Responses to RC1 for interactive discussion:

The referee's suggestions are quite helpful, especially those concerning the need for additional explanations to clarify the usage of certain terms, the misleading use of certain terms which are employed interchangeably throughout the manuscript, the need for additional practical information about the algorithm's functionality, the need for additional transparency regarding the calculation of the seasonal cycle and trend components by CCGCRV, and the suggestion to scale back the interpretation of the results with regard to seasonal CO_2 anomaly detection. Some specific responses follow:

- Line 85: Sentence changed to note that the algorithm is designed for use with daily datasets.

- Line 96: Sentence changed to read "In our case, we select two different settings for the short-term filter…" This is meant to impart to the reader that the smoothing span and filter length are user-definable.

- We have no best estimate for the minimum time series length required to produce a "reasonable" fit. By default, the minimum required length in our study is that of the shortest record in our analysis (4 years at NOR), since we effectively assume that the results we obtain at these sites are generally as valid/useful as those obtained at sites with longer historical records. Admittedly, it would be nice to have a uniform setting for the historical record length that we use in all cases, but the problem is that many ICOS sites have come online only recently, and using only the past 4 years of data just for the sake of making the analysis uniform would have meant ignoring results prior to 2017 at the other 9 sites. This would mislead our readers as to the breadth of the algorithm's applications, and moreover, 4 years seems objectively too short for a good fit at the sites which have almost 10 years of usable data. We have now noted in the first paragraph of section 2.3 that the time period used for the calculations is user-settable (there is a line in the code for this purpose). In addition, there is a paragraph in the discussion section noting that the length of the historical record surely makes a difference for the results. We have added to this paragraph a sentence noting that in the future, as our stations accumulate longer and longer records, a "standard" record length (e.g. 10 years) may be adopted for added consistency.

- Regarding end effects, we feel that we have adequately circumvented this problem by calculating our $\pm 2\sigma$ envelope according to the procedure described in equation (3), where σ_d for any day, including the very last day in the record, depends on a sufficiently large residual dataset in the vicinity of that same calendar day over several years of data. Any end effects should thus be negligible.

- Line 105: While it is true that synoptic events do not occur only in winter and seasonal anomalies are not best characterized as occurring only in summer, dividing the analysis in this way seemed a logical choice. Firstly, as we mention in the discussion section, anomaly patterns appearing to be "seasonal" in length may simply reflect the unusual persistence or frequency of synoptic-scale weather patterns. This might conceivably be the case for unusually warm winters over Europe, where we would expect frequent or persistent positive NAO regimes to dominate. However, when we apply the 90-day smoothing span to the whole year and examine winter periods, we find that such winters appear mostly as flat, black lines on our anomaly graphs. This makes them, by our definition, not very anomalous. The more interesting winters are those featuring frequent cold weather patterns from the east, which do show positive CO_2 anomalies at the 90-day bandwidth. However, we feel that these winters are better viewed as just that (dominated by more frequent atmospheric transport patterns emerging from the east) and not as winters with an atypical seasonal suppression of photosynthetic activity. Terrestrial carbon cycle anomalies such as GPP anomalies are events to which we ascribe “seasonal” anomalies, and these do not appear to be the driving force of the excursions that we detect in the wintertime. If they were, we would also expect that warmer winters would coincide with increased GPP, enough so that we might observe negative seasonal-length CO2 anomalies during warm winters. We did not observe any such wintertime 90-day anomalies, at least not with our $\pm 2\sigma$ envelope definition.

In the summer, a span of 30 days rendered the figures a bit too cluttered to be able to clearly see the signals we were hoping to detect. Indeed, this is likely due to the fact that synoptic and seasonal signals can occur simultaneously and overlap with one another, as you mention. Our intention in using a 90-day smoothing span for the summer was actually to filter out synoptic signals to the extent possible. Synoptic anomalies certainly do occur in the summertime, but they are more difficult to tease out than they are in the winter, due to the contemporaneous effects of terrestrial carbon cycling processes which often occur at longer timescales. In addition, the NAO index is slightly less well-defined in the summertime, which increases uncertainty insofar as interpreting the causes of any synoptic signals observed. We therefore chose not to examine the algorithm’s ability to detect synoptic length events in the summer. We agree that the reasons behind our decision could be explained a bit more thoroughly in the manuscript and have added an explanation for our rationale in the third paragraph of section 2.

- Line 134: Yes, there is a line in the code which extracts only afternoon hours for non-mountain sites and nighttime hours for mountain sites prior to the analysis. The user could skip this step and keep all 24 readings for each day (or however many were available), or select a different time range. We have now noted this in section 2.1. In general, studies in our field tend to use afternoon values for non-mountain sites (e.g. Morgan et al., 2015; El Yazidi et al., 2018; Wang et al., 2018), and we decided to keep with this convention. We did run the analysis using all daily readings at all sites, and the anomaly patterns were not drastically different overall. However, using the mean of all 24 daily measurements renders the results less useful for flux inversion estimations, which generally rely on samples taken when vertical CO_2 gradients are lowest (Monteil et al., 2019), and would be inconsistent with ICOS flask sampling protocols, which recommend afternoon sampling at non-mountain sites and nighttime sampling at mountain sites.

- Figure 3: Plots of the deseasonalized anomaly patterns (δC_{30} and δC_{90}) for the TRN station have now been added to figure 3. The long-term trend for 2013-2020 at TRN has been added to figure 2. Our usage of the terms “seasonal cycle” and “long-term trend” has now been explicitly defined in the first paragraph of section 2.2. Our usage of the term “synoptic” is already defined in the ntroduction (paragraph 4). As to the specific example the reviewer gives (“Here, we have assigned variability longer than XX days to the trend component of the decomposition procedure, thus excluding it from the seasonal cycle”), we feel that this adequately discussed in the third paragraph of section 2.2, where we state that we use 667 days as the cutoff value for the long-term filter. As we understand it, variations at frequencies greater than 1 cycle per year are by default not considered a part of the seasonal cycle. It is thus implied that variations at frequencies longer than 365 days and up to a maximum of 667 days are assigned to the trend component. The term “localized” to describe nearby point sources of pollution has now been clarified in paragraph 6 of the introduction.

- It is true that we chose to examine “seasonal” events in summer and “synoptic” events in winter partially to illustrate the scope of algorithm, and some of what we thought were its most useful applications. As mentioned above, we also found that after trial and error, this was the most logical way to divide the analysis and interpret the results. We have now explained this rationale in the third paragraph of section 2.

- Line 250: Localized fluctuations were counted manually as a sort of ad-hoc way to score the skill of the algorithm. We might consider removing this step from the manuscript altogether, since it is not really a part of the algorithm per se. In the meantime, we have noted the manual nature of this step in section 3.1.

Line 321: It does seem a bit unfeasible at this point to conduct back-trajectory analyses over several months at ten separate sites. Less labor-intensive would be using monthly average NAO indices to test whether, for example, the CO_2 spike observed in July of 2018 corresponded with unusually intense or persistent blocking conditions. Then again, we already know that this was the case, to some extent. Thus, we agree that it is probably better to note that the algorithm merely produces the signals we expect to see, and not extrapolate too much as to why, or at least to mention that atmospheric transport may have also played a significant role in the summer of 2018. We have scaled back our interpretation of the results in section 3.2 and included a new paragraph noting that, although the timing of the seasonal anomaly patterns we observe in 2018 corresponds to the terrestrial biospheric signals observed across Europe in that year, we can not definitively say that these signals are what we detect, and have not established a causal link.

Line 333: Regarding long-term trends in variability, we came to the same conclusion as you in analyses conducted separately. Best we do not read too far into any trends in variability or anomalies that we think we observe. The final paragraph of section 3.2 has been removed, as it was largely based on speculation and not any scientific analysis, as noted.

Line 68: Sentence changed to note that the methodology is designed for application to ICOS station data specifically.

Line 410: The first paragraph has been edited to reiterate that the effects of summertime atmospheric transport have not been quantified in this study and as such, the algorithm’s capacity to detect exceptional biospheric episodes at the seasonal bandwidth has not be definitely determined.

Technical corrections:

- The opening sentence of the abstract has been re-worded to make it less ambiguous, as suggested by Referee 1.

- The term “weather” has been replaced with “atmospheric transport” in all instances. The lone use of the phrase “climatological occurrences” has been replaced by “meteorological occurrences.” The term “swath” used once in the literature review section to describe a window of time series measurements has been changed to “span.”

- Notation change: σ, 2σ and 3σ have been changed to ±σ, ±2σ and ±3σ where appropriate.

- In response to Referee 1, the terms “Gaussian” and “Gamma” have been changed these to “Gaussian curve” and “Gamma curve” for added clarity.

- Line 78 wording changed to read “at multiple European sites.”

- Line 227: No changes made. The sentence is intended to read “such as when NAO- or BLO regimes prevail.”

- The y-axis bounds in figure 5 have been adjusted to contain the full range of CH_4 readings at all stations.

- No readings exist for either CH_4 or CO_2 at GAT from Oct. 23 to Nov. 21, 2018, likely due to an instrument malfunction at this time. Before the CCGCRV fitting procedure is applied to the raw daily measurements, data gaps such as this are filled using a simple linear interpolation, as mentioned in the first paragraph of the discussion section and as recommended by, e.g., Pickers and Manning (2015). This may, in some cases, lead to the selection of “false positives” such as the one seen at GAT. This is, admittedly, a drawback of the algorithm, albeit one without an easy solution other than manual inspection. In the first paragraph of the discussion section, we have now alluded specifically to the gap at GAT as an illustrative example of the data gaps problem.

- Pickers and Manning (among others) note that the FFT algorithm used by CCGCRV requires time series data to be evenly spaced and without gaps. Their work has now been cited in the first paragraph of the discussion section.

<u>Responses to RC2 for interactive discussion</u>:

Thank you for your very helpful comments and suggestions. Specific responses follow:

- Line 321: It is true that additional analysis using meteorological data would help to better isolate the contributions of different factors to anomalous signal events. Although we feel that the similarities between the wintertime 30-day anomaly patterns in CO2 and CH4 strongly suggest a link with prevailing wind regime, pinpointing the causes of summertime 90-day anomalies is less straightforward. More conclusive results might be obtained by an atmospheric back-trajectory analysis, but as such an analysis would have to be conducted over several months at ten sites, its scope is a bit unfeasible, especially given the short duration of the lead author’s contract at LSCE (Alex Resovsky is now employed at ARIA Technologies, Inc.). As a compromise, we have scaled back our interpretation of the summertime anomaly patterns, particularly the anomaly patterns observed in the summer of 2018, which we admit we can not readily attribute solely to terrestrial carbon cycle imbalances. We now note, in section 3.2, that although the algorithm produces the signals we would expect to see for the summer of 2018 based on the timing and magnitude of continent-wide NBP anomalies known to have occurred during that period, we have not established a causal link in this manuscript.

- Line 338: Regarding data gaps, the best illustrative example of a problematic gap in our data would be the CH4/CO2 gap observed in October/November of 2018 at GAT. No readings exist for either trace gas at GAT from Oct. 23 to Nov. 21, 2018, likely due to an instrument malfunction at this time. As mentioned in the first paragraph of the discussion section, data gaps such as this are filled using a simple linear interpolation before the CCGCRV fitting procedure is applied, as recommended by, e.g., Pickers and Manning (2015). This interpolation procedure may, in some cases, lead to the selection of "false positives" such as the slight negative anomaly seen at GAT in late October of 2018. One can see from Figure 5 that this erroneous anomaly is linked to the fit of the LOESS curve to the artificially interpolated raw data; since the measurements on the left side of the data gap are near the lower bounds of the 2σ-envelope, the linearly interpolated datapoints just afterward are weighted downward, causing the LOESS curve to fall slightly outside the envelope. Using multi-year averaged seasonal cycles instead of harmonic functions from CCGCRV would likely not drastically affect the results, at least not in the case of the GAT example, since the false positive here is linked to the linear interpolation and not the envelope definition. The interpolation procedure could perhaps be improved by interpolating large data gaps with multi-year averaged seasonal cycle values, but this would raise a couple concerns: 1) any specific improvement would be difficult to ascertain; since the data themselves do not exist, any sort of imputation procedure ports some degree of uncertainty to the results, and 2) determining the multi-year seasonal cycle curve with CCGCRV would require an *a priori* interpolation of any data gaps, i.e. using a linear interpolation, followed by an *a posteriori* re-interpolation of the same gaps, followed by a secondary application of the CCGCRV fitting procedure. It is unclear whether any reduction in false positive detection, which appears to be quite rare overall, justifies this trade-off in efficiency. Nonetheless, we have now noted in the first paragraph of the discussion section that interpolation of large data gaps with multi-year average seasonal cycle values represents one potential way to reduce the detection of rare false positives linked to large data gaps. We also note here that we are prepared to test and implement this change in the methodology if absolutely required. We have alluded specifically to the 2018 data gap at GAT as an illustrative example of the data gaps problem.

Technical changes in response to comments:

- Line 142: Sentence changed to reflect the current nomenclature of the Global Monitoring Laboratory at NOAA (CCGG/GML). The link to the R code for CCGCRV has also been edited (line 145).

- Line 149: Sentence changed to read "Basically, a fit to a time series is first obtained using a linear least squares regression following the 'LFIT' protocol, in which a linear function describing the data is determined from an x^2 minimization of the residuals (Press et al., 1996)."

- End effects should be minimal when using our algorithm, which we feel is one of its strongest advantages. We feel that we have adequately circumvented the problem of end effects, which is well-known to CCGCRV users, by calculating our ±2σ envelope according to the procedure described in equation (3), whereby σ_d for any day, including the very last day in the record, depends on a sufficiently large residual dataset in the vicinity of that same calendar day over several years of data, and only minimally on that day's measurement. Thus, although the smoothed, detrended seasonal cycle, which is based on the polynomial part of equation (1), may be sensitive to end effects, these effects should only minimally affect the envelope calculation.

<u>Other technical notes</u>:

After an unfortunate technical incident which resulted in the loss of all figures for this manuscript (https://answers.microsoft.com/en-us/msoffice/forum/all/the-image-part-with-relationship-id-rid8-was-not/8f654849-8267-421b-a9cc-c73d5639cc25), the original data used in the analysis had to be re-obtained, which two former colleagues at LSCE were kind enough to help out with. The recovered data sets were not exactly the same as the original ones, specifically for the OPE and PUY sites, which had previously had slightly different starting dates. The new starting dates did not fundamentally affect the algorithm's performance or the results obtained at OPE or PUY, so although the starting dates for those two sites have now been updated in table 1, the results and analysis are the same as before. However, during the ordeal it became apparent that the data that had been associated with the HPB site in our previous submission was erroneous. The starting date for the L2 data at HPB has now been amended in table 1, and figures 4 through 9 now show the correct data for HPB. This is also the reason why there are some slight changes to the wording in section 3, mostly in the analysis of the results.

We also discovered that we had been using pre-L2 data at more sites than previously noted. This is the reason for the change in lines 129-132.