Review of Munassar et al. 2021 submitted to Atmospheric Chemistry and Physics

General

The study presents inverse modelling estimates of the net ecosystem exchange across Europe for the period 2006 to 2019 with a special focus on the exceptional years 2018/2019 and the analysis of influence of environmental drivers (temperature, moisture availability) on NEE. The applied methods have been established in previous work and are generally sound. However, a few general questions concerning the method should be addressed (see comments below). The manuscript is well organized and generally well written. Nevertheless, in some cases the description of results could be more precise. Overall, I recommend the manuscript for publication after a series of rather minor issues have been addressed.

General comments

Title: In the title and elsewhere the inversion system is labeled 'pre-operational'. However, there is no discussion whatsoever, why this term is used. Furthermore, if the current system is the pre-operational system, what do the authors envision as the operation system? Either this label has to be discussed in more detail or it should be removed from the text and title.

Anthropogenic emissions: Anthropogenic CO2 emissions are assumed to be well-known for the current study (L51) and are not updated as part of the inversion. How sure can we be about this? While the sensitivity of the inversion towards prior biospheric fluxes and prescribed ocean fluxes is investigated, no such analysis is done for the anthropogenic fluxes. Some kind of discussion of this assumption should be included in the manuscript. What is the potential uncertainty added by this assumption. This, especially in the light of inter-annual variability and the timing of anthropogenic emissions that exhibit larger uncertainties than annual national totals.

Specific comments

L44-46: Consider adding references to this statement.

L53 and elsewhere: The terms 'error' and 'uncertainty' seem to be used interchangeable throughout the manuscript. Consider using 'uncertainties' throughout and distinguish between random and non-random uncertainties, if that was the intention when using error (systematic) and uncertainties (random).

L83/84: Since 2006 IFS HRES has undergone several configuration changes, some of them affecting horizontal and vertical resolution. Please mention at which resolution the input data was available (spatial and temporal) for STILT and if and when configurations changed.

L85: The number of released particles seems very small. Over which time interval are these particles released? Other inverse modelling studies using Lagrangian transport models employ much larger particle numbers (e.g. Lauvaux et al. 2016) even though their domains and transport times are smaller. If only 100 particles are released per hour of measurements but residence times are evaluated in a grid of approximately (160 x 160 = 25'600) cells, it seems unlikely to get a statistically robust estimate of residence times.

L110: Depending on the height of the mountain, I would rather say that these sites experience free tropospheric conditions than residual layer conditions.

L145/146: What is this choice based on? Which values did these length scales take? Are these decay distances constant over the whole domain?

L150: Were the CarboScope ocean fluxes updated since the 2013 publication or were these also climatologies?

L152: What is the benefit of using EDGAR 4.3 over the much more recent EDGAR 6.0, which provides more temporally resolved emissions up to 2018? Since these emissions are taken as 'truth', I wonder if a sensitivity inversion with alternative anthropogenic emissions should have been conducted. Would it be possible to make an estimate of how large an uncertainty may be added by the prescribed anthropogenic emissions?

L165: The sample standard deviation when using only two or three samples is not a very robust estimator of the true standard deviation and is generally biased low. This should be stated as a warning when comparing these ensemble spreads with 'true' standard deviations as used in prior uncertainty.

L167/168: In Bayesian inversions, it is usually possible to calculate the posterior covariance directly. It sounds as if this is not the case here. The posterior uncertainty will also depend on the prescribed prior uncertainty. Here, it seems that the prior uncertainty is not considered but only the data-mismatch uncertainty is used. Please clarify.

Section 3.1: Results in this section only span the period 2006-2018 and exclude 2019. Although, this is correctly listed in Table 2, this limitation is not mentioned in the text and the title of the manuscript suggests that results would include 2019. Please clarify in more detail that different periods are used and why.

Section 3.1: Results in this section are only presented for Central, Northern and all of Europe. It would be interesting to see the results for the other regions as well. Could similar figures as Fig 3 and 5 also be provided for the other regions as part of a supplement? Especially since there are some references in the text to other regions (e.g. L191).

Table 3: Provide results for all of Europe. How is it possible that spread reduction is 95.1 % across all of Europe if Central Europe has a spread reduction of 96% and all other regions will have considerable less reduction? Without seeing the results for all European regions, the discussion remains unsatisfactory.

L216: This is already discussed in L195, where a reference to table 3 (with the same numbers) is given. Please consolidate.

Figure 7, L242-247: How is it possible that the innovation over south-eastern Europe is almost as large as in Central Europe even though there are no observations in this area (B0, B1)? Should the posterior in that case not stay very close to the prior?

L249: This sounds as if a different inversion system as compared to the previous section was used. Please clarify.

L250: It is the same starting time as in the previous section. Only the additional year 2019 seems different.

L259: 'Seasonal NEE': Did you mean summer? Otherwise, the described years with positive anomalies don't make sense.

L274/275: The biospheric signal is generally weak in winter (as predicted by the biosphere parameterisations). The IAV seen in the posterior may also be attributable to IAV in fossil fuel emissions that are not well represented in the used inventory. Please comment.

L283: Why was this temperature dataset selected over ERA-5, which was used to drive the FLUXCOM estimates?

L288/289: I do not see this at all. The summer anomalies in earlier years (2007, 2010, 2012) were much more dramatic. Even if we discard these because of the poorer data coverage (but why show them then in the figure?), summer 2018 does not seem too exceptional. Wouldn't it make more sense to discuss the growing season as a whole instead of summer and spring separately? Spring 2018 and 2019 look more exceptional to me than the summer months. Or limit the discussion to the whole year as is done below.

L312: Could this also be driven by a generally earlier start of the growing season (not just in evergreens) already towards the end of February? Especially in southern Europe many crops start developing around that time already. Finally, this may once again be a misattribution of anthropogenic emissions. Warmer winters mean less anthropogenic emissions, if these are not considered correctly or fully the "missing" CO_2 may be attributed to biospheric uptake!

L321: 'identical observations'. Do you mean an identical set of observation sites? Otherwise, this could be misunderstood in the sense that identical observations were repeatedly used each year.

L321: Why are none of the German ICOS sites used here? Were they not available in 2018 and 2019? Why restrict to only 16 sites when trying to analyse the spatial differences?

L329-331: This is not correct. Over the UK we can see basically no temperature anomaly in summer 2019, a positive SPEI anomaly but still largely increased posterior NEE. Please be more precise in the description.

L400: Most of western and central Europe do not experience any lasting snow cover anymore! Periods with snow cover are mostly limited to a few days. Colder winters also don't necessarily mean more snow as cold periods in central Europe are usually connected to easterly advection in high pressure systems with little precipitation. Overall, this 'theory' would need to be evaluated with additional datasets (snow cover, soil temperatures, etc.).

Figure 11: Why exclude fall here?

Definition of winter season: Which months are incorporated into the winter estimate of a specific year (X)? Jan X, Feb X and Dec X, or Dec X, Jan X+1, Feb X+1? Please clarify. If the first definition is used then there is no connection in the climatological sense and the interpretation may be more difficult.

Technical comments

L38: NEE was only defined in abstract. Please redefine in main text.

L103: Remove line. Seems to be a mistake.