Referee #1

Major comments

1. All three datasets are of dubious accuracy in representing interannual variability. The annual totals computed from eddy covariance sum much larger fluxes of opposing signs with likely systematic biases, especially in nighttime. The empirical upscaling was found to have relatively weak performance in representing interannual variability in a synthetic data experiment (without even accounting for any measurement or representativeness error in the training set) reported by Jung et al. (2009), for which the absence of soil moisture as a predictor is given by them as one reason. The inversion estimate, as the authors point out, is dominated at sub-continental scales by the (reasonable) prior assumption that variability scales with modeled NPP, and it probably contains little actual information from the CO2 time series at those scales. Could it makes sense to run the inversion with a more 'flat' prior, or a prior based on the MPI-MTE IAV, to get different IAV estimates?

   We agree with the reviewer that all the datasets used in the present analyses present weaknesses and lack of accuracy in representing the inter-annual variability. On the other side, this is what is currently available in terms of global-scale data of CO_2 land fluxes derived from inversions of atmospheric measurements or from the upscaling of surface flux observations. Following the reviewer suggestion, the limits of each product will be better discussed in the Materials and Methods section and strengths and weaknesses will be taken into account when evaluating results. Concerning the prior used for the Jena Inversion, it has indeed a seasonal pattern, however this is constant from one year to another, hence there is no influence of the prior on the IAV. The prior can only influence the fine-scale spatial pattern of IAV, since in the optimization the fluxes scale in space with the average prior flux. On the contrary the temporal IAV derives fully from the atmospheric signal. Using an MPI-MTE based prior for the Jena Inversion product would contaminate the IAV estimation, mostly because MPI-MTE varies in time, hence MPI-MTE IAV would influence the IAV derived from the Jena Inversion with the result that the two products wouldn't be independent any more.

2. Figure 4 shows the dependence of median(?) IAV on resolution for the two gridded products. I wonder if something like this could be done with the available Fluxnet stations as well, for example with the help of a variogram (mean covariance of de-seasonalized NEE time series as a function of inter-station distance). This could help in deciding whether the lower IAV in the gridded products compared to Fluxnet is only because of the difference in spatial scale or is more intrinsic.

   We thank the reviewer for the interesting suggestion. We will add a new set of results to Fig 4 which will explore the dependence of IAV on the spatial averaging of the Fluxnet dataset, following the scheme used for the gridded product. The new series represents the IAV calculated from the Fluxnet database as a function of the area of aggregation of the sites, starting from single sites and then proceeding with averaging time series for groups of sites located within an increasing distance. This procedure applied to flux sites mimics a decreasing resolution as done for the gridded products.

3. I didn't see any analysis of to what extent the IAV between the three products is actually in phase (i.e. the correlation of the deseasonalized NEE time series between the datasets). It would probably be relevant to show this.
   We will consider this point together with point #1 raised by reviewer #3 and will perform an analysis on global averages of the two global products and of the Global Carbon Project estimates.

4. Also, forest inventories and crop yield statistics provide more reliable direct measurements of (at least above-ground) NPP and its IAV in many countries, potentially with rather good spatial coverage. Would there be any way to compare these to the IAV in the data sets reported here?
   Following the suggestion of the reviewer we considered other possible data streams for the analysis but ultimately concluded that neither forest inventory nor yield statistics are appropriate for the present analysis. In fact, forest inventories are typically performed every 10-15 years, therefore they report NPP as a time average and for this reason they cannot be used in an inter-annual variability analysis. Crop yields are not necessary correlated to primary productivity, as they may be affected by events that do not affect GPP like for example a storm or frost at the end of the growing season that can fully compromise the yield but do not substantially change GPP.

Minor points

1. The element "carbon" is not capitalized (title and line 287).
   The typo will be corrected.
2. Line 25: no comma before "that"
   Comma will be cancelled
3. Figure 1c: It would be good to show the station network on the map.
   Following the reviewer suggestion the station network will be plotted in Figure 1c
4. "Anomalies" sounds strange as a description of the IAV residuals from linear trend shown in Figure 5 and discussed in the text. Perhaps there is a better term.
   As suggested by the reviewer we will use the term residuals.
5. The Jung et al. (2009) citation should be to the final paper, not the discussion paper.
   The citation will be replaced with that of the final paper.
6. Formatting in the bibliography needs to be fixed, e.g. for Morgenstern et al. (2004) and others.
   Bibliography will be checked and fixed.

Referee #2

General comments

1. There are some weaknesses. Some areas of the text, and a critical point or two in the methods, are unclear. Neither the MPI-MTE nor the inversion products seem ideal for this kind of IAV analysis, although I recognize that this is all there is to work with; still, the authors should address this.
   As stated by the reviewers the dataset used in the analysis are those available nowadays for the land CO_2 fluxes, namely i) site observations based on eddy covariance, ii) statistically upscaled products derived from site level measurements as MPI-MTE, or iii) inversion modeling products. We are aware of the weaknesses of the products used in this analysis and we plan to better discuss them together with their pros both in product descriptions and in the result discussion. Refer also to Referee #1 comment 1.

2. In addition, the conclusion should be re-done or removed; on a related note, the strengths and weaknesses of these NEE data products might be better, and more succinctly, summarized based on the analyses performed.
   We will prepare a new version of the conclusions following the suggestions of reviewer #2 and #3.

Specific comments

1. Lines 118-120: not as clear as it should be. Interannual variability computed with a 12-month window? How is this possible, as that's only 1 year?
   Analysis of IAV was based on the entire time series. Annual values were calculated not only for the "solar" years which were available in the dataset, but additional "years" were generated using a 12-month moving window which was shifted one month a time (Luyssaert et al. 2007).
2. L. 171-172: move to figure caption, or methods
   The sentence will be moved to Materials and Methods section 2.2
3. L. 197: "area of"
   The typo will be corrected
4. L. 241-243: unclear
   We will better clarify the concept in the revised text on the basis of what follows.
   The impact of climate drivers on IAV is based on a spatial analysis and not a temporal one. Spatial analyses of IAV in the inversion product are critical because at fine scale the spatial variability of the fluxes is mainly controlled by priors. In fact, the optimization algorithm of the inversion spatially allocates the fluxes proportionally to the prior; hence grid cells with higher productivity will change more if compared to cells with lower prior value (i.e. IAV at fine scale is proportional to the prior). For this reason we did not perform the spatial analysis on the inversion. On the contrary, prior does not affect the temporal analysis of IAV performed on the inversion product throughout the paper.
5. L. 250-: separating paragraphs, or indenting their first lines, would make this easier to read

Following the reviewer suggestion paragraph first lines were indented.

6. L. 286-: these aren’t conclusions, just a recapitulation of results; remove

   As stated above we will reformulate the conclusions in the new version of the manuscript.

7. Figure 2: Rain (in axis title) or Precipitation (in caption)?

   Axis title will be modified in order to be consistent with the figure caption

Referee #3

General comments

1. I feel like the paper is missing the bigger take home message I was looking for, to the globally (or Fluxnet) integrated anomalies in NEE match up with 1) each other and 2) anomalies in the land C sink the global carbon project (Le Quéré et al. 2014; these data are available in a downloadable spreadsheet at http://www.globalcarbonproject.org/carbonbudget)
   Even though the focus of the paper is on the pattern of IAV, we agree with the reviewer on the usefulness of a global inter-comparison of anomalies between products and with the GCP.
   In the new version of the manuscript we will therefore provide such a comparison, bearing in mind that GCP land fluxes are estimated as residual term from the atmospheric CO_2 budget and are therefore not completely independent from the inversion product.

2. Since the paper is ostensibly about inter-annual variability in the terrestrial C cycle (NEE) what aren't all data products detrended first (these are weak responses anyway caused by different assumptions made with each approach)? Then the authors would be better able to address the IAV (or anomalies) which seem to be the focus of the paper.
   IAV is generally defined as the temporal variability of the annual flux as generated by trend and residuals (Yuan et al. 2009), for this reason in the manuscript we analyzed both components and quantified the relative magnitude of the two (e.g. Fig 5 show that IAV is dominated by the anomalies). We will make this clearer in the new version of the manuscript.

3. What climate or weather data are used in MTE or the Jena inversion. Presumably neither used CRU (temperature) and GPCC (precipitation), as the authors of this paper chose to do? Thus, are analyses of climate drivers on IAV of NEE actually really just comparisons of distinct climate reanalysis products? Also, why not use the CRU precipitation product for consistency with the temperature data being used?
   MPI-MTE is based on the same climate drivers adopted in this analysis, namely CRU for temperature and GPCC for precipitation (Jung et al. 2011), while Jena-Inversion is not using any climate data in the flux calculation (with the exception of the wind field), being purely based on the atmospheric concentration measurements and an inversion transport model. GPCC precipitation was used instead of CRU for consistency with MPI-MTE, besides nowadays it is considered a better product as far as precipitation is concerned.

4. Much of the text in section 3 is heavy on the results with little discussion and interpretation of the key findings. Although some sections do communicate broader statements about the findings (e.g. lines 197-206), similar thoughtful development of ideas should be included throughout this section
   In the revised version we will improve the discussion of results.

5. Why aren't correlations of IAV with site – level or global-scale climate drivers shown for Fluxnet or Jena inversion products?
   The analysis of the global climate drivers of IAV was performed with the MPI-MTE because it is the only gridded product suitable for this purpose. The analysis has not been performed on the Jena Inversion products for the reasons explained in Reviewer #2 Specific Comments #4. Besides, a site level analysis is beyond the scope of the paper since it has already been performed in other papers (Luyssaert et al. 2007; Yuan et al. 2009; Wu et al. 2012).

6. I'm unclear what value is communicated by the calculation of CUP and CRP and would suggest removing these analyses from the paper. The finding that temperate and boreal systems have a stronger seasonal cycle in their CO2 drawdown seems obvious from atmospheric CO2 growth curves. Instead, if the purpose of these analysis is to "identify the role of photosynthesis and respiration as sources of IAV_NEE" (line 67), then it seems much more straightforward to just look at the IAV (or anomalies) of GPP and TER from the Fluxnet and MTE products directly. Then they could be correlated with climate drivers too? For example, at high latitudes do GPP and TER show strong temperature sensitivities, with anomalies GPP outpacing TER in warm years? Conversely, are Tropical GPP anomalies largely temperature related too, whereas TER shows less inter annual variability & climate sensitivity?
   Since MPI-MTE and Fluxnet come from the same data source, while the Jena Inversion is a completely independent product, we think that it can be of interest to see if patterns are consistent. Since atmospheric inversion does not allow to separate GPP and TER, we used CUP and CRP as their proxies and we tested the validity of this assumption. Results of this analysis are shown in Fig 9, where we can infer that CUP and CRP are dominated by GPP and TER respectively. Although not being a precise GPP and TER estimation, NEE_{CUP} and NEE_{CRP} are highly correlated with them. We believe that the analysis of CUP and CRP brings additional information when performed on the inversion product. In particular at high latitudes where GPP/TER partitioning performed as CUP/CRP is particularly clear. Please refer also to point 13 of the Specific comments. In addition, the separation of the ecosystem CO_2 fluxes in these two terms is becoming increasingly common since it allows the description of a plant phenology based on carbon fluxes instead of greenness indexes.

Specific comments

1. I'm not used to seeing citations in the abstract. Is that the format for this journal?
   Citation will be removed.
2. I'm used to seeing ecosystem respiration referred to as ER, but maybe the authors are used to using different conventions?
   Both symbols can be found in the existing literature.
3. Line 55 This single paragraph is a single sentence consisting of a very long list of NEE estimation approaches. Why not break this into a sentence about each approach and discuss strengths/ weaknesses of each?

We take this point and will discuss further the strengths and weaknesses of the different approaches in the Materials/Methods sections.

4. Line 67 organization of objectives i), ii), and iii) don't align with the organization of methods and sections 3.1, 3.2 and 3.3. Can the objective reflect the broader layout of the paper?
   Objective order will be reorganized in accordance with the other sections of the manuscript.
5. I'd suggest Line 73 are "LaThuile and 2015" two distinct references?
   These are two subsequent releases of the Fluxnet dataset namely La Thuille and the 2015 releaese which are available at: http://fluxnet.fluxdata.org/data/la-thuile-dataset/ http://fluxnet.fluxdata.org/data/fluxnet2015-dataset/ We will add the links to the manuscript.
6. It's not clear if or how data were re-gridded (e.g. [1] subtracting finer scale RETRO and GFED4 fire fluxes from the Jena inversion, or [2] for temperature and precipitation in Fig. 2)
   Fire and Meteo Data were regridded using the aggregate function of the R-package raster, a sentence explaining this will be added to the manuscript.
7. More broadly, is subtracting for fire fluxes even necessary? Do the 14 observations extrapolated to this global product even 'see' the effect of forest fires? Don't the atmospheric inversion products the global carbon project implicitly see the effects of these fires? If so, why should they be subtracted out here?
   Inversion based estimates of land CO_2 fluxes include the signal of forest fires while MPI-MTE and Fluxnet don't. To maintain consistency in the analysis and to allow a proper comparison between products we decided to exclude fire driven IAV from the Jena Inversion product. We will make this clearer in the revised text.
8. Line 80. There are enough abbreviations in the text already. Are these needed too? Their use in lines 210-219 makes the text very hard to follow.
   Since the acronyms were only used in Fig 6, following the reviewer suggestion, they will be removed from the text and explained in Figure 6 caption.
9. Line 109 Air should not be capitalized.
   The typo will be corrected.
10. 'Jena inversion' or 'Jena Inversion' should be used consistently throughout the text.
    The spelling of the product name will be homogenized throughout the manuscript.
11. Were any lagged correlations explored to see if climate variability affected NEE in the subsequent season / year?
    Lagged correlations were beyond the scope of the paper. In the present paper only spatial patterns of the IAV dependence on climate drivers were analyzed.
12. Standard deviation and IAV are used interchangeably throughout the manuscript, but I think they mean the same thing? If so, just one term should be used for consistency. If they are different, it should be clarified in the text.
    The reviewer is correct the two terms can be used interchangeably to identify inter-annual variability as stated in Section 2.2 L1-2.
13. Line 133 I have no idea what this means "the difference between the two determination coefficients was computed" or where this analysis is presented (Fig 9)? More broadly, I'm unclear how / why the authors tried to infer something about GPP and TER from the inversion product.

We agree with the reviewer that the sentence was not clear. We will therefore improve the description of the analysis based on what follows.

Results of this analysis are presented in Figure 10 and 11. NEE was linearly correlated with GPP and TER (for Fluxnet sites and MPI-MTE, for which GPP and TER are available) and with NEE_{CUP} (where CUP stands for Carbon Uptake Period), and NEE_{CRP} (where CRP stands for Carbon release period) for the Jena Inversion product (for which GPP and TER are not available, and CUP and CRP were used as their proxies) to detect which of the two processes drives the IAV of NEE. The difference between the R2 of the two regressions calculated for each pixel was plotted on maps in Fig 10 and in the climate space in Fig 11. The goodness of the assumption of CUP and CRP as proxies of GPP and TER was tested with the analysis shown in Fig 9.

14. Line 171. Why was IAV normalized using GPP estimates and not NEE, the later giving a real coefficient of variation (CV; grid cell standard deviation NEE / mean grid cell NEE). This should be clarified both in the text and caption. Also, shouldn't grid cell CV be calculated first, and then averaged over each climate bin?

    IAV was not normalized with NEE because the latter fluctuates around zero and can lead to unreliably high values of CV. Normalization using GPP (which is always positive) offers a more robust metric of relative IAV.

    The ratio of the means is a more robust estimation since mean of ratios is more sensible to outliers if compared to the ratio of the means, besides the latter gives more weight to points that bear more information.

    We will clarify these methodological details in the revised document.

15. Line 180 & Fig. 3 I am unclear what insight this figure provides to the manuscript and it's sparingly discussed in the text. It's used to justify the CV calculation in Fig. 3 (line 173), but as this is a standard statistical approach I'm not sure it's warranted? As such, should the display item just be removed?

    Fig 3 shows that in two datastreams (Fluxnet sites and Jena inversion) IAV increases monotonically and almost linearly with the productivity of the site. On the contrary MPI-MTE shows a different pattern, with a clear maximum followed by a decline of IAV in high productive sites. We think that this is due to the prominent role that FaPAR has in the MTE approach. Canopy greenness is particularly stable in the tropical humid forests (that are the most productive one) generating this unusual pattern of low relative IAV. We will discuss this aspect in further details in the revised version, since it is relevant to understand the general performance of the MTE model in the representation of the global IAV patterns.

16. Line 200. It seems like 'trends' in IAV should be driven mainly by environmental presses like atmospheric CO2 concentrations or broad-scale / chronic N deposition inputs. By contrast, climate variability, land use change, and fires should be responsible for 'anomalies' the dataset. Given that the Jena inversion depends strongly on modeled NPP products it's not surprising that is shows stronger 'trends' (see suggestion to detrend data, above). Also, it would be interesting to see if fire fluxes were not backed out of the Jena inversion (again mentioned above) how the magnitude and timing of anomalies from these two data products compared to anomalies in the atmospheric CO2 growth rate. This also could provide a better opportunities for the authors to illustrate the differences between the data products that are currently in the discussion.

As previously stated (Ref #1, General comment #1; Ref #2, Specific comment #4), the temporal dynamic of the land fluxes in the Jena inversion is totally driven by the atmospheric signal and fully independent from the prior, since the latter in this inversion scheme are time invariant. We argue that in the MTE product the lower IAV due to "trend" is due to the poor or lacking representation of environmental drivers like CO_2 or N deposition in this data product. We will make this clearer in the revised version.

17. Line 296 Carbon should be lowercase

    It will be corrected.

18. The conclusion is really just a summary of results already presented (and repeated from the abstract). I'd omit this text, or say something more broadly about what we can infer from the study.

    Conclusions will be reformulated in the new version of the manuscript as requested by the reviewers.

19. Fig. 6 & 8 I know abbreviations for each plant functional type are given in the text, but not using them in the caption or x-axis label bar make this figure hard to understand.

    Following the reviewer suggestion we will remove the acronyms from the text and we will adde their explanation to the figure caption.

20. Fig. 6 Aren't there enough observations to include error estimates (or box-wisker plots) for Fluxnet sites?

    Standard errors will be plotted for Fluxnet PFT IAV values.

21. Fig. 7 Caption and text should use the same (consistent) terminology here. I'm not really clear what is being compared here? How does one calculate a spatial correlation coefficient on two single values (e.g., correlation of IAV∼ mean temperature)?

    The correlation was calculated in a moving spatial window of more than 600 points, we retrieved a IAV value and a temperature/precipitation value for each pixel. This will be better clarified in the revised text.

22. Fig 7 The use of red-blue color bar on the left plots to show +/- correlation is confusing when on the right panels red-blue shows zonal mean correlations with trends or anomalies?

    We take this point and will change the colors in the barplot to avoid misinterpretation of the figure.

23. Fig 8 If this part of the analysis stays in the revised manuscript, I'd suggest the caption should be more descriptive (what are red and green bars).

    Additional information will be added to the figure caption to make the figure more readable.

24. Fig 9 I really don't understand what this figure is showing. The text & figure caption are not clear. More, the inset showing Western Europe seems strange. If this figure remains in the paper at all, would it make more sense to 1) omit the inset or 2) put it into supplementary material?

    This figure will be better explained in the new version of the manuscript based on what follows.

    The aim of the figure is to highlight the role of GPP and TER (for MPI-MTE and Fluxnet) and of their proxy NEE_{CUP} and NEE_{CRP} (for Jena Inversion) in building up IAV_{NEE}. The determination coefficients were calculated for each pixel (for the gridded products) or site (for Fluxnet) fitting linear regressions of IAV_{NEE} vs either GPP (TER), or NEE_{CUP} (NEE_{CRP}). The difference in determination coefficients of GPP and TER linear regressions (the same holds for NEE_{CUP} and

NEE_{CRP}) was used as a measure of which driver affects more IAV_{NEE}. Blue zones are GPP/CUP driven zones being the difference R_{GPP}^2-R_{TER}^2 (or R_{CUP}^2-R_{CRP}^2) positive while red zones are TER/CRP driven. See also Reviewer#3 answer #13. The inset was included in the graph because of the high site density of flux sites that characterizes Europe. Plotting an enlarged map allows in our opinion a better visualization of results.

25. Fig 10 I also cannot understand I'm unclear what the color bar signifies (DRˆ2)? Is this the difference between TER/GPP when NEE < 0 during uptake periods and GPP/TER when NEE > 0 for MTE? If so, what does this difference of ratios really less us? I also still unclear how this is translated onto the Jena data?

    Figure 10 summarizes results plotted on maps in Figure 9 in a temperature/precipitation space. Blue pixels are GPP/CUP driven climate classes, red pixels are TER/CRP driven climate classes.

BIBLIOGRAPHY

Jung M, Reichstein M, Margolis H a., Cescatti A, Richardson AD, Arain MA, Arneth A, Bernhofer C, Bonal D, Chen J, Gianelle D, Gobron N, Kiely G, Kutsch W, Lasslop G, Law BE, Lindroth A, Merbold L, Montagnani L, Moors EJ, Papale D, Sottocornola M, Vaccari F, Williams C (2011) Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. J Geophys Res 116:G00J07. doi: 10.1029/2010JG001566

Luyssaert S, Janssens I a., Sulkava M, Papale D, Dolman a. J, Reichstein M, Hollmén J, Martin JG, Suni T, Vesala T, Loustau D, Law BE, Moors EJ (2007) Photosynthesis drives anomalies in net carbon-exchange of pine forests at different latitudes. Glob Chang Biol 13:2110–2127. doi: 10.1111/j.1365-2486.2007.01432.x

Wu J, van der Linden L, Lasslop G, Carvalhais N, Pilegaard K, Beier C, Ibrom A (2012) Effects of climate variability and functional changes on the interannual variation of the carbon balance in a temperate deciduous forest. Biogeosciences 9:13–28. doi: 10.5194/bg-9-13-2012

Yuan W, Luo Y, Richardson AD, Oren R, Luyssaert S, Janssens I a., Ceulemans R, Zhou X, Grunwald T, Aubinet M, Berhofer C, Baldocchi DD, Chen J, Dunn AL, Deforest JL, Dragoni D, Goldstein AH, Moors E, William Munger J, Monson RK, Suyker AE, Starr G, Scott RL, Tenhunen J, Verma SB, Vesala T, Wofsy SC (2009) Latitudinal patterns of magnitude and interannual variability in net ecosystem exchange regulated by biological and environmental variables. Glob Chang Biol 15:2905–2920. doi: 10.1111/j.1365-2486.2009.01870.x