**Reply to 2nd review of Anonymous Referee #1**

In the following we repeat the comments of the referee in italics (and black) and add our replies in blue and regular fonts.

*I appreciate the authors taking the time to address my previous comments. I had, however, hoped that the authors would have been a bit more creative in addressing some of my comments, which would not have required additional analysis. Here, I provide small suggestions on further addressing some comments that I think deserve more attention.*

We thank anonymous referee #1 for his/her 2nd review and reply to the specific 3 points below.

*For a new system aimed at emission verification, it seems strange to not in any way mention trends in the manuscript. Emission targets are generally set in terms of relative emission reductions, rather than absolute emissions, so it seems that that would be where verification efforts are focused. I would like the authors to reconsider adding a few sentences on the practicalities of moving to a longer timeseries, and on possible implications on the results. This is partly done in the new L765-774 paragraph, but this is for some reason focused on Switzerland, whereas the presented study is for Europe. For a reader who is not completely up-to-date on existing verification efforts, it is hard to otherwise place this study in context.*

We had added in the previous revised version the requested information about the practicalities to apply FLEXVAR to longer time series as follows:

"FLEXVAR inversions with the configuration presented in this paper could be performed for the years 2002 to 2021, the period for which meteorological fields from the COSMO-7 model at 7 km × 7 km resolution are available. For analysis periods after 2021, the use of different high-resolution meteorological input fields could be considered, such as e.g., the operational analysis data from the ECMWF IFS model at high resolution (0.1° x 0.1°) or the operational MeteoSwiss COSMO-1 analysis at horizontal resolution of 1 km × 1 km. COSMO-1, however, is limited to the larger Alpine area, but can be nested into FLEXPART-IFS. A FLEXPART-COSMO modelling system using COSMO-1 has already been developed by Empa, including a modification of the turbulence parameterization [Katharopoulos et al., 2022], which is required owing to the very high resolution of 1 km × 1 km."

Only the COSMO-1 data are limited to the larger Alpine area, while the ECMWF IFS model at high resolution (0.1° x 0.1°) are available even globally. Therefore, potential future applications of FLEXVAR for other time periods are certainly not limited to Switzerland. With the existing COSMO-7 data, FLEXVAR inversions could be directly performed for the years 2002 to 2021 (on the COSMO-7 domain as presented in the current paper for 2018).

As stated in our previous reply, however, the analysis of emission trends is beyond the scope of the present paper. We emphasize again, that in particular the analysis of the uncertainties in derived emission trends is rather challenging.

*L672-674: This is a very implicit way of stating that the different model systems use different spatial correlations. I suggest making it more explicit, since it seems important.*

*In addition, an easy way to check the influence of the different spatial correlation lengths on this comparison would be to have a look at the same observational error statistics for inversion INV-E1-O1-S3.2. Have the authors done this? Is FLEXVAR-200km still better performing than FLExKF?*

*I understand that the difference in performance is relatively small, but the authors themselves raise the point, so I would like them to make these small efforts to find more clearly where the difference comes from.*

As stated in our previous reply we had further investigated the impact of the covariance settings on the achieved correlation, including the analysis of inversion INV-E1-O1-S3.2 (increasing the correlation length from 100 km to 200 km) and INV-E1-O1-S4.2 (increasing the prior uncertainty from 100% to 200%). However, these sensitivity inversions have been performed using emission data set E1 and observation data set O1 and should therefore be compared with the corresponding FLEXVAR and FLExKF inversions (i.e., FLEXVAR INV-E1-O1 and FLExKF E1-O1).

Following the request of the reviewer, we have added now a short summary of this additional analysis also the revised manuscript:

"E.g., the FLEXVAR inversion INV-E3-O2 used for the model comparison applies a smaller spatial correlation length ($L_{\mathrm{corr}}$ = 100 km) compared to FLExKF ($L_{\mathrm{corr}}$ = 200 km). Comparison of FLEXVAR inversions INV-E1-O1 and INV-E1-O1-S3.2 shows that increasing the correlation length from 100 km to 200 km is indeed slightly deteriorating the statistical performance (mean correlation coefficient and mean rms difference), but nevertheless FLEXVAR (INV-E1-O1-S3.2) still performs slightly better compared to FLExKF (inversion FLExKF E1-O1). On the other hand, FLExKF applies a higher prior uncertainty than FLEXVAR (Table S4) in the model comparison discussed in the paper. For FLEXVAR, increasing the prior uncertainty from 100% to 200% (INV-E1-O1-S4.2 vs. INV-E1-O1), is slightly improving the statistical performance, i.e., partly compensating the effect of a larger correlation length (results not shown)."

*The authors go through considerable effort to compare different inverse modeling systems. I understand that it is hard to compare explicitly and quantitatively the computational costs of the different systems. However, I would like to see some small discussion of the practical (dis)advantages of the different inverse systems (as in the author's reply to the final point of my first review), since this is an important part of choosing which inverse system to use.*

As stated in our previous reply, the FLEXPART-COSMO back trajectories were computed on a different computing platform than the FLEXVAR and TM5-4DVAR inversions. Therefore, we cannot compare quantitatively the required computational resources. Moreover, such a comparison strongly depends on the specific application, e.g., number of stations used (e.g., computational costs for FLEXPART back trajectories scale directly with number of stations, while the number of stations has only a minor impact on the costs of TM5-4DVAR inversions) and number of FLEXVAR inversions to be performed (since FLEXPART-COSMO back trajectories and TM5 baselines need to be computed only once). Apart from

this, it is not clear, what exactly the reviewer means by "practical (dis)advantages of the different inverse systems"