

Response to Astrid Kerkweg's Comments

Dear authors,

In my role as executive editor I ask you to move the Code Availability Section to its regular place after the conclusion but in front of the Appendix when revising your article.

Thanks, Astrid Kerkweg

Response: We thank you very much for managing our manuscript. The Code Availability Section has been moved according to your suggestion.

Response to Diego Miralles's Comments

Dear authors and editor,

I have read this manuscript with great interest. The method has a good potential, and I am aware that the authors have put several years of efforts behind the development of this methodology. I think that the consideration of the atmospheric boundary layer dynamics and the consistent radiative fluxes, set the uniqueness of the model.

Nonetheless, there are a few points that need to be reconsidered, better put in context, or simply clarified, to make the paper a better manuscript. In the following I provide a summary of my assessment.

Response: Dear Diego Miralles, we thank you very much for the constructive comments and suggestions. In the following, we provide an item-by-item response to your specific comments. Your comments are written in italic black color; our responses are shown in upright font blue color.

Some general comments:

1. – On the satellite nature

The title 'High resolution land surface fluxes from satellite data', or sentences like 'The required drivers for HOLAPS comprise satellite data at different processing levels as well as re-analysis data for a limited number of variables' may feel to the reader a bit... 'optimistic'. The truth is that about half of your forcing variables (i.e., air temperature, air humidity, windspeed, air pressure) come from ERA-Interim (which is fine!). I would, therefore, try to soften these statements all across the manuscript, starting from the title. Perhaps 'observational-based' is a more accurate and still relevant way to refer to the data sets, as it reflects the nature of both satellite and reanalysis forcing.

Response: Thanks for the comments. You are right, the HOLAPS is driven by both satellite and reanalysis data. The HOLAPS is designed to maximize the use of satellite data. However, there are no direct satellite observations for variables such as windspeed, surface pressure as well as air temperature. The reanalysis dataset is then used as a substitute. In that sense, HOLAPS is similar to many other global ET datasets. To reflect the current status of HOLAPS, the title has been changed to 'High resolution land surface fluxes from satellite and reanalysis data'. In addition, the phrase 'observational-based' has been used in the manuscript to replace the phrase of satellite data.

Along those lines as well, I feel a bit uncomfortable with the recurrent sentence (appearing up to four times...) stating that HOLAPS is designed to maximise the use of satellite data. If it was, it would use no reanalysis data at all. This is possible by just using a Priestley and Taylor approach, which in principle does not rely on humidity or windspeed data – see e.g. some GLEAM versions using satellite data only in Miralles et al., 2014 (S.I.). However, I am not claiming that

considering humidity or windspeed is not a good idea; yet, I do not think it is steered by the rationale of maximising the satellite nature of the model. On the other hand, if by 'maximising' the authors actually mean that all relevant satellite observable E drivers are exploited, I tend to disagree once more, since important variables for evaporation – like soil moisture, vegetation optical depth, or fluorescence – are observed from satellites and not included in HOLAPS. Therefore, I would avoid the use of 'maximising' in this context, and try to find a less ambitious but more accurate way of phrasing it. Perhaps what you achieved is a physically-sound algorithm that relies to a large extent on observable forcing. I'd be supportive of statements along those lines.

Response: Thanks for the comments. We agree that the current version of HOLAPS has still potential to make further use of more satellite information, like e.g. surface soil moisture, like it is done within the GLEAM model. What we were aiming for is to actually maximize the consistency across different used forcing datasets. The rationale for using wind speed and humidity information is twofold. The total water vapor content from ERA-interim is required to calculate the surface solar and thermal radiation fluxes. In principle this can be also replaced by climatology data. In addition, the wind speed is required for the calculation of the sensible heat fluxes, which had not been further analyzed in the original version of the manuscript.

According to your suggestions, the sentences with the description 'maximizing the use of satellite' in the manuscript have been changed to 'makes use of satellite and reanalysis datasets'. We have also included a further discussion on the usage of other satellite data in future HOLAPS version in the discussion section.

2. – *The succinct abstract*

The abstract could be a lot more comprehensive and include: (a) the period covered by the data set, (b) the statistics of the validation (avoiding the use of 'very good' in reference to these statistics and noting the numerical value of the inferences), (c) state the spatial and temporal resolution, (d) can anything be said about the global results in Fig. 1?

Response: Thank you for the suggestions. We have revised the abstract and added more information about the obtained error statistics from the present study. However, we have not added any further details regarding a global dataset. The major purpose of the present study is to provide a point like evaluation of HOLAPS results and perform a sensitivity analysis regarding the impact of different forcing and forcing combinations used. A global demonstrator product had been generated for a time period of 2001-2005, but this was based on a previous model version. The model has undergone major improvements since then and the demonstrator product results are therefore outdated.

We are currently in the phase of re-running the model at global scales which will lead to a dataset with a spatial resolution in the order of 1km and an hourly temporal resolution. The dataset will be produced and analyzed within the year 2016 and then released to the public. It will be based on the code used for this manuscript.

Our motivation is to publish first comprehensive local validation results (this paper) before we go to a large scale production and assessment. That's why we did not provide further details regarding the global dataset.

3. – The way to frame the need for HOLAPS

This has to be rephrased in the Introduction to some extent. The argument of higher spatial and temporal resolution was very valid five years ago, when this effort started. However, as of 2016, existing global heat flux models have been run already at finer resolutions than 5km, and at sub-daily scales. Moreover, statements in the lines of 'The currently existing datasets have spatial resolutions between 0.25 and 2.5 and are focused on daily to monthly timescales (Miralles et al., 2011; Mu et al., 2007; Vinukollu et al., 2011)' are in fact not really accurate. Global datasets have currently resolutions of 1km–25km, and most have been run at 3-hourly resolution or even higher (see e.g. McCabe et al., 2015; Michel et al., 2015; Miralles et al., 2015). In fact, even the papers cited by the authors – which date from at least five years ago – report results at resolutions that differ from the ones stated in the manuscript: Miralles et al. (2011) was daily 0.25 degrees, Vinukollu et al. (2011) was daily 5km, and Mu et al. (2007) was 8-daily 0.05 degree resolution.

Therefore, based on these citations, the reported range 'between 0.25 and 2.5' and 'daily to monthly timescales' should be 'between 0.05 and 0.25' and 'daily to 8-daily timescales'. This is choosing these early efforts and without taking into consideration any progress over the last five years. In summary: please, highlight the high resolution of HOLAPS without stating that it is notoriously higher than for any existing analogous data sets, and specially try to reflect more accurately the state of the art as of 2016.

Response: Thanks for the comments. The descriptions of the current status have been updated through integrating the recent progress of the estimation of global heat flux within the revised manuscript.

4. – The added value upon ERA-Interim

The authors design several interesting experiments altering the observations used as radiative forcing in HOLAPS. The reader may wonder at this stage what the result of using ERA-Interim radiation in HOLAPS would be. This is an important question, because the added value of having a radiative module in HOLAPS should be explored and understood, and ERA-Interim is used for several other forcing variables making it consistent with the current setting. My feeling is that results should deteriorate in general when using ERA-Interim, even if it is just due to tower-to-footprint errors.

Then, an equally interesting question for most readers would be how the validation of the latent heat flux from HOLAPS compares to the validation results of the latent heat flux from ERA Interim (e.g. adding a sentence in the description of the results, or even a new panel to Figure 7). This will not only provide an estimate of the added value of having a radiative module but of the entire methodology in general. These two experiments appear important, even if HOLAPS does not outperform ERA-Interim dramatically.

Response: Thanks for the comments and suggestions. The use of surface solar radiation flux from ERA-Interim would give further information on the model sensitivity to radiation forcing, we agree. However, like you said, ERA-interim radiation is not expected to provide better results compared to the CM SAF-SIS radiation forcing. A major reason for this might be that the ERA-interim radiation data has a rather coarse spatial resolution and can not properly resolve clouds. Through inter-comparison and validation against ground-based measurements, Posselt et al., 2012 found that CM SAF radiation data outperforms ERA-Interim surface solar irradiance data at both daily (monthly) time scales with BIAS of 4.58 w/m^2 (4.4 w/m^2), standard deviation (SD) of 23.73 w/m^2 (8.14 w/m^2) for CM SAF, and BIAS of 5.64 w/m^2 (5.71 w/m^2), SD of 39.14 w/m^2 (12.15 w/m^2) for ERA-Interim.

We therefore did not expect that using ERA interim radiation data would improve the results in any sense. We have therefore decided to not include an additional experiment to the study, but rather discuss the above points in the discussion section of the manuscript which has been extended accordingly.

We also agree that an inter-comparison of HOLAPS results against other LE estimates would be very interesting. Like we outlined above, the current manuscript focuses however on a point scale evaluation and sensitivity analysis of the HOLAPS model itself and NOT on a cross-comparison with other datasets. We believe that it does not make too much sense to compare ERA-interim latent heat flux estimates, which are gridded at coarse spatial resolution, against HOLAPS point-like estimates from the present study. One would compare apples with oranges due to the different spatial scales and land cover heterogeneity.

We plan however a comprehensive cross-comparison study of HOLAPS global results against a multitude of other datasets (e.g. ERA-interim, GLDAS, GLEAM, GEWEX ...) in a separate study. Comparing HOLAPS results at the regional or global scale against these other dataset would be much more beneficial than comparing ERA-interim against local HOLAPS data and **FLUXNET** measurements. We have therefore decided to not include such a comparison in the revised version of the manuscript.

5. – On the low values in tropics

I couldn't help noticing that the values in the tropics are low; 60–70 W m^{-2} is about 800–900 mm year⁻¹, which it is almost a two-fold difference from what I would expect based on literature. As an example, the Amazonia evaporation is about the same as the evaporation in the Pampas region now. That is not normal, and it may be partly due to the calculation of interception loss based on a Priestley and Taylor equation, which can lead to a several-fold underestimation of the flux. This is a practice that has long been criticized (see e.g. Shuttleworth and Calder, 1979), given that the energy used for evaporation in wet canopies does not come from net radiation (Holwerda et al., 2011; van Dijk et al., 2015; Savenije, 2004). This may explain the low gradient of evaporation in the tropics observed in Figure 1. In addition, since there are no towers in tropical regions this issue would not show in current validations.

Response: Thanks for the comments. We agree that the values in the tropics are too low in Figure 1. We should confess that the results shown in Figure 1 are based on old HOLAPS simulation, which actually had a severe error in the calculation of the surface fluxes. This error had in particular an impact for regions with large radiation and dense vegetation. This bug has been fixed now. Results shown in the present study were all based on a HOLAPS version where the bug had already been fixed – except for Figure 1. Like outlined above, we are currently running the HOLAPS over global scale. To avoid any misunderstanding, we have decided to remove Figure 1 for the revised version of the manuscript, as it is not based on the same code version like all other results in the manuscript.

Thanks also for your suggestion regarding the impact of interception ET in the tropics. As HOLAPS is taking explicitly into account interception evaporation (similar to e.g. GLEAM) we actually believe that this effect should be of less of importance for HOLAPS.

6. – No sensible heat flux results

The title is very promising, but the paper soon tilts towards evaporation only. If latent and sensible turbulent heat fluxes are calculated by the model separately, and they are available at the tower level, why do validations and maps refer to the latent heat flux only?

Response: Thanks for the questions, we agree with your statement. The reason why we focused in particular on LE is that many of the available datasets are ET only datasets. As HOLAPS calculates both, we agree that it would be beneficial to provide this additional information. The results for sensible heat flux are therefore added in the manuscript and are being discussed.

Comments on Figures and Tables:

– Figure 1: I would add a colormap with the units in mm/year. That would make it more comparable to other studies. How does this map qualitatively compare to other datasets? (just one line or two in the text). I would also be interested in seeing the analogous map of H. Also, why only 2001–2005?

Response: Thanks. As we said before, the Figure 1 was based on old simulation that contained a bug. The HOLAPS results provided in the manuscript are only based on tower-scale simulation. Therefore, we decided to remove Figure 1.

– Table 1: It may be a good idea to indicate the temporal coverage and spatial domain of the forcing variables in the table.

Response: Thanks for the suggestion. Temporal and spatial coverage of the datasets have been added to the Table 1.

– Table A1: It is very useful to have this table of acronyms. There are however several acronyms that are not defined in the main text. I am not sure whether that is fine with the policies of the

journal, but it does make the reader need to rely on this table. I pointed to some of the acronyms that are undefined under the list of minor comments below.

Response: We thank you for the comment. We had decided to use the acronym table, as the manuscript contains a lot of different symbols. We therefore wanted to have a central place to allow for the reader to easily look-up variables and unit definitions. Some of the variables are explained in addition in the text or, if there are only constant values that are very specific to a particular equation, the variables had not been included into the acronym table.

We thought that this would allow the reader to have most efficient reading through the formulas without the need to explain each variable explicitly in the text, so this might become quite repetitive and hard to read.

We have adopted the manuscript following your comments to ensure that all relevant acronyms are contained in the acronym table first and that they are also explained in the text where suitable to improve the readability of the text itself.

– Figure 2. Within the diagram of HOLAPS estimated fluxes and modules, one can find precipitation now. It may be an error.

Response: The caption might have been a bit confusing. What the figure should illustrate are the HOLAPS drivers as well as estimated fluxes. Thus, precipitation is used as a driver for HOLAPS like discussed in the manuscript. The figure caption was adapted accordingly.

– Figure 4. Note the experiment name in the caption. It seems that the issues with the radiation come at nighttime; could you comment somewhere on that? Other than nighttime it seems that the correspondence is extremely good.

Response: The experiment name has been added in the caption. Nighttime values are included in the comparison now (see next comment).

– Figure 7. It seems that evaporation is always positive. Can you comment on that as well? Nonetheless, the axes are cropped to zero in b) but not in a); any reason for that?

Response: We had explicitly excluded the nighttime from the comparison, assuming that FLUXNET measurements are more uncertain under these more stable conditions. For the revised version of the manuscript, we removed this condition and also include nighttime values now. Including the nighttime values actually even did slightly increase the skill scores. All figures and their scales are adapted in the revised version of the manuscript.

– A general comment about the figures and tables. I have the impression that the statistics are now calculated by plotting together all the stations at the same time, in a single scatter plot. My apologies if this isn't the case. That is problematic, because the energy imbalance in FLUXNET towers is rather systematic but location-dependent. That would penalise the statistics. On the other hand, another confounding effect will unfairly favour the statistics instead: there are towers at the wet-end and at the dry-end of your scatters, meaning that a correlation now does not necessarily mean a correlation at the tower level. As the authors know, a better idea to

evaluate the quality of the model is to provide the mean of the statistical inferences calculated at each tower. In other words: the mean of the 49 correlations is better than the correlation of the scatterplot constructed by adding together the 49 towers with their own climatological mean and biases (as in Figure 7). This would in addition make the validation comparable to most literature studies. If this is the way it is done now, please describe it in the paper and captions. I have the impression that that may be the way the whiskers plots are designed but I could not find the information.

Response: Thanks for the comments. In the current manuscript, both kinds of statistics are analyzed and shown. You are right, the mean of the statistics that are calculated at each tower are shown in the form of box plots. To avoid confusion, the captions and the descriptions in the manuscript are improved.

– *One final comment for the scatterplots: labels should be bigger!*

Response: Done. Thanks.

Comments on the model:

– *The terms used in equation B1, B2, etc. Do they need to be defined after the equations, or is it clear enough with the acronyms table?*

Response: see our comment above regarding the acronym table usage.

– *L18 P10804: Italics in 'L'.*

Response: Done. Thanks.

– *As mentioned below, calculating interception loss based on a Priestley and Taylor equation can lead to a several-fold underestimation of the flux, and this may partly explain the low values in rainforests (Figure 1).*

Response: Thanks for the comment. The Figure 1 has been deleted like discussed above. Like outlined already in previous responses, HOLAPS calculates explicitly the interception evaporation. We therefore believe that the interception loss should not be a problem, but would be in general interested to validate in more detail the particular interception evaporation component of HOLAPS if possible.

– *I may be missing where windspeed enters the calculation of LE, or whether it is only used in the calculation of H, given that the aerodynamic resistance does not appear directly in the latent heat flux calculation. Since the input variable is already there, wouldn't it be interesting to use this resistance in the latent heat calculation? – maybe even within your inhibition function? My apologies if I am just not finding the equation.*

Response: You are absolutely right. The windspeed is currently only used to calculate the aerodynamic resistance for the sensible heat flux calculations. It is not used to calculate any surface resistance terms for the latent heat flux estimates so far. In principle we agree that this

would be interesting to do as it would lead to a more consistent estimate of LE and H within HOLAPS, but we plan to conduct this work for a future version of the model code, which will also minimize the dependency on the surface air temperature in the calculation of the surface fluxes (see e.g. Mallick et al, 2015¹).

– L10 P10807: *in the partitioning of net radiation within the fractional covers, is the albedo of vegetation and bare soil considered to be the same? – How are snow times considered?*

Response: Thanks for the comment. Currently, a single albedo value is used to calculate the net surface radiation which is then disentangled into soil and vegetation components, like described in the manuscript. We will consider the use of data which provides information about the (soil) background albedo like e.g. provided by the TIP product (Pinty et al., 2011²) in a future version of HOLAPS. Snow times are not considered in the present study. These are excluded by proper choice of the investigation periods.

– *What does the model do at nighttime?*

Response: The model is also running during nighttime. Shortwave radiation forcing is obviously zero during the nighttime hours.

Other minor comments:

– L14 P10784: *remove 'global'.*

Response: Done.

– *The model classification in the second paragraph of the Introduction: many of the current global models that are based on Penman-Monteith or Priestley and Taylor approaches (e.g., Mu et al. 2007; Fisher et al., 2008; Miralles et al., 2011a,b; Zhang et al., 2015; Zhang et al., 2016) do not fit within any of these categories. Note that Fisher et al. (2008) is certainly not an energy balance model, as it does not close the energy balance.*

Response: We have added a new category: (4) the methods based on Penman-Monteith or Priestley and Taylor equations.

– *The part mentioning WACMOS-ET in the Introduction feels like if it was added afterwards, and needs to be better integrated within the description of the state of the art. Again, within this description I would try to be more inclusive in what comes to other existing efforts (e.g. Zhang et al., 2015; Zhang et al., 2016; Anderson et al., 2007), and highlight the two main features of HOLAPS: (a) a consistent radiative forcing thanks to the ABL modelling, (b) a high spatiotemporal*

¹ Mallick, K., E. Boegh, I. Trebs, J. G. Alfieri, W. P. Kustas, J. H. Prueger, D. Niyogi, N. Das, D. T. Drewry, L. Hoffmann, and A. J. Jarvis (2015), Reintroducing radiometric surface temperature into the Penman-Monteith formulation, *Water Resour. Res.*, 51, 6214–6243, doi:10.1002/2014WR016106.

² Pinty, B., I. Andreadakis, M. Clerici, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J.-L. Widlowski (2011), Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties, *J. Geophys. Res.*, 116, D09105, doi:10.1029/2010JD015372.

resolution allowed by the use of geostationary satellites. There is certainly room (and need) for new approaches like HOLAPS, but the way it should be embedded within the existing body of models needs to be phrased more precisely.

Response: Thanks for the suggestions. The introduction will be rewritten and we will take this comment into account in the revised version of the manuscript.

– L5 P10788: 'is' for 'are'.

Response: The typo mentioned by the reviewer could not be found at this page and line.

– L15 P10788: same thing.

Response: We could not identify an issue here neither.

– Sentence: *'The drivers required to force HOLAPS are summarized in Table 1. These consist of satellite remote sensing data products'. Again, try to be more accurate and objective in these statements. Half of the driving data sets are not satellite-based (note also that you miss ERA-Interim pressure from Table 1, add). It is not something to be ashamed of... I personally do not see the issue with this. Try to be a bit more clear about it.*

Response: Thanks. The description has been improved to 'These consist of satellite remote sensing and reanalysis datasets'. In addition, the missed ERA-Interim pressure has been added to Table 1.

– L5 P10789: you may need to define these symbols.

Response: The definition of the symbols and a cross reference to the acronym table has been added.

– L10 P10789: *'a simple energy balance correction is applied', put in brackets what it consists on.*

Response: The Bowen ratio method is added in the bracket.

– *'stable conditions. As the eddy covariance' for 'stable conditions, as the eddy covariance'.*

Response: Done.

– L12 P10789: *'dataset' for 'datasets'.*

Response: Done.

– *'maximum consistency between the shortwave and longwave radiation fluxes is ensured as the same ancillary data is used'. The authors state that this is a unique feature of HOLAPS, but could you add a sentence stating why this same feature does not apply to other net radiation datasets from satellite observations, like SRB or CERES (or even reanalysis data, like ERA-Interim). My feeling is that other turbulent flux models that calculate air temperature, like ALEXI*

(Anderson et al., 2007), may be even more consistent in their shortwave and longwave forcing. I am quite sure I did not understand well this description; since this feature is a uniqueness of the model it may be appropriate to elaborate this part a bit more in the main text.

Response: SRB, CERES and reanalysis data are at much coarser spatial resolutions. They are nevertheless consistent in their short- and longwave flux estimates like the reviewer is saying. We agree on that point. However, if going to high temporal and spatial resolutions, there is to our best knowledge currently no product existing ensuring consistency between the shortwave and longwave radiation fluxes. The products from the CMSAF and Land SAF for instance are estimating these flux components independently (only a common cloud mask is used).

The HOLAPS approach is rather similar to ALEXI, thus these two would be somehow comparable. We mention this in the revised version of the manuscript.

– P10791: *If the authors are using TRMM data, the references to global model output should be corrected. I don't think that 50N–50S can be considered as global... living in Brussels I find a bit excluded by that statement ;). Maybe 'quasi-global' is the right way to refer to TRMM data; we have used this in the past in the same context (see e.g. Crow, Miralles and Cosh, 2010 'A Quasi-Global Evaluation System for Satellite-Based Surface Soil Moisture Retrievals').*

Response: Thanks for the suggestion. The term 'quasi-global' has been used to replace 'global' in the manuscript.

– P10792: *there are several one-line paragraphs in the paper that may be grouped into longer paragraphs.*

Response: Done.

– P10792: *'A limited number of additional fields (temperature, wind speed, total column water vapor path, pressure) are required from global re-analysis'. Please, remove 'limited'. Referring to 4 satellite input variables as 'maximising the use of satellite data', while then stating that 4 reanalysis variables are 'a limited number', does not feel too objective... Again, there is no reason to feel uncomfortable about using reanalysis data, especially since there are no 3-hourly global satellite fields for these variables.*

Response: The term 'limited' has been removed.

– *'which corresponds to a spatial sampling of 0.7', despite this forcing resolution the final output of HOLAPS comes at more than an order of magnitude higher resolution. Therefore, I think a sentence is required somewhere to explain this issue or some readers may get a bit confused at this stage.*

Response: We agree. We have therefore added the following sentence: "The scale mismatch between the used reanalysis data and the local scale HOLAPS simulations might result in additional uncertainty in the simulations." It is emphasized that the scale mismatch between different forcings and the local scale simulations will directly affect the validation statistics. One motivation to do the different experiments like shown in the paper was actually to illustrate the

differences between local forcing only (CTRL simulations) and large scale forcing from either single forcing fields or a combination of forcing fields from satellite and reanalysis data.

– *‘However for the present study, no global land cover dataset is used as the experiments conducted are only performed on the point scale’, but were not they used in the global map in Figure 1?*

Response: As stated before, the current HOLAPS simulation is only based on tower-scale. The global map (not used in the manuscript any more) was based on ESA GlobCover data and the forthcoming HOLAPS dataset will be based on the ESA CCI global landcover dataset.

– *‘Currently the HWSD is the only globally available soil information.’ This is a bit of a weird (and inaccurate) statement. Could the authors elaborate a bit more on what they mean by ‘information’? Otherwise just delete, because there are quite a few global datasets of soil characteristics available out there...*

Response: We agree that there is more than HWSD (and the commonly used FAO) dataset. The sentence has been therefore deleted like suggested.

– *L14 P10793: ‘related to the usage of satellite data...’. Add: ‘and reanalysis’.*

Response: Done.

– *Forcing the model with FLUXNET data ‘allows to quantify HOLAPS accuracies without additional uncertainties from the driver variables’. The authors spent quite some effort discussing the uncertainties FLUXNET measurements, and the plenty of representativeness and resolution issues in the forcing and ancillary datasets. This statement needs to be deleted or corrected.*

Response: The sentence has been changed to ‘allows to quantify HOLAPS accuracies without additional uncertainties from the satellite and reanalysis datasets’.

– *P10795: a fourth source of errors is the one I just mentioned, the differences in footprint depending on the measured variable, etc. The authors may add this as ‘d)’. This error occurs even if each measurement is perfect, so it is not to be considered as the same as forcing uncertainty...*

Response: Thanks for this comment. We are fully aware of the representativeness error when comparing FLUXNET against gridded datasets. However in the present case, the simulations are done at local (point-like) scale and the representativeness error is in that case not applicable for the comparison of the FLUXNET measurements against HOLAPS estimates themselves, but is rather included as part of the forcing uncertainties.

We have clarified this in the revised version of the manuscript as follows:” The variance of the difference between the model simulations and FLUXNET data is a function of a) the uncertainties of the HOLAPS model itself, b) the sensitivity of the HOLAPS model to uncertainties in the forcing data (including representativeness error) as well as c) uncertainties in the FLUXNET reference data.”

– I do not understand the need to refer to the MSD, RMSD, R, etc. as E^2 , E' ... For me

E^2 is evaporation squared (?). Is this needed?

Response: We agree that using the letter 'E' might be confusing in this context, like you say, some reader might refer to it as the power of evaporation. We followed here the original notation given by Taylor (2001). To avoid any confusion, the letter E in the equations has been removed and the paragraph has been revised.

– 'A monthly mean is calculated if at least 2/3 of the days of a month contained valid values'. Are these times also deleted from the gridded forcing when performing comparisons between tower-forced output and gridded-forced output?

Response: Yes, these days are also deleted when gridded forcing is used. Thus the same temporal masking applies for all experiments to ensure consistency of the statistical values obtained.

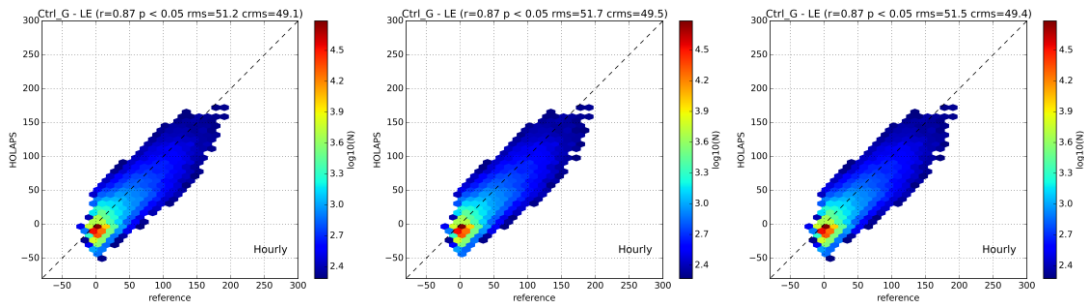
– What do you do with rainfall times in the validations? The latent heat flux measurements from eddy covariance are typically removed when it rains in this kind of comparisons (see e.g. Ershadi et al., 2014; Miralles et al., 2011; Michel et al., 2015), because of their little reliability when humidity approaches saturation (see e.g. van Dijk et al., 2015). This also means that the interception loss of the model cannot be validated through comparisons to FLUXNET data... These issues should be acknowledged somewhere.

Response: We agree that eddy covariance measurements are less reliable under rainfall conditions, like the reviewer is saying. The baseline for the comparison in the paper is, that we apply the FLUXNET quality flags and only perform comparison with data that is assumed to be of highest accuracy regarding to the FLUXNET data providers. This covers already many rainfall events.

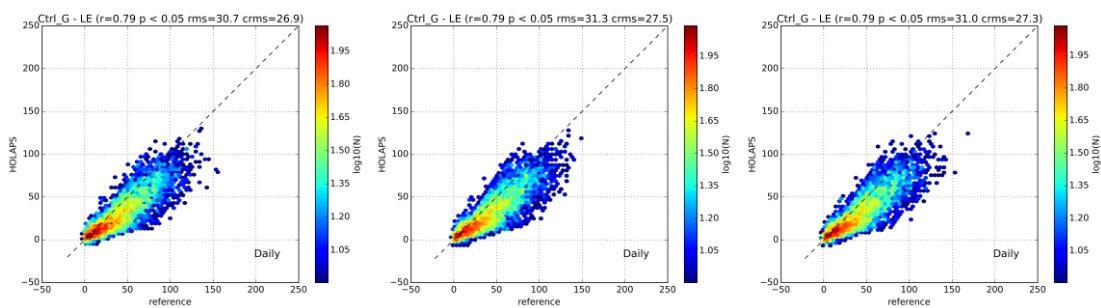
Following the suggestion of the reviewer, we performed a sensitivity analysis where we excluded rainfall events, using different thresholds for the 30' rainfall intensity recorded at the FLUXNET stations. All samples where the rainfall intensity was above the threshold were excluded from the comparison. We used different thresholds, as we could not find applicable threshold values in the references provided by the reviewer.

This additional filtering did not have major impact on the comparisons. This is due to the fact, that many precipitation events are already covered by the flags already applied. The following figures show the overall LE statistics for the data with FLUXNET flags applied (left) and an additional rainfall filter for intensities of > 0.5 mm/h (middle) and > 1mm/h (right).

We have added a brief discussion on the principle need for precipitation filtering in the revised version of the manuscript.



Hourly



Daily

– Again, there is a bias towards latent heat, which in the case of FLUXNET is in fact more unreliable than the sensible heat flux measurement. Is there any way to include a figure on H , apart from a global map, showing the statistics of the validation? Note that the current title and the name of the model do not reflect this bias towards evaporation.

Response: Like stated in a previous answer, we have followed the reviewer’s suggestion and have added the sensible heat flux results to the manuscript.

– P10799: ‘Error statistics for all experiments is’, change ‘is’ for ‘are’.

Response: Done.

– ‘are smallest’ for ‘are smaller’ or ‘are the smallest’.

Response: Done.

– Discussion section: I believe you are comparing your HOLAPS R to the $R2$ reported by Michel et al. (2015). That may explain some differences. Please double-check this entire discussion...

Response: We thank you for this comment. We will check this again and correct the manuscript accordingly.

– P10801: 'similar approach like (Michel et al., 2015)' for 'similar approach to the one by Michel et al. (2015)'.

Response: Done.

– Last few sentences of the Discussion, the word 'further' is repeated too many times.

Response: The second further has been deleted.

– Conclusion section: I would try to elaborate a bit more again on the main characteristics of HOLAPS and what it a priori adds upon existing methods. I think doing it in the form of a list (e.g., 'a', 'b', 'c') may help.

Response: Thank you for the suggestion; we will revise the conclusion section accordingly.

– Regarding the comparison to other datasets mentioned in the Conclusion, I would rather recommend the comparison to the more recent and carefully developed WACMOS-ET (Michel et al., 2015; Miralles et al., 2015) and GEWEX LandFlux (McCabe et al., 2015) databases.

Response: Thanks for the comment. We have compared our results to your suggested references.

I hope some of these suggestions help improve the manuscript. Despite the length of the review I believe the authors have done a very good job at developing the model.

Response: We thank you very much for devoting a lot of time to our manuscript. We believe all these comments are very constructive and useful and help to improve the manuscript. Very much appreciated!

Best regards, Diego

REFERENCES

Anderson, M., Norman, J., Mecikalski, J., Otkin, J., and Kustas, W.: A climatological study of evapotranspiration and moisture stress across the continental United States based on thermal remote sensing: 1. Model formulation, *J. Geophys. Res.*, 112, D10117, doi:10.1029/2006JD007506, 2007. 10746.

Crow, W. T., D. G. Miralles, and M. H. Cosh (2010), A quasi- global evaluation system for satellite-based surface soil moisture retrievals, *IEEE Trans. Geosci. Remote Sens.*, 48(6), 2516–2527, doi:10.1109/TGRS.2010.2040481.

Ershadi, A., McCabe, M. F., Evans, J. P., Chaney, N. W., and Wood, E. F.: Multi-site evaluation of terrestrial evaporation models using FLUXNET data, *Agric. For. Meteorol.*, 187, 46–61, doi:10.1016/j.agrformet.2013.11.008, 2014.

Fisher, J. B., Tu, K. P., and Baldocchi, D. D.: Global estimates of the land-atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites, *Remote Sens. Environ.*, 112, 901–919, 2008.

Holwerda, F., Bruijnzeel, L. A., Scatena, F. N., Vugts, H. F., and Meesters, A. G. C. A.:

Wet canopy evaporation from a Puerto Rican lower montane rain forest: The importance of realistically estimated aerodynamic conductance, *J. Hydrol.*, 414–415, 1–15, doi:10.1016/j.jhydrol.2011.07.033, 2011.

McCabe, M. F., Ershadi, A., Jiménez, 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, *Geosci. Model Dev. Discuss.*, 8, 6809–6866, doi:10.5194/gmdd-8-6809-2015, 2015.

Michel, D., Jiménez, C., Miralles, D. G., Jung, M., Hirschi, M., Ershadi, A., Martens, B., McCabe, M. F., Fisher, J. B., Mu, Q., Seneviratne, S. I., Wood, E. F., and Fernández-Prieto, D.: The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms, *Hydrol. Earth Syst. Sci. Discuss.*, 12, 10739–10787, doi:10.5194/hessd-12-10739-2015, 2015.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., and Dolman, A. J.: Magnitude and variability of land evaporation and its components at the global scale, *Hydrol. Earth Syst. Sci.*, 15, 967–981, 2011a.

Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, *Hydrol. Earth Syst. Sci.*, 15, 453–469, 2011b.

Miralles, D. G., et al.: The WACMOS-ET project – Part 1: Tower-scale evaluation of four remote sensing-based evapotranspiration algorithms, *Hydrol. Earth Syst. Sci. Discuss.*, 12, 10739–10787, doi:10.5194/hessd-12-10739-2015, 2015.

Miralles, D. G., van den Berg, M. J., Gash, J. H., Parinussa, R. M., De Jeu, R. A. M., Beck, H. E., Holmes, T. R. H., Jiménez, C., Verhoest, N. E. C., Dorigo, W. A., Teuling, A. J., and Dolman, A. J.: El Niño–La Niña cycle and recent trends in continental evaporation, *Nat. Clim. Change*, 4, 122–126, 2014b.

Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, *Remote Sens. Environ.*, 111, 519–536, doi:10.1016/j.rse.2007.04.015, 2007.

Savenije, H. G.: The importance of interception and why we should delete the term evapotranspiration from our vocabulary, *Hydrol. Process.* 18, 1507–1511, doi:10.1002/hyp.5563, 2004.

Shuttleworth, W. J., and Calder, I. R.: Has the Priestley-Taylor equation any relevance to forest evaporation?, *J. Appl. Meteorol.*, 18, 639–646, 1979.

van Dijk, A. I. J. M., Gash, J. H., van Gorsel, E., Blanken, P. D., Cescatti, A., Emmel, C., Gielen, B., Harman, I., Kiely, G., Merbold, L., Montagnani, L., Moors, E., Roland, M., Sottocornola, M., Varlagin, A., Williams, C.A., and Wohlfahrt, G.: Rainfall interception and the coupled surface water and energy balance, *Agric. For. Meteorol.*, 214–215, 402–415, doi:10.1016/j.agrformet.2015.09.006, 2015.

Vinukollu, R. K., Sheffield, J., Wood, E. F., Bosilovich, M. G., and Mocko, D.: Multimodel analysis of energy and water fluxes: intercomparisons between operational analyses, a land surface model, and remote sensing, *J. Hydrometeorol.*, 13, 3–26, 2011.

Zhang, K., Kimball, J. S., Nemani, R. R., Running, S. W., Hong, Y., Gourley, J. J., and Yu, Z.: Vegetation greening and climate change promote multidecadal rises of global land evapotranspiration, *Sci. Rep.*, 5, 15956, doi:10.1038/srep15956, 2015.

Zhang, Y., Peña-Arancibia, J. L., McVicar, T. R., Chiew, F. H. S., Vaze, J., Liu, C., Lu, X., Zheng, H., Wang, Y., Liu, Y. Y., Miralles, D. M., and Pan M.: Multi-decadal trends in global terrestrial evapotranspiration and its components, *Sci. Rep.*, 5, 19124; doi:10.1038/srep19124, 2016.

Response to Referee #2's Comments

A new framework (HOLAPS) for the estimation of global surface energy and water fluxes at the land surface has been described and verified with flux towers. The model is designed to achieve usage of globally available satellite data. Several options for global simulation has been suggested. Since this paper is the first one on HOLAPS model, it is suggested to specify more technical details and let the readers understand the physical process included in the model. The paper is encouraged for publication. Below is some minor comments.

Response: We thank you very much for the encouragement and suggestions. In the following, we provide an item-by-item response to your specific comments. Your comments are written in italic black color; our responses are shown in upright font blue color.

In your initial comments above you suggest to provide more technical details to let the reader understand the physical process included in the model. We agree that a detailed technical description is required, which is the reason why we actually provide the full technical description in the Annex of the manuscript to a level of detail that would in principle allow the reader to code the model itself. Thus we do not understand what the reviewer is further missing in terms of technical detail information. One objective of the present manuscript is the actual technically comprehensive documentation of the HOLAPS approach.

ΔS is the thermal inertia which is estimated as function of soil moisture conditions (Murray and Verhoef, 2007).-à rs 'H','D','M' Table 3 should be 'h', 'd','m' to make a consistent format as the table title. S is not found in the table. Is RG means S?

Response: Thanks for the comments. The corresponding parts have been updated where possible. We are however not sure what the reviewer meant by "S is not found in the table. Is RG means S?" Our interpretation of the question is that the reviewer meant the thermal inertia SIGMA which is calculated based on the reference provided in the manuscript. The symbol for SIGMA is also included in Table A1. If the reviewer did refer to the caption of Table 3, then this contained an error, where "S" was used instead of "Rg" for the radiation. We have fixed this in the revised version.

In equation B17 and B16, I am wondering how did 'z' is set for sensible heat flux calculation, since ERA-interim temperature and wind speed are used for sensible heat flux calculation.

Response: The wind speed $u(z)$ at vegetation height level is obtained by assuming a logarithmic wind profile with displacement height of 2/3 of the vegetation height. This has been clarified in the revised version of the manuscript.

Sensible heat flux is also one important output variable in HOLAPS. I did not find any evaluation of this variable. The accuracy of sensible heat could influence that of latent heat. Table B1 also list z_{0h} and z_{0m} for different land covers, it maybe interesting to verify these parameters by comparing calculated and observed sensible heat. Otherwise please discuss the possibilities for model accuracy improvement in future.

Response: The addition of an evaluation of the sensible heat flux has been also suggested by other reviewers. We have therefore added a corresponding section on evaluation of the sensible heat flux to the revised version of the manuscript which compares the estimated and measured sensible heat fluxes. The roughness length parameters currently used within HOLAPS are based on literature values. We are not sure what the reviewer means by “verifying these parameters”. Something one could do in principle is to use the FLUXNET sensible heat flux estimates and obtain the roughness lengths parameters by inversed modelling. However, we actually try to avoid such a model calibration procedure as it would limit our validation dataset.

I did not find the meaning of \tilde{A} in equation B6 in the text or in the symbol table. Please specify it.

Response: The missing parameter ω in equation B6 is the diurnal angular frequency. This has been clarified now in the manuscript.

I also did not find the clues on how did HOLAPS solve runoff (Q) in equation B8? The author gives surface water balance equation, but did not tell us how did they process water balance for other soil layers. If possible, please also add water balance for each soil layer. Can they also specify how did they solve soil suction pressure head Ψ in equation B9?

Response: Thanks for the question. We have clarified this point in the revised manuscript as follows: “The water fluxes between the different soil layers is solved using a numerical approach. The net soil water flux in a soil layer is hereby determined by the fluxes into and from the layers above any below, whereas the model allows for both downward (percolation) as well as upward (capillary rise) fluxes. Surface runoff Q is obtained as the excess of water that can not infiltrate the soil when maximum infiltration capacity is reached. The relationship between volumetric soil moisture content and soil suction head ψ is calculated using the model of van Genuchten (1980).”

Response to Matthew McCabe's Comments

Overview. I have been familiar with the HOLAPS effort for a number of years now and was very pleased to see this first publication put forward for consideration in GMDD. The HOLAPS project has been a leading effort in large scale flux estimation, utilizing geostationary data in a novel modeling framework to provide high spatial and temporal resolution flux estimates. Over the course of this ongoing project, there have been a number of related flux products put forward in the literature and also being distributed as products to the community. As such, it would be instructive if this HOLAPS effort could be better placed in this context, outlining some of the previous efforts and highlighting the novelty and innovation of the present contribution.

Response: We thank you very much for the encouragement and suggestions. In the following, we provide an item-by-item response to your specific comments. Your comments are written in italic black color; our responses are shown in upright font blue color.

We had already included some of the more recent efforts in global ET estimates in the manuscript (e.g. WACMOS-ET, GEWEX LandFlux), but have given further credits to these unique projects and data products in the revised version of the manuscript. Further details below.

As it stands, the manuscript only partly delivers on the stated objectives to 1) introduce and validate the fluxes at a global scale and 2) perform a thorough uncertainty assessment. The validation and assessment of flux behavior could be strengthened by additional analysis, coupled with a more detailed interpretation of results. I would encourage the authors to explore aspects of the analysis beyond the use of entire-period (or global) statistics and to really disentangle some of the variability in model performance as a function of time (and space). In terms of the uncertainty analysis, while this is certainly an interesting element of the work, it is more an assessment of the impact of forcing variability rather than a concerted effort to characterize model uncertainty or even sensitivity. It may be worth rephrasing this objective or alternatively exploring it in a deeper manner. Such efforts are a badly needed element of flux assessment in the community.

Response: The aim of the manuscript is to introduce the HOLAPS framework, and validate it against FLUXNET observations. The uncertainty analysis aims to quantify the model sensitivity to a particular forcing. We agree that the experiments conducted in the current study are more like a model sensitivity analysis, from which we diagnosis the model sensitivity to different forcing data. We have further stratified the analysis like suggested by the reviewer in the following comments. Details follow below.

Following is a summary of some comments and suggestions that might help to strengthen or re-focus the manuscript. With attention to identifying the key contributions and novelty, as well as teasing out the underlying model behavior and response through additional analysis, I believe that this work has potential to advance our understanding of global water and energy cycles. With some fine-tuning, it will make a valuable contribution to flux estimation efforts.

Response: Thanks a lot for the encouragement which is very much appreciated. We agree that the manuscript could still be sharpened and very much appreciate your constructive comments.

Introduction.

The Introduction provides a rather narrow review of the literature and misses a number of recent (last 5 years) contributions towards global flux estimation. It would be helpful to provide a more thorough review of this literature to define the context within which HOLAPS is being proposed. For example, what are the current knowledge gaps or limitations in global flux estimation; what are the key advantages and contributions of this dataset; how does HOLAPS advance or improve upon these past efforts? Basically, a clearer expression of how this work progresses upon recent efforts is needed. The fact that HOLAPS is proposing high temporal (<1 hour) and spatial (5km) flux estimates globally is clearly novel (although the authors should review related ALEXI/disALEXI research) and needs to be highlighted further. So do the advantages of such resolutions for studies that other existing global products (e.g. GEWEX Landflux) are unable to offer insights into (i.e. state the knowledge gap that HOLAPS is filling).

Response: Thank you for the suggestions. We have improved the introduction to integrate the most recent progress of the estimation of heat flux within the research community.

Page 10784, Line 19: perhaps rephrase by removing “In the last years” from the start of this sentence – just state that local scales fluxes are (predominantly) measured by EC systems – in terms of the FLUXNET collection at least.

Response: Done, the sentence has been modified to ‘At the ecosystem scale, the land-atmosphere fluxes have been mainly measured by a network of flux tower sites within the frame of FLUXNET’

Page 10785, Line 1. I’m not sure Fisher et al. 2008 is best described as a surface energy balance approach.

Response: We have added a new category: (4) the methods based on Penman-Monteith or Priestley and Taylor equations as reviewer #1 (Diego Miralles) had similar comment.

Page 10785, Line 2. Ultimately, almost all approaches can be described by point 4, since E cannot be inferred directly (hence rely on spatially variable surface parameters as proxies). If you are referring to techniques such as the triangle method or SEBAL/METRIC type approaches, perhaps better to rephrase this to make it more explicit to this family of techniques.

Response: The point 4 has been changed to ‘spatial variability methods’.

Page 10785, Line 5. You can reference the most recent GEWEX-Landflux paper:

McCabe et al. (2016) “The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data”, GMD (accepted).

Response: Thanks for the paper. It has been cited.

Page 10785, from Line 15. There are quite a few “high-resolution” data sets currently available that exceed the stated resolutions (0.25 – 2 degrees). It would be worth examining the recent literature to update these values and place the HOLAPS contribution in context.

Response: The sentence has been improved to ‘The currently existing datasets have spatial resolutions between 0.01 degrees and 2.5 degrees and are focused on hourly to monthly timescales (Vinukollu et al., 2011; Mu et al., 2007; Miralles et al., 2011; McCabe et al., 2016; Michel et al., 2016; Miralles et al., 2016).’

Page 10785, Line 29. Probably remove “exclusively”, as again, most products derive their data from remote sensing and some form of meteorological forcing, so it’s not clear what is exclusive about this.

Response: Done.

Page 10786, Line 8. Rephrase this sentence (perhaps by full-colon after question?)

Response: Done.

Model.

Page 10786, Line 19. I would advise including some clearer description of what precisely is “state-of-the-art” about the land surface model underlying HOLAPS. The coupling of the land surface scheme to a 1D mixed layer model of the PBL is a nice feature of the HOLAPS approach: is this the state-of-the-art aspect?

Response: We have rephrased the paragraph and clarified where we see new contributions by HOLAPS. The usage of a PBL scheme is not very common and a special feature of the HOLAPS approach. It is however not unique, as the reviewer points out in his next comment. We have acknowledged the previous works in that field in the manuscript.

It would be helpful to see how this coupling methodology relates to similar approaches used by researchers over the years (e.g. McNaughton and Spriggs, 1986 for an early example, but up to and including the referenced Anderson et al. 2007 and more recent works by those authors). After reviewing, I noted that much of the model description is listed in Appendix B, including the PBL component. However, to justify use of “state of the art” it would still be good to explicitly describe these distinguishing model features relative to other approaches (this may also be an element that is reflected in the Introduction).

Anderson, MC, Norman JM, Diak GR, Kustas WP and Mecikalski JR (1997). "A two- source time-integrated model for estimating surface fluxes using thermal infrared remote sensing." Remote Sensing of Environment 60(2): 195-216.

McNaughton, KG and Spriggs TW (1986). "A mixed-layer model for regional evaporation." Boundary-Layer Meteorology 34(3): 243-262.

Response: Thanks for the comment. In fact, the PBL model used within HOLAPS is a mixed 1D boundary layer model which is very similar to the approaches used by McNaughton and Spriggs (1986) or Anderson et al. (2007). In that sense, the HOLAPS approach is not “unique”, but we would argue, that coupling a land surface model with a PBL model is typically NOT done for ET estimation. We see however, that stating this as “state-of-the-art” might be a bit misleading for the reader. We have therefore rewritten the paragraph to make clear how HOLAPS builds on previous approaches.

After reading further sections, I think it is important to introduce some model description into the main-body of the text, leaving the more explicit technical details in the Appendix. An overview of the approach, a schematic of model elements and processes, together with a description of flux product development would be valuable to the reader (especially as I believe this is the first HOLAPS publication). I note that Figure B1 presents a runtime environment, but this does not describe the schematic in terms of model components and their inter-relations.

Response: Thanks for this comment. Our original idea was to put all technical details in the Annex as it is not so easy to draw a line between a high level and detailed technical description of the model used. As it is our objective to provide the full document of the HOLAPS approach in this manuscript, we had decided for presentation in Annex form. We see however the issue, that the reader might like to get a better impression of the actual functionality when reading through the main part of the manuscript. We have therefore decided to follow the reviewer’s suggestion and included a figure and more technical description on the HOLAPS components within the main text of the manuscript.

What is precipitation used for exactly: interception, presumably, but I assume also the water balance component of the model? Given the variability in available global precipitation products, wouldn’t this also be an important contribution to examining product “uncertainty”? Certainly this will impact considerably on flux-partitioning!

Response: The precipitation is used in particular for the surface water balance (including interception); see in particular section B2.4. In the present study we use a single precipitation product (TMPA) which was chosen due to its comparably high temporal resolution. The impact of using the satellite precipitation estimates instead of local precipitation data has been already quantified within the study (experiment TMPA_50) and has shown comparably low impact (instead of the radiation impact). We agree that we could have added also other precipitation datasets (e.g. also from reanalysis data) or geostationary satellite data, but this would have led even more experiments than we are dealing with in the present study. While this would be technically straightforward to do, we had actually explicitly decided to keep the present study focused on a limited set of datasets. As the impact of TMPA precipitation is small compared to the impact of the radiation data we expect that using an alternative precipitation product would actually not change too much of the results. We had therefore not included additional precipitation datasets in the study so far.

Data.

It has not been stated yet, but what is the time period over which these HOLAPS simulations were run (such information would also be helpful for the products in Table 1). I understood (perhaps incorrectly) that HOLAPS was an operational product and that long-term simulations

were available: if so, this should certainly be one feature that is highlighted in the Introduction, as a global long-term high spatial/temporal product is certainly an advance.

Response: The present paper describes only the HOLAPS framework and provided documentation and validation at the local scale. It does NOT present a data product. This seems to be a misunderstanding. The team is currently processing a global high resolution data product which is planned to be released in 2016. For each of the FLUXNET stations, different periods were used due to different data coverage. This used data for each station is traceable from the Table in Appendix C.

The temporal coverage of the used input datasets has been added to Table 1 like suggested by the reviewer.

Section 3.1.

Was there a reason why the evaluation period only covered 2003-2005? What other constraints on the quality of the Fluxnet forcing were used (i.e. was the data filtered for rainfall, freezing conditions, night-time etc..)?

Response: The period 2003-2005 was originally motivated for the validation of a gridded demonstrator dataset. As we had the FLUXNET data already preprocessed for this period it was chosen for the present study. In principle the period could be extended, dependent on the available data coverage of each FLUXNET stations. However, as the present study covers a large range of biomes and conditions we do not expect any major changes in the results if we would extend the studies to more years of the same stations. Regarding the filtering of the FLUXNET data, we applied the quality flags provided by the FLUXNET data as was already discussed in the manuscript. This eliminates already conditions with uncertain FLUXNET reference data. Nighttime values are included in the revised version of the manuscript like was discussed before. Rainfall data is not explicitly filtered. The first reviewer (Diego Miralles) also asked about the rainfall filtering. We have therefore performed a sensitivity study regarding the impact of precipitation filtering. Results indicated that an explicit rainfall filtering has no major impact on the results of this study (see details in response to reviewer #1).

Table C1 lists the Fluxnet towers, but it is not clear from the caption (or to the reader up to this point in the manuscript) what is meant by the Coverage option in this table? Fluxnet are obviously point scale locations, so it is unclear what the Coverage information refers to: perhaps state more explicitly in the caption (likewise for other Figures, were necessary).

Response: Thanks for the suggestions. The captions have been improved with the description: 'The coverage term specifies the location of each FLUXNET station. $\pm 50^\circ$ refers to that the station is within the latitudes between 50° N and 50° S, while Meteosat indicates the station is within the coverage of Meteosat'. These coverages are also illustrated in Figure 3.

Page 10790. 3.2.1. Surface radiation data. The internal consistency of the radiation components is a nice feature of this approach. The downward shortwave is available from a number of

different sources and is the focus of uncertainty assessments. However, from a flux partitioning perspective, knowledge of the surface albedo (and subsequent outgoing SW flux) is one of the most important aspects of flux-estimation. Was there any effort to assess the impact of uncertainties in this variable on flux estimation?

Response: Thanks for the question. So far we have not conducted a systematic uncertainty assessment regarding the impact of surface albedo data on the HOLAPS estimates. The only (limited) study we performed so far in this respect is that we compared results obtained using MODIS surface albedo and surface albedo from the ESA GlobAlbedo project. In fact, results were quite similar, which is not astonishing, as GlobAlbedo was using MODIS data as a prior in their retrieval approach. We agree however that it would be worth to include more permutations of different datasets which go beyond the current set of variables used in the present study. One could also think of using different vegetation datasets, soil data ...

For the present study we had decided to limit ourselves to the experiments given. In the future, we plan to use products that are going to be generated by the FP7 project QA4ECV, which is based on the heritage of GlobAlbedo and which will deliver physically consistent surface properties (albedo, faPAR, LAI) from a single retrieval approach, which would be unique and maximize consistency of these parameters within HOLAPS. As these products are expected to be available by mid of 2016 we had decided to not further extend the sensitivity studies at the moment.

Page 10791. 3.2.2. Precipitation data. I'm a little unclear on the global extent of the HOLAPS product if it is using TRMM satellite data? Are other data being used to ensure that this is a global product, or is it constrained (as the text suggests in Line 17) to the same geographic restrictions as TRMM? If so, some adjustment of the claim of a global product is probably required.

Response: Yes, the current HOLAPS version and the present study is restricted to the TRMM coverage. To make it clearer, the term quasi-global is used to replace 'global' in the revised manuscript.

Page 10792. Line 3. Surface albedo is derived from the ESA GlobAlbedo project. I'm not familiar with the mechanics of this approach, but it presumably uses some other shortwave dataset to assist in deriving the albedo. How does this affect the internal consistency of your radiation flux estimates (Page 10789) given that you are using a different shortwave product to the albedo employed in the model?

Response: We are currently using existing L2 products for albedo and vegetation properties. It is correct, that these product make their own atmospheric correction and that a maximum consistency could only be achieved if these products would not be taken. This would however imply that the entire processing would need to be based on L1 radiance data, which basically also requires using appropriate BRDF kernels etc. within HOLAPS. While this is in principle possible, we don't see the real merit at the moment. GlobAlbedo and other albedo products (e.g. MODIS) have been thoroughly validated. We don't believe that the inconsistencies in the downwelling shortwave radiation fluxes has any major impact on HOLAPS results. We expect uncertainties due to other factors like e.g. inaccurate soil information to be much more of an importance.

Page 10792. 3.2.4. Reanalysis data. The data are available every 6 hours, yet E retrievals are provided approximately every 1-hour? Some model description in the main part of the text is probably required, especially in order to describe how the 6- hourly meteorological data are used to force this required model resolution? Has there been any attempt to compare the reanalysis forcing with the tower data (e.g. some simple statistical comparison as was done in McCabe et al. 2016 for instance).

Response: Thanks for the comment. Different temporal interpolation schemes are employed in the HOLAPS framework, dependent on the variable being used. As a default, linear interpolation is used. This is however not appropriate for radiation data which has a sinusoidal diurnal cycle. What we do in practice that we invert the so-called cloud index (CI) from the radiation data and then interpolate this one linearly and used the interpolated CI then again to re-calculate the surface radiation. The advantage of this approach is that the non-linear temporal evolution of the solar radiation data is better captured. To clarify our approach in that respect, but also to better document how we treat data gaps we have added a subsection in the Annex describing the approaches in more detail.

Regarding the comparison with ERA-interim radiation, see response to next comment below.

Given that ERA Interim data are already being used, it would be interesting to run the HOLAPS data using the full-suite of available forcing (i.e. radiation) and then compare against 1) the tower data and 2) the reanalysis flux estimate. This would also allow some separation of model versus forcing uncertainty to be examined.

Response: Thanks for the suggestions. The use radiation data from ERA-Interim would give further information on the model sensitivity to radiation forcing. However, we do not expect better results compared to the CM SAF-SIS radiation forcing. Through inter-comparison and validation against ground-based measurements, Posselt et al., 2012 found that CM SAF outperforms ERA-Interim surface solar irradiance data at both daily (monthly) time scales with BIAS of 4.58 w/m² (4.4 w/m²), standard deviation (SD) of 23.73 w/m² (8.14 w/m²) for CM SAF , and BIAS of 5.64 w/m² (5.71 w/m²), SD of 39.14 w/m² (12.15 w/m²) for ERA-Interim.

This is the reason why we had not considered the use of ERA-interim radiation data in the present study.

Methods

Page 10793. Experimental set-up. The “uncertainty analysis” is more akin to the parallel forcing study undertaken by McCabe et al. 2016, whereby gridded forcing and local scale forcing are compared, along with the impact on flux estimation. The terms “sensitivity” and “uncertainty” seem to be used interchangeably: the authors may wish to define what is meant by these terms early in the manuscript (the title has uncertainty analysis, but much of the methods mentions a sensitivity analysis).

Response: Thanks for the suggestions, we have clarified these points (see also next comment) and have also given cross-reference to McCabe al. (2016).

While different “scales” of forcing are examined (this is certainly worth highlighting in the Introduction and Discussion) it is not really a true uncertainty analysis, as there is no capacity to actually attribute model sensitivity to a particular forcing. When mentioned in the Introduction, I had imagined a rigorous uncertainty assessment that sought to disentangle the issues of forcing uncertainty (distinct from variability in the type of forcing data used) on flux estimation. This is a much needed (and mostly missing) aspect of global flux estimation. It may be worth rephrasing the discussion of “uncertainty assessment” in line with what is actually undertaken here (this is not suggesting that the analysis is not useful, just that it does not really identify impacts of actual uncertainty in forcing, as opposed to impacts of changing forcing data). These ideas are stated well in the first paragraph of Section 4.2, but the analysis does not discriminate between them (which would provide a true uncertainty analysis).

Response: Thanks for the suggestions. We agree that the analysis that we did in this study is more like a sensitivity analysis and does not allow to disentangle different uncertainty components individually. We agree that such an analysis would be even more interesting to conduct, but this would require to perform a different set of experiments, with controlled noise for different forcings, thus a synthetic study using synthetic data. We think this is worthwhile doing, but see this beyond the scope of the present study, where we were more focused on the question what would be the impact of using different EO based drivers instead of local measurements on HOLAPS results.

We believe that it is important to manage the expectations of the reader here, as it seems that the reviewer had different expectations like we envisaged when reading the introduction We have therefore clarified the objectives in the revised version of the manuscript.

Results.

The section presents a rather standard statistical assessment of the HOLAPS flux retrievals, essentially reconfiguring HOLAPS forcing using a number of available sources and providing a brief summary of subsequent model response. No real understanding of the impact of uncertainty is able to be determined here: rather it is more a perturbation-simulation experiment (see the earlier points above on uncertainty). I was hoping to see a detailed sensitivity analysis, but instead we see that using different forcing results in different responses. I do not mean to be flippant here, but it is a common criticism of such papers (see reviewer comments and response to the GEWEX Landflux paper in GMDD for example) and I am sympathetic to the effort involved in product assessment being undertaken. But there are many possible causes for “error” and “uncertainty” in flux estimation: variable forcing data sets being just one. It would be good to see some of these issues expanded upon further, either here or in the discussion sections.

Response: Thanks for the suggestions. We believe that it is important to keep the major objectives of this manuscript in mind that are a) to document HOLAPS and b) provide information about its accuracies and effect of different forcing data.

We totally agree with the reviewer, that there are many reasons for deviations between observations and model results and the internal variability of a model. Some of which we can think of in this context are e.g.

Model related

- Parameterization uncertainty
- Detail of model process description
- Uncertainty of ancillary data (e.g. soil information)

Driver related

- Accuracy and representativeness of different drivers

Validation related

- Uncertainties of reference data
- Scale mismatch between observations and model results

These are all very important aspects. As this is the first HOLAPS paper, we actually tried to avoid too much of complexity and focused mainly on the impact of the forcing drivers, while keeping the model setup fixed.

We believe that it would be far beyond the scope of the present manuscript to discuss all these points, but could think of a follow up paper investigating different aspects (e.g. parameter uncertainty for HOLAPS).

We think nevertheless that it is important to discuss the need for these further investigations and make clear that the present study provides only a limited set of the above points. We have therefore adapted the manuscript to clarify these points.

Likewise, it would be great to see something other than just “whole-period” summary statistics. How do night-time versus day-time fluxes compared? What is the impact of seasonal cycles? What about cloudy-versus non-cloudy conditions (in comparing the radiative fluxes especially, it would be good to see some capacity for distinction of these between the different products)? Dry climates versus wet-climates? There are many ways to make the statistical assessment of these towers more informative than just providing such “global” evaluation and I would encourage the authors to really be creative in this aspect. One of the common criticisms of flux evaluation papers is how does it advance upon the study of XYZ et al. – so here is an opportunity to avoid that.

Response: We agree with the statements of the reviewer. We actually had also stratified our results by landcover type of the different FLUXNET stations, but had decided to not include it into the previous version of the manuscript as we had the feeling that the paper had already an appropriate length. We agree however with the reviewer that it would be good for the reader to

get more detailed insight about the accuracy of HOLAPS under different conditions. We have therefore decided to follow the reviewer comments and have added further analysis details in the revised version of the manuscript.

Given HOLAPS provides estimates of sensible-heat fluxes, it would be interesting to see how the partitioning results compare with those estimates at the tower (as well as comparisons with H itself).

Response: We agree and the other reviewers had similar comments. We have therefore added the comparison results of H to the revised version of the manuscript.

I note the negative fluxes in Figure 4 (for hourly) but not so in the corresponding evaporation values in Figure 7? Is there some filtering of model results for this period?

Response: A filtering had been done for nighttime (negative) LE values, as we assumed that the nighttime FLUXNET reference data would be more uncertain due to less turbulent conditions. We have removed this condition now and have redone the entire figures including also the nighttime values. Results became even slightly better using this approach.

I do not recall seeing any mention of ground heat flux, outside of its appearance in equation B1? How is this accounted for in the model (and how is it calculated). This would seem to be important, especially with the focus on sub-daily simulation, where the role of G is not insignificant?

Response: The HOLAPS model simulates explicitly the soil temperature evolution using a force restore approach. This soil module is explicitly coupled with the model to calculate the surface turbulent fluxes and provides information on T_s , which then affects longwave emission. The ground heat flux is estimated as the residuum of the surface energy budget ($R_N - LE - H - G = 0$) and then also used in the prognostic equation for the soil temperature model (see section B2.3). To allow for reasonable estimates of the ground heat flux, a numerical oversampling is applied, which means that the actual coupling between the surface fluxes and the soil heat fluxes is done at a temporal resolution higher than the actual model timestep.

Figure 5 and Figure 6 need to be scaled to make it easier to read and interpret. Likewise the similar Figures in the Appendices.

Response: The figures have been rescaled.

Discussion.

Page 10800, Line 10. Perhaps provide some key references and benchmarks for the statement “. . .to those obtained in other studies”.

Response: The references Ershadi et al. (2014), Michel et al. (2016) and McCabe et al. (2016) have been added.

In addition to the WACMOS (not WACHMOS) study of Michel et al. 2015, you may also wish to compare your results to the McCabe et al. 2016 GMDD study, as the analysis reflects a similar approach to that undertaken here. For a comprehensive multi-model evaluation at the tower-scale, the work of Ershadi et al (2014) may also be of interest (not just shameless plugging: they are quite pertinent to this analysis).

McCabe et al. (2016) "The GEWEX LandFlux project: evaluation of model evaporation using tower-based and globally-gridded forcing data", GMD (accepted).

Ershadi et al. (2014). "Multi-site evaluation of terrestrial evaporation models using FLUXNET data." Agricultural and Forest Meteorology 187: 46-61.

[Response: Thank you very much. The results from Ershadi et al., \(2004\) McCabe et al., 2016 are also compared to our results now.](#)

It would be good to see some further insight and discussion of some of the issues related to global flux development, and how the HOLAPS effort is addressing these, in the Discussion section. What about other sources of uncertainty? What is the (realistic) potential for an operational product and its accuracy? What are the impacts and importance of these issues on flux development and the ultimate utility of such products? What remains to be done to achieve some (stated?) objectives? Where are the major challenges and how might these be addressed? Basically, this section needs some implications and further analysis (of this and related works) to provide a useful summary of the HOLAPS contribution and a context within which the effort can be placed.

[Response: We thank you for the comment. We have further extended the discussion and conclusion section accordingly. In particular we discuss the potential and challenges of estimating high resolution land surface fluxes at the global scale. In principle there remain a lot of challenges like the reviewer knows. We stress in particular the need for appropriate fundamental climate data records for the entire GEORING \(currently not existing\), improved ancillary data, like e.g. soil maps, challenges in the validation. We provide also an outlook of potential integration into operational services, like e.g. the Copernicus land service.](#)

Conclusions.

Page 10802, Line 10. I remain a little unclear on the "consistent global water and energy fluxes" statement. I do not see how albedo (derived from ESA GlobAlbedo?) is assured to be consistent with varying vegetation cover from a separate product? Surely this will have a considerable impact on derived radiative fluxes and inevitably related water and energy flux retrievals? Perhaps this would be clearer with a more detailed modeling description section.

[Response: We agree that consistency between the used surface datasets is essential \(e.g. albedo, LAI, faPAR ...\). So far, this consistency is not achieved in existing products. Also the products that are being used within HOLAPS do not allow for this consistency so far. However, already within the GLOBALbedo project, internally consistent faPAR estimates were produced that are based on the GlobAlbedo broadband albedo data product. The approach is based on](#)

the Two-Stream Inversion package (TIP, Pinty et al., 2011³). In an ongoing EU project (QA4ECV), multidecadal datasets of albedo, faPAR and effective LAI will be produced which will maximize the consistency of these important HOLAPS parameters.

A discussion on these current limitations has been added to the revised manuscript.

³ Pinty, B., I. Andreadakis, M. Clerici, T. Kaminski, M. Taberner, M. M. Verstraete, N. Gobron, S. Plummer, and J.-L. Widlowski (2011), Exploiting the MODIS albedos with the Two-stream Inversion Package (JRC-TIP): 1. Effective leaf area index, vegetation, and soil properties, *J. Geophys. Res.*, 116, D09105, doi:10.1029/2010JD015372