REPLY TO THE REVIEWERS COMMENTS

We thank the reviewers for providing their comments to our manuscript. Their feedback comments have been very useful in further improving the manuscript. Responses to the comments are provided in detail below. We are happy to provide more details or incorporate any further suggestion in any aspect of our work, where it might be required.

REVIEWER 1:

R1C1: *The introduction could be greatly shortened. The authors give a motivation for the study of land-atmosphere interactions, as well as a history of land surface modeling, neither of which are necessary. Basically the last paragraph of the introduction (p2444) would suffice. It describes the paper's objectives and the methodology used to meet those objectives. The previous few pages are mainly superfluous.*

ANS: We shortened the introduction as requested by the reviewer above, although we have kept some reference to the motivation for the study of land-atmosphere interactions and previous works on the model to provide some background to the work undertaken.

R1C2:*The authors clearly describe the metrics they are using in section 4.3, but none of the metrics are relative, which I found an impediment to judging the quality of simulations. Specifically, what does a mean bias error of say 50 W/m^2 imply? If the mean energy flux is 500 W/m2, then perhaps that is a 'good' simulation; if the mean is 50 W/m2, it is likely not. Without a reference to the actual observed value, the reader cannot determine this. For example, in section 5.3, the authors state that because the RMSD was the lowest of the examined fluxes, and the average mean bias error is 2.84 W/m2, latent heat (LE) was well reproduced. However, an examination of figure 6 shows LE to exhibit quite a lot of scatter compared to some of the other variables. In addition, most of the values are 150 W/m2 or less, indicating that the RMSD of _40 W/m2 is quite significant. This undermines statements such as "The model showed excellent precision in reproducing daily trends of LE fluxes in most sites evaluated"*

ANS: We appreciate the reviewers point, and to address it we have estimated a few additional statistics which we have added to the updated tables of our revised manuscript. We have also added comments in the main text of our manuscript in different sections (results, discussion, conclusions) related to the new information. By including information on the daily average observed and modelled mean energy fluxes for use as reference, we have tried to remove subjective assessments of accuracy and based our interpretation of the results on more objective statements. We believe that it is now much easier for someone to appreciate the discrepancies between the model predictions and the corresponding in-situ for the different days on which the model was evaluated.

R1C3: *Some statements were confusing: p 2454 "A systematic underestimations of Rnet was evident, leading to an overall satisfactory agreement between the model predictions and in situ observations"; why is a systematic underestimation considered satisfactory?*

ANS: Indeed, the way the sentence was written was confusing to the reader. We have amended the sentence to reflect essentially that a constant or systematic and more pronounced underestimation by the model leads to poorer agreement with the in-situ data over this site. We hope is clearly explained now.

R1C4: *Another issue is the use of correlation; when looking at signals having a strong diurnal cycle, high correlations are to be expected (incidentally, are three significant digits for R2 and NASH necessary?). Are there metrics that the authors could use that would account for this effect?*

ANS: We thank the reviewer for their comment, however, we believe that the statistical metrics used in the study provide complimentary information so that people can understand the results in regards to the models' accurate simulation of the diurnal cycle and cover this information in detail within the manuscript. If the reviewer wishes for us to provide specific statistical metrics for this purpose, we kindly ask to let us know what specifically they wish to be added in the manuscript.

R1C5: *End of section 5.1, the authors note that larger errors for Oz sites occur feb-june, while the converse is true in the Ameriflux sites; the authors may wish to add that these time periods correspond to summer for each region, and are therefore consistent.*
ANS: We have added the required information to this section as per reviewer request.

R1C6: *The paper contains no figures showing actual simulated fluxes. This would helpful in understanding the characteristics of the errors, as well as the observations.*
ANS: We thank the reviewer for their comment; we have now added a figure showing an example of two days of simulated and observed fluxes for as well as information on the RMSD in relation to the percentage of observed fluxes (e.g. RMSD for LE is within 10% of the observed fluxes) in each table. We have also made reference to the trends seen in the discussion section.

R1C7: *The authors use subjective descriptions throughout the paper. For example, in the abstract: "A good to excellent agreement between the model predictions and the in situ measurements was reported,..." good/excellent are not defined, and the reader may not agree with these subjective measures. Another, pg 2454: "leading to an overall satisfactory agreement", what is the authors' definition of satisfactory? p2456: "The latter was suggestive that model predictions were in good to excellent agreement to the in situ measurements", what differentiates "good" from "excellent"? I would encourage the authors to replace such statements in the text with statements having clearly defined meaning (e.g. "within 10 per cent of observed", etc...).*
ANS: We have revised the manuscript and where possible have tried to replace the subjective descriptions with objective assessments of model prediction accuracy based on statistical trends. In this respect, for example, we have replaced "subjective descriptions" with sentences which actually utilise the new statistics we computed suggested in R1C1 to compare the model predictions against the in-situ and this was done across the manuscript, where possible.

REVIEWER 2:

This paper describes the validation of the one dimensional SimSphere model against eight Fluxnet sites in the US and Australia. The validation period covers 72 selected cloud free days during 2011. This paper is fairly well written and likely to be of interest to the land surface modelling community.

Major Comments

R2C1: *The authors should avoid the use of subjective assessments such as 2438L15 "good to excellent agreement". Readers may not consider a RMSD of 3 to 4 Kelvin for atmospheric temperature to be good, excellent or even satisfactory.*
ANS: We thank the reviewer for their comment. This is a similar, if not identical comment to comment R1C1. As mentioned above, we have revised the manuscript and where possible, have tried to replace subjective descriptions like the ones mentioned by the reviewer with more objective assessments. In this respect, for example, we have replaced "subjective descriptions" with sentences which actually utilise the new statistics we computed suggested in R1C1 to compare the model predictions against the in-situ and this was done across the manuscript, where possible

R2C2: *The authors should also provide validation statistics (e.g. RMSD) for a "zero skill" model that only uses persistence or a monthly climatology based on observations. This would aid the reader to determine the actual skill of the SimSphere model. A recent paper compares 13 Land Surface Models (LSMs) at 20 Fluxnet sites and finds that the LSMs are less skilful at predicting sensible heat flux than simple linear regression against incoming surface shortwave radiation. Best, et al. "The plumbing of land surface models: benchmarking model performance." Journal of Hydrometeorology 2015, http://journals.ametsoc.org/doi/abs/10.1175/JHM-D-14-0158.1. 3)*

ANS: With respect, we disagree with the implementation of this suggestion in our work. The aim of our study hasn't been to perform an intercomparison of SimSphere against other land surface modelling schemes, no matter how complicated or simple they could be, but to provide insights on this specific models' ability in simulating key parameters at a range of ecosystems conditions in the USA and Australia. We are aware of several other studies similar to ours – yet of much smaller set of validation days in total - already published validating SimSphere, and some of which have actually been published to very respected journals in the field. Also, we do feel strongly that the inclusion of any other model against which SimSphere predictions would be compared to would dramatically increase the paper length, and would the same time jeopardise the focal length of this paper which is clearly described.

R2C3: *The authors should discuss more deeply why they have chosen to validate SimSphere against only 8 Fluxnet sites during 2011. How many Fluxnet sites are available for validation? They should also discuss whether it would be possible to validate SimSphere over a longer time period using Fluxnet data from other years.*
ANS: We thank the reviewer for the opportunity to provide more details on this aspect of our work. Actually, we explore the potential use of several Fluxnet sites in this study before concluding to the ones we have finally included. We basically wanted initially to use sites from as many as possible commonly found different ecosystem types in the USA and Australia. At the same time we need to have sites which satisfied other criteria, such as findings sites that are as much homogeneous as possible, sites with relatively invariable topography as well as sites which during the year didn't have a lot of human interventions and sites on which all of the parameters we were interested in validating were available at the same year and days. So, after personal communications with the site PIs we concluded to the sites we have concluded to use in this study. Also, specifically for the Australian sites the sites we used were the only ones on which we could find satisfying criteria such as the availability of in-situ data acquired for all the parameters were interested in validating.

R2C4: *The Introduction can be improved by having a greater focus on the utility and usefulness of the SimSphere model. 2443L10 states that SimSphere is used to downscale SMOS soil moisture to 1km resolution. What features of SimSphere make it attractive for such applications? How does SimSphere differ from single column versions of weather and climate models?*
ANS: We have updated the introduction section and have also made it shorter, as was suggested by reviewer 1 (comment R1C1). A very important point we clarified was the fact that SimSphere is not actually used in this downscaling approach since the method used for this purpose is a variant of the "triangle" on which SimSphere is used, which was not so clear before. We also provided there as reference the overview paper on SimSphere use published not long ago by one of the co-authors on which interested readers can go and read more about the studies using the model. With regards to the last part of the reviewers' comment, we believe we provide in the introduction (paragraph 3) an explanation of what SVAT models such as SimSphere aim to simulate, which makes, we believe, the difference between SimSphere and single column weather and climate models obvious to the readers understand. If the reviewer wishes for us to provide more information on this, we kindly ask to let us know what specifically they wish to be added in the manuscript.

Minor Comments
R2C5: *2449L1: Explain the symbols G and S used in equation 1.*
ANS: 'G' is the soil surface heat flux and 'S' is the above ground heat storage in the vegetation. We have added this information to the text.

R2C6: *Table 1: RKS parameter: Please check whether Cosby et al 1984 provide estimates of saturated hydraulic conductivity or saturated thermal conductivity. What are the units of the RKS, THM and PSI parameters?*
ANS: We have added the units of the 3 parameters, and cited references where required.

R2C7: *Sections 5.1 and 5.2: Please make it clear that Rg and Rnet are at the surface.*
ANS: We have amended the text to reflect this.

REVIEWER 3:

R3C1: *The authors claimed that "SimSphere's use is rapidly expanding worldwide as both a research and educational tool alike". However, I could not find many studies using this model in the published literature, except for the papers by Petropoulos. Therefore, this is overstated and is understandable that "to our knowledge, validation studies involving direct comparisons of model predictions against in situ observations have as of now been scarce and incomprehensive."*
ANS: We appreciate the reviewers concern here and we have tried to reduce the use of descriptions on the model in superlative terms throughout the manuscript. However, we would like to note that work on the model has not only been done by one of the co-authors, but there are also other groups worldwide which use this particular model, especially when also accounting for the use of model as an educational tool in many Universities worldwide. Also, in regards to the last comment of the review on validation studies on the model, clearly the SimSphere overview paper of Petropoulos et al. (2009) refers to other validation studies done on the model in the past by other researchers independently, and furthermore, in the same work, the need to further validate key model outputs in a wide range of land use/cover types is underlined, which has been a key aspect of this study.

R3C2: *My question to the authors is that, what is the purpose of using this model instead of using other more popular models such as JULES in the U.K., CABLE in Australia, and many others in the U.S.?*
ANS: SimSphere is a relatively simple, easy to understand and freely distributed model that doesn't require a lot of computational power and doesn't require a significant number of input parameters to be initialised. It is also written in a user-friendly interface and in Java programming language, which allows relatively easy interventions to the model code when making changes to any of the model components which might be required. Also, the model architectural design includes a dynamic boundary layer modelling which is not common in SVAT models.

R3C3: *It is very overwhelmed to read so many numbers (statistics) in the Results Section. I highly suggest to list less numbers. Instead, it would be better to include some in-depth interpretation.*
ANS: When we wrote the paper we thought that the provision of the detailed results per day and also of the summarised ones per site was useful as they can offer additional insights on the model agreement with the observations for the individual days (e.g. to see in detail what are days of poor agreement and if they are related to specific parameters such as season). We do however agree with the reviewer that the layout and formatting of the tables may make the inclusion of a large number of statistics overwhelming. To improve the visual appearance of those numbers, we have changed the table layouts which we hope provide more clarity and readability, highlighting the important statistics without losing any of the detail.

R3C4: *Figure 1 was published in several papers by the authors already. Will there be a copyright issue to publish it again? Is it necessary to include it here?*
ANS: We agree with the reviewer that the figure may not be necessary and have removed it from the manuscript.

R3C5: *If I understand correctly, the model was initialized with observed values, which is problematic. This might be the reason that the model shows high performance skills during the several months of simulation. If the model was run for additional years, the influence of initialization will be small and hence the model is expected to show poor skill.*
ANS: We are not sure why the reviewer raises this comment; the approach we followed is one that has been used in other validation studies of SimSphere, but also in similar validation experiments done on other SVAT models. Also, as we state in the manuscript methodology

section, we used primarily observed values but we also used other sources when available such as information from PI, literature or in some rare occasions indeed educated guesses.

R3C6: *Why does the model need to simulate incoming solar radiation and air temperature? Since these two variables are commonly measured, why can you treat them as model inputs?*
ANS: SimSphere's architectural design (briefly described in section 2 of paper) requires the model to compute solar radiation and Tair (at different heights). Both of those parameters are not provided as inputs to the model, but are computed; basically Tair is computed using the sounding profile data which are given as input to the model whereas the shortwave incoming radiation is computed from other inputs provided to the model, such as the geographical location, slope and aspect. This is the way the model is built and we cannot make those parameters inputs to the model.

R3C7: *This study evaluated the model for only 72 days. This is definitely not enough. It has to be at least several years.*
ANS: We appreciate the reviewers concern expressed in the comment above. We would like to underline here that only days of complete measured data were included in the EBC estimations, days with gapfilled data were rejected. Furthermore, in regards to the other comment of the reviewer related to: "For the stated aims of the manuscript (an in-depth validation of the model), simulations should be undertaken for all periods (day-time, night-time, clear skies, cloudy skies, precipitation, all seasons, etc.) with valid observational data." we also agree that this is of course a valid criticism although though unfortunately unavoidable since reliable validation data under all conditions would be unavailable. Eddy Covariance data (LE and H components) used as observational validation data are subject to strict assumptions such as sufficient turbulent mixing, appropriate atmospheric thermal structure etc. Particularly for open path sensors scattering of infra-red signals by water droplets precludes measurements during precipitation events being retained for example and nighttime data are often plagued by insufficient mixing due to low friction velocities. Strict quality control typically rejects data collected under unfavourable conditions resulting in no data being available for model validation during these times. Continuous long term Eddy Covariance datasets that extend across these conditions do so only by being themselves modelled (gapfilled) from higher quality measurements. It is these higher quality measurements that have been used in the validations in this paper with short term assessments of energy balance closure being used to determine the suitability of these validation days. It is only by using these data that uncertainties in the observation data can be minimised and validations can be judged. Finally, in overall, many of the previously validation exercises on SimSphere which we have cited in our manuscript herein (but also in other similar studies to ours implemented to other models) have used "selected" days only to validate the model performance (e.g. days of stable atmospheric condition, non-convective conditions etc) and our practice here is in line to those studies as well and we do believe it is only fair to the model to validate it under conditions which it is able to simulate or take into consideration as otherwise cannot be expecting the model to replicate a reality which hasn't been taken into consideration into its architectural design in the first place.

R3C8: *Table 3 to Table 8: Why did you calculate the statistics for each day? Is this necessary? These tables are difficult to read. I suggest the authors find a better way to show these results.*
ANS: This comment is similar to comment R3C3, so we would kindly refer the reviewer to the reply we have already provided in that comment. In short, we have attempted to change the layout of our tables and make them easier to read.

R3C9: *Figure 4 to Figure 9: These figures can be combined into just one figure.*
ANS: We agree with the reviewers' suggestion and we have now combined all our figures into a single one and have used individual letters to refer to the individual descriptions of each figure within.

REVIEWER 4 (ANDY PITMAN):

SC1: *My interpretation of this paper suggests it has screened cloudy days in the analysis - and it has limited the analysis to 72 data in 2011. If I am right in this interpretation the conclusion "The model presents itself as an important tool to acquire regional specific data, essential for numerous hydrological modelling, agriculture and water resource management applications" seems flawed. Each of these applications need to accommodate inter-annual and seasonal variability. If there is anywhere in the world a model needs to account for inter-annual variability it is Australia. It does not seem clear to me that evaluation over 72 days provides any real capacity to make robust conclusions. The model also has to be able to cope with cloudy days (!) since in any application for hydrological modelling, agriculture or water resource management one might presume that periods of cloud cover and potential rainfall are rather profoundly important to the resulting applicability of the model. I may well have misunderstood of course but I suggest these issues need to be addressed.*

ANS: We appreciate the reviewers concern expressed via this comment. We have changed the text and have rectified now we believe this issue. As already mentioned when replying to the comment from another reviewer, we would like to underline here as well that only days of complete measured data were included in the EBC estimations, days with gapfilled data were rejected. Furthermore, in regards to the other comment of the reviewer related to: "For the stated aims of the manuscript (an in-depth validation of the model), simulations should be undertaken for all periods (day-time, night-time, clear skies, cloudy skies, precipitation, all seasons, etc.) with valid observational data." we also agree that this is of course a valid criticism although unfortunately unavoidable since reliable validation data under all conditions would be unavailable. Eddy Covariance data (LE and H components) used as observational validation data are subject to strict assumptions such as sufficient turbulent mixing, appropriate atmospheric thermal structure etc. Particularly for open path sensors scattering of infra-red signals by water droplets precludes measurements during precipitation events being retained for example, and nighttime data are often plagued by insufficient mixing due to low friction velocities. Strict quality control typically rejects data collected under unfavourable conditions resulting in no data being available for model validation during these times. Continuous long term Eddy Covariance datasets that extend across these conditions do so only by being themselves modelled (gapfilled) from higher quality measurements. It is these higher quality measurements that have been used in the validations in this paper with short term assessments of energy balance closure being used to determine the suitability of these validation days. It is only by using these data that uncertainties in the observation data can be minimised and validations can be judged. Finally, in overall, many of the previously validation exercises on SimSphere which we have cited in our manuscript herein (but also in other similar studies to ours implemented to other models) have used "selected" days only to validate the model performance (e.g. days of stable atmospheric condition, non-convective conditions etc) and our practice here is in line to those studies as well and we do believe it is only fair to the model to validate it under conditions which it is able to simulate or take into consideration as otherwise cannot be expecting the model to replicate a reality which hasn't been taken into consideration into its architectural design in the first place.

SC2: *There is also a lot of activity in Australia linked to model evaluations of land surface models not cited here - perhaps see papers by Haverd or Abramowitz or Ying Ping Wang .... all in the international literature.*

ANS: As this was not a comparison study of land surface models we have kept the reference to other models to a minimum and concentrated on validating the results only in relation to previous validation studies of SimSphere. However, we have added the suggested references in our revised manuscript at the introduction section which have also modified in an attempt to address comments we received by 2 of the other reviewers related to this section of our manuscript.