

Review of

Background error covariance with balance constraints for aerosol species and applications in data assimilation

by

Z. Zang et al.

Summary and general comments

This paper discusses implementation of cross-correlations between aerosol variables in a variational data assimilation (DA) system via balance constraint. The authors describe their methodology and then apply their new developments for a single 24-hr forecast over Southern California. Results suggest that incorporating cross-correlations within the DA system was beneficial, especially for 3- to 18-hr forecasts.

This paper is generally interesting and good, although there are some shortcomings that I believe should be addressed before publication. My biggest concerns regard that only one forecast was produced and lack of discussion about other methods of dealing with cross-correlations for aerosols, such as ensemble-based DA methods. Additionally, there are many areas of text that I believe require some clarification.

Bigger comments, questions, and concerns

1. I appreciate that you actually implemented your developments in a DA system to see the real-world impacts. However, you only showed results from one forecast, which does not give much confidence regarding the generality or strength of the results. If possible, I strongly urge you to add more cases. I know that adding more cases requires more work, but doing so would not add much to the length of the paper and would make the conclusions much more robust.

2. In my opinion, you neglected to discuss another (and easier) method of dealing with cross-correlations between aerosol species: ensemble-based DA methods (such as the ensemble Kalman filter) that naturally handle cross-correlations. Thus, I strongly believe you should mention ensemble DA methods in the introduction, and you should cite and briefly discuss Pagowski and Grell (2012) and Schwartz et al. (2014), who assimilated aerosol observations, including PM_{2.5}, with ensemble-based DA methods. There are other references that have also assimilated aerosol observations with ensemble DA, but I believe those two are the most relevant, and without this material regarding ensemble DA, I believe your work is not placed within its proper context.

3. In light of the above comment, I believe your title should be more specific, and I suggest adding the word “variational” before “data assimilation”.

4. You left out a few important details about the DA system. For example, what DA system were you using? Was it GSI or some other system? Please briefly explain somewhere in the text. Additionally, for your 24-hr forecast you described in section 5, what was the background for DA? Finally, please briefly state the observation errors that you used.
5. I believe some aspects regarding Eqs. (6-13) need clarification.
 - a) Page 8, line 9: Please clarify what you mean by “first variable”.
 - b) Page 8, Eq. (7): Please fill-in the upper triangle of \mathbf{K} . Are all upper-triangle elements zero?
 - c) Page 8, line 19: Please clarify what you mean by “a one regression coefficient.”
 - d) Some more details about how you compute ρ_{ij} would be beneficial.
 - e) Additionally, I think it would be nice if you provided some details on how to interpret ρ_{ij} to bolster the discussion on page 14.
 - f) What would happen if the regression was not “based” on EC? In other words, what would happen if you listed the control vector species in reverse [such that OTR was in the first row on the LHS of Eq. (7) and EC was in the last row]? You mention some of this on page 14 lines 6-8, but I believe a clear description about the “order” or “first and second variables” would be greatly beneficial. You also mention using OTR as the “last variable” (page 14, line 19), but the rationale for this choice is not obvious to me. Please clarify.
 - g) Page 9, line 1: I feel like the word “deduced” to describe ρ_{21} is inaccurate. How exactly are you obtaining ρ_{ij} ?
 - h) Page 9, lines 3-10: I found ϵ confusing and also unnecessary. By definition, $\epsilon = \delta\text{OC}_u$, so why not just use δOC_u directly in place of ϵ ? Thus, I suggest removing all instances of ϵ .
 - i) Page 9, Eq. (11). I believe you’re missing “ δ ” on EC and OC_u .

Smaller comments, questions, and concerns

1. Page 2, line 14: Please clarify what you mean by “coincident”.
2. Page 2, line 17: Please omit the word “significant” because you did not perform any statistical significance testing, and you only showed results from one forecast.
3. Page 2, line 21: Again, omit “significantly”.
4. Page 2, line 26: Technically, the observation errors also determine the analysis increments.
5. Page 3, line 5: Most models now have a state size $O(10^7)$. Suggest modifying.
6. Page 3, lines 3-8: Note that with ensemble DA methods, these issues are not as difficult to deal with.
7. Page 3, line 12: Please define in words what you mean by $\text{PM}_{2.5}$.
8. Page 3, line 13: Suggest spelling out GSI and adding a reference.
9. Page 4, lines 9-11: Do these assumptions only apply to variational approaches?

10. Page 4, line 20: This might be a good place to mention Pagowski and Grell (2012) and Schwartz et al. (2014).
11. Page 5, line 1: Please spell out “AOD”.
12. Page 5, line 4: Please clarify what you mean by “not adjacent”.
13. Page 5, line 10: Please clarify what you mean by “eight/four”.
14. Page 5, line 12: Suggest “developed” rather than “estimated”.
15. Page 6, Eq. (1): It should be $J(x)$ not $J(\delta x)$.
16. Page 6, lines 20-25 and Eq. (2): You’ve ignored non-linear H and its linearization about the background to derive the linear \mathbf{H} . In Eq. (1), H is non-linear, but in Eq. (2) it’s linear, because you’ve linearized H about \mathbf{x}^b . Please be more precise.
17. Page 7, line 2: In the expression for the innovation, here H should be non-linear (H).
18. Page 7, line 7: Again, it should probably be 10^7 rather than 10^6 . Also, 10^{12} should probably be 10^{14} .
19. Page 7, line 14: Please clarify what you mean by “is commonly simplified with vertical levels.”
20. Page 10, lines 8-9: It was unclear to me how you got Eq. (17) from Eq. (6). Please add some steps or clarify.
21. Page 11, line 1: In Eq. (20), it appears you used $\delta x = \mathbf{B}^{1/2}\delta z$. Thus, I believe line 1 on page 11 should read $\delta z = \mathbf{B}^{-1/2}\delta x$ (note the negative sign on the exponent of \mathbf{B}). Please double-check.
22. Page 12, line 5: Should be “horizontal grid spacing” not “resolution”...they mean different things.
23. Page 12, lines 10-12: Please clarify what you mean by “former forecast”. Additionally, where do the initial meteorological conditions come from? Are these also from NARR?
24. Page 12, line 26 and page 13, line 2: I wonder if you might want to rename “E_ORG” to “E_OC” and “E_PM25” to “E_OTR” to be consistent with the nomenclature of the control variables. If so, please also change on the relevant figure (Fig. 2) and elsewhere in the text.
25. Page 13, line 5: “With the exception” is misleading and suggests that the diagonal correlations will be < 0.5 . Please modify.
26. Page 13, line 9: Suggest “high” rather than “close”.
27. Page 14, line 1: I believe it should be Eqs. (6-13) rather than Eqs. (6-12).
28. Page 14, line 2: I believe Eq. (7) is more correct than Eq. (6).
29. Page 14, lines 9-19: Should the control variables here have subscripts “u”? I’m not sure. Please double-check.
30. Page 14: Just a comment—I really like Fig. 3.
31. Page 15, line 2: Suggest “obtained” rather than “performed”.
32. Page 15, lines 2, 3, and 8: In all these locations, it should be Fig. 4, not Fig. 2.
33. Page 15, lines 10-11: I believe OTR and NO_3 should be OTR_u and NO_{3u} , respectively.
34. Page 15, lines 10-11: Please clarify with what the “decreases” are with respect to.
35. Page 15, line 17: I believe it should be Eq. (22), not Eq. (21).

36. Page 15, lines 18-25: Please explain how you get the horizontal correlation scale (L_s) from Fig. 5. Is L_s defined as an e-folding distance? Overall, I was a bit confused by your description of L_s —please clarify.
37. Page 15, line 27: I believe OC, NO₃, SO₄, and OTR should have subscript “u”.
38. Page 16, line 1: Please clarify what you mean by “common factors in regression equations”.
39. Page 16, lines 4-16: Similar to my above comment, please explain how you get the vertical correlation length-scales from Fig. 6.
40. Page 16, lines 13-16: I only see very small differences regarding the vertical correlations between the full and unbalanced variables. Perhaps you may wish to modify the text.
41. Page 17, line 23: Please clarify that DA-balance assimilates the same observations as “DA-full”.
42. Page 17, line 25: “WRF” not “Wrf”.
43. Page 17, line 28: Please clarify what you mean by “the initial time”.
44. Page 18, lines 4-26: I feel like this discussion slightly misses the main points. In my opinion, the main point is that the balance constraints can allow *observations of a specific species* to impact other variables. Even with PM2.5 observations, because the model-simulated PM2.5 is a function of all the control variables, the individual species’ fields are adjusted through the BECs, even without a direct observation of the individual species. Thus, without multivariate correlations, an aircraft observation of OC can only impact OC (because the forward operator for OC is only a function of OC), but with the multivariate BECs, an OC observation can now impact OTR or EC. Perhaps you might wish to clarify some aspects of the text along these lines.
45. Page 19, lines 1-5: I don’t believe it is appropriate to describe the smaller RMSEs as “improvements”. You’re simply looking at *fits* to observations, which, on their own, do not tell you anything about the relative goodness of your DA system.
46. Page 19, line 17: The description here of Fig. 11 is incorrect.
47. Page 19, line 19: It should be Fig. 11a, not 1a.
48. Page 19, line 20: Omit “significantly”. You can maybe replace it with “substantially”.
49. Page 20, lines 16-17: Please clarify what you mean by these lines.
50. Page 20, line 27: Please clarify what you mean by “mutual spread”.
51. Page 21, line 6: I don’t agree with this line. You’re only looking at the analysis fits, which does not mean your analysis fields are necessarily better.
52. Page 21, line 20: Please clarify what you mean by a “universal balance constraint”.
53. Table 1: Suggest also pointing to Eq. (7) in the caption. Also, you should annotate the various species on this figure somehow, because it’s difficult to look back to Eq. (7).
54. Fig. 2 caption: Suggest “NEI05” rather than just “NEI”
55. Fig. 4 caption: In my opinion, this figure isn’t that close to Fig. 3 so I suggest elaborating.
56. Fig. 5 caption: Suggest pointing to Fig. 4 rather than Fig. 3.

57. Fig. 6: Suggest adding labels of “Height” to the axes.
58. Fig. 7: Suggest adding a unit (meters) to the colorbar.
59. Figs. 8 and 9: The labels above/below the panels are very small. Can these be enlarged?

References

Pagowski, M., and G. A. Grell, 2012: Experiments with the assimilation of fine aerosols using an ensemble Kalman filter, *J. Geophys. Res.*, **117**, D21302, doi:10.1029/2012JD018333.

Schwartz, C. S., Z. Liu, H.-C. Lin, and J. D. Cetola, 2014: Assimilating aerosol observations with a “hybrid” variational-ensemble data assimilation system. *J. Geophys. Res. Atmos.*, **119**, doi: 10.1002/2013JD020937.