

All referee comments are given in ***bold italics***, replies in plain font.

Anonymous Referee #1

General Comments:

The prior emissions the authors use are always biased high. I have concerns whether these biased priors may result in posterior emission estimates that are also biased high, especially for an inversion configuration that do not consider correlated errors in prior emissions (e.g., S-MS). The authors may consider including priors that are biased low and see if they obtain similar posterior emissions.

We performed additional sensitivity inversions with 30 % lower and 30 % higher priors compared to our BASE case. Although the new low priors are still often larger than the a-posteriori values, we could not observe a strong influence of the total prior emissions onto the posterior emissions. As an indicator, the ratio between the high and the low prior sensitivity inversions was calculated for the prior and the posterior emissions. The ratio was 1.85 for the prior emissions, whereas it ranged from 0.93 to 1.25 for the posterior emissions and most regions and compounds. Consequently, in most cases the range in posterior emissions spanned by these variations in the prior was smaller than the estimated uncertainties of the posterior emissions. Exceptions were HFC-152a emissions from Greece, which were significantly larger for the high prior, and HCFC-142b emissions from Greece and Turkey. However, in the latter case the posteriori uncertainties were still larger than the range of these sensitivity runs. This clearly indicates that especially for the well simulated species the dependency on the prior emission level is not the main source of uncertainty. A discussion of these additional sensitivity inversions was included in the revised manuscript.

The authors derive temporally constant emissions from their inversion framework. This may result in larger model-data mismatch errors and an under-fit of their atmospheric data comparing to inversions that estimate time-varying emissions. This problem could be more severe for chemicals that have strong seasonal cycles in their emissions (e.g. HFC-134a, HCFC-22, HFC-125). Conducting an inversion with deriving varying emissions could be beneficial to further test the robustness of their results.

The seasonality of the emissions was initially not targeted, because our observations in the Eastern Mediterranean do not cover a complete annual cycle and temporally variable posterior emission estimates may suffer from this lack of observations. However, we performed one additional inversion with seasonally variable emissions of HFC-134a (the most abundant and best simulated compound). As expected, we find mixed results for the Eastern Mediterranean, where the maximum posterior emissions were derived for the fall (SON), not the summer (JJA). However, this is mainly due to the lack of observations in this period and the posterior staying close to the prior. In Western Europe we observed a clear seasonality with the expected summer peak for Italy, Germany (Central W), the Iberian Peninsula and the British Isles, but not for France and the Benelux region. Total annual emissions in the regions experiencing a seasonal cycle were slightly enhanced compared to our BASE scenario. In the revised manuscript we discuss seasonality in the emissions in general, reviewing previous studies, and specifically for HFC-134a along with the results described here.

Specific comments:

Lines 25-27. "The eastern Mediterranean is home to...under the Kyoto and Montreal Protocol". This sentence seems unnecessary in the abstract.

In our opinion, it is important to point out to the reader the diverse character of the analyzed area, due to both, countries with long established emission bans and countries where regulations have only become effective in recent years.

Lines 42-44. “reduction of the uncertainties by 40 – 80%”. This reduction is for which region?

This number describes the range of uncertainty reduction over the entire inversion domain. For clarification purposes, the wording has been adjusted in the manuscript.

Line 51. “all long-lived halocarbons are potent greenhouse gases”. This is inaccurate. Gases such as CH₃Br and CH₃Cl are not potent greenhouse gases.

That’s correct, thank you. The wording was changed to “most” in the revised manuscript.

Lines 57 – 59. “To track the development of CFC and HCFC emissions to the atmosphere,...” I don’t think the reporting required by the MP is to specifically track emissions.

It’s correct that the reporting does only indirectly track the possible emissions of halogenated substances to the atmosphere, due to the non-emissive use of a lot of these compounds. It is not the specific reason of the reporting to the MP which is interested in the production, consumption and trade of these substances. We therefore changed this sentence in the revised manuscript to agree with the correct purpose.

Line 71. “15% of the original value”. Inaccurate. Maybe consider mentioning 15% of the baseline value during the years of xx?

Because “original value” is very unspecific, we changed the wording to “baseline value” in the revised manuscript, according to the referee’s suggestion. Due to different groups having different baseline periods and because these periods are not yet important for the results of this paper, we just mentioned the underlying 3-year baseline period though.

Line 168. Please spell out “NNE-E”

This was changed in the revised manuscript.

Line 183. “three-hourly aggregates”. Do you mean “three-hourly averages”?

We used “three-hourly aggregates” due to consistency, as it has been used in other studies based on the same methods before. Because the aggregation method is based on averaging, we changed the wording to “three-hourly averages” in the revised manuscript to be more specific.

Lines 200 – 207. Transport simulation. Why do the authors choose to run their transport simulation at altitudes that are different from their actual sampling altitudes? How much different are the derived sensitivities at their simulation altitudes and the actual sampling altitudes?

The simulation altitudes are different from the sampling altitudes because of the terrain’s representation in the underlying ECMWF model. Due to the resolution of 0.2° by 0.2° the mountainous terrain is smoothed. Therefore, the release heights should be lower than the actual sampling heights in order to catch the effect of the topography on the atmospheric flow. The release altitudes for Jungfrauoch and Monte Cimone were chosen based on tests and experience of previous studies (e.g. Keller et al., 2012), which used the same transport model and similar inversion setups. The sensitivities of Finokalia were simulated using release heights of 150 m.a.s.l. and 250 m.a.s.l. with no significant difference, which is based on the fact, that the source regions for this station are mostly located far away across the Mediterranean Sea, providing a sufficient timeframe for mixing and dispersion.

Lines 210 – 212. 1.7% degradation of HFC-152a emissions. Is this for the summer or an average for the whole year? Please be clear.

As mentioned in the manuscript this calculation is based on an average lifetime of HFC-152a and is therefore to be understood as an average degradation rate during the 10-day transport period, not specifically during summer. We clarified this in the revised manuscript.

Lines 289 – 290. HCFC-142b is mostly used as a foam blowing agent whereas HCFC-22 is used as refrigerant in air conditioning. So, they are not necessarily collocated. Assuming they have the same regional emission shares is not a good assumption for creating HCFC-142b prior emissions.

In the absence of better alternatives, we used the collocation of HCFC-142b and HCFC-22 based on the theoretical assumption, that emissions of these compounds are likely to be collocated on a larger scale (order of 100 km) due to existing regulations.

Line 293. Which version of the EDGAR inventory do you mean? Be clear.

We used EDGAR v4.2 and EDGAR v4.2 FT2010. The description was changed in the respective paragraphs and in the references.

Lines 307 – 321. Uncertainty covariance matrix B. Please explain why giving correlation lengths of 200 km and 5 days? How sensitive are the derived emissions to these correlation lengths? Also be clear on whether you considered anti-correlated errors between BE and BB.

The spatial correlation length scale was chosen based on findings in previous studies where a maximum likelihood optimization was used to estimate covariance parameters (Brunner et al., 2012, Henne et al., 2016). In this study, no such parameter optimization could be carried out, therefore, we followed the range used in previous work. The spatial correlation length scale does not largely impact region total emissions when varied within reasonable range (100 – 500 km). Lower values often lead to undesired dipole patterns in the posterior adjustments, whereas larger values lead to very smooth patterns that do not allow the inversion to pick up differences in neighboring regions. 200 km seems to present a reasonable value in the golden middle. Similar arguments can be found for the temporal correlation length of the baseline. One important argument for this value is the fact that synoptic systems, which are responsible for the advection of different baseline concentrations, exhibit time scale of about 5 days as well. In the case of the baseline concentrations of the compounds discussed here the influence of the correlation time scale is not very large because the baseline uncertainty itself is relatively small, so that adjustments to the baseline are also relative small and do not alter strongly with the temporal correlation time scale. No covariance between emissions and baseline were considered in the prior. However, we checked the posterior covariance between emissions and baseline to rule out that large negative covariance existed.

Lines 325 – 327. Calculation of σ model was derived based on the prior simulation and observations. This approach would likely overestimate the model-data mismatch errors and result in an under-fit of atmospheric data.

The procedure to estimate σ_{model} was not described correctly in the manuscript. Actually, we are not using the RMSE from the prior simulation, but iterate the inversion 3 times using the RMSE from the previous posterior simulation in the current iteration while keeping all other parameters and prior fields the same. This approach is similar to that suggested by Stohl et al. (2009) and results in a more reasonable model-data mismatch error. The description was updated in the revised manuscript.

Lines 368 – 375. Please indicate in the paragraph whether you consider results from S-NFKL and S-OFKL into the final emission uncertainty estimation

The results of S-NFKL and S-OFKL were NOT used in the final uncertainty emissions but were solely run to evaluate the benefit of additional measurements in Finokalia and the influence of additional measurements at the existing AGAGE sites. This was already pointed out in the original manuscript in section 3.4 (L572-580) but is now mentioned at this location in the revised manuscript.

Lines 403 – 404. Inaccurate description on HFC-134a and HFC-125 usage. Only HFC-134a is mainly used in mobile air conditioning. HFC-125 is mainly used in commercial refrigeration and residential air conditioning.

This was corrected in the respective paragraph of the revised manuscript.

Lines 416 – 418. It seems that the baseline was shifted at FKL for HCFC-142b in Fig.3. What caused it?

We assume that the change of the general direction of advection caused the shift in the baseline of HCFC-142b. However this doesn't have a large impact on the determination of regional emission values, as the baseline for each site is adjusted during the inversion and only the signal above the baseline is relevant.

Lines 428 – 429. "simulated a priori mole fractions reproduced the variability of the observations". Please provide the correlation (r2) value between the simulated and observed mole fractions here. In this way, it gives quantitative information on how much variability the prior simulation explained the observed variability

The coefficient of determination (R^2) and also the difference between simulated signals of the prior and the posterior simulation for the four sites is discussed in the following paragraph, showing that the inversion performance is comparable to other studies with similar inversion schemes (e.g. Stohl, et al., 2009; Keller, et al., 2012). In this paragraph, a reference to Table 2 is given, which contains all the R^2 values. To make this clearer, a reference to Table 2 has been added to the first paragraph mentioned by the referee.

Lines 495 – 497. Error reductions were expected after an inversion. This cannot demonstrate that you achieved satisfactory emission estimates.

We are aware that uncertainty reductions are not a proof of satisfactory posterior emission estimates. However, these error reductions are an indication that the inversion is working correctly which we tried to express with this sentence. Because we understand, that this sentence does not improve the understanding but is rather confusing, we deleted it in the revised manuscript.

Lines 510 – 520. The authors discussed the differences in RMSE, R2 and Taylor Skill Scores for inversion results obtained from the "global" and "local" approaches. But how different are the covariance parameters derived from both approaches? It would be useful to show and discuss those differences first.

We agree with the reviewer that the differences in the covariance matrices were not sufficiently discussed. The main difference, as mentioned in the methods section, was the scope of the temporal correlation length scale, τ_c . In the "global" method only a single correlation length scale was used for each site, whereas for the "local" approach a local correlation length scale was used the varied for each time and site and was based on a local window of 10 days. The variances on the diagonal of the data-mismatch covariance matrix were the same in both methods; at least for the first iteration of the RMS estimator. Since the iteration uses the posterior emissions for a new estimate of the RMS error, the final values may be different for the local and global method. We added this detail, which we had missed in the original version, in the revised manuscript (see also comment above on RMS

error). Having said this, one can first compare the total covariance by site that is contained in \mathbf{R} by calculating

$$\sigma_k = \sqrt{\frac{\mathbf{R}_k \cdot \mathbf{R}_k^T}{N_k}},$$

where \mathbf{R}_k is the block matrix belonging to all N_k observations/simulations of a single site.

In the case of the HFC-134a inversion, discussed at the mentioned location in the text, σ_k took values of 4.2, 7.3, 8.7 and 8.7 ppt for our BASE inversion (global τ_c) and the sites FKL, JFJ, MHD and CMN, respectively. For the S-ML sensitivity inversion (local τ_c) these values only differed slightly for the sites FKL and CMN, but were 8.3 and 9.0 for the sites JFJ and MHD, respectively. As a consequence less (more) weight was given to the observations from JFJ (MHD) in S-ML than in the BASE inversion. Especially for MHD one would expect that the posterior performance would be increased in the S-ML case compared to the BASE inversion. As described in the original text, this was not the case. The reason can be found in the distinctly different temporal pattern of the temporal correlation length scale. Panel d) of the figure below shows the differences between the empirical auto correlation function for a running window width of 10 days and the fitted auto correlation function with a constant (global) correlation length scale for the site MHD. MHD infrequently receives pollution events from the European continent. These episodes are usually characterized by relatively large model residuals (panel a). Also the auto correlation of the residuals during these periods is enhanced (panel c). The global estimate of τ_c then leads to an underestimation of auto-correlation during these periods (indicated by positive values in panel d). Finally, this means that in the BASE inversion more weight (smaller auto correlation, and, hence, smaller covariance) is given to the observations from MHD during the pollution events as compared to the sensitivity inversion with local τ_c . In turn, the posterior adjustments for MHD have a larger impact for the BASE inversion and performance improves more than in the S-ML case.

The above discussion for our example compound HFC-134a was added to the revised manuscript and the plot added to the supplementary material.

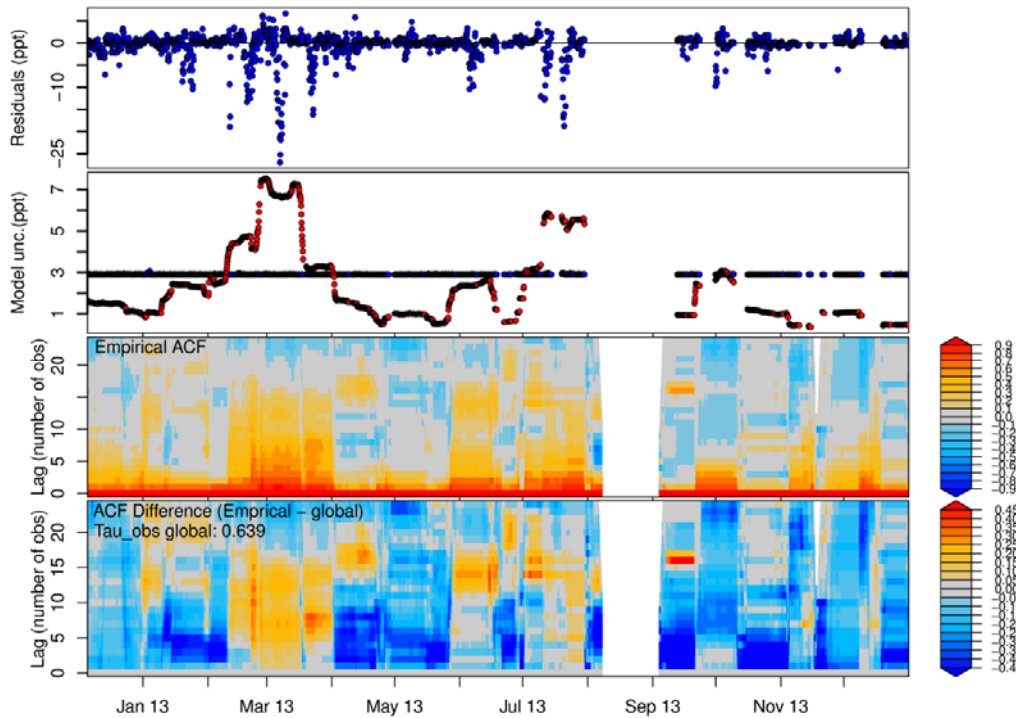


Figure 1: Time series of a) “prior” model residuals, b) data-mismatch uncertainty, blue symbols σ_0 and running RMS, red symbols, c) empirical auto correlation function based on 10 day moving window, d) difference between empirical ACF and fitted auto correlation with constant (global) correlation length scale. All given for the site MHD and for HFC-134a.

Lines 521 – 529. The results show the S-MS setup improves RMSE, R2 and Taylor Scores. This is, to some extent, expected and pre-defined, because the S-MS scenario did not include correlated errors in the R matrix whereas the other two scenarios (base and S-ML) did. To really test and understand the advantages or disadvantages of having correlated errors in the prior fluxes or atmospheric data, it is better to isolate the problem. In another word, it is better to have a scenario that has a same R matrix as the S-MS scenario and a same B matrix as the base scenario.

We thank the reviewer for this comment and addressed his concern in the discussion. However, we could not change our previous setup in the limited time available for the revision of this paper but will address this issue in future analysis.

Lines 530 – 540. Results seem to indicate emissions derived from the S-MS may be more biased toward the prior.

This seems to be unlikely, since the total data-mismatch covariance for the S-MS was the lowest of all three sensitivity inversions and, hence, the posterior simulations should be closest to the observations in this case, which agrees with the performance analysis in the manuscript. The reason for relatively large posterior emissions (closer to the prior) more likely is due some unrealistic adjustments of individual grid cells that are not well constrained due to the absence of off-diagonal elements in the covariance matrices.

Lines 605 – 607. It is a little dangerous to conclude a trend from two different studies given unknown differences from different methodologies and different atmospheric data.

We completely agree with the referee, that a trend cannot be deduced from 2 data points. To avoid making readers think that we indicate a trend based on our findings, we changed the sentence using

a less “impactful” wording to still show the differences in these two studies and what this “could” possibly mean.

Line 628. “account for 39.7%”. Should this be ~50% (0.53 / 1.0)?

Right, the wrong numbers have mistakenly been used to calculate those percentages. We updated the values in the revised manuscript.

Lines 628 – 632. “...a reduction by a factor of 2”. Again, it is not a good idea to conclude a quantitative trend from two different studies.

It was not our intention to show a quantitative trend in actual emissions here but to show the difference of the estimated values between our study and the one by Keller et al. (2012). To further avoid readers to think about a proposed trend, the wording was changed in the revised manuscript.

Lines 644 – 646. Inaccurate description on the HFC usage. HFC-125 is often used with HFC-143a in commercial refrigeration or with HFC-32 as a refrigerant blend in residential air conditioning.

The description was adjusted in the revised manuscript.

Lines 666 – 668. Citation of Brunner et al. [2016] here and thereafter. It is better not to cite an article that is not publicly available.

Meanwhile the article has been published and changes made to the discussion paper do not affect our manuscript. We now cite the finally published article and changed citations and the bibliography in the revised manuscript accordingly.

Brunner, D., T. Arnold, S. Henne, A. J. Manning, R. L. Thompson, M. Maione, S. O'Doherty, and S. Reimann (2016). Comparison of four inverse modelling systems applied to the estimation of HFC-125, HFC-134a and SF6 emissions over Europe, *Atmos. Chem. Phys.*, 17(17), 10651-10649, doi: 10.5194/acp-17-10651-2017.

Line 695. HFC-143a is mainly used in commercial refrigeration, not in air conditioners.

The description was changed in the revised manuscript.

Lines 717 – 719. The incomplete reporting to the UNFCCC does not seem to be an appropriate explanation for much larger reported emissions than atmosphere-derived emissions.

The sentence has been reformulated to avoid the impression, that incomplete reporting to the UNFCCC is the explanation for much larger reported emissions than the ones derived from atmospheric measurements. The main focus is on the issue, that emissions are reported in the country of manufacturing, although most of the emissions may occur during the usage of the product, which does not necessarily have to be the same country.

Lines 727 – 733. The authors compared their emission estimates with a few previous top-down emission estimates. Although the authors noted the differences, they did not provide an explanation why they are different.

Implicitly it is assumed, that any differences in top-down emission estimates originate in emission changes. However, we are aware that with the current network of three routine monitoring sites in Europe, complemented with the observations at FKL, it is difficult to determine smaller scale emission patterns. It was also shown by Brunner et al. (2017) that the use of different inversion systems can lead to significantly different results on the country scale, which makes the comparison of different European studies difficult, as long as there is no denser measurement network available.

Lines 806 – 808. Why including FKL increased the estimation of the baseline mixing ratios at JFJ and CMN?

FKL is sensitive to emissions from Italy, given the appropriate flow conditions. As a consequence, including measurements in FKL will impact the estimated regional emissions from Italy. A change in emissions from Italy, which do influence the simulations at JFJ and especially CMN, can thus have an influence on the baseline of these stations as well. Hence, the occurrence of a long-distance effect of the measurements in FKL on the whole inversion system can be understood qualitatively. However, we cannot assess if the magnitude of the effect is correct.