Response to Anonymous Referee #1

We thank the referee for providing a review of the manuscript and agree that the suggested changes and clarifications improve it. We have made the changes outlined below in the revised manuscript. Each item starts with the reviewer's comment followed by the changes to the manuscript. The text in blue is a re-written or new paragraph/sentence which was added to the manuscript. The page and line numbers of where changes have been made to the updated manuscript are included at the end of each reply.

Comments

- **General comments; I can't understand the novelty of this manuscript. I agree that the novelty is the performance confirmation of JULES. Please rethink why the authors would like to show others the original results via this manuscript. And, for all things, if the content is related to just JULES unique performance confirmation, it might not directly help the reader's scientific knowledge. At such times the authors need to improve the explanation by changing the standpoint. Please rewrite the manuscript to serve to help the readers in getting maximum benefit from what the authors revealed.**
  This study provides an evaluation of JULES at global and regional scales and provides details on which ecosystems/regions to focus on for future improvements of the model. Changes have been made to the manuscript as suggested and we have added extra text to the manuscript in order to improve and clarify it.

- **Please organize all the information of the model introduction. The authors wrote them in 1. Introduction section and 2.1 Model description. Naturally the 2.1 section should be included contents directly related to this study's discussion, and omit the explanation that had little to do with this study. For example, the authors wrote the interminable explanation for the GPP calculation method, but the reader can understand several author statements in discussion section without such knowledge; there is no explanation about spatial resolution as model structure . . . etc.**
  The model description section has been re-written in order to include only the contents directly related to this study and to explain the effect of the meteorological data on photosynthesis and thus GPP (Pages 3–4). The following text was added:

  JULES is driven by the downward shortwave and longwave radiation fluxes, rainfall and snowfall rates, surface air temperature, wind speed, surface pressure and specific humidity. The downward shortwave and longwave radiation fluxes play an important role in the surface energy balance, where the downwelling radiation fluxes must equal the outgoing fluxes of sensible heat, latent heat, ground flux, reflected shortwave radiation and upwelling thermal energy, and the calculation of photosynthesis (Best et al., 2011; Clark et al., 2011). GPP is the total C used by plants in photosynthesis at the canopy scale with potential (without water and ozone stress) leaf-level photosynthesis calculated as the smoothed minimum of three limiting rates: (1) Rubisco-limited rate (determined using surface air temperature and atmospheric CO_2 concentrations), (2) Light-limited rate (determined using downward radiation fluxes) and (3) Rate of transport of photosynthetic products (C_3 plants) and PEP-Carboxylase limitation (C_4 plants) (determined using surface air temperature and pressure) (Clark et al., 2011). By taking soil moisture stress into account, leaf-level photosynthesis is calculated by multiplying the potential leaf-level photosynthesis by a soil moisture factor (determined using mean soil moisture concentration in the root zone and thus, precipitation).

In JULES, there are two options available for radiation interception and the scaling of photosynthesis from leaf-level to canopy-level: (i) big leaf approach and (ii) multi-layer approach. For all model simulations performed in this study, the multi-layer approach was used which takes into account the vertical gradient of canopy photosynthetic capacity (decreasing leaf nitrogen from top to bottom of canopy) and includes light inhibition of leaf respiration (Option 4 in Table 3 of Clark et al. (2011)). Canopy-scale fluxes are estimated to be the sum of the leaf-level fluxes in each canopy layer, scaled by leaf area. LAI is calculated for each canopy level (default number is 10), with a maximum LAI prescribed for each PFT.

- **Please more explain why the authors used different climate dataset. What of the JULES GPP estimate do the authors reveal? Why did you examine just sensitivity to each dataset? (why didn't you choose the sensitivity to each meteorological parameter?). Please add the comparison among three climate datasets into results. I can't understand the impact of climate dataset on GPP (e.g., fig. 2, 3 . . .), because I don't know the difference of the climate dataset specific feature related to this study. Moreover, please add the explanation of the relationship between JULES and the meteorological parameters in 2.1 Model description section. It means, the reader would like to know the model structural interpretation in discussing what types of calculation approach to choose.**

  Reasons for why we used different meteorological datasets to drive JULES were added to the experimental design section (Page 4, lines 22–25).

  A general overview is provided of how sensitive JULES GPP is to the meteorological dataset used at global scales rather than for each meteorological variable. By analysing the models sensitivity to each meteorological dataset, different analyses of the global climate are compared and therefore a multi-factor analysis of combined changes in meteorological variables can be performed.

  The sensitivity to meteorological parameters was performed in Chapter 6 of the PhD thesis of Darren Slevin(Slevin, 2016). A brief summary of this sensitivity study is provided in section 4.4 (Page 15, lines 23–32).

  A simple sensitivity study of the model to changes in climate (surface (2m) air temperature, precipitation and atmospheric CO_2 concentrations) when simulating GPP at global and regional scales for 2000–2010 was performed in Chapter 6 of the PhD thesis of Darren Slevin(Slevin, 2016) . Only changes to one climate variable were made at a time due to complex interactions associated with multiple changes in climatic factors resulting in complex non-linear ecosystem responses which can be difficult to explain. JULES GPP was found to be sensitive to changes in all three climate variables with modelled LAI only sensitive to changes in surface air temperature (Slevin, 2016). At the regional scale, for model simulations with varying air temperature, GPP increased with increasing temperature in the extratropics, but decreased with increasing temperature in the tropics. Model simulations with varying precipitation at regional scales show the same trend as those at global scales with GPP increasing with increasing precipitation and decreasing with decreasing precipitation except for the magnitude of the effect observed.

  Information on how differences in the three climate datasets (WFDEI-GPCC, WFDEI-

CRU and PRINCETON) affect GPP simulations has been included in various parts of the manuscript (Page 11, lines 19–24; Page 15, lines 1–16). In the model description section (2.1), a paragraph has been added which provides an explanation of the relationship between JULES and the meteorological parameters (Page 3, lines 25–30; Page 4, lines 1–5). This relationship between JULES and the meteorological parameters is also included in the discussion section (Page 12, line 30–Page 13, line 10; Page 15, lines 1–8; Page 15, line 33–Page 16, line 9).

- **The authors should organize first and second paragraph of "1. Introduction". The authors should integrate the two paragraphs into one. P1 L19-20: delete the sentence (Changes in atmospheric CO2 ...). P2 L2 and L4: "location of" → reservoir in? P2 L3: "Changes in the land surface" is not clear. P2 L7: "models and observations (Friedlingstein" → the existing studies (e.g., Friedlingstein ...**
  The first two paragraphs of the introduction have been combined into one with changes made to the text and repetitive text removed. The new paragraph follows (Page 1, line 18–Page 2, line 8).

  The land surface is an important component of the climate system, provides the lower boundary for the atmosphere and exchanges energy, water and carbon (C) with the atmosphere (Pielke et al., 1998; Pitman, 2003; Seneviratne and Stöckli, 2008). It also controls the partitioning of available energy (into latent and sensible heat) and water (into evaporation and runoff) at the surface (Bonan, 2008). Changes in the land surface due to human activities, such as those from tropical deforestation, can influence climate at various time and spatial scales and since the land surface is the location of the terrestrial C cycle, it's ability to act as a C source or sink can influence atmospheric CO_2 concentrations (Le Quéré et al., 2009; Pan et al., 2011; Le Quéré et al., 2013; Tian et al., 2016). The reduced ability of the land surface to absorb increased anthropogenic CO_2 emissions in the future has been shown by models and inferred from observations (Friedlingstein et al., 2006; Canadell et al., 2007; Friedlingstein et al., 2014; Sitch et al., 2015). Friedlingstein et al. (2006) and Friedlingstein et al. (2014) have suggested that a major source of model uncertainty is the land C cycle and this can affect the ability of earth system models (ESMs; also known as coupled carbon-cycle–climate models) to reliably simulate future atmospheric CO_2 concentrations and climate (Dalmonech et al., 2014).

- **The explanation relevant to data used is strange format (P5 L10-P7 L21). For example, why is parameter's unit necessary here? most explanation of "P6 L28-P7 L9" is for the Zhao's work, not this study. After downloading the data, what did the author do as the data pre-processing? The explanation directly related to this study (P7 L13-21) should be written at the start of the paragraph ... etc.**
  The units of the meteorological variables used to drive JULES has been tidied up in the Data section (Section 2.3) (Page 5, lines 26–28 and lines 32–33). The information provided regarding Zhao's work has been shortened (Page 6, lines 14-26). Information on how the data was pre-processed has been included at the end of the paragraph (Page 6, lines 24-26). The paragraph regarding CARDAMOM was structured in such a way that general information on the framework was put first followed by the model output used in this study (Page 6, lines 27–34; Page 7, lines 1–4).

- **P10 L19-21: The statement does not match with fig. 3. It is significant mistake.**

The statement now reads (Page 8, lines 24–26)

This value is greater than that estimated by MODIS, FLUXNET-MTE and CARDAMOM with annual average global GPP estimated to be 112, 130 and 114 $\mathrm{Pg\,C\,year^{-1}}$, respectively, for the same period (Figures 2a, b and d).

- **P15 L13-14: "In general, CARDAMOM was better at simulating GPP than JULES.". Please present factual evidence if the statement is correct. The dataset is created with ground observations, and the empirical method is used to expand it from point to spatial data; CARDAMOM may include some significant error.**
  The statement "In general, CARDAMOM was better at simulating GPP than JULES." was used since global GPP simulated by CARDAMOM GPP (Figure 2) and the pattern of zonal means of total annual model simulated GPP (Figure 5) was between that of MODIS and FLUXNET-MTE. However, we removed this sentence from the Conclusions section (Page 17, lines 1–7). All GPP estimates have errors, but these are not always quantified and provided.

  In the conclusions, the following paragraph was added which discusses the sources of error in the three benchmarking datasets (Page 17, line 30–Page 18, line 4).

  The three benchmarking datasets all contain sources of error. Since observations of GPP do not exist at global scales, the MODIS and FLUXNET-MTE datasets are referred to as observation-based estimates of GPP as they are generated using observations and models. CARDAMOM may contain significant error from the assimilated data and model structure (number of pools, fire resilience of ecosystems), but so do the empirically based FLUXNET-MTE data (up-scaling of a partitioning algorithm) and MODIS GPP (a model based on PFT specific light-use efficiency). The advantage of CARDAMOM is that it is a process-based model and it ensures that the whole ecosystem functioning is coherent, while the observation-based datasets are only empirically based representations of GPP. In Figure S4 of the Supplementary Information of Bloom et al. (2016), there is a detailed study of the sensitivity of CARDAMOM to these various factors at 4 selected pixels representing temperate, boreal, wet and dry tropical ecosystems. Overall, there is not much difference in retrieved parameters because of the large error/uncertainty terms used when computing the likelihood.

- **Fig. 7: As everybody knows, accuracy of the satellite observations is essentially not good at low latitude because of bad observed condition by cloud cover. The authors should represent the difference of GPP in not only low latitude but also other region. Since the evaluation data is global scale, you can do the comparison at global scale. If you keep the way to compare your results with others at just low latitude, please explain the reason.**
  The reason for only examining the difference in GPP fluxes between 15°N–30°N (Figure 5) was to find out which region contributed most to this difference and the possible reasons behind it. We suggest that this difference in GPP was due to incorrect simulation by JULES in Mexico (Figure 7). Even when JULES was driven with multiple meteorological datasets, it was unable to simulate GPP in this region (Figure 5) (Page 12, lines 14–29; Page 15, lines 6–8).

- **Abstract; L6: delete "it was found that" L8: delete "fluxes" L9: delete "It was found that L9: between → among L9-11: this sentence is not clear. L12: what is the meaning of "no impact" → Please add the quantitative interpretation.**

The words were deleted as suggested (Page 1, lines 5–7; Page 1, line 8; Page 1, lines 9–10). The sentence at lines 9-11 was re-written (Page 1, lines 9–10). A quantitative interpretation was added to Line 12 (Page 1, lines 10–12).

Bibliography

M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, R .L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The Joint UK Land Environment Simulator (JULES), Model description–Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4:677–699, 2011. doi: 10.5194/gmd-4-677-2011.

A. A. Bloom, J.-F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, 113:1285–1290, 2016. doi: 10.1073/pnas.1515160113.

G. B. Bonan. Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests. *Science*, 320:1444–1449, 2008. doi: 10.1126/science.1155121.

J. G. Canadell, C. Le Quéré, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R. A. Houghton, and G. Marland. Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences of the United States of America*, 104: 18866–18870, 2007. doi: 10.1073/pnas.0702737104.

D. B. Clark, L. M. Mercado, S. Sitch, C. D. Jones, N. Gedney, M. J. Best, M. Pryor, G. G. Rooney, R. L. H. Essery, E. Blyth, O. Boucher, R. J. Harding, C. Huntingford, and P. M. Cox. The Joint UK Land Environment Simulator (JULES), model description–Part 2: Carbon fluxes and vegetation dynamics. *Geoscientific Model Development*, 4:701–722, 2011. doi: 10.5194/gmd-4-701-2011.

D. Dalmonech, S. Zaehle, G. J. Schürmann, V. Brovkin, C. Reick, and R. Schnur. Separation of the Effects of Land and Climate Model Errors on Simulated Contemporary Land Carbon Cycle Trends in the MPI Earth System Model version 1. *Journal of Climate*, 28:272–291, 2014. doi: 10.1175/JCLI-D-13-00593.1.

P. Friedlingstein, P. Cox, R. Betts, L. Bopp, W. Von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, et al. Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, 19:3337–3353, 2006. doi: 10.1175/JCLI3800.1.

P. Friedlingstein, M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27:511–526, 2014. doi: 10.1175/JCLI-D-12-00579.1.

C. Le Quéré, M. R. Raupach, J. G. Canadell, and G. Marland. Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, 2:831–836, 2009. doi: 10.1038/ngeo689.

C. Le Quéré, R. J. Andres, T. Boden, T. Conway, R. A. Houghton, J. I. House, G. Marland, G. P. Peters, G. R. van der Werf, A. Ahlström, R. M. Andrew, L. Bopp, J. G. Canadell, P. Ciais, S. C. Doney, C. Enright, P. Friedlingstein, C. Huntingford, A. K. Jain, C. Jourdain, E. Kato, R. F. Keeling, K. Klein Goldewijk, S. Levis, P. Levy, M. Lomas, B. Poulter, M. R. Raupach, J. Schwinger, S. Sitch, B. D. Stocker, N. Viovy, S. Zaehle, and N. Zeng. The global carbon budget 1959–2011. *Earth System Science Data*, 5:165–185, 2013. doi: 10.5194/essd-5-165-2013.

Y. Pan, R. A. Birdsey, J. Fang, R. Houghton, P. E. Kauppi, W. A. Kurz, O. L. Phillips, A. Shvidenko, S. L. Lewis, J. G. Canadell, P. Ciais, R. B. Jackson, S. W. Pacala, A. D. McGuire, S. Piao, A. Rautiainen, S. Sitch, and D. Hayes. A Large and Persistent Carbon Sink in the World's Forests. *Science*, 333:988–993, 2011. doi: 10.1126/science.1201609.

R. A. Pielke, R. Avissar, M. Raupach, A. J. Dolman, X. Zeng, and A. S. Denning. Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate. *Global Change Biology*, 4:461–475, 1998. doi: 10.1046/j.1365-2486.1998.t01-1-00176.x.

A. J. Pitman. The evolution of, and revolution in, land surface schemes designed for climate models. *International Journal of Climatology*, 23:479–510, 2003. doi: 10.1002/joc.893.

S. I. Seneviratne and R. Stöckli. *Climate Variability and Extremes during the Past 100 Years* , chapter The Role of Land-Atmosphere Interactions for Climate Variability in Europe, pages 179–193. Springer, 2008.

S. Sitch, P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, A. Ahlström, S. C. Doney, H. Graven, C. Heinze, C. Huntingford, S. Levis, P. E. Levy, M. Lomas, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, A. Arneth, G. Bonan, L. Bopp, J. G. Canadell, F. Chevallier, P. Ciais, R. Ellis, M. Gloor, P. Peylin, S. L. Piao, C. Le Quéré, B. Smith, Z. Zhu, and R. Myneni. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, 12:653–679, 2015. doi: 10.5194/bg-12-653-2015.

D. Slevin. *Investigating sources of uncertainty associated with the JULES land surface model.* PhD thesis, School of GeoSciences, University of Edinburgh, 2016. URL `http://hdl.handle.net/1842/18757`.

H. Tian, C. Lu, P. Ciais, A. M. Michalak, J. G. Canadell, E. Saikawa, D. N. Huntzinger, K. R. Gurney, S. Sitch, B. Zhang, J. Yang, P. Bousquet, L. Bruhwiler, G. Chen, E. Dlugokencky, P. Friedlingstein, J. Melillo, S. Pan, B. Poulter, R. Prinn, M. Saunois, C. R. Schwalm, and S. C. Wofsy. The terrestrial biosphere as a net source of greenhouse gases to the atmosphere. *Nature*, 531:225–228, 2016. doi: 10.1038/nature16946.

Response to Anonymous Referee #2

We thank the referee for providing a review of the manuscript and agree that the suggested changes and clarifications improve it. We have made the changes outlined below in the revised manuscript. Each item starts with the reviewer's comment followed by the changes to the manuscript. The text in blue is a re-written or new paragraph/sentence which has been added to the manuscript. The page and line numbers of where changes have been made to the updated manuscript are included at the end of each reply.

Major Comments

1. **Results: I think the discussion of results in Section 3 needs some improvement, with more detail on the processes behind the modelled and observed patterns in GPP. The focus of the paper is on comparing JULES to these datasets, but it would be more interesting to first explain what the datasets show. Lead each section with a brief explanation of the observed pattern in GPP, and explain differences between the datasets. Then the results of JULES can be given within the context of the observations and CARDAMOM.**
As suggested, in the discussion section, We have started each subsection with a brief explanation of the pattern in observed and CARDAMOM GPP, followed by differences between the datasets. Finally, JULES GPP is given within the context of the observations and CARDAMOM. This has been done for subsections 4.1 and 4.2 (Pages 11–14). This has not been done for subsections 4.3 and 4.4 since the performance of JULES is being evaluated against itself. An extra paragraph was added to subsection 4.3 regarding the effect of spatial resolution on GPP simulations (Page 14, lines 22–26).

   Using a different soil ancillary dataset or land cover map (which specifies the PFT fractions) may have a larger impact than changing the spatial resolution. The regridding method used in this study was the conservative method, which preserves the same information when interpolating from $0.5^\circ \times 0.5^\circ$ to $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$ spatial resolutions, and results in only small differences in global GPP between the model simulations with varying spatial resolution. These small differences are due to differences in the PFT fractions of the land cover map after regridding.

2. **For example, in Figure 2b, JULES does very well if you are only comparing to MODIS. But it overestimates the variability of GPP during winter months compared to the other two datasets. So does this mean that JULES captures the interannual variability, or not?**
In Figure 3b, JULES does very well if it is only compared to MODIS and overestimates the variability of GPP during winter months compared to the other two datasets (FLUXNET-MTE and CARDAMOM). We would say that JULES captures interannual variability since the coefficient of variation (CV) expressed as percentages of the mean monthly GPP for JULES lies between the CV values for the three observation-based estimates (Page 12, lines 3–5).

3. **Another example: Page 11, Lines 29–33 (Discussion of figure 6): Why are the results for the extratropics the only ones discussed? I think much more could be said here - instead of just listing the differences it would be better to provide some more evaluation. For example, it was already stated that**

JULES overestimates GPP in the tropics, and this analysis shows that the overestimation occurs in all tropical land areas. That is a useful thing to note. On the other hand, JULES does reasonable in the extratropics - but it is consistently lower than all three datasets in Northern Asia.
The results for the tropics has been added including some suggestions for improving simulated GPP. The following paragraphs were added to sections 3.3 and 4.2, respectively (Page 10, lines 15–21; Page 13, lines 17–19).

JULES overestimates GPP in all three tropical land areas compared to MODIS, FLUXNET-MTE and CARDAMOM (Figures 6c, e and f). Differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP with average annual GPP range from 7.4–12.1 $\mathrm{Pg\,C\,year^{-1}}$, 7.7–13 $\mathrm{Pg\,C\,year^{-1}}$ and 1–1.3 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively, in South and South-East Asia, 9.5–13.7 $\mathrm{Pg\,C\,year^{-1}}$, 8.4–12.3 $\mathrm{Pg\,C\,year^{-1}}$ and 1.7–2.1 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively, in Africa and 18-23.2 $\mathrm{Pg\,C\,year^{-1}}$, 9–12.9 $\mathrm{Pg\,C\,year^{-1}}$ and 1.4–1.8 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively, in Central and South America (Figures 6c, e and f, respectively).

JULES simulated average annual GPP to be 61, 54 and 7 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively. JULES (JULES-WFDEI-GPCC) simulates higher GPP than MODIS, FLUXNET-MTE and CARDAMOM at global scales and this was found to be due to higher GPP simulated by JULES for forests and grasslands in the tropics (Figure 4b).

Yes we found that JULES performs reasonably well in the extratropics (Europe, Northern Asia, North America and Greenland and the Extratropical Southern Hemisphere), with the exception of Northern Asia and North America and Greenland, where the model is either equal to or lower than all three datasets. This may be due to the inability of this version of JULES to accurately simulate GPP in boreal regions where permafrost exists. It may also due to a different land cover map being used by JULES, MODIS and FLUXNET-MTE. The following paragraph were added to section 4.2 (Page 14, lines 4–8).

In the four extratropical regions (Europe, Northern Asia, Extratropical Southern Hemisphere and North America and Greenland), JULES simulated similar GPP to MODIS, FLUXNET-MTE and CARDAMOM for the three biomes in Europe and the Extratropical Southern Hemisphere (Figures 6a and d), with the exception of Northern Asia and North America and Greenland, where the model is either equal to or lower than all three datasets (Figures 6b and g). This is due to the inability of this version of JULES to accurately simulate GPP in boreal regions where permafrost exists.

4. **Robustness of results: A potential strength of this manuscript is the comparison of JULES using different datasets, however I found the discussion of this topic a bit thin. Could the authors provide some more detailed discussion and context of the results? Here are some examples where further information could be provided: It's interesting that the results were insensitive to the spatial resolution (Page 12, Lines 9-12). This is an important conclusion of the analysis, and as the authors point out, using courser resolutions can save computational resources. But is the result surprising given that the same soil ancillary data was used for all experiments? The JULES parameterizations are not scale-dependent (for example, this isn't the same as comparing scales in a model that uses cloud microphysical processes). I think using a different soil ancillary data set would have a larger impact than changing the resolution.**
Yes, we found it interesting that the results were insensitive to spatial resolution. This is

a useful since lower resolution global simulations can be performed to save computational resources. Using a different soil ancillary dataset or land cover map would have a larger impact than changing the resolution. A paragraph discussing these points has been added to section 4.3 (Page 14, line 22–26).

Using a different soil ancillary dataset or land cover map (which specifies the PFT fractions) may have a larger impact than changing the spatial resolution. The regridding method used in this study was the conservative method, which preserves the same information when interpolating from $0.5^\circ \times 0.5^\circ$ to $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$ spatial resolutions, and results in only small differences in global GPP between the model simulations with varying spatial resolution. These small differences are due to differences in the PFT fractions of the land cover map after regridding.

5. **Also the meteorological dataset did not strongly change the results. However this is dependent on two things: 1) Maybe there were not large differences in climate between the data sets? IE: Page 15, Lines 2-6: Why are these differences in GPP occurring? Is the temperature and precipitation (or other variables) very different between the datasets in these regions? Are there other regions where the climate is very different, but the JULES simulations do not show dramatically different GPP? It would be good to provide some more information on the climates from the different driving data sets. 2) Since JULES was run with prescribed PFTs, there was no feedback between NPP and the land cover. It's possible that the GPP would be much more sensitive to the meteorology if competition between PFTs were allowed. Could the authors provide two additional experiments where the competition is allowed (e.g. one with either WFDEI product and one with the PRINCETON dataset)? Or at least provide the caveat that these results are possibly only valid when TRIFFID is not turned on. Although it's more work, I do think the additional simulations with TRIFFID would make this paper more relevant to a larger audience, as it seems most investigations using JULES have TRIFFID predicting PFTs (for example in TRENDY, the HELIX project, ISIMIP, and most CMIP5 and upcoming CMIP6 experiments).**
1) When JULES was driven with different meteorological datasets, differences in simulated GPP occurred mostly in the tropics (between 5°N-5°S) with JULES driven with WFDEI-GPCC-1degree simulating higher GPP than JULES driven with PRINCETON and slightly higher GPP in the extratropics was simulated by JULES was driven with PRINCETON (Figure 5). There are differences in climate between the two datasets. Positive biases in the downward longwave radiation fluxes and surface air temperatures in the meteorological datasets are the reason for these differences (Figures G.5 and G.6 in Slevin (2016)). In general, precipitation in the WFDEI-GPCC dataset is higher than that of PRINCETON (Figures G.6b and d in Slevin (2016)) with surface air temperatures higher in PRINCETON (Figures G.6a and c in Slevin (2016)). However, since JULES is more sensitive to downward longwave radiation and surface air temperature than precipitation when simulating GPP, the main reason for differences in simulated GPP when JULES was driven with two different meteorological datasets is due to differences in downward longwave radiation fluxes and surface air temperatures. There are differences in northern Eurasia (above 60°N) in the meteorological datasets with slightly higher radiation fluxes (downward shortwave and longwave) and surface air temperatures in the PRINCETON dataset with little difference between the JULES simulations driven with WFDEI-GPCC and PRINCETON in this region (Figure 5). Information on differences in the meteoro-

logical dataset (WFDEI-GPCC and PRINCETON) led to differences in simulated GPP has been added to section 4.4 on Page 15, lines 1–16.

The higher simulated GPP in the tropics when JULES was driven with WFDEI-GPCC is due to positive biases in downward longwave radiation fluxes in WFDEI-GPCC in the Amazonian, African and South-East Asian tropics (Figures G.5b and d in Slevin (2016)) and the higher GPP simulated by JULES (driven with PRINCETON) in the extratropics are a result of positive biases in downward longwave radiation in the PRINCETON dataset in North America and Northern Asia (Figure G.5b in Slevin (2016)) and positive biases in surface air temperature in the PRINCETON dataset in the Northern Hemisphere (Figures G.6a and c in Slevin (2016)). As with the JULES-WFDEI-GPCC simulations, there are also differences in GPP between the PRINCETON driven JULES simulation and the observation-based and CARDAMOM estimates at latitudes 15°N-30°N (Figure 5). There was no improvement in simulated GPP when a different meteorological dataset was used.

In general, precipitation in the WFDEI-GPCC dataset is higher than that of PRINCETON (Figures G.6b and d in Slevin (2016)) with surface air temperatures higher in PRINCETON (Figures G.6a and c in Slevin (2016)). However, since JULES is more sensitive to downward longwave radiation and surface air temperature than precipitation when simulating GPP(Alton et al., 2007), the main reason for differences in simulated GPP when JULES was driven with two different meteorological datasets is due to differences in downward longwave radiation fluxes and surface air temperatures. There are differences in northern Eurasia (above 60°N) in the meteorological datasets with slightly higher radiation fluxes (downward shortwave and longwave) and surface air temperatures in the PRINCETON dataset with little difference between the JULES simulations driven with WFDEI-GPCC and PRINCETON in this region (Figure 5).

and on Page 15, line 33–Page 16, line 9.

When JULES was driven with the PRINCETON dataset, it was found that simulated photosynthesis was mostly Rubisco-limited (Figure 5.25 in Slevin (2016)). A similar trend was found when JULES was driven with the WFDEI-GPCC dataset (Figure 5.6 in Slevin (2016)). Similar trends in transport limitation were found with the JULES-PRINCETON model simulation, though the number of model gridboxes in which transport limitation dominated was less than that for the JULES-WFDEI-GPCC-1degree model simulation (Figures 5.25 and 5.28 in Slevin (2016)). When comparing the model gridbox fractions for the JULES-WFDEI-GPCC-1degree and JULES-PRINCETON model simulations, it was found that when JULES was driven with the PRINCETON dataset, simulated photosynthesis was more Rubisco-limited than when the model was driven with WFDEI-GPCC (Figure 5.26 in Slevin (2016)). Light-limitation was more important in simulating photosynthesis when JULES was driven with WFDEI-GPCC than PRINCETON (Figure 5.27 in Slevin (2016)). The percentage of model gridboxes which are transport-limited show a pronounced geographical variation with the WFDEI-GPCC driven simulation being more transport-limited in the Southern Hemisphere and the PRINCETON driven simulation being more transport-limited in the Northern Hemisphere (Figure 5.28 in Slevin (2016)).

2) Yes, since JULES was run with prescribed PFTs, there was no feedback between NPP and the land cover and there is a possibility that GPP could be more sensitive to the meteorology if competition between PFTs were allowed. These additional simulations with TRIFFID would make this paper more relevant to a larger audience. Therefore, two more model simulations were carried out where vegetation competition (and TRIFFID) were switched on. This was done with the WFDEI-GPCC and PRINCETON datasets

(both at $1^\circ \times 1^\circ$ spatial resolution). A new figure was added showing the results from these extra model simulations (Page 32, Figure 8). A paragraph describing the results from these simulations was added to section 4.4 (Page 16, lines 10-24)

In this study, the model simulations were performed with prescribed PFTs (i.e. no vegetation competition). If competition between PFTs was allowed (i.e. vegetation competition), the annual average global GPP would be higher by 15 % and 17 %, for the WFDEI-GPCC and PRINCETON driven simulations, respectively (Figures 8b and e). In general, with vegetation competition switched on, higher GPP was simulated by JULES when driven with both datasets (Figures 8c and f). Higher GPP occurred mostly in Europe, south-eastern US, and in the tropical regions of Central and South America, Africa and South and South-East Asia (Figures 8c and f). This increased GPP in tropical regions is due to the tree-shrub-grass dominance heirachy in TRIFFID with dominant types (trees) limiting the expansion of subdominant types (shrubs and grasses). In savanna regions, such as the Sudanian Savanna, which stretches from the Atlantic Ocean in the west to the Ethiopian Highlands in the east of Africa, and northern Australia, there is higher GPP with prescribed PFTs (Figures 8c and f). These are also fire-prone regions. The version of JULES used in this study has no fire module and TRIFFID may overestimate woody cover and therefore GPP.

In terms of global GPP, the WFDEI-GPCC and PRINCETON driven simulations produce similar increases (Figures 8b and e). However, the spatial pattern is slightly different with higher GPP simulated in the Amazon region when JULES was driven with the WFDEI-GPCC dataset and higher GPP in southern Brazil and Argentina and Southeast Asia when JULES was driven with the PRINCETON dataset (Figures 8c and f). The spatial pattern of simulated GPP is more sensitive to the meteorological data than the annual average global GPP if competition between PFTs is allowed. This may be due to compensating differences in the sensitivity of the model to the two meteorological datasets.

and to the conclusions (Page 17, lines 26–29).

The model simulations in this study were largely performed with prescribed PFTs (i.e. no competition between PFTs was allowed). With competition between PFTs, the annual average global GPP was higher by 15 % and 17 %, for the WFDEI-GPCC and PRINCETON driven simulations, respectively, with the spatial pattern of simulated GPP more sensitive to the meteorological data used.

Other Comments

1. **There are several places where the text is repetitive:**
   The text has been updated to avoid repetition (see below).

   - **GPP is important because errors in its calculation can propagate through the model and affect biomass and other flux calculations: Page 2, Lines 27–28; Page 3, Line 5; Page 4 Lines 31–33.**
     Done (Page 2, lines 19–24).
   - **JULES is compared against FLUXNET-MTE, MODIS GPP, and CARDAMOM: Page 5, Lines 1–2; Page 5, Lines 6–7; Page 5, Line 11.**
     Done (Page 4, lines 27–28).
   - **Simulations are 2001–2010 because of availability of data: Page 4, Lines 33–34; Page 5, Lines 6–7.**
     Done (Page 4, lines 26–27).

- **The list of driving meteorological variables is given three times on pages 5–6. Even though there are differences between what is available from WATCH vs PRINCETON, this information could be given in a more concise manner.**
  Done (Page 5, lines 26–28 and 32–33). The driving meteorological variables is also listed in the model description section since it is required when explaining the connection between the meteorological variables and GPP (Page 3, lines 25–26).
- **The FLUXNET-MTE is described as being derived from a machine learning technique/ model tree ensemble twice in lines 14–20 of Page 6.**
  Done (Page 6, lines 2–6).
- **Section 2.4 - There are more examples of this, please proofread the text and remove all repetition.**
  This section was re-written in order to avoid repetition (Page 7, lines 5–29).

2. **Page 2, Lines 6–7: It would be incorrect to say the reduced ability of land to absorb CO2 in the future has been observed. Perhaps better to say "...has been shown by models and inferred from observations ..."**
   This sentence has been changed to (Page 2, lines 3–5)

   The reduced ability of the land surface to absorb increased anthropogenic CO_2 emissions in the future has been shown by models and inferred from observations (Friedlingstein et al., 2006; Canadell et al., 2007; Friedlingstein et al., 2014; Sitch et al., 2015).

3. **Page 3, Lines 5–9: This paragraph needs some revision. The comparison of JULES to these precise datasets is not an important part of model development in general. Would be better to say that evaluating the simulated GPP at a range of scales and its sensitivity to spatial resolution and meteorological data is essential for informing future model developments. The specific datasets can be mentioned next, ie "In this manuscript, we do this using the FLUXNET-MTE etc."**
   This paragraph has been re-written (Page 2, line 33–Page 3, line 2).

   JULES has been evaluated at various scales: point (Blyth et al., 2010, 2011; Slevin et al., 2015; Ménard et al., 2015), regional (Galbraith et al., 2010; Burke et al., 2013; Chadburn et al., 2015) and globally as part of model-intercomparison studies (Anav et al., 2015; Sitch et al., 2015). Evaluating simulated GPP at a range of scales and its sensitivity to spatial resolution and meteorological data is essential for informing future model developments. In this manuscript, we do this using two observation-based datasets (FLUXNET-MTE and MODIS) and the CARbon DAta MOdel fraMework (Bloom et al., 2016, CARDAMOM).

4. **Page 3, Line 25: I suggest removing "In LSMs"**
   Done (Page 3, lines 18–19).

5. **Page 5, Lines 11–12: Please specify what information is provided by the soil dataset.**
   The following information is provided by the soil dataset (Page 5, lines 3–6).

   The soil dataset used was the Harmonized World Soil Database version 1.2 (Nachtergaele et al., 2012, HWSD) and contains soil property data such as soil texture fractions, water storage capacity, soil depth and pH (Nachtergaele et al., 2012). In this study, the soil texture fractions (% of sand, silt and clay) were used to calculate the soil thermal and hydraulic conductivity parameters listed in Table 3 of Best et al. (2011).

6. **Page 5, Lines 17-19: I don't see why the requirement for data at 6 hourly intervals or less leads to the need for a number of datasets. However, there is value in evaluating model response to a number of datasets - for example JULES is currently run with different datasets for a number of projects and MIPs, and it is not known to what extent these different datasets affect the results.**
Yes, I agree that evaluating JULES' response to various datasets (soil, vegetation and meteorological) can help to explain its behaviour when used as part of a multi-model inter-comparison project. This sentence has been changed to (Page 5, lines 8–10)

   Two meteorological datasets were used to drive the model offline (i.e. run separately from its host Earth System Model) at global scales; WFDEI (Weedon et al., 2014) and PRINCETON (Sheffield et al., 2006).

7. **Page 7, Lines 10-11: What is meant by modelling "quality"?**
The sentence is missing the word "improve". It now reads (Page 6, lines 27–29)

   The CARbon DAta MOdel fraMework (CARDAMOM) is a model-data fusion approach which consists of merging observational data with models in order to improve model quality and characterise its uncertainty.

8. **All evaluation of GPP is based on area-weighted GPP, correct? I think this could be said once in Section 2.5 and then it does not need to be repeated throughout the remainder of the text.**
Yes, all evaluation of GPP is based on area-weighted GPP. Since it is mentioned in Section 2.5, it has been removed from the remainder of the text.

9. **I would lead the results with the evaluation of the global GPP, then examine seasonal and interannual variation (ie switch sections 3.2 and 3.1). The seasonal cycle discussion does not belong in the section on interannual variability. This section should be renamed "Seasonal and interannual variability." Each section in the results ends with a one sentence summary - consider moving this sentence to the beginning of each section instead.**
Sections 3.2 and 3.1 were switched (Pages 8–9). This also required that the abstract (Page 1, lines 5–8), the study questions (Page 3, lines 6–10), the list of experiments (Section 2.4; Page 7) and parts of Section 4.1 (Discussion; Page 12–14) and the Conclusions (Section 5; Page 16–18) be slightly re-written. Section 3.2 has been renamed to "Seasonal and interannual variability of GPP." (Page 9). The one sentence summary at the end of each section in the results section has been moved to the beginning.

10. **Page 10, Lines 8-9: I would move the last sentence of this paragraph to earlier in the paragraph since it explains how the reader should interpret the CV plot.**
The last sentence of this paragraph has been moved to earlier in the paragraph as suggested (Page 9, lines 13–15).

11. **Page 10, Lines 13-15: This sentence is unclear.**
This sentence has been rewritten (Page 9, lines 22–23).

    The model is able to capture simulated monthly anomalies from 2001 to 2010 with the exception of those in 2002 (Figure 3c).

12. **Page 10, Lines 21-23: These numbers are different from what's given in Figure 3.**
The numbers have been changed to reflect those given in Figure 2 (Page 8, lines 24–27).

This value is greater than that estimated by MODIS, FLUXNET-MTE and CARDAMOM with annual average global GPP estimated to be 112, 130 and 114 Pg C $year^{-1}$, respectively, for the same period (Figures 2a, b and d). The higher global GPP simulated by the JULES-WFDEI-GPCC driven simulations is greater than the MODIS, FLUXNET-MTE and CARDAMOM estimates by 25 %, 8 % and 23 % on average, respectively.

13. **In Section 3.3, it's a bit unusual to give total over the 10 year period, instead of annual fluxes, which is what is more usually reported in global-scale evaluations of GPP.**
Annual fluxes have been provided for GPP in Section 3.3 (Pages 9–10).

14. **Throughout the results, it would be much easier to read through if a range of the results are given instead of listing each GPP value every time. For example, Page 11, Line 15: Replace with "JULES overestimates total annual GPP by 20-41%"**
The results section has now been changed so that a range of results are given instead of listing each GPP value every time (Pages 8–11).

15. **Page 12, Lines 22, 24: I think it would be more appropriate to refer to the "pattern" of zonal means rather than the "trend" in zonal means, as trends typically refer to change in time, rather than change in space.**
Yes, you are correct. This has been changed (Page 11, lines 15–18).

16. **It's difficult to distinguish between the reds and pinks in Figures 2, 3, and 5; and between the shades of blue/green in Figures 4 and 6. Could a different set of colors be used?**
A different set of colors has been used to distinguish between the various model simulations in order to make it easier for the reader (Pages 26–30).

Bibliography

P. Alton, L. Mercado, and P. North. A sensitivity analysis of the land-surface scheme JULES conducted for three forest biomes: Biophysical parameters, model processes, and meteorological driving data. *Global Biogeochemical Cycles*, 20:GB1008, 2007. doi: 10.1029/2005GB002653.

A. Anav, P. Friedlingstein, C. Beer, P. Ciais, A. Harper, C. Jones, G. Murray-Tortarolo, D. Papale, N. C. Parazoo, P. Peylin, S. Piao, S. Sitch, N. Viovy, A. Wiltshire, and M. Zhao. Spatio-temporal patterns of terrestrial gross primary production: A review. *Reviews of Geophysics*, 2015. doi: 10.1002/2015RG000483.

M. J. Best, M. Pryor, D. B. Clark, G. G. Rooney, R .L. H. Essery, C. B. Ménard, J. M. Edwards, M. A. Hendry, A. Porson, N. Gedney, L. M. Mercado, S. Sitch, E. Blyth, O. Boucher, P. M. Cox, C. S. B. Grimmond, and R. J. Harding. The Joint UK Land Environment Simulator (JULES), Model description–Part 1: Energy and water fluxes. *Geoscientific Model Development*, 4:677–699, 2011. doi: 10.5194/gmd-4-677-2011.

A. A. Bloom, J.-F. Exbrayat, I. R. van der Velde, L. Feng, and M. Williams. The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, 113:1285–1290, 2016. doi: 10.1073/pnas.1515160113.

E. Blyth, J. Gash, A. Lloyd, M. Pryor, G. P. Weedon, and J. Shuttleworth. Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data. *Journal of Hydrometeorology*, 11:509–519, 2010. doi: 10.1175/2009JHM1183.1.

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: 255–269, 2011. doi: 10.5194/gmd-4-255-2011.

E. J. Burke, R. Dankers, C. D. Jones, and A. J. Wiltshire. A retrospective analysis of pan Arctic permafrost using the JULES land surface model. *Climate Dynamics*, 41:1025–1038, 2013. doi: 10.1007/s00382-012-1648-x.

J. G. Canadell, C. Le Quéré, M. R. Raupach, C. B. Field, E. T. Buitenhuis, P. Ciais, T. J. Conway, N. P. Gillett, R. A. Houghton, and G. Marland. Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks. *Proceedings of the National Academy of Sciences of the United States of America*, 104: 18866–18870, 2007. doi: 10.1073/pnas.0702737104.

S. Chadburn, E. Burke, R. Essery, J. Boike, M. Langer, M. Heikenfeld, P. Cox, and P. Friedlingstein. An improved representation of physical permafrost dynamics in the JULES land-surface model. *Geoscientific Model Development*, 8:1493–1508, 2015. doi: 10.5194/gmd-8-1493-2015.

P. Friedlingstein, P. Cox, R. Betts, L. Bopp, W. Von Bloh, V. Brovkin, P. Cadule, S. Doney, M. Eby, I. Fung, et al. Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison. *Journal of Climate*, 19:3337–3353, 2006. doi: 10.1175/JCLI3800.1.

P. Friedlingstein, M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27:511–526, 2014. doi: 10.1175/JCLI-D-12-00579.1.

D. Galbraith, P. E. Levy, S. Sitch, C. Huntingford, P. Cox, M. Williams, and P. Meir. Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change. *New Phytologist*, 187:647–665, 2010. doi: 10.1111/j.1469-8137.2010.03350.x.

C. B. Ménard, J. Ikonen, K. Rautiainen, M. Aurela, A. N. Arslan, and J. Pulliainen. Effects of Meteorological and Ancillary Data, Temporal Averaging, and Evaluation Methods on Model Performance and Uncertainty in a Land Surface Model. *Journal of Hydrometeorology*, 16: 2559–2576, 2015. doi: 10.1175/JHM-D-15-0013.1.

F. Nachtergaele, H. van Velthuizen, L. Verelst, D. Wiberg, N. Batjes, K. Dijkshoorn, V. van Engelen, G. Fischer, A. Jones, L. Montanarella, M. Petri, S. Prieler, E. Teixeira, and X. Shi. Harmonized World Soil Database v1.2. Technical report, International Institute for Applied Systems Analysis (IIASA), Food and Agriculture Organization of the United Nations (FAO), February 2012.

J. Sheffield, G. Goteti, and E. F. Wood. Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, 19:3088–3111, 2006. doi: 10.1175/JCLI3790.1.

S. Sitch, P. Friedlingstein, N. Gruber, S. D. Jones, G. Murray-Tortarolo, A. Ahlström, S. C. Doney, H. Graven, C. Heinze, C. Huntingford, S. Levis, P. E. Levy, M. Lomas, B. Poulter, N. Viovy, S. Zaehle, N. Zeng, A. Arneth, G. Bonan, L. Bopp, J. G. Canadell, F. Chevallier, P. Ciais, R. Ellis, M. Gloor, P. Peylin, S. L. Piao, C. Le Quéré, B. Smith, Z. Zhu, and R. Myneni. Recent trends and drivers of regional sources and sinks of carbon dioxide. *Biogeosciences*, 12:653–679, 2015. doi: 10.5194/bg-12-653-2015.

D. Slevin. *Investigating sources of uncertainty associated with the JULES land surface model.* PhD thesis, School of GeoSciences, University of Edinburgh, 2016. URL `http://hdl.handle.net/1842/18757`.

D. Slevin, S. F. B. Tett, and M. Williams. Multi-site evaluation of the JULES land surface model using global and local data. *Geoscientific Model Development*, 8:295–316, 2015. doi: 10.5194/gmd-8-295-2015.

G. P. Weedon, G. Balsamo, N. Bellouin, S. Gomes, M. J. Best, and P. Viterbo. The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data. *Water Resources Research*, 50:7505–7514, 2014. doi: 10.1002/2014WR015638.

Global Evaluation of Gross Primary Productivity in the JULES Land Surface Model ~~.~~v3.4.1

Darren Slevin[1], Simon F. B. Tett[1], Jean-François Exbrayat[1, 2], A. Anthony Bloom[1, 2, 3], and Mathew Williams[1, 2]

[1]School of GeoSciences, The University of Edinburgh, Crew Building, Alexander Crum Brown Road, Edinburgh, EH9 3FF, UK

[2]National Centre for Earth Observation, The University of Edinburgh, Crew Building, Alexander Crum Brown Road, Edinburgh, EH9 3FF, UK

[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA

Correspondence to: Darren Slevin (d.slevin@ed.ac.uk)

Abstract. This study evaluates the ability of the JULES Land Surface Model (LSM) to simulate Gross Primary Productivity (GPP) at regional and global scales for 2001–2010. Model simulations, performed at various spatial resolutions and driven with a variety of meteorological datasets (WFDEI-GPCC, WFDEI-CRU and PRINCETON), were compared to the MODIS GPP product, spatially gridded estimates of upscaled GPP from the FLUXNET network (FLUXNET-MTE) and the CARDAMOM terrestrial carbon cycle analysis. Firstly, ~~JULES was found to simulate interannual variability (IAV) at global scales. When JULES was~~ when JULES was driven with the WFDEI-GPCC dataset (at $0.5° \times 0.5°$ spatial resolution), ~~it was found that~~ the annual average global GPP simulated by JULES for 2001–2010 was higher than the observation-based estimates (MODIS and FLUXNET-MTE), by 25 % and 8 %, respectively, and CARDAMOM estimates by 23 %. JULES was found to simulate interannual variability (IAV) at global scales. Secondly, GPP ~~fluxes~~ simulated by JULES for various biomes (forests, grasslands and shrubs) at global and regional scales were compared. ~~It was found that differences between~~ Differences among JULES, MODIS, FLUXNET-MTE and CARDAMOM at global scales were ~~mostly~~ due to differences in ~~the tropics with CARDAMOM performing better than JULES in this region~~simulated GPP in the tropics. Thirdly, it was shown that spatial resolution ($0.5° \times 0.5°$, $1° \times 1°$ and $2° \times 2°$) had ~~no~~ little impact on simulated GPP ~~on~~ at these large scales with global GPP ranging from 140–142 $\mathrm{Pg\,C\,year^{-1}}$. Finally, the sensitivity of JULES to meteorological driving data, a major source of model uncertainty, was examined. Estimates of annual average global GPP were higher when JULES was driven with the PRINCETON meteorological dataset than when driven with the WFDEI-GPCC dataset by ~~4~~3 $\mathrm{Pg\,C\,year^{-1}}$. At regional scales, differences between two were observed with the WFDEI-GPCC driven model simulations estimating higher GPP in the tropics (at 5°N–5°S) and the PRINCETON driven model simulations estimating higher GPP in the extratropics (at 30°N–60°N).

1 Introduction

~~Changes in atmospheric concentrations and water vapour can alter the energy balance of the atmosphere and thus climate. One important influence on these greenhouse gases is the land surface.~~ The land surface is an important component of the climate system, provides the lower boundary for the atmosphere and exchanges energy, water and carbon (C) with the atmosphere (Pielke et al., 1998; Pitman, 2003; Seneviratne and Stöckli, 2008). It also controls the partitioning of available energy (into latent and sensible heat) and water (into evaporation and runoff) at the surface ~~, is the location of the terrestrial C sink and influences weather and climate and vice versa~~ (Bonan, 2008). Changes in the land surface due to human activities, such as those from tropical deforestation, can influence climate at various time and spatial scales and since the land surface is the location of the terrestrial C cycle, it's ability to act as a C source or sink can influence atmospheric CO_2 concentrations (Le Quéré et al., 2009; Pan et al., 2011; Le Quéré et al., 2013; Tian et al., 2016).

The reduced ability of the land surface to absorb increased anthropogenic CO_2 emissions in the future has been shown by models and inferred from observations (Friedlingstein et al., 2006; Canadell et al., 2007; Friedlingstein et al., 2014; Sitch et al., 2015). ~~The Coupled Climate–Carbon Cycle Model Intercomparison Project (Friedlingstein et al., 2006, C4MIP) and phase 5 of the Coupled Model Intercomparison Project (Arora et al., 2013; Friedlingstein et al., 2014, CMIP5) both showed a large spread in the future projections of atmospheric by coupled climate carbon-cycle models using the same emission scenario. Using observations of atmospheric concentrations since the 1960s, Canadell et al. (2007) showed a reduction in the efficiency of CO_2 sinks on land and oceans to store anthropogenic emissions.~~ Friedlingstein et al. (2006) and Friedlingstein et al. (2014) have suggested that a major source of model uncertainty is the land C cycle and this can affect the ability of earth system models (ESMs; also known as coupled carbon-cycle–climate models) to reliably simulate future atmospheric CO_2 concentrations and climate (Dalmonech et al., 2014).

Plants fix CO_2 as organic compounds through photosynthesis at the leaf scale and Gross Primary Productivity (GPP) is the total amount of C used in photosynthesis by plants at the ecosystem level (Beer et al., 2010; Chapin III et al., 2012). Photosynthesis at the leaf and canopy scale vary in response to changes in climate (temperature, precipitation, humidity and downward radiation fluxes) and nutrient availability (Anav et al., 2015). Terrestrial GPP is an important (and the largest) C flux since it drives several ecosystem functions such as respiration and growth (Beer et al., 2010). GPP contributes to the production of food, fibre, and wood for humans and along with respiration, is one of the major processes controlling the exchange of CO_2 between the land and atmosphere (Beer et al., 2010). It also plays an important role in the global C cycle helping terrestrial ecosystems to partially offset anthropogenic CO_2 emissions (Janssens et al., 2003; Cox and Jones, 2008; Battin et al., 2009; Anav et al., 2015)

However, at the global scale, there are no direct measurements of GPP (Anav et al., 2015). Global estimates of GPP exist, but are not solely based on measurements and, therefore, large uncertainties exist in these estimates (Anav et al., 2015). In LSMs, the correct simulation of GPP is important since errors in its calculation can propagate through the model and affect biomass and other flux calculations, such as Net Ecosystem Exchange ~~(Schaefer et al., 2012, NEE)~~(NEE) (Schaefer et al., 2012). In JULES, NEE is not a model output and is calculated as total ecosystem respiration minus GPP. The correct representation of leaf level stomatal conductance influences GPP and transpiration and errors in GPP can also introduce errors into simulated latent and sensible heat fluxes.

Land surface models (LSMs) have become considerably more complex since the simple "bucket" model of Manabe (1969). Deardorff (1978) developed a model which could simulate temperature and moisture for two soil layers and included a vegetation layer. Sellers et al. (1986) built on the work of Deardorff (1978) by developing a globally applicable LSM. Foley et al. (1996) incorporated vegetation dynamics into an LSM. These developments have led to LSMs which can realistically represent complex vegetation responses to meteorology, the climate effect of snow and biogeochemical processes (Pitman, 2003; van den Hurk et al., 2011). Therefore, as LSMs become more complex, their accuracy must be evaluated.

JULES has been evaluated at various scales: point (Blyth et al., 2010, 2011; Slevin et al., 2015; Ménard et al., 2015), regional (Galbraith et al., 2010; Burke et al., 2013; Chadburn et al., 2015) and globally as part of model-intercomparison studies (Anav et al., 2015; Sitch et al., 2015). ~~Since biased simulations of GPP can introduce errors into other ecosystem processes in JULES, examining how the model performs at global and regional scales compared to GPP estimates from~~ Evaluating simulated GPP at a range of scales and its sensitivity to spatial resolution and meteorological data is essential for informing future model developments. In this manuscript, we do this using two observation-based datasets (FLUXNET-MTE and MODIS) and the CARbon DAta MOdel fraMework (Bloom et al., 2016, CARDAMOM)~~and how sources of uncertainty, such as the sensitivity of the model to spatial resolution and the meteorological data used to drive the model, affect model performance is an important part of model development~~.

In this study, the ability of JULES version 3.4.1 to simulate global and regional fluxes of GPP for various biomes, spatial resolutions and using different meteorological data to drive the model is evaluated. In particular, the following research questions are addressed:

- ~~Can JULES capture interannual variability of GPP at the global scale?~~ How do estimates of global GPP compare to those from ~~observational-based~~ the observation-based datasets and the CARDAMOM framework? Can JULES capture interannual variability of GPP at the global scale?
- How does JULES GPP compare for various biomes at the global and regional scales?
- How sensitive are fluxes of GPP to the spatial resolution of the model?
- Is the meteorological data set used to drive the model important at the global scale?

2 Methods and model

2.1 Model description

The Joint UK Land Environment Simulator (JULES) is the land surface scheme of the UK Met Office Unified Model (MetUM~~, current version 10.2~~), which is a single model family used to simulate weather and climate across a range of timescales ~~(Walters et al., 2014)~~(Walters et al., 2016). JULES is a community land surface model which has evolved from the Met Office Surface Exchange Scheme (MOSES) (Cox et al., 1999) and is used for modelling all of the processes at the land surface, in the sub-surface soil, and surface exchange processes (Best et al., 2011; Clark et al., 2011). JULES can be used *offline* (i.e. outside of the host ESM, MetUM) and model simulations can be performed at point, regional and global scales. ~~In LSMs,~~ Plant Functional Types (PFTs) are used to represent broad groupings of plant species with similar ecosystem functions and resource use. In the version of JULES used in this study (version 3.4.1), each model gridbox consists of 9 different surface types; 5 PFTs (broadleaf trees, needleleaf trees, C3 (temperate) grass, C4 (tropical) grass and shrubs) and 4 non-vegetation surface types (urban, inland water, bare soil and land-ice). Model gridboxes can consist entirely of a mixture of the first 8 surface types or only land-ice. Since model version 4.2, each JULES gridbox can contain nine PFTS (tropical broadleaf evergreen, temperate broadleaf evergreen, broadleaf deciduous, needleleaf evergreen, needleleaf deciduous, C3, C4, evergreen shrub, deciduous shrub) (Harper et al., 2016).

~~The~~ JULES is driven by the downward shortwave and longwave radiation fluxes, rainfall and snowfall rates, surface air temperature, wind speed, surface pressure and specific humidity. The downward shortwave and longwave radiation fluxes play an important role in the surface energy balance, where the downwelling radiation fluxes must equal the outgoing fluxes of sensible heat, latent heat, ground flux, reflected shortwave radiation and upwelling thermal energy, and the calculation of photosynthesis (Best et al., 2011; Clark et al., 2011). GPP is the total C used by plants in photosynthesis at the canopy scale with potential (without water and ozone stress) leaf-level photosynthesis ~~(A) is calculated using the Cand Cphotosynthesis models of Collatz et al. (1991) and Collatz et al. (1992), respectively. This is calculated~~ calculated as the smoothed minimum of three limiting rates: (1) Rubisco-limited rate (determined using surface air temperature and atmospheric CO_2 concentrations), (2) Light-limited rate (determined using downward radiation fluxes) and (3) Rate of transport of photosynthetic products (C_3 plants) and PEP-Carboxylase limitation (C_4 plants) ~~(Clark et al., 2011). Leaf photosynthesis is linked to stomatal conductance via the diffusion equation using the Jacobs (1994) formulation.~~ (determined using surface air temperature and pressure) (Clark et al., 2011). By taking soil moisture stress into account, leaf-level photosynthesis ~~(A)~~ is calculated by multiplying the potential leaf-level photosynthesis ~~(A)~~ by a soil moisture factor ~~, β (A= Aβ). The effect of on leaf photosynthesis can also be included when calculating A, but it is not shown here.~~ (determined using mean soil moisture concentration in the root zone and thus, precipitation).

~~There~~ In JULES, there are two options available ~~in JULES~~ for radiation interception and the scaling of photosynthesis from leaf-level to canopy-level: (i) big leaf approach and (ii) multi-layer approach. For all model simulations performed in this study, the multi-layer approach was used ~~. With the multi-layer approach, there are four variations (Table 3 of Clark et al. (2011)) that consider the vertical profile~~ which takes into account the vertical gradient of canopy photosynthetic capacity ~~, light inhibition of leaf respiration, the inclusion of sunfleck penetration and the division of canopy layers into sunlit and shaded leaves. Option 4 (from Table 3 of Clark et al. (2011)) was used for the model simulations performed in this study. This option includes the decrease in photosynthetic capacity from~~ (decreasing leaf nitrogen from top to bottom of ~~the canopyand the~~ canopy) and includes light inhibition of leaf respiration ~~in light. In the multi-layer approach, the amount of radiation absorbed and photosynthesis are estimated for a number of user defined canopy layers (dL= L/n, where Lis the canopy leaf area and dLis the canopy layer leaf area~~(Option 4 in Table 3 of Clark et al. (2011)). Canopy-scale fluxes are estimated to be the sum of the leaf-level fluxes in each canopy layer, scaled by leaf area. ~~Clark et al. (2011) contains a more detailed description of leaf-level photosynthesis and its scaling up to the canopy-level~~LAI is calculated for each canopy level (default number is 10), with a maximum LAI prescribed for each PFT.

Phenology (bud burst and leaf senescence) in JULES is usually updated once per day by multiplying the annual maximum LAI by a scaling factor (calculated using accumulated temperature-dependent leaf turnover rates). For each PFT, the C fluxes are calculated using a coupled photosynthesis-stomatal conductance model on each model timestep (typically 30 to 60 minutes) (Cox et al., 1998). These fluxes are then time-averaged (usually every 10 days) before being passed to TRIFFID (Top-down Representation of Interactive Foliage and Flora Including Dynamics), JULES' dynamic global vegetation model, which updates the vegetation distribution, based on the net C available to it and competition with other vegetation types, and soil C in each model gridbox on a longer timestep (usually every 10 days) (Cox, 2001). ~~Biophysical properties of the land surface, such as albedo and roughness length, are updated based on the new vegetation state, so that these properties can affect the atmosphere (Clark et al., 2011).~~ Clark et al. (2011) and Best et al. (2011) contain a more detailed description of JULES.

2.2 Experimental design

Offline simulations of GPP were performed at the global scale for the 2001–2010 period using various meteorological datasets and spatial resolutions (Table 1). A general overview is provided of how sensitive JULES GPP is to the meteorological dataset used at global scales rather than for each meteorological variable. By analysing the models sensitivity to each meteorological dataset, different analyses of the global climate are compared and therefore a multi-factor analysis of combined changes in meteorological variables can be performed. The land cover was kept constant at values for the year 2000 (Loveland et al., 2000) and annual atmospheric CO_2 concentrations were varied as in the historical record. ~~In LSMs, the correct simulation of GPP is important since errors in its calculation can lead to errors in biomass and other land-atmosphere flux calculations, such as Net Ecosystem Exchange (NEE), and latent and sensible heat fluxes (Schaefer et al., 2012; Slevin et al., 2015).~~ The 2001–2010 time period was used to due to the availability of ~~two global meteorological datasets used to drive JULES and estimates of GPP , from two observation-based datasets and the CARDAMOM framework, used to evaluate model performance~~multiple global meteorological and GPP datasets for this period. JULES is ~~being benchmarked against the upscaled FLUXNET~~compared against FLUXNET-MTE, MODIS and CARDAMOM GPP.

Prior to performing the global scale model simulations, the soil moisture was brought to equilibrium using a 40 year global spin-up by cycling 10 years of meteorological data (1979–1989) twice and 10 years of meteorological data (1989–1999) twice (in equilibrium mode), followed by a 12 year spin-up by cycling 12 years of meteorological data (1999–2010) once (in dynamical mode). ~~Finally, the actual model simulations were performed for 2001–2010 due to the availability of multiple global meteorological datasets and the observation-based (upscaled FLUXNET and MODIS) and CARDAMOM estimates of global GPP for this period.~~ Clark et al. (2011) contains more information on spinning up the soil C pools.

2.3 Data

The datasets used in this study include those used as input to JULES (soil, vegetation and meteorological data) and the benchmarking data~~(FLUXNET-MTE, MODIS and CARDAMOM GPP) against which model performance is compared~~. The soil dataset used was the Harmonized World Soil Database version 1.2 (Nachtergaele et al., 2012, HWSD) and ~~the~~ contains soil property data such as soil texture fractions, water storage capacity, soil depth and pH (Nachtergaele et al., 2012). In this study, the soil texture fractions (% of sand, silt and clay) were used to calculate the soil thermal and hydraulic conductivity parameters listed in Table 3 of Best et al. (2011). The land cover classification scheme used for specifying the PFT fractions for each model gridbox at the global scale was Global Land Cover Characterization database version 2.0 (Loveland et al., 2000, http://edc2.usgs.gov/glcc/glcc.php). ~~The meteorological data~~ Two meteorological datasets were used to drive ~~JULES~~ the model offline (i.e. run separately from its host Earth System Model) ~~includes the downward shortwave and longwave radiation fluxes (), rainfall and snowfall rates (), surface (2 m) air temperature (K), wind speed (), surface air pressure (Pa) and specific humidity (). Since JULES requires meteorological data at 6 hourly intervals or less in order to drive the model offline, a number of datasets were used~~at global scales; WFDEI (Weedon et al., 2014) and PRINCETON (Sheffield et al., 2006).

Global gridded estimates of GPP derived from the upscaling of observations from the FLUXNET network of tower sites (Jung et al., 2009), estimates from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, aboard the U.S. National Aeronautics and Space Administration (NASA) Earth Observation System (EOS) satellites, Terra and Aqua (Yang et al., 2006), and GPP simulated by the CARbon DAta MOdel fraMework (Bloom et al., 2016, CARDAMOM) ~~are~~ were used to evaluate model performance. These global gridded estimates of GPP provide a means to evaluate LSMs at large scales (Jung et al., 2009, 2010; Beer et al., 2010; Zhao and Running, 2010; Bonan et al., 2011; Lei et al., 2014).

2.3.1 Forcing data

~~The WFDEI meteorological driving data was created using the same methodology as the WATCH Forcing Data (WFD) applied to the ERA-Interim reanalysis data for the 1979–2012 period (Weedon et al., 2014). The WFD data set (1901–2001) was created as part of the EU Water and Global Change (WATCH) project (Harding et al., 2011, http://www.eu-watch.org/) and was derived from the European Centre for Medium-range Weather Forecasts (ECMWF) ERA-40 reanalysis for 1958–2001 with data for 1901–1957 derived using randomly selected years from the ERA-40 data (Weedon et al., 2010, 2011).~~ As part of the EMBRACE EU FP7 programme (http://www.embrace-project.eu/), the ~~WFD~~ WATCH Forcing Data (WFD) methodology was applied to the ERA-Interim reanalysis data for the 1979–2013 period to generate the WFDEI meteorological forcing data (Weedon et al., 2014). As for the WFD, WFDEI has two precipitation products, corrected using either CRU (Climate Research Unit at the University of East Anglia) or GPCC (Global Precipitation Climatology Centre) precipitation totals (Weedon et al., 2014) and are referred to as WFDEI-CRU and WFDEI-GPCC, respectively. The GPCC data product is a gridded gauged precipitation dataset and provides a higher resolution dataset (i.e. better station coverage, particularly at high latitudes, and especially for the end of the 20$^{\text{th}}$ century) than the CRU precipitation totals (Weedon et al., 2014). The WFDEI dataset consists of 3 hourly, regularly gridded data at half-degree ($0.5° \times 0.5°$) spatial resolution and is only available for land points including Antarctica. The dataset contains the following meteorological variables: downward shortwave and longwave radiation fluxes ($\mathrm{W\,m^{-2}}$), rainfall rate ($\mathrm{kg\,m^{-2}\,s^{-1}}$), snowfall rate ($\mathrm{kg\,m^{-2}\,s^{-1}}$), $2\,\mathrm{m}$ temperature (K), $10\,\mathrm{m}$ wind speed ($\mathrm{m\,s^{-1}}$), surface pressure (Pa) and $2\,\mathrm{m}$ specific humidity ($\mathrm{kg\,kg^{-1}}$).

The PRINCETON dataset is a global 62 year near-surface meteorological data set used for driving land surface models and was created by Princeton University's Terrestrial Hydrology Group (Sheffield et al., 2006, http://hydrology.princeton.edu/home.php). The PRINCETON data set consists of 3 hourly, regularly gridded data at 1-degree ($1° \times 1°$) spatial resolution for the 1948–2010 period and is only available for land points excluding Antarctica. The dataset contains the ~~following meteorological variables : downward shortwave and longwave radiation fluxes (),~~ same meteorological variables as WFDEI with the exception of rainfall and snowfall rates summed as total precipitation ($\mathrm{kg\,m^{-2}\,s^{-1}}$)~~, air temperature (K), 10 wind speed (), surface pressure () and specific humidity ()~~.

2.3.2 Benchmarking data

The upscaled FLUXNET GPP (hereafter referred to as FLUXNET-MTE) was derived using a model tree ensemble (MTE) approach, a type of machine learning technique that can be trained to predict land-atmosphere fluxes (Jung et al., 2009)~~and provides a means to evaluate LSMs at large scales (Jung et al., 2009, 2010; Beer et al., 2010; Jung et al., 2011; Bonan et al., 2011)~~. Based on observed meteorological data, land cover data and remotely sensed vegetation properties (fraction of absorbed photosynthetic active radiation), the upscaling principle ~~uses a machine learning algorithm (model tree ensembles) to~~ can predict estimates of C fluxes at FLUXNET sites with available quality-filtered flux data and the trained model is then applied spatially using grids of the input data (Jung et al., 2009, 2011). However, these machine learning algorithms are typically data limited due to the quantity, quality and representativeness of the training dataset (Jung et al., 2009). There are two upscaled FLUXNET GPP datasets available depending on the flux partitioning method used to separate net ecosystem exchange of CO_2 (NEE) into GPP and terrestrial ecosystem respiration (TER) (Reichstein et al., 2005; Lasslop et al., 2010). In this study, GPP based on the work by Reichstein et al. (2005) was used (this is the flux partitioning method used by the FLUXNET network). However, differences between the two upscaled FLUXNET GPP datasets are small. FLUXNET-MTE is a $0.5^\circ \times 0.5^\circ$ spatial and monthly temporal resolution data set for the 1982–2011 period and is available from the Max Planck Institute for Biogeochemistry Data Portal (https://www.bgc-jena.mpg.de/geodb/projects/Home.php).

The MOD17 MODIS Gross/Net Primary Productivity (GPP/NPP) product provides continuous estimates of GPP/NPP for the Earth's entire land surface and is produced as part of the NASA's Earth Observing System (EOS) program. The MOD17 algorithm produces two subproducts, MOD17A2 (which stores 8-day composite GPP, net photosynthesis and QC flags) and MOD17A3 (annual NPP and QC flags) (Zhao et al., 2005). The resulting datasets contain regular gridded global estimates of GPP and NPP for the terrestrial land surface at the 1 km spatial resolution (Running et al., 2000). ~~Sources of error in the Collection 4 primary production include mismatching spatial resolution between the gridded meteorological data (1° × 1.25°) and MODIS pixels (1 km), contaminated or missing 8-day LAI/FPAR (MOD15A2 MODIS product) due to cloud cover or sensor malfunction and misclassified land cover from the MODIS land cover product (MOD12Q1) which can result in incorrect parameters from the MOD17 Biome Property Look-Up Table (BPLUT) and therefore leads to incorrect GPP estimates (Zhao et al., 2005).~~ The Numerical Terradynamic Simulation Group (NTSG) (http://www.ntsg.umt.edu/project/mod17) at the University of Montana ~~rectified these problems~~ corrected problems associated with GPP estimates by spatial interpolation of the coarse resolution meteorological data, temporal infilling of cloud-contaminated MOD15A2 LAI/FPAR data and modification of BPLUT (Biome Property Look-Up Table) parameters based on observed GPP from flux tower measurements in order to create an improved MOD17 GPP product (Zhao et al., 2005). The global monthly MODIS GPP (version 55) dataset at $0.05° \times 0.05°$ spatial resolution for the 2001–2010 period was downloaded from ~~the NTSG ftp server~~ (ftp://ftp.ntsg.umt.edu/pub/MODIS/NTSG_Products/). For the purposes of this study, the data was regridded to $0.5° \times 0.5°$ spatial resolution using the first order conservative remapping function (remapcon) of the Climate Data Operators (CDO) software package (https://code.zmaw.de/projects/cdo).

The CARbon DAta MOdel fraMework (CARDAMOM) is a model-data fusion approach which consists of merging observational data with models in order to improve model quality and characterise its uncertainty (Bloom and Williams, 2015; Bloom et al., 2016). CARDAMOM relies on a Bayesian Markov Chain Monte Carlo (MCMC) algorithm to explore the parametric uncertainty of the ecosystem C balance model Data Assimilation Linked Ecosystem Carbon Model version two (Bloom et al., 2016, DALEC2) according to available C relevant data-streams (fluxes, leaf area index, changes in biomass, etc.). CARDAMOM can be applied at the point-scale and spatially with available remote-sensing based products such as MODIS LAI, biomass and soil carbon maps. When the framework is applied spatially, the Bayesian model-data fusion approach is performed in every model gridbox independently without using pre-defined biome maps. C fluxes, pool increments and parameter values with explicit confidence intervals attached to them are output from the MCMC algorithm. In this study, MODIS LAI, a tropical biomass map (Saatchi et al., 2011), a soil C dataset (Hiederer and Köchy, 2011), MODIS burned area (Giglio et al., 2013) and the ERA-Interim reanalysis data have been used as input to CARDAMOM in order to produce a global monthly mean GPP dataset at $1° \times 1°$ spatial resolution for the 2001–2010 period (Bloom et al., 2016).

2.4 Outline of experiments

This section describes the model simulations performed in this study (Table 1). For the JULES model simulations, the first part of the model simulation name refers to JULES version 3.4.1 and the second part refers to the global gridded meteorological dataset used to drive the model (Table 1). The spatial resolution of the model grid is appended to the end of the model simulation name. Model simulation names without an attached spatial resolution mean that the model simulation was performed at $0.5° \times 0.5°$ spatial resolution. ~~The TRIFFID~~Vegetation competition (simulated by TRIFFID, JULES' dynamic global vegetation model~~(a submodel of JULES which simulates vegetation competition) and vegetation competition have~~) has been switched off for ~~all JULES' simulations. Vegetation competition was disabled~~ the majority of model simulations. This was done in order to prevent unrealistic vegetation fractions in model gridboxes for global scale simulations of GPP. Differences between having prescribed PFTS (no vegetation competition) and allowing competition between PFTs was also examined. For the CARDAMOM simulation, the ERA-Interim reanalysis product was used to drive the DALEC2 model at $1° \times 1°$ resolution. Model results were compared to FLUXNET-MTE, MODIS and CARDAMOM GPP.

2.4.1 ~~Interannual variability of GPP~~

Firstly, model estimates of total annual GPP (JULES-WFDEI-GPCC) were integrated globally. The ability of JULES to simulate the interannual variability (IAV) of GPP at ~~the global scale was examined . Model simulations were performed for~~ global scales was examined from 2001–2010 ~~using global parameter and meteorological datasets~~ (JULES-WFDEI-GPCC; Table 1)~~with the results compared to GPP from observation-based estimates (FLUXNET-MTE and MODIS) and CARDAMOM.~~

2.4.1 ~~Global GPP~~

~~Model estimates of total annual GPP (JULES-WFDEI-GPCC) integrated across the globe were compared to FLUXNET-MTE, MODIS and CARDAMOM GPP. FLUXNET-MTE (global flux datasets derived from individual flux tower sites) and MODIS (satellite datasets) provide a means to evaluate JULES (and other LSMs) at global and regional scales (Bonan et al., 2011).~~

2.4.1 ~~Global and regional comparison for various biomes~~

~~In addition to deriving estimates of globally integrated GPP fluxes, the modelled (JULES-WFDEI-GPCC), FLUXNET-MTE, MODIS and CARDAMOM GPP were~~ . Secondly, the modelled and observation-based estimates of GPP were further compared by biome type (Forest, Grassland and Shrub) at ~~the~~ global and regional scales (Global, Tropics and Extratropics). ~~The global GPP was further~~ GPP was analysed by biome type at ~~the regional scale~~ regional scales by dividing the global land area into seven regions (Figure 1; Table 2).

2.4.1 ~~Sensitivity to the spatial resolution of the input data~~

~~The~~ Thirdly, the sensitivity of the model to the spatial resolution of the input data was evaluated by varying the resolution of the ancillary data (soil and vegetation) and meteorological data (WFDEI-GPCC) ~~and re-running the model simulations for 2001–2010~~(Table 1). The input data was regridded from $0.5° \times 0.5°$ to $1° \times 1°$ spatial resolution (JULES-WFDEI-GPCC-1degree; Table 1) and from $0.5° \times 0.5°$ to $2° \times 2°$ spatial resolution (JULES-WFDEI-GPCC-2degree) using ~~the first order conservative remapping function (remapcon) of the Climate Data Operators (CDO) software package (). The observation-based (FLUXNET-MTE and MODIS)~~ CDO. Further information on how the datasets were regridded ~~using this method. The output from these simulations were compared to those at $0.5° \times 0.5°$ spatial resolution (JULES-WFDEI-GPCC; Table 1).~~ can be found in Appendix D of Slevin (2016).

2.4.1 ~~Sensitivity to the meteorological driving data set~~

~~The~~ Finally, the sensitivity of JULES to the meteorological driving data was evaluated by comparing model simulations driven using the WFDEI-GPCC (JULES-WFDEI-GPCC-1degree; Table 1) and PRINCETON datasets (JULES-PRINCETON; Table 1) at $1° \times 1°$ spatial resolution ~~. In these model simulations, the same ancillary datasets are~~ (the same soil and vegetation ancillary datasets were used by both~~with the only difference in the modelsimulations being the meteorological data used to drive the model~~). The model's sensitivity to precipitation was examined by driving it with the WFDEI-CRU dataset (JULES-WFDEI-CRU; Table 1) at $0.5° \times 0.5°$ spatial resolution.

2.5 Model analyses

In order to quantify how the model performs at the global scale, the following metrics were used: global area-weighted mean ($\bar{x}$; Equation 1), Coefficient of Variation (CV; Equation 2) and monthly anomalies (Equation 3).

$$\bar{x} = \frac{\sum_{i,j=1}^{i=m,j=n} a_{i,j}\, x_{i,j}}{\sum_{i,j=1}^{i=m,j=n} a_{i,j}} \tag{1}$$

The global area-weighted mean is calculated by multiplying the monthly GPP flux for each grid box ($x_{i,j}$) by the area of its grid box ($a_{i,j}$) and dividing the sum of these values by the total land surface area. m and n are the total number of grid boxes in the x- and y-direction, respectively. For example, when running a global scale model simulation at half-degree ($0.5° \times 0.5°$) spatial resolution, $m = 720$ (number of grid boxes in the west-east direction) and $n = 360$ (number of grid boxes in the north-south direction).

$$\mathrm{CV} = \frac{\sigma}{\mu} \times 100 \tag{2}$$

CV (also known as relative variability) is a measure of the relative magnitude of the standard deviation (σ) and is calculated by dividing the standard deviation by the mean (μ). It is expressed as a percentage and is always positive. CV is a useful statistic since it allows the degree of variation of various datasets to be compared even if the means are quite different from each other. It is also dimensionless which means that CVs can be used to compare the dispersion (variability) of the data when other measures like standard deviation or root mean squared error cannot.

To quantify model performance at the global scale, CV was calculated by first computing the standard deviation and means of the global area-weighted means for each month and then dividing the average of the standard deviations by the average of the means for each month.

$$\text{Monthly anomaly} = x - \bar{x}_{clim} \tag{3}$$

The monthly anomaly is defined as the departure of the observed monthly values (x) from the long-term (climatological) average for that month ($\bar{x}_{clim}$).

3 Results

3.1 ~~Interannual variability of~~ Global GPP

~~JULES simulates the seasonal~~ In general, JULES simulates higher annual average global GPP than MODIS, FLUXNET-MTE and CARDAMOM with JULES GPP closer to FLUXNET-MTE estimates. When driven with the WFDEI-GPCC dataset (JULES-WFDEI-GPCC; Table 1), JULES simulates global GPP with an annual average of 140 $\mathrm{Pg\,C\,year^{-1}}$ (the combined GPP of all terrestrial ecosystems) over the 2001–2010 period (Figure 2c). This value is greater than that estimated by

MODIS, FLUXNET-MTE and CARDAMOM with annual average global GPP estimated to be 112, 130 and 114 $\mathrm{Pg\,C\,year^{-1}}$, respectively, for the same period (Figures 2a, b and d). The higher global GPP simulated by the JULES-WFDEI-GPCC driven simulations is greater than the MODIS, FLUXNET-MTE and CARDAMOM estimates by 25 %, 8 % and 23 % on average, respectively.

The difference in average annual global GPP between JULES-WFDEI-GPCC and MODIS (both at $0.5° \times 0.5°$ spatial resolution) is greater (28 $\mathrm{Pg\,C\,year^{-1}}$) than that between JULES-WFDEI-GPCC and FLUXNET-MTE (10 $\mathrm{Pg\,C\,year^{-1}}$) and between JULES-WFDEI-GPCC and CARDAMOM (26 $\mathrm{Pg\,C\,year^{-1}}$). This difference between the model simulated and observation-based GPP estimates is also shown in the zonal mean of the total annual JULES-WFDEI-GPCC, MODIS, FLUXNET-MTE and CARDAMOM GPP with the largest differences between datasets found in the tropics at 10°S-10°N and 15°N-30°N (Figure 2e).

3.2 Seasonal and interannual variability of GPP

Overall, it was found that JULES can simulate simulate seasonal and interannual variability of GPP at global scales. JULES simulates the seasonal cycle of GPP (JULES-WFDEI-GPCC; Table 1) ~~at the global scale~~ (Figure 3a) with the global ~~area-weighted~~ average of its monthly GPP for 2001–2010 falling within range of the observation-based estimates (FLUXNET-MTE and MODIS) for much of the year (between 64 and 107 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$). A similar trend can be found with the CARDAMOM GPP (Figure 3a). The exception to this are the Northern Hemisphere winter months (January, February, March and December) with JULES simulating higher global mean GPP by 2 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$ on average compared to FLUXNET-MTE. The MODIS GPP means are lower than FLUXNET-MTE for each of the monthly climatologies by 10 $\mathrm{g\,C\,m^{-2}\,month^{-1}}$ on average (Figure 3a).

The standard deviation of the monthly GPP fluxes is used to measure interannual variability and this is expressed as a percentage of the mean monthly GPP fluxes using coefficient of variation (CV). Low values of CV mean that differences between the monthly GPP fluxes and the mean monthly GPP fluxes are small and larger CV values mean the opposite. The CV of the model simulated and observation-based GPP fluxes range between 0.8–4 % for the mean monthly GPP with the highest differences between the monthly values being for Northern Hemisphere winter and spring (February, March, April, November and December) (Figure 3b). This pattern is similar to the global ~~area-weighted~~ average of the monthly climatologies (Figure 3a). ~~Low values of CV mean that differences between the monthly GPP fluxes and the mean monthly GPP fluxes are small and larger CV values mean the opposite.~~

The monthly anomalies (computed using the global ~~area-weighted~~ mean values) expressed as percentages of the global ~~area-weighted mean of the model simulated~~ mean of model simulated monthly GPP (JULES-WFDEI-GPCC) compare equally well to both FLUXNET-MTE and MODIS GPP for 2001–2010 with both having Root Mean Squared Errors (RMSEs) of 2.4 % with CARDAMOM having much lower year to year variation (Figure 3c). ~~However, the high variation in JULES GPP at the beginning and end of the year is observed in the~~ The model is able to capture simulated monthly anomalies from 2001 to 2010 ~~which in some years, such as~~ with the exception of those in 2002 ~~, the model is unable to capture monthly GPP~~ (Figure 3c).

~~Overall, it was found that JULES can simulate interannual variability at global scales.~~

~~3.3 Global GPP~~

~~When driven with the WFDEI-GPCC dataset (JULES-WFDEI-GPCC; Table 1), JULES simulates global GPP with an annual average of 140 (the combined GPP of all terrestrial ecosystems) over the 2001–2010 period (Figure 2e). This value is greater than that estimated by MODIS, FLUXNET-MTE and CARDAMOM with annual average globl GPP estimated to be 130, 112 and 114 , respectively, for the same period (Figures 2a, b and d). The higher global GPP simulated by the JULES-WFDEI-GPCC driven simulations is greater than the MODIS, FLUXNET-MTE and CARDAMOM estimates by 25 %, 8 % and 23 % on average, respectively.~~

~~The difference in average annual global GPP between JULES-WFDEI-GPCC and MODIS (both at 0.5° × 0.5° spatial resolution) is greater (28) than that between JULES-WFDEI-GPCC and FLUXNET-MTE (10) and between JULES-WFDEI-GPCC and CARDAMOM (26). This difference between the model simulated and observation-based GPP estimates is also shown in the zonal mean of the total annual JULES-WFDEI-GPCC, MODIS, FLUXNET-MTE and CARDAMOM GPP with the largest differences between datasets found in the tropics at 10°S-10°N and 15°N-30°N (Figure 2e).~~

~~In general, JULES simulates higher annual average global GPP than MODIS, FLUXNET-MTE and CARDAMOM with JULES GPP closer to FLUXNET-MTE GPP estimates.~~

3.3 *Global and regional comparison of simulated GPP for various biomes*

In addition to examining the ability of JULES to simulate global GPP (integrated across all ecosystem types), the total annual GPP for 2001–2010 was compared for various biomes (forests, grasslands and shrubs) at global and regional scales (Figure 4). This means that areas for model improvement can be identified at scales smaller than the global. JULES overestimates GPP in all tropical land areas (Central and South America, Africa and South and South-East Asia), but is able to simulate it in the extratropics (Europe, Northern Asia, North America and Greenland and the Extratropical Southern Hemisphere) (Figure 4).

When JULES was driven with WFDEI-GPCC (JULES-WFDEI-GPCC), JULES simulated ~~total~~ average annual GPP to be ~~612~~61~~, 536~~Pg C year^{-1}, 54 ~~and 74~~Pg C year^{-1} and 7 Pg C year^{-1} for forests, grasslands and shrubs, respectively ~~with average annual GPP for forests, grasslands and shrubs being 61 , 54 and 7 , respectively~~ (Figure 4a). With the exception of shrubs, JULES overestimates ~~total~~ average annual GPP by 31 %, 13 % and 22 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for forests and by 27 %, 10 % and 31 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively, for grasslands (Figure 4a). Differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP for shrubs are small with ~~total~~ average annual GPP ranging within ~~68–78~~7–8 Pg C year^{-1} (Figure 4a).

The differences in total annual GPP at the global scale is mainly due to differences between model simulated (JULES and CARDAMOM) and the observation-based estimates (MODIS and FLUXNET-MTE) in the tropics (30°S–30°N) (Figure 4b). In the tropics, JULES simulates total annual GPP to be ~~501~~50~~, 387~~Pg C year^{-1}, 39 ~~and 52~~Pg C year^{-1} and 5 Pg C year^{-1} for forests, grasslands and shrubs, respectively, for the 2001–2010 period. JULES overestimates total annual GPP by ~~41~~20–41 % ~~, 20 % and 35 %~~ compared to MODIS, FLUXNET-MTE and CARDAMOM GPP ~~, respectively,~~ for forests and by ~~47~~24–54 %

~~, 24 % and 54 % compared to MODIS, FLUXNET-MTE and CARDAMOM GPP, respectively,~~ for grasslands in the tropical regions (Figure 4b). Differences between model simulated and observation-based estimates of GPP are small in the tropics for shrubs with total annual GPP ranging from ~~44–52~~4–5 Pg C year^{-1} (Figure 4b). In the extratropics (30°N–90°N and 30°S–90°S), differences between model and observed GPP are small with ~~total~~ average annual GPP for forests, grasslands and shrubs found to be ~~111–128~~11–13 ~~, 148–178~~Pg C year^{-1}, 15–18 ~~and 21–28~~Pg C year^{-1} and 2–3 Pg C year^{-1}, respectively (Figure 4c).

Total annual GPP at the regional scale was examined by ~~dividing~~ splitting the land area into seven regions (Figure 1; Table 2). The tropical regions (30°S–30°N) have been further divided up into three regions; Central and South America, Africa and South and South-East Asia. The extratropics (30°N–90°N and 30°S–90°S) have been divided into four regions; Europe, Northern Asia, North America and Greenland and the extratropical Southern Hemisphere. ~~By normalising the~~ JULES overestimates GPP in all three tropical land areas compared to MODIS, FLUXNET-MTE ~~and MODIS GPP by JULES GPP for these three regions (Figure 6), it is easier to see the differences between the model simulated and observation-based estimates of GPP . The dashed line at y=1 for each of the seven regions in Figure 6represents where model and observation-based total annual GPP match.~~

and CARDAMOM (Figures 6c, e and f). Differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP with average annual GPP range from 7.4–12.1 Pg C year^{-1}, 7.7–13 Pg C year^{-1} and 1–1.3 Pg C year^{-1} for forests, grasslands and shrubs, respectively, in South and South-East Asia, 9.5–13.7 Pg C year^{-1}, 8.4–12.3 Pg C year^{-1} and 1.7–2.1 Pg C year^{-1} for forests, grasslands and shrubs, respectively, in Africa and 18-23.2 Pg C year^{-1}, 9–12.9 Pg C year^{-1} and 1.4–1.8 Pg C year^{-1} for forests, grasslands and shrubs, respectively, in Central and South America (Figures 6c, e and f, respectively). In the extratropics, differences between JULES, MODIS, FLUXNET-MTE and CARDAMOM GPP are small with ~~total~~ average annual GPP ranging from ~~16–20~~1.6–2 ~~and 42–55~~Pg C year^{-1} and 4–5 Pg C year^{-1} for forests and grasslands, respectively, in Europe, ~~42–53~~4–5 ~~and 46–59~~Pg C year^{-1} and 4–6 Pg C year^{-1} for forests and grasslands, respectively, in Northern Asia, ~~6–8~~0.6–0.8 Pg C and ~~13–18~~1.3–1.8 Pg C year^{-1} for forests and grasslands, respectively, in the Extratropical Southern Hemisphere and ~~42–49~~4–5 ~~and 37–44~~Pg C year^{-1} and 3–5 Pg C year^{-1} for forests and grasslands, respectively, in North America and Greenland (Figures 6a, b, d and g, respectively).

~~In general, JULES overestimates GPP in the tropics, but is able to simulate it in the extratropics.~~

3.4 Sensitivity to spatial resolution

When simulating GPP at global and regional scales, there was little impact from varying spatial resolution ($0.5° \times 0.5°$, $1° \times 1°$ and $2° \times 2°$) (Figure 5). When simulations of GPP were performed at lower spatial resolutions (JULES-WFDEI-GPCC-1degree and JULES-WFDEI-GPCC-2degree; Table 1), the average annual global GPP at $0.5° \times 0.5°$, $1° \times 1°$ and $2° \times 2°$ spatial resolutions was 140 Pg C year^{-1}, 141 Pg C year^{-1} and 142 Pg C year^{-1}, respectively. The percentage differences between JULES and the observation-based GPP estimates (MODIS and FLUXNET-MTE) at the various spatial resolutions are approximately equal with JULES differing from MODIS and FLUXNET-MTE by 8 % and 25 %, respectively, at $0.5° \times 0.5°$ spatial resolution, by 8 % and 26 %, respectively, at $1° \times 1°$ resolution and by 9 % and 26 %, respectively, at $2° \times 2°$ resolution.

The zonal mean of modelled total annual GPP at various spatial resolutions are approximately equal (Figure 5). This insensitivity to spatial resolution is also found at regional scales (Figure 6). This insensitivity to spatial resolution is a useful result since it means that model simulations can be performed at $2° \times 2°$ resolution with little difference to model output from the simulations at $0.5° \times 0.5°$ and due to the lower computational cost, model run times (at $2° \times 2°$ resolution) are short (approximately $16\times$ faster than the $0.5° \times 0.5°$ resolution simulations).

~~When simulating GPP at global and regional scales, there was little impact from varying spatial resolution ($0.5° \times 0.5°$,$1° \times 1°$ and$2° \times 2°$).~~

3.5 Sensitivity to meteorological data set

When JULES was driven with the PRINCETON dataset, simulated global GPP was found to be higher than that simulated using WFDEI-GPCC by 3 Pg C year^{-1} on average with the largest differences occurring in the tropics (Figures 5, 8a and 8d). When driven with the PRINCETON dataset (JULES-PRINCETON; Table 1), JULES simulates global GPP with an annual average of 144 Pg C year^{-1} for the 2001–2010 period ~~.~~ (Figure 8d).

As observed when driving JULES with the WFDEI-GPCC dataset (Figure 2), JULES-PRINCETON simulates higher global GPP than MODIS, FLUXNET-MTE and CARDAMOM at $1° \times 1°$ spatial resolution by ~~29~~11–29 %~~, 11 % and 26 % on average, respectively~~. This compares quite well to global GPP simulated by JULES when driven with the WFDEI-GPCC dataset, which had an annual average global GPP of 140 Pg C year^{-1}. GPP simulated by JULES-WFDEI-GPCC was only higher than that of MODIS, FLUXNET-MTE (both at $0.5° \times 0.5°$ spatial resolution) and CARDAMOM (at $1° \times 1°$ resolution) by ~~25~~8–25 %~~, 8 % and 23 % on average, respectively. The trend~~ . The pattern in zonal mean of total annual GPP simulated by the model (when driven with PRINCETON) is similar to that when driven with WFDEI-GPCC (at $1° \times 1°$ spatial resolution) with differences ~~being mostly in the tropics (Figure 5). The trend in differences~~ between JULES-PRINCETON and JULES-WFDEI-GPCC-1degree and the observation-based estimates (MODIS and FLUXNET-MTE) ~~is similar with model output from both simulations overestimating GPP~~ being mostly in the tropics (Figure 5).

There is little difference in simulated GPP when using either WFDEI-GPCC or WFDEI-CRU (which differ only in the precipitation product used) to drive JULES (Figure 4; Figure G.2 in Slevin (2016)). When driven with the WFDEI-CRU dataset, JULES simulates global GPP with an annual average of ~~141~~142 Pg C year^{-1} (the combined GPP of all terrestrial ecosystems) over 2001–2010 (Figure G.3 in Slevin (2016)). This is ~~1~~2 Pg C year^{-1} higher than that simulated when JULES is driven with WFDEI-GPCC (140 Pg C year^{-1}). The small differences in global GPP can also found at regional scales in both the tropical and extratropical regions (Figures 4b and c, respectively).

~~In general, when JULES is driven with the PRINCETON dataset, simulated global GPP was found to be higher than that simulated using WFDEI-GPCC by 3,on average with the largest differences occurring in the tropics. There is little difference in simulated GPP when using either WFDEI-GPCC or WFDEI-CRU (which differ only in the precipitation product used) to drive JULES.~~

4 Discussion

4.1 ~~Can JULES capture interannual variability of GPP at the global scale?~~ How do estimates of total annual GPP compare to those from observational datasets? Can JULES capture the seasonal and interannual variability of GPP at global scales?

~~When JULES was driven with the WFDEI-GPCC dataset (at $0.5^\circ \times 0.5^\circ$ spatial resolution), the model was able to capture interannual variability at the global scale (Figure 3c). This was also found when simulating GPP at lower spatial resolution ($1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$).~~ At global scales, ~~JULES estimates~~ the annual average GPP ~~to be 140~~ (combined GPP of all terrestrial ecosystems) ~~over 2001–2010, which is greater than~~ of the MODIS, FLUXNET-MTE and CARDAMOM ~~GPP by 25 %, 8 % and 23 % on average~~estimates over 2001–2010 are 112, 130 and 114 $\mathrm{Pg\,C\,year^{-1}}$, respectively (Figure 2). ~~These differences~~ Differences in GPP estimates are due to differences ~~in GPP between JULES, the observation-based estimates (MODIS and FLUXNET-MTE) and CARDAMOM~~ for forests and grasslands in the tropics (Figure 4b). MODIS and CARDAMOM GPP estimates are similar at global and regional scales since the CARDAMOM framework assimilates MODIS LAI data (Figure 2). JULES estimates the annual average GPP to be 140 $\mathrm{Pg\,C\,year^{-1}}$, which is greater than MODIS, FLUXNET-MTE and CARDAMOM GPP by 8–25 % (Figure 2). In the extratropics, JULES was able to simulate GPP compared to MODIS, FLUXNET-MTE and CARDAMOM due to its phenology model and associated model parameters being designed for temperate regions. When JULES was driven with the WFDEI-GPCC dataset (at $0.5^\circ \times 0.5^\circ$ spatial resolution), the model was able to capture interannual variability at the global scale (Figure 3b).

The main difference between ~~model simulated (~~JULES and CARDAMOM ~~) and observation-based (MODIS and FLUXNET-MTE) estimates of GPP~~ GPP estimates was found in the tropics with CARDAMOM GPP ~~estimates~~ being between the two observation-based datasets (Figure 2e). ~~Photosynthesis is~~ The reason for this is that MODIS LAI is used as input to CARDAMOM to constrain LAI. Photosynthesis is also modelled differently in JULES and CARDAMOM. In JULES, leaf-level photosynthesis is calculated as the minimum of three limiting rates which is then scaled up to canopy level using the sum of the leaf-level fluxes in each canopy layer, scaled by leaf area (Clark et al., 2011). In CARDAMOM, GPP is calculated as a function of LAI, air temperature and radiation using the Aggregated Canopy Model (Williams et al., 1997, ACM). ACM is an emulator of the Soil Plant Atmosphere ~~model (Williams et al., 1996, SPA)~~(SPA) model and uses a set of equations to simulate daily GPP estimates produced by SPA (Williams et al., 1996).

~~When JULES was driven with WFDEI-GPCC (JULES-WFDEI-GPCC), it~~ JULES simulates lower GPP than MODIS, FLUXNET-MTE and CARDAMOM at 15°N–30°N (Figures 5 and 7). This difference in GPP was due to the incorrect simulation of GPP by JULES in Mexico (Figure 7). No improvement in model performance was found when JULES was driven with different meteorological datasets (Figure 5). The total annual MODIS and FLUXNET-MTE GPP estimates for 2001–2010 are higher than that simulated by JULES by 1.0% and 6.7%, respectively, for Mexico, with CARDAMOM GPP estimates for the same period being lower than JULES GPP by 5.9%. One of the major vegetation types in Mexico is drought-deciduous plants (drought-deciduous plants lose their leaves during the dry season or periods of dryness as opposed to temperate deciduous plants which lose their leaves during periods of cold weather) and JULES does not contain this PFT.

Drought-deciduous plants can be found in the seasonally dry tropical forests of Mexico, Central America and northwestern South America. The implementation of drought-deciduous forest and shrub PFTs would help improve simulated GPP at latitudes $15°$N-$30°$N. In JULES, phenology is updated once per day by multiplying the annual maximum LAI by a scaling factor, which is calculated using temperature-dependent leaf turnover rates. Leaf turnover rates are a function of surface air temperature and increase when the temperature drops below a certain value (this varies depending on the PFT). While this is suitable for deciduous broadleaf forests in temperate regions, such as Northern Europe, it will lead to inaccurate modelled LAI for drought-deciduous forests. Instead of modifying modelled LAI using a temperature-derived scaling factor, the scaling factor could be calculated by using periods of dryness as the controlling factor.

In general, when JULES was driven with the WFDEI-GPCC dataset at global scales (JULES-WFDEI-GPCC-1degree), it was found that simulated photosynthesis was Rubisco-limited (Figures 5.6 and 5.7 in Slevin (2016)). Under saturated irradiance and limited atmospheric CO_2 concentrations, the Rubisco limiting rate is considered the main limiting factor (Marcus et al., 2008). Since the multi-layer approach for radiation interception and scaling from leaf-level to canopy-level photosynthesis was used by JULES in this study, the model simulates competition between Rubisco-limited and light-limited photosynthesis for each canopy layer (Clark et al., 2011). This means that lower in the canopy, there is increased light limitation and in the upper layers of the canopy, Rubisco limitation dominates (Clark et al., 2011). A description of the methods used to determine which limiting rate dominates each model gridbox when calculating potential leaf-level photosynthesis is provided in Appendix F of Slevin (2016).

In regions dominated by grasses and shrubs, photosynthesis was found to be transport-limited (Figure 5.6 in Slevin (2016)), which refers to the rate of transport of photosynthetic products (for C3 plants) and PEPCarboxylase limitation (for C4 plants). Transport limitation occurs mostly in Northern Eurasia and North America during the Spring and Summer months (March–September) and during the Autumn and Winter months (October–February) in Central Asia (Figures 5.6 and 5.9 in Slevin (2016)). The percentage of model gridboxes that were found to be Rubisco-limited was high (40– 100%), whereas the percentage of model gridboxes that were found to be light-limited were small (0–20%) (Figures 5.7 and 5.8, respectively, in Slevin (2016)).

4.2 How do fluxes of GPP simulated by JULES compare for various biomes at the global and regional scales?

~~JULES~~ At global scales, differences between MODIS and CARDAMOM estimates of average annual GPP are similar with MODIS simulating average annual GPP to be 46.6, 42.1 and 7.0 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively, and CARDAMOM simulating average annual GPP to be 50.1, 40.8 and 6.8 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively (Figure 4a). MODIS and CARDAMOM GPP estimates are similar due to MODIS LAI being assimilated into the simulation of GPP by CARDAMOM. FLUXNET-MTE GPP is higher than these estimates for all biomes (Figure 4).

JULES simulated average annual GPP to be 61, 54 and 7 $\mathrm{Pg\,C\,year^{-1}}$ for forests, grasslands and shrubs, respectively. JULES (JULES-WFDEI-GPCC) simulates higher GPP than MODIS, FLUXNET-MTE and CARDAMOM at global scales and this was found to be due to higher GPP simulated by JULES for forests and grasslands in the tropics (Figure 4b). The ~~total annual~~ average annual JULES GPP for shrubs globally and in the tropics and extratropics are approximately equal (Figures 4a and c).

~~JULES simulated average annual GPP to be 61 , 54 and 7 for forests, grasslands and shrubs, respectively.~~ This higher GPP in the tropics is due to the incorrect simulation of GPP by the version of JULES (version 3.4.1) used in this study. In this version, the PFT used to represent tropical forests is the broadleaf tree, which is used to simulate GPP in both tropical and temperate regions. This means that the model parameters used for the broadleaf tree PFT may not be suitable for simulating GPP in the tropics. The addition of extra PFTs more suited to tropical regions, such as tropical broadleaf evergreen (in version 4.2) and a drought-deciduous PFT, and a phenology model which simulates LAI in tropical regions would both improve GPP simulations.

Simulated GPP for forests is similar to that calculated by Beer et al. (2010) (sum of the values for tropical, temperate and boreal forests) with average annual GPP being 59 $\mathrm{Pg\,C\,year^{-1}}$. Since Beer et al. (2010) provides average annual GPP values for tropical savannahs and grasslands, temperate grasslands and shrublands and croplands, these are summed in order to obtain average annual global GPP for grasslands and shrubs 54.6 $\mathrm{Pg\,C\,year^{-1}}$, which is lower than the model simulated value of 61 $\mathrm{Pg\,C\,year^{-1}}$.

By further dividing the global land area into seven regions (Table 2), it was found that for ~~the~~ all three tropical regions (Central and South America, Africa and South and South-East Asia), JULES overestimated total annual GPP for forests, grasslands and shrubs (Figures 6c, e, and f). ~~The~~ Model version 4.2 of JULES contains PFTs, such as tropical broadleaf evergreen and evergreen shrub, which would improve GPP simulations in the tropics (Harper et al., 2016). Improved simulation of LAI in tropical regions would also aid in reducing differences between model simulated and observation-based estimates of GPP in these regions.

In the four extratropical regions (Europe, Northern Asia, Extratropical Southern Hemisphere and North America and Greenland)~~simulate similar GPP for JULES, MODISand~~ , JULES simulated similar GPP to MODIS, FLUXNET-MTE and CARDAMOM for the three biomes ~~with shrubs in North America and Greenland, Northern Asia and the extratropical Southern Hemisphere being underestimated by JULES~~ in Europe and the Extratropical Southern Hemisphere (Figures 6a ~~, b, d~~ and d), with the exception of Northern Asia and North America and Greenland, where the model is either equal to or lower than all three datasets (Figures 6b and g). This is probably due to the inability of this version of JULES to accurately simulate GPP in boreal regions where permafrost exists.

4.3 How sensitive are fluxes of GPP to the spatial resolution of the model?

~~When JULES was driven with the WFDEI-GPCC dataset at three different spatial resolutions (JULES-WFDEI-GPCC, JULES-WFDEI-GPCC-1degree and JULES-WFDEI-GPCC-2degree; Table 1), it was found that for GPP simulations, the model was~~ JULES was insensitive to spatial resolution with average annual global GPP being 140 $\mathrm{Pg\,C\,year^{-1}}$, 141 $\mathrm{Pg\,C\,year^{-1}}$ and 142 $\mathrm{Pg\,C\,year^{-1}}$ at $0.5^\circ \times 0.5^\circ$, $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$ spatial resolutions, respectively. This ~~trend~~ pattern was also observed in the zonal mean of total annual GPP (Figure 5). The insensitivity of the model to spatial resolution at the global scale was also observed at the regional scale when comparing simulated GPP fluxes for forests, grasslands and shrubs in the tropics and extratropics (Figure 6).

Little research has been performed on the effects of spatial resolution on JULES simulations (as well as other LSMs). Studies using atmospheric chemistry models have shown that the spatial resolution of the input meteorological data can affect model output (Ito et al., 2009; Pugh et al., 2013; Schaap et al., 2015). The results found here agree with those from Compton and Best (2011). Compton and Best (2011) showed that JULES was insensitive to spatial resolution when the WFD dataset was regridded from half-degree to 1-degree and 2-degree when simulating the terrestrial hydrological cycle. It was found that spatial resolution had little or no effect on simulations of global mean total evaporation and total runoff. However, the study showed that JULES was sensitive to temporal resolution when simulating the same hydrological components.

Using a different soil ancillary dataset or land cover map (which specifies the PFT fractions) may have a larger impact than changing the spatial resolution. The regridding method used in this study was the conservative method, which preserves the same information when interpolating from $0.5° \times 0.5°$ to $1° \times 1°$ and $2° \times 2°$ spatial resolutions, and results in only small differences in global GPP between the model simulations with varying spatial resolution. These small differences are due to differences in the PFT fractions of the land cover map after regridding.

4.4 *Is the meteorological dataset used to drive the model important at the global scale?*

When JULES was driven with the PRINCETON dataset at $1° \times 1°$ spatial resolution (Table ~~1)~~2, the annual average global GPP was slightly higher by ~~by 4~~3 $\mathrm{Pg\,C\,year^{-1}}$ than that simulated by JULES when driven with the WFDEI-GPCC dataset at the same resolution. In general, differences in GPP fluxes for model simulations driven using WFDEI-GPCC and PRINCETON are mainly in the deep tropics (at 5°N–5°S) with JULES-WFDEI-GPCC-1degree simulating higher GPP than JULES-PRINCETON and in the extratropics at 30°N–60°N, JULES-PRINCETON simulates slightly higher GPP (~~Figure~~Figures 5 and 6).

The higher simulated GPP in the tropics when JULES was driven with WFDEI-GPCC is due to positive biases in downward longwave radiation fluxes in WFDEI-GPCC in the Amazonian, African and South-East Asian tropics (Figures G.5b and d in Slevin (2016)) and the higher GPP simulated by JULES (driven with PRINCETON) in the extratropics are a result of positive biases in downward longwave radiation in the PRINCETON dataset in North America and Northern Asia (Figure G.5b in Slevin (2016)) and positive biases in surface air temperature in the PRINCETON dataset in the Northern Hemisphere (Figures G.6a and c in Slevin (2016)). As with the JULES-WFDEI-GPCC simulations, there are also differences in GPP between the PRINCETON driven JULES simulation and the observation-based and CARDAMOM estimates at latitudes 15°N-30°N (Figure 5). There was no improvement in simulated GPP when a different meteorological dataset was used.

In general, precipitation in the WFDEI-GPCC dataset is higher than that of PRINCETON (Figures G.6b and d in Slevin (2016)) with surface air temperatures higher in PRINCETON (Figures G.6a and c in Slevin (2016)). However, since JULES is more sensitive to downward longwave radiation and surface air temperature than precipitation when simulating GPP(Alton et al., 2007), the main reason for differences in simulated GPP when JULES was driven with two different meteorological datasets is due to differences in downward longwave radiation fluxes and surface air temperatures. There are differences in northern Eurasia (above 60°N) in the meteorological datasets with slightly higher radiation fluxes (downward shortwave and longwave) and surface air temperatures in the PRINCETON dataset with little difference between the JULES simulations driven with WFDEI-GPCC and PRINCETON in this region (Figure 5).

Other studies have shown that the meteorological dataset used to drive LSMs is a large source of uncertainty in global land surface modelling (Hicke, 2005; Jung et al., 2007; Poulter et al., 2011). Different methods are used to create time series of global gridded climate data in order to drive LSMs and this can introduce uncertainty that can propagate through model simulations (Zhao et al., 2006). Even at the point scale, differences in simulated GPP were observed when driving JULES with the WFDEI-GPCC and PRINCETON datasets (Slevin et al., 2015). As in this study, it also occurred in the tropics. The choice of meteorological dataset used to drive JULES has an important influence on GPP simulations.

A simple sensitivity study of the model to changes in climate (surface (2m) air temperature, precipitation and atmospheric CO_2 concentrations) when simulating GPP at global and regional scales for 2000–2010 was performed(Slevin, 2016). Only changes to one climate variable were made at a time due to complex interactions associated with multiple changes in climatic factors resulting in complex non-linear ecosystem responses which can be difficult to explain. JULES GPP was found to be sensitive to changes in all three climate variables with modelled LAI only sensitive to changes in surface air temperature (Slevin, 2016). At the regional scale, for model simulations with varying air temperature, GPP increased with increasing temperature in the extratropics, but decreased with increasing temperature in the tropics. Model simulations with varying precipitation at regional scales show the same trend as those at global scales with GPP increasing with increasing precipitation and decreasing with decreasing precipitation except for the magnitude of the effect observed. More detailed information on the sensitivity study is provided in Chapter 6 of the PhD thesis of Darren Slevin(Slevin, 2016).

When JULES was driven with the PRINCETON dataset, it was found that simulated photosynthesis was mostly Rubisco-limited (Figure 5.25 in Slevin (2016)). A similar trend was found when JULES was driven with the WFDEI-GPCC dataset (Figure 5.6 in Slevin (2016)). Similar trends in transport limitation were found with the JULES-PRINCETON model simulation, though the number of model gridboxes in which transport limitation dominated was less than that for the JULES-WFDEI-GPCC-1degree model simulation (Figures 5.25 and 5.28 in Slevin (2016)). When comparing the model gridbox fractions for the JULES-WFDEI-GPCC-1degree and JULES-PRINCETON model simulations, it was found that when JULES was driven with the PRINCETON dataset, simulated photosynthesis was more Rubisco-limited than when the model was driven with WFDEI-GPCC (Figure 5.26 in Slevin (2016)). Light-limitation was more important in simulating photosynthesis when JULES was driven with WFDEI-GPCC than PRINCETON (Figure 5.27 in Slevin (2016)). The percentage of model gridboxes which are transport-limited show a pronounced geographical variation with the WFDEI-GPCC driven simulation being more transport-limited in the Southern Hemisphere and the PRINCETON driven simulation being more transport-limited in the Northern Hemisphere (Figure 5.28 in Slevin (2016)).

In this study, the model simulations were performed with prescribed PFTs (i.e. no vegetation competition). If competition between PFTs was allowed (i.e. vegetation competition), the annual average global GPP would be higher by 15 % and 17 %, for the WFDEI-GPCC and PRINCETON driven simulations, respectively (Figures 8b and e). Higher GPP occurred mostly in Europe, southeastern US, and in the tropical regions of Central and South America, Africa and South and South-East Asia (Figures 8c and f). This increased GPP in tropical regions is due to the tree-shrub-grass dominance heirachy in TRIFFID with dominant types (trees) limiting the expansion of subdominant types (shrubs and grasses). In savanna regions, such as the Sudanian Savanna, which stretches from the Atlantic Ocean in the west to the Ethiopian Highlands in the east of Africa, and northern Australia, there is higher GPP with prescribed PFTs (Figures 8c and f). These are also fire-prone regions. The version of JULES used in this study has no fire module and TRIFFID may overestimate woody cover and therefore GPP.

In terms of global GPP, the WFDEI-GPCC and PRINCETON driven simulations produce similar increases (Figures 8b and e). However, the spatial pattern is slightly different with higher GPP simulated in the Amazon region when JULES was driven with the WFDEI-GPCC dataset and higher GPP in southern Brazil and Argentina and Southeast Asia when JULES was driven with the PRINCETON dataset (Figures 8c and f). The spatial pattern of simulated GPP is more sensitive to the meteorological data than the annual average global GPP if competition between PFTs is allowed. This may be due to compensating differences in the sensitivity of the model to the two meteorological datasets.

5 Conclusions

An evaluation of JULES was performed at global and regional scales with simulated GPP compared to global gridded ($0.5° \times 0.5°$ spatial and monthly temporal resolution) estimates of GPP derived from upscaled FLUXNET observations (FLUXNET-MTE), satellite observations from the MODIS sensor and that produced by the CARDAMOM data assimilation framework. ~~In general, it was found that JULES was able to capture interannual variability at the global scale. JULES simulated higher~~ JULES simulated higher average annual global GPP than FLUXNET-MTE, MODIS and CARDAMOM but at ~~the regional scale, these differences were due to differences between model simulated and observation-based estimates~~ regional scales, differences arose in the tropics. ~~In general, CARDAMOM was better at simulating GPP than JULES~~ It was found that JULES was able to capture interannual variability at the global scale.

Differences in GPP between JULES and the benchmarking datasets (FLUXNET-MTE, MODIS and CARDAMOM) at 15°N–30°N is due to higher FLUXNET-MTE, MODIS and CARDAMOM GPP for Mexico because of a lack of drought-deciduous PFTs in JULES. The inclusion of these PFTs would improve GPP simulations at latitude 15°N-30°N (mostly in Mexico). By dividing the global land area into seven regions, it was found that all three tropical regions (Central and South America, Africa and South and South-East Asia) contribute to model-observation differences at the global scale compared to FLUXNET-MTE and MODIS. The model is able simulate GPP estimates ~~at~~ in the four extratropical regions (Europe, Northern Asia, North America and Greenland and the extratropical Southern Hemisphere).

Improved GPP simulations in the tropics can be attained with the introduction of more PFT classes and their associated model parameters. In this study, the version of JULES used was 3.4.1. In this version, each model grid box is composed of nine different surface types and five of these are PFTs. Since model version 4.2, each JULES gridbox contains nine PFTS (tropical broadleaf evergreen, temperate broadleaf evergreen, broadleaf deciduous, needleleaf evergreen, needleleaf deciduous, C3, C4, evergreen shrub, deciduous shrub) ~~Harper et al. (2016)~~(Harper et al., 2016). In addition to these PFTs, a phenology model which can simulate LAI in both temperate and tropical regions, would help to reduce differences between model simulated and observation-based estimates of GPP in the dry and wet tropics.

When JULES was driven at the global and regional scale with the WFDEI-GPCC dataset at various spatial resolutions ($0.5^\circ \times 0.5^\circ$, $1^\circ \times 1^\circ$ and $2^\circ \times 2^\circ$), it was found that the model was insensitive to spatial resolution. Similar results were shown by Compton and Best (2011) when simulating components of the terrestrial hydrological cycle. Differences between high ($0.5^\circ \times 0.5^\circ$) and low ($2^\circ \times 2^\circ$) spatial resolution simulations of GPP are very similar. This means that low spatial resolution model simulations at these scales can be performed in place of high resolution when simulating GPP and results in shorter model run times.

The meteorological dataset used to drive LSMs at the global scale is an important source of model uncertainty (Poulter et al., 2011). By using a different meteorological dataset (PRINCETON) to drive the model, it was found that simulated GPP was similar to that when the model was driven with the WFDEI-GPCC dataset (at $1^\circ \times 1^\circ$ spatial resolution) with exceptions to this being in the tropics and the northern extratropics. These differences are due to biases in the downward radiation fluxes and surface air temperature in the meteorological data (WFDEI-GPCC and PRINCETON). When JULES was driven with the WFDEI-GPCC and PRINCETON datasets (both at $1^\circ \times 1^\circ$ spatial resolution), simulated photosynthesis was Rubisco-limited. Differences in precipitation, and hence soil moisture stress, did not play a role in differences between the two model simulations. When JULES was driven with the WFDEI-CRU dataset instead of WFDEI-GPCC, differences in simulated GPP were very small. The model simulations in this study were largely performed with prescribed PFTs (i.e. no competition between PFTs was allowed). With competition between PFTs, the annual average global GPP was higher by 15 % and 17 %, for the WFDEI-GPCC and PRINCETON driven simulations, respectively, with the spatial pattern of simulated GPP more sensitive to the meteorological data used.

The three benchmarking datasets all contain sources of error. Since observations of GPP do not exist at global scales, the MODIS and FLUXNET-MTE datasets are referred to as observation-based estimates of GPP as they are generated using observations and models. CARDAMOM may contain significant error from the assimilated data and model structure (number of pools, fire resilience of ecosystems), but so do the empirically based FLUXNET-MTE data (up-scaling of a partitioning algorithm) and MODIS GPP (a model based on PFT specific light-use efficiency). The advantage of CARDAMOM is that it is a process-based model and it ensures that the whole ecosystem functioning is coherent, while the observation-based datasets are only empirically based representations of GPP. In Figure S4 of the Supplementary Information of Bloom et al. (2016), there is a detailed study of the sensitivity of CARDAMOM to these various factors at 4 selected pixels representing temperate, boreal, wet and dry tropical ecosystems. Overall, there is not much difference in retrieved parameters because of the large error/uncertainty terms used when computing the likelihood.

In general, differences between JULES GPP and the benchmarking datasets (FLUXNET-MTE, MODIS and CARDAMOM) occur mostly in the tropics with differences at 15°N–30°N due to a lack of drought-deciduous PFTs in JULES. ~~There was little difference in JULES GPP when the model~~ When JULES was driven with different meteorological datasets ~~. Finally~~(WFDEI-GPCC and PRINCETON), the WFDEI-GPCC driven model simulations estimated higher GPP in the tropics (at 5°N–5°S) and the PRINCETON driven model simulations estimating higher GPP in the extratropics (at 30°N–60°N). The meteorological dataset used to drive JULES was found to be a source of model uncertainty in the tropics, though this may be due to model error. Finally, when model simulations of GPP were performed at various spatial resolutions ($0.5° \times 0.5°$, ~~at large scales JULES GPP was~~ $1° \times 1°$ and $2° \times 2°$), JULES was found to be insensitive to spatial resolution.

6 Code and/or data availability

The JULES model code (v3.4.1) is stored at the Met Office Science Repository Service in the JULES repository (https://code.metoffice.gov.uk/trac/jules) and access to the code can be requested from the official website of JULES (https://jules.jchmr.org/software-and-documentation). The outputs from the JULES model simulations reported in this paper have been deposited online at DataShare, the University of Edinburgh's digital repository of multidisciplinary research datasets, http://dx.doi.org/10.7488/ds/1461.

Author contributions. D.S., S.F.B.T. and M.W. designed the research. D.S., J.-F.E. and A.A.B performed the model simulations. D.S., S.F.B.T. and M.W. analysed the data. D.S. prepared the manuscript with contributions from all co-authors.

Acknowledgements. ~~I~~ The authors would like to thank the anonymous reviewers and editor for their valuable comments which improved the quality of the paper. We would like to gratefully acknowledge the School of GeoSciences, University of Edinburgh, which provided ~~me~~ D. Slevin with a School Scholarship that allowed ~~me to pursue my~~ him to pursue his PhD research, during which this work was carried out. Part of this work was conducted at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA. This work would not have been possible without help from the following people. Rich Ellis at the Centre for Ecology and Hydrology (CEH) provided the soil and vegetation ancillary files and advice on modifying the namelist files in order to set up the model simulations to run on a 2d grid. Doug Clark (also at CEH) provided advice on model spin-up at global scales. Information on the snow schemes used by JULES and advice on which one to use at global scales was provided by Richard Essery. Graham Weedon provided invaluable advice on using the WFDEI meteorological dataset.

References

Alton, P., Mercado, L., and North, P.: A sensitivity analysis of the land-surface scheme JULES conducted for three forest biomes: Biophysical parameters, model processes, and meteorological driving data, Global Biogeochemical Cycles, 20, GB1008, doi:10.1029/2005GB002653, 2007.

Anav, A., Friedlingstein, P., Beer, C., Ciais, P., Harper, A., Jones, C., Murray-Tortarolo, G., Papale, D., Parazoo, N. C., Peylin, P., Piao, S., Sitch, S., Viovy, N., Wiltshire, A., and Zhao, M.: Spatio-temporal patterns of terrestrial gross primary production: A review, Reviews of Geophysics, doi:10.1002/2015RG000483, 2015.

Arora, V. K., Boer, G. J., Friedlingstein, P., Eby, M., Jones, C. D., Christian, J. R., Bonan, G., Bopp, L., Brovkin, V., Cadule, P., Hajima, T., Ilyina, T., Lindsay, K., Tjiputra, J. F., and Wu, T.: Carbon–Concentration and Carbon–Climate Feedbacks in CMIP5 Earth System Models, Journal of Climate, 26, 5289–5314, doi:10.1175/JCLI-D-12-00494.1, 2013.

Battin, T. J., Luyssaert, S., Kaplan, L. A., Aufdenkampe, A. K., Richter, A., and Tranvik, L. J.: The boundless carbon cycle, Nature Geoscience, 2, 598–600, doi:10.1038/ngeo618, 2009.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rödenbeck, C., Arain, M. A., Baldocchi, D., Bonan, G. B., Bondeau, A., Cescatti, A., Lasslop, G., Lindroth, A., Lomas, M., Luyssaert, S., Margolis, H., Oleson, K. W., Roupsard, O., Veenendaal, E., Viovy, N., Williams, C., Woodward, F. I., and Papale, D.: Terrestrial Gross Carbon Dioxide Uptake: Global Distribution and Covariation with Climate, Science, 329, 834–838, doi:10.1126/science.1184984, 2010.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), Model description–Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, doi:10.5194/gmd-4-677-2011, 2011.

Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model–data fusion framework, Biogeosciences, 12, 1299–1315, doi:10.5194/bg-12-1299-2015, 2015.

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., and Williams, M.: The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times, Proceedings of the National Academy of Sciences, 113, 1285–1290, doi:10.1073/pnas.1515160113, 2016.

Blyth, E., Gash, J., Lloyd, A., Pryor, M., Weedon, G. P., and Shuttleworth, J.: Evaluating the JULES Land Surface Model Energy Fluxes Using FLUXNET Data, Journal of Hydrometeorology, 11, 509–519, doi:10.1175/2009JHM1183.1, 2010.

Blyth, E., Clark, D. B., Ellis, R., Huntingford, C., Los, S., Pryor, M., Best, M., and Sitch, S.: 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, 255–269, doi:10.5194/gmd-4-255-2011, 2011.

Bonan, G. B.: Forests and Climate Change: Forcings, Feedbacks, and the Climate Benefits of Forests, Science, 320, 1444–1449, doi:10.1126/science.1155121, 2008.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, Journal of Geophysical Research, 116, G02 014, doi:10.1029/2010JG001593, 2011.

Burke, E. J., Dankers, R., Jones, C. D., and Wiltshire, A. J.: A retrospective analysis of pan Arctic permafrost using the JULES land surface model, Climate Dynamics, 41, 1025–1038, doi:10.1007/s00382-012-1648-x, 2013.

Canadell, J. G., Le Quéré, C., Raupach, M. R., Field, C. B., Buitenhuis, E. T., Ciais, P., Conway, T. J., Gillett, N. P., Houghton, R. A., and Marland, G.: Contributions to accelerating atmospheric CO_2 growth from economic activity, carbon intensity, and efficiency of natural sinks, Proceedings of the National Academy of Sciences of the United States of America, 104, 18 866–18 870, doi:10.1073/pnas.0702737104, 2007.

Chadburn, S., Burke, E., Essery, R., Boike, J., Langer, M., Heikenfeld, M., Cox, P., and Friedlingstein, P.: An improved representation of physical permafrost dynamics in the JULES land-surface model, Geoscientific Model Development, 8, 1493–1508, doi:10.5194/gmd-8-1493-2015, 2015.

Chapin III, F. S., Matson, P. A., and Vitousek, P.: Principles of Terrestrial Ecosystem Ecology, Springer-Verlag New York, doi:10.1007/978-1-4419-9504-9, 2012.

Clark, D. B., Mercado, L. M., Sitch, S., Jones, C. D., Gedney, N., Best, M. J., Pryor, M., Rooney, G. G., Essery, R. L. H., Blyth, E., Boucher, O., Harding, R. J., Huntingford, C., and Cox, P. M.: The Joint UK Land Environment Simulator (JULES), model description–Part 2: Carbon fluxes and vegetation dynamics, Geoscientific Model Development, 4, 701–722, doi:10.5194/gmd-4-701-2011, 2011.

Collatz, G. J., Ball, J. T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agricultural and Forest Meteorology, 54, 107–136, doi:10.1016/0168-1923(91)90002-8, 1991.

Collatz, G. J., Ribas-Carbo, M., and Berry, J. A.: Coupled photosynthesis-stomatal conductance model for leaves of C_4 plants, Functional Plant Biology, 19, 519–538, doi:10.1071/PP9920519, 1992.

Compton, E. and Best, M.: Impact of spatial and temporal resolution on modelled terrestrial hydrological cycle components, Tech. Rep. 44, WATCH Technical Report, 19 pp [Available at http:// www.eu-watch.org.], 2011.

Cox, P. and Jones, C.: Illuminating the Modern Dance of Climate and CO_2, Science, 321, 1642–1644, doi:10.1126/science.1158907, 2008.

Cox, P. M.: Description of the "TRIFFID" Dynamic Global Vegetation Model, Tech. Rep. 24, Hadley Centre, Met Office, London Road, Bracknell, Berkshire, RG12 2SY, UK, 2001.

Cox, P. M., Huntingford, C., and Harding, R. J.: A canopy conductance and photosynthesis model for use in a GCM land surface scheme, Journal of Hydrology, 212–213, 79–94, doi:10.1016/S0022-1694(98)00203-0, 1998.

Cox, P. M., Betts, R. A., Bunton, C. B., Essery, R. L. H., Rowntree, P. R., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Climate Dynamics, 15, 183–203, doi:10.1007/s003820050276, 1999.

Dalmonech, D., Zaehle, S., Schürmann, G. J., Brovkin, V., Reick, C., and Schnur, R.: Separation of the Effects of Land and Climate Model Errors on Simulated Contemporary Land Carbon Cycle Trends in the MPI Earth System Model version 1, Journal of Climate, 28, 272–291, doi:10.1175/JCLI-D-13-00593.1, 2014.

Deardorff, J. W.: Efficient Prediction of Ground Surface Temperature and Moisture, With Inclusion of a Layer of Vegetation, Journal of Geophysical Research, 83, 1889–1903, doi:10.1029/JC083iC04p01889, 1978.

Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A.: An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics, Global Biogeochemical Cycles, 10, 603–628, doi:10.1029/96GB02692, 1996.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W., Brovkin, V., Cadule, P., Doney, S., Eby, M., Fung, I., et al.: Climate-Carbon Cycle Feedback Analysis: Results from the C4MIP Model Intercomparison, Journal of Climate, 19, 3337–3353, doi:10.1175/JCLI3800.1, 2006.

Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., and Knutti, R.: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks, Journal of Climate, 27, 511–526, doi:10.1175/JCLI-D-12-00579.1, 2014.

Galbraith, D., Levy, P. E., Sitch, S., Huntingford, C., Cox, P., Williams, M., and Meir, P.: Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change, New Phytologist, 187, 647–665, doi:10.1111/j.1469-8137.2010.03350.x, 2010.

Giglio, L., Randerson, J. T., and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), Journal of Geophysical Research: Biogeosciences, 118, 317–328, doi:10.1002/jgrg.20042, 2013.

Harding, R., Best, M., Blyth, E., Hagemann, S., Kabat, P., Tallaksen, L. M., Warnaars, T., Wiberg, D., Weedon, G. P., van Lanen, H., Ludwig, F., and Haddeland, I.: WATCH: Current Knowledge of the Terrestrial Global Water Cycle, Journal of Hydrometeorology, 12, 1149–1156, doi:10.1175/JHM-D-11-024.1, 2011.

Harper, A. B., Cox, P. M., Friedlingstein, P., Wiltshire, A. J., Jones, C. D., Sitch, S., Mercado, L. M., Groenendijk, M., Robertson, E., Kattge, J., Bönisch, G., Atkin, O. K., Bahn, M., Cornelissen, J., Niinemets, U., Onipchenko, V., Peñuelas, J., Poorter, L., Reich, P. B., Soudzilovskaia, N. A., and Bodegom, P. V.: Improved representation of plant functional types and physiology in the Joint UK Land Environment Simulator (JULES v4.2) using plant trait information, Geoscientific Model Development, 9, 2415–2440, doi:10.5194/gmd-9-2415-2016, 2016.

Hicke, J. A.: NCEP and GISS solar radiation data sets available for ecosystem modeling: Description, differences, and impacts on net primary production, Global Biogeochemical Cycles, 19, GB2006, doi:10.1029/2004GB002391, 2005.

Hiederer, R. and Köchy, M.: Global Soil Organic Carbon Estimates and the Harmonized World Soil Database, Tech. Rep. EUR 25225 EN, Publications Office of the European Union, Luxembourg, 2011.

Ito, A., Sillman, S., and Penner, J. E.: Global chemical transport model study of ozone response to changes in chemical kinetics and biogenic volatile organic compounds emissions due to increasing temperatures: Sensitivities to isoprene nitrate chemistry and grid resolution, Journal of Geophysical Research: Atmospheres (1984–2012), 114, D09 301, doi:10.1029/2008JD011254, 2009.

Jacobs, C. M. J.: Direct impact of atmospheric CO_2 enrichment on regional transpiration, Ph.D. thesis, Wageningen Agricultural University, 1994.

Janssens, I. A., Freibauer, A., Ciais, P., Smith, P., Nabuurs, G.-J., Folberth, G., Schlamadinger, B., Hutjes, R. W. A., Ceulemans, R., Schulze, E.-D., Valentini, R., and Dolman, A. J.: Europe's Terrestrial Biosphere Absorbs 7 to 12% of European Anthropogenic CO_2 Emissions, Science, 300, 1538–1542, doi:10.1126/science.1083592, 2003.

Jung, M., Vetter, M., Herold, M., Churkina, G., Reichstein, M., Zaehle, S., Ciais, P., Viovy, N., Bondeau, A., Chen, Y., Trusilova, K., Feser, F., and Heimann, M.: Uncertainties of modeling gross primary productivity over Europe: A systematic study on the effects of using different drivers and terrestrial biosphere models, Global Biogeochemical Cycles, 21, GB4021, doi:10.1029/2006GB002915, 2007.

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009.

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010.

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, Journal of Geophysical Research, 116, G00J07, doi:10.1029/2010JG001566, 2011.

Lasslop, G., Reichstein, M., Papale, D., Richardson, A. D., Arneth, A., Barr, A., Stoy, P., and Wohlfahrt, G.: Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: critical issues and global evaluation, Global Change Biology, 16, 187–208, doi:10.1111/j.1365-2486.2009.02041.x, 2010.

Le Quéré, C., Raupach, M. R., Canadell, J. G., and Marland, G.: Trends in the sources and sinks of carbon dioxide, Nature Geoscience, 2, 831–836, doi:10.1038/ngeo689, 2009.

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., van der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., Jourdain, C., Kato, E., Keeling, R. F., Klein Goldewijk, K., Levis, S., Levy, P., Lomas, M., Poulter, B., Raupach, M. R., Schwinger, J., Sitch, S., Stocker, B. D., Viovy, N., Zaehle, S., and Zeng, N.: The global carbon budget 1959–2011, Earth System Science Data, 5, 165–185, doi:10.5194/essd-5-165-2013, 2013.

Lei, H., Huang, M., Leung, L. R., Yang, D., Shi, X., Mao, J., Hayes, D. J., Schwalm, C. R., Wei, Y., and Liu, S.: Sensitivity of global terrestrial gross primary production to hydrologic states simulated by the Community Land Model using two runoff parameterizations, Journal of Advances in Modeling Earth Systems, 6, 658–679, doi:10.1002/2013MS000252, 2014.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303–1330, doi:10.1080/014311600210191, 2000.

Manabe, S.: Climate and the ocean circulation: 1, the atmospheric circulation and the hydrology of the Earth's surface, Monthly Weather Review, 97, 739–805, doi:10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2, 1969.

Marcus, Y., Altman-Gueta, H., Snir, A., Wolff, Y., and Gurevitz, M.: Photosynthesis. Energy from the Sun: 14th International Congress on Photosynthesis, chap. Does Rubisco Limit the Rate of Photosynthesis?, pp. 863–866, Springer Netherlands, doi:10.1007/978-1-4020-6709-9_191, 2008.

Ménard, C. B., Ikonen, J., Rautiainen, K., Aurela, M., Arslan, A. N., and Pulliainen, J.: Effects of Meteorological and Ancillary Data, Temporal Averaging, and Evaluation Methods on Model Performance and Uncertainty in a Land Surface Model, Journal of Hydrometeorology, 16, 2559–2576, doi:10.1175/JHM-D-15-0013.1, 2015.

Nachtergaele, F., van Velthuizen, H., Verelst, L., Wiberg, D., Batjes, N., Dijkshoorn, K., van Engelen, V., Fischer, G., Jones, A., Montanarella, L., Petri, M., Prieler, S., Teixeira, E., and Shi, X.: Harmonized World Soil Database v1.2, Tech. rep., International Institute for Applied Systems Analysis (IIASA), Food and Agriculture Organization of the United Nations (FAO), 2012.

Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L., Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire, A. D., Piao, S., Rautiainen, A., Sitch, S., and Hayes, D.: A Large and Persistent Carbon Sink in the World's Forests, Science, 333, 988–993, doi:10.1126/science.1201609, 2011.

Pielke, R. A., Avissar, R., Raupach, M., Dolman, A. J., Zeng, X., and Denning, A. S.: Interactions between the atmosphere and terrestrial ecosystems: influence on weather and climate, Global Change Biology, 4, 461–475, doi:10.1046/j.1365-2486.1998.t01-1-00176.x, 1998.

Pitman, A. J.: The evolution of, and revolution in, land surface schemes designed for climate models, International Journal of Climatology, 23, 479–510, doi:10.1002/joc.893, 2003.

Poulter, B., Frank, D. C., Hodson, E. L., and Zimmermann, N. E.: Impacts of land cover and climate data selection on understanding terrestrial carbon dynamics and the CO_2 airborne fraction, Biogeosciences, 8, 2027–2036, doi:10.5194/bg-8-2027-2011, 2011.

Pugh, T. A. M., Ashworth, K., Wild, O., and Hewitt, C. N.: Effects of the spatial resolution of climate data on estimates of biogenic isoprene emissions, Atmospheric Environment, 70, 1–6, doi:10.1016/j.atmosenv.2013.01.001, 2013.

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havránková, K., Ilvesniemi, H abd Janous, D., Knohl, A., Laurila, T., Lohila, A., Loustau, D., Matteucci, G., Meyers, T., Miglietta, F., Ourcival, J.-M., Pumpanen, J., Rambal, S., Rotenberg, E., Sanz, M., Tenhunen, J., Seufert, G., Vaccari, F., Vesala, T., Yakir, D., and Valentini, R.: On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Global Change Biology, 11, 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x, 2005.

Running, S. W., Thornton, P. E., Nemani, R., and Glassy, J. M.: Global Terrestrial Gross and Net Primary Productivity from the Earth Observing System, in: Methods in Ecosystem Science, pp. 44–57, Springer, 2000.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A.: Benchmark map of forest carbon stocks in tropical regions across three continents, Proceedings of the National Academy of Sciences, 108, 9899–9904, doi:10.1073/pnas.1019576108, 2011.

Schaap, M., Cuvelier, C., Hendriks, C., Bessagnet, B., Baldasano, J. M., Colette, A., Thunis, P., Karam, D., Fagerli, H., Graff, A., Kranenburg, R., Nyiri, A., Pay, M. T., Rouïl, L., Schulz, M., Simpson, D., Stern, R., Terrenoire, E., and Wind, P.: Performance of European chemistry transport models as function of horizontal resolution, Atmospheric Environment, 112, 90–105, doi:10.1016/j.atmosenv.2015.04.003, 2015.

Schaefer, K., Schwalm, C. R., Williams, C., Arain, M. A., Barr, A., Chen, J. M., Davis, K. J., Dimitrov, D., Hilton, T. W., Hollinger, D. Y., Humphreys, E., Poulter, B., Raczka, B. M., Richardson, A. D., Sahoo, A., Thornton, P., Vargas, R.and Verbeeck, H., Anderson, R., Baker, I., Black, T. A., Bolstad, P., Chen, J., Curtis, P. S., Desai, A. R., Dietze, M., Dragoni, D., Gough, C., Grant, R. F., Gu, L., Jain, A., Kucharik, C., Law, B., Liu, S., Lokipitiya, E., Margolis, H. A., Matamala, R., McCaughey, J. H., Monson, R., Munger, J. W., Oechel, W., Peng, C., Price, D. T., Ricciuto, D., Riley, W. J., Roulet, N., Tian, H., Tonitto, C., Torn, M., Weng, E., and Zhou, X.: A model-data comparison of gross primary productivity: Results from the North American Carbon Program site synthesis, Journal of Geophysical Research, 117, G03 010, doi:10.1029/2012JG001960, 2012.

Sellers, P., Mintz, Y., Sud, Y., and Dalcher, A.: A Simple Biosphere Model (SiB) for Use within General Circulation Models, Journal of the Atmospheric Sciences, 43, 505–531, doi:10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2, 1986.

Seneviratne, S. I. and Stöckli, R.: Climate Variability and Extremes during the Past 100 Years , chap. The Role of Land-Atmosphere Interactions for Climate Variability in Europe, pp. 179–193, Springer, 2008.

Sheffield, J., Goteti, G., and Wood, E. F.: Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling, Journal of Climate, 19, 3088–3111, doi:10.1175/JCLI3790.1, 2006.

Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quéré, C., Smith, B., Zhu, Z., and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon dioxide, Biogeosciences, 12, 653–679, doi:10.5194/bg-12-653-2015, 2015.

Slevin, D.: Investigating sources of uncertainty associated with the JULES land surface model, Ph.D. thesis, School of GeoSciences, University of Edinburgh, http://hdl.handle.net/1842/18757, 2016.

Slevin, D., Tett, S. F. B., and Williams, M.: Multi-site evaluation of the JULES land surface model using global and local data, Geoscientific Model Development, 8, 295–316, doi:10.5194/gmd-8-295-2015, 2015.

Tian, H., Lu, C., Ciais, P., Michalak, A. M., Canadell, J. G., Saikawa, E., Huntzinger, D. N., Gurney, K. R., Sitch, S., Zhang, B., Yang, J., Bousquet, P., Bruhwiler, L., Chen, G., Dlugokencky, E., Friedlingstein, P., Melillo, J., Pan, S., Poulter, B., Prinn, R., Saunois, M., Schwalm, C. R., and Wofsy, S. C.: The terrestrial biosphere as a net source of greenhouse gases to the atmosphere, Nature, 531, 225–228, doi:10.1038/nature16946, 2016.

van den Hurk, B., Best, M., Dirmeyer, P., Pitman, A., Polcher, J., and Santanello, J.: Acceleration of land surface model development over a decade of GLASS, Bulletin of the American Meteorological Society, 92, 1593–1600, doi:10.1175/BAMS-D-11-00007.1, 2011.

Walters, D., Brooks, M., Boutle, I., Melvin, T., Stratton, R., Vosper, S., Wells, H., Williams, K., Wood, N., Allen, T., Bushell, A., Copsey, D., Earnshaw, P., Edwards, J., Gross, M., Hardiman, S., Harris, C., Heming, J., Klingaman, N., Levine, R., Manners, J., Martin, G., Milton, S., Mittermaier, M., Morcrette, C., Riddick, T., Roberts, M., Sanchez, C., Selwood, P., Stirling, A., Smith, C., Suri, D., Tennant, W., Vidale, P. L., Wilkinson, J., Willett, M., Woolnough, S., and Xavier, P.: The Met Office Unified Model Global Atmosphere 6.0/6.1 and JULES Global Land 6.0/6.1 configurations, Geoscientific Model Development Discussions, 2016, 1–52, doi:10.5194/gmd-2016-194, http://www.geosci-model-dev-discuss.net/gmd-2016-194/, 2016.

Walters, D. N., Williams, K. D., Boutle, I. A., Bushell, A. C., Edwards, J. M., Field, P. R., Lock, A. P., Morcrette, C. J., Stratton, R. A., Wilkinson, J. M., Willett, M. R., Bellouin, N., Bodas-Salcedo, A., Brooks, M. E., Copsey, D., Earnshaw, P. D., Hardiman, S. C., Harris, C. M., Levine, R. C., MacLachlan, C., Manners, J. C., Martin, G. M., Milton, S. F., Palmer, M. D., Roberts, M. J., Rodríguez, J. M., Tennant, W. J., and Vidale, P. L.: The Met Office Unified Model Global Atmosphere 4.0 and JULES Global Land 4.0 configurations, Geoscientific Model Development, 7, 361–386, doi:10.5194/gmd-7-361-2014, 2014.

Weedon, G. P., Gomes, S., Viterbo, P., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: The WATCH Forcing Data 1958–2001: A meteorological forcing dataset for land surface- and hydrological-models, Tech. Rep. 22, WATCH Technical Report, 41 pp [Available at http:// www.eu-watch.org.], 2010.

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, Journal of Hydrometeorology, 12, 823–848, doi:10.1175/2011JHM1369.1, 2011.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resources Research, 50, 7505–7514, doi:10.1002/2014WR015638, 2014.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Wofsy, S. C., Shaver, G. R., Melillo, J. M., Munger, J. W., Fan, S.-M., and Nadelhoffer, K. J.: Modelling the soil-plant-atmosphere continuum in a Quercus–Acer stand at Harvard Forest: the regulation of stomatal conductance by light, nitrogen and soil/plant hydraulic properties, Plant, Cell & Environment, 19, 911–927, doi:10.1111/j.1365-3040.1996.tb00456.x, 1996.

Williams, M., Rastetter, E. B., Fernandes, D. N., Goulden, M. L., Shaver, G. R., and Johnson, L. C.: Predicting Gross Primary Productivity in Terrestrial Ecosstems, Ecological Applications, 7, 882–894, doi:10.1890/1051-0761(1997)007[0882:PGPPIT]2.0.CO;2, 1997.

Yang, W., Tan, B., Huang, D., Rautiainen, M., Shabanov, N. V., Wang, Y., Privette, J. L., Huemmrich, K. F., Fensholt, R., Sandholt, I., et al.: MODIS leaf area index products: From validation to algorithm improvement, Geoscience and Remote Sensing, IEEE Transactions on, 44, 1885–1898, doi:10.1109/TGRS.2006.871215, 2006.

Zhao, M. and Running, S. W.: Drought-induced reduction in global terrestrial net primary production from 2000 through 2009, Science, 329, 940–943, doi:10.1126/science.1192666, 2010.

Zhao, M., Heinsch, F. A., Nemani, R. R., and Running, S. W.: Improvements of the MODIS terrestrial gross and net primary production global data set, Remote Sensing of Environment, 95, 164–176, doi:10.1016/j.rse.2004.12.011, 2005.

Zhao, M., Running, S. W., and Nemani, R. R.: Sensitivity of Moderate Resolution Imaging Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological reanalyses, Journal of Geophysical Research: Biogeosciences, 111, doi:10.1029/2004JG000004, 2006.

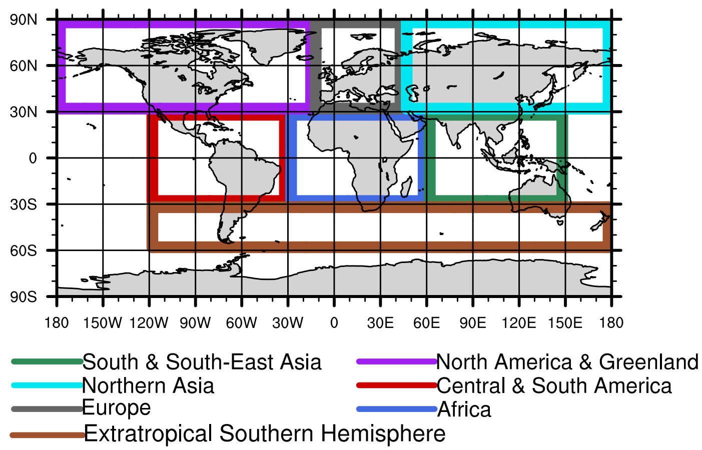
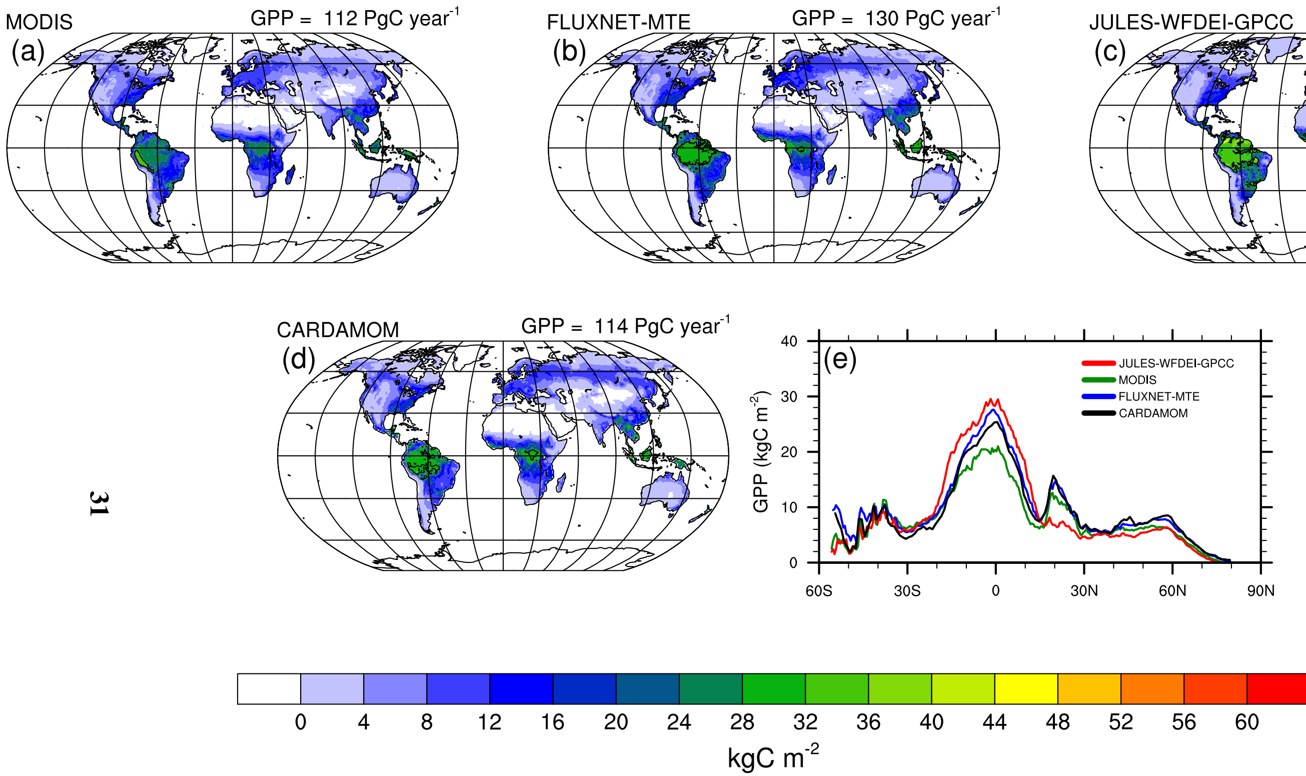


Figure 1. Map showing the regions specified in Table 2.

~~Comparison of JULES, observation-based (FLUXNET-MTE and MODIS) and CARDAMOM (Table 1) GPP fluxes for the 2001–2010 period at global scales. **(a)** shows the global area-weighted average of the mean monthly GPP, **(b)** shows the coefficient of variation (CV) expressed as percentages of the mean monthly GPP and **(c)** shows the monthly anomalies (global area-weighted mean) expressed as percentages of the mean monthly GPP (global area-weighted mean) for each month.~~



igure 2. Total annual and zonal mean model simulated (JULES-WFDEI-GPCC), observed (FLUXNET-MTE and MODIS) and CARDAMOM GPP fluxes for the 001–2010 period at the global scale. JULES, FLUXNET-MTE and MODIS GPP are at $0.5^\circ \times 0.5^\circ$ spatial resolution and CARDAMOM is at $1^\circ \times 1^\circ$ resolution. **(a)**, b), **(c)** and **(d)** show the total annual GPP of JULES-WFDEI-GPCC, FLUXNET-MTE, MODIS and CARDAMOM GPP, respectively. At the top right of each map ubplot, the average global annual GPP for 2001–2010 is displayed. **(e)** shows the zonal mean of the total annual JULES-WFDEI-GPCC, FLUXNET-MTE, MODIS nd CARDAMOM GPP, respectively. Included in each map subplot are contour lines for the tropical regions.

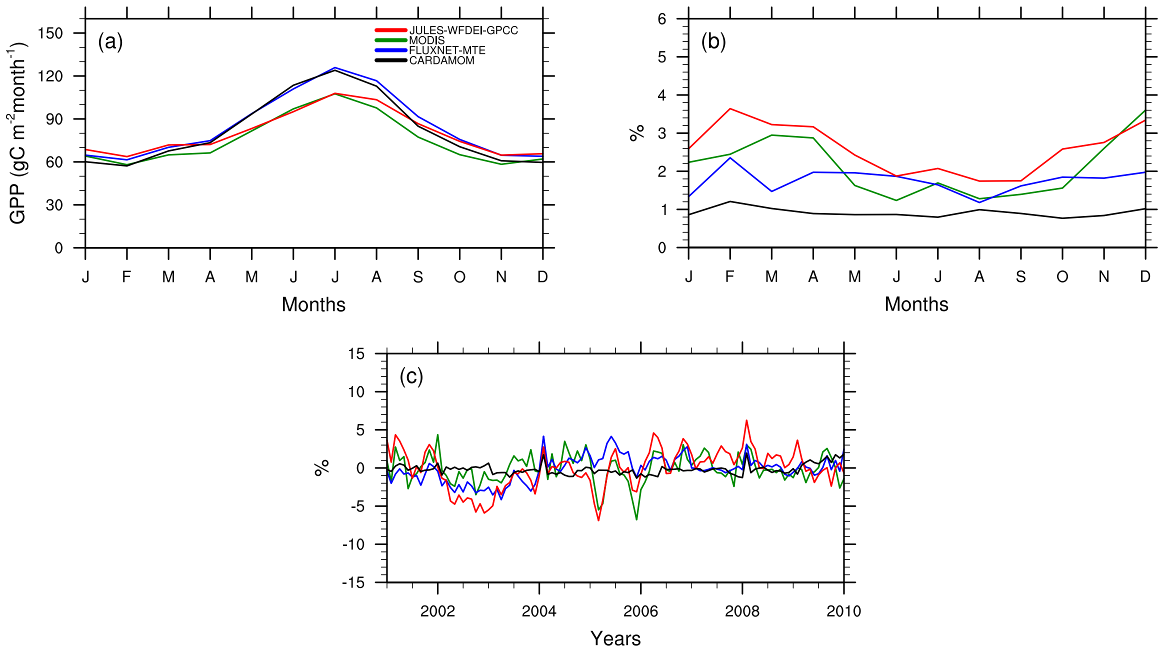


Figure 3. Comparison of JULES, observation-based (FLUXNET-MTE and MODIS) and CARDAMOM (Table 1) GPP fluxes for the 2001–2010 period at global scales. **(a)** shows the global average of the mean monthly GPP, **(b)** shows the coefficient of variation (CV) expressed as percentages of the mean monthly GPP and **(c)** shows the monthly anomalies expressed as percentages of the mean monthly GPP for each month.

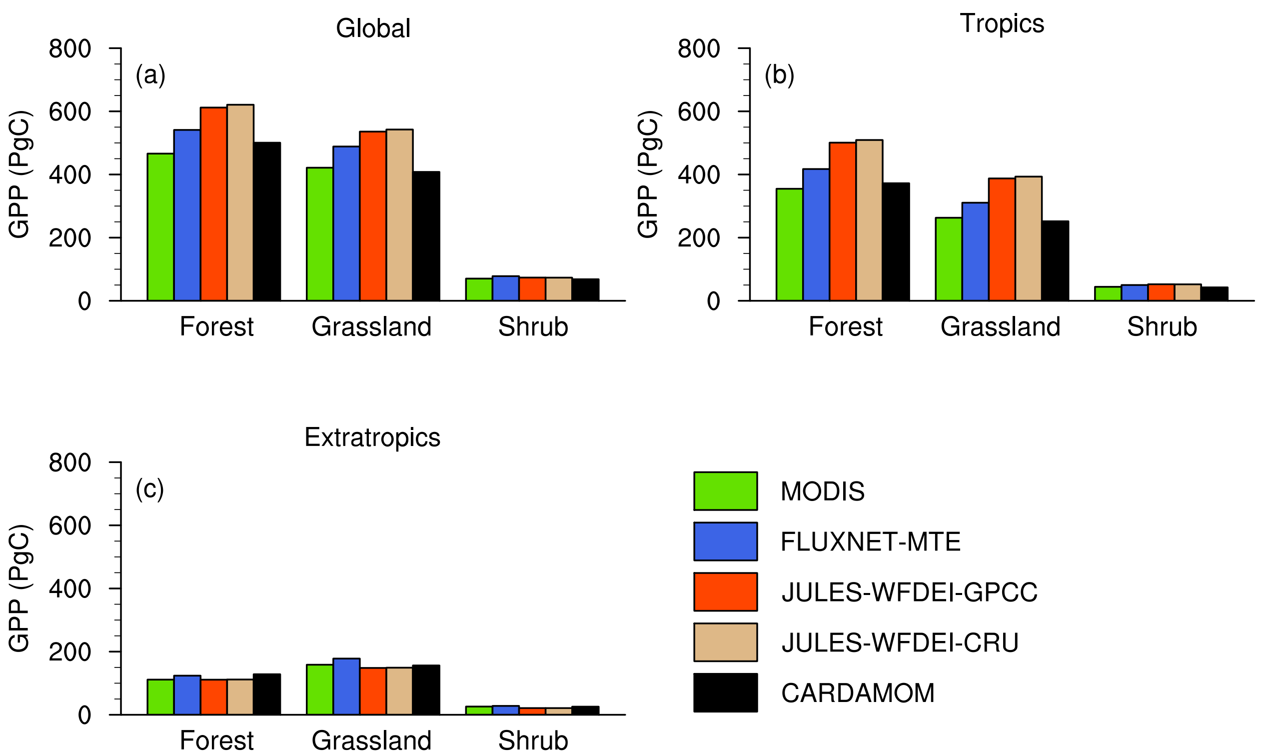


Figure 4. Total (summed over 10 years) model simulated (JULES-WFDEI-GPCC, JULES-WFDEI-CRU and CARDAMOM), observation-based (FLUXNET-MTE and MODIS) GPP fluxes for the 2001–2010 period at global and regional scales (tropics and extratropics) for 3 biome types (Forest, Grassland and Shrub). **(a)** shows the global total annual GPP, **(b)** for the tropics (30°S–30°N) and **(c)** for the extratropics (30°N–90°N and 30°S–90°S) for forests, grasslands and shrubs.

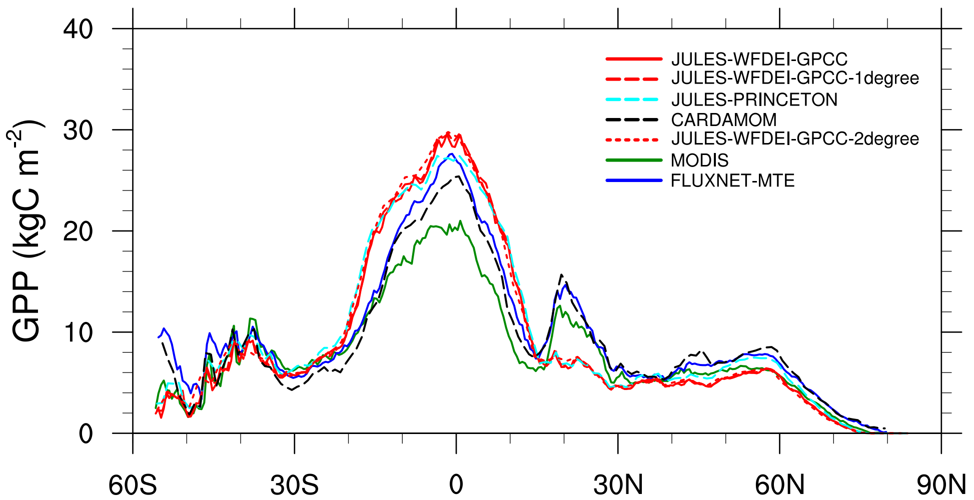


Figure 5. Zonal mean of total annual model simulated (JULES-WFDEI-GPCC, JULES-WFDEI-GPCC-1degree, JULES-PRINCETON, CARDAMOM and JULES-WFDEI-GPCC-2degree) and observed (FLUXNET-MTE and MODIS) GPP fluxes for 2001–2010. JULES-WFDEI-GPCC, FLUXNET-MTE and MODIS are at $0.5^{\circ} \times 0.5^{\circ}$ spatial resolution.

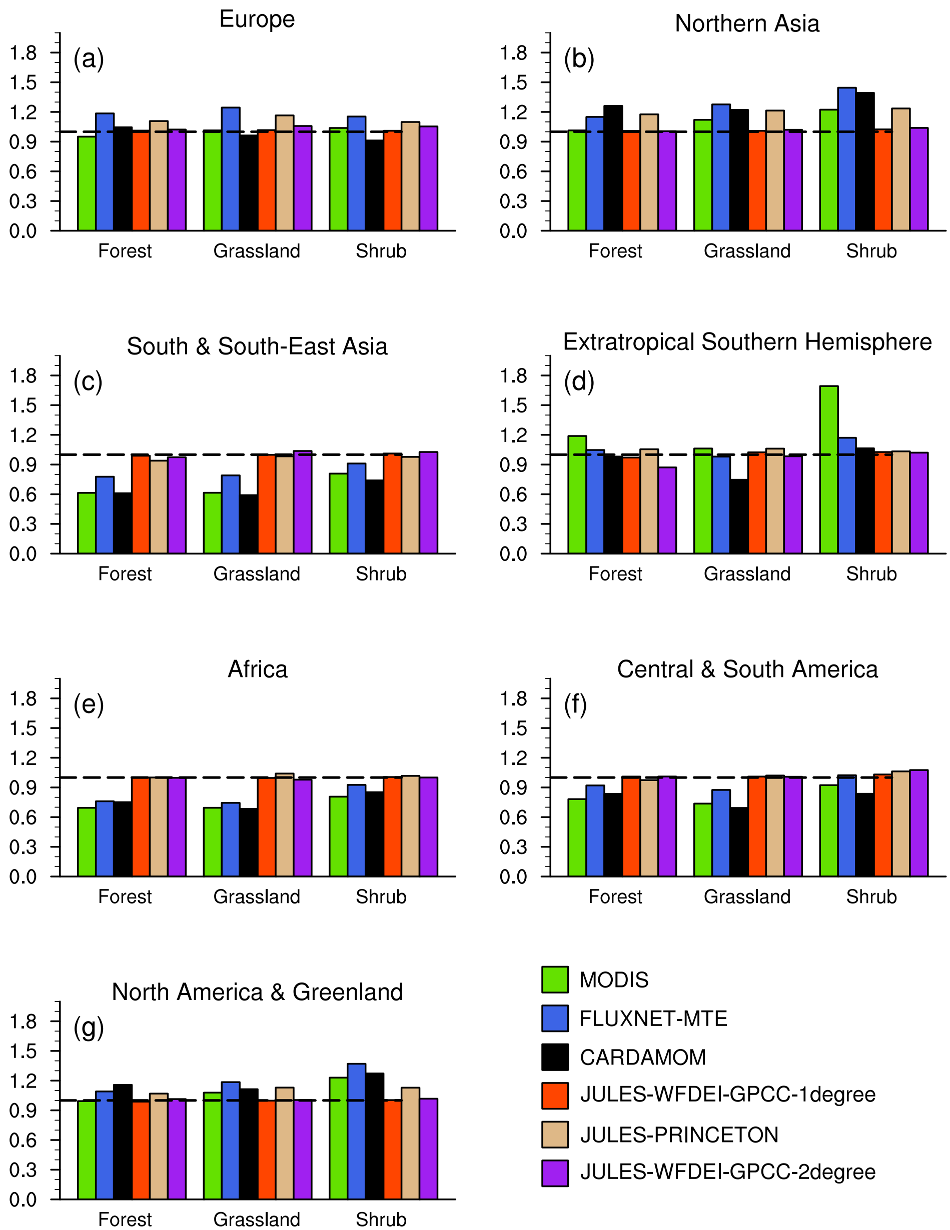


Figure 6. Total annual model simulated (JULES-WFDEI-GPCC-1degree, JULES-PRINCETON, CARDAMOM and JULES-WFDEI-GPCC-2degree) and observed (FLUXNET-MTE and MODIS) GPP fluxes for the 2001–2010 period normalised by model simulated (JULES-WFDEI-GPCC) GPP for various regions (Table 2) for 3 biome types (Forest, Grassland and Shrub). **(a)** shows normalised GPP for Europe, **(b)** for Northern Asia, **(c)** for South & South-Asia, **(d)** for extratropical Southern Hemisphere, **(e)** for Africa, **(f)** for Central & South America and **(g)** for North America & Greenland. The dotted line at y=1 represents where the model and observations match.

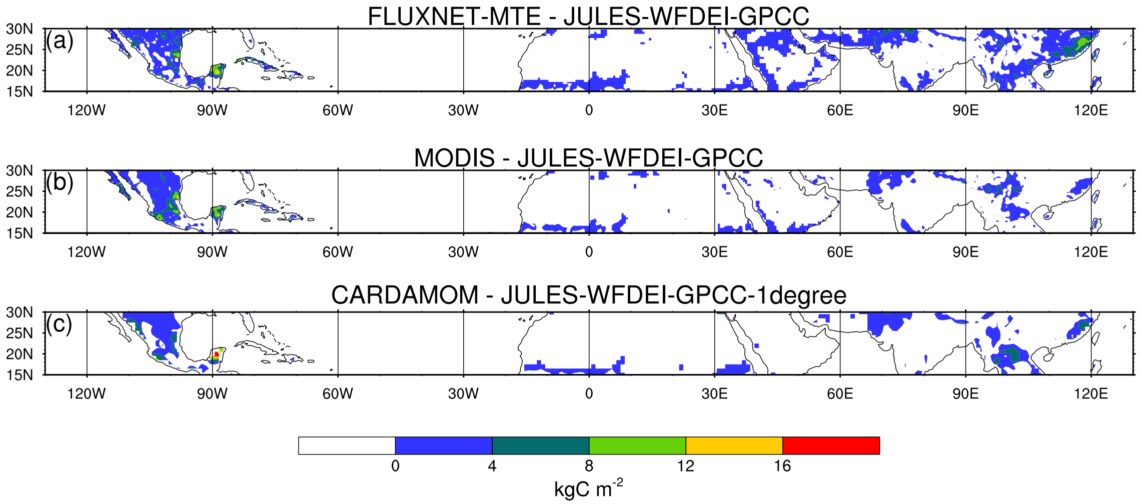


Figure 7. Difference in total annual GPP between JULES-WFDEI-GPCC and the observation-based (FLUXNET-MTE and MODIS) and CARDAMOM estimates of GPP for the 2001–2010 period at latitudes 15°N-30°N. **(a)** shows the difference between FLUXNET-MTE and JULES, **(b)** between MODIS and JULES and **(c)** between CARDAMOM and JULES. A positive change in GPP means the observation-based estimates (FLUXNET-MTE and MODIS) or CARDAMOM estimate are higher than the model.

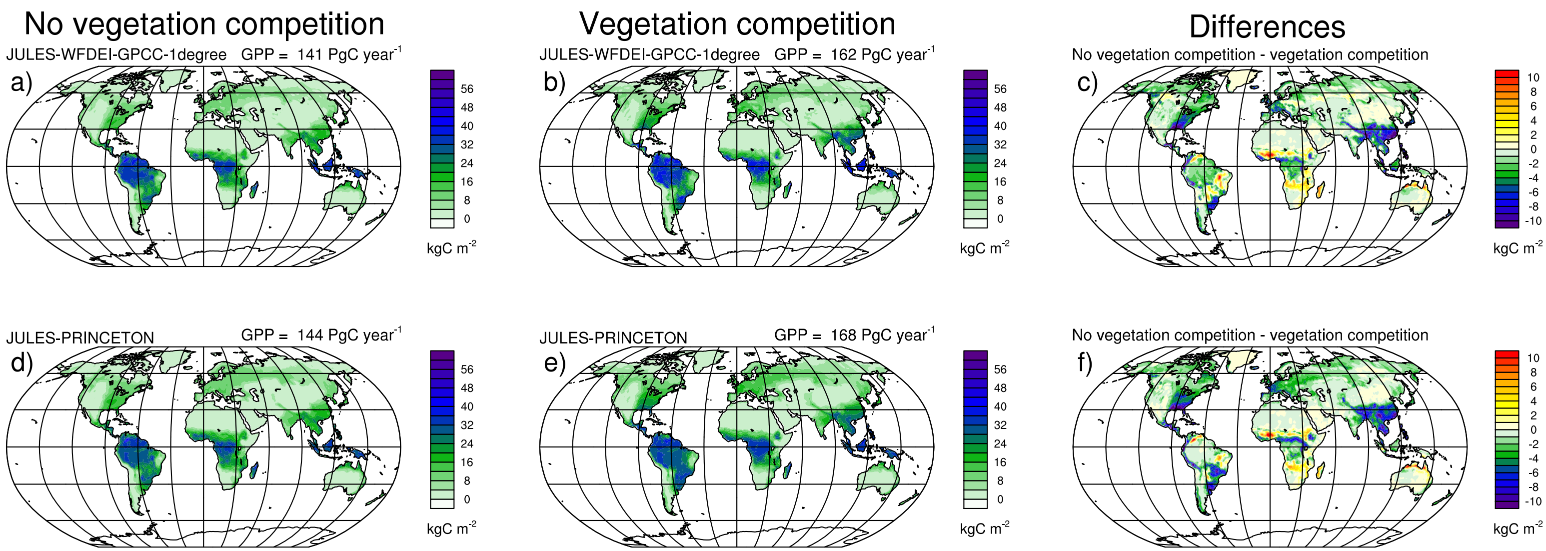


Figure 8. Total annual model simulated (JULES-WFDEI-GPCC-1degree and JULES-PRINCETON) GPP when simulations were performed with prescribed PFTs (vegetation competition switched off) and with different PFTs competing against each other (vegetation competition switched on) for 2001–2010. **(a)** and **(b)** show the total annual JULES-WFDEI-GPCC-1degree GPP with vegetation competition switched off and on, respectively, and **(c)** shows the difference. **(d)** and **(e)** show the total annual JULES-PRINCETON GPP with vegetation competition switched off and on, respectively, and **(f)** shows the difference. At the top right of **(a)**, **(b)**, **(d)** and **(e)**, the average annual global GPP for 2001–2010 is displayed.

Table 1. Types of global scale model simulations performed.

Model simulations	Meteorological forcing	Spatial resolution	Grid dimensions[a]
JULES-WFDEI-GPCC	WFDEI-GPCC	$0.5^\circ \times 0.5^\circ$	720×360
JULES-WFDEI-CRU	WFDEI-CRU	$0.5^\circ \times 0.5^\circ$	720×360
JULES-WFDEI-GPCC-1degree	WFDEI-GPCC	$1^\circ \times 1^\circ$	360×180
JULES-PRINCETON	PRINCETON	$1^\circ \times 1^\circ$	360×180
JULES-WFDEI-GPCC-2degree	WFDEI-GPCC	$2^\circ \times 2^\circ$	180×90

[a] Grid dimensions are given as the number of grid boxes in the longitudinal direction by the number of grid boxes in the latitudinal direction.

Table 2. List of regions used. Only land grid points are used in the analysis.

Name	Latitude ($^\circ$)	Longitude ($^\circ$)
Europe	30N–90N	15W–45E
Northern Asia	30N–90N	45E–180E
South & South-East Asia	30S–30N	60E–150E
Extratropical Southern Hemisphere	60S–30S	120W–180E
Africa	30S–30N	30W–60E
Central & Southern America	30S–30N	120W–30W
North America & Greenland	30N–90N	180W–15W