Dear Referee #3
Thank you for thoroughly reading and commenting the manuscript. Please find below the replies to your suggestions; each of your suggestions is followed by the corresponding reply in bold letters and (where appropriate) the actions taken to address it in the updated version of the manuscript in italics.

General comments:

The paper presents a multi-species inversion framework tested using pseudo-data experiments. Various assumptions are made to evaluate the sensitivity of the inversion, with an emphasis on the impact of error correlations across species and sectors. Overall, the paper presents an innovative approach to assimilate various atmospheric species in a single inversion framework. This study is clearly worthwhile publishing but lacks a better evaluation of the aggregation operator assumption (perfect prior emission distribution) and the impact of systematic errors in the system affecting the correlations in the gas-sector attribution problem. The Observing System Simulation Experiments (OSSE's) cover some of the assumptions with varying levels of uncertainties but several components are not carefully considered. The two major concerns here are the aggregation operator, that remains perfectly known and so the spatial distribution of the prior fluxes, and the assessment of correlations among sectors and across trace gases for the different species that remain very unclear. A last but less critical concern is related to the assumption that transport errors are similar across species, which is unlikely for $CH_4$ and $CO_2$ for example, rarely co-emitted (only $CO_2$-CO is discussed) and therefore affected by different problems in different parts of the domain. The work focuses primarily on random errors and ignores systematic errors that remain the main limitations in atmospheric inversions. Therefore, this study requires some additional experiments before publication, specifically addressing the error associated with the aggregation operator and errors in gas ratios for the different sectors.

- The use of an aggregation operator needs to be discussed. Hyper-parameters (here scaling factors for the sectors) are used to reduce the dimension of the problem but correspond to an assumption of perfectly-known distributions. The system should be evaluated not only under the "perfect spatial distribution" assumption, especially for $CO_2$ biogenic fluxes which are clearly not well-known. One suggestion to clarify the concern here would be to use VPRM as truth but assumes a different distribution when constructing the aggregated solution such as the posterior fluxes from Panagiotis et al. (2016). Other experiments could be designed here to test the aggregation problem. Similarly, the area defined by half of the total footprint is arbitrary and never tested nor justified. Why 50% was used? How much variations are expected within that area which would affect the error correlations? If a power plant is located near an airport, how would that affect the CO/$CO_2$ correlations and therefore the homogeneity within the aggregated area?

**Note that we do actually not focus on the domain total, as we believe it is not reasonable to constrain the whole European domain when pseudo-observations are focused only around a single city; for this reason we chose the region marked by the 50% footprint area, that contains most of the surface influence. We suggest to add the following sentence at page 8 – line 30:**

*… into physically representative quantities. As the pseudo-observations are clustered around a single location (Frankfurt), fluxes over the whole European domain can very likely not be constrained. Therefore, as spatial aggregation scale we chose a domain…*

**To the main point of this comment, we actually do not exactly assume perfect knowledge of the spatial distribution of total emissions; it is only within each sector and fuel type the spatial pattern of the emissions are assumed to be known.**

**We admit that the modeling framework that we set up is not particularly well suited to investigate the aggregation error. However, the chosen domain is quite small, and the total fossil fuel fluxes are divided according to species, emission categories, fuel types and months. This result in numerous degrees of freedom available to resolve biosphere fluxes, and for this reason we expect the aggregation error not to be a particularly important source of uncertainty.**

**In our inversion, as in all inversions, the near field is a critical domain in the arising of systematic errors. The better way to address systematic errors is of course by comparing model outputs with real observations, which are currently unavailable. The bias errors in atmospheric inversions making use of airborne measurements will have to be addressed anyway, once real observations from IAGOS will be available. For this reason, in this paper we chose to focus on random errors instead.**

**We suggest to add the following sentence at page 8 – line 31:**

*… 2011 (Fig. 1). Note that by using this aggregation scale we assume perfectly-known distribution within a given flux category that can result in aggregation error, especially with respect to biogenic fluxes, that are not so well known as anthropogenic fluxes. However, the chosen domain of aggregation is quite small, and the total anthropogenic fluxes are divided according to species, emission categories, fuel types and months. This result in 69 numerous degrees of freedom per month for each anthropogenic species and 10 degrees of freedom per month for the biospheric fluxes; for this reason we expect the aggregation error not to be a particularly important source of uncertainty.*

- The discussion about error correlations across species is confusing. How did you define the emissions for the different sectors? Have you assigned gas ratios to various sectors? If so, what are these ratios? Some of the discussions are related to using CO2 and CO data to diagnose gas-to-gas correlations, but the exact definition of the emissions of the different gases for each sector has been defined in the inversion system. Or maybe the sectors are unrelated for each gas? The different sectors have ratios in terms of trace gas emissions but these emission ratios vary regionally. This section needs to be explained in more details. The assumptions made here should also be tested in the inversion framework.

**Emission ratios are not used here, but we used instead bottom-up calculated emissions for each of the three gases, using different emission sector-specific factors, which are for CO also region-specific. These country emissions are then gridded consistently with geospatial proxy data that are representative for the emitting activity, common to all species for the multi-species sources.**

**We suggest the following changes to the text:**

**Add at page 7, line 1:**
*…on our regional European domain. For each of the three anthropogenic modeled species ($CO_2$, CO and $CH_4$), different emission maps are used as input. Temporal profiles are then applied to these sector- and fuel-specific emission maps.*

**Replace at page 14, from line 5 to line 24:**
*CO2 and CO are dominated by combustion sectors. The most important emission sectors for CO2 are energy, industry, transport and building, each contributing 7-10 MtC y-1 in July and 6-14 MtC y-1 in December. Dominant fuels for CO2 are coal, gas and oil, whose prior fluxes (pseudo data) have a magnitude of 6-11 Megatons of carbon per year (MtC y-1) in July and 8-14 MtC y-1 in December. For CO the most important emission sector is heating of buildings during winter contributing a 0.19 MtC y-1 flux with only secondary contributions from industry and transport with a magnitude of 0.04 MtC y-1 and 0.05 MtC y-1 respectively (during July and December). The dominant fuel for CO is biofuel with 0.19 MtC y-1 emissions during winter. The secondary industrial and transport contributions originate in summer from oil and biofuels with a magnitude of 0.06-0.08 MtC y-1 and from agricultural waste burning with a magnitude of 0.06-0.11 MtC y-1.*

*Contrary to CO2 and CO, CH4 is determined by non-combustion sectors, more specifically by a contribution of 0.15 MtC y-1 flux from agriculture (manure management and rice cultivation) in July with secondary contributions from waste and energy with a magnitude of roughly 0.06-0.08 MtC y-1 in both July and December. Other non-combustion sectors, in particular wastewater treatment and landfills contribute to a total of 0.16-0.24 MtC y-1 of emissions. These non-combustion sectors contribute to less than 20% of total CO2 emissions, with 1.13 MtC y-1 from the cement and lime industry and less than 20% to the total CO emissions (0.03 MtC y-1 from the metal industry).*
*The contribution to $CO_2$ from biospheric primary production (a sink for atmospheric $CO_2$) is about 100 MtC y-1 in July, which drops to almost zero in*

*December, while respiration values are 50 MtC y-1 in July and roughly 150 MtC y-1 in December.*

- CO biogenic fluxes: the paper does not address the problem of CO biogenic fluxes during the growing season. Warm days in summer correspond to large amount of biogenic VOC's being emitted from the vegetation, producing CO to non-negligible levels. This issue should be discussed if not addressed. How would this problem affect the ability to retrieve the truth?

**To discuss this issue we propose to add the following at page 14, line 4:**

*Note that our modeling framework does not allow for simulating CO biogenic fluxes during the growing season. Warm days in summer correspond to large amount of biogenic VOC's being emitted from the vegetation, producing CO to non-negligible levels. According to Hudman (2008), anthropogenic emissions accounts for only 31% of CO emissions in the US during summer. Conversely, according to estimates from EDGAR, CO anthropogenic emissions during summer are about 18% of the annual anthropogenic emissions. Combining these two results, one could conclude that CO production from biogenic sources accounts for roughly 42% of total annual CO emissions.*

*In general, the absence of some emission sources in an inventory is equivalent to the assumption of having point sources not included in the emission map, but still contributing to the measurements. The inversion scheme would typically react to this by assigning such point sources in some other sector other fuel type. As a result, the posterior enhancements would be biased low in proximity of that point sources, and (slightly) biased high for influences from other regions with the same sector or fuel type. This issue should definitely be considered in a future study making use of actual CO, $CO_2$ and $CH_4$ observations from IAGOS but has limited effects on this paper, as our main focus is on the benefits of inter-species correlation on the posterior uncertainty in the frame of a synthetic experiment.*

- When you constructed your error correlations for CH4, transport errors are unlikely to be highly correlated as CH4 is only partially co-emitted with CO2 and CO. Large emissions from NG production and farming activities are uncorrelated with biogenic or fossil fuel consumption. This problem should be addressed here. If transport errors, which are spatially variable, affect CH4 and CO2/CO in different ways, the error correlation would be affected. Additional experiments using incorrect error correlations would quantify the sensitivity of the inverse fluxes to the assumptions made in prior errors.

**This is a very useful suggestion, which we followed now. We propose to add the following at Page 14, Line 29**

*"Improper characterization of the error correlation may result in systematic bias in the posterior estimate. As mentioned in Sect. 2.1.6, inter-species correlation, the correlation between different fuel types and the correlation between different emission sectors in **Sprior** is assumed equal to 0.7 (Sect. 2.1.4). To assess how well the system will reproduce the 'true' fluxes with incorrectly specified correlations, a series of experiment was performed in which the inter-species correlation in*

***Sepsilon*** *remains equal to 0.7, while the three correlation coefficients in* ***Sprior*** *assume different values ranging from 0.1 to 0.9. Table 5 shows the residuals between total annual posterior fluxes and total annual true fluxes for the five simulated species, derived similarly as for Table 4.*
*We found that for all species the uncertainty reduction increases with correlation. More precisely, from correlation 0.1 to 0.9, the annual uncertainty reduction for anthropogenic $CO_2$ increases from 6.5% to 40.9%, while the increase is lower for GEE (from 64.6% to 65.2%) and respiration (from 35.1% to 36.8%) because the biospheric fluxes are independent from other species. For CO, the uncertainty reduction increases from 40.6% (with correlation 0.1) to 57.5 (with correlation 0.9). The annual uncertainty reduction for CH4 increases from 32.6% to 59.0%.*

*In addition, the posterior-truth biases are always lower than the prior-truth biases. The posterior uncertainty values are almost always larger then the corresponding bias values, except for CO with prior correlation equal 0.8, and fossil fuels $CO_2$ with prior correlations equal to 0.6, 0.7 and 0.9. Thus, except for these few cases, the posterior is not significantly different from the truth. Conversely, the prior (not shown) is significantly different than the truth in the majority of cases for fossil fuel fluxes, and in some cases also for biogenic fluxes. The effect of assuming the incorrect error correlations appears to be in general small. Following this result, the fact that $CH_4$ is only partially co-emitted with $CO_2$ and CO should not affect the inversion in a strong way.*

| Correlation | Post-Truth $CO_2$ ff | Post-Truth CO | Post-Truth $CH_4$ | Post-Truth GEE | Post-Truth Respiration |
|---|---|---|---|---|---|
| 0.1 | -17.2 ±17.6 | -0.1 ± 0.3 | -0.2 ± 0.4 | -21.0 ±27.5 | 12.9 ±26.1 |
| 0.2 | -13.6 ±14.4 | 0.1 ± 0.3 | -0.2 ± 0.4 | 4.8 ±27.5 | -5.8 ±26.1 |
| 0.3 | -12.8 ±12.0 | -0.2 ± 0.2 | -0.2 ± 0.3 | -1.2 ±27.4 | 0.5 ±26.1 |
| 0.4 | 2.9 ±10.1 | -0.1 ± 0.2 | -0.1 ± 0.3 | -21.1 ±27.4 | 23.1 ±26.0 |
| 0.5 | 0.5 ± 8.6 | -0.1 ± 0.2 | -0.2 ± 0.3 | 17.8 ±27.3 | -8.3 ±26.0 |
| 0.6 | 25.2 ± 7.3 | 0.1 ± 0.2 | -0.3 ± 0.2 | 5.7 ±27.3 | 6.6 ±25.9 |
| 0.7 | 25.6 ± 6.1 | -0.1 ± 0.2 | 0.1 ± 0.2 | 16.8 ±27.3 | -7.2 ±25.8 |
| 0.8 | -0.9 ± 5.0 | -0.2 ± 0.1 | 0.1 ± 0.2 | -5.8 ±27.3 | 23.3 ±25.7 |
| 0.9 | 13.8 ± 3.7 | 0.1 ± 0.1 | 0.2 ± 0.1 | -10.2 ±27.2 | 14.5 ±25.5 |

Table 5: Residuals between total annual posterior fluxes and total annual true fluxes for the five simulated species (in MtC/yr) and different inter-species correlation values in the prior error covariance matrix (first column). The corresponding posterior uncertainty was added for each Post-Truth value.

- The problem of unreported sources in CH4 inventory is not addressed at all. Recent papers have discussed the lack of information for natural gas and oil production operations, or from recent and old mining areas. How would unreported sources affect the inverse solutions? This question comes back to the aggregation operator.

**To discuss this issue we propose to add the following at page 14, line 4:**

*Our modeling framework is currently not well suited to account for unreported sources of* $CH_4$ *due to the lack of information about natural gas and oil production operations, or from recent and old mining areas.. Many recent studies have discussed the problem, mainly referring to shale basins exploited via hydraulic fracturing in the US (e.g. Kort et al., 2016; Karion et. al, 2015; Lyon et al., 2015). For example, Karion (2015) concludes that EDGAR underestimates methane emissions associated with oil and gas industry by a factor of 5 in the US.*

*However, the situation over the European continent may be quite different. In a review about risk assessment of shale gas development in the UK, Prpich (2015) reports that the European Union is generally much more cautious about unconventional oil and gas sources, while a recent study on a methane plume over the North Sea (Cain et al., 2017) concluded that the bulk signature of said plume originated from on-shore coal mines and power stations in the Yorkshire area.*

*In general, the absence of some emission sources in an inventory is equivalent to the assumption of having point sources not included in the emission map, but still contributing to the measurements. The inversion scheme would typically react to this by assigning such point sources in some other sector other fuel type. As a result, the posterior enhancements would be biased low in proximity of that point sources, and (slightly) biased high for influences from other regions with the same sector or fuel type. This issue should definitely be considered in a future study making use of actual CO,* $CO_2$ *and* $CH_4$ *observations from IAGOS but has limited effects on this paper, as our main focus is on the benefits of inter-species correlation on the posterior uncertainty in the frame of a synthetic experiment*

- The utility of the figures showing the multiple error covariance matrices for the different cases remains limited. The information content would be better described with words or mathematically. Readers cannot extract useful information from contour plots of covariance matrices. They could remain part of the paper but as part of the supplementary information. A table could also synthesize the various assumptions tested in the inversion system.

**We propose to add axis label to Fig. 2,3,4 to increase readability. Such axis should identify different species, emission sectors, fuel types and vegetation categories. In addition, we suggest to introduce two different equations (see below) to describe mathematically the error structure in the different cases.**

*Eq6.2:* $S_{multi} =$

$$\begin{matrix} X_{co,co} & X_{co,co2} & X_{co,ch4} \\ X_{co2,co} & X_{co2,co2} & X_{co2,ch4} \\ X_{ch4,co} & X_{ch4,co2} & X_{ch4,ch4} \end{matrix}$$

*Eq6.3:* $S_{single} =$

$$\begin{matrix} X_{co,co} & 0 & 0 \\ 0 & X_{co2,co2} & 0 \\ 0 & 0 & X_{ch4,ch4} \end{matrix}$$

*Note that each element in Eq 6.2 and 6.3 is a sub-matrix. In the case of $S_{prior}$, each element of such sub-matrices indicates the covariance between different flux categories. Conversely, in the case of $S_{epsilon}$, the sub-matrices show the covariance between different observations.*

Technical comments:

3-1: Consequently, intercomparisons...
**The text was edited according to the suggestion**

3-3: the international level
**The text was edited according to the suggestion**

3- 1st paragraph: This paragraph is confusing and not always following a logical path. Prediction skills and emission reduction are two different problems not directly connected to each other. Explain better the broad context of this study by focusing on the main general issues and clarify which one you are trying to address here.
**The paragraph was rephrased as follows:**
*As widely recognized at the international level, there is a need for reduction in anthropogenic emissions (IPCC). This however implies the necessity for reliable climate predictions from atmospheric models in order to allow policymakers to take informed decisions. Unfortunately, current climate predictions are hampered by excessive uncertainties; for example intercomparisons of different models show important differences on their predictions as shown in Friedlingstein (2016). This makes it difficult to assess the better environmental policies to implement. Because most biogenic fluxes …*

3-10: A commonly used approach to estimate...
**The text was edited according to the suggestion**

3-13: Actually, the uncertainty reduction relies purely on the assumptions made in the system and not on the effective ability of the system to produce a reliable solution. Bayesian system assumes that data will improve the a priori by construction. Explain better what you mean here.
**The text was modified as follows.**
*As the main goal of this study is to assess the benefit of inter-species correlations in reducing the uncertainty of the posterior state space, we are particularly interested*

*in the effects of such correlations on the uncertainty reduction, defined as the difference between prior and posterior uncertainty normalized by the prior.*

3- 2nd paragraph: Several papers are missing here. For example, CO2-CH4 inversion using satellite data (Pandey et al., 2015) or the optimization of co-emitted species (Brioude et al., 2012), and early work on delta 13-CO2 by Enting et al. (1995). The authors should dig into atmospheric chemistry studies where several studies have addressed the use of multiple co-emitted species to constrain emissions at small scales.
Previous studies using multiple species to constrain emissions should be introduced here, even without having used a formal inversion framework, such as urban studies over Los Angeles (e.g. Peischl et al., 2013). The optimization problem is equivalent and relies on similar ideas to constrain the emissions.
**We replaced the text at Pag. 3, lines 20-22 with:**
*Several studies have made use the correlations among different species. One of the first example is the work from Enting (1995) on $CO_2$ and $^{13}CO_2$, while Brioude (2012) attempted to derive a $CO_2$ emission inventory without a prior emission estimate, instead using inventories of CO, $NO_y$ and $SO_2$. Similarly, Peischl (2013) made use of CO and $CO_2$ inventories to help quantifying sources of $CH_4$ in the Los Angeles basin. The ability of measuring multiple species has been proved useful also in remote sensing. For example, Pandey (2015) made use of simultaneously retrieved $CO_2$ and $CH_4$ total column to reduce scattering effect. Further examples of studies making use of co-emitted species can be found in the frame of atmospheric chemistry (Konovalov et al., 2014; Berezin et al., 2013; Pison et al., 2009). More focused on exploiting inter-species correlation to reduce uncertainty in the frame of Bayesian Inversion, Palmer (2006) made use of CO2-CO correlations to improve an inversion using data from the TRACE-P aircraft mission, while Wang (2009) employed a similar method using satellite data, obtaining a reduction in the flux error of a $CO_2$ inversion.*

5-24: This technique assumes that the wind direction and speed are comparable near the surface and at 2km high. Mass-balance studies have shown that this is often not the case (e.g. Karion et al., 2015). Free troposheric air represents different air masses due to the wind direction and speed gradients in the vertical. This assumption would need to be tested with the particle model.
**This is a misunderstanding. We do not rely on winds within the mixed layer and the wind above to be comparable, as our transport operator H represents the mixed layer enhancements appropriately. We added the following text to**

*... a single footprint is derived. To represent the mixed layer enhancements, the footprints for receptors within the boundary layer are averaged, and the footprint for the free tropospheric receptor is subtrated from this, resulting in a footprint for the mixed layer enhancements. This footprint is then matrix-multiplied ...*

7-3: What about CO biogenic fluxes? During warm summer times, biogenic CO

fluxes represent a significant fraction of the signals. Did you ignore this contribution in your study?
**A similar comment has already been addressed (see above).**

2.1.3 To reduce the dimension of the state vector, you assume here that the spatial distribution of the prior fluxes and emissions are perfect, using an aggregation operator. This approach is reasonable for fossil fuel emissions but less convincing for biogenic fluxes.
**A similar comment has already been addressed (see above).**

12-28: How did you take into account the truncation of the prior errors? Did you adjust the truncated random perturbations to match the non-truncated assumption made in the prior error covariance matrix?
**The error realization is obtained by multiplying a randomly generated, normally distributed vector with the prior error covariance matrix. This ensures that such realization has the same error correlation of the prior uncertainty. Where the result of such matrix-vector product is negative, the same operation is performed recursively until all elements of the state vector are positive. We suggest adding the following text at page 12 line 24:**
*… to avoid negative state vector values. In detail, the error realization is obtained by multiplying a randomly generated, normally distributed vector with the prior error covariance matrix. This ensures that such realization has the same error correlation of the prior uncertainty. Where the result of such matrix-vector product is negative, the same operation is performed recursively until all elements of the state vector are positive. This ensures that the difference …*