Hydrology and
Earth System
Sciences
Open Access
Discussions

Interactive comment on "A global water cycle reanalysis (2003–2012) reconciling satellite gravimetry and altimetry observations with a hydrological model ensemble" *by* A. I. J. M. van Dijk et al.

Anonymous Referee #1

Received and published: 13 February 2014

The paper by van Dijk et al entitled "A global water cycle reanalysis (2003–2012) reconciling satellite gravimetry and altimetry observations with a hydrological model ensemble" attempts to assimilate satellite gravimetry data, with some additional in-situ and altimetry derived data sets, with model-based estimates of global total water storage. The different hydrological models and GRACE data sets are merged together using different strategies before being combined in a Kalman-filter like approach. Estimates from the different hydrological stores are then assessed, with some compared against external observations.

The study itself attempts to address the extremely ambitious problem of estimating global water storage variations, which involve everything from depleting continental groundwater stores to accelerating glaciers in the cryosphere to sea level change. The data sets simply don't exist to measure this with the same fidelity across the globe, and modeled estimates are too incomplete or inaccurate. For this reason, the current paper's attempt to combine the two in a data assimilation approach is relevant, but as I interpreted the methodology and results, the conclusions are not well supported by the work shown, and the uncertainties and assumptions made make the reader question their validity. Below is a summary of my concerns:

1) First, I would recommend changing the title to remove the word "reconciling". To me, reconciling implies the resolution of a long-standing difference between two or more camps of thought. This isn't what's done here, and reconciling is used more in the context of incorporating/combining/assimilating the two quantities (models & data).

2) I had a difficult time understanding the methodology used. The choice of variable annotation and terminology made it difficult to follow in places (e.g., a Gaussian smoother was termed an "observational model"; see detailed comments below). Many key aspects of the methodology were left up to the reader to explore in the literature (triple collocation, groundwater estimates, surface water use estimates, generation of nearly all satellite data sets and their uncertainties, generation of the hydrological models). To the readers, these critical items are like black boxes, that the reader would have to spend considerable extra time to understand. I realize that the authors can't replicate all of the work previously done, but I think more can be done to explain or visualize the data sets involved, and their general characteristics.

3) More specific to the methodology, I have concerns about the underlying premise behind the ensemble approach. Four variants of GLDAS were included, which all have similar underlying physics, in addition to an independent W3RA model. The GLDAS variants do not model deep soil or groundwater, so these values were "patched" in using groundwater depletion/recharge estimates from Wada et al (2012), which used the

PCRGLOBWB model. Adding the groundwater to the GLDAS models seems inconsistent, and guaranteed to generate model errors, since the physics of the two models are not linked in any way. Plus, this means there is only one real variant of the groundwater estimates. Why wasn't PCRGLOBWB used as a model variant? And my idea of a traditional ensemble approach is to vary the parameters within a single model, given the uncertainty of the parameters involved. What the authors do looks more like a (weighted) averaging of disparate model sets. What justification is there that this will generate a more accurate overall model? Why is just taking the average of a group of separate publicly available models at each time step the best approach? Same for the GRACE data sets? Where is it justified that averaging the results of a handful of GRACE solutions is optimal? In both cases, the results of the entire ensemble can be diminished by the inclusion of one or more bad models or data sets. If I have misinterpreted the methodology, then I would ask the authors to provide more explanation and/or derivations of the technique in the text.

4) The number of assumptions and adjustments that went into the analysis were numerous, and didn't really provide much confidence that the conclusions were reliable. One example is the triple collocation. Four important assumptions were listed, of which I thought only one was really satisfied. Another is that "Storage in water bodies without altimetry data was assumed negligible," although the altimetry only covered 62 lakes globally. Seemingly arbitrary adjustments were made that I felt impacted the interpretation of the results. Examples include the additional 5 mm error added to "correct for potential covariance in errors between the GRACE products...", as well as the -83 Gt/yr "adjustment" made to make the GRACE glacier mass estimates more in line with the Jacob et al results. Combine this with the extra +87 Gt/yr adjustment from new reservoir impoundments (that was first introduced in Sec 4.4, just before the conclusions), and it felt like the numbers used for the total water cycle estimates in Table 3 were not directly supported by the work presented in the paper, and in reality can have large volume/mass swings that meet or exceed the 0.39 mm/yr SLR discrepancy discussed in the conclusions.

<div align="center">C7880</div>

5) My last major concern involved the validation of the results. As I understand it, the results of the validation efforts were as follows:

vs regional storage trends: increased variability seen (could also be noise), along with amplified trends (again, could also be errors), and some dramatic trend changes (mainly in arctic, where models known to be poor).

vs river discharge: done, but comparisons inconclusive – only a handful of major rivers evaluated

vs SWE: done, but comparisons inconclusive

vs glacier mass balance: results similar to other solutions – not surprising, since the Tellus solutions are generated by the same co-authors (Wahr, etc.) behind the Gardner et al and Jacob et al works used for comparison.

vs groundwater: validation was not done.

Given this, it can be argued that the comparisons to the independent observations don't contribute much to the validation of the results.

I also had a number of more specific comments:

P15477L19: term offline used here, but defined later

P15480L08: As I suggest above, I interpret this as meaning that the groundwater store is modeled for all five models using the PCRGLOBWB model (with depletion rates from IGRAC)?

P15481L06: The streamflow editing criteria seemed odd – why not choose those records with values over the study timeframe (2002-2010)?

P15482L27: According to the Tellus website, the processing and filtering of the land and ocean products were different, e.g., the ocean products have 500km smoothing applied. Please comment.

P15482L28: not clear how assimilating the retrievals means you should not correct for leakage effects

P15483L3: should specify GIA model used; wording suggests the correction was not the same as that applied to the GRGS solutions

P15483L08: Do you mean long-term trend? The earthquake co-seismic response is essentially a step function, with post-seismic changes being non-linear, but occurring over many years. "Seasonal" signal to me implies semi-annual periodic signals.

P15483L26: why isn't the definition of w_l shown here, instead of later in Eqn 8?

P15484Eqn3: would recommend using a different super/subscript to distinguish this definition of s_t^b from that of Eqn 2

P15484L19: I find this terminology strange. An observation model in my mind represents a functional model that relates the observational data to the system dynamics and parameters. Here, it is used to describe a Gaussian smoother, which is a generic convolution operator that has no dependence on the observations or system dynamics.

P15484L23: the Gaussian filter used for most GRACE solutions in the literature (and I assume that for those on the Tellus site) is based on that described by Jekeli et al (1981), which has a slightly different "bell-curve" shape than a traditional Gaussian curve, since it is optimized for geodetic applications. It's not clear that you are smoothing your total storage estimates with the same filter kernel – this could change the comparison values, and hence your interpretation of the results.

P15485L09: read literally, L can only equal 5. L should also be in lower case to match that in the equation. Same for M.

P15485L14: Do the uncertainties vary significantly for the various GRACE solutions? Please comment.

P15485L17: The term "disaggregate" can have different meanings, so I would recom-

C7882

mend clarifying throughout the paper that you are spatially disaggregating the solutions

P15486L25: How do you transform model-derived storage into TWS as derived from GRACE? It is either derived from models, or derived from GRACE. Please reword.

P15487L06: To both the ocean and land products? As mentioned earlier, the ocean products already have 500km applied according to the Tellus website.

P15487L08: According to the GRGS website: "It is reminded to the users of the GRGS products that NO SMOOTHING OR FILTERING is necessary when using them, since they have already been stabilized during their generation process." The extra smoothing seems to violate this.

P15487L11: Is this correct? There are five land models, three Tellus solutions, and one GRGS solution. Where do the 15 GRGS solutions come from?

P15487L14: I can easily see the data sets violating assumptions 1-2 (maybe 3 as well). You would have no way of knowing whether the data sets are biased to each other, but you have no reason to assume they are not. We know GRACE errors vary in time, depending on time frame (< June 2003 vs > June 2003) or proximity to near-resonance orbits. Whether the error is time-correlated is debatable.

P15487L25: Not clear what this has to do with the discussion on the triple collocation assumptions. Please clarify.

P15488L01: The LAGEOS data they use only contributes to the C20 coefficient, nothing else (as stated on the GRGS website). While the retrieval methods is slightly different than the other centers, they still use the same background models (ocean tide, solid earth) and their static reference field incorporates the EIGEN (GFZ) mean field. Not sure what they do regarding aliasing, but I assume GRGS uses the same de-aliasing product as the other centers. This all suggests to me that the correlation might be stronger than suspected. Why can't the GRGS fields simply be lumped into the analysis with the other GRACE solutions?

C7883

P15491L24: Is this due to the extra smoothing applied, as well as the fact that the GRGS solutions themselves extend only to deg/ord 50? This extra smoothing/reduced resolution would diminish trends and variations.

P15496L03: I was also expecting this latitudinal dependency. The fact you did not see this makes me wonder whether some of the variability seen in the regional storage trends isn't partially due to this.

———————————————————

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 15475, 2013.