Referee #1

• Figure 4: with ranges, I meant range of XCH4 values within a month and the domain. Could be standard deviations or quantiles. I suppose you can possibly add for both observations and modelled values.

Thanks for the clarification. We have added the standard deviations to both the observed and modelled mixing ratios in Fig. 4.

• Following the comments by the reviewer 3, please shortly add reasons for the choice of the FLEXPART version in Introduction or Method section and your plan to implement the code in FLEXPARTv11 in Conclusion.

We implemented this methodology in FLEXPART v10.4 simply because v11 was not ready (also with no outlook of when it would be ready) when we started this work. (Note that the FLEXPART-v11 paper was only published 30 October 2024) . Since the methodology is general (it could be implemented in other LPDMs) we do not want to focus too much on FLEXPART in the paper. We, however, include a brief description of how the methodology can be used with FLEXPART-v11 in the Supplement.

Reply to Reviewer 2

We thank the reviewer for these comments and reply to them point-by-point below. Please note our responses are indicated in blue font.

I thank authors for revising the manuscript by taking into account of my comments. But, I noticed that the authors' responses to my comments differ in format from those addressed to Reviewers 1 and 2. While their replies are structured as point-by-point responses with reviewer comments quoted in blue, mine are listed by number without the original comments included. This formatting made it difficult to track and cross-reference of my original concerns. For future revisions, I suggest maintaining a consistent structure that includes the reviewer's comments, as this facilitates the process for both reviewers and the editor.

The different format was simply because your review comments were submitted in a PDF format that was not recognizable as text. This meant that we could not simply copy the text from the PDF into a new document and include our replies. This time round a work around was found by scanning the PDF and converting it to text. Despite the different format in the previous response, please note that we did take care to respond to all comments and our responses were numbered according to the comment they addressed.

Even though the manuscript improved significantly from the previous version, not all of the concerns I raised during the initial review have been fully addressed. While the revised methods section is overall clearer, several aspects of the FLEXPART implementation remain vague. I appreciate that the authors corrected and rewrote Equations (6) to (11), as the original formulations contained significant errors. However, it is still unclear whether these errors were confined to the manuscript or if they were also present in the model implementation. I wonder whether the discrepancies arose from a miscommunication between the code developer and the person who wrote the article. If that is the case, it raises the concern that the same misinterpretation may have affected the code implementation itself. I hope this is not the case. Nonetheless, since no new simulations have been provided in this revision, it remains uncertain whether the code implementation is consistent with the corrected equations.

The same person (Rona Thompson) wrote the code and the manuscript, so there is no miscommunication. In the first version of the manuscript, we provided simplified versions of the equations, which we thought it would make it easier for the readers to follow, e.g., the transmission function was omitted in Eq. 6 as this term is does not significantly differ from one. The correct equations were implemented in the code and are consistent with the revised version of the manuscript.

A key unresolved concern in the methods section is with the description of the affine transformation applied to particle coordinates. In my initial review, I requested a clear mathematical description of this transformation. In the revised manuscript, the authors describe the process only qualitatively and refer readers to the Fortran source code in a public repository. This is not an adequate substitute for proper documentation in the manuscript or supplementary material. A clear and explicit mathematical description of the transformation is essential for transparency and reproducibility, particularly because this step directly affects the spatial accuracy of the satellite pixel representation.

We did not previously include the explicit calculations for the affine transformation because this is an algorithm and not just a few equations. We now include the code excerpt for this with explanations in the Supplement.

Regarding the results and discussion, the authors state that they performed sensitivity tests but chose not to include them, as the paper's emphasis is on methodology. However, if the primary contribution of the manuscript is methodological, it becomes even more important to describe all technical steps, including the transformation, particle release setup, and super-observations procedure, rigorously. Providing this level of detail is critical to ensure that the approach can be reproduced and built upon by other researchers.

The main focus on the paper is the methodology for modeling column observations, such as those from satellites, in an efficient way using a Lagrangian particle dispersion model. The case study is given to demonstrate the application of the methodology but is not intended to be the focus of the manuscript. We have, however, included more technical details in the paper, specifically, we now include the excerpt of code for the affine transformation in the Supplement.

To be clear, my goal is not to obstruct publication. I recognize that the manuscript presents valuable developments and has strong potential. However, given that the stated focus is methodological, I believe the technical content should be presented with greater clarity and rigor to meet the standards of transparency and reproducibility expected in the field.

We have now added further explanations on the methodology in the Supplement. We chose to put it there as we feel that having these extra explanations in the main part of the manuscript would make the paper too long and distract from the main focus.

Further comments are below based on the response of the authors:

Comment 1:

I recommend adding a brief statement to the main text, perhaps in the Introduction, as there is no dedicated discussion section, to clarify how the developments presented in this study relate to FLEXPART v11. Specifically, while FLEXPART v11 introduces the option for custom particle initialization, it does not currently support simulation of total column averages from satellite retrievals. In contrast, the work presented here, based on FLEXPART v10.4, provides an operational method for calculating total column source-receptor relationships (SRRs) from satellite observations.

It would also be valuable to note that the developments described in this paper are planned for future integration into FLEXPART vII. Including this clarification will help readers and potential users select the appropriate FLEXPART version for their applications and will underscore the significance of the methodological contribution presented here.

We have now included a short section in the Supplement about how these developments could be used in conjunction with FLEXPART-v11. In FLEXPART-v11, the particle initial positions can be specified by reading in a NetCDF file (this was not an option with previous FLEXPART versions). The developments described in this paper can be used to create such a NetCDF file specifying the particle positions for the releases. This would involve the steps from the pre-processor developed for FLEXPARTv10.4 and FLEXINVERT, which reads the satellite observations (L2 data) and performs optional averaging of retrievals, and then use the code "releaseparticles_satellite.f90" written for FLEXPART-v10.4, which performs the affine transformation and calculates the release positions for all particles.

We think this description does not belong in the main part of the paper because the focus of the paper is not about the development of FLEXPART. Instead, the focus is on the methodology of calculating source receptor relationships for satellite observations, which is general in the sense that it could be implemented in any Lagrangian transport model.

Comment 3:

In my initial review I raised concerns regarding the formulation of Equation (6), which originally appeared in the manuscript as:

$$H_{ijk}^{col}\sum_{n=1}^{N} a_n w_n \frac{t}{m_n \rho_{ijk}} \sum_{p=1}^{J_{i,jk,n}} m_p$$

At that time, I pointed out that the variable t should represent the residence time of particles within the surface layer as defined in Seibert and Frank (2004, Eq. 8) and Wu et al. (2018, Eq. 4), rather than a fixed sampling duration. I now see that the authors have updated Equation (6) as follows:

$$H_{in} = \frac{1}{P_n}\sum_{p=1} \frac{l_{pin}\Delta t_{pin}}{\rho_i}$$

While I appreciate the authors' revision of Equation (6) and subsequent equations and their explanation of the transmission function l_pin, I suggest improving the clarity of the description. The current phrasing : l_pin "represents the fraction of the mass remaining in the particle" could be misinterpreted by readers, particularly those less familiar with Lagrangian particle dispersion models operating in backward mode. In this context, particles are tracers and do not carry real mass. Rather, the transmission function serves as a scaling factor to account for atmospheric processes (e.g., chemical lasses), modifying each particle's residence time contribution to the SRR.

In FLEXPART particles are assigned an initial (arbitrary) mass, which can change e.g. due to chemical loss, and is the implementation of the transmission function.

Comment 5:

In my initial review, I suggested the authors to provide the explicit form of the affine transformation applied to particle coordinates (i.e., the mathematical equations used). However, the revised manuscript still describes the method only in qualitative terms, and the authors refer readers to the subroutine releaseparticles_satellite.190 in the public Fortran code base to understand the implementation.

While I appreciate that the code is openly available, I would like to reiterate that it is neither feasible nor expected for a voluntary reviewer to examine and interpret a large Fortran code base to assess a methodological detail that could, and should, be documented clearly in the manuscript or its supplementary material. Given that the affine transformation plays a critical role in accurately representing satellite pixel geometry, I strongly recommend that the authors include the mathematical formulation of this transformation for transparency and reproducibility.

We now include excerpts of the code for the affine transformation (with explanation) in the Supplement.

Comment 6:

It would significantly improve the manuscript if the authors provided the explicit equations used to perform the super-observation (super-orbiting) processing of TROPOMI data. Currently, this procedure is only described qualitatively, which limits the reproducibility of the method and may confuse readers unfamiliar with the approach.

The code that performs the calculation of the optimal grid for the retrievals (note that it is not specific to TROPOMI) and the averaging of the retrievals to this grid is over 400 lines. We have never seen a paper in Atmos. Chem. Phys. that includes entire algorithms or code. We describe the principle of how the algorithm calculates the optimal grid in Section 2.2, and we give the equations for the averaging of the retrievals to super-observations in Equations 12 to 14. The code is open source and anyone who is interested in how the algorithm is implemented can download the code from the Git repository.

Comment 8:

Indeed the reduced-chi-square criterion can be ambiguous and alone is not a sufficient criterion for assessing the appropriateness of the uncertainties. However, as stated by Chevallier et al 2007, in an idealized system (OSSE experiment), the cost function J(x) converges toward N (number of observations). Hopefully, the author made this test with TROPOMI before attempting to run real case inversion.

We did run synthetic data experiments to check the validity of the code and settings. For the TROPOMI inversions the reduced chi-square value is 1.08 and is very close to the “ideal” value of 1.0.

The reduced chi-square value of 4.86 for the ground-based observations suggests that the observation or model errors may be underestimated, or that there is unaccounted model-data mismatch. It would be helpful if the authors clarified how the observation uncertainty was defined for the in situ network. particularly whether representativeness and model transport errors were included, and at what magnitude.

We agree that the reduced-chi square value is high for the ground-based observation inversion. The observation space uncertainties include the measurement uncertainty, a proxy for the transport uncertainty was taken as the standard deviation of the measurement in one hour for continuous measurements and a set value of 5 ppbv for flask measurements. For continuous observations, if the standard deviation was < 5ppb the minimum estimate of 5 ppbv for the transport uncertainty was used. In addition, we calculated an estimate for the background uncertainty.

In this revision, we re-ran the inversion increasing the observation space uncertainties by 5 ppbv with respect to their former values. The reduced-chi square value decreased to 2.16. Also, the median of the uncertainty is now 11.9 ppbv, while the median of the absolute posterior model – observation error is 10.6 ppbv, indicating a reasonable estimation of the observation space uncertainty.

I made this comment because, the authors mention that *'the satellite observations are more uncertain compared to the ground-based observations (the median uncertainty for XCH4 in our study was 16 ppb compared to 8 ppb for the ground-based observations) and the model-observation errors are weighted by the inverse square of the uncertainty'.* It could be possible that the uncertainties of ground based observations in combination with the model errors are larger that 8 ppb, hence the reduced chi-square value is super high (4.86).

We have revised our observation space uncertainties upwards resulting now in a median of 12 ppb. The posterior fluxes from the inversion with this larger observation space uncertainty are very similar to those from the original inversion. The mean posterior source over the domain is now 34.6 compared to the previous value of 33.9 Tg/y. The posterior uncertainties are also very similar to the original inversion. We have updated figures 5, 6 and 8 and figure S3b and the text for the new inversion results and uncertainty estimates.

The ground-based observation space uncertainties still remain lower than those for the satellite observations, 12 ppb versus 16 ppb for the median uncertainty, respectively. Moreover, in our discussion we give two principal reasons for why the inversions using TROPOMI have a lower uncertainty reduction and why the posterior fluxes remain very close to the prior ones. The first is due to the higher uncertainty in the TROPOMI observations compared to the ground-based ones (which still holds after the upward revised uncertainties in the ground-based observation inversion). The second, and even more importantly, is the fact that each single retrieval (or TROPOMI observation) is much less sensitive to the fluxes in the domain leading to smaller model-observation differences. Since the cost function depends on the square of the model-observation differences, a few large differences have more influence on the cost than many small ones.

Comment 9:

The authors note that they performed several sensitivity tests but chose not to include them in the manuscript, as the focus of the paper is "primarily on the methodology of using satellite observations in an inversion framework based on a Lagrangian transport model." However, if the manuscript is intended to be primarily methodological, then the methodological components should be described in greater detail.

We describe the methodology for the calculation of the source receptor relationships for satellite observations, and the methodology for the averaging of the retrievals, in detail. Furthermore, we discuss the key factors that the methodology is sensitive to. Specifically, we discuss the sensitivity to the initial mixing ratios, which are used to calculate the background column mixing ratios, and is the most important factor for determining the results. We present results using different initial mixing ratios and discuss how this sensitivity can be resolved by optimizing scalars of the initial mixing ratios (see section 3.2.2). We also discuss the choice of uncertainty for the scalars on the initial mixing ratios, and how these uncertainties were increased based on the results of sensitivity tests (see L226). We reiterate, that the focus of the paper is the methodology for the calculation of the source receptor relationships for satellite observations, and that the case study is to demonstrate the application of the methodology but is not intended to be the focus itself.

Given the concise length and the technical nature of the work, specifically, the adaptation of FLEXPART vl0.4 for use with satellite-based retrievals, the manuscript should contain the formal presentation of the algorithmic steps, including the mathematical formulation of transformations and other implementation choices, which are currently either omitted or described only qualitatively.

We include the key equations for the methodology of calculating source-receptor relationships using an LPDM in the manuscript (Eq. 1 to 11), and the key equations for the averaging of the retrievals (Eq 12 to 14). We have also in this revision included the algorithm for the affine transformation in the Supplement. It is not normal practice in Atmos. Chem. Phys. to include entire algorithms and code in papers. Specifically, the algorithm for the optimal averaging for super observations is over 400 lines long, and those readers who are interested in exactly how this is implemented can access the code from the open Git repository (details are in the section “Data Availability”) and they are also welcome to contact the corresponding author (Rona Thompson).