SUMMARY

This revised manuscript reads infinitely better that the first version. My compliments for that!

This paper addresses a valid tool, not new but new to this application, to analyse the consistency of water variables. Given all the difficulties to address this inconsistency, encountered in other studies, it is therefore worth publishing. However, the conclusion that 'large-scale evaporation products are inconsistent', both among themselves as with other water budget components, needs more clarification. After reading through it a couple of times, I think the main message is to make sure that there is a discussion or conclusion that the problem of inconsistency does not necessarily lie in the satellite products, but in their underlying input components. Therefore, I advise 'acceptation with major revision'.

MAJOR COMMENTS

Even though the authors test a new comparison tool, one has to be aware of what is compared. In my opinion, this analyses could have gone a bit deeper and explain what the origin of the inconsistencies could be (for example, input meteo data). Let me explain with two examples:

1) Differences caused by the different ET data can either be caused by the method, or their input data. In the case of MOD16, for example, input data comes from the GMAO data. But correlations are made with the GPCP data. How do the GPCP data and the GMAO data correlate, for example in terms of rainfall, RH etc. That can be a cause for bad correlation. In my opinion, such causes need to be mentioned, because it could well be that the satellite observations of one method are better than another, but the other input data messes up the correlation. In my opinion, this needs to be discussed, even with an example.
2) Following from 1), in my opinion, one should have better compared the GPCP data with some components of the GMAO dataset. If inconsistencies can be found there, they can explain inconsistencies between MOD16 and other products not based on GMAO.

Therefore, this paper reads a bit like 'we have a really cool tool (which it really is!) and we use it compare water variables'. But for me, the conclusion that there is inconsistency that needs to be worked on (although a true statement) sounds a bit too easy. What are these underlying causes? Are they actually caused by the satellite data? Or by the model input components? This needs to be properly discussed.

Another point, but similar, that still worries me about Figs 4-7 is the outflow part. Again, this question pops to my mind: "are we comparing the right products?". Any remaining P-ET will either come out of the catchment as quickflow or baseflow. Baseflow can take months to years to come to the surface, whereas quickflow can leave the catchment in days. Is GRACE data corrected for outflow and how? If it is, ok, stop reading, my bad. If not, GRACE does not only show the inflow (P-ET), but also the outflow. And this TWSA needs to compared to P-ET-Q (i.e., Q being all outflow out of the catchment). This means that Figs 4-7 could either show an inconsistency in measured water volumes, or the difference between outflow and inflow, or (probably) a combination of both.

I am not saying that the above points need to be part of this paper (although it would be nice), but they need to be at least properly addressed in the discussion, conclusion and abstract.

MINOR COMMENTS

Only two typos found:

P1: line 15: after ‘environments’, remove ‘including’ and put the basin names in between brackets.

P2, line 3: ‘wide range’, not ‘wide-range’.