OVERVIEW

This is a well-written manuscript whose conclusions are derived via a methodologically evaluation analysis. I think that the (clear) description and evaluation of the version-to-version differences constitutes a significant enough contribution to warrant publication. However – prior to that point – the following two major points should be addressed.

We thank the referee for assessing the quality of the paper and giving some interesting comments. Below, we give a point-to-point reply to the comments posted by the reviewer.

MAJOR COMMENTS

1. In the introduction, the author's argue that GLEAM is unique in that it is "primarily driven by microwave remote sensing observations." So the novelty here really seems to spring from 1) the assimilation of microwave-based soil moisture and 2) the use of microwave-based vegetation optical depth in the canopy stress formulation. If you take away these two aspects, the approach really just collapses down into a basic rain-driven soil water balance approach (which is relatively simple compared to the combined water/energy balance land surface already being run globally in e.g. GLDAS).

   We thank the referee for this comment. We would like to emphasize that (a) not only the soil moisture and vegetation optical depth datasets are based on microwave observations, but that the entire dynamic forcing dataset of the GLEAM v3b and c are based on satellite observations, and (b) a third important feature of the model is the detailed estimation of interception loss via the modified Gash's analytical model (Miralles et al., 2010), which was (for instance) used to benchmark the MERRA reanalysis and correct its interception estimates in new releases (Reichle at al., 2017). Therefore, we claim that the model can be primarily driven by satellite observations, being in their vast majority of microwave nature (soil moisture, precipitation, vegetation optical depth); thus available also during cloudy conditions, which is a unique feature for this type of models dedicated to estimate terrestrial evaporation from remotely-sensed data, since other models (such as e.g. Zhang et al., 2010; Fisher et al., 2008; Mu et al., 2007) need to rely on reanalysis meteorology, due to the requirements of atmospheric humidity, and on optical greenness data.

   We fully agree with the referee that the approach is simpler than land-surface models such as GLDAS, which provide a more complete representation of land surface processes: the added value of GLEAM is that it is specifically designed to estimate terrestrial evaporation and that it has been thoroughly evaluated and validated in regards to its skill to perform this very specific task. Needless to say that the variability in the representation of evaporation in more complex models is actually very large (Jimenez et al., 2011), mostly due to the fact that these models

are not specifically developed to estimate the evaporation flux accurately, independently of their complexity. Nonetheless, it should be noted that it is not our intention to present these features of GLEAM as innovative, as the core of the model was developed in 2011 based on this same rationale. We will do an effort to incorporate these points in the revised version.

So it would strengthen the paper if there were more support for the assertion that GLEAM is driven "primarily" by surface microwave observations.

Figure 5 and 6 are clearly an attempt to do this… but the results are not very compelling. The second and third columns of Figure 5 show that the background water balance model is generally superior to the assimilated observations. So naturally, more weight is (generally) placed on the water balance model background. This is ok… but it is really consistent with GLEAM being "primarily" driven by the microwave surface observations? Instead, it seems more accurate to say that GLEAM is being "primarily" driven by water balance considerations and these balance considerations are being nudged by "secondary" considerations derived from microwave DA.

We would first like to stress that the water balance is primarily driven by microwave-based precipitation, and evaporation estimates based on microwave precipitation, microwave vegetation optical depth and satellite-based (or reanalysis) meteorology. We agree nonetheless that the text should state that GLEAM is mostly driven by satellite data, which are primarily derived from microwave sensors. Further, Figures 5 and 6 mainly show that the impact of the DA on the modelled surface soil moisture strongly depends on the quality of the model open loop soil moisture, which on his turn is highly impacted by the quality of the precipitation forcing. We will clarify these points further in the manuscript.

No comparable results are shown for either root-zone soil moisture or ET… presumably because the impact of microwave DA is even less for these outputs.

The impact of the data assimilation system on the estimated evaporation is indeed limited and not discussed here. For a more detailed discussion about the impact of the data assimilation on evaporation, we would like to point the referee to Martens et al. (2016).

I realize that some of this is just semantics (i.e. what constitutes "primary" versus "secondary")… but I do think that the authors should either: 1) present better evidence for the "primary" role of the microwave observations in GLEAM or 2) be more objective in describing the novelty of their approach… particularly the impact of their novel methodological elements relative to approaches (like a classical soil water balance model) which have been around for quite some time.

We note again that the precipitation is also microwave-based (to the largest extent). In the revised version of the manuscript, we will try to be more specific about what we can consider novel features of our approach (see first response).

2. Some type of statistical significance analysis is needed to assess the noted version-to-version differences. I do not think that "statistically-significant" differences should be a requirement for publication. Nevertheless, the reader should be given a sense as to how large the stated performance differences are relative to expected levels of sampling noise.

   We agree with the reviewer that we need to support the results with statistical significance tests. Therefore, in the revised version of the manuscript we will include the results of a statistical test to verify whether differences in correlations are significant or not. The discussion of the results will be based on these tests as well. We note, nevertheless, that the two versions are similar on their estimates, and that the rationale for updating the method has been to make it more realistic while keeping the simplicity of the algorithm, as well as extending substantially the dataset temporal record based on the adoption of a new range of forcing data.

MINOR COMMENTS

1. Page 1, Line 7… I'd stay away from subjective statements like "most of these variables can be relatively easily observed at different spatial scales"… it is a stretch to call the remote estimation of rainfall (for example) "easy"… much safer to say from the remote retrieval of ET is difficult relative to other water balance components.

   This will be updated.

2. Figure 6 does not seem to be references in the manuscript. Also, unclear why case 3c is dropped when moving from Fig. 5 to Fig. 6.

   The results in Figure 6 are only briefly referred to at P13-L17-18 of the original paper, since the conclusions are analogous to the ones that may be drawn from Figure 5. In the new manuscript, we will further elaborate on the results in Figure 6. Anomaly correlations are not calculated for the GLEAMv3.0c (and thus not shown) as the period covered by this product is only 5 years (2011–2015). Given that none of the *in situ* stations fully covers this period with measurements, it is believed that this period is too short to calculate a robust climatology. This will be mentioned in the revised manuscript.