Reply to Anonymous Referee #2

We thank anonymous referee #2 for the very positive overall evaluation of our manuscript and the very constructive comments.

In the following we repeat the specific comments of the referee in italics and add our replies in regular fonts.

P2 L42: "of the derived inversion increments" – please clarify what this means

we mean the difference between posterior and prior emissions. This should be a commonly used term. But for clarity we could add a definition in the text (e.g., P16, L452, where this term is used the first time in the main text).

P6 L189: "ignores any error correlation between different observations" – Please briefly discuss the impact of making this assumption between different 3hr periods. It is highly likely that a measurement is strongly correlated with the 3hr measurements either side. How can this effect be minimised? Please explain why this assumption is necessary.

We agree that that a strong correlation should be expected between consecutive 3-hourly time windows, e.g., between [12:00 - 15:00] and [15:00 - 18:00]. However, we use only one 3-hourly time window per day (see section 3.1), and the error correlation between the 3-hourly time windows of two consecutive days should be much weaker (and should be largely determined by the actual synoptic situation). For simplicity, we have chosen here the simple assumption of ignoring any error correlation, as assumed also in many previous studies.

P6 L196 Eq10: Please clarify the equation with the use of brackets especially where items are divided. Please also provide the units of e (I assumed g/m2/s), it would be useful to have this stated.

We would prefer not to use additional brackets (in order to avoid confusion with the use of brackets for parameters of certain variables), but to use instead a presentation with numerator and denominator to improve the readability of the formula. We will add the units of e.

P10 L301: "The matrix was scaled" – Please be more specific, which matrix? I assumed B the error covariance matrix? Fixing this to 20%, how did this compare to the other inversion setups?

Yes, the prior error covariance matrix is meant here. For the FLEXVAR inversions presented in the manuscript the aggregated total uncertainty depends on the corresponding covariance settings. E.g., for 'INV-E1-O1', the aggregated total uncertainty is 12% (1-sigma).

P10 L312: "using 25 vertical layers" – I assume these are concentrated near to the surface? Please add a brief sentence describing this selection.

Around 5 layers are within the boundary layer, 10 layers within the free troposphere, and 10 layers in the stratosphere. We will add a short sentence in the revised manuscript.

P10 L319: Please describe the rational behind the choice of the temporal correlation time scales.

The rationale behind this choice is that wetlands, rice, and biomass burning have pronounced seasonal cycles, while the "remaining sources" are assumed to have no or only small seasonal variations.

P10 L321: "function of local emissions" – Are these the prior emissions? What distance is 'local'?

Local emissions are the emissions of the grid cell in which the corresponding monitoring station is located. We use here the actual emissions (i.e., prior or posterior) of the corresponding model simulation.

P10 L323 – 325: This last sentence seems out of place to me. How does it relate to the actual "TM5-4DVAR" inversion being discussed in this section?

The TM5-4DVAR inversions are used both to calculate the baselines for FLEXVAR and for the model comparison.

P12 L369: Please specify which stations are classed as "mountain stations" as some are obvious others are less so e.g. Ochsenkopf, Beromunster etc? Also please describe which stations have the >200m difference between the model and actual orography imposed and what these release heights actually are, maybe simply add extra columns in Table 1 describing this height and class of station.

We will add in Table 1 a column indicating which stations are classified as "mountain stations" and a further column with the applied release heights.

P12 Tab1: Why are Tacolneston 100m data used? In 2018, the 185m inlet samples much more frequently and is obviously higher and better able to be simulated?

Unfortunately, the 185m sampling height was missing in our list of the FLEXPART simulations. It would have been a significant additional effort to re-run the FLEXPART simulations just for one further station level - therefore we had decided to use for that station just the available 100m level.

Indeed, it would have been preferable to use the data from the 185m sampling height. Nevertheless, the measurements from the 100m level should also be quite representative and well suited for the inverse modelling. Also, the station statistics (comparison of model simulations with measurements) are excellent for the 100m level of this station (see Figure S6).

P13 L373: "measurement uncertainty is set to 3 ppb" – How has this been derived? Most if not all observations come with an understanding of this quantity and this can vary between sites and over time. For instance why not use the variability in the CH4 observations across the 3hr period, the data are reported at up to 1 minute resolution? Also there are repeated measurements against standards, the repeatability of these observations also indicate how uncertain the measurements are.

The applied value of 3 ppb is a conservative estimate which should include also potential additional errors (e.g. due to sampling). In any case, however, the modelling errors (model representation

error) are usually much larger - therefore the assumed value for the measurement uncertainty has probably an only minor impact on the inversion results.

P13 L391: "Natural CH4 emissions were generally used" – When were these not used? The word 'generally' implies that in some instance they were not used, when they are not used, what was used? Table 2 implies they are always used.

Yes, we always used the natural CH_4 emissions from the GCP-CH4 data set. We will delete "generally" in the updated version of the manuscript.

P15 L415: "Offshore emissions over the sea are not included in the country totals" – Please explain the impact of this decision? The UK, Netherlands, Norway have significant emissions offshore in the North Sea. In the prior inventories how significant are these, and how does this impact on the conclusions that the UK+Ire totals are similar to what is reported given that the reported totals include these emissions?

Unfortunately, we do not have the information from the gridded emission inventories about the attribution of the offshore emissions to individual countries. However, it is interesting to note that the inversions generally significantly reduce the offshore emissions of the prior inventories (see e.g., Figure 6, where this reduction is clearly visible in the inversions of all three models).

P16 L455: Figure 5 is referred to here but is not shown until page 24. I think it should appear earlier in the document.

Figure 5 could be shown indeed earlier, e.g., directly after Figure 1.

P20 L530: Maybe a similar comment also could be made about NW France which always appears to have very enhanced emissions but is relatively far from emissions? Or do you think these are real?

(remark: we assume that the referee means here "is relatively far from *observations*")

The derived enhancements over NW France seems to be a much more robust feature of the inversions compared to the enhancements at the eastern domain boundary. The latter depend strongly on the chosen prior uncertainties (Figure S5), while the enhancements over NW France are visible basically in all inversions (and from all three models). However, in the absence of additional studies (e.g., regional measurement campaigns) it remains difficult to judge how realistic these derived emission patterns are. Clearly further independent validation studies will be required to evaluate the quality of the inverse modelling results.

P22 L575: "INV-E1-O2 compared to INV-E1-O1" – Please provide the actual values for both simulations as well, along with the +- uncertainties, Figure 5 is too crowded to really extract values.

We will add the actual values and their uncertainties in the updated version of the manuscript.

P24 Fig5: These figures are just too crowded and the different colours are impossible to discern e.g. inv-E1-O1-S2.1 and inv-E1-O1-S1 are indistinguishable. There is just too much information on each plot. On the RHS plots why are the E1 data repeated multiple times? The text for the range lines are blurred onto the lines. Please can this plot be improved?

We will update the Fig. 5. However, we plan to keep the data of the prior emissions for each inversion. Even if the prior values are identical for a given prior inventory (E1, E2, E3), their uncertainty depends on the chosen model covariance settings.

P25 L606: Please mention that Fig6 resolution has been downgraded compared to earlier, I assume to match TM5?

No, the resolution has not been downgraded. For the model comparison shown in this figure we used E3 as prior inventory, which had been provided at horizontal resolution of 1° × 1°.

P25 L613: Please provide the actual emission numbers rather than just the % change.

We will add the actual emission numbers in the revised version of the manuscript.

P27 Fig7: Similar comment to Fig5, it is hard to read the words/numbers in the RH plots, the plots themselves obscure the letters. Please can these be made clearer in some way? Although it is useful to see different inversions compared.

We will update also the Fig. 7 (consistently with the planned update of Fig. 5).

P29 L688: Summarising the results for each country grouping in a table would be very useful here.

We will consider to add a table in the supplementary material.

P29 L710: "emissions in September (Fig. 5)." - I found this impossible to see as there are too many lines.

We will try to improve the visibility of the seasonal variation of the prior emissions

Minor Text Comments

We will directly adopt most of the suggested minor text comments.