Anonymous Referee #2
Received and published: 5 December 2019

We wish to thank the referee for his/her helpful comments. The full reviews are copied hereafter and our responses are inserted. The comments of the reviewer are in normal black and our answers in bold.

The paper describes the variational data assimilation version of the CHIMERE,PYVAR-CHIMERE, which is capable of inversions of reactive gases. As a demonstration to the newly developed code, the inversion of CO and NO2 is shown for two different days in late winter/early spring 2015. The papers topic is of good relevance for GMD and contribute to a documented open source regional data assimilation system for reactive chemistry. Although the paper is generally well written, major changes are requested before publishing the manuscript in GMD.

These major changes are:
- The quality of figures and formulas is unacceptable. Arrows should be larger/thicker (Fig. 1 and 2), annotations in Figs. 5, 6, 7, and 8 are too small, separation of subplots and 7 should be clearer.
The quality of the all the figures have been improved.

- The description of the inversion is unsatisfactory. The cost function and its gradient should explicitly show the model operator M, which is currently included in the state vector x.
We do not agree with this statement, the model operator is not included in the state vector x.

- Further, it is unclear how the emissions are corrected. How can negative emissions be avoided?
We now answered to this remark in "Section 4.2.2. Covariance matrices B and R": "With such a set-up, in theory, we could obtained negative posterior emissions since the inversion system does not impose a constraint of positivity in the results. Nevertheless, even 100% of uncertainty lead to a prior distribution mostly (>80%) on the positive side. The assimilation of data showing an increase above the background (at the edges of the domain; not shown) further drive the inversion towards positive emissions for both CO and NO_x inversions. In practice, our inversion does not lead to negative posterior emissions (Figure 7b). Spatial and temporal correlations in B would further limit the probability to get negative emissions locally by smoothing the posterior emissions at a spatial scale at which the "aggregated" prior uncertainty is smaller than 100%. However, a positivity constraint should be implemented in future versions of the system."

- Are the emissions optimized for each time step or for the whole assimilation window?
The user can chose the time resolution at which the emissions are optimized. In our illustrations, we now present inversions at 7-day and at 1-day resolutions.

-Are the emissions constant for the simulation time or does the inversion result in correction factors for the emissions?
Indeed, the emissions are inverted, i.e, the inversion results in correction factors for the emissions at the specified time and spatial scales.

-Then, the special treatment of 4D-var for emission factor optimization should be shown, e. g. how the positive definiteness of the correction factors is ensured. The manuscript must be more precise in this context.
We do not agree. We would like to emphasize that the PYVAR-CHIMERE system for inversion is not a 4D-VAR one.

-The calculation of the size of the control vector is erroneous. The vertical dependence of the initial conditions is missing in the calculation.
The vertical dependence is indeed taken into account. The number of components in the control vector has been corrected.

-In the experiment section (section 4) no information on the initial and boundary conditions is given.
We now provide more information about the initial, lateral and top boundary conditions in the new "Section 4.1.2. CHIMERE set-up": "Different climatological values from the LMDZ-INCA global model [Szopa et al., 2008] or from a MACC reanalysis are used to prescribe concentrations at the lateral and top boundaries and the initial atmospheric composition in the domain."

-It should be illustrated to what degree both are changed during the inversion. Further, a comparison or sensitivity test should be shown on what the impact of emission optimization is compared to a joint optimization with initial and boundary conditions.
We choose not to perform such a sensitivity test. We have added text in "Section 4.2.2. Covariance matrices B and R" to explain this choice: "Based on the sensitivity test in Figure 5, the errors assigned to the CO and NO_2 lateral boundary conditions and to their initial conditions are set at 15%. As these errors are significantly lower than those of the emissions and as CO surface concentrations are mainly due to emissions (Figure 4), we assume a small relative influence of the correction of initial and boundary conditions on our results."

-Although the two test cases show a reduction of the difference between the assimilated observation and the analysis, this is not a proof of the successful operation of the data assimilation algorithm. A comparison with independent observations and a table with quality measures (e. g. bias, root mean square error, cost reduction) is necessary.
Indeed, but the main purpose of a GMD paper is less to present specific inversion studies than to describe the general modeling framework. We chose not to present evaluation for our illustration, as this is not the scope of this paper. Nevertheless, we have added sentences about the reduced mean bias between the observations and the simulation using the posterior emissions instead of the prior ones in Section 4.2.3 and in Section 4.2.4 to show the successful operation of the inversion system.

- It is advised to perform the analysis on a few consecutive days to assess the stability and quality of the inversion on different days.
We agree, to assess the stability and quality of the inversion, we now present a period of 7 days for the CO inversion.

- in the description of the test cases both, initial values and boundary conditions are included in the control vector, thus, the analysis is not complete without showing these two variables. A discussion is needed about the correction for all three variables, i.e. emissions, initial values, and boundary conditions, their relative influence on the analysis and about potential limitations of the inversion.
As already explained above, we choose not to perform such a sensitivity test. We have added text in "Section 4.2.2. Covariance matrices B and R" to explain this choice: "Based on the sensitivity test in Figure 5, the errors assigned to the CO and NO_2 lateral boundary conditions and to their initial conditions are set at 15%. As these errors are significantly lower than those of the emissions and as CO surface concentrations are mainly due to emissions (Figure 4), we assume a small relative influence of the correction of initial and boundary conditions on our results."

-If the full adjoint of the chemical processes is used there should be an adjoint signal for other species than CO and NO2 as well. This must be clarified. Are these signals simply not considered or not discussed? What is the reason for not optimizing NO emissions then?
The adjoint compute the sensitivity to all the components of the x vector. In the inversion we made for the former version of this paper, we chose to only infer NO_2 emissions. As both the reviewers have been disturbed by this illustration, CO, NO and NO_2 are now included in the x vector and we now present inversion for NO_x emissions. Other species could be considered for other studies.

-the model resolution of 0.5 x 0.5 square degrees seems to be a bit coarse for anthropogenic emission assessments. Is nesting available? A discussion on this point is needed.
Yes, nesting is available in CHIMERE. A number of studies for anthropogenic emission assessments have been done at even coarser resolution than 0.50x0.5°: for example, Miyazaki et al. [2017] using an approximately 2.8◦×2.8◦ resolution with the global CTM MIROC-Chem or Wang et al. [2019] using a 2.5° or 2.5° resolution with Geos-Chem.

We discussed about finer resolutions in the perspectives of the study.

Miyazaki, K., Eskes, H., Sudo, K., Boersma, K. F., Bowman, K., and Kanaya, Y.: Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, Atmos. Chem. Phys., 17, 807–837, https://doi.org/10.5194/acp-17-807-2017, 2017.

Wang, Y., Wang, J., Xu, X., Henze, D. K., and Qu, Z.: Inverse modeling of SO_2 and NO_x emissions over China using multi-sensor satellite data: 1. formulation and sensitivity analysis, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-879, in review, 2019.

-a better description of the B-matrix is needed in section 4.1.3.
We agree, we have added information about the B-matrix, in Section 3.3: " Different simple but efficient ways of building the error covariance matrix B are implemented in PYVAR-CHIMERE. The variances and correlations are defined independently. The variances are specified by the user through standard deviation coefficient (Table 1), which can be a fixed value ("fx") or a percentage ("pc") to define the diagonal standard deviation matrix $\sum$. For correction types "mult" and "scale", as well as for correction type "add" with a fixed value, the value is directly used as the standard deviation of the uncertainty in the corresponding components of the control vector. For correction type "add" with a percentage provided, maps of standard deviation of uncertainty are built by applying this percentage to the matching input fields (fluxes, initial conditions, boundary conditions). The user may also provide a script to build personalized maps of variances.

Potential correlations between uncertainties in different types of control variables, e.g. between fluxes and boundary conditions, and correlations between uncertainties in different species, e.g. between fluxes of CO and NO_x, are not coded yet. Only correlations for a given type of control variable and a given species are so far taken into account so that the B matrix is block diagonal. For a given type of control variable and a given species (in the illustration in section 4.2.2: CO, NO or NO_2 fluxes), spatial and temporal correlations can be defined using correlation lengths through time Lt and space Ls. Those lengths are used to model temporal and/or spatial auto-correlations using an exponentially decaying function: the correlation r between parameters and at a given location but separated by duration $d(x_i, x_j)$, or at a given time but distant by $d(x_i, x_j)$is given by $r(x_i, x_j) = exp\left(\frac{-d(x_i,x_j)}{L}\right)$where L $(= L_T \; or L_S)$ is the corresponding correlation length. There is no correlation between uncertainties in land and

ocean flux. . Note that the spatial correlations are computed for each vertical level independently when dealing with control variables with vertical resolution (3D fields of fluxes when accounting for emission injection heights, or boundary/initial conditions). Vertical correlations in the uncertainties in such variables have not been coded yet. Apart from this, the system assumes that temporal correlations and spatial correlations depend on the time lag and distance but not on the specific time and location of the corresponding parameters. It also assumes that the correlation between uncertainties at different locations and different time can be derived from the product of the corresponding autocorrelation in time and space.

Each block of B can thus be decomposed based on Kronecker products: $B=\sum Ct \otimes C_s \sum$where $\otimes$ is the Kronecker product, C_t and C_s are the temporal and spatial correlations, respectively. The calculations involving $B^{1/2}$ are simplified in PYVAR-CHIMERE using the Eigen-decomposition of Ct and Cs. Its square root can be calculated according to: $C_t^{1/2}= V_{Ct}D_{Ct}^{1/2}V_{Ct}^{T}$ (and similarly for C_s) (Eq 4) where V_{Ct} is the matrix with the Eigenvectors as columns, and D_{Ct} is the diagonal matrix of Eigenvalues of C_t. It is possible to chose a threshold under which the eigenvalues are truncated when computing the spatial correlations in order to save computation and memory, but not when computing the temporal correlations."

-What about correlations for initial and boundary conditions?
There is no correlation for initial and boundary conditions in our inversions.

Further minor comments:
- line 30: (VOCs) instead of "(VOCs))"
It has been changed.

-line 39: reference for (LRTAP) would be appreciated
We added the UNECE website as a reference for LRTAP.

-line 43: no commas
The commas have been removed.

- line 85: CO and NOx (instead of "CO, NOx")
It has been changed.

-line 88: citation van der A. [2008] is not appropriate (van der A et al. [2008]), also in the reminder of the manuscript
It has been changed.

-line 93/94: ".. for which variational methods are more suitable than KFs by design": a reference would be appreciated for this statement.
This sentence has been removed.

-line 122/123: of the current inversion (instead "of the inversion")
It has been changed.

-line 163: quasi-Newton (instead "quasi-Newtonian")
It has been changed.

-line 165: Reference for incremental 4D-var approach is appreciated
We would like to emphasize that the PYVAR-CHIMERE system for inversion is not a 4D-VAR one. We do not need reference for incremental 4D-var approach.

- line 203: It would be appreciated if the manuscript contains a table with the available (and adjoint) processes of CHIMERE
The available processes of CHIMERE are already listed in Figure 1 (emissions, transport, chemistry and deposition).

- line 227/228: better: "PYVAR, CHIMERE, and text sources are displayed in blue, orange, and grey boxes, respectively."
It has been changed.

- caption of figure 4: better: "Simplified scheme of how PYVAR scripts prepare the observations y using satellite data. PYVAR and text sources are displayed in blue and grey boxes, respectively."
It has been changed.

- line 264: Equation "Cm =" is not a correct mathematical formulation, Cm(o) is a column, xa is the state vector (a profile in this context).
Here, Cm(0) is not a column but the vertical distribution in partial subcolumns from a chemistry-transport model at the same satellite pressure levels. xa is not the state vector but the a priori profile provided together with the averaging kernels when relevant. We kept this formulation.

- line 290: ... days for CO and NO_2, respectively (instead of "... days, respectively for CO and NO2")
It has been changed.

- line 302: Table 1 is not control vector specific. This sentence can be removed
Indeed, we move this sentence in Section 4.

- line 304: for one day (instead "at a 1-day"); resolution (instead "resolutions")
It has been changed.

- line 313/314: a spin-up for the initial values is needed for an appropriate analysis, otherwise the model maybe to far off the observations for a suitable correction.
We agree, we indeed performed runs with a spin-up of 10 days. We have added this information in "Section 4.1.2. CHIMERE set-up": "In order to ensure realistic fields of simulated CO and NO_2 concentrations from the beginning of the inversion period, runs have been preceded with a 10-day spin-up."

- line 317: Reference for MOPITT is missing
Indeed, we have added a reference.

- line 328: MOPITT instead of "OMI"
It has been changed.

- page 12, line 4: flown instead of "flying"
It has been changed.

- line 368: parts (instead of "part"); present (instead of "presents")
It has been changed.

- page 14, last line: particularly over the Po Valley (instead ", and particularly over PoValley")
It has been changed.

- caption Fig. 7 d: is it really the difference between prior and posterior? Inconsistency with text (see next point)
- line 374: Fig. 5c seems to be wrong here. Is it Fig. 7d?
Indeed, this has been corrected.

- line 380/381: Using the full adjoint of CHIMERE, this must already be available. Please check for adjoint NO signals
As already explained above, we now present inversion of NOx emissions.

- line 399: remove "for example"
It has been removed.