Variability of basin-scale terrestrial water storage from a P-E-R water budget method: the Amazon and Mississippi

Ning Zeng\textsuperscript{1,2}, Jin-Ho Yoon\textsuperscript{1}, Annarita Mariotti\textsuperscript{2,3} and Sean Swenson \textsuperscript{4}

\textsuperscript{1}Department of Atmospheric and Oceanic Science
University of Maryland, USA

\textsuperscript{2}Earth System Science Interdisciplinary Center
University of Maryland, USA

\textsuperscript{3}ENEA Climate section, Rome, Italy

\textsuperscript{4}University of Colorado, USA

\textsuperscript{1}Corresponding author address: Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD 20742-2425, USA; email: zeng@atmos.umd.edu; http://www.atmos.umd.edu/~zeng
Abstract

In an approach termed the P-E-R (or simply PER) method, we apply the basin water budget equation to diagnose the long-term variability of the total terrestrial water storage (TWS). The key input variables are observed precipitation (P) and runoff (R), and estimated evaporation (E). Unlike typical offline land-surface model estimate where only atmospheric variables are used as input, the direct use of observed runoff in the PER method imposes an important constraint on the diagnosed TWS. Although there lack basin-scale observations of evaporation, the tendency of E to have significantly less variability than the difference between precipitation and runoff (P-R) minimizes the uncertainties originating from estimated evaporation. Compared to the more traditional method using atmospheric moisture convergence (MC) minus R (MCR method), the use of observed precipitation in PER method is expected to lead to general improvement, especially in regions atmospheric radiosonde data are too sparse to constrain the atmospheric model analyzed MC such as in the remote tropics.

We show the results for the tropics using the example of the Amazon basin. The seasonal cycle of diagnosed TWS over the Amazon is about 200 mm, compares favorably with satellite gravity estimate from NASA’s GRACE mission, thus lending confidence both in this method and the usefulness of space gravity based large-scale soil moisture estimates. The interannual variability of TWS in the Amazon is about 150 mm, also consistent with GRACE data, but significantly larger than model results which typically represent near-surface soil moisture. We also apply this P-E-R method to the midlatitude Mississippi basin and discuss the impact of major 20th century droughts such as the Dust Bowl period on the long-term soil moisture variability. The multidecadal TWS variability for both basins has an amplitude up to 600 mm, much larger than model simulated soil moisture changes. While we currently lack
independent means to verify these long-term changes, such large variability implies the remarkable capacity of land-surface in storing and taking up water. The results also suggest the existence of water storage memories on decadal time scales, significantly longer than typically assumed seasonal timescales associated with surface soil moisture.
1 Introduction

Freshwater stored on the continents in the soil, at the surface or underground is fundamental for life on land. These water reservoirs are also important for climate as they provide potential feedback mechanisms for climate variability (e.g., Yeh et al. 1984; Delworth and Manabe 1993; Zeng et al. 1999; Koster et al. 2004). Motivated in part by the prospect of improving seasonal-interannual climate prediction using the knowledge of soil moisture state, there has been significant interest in soil moisture variability in recent years.

Several methods have been used in obtaining soil moisture information including in situ observations, satellite remote sensing, offline land-surface model simulations, land data assimilation, and basin-scale budget analysis. Table 1 lists some of these methods and examples as well as their main characteristics. While being ground truth, in situ observations are limited in spatial and temporal coverage, mostly in a few regions including the US, former soviet Union and China. Models resolve soil moisture at high spatiotemporal resolution, but are less constrained by observation. Satellite gravity measurements only have reliable information on scales larger than a few 100,000 km$^2$, i.e., the size of a medium-size river basin. The two basin budget methods only give basin-scale soil water storage, but there is no fundamental limit on the basin size.

Each of these methods has its own advantages and limitations. There is limited understanding of the consistency or agreement of these often independent methods. While there is a reasonable understanding of the climatological seasonal cycles of all aspects of the hydrological cycle, there is a significant lack of knowledge on the interannual variability of terrestrial hydrological variables including soil moisture.

While most other methods in Table 1 estimate near surface soil moisture, the two basin budget approaches and the satellite gravity sensor measure total terrestrial water storage (TWS). This water storage includes moisture near the surface, but also at the surface (snow and reservoir), deeper soil moisture, and ground water (Fig. 1). Deep soil and ground water variability may be particularly large (Rodell and Famiglietti, 2001; Seneviratne et al., 2004). There is some confusion
in literature as to the terminology, and here we make a distinction between the various components of the water storage vs. the total (TWS) when needed, but will also use the term ‘soil moisture’ for general discussion as widely used and understood.

Much emphasis has been on near-surface soil moisture, partly because seasonal crops in agriculture are typically shallow rooted. However, trees, shrubs and some natural grasses can have very deep roots that have been observed to take water from deep soil and ground water, for instance, below 8 m over the Amazon (Nepstad, 1994), 5-20 m in Edwards Plateau, Texas (Jackson et al., 1999), 7 m in Arizona (Davis and Pase, 1977). Recent research has suggested the potential importance on land hydrology and climate of such variability through plant deep root water uptake (Kleidon and Heimann, 2000; Jackson et al., 2000; Amenu et al., 2005). In addition to the apparent effects of irrigation and drinking water usage, the role of underground water in climate may also be more important than once thought. There is thus a strong need in knowing how water storage, especially in the lower part, varies over time and space.

In general, observations have been too short to demonstrate decadal soil moisture variability, but such information is becoming available for a few places such as Ukraine (Robock et al., 2005) and Illinois (Yeh et al., 1998; Rodell and Famiglietti, 2001). In addition to seasonal-interannual variability, the PER method presented here can provide useful information on decadal and longer-time scale water storage variability, limited mainly by the length of precipitation and runoff data. This method is to apply the simple basin water budget equation to diagnose the long-term variability of TWS. We present the method in section 2 and contrast it with the more traditional method using atmospheric moisture convergence. We then discuss the seasonal cycle and interannual variability from this method for the Amazon basin in sections 3 and 4, and compare the results with the moisture convergence method and satellite gravity based observations. In section 5, the long-term soil moisture variability is presented for the Mississippi basin. Potential limitations of the method are discussed in section 6, followed by conclusions in section 7.
2 Methodology

2.1 Basin budget methods for total water storage: MCR and PER

The traditional moisture convergence method (termed MCR here; Rasmusson 1968; Roads et al. 1994; Oki et al., 1995; Zeng, 1999; Seneviratne et al., 2004) considers the atmosphere and land-surface over a drainage basin as one single box, thus precipitation and evaporation vanish as interior fluxes for the total water budget. In this method (Fig. 2), moisture convergence ($C$; the vertically integrated water vapor flux) and observed streamflow for the drainage basin (runoff integrated over the whole basin) are integrated to obtain the change in atmosphere ($W$) and soil water storage ($S$):

$$\frac{d(W + S)}{dt} = C - R$$  \hspace{1cm} (1)

Using the recent atmospheric reanalyses, this method appears to produce reasonable estimates of the seasonal cycles and in some cases year-to-year variability over several basins around the world (Roads et al. 1994; Zeng 1999; Seneviratne et al. 2004). However, on decadal and longer time scales, the results are less robust especially over remote tropical regions such as over the Amazon. Our analysis (not shown) suggests that the lack of radiosonde data in these regions left the host atmospheric model poorly constrained and the simulation of convective rainfall is one of the weaker aspects of the models. Artificial jumps in the reanalysis systems when different observations are injected (Kalnay et al., 1996; Betts et al., 2005) may also be important. Such problems may be significantly alleviated if the observed precipitation is used in place of moisture convergence because precipitation is generally a better observed quantity over longer period of time, and this leads to the P-E-R (or simply PER) method.

In the PER method, only the land surface is considered (Fig. 3). The water budget equation for the land box is:

$$\frac{dS}{dt} = P - E - R$$  \hspace{1cm} (2)

where $S$ is the total terrestrial water storage and $P$ is precipitation, $E$ is evaporation (for simplicity, we use evaporation and evapotranspiration inter-exchangeably here).
In this method, precipitation and runoff are observed, and evaporation may be estimated using a land surface model driven by observed precipitation and other atmospheric variables (see caveats below in section 2.2). Thus the water budget Eq. 2 can be explicitly written for this method as:

$$\frac{dS}{dt} = P_{\text{obs}} - E_{\text{est}} - R_{\text{obs}}$$  \hspace{1cm} (3)

where the subscript ‘obs’ denotes ‘observation’, and ‘est’ denotes ‘estimate’. Compared to the moisture convergence method, the PER method uses observed P and R, thus more observational constraint. Offline land-surface model simulation uses only observed P as input, and in contrast, the PER method uses both observed P and observed R as an input.

Similar to the moisture convergence method (Rasmusson 1968; see discussions of this technique in Zeng 1999), a constant correction is added to E such that $P - E^* - R$ ($E^* = E + \text{correction}$) integrated over the analysis period is zero. As a result, the diagnosed TWS has the same value at the beginning and the end of the integration. This is equivalent to removing a linear trend in TWS. This correction is necessary as typical estimation of E tends to have systematic bias when compared to P-R, as indicated by the vertical shifts of E estimates relative to P-R in Fig.4. This bias can easily result in unrealistically large drift (trend) in the integrated S (e.g., Fig.6c of Zeng, 1999 and Fig. 9a of Seneviratne et al., 2004 for similar situations in MCR method). For instance, because the NCEP2 E is about 1 mm d$^{-1}$ larger than P-R (Fig.4), the imbalance without E correction would lead to about 365 mm drift at the end of the first year, so that the diagnosed S would go out of range of the plot (Fig.6) in 2-3 years. Although SLand has rather small bias in the Amazon, it is larger in the Mississippi. Thus this is a fundamental limitation of this method in that only the relative changes within the integration period can be inferred from such methods. But the relative variations within the period still provide valuable information not readily available otherwise.
2.2 Uncertainties due to evaporation estimate

The main potential weakness of the PER method is that basin-scale evaporation is generally not available and needs to be estimated. Should such estimates introduce large error, uncertainty would be large in the diagnosed water storage. Equation 3 indicates that a sufficient criterion is for the variation of $E_{est}$ to be significantly smaller than that of $P_{obs} - R_{obs}$. Because the magnitude of variation is often timescale dependent, this criterion may differ, for instance, on seasonal vs. interannual timescales.

Observational evidence suggests that evaporation indeed tends to have relatively small variation. Field measurements at the heart of the Amazon rainforest show a rather small seasonal amplitude in evaporation despite the large seasonal cycle in precipitation and soil moisture (Shuttleworth, 1988; Werth and Avissar, 2004). In the extra-tropics such as the Mississippi basin where radiation (thus potential evaporation) has large seasonal cycle, the seasonal amplitude of $E$ can be large, but the interannual variation is much smaller. To assess the uncertainty introduced by estimated evaporation, we have analyzed the evaporation from the model SLand (Zeng et al., 2000) and reanalysis products ERA40 (Gibson et al., 1997), NCEP/DOE (NCEP2; Kanamitsu et al., 2002), NARR (North American Regional Reanalysis; Mesinger et al., 2006). The reanalysis $E$ was also simulated by the embedded land-surface model which are typically more sophisticated than SLand. The results (Fig.4) show that the variance of interannual variability in $E$ is smaller than $P_{obs} - R_{obs}$ by a factor of 3 to 8 (as measured by standard deviation) over the Amazon, while for the Mississippi it is a factor of 2 to 3 smaller. This of course does not exclude possible larger uncertainties for particular events. Thus, in the worst-case scenario, even if the estimated $E_{est}$ is completely out of phase with the (unknown) truth, the diagnosed water storage would still reflect the dominant signal from $P_{obs} - R_{obs}$. In practice, the uncertainty would be smaller than the worst-case scenario because the model estimate can capture the variability in $E$ to some degree because it is mainly driven by observed precipitation and radiation. The resulting uncertainties in diagnosed TWS will be assessed below for the Amazon and Mississippi.
2.3 Forcing Data

The observed precipitation for 1901-2000 from the Climate Research Unit of University of East Anglia (CRU; New et al., 1999; Mitchell and Jones, 2005) was used in Eq. 3. The precipitation and temperature from the same dataset was also used in conjunction with the climatological values of surface wind and vapor pressure, along with radiation from NCEP/NCAR reanalysis to drive an offline model SLand (Simple-Land, Zeng et al. 2000) coupled to a dynamic vegetation model VEGAS (Zeng et al. 2005). The model was run at $1^\circ \times 1^\circ$ resolution at daily timestep and the results were then aggregated over the studied basin for the budget analysis at monthly timesteps. We also used the precipitation data from another gauge-based dataset PRECL (Chen et al., 2002), and the satellite-gauge blend of CMAP (Xie and Arkin, 1996). The results show broad agreement but with some differences especially on longer timescales, and this will be discussed in section 6. The Southern Oscillation Index (SOI) is used as an index for the atmospheric variability over the tropical Pacific Ocean for comparison purpose because the Amazon climate and hydrological variability are largely controlled by ENSO.

The monthly historical streamflow records for the Amazon River at Obidos, and for the Xingu River at Altamira were used to reconstruct the Amazon basin runoff, following the method used by Zeng (1999). In the following three sections, we present the seasonal cycle and variability from 1970 to 1997 for the Amazon basin, and long-term variability from 1928 to 1998 for the Mississippi basin using streamflow data from Vicksburg.

Evaporation is estimated using an offline simulation of SLand (above). For comparison, we will also show results using the ERA40, NCEP2 and NARR evaporation products as discussed in section 2.2 above. Equation 3 was integrated at monthly timestep to obtain $S$, with an arbitrary integration constant.
3 Seasonal cycle over the Amazon basin

A climatological seasonal cycle was derived as the average of the 28 year diagnosed total land water storage. Figure 5 shows the seasonal cycle for the Amazon basin. The 3 MC method analyses (Fig. 5a) have similar seasonal amplitude of 150mm, while the satellite gravity based estimate from GRACE and the PER method have an amplitude of about 300mm (Fig. 5b). The two offline models, SLand (Zeng et al. 2000) and the CPC Leaky Bucket model (Fan and van den Dool, 2004), also have an amplitude of about 150mm. Both GRACE and the PER method give a maximum in April-May and minimum in October-November after the drier boreal summer (the basin averages tend to be dominated by the larger southern Amazon). The models and reanalyses produced maximum and minimum somewhat earlier by 1-2 months.

To the extent that GRACE measurement can be considered as a good observation of the basin scale water storage, the PER method appears to capture this observed change. The reanalyses and the two offline models thus tend to underestimate somewhat the seasonal cycle amplitude in the Amazon. Given the uncertainties in all these methods and large interannual variability (below), the seasonal cycle of Amazon water storage can be given as 250±100 mm. However, the two basin budget methods (GRACE and P-E) include all the changes from surface to underground water, thus providing an upper limit to the models which normally include only a fraction of the active soil moisture as discussed further below. A caveat we emphasize is that such conclusion can differ for different basins as data quality and model may behave very differently at different places.

4 Amazon interannual and decadal variability 1970-1997

Figure 6a shows the water input (P-E*) and the observed runoff R of the Amazon basin at monthly resolution. There is a robust seasonal cycle over which P-E surpasses R during winter and spring, when soil moisture is recharged. P-E is less than R from early summer to fall when soil moisture is discharged (Fig.6b). Overall, R has a seasonal amplitude about factor of two smaller and a phase lag of 3-4 months relative to P-E* (also see Zeng 1999). This reduced amplitude and phase lag is
typical because soil moisture is a damped and delayed response to the driving precipitation due to its memory effect.

The diagnosed soil moisture shows large interannual to interdecadal variability on which the seasonal cycle is superimposed. The long-term variability can be seen more clearly by filtering out the seasonal cycle using a simple 12 month running mean (Fig. 7b, solid line). On multidecadal timescales, there is a major recharge period from 1971 to 1985, followed by a discharge afterwards (Fig. 7a,b). The overall amplitude over the 27 years is about 600 mm. We note again that, the correction in E* (section 2.2) removes any long term trend so that over the whole analysis period there is no net gain in soil water storage. Thus the lowest frequency change can only be viewed as relative, i.e., the decrease in the latter half is only relative to the increase in the first half of the 27 years.

Figure 7b also shows the results from using different evaporation estimates. Note that the shorter NCEP2 result was ‘tilted’ to match that of SLand because the overall trend is undetermined (section 2.1). Because of the cumulative effect on S in any small but consistent difference in P-E-R, the difference using different evaporation is larger at longer timescales. In a limited sense, results from these three evaporation give an error range for the PER method of about 30% on multidecadal timescale, and 15% on interannual timescale, and somewhat smaller error range due to precipitation (section 6).

The large change of 600 mm in TWS is remarkable, as many current land-surface models have field capacity (the maximum change in soil moisture a model can produce) comparable to this, and the earlier bucket model had a field capacity of only 150 mm (Manabe et al., 1965). Although we currently have no other means to validate the magnitude of such long-term change, the general variation (up and downs) can be assessed because Amazon hydrological cycle is dominated by ENSO related interannual variability. For instance, the period with largest recharge during 1974-75 corresponds to two major La Niña events before the 1976-77 decadal climate shift in the Pacific
Ocean, as indicated by the Southern Oscillation Index (SOI). In the other direction, the major discharge period of early 1990s was caused by the protracted El Niño of 1991-1993.

A simple high-pass filter was applied to the diagnosed soil moisture in Fig. 7b to remove the frequencies lower than 7 years. The remaining signal is mostly interannual (Fig. 7c), showing decreasing soil moisture during events such as 1982-83 El Niño. However, even in this case, the major peaks reflect the lower frequency variations such as the two La Niña events around 1975 and early 1990 El Niño. Plotted in Fig. 7d are the model simulated soil moisture from SLand and the CPC Leaky Bucket model, and both show significantly smaller variability at about 1/3 of the diagnosed amplitude on interannual time scales, while the decadal and longer-term variability is even smaller (cf. Fig. 7b).

The advent of recent satellite gravity based measurement from NASA’s GRACE mission provides an independent means to validate the interannual variability of the diagnostic method. Figure 8 shows the GRACE measurement of Amazon water change for the two year period of April 2002 to May 2004. The first half of this period has a larger seasonal variation of about 400 mm, while the second half has a smaller seasonal amplitude of about 250 mm (the seasonal amplitude in Fig.5 is an estimate over the whole 2 yr period), following the El Niño of 2002-2003. Thus the interannual change is about 150 mm, comparable to the amplitude of interannual variability derived using the PER method (Fig. 7c). Since the 2002-2003 El Niño is a relatively small one, the GRACE data suggests a potentially very large interannual variability in the land-surface water storage, consistent with our diagnostic approach (with the caveat that the GRACE data has its own uncertainties and the observation period is too short to statistically verify the water budget approach).

Such large differences especially on longer time scales are striking. While the diagnosis may overestimate the amplitude of these slow variations (section 6), the models appear to significantly underestimate it. One contributor of the smaller model changes is that current land-surface models typically only represent the water holding capacity of the top 1-2 meters of soil. Among the two
models, SLand has a field capacity of 500 mm, while the Leaky Bucket model has 750 mm. Thus it is understandable that these models underestimate the multi-decadal change on the order of 600 mm. However, simple increase in model soil depth may not be sufficient if deep soil water can not be utilized by vegetation. We only analyzed two simple land-surface models with a single soil layer, it remains to be seen how more sophisticated land models compare with the water budget and satellite results.

5 Long-term variability in the Mississippi basin

Data quality and model may behave very differently from basin to basin, especially across different climatic regimes. It is thus of great interest to see how the PER method works for midlatitude regions. We have applied the method to the Mississippi basin (Fig.9). The total water storage (Fig.9b) in the Mississippi decreases by about 400 mm from the 1920s to the end of the 1930s (the Dust Bowl period), followed by a recharge period in the 1940s. The drought in the 1950s plunged soil moisture to the lowest level, and then recovered to high level during the following two decades of pluvial period (Seager et al. 2005). Smaller drought events also left their impact on the soil moisture, such as 1988. Further analysis (not shown) suggests that these changes are largely influenced by the west part of the basin, in particular, the Great Plain region.

Similar to the Amazon case, because the overall trend is undetermined over the period each data is available, the S using ERA40 and NARR evaporation was ‘trend-matched’ to the S diagnosed using the longer E from SLand (Fig.9b), so that the comparison is only meaningful for the relative variations within the shorter period of any two curves. The different evaporation estimates give generally similar S, especially on interannual timescales (Fig.9c). The amplitude of interannual variability of about 100 mm is comparable to the total water storage variability based on in situ observations of the major components of the TWS for Illinois (Rodell and Famiglietti, 2001), if the assumption can be made that Illinois is representative of the whole Mississippi basin. Interestingly, the two model simulated soil moisture has interannual amplitude much closer to the diagnosed
one than in the Amazon case (Fig.9d). In both cases, the phase relationships are generally in good agreement (not surprisingly because they all reflect the signal in the precipitation forcing). However, the decadal and interdecadal variabilities in the two models are still much smaller than the diagnosed.

In general, in situ observations have been too short to demonstrate decadal and longer-term soil moisture variability, but such information is becoming available for places such as Ukraine (Robock et al., 2005) and Illinois (Yeh et al., 1998; Rodell and Famiglietti, 2001). We plotted in Fig.9b a point measurement from 1983 to 1998 at Freeport, Illinois (Hollinger and Isard, 1994). The data shows an increasing trend over the 15 year period, consistent with the diagnosed $S$, but with significantly smaller change (100 mm compared to about 250 mm). The smaller trend in the Freeport data is likely due to the fact that the measurement is only for the top 2 meter soil while the diagnosis is for the whole column. In addition, one cannot really compare a point measurement with the whole basin. Nonetheless, the data at least does not contradict the possibility of large decadal-scale change in water storage.

One implication of the diagnosed variability in Fig.9b is potentially very long soil moisture memory. For instance, the discharge during the Dust Bowl period is somewhat larger than that of the 1950’s drought, as seen by the larger shaded area with negative P-E*-R values for the 1930s in Fig.9a, yet soil moisture is at a minimum at the end of the 1950’s drought. This is because the recharge in the 1940s is not sufficient to recover the water loss in the 1930s. Thus in a way the memory of the Dust Bowl period was not forgotten until a few decades later. While uncertainties in the data and methodology may hamper the accuracy of such detailed interpretation, the results nonetheless suggest a significantly longer water storage memory than typically assumed.

6 Discussion

At first sight, it is surprising that the mere use of observed runoff would lead to much larger $S$ variability in the PER method, compared to a typical offline land model simulation. This is seen
clearly by comparing the solid line in Fig. 7b with that in Fig. 7d for the Amazon, and Fig. 9b with Fig. 9d for the Mississippi. Both methods use identical P and E, with the only difference that the model simulates its own runoff and the PER method uses observed runoff.

Insight comes from a comparison of how R follows P differently in the two cases. Theoretically speaking, because P is the main driving force of land-surface hydrology and the soil moisture ‘buffering’ effect acts as a low-pass filter of the precipitation signal, the subsequent outgoing fluxes including R (Fig. 3) are damped and delayed responses to P. Indeed, Fig. 10 and Fig. 11 show that observed R has an overall interannual amplitude of about 2/3 of P, with a typical lag of about a few months. This damping is larger on seasonal time scale (Fig. 6a). Such relationship can be shown with mathematical rigor easily in simple first-order force-damped equations such as Eq. (2) in the special case of a single intrinsic timescale. The most important point here is that the modeled R follows P much too closely compared to observed R, both in terms of amplitude and phasing. As a result, modeled R has a stronger tendency to cancel out changes in precipitation so that the difference P-R is much smaller than observed (Figs. 10, 11), thus leading to smaller variability in the diagnosed S (neglecting E variability; section 2.2). This is especially the case for the Amazon, and explains why the models severely underestimate TWS variability there (section 4 and Fig. 7d). For the Mississippi, the model simulated amplitude in S (Fig. 9d) is closer to the diagnosed (Fig. 9c) on interannual timescale (though still largely underestimated on decadal and longer timescales), consistent with the fact that the model P-R is much closer to observed (Figs. 11). It is not clear why the model does better over the Mississippi than the Amazon.

Such damped and significantly delayed response of runoff to precipitation suggests the large capacity of land-surface in storing and taking-up water, a capacity models appear to significantly under-represent. To put it another way, real land surface seems to be able to hold up more water from wet episodes, then uses it more slowly during dry periods, while models tend to have overly sensitive runoff response that flushes out ‘excessive’ water too quickly.
We argued in section 2 that estimated evaporation may be the main potential source of uncertainty because precipitation and runoff are observed (Eq. 3). We have used several different estimates to show that because of the generally small variability in E compared to P-R, the differences in diagnosed S due to different E estimates, though non-negligible, are small enough to retain the main decadal and especially interannual variabilities. A caveat is that all these evaporation estimates are model based, thus possibilities remain that actual E may have larger variability than depicted by these estimates.

To quantify impact on S due to the uncertainties in observed precipitation, we used three different precipitation datasets (but with the same E) in Eq. 3 to diagnose S. They show general agreement for both Amazon and Mississippi, especially between the two gauge-based datasets CRU and PRECL, while the satellite-gauge blended CMAP results are more different. Overall the uncertainty due to precipitation is somewhat smaller than that due to evaporation.

For decadal and longer timescales, a potentially significant source of uncertainty may come from the fact that the observational systems of precipitation and runoff may have changed over time, for instance, due to the replacement of instrumentation or changes in measurement protocol. We have seen that the diagnosed S is relatively sensitive to persistent imbalance in P-R which contributes to S cumulatively. Should R be ‘artificially shifted’ relative to P from one sub-period to the next, the error would also accumulate together with the real signal. The correction procedure in E* (section 2.1) removes the linear trend (in both signal and error) for the whole period, but the variations are still subject to ‘sub-period’ systematic errors in precipitation and runoff. Such errors are difficult to quantify and will have to be analyzed case by case for individual basins.

7 Conclusions

Applying the basin water budget equation in the PER method for the Amazon and the Mississippi basin, we found changes of 100-200 mm on interannual timescales, and 500-600 mm over multidecadal timescales in terrestrial total water storage for these two basins. Such large change
especially on longer timescales is remarkable, as many land-surface models have field capacity (the maximum change in soil moisture a model can produce) comparable or smaller than this.

Theoretically speaking, the diagnosed TWS variability is larger than or equal to any modeled soil moisture variability because the diagnosis includes all the possible changes in the basin, including surface and underground water and water stored in vegetation, while models typically only simulate soil moisture change in the top 1-2 meter. For example, during flooding season, Amazon river expands into adjacent forest and the surface water can account for about 10% of the seasonal soil moisture change. However, the major contributor missing in simple land-surface models is likely the deeper soil moisture storage and ground water. Such deep water storage is utilized by deep roots. In one instance, Nepstad et al. (1994) found deep root water uptake down to 8 m below surface that sustained normal growth during a prolonged dry period. Even deeper roots have been observed in many other regions (Schenk and Jackson 2005). More sophisticated models may include several soil layers which would increase the effective field capacity and thus the amplitude of variation. However, the lack of deep roots may prevent the model to utilize deep water storage as efficiently as Nature does. Our analysis with a simple model suggests that model runoff may respond too quickly to remove excessive precipitation such that soil moisture variability is small, regardless of the specified field capacity. It is worth noting that even if a model simulates a good mean seasonal cycle in runoff (as is often validated in model intercomparison project such as GSWP), it does not guarantee a good simulation of interannual variability, as it is the difference in P and R that matters most to soil moisture (section 6 above). Projects such as GSWP2 (http://www.iges.org/gswp2/) is expected to make available longer term simulations where interannual variability can be assessed. The large changes on decadal to multidecadal timescales found by the current diagnostic approach suggest that current models may significantly underestimate such variations. An important implication is that land-surface may have a memory beyond one
year related to the change in the total water storage\(^1\), significantly longer than the typically cited 1 month to 1 season.

The PER basin budget method uses observed precipitation and runoff, combined with estimated evapotranspiration to estimate the change in total land water storage. The land water budget equation (Eq. 2) is simple and has been used for various purposes (e.g., Mintz and Walker, 1993). But to our knowledge, it had not been applied for long-term water storage variability in a way similar to the PER method discussed here. Our preliminary analysis suggests its ability in depicting long term water storage change, in a way significantly more robust than the moisture convergence method. The results will thus provide an important means to cross-validate other methods such as GRACE data. If such comparison leads to confidence in both methods over the period when satellite is available (recent years and near future) as suggested by this work, they will provide important information on water storage variability in major periods of the 20th century when the basin-budget method is applicable. They can also provide validated approaches for long-term land water monitoring in the future.

**Acknowledgments.** We thank discussions with H. van den Dool, A. Robock, J. Wahr, P. Dirmeyer, and M. Rodell. Comments from an anonymous reviewer led to significant improvement of the manuscript. This research was supported by NSF grant ATM-0328286, NOAA grants NA04OAR4310091 and NA04OAR4310114.

\(^1\) In a simplistic way, time scale can be estimated as capacity divided by flux. Assuming a field capacity of 1000 mm for the Amazon, then the time scale is 1000 mm / (5 mm d\(^{-1}\)) = 200 days
References


Table 1: Methods commonly used to estimate soil moisture variability.

<table>
<thead>
<tr>
<th>Method</th>
<th>Examples</th>
<th>Advantages</th>
<th>Limitations</th>
<th>Degree of observational constraint</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ</td>
<td>Global soil moisture data bank(^a)</td>
<td>Direct measurement</td>
<td>Point data; limited coverage; near surface</td>
<td>High</td>
</tr>
<tr>
<td>Remote sensing Microwave</td>
<td>Microwave(^b)</td>
<td>High spatial and temporal resolution</td>
<td>Surface; area sparsely vegetated</td>
<td>High</td>
</tr>
<tr>
<td>Remote sensing Gravity</td>
<td>GRACE(^c)</td>
<td>Independent; total column</td>
<td>Low resolution; only since 2002</td>
<td>High</td>
</tr>
<tr>
<td>Offline land model forced by atmo vars</td>
<td>GSWP(^d); PDSI(^e)</td>
<td>High spatial and temporal resolution</td>
<td>Model dependent; top soil layers</td>
<td>Medium</td>
</tr>
<tr>
<td>Data assimilation Model+observation</td>
<td>GLDAS(^f)</td>
<td>High spatial and temporal resolution</td>
<td>Model dependent; top soil layers</td>
<td>Med-High</td>
</tr>
<tr>
<td>Basin-scale budget Moist.Conv.(MCR)</td>
<td>Rasmusson (1968)</td>
<td>Long-term; total column</td>
<td>Basin only; sensitive to quality of atmospheric data</td>
<td>Medium</td>
</tr>
<tr>
<td>Basin-scale budget P-E-R (PER)</td>
<td>This study</td>
<td>Long-term; total column</td>
<td>Basin only; some uncertainty in evaporation</td>
<td>Med-High</td>
</tr>
</tbody>
</table>

---

a. Robock et al. (2000).
c. Gravity Recovery and Climate Experiment: Tapley et al. (2004); Wahr et al. (2004).
e. Palmer Drought Severity Index.
Figure 1: Major components of total land water storage. Traditional methods including in situ soil moisture measurements and land-surface modeling have focused on the near-surface component. This study focuses on the total water storage using basin-budget PER method.
Figure 2: The moisture convergence method in diagnosing basin scale soil moisture variability. The atmosphere and land are treated as one single box such that the total water storage (atmosphere water $W$ plus soil water storage $S$) can be diagnosed as a time integral of moisture convergence ($C$; from atmospheric analysis) minus runoff ($R$; observed). Precipitation ($P$) and evaporation ($E$) are not needed because they are interior fluxes. Adopted from Zeng (1999).
Figure 3: The P-E-R (PER) method in diagnosing basin scale soil water storage variability. Only the water budget for the land-surface is needed. Soil water storage (S) is a time integral of precipitation (P) minus evaporation (E) minus runoff (R). Precipitation and runoff are from observations while evaporation needs to be estimated using model.
Figure 4: Evaporation (E) estimated from the SLand model and ERA40, NCEP2, NARR reanalyses compared to observed P-R. The uncertainty in diagnosed water storage (Eq. 3) due to the estimated E would be small if E variability is significantly smaller than that of P-R. Seasonal cycle has been removed using a 12 month running mean.
Figure 5: Seasonal cycle of Amazon soil moisture using: (a) the MCR method with moisture convergence from three reanalyses; (b) the PER method using observed P, R and model estimated E, and that derived using two years of GRACE satellite observations (Fig. 8); (c) simulated by two offline land surface models (SLand and CPC Leaky Bucket) forced by observed atmospheric variables, and from the Global Land Data Assimilation System (GLDAS).
Figure 6: (a) Variabilities of P-E* and R in mm d\(^{-1}\); (b) The diagnosed Amazon soil moisture in mm. When P-E* is greater than R, the soil water storage increases, undergoing a recharge period (heavy shading); and when P-E* is less than R, the soil undergoes discharge (light shading). There is a strong seasonal cycle, but also with comparable or larger interannual-interdecadal variabilities.
Figure 7: (a) Variabilities of P-E* and R in mm d⁻¹, and the southern oscillation index (SOI; labeled on the right in mb); (b) The diagnosed Amazon water storage S in mm, using different evaporation estimates. The shorter NCEP2 result was ‘tilted’ to match that of SLand because the overall trend is undetermined (section 2.1); (c) as in (b), except for the high frequency component (higher than 7 years, i.e., interannual but not decadal and longer); (d) S simulated by two models: SLand and the CPC leaky bucket model. A major recharge period occurred during two large La Niña events in 1974-75, and a major discharge period was associated with the protracted El Niño of 1991-93. Seasonal cycle was removed by a 12 month running mean.
Figure 8: Two years of observed changes in Amazon soil water storage (mm) from the GRACE satellite gravity measurements. A much smaller seasonal cycle was seen following the 2002-2003 El Niño compared to the year before.
Figure 9: As in Fig. 7, but for the Mississippi basin. The major discharge periods are the Dust Bowl in the 1930s, drought in the 1950s. A major recharge took place during the pluvial period from the 1960s to the 1990s.
Figure 10: (a) Anomalies of observed and modeled (SLand) runoff against precipitation (mm d$^{-1}$); (b) Differences between observed precipitation and observed (dashed) and modeled (line with open circles) runoff. The observed runoff is more different from observed precipitation especially over the Amazon, implying the large capacity of land to store water that is under-represented by models.
Figure 11: As in Fig. 10, but for the Mississippi basin.
Figure 12: Relative total water storage $S$ (mm) diagnosed using three different precipitation datasets CRU, PRECL, and CMAP. Evaporation was estimated using the SLand model forced by the corresponding precipitation.