EGUSPHERE–2026–3458

Towards robust inversions: Parametric weighting of high-resolution TROPOMI observations for global methane emission estimates with TMVar (TM5-MP/4DVAR)

Author comments to RC2
https://doi.org/10.5194/egusphere-2026-3458-RC2

We would like to thank both referees for their constructive comments and corrections that helped to improve our manuscript. Hereby, we provide answers and comments on each of the points raised by the referees. Each comment of the reviewers is numbered and quoted in a gray box, followed by our replies and quoted text in a yellow box.

Anonymous Referee #2

(1) Missing distinction from super-observations
I am surprised that the authors do not relate their method to super-observations (Chen et al., 2023). Although super-observations differ in some respects from the parametric regularization presented here, they are also a way of reducing the weight of densely spaced observations. However, the authors only mention them in Sect. 2.3 as a possible extension, not as an alternative approach and only compare their method with inversions using a constant factor.

This issue has been raised by both referees #1 and #2. The same modifications apply. We included the following sentences in Section 1: Introduction.

> Super-observations can provide an alternative by aggregating individual observations and accounting for their error correlations, thereby partly addressing local sampling density (Miyazaki et al., 2012; Chen et al., 2023; Nüß et al., 2025; Rijsdijk et al., 2025). This is limited to the observations contributing to each super-observation, and aggregation inevitably reduces the spatial and temporal detail retained from the original observation.

(2) Missing application of statistical diagnostics
There are objective criteria for defining the error covariance matrices, as well as statistical tests, for example the chi-squared test to check whether the assigned uncertainties are consistent with the observed residuals. Unfortunately, this is completely missing. In my opinion, an additional scaling of R in all inversions to achieve statistical consistency would be sensible and would provide the fairest comparison between the inversions. Also, having the posterior uncertainty available (through an alternative method) would be desirable.

We agree with the referee that applying a chi-squared test to find a global regularization factor could improve inversion results. However, as we mentioned in the abstract, the scope of the current work is to address the fact that satellites provide large numbers of heterogeneously distributed observations, causing regions with fewer observations to have lower impact over the total observational cost. Therefore, the outcome of our study is a regularization factor per observation, instead of a global one for all observations.

Due to the number of observations, applying individual chi-squared tests to each of them is computationally unfeasible. A single inversion in the setup presented here takes approximately 3 months to converge on our current HPC system, and an iterative chi-squared scaling, or an ensemble-based posterior uncertainty estimate, would require several additional runs per

inversion. Nonetheless, we acknowledge that our choice leaves the prescribed uncertainties unverified against the residuals.

(3) Temporal dimension of the weighting
The abstract (lines 6–7) and the introduction (e.g. line 50) state that the weighting factors are based on the spatial and temporal distribution of the observations. As far as I understand, however, the kernel in Eq. (5) depends only on the spatial distances. Time enters only as a hard cutoff that selects the neighboring bands. As the authors show, extending the window has a negligible effect. I would suggest to rephrase the description of the method in the abstract and introduction accordingly.

The understanding of the referee is correct. Accordingly, we adapted the phrasing when we considered it necessary.

Abstract
We present a parametric regularization method that computes observation-specific weighting factors based on the spatial distribution of satellite observations within a defined temporal window.

Section 1. Introduction
In this study, we introduce a parametric method that computes observation-specific regularization factors from the local spatial density of satellite measurements.

(4) Framing of the problem and objectives
In the introduction, the authors present the problem solely as unevenly distributed observations. However, it is not until the methods section where they write that the actual problem is the neglected error correlations. Dense or unevenly distributed observations are not a problem in and of themselves. After all, the correlated errors would theoretically be accounted for in the error covariance matrix. The authors also state a "more homogeneous spatial distribution of observational weights" (lines 380–381) as their "primary objective" and measure success by the decrease in CV. Compared to a simple, constant approach, a lower CV may represent an improvement, but when errors are correctly described, a densely observed region does indeed contain more information than a sparsely observed one. Just not proportionally to the number of observations. Homogeneous weights are therefore not a goal in and of themselves. The goal would be to approximate the correlated error covariance as closely as possible. I also do not entirely agree with the characterization of "vanishing gradients" as a problem. In the Bayesian framework, these are not a malfunction. Sparsely observed regions are simply less constrained and remain closer to the prior.

We agree that heterogeneous observation density is not inherently a problem, and that the underlying issue is the neglect of observational error correlations, which can lead to an overestimation of the information content of densely sampled regions when a diagonal observational error covariance matrix is used. We have revised the Introduction to clarify this distinction as follows.

This approach homogenizes the effective influence of densely and sparsely sampled regions while remaining computationally efficient, partially addressing the limitation of neglecting the correlated errors when using a diagonal observational error covariance matrix.

We also agree that a homogeneous spatial distribution of observational weights is not an objective

in itself, and that a reduction in the coefficient of variation (CV) does not demonstrate that the prescribed weights more accurately represent the true observational error covariance. We use the CV as a diagnostic of how the proposed method redistributes observational weights relative to the constant regularization approach, rather than as a measure of statistical optimality. We clarify this distinction in the manuscript.

> The purpose of the adaptive regularization is to reduce the disproportionate influence that densely sampled regions may exert when observational error correlations are not explicitly represented. We use the coefficient of variation (CV) of grid-cell weights to quantify the redistribution produced by the adaptive regularization relative to constant regularization. For 2019, the CV decreases from 1.62 to 1.30, corresponding to an approximately 20% reduction in relative variation (Table 3). This indicates that the method reduces the spatial variability of the effective observational weights. Note that, in the constant regularization factor case, the magnitude of γ does not affect the calculated variation.

Regarding vanishing gradients, we agree that weak gradients in sparsely observed regions are not inherently problematic in a Bayesian inversion, as these regions should generally remain closer to the prior when observational constraints are weak. However, our concern is the potential for weakly constrained regions to adjust emissions in order to reconcile mismatches elsewhere in the domain. Because adjustments in these regions may incur a relatively small observational cost, they can act as "garbage collectors" for the inversion, as discussed in Section 3.3.2. This behavior is not inevitable, nor do we suggest that all sparsely observed regions should receive the same weight as densely observed regions. Rather, our method explores whether a density-based adjustment of the observational weights can mitigate this effect while retaining the computational efficiency of a diagonal covariance matrix.

> (5) Availability of the output
> Although the authors are not obliged to make the inversion results publicly available, doing so would be highly desirable to facilitate access for other scientists. It would be great, for instance, if the results could also be compared with regional inversions.

Because the setup is tailored to be a case study of the regularization scheme, we considered the results not suitable for a direct comparison with regional estimates. Nevertheless, they can be provided upon request, to ensure a proper interpretation of them.

Specific comments

> Page 2, line 46: The problem is not restricted to variational inversions, but affects all types of inversions.

We changed the sentence to:

> However, in both analytical and variational inversions...

> Page 9, line 237: Does the minimum observation error of 3 ppb apply before computing the weighted mean, i.e. to the individual observations, or afterwards to the weighted mean?

Clarified that the minimum is applied after calculating the weighted mean.

> Afterwards, a minimum measurement error of 3.0 ppb is imposed...

Page 14, lines 348–349: Two things are unclear to me here. First, is the transport error already part of R when R is inflated, or is the observation error in R inflated first and the transport error added afterwards? Second, it this statement is unclear to me: "an estimate of the spatial and temporal transport errors from the a priori runs". What is meant here, and how is the transport error estimated? I recommend clarifying this paragraph.

General clarification of the paragraph, now it reads as follows.

> For the observation error covariance matrix $\mathbf{R}$, the observation uncertainties are specified according to the observation type. For the NOAA surface observations, the uncertainty follows the procedure described in Bergamaschi et al. (2010), using a minimum standard deviation of 3.0 ppb to represent instrumental uncertainty, with additional contributions accounting for temporal and spatial representation errors. These contributions are estimated from model simulations, considering variability within 3-hourly intervals and the impact of local emissions. For the TROPOMI observations, the retrieval uncertainty provided with the data is used, with a minimum standard deviation of 3.0 ppb. No additional spatial or temporal representation errors are included for the satellite observations. Observation errors are assumed to be uncorrelated; thus, $\mathbf{R}$ is constructed as a diagonal matrix containing the resulting observation variances and subsequently inflated using the weighting matrix $\mathbf{\Gamma}$.

Page 22, line 486 – page 23, line 488: I do not understand the statement "neither the satellite nor the station observations are bias-corrected". To my knowledge, NOAA measurements are not bias-corrected anyway. Moreover, at least for S2/S3 there are no "competing constraints", since only satellite data are assimilated. These results only show that the solution that fits the columns does not fit the surface observations. I suggest clarifying these statements.

Generally rephrased the paragraph to clarify.

> Posterior simulations generally reduce normalized mismatches relative to the a priori across most stations, as expected, since NOAA measurements are assimilated in both the baseline and combined inversions (see Figure S1). Overall, simulations assimilating satellite observations show larger mismatches against NOAA flask measurements, with the discrepancy increasing when adaptive regularization is applied. Since the satellite retrievals are not bias-corrected, differences between the satellite and in-situ observations may contribute to these contrasting results. However, the larger NOAA mismatches do not necessarily indicate inconsistencies between the datasets, but rather show that a solution that improves agreement with satellite total columns does not necessarily improve agreement with near-surface methane concentrations. This may also reflect limitations in the representation of atmospheric transport, vertical mixing, or methane sources and sinks.

Page 24, line 507: From Fig. S8, I cannot see that the absolute mismatch is reduced in several regions when adaptive weighting is applied. A quantitative measure would be needed here.

We believe that Table S4 may already provide what the referee suggests.

Technical comments

Page 11, line 286: “observation a priori mixing ratios” → “retrieval a priori mixing ratios”?

Changed accordingly.

Additional changes

Motivated by the comments from both referees, the discussion of the limitations of the method has been improved in Section 2.3, which now reads as follows:

> The proposed approach has two potential limitations. First, the representativeness effects are accounted for only indirectly, through density-based inflation rather than explicit covariance modeling. Nevertheless, this simplification allows the method to scale efficiently to the large observational datasets typically used in 4DVar inversions. Second, due to the nature of a kernel density estimation, the method does not ensure locally monotonical increasing behavior of the cost function. In practice, for the here used dense TROPOMI data, this behavior has not been observed.

References

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., ..., and Dlugokencky, E. J. (2010). Inverse modeling of European CH_4 emissions 2001-2006. *Journal of Geophysical Research Atmospheres*, *115*(22), 22309. https://doi.org/10.1029/2010JD014180

Chen, Z., Jacob, D. J., Gautam, R., Omara, M., Stavins, R. N., ..., and Hancock, S. (2023). Satellite quantification of methane emissions and oil-gas methane intensities from individual countries in the Middle East and North Africa: implications for climate action. *Atmospheric Chemistry and Physics*, *23*(10), 5945–5967. https://doi.org/10.5194/ACP-23-5945-2023

Miyazaki, K., Eskes, H. J., and Sudo, K. (2012). Global NOx emission estimates derived from an assimilation of OMI tropospheric NO2 columns. *Atmospheric Chemistry and Physics*, *12*(5), 2263–2288. https://doi.org/10.5194/acp-12-2263-2012

Nüß, J. R., Daskalakis, N., Piwowarczyk, F. G., Gkouvousis, A., Schneising, O., ..., and Vrekoussis, M. (2025). Top-down CO emission estimates using TROPOMI CO data in the TM5-4DVAR (r1258) inverse modeling suit. *Geoscientific Model Development*, *18*(10), 2861–2890. https://doi.org/10.5194/gmd-18-2861-2025

Rijsdijk, P., Eskes, H., Dingemans, A., Boersma, K. F., Sekiya, T., ..., and Houweling, S. (2025). Quantifying uncertainties in satellite NO_2 super-observations for data assimilation and model evaluation. *Geoscientific Model Development*, *18*(2), 483–509. https://doi.org/10.5194/gmd-18-483-2025