Second review of Global Evaluation of Gross Primary Productivity in the JULES Land Surface Model by D. Slevin et al.

Overall the authors have addressed most of the points I raised in my initial review. I had two major comments which the authors responded to in detail. I am satisfied with most responses but below are two further questions based on these responses.

1. [My original comment] (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.

[Author Response] 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 Hemi- sphere and North America and Greenland), JULES simulated similar GPP to MODIS, FLUXNET-MTE and CARDAMOM for the three biomes in Europe and the Extratrop- ical 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.

[New comment] Can more evidence be provided as to why permafrost would explain these differences? I can see how lack of frozen soils in JULES might create more water available for plants and increase GPP, but this is opposite to what was found in the study. It looks like most of the bias is in the shrub-dominated regions (fig 6). So the problem could be a lack of appropriate PFTs for these cold environments. In the response the authors state that a different land cover map is assumed in JULES, MODIS, and FLUXNET-MTE – what about this source for explaining some of the differences?

5. [My original comment] 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.

[New comment] I disagree with the conclusion that longwave radiation is a cause for differences between the model results with different datasets. Alton et al. (2007) found only 0.6% impact of LW radiation on simulated GPP, compared to 5% for SW radiation. There is no physiological reason to expect a strong link between longwave radiation and GPP. Also I don't see justification for the following statement in the cited Alton et al. paper:

"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."

Alton et al. (2007) used constant soil moisture stress and soil resistance terms and so did not represent sensitivity of model fluxes to the cumulative effect of precipitation (see their section 4.5). Aren't there clues to the reasons for difference between the meteorological datasets in the discussion of limiting factors for photosynthesis? The fact that the WFDEI-GPCC runs are more light-limited agrees with Fig. G5 in the PhD thesis cited – this dataset also has lower downwelling SW radiation than the Princeton data.

The authors state: "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))" (~Line 10, page 15). As the authors have pointed out elsewhere in the paper and references cited within, JULES tends to decrease GPP with higher temperatures in the tropics, and moisture availability can directly impact the GPP through the soil moisture stress function. Therefore, both of these biases (higher precipitation and lower temperatures in GPCC) could also explain the higher GPP simulated with WFDEI-GPCC.

Other comments
1. Is there a relationship between the high bias in the tropics (Fig. 2) and the high bias during DJFM in Fig. 3? If so, that would imply that GPP is too high in the tropics during the wet season, and could give some clues to the reason for the over-estimation.

2. The added description of driving data (beginning at the end of Page 3) and how it influences JULES GPP is helpful but I think there are a few points to clarify:
    - Relating to my above comment, I think it is misleading to say that downwelling shortwave *and longwave* radiation play an important role in the calculation of photosynthesis.
    - The light-limited rate is only a function of downwelling shortwave radiation
    - The soil moisture stress is definitely affected by precipitation but it is not part of the calculation, which the new sentence at the end of the paragraph seems to imply.

3. It's great that the model outputs have been made available. Is it possible to share other data used to drive JULES, for example the soil data and PFT distribution?

4. How were the biomes determined? Do the "forest", "grassland", and "shrub" biomes correspond to grid cells where these vegetation types are dominant?

5. In the zonal mean plot, there is a large low bias in JULES in the sub-tropics (15-30N), but this is not mentioned until the discussion. I think this should be mentioned earlier, especially because the low bias is apparently overwhelmed by the tropical high bias in the biome-scale plots for the Tropics (30S-30N). Also if the trees/shrubs/grasses for sub-tropics were included in Figure 4, this would back up the claim that a drought-deciduous tree or shrub PFT would help improve this large model bias. This is just a suggestion if the authors agree it would be useful to add this distinction.

6. Discussion: There is still some repetition in the discussion, I would advise additional proof-reading to see where repetition can be removed and similar threads of the discussion can be joined together. I think there should be more of a link between the discussion of the transport vs rubisco limited rates and the biases found in this study.