

[Interactive
Comment](#)

Interactive comment on “Multi-site evaluation of the JULES land surface model using global and local data” by D. Slevin et al.

Anonymous Referee #2

Received and published: 29 September 2014

This manuscript presents the results from a series of simulations that apply the JULES dynamic vegetation model to the specific conditions at 12 field sites encompassing a range of geographical locations and ecosystem types. For each site the effect of driving JULES with different meteorological data and/or model parameters on the model's skill in estimating gross primary productivity (GPP) was assessed. Such a study was long over-due as land surface models such as JULES are traditionally evaluated and benchmarked using site-specific (i.e. local data) but often used in practice with generic ecosystem parameters and global meteorological data. An understanding of the implications of such practice is important as the modelling community attempts to address the issue of model uncertainty.

Overall, the experiments are well designed and executed, and the results timely. The

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

[Discussion Paper](#)



subject matter fits centrally within the scope of Geoscientific Model Development, and is of interest to researchers from a wide range of disciplines, from plant physiology to climate and Earth System science, and from both the modelling and experimentalist communities. However, I find that the analysis of the results is superficial and would benefit from further depth (see below). I would recommend that the paper be accepted for publication in GMD, but only once the concerns outlined below are satisfactorily addressed.

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.

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-

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



specialist colleagues to read through and highlight sections that are overly technical or insufficiently well-explained.

Introduction, p5345 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.

p5345, L2-4 – 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).

p5345, L17-18 – 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?

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

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



estimates, how this baseline is derived and the uncertainties of “observed” GPP.

Introduction

p5343, L2 – Split into two sentences: “1998). Although”

p5343, L7-10 – Please make it clearer which of these processes are sources and which are sinks.

p5343, L13-17 – Please split this sentence which is over-long and rather impenetrable. Perhaps: “2006). One of its” and “anthropogenic CO₂. The magnitude of this”

p5343, L28 – Please outline briefly what a “bucket” model is in the context of land surface modeling.

p5344, L8 – Is Sellers et al, 1997 really the best and most recent reference for the potential of the current generation of land surface models?

p5345, L2-4 – Inconsistent tenses; replace “compare” with “compared”

2 Methods and model

2.1 Model description

p5345, L21 – Please insert “has” between “and” and “evolved”

p5346, L1-4 – Please explain the tiles system that JULES uses and how this translates to single-point (site) modelling, such as that conducted in this study.

p5346, L1-4 – Presumably this version of the JULES model does not include JULES-CROP, and it appears that none of the PFTs is cropland. Why have the authors chosen to include a cropland site in this evaluation?

p5346, L9 – Please insert “ “ between “component” and “(“

2.2 Experimental design

p5346, L1-2 – As requested above, please explain clearly how the JULES model de-

C1872

GMDD

7, C1869–C1882, 2014

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



scribed 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?

p5346, L28-5347, L2 – Have these sites been used in previous model evaluations or benchmarking studies? If so, please give appropriate references here.

p5347, L2-L5 – Does this mean that for each site a single year was chosen, but that the chosen year differs between sites?

p347, L4 – “gapfilled” should read “gap-filled”

p5347, L13-p5348, L12 – 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.

p5347, L13-p5347, L17 – 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.

p5347, L17 – 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.

p5348, L2-4 – 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

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



default values? Are these only used for the so-called model parameter data?

p5348, L6-12 – 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.

Section 2.3

Please move this section so that the acronyms are explained and the datasets described before the model simulations are outlined.

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.

p5348, L15-16 – 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.

p5349, L2, and throughout – Please use either site-specific or local, or define clearly what is meant by each term if they are in fact different.

p5350, L3-6 – 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).

p5350, L12 – 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.

p5351, L15- Section 2.2.4 states that JULES is driven with daily MODIS LAI data. Here

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



it is stated that the data is actually an 8-day composite. Please clarify this apparent discrepancy.

p5351, L16-17 – Listing the pixel numbers used to create a 3 x 3 gridbox centred on the flux tower is unnecessary.

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.

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.

p5352, L17 – As noted previously, the source, derivation and uncertainties of the observations should be explicitly described somewhere in the paper.

p5352, L17 – 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.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



p5352, L17-L23 – 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?

p5353, L5 – Please make clear to non-specialists, what a sclerophyll forest is.

p5353, L15-p5353, L8 – 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.

p5353, L9-p5354, L9 – Surely, the results of the simulations comparing the use of global meteorological data belongs in sub-section 3.2 (titled “Global meteorological data”)?

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.

p5354, L11 – Should this read “global parameters” rather than “global data”?

p5354, L24-25 – Are these additional errors that are introduced? And does this mean that the model performance has deteriorated?

p5354, L25 – Please define and quantify “small”.

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.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



p5355, L5-6 – 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.

p5355, L11-13 – 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?

p5355, L15 – Please quantify and define “best” in this context.

p5356, L3-8 – Please specify the relative differences in addition to the absolute changes.

p5356, L17 – 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.

p5356, L28-p5357, L1 – Please define and quantify the terms “better” and “quite well”.

p5356, L5 – 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.

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?

p5357, L15,18, 22 – Please quantify “small”.

p5358, L3 – Please remove the first “,” so that it reads “Of the 7 sites where JULES’ performance improved using MODIS data,”

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



p5358, L9 – There seem to be too many “)” in this sentence.

p5358, L18-22 – 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.

p5358, L26 – Please quantify “small”.

p5358, L29 – Please quantify “equally well”.

p5359, L1 – please justify the use of the word “significant” by performing rigorous statistical analyses to the results of this study.

4 Discussions

p5359, L6 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.

4.1 How well doe JULES perform . . .

p5359, L6-8 – Please specify which panel(s) of Fig. 2 show this.

P5359, L8-9 – Please specify which figure and which panel(s) show this.

p5359, L9 – The Mediterranean site ES also appears to be an exception.

p5359, L10-16 – Please make clear which figures and panels show this.

p5359, L20 – As noted previously, these biases and RMSEs need to be put into context.

p5359, L26-p5360, L1 – Please re-phrase to remove the split infinitive, e.g “tend to match more closely the local . . .”

p5360, L7-8 – Please expand on this statement. What exactly do the authors mean

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



when they refer to model error?

p5360, L8-11 – 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?

4.2 How much error . . .

p5360, L18 – Again, please quantify and define “similar” in the context of model skill.

p5360, L19 – 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.

p5360, L19-21 – 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.

4.3 Of the global . . .

p5360, L24 – As noted above, the authors now appear to contradict themselves and recommend the use of the WFDEI dataset.

p5360, L24 – Again, please define and evaluate “best”.

p5360, L26-p5361, L1 – Please see previous comments regarding dataset resolution.

p5361, L5 – Again, please define and evaluate “quite small”.

p5361, L6 – The word “significantly” has a precise statistical meaning. Please perform the necessary statistical analyses to justify its use.

p5361, L6-12 – The authors use the phrase “associated with” four times in this short

section. Please consider re-wording this paragraph.

p5361, L11-12 – Please elucidate which sites are being referred to.

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.

p5361, L15-p5362, L3 – 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.

p5362, L11 – “throughout”

p5362, L12-L14 – Please define and quantify the phrases “quite well”, “reasonably well”, “poor”.

p5362, L16 – Please explain what is meant by “temperate-dependent for the BL PFT class”. Is this ‘model-tuning’?

p5362, L21 – “could” or “would” be possible?

p5362, L22 – i.e. the model is tuned to give the correct GPP at temperate sites, but not at other ecosystems.

5 Conclusions

p5362, L25-26, and throughout – Please be consistent with the use of “local” and “site-specific” which appear to be used interchangeably at present.

p5362, L27 – Please define and evaluate “quite well”.

p5363, L1 – Please quantify the worsening in performance.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



p5363, L5 – Please split into two sentences: replace “correct and at” with “correct. At”

p5363, L11-15 – 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?

p5363, L17 – Please quantify “small”

p5363, L17-18 – Please explain how the use of MODIS LAI data should lead to an improvement compared to local meteorological data.

p5363, L21 – Please define and quantify “reasonably well”.

p5363, L22-25 – 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.

p5363, L29 – Remove “and is”.

p5364, L1-4 – 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?

p5364, L1-4 – 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.

Tables

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

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

include a table showing clearly precisely which parameters and variables are altered as a result of switching from site-specific to global data.

Table 4 is referred to before Table 3 in the text. Please consider re-ordering the tables.

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.

Figures

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 – Please refer to panel (a) and panel (b) in the caption.

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.

Interactive comment on Geosci. Model Dev. Discuss., 7, 5341, 2014.

GMDD

7, C1869–C1882, 2014

Interactive
Comment

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper