Accuracy of GRACE mass estimates

John Wahr,[1] Sean Swenson,[1] and Isabella Velicogna[1]

Received 23 November 2005; revised 11 January 2006; accepted 7 February 2006; published 18 March 2006.

[1] The GRACE satellite mission is mapping the Earth's gravity field at monthly intervals. The solutions can be used to determine monthly changes in the distribution of water on land and in the ocean. Most GRACE studies to-date have focussed on producing maps of mass variability, with little discussion of the errors in those maps. Error estimates, though, are necessary if GRACE is to be used as a diagnostic tool for assessing and improving hydrology and ocean models. Furthermore, only with error estimates can it be decided whether some feature of the data is real, and how accurately that feature is determined by GRACE. Here, we describe a method of constructing error estimates for GRACE mass values. The errors depend on latitude and smoothing radius. Once the errors are adjusted for these factors, we find they are normally-distributed. This allows us to assign confidence levels to GRACE mass estimates. **Citation:** Wahr, J., S. Swenson, and I. Velicogna (2006), Accuracy of GRACE mass estimates, *Geophys. Res. Lett.*, *33*, L06401, doi:10.1029/2005GL025305.

1. Introduction

[2] GRACE (Gravity Recovery and Climate Experiment), managed jointly by NASA and DLR, was launched in March, 2002 [*Tapley et al.*, 2004]. As of January, 2006, 22 monthly gravity field solutions had been released to users. Solutions consist of spherical harmonic (Stokes) coefficients, C_{lm} and S_{lm}, complete to degree and order (l and m) 120. GRACE does not recover $l = 1$ terms, and the C_{20} coefficients show anomalously large variability. We do not include those terms in this analysis.

[3] Time variations in the gravity field can be used to determine changes in the Earth's mass distribution. At the temporal and spatial scales of GRACE, the gravity signal mostly reflects mass variations within the atmosphere, oceans, and water stored on land. GRACE has no vertical resolution. It is not possible to tell whether a mass variation inferred for some region on land is caused by changes in water on the surface, in water below ground, or in atmospheric mass above the region. Users must employ independent means to separate those contributions.

[4] Suppose the goal is to use GRACE mass variations to assess a regional water storage model. The GRACE results would be interpreted as estimates of total water storage variability. The errors in those estimates fall into two categories: (i) those due to errors in the monthly GRACE gravity field solutions; (ii) those due to changes in the true monthly mass averages caused by things other than continental water storage. Measurement and processing errors contribute to (i). Contributions to (ii) could include gravity signals caused, for example, by unmodeled mass variations in the Earth's interior.

[5] The atmosphere contributes to both categories. The GRACE Project uses ECMWF meteorological fields to remove atmospheric effects from the raw data before constructing gravity fields. But there are errors in the ECMWF fields. The atmospheric mass signal during a month can be decomposed into the average for that month, plus variability about the monthly average. An error in the ECMWF monthly average would contribute to (ii). Errors in the variability during the month do not cause errors in the true monthly signal. But they can alias into the GRACE monthly fields in a way that depends on the GRACE ground track during that month. These aliasing errors fall into category (i). We use a similar method to classify contributions from errors in the tidal and non-tidal ocean models used to reduce oceanic effects prior to constructing gravity fields.

[6] In principle, category (ii) errors can be reduced by GRACE users once better monthly averaged atmospheric or oceanic results become available. Category (i) errors (measurement, processing, and aliasing errors) can not be reduced without re-generating the gravity field solutions. Our goal is to estimate the sum of all category (i) errors.

[7] Error estimates for remote sensing missions often rely on ground truth validation. This is a prerequisite for obtaining meaningful error estimates for missions where the observed variables (e.g., brightness temperatures) are related to the quantities of interest (e.g., surface characteristics) through a set of initially uncalibrated parameters.

[8] It is difficult, at present, to use ground truth to validate GRACE mass estimates. It is not clear there is any place where monthly variations of vertically integrated mass, averaged over the area of a GRACE footprint (probably at least 10^6 km^2 given the present GRACE accuracy levels), are monitored well enough. Fortunately, the physics that relates the GRACE observables (orbital motion) to the end products (mass anomalies) is well known, following from Newton's law of gravity and the second law of motion. This reduces the need for validation, and makes it feasible to construct error estimates using only the GRACE fields. Our method is an extension of that described by *Wahr et al.* [2004].

2. Gravity Field Errors

[9] The construction of mass error estimates is a 2-step process. First, and discussed in this section, we derive Stokes coefficient uncertainties. Second, those uncertainties are used to estimate mass errors.

[10] The GRACE data release includes estimates of "calibrated errors" in the Stokes coefficients: diagonal

[1]Department of Physics and Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado, USA.

0094-8276/06/2005GL025305

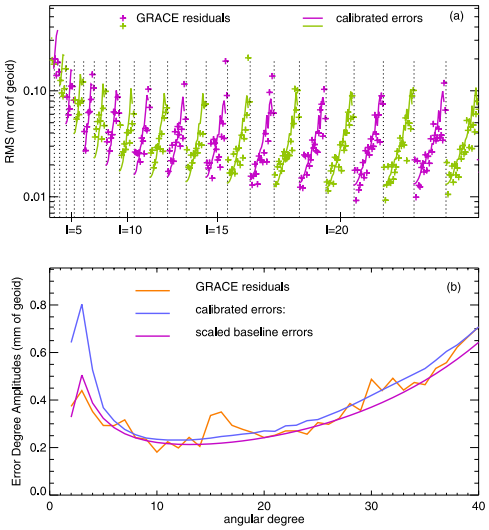

Figure 1. Compares Stokes coefficients obtained from all 22 GRACE fields, after removing a constant and an annual cycle, with the average of the 22 calibrated error fields. Results converted to mm of geoid by multiplying by the Earth's radius. (a) Compares RMS values of the GRACE residual coefficients (plus signs) with the calibrated errors (solid lines). The vertical dotted lines delineate values of the degree l, and the order m increases from left to right within those lines. (b) Compares degree amplitudes of the calibrated errors, the GRACE residuals, and a smoothed version of the GRACE residuals computed by scaling the baseline errors.

elements of the covariance matrix, rescaled by the Project to match certain characteristics of the fields. These error estimates are meant to represent all category (i) errors in the gravity field solutions.

[11] We independently assess the calibrated errors by fitting and removing a constant and an annual cycle from the 22 monthly values of each GRACE Stokes coefficient. We assume the residuals are due entirely to category (i) gravity field errors. This overestimates those errors, since some of the non-annual variability is surely real. Conversely, if the GRACE errors include systematic annual components our removal of the annual cycle would cause an underestimate of the error. At present the errors are believed to be largely free of annual components (S. Bettadpur, personal communication). We find the RMS of the 22 residuals for each Stokes coefficient, and compare with the calibrated errors. Before comparison, we multiply our RMS values by 1.05, to compensate for the fact that when a constant and annual cycle are removed from 22 random numbers with the GRACE temporal spacing, the average RMS is reduced by 5% (deduced by fitting to simulated random numbers).

[12] Results are shown in Figure 1a for degrees $l \leq 25$ (C_{20} not included). The plus signs are our RMS values, and the solid lines are the calibrated errors. The general agreement is good, except for degrees $l \leq 5$ where the calibrated errors are too large. This is evident in Figure 1b, which shows degree amplitudes ($\sqrt{\sum_m (C_{lm}^2 + S_{lm}^2)}$) of both our residual values (orange) and the calibrated errors (blue).

[13] To obtain our best estimate of the Stokes coefficient errors, we retain the m-dependence of the calibrated errors, but multiply all m-dependent values of the same degree by a degree-dependent scaling factor. The factor is chosen so the degree amplitudes of the results are in good agreement with the degree amplitudes of our RMS values. We could use a scaling factor that gives exact agreement at every degree, so that the degree amplitudes of the scaled calibrated errors agreed with every variation of the orange line in Figure 1b. But it is possible that much of that variability is a consequence of sampling the underlying error distribution function with just 22 points. Instead, we choose the scaling factor so the scaled degree amplitudes agree with a smoothed version of the orange line, shown as the purple line in Figure 1b. This smoothed version is equal to a constant times the error degree amplitudes of the baseline performance target [*Jet Propulsion Laboratory*, 2001]. The constant is determined by fitting the baseline degree amplitudes to the orange line.

3. Relating Mass and Gravity Field Errors

[14] Let C_{lm}^i and S_{lm}^i be the GRACE Stokes coefficients for monthly field i (i = 1, .., N, where N = 22), after removing the long-term average. Let σ_i be a GRACE mass estimate obtained from those coefficients; σ_i could be a Gaussian-smoothed mass anomaly about a point (equation (30) of *Wahr et al.* [1998]), or an optimized regional average (equation (2) of *Swenson et al.* [2003]), etc. σ_i is linearly related to the Stokes coefficients as:

$$\sigma_i = \sum_{l,m} \left[F_l^m C_{lm}^i + G_l^m S_{lm}^i \right] \tag{1}$$

where F_l^m and G_l^m are time-independent coefficients defining the averaging kernel.

[15] Let $\delta\sigma_i$, δC_{lm}^i, and δS_{lm}^i represent errors in the mass values and in the Stokes coefficients. Then

$$\delta\sigma_i = \sum_{l,m} \left[F_l^m \delta C_{lm}^i + G_l^m \delta S_{lm}^i \right] \tag{2}$$

The RMS of the mass errors is:

$$\Delta = \sqrt{\sum_{i=1}^{N} \frac{\delta\sigma_i^2}{N}} = \left[\sum_{l,m,p,q} \left(F_l^m F_p^q \left[\sum_{i=1}^{N} \frac{\delta C_{lm}^i \delta C_{pq}^i}{N} \right] + 2 F_l^m G_p^q \left[\sum_{i=1}^{N} \frac{\delta C_{lm}^i \delta S_{pq}^i}{N} \right] + G_l^m G_p^q \left[\sum_{i=1}^{N} \frac{\delta S_{lm}^i \delta S_{pq}^i}{N} \right] \right) \right]^{\frac{1}{2}} \tag{3}$$

The quantities within the square brackets are elements of the covariance matrix.

[16] The GRACE Project provided us with a covariance matrix for a single month (August, 2003). We find that the inclusion of off-diagonal elements (i.e., where $(p, q) \neq (l, m)$) has little impact on Δ. If we ignore those elements,

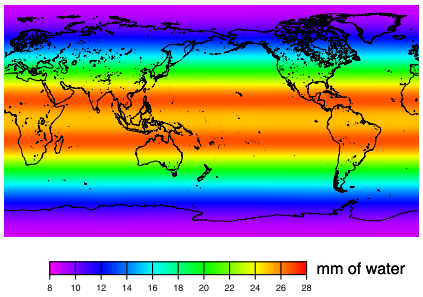

Figure 2. Our estimated uncertainties in the GRACE mass estimates, in mm of water thickness, for 750-km Gaussian averages and averaged over all 22 months. Obtained by propagating the Stokes coefficient errors through (4).

equivalent to ignoring correlations between different Stokes coefficients, (3) reduces to

$$\Delta = \sqrt{\sum_{l,m} \left(F_{lm}^2 \left[\sum_{i=1}^{N} \frac{\delta C_{lm}^{i^2}}{N} \right] + G_{lm}^2 \left[\sum_{i=1}^{N} \frac{\delta S_{lm}^{i^2}}{N} \right] \right)} \qquad (4)$$

4. Mass Errors

[17] We estimate mass errors by using our Stokes coefficient errors (section 2) in place of the quantities inside the square brackets in (4). Figure 2 shows the resulting mass errors (l = 1 and C_{20} omitted), expressed in mm of water thickness, when F_l^m and G_l^m represent Gaussian averages with a 750-km smoothing radius. The errors are nearly longitude-independent, and are smaller near the poles (8 mm) than at low latitudes (25–27 mm), due presumably to denser ground track coverage near the poles. The global, area-weighted mean is 21 mm.

[18] The errors vary from one month to another. To estimate the errors for an individual month we follow the section 2 procedure for finding Stokes coefficient errors, except we scale the calibrated errors to match the (smoothed) degree amplitudes of the residuals for that month only. Our error estimate for the Gaussian averages for each month has the spatial dependence shown in Figure 2, but a different overall amplitude. Figure 3 (circles) shows the global, area-weighted mean of the results for each month.

[19] The mass errors vary with the size of the region. Figure 3 (solid line) shows the global, area-weighted mean of Gaussian averages, averaged over all 22 months, as a function of smoothing radius. The errors decrease as the radius increases, falling from 38 mm at 500 km to 15 mm at 1000 km.

[20] Figures 2 and 3 show errors for Gaussian-averages of mass. Errors for specially constructed regional mass averages can be computed as described above, except using F_l^m and G_l^m coefficients appropriate for those averaging kernels.

5. Confidence Levels

[21] RMS values, such as those shown in Figures 2 and 3, are of practical value only if they can be used to assign confidence levels. That requires knowledge of the underlying error probability distribution. To find that distribution for 750-km Gaussian-averages, we use the residual Stokes

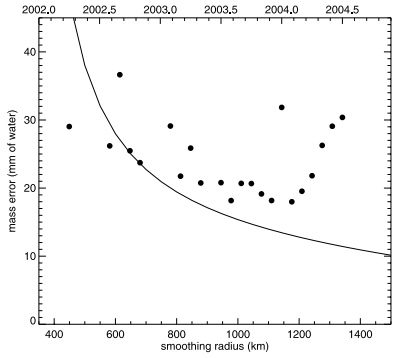

Figure 3. Global means of uncertainties in Gaussian-smoothed mass estimates; for individual GRACE fields and 750-km smoothing (circles, x-axis scale at the top); and as a function of smoothing radius, averaged over all 22 months (solid line, x-axis scale at the bottom).

coefficients in (1) to find mass values at every point in a 2° × 2° grid for each of the 22 months. As before, we assume each of these values is solely a consequence of gravity field errors. A histogram of values for all grid points and all 22 times should then reveal the probability distribution function. To construct this histogram, we divide each mass value by the expected error at its location and time: i.e., by the Figure 2 results, scaled to account for the monthly differences shown in Figure 3. This is necessary when combining values, since the results above demonstrate that the errors have significant latitude and (to a lesser extent) time dependence. We weight each grid point by its area-average when counting its contribution to the histogram, to equalize contributions from high- and low-latitudes.

[22] Figure 4 shows the histogram, along with a normal distribution with the same variance. The agreement is encouraging. For a normal distribution, 68.3% of all values lie within one RMS of the mean (95.4% lie within two RMS). This suggests that for 750-km Gaussian-averaged mass estimates, we can be 68.3% sure that the errors at any location and any time are smaller than the RMS values shown in Figure 2, scaled to account for the Figure 3 monthly differences.

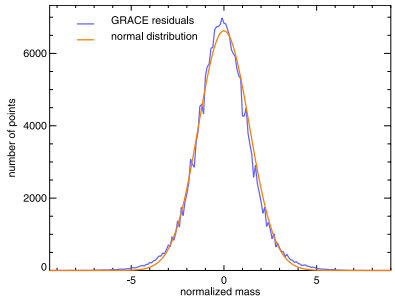

Figure 4. Histogram of all 750-km, Gaussian-smoothed mass residuals at each time and at every grid point in a 2° × 2° grid. The residuals have been normalized by dividing by the estimated error at that latitude and time; and each grid point contributes to the histogram in proportion to its area. Also shown is the probability distribution of a normal distribution with the same variance.

[23] This conclusion needs qualification. Figure 5 shows RMS values of the 750-km Gaussian mass residuals determined from the 22 monthly values at each location, with each month weighted to offset the monthly-dependence shown in Figure 3. If the errors at every location were normally-distributed with the RMS shown in Figure 2, then Figure 5 and Figure 2 would look similar. But there is too much longitudinal variability in Figure 5, even recognizing that RMS values computed from just 22 samples of a normal distribution will not yield the exact RMS of the underlying distribution. Models (not shown) indicate that some of the variability, such as in South America and the Indian subcontinent, as well as much of the increased RMS at high northern latitudes, are caused by real non-annual signals. But even over the ocean, where the true mass signals are likely to be smaller than over land, there is significant longitudinal variability. Figure 6 shows a histogram of the Figure 5 RMS values at all grid points, normalized by dividing by the RMS values shown in Figure 2 and contributing to the histogram in proportion to the grid point area. The histogram is compared with the expected distribution of RMS values if each data point consisted of 22 values taken from a normal distribution with the same variance. Results are shown for the entire Earth and for the oceans alone. Although the agreement is reasonably good, there is clearly a larger spread in the GRACE RMS values than expected for a normal distribution. The fact that the GRACE RMS distribution looks similar in the global and ocean-only cases, suggests the presence of real hydrology signals is not causing the non-normal shape.

[24] Our interpretation is that during these 22 months the errors at some locations had more variability than at other locations; more than expected given the natural variability of sets of 22 numbers taken from a normal distribution. The implication is that the RMS estimates shown in Figure 2, which are used to normalize the values used in the histogram, are an imperfect representation of the true errors. This could be due to incorrect m-dependence of the calibrated errors, which we scaled and used in (4) to find the mass errors. Or it could be that the effects of off-diagonal elements in the covariance matrix are larger than we deduce from the August, 2003 covariance matrix (section 3).

6. Discussion

[25] We have argued that the errors in Gaussian-smoothed mass estimates are normally distributed. Although we have shown figures only for 750-km radii, results for other radii are similar. The implication is that when a Gaussian-averaged mass value is computed for one of the GRACE fields, the user can be 68.3% confident the true mass value lies within one RMS of the computed value, where one RMS for that location and month can be found from Figures 2 and 3 for a 750-km radius, and can be computed for other radii as described in this paper (and as summarized in Figure 3). We believe these error estimates are conservative, since they are constructed assuming all non-annual components of the data are caused by errors. Confidence levels for other types of averaging kernels can be found by repeating the steps described in this paper using different F_l^m's and G_l^m's in (4). Whatever the averaging kernel, the monthly RMS estimates can be used when fitting annual, secular, etc terms to the monthly mass values, to construct 68.3% confidence limits on those terms as well.

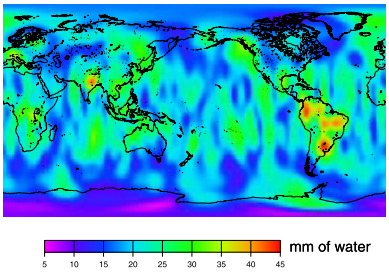

Figure 5. The RMS, as a function of location, of the 22 monthly mass fields (Gaussian-smoothed with a 750-km radius), after removing the annual cycle.

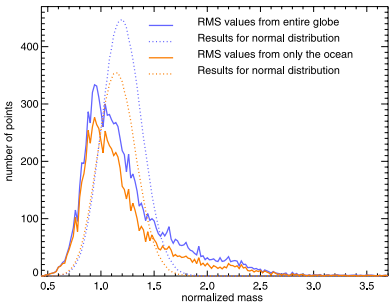

Figure 6. Histograms of the RMS values shown in Figure 5, after normalizing those values by dividing by the uncertainties shown in Figure 2, and with each grid point contributing to the histogram in proportion to its area. Results are shown for all grid points, and for all ocean grid points. Also shown are corresponding results for normal distributions with the same variances.

[26] The RMS values do not include all the position-dependent complexities of the true errors. As a result, locations where the errors exceed our 68.3% confident levels tend to be grouped together, as do locations where the errors are much smaller than that level. One might be tempted to use the RMS values shown in Figure 5, instead of the errors shown in Figure 2, to construct confidence levels. But this is not correct either, since many of the prominent features shown in Figure 5 are real geophysical signals. Until more is known about the true m-dependence of the Stokes coefficient errors and the correlations between errors in different Stokes coefficients, the RMS values shown in Figure 2 provide relatively simple and realistic error bars for the fields released to-date. Our method of estimating those error bars should remain useful in the future as the fields improve. Though it may be that terms besides an annual cycle will eventually also have to be removed, to avoid overestimating the errors.

[27] **Acknowledgments.** We thank S. Bettadpur, J. Ries, V. Zlotnicki, and an anonymous referee for discussions and helpful reviews. This work was supported by NASA grant NNG04GF02G, and NSF grants OPP-0324721 and EAR-0309678.

References

Jet Propulsion Laboratory (2001), GRACE science and mission requirements document, *JPL Publ., D-15928, Rev. D.*, Pasadena, Calif.

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

Tapley, B. D., S. Bettadpur, M. Watkins, and C. Reigber (2004), The Gravity Recovery and Climate Experiment: Mission overview and early results, *Geophys. Res. Lett.*, *31*, L09607, doi:10.1029/2004GL019920.

Wahr, J., M. Molenaar, and F. Bryan (1998), Time-variability of the Earth's gravity field: Hydrological and oceanic effects and their possible detection using GRACE, *J. Geophys. Res.*, *103*, 30,205–30,230.

Wahr, J., S. Swenson, V. Zlotnicki, and I. Velicogna (2004), Time-variable gravity from GRACE: First results, *Geophys. Res. Lett.*, *31*, L11501, doi:10.1029/2004GL019779.

J. Wahr, S. Swenson, and I. Velicogna, Department of Physics and CIRES, University of Colorado, Campus Box 390, Boulder, CO 80309-0390, USA. (wahr@lemond.colorado.edu)