This paper puts together a wide range of spatially explicit bottom-up surface-atmosphere CO2 flux data sets aiming to reconcile the carbon budget from bottom-up estimation and the atmospheric CO2 growth rate. While this type of research is needed for improving our understanding of carbon cycle, this study has serious flaws in generating the data and is lack of validation and deep analysis of the combined data set. The language is vague in many places. At this stage, I don't recommend publishing the paper.

Major comments:
1. The added value of the new combined dataset is very limited.
The authors simply put different data streams together, and there is no effort trying to harmonize the data, even though some of the datasets do not cover the same time period, e.g., the crops cover 2005 to 2010.

We strongly disagree with the impression that "there is no effort trying to harmonize… ". On the contrary, we have made major efforts to homogenize the various datasets comprising the current knowledge of spatiotemporally explicit, data-driven surface-atmosphere CO2 exchange. The chosen time period (2001-2010), spatial (1x1 degree) and temporal (monthly) resolution are a compromise arising from the availability of the different datasets. For several datasets, only one (annual mean) estimate for the chosen time period is available, including for Shelves, Estuaries, Rivers, Lakes, Wood harvest and the land use change flux (Eluc). All other datasets cover the entire time period with at least monthly time resolution and an original spatial resolution of 1x1 degree or finer such that resampling does not induce inconsistencies. Crop respiration data was extended backwards through linear extrapolation at each pixel. This is explained on p. 9 line 3.

2. The paper compares the bottom-up estimations with the top-down inversion results (Figures 3 and 4, section 3.4), but it is lack of discussion about why these two approaches have different results, and which estimation is closer to reality.

We will add a more in-depth discussion regarding these differences. We believe that our estimates overestimate carbon uptake in tropical land areas and carbon release in tundra regions. This may explain many of the differences visible in Figs. 3 and 4. For areas where the different estimates converge (mid-latitudes, Europe, Russia, South Asia, East Asia, Australia) we can state with some confidence that we know net carbon exchange. However, an overall judgment which estimates are closer to reality cannot be made given current knowledge.

3. In section 3.4, it says that "both estimates agree well in the extratropics", but the figure 3d shows that the NCE-FF and atmospheric inversion results also have large differences in the NH high latitudes (between 60N and 75N), with the NCE-FF indicating a source to the atmosphere, while the atmospheric inversion indicating a weak sink.

We agree with the reviewer and will add more discussion on this point in the revised version. At very high latitudes, in the tundra region, very few flux tower observations are available. Hence the FLUXCOM runs are not well constrained in those regions. In contrast to the tropics where this leads to an unrealistically large carbon sink, in the high latitudes the FLUXCOM runs show a strong source.

4. Even though the latitudinal pattern of the inversion results follows a pattern similar to that of the aquatic fluxes (Figure 3c), there is no direct evidence indicating the propagation of the marine signal into continents during atmospheric inversion. I suggest removing the discussion on the pattern comparison between aquatic fluxes and atmospheric flux inversion results in section 3.4.

We agree and will omit this discussion.

5. Section 4.5 discusses the possible application of the combined dataset in model-data integration studies. It is an interesting idea. However, with such large uncertainties (with more than 10GtC disagreement with the atmospheric CO2 growth rate) in the combined dataset and a mixture of all different carbon flux components, it is not clear how such product can be used in carbon cycle data assimilation that focuses primarily on land carbon fluxes. What is the added value of using such data set compared to directly using flux tower observations? In addition, if such product were to be used as "observations" in a data assimilation system, a rigorous validation against independent observations is needed.

Our main aim is to exploit the spatiotemporal explicitness of the NCE flux in tandem with the spatially explicit uncertainties for model-data fusion. Probably the most relevant application would be using these data at the regional scale, as one goal of the study is to pinpoint regions of small and large uncertainties in the NCE estimates. In some regions, uncertainties are so large that nearly no meaningful information on the mean NCE flux can be obtained with currently available observational networks and statistical approaches. This is, for example, the case for many tropical land regions. But, and we see this as a key advantage of our study, the included uncertainties clearly indicate the merit of such a data compilation, especially in contrast to flux tower observations: our study includes all the major fluxes, such as fire emissions, inland aquatic fluxes, tropical land use change estimates, and emissions related to harvested wood and crop products. This is much closer related (and more directly comparable) to the actual net carbon exchange fluxes as they are resolved by inversions (if fossil fuel emissions are omitted). All the datasets used in this study were validated individually against independent sources, and those studies are referenced in the respective sections. We don't know of any independent observation that can be used to validate the obtained NCE flux at such high spatial and temporal scale. An exception may be inversions and the regional aggregates obtained in the RECCAP synthesis, and we compare our estimates to RECCAP in the manuscript. We will include a similar comparison for the ocean regions in the revised version. At finer spatial and temporal scale, and in some regions, especially the tropics and northern high latitudes, independent trustworthy references are lacking.

6. In section 3.2, it says that "13% of our runs we obtain a global C source that is consistent with the atmospheric growth rate", what are the spatial distributions of the fluxes from these 13% runs?

As mentioned also in the response to reviewer 1, we have decided to omit this from the revised version. We base this on the following: (1) There is a large uncertainty related to missing fluxes (see page 16). (2) We cannot be sure that our assessment includes all relevant fluxes in all regions. (3) Constraining the NEP ensemble could thus be misleading because the constraint is highly uncertain. Furthermore, if we assume that relevant drivers are missing in the set of predictors in FLUXCOM (e.g. forest age, see Sect. 4.1), all members in FLUXCOM are biased in the same way (all members use the same set of predictors). Constraining the ensemble thus cannot provide us with new insights into the processes that need further investigation or drivers that are missing in the set of predictors.

7. L26 (P14): what is the distribution of the different age classes of forests in FLUXNET? Is there solid evidence showing that the year and regrowing forests are overrepresented in FLUXNET?

So far this is only a hypothesis and it has not been shown. This hypothesis is a strong candidate in explaining the overestimation of the carbon sink in the tropics. Future research has to demonstrate whether these hypotheses are valid. We will make this clearer in the revised version.

8. Section 3.5 discusses seasonal cycle and monthly variability. It would be helpful to put this discussion in perspective, e.g., comparing to other independent estimations, so that the readers would know the credibility of this result. It is not clear what are the latitude ranges for the NH and SH in Figure 6.

The ranges are 90 S-0 for SH and 0-90 N for NH. We will include a comparison with inversions for the seasonal cycle, which are the only independent estimates based on observational data. As annual variability is already compared to inversions demonstrating large discrepancies (Figure 5) we refrain from comparing monthly variability.

9. Line 6 (p15), what is the basis for the 50% uncertainty?

Here we refer to a recent paper on global fire emissions by van der Werf et al, currently in discussion for Earth System Science Data (van der Werf et al., 2017). Estimating uncertainties in fire emission estimates is notoriously difficult. Assuming 50% uncertainty for estimated fire emissions is a best guess assessment, and better quantifying this uncertainty requires an assessment of the burned area estimates as well as new field data on fuel consumption and emission factors.

Minor comments

1. In the abstract, "would require an offsetting surface C source of 4.27 +- 0.10 PgC/yr", should the offset be 4.27 + 6.07 PgC/yr in order to have 4.27 PgC atmospheric CO2 growth rate?

Yes, that is correct. We will reformulate this sentence to make this clearer.

2. Line 13-16 (p3), it is not clear what the "background CO2 fluxes" means.

With background fluxes we mean the fluxes before human disturbance (i.e., before the large increase in fossil fuel emissions). Those are not included in the estimates of the Global Carbon Project which only discusses the human perturbation. In the revised version we will avoid the term to avoid confusion.

3. Line 23 (p4), "goal of this study the" should be "goal of this study to"

Thank you.

4. Line 16 (p6): what is "schused"?

Should be "used", will be changed in the revised version.

5. Line 21 (p14): What does the "relevant drivers" refer to? Be more specific.

Additional drivers relevant for upscaling NEP could be, for instance, the disturbance history (e.g. time since the last disturbance) or, closely related, forest age. This is mentioned a few lines higher up. We will be more specific here and put the most likely missing drivers.

6. Line 29 (p14): what does the "global driver" refer to?

By this we refer to the fact that there is no global map of forest age, which could be used as an additional driver for upscaling NEP (see also comment above). We will clarify this in the revised version.

Additional References:

van der Werf et al.: Global fire emissions estimates during 1997–2015, Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-62, in review, 2017.