Editor's comments

We thank the editor for providing a final review of the manuscript and agree that the suggested changes and clarification improve it. We have made the changes outlined below in the revised manuscript. The page and line numbers of where changes have been made to the updated manuscript are included.

1. It is apparent that much effort has gone into this manuscript in responding to the reviewer's concerns and incorporating the suggested changes. I am almost satisfied that the manuscript has reached maturity and can be published in GMD. I would like to ask the authors for one final effort before we can go ahead with publication: following the comments of reviewer #2 the authors should try to use more quantitative assertions when discussing model performance.

To give an example: In the discussions (Page 5346, line 22 and onwards) the authors state that "the phenology model performed _quite well_ at temperate sites...", and similar qualitative terms are used throughout the text. While I appreciate that this "policy" makes the text more readable and approachable it also undermines somewhat the author's efforts in quantitatively assessing the JULES model's performance. I suggest to amend the text by introducing a consistent terminology for the very same terms. E.g., the authors could define "quite well" as meaning that the bias is less than 1 gC/m2/s and the RMSE is less than 2 $gC/m^2/s$. Similar definitions can be found for other terms like "poor", etc., and other metrics could be applied, too. My numbers are fairly arbitrary at this point and the authors will have to choose their own. The main advantage in my view, however, is that the qualitative terms also gain a quantitative or at least semi-quantitative meaning which can than be debated and discussed in the follow-up literature. The definition of a consistent terminology can be achieved either by adding a table or by adding the metrics in parenthesis at least at their first introduction in the text. For instance, a single paragraph fairly early on could be enough to define the terms.

I would like to ask the authors for this "final push" and then we shall go ahead with publication. At this point I would also like to thank the anonymous reviewers for their sterling job and all their efforts.

Yes, a quantitative definition of qualitative terms used to describe model performance would be a useful aid to readers and could be used and discussed in the follow-up literature. This was achieved by providing a paragraph in the model analyses section (Page 5354, lines 16-19) which describes a simple, but subjective, ranking system using qualitative terms (*Very well, Good* and *Poorly*) to describe the model's ability to simulate GPP. Table 4 (Page 5380) provides definitions of these qualitative terms (Page 5387). The results (Page 5355, lines 1-3, 5-7, 12, 17, 23, 27; Page 5357, line 15; Page 5359, lines 12, 17; Page 5360, line 22), discussions (Page 5361, line 20; Page 5363, line 4; Page 5365, lines 1-3) and conclusions sections (Page 5365, lines 15; Page 5366, line 11) have been updated with the new qualitative terms.