Review of

Generalized Background Error Covariance Matrix Model (GEN_BE v2.0)

Descombes, Auligné, Vandnberghe and Barker

Reviewer: R. Bannister (waiving anonymity)

1 General comments

This paper provides scientific and technical documentation for v2.0 of the GEN_BE system for modelling background error covariances in meteorological data assimilation. The paper gives a general background to the steps that make up the change of variable (control variable transform) used to represent compactly the B-matrix in 3D-VAR and summarizes the options that GEN_BE v2.0 provides. A selection of statistics (e.g. correlations, length-scales, eigenmodes) and pseudo observation tests are shown to demonstrate the performance and capabilities of the suite of code.

The methods themselves used in the control variable transform are not new - most are used in other systems (like empirical orthogonal function decompositions, digital filters and statistical regressions), but the flexible way that they are adopted in this system is innovative and potentially useful to many forecasting systems other than the ones used in this paper. The flexibility includes the extension of the control vector to include hydrometeors like snow, rain, ice and cloud, which will inevitably be useful to other data assimilation researchers across the world who work in, e.g., radar assimilation or convective-scale data assimilation in general.

The presentation in general requires some attention. There are far too many grammatical errors, spelling errors, and mathematical errors. I have highlighted in Section 3 of this report as many errors as I can. There is sometimes a lack of consistency in presenting information throughout the paper (e.g. in representing elevation in the atmosphere, sometimes model level is used, other times pressure is used - this makes it difficult to compare plots even with the pressure/level plot - Fig. 14, which in any case appears far too late in the paper). It should be clear from each Fig. that shows statistics, the source of the data used, its type (ensemble or NMC), the averaging done (over how many days), and the modelling method used to produce the plot (e.g. EOFs or recursive filters). This is not always done in the current version. There is scope to discuss any obvious limitations of the software suite, e.g. can it deal with models on different grids, can it cope with reversal of order of the transforms (e.g. the transforms $\mathbf{U}_\mathrm{v}$ and $\mathbf{U}_\mathrm{h}$). These points are raised in more detail in section 2.

2 Specific comments

1. In Section 1 of the paper the flexible nature of the suite is discussed, especially that it allows input from a range of models. Firstly, what computational grid does the system use (e.g., Arakawa A, B, C, Lorenz, Charney-Phillips, an irregular grid, etc.)? How does the software deal with input data held on grids different to the one used?

2. I was wondering if the suite has the capability of dealing with the following:

   (a) In Section 2.2.2 the control variable transform is shown. If a user wishes to experiment with alternative orderings of the transforms, e.g. $\delta\mathbf{x} = \mathbf{S}\mathbf{U}_\mathrm{p}\mathbf{U}_\mathrm{h}\mathbf{U}_\mathrm{v}\mathbf{u}$ instead of $\delta\mathbf{x} = \mathbf{S}\mathbf{U}_\mathrm{p}\mathbf{U}_\mathrm{v}\mathbf{U}_\mathrm{h}\mathbf{u}$, is this possible?

   (b) Can the suite deal with other methods of modelling horizontal correlations such as spectral methods or diffusion operators?

   (c) Can the suite deal with dynamical balance operators instead of purely statistical ones?

   (d) Is the user tied to streamfunction, potential and temperature (e.g. vorticity/divergence/pressure/PV, etc. might be desired)?

   (e) What about background errors that are distributed in time as might be used in weak constraint 4D-Var (so that $\delta\mathbf{x}$ and $\mathbf{u}$ are 4-D fields instead of 3-D)?

(f) Note: I realise that the above might be beyond the objectives of GEN_BE v2.0, but these issues are worth noting (at least to state the limitations of the software and/or any future development work that might be planned).

3. P.8, 2nd bullet: an ensemble can be worse than NMC due to incorrect spread of the ensemble.

4. Section 3.1.1 in general: It's not immediately clear what distinguishes stages 0 and 1. Fig. 2 says that stage 0 computes error perturbations and stage 1remove mean. This should be made absolutely clear in the text.

5. P.10: "... and then directly calculates the regression coefficient as a product" - a product of what with what?

6. Equation 6: I would say that this equation comes from the finite difference formula rather than "Taylor development". It also relies on symmetry of ρ about the origin.

7. Last equation on P.11: presumably $\rho(\mathbf{x})$ should be $\rho(\delta\mathbf{x})$, and the $L_{\rm vg}$ should be $L_{\rm vg}^2$.

8. Equation 6 is just the previous equation rearranged.

9. P. 12: last para.: the EOF representation of the vertical covariance matrix is exact if all EOFs are used (they will include inhomogeneity e.g.).

10. P.12, L.22: "sparse repartition" - what does this mean?

11. Equation 8: what is the significance of the factor of 8 in the denominator?

12. Equation 10: shouldn't the right hand side be square-rooted?

13. P.15, L.16: "precising"?

14. Notes on figures that involve model level: where does the boundary layer top and tropopause relate to the model levels? What is the data used to compute them (ensemble? NMC? time period?)

15. P.16, L.19: What is nebulosity?

16. P.16, L.21 and P.17, L.9: what is the relative humidity rate?

17. P. 17, description of covar6: there seems to be 0, 1, and 2, meaning 'no regression', 'full regression' and 'diagonal only'. The last one I assumed to be the meaning. The key should be pointed out explicitly, perhaps in one of the tables. Are there any other options beyond 2?

18. P.21, 1st para.: This needs to be written in a more lucid style.

19. Fig. 14 (pressure vs. level) is out of place - this should be placed earlier in the paper (e.g. immediately after Fig. 3).

20. It would be useful to show statistics that come directly from the sample (i.e. not the statistics implied by the transforms) for comparison with Figs. 11-13. Also on these Figs. please include axis labels.

21. P.23, L.24: The text refers to 1-D variance in connection with Fig. 18a, but this Fig. looks to me like a slice through 2-D data.

22. P.24, L.23: "Methods that combine general statistics of the background errors and local balance are found to perform better when the ensemble size is small." As it reads this statement says that these hybrid techniques do better with smaller ensembles than with larger ensembles. Is this what the authors want to say? Do they mean, "When the ensemble size is small, ,methods that combine general statistics of the background errors and local balance are found to perform better".

23. P.25, L.21 onwards: The authors talk about the ensemble of the day. Does this mean that regression coefficients have to be recalculated each time as new ensemble is used?

24. Table 4 (and text that refers to this table): Given that $q_{\text{cloud}}$and q_{ice} are two separate variables, does this mean that the former refers only to liquid water?

25. Please include in each Fig. that shows statistics the following: the source of the data used, its type (e.g. the model and ensemble or NMC), the averaging done (e.g. over how many days), and the modelling method used to produce the plot (e.g. EOFs or recursive filters).

3 Technical corrections

Things crossed out should be removed and things in italics should be added.

1. P.3, L.4: "~~underlaying~~ *underlying*".
2. P.3, L.23: "~~this~~ *these*".
3. P.4, L.10: please define "CONUS" - this might not be known outside of N.America.
4. P.4, L.13: "~~Test~~ *Testbed*".
5. P.5, L.1: "~~mapper~~ *map*".
6. P.5, L.6: "~~In general~~ *Often*".
7. P.5, L.10: "B matrix~~,~~ *is* comprised".
8. P.6, L.20: "application~~s~~ of *a* recursive filter".
9. P.7, L.8: "modeling *of* a".
10. P.7, L.19: "In ~~the~~ version 2.0".
11. P.10, L17: "balanced part ~~for~~ *from* each other variable".
12. P.10, L19: "bloc*k*".
13. Equation 5: It is unusual that ∇^2 is used here for a 1-D vertical derivative. It would be more informative to have $\partial^2\rho/\partial z^2$ (or whatever the derivative is respect to).
14. P.11, L.15: "~~subsisting~~ *substituting*".
15. P.12, L.16: "~~grib~~ *grid*".
16. P.12, L.19 and P.19, L.18: "potential velocity" change to "velocity potential".
17. P.12, L.25: "plan*e*".
18. P.13, L.1: "point*s*".
19. P.13, Ls.7 and 12: "~~the~~".
20. P.14, L.16: "control~~s~~ variables".
21. P.15, L.25: "~~with~~ *and*".
22. P.16, L.18: "~~statically~~ *statistically*".
23. P.18, first para.: add "the" before each compiler name.
24. P.19, Ls. 1 and 3: "~~of~~ resolution".
25. P.20, L.26: "... horizontal slice ~~done~~ at the 500hPa *level* for ~~the~~ temperature ...".
26. P.21, L.2: "*is* employed *in* the".
27. P.21, L.11: "... is spread~~ed~~ out by the ~~recursive filter~~ *EOF decomposition* ...". [At least the caption of Fig. 11 refers to EOF.]
28. P.21, L.20: "~~This~~ *These* ~~kind of~~ *ensemble-based* background error statistics ~~has~~ *have* potentially more skill~~s~~ ...".
29. P.22, L.4: "Modifications code" change to "Code modifications".
30. P.23, L.26: "It is *used* most of the time ~~used~~ ...".
31. P.24, L.7: "... the possibility *of adding* ~~to add~~ clouds in the dry area or *of removing* ~~to remove~~ clouds in the ...".
32. P.24, L.9: "... ~~including~~ *the inclusion of* ...".
33. P.26, L.14: "~~amount~~ *number*".
34. P.26, L.19: The Barré et al. reference does not appear in the reference list.
35. P.26, L.26: "aerosol~~s~~".
36. P.27, L.1: "~~the~~ optical depth".
37. Table 1: "Allow*s* GEN_BE to read ...".
38. Table 1: "... historical ~~date~~ *data* available, defined in hour*s* ...".
39. Table 2: "... hgt define~~d~~ the width ...".
40. Table 2 (two occurrences): "level model" change to "model level".
41. Table 3: In the first item in this table, should "1-8" be "0-7"?
42. Table 6: "First variable ~~do~~ *does* not ...".
43. Table 7, last three rows: is there a reason for upper case in "Cov"?
44. Fig. 16: The caption and Fig. itself don't match (horizontal shown instead of vertical).