

Estimating Large-Scale Precipitation Minus Evapotranspiration from GRACE Satellite Gravity Measurements

SEAN SWENSON AND JOHN WAHR

Department of Physics, and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado

(Manuscript received 29 November 2004, in final form 25 July 2005)

ABSTRACT

Currently, observations of key components of the earth's large-scale water and energy budgets are sparse or even nonexistent. One key component, precipitation minus evapotranspiration ($P - ET$), remains largely unmeasured due to the absence of observations of ET . Precipitation minus evapotranspiration describes the flux of water between the atmosphere and the earth's surface, and therefore provides important information regarding the interaction of the atmosphere with the land surface. In this paper, large-scale changes in continental water storage derived from satellite gravity data from the Gravity Recovery and Climate Experiment (GRACE) project are combined with river discharge data to obtain estimates of areally averaged $P - ET$.

After constructing an equation describing the large-scale terrestrial water balance reflecting the temporal sampling of GRACE water storage estimates, GRACE-derived $P - ET$ estimates are compared to those obtained from a reanalysis dataset [NCEP/Department of Energy (DOE) reanalysis-2] and a land surface model driven with observation-based forcing [Global Land Data Assimilation System (GLDAS)/Noah] for two large U.S. river basins. GRACE-derived $P - ET$ compares quite favorably with the reanalysis-2 output, while $P - ET$ from the Noah model shows significant differences. Because the uncertainties in the GRACE results can be computed rigorously, this comparison may be considered as a validation of the models.

In addition to showing how GRACE $P - ET$ estimates may be used to validate model output, the accuracy of GRACE estimates of both the seasonal cycle and the monthly averaged rate of $P - ET$ is examined. Finally, the potential for estimating seasonal evapotranspiration is demonstrated by combining GRACE seasonal $P - ET$ estimates with independent estimates of the seasonal cycle of precipitation.

1. Introduction

A number of studies have attempted to quantify and understand the large-scale aspects of the earth's water and energy budgets (Seneviratne et al. 2004; Roads et al. 2002, 2003; Berbery and Barros 2002; Maurer et al. 2001; Yarosh et al. 1999; Yeh et al. 1998; Ropelewski and Yarosh 1998). These studies have been hindered by a lack of consistent, high quality data having adequate spatial resolution and coverage to quantitatively describe the large-scale water cycle. Accurately closing the water budget at large scales may therefore be considered an open problem, and the subject of considerable current research.

Partitioning the water balance into atmospheric and terrestrial components, one may express the areally averaged water budget as

$$\frac{\partial Q}{\partial t} = -(P - ET) + MC, \quad (1)$$

$$\frac{\partial S}{\partial t} = P - ET - R, \quad (2)$$

$$\frac{\partial(Q + S)}{\partial t} = MC - R, \quad (3)$$

where Q is total precipitable water content of the atmosphere; P is precipitation; ET is evapotranspiration; $MC = -\nabla \cdot \{vq\}$ is moisture convergence, where $\{vq\}$ is the vertically integrated atmospheric moisture flux, and v and q are the horizontal wind and specific humidity, respectively; S is terrestrial water storage; and R is net surface and subsurface discharge (Gutowski et al.

Corresponding author address: Dr. Sean Swenson, Dept. of Physics, and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Campus Box 390, Boulder, CO 80309.
E-mail: swensosc@colorado.edu

1997). Note that all terms represent the areally averaged value for a specific region.

Of the six components of the water cycle described by (1) and (2), the most rigorously measured are precipitation (P) and discharge (R). For example, the Global Precipitation Climatology Project (GPCP) currently merges satellite precipitation estimates with data from over 6000 rain gauges to provide monthly, global-gridded precipitation maps having $1^\circ \times 1^\circ$ resolution (Huffman et al. 1997). In the United States, agencies such as the National Centers for Environmental Prediction (NCEP) combine observations from more than 3000 automated rain gauges with Doppler radar estimates of precipitation to create hourly multisensor precipitation maps having 4 km by 4 km resolution (Seo 1998). Globally, a gauge-only precipitation dataset is produced at the Global Precipitation Climatology Center (GPCC) (Rudolf 1995). This dataset consists of monthly areally averaged rainfall totals, on a $1^\circ \times 1^\circ$ global grid.

Surface discharge (R) is also regularly measured over much of the world. The Global Runoff Data Center (GRDC) produces a database consisting of data from more than 3600 stations monitoring about 2900 rivers worldwide. The U.S. Geological Survey (USGS) operates and maintains about 7000 gauges in the United States. Subsurface discharge is less well understood, but its flow patterns are typically assumed to be consistent with those at the surface (Zekster and Loaiciga 1993).

Radiosonde measurements of vertical profiles of pressure, temperature, relative humidity, and wind speed are used to estimate precipitable water (Q) and moisture convergence (MC). These observations are made much less frequently, and with a much lower spatial density, than those of precipitation and discharge. In the United States, for example, the National Weather Service (NWS) operates approximately 100 radiosonde stations, making twice daily measurements. Globally, the National Climatic Data Center (NCDC) catalogues data at 2300 upper-air stations.

In principle, S in (2) describes changes in water storage throughout the entire water column, including surface water, soil moisture, groundwater, snow, and ice (Rodell and Famiglietti 2001). Currently, there exist some regional networks for measuring soil moisture in the United States [Oklahoma Mesonet (<http://okmesonet.ocs.ou.edu/overview/default.html>); Illinois State Water Survey (Hollinger and Isard 1994); and the Global Soil Moisture Data Bank in Asia (Robock et al. 2000)]. However, operational measurements of terrestrial water storage at a continental scale are minimal to nonexistent in most areas. Estimates of changes in groundwater primarily come from monitoring of water

levels in wells. The USGS monitors a national network consisting of about 150 wells as part of the Groundwater Resources Program, supplemented by thousands of wells in some states as part of the Cooperative Water Program. However, some wells are monitored only infrequently (i.e., 1–2 times per year). Additionally, the use of well-level information to describe regional groundwater changes requires accurate measurements of subsurface parameters such as porosity or specific yield, as well as knowledge of their spatial variability, neither of which is well known (Alley et al. 2002). Finally, water in the unsaturated zone between the surface soil layer and the water table is measured neither by soil moisture techniques (i.e., gravimetric or neutron probes) nor by well-level monitoring.

Precipitation minus evapotranspiration ($P - ET$) is the net flux of water from the atmosphere to the earth's surface and is therefore present in both (1) and (2). While precipitation is routinely measured and archived, no such network of evapotranspiration measurements exists. Regional estimates of ET are typically made by incorporating satellite-measured estimates of brightness temperatures and surface meteorological variables into complex land surface models (Diak et al. 2004; Boegh and Soegaard 2004), and as such are limited by errors in the radiometric temperature observations, errors in the model parameterizations, and representativeness of the meteorological observations.

Because accurate estimates of evapotranspiration and water storage are unavailable, water budget studies typically begin by using estimates of $(\partial Q/\partial t)$ and MC in (1) to compute residual values of $P - ET$. For example, Ropelewski and Yarosh (1998) combined estimates of vertically integrated flux divergence and water vapor derived from twice-daily radiosonde measurements with measurements of river discharge and precipitation to estimate the mean annual cycle of both $P - ET$ and evaporation for the 20-yr period spanning 1973–92. Yarosh et al. (1999) concluded that, due to inadequate spatial and temporal sampling, the network of 15 radiosondes could not adequately describe the large-scale atmospheric water budget.

Noting that “[m]any relevant hydrologic processes either are not measured or are poorly measured,” Roads et al. (1994) instead used atmospheric analyses, which attempt to overcome the observational shortfall by combining observed meteorological variables with short-range model forecasts, from the National Meteorological Center (now NCEP) to estimate a climatological hydrologic budget for the United States. Basin averages of residual evaporation from this climatology were then compared to basin-averaged measurements of adjusted pan evaporation from an earlier study by

Guetter and Georgakakos (1993). The two estimates of ET showed significant differences. Water storage estimates, calculated as residuals in both studies, showed similar discrepancies. In a later study, Roads et al. (2002) examined the vertically integrated water and energy budgets of large regions using NCEP/Department of Energy (DOE) reanalysis-2 (R-2) datasets. R-2 contains many improvements over previous analyses. Because of imperfect parameterizations of certain model processes, the model state variables must be nudged close to observations. The budget residuals shown in this study, which are due primarily to nudging, were often a significant component of the regional water budgets.

Using the soil moisture dataset described in Hollinger and Isard (1994), Seneviratne et al. (2004) assessed terrestrial water storage inferred from the European Centre for Medium-Range Weather Forecasts (ECMWF) 40-yr Re-Analysis (ERA-40) dataset and found the reanalysis water storage estimates agreed well with the observed soil moisture. While these results show the ability of reanalyses to quite accurately represent the land state, it should be noted that this assessment focused on a single region (Illinois) that is both smaller than the basin scales examined here and very well observed relative to most of the world, and therefore may not be suitably general to extrapolate to larger spatial scales, or to other regions where sources of data for assimilation are sparse.

Currently, our understanding of the water budget is limited by a lack of observational data and by inadequate parameterizations of the underlying physics in models meant to simulate the global climate. Our ability to close the water budget observationally will be improved by the addition of data from the Gravity Recovery and Climate Experiment (GRACE). GRACE provides global, monthly estimates of the earth's gravity field, from which estimates of vertically integrated, terrestrial water storage can be inferred.

The potential for estimating $P - ET$ and ET by incorporating GRACE data in a water balance approach has been demonstrated by Swenson et al. (2001) and Rodell et al. (2004b), respectively. In this paper, we examine the accuracy of estimates of areally averaged precipitation minus evapotranspiration, $P - ET$, obtained by combining GRACE estimates of changes in regional water storage, S , with measurements of surface runoff, R . Because of the temporal sampling characteristics of GRACE, (2) cannot be used directly to relate changes in water storage derived from GRACE to the fluxes of precipitation, evapotranspiration, and runoff. Therefore, we first develop an equation that relates changes in temporally averaged water storage to the

appropriate representations of P , ET , and R . We then show results for two North American river basins, the Mississippi and the Ohio-Tennessee, and compare our $P - ET$ results with those obtained from a general circulation model (NCEP/DOE reanalysis-2) and a land surface model [Global Land Data Assimilation System (GLDAS)/Noah]. Finally, we examine the accuracies with which GRACE data can be used to infer the monthly averaged rate of $P - ET$ and the seasonal cycle of $P - ET$.

To assess the accuracy of water storage estimates made by GRACE, we follow earlier works (Wahr et al. 1998; Swenson et al. 2003) that used simulations to characterize the errors inherent in the use of GRACE data to recover surface-mass variability. Simulations provide a crucial tool for determining the errors due to leakage from unwanted sources of surface-mass variability, as well as errors resulting from the imperfect removal of the atmospheric gravity signal from GRACE data. Estimates of the errors in GRACE water storage estimates based on satellite measurement error alone are inadequate. It is for this reason that GRACE simulations are a necessary component of any assessment of the accuracy of actual GRACE water storage estimates.

2. GRACE background

GRACE is a partnership between the National Aeronautics and Space Administration (NASA) and Deutsche Forschungsanstalt für Luft und Raumfahrt (DLR) designed to accurately map variations in the earth's gravity field at monthly time scales over its 5-yr lifespan (Tapley et al. 2004a). Launched in March 2002, under the auspices of the NASA Earth System Science Pathfinder (ESSP) Program, the GRACE mission consists of a pair of satellites in tandem formation. Variations in the gravity field cause the range between the two satellites to vary; these variations are corrected for nongravitational effects via onboard accelerometers, and combined with onboard GPS data to create a map of the earth's gravity field. At approximately monthly intervals, the GRACE Project constructs a new, monthly averaged estimate of the earth's gravity field.

The month-to-month gravity variations obtained from GRACE provide information about changes in the distribution of mass within the earth and at its surface. Rodell and Famiglietti (1999) compared the output from 12 modeled hydrologic datasets to preliminary estimates of the GRACE error budget, and predicted that monthly terrestrial water storage variations for river basins having areas of $2.0 \times 10^5 \text{ km}^2$ and larger would be detectable. Swenson et al. (2003) used simu-

lated GRACE data to estimate the accuracy with which monthly variations in areally averaged terrestrial water storage could be recovered from GRACE gravity fields. Using mission baseline estimates of errors in the gravity fields, that study showed that changes in water storage, averaged over regions having areas of roughly $4.0 \times 10^5 \text{ km}^2$ and larger, could be recovered to an accuracy of a few millimeters of equivalent water thickness. While the initial gravity fields released by the GRACE Project have not attained the baseline error levels, the 2003 fields show improvement over those from 2002 (Wahr et al. 2004), and continued improvement is anticipated. In the future, the entire dataset will be reprocessed as the GRACE gravity field estimation process is refined.

3. Estimating changes in areally averaged water storage

Given a field of water storage values, one can represent the change in areally averaged water storage, ΔS , for a given region with the following expression:

$$\Delta S = \frac{1}{\Omega_{\text{region}}} \int \Delta \sigma(\theta, \phi) \vartheta(\theta, \phi) d\Omega, \quad (4)$$

where $d\Omega = \sin\theta d\theta d\phi$ is an element of solid angle, and the integration is performed over the sphere. Here, $\vartheta(\theta, \phi)$ which we call the exact averaging kernel, determines the relative contribution of each point in the field to the integrated value. In this case, ϑ has a value of 1 for all points inside the region of interest, and a value of 0 for all points outside that region. Integrating ϑ over the sphere gives Ω_{region} , the angular area of the region.

Because of the errors inherent in the GRACE gravity fields, using an exact averaging kernel to extract regional water storage variations leads to highly inaccurate estimates (Swenson et al. 2003). Instead, a modified averaging kernel can be constructed (Swenson and Wahr 2002), which attempts to minimize the effects of satellite measurement errors while faithfully representing the water storage signal in the region of interest. To optimize the averaging kernel, one requires an estimate of the GRACE measurement uncertainty, which the GRACE Project provides, as well as initial guesses for the approximate amplitude and spatial scale of the water storage variability. The different sampling properties of the exact and modified averaging kernels leads to an additional error called ‘‘leakage.’’ Despite this additional error, using such an approximate averaging kernel, instead of an exact averaging kernel, can reduce the total error in the recovered water storage anomaly by as much as an order of magnitude (Swenson et al. 2003).

4. Relating changes in GRACE water storage estimates to precipitation minus evapotranspiration

Equation (2), which describes the large-scale terrestrial water balance, relates the instantaneous values of the individual water balance components. Because GRACE estimates of water storage are not instantaneous, but rather temporally averaged quantities, one must modify (2) to reflect the temporal sampling appropriate to GRACE. Here we generalize the derivation of the results of Rodell et al. (2004).

Defining the accumulation rate, $A(t)$, by

$$A(t) = P(t) - ET(t) - R(t), \quad (5)$$

Eq. (2) can be written as

$$\frac{dS}{dt} = A(t). \quad (6)$$

Integrating both sides of the equation gives the change in storage, or net accumulation,

$$S(t_2) - S(t_1) = \int_{t_1}^{t_2} A(t) dt, \quad (7)$$

between two times, t_1 and t_2 . The average accumulation rate is simply the change in storage divided by the amount of time between t_1 and t_2 ,

$$\frac{S(t_2) - S(t_1)}{T_{21}} = \langle A \rangle_{21}, \quad (8)$$

where $T_{21} = t_2 - t_1$, and brackets, $\langle \rangle$, represent a time-averaged quantity.

An expression for the average storage for the time period t_1 to t_2 may be obtained by letting $t_2 \rightarrow t$ in (7) and integrating

$$\langle S \rangle_{21} = \frac{1}{T_{21}} \int_{t_1}^{t_2} S(t) dt = S(t_1) + \frac{1}{T_{21}} \int_{t_1}^{t_2} dt \int_{t_1}^t A(t') dt'. \quad (9)$$

Thus the average storage is the initial storage plus the average accumulation during that time period.

Relative changes in time-averaged water storage, such as those provided by GRACE, can be expressed as

$$\langle S \rangle_{43} - \langle S \rangle_{21} = S(t_3) - S(t_1) + \langle \Phi \rangle_{43} - \langle \Phi \rangle_{21}, \quad (10)$$

where $\langle \Phi \rangle$ is the average accumulation for that time period:

$$\Phi(t) = \int_{t_1}^t A(t') dt' \quad \text{and} \quad \langle \Phi \rangle_{21} = \frac{1}{T_{21}} \int_{t_1}^{t_2} \Phi(t) dt, \quad (11)$$

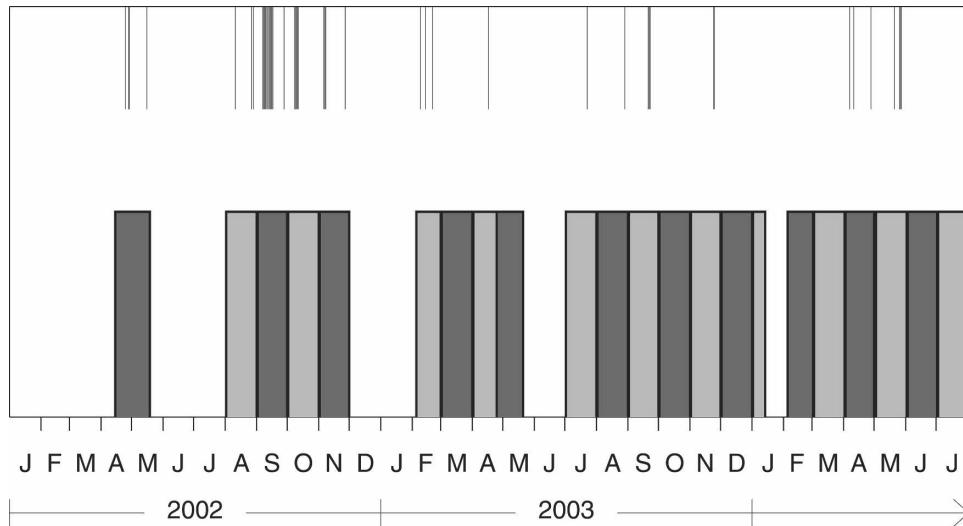


FIG. 1. Shaded bars along the bottom axis of the figure represent time periods during which data were used to construct GRACE gravity fields. Lines along the top axis of the figure represent days that were not included in each time period.

and the first period is bounded by times t_1 and t_2 , and the second period is bounded by times t_3 and t_4 . Combining Eqs. (10) and (7) gives

$$\langle S \rangle_{43} - \langle S \rangle_{21} = T_{31} \langle A \rangle_{31} + \langle \Phi \rangle_{43} - \langle \Phi \rangle_{21}. \quad (12)$$

The first term on the rhs of (12) is the total accumulation from the beginning of the first month to the beginning of the second month. The second and third terms represent the difference between the average accumulations during the first and second months. Note that this formulation makes no assumption on the continuity of the GRACE data; it may be used both for consecutive months and for months that are separated by time periods when GRACE data are unavailable, such as those that occur in 2002.

Equation (12), which is nearly equivalent to Eq. (6) of Rodell et al. (2004), represents the appropriate relation between monthly averaged changes in water storage and the precipitation, evapotranspiration, and runoff fluxes that comprise the land surface water balance. To validate GRACE water storage estimates from observations of P, E, and R, (or vice versa), (12) must be used to properly account for the inherent temporal sampling of GRACE. Given adequate temporal resolution of the flux terms (e.g., daily meteorological observations, 6-hourly model output), evaluation of the right-hand side of (12) is straightforward.

In this study, we make the additional step of dividing both sides of (12) by T_{31} ,

$$\frac{\langle S \rangle_{43} - \langle S \rangle_{21}}{T_{31}} = \langle A \rangle_{31} + \frac{\langle \Phi \rangle_{43} - \langle \Phi \rangle_{21}}{T_{31}}. \quad (13)$$

The quantities represented in (12) are integrated quantities, and therefore, differences in the underlying flux terms may lead to accumulations that diverge greatly for large values of T_{31} , the period of time between successive averaging periods. Equation (13) avoids this effect by representing the flux terms as an average rate plus an additional term that depends on the length of the averaging periods, T_{21} and T_{43} . [Note that (13) reduces to (8) as $t_2 \rightarrow t_1$ and $t_4 \rightarrow t_3$, as it should.]

5. Datasets

a. GRACE data

At present, the GRACE Project has released 22 gravity field solutions; each solution represents the earth's average gravity field for an approximately 30-day time period. Each solution consists of a set of Stokes coefficients complete to degree and order 120. Figure 1 shows the time span of the currently available data, as well as the days of data that were not included in each solution. During the processing of the raw data, effects of short-period mass variability are removed using models of the atmosphere, ocean, and tides. Presently, no hydrologic model is used to de-alias the gravity fields.

With the gravity field solutions, the GRACE Project also produces a calibrated formal error estimate that accounts for the instrument measurement errors and processing errors. The calibrated error estimate is derived in part from internal consistency tests examining the variability of a number of subset solutions (Tapley

et al. 2004b). At low degrees, the calibrated error estimate is quite conservative; in fact, at the lowest degrees, it is larger than the root-mean-square variation in the gravity fields themselves. In addition, the calibrated error estimates are not provided for each field individually; there is one error estimate made for the 2002 fields, and another for all subsequent fields.

The fact that the error estimates at the lowest degrees are larger than the variation in the actual GRACE fields indicates that these error estimates include some systematic effects, which will not be present in the time-variable component of the GRACE fields. Therefore, we seek an error estimate that represents just the errors in the time-variable component of the GRACE fields. As an alternative to the calibrated formal error estimates, we have constructed a time series of GRACE error fields based on a combination of the formal error estimates, the prelaunch baseline error estimates, and the nonannual component of the GRACE gravity fields. These estimates allow us first to remove possible systematic effects and second to characterize the time evolution of the errors in the GRACE solutions.

In brief, we assume the degree amplitudes of the actual errors for each field are given by the baseline error estimate multiplied by a scale factor. To obtain the scale factor, we first compute the nonannual residual field for each month. Next, the degree amplitudes of the baseline error estimates are fit with a scale factor to those of each residual field. The residuals, obtained by fitting and removing the best-fitting mean and annual cycle of the entire time series of gravity field coefficients at each degree and order, are assumed to be entirely the consequence of GRACE gravity field errors. By assuming that all of the nonannual residual is comprised of error we are also making a conservative error estimate; there is almost certainly signal remaining in the residual, and therefore the errors are likely even smaller than our estimates. Finally, to describe the error as a function of order, m , for each degree, l , we apply the m distribution of the formal error estimates to our scaled error estimates.

Leakage, which refers to the influence of mass variability from outside the region of interest, is an additional source of error in a GRACE estimate of water storage. The process of creating region-specific averaging kernels, which minimize this error source, is described in Swenson and Wahr (2002) and Swenson et al. (2003). To estimate the amount of leakage error present in a GRACE water storage estimate, we perform simulations by applying the averaging kernel to a synthetic GRACE dataset. The synthetic data are con-

structed by combining models of sources of mass variability (such as terrestrial water storage, including soil moisture, groundwater, snow, ice, ocean bottom pressure, and postglacial rebound) with models of GRACE processing and measurement errors (Wahr et al. 1998). These simulations allow us to identify the relative contributions of each of these components in the synthetic GRACE-derived water storage estimate, and to compare the errors to the true signal in the synthetic data. While we cannot directly estimate the amount of leakage error in an actual GRACE water storage estimate, the results of the simulations allow us to infer their values relative to the signal. Furthermore, as the models used to construct the synthetic datasets are improved, these error estimates will also improve.

For this study, the synthetic GRACE dataset was created using the Climate Prediction Center's (CPC) "Leaky Bucket" soil moisture model (Fan and Van den Dool 2004). The model takes observed precipitation and temperature and calculates soil moisture, evaporation, and runoff. The CPC model does not include water storage variability beneath the soil layer, and does not fully account for the latent heat of fusion of snow, thereby causing the model to remove snow loads prematurely. The ocean model is a Jet Propulsion Laboratory (JPL) run of the estimating the circulation and climate of the ocean (ECCO) general circulation model (Lee et al. 2002), forced with NCEP reanalysis winds, as well as thermal and salinity fluxes; it assimilates sea surface heights from the Ocean Topography Experiment (TOPEX), and temperature profiles from XBT, Tropical Atmosphere Ocean (TAO) array, and World Ocean Circulation Experiment (WOCE) cruises. To mimic the process used by the GRACE Project to dealias the raw data, we remove the output of a barotropic ocean model (Ali and Zlotnicki 2003) from the ECCO results. ECMWF meteorological fields have been used to remove atmospheric mass variability from the GRACE field during the processing stage. We model errors due to residual atmospheric effects by taking the difference between the ECMWF and NCEP global-gridded surface pressure fields, divided by $\sqrt{2}$.

b. Discharge

The water balance of two river basins is studied in this work: the Mississippi River basin and the Ohio-Tennessee River basin. Measurements of river discharge for the Mississippi River are made by the U.S. Army Corp of Engineers (USACE), while those for the Ohio-Tennessee River are made by the U.S. Geological Survey. It is assumed that all the discharge through the basin boundaries is measured by gauging stations. However, because of tidal effects, it is not possible to

obtain discharge data from a station at the mouth of the Mississippi River. To avoid tidal effects, discharge data were chosen from gauging stations that were roughly 150 miles inland from the Gulf of Mexico. Runoff originating after this point will therefore be unmeasured, but we assume that its contribution to the total basin discharge is negligible. The Mississippi River is heavily controlled to obviate flooding risks. As part of this system of control structures, water from the Mississippi is at times diverted from the Mississippi to the Atchafalaya River. Therefore, discharge measurements from both the Mississippi, at Tarbert Landing, and the Atchafalaya, at Simmesport, were combined to estimate the total outflow from the Mississippi River basin. At these stations, measurements of discharge are made approximately every few days using acoustic Doppler current profiler (ADCP) meters; a stage–discharge relation is not employed. Discharge estimates at times between measurements are obtained by interpolation. Morlock (1996) estimated the errors in ADCP discharge measurements from 12 evaluation sites to be 1%–7% of the measured discharge. Simpson and Oltmann (1993) determined the typical accuracy in an ADCP measurement to be better than 2% of the mean discharge of large rivers. Based on these studies, we chose to model the errors in the daily discharge estimates for the Mississippi River as a normally distributed random number having a value of 2% of the discharge estimate. Because these discharge estimates are made very frequently without the use of a stage–discharge relation, we assume that there are no biases in the data.

Discharge estimates for the Ohio–Tennessee at Metropolis, Illinois, are also currently made with ADCP meters; however, these measurements are made much less frequently than those for the Mississippi. Typically, a few measurements are made each year, from which a discharge rating curve is derived. At the Metropolis gauge, measurements of the gauge height are combined with measurements of the fall ratio between Metropolis and dam 53 in a rating table to determine the discharge. Because the rating curve is only adjusted when measurements are made, errors in discharge estimates derived from the rating curve are likely to be correlated in time; that is, the discharge estimates are likely to contain systematic errors. Rather than attempt to estimate possible systematic effects in the discharge data, which is beyond the scope of this study, we model the error in the daily discharge estimates at the Metropolis gauge as a normally distributed random number having a value of 7% of the discharge estimate. In addition, we note that the largest source of uncertainty in the estimation of the seasonal cycle of discharge is due to unmodeled

variability at periods other than the annual and semi-annual.

c. Precipitation

The GPCP produces global analyses of monthly precipitation derived from satellite and gauge measurements. Infrared and microwave satellite estimates of precipitation are merged with rain gauge data from more than 6000 stations to produce global monthly mean precipitation data on a $2.5^\circ \times 2.5^\circ$ grid (Huffman et al. 1997). We chose to use the GPCP dataset because in addition to the analyses themselves, the GPCP provides gridded, monthly error estimates for the precipitation fields. The error estimation procedure (Huffman 1997) takes into account both sampling and measurement-algorithm effects. Because systematic effects are removed from the individual datasets prior to the merging process, bias effects in the final product are considered small relative to random errors, and are therefore neglected. This result is supported by comparisons with estimates from fifteen $2.5^\circ \times 2.5^\circ$ validation cells, which showed bias errors to be approximately an order of magnitude smaller than random errors (Huffman et al. 1997).

6. Results

Comparison with model output

In this section we use (12) to compare the output from a general circulation model (GCM) and a land surface model (LSM) to areally averaged estimates of precipitation minus evapotranspiration made from a combination of GRACE water storage estimates and daily observations of river discharge for the Mississippi and Ohio–Tennessee River basins. Because the model output spanned the period from the beginning of 2002 to April 2004, our results incorporated only the 18 GRACE gravity fields corresponding to this time period.

Figure 2 shows the averaging kernels used to infer the GRACE water storage estimates and errors for the Mississippi (top) and Ohio–Tennessee (bottom) River basins. The exact spatial average of a region is described by an averaging kernel having a value of 1 everywhere inside the region, and 0 everywhere outside the region. However, the presence of GRACE measurement errors requires the deweighting of the short wavelength features of the averaging kernel [see Swenson and Wahr (2002) for a detailed discussion on the construction of optimal GRACE averaging kernels], resulting in a smoothed version of the exact averaging kernel. The smoothness of an averaging kernel is de-

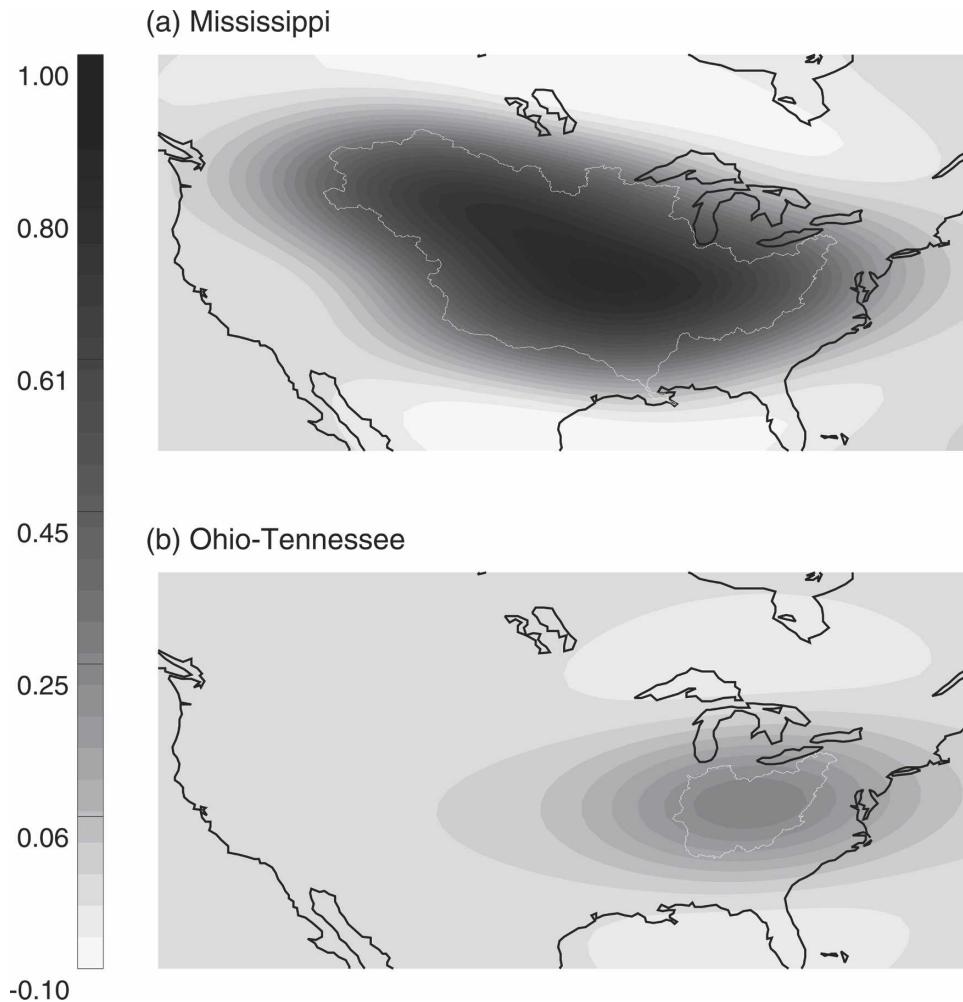


FIG. 2. Averaging kernels used to extract spatially averaged terrestrial water storage estimates from (a) the Mississippi River basin and (b) the Ohio–Tennessee River basin.

terminated by the dominant spatial scale of the basin and the relative strengths of the water storage signal and the GRACE errors. Thus, while the Mississippi averaging kernel requires relatively little smoothing, the averaging kernel for the Ohio–Tennessee River basin, with an area roughly 5 times smaller than the area of the Mississippi River basin, requires a much greater degree of smoothing to produce an averaging kernel capable of extracting a water storage estimate that is not dominated by GRACE measurement errors.

A time series of observed $P - ET$ was constructed by combining daily observations of discharge with water storage estimates computed from GRACE data, using these averaging kernels. Each point in the time series was determined by first taking the difference between two consecutive GRACE water storage estimates divided by the period of time between the first and second averaging periods, that is, the lhs of (13), then add-

ing to that value the accumulated discharge quantities, that is, the rhs of (13) where A now represents discharge. A time series for each model was also constructed by using the models' estimates of the flux of $P - ET$ in the rhs of (13). The expression " $P - ET$ " in this section thus should be understood by the reader to refer to a quantity calculated from (13).

Following Stepaniak (2004), we used GCM output in the form of 6-hourly, global, gridded, sigma-level fields of pressure, specific humidity, and wind velocity from the NCEP/DOE R-2 (Kanamitsu et al. 2002) to compute $P - ET$ from the vertically integrated moisture convergence and precipitable water tendency terms in the atmospheric water budget [Eq. (1)]. Calculation of the moisture convergence was done in the spectral domain, with a final spatial resolution of 128×64 .

Estimates of $P - ET$ were also obtained from 3-hourly, $1^\circ \times 1^\circ$ output of the Noah LSM (Ek et al.

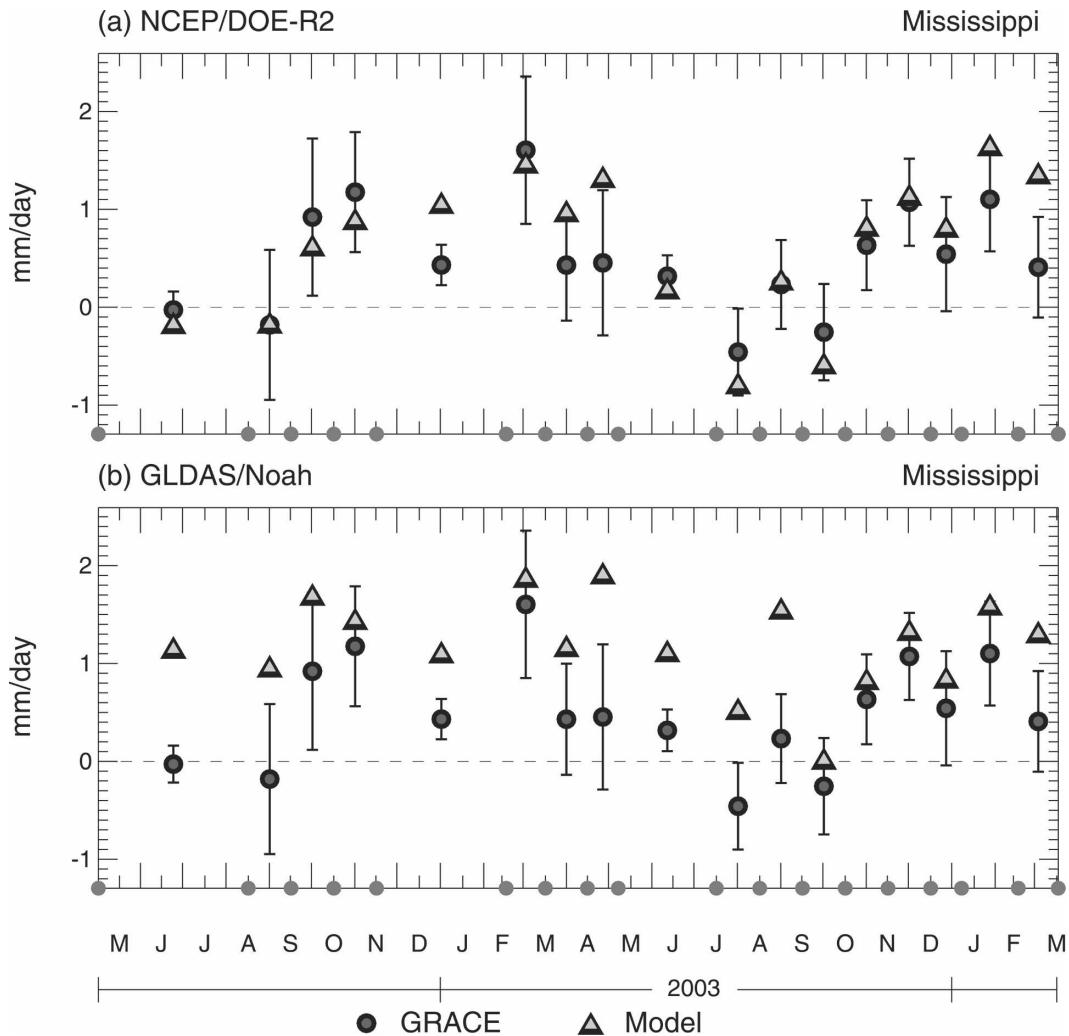


FIG. 3. Comparison of GRACE plus observed discharge-derived estimates of $P - ET$ (circles) to model-derived estimates of $P - ET$ (triangles) for the Mississippi River basin. (a) Estimates computed from NCEP/DOE R-2 GCM output; (b) estimates computed from GLDAS/Noah LSM output. The x axis shows time in months from the beginning of 2002; y axis shows $P - ET$ in mm day^{-1} of water. Small circles on the x axis represent the months for which GRACE water storage estimates are available.

2003) produced with the GLDAS (Rodell et al. 2004a). The components of the terrestrial water budget provided by the Noah LSM were accumulated rainfall and snowfall, baseflow-groundwater, and storm-surface runoff, evapotranspiration, plant-canopy surface water, soil moisture content (to 2-m depth), and accumulated snow.

After temporally aggregating the 3- and 6-hourly values using (12), each $P - ET$ dataset was spatially averaged to produce a time series for both the Mississippi and Ohio-Tennessee River basins. Figure 3 shows the time series of $P - ET$ for the Mississippi River basin computed from each of the datasets. The NCEP/DOE R-2 time series is plotted in the top panel, while the

GLDAS/Noah time series is plotted in the bottom panel. Both model results are represented by blue triangles.

The GRACE plus observed discharge results, which we refer to as GRACE/D, are represented by red circles in both the top and bottom panels. The R-2 time series shows a strong seasonal pattern, with positive $P - ET$ from the beginning in the fall, reaching a maximum during the winter, and declining until midsummer, when a shorter period of negative $P - ET$ begins. Similar behavior is exhibited by the GRACE/D time series. The mean of the R-2 time series is 0.55 mm day^{-1} with an rms variation of 0.69 mm day^{-1} , and the GRACE time series has a mean of 0.49 mm day^{-1} with an rms of

TABLE 1. $P - ET$ (mm day⁻¹).

Mississippi	GRACE/D	R-2	Noah
Mean	0.49	0.55	1.22
Rms	0.54	0.69	0.47
Ohio-Tennessee	GRACE/D	R-2	Noah
Mean	1.25	1.44	2.68
Rms	0.90	1.17	0.85

0.54 mm day⁻¹ (Table 1). The rms of the total error in the GRACE/D time series is 0.55 mm day⁻¹, and is dominated by error in the GRACE water storage estimates, error in the accumulated discharge estimates being negligible (Table 2). Of the three components of error in the water storage estimates, instrument error (or satellite measurement error) is the largest (0.53 mm day⁻¹), followed by leakage error (0.22 mm day⁻¹), and error due to residual atmospheric gravity signals (0.11 mm day⁻¹). All but three of the R-2 data points lie within the error bars of the GRACE/D results, and there appears to be no systematic offset relative to the GRACE/D results.

The Noah time series, plotted in the bottom panel, agrees less well with the GRACE/D results. A positive bias is seen in the Noah time series relative to both the R-2 and GRACE/D results. In addition, the seasonality exhibited by the R-2 and GRACE/D time series is not readily apparent in the Noah results. The mean of the Noah time series is 1.22 mm day⁻¹ and the rms is 0.47 mm day⁻¹. The fact that all the Noah $P - ET$ values are larger than the GRACE/D $P - ET$ values indicates that the Noah model systematically overestimates the flux of $P - ET$ relative to GRACE/D.

The $P - ET$ time series for the Ohio-Tennessee River basin exhibit seasonality similar to that of the Mississippi River basin, but with greater amplitudes (Fig. 4). The mean of the R-2 time series is 1.44 mm day⁻¹ with an rms variation of 1.17 mm day⁻¹, and the GRACE/D time series has a mean of 1.25 mm day⁻¹

with an rms of 0.90 mm day⁻¹. The rms of the total error in the GRACE/D time series is 0.63 mm day⁻¹. The mean of the Noah time series is 2.68 mm day⁻¹ and the rms is 0.65 mm day⁻¹. In both cases, the agreement with the GRACE/D results is poorer, and more points of the model time series lie outside the GRACE error bars.

One source of the apparent positive bias shown in the Noah results may be the precipitation data used to force the model. The output used in this study was obtained by forcing the Noah LSM with precipitation from the GDAS, NCEP's operational global atmospheric data assimilation system. Figure 5 compares estimates of precipitation rate from GDAS to the CPC's gauge-only, real-time precipitation analysis (CPC-G), the CPC's Merged Analysis of Precipitation (CMAP), and the GPCP satellite-gauge product. The CPC-G data is available at daily intervals, while CMAP and GPCP data are available as monthly averages. The three alternate precipitation datasets generally agree with one another, while the GDAS precipitation is typically larger, especially during the spring and summer months. The greater GDAS spring/summer precipitation rate is consistent with the greater Noah spring/summer $P - ET$ if the river basins considered here are energy-limited domains, in which case evaporation already occurs at near the potential rate and excess precipitation cannot be evaporated.

7. Estimating monthly averaged fluxes

In situations such as when the agreement between model output and GRACE/D results is poor or when observations lack the temporal resolution needed to compute the quantities in (12), it would be desirable to use GRACE and observed discharge data to directly estimate the flux of $P - ET$. Changes in instantaneous water storage values possess a direct relationship to the flux terms in the water balance equation; the difference in storage at two times divided by the period between those times is simply the time-averaged flux of $P - ET - R$. Changes in GRACE water storage, on the other hand, lack a unique, physically meaningful relationship with the fluxes of precipitation, evapotranspiration, and runoff; this is expressed by the rhs of (12), which is composed of terms with different physical interpretations. The first term represents the total accumulation over a time period from the beginning of one GRACE averaging period to the beginning of a second GRACE averaging period, while the second and third terms represent the difference between the average accumulations during the first and second averaging periods. If the instantaneous values of water storage in (8) are

TABLE 2. GRACE/D components and errors (mm day⁻¹).

$P - ET$ components	Mississippi	Ohio-Tennessee
GRACE ($\Delta S/\Delta T$) mean	0.07	0.07
GRACE ($\Delta S/\Delta T$) rms	0.50	0.60
Discharge mean	0.42	1.18
Discharge rms	0.18	0.60
$P - ET$ errors	Mississippi	Ohio-Tennessee
Total	0.55	0.63
Discharge	0.01	0.02
GRACE ($\Delta S/\Delta T$)	0.55	0.63
Instrument	0.53	0.53
Leakage	0.22	0.38
Residual atmosphere	0.11	0.08

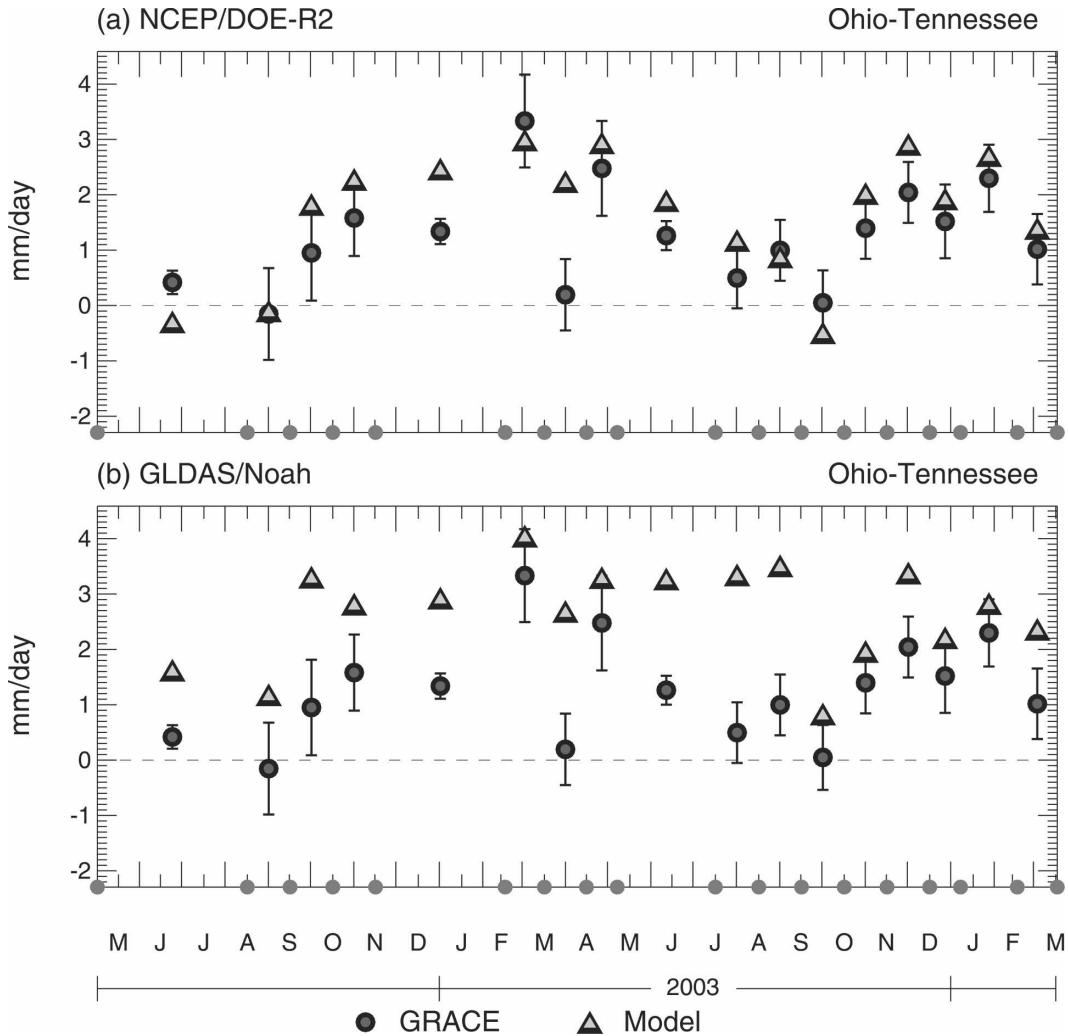


FIG. 4. Same as Fig. 3 except for the Ohio-Tennessee River basin.

replaced by temporally averaged values determined from GRACE, one can infer approximate values of the time-averaged flux terms,

$$\langle A \rangle_T \approx \frac{\langle S \rangle_{43} - \langle S \rangle_{21}}{T}, \quad (14)$$

where $t_{43} = (t_4 + t_3/2)$ and $t_{21} = (t_2 + t_1/2)$ are the times corresponding to the middle of each GRACE averaging period, and $T = t_{43} - t_{21}$. The accuracy of (14) will depend on the degree to which the differences between temporally averaged values represent the differences in the instantaneous values. In this section, we examine the accuracy of this approximation using model output. Because the GLDAS/Noah model explicitly computes storage, the approximation may be assessed by comparing the storage terms directly:

$$\frac{S(t_{43}) - S(t_{21})}{T} \approx \frac{\langle S \rangle_{43} - \langle S \rangle_{21}}{T}. \quad (15)$$

Water storage is not part of NCEP/DOE R-2 model output; therefore we assess the accuracy of (14) via the corresponding flux terms from Eqs. (8) and (12),

$$\begin{aligned} \langle P - ET - R \rangle_T \approx & \frac{T_{31}}{T} \langle P - ET - R \rangle_{31} \\ & + \frac{\langle \Phi^{P-ET-R} \rangle_{43} - \langle \Phi^{P-ET-R} \rangle_{21}}{T}, \end{aligned} \quad (16)$$

using R-2 model output and observed discharge.

Figures 6a and 6b compare the changes in instantaneous storage at the middle of each month to the changes in the monthly averaged storage for each month for the Mississippi and Ohio-Tennessee basins respectively, computed from GLDAS/Noah output. The root-mean-squares of the instantaneous storage changes [lhs of (15)] are 0.45 mm day^{-1} for the Missis-

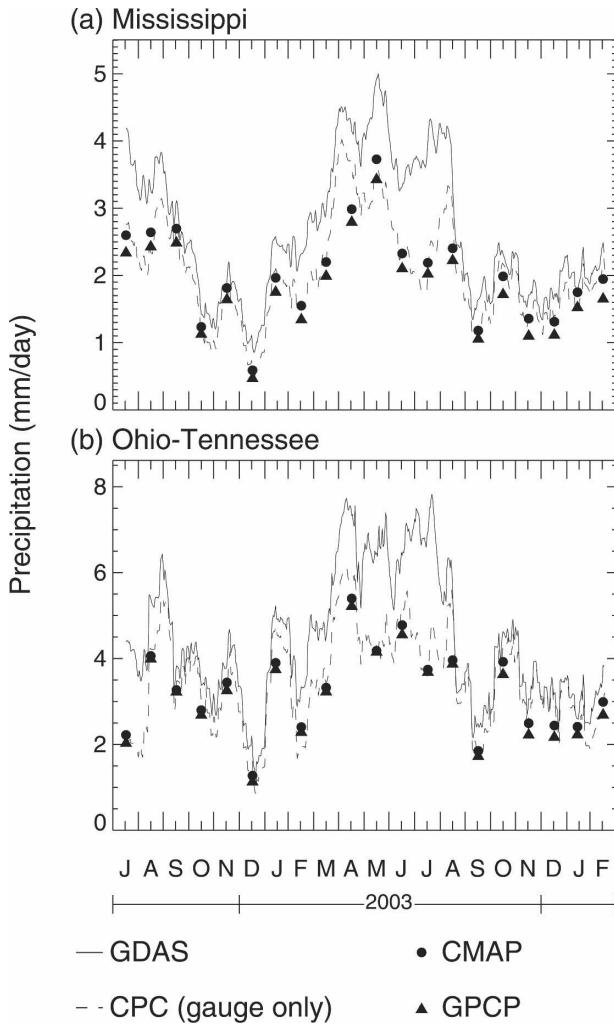


FIG. 5. Comparison of precipitation estimates from GDAS operational analyses (solid line), CPC gauge-only analysis (dashed line), CMAP satellite-gauge merged analysis (circles), and GPCP satellite-gauge merged analysis (triangles). Estimates for the (a) Mississippi River basin and (b) Ohio-Tennessee River basin. The x axis shows time in months from July 2002; y axis shows precipitation rate in mm day^{-1} .

sippi and 0.69 mm day^{-1} for the Ohio-Tennessee. The rms differences between the instantaneous [lhs of (15)] and monthly averaged [rhs of (16)] storage time series are 0.19 mm day^{-1} for the Mississippi and 0.42 mm day^{-1} for the Ohio-Tennessee. Figures 6c and 6d compare the monthly averaged flux of $P - ET - R$ to the corresponding quantity expressed by the rhs of (16). The rms values of the monthly averaged fluxes [lhs of (16)] are 0.79 mm day^{-1} for the Mississippi and 1.28 mm day^{-1} for the Ohio-Tennessee. The rms differences between the respective time series are 0.25 mm day^{-1} for the Mississippi and 0.54 mm day^{-1} for the Ohio-Tennessee. Fitting and removing a seasonal cycle

to the difference of the time series in each of the four cases reduces the variance negligibly. Therefore, the error in approximation (14) may be considered essentially random, with an rms between 32% and 42% of the rms of the month-to-month changes in the instantaneous water storage for the Mississippi and between 42% and 61% for the Ohio-Tennessee. Adding this error in quadrature to the errors in GRACE water storage changes and the errors in the discharge estimates for these basins (Table 2) would result in estimates of monthly averaged $P - ET$ having a total rms error of $0.58\text{--}0.60 \text{ mm day}^{-1}$ for the Mississippi and $0.69\text{--}0.75 \text{ mm day}^{-1}$ for the Ohio-Tennessee. The errors in the GRACE water storage changes are larger than the additional errors due to approximating the instantaneous storage changes by the corresponding time-averaged storage changes, and therefore are the limiting source of error. Should the accuracy in the GRACE data improve to prelaunch baseline estimates, which are at the subcentimeter level for these basins (Swenson et al. 2003), the approximation error would then become the limiting error source in estimates of the time-averaged rate of precipitation minus evapotranspiration.

8. Estimating seasonal fluxes

In this section, we use GRACE data and river discharge observations to determine the seasonal cycle of $P - ET$. The coefficients of the seasonal cycle of $P - ET$ can be estimated with more certainty than monthly averaged values, without the need for the approximation used in the previous section. We use five terms to decompose the seasonal cycle of any variable, $x(t)$, into a mean, an annual cosine, an annual sine, a semiannual cosine, and a semiannual sine:

$$x(t) = x_0 + \sum_{i=1}^2 [C_i^x \cos(\omega_i t) + S_i^x \sin(\omega_i t)], \quad (17)$$

where x_0 is the variable's time-mean value, $\omega_1 = (2\pi/\text{year})$ is the annual frequency, $\omega_2 = 2\omega_1$ is the semiannual frequency, and C_i^x and S_i^x are the amplitudes of the cosine and sine terms, respectively. The parameters x_0 , C_i^x , and S_i^x are estimated by a least squares fitting procedure.

Equation (17) can be combined with the equation for the instantaneous water balance,

$$\frac{dS}{dt} = P - ET - R, \quad (18)$$

to relate the amplitudes of the seasonal cycles of water storage, discharge, and $P - ET$:

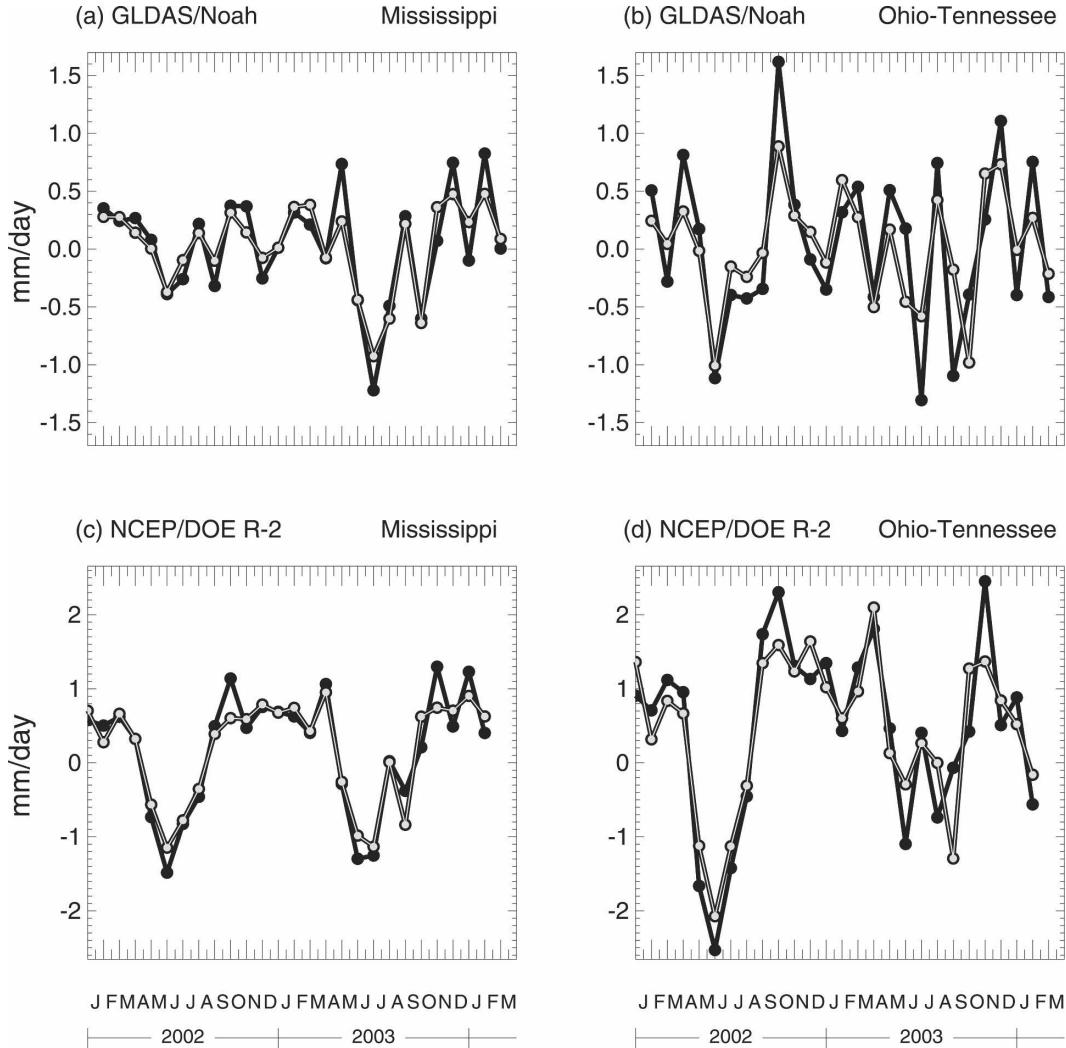


FIG. 6. (top) Comparison of the changes in instantaneous water storage to changes in monthly averaged water storage for the (a) Mississippi and (b) Ohio-Tennessee River basins, estimated from the Noah LSM. (bottom) Comparison of the monthly averaged flux of $P - ET - R$ to the quantity described by the rhs of (16) for the (c) Mississippi and (d) Ohio-Tennessee River basins, estimated from the R-2 GCM. Black lines represent changes in instantaneous water storage; gray lines represent changes in temporally averaged water storage. The x axis shows time in months from the beginning of 2002; y axis has units of mm day^{-1} of water.

$$\begin{aligned}
 P_0 - ET_0 &= R_0 + S_t, \\
 C_i^{P-ET} &= C_i^R + \omega_i S_i^{\text{WS}}, \\
 S_i^{P-ET} &= S_i^R - \omega_i C_i^{\text{WS}}. \quad (19)
 \end{aligned}$$

The term S_t , which describes a linear trend in the storage time series, is added because the storage term is related to the fluxes through its derivative.

River discharge observations and GRACE data both represent time-averaged quantities. Because their averaging period is small compared to both the annual and semiannual time scales, daily discharge observations may be used with (17) to determine the coefficients of

the seasonal cycle of discharge. The averaging periods of GRACE data are typically a month or longer, and therefore (17) must be modified by first temporally averaging each term over the time periods corresponding to each GRACE field before fitting

$$\begin{aligned}
 S_n &= S_0 + S_t \bar{t}_n + \sum_{i=1}^2 \\
 &\times \left[C_i^{\text{WS}} \frac{1}{T_n} \int dt \cos(\omega_i t) + S_i^{\text{WS}} \frac{1}{T_n} \int dt \sin(\omega_i t) \right], \quad (20)
 \end{aligned}$$

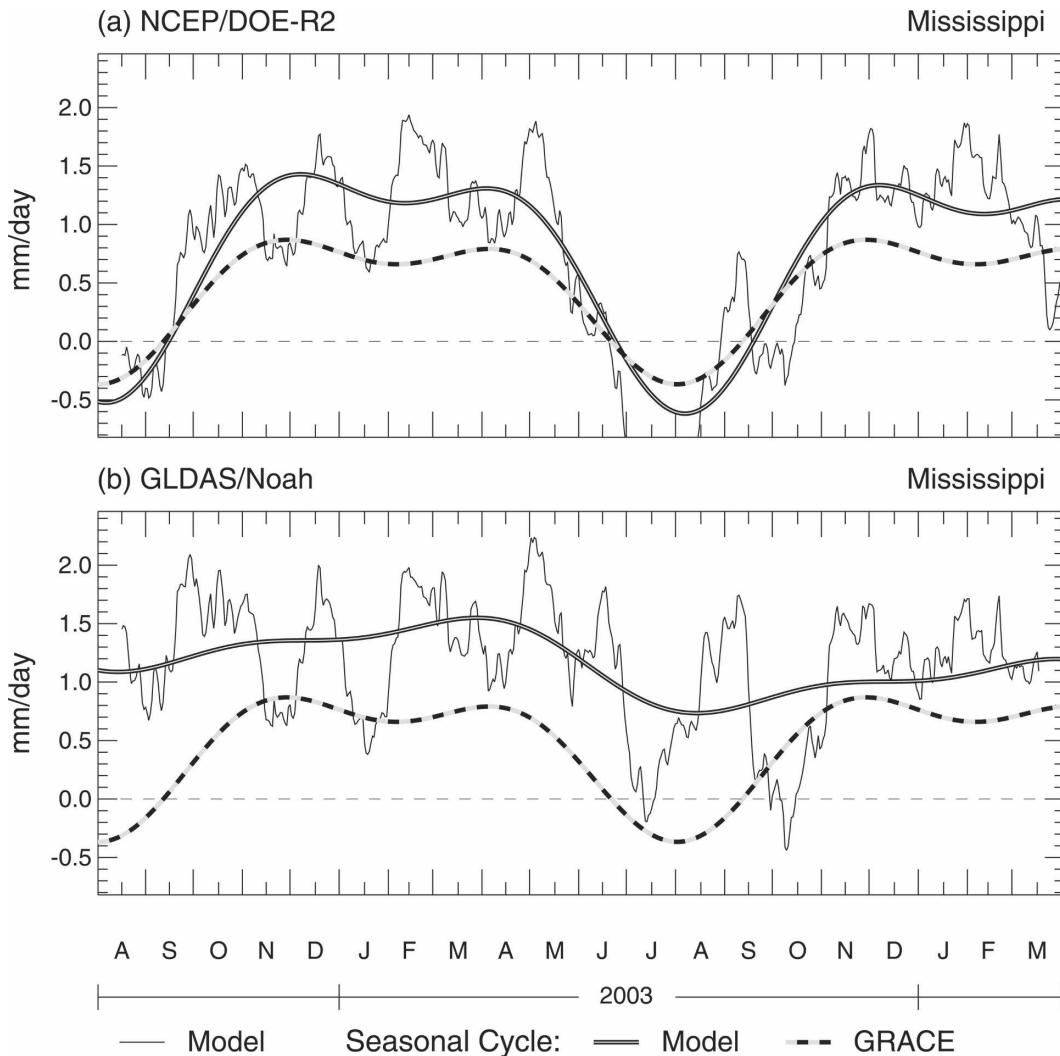


FIG. 7. Comparison of the best-fitting seasonal cycles of $P - ET$. Results for (a) R-2 and GRACE/D and (b) Noah and GRACE. Black lines are daily model values, smoothed with a 30-day running filter. Gray lines are best-fitting seasonal cycles from model output; dashed lines are those derived from a combination of GRACE and observed discharge data. The x axis shows time in months from the beginning of 2002; y axis shows flux of $P - ET$ in mm day^{-1} .

where the index n represents each GRACE field, T_n is the time period sampled by each GRACE field, S_i is a linear trend, and \bar{t}_n is the average time during the period T_n . The amplitudes C_i^{WS} and S_i^{WS} obtained by fitting the resulting S_n time series to the GRACE data thus represent the seasonal cycle of instantaneous water storage, not the seasonal cycle of time-averaged water storage, and may be used in (19).

Figure 7 compares the best-fitting seasonal cycles computed from GRACE and observed discharge data with those computed from model output for the Mississippi River basin. The R-2 results, shown in the top panel, generally agree with the GRACE/D results. The black line represents the daily values of $P - ET$,

smoothed with a running 30-day filter. The blue line shows the seasonal cycle of $P - ET$ from the R-2 model output, while the red line shows the seasonal cycle obtained from a combination of GRACE water storage estimates and observed river discharge data. The seasonal coefficients of $P - ET$ for GRACE/D and each of the models, as well as δG , the uncertainty in each of the GRACE/D seasonal coefficients, are shown in Table 3. Table 3 shows the $P - ET$ seasonal components expressed as amplitude and phase. The phase convention used here is given by $x(t) = \cos(\omega t - \phi)$, and the units are converted to days from 1 January. From Table 3 it is apparent that both the annual and semiannual phases of the GRACE/D seasonal cycle compare very well

TABLE 3. GRACE/D seasonal $P - ET$ amplitude (mm day^{-1}) and phase (days).

Mississippi	GRACE/D	δG	R-2	Noah
Annual amplitude	0.52	0.11	0.95	0.22
Annual phase	27	19	17	62
Semi-annual amplitude	0.29	0.27	0.38	0.12
Semi-annual phase	-60	30	-66	-70
Mean	0.44	0.01	0.65	1.16
Ohio-Tennessee	GRACE/D	δG	R-2	Noah
Annual amplitude	0.70	0.13	1.21	0.41
Annual phase	44	29	35	86
Semi-annual amplitude	0.44	0.34	0.82	0.49
Semi-annual phase	-39	27	-45	-34
Mean	1.15	0.03	1.84	2.70

with those of the R-2 seasonal cycle. The GRACE/D $P - ET$ annual phase for the Mississippi has a value of 30 days and the R-2 annual phase has a value of 28 days. The values for the semiannual phase are -63 and -66 days, respectively. The semiannual amplitudes also agree, having values of 0.85 mm day^{-1} for GRACE/D and 1.15 mm day^{-1} for R-2. The annual amplitudes are more disparate; the GRACE results have an amplitude of 1.58 mm day^{-1} , while the R-2 results have an amplitude of 2.86 mm day^{-1} . The annual cycle of $P - ET$ derived from the Noah model output, shown in the bottom panel of Fig. 7, compares well with neither GRACE/D nor R-2. Its annual amplitude and phase are 0.66 mm day^{-1} and 62 days. The Noah semiannual amplitude, 0.37 mm day^{-1} , is much smaller than either GRACE/D or R-2, but its phase, -70 days, is nearly the same. The means of each time series are 1.32, 1.93, and 3.48 mm day^{-1} for GRACE/D, R-2, and Noah, respectively. The GRACE/D mean is dominated by the mean discharge, the trend in the GRACE storage time series, S_t , being negligible.

Best-fitting seasonal cycles for GRACE/D and model data for the Ohio-Tennessee River basin are shown in Fig. 8. Again, the GRACE/D and R-2 seasonal cycles of $P - ET$ agree well, especially in the annual (41/35 days) and semiannual (-37/-45 days) phases. The R-2 amplitudes ($3.62/2.46 \text{ mm day}^{-1}$) are again larger than those of GRACE/D ($2.05/1.47 \text{ mm day}^{-1}$). The Noah time series has a smaller annual amplitude, 1.24 mm day^{-1} , and greater annual phase, 86 days, than GRACE/D, but its semiannual cycle agrees well with GRACE, having an amplitude of 1.46 mm day^{-1} and a phase of -34 days.

A visual inspection of the time series of daily $P - ET$ from the two models, shown in Figs. 7 and 8, reveals that the R-2 and Noah results agree better in the winter months than in the summer months. The absence in the Noah results of a period of negative net precipitation in

the summer explains the lack of agreement between the Noah seasonal cycle of $P - ET$ and those for R-2 and GRACE/D. As mentioned earlier, this may indicate that the precipitation used to force the land surface model is too large in the spring and summer months.

Potential accuracy in estimating seasonal evapotranspiration as a residual

After one estimates areally averaged $P - ET$, it is a straightforward matter to subtract those estimates from areally averaged measurements of precipitation to obtain evapotranspiration as a residual:

$$ET = P - R - \frac{dS}{dt}. \quad (21)$$

Rodell et al. (2004b) compared monthly GRACE-derived estimates of ET to that from a number of models, finding good agreement. However, because of the relatively large error bars associated with the monthly ET values, discriminating between the models is difficult.

Here, we examine the seasonal cycle of ET, which is related to the seasonal coefficients of P , R , and water storage by

$$\begin{aligned} ET_0 &= P_0 - R_0 - S_t, \\ C_i^{ET} &= C_i^P - C_i^R - \omega_i S_i^{WS}, \\ S_i^{ET} &= S_i^P - S_i^R + \omega_i C_i^{WS}. \end{aligned} \quad (22)$$

We first estimated the seasonal cycle of precipitation by fitting the monthly GPCP precipitation estimates using (20). From this seasonal cycle of P , we removed the seasonal cycle of $P - ET$ derived from GRACE and observed discharge; we refer to this estimate of seasonal ET as "GDP." The seasonal cycle of ET from GDP was then compared to that computed from the Noah model. (We did not compare GDP ET with R-2 data because ET is not a basic output of R-2. Evapotranspiration is estimated from R-2 or GRACE/D $P - ET$ by the same process, and therefore any differences between the R-2 ET results and the GDP ET results are already present in the respective $P - ET$ time series.)

Figure 9 compares the ET estimates for the (a) Mississippi and (b) Ohio-Tennessee. The black line represents the daily values of ET taken from Noah, smoothed with a running 30-day filter. The blue line shows the best-fitting Noah seasonal cycle of ET, while the red line shows the seasonal cycle obtained from removing the GRACE estimate of $P - ET$ from the GPCP precipitation data. Units are centimeters per month. While the GRACE $P - ET$ estimates did not agree well with those of Noah (Figs. 7 and 8), the re-

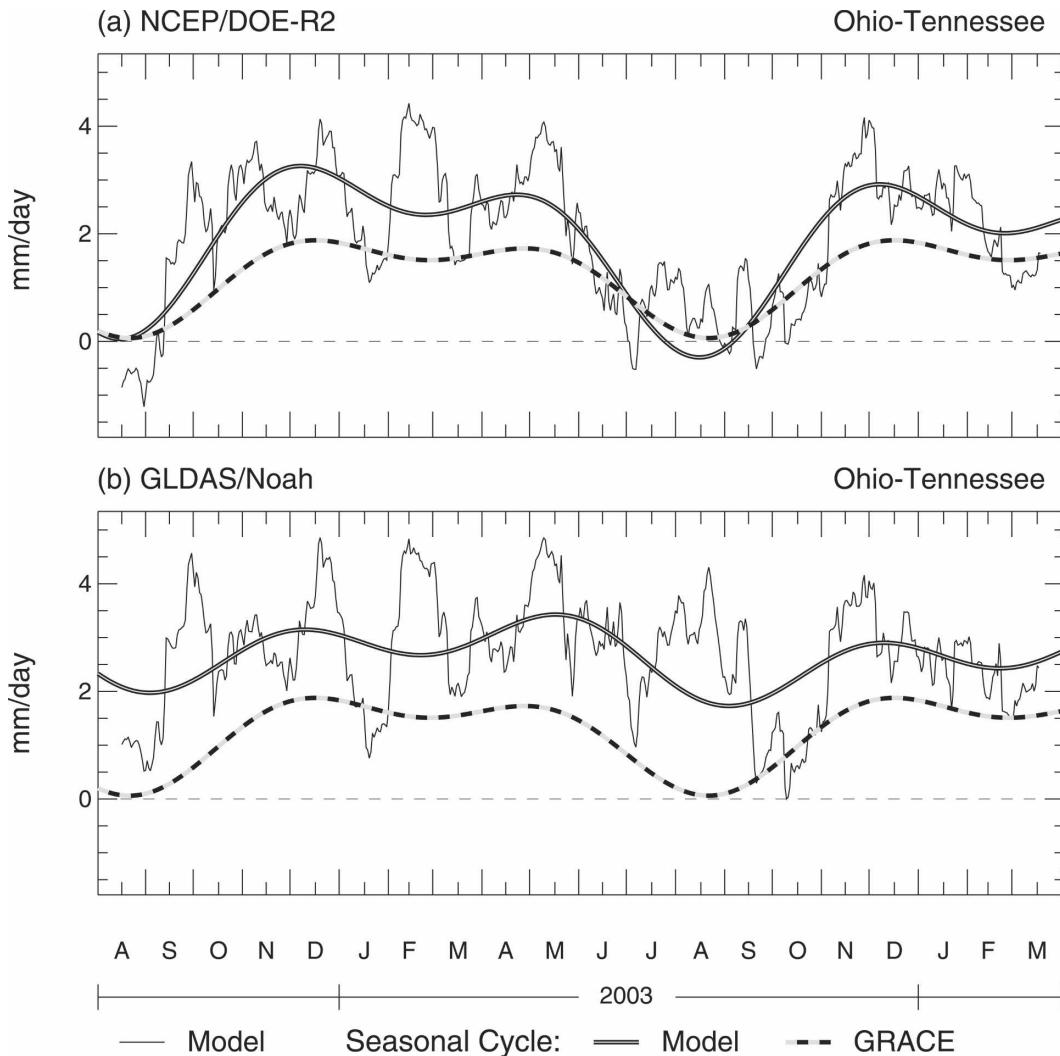


FIG. 8. Same as Fig. 7, except that it is for the Ohio–Tennessee River basin.

spective ET time series do show good agreement. In each river basin, the seasonal cycle of ET is dominated by the annual cycle. For the Mississippi, the GRACE ET time series has an annual amplitude and phase of 4.07 mm day^{-1} and -176 days compared to the Noah time series values of 4.50 mm day^{-1} and -178 days (Table 4). The semiannual amplitudes and phases are 0.38 mm day^{-1} and 10 days for GRACE and 0.37 mm day^{-1} and 7 days for Noah. The Noah time series has a somewhat larger mean, 5.25 mm day^{-1} , than the GRACE time series, 4.80 mm day^{-1} .

Similar agreement was obtained for the Ohio–Tennessee, although semiannual phases do not agree. This is not surprising, given the small amplitude of the semiannual cycle. The annual amplitude of the Noah time series, 5.51 mm day^{-1} , is again larger than that of GRACE, 4.62 mm day^{-1} , while the annual phases are

nearly equal, having values of -175 and -174 days, respectively.

9. Summary

In this study, we examine the accuracies of $P - ET$ estimates made from a combination of GRACE and observed river discharge, and use these observationally based $P - ET$ estimates to validate model output. Because of its connection to both the atmospheric and terrestrial water budgets, $P - ET$ can be used to assess both general circulation models and land surface models. Models that simulate the interaction between the earth’s land surface and atmosphere continue to evolve, incorporating more sophisticated process physics and finer temporal and spatial resolutions at which those processes are represented. These new capabilities are

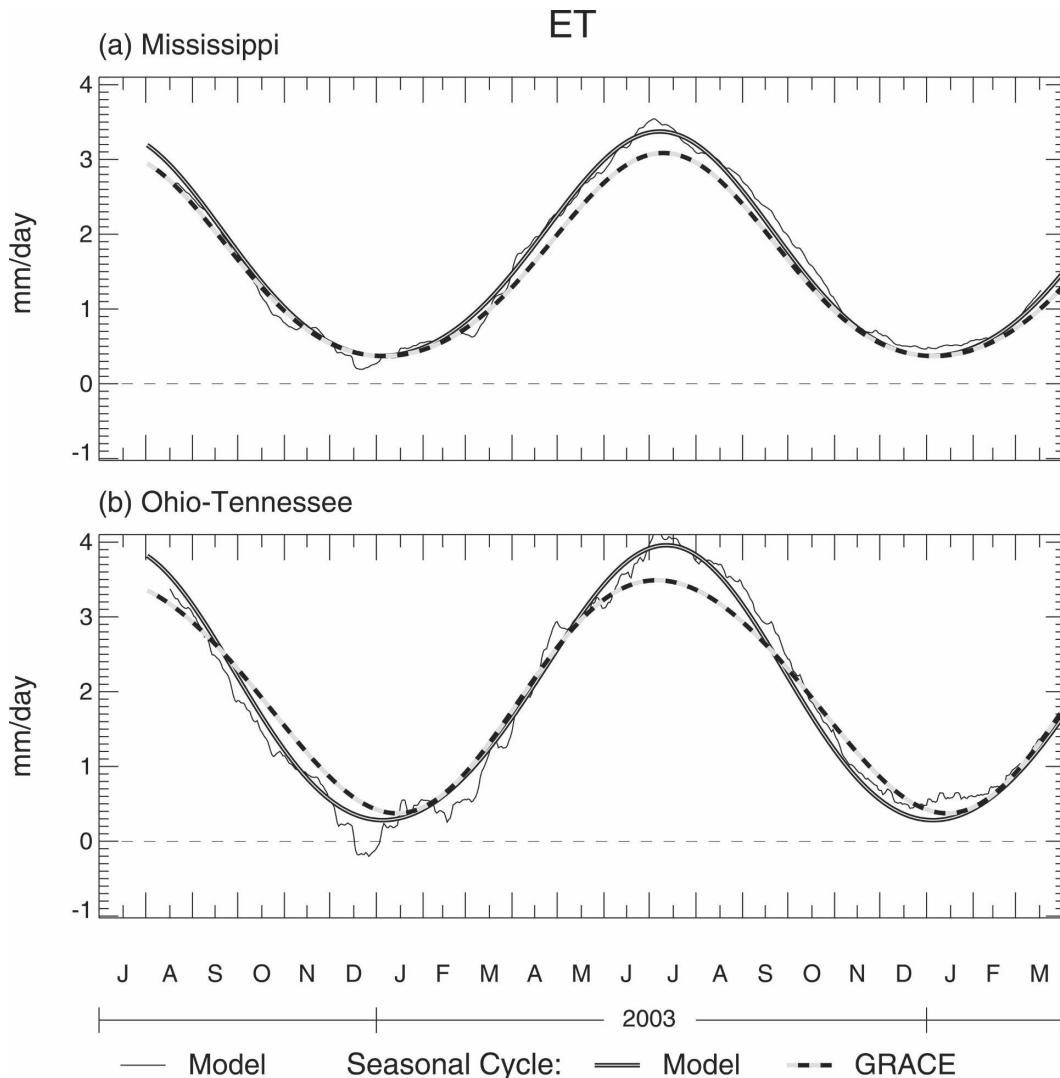


FIG. 9. Comparison of the best-fitting seasonal cycles of ET from GRACE and Noah. Results for the (a) Mississippi River basin and (b) Ohio-Tennessee River basin. Gray lines are best-fitting seasonal cycles from model output; dashed lines are those derived from a combination of GRACE and observed precipitation and discharge data. The x axis shows time in months from the beginning of 2002; y axis shows flux of $P - ET$ in cm of water per month.

useful if they provide more accurate descriptions of the present and future state of the earth's water and energy budgets. However, uncertainties in model output are difficult to assess because of a lack of observations. Spatially averaged water storage changes from GRACE combine the attractive features of global, homogeneous spatial coverage and the ability to provide quantitative error estimates.

We first present a water balance equation corresponding to the temporal sampling characteristics of GRACE. Using this water balance formulation, we found GRACE/D $P - ET$ estimates compare favorably with those computed from reanalysis-2 output. Signifi-

cant differences between the GRACE/D results and those computed from the Noah LSM exist, most notably due to the presence of a positive bias in the Noah results relative to GRACE/D.

We also examine the likely error introduced by the interpretation of changes in GRACE water storage estimates as a time-averaged flux. In addition to being a validation tool, using the difference between two GRACE water storage estimates to approximate the difference in the instantaneous water storage between two times allows one to reinterpret the GRACE results as a time-averaged flux of $P - ET - R$. Based on a comparison using model output, we estimate that this

TABLE 4. Evapotranspiration amplitude (mm day⁻¹) and phase (days).

Mississippi	GRACE	δG	Noah
Annual amplitude	1.36	0.21	1.50
Annual phase	-176	20	-178
Semi-annual amplitude	0.13	0.31	0.12
Semi-annual phase	10	37	7
Mean	1.60	0.12	1.75
Ohio-Tennessee	GRACE	δG	Noah
Annual amplitude	1.54	0.25	1.84
Annual phase	-174	28	-175
Semi-annual amplitude	0.08	0.40	0.11
Semi-annual phase	-58	53	19
Mean	2.05	0.14	2.02

approximation introduces an error of roughly 1 cm or less.

Finally, we estimated the best-fitting seasonal cycles of $P - ET$ and ET from GRACE data and observed runoff and precipitation. The R-2 $P - ET$ results agreed very well with the GRACE/D $P - ET$ results, although the amplitudes were about 80% larger and the mean was about 60% larger. The Noah output had a smaller seasonal cycle of $P - ET$ than did GRACE, but was dominated by a mean that was over twice as large as that of the GRACE/D results. Both the Noah and GRACE/D seasonal cycles of ET were dominated by the annual component and compared quite well, especially in the annual phase. The Noah ET annual amplitude was slightly larger than that of the GRACE/D results.

The application of GRACE data to the estimation of surface fluxes will benefit greatly from improvement in two areas. The first is the continued increase in the resolution of the GRACE data themselves. At the current accuracy levels, this method is only applicable to large river basins, but if GRACE reaches its baseline accuracy levels, many smaller river basins around the world could be studied. The second source of improvement in the ability to widely apply this method is the availability of near-real-time river discharge data. Currently, most large rivers are gauged, but an infrastructure does not exist to supply the scientific community with this data. As discharge data become more accessible, we will be able to perform more studies of this kind.

Acknowledgments. This work was partially supported by NASA Grants NAG5-9450 and NNG04GF02G and NSF Grant EAR-0309678 to the University of Colorado. We thank Harley Winer of the USACE and Steve Darnell of the USGS for their helpful discussions regarding river discharge measurements. We would also

like to thank Matt Rodell for providing us with the GLDAS/Noah model, and for his insight into the details of the modeling process. Finally, we thank Chris Milly and two anonymous reviewers for their advice. The river basin boundaries were extracted from the HYDRO 1K Elevation Derivative Database maintained by the U.S. Geological Survey EROS Data Center from their Web site at <http://edcaac.usgs.gov/gtopo30/hydro/>.

REFERENCES

- Ali, A. H., and V. Zlotnicki, 2003: Quality of wind stress fields measured by the skill of a barotropic ocean model: Importance of stability of the marine atmospheric boundary layer. *Geophys. Res. Lett.*, **30**, 1129, doi:10.1029/2002GL016058.
- Alley, W. M., R. W. Healy, J. W. LaBaugh, and T. E. Reilly, 2002: Flow and storage in groundwater systems. *Science*, **296**, 1985–1990.
- Boegh, E., and H. Soegaard, 2004: Remote sensing based estimation of evapotranspiration rates. *Int. J. Remote Sens.*, **25**, 2535–2551.
- Berbery, E. H., and V. R. Barros, 2002: The hydrologic cycle of the La Plata basin in South America. *J. Hydrometeorol.*, **3**, 630–645.
- Diak, G. R., J. R. Mecikalski, M. C. Anderson, J. M. Norman, W. P. Kustas, R. D. Torn, and R. L. DeWolf, 2004: Estimating land surface energy budgets from space. *Bull. Amer. Meteor. Soc.*, **85**, 65–78.
- Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley, 2003: Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta Model. *J. Geophys. Res.*, **108**, 8851, doi:10.1029/2002JD003296.
- Fan, Y., and H. Van den Dool, 2004: The Climate Prediction Center global monthly soil moisture data set at 0.5° resolution for 1948 to present. *J. Geophys. Res.*, **109**, D10102, doi:10.1029/2003JD004345.
- Guetter, A. K., and K. P. Georgakakos, 1993: River outflow of the conterminous United States. *Bull. Amer. Meteor. Soc.*, **74**, 1873–1891.
- Gutowski, W. J., Y. Chen, and Z. Otles, 1997: Atmospheric water vapor transport in NCEP-NCAR reanalyses: Comparison with river discharge in the central United States. *Bull. Amer. Meteor. Soc.*, **78**, 1957–1969.
- Hollinger, S. E., and S. A. Isard, 1994: A soil moisture climatology of Illinois. *J. Climate*, **7**, 822–833.
- Huffman, G. J., 1997: Estimates of root-mean-square random error for finite samples of estimated precipitation. *J. Appl. Meteorol.*, **36**, 1191–1201.
- , and Coauthors, 1997: The Global Precipitation Climatology Project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, **78**, 5–20.
- Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP/DOE AMIP-II reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1643.
- Lee, T., I. Fukumori, D. Menemenlis, Z. Xing, and L.-L. Fu, 2002: Effects of the Indonesian Throughflow on the Pacific and Indian Oceans. *J. Phys. Oceanogr.*, **32**, 1404–1429.
- Maurer, E. P., G. M. O'Donnell, and D. P. Lettenmaier, 2001: Evaluation of the land surface water budget in NCEP/NCAR

- and NCEP/DOE reanalyses using an off-line hydrologic model. *J. Geophys. Res.*, **106** (D16), 17 841–17 862.
- Morlock, S. E., 1996: Evaluation of acoustic Doppler current profiler measurements of river discharge. U.S. Geological Survey Water-Resources Investigations, Rep. 95-4218, 37 pp.
- Roads, J. O., S.-C. Chen, A. K. Guetter, and K. P. Georgakakos, 1994: Large-scale aspects of the United States hydrologic cycle. *Bull. Amer. Meteor. Soc.*, **75**, 1589–1610.
- , M. Kanamitsu, and R. Stewart, 2002: CSE water and energy budgets in the NCEP–DOE reanalysis II. *J. Hydro-meteor.*, **3**, 227–248.
- , and Coauthors, 2003: GCIP water and energy budget synthesis (WEBS). *J. Geophys. Res.*, **108**, 8609, doi:10.1029/2002JD002583.
- Robock, A., Y. V. Konstantin, G. Srinivasan, J. K. Entin, S. E. Hollinger, N. A. Speranskaya, S. Liu, and A. Namkhai, 2000: The global soil moisture data bank. *Bull. Amer. Meteor. Soc.*, **81**, 1281–1299.
- Rodell, M., and J. S. Famiglietti, 1999: Detectability of variations in continental water storage from satellite observations of the time dependent gravity field. *Water Resour. Res.*, **35**, 2705–2723.
- , and —, 2001: An analysis of terrestrial water storage variations in Illinois with implications for GRACE. *Water Resour. Res.*, **37**, 1327–1339.
- , and Coauthors, 2004a: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.*, **85**, 381–394.
- , J. S. Famiglietti, J. Chen, S. I. Seneviratne, P. Viterbo, S. Holl, and C. R. Wilson, 2004b: Basin scale estimates of evapotranspiration using GRACE and other observations. *Geophys. Res. Lett.*, **31**, L20504, doi:10.1029/2004GL020873.
- Ropelewski, C. F., and E. S. Yarosh, 1998: The observed mean annual cycle of moisture budgets over the central United States. *J. Climate*, **11**, 2180–2190.
- Rudolf, B., 1995: The Global Precipitation Climatology Centre. *WMO Bull.*, **44**, 77–78.
- Seneviratne, S. I., P. Viterbo, D. Luthi, and C. Schar, 2004: Inferring changes in terrestrial water storage using ERA-40 reanalysis data: The Mississippi River basin. *J. Climate*, **17**, 2039–2057.
- Seo, D. J., 1998: Real-time estimation of rainfall fields using radar rainfall and rain gauge data. *J. Hydrol.*, **208**, 37–52.
- Simpson, M. R., and R. N. Oltmann, 1993: Discharge-measurement system using an acoustic Doppler current profiler with applications to large rivers and estuaries. U.S. Geological Water-Supply Paper 2395, 32 pp.
- Stepaniak, D. P., cited 2004: Vertically integrated mass, moisture, heat, and energy budget products derived from the NCEP/NCAR reanalysis. [Available online at <http://www.cgd.ucar.edu/cas/catalog/newbudgets/index.html>.]
- Swenson, S., and J. Wahr, 2002: Methods for inferring regional surface-mass anomalies from GRACE measurements of time-variable gravity. *J. Geophys. Res.*, **107**, 2193, doi:10.1029/2001JB000576.
- , —, I. Velicogna, and P. C. D. Milly, 2001: Large-scale hydrology inferred from GRACE estimates of time-variable gravity. *Eos, Trans. Amer. Geophys. Union*, **82** (Fall Meeting Suppl.), Abstract H42F-12.
- , —, and P. C. D. Milly, 2003: Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE). *Water Resour. Res.*, **39**, 1223, doi:10.1029/2002WR001808.
- Tapley, B. D., S. Bettadpur, J. C. Ries, P. F. Thompson, and M. M. Watkins, 2004a: GRACE measurements of mass variability in the Earth system. *Science*, **305**, 503–505.
- , —, M. Watkins, and Ch. Reigber, 2004b: The Gravity Recovery and Climate Experiment; Mission overview and early results. *Geophys. Res. Lett.*, **31**, L09607, doi:10.1029/2004GL019920.
- 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** (B12), 30 205–30 229.
- , S. Swenson, V. Zlotnicki, and I. Velicogna, 2004: Time-variable gravity from GRACE: First results. *Geophys. Res. Lett.*, **31**, L11501, doi:10.1029/2004GL019779.
- Yarosh, E. S., C. F. Ropelewski, and E. H. Berbery, 1999: Biases of the observed atmospheric water budgets over the central United States. *J. Geophys. Res.*, **104** (D16), 19 349–19 360.
- Yeh, P. J. F., M. Irizarry, and E. A. B. Eltahir, 1998: Hydroclimatology of Illinois: A comparison of monthly evaporation estimates based on atmospheric water balance and soil water balance. *J. Geophys. Res.*, **103**, 19 823–19 837.
- Zekster, I. S., and H. A. Loaiciga, 1993: Groundwater fluxes in the global hydrologic cycle: Past, present, future. *J. Hydrol.*, **144**, 405–427.