General Comments: "The paper Quantifying the carbon uptake…" by Wisskirchen et al., adds to the growing literature on model/data comparisons. This is a fairly recent trend and a worthwhile task, as far too little effort is placed on model uncertainty. The results from this paper clearly point to the difficulties that these models face in matching in-situ validations. The paper however detects rather critical shortcomings in the model but does not go far enough in discussing these problems. Both the discussion and conclusions need improvement in this regard.

First of all thank you for your review. Since a second reviewer also suggested to elaborate more on the discussion and conclusion, we will do so.

What do these rather large errors (ie. agriculture) imply in the Bethy model? Can they be easily explained and improved. What concrete changes to the model do the authors propose?

We are not completely clear about which errors you refer. Is it the discrepancy of BETHY/DLR to the FLUXNET measurements? If so, we are not sure if the term "error" is the right to use, because it would imply that one of the data sources is "correct". We already stated some ideas on why these discrepancies occurred, and will further elaborate on them in the discussion.

Results from this comparison should also be described in the context of other comparisons such as the paper by Schaefer et al., 2012 doi:10.1029/2012JG001960

A similar issue was pointed out by a second reviewer, and thus will find input to our revisions. Thanks for the source!

Specific Comments:
How do you map the flux tower data onto the model data – i.e. how do you match the different resolution of the datasets? Discussion here is warranted.

We assumed that the pixel which contains the flux tower is representative. We added a paragraph to explain this in more detail.

I think a map of the tower locations is warranted, as there are so many. Simply looking at tables is not enough.

Thanks for your suggestion. We added a map (new Fig 1).

I would like to see a difference or anomaly map. Figure 1 does not really bring across the variations from year to year. One suggestion would be to create a long-term average, and subtract each year from this average.

We followed your suggestion and changed figure 1 to a difference map. It is indeed the better option.

Figure 2d and 2e both show rather large discrepancies, but the r2 is still 0.6, i.e. rather good. Please confirm that this is really true, as judging from the figure I am skeptical.

We confirm that our figure is correct. We agree that discrepancies are obvious, however since r2 expresses the overall agreement of two curves and not the amplitude of offsets a r2 of 0.6 expectable.

Technical Comments:
Figure 3 is not referenced in the paper I think. If it should be included, justify why and please convert the x-axis so that we may more easily interpret the years.

Fig 3 (now Fig 4) was (and is) referenced (Page 2469; Line 28). We changed the x-axis.

There are numerous spelling mistakes in the paper. It would benefit from a native speaker edit.

We will carefully check the manuscript to minimize spelling errors.

Page 2469 line 13, low should in fact be high?

Yes → Fixed

Page 2472 line 17 do not use measured GPP, as GPP is not measured. It is modeled or estimated.

Fixed

Page 2472 line 20 makes no sense.

Indeed. In fact we wanted to say … qualitative statements …

Page 2473 line 7, noticeable climate change needs explanation.

We added examples.