Integrating multiple satellite observations into a coherent dataset to monitor the full water cycle – Application to the Mediterranean region

Reviewer 2

MAJOR COMMENTS

• Overall, I found this to be a strong and well-written paper. It makes two (worthwhile) contributions. First, a methodological contribution regarding the appropriate approach for simultaneously enforcing water closure within land, atmosphere and ocean domains. Second, it advances the state-of-the-art in terms of water balance estimates for the Mediterranean Basin. These contributions are significant and worthy of publication in HESS. Nevertheless, there are three major points that the authors should address before final publication. I suspect that some of my concerns arise from misunderstandings on my part (rather than actual flaws) and can be addressed via re-writing for improved clarity.

- Thank you for your valuable comments, we hope that the new version of the manuscript is now in a better shape.

The paper needs to do a better job of describing the INTegration (INT) methodology and its impact on subsequent stock and flux predictions. My understanding is that the INT approach is applied with two aims: 1) to downscale sub-basin scale results down to the pixel scale and 2) to extrapolate (balance-constrained) results OUTSIDE of the Mediterranean. This raises two questions:

• If INT down-scales and extrapolates outside of the Mediterranean Basin , why does it have any impact on sea-level results in Figure 3 (which presumably reflect spatially averaged conditions within the Basin...neither of which are impacted by INT)? To me,

it seems like SW+PF and INT should yield the same results for a sea-level metric. However, these results are used specifically to motivate the added value of INT (line 15 of page 17). Is the improvement in INT versus SW attributable to INT? Or would it also occur for SW+PF?

- You are right, the text is not clear enough in this subsection. The altimeter measures are available only for the Mediterranean Sea, so when the closure constraint is applied over the two seas at once, only the Mediterranean Sea (not the black sea) level is used for the evaluation. You are right, INT and SW+PF would give the same sea level estimates if the Black sea was considered within the Mediterranean Sea in Figure 3. There is no inter/extrapolation of the closure for the Mediterranean Sea and the improvement in INT versus SW is attributable to SW+PF, however the representation of the closure impact of the two seas in the Mediterranean Sea is attributable to INT since SW+PF represents the spatial average over the Mediterranean within the Black Sea. We added additional comments in the associated section: *no inter/extrapolation have been used for the Mediterranean Sea sub-basin (Mediterranean Sea plus the Black Sea) and the improvement of INT versus SW is due only to the closure constraint of the PF. But the representation of the SW+PF estimate cannot be obtained for the Mediterranean Sea only since SW+PF close the water cycle over the Mediterranean within the Black Sea (no information about the Bosporus netflow).*

- **What exactly is the rational for the extrapolation portion of INT? Why would you ever want to extrapolate?**

- This has been a long discussion between the co-authors. The rational is twofold: (1) The extrapolation of a closure constraint is interesting at the technical level because for other regions, or when working at the global scale, some form of inter/extrapolation is required (See for instance: Munier and Aires, A new global method of satellite dataset merging and quality characterization constrained by the terrestrial water cycle budget, RSE 205, 119-130, 2018). (2) The extrapolation outside of the Mediterranean region also allow us to use more in situ observation for the evaluation, and this helps better testing the generalisation ability of our extrapolation scheme. Another minor justification is that users often prefer to use a simpler dataset with a rectangular domain, especially in the modelling community.
The justification of this interpolation is based on the assumption that most of the water cycle imbalance is due to satellite errors (this assumption is used for the CAL methodology too). The closure constrain is supposed to

improve the satellite estimate by reducing the bias and random errors. If no other information is used (such as surface type, see Munier and Aires 2018), the EO errors should have a spatial continuity and it then makes sense to extrapolate results based on this spatial continuity. We added this discussions in section 3.5: *The extrapolation of a closure constraint is interesting at the technical level because for other regions, or when working at the global scale, some form of inter/extrapolation between the monitored sub-basins is required (Munier and Aires, 2018). The extrapolation outside of the Mediterranean region will also allow us to use more in situ observation for the evaluation, this will help better testing the generalisation ability of our extrapolation scheme. The justification of this inter/extrapolation is based on the assumption that most of the water cycle imbalance is due to satellite errors (this assumption is used for the CAL methodology too). The closure constrain is supposed to improve the satellite estimate by reducing the bias and random errors. If no other information is used (such as surface type, see (Munier and Aires 2018)), the EO errors should have a spatial continuity and it then makes sense to extrapolate results based on this spatial continuity.*

• **Why not just apply terrestrial closure (at a minimum) to Northern Europe SW results?**
- You are right, Northern Europe is better monitored and river discharge could had been used to constrain Northern Europe basins at the SW+PF stage. As mentioned earlier, we prefer here to perform the main analysis over the Mediterranean basin and then test the extrapolation scheme over well monitored locations. This is now clearer in the text

• **Also, how does this extrapolation contribute to the (bottom-line) analysis in Figure 6?**
- The extrapolation does not contribute to the analysis in Figure 6 since only the Mediterranean basins are considered for computing the annual fluxes. This is now clearer in the text, Section 5.1: *The water cycle is analysed over its natural sub-basin's boundaries.*

• **I presume it facilitates the application of a larger atmospheric water balance analysis, but , given that this is a Mediterranean Basin analysis - it seems strange to extrapolate BEYOND the Mediterranean Basin. It would improve the manuscript if this extrapolation step was better motivated. -** We understand your concerns. We hope that our motivation in the extrapolation is now better explained.

I feel like the paper could do a better job describing its approach to error estimation (and the effect of these estimates on its merging results).
- There are two error estimations in our paper: *a priori* EO uncertainty assumption, before the merging, and the *a posteriori* uncertainties estimated after the merging. we hope we will not be confused in the following.

- **I was confused by the treatment of EO uncertainty throughout the manuscript. First, in Line 8 of page 8, the manuscript says states "...we considered the same uncertainty for all data sets of given parameters following de-biasing..." However, later in Section 3.3, it seems as if a different uncertainty is assigned to various precipitation estimates when applying Equation (5). Can these descriptions be made more consistent?**

- Sorry for this ambiguity, Eq. (5) gives the general formula, with different uncertainties, but we considered, you are right, same uncertainties in this application over the Mediterranean basin. This is now clearer in section 3.3: *Since no specific uncertainty estimates were available in the literature for the Mediterranean basin, the uncertainties are assumed to be identical (i.e. same σ_i) in the following.*

- **in Section 3.4, the authors invoke a filter-based closure constraint that varies as a function of a Σ matrix but do not describe how this matrix was derived. If would be helpful if this was clarified.**

- In our approach, we decided to close the water budget with a relaxation term: we assume an uncertainty in the closure. Such a relaxation on a constrain is commonly used in optimization, it generally follows a Gaussian distribution centred with a standard deviation Σ chosen *a priori* in a had hoc way. The matrix Σ must include the uncertainty for the continental, oceanic and atmospheric water cycles, it is not provided explicitly in Eq. (2):

$$\Sigma = \begin{pmatrix} \sigma_l & \mathbf{0} \\ \mathbf{0} & \sigma_o \end{pmatrix}$$

where $\sigma_l = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ represents the standard deviation of the constrained terrestrial and atmospheric water budget residual over land; and $\sigma_o = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$ represents the standard deviations of the constrained oceanic and atmospheric water budget residual over sea. Σ assumes no correlation in the imbalance of the 3 water cycles at monthly and annual scales, at sub-basin or entire basin scales.

• **Equation (5) appears to use the temporal standard deviation of individual products (precipitation products in the example given) after seasonal bias correction as a proxy for the magnitude of their random error. This seems like a dangerous assumption. Assume, for example, that you had a precipitation product that simply mimicked the TMPA seasonal climatology (used here as the de-biasing reference). Given large interannual variability in rainfall, this product would be a poor rainfall product to use in a water balance context. However, it would have a low temporal standard deviation, and (therefore) be heavily weighted by Equation (5).**
- Sorry for the ambiguity. Eq. (5) represents the uncertainty of the EO products, not the temporal standard deviation. This is now clearer in the text: "*let us consider the p precipitation observations P_i associated with Gaussian errors $\epsilon_i \sim \mathcal{N}(O, \sigma_i)$* "

• **Also - on a related point - after you de-bias the precipitation products, does it really matter (for a long-term water balance study) how - or even IF - you merge the products? After de-biasing, they all have the same long-terms means and will thus produce the same long-term water balance analysis.**
- You are right, the seasonal de-biasing is an important step, especially for the precipitation. Although the season de-biased products will have same season, their inter annual, long-term or short term variations will not be the same. This is now clearer in the text: *After the seasonal de-biasing, all the precipitation products will have a similar season, but their long-term trend, inter-annual or monthly variations will still be different. In particular, the seasonal de-biasing will not change the trend of the EO products.*

• **Page 16, Line 3. The PF approach is designed explicitly to reduce closure residuals. So, it is questionable to use the reduction of closure residuals as evidence that that PF approach is "working" or that a closure constraint is necessary. Another possibility is that not all flux and stock terms are being accounted for him. That is, the flux/stock estimates utilized here are actually accurate but nevertheless should not close. Some discussion of this possibility is needed. The same issue comes up in Section 6, first paragraph. By design, the author's approach reduces residuals (that is a given).**

However, can this really be taken as objective evidence that flux or stock predictions have actually been improved?
- Yes, closure could happen for the wrong reasons, and we could correct fine EO products for compensating for missing components. The assumption is here that the missing components are random and that the merging will reduce their impact. The only way to make the closure constraint is beneficial is to evaluate the process using independent *in situ* data. This is done in our paper for precipitation and evapotranspiration, it was done also in: Combining data sets of satellite-retrieved products for basin-scale water balance study: 2. Evaluation on the Mississippi Basin and closure correction model, Munier, Aires, Schlaffer, Prigent, Papa, Maisongrande, and Pan, JGR Atmospheres, 2014. What we actually test in section 4.1 is that our methodology is doing what it was designed to do, close the water budget. This is now clearer in the text: *As a closure enforcing, the constraint approaches could yield to a closure of the water cycle, but in degrading fine EO estimate for compensating the imbalance. In order to evaluate the performance of the methodologies in improving the EO estimate, the constrained products will be compared with in situ measurements. The following Section is for assessing that the methodologies do close the water cycle as they suppose to do. The impact of hydrological constraint (PF) as well as the INTegration (INT) and CALibration (CAL) processes on the spatial averaging of the water component estimates and the WC budget residuals, over the several Mediterranean sub-basins, is summarized on Figure A.1 in the Apendix.*

• **Why are there so few rain gauge stations applied to the precipitation analysis in Figure 5? It's difficult to believe that only 7 rain gauges are available in the Mediterranean basin.**
- The rain gauge stations came from the FLUXNET database. In this way, precipitation and evapotranspiration evaluation are performed in the same network (36 gauges for precipitation). However, precipitation is also evaluate using the *in situ* gridded dataset Eobs at the basin scale (Section 4.2).

• **Also, why is it that the best results (for INT) in Figure 5 occur OUTSIDE of the basin (where INT results are based on an approximate extrapolation)? This seems odd. If INT is performing an accurate downscaling, it seems like it would be more effective WITHIN the basin.**
- The fact that INT can have best results outside of the basin can be explained by the poor performance of the precipitation estimate over particularly complex topography (mountainous rainfall) or coastal pixels with

land/sea contamination due to the coarse spatial resolution of satellite estimates. This have been added in the text: *The evaluation of EO estimate at 0.25° spatial resolution using tower sites should be taken with caution. The poor performance of satellite estimate over particularly complex topography (mountainous rainfall) or coastal pixels with land/sea contamination could explain the difference between the INT estimate and the FLUXNET measurement at this particular location.*

• In Section 3.2 (on the "optimal selection" (OS) approach) necessary? The methodology section is already quite long and the OS results do not seem to make a significant contribution to the manuscript's results.

- Thank you for this comment. Following your suggestion, we suppressed the OS section, the method is simply explained at the beginning of section 4.4 on SW: *A general approach to deal with EO datasets in the analysis of the WC is to chose the best individual dataset for each one of the water components. This is the approach developed, for example, in the GEWEX project. In Pellet et al. (2018) an Optimal Selection (OS) was based on the minimization of the water budget residuals to select the best combination of individual dataset. On the contrary, the SW approach relies on the merging of several EO datasets for each water component, in order to reduce their uncertainty.*

• What is meant by "quasi-triangular balance" in Section 5.1? This terminology will likely be unfamiliar for some HESS readers (it was to me).

- Sorry, the term "quasi-triangular balance" was really clumsy. We just wanted to mention that Mediterranean WC is mainly driven by the European sub-basins and that African coasts are not contributing so much. We have replaced the title of section 5.1 by a straightforward description: "Analysis of the Mediterranean WC".

• Figures 4 and 6. A better use of colour would be to differentiate between the INT and CAL cases (which are very difficult to distinguish). Also, the INT+/- and CAL+/- notation should be explained in the figure caption.

- The reviewer might mean Figures 4 and 5. The colours are now changed and the caption explicits the notations.

• Some discussion of the statistical significant of differences in Fig-

ure 3 would be useful
- The correlation difference is statistically significant at the 70%-level based on the T-test. This has been added to the caption of Figure 3.

• **Overall the paper is quite well-written, but it does suffer from an excess of minor English usage errors. Superficial proof-reading in this regard would help.**
- The typos and English writing have been improved, we hope that the manuscript is now in a better shape.