

We thank the reviewer for providing constructive comments on our manuscript. The reviewer highlights some key challenges involved in a study like ours. We present our responses below. Reviewer's comments are in italics.

[MaC1]: Authors treat GPP from FLUXNET La Thuile data set as ground truth, and their conclusion need to take cautions. Firstly, GPP from eddy flux towers are not directly measured but derived by combining empirical models with measured NEE and other environmental variables. As to the uncertainties in La Thuile GPP dataset, there has been debate between scientists who are in charge of individual eddy flux towers and those who want to generate a sort of “uniform” GPP data set from eddy flux towers, such as FLUXNET La Thuile dataset. The formal argued that they know their sites and thus a unique method and night-time wind speed threshold should be applied to derive GPP for their tower sites; whereas scientists who proposed La Thuile dataset use a uniform method across all towers to derive GPP. Thus how much confidence we have in the GPP from La Thuile is an open question.

The reviewer correctly points out that GPP in the La Thuile dataset is not measured. The reviewer is also right that some scientists managing eddy covariance towers disagree with the common method used in processing the La Thuile dataset. We will acknowledge these in the revised manuscript and will replace L 7-13 on page 11632 with the following paragraph:

“Our analysis is based on measurements included in the FLUXNET “La Thuile” dataset. (<http://www.fluxdata.org/SitePages/AboutFLUXNET.aspx>). This dataset contains measurements of net ecosystem exchange (NEE) and near surface meteorology for 247 sites encompassing approximately 850 site-years of data since 2000. The dataset uses an empirical temperature response function to model ecosystem respiration (Reichstein et al., 2005), and estimates GPP as the residual of the sum of measured NEE and modeled respiration. The temperature response function is calibrated using nighttime data when winds are usually low and assumes that the calibrated relationship holds during daytime. It is worth noting that some site investigators are not in full agreement regarding the method used to model respiration in La Thuile dataset. To address these concerns, efforts to refine and improve respiration estimates are underway. Until these revised data are available, however, the La Thuile dataset is the only reference dataset available for this type of study. Most importantly, the La Thuile data set has been widely used, including in a number of high-profile synthesis studies (e.g., Beer et al., 2010). We therefore believe that the GPP data used here are of sufficient quality to meet the needs of this study, although we recognize that it includes errors and uncertainties associated with modeled respiration. To minimize these errors, we only included sites with high quality data and identified a subset of 176 sites with 515 site-years of data where each site-year satisfied two conditions: (i) more than 95 % of the days had daily GPP data, and (ii) the mean daily quality flag was more than 0.75 (Richardson et al., 2010).”

[MaC2]: *Secondly, there is a scale issue involved in comparing MODIS data with GPP from eddy flux towers. The location and size of footprint of the fluxes measured at towers are different from each other and highly dynamics which is influenced by i) site topography and homogeneity; ii) wind speed and direction; and iii) height of eddy flux towers. On the other hand, MODIS 500-m or 1-km data is in fact not 500-m or 1-km due to low frequency of nadir pixels (Tan et al., 2006). Therefore, in some cases, a direct comparison between MODIS data and GPP from eddy flux towers is problematic though almost all related studies follow such direct comparison.*

The reviewer is right in pointing out gridding artifacts in MODIS and mismatch in scale between MODIS pixel and tower footprint. We specifically recognise this issue in our conclusion and suggest that Landsat data can provide an improved basis for assessing models with eddy flux data. However, high frequency global scale assessments are currently not feasible with Landsat and MODIS is far and away the most widely used source of remotely sensed data for studies of this nature. Thus, the community needs to get an assessment of the accuracy and reliability of MODIS-based estimates. The challenge in any study that combines MODIS and FLUXNET data is to minimize the effects of gridding and mismatch in scale.

One of the many advantages of taking a 3 by 3 window is that the mean quantity is less likely to be influenced by signal from outside the window due to gridding. Moreover, as the reviewer also suggests, mismatch in scale in itself is not an issue in validating predictions of models that use MODIS data. In a homogeneous landscape even if the tower footprint is significantly smaller than MODIS pixels we can compare the two with a great deal of confidence. The mismatch in scale becomes an issue because of landscape heterogeneity. As we describe in the text, we attempt to reduce the effect of landscape heterogeneity by removing sites that are excessively heterogeneous. For the sites included in the study, we provide estimates of site heterogeneity in every biome in Figure 2. This information can be combined with the results to get an idea of the role of landscape heterogeneity. However, our analysis suggests that landscape heterogeneity is unlikely to be a significant source of disagreement between MODIS results and the in-situ GPP data.

[MaC3]: *But for the study of inter-annual variability, a subtle year-to-year change, both uncertainties in GPP from La Thuile and mismatch in scale become a major issue. Therefore, authors should state the uncertainties of using “exceeded +/-10% of mean annual GPP at each site” to represent the large anomalies since they don’t know how much uncertainties in the GPP from La Thuile dataset as to inter-annual variability. Therefore, authors should state the uncertainties of using “exceeded +/-10% of mean annual GPP at each site” to represent the large anomalies since they don’t know how much uncertainties in the GPP from La Thuile dataset as to inter-annual variability.*

We acknowledge that there is uncertainty in GPP and refer to relevant studies (L 25-28 on page 11636 and L 1-11 on page 11637). Our assertion is based on these studies. For example, Papale et al. (2006) also show that any GPP anomaly greater than 100 gCm^{-2} at

annual scale reflects a true signal. The mean annual site GPP in every biome is close to or above 1000 gCm⁻² (Table 4). Thus, the threshold of 10% corresponds to a mean absolute anomaly greater than 100 gCm⁻². We recognise that these general statistics do not provide conclusive evidence. However, characterizing uncertainty in flux tower data is an active field of research that is well outside the scope of this study. Our argument that large anomalies (>10%) have high signal to noise ratio is thus based on the best published evidence that is available.

[MaC4]: They also need to clearly state the cautions of their conclusions in the abstract due to the two major issues I stated above, though continuous refinement of remote sensing-based methods for monitoring GPP is another issue.

We concur with the reviewer and will modify the abstract appropriately in the revised manuscript.

REFERENCES

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., Rodenbeck, 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, 2010.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., and Yakir, D.: Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation, *Biogeosciences*, 3, 571-583, doi:10.5194/bg-3-571-2006, 2006

Reichstein, M., Falge, E., Baldocchi, D., Papale, D., Aubinet, M., Berbigier, P., Bernhofer, C., Buchmann, N., Gilmanov, T., Granier, A., Grünwald, T., Havráneková, K., Ilvesniemi, H., 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. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, *Global Change Biology*, 11: 1424–1439.

Richardson, A. D., Black, T. A., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., Hollinger, D. Y., Kutsch, W. L., Longdoz, B., Luyssaert, S., Migliavacca, M., Montagnani, L., Munger, J. W., Moors, E., Piao, S. L., Rebmann, C., Reichstein, M., Saigusa, N., Tomelleri, E., Vargas, R., and Varlagin, A.: Influence of spring and autumn phenological transitions on forest ecosystem productivity, *Philos. T. Roy. Soc. B*, 365, 3227–3246, 2010a.