



EGUsphere, referee comment RC1
<https://doi.org/10.5194/egusphere-2022-719-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on egusphere-2022-719

Anonymous Referee #1

Referee comment on "Estimation of CH₄ emission based on an advanced 4D-LETKF assimilation system" by Jagat S. H. Bisht et al., EGUsphere,
<https://doi.org/10.5194/egusphere-2022-719-RC1>, 2022

General comments

Bisht et al. present in their manuscript a new data assimilation system based on the Local Ensemble Transform Kalman Filter (LETKF) method with atmospheric transport described by the atmospheric transport model, MIROC4. The system is applied to the estimation of surface fluxes of methane, using both a network of surface observations and GOSAT satellite retrievals. This study describes the method and tests it using Observing System Simulation Experiments (OSSEs) consisting of performing inversions with synthetic observations and for which the true fluxes are known. A number of sensitivity tests are presented to test the system.

On the whole the methodology is scientifically sound and based on previously published models and algorithms. However, in parts the manuscript is difficult to follow and the text unclear or ambiguous. In particular, I suggest improving the description of the methodology especially regarding the preparation and selection of the pseudo observations (see specific comments). In addition, the results and discussion section could be improved to make it easier to follow.

Based on this, I think the manuscript could become acceptable after minor revisions.

Specific comments

L10: "which is substantially" should be "which has substantially"

L23: suggest removing "absolute" before normalized, since if normalization is done the

value is always relative

L32: "much high" should be "much higher"

L35: "that have anthropogenic" should be "that has anthropogenic" (i.e., singular form)

L36, "the global CH₄ budget"

L37: suggest stating that the given range is for the total of all sources and not to put it in parentheses since it is quite important information

L38: suggest changing "remaining CH₄ emissions" to "main anthropogenic CH₄ emissions" since you list only anthropogenic ones and not all (e.g., the minor source from incomplete combustion of bio and fossil fuels is not mentioned)

L43: I think the reaction with Cl radicals actually mostly occurs in the troposphere where Cl is more abundant, see e.g.:

Wuebbles, D., Hayhoe, K. and Kotamarthi, R. (1999), Atmospheric Methane in the Global Environment. In: Atmospheric Methane: Sources, Sinks, and Role in Global Change. (Eds. M. Khalil), Springer-Verlag, New York, NY.

Allan, W., Struthers, H., and Lowe, D. C. (2007), Methane carbon isotope effects caused by atomic chlorine in the marine boundary layer: Global model results compared with Southern Hemisphere measurements, J. Geophys. Res., 112(D4), doi:10.1029/2006JD007369.

L63: The resolution of the control vector in EnKF methods is strongly limited by the ensemble size, if the number of ensemble members is much smaller than the rank of the error covariance matrix, then this method can give spurious results, see e.g.:

Houtekamer, P. L., & Zhang, F. (2016). Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation. Mon. Wea. Rev., 144(12), 4489–4532. <http://doi.org/10.1175/MWR-D-15-0440.1>

This limitation is not present in variational methods.

L70: replace "in the" with "for", i.e., "for carbon cycle research"

L73: Remove "The" before "assimilation" and change "window" to "windows" since you are not referring to one specific assimilation window, but to them generally.

L74: change "hour" to "hours"

L75: The time resolution of the control vector is not the only consideration in the assimilation time window, but the time frame over which the system behaves linearly, and in what time frame the observations respond to the control variables (in this case, determined by atmospheric transport)

L80: change "estimate" to "estimates"

Eq. 1: This equation should be re-written to express x^b and x_{mean}^b (column vectors) as matrices with the same dimensions as X^b (or alternatively for any i th member of the ensemble using the i th column of X^b)

L99: "and is derived" (missing "is") and change "with" to "using"

Figure 1: Please change "broken line" to "dotted line" as "broken" could also be confused with the dashed line used.

L140: Please spell-out RTPS

L144: Please specify that Eq. 8 is referring to RTPS and not RTPP.

L169: change "accelerates" to "accelerate"

L170: change "observation" to "observations"

L191: change to "applied to the"

L198: change to "initial perturbations are applied"

Section 3.3: I don't see where the locations of the surface network sites are given. It would be helpful to include a figure of these.

L205: change to "Errors in the estimated fluxes could arise..." I think the authors should also specify that this is in the context of the OSSE. In real-data inversions there are additional sources of potential error, e.g., modelled transport, inappropriate prior or observation uncertainties.

L205: Please clarify if "inflation used" the authors refer to the inflation of the covariance matrix (as described in section 2.1), and if so, is this not coupled to insufficient ensemble size since the inflation is to account for an under dispersive ensemble?

L210-212: I'm not sure what the authors mean by the following:

"We have estimated the CH₄ flux for each grid by choosing the observation that influence the grid point using optimal cutoff radius (horizontal covariance localization) of 2200 km and vertical covariance localization of 0.3 in the natural logarithmic pressure (ln P) coordinate." Could the authors please explain in more detail how observations were selected for assimilation?

In addition could the authors please explain:

"The optimized value of horizontal and vertical localizations..."

The localizations of which variables?

L229-234: I suggest removing the discussion of the assimilation window here and adding the new information to where this is discussed in section 2 (note, the assimilation window is discussed in section 2 (not 2.1)).

L242: XCH4 is not weighted by the prior and averaging kernel, but rather it is a weighting of the prior and the modelled mixing ratios, where the weighting is given by the averaging kernel.

L247-251: Similar to my comment above, I think the selection of observations needs further explanation.

L258: Please change "It could be noticed that the..." to e.g. "Noteworthy is that the..."

L259: change to "15% larger error"

L315: change "discussed" to "discuss" and add "of" before "GOSAT"

Section 4.2, L315-329: I find these paragraphs quite confusing. If I understand well, these paragraphs should introduce the sensitivity tests carried-out in this section? If so, please start with the description of these tests, and simply state if the same set-up was used (or not) as Experiment 1. Why were the experiments for assimilation window and ensemble size performed on the satellite observation dataset and not on the "dense surface observation" dataset? Would the results, e.g., for assimilation window, change for surface observations compared to satellite ones?

L359: By "the larger coverage of CH4 observations" in the 8 day assimilation window do the authors mean the greater sensitivity or "footprint" of the observations through the longer computation of atmospheric transport? This should be made clearer.

L375: I think by "relatively diluted flux signal" the authors mean the weaker constraint on surface fluxes provided by satellite observations or the weaker connection of the satellite signal to surface fluxes. I think the term "diluted" is a bit vague.

Section 4.2.3: I think the chi-square test needs a bit more explanation. For instance, which normal variables are being summed in this test? It would be helpful to write the equation. Also, if I understand correctly what is being tested here, would a value greater or lower than one possibly be also due to an under estimation of the observation error covariance?

L387: "Our results suggest that, background error covariance matrix is highly underestimated in both RTPS and FM covariance inflation methods (Fig. 7b) and indicates much confidence in the model" – please explain why this gives the authors "much

confidence”

L395-407: This section is difficult to follow. To start with, by “the flux estimation is extremely sensitive...” do the authors mean that the analysis fluxes are sensitive to the prior uncertainty, and by “provide larger prior uncertainty” do the authors mean to generate the prior error covariance matrix, or do they mean the perturbation to generate the prior fluxes? And in L402, by “the flux estimated error” do the authors mean the error between the analysis and true fluxes, and that this error would be larger when the inflation parameter is calculated grid-wise compared to region-wise?

L406: The authors mention that machine learning could be used to determine the spread of the initial ensemble. I think this needs to be explained, i.e., how could machine learning help?