Review of

Four-dimensional Variational Assimilation for SO_2 Emission and its Application around the COVID-19 lockdown in the spring 2020 over China

by

Yiwen Hu et al.

Overall comments

This manuscript describes a four-dimensional variational (4DVAR) data assimilation (DA) system designed to update SO_2 emissions by assimilating surface SO_2 concentrations. The system was applied to study emissions reductions over China during the beginning of the COVID-19 pandemic. Additionally, the authors showed how obtaining SO_2 emissions from 4DVAR can potentially lead to better subsequent SO_2 concentration forecasts relative to using static SO_2 emissions.

I think the topics in this manuscript are interesting and important, and I have few concerns with the overall objectives of this work. However, I think there are shortcomings regarding descriptions of the DA system and the experiments, most of which are related to lack of clarity. In addition, I don't think the OSSE part of the manuscript is necessary, and I instead suggest replacing this material with other metrics demonstrating that the 4DVAR DA system works as expected. Finally, the writing is mostly understandable, but there are many grammatical errors that should be fixed prior to publication in ACP.

Major comments

1. I think more descriptions about the mechanisms through which your 4DVAR DA system updates SO_2 emissions are needed. It isn't entirely obvious how this adjustment is achieved using 4DVAR. Did you somehow model cross-covariances between SO_2 concentrations and emissions? Or is the tangent-linear/adjoint model effectively altering the emissions? I was left unsure of exactly how you updated SO_2 emissions, and I think this key point of your method needs to be explicitly described. It might also be worth contrasting your method with how an EnKF works; in the EnKF it would be straightforward to update SO_2 emissions, as SO_2 emissions could simply be added to the state vector and be naturally updated by ensemble-based covariances.

2. I believe that when harvesting synthetic observations from "truth" in an OSSE it is customary to add noise drawn from an "observation error" distribution to the synthetic observations. However, your methods don't mention doing this, which makes me wonder if the OSSE was properly performed. By not adding noise to harvested observations, you will likely get analysis fits much closer to "truth" than if you add noise to the synthetic observations you eventually assimilate.

   Regardless, I don't feel the material about the OSSE adds much to the manuscript, and it is little more than a "sanity check" that I didn't find very convincing. I think a better

demonstration of the 4DVAR system's efficacy would be to show plots of: 1) The cost function reduction from a 4DVAR analysis that assimilated real observations; 2) Analysis increments directly showing how SO_2 emissions were modified by assimilating SO_2 observations; and 3) "observation minus background" and "observation minus analysis" statistics. I think presenting these types of plots would increase confidence that your 4DVAR DA system works as expected and can provide stronger evidence of system robustness than the OSSE.

Additionally, I'm somewhat concerned with Fig. 5b, which shows that the cost function increased between the 8th and 9th iterations. Variational cost functions must monotonically decrease with each inner-loop iteration, so there is a chance that something went wrong. Please look into this or offer an explanation.

3. I think section 2.4 needs to be more specific about what you actually did, rather than making fairly general statements. What specific observation error did you use? What were your values of ε_r and ε_o and how did you arrive at these values?

   Similarly, in lines 201-204, please state how you produced these 48- and 24-h forecasts. What model configuration did you use? You should also cite the "NMC method" (Parrish and Derber 1992) for this approach. Moreover, in line 202, should "state variables" be "background errors"? How did you model the correlations in *C*, especially for the emissions? Overall, please be more specific about your background and observation error covariance construction.

4. Several aspects of the experimental design were not initially clear to me and caused confusion. Although some elements became clearer with time, I think descriptions of the experiments should be clarified:

   a. From Table 3, it appears that you performed DA experiments for ~3 weeks. What was the cycling period of your experiments (i.e., how often did you produce new 4DVAR analyses)? Was it 6 hours? Did you continuously cycle both the chemistry and meteorology, or did you periodically update meteorology from an external source, like the GFS model? What did you do for chemical boundary conditions (my apologies if I missed it)? Furthermore, Table 3 states 24-h forecasts were produced, but how often did you initialize these 24-h forecasts? Overall, the temporal aspect of the experiments should be clarified.

   b. Fig. 3 didn't seem clear to me. What specific field(s) are being updated? Just SO_2 emissions? Or both SO_2 emissions and concentrations? Additionally, this figure might be clearer if you annotated the mathematical symbols from Eq. (1) or (3) on it so readers can link this figure to the equations. It might also be nice if you added another panel to the figure showing the temporal progression of the DA system (per above comment). Finally, in the top-left box, there's a typo (it should be "field").

c. It appears that you ran the 4DVAR DA system over two separate periods to estimate SO_2 emissions: 1) 17 Jan – 7 Feb 2019; and 2) 17 Jan – 7 Feb 2020. However, you never explicitly stated this! Thus, there are really two parts to this work. The first is estimating emissions in 2019 and 2020 from the 4DVAR DA system. The second is using different emissions estimates to drive various sets of forecasts over a common period in 2020. This distinction was not always clear, which caused me confusion.

   Please explicitly state these various experiments and their purposes. It is important that you do so because of places like line 328, where you simply stated the *years* of the emissions, and not the names in Table 3 (which contain years); more distinction needs to be made for data *for a given year* versus various experiments *in 2020* that used emissions generated from various years.

d. For the DA_2019 experiment, you effectively seemed to use a "pre-processing" step, where you ran the 4DVAR DA system over 2019 and then used those emissions when simulating a period in 2020. It might be worth noting that in contrast, the DA_2020 experiment did everything all at once without the need for a "pre-processing" step.

5. Section 3.2: Please clarify that the data (i.e., emissions) discussed in this subsection were obtained from the 4DVAR analyses (I think), and not some other source. In general, please be more precise about from where the data on each figure come from.

6. The writing is understandable but there are many grammatical errors that I found distracting. Please carefully proofread the manuscript.

Minor comments

1. Line 48: "most explored algorithms" for what? Please be specific.
2. Lines 50-51: The second instance of "to estimate" should be removed, and perhaps "total regional and global emissions" should be moved to after the first instance of "to estimate".
3. Lines 60-63: This statement is too broad. There are techniques to handle this problem, like inflation, that are well established at least for meteorological EnKF DA. Please refine this statement.
4. Line 89: Please add a reference for WRF-Chem.
5. Line 105: Suggest "covering all of China" instead of "covering the entire country".
6. Lines 105, 158, 159: "resolution" should be "grid spacing", as the two are not the same.
7. Lines 111, Table 1: Please add a reference for the Grell-3D scheme.
8. Fig. 1 caption: Please state that this figure also shows the WRF modeling domain!
9. Lines 122-125: Somewhere in here, please specifically define n.
10. Somewhere in section 2.2, please be more precise about which "control variables" are included in c_o. Is it just SO_2 concentration?
11. Line 131: Here, is H nonlinear? (probably it is). Please state.
12. Lines 131-132: Please state that R is the observation error covariance matrix.

13. Line 133: Please define *f*.
14. Eq. (3): Some more explanation is needed about how you go from Eq. (1) to Eq. (3) for the observation term. Specifically, please note the linearization about the background.
15. Lines 150-152: How does this relate to the cycling period of the DA system? Does this mean you produced new analyses every 6?
16. Line 161: "large horizontal resolution study" is unclear. Are you referring to your specific study or something else? Please clarify.
17. Eq. (10): L_{turb} doesn't appear in the list of quantities in line 169, and L_{dry}, which does appear in line 169, doesn't appear in the list of equations. Please clarify. Also please double check Eq. (14).
18. Lines 170-185: Should L_{trub} be L_{turb}?
19. Line 180: Should it be Eqs. 9–13 instead of Eqs. 9–12?
20. Lines 168-185: I'm not an expert about adjoint modeling, but I had the feeling that these lines aren't precise enough about the adjoint model formulation. Shouldn't there be more derivatives in there?
21. Line 192: Please omit "the assimilation variable, which is the"…it's confusing, because that phrase is somewhat referring to a state/control variable, even though you're really talking about observation errors.
22. In lines 220-228, please be very precise about "emission" vs. "concentration" in your descriptions.
23. Line 243: Please change February 6 to February 7 for consistency with Table 3.
24. Line 246: Typo: it should be "physiochemical".
25. Line 249: Suggest "…based on the spin-up forecasts initialized at 0000 UTC…".
26. Line 250: Please be more precise about "the previous day" (this comment relates to earlier comments about the cycling period).
27. Line 256: There seem to be more than 13 x 9 points in Fig. 4, so I was confused about this statement concerning "arrays and columns". Please clarify.
28. Fig. 5: The legend in the left panel is covering data and should be moved, and the y-axis in the right panel should probably be "J" not "J_b".
29. Line 282: I'm not sure I agree with this statement, especially in (a) and (b); the 2019 SO_2 concentration decreases with time but the SO_2 emissions seem steady. Please revise.
30. Line 283: To my eyes, it looks like the lowest emissions were on 1 February, not 3 February (per Fig. 6b).
31. Lines 289-292: Please point to Figs. 6c,d here.
32. Fig. 6 caption: Please state the meanings of the vertical lines.
33. Line 311: I believes "rates" should be "ratios".
34. Fig. 7: What are the insets in the lower right corner of each panel? Additionally, please be more precise about the subtraction convention. Above (c) and (d), it says "2020 – 2019" but the caption says, "differences between 2019 and 2020", which implies "2019 – 2020". It might be clearest to just write out "2020 minus 2019". Finally, please state in the caption whether these statistics are averaged over the entire period. Similar comments also apply to Fig. 8.
35. Line 315: I don't think "observations" is the correct word. Is "analyses" more accurate? Please also see line 363.
36. Line 320: Please remove "slightly"; it's too subjective.

37. Line 321: Please remove “Remarkably”, which is also subjective, and furthermore, the differences don’t seem “remarkable”.
38. Line 333: Please remove “slightly”. Also, it seems that this behavior was only evident in Fig. 9a, so please clarify the region you are discussing.
39. Fig. 9 caption: Are these statistics averaged/aggregated over the entire period and over all sites or grid points? Please clarify.
40. Lines 352-355: I found this chunk troublesome. The explanation you offered didn’t make sense to me, and I’m not sure all your statements are accurate. Please clarify or omit.
41. Throughout, including figure captions: “Skill” should be “accuracy”. Skill is “accuracy relative to a baseline”, and all of the metrics you are showing are measures of accuracy, not skill. I believe every instance of “skill” needs to be changed to “accuracy”.
42. Fig. 10 caption: Please clarify whether these statistics are averaged/aggregated over the entire time period and all sites. Same comment for Fig. 11.
43. Lines 369-370: Please omit “compared with the Ctrl_2016 experiment”.
44. Lines 373-374: Please clarify what you mean by the “background field”. Do you mean the field at the very start of the period (0000 UTC 17 January 2020)?
45. Line 390: Can you point to a figure for this key result about the decrease of optimized emissions? Also, did you ever state these values in the results section (sorry if I missed it)?
46. Figs. 1, 4, 5, 8, 9: Please add annotations (e.g., “a”, “b”) to all these figures.