New Estimates of Variations in Water Flux and Storage over Europe Based on Regional (Re)Analyses and Multisensor Observations

Anne Springer, Jürgen Kusche, Kerstin Hartung, Christan Ohlwein, Laurent Longuevergne

New Estimates of Variations in Water Flux and Storage over Europe Based on Regional (Re)Analyses and Multisensor Observations

ANNE SPRINGER

Institute of Geodesy and Geoinformation, Bonn University, and Centre for High-Performance Scientific Computing in Terrestrial Systems, Geoverbund ABC/J, Bonn, Germany

JÜRGEN KUSCHE

Institute of Geodesy and Geoinformation, Bonn University, Bonn, Germany

KERSTIN HARTUNG

Meteorological Institute, Stockholm University, Stockholm, Sweden

CHRISTAN OHLWEIN

Hans-Ertel Centre for Weather Research, Climate Monitoring Branch, Meteorological Institute, Bonn University, Bonn, Germany

LAURENT LONGUEVERGNE

CNRS, UMR 6118, Geosciences Rennes, Rennes, France

(Manuscript received 7 March 2014, in final form 10 July 2014)

ABSTRACT

Precipitation minus evapotranspiration, the net flux of water between the atmosphere and Earth's surface, links atmospheric and terrestrial water budgets and thus represents an important boundary condition for both climate modeling and hydrological studies. However, the atmospheric–terrestrial flux is poorly constrained by direct observations because of a lack of unbiased measurements. Thus, it is usually reconstructed from atmospheric reanalyses. Via the terrestrial water budget equation, water storage estimates from the Gravity Recovery and Climate Experiment (GRACE) combined with measured river discharge can be used to assess the realism of the atmospheric–terrestrial flux in models. In this contribution, the closure of the terrestrial water budget is assessed over a number of European river basins using the recently reprocessed GRACE release 05 data, together with precipitation and evapotranspiration from the operational analyses of high-resolution, limited-area NWP models [Consortium for Small-Scale Modelling, German version (COSMO-DE) and European version (COSMO-EU)] and the new COSMO 6-km reanalysis (COSMO-REA6) for the European Coordinated Regional Climate Downscaling Experiment (CORDEX) domain. These closures are compared to those obtained with global reanalyses, land surface models, and observation-based datasets. The spatial resolution achieved with the recent GRACE data allows for better evaluation of the water budget in smaller river basins than before and for the identification of biases up to 25 mm month^{-1} in the different products. Variations of deseasoned and detrended atmospheric–terrestrial flux are found to agree notably well with flux derived from GRACE and discharge data with correlations up to 0.75. Finally, bias-corrected fluxes are derived from various data combinations, and from this, a 20-yr time series of catchment-integrated water storage variations is reconstructed.

1. Introduction

Redistribution and phase changes of water have a strong impact on Earth's climate. The hydrological cycle is linked to the energy cycle via fluxes, such as evapotranspiration E and precipitation P. Water evaporates

Corresponding author address: Anne Springer, Astronomical, Physical and Mathematical Geodesy Group, Institute of Geodesy and Geoinformation (IGG), Bonn University, Nußallee 17, 53115 Bonn, Germany.
E-mail: springer@geod.uni-bonn.de

DOI: 10.1175/JHM-D-14-0050.1

from the oceans and the land surface because of solar radiation, thereby cooling them, and is transported through the atmosphere, released as precipitation, and finally returns to the oceans via surface runoff and river discharge R.

An "intensification" of the water cycle (Huntington 2006) should be associated with observable positive rates of continental P and E. In fact, an increase of land precipitation has been observed in the higher latitudes, that is, over North America and Eurasia (Trenberth 2011). For these regions, changes in P usually implicate an increase of extreme precipitation events (Coumou and Rahmstorf 2012) leading to more frequent and severe flooding (Hall et al. 2013). On the other hand, observed positive trends in E have been attributed to rising temperatures and radiation and are found to be constrained by moisture supply (Jung et al. 2010). Precipitation minus evapotranspiration ($P - E$) controls the amount of water stored in soils, groundwater, and surface water bodies. With freshwater representing a major natural resource, monitoring systems are indispensable in order to manage the various components of water consumption. Quantifying spatiotemporal variations in fluxes and storages is thus a necessity in order to assess their impact on local and global climatic, ecological, and environmental conditions. For all these reasons, long, unbiased, and consistent time series of fluxes P, E, R, and of total water storage (TWS) are required.

At all spatial scales (i.e., grid cell, catchment, and continent) water storage changes ΔS are linked to P, E, and R via the terrestrial water budget equation $\Delta S = P - E - R$.

These terms can be evaluated from different datasets. Since 2002, the Gravity Recovery and Climate Experiment (GRACE) twin-satellite mission enables the direct observation of total water storage variations from space. At catchment scale, river discharge is usually derived from gauging stations, and P and E are either obtained from observation-based datasets or from numerical weather prediction (NWP) model analyses or reanalyses. However, the representation of P and E in global reanalyses is subject to large uncertainties (e.g., Lorenz and Kunstmann 2012).

At the core of this study is the closure of the water budget at river basin scale, in order to assess the realism of the representation of the P and E fluxes in NWP models. In particular, high-resolution output fields from the Consortium for Small-Scale Modelling, German version (COSMO-DE) and European version (COSMO-EU), which are analysis runs of the nonhydrostatic regional atmospheric COSMO model developed by the Deutscher Wetterdienst (DWD), are evaluated and compared to

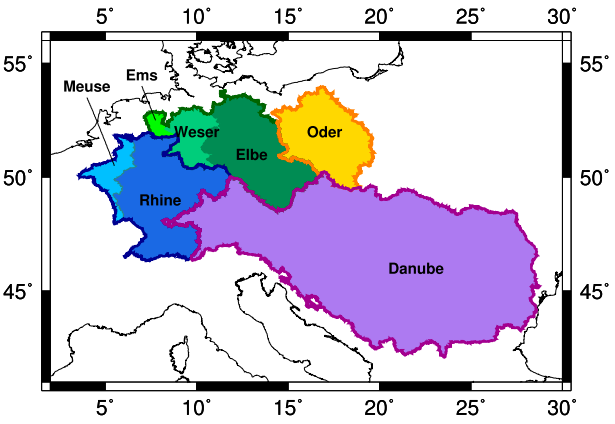

FIG. 1. Central European river basins: Rhine–Meuse, Ems–Weser–Elbe, Oder, and Danube.

global models and observation-based datasets. Our study area covers the catchments of the Rhine–Meuse (189 780 km^2 combined), Ems–Weser–Elbe (178 040 km^2 combined), Oder (109 560 km^2), and Danube (807 000 km^2) as illustrated in Fig. 1. First, the water budget equation is solved for the atmospheric–terrestrial fluxes $P - E$ and E. Then, the residuals $\epsilon = \Delta S - (P - E - R)$ of the catchment-averaged water budget are quantified and an error model is derived for individual combinations of datasets. Finally, the storage term S is assessed through comparing time-integrated, error-corrected $P - E - R$ to the reprocessed GRACE release 05a (RL05a) data.

Prior to the launch of the GRACE mission, investigations of catchment-scale water storage were based on modeling (e.g., Döll et al. 2003), but with the availability of the GRACE products, comparisons between terrestrial water storage derived from hydrological models and those measured from space were enabled (e.g., Tapley et al. 2004). Subsequently, a large number of studies began to use GRACE in combination with hydrological models for validation, calibration, or assimilation purposes (e.g., Zaitchik 2008; Werth and Güntner 2010).

Closure of the water budget in central European river basins was previously investigated by Hirschi et al. (2006) and Seitz et al. (2008). Hirschi et al. (2006) found the agreement of GRACE-derived TWS and integrated European Centre for Medium-Range Weather Forecasts (ECMWF) operational moisture flux convergence data to be rather poor because of the very limited resolution of early GRACE releases. Seitz et al. (2008) identified a phase shift of 30 days between integrated moisture flux convergence data from National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) and GRACE

for the years after 2006. Compared to Hirschi et al. (2006), they obtained better agreements between GRACE and model data because of reprocessed GRACE solutions and a larger study area. In particular, the Danube, the largest of the European river basins, was subject to several investigations in the context of water budget studies: for example, Ramilien et al. (2006) and Rodell et al. (2011) used GRACE data in conjunction with observed P and R in order to estimate E, which was then compared to results from different atmospheric and land surface models. Rodell et al. (2011) confirmed the benefit of GRACE-based E estimates for the assessment of the phase and amplitude of the annual cycle, especially for regions where model-derived E estimates disagree.

The global NWP models and observation-based datasets, which we employ as a reference for comparison and validation purposes in this study, have been widely used in the context of water budget studies. For example, Hirschi et al. (2006), Rodell et al. (2011), and Syed et al. (2009) worked with the ECMWF reanalysis and the Modern-Era Retrospective Analysis for Research and Applications (MERRA). In addition, Fersch et al. (2012) included high-resolution regional atmospheric simulations produced with the Weather Research and Forecasting (WRF) Model for six different climatic regions. They found dynamic downscaling of global atmospheric fields generally useful for reducing the bias in these fields, but time series of basin-averaged TWS change estimates did not improve with respect to the global models. Outputs of different models and observational datasets were merged using assimilation techniques by Pan et al. (2012) to derive trends and annual water budget terms over 32 major river basins. Furthermore, the quality of observation-based P and E datasets has been assessed in several studies. For instance, the Global Precipitation Climatology Centre (GPCC) dataset and the European daily high-resolution gridded dataset (E-OBS) are discussed in detail by Roebeling et al. (2012). Acquisition of evapotranspiration fields on an observational basis is considered more challenging (Wang and Dickinson 2012). An overview and evaluation of recently derived global evapotranspiration datasets can be found, for example, in Mueller et al. (2011) and Jiménez et al. (2011), including a dataset of upscaled FLUXNET data derived by Jung et al. (2011), which is also used in our study. Furthermore, Long et al. (2014) used the three-cornered hat method in order to compute uncertainties of evapotranspiration from three different methods, including the water budget approach.

Continental-scale atmospheric fluxes of P, E, and $P - E$ produced by three global reanalyses, including MERRA and Interim ECMWF Re-Analysis (ERA-Interim), were assessed by Lorenz and Kunstmann (2012). They determined inconsistent continental and oceanic $P - E$ for MERRA, whereas ERA-Interim showed a good closure of the water balance. Miralles et al. (2011) provide P, E, and $P - E$ estimates from a combination of remote sensing observations.

In this contribution, we focus on the atmospheric part of the water cycle. However, we found moisture flux convergence derived from the high-resolution COSMO models as being very noisy; therefore, we directly evaluate P and E fields. We will demonstrate that the COSMO model shows superior skills in closing the water budget over Europe compared to the global models. Until now, most water budget studies have concentrated on large river basins and/or the annual cycle of E, $P - E$, or S only; here, we will include month-to-month variability and trends. Moreover, we will evaluate the water budget equation for smaller European river basins. We find this possible now thanks to the increased accuracy of the reprocessed GRACE RL05a data and to a consistent and thorough postprocessing of GRACE and flux deficit data. This contribution is organized as follows. In section 2, we introduce the investigated datasets and the methods. The results are presented in section 3. First, the characteristics of the four components of the water budget equation are assessed. Then, the water budget equation is solved for $P - E$ and compared to the independent GRACE and discharge observations. As the observation and modeling of E is particularly challenging, the derivation of E via the water budget equation is evaluated. In the next step, the residuals of the water budget are quantified and error models for the individual datasets of $P - E$ are derived. Finally, error-corrected fluxes are computed and used in order to extend the GRACE-derived time series of TWS change and TWS (since 2002) backward in time (to 1990).

2. Data and methods

a. Data

Assessing the closure of the terrestrial water budget requires that datasets of P, E, R, and ΔS are combined. Evapotranspiration is related to latent heat flux Q_L via the relation

$$E = \lambda^{-1} Q_L, \tag{1}$$

with the latent heat of vaporization for water $\lambda = 2.5 \times 10^3$ kJ kg^{-1}. Table 1 provides an overview of the datasets that are described in the following.

TABLE 1. Components of the water budget are available from the datasets with different time frame and horizontal resolution. GRACE data are available since 2003 with gaps in June 2003; January, June, and November 2011; and May and October 2012. GRDC provides river discharge data until 2009 for Ems, Weser, Elbe, and Oder and until 2010 for Meuse, Rhine, and Danube. BfG provides discharge data until 2012 for all German rivers including the Meuse and until 2010 for the Danube.

Dataset	Type (version)	Horizontal resolution	Time frame	Parameter(s)
COSMO-DE	Analysis	0.025° × 0.025°	Jan 2007–present	P, E
COSMO-EU	Analysis	0.0625° × 0.0625°	Jan 2006–present	P, E
CORDEX	Reanalysis	0.055° × 0.055°	Jan–Dec 2011	P, E
ERA-Interim	Reanalysis	1.5° × 1.5°	Jan 1979–present	P, E
MERRA	Reanalysis	0.67° × 0.5°	Jan 1980–present	P, E
E-OBS	Observations (v8.0)	0.25° × 0.25°	Jan 1950–Dec 2012	P
GPCC	Observations (v6.0)	0.5° × 0.5°	Jan 1901–Dec 2010	P
MPI	Observations	0.5° × 0.5°	Jan 1982–Dec 2011	E
GRACE	Observations (RL05a)	—	Jan 2003–present	S, ΔS
GRDC	Observations	—	Jan 1857–Dec 2009 (2010)	R
BfG	Observations	—	Jan 1951–Dec 2010 (2012)	R

1) GRACE

The primary objective of the National Aeronautics and Space Administration (NASA)–German Aerospace Center (DLR) GRACE mission, launched in 2002, is the measurement of Earth's time-variable gravity field. In a seminal paper, Wahr et al. (1998) anticipated the benefits and limitations of the mission for resolving water storage variations from space; this was later confirmed by a great number of applications in large-scale hydrology as described, for example, in the review by Güntner (2008). With GRACE, a dual one-way K-band microwave ranging system measures the distance and its rate between two satellites in coplanar orbit, with accuracy of a few micrometers. From the distance change and from complementary tracking of the individual spacecraft positions with GPS, Earth's changing gravity field is inferred on a monthly basis. Many users are interested in the temporal variations due to changes in Earth's terrestrial hydrosphere, which are obtained in a straightforward way since other gravity signals like Earth and ocean tides, nontidal ocean mass variability, and atmospheric pressure changes have already been removed in the gravity recovery processing by using so-called background models.

With release 05 (Dahle et al. 2013), the GeoForschungsZentrum (GFZ) Helmholtz-Zentrum in Potsdam provides improved (Chambers and Bonin 2012) monthly gravity solutions in the form of spherical harmonic coefficients starting in January 2003. Release 05 was recently reprocessed because of some shortcomings in the satellite orbit determination, and the new dataset, which is applied in this contribution, has been distributed as release 05a. We will compare the GFZ solution to those generated at the Jet Propulsion Laboratory (JPL), California Institute of Technology, and the Center for Space Research (CSR), The University of Texas at Austin. In addition, we will make use of error estimates for GFZ coefficients in the form of so-called calibrated errors; these do account for measurement and processing errors, but they neglect error correlations and cannot account for errors introduced by background model imperfections.

The same postprocessing is performed for each of the three GRACE solutions. First, the gravity effect resulting from ongoing isostatic adjustment of Earth's crust and mantle, in response to past ice age deglaciation, is removed from the harmonic coefficients using the model of Klemann and Martinec (2011). Then, since coefficients of degree 1 cannot be measured by GRACE, these are substituted from a time series of geocenter motion provided by Rietbroek et al. (2012). Furthermore, the zonal $c_{2,0}$ coefficient, related to the flattening of Earth, is known to be not well determined by GRACE. Therefore, it is replaced using a time series derived from satellite laser ranging (SLR) by Cheng and Tapley (2004).

2) WGHM

As an alternative to GRACE, the WaterGAP Global Hydrology Model (WGHM; Döll et al. 2003) is used to obtain TWS anomalies. WGHM represents a global conceptual model of the terrestrial water cycle, with the objective of providing indicators of freshwater availability worldwide. All storage compartments relevant for calculating TWS are modeled, including surface water and groundwater. In WGHM, the water balance is solved at daily time steps with a spatial resolution of 0.5°. Monthly precipitation forcing data from GPCC are disaggregated to daily rain rates. The computation of daily potential evapotranspiration is based on Priestley and Taylor (1972). The model has been calibrated against long-term discharge. Special emphasis is paid to anthropogenic effects like water consumption.

3) TWS CHANGE FROM PAN ET AL. (2012)

A reference dataset for TWS change was provided to us by Ming Pan. Pan et al. (2012) evaluated the components of the water budget by merging a number of different datasets using a data assimilation technique to enforce the water balance. As a prerequisite, they had performed a thorough error investigation. Thus, the estimation of the water budget components is optimized and includes error information derived from confronting independent datasets. Pan et al. (2012) utilized in situ observations, remote sensing retrievals, and model outputs, among other datasets. For TWS change, GRACE data and outputs from the Variable Infiltration Capacity (VIC) model were applied. The study was performed for 32 river basins and for the time span 1984–2006. In our contribution, we use the TWS change estimate provided for the Danube basin.

4) REGIONAL ANALYSES AND REANALYSES

The COSMO-EU and -DE analyses and the COSMO 6-km reanalysis (COSMO-REA6) are based on the COSMO NWP model developed by the DWD. COSMO represents a nonhydrostatic regional atmospheric model, where atmosphere–surface coupling is realized with the soil and vegetation model TERRA_ML. Lateral boundary conditions are obtained for COSMO-EU from the operational global numerical weather prediction model (GME) at DWD and for COSMO-DE from the COSMO-EU model. Continuous data assimilation is performed based on observation nudging (Schraff 1997) with hourly model outputs. COSMO-EU output fields are available to us from 2006, and COSMO-DE outputs are available from 2007 onward.

COSMO-EU covers the eastern Atlantic and Europe with a grid resolution of 0.0625° (~7 km) and 40 vertical layers. COSMO-DE has a higher resolution of 0.025° (~2.8 km) and 50 vertical layers and covers Germany, Switzerland, Austria, and parts of neighboring countries with a total area of about $1300 \times 1200\,\mathrm{km}^2$. The major differences to COSMO-EU are that COSMO-DE resolves deep convection and assimilates surface precipitation rates from the German radar network. Additionally, we investigate the COSMO-based reanalysis COSMO-REA6 for the European domain (EUR-11) of the Coordinated Regional Climate Downscaling Experiment (CORDEX) but with 0.055° (6.2 km) horizontal grid spacing (Bollmeyer et al. 2014, manuscript submitted to *Quart. J. Roy. Meteor. Soc.*). It is currently being developed and processed within the Hans-Ertel Centre for Weather Research and has been completed for the years 2007–12.

5) GLOBAL REANALYSES

For comparison and validation, we use the global ERA-Interim and MERRA. ERA-Interim was initiated in 2006 and covers the period since 1979. It is the successor of the 40-yr ECMWF Re-Analysis (ERA-40) and was designed to improve, for instance, the representation of the hydrological cycle, the quality of the stratospheric circulation and the handling of biases and changes in the observing system (Berrisford et al. 2011). A sequential data assimilation scheme with 12-hourly analysis cycles is performed, and every 3 h output fields of the 2D variables are produced (Dee et al. 2011). In the vertical dimension, 37 levels are considered. Synoptic monthly means of precipitation and latent heat flux evaluated in this study have a resolution of $1.5° \times 1.5°$.

MERRA is from NASA's Global Modeling and Assimilation Office (GMAO) and covers the period from 1979 until today. The Goddard Earth Observing System, version 5 (GEOS-5), is applied for data assimilation, with special attention paid to improving the hydrological cycle (Rienecker et al. 2011). Uncertainties in precipitation and high-frequency variability are reduced (Lucchesi 2012), rendering the model more adequate for climate studies. We use monthly output products that are archived and distributed by the Goddard Earth Sciences Data and Information Services Center (DISC). The horizontal grid resolution is $\delta\lambda = 0.67°$ in longitude and $\delta\phi = 0.5°$ in latitude; 72 vertical levels are considered.

6) OFFLINE LAND SURFACE MODEL

Since MERRA's land surface fluxes are known to contain errors, we also consider its land component from the revised, land-only, offline version, MERRA-Land (Reichle et al. 2011). MERRA is prone to errors in the intensity of precipitation, resulting from the atmospheric general circulation model and the land surface model itself (Reichle et al. 2011). MERRA-Land is driven with hourly atmospheric fields from MERRA, with the precipitation forcings corrected using an observation-based product of the Global Precipitation Climatology Project (GPCP). Furthermore, MERRA-Land contains an updated catchment land surface model with revised model parameters. These changes were shown to lead to improved land surface fluxes (Reichle et al. 2011).

7) OBSERVATIONAL DATASETS

In this study, we compare analysis and reanalysis results to three gridded, observation-based datasets. Global monthly latent heat flux grids are provided by the Max Planck Institute (MPI) for Biogeochemistry (Jung et al. 2011) at a resolution of $0.5° \times 0.5°$. They used

a machine learning approach to upscale observations from FLUXNET together with meteorological and remote sensing observations. FLUXNET is a global network of more than 570 tower sites (200 stations in Europe) where the eddy covariance method is applied to assess atmospheric variables such as latent heat flux.

Second, we use a precipitation product from the GPCC, established in 1989 at the DWD. It is based on 67 200 rain gauge stations worldwide. Monthly total precipitation on 0.5° × 0.5° grids was downloaded from the GPCC website. We use the GPCC Full Data Reanalysis, version 6.0, the most accurate GPCC in situ precipitation reanalysis dataset (Schneider et al. 2011) that covers the period 1901–2010. Using GPCC data for assessing the global water cycle is discussed in Schneider et al. (2014). Finally, we consider the European gridded rain gauge dataset E-OBS from the European Union 6th Framework Programme (EU-FP6) project Ensemble-Based Predictions of Climate Change and Their Impacts (ENSEMBLES) and from the data providers in the European Climate Assessment & Dataset (ECA&D) project (www.ecad.eu). Daily gridded observational precipitation data cover the European land surface and are available at different resolutions. We use version 7.0 with a grid resolution of 0.25° × 0.25°, which contains data from January 1950 to December 2012, collected from 2316 stations.

8) River discharge data

Here we use monthly discharge data measured at the gauging stations located at the most downstream part of each river provided by the Global Runoff Data Centre (GRDC) and Bundesanstalt für Gewässerkunde (BfG). The BfG time series agree with the GRDC time series but provide more complete and recent information for German rivers. In a few cases, missing data have been inferred from upstream gauging stations. Data are available up to 2012 for Rhine–Meuse, Ems–Weser–Elbe, and Oder and up to 2010 for the Danube.

b. Methods

1) The water budget equation

At all spatial scales, ΔS within a given region is balanced by inflow (P) and outflow (E and R):

$$\Delta S = P - E - R. \tag{2}$$

While Eq. (2) describes the terrestrial water budget, the atmospheric water balance can be expressed by approximating $P - E$ through atmospheric water storage and vertically integrated moisture flux convergence (e.g., Fersch et al. 2012). However, fields of horizontal moisture flux represented in high-resolution atmospheric models such as COSMO are relatively noisy. We found that computing the divergence amplifies noise and leads to results that are contaminated with large errors. Therefore, in this contribution, we use the flux deficit $P - E$ as represented in Eq. (2).

The GRACE solutions determine the temporal resolution with which we evaluate the water budget equation at catchment scale. Monthly mean, spatially averaged TWS from GRACE has to be numerically differentiated in order to approximate ΔS. Performing this in a way consistent with the flux terms involves some subtle issues that are often overlooked in the literature; therefore, we will provide a full account here. Moreover, as an alternative, we will discuss an approach that seeks to consistently time integrate the spatially averaged fluxes.

Monthly averaged ΔS within an arbitrary (catchment) area is related to the monthly accumulated P, E, and R according to

$$\Delta S = \sum_{i=1}^{n_i} P(t_i) - \sum_{j=1}^{n_j} E(t_j) - \sum_{k=1}^{n_k} R(t_k), \tag{3}$$

where t is discretized with daily or subdaily time steps and t_1 and t_n denote the first and the last value of one particular month. The indices i, j, and k correspond to the individual time steps of the P, E, and R datasets. The monthly water budget of a catchment can then be represented as in Eq. (2). Monthly accumulated fluxes are obtained from the data providers or calculated through Eq. (3).

In Eq. (2), ΔS is usually obtained by numerically differentiating GRACE-derived S. Instead, the integral of the right side of the water budget equation can be equated at time τ, after discretizing according to

$$S_F(\tau) = \sum_{i=1}^{m} [P(\tau_i') - E(\tau_i') - R(\tau_i')], \tag{4}$$

where τ_i' represents monthly intervals; $P(\tau_i')$, $E(\tau_i')$, and $R(\tau_i')$ represent the monthly sums; and S_F represents TWS inferred from fluxes. Variable $S_F(\tau)$ equates to TWS, derived from integrating fluxes, at time τ with respect to (unknown) TWS at time τ_1. However, as we will show, spatially averaged $P - E$ is usually contaminated with nonnegligible errors. Thus, we prefer to use error-corrected precipitation minus evapotranspiration ($\widetilde{P - E}$). The estimation and modeling of this error will be described in section 2b(3). TWS from error-corrected fluxes $\widetilde{S_F}$ is then defined as

$$\widetilde{S_F}(\tau) = \sum_{i=1}^{m} [\widetilde{(P - E)}(\tau_i') - R(\tau_i')]. \tag{5}$$

In the following, we will limit our deliberations to error-corrected TWS estimates; therefore, the tilde will be omitted and the left-hand side of Eq. (5) will be abbreviated

by $S_F(\tau)$. Flux-integrated TWS anomalies $S_F^a(\tau)$ are obtained by removing the long-term mean over a time span from τ_a to τ_e, for example, the GRACE period, according to

$$S_F^a(\tau) = S_F(\tau) - \frac{1}{\tau_e - \tau_a} \sum_{i=a}^{e} S_F(\tau_i'). \tag{6}$$

Correspondingly, GRACE-derived TWS anomalies $S^a(\tau)$ for the same time frame are obtained from

$$S^a(\tau) = S(\tau) - \frac{1}{\tau_e - \tau_a} \sum_{i=a}^{e} S(\tau_i'). \tag{7}$$

To be consistent, the long-term mean removed in both approaches must refer to the same time period. Apart from observation errors and incorrect representation of analysis and reanalysis fields, TWS anomalies obtained from integrating fluxes [Eq. (6)] and from GRACE [Eq. (7)] should match. In section 3e, these two approaches are compared.

Both direct and time-integrated evaluations of the water budget equation have advantages and drawbacks. Because of the integration of fluxes in Eq. (4), errors in P, E, and R accumulate. While white noise leads to random walk–type patterns, constant biases introduce trends into S_F^a. Therefore, for evaluating primarily the fluxes, the approach of Eq. (2) is well suited as it employs flux information directly. On the other hand, storage change exhibits much variability, whereas storage evolves at longer time scales. Short-term variations are easier to interpret analyzing TWS instead of TWS change. Furthermore, the GRACE signal-to-noise ratio is better when TWS is evaluated and not its derivative. We will show that, despite unfavorable error propagation, fluxes from NWP models enable us to infer storage variation. This is the key to reconstructing realistic spatially averaged TWS prior to the GRACE period, which may be then spatially disaggregated through land surface models. However, this requires a well-determined error model for the fluxes from storage data collected within the GRACE period. Yet we found that, especially in global reanalyses, biases in $P - E$ may change with time.

2) GENERATING CONSISTENT TIME SERIES

Assessing closure of the water budget, that is, studying the residuals $\epsilon = \Delta S - (P - E - R)$, requires consistent time series of TWS change, P, E, and R. Spatial averages of P and E are simply accumulated over all output grid cells of the target region, defined as the catchment area down to the gauging station where river discharge is measured.

For deriving TWS from the GRACE data products, it is possible to accumulate over gridded TWS maps, as provided online through the NASA Tellus, GFZ International Centre for Global Earth Models (ICGEM), or Groupe de Recherche de Géodésie Spatiale (GRGS) GRACE Plotter websites, in a similar way as with the fluxes. However, it is more consistent to map the original spherical harmonic coefficients into spatial averages for a number of reasons: (i) TWS maps have to be filtered to suppress "striping," but less filtering is required to derive catchment averages, so using gridded products usually involves a loss of information; (ii) filtering reduces amplitude, but counteracting this effect for TWS maps is more difficult than for region averages; and (iii) for TWS maps, usually no consistent error information is provided.

First, the GRACE potential coefficients $c_{n,m}$ and $s_{n,m}$, postprocessed as described in section 2a(1), are reduced by a long-term mean field corresponding to the period τ_a, τ_e to $\Delta c_{n,m}$ and $\Delta s_{n,m}$, and converted to coefficients of surface density $\Delta c_{n,m}^{\sigma}$ and $\Delta s_{n,m}^{\sigma}$ (Wahr et al. 1998) via the relation

$$\begin{pmatrix} \Delta c_{n,m}^{\sigma} \\ \Delta s_{n,m}^{\sigma} \end{pmatrix} = \frac{M}{4\pi R_E^2} \frac{2n+1}{1+k_n'} \begin{pmatrix} \Delta c_{n,m} \\ \Delta s_{n,m} \end{pmatrix}. \tag{8}$$

In Eq. (8), M is Earth's mass, R_E is its radius, and $(2n+1)/(1+k_n')$ accounts for the elastic deformation of the solid Earth under a changing water storage load through applying the load Love numbers k_n' [we use LLNs from Gegout (2014)]. The total mass variation over a given area—which corresponds to TWS since atmospheric mass variations have been removed already—is calculated directly from the coefficients of surface density according to

$$S = 4\pi R_E^2 \sum_{n=1}^{n_{\max}} \sum_{m=0}^{n} \left[f_{n,m}^{c} \sum_{n'=1}^{n} \sum_{m'=0}^{n'} w_{n,m}^{(c,c),n',m'} \Delta c_{n',m'}^{\sigma} + w_{n,m}^{(c,s),n',m'} \Delta s_{n',m'}^{\sigma} + f_{n,m}^{s} \sum_{n'=1}^{n} \sum_{m'=0}^{n'} w_{n,m}^{(s,c),n',m'} \Delta c_{n',m'}^{\sigma} + w_{n,m}^{(s,s),n',m'} \Delta s_{n',m'}^{\sigma} \right]. \tag{9}$$

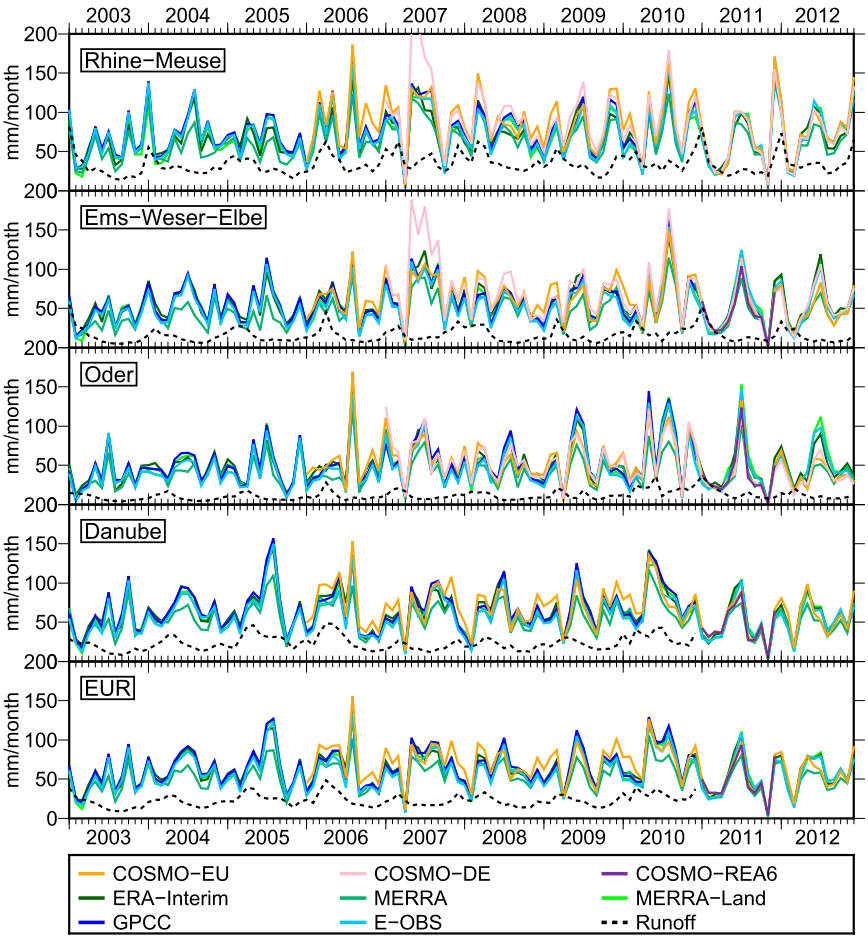

FIG. 2. Monthly total P for different river basins and models. The black dashed line represents monthly river discharge.

The geographical extent of the target region is represented by the shape coefficients $f^c_{n,m}$ and $f^s_{n,m}$ (Wahr et al. 1998). Applying filter coefficients $w^{(...),n',m'}_{n,m}$ is generally deemed necessary to reduce correlated noise contained in the potential coefficients of high degree. In this contribution, we use the nonisotropic DDK4 filter (Kusche 2007) that seeks to decorrelate and smooth the GRACE solutions simultaneously. Finally, mass variations are converted to equivalent water height by dividing Eq. (9) by a reference density of water.

Spatially averaged TWS from maps or from Eq. (9) includes an error due to the limited spectral bandwidth recovered by GRACE and the necessity to smooth the solutions through filtering. The resulting amplitude underestimation, or in some cases, overestimation, has been termed the leakage effect (Klees et al. 2007; Longuevergne et al. 2010). Depending on the mass distribution inside and outside of the basin (i.e., water storage variability and, in coastal regions, ocean mass change), the filter averaging width, and the extent of the region, the signal is either transported out of the target area (leakage out) or leaks into it (leakage in). Therefore, it is imperative to rescale water storage signals derived from GRACE. Rescaling factors are derived as multiplicative corrections separately for every river catchment and for each month through assuming the spatial distribution of $P - E$ as "true TWS change" and monitoring the effect of the truncation of the spherical harmonic expansion and the damping effect of GRACE filtering. Magnitudes of the rescaling factors and the associated TWS variability are discussed in section 3a(4).

The time derivative of storage $\Delta S(\tau)$ for the month τ is obtained from central differences using storage of the following month $S_{\tau+1}$ and the previous month $S_{\tau-1}$ as

$$\Delta S(\tau) = \frac{1}{2}(S_{\tau+1} - S_{\tau-1}). \tag{10}$$

TABLE 2. Mean value μ, annual amplitude A, and the trend of P for the time period 2006–10 for individual river basins are compared. The RMSE is calculated for each model with reference to the mean precipitation of all datasets. The total RMSE is a measure for the variability between all products.

Basin	Model	μ (mm month^{-1})	A (mm month^{-1})	Trend [(mm month^{-1}) yr^{-1}]	RMSE (mm month^{-1})	Total RMSE (mm month^{-1})
Rhine–Meuse	COSMO-EU	90	2	−0.5	16	9
	ERA-Interim	75	9	0.7	6	
	MERRA	69	6	0.4	11	
	MERRA-Land	74	13	−0.4	6	
	GPCC	81	11	−0.4	5	
	E-OBS	76	12	−0.8	5	
Ems–Weser–Elbe	COSMO-EU	66	4	1.2	11	8
	ERA-Interim	62	12	1.0	5	
	MERRA	52	6	1.3	11	
	MERRA-Land	60	15	2.1	5	
	GPCC	61	15	2.1	5	
	E-OBS	56	15	1.9	6	
Oder	COSMO-EU	57	6	1.0	9	8
	ERA-Interim	56	13	3.4	5	
	MERRA	45	4	1.4	13	
	MERRA-Land	53	20	4.2	6	
	GPCC	56	22	3.9	7	
	E-OBS	51	19	3.8	6	
Danube	COSMO-EU	73	9	0.9	11	8
	ERA-Interim	68	20	3.1	5	
	MERRA	57	11	2.0	11	
	MERRA-Land	64	20	3.4	5	
	GPCC	68	21	3.6	6	
	E-OBS	62	18	3.7	5	
EUR	COSMO-EU	73	6	0.8	11	7
	ERA-Interim	67	17	2.5	4	
	MERRA	57	8	1.6	11	
	MERRA-Land	64	19	2.7	5	
	GPCC	68	19	2.8	5	
	E-OBS	62	17	2.8	5	

Unlike with backward or forward difference operators, central difference schemes avoid introducing a phase shift in GRACE storage change time series.

3) ERROR MODEL OF THE FLUX DEFICIT

We find that the residuals of the water budget equation differ most notably for every set of model output fields and for every catchment. Since R is small, ϵ can thus be safely assumed to be associated with errors in $P - E$, which means that we will use $\Delta S + R$ as a reference for fitting an error model to the catchment residuals for each model output.

First, different models for representing the residuals were assessed. We observe that the difference of monthly P datasets appears to be characterized primarily by an offset, and E is generally smoother, mainly characterized by an annual signal. Since we believe that both errors in P and in E contribute to the residual of the water budget, these characteristics suggest the consideration of an offset and the modeling of an annual and a semiannual signal. The offset can be considered using an additive model according to

$$\epsilon = (P - E) - (\Delta S + R) + a + b\sin\left(\frac{2\pi}{T}t\right) + c\cos\left(\frac{2\pi}{T}t\right) + d\sin\left(\frac{2\pi}{T/2}t\right) + e\cos\left(\frac{2\pi}{T/2}t\right), \tag{11}$$

or a multiplicative model according to

$$\epsilon = a(P - E) - (\Delta S + R) + b\sin\left(\frac{2\pi}{T}t\right) + c\cos\left(\frac{2\pi}{T}t\right) + d\sin\left(\frac{2\pi}{T/2}t\right) + e\cos\left(\frac{2\pi}{T/2}t\right), \tag{12}$$

where T is the annual period. The estimated error model ϵ is applied to $P - E$ of Eqs. (3) and (4) in order to obtain error-corrected TWS change and TWS estimates.

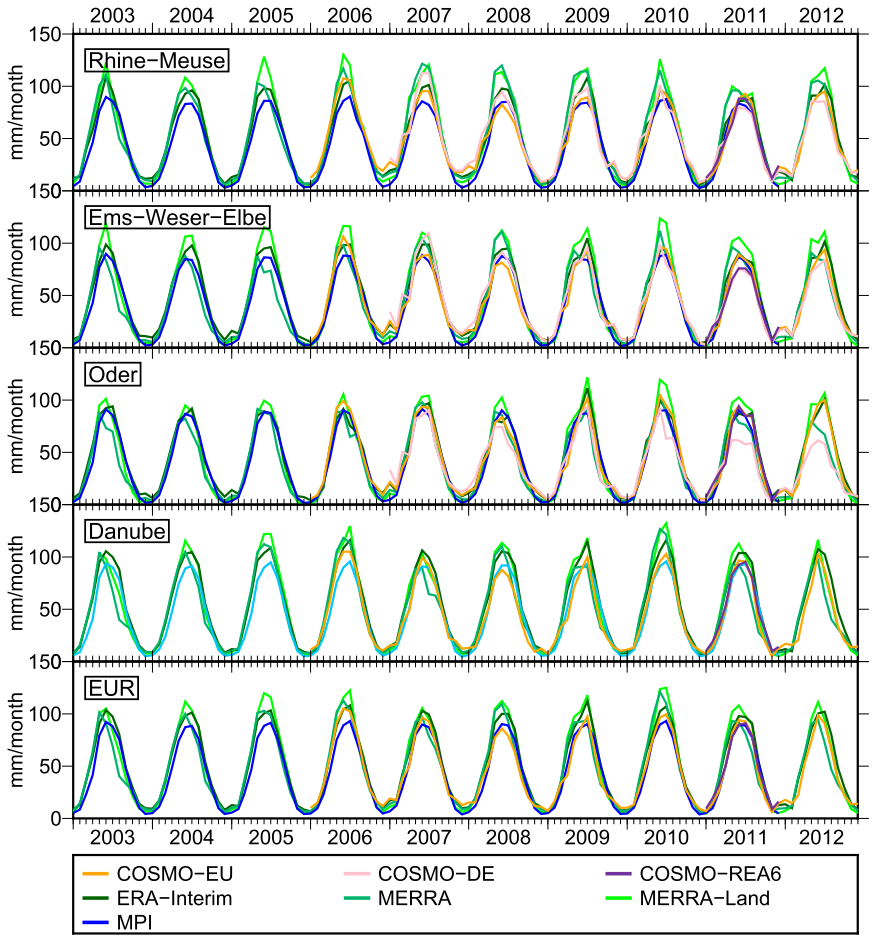

FIG. 3. Monthly total E for different river basins and models.

4) ERROR ASSESSMENT

The GFZ provides calibrated errors [assessed, for example, by Wahr et al. (2006)] for the monthly spherical harmonic coefficients that are mapped to error estimates for spatially averaged TWS change. These error estimates include the effects of noise in the distance measurement between the satellites as well as those of the precise orbit determination, while they disregard noise of some other relevant instruments such as accelerometers and errors in the background models. However, since spatial correlations (i.e., correlations between different coefficients) are not provided, the result is expected to somewhat underestimate the real error. Besides, the standard deviation of discharge is assumed to be 15% of the discharge value itself. We believe this is a conservative assumption, compared to the relative uncertainties between 5% and 10% assumed by Rodell et al. (2011) and Pan et al. (2012). Finally, uncertainties of P and E fields are inferred from the spreading of these datasets around the mean of the various products.

3. Results

First, in section 3a, we describe the variability of the fluxes and storages and the contribution of each component to the water budget. In section 3b, the atmospheric–terrestrial flux as represented by the models and observational datasets is evaluated using GRACE and river discharge data as an independent reference. The observation and modeling of E is particularly challenging; this is why we assess the approach of solving the water budget for E in section 3c. Next, in section 3d, the residuals of the water budget are analyzed and an error model for $P - E$ is established. Finally, in section 3e, this error model is used to correct the flux $P - E$ in order to reconstruct TWS change and TWS prior to the GRACE period.

All results are presented for the combined river basins Rhine–Meuse, Ems–Weser–Elbe, the Oder and Danube catchments, and the total catchment area of these four regions, referred to here as EUR. The statistics we provide refer to the time period 2006–10. COSMO-DE (available from 2007 to 2010) and COSMO-REA6 (at the time of the

TABLE 3. As in Table 2, but for *E*.

Basin	Model	μ (mm month^{-1})	A (mm month^{-1})	Trend [(mm month^{-1}) yr^{-1}]	RMSE (mm month^{-1})	Total RMSE (mm month^{-1})
Rhine–Meuse	COSMO-EU	49	40	−0.8	8	8
	ERA-Interim	49	46	0.7	3	
	MERRA	54	54	0.4	8	
	MERRA-Land	53	57	−0.4	8	
	MPI	42	43	−0.4	10	
Ems–Weser–Elbe	COSMO-EU	44	41	1.9	7	6
	ERA-Interim	48	45	1.0	4	
	MERRA	46	48	1.3	6	
	MERRA-Land	48	56	2.1	9	
	MPI	41	44	2.1	7	
Oder	COSMO-EU	44	43	3.8	5	6
	ERA-Interim	45	44	3.4	4	
	MERRA	40	44	1.4	6	
	MERRA-Land	44	53	4.2	7	
	MPI	41	45	3.9	5	
Danube	COSMO-EU	46	44	3.7	7	7
	ERA-Interim	53	52	3.1	5	
	MERRA	50	49	2.0	7	
	MERRA-Land	53	55	3.4	7	
	MPI	44	46	3.6	8	
EUR	COSMO-EU	46	43	2.8	6	6
	ERA-Interim	51	49	2.5	4	
	MERRA	49	49	1.6	6	
	MERRA-Land	52	55	2.7	7	
	MPI	43	45	2.8	7	

computations available for 2011) do not cover the entire time span, and therefore, these results will only be illustrated in the figures. Throughout the analysis we pay particular attention to the high-resolution COSMO-EU model, as this relatively new model has not been discussed much in the literature so far.

a. Fluxes and storages: The components of the terrestrial water budget

1) PRECIPITATION

Monthly total precipitation over Europe is highly variable (Fig. 2). Spatial averaging causes smaller variability in large target areas (e.g., EUR) than in small target areas (e.g., Rhine–Meuse). The *P* time series are controlled by the intensity of individual precipitation events and not by the annual cycle. Table 2 details similarities and differences between the individual basins and precipitation products.

Over EUR, Danube, and Ems–Weser–Elbe, average *P* amounts to about 65 mm month^{-1}. In contrast, mean *P* is significantly smaller in the Oder basin, with about 55 mm month^{-1}, and higher in the Rhine–Meuse basin, with about 80 mm month^{-1}. While over the catchment of Rhine–Meuse no significant precipitation trend can be seen during 2006–10, for the other basins we do find positive trends between 2 and 4 mm month^{-1}. As already mentioned, the annual cycle is rather small, with amplitudes between 10 and 20 mm month^{-1} for the individual basins.

Differences among the individual *P* datasets are mainly characterized by an offset (Fig. 2). Generally, mean *P* from the different datasets has a similar order in every basin (Table 2). Most of the time, COSMO-EU exhibits the highest values followed by GPCC, ERA-Interim, MERRA-Land, E-OBS, and MERRA. As a gridded dataset, which is based on spatial interpolation of observed precipitation, GPCC is naturally less prone to a model-induced bias. This is complemented by thorough quality control as described by Schneider et al. (2014). GPCC and E-OBS differ by an offset of 5–6 mm month^{-1}. ERA-Interim and MERRA-Land tend to underestimate *P*, yet amplitudes and trends are consistent with the observation-based datasets. In agreement with Lorenz and Kunstmann (2012), we find that MERRA generally predicts significantly less *P* than the other datasets. Furthermore, compared to the other models, the annual amplitude of MERRA is only half as large and the trend also appears to be notably smaller.

In total, COSMO-EU overestimates *P* by about 5 mm month^{-1} compared to GPCC. This is mainly because of high precipitation rates in the first years of the model run, which also lead to a large root-mean-square

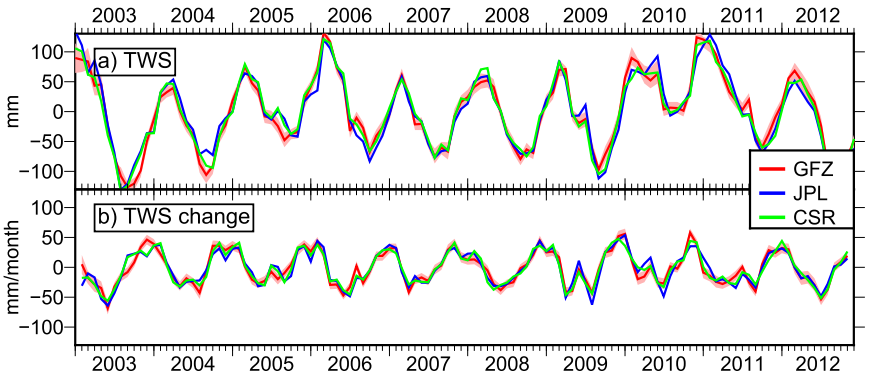

FIG. 4. (a) TWS and (b) TWS changes (TWSC) of the EUR catchment from different GRACE processing centers. For the GFZ solution, error information is propagated from calibrated errors (shaded in red).

error (RMSE). However, the agreement of COSMO-EU with the observation-based datasets improves with time, which is likely because of ongoing changes in the data assimilation scheme. We find that one year of reanalysis data from COSMO-REA6 in 2011 closely resembles COSMO-EU. We further note that in 2007, COSMO-DE differs significantly with respect to the other models in the catchments of Rhine–Meuse and Ems–Weser–Elbe. We speculate that this effect is likely due to the introduction of radar-derived rain rates via latent heat nudging.

2) EVAPOTRANSPIRATION

Evapotranspiration variability over EUR is rather smooth, with a clear annual cycle with large values in summer and limited signal in winter (Fig. 3). Mean values and amplitudes are similar in all river basins (Table 3). The E trend is nearly zero in the catchment of Rhine–Meuse, whereas the other basins have clear positive trends.

Especially in summer, large biases of up to 45 mm month^{-1} between the models persist (Fig. 3). The only observational dataset, MPI, indicates the smallest mean E, followed by E from COSMO-EU (Table 3). ERA-Interim, MERRA, and MERRA-Land predict larger mean E in equal measure. The characteristic seasonal cycle has amplitudes of about 50 mm month^{-1}, which differ largely for the individual models. MERRA-Land, in particular, overestimates the annual cycle in comparison to MPI with amplitude differences of up to 13 mm month^{-1}, that is, more than 20% of the amplitude. This spread reflects the large uncertainties of E datasets.

TABLE 4. This table shows the accuracy of the time series of $\Delta S + R$ and quantifies the error of the individual $P - E$ time series. The value $\hat{s}_{\Delta S+R}$ is the std dev of monthly $\Delta S + R$ values, $\bar{s}_{\Delta S+R}$ represents the std dev of the temporal mean of $\Delta S + R$, and h is the correction of mean ΔS due to the rescaling factor. The constant part of the error model for each $P - E$ dataset is represented by the bias a. Finally, the bottom part of the table shows the RMSE of remaining residuals after applying the error model to the fluxes.

Basin		Rhine–Meuse	Ems–Weser–Elbe	Oder	Danube	EUR
Accuracy of $\Delta S + R$ (mm month^{-1})	$\hat{s}_{\Delta S+R}$	24	23	25	11	9
	$\bar{s}_{\Delta S+R}$	3.1	3.0	3.3	1.4	1.1
	h	3	3	4	2	1
a (mm month^{-1})	COSMO-EU	−2	−3	0	−2	−2
	ERA-Interim	13	5	3	10	9
	MERRA	25	13	9	17	18
	MERRA-Land	19	7	5	14	13
	GPCC + MPI	0	−1	−2	0	0
	E-OBS + MPI	5	4	3	6	6
RMSE residuals (error model applied) (mm month^{-1})	COSMO-EU	26	27	21	18	16
	ERA-Interim	23	25	20	16	14
	MERRA	21	23	19	16	13
	MERRA-Land	26	27	22	17	16
	GPCC + MPI	25	26	21	16	15
	E-OBS + MPI	25	25	21	16	15

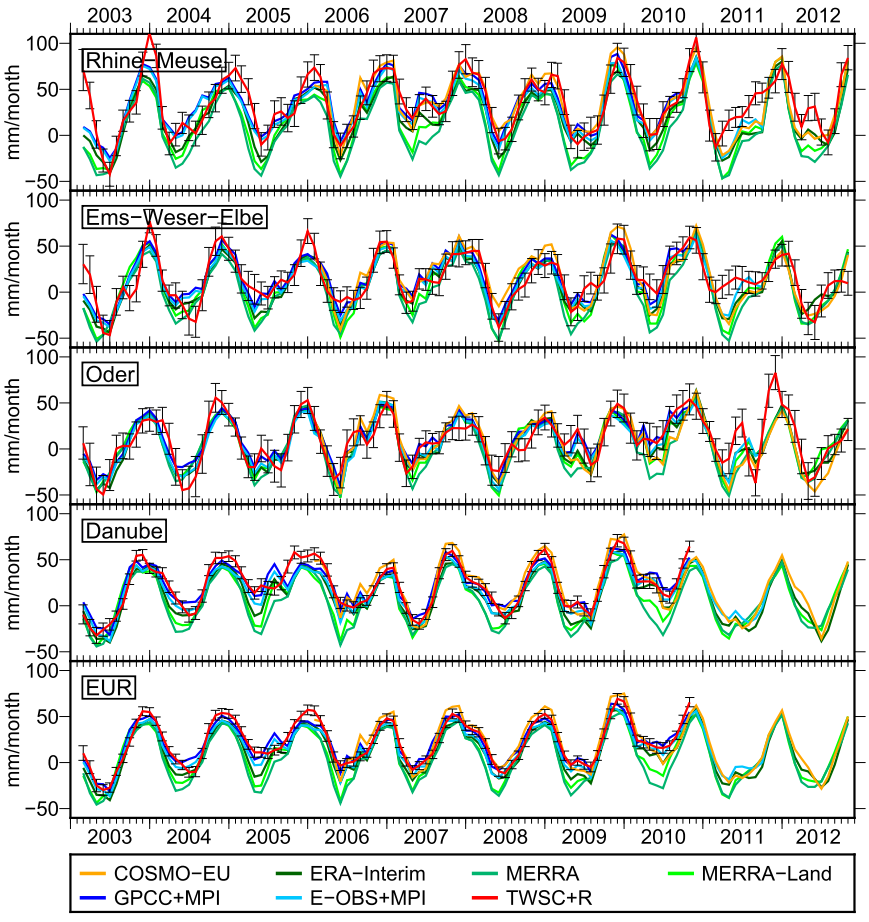

FIG. 5. The water budget equation is solved for $P - E = \Delta S + R$. The right side of the equation is represented by the red line (TWSC = ΔS) and includes error bars. For Danube discharge, data are only given until 2010, resulting in a shorter time series of $\Delta S + R$ in the basins of Danube and EUR. Time series are smoothed with a 3-month moving average to facilitate interpretation.

3) RIVER DISCHARGE

Measured river discharge was obtained from the GRDC and the BfG. Monthly discharge amounts to about one-third of precipitation (Fig. 2). The monthly values lead to a rather smooth time series with strong annual signal with its maximum in winter or spring.

4) TOTAL WATER STORAGE CHANGE

Three GRACE solutions are compared for the EUR study area in Fig. 4. We find that the reprocessed GFZ release 05a fits much better to the JPL and CSR solutions than the previous GFZ release 05 (not shown). Figure 4a shows TWS and Fig. 4b shows TWS change, derived using central differences. We find that the computation of backward derivatives instead (not shown) would introduce a phase shift of at least 8 days.

Table 4 provides, in the first three rows, information about the error budget of the model-independent part of the water budget, $\Delta S + R$. Depending on the basin size, discharge contributes 4–20 times less to the error budget as compared to GRACE; that is, its effect is small or even negligible in the error budget. Also, we find that the rescaling procedure has a small influence of only 1–4 mm month^{-1} on the TWS change estimate. We estimate that in the EUR basin, $\Delta S + R$ exhibits a standard deviation of 9 mm month^{-1}, and that for the smaller basins the error of $\Delta S + R$ amounts to about 24 mm month^{-1}. In comparison, the time series of $P - E$ have a standard deviation of about 23 mm month^{-1} for EUR.

b. *Evaluation of atmospheric–terrestrial flux $P - E$*

The water budget is now solved for the atmospheric–terrestrial flux $P - E$ in order to validate the individual datasets. In Fig. 5, both sides of the water budget equation are pictured. We find that, in particular,

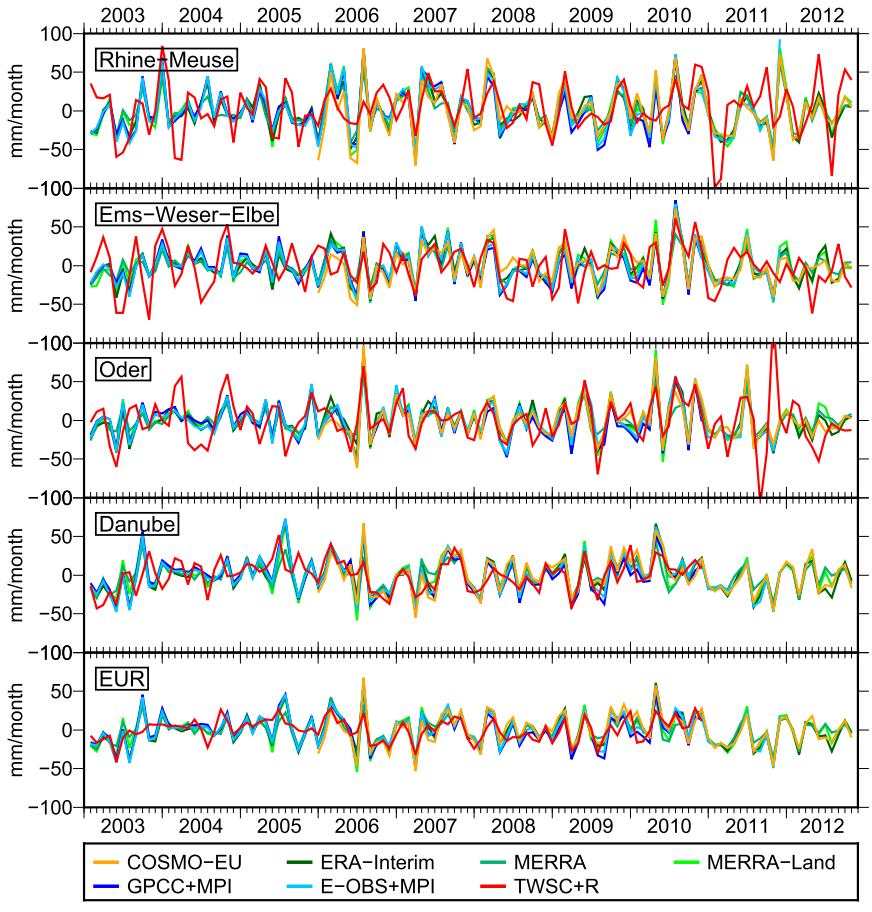

FIG. 6. Deseasoned and detrended variability of $P - E$ and $\Delta S + R$ for the individual basins.

COSMO-EU and the observational datasets are in good agreement with $\Delta S + R$, even for smaller basins. Also, ERA-Interim performs quite well, whereas both MERRA models drift, that is, they systematically underestimate $P - E$. A closer look at deseasoned and detrended variability of the time series is obtained by removing mean value, trend, and signals of 6 and 12 months; this result is shown in Fig. 6. The deseasoned and detrended time series reflect the influence of specific meteorological conditions on $P - E$. Correlations of the deseasoned and detrended variations of $P - E$ and $\Delta S + R$ reach values of about 0.6–0.75. An exception is the basin of Ems–Weser–Elbe, for which we obtain correlations of only 0.5. In this context, it should also be mentioned that the application of central (instead of backward) differences to GRACE time series significantly increases the correlations by about 0.2.

A more detailed interpretation of $P - E = \Delta S + R$ is provided in Table 5. Mean $P - E$ for the complete EUR region agrees well with the estimate of Miralles et al. (2011) of 22.4 mm month^{-1} for mean European $P - E$ and the estimate of globally averaged $P - E$ computed by Lorenz and Kunstmann (2012) of 26.6 mm month^{-1}. In contrast, the basin of Rhine–Meuse shows a large flux deficit (about 30 mm month^{-1}) and the Oder basin shows a rather small flux deficit (about 10 mm month^{-1}). All basins have a positive trend in $P - E$ and $\Delta S + R$ between 2 and 5 mm month^{-1}; only the trend in the combined catchment area of Rhine–Meuse appears smaller.

We find, in particular, that the mean of COSMO-EU in all basins is in good agreement with the GRACE-derived time series. In contrast, ERA-Interim and MERRA underestimate mean $P - E$ by up to 12 and 25 mm month^{-1}, respectively. Additionally, we find that MERRA and MERRA-Land are shifted with respect to $\Delta S + R$ by up to 6 and 10 days. All models overestimate the annual amplitude, whereas the amplitudes of the observational datasets are close to $\Delta S + R$. Among the models, the amplitude of ERA-Interim agrees best with $\Delta S + R$. In all basins, ERA-Interim and the observational datasets have the highest correlation of detrended and deseasoned $P - E$ and $\Delta S + R$, but the difference to the other models is small. Obviously, the correlations depend primarily on the basin size and less on the

TABLE 5. The water budget equation is solved for $P - E$ and compared to GRACE and observed discharge $\Delta S + R$. The mean value over the time span 2006–10, the annual amplitude, and the trend are considered. Moreover, the phase shift ϕ and the RMSE of the two sides of the equation are presented, and the correlation coefficient r refers to the deseasoned and detrended time series.

Basin	Model	μ (mm month^{-1})	A (mm month^{-1})	Trend [(mm month^{-1}) yr^{-1}]	ϕ (days)	RMSE (mm month^{-1})	r
Rhine–Meuse	COSMO-EU	41	38	1.5	−6	26	0.61
	ERA-Interim	26	37	1.3	−4	27	0.62
	MERRA	14	48	1.2	−7	34	0.62
	MERRA-Land	20	45	0.2	−10	34	0.59
	GPCC + MPI	39	33	−0.1	−7	26	0.64
	E-OBS + MPI	34	32	−0.5	−7	26	0.63
	$\Delta S + R$	39 ± 3.1	39 ± 5.2	1.3 ± 2.0	—	—	—
Ems–Weser–Elbe	COSMO-EU	22	38	2.2	−4	28	0.40
	ERA-Interim	14	34	1.9	−5	27	0.48
	MERRA	6	44	1.3	−7	29	0.48
	MERRA-Land	12	44	1.8	−8	32	0.46
	GPCC + MPI	21	32	2.3	−6	27	0.49
	E-OBS + MPI	15	32	2.1	−6	26	0.49
	$\Delta S + R$	19 ± 3.0	28 ± 4.9	3.6 ± 2.0	—	—	—
Oder	COSMO-EU	13	38	1.4	4	25	0.66
	ERA-Interim	11	32	2.9	2	22	0.68
	MERRA	5	41	1.6	−1	26	0.63
	MERRA-Land	9	35	2.9	−2	26	0.67
	GPCC + MPI	15	25	4.0	−1	21	0.69
	E-OBS + MPI	11	28	3.9	1	22	0.67
	$\Delta S + R$	14 ± 3.3	22 ± 4.8	5.0 ± 2.2	—	—	—
Danube	COSMO-EU	27	36	2.0	−1	19	0.68
	ERA-Interim	15	31	3.0	−2	19	0.71
	MERRA	7	38	1.6	−6	25	0.63
	MERRA-Land	11	36	2.8	−9	24	0.66
	GPCC + MPI	24	25	3.5	−3	17	0.71
	E-OBS + MPI	18	28	3.6	−3	18	0.71
	$\Delta S + R$	24 ± 1.4	32 ± 2.3	3.8 ± 1.0	—	—	—
EUR	COSMO-EU	27	36	1.9	−2	17	0.66
	ERA-Interim	16	32	2.6	−2	17	0.74
	MERRA	8	41	1.5	−6	24	0.69
	MERRA-Land	12	38	2.3	−8	23	0.69
	GPCC + MPI	25	27	2.8	−4	15	0.73
	E-OBS + MPI	19	29	2.8	−4	16	0.72
	$\Delta S + R$	25 ± 1.1	32 ± 1.9	3.3 ± 0.8	—	—	—

dataset. Finally, the RMSE was derived; this represents the deviation of $P - E$ from the time series derived from GRACE and discharge data. For MERRA and MERRA-Land, the RMSE is significantly larger than for the other datasets.

c. *Evapotranspiration: Comparison of two approaches*

We now have two ways of computing spatially averaged evapotranspiration: first, based on E obtained from the NWP models and observational datasets, and second, based on E derived from the water budget equation via $P - \Delta S - R$. In Fig. 7, dashed lines represent the E model outputs and solid lines represent E estimates from the water budget equation. The figure shows the annual cycle of E (annual amplitude and phase shift). The spread of the E datasets is particularly large during summer. During winter, estimates via the water budget equation tend to be smaller than model outputs and MPI. However, during summer both sides of the water budget equation agree very well for the observational datasets and for COSMO-EU. Also, the estimate of $P - \Delta S - R$ from ERA-Interim fits well to the observational datasets, while modeled E slightly overestimates E during summer. MERRA clearly underestimates E when derived via the water budget approach, while MERRA-Land performs well for this way. Modeled E of both MERRA models is too high during summer and slightly lagged compared to the other datasets.

We suggest that, while over Europe, the two approaches do not lead to very different results (except for the MERRA models); the water budget approach has the potential to significantly improve E estimates in regions with limited availability of observational data.

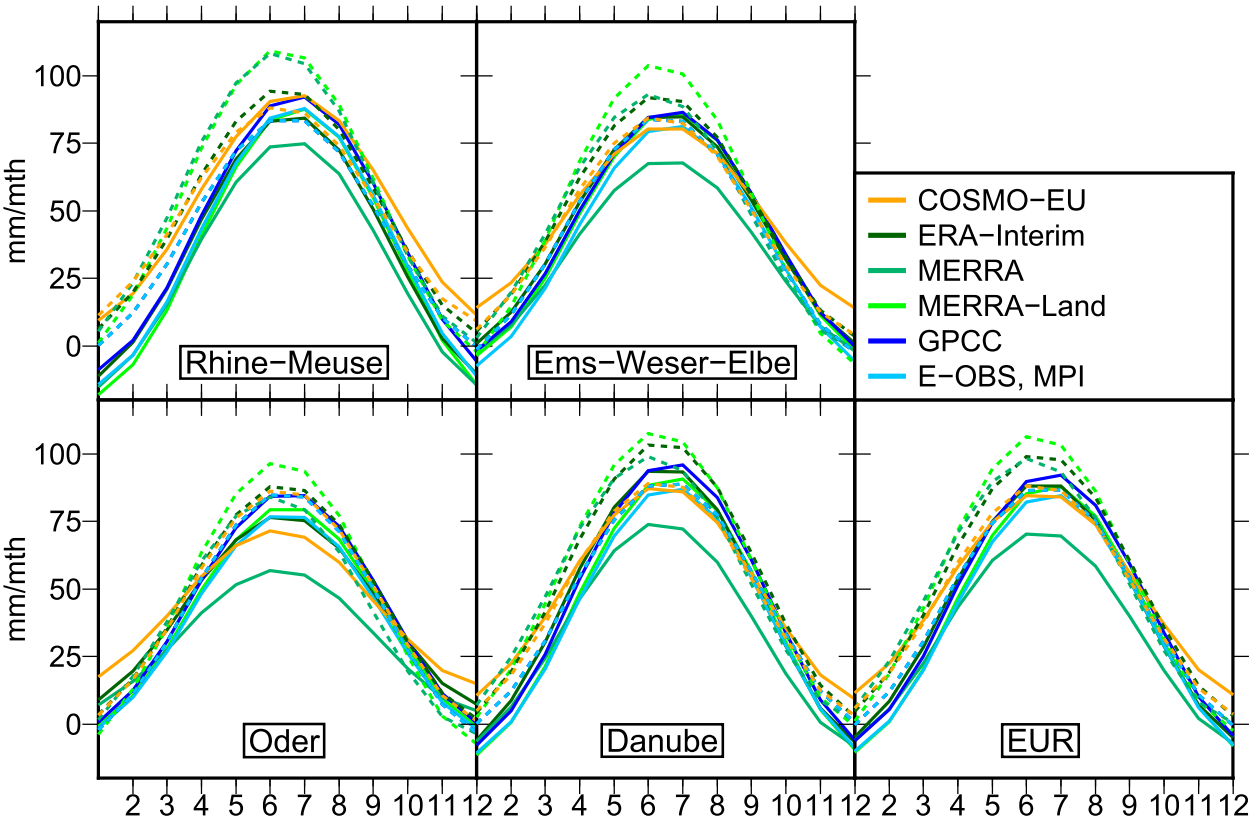

FIG. 7. The annual cycle of E (A and ϕ) is represented. The solid lines show E calculated via the water budget equation according to $P - \Delta S - R$. The dashed lines show the E outputs from the models and from MPI.

d. The residuals of the water budget

In theory, perfect closure of the water budget equation implies $\epsilon = \Delta S - (P - E - R) = 0$. Yet this is not the case, as we have shown above, suggesting that model outputs and observational datasets are contaminated with errors. In Fig. 8, we display the residuals or nonclosure of the water budget equation for the different models and the individual basins. In the first step, the two error models introduced in section 2b(3) were assessed. We find that the error model using a constant offset reduces the water budget residuals by 47% whereas a multiplicative model reduces the residuals only by 35%. Therefore, we choose the additive model as represented in Eq. (11) for the following investigations.

The estimated time-invariant bias a of the error model is specified in the middle of Table 4. It is notably large for MERRA with values of up to 25 mm month^{-1}. Besides, MERRA-Land and ERA-Interim also appear to have large biases of up to 19 and 13 mm month^{-1}, respectively. In contrast, we find COSMO-EU relatively unbiased with values between 0 and -3 mm month^{-1}. The combination of GPCC and MPI data leads to the best closure of the water budget. Whether this bias is statistically significant is assessed with respect to the accuracy of the GRACE and discharge data (given in the upper part of Table 4). Mean values of the time series of $\Delta S + R$ have standard deviations of 1–4 mm month^{-1}; consequently, only the bias estimates of MERRA, MERRA-Land, and ERA-Interim can be considered as significant in all of the basins except for the Oder catchment. We find it surprising that the bias for the small Oder catchment is very small, with a maximum of 10 mm month^{-1} for MERRA.

With respect to Tables 2 and 3, we hypothesize that the bias, observed in the nonclosure of the water budget, originates from P and E fields in ratios depending on the specific data products. Additionally, the annual amplitude of E is more than twice as large as the annual amplitude of P. Phase shifts of E and the annual amplitudes themselves will therefore contribute largely to the annual component of the residual. These are revealed in a power spectral density (PSD) computation shown in Fig. 9. Distinct annual frequencies are contained in MERRA and MERRA-Land, which may be due to

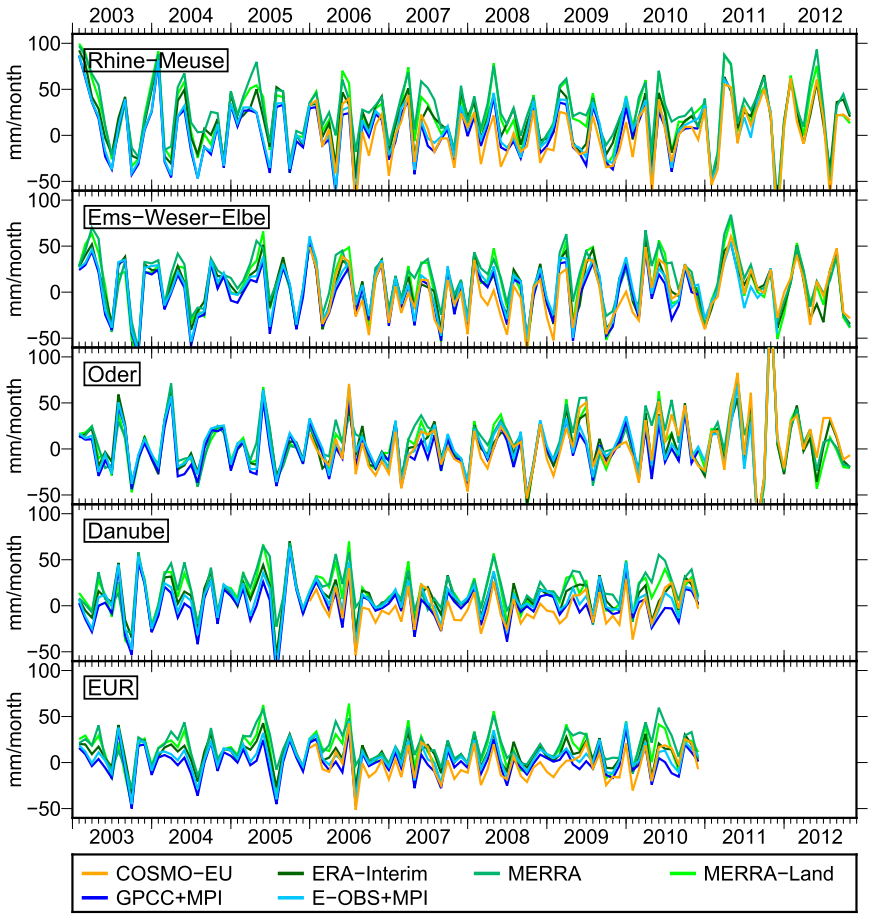

FIG. 8. The bias ϵ of the water budget equation for individual basins and models.

a phase shift in the models. Besides, we observe similar features for all of the models in the higher frequencies. In general, the detected seasonal signal of the residuals decreases with increasing catchment area.

Applying the individual error models to the fluxes $P - E$, we can again assess the closure of the water budget. The lower part of Table 4 provides the RMSE of the remaining residuals. It can be interpreted as a measure for the performance of the models at the higher frequencies. MERRA has the smallest RMSEs, followed by ERA-Interim. This shows that, although the mean of the global models is biased, they perform well for the higher frequencies. The RMSEs of the other models and observational datasets are a bit larger. Overall, the RMSE of the remaining residuals mainly depends on the basin size.

e. Flux-derived total water storage

As already discussed, the best use of GRACE information content is achieved when working with TWS instead of TWS change; therefore, we pursue the evaluation of the integrated water budget equation here. First, GRACE-derived TWS change (Fig. 10a) and TWS (Fig. 10b) are illustrated with red curves exemplarily for the Danube basin. Then, GRACE is contrasted against (catchment integrated) TWS change and TWS derived from NWP model outputs and observational datasets. Furthermore, we consider the dataset of Ming Pan (Pan et al. 2012) as a reference, and for TWS we additionally use storage information from the WGHM for comparison. Error-corrected $P - E - R$ data are evaluated for the time span since 1990; this provides storage information for the period before the start of the GRACE mission and can be termed a ''GRACE/model reconstruction'' of TWS.

We find that error-corrected model estimates of TWS change are in good agreement with each other and also fit to GRACE and TWS change from the multisource approach of Pan et al. (2012) (magenta line). As errors in $P - E$ accumulate through integration, the differences are larger in Fig. 10b. It is difficult to clearly distinguish specific meteorological conditions, such as the flood in

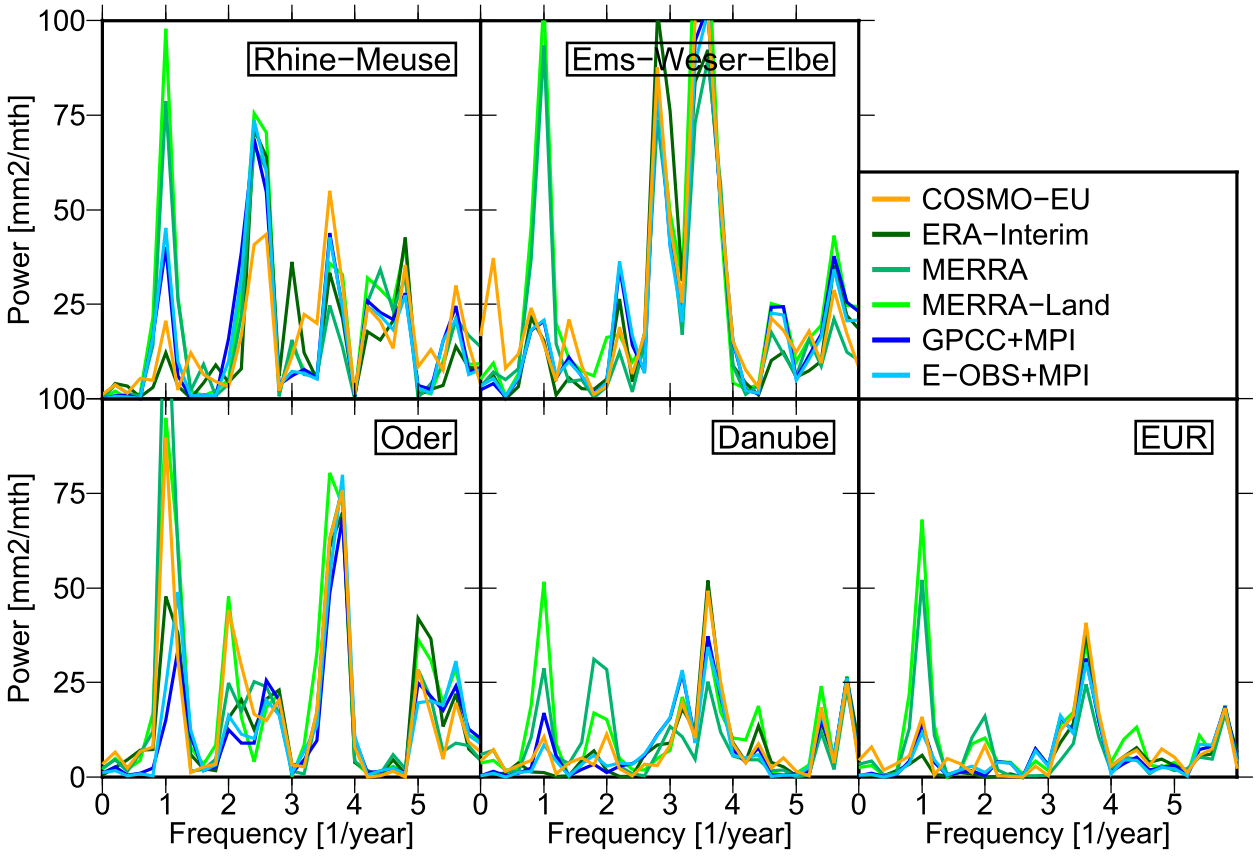

FIG. 9. PSD of the residuals of the water budget equation.

summer 1999, high precipitation rates at the beginning of 2000, or the hot summers of 2000 and 2003. In fact, the reconstructed time series show the evolution of long-term storage mixed with the remaining errors of the flux-based datasets.

In all cases, the error models for the fluxes were derived from data of the time span 2006–10. In Fig. 10b, we observe that these models lead to surprisingly close flux-based TWS estimates for 2002–10, as described before. Before this period, the models drift, likely because of temporal changes of the bias that we did not account for. The propagation of the MERRA time series, for example, indicates a negative bias from 1990 to 2002, which leads to a positive trend. In contrast, ERA-Interim first has a negative trend that changes to a positive trend in 1995 (Fig. 10c). Although in the approach of Pan et al. (2012) the water balance is constrained, remaining errors also lead to a trend in storage. In total, drifts exist that amount to up to 5 times the size of the annual amplitude. If the error model is not applied, large biases lead to drifts of more than 40 times the size of the annual amplitude (not shown here). This shows the value of GRACE for better constraining the long-term evolution of storage. To our knowledge, this is the first attempt to reconstruct TWS from flux information for a time span of 20 years in the Danube basin.

4. Conclusions and outlook

In this contribution, the terrestrial water budget was (almost) closed over European river basins using GRACE-derived TWS change, river discharge, and atmospheric fluxes P and E from a number of different datasets and by developing a catchment-wide error model. Output fields from the analyses COSMO-EU and COSMO-DE and the reanalysis COSMO-REA6; the global reanalyses ERA-Interim and MERRA; the land surface model MERRA-Land; and the observation-based datasets GPCC, E-OBS, and MPI were investigated. The quality of $P - E$ in the NWP models was assessed by analyzing the residuals of the water budget equation. Furthermore, an error model for the fluxes was derived. Finally, model-derived, error-corrected fluxes were used to reconstruct TWS change and TWS previous to the GRACE period.

In line with other studies, we find that the GRACE data offer great potential for assessing the realism of atmospheric–terrestrial fluxes. The representation of

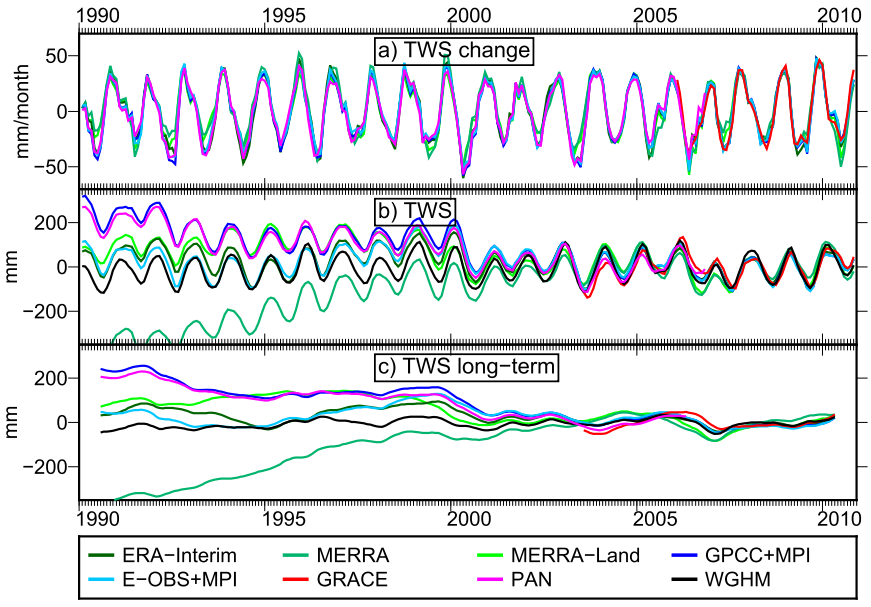

FIG. 10. (a) TWS change and (b) TWS from GRACE is compared to $P - E - R$ for the Danube catchment (smoothed with a 3-month moving average). From NWP models and observational data of $P - E$, storage information can also be obtained for the time before the start of the GRACE mission. (c) The long-term evolution of TWS. As a reference, we use TWS change provided by Ming Pan (magenta) and TWS from WGHM (black).

$P - E$ in the high-resolution COSMO-EU model was identified as relatively unbiased in all river basins. Mean $P - E$ from ERA-Interim was found as biased with values of 3–13 mm month^{-1}, whereas we found that MERRA and MERRA-Land contain offsets of up to 18 and 24 mm month^{-1} and a phase shift of up to 9 days. In contrast, MERRA seems to perform better for the month-to-month signals than the other models. The value of P from COSMO-EU differs, at the beginning of the model run, from the other datasets. Therefore, the evaluation of the corresponding COSMO-REA6, which is available now, appears as a very promising endeavor. COSMO-DE P was found to be contaminated with errors at the beginning of the model run, likely due to the introduction of radar-derived rain rates, and this problem was apparently removed over time.

Detrended and deseasoned variations derived from GRACE and discharge data were found to be highly correlated with the outputs of NWP models, with correlations of about 0.6–0.75. In this context, it should be mentioned that the residuals of the water budget equation can be significantly reduced when using central differences for the GRACE derivative instead of backward differences. At the same time, this increases the correlation between the detrended and deseasoned variations by about 0.2. We dare to say that with the improved quality of release 05, GRACE products can now be used to assess smaller basins than before (e.g., the Oder catchment with 110 000 km^2) and to study the evolution of the water storage clearly beyond its annual behavior, even for European regions.

The water budget equation was solved for evapotranspiration in order to obtain an independent estimate for this most uncertain component of the water budget. An even better estimate of E could be obtained by separating the bias of P and E. Especially for regions with poor coverage of observational data, the water budget equation represents a valuable tool for assessing P and in particular E.

In the last step, we used the fluxes P, E, and R in order to reconstruct catchment-integrated water storage (change) prior to the GRACE period and to identify the effect of yearly weather conditions on these time series. We believe that extending this approach to an even longer time span could become possible, whenever suitable models of a time-variable bias of $P - E$ can be found.

Finally, we acknowledge that the GRACE Follow-On Mission, to be launched in 2017 and equipped with a more precise laser ranging interferometer instrument, will enable us to reduce TWS and TWS change uncertainties and to evaluate the terrestrial water budget at smaller spatial scales. At the same time, improved regional reanalyses such as COSMO-REA6

will hopefully provide more realism to model-derived fields.

Acknowledgments. We thank the Global Runoff Data Centre, 56068 Koblenz, Germany for providing river discharge data and watershed boundaries, and we acknowledge discharge data obtained from the Bundesanstalt für Gewässerkunde. We are grateful to Ming Pan for providing the reference dataset applied in Fig. 10, as well as for his useful comments. Moreover, part of this research was carried out in the Hans-Ertel Centre for Weather Research, a research network of universities, research institutes, and the Deutscher Wetterdienst funded by the BMVBS (Federal Ministry of Transport, Building and Urban Development). Finally, the authors thank editor Ruby Leung and the reviewers for their helpful comments.

REFERENCES

Berrisford, P., and Coauthors, 2011: The ERA-Interim archive, version 2.0. ERA Rep. Series 1, European Centre for Medium Range Weather Forecasts, 23 pp.

Chambers, D. P., and J. A. Bonin, 2012: Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean. *Ocean Sci.,* **9,** 2187–2214, doi:10.5194/os-8-859-2012.

Cheng, M., and B. D. Tapley, 2004: Variations in the Earth's oblateness during the past 28 years. *J. Geophys. Res.,* **109,** B09402, doi:10.1029/2004JB003028.

Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Climate Change,* **2,** 491–496, doi:10.1038/nclimate1452.

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 product release 0005: Revised edition. Scientific Tech. Rep. STR12/02, 21 pp., doi:10.2312/GFZ.b103-1202-25.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* **137,** 553–597, doi:10.1002/qj.828.

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

Fersch, B., H. Kunstmann, A. Bárdossy, B. Devaraju, and N. Sneeuw, 2012: Continental-scale basin water storage variation from global and dynamically downscaled atmospheric water budgets in comparison with GRACE-derived observations. *J. Hydrometeor.,* **13,** 1589–1603, doi:10.1175/JHM-D-11-0143.1.

Gegout, P., cited 2014: Load Love numbers. [Available online at http://gemini.gsfc.nasa.gov/aplo/Load_Love2_CM.dat.]

Güntner, A., 2008: Improvement of global hydrological models using GRACE data. *Surv. Geophys.,* **29,** 375–397, doi:10.1007/s10712-008-9038-y.

Hall, J., and Coauthors, 2013: Understanding flood regime changes in Europe: A state of the art assessment. *Hydrol. Earth Syst. Sci.,* **10,** 15 525–15 624, doi:10.5194/hessd-10-15525-2013.

Hirschi, M., P. Viterbo, and S. I. Seneviratne, 2006: Basin-scale water-balance estimates of terrestrial water storage variations from ECMWF operational forecast analysis. *Geophys. Res. Lett.,* **33,** L21401, doi:10.1029/2006GL027659.

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

Jiménez, C., and Coauthors, 2011: Global intercomparison of 12 land surface heat flux estimates. *J. Geophys. Res.,* **116,** D02102, doi:10.1029/2010JD014545.

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.

——, and Coauthors, 2011: Global patterns of land–atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.,* **116,** G00J07, doi:10.1029/2010JG001566.

Klees, R., E. A. Zapreeva, H. C. Winsemius, and H. H. G. Savenije, 2007: The bias in GRACE estimates of continental water storage variations. *Hydrol. Earth Syst. Sci.,* **11,** 1227–1241, doi:10.5194/hess-11-1227-2007.

Klemann, V., and Z. Martinec, 2011: Contribution of glacial-isostatic adjustment to the geocenter motion. *Tectonophysics,* **511,** 99–108, doi:10.1016/j.tecto.2009.08.031.

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

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

Longuevergne, L., B. Scanlon, and C. R. Wilson, 2010: GRACE hydrological estimates for small basins: Evaluating processing approaches on the High Plains Aquifer, USA. *Water Resour. Res.,* **46,** W11517, doi:10.1029/2009WR008564.

Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.,* **13,** 1397–1420, doi:10.1175/JHM-D-11-088.1.

Lucchesi, R., 2012: File specification for MERRA products. GMAO Office Note No. 1 (version 2.3), Global Modeling and Assimilation Office, 83 pp. [Available online at http://gmao.gsfc.nasa.gov/pubs/docs/Lucchesi528.pdf.]

Miralles, D. G., R. A. M. De Jeu, J. H. Gash, T. R. H. Holmes, and A. J. Dolman, 2011: Magnitude and variability of land evaporation and its components at the global scale. *Hydrol. Earth Syst. Sci.,* **15,** 967–981, doi:10.5194/hess-15-967-2011.

Mueller, B., and Coauthors, 2011: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations. *Geophys. Res. Lett.,* **38,** L06402, doi:10.1029/2010GL046230.

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. *J. Climate,* **25,** 3191–3206, doi:10.1175/JCLI-D-11-00300.1.

Priestley, C., and R. Taylor, 1972: On the assessment of surface heat flux and evaporation using large-scale parameters. *Mon. Wea. Rev.,* **100,** 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2.

Ramilien, G., F. Frappart, A. Günter, 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 Resour. Res.,* **42,** W10403, doi:10.1029/2005WR004331.

Reichle, R. H., R. D. Koster, G. J. M. D. Lannoy, B. A. Forman, Q. Liu, S. P. P. Mahanama, and A. Tourè, 2011: Assessment and enhancement of MERRA land surface hydrology estimates. *J. Climate,* **24,** 6322–6338, doi:10.1175/JCLI-D-10-05033.1.

Rienecker, M. M., and Coauthors, 2011: MERRA: NASAs Modern-Era Retrospective Analysis for Research and Applications. *J. Climate,* **24,** 3624–3648, doi:10.1175/JCLI-D-11-00015.1.

Rietbroek, R., M. Fritsche, S.-E. Brunnabend, I. Daras, J. Kusche, J. Schröter, F. Flechtner, and R. Dietrich, 2012: Global surface mass from a new combination of GRACE, modelled OBP and reprocessed GPS data. *J. Geodyn.,* **59–60,** 64–71, doi:10.1016/j.jog.2011.02.003.

Rodell, M., E. B. McWilliams, J. S. Famiglietti, H. K. Beaudoing, and J. Nigro, 2011: Estimating evapotranspiration using an observation based terrestrial water budget. *Hydrol. Processes,* **25,** 4082–4092, doi:10.1002/hyp.8369.

Roebeling, R. A., E. L. A. Wolters, J. F. Meirink, and H. Leinse, 2012: Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *J. Hydrometeor.,* **13,** 1552–1566, doi:10.1175/JHM-D-11-089.1.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese, 2011: GPCC full data reanalysis version 6.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. Global Precipitation Climatology Centre, Offenbach, Germany, doi:10.5676/DWD_GPCC/FD_M_V6_050.

——, ——, ——, ——, 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. *Theor. Appl. Climatol.,* **115,** 15–40, doi:10.1007/s00704-013-0860-x.

Schraff, C. H., 1997: Mesoscale data assimilation and prediction of low stratus in the Alpine region. *Meteor. Atmos. Phys.,* **64,** 21–50, doi:10.1007/BF01044128.

Seitz, F., M. Schmidt, and C. K. Shum, 2008: Signals of extreme weather conditions in Central Europe in GRACE 4-D hydrological mass variations. *Earth Planet. Sci. Lett.,* **268,** 165–170, doi:10.1016/j.epsl.2008.01.001.

Syed, T. H., J. S. Famiglietti, and D. P. Chambers, 2009: GRACE-based estimates of terrestrial freshwater discharge from basin to continental scales. *J. Hydrometeor.,* **10,** 22–40, doi:10.1175/2008JHM993.1.

Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, 2004: GRACE measurements of mass variability in the earth system. *Science,* **305,** 503–505, doi:10.1126/science.1099192.

Trenberth, K. E., 2011: Changes in precipitation with climatic change. *Climate Res.,* **47,** 123–138, doi:10.3354/cr00953.

Wahr, J., M. Molenaar, and F. Bryan, 1998: Time variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE. *J. Geophys. Res.,* **103,** 30 205–30 229, doi:10.1029/98JB02844.

——, S. Swenson, and I. Velicogna, 2006: Accuracy of grace mass estimates. *Geophys. Res. Lett.,* **33,** L06401, doi:10.1029/2005GL025305.

Wang, K., and R. E. Dickinson, 2012: A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Rev. Geophys.,* **50,** RG2050, doi:10.1029/2011RG000373.

Werth, S., and A. Güntner, 2010: Calibration analysis for water storage variability of the global hydrological model WGHM. *Hydrol. Earth Syst. Sci.,* **14,** 59–78, doi:10.5194/hess-14-59-2010.

Zaitchik, B. F., 2008: Assimilation of GRACE terrestrial water storage data into a land surface model: Results for the Mississippi River basin. *J. Hydrometeor.,* **9,** 535–548, doi:10.1175/2007JHM951.1.