Hydrol. Earth Syst. Sci. Discuss., 10, C8117–C8123, 2014
www.hydrol-earth-syst-sci-discuss.net/10/C8117/2014/

Hydrology and Earth System Sciences
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.

A. I. J. M. van Dijk et al.

albert.vandijk@csiro.au

Received and published: 6 March 2014

We thank the referee for these valuable comments. The reviewer makes several specific comments that we will address in revising the m/s. Below we respond to the 5 main concerns raised (letters between brackets are sometimes added for cross-referencing where there are multiple sub-comments).

COMMENT 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).

RESPONSE 1) We referred to 'data reconciliation', a statistical term with a fairly specific meaning that describes what we did. We don't feel 'reconciling' is confusing but will await the editor's advice. 'Assimilating' would be a reasonable alternative.

COMMENT 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.

RESPONSE 2) We appreciate that the methodology used is fairly elaborate, and full replication of the experiment would probably require reading much of the cited literature. The complexity is further increased because a model ensemble and multiple observations were used, but that is how we were able to provide better constraints on the assimilation. Unfortunately complex methodologies have become an inevitable part of our research (consider for example the complexity of numerical weather model assimilation schemes). We did our best to describe all aspects with the detail needed and cited data and literature references, as well as providing a visual diagram illustrating the methodology. We value the specific suggestions the referee provides to improving the methodology description and will take them into account in revising the m/s.

COMMENT 3) More specific to the methodology, I have concerns about the underlying

premise behind the ensemble approach. (a) Four variants of GLDAS were included, which all have similar underlying physics, in addition to an independent W3RA model. (b) 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. (c) Plus, this means there is only one real variant of the groundwater estimates. (d) Why wasn't PCRGLOBWB used as a model variant? (e) 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. (f) 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? (g) 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.

RESPONSE 3) (a) We do not necessarily agree that the four GLDAS models all have similar physics but that may be a matter of definition. (b) Combining the Wada et al (2012) groundwater depletion estimates with the GLDAS models would be conceptually inconsistent if extractions from an unconfined aquifer were also incorrectly assumed to discharge as streamflow (i.e. the water would be counted twice). It would be easy to correct for this if we had information on whether extraction was from a discharging shallow aquifer or not, but we lacked this information. Fortunately, in practice, the error associated with this is likely to be small where (i) groundwater extraction is negligible compared to discharge, as is typical for humid regions, or (ii) groundwater discharge is negligible compared to extraction, which is typical for dry regions. (c) Correct, although with uncertainty estimates. We agree with the referee that ideally more global land sur-

face models would better represent groundwater dynamics and that ideally additional, independent estimates of global groundwater depletion would be available, and hopefully this will happen in future. (d) PCRGLOBWB was not used because estimates for the full assimilation period were not available. (e) Probably depends on one's frame of reference. Someone close to Ensemble Kalman Filter approaches might primarily think of an ensemble created by perturbing forcing. Here we use the term with reference to the literature on multi-model ensemble approaches. We will take advice from the editor - we could change the last words of the title to "multi-model ensemble" for example. (f) A simple ensemble average is justified if the errors in the individual estimates are dominated by noise of similar magnitude. In this case we could not be sure that the error magnitude was indeed similar and hence took a different approach, characterising the error in each of the ensemble members (for models as well as GRACE products) using the triple collocation approach, and incorporating those errors in the assimilation scheme. A manuscript has just been published (Ensemble prediction and intercomparison analysis of GRACE time-variable gravity field models, C. Sakumura, S. Bettadpur and S. Bruinsma, GRL, DOI: 10.1002/2013GL058632) that demonstrates that the different GRACE retrievals used here indeed do have independent noise, which provides some additional justification for our approach (g) Correct. However we used the member-specific error estimates. Therefore, wherever a member is particularly 'bad' (i.e. has a comparatively large error) it will exert correspondingly less influence on the assimilation result.

COMMENT 4) (a) 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. (b) Another is that Storage in water bodies without altimetry data was assumed negligible, although the altimetry only covered 62 lakes globally. (c) 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...,

(d) 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.

RESPONSE 4) (a) Characterising errors is inherently difficult and uncertain, but the strength of a formal data assimilation approach is that it demands error estimates and so exposes all assumptions, producing assimilation results with quantified uncertainty. We intended to document, motivate and discuss each assumption we needed to make with some care. For example we do discuss which of the triple collocation assumptions are more or less likely to affect the analysis. Where improvements on the methods were currently not yet possible an opportunity for future research is indicated. (b) This was an inevitable caveat given limits on the observations available (however see further below). In the discussion we address the possible role of surface water and an opportunity to make better use of satellite observations in future. (c) The 5 mm was not quite arbitrary and argued the case. Fundamentally, we wanted to make a conservative assumption. We note that the influence of the added error on the calculated gain matrix was actually small because the model and GRACE errors are generally of quite different magnitude (see Fig 2). (d) The referee is correct that these numbers were not derived directly from the data assimilation, which is why they are raised in the interpretation and discussion. We were not able to use these underlying data as a priori constraint, and so needed to leave them as uncertainties for the discussion with best effort post hoc adjustments. However, for the case of the reservoir impoundment effects; since this m/s was published we have developed a way to ingest prior estimates into the assimilation scheme, and intend to revise the m/s accordingly. This is not possible for the glacier adjustments, and we identify this as an area of future research.

COMMENT 5) My last major concern involved the validation of the results. As I understand it, the results of the validation efforts were as follows: (a) 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). (b) vs river discharge: done, but comparisons inconclusive – only a handful of major rivers evaluated (c) vs SWE: done, but comparisons inconclusive (d) 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. (e) vs groundwater: validation was not done. (f) Given this, it can be argued that the comparisons to the independent observations don't contribute much to the validation of the results.

RESPONSE 5) (a) The interpretation of regional storage trends was to confirm that the assimilation scheme behaved as intended, and the patterns are of interest in their own right. However this was not part of the validation. (b) 450 river basins in addition to 445 river altimetry sites is perhaps more than a handful and the results were not inconclusive: there were improvements in some regions and degradation in others. An improvement across the board would have been great but was not to be expected – however it is encouraging that there were some strong agreements for large rivers with a strong bearing on the GRACE signal, such as the Amazon system. (c) A similar answer: the results were not inconclusive and improvements everywhere were not expected, but importantly agreement improved in several regions where there are large snowpack variations. (d) For several glaciers independent observations were used, and therefore in Section 3.8 and Table 5 we distinguish those from the glaciers for which estimates are GRACE-derived and therefore not independent. (e) Correct, unfortunately there are no suitable groundwater observations that would allow validation, and in any case that would be conceptually different from sub-surface (ground + soil water) storage. (f) So in summary, we disagree that the validation was inconclusive or did not contribute much. Of course had more observations been available we could and would have used these in either assimilation or evaluation as well.

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