Reviewer's comments No. 1

We thank the referee for providing a review of the manuscript and agree that the suggested changes and clarification improve it. We have made the changes outlined below in the revised manuscript. Each item starts with the reviewer's comment. The page and line numbers of where changes have been made to the updated manuscript are included at the end of each reply.

1. **Page 5343, line 8: As done for soil processes, could you specify by which processes CO2 fertilisation leads to an increase in plant CO2 uptake? Photosynthesis, LAI, etc. increase?**
   The processes affected by CO2 fertilisation have been added (Page 5343, lines 8-14).

2. **Page 5346, lines 1-4: Is JULES considering only natural vegetation types or also agricultural ones? In order to have a better idea of which PFTs are considered in JULES, could you please list their names?**
   JULES version 3.0 (version used in this study) has only natural vegetation types and does not have agriculture PFTs (such as croplands). Nine surface types are used in JULES (version 3.0): five Plant Functional Types (broadleaf trees, needleleaf trees, C3 (temperate) grass, C4 (tropical) grass and shrubs) and four non-vegetation types (urban, inland water, bare soil and land-ice) (Page 5346, lines 9-13).

3. **Page 5346, lines 10-15: Is the LAI in JULES calculated for each canopy layer or is it based on a big-leaf approach? Is there a maximum LAI prescribed for each PFT that could affect the value calculated eventually, and therefore the model data comparison?**
   LAI in JULES is calculated for each canopy layer and is updated once per day by multiplying the annual maximum LAI by a scaling factor, which is calculated by using temperature-dependent leaf turnover rates (Page 5346, line 27 - Page 5347, line 5).

4. **Page 5346, line 24: Could you specify which fluxes are directly, or less directly affected by an error in GPP estimates?**
   Errors in the calculation of GPP can lead to errors in simulated biomass, Net Ecosystem Exchange (NEE) and latent and sensible heat fluxes (Page 5347, lines 17-22).

5. **Page 5353, section 3.1: Site values for parameters such as Vcmax or maximum LAI were adjusted to global or local values, depending on simulations performed, for this evaluation. Do the authors have an idea of the model performance when used in its standard configuration, including original values for parameters?**
   Yes, site values for parameters such as Vcmax or maximum LAI were adjusted to global or local values, depending on simulations performed. The original values for the parameters, i.e. in its standard configuration, are the global values (Page 5354, section 3.1).

6. **Page 5357, section 3.4: Regarding satellite data derived LAI, could the authors give a quick overview of the different datasets available, and known discrepencies?**
   A brief overview of the different satellite LAI datasets available and the reason for choosing the MODIS dataset has been provided (Page 5348, line 25 - Page 5349, line 3).

Reviewer's comments No. 2

We thank the referee for providing a review of the manuscript and agree that the suggested changes and clarification improve it. We have made the changes outlined below in the revised manuscript. Each item starts with the reviewer's comment. The page and line numbers of where changes have been made to the updated manuscript are included at the end of each reply.

Major concerns:

1. **Given the aim of the study (to evaluate the relative performance of the model when driven with different meteorological and vegetation characteristic datasets) I find the lack of rigorous statistical analysis of the model output data of concern. Rather than performing significance tests, for example, the authors employ phrases such as "similar", "significantly different", "relatively well" when describing model performance. Given the current efforts within the modelling community to evaluate, constrain and improve the uncertainties associated with model performance and model projections, this is in my view not acceptable. Furthermore, the simulations are carried out for a single year with the result that the errors and biases shown for the different datasets cannot be put into context against the inter-annual variability in the datasets themselves and the JULES model. Both of these concerns should be adequately addressed before the manuscript is accepted for publication in GMD.**
   The phrases "similar", "significantly different" and "relatively well" were used when describing model performance in the summary paragraph at the end of each results subsection in order to promote readability. RMSE and Bias are standard metrics used in the land surface modelling community to compare model output to observations. The primary reasons for carrying out simulations for single years was that this study was a continuation of an evaluation of JULES by Blyth et al., (2011) and observations, such as leaf area index and Vcmax were only available for certain years. Single-year model simulations were performed using the best available information for that year and data from different years were not used together. The model's ability to capture interannual variability is discussed in sub-section 3.1 (Effect of local data on simulated GPP, Page 5354-5355).

2. **Given the wide audience of GMD, and the relevance of this paper to such a large section of the Geoscience community I feel that the manuscript would benefit from a thorough overhaul of the explanation of the model set-up and simulations, as well as in the presentation and discussion of results. The language and phraseology used is inconsistent and at times confusing, and the selection of results presented in the text not sufficiently comprehensive. I would advise the authors to give the manuscript to non-specialist colleagues to read through and highlight sections that are overly technical or insufficiently well-explained.**
   The manuscript was given to non-specialist colleagues (with no knowledge of land surface modelling) who provided feedback on the readability of the manuscript and the relevant changes were made. The language and phraseology used in the original manuscript has been changed where appropriate.

3. **Introduction, Page 5345 and throughout the remainder of the text: It is not**

always clear precisely what is meant when the authors refer to model parameters, meteorological data and atmospheric conditions. This confusion is due in part to the authors' failure to present a clear list of the altered data, and their inconsistent use of words such as local, site specific, meteorological and atmospheric. The confusion is compounded with an inconsistent use of acronyms, particularly with regard to the global meteorological datasets which are at times referred to by their full names and at others by an abbreviation of this, and yet others by the name of the simulation in which they were applied. In addition, acronyms are often used earlier in the text than they are explained, presumably a result of the order in which the sections of the manuscript were written.

A table (Table 1, Page 5377) listing the altered data (model parameters and meteorological variables) has now been provided. "local" and "site-specific" have the same meaning and "local" is referred to only in the manuscript. "meteorological" and "atmospheric" have the same meaning and "meteorological" is referred to only in the manuscript. We have moved sub-section outlining the experiments (it is now sub-section 2.4) to near the end of section 2 (Methods and model). This makes more sense since the datasets (and associated acronyms) have been explained before they are referred to.

4. **Page 5345, lines 2-4: Please make explicit at this stage precisely which model parameters are altered in this series of experiments, as well as the methods by which the values of these parameters were derived for the "local" datasets (i.e. from a model, from observations, or by ecosystem or species type).**
   This paragraph has been re-written. As mentioned in the previous comment, a table (Table 1, Page 5377) listing which model parameters and meteorological variables has been included in the manuscript. The definition of local and global and the satellite dataset used has been provided.

5. **Page 5345, lines 17-18: This fourth science question does not match the stated motivation for the study. While it certainly fits with a potential source of "global" data, LAI is not the only phonological/physiological satellite retrieval available. Why select this parameter to test rather than any other?**
   Yes, it is correct that LAI is not the only physiological satellite retrieval available. However, in JULES, LAI is an important parameter when calculating GPP and the JULES' phenology model is important for modelling the seasonal cycle of LAI. This fourth science question concerns testing the JULES' phenology model i.e. how do predictions of GPP compare when using the default phenology model to compute LAI to those using satellite LAI. This has been clarified in the manuscript by updating the abstract and the outline of experiments (Daily satellite phenology). Reasons for using the MODIS data are included in sub-section 2.3 (Data).

Minor and technical comments:

Abstract

The abstract is too long, too detailed and too specialised. It contains a number of jargon words and unexplained acronyms. It should be concise and aimed at an audience without specialist knowledge of the JULES model and/or global datasets. The authors should also make explicit the baseline against which they compare their model estimates, how this baseline is derived and the uncertainties of "observed" GPP.

The abstract was given to a climate researcher (with no specialist knowledge of land surface modelling) to read and feedback was provided, which was then used to re-write the abstract (Page 5342). The baseline against which the model estimates are compared to has been included in the abstract (Page 5342). The source, derivation and the uncertainties of "observed" GPP has been added as a new section (2.3.2 Observational data, Page 5350).

1 Introduction

1. **Page 5343, line 2: Split into two sentences: "1998). Although"**
   Done (Page 5342, line 23).

2. **Page 5343, lines 7-10: Please make it clearer which of these processes are sources and which are sinks.**
   A list of the sources and sinks has been added (Page 5342, line 25 - Page 5343, line 7).

3. **Page 5343, lines 13-17: Please split this sentence which is over-long and rather impenetrable. Perhaps: "2006). One of it" and "anthropogenic CO2. The magnitude of this"**
   Done (Page 5343, lines 18-22).

4. **Page 5343, line 28: Please outline briefly what a "bucket" model is in the context of land surface modeling.**
   Done (Page 5344, lines 4-8).

5. **Page 5344, line 8: Is Sellers et al, 1997 really the best and most recent reference for the potential of the current generation of land surface models?**
   This brief summary of land surface modelling has been re-written (Page 5344, lines 4-15).

6. **Page 5345, lines 2-4: Inconsistent tenses; replace "compare" with "compared".**
   Done (Page 5345, line 7).

2 Methods and model

2.1 Model description

7. **Page 5345, line 21: Please insert "has" between "and" and "evolved".**
   This line has been re-written (Page 5346, lines 3-7).

8. **Page 5346, line 1-4: Please explain the tiles system that JULES uses and how this translates to single-point (site) modelling, such as that conducted in this study.**
   An explanation of the tiling system used by JULES and how this translates to single-point modelling has been provided (Page 5346, lines 9-15).

9. **Page 5346, lines 1-4: Presumably this version of the JULES model does not include JULESCROP, and it appears that none of the PFTs is cropland. Why have the authors chosen to include a cropland site in this evaluation?**
   This version of the JULES model (version 3.0) does not include JULES-Crop. Since this study is an evaluation of the JULES model, we decided to include a cropland site to test if better model parameter and meteorological data could improve GPP simulations.

10. **Page 5346, line 9: Please insert " " between "component" and "("**
Done (Page 5346, line 20).

2.2 Experimental design

11. **Page 5346, lines 1-2: As requested above, please explain clearly how the JULES model described in section 2.1 is applied at a single point, in particular with regard to the tile system used to assign land cover in JULES and the vertical structure. How do the global and local operational versions of JULES differ? How are local parameters and variables determined for a single point? How are model parameters from global datasets applied to the 10 canopy layers?**
An explanation of the tiling system used by JULES and how this translates to single-point modelling has been provided (Page 5346, lines 9-15). Differences between the local (standalone) and global operational versions are provided on Page 5347, lines 9-13. An explanation of where the local parameter datasets have been obtained from is found on Page 5348, lines 17-19. The model parameters from global datasets are not applied to the 10 canopy layers. We should have been more clear on this. In the model description section (2.1 Model description), we state that GPP is calculated first at leaf-level and then scaled up to canopy-level using LAI for each of the 10 canopy layers.

12. **Page 5346, line 28; Page 5347, line 2: Have these sites been used in previous model evaluations or benchmarking studies? If so, please give appropriate references here.**
References to previous model evaluations/benchmarking studies have been added (Page 5347, line 23).

13. **Page 5347, lines 2-5: Does this mean that for each site a single year was chosen, but that the chosen year differs between sites?**
Yes, for each site a single year was chosen. The reason for this is that local site observations, such as annual maximum LAI, may only be available for 1 year. Running the model with local data for a specific year meant that the model had the best available local data for that year. We did not want to mix and match data from various years at a study site.

14. **Page 5347, line 4: "gapfilled" should read "gap-filled"**
Done (Page 5347, line 24).

15. **Page 5347, line 13 - Page 5348, line 12: While I can appreciate the sense behind outlining specific simulations within the section describing the design of the experiment, this section would more properly belong after the datasets have been introduced (i.e. after section 2.3) and would certainly make it far easier for the reader to understand the differences between the simulations.**
The brief outline of specific simulations has been moved to a section of its own (it is now sub-section 2.4) after the datasets have been introduced (Pages 5352, 5353).

16. **Page 5347, line 13 - Page 5347, line 17: Although the HadGEM model was mentioned in the Abstract it has not been alluded to since, and it is not clear how this model is connected to any of the data being used in these simulations. The acronyms of the simulations have likewise not been explained. Here, it sounds as if the model output of the different simulations is being compared to other model output data, whereas the Abstract implied that model output**

was compared against observational data.
Section 2.3 has been updated with how the HadGEM model is connected to the global data used in these simulations (Page 5348, lines 19-22). The acronyms of the simulations are now explained in sub-section 2.4 (Outline of experiments) (Page 5352, lines 10-12) and Table 3 (Page 5379). The model output of the different simulations (local-F, global-WEIG, global-WEIC and global-P) is being compared to observational data, not other model output (Page 5352, line 16 and Page 5353 lines 2-4).

17. **Page 5347, line 17: This is the first mention of vegetation competition. Please explain what this means in the context of the JULES model (perhaps in section 2.1) and how turning it on/off may alter model performance.**
Section 2.1 now includes a brief description of the role of vegetation competition in this study and how turning it on/off may affect model performance (Page 5346, lines 24-26).

18. **Page 5348, lines 2-4: Should this not have been the first step? How do the acronyms employed here relate to those for the simulations as used on p5347? What about the HadGEM default values? Are these only used for the so-called model parameter data?**
No, the model simulations using the various datasets (local and global) were compared to the observational datasets first. Once this had been done, we then wanted to know which input data (model parameters or meteorological data) played a more important role when determining GPP fluxes. When it was discovered that the meteorological data was the most important factor, we decided to compare the various meteorological forcing datasets to the local observations (FLUXNET) in order to find out if the improved results were a result of biases in the meteorological data. This section has now been placed after the section describing the datasets and the acronyms should make more sense now (Page 5352, 5353). The HadGEM default values are used for the model parameter data (Page 5348, lines 19-23).

19. **Page 5348, lines 6-12: As both simulations are described as using MODIS LAI data in some form, please could the authors make it explicitly clear how the TRIFFID and phenology modules in JULES use MODIS data, and how this differs from how it is used when they are turned off.**
A more clear explanation of how JULES uses MODIS data and what happens when the phenology and TRIFFID modules have been switched off has been provided (Page 5353, lines 12-21).

2.3 Data

Please move this section so that the acronyms are explained and the datasets described before the model simulations are outlined.
Done (Page 5348).

Please explain how the different units used to describe spatial resolution compare. Roughly, what land area (i.e. km x km) would 1 degree x 1 degree data map to in the tropics and in temperate regions.
The land area for which the spatial resolution of each dataset (WFDEI and PRINCETON) equates to has been provided (Page 5349, lines 22-26 and Page 5350, lines 2-5).

20. **Page 5348, lines 15-16: Previous studies have shown that the resolution (both spatial and temporal) of meteorological data can affect the output of land surface and atmospheric chemistry models (e.g. Ito et al., 2009; Pugh et al., 2013; Gego et al., 2005; Ashworth et al., 2010; Colette et al., 2013) and may**

even introduce a systematic bias.
References to previous studies of how the spatial and temporal resolution of the meteorological data can affect output from land surface and atmospheric chemistry models has been included (Page 5350, lines 5-7).

21. **Page 5349, line 2, and throughout: Please use either site-specific or local, or define clearly what is meant by each term if they are in fact different.**
Site-specific and local have the same meaning. In the introduction, the definition of site-specific (local) and global is given (Page 5345, lines 8-12). From this definition onwards, local is used.

22. **Page 5350, lines 3-6: The authors describe PRINCETON as a 50 yr dataset, but then state that it consists of data for the 1948-2008 period, which is 60 years. Also please state explicitly that the data covers the full globe and not just land points (if indeed that is the case).**
The time period of the PRINCETON dataset has been corrected and further information on the dataset has been included (Page 5350, line 27).

23. **Page 5350, line 12: Please explain more clearly how the mapping of the 17 land categories in the IGBP landcover scheme was mapped to the 5 PFTS and 4 other land cover types in JULES.**
A more detailed description of how the 17 land categories in the IGBP landcover scheme are mapped to the 5 vegetation and 4 non-vegetation land cover types in JULES has been included (Page 5350, line 24 - Page 5351, line 3 and Appendix A).

24. **Page 5351, line 15: Section 2.2.4 states that JULES is driven with daily MODIS LAI data. Here it is stated that the data is actually an 8-day composite. Please clarify this apparent discrepancy.**
Yes, JULES was driven with daily MODIS LAI data. The MODIS Land Product Subsets contain LAI data at 8-day composite intervals. At the end of the paragraph, it is stated that each time-series of 8 day composite values was linearly interpolated to obtain a daily LAI time-series (Page 5352, line 7-8).

25. **Page 5351, lines 16-17: Listing the pixel numbers used to create a 3 x 3 grid-box centred on the flux tower is unnecessary.**
The list of pixel numbers has been removed (Page 5351, line 28).

3 Results

This section is poorly presented and poorly explained. There appears to be substantial overlap between sections and the terminology used to describe the different simulations and different datasets or parameter sets used is inconsistent. Furthermore, only a small sample of the results is presented in the text and inconsistently in terms of whether absolute or percentage values are stated. Please add a further table giving both absolute and percentage differences for each simulation, and state clearly the absolute values observed in the baseline case. As noted previously, it is not sufficiently clear what is being taken as the baseline case and how the value of GPP has been determined for this case.
A table (Table 6, Page 5382) giving both absolute and percentage differences between the model and observed (FLUXNET) total annual GPP has been provided. Model total

annual GPP has been compared to the observed (FLUXNET) values (baseline case).

3.1 Global vs. local fluxes

Please alter the title of this sub-section to better reflect the different simulations that are being presented here. Given the titles of the other sub-sections, this section is presumably included to state the differences between simulations using local (site-specific) model parameters and those derived from global datasets. However this is neither clear from the title, nor the text as this sub-section then drifts off into a discussion of the effects of different meteorology, which surely should be covered in sub-section 3.2.
The title of this sub-section has been changed to "Effect of local data on simulated GPP" (Page 5354, line 17). This sub-section compares the results from model simulations using local data to observations of GPP from the FLUXNET network. The discussion of the effects of the meteorology data has now been moved to sub-section 3.2 (Page 5356, lines 16-23).

26. **Page 5352, line 17: As noted previously, the source, derivation and uncertainties of the observations should be explicitly described somewhere in the paper.**
The source, derivation and uncertainties of the observations has been added as a new section (2.3.2 Observational data, Page 5350).

27. **Page 5352, line 17: The local data being referred to here should be clearly and explicitly listed somewhere in the paper, i.e. precisely which model parameters are altered between the different simulations when local data is replaced with global data.**
A table (Table 1, Page 5377) listing the model parameters (vegetation and soil) and meteorological data, which are altered between simulations when using either global or local data, has been included and is referenced in the Introduction (Page 5345, lines 6-7) and Data section (Page 5348, line 12 and 16).

28. **Page 5352, lines 17-23: Please put this into context (ideally by presenting the data in a table as suggested previously). What percentage of the absolute GPP do the RMSE and bias represent?**
A table (Table 6, Page 5382) giving both absolute and percentage differences between the model and observed (FLUXNET) total annual GPP has been provided as suggested previously. In addition to RMSE and bias, the absolute difference is included when comparing the results from model simulations to observations (sub-section 3.1).

29. **Page 5353, line 5: Please make clear to non-specialists, what a sclerophyll forest is.**
The definition of a sclerophyll forest has been added (Page 5355, lines 9-10).

30. **Page 5353, lines 8-15: Please ensure that the results from each of the sites, as well as the average across all sites, are presented either in the text or in tabular form.**
Table 6 (Page 5382), which provides the observed (FLUXNET) total annual GPP and the absolute and percentage differences between the model and these observed values, now includes totals. Note that the totals for the model simulations is the total of the differences (not absolute differences) between the observed and model total annual GPP.

31. **Page 5353, line 9; Page 5354, line 9: Surely, the results of the simulations comparing the use of global meteorological data belongs in sub-section 3.2 (titled "Global meteorological data")?**
The results of the simulations comparing the use of global meteorological data has been moved to sub-section 3.2 as suggested. Sub-section 3.2 has been renamed as "Effect of global data on simulated GPP".

3.2 Global meteorological data

Throughout - As noted above, please move the results and discussions of the use of global meteorological data from sub-section 3.1 to this sub-section. Again, please could the authors make clear what the baseline is against which each of these simulations is being evaluated.
This sub-section has been renamed as "Effect of global data on simulated GPP". The results and discussions of the use of global meteorological data has been moved from sub-section 3.1 to this sub-section (Page 5356). Simulations using global data (global-WEIG, global-WEIC and global-P) are compared to observed GPP from the FLUXNET network (Page 5355, line 28).

32. **Page 5354, line 11: Should this read "global parameters" rather than "global data"?**
No, this should read "global data" (Page 5356, line 8). It refers to the previous sub-section (3.1) where simulations used either local and the previous paragraphs where simulations used global data. Results from model simulations using global data (model parameters and meteorological data) and those local parameters with global meteorological data are now described in the same sub-section (3.2).

33. **Page 5354, lines 24-25: Are these additional errors that are introduced? And does this mean that the model performance has deteriorated?**
No, these are not additional errors that are introduced. At the sites mentioned (Tharandt, Kaamanen and Hyytiala), the use of global meteorological data has not introduced large negative biases into GPP predictions (Page 5356, lines 20-23).

34. **Page 5354, line 25: Please define and quantify "small".**
This line has been re-written (Page 5356, lines 20-23).

3.3 Global vs. local meteorological data

Surely an assessment of the closeness of the global data (both meteorological and vegetation characteristics) to the site-specific data should have been the first analysis performed and presented. Furthermore, I would expect to see a rigorous statistical analysis of the goodness of fit between the site-specific and global datasets. As noted previously, the authors need to be far more rigorous in their terminology with regard to local vs. site-specific, meteorological vs. atmospheric, etc.
As mentioned previously we wanted the first analysis to be of the model simulations using local and global data. Then we wanted to try to find out what was causing the differences i.e. the meteorological or parameter datasets. 'Local' and 'site-specific' have the same meaning, as do 'meteorological' and 'atmospheric' datasets. 'local' and 'meteorological' are used throughout the paper.

35. **Page 5355, lines 5-6: Please could the authors explain clearly what the difference is between global atmospheric forcing data and global meteorological data in the context of these simulations.**
There is no difference between global atmospheric forcing data and global meteorological data in the context of these simulations. The term "atmospheric forcing data" has been changed to "meteorological data" throughout the paper (Page 5357, lines 14-15).

36. **Page 5355, lines 11-13: By "converted to dimensionless quantities by dividing the daily time series by the annual mean" do the authors mean that the data was normalised against the annual mean for each site?**
Yes, the data was normalised against the annual mean for each site (Page 5357, line 19-20).

37. **Page 5355, line 15: Please quantify and define "best" in this context.**
We state that the WFDEI dataset compares better than PRINCETON to the FLUXNET data due to lower RMSEs and biases, which have been quantified in paragraph 2 and 3 of this sub-section (Page 5357, lines 21-23).

38. **Page 5356, lines 3-8: Please specify the relative differences in addition to the absolute changes.**
RMSE and average bias were used to quantify differences between global and local meteorological data.

39. **Page 5356, line 17: Please could the authors explain how this attribution was performed given that the simulations they have described suggest that all the variables of a dataset were altered at the same time. If they have performed further simulations or sensitivity studies these should also be described, and the results clearly presented.**
On re-reading this line, it was realised that what was written meant that all the variables of a dataset were altered at the same time. However, we meant swapping only the local meteorological data (FLUXNET) with the global meteorological data (PRINCETON) for these model simulations. This line has now been updated (Page 5358, lines 22-28).

40. **Page 5356, line 28-Page 5357, line 1: Please define and quantify the terms "better" and "quite well".**
We use the qualitative terms "better" and "quite well" when comparing the WFDEI dataset to PRINCETON and when describing which variables of the meteorological data compare best to the local values. The preceding paragraphs provide evidence for their use. At the end of the results and discussions section, a summary of each sub-section was provided to improve readability and so this paragraph does not contain quantifiable results (Page 5359, lines 7-12).

41. **Page 5357, line 5: The statement that the improvement in model performance at the tropical sites was due to biases in the meteorological data appears inconsistent with the authors' subsequent conclusion that better parameterisation of tropical sites is required in order to improve the representation of the carbon cycle at tropical biomes by JULES.**
The improvement in model performance at the tropical sites is due to biases in the meteorological data and is a case of the right answer for the wrong reason. In order to improve model predictions at tropical sites, more PFTs (e.g. deciduous evergreen Broadleaf) will be need to introduced with the appropriate parameter values. In addition to this, the phenology model in JULES will need to modified to deal with tropical PFTs and the

inclusion of tropical processes into the model.

3.4 Forcing JULES with daily satellite phenology

Throughout-Please could the authors make clear how the JULES phenology module currently calculates the daily values of LAI required by the model. Are they simply computed as fractions of the annual LAI?
Yes, daily values of LAI are computed as fractions of the annual maximum LAI. The model description sub-section (2.1) has been updated with more detail on JULES' phenology model (Page 5346, line 27 - Page 5347, line 6).

42. **Page 5357, lines 15,18 and 22: Please quantify "small".**
"small" has been quantified (Page 5359, line 22-24).

43. **Page 5358, line 3: Please remove the first "," so that it reads "Of the 7 sites where JULES' performance improved using MODIS data,"**
Done (Page 5360, line 9).

44. **Page 5358, line 9: There seem to be too many ")" in this sentence.**
This line has been re-written (Page 5360, lines 13-15).

45. **Page 5358, lines 18-22: This appears to contradict the authors' discussions and conclusions that for some of the sites the MODIS LAI data is a poor match to that observed locally.**
We found that the MODIS LAI values tended to be a closer match to the local values than the global ones. The model simulations using the global data are using the global parameter values, which is a reason for the poor model performance. However, at some sites, the MODIS LAI could be noisy (large day), but over the course of the year, the total annual GPP may be similar to the simulations using local data (Page 5360, lines 22-25).

46. **Page 5358, line 26: Please quantify "small".**
Done (Page 5361, lines 1-2).

47. **Page 5358, line 29: Please quantify "equally well".**
Done (Page 5361, lines 5).

48. **Page 5359, line 1: Please justify the use of the word "significant" by performing rigorous statistical analyses to the results of this study.**
This line has been re-written (Page 5361, lines 5-7).

4 Discussions

49. **Page 5359, line 6 and throughout: Please define and quantify what "very well" means in the context of model performance, preferably in a statistically rigorous way as outlined in Major concerns above. As previously noted, quoting absolute values of RMSE and bias is of limited use in comparing the skill of the model across different sites, and showing the results graphically only can be misleading as the impression is very much dependent on the scales of the axes.**
This section has been restructured. The first sub-section (4.1) contains the discussion for sub-sections 3.1 and 3.2. This has been done because it was decided that discussing the local and global model results together made more sense. RMSE and bias were plotted

since it is easier to see trends in the results as opposed to having the results displayed in a table. Readers are more likely to examine a figure than a table.

4.1 How well does JULES perform...

50. **Page 5359, line 6-8: Please specify which panel(s) of Fig. 2 show this.**
Done (Page 5361, lines 11-12).

51. **Page 5359, lines 8-9: Please specify which figure and which panel(s) show this.**
Done (Page 5361, lines 16-18).

52. **Page 5359, line 9: The Mediterranean site ES also appears to be an exception.**
This line has been updated to include El Saler (Page 5361, lines 17).

53. **Page 5359, lines 10-16: Please make clear which figures and panels show this.**
The relevant figure in Blyth et al. (2011) has been referred to (Page 5361, line 23 - Page 5362, line 3).

54. **Page 5359, line 20: As noted previously, these biases and RMSEs need to be put into context.**
This line has been re-written (Page 5362, lines 3-6).

55. **Page 5359, line 26 - Page 5360, line 1: Please re-phrase to remove the split infinitive, e.g "tend to match more closely the local..."**
This line has been re-written (Page 5362, lines 12-13).

56. **Page 5360, lines 7-8: Please expand on this statement. What exactly do the authors mean when they refer to model error?**
We performed a temperature sensitivity study at Tumbarumba using local parameter and meteorological datasets (local-F). We increased the winter and spring surface air temperatures (May-October) of the FLUXNET data by increments of 1 degree Celsius and re-ran the model each time. Improvements in the magnitude of the model seasonal cycle were observed, but only at high surface air temperatures (an increase in 7 degree Celsius). Since the model performed poorly when using both global and local meteorological data, we can assume that this is due to the model itself rather than the forcing data. Tumbarumba is classified as a sclerophyll forest and JULES does not have this land cover type. We assigned the Needleleaf (NL) PFT to JULES at this site. The introduction of the correct PFT and associated parameters may improve the results at this site (Page 5362, lines 17-28).

57. **Page 5360, lines 8-11: Please give further details of the temperature sensitivity study that was performed at the Tumbarumba site. How were the local data modified and precisely what were the findings of this study?**
The response to this comment is included with the answer to the previous (Page 5362, lines 17-28).

4.2 How much error...

58. **Page 5360, line 18: Again, please quantify and define "similar" in the context of model skill.**
References to sub-figures 2a and 2f have been added (Page 5363 lines 4, 6-7).

59. **Page 5360, line 19: Please do not use the word "significant" which has a very specific statistical meaning when the significance of the results have not been statistically analysed and verified.**
This line has been removed.

60. **Page 5360, lines 19-21: Please re-phrase this statement. What exactly do the authors suggest that the modelling community does if it wishes to perform model simulations at sites with limited or no meteorological data if it "may not" use a widely available global met dataset? Particularly given that the authors have also claimed that this is the best available data and they have not demonstrated that over a long time period model performance is compromised.**
This line has been removed. A further study, over a longer time period, is required in order to ascertain if the WFDEI dataset may be used in place of FLUXNET data.

4.3 Of the global...

61. **Page 5360, line 24: As noted above, the authors now appear to contradict themselves and recommend the use of the WFDEI dataset.**
In this line, we only state that it compares better to the FLUXNET data than PRINCETON does (Page 5363, lines 3-4).

62. **Page 5360, line 24: Again, please define and evaluate "best".**
We have shown in section 3.3 that the WFDEI compares better to FLUXNET data than to PRINCETON. This line has been altered (Page 5363, lines 3-4).

63. **Page 5360, line 26 - Page 5361, line 1: Please see previous comments regarding dataset resolution.**
Information regarding the resolution of the re-analysis datasets, upon which the WFDEI and PRINCETON datasets are based, has been added (Page 5363, lines 5-10).

64. **Page 5361, line 5: Again, please define and evaluate "quite small".**
Done (Page 5363, lines 15-16).

65. **Page 5361, line 6: The word "significantly" has a precise statistical meaning. Please perform the necessary statistical analyses to justify its use.**
The use of the word "significant" has been removed (Page 5363, line 16).

66. **Page 5361, lines 6-12: The authors use the phrase "associated with" four times in this short section. Please consider re-wording this paragraph.**
This paragraph has been re-worded (Page 5364, lines 17-22).

67. **Page 5361, lines 11-12: Please elucidate which sites are being referred to.**
We refer to the tropical sites, Santarem Km67 and Santarem Km83 (Page 5364, lines 21-22).

4.4 Are improvements in...

Throughout - Perhaps the authors could explain why it MODIS data should be seen as a valuable source of information when they have explicitly declared that modellers should not use global meteorological data, even when no alternative is available, and have described MODIS data as noisy at a number of

the sites.
MODIS data should be seen as a valuable source of information since it provides a high resolution and high frequency dataset. It provides a better understanding of plant response to weather and climate and is a useful aid to modellers. Even though we have described the MODIS data as being noisy at a number of sites, it is still an important source of information (Page 5366, lines 5-8).

68. **Page 5361, line 15 - Page 5362, line 3: Please explain and quantify what is meant by "quite noisy", and how this might lead to an under-estimation of GPP at the boreal sites. Further, please explain why the noise in the LAI data at the tropical sites, by contrast, did not lead to a deterioration in model performance.**
Noisy is defined as large day-to-day variations in the MODIS LAI daily timeseries (Page 5364, lines 6-7). The MODIS LAI at the boreal sites was less than the LAI computed by the phenology model, whereas for the tropical sites, the MODIS LAI was greater than that computed by the phenology model.

69. **Page 5362, line 11: "throughout"**
"through out" changed to "throughout" (Page 5364, line 21).

70. **Page 5362, lines 12-14: Please define and quantify the phrases "quite well", "reasonably well", "poor".**
In section 3.4, it was shown that JULES' phenology model performed quite well at the temperate sites, but not at the tropical and cropland sites. At the end of each section, we have a paragraph which summarises the main points of the section. This is for readability and so qualitative phrases are used (Page 5364, lines 22-25).

71. **Page 5362, line 16: Please explain what is meant by "temperate-dependent for the BL PFT class". Is this 'model-tuning'?**
This should read "temperature-dependent for the BL/NL PFT classes". It means that the phenology module (which updates daily LAI) is predominantly influenced by temperature and the parameters chosen for these PFTs are tuned for temperate regions. Yes, this is model tuning. This line has been re-written (Page 5364, lines 26-27).

72. **Page 5362, line 21: "could" or "would" be possible?**
"could" (Page 5365, line 1).

73. **Page 5362, line 22: i.e. the model is tuned to give the correct GPP at temperate sites, but not at other ecosystems.**
The inclusion of tropical PFTs (e.g. tropical evergreen broadleaf and tropical deciduous broadleaf) is important because you can then include important processes which you cannot do with a generic broadleaf tree (rooting depth, Vcmax and photosynthesis-temperature relationships being the main differences). This line has been re-written (Page 5365, lines 1-4).

5 Conclusions

74. **Page 5362, lines 25-26 and throughout: Please be consistent with the use of "local" and "site-specific" which appear to be used interchangeably at present.**
"local" and "site-specific" have the same meaning. As previously mentioned, "site-specific" has been replaced with "local" throughout the manuscript (Page 5365, line 6).

75. **Page 5362, line 27: Please define and evaluate "quite well".**
In section 3.1, it was shown that with the use of local data, improvements in model performance were observed. For readability, we want the conclusions to contain a broad overview of the results and so use qualitative terms to summarise the results (Page 5365, line 8).

76. **Page 5363, line 1: Please quantify the worsening in performance.**
In section 3.1, it was shown that with the use of global data, model performance worsened. As mentioned in the answer to the previous comment, we use qualitative terms to summarise the results in the conclusions (Page 5365, lines 12-13).

77. **Page 5363, line 5: Please split into two sentences: replace "correct and at" with "correct. At"**
Done (Page 5365, line 15).

78. **Page 5363, lines 11-15: Please explain why the global meteorological data that is a poorer match to local data should lead to an improved performance due to higher radiation and temperatures. Does this not simply suggest that there are compensating errors which should be identified and addressed within the model?**
JULES is very sensitive to the downward radiation fluxes and surface air temperatures and so over/under-estimation of these variables in the global meteorological data can affect the model output. So yes, this does mean that there are compensating errors within the model which should be identified and addressed (Page 5365, lines 21-25).

79. **Page 5363, line 17: Please quantify "small".**
This line has been re-written (Page 5365, line 26 - Page 5366, line 1).

80. **Page 5363, lines 17-18: Please explain how the use of MODIS LAI data should lead to an improvement compared to local meteorological data.**
This line has been re-written. It should have said that model simulations using local and MODIS data displayed improvements in modelled GPP compared to using only local data. The beginning and ending of the growing season was more correctly modelled when JULES was forced with daily MODIS LAI (Page 5365, line 26 - Page 5366, line 2).

81. **Page 5363, line 21: Please define and quantify "reasonably well".**
It was shown in sub-section 3.4 that JULES was able to simulate GPP in the temperate regions using the default phenology model. "reasonably well" is used qualitatively for readability (Page 5366, line 4).

82. **Page 5363, lines 22-25: Please elucidate how the improvement in model performance observed when using MODIS data suggests that the model parameters for C3 grasses are more accurate than those for other PFT classes within the model.**
This line has been re-phrased (Page 5366, lines 5-6).

83. **Page 5363, line 29: Remove "and is".**
Done (Page 5366, line 13).

84. **Page 5364, lines 1-4: Please would the authors explain where precisely they have shown that introducing more PFT classes into the JULES model would lead to an improvement in the modelling of the terrestrial carbon cycle, and**

if that is the case, how many and which PFTs would need to be included?
We have not shown that introducing more PFT classes into the JULES model will lead to an improvement in modelling GPP fluxes. We suggest it may lead to improvements. The introduction of a tropical evergreen broadleaf and tropical deciduous broadleaf PFT would improve carbon cycle simulations in tropical regions (Page 5366, lines 14-18).

85. **Page 5364, lines 1-4: Please would the authors explain why they are now arguing that improvements in model parameters and the phenology module would lead to such an improvement in model performance when their simulations appear to have demonstrated that the meteorological data used has a bigger impact on model output than the model vegetation parameters.**
Since our simulations showed that the meteorological data used has a larger impact on model output than model parameters. This would imply that improving the values of the model parameters would only have a small impact on GPP predictions. The improvement in model performance when using global meteorological data means that there are model errors when need to be identified and addressed (as previously mentioned). We only suggest that the introduction of new PFTs and associated model parameters would improve carbon cycle simulations. It is difficult to choose one Vcmax value for all broadleaf PFTs. Also the photosynthesis-temperature relationship is different for plants in temperate and tropical regions.

Tables

86. **As stated above, the authors should include a table listing the results of each model simulation (both absolute and relative values). As stated above, the authors should include a table showing clearly precisely which parameters and variables are altered as a result of switching from site-specific to global data.**
A table listing the results of each model simulation (absolute and percentage differences; Table 6, Page 5382) has been added. As requested previously, a table (Table 1, Page 5377) showing which model parameters are altered as a result of switching from local to global has been included.

87. **Table 4 is referred to before Table 3 in the text. Please consider re-ordering the tables.**
A description of how global model parameters are derived from the IGBP land cover categories has now been included (section 2.3.3: Ecological and soil data, Page 5350-5351) and more detail in included in Appendix A (Page 5376). As a result of this, the tables are referred to in the correct order.

88. **Table 3 - How do the fractions of PFT in each gridcell translate to the model parameters used in these single-point simulations? I assume these are the fractions given in the IGBP dataset, although this is not clear.**
Section 2.3.3 (Ecological and soil data) and Appendix A contains a more detailed description of how the IGBP land cover categories are used to obtain the gridcell PFT fractions and other model parameters, such as annual maximum LAI and canopy height, for model simulations using global data. In Table 4, the first row for each site refers to global data and the second row refers to the local values. For the global data, the IGBP value and class is included and this value is used to obtain the annual maximum LAI (Table 7) and the canopy height factor (Table 8), which is used to compute canopy height.

Figures

89. **Figure 5 - It seems rather self-evident which side of the 1:1 line represents a model over-estimation and which an under-estimation.**
Figure 5 is now Figure 6 (Page 5390). The arrows and under-/over-estimation labels have been removed.

90. **Figure 5 - Please refer to panel (a) and panel (b) in the caption.**
Done (Page 5390).

91. **Figure 5 - In panel (a), which points are MODIS and which are global? I assume that the lighter shades are MODIS but this should be clearly stated in the caption.**
Yes, in panel (a) the lighter shades are MODIS data (Page 5390).

Bibliography

E. Blyth, D. B. Clark, R. Ellis, C. Huntingford, S. Los, M. Pryor, M. Best, and S. Sitch. A comprehensive set of benchmark tests for a land surface model of simultaneous fluxes of water and carbon at both the global and seasonal scale. *Geoscientific Model Development*, 4(2): 255–269, 2011.