Response to Reviewer #2 (Anonymous)

**It has to be said that a related piece of work has recently appeared (Talpe et al. 2017)…*

Yes. I also recently read this paper (and talked to the author), and put a brief note in the methods section of my appendix about the differences of method and similarities of general results that we found.

**Replacement of C20 by SLR derived estimates This issue has already been mentioned by Matt King in his short comment. So this is to reconfirm that this issue also stroke me as somewhat tricky. By replacing C20 by an SLR-estimate a dependency is introduced which may be favorable for the CSR-SLR solution in the comparison. To clear this up, maybe the authors could show how much C20 contributes to the estimated time series.*

As I also said to Matt, I think this is a good point, since the influence of C20 is so large at the poles. You're both right that using very similar C20 terms for GRACE and the Cheng SLR series might bias them toward each other for reasons that have nothing to do with GRACE itself. However, because the C20 terms are such a big part of the final signal, I didn't really want to produce this paper by totally excluding it. Instead, to answer your question, I decided to test what the impact of removing it was, to see if it was reducing the divergence I see between GRACE and the Cheng SLR series.

So I recreated each of the three main series (GRACE, Cheng 5x5 SLR and Sosnica 10x10 SLR) and totally omitted the C20 terms, then inverted each and took a look at the time series. If the C20 term was causing falsely alignment with GRACE, I would see a larger divergence between GRACE and the Cheng series, in which case, my paper would require revising.

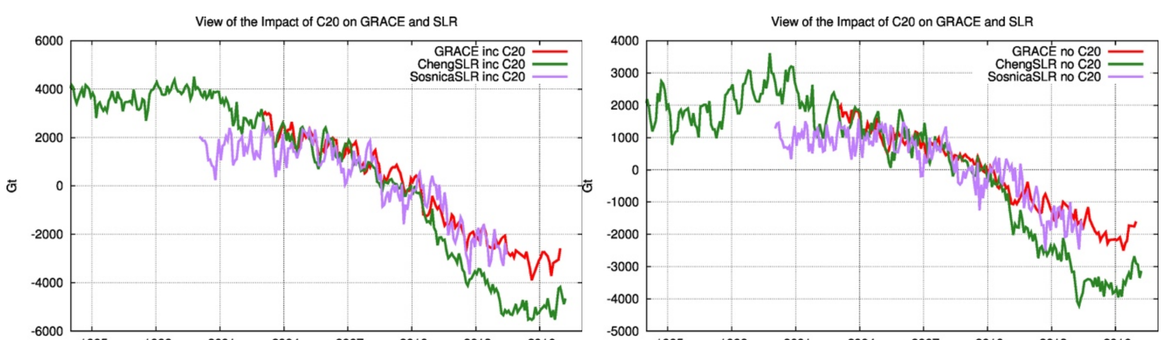


Figure 1: Left-hand image is the inversion with C20 included. Right-hand image is the inversion with C20 totally removed.

However, I see no notable changes in terms of divergence. There are three main effects of removing the C20 terms. First, the overall trend of all three of the series dropped like a rock. (No surprise, given the geometry of the situation.) Second, the month-to-month jitter in all three of the series changed. Third, most oddly, removing the C20 term from the

Cheng series produced a large, visible annual signal before about 2007. The other series (including GRACE, using a similar C20) didn't show this impact. So that's bizarre. I assume that the C20 term in the Cheng series is coupled with some other term to produce this (which wouldn't especially surprise its creators, since they're aware of the general coupling between harmonics caused by a barely solvable problem).

In any case, there was not any significant change in the interannual signal divergence. So in practice, the replacement of the GRACE C20 should bias GRACE towards the Cheng SLR series doesn't seem to have any major effect on the part of the spectrum that I'm worried about. That's a relief.

I have created the following commentary for the final version of the paper, briefly discussing this:

> We did consider the impact of replacing the GRACE C_{20} term with that from a series related to the Cheng 5x5 SLR data. To test whether this unfairly biased the Cheng 5x5 SLR results toward GRACE, we removed the C_{20} terms completely from all of the GRACE and SLR series, then inverted each of them again. Removing the impact of the equatorial bulge did greatly reduce the trend of each Greenland+Antarctica inverted series, but it did not significantly impact the interannual differences between GRACE and any SLR series. We thus conclude that the replacement of GRACE's C_{20} values is not a large contributing factor to these results.

** Neglecting degree 1 contributions I understand the decision of the authors to not account for the degree 1 signal, based on remaining errors in the SLR data. However, the potential influence of degree 1 neglection may be too large to ignore. As an alternative, maybe the authors can treat the degree 1 signal as noise and assess its influence on the results by producing an ensemble of realistic variations and propagating this through the inversion?*

I previously ran a comparison of the Cheng SLR geocenter terms compared to those computed over the GRACE time-period with the technique of Swenson et al. I was surprised to find that the difference between the two geocenter estimates (in terms of their impact on the inverted timeseries) was about the same size as the difference between not using any geocenter and using the Swenson version. According to my coauthor Minkang Cheng and his colleague John Ries, much of this difference is likely to be an error in the SLR C_{10} term caused by the uneven distribution of the ground station network. Also, the difference (after inversion) was small over the combination of Greenland and Antarctica. Certainly, the geocenter term is not what is causing the divergence of SLR from GRACE after 2010, for example (I checked).

I do agree that if one was actually trying to measure the total mass loss of Greenland/Antarctica with this method, so as to compare to other similar estimates, a geocenter would be required. However, in this case, the discrepencies between SLR and GRACE are so large that the main point of the article is actually that one should NOT use 5x5 monthly SLR to push the estimate of mass change back in time. That being the case, the comparison can be run without geocenter being added (in either GRACE or SLR, to keep things equal).

**The supplement has a *.zip ending but actually is in *tgz format*

Sorry; I'll fix that.

** abstract: maybe add some numbers in the abstract to quantify things a bit more*

Which particular details would the reviewer care to have quantified?

** Does the average TC reader knows what is meant by 5x5, 10x10?*

Good point. I'll make sure that's defined initially.

** eq 1 shouldn't the '1-' be outside of the fraction?*

Corrected.

** " indicative of a systematic interannual-scale error in the SLR inversion" What is meant by this? Maybe add a reference, which illustrates the problem at hand?*

What I mean is that, while the GRACE-ChengSLR trend difference is 40% the size of the total trend for 2003-2014, I do not believe this really represents an inability of SLR to represent the long-term trend. Rather, I believe this to be a symptom of SLR's tendency to veer away from the GRACE "truth" for multiple years in a row, then correct itself and come back into alignment (as it seems to be doing in 2017, and as it also may have done back in 2002). Neither I nor Minkang Cheng know of no reference which discusses this, since the accuracy of SLR's interannual variability is very hard to quantify, particularly pre-GRACE. The 15-year record since GRACE started may not be long enough to quantify deviations which take 5+ years to resolve.

This line now reads: "So instead of representing a true, long-term error in trend, the large interannual differences between GRACE and the Cheng 5x5 SLR series are probably indicative of a systematic interannual-scale error in the SLR inversion, which cannot be well quantified given the relatively short length of the GRACE record." I hope that's clearer.

** "451 + 28 Gt/yr" I assume this is for Antarctica and Greenland? Maybe explicitly mention this again*

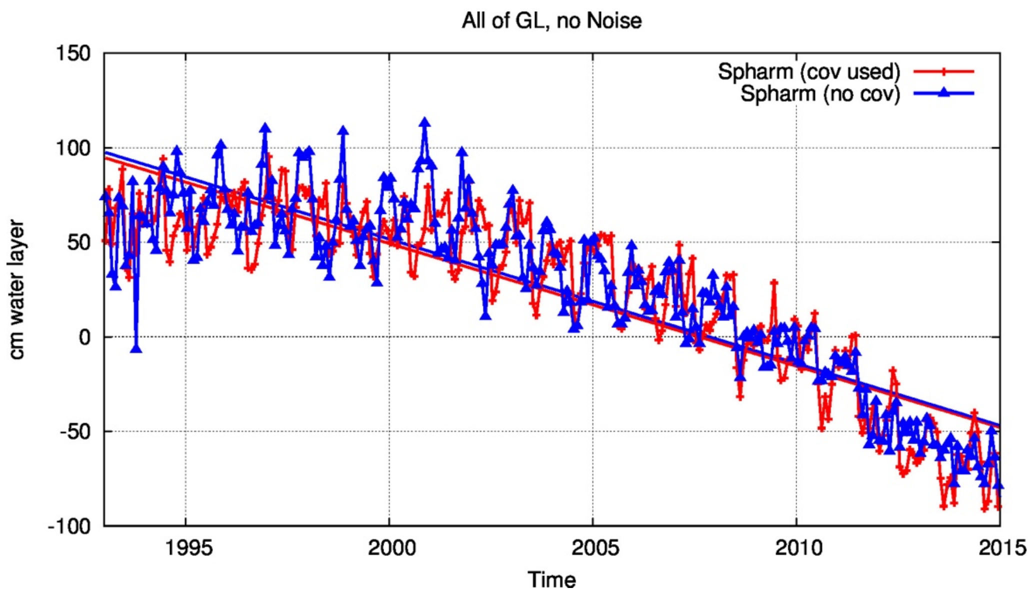
Yes, and done.

---

* Use of diagonal SLR and GRACE error-covariances , and thus neglecting off-diagonal error-covariance. I think this is the most serious issue I can find in the paper. Since I don't know whether this is going to have a large impact on the results I'm recommending a major revision to allow the authors to clarify this. I suspect that in particular SLR may have significant off-diagonal components in its error-covariances. The SLR network is very sparse and may not be optimal for the retrieval for ice mass change signals at higher latitudes. To account for this, one would in principle need to propagate the full SLR error-covariance on the 1x1 degree grid used as observations. The associated error-covariance matrix of the gridpoints will consequently be quite unstable (e.g. from 36 SLR 'observations' one produces 360x180 observations, without adding more information), which potentially could break down the inversion scheme as it is implemented now. In the current setup, the authors ignored error-covariances and by choosing an equidistant 1x1 grid also artificially increased the density of observations at higher latitudes inversely proportional to cosine(lat). In a broad sense, ignoring offdiagonal contributions and artificial increase of observations can be interpreted as a regularization, which the authors should justify. I therefore, propose that the authors either justify their choices for the 1x1 grid in combination with a diagonal error-covariance or better: that the authors replace matrix H (see eq S1) by an operator which directly maps Stokes coefficients to the unknown vector a. When full error-covariances are available these can then also be implemented with hopefully relatively little effort.

** "and thus heavily dependent on the same very low degree spherical harmonics"*
Maybe quantify this with formal error correlations?

I tested the impact of the Cheng 5x5 SLR covariance matrices a year or so ago, using the uncorrelated inversion method only (see example for Greenland below). The red line is the inversion with the covariance matrix included, while the blue line uses the identity matrix instead. As you can see, the subannual part changes – but that's also the least accurate part of the signal. The long-period and interannual signals remain effectively the same. Now, this was done with an older, simpler version of the code, without the correlations between sub-regions included. But I've no reason to believe that results using the modern code would look all that different. (I should add, by the way, that one reason the error covariances make little difference is probably because the Cheng SLR series uses a consistent number of satellites in it: always just the five. So the accuracy of the combined solution doesn't drastically change when a new satellite is added/removed, as is the case in the Talpe paper.)



I just talked to Matt Talpe. Neither he nor I can find any way to take these spherical harmonic covariances and propagating them onto a 1x1 degree grid. The only way to take the error covariance into account would be to switch to a spherical harmonic representation from the top-down. That's simple in concept, but involves a total rewrite of the correlated inversion computer code in practice. Moreover, spherical harmonics naturally force the use of global data, while the gridded technique also allows us to use the same code on limited regions of the world. For 5x5 harmonics like SLR, that isn't useful, but it is valuable for other purposes using 60x60 GRACE data. Given the lack of clear improvement in the inverted results above, we chose to stick with the easier-to-use system already in place.

As I mentioned to the other reviewers, Minkang Cheng has very recently created an SLR series using the exact same input data, but estimating over 6 months rather a single month. When I repeat the inversion process with that data, I find that the divergence after 2010 vanishes (proving that it really was just a numerical instability). Once we tidy up these new results, we will surely write another paper. I feel it would be more meaningful to incorporate the detailed error analyses you mention, in that future paper instead, which will (hopefully) contain a timeseries which we feel is stable and trust-worthy enough for others to use. Given that the conclusion of this current paper is basically "the monthly SLR data is not accurate enough to use for this purpose", extended error calculations seem noncritical, to me. I hope that a promise to look into the impact of the error covariances for the future will be sufficient for you and the editor.

Thank you very much for your assistance and helpful thoughts,
Jennifer Bonin