Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-219-RC1, 2020

Geoscientific Model Development Discussions
Open Access

Interactive comment on "The Environment and Climate Change Canada Carbon Assimilation System (EC-CAS v1.0): demonstration with simulated CO observations" *by* Vikram Khade et al.

Anonymous Referee #1

Received and published: 15 September 2020

General comments:

The paper describes the new development of the coupled weather and atmospheric composition system based on the Environment and Climate Change Canada's (ECCC's) operational Ensemble Kalman Filter (EnKF). While the paper describes this new configuration as an important advance for the ECCC system it misses important points to provide an accurate and complete description that such system should deserves.

Printer-friendly version

Discussion paper

The first major point that needs to be addressed is that the paper advertises in several places that it is a greenhouse gases (GHG) atmospheric data assimilation and surface flux inversion system. However only CO atmospheric data assimilation is showcased. I would strongly recommend that the authors remove all claims that a flux inversion GHG system has been setup and then use a different terminology such as simply “atmospheric composition data assimilation” or “atmospheric carbon data assimilation” as in the title. The study uses synthetic observation to evaluate the system. Therefore, why the authors did not simulate the HYPNET CO2 and CH4 observations and perform the assimilation of such to at least justify the GHG component of the system?

It seems that the added value of the paper is the extension of the ECCC operational system to atmospheric composition using CO assimilation as a proof of concept. While the focus is on CO assimilation, very little importance is given to the meteorology assimilation evaluation in such configuration. How does this compare to the actual operational ECCC system? Almost no references are given to reader to refer to the NWP system and its evaluation. I would recommend the authors to give a short summary on the meteorological data assimilation rather than ascertaining that the meteorological data assimilation is working as expected.

The overall presentation of the paper requires strong efforts to improve clarity. Almost all parts of the paper lack clarity. Some parts are over emphasising some aspects that are not relevant for the evolution of the system while other parts that are important are covered very briefly. To give few examples:

- Very little is explained about the simulation of MOPITT synthetic observations, averaging kernels and their errors. It seems that a paragraph is maybe missing.

- Extensive description of the meteorological setup is given but very little is described and showed about the actual meteorological data assimilation results.

- Some of the terminology used is not really common for atmospheric data assimilation, I would encourage the author to revise this throughout the text.

Printer-friendly version

Discussion paper

- Several misleading statements about data assimilation and atmospheric composition need to be corrected. Please refer to the specific comments for details.

Specific comments:

Line 38: Be consistent, so maybe replace by air quality. Or explain that air quality is partly driven by weather.

Lines 39-41: This sentence has some shortcomings that could mislead the reader. Be consistent with the previous sentence and please develop this statement in more precise information. Air quality is a bit different from tropospheric pollution. Tropospheric atmospheric composition prediction is essential to air quality prediction which is looking at surface levels of pollutants. Tropospheric pollution prediction relates to longer time scales than 5 days, especially for CO. Air quality is driven by emissions variations and synoptic variations of weather regimes.

Line 41: Which data assimilation systems are we talking about here?

Line 63 and line 65: Swap years to chronological order

Line 78: This system now can estimate emissions using state augmentation as described in Gaubert et al., 2020 (Gaubert, B., Emmons, L. K., Raeder, K., Tilmes, S., Miyazaki, K., Arellano Jr., A. F., Elguindi, N., Granier, C., Tang, W., Barré, J., Worden, H. M., Buchholz, R. R., Edwards, D. P., Franke, P., Anderson, J. L., Saunois, M., Schroeder, J., Woo, J.-H., Simpson, I. J., Blake, D. R., Meinardi, S., Wennberg, P. O., Crounse, J., Teng, A., Kim, M., Dickerson, R. R., He, H., and Ren, X.: Correcting model biases of CO in East Asia: impact on oxidant distributions during KORUS-AQ, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-599, in review, 2020.)

Lines 81-82: Maybe this is a bit misleading as the paper seems to focus on CO (even if CO is important for GHG estimations). Also, the term "estimate GHGs" is a bit vague in my opinion. Maybe replace to something more specific such as "estimate CO atmospheric distribution".

Lines 88-91: This paragraph is not necessary here as some of it should be moved to the introduction.

Line 88: “Trial fields” is quite uncommon data assimilation terminology. Maybe replace by forecast, background, prior or first guess fields depending on what you are meaning by trial here.

Lines 88-90: The first and second stage are not explicitly mentioned. I would re-write those two general sentences with a more traditional way to introduce the general concepts of data assimilation.

Lines 94-95: The sentence “The model is initialized. . .” is confusing please rephrase.

Line 95: Please “trial fields” replace with appropriate traditional data assimilation terminology throughout the text.

Line 97: “Blending” is not really the correct word for the data assimilation procedure. I would recommend the author to use the appropriate vocabulary for data assimilation in the literature that tackles atmospheric data assimilation.

Lines 99-101: You do not really need to specify what will be the sections to come here. Consider removing.

Line 108: I do not think that “lib” is the appropriate terminology here. Please again replace with, for example: "... 80 vertical levels from the surface to 0.1 hPa."

Line 109: what type of hybrid coordinate? There are several of them.

Line 114-155: Please be more specific and add diffusivity in this sentence.

Line 120: Not correctly written. The atmospheric chemistry scheme is not removed for CO2. You remove the reactive chemistry in a model to increase its performance. Please rephrase.

Line 130: Start a new paragraph here as you now write about CH4 surface fluxes.

Line 135: Start a new paragraph here as you now write about CO emissions.

Line 148: Please define xf and xa here. xf and xa are commonly called the prior and posterior state respectively in the EnKF terminology. Alternatively, you could call them forecast (hence the superscript f) and analysis (hence the superscript a). Please consider using the commonly used atmospheric data assimilation vocabulary throughout the text for more clarity.

Line 148: Consider directly defining the other elements of the equation 1 before going into explanations.

Line 152-153: The sentence "Pf is the forecast error. . ." is a bit vague, please be more specific in the definitions.

Line 196: I think there are more relevant papers for this statement. In Inness et al., 2015 the system used was a CTM configuration where the meteorological fields are forced by external meteorological fields. In that sense the DA system could not drive any constrain on the meteorology. Please cite instead Barré et al., 2015 and/or Gaubert et al., 2016 and/or Kang et al., 2012 and so on... Those papers are using EnKF with this variable localisation between atmospheric composition and meteorological variables.

Line 200-201: The sentence "The spatial correlation. . ." seems to have no link with the previous ones. Please remove or develop in a new paragraph.

Lines 209-210: The sentence "To simulate model . . ." is unclear. Please rephrase and possibly add a reference for this error representation method.

Lines 214-215: But this paper is not doing flux estimation. Maybe consider changing to atmospheric composition data assimilation and change to the appropriate references.

Lines 218-219: "In EC-CAS, for the meteorological assimilation, the same scheme is used, but for GHGs, no such additive error is present." Is this the configuration used in this paper? If yes, why then bother going though all these details above?

Line 224: Then why not using synthetic GHG observation of CO_2 and CH_4 (amongst other GHGs)?

Line 236: change to “the use of a surface flux”. I would recommend the authors to be consistent with this terminology as fluxes are not necessarily at the surface in the atmosphere.

Lines 269-270: “. . . its retrieved profiles are sensitive to CO in the lower troposphere where . . .” MOPITT retrievals are sensitive throughout the entire troposphere. The multispectral retrievals allow an enhanced sensitivity towards the surface over land only when the conditions are favourable. Please correct and amend the text accordingly.

Line 271: What are those data assimilation systems? This statement is not true. Number of air quality DAS only assimilate surface stations. Please be more accurate here.

Line 273-274: This statement is misleading. You do not use the averaging kernel to construct the observation operator. You feed the observation operator with the averaging kernel to sample the first guess.

Line 276: This is unclear. Does this mean you discard all observations that have a retrieval surface pressure below 1000 hPa? I do hope you are not doing this. Please clarify the sentence.

Lines 277-278: This is not the proper definition of the averaging kernel matrix. Please use the common definition given by Rodgers 2000.

Inverse Methods for Atmospheric Sounding. Theory and Practice. https://doi.org/10.1142/3171 | July 2000. Pages: 256. By (author):; Clive D Rodgers (Oxford).

Line 278: H is not a forward operator but only an observation operator in the Kalman filter as it does not need to generate a forward model prediction to get a model equivalent quantity. It is true in for example the 4D Var formulation. Please correct.

Line 281: The authors do not use the same system as in Jiang et al., 2015a. If they do this needs to be clearer earlier in the paper. If not, please recall a bit more of the methodology or use the appropriate reference to the system used in this paper.

Line 285: “varied between 10-16%” is this the value that the authors use to set up the observation errors. It seems that few sentences are missing to explain the setup on MOPITT observation errors.

Lines 289-290: This sentence is hard to understand. Please rewrite.

Line 290: “other issues” Please be specific of what other issues.

Lines 293-299: So why do the authors bother simulating observations then? Why not testing the DAS in real conditions? Please justify more clearly the choices here and certainly earlier in the paper.

Lines 308-309: The statement “An ensemble of forecasts…” is incomplete as is, I would remove it as this would need couple sentence to make this point clear and this paragraph is not the place for that.

Lines 311-312: This was already mentioned earlier. Remove.

Line 320: Regarding the reference to Pires et al., 1996, I think numerical weather prediction and predictability ranges have evolved since the mid-90’s. Please use a more recent reference. Also, the time of the DAS RMSE stabilisation is not due to weather predictability but mostly due to the DAS setup, i.e. background error, observation density, type and error and so on... Please rewrite the related statements.

Lines 321-322: The authors could add winds, surface pressure and RH (or another NWP variable of your choice) in a four-panel plot to make your point stronger and avoid such statement.

Line 328: What is the “additive model error term”? Is it inflation? Please refer to the section where it is defined and explained? If not define here and/or add the appropriate reference.

Lines 332-333: This is statistically expected considering Gaussianity and the truth being drawn from the prior distribution itself. Please modify the statement accordingly.

Line 337: change “establishes” by “is”

Line 338-339: The authors should stop recalling what would be the next sub-sections at the end of each sub-sections.

Line 341: Please use more common vocabulary; “trial” is not used in atmospheric data assimilation.

Line 347: what is the matrix inverse. Is it the inverse of P?,H?, R? Or K? Please be more specific.

Lines 351-352: The sentence “The scaling factor. . .” is hard to understand. What is the scaling factor of the innovation? Please define.

Lines 356-357: The statement “In theory a given. . .” is misleading statement. In the theoretical case of a perfect ensemble with an infinite number of members, the spurious correlation would not exist, and you would not need to localise. Hence you would apply the filter globally. Please remove or change accordingly.

Line 359: The statement “small correlations cannot be trusted” is misleading. A GC localisation is not applied to remove small correlations but spurious correlations that are far from the observation location. Small correlations are not necessarily spurious. This also depends on the ensemble size and nature of the state (e.g. lifetime and transport). Please remove or change accordingly.

Line 363: What the meteorological cut off values? Please detail and/or provide reference.

Line 365: Change “has a peak” to “has its maximum”.

Line 407: Change “blob” to “area”.

Lines 409-410: This is incomplete. The transport of corrected concentration plays a major role as well. I would say this is the combination of both in your case. Please update the text accordingly.

Line 411: If the surface only concentrations and not the 0-5km column were displayed different results might appear as the observation network is at the surface. Also, it is hard to tell in figure 6 that the RMSE is much lower in Western Canada as this is at the edge of the colour scale. The authors should zoom and adjust the plot to make the point clearer.

Lines 441-442: Again, this is not only the EnKF but also the transport of corrected concentration by the model itself that improves the RMSE. Please correct the text.

Line 471: I would disagree with that statement. The vertical information content in the MOPITT retrieval as opposed to HYPNET is not precisely located but spreads across the vertical. So, this is not because the degrees of freedom on the vertical are comparable that the vertical information is similar. Please correct the statement.

Line 473: The authors do not show this as a not directly GHG gas has been assimilated. I would suggest removing GHG but change to something as "atmospheric composition" as only CO assimilation has been demonstrated in the paper.

Lines 484: This is true, but this needs to be reformulated correctly. Please mention atmospheric transport.

Lines 494-495: I am not convinced this is a conclusion from Miyazaki et al., 2012. CO surface flux errors can be correlated with other fields if you consider the co-emission of different species through a given sector. Remove or change the statement accordingly.

Line 506: The authors did not show anything about flux inversions. Please remove this statement.

Line 508: Please define smoother. Add a reference. Use the book from Bocquet et al., 2016 for definitions of the smoother.

Line 509: Future observations? That do not exist unfortunately... Please remove or correct.

Figure 7: What do the numbers in the x-axis indicate? Please clarify or change.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-219, 2020.