Response to Reviewer #2.

Thank you very much for your helpful review. We'll respond to each point in kind:

0. Another major item is: "Half the References section is missing!" (Perhaps a problem with the OS website?).

This error was noted by the first reviewer and has been corrected. We honestly have no idea how the second half of the references vanished, but it's been fixed. Oops. Sorry.

1. My major complaint is not with the GRACE processing, but with the altimetry – and more specifically it is with Figure 1, which reports a large "bias offset" between Jason-1 and Jason-2. This certainly will be a surprise to the altimeter community, and it contradicts what has been previously published – see, for example, papers by Ablain et al. (doi: 10.1080/01490419.2010.487805) and Beckley et al. (doi: 1.1080/01490419.2010.491029).

The Jason project teams and most users would be very concerned to see Figure 1 published as is, and for good reason. In fact, this "bias" is merely caused by use of inconsistent versions of Jason GDRs. One cannot blindly combine different GDRs, based on different corrections and possibly other things (retracking?), and expect consistency. The authors should not rely on the "experienced aid of Don Chambers," but should carefully examine user handbooks and other documentation. They will find that there are other differences, too, not just the MSS model. After I did some digging, I can add one thing in the authors' defense, which is a point about better data documentation. For some reason, the CLS group uses a naming (or non-naming) convention that is confusing. Their MSS evidently comes with a rate, and by changing the "reference time period" the MSS obtains different values, even though the fundamental MSS model is the "same" and retains the same name. (It is not a matter of the time span of data going into the determination of the MSS.) It would be much better if CLS didn't confuse users in this way, but that's the way matters stand. The GDR attributes give no hint of this problem, but the data handbooks do.

What the authors should have done, and should have written, is something like the following:

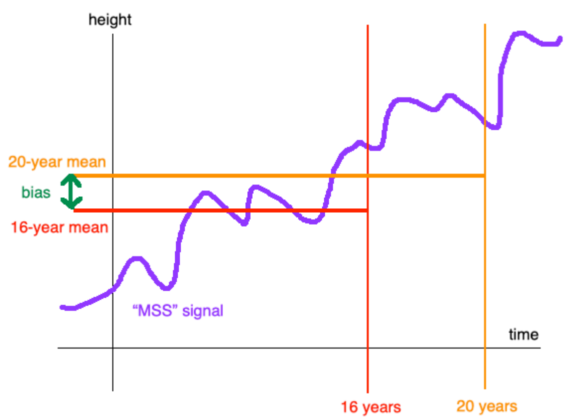
"We have used the best available Geophysical Data Records (GDRs) from Jason-1 and Jason-2, and applied consistent geophysical models to ensure a self-consistent time series of sea surface height anomalies across the missions. The source data are from Jason-1 version "E" and Jason-2 version "D" GDRs. Documentation for these different version numbers indicate the use of different processing standards, in particular ancillary geophysical models in the two sets of products. Most important for our investigation, we have used a consistent mean sea surface and ocean tide model. We have also used the ECMWF Reanalysis for the dry troposphere and inverse barometer corrections, as provided on the Jason-1 GDR-E, to mitigate any changes to the ECMWF operational analysis during our period of interest."

This does require some data processing. An alternative approach is not to use the GDRs at all, but instead use DUACS(Aviso) or MEASURES products, which are reprocessed data with consistent data handling since 1993.

We entirely agree as to what the problem is, and the ideal way to fix it. Unfortunately, as you say, the issue is not well-documented at all, even in the handbooks, much less on the AVISO website – which means we will not be the only people to run into this problem and be baffled by it. Anyone downloading modern GDR data will run into it, because the older versions of the data are no longer available. We cannot find a copy of the old MSS model anywhere, though the most recent MSS model data can be downloaded from the AVISO website after some contortions. Similarly, the GDR-D data for Jason1 simply isn't online anymore at either PODAAC or AVISO. GDR-E only exists for Jason2 and Jason3, not Jason1. Which makes using matching GDR versions or official MSS models effectively impossible.

We chose to make a point of this in this paper largely to make others aware of the potential problem. It's very easy to miss, because it's natural to assume that since the GDR versions out there are the only ones available, they can be strung together safely. People have read the papers you suggested, and are thus using codes which assume the mission-to-mission offset is spatially-uniform (as it otherwise would be). But with the jump between GDR-D and GDR-E, it's not.

The big problem is not that the SSH model changed between versions (that's expected), but that they didn't level it so that the (arbitrary) time-means were the same between versions. I hand-drew a cartoon of the issue below, in case that better helps explain what I'm describing. Even assuming the two MSS models were perfectly identical, because of the 16-year vs. 20-year time span, they're going to see different MEAN values. That's where the bias jump between missions (or rather, between MSS models) is coming from. It's way too big to simply be from real model improvements. They just didn't recenter the bias to the same timespan as the old MSS model, or provide a way for others to do so afterwards.



The resulting bias will result in incorrect results on the order of +/- 4cm heights in some areas, so it's important that others know that it needs to be handled. Ideally, of course, the producers of the GDRs would update all the Jason satellites together, or use MSS models with identical global means, to avoid this issue altogether. But they haven't, so GDR users have to deal with the issue on their own. Which they certainly can't do, if they aren't told there's an issue to begin with!

Now, as you say, the best method to handle this would be to reprocess with my own, consistent MSS model, not the differing versions inside the GDR files. Most likely we'll end up doing that for future work, if the Jason2 and 3 GDR products don't come out with a version E soon (as we keep hoping they will!). But that's going to take a lot of time and effort to code up from scratch and run, and the mean-bias correction we've already made corrects for the worst of the problem – the bias jump between missions – already. (The E-version MSS may also be more accurate on a point-by-point basis, but that just means a possible quality degradation between missions – and we already have a degradation over time with GRACE anyhow, so that's tolerable.)

As you say, DUACS is another option, but we would prefer to avoid using a premade gridded product. While that would fix this MSS issue, we have no idea how the optimal interpolation used in the combination of multiple satellites will alter the high-frequency data, particularly in areas with less good coverage. That seems to us a harder to handle problem than the change in MSS model (once the jump between missions is handled).

Because of the confusion both reviewers showed for this subject, we have lengthened and clarified this section in the text. Hopefully the details will make more sense now.

2. The reference "Eumetsat,... (2016) for Jason-1 products isn't right, as Eumetsat had nothing to do with Jason-1.
You're right. But actually, when altering this section of text, we realized that the Jason-1 handbook has not been correctly updated to list the current MSS information, so we instead pointed to the AVISO website, which we presume has the most up-to-date information.

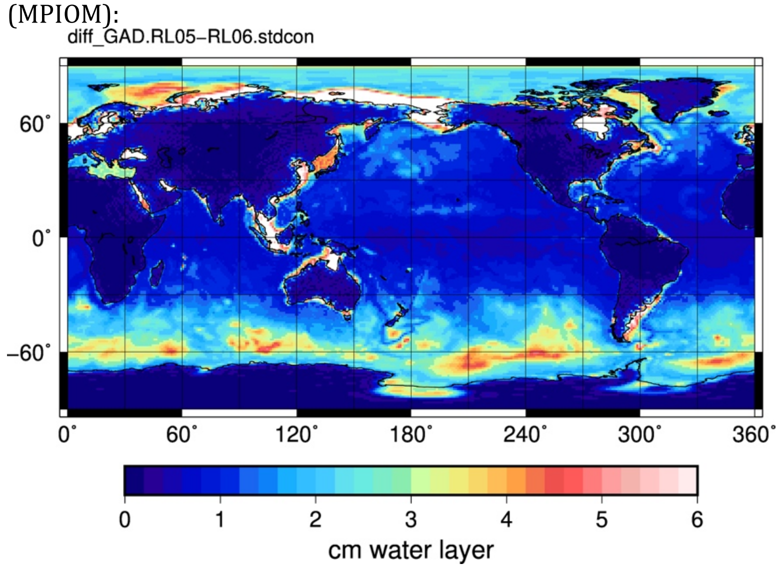
3. Line 21: "as large as" -> "even larger than"

Corrected

4. Line 23: How do you know the ocean models are poor in the Southern Ocean? If data assimilation has been used in their development, then I'd agree, but I thought OMCT and MPIOM had no assimilation. Is there another reason to think models are poor there?

It's true that neither OMCT nor MPIOM use data assimilation, but at the same time, their creators do test them (or maybe even "tune" them?) by checking against whatever data exists. So in areas where data is limited, it's hard for them to determine if the models go amiss in some way, and correct for it. Thanks to its depth and difficulty to get to, the southern ocean is one of those places which is not well observed (few bottom pressure recorders, limited XBT drops, etc). And thanks to its uneven topography and global zonal circulation, its physics is complex. All of which make errors more likely than elsewhere.

The easiest way for me to demonstrate the heightened uncertainty in this area is simply to show you the standard deviation of the submonthly differences between AOD1B RL05 (OMCT) and AOD1B RL06 (MPIOM):



Notice that some of the biggest submonthly differences – up to 5cm in equivalent water height – are in the Southern Ocean. The larger uncertainties in this area are confirmed by the AOD1B modelers (ie: Dobslaw et al 2017; doi:10.1093/gji/ggx302) who similarly see large differences in their models and improved GRACE KBRR residuals in the area. Other modelers recognize the area as similarly less well-known.

The usual assumption is that a newer model is generally better – and in fact, we've proved that in this paper, in terms of AOD1B. But, as we also showed here, there are remaining submonthly errors in the model. It's hard to say exactly how large those errors are (I couldn't find any good paper on it for any of the models used here), due to the paucity of observed data in the region, but it's telling that neither the OMCT nor MPIOM modelers were surprised to hear about potential weaknesses in their models in this area.

5. Line 27: "predicated" is the wrong word to use here.

Corrected

6. Line 163: "signal" -> "signal was"

Corrected

7. Line 170: The authors here might wish to cite published work that has examined the barotropic circulation in this region. For example, work by Chris Hughes: doi:10.1029/2006JC003679

Yes, thank you, done. That was, in fact, the paper that got me (Jennifer) interested in this region.

8. Line 210: Is MPIOM also forced by pressure? If not, how does this affect the C3 comparisons? Line 203 already notes that OMCT uses pressure forcing.

Yes, both models use comparable types of forcing, including pressure. We've altered the text to make this more explicit.

9. Line 214: The Lynch-Gray reference should be augmented (or even replaced by) Carrere et al.: doi:10.1029/2002GL016473

Done, thanks.

10. Caption to Figure 3. It would be useful to give the time intervals over which these standard deviations were computed. (In fact, I don't think I saw this in the main text anywhere either, but I may have missed it.)

Oops. You're right. That's now fixed, both in Fig 3 and in the main text. The span used was April 2002 to January 2016 (the periods when we had data for all the series). Sorry.

11. Line 247. I would add "except for the middle and North Atlantic". It seems GRACE is not improving the prior model there.

Agreed, corrected. It's interesting that there's such a neutral response in the Atlantic, actually.

12. Lines 280, 288: Are Figures 6a and 6b reversed?

Yes, thanks. Corrected.

13. Line 308: Could slightly more explanation be added here, or at least a reference? It is not obvious to me how Gaussian temporal windows are being used to form a band-pass filter.

Done. Basically, you can use two windows of different lengths to create a band-pass filter. For example, if you process the same series with first a 30-day boxcar sliding window and (separately) a 20-day boxcar, then you can difference the 20-day series from the 30-day one to get a bandpass filter between 20-30 days. It's the old game where windowing in the time domain is equivalent to filtering in the frequency domain. We're using Gaussian windows to avoid ringing due to sharp cutoffs, but the same principle applies. (Yes, we've checked this with an FFT in the past.) The benefit of using windows rather than frequency-based techniques is that it allows you to work with gappy data.

14. Line 345: I would again suggest that it is mentioned that the Middle and North Atlantic are problem areas.

Done.

15. Section 7. Since this section already lists long URL addresses for data used, those things could be eliminated in the main text.

Good idea. Done.

16. I much appreciate the color scales in (for example) Figure 5, where arrows point which way which model is superior. Very useful. Some of the figures are a bit hard to read, however, and a bit cramped. The fonts/resolution of Figure 2 seems especially fuzzy – a Word feature?

Yes, it annoys us, too, and yes, Word makes it worse than the file is on its own. Figure 2 is a combination of a line plot (made in gnuplot) and a map (made in GMT). We can't find a "pretty" way of joining the two, except doing so manually in a graphics program, which reduces the resolution – and forces us to save in PNG format (since the journal won't accept jpgs). Word, apparently, doesn't like PNG format. We've redone Figure 2 with a clearer-looking font to match the others, which makes it less obvious. The final plot should look better, since it won't have the issue with Word on top of the limited resolution.