

Review of "Towards robust inversions: Parametric weighting of high-resolution TROPOMI observations for global CH₄ emission estimates with TMVar (TM5-MP/4DVAR)"

The manuscript aims to mitigate the effects of the uneven spatial distribution of satellite methane observations on inversions. Since an error covariance matrix including off-diagonal covariances is often computationally infeasible, the diagonal elements have to be adjusted accordingly. To achieve this, the authors introduce a parametric regularization that depends on the spatial and temporal distribution of the observations (although in practice only the spatial distribution affects the regularization, while time only acts as a hard cutoff). The more observations are located near a given observation, and the closer they are, the lower its weight is (i.e. the larger its uncertainty in the inversion).

For global inversions assimilating satellite data (the TROPOMI WFMD product), both alone and combined with NOAA surface measurements, the authors show the effects of this regularization and interpret the results. The results are further evaluated against NOAA measurements and non-assimilated TCCON observations. In addition, the authors present sensitivity tests regarding the bandwidth of the regularization term and the time window of the observations considered.

The study is scientifically sound, well written and mostly clearly presented. In my opinion, it makes a valuable contribution towards more robust uncertainty descriptions in inversions with large numbers of densely spaced observations, which is important given the growing number of methane-observing satellites. I therefore recommend the manuscript for publication after minor revisions.

General comments

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.

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.

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.

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.

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.

Specific comments

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

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.

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.

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.

Technical comment

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