This is confirmed by the reanalyses' intra-annual spatial correlations, with the lowest correlation coefficients occurring in May (Europe; Figure 2.11f) and from July to August (North America; Figure 2.11e).

Over Asia, South America, and Africa, the differences between CFSR and GPCC decrease significantly, which might be explained by the assimilation of AMSU data (Figures 2.9g,h,i). After 1998, the wet bias of CFSR over Asia is constantly reduced

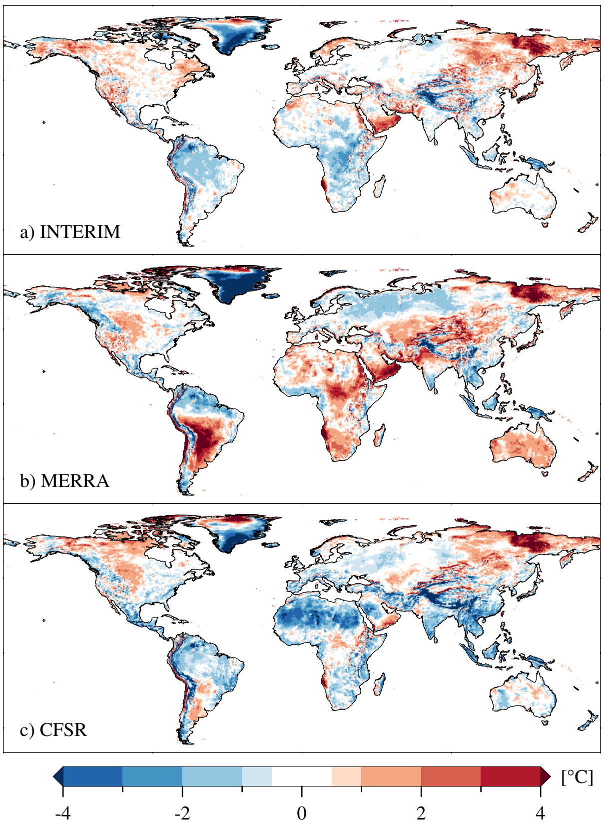

FIGURE 2.13: Differences in the mean annual temperatures at 2 m from ERA-Interim (a), MERRA (b), and CFSR (c) and CRU in [°C]. The largest differences between MERRA and CRU can be observed over South America, where MERRA shows a warm bias $> 4\,^\circ$C. The deviation pattern agrees well with the differences in the long-term mean precipitation estimates, where MERRA revealed a dry bias in these regions. This is also true for ERA-Interim, which assumes too cold and too wet conditions over the Congo basin. All three reanalyses show a warm bias over Siberia, while large parts of Greenland and the mountain ranges are generally too cold.

to 0.6 mm day^{-1} while bias reduction over the other continents is only temporary, as the differences between CFSR and GPCC increase toward the end of the time period.

Over North America, the MERRA precipitation estimates show the smallest deviations from the GPCC observations on both interannual and intra-annual time scales even though the slightly biased ERA-Interim estimates tend to display higher spatial correlations (Figures 2.8e and 2.11e). Over Europe, the precipitation estimates from ERA-Interim are superior to the other two reanalyses (Figures 2.8f, 2.9f, 2.10f, and 2.11f), while ERA-Interim has a wet bias over Africa because of the overestimation of precipitation in the Congo basin (Figure 2.6c). A general dry bias of the CPC observations can be noticed over all regions except for North America and Australia.

It is mentioned in the dataset description that especially over large parts of Africa and South America the observations should be treated carefully, as there is a very sparse distribution of gauges, even if the spatial correlations are in good agreement with the other datasets.

Over Australia, it can be seen that the three reanalyses and the observations from CRU show similar spatial correlations, which differ from GPCC especially in April and November (Figure 2.11j). These drops are also evident for GPCP and CPC, but to a smaller extent. As CRU and CPC are based solely on gauge observations, the Australian rainfall patterns from GPCC have to differ from the other datasets. However, this difference had not yet been detected.

2 m Temperature

The mean annual differences of the reference temperatures given by the CRU dataset and the three reanalyses are shown in Figure 2.13. The patterns of larger temperature differences are closely related to the differences in the precipitation fields (Figure 2.6). The MERRA temperature estimates (Figure 2.13b) seem to have a warm bias especially in South America, where the difference between MERRA and CRU reaches values of up to 6 °C. This warm bias may cause an increased saturation deficit of the air, which might explain the underestimation of South American precipitation in the MERRA dataset. A similar effect can be noticed over central Africa, where MERRA predicts too-warm conditions and too-little rainfall. The ERA-Interim field (Figure 2.13a) shows a cold bias in central Africa and South America and, thus, a decreased saturation deficit, which results in larger rainfall compared to the other datasets. The relation between the temperature and precipitation biases might also be explained by the reduced clouds and precipitation in these regions, leading to excess solar radiation reaching the surface, which results in an increased temperature.

In general, it can be concluded that the temperature range (Figure 2.7d) of the three reanalyses is similar to that of the precipitation fields (Figure 2.7c). The widest range of temperature observations (Figure 2.7b) can be detected over the large mountain ranges. This might be due to an elevation correction performed in the DEL dataset, but not in CRU. Over the largest parts of the continents, a general uncertainty of about 1 °C can be expected. When considering this value as an uncertainty bound, the large-scale deviations of MERRA and ERA-Interim over South America or the Congo basin and the general cold bias in the CFSR dataset over the whole Sahara indicate significant inaccuracies in the reanalyses. On the other hand, only slight deviations are encountered over North America, Europe, and Australia. Overall, ERA-Interim shows the best agreement with CRU.

Closure of the Water Budgets

Surface Water Budget

Table 2.3 summarizes the computed long-term mean values of the different quantities contributing to the global and continental-scale water budgets. The estimates from Trenberth et al. [2007] and Oki and Shinjiro [2006] are presented here as well for reference. In the long-term mean, ERA-Interim and MERRA show a reasonable closure of the global surface water balance, as $P-E$ over land equals the divergence of moisture $E-P$ over the oceans. ERA-Interim generally predicts more oceanic precipitation and evaporation. Both datasets achieve a closure of the combined continental–oceanic water budget (2.4) with a remaining residual of about 1 % (ERA-Interim) and 5 % (MERRA) of the continental $P-E$ moisture budget. Similar values for ERA-Interim and MERRA were also reported by Trenberth et al. [2011]. Jung et al. [2010] estimated a mean total land surface evapotranspiration of $65 \pm 3 \times 10^{15}$ kg year^{-1} between 1982 and 2008, which agrees with the estimates from Oki and Shinjiro [2006]. It should be noted that small deviations of the estimates may be due to differences in the used land–sea mask or, when compared to the estimates from, for example, Bosilovich et al. [2011], a different time period.

On the other hand, CFSR leaves an imbalance of about 80 % of the continental surface water budget due to an overestimation and underestimation of continental and oceanic $P-E$ values, respectively. It can be assumed that the too-small oceanic

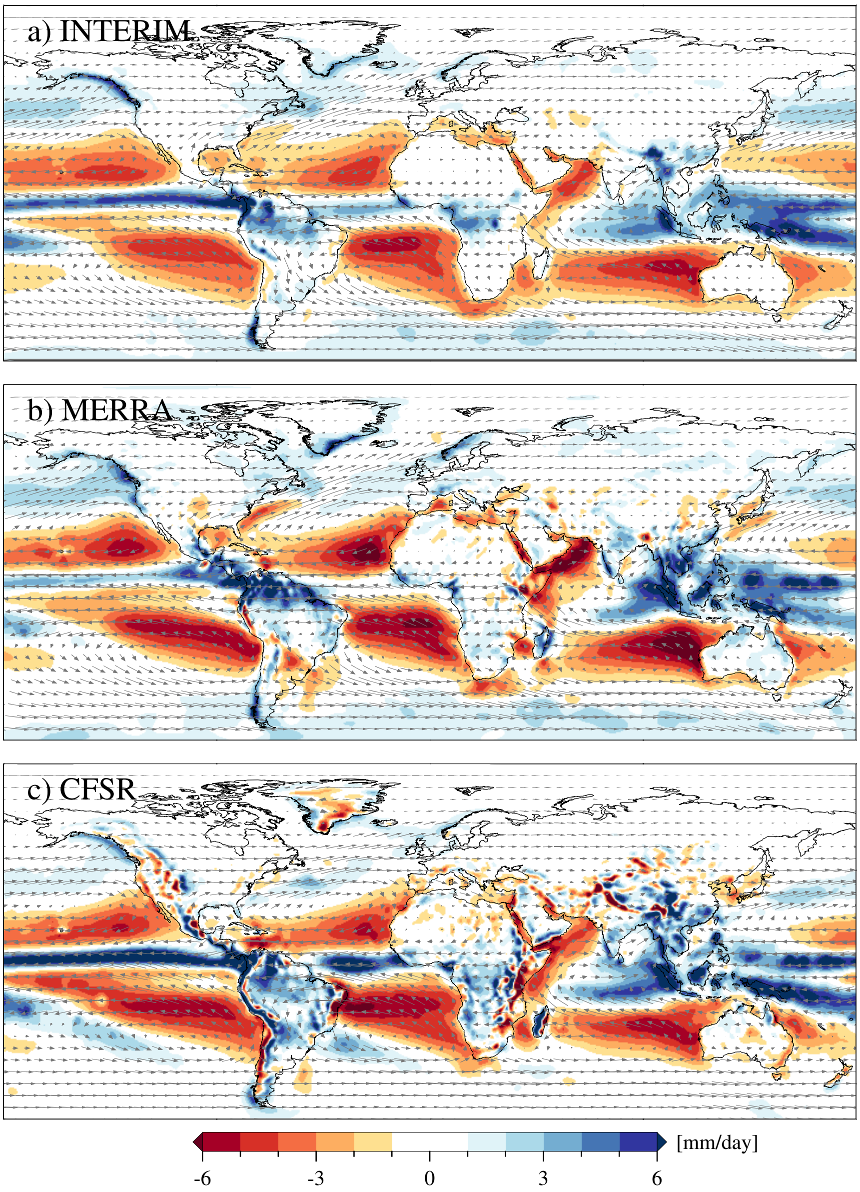

FIGURE 2.14: Vertically integrated moisture fluxes and moisture flux convergences in [mm day^{-1}] from ERA-Interim (a), MERRA (b), and CFSR (c). Positive values depict areas with a surplus of precipitation (i.e., $P > E$) while evaporation is larger than precipitation over regions with negative values. The direction and the amount of moisture transported are represented by the vector field. Large differences of the reanalyses can be observed along the ITCZ over the oceans, where CFSR shows larger moisture flux convergences than ERA-Interim and MERRA. There is also a positive pattern east of Brazil, which is absent in the other reanalyses. This results in a generally increased depletion of water in the CFSR over the oceans.

$P - E$ value of CFSR mainly arises from an overestimation of rainfall, as both ERA-Interim and MERRA assume an evaporation surplus of about 8 % with respect to the water that precipitates over the oceans, while CFSR predicts only 2 %. This is confirmed by the evaluation with the HOAPS dataset (Table 2.4), as CFSR predicts

TABLE 2.3: Mean global water cycle components over land and ocean between 1989 and 2006 in [10^{15} kg year^{-1}]; the values in the right most columns are the long-term estimates from Trenberth et al. [2007] (fifth column) and Oki and Shinjiro [2006] (sixth column) and are printed here as a reference.

	ERA-Interim	MERRA	CFSR	TB	Oki
P_{land}	119.7	115.4	127.7	113	111.0
E_{land}	82.1	86.7	70.4	73	65.5
P_{ocean}	418.0	381.3	459.3	373	391.0
E_{ocean}	455.3	411.5	470.0	413	436.5
R	46.6	31.3	38.3	40	45.5
$(P-E)_{\text{land}}$	37.6	28.7	57.3	40	45.5
$(P-E)_{\text{ocean}}$	−37.4	−30.2	−10.7	−40	−45.5
$(P-E)_{\text{ocean, before 98}}$	−29.1	−48.9	−26.2	–	–
$(P-E)_{\text{ocean, after 98}}$	−47.7	−6.8	8.6	–	–
$\nabla \cdot \mathbf{Q}_{\text{land}}$	32.3	38.2	34.6	40	45.5
$\nabla \cdot \mathbf{Q}_{\text{ocean}}$	−32.6	−38.2	−34.6	−40	−45.5
$RES_{s,\text{land}}$	9.0	2.6	19.0	–	–
$RES_{w,\text{land}}$	5.3	−9.5	22.7	–	–

significantly more rainfall than the other datasets. This might be due to the high moisture convergence in the oceanic domain of the intertropical convergence zone (ITCZ) (Figure 2.14c). There are also patterns of large positive $P-E$ values south east and west of South America, which are absent in ERA-Interim and MERRA, assuming a significantly larger depletion of water over the oceans.

According to Figure 2.15a, there is a shift in the global $P-E$ moisture budgets of CFSR and MERRA in 1998. Both depict a significant increase of oceanic $P-E$ (Figure 2.15c), with CFSR reaching values of about 20×10^{15} kg year^{-1} in 2001. In both cases, this is caused by an increase in oceanic rainfall, while ERA-Interim predicts a decrease (not shown). The increase in MERRA and CFSR is likely due to the assimilation of sounding radiances from AMSU-A on the NOAA-15 satellite from 1998 [Bosilovich et al., 2011, Nicolas and Bromwich, 2011]. Robertson et al. [2011] detected that the assimilation of AMSU-A data has a significant impact on the MERRA water vapor increments, leading to an increased amount of moisture in the model. This agrees with the time evolution of the total atmospheric water vapor content (Figures 2.17a,c), which shows a sudden increase of both MERRA and CFSR in 1998. Robertson et al. [2011] further concluded that the additional moisture causes an increase of precipitation especially over the tropic oceans. If so, there should also be an increase in oceanic evaporation for compensating the shift in oceanic rainfall, which cannot be detected in MERRA and CFSR. The significant changes of many CFSR variables in 1998 are discussed by Wang et al. [2010] and Xue et al. [2010].

The continental $P - E$ estimates (Figure 2.15b) are in much better agreement than the oceanic moisture budgets. In 1998, however, changes are significant, as MERRA predicts an increase of 6×10^{15} kg year^{-1}. The CFSR budgets decrease by about 12×10^{15} kg year^{-1} between 1998 and 2000. The gap in the continental $P - E$ budgets is less distinct and becomes smaller toward the end of the time series. The differences and spatial correlations of precipitation estimates from MERRA and GPCC show a sudden change in 1998 over South America only (Figures 2.8h, 2.9h). CFSR precipitation deviations from GPCC, however, exhibit significant gaps over Asia, South America, and Africa (Figure 2.9g,h,i). The reason has not yet been revealed, but if the assimilation of AMSU data causes these changes in the precipitation estimates, the impact of assimilation would differ significantly for both MERRA and CFSR.

As a result of the oceanic $P - E$ increase, the annual budgets between the oceans and the continents are highly distorted in MERRA and CFSR. CFSR shows a positive oceanic $P - E$ average of 8.6×10^{15} kg year^{-1} between 1999 and 2006, which obviously is not reasonable. The MERRA $P - E$ oceanic moisture budgets exhibit a change in sign after 2005. According to Rienecker et al. [2011], ERA-Interim does not use these observations and, thus, shows no shift in 1998. ERA-Interim's oceanic $P - E$ moisture budgets reveal a permanent downward trend until 1998, after which the budgets fluctuate around -45×10^{15} kg year^{-1}, which agrees with the reference values in Table 2.3.

Another significant shift in both MERRA and CFSR is assumed to occur in 2001 when data from the NOAA-16 satellite are introduced. Indeed, there is a distinct increase in the oceanic $P - E$ estimates of MERRA between 2000 and 2001. In the CFSR

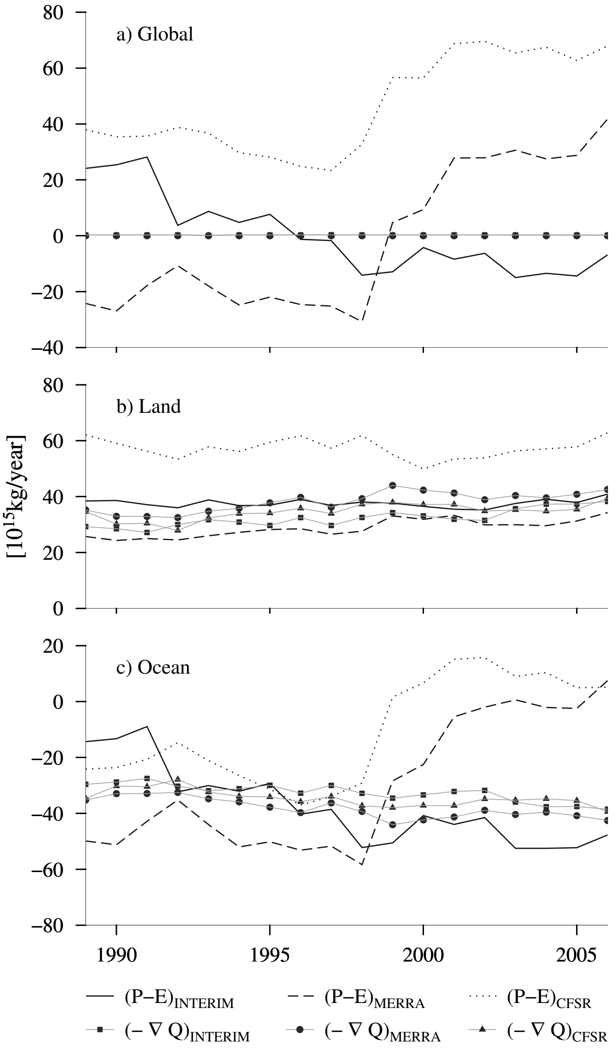

FIGURE 2.15: (a) Global, (b) continental, and (c) oceanic annual water balance in 10^{15} kg year^{-1}. $P-E$ budgets are plotted as black lines, while the atmospheric budgets are represented by the dotted gray lines. On annual time scales, the difference between $P-E$ and the moisture flux divergences is an estimate of the atmospheric water forcing increment in the reanalysis models. The closure of the combined atmospheric-terrestrial water budget would require both $P-E$ and $-\mathbf{\nabla}\cdot\mathbf{Q}$ to be equal. MERRA and CFSR show an unrealistic increase of oceanic $P-E$ from 1998, which is likely due to changes in the assimilated observations. Over the continents, the $P-E$ budgets from ERA-Interim and MERRA are in better agreement with the moisture fluxes, leading to a nearly closed continental atmospheric-terrestrial water budget. CFSR overestimates continental $P-E$ as well, which causes an imbalanced residual of about 30×10^{15} kg year^{-1}.

dataset, there also is an upward shift of oceanic $P-E$ estimates between 1999 and 2001, but the effect seems to weaken at the end of the time series.

When analyzing the continental surface water balance (2.3), MERRA shows the best performance in closing the long-term water balance. The surplus of evaporation

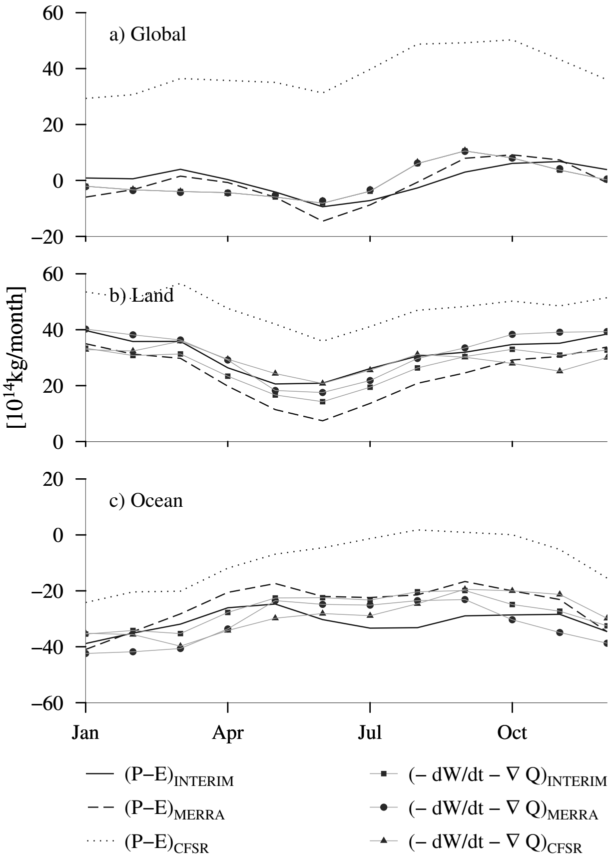

FIGURE 2.16: (a) Global, (b) continental, and (c) oceanic intra-annual water balance in 10^{14} kg/month. $P-E$ budgets are plotted as black lines, while the atmospheric budgets are represented by the dotted gray lines. For the monthly budgets, changes in the atmospheric water vapor content dW/dt were considered as well. ERA-Interim and MERRA show a good closure of the global atmospheric-terrestrial intra-annual budgets. Over the continents, there is a clear cycle with its minimum (maximum) in June (January) in both $P-E$ and the atmospheric moisture budgets, which might be explained by the decrease (increase) of evaporation (precipitation) during the boreal winter (summer) over the large continental areas of northern hemisphere. The CFSR $P-E$ estimates show a bias both over the oceans and the continents, causing a significant imbalance of the global intra-annual atmospheric-terrestrial moisture budget.

over the continents is balanced by reduced runoff compared to the other reanalyses and the reference estimates. This leads to a significantly smaller surface water forcing residual in the terrestrial water storage of 2.6×10^{15} kg year^{-1}. As the time evolution of both the annual $P-E$ moisture budgets and the total annual runoff (not shown) are not constant over time, however, RES_s changes as well. For water budget studies on shorter time scales, this changing imbalance should not be neglected. As reported by Roads et al. [2002], the storage change dS/dt may be significant during shorter periods and, hence, a large part of RES_s might be due to natural processes rather than artificial forcing increments. CFSR and ERA-Interim have larger residuals of 19.0×10^{15} kg year^{-1} and 9×10^{15} kg year^{-1}, respectively, between runoff and continental $P - E$ even if the runoff estimates seem to be more realistic compared to MERRA.

The global intra-annual water budgets (Figure 2.16a) show a clear annual cycle with the minimum $P - E$ in June due to the increased evaporation and reduced precipitation during the Northern Hemispheric summer months. Compared to the annual $P - E$ moisture budgets, the intra-annual variations of MERRA and ERA-Interim are in much better agreement. Even if CFSR reproduces a similar annual cycle, there is a significant deviation of about 30×10^{14} kg month^{-1} from the other reanalyses, which causes a remaining imbalance during an intra-annual cycle of the global $P - E$ moisture budgets.

Atmospheric and Combined Atmospheric-Terrestrial Water Balance

The global atmospheric water budgets $\boldsymbol{\nabla} \cdot \mathbf{Q}$ (Figure 2.15a) are nearly constant during the analyzed period. Consequently, the atmospheric moisture exchange between the oceans and the continents is a fully closed cycle in the three reanalyses. As the time series of moisture flux divergences obtained by the three reanalyses are in good agreement, it is concluded that using the atmospheric budgets for quantifying the exchange of moisture between the oceans and land masses is more reliable than the modeled $P - E$ moisture budgets. The closure of the global combined atmospheric-terrestrial water balance (2.8) reveals some shortcomings: ERA-Interim predicts too high (too low) $P - E$ estimates until 1996 (from 1996). CFSR has a significant moist bias (i.e., too high global $P - E$ values) over the whole time series with a sudden increase likely due to the assimilation of AMSU data in 1998, while MERRA shows the same gap with too dry (too wet) conditions until 1999 (from 1999). Over the continents (Figure 2.15b), both the atmospheric and terrestrial budgets are in good agreement, leading to the closure of the combined atmospheric-terrestrial water balance. Hence, the largest part of the global imbalance comes from the large gaps between the $P - E$ moisture budgets and the moisture flux divergences over the

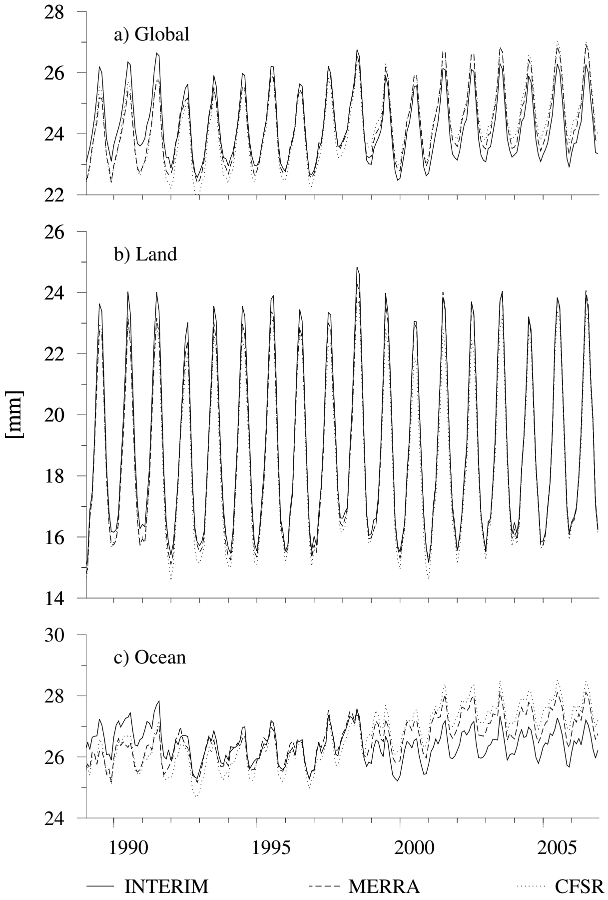

FIGURE 2.17: (a) Global, (b) continental, and (c) oceanic precipitable water estimates from the three reanalyses in [mm]. Globally, ERA-Interim shows higher values until 1995. Between 1992 and early 1998, the three reanalyses are in very good agreement, which is likely due to the assimilation of similar observations during that period. After 1998, the MERRA and CFSR estimates are higher when compared to ERA-Interim. This also holds for the time evolution of the atmospheric water vapor over the oceans. Again, it is presumed that the increase in 1998 is due to changes in the assimilated observations in both MERRA and CFSR, but not in ERA-Interim. When compared to other components of the large-scale water cycle, the water vapor estimates are in much better agreement, which is likely due to the forcing of the estimates towards observations in all three reanalyses.

oceans (Figure 2.15c). As the changes in the tendency terms of atmospheric and terrestrial water storage can be neglected on annual time scales, the differences between the moisture flux divergences and the $P-E$ budgets are an estimate of the atmospheric water forcing due to the assimilation of observations. Hence, the impact on MERRA and ERA-Interim is less pronounced over the continents than over the oceans.

As regards long-term monthly moisture budgets, the global, continental, and oceanic

ERA-Interim and MERRA $P-E$ values and atmospheric moisture fluxes are in very good agreement (Figure 2.16). We assume that even if the annual variations of the water cycle show significant shortcomings, the modeled processes are balanced well on a monthly time scale. The CFSR budgets show a significant overestimation of $P-E$ over both the continents and the oceans, resulting in a significant imbalance which has its maximum between September and October, where the global water budget leaves a monthly imbalance of up to 48×10^{14} kg month^{-1} (i.e., the intra-annual water cycle is not closed in CFSR). It should be noted, however, that CFSR does not provide fields of evapotranspiration. The imbalance might be affected largely by the approximation of E from fields of latent heat flux and sublimation (2.13). As the long-term average of continental evapotranspiration agrees with the model estimates from Trenberth et al. [2007], however, it is likely that the too high $P-E$ values arise from the CFSR precipitation.

Atmospheric Water Vapor

Figure 2.17 shows the monthly mean of total precipitable water over the complete time series. Globally, ERA-Interim predicts more atmospheric vapor before 1998 and less vapor after 1998 compared to MERRA and CFSR (Figure 2.17a). The main differences between the datasets result from deviations over the oceans (Figure 2.17c) that can be divided clearly into three periods. Before 1992, the ERA-Interim water vapor exceeds the estimates from CFSR and MERRA. Between 1992 and 1998, the three reanalyses are in good agreement, as the models use similar observational data in this period. After 1998, CFSR and MERRA show an increase of the precipitable water over the oceans, which has already been discussed in section *Surface Water Budget*. Compared to the differences in the reanalyses' water budgets, the time series of precipitable water are in good agreement. This is emphasized by Figure 2.18 where also satellite observations from the HOAPS dataset representing the total atmospheric water vapor over the ice-free ocean are shown. Especially during the period between 1992 and 1998, the reanalyses successfully reproduce the annual cycle of water vapor over the oceans. After 1998, MERRA and CFSR overestimate the

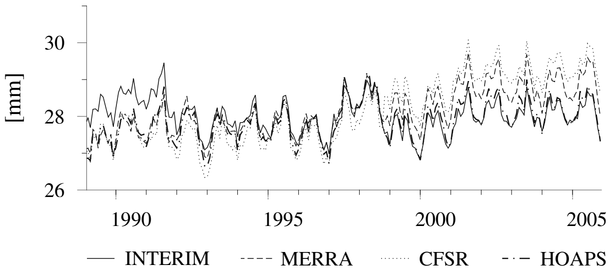

FIGURE 2.18: Precipitable water in [mm] from the reanalyses and the HOAPS dataset over the ice-free ocean. Between 1992 and 1998, the three reanalyses are in good agreement with the HOAPS observations. It is obvious that the increase in 1998 of both MERRA and CFSR causes an overestimation of precipitable water when compared to HOAPS. ERA-Interim shows an overestimation before 1992, but reveals a good agreement with HOAPS for the rest of the time period.

amount of precipitable water, while ERA-Interim still shows a good agreement with HOAPS. On the other hand, ERA-Interim clearly overestimates atmospheric water vapor before 1992, while MERRA and CFSR agree well with HOAPS. This changing level of agreement is likely due to the assimilation of different data sources, as all three reanalyses use similar observations between 1992 and 1998 only.

Changes in the atmospheric water vapor content dW/dt were considered when computing the monthly water budgets. There is a strong annual cycle especially over the continents, leading to maximal values of dW/dt in spring and autumn. This agrees with, for example, Rasmusson [1968] and is also considered in Seneviratne et al. [2004]. However, changes in the vertically integrated water vapor usually are smaller by several orders of magnitude compared to other quantities of the hydrological cycle and do not vary on annual time scales, as the intra-annual cycle of dW/dt is closed with sufficient accuracy in all three reanalyses (not shown). It is therefore proposed to neglect dW/dt for large-scale and long-term water budget studies.

Evaluation of Hydrological Variables over the Oceans

Mean estimates of modeled precipitation, evaporation, and $P-E$ over the ice-free oceans are presented and compared with satellite observations from HOAPS in Table 2.4. As another reference, the GPCP dataset predicts a mean precipitation rate

TABLE 2.4: Mean oceanic precipitation, evaporation, and $P-E$ between 1989 and 2006 from ERA-Interim, MERRA, CFSR, and satellite observations from the HOAPS dataset in [mm day^{-1}]. The numbers in the brackets denote the standard deviations.

	P	E	$P-E$
HOAPS	2.86 (±0.16)	3.54 (±0.18)	−0.68 (±0.23)
ERA-Interim	3.29 (±0.10)	3.71 (±0.09)	−0.43 (±0.13)
MERRA	2.99 (±0.20)	3.37 (±0.08)	−0.39 (±0.20)
CFSR	3.56 (±0.16)	3.81 (±0.11)	0.08 (±0.15)

of about 3.0 mm day^{-1}. CFSR shows an overestimation of oceanic rainfall of 0.6 mm day^{-1}. As evaporation from CFSR is only about 0.3 mm day^{-1} higher compared to HOAPS, the mean $P-E$ moisture budget is positive, showing an overestimation of 0.7 mm day^{-1}. All three reanalyses overestimate the $P-E$ moisture budget, but in ERA-Interim, this is due to an overestimation of precipitation and evaporation, while MERRA underestimates evaporation. Hence, it is impossible to make a general statement about the origins of the too large oceanic $P-E$ moisture budgets in the analyzed reanalyses.

Summary and Conclusions

The present study demonstrated major differences of differences between the three reanalyses ERA-Interim from ECMWF, MERRA from NASA, and CFSR from NCEP. Precipitation is one of the most important quantities of the water cycle. Its estimates are highly uncertain in terms of spatial variability and total amount. The largest discrepancies occur in the summer months of the respective hemisphere, because convective effects still are a large source of uncertainty. However, a validation of, for example, tropical or mountainous rainfall also remains difficult because of the large differences of the observation datasets in these areas. This can be attributed to the irregular distribution of gauges especially in such complex and highly variable regions. Hence, there are large parts of South America and Africa that are completely ungauged. For regions with a dense network of gauges like North America, Australia, or Europe, it may be concluded that ERA-Interim still provides the most reliable rainfall estimates. The largest problem in these areas is the decrease of active gauges during the period considered. The quality of an interpolated product over a continuously changing network of observations remains questionable. On the other hand, validation of the coarsely resolved GPCC and reanalysis precipitation estimates with *in situ* rainfall observations is not yet meaningful due to the precipitation's high spatial variability and dependence on surrounding terrain.

In data-sparse regions, the general tendency of the datasets considered can be regarded examined only. It may be concluded that major shortcomings exist in the spatial patterns of South American rainfall in MERRA and in the total amount of mid- to high-latitude precipitation in CFSR where a significant bias was detected. In the case of the MERRA reanalysis, these shortcomings are well known. Thus, NASA is currently performing a land-only rerun of the MERRA reanalysis in which observational data from GPCP and other independent global data products are used to correct and evaluate MERRA's land surface hydrology [Rienecker et al., 2011]. The corrected precipitation estimates are already available and will replace the original MERRA products in the near future [Reichle et al., 2011].

The differences in the amount and spatial patterns of continental rainfall between GPCC and the other datasets remain more or less constant during the whole period (except for South American precipitation in the MERRA dataset). This is important, as the assimilation of observations, which became available during the period considered, does not seem to significantly affect the precipitation estimates over the landmasses. The situation is clearly different for oceanic precipitation that exhibits a significant shift in late 1998 when sounding radiances from AMSU were assimilated into both MERRA and CFSR. Atmospheric reanalysis models are still sensitive to the introduction of observational data. Similar findings were also presented in Bengtsson [2004] for the ERA-40 reanalysis.

The uncertainties in the precipitation estimates of the three reanalyses are highly correlated with the variability in the temperature fields. An obvious connection was found between the large-scale underestimation of South American precipitation from MERRA and a general warm bias in these regions. To make statements about the validity of an atmospheric reanalysis, it is not sufficient to consider one quantity like, for example, precipitation only, but also other variables must be taken into account. This is crucial to the closure of the water budgets of these datasets. When introducing new observations, CFSR and MERRA became significantly imbalanced after 1998. Oceanic precipitation increased because of the assimilation of sounding radiances from AMSU. In both CFSR and MERRA, the increase in total water is not balanced by an increase in oceanic evaporation or a decrease of continental $P-E$ estimates, leaving a large gap in the global water balance. The atmospheric budgets do not show such a sudden shift, but tend to increase during the time series considered. Furthermore, the differences of the $P-E$ estimates are much larger than those of their atmospheric counterpart. It may therefore be concluded that they are still more reliable than the terrestrial $P-E$ values.

Because of the limitations presented, the performance of all three reanalyses in reproducing the hydrological cycle still causes doubts in the use of such models for climate trend analyses and long-term water budget studies.

Acknowledgments

We would like to thank the European Centre for Medium-Range Weather Forecast, the Global Modeling and Assimilation Office, the Goddard Earth Sciences Data and Information Services Center, the National Centers for Environmental Prediction, and the National Center for Atmospheric Research for creating and providing the reanalysis data. GPCC data were obtained from `http://gpcc.dwd.de`. GPCP rainfall observations are developed and computed by the NASA/Goddard Space Flight Center's Laboratory for Atmospheres as a contribution to the GEWEX Global Precipitation Climatology Project (`http://precip.gsfc.nasa.gov/`). The CRU high-resolution precipitation and temperature data were downloaded from the British Atmospheric Data Centre (`http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_1256223773328276`). The daily rainfall observations from the Climate Prediction Center are provided at `ftp.cpc.ncep.noaa.gov/precip/CPC_UNI_PRCP/GAUGE_GLB/`. Precipitation and temperature observations from the University of Delaware were created at the Center for Climatic Research at the Department of Geography (`http://climate.geog.udel.edu/~climate/`). We would also like to thank Prof. Dr. Tonie van Dam (University of Luxemburg), Dr. Patrick Laux (Karlsruhe Institute of Technology), and M. Sc. Balaji Devaraju (University of Stuttgart), who improved this paper with their comments and corrections. This work is part of the *Direct Waterbalance* subproject of the DFG priority programme 1257 *Mass Transport and Mass Distribution in the System Earth* `http://www.massentransporte.de`.

Chapter 3

Article II: Large-scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances

Lorenz, C., B. Devaraju, M. J. Tourian, J. Riegger, N. Sneeuw, and H. Kunstmann, 2014: Large-scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances, *Journal of Hydrometeorology*, **15**, 2111–2139, doi: 10.1175/JHM-D-13-0157.1[1]

Abstract

The performance of hydrological and hydrometeorological water-balance-based methods to estimate monthly runoff is analyzed. Such an analysis also allows for the examination of the closure of water budgets at different spatial (continental and catchment) and temporal (monthly, seasonal, and annual) scales. For this analysis, different combinations of gridded observations [Global Precipitation Climatology Centre (GPCC), Global Precipitation Climatology Project (GPCP), Climate Prediction Center (CPC), Climatic Research Unit (CRU), and University of Delaware (DEL)],

---

atmospheric reanalysis models [Interim ECMWF Re-Analysis (ERA-Interim), Climate Forecast System Reanalysis (CFSR), and Modern-Era Retrospective Analysis for Research and Applications (MERRA)], partially model-based datasets [Global Land Surface Evaporation: The Amsterdam Methodology (GLEAM), Moderate Resolution Imaging Spectroradiometer (MODIS) Global Evapotranspiration Project (MOD16), and FLUXNET Multi-Tree Ensemble (FLUXNET MTE)], and Gravity Recovery and Climate Experiment (GRACE) satellite-derived water storage changes are employed. The derived ensemble of hydrological and hydrometeorological budget–based runoff estimates, together with results from different land surface hydrological models [Global Land Data Assimilation System (GLDAS) and the land-only version of MERRA (MERRA Land)] and a simple predictor based on the precipitation–runoff ratio, is compared with observed monthly in situ runoff for 96 catchments of different sizes and climatic conditions worldwide. Despite significant shortcomings of the budget-based methods over many catchments, the evaluation allows for the demarcation of areas with consistently reasonable runoff estimates. Good agreement was particularly observed when runoff followed a dominant annual cycle like the Amazonas. This holds true also for catchments with an area far below the spatial resolution of GRACE, like the Rhine. Over catchments with low or nearly constant runoff, the budget-based approaches do not provide realistic runoff estimates because of significant biases in the input datasets. In general, no specific data combination could be identified that consistently performed over all catchments. Thus, the performance over a specific single catchment cannot be extrapolated to other regions. Only in few cases do specific dataset combinations provide reasonable water budget closure; in most cases, significant imbalances remain for all the applied datasets.

Introduction

With water, the most essential natural resource for sustaining life, coming under threat [Vörösmarty et al., 2010], it becomes imperative to understand the hydrological cycle and monitor its constituents. This urgent need, however, has not been able to stem the steady decline in the number of precipitation and runoff gauges (Figure 3.1),

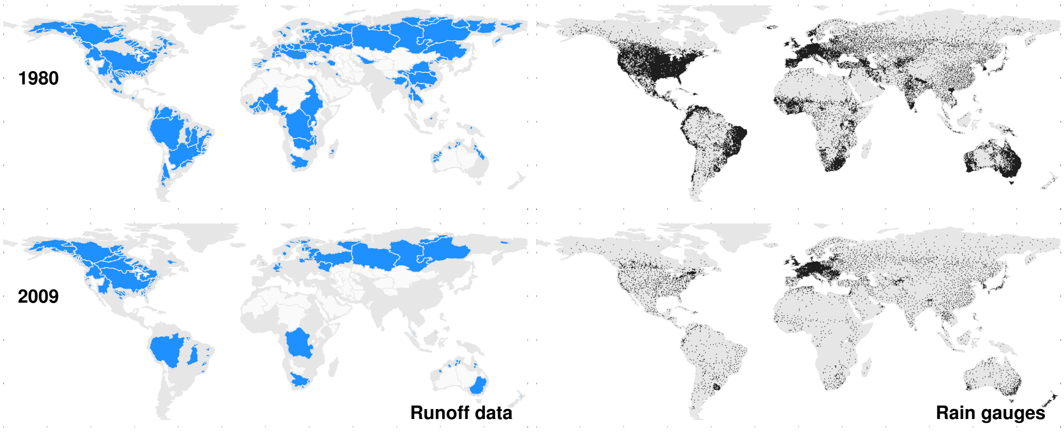

FIGURE 3.1: The maps show the decline in the number of runoff (left) and rain (right) gauges. The catchments whose runoff measurements are publicly available (blue), that do not discharge into the oceans or other open water bodies (white) and those that are not gauged or whose runoff measurements are not publicly available (gray) are shown on the left column. In the right column the individual dots represent one or more rain gauges in a 0.5° × 0.5° gridcell.

which still provide the only direct observations of terrestrial water cycle components. This is a setback both for our understanding of the evolution of the hydrological cycle and for the calibration and validation of hydrological and atmospheric models and *ex situ* measurements of those components. The situation is further aggravated by the lack of reliable data for evapotranspiration and water storage changes, which, together with precipitation and runoff, form the major components of large-scale hydrological cycle.

Over the years, in the absence of measurements of the hydrological cycle components, it has been normal practice to resort to sophisticated hydrological and atmospheric reanalysis models to study the global water balance [e.g., Trenberth et al., 2007]. However, the hydrological or land surface models suffer from the steady changes in the spatial distribution and number of precipitation or runoff gauges. For such models, observations are needed both as an input and also as a validation measure. Therefore, there is a desperate need for new and alternative methods.

Spaceborne measurements seem to provide the necessary respite by a plethora of satellites providing nearly all the components of the global hydrological cycle: precipitation from the Tropical Rainfall Measuring Mission [TRMM; Kummerow et al., 2000], water storage changes from the Gravity Recovery and Climate Experiment

[GRACE; Schmidt et al., 2008], runoff from satellite altimetry and synthetic aperture radar [Alsdorf et al., 2007], soil moisture from the Soil Moisture and Ocean Salinity (SMOS) satellite [Kerr et al., 2010] or (beginning in late 2014) the Soil Moisture Active Passive (SMAP) satellite [Entekhabi et al., 2010], and a host of variables concerning the hydrological cycle from *Aqua* [Parkinson, 2003]. The space-borne observations bring with them the advantages of global coverage and homogeneous accuracy, but with limited spatial and temporal resolutions. Hrachowitz et al. [2013] acknowledge these advancements and provide examples of spaceborne measurements becoming indispensable to hydrological research, especially in data-poor regions like West Africa, Asia, and the vast open oceans.

In the context of runoff, Alsdorf et al. [2007] provide a broad overview of the types of spaceborne runoff measurements, their relative merits and challenges. Recently, but not for the first time, Tourian et al. [2013] demonstrated the potential and accuracy of runoff derived from satellite altimetry. However, an application of this method to catchments still depends on the existence of contemporaneous or previous *in situ* runoff measurements. Therefore, our study focuses on independent methods using mass estimates of the water storage in river systems from GRACE, while observed runoff is used as a validation measure only.

The use of GRACE water storage changes entails the use of the water balance equations, wherein runoff is computed from a combination of precipitation and evapotranspiration in addition to water storage changes. Such water budget based methods have already been applied by Sheffield et al. [2009] who used GRACE and other remotely sensed data for computing streamflow from the Mississippi basin. Syed et al. [2005] used an atmospheric water balance for estimating discharge over the Amazonas and Mississippi basins and, in the following years, total freshwater discharge from the entire Pan-Arctic [Syed et al., 2007] or continents and large river basins [Syed et al., 2009]. Other methods of estimating runoff from GRACE are also being proposed [e.g., Riegger and Tourian, 2014] but they are still in a very nascent state of development. For a long time, evapotranspiration has been elusive and has been the thorn in the flesh in global hydrological studies [e.g., Schlosser and Houser, 2007, Trenberth et al., 2007]. Lately, attempts have been made to provide global evapotranspiration estimates apart from what has been available from various hydrological and atmospheric reanalysis models [Jung et al., 2009, Miralles et al., 2011b, Mu et al., 2007, Salvucci and Gentine, 2013]. Precipitation and runoff datasets are the most mature of all the hydrological datasets, but there are significant differences between the different precipitation datasets [Lorenz and Kunstmann, 2012]. With an accuracy of 10–20 %, observed runoff is the most accurately measured component of the hydrological cycle [Fekete et al., 2012], but suffers from variations in the time period of observation, substantial delays in data access and large declines in the monitoring capacity [Fekete and Vörösmarty, 2002, Fekete et al., 2002, Shiklomanov et al., 2002, Vörösmarty et al., 2001].

To analyze the performance of runoff estimations independent of gauge-based observations, we thus set out to derive large-scale runoff from catchments using water-balance-based methods. Therefore, we use the most recent versions of widely used data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences. By combining all these datasets, we end up with a total of 90 possible runoff estimates from a terrestrial water balance and 15 from an atmospheric–terrestrial water balance for each catchment. These estimates are then validated against observed runoff data from the Global Runoff Data Centre [GRDC; GRDC, 2013] over 96 river basins. The performance of these estimates is also compared with the performances of sophisticated hydrological (land surface) models and an empirical method involving the precipitation-runoff ($P-R$) ratio. This allows us to:

- evaluate the potential of runoff estimation from water-balance-based methods, land surface models and an empirical predictor,
- analyze the quality of the different contributing datasets, and
- evaluate the current status of the closure of catchment scale water budgets

Until now, only a few studies existed where runoff from such a large number of river basins was evaluated. Some examples of these are Zaitchik et al. [2010], which focuses on the validation of Global Land Data Assimilation System (GLDAS) over

66 larger river basins, or Zhou et al. [2012], where runoff estimates from different land surface models are compared against mean annual runoff observations over 150 catchments. A validation of runoff from global climate models using the Total Runoff Integrating Pathways (TRIP) river flow model is presented in Falloon et al. [2011]. In most cases, such studies focus on hydrological models, while we estimate runoff from simple budget equations where the most recent gridded observation datasets, land surface, and atmospheric reanalysis models serve as inputs.

Our study is outlined as follows: in section *Methods*, we elaborate on the water-balance-based methods and the empirical method for estimating runoff from precipitation; in section *Data*, the datasets that we have used in the different approaches are detailed; in section section *Results and Discussion*, we investigate the utility of the runoff estimates from water-balance-based methods as an alternative to observed runoff and also exploit those estimates to study the global water balance closure; and finally, we draw our conclusions in section *Summary and Conclusions*.

Methods

The primary concern in this study is to estimate runoff globally from the landmasses, and we have chosen to use the water balance equations as the means to our end. Water balance equations define the hydrological cycle in quantitative terms by taking into account the amount of water entering and leaving a river basin within, for example, 1 month. This budgeting can be approached in several different ways, and here we have chosen both *large-scale terrestrial* and *atmospheric-terrestrial* water balance equations. Our choice was driven mainly by the availability of water storage change data from the GRACE satellite mission. All our calculations are performed on a monthly basis in order to be as consistent as possible with the monthly runoff observations as provided from the GRDC.

Hydrological Budget Approach

At continental scales, the water entering a river basin is described by precipitation P, the water leaving it is described by evapotranspiration ET_a and runoff R, and whatever remains is the water storage change $\mathrm{d}S/\mathrm{d}t$. Mathematically, this can be written as

$$P - ET_\mathrm{a} - R = \frac{\mathrm{d}S}{\mathrm{d}t}, \tag{3.1}$$

which is the large-scale terrestrial water balance equation, and it immediately allows us to compute the river runoff as

$$R = P - ET_\mathrm{a} - \frac{\mathrm{d}S}{\mathrm{d}t} \equiv R_\mathrm{hyd}. \tag{3.2}$$

Since this method uses only hydrological variables to arrive at runoff estimates, it will henceforth be called the *hydrological approach*. This approach does not require runoff routing as it is taken care of by the water storage changes, and it comes with the added benefit that no matter which component we estimate, the closure of the water balance can be analyzed. The major pitfall of such an approach is that the accuracy of the runoff estimates will be only as good as the least accurate dataset. Because of the law of error propagation, errors in the individual water cycle components might thus add up and result in very uncertain estimates of runoff. Further, since at any point there is a need for three different datasets to study one of the variables of interest, consistency in terms of spatial and temporal resolution are of paramount importance. On the other hand, biases in the individual input datasets might get canceled out as runoff is computed as a residual.

Hydrometeorological Budget Approach

The number of inputs in (3.1) can be reduced to two by linking the terrestrial water balance with the atmospheric-terrestrial water balance equation [Roads et al., 1994]

$$\frac{\mathrm{d}W}{\mathrm{d}t} + \boldsymbol{\nabla} \cdot \mathbf{Q} = ET_\mathrm{a} - P, \tag{3.3}$$

where W denotes the total column water content in the atmosphere and $\nabla \cdot \mathbf{Q}$ is the net balance of moisture flux (i.e., moisture flux divergence). The moisture flux is defined as

$$\mathbf{Q} = \frac{1}{g} \int\limits_{p=0}^{p_{\text{sfc}}} \boldsymbol{v}_{\text{h}}(p)\, q(p)\, \mathrm{d}p \tag{3.4}$$

with air pressure p (Pa), the gravitational acceleration g (m s^{-2}), the horizontal wind vector $\boldsymbol{v}_{\text{h}}$, and the specific humidity q (kg kg^{-1}) [e.g., Fersch et al., 2012].

According to Peixoto and Oort [1992], changes in the atmospheric water content $\mathrm{d}W/\mathrm{d}t$ can be neglected on monthly or longer time scales, which gives us the result

$$\nabla \cdot \mathbf{Q} \approx ET_{\text{a}} - P\,. \tag{3.5}$$

By combining (3.1) and (3.5), runoff can be derived from the atmospheric-terrestrial water balance

$$R_{\text{atm}} = -\nabla \cdot \mathbf{Q} - \frac{\mathrm{d}S}{\mathrm{d}t}\,. \tag{3.6}$$

We will call this method the *hydrometeorological approach* as it takes both hydrological and atmospheric datasets as inputs. This method also inherits the advantages and disadvantage of the hydrological approach.

Runoff-Precipitation Ratio

Since there is a strong relationship between runoff and precipitation, runoff can be estimated directly from precipitation by scaling it with an empirically derived scaling factor. This factor is the ratio between runoff and precipitation, which is computed for each calendar month (January—December) individually. For each catchment, the runoff value of the j^{th} calendar month in the i^{th} year $R_{i,j}$ is divided by the corresponding precipitation value $P_{i,j}$

$$\kappa_{i,j} = \frac{R_{i,j}}{P_{i,j}}\,. \tag{3.7}$$

Then, an average $\bar{\kappa}_j$ of the ratios $\kappa_{i,j}$ for each calendar month j is computed over the available N number of years:

$$\bar{\kappa}_j = \frac{1}{N}\sum_{i=1}^{N}\kappa_{i,j}\,. \tag{3.8}$$

The runoff is then computed for each month by scaling precipitation Pi, j with $\bar{\kappa}_{i,j}$

$$R_{i,j} = \bar{\kappa}_j\, P_{i,j}\,. \tag{3.9}$$

This empirical model depends on previous runoff data, and the past and current precipitation data, which makes it more of a predictor model rather than an estimator. The computation of $\bar{\kappa}_j$ for individual calendar months takes care of the mean annual cycle, and by scaling them with the concurrent precipitation data, the associated climatic variations are also imbued into the runoff estimates.

Due in part to the simplicity and empirical nature of the model, its success is entailed to the dynamic behavior of the river system whose runoff is estimated. Runoff of a river system more prone to vagaries in the mean annual hydrological cycle, or in which there is a lot of anthropogenic intervention, might not be estimated via this model. Furthermore, evapotranspiration and water storage do not appear explicitly in these formulas, but they certainly affect the relationship between precipitation and runoff (i.e., their impacts are "hidden" within the ratios). It is thus essential that both evapotranspiration and water storage changes follow a stable annual cycle, which does not change significantly from year to year. That being said, estimating runoff from catchments with, for example, a trend in one of these two variables might lead to large uncertainties.

To evaluate the performance of this empirical model as a predictor (i.e., for ungauged regions or runoff observations with large data-gaps), the values of $\bar{\kappa}_j$ are computed from runoff and precipitation data only until the end of 2002. From 2003, runoff is estimated by multiplying current precipitation data with the mean monthly scaling factor $\bar{\kappa}_j$ of the corresponding month j. For analyzing the reliability of this approach, different precipitation datasets (see section *Data*) have been used for both the estimation of $\bar{\kappa}_j$ and the prediction of runoff.

Water Budget Imbalance

The GRACE-based methods that we use here are based on the water balance equations, which allow us to compute the imbalance in the water budget:

$$\delta_{\mathrm{hyd}} = R - R_{\mathrm{hyd}} = R - P + ET_{\mathrm{a}} + \frac{\mathrm{d}S}{\mathrm{d}t}, \tag{3.10}$$

$$\delta_{\mathrm{atm}} = R - R_{\mathrm{atm}} = R + \boldsymbol{\nabla} \cdot \mathbf{Q} + \frac{\mathrm{d}S}{\mathrm{d}t}, \tag{3.11}$$

where δ_{hyd} and δ_{atm} are the water budget imbalances of the hydrological and atmospheric water balance equations, respectively.

Data

The water balance equations (3.2) and (3.6) and the empirical model (3.9) use all the components of the hydrological cycle. We have ensured that all the datasets that we have used for our analysis here are observation-based. In cases where direct observations are not possible, for example, moisture flux divergence, only partially modeled datasets are used. In addition to this, we have also taken care to use only datasets that are publicly available. An overview of the different data sources of precipitation, evapotranspiration, water storage changes, moisture flux divergences, and runoff is given in Table 3.1.

For the comparison of the runoff estimates with the observations, we have remapped all the input fields to the same spatial resolution (0.5° × 0.5°) using a first-order conservative interpolation [Jones, 1999]. From these fields, area-weighted averages were computed over the study regions shown in Figure 3.2. To be temporally consistent with water storage changes from GRACE, Landerer et al. [2010] proposed to

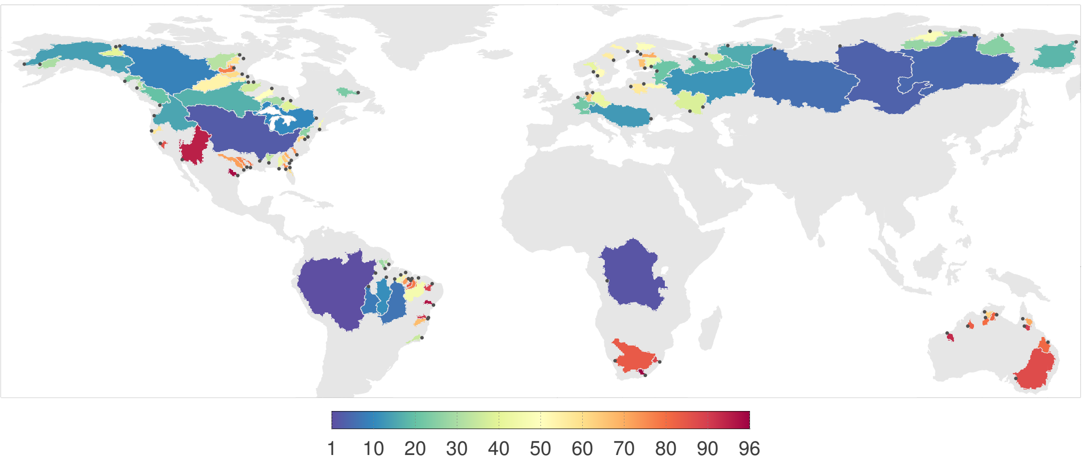

FIGURE 3.2: Location of the basins given in Table 3.2 where the dot in each basin is located at the respective river gauge. As the catchments are ordered with respect to their mean annual discharge, the blueish (reddish) basins further indicate high (low) discharge.

smooth the time series of the other water cycle variables according to

$$\tilde{F}_i = \frac{1}{4}F_{i-1} + \frac{1}{2}F_i + \frac{1}{4}F_{i+1}\,, \tag{3.12}$$

where the F variables are time series of precipitation, evapotranspiration, moisture flux divergences, and observed and modeled runoff. The indices i, $i-1$, and $i+1$ refer to the previous, current, and following month, respectively. We have applied this filter to all our estimates.

Runoff

In order to attain maximum data coverage, both in space and time, data are collected from different sources: the GRDC; the Arctic Rapid Integrated Monitoring System (ArcticRIMS) project; Water Survey Canada; the U.S. Geological Survey (USGS); the U.S. Army Corps of Engineers (USACE); the Environmental Research Observatory (ORE) Geodynamical, Hydrological, and Biogeochemical Control of Erosion/Alteration and Material Transport in the Amazonas Basin (HYBAM) project; the Department of Water, Land and Biodiversity Conservation, Government of Australia; and the Department of Water Affairs and Forestry, Republic of South Africa.

TABLE 3.1: Summary of the observation and model based datasets containing precipitation (P), actual evapotranspiration (ET_a), vertically integrated moisture flux divergences ($\nabla \cdot \mathbf{Q}$), water storage changes ($\mathrm{d}M/\mathrm{d}t$), and runoff (R_{obs} and R_{mod}).

Variable	Dataset	Version	Resolution		Time period
			Spatial	Temporal	
P	GPCC	6.0	0.5° × 0.5°	1 mo	1901–2010
	GPCP	2.2	2.5° × 2.5°	1 mo	1979–present
	CPC	1.0	0.25° × 0.25°	1 d	1979–present
	CRU	3.1	0.5° × 0.5°	1 mo	1901–2009
	DEL	2.01	0.5° × 0.5°	1 mo	1900–2008
ET_a	ERA-Interim	–	0.75° × 0.75°	1 mo, 1 d, 6 h	1979–present
	GLDAS	Noah 2.7	0.25° × 0.25°	1 mo, 3 h	1979–present
	GLEAM	v1B	0.25° × 0.25°	1 d	1984–2007
	MODIS	–	0.5o × 0.5°	1 d	–
	MERRA	–	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 mo, 1 d, 6 h, 3 h, 1h	1979–present
	MERRA Land	1.0	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 mo, 1 d, 1 h	1980–present
$\nabla \cdot \mathbf{Q}$	ERA-Interim	–	0.75° × 0.75°	1 mo, 1 d, 6 h	1979–present
	MERRA	–	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 mo, 1 d, 6 h, 3 h, 1 h	1979–present
	CFSR	–	0.5° × 0.5°	1 mo, 1 d, 6 h	1979–2011
$\mathrm{d}M/\mathrm{d}t$	GRACE GFZ	R5	–	1 mo	2005–present
	GRACE CSR	R5	–	1 mo	2005–present
R_{obs}	GRDC	–	–		
R_{mod}	GLDAS	VIC	1.0° × 1.0°	1 mo, 3 h	1979–present
	GLDAS	CLM 2.0	1.0° × 1.0°	1 mo, 3 h	1979–present
	GLDAS	Noah 2.7	0.25° × 0.25°	1 mo, 3 h	1979–present
	GLDAS	Noah 3.3	1.0° × 1.0°	1 mo, 3 h	1979–present
	GLDAS	Mosaic	1.0° × 1.0°	1 mo, 3 h	1979–present
	MERRA Land	1.0	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 mo, 1 d, 1h	1980–present

The GRACE-based approaches (i.e., R_{hyd} and R_{atm}) are essentially global methods capable of providing runoff estimates of all catchments without the necessity of any runoff observations. To validate these estimates, we need contemporaneous (2003-–10) observed runoff data for all the catchments. Because of data gaps both in space and time, we choose catchments that have a minimum of five years of observed runoff data within the period 2003—2010 so that we can perform a meaningful statistical assessment. Applying this restriction, we are left with 96 catchments out of a total of 403 catchments to perform our investigation. An overview over the different study regions, the respective river gauges, the area of the basins, and the mean annual discharge is given in Table 3.2 and Figure 3.2. The number of available runoff observations is shown in Figure 3.3, where the white and gray boxes indicate available and missing data, respectively.

Spatially, the collection of catchments covers a wide range of climatic and geographic conditions. This includes some tropical catchments with the highest discharge worldwide (e.g., Amazonas and Congo) and also the vast high-latitude basins (e.g., Lena, Ob, and Yana) with typically low discharge values during the winter months due to frozen conditions and peak values between June and August during and after the melting season. Also, the areal extent of the selected catchments varies from 4.7×10^6 to $2.1 \times 10^4\,\mathrm{km}^2$, where we have deliberately included the small catchments ($< 200,000\,\mathrm{km}^2$) as well. This allows the quantification of the area dependence of GRACE-based estimates. To better characterize the study regions, Figures 3.4 and 3.5 show the annual cycle of precipitation, evapotranspiration, runoff, water-storage changes, and moisture flux convergences for all the basins.

Precipitation

Precipitation is traditionally measured using various types of rain gauges, such as the nonrecording cylindrical container, the recording weighing gauge, floating gauges, and tipping-bucket-type gauges. In the recent past, precipitation was also measured with weather radar and weather satellites. Generally, because of the adequate amount of sensors, global gridded precipitation data are available from different data providers. In this study, we used precipitation data from the following:

- Global Precipitation Climatology Center [GPCC; Schneider et al., 2008]
- Global Precipitation Climatology Project [GPCP; Adler et al., 2003]
- Climate Prediction Center (CPC) "Unified" precipitation [Chen et al., 2008]
- Climatic Research Unit [CRU; Harris et al., 2014], and
- University of Delaware [DEL; Matsuura and Willmott, 2012]

From these datasets, GPCC, CPC, CRU, and DEL are derived from gauge observations only. The largest number of stations are included in the GPCC product, while CPC archives a higher station density, for example, over North America. However, because of the very nature of such data products, the precipitation estimates can be highly uncertain, especially in regions with only a few or no stations. Figure 3.1 clearly shows that large parts over, for example, South America, Africa, or Siberia are completely unobserved. Moreover, the number and distribution of rain gauges varies significantly over time, which introduces a further source of uncertainty. On the other hand, observations are provided in a relatively high spatial resolution and can be assumed to be of high quality over data-rich regions such as Europe or large parts of North America. For the spatially lower resolved GPCP product, data from different rainfall sensors have been merged in order to provide a more consistent data source for global (land and ocean) precipitation. Currently, the GPCP includes microwave, infrared, and gauge-based observations of precipitation [e.g., Huffman et al., 2009].

Evapotranspiration

A wide variety of model approaches for estimating evapotranspiration exist, which range from simple empirical ones to complex ones including radiative energy balance. Many researches have focused on remote sensing approaches, in which the observation of surface temperature and vegetation indices are used for estimation of evapotranspiration. In this study, we have used data from the following:

- Global Land-surface Evaporation: the Amsterdam Methodology [GLEAM; Miralles et al., 2011b]
- GLDAS, version 2 [GLDAS2; Rodell et al., 2004b]
- Interim European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis [ERA-Interim; Berrisford et al., 2009, Dee et al., 2011]
- the land-only version of the Modern-Era Retrospective Analysis for Research and Applications [MERRA Land; Reichle et al., 2011]
- FLUXNET Multi-Tree Ensemble [FLUXNET MTE; Jung et al., 2009, 2010], and
- Moderate Resolution Imaging Spectroradiometer (MODIS) Global Evapotranspiration Project [MOD16; Mu et al., 2007, 2011]

GLEAM, FLUXNET MTE, and MOD16 are evapotranspiration-only products, that is, they should, in principle, provide the most reliable estimates of global evapotranspiration. GLEAM is based on multi-satellite observations and combines a wide range of remotely sensed observations within a Priestley–Taylor-based framework [Miralles et al., 2011b]. In contrast to this multi-satellite approach, MOD16 is based on the Penman–Monteith method and is derived by combining remote sensing data from MODIS and global meteorological data [Mu et al., 2007]. The dataset is available in a very high spatial resolution of 1 km. For this study, we use the coarser $0.5^\circ \times 0.5^\circ$ version, which is also provided by the data center. The FLUXNET MTE dataset provides global evapotranspiration by empirical upscaling of eddy-covariance measurements from the FLUXNET network [Jung et al., 2009].

We also use the evapotranspiration from GLDAS2 (i.e., the Noah3.3 realization) and MERRA Land. A description of these datasets is given in section *Land Surface Models*.

Finally, ERA-Interim is a reanalysis model where a global atmospheric model is assimilated with numerous observations in order to provide reliable estimates of a wide variety of hydrometeorological and hydrological variables (see section *Moisture Flux Divergence*). A comparison of some of these datasets is presented in, for example, Mueller et al. [2011].

To get an impression of the level of agreement between the different precipitation and evapotranspiration datasets, the top row in Figure 3.6 shows scatterplots between the mean annual and seasonal precipitation from GPCC and the other datasets. For the bottom row, evapotranspiration from MOD16 served as reference.

Water Storage Changes

Water storage changes are computed from GRACE observations of the temporal gravity field changes. Eight years of GRACE (release 5) data have been used in this study, and they come from two data processing centers: GeoForschungsZentrum Potsdam [GFZ; Dahle et al., 2013] and Center for Space Research, The University of Texas at Austin [CSR; Bettadpur, 2012]. The data are provided as spherical

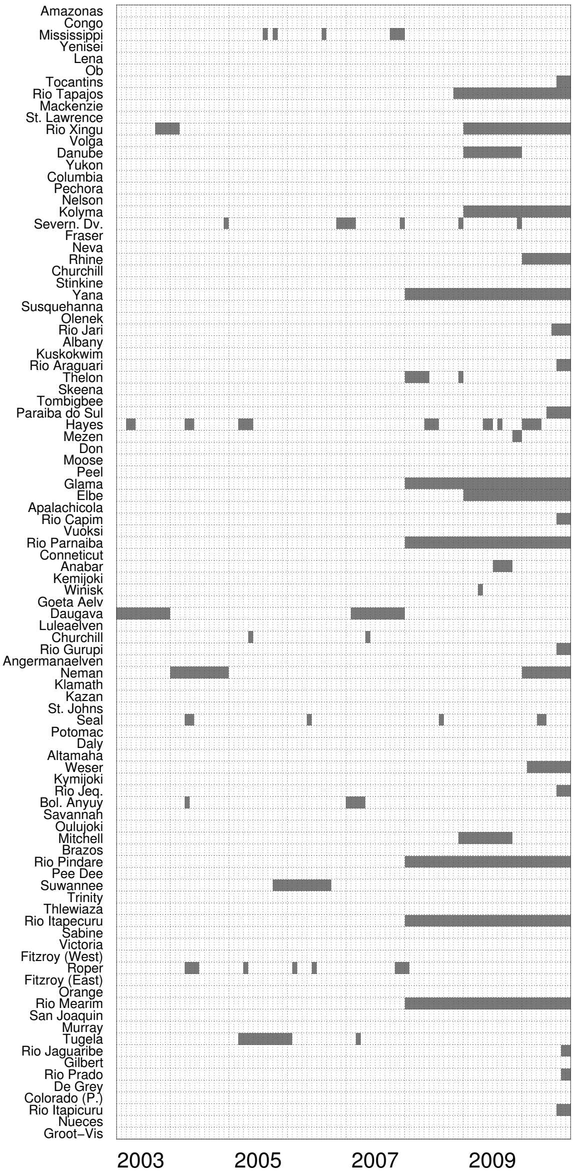

FIGURE 3.3: Overview of the amount of runoff observations for every catchment. Each box corresponds to a single month of a specific catchment. White boxes indicate that observations are available for that month while gray boxes depict missing data.

harmonic coefficients of the monthly gravity field snapshots. Wahr et al. [1998] provide the mathematics to invert the monthly snapshots to surface mass anomalies ΔM. According to them, those surface mass anomalies represent anomalies in water storage over the continental landmasses. For this reason, we will denote the water storage changes estimated from GRACE as $\mathrm{d}M/\mathrm{d}t$. The GRACE data as such are plagued by a lot of errors, which necessitates further processing prior to their usage (see section *Treating the Errors in GRACE Mass Anomalies* in this chapter's appendix for details). The postprocessed ΔM are taken on a 0.5° × 0.5° grid and then aggregated over catchments to derive catchment-specific time series. The time series are then numerically differentiated using three-point central differences to arrive at the required flux quantity, surface mass changes $\mathrm{d}M/\mathrm{d}t$.

In this study, we use three different GRACE-based datasets in order to gain information on differences between products from different data providers or filtering methods. Therefore, the GFZ data have been filtered with both a regularization filter ($\mathrm{GFZ}_{\mathrm{reg}}$; see section *High-Frequency Noise* in this chapter's appendix for details) as well as the widely used decorrelation (or destriping) filter in combination with a Gaussian filter ($\mathrm{GFZ}_{\mathrm{dstr}}$). The data from the CSR have been filtered the regularization filter only ($\mathrm{CSR}_{\mathrm{reg}}$).

Moisture Flux Divergence

Vertically integrated moisture flux divergences can be obtained from global atmospheric reanalysis models. These models are combinations of a numerical model and an analysis scheme where the numerical model predicts the state of the earth's atmosphere, oceans, land surface, and upper soil layers. The analysis scheme, on the other hand, assimilates a large number of observations into the numerical model. Because of the very nature of assimilation, the combination of an analysis scheme forces the estimates to stay both in reasonable physical and numerical bounds, which, on the flip side, disturbs the model's interior mass budgets [Trenberth and Fasullo, 2013, Trenberth et al., 2011].

The present-day reanalyses models assimilate nearly 7×10^6–9×10^6 observations every 6–12 h, which goes to show their data-intensive nature. These reanalyses usually cover a period of multiple decades and provide estimates of the most important hydrological, hydrometeorological, and atmospheric quantities globally. Kalnay et al. [1996] and Kistler et al., 2001 categorize the gridded variables from reanalyses into three distinct classes according to the influence of the assimilated observations on them. The type A variables are mainly influenced by observations (e.g., upper-air temperatures or horizontal winds). Type B variables are influenced by both the observations as well as the model (e.g., surface and 2-m temperatures), while type C variables are solely derived from the model (e.g., precipitation and runoff).

The vertically integrated moisture flux divergences are derived from horizontal wind fields—a type A variable—and specific humidity—a type B variable [see Eq. (3.4)]. Therefore, the moisture flux divergence can be assumed to be more reliable and stable than, for example, modeled evapotranspiration. Here, data from three widely used state-of-the-art reanalysis models have been used: ERA-Interim; MERRA from the Goddard Space Flight Center (GSFC), National Aeronautics and Space Administration [NASA; Rienecker et al., 2011]; and Climate Forecast System Reanalysis [CFSR; Saha et al., 2010] from the National Centers for Environmental Prediction (NCEP), National Oceanic and Atmospheric Administration (NOAA).

Land Surface Models

In this study, we also used monthly runoff and evapotranspiration from five different land surface models driven by the GLDAS [Rodell et al., 2004b] and the MERRA Land [Reichle et al., 2011]. In contrast to traditional hydrological or land surface models, these estimates are strongly constrained by observations (similar to atmospheric reanalyses) in order to provide more reliable estimates of hydrological variables. MERRA Land is an offline (without atmosphere) rerun of a revised version of MERRA's land component, which is forced by the atmospheric reanalysis near-surface fields (air temperature, wind, and humidity), surface pressure, radiation, and observed precipitation based on the CPC precipitation product (see section *Precipitation*). MERRA Land thus provides a supplemental and improved set of land surface hydrological fields [Reichle et al., 2011].

GLDAS can be seen as a data assimilation framework, where different land surface models are combined with satellite- and ground-based observations in order to improve and constrain the simulated hydrological processes. In particular, the models currently driven by the GLDAS are the Noah Land Surface Model [Noah LSM; Chen et al., 1996, Ek, 2003, Koren et al., 1999], the Common Land Model [CLM; Bonan, 1998, Dai et al., 2003, Dickinson et al., 2009], Mosaic [Koster and Suarez, 1996], and the Variable Infiltration Capacity (VIC) Model [Liang et al., 1994]. These four realizations are usually referred to as GLDAS1 and share the same forcing datasets:

- meteorological data from the Global Data Assimilation System [GDAS; Derber et al., 1991]
- downscaled precipitation from the CPC's Merged Analysis of Precipitation [CMAP; Xie and Arkin, 1997], and
- satellite-derived downward radiation from the Air Force Weather Agency Agricultural Meteorology Modeling System [AFWA AGRMET; Kopp and Kiess, 1996].

An overview of the different models and model forcings for GLDAS1 is given in Rodell et al. [2004b]. The first GLDAS2 dataset was recently released and provides hydrological variables for the period 1948-–2010. This version is currently based on the Noah3.3 LSM and uses forcing data from the Princeton meteorological forcing data [Sheffield et al., 2006] only. In this study, we use the total runoff estimates from the four GLDAS1 and the single GLDAS2 realizations.

It should be noted that these models do not consider a routing scheme for estimating the runoff (R. H. Reichle 2013, personal communication; M. Rodell 2012, personal communication). Therefore, integrating fields of surface and subsurface runoff over a large area on monthly or shorter time scales might lead to inaccuracies in, for example, peak runoff as varying flow directions and velocities (due to the topography of the basin) are not considered. As a consequence, the impact of such routing schemes reduces when considering longer time scales (seasonal and annual). Therefore, for this study, we have used the modeled monthly runoff as provided by the data centers without applying any routing scheme in order to analyze the quality of these estimates on monthly and longer time scales.

TABLE 3.2: River basins which were used in this study including the respective river gauge, its location, the area of the basin, and the mean annual discharge as provided from GRDC.

	River	Station	Lat.	Lon.	Area [km^2]	Discharge [m^3/s]
1	Amazonas	Obidos	-1.95	-55.51	4672876	176498
2	Congo	Kinshasa	-4.3	15.3	3615546	40108
3	Mississippi	Vicksburg	32.32	-90.91	2938538	20793
4	Yenisei	Igarka	67.48	86.5	2454961	20030
5	Lena	Kyusyr	70.7	127.65	2417932	19261
6	Ob	Salekhard	66.57	66.53	2926321	12483
7	Tocantins	Tucurui	-3.76	-49.65	752993	10291
8	Rio Tapajos	Fortaleza	-6.05	-57.64	366843	9773
9	Mackenzie	Arctic Red River	67.46	-133.75	1666073	9707
10	St. Lawrence	Lasalle	45.42	-73.62	943769	8106
11	Rio Xingu	Altamira	-3.22	-52.21	445289	7982
12	Volga	Volgograd Power Plant	48.81	44.59	1345070	7778
13	Danube	Ceatal Izmail	45.22	28.72	771277	7091
14	Yukon	Pilot Station, Ak	61.93	-162.88	819635	6403
15	Columbia	Beaver Army Terminal	46.18	-123.18	662542	6102
16	Pechora	Oksino	67.6	52.2	304670	5139
17	Nelson	Long Spruce Gen. Stat.	56.4	-94.37	1126480	3822
18	Kolyma	Kolymskaya	68.73	158.72	421802	3546
19	Severnaya Dvina (N)	Ust-Pinega	64.15	41.92	330709	3453
20	Fraser	Hope	49.38	-121.45	228874	2533
21	Neva	Novosaratovka	59.84	30.53	225651	2473
22	Rhine	Rees	51.75	6.4	169706	2096
23	Churchill, Fleuve (Lab.)	Above upper Muskrat Falls	53.25	-60.79	86658	1736
24	Stinkine	Near Wrangell	56.71	-132.13	52894	1562
25	Yana	Ubileynaya	70.75	136.08	220949	1300
26	Susquehanna	Conowingo	39.66	-76.18	69694	1288
27	Olenek	7.5 km d/s of mouth of Pur	72.12	123.22	199723	1255
28	Rio Jari	Sao Francisco	-0.57	-52.57	46355	1208
29	Albany	Near Hat Island	51.33	-83.84	97926	1154
30	Kuskokwim	Crooked Creek	61.87	-158.1	82074	1145
31	Rio Araguari	Porto Platon	0.71	-51.44	21630	1050
32	Thelon	below outlet of Schultz Lake	64.78	-97.05	171346	912
33	Skeena	Usk	54.63	-128.43	31555	886
34	Tombigbee	Coffeeville	31.76	-88.13	51622	854
35	Paraiba do Sul	Campos - Ponte Municipal	-21.75	-41.3	60146	728
36	Hayes (trib. Hud. Bay)	Below Goods River	56.43	-92.79	97802	707
37	Mezen	Malonisogorskaya	65.03	45.62	54125	693
38	Don	Razdorskaya	47.54	40.65	378180	669
39	Moose (trib. Hud. Bay)	Above Moose River	50.74	-81.46	91157	661
40	Peel (trib. Mackenzie)	Above Fort McPherson	67.25	-134.88	63040	661
41	Glama	Langnes	59.61	11.12	38534	658
42	Elbe	Neu Darchau	53.23	10.89	134037	645

Continued on next page

Table 3.2 – *Continued from previous page*

	River	Station	Lat.	Lon.	Area [km^2]	Discharge [$^{\mathrm{m}^3}/_{\mathrm{s}}$]
43	Apalachicola	Near Sumatra	29.95	-85.02	57546	633
44	Rio Capim	Badajos	-2.51	-47.77	37004	615
45	Vuoksi	Tainionkoski	61.22	28.78	66393	612
46	Rio Parnaiba	Luzilandia	-3.45	-42.37	297049	606
47	Conneticut	Thompsonville	41.99	-72.61	24596	590
48	Anabar	Saskylakh	71.98	114.06	79786	578
49	Kemijoki	Isohaara	65.78	24.55	53763	543
50	Winisk	Below Ashweig River Trib.	54.52	-87.23	84526	535
51	Goeta Aelv	Vargoens KRV	58.36	12.38	48265	530
52	Daugava	Daugavpils	55.88	26.53	61576	501
53	Lule	Bodens KRV	65.81	21.67	22885	497
54	Churchill	Above Red Head Rapids	58.12	-94.63	299391	489
55	Rio Gurupi	Alto Bonito	-1.8	-46.32	33933	473
56	Angermanaelven	Solleftea KRV	63.17	17.27	33376	467
57	Nemunas - Neman	Smalininkai	55.08	22.58	89806	458
58	Klamath	Near Klamath	41.51	-124.0	36896	451
59	Kazan	Above Kazan Falls	63.65	-95.85	40945	450
60	St. Johns	Jacksonville	30.32	-81.67	27007	427
61	Seal	Below Great Island	58.89	-96.28	49662	397
62	Potomac	Near Washington, D.C.	38.95	-77.13	31151	380
63	Daly	Mount Nancar	-13.83	130.74	44883	357
64	Altamaha	Doctortown	31.66	-81.83	33706	346
65	Weser	Intschede	52.96	9.13	36293	311
66	Kymi	Anjala	60.7	26.82	33195	281
67	Rio Jequitinhonha	Itapebi	-15.95	-39.52	65067	274
68	Bolshoy Anyuy	Konstantinovo	68.15	161.17	50025	272
69	Savannah	Near Clyo	32.53	-81.27	23055	268
70	Oulujoki	Meriskoski	65.02	25.52	26517	259
71	Mitchell	Koolatah	-15.95	142.38	44381	247
72	Brazos	Richmond	29.58	-95.76	106914	243
73	Rio Pindare	Pindare-Mirim	-3.66	-45.46	33895	237
74	Pee Dee	PeeDee	34.2	-79.55	22686	236
75	Suwannee	Near Bell	29.79	-82.92	21279	233
76	Trinity (Texas)	Romayor	30.43	-94.85	44365	225
77	Thlewiaza	Above outlet Sealhole Lake	60.79	-98.78	53705	219
78	Rio Itapecuru	Cantanhede	-3.63	-44.38	58451	213
79	Sabine	Near Ruliff	30.3	-93.74	23626	195
80	Victoria	Coolibah Homestead	-15.56	130.96	44380	191
81	Fitzroy (west Australia)	The Gap	-23.09	150.11	126986	190
82	Roper	Red Rock	-14.7	134.42	50687	188
83	Fitzroy (east Australia)	Fitzroy Crossing	-18.21	125.58	44147	144
84	Orange	Vioolsdrif	-28.76	17.72	828475	104
85	Rio Mearim	Bacabal	-4.22	-44.77	27688	102
86	San Joaquin	Near Vernalis	37.68	-121.27	29459	97
87	Murray	Overland Corner	-34.17	140.28	1022767	88

Continued on next page

Table 3.2 – *Continued from previous page*

	River	Station	Lat.	Lon.	Area [km^2]	Discharge [$^{m^3}/_{s}$]
88	Tugela	Mandini	-29.14	31.39	29848	76
89	Rio Jaguaribe	Peixe Gordo	-5.23	-38.2	55288	56
90	Gilbert	Rockfields	-18.2	142.88	32248	53
91	Rio Prado	Mascote	-15.56	-39.31	38733	52
92	De Grey	Coolenar Pool	-20.31	119.25	51695	33
93	Colorado (Pacific)	At Yuma	32.73	-114.63	636508	29
94	Rio Itapicuru	Usina Altamira	-11.74	-37.8	36421	22
95	Nueces	Calallen	27.88	-97.63	43371	16
96	Groot-Vis	Matolemas Location Out.	-33.24	27.0	28741	11

Results and Discussion

Performance Metrics

We compare the time series of runoff estimates from GRACE-based approaches with the observed runoff. To get a holistic understanding, we also compare them with the estimates from hydrological models and the runoff–precipitation ratio method. Such an analysis should provide us a balanced view on the utility of the GRACE-based estimates in terms of already available global-scale alternatives. In the current setup, we use the following three metrics to carry out our evaluation: correlation coefficient, percentage bias [PBIAS; Gupta et al., 1999] and Nash-Sutcliffe efficiency (NSE) coefficient [Nash and Sutcliffe, 1970].

The correlation coefficient quantifies the level of common information content between observed and estimated runoff, although, as is well known, it is insensitive to the amplitude and the mean value. For the correlation analysis, we only show correlations above 0.25, which is approximately the 5 % significance level with respect to 60 samples (i.e., 5 years of data). The PBIAS,

$$\mathrm{PBIAS} = \frac{\sum_{t=1}^{T}\left(R_{t,\mathrm{obs}} - R_{t,\mathrm{est}}\right)}{\sum_{t=1}^{T} R_{t,\mathrm{obs}}} = 1 - \frac{\overline{R}_{\mathrm{est}}}{\overline{R}_{\mathrm{obs}}}, \tag{3.13}$$

on the other hand, gives precisely this information about the relative difference between the long-term mean of the observations $\overline{R}_{\mathrm{obs}}$ and estimates $\overline{R}_{\mathrm{est}}$. Finally, correlation and PBIAS are summarized by the NSE coefficient:

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T} \left(R_{t,\mathrm{obs}} - R_{t,\mathrm{est}}\right)^2}{\sum_{t=1}^{T} \left(R_{t,\mathrm{obs}} - \overline{R}_{\mathrm{obs}}\right)^2} . \qquad (3.14)$$

The NSE is a conservative quantity, and it is highly sensitive to the overall agreement between the observations and estimates: phase, amplitude, and mean. It ranges between $-\infty$ and 1, where a value of 1 indicates that the estimated and observed runoff agree in terms of all the three variables mentioned above. Further, the NSE can be treated as relative imbalance at monthly scales for the R_{hyd} and R_{atm} methods as it involves the scaled mean of the monthly squared error.

Notes on the Interpretation of the Results

The sheer number of combinations, due to the choice of multiple datasets for each of the budget components, forms an ensemble that helps to provide a broad overview of the approach in hand, but also makes the assessment a little cumbersome. The contributing datasets are shown in Figure 3.7 together with a color code. The performance of the different approaches and data combinations at the monthly time scale is summarized in Figures 3.8, 3.11, and 3.12. The upper part of these figures collects the values of the performance metrics into a matrix, where each cell denotes the individual performance values for each catchment and each estimate separately. In other words, each row of the matrix provides an overview of the performance of different combinations for a particular catchment, and each column provides an overview for a particular combination. The lower part summarizes each column of the matrix by sorting the performance metric values within prescribed intervals. The color scales in these matrices have been chosen in such a manner that darker colors indicate better performance and lighter colors indicate poorer performance. The

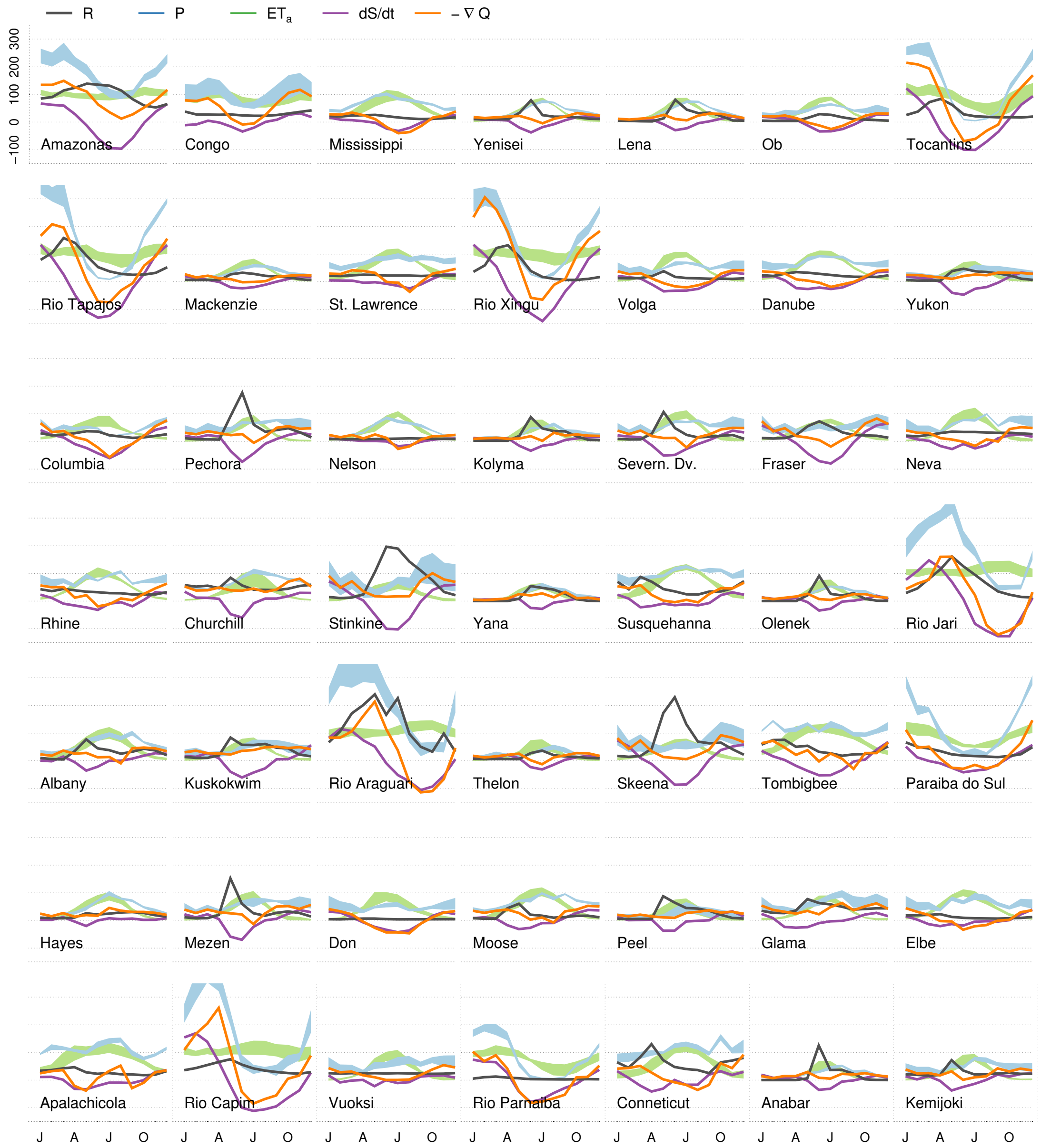

FIGURE 3.4: Mean annual cycle of the different budget quantities. The precipitation bands are derived from the maximum and minimum estimates of GPCC, GPCP, CRU, DEL, and CPC while the evapotranspiration bands are given by GLDAS, ERA-Interim, GLEAM, MERRA Land, MODIS, and FLUXNET MTE. Water storage changes, moisture flux convergences (i.e., negative divergences), and runoff are taken from the CSR$_{reg}$, ERA-Interim, and GRDC dataset, respectively.

datasets of a single combination can be identified with the aid of the color code and the legend (Figure 3.7).

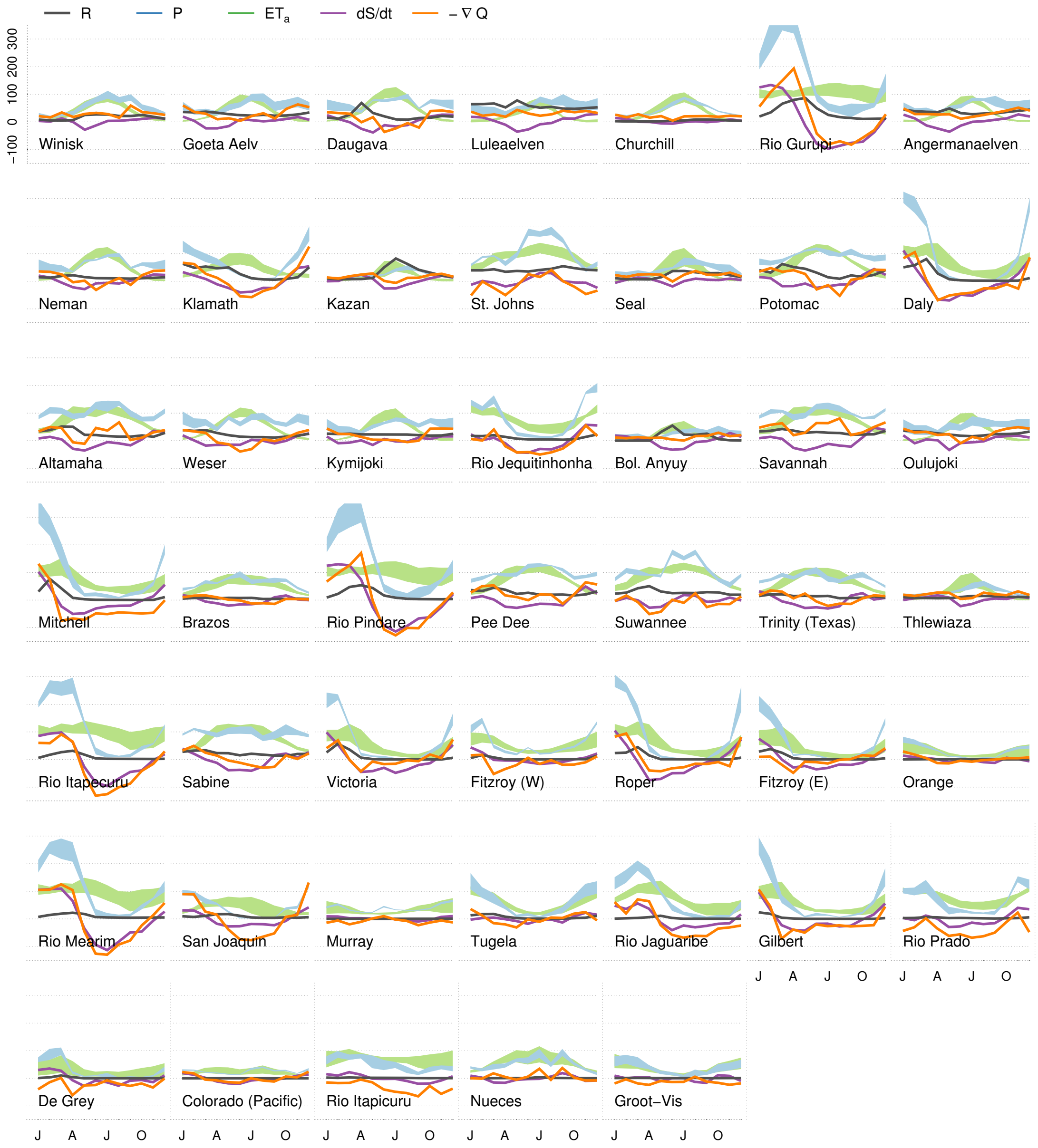

FIGURE 3.5: As in Figure 3.4, but for catchments 50–96.

Runoff from Terrestrial Water Budget: R_{hyd}

The runoff estimates from the terrestrial water budget equation is the largest group of the four different runoff estimates compared here. This is obvious because the runoff estimates from the terrestrial water balance equation require three inputs: precipitation, evapotranspiration, and water storage change. The performance metric PBIAS (Figure 3.8) shows a random behavior with no data combination providing an outright good performance consistently over all the catchments. However, for some catchments (Amazonas, Rio Tapajos, and Rio Jari) R_{hyd} provides consistently good

PBIAS values ($\leq 25\,\%$). Also, for most of the catchments ($\approx$ 60) there is at least one data combination that provides a PBIAS value of nearly zero (see Figure 3.9). In other words, there is at least one data combination that provides a reliable mean runoff value for most of the catchments. In all this, the MERRA Land dataset performs consistently poorly except when it is combined with GPCP (see stacked bars in Figure 3.8). The possible reason could be that GPCP is the wettest of all the precipitation datasets and MERRA Land is the wettest of all the evapotranspiration datasets (see Figure 3.6), and therefore their biases get canceled out. But when combined with the other datasets, the bias in MERRA Land becomes visible. Similarly, GPCP does not perform very well when combined with MOD16 and FLUXNET MTE, while other precipitation datasets perform very well with them. The other interesting feature of the PBIAS matrix is the consistently poor performance in the Arctic catchments (Yenisei, Lena, Ob, Yukon, Thelon, Anabar, Kazan, Mackenzie, Kolyma, Churchill, Yana, Luleaelven, Olenek, Kuskokwim, and Peel), where there is a substantial underestimation of the runoff values (also see Figure 3.10). However, we also observe that in these catchments the data combinations that involve GPCP are relatively well behaved. This better performance might be due to the gauge undercatch correction performed in the GPCP dataset, which provides higher values for the precipitation values in the Arctic catchments.

In contrast to PBIAS, the correlation matrix (Figure 3.11), though random to an extent, shows a homogeneous picture. In nearly 50 % of the catchments, the correlation values are consistently more than 0.5 irrespective of the data combination, which demonstrates that most of the data combinations capture the dynamics of the catchment runoff to a large extent. Further, the Arctic and tropical catchments provide strong correlations with many data combinations, having a correlation upward of 0.75. Some catchments, however, show strong dependency on the data combination (e.g., Mackenzie and Bolshoy Anyuy provide very good correlation when using data combinations involving $\text{GLDAS}_{\text{Noah3.3}}$). Similarly, the tropical catchments are strongly influenced by the choice of precipitation dataset, where the best performance is given by the use of GPCC and GPCP. As in the case of PBIAS, the MERRA Land dataset produces the lowest correlation values. Lowest correlations

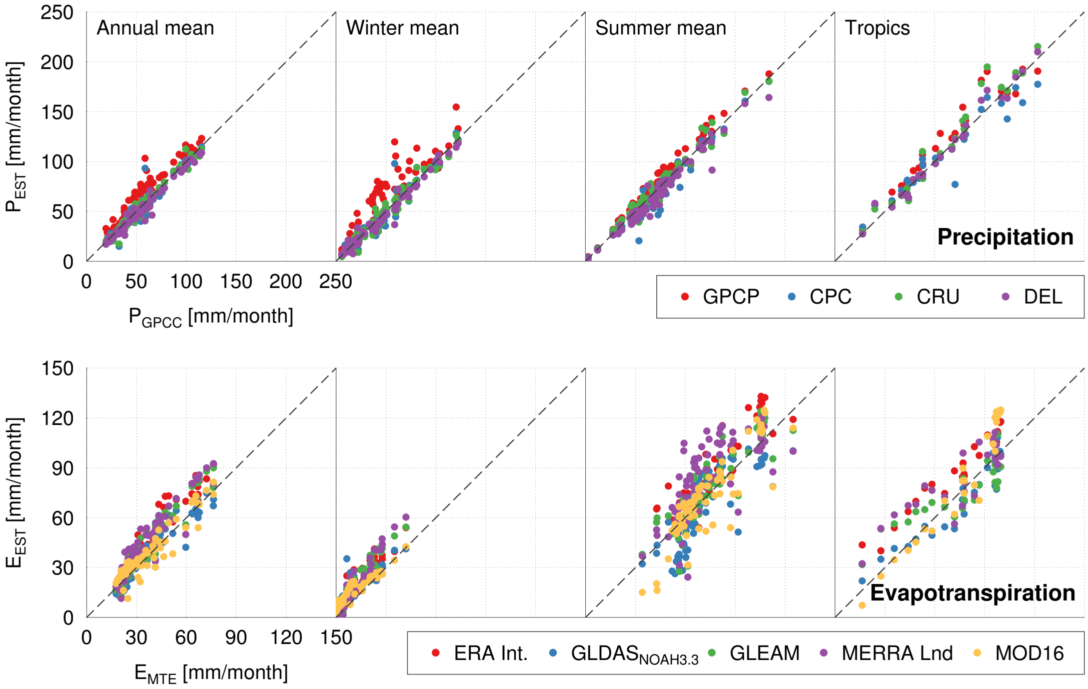

FIGURE 3.6: Scatter plots between the different datasets of precipitation and evapotranspiration over the study regions with respect to GPCC (precipitation) and FLUXNET MTE (evapotranspiration). The first three columns show three different time scales: Long-term annual mean, winter-mean, and summer-mean. The tropical catchments are removed from these scatter plots. The winter-mean was derived from the average of DJF for the northern hemisphere catchments and JJA for southern hemisphere catchments. Similarly, depending on the hemispherical location of the catchments, their respective summer months were used for computing the summer-mean. The fourth column shows the long term mean for the 25 tropical catchments.

are also found in catchments (e.g., St. Lawrence, Neva, Don, Vuoksi, San Joaquin, Orange, Murray, and Colorado) where the runoff does not have a clear seasonal cycle and is constant throughout the year. These catchments are affected by the biases and errors in the datasets, which produce short-term artifacts, thus reducing the correlation with observed runoff data (see supplemental material). An interesting aspect of Figure 3.11 is the influence of the choice of GRACE dataset in the Arctic catchments (e.g., Yenisei, Ob, Pechora, Yana, Anabar, Stinkine, and Peel). The reason for the influence of the water storage dataset is that there is a clear time shift between precipitation and evapotranspiration in the Arctic catchments (see supplementary material), and therefore, the water storage changes have an important role to play in the runoff estimation. Nevertheless, it should also be noted that there is no single GRACE dataset that consistently performs well in all of the Arctic catchments. This

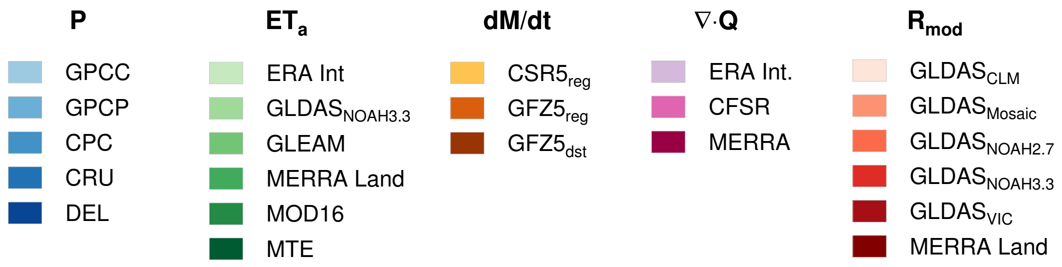

FIGURE 3.7: Legend for the color codes in Figures 3.8, 3.11, and 3.12.

is also corroborated by the histograms in Figure 3.13.

NSE, as mentioned before, summarizes the PBIAS and correlation metrics, where a data combination should have less bias, and also must capture the dynamics of the catchment runoff. It is clearly evident that no data combination fulfills the twin requirements over all the catchments (see Figure 3.12). The Amazonas is the only catchment that provides a value of more than 0.75 and also only for select combinations. For example, the combinations involving CPC do not even produce a positive NSE value. Some reasonable (positive) values are provided by Rio Tapajos, Rio Xingu, Pechora, Severnaya Dvina, Stinkine, Tombigbee, Mezen, and Anabar over a variety of data combinations. Interestingly, many of these catchments are quite small (see Table 3.1) and are considered indiscernible by the GRACE mission. In spite of this optimism, Figure 3.12 clearly demonstrates that the current status of global hydrological datasets does not allow a reasonable estimation of runoff globally. The main weakness of these datasets is the biases in the individual datasets, which limit the ability of the terrestrial water budget equation to estimate runoff.

To study any missing patterns in the matrix plots, we also study the impact of a particular dataset in all the data combinations in which it was involved. This was carried out via histogram analysis as shown in Figure 3.13. The histogram clearly shows that for all the datasets involved in the R_{hyd} approach, nearly 40 % of the dataset provides a correlation of more than 0.5, but only 30 % of the datasets have a relative bias of less than 25 %. This again confirms that the major problem of the R_{hyd} method lies in the biases in the individual datasets. Further, these histograms also indicate that CPC precipitation and MERRA Land evapotranspiration do not provide good estimates as they consistently perform poorly in all three metrics. The

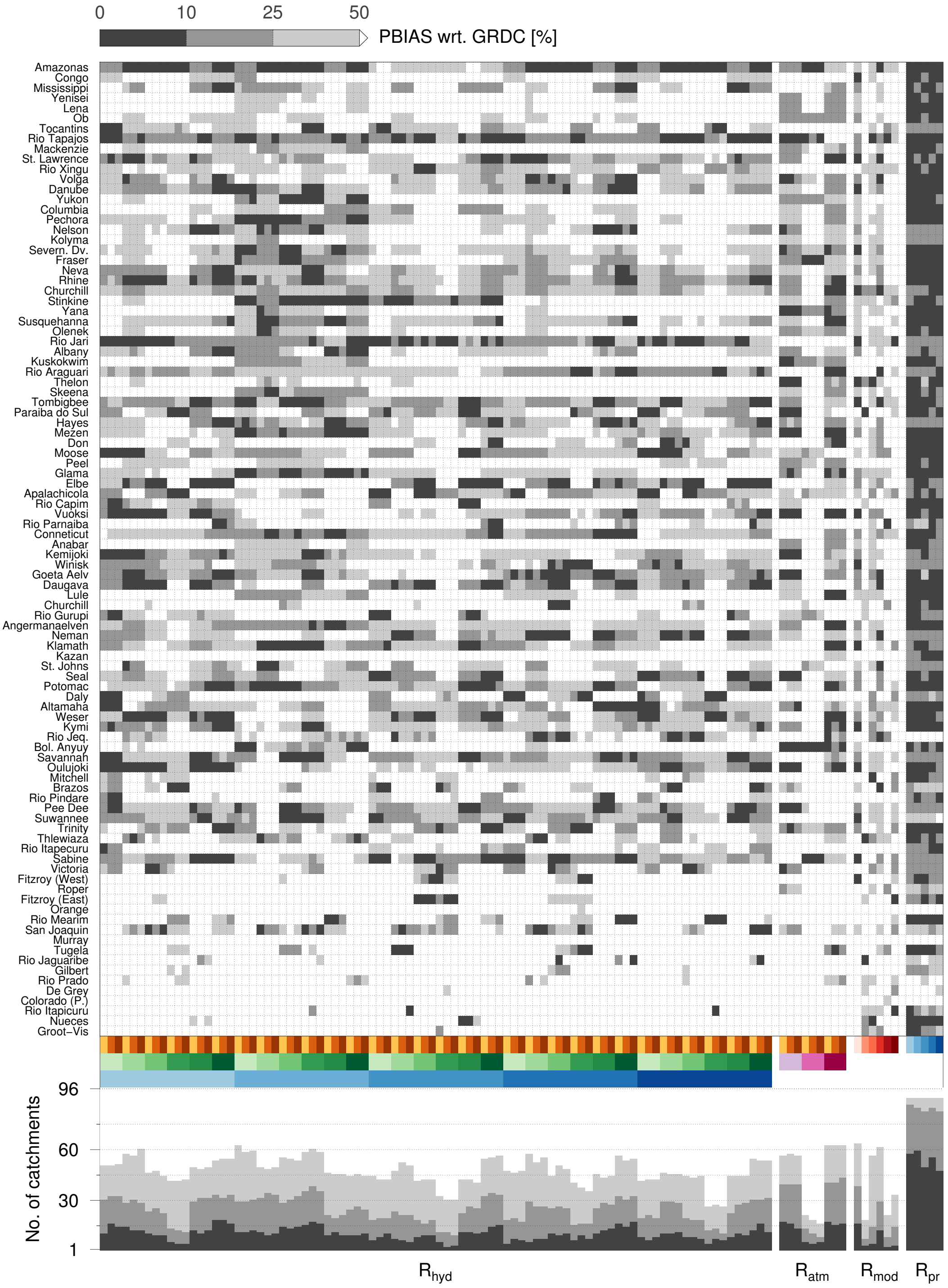

FIGURE 3.8: PBIAS for the 96 study regions between observed and estimated runoff. A guide line on how to read the figure is given in section *Notes on the Interpretation of the Results* while the color codes can be identified with Figure 3.7.

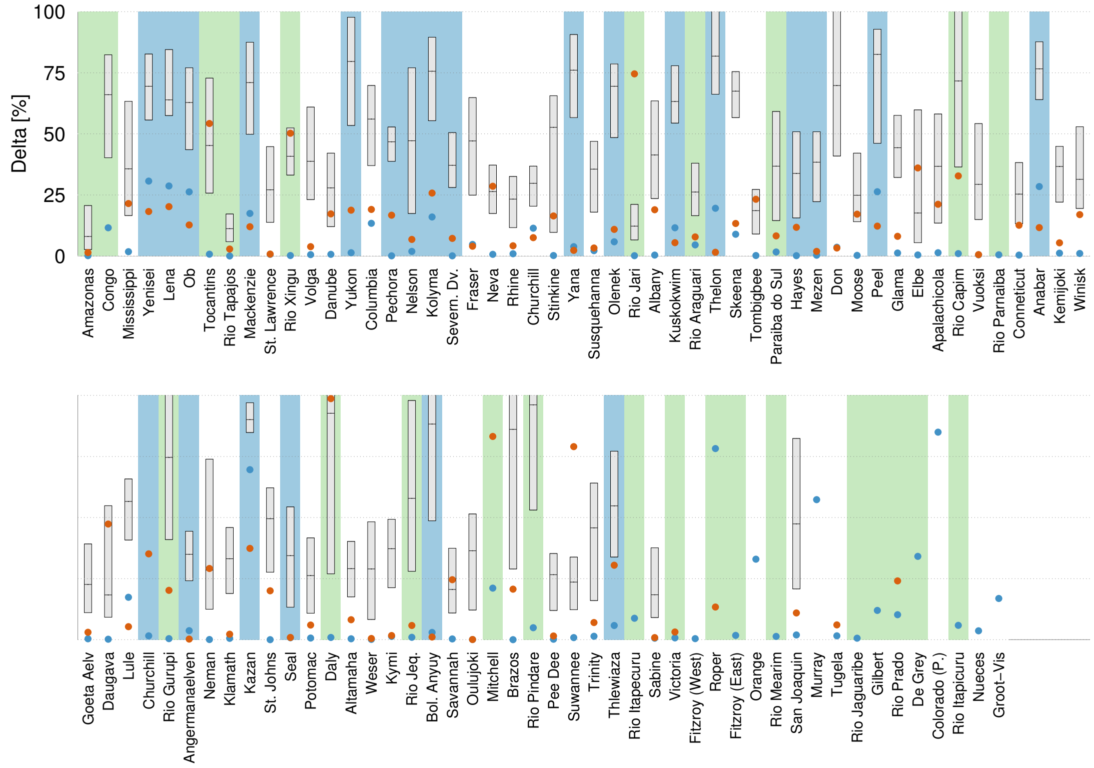

FIGURE 3.9: Box-plot showing the relative water cycle imbalance with respect to the observed mean annual runoff. The lower and upper bounds of the gray boxes depict the 25th and 75th percentile of the 90 ensemble members of R_{hyd} while the line within the boxes shows the median imbalance. The blue (red) dots are the minimum imbalances from R_{hyd} (R_{atm}). The background colors indicate arctic (blue) and tropical (green) catchments. The arctic catchments have been chosen according to the entries in the ArcticRIMS database. For the tropical regimes, we have selected the basins within $-23.5°$ and $+23.5°$ latitude.

GRACE datasets differ predominantly in the way they help capture the dynamics, but not otherwise. The biggest differences between the datasets of one variable is seen in evapotranspiration, where $\text{GLDAS}_{\text{Noah3.3}}$, MOD16, and FLUXNET MTE perform consistently good in all three metrics.

Runoff from Atmospheric Water Budget: R_{atm}

The R_{atm} approach uses only two data sources, and therefore, we can expect a better performance when compared with R_{hyd} as there is one less source of potential errors. However, the PBIAS values from R_{atm} (Figure 3.8) show a similar random behavior like the estimates from R_{hyd}. Also, the number of catchments with PBIAS values $\leq 10\,\%$ is similar compared to R_{hyd} (see stacked bars in Figure 3.8). On the other

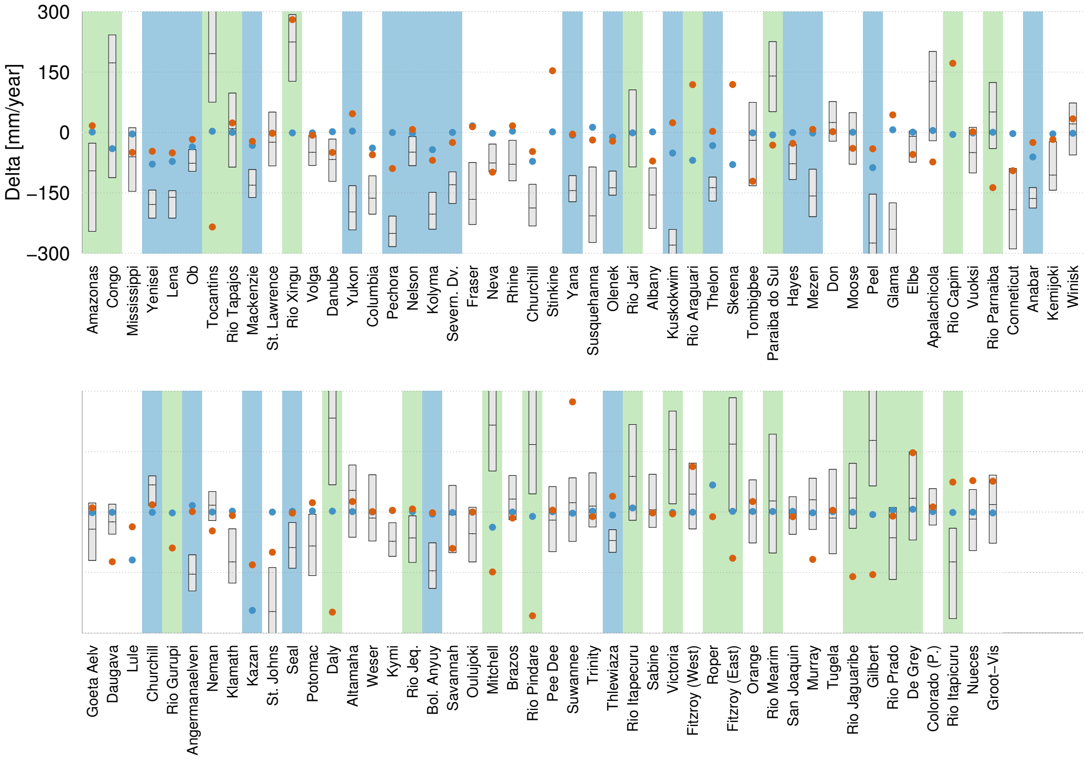

FIGURE 3.10: As in Figure 3.9, but for the absolute water cycle imbalance with respect to the observed mean annual runoff.

hand, there are slightly more catchments with bias values between 10 and 25 %, which is mainly due to a better performance over Arctic catchments (e.g., Yenisei, Lena, Ob, Kuskokwim, and Thelon). Here, while R_{hyd} tends to underestimate runoff (cf. blue and orange dots in Figure 3.10), R_{atm} shows a better water budget closure with remaining relative imbalances of less than 25 %. As many precipitation datasets suffer from an underestimation of precipitation in high-latitude regions [e.g., due to gauge undercatch errors; Lorenz and Kunstmann, 2012], moisture flux divergences might be more reliable than those derived from precipitation and evapotranspiration over these regions. Of the three moisture flux divergence datasets, ERA-Interim and MERRA show a similar overall performance while CFSR reaches PBIAS values ($\leq$ 25 %) over less than 30 catchments, indicating large biases in the reanalysis model.

In terms of correlation (Figure 3.11), R_{atm} shows slightly worse results than the best combinations from R_{hyd}. However, there are still about 45 catchments where moisture fluxes from ERA-Interim and MERRA reach correlations ≥ 0.5, indicating

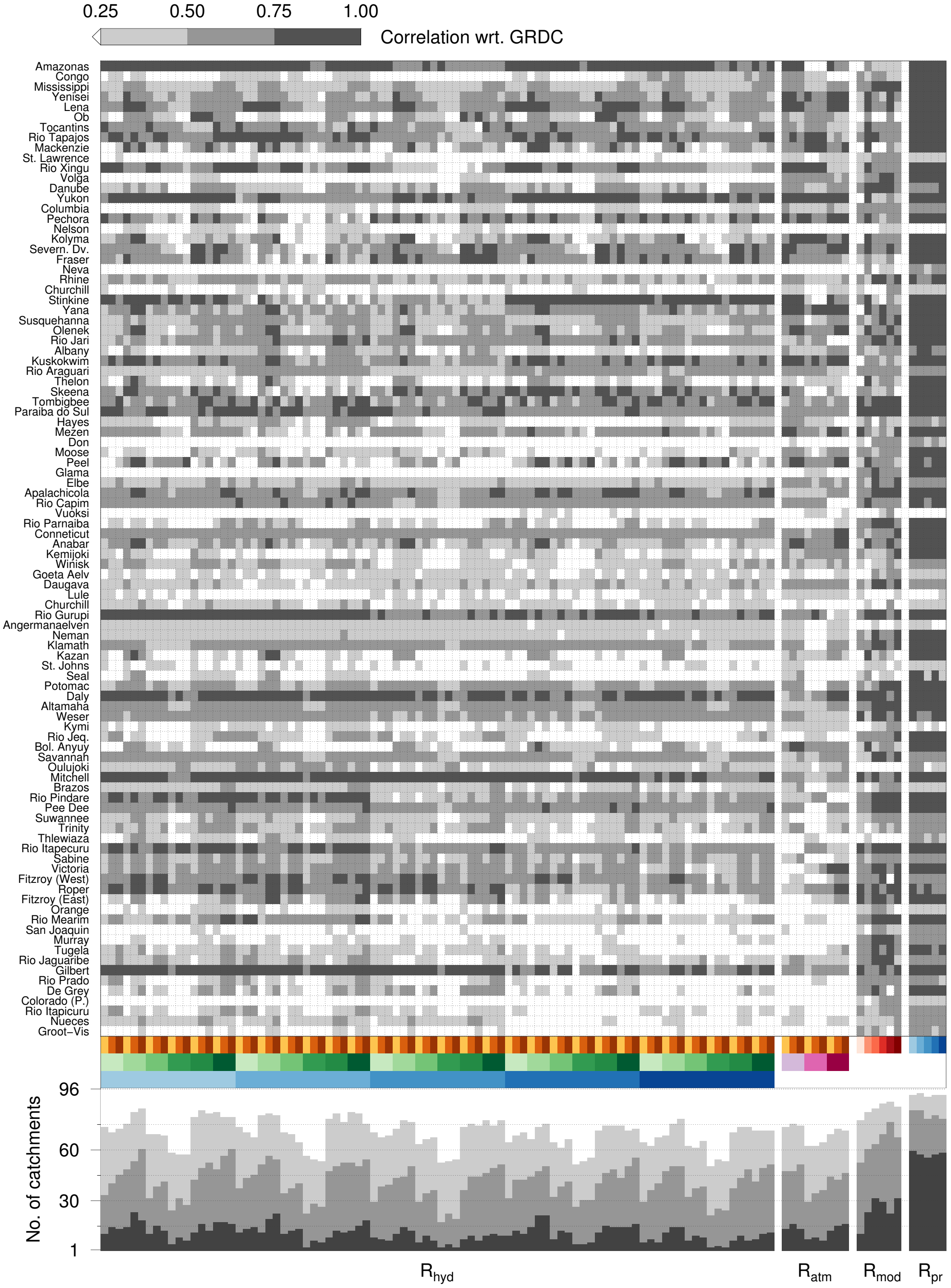

FIGURE 3.11: Correlation for the 96 study regions between observed and estimated runoff.

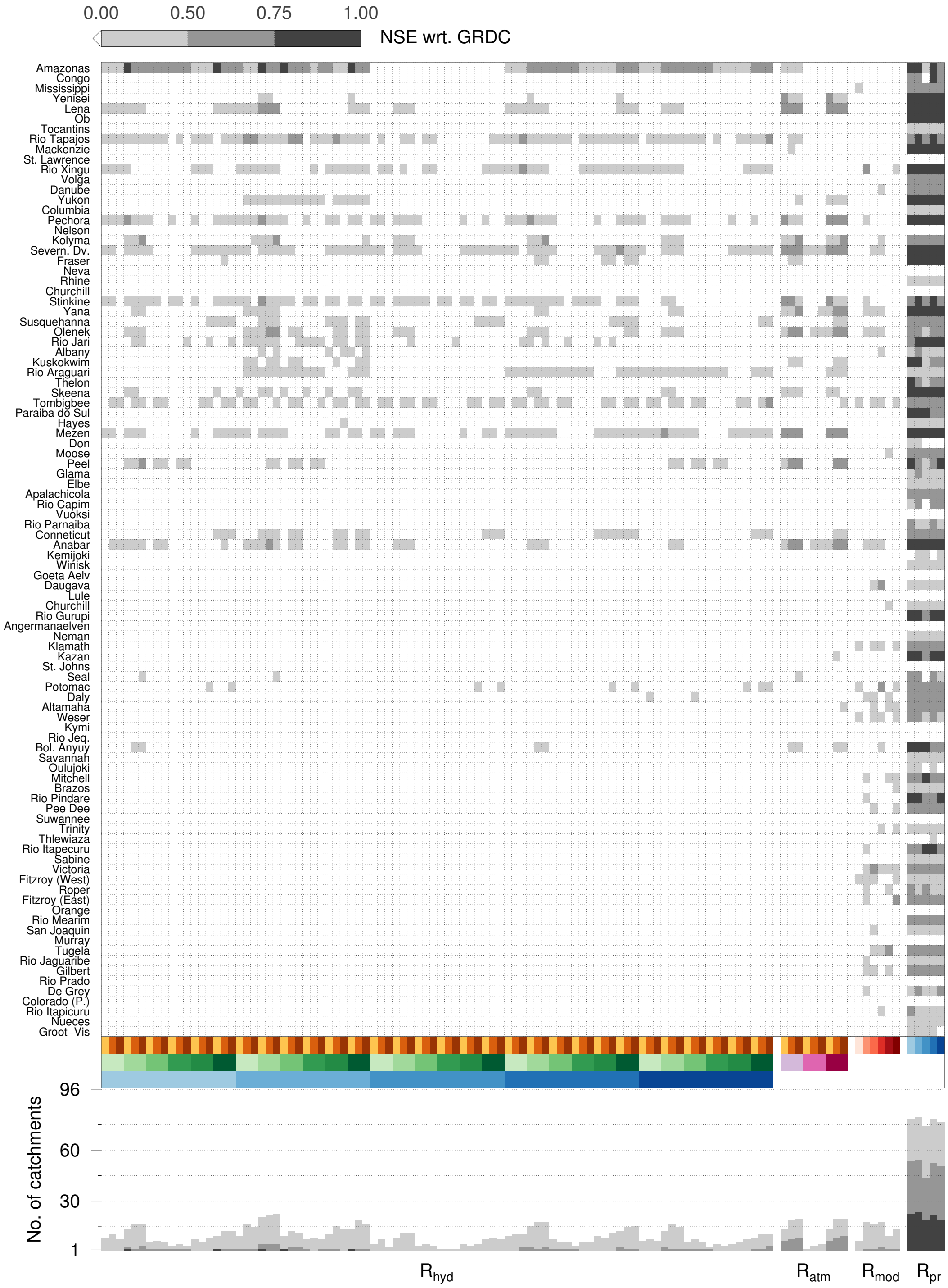

FIGURE 3.12: Nash-Sutcliffe Efficiency coefficient for the 96 study regions between observed and estimated runoff.

reasonable dynamics in the runoff estimates from these combinations (see stacked bars in Figure 3.11). Even if the better results over Arctic basins are not visible in the correlation matrix, there are still some catchments where R_{atm} shows higher correlations than R_{hyd} (e.g., Volga, Columbia, and Conneticut).

Finally, the matrix showing the NSE values (Figure 3.12) reveals that the combinations using moisture fluxes from ERA-Interim and MERRA perform slightly better than the R_{hyd}-estimates. Positive NSE values can be observed over at lest 15 catchments, while there are only 8 catchments with values ≥ 0.5 (see stacked bars in Figure 3.12). The matrix shows that R_{atm} provides better results mainly over Arctic catchments (e.g., Yenisei and Yana). Here, all combinations (except for these using CFSR moisture fluxes) reach values greater than 0.5, which indicates good agreement with the runoff observations.

The histograms (bottom row in Figure 3.13) confirm the similar performance of moisture fluxes from ERA-Interim and MERRA and the significant biases in the CFSR dataset. Over roughly 50 % of all the R_{atm} combinations and catchments involving moisture fluxes from ERA-Interim, correlations ≥ 0.5 with respect to observed runoff can be observed. In terms of PBIAS, MERRA has slightly more data points with biases $\leq 50\,\%$. Finally, the NSE shows that there is no superior data source of moisture flux divergences as both ERA-Interim and MERRA reach values between 0 and 0.5 (0.5 and 0.75) over about 10 % (8 %) of the data points.

Runoff from Land Surface Models: R_{mod}

In terms of PBIAS (Figure 3.8), the R_{mod} estimates from $\text{GLDAS}_{\text{CLM}}$ and $\text{GLDAS}_{\text{Noah3.3}}$ show lower biases than the other land surface models. However, compared with R_{hyd} and R_{atm}, the models seem to suffer from some large biases over many catchments. For large basins like Amazonas this can be partly explained with the lack of a routing scheme which denies the correct estimation of peak-runoff and thus creates large differences with respect to the observations. However, even over these large basins, there are single models which show PBIAS values $\leq 10\,\%$ (e.g., $\text{GLDAS}_{\text{Noah3.3}}$ over

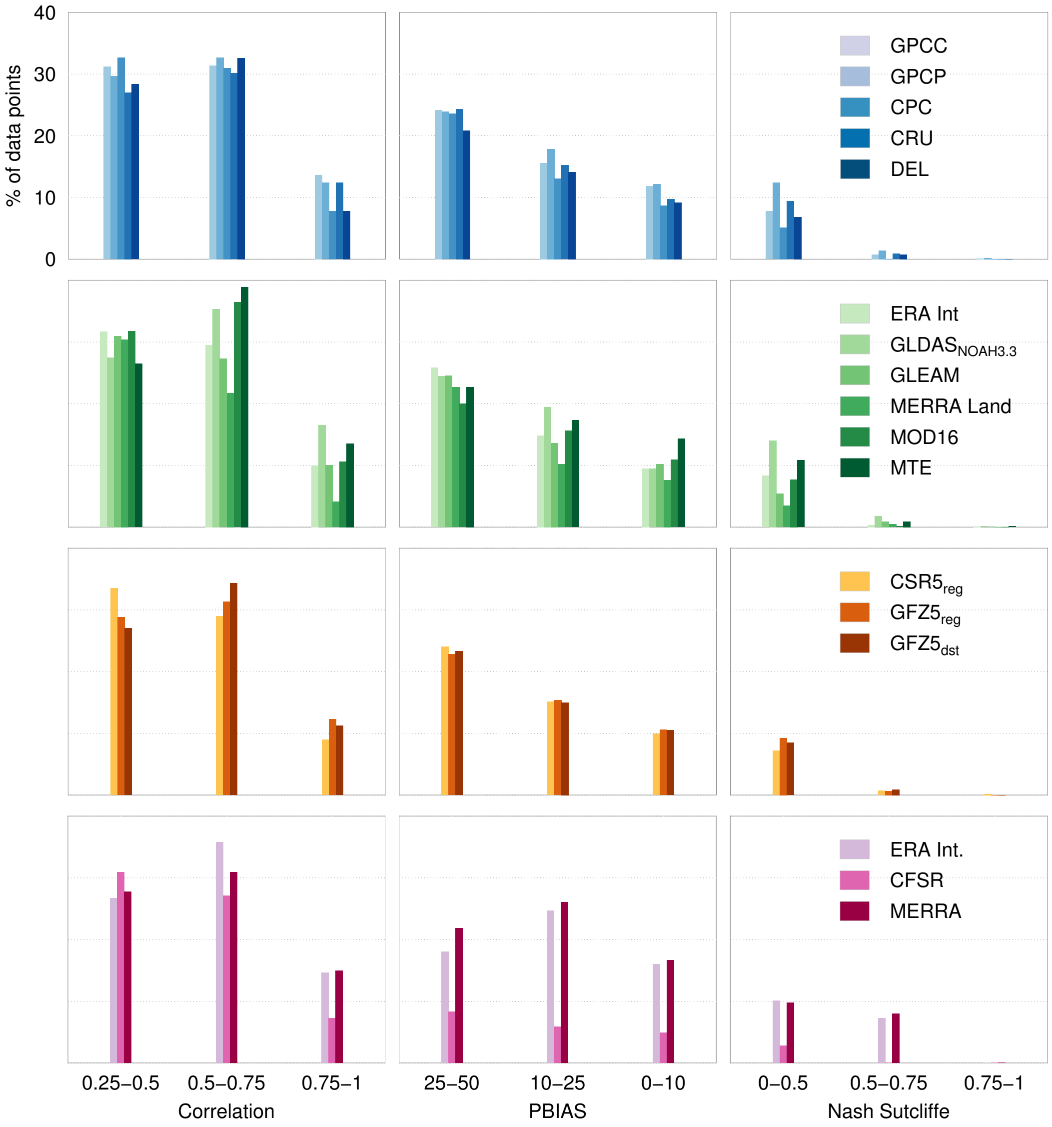

FIGURE 3.13: Histograms showing the percentage of data points (i.e., all combinations and all catchments using a specific dataset) within the intervals from Figures 3.8, 3.11, and 3.12. They thus serve as a summary of the matrix plots and show the individual performance of every dataset. The plot shows separate histograms for precipitation (first row, 1728 data points), evapotranspiration (second row, 1440 data points), water storage changes (third row, 2880 data points), and moisture flux divergences (bottom row, 288 data points). The number of data points is the product of the number of catchments and the number of combinations using a single dataset (e.g., 96 catchments times 18 combinations using GPCC precipitation = 1728 data points)

.

Amazonas and Volga or GLDAS$_{CLM}$ over Congo). This indicates that the lack of a routing scheme cannot fully explain the large deviations from the observations. Interestingly, even if GLDAS$_{CLM}$, GLDAS$_{Mosaic}$, GLDAS$_{Noah2.7}$, and GLDAS$_{VIC}$ are driven by the same forcing data, the PBIAS shows some significant scattering as there is not a single catchment where all these models perform consistently good. Overall, $\text{GLDAS}_{\text{Mosaic}}$ and $\text{GLDAS}_{\text{VIC}}$ show the largest biases with values smaller than 50 % over less than 40 catchments (see stacked bars in Figure 3.8).

In contrast to the PBIAS analysis, the correlations between R_{mod} and the observations draw a much more homogeneous picture (Figure 3.11). There are many catchments with correlation values greater than 0.5 from all models. This holds true for large catchments (e.g., Mississippi, Volga, and Danube) but also for very small basins with low discharge (e.g., Rio Mearim, Gilbert, or Rio Prado). Thus, there is no dependency on a catchment's area or mean discharge, if land surface models are able to provide reasonable runoff dynamics. This performance is even more impressive when compared with the results from R_{hyd} and R_{atm} as most land surface models show correlations of at least 0.5 (0.75) over 60 (30) and more catchments. Moreover, R_{mod} perform quite well over regions where R_{hyd} and R_{atm} do not capture the runoff dynamics. These are either catchments where runoff has a much smaller amplitude compared to the other water cycle variables (e.g., Mississippi, Volga, Rhine, Don, and Neman) and/or no clear distinct annual cycle in precipitation (e.g., Rhine, Don, Daugava, Neman, Potomac, Altamaha, Weser, or Savannah). In these cases, the biases of the data sources of precipitation and evapotranspiration create short-term variations that result in low correlation values.

The NSE matrix (Figure 3.12) shows a similar performance of R_{mod} and R_{hyd}. However, the two methods often succeed in totally different catchments. Especially over smaller catchments with low discharge, R_{mod} seems to provide reasonable runoff estimates (e.g., Potomac, the western Fitzroy, Victoria, or Tugela). Here, we see a very good agreement in the time series between the observed and estimated runoff. Furthermore, there are several catchments where most of the R_{mod} estimates provide NSE values greater than 0 (e.g., Tombigbee, Klamath, or Victoria) which should give evidence in the reliability of the modeled runoff in these regions.

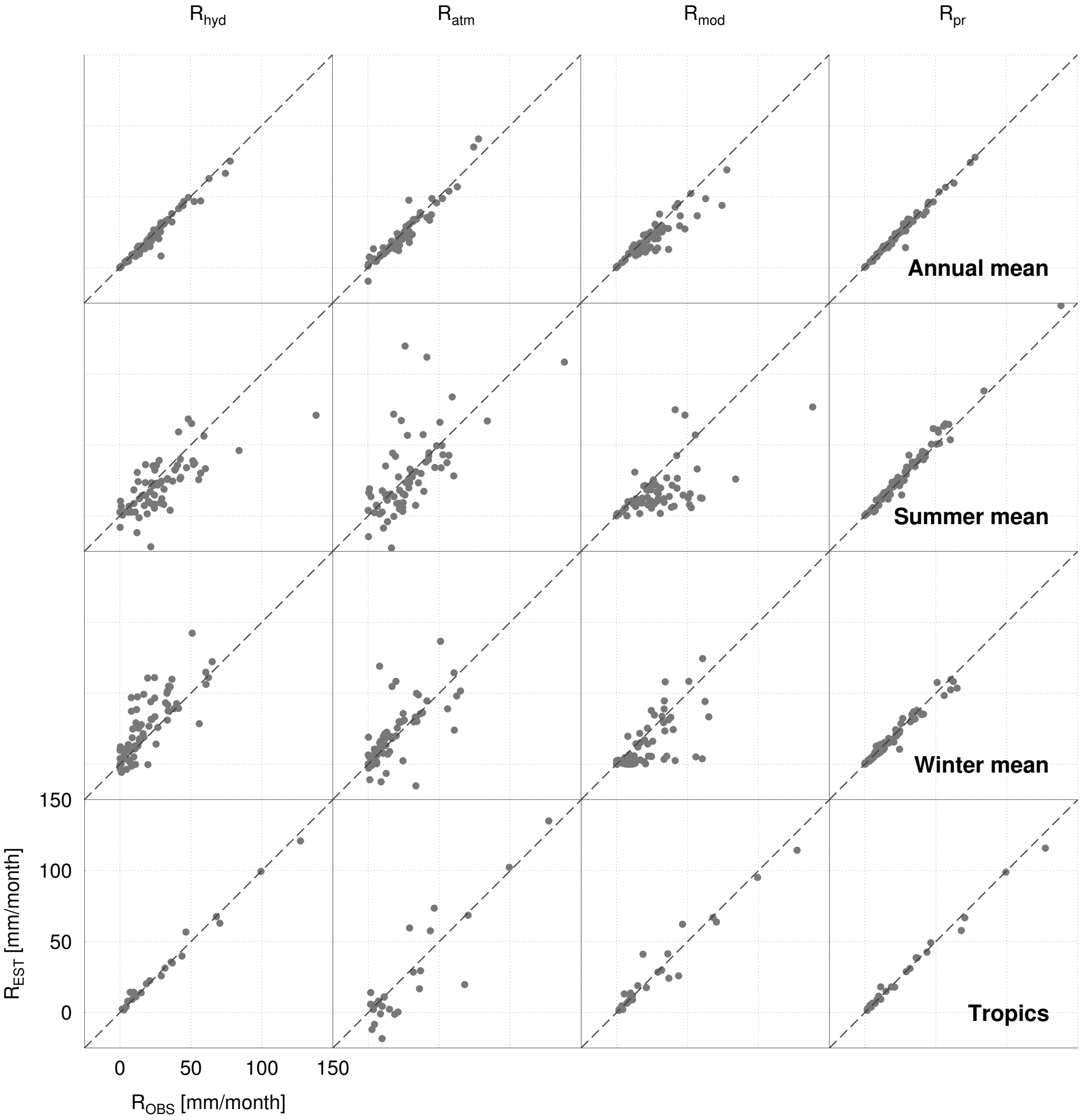

FIGURE 3.14: Scatter plots of the mean annual runoff (top row), summer and winter averages (second and third row), and mean annual runoff over the tropical catchments (fourth row). Each dot represents a single catchment. The x-values are given by the GRDC-observations while the y-values are derived from the four approaches best possible combinations for each catchment individually in terms of the PBIAS. The summer means are computed from the JJA-averages over the northern and the DJF-averages over the southern hemispheric catchments.

Runoff from Precipitation-Runoff Ratio: R_{pr}

The rudimentary approach based on precipitation-runoff ratio is by far the most bias-free approach with more than 45 (75) catchments that provide a PBIAS value of less than 10 % (25 %; see stacked bars in Figure 3.8). Surprisingly, these values hold well irrespective of the precipitation dataset being used. Nevertheless, this method also fails in some catchments (St. Lawrence, Nelson, Churchill, Vuoksi, Rio Jequitinhonha, Suwanee, Orange, Murray, De Gray, and Rio Prado), wherein there is not a clear relationship between precipitation and runoff. The correlation values (Figure 3.11) also demonstrate that the R_{pr} method captures the dynamics accurately, leading to a very good performance. Here again the choice of the precipitation dataset does not matter at all for performance of the method as nearly 60 catchments provide a correlation of more than 0.75. The catchments that returned poor PBIAS values also perform poorly in terms of correlation. The NSE values (Figure 3.12) paint the same picture as PBIAS and correlation values. However, the CPC dataset performs relatively worse than the other precipitation datasets. In spite of this, the R_{pr} method performs the best of all the methods analyzed here.

This method is successful in catchments where there is a strong annual cycle (cyclostationarity) in both the precipitation and runoff datasets, which directly translates into a strong relationship in the $P-R$ ratio. This method fails when the runoff has a nonstationary behavior (e.g., see the time series of Nelson, St. Lawrence, or Vuoksi in the supplemental material), thus resulting in a lack of relationship in said ratio. Another interesting attribute of this method is that the ratio absorbs the biases in the precipitation datasets: the higher the biases, the smaller the ratio and vice versa. Although the $P-R$ ratio deals only with the stable relationship between runoff and precipitation, it implicitly demands the stability of the other two components of the water cycle: evapotranspiration and water storage change. For example, it is precisely because of the instability in water storage change that the method fails in catchments like St. Lawrence and Nelson, where the precipitation and evapotranspiration have a very strong annual cycle. Therefore, it can be said that the influence of evapotranspiration and water storage change is implicit in the $P-R$ ratio.

A theoretical advantage of R_{hyd} and R_{atm} over models (without routing) and R_{pr} is that anthropogenic influences should not play a role when considering the level of agreement with respect to GRDC. As the predictions of R_{pr} are based on the *mean monthly* runoff-precipitation-ratio, anthropogenic effects have to be treated separately, or, more precisely, those effects cannot be taken into account when no runoff observations are available. However, we could not detect a better performance of the budget approaches compared to R_{pr} and R_{mod} over such catchments.

Water Budget Closure

As R_{hyd} and R_{atm} are based on the terrestrial and atmospheric-terrestrial water balance, they enable us to study the water budget (mis)closure at different time scales. The PBIAS value, which contains the mean of the differences between the observations and the estimates, provides the relative long-term imbalance of the water budget from these two methods. This can be expressed as follows:

$$\begin{aligned}\text{PBIAS}_{\text{hyd}} &= \frac{\sum\limits_{t=1}^{T}\left(R_{t,\text{obs}} - R_{t,\text{hyd}}\right)}{\sum\limits_{t=1}^{T} R_{t,\text{obs}}} \\ &= \frac{\sum\limits_{t=1}^{T} \delta_{t,\text{hyd}}}{\sum\limits_{t=1}^{T} R_{t,\text{obs}}} \\ &= \frac{\overline{\delta}_{\text{hyd}}}{\overline{R}_{\text{obs}}},\end{aligned} \tag{3.15}$$

similarly,

$$\text{PBIAS}_{\text{atm}} = \frac{\overline{\delta}_{\text{atm}}}{\overline{R}_{\text{obs}}}. \tag{3.16}$$

At the annual scale, for almost all of the catchments, there is at least one data combination from either method that enables us to close the water budget (cf. Figure 3.10 and annual mean and tropics of R_{hyd} and R_{atm} in Figure 3.14), but as mentioned in section *Runoff from Terrestrial Water Budget:* R_{hyd}, the data combination differs from catchment to catchment. It is also clear that the R_{hyd} method, despite dealing with biases from three water balance equation variables, is able to close water budget better than the R_{atm} method (except for the Arctic catchments). This is conspicuous, especially in the tropics. The PBIAS values also show huge imbalances for the low discharge catchments (bottom part of Figure 3.8), but care must taken to

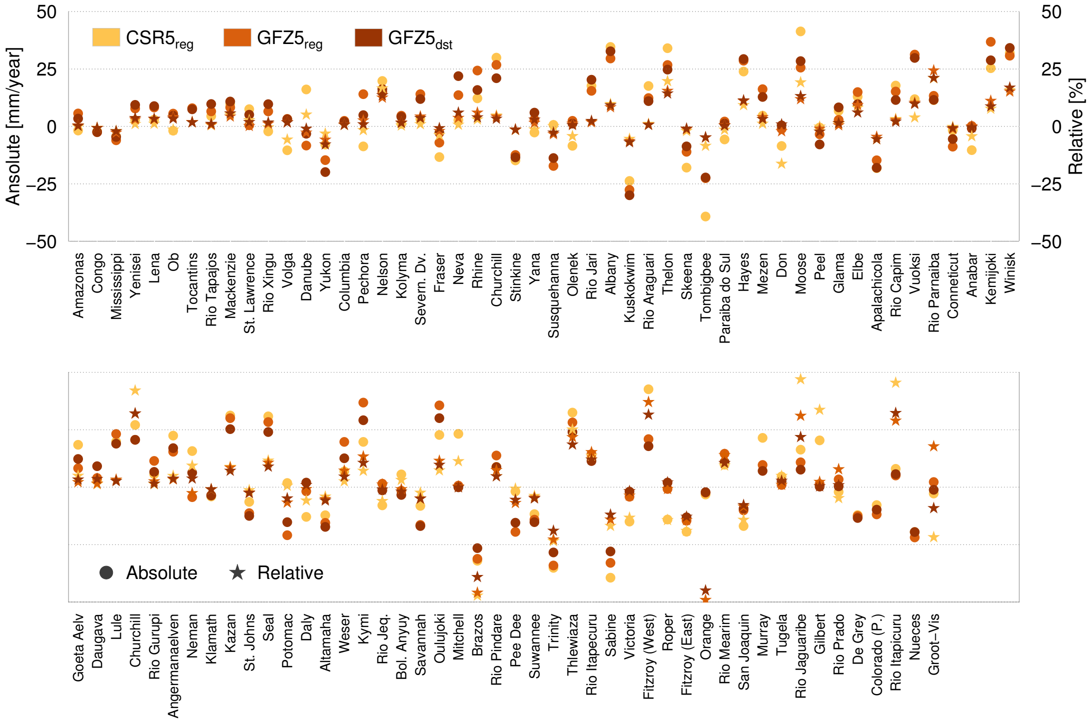

FIGURE 3.15: Contribution of water storage changes from the three GRACE datasets to the long-term water budget over the 96 study regions. The circles show the absolute contribution in units of [mm/year] while the stars depict the relative contribution with respect to the mean annual runoff.

interpret these values. In these catchments, the imbalances are effectively divided by values close to zero, and therefore, those PBIAS values end up being orders of magnitude higher than for the catchments with significant runoff (catchments on the right-hand side in the bottom rows of Figures 3.9 and 3.10). To this end we also look at the absolute water budget imbalance.

Each of the data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences might be biased over certain regions and must thus be considered as a potential source of errors. In some cases, such biases can get canceled out when computing R_{hyd} or R_{atm}, leading to random errors which scatter around zero. This is the case for catchments like Rio Tapajos, St. Lawrence, Rio Jari, Tombigbee, or even small basins like San Joaquin (Figure 3.10). In these cases, the scattering gives information on the uncertainty level of the water cycle variables.

Over most catchments, however, either a positive or negative bias for most combinations can be observed. This indicates general shortcomings in the current quality of the different water cycle variables. Most obviously, this can be observed for the Arctic (tropical) catchments, where the majority of R_{hyd} combinations consistently tend to underestimate (overestimate) runoff (Figure 3.10).

The main reason for these heavily biased runoff estimates can be assumed to originate from uncertainties in the precipitation and evapotranspiration datasets. According to Figure 3.15, there is no clear contribution of water storage changes from GRACE to the catchments showing large remaining imbalances. For example, for Tocantins, where R_{hyd} is significantly overestimated, the GRACE datasets show an increase of water storage of less than 10 mm/year. The same holds true for several Arctic catchments (e.g., Yenisei or Lena). In general, the total annual water storage as depicted by GRACE is less than ±50 mm/year for all catchments, but for most of the catchments, the total annual water storage values are within ±25 mm/year. Therefore, for majority of the analyzed catchments, as GRACE does not contribute a lot to the imbalances, the biases in the precipitation and evapotranspiration datasets contribute heavily to the over- and underestimation.

At the monthly time scales, use can be made of the NSE to study the water budget closure because it contains the mean squared error (mse) of estimates with respect to the observations in the numerator and the variance of the observations in the denominator. Thus, it is the normalized mean squared error of the estimates:

$$\begin{aligned} \text{NSE} &= 1 - \frac{\sum_{t=1}^{T}\left(R_{t,\text{obs}} - R_{t,\text{est}}\right)^2}{\sum_{t=1}^{T}\left(R_{t,\text{obs}} - \overline{R}_{\text{obs}}\right)^2}, && (3.14) \\ \text{MSE} &= \frac{1}{T}\sum_{t=1}^{T}\left(R_{t,\text{obs}} - R_{t,\text{est}}\right)^2, && (3.17) \\ \sigma_{\text{obs}}^2 &= \frac{1}{T}\sum_{t=1}^{T}\left(R_{t,\text{obs}} - \overline{R}_{\text{obs}}\right)^2, && (3.18) \\ \therefore \quad \text{NSE} &= 1 - \frac{\text{MSE}}{\sigma_{\text{obs}}^2}, && (3.19) \end{aligned}$$

where σ_{obs}^2 is the variance of the observed runoff. Like PBIAS, when the catchment runoff is close to zero, the NSE values become bloated. Nevertheless, the closure at monthly time scales is far from being acceptable. Only around 15 catchments provide a positive value (Figure 3.12), which clearly indicates that the datasets are not mature enough to close the water balance at monthly time scales.

Summary and Conclusions

In this contribution, we have estimated runoff from water balance equations mainly using GRACE water storage changes. A wide variety of hydrological and hydrometeorological datasets have been used to arrive at the runoff estimates. The estimates were then compared to observed runoff. This was only possible in about 96 catchments because of the limited time span of the GRACE dataset and the limited data availability and large data gaps in the GRDC database. To quantify the estimates holistically, we included runoff estimates from sophisticated hydrological land surface models and an empirical model involving the ratio between precipitation and runoff in the analysis. The evaluation of the estimates was carried out using three performance metrics, namely, PBIAS, correlation coefficient, and NSE.

Based on the assessment, the following conclusions are drawn.

1. Overall, the estimates from the precipitation–runoff ratio, though being a simple method, provide by far the best results when compared to the budget-based methods and the land surface models. This also implies that for most catchments, the impact of evapotranspiration and water storage changes on runoff underlies a stable annual cycle, which remains constant from year to year. However, the method can only succeed when the monthly relationship between runoff and precipitation remains constant over the years. For catchments with more or less random time series of precipitation or runoff, no stable precipitation–runoff ratio can be derived. For such catchments, the member of R_{pr} showed consistently bad results.
2. In the water-balance-based estimates, there is no combination of datasets that is consistently standing out as the best, and there is also no combination

that behaved consistently over different catchments. In particular, the PBIAS analysis shows highly variable results from the R_{hyd} and R_{atm} estimates. The results further indicate that R_{hyd} tends to underestimate (overestimate) runoff over Arctic (tropical) catchments. For Arctic catchments, however, R_{atm} shows better PBIAS values. As a consequence, the quality of different data sources of precipitation, evapotranspiration, water storage changes, and moisture fluxes differ significantly from catchment to catchment. The analysis further shows that the PBIAS is very sensitive to the choice of a member of the R_{hyd}-ensemble.

3. This indicates that one cannot extrapolate the performance of a combination or method to other basins without thorough validation. To a certain extent, this holds also true for the estimates from the precipitation-runoff ratio. Performance of combinations thus requires scrutiny over individual catchments. This also has strong implications for the regionalization of these estimates to adjacent catchments.

4. The estimates from R_{hyd} and R_{atm} perform well in catchments where the runoff exhibits a strong annual cycle combined with a large volume of runoff. Over such catchments, most combinations reach correlations of at least 0.75. Basins where the magnitudes of the variations in the time series of runoff are far below the other water cycle variables suffer from biases in the different data sources of precipitation, evapotranspiration, water storage changes, and moisture flux divergences. These biases create short-time variations, which deny a reasonable estimation of runoff.

5. Further, the water-balance-based methods are reliant on GRACE datasets, which are reported to be of inferior spatial resolution compared to the other datasets used here. In spite of this shortcoming, the R_{hyd} estimates still provide good performance in small catchments, but only those with a strong annual cycle. Moreover, the different GRACE datasets show significantly different results in catchments with a time shift between peak runoff, precipitation, and evapotranspiration. In these cases, GRACE has to take care of the water storage and thus has a major influence on the performance of the R_{hyd} estimates.

6. For most catchments, there is still a significant relative water budget imbalance with respect to the mean runoff. The majority of combinations from R_{hyd} suffers from a remaining imbalance of more than 25 % over many basins. However, over most catchments, there is at least a single combination which leads to a relative imbalance close to 0 % indicating a reasonable water budget closure. Moreover, over several catchments, the absolute imbalance scatters around zero. Here, the water budget residual can be assumed to be randomly distributed between the different combinations of R_{hyd}. This emphasizes the use of ensemble-based methods for analyzing the water cycle over such catchments.

7. The models performed comparatively poorly when compared with the other approaches. A reason for this might be due to the lack of a routing scheme. However, even on annual time scales, significant biases could be observed when compared with the observations. It could be also detected that over several catchments, modeled runoff might show a significant time lag. However, this is again not consistent throughout all the study regions, which makes it inevitable to evaluate such data sources both in the individual catchments and also globally.

In general, none of the analyzed methods is able to provide reasonable runoff estimates globally. There are, however, several regions where the independent estimates from R_{hyd}, R_{atm}, and R_{mod} are able to successfully reproduce observed runoff. But as the data combinations and models that perform well differ from catchment to catchment, we cannot rely on a single dataset for analyzing the basin-scale water cycle. Therefore, there is still an urgent need for *in situ* observations of the major water cycle variables, which are mandatory for further necessary improvements of global hydrological datasets.

Acknowledgments

We gratefully acknowledge the support of projects SN13/1, BA1150/11, KU2090/1 by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program SPP1257 *Mass Transport and Mass Distribution in the System Earth*. We would like to thank the following data providers: Global Runoff Data Centre (GRDC) `www.bafg.de/GRDC`, Global Precipitation Climatology Project (GPCP) `precip.gsfc.nasa.gov`, European Centre for Medium-Range Weather Forecasts (ECMWF) `www.ecmwf.int`, University of Texas at Austin, Center for Space Research (CSR) `www.csr.utexas.edu/`, NASA's Earth Science Division and Goddard Earth Sciences (GES) Data and Information Services Center (DISC) `ldas.gsfc.nasa.gov`. Further thanks are due to Dr. Diego Miralles (University of Bristol) for kindly providing the GLEAM evapotranspiration data. All graphics in the document were produced with the Generic Mapping Tools (GMT) `gmt.soest.hawaii.edu` [Wessel and Smith, 1991].

Appendix

Treating the Errors in GRACE Mass Anomalies

The major sources of errors in the GRACE data are aliasing of residual tidal signal [Seo et al., 2008], poor estimates of C_{20} coefficient [Chen et al., 2005], and high-frequency noise in the spherical harmonic coefficients due to orbit geometry [Kusche et al., 2009]. These errors are treated at the spherical harmonic coefficient level.

Tidal Residuals

The tidal residuals alias frequencies that depend on the orbit geometry of GRACE, sampling rate of GRACE and the natural frequency of the corresponding tides, which enables them to be modeled and removed from the signal to a large extent. This is accomplished by performing a least squares spectrum analysis on the residuals of GRACE temporal geoid anomalies, which are obtained after removing the mean annual behavior The mean annual behavior is computed by taking the average of the twelve calendar months as is done for the precipitation-runoff ratio (3.8) [see Tourian et al., 2011, for the efficacy of this method]. The spectral analysis is performed only for select tidal frequencies: M_1, S_1, S_2, P_1, Q_1, O_1 and N_2. The estimates of the aliased frequencies are subsequently subtracted from the residuals of the spherical harmonic coefficients, which are then added back to the mean annual behavior to obtain the corrected monthly anomalies.

Erroneous C_{20} Coefficients

The erroneous C_{20} coefficients are replaced with those estimated from Satellite Laser Ranging (SLR) in the case of the CSR monthly solutions as suggested by Cheng and Ries [2013]. However, the C_{20} coefficients in the GFZ monthly solutions are used as it is as they are well-behaved. This is due to the use of EIGEN-6C gravity field model as a background model in the processing scheme [Dahle et al., 2013], which includes data from GRACE, GOCE, and LAGEOS (SLR) satellites and also data from airborne and terrestrial gravimetry.

High-Frequency Noise

The high-frequency noise in the spherical harmonic coefficients manifest themselves as stripe-like patterns in the spatial domain. This is dealt with by filtering, and here, we apply two methods of filtering: decorrelation filter in combination with a Gaussian filter with a filter radius of 350 km [Swenson and Wahr, 2006], and regularization filter as proposed by Kusche [2007], but with certain modifications. The modifications are that while Kusche [2007] computes the signal covariance from a combination of different models, we compute the signal covariance from the GRACE data itself as described by Sasgen et al. [2006]; and Kusche [2007] uses stationary signal and noise covariances, whereas we use cyclo-stationary signal covariances and non-stationary noise covariances. For the GFZ dataset, standard deviations of the spherical harmonic coefficients are provided, which were used for the simulation of noise covariances. The standard deviations for the CSR dataset are not provided, and therefore, we have used standard deviations from release 4 dataset for the noise covariance simulation.

Chapter 4

Article III: Estimating Runoff Using Hydro-Geodetic Approaches

Sneeuw, N., C. Lorenz, B. Devaraju, M. J. Tourian, J. Riegger, H. Kunstmann, and A. Bárdossy, 2014: Estimating Runoff Using Hydro-Geodetic Approaches, *Surveys in Geophysics*, **35**, 1333–1359, doi: 10.1007/s10712-014-9300-4[1]

Abstract

Given the continuous decline in global runoff data availability over the past decades, alternative approaches for runoff determination are gaining importance. When aiming for global scale runoff at a sufficient temporal resolution and with homogeneous accuracy, the choice to use spaceborne sensors is only a logical step. In this respect, we take water storage changes from the Gravity Recovery and Climate Experiment (GRACE) and water level measurements from satellite altimetry, and present a comprehensive assessment of five different approaches for river runoff estimation: hydrological balance equation, hydrometeorological balance equation, satellite altimetry with quantile function based stage-discharge relationships, a rudimentary instantaneous runoff-precipitation relationship, and a runoff-storage relationship that takes time-lag into account. As a common property these approaches do not

---

rely on hydrological modeling; they are either purely data driven or make additional use of atmospheric reanalyses. Further, these methods, except runoff-precipitation ratio, use geodetic observables as one of their inputs, and therefore, they are termed hydro-geodetic approaches. The runoff prediction skill of these approaches are validated against *in situ* runoff and compared to hydrological model predictions. Our results show that catchment-specific methods (altimetry and runoff-storage relationship) clearly outperform the global methods (hydrological and hydrometeorological approaches) in the six study regions we considered. The global methods have the potential to provide runoff over all landmasses, which implies gauged and ungauged basins alike, but are still limited due to inconsistencies in the global hydrological and hydrometeorological datasets that they use.

Introduction

Against the backdrop of global change, both in terms of climate and demography, there is an increasing need for monitoring global water cycle. Recognizing this need a number of initiatives have been taken in the recent decades to monitor the global water cycle and other climatic phenomena, for example, Global Energy and Water Cycle Experiment [GEWEX; Chahine, 1992a], Climate Variability and Predictability [CLIVAR; CLIVAR Scientific Steering Group and World Climate Research Programme, 1995], Global Climate Observing System [GCOS; Fellous, 2014], FLUXNET [Baldocchi et al., 2001] and the International Soil Moisture Network [ISMN; Dorigo et al., 2011]. These initiatives are being ably supported by the satellite-based Earth observation programmes of National Aeronautic and Space Agency [NASA; Asrar et al., 2001], European Space Agency [ESA; Southwood, 1999] and other space agencies [Gleick et al., 2013]. All in all, this has resulted in great advancements in measurement technology both *in situ* and *ex situ*, and datasets of different quantities involved in the global hydrological cycle [Gleick et al., 2013].

Runoff is one of the most accurate (10–20 % of the observation) of the global hydrological cycle quantities [Fekete et al., 2012]. The publicly available global runoff

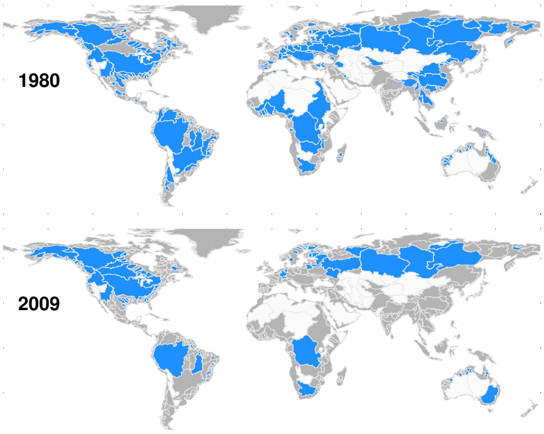

FIGURE 4.1: The maps show the catchments whose runoff measurements are publicly available (blue), that do not discharge into the oceans or other open water bodies (white) and those that are not gauged or whose runoff measurements are not publicly available (gray).

database (e.g., Global Runoff Data Center, GRDC) is very limited in its spatial and temporal coverage of catchments worldwide. Moreover, the acquisition of discharge data and their delivery to the database are on the decline since the late 1970s [cf. Tourian et al., 2013, Figure 1 or Lawford et al., 2013, Figure 3], be it for economical, political or other reasons. The number of available stations went down from about 8, 000 (pre-1970) to roughly 2, 000 (around the year 2010). The total monitored annual streamflow has dropped accordingly by about 75%. Figure 4.1 visualizes a clear decline between 1980 and 2009 in the number of catchments whose measured runoff is both available and accessible. Prediction of runoff by modeling has so far been successful mostly in the gauged basins rather than in the ungauged basins [Hrachowitz et al., 2013], thus unable to alleviate the problem. To this end, Sheffield et al. [2009] and Sahoo et al. [2011] demonstrate the potential of satellite-based data for studying the water cycle at basin scales.

Given the insufficient monitoring from *in situ* gauge networks, and without any outlook of improvement, spaceborne approaches are currently being investigated. Satellite-based earth observation with its global coverage and homogeneous accuracy has been demonstrated to be a potential alternative to *in situ* measurements, but it does not yet provide the desired accuracy and require *in situ* data for calibration and validation [Trenberth, K. E. and Asrar, 2014]. Nevertheless, they are the only source of information in data poor regions and the open oceans [Hrachowitz et al., 2013]. To this end, Sheffield et al. [2009] and Sahoo et al. [2011] demonstrate the potential of satellite-based data for studying the water cycle at basin scales.

Alsdorf et al. [2007] reviewed the potential of several spaceborne observation techniques in contributing to the tracking of surface freshwater from space. A diversity of satellite sensors was classified in terms of their ability to monitor water surface elevation (h), slope ($\partial h/\partial x$), its time variation ($\partial h/\partial t$), water surface area (A) or mass of the total water storage (S). Although they acknowledge an enormous improvement in the spaceborne monitoring of these hydraulic variables over the recent past, they do point out obvious limitations, amongst others: dependence on calibration by *in situ* measurements, poor resolution (spatial and temporal) and accuracy, and error sources proprietary to the technique involved.

The observables that have been identified above are all geometric except water storage. Apart from water surface area, all the other geometric variables are provided by spaceborne geodetic sensors (satellite altimetry and synthetic aperture radar), and the physical variable of water storage is observed by the GRACE satellite mission, which is again a spaceborne geodetic sensor. Several studies make use of geodetic spaceborne sensors for calibrating and validating hydrological models. Wilson et al. [2007] used data from satellite altimetry for validating floodplain predictions from a hydrodynamic model over the central Amazon floodplain. Getirana [2010] was the first to apply spatial altimetry data for automatic calibration of a hydrological model over the Branco River basin, which is located in the Northern Amazon basin. In Paiva et al. [2013], results from the MGB-IPG hydrodynamic model were compared against water levels derived from ENVISAT altimetry data and terrestrial water storage anomalies from GRACE.

The water storage changes estimated from GRACE are dominated by noise requiring filtering [e.g., Kusche, 2007]. An assessment on the performance of different filter tools for GRACE has been presented in Werth et al. [2009]. The authors compared a multi-model mean of the three global hydrological models WGHM, GLDAS, and LaD with GRACE-based water mass variations over 22 of the world's largest river basins. Ferguson et al. [2010] discussed the estimation of evapotranspiration from remote sensing data using the Penman-Monteith model and validated their results with total water storage variations derived from different GRACE products.

In this contribution we compare the skill of four different satellite-based remote sensing methods in determining runoff from gauged and ungauged basins alike. Three different methods make use of GRACE data: one based on the hydrological water balance, and one based on the atmospheric water balance. A third GRACE-related method makes use of relationship between runoff and storage. The fourth hydro-geodetic method makes use of satellite radar altimetry. These methods have in common that they use geodetic space sensors, which is the reason for coining them *hydro-geodetic* approaches. For further comparison we incorporate another elementary approach, namely a basic relationship between precipitation and runoff, and the output from a hydrological model. The skill score is evaluated over gauged catchments with available runoff data. Since all the other quantities are in terms of change in water column (mm/month), we have considered the river discharge of gauging stations at the outlet of catchments and divided them with the area of their catchments. This means that we virtually spread the discharged water over the surface area of the catchment. This quantity is often called runoff or specific discharge, and this will be the definition that will be used in the rest of the document.

The use of satellite radar altimetry for monitoring rivers was first demonstrated by Koblinsky et al. [1993], after which many studies brought forth improvements [e.g., Berry, 2005, Birkett, 1998, Birkinshaw et al., 2010, Bjerklie et al., 2003, Frappart et al., 2006, Papa et al., 2010, Tourian et al., 2013], which makes it a mature technique. Satellite altimetry provides river water level, which needs to be converted to runoff, which we will term as *altimetric river runoff*. The altimetric river runoff is of great benefit for hydrological community as the reported accuracies of obtained altimetric runoff over different rivers vary between 5 and 15 % [Birkinshaw et al., 2010, Tourian et al., 2013, Zakharova et al., 2006]. However, its temporal and spatial resolution limit the use of satellite altimetry for river runoff estimation. Satellite altimetry provides snapshot measurements every 10 or 35 days (depending on the mission), which does not correspond to the daily or monthly sampling used for river gauges. Its spatial resolution would also limit the runoff estimation as its footprint size is ca.

8.5 km [Fu and Cazenave, 2001], which covers river surface water together with the neighboring topography and indeed the vegetation canopy. Moreover, availability of altimetric water level measurements over rivers also limit the use of altimetry for runoff estimation as it is dictated by satellite's orbital configuration. The inter-track distance of ENVISAT at the equator is approximately 80 km, which is better than the spacing of 311 km of TOPEX/Poseidon and Jason 1 & 2. These limitations are expected to be overcome with the launch of the Surface Water and Ocean Topography (SWOT) satellite altimetry mission in 2020 [Durand et al., 2010].

The water-budget based methods to estimate runoff received an impetus with the availability of total water storage changes from the GRACE satellite mission. Syed et al. [2005] showcased the potential of such methods by estimating discharge from the Amazon and Mississippi basins. They also assumed that differences between estimated and observed runoff might be due to unrecorded groundwater and surface water inflows. Thereafter Syed et al. [2007] estimated the freshwater discharge from the Pan-Arctic region into the Arctic sea. Syed et al. [2009] culminated these studies by estimating the discharge from the continents to the oceans. Concurrently, Sheffield et al. [2009] estimated streamflow over the Mississippi basin, but concluded that the errors from the budget-based methods are rather prohibitive, and hence not viable. Lorenz et al. [2014] also come to the same conclusion after estimating runoff via the budget-based methods for over 96 gauged catchments. Riegger and Tourian [2014] demonstrate another method to utilize the water storage changes to estimate runoff that the hydraulic relationship between runoff and water storage. However, the limiting factor of GRACE is its accuracy, which hovers between 10 and 30 mm/month [Fersch et al., 2012, Long et al., 2014, Riegger et al., 2012]. Although Longuevergne et al. [2010] put a limit on the spatial resolution of GRACE at $\approx 200\,000\,\mathrm{km}^2$, Lorenz et al. [2014] showed that GRACE can be useful for basins with an area far below $200\,000\,\mathrm{km}^2$, but only when their water storage changes show a strong seasonality.

Depending on the choice of hydrological, hydrometeorological and geodetic data products that serve as input to each of the methods, in principle, a host of permutations need to be assessed. This has been performed by Lorenz et al. [2014], but in contrast, a preselection of data products and a choice of six relevant catchments allows us, here, to focus on the methods instead. In line with this aim all methods are explained in section *Hydro-Geodetic Approaches*. The preselected products (data and models) that serve as input to the methods are discussed in section *Data and models*. The results of validating all six approaches against observed runoff are provided in section *Results and Discussion* and summarized in section *Summary and Conclusions*.

Hydro-Geodetic Approaches

The primary characteristic of the hydro-geodetic approaches is that they exploit the geodetic observables that are directly or indirectly affected by the hydrological phenomena. In this study, we will be concerned with two such geodetic observables: temporal changes in the gravity field that are caused by water storage changes over the continents as observed by GRACE, and water-level changes in rivers as observed by the ENVISAT satellite mission.

Runoff from Altimetry

The river runoff at the selected gauges is typically determined from an empirical functional relation between water level estimated by satellite altimetry H at a virtual station close to a runoff gauging station and measured runoff R at the gauge [Birkinshaw et al., 2010, Coe and Birkett, 2004, Getirana and Peters-Lidard, 2013, Kouraev et al., 2004, Leon et al., 2006, Papa et al., 2010, Zakharova et al., 2006]. This relation, referred to as a *rating curve* $R = T(H)$, is specific to each gauging station. This method conventionally needs simultaneous measurements of water level from altimetry and *in situ* runoff measurements. In order to avoid the limitation of having synchronous datasets, a statistical approach based on quantile function mapping was developed [Tourian et al., 2013]. The statistical approach provides the opportunity of extracting runoff values from altimetry data over rivers for which the runoff measurements at the selected gauges were made before the age of satellites.

The quantile function specifies, for a given probability $0 < p < 1$, the maximum value that R or H can attain with that probability. As water level and measured runoff should follow a monotonic relationship, it can be inferred that $T(\cdot)$ is a non-decreasing function. Hence, according to the Q-transformation rule [Gilchrist, 2000] $T(Q_{\mathrm{H}})$ must be a quantile function, because a non-decreasing function of a non-decreasing function must itself be a non-decreasing function. Therefore, the functional relationship $T(\cdot)$ can be defined through quantile function of water level measurements (Q_{H}) and quantile function of old runoff Q_{R}

$$Q_{\mathrm{R}} = T(Q_{\mathrm{H}}). \tag{4.1}$$

Since the time coordinate is not involved here explicitly, we eliminate the requirement of synchronous datasets.

Figure 4.2 shows the empirical quantile functions for water level and monthly discharge for the Mekong river and the resulting statistical discharge-water level relationship, that leads to a statistical rating curve for the virtual station on the Mekong River by fitting a quadratic curve. In fact, the rating curve is constructed here by eliminating the probability-coordinates, whereas conventionally the time-coordinate is eliminated. We are thus able to salvage pre-satellite altimetry runoff data and turn them into utilizable data for the satellite altimetry time frame.

Despite a good performance overall, the main limitation of our statistical approach is that it leads to erroneous discharge when discharge behavior of river is not stationary over time. Moreover, altimetry-based methods have inherent challenges: at times the virtual stations are located far away from the river outlet making the rating curve less effective; altimetry can only provide snapshots of the water level, which is not representative of daily or monthly runoff; the temporal resolution is dictated by the repeat-period of the altimetry satellites, which is usually not daily or monthly (35 days for the case of ENVISAT); and many of the current algorithms for retrieving runoff from water levels depend on *in situ* datasets [Tourian et al., 2013].

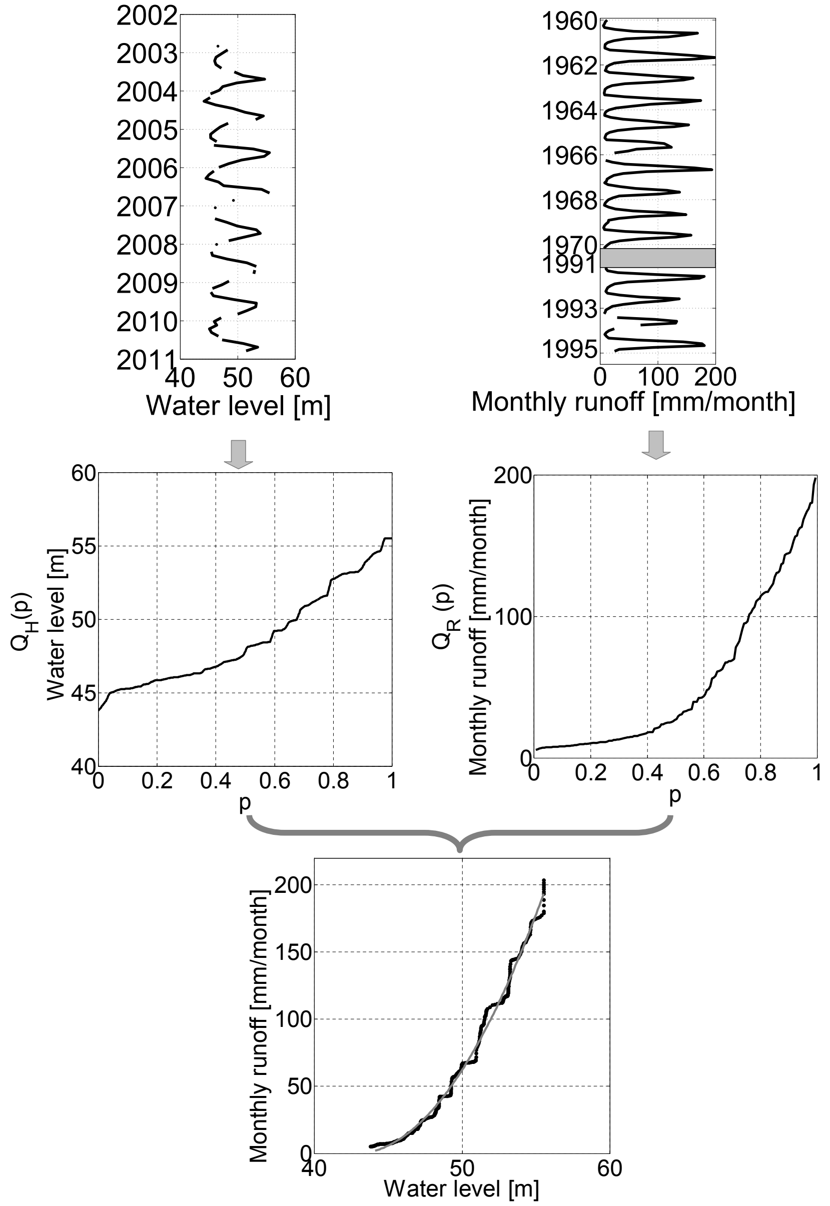

FIGURE 4.2: From Tourian et al. [2013]: Estimated water level from satellite altimetry and available discharge for the Mekong River (top) are transferred to the quantile functions (middle). From the corresponding probabilities the scatter plot of discharge versus water level is constructed . A smoothened rating curve is then obtained by fitting a quadratic line to the scatter (bottom). Note the dissimilar time axes of the two datasets (top).

GRACE-Based Approaches

The twin satellite system of GRACE provides weekly/monthly snapshots of the static geoid as spherical harmonic coefficients. After removing a long-term mean of the static field from the monthly snapshots, monthly anomalies of the geoid become visible. These monthly anomalies represent the sum total of all the geoid changes induced by a variety of geophysical phenomena, one of which is the continental water storage anomalies. In order to retrieve the continental water storage anomalies from the monthly geoid anomalies, contributions of the well-known phenomena like tides (ocean, solid earth and atmospheric), atmospheric and oceanic mass changes are removed. The remnant signal contains contributions from continental water storage, glacier, Arctic and Antarctic ice mass changes, and solid earth phenomena like glacial isostatic adjustment [Wahr et al., 1998]. For hydrology, continental water storage anomalies are the quantities of interest, and Wahr et al. [1998] derived a relationship between geoid anomalies and the water storage anomalies, which is given as

$$\begin{aligned}\Delta M(\theta,\lambda)[t] \;=\; & \frac{a_{\mathrm{e}}\,\rho_{\mathrm{ave}}}{3\rho_{\mathrm{w}}} \sum_{l=0}^{L} \frac{2l+1}{1+k_l} \times \\ & \sum_{m=-l}^{l} P_{lm}(\cos\theta)\,\Delta K_{lm}[t]\,e^{im\lambda}\,, \end{aligned} \tag{4.2}$$

where $\Delta M(\theta,\lambda)[t]$ is the water storage anomaly at co-latitude θ, longitude λ and time-point t expressed in terms of equivalent water height (m), ρ_{ave} is the average density of Earth (5515 kg/m^3), ρ_{w} is the average density of water (1000 kg/m^3), a_{e} is the semi-major axis of the Earth (6378.1363 km), k_l is the load Love number for spherical harmonic degree l, $P_{lm}(\cos\theta)$ is the normalized associated Legendre polynomial of degree l and order m, and $\Delta K_{lm}[t]$ are the complex spherical harmonic coefficients of temporal anomalies of the geoid after removing a long-term mean. In the above equations we have used square brackets for the temporal-coordinate to indicate the discrete nature of the observations. The water storage anomalies must be numerically differentiated to get the flux quantities as all the hydrological quantities are flux quantities. Therefore,

$$\frac{\mathrm{d}M(\cdot)[t]}{\mathrm{d}t} \approx \frac{1}{2\Delta t}\left(\Delta M(\cdot)[t+\Delta t] - \Delta M(\cdot)[t-\Delta t]\right), \tag{4.3}$$

where $\mathrm{d}M(\cdot)[t]/\mathrm{d}t$ is the water storage change quantity. Now that we have arrived at the quantity that can be used with the hydrological quantities, we will discuss the methods that employ GRACE data to estimate runoff.

Hydrological Approach

Water balance equations define the hydrological cycle in quantitative terms. In general, in any river basin the amount of water entering and leaving would lead to a change in the storage. This can be translated mathematically as

$$P - ET_{\mathrm{a}} - R = \frac{\mathrm{d}M}{\mathrm{d}t}, \tag{4.4}$$

where P is precipitation, ET_{a} represents the actual evapotranspiration, R denotes the river runoff and $\mathrm{d}M/\mathrm{d}t$ is the rate of water storage changes over a catchment. From this equation, one can compute the river runoff as

$$R_{\mathrm{hyd}} = P - ET_{\mathrm{a}} - \frac{\mathrm{d}M}{\mathrm{d}t}. \tag{4.5}$$

The benefit of such an approach, hereafter hydrological approach, is that the additional runoff routing component is not required. This is actually the case as the changes in the storage provided by GRACE represent the routing. Nevertheless, the method suffers from the accumulation of errors and biases from the three datasets, and also the resolution and accuracy will be limited to the weakest of the three datasets. Therefore, the quality of runoff estimates depends on the quality of the data sources of precipitation, evapotranspiration, and water storage changes.

Hydrometeorological Approach

According to Roads et al. [1994], the terrestrial water budget [Eq. (4.4)] can be linked to the atmosphere by the atmospheric water balance equation

$$\frac{\mathrm{d}W}{\mathrm{d}t} + \nabla \cdot \mathbf{Q} = ET_\mathrm{a} - P\,, \tag{4.6}$$

where W denotes the total column water content in the atmosphere. $\nabla \cdot \mathbf{Q}$ is the net balance of moisture flux (i.e., moisture flux divergence), which is defined as

$$\nabla \cdot \mathbf{Q} \equiv \frac{1}{g} \nabla \cdot \int_0^{p_\mathrm{sfc}} v_\mathrm{h}(p)\, q(p)\, dp \tag{4.7}$$

with air pressure p (Pa), the gravitational acceleration g (m s^{-2}), the horizontal wind vector $\vec{v}_\mathrm{h}$, and the specific humidity q (kg kg^{-1}). According to Peixoto and Oort [1992], changes in the atmospheric water content $\mathrm{d}W/\mathrm{d}t$ can be neglected on monthly or longer time scales—that is,

$$\nabla \cdot \mathbf{Q} = ET_\mathrm{a} - P\,. \tag{4.8}$$

By combining Eqs. (4.4) and (4.8), runoff can be derived from the atmospheric-terrestrial water balance

$$R_\mathrm{atm} = -\nabla \cdot \mathbf{Q} - \frac{\mathrm{d}M}{\mathrm{d}t}\,. \tag{4.9}$$

A potential advantage of this approach compared to the hydrological approach is that only two datasets are required for estimating runoff. Nevertheless, the performance of the hydrometeorological approach again depends on the quality of the data sources of moisture flux divergences and water storage changes.

Runoff-Storage Relationship

The direct comparison of measured runoff and storage mass determined by GRACE shows a distinct, yet seasonally characteristic, behavior for different climatic (tropical, boreal) and physiographic (surface, vegetation, soil etc.) conditions [Figure 4.3 (top left) for Amazon and Figure 4.3 (bottom left) for Ob]. Detailed investigations on the runoff-storage relationship [Riegger and Tourian, 2014] show that for Amazon the runoff–storage relationship can be characterized as a linear time invariant system with a phase shift, that corresponds to the time lag between runoff and storage. Thus a linear relationship is revealed between runoff and the hydraulically coupled liquid storage, once the time lag between runoff and the coupled liquid storage is taken into account and adapted [Figure 4.3 (top right) for Amazon]. Similar results were found for the relationship of mass storage from GRACE with surface water extent [Papa et al., 2008] and with rainfall [Frappart et al., 2013].

The time lag between runoff and water storage is caused by the finite transition time between input and hydraulic coupling to the discharge network. The data-based adaption solves the problem of runoff routing. Deviations from the linear behavior as in Figure 4.3 (bottom right) for Ob are interpreted as being caused by storage components which are not hydraulically coupled to the drainage system like floods, snow, etc. For a quantification of the hydraulically coupled liquid storage, which actually contributes to runoff we use satellite imagery data. For boreal catchments, MODIS snow coverage allows us to separate the coupled liquid storage from the decoupled solid storage, and allows us to assign the corresponding solid and liquid masses. We apply two conceptual approaches to utilize the snow coverage ratios for the quantification of solid and liquid storage:

1. The *solid approach* assumes a horizontal separation into a totally frozen part underneath the snow covered area of the catchment and liquid part in the rest. Such an approach is totally based on input from GRACE and MODIS snow coverage data.

2. The *snow approach* assumes a vertical separation into a liquid and a snow storage. For this case, the aggregation and release of snow or liquid recharge has to be calculated by a model based on recharge ($P - ET_\mathrm{a}$) for mass input and MODIS snow coverage for separation.

This mass separation together with the appropriate adjustment of the time lag leads to obtaining a high correlation coefficient between runoff and coupled liquid mass (Figure 4.3).

For unmanaged catchments, the linearity assumption of the runoff and liquid storage (RS-relationship) can in principle be used for the calculation of runoff from storage, and vice versa. For a full description of the linear relationship of runoff and storage, the absolute value of water storage is needed in order to achieve a runoff being proportional to storage. Thus, an unknown constant storage M_c has to be added to the monthly storage anomalies $\Delta M[t]$ to obtain the total storage mass:

$$M[t] \;=\; \Delta M[t] + M_\mathrm{c}\,. \tag{4.10}$$

The storage offset M_c corresponds to the long-term mean storage, which is determined by the steady-state condition in a catchment that long-term runoff must balance long-term recharge. As it cannot be directly observed, it has to be estimated in order to achieve proportionality. The scaling factor between runoff and adapted coupled mass corresponds to $\frac{1}{\tau}$ where τ is the hydraulic time constant. The time lag can be optimized by a combination of n discrete months ($n \geq 0$) and a continuous time shift $\omega \in [0,1]$ within one month backwards in time applied on coupled liquid mass, from which runoff can be determined as

$$R_\mathrm{rs}[t] \;=\; \frac{1}{\tau}\Big(\omega M[t-n-1] \;+\; (1-\omega) M[t-n]\Big). \tag{4.11}$$

Prerequisite is that the parameters—hydraulic time constant, time lag and mass offset—are determined by an adaption of coupled liquid mass from GRACE mass deviation and MODIS snow coverage to runoff from another time period. The adapted parameters are expected to depend only on hydraulic condition of the

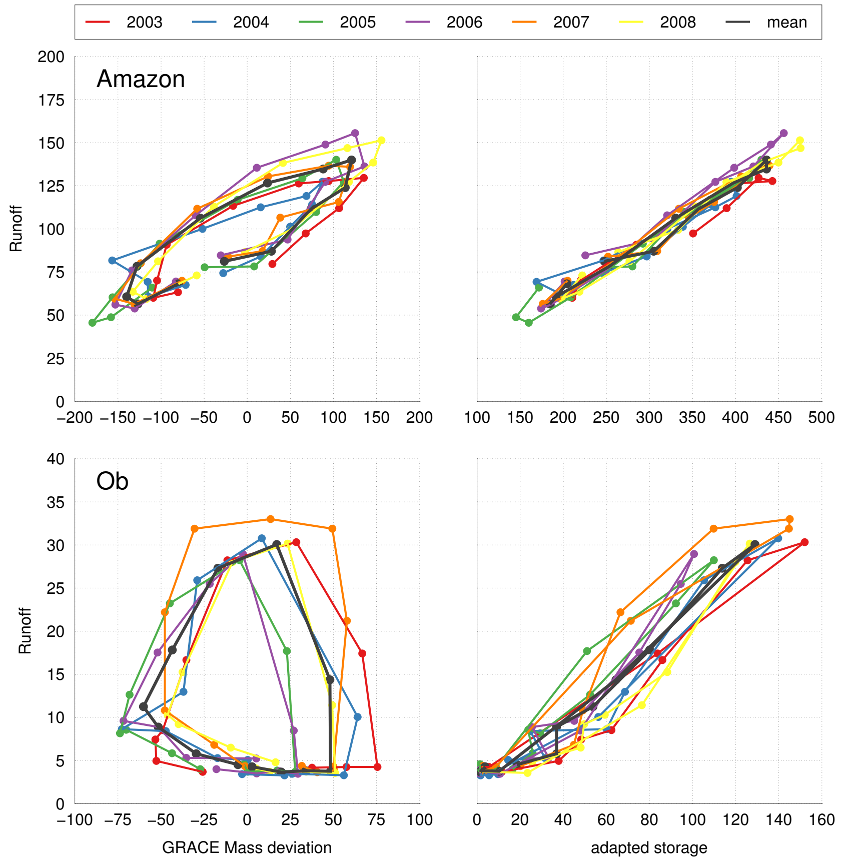

FIGURE 4.3: Monthly runoff observations versus GRACE mass deviation (left column) and adapted mass (right column), respectively, for the Amazon (top row) and Ob (bottom row) basin. The figures show both a comparison for individual years and for the mean monthly behavior.

drainage network and not on climatic conditions of the catchment and thus should remain constant as long as the drainage network is not changed. This means that they can be used for the determination of an unknown recharge from given GRACE mass deviation and remote sensing data. The detailed description of methodology and results can be found in Riegger and Tourian [2014].

Precipitation-Runoff Ratio

In addition to the hydro-geodetic approaches, we also use an empirical and rudimentary approach for estimating runoff of catchments, which depends on precipitation data and at least three years observed runoff data. We include this method in our comparison in order to verify if the hydro-geodetic approaches are able to provide as good estimates as the precipitation–runoff ratio approach, if not better.

It is well known that there is a strong relationship between precipitation and runoff [Zhou et al., 2012]. In this method, runoff is computed directly from precipitation. To that end, a ratio between runoff and precipitation is computed for each month individually. Therefore for each catchment the monthly runoff value at the j^{th} month (Jan.–Dec.) of the i^{th} year, R_{ij} is divided by corresponding precipitation value P_{ij}:

$$\kappa_{ij} = \frac{R_{ij}}{P_{ij}} . \tag{4.12}$$

Then for each month j, the ratio is computed by computing the monthly mean:

$$\bar{\kappa}_j = \frac{1}{N} \sum_{i=1}^{N} \kappa_{ij} , \tag{4.13}$$

where N is the number of years. The runoff is then computed for each month by:

$$R_{\text{pr}} = P_{ij} \; \bar{\kappa}_j . \tag{4.14}$$

In this study, for computing the $\bar{\kappa}_j$ values, precipitation and runoff data until the end of 2004 are used. The computed ratio is then multiplied with precipitation data of corresponding month after 2004.

The hydrological approach, hydro-meteorological approach, the runoff-storage relationship method and the precipitation-runoff ratio method all use global hydrological and hydro-meteorological datasets, but of these methods only the runoff-storage method is fine-tuned to the runoff of the particular catchment while the other methods are not. A similar argument can be placed for the precipitation-runoff ratio method, but the model is way too simplistic, and there is no fine-tuning of the ratio toward the catchment behavior. In the altimetry approach, the water level measurements, but are *in situ* measurements, but satellite-based, they are also catchment-specific methods. In view of this, the methods presented above can be coarsely classified as *catchment-specific* and *global* methods.

TABLE 4.1: Summary of the observation and model based datasets containing precipitation (P), actual evapotranspiration (ET_a), vertically integrated moisture flux divergences ($\boldsymbol{\nabla} \cdot \mathbf{Q}$), water storage changes ($\dot{M}$), and runoff (R_obs and R_mod). The references for the used datasets are as follows: GPCP: Adler et al. [2003], GLEAM: Miralles et al. [2011b], ERA-Interim: Simmons et al. [2006], GRACE-CSR: Bettadpur [2007], GLDAS: Rodell et al. [2004b]

Variable	Dataset	Version	Resolution		Time period
			Spatial	Temporal	
P	GPCC	2.2	2.5° × 2.5°	1 mo	1979–present
ET_a	GLEAM	–	0.25° × 0.25°	1 d	1984–2011
$\boldsymbol{\nabla} \cdot \mathbf{Q}$	ERA-Interim	–	0.75° × 0.75°	1 mo, 1 d, 6 h	1979–present
$\mathrm{d}M/\mathrm{d}t$	GRACE-CSR	R5	–	1 mo	2005–2010
R_obs	GRDC	–	–		
R_mod	GLDAS	Noah2 2.7	0.25° × 0.25°	1 mo, 3 h	1979–present

Data and Models

Gathering from (4.1), (4.4), (4.9) and (4.11), it is clear that we need datasets for precipitation, evapotranspiration, water levels from altimetry, observed runoff, vertically integrated moisture flux divergences and water storage changes from GRACE. Apart from geodetic datasets, the rest of the datasets are used as is without any modifications, while there is some processing involved prior to the usage of the geodetic datasets. The processing of geodetic datasets will be detailed in the subsequent sections, and the details of other datasets are summarized in Table 4.1, where the choice of the datasets was dictated by a global study done elsewhere [Lorenz et al., 2014]. The major conclusion of this study was that the runoff estimates from the different data combinations within the class of GRACE-based methods (i.e., R_hyd and R_atm) did not provide a combination that performed consistently over all the catchments. This led us to choose a combination that was overall the best among the combinations that were assessed (see Table 4.1).

Water Level from Satellite Altimetry

In this study, altimetry data from the Environmental Satellite (ENVISAT) is used from 2002 to 2010. We use ENVISAT data as its good performance over inland water

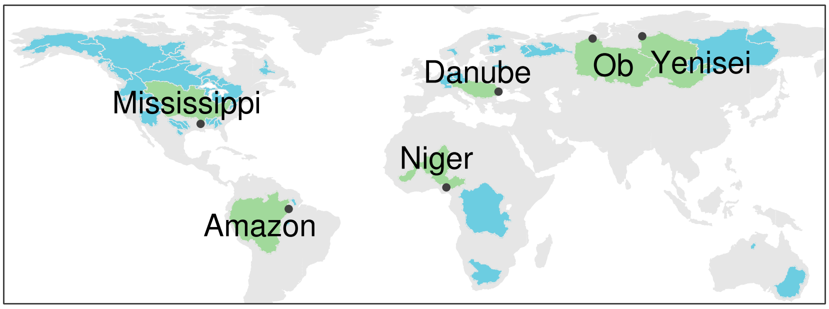

FIGURE 4.4: Location of the 6 study areas (green) with the location of the respective river gauges (black dot). The blue areas are 51 additional catchments, where at least three years of runoff observations between 2005 and 2010 are available. All these catchments were used in the correlation analysis shown in Figure 4.7.

bodies was acknowledged by previous studies [e.g., Da Silva et al., 2010, Frappart et al., 2006]. ENVISAT was launched in March 2002 as a successor to ERS-1 and ERS-2 by the European Space Agency (ESA). Where the ENVISAT groundtrack crosses the river the virtual station is defined. The water level time series for the virtual stations close to runoff gauging stations of ENVISAT-RA2 are obtained by employing an algorithm, which conforms to a large extent to the standard processing of altimetry data in hydrological applications [Da Silva et al., 2010]. For this study, we have selected virtual stations in six major catchments (Figure 4.4).

Within a certain radius around the virtual station (depends on the water body), a time series is generated from successive altimeter measurements at each satellite pass. As the along-track distance of 18 Hz measurements is ca. 370 m, the radius of virtual station should be defined carefully to assign water-reflected measurements, only. The water level time series H are then computed by subtracting the estimated range from retracking algorithm ρ and corresponding corrections c from the altitude of satellite h ($H = h - \rho - c$).

Our algorithm for obtaining the water level time series H mainly relies on the Ice-1 retracker, which according to Frappart et al. [2006] and Da Silva et al. [2010], provides the best estimate of inland water level variations. However, the water level measurements are optimized by employing information from the other retrackers as well (Figure 4.5). In fact, measurements from different retrackers can be combined as they are originally from the same waveforms. The algorithm for the combination

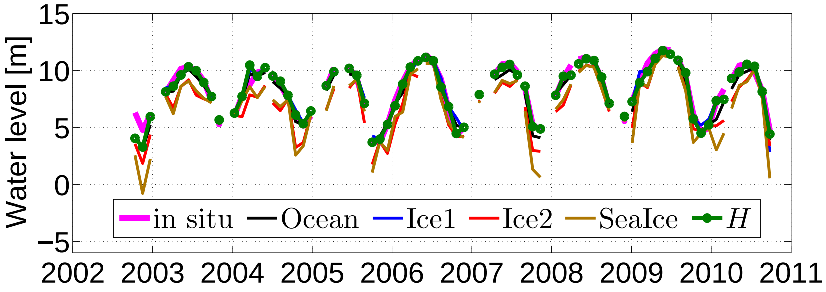

FIGURE 4.5: Water level time series from different retracking methods, combination approach and the *in situ* measurements for Amazon River at Obidos.

TABLE 4.2: Correlation and σ of residual of obtained water level time series from different retracking methods and the combination approach with *in situ* measurements together with percentage of contribution of different retrackers for creating H for Amazon.

	Ocean	*Ice-1*	*Ice-2*	*SeaIce*	H
Correlation	0.97	0.96	0.92	0.89	**0.98**
σ of residual [m]	0.54	0.58	0.89	1.29	**0.44**
Contribution to H	27%	69%	0%	4%	–

of retrackers implements the following criterion: over each pass/time pick up the water level from the retracker with the lower variation in terms of standard deviation within the virtual station and assign it to the water level time series. In case of the minimum standard deviation larger than one, the algorithm chooses the median of Ice-1 retrackers. Table 4.2 lists the computed correlation and standard deviation (σ) of water level residual time series from different retrackers and combination approach and contribution of each retracker to H for the Amazon's virtual station at Obidos.

GRACE Data

We have used 6 years (2005–2010) of GRACE release 5 datasets from the GeoForschungsZentrum [GFZ; Dahle et al., 2013] and Center for Space Research (CSR), The University of Texas at Austin [Bettadpur, 2012] processing centres. These are monthly snapshots of the static geoid given in terms of spherical harmonic coefficients, which after removing a long-term mean describe the monthly anomalies in the geoid.

These changes are ridden with noise from different sources: aliasing of residual tidal signal [Seo et al., 2008], poor estimates of C_{20} coefficient, and high-frequency noise in the spherical harmonic coefficients due to orbit geometry.

Aliasing of Residual Tidal Signal The residual tidal signal aliases into frequencies that depend on the orbit geometry of GRACE, sampling rate of GRACE and the natural frequency of the corresponding tides; they can be modeled and removed from the signal to a large extent. To this extent, we compute and remove the mean annual behavior of each of the spherical harmonic coefficients, after which we are left with residuals replete of any seasonal behavior. Then, a least squares spectrum analysis is carried out to estimate the coefficients of the aliased frequencies, which are then subsequently subtracted from the residuals of the spherical harmonic coefficients. The corrected spherical harmonic residuals are then added back to the mean annual behavior to obtain the corrected monthly anomalies.

Erroneous C_{20} Coefficient Due to the orbit configuration of GRACE satellites, they are insensitive to the C_{20} coefficient of the gravity field. Thus, the erroneous C_{20} coefficients are replaced with those estimated from Satellite Laser Ranging (SLR) in the case of the CSR monthly solutions [Cheng and Ries, 2013], while the C_{20} coefficients in the GFZ monthly solutions are used as-is as they are well-behaved. This, we believe, is due to the use of the EIGEN-6C gravity field model as a background model in the processing scheme [Dahle et al., 2013]. The EIGEN-6C gravity field model includes data from GRACE, GOCE and LAGEOS (SLR) satellites and also data from airborne and terrestrial gravimetry [ICGEM, 2013].

High-Frequency Noise It is well-known that the high-frequency noise in GRACE, which manifests itself as stripe-like patterns in the spatial domain, has to be dealt with by filtering. We apply regularization filter as proposed by Kusche [2007] but with certain modifications. The regularization filter requires signal and noise covariance matrices, and the modifications we make pertain to these covariance matrices. Kusche [2007] computes the signal covariance from a combination of different models, while

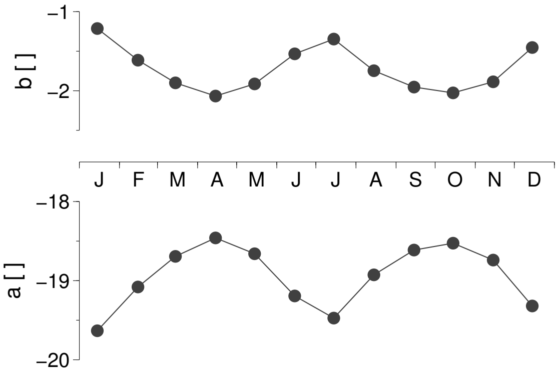

FIGURE 4.6: Estimates of a and b in (4.15), which define the cyclo-stationary signal covariance models. Here, we see a clear annual cycle with a characteristic double peak as we estimate these coefficients from quadratic quantities.

we compute the signal covariance from the GRACE data itself as described by Lorenz [2009]. The signal covariance so calculated is a power-law

$$\sigma_{lm}^2 = \frac{10^a\, l^b}{2l+1}, \tag{4.15}$$

where σ_{lm}^2 is the signal variance of a single spherical harmonic coefficient of degree l and order m, and a and b are the parameters that define the signal covariance model. We compute cyclo-stationary signal covariance models, which is achieved by calculating degree variances of the seasonal cycle of spherical harmonic coefficients. In other words, we compute a and b for every single calendar month in a year. The signal covariances so computed have a dominant annual behavior (Figure 4.6), which is reflective of the dominant annual cycle present in the temporal geoid anomalies of GRACE. As to the noise covariance, Kusche [2007] uses the same simulated noise covariance for all the monthly solutions, whereas we simulate and use the noise covariances for every month.

After post-processing the spherical harmonic coefficients, they are transformed into spatial grids of 0.5° × 0.5°. In order to arrive at the flux quantity, we compute the three-point central differences of the time series of mass variations. Further, we are only interested in catchment-wise values and not pixel-wise values, and therefore, we compute area-weighted averages for each of the catchments. This is mathematically expressed as

$$\Delta M(\chi)[t] = \sum_{i=1}^{N} \Delta M(\theta_i, \lambda_i)[t] \, \frac{A_i}{A_\chi} \tag{4.16}$$

$\Delta M(\chi)[t]$ refers to the area aggregated mass changes of individual catchments χ with area A_χ and i denote the individual pixels in the half-a-degree spatial grid with N being the total number of pixels in a given catchment. In the above equations we have used square brackets for the temporal-coordinate to indicate the discrete nature of the observations.

Precipitation

Precipitation, along with runoff, is one of the directly observable quantities amongst those involved in the hydrological cycle, and it serves as an input to hydrological and hydrometeorological models [e.g., Rodell et al., 2004b, Simmons et al., 2006]. Lorenz and Kunstmann [2012] show that the spatial distribution and change in the number of rain gauges over time is a major source of uncertainty in the precipitation observations. Moreover, large parts of, for example, Africa or Asia are completely unobserved, which leads to highly uncertain precipitation estimates over these regions. In this study, we use precipitation data from the Global Precipitation Climatology Project [GPCP; Adler et al., 2003]. The monthly product, which is provided in a rather coarse resolution of 2.5° × 2.5°, is a combination of remotely sensed precipitation over the oceans and gauge based estimates from the Global Precipitation Climatology Center [GPCC; Schneider et al., 2014] over land. Here, we use data from the most recent version 2.2 of the GPCP.

Evapotranspiration

Evapotranspiration has been one of the quantities that was difficult to observe, and more often sophisticated models were used in their place [e.g., Schlosser and Houser, 2007, Trenberth et al., 2007]. Recently, inroads have been made in both satellite-based [e.g., Long et al., 2014, Miralles et al., 2011b, Mu et al., 2007, Rodell et al., 2004b, Vinukollu et al., 2011] as well as terrestrial observations [Jung et al., 2010, Salvucci and Gentine, 2013]. Jung et al. [2010] were able to find meaningful trends in the rates of evapotranspiration from the terrestrial observations and reported an uncertainty of 3,000 km^3 for the mean global surface evapotranspiration. For the satellite-based estimates, Vinukollu et al. [2011] find that the estimates are in good agreement with the terrestrial observations at the annual scales, and also capture seasonal variations including the extreme events-suppression of evapotranspiration during droughts. Long et al. [2014] report an uncertainty of 10–30 mm/month depending on the satellite data after comparisons with FLUXNET data in southwest of United States of America. In this study, we use data from the Global Land-surface Evaporation: the Amsterdam Methodology [GLEAM; Miralles et al., 2011a], where evapotranspiration is estimated through a combination of a wide range of remotely sensed observations within a Priestley and Taylor-based framework.

Hydrological Models

As we know that hydrological models are also used in many studies, especially in the ungauged basins, their inclusion in the analysis, in spite of the presence of observed datasets, will only complement our analysis here. In this respect, we have included runoff estimates from the Global Land Data Assimilation System [GLDAS; Rodell et al., 2004b]. In particular, we used the estimates from the Noah LSM [Chen et al., 1996, Ek, 2003, Koren et al., 1999] which is driven by the GLDAS. This includes the following forcing data:

- Meteorological data from the Global Data Assimilation System [GDAS; Derber et al., 1991]

- Downscaled precipitation from the Climate Prediction Center's Merged Analysis of Precipitation [CMAP; Xie and Arkin, 1997]

- Satellite-derived downward radiation from the Air Force Weather Agency Agricultural Meteorology modeling system [AFWA AGRMET; Kopp and Kiess, 1996].

Zaitchik et al. [2010] compared runoff estimates from the different GLDAS members with observed runoff over several river basins. They concluded that there is no best realization which performs consistently good on a global scale. Similar results were presented in Lorenz et al. [2014]. Due to this reason, we selected only a single member from the GLDAS as a representative for land surface models. This should give an impression on the performance which can be expected from such models.

It should be noted that a major drawback of many global hydrological models is the lack of a routing scheme. Thus, the runoff output of a grid cell near the outlet can not be interpreted as the catchment's runoff—that is, there might be significant time shifts between modeled and true peak runoff. By temporal averaging, the impact of a routing scheme might become negligible, depending on the hydrological behavior of a catchment.

Results and Discussion

For evaluating the performance of the different approaches, we use the following metrics: the correlation coefficient between observed and estimated runoff, the Nash Sutcliffe coefficient [Nash and Sutcliffe, 1970] and percentage bias PBIAS [e.g., Gupta et al., 1999]. Analyzing the correlation coefficient helps in quantifying the common information content between observed and estimated runoff, although it is insensitive to amplitude and mean value. Therefore, the percentage bias

$$\mathrm{PBIAS} = 100 \frac{\sum\limits_{t=1}^{T} \left(R_{t,\mathrm{obs}} - R_{t,\mathrm{est}}\right)}{\sum\limits_{t=1}^{T} R_{t,\mathrm{obs}}} \tag{4.17}$$

is calculated, which gives relative information about the mean difference between the observations and estimates. In addition, we also compute the Nash Sutcliffe

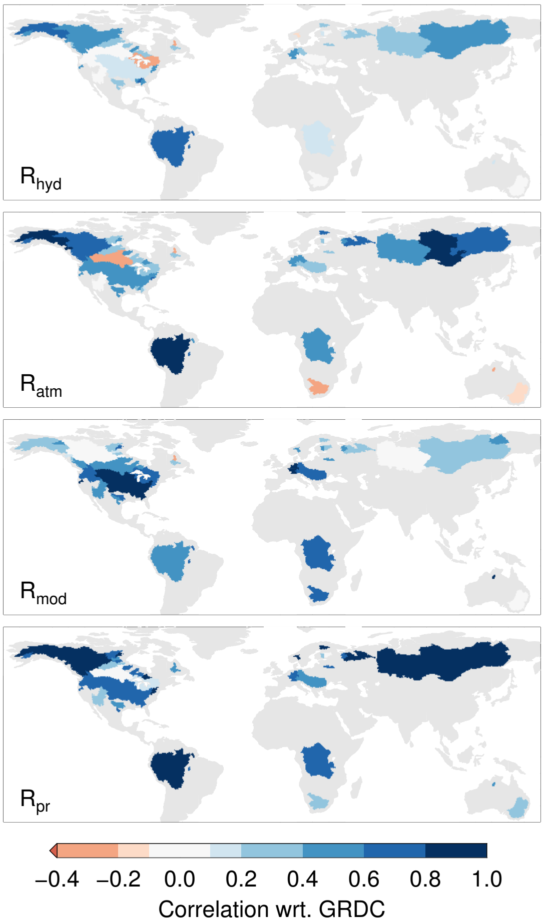

FIGURE 4.7: Correlation map between the runoff observations from GRDC and the four global approaches evaluated over 57 catchments.

Efficiency coefficient

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T}\left(R_{t,\mathrm{obs}} - R_{t,\mathrm{est}}\right)^2}{\sum_{t=1}^{T}\left(R_{t,\mathrm{obs}} - \overline{R}_{\mathrm{obs}}\right)^2} \tag{4.18}$$

with $\overline{R}_{\mathrm{obs}}$ being the long-term mean observed runoff. The NSE summarizes the overall agreement between the observations and estimates. It ranges between $-\infty$ and 1, where the coefficient tends to one if the estimated and observed runoff agree in phase, amplitude, and mean.

For an overview of other commonly used performance measures see, for example, Moriasi et al. [2007]. Table 4.3 summarizes the performance of the different approaches together with the mean annual runoff.

TABLE 4.3: Performance measures between the ensemble approaches, the catchment specific runoff estimates, and observations from GRDC for 5 selected catchments Amazon, Mississippi, Ob, Yenisei, and Danube where runoff estimates from altimetry or the *RS*-relationship are available. We have excluded the Niger basin as there is only one year of observations during the period 2005 to 2010 available.

		AM	MI	OB	YE	DA
	R_{obs}	101.1	17.5	11.5	21.6	25.9
Mean	R_{hyd}	101.7	16.0	17.9	11.1	34.1
	R_{atm}	85.3	7.4	10.5	17.6	15.6
	R_{mod}	63.7	6.8	4.9	7.8	14.0
	R_{pr}	96.7	16.4	12.1	21.9	24.8
	R_{alt}	102.6		10.8		22.9
	R_{rs}	104.6		10.4	20.1	
Corr	R_{hyd}	0.89	0.34	0.16	0.56	0.45
	R_{atm}	0.87	0.72	0.60	0.86	0.73
	R_{mod}	0.39	0.86	-0.12	0.51	0.84
	R_{pr}	0.93	0.78	0.96	0.98	0.74
	R_{alt}	0.98		0.94		0.94
	R_{rs}	0.97		0.97	0.88	
PBIAS [%]	R_{hyd}	-0.6	2.9	-57.0	48.6	-28.6
	R_{atm}	15.6	54.9	4.8	18.6	39.2
	R_{mod}	37.0	61.2	56.5	63.9	44.9
	R_{pr}	4.3	5.0	-6.1	-1.2	4.1
	R_{alt}	1.9		-0.7		8.7
	R_{rs}	-3.5		12.5	9.5	
NSE	R_{hyd}	0.79	-0.66	-2.48	-0.36	-1.60
	R_{atm}	0.41	-1.56	0.06	0.60	-1.95
	R_{mod}	-2.20	-0.86	-1.85	-0.81	-0.96
	R_{pr}	0.82	0.57	0.88	0.90	0.53
	R_{alt}	0.95		0.89		0.75
	R_{rs}	0.91		0.74	0.33	

The starting point of our discussion is the inconsistency of the GRACE-based approaches, which is expressed in terms of the correlation with observed runoff. While some catchments are highly correlated, there are catchments with low or even negative correlation, independent of their size (Figure 4.7). For example, R_{mod} performs well over the Mississippi basin, while R_{hyd} shows rather low correlation. Over boreal catchments, R_{pr} and R_{atm} seem to outperform the other two global approaches. Such

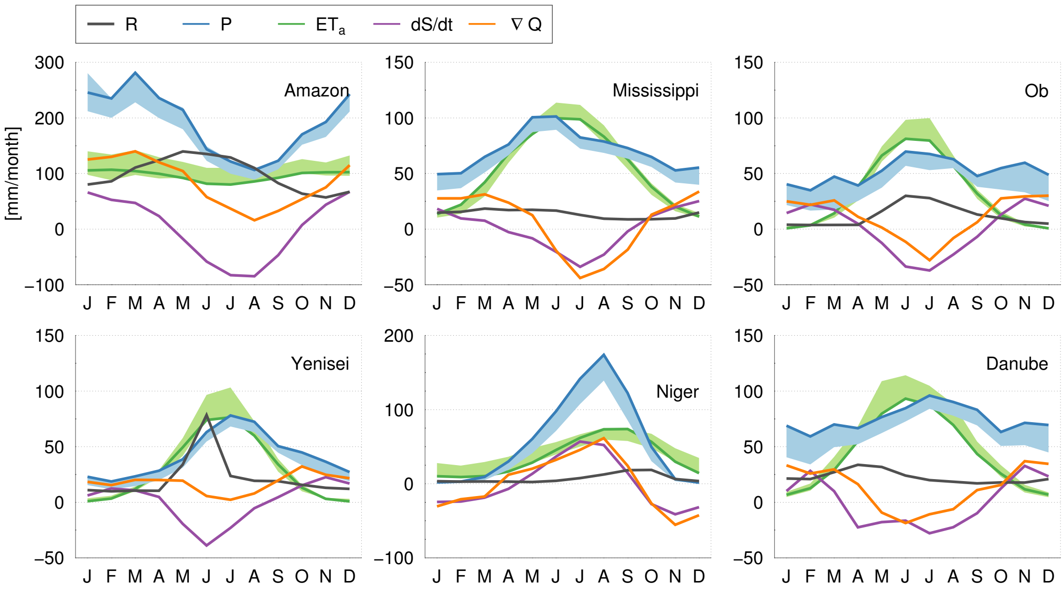

FIGURE 4.8: Seasonal cycle of precipitation (blue), evapotranspiration (green), runoff (black), and water storage changes (purple) for the 18 largest analyzed river basins. The bands are defined by the maximum and minimum value of an ensemble of P (GPCC, GPCP, DEL, CRU, and CPC) and ET_a (GLEAM, ERA-Interim, GLDAS, MERRA, MERRA Land) datasets. The blue and green lines are derived from GPCC and GLEAM, respectively.

inconsistencies underline the limitations of the global approaches and the need for catchment-specific methods like altimetry and the *RS*-relationship.

Here, we assess the estimated runoff from the hydro-geodetic approaches in six selected catchments, which vary in their climatic and hydrological behavior: Amazon, Mississippi, Ob, Yenisei, Niger, Danube. The hydrological behavior is characterized by the seasonal cycle of precipitation, evapotranspiration, moisture flux divergence, runoff, and water storage changes (Figure 4.8). The Amazon catchment with high precipitation rates of up to 250 mm/month during the rainy season show almost constant evapotranspiration throughout the year with higher amount of runoff in southern hemisphere winter. The warm summer in the Arctic tundra catchment of Ob (especially in its southern part) leads to high evapotranspiration and runoff between May to September. Over the Yenisei basin, the climate considerably changes along the river. However, high runoff driven by high precipitation and melting occurs during the summer. Unlike Yenisei, the peak runoff values for Niger, Danube, and Mississippi do not significantly contribute in their storage behavior.

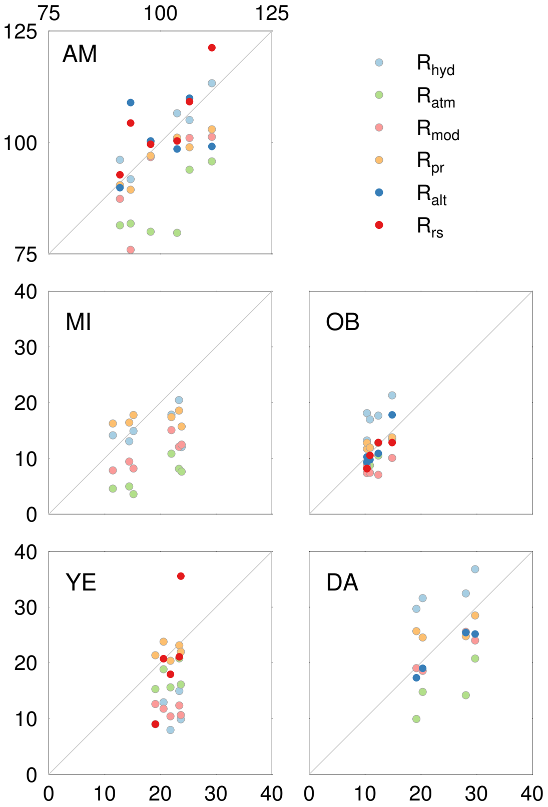

FIGURE 4.9: Scatterplots of annual mean runoff in [mm/month] derived from the global approaches R_{hyd}, R_{atm}, R_{mod}, and R_{pr}, as well as the catchment specific approaches R_{alt} and R_{rs} with respect to GRDC. The observations refer to the x-axis, while the y-values are given by the runoff estimates. We have excluded the Niger basin as there is only one year of observations during the period 2005 to 2010 available.

Overall, our assessment shows that the catchment-specific estimates R_{alt} and R_{rs} together with R_{pr} perform consistently well in terms of the long-term mean runoff (mean and PBIAS in Table 4.3), although their annual means might scatter from the observations (see Figure 4.9). On the other hand, R_{atm} and R_{mod} seem to underestimate consistently (e.g., positive PBIAS of R_{atm} and R_{mod}). For R_{hyd}, we can observe both over- and underestimation of the mean runoff, depending on the catchment. It should be mentioned that long-term and annual mean storage changes tend to be negligible (Figure 4.8), except in catchments with a significant storage gain or loss like, for example, Ganges [Rodell et al., 2009]. Thus, deviations from observed runoff are mainly caused by biases in the precipitation, evapotranspiration, or moisture flux divergence datasets.

In order to avoid these biases within the runoff estimates, we subsequently removed the long-term average from each time series, hereafter referred as *runoff anomalies* (Figure 4.10, right column). We also computed the seasonal cycle of each runoff estimate (Figure 4.10, left column).

Amazon Basin All approaches (except for R_{mod}) estimate the seasonal cycle of the runoff anomalies over Amazon very well in terms of amplitude and phase. This is not surprising as the monthly time series are in very good agreement with the observed runoff as well, resulting in correlation coefficients larger than 0.8 for all approaches (Table 4.3). The exception to this agreement is R_{mod}, where the low correlation of 0.4 is due to a time-shift of about two months. However, Syed et al. [2005] also estimated runoff from a terrestrial-atmospheric water budget for the Amazon basin. While their results refer to a different study period, they could also observe a significant underestimation of R_{atm} (62.2 mm/month) compared to the observations (82.2 mm/month), which agrees with our computations ($R_{\mathrm{obs}} = 82.2$ mm/month vs. $R_{\mathrm{atm}} = 62.2$ mm/month).

The catchment-specific time series R_{alt} and R_{rs} are much smoother than those from the other approaches. The short-term fluctuations are mainly due to the imbalances in the global hydrological and hydrometeorological datasets that are used in these methods. Interestingly, a considerable discrepancy between R_{obs} and R_{rs} in 2009 can be noticed, which can not be associated to errors in the GRACE data due to the excellent signal-to-noise ratio over Amazon. This discrepancy, which might be due to inaccurate runoff measurements during a severe flood event, is also confirmed by Chen et al. [2010].

Mississippi and Danube Basins Compared with the Amazon basin, the amplitudes of the seasonal variations over the mid-latitude basins Mississippi and Danube are not pronounced. These subtle seasonal variations are mainly due to the comparable and large amplitudes of precipitation and evapotranspiration, which effectively get differenced in the balance equations (Figure 4.8). In the presence of a small but significant storage, this results in a rather constant runoff throughout the year. The differencing leads furthermore (a) to noisy behavior in the monthly time series from the global approaches, resulting in negative NSE of R_{hyd} and R_{atm}, and (b) sometimes negative runoff estimates. Similar to the Amazon basin, Syed et al. [2005] also estimated too low runoff from a hydrometeorological approach over the Mississippi basin. As even R_{pr} does not correlate substantially, only the catchment-specific methods R_{alt} of R_{rs} seem to provide reliable runoff estimates at mid-latitude catchments.

Ob and Yenisei Basins The Siberian catchments Ob and Yenisei show a pronounced seasonal cycle. However, in the case of Ob basin the behavior of R_{hyd} and R_{atm} is noisy, which is, again, due to the low amplitude of runoff. In contrast, the R_{pr} estimates, which also use the global precipitation datasets, show very good agreement with the observations over both the Ob and Yenisei basin. This is also indicated by the performance metrics (Table 4.3), where R_{pr} can even compete with the catchment-specific methods, which again provide the most reliable runoff estimates. In contrast to the Ob basin, the R_{atm} estimates over the Yenisei basin agree very well with the observations that could be explained with a very distinct peak of the runoff-anomalies of more than 50 mm/month during June. However, R_{atm} significantly overestimates the peaks in 2005 and 2008. As the other two GRACE-based estimates from R_{rs} and R_{hyd} do not show such inconsistencies, these peaks are likely due to too large moisture flux convergences from ERA-Interim. Similar to the Amazon basin, GLDAS-Noah2 again shows a time-shift of about two months and a consistent underestimation of runoff in both catchments.

Niger Basin As the observations cover only the first year of the study period, the Niger basin can be considered as *quasi-ungauged*. R_{pr} and R_{alt} (even with some data gaps) provide a smooth annual cycle, which matches the observations of 2005. R_{hyd} catches the seasonal variations as well, but estimates too early and too high peak runoff. This might be explained by an erroneous evapotranspiration, because (a) the precipitation-based R_{pr} estimates agree very well with catchment-specific R_{alt} estimates, and (b) an unlikely time-shift in GRACE. R_{atm} also shows a very noisy

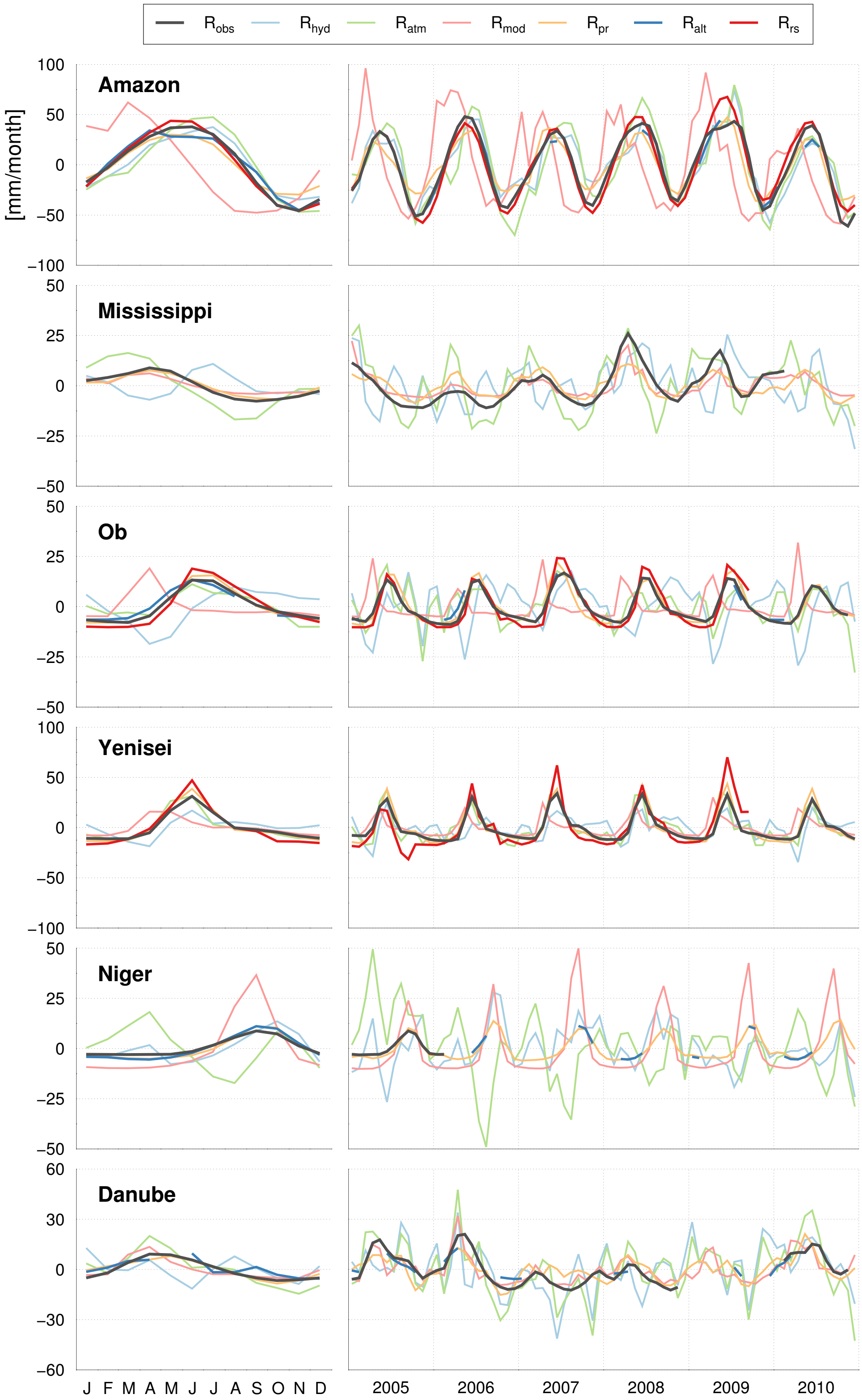

FIGURE 4.10: Runoff anomalies depicted as time series of monthly values (right column) and the seasonal cycle (left column) derived from GRDC (R_{obs}), hydrological (R_{hyd}) and atmospheric (R_{atm}) water budgets, GLDAS-Noah2 (R_{mod}), precipitation-runoff ratio (R_{pr}), altimetry (R_{alt}), and a runoff-storage relationship (R_{rs}).

behavior and does not, therefore, provide reasonable runoff estimates over the Niger basin.

Summary and Conclusions

In this contribution, we assessed the skill of four different hydro-geodetic approaches for estimating runoff. In addition, a primitive method that exploits the relationship between precipitation and runoff was also assessed. The skill was evaluated against observed runoff data from GRDC. In order to show the skill of these methods in comparison with recent hydrological models, they were also evaluated against the publicly available GLDAS-Noah2 model. The assessment was based on certain performance metrics, namely correlation, Nash Sutcliffe Efficiency (NSE) coefficient, and relative bias (PBIAS).

Our results show that catchment-specific methods (altimetry and runoff-storage relationship), despite their limitations, clearly outperform the global methods (hydrological and hydrometeorological approaches). The relatively simple method of the precipitation-runoff ratio showed comparable performance with the catchment-specific methods over the six study regions. Although the global methods have the potential to provide runoff over all landmasses that implies gauged and ungauged basins alike, they are ridden with accumulated errors and biases from the GRACE data and the global hydrological and hydrometeorological datasets that they use. It was also clear from the assessment that the hydro-geodetic approaches and the precipitation-runoff ratio approach outperform the modeled runoff, where the primary problem seems to be a phase shift. The phase shift as suspected seems to be an artifact of the lack of a routing scheme, which is a prerequisite for modeled runoff, especially at shorter time scales. Apparently, as the GRACE-based approaches take water storage changes as an input, they do not require a routing scheme and, hence, do not suffer from such a phase shift.

A major drawback of the global GRACE-based approaches is that the performance changes significantly from basin to basin. Our results show that, especially those catchments where runoff is low and the seasonal variations are not pronounced (here: Niger, Mississippi, Danube), inconsistencies in the hydrological and hydrometeorological datasets might lead to large biases and unrealistic short-term fluctuations in the runoff estimates. With improvements in the quality of global hydrological and hydrometeorological datasets, global methods will provide better runoff estimates. Hence, reliable estimates for ungauged basins can also be expected. It should be further noted that the performance of the GRACE-based methods, despite their shortcomings, is still comparable to other runoff-specific approaches [e.g., Vergnes and Decharme, 2012, Zaitchik et al., 2010] or purely model-based results [e.g., Lohmann et al., 2004].

While the catchment-specific methods outperform the global approaches, it must be noted that they rely heavily on at least an asynchronous runoff time series. These methods might also be inadequate in catchments with significant anthropogenic influences and those with non-stationary behavior. This really calls for catchment-specific methods that are independent of runoff observations. Nevertheless, the altimetry and runoff–storage relationship methods hold promise for filling data gaps in observed runoff time series and also extending now defunct gauged stations.

Acknowledgments

We gratefully acknowledge the support of projects SN13/1, BA1150/11, KU2090/1 by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the priority program SPP1257 *Mass Transport and Mass Distribution in the System Earth*. We would like to thank the following data providers: Global Runoff Data Centre (GRDC) `www.bafg.de/GRDC`, Global Precipitation Climatology Project (GPCP) `precip.gsfc.nasa.gov`, European Centre for Medium-Range Weather Forecasts (ECMWF) `www.ecmwf.int`, University of Texas at Austin, Center for Space Research (CSR) `www.csr.utexas.edu/`, NASA's Earth Science Division and Goddard Earth Sciences (GES) Data and Information Services Center (DISC) `ldas.gsfc.nasa.gov`.

Further thanks are due to Dr. Diego Miralles (University of Bristol) for kindly providing the GLEAM evapotranspiration data. All graphics have been produced with the Generic Mapping Tools [GMT; Wessel and Smith, 1991].

Chapter 5

Article IV: Basin-Scale Runoff Prediction: An Ensemble Kalman Filter Framework based on Global Hydrometeorological Datasets

---

Lorenz, C., M. J. Tourian, B. Devaraju, N. Sneeuw, and H. Kunstmann, 2015: Basin-Scale Runoff Prediction: An Ensemble Kalman Filter Framework based on Global Hydrometeorological Datasets, *Water Resources Research*, **51**, 8450–8475, doi: 10.1002/2014WR016794[1]

Abstract

In order to cope with the steady decline of the number of *in situ* gauges worldwide, there is a growing need for alternative methods to estimate runoff. We present an Ensemble Kalman Filter based approach that allows us to conclude on runoff for poorly or irregularly gauged basins. The approach focuses on the application of publicly available global hydrometeorological datasets for precipitation (GPCC, GPCP, CRU, DEL), evapotranspiration (MODIS, FLUXNET, GLEAM, ERA-Interim, GLDAS), and water storage changes (GRACE, WGHM, GLDAS, MERRA Land). Furthermore, runoff data from the GRDC and satellite altimetry derived estimates are used. We

---

follow a least squares prediction that exploits the joint temporal and spatial auto- and cross-covariance structures of precipitation, evapotranspiration, water storage changes, and runoff. We further consider time-dependent uncertainty estimates derived from all datasets. Our in-depth analysis comprises 29 large river basins of different climate regions, with which runoff is predicted for a subset of 16 basins. Six configurations are analyzed: the Ensemble Kalman Filter (Smoother) and the hard (soft) Constrained Ensemble Kalman Filter (Smoother). Comparing the predictions to observed monthly runoff shows correlations larger than 0.5, percentage biases lower than ±20 %, and NSE values larger than 0.5. A modified NSE metric, stressing the difference to the mean annual cycle, shows an improvement of runoff predictions for 14 of the 16 basins. The proposed method is able to provide runoff estimates for nearly 100 poorly gauged basins covering an area of more than 11,500,000 km^2 with a freshwater discharge, in volume, of more than 125,000 $^{\mathrm{m}^3}/_{\mathrm{s}}$.

Introduction

The percentage of geographically and temporally accessible runoff is of crucial importance for sufficient freshwater supply. Early estimates on the human appropriation of accessible runoff have been presented in Postel et al. [1996], and they raised awareness on the problem of declining water availability per capita, primarily due to a growing world population. The increase in the population of 45 % stands opposed to an increase of accessible runoff of only 10 % within 30 years. Falkenmark and Rockström [2004] estimated an increase of water requirements by about 1,300 m^2 per capita per year for each additional person. Vörösmarty et al. [2010] stress that 80 % of the world's population is exposed to high levels of threat to water security and that 65 % of the total continental discharge is classified as moderately to highly threatened. Accordingly, recent hydrological and hydrometeorological research must aim at an improved observation, modeling, and understanding of the terrestrial water cycle, in particular runoff.

The importance of long-term monitoring of hydrological variables was recognized by the World Meteorological Organization (WMO) already in 1980, which led to the initiation of the Global Runoff Data Centre (GRDC). The GRDC collects and harmonizes global runoff observations from national hydrological services and makes them available to the public. However, over many catchments around the world, runoff is not gauged [Blöschl et al., 2013].

In Dai and Trenberth [2002], it is reported that the annual runoff rate over the unmonitored areas is comparable to that over the monitored areas. Furthermore, studies by Sivapalan et al. [2003], Shiklomanov et al. [2002], Milzow et al. [2011], and Fekete et al. [2012] stress that the number of active river gauges is steadily decreasing. Figure 5.1a exemplary shows river basins without any measurements and basins with more than five years of missing data during the period 2000 and 2010. Since a similar decline can be observed for rainfall gauges [Lorenz and Kunstmann, 2012], the hydrological community has to face a significantly shrinking database for the most important water cycle variables.

The serious problem of runoff data availability and its importance for hydrological cycle studies led the International Association of Hydrological Sciences (IAHS) to launch the study framework Prediction in Ungauged Basins (PUB). The goal of 10 years of PUB was to predict the hydrological cycle in ungauged basins by improving the understanding of climatic and landscape controls on hydrological processes [Blöschl et al., 2013]. It has been realized that on top of data quantity, data quality also tremendously affects the performance of modeling [McMillan et al., 2010].

With respect to river runoff, one of the motives of PUB was to seek for spaceborne alternatives, as traditional runoff data acquisition at *in situ* gauges is typically costly and time consuming. These alternatives indeed have the potential to be highly valuable for hydrology [Alsdorf et al., 2007]. In terms of runoff, it has been already demonstrated by Koblinsky et al. [1993] that satellite-based altimetry can be used to describe river level variations. Therefore, many studies made use of satellite data in order to analyze the water cycle over ungauged basins [e.g., Khan et al., 2011]. However, Alsdorf et al. [2007] already indicated high uncertainties in satellite derived

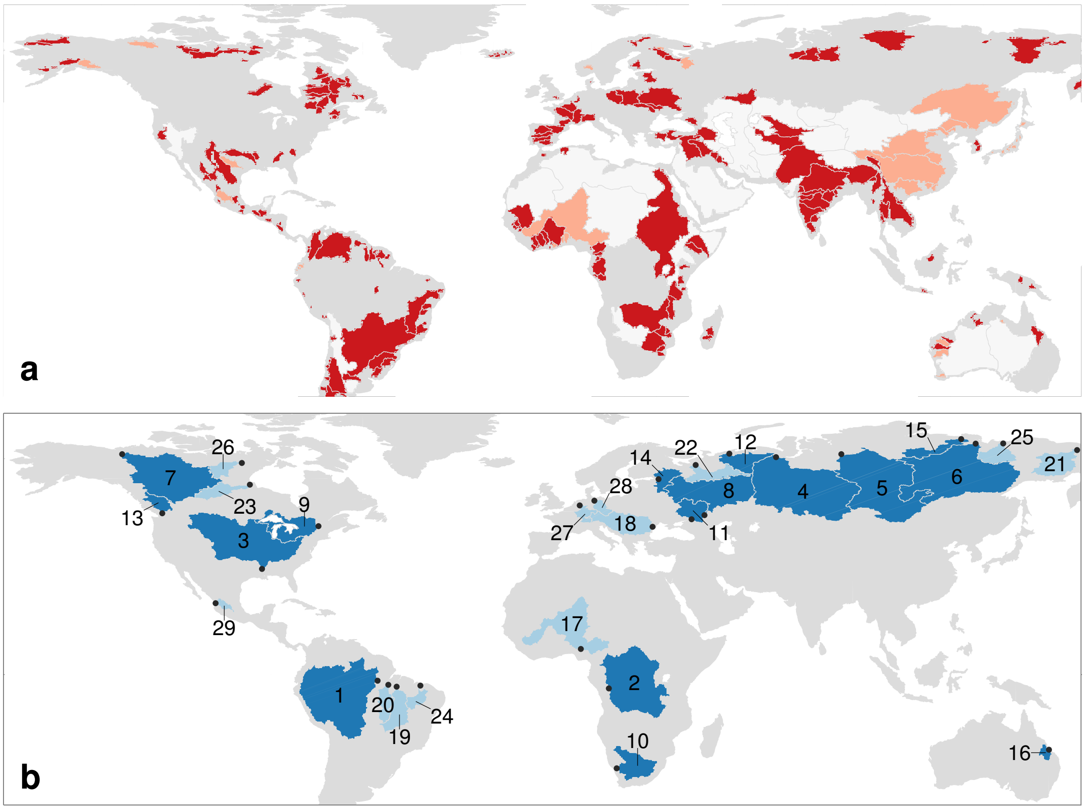

FIGURE 5.1: (a) Distribution of ungauged and poorly gauged catchments. The dark red areas indicate catchments without any measurements between 2002 and 2012 while light red basins contain more than five years of missing values during that period. The white catchments can be considered to be dischargeless (e.g., deserts). (b) Location of the 29 basins given in Table 5.2 where the dot in each basin is located at the respective river gauge. The 16 dark blue shaded basins are used for validating the runoff predictions in section *Results and Discussion*.

hydrological data due to poor spatial and temporal resolutions and the inabilities to penetrate clouds or smoke. In terms of precipitation, for example, Aghakouchak et al. [2012] analyzed different systematic and random errors in satellite-derived precipitation datasets over the United States. It was further stressed by Fekete et al. [2012] that there is still no substitute for *in situ* discharge monitoring. Thus, the combination and harmonization of spaceborne and terrestrial data is a big challenge of the current hydrological research.

Another major task of the hydrological community is the closure of the terrestrial water budget. Even if the basic equation of the terrestrial water cycle

$$P - ET - R - \dot{M} = 0 \tag{5.1}$$

with precipitation P, evapotranspiration ET, runoff R, and water storage changes $\dot{M}$ appears simple, it is known from many studies that a real closure on larger spatial scales can not be achieved with current data sources.

For a long time, a major issue of analyzing the hydrological cycle was missing observation based data of the terrestrial water storage. In essence, direct observation of the water storage components (e.g., snow, ice, surface water, soil, (un)saturated underground storages) was not possible on large spatial scales [Rodell and Famiglietti, 1999]. Measurements of water storage changes by measurements of changes in groundwater levels and in soil water saturation are not reliable due to the insufficiently known storage coefficients and also the inadequate density of monitoring points on large spatial scales [Riegger et al., 2012].

Only after the launch of the twin satellites of the Gravity Recovery and Climate Experiment GRACE, closing the water budget on shorter time scales became possible [Tapley, 2004]. Sheffield et al. [2009] combined satellite-based precipitation, model-based evapotranspiration and GRACE data to close the water budget over the Mississippi River basin over the time period of 2003—2005. Rodell et al. [2004a] closed the water budget using GRACE at basin-scales to estimate evapotranspiration as the residual of water budget. In Syed et al. [2009, 2005, 2007], GRACE data has been used for deriving freshwater discharge using a coupled land-atmosphere water balance.

In a similar study, Sneeuw et al. [2014] compared the performance of estimated runoff derived from water balance closure with those from models and *in situ* data. In a more extensive analysis, Lorenz et al. [2014] analyzed combinations of different gridded observations, atmospheric reanalysis models, land surface hydrological models, partially-model-based datasets and GRACE to derive and evaluate an ensemble of hydrological and hydrometeorological budget-based runoff estimates. However, most of these studies did not close the water budget explicitly, but estimated a single water cycle variable from the remaining ones. Thus, all the errors and uncertainties in the input data sources inevitably propagate to the estimated variable.

Other approaches, which make use of data assimilation techniques, are presented in, for example, Pan and Wood [2006]. The authors developed a Constrained Ensemble Kalman Filter (CEnKF), which enforced water budget closure, thereby, providing a constrained best estimate of the basin-scale water budget. Similarly, Pan et al. [2012] combined estimates of the terrestrial water budget from different data sources and to enforce the water balance constraint using the constrained Kalman filter technique.

In general, the term data assimilation is used for combining observation data with hydrological, hydrometeorological, or land surface models. The application usually ranges from extensive global reanalysis models, which simulate the whole atmosphere and the land surface [e.g., Dee et al., 2011], to models which focus on the land surface only [e.g., Reichle et al., 2002, Rodell et al., 2004a]. However, it is well known that our current model systems suffer from biases or errors in the simulated water cycle variables, which finally lead to imbalances in the water budgets [e.g., Lorenz and Kunstmann, 2012]. We therefore propose a data assimilation framework, where the mathematical description is kept as simple as possible. In fact, the approach is based on the terrestrial water budget equation (5.1) only, which can be considered as the most simple hydrological model. On the other hand, we want to exploit all the advantages and strengths from widely used global available datasets for the major water cycle variables, and thus, use real data instead of complex model equations within an Ensemble Kalman Filtering framework. The algorithm is, therefore, proposed as an alternative post-processing tool, which can be used for combining, correcting, and predicting basin-scale time series of the four major water cycle variables. In this study, however, we focus on the estimation of runoff as an performance evaluation step of the proposed approaches.

The datasets that are used are presented in section *Data*, while we describe the methods in section *Methods*. In section *Results and Discussion*, different configurations of the assimilation framework are analyzed in order to obtain the best possible estimates. These estimates are then compared against monthly observations over 16 large catchments during the period 2005 until 2010. The performance of the estimates is analyzed using the common performance metrics correlation, PBIAS, and NSE. However, as runoff over many catchments follows a dominant annual cycle, we further analyze a NSE-type metric, which relates the estimates to the mean annual cycle from historic data. In section *Comparison to similar Studies,* the performance of the presented approach is compared with similar studies. Finally, section *Conclusion* contains a conclusion of the major findings and an outlook.

Data

A detailed description of the applied datasets is given in Lorenz et al. [2014] and Sneeuw et al. [2014]. Here, only a brief overview of the different datasets of the four water cycle variables is provided. The most important information is further summarized in Table 5.1.

Runoff

In order to attain maximum data coverage, both in space and time, data are collected from different sources, namely the Global Runoff Data Center (GRDC), the ArcticRIMS project, Water Survey Canada, the United States Geological Survey, United States Army Corps of Engineers (USACE), the ORE HYBAM project, the Department of Water, Land and Biodiversity Conservation, Government of Australia, and the Department of Water Affairs and Forestry, Republic of South Africa.

The runoff database is further enhanced by using estimates from satellite altimetry over the Amazon, Ob, Don, and Danube basins. The methods for deriving runoff from altimetry data are described in Tourian et al. [2013]. They proposed a statistical approach based on quantile functions to infer a functional relation between altimetric water level and historic river gauge data. Using such a statistical function, the water level measurements from satellite altimetry is then mapped to an estimation of runoff at river gauges.

TABLE 5.1: Summary of the observation and model based datasets containing precipitation (P), actual evapotranspiration (ET), water storage changes ($\dot{M}$), runoff from river gauges (R_{obs}), and satellite altimetry (R_{alt}).

Variable	Dataset	Version	Resolution		Time period
			Spatial	Temporal	
P	GPCC	6.0	0.5° × 0.5°	1 month	1901–2010
	GPCP	2.2	2.5° × 2.5°	1 month	1979–present
	CRU	3.22	0.5° × 0.5°	1 month	1901–2013
	DEL	3.02	0.5° × 0.5°	1 month	1900–2010
ET	ERA-Interim	–	0.75° × 0.75°	1 month, 1 day, 6 h	1979–present
	GLDAS	Noah3.3	1.0° × 1.0°	1 month, 3 h	1948–present
	GLEAM	v1B	0.25° × 0.25°	1 day	1984–2007
	MOD16	A2	0.5° × 0.5°	1 year,1 month, 8 days	2000–2013
	FLUXNET MTE		0.5° × 0.5°	1 month	1980–present
$\dot{M}$	GRACE-GFZ	R5	–	1 month	2002–present
	GRACE-CSR	R5	–	1 month	2002 – present
	MERRA Land	1.0	$\frac{1}{2}^\circ \times \frac{2}{3}^\circ$	1 month, 1 day, 1 h	1980–present
	GLDAS	Noah3.3	1.0° × 1.0°	1 month, 3 h	1948–present
	WGHM	nouse	0.5° × 0.5°	1 month	1960–2009
R_{obs}	GRDC	–	–		
R_{alt}	Tourian et al. [2013]	–	–		

Precipitation

Four different observation based precipitation datasets, namely the data from the Global Precipitation Climatology Center [GPCC; Schneider et al., 2008], the Global Precipitation Climatology Project [GPCP; Adler et al., 2003], the Climatic Research Unit [CRU; Harris et al., 2014] and the University of Delaware [DEL; Matsuura and Willmott, 2012] are used. From these datasets, GPCC, CRU, and DEL are based on gauge observations only. The low spatial resolution GPCP product is a combination of different rainfall sensors. Currently, the GPCP includes microwave, infrared, and gauge based observations of precipitation [e.g., Huffman et al., 2009]. The largest number of stations are included in the GPCC product, while CRU and DEL are based on a much smaller number of rainfall gauges. Despite the pure number of stations, also the spatial coverage is significantly different between these three datasets. An analysis of the different precipitation products is presented in, for example, Lorenz and Kunstmann [2012].

Evapotranspiration

In contrast to the precipitation data (which are all based on similar observations), evapotranspiration data are derived from quite different modeling and data merging approaches. A comprehensive overview over different global evapotranspiration datasets can be found in Mueller et al. [2011] or Mueller et al. [2013]. Here, we use three evapotranspiration-only products, namely the Global Land-surface Evaporation: the Amsterdam Methodology [GLEAM; Miralles et al., 2011b], the Fluxnet Multi-Tree Ensemble [FLUXNET MTE; Jung et al., 2009, 2010], and data from the Moderate-resolution Imaging Spectroradiometer [MOD16; Mu et al., 2007, 2011].

Besides these products, we also use modeled evapotranspiration from version 2 of the Global Land Data Assimilation System [GLDAS; Rodell et al., 2004b] (i.e., the Noah3.3 realization) and the ERA-Interim Reanalysis from the European Centre for Medium-Range Weather Forecasts [Berrisford et al., 2009, Dee et al., 2011, Simmons et al., 2006].

GLEAM is based on multi-satellite observations and combines a wide range of remotely sensed observations within a Priestley and Taylor-based framework [Miralles et al., 2011b]. In contrast to this multi-satellite-approach, MOD16 is based on remote sensing data from the MODIS satellites and global meteorological data. The datasets are combined using the Penman-Monteith method [Mu et al., 2007]. Finally, the MTE dataset provides global evapotranspiration by empirical upscaling of eddy-covariance measurements from the FLUXNET network [Jung et al., 2009].

Water Storage Changes

GRACE Data

In this study, GRACE observations of the temporal gravity field changes are applied for deriving basin-scale water storage changes. In particular, 10 years (2003 to 2012) of GRACE release 5 data from the two data processing centers GeoForschungsZentrum

(GFZ) Potsdam [Dahle et al., 2013] and the Center for Space Research (CSR), The University of Texas at Austin [Bettadpur, 2012] is used. The spherical harmonic coefficients are filtered with a regularization type filter, as described in Sneeuw et al. [2014]. Then, the spectral data are transformed to a 0.5° × 0.5° grid using the equations from Wahr et al. [1998] and then aggregated over catchments to derive catchment-specific time series. As these time series represent water storage *anomalies*, the data has to be numerically differentiated in order to arrive at the required flux quantity, water storage *changes*. This is done by applying the method of three-point central differences. Further details on the pre-processing of GRACE data can be found in Sneeuw et al. [2014] and Lorenz et al. [2014].

Land Surface Models

As GRACE data is only available from 2002 on, we also need some legacy data for deriving correlations between water storage changes from different catchments (see section *Derivation of the Prediction Equation*). Therefore, we use data from GLDAS. Additionally, we also use data from the land-only replay of the Modern-Era Retrospective analysis for Research and Applications [MERRA Land; Reichle et al., 2011] and the WaterGAP Global Hydrology Model [WGHM; Döll et al., 2003, 2014]. In contrast to the two land surface models MERRA Land and GLDAS, which do not contain a routing scheme (R. Reichle (2013) and M. Rodell (2012), personal communication), WGHM is tuned against observed discharge at 724 gauging stations [Döll et al., 2003]. It is assumed that this improves the reliability of the water availability estimates from the model. Data from WGHM has been used for various water budget studies [e.g., Döll et al., 2014, Forootan et al., 2012, Kusche et al., 2009, Ramillien et al., 2006, Schmidt et al., 2006].

TABLE 5.2: River basins which were used in this study including the respective river gauge, its location, the area of the basin, and the mean annual discharge between the period 1980 to 2010 as provided from GRDC. The location of the basins is shown in Figure 5.1b. From the 29 catchments, the first 16 are used for validating the runoff predictions.

	River	Station	Lat.	Lon.	Area [km^2]	Discharge [m^3/s]
1	Amazon	Obidos	-1.95	-55.51	4672876	171915
2	Congo	Kinshasa	-4.3	15.3	3615546	38652
3	Mississippi	Vicksburg	32.32	-90.91	2938538	18661
4	Ob	Salekhard	66.57	66.53	2926321	12939
5	Yenisei	Igarka	67.48	86.5	2454961	19388
6	Lena	Kyusyr	70.7	127.65	2417932	17761
7	Mackenzie	Arctic Red River	67.46	-133.75	1666073	9194
8	Volga	Volgograd Power Plant	48.81	44.59	1345070	8143
9	St. Lawrence	Lasalle	45.42	-73.62	943769	8575
10	Orange	Vioolsdrif	-28.76	17.72	828475	169
11	Don	Razdorskaya	47.54	40.65	378180	685
12	Pechora	Oksino	67.6	52.2	304670	4857
13	Fraser	Hope	49.38	-121.45	228874	2657
14	Neva	Novosaratovka	59.84	30.53	225651	2490
15	Olenek	7.5 km d/s of mouth of Pur	72.12	123.22	199723	1257
16	Fitzroy (west Australia)	The Gap	-23.09	150.11	126986	138
17	Niger	Lokoja	7.8	6.76	2100508	5070
18	Danube	Ceatal Izmail	45.22	28.72	771277	6490
19	Tocantins	Tucurui	-3.76	-49.65	752993	10966
20	Rio Xingu	Altamira	-3.22	-52.21	445289	8062
21	Kolyma	Kolymskaya	68.73	158.72	421802	3262
22	Severnaya Dvina (northern)	Ust-Pinega	64.15	41.92	330709	3392
23	Churchill	Above Red Head Rapids	58.12	-94.63	299391	337
24	Rio Parnaiba	Luzilandia	-3.45	-42.37	297049	692
25	Yana	Ubileynaya	70.75	136.08	220949	1127
26	Thelon	below outlet of Schultz Lake	64.78	-97.05	171346	924
27	Rhine	Rees	51.75	6.4	169706	2392
28	Elbe	Neu Darchau	53.23	10.89	134037	701
29	Rio Santiago	El Capomal	21.83	-1.05	126986	161

Data Consistency

For the comparison of the runoff estimates with observations, we have re-mapped all the input fields to the same grid resolution ($0.5^\circ \times 0.5^\circ$) using a first-order conservative interpolation [Jones, 1999]. From these fields, area-weighted averages were computed over the study regions shown in Figure 5.1.

The water storage changes $\dot{M}$, computed from GRACE via the central difference scheme, are an approximation of the *true* derivatives. In the spectral domain they differ by a *sinc* function. In fact, the numerical derivatives from the central difference scheme are not compatible yet to the hydrological signals (P, ET and R in Eq. 5.1). The hydrological water storage changes should be filtered using a filter that resembles the *sinc* function in the frequency domain. Swenson and Wahr [2006] proposed to smooth the time series of the other water cycle variables according to

$$\widetilde{F}_t = \frac{1}{4}F_{t-1} + \frac{1}{2}F_t + \frac{1}{4}F_{t+1}\,, \tag{5.2}$$

where F are time series of precipitation, evapotranspiration, and observed runoff. The indices $t-1$, t, and $t+1$ refer to the previous, current, and following month. We apply this filter to all the time series of precipitation, runoff and evapotranspiration.

Methods

Overview

For estimating basin-scale water cycle variables, different formulations of an Ensemble Kalman Filter [EnKF; Evensen, 1994, 2003, Houtekamer and Mitchell, 1998] framework are applied, which successively predict and correct the state vector containing precipitation, evapotranspiration, runoff, and water storage changes

$$\mathbf{X}_t = \left[\mathbf{P}_t^{\mathsf{T}} \quad \mathbf{ET}_t^{\mathsf{T}} \quad \mathbf{R}_t^{\mathsf{T}} \quad \dot{\mathbf{M}}_t^{\mathsf{T}}\right]^{\mathsf{T}} \tag{5.3}$$

with t being the time-index, $\mathbf{P}_t$, $\mathbf{ET}_t$, $\mathbf{R}_t$, and $\dot{\mathbf{M}}_t$ four sub-vectors with dimensions $[N \times 1]$ and N the number of catchments. A summary of the different input parameter and formulas for the Ensemble Kalman Filter (EnKF), the hard (CEnKF_h) and soft (CEnKF_s) constrained Ensemble Kalman Filter, the Ensemble Kalman Smoother (EnKS), and the hard (CEnKS_h) and soft (CEnKS_s) constrained Ensemble Kalman Smoother are presented in Tables 5.3 and 5.4, respectively.

It can be construed that all the filter equations are linear. With the assumption that the errors of the variables are normally distributed, a single-state Kalman filter (SSKF) is sufficient to achieve the results. Instead we present an ensemble Kalman filter (EnKF) approach, which will allow for its extension to assimilate grid-point values on the global scale. Predicting and estimating grid-point values leads to huge (spatial) covariance matrices, which might cause computational issues, for example, during inversion. Therefore, we propose and proceed with the EnKF framework.

Derivation of the Prediction Equation

On monthly time scales, it can be assumed that the basin-scale water cycle variables follow a (more or less) distinct annual cycle. Therefore, the *anomalies* at time-step t are expressed as

$$\mathbf{r}_t = \mathbf{X}_t - \widetilde{\mathbf{X}}_t \,, \tag{5.4}$$

where $\mathbf{X}_t$ is the state vector from (5.3). $\widetilde{\mathbf{X}}_t$ represents the long-term mean annual cycle for each of the four water cycle variables for every basin through

$$\widetilde{\mathbf{X}}_m = \frac{1}{T/12} \sum_{i}^{T/12} \mathbf{X}_{i,m} \,, \tag{5.5}$$

where the indices i and m refer to different years and months, respectively. It is further assumed that the prediction from time-step $t-1$ to t is expressed through a stochastic process of first order

$$\mathbf{r}_t = \mathbf{A}\mathbf{r}_{t-1} + \mathbf{e}_t \,, \tag{5.6}$$

where $\mathbf{A}$ is a (yet unknown) prediction function and $\mathbf{e}_t$ some zero-mean white prediction noise with covariance matrix $D\{\mathbf{e}\} = E\left\{\mathbf{e}\mathbf{e}^{\mathrm{T}}\right\} = \mathbf{Q}_{\mathrm{P}}$ and with the auto- and cross-covariance matrices

$$\boldsymbol{\Sigma} = D\{\mathbf{r}_t\} \quad \text{and} \quad \boldsymbol{\Sigma}_{\Delta} = D\{\mathbf{r}_t, \mathbf{r}_{t-1}\} \,. \tag{5.7}$$

The prediction matrix $\mathbf{A}$ is given through the so-called *least squares prediction* or *multiple ordinary least squares regression*, which is described in, for example, Moritz [1980] or Kurtenbach et al. [2012].

As the *true* process dynamics $\mathbf{A}$ are not known, we are looking for an approximation $\widehat{\mathbf{A}}$. This approximation should minimize the prediction error

$$\mathbf{e} = \mathbf{r}_t - \widehat{\mathbf{r}}_t. \tag{5.8}$$

where $\widehat{\mathbf{r}}_t$ is an estimate of the true state. According to (5.8), we therefore compute the outer product of the error vectors

$$\begin{aligned}\mathbf{e}\mathbf{e}^\mathsf{T} &= (\mathbf{r}_t - \widehat{\mathbf{r}}_t)(\mathbf{r}_t - \widehat{\mathbf{r}}_t)^\mathsf{T} \\ &= (\mathbf{r}_t - \mathbf{A}\mathbf{r}_{t-1})(\mathbf{r}_t - \mathbf{A}\mathbf{r}_{t-1})^\mathsf{T} \\ &= \mathbf{r}_t\mathbf{r}_t^\mathsf{T} - \mathbf{r}_t\mathbf{r}_{t-1}^\mathsf{T}\mathbf{A}^\mathsf{T} - \mathbf{A}\mathbf{r}_{t-1}\mathbf{r}_t^\mathsf{T} + \mathbf{A}\mathbf{r}_{t-1}\mathbf{r}_{t-1}^\mathsf{T}\mathbf{A}^\mathsf{T}\,. \end{aligned} \tag{5.9}$$

The corresponding error covariance matrix is given through

$$D\{\mathbf{e}\} = E\left\{\mathbf{e}\mathbf{e}^\mathsf{T}\right\} = \mathbf{\Sigma} - \mathbf{\Sigma}_\Delta\mathbf{A}^\mathsf{T} - \mathbf{A}\mathbf{\Sigma}_\Delta^\mathsf{T} + \mathbf{A}\mathbf{\Sigma}\mathbf{A}^\mathsf{T}\,. \tag{5.10}$$

An estimate for $\mathbf{A}$ is found by minimizing the trace of the error covariance matrix, i.e.

$$\widehat{\mathbf{A}} = \arg\min \operatorname{Tr}\left[D\{\mathbf{e}\}\right], \tag{5.11}$$

which is the case for

$$\widehat{\mathbf{A}} = \mathbf{\Sigma}_\Delta\mathbf{\Sigma}^{-1}\,. \tag{5.12}$$

With this estimator, the final prediction equation from time-step $t-1$ to t reads as

$$\widehat{\mathbf{r}}_t = \widehat{\mathbf{A}}\mathbf{r}_{t-1}\,. \tag{5.13}$$

For a more detailed derivation, see, for example, Kurtenbach et al. [2012]. As it is assumed that this prediction matrix remains constant over time, the time-index has been omitted. Inserting (5.12) into (5.10) yields

$$D\{\mathbf{e}\} = \mathbf{\Sigma} - \mathbf{\Sigma}_\Delta\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_\Delta = \mathbf{Q}_\mathrm{P}\,, \tag{5.14}$$

which is the error covariance matrix of the predictor $\widehat{\mathbf{A}}$ or, in the common Kalman notation, the covariance matrix of the prediction noise. As the *true* auto- and cross-covariance between anomalies of precipitation, evapotranspiration, runoff, and water storage changes are also unknown, they are approximated by the empirical sample covariance matrices

$$\widehat{\boldsymbol{\Sigma}} = \frac{1}{T-1}\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t^{\mathsf{T}}\,, \text{and} \tag{5.15}$$

$$\widehat{\boldsymbol{\Sigma}}_{\Delta} = \frac{1}{T-2}\sum_{t=2}^{T}\mathbf{r}_t\mathbf{r}_{t-1}^{\mathsf{T}}\,. \tag{5.16}$$

The two matrices are shown in Figures 5.3a,b.

Putting (5.4) into (5.6) and replacing the true dynamics with the approximated prediction matrix $\widehat{\mathbf{A}}$, the process model reads as [Tourian, 2013]

$$\begin{aligned}
\left(\mathbf{X}_t - \widetilde{\mathbf{X}}_t\right) &= \widehat{\mathbf{A}}\left(\mathbf{X}_{t-1} - \tilde{\mathbf{X}}_{t-1}\right) + \mathbf{e}_t \\
&= \widehat{\mathbf{A}}\mathbf{x}_{t-1} - \widehat{\mathbf{A}}\widetilde{\mathbf{X}}_{t-1} + \mathbf{e}_t \\
\Rightarrow \mathbf{X}_t &= \widehat{\mathbf{A}}\mathbf{X}_{t-1} - \widehat{\mathbf{A}}\widetilde{\mathbf{X}}_{t-1} + \widetilde{\mathbf{X}}_t + \mathbf{e}_t\,,
\end{aligned} \tag{5.17}$$

which can be re-written as

$$\mathbf{X}_t = \widehat{\mathbf{A}}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{U}_{t-1} + \mathbf{e}_t \tag{5.18}$$

with

$$\mathbf{B} = \begin{bmatrix} -\widehat{\mathbf{A}} & \mathbf{I} \end{bmatrix} \quad \text{and} \quad \mathbf{U}_t = \begin{bmatrix} \widetilde{\mathbf{X}}_{t-1} \\ \widetilde{\mathbf{X}}_t \end{bmatrix} \tag{5.19}$$

and $\mathbf{I}$ being the identity matrix. Thus, the prediction is based on the covariances between the state variables (through the least squares prediction matrix) and the mean annual cycle through the control input $\mathbf{B}_{t-1}\mathbf{U}_{t-1}$.

The control input is computed from the collection of datasets, which were described in section *Data*. For each water cycle variable, the annual cycle is given through the *mean* of all respective data sources. As we are mainly interested in the performance of the methodology as a predictor, only past data from 1980 to 2002 is used for computing the mean annual cycle.

Derivation of the Observation Equation

The first observation group relates the state vector to observed precipitation, evapotranspiration, runoff, and water storage changes. Thus, the framework allows the assimilation of an arbitrary number of observation data of water cycle variables. These variables can be further used as constraints, if they are assumed to be highly precise.

For assimilating runoff observations from satellite altimetry $\mathbf{R}_{t,\mathrm{alt}}$, another observation group is added to the observation vector. The runoff estimates in the state vector are directly related to the altimetry observations, which then gives the full observation equation through

$$\underbrace{\begin{bmatrix} \mathbf{P}_{t,\mathrm{obs}} \\ \mathbf{ET}_{t,\mathrm{obs}} \\ \mathbf{R}_{t,\mathrm{obs}} \\ \dot{\mathbf{M}}_{t,\mathrm{obs}} \\ \mathbf{R}_{t,\mathrm{alt}} \end{bmatrix}}_{\mathbf{Y}_t} = \underbrace{\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \end{bmatrix}}_{\mathbf{H}_t} \underbrace{\begin{bmatrix} \mathbf{P}_t \\ \mathbf{ET}_t \\ \mathbf{R}_t \\ \dot{\mathbf{M}}_t \end{bmatrix}}_{\mathbf{X}_t} + \boldsymbol{\nu}_t \tag{5.20}$$

or, as in the common Kalman literature,

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \boldsymbol{\nu}_t \,, \tag{5.21}$$

where $\mathbf{I}$ and $\mathbf{0}$ are $[N \times N]$ identity and zero matrices, respectively, and $\boldsymbol{\nu}_t$ is the zero-mean white observation noise with covariance matrix $D\{\boldsymbol{\nu}\} = \mathbf{Q}_\mathrm{O}$. The computation of this covariance matrix is discussed in section *A Multi-Data Approach for Estimating Basin-Scale Errors*. If the observed time series contain missing values, the observation relation matrix has to be adjusted accordingly. Therefore, the time-index t is used here to indicate a *dynamic* observation relation matrix, which changes depending on the number of available observations.

Similar to the control input in section *Derivation of the Prediction Equation*, the observations are computed from the collection of input data sources shown in Table 5.1. It is common practice in hydrology to use an ensemble mean instead of single models or datasets [e.g., Kirtman et al., 2013]. Nevertheless, high-quality observations over, for example, a single catchment (if available) could be used here as well. This will be further analyzed in future studies.

Closure of Catchment-Scale Water Budgets

Water budget closure between the four water cycle variables can be achieved by adding state constraints to the assimilation scheme. These constraints adjust the estimates of precipitation, evapotranspiration, runoff, and water-storage changes in a way that the catchment-scale water budgets are closed. There are several methods for adding such constraints to the assimilation scheme [e.g., Simon, 2010, Simon and Chia, 2002]. Here, we discuss and apply two straightforward approaches for adding water budget constraints.

Both methods are based on the augmentation of the observation vector by adding some *pseudo observations* of the water balance closure. The terrestrial water balance equation (5.1) can be written as a state constraint through

$$\begin{bmatrix} \mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{P}_t \\ \mathbf{ET}_t \\ \mathbf{R}_t \\ \dot{\mathbf{M}}_t \end{bmatrix} = \mathbf{0}, \tag{5.22}$$

where $\mathbf{0}$ is a $[N \times 1]$ vector which contains only zeros and $\mathbf{I}$ are $[N \times N]$ identity matrices.

When we write these constraints in shorter matrix notation

$$\mathbf{G}\mathbf{X}_t = \mathbf{0} \tag{5.23}$$

with $\mathbf{G} = \begin{bmatrix} \mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I} \end{bmatrix}$, we can augment the observation equation (5.20) through

$$\begin{bmatrix} \mathbf{Y}_t \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_t \\ \mathbf{G} \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} \boldsymbol{\nu}_t \\ \boldsymbol{\omega}_t \end{bmatrix}, \tag{5.24}$$

where $\boldsymbol{\omega}_t$ is some zero-mean white noise with covariance $\mathbf{Q}_{\mathrm{WB}}$ for the pseudo observations of the water budget closure.

If the water budget should be perfectly closed, the method is usually referred to as *perfect observations* or *hard constraints*. This can be easily achieved by assuming the pseudo observations to be noise free—that is, $\mathbf{Q}_{\mathrm{WB}} = \mathbf{0}$. However, this assumption can lead to numerical issues, as the augmented observation error covariance matrix is singular. Furthermore, by considering the large differences between the data sources of precipitation, evapotranspiration, and water storage changes, such a constraint might be too strict.

Therefore, one could allow some small imbalances. Such constraints are thus called *imperfect observations* or *soft constraints*. In this case, we add some small non-zero observation errors—that is, $\boldsymbol{\omega}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{\mathrm{WB}})$. In order to be consistent with the basin-specific characteristics, this error is chosen to be 10 % of the mean annual cycle of runoff [Fekete et al., 2012].

A Multi-Data Approach for Estimating Basin-Scale Errors

As there are plenty of data sources, it is difficult to identify the "best" datasets for each of the water cycle variables. It was shown in Lorenz et al. [2014] that there is no such dataset that performs best globally. Moreover, a proper description of the observation errors is essential in order to successfully apply any Kalman Filtering approach. Therefore, an ensemble-based approach has been chosen, where numerous realizations of observations are estimated, based on the range of the input data sources.

First, the *range* of the collection of input datasets for the water cycle variable F (which is either precipitation, evapotranspiration, or water storage changes) is computed for every catchment n and every time-step t. This range is assumed to be a proxy for the *uncertainty* of each of the four water cycle variables. Therefore, all precipitation anomalies from K different datasets are stored in a vector:

$$\mathbf{f}_{n,t} = \begin{bmatrix} f_{n,t,1} & f_{n,t,2} & \cdots & f_{n,t,K} \end{bmatrix}^{\mathrm{T}}, \tag{5.25}$$

where the $f_{n,t,k}$ are the anomalies with respect to the mean annual cycle $\widetilde{F}_t$. From this vector, the range is given through

$$\Delta f_{n,t} = \frac{1}{2}\left[\max\left(\mathbf{f}_{n,t}\right) - \min\left(\mathbf{f}_{n,t}\right)\right]. \tag{5.26}$$

The sample covariance and correlation matrices are then estimated through

$$\boldsymbol{\Sigma}_{\Delta f} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\boldsymbol{\Delta}\mathbf{f}_t - \overline{\boldsymbol{\Delta}\mathbf{f}}\right)\left(\boldsymbol{\Delta}\mathbf{f}_t - \overline{\boldsymbol{\Delta}\mathbf{f}}\right)^{\mathrm{T}} \tag{5.27}$$

$$\boldsymbol{\rho}_{\Delta f} = \left(\boldsymbol{\Sigma}_{\Delta f}^{(\mathrm{diag})}\right)^{-\frac{1}{2}} \boldsymbol{\Sigma}_{\Delta f} \left(\boldsymbol{\Sigma}_{\Delta f}^{(\mathrm{diag})}\right)^{-\frac{1}{2}}, \tag{5.28}$$

where $\boldsymbol{\Sigma}_{\Delta f}$ (diag) is the diagonal variance matrix of $\boldsymbol{\Sigma}_{\Delta f}$. These correlation matrices thus describe the relationship between the *uncertainties* of the water cycle anomalies from different catchments. In order to use as much data as possible, the matrices are computed from past data only (i.e., until 2002). Then, the correlation matrices are re-scaled with the ensemble standard deviation from actual data (i.e., from 2003). Although multiple datasets for each of the hydrological variables are used, there is not enough data for estimating reliable ensemble standard deviations for *every* time-step. Therefore, a *cyclostationary* approach has been chosen:

$$\sigma_{F,n,m} = \sqrt{\frac{1}{K(T/12)-1}\sum_{k=1}^{K}\sum_{i=1}^{T/12}\left(F_{n,k,i,m} - \widetilde{F}_{n,k,m}\right)^2}, \tag{5.29}$$

where the index i represents a single year and m a single calendar month (e.g., January, February, etc.). K is the number of datasets for a single variable and $\widetilde{F}_{n,k,m}$ the long-term mean annual cycle of the kth dataset over catchment n for month m. The standard deviations for the other calendar months are computed likewise. The error covariance matrix is then given by

$$\mathbf{\Sigma}_{F,m} = \left(\mathbf{I}\,\boldsymbol{\sigma}_{F,m}\right)\boldsymbol{\rho}_{\Delta f}\left(\mathbf{I}\,\boldsymbol{\sigma}_{F,m}\right), \tag{5.30}$$

where $\mathbf{I}$ is the $N \times N$ unit matrix and

$$\boldsymbol{\sigma}_{F,m} = \begin{bmatrix} \sigma_{F,1,m} & \sigma_{F,2,m} & \cdots & \sigma_{F,N,m} \end{bmatrix}. \tag{5.31}$$

The matrices for each water cycle variable are then combined in the full observation error covariance matrix

$$\mathbf{Q}_{\mathrm{O},t} = \begin{bmatrix} \mathbf{\Sigma}_{\mathrm{P},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathrm{ET},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{obs}},t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\dot{\mathrm{M}},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{alt}},t} \end{bmatrix}, \tag{5.32}$$

where $\mathbf{\Sigma}_{\mathrm{R_{obs}}}$ and $\mathbf{\Sigma}_{\mathrm{R_{alt}}}$ are the covariance matrices for the runoff observations from GRDC and satellite altimetry, respectively. The mean monthly index m has been replaced with the time-index t in order to account for the monthly errors in the runoff observations from GRDC and satellite altimetry. According to, for example, Fekete et al. [2012], traditional discharge measurements usually have an accuracy of 5–10 %. We thus compute the errors of the runoff observations from GRDC to be 5 % of the reported values. The errors from satellite altimetry are obtained by propagating the errors of the altimetric water level measurements through the rating curve model, as described in Tourian et al. [2013].

Obviously, correlations between the uncertainties of different water cycle variables are neglected. As the data comes from totally different sensors, we think that this is a reasonable assumption. However, it should be further investigated how, for example, a fully populated observation error covariance matrix affects the performance of the assimilation framework.

Figure 5.2 shows histograms of random perturbations derived from the covariance matrix in (5.32). These random values are used for computing an ensemble of perturbed observations within the EnKF. Here, we only show some exemplary realizations for the three catchments Amazon, Danube, and Yana for January and July. It is obvious that the chosen methodology is able to produce different perturbations for different catchments, variables, and months. For example, the precipitation perturbations in January over the Amazon basin show a wide spread as only about 73 % are within the range between -50 and 50 mm/month. On the contrary, in July, almost 100 % are within the same interval. Over the Yana basin, almost 100 % of the precipitation perturbations are between -10 and 10 mm/month. The method is thus able to provide perturbations which depend on the climatic conditions, but also on the spread of the different datasets over the basins. However, it should be mentioned that a limited number of rainfall gauges in a basin can lead to very similar precipitation estimates from different products. This suggests a good agreement between these data sources, even if the true precipitation can be totally different.

Computation of the Least Squares Prediction Matrix

According to equations (5.12), (5.14), (5.12), and (5.15) the least squares prediction matrix $\widehat{\mathbf{A}}$ and its error covariance matrix $\mathbf{Q}_\mathrm{P}$ are computed from the auto- and cross-covariance matrices between anomalies of precipitation, evapotranspiration, runoff, and water storage changes. In order to capture reliable covariance structures, a collection of different data sources (which are listed in Table 5.1) has been used instead of singe datasets. As it is avoided to use the same data twice, the covariance matrices are derived from a different period (1980–2002) than the estimated period (2003–2010). The two matrices $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_\Delta$ are shown in Figures 5.3a and b.

It is emphasized that these matrices are derived from the anomalies, which consist of random errors, climatic variability and extreme events. They thus represent

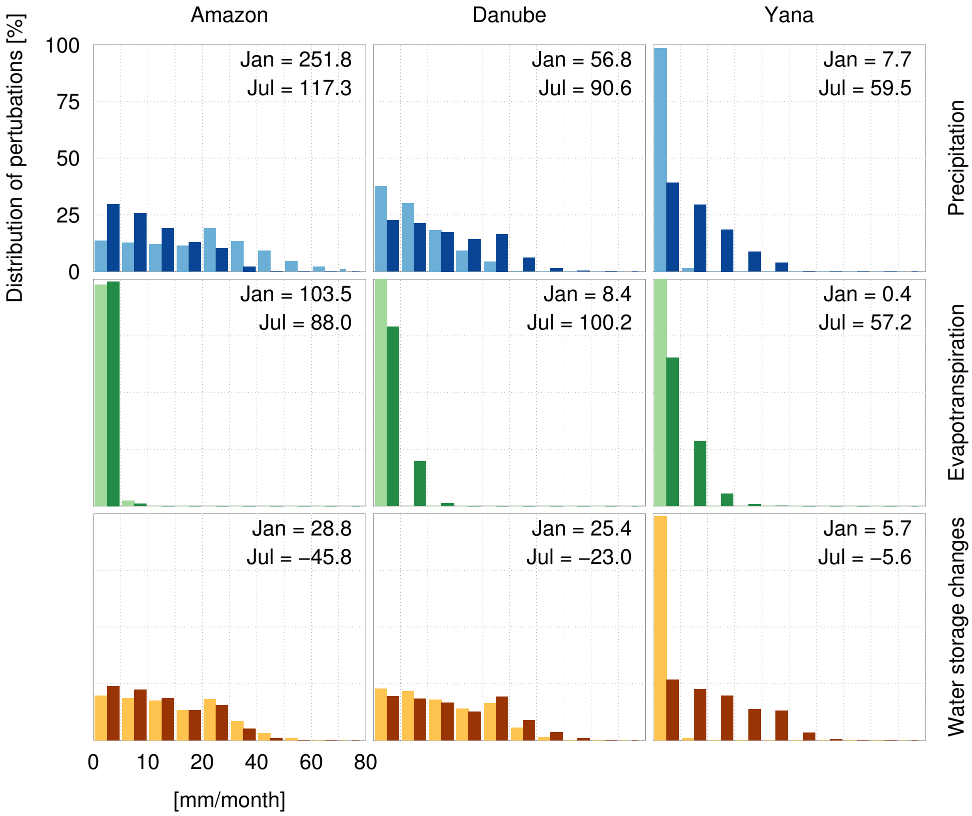

FIGURE 5.2: Histograms of 10000 random realizations drawn from a multivariate normal distribution with zero mean and the observation error covariance matrix from (5.32). The histograms show the distributions of the random perturbations for the three catchments Amazon (left column), Danube (middle column), and Yana (right column) for precipitation (top row), evapotranspiration (center row), and water storage changes (bottom row). The left bars (bright colors) correspond to the perturbations of January while the right bars (dark colors) show the perturbations in July. The figure only shows the absolute value of the perturbations as they are symmetric around zero. The numbers in the top right corner of every histogram are the ensemble mean values of precipitation, evapotranspiration, and water storage changes for January and July, respectively, in mm/month.

covariances mainly due to features which deviate significantly from the annual cycle. Both the auto- and cross-covariance matrix show high positive inter-catchment covariances throughout most basins especially for precipitation, evapotranspiration, and water storage changes. These characteristics have been already addressed in Riegger et al. [2012], where it is concluded that inter-catchment correlations can be expected, for example, for neighboring catchments due to local climatic conditions.

The off-diagonal elements in $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_{\Delta}$ for runoff are comparatively small. This indicates that there are no such strong inter-catchment correlations (or covariances) for runoff. Thus, at the anomaly level, climatic and physical variations in time series of runoff are rather local phenomena.

The inter-variable covariances are generally lower. Some strong relationship between precipitation, water storage changes, and, to a certain extent, runoff can be detected. This is indicated by relatively high covariances on the main-diagonals of the corresponding matrix blocks. At the anomaly level, this is intuitive as, for example, an exceptionally strong precipitation event might lead to high runoff and an increased water storage. The second case can be clearly identified by the high covariances on the main diagonals of the upper right and lower left sub-matrices, respectively. In terms of precipitation and runoff, such a strong relationship can only be observed over some catchments. As an example, the gauges of the four basins Amazon, Tocantins, Rio Xingu, and Rio Parnaiba are all located at the North-Eastern part of South America. Over these catchments, there are high inter-catchment and inter-variable covariances between precipitation, runoff, and water storage changes.

On the contrary, evapotranspiration generally does not show such high covariances with any of the other water cycle variables. This indicates that there is no strong coupling between the residuals of evapotranspiration and other water cycle variables. The reason for this is the stable periodic characteristic of the evapotranspiration time series. Compared to the other water cycle variables, the total signal power (i.e., the root mean square RMS) of the evapotranspiration anomalies are very low and the dynamics possess a random behavior (not shown here).

When comparing the auto- and cross-covariance-matrices in Figures 5.3a and b, they both show similar structures, even if the covariances in $\mathbf{\Sigma}_{\Delta}$ are generally lower. The top right sub-matrix in Figure 5.3b shows, as hydrologically expected, a strong cross-covariance between precipitation and water storage changes with a time-lag of one month. On the other hand, the bottom left sub-matrix does not show such high covariances. Thus, after a strong precipitation event, the impact on the water storage can be seen even one month later. There is no such strong coupling between, for example, anomalies of precipitation and runoff, even if that could have been expected. This suggests that, based on the empirical covariances, significant deviations from

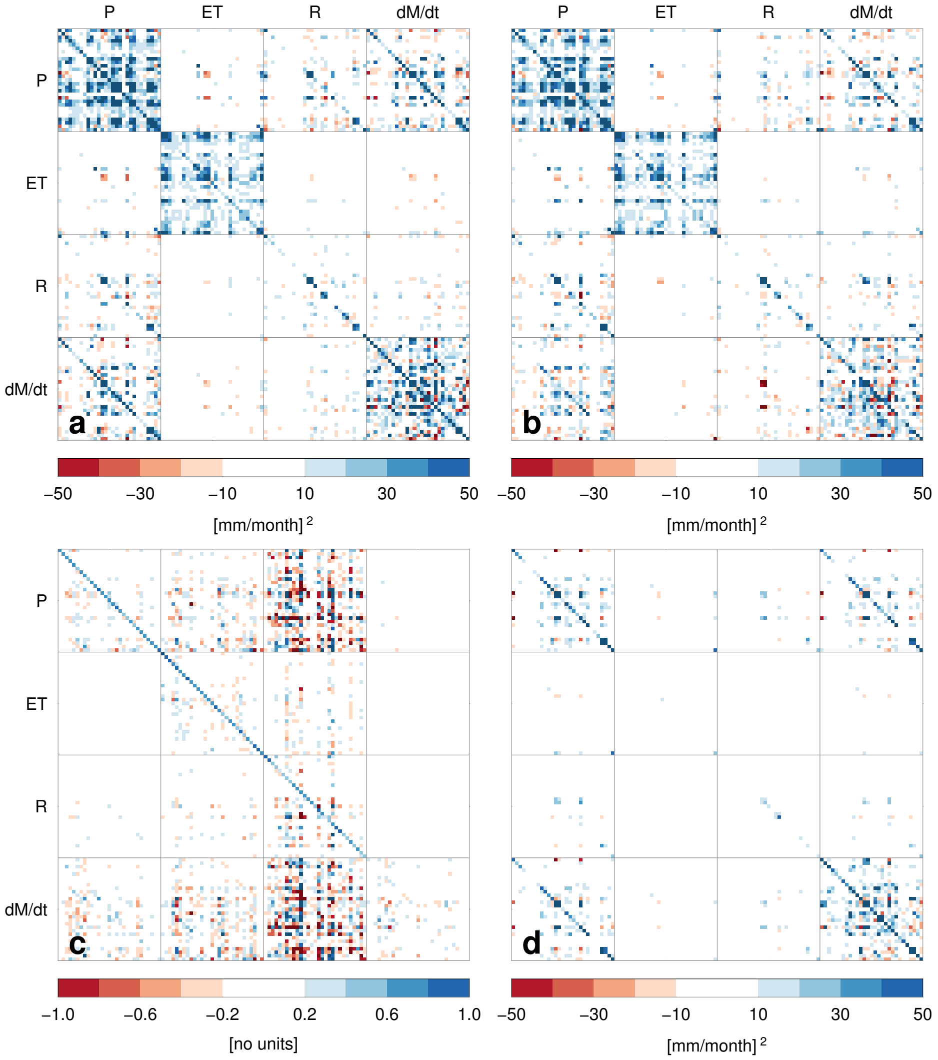

FIGURE 5.3: Temporal (a) auto-covariance and (b) cross-covariance matrix [$\mathbf{\Sigma}$ and $\mathbf{\Sigma}_\Delta$, see equations (5.15) and (5.16)] between anomalies of precipitation, evapotranspiration, runoff, and water storage changes. The anomalies are computed with respect to the ensemble mean annual cycle from all available data sources. The matrices thus represent the overall mean covariance structure of the water cycle variables. Both matrices consist of 16 sub-matrices, of which each has the dimension [29 × 29]. These sub-matrices therefore represent the auto- and (temporal) cross-covariance between the different study regions and water cycle variables. (c) Least squares prediction matrix $\widehat{\mathbf{A}}$ and (d) the corresponding error matrix $\mathbf{Q}_\text{P}$.

the mean annual cycle of precipitation do not necessarily lead to exceptional runoff months.

The least squares prediction matrix and its corresponding error covariance matrix can be derived from the auto- and cross-covariance matrices according to equations (5.12) and (5.14). The two matrices $\widehat{\mathbf{A}}$ and $\mathbf{Q}_{\mathrm{P}}$ are shown in Figures 5.3c and d.

Instead of just using fully populated covariance matrices, we can distinguish between three cases:

1. Covariances between catchments and variables (case 1)
2. Covariances between catchments (case 2)
3. Covariances between variables (case 3)

In the first case, it is assumed that there are covariances between both the catchments and the water cycle variables. In case two and three, either the covariances between the water cycle variables (case 2) or the catchments (case 3) are neglected. In Figures 5.3c and d, only the matrices for case 1 are shown. The other two cases are derived from these matrices by removing all sub-matrices except for those on the main diagonal (case 2) or all elements except for those on the main diagonals of each sub-matrix (case 3). It will be discussed in section *Structure of the Least Squares Prediction Matrix* if these three cases lead to different predictions. Besides that, initial tests confirmed that using equation (5.13) with different configurations of the least squares prediction matrices can explain about 70 % of the runoff anomaly variance (not shown). However, this must not hold true for the other water cycle variables, which might be less predictable through such a covariance-based least squares approach.

Performance Metrics

For validating the estimated time series of precipitation, evapotranspiration, runoff, and water storage changes, numerous performance metrics can be evaluated. In

TABLE 5.3: Overview of the different parameters and variables for the assimilation framework.

Parameter	Equation	Dimension
State vector	$\mathbf{X}_t = \begin{bmatrix}\mathbf{P}_t^{\mathsf{T}} & \mathbf{ET}_t^{\mathsf{T}} & \mathbf{R}_t^{\mathsf{T}} & \dot{\mathbf{M}}_t^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$	$4N \times 1$
Annual cycle	$\widetilde{\mathbf{X}}_m = \frac{1}{T/12}\sum_{i,m}^{T/12}\mathbf{X}_{i,m}$	$4N \times 1$
Residuals	$\mathbf{r}_t = \mathbf{X}_t - \widetilde{\mathbf{X}}_t$	$4N \times 1$
Auto-covariance	$\mathbf{\Sigma} = \frac{1}{T-1}\sum_{t=1}^{T}\mathbf{r}_t\mathbf{r}_t^{\mathsf{T}}$	$4N \times 4N$
Cross-covariance	$\mathbf{\Sigma}_{\Delta} = \frac{1}{T-2}\sum_{t=2}^{T}\mathbf{r}_t\mathbf{r}_{t-1}^{\mathsf{T}}$	$4N \times 4N$
Prediction matrix	$\widehat{\mathbf{A}} = \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}^{-1}$	$4N \times 4N$
Prediction noise	$\mathbf{Q}_{\mathrm{P}} = \mathbf{\Sigma} - \mathbf{\Sigma}_{\Delta}\mathbf{\Sigma}^{-1}\mathbf{\Sigma}_{\Delta}^{\mathsf{T}}$	$4N \times 4N$
Control input	$\mathbf{B} = \begin{bmatrix}-\widehat{\mathbf{A}} & \mathbf{I}\end{bmatrix}$	$4N \times 8N$
	$\mathbf{U}_t = \begin{bmatrix}\widetilde{\mathbf{X}}_{t-1} & \widetilde{\mathbf{X}}_t\end{bmatrix}^{\mathsf{T}}$	$8N \times 1$
Observation vector	$\mathbf{Y}_t = \begin{bmatrix}\mathbf{P}_{t,\mathrm{obs}}^{\mathsf{T}} & \mathbf{ET}_{t,\mathrm{obs}}^{\mathsf{T}} & \mathbf{R}_{t,\mathrm{obs}}^{\mathsf{T}} & \dot{\mathbf{M}}_{t,\mathbf{obs}}^{\mathsf{T}} & \mathbf{R}_{t,\mathrm{alt}}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$	$5N \times 1$
Observation covariance	$\mathbf{Q}_{\mathrm{O},t} = \begin{bmatrix}\mathbf{\Sigma}_{\mathrm{P},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_{\mathrm{ET},t} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{obs}},t} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\dot{\mathrm{M}},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{\Sigma}_{\mathrm{R_{alt}},t}\end{bmatrix}$	$5N \times 5N$
Observation relation matrix	$\mathbf{H}_t = \begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0}\end{bmatrix}$	$5N \times 4N$
Constraints	$\mathbf{G}\mathbf{X}_t = \mathbf{0}$	
	$\mathbf{G} = \begin{bmatrix}\mathbf{I} & -\mathbf{I} & -\mathbf{I} & -\mathbf{I}\end{bmatrix}$	$N \times 4N$
Augmented obs. rel. matrix	$\mathbf{H}_{t,\mathrm{aug}} = \begin{bmatrix}\mathbf{H}_t^{\mathsf{T}} & \mathbf{G}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$	$6N \times 4N$
Augmented observation vector	$\mathbf{Y}_{t,\mathrm{aug}} = \begin{bmatrix}\mathbf{Y}_t^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}}\end{bmatrix}^{\mathsf{T}}$	$6N \times 1$
Augmented observation covariance	$\mathbf{Q}_{\mathrm{O},t,\mathrm{aug}} = \begin{bmatrix}\mathbf{Q}_{\mathrm{O},t} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{\mathrm{WB},t}\end{bmatrix}$	$6N \times 6N$

hydrology, it is common to use correlation, the percentage bias [PBIAS; Gupta et al., 1999]), and the Nash-Sutcliffe Efficiency [NSE; Nash and Sutcliffe, 1970]). From these, correlation describes the level of common information content between two time series. As it is insensitive to the amplitude and the mean value, the PBIAS between an observed (Y_t) and predicted (X_t) time series,

$$\mathrm{PBIAS} = \frac{\sum_{t=1}^{T}(X_t - Y_t)}{\sum_{t=1}^{T} Y_t} = 1 - \frac{\overline{X}}{\overline{Y}}, \tag{5.33}$$

TABLE 5.4: Overview of the assimilation algorithm. For the constrained predictions, the observation vector ($\mathbf{Y}_t$), the observation relation matrix ($\mathbf{H}_t$), and the observation covariance matrix ($\mathbf{Q}_{\mathrm{O},t}$) are replaced with the augmented parameters ($\mathbf{Y}_{t,\mathrm{aug}}$, $\mathbf{H}_{t,\mathrm{aug}}$, $\mathbf{Q}_{\mathrm{O},t,\mathrm{aug}}$). A particular ensemble member is identified by the index d.

Parameter	Equation		
Initial state	$\mathbf{X}^{+}_{0,d}$	$=$	$\mathbf{X}_0 + \mathbf{e}_d$ with $\mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\mathrm{P})$
Prediction step	$\mathbf{X}^{-}_{t,d}$	$=$	$\widehat{\mathbf{A}}\mathbf{X}^{+}_{t-1,d} + \mathbf{B}\mathbf{U}_{t-1} + \mathbf{e}_d$ with $\mathbf{e}_d \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_\mathrm{P})$
Prediction covariance	$\widehat{\mathbf{\Sigma}}^{-}_{X,t}$	$=$	$\frac{1}{D-1}\sum_{d=1}^{D}\left(\mathbf{X}^{-}_{t,d} - \overline{\mathbf{X}}^{-}_{t}\right)\left(\mathbf{X}^{-}_{t,d} - \overline{\mathbf{X}}^{-}_{t}\right)^{\mathsf{T}}$
Observation Innovations	$\mathbf{V}_{t,d}$	$=$	$\mathbf{Y}_t - \mathbf{H}_t\mathbf{X}^{-}_{t,d} + \boldsymbol{\nu}_{t,d}$ with $\boldsymbol{\nu}_{t,n} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{\mathrm{O},t}\right)$
Kalman Gain	$\mathbf{K}_t$	$=$	$\widehat{\mathbf{\Sigma}}^{-}_{X,t}\mathbf{H}_t^{\mathsf{T}}\left(\mathbf{H}_t\widehat{\mathbf{\Sigma}}^{-}_{X,t}\mathbf{H}_t^{\mathsf{T}} + \mathbf{Q}_{\mathrm{O},t}\right)^{-1}$
Correction step	$\mathbf{X}^{+}_{t,d}$	$=$	$\mathbf{X}^{-}_{t,d} + \mathbf{K}_t\mathbf{V}_{t,d}$
Corr. covariance	$\widehat{\mathbf{\Sigma}}^{+}_{X,t}$	$=$	$\frac{1}{D-1}\sum_{d=1}^{D}\left(\mathbf{X}^{+}_{t,d} - \overline{\mathbf{X}}^{+}_{t}\right)\left(\mathbf{X}^{+}_{t,d} - \overline{\mathbf{X}}^{+}_{t}\right)^{\mathsf{T}}$
Kalman Gain (smoother)	$\mathbf{K}_t$	$=$	$\widehat{\mathbf{\Sigma}}^{+}_{X,t}\widehat{\mathbf{A}}\left(\widehat{\mathbf{\Sigma}}^{-}_{X,t+1}\right)^{-1}$
Smoothing step	$\mathbf{X}_{t,d}$	$=$	$\mathbf{X}^{+}_{t,d} + \mathbf{K}_t\left(\mathbf{X}_{t+1,d} - \mathbf{X}^{-}_{t+1,d}\right)$
Smoothed covariance	$\widehat{\mathbf{\Sigma}}_{X,t}$	$=$	$\widehat{\mathbf{\Sigma}}^{+}_{X,t} + \mathbf{K}_t\left(\widehat{\mathbf{\Sigma}}_{X,t+1} - \widehat{\mathbf{\Sigma}}^{-}_{X,t+1}\right)\mathbf{K}_t^{\mathsf{T}}$

gives precise information about the relative difference between the long-term means. Finally, both metrics are summarized in the $\mathrm{NSE}_{\mathrm{mean}}$-coefficient

$$\mathrm{NSE}_{\mathrm{mean}} = 1 - \frac{\sum_{t=1}^{\mathsf{T}} (X_t - Y_t)^2}{\sum_{t=1}^{T} \left(Y_t - \overline{Y}\right)^2}, \tag{5.34}$$

which is highly sensitive to the agreement in phase, amplitude, and mean between two datasets. The $\mathrm{NSE}_{\mathrm{mean}}$ further represents the normalized mean squared error between, for example, an observed and predicted time series [Lorenz et al., 2014].

In this study, an alternative formulation of the NSE-coefficient, which takes the long-term annual cycle of a variable into account, is proposed:

$$\mathrm{NSE}_{\mathrm{cycle}} = 1 - \frac{\sum_{t=1}^{T} (X_t - Y_t)^2}{\sum_{t=1}^{T} \left(Y_t - \widetilde{Y}\right)^2}, \tag{5.35}$$

where $\widetilde{Y}$ is not the long-term mean, but the long-term mean annual cycle. Similar to the common Nash-Sutcliffe Efficiency, (5.35) gives information if the estimated time series X_t is a better predictor than the long-term annual cycle from historic data $\widetilde{Y}$ ($\mathrm{NSE}_{\mathrm{cycle}} > 0$). If $\mathrm{NSE}_{\mathrm{cycle}} < 0$, the estimated time series can not improve the long-term mean annual cycle with respect to the observations. This indicates that the estimated residuals (with respect to the mean annual cycle) then do not capture the short- and long-term climatic variability in the observations.

Results and Discussion

In this section, the performance of the approach presented here will be discussed. First, different set-ups of the assimilation framework are analyzed in order to find an appropriate configuration. Therefore, several tests are run with both the unconstrained and hard Constrained Ensemble Kalman Filter (hereafter EnKF and $\mathrm{CEnKF}_\mathrm{h}$). For these tests, the framework estimates runoff over 16 of the 29 catchments, where continuous runoff observations during the period 2005 to 2010 are available (see Figure 5.1 and Table 5.2). The runoff observations from the GRDC for these 16 catchments are removed from 2005—that is, they are not assimilated. Thus, all following performance metrics are derived from the comparison between observed runoff and the estimates from the proposed data assimilation framework during the period 2005 to 2010.

Sensitivity Analysis

Ensemble Size

In an Ensemble Kalman Filtering framework, the *true* covariance matrices are approximated by empirical sample covariance matrices. Therefore, it has to be investigated if and how different ensemble sizes have an impact on the estimates. Studies like, for example, Mitchell et al. [2002] provide a comprehensive analysis of this issue.

Here, estimated and observed runoff are compared with respect to the ensemble size. Figure 5.4a thus shows the $\mathrm{NSE}_{\mathrm{mean}}$ between observed and estimated runoff for the first five catchments out of the 16 study regions, namely the Amazon, Congo, Mississippi, Ob, and Yenisei basins.

From these catchments, there are catchments with (Amazon and Ob) and without (Congo, Mississippi, and Yenisei) runoff observations from altimetry. Furthermore, these catchments cover the main climatic regimes—that is, the tropics (Amazon and Congo), mid- (Mississippi) and high-latitudes (Ob and Yenisei). This allows a basic analysis with respect to different runoff characteristics.

The figure clearly shows that $\mathrm{NSE}_{\mathrm{mean}}$ remains rather constant throughout all ensemble sizes, especially over the Amazon, Ob, and Yenisei basins. The largest variations can be observed for the Mississippi basin when ensemble sizes smaller than 5000 are used. However, these variations are negligible compared to the magnitude of the performance metric. The impact of the ensemble size on the unconstrained (dark colors) and constrained (bright colors) are very similar. From this analysis we have chosen an ensemble size of 10000 as Figure 5.4a clearly shows that using larger ensembles does not lead to improved results.

Perturbation Analysis

As any Ensemble Kalman Filtering approach involves random drawings of perturbations, the estimates from repeated runs of the EnKF can show a certain spread. In order to analyze this uncertainty within the proposed framework, the EnKF and $\mathrm{CEnKF}_{\mathrm{h}}$ are run with a fixed ensemble size of 10000 for 500 times. Figure 5.4b shows the distribution of the Nash-Sutcliffe efficiency between observed and estimated runoff for the 16 study regions.

Foremost, the NSE values show a very small spread, even if there are large differences between the catchments. For most catchments, the interquartile range (IQR) of the NSE is below 0.04. Larger values can be observed for the Orange or Don basins, where the IQR can reach values up to 0.06. Runoff over these catchments usually does not follow a typical annual cycle and has a lower overall magnitude. Therefore, even small changes in the predicted time series can result in significantly different performance metrics. But compared to the inter-catchment variability of NSE values, this spread appears negligible.

Comparing the performance metrics from the unconstrained (EnKF) and the constrained (CEnKF_h) estimates clearly shows that there is no significant difference in terms of spread. Thus, even if the water budgets are forced to be closed, the estimated runoff time series from multiple runs of the CEnKF_h show a (negligible) variability.

Structure of the Least Squares Prediction Matrix

It is investigated how the different covariance structures in the least squares prediction matrix influence the results. This analysis thus helps to identify if the estimates benefit from fully populated covariance matrices (covariances between both catchments and water cycle variables, case 1) or if covariances between catchments only (case 2) or water cycle variables only (case 3) provide better results.

Figure 5.4c shows the NSE between observed runoff and the estimates from the EnKF and CEnKF_h over 16 catchments for different configurations of the prediction matrix. Most catchments show positive values for all three cases, while the highest values are clearly achieved by case 3. In addition, using an unconstrained (darker dots) or constrained (brighter dots) does not lead to significantly different metrics. The worst performance can be observed for the Orange, Don, and Fitzroy basins, where case 1 and 2 lead to negative NSE values. The reasons for this are again the low overall magnitudes and irregular dynamics of the runoff time series over these catchments.

The best performance can be achieved when correlations only between the water cycle variables are assumed (case 3). This can be understood from case 2, where the runoff estimates for a specific catchment depend only on runoff from other basins. From a climatic point of view, this would make sense as for example, the tropical

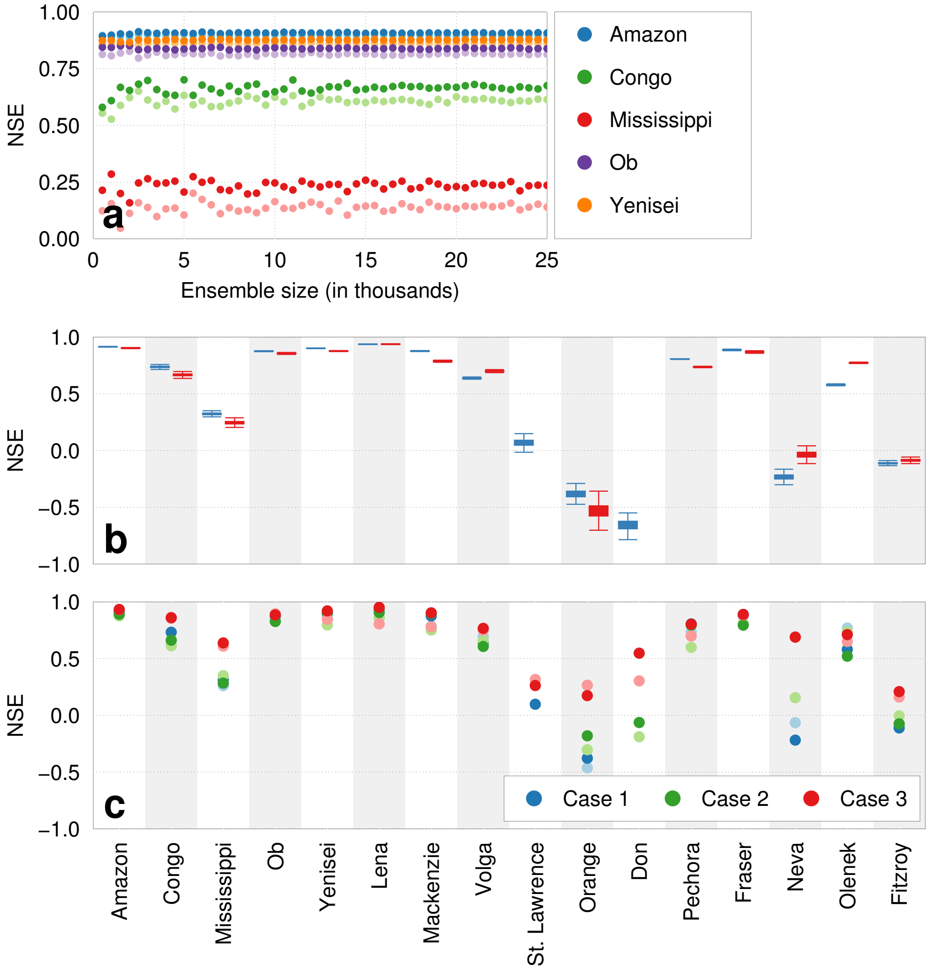

FIGURE 5.4: Sensitivity analysis showing the NSE values between observed and predicted runoff. (a) NSE derived from multiple runs of the filter using different ensemble sizes (the darker and brighter colors are derived from the EnKF and $\mathrm{CEnKF_h}$, respectively). (b) Box-and-whisker plots from 500 runs of the EnKF (blue) and the $\mathrm{CEnKF_h}$ (red) with the same settings and an ensemble size of 10000. The boxes represent the 25 % and 75 % percentile of the distribution of performance metrics. The top and bottom whiskers cover approximately 99.3 % of the data, if a normal distribution is assumed. (c) NSE from the EnKF (dark colors) and the $\mathrm{CEnKF_h}$ (bright colors) using a fully populated prediction matrix (case 1, blue), and prediction matrices which neglect correlations between water cycle variables (case 2, green) or catchments (case 3, red).

catchments often show similar runoff characteristics. However, this relationship is obviously not strong enough for providing reasonable estimates.

On the other hand, it is noted that the use of fully populated prediction matrix (case 1) does not lead to better results. For instance, the NSE values for case 1 do not always lie between those from case 2 and 3. The reason for this behavior is not yet fully understood, and therefore, needs further investigation. Henceforth, we will use covariance structures as in case 3 for the proposed methods.

Prediction Performance

After the identification of a suitable set-up, this section discusses the overall prediction performance of the proposed method. The filter is run with an ensemble size of 10000 and case 3 of the prediction matrix in six different configurations (see section *Overview*). The performance metrics of these different configurations with respect to observed monthly runoff are summarized in Figure 5.7. The corresponding time series for the 16 study regions are shown in Figures 5.5 and 5.6.

In terms of correlation, an overall good agreement between the dynamics of estimated and observed runoff can be identified. Over most catchments (except for Orange and Fitzroy), the different configurations led to correlations larger than 0.6. The catchments with high amplitudes and a dominant annual cycle in the runoff time series (Amazon, Congo, Mississippi, Ob, Yenisei, Lena, Mackenzie, Volga, Pechora, Fraser, Neva, and Olenek) show correlations of at least 0.8. This good agreement can be explained by the repeating annual cycle of runoff, which is reproduced well by all six set-ups (see Figures 5.5 and 5.6). The main reason for this is the strong control input, which forces the runoff estimates to follow the mean annual cycle. However, several catchments show significant climatic variations, which are revealed by the anomaly time series in the right columns of Figures 5.5 and 5.6. E.g. over the Amazon basin, even if there is a dominant annual cycle, the anomalies still show some significant dynamics, which are reproduced quite well by the estimates. Also over Congo, Mississippi, Ob, Volga, Fraser, Neva, and Olenek, the estimated

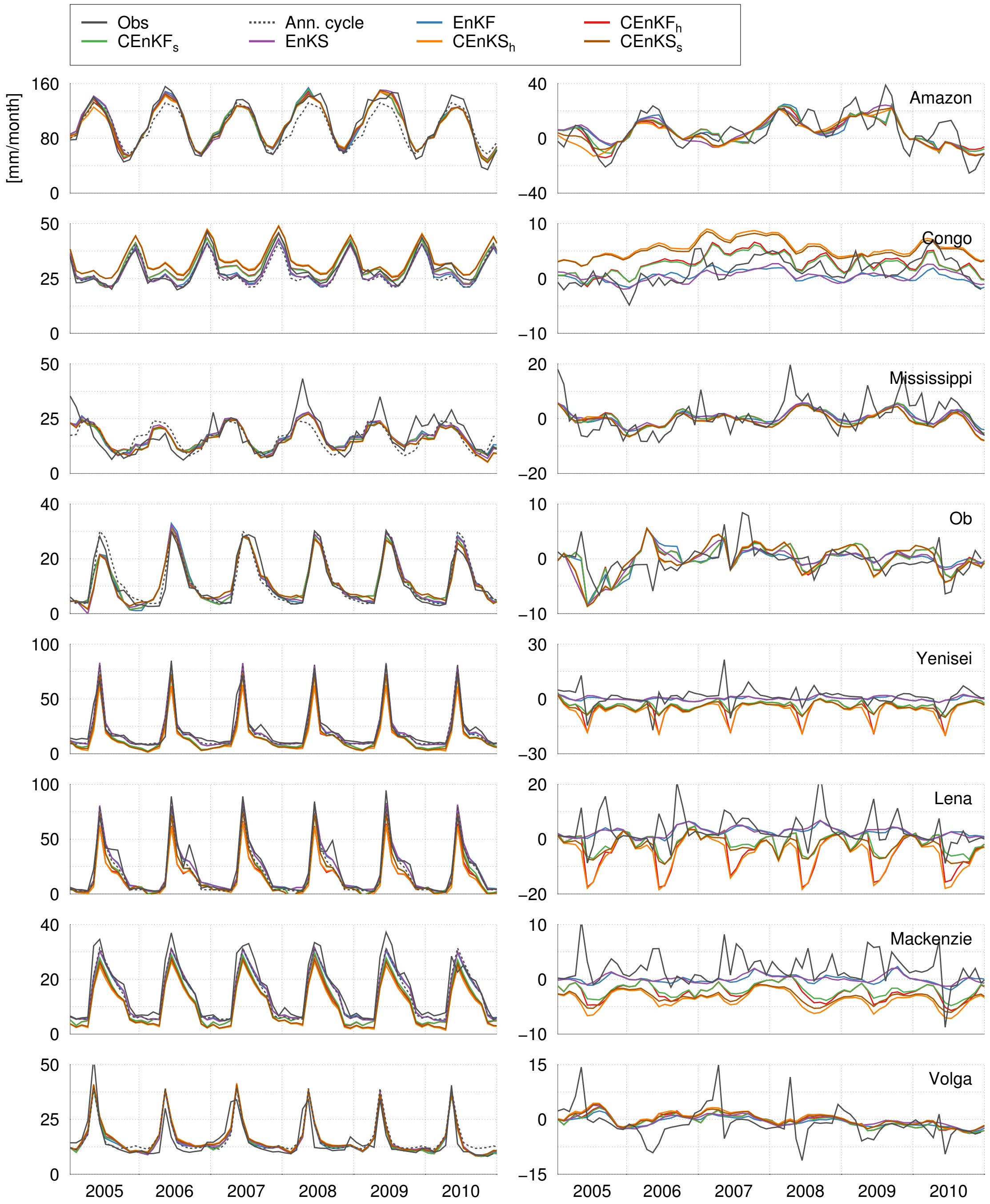

FIGURE 5.5: Time series of observed (black) and predicted runoff from different configurations of the Kalman Filter and Smoother for the study regions 1–8. The left column shows the full runoff time series while for the right column, the mean annual cycle from historic runoff data has been removed.

runoff anomalies show significant long-term variations, which agree well with the observations.

For Lena, Pechora, and Olenek, the runoff observations show a significant second peak during the late summer months, that can not be captured by the estimates.

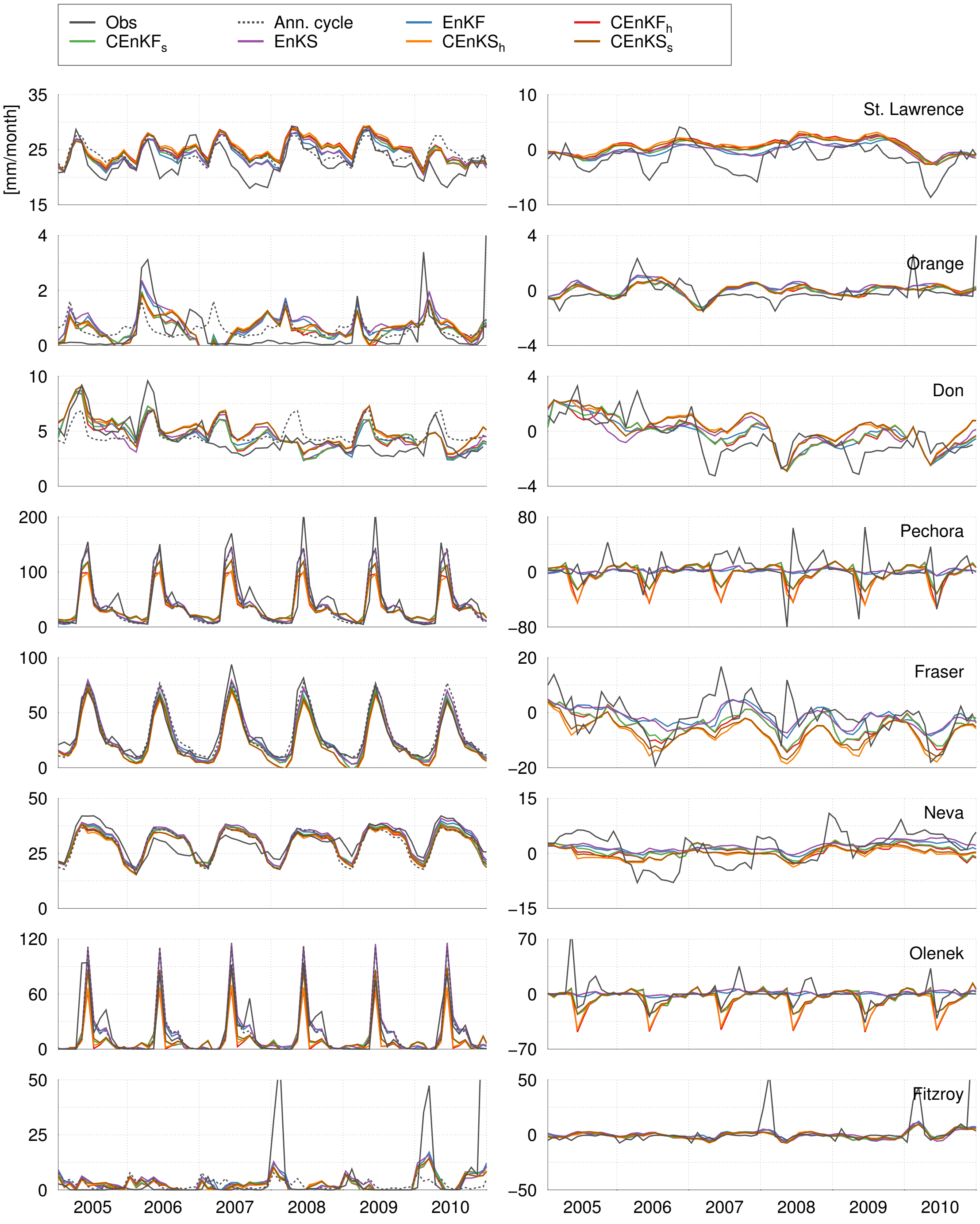

FIGURE 5.6: Time series of observed (black) and predicted runoff from different configurations of the Kalman Filter and Smoother for the study regions 9–16. The left column shows the full runoff time series while for the right column, the mean annual cycle from historic runoff data has been removed.

Over the Mackenzie basin, the estimates can only capture the annual cycle but none of the exceptional events. Additionally, the sudden drop in the constrained anomalies during 2008, for example, can not be found in the observations.

It must be noted that the anomalies for many of the catchments with a dominant annual cycle are rather small, which complicates the estimation of exceptional events. This also explains why the correlations over these basins, despite the disagreement of the anomalies, are still very high.

Over catchments with lower overall magnitudes and irregular time series dynamics (St. Lawrence, Orange, Don, Fitzroy), several configurations achieve correlations of 0.5 and more. Thus, the runoff estimates are in principle able to capture a large part of the climatic variability in the observations. But when looking at the time series for these four basins (Figures 5.5 and 5.6), the estimates do not agree well with the observations. Over Don, both the estimates and the observations show a peak in runoff during the summer months during the years 2005 and 2006. In 2007 and 2009, there are no such peaks in the observations, but still in the estimated runoff. Nevertheless, on the anomaly level (right columns in Figures 5.5 and 5.6), observed runoff shows a slight decreasing trend during the period 2005 until 2010 which is also roughly reproduced by the estimates. That being said, if runoff over such basins is estimated from a water balance approach, it is mandatory that all input variables are highly consistent in both their magnitudes and dynamics. Otherwise, due to the biases in the estimates of the other water cycle variables, some artificial short-term variations are introduced, which finally result in both low correlation values and high relative biases.

Overall, the estimates are not able to reproduce extreme events in the runoff time series (e.g., Mississippi in 2008). This can be explained by several reasons. First of all, runoff is predicted using a least squares prediction, which exploits covariance information. In fact, this information reflects only the averaged statistical relationship between the variables and basins. The statistical relationships of extreme events are thus not well represented. Moreover, the filtering of the input data, as described in section *Data Consistency*, significantly dampens the peaks in the time series. An extreme event could disappear simply due to the smoother input time series. On the other hand, the predictions show a significant increase during heavy flooding events over the Fitzroy river in January 2008, February/March 2010 and December 2010/January 2011 (source: `bom.gov.au`). These were induced by heavy rain events, which can be also detected in the input precipitation datasets (not shown). The presented approach can therefore capture such extreme events, if these also occur in, for example, the precipitation time series. This, however, must not always be the case: persistent rainfall during the end of the previous and the beginning of the next month can lead to a flooding event, even if (due to the monthly resolution) there are no exceptional features in the precipitation time series.

When analyzing the performance of the different configurations, there is hardly any difference between the unconstrained, constrained, filtered, and smoothed estimates. The largest spread in terms of correlation can be observed over the Don basin (CEnKF_h and CEnKF_s: 0.72, CEnKS_h and CEnKS_s: 0.60), where the correlations of the filtered are significantly higher than the smoothed predictions. This holds also for the unconstrained and constrained estimates. Other significant differences can only be observed over the Fitzroy basin (EnKF: 0.54, CEnKS_h and CEnKS_s: 0.48). For most catchments, however, the different configurations have only little impact on the dynamics of the estimated time series.

Despite a slight underestimation of runoff over most catchments, the PBIAS values in Figure 5.7 show a different picture as there are significant differences between the metrics from the six set-ups. Even if there are many catchments where most configurations provide PBIAS values $< \pm 10\,\%$ (Amazon, Congo, Mississippi, Ob, Volga, St. Lawrence, Don, Pechora, Fraser, and Neva), it is obvious that the constrained estimates show generally larger biases. This can be particularly observed over the high-latitude catchments Yenisei, Lena, Mackenzie, Don, Pechora, Fraser, Olenek, and Fitzroy, where the PBIAS values of the unconstrained and constrained estimates differ by more than 10 %. An explanation for this can be a high uncertainty of the runoff estimates. When using such constraints within an assimilation framework, each variable in the state vector receives a correction, so that the water budgets are closed. This correction is based on the relative uncertainty of the respective variable.

Thus, uncertain variables receive stronger corrections. The runoff estimates over the catchments showing large differences between the unconstrained and constrained predictions can therefore be assumed to be more uncertain than, for example, the predictions over the Amazon, Ob, or Volga basins. As a consequence, forcing the water budgets to be perfectly closed seems to be too strict. This assumption is supported by the better performance of the soft constrained estimates over most basins.

Over some of the catchments, however, the anomalies of the unconstrained estimates are small or almost zero throughout the whole period (Yenisei and Mackenzie)—that is, they are based mainly on the control input. When compared with observed runoff, the peaks in the unconstrained estimates are sometimes over- and sometimes underestimated. Thus, the errors cancel out, which leads to an overall small PBIAS-value.

In contrast to most other catchments, the estimates over the Congo, Mackenzie, and Fraser basins show some distinct biases. For Mackenzie and Fraser, the constrained estimates significantly underestimate runoff over nearly the whole study period. The estimates from the CEnKS_h and CEnKS_s predict the lowest runoff over both basins. In these cases, constraining and smoothing results in a significant bias with respect to the observations. The same holds true for the Congo basin, where especially the estimates from CEnKS_h and CEnKS_s overestimate runoff.

In the hydrological community, NSE_mean values > 0.5 usually indicate an acceptable performance of runoff estimates [e.g., Moriasi et al., 2007, Santhi et al., 2001]. This is the case for 13 of the 16 estimated catchments. Several catchments even show NSE_mean values > 0.8 (Amazon, Congo, Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser). These are mainly basins with large amplitudes and a rather stable annual cycle in the runoff time series. For such catchments, it is thus suggested to analyze the NSE_cycle metric (bottom row of Figure 5.7). This metric obviously shows significantly smaller values than the NSE_mean. Still, there are positive values over 14 of the 16 study regions. This inherently means that the estimates over these 14 catchments agree clearly better with the runoff observations than the values of the mean annual cycle.

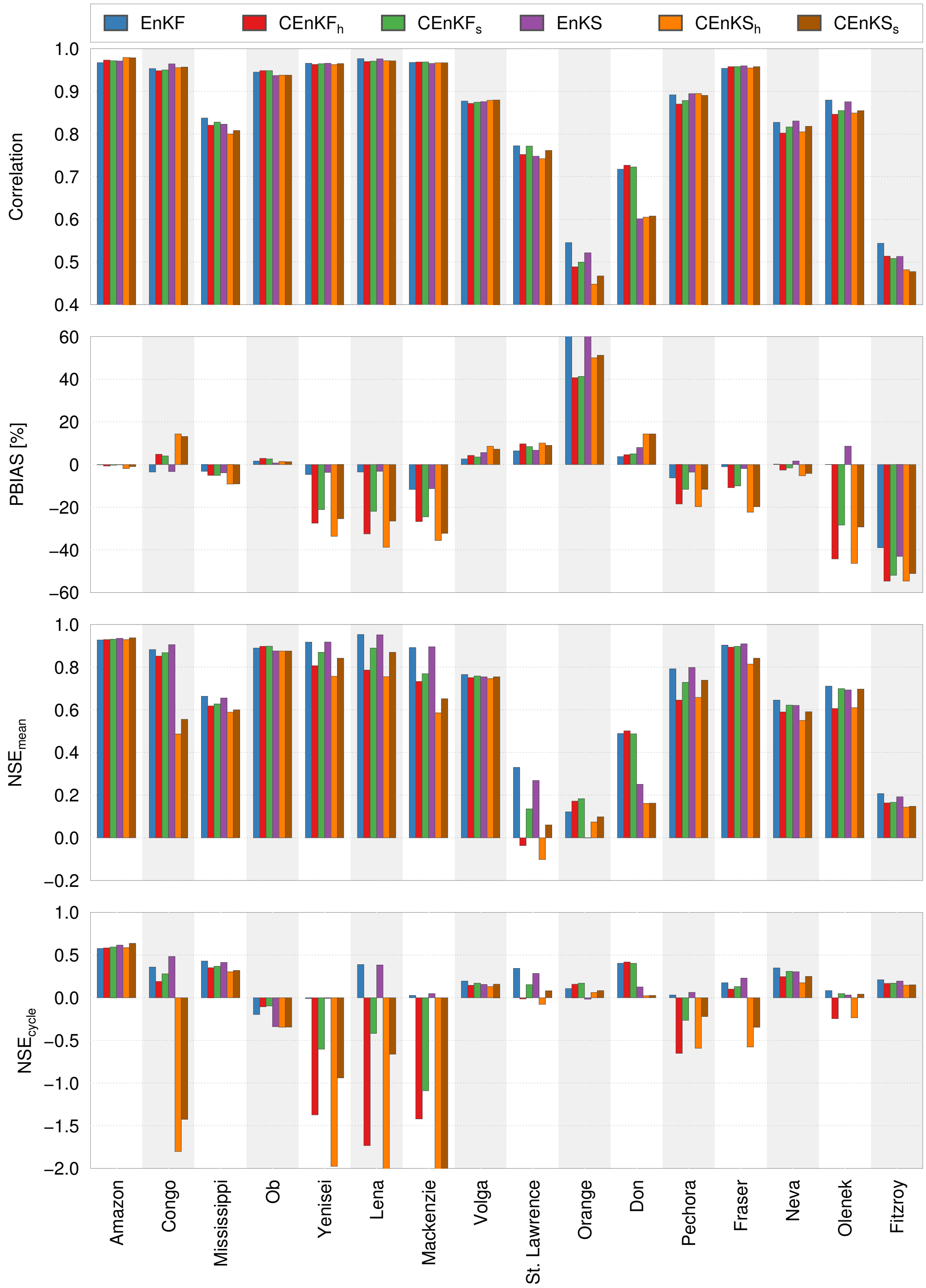

FIGURE 5.7: Performance metrics between observed and predicted runoff over all 16 study regions. The statistics are based on the period between 2005 and 2010. The colors indicate different configurations of the assimilation framework.

It must be mentioned that over catchments with a less dominant annual cycle (St. Lawrence, Orange, Don, or Fitzroy) the annual cycle from historic data is obviously a bad predictor. Thus, positive $\text{NSE}_{\text{cycle}}$ values can be expected even if the estimates show only good agreement during a limited period or for distinct peaks. In such cases, using the NSE_{mean} gives more information about the performance of an estimated time series.

Significant improvements in terms of positive $\text{NSE}_{\text{cycle}}$ values can be found over Amazon, Congo, Mississippi, Lena (EnKF and EnKS), Volga, St. Lawrence, Don, Neva, and Fitzroy. Even more than the classical NSE_{mean} metric, the $\text{NSE}_{\text{cycle}}$ is very sensitive with respect to the agreement of extreme events and long-term climatic variations. From the anomaly time series in Figures 5.5 and 5.6 it is obvious that the predictions successfully reproduce the long-term variations in the residual runoff observations over Amazon, Congo, Mississippi, Lena (EnKF and EnKS), Don, and Neva.

It becomes evident that even with high NSE_{mean} values over Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser, several configurations provide worse estimates than the mean annual cycle from historic data. Thus, the variability on top of the annual cycle, which is induced by the assimilation, does not improve the estimates. Ob seems to be a special case, as the anomalies of the estimates and the observations show similar long-term variations. In the end, the large differences, especially during 2005 and 2006, between the estimates and observations seem to result in a negative $\text{NSE}_{\text{cycle}}$-value.

In the case of Yenisei, Lena, Pechora, Fraser, and Olenek, constraining the estimates dampens the maximum of runoff during the summer months. This can be identified by the negative peaks in the anomaly time series in Figures 5.5 and 5.6, where the hard constrained estimates from CEnKF_{h} and CEnKF_{s} assume the lowest runoff. Over the Olenek basin, using soft constraints actually leads to a better agreement, even if the secondary peaks during the late summer months are still missing. This is not the case for the Yenisei, Lena, Pechora, and Fraser basins, where the unconstrained estimates provide better results. Especially for the Lena basin, the long-term variability in the anomaly estimates from the EnKF and EnKS agrees well with the observations, even if the peaks in the anomalies are higher. A closer look reveals similar features over the Yenisei, Pechora, and Fraser basin. Over all these basins, the soft constrained estimates are able to capture some of the negative peaks in the residual time series, but also show several unrealistic variations. Thus, using such constraints can lead to a better representation of short-term features, but the unconstrained estimates are able to capture the climatic long-term variations.

Interestingly, over all high-latitude catchments with a distinct snow accumulation phase (i.e., where runoff is close to zero), most configurations are able to reproduce the low runoff during the winter months. It can be assumed that these phases do not change drastically from year to year as the observations show only little deviations from the long-term annual cycle. Thus, during these phases, the estimates are mainly based on the control input (i.e., the mean annual cycle), which explains the good agreement with the observations.

Comparison to Similar Studies

In Troy et al. [2011], the water budgets of several basins over Northern Eurasia have been estimated using multiple data sources. They further estimate basin-scale runoff from a weighted water-balance approach, which is roughly comparable to our method. During their validation period between 1984 and 1999, they achieve $\mathrm{NSE}_{\mathrm{mean}}$ values of 0.94 over the Lena and Yenisei and 0.89 over the Ob basin with respect to observed monthly runoff. Even if our performance metrics are based on a different period, we achieve similar values (Lena: 0.95 from EnKF and EnKS, Yenisei: 0.92 from EnKF and EnKS, and Ob: 0.90 from $\mathrm{CEnKF}_{\mathrm{h}}$).

Recently, Gudmundsson and Seneviratne [2014] published an approach for estimating gridded runoff over Europe. They also present a catchment-scale validation for several European river basins, including the Rhine and Elbe basin. These two basins are not included in our performance analysis due to missing data during the selected study period. However, when estimating runoff using our ensemble based approaches, we obtain $\mathrm{NSE}_{\mathrm{mean}}$ values of 0.66 (Rhine, EnKF) and 0.62 (Elbe, $\mathrm{CEnKF}_{\mathrm{h}}$ and $\mathrm{CEnKF}_{\mathrm{s}}$), which is in good agreement with the findings from Gudmundsson and Seneviratne [2014].

In Kang et al. [2014], the Variable Infiltration Capacity (VIC) model [Liang et al., 1994] model has been applied to the Fraser river basin. During their validation period from 1969 to 2006, the model estimated monthly runoff with respect to the observations from the gauging station at Hope with a NSE_{mean} of 0.85. In our study, the runoff estimates for the Fraser basin during the period 2005 to 2010 show a comparable performance ($\text{NSE}_{\text{mean}} = 0.90$ from EnKF, CEnKF_{s}, and EnKS).

Riegger and Tourian [2014] proposed a methodology, that is based on the characterization of the relationship between runoff and water storage changes. They simulated monthly runoff over five catchments, namely the Mackenzie ($\text{NSE}_{\text{mean}} = 0.93$), Ob (0.95), Lena (0.87), Yukon (0.81), and Yenisei (0.63) basins. Our approach is able to estimate runoff with a similar performance (Mackenzie: 0.90 from EnKF and EnKS, Ob: 0.90 from CEnKF_{h}, Lena: 0.95 from EnKF and EnKS, and Yenisei: 0.92 from EnKF and EnKS).

Conclusion

In this study, a data-assimilation framework is proposed, which can be employed for predicting and correcting catchment-scale time series of runoff. One of the major design parameters of the framework is the use of as much real data as possible instead of using complex model equations. Thus, the approach is based on the terrestrial water budget equation only, which is included in an Ensemble Kalman Filtering framework. The prediction scheme predicts precipitation, evapotranspiration, runoff, and water storage changes using the so called least squares prediction method. By this, we can exploit temporal and spatial covariance structures between different catchments and water cycle variables. For the observation equations of the dynamic process model, the most recent versions of widely used data sources for precipitation, evapotranspiration, runoff, and water storage changes are applied. We further use

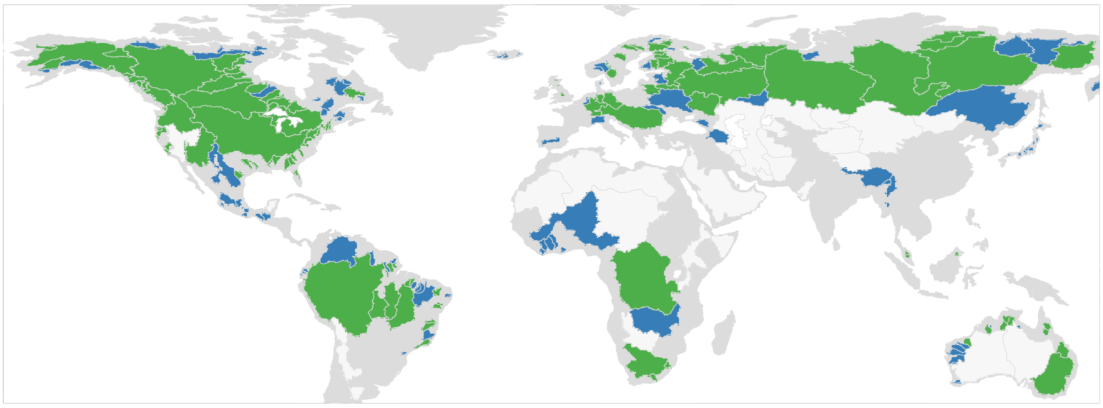

FIGURE 5.8: River basins for which runoff can be predicted using the Ensemble Kalman Filter approach based on global hydrometeorological datasets. For the green areas, more than 5 years of data after 2002 are available. The blue areas are (currently) poorly gauged basins with less than 5 years of data after 2002, but more than 5 years of continuous runoff measurements during the period 1980 to 2002. These catchments cover an area of more than 11,500,000 km^2 and provide a mean annual discharge volume of more than 125,000 $\mathrm{m^3/s}$ of freshwater resources.

estimated runoff from satellite altimetry in order to both fill the gaps in the time series of gauge-based runoff observations but also to improve the predictions.

The filter is first run with different ensemble sizes and three different structures of the prediction matrix for finding a suitable set-up of the framework. The predictions from the filter are then compared with observed runoff from GRDC on a monthly basis. So far, the best results can be achieved by assuming correlations between water cycle variables only (case 3).

In order to enforce water budget closure between the estimated parameters, we further add appropriate water budget constraints to the framework. Therefore, it is distinguished between hard constraints, which assume perfect closure of the water budgets, and soft constraints, which allow a small well defined imbalance. In the end, the framework is run in six different configurations: the Ensemble Kalman Filter, the hard and soft Constrained Ensemble Kalman Filter, the Ensemble Kalman Smoother, and the hard and soft Constrained Ensemble Kalman Smoother.

In order to assess the performance of our data-assimilation approach, runoff is predicted over 16 large river basins and validated against *in situ* data. The performance analysis shows that the proposed method is able to estimate runoff with correlations larger than 0.8 for 12 of the 16 study regions. In terms of the PBIAS, values are less than ±20 % for most of the catchments (except for Orange and Fitzroy). Finally, the NSE_{mean} values are larger than 0.5 for 13 of the 16 catchments. As the runoff time series of several catchments are dominated by a strong annual cycle, we further analyze the $\text{NSE}_{\text{cycle}}$, which relates the performance of the predictions to the long-term mean annual cycle. For 14 of the 16 study basins, at least several configurations achieve $\text{NSE}_{\text{cycle}}$ values larger than 0. The time series clearly show that the reason for this good performance is a promising agreement between the observed and predicted long-term variations in the runoff time series.

The validation further shows that the different aforementioned configurations lead to similar time series and performance metrics for most of the studied catchments. The different configurations thus have only a minor impact on the dynamics of the predicted time series. As a general pattern, the constrained configurations seem to suffer from larger biases. This holds especially true for the Yenisei, Lena, Mackenzie, Pechora, Fraser, and Olenek basins. This results in rather negative $\text{NSE}_{\text{cycle}}$ values. The soft constrained provide slightly better results, which highlights the uncertainty in the data sources of water cycle variables. Forcing an absolute closure of water cycle from uncertain data might thus lead to numerical artifacts.

As a conclusion, the method is able to provide runoff estimates over the catchments shown in Figure 5.8, where only few or even no runoff observations are available after the year 2002. These catchments cover an area of more than 11,500,000 km^2 and provide a discharge volume of more than 125,000 m^3/s of freshwater resources.

The performance analysis emphasizes the promising performance of the proposed method for predicting runoff. There is still room for improvement. Over Ob, Yenisei, Lena, Mackenzie, Pechora, and Fraser, several configurations provide worse predictions than the mean annual cycle from historic data, despite the high NSE_{mean} values. The reason for this has to be analyzed in more detail, as the bad performance of the constrained predictions indicates large inconsistencies in the data sources of the water cycle variables.

Furthermore, the predictions are not able to reliably reproduce extreme events in the runoff time series as it is not appropriate to predict, for example, flooding events from monthly data. Also because the least squares prediction further exploits covariances, which describe the overall mean relationships between water cycle variables and catchments, Future studies must address an appropriate statistical characterization of exceptional events in the time series of the water cycle variables.

Overall, the presented configurations of the data assimilation framework allow, despite their limitations and shortcomings, to both fill data gaps and extend the streamflow time series for basins with discontinued observations. Thereby, the framework is an alternate option for predicting runoff of ungauged and poorly gauged basins.

Acknowledgments

The authors would like to thank the Helmholtz Alliance Remote Sensing and Earth System Dynamics and the priority program 1257 *Mass Transport and Mass Distribution in the System Earth* of the Deutsche Forschungsgemeinschaft DFG for funding this research. We would further like to thank the following data providers: Global Runoff Data Centre GRDC, Global Precipitation Climatology Project GPCP, Global Precipitation Climatology Centre GPCC, University of Delaware UDEL, Numerical Terradynamic Simulation Group at the University of Montana NTSG UMT, Max Planck Institute for Biogeochemistry MPI BGC, European Centre for Medium-Range Weather Forecasts ECMWF, University of Texas at Austin, Center for Space Research CSR, NASA's Goddard Earth Science Data and Information Services Center GES DISC, NOAA's Earth System Research Laboratory Physical Sciences Division ESRL PSD and NCAR's Computational and Information Systems Laboratory Research Data Archive CISL RDA. Further thanks are due to Dr. Diego Miralles (University of Bristol) for kindly providing the GLEAM evapotranspiration data. All graphics in the document were produced with the Generic Mapping Tools [Wessel and Smith, 1991].

Chapter 6

Summary and Synthesis

In this work, various global hydrological and hydrometeorological datasets are used in order to comprehensively analyze the large-scale water budget. This is done over different spatial and temporal scales. First, it is analyzed how well three widely used atmospheric reanalyses agree with gridded observations of precipitation and temperature and if these comprehensive models are able to close the global long-term water budgets [chapter 2, Lorenz and Kunstmann, 2012]. It is further analyzed how well different observation based datasets agree in terms of precipitation and temperature. This, in the end, allows to quantify the uncertainty with which we have to cope when using such data. In the second article [chapter 3, Lorenz et al., 2014], the focus is on the closure of the basin-scale water budgets. It is thus the logical consequence from the global long-term analyses in the first publication. Within this assessment, the closure of the basin-scale water budgets is directly quantified by analyzing the difference between observed runoff and runoff estimates from a terrestrial and atmospheric-terrestrial water balance. The choice of study regions in this article is driven by the availability of runoff observations, which already shows the importance of runoff data for water budget studies. Thus, the third article [chapter 4, Sneeuw et al., 2014] analyzes different hydro-geodetic methods, which which runoff can be estimated. The performance of these methods is then evaluated by comparing the estimates to observed runoff. It is distinguished between global methods, which do not need runoff data, and the catchment-specific methods, which require (at least historic) runoff observations. In the fourth article [chapter 5, Lorenz et al., 2015], it is investigated, if and how the global budget-based methods can be improved in order to achieve a reasonable water budget closure on the catchment scale and, as a consequence, provide realistic runoff estimates. The four articles therefore present a broad overview of the strengths and weaknesses of our current state-of-the-art hydrological and hydrometeorological data sources for the four major water cycle variables. The major findings of each of the articles are summarized in the following.

Chapter II In the article *The Hydrological Cycle in Three State-Of-The-Art Reanalyses: Intercomparison and Performance Analysis*, three widely used atmospheric reanalysis models, namely ERA-Interim, MERRA, and CFSR, are evaluated and compared against gridded observations of precipitation and temperature from GPCC or CRU. It is demonstrated that even on annual and longer time scales, there are significant differences between the spatial precipitation and temperature patterns of the models. In terms of precipitation, the largest differences are detected over the tropics, high-latitude regions, and mountain ranges. Overall, the range of modeled precipitation reaches values of more than 5 mm day^{-1}, which, in the end, adds up to an annual uncertainty of more than $1,500$ mm year^{-1}. The analyzed observation datasets, however, also show differences of more than 3 mm day^{-1} over, for example, the Congo basin. One reason for this is the very limited number of rain gauges which are located in these areas. It is demonstrated that the number of active gauges over different continents are significantly decreasing. Over Africa, for example, the GPCC dataset used more than $3,000$ gauges before the year 1990. In 2005, the number dropped to only about $1,000$ gauges for the whole continent. An inspection of the spatial distribution of rain gauges reveals that there are huge areas without a single observation station. This, obviously, makes it rather difficult to reasonably validate modeled precipitation over such areas. But even over well observed areas like, for example, Europe or North America, significant differences between the models and the observations can be detected. The largest discrepancies occur during the summer months of the respective hemisphere as convective effects remain a significant source of uncertainty. On the other hand, the differences between the gridded precipitation observations and the reanalyses remain rather constant during the whole study period from 1989 to 2006. This is an important finding as changes in the observation system, which is assimilated into the reanalysis models, do not have a significant impact on the precipitation estimates. However, the situation changes for oceanic precipitation, where both MERRA and CFSR show a significant shift during 1998, when sounding radiances from AMSU were assimilated into both reanalyses.

The article further discusses the water budget closure of the three reanalysis models. Therefore, area-averaged values are calculated from reanalysis-based maps of precipitation, evapotranspiration, runoff, moisture flux divergence, and total atmospheric water vapor. These values are then used for evaluating different long-term water budget equations. The analysis shows that especially MERRA and CFSR suffer from significant imbalances in their oceanic-continental water budgets, which are mainly due to unbalanced estimates of oceanic precipitation and evapotranspiration. This is partly caused by changes in the observation system, which is assimilated into the reanalysis models. Even if these effects can not clearly identified over the continents, it can be assumed that such changes might also impact the modeled climate variables over land.

The results indicate that at least these comprehensive atmospheric reanalysis models still suffer from significant biases in their estimates and thus from imbalances in the representation of the water cycle. These biases and the impact of changes in the observation system on the modeled climate variables further deny the use of such models for water budget or climate change studies.

Chapter III The water budget closure on basin scales is discussed in the second article *Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances*. The fundamental approach is the estimation of runoff from the terrestrial and atmospheric-terrestrial water budget equations, respectively. The remaining water cycle variables are given by state-of-the-art and widely used hydrological and hydrometeorological datasets. This allows to estimate runoff from different combinations of gridded observations (GPCC, GPCP, CPC, CRU, DEL), atmospheric reanalysis models (ERA-Interim, MERRA, CFSR), partially-model-based datasets (GLEAM, MOD16, FLUXNET MTE) and GRACE-satellite derived water storage changes. This, in total, gives 90 runoff estimates from the terrestrial and 9 from the atmospheric-terrestrial water budget for each of the 96 study regions.

The comparison of the budget-based methods, as a matter of fact, directly quantifies the imbalance over different catchments. In general, basin-scale long-term water budget closure, which is represented by a small PBIAS value, can be achieved only over a small number of study regions and from a small number of data combinations. It is concluded that the runoff estimates from the terrestrial water budget tend to underestimate (overestimate) runoff over Arctic (tropical) catchments. As the PBIAS shows highly variable results for different catchments and different data combinations, it can be further concluded that each dataset suffers from it's biases and uncertainties. But even if the majority of runoff estimates show large imbalances of more than 25 % of the mean annual runoff, there are certain combinations, which lead to satisfactory water budget closure over most catchments. However, the preferable datasets change from catchment to catchment, which clearly denies the extrapolation of the performance of a method (i.e., data combination) to other catchments.

The level, to which the water budgets can be closed, seems to depend significantly on the dynamics of the water cycle variables. Good results can be achieved for catchments where runoff exhibits a strong annual cycle with a large volume of runoff. In these cases, the biases of the individual data sources for precipitation, evapotranspiration, and water storage changes carry less weight. This holds even true for catchments which are far below the resolution of GRACE. On the other hand, closure over catchments with very low magnitudes in the runoff time series is not yet possible with our current data sources.

Interestingly, the estimates from the precipitation-runoff ration shows by far the best agreement with the runoff observations. This indicates that, despite their magnitude, evapotranspiration and water storage changes follow a stable annual cycle, which remains constant from year to year. But the method can only succeed over catchments where the ratio between precipitation and runoff remains constant over a longer period, which also prohibits the application of the approach to catchments with strong anthropogenic impacts.

As a summary, the paper shows that our current data sources for the major water cycle variables are not mature enough for analyzing the water cycle on the global scale. The differences in their spatial and temporal resolution, but also their respective uncertainties, clearly deny a general global application. Therefore, it is of major importance to overcome the mismatches between datasets from different observation systems or models.

Chapter IV As the choice of the study regions in the second article is mainly dictated by the availability of runoff observations, it is obvious that there is still an urgent need for *in situ* runoff observations if we want to study the catchment-scale water cycle. However, it is well known that the number of streamflow gauges is decreasing and that the annual runoff rate over the unmonitored areas is comparable to that over the monitored areas [Dai and Trenberth, 2002]. The third article thus provides an overview over several alternative methods which can be applied for estimating runoff. The budget-based approaches, the precipitation-runoff ratio, and several land surface models, which were already discussed in the second article, are compared against runoff estimates from satellite-altimetry and an empirical relationship between runoff and water storage changes. The estimates are then evaluated with runoff observations from GRDC over the six catchments Amazon, Danube, Mississippi, Niger, Ob, and Yenisei.

The analysis clearly shows the high performance of the catchment specific approaches from satellite altimetry and the runoff-storage relationship. It can be assumed that especially satellite altimetry will be a highly promising runoff estimation alternative with the advent of a new generation of altimetry satellites. On the other hand, the water budget equations can be evaluated globally which, in principle, allows the estimation of runoff for all catchments. But, as already discussed in the second article, the biases of the individual data sources for the water cycle variables lead to imbalances in the water budgets, which explains the weak performance of these estimates especially for catchments with low magnitudes in their runoff time series.

Nevertheless, it is shown that the approaches under discussion are indeed promising alternatives to, for example, classical hydrological modeling approaches for estimating runoff. Even if each method still has its shortcomings, it can be assumed that such runoff estimates will improve with new and better data sources for the different water cycle variables and new satellite altimetry missions.

Chapter V As already mentioned, the mismatch between the current data sources for precipitation, evapotranspiration, and water storage changes deny the empirical evaluation of the terrestrial water budget equation for estimating runoff. It is therefore investigated, if a data assimilation approach is able to improve the estimates of the different water cycle variables in order to close the water budgets on the catchment scale and and if this will therefore lead to reasonable runoff predictions. The approach is based on an Ensemble Kalman filter framework, which can be applied for estimating and predicting runoff over poorly or irregularly gauged basins. The prediction step within the Kalman filter makes use of the so-called least squares prediction, where the joint temporal and spatial auto- and cross-covariance structures of precipitation, evapotranspiration, runoff, and water storage changes are taken into account. The required covariance matrices are derived from a wide collection of state-of-the-art global hydrological and hydrometeorological datasets, which have been already applied and analyzed in the previous articles. A preliminary analysis of the least squares predictor, which is similar to an autoregressive model with a time-lag of one month or the multiple ordinary least squares regression, reveals that the approach is able to reproduce about 70 % of the variance of the runoff anomalies. This already indicates that the short- and long-term correlations between runoff and the other water cycle variables contain enough information for predicting basin-scale runoff at the anomaly-level (i.e., the deviations from the mean annual cycle).

The different hydrological and hydrometeorological datasets are further used during the correction step, where the predicted state is combined with the ensemble mean of each of the four water cycle variables under consideration of time- and space-dependent uncertainty estimates. In total, the article discusses six different configurations of the proposed data assimilation framework: the Ensemble Kalman Filter

(Smoother) and the hard (soft) constrained Ensemble Kalman Filter (Smoother). It is discussed if hard constraints, which enforce perfect water budget closure, or soft constraints, which allow some well-defined imbalances, change the runoff estimates and if the application of a smoothing step is able to improve the results. The framework is run for 29 large river basins of different climate regions, with which runoff is predicted for a subset of 16 basis.

For the in-depth analysis of the framework, the runoff predictions from the six configurations are compared against runoff observations from the GRDC during the period 2005–2010. First of all, there are no significant differences between the runoff dynamics from the analyzed approaches. But a closer look reveals some deviations mainly in the peaks and troughs of the estimated runoff time series. The different performance metrics indeed show that especially the constrained estimates suffer from slight biases. As the soft-constrained yield better results than the hard-constrained predictions, it can be assumed that the uncertainties and mismatches (e.g., due to different spatial or temporal resolutions) between the datasets for the water cycle variables do not allow a reasonable water budget closure. In other words, forcing the water budgets to be closed leads to unrealistic corrections of the different water cycle variables.

Nevertheless, the approach is able to successfully reproduce the runoff observations over most of the study regions. The correlation between the estimates and the GRDC observations is larger than 0.8 for 12 of the 16 catchments, independent of the configuration. An analysis of the Nash Sutcliffe Efficiency, which is a rather sensitive performance metric, reveals positive values for all study regions. Moreover, the NSE_{mean} is larger than 0.5 for 13 catchments, even if the smoothed and constrained estimates are slightly worse. As runoff follows a dominant annual cycle over most catchments, a further metric, which takes the long-term-mean annual cycle into account, is evaluated as well. The $\text{NSE}_{\text{cycle}}$ also shows positive values for 14 of the 16 study regions. This clearly indicates that, despite the reasonable estimation of the annual cycle, the approach is further able to reproduce irregular long- and short-term variations in the runoff time series. In other words, as these results are based on the prediction from other water cycle variables only, there are enough information and similarities in the basin-scale time series of precipitation, evapotranspiration, and water storage changes for deriving realistic runoff predictions. It will be part of future work to exploit these relationships in order to provide runoff estimates for currently poorly or ungauged catchments, which cover an area of 11,500,000 km^2 and provide a discharge volume of more than 125,000 $^{m^3}/_{s}$ of freshwater resources. It will be also investigated if such an approach is also able to provide runoff predictions from, for example, climate models, which will then allow a consistent global-scale analysis of trends in the catchment-scale runoff time series.

.

Chapter 7

Conclusion and Outlook

On the global scale, the study clearly shows that there are significant differences between the three widely used atmospheric reanalysis models and multiple gridded observation datasets. It is further stressed that there are large imbalances in their oceanic and continental moisture budgets. One reason for such inconsistencies are the variations in the observations, which are assimilated into the models. The time series of oceanic $P - E$ from the MERRA and CFSR reanalyses show a significant increase during 1998, which coincides with the assimilation of satellite-based radiance sounding data from that year. Besides such highly unrealistic gaps in the time series, it can be assumed that changes in the number and type of assimilated observations can also cause smaller imbalances, which, in principle, hinder the use of such models for climate trend studies. The same holds true for the gridded observations from, for example, GPCC or CRU, as these are also based on a constantly changing number of *in situ* rain gauges. These changes occur both in the number and spatial distribution of gauges, while large areas over less developed regions remain completely unobserved. The analysis of the long-term mean patterns of precipitation and the water budget closure on the catchment-scale clearly shows that there are significant differences between all analyzed datasets, which cause an imbalance of the water budgets of more than 25 % of the mean annual runoff over most study regions. Similar results are reported from Sahoo et al. [2011]. They analyzed the water budget closure from combined satellite and *in situ* discharge products over 10 study regions. It is stressed that water budget closure is not achievable with errors of the order of 5–25 % of mean annual precipitation.

It is further shown that there is no single dataset which performs consistently well on the global scale. This point has to be stressed even further. Even if the analyzed data sources provide global estimates of each of the four water cycle variables, we can not assume a consistent quality at every single location. This, however, is due to numerous reasons. The state-of-the-art hydrological and hydrometeorological data originates from completely different sensors, modeling approaches, and other sources, which differ in terms of spatial and temporal resolution, but also in their respective uncertainties and error levels. All these differences in the data quality and quantity obviously lead to biases in the analysis of the global water budget.

On the other hand, the study shows that there is huge potential in our current data sources for the major water cycle variables. Over several regions, the time series for precipitation, evapotranspiration, runoff, and water storage changes from publicly available global datasets already allow a reasonable empirical analysis of the water cycle and the respective interactions of the four major variables. Using data correction algorithms like the Ensemble Kalman Filter, which is proposed in this work, can also lead to improved and consistent estimates of the water cycle variables on the global scale. It will be therefore analyzed, if the presented algorithm can be further improved. It is also shown that runoff estimates from satellite altimetry provide a reasonable alternative to, for example, classical hydrological modeling approaches. Promising results are further presented in, for example, Rodell et al. [2015], where optimization algorithms are applied in order improve the water budget closure on different spatial and temporal scales. It is reported that the global surface water budget closure error from observational estimates is 3.9 % of precipitation with an uncertainty of 6.1 %.

Besides such mathematical correction approaches, there are multiple new satellite missions, which will allow a significantly improved space-borne observation of different water cycle variables: Since its launch on January 31st, 2015, NASA's Soil Moisture Active Passive [SMAP; Entekhabi et al., 2010] mission provides global high-resolution soil moisture estimates. A new era in the satellite-based observation of precipitation was started with the launch of the core observatory for NASA's Global Precipitation Measurement [GPM; Hou et al., 2014] mission on February

27th, 2014. The mission uses an international network of satellites, from which precipitation observations are collected and combined. It is indented to provide global high-resolution precipitation maps every three hours. The successful concept of the GRACE mission will be continued by the GRACE Follow-On (GRACE-FO) mission, whose launch is scheduled for August 2017. In the context of runoff, it is assumed that especially the dedicated Surface Water and Ocean Topography [SWOT; Fu et al., 2009] mission or multi-satellite altimeter approaches will significantly increase number of catchments, where runoff can be estimated, and the quality of these estimates.

However, for most regions, we currently have to face large uncertainties and biases between the different datasets on the one, and a significantly decreasing number of *in situ* observations, on the other hand. In the context of the projected changes in the global water cycle, it is thus of major importance to call attention to the strengths and, maybe even more important, weaknesses of our current datasets. Furthermore, if the decrease in the number of *in situ* stations for the most important water cycle variables will continue, it will become more and more difficult to describe, analyze and understand the large-scale water cycle and the interactions between it's variables.

On the other hand, the new approaches and developments in the global monitoring of the major water cycle variables are without any doubt a step towards the global water balance closure across different spatial and temporal scales. They further have to be accepted and considered as a promising alternative to, for example, classical hydrological and hydrometeorological modeling approaches. In fact, such observation-based empirical approaches for analyzing the water budgets require far less computational effort and might be more straightforward and easier to apply while less parameterized compared to highly complex models. That being said, it has to be stressed that the hydrological community has to focus on both branches for further understanding the dynamics and interactions of the climate system. Thus, one of the major efforts of the modern hydrological community has to be the combination and joint analysis of all the different data-sources from numerous satellites, new and innovative terrestrial observation systems, and modeling approaches. It is therefore highly notable that international initiatives like, for example, GEWEX provide a platform which brings together scientists from different hydrological and hydrometeorological branches with a common goal. Only such joint initiatives and collaborations allow to further improve our current knowledge of the global water cycle and to finally conclude on the future development of our climate system.

Appendix A

Contribution of the Author to the Different Papers

Chapter 2: The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis The study was conceived and designed by Christof Lorenz and Harald Kunstmann. All computations, the final analysis and interpretation of the results, and the preparation of the figures were performed by Christof Lorenz. The paper was mainly written by Christof Lorenz with contributions from Harald Kunstmann.

Chapter 3: Large-Scale Runoff from Landmasses: A Global Assessment of the Closure of the Hydrological and Atmospheric Water Balances The paper was mainly written by Christof Lorenz and Balaji Devaraju with contributions and corrections from Harald Kunstmann and Nico Sneeuw. Christof Lorenz further collected all the data, performed the water budget calculations from the hydrological and hydrometeorological datasets, and designed the evaluation strategy for the large amount of runoff estimates from different approaches. The section about the runoff-precipitation ratio, together with the corresponding runoff estimates, was provided from Mohammad J. Tourian and Johannes Riegger. The water storage changes from GRACE were computed by Balaji Devaraju. The comparison and evaluation of the runoff estimates from the different approaches was performed by Christof Lorenz, while Balaji Devaraju contributed to the analysis and interpretation of the results. The figures in the paper were generated by Christof Lorenz and Balaji Devaraju.

Chapter 4: Estimating Runoff Using Hydro-Geodetic Approaches The paper was mainly written by Nico Sneeuw, Christof Lorenz, and Balaji Devaraju with contributions from Mohammad J. Tourian and Harald Kunstmann. The collection of data, the computation of the budget-based runoff estimates, the comparison of the different approaches, and the interpretation of the results was mainly performed by Christof Lorenz with contributions from Balaji Devaraju. The section about altimetry-based runoff estimation and the runoff-precipitation ratio was provided by Mohammad J. Tourian. The figures were generated by Christof Lorenz, Mohammad J. Tourian, and Balaji Devaraju.

Chapter 5: Basin-Scale Runoff Prediction: An Ensemble Kalman Filter Framework based on Global Hydrometeorological Datasets The study and the corresponding data assimilation approach under discussion was developed and designed by Christof Lorenz and Mohammad J. Tourian. Christof Lorenz developed the mathematical framework, while Mohammad J. Tourian contributed to the process model and provided runoff estimates from satellite altimetry. The collection of data, all computations, and the final assimilation and prediction of runoff was performed by Christof Lorenz. The water storage changes from GRACE were provided by Balaji Devaraju. Christof Lorenz further generated all the figures and performed the final analysis and interpretation of the results. The text was mainly written by Christof Lorenz with contributions from Mohammad J. Tourian, Balaji Devaraju, and Harald Kunstmann.

Bibliography

Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *Journal of Hydrometeorology*, **4**, 1147–1167, doi:10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2.

Aghakouchak, A., A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, **39**, L09 406, doi:10.1029/2012GL051592.

Alsdorf, D. E., E. Rodríguez, and D. P. Lettenmaier, 2007: Measuring surface water from space. *Reviews of Geophysics*, **45**, RG2002, doi:10.1029/2006RG000197.

Andersson, A., K. Fennig, C. Klepp, S. Bakan, H. Graßl, and J. Schulz, 2010: The Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data – HOAPS-3. *Earth System Science Data*, **2**, 215–234, doi:10.5194/essd-2-215-2010.

Asrar, G., J. A. Kaye, and P. Morel, 2001: NASA Research Strategy for Earth System Science: Climate Component. *Bulletin of the American Meteorological Society*, **82**, 1309–1329, doi:10.1175/1520-0477(2001)082<1309:NRSFES>2.3.CO;2.

Baldocchi, D., and Coauthors, 2001: FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, **82**, 2415–2434, doi:10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2.

Bengtsson, L., 2004: Can climate trends be calculated from reanalysis data? *Journal of Geophysical Research*, **109**, D11 111, doi:10.1029/2004JD004536.

Berrisford, P., D. Dee, K. Fielding, M. Fuentes, P. Kallberg, S. Kobayashi, and S. Uppala, 2009: The ERA-Interim archive Version 1.0. Era report series, ECMWF, 16 pp.

Berry, P. A. M., 2005: Global inland water monitoring from multi-mission altimetry. *Geophysical Research Letters*, **32**, L16 401, doi:10.1029/2005GL022814.

Bettadpur, S., 2007: Gravity Recovery and Climate Experiment, Level-2 Gravity Field Product User Handbook. *Center for Space Research, The University of Texas at Austin*, **CSR-GR-03-01**, 1–18.

Bettadpur, S., 2012: UTCSR level-2 processing standards document (for level-2 product release 0005). Tech. Rep. Rev. 4.0, Center for Space Research, The University of Texas at Austin.

Betts, A. K., J. H. Ball, and P. Viterbo, 1999: Basin-scale surface water and energy budgets for the Mississippi from the ECMWF reanalysis. *Journal of Geophysical Research*, **104**, 293–306, doi:10.1029/1999JD900056.

Betts, A. K., J. H. Ball, and P. Viterbo, 2003: Evaluation of the ERA-40 Surface Water Budget and Surface Temperature for the Mackenzie River Basin. *Journal of Hydrometeorology*, **4**, 1194–1211, doi:10.1175/1525-7541(2003)004<1194:EOTESW>2.0.CO;2.

Betts, A. K., J. H. Ball, P. Viterbo, A. Dai, and J. Marengo, 2005: Hydrometeorology of the Amazon in ERA-40. *Journal of Hydrometeorology*, **6**, 764–774, doi:10.1175/JHM441.1.

Betts, A. K., M. Köhler, and Y. Zhang, 2009: Comparison of river basin hydrometeorology in ERA-Interim and ERA-40 reanalyses with observations. *Journal of Geophysical Research*, **114**, D02 101, doi:10.1029/2008JD010761.

Birkett, C. M., 1998: Contribution of the TOPEX NASA Radar Altimeter to the global monitoring of large rivers and wetlands. *Water Resources Research*, **34**, 1223–1239, doi:10.1029/98WR00124.

Birkinshaw, S. J., G. M. O'Donnell, P. Moore, C. G. Kilsby, H. J. Fowler, and P. A. M. Berry, 2010: Using satellite altimetry data to augment flow estimation techniques on the Mekong River. *Hydrological Processes*, **24**, 3811–3825, doi:10.1002/hyp.7811.

Bjerklie, D. M., S. Lawrence Dingman, C. J. Vorosmarty, C. H. Bolster, and R. G. Congalton, 2003: Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, **278**, 17–38, doi:10.1016/S0022-1694(03)00129-X.

Blöschl, G., M. Sivapalan, and T. Wagener, 2013: *Runoff prediction in ungauged basins: Synthesis across processes, places and scales*. Cambridge University Press.

Bonan, G. B., 1998: The land surface climatology of the NCAR Land Surface Model coupled to the NCEP Community Climate Model. *Journal of Climate*, **11**, 1307–1326, doi:10.1175/1520-0442(1998)011<1307:TLSCOT>2.0.CO;2.

Bosilovich, M., J. Chen, F. R. Robertson, and R. F. Adler, 2008: Evaluation of Global Precipitation in Reanalyses. *Journal of Applied Meteorology and Climatology*, **47**, 2279–2299, doi:10.1175/2008JAMC1921.1.

Bosilovich, M. G., F. R. Robertson, and J. Chen, 2011: Global Energy and Water Budgets in MERRA. *Journal of Climate*, **24**, 5721–5739, doi:10.1175/2011JCLI4175.1.

Bytheway, J. L., and C. D. Kummerow, 2013: Inferring the uncertainty of satellite precipitation estimates in data-sparse regions over land. *Journal of Geophysical Research: Atmospheres*, **118**, 9524–9533, doi:10.1002/jgrd.50607.

Carpenter, S. R., E. H. Stanley, and M. J. V. Zanden, 2011: State of the World's Freshwater Ecosystems: Physical, Chemical, and Biological Changes. *Annual Review of Environment and Resources*, **36**, 75–99, doi:10.1146/annurev-environ-021810-094524.

Chahine, M. T., 1992a: GEWEX: The Global Energy and Water Cycle Experiment. *Eos, Transactions American Geophysical Union*, **73**, 9–14, doi:10.1029/91EO00007.

Chahine, M. T., 1992b: The hydrological cycle and its influence on climate. *Nature*, **359**, 373–380, doi:10.1038/359373a0.

Chen, F., and Coauthors, 1996: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations. *Journal of Geophysical Research*, **101**, 7251–7268, doi:10.1029/95JD02165.

Chen, J. L., M. Rodell, C. R. Wilson, and J. S. Famiglietti, 2005: Low degree spherical harmonic influences on Gravity Recovery and Climate Experiment (GRACE) water storage estimates. *Geophysical Research Letters*, **32**, L14 405, doi: 10.1029/2005GL022964.

Chen, J. L., C. R. Wilson, and B. D. Tapley, 2010: The 2009 exceptional Amazon flood and interannual terrestrial water storage change observed by GRACE. *Water Resources Research*, **46**, doi:10.1029/2010WR009383.

Chen, M., W. Shi, P. Xie, V. B. S. Silva, V. E. Kousky, R. W. Higgins, and J. E. Janowiak, 2008: Assessing objective techniques for gauge-based analyses of global daily precipitation. *Journal of Geophysical Research*, **113**, D04 110, doi: 10.1029/2007JD009132.

Cheng, M., and J. Ries, 2013: Monthly estimates of C20 from 5 SLR satellites based on GRACE RL05 models. GRACE Technical Note 7, Center for Space Research, The University of Texas at Austin.

Chido, G., and L. Haimberger, 2009: Interannual changes in mass consistent ernergy budgets from ERA-Interim and satellite data. *Journal of Geophysical Research*, **115**, D02 112, doi:10.1029/2009JD012049.

CLIVAR Scientific Steering Group and World Climate Research Programme, 1995: *CLIVAR, a study of climate variability and predictability: science plan*, Vol. 89. World Climate Research Programme.

Coe, M. T., and C. M. Birkett, 2004: Calculation of river discharge and prediction of lake height from satellite radar altimetry: Example for the Lake Chad basin. *Water Resource Research*, **70**, W102 205, doi:10.1029/2003WR002543.

Da Silva, J. S., S. Calmant, F. Seyler, O. C. R. Filho, G. Cochonneau, and W. J. Mansur, 2010: Water levels in the Amazon basin derived from the ERS 2 and ENVISAT radar altimetry missions. *Remote Sensing of Environment*, **114**, 2160–2181, doi: 10.1016/j.rse.2010.04.020.

Dahle, C., F. Flechtner, C. Gruber, D. König, R. König, G. Michalak, and K.-H.Neumayer, 2013: GFZ GRACE level-2 processing standards document (for level-2 productrelease 0005). Scientific Technical Report Rev. 1.1, GFZ German Research Centre for Geosciences.

Dai, A., and K. E. Trenberth, 2002: Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations. *Journal of Hydrometeorology*, **3**, 660–687, doi:10.1175/1525-7541(2002)003<0660:EOFDFC>2.0.CO;2.

Dai, Y., and Coauthors, 2003: The Common Land Model. *Bulletin of the American Meteorological Society*, **84**, 1013–1023, doi:10.1175/BAMS-84-8-1013.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**, 553–597, doi:10.1002/qj.828.

Derber, J. C., D. F. Parrish, and S. J. Lord, 1991: The new global operational analysis system at the National Meteorological Center. *Weather and Forecasting*, **6**, 538–547, doi:10.1175/1520-0434(1991)006<0538:TNGOAS>2.0.CO;2.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy, 2009: Biosphere-Atmosphere Transfer Scheme (BATS) version 1e as coupled to the NCAR Community Climate Model. *Journal of Geophysical Research*, **115**, D02 112, doi: 10.1029/2009JD012049.

Döll, P., F. Kaspar, and B. Lehner, 2003: A global hydrological model for deriving water availability indicators: model tuning and validation. *Journal of Hydrology*, **270**, 105–134, doi:10.1016/S0022-1694(02)00283-4.

Döll, P., H. M. Schmied, C. Schuh, F. T. Portmann, and A. Eicker, 2014: Global-scale assessment of groundwater depletion and related groundwater abstractions: Combining hydrological modeling with information from well observations and GRACE satellites. *Water Resources Research*, **50**, 5698–5720, doi:10.1002/2014WR015595.

Dorigo, W. A., and Coauthors, 2011: The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, **15**, 1675–1698, doi:10.5194/hess-15-1675-2011.

Durand, M., L.-L. Fu, D. P. Lettenmaier, D. E. Alsdorf, E. Rodriguez, and D. Esteban-Fernandez, 2010: The Surface Water and Ocean Topography Mission: Observing Terrestrial Surface Water and Oceanic Submesoscale Eddies. *Proceedings of the IEEE*, **98**, 766–779, doi:10.1109/JPROC.2010.2043031.

Ek, M. B., 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. *Journal of Geophysical Research*, **108**, 8851, doi:10.1029/2002JD003296.

Entekhabi, D., and Coauthors, 2010: The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, **98**, 704–716, doi:10.1109/JPROC.2010.2043918.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**, 10 143, doi:10.1029/94JC00572.

Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367, doi:10.1007/s10236-003-0036-9.

Falkenmark, M., and J. Rockström, 2004: Balancing water for humans and nature: the new approach in ecohydrology. Tech. rep., Earthscan.

Falloon, P., R. Betts, A. Wiltshire, R. Dankers, C. Mathison, D. McNeall, P. Bates, and M. Trigg, 2011: Validation of River Flows in HadGEM1 and HadCM3 with the TRIP River Flow Model. *Journal of Hydrometeorology*, **12**, 1157–1180, doi: 10.1175/2011JHM1388.1.

Fekete, B. M., U. Looser, A. Pietroniro, and R. D. Robarts, 2012: Rationale for Monitoring Discharge on the Ground. *Journal of Hydrometeorology*, **13**, 1977–1986, doi:10.1175/JHM-D-11-0126.1.

Fekete, B. M., and C. J. Vörösmarty, 2002: The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *Predictions in Ungauged Basins: PUB kick-off (Proceedings of the PUB kick-off meeting held in Brasilia, 20–22 November 2002). IAHS Publication*, Vol. 349.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, 2002: High-resolution fields of global runoff combining observed river discharge and simulated water balances. *Global Biogeochemical Cycles*, **16**, 15,1–15,10, doi:10.1029/1999GB001254.

Fellous, J.-L., 2014: Global Climate Observing System. *Encyclopedia of Remote Sensing*, E. Njoku, Ed., Encyclopedia of Earth Sciences Series, Springer New York, 254–257, doi:10.1007/978-0-387-36699-9_56.

Ferguson, C., J. Sheffield, E. Wood, and H. Gao, 2010: Quantifying uncertainty in a remote sensing-based estimate of evapotranspiration over continental USA. *International Journal of Remote Sensing*, **31**, 3821–3865, doi:10.1080/01431161.2010.483490.

Fersch, B., and Coauthors, 2012: Continental-Scale Basin Water Storage Variation from Global and Dynamically Downscaled Atmospheric Water Budgets in Comparison with GRACE-Derived Observations. *Journal of Hydrometeorology*, **13**, 1589–1603, doi:10.1175/JHM-D-11-0143.1.

Field, C. B., V. Barros, T. F. Stocker, and Q. D. Et al., Eds., 2012: *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Forootan, E., J. L. Awange, J. Kusche, B. Heck, and A. Eicker, 2012: Independent patterns of water mass anomalies over Australia from satellite data and models. *Remote Sensing of Environment*, **124**, 427–443, doi:10.1016/j.rse.2012.05.023.

Frappart, F., S. Calmant, M. Cauhopé, F. Seyler, and A. Cazenave, 2006: Preliminary results of ENVISAT RA-2-derived water levels validation over the Amazon basin. *Remote Sensing of Environment*, **100**, 252–264, doi:10.1016/j.rse.2005.10.027.

Frappart, F., G. Ramillien, and J. Ronchail, 2013: Changes in terrestrial water storage versus rainfall and discharges in the Amazon basin. *International Journal of Climatology*, **33**, 3029–3046, doi:10.1002/joc.3647.

Frauenfeld, O. W. and Zhang, T. and Serreze, M. C., 2005: Climate change and variability using European Centre for Medium-Range Weather Forecasts reanalysis (ERA-40) temperatures on the Tibetan Plateau. *Climate and Dynamics*, **110**, D02 101, doi:10.1029/2004JD005230.

Fu, L.-L., D. Alsdorf, E. Rodriguez, R. Morrow, N. Mognard, J. Lambin, P. Vaze, and T. Lafon, 2009: The SWOT (Surface Water and Ocean Topography) Mission: spaceborne radar interferometry for oceanographic and hydrological applications. *OCEANOBS'09 Conference*.

Fu, L.-L., and A. Cazenave, 2001: *Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications*. International Geophysical series.

Fuchs, T., J. Rapp, F. Rubel, and B. Rudolf, 2001: Correction of synoptic precipitation observations due to systematic measuring errors with special regard to precipitation phases. *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, **26**, 689–693, doi:0.1016/S1464-1909(01)00070-3.

Getirana, A. C. V., 2010: Integrating spatial altimetry data into the automatic calibration of hydrological models. *Journal of Hydrology*, **387**, 244–255, doi:10.1016/j.jhydrol.2010.04.013.

Getirana, A. C. V., and C. Peters-Lidard, 2013: Estimating water discharge from large radar altimetry datasets. *Hydrology and Earth System Sciences*, **17**, 923–933, doi:10.5194/hess-17-923-2013.

Gilchrist, W. G., 2000: *Statistical modelling with quantile functions*. London: Chapman & Hall.

Gleick, P. H., H. Cooley, J. S. Famiglietti, D. P. Lettenmaier, T. Oki, C. J. Vörösmarty, and E. F. Wood, 2013: Improving Understanding of the Global Hydrologic Cycle.

Climate Science for Serving Society, G. R. Asrar, and J. W. Hurrell, Eds., Springer Netherlands, Dordrecht, 151–184, doi:10.1007/978-94-007-6692-1_6.

GRDC, 2013: Report of the Tenth Meeting of the GRDC Steering Committee, 15 –17 June 2011, Koblenz, Germany. Tech. rep., GRDC, 42 pp.

Gudmundsson, L., and S. I. Seneviratne, 2014: Towards observation based gridded runoff estimates for Europe. *Hydrology and Earth System Sciences Discussions*, **11**, 12 883–12 932, doi:10.5194/hessd-11-12883-2014.

Gupta, H. V., S. Shrooshian, and P. O. Yapo, 1999: Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *Journal of Hydrologic Engineering*, **4**, 135–143, doi:10.1061/(ASCE)1084-0699(1999)4:2(135).

Hagemann, S., K. Arpe, and L. Bengtsson, 2005: Validation of the hydrological cycle of ERA-40. *ERA Report Series*, ECMWF, Reading, U. K., Vol. 24, 42pp.

Harris, I., P. Jones, T. Osborn, and D. Lister, 2014: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset. *International Journal of Climatology*, **34**, 623–642, doi:10.1002/joc.3711.

Higgins, R. W., V. E. Kousky, V. B. S. Silva, E. Becker, and P. Xie, 2010: Intercomparison of Daily Precipitation Statistics over the United States in Observations and in NCEP Reanalysis Products. *Journal of Climate*, **23**, 4637–4650, doi:10.1175/2010JCLI3638.1.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bulletin of the American Meteorological Society*, **95**, 701–722, doi:10.1175/BAMS-D-13-00164.1.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Hrachowitz, M., and Coauthors, 2013: A decade of Predictions in Ungauged Basins (PUB)—a review. *Hydrological Sciences Journal*, **58**, 1198–1255, doi:10.1080/02626667.2013.803183.

Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP Version 2.1. *Geophysical Research Letters*, **36**, L17 808, doi:10.1029/2009GL040000.

Huntington, T. G., 2006: Evidence for intensification of the global water cycle: Review and synthesis. *Journal of Hydrology*, **319**, 83–95, doi:10.1016/j.jhydrol.2005.07.003.

ICGEM, 2013: International Center for Global Gravity Field Models. Tech. rep., GFZ.

IPCC, 2007: Chapter 3—Observations: Surface and Atmospheric Climate Change. *Climate Change 2007: The Physical Science Basis – Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, P. D. et al. Trenberth, K.E., Jones, Ed., Cambridge University Press, 996.

IPCC, 2014: Summary for policymakers. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, C. B. Field, V. R. Barros, D. J. Dokken, K. J. Mach, M. D. Mastrandrea, and T. E. B. Et al., Eds., Cambridge University Press, 1–32.

Jacobson, M. Z., 2007: *Fundamentals of Atmospheric Modeling*. 2nd ed., Cambridge University Press, 813 pp.

Janowiak, J. E., A. Gruber, C. R. Kondragunta, R. E. Livezey, and G. J. Huffman, 1997: A Comparison of the NCEP-NCAR Reanalysis Precipitation and the GPCP Rain Gauge-Satellite Combined Dataset with Observational Error Considerations. *Journal of Climate*, **11**, 2960–2979, doi:10.1175/1520-0442(1998)011<2960:ACOTNN>2.0.CO;2.

Jones, P., 1999: First-and second-order conservative remapping schemes for grids in spherical coordinates. *Monthly Weather Review*, **127**, 2204–2210, doi:10.1175/1520-0493(1999)127<2204:FASOCR>2.0.CO;2.

Jung, M., M. Reichstein, and A. Bondeau, 2009: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model. *Biogeosciences*, **6**, 2001–2013, doi:10.5194/bg-6-2001-2009.

Jung, M., and Coauthors, 2010: Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature*, **467**, 951–954, doi:10.1038/nature09396.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, **77**, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kang, D. H., X. Shi, H. Gaon, and S. J. Déry, 2014: On the Changing Contribution of Snow to the Hydrology of the Fraser River Basin, Canada. *Journal of Hydrometeorology*, **15**, 1344–1365, doi:10.1175/JHM-D-13-0120.1.

Kerr, Y. H., and Coauthors, 2010: The SMOS Mission: New Tool for Monitoring Key Elements ofthe Global Water Cycle. *Proceedings of the IEEE*, **98**, 666–687, doi:10.1109/JPROC.2010.2043032.

Khan, S. I., and Coauthors, 2011: Satellite Remote Sensing and Hydrologic Modeling for Flood Inundation Mapping in Lake Victoria Basin: Implications for Hydrologic Prediction in Ungauged Basins. *Geoscience and Remote Sensing, IEEE Transactions on*, **49**, 85–95, doi:10.1109/TGRS.2010.2057513.

Kirtman, B. P., and Coauthors, 2013: The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bulletin of the American Meteorological Society*, **95**, 585–601, doi:10.1175/BAMS-D-12-00050.1.

Kistler, R., and Coauthors, 2001: The NCEP-NCAR 50-Year Reanalysis: Monthly Means CD-ROM and Documentation. *Bulletin of the American Meteorological Society*, **82**, 247–268, doi:10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2.

Kleidon, A., and S. Schymanski, 2008: Thermodynamics and optimality of the water budget on land: A review. *Geophysical Research Letters*, **35**, L20 404, doi: 10.1029/2008GL035393.

Koblinsky, C. J., R. T. Clarke, A. C. Brenner, and H. Frey, 1993: Measurement of river level variations with satellite altimetry. *Water Resources Research*, **29**, 1839–1848, doi:10.1029/93WR00542.

Kopp, T. J., and R. B. Kiess, 1996: The air force global weather central cloud analysis model. *15th Conf. on Weather Analysis and Forecasting*, Norfolk, VA, Amer. Meteor. Soc., 220–222.

Koren, V., J. Schaake, K. Mitchell, Q. Y. Duan, F. Chen, and J. M. Baker, 1999: A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research*, **104**, 569–585, doi:10.1029/1999JD900232.

Koster, R. D., and M. J. Suarez, 1996: Energy and water balance calculations in the mosaic lsm. *NASA technical memorandum*, National Aeronautics and Space Administration, Goddard Space Flight Center, Laboratory for Atmospheres, Data Assimilation Office: Laboratory for Hydrospheric Processes.

Kouraev, A. V., E. A. Zakharova, O. Samain, N. M. Mognard, and A. Cazenave, 2004: Ob river discharge from TOPEX/Poseidon satellite altimetry (1992-2002). *Remote Sensing of Environment*, **93**, 238–245, doi:10.1016/j.rse.2004.07.007.

Kummerow, C., and Coauthors, 2000: The Status of the Tropical Rainfall Measuring Mission (TRMM) after Two Years in Orbit. *Journal of Applied Meteorology*, **39**, 1967–1982, doi:10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2.

Kurtenbach, E., A. Eicker, T. Mayer-Gürr, M. Holschneider, M. Hayn, M. Fuhrmann, and J. Kusche, 2012: Improved daily GRACE gravity field solutions using a Kalman smoother. *Journal of Geodynamics*, **59-60**, 39–48, doi:10.1016/j.jog.2012.02.006.

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

Kusche, J., R. Schmidt, S. Petrovic, and R. Rietbroek, 2009: Decorrelated GRACE time-variable gravity solutions by GFZ, and their validation using a hydrological model. *Journal of Geodesy*, **83**, 903–913, doi:10.1007/s00 190–009–0308–3.

Landerer, F. W., J. O. Dickey, and A. Güntner, 2010: Terrestrial water budget of the Eurasian pan-Arctic from GRACE satellite measurements during 2003–2009. *Journal of Geophysical Research*, **115**, D23 115, doi:10.1029/2010JD014584.

Landerer, F. W., and S. C. Swenson, 2012: Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, **48**, W04 531, doi:10.1029/2011WR011453.

Lawford, R., A. Strauch, D. Toll, B. Fekete, and D. Cripe, 2013: Earth observations for global water security. *Current Opinion in Environmental Sustainability*, **5**, 633–643, doi:10.1016/j.cosust.2013.11.009.

Legates, D. R., and C. J. Willmott, 1990: Mean seasonal and spatial variability in gauge-corrected, global precipitation. *International Journal of Climatology*, **10**, 111–127, doi:10.1002/joc.3370100202.

Leon, J. G., S. Calmant, F. Seyler, M.-P. Bonnet, M. Cauhop, F. Frappart, N. Filizola, and P. Fraizy, 2006: Rating curves and estimation of average water depth at the upper Negro River based on satellite altimeter data and modeled discharges. *Journal of Hydrology*, **328**, 481–496, doi:10.1016/j.jhydrol.2005.12.006.

Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges, 1994: A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, **99**, 14 415–14 428, doi:10.1029/94JD00483.

Lohmann, D., and Coauthors, 2004: Streamflow and water balance intercomparisons of four land surface models in the North American Land Data Assimilation System project. *Journal of Geophysical Research: Atmospheres (1984–2012)*, **109**, D07S91, doi:10.1029/2003JD003517.

Long, D., L. Longuevergne, and B. R. Scanlon, 2014: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites. *Water Resources Research*, **50**, 1131–1151, doi:10.1002/2013WR014581.

Longuevergne, L., B. R. Scanlon, and C. R. Wilson, 2010: GRACE Hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resources Research*, **46**, W11 517, doi:10.1029/2009WR008564.

Lorenz, C., 2009: Applying stochastic constraints on time-variable GRACE data. Diploma thesis, University of Stuttgart, Germany.

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

Lorenz, C., and H. Kunstmann, 2012: The Hydrological Cycle in Three State-of-the-Art Reanalyses: Intercomparison and Performance Analysis. *Journal of Hydrometeorology*, **13**, 1397–1420, doi:10.1175/JHM-D-11-088.1.

Lorenz, C., M. J. Tourian, B. Devaraju, N. Sneeuw, and H. Kunstmann, 2015: Basin-scale runoff prediction: An Ensemble Kalman Filter framework based on global hydrometeorological data sets. *Water Resources Research*, **51**, 8450–8475, doi:10.1002/2014WR016794.

Matsuura, K., and C. J. Willmot, 2009: *Terrestrial Air Temperature: 1900-2008 Gridded Monthly Time Series (Version 2.01)*. Center for Climatic Research, University of Delaware.

Matsuura, K., and C. J. Willmott, 2012: *Terrestrial Precipitation: 1900-2010 Gridded Monthly Time Series (Version 3.02)*. Center for Climatic Research, University of Delaware.

McMillan, H., J. Freer, F. Pappenberger, T. Krueger, and M. Clark, 2010: Impacts of uncertain river flow data on rainfall-runoff model calibration and discharge predictions. *Hydrological Processes*, **24**, 1270–1284, doi:10.1002/hyp.7587.

Milzow, C., P. E. Krogh, and P. Bauer-Gottwein, 2011: Combining satellite radar altimetry, SAR surface soil moisture and GRACE total storage changes for hydrological model calibration in a large poorly gauged catchment. *Hydrology and Earth System Sciences*, **15**, 1729–1743, doi:10.5194/hess-15-1729-2011.

Miralles, D. G., T. R. H. Holmes, R. A. M. De Jeu, J. H. Gash, A. G. C. A. Meesters, and A. J. Dolman, 2011a: Global land-surface evaporation estimated from satellite-based observations. *Hydrology and Earth System Sciences*, **15**, 453–469, doi:10.5194/hess-15-453-2011.

Miralles, D. G., R. A. M. D. Jeu, J. H. Gash, T. R. H. Holmes, and A. J. Dolman, 2011b: An application of GLEAM to estimating global evaporation. *Hydrology and Earth System Sciences Discussions*, **8**, 1–27, doi:10.5194/hessd-8-1-2011.

Mitchell, H. L., P. L. Houtekamer, and G. Pellerin, 2002: Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. *Monthly Weather Review*, **130**, 2791–2808, doi:10.1175/1520-0493(2002)130<2791:ESBAME>2.0.CO;2.

Mitchell, T. D., and P. D. Jones, 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, **25**, 693–712, doi:10.1002/joc.1181.

Moriasi, D. N., J. G. Arnold, M. W. V. Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, 2007: Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, **50**, 885–900.

Moritz, H., 1980: *Advanced physical geodesy*. Wichmann.

Mu, Q., F. A. Heinsch, M. Zhao, and S. W. Running, 2007: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data. *Remote Sensing of Environment*, **111**, 519–536, doi:10.1016/j.rse.2011.02.019.

Mu, Q., M. Zhao, and S. W. Running, 2011: Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, **115**, 1781–1800, doi:10.1016/j.rse.2011.02.019.

Mueller, B., M. Hirschi, and S. I. Seneviratne, 2010: New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data. *Hydrological Processes*, **25**, 996–1008, doi:10.1002/hyp.7652.

Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophysical Research Letters*, **38**, L06 492, doi:10.1029/2010GL046230.

Mueller, B., and Coauthors, 2013: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis. *Hydrology and Earth System Sciences*, **17**, 3707–3720, doi:10.5194/hess-17-3707-2013.

Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models: Part 1. A discussion of principles. *Journal of Hydrology*, **10**, 282–290, doi:10.1016/0022-1694(70)90255-6.

Nicolas, J. P., and D. H. Bromwich, 2011: Precipitation Changes in High Southern Latitudes from Global Reanalyses: A Cautionary Tale. *Surveys in Geophysics*, **32**, 475–494, doi:10.1007/s10712-011-9114-6.

NSIT, 2007: NASA Earth Science Implementation Plan for Energy and Water Cycle Research Predicting Energy and Water Cycle Consequences of Earth System Variability and Change. Tech. rep., NASA.

Oki, T., and K. Shinjiro, 2006: Global Hydrological Cycles and World Water Resources. *Science*, **313**, 1068–1072, doi:10.1126/science.1128845.

Paiva, R. C. D., D. C. Buarque, W. Colischonn, M. P. Bonnet, F. Frappart, S. Calmant, and C. A. B. Mendes, 2013: Large-scale hydrologic and hydrodynamic modelling of the Amazon River basin. *Water Resources*, **49**, 1124–1216, doi:10.1002/wrcr.20067.

Pan, M., A. K. Sahoo, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 2012: Multisource estimation of long-term terrestrial water budget for major global river basins. *Journal of Climate*, **25**, 3191–3206, doi:10.1175/JCLI-D-11-00300.1.

Pan, M., and E. F. Wood, 2006: Data Assimilation for Estimating the Terrestrial Water Budget Using a Constrained Ensemble Kalman Filter. *Journal of Hydrometeorology*, **7**, 534–547, doi:10.1175/JHM495.1.

Papa, F., F. Durand, W. B. Rossow, A. Rahman, and S. K. Bala, 2010: Satellite altimeter-derived monthly discharge of the Ganga-Brahmaputra River and its seasonal to interannual variations from 1993 to 2008. *Journal of Geophysical Research*, **115**, C12 013, doi:10.1029/2009JC006075.

Papa, F., A. Güntner, F. Frappart, C. Prigent, and W. B. Rossow, 2008: Variations of surface water extent and water storage in large river basins: A comparison of different global data sources. *Geophysical Research Letters*, **35**, L11 401, doi:10.1029/2008GL033857.

Parish, E. S., E. Kodra, K. Steinhaeuser, and A. R. Ganguly, 2012: Estimating future global per capita water availability based on changes in climate and population. *Computers & Geosciences*, **42**, 79–86, doi:10.1016/j.cageo.2012.01.019.

Parkinson, C. L., 2003: Aqua: an Earth-Observing Satellite mission to examine water and other climate variables. *Geoscience and Remote Sensing, IEEE Transactions on*, **41**, 173–183, doi:10.1109/TGRS.2002.808319.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate*. American Institute of Physics, New York, 520 pp.

Poccard, I., S. Janicot, and P. Camberlin, 2000: Comparison of rainfall structures between NCEP/NCAR reanalyses and observed data over tropical Africa. *Climate Dynamics*, **16**, 897–915, doi:10.1007/s003820000087.

Postel, S. L., G. C. Daily, and P. R. Ehrlich, 1996: Human Appropriation of Renewable Fresh Water. *Science*, **271**, 785–788, doi:10.1126/science.271.5250.785.

Qian, T., A. Dai, K. E. Trenberth, and K. W. Oleson, 2006: Simulation of Global Land Surface Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations. *Journal of Hydrometeorology*, **7**, 953–975, doi:10.1175/JHM540.1.

Ramillien, G., F. Frappart, A. Güntner, T. Ngo-Duc, A. Cazenave, and K. Laval, 2006: Time variations of the regional evapotranspiration rate from Gravity Recovery and Climate Experiment (GRACE) satellite gravimetry. *Water Resources Research*, **42**, W10 403, doi:10.1029/2005WR004331.

Raschke, E., and Coauthors, 1998: The Baltic Sea Experiment BALTEX: A brief overview and some selected results of the authors. *Surveys in Geophysics*, **19**, 1–22, doi:10.1023/A:1006567924820.

Rasmusson, E. M., 1968: Atmospheric Water vapor transport and the water balance of North America. II. Large-scale water valance investigations. *Monthly Weather Review*, **96**, 720–734, doi:10.1175/1520-0493(1968)096<0720:AWVTAT>2.0.CO;2.

Redelsperger, J.-L., C. D. Thorncroft, A. Diedhiou, T. Lebel, D. J. Parker, and J. Polcher, 2006: African Monsoon Multidisciplinary Analysis: An International Research Project and Field Campaign. *Bulletin of the American Meteorological Society*, **87**, 1739–1746, doi:10.1175/BAMS-87-12-1739.

Reichle, R., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic Data Assimilation with the Ensemble Kalman Filter. *Journal of Hydrometeorology*, **130**, 103–114, doi:10.1175/1520-0493(2002)130<0103:HDAWTE>2.0.CO;2.

Reichle, R. H., R. D. Kostera, G. J. M. D. Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Touré, 2011: Assessment and Enhancement of MERRA Land Surface Hydrology Estimates. *Journal of Climate*, **24**, 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

Riegger, J., and M. J. Tourian, 2014: Characterization of Runoff-Storage Relationships by Satellite Gravimetry and Remote Sensing. *Water Resource Research*, **50**, 3444–3466, doi:10.1002/2013WR013847.

Riegger, J., M. J. Tourian, B. Devaraju, and N. Sneeuw, 2012: Analysis of GRACE uncertainties by hydrological and hydro-meteorological observations. *Journal of Geodynamics*, **59**, 16–27, doi:10.1016/j.jog.2012.02.001.

Rienecker, M. M., and Coauthors, 2011: MERRA – NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, **24**, 3624–3648, doi: 10.1175/JCLI-D-11-00015.1.

Roads, J. O., S. C. Chen, A. K. Guetter, and K. P. Georgakaos, 1994: Large-Scale Aspects of the United States Hydrologic Cycle. *Bulletin of the American Meteorological Society*, **75**, 1589–1610, doi:10.1175/1520-0477(1994)075<1589:LSAOTU>2.0.CO;2.

Roads, J. O., M. Kanamitsu, and R. E. Stewart, 2002: CSE water and energy budgets in the NCEP-DOE Reanalysis II. *Journal of Hydrometeorology*, **3**, 248–277, doi: 10.1175/1525-7541(2002)003<0227:CWAEBI>2.0.CO;2.

Robertson, F., M. Bosilovich, J. Maryland, and T. Miller, 2011: The Effect of Satellite Observing System Changrs on MERRA Water and Energy Fluxes. *Journal of Climate*, **24**, 5197–5217, doi:10.1175/2011JCLI4227.1.

Rodell, M., J. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004a: Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophysical Research Letters*, **31**, L20 504, doi:10.1029/2004GL020873.

Rodell, M., and J. S. Famiglietti, 1999: Detectability of variations in continental water storage from satellite observation of time dependent gravity field. *Water Resource Research*, **35**, 2705–2723, doi:10.1029/1999WR900141.

Rodell, M., I. Velicogna, and J. Famiglietti, 2009: Satellite-based estimates of groundwater depletion in India. *Nature*, **460**, 999–1002, doi:10.1038/nature08238.

Rodell, M., and Coauthors, 2004b: The Global Land Data Assimilation System. *Bulletin of the American Meteorological Society*, **85**, 381–394, doi:10.1175/BAMS-85-3-381.

Rodell, M., and Coauthors, 2015: The Observed State of the Water Cycle in the Early 21st Century. *Journal of Climate*, **28**, 8289–8318, doi:10.1175/JCLI-D-14-00555.1.

Rudolf, B., and F. Rubel, 2005: Global precipitation. *Observed Global Climate*, M. Hantel, Ed., Landolt-Börnstein - Group V Geophysics, Vol. 6, Springer Berlin Heidelberg, chap. 11, 1–53.

Rudolf, B., and U. Schneider, 2005: Calculation of gridded precipitation data for the global land-surface using in-situ gauge observations. *Proceedings of the 2nd workshop of the international precipitation working group IPWG, Monterey, October 2004*, 231–247.

Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057, doi:10.1175/2010BAMS3001.1.

Sahoo, A. K., M. Pan, T. J. Troy, R. K. Vinukollu, J. Sheffield, and E. F. Wood, 2011: Reconciling the global terrestrial water budget using satellite remote sensing. *Remote Sensing of Environment*, **115**, 1850–1865, doi:10.1016/j.rse.2011.03.009.

Salvucci, G. D., and P. Gentine, 2013: Emergent relation between surface vapor conductance and relative humidity profiles yields evaporation rates from weather data. *Proceedings of the National Academy of Sciences*, **110**, 6287–6291, doi:10.1073/pnas.1215844110.

Santer, B. D., and Coauthors, 2004: Identification of anthropogenic climate change using a second-generation reanalysis. *Journal of Geophysical Research D: Atmospheres*, **109**, D21 104, doi:10.1029/2004JD005075.

Santhi, C., J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck, 2001: Validation of the SWAT model on a large river basin with point and nonpoint sources. *Journal of the American Water Resources Association*, **37**, 1169–1188, doi:10.1111/j.1752-1688.2001.tb03630.x.

Sasgen, I., Z. Martinec, and K. Fleming, 2006: Wiener optimal filtering of GRACE data. *Studia Geophysica et Geodaetica*, **50**, 499–508.

Schlosser, C. A., and P. R. Houser, 2007: Assessing a Satellite-Era Perspective of the Global Water Cycle. *Journal of Climate*, **20**, 1316–1338, doi:10.1175/JCLI4057.1.

Schmidt, R., F. Flechtner, U. Meyer, K.-H. Neumayer, C. Dahle, R. König, and J. Kusche, 2008: Hydrological Signals Observed by the GRACE Satellites. *Surveys in Geophysics*, **29**, 319–334.

Schmidt, R., and Coauthors, 2006: GRACE observations of changes in continental water storage. *Global and Planetary Change*, **50**, 112–126, doi:10.1016/j.gloplacha.2004.11.018.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, M. Ziese, and B. Rudolf, 2014: GPCC's new land surface precipitation climatology based on quality-controlled in situ data and its role in quantifying the global water cycle. *Theoretical and Applied Climatology*, **115**, 15–40, doi:10.1007/s00704-013-0860-x.

Schneider, U., T. Fuchs, A. Meyer-Christoffer, and B. Rudolf, 2008: Global precipitation analysis products of the GPCC. Tech. rep., Deutscher Wetterdienst.

Seneviratne, S. I., P. Viterbo, D. Lüthi, and C. Schär, 2004: Inferring changes in terrestrial water storage using ERA-40 reanalysis data: The Mississippi River basin. *Journal of Climate*, **17**, 2039–2057, doi:10.1175/1520-0442(2004)017<2039:ICITWS>2.0.CO;2.

Seo, K. W., C. R. Wilson, S. C. Han, and D. E. Waliser, 2008: Gravity recovery and climate experiment (GRACE) alias error from ocean tides. *Journal of Geophysical Research*, **113**, B03 405, doi:10.1029/2006JB004747.

Sheffield, J., C. R. Ferguson, T. J. Troy, E. F. Wood, and M. F. McCabe, 2009: Closing the terrestrial water budget from satellite remote sensing. *Geophysical Research Letters*, **36**, L07 403, doi:10.1029/2009GL037338.

Sheffield, J., G. Goteti, E. F. Wood, J. Sheffield, G. Goteti, and E. F. Wood, 2006: Development of a 50-Year High-Resolution Global Dataset of Meteorological Forcings for Land Surface Modeling. *Journal of Climate*, **19**, 3088–3111, doi: 10.1175/JCLI3790.1.

Shiklomanov, A. I., R. B. Lammers, and C. J. Vörösmarty, 2002: Widespread decline in hydrological monitoring threatens Pan-Arctic Research. *Eos, Transactions American Geophysical Union*, **83**, 13, doi:10.1029/2002EO000007.

Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2006: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, **110**, 25–35.

Simon, D., 2010: Kalman Filtering with State Constraints: A Survey of Linear and Nonlinear Algorithms. *IET Control Theory & Applications*, **4**, 1303–1318, doi: 10.1049/iet-cta.2009.0032.

Simon, D., and T. Chia, 2002: Kalman Filtering with State Constraints. *IEEE Transactions on Aerospace and Electronic Systems*, **39**, 128–136, doi:10.1109/7.993234.

Sivapalan, M., and Coauthors, 2003: IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, **48**, 857–880, doi:10.1623/hysj.48.6.857.51421.

Smith, E., and Coauthors, 2007: International Global Precipitation Measurement (GPM) Program and Mission: An Overview. *Measuring Precipitation From Space*, Advances In Global Change Research, Vol. 28, Springer Netherlands, 611–653, doi:10.1007/978-1-4020-5835-6_48.

Sneeuw, N., C. Lorenz, B. Devaraju, M. J. Tourian, J. Riegger, H. Kunstmann, and A. Bárdossy, 2014: Estimating Runoff Using Hydro-Geodetic Approaches. *Surveys in Geophysics*, **4**, 1303–1318, doi:10.1007/s10712-014-9300-4.

Southwood, D., 1999: Introducing the 'Living Planet' programme – the ESA strategy for earth observation. Tech. rep., European Space Agency, ESA Publications Division, Noordwijk, The Netherlands.

Swenson, S., and J. Wahr, 2006: Estimating large-scale precipitation minus evapotranspiration from GRACE satellite gravity measurements. *Journal of Hydrometeorology*, **7**, 252–270, doi:10.1175/JHM478.1.

Syed, T. H., J. S. Famiglietti, and D. P. Chambers, 2009: GRACE-Based Estimates of Terrestrial Freshwater Discharge from Basin to Continental Scales. *Journal of Hydrometeorology*, **10**, 22–40, doi:10.1175/2008JHM993.1.

Syed, T. H., J. S. Famiglietti, J. Chen, M. Rodell, S. I. Seneviratne, P. Viterbo, and C. R. Wilson, 2005: Total basin discharge for the Amazon and Mississippi River basins from GRACE and a land-atmosphere water balance. *Geophysical Research Letters*, **32**, L24 404, doi:10.1029/2005GL024851.

Syed, T. H., J. S. Famiglietti, V. Zlotnicki, and M. Rodell, 2007: Contemporary estimates of Pan-Arctic freshwater discharge from GRACE and reanalysis. *Geophysical Research Letters*, **34**, L19 404, doi:10.1029/2007GL031254.

Sylla, M. B., E. Coppola, L. Mariotti, F. Giorgi, P. M. Ruti, A. Dell'Aquila, and X. Bi, 2010: Multiyear simulation of the African climate using a regional climate model (RegCM3) with the high resolution ERA-interim reanalysis. *Climate Dynamics*, **35**, 231–247, doi:10.1007/s00382-009-0613-9.

Szeto, K. K., H. Tran, M. D. MacKay, R. Crawford, and R. E. Stewart, 2008: The MAGS Water and Energy Budget Study. *Journal of Hydrometeorology*, **9**, 96–115, doi:10.1175/2007JHM810.1.

Tapley, B. D., 2004: GRACE Measurements of Mass Variability in the Earth System. *Science*, **305**, 503–505, doi:10.1126/science.1099192.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, **106**, 7183–7192, doi:10.1029/2000JD900719.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. *Bulletin of the American Meteorological Society*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tourian, M. J., 2013: Application of spaceborne geodetic sensors for hydrology. Ph.D. thesis, University of Stuttgart, Germany.

Tourian, M. J., J. Riegger, N. Sneeuw, and B. Devaraju, 2011: Outlier identification and correction for GRACE aggregated data. *Studia Geophysica et Geodaetica*, **55**, 627–640, doi:10.1007/s11200-009-9007-z.

Tourian, M. J., N. Sneeuw, and A. Bárdossy, 2013: A quantile function approach to discharge estimation from satellite altimetry (ENVISAT). *Water Resources Research*, **49**, 1–13, doi:10.1002/wrcr.20348.

Trenberth, K. E., 2007: Observations: Atmospheric surface and climate change. *Climate Change 2007: The Physical Science Basis*, S. et al. Solomon, Ed., Cambridge University Press, 235–336.

Trenberth, K. E., 2011: Changes in precipitation with climate change. *Climate Research*, **47**, 123–138, doi:10.3354/cr00953.

Trenberth, K. E., and J. T. Fasullo, 2013: Regional Energy and Water Cycles: Transports from Ocean to Land. *Journal of Climate*, **26**, 7837–7851, doi:10.1175/JCLI-D-13-00008.1.

Trenberth, K. E., J. T. Fasullo, and J. Mackaro, 2011: Atmospheric Moisture Transports from Ocean to Land and Global Energy Flows in Reanalyses. *Journal of Climate*, **24**, 4907–4924, doi:10.1175/2011JCLI4171.1.

Trenberth, K. E., L. Smith, T. Qian, A. Dai, and J. Fasullo, 2007: Estimates of the global water budget and its annual cycle using observational and model data. *Journal of Hydrometeorology*, **8**, 758–769, doi:10.1175/JHM600.1.

Trenberth, K. E. and Asrar, G. R., 2014: Challenges and Opportunities in Water Cycle Research: WCRP Contributions. *Surveys in Geophysics*, **35**, 515–532, doi:10.1007/s10712-012-9214-y.

Troy, T. J., J. , and E. F. Wood, 2011: Estimation of the Terrestrial Water Budget over Northern Eurasia through the Use of Multiple Data Sources. *Journal of Climate*, **24**, 3272–3293, doi:10.1175/2011JCLI3936.1.

Uppala, S. M., P. W. Kålberg, A. J. Simmons, U. Andrae, V. D. C. Bechtold, M. Fiorino, J. K. Gibson, and J. H. Et al., 2005: The ERA-40 Re-Analysis. *Quarterly Journal of the Royal Meteorological Society*, **131**, 2961–3012, doi:10.1256/qj.04.176.

Van der Ent, R. J., H. H. G. Savenije, B. Schaefli, and S. C. Steele-Dunne, 2010: Origin and fate of atmospheric moisture over continents. *Water Resources Research*, **46**, W09 525, doi:10.1029/2010WR009127.

Vergnes, J. P., and B. Decharme, 2012: A simple groundwater scheme in the TRIP river routing model: global off-line evaluation against GRACE terrestrial water storage estimates and observed river discharges. *Hydrology & Earth System Sciences Discussions*, **9**, 3889–3908, doi:10.5194/hess-16-3889-2012.

Vinukollu, R. K., E. F. Wood, C. R. Ferguson, and J. B. Fisher, 2011: Global estimates of evapotranspiration for climate studies using multi-sensor remote sensing data: Evaluation of three process-based approaches. *Remote Sensing of Environment*, **115**, 801–823, doi:10.1016/j.rse.2010.11.006.

Vörösmarty, C. J., and Coauthors, 2001: Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union*, **82**, 54–58, doi:10.1029/01EO00031.

Vörösmarty, C. J., and Coauthors, 2010: Global threats to human water security and river biodiversity. *Nature*, **467**, 555–561, doi:10.1038/nature09440.

Wahr, J., M. Molenaar, and F. Bryan, 1998: The Time-Varibility of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *Journal of Geophysical Research*, **103**, 30 205–30 230, doi:10.1029/98JB02844.

Wang, W., P. Xie, S.-H. Yoo, Y. Xue, A. Kumar, and X. Wu, 2010: An assessment of the surface climate in the NCEP climate forecast system reanalysis. *Climate Dynamics*, **37**, 1601–1620, doi:10.1007/s00382-010-0935-7.

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

Wessel, P., and W. H. F. Smith, 1991: Free software helps map and display data. *Eos, Transactions American Geophysical Union*, **72**, 441–441, doi:10.1029/90EO00319.

Willmott, C. J., 1985: Climatology of the seasonal terrestrial water cycle. *International Journal of Climatology*, **5**, 589–606, doi:10.1002/joc.3370050602.

Wilson, W., P. Bates, D. Alsdorf, B. Forsberg, M. Horritt, J. Melack, F. Frappart, and J. Famiglietti, 2007: Modeling large-scale inundation of Amazonian seasonally flooded wetlands. *Geophysical Research Letters*, **34**, L15 404, doi:10.1029/2007GL030156.

Xie, P., and P. A. Arkin, 1996: Analyses of Global Monthly Precipitation Using Gauge Observations, Satellite Estimates, and Numerical Model Predictions. *Journal of Climate*, **9**, 840–858, doi:10.1175/1520-0442(1996)009<0840:AOGMPU>2.0.CO;2.

Xie, P., and P. A. Arkin, 1997: Global Precipitation: A 17-Year Monthly Analysis Based on Gauge Observations, Satellite Estimates, and Numerical Model Outputs. *Bulletin of the American Meteorological Society*, **78**, 2539–2558, doi:10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2.

Xue, Y., B. Huang, Z.-Z. Hu, A. Kumar, C. Wen, D. Behringer, and S. Nadiga, 2010: An assessment of oceanic variability in the NCEP climate forecast system reanalysis. *Climate Dynamics*, **37**, 2511–2539, doi:10.1007/s00382-010-0954-4.

Yeh, P. J.-F., and J. S. Famiglietti, 2008: Regional terrestrial water storage change and evapotranspiration from terrestrial and atmospheric water balance computations. *Journal of Geophysical Research*, **113**, D09 108, doi:10.1029/2007JD009045.

Zaitchik, B. F., M. Rodell, and F. Olivera, 2010: Evaluation of the Global Land Data Assimilation System using global river discharge data and a source-to-sink routing scheme. *Water Resources Research*, **46**, W06 507, doi:10.1029/2009WR007811.

Zakharova, E. A., A. V. Kouraev, A. Cazenave, and F. Seyler, 2006: Amazon River Discharge estimated from TOPEX/Poseidon altimetry. *Comptes Rendus Geoscience*, **338**, 188–196, doi:10.1016/j.crte.2005.10.003.

Zhou, X., Y. Zhang, Y. Wang, H. Zhang, J. Vaze, L. Zhang, Y. Yang, and Y. Zhou, 2012: Benchmarking global land surface models against the observed mean annual runoff from 150 large basins. *Journal of Hydrology*, **470-471**, 269–279, doi:10.1016/j.jhydrol.2012.09.002.