Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-269-AC2, 2016

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Evaluating the hydrological consistency of satellite based water cycle components" *by* O. López et al.

O. López et al.

oliver.lopez@kaust.edu.sa

Received and published: 26 July 2016

Author Introduction. Thank you for your positive comments and thoughtful review. In addressing these, we have itemized the various comments in order of their appearance in the letter.

Comment 1. This paper describes an evaluation method to explore correlation of satellite-derived water budget components P-ET and Terrestrial Water Storage. It uses a spherical harmonic analysis to analyse correlation and differences between multiple time series of P-ET and GRACE Terrestrial Water Storage. The method results cannot explain differences of the analyses in the three large-scale basin studies. This study uses a novel approach to estimate differences between two time series. The fact that this analyses does not lead (yet) to valuable results, is no reason to disqualify it in any

way. Therefore, it makes this paper a valid and potentially useful addition to the journal. However, the descriptions and approach are, in my opinion, not fully crystallised and require more research. Some of the descriptions are also rather confusing and need a better structuring (see further comments below). I get the feeling that recommendations described in this work should have actually be part of the paper.

Author response. The overriding objective of this work is to determine whether hydrological consistency can be achieved using independent remote sensing data over regions that exhibit "idealized" conditions (i.e. where hydrological processes are reasonably well defined). The fact that a high degree of consistency was not observed is an important result, as it highlights the challenge and disparity of these remote observations. The reason that we explored this approach was a means to independently assess a selection of evaporation models. Assessment of such large scale products is challenging, and new approaches are required to provide a more holistic evaluation strategy. This point may not have been well articulated, so we will re-examine the descriptions and rationale behind the approach to ensure greater clarity of purpose.

Comment 2. The whole term 'hydrological consistency', which is everywhere in the paper, e.g. the title, is not very clearly explained. Could you consider a better way to describe what you want to research? For example, 'the ability to balance the water budget' or something containing the words 'hydrological closure'?

Author response. This point is well taken. Remote sensing offers a number of independent means with which to retrieve various components of the hydrological cycle (i.e. rainfall, soil moisture, evaporation). Ideally, these observations should be hydrologically consistent: that is, an observed rainfall event should cause a corresponding change in soil moisture, for instance. Likewise, a reduction in soil moisture should be reflected by an increased flux of evaporation. Consistency is just another term that encompasses the expectation of a water budget: changes in one term should be reflected in others. While this has been explored qualitatively in the past (McCabe et al. 2008), here we wanted to determine if the method (using spherical harmonics) could reveal some level of agreement between spatial (and temporal) patterns of these independent hydrological variables. We will ensure that this term is better defined so that the reader understands the intent (and limitations) of the research.

References

McCabe, M., Wood, E., Wójcik, R., Pan, M., Sheffield, J., Gao, H. and Su, H.: Hydrological consistency using multi-sensor remote sensing data for water and energy cycle studies, Remote Sensing of Environment, 112(2), 430-444, doi:http://dx.doi.org/10.1016/j.rse.2007.03.027, 2008.

Comment 3. In terms of concise descriptions: in my opinion, this study knows too many research questions and too little answers. It needs serious work on structure, correct descriptions and conciseness. The authors describe several paragraphs saying: 'the objective of this work', or 'a secondary objective of this work was..', or 'one key motivation of the study..'. Or they reword their aims in questions, such as: 'how accurate are the hydrological components derived from satellite observations?' or 'is hydrological consistency achieved with a particular product....?, etc. To me, that makes it confusing. Can you please clearly state the objectives and motivation at first, and then relate back to exactly those? That would make this work better readable. Furthermore, the many different aims and questions also makes you wonder about the conciseness and correctness of this study. For example, the questions "how accurate are the hydrological components derived from satellite observations?" cannot be answered with the results of this study. If all these research objectives and questions would be compiled into only two research questions, I think the paper would be more understandable.

Author response. We will work on restructuring the paper so that the main purpose and objectives can be identified more easily. As mentioned above, the goal is to evaluate the "hydrological consistency" method in a close-to-ideal scenario, keeping in mind that the desired potential use of the approach was to evaluate different evaporation products. How to evaluate large scale remote sensing products is an important ques-

tion that has some philosophical, as well as analytical, challenges. However, it may be better to ensure that these are contained in the discussion section rather than the introduction.

Comment 4. It is unclear what the added value of the spherical component analysis is. After all, correlations between water budget components can also be made in a different way, and the spherical component analyses does not lead to any new insights.

Author response. Undertaking the analysis in spherical harmonic space does seem like an unusual approach: but this is part of the novelty of the work. The reason for doing this was to ensure that a fair comparison between GRACE data and satellite products could be undertaken. Since GRACE data is filtered in spherical harmonics (unlike more traditional remote sensing variables in hydrology), a comparison between this and the other spatial maps of hydrological retrievals is likely to be imprecise (see Tapley et al., 2004 and supporting online material). Ensuring a consistency in spatial comparisons is one of the main reasons for doing the analysis in this way. While scaling the GRACE signal to account for differences has been proposed as an alternative solution to this problem (Landerer and Swenson, 2012), it has recently been shown to affect results (Long et al., 2015). By removing the impact (and model dependence) of this scaling term on the GRACE data, a more reasonable intercomparison of hydrological variables can be undertaken.

References

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F. and Watkins, M. M.: GRACE measurements of mass variability in the earth system, Science, 305(5683), 503-505, doi:10.1126/science.1099192, 2004. Landerer, F. W. and Swenson, S. C.: Accuracy of scaled GRACE terrestrial water storage estimates, Water Resources Research, 48(4), doi:10.1029/2011WR011453, 2012. Long, D., Longuevergne, L. and Scanlon, B. R.: Global analysis of approaches for deriving total water storage changes from GRACE satellites, Water Resources Research, doi:10.1002/2014WR016853, 2015.

Comment 5. Moreover, the lag between some data make the results of the analysis method less obvious. The method does not lead to any quantifications of if the water budget is in balance, or by how much it is off (for example, in percentages of the total water budget). That makes this study for me hard to judge: it evaluates water budgets [but] this does not lead to any new insights (unless you describe that part better), and cannot answer obvious explanations for imbalance of a water budget.

Author response. It is true that at this exploratory stage of investigation, the approach does not provide a quantified metric of water budget closure, since it is difficult to relate the correlation between the two sets of spherical harmonics into a measurable quantification of the water budget imbalance. However (and this relates a little to the philosophical aspect of product evaluation mentioned above), to attempt to do this by using observations alone (i.e. not involving a hydrological model) requires that the individual products are themselves well quantified (or "validated"). The reality is that at the large scales studied here (and even at much smaller scales), they are not. Large scale retrievals of evaporation, soil moisture and rainfall products do not come with well-defined accuracy metrics, let alone uncertainty bounds. The question of how to evaluate such datasets remains an outstanding one – and one that requires examining a range of approaches. Determining whether these individual products are at the least consistent with each other (i.e. they reflect hydrological expectation) is a needed first step in product assessment. That is essentially what we attempt to do here: and find (perhaps not surprisingly) that we are not able to do this yet, even in relatively simple systems. We believe that this is an important, if under-appreciated, insight that provokes a need for both improved products and evaluation strategies.

Comment 6. It is unclear why the focus is put on the ET component. You should pitch that better. After all, any uncertainty of the P component would result in larger uncertainty. I think the discussion and conclusion of this study need to point out that comparisons in catchment study need to be undertaken using the regional information on hydro(geo)logy and ground estimates of P, ET and streamflow. My guess is that you

want to say that, despite being the second-largest component of the water budget, ET is most uncertain?

Author response. The focus on the ET component comes from a desire to evaluate some recently developed global satellite evaporation products (see McCabe et al. 2016). Given the spatial mismatch between ground observations (and the lack of continuous large-scale coverage of in-situ data in remote regions), it is difficult (moreover, inappropriate) to validate these large-scale products in such a traditional manner: hence the comment on product evaluation being an outstanding problem in hydrology mentioned above. In the recent literature, inter-comparison between evaporation products has mostly been done by providing estimates of the uncertainty in terms of the variance among the products, sometimes including other types of datasets as well (from land surface models and climate reanalysis) [Jimenez et al., 2011; Mueller et al., 2011; Long et al., 2014]. While this is a good first order approach, it also recognises the challenge and lack of a benchmark evaluation set. Furthermore, rather than comparing the uncertainties between the evaporation products and the other hydrological components (which are poorly defined), we attempted to distinguish between the different evaporation products relative to their consistency with precipitation and storage. That is, are observed changes or patterns in the evaporation datasets reflected in these other hydrological variables. We explore this approach precisely because of the challenges in quantifying uncertainty based upon traditional in-situ methods.

In terms of exploring a range of other datasets, the GPCP product was chosen due to its global coverage as well as being widely used in the literature. However, as the reviewer notes, other precipitation products could be included to examine their impact. We have recently investigated this using the PERSIANN global data set (Hsu et al. 1997), but have not seen any significant change that would alter the conclusions of the study (see these early results and a preliminary Figure S1 below, that can be included as supplementary material). Further data sets can be considered if required, but ideally we would prefer that the focus remains on determining variability in the evaporation products rather than including a greater number of variables and complicating the analysis and interpretation of results.

References

Hsu, K.-l., Gao, X., Sorooshian, S. and Gupta, H. V.: Precipitation estimation from remotely sensed information using artificial neural networks, Journal of Applied Meteorology, 36(9), 1176-1190, 1997. Jimenez, C., Prigent, C., Mueller, B., Seneviratne, S. I., McCabe, M. F., Wood, E. F., Rossow, W. B., Balsamo, G., Betts, A. K., Dirmeyer, P. A., Fisher, J. B., Jung, M., Kanamitsu, M., Reichle, R. H., Reichstein, M., Rodell, M., Sheffield, J., Tu, K. and Wang, K.: Global intercomparison of 12 land surface heat flux estimates, Journal of Geophysical Research: Atmospheres, 116(D2), doi:10.1029/2010JD014545, 2011. Long, D., Longuevergne, L. and Scanlon, B. R.: Uncertainty in evapotranspiration from land surface modeling, remote sensing, and GRACE satellites, Water Resources Research, doi:10.1002/2013WR014581, 2014. McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D. G., Michel, D. and Wood, E. F.: The GEWEX landFlux project: Evaluation of model evaporation using tower-based and globally gridded forcing data, Geoscientific Model Development, 9(1), 283-305, doi:10.5194/gmd-9-283-2016, 2016. Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F. and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophysical Research Letters, 38(6), doi:10.1029/2010GL046230, 2011.

Comment 7. The word groundwater is mentioned in the description of GRACE data. However, it is surprising that the word groundwater is not mentioned when discussing the phase lag to GRACE data, nor the separation of the P-ET(=Q) into streamflow and groundwater flow. One could for example compare global data of baseflow (BFI) and look if these compare to the differences in lag. See for example the wonderful work of Beck et al. (2013): Beck, H. E., A. I. J. M. van Dijk, D. G. Miralles, R. A. M. de Jeu, L.

A. Bruijnzeel, T. R. McVicar, and J. Schellekens (2013), Global patterns in base flow index and reces- sion based on streamflow observations from 3394 catchments, Water Resour. Res., 49, 7843–7863, doi:10.1002/2013WR013918.

Author response. We agree that this is a very nice study and appreciate the reviewer drawing this paper to our attention. After examining the paper, it may prove useful in aiding the interpretation of the physical aspects of incorporating a lag term into this analysis i.e. comparing the BFI of the four regions of study (and streamflow partitioning), against the observed lag seen in our method to determine any similarity (or differences). Even though the BFI from this study is a static variable in time, we agree that it will likely add value to our discussion.

Comment 8. It is also surprising that nothing is said on snow storage in the discussion on P-ET.

Author response. Snowmelt resulting from snow storage in two of the study regions was recognized as one of the complicating elements in our study. Unfortunately, the type of basin that would yield the most useful results to our study is not, in practise, easy to find e.g. one that is large enough for GRACE to detect the change in mass, that has sufficient variability in water storage (a requirement for GRACE), while having a small or negligible runoff component, no snow component, and strong precipitation and evaporation fluxes! The most limiting factor in using GRACE data in the study is the size requirements for the studied catchment. Because of their large extent and geographical features, the Colorado River and Aral Sea basins do include regions where snow storage plays an important role. Snow storage itself is not a problem, because GRACE detects changes in storage irrespective of their nature (snow, groundwater, soil moisture, etc). However, snowmelt may contribute to delayed changes in storage that can affect the results. The influence that snowmelt, as well as other potential sources of lag in the system is not known, and forms part of reason to explore the inclusion of a lag in the GRACE data.

Comment 9. Furthermore, the description of input data is not clear. Especially the ET part should contain what part of the data we use. For example, do we use PET from MOD16 or AET? Also, MOD16 contains an unclear description that first suggests it is a Penman- Monteith method, and then suggests it is Priestley-Taylor. Furthermore, the spatial detail in Table 1 on MOD16 is not correct to my knowledge and resolution should be 1km (at least if the data from Mu et al from http://www.ntsg.umt.edu/project/mod16 is used). See more detailed text comments below.

Author response. In this work, actual evapotranspiration (AET) was used for all evaporation products: this should have been more clearly articulated in the text. The particular sentence the reviewer refers to (Page 5, line 10) will be rephrased in a revised version to clearly indicate that Priestley-Taylor is only used for plant transpiration. In terms of the resolution, MOD16 (as found in http://www.ntsg.umt.edu/project/mod16) is available as a 1 km resolution product in a sinusoidal projection. In this study, the 1 km sinusoidal product was reprojected to a regular 0.05 degree-resolution grid using the MODIS Reprojection Tool (MRT). While a 0.05 degree resolution monthly product was available, we summed the original product to get monthly estimates in order to be time-consistent with the GRACE CSR product months.

We appreciate the potential confusion, so will improve the explanation in a revised version.

Comment 10. Methods and results are in my opinion too intertwined, and should be separated more. For example, in the results, there are some descriptions of the basins. For example, descriptions of the CRB, ASB and LEB basins contain texts that should in my opinion be put into 2.5 Study Regions.

Author response. We will review the manuscript and where needed, better separate background methods from results. Some of the description of the basins found in the results relate to the observed trends in water storage and precipitation observed during the study period. This was done in order to relate the changes in degree correlation to

the changes during these trends.

Comment 11. I am not an expert in spherical harmonic analyses, so I cannot judge on the correctness of the method. However, I think it would make the text a bit clearer if the term 'degree' (l) would be explained a bit more in detail, so non-experts in spherical harmonics would also understand. In that way, your plots 4-7 would be easier to understand.

Author response. We acknowledge that this is an abstract concept and we will work on the phrasing in the methodology to better explain it to the broad readership of HESS.

Comment 12. A weakness of this paper is that it does not incorporate uncertainty. After all, estimates of ET using ground observations already contain quite a bit of uncertainty (see e.g. Westerhoff, 2015, with work on Uncertainty of Penman and Penman-Monteith estimates). Another weakness of this paper is that it is not clear what specific ET has been used (PET, AET, P-T, P-M). A quick fix could be a better description of the input data. However, a discussion of different methods used (e.g. P-T or P-M, or PET or AET etc) should be clear on pointing out differences in methods. It would be unfair to judge that satellite data is not hydrologically consistent, if the ground-estimates already are so uncertain and if the different methods have caused this uncertainty.

Author response. As mentioned above, actual evaporation from all three evaporation products is used. We fully appreciate the comment regarding the uncertainty of both satellite and ground based measurements of evaporation. Indeed, our group have published a number of papers on this precise topic. Given the already extensive literature on this, we are not convinced that further repetition (or additional model descriptions) will add much to the paper. We will certainly attempt to clarify the precise nature of the products used by ensuring a thorough review of the input descriptions and can add detail to data and model descriptions where required.

The quality issue of ground based observations that you correctly raise is one reason why alternative approaches to product evaluation are required. The quality (or otherwise) of the satellite product should not be judged (solely) on its agreement with unrepresentative point scale approaches. It should also be judged on whether it can reflect hydrological expectation, as observed in independently observed and hydrologically linked variables (e.g. rainfall, soil moisture or groundwater storage). It is this approach that we have adopted to explore here. Certainly a more holistic evaluation strategy is required to ensure greater confidence in large scale products.

Comment 13. It is also unfair to put 'the blame' on ET, because P is a much more likely candidate for this uncertainty. You need to describe better why ET is so important in this analyses. My guess is that you want to say that, despite being the second-largest component of the water budget, ET is most uncertain. Can you derive uncertainties of all estimates? This study would be much stronger if it could quantify what absolute values of inconsistencies we are talking about.

Author response. As mentioned earlier, the focus on evaporation in this study was based on a need for an alternative way to evaluate satellite evaporation products (since in-situ observations are not the best way forward). Therefore, only a single precipitation product was explored: albeit one that has been well studied and used in global analysis. But this is also true for the water storage variations: although there are a number of water storage products based on GRACE satellite data (Bruinsma et al., 2010; Liu et al., 2010; Rowlands et al. 2005), the choice of a particular product (CSR) was based on its widespread use in a number of previous studies. Having said this, other precipitation products can certainly be considered and can be explored in a revised version (see earlier comment and preliminary results with PERSIANN).

References

Bruinsma, S., Lemoine, J.-M., Biancale, R. and Valès, N.: CNES/GRGS 10-day gravity field models (release 2) and their evaluation, Advances in Space Research, 45(4), 587-601, doi:http://dx.doi.org/10.1016/j.asr.2009.10.012, 2010. Liu, X., Ditmar, P., Siemes, C., Slobbe, D. C., Revtova, E., Klees, R., Riva, R. and Zhao, Q.: DEOS mass transport

model (DMT-1) based on GRACE satellite data: Methodology and validation, Geophysical Journal International, 181(2), 769-788, doi:10.1111/j.1365-246X.2010.04533.x, 2010. Rowlands, D. D., Luthcke, S. B., Klosko, S. M., Lemoine, F. G. R., Chinn, D. S., McCarthy, J. J., Cox, C. M. and Anderson, O. B.: Resolving mass flux at high spatial and temporal resolution using GRACE intersatellite measurements, Geophysical Research Letters, 32(4), doi:10.1029/2004GL021908, 2005.

Comment 14. If the document would be structured better, I think it would improve the quality and probably I would understand better what the exact goal is of this paper. Reducing the number of research questions/objectives would probably clarify why this paper should be published. Therefore, I recommend a major revision.

Author response. We appreciate your thoughtful and constructive comments. We will make sure to structure the text throughout the document so that the main goal of the study can be understood clearly, as well as better articulating the results and novelty of this work.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-269, 2016.

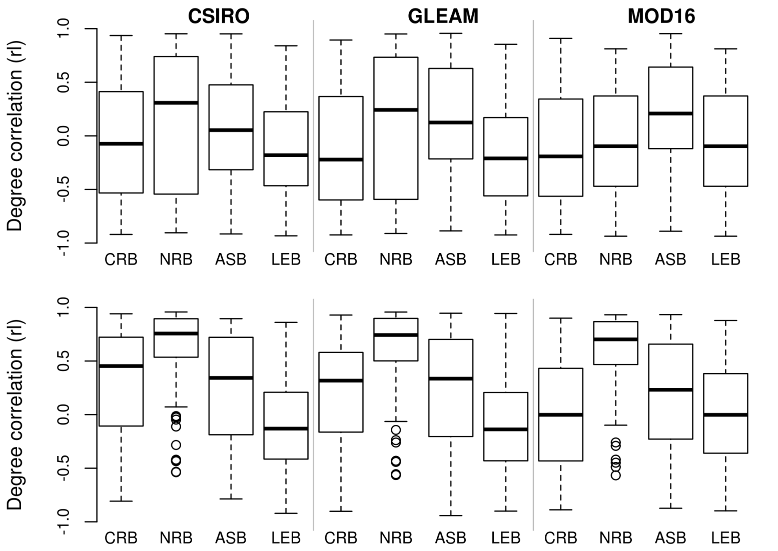

Figure S1. Top: average degree correlation statistics per study region and evaporation product. Bottom: GRACE data were shifted by two months to match the phase with P-E anomalies. The boxplots show the first, second (median) and third quartiles. Outliers, defined as data outside the 1.5 inter-quartile range (IQR) whiskers below or above the first and third quartiles are shown as circles. This figure represents a summary of the analysis using the PERSIANN product as precipitation. **The results are very similar to those in Figure 8.**

Fig. 1. Supplementary Figure 1