Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE)

Sean Swenson and John Wahr

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

P. C. D. Milly

U. S. Geological Survey and Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, New Jersey, USA

Received 28 October 2002; revised 30 April 2003; accepted 16 May 2003; published 28 August 2003.

[1] The satellite Gravity Recovery and Climate Experiment (GRACE) provides data describing monthly changes in the geoid, which are closely related to changes in vertically integrated terrestrial water storage. Unlike conventional point or gridded hydrologic measurements, such as those from rain gauges, stream gauges, rain radars, and radiometric satellite images, GRACE data are sets of Stokes coefficients in a truncated spherical harmonic expansion of the geoid. *Swenson and Wahr* [2002] describe techniques for constructing spatial averaging kernels, with which the average change in vertically integrated water storage within a given region can be extracted from a set of Stokes coefficients. This study extends that work by applying averaging kernels to a realistic synthetic GRACE gravity signal derived in part from a large-scale hydrologic model. By comparing the water storage estimates inferred from the synthetic GRACE data with the water storage estimates predicted by the same hydrologic model, we are able to assess the accuracy of the GRACE estimates and to compare the performance of different averaging kernels. We focus specifically on recovering monthly water storage variations within North American river basins. We conclude that GRACE will be capable of estimating monthly changes in water storage to accuracies of better than 1 cm of water thickness for regions having areas of $4.0 \cdot 10^5$ km^2 or larger. Accuracies are better for larger regions. The water storage signal of the Mississippi river basin (area = $3.9 \cdot 10^6$ km^2), for example, can be obtained to better than 5 mm. For regional- to global-scale water balance analyses, this result indicates that GRACE will provide a useful, direct measure of seasonal water storage for river-basin water balance analyses; such data are without precedent in hydrologic analysis. *INDEX TERMS:* 1214 Geodesy and Gravity: Geopotential theory and determination; 1223 Geodesy and Gravity: Ocean/Earth/atmosphere interactions (3339); 1645 Global Change: Solid Earth; 1655 Global Change: Water cycles (1836); 1829 Hydrology: Groundwater hydrology; *KEYWORDS:* GRACE, time-variable gravity, groundwater, soil moisture, water storage, water budget

Citation: Swenson, S., J. Wahr, and P. C. D. Milly, Estimated accuracies of regional water storage variations inferred from the Gravity Recovery and Climate Experiment (GRACE), *Water Resour. Res.*, *39*(8), 1223, doi:10.1029/2002WR001808, 2003.

1. Introduction

[2] The NASA/DLR (Deutsches Zentrum fur Luft und Raumfahrt) satellite mission GRACE (Gravity Recovery and Climate Experiment), launched in March, 2002, is intended to provide highly accurate monthly solutions for the Earth's gravity field at scales of a few hundred km and larger. The mission, which has a 5-year lifetime, consists of two identical satellites in identical Earth orbits, one following the other at a distance of about 220 km. The satellites use microwaves to continually monitor their separation distance. As the satellites pass through gravity highs or lows, that distance changes. Thus, after removing the effects of nongravitational accelerations as detected by on-board accelerometers, the distance measurements can be used to solve for the gravity field.

[3] The month-to-month gravity variations obtained by differencing the GRACE gravity fields provide information about changes in the distribution of mass within the Earth and at its surface. There are a surprisingly large number of Earth-based processes likely to produce a time-variable gravity signal detectable by GRACE, including mass variations in the oceans, the atmosphere, the polar ice sheets, and the solid Earth. In general, the largest time-variable gravity signals observable in the GRACE data are expected to come from changes in the distribution of water and snow stored on land [*Wahr et al.*, 1998]. As a result, GRACE promises to

0043-1397/03/2002WR001808

provide a wealth of new and useful hydrologic information: namely, estimates of monthly changes in continental water storage, averaged over scales of several hundred km or larger anywhere in the world [*Dickey et al.*, 1997; *Wahr et al.*, 1998; *Rodell and Famiglietti*, 1999, 2001, 2002; *Swenson and Wahr*, 2002]. The primary purpose of this paper is to describe optimal methods for constructing those estimates, and to determine their probable accuracy.

[4] To determine whether GRACE is able to detect water storage variability within a particular region, two issues must be addressed. One is whether the gravitational effects are greater than the GRACE observational errors. Because of the decreased sensitivity of GRACE at short wavelengths, accurate estimates can be obtained only for regions that are several hundred km or more in scale. The other issue is whether the gravitational signal from the region of interest can be separated from all the other time-varying gravitational signals acting on GRACE.

[5] There are two types of separation problems. One is when the contaminating signal comes from a different region on the surface, such as from the water storage variability in a neighboring region, or from the ocean. In principle, GRACE could remove the effects of any such signals if it had perfect spatial resolution. One of the main concerns raised in this paper, and in an earlier paper [*Swenson and Wahr*, 2002] extended here, is to find a method of analysis that can optimally resolve the competing criteria of needing short scales to minimize the separation problem, but preferring longer scales to reduce the effects of GRACE measurement errors.

[6] The other separation problem arises when there are mass variations either in the atmosphere above the region, or in the solid Earth below it. Gravity has no vertical resolving power. If there is unmodeled mass variability in the atmosphere or solid Earth, its effects will be mapped directly into the water storage estimates. The GRACE Project is removing the effects of the atmospheric signal from the GRACE measurements prior to constructing each gravity field, using 6-hourly global surface pressure and geopotential height fields provided by the European Center for Medium Range Weather Forecasts (ECMWF). However, there are errors in the ECMWF fields, and those errors will affect the water storage estimates. The atmospheric errors are likely to be the limiting error source at scales of about 700 km and larger, but their effects on the regionally averaged water storage recovery are generally not apt to be larger than a few mm of equivalent water thickness [*Velicogna et al.*, 2001].

[7] In the solid Earth, by far the largest inadequately modeled contributions come from postglacial rebound (PGR), the ongoing, viscoelastic response of the solid Earth to the deglaciation at the end of the last ice age (for recent reviews, see *Wu* [1998] and *Mitrovica and Vermeersen* [2002]). This signal is linear in time, and is concentrated under Canada, Scandinavia, Antarctica, and Greenland. Thus it will contaminate GRACE estimates of the linearly varying water storage signals in those regions. Its effects will spill over somewhat into neighboring regions, and this issue is addressed in section 5.2. PGR should have no effect on the recovery of nonlinear temporal variability.

[8] The remainder of the gravity anomaly signal can then be attributed to changes in continental water storage, snow, and ice. Because water storage variability occurs in a thin layer at the surface of the Earth, GRACE is insensitive to the vertical distribution of mass [*Swenson and Wahr*, 2002]. GRACE thus senses changes in vertically integrated water storage, i.e., the total change in groundwater, soil moisture, surface water, snow and ice.

[9] The GRACE data set, which is provided to users every month, is not a set of point measurements. Instead, it is a finite set of spherical harmonic (Stokes) coefficients. These coefficients can be used to construct water storage averages over regions of any size and shape anywhere on the globe. However, the problem of constructing these averages is complicated by the competing requirements of reducing the effects of GRACE measurement errors (caused by such things as system-noise error in the inter-satellite range-rate, accelerometer error, error in the ultrastable oscillator, and error in the orbits) while minimizing the contaminating signal from adjacent regions.

[10] *Swenson and Wahr* [2002] developed a method of estimating water storage variability in an arbitrary region that minimizes the combined effects of satellite measurement errors and unwanted signals from nearby regions (referred to as ''leakage''). This method uses measurement error covariance matrices provided by the GRACE Project to construct averaging kernels for the region of interest. The errors in these water storage results due to leakage from surrounding regions can be reliably estimated only if there is a priori knowledge of certain characteristics of mass variability within those regions, such as the amplitude of monthly variability, the temporal correlation between the region of interest and neighboring regions, and the rate at which temporal correlations change with distance.

[11] One way of estimating the uncertainty in GRACE water storage estimates is to simulate GRACE data by combining large-scale (preferably global), gridded models of the expected monthly water storage signal with models of errors in the meteorological fields used to remove atmospheric effects and GRACE measurement errors. Regional water storage estimates can be constructed from those simulated data, and those estimates can be compared with the true regional averages obtained from the hydrology model alone. The differences are a measure of the error in the GRACE estimates. We construct such simulations in this paper.

[12] Our objectives are as follows: (1) to demonstrate how the methods developed by *Swenson and Wahr* [2002] can be applied and optimized when dealing with the signals expected in actual GRACE data and (2) to illustrate how simulations can be used to estimate the uncertainties in the water storage estimates. Our simulations assume a specific global hydrology model and adopt the prelaunch estimates of the GRACE measurement errors. Although users will certainly use other hydrology models, and the error covariance matrices for real GRACE data are likely to differ, perhaps substantially, from the prelaunch estimates, the simulation methods described here will still be appropriate. In addition, preliminary predictions are presented of GRACE water storage accuracies for specific North American river basins.

2. Synthetic GRACE Data

[13] The process of creating synthetic GRACE gravity data is described in detail by *Wahr et al.* [1998]. Here we

summarize this procedure. Five years of synthetic, monthly GRACE gravity solutions, represented in the form of monthly Stokes coefficients (see section 3, below), were constructed from models of surface-mass variability. Modelled quantities include terrestrial water storage (soil moisture, groundwater, surface water, and snowpack), oceanographic processes, errors resulting from the removal of the contribution from atmospheric mass redistribution, and expected GRACE measurement errors.

[14] Continental storage of root-zone soil water, groundwater, and snowpack were simulated on a 1-degree longitude/latitude grid by the water and energy balance model of *Milly and Shmakin* [2002]. The model was driven by observed precipitation for the period 1979–1998 as given by the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP); further details of the methodology and an evaluation of the model accuracy are given by *Shmakin et al.* [2002]. The five-year period from 1994–1998 was used in this analysis. Output from the ocean model, a derivative of the Los Alamos Parallel Ocean Program ocean general circulation model [*Dukowicz and Smith*, 1994] run at the National Center for Atmospheric Research (NCAR), was provided by M. Molenaar and F. Bryan (personal communication, 1999). Errors in atmospheric mass corrections were estimated as the de-meaned difference between the ECMWF and NCEP (National Centers for Environmental Prediction) global gridded surface pressure fields, divided by $\sqrt{2}$. The variance of the satellite measurement errors as a function of wavelength, consistent with the GRACE Science and Mission Requirements Document [*Jet Propulsion Laboratory*, 2001], was provided by B. Thomas and M. Watkins at JPL (personal communication, 2001). A random number generator was used to create a representation of satellite measurement errors for individual Stokes coefficients corresponding to the given variances.

[15] One source of surface-mass variability, postglacial rebound, was excluded from our simulated data. The results shown below should therefore not be used to assess GRACE's ability to recover linearly varying water storage signals at high northern latitudes, particularly in Canada and Scandinavia. This issue is discussed in more detail in section 5.2.

3. Estimating Changes in Regional Averages of Continental Water Storage

[16] The gravitational equipotential surface most closely coinciding with mean sea level over the ocean is known as the geoid. Changes in the distribution of mass in the Earth, e.g., continental water storage variation, are responsible for spatiotemporal changes in observations of the geoid, N. Typically, the geoid is expressed as a sum of normalized associated Legendre functions, $\tilde{P}_{lm}$, [see, e.g., *Chao and Gross*, 1987]:

$$N(\theta,\phi) = a\sum_{l=0}^{l_{max}}\sum_{m=0}^{l}\tilde{P}_{lm}(\cos\theta)\{C_{lm}\cos m\phi + S_{lm}\sin m\phi\}, \quad (1)$$

where θ is co-latitude, ϕ is longitude, a is the mean radius of the Earth, and the C_{lm} and S_{lm} are dimensionless Stokes coefficients. The spatial scale (half-wavelength) of a Legendre function is roughly (20,000/l) km. GRACE Project personnel will use the GRACE measurements to solve for the Stokes coefficients up to degree $l \sim 100$ every 30 days, and these coefficients will be made available to users.

[17] *Swenson and Wahr* [2002] showed that changes in the GRACE Stokes coefficients, ΔC_{lm} and ΔS_{lm}, can be used to estimate $\Delta\tilde{\sigma}_{region}$, the water storage anomaly within a region, as:

$$\Delta\tilde{\sigma}_{region} = \sum_{l=0}^{l_{max}}\sum_{m=0}^{l}\frac{1}{\Omega_{region}}\frac{a\,\rho_E}{3}\frac{(2l+1)}{(1+k_l)}\left(W_{lm}^{c}\Delta C_{lm} + W_{lm}^{s}\Delta S_{lm}\right). \quad (2)$$

where ρ_E is the average density of the Earth, Ω_{region} is the angular area of the region (the area divided by a^2), l_{max} = 100 is the largest angular degree in the GRACE gravity solution, and k_l are load Love numbers representing the effects of the Earth's response to surface loads. Values of the Love numbers are given in *Wahr et al.* [1998]. W_{lm}^{c} and W_{lm}^{s}, the averaging kernel coefficients, are specifically constructed for the region to be isolated, and they determine the relative amounts of satellite and leakage errors. *Swenson and Wahr* [2002] outlined four methods of creating averaging kernels: choosing an exact (or block) average, Gaussian convolution, the use of Lagrange multipliers, and minimization of the sum of satellite and leakage errors.

3.1. Examples

3.1.1. Exact Averaging Kernel

[18] For the first of two examples, we use the exact averaging kernels of the Mississippi and Ohio river basins to recover monthly basin averages from the synthetic GRACE Stokes coefficients described in section 2. The weighting coefficients of an exact averaging kernel consist of the spherical harmonic coefficients of the basin function, which is defined such that it takes values of 1 inside the basin and 0 outside. The term "exact" is applied to this averaging kernel because it would provide the exact value of the basin average in the absence of satellite measurement errors. It will be shown in section 3.1.3 that measurement errors lead to inaccurate water storage estimates when the exact averaging kernel is employed (see *Swenson and Wahr* [2002] for a detailed explanation of this result).

3.1.2. Fixed Satellite Measurement Error Averaging Kernel

[19] *Swenson and Wahr* [2002] described alternative averaging kernels constructed using the method of Lagrange multipliers. This technique creates an averaging kernel that minimizes the contribution of either the measurement error or the leakage error, while constraining the other error (leakage or measurement, respectively) to some desired value. In this case, *Swenson and Wahr* used a measure of leakage error based solely on the difference in shape between the exact and smoothed averaging kernel, and therefore required no a priori knowledge of the signal. As a second example, an a priori model for the satellite measurement error is used to construct an averaging kernel such that the variance of the expected measurement error is constrained to a value of 0.5 cm.

3.1.3. Comparison

[20] The monthly basin average recovered with the exact averaging kernel for each basin (dashed line) is compared to

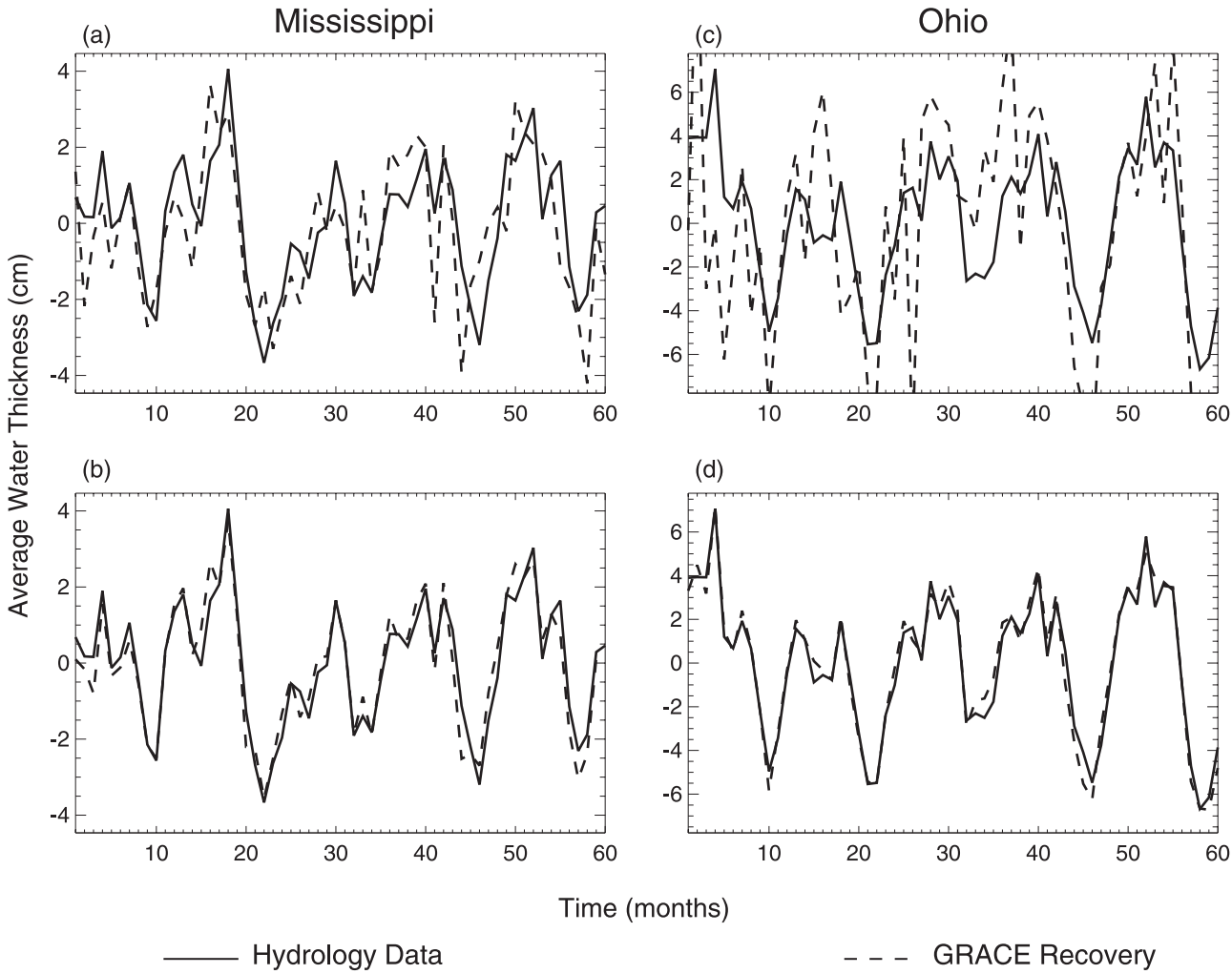


Figure 1. Recovered monthly anomalies of basin-averaged continental water storage for the Mississippi and Ohio river basins. (a) The results for the exact averaging kernel of the Mississippi basin. (b) The results for an averaging kernel designed to include no more than 0.5 cm of satellite measurement error. (c and d) The same as Figures 2a and 2b except that they are for the Ohio basin. The y axis is in units of centimeters of equivalent water thickness, and the x axis is in units of months. Solid line represents the true change, and dashed line represents the value estimated from synthetic GRACE data.

the true basin average (solid line) for each basin in Figures 1a and 1c. Table 1 (under the heading "Exact") gives RMS values of the recovered water storage signal, total recovery error, i.e., the differences between the two time series shown in Figure 1, as well as the RMS contributions from each of the individual error sources that comprise the total error: satellite measurement error, leakage from water storage signals outside the basin, error resulting from the removal of the atmospheric signal, and leakage from mass variation in the oceans. Note that in any column of Table 1, the squares of the individual errors do not necessarily sum to the square of the total error, because individual errors may be correlated.

[21] Satellite measurement error is the largest error component when an exact averaging kernel is used (Table 1). The error is significantly larger for the Ohio Basin than for the Mississippi Basin, because the Ohio Basin is smaller, and the errors in the Stokes coefficients are larger at smaller scales (i.e., at larger angular degrees). The satellite measurement error would continue to increase if the basin size were decreased further. The ocean and water storage leakage errors are not exactly zero, because the sum over l in (2) is truncated at $l = l_{max}$, rather than going to infinity. This truncation causes the equivalent averaging kernel in the spatial domain to differ slightly from the true basin function.

Table 1. Recovery RMS Water Storage Signal and Errors[a]

	Mississippi		Ohio	
Basin Kernel Type	Exact	Fixed Satellite Error	Exact	Fixed Satellite Error
Water Storage Signal	1.82	1.69	5.60	3.34
Total Error	1.27	0.50	4.23	0.59
Satellite Measurement Error	1.24	0.42	4.19	0.56
Water Storage Leakage	0.02	0.02	0.06	0.11
Atmospheric Removal Error	0.26	0.26	0.20	0.21
Ocean Leakage	0.02	0.02	0.01	0.01

[a]In centimeters of equivalent water thickness.

[22] The results of using an averaging kernel constrained to give 0.5 cm of measurement error for the Mississippi and Ohio river Basins are shown in Figures 1b and 1d. The RMS contributions from the individual error sources are shown in Table 1, in the columns labeled ''Fixed Satellite Error''. In each case, the satellite measurement errors account for about 0.5 cm of the total error, as they should, because that was the contribution specified when constructing the averaging kernels. (The satellite measurement error contributions are not exactly 0.5 cm, because the synthetic data were constructed using a statistical realization of the expected satellite errors, while the 0.5 cm requirement refers to the ensemble average of all possible statistical realizations). The leakage errors remain small and are not much different than they were for the exact averaging kernel. The RMS of the total error has thus decreased dramatically for each basin: from 1.27 cm to 0.50 cm for the Mississippi, and from 4.23 to 0.59 for the Ohio.

[23] These total errors, however, and the total errors obtained with the other variation of the Lagrange multiplier approach (not shown here) where the leakage is specified and the effects of satellite measurement errors are minimized, are larger than those we will obtain in the next section by minimizing the total error. Our rationale for showing the Lagrange multiplier method is to offer an alternative to the total error minimization method, described below, that is independent of assumptions about the specific nature of the temporal and spatial characteristics of the signal. In regimes where the spatial and/or temporal characteristics of the expected signal are poorly known, or where the assumption of an azimuthally symmetric covariance function is invalid, these methods may be preferable to the technique of minimizing an a priori estimate of the total error. However, it will be shown that for the purpose of recovering continental water storage variability, the method described in section 3.2 provides more accurate results.

3.2. Minimizing the Total Error: Azimuthally Symmetric Covariance Function Assumption

[24] When a priori signal information is available, it becomes possible to construct an averaging kernel that minimizes the sum of the variance of the satellite and leakage errors. To minimize these contributions, one requires a priori estimates of the satellite measurement error and of the signal to be measured. Satellite error covariance matrices will be made available concurrently with GRACE Stokes coefficients. Estimates of the signal may be obtained from models, observations, or some combination of the two. Constructing an averaging kernel with this method is simplified if one assumes the signal covariance function (CF) is azimuthally symmetric. Here we assess the accuracy of this approximation, and outline an iterative approach for creating an optimal averaging kernel.

[25] We will use two models for our azimuthally symmetric CF: a Gaussian and a decaying exponential. In both cases, the CF has the form:

$$cov_\sigma(\theta, \phi, \theta', \phi') = \sigma_0^2 \, G(\gamma, d) \tag{3}$$

where γ is the angular distance between two points (θ, ϕ) and (θ', ϕ'), i.e., $\cos\gamma = \cos\theta\cos\theta' + \sin\theta\sin\theta'\cos(\phi - \phi')$, σ_0^2 is the local signal variance, and d is the function half-width, which we use as a measure of the correlation length. We assume that both σ_0^2 and d are spatially uniform.

[26] The angular dependence of a Gaussian CF can be expressed as

$$G(\gamma, d) = \exp[-b(1 - \cos\gamma)], \tag{4}$$

where

$$b = \frac{\ln(2)}{(1 - \cos(d/a))}, \tag{5}$$

and an exponential CF can be expressed as

$$G(\gamma, d) = \exp[-b\gamma], \tag{6}$$

where

$$b = \frac{\ln(2)}{(d/a)}. \tag{7}$$

[27] With the azimuthally symmetric CF approximation, the optimal averaging kernel coefficients become [see *Swenson and Wahr*, 2002]:

$$\begin{Bmatrix} W_{lm}^c \\ W_{lm}^s \end{Bmatrix} = \left[1 + \frac{2K(2l+1)B_l^2}{\sigma_0^2 G_l (1 + k_l)^2}\right]^{-1} \begin{Bmatrix} \vartheta_{lm}^c \\ \vartheta_{lm}^s \end{Bmatrix}. \tag{8}$$

Here $K = a\rho_E/3$, B_l are degree amplitudes of the satellite measurement errors, and G_l are the Legendre coefficients of $G(\gamma, d)$: $G(\gamma, d) = \Sigma_l G_l P_l(\gamma)$, where P_l are Legendre polynomials. For a Gaussian CF, a recursion relation for G_l can be obtained from equation (34) of *Swenson and Wahr* [2002]. Note, however, that the Gaussian used by *Swenson and Wahr* [2002] is normalized so that its global integral is 1. The coefficients describing $G(\gamma, d)$ may be produced by dividing those in Equation (34) by the normalization factor $\frac{b}{2\pi}\frac{1}{1-e^{-2b}}$ from Equation (30). The Legendre coefficients of an exponential must be computed numerically.

[28] The competing influences of satellite measurement error and leakage error can be understood by examining (8). When the satellite measurement errors at some specified value of l are large compared to the water storage signal in and around the region of interest, the inverse of the bracketed quantity in (8) becomes small, thereby reducing the effects of satellite error for that value of l. This reduction typically increases as l increases, i.e., at short wavelengths, because B_l increases at large l. On the other hand, when the signal is large compared to measurement errors, the bracketed quantity approaches 1. The result is an averaging kernel that looks more like the exact averaging kernel, and therefore reduces leakage error.

[29] W_{lm}^c and W_{lm}^s from (8) can be used in (2) to estimate $\Delta\widetilde{\sigma}_{region}$ from GRACE data. A problem, though, is that numerical values must first be chosen for σ_0^2 and d (G_l depends on d). In principle, the values of these parameters can be estimated from GRACE data, using an iterative approach. We will adopt such an approach for estimating σ_0^2. First, a plausible value of d, which will be kept constant throughout the analysis, and an initial value for σ_0^2 are chosen. Using (8) in (2), $\Delta\widetilde{\sigma}_{region}$ is estimated from GRACE

data for each month's set of Stokes coefficients. It is straightforward to show that, with the assumption of an azimuthally symmetric CF, the variance of the basin-averaged surface-mass anomaly can be related to σ_0^2 and G_l by

$$var(\tilde{\sigma}) = \frac{\sigma_0^2}{\Omega_{region}^2} \sum_{l=0}^{\infty} \sum_{m=0}^{l} \frac{G_l}{2} \left[\vartheta_{lm}^{c2} + \vartheta_{lm}^{s2} \right]. \quad (9)$$

The variance of the basin averages is computed as $var(\tilde{\sigma}) = \frac{1}{N} \sum_{t_i=1}^{N} \left[\Delta\tilde{\sigma}(t_i)_{region} \right]^2$, where t_i is the time index, and N is the number of estimates of $\Delta\tilde{\sigma}_{region}$ in the time series. After putting $var(\tilde{\sigma})$ in (9), the resulting equation can be used to solve for an improved estimate of σ_0^2, which is then used in (8) to obtain a better estimate of W_{lm}^c and W_{lm}^s. These steps are repeated until we obtain convergence. A similar iterative approach could be constructed to find d, but the inversion is nonlinear and so is more difficult. As shown below, we have obtained good results by iterating only on σ_0^2.

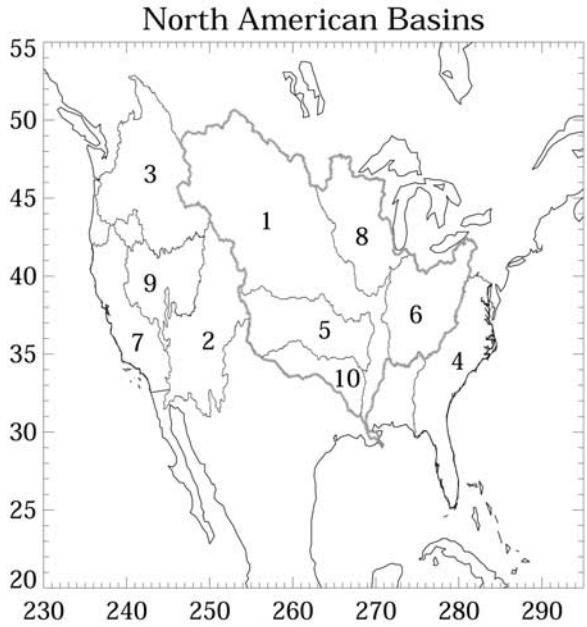
4. Results

[30] To determine what a "plausible" correlation length might be, we first examine the dependence of the total estimation error on d, the correlation length used to create the averaging kernel. We choose a region and compute $var(\tilde{\sigma})$ for that region directly from the hydrologic model; this value of $var(\tilde{\sigma})$ will be taken as the "correct" value. For each value of d, we use (9), together with our value of $var(\tilde{\sigma})$, to solve for ${\sigma_0}^2$. Thus we are effectively parameterizing each averaging kernel by d, which is allowed to vary, and $var(\tilde{\sigma})$, which is held constant at its correct value. We apply each averaging kernel to the synthetic GRACE data set, and compare the resulting estimates of $\Delta\tilde{\sigma}$ with the correct values. The 10 basins for which averaging kernels were constructed in this manner, and the area of each basin in km^2 are shown in Figure 2. Note that basin 1, the Mississippi river basin (outlined with a thick shaded line), includes the Upper Mississippi, Ohio, Arkansas, and Red river basins.

4.1. Gaussian Covariance Function

[31] The total error in the estimate of the basin-averaged surface-mass anomaly, and the relative contributions of satellite measurement error, leakage error, and pressure errors, when a Gaussian CF (4) is used to determine G_l in (8), are shown in Figure 3. The general characteristics of each plot are as follows. As the correlation length increases, total error (solid line) initially decreases to a minimum value, after which it increases. The minimum total error generally occurs at values of d between 200 and 600 km. Satellite measurement error (dashed line) decreases as d increases. Leakage error (dotted-dashed line) increases as d increases. Pressure error (dotted line) is relatively constant as a function of d. For this reason, pressure errors are not included in the minimization procedure.

[32] Note that the pressure error is the limiting source of error for all but the smallest river basins considered here. In some cases, certain error components are anti-correlated; the variance of the total error is therefore less than the sum of the variances of its constituents. The variance of the total error in the Colorado basin recovery, for example, is less than the variance of the pressure error for values of d between 500 and 1000 km.



1. Mississippi	388 x 10^4 km^2	6. Ohio	73 x 10^4 km^2
2. Colorado	101 x 10^4 km^2	7. California	71 x 10^4 km^2
3. Columbia	92 x 10^4 km^2	8. Upper Mississippi	67 x 10^4 km^2
4. East Coast	87 x 10^4 km^2	9. Great Basin	59 x 10^4 km^2
5. Arkansas	76 x 10^4 km^2	10. Red	40 x 10^4 km^2

Figure 2. (top) Boundaries of North American basins for which averaging kernels were created. Note that the Mississippi River basin, (basin 1, thick shaded outline), includes basins 5, 6, 8, and 10. (bottom) Areas of basins, ranked largest to smallest by area.

[33] Because this technique for constructing the averaging kernel is based on minimizing the variance of the leakage and measurement errors, an averaging kernel derived from the correlation length that most closely represents the true covariance function should produce leakage and measurement errors that are nearly equal. Averaging kernels based on correlation lengths that poorly approximate the real covariance function will produce unequal amounts of leakage and measurement error. However, in cases such as the Colorado basin, anti-correlation between pressure errors and the water storage signal will cause the minimum error to occur at a correlation length that does not best represent the true covariance function.

4.2. Decaying Exponential Covariance Function

[34] The relative error contributions when averaging kernels constructed using a decaying exponential CF, (6), are used to recover the basin-averaged surface-mass anomaly (Figure 4). The general behavior of the error components as a function of d is the same as that in the Gaussian case. However, the error components change less rapidly with d than do those in the case of a Gaussian CF, resulting

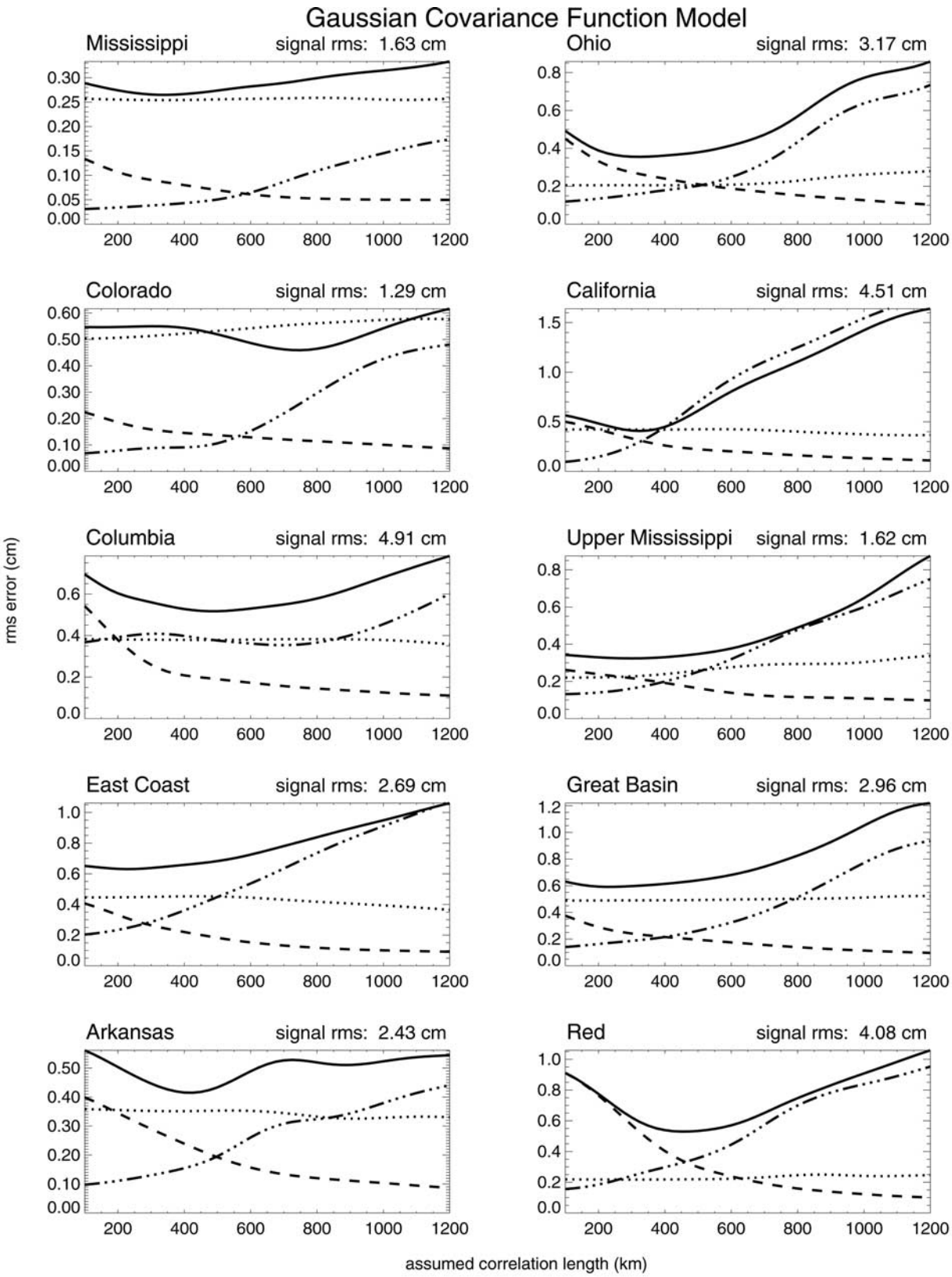


Figure 3. Recovery errors as a function of correlation length for North American basins. Solid line is total error; dotted line is error from removing the atmospheric signal; dashed line is measurement error; dotted-dashed line is leakage error. Note that the y axis is not the same for all plots.

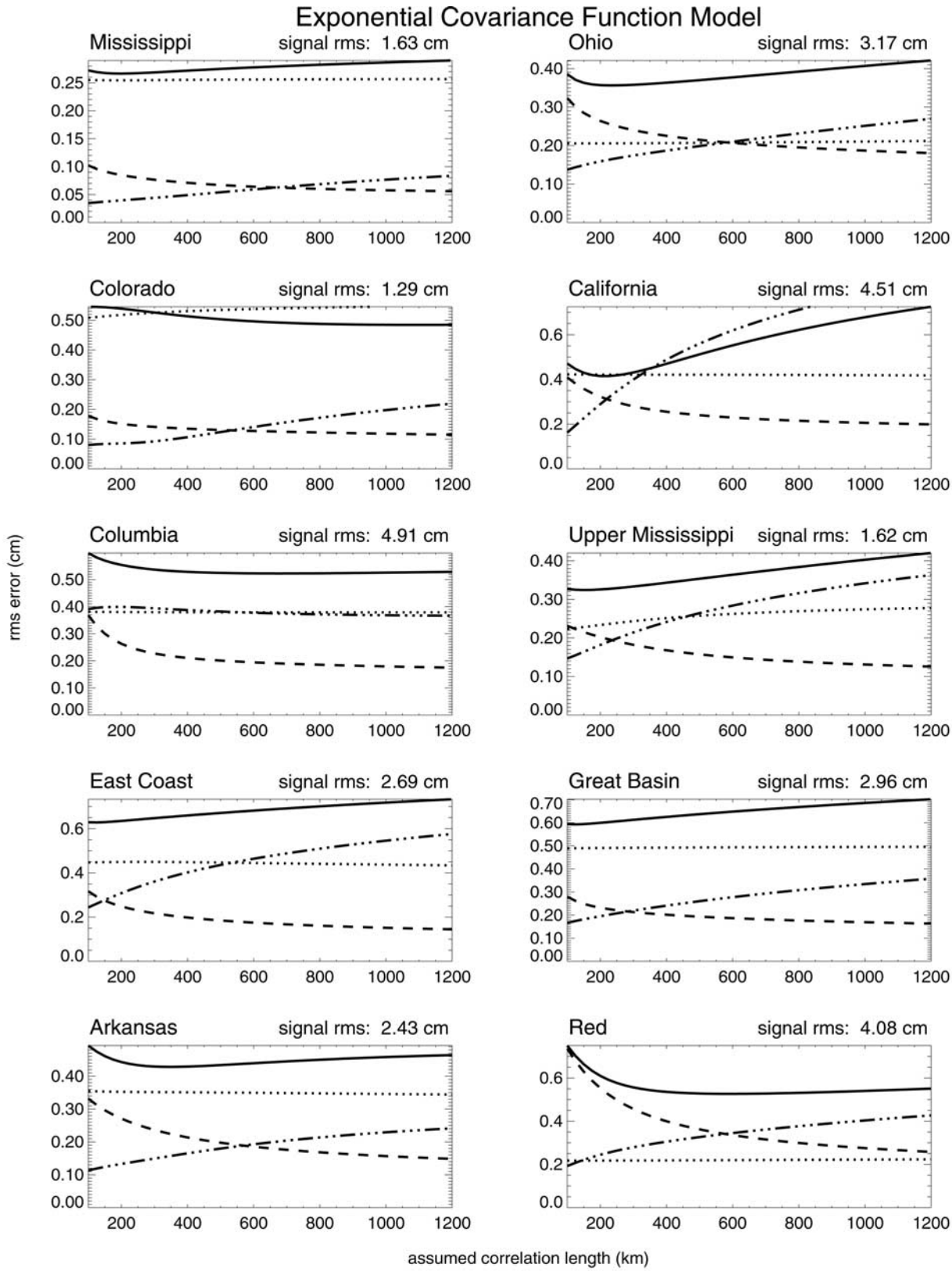


Figure 4. This plot is the same as Figure 3 except that the covariance function model used to construct the averaging kernels is a decaying exponential. Note that the y axis is not the same for all plots.

Table 2. Minimum RMS Basin Average Water Storage and Recovery Errors[a]

Basin	d_{min}	Water Storage	Errors				
			Total	Atmosphere	WS Leakage	Satellite	Ocean
Mississippi	200	1.63	0.27	0.25	0.08	0.04	0.01
Colorado	1100	1.29	0.48	0.55	0.12	0.21	0.01
Columbia	675	4.91	0.52	0.38	0.19	0.38	0.00
East Coast	125	2.69	0.63	0.45	0.29	0.26	0.24
Arkansas	350	2.43	0.43	0.35	0.22	0.16	0.01
Ohio	225	3.17	0.36	0.21	0.26	0.16	0.02
California	225	4.51	0.42	0.42	0.31	0.32	0.08
Upper Mississippi	150	1.62	0.32	0.23	0.22	0.16	0.00
Great Basin	125	2.96	0.59	0.49	0.26	0.17	0.01
Red	600	4.08	0.53	0.22	0.34	0.34	0.04

[a]Exponential covariance function. In centimeters of equivalent water thickness, except d_{min}, which is in kilometers.

in total error curves that are generally much shallower than those for the Gaussian case.

4.3. Model Comparison/Analysis

[35] The minimum errors, as well as d_{min}, the correlation length at which the total error is a minimum, obtainable under the assumption of an exponential CF are shown in Table 2. The minimum errors in the Gaussian case are all within 1% of those in the exponential case. In both cases, d_{min} is typically between 200 and 800 km. This range of length scales is consistent with studies of the spatial coherence of soil moisture [*Entin et al.*, 2000; *Vinnikov et al.*, 1996]. However, the error curves in the Gaussian case change more rapidly with d than those in the exponential case. Thus estimates of $\Delta\widetilde{\sigma}$ produced using averaging kernels constructed with $d \neq d_{min}$ will typically be less accurate in the Gaussian case than those produced under the assumption of an exponential CF.

[36] The difference between the rates of change of total error for the two CFs can be understood by examining the averaging kernel coefficients. The dependence of W_{lm}^c and W_{lm}^s on d comes from the spectrum of the CF, G_l. The Legendre expansion, G_l, for each model of the covariance function is shown in Figure 5. The CF correlation length ranges from 200 km (darkest line) to 1000 km (lightest line). For l greater than about 20, G_l for the exponential model (Figure 5, top) varies as a function of d relatively less than G_l does for the Gaussian (Figure 5, bottom). For example, the ratio of G_{21} for d = 200 km to G_{21} for d = 1000 km in the exponential case is about 1.8, although the ratio in the Gaussian case is about 1.9. The same ratios for G_{40} are 3.4 and $3.7 \cdot 10^4$, respectively.

[37] W_{lm}^c and W_{lm}^s computed using an exponential CF therefore vary more slowly with d than do those computed from a Gaussian model of the CF, meaning that the basin-averaged surface-mass anomalies computed with averaging kernels derived from the exponential CF will also vary relatively slowly with d compared to those derived from the Gaussian CF.

[38] From a practical standpoint, this means that an inaccurately determined value of d will lead to greater errors when using a Gaussian CF averaging kernel. However, it should be noted that the better accuracies obtainable with an exponential CF averaging kernel for $d \neq d_{min}$ are the result of the slowly varying nature of the Legendre spectrum as a function of correlation length, and not an indication that a decaying exponential is a better model for the true water storage signal than a Gaussian. If that were the case, the estimates obtained when $d = d_{min}$ would be more accurate for the exponential case than the Gaussian case.

4.4. Estimating $var(\widetilde{\sigma})$

[39] In the previous section, we examined the accuracies of simulated GRACE recoveries for two models of the covariance function over a range of correlation lengths. $var(\widetilde{\sigma})$ was set to its "true" value, and (9) was used to solve for a value of σ_0^2 consistent with both $var(\widetilde{\sigma})$ and d. In practice, one will not know $var(\widetilde{\sigma})$. However, once d is chosen, we will show that $var(\widetilde{\sigma})$ can be obtained through the iterative procedure described at the end of section 3.2. This is especially promising for the exponential CF averaging kernel because of its relative insensitivity to correlation length; choosing a value of $d \neq d_{min}$ will not significantly degrade the recovery.

[40] For each basin, we created three exponential CF averaging kernels based on correlation lengths, d, of 200, 400, and 800 km, respectively. Legendre coefficients, G_l, were computed for each value of d. $var(\widetilde{\sigma})$ was given an initial value of 400 cm^2 (20 cm rms). Equation (9) was then used to calculate a value of σ_0^2 that was consistent with the specified values of d and $var(\widetilde{\sigma})$. W_{lm}^c and W_{lm}^s were computed from these values of G_l and σ_0^2 using (8). The averaging kernel described by W_{lm}^c and W_{lm}^s was then applied to the synthetic GRACE data set to estimate a new value of $var(\widetilde{\sigma})$. After each iteration, σ_0^2 was reset to a value determined by the new value of $var(\widetilde{\sigma})$, and a new averaging kernel was constructed. The solution converged after five iterations; successive values of $var(\widetilde{\sigma})$ differed by less than 0.01 cm^2. The true value of the RMS of the basin-averaged change in water storage, denoted as "Water Storage Amplitude", and the final RMS values of the total error obtained after five iterations, for each value of d, are shown in Figure 6. The total error in all cases is less than 0.75 cm. The minimum total error possible using an exponential CF averaging kernel is represented by an arrow to the right of the column. The RMS total error and the RMS of the true signal are shown in Table 3. Fractional errors range from roughly 40% for the Colorado basin, to roughly 10% for basins such as the Columbia, California, and Ohio. For the remainder of the basins, the total error is approximately 15 to 25% of the amplitude of water storage variation.

5. Caveats

5.1. Basin Size Limits

[41] The synthetic GRACE data used in this study were derived from a global-gridded hydrologic model, at a resolution of 1 degree, which corresponds to about 110 km at the equator. Correlations at shorter length scales, which may occur in the true water storage signal, cannot be resolved with this model. For this reason, we have shown results only for basins having areas 1 to 2 orders of magnitude greater than this resolution. Small-scale correlations then can be ignored, because their contribution to the

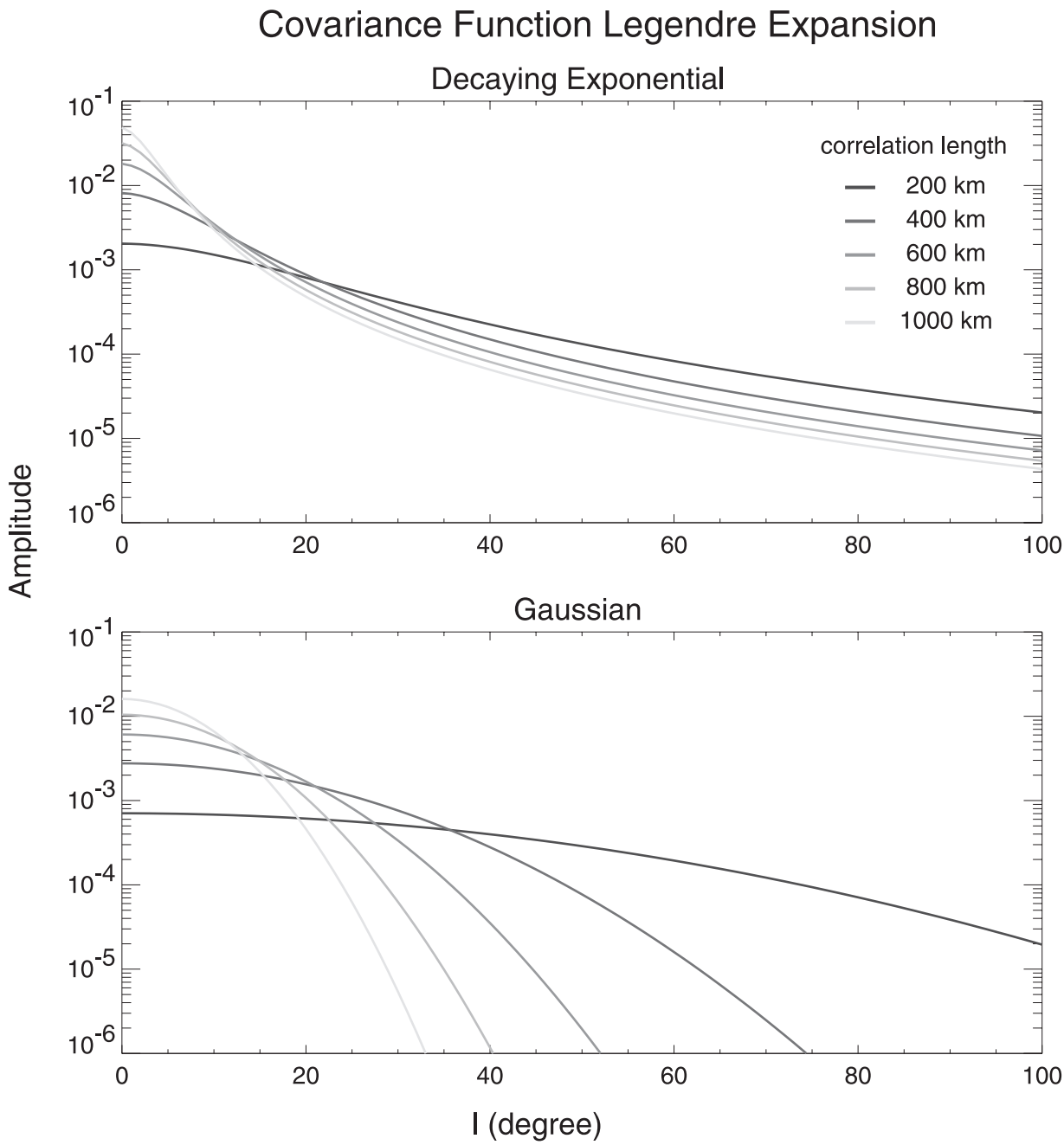


Figure 5. Comparison of the Legendre spectrum, G_l, for a decaying exponential CF and a Gaussian CF, for different values of d, the correlation length.

average approaches zero for large areas. It may be possible to recover changes in water storage for smaller basins; however, accuracy assessments must include models having sufficiently fine resolutions to adequately represent possible small-scale spatial variability.

5.2. Postglacial Rebound

[42] Because GRACE has no vertical resolving power, any unmodeled time-variable gravity signal originating within the solid Earth could be mistakenly interpreted as a surface mass signal, and so could contaminate continental water storage estimates. The solid Earth signal likely to cause the most serious problems is postglacial rebound (PGR), which is the ongoing, viscoelastic response of the solid Earth to the deglaciation that occurred at the end of the last ice age (for recent reviews, see *Wu* [1998] and *Mitrovica and Vermeersen* [2002]). A great deal is known about the PGR process, but the signal cannot yet be modeled well enough to allow it to be adequately removed from GRACE data. PGR deformation depends on the Earth's viscosity profile and the deglaciation history, and there are simply not enough PGR observations to constrain these quantities with complete confidence. Moreover, even the level of uncertainty in the PGR models is not well understood.

[43] PGR signals were not included in our simulated GRACE data, so their effects were not considered in the results described above. Here we try to quantify those effects. We note that because the PGR process has charac-

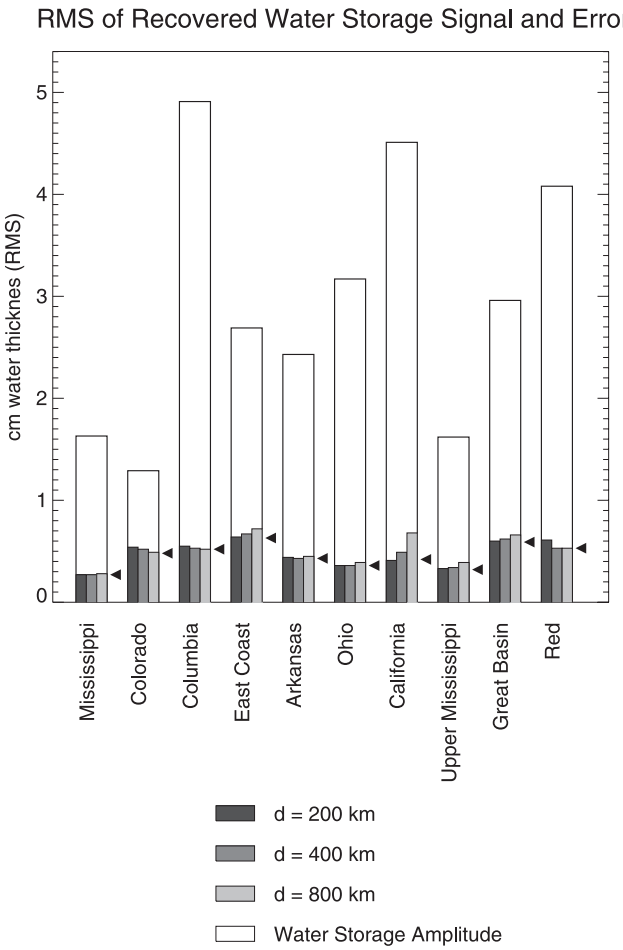


Figure 6. Plot showing amplitude of water storage variation and total error for averaging kernels based on decaying exponential CFs having correlation lengths of 200, 400, 800 km, respectively.

teristic timescales of several centuries and longer, its gravity signal will appear as a linear trend over the 5-year lifetime of GRACE. It will thus contaminate GRACE estimates of the linearly varying components of water storage, but will have no impact on the recovery of seasonal or other non-secular signals.

[44] To estimate the possible effects on GRACE estimates of the linearly varying water storage, we modeled the PGR contributions to the Stokes coefficients using a collocation technique described by *Wahr et al.* [2001]. We used the ICE-3G deglaciation model of *Tushingham and Peltier* [1991] and assumed a viscosity profile for the Earth of 10^{22} Pa s in the lower mantle (below 670 km depth) and 10^{21} Pa s in the upper mantle (between 670 km depth and the base of the elastic lithosphere), with a lithospheric thickness of 120 km.

[45] Suppose no attempt were made to remove a PGR model from the GRACE data, prior to determining the water storage variability. The resulting contamination of a regional water storage estimate can be estimated by setting ΔC_{lm} and ΔS_{lm} in (2) equal to our predicted PGR Stokes coefficients. Rather than looking at specific basins, we have elected to estimate surface mass anomalies at every point on the globe, without spatial averaging. Thus we have computed surface mass anomalies at every point (θ, ϕ), using [see *Swenson and Wahr*, 2002]:

$$\Delta\sigma(\theta,\phi) = \frac{a\,\rho_E}{3}\sum_{l=0}^{\infty}\sum_{m=0}^{l}\frac{(2l+1)}{(1+k_l)}\tilde{P}_{lm}(\cos\theta) \cdot\{\Delta C_{lm}\cos m\phi + \Delta S_{lm}\sin m\phi\}, \quad (10)$$

[46] The resulting linear trends in surface mass in cm/yr of equivalent water thickness, over the entire Northern Hemisphere and over just North America, are shown in Figures 7a and 7b, respectively. These trends can be up to a few cm/yr in the vicinity of Hudson Bay and in Scandinavia, but are smaller than 0.5 cm/yr over most of the globe, including over most of the US. The trends over land in the Southern Hemisphere (not shown) are on the order of ±0.25 cm/yr or smaller everywhere except on Antarctica where they can be much larger. This is the level of PGR error a user must be prepared to accept, if no PGR model is removed prior to solving for the water storage. Because this error varies linearly with time, it can be fit and removed but this would also remove any real linear trend in the true water storage signal.

[47] Suppose a PGR model is removed from the GRACE gravity data prior to the construction of a water storage estimate. How much improvement could be expected? To address this question, we constructed a second PGR model for the Stokes coefficients, using a different deglaciation history (ICE-4G from *Peltier* [1994]), and a different lower mantle viscosity (5×10^{22} Pa s). We subtracted this second model from the first model, and used those differenced Stokes coefficients in (10). The results over North America are shown in Figure 7c. We interpret these results as a crude measure of the errors that might remain in the linearly varying GRACE water storage estimates after a PGR model has been removed. It appears as though the removal of a PGR model might reduce the water storage errors by about a factor of two.

[48] This combination of ice sheet and viscosity parameters was chosen to maximize the difference between models, in the sense that the trends in the Stokes coefficients are larger for ICE-3G than for ICE-4G, and are also larger for a lower mantle viscosity of 1×10^{22} Pa s than for 5×10^{22} Pa s. We certainly can not claim that these two models are end-members of the set of all plausible models. On the

Table 3. RMS Basin Average Water Storage and Total Recovery Error[a]

Basin	Water Storage	Correlation Length		
		d = 200 km	d = 400 km	d = 800 km
Mississippi	1.63	0.27	0.27	0.28
Colorado	1.29	0.54	0.52	0.49
Columbia	4.91	0.55	0.53	0.52
East Coast	2.69	0.64	0.67	0.72
Arkansas	2.43	0.44	0.43	0.45
Ohio	3.17	0.36	0.36	0.39
California	4.51	0.41	0.49	0.68
Upper Mississippi	1.62	0.33	0.34	0.39
Great Basin	2.96	0.60	0.62	0.66
Red	4.08	0.61	0.53	0.53

[a]In centimeters of water.

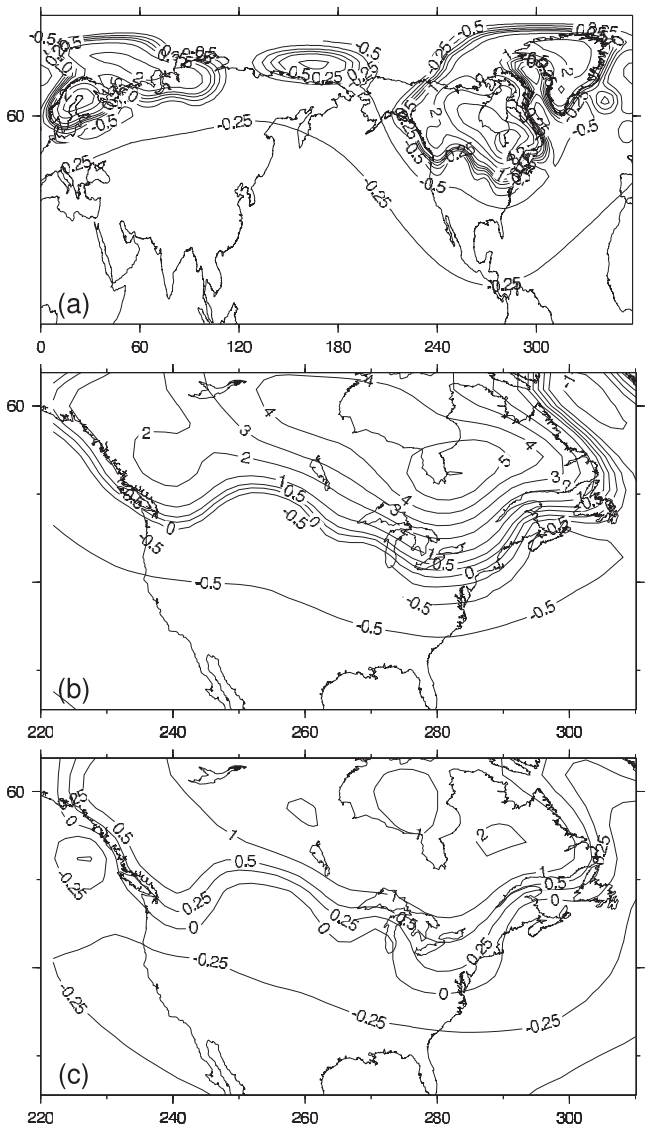


Figure 7. The predicted error in the linearly varying, GRACE water storage signal, due to postglacial rebound. (a and b) The error predicted using the rebound model described in the text. (c) An estimate of how large the error over North America might be if a rebound model is first removed from the GRACE data. In each panel, the contour interval is 1 cm/yr of water thickness, although contours at ±0.25 and ±0.5 cm/yr are also included.

other hand, it is possible that this difference could overestimate the errors because the ICE-3G and ICE-4G models were constructed to be consistent with specific viscosity profiles, which differ from the profiles we have somewhat arbitrarily chosen for this comparison. Furthermore, the secular Stokes coefficients determined from GRACE by the end of its five-year lifetime should help considerably in improving the PGR models [*Velicogna and Wahr*, 2002].

6. Summary

[49] The accuracy of GRACE estimates of water storage variability within a region depends on the GRACE measurement errors, and the degree to which the gravity signal from the water storage can be separated from other time-variable gravity signals. The signal separation problem can be severe. Gravity measurements made outside the Earth, whether from satellites or surface gravimeters, provide no information about the vertical distribution of mass. Using GRACE gravity data alone, there is no way to tell whether a time-variable gravity signal is caused by mass variability at the Earth's surface, in the atmosphere, or deep within the mantle. Thus GRACE data can be used to constrain only the vertically integrated water storage variability, and cannot separate soil moisture from surface water or from water deeper underground. For this reason, the GRACE Project will use ECMWF pressure and geopotential height fields to remove the effects of atmospheric mass from the data before releasing those data to the public. In our simulated data we have included estimates of the errors likely to be present in those atmospheric corrections.

[50] Gravity data are capable of separating mass signals from different horizontal positions. If GRACE could provide error-free Stokes coefficients to arbitrarily large angular degree, a linear combination of those coefficients could be constructed that could perfectly separate the water storage signal in the region of interest from water storage signals in surrounding regions. GRACE data do have errors, however, and those errors increase rapidly with decreasing length scale. As shown in section 3.1.1, a regional average made with the exact averaging kernel will display the effects of these measurement errors, and those effects can rapidly become cripplingly large as the size of the region decreases.

[51] Several methods for analyzing GRACE data to obtain regional averages of water storage variability, while reducing the effects of measurement errors and contamination from mass variability in surrounding regions, are described by *Swenson and Wahr* [2002]. Section 3.1.2 shows an example of an averaging kernel constructed in the absence of a priori assumptions regarding the water storage signal. The leakage definition used to construct the averaging kernel in this example is geometric; it measures the departure of the shape of the smoothed averaging kernel from that of the exact averaging kernel [see *Swenson and Wahr*, 2002, equation 42]. However, unless the leakage error can be expressed in terms of the physical variable being measured (in this case, the vertically integrated water storage), the accuracy of the GRACE recovery cannot be directly assessed. In this study, for example, we use global-gridded models of surface-mass variability and error sources to construct synthetic GRACE data. Comparing the estimates of changes in water storage, recovered from this synthetic data set using an approximate averaging kernel, to the value derived directly from the model allows one to estimate the leakage error, as well as the satellite measurement error.

[52] When independent hydrologic information, such as the modeled data used in this study, is available, it is possible to construct an averaging kernel that is optimized to minimize the total error in the recovery. *Swenson and Wahr* [2002] showed how to create an optimal averaging kernel with an azimuthally symmetric model of the signal covariance function. The use of this model is encouraged by studies such as *Rodell and Famiglietti* [2001], who determined that the largest component of variability in the American Midwest is soil moisture, and *Entin et al.*

[2000], who showed that spatial correlation functions made from midlatitude observations of soil moisture may be approximated by an exponential function that decays with distance. Although water storage in the intermediate/unsaturated zone and groundwater are both included in the LaD model, we know of no observations with which to compare the large-scale spatial correlation properties of these deeper regions. An additional reason for using the azimuthally symmetric covariance function is that the functional form may be described by just two parameters: the correlation length, d, and the local signal variance, σ_0^2.

[53] In this study, we have applied averaging kernels based on the azimuthally symmetric covariance function model to simulated GRACE data to extract the water storage signal in various North American river basins. Our objective is to assess the ability of GRACE to recover the water storage variability in those regions. Averaging kernels based on two models of the covariance function, one a Gaussian and the other a decaying exponential, were able to recover the water storage signal with nearly equal accuracies. However, results obtained using the Gaussian model varied relatively more rapidly as a function of the assumed correlation length than did results obtained with the exponential model. The relative insensitivity of the exponential model to correlation length implies that averaging kernels based on this model are more robust with inaccurately determined values of d.

[54] After the correlation length is specified, the local signal variance, σ_0^2, and the basin-averaged water storage anomaly, $\widetilde{\sigma}$, can be determined concurrently using an iterative approach. Thus d is the only model parameter that must be determined from independent information. Our results indicate that with this approach, GRACE will be able to provide estimates of water storage change to an accuracy of about 0.7 cm equivalent water thickness for a basin with an area of $4.0 \cdot 10^5$ km^2, and an accuracy of about 0.3 cm equivalent water thickness for a basin with an area of $3.9 \cdot 10^6$ km^2.

[55] **Acknowledgments.** We thank Brooks Thomas and Mike Watkins for providing their estimate of the GRACE measurement errors, Mery Molenaar and Frank Bryan for their estimates of the oceanic Stokes coefficients, and Dick Peltier for providing his ICE-4G ice deglaciation model. We also thank Kirsten Findell, Don Pool, and Dorothy Tepper for providing helpful reviews of the manuscript. This work was partially supported by NASA grants NAG5-9450 and NAG5-7703 and NOAA grant NA96GP045 to the University of Colorado, by a CIRES Innovative Research Grant, and a NASA Earth Systems Science Fellowship awarded to Sean Swenson.

References

Chao, B. F., and R. S. Gross, Changes in the Earth's rotation and low-degree gravitational field induced by earthquakes, *Geophys. J. R. Astron. Soc.*, *91*, 569–596, 1987.

Dickey, J. O., et al., *Satellite Gravity and the Geosphere: Contributions to the Study of the Solid Earth and Its Fluid Envelope*, 112 pp., Natl. Acad. Press, Washington, D. C., 1997.

Dukowicz, J. K., and R. D. Smith, Implicit free-surface method for the Bryan-Cox-Semtner ocean model, *J. Geophys. Res.*, *99*, 7991–8014, 1994.

Entin, J., A. Robock, K. Vinnikov, S. Hollinger, S. Liu, and A. Namkhai, Temporal and spatial scales of observed soil moisture variations in the extratropics, *J. Geophys. Res.*, *105*(D9), 11,865–11,877, 2000.

Jet Propulsion Laboratory, GRACE Science and Mission Requirements Document, *GRACE 327-200, JPL Publ. D-15928, Rev. D*, Pasadena, Calif., 2001.

Milly, P. C. D., and A. B. Shmakin, Global modeling of land water and energy balances. part I: The land dynamics (LaD) model, *J. Hydrometeorol.*, *3*(3), 283–299, 2002.

Mitrovica, J. X., and L. L. A. Vermeersen (Eds.), *Glacial Isostatic Adjustment and the Earth System: Sea-Level, Crustal Deformation, Gravity and Rotation*, *Geodyn. Ser.*, vol. 29, AGU, Washington, D. C., 2002.

Peltier, W. R., Ice-age paleotopography, *Science*, *265*, 195–201, 1994.

Rodell, M., and J. S. Famiglietti, Detectability of variations in continental water storage from satellite observations of the time dependent gravity field, *Water Resour. Res.*, *35*, 2705–2723, 1999.

Rodell, M., and J. S. Famiglietti, An analysis of terrestrial water storage variations in Illinois with implications for GRACE, *Water Resour. Res.*, *37*, 1327–1339, 2001.

Rodell, M., and J. S. Famiglietti, The potential for satellite-based monitoring of groundwater storage changes using GRACE: The High Plains aquifer, central US, *J. Hydrol.*, *263*, 245–256, 2002.

Shmakin, A. B., P. C. D. Milly, and K. A. Dunne, Global modeling of land water and energy balances. part III: Interannual variability, *J. Hydrometeorol.*, *3*(3), 311–321, 2002.

Swenson, S., and J. Wahr, Methods for inferring regional surface-mass anomalies from GRACE measurements of time-variable gravity, *J. Geophys. Res.*, *107*(B9), 2193, doi:10.1029/2001JB000576, 2002.

Tushingham, A. M., and W. R. Peltier, ICE-3G: A new global model of late Pleistocene deglaciation based upon geophysical predictions of postglacial relative sea level change, *J. Geophys. Res.*, *96*, 4497–4523, 1991.

Velicogna, I., and J. Wahr, Postglacial rebound and Earth's viscosity structure from GRACE, *J. Geophys. Res.*, *107*(B12), 2376, doi:10.1029/2001JB001735, 2002.

Velicogna, I., J. Wahr, and H. Van den Dool, Can surface pressure be used to remove atmospheric contributions from GRACE data with sufficient accuracy to recover hydrological signals?, *J. Geophys. Res.*, *106*, 16,415–16,434, 2001.

Vinnikov, K., A. Robock, N. Speranskaya, and C. Schlosser, Scales of temporal and spatial variability of midlatitude soil moisture, *J. Geophys. Res.*, *101*, 7163–7174, 1996.

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

Wahr, J., T. van Dam, K. Larson, and O. Francis, Geodetic measurements in Greenland and their implications, *J. Geophys. Res.*, *106*, 16,567–16,582, 2001.

Wu, P., (Ed.), *Dynamics of the Ice Age Earth: A Modern Perspective*, Trans Tech Publ., Zurich, Switzerland, 1998.

P. C. D. Milly, Geophysical Fluid Dynamics Laboratory, NOAA, Princeton, NJ 08542, USA.

S. Swenson and J. Wahr, Department of Physics and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Campus Box 390, Boulder, CO 80309-0390, USA. (swensosc@colorado.edu; wahr@lemond.Colorado.edu)