General comments on corrections done

We would like to thank all five reviewers for their valuable input. A major revision of the structure of the document has been done as suggested by most. The technical details related to the code were moved to the Appendix.

Also, their remarks lead to the rephrasing of three paragraphs and the discussion part in the conclusions was extended. More details have been provided on the setup of the experiments with different modeling of B. Some plots and captions have been corrected and completed. The first part of this document gives an overview of the modifications done, followed by the answers to the comments of each reviewer.

(1) Modifications of the structure of the document.

Reviewers asked modifications of the structure of the document:

In Sect. 2.0, Sect. 2.2.3 doesn't exist anymore.

As asked by different reviewers, the technical details of Sect 3. has been moved in Appendices. Thus, Sect 3. is renamed "Five stages to generate the background error covariance statistics (GEN_BE code version 2.0). and subsections from 3.1.1 to 3.1.4 renumbered from 3.1 to 3.4.

The previous Sect. 3.2 does not exist anymore:

- Sect. 3.2.1 have been included in the Appendix A (FORTRAN code and input/output description)
- The first part of the Sect. 3.2.2 has been merged to the new Sect 3.2.
- Section 3.2.3 becomes Appendix C (Installation, compilation, set up and visualization).
- The description of the namelist options goes in Appendix B (Description of the namelist options)

Section 5.0 includes now the results related to chemistry data assimilation previously shown in Appendix A. Sect. 5.0 is renamed "Cloud and chemistry variational data assimilation"

- Sect. 5.1 is named "Generation of a multivariate background error covariance for hydrometeors.
- Sect. 5.1.1 is added and is composed by the part related to the balance operator previously presented in Sect. 3.2.2. Section 5.1.1 is named "Generation of a multivariate background error covariance for hydrometeors.
- Previous Sect 5.1, and 5.2 becomes 5.1.2 and 5.1.3
- Previous Appendix B becomes Sect. 5.2 and is named "Background Error for Chemical Species"

(2) Modification Equations

Some Equations has been corrected, added and renumbered. We give an update below of the different modifications done.

- Eq. (1) J_b and J_o terms are added
- Eq. (2) new equation added to present a general definition of B
- Eq. (3) the definition of $\delta x=(x_b-x)$ added and renumbered
- Eq. (4) renumbered
- Eq. (5) $B^{1/2}$ is presented instead of δx
- Eq. (6) new equation to present the calculation of the regression coefficient.
- Eq. (7) presents of the calculation of the unbalanced part of the perturbations δt_u
- Eq. (8a) presentation of the Daley's formula that define the vertical length scale for one dimension along the vertical (z).
- Eq. (8b) presentation of an approximation of the formula of Daley along the vertical
- Eq. (9a) presentation of the Gaussian formula that define the vertical length scale for one dimension along the vertical (z).
- Eq. (9b) inverted expression of 9(a)
- Eq. (10a) corrected and renumbered
- Eq. (10b) corrected and renumbered
- Eq. (11) corrected and renumbered
- Eq. (12a-c) Identical

(3) Modification Figures

Previous Fig. 14, that shows the distribution of the vertical model level in function of pressure level, is presented earlier in the document (in the first paragraph of section 3.0 and becomes Fig. 3).
It allows visualizing the density of the vertical model in function of pressure and switch from vertical model level to pressure accurately when results are presented in sect 3.0, 4.0 and 5.1.

Fig. 9, 10,11,12,13,15,16,18a added right vertical axis in hPa pressure levels.

(4) Modification Tables

Table are renumbered:
Table 4 becomes Table 1
Table 2 is created to gather the setup information about the different modeling of B.
The other Tables are moved into the appendix:
- Previous Tables B1, B2, and B3 become Tables A1, A2 and A3.

- Previous Tables 1, 2, 3, 6, 7 and 5 become respectively Tables B1, B2, B3, B4, B5, and B6.

(5) Major revision in the text

Description of the experiments:

(a) The description of the D-ensemble dataset (50 members over the CONUS domain) coming from DART is done in the second paragraph of Sect. 3. : "Figures shown in … Romine et al. (2014) to generate the ensemble and … Table contains detailed information setup of the data assimilation experiment."

Reference about DC3 experiment of Romine et al. 2012 is replaced by:
Romine G., S., Schwartz C., S., Berner J., Fossell, R., K., Snyder C., Anderson J. and Weisman M., L.: Representing forecast error in a convection-permitting ensemble system, Mon. Weather Rev., doi: http://dx.doi.org/10.1175/MWR-D-14-00100.1, 2014.

(b) A new table 2 is presented Section 4.0, to give details about the benchmark performed.

Table 2: Description of the setup of the background error matrix modeling diagnosed over the CONUS Domain. $\mathbf{B}_{eof}$ and $\mathbf{B}_{rcf}$ are diagnosed using GEN_BE code version 2.0 and the D-Ensemble method while $\mathbf{B}_{nam}$ is performed by NCEP using the NMC method.

Paragraphs rephrased:

(a) In the introduction, the first paragraph has been corrected, the second and the third rephrased following the remarks of the different reviewers.
(b) Section 2.2.2, the order of the description of the different transform match the Eq. 5:
  - The $\mathbf{U}_p$ matrix, called physical transform or balance operator, …
  - The $\mathbf{S}$ matrix is …
  - The $\mathbf{U}_v$ matrix, called vertical transform, …
  - The $\mathbf{U}_h$ matrix, called horizontal transform, …
(c) First paragraph of Section 3.0 has been rephrased.
(d) Section 3.2 has been rephrased (merge of previous sections).
(e) First paragraph of Sect 4.0 is rephrased and additional information is given to the general setup of the different modeling of B ($\mathbf{B}_{eof}$, $\mathbf{B}_{rcf}$ and $\mathbf{B}_{nam}$). References have been added: Romine et al. 2014, Rogers et al. 2009 and Wu 2005.
(f) Section 4.2 has been rephrased
(g) Section 5.1.1 coming from the previous Sect. 3.2.2 is partially rephrased to become independent.

(h) The discussion has been extended in Section 6, which is partially rephrased.

(6) Direct answers are given on the different referee below, in the following document.

Corrections Referee 2: 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 rep- resent 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 Uv and Uh). These points are raised in more detail in section 2.

Answers:

We want to thank R. Bannister for his review and advice to improve the discussion about the limitations and future of the GEN_BE code version 2.0, and the presentation of the document.

- *The figures that show statistics contained additional information: the source of the data used (WRF-ARW/WRF-NMM/WRF-CHEM), its type (D-ensemble or NMC), the averaging done (over how many days), and the modelling method used to produce the plot (e.g. EOFs or RFs)*

- *In Sect. 6.0, one paragraph has been added to discuss obvious limitations of the software and suite: it deals with models on different grids.*

- *Corrections has been done following the remarks mentioned in the specific comments*

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?

The answer is contained in a paragraph added in Sect 6.0:
"In these previous examples, GEN_BE code version 2.0 can handle input datasets coming from WRF, a model defined on a C-Arakawa grid, and the background error statistic outputs are computed on unstaggered A-Arakawa grid. Within minor modifications, the code would be able to handle other horizontal grids. Also, statitics could easily be done on models with different vertical grid definition. If we consider performing the background errors statistics on an unstructured grid, the structure of the code can remain the same but few mathematical operators, such as differential and laplacian, and estimation of the distance between two grid points, would need to be re-defined according to the grid. In fact, the Up transform needs to be performed in the unstructured grid according to the user's choice of control variables. Uv transform will remain identical and Uh transform would be modified according to the mathematical operators. Another option would be to interpolate first the input dataset on a regular grid according to the data assimilation system used and then compute the statistics. Thus, implementation of models with different grid can be done in the GEN_BE v2.0 code based on its general framework and may be completed by adding new diagnostics."

We are currently implementing a Model with an unstructured grid (MPAS-GLOBAL) into GSI. We performed the data assimilation process interpolating the meteorological field on a unstaggered A-Arakawa Gaussian grid. In this case, the calculations to perform the statistics that model B using GEN_BE are straightforward.

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. $\delta x = SUpUhUvu$ instead of $\delta x = SUpUvUhu$, is this possible?
The actual order of the transforms are $\delta x = UpSUvUhu$ *(equation 4 is* $UpSUvUhu$ *and not* $SUpUhUvu$ *as previously written, see modification p4296, Eq 4). It is possible for a*

user to invert the order of the Uh and Uv transforms. (Wlasak and Cullen 2014 study the impact of the defined order of vertical and horizontal operators).

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

The original version GEN_BE V1.0 contained spectral method for global applications. The code can be provided and minor update need to be according to the new framework. We can provide it as it is if necessary.

We have no particular plan to use diffusion operators yet.

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

We do not have plan in next future to use balance operators. Variational-Ensemble hybrid methods will be used to add some flow dependence in the background error covariances.

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

The GEN_BE V.2 code has been designed to handle various control variables defined by the user (Table 4). Vorticity/divergence are already implemented. Adding new control variables is a minor development. In this case, the user has to define in the module io_input_model.f90 how to compute this variable according to the model variables. All the stages of GEN_BE can be applied directly and the user can define linear correlated errors with other control variable via the namelist.input file.

(e) What about background errors that are distributed in time as might be used in weak constraint 4D-Var (so that δx and u are 4-D fields instead of 3-D)?

For example, it would be possible to apply GEN_BE on the tendency of the fields too to have an estimate.

(f) Note: I realise that the above might be beyond the objectives of GEN_BE v2.0, but these issues are worth nothing (at least to state the limitations of the software and/or any future development work that might be planned).

We have extended the discussion part in Sect. 6 including some answers of the previous remarks

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

We agree: an incorrect low spread of the ensemble method will give to much confidence on the background error covariance and likely less observations would be assimilated. Adaptative inflation is used in the DART-EAKF D-ensemble. Moreover, the spread and the correlations in the NMC method, may be inappropriate specially in the area poorly observed.

Also, we add the reference Fisher (2003) in this paragraph and refer about some drawbacks of the D-ensemble method in the sentence: "However, more computational resources are required to run an ensemble simulation and it may not provide automatically the optimum B for a particular system (Fisher 2003)."

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 1 remove mean. This should be made absolutely clear in the text.
An other reviewer asked us to introduce clearly the goal of each stage before going further. It has been done before section 3.1 writing "Stage 1 removes the mean of these perturbations and defined the binning applied."

5. P.10: ... and then directly calculates the regression coefficient as a product - a product of what with what?
Section 3.2, Eq. 6 has been added and commented.

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.

After introducing the Daley's formula, previous version has been changed by:
"Approximating Eq (7a) with finite difference to the second order derivatives of $\rho(\delta z)$ *and assuming ρ symmetric around the origin results in:"*

7. Last equation on P.11: presumably ρ(x) should be ρ(δx), and the Lvg should be L2vg.
Corrected in the text

8. Equation 6 is just the previous equation rearranged.
(see correction 6)

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?
The sentence :
"Hydrometeors mixing ratio show even more local structures due to their sparse repartition on the horizontal and the vertical" is replaced by
"Hydrometeors mixing ratio show even more local structures due to sparse locations"

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

The formula has been replaced by $\rho(r) = \exp(-\frac{r^2}{2L})$

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

13. P.15, L.16: precising ?
This sentence has been rephrased in the text: "The parameters equal to ... by subtracting their balanced part coming from the stream function (psi).

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?).
Table 2, presented section 4.0, has been created to gather the information of the set up of the different modeling of B. Also, we add some information at the end of captions to reference the method used:
According to the renumbered figures:
Fig3: added "(WRF, Res. 15 km)."
Fig4: added "(WRF, Res. 15 km, D-ensemble)"
Fig5: added "(WRF, Res. 15 km, D-ensemble)."
Fig6: added "(WRF, Res. 15 km, D-ensemble, EOFs)"
Fig7: added "(WRF, Res. 15 km, D-ensemble, EOFs)"
Fig8: added "(WRF, Res. 15 km, D-ensemble, EOFs)."
Fig9: added "(WRF, Res. 15 km, D-ensemble, RFs)."
Fig10: added "(WRF, Res. 15 km, D-ensemble, RFs)."
Fig11: added "(Beof: WRF Res. 15 km, D-ensemble, EOFs)"
Fig 12: added "(Brcf: WRF Res. 15 km, D-ensemble, RFs)."
Fig 13: added "(Bnam:WRF-NMM Res. 12 km, NMC, RFs)."
Fig 14: added "(WRF, Res. 15 km, D-ensemble)."
Fig 15: added "(WRF, Res. 15 km, D-ensemble)"
Fig 16: added "(WRF, Res. 15 km, D-ensemble)"
Fig 17: added "(WRF, Res. 15 km, D-ensemble, RFs)"
Fig 18: added "(WRF, Res. 15 km, D-ensemble)"
Fig19: added "(WRF-CHEM, Res. 36 km, D-ensemble)"
Fig20: added "(WRF-CHEM, Res. 36 km, D-ensemble)"
Fig21: added "(WRF-CHEM, Res. 36 km, D-ensemble)"

15. P.16, L.19: What is nebulosity?
Nebulosity is a parameter diagnosed to define the presence of cloud in the low atmosphere. In the study of Ménétrier et al. 2011, the authors estimate the nebulosity based on the cloud fraction calculation of AROME coming from the first vertical model levels.

16. P.16, L.21 and P.17, L.9: what is the relative humidity rate?
P16 L21
In the new paragraph, "relative humidity rate" is replaced by "humidity"

P.17, L.9 Replaced by “as their presence or absence is directly related to the humidity rate” by
“as their presence or absence is directly related”

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?
No other option beyond 2 are available at present.
Table 6: added “At present, the parameter covar can take three values: 0, 1, and 2, meaning “no regression”, “full regression” and “diagonal only”. ”

18. P.21, 1st para.: This needs to be written in a more lucid style.
The paragraph has been rephrased.

19. Fig. 14 (pressure vs. level) is out of place - this should be placed earlier in the paper (e.g. immediately after Fig. 3).
This figure is now presented just after the figure of the CONUS domain.

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.
Some information have been added to compare horizontal length scale coming from NAM in Sect 4.1.2 : “Direct comparison of … compare statistics from forecast of different length.”
We redid the Figs (11-13) adding the axis.

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.
The order of the figures has been inverted. The legend is replaced by:
“(a) Profile of standard deviation of liquid water condensate mixing ratio (qcloud in g kg-1) averaged along the vertical and (b) horizontal cross-section of standard deviation of qcloud at the vertical model level 5 (950 hPa). Both plots indicate the presence of low maritime clouds noted by high standard deviation.”

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.
The sentence “Methods that combine general statistics of the background errors and local balance are found to perform better when the ensemble size is small”

is replaced by "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 coeffcients have to be recalculated each time as new ensemble is used?
Only the geographical mask could be updated and the variance 3-D. If the goals of DA and the meteorological situation remain the same (same specific event) for the next cycle of analysis, we can keep these regression coefficients. Moreover, it should be possible to re-computed every cycle these coefficients are they are not in the critical path for operational center: these statistics are not so CPU time expansive and can be launched as soon the forecast is available.

In Sect. 6 renamed summary and discussions the following sentences have been added: "The regression coefficients calculated, can be conserved for a next cycle analysis as they are averaged by bins or recalculated as they are not so expansive with regard to CPU (central processing unit) time."

24. Table 4 (and text that refers to this table): Given that qcloud and qice are two separate variables, does this mean that the former refers only to liquid water?
In table 4, I replaced "Cloud mixing ratio" by "Cloud water mixing ratio" and "Rain mixing ratio by "Rain water mixing ratio"

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).
Done and show in remark (14)

3 Technical corrections Things crossed out should be removed and things

1. P.3, L.4: *The sentence has been split and rephrase: "The probability errors are supposed to be normally distributed"*

2. P.3, L.23: this → these .
"these different efforts" replaced by "this different efforts"

3. P.4, L.10: please define CONUS - this might not be known outside of N.America.
Replaced by CONUS (CONtiguous United States) and moved in the first paragraph of Sect 3.0

4. P.4, L.13: Test→ Testbed .
Replaced

5. P.5, L.1: mapper → map.

Replaced

6. P.5, L.6: In general → Often .
Replaced

7. P.5, L.10: B matrix, is comprised.
Changed by Replaced "B matrix, being comprised of nearly 10^8 x 10^8 = 10^{16} entries, is too."

8. P.6, L.20: applications of a recursive filter.
Add a before recursive filter.

9. P.7, L.8: modeling of a .
The paragraph has been merged to another section and this sentence doesn't exist anymore.

10. P.7, L.19: In version 2.0 .
The all paragraph has been re-written.

11. P.10, L17: balanced part for from each other variable.
The sentence has been replaced by "Linear regressions are performed to derive uncorrelated (i.e. unbalanced) perturbations by removing the balanced part from each other perturbation variable"

12. P.10, L19: block.
Corrected to block

13. Equation 5: It is unusual that ∇2 is used here for a 1-D vertical derivative. It would be more informative to have ∂2ρ/∂z2 (or whatever the derivative is respect to).
We corrected the equation.

14. P.11, L.15: subsisting -> substituting .
The sentence has been rephrased (see point 6 of the specific comments)

15. P.12, L.16: grib-> grid .
Changed grib to grid

16. P.12, L.19 and P.19, L.18: potential velocity change to velocity potential .
Done

17. P.12, L.25: plane.
"Plan" changed to "plane"

18. P.13, L.1: points.

Point changed to points

19. P.13, Ls.7 and 12: the .
Removed "the" before Eq.

20. P.14, L.16: controls variables.
"controls variables" replaced by "control variables"

21. P.15, L.25: with and .
These sentences have been rephrased in the new section 3.2.

22. P.16, L.18: statically→statistically .
These sentences have been rephrased in the new section 3.2.

23. P.18, first para.: add the before each compiler name.
Done

24. P.19, Ls. 1 and 3: of resolution .
Replaced by "domain at 15 km resolution"

25. P.20, L.26: ... horizontal slice done at the 500hPa level for the temperature ... .
Replaced by "horizontal cross-section at the 500hPa level for temperature"

26. P.21, L.2: is employed in the .
The sentence is replaced by "When the operator (Uv) employs EOF decomposition, the Jb term of the cost function is weighted by the variance coming from the eigenvalues of Beof"

27. P.21, L.11: ... is spreaded out by the recursive filter EOF decomposition ... . [At least the caption of Fig. 11 refers to EOF.]
We removed it in the new paragraph.

28. P.21, L.20: This→ These
The all sentence has been replaced by "These ensemble based background error statistics have potentially more skill to estimate correlated errors related to the present meteorological event."

29. P.22, L.4: Modifiations code change to Code modifcations
Done

30. P.23, L.26: It is used most of the time used ... .
Sentence replaced by "It is most of the time used when the perturbations come from the NMC method or when the variance is not diagnosed at the analysis time".

31. P.24, L.7: ... the possibility of adding

The sentence has been replaced by :" The covariance between mixing ratio of cloud water condensate and relative humidity, described in Sect. 5.1.1, can reinforce the ability of adding clouds in the dry area or removing clouds in the cloudy area."

32. P.24, L.9: "is beneficial at the analysis time as it allows including increments of hydrometeors directly at the analysis time"
Replaced by "The univariate version of the balance operator for hydrometeors may be beneficial at the analysis time as hydrometeors can be directly assimilated"

33. P.26, L.14: amount → number .
Amount replaced by number

34. P.26, L.19: The Barré et al. reference does not appear in the reference list.
The reference has been added.

35. P.26, L.26: aerosols .
The reference of Benedetti and fisher 2007 and the sentence about aerosols has been removed.

36. P.27, L.1: the optical depth .
Sentence removed

37. Table 1: Allows GEN_BE to read ... .
The all sentence is corrected as follow "Set up the acronym for the model input allows GEN_BE to read different input model in the stage 0."

38. Table 1: ... historical date data available, defined in hours ... .
Replaced date available by "date data available"

39. Table 2: ... hgt defined the width ... .
I replaced "binwidth_hgt defined the width that splits the bins" by "binwidth_hgt define the width that splits the bins"

40. Table 2 (two occurrences): level model change to model level.
Line bin_type 3: I replaced "vertical level model" by "vertical model level"
Line bin_type 4: I replaced "model vertical level" by "vertical model level"
Line bin_type 5: I replaced "vertical level model" by "vertical model level"
Line_bin_type 7 : I replaced "vertical levels" by "vertical model level"

41. Table 3: In the first item in this table, should 1-8 be 0-7 ?
Yes, we corrected the line bin_type to 0.

42. Table 6:   First variable do →does not ... .
Line covar1, we replaced do by does.

43. Table 7, last three rows: is there a reason for upper case in Cov?
All has to be lower case. We corrected.

44. Fig. 16: The caption and Fig. itself don't match (horizontal shown instead of vertical)
It has been corrected.

Corrections Referee 4

General comments

This paper gives a detailed description of the GEN_BE 2.0 system including theoretical discussion, equations, options and the code. I found this paper is very useful for the readers who wants to learn and use this tool and for readers just want to learn more details on how to model a BE in variational analysis. Those details are necessary information when apply this tool in data assimilation but, on other hands, I think those details make this paper hard to follow and read through. My suggest is to do a major revision of the paper structure: put all the details on code and namelist options to the appendix as reference for readers who want to apply the tools but leave the theoretical analysis, practical discussions, and test results of modeling BE in the paper.

(1) Answers to the general comments

We want to thank Referee 4 for his or her valuable advice to modify the structure of the code and to improve the introduction of the manuscript.

A major revision of the structure of the paper has been done to present the GEN_BE code Version 2.0 and its application. The details on code and namelist options have been moved in 3 appendices. (See general comments at the beginning of the document)

(2) Answers to the detailed comments

1. Page2, Line14-16: the statement on a“multivariate approach” Is not clear. It can refer to adding new control variables for the cloud analysis, or to the GEN-BE for providing covariance among all the analysis variables.

It has been rephrased:
“Different choices of control variables and their correlated errors used to mimic general physical balance (geostrophic, hydrostatic, ...) in the atmosphere have been largely investigated by different operational centers and referenced in Banister (2008b). Since then, such multivariate relationship approaches has been studied …”

2. Page3, Line2: change “or the UK Met office ”to“ and the UK Met office”.
Done

3. Page3, Line3: add “,” after“ techniques”.
Done

4 Page3, Line6: change “that minimize” to “, to minimize”

This part of the sentence has been removed.
More details have been written in the paragraph 2.2.2 that describes Up transform.

5. Page3, Line9-12: More available observations are not the only reason why cloud and chemical data analysis are needed. I think the needs of improving the cloud forecast and chemical (pollution) forecast are major drivers of the development of the cloud and chemical data assimilation.

This part has been rephrased in the first paragraph of the introduction.

6. Page3, Line27:"the two first sections". This is confusion.
Each section is described separately now.

7. Page4, Line14: Please give more details on which kind of "results" author will give in the Appendix to give reader an idea what in the Appendix.

The document has been restructured to gather all the technical details in the appendices and are not anymore presented in the introduction. They are presented for the first time in the first paragraph sect. 3.0.

8. Page5, Line9: "nor be stored" changes to "and to"
Done

9. Page5, Line12: I think "parameterized" has the same meaning as "Modelling" in the next line. If this is the case, please use the "modeling" just as other part of the paper.
Replaced "parameterized" by "modeled"

10.Page 5, last line" "linear operator" changes to " linear observation operator".
Corrected, the sentence begins now as follow:
"H is the linearized observation operator …."

11.Page 6, equation 4. This equation can be expressed as square root of B equals to ...
replaced the equation 4 by

$$B^{1/2} = U_p S U_v U_h$$

12.Page 7, Line 7: "and make to" changes to " and to make"
This paragraph doesn't exist anymore.

13.Page 7, Line 8-10: “the new version ... a new model of B”. I don't understand what this sentence means.
This paragraph doesn’t exist anymore

14.Page 7, Line 20: add ‘the’ before ‘modeling’. The same line: add “background” before “error covariance” and change “become” to “becoming’.
The sentence is rephrased: “The five steps, from stage 0 to 4, that model a background error covariance matrix, become independent of the choice of control variables and model input, which allows for more flexibility (Fig. 2).”

15.Page 8 Line 4-8: please list the functions of each stage more clear the specific to help readers go through the details of each stage smoothly.

The first paragraph of section 3.0 has been rewritten to take into account this remark : “The general structure of the GEN_BE code … Appendix C explains how to compile and run the code.”

16.Page 8, Line 10-11: “sample of model forecasts” changes to “ sample of perturbations”
We changed to “sample of perturbations”.

17.Page 8, Line 20: “ an ensemble of ” changes to “ ensemble perturbations of”
We replaced “an ensemble of perturbations of previous forecasts valid at the same time” by “an ensemble of perturbations valid at the same time”.

18.Page 11, Line 2: “After”, should be “when”
Two successive steps are necessary to estimate the vertical auto-correlation parameters. First, the vertical auto-covariance matrix averaged by vertical levels is computed. Then, two different techniques can be used to diagonalize this matrix.

19.Page 12, Line 24: “we estimate length scales” means horizontal or vertical or both. Needs to clear define which part of length scales here and in other parts of the paper.
We replaced “In stage 4, we estimate length scales averaged” By “In stage 4, we estimate horizontal length scales in a 2-D plan defined by vertical level or EOF mode.”

20.Page 13, equation 8. Please define “r”
r has been defined in the sentence that follow the Equation:
“where $\rho(r)$ is the correlation calculated for a distance r between two grid points”

21.Page 13, the paragraph starts from “Usually, ...”: This is paragraph is very helpful for readers to understand the advantage and disadvantage of the each option in global_bin but it also mixed with too much details on exact number of the option. The other parts of the paper also have the same issue as I described in summary. Please think how to keep the useful discussion of the BE option in paper but leave the details to the appendix.

The last paragraph of section 3.4 has been rewritten: “The horizontal length scale ... normalization issues (Michel and Auligne 2010).”
The technical details about global_bin flag is a part of appendix B (namelist section “&gen_be_lenscale”).

22.Page 15, Line 12-15: please revise this sentence to make it easy read and understand.
The all section 3.2.2 has been merged to section 3.1.2 and rephrased.

23.Page 15, Line 25-26: please give more explanation on the purpose of showing the correlation of T with both specific humidity and relative humidity.
The all section 3.2.2 has been merged to section 3.1.2
A sentence have been added to explain the purpose of such diagnostic:” Diagnostics such as vertical cross-covariance (Fig. 4) or vertical cross-correlation are helpful to analyze the relationship between variables and can be done by using stage 2.”

24.Page 19, Line 2-3: RAP is not using NAM BE directly. The BE for RAP is a combination of the global BE and NAM BE with some tunings. Also, NAM BE should be 1 degree of the resolution instead of 0.1 degree.
The general setup is described in Table 2 (added): forecast used to construct B, in this case, is at a 12 km resolution. Statistics of B are also binned within a latitude band of 1 degree.

25.Page 20, last line: “ the pseudo observation of 1K”: should be “the innovation of 1K.
Done

26.Figure 1 caption: what is “DC3”? Please give explanation or delete it.
We removed it.

27.Figure 8 and 9 and 10 are whole domain results, right?
Yes, for the entire domain, by vertical levels or EOF modes.
It appears now in the new Table 2 that describes the different modeling of B

Corrections Referee 3

(1) General comments

This paper presents a code called GENerate the Background Errors version 2.0 (GEN_BE v2.0), which goes together with data assimilation tools using the Weather Research and Forecasting (WRF) model. It allows different modeling of the background error covariance matrix B. Those elements can be very useful and this should be published in GMD. However, this paper lacks of scientific (or technical and research reports) references for both the description of the system and results interpretation. I think this is an important issue that needs to be addressed, some elements can be found in both minor and major comments.

(2) Major comments

Firstly, the main objectives and methods need to be clarifying in the introduction section. The authors present a "D-Ensemble" in the introduction. It seems that it is used as a benchmark to be compare to. It is important to present how this setup is different to other simulations conducted using the variational approach and in different way of modeling B. Please provide some insight on first of the data assimilation setup, which and how many observations are effectively assimilated (even if it is only one), the analysis period and what are the differences in model simulations (grid spacing, boundary conditions, physics). This can be done using a table that summarizes the different experiments and set-up. Figures about horizontal and vertical domain dimensions can be presented at the same time. Also, the authors might consider some minor reorganization of the paper as indicated below:

Present a brief review of the 5 stages (Sect. 3) and the code structure in the introduction. I suggest to move all the code/software aspects in an appendix, especially "Sect. 3.2.1 FORTRAN code and input/output" and "3.2.3 Installation, compilation, set up and visualization". This appendix must expose the general structure of the code (options, names of files . . .). It must be an intermediate between results and methods exposed in the main text (add references to the appendix if necessary) and the code in the supplementary material.

I think the appendix on chemistry can be really interesting for that community, and those results should be presented in the main text if it goes with appropriate references (in addition to the one already presented in the paper). First, reviews of chemical data assimilation applied to air quality modeling can be found in Carmichael et al. (2008), Sandu and Chai (2011), and Lahoz et al. (2007) for stratospheric application. Then, you should refer to publications where assimilation has been done using WRF/CHEM and GSI or a 3D-Var (e.g. Pagowski et al. 2010, Schwartz et al. 2012, Li et al. 2013, Pagowski et al. 2014). Finally, some papers gives some estimation of similar quantities presented in figure A1, A2, and A3, such as errors length scale and variances (e.g. Constantinescu et al. 2007, Schwinger and Elbern 2010, Jaumouillé et al. 2012, Gaubert et al. 2014, Robichaud and Ménart 2014).

(3) Answer to major comments

We want to thank Referee 3 for the numerous remarks that lead to major revision in the structure and presentation of the document.

- A table is added Sect 4 to present the different setup that lead to different modeling of model background error in the benchmark.

- We followed the advices to move the technical details in appendix.

- We agree with the reviewer and we now moved the appendix A in the main text. The chemical section has been now improved: and refer to various publications that pointed out the need of different characterization of the BECM of data assimilation in atmospheric chemistry.

(4) Answers to minor comments

P4293 L3-4: This sentence is not clear to me: "assuming that the underlaying probabil- ity errors are normally distributed", I would suggest "assuming that errors are normally distributed".

The sentence has been split and replaced by: "The probability errors are supposed to be normally distributed and B is determined for a limited set of variables, called control variables."

P4293 L5-7: "that minimize the error covariance between variables", which variables is it, the control variable? The verb "minimize" is confusing since this "determination" is done a priori.

This sentence has been removed from the introduction and more details are written in section 2.2.2 in the paragraph for Up.
Also, some Diagnostics exist, as explained in section 3.2, (vertical cross-correlation) to estimate the error correlated between variables a priori.

P4293 L19: Please provide a reference for WRF, UM and WRFDA as it is done for GSI.
L19: added reference (UM, Davies et al., 2005)
L19: added reference (WRF, Skamarock et al., 2008)
L21: added reference (WRFDA, Barker et al., 2012)

P4294 L13 Can you add an appropriate reference for DART, e.g. Anderson et al. (2009). Ensemble Kalman filters for large geophysical applications
Done

P4295 L2: "to the irregurlarly distributed observation locations", you can remove "irre- gurlarly distributed".

It has been removed.

P4295 L3: Note that the exact knowledge of R and B would theoretically require the knowledge of the true state of the atmosphere. . .”. I would say “By definition, exact values of B and R would requires the knowledge. . .”.

Replaced by “By definition, exact values of R and B would require the knowledge”

P4295 L7: “i.e. uncorrelated observations, . . .”, you can say “i.e. uncorrelated observation errors” or “i.e. observations are assumed to be independent, . . .”.

Corrected by “i.e. uncorrelated observation errors”

P4295 L9 to 15: This paragraph should be more detailed and presented before the description of the different section. “All the results presented in the different sections were obtain from a numerical experiment with the WRF model”, this statement does not seems to be true, see for example in Sect. 5.2 (P4313, L17): “. . .we conducted a series of tests in which pseudo-observations of hydrometeors were assimilated into WRF-DA. . .”.

The modeling of B are based on datasets coming from WRF and WRF-CHEM forecasts and the data assimilation system used to test B are WRFDA and GSI.
This sentence doesn’t exist anymore in the introduction.

P4295 L23: “using a non linear observation operator. H is the tangent linear operator”. Please clarify, H is the observation operator and can be linear.

Replaced by : “ … using a non linear observation operator H. H is the linearized observation operator which makes the cost function quadratic and easier to minimize.”

P4297 L8-9: “The new version of the code allows modeling a real time configuration of B like NCEP does using five control variables”. Can you clarify what do you mean by real time configuration”? Please remove “like NCEP does” or give a reference.

In the GSI code developed at NCEP, the stream function, velocity potential, temperature, surface pressure and normalized relative humidity are the variable used. Kleist et al. (2009) used these controls variables for the Global Forecast System (GFS).
This section doesn’t exist anymore.

P4297 L14: “statitics of chemistry species to model B”, it is “error statistics of chemical species needed to model B”. “The community system Data Assimilation Research Test (DART)”, it is “Data Assimilation Research Testbed”

“statitics of chemistry species to model B” replaced by “error statistics of chemical species needed to model B”.
The reference to DART has been corrected to “Data Assimilation Research Testbed” in the first paragraph of Sect 3.

P4297 L24: “The version 2.0 of the code includes more physics options and flexibility has been added making all the algorithm in the different stages independent of the choice of control variable and model input”. This sentence is not clear and should be split like: “The version 2.0 of the code includes more physics options. In addition, the use of different stages, independent of control variable and model input allows more flexibility.

The sentence has been replaced by “The five steps, from stage 0 to 4, that model a background error covariance matrix, become independent of the choice of control variables and model input, which allows for more flexibility.”

P4298 L5: please correct “proxi”.
The all paragraph has been rephrased. The word “proxi” is not used anymore.

P4298 L20: Add some references (e.g. fisher 2003, Pereira et al. 2006).
Fisher, M., 2003: Background error covariance modelling. Proceedings of the ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean, 8-12 September 2003, 45-63.
The references are added.

P4299 L28 to P4300 L5: “Stage 1 creates the NetCDF file bin.nc ... module io_input.f90” Please move theses sentences in an appendix dedicated to the code description.
It has been moved in Appendix B.

P4300 L12 “The NCEP method”, please provide a reference.
Section 3.2 has been rephrased and the explanation is a part of the appendix B now: “Furthermore, when the regression coefficients … the io_output_applications.f90 Fortran module”.
No specific reference has been found on the filtering applied.

P4302 L10-L15 Please rephrase and move the algorithm description in an appendix dedicated to the code description.

It has been rephrased as follow:” The last paragraph of section 3.4 has been rewritten: “The horizontal length scale ... normalization issues (Michel and Auligne 2010).”The technical part about global_bin flag is a part of appendix B (namelist section “&gen_be_lenscale”).”

P4302 L18: Can you provide values in their unit and the level in parentheses?

I replaced “of the WRF computational domain at level 5 (~500m above the ground)” by “of the WRF computational domain around 500 m above the ground (model level 5)”

P4303 L6: Please rephrase “horizfunc = gauss”, put this options in parentheses.
P4303 L10: Idem.

“The first method (ls_method=1) employs a distribution function to fit the correlation for a 2-D field by vertical level or by EOF mode as defined in Sect.3.3 If a Gaussian function is chosen, the length scale is determined by solving Eq. (10a):

…
where $\rho(r)$ is the correlation calculated for a distance r between two grid points.
If a second order autoregressive (SOAR) correlation function is used, the length scale L is determined by solving Eq. (10b):"

P4304 L1 to L7: Please avoid the use of codes variables. You need to clarify the different available options.
This paragraph has been rephrase and details about the different available options moved in Appendix B.

P4304-4305 Sect. 3.2.1: please move this paragraph dedicated to the code description.
This paragraph has been moved in Appendix A.

P4305 L3: Change "variational a data assimilation" to "a variational data assimilation.
Done.

P4305 L12: "For example, NCEP operates", please give a reference.
The all paragraph has been rephrased (merge with other parts) and this sentence doesn't exist anymore.

P4305 L15-19: Please clarify and describe code options in parentheses.
Done.

P4306 L22: "A univariate version . . .", please provide a reference for this statement.
The paragraph has been rephrased and this sentence doesn't exist anymore.

P4308 Sect. 3.2.3: Please move that section in an appendix dedicated to the code description.
This section is referenced as Appendix C.

P4309 L3: can you describe the NAM acronym.
North American Mesoscale, done in the first paragraph of Sect. 4."

P4309 L12: "The first five eigenvectors are shown Fig. 6", "are shown in Fig. 6".
Done.

P4310 L9: Can you indicate distances first and the grid point in parentheses?

We replaced by "150 km (10 grid points) for all the vertical model levels, while the length scales of temperature and relative humidity remain in a range of 15 km to 30 km (1 to 2 grid points) below 200 hPa level."

P4310 L15: "parabolic approximation Eq. (6)." Please rephrase.

We replaced "the formula of Daley (1991, p110) and using the parabolic approximation Eq. (6)" By "coming from Eq. (8b)".

P4310 L24: “and the observation error of 1K.” Is it not “an observation error of 1K”?

Replaced by “a pseudo observation test of temperature with an innovation and an obervation error of 1 Kelvin”.

P4311 L9: “Bnam matrix coming from NAM”. Can you specify, like “constructed from NAM forecast error statistics”.
The all paragraph has been rephrased. We took into account this remark.

P4312 L12: The Fig. 14 should have been presented before.
This figure is presented now just after the figure of the CONUS domain.

P4312 L20: Please indicate the distance in km first.

This remark is applied:
P4309 L19 replace “39 grid points for the first EOF mode, i.e. close to 600 km”
By “600km (39 grid points) for the first EOF mode ”
P4309 L21 replace from 9 to 2 grid points
By from 135 km to 30 km (9 to 2 grid points)
P4309 L23 replace 15 km by 15 km (1 grid point)
P4310 L7 replace “above 150 km for all the vertical model levels, while the length scales of temperature and relative humidity remain in a range of one to two grid points under 200 hPa (i.e. 15 km and 30 km).”
By “above 150 km (10 grid points) for all the vertical model levels, while the length scales of temperature and relative humidity remain in a range of 15 km to 30 km (one to two grid points) below 200 hPa level.”
150 km by 150 km (10 grid points).
P4312 L20: The sentence is rephrased “…(less than 30 km, 2 grid points).”
P4312 L21: replace “one grid point (15 km)” by “15 km (1 grid point)”

P4314 L15: Define the acronym NWP.
It has been defined now in the first paragraph Sect. 3 by “Numerical Weather Prediction (NWP)”

P4314 L17-18: Give references about Meteo-France and the Met-Office system.
We added
“such as Météo-France with the Application of Research to Operations at Mesoscale system (AROME, Seity et al., 2011) and the Met Office with the Met Office Global and Regional Ensemble Prediction System (MOGREPS, Bowler et al., 2008; Migliorini et al., 2011)”

P4315 L28: “even if data assimilation of chemical species and aerosols remains difficult due to strong non-linearities”. This statement is imprecise and needs referencing.
We now removed this statement from the text.

Table B2: Last row, last column, ‘readble’, you mean readable.
Replaced

Figure 1: the last word is statistics.
Corrected

Figure 3: ‘Sparecly’. The sentence is not clear.

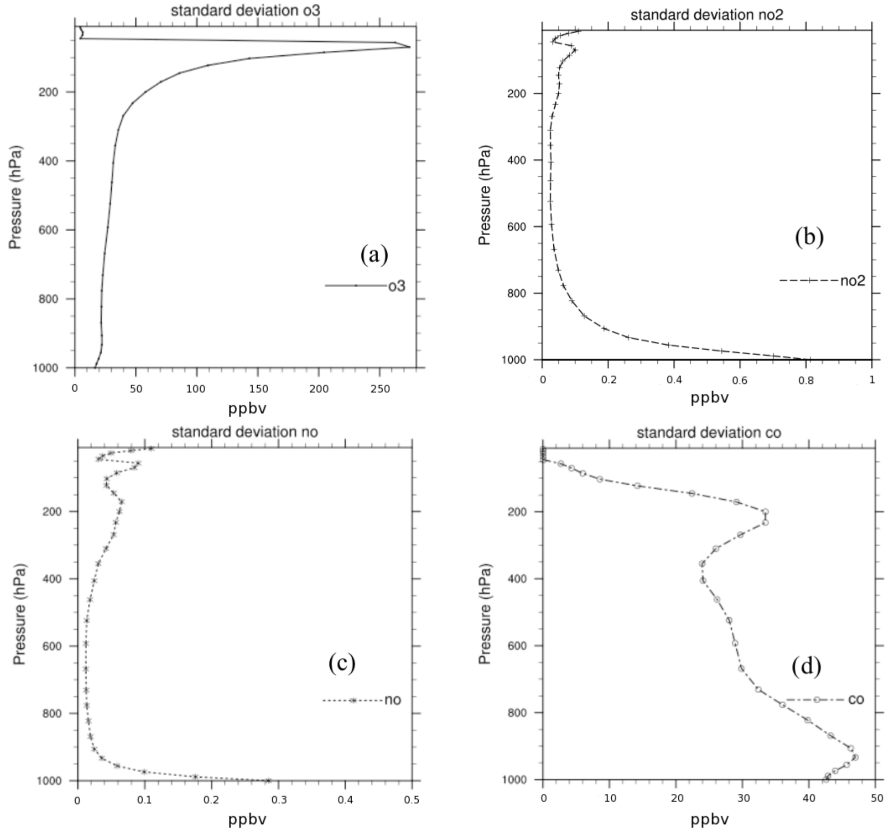
The all caption have been replaced by “Horizontal autocorrelation performed at the center of each square grid over vertical model level 5, around 950 hPa, for the control variables (a) stream function (psi), (b) temperature (t), (c) relative humidity (rh), and (d) Cloud mixing ratio (qcloud). Larger correlations are observed for stream function compared to temperature and relative humidity. Cloud mixing ratio has the smallest correlation due to sparce distribution of hydrometeors”

Figure 6-8-9-10-16-A2-A3: Can you indicate (remind), at least an approximate value, how much distance is representing by a grid point.

We did it in the caption of the figures.

Figure A1, can you redo this figure using ppbv instead of ppmv, especially for the NOx.

Done



The model resolution is 36 km and is already mentioned in the text. We then add this information in the figure caption for more clarity.

Corrections Referee 5

Comments on “Generalized Background Error covariance matrix model (GEN_BE v2.0)” by G. Descombes, T. Auligné, F. Vandenberghe, and D. M. Barker

The paper ‘Generalized Background Error covariance model (GEN_BE v2.0)’ presents a tool for the diagnosis of the background error covariance matrix for meteorological and atmospheric chemistry data assimilation applications. The code is based on existing techniques and does not present novel algorithms. However, GEN_BE v2.0 is of potential interest for many researchers in the field of geophysical data assimilation and the presentation is supported by several examples of scientific interest.

The paper lacks of scientific rigour in some sections, the structure is not optimal and it contains multiple language mistakes or approximations. Therefore, I recommend a major revision prior to publication in GMD. The main comments are detailed below.

General comments:

Introduction

no particular emphasis on the scientific aspects that are examined later in the paper (e.g. the analysis of meteorological and chemistry error covariances). These applications are listed in the content of sections, with lack of important details, like the ensemble specifications, or too much detail, like the specification of the CV5 set of variables or the CONUS domain. I suggest to the authors to better introduce the scientific framework of the examined cases (e.g. multivariate meteorological analyses), with corresponding references, then introduce the numerical experiments. The reader should understand why those experiments are done at the introduction level. Details about the single experiences (e.g. the geographical domain, the ensemble...) could be given later in the corresponding sections.

Section 3

Section 3 describes the details of the employed algorithms, the code utilization and presents some results from the numerical experiments (mostly error correlation plots). This makes a very long section, difficult to be read. I suggest the authors to remove all the technical details like names of FORTRAN variables and routines from the text. Some sub-sections could also be removed (e.g. 3.2.1 and 3.2.3). The code instructions should be moved in an appendix and reference the main text when needed. Second, I suggest to move the discussion of the correlation plots (Fig 3,4,5) to section 4, adding a detailed description of the model configuration used to calculate the ensemble statistics, which was missing in Section 3. In this way the reader can find the complete discussion of the numerical experiments in the same section. Moreover, the analysis of error correlations will be directly followed by the length scale/EOF analysis.Finally, section 3.2.2 could be merged with section 3.1.2, since they are strictly related.

Section 4
Please avoid switching frequently from grid point to km when discussing the length scales (e.g. page 4309, lines 20-22, page 4313, lines 7-9). Physical units like km for horizontal distances or hPa for vertical distances are preferable. Otherwise put always grid points and corresponding physical values in brackets. All plots should provide axes in physical units as well (Figure 4-5-6-9-11-12-13-15-16-18). Figure 14 would not be necessary anymore.

Answers to the general comments

We want to thank Referee 5 for the numerous remarks that lead to major revisions in the structure and presentation of the document. Additional information has been given to improve the presentation and discussion of the section about chemistry data assimilation.

- ***Introduction***
  Several part of the introduction has been rephrased to introduce why we do a focus on cloud and chemistry data assimilation.

- ***Section 3.0***
  We follow most of the recommendation to change the structure of the document: technical information are moved in three appendices and some paragraph have been merged. (see the structure presented at the beginning of the document)

- ***Section 4.0***
  Discussion on length scale are done preferably done in their physical units. Additional information to switch easily from grid units to physical units for the Figure 8-9-11-12-13-15-16-18.

Detailed comments

1) Page 4292, lines 5-10: This sentence is too long and does not clarify what the GEN_BE does. From the title and the previous lines (3-6) the reader expects a generic or generalized code conceived to model background error covariances for data assimilation applications. Here GEN is used for GENerate, which is

indeed the purpose of the presented code e.g. generate B parameterizations for further use in data assimilation systems (like WRFDA and GSI). The abstract should clearly state this and the authors should decide between ‘generate’ and ‘generalized’.

The title “Generalized Background Error Covariance Matrix Model (GEN_BE v2.0)” refers the ability of the new code version 2.0. The first version was designed mainly to handle variables and linear regression coefficient hard coded in the code. As explained in the abstract and the introduction, the new framework allows to handle different control variables defines and cross-correlated errors defined as an input. Also, implementation of new models should be straightforward. Finally, the code version 2.0 gathers different transform such Uv defined by EOF or with recursive filter, variance 3D (S).

Also, the sentence have been spit: ” …Forecasting (WRF) community model. GEN_BE allows for a simpler, flexible, robust, and community-oriented framework that gathers methods used by some meteorological …”

2) Page 4292, lines13:‘...performing benchmarks...’, please precise what kind of benchmark you considered in the study (e.g. multivariate meteorological analyses) before introducing the hydrometeors and atmospheric chemistry applications.

The benchmark involves different modeling of **B**.*The sentence has been completed: “ … by performing benchmarks of different modeling of* ***B*** *and showing …”*

Additional modifications are done:
L17: “a tool flexible enough to involve” replaced by “a tool flexible enough to implement”

3) Page 4292, line20:‘...chosen as a testbed for diagnostic and new modelling of B’ Do you mean that GEN_BE can be used to verify the results of similar codes? Or that new variables and error covariances can be implemented and tested easily? Please clarify or remove.

The sentence has been replaced by: “L20:” replaced by “ … (GEN_BE v2.0) can be easily applied to other domains of science and be chosen to diagnose and to model B.”

.

4) Page 4292, lines 25-26 and page 4293 lines 1-5: I find this affirmation too

strong, the performances of data assimilation can be improved also by considering more advanced assimilation algorithms or by improving observation error estimations. Moreover, I can't see the logical link with the end of the sentence '...assuming that the underlying probability errors are normally distributed'.

The first sentence has been modified and split: "Since the best estimate of the background error covariances matrix (B) is a key component for data assimilation improvements, various … within a variational framework"
The second sentence is: "The probability errors are supposed to be normally distributed and B is determined for a limited set of variables, called control variables"

5) Page 4293, lines 5-7: '...are usually...' Please either add a reference or explain the reason of choosing variables with uncorrelated errors.

The reason is that we want B block diagonal after Up to be able to model a full B matrix. When the control variables are uncorrelated, there is no need to model their cross-correlated errors. Otherwise, they will be model by linear regressions in our case. This part has been moved into section 2.2.2, in paragraph of Up, to explain more in details.

6) Page 4293, line17-18-19:'MM5, NCAR, WRF', Please add the full name of every model or institute the first time they appear in the text, and possibly a reference in case of a model (e.g. for WRFDA).

We added the description and the references.

7) Page 4293, line26: '...unite them'. Clarify what should be unified.

It has been rephrased:
"This new flexibility associated with the possibility to define a set of control variables and their covariance errors as an input should reduce future developments of the code considerably and should benefit to a larger community in geophysical science."

8) Page 4294, line5: '...using different transforms...'.
The concept of transform was not introduced before, which makes the sentence obscure for the reader.

It has been defined now in the sentence. "Section 2.0 presents the role of the background error covariance and how a series of different operators (i.e. balance, vertical and horizontal transforms) can model B."

9) Page 4295, line7: the errors are supposed uncorrelated, not the observations themselves.

We replaced “uncorrelated observations” by “uncorrelated observation errors”.

10) Page 4295, line 17: please specify that $x=(\mathbf{x}_b-\mathbf{x})$

We replaced $\delta x = B^{1/2}u$ *by* $\delta x = (x_b - \mathrm{x}) = \mathbf{B}^{1/2}u$ *.*

11) Page 4295, line 24: you could probably mention that the rewritten cost function in Eq. 3 is quadratic, which allows a global minimization

We replaced “H is the tangent linear operator” by “H is the linearized observation operator which makes the cost function quadratic and easier to minimize.”

12) Page 4296, line 10-14: please define what does it mean balanced and unbalanced before, or add a reference.

The paragraph describing the Up matrix is rephrased and reference to section 3.2 is added.

13) Page 4296, line 17: please clarify how horizontal diffusion is used in the framework of B modelling or remove it. The reader is anyhow addressed to other studies on the subject of covariance modelling few lines later.

We removed “which are affordable approximations of horizontal diffusion.”

14) Page 4297: Section 2.2.3 seems more as part of the introduction or should be reduced and merged with 2.2.2.

The section 2.2.3 doesn’t exist anymore.

15) Page 4297, lines 23-27: This was already said at line 2-3 and in the introduction. Please consider removing it.

The section 2.2.3 doesn’t exist anymore.

16) Page 4298, line 4: define ‘raw model perturbations of the analysis variables’. Do ‘analysis variables’ correspond to the ‘control variables’?

We replaced “raw model perturbations of the analysis variables” by “model perturbations of the control variables.”

17) Page 4299, lines 8-16: The explication of the reasons to perform spatial

averaging, or ‘binning’, are not clear. I don’t see how spatial averaging can ‘increase the number of samples’ or ‘reduce the dimensional of statistical output parameter’ or ‘add heterogeneity and anisotropy in B’. I suppose that the authors want to say that, since the number of samples of the ensemble is limited, a strategy to filter the sampling noise is needed. The paragraph should be rephrased with the aid of some of the numerous references that exists in term of ensemble filtering.

Answer: In variational methods, B needs to be estimated for the entire domain. Since it is not possible to compute a full rank B matrix, different hypothesis are taken to filter the sampling noise coming from a limited number of perturbations and to reduce its dimensions in order to model a static error covariances. The focus of this paragraph is to have a background error statistics (coming from an ensemble or a NMC method) model for the entire domain: Binning is also a way to model a B matrix for specifics needs, to filter the statistics, and reduce its dimensions (Regression coefficients computed by grid point Reg_coeff(i,j,k1,k2) becomes Reg_coeff(b,k1,k2) where b in the bin class of a grid point at the location (i,j))

Correction: the paragraph is rephrased
“Since the number … characterize convection events”

18) Page 4299, line 20. Please add a reference about the resulting skewness of hydrometeors statistics.

We added the reference “can be skewed (Michel et al. 2011)”.

19) Page 4300, line 8. What do you mean by ‘estimation error’?

The corrected sentence is “Analysis increment for one variable may impact an another if they have correlated errors.”

20) Page 4300, lines 11-15. Either give a reference to the NCEP method or write more clearly the steps that lead to the calculation of the regression coefficients.
In section 3.2, Eq. (5) is added and the appendix B gives some details about the calculation (no references found).

Similarly for lines 16-19. Are linear regressions calculated on perturbations or variables themselves?

Linear regressions are applied to the perturbations to compute the statistics on uncorrelated control variables. They are also applied later in the data assimilation *process on the variables themselves.*

Is Up block diagonal or Uh and Uv? Please clarify.
Up is block diagonal.

21) Page 4300, line 20. Stage 2 has changed with respect to GEN_BE v1.0? Is it necessary to be written?
It has been removed.

22) Page 4301, line 20. L should be squared, x should be $\delta\mathbf{x}$ and the equation seems not numbered.
Corrected to δz and numbered.

23) Page 4302, line 2. The correct equation seems 5 or the one which is not numbered.
The equation has been numbered since.

24) Page 4302, line 8-9. What does it means 'by bin'? Do you mean, without spatial averaging? And why it is not useful for data assimilation? Please clarify.

We removed this sentence and rephrase this paragraph:
Pannekoucke et al. (2008) studied … the horizontal length scale for stage 4.

25) Page 4302, line 9-11. Which regression coefficient? Does it mean that the binning can be decided independently at each stage? Please clarify

For example, the regression coefficients can be binned and the vertical length scale computed uniform by vertical level. This is the case for the Bnam defined for regional applications in GSI provided by NCEP.

26) Page 4302, lines 19-24. Quantify larger, smaller and local in term of kilometres.

This has been included in the text:
"The stream function (3a) and velocity potential control variables have larger and more isotropic spatial correlations while the temperature (3b) and the humidity (3c) control variables show smaller and anisotropic correlations at different locations. The radius of the area where the correlation overpasses 0.9 is within a range of 100 km to 400 km for stream function while this radius reaches its maximum around 100 km for temperature and humidity. Hydrometeors mixing ratio show even more local structures due to their sparse location on the horizontal and the vertical (3d)."

27) Page 4303, line 5. I could not find the explanation in Sect. 3.1.2
Wrong reference, it should be Sect. 3.3 where we introduced the decomposition by EOF modes.
We replaced "by EOF mode or by vertical level as explained in Sect. 3.1.2 by "by vertical level or by EOF mode as explained in Sect 3.3"

28) Page 4303, lines 6-7: Is the solution calculated considering the nearest grid points?

A radius r0 can be defined to consider only the points which are distant of a distance r inferior to r0. Moreover as it is mentioned in the same section, the use of the second formula (ls_method=2,Wu et al. (2002)) is advised.

29) Page 4303, line 11: what is it meant by 'pseudo correlation'?
We removed pseudo.

30) Page 4303, line 20-21. What does it mean 'at best it can be statistically binned'? Moreover, horizontal length scales for a given vertical level are 'usually' not uniform, as also shown in the example in Figure 3. Please clarify.

We agree, length scale can be computed uniform or binned (which include diagnosed by grid point). Moreover, in practice, operational centers such as NCEP, used statistics averaged by vertical level and binned for some of them. There are potentially some issues to handle heterogeneous length scales with recursive filters as mentioned in this rephrased paragraph: "The horizontal length scale can be … be required because of recursive normalization issues (Michel and Auligne 2010). "

31) Page 4304, line 4. Please add a reference about the poor results of recursive filters.
See point (30)

32) Page 4305, line 9. What does 'Generalized' stands for in the section title? As suggested in the general comment I would merge this section with the 3.1.2.

This paragraph has been merged to the section mentioned and rephrased. Also, the part related to data assimilation of the multivariate hydrometeors experiment is presented now in the new section 5.1. This merged section has been rephrased.

33) Page 4305, lines 19-21. The sentence is not clear, what kind of benchmark is done? Which are the other series of operators?

The sentence has been replaced by "Benchmark results of pseudo temperature test involving different modeling of B and the same Up transform (CV5) are shown Sect 4."

34) Page 4305, line 22. 'Recent studies' should be referenced.
As mentioned point (32), the paragraph has been rephrased.

Now the studies dedicated to better estimate the background error covariances matrix in cloudy areas are first presented and then discussed.
They are introduced by the sentence "Thus, various studies have been dedicated to better estimate the background error of humidity in cloudy areas (Carron and Fillon 2010, Montmerle et Berre 2010, Ménétrier and Montmerle 2011)."

35) Page 4305, lines 26-28. The statement is not really supported by Figure 4 because, as far as I understood, the statistics are shown for the entire CONUS domain (dry and wet areas). Or does the statement refer only to the cited study?

The statistics are shown Fig. 4 for the entire CONUS domain. The statement refers only to the cited studies. The new paragraph presents first these figures and then discusses about the application of binning.

36) Page 4306, line 1. Please avoid using probably, if the results are suggesting the conclusion that condensation and precipitation process determine the observed statistics clarify it, add a reference otherwise.

The sentence is now:
"At saturation, these statistics likely rely on processes of condensation and precipitation when the released latent heat flux warms the atmosphere (Holm 2002)."

37) Page 4306, line 5. 'They explain that imbalance in precipitating areas'. Please clarify the imbalance between which variables.

This sentence has been completed: "For a winter test-case where stratiform-type precipitation is predominant, they explain that geostrophic imbalance in precipitation areas, can be characterized by the linear balance operator between the stream function and the mass fields (t and ps)"

38) Page 4306, lines 17-18. ‘As the dynamic control variable...do not explain statically the presence of fog’ The authors probably want to say that dynamical variables such as vorticity and divergence do not drive fog formation processes.

The sentence is rephrased: ” Dynamical variables such as vorticity and divergence are not included in the balance humidity operator since they do not drive fog formation processes.”

39) Page 4306, line 22. ‘...dry and humid atmosphere’ . I imagine the authors mean for both a dry and a humid atmosphere. Again, is this statement supported by the Figure 4, and if yes please clarify. Otherwise add a reference.
The rephrased paragraph should clarify it.

The paragraph has been rephrased and the sentence is:
“For example, Fig. 4 shows the cross-correlation between humidity and temperature for all atmosphere conditions (mixing dry and wet conditions).”

40) Page 4306, lines 24-25. Which is the transform used in real time at NCEP? For real time do the authors mean operational analyses?
The same Up transform is used. Kleist et al. (2009) described this transform used in GFS-GDAS system. This sentence has been removed.

41) Page 4308, lines 10-18. As far as I understood a non-cloudy/cloudy mask is used to restrict the statistical sample of perturbations. Which values of cloudiness or other relevant variables are considered to perform this filtering? ‘Such filter may overestimate the vertical correlation around a given vertical level’. Please clarify the reason and which levels are affected by this issue.

As suggested in previous comment, this part has been moved to a new section 5.1.1. The sentence has been replaced by: ”However, we may want to localized this balance around a given vertical model level.”

42) Page 4308, lines 21-26 and page 4309 lines 1-4. In the general comments I suggested to move here the description of the numerical experiences setting. Some additional details should however be given or appropriate references should be provided for a better interpretation of the results. Which is the NCEP real time configuration (e.g. assimilated datasets)? What are the main features of the WRF ensemble (type and magnitude of perturbations, initialization...)? What kind of horizontal and vertical grid do GSI and WRFDA use (degrees, hybrid sigma-pressure levels, resolution)? What does NAM stands for? Is the NCEP real time configuration differing also on the vertical grid? What kind of data is assimilated in the NCEP operational system?
Table 2 has been added section 4.0 to explain in details the different modeling of

B. The Acronym NAM has been described.

43) Page 4309, line 22. '...decreases more monotonically' is not a clear statement, unless a degree of 'monotonicity' is defined. Please rephrase.
The new sentence is "Relative humidity length scale remains small, decreasing from approximately 30 km to 15 km as a function of the EOF mode."

44) Page 4310, line 12. '... representing more synoptic events at high altitude' is not scientifically sound. What it is meant by 'more synoptic' and 'high altitude'? Please rephrase.
The end of the sentence has been replaced by "from the bottom to the top of the model as they represent larger scale events."

45) Page 4310, lines 22-24. First define the experiment setting (innovation and observation error values, location of the observation) then describe briefly what do the plots represent (horizontal and vertical slices of the resulting increment).
The full section 4.2 has been rephrased.

46) Page 4311, lines 4-7. The sentence is too long and not very clear. What is the link with the fact that the domain is of limited area? Please rephrase.
The full section 4.2 has been rephrased.

47) Page 4311, line 9. '...show close results'. It is difficult to verify this statement on the plots. Values of contour lines in Fig. 11-12-13 are in very small letters and it seems that the contour ranges are different among the different experiences. The plot range should be uniformed, the physical units for the contour lines added and I might suggest adding a color scale to ease the evaluation of the maximum and minimum values of the increment.
The full section 4.2 has been rephrased.

48) Page 4311, line 12. Can you provide some insights about the observed differences in the horizontal length scale between the EOF and the level by level estimation?
The full section 4.2 has been rephrased.

49) Page 4311, line 15-21. 'More climatological' is not scientifically sound, please rephrase. I also think that a deeper discussion of the differences between the NMC derived B and the ensemble derived B would greatly improve the paper. But this should be probably done when horizontal and vertical length scales are discussed (currently Sec. 3.1.4 and 4.1.2 currently).

The full section 4.2 has been rephrased.

50) Page 4311, line 24-25. 'The XZ plan follows the isocontour of 0 m s-1 for U' means that the U increment is negligible? Are the 'complex structures' observed for V realistic in term of the modelled balance?
The full section 4.2 has been rephrased.

51) Page 4311, line 28. As noted in points 48-49-50, these differences should be better presented and discussed before affirming that they are well explained.
The full section 4.2 has been rephrased.

52) Page 4312, line 16. Please clarify why recursive filters make the analysis of length scale 'easier'.

The sentence is rephrased: "The vertical and horizontal transforms retained are the recursive filters making the interpretation of the length scale parameter easier as they are directly associated to a vertical model level."

In addition, in section 4.1.1, a sentence has been added first at the end of the first paragraph: "Also, the EOF decomposition allows optionally some filtering as the largest variances (i.e. eigen values) are associated with the first EOFs, the latest EOFs may be not taken into account if they mostly represent vertical noise in the system"
Then a second sentence at the end of the second paragraph:
"As the horizontal length scale is associated to EOF mode and not directly related to a vertical model level, futher discussions on the association of length scale with physiscal event may be difficult."

53) Page 4313, lines 3-14 Figure 16 seems to be identical to figure 15. Please check.
Figures should be different, mistake corrected.

54) Page 4313, line 24. Change Fig 18a with 18b (and b with a at page 4314 line 2). Is the variance profile in Fig. 18 coming from the ensemble?
We changed Fig 18a with 18b. Both figures come from the same D-ensemble.

55) Page 4314, line 3-5. 'The increment is most likely important' is not correct. Please put larger, smaller or significant and quantification in physical units. Are the observed increments over the dry area not realistic?

The sentence: "The increment is most likely important where the variability of cloud presence exists …" is replaced by 'The increment is most likely greater than 10^{-3} g/kg where the variability of cloud presence exists."

Answer to the question: There is no increment if the background error standard deviation (diag(B)) are equal to zero. This is the reason of the sentence that follows "A minimum value would likely certainly need to be set to retain the possibility of increments in the dry area". If a minimum value is set up, increment will be possible.

56) Page 4314, line 12-13. Is this result specific to the examined case or is it expected in general?
The sentence is removed.

57) Page 4315, line 16. 'similar results with comprehensive differences' is not correct. Please rephrase considering the new elements arising from the discussion in Sec. 4.2
It has been rephrased in the text: "second, .. using Bnam."

58) Page 4315, lines 26-27. This statement is too generic. Non-linearity exists in meteorology as well and it does not hamper data assimilation.

We removed this statement from the text.

59) Page 4316, lines 19-21. The reference to Barré et al. 2013 is not very pertinent to the discussion. Either add a comprehensive list of studies that performed chemical data assimilation or cite only the studies that focused on the modelling of the B matrix (e.g. Massart et. al 2012, Jaumouillé et al 2013, Gaubert et al. 2014). Since this is not a review article the second option should be considered.

We add a short review of the B matrix characterization for atmospheric chemical data assimilation. We mention studies that use different orders of detail for the B matrix modelization: from using static estimated B matrix to hourly length scales variations. We should mention that Gaubert et al. 2014 uses a ensemble Kalman filter technique that implicitly characterize the B matrix, as opposed to variational technique where B has to be specified.

60) Page 4316, lines 23-25. Taking a realistic background error into account does not depend on the complexity and the accuracy of the chemical models.

Consider removing this sentence.

The statement has been removed from the text.

61) Page 4317, line 1 and previous line. Either detail how the aerosol optical depth is used or do not mention it.

The statement has been removed from the text.

62) Page 4317, lines 9-14. Please detail what kind of chemical scheme is used and/or add a reference for WRF-CHEM, MOZART and MEGAN. Provide also some information or reference about the ensemble perturbations (variance, spatial/temporal correlation etc.)

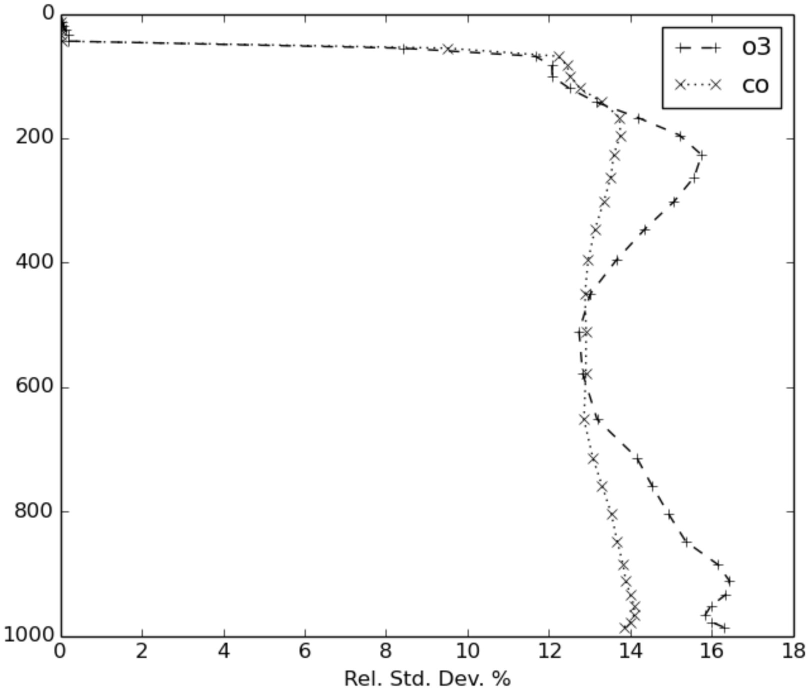
The text has been detailed.

*In the last following comments the reviewer arises important questions on chemical **B** matrix characterization. Since this paper is a general presentation on the GENBV2 system and the chemistry part a proof of concept that the code can be employed on specific cases as chemistry, we do not decide to dig into details this part. We agree with the reviewer that is an interesting topic and needs further diagnosing in a possible following paper. This has been clarified at the end of the section.*

63) Page 4317, lines 14-16. The relative variability should also be displayed in Figure A1, at least for ozone. It would allow to better detect the boundary layer variability of ozone due to the perturbed emissions.

In this section, we chose to put the standard deviations on their original units to show if the calculated standard deviation looked physical and (also because a data assimilation system will deal with those absolute values and not the relative ones).

Enhanced boundary layer values due to perturbed emissions appear close to the source regions (mostly anthropogenic sources over urbanized regions). Where emissions are strong enough, the emission perturbation will produce a standard deviation that is stronger than the standard deviation produced by the lateral boundary conditions. This is most likely the case for relatively long-lived species as ozone and carbon monoxide (couple weeks of life-time). The spread values relative to the ensemble mean averaged along the domain will not necessarily reveal enhanced values on the boundary layer more than absolute spread values. To convince the reviewer we provide below a plot showing relative *standard deviations profiles for ozone and carbon monoxide. Enhanced boundary layer relative variability is not obviously observed for carbon monoxide. Only slight increase for ozone relative variability is observed toward the surface. Because of reason stated above ozone standard deviations is not showing a clear enhancement due to the averaging of different regions (e.g. sea versus land, high emission regions versus remote region, high PBL height versus low PBL height). This point needs further regional detailed investigations that are out of the scope of the paper.*



64) Page 4317, lines 22-26. Vertical mixing in the planetary boundary layer is supposed to introduce a vertical error correlation, not to decrease it. Since the vertical mixing decreases above the boundary layer, this is probably the reason of the decrease of the vertical length scale above 850 hPa. On the other hand, surface emissions are generally injected over the first levels of chemical transport models, which might increase the error correlation close to the surface. The authors should verify the way emissions are treated in WRF-CHEM.

We agree with the reviewer with the first part of this comment and we have clarified the text accordingly. However we do not fully explain the strong decrease of vertical correlations close to the surface, since emissions mostly impact the model lowest level (closest to the surface) in WRF-Chem. The text has been clarified accordingly.

65) Page 4318, line 10. Since one of the main content of the paper is the balance between control variables it would have been very interesting to check whether the linear regression approach provides meaningful results applied to interacting chemical species like NOx, CO and O3. Can the authors comment on this?

We agree on the reviewer that the chemical balance between variables is a very important and interesting topic for chemical data assimilation purposes. However as stated above, this section is a proof of concept that the GEN_BE v2.0 code can be directly adapted to chemical variables. Diagnosing the chemical balance would require an extensive study on the B matrix for tropospheric chemistry, which is not the scope of this paper. Chemical balance (on various atmospheric chemical models at different scales) then could be diagnosed by using GEN_BE V2.0 in following studies.

Minor corrections:

1) Page 4293, line 9: change ‘dataset observations’ to ‘observational datasets’
changed

2) Page 4293, lines11-12: do you mean that the availability of more observations involve the control of new model variables? Please rephrase.
Cloud and Chemistry data assimilation may involve new variables such as hydrometeors and chemical species. It has been added the word may “ … large set of sensors that may involve more variables, which are ..”

3) Page 4293, line 27. Change ‘the two first sections’ with ‘Section 2.1 and 2.2’
It has been rephrased.

4) Page 4295, line 10: ‘comprised of’ should be ‘being comprised of’
done

5) Page 4296, line 3: Change ‘decomposed to’ in ‘decomposed into’
changed “decomposed into”

6) Page 4296, line9: Please add ‘foreach grid point’.
It has been added (grid point space)

7) Page 4298, line 15: change ’24 minus...’ with ’24 h minus...’
changed to “e.g. 24 hour minus 12 hour forecasts”

8) Page 4300, line 24: change ‘do not depend of the control variables’ to ‘do not *depend on the particular choice of the control variables’*

9) Page4301,line 5: specify that the length scale is horizontal

10) Page 4303, line 5: change ‘by EOF mode’ with ‘for each EOF mode’

11) Page 4305, line 25: change ‘correlated errors between...’ with ‘error correlation between...’

12) Page 4310, line 3. Change ‘applied by vertical level...’ to ‘applied for each vertical level...’
changed by “applied at every vertical model level for each variables”

Bibliography

Gaubert, B., Coman, a., Foret, G., Meleux, F., Ung, a., Rouil, L., ... Beekmann, M. (2014). Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model. Geoscientific Model Development, 7(1), 283–302. doi:10.5194/gmd-7-283-2014
Jaumouillé, E., Massart, S., Piacentini, A., Cariolle, D., & Peuch, V.-H. (2012). Impact of a time-dependent background error covariance matrix on air quality analysis. Geoscientific Model Development, 5(5), 1075–1090. doi:10.5194/gmd-5-1075-2012
Massart, S., Piacentini, A., & Pannekoucke, O. (2012). Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses. Quarterly Journal of the Royal Meteorological Society, 138(665), 889–905. doi:10.1002/qj.971

Corrections Referee 1

(1) General comments:

Overall: This is a study describing a resource that is useful to the data assimilation community. I hope the authors make the code available – however, I do not see any information about this in the paper (maybe I have missed this). Perhaps the authors should consider providing this information. As such, I think the paper is of interest to GMD. However, before the paper is acceptable for publication the authors must at least do the following.

- (i) Improve the presentation – the paper has a lot of sloppy writing (see examples in specific comments).
- (ii) Provide more details linking the approach discussed and the general move to hybrid systems to represent background error covariances (e.g., WMO Data Symposium, 2013), e.g., in the conclusions section.

Answers:

(i) The structure of the paper changed and numerous corrections in the text have been done.
(ii) Additional information has been discussed in Sect. 6 to move towards hybrid systems (see paragraph, "Wang et al. (2008a, 2008b) performed ... useful in other geophysical applications."). References have been added.

We want to thank Referee 1 for the numerous remarks that lead to modify the structure and presentation of the document. Additional information has been given to improve the discussion about hybrid data assimilation systems.

(2) Specific comments:

The authors should also address the specific and style comments below.
Specific comments:

P. 4293 L. 1: It is "Forecasts".
We Replaced "Forecast" by "Forecasts".

L. 18: Introduce acronym for NCAR (and other abbreviations used, e.g., CONUS, SOAR, NAM).
P 4293 L18: we replaced NCAR by National Center for Atmospheric Research (NCAR)
P4294, L10 : we replaced "CONUS" by "CONUS (CONtiguous United States)"
P13, L11 : we replaced "a soar function" by "a second order autoregressive (SOAR)" function
NAM: North American Mesoscale

P. 4295 L. 1: You should mention that H is a non-linear operator.
We replaced "H called the observation operator" by "H, called the non-linear observation operator,"

L. 18: To help the reader, please explain what is meant by "conditioning" of the B matrix.
Reference has been added (Courtier et al. 1994).

P. 4297
L. 13: Perhaps mention that other variables, dependent on the control variables, can be derived from them. These variables do not need to be in the control vector.
We were speaking in this part more about when the choice of the control variable is already made.
This Sect. does not exist anymore.

P. 4300
L. 13: I think the more common spelling is "Cholesky".
Corrected to "Cholesky".

P. 4302 L. 9: Why is this not useful for data assimilation?
The paragraph has been rephrased: "Pannekoucke et al. … and the horizontal length scale for stage 4.

P. 4303 L. 19: Incorrect spelling of Pereira (according to references) and Pannecoucke. P.
Corrected according to the references i.e.: "Pereira" and "Pannekoucke"

4305 L. 17: What do you mean by "precising"?
This sentence has been rephrased in the text: "The parameters equal to 1… by subtracting their balanced part coming from the stream function (psi).

P. 4306
L. 1: I suggest you replace "probably" with "likely" (here and elsewhere).
Done P. 4306, L1
Done P. 4314, L5

L. 4: What is "ps"?
L. 14: What is "qs"?
ps is the surface pressure and qs is the specific humidity. Section 3.2 has been corrected defining first these abbreviations.

P. 4307 L. 3: Identify the variables in Eq. (12a).
We replaced the sentence:
"In a first approach, relative humidity is balanced with the mass fields and does not include dynamic variables such as the stream function and potential velocity as in the following Eq. (12a):"

by
“In a first approach, relative humidity (rh) is balanced in Eq. (12a) with the mass fields of unbalanced temperature (tu), unbalanced surface pressure (psu) and does not include dynamic variables such as the stream function (psi) and unbalanced potential velocity (chiu):”

L. 6: Explain the convention behind the writing out of “covar5”.
We replaced the sentence: “In this case, the line describing covariances with the humidity becomes: covar5 = 0, 0, 1, 1, 0, 0, 0, 0, 0, 0.”
by
“In this case, the line covar5 of Table B5 that describes the covariances between the fifth control variable, (relative humidity), with the third control variables tu and the fourth psu is: covar5 = 0, 0, 1, 1, 0, 0, 0, 0, 0, 0.”

P. 4309
L. 24: I suggest: “...a relatively small...”.
Done

P. 4310
L. 5: By “precedent” do you mean “previous”? If then, I think it would better to write “previous section”.
We replaced the sentence “The diagnostic of horizontal length scale shows similar characteristics to the one presented on the previous section 4.1.1 performed by EOF modes.”
by “The horizontal length scales diagnosed for each control variable by vertical level (Fig. 9) or by EOF mode (Fig. 8) have the same range of value”

L. 18: Do you mean “...decreasing the vertical correlation”?
We replaced “lowering down” by “decreasing”

L. 21: make -> provide.
Done

L. 24: Do you mean that the innovation is also 1 K?
We replaced by “with an innovation of 1 Kelvin and an obervation error of 1 Kelvin”

P. 4311 L. 2: Identify the J_b term.

Eq. (1) is modified page p4294 to include the Jo and Jb terms now.
In the paragraph that follows Eq. (1), the sentences: “R is the observational error covariance matrix. B is the background error covariance matrix.” are replaced by “The Jo term contains R, the observational error covariance matrix. The Jb term contains B, the background error covariance matrix covariance matrix defined in Eq. (2):”

L. 21: Why does it have more skill to estimate the correlated errors?

We gave explanations extended the sentence by:
"These ensemble based background error statistics have potentially more skill to estimate correlated errors related to the present meteorological event."

L. 24: What do you mean by the XZ plan?

We replaced "XZ plan" by "vertical cross-section (XZ)"
Also, we replaced L15 "On the vertical slice XZ" by "On the vertical cross-section (XZ)"

P. 4312 L. 21, 23: By "under" do you mean pressure levels at lower heights?

We replaced the expression in the sentence: "it occurs below 150 hPa pressure level for … below 400 hPa pressure level for…"

P. 4314 L. 25: Should be Hamill (here and elsewhere).

Hamil corrected to Hamill (here and line 7, page 4316)

P. 4315 L. 16: comprehensive differences between what parameters?

The conclusion has been partially rephrased: "Second, … area using Bnam."

L. 28: Please provide a reference for this statement about non-linearities. There has been lots of work on chemical data assimilation, so you could reference some papers here (you reference papers in Appendix A, but the concept of chemical data assimilation is discussed first here).

This statement has been removed.

P. 4316
L. 19: Barré et al. (2013) not in reference list.

The reference has been added

P. 4318
L. 8: Please rephrase "not for the good reason".

The sentence is rephrased: "For these reasons, the analysis may fit the observation even if the data assimilation does not involve the origin of the mismatch."

L. 10: Provide references for the work on hybrid and ensemble methods.

Additional references have been added in section 6.0 where ensemble and hybrid methods are discussed.

P. 4319
L. 29: Should be "Ebel".

Done

P. 4334
Fig. caption: Identify the variables (psi, ..., q_{cloud}). What is “sparecly”? Distrubuted -> distributed.

We changed the legend of figures 3 by “… (a) stream function (psi), (b) temperature (t), (c) relative humidity (rh), and (d) Cloud mixing ratio (q_{cloud}). Larger correlations are observed for stream function compared to temperature and relative humidity. Cloud mixing ratio has the smallest correlation due to sparce location of hydrometeors.”
Note:
“observed” → “observed”,
“sparecly distrubuted” →” due to sparce location of hydrometeors”

P. 4336 Fig. caption: Identify variables. Intensify -> intensifies.
Variable of relative humidity (rh) and cloud mixing ratio (q_{cloud}) are identified.
“Intensify” is replaced by “intensifies.”
Also temperature (t) specific humidity (qs) and relative humidity (rh) are identified
P4336

P. 4338 Fig. caption: Which are the former variables?
We replaced “former” by “control”

P. 4342, 4343, 4344 Fig. caption: What does each panel show?

P. 4342: replaced the caption by “
Pseudo observation test of temperature (innovation of +1 K) from the WRFDA application. The three plots on the left panel show, from top to bottom, horizontal cross-section (XY) of t (K), U and V wind component (m s-1) respectively. Then, the right panel shows the corresponding cross-section (XZ) of the former variables (Beof: WRF Res. 15 km, D-ensemble, EOFs).

p4343
Pseudo observation test of temperature (innovation of +1 K) from the GSI application. The three plots on the left panel show, from top to bottom, horizontal cross-section (XY) of t (K), U and V wind component (ms-1) respectively. Then, the right panel shows the corresponding cross-section (XZ) of the former variables. (Brcf: WRF Res. 15 km, D-ensemble, RFs).

P4344
Pseudo observation test of temperature (innovation of +1 K) from the GSI application. The three plots on the left panel show, from top to bottom, horizontal cross-section (XY) of t (K), U and V wind component (ms-1) respectively. Then, the right panel shows the corresponding cross-section (XZ) of the former variables. (Bnam:WRF-NMM Res. 12 km,

NMC, RFs).

P. 4346 Fig. caption: Say something like “The plots show similar behavior”.
Caption replaced by:
“Horizontal length scale for the hydrometeors using (a) 50 members and (b) using 5 members. The plots show similar characteristics regardless to the ensemble members.”

P. 4347 Fig. caption: Rephrase – text is not clear.
Legend replaced by :
“Vertical length scale for the hydrometeors using (a) 50 members and (b) using 5 members. The plots show similar characteristics regardless to the ensemble members.”

P. 4348 Fig. caption: What is panel (b)? Slide –> slice (I presume).
Replaced “slide” by “slice”

P. 4349
Fig. caption: Explain how the presence of low maritime cloud is indicated.
Caption has been replaced by “(a) Profile of standard deviation of liquid water condensate mixing ratio (qcloud in g kg^{-1}) averaged along the vertical and (b) horizontal cross-section of standard deviation of qcloud at the vertical model level 5 (950 hPa). Both plots indicate the presence of low maritime clouds noted by high standard deviation.”

P. 4350, 4351, 4352 Fig. caption: What are the units? Same for similar figures. Style comments:
Caption Corrected: units in ppmv added, (WRF-CHEM, Res. 36 km, D-Ensemble).

(2) Style comments

P. 4296 L. 3: to -> in.
done

P. 4300 L. 19: bloc -> block.
done

P. 4301
L. 5: Avoid the use of abbreviations like aka.
“computing eigenvectors (aka. EOFs) and eigenvalues.” has been replaced by “The first method diagonalizes the VACM performing an EOF decomposition (i.e. computing eigenvectors and eigenvalues).’’

L. 11: “...from Daley’s...”.

Done

L. 15: development -> expansion.
The sentence is replaced by "Approximating Eq (8a) with finite difference to the second order derivatives of $\rho(\delta z)$ *and assuming* ρ *is symmetric around the origin results in:"*

L. 18: named –> call.
The sentence is replaced by "where Lvp represents the vertical length scale approximate by a parabolic function."

L. 19: following -> follows.
The sentence is replaced by "If the correlation is approximated at the origin by a Gaussian function as follows:"

P. 4302
L. 4: close -> similar.
Done L5

L. 16: grib -> grid.
Done

P. 4303
L. 1: I suggest "...grid point space..."...
Done
L. 7: "...solving Eq. (8)...".
Done, removed the done
L. 14: "...procedure is both...".
Done

P. 4304
L. 2: spreaded -> spread.
Done

L. 3: Morever -> Moreover.
Done

P. 4307
L. 18: spreaded -> spread.
Done

P. 4308
L. 17: "…presented in Tables…".
We replaced by "... presented in Appendix B."

P. 4309
L. 20: EOF -> EOFs.
Done

P. 4311
L. 9: close -> similar.
Done

L. 21: skills -> skill.
Done

L. 23: adjustement -> adjustment.
Done

P. 4312
L. 4: “Modifications of the code have...”.
Replaced by “Code modifications have been done”

L. 20: I think you mean “...is a length scale smaller than...”.
The sentence “The horizontal length scale values of the different hydrometeors shown in Fig 15a do not overpass 30 km (2 grid points) which is smaller than that of other control variables.”
has been replaced by
The horizontal length scale values of the different hydrometeors shown in ***Error! Reference source not found.*** *are smaller in comparison of other control variables (less than 30 km, 2 grid points).*
Is smaller than two grid points i.e. 30 km. and also in comparison to that of other control variables.

P. 4313
L. 1: Perhaps replace “remarkable” with “noted”?
Replaced

P. 4314
L. 16: “...from the D-ensemble...”.
Done, added “the”

P. 4315
L. 19: precedent -> previous.
Done

P. 4316
L. 13: meteorlogy -> meteorology.

Done

P. 4330
Comments column:
“This file exists...”;
done
“...input file is split...”;
done
“...binary files split...”;
done
“...the horizontal lenscale...”.
replaced by “length scale”
corrected “readble” by readble

P. 4331
Description column:
lenscale
Replaced by length scale;
eigen value
Replaced by eigenvalue
eigen vector
Replaced by eigenvector

P. 4333 Fig. caption: “...model B are in blue”.
“is” replaced by “are”

P. 4337 Fig. caption: “...resulting from the...”.
“of” replaced by “from”

P. 4339 Fig. caption: larger -> longer.
Replaced

P. 4430 Fig. caption: smaller -> shorter.
Done

P. 4345 Fig. caption: I suggest “Vertical model level”.
Done

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

G. Descombes[1], T. Auligné[1], F. Vandenberghe[2], D. M. Barker[3] and J. Barré[4]

[1]{National Center for Atmospheric Research/MMM, Boulder, Colorado}

[2]{National Center for Atmospheric Research/RAL, Boulder, Colorado}

[3]{Met Office, Exeter, United Kingdom}

[4]{National Center for Atmospheric Research/ACD, Boulder, Colorado}

Correspondence to: G. Descombes (gael@ucar.edu)

Abstract

The specification of state background error statistics is a key component of data assimilation since it affects the impact observations will have on the analysis. In the variational data assimilation approach, applied in geophysical sciences, the dimensions of the background error covariance matrix (**B**) are usually too large to be explicitly determined and **B** needs to be modeled. Recent efforts to include new variables in the analysis such as cloud parameters and chemical species have required the development of the code to GENerate the Background Errors (GEN_BE) version 2.0 for the Weather Research and Forecasting (WRF) community model. GEN_BE allows for a simpler, flexible, robust, and community-oriented framework that gathers methods used by some meteorological operational centers and researchers.

We present the advantages of this new design for the data assimilation community by performing benchmarks of different modeling of **B** and showing some of the new features on data assimilation test cases. As data assimilation for clouds remains a challenge, we present a multivariate approach that includes hydrometeors in the control variables and new correlated errors. In addition, the GEN_BE v2.0 code is employed to diagnose error parameter statistics for chemical species, which shows that it is a tool flexible enough to implement new control variables. While the generation of the background errors statistics code has been first developed for atmospheric research, the new version (GEN_BE v2.0) can be easily applied to

titi toto 1/9/15 4:07 AM
Formatted: Title

titi toto 1/9/15 4:07 AM
Deleted: and

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English (UK)

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Line spacing: 1.5 lines, Widow/Orphan control, Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Arial, Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: to allow

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: involve

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: extended

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

other domains of science and be chosen to diagnose and to model **B**. Initially developed for variational data assimilation, the model of the **B** matrix may be useful for variational ensemble hybrid methods as well.

titi toto 1/9/15 4:07 AM
Deleted: as a testbed for diagnostic

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: new modeling of B

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

1 Introduction

Since the best estimate of the background error covariances matrix (**B**) is a key component for data assimilation improvements, various meteorological operational centers such as the European Centre for Medium‑Range Weather Forecasts (ECMWF), the National Centers for Environmental Prediction (NCEP), and the UK Met office, continue to develop new algorithms, techniques, and tools (Bannister, 2008a, b) to model **B** within a variational framework. The probability errors are supposed to be normally distributed and **B** is determined for a limited set of variables, called control variables. The dimensions of **B** are also reduced by diagnosing several parameters that drive a series of operators to model **B.**

However, necessities to extend the capabilities of **B** subsist. For example, improving cloud (Auligné et al., 2011) and pollution forecast are major drivers of development of cloud and chemical data assimilation. In the meantime, as more and more observational datasets coming from radars, satellites, airplanes, and ground stations become available in real time, there is a tendency to generalize data assimilation to a large set of sensors that may involve more variables, which are present in geophysical numerical models.

The opportunity has been taken to redesign the GEN_BE code by extending its capabilities to investigate and to estimate new error covariances. Originally, the GEN_BE code was developed by Barker et al. (2004) as a component of a three-dimentional variational data assimilation (3DVAR) method to estimate the background error of the fifth-generation Penn State/NCAR Mesoscale Model (MM5, Grell et al., 1994) for a limited-area system. Since this initial version, various branches of code have been developed at the National Center for Atmospheric Research (NCAR) and at the UK Met Office to address specific needs using different models such as the Weather Research Forecast (WRF, Skamarock et al., 2008) and the Unified Model (UM, Davies et al, 2005) on different data assimilation platforms such as the Weather Research Forecast Data Assimilation system (WRFDA, Barker et al., 2012) and the Grid point Statistical Interpolation system (GSI, Kleist et al., 2009). Different choices of control variables and their correlated errors used to mimic general physical balance (geostrophic, hydrostatic, ...) in the atmosphere have been largely investigated by different operational centers and referenced in Banister (2008b). Since then, such multivariate

Formatted ... [1]
titi toto 1/9/15 4:07 AM
Deleted: improvements in data assimilation ... [2]
titi toto 1/9/15 4:07 AM
Deleted:)
titi toto 1/9/15 4:07 AM
Formatted ... [3]
titi toto 1/9/15 4:07 AM
Formatted ... [4]
titi toto 1/9/15 4:07 AM
Deleted: Forecast
titi toto 1/9/15 4:07 AM
Formatted ... [5]
titi toto 1/9/15 4:07 AM
Deleted: or
titi toto 1/9/15 4:07 AM
Formatted ... [6]
titi toto 1/9/15 4:07 AM
Deleted: and 2008b
titi toto 1/9/15 4:07 AM
Formatted ... [7]
titi toto 1/9/15 4:07 AM
Deleted: assuming that the underlaying
titi toto 1/9/15 4:07 AM
Formatted ... [8]
titi toto 1/9/15 4:07 AM
Deleted: . Statistics of the background error ... [9]
titi toto 1/9/15 4:07 AM
Deleted: that minimize the error covarian ... [10]
titi toto 1/9/15 4:07 AM
Formatted ... [11]
titi toto 1/9/15 4:07 AM
Formatted ... [12]
titi toto 1/9/15 4:07 AM
Deleted: need to be diagnosed to
titi toto 1/9/15 4:07 AM
Formatted ... [13]
titi toto 1/9/15 4:07 AM
Deleted: the
titi toto 1/9/15 4:07 AM
Formatted ... [14]
titi toto 1/9/15 4:07 AM
Deleted: that
titi toto 1/9/15 4:07 AM
Formatted ... [15]
titi toto 1/9/15 4:07 AM
Deleted: .
titi toto 1/9/15 4:07 AM
Formatted ... [16]
titi toto 1/9/15 4:07 AM
Deleted: dataset observations
titi toto 1/9/15 4:07 AM
Formatted ... [17]
titi toto 1/9/15 4:07 AM
Deleted: involves
titi toto 1/9/15 4:07 AM
Formatted ... [18]
titi toto 1/9/15 4:07 AM
Deleted: Necessities for cloud data assimi ... [19]
titi toto 1/9/15 4:07 AM
Formatted ... [20]
titi toto 1/9/15 4:07 AM
Deleted: and 2012
titi toto 1/9/15 4:07 AM
Formatted ... [21]
titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM
Formatted ... [22]
titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM

relationship approaches has been studied to characterize hererogeneous background errors in precipitating and nonprecipitating areas for regional applications (Fillon and al. 2010; Montmerle and Berre, 2010). Special emphasis is made in Michel et al. (2011) to include hydrometeors in the background error statistics as their direct analysis increment can come from data assimilation of radar reflectivity and satellites radiances. The framework of the GEN_BE code version 2.0 has been developed to merge these different efforts using linear regression to model the balance between variables, Empirical Orthogonal Functions (EOFs) decomposition techniques and diagnostic of length scales to apply recursive filters (RFs). It allows reading input from different models and providing output for different data assimilation platforms. This new flexibility associated with the possibility to define a set of control variables and their covariance errors as an input should reduce future developments of the code considerably and should benefit to a larger community in geophysical science.

This document describes the methods included in the GEN_BE code version 2.0 to investigate modeling of B for cloud and chemical data assimilation applications. Section 2.0 presents the role of the background error covariance and how a series of different operators (i.e. balance, vertical and horizontal transforms) can model **B**. The third section describes the general structure of the code, the methods to estimate the different parameters that model B and their role in the data assimilation processes. It explains how to modify, to extend the control variables and to define multivariate background errors when correlated errors between variables are modeled by linear regression (i.e. balance transform Up). Section 4 presents results of a benchmark performed on two different systems of data assimilation (WRFDA and GSI) using different model of B based on WRF model forecast involving the same set of five control variables (referenced as CV5 hereafter) available in GSI (Kleist et al., 2009). Finally, Sect. 5 presents results of a multivariate cloud data assimilation approach that includes hydrometeors as control variables (referenced as CV9 hereafter) and their correlated error with humidity. In addition, the diagnostic of parameters such as standard deviation, vertical and horizontal length scales are discussed for carbon monoxide (CO), nitrogen oxides (NO_x) and ozone (O_3) chemical species in a variational data assimilation framework.

titi toto 1/9/15 4:07 AM
Deleted: provide output for different data assimilation platforms. The possibility to define the set of control variables and their covariance errors as an input should reduce considerably future developments of the code and unite them.
This document is organized as follows: the two first sections present the role of the background error covariance and how a series of different operators can model B. The third section describes the general structure of the code

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: gives key information to model B

titi toto 1/9/15 4:07 AM
Deleted: a specific application. The estimation of the different parameters and their role in the data assimilation processes are discussed. It contains technical information that explains how to modify

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: extend the control variables

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English (US)

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: their error covariances. Section 4 presents results of a benchmark performed on two different systems of data assimilation (WRFDA and GSI) using different transforms involving the same set of five control variables (CV5) as defined for real time at NCEP on the rapid refresh domain. Finally, Sect. 5 presents an expansion of the control variable set on a test case that includes cloud hydrometeors in a multivariate approach (CV9). All the results presented in the differents sections were obtained from a numerical experiment

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: WRF model involving an ensemble of 50 members over the CONUS domain at 15 km resolution. Figure 1 shows the extension of the WRF computational domain. Each member is a si ... [23]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: system Data Assimilation Research Test (DART) was used to generate the ensemble ... [24]

titi toto 1/9/15 4:07 AM
Deleted: different dataset and it illustrates an application to chemical species.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

2 Role of the background error covariance matrix in the variational data assimilation method

2.1 The Variational method

The solution of three-dimensional variational data assimilation (3DVAR) is sought as the minimum of the following cost function (Courtier et al. 1994):

$$J(x)=J_b(x)+J_o(x)=\frac{1}{2}(x_b-x)^T B(x_b-x)+\frac{1}{2}[y_o-H(x)]^T R^{-1}[y_o-H(x)] \qquad (1)$$

Where ***x*** is the state vector composed of the model variables to analyse, at every grid point of the 3-dimensional (3-D) model computational grid. ***x***$_b$ is the background state vector, and usually provided by a previous forecast. ***y***$_o$ is the vector of observations and *H*, called the non-linear observation operator, is a map from the gridded model variables to the observation locations. The J_o term contains **R**, the observational error covariance matrix. The J_b term contains **B**, the background error covariance matrix defined in Eq (2):

$$B=\overline{(x_b-x_t)(x_b-x_t)^T} \qquad (2)$$

where x_t is the true state vector and the overbare represent an average over a number of forecasts.

By definition, exact values of **R** and **B** would require the knowledge of the true state of the atmosphere at all times and everywhere on the model computational grid. This is not possible, and both matrices have to be estimated in practice. Often, the **R** matrix is assumed to be diagonal, i.e. uncorrelated observation errors, with empirically prescribed variances. Notice also that the dimension of the **B** matrix is the square of the 3-D model grid multiplied by the number of analyzed variables. For typical geophysical applications as in meteorology, the size of the **B** matrix, being comprised of nearly 10^8 x 10^8 = 10^{16} entries, is too large to be calculate explicitly and to be stored in present day computer memories. As a result, the **B** matrix needs to be modeled.

Formatted ... [25]
titi toto 1/9/15 4:07 AM
Formatted ... [26]
titi toto 1/9/15 4:07 AM
Formatted ... [27]
titi toto 1/9/15 4:07 AM
Deleted: ... [28]
Unknown
Formatted ... [29]
titi toto 1/9/15 4:07 AM
Formatted ... [30]
titi toto 1/9/15 4:07 AM
Formatted ... [31]
titi toto 1/9/15 4:07 AM
Deleted: per
titi toto 1/9/15 4:07 AM
Formatted ... [32]
titi toto 1/9/15 4:07 AM
Deleted: irregularly distributed
titi toto 1/9/15 4:07 AM
Formatted ... [33]
titi toto 1/9/15 4:07 AM
Deleted: is
titi toto 1/9/15 4:07 AM
Formatted ... [34]
titi toto 1/9/15 4:07 AM
Deleted: is
titi toto 1/9/15 4:07 AM
Deleted: . Note that
titi toto 1/9/15 4:07 AM
Formatted ... [35]
titi toto 1/9/15 4:07 AM
Formatted ... [36]
titi toto 1/9/15 4:07 AM
Deleted: knowledge
titi toto 1/9/15 4:07 AM
Formatted ... [37]
titi toto 1/9/15 4:07 AM
Deleted: theoretically
titi toto 1/9/15 4:07 AM
Formatted ... [38]
titi toto 1/9/15 4:07 AM
Deleted: In general
titi toto 1/9/15 4:07 AM
Formatted ... [39]
titi toto 1/9/15 4:07 AM
Deleted: observations
titi toto 1/9/15 4:07 AM
Formatted ... [40]
titi toto 1/9/15 4:07 AM
Deleted: nor
titi toto 1/9/15 4:07 AM
Formatted ... [41]
titi toto 1/9/15 4:07 AM
Deleted: parameterized
titi toto 1/9/15 4:07 AM
Formatted ... [42]

2.2 Modelling of the background error covariance matrix

2.2.1 Control variable transform

The cost function as defined in Eq. (1) is usually minimized after applying the change of a variable:

$$\delta x = (x_b - x) = B^{1/2}u \qquad (3)$$

as it improves the conditioning (Courtier et al. 1994) and therefore accelerates the convergence. $\mathbf{B}^{1/2}$ is the square root of the background error covariance matrix. The variable u is called the control variable and the cost function becomes:

$$J(u) = \frac{1}{2}u^T u + \frac{1}{2}(d - HB^{1/2}u)^T R^{-1}(d - HB^{1/2}u) \qquad (4)$$

Where *d* is the innovation vector defined as $\boldsymbol{d}=(\boldsymbol{y}_o-H(\boldsymbol{x}_b))$ and it represents the difference between observations and their modeled values using a non-linear observation operator *H*. H is the linearized observation operator, which makes the cost function quadratic and easier to minimize.

2.2.2 Background errors covariance matrix modelled by a succession of operators.

The square root of the **B** matrix as defined in Eq. (3) is decomposed to a series of sub-matrices, each corresponding to an elemental transform that can be individually modeled:

$$B^{1/2} = U_p S U_v U_h \qquad (5)$$

where:

- The $\mathbf{U_p}$ matrix, called physical transform or balance operator, defines the set of control variables and their relationships. In practice, the control variables are calulated using the model variables and selected to minimize their cross-correlations. Also, the existing cross-correlations, called balanced part, can be reduced by appling statistical linear regressions (explained Sect. 3.2). The idea is that those new variables are less correlated with each other and so the corresponding off diagonal terms in the matrix vanish.

- The S matrix is diagonal and composed of the standard deviations of the background errors.

Formatted ... [43]
titi toto 1/9/15 4:07 AM
Formatted ... [44]
titi toto 1/9/15 4:07 AM
Deleted: ... [45]
Unknown
Formatted ... [46]
titi toto 1/9/15 4:07 AM
Formatted ... [47]
titi toto 1/9/15 4:07 AM
Formatted ... [48]
titi toto 1/9/15 4:07 AM
Deleted: $J(u) = \frac{1}{2}u^T u + \frac{1}{2}(d - HB^{1/2}u$ ... [49]
titi toto 1/9/15 4:07 AM
Formatted ... [50]
titi toto 1/9/15 4:07 AM
Formatted ... [51]
titi toto 1/9/15 4:07 AM
Deleted: tangent linear
titi toto 1/9/15 4:07 AM
Formatted ... [52]
titi toto 1/9/15 4:07 AM
Deleted: .
titi toto 1/9/15 4:07 AM
Formatted ... [53]
titi toto 1/9/15 4:07 AM
Formatted ... [54]
titi toto 1/9/15 4:07 AM
Deleted: 2
titi toto 1/9/15 4:07 AM
Formatted ... [55]
titi toto 1/9/15 4:07 AM
Formatted ... [56]
titi toto 1/9/15 4:07 AM
Deleted: ... [57]
titi toto 1/9/15 4:07 AM
Formatted ... [58]
titi toto 1/9/15 4:07 AM
Deleted: - The matrix S is diagonal and co ... [59]
titi toto 1/9/15 4:07 AM
Formatted ... [60]
titi toto 1/9/15 4:07 AM
Deleted: cross-correlations between different
titi toto 1/9/15 4:07 AM
Formatted ... [61]
titi toto 1/9/15 4:07 AM
Deleted: unbalanced
titi toto 1/9/15 4:07 AM
Formatted ... [62]
titi toto 1/9/15 4:07 AM
Deleted: obtained
titi toto 1/9/15 4:07 AM
Formatted ... [63]
titi toto 1/9/15 4:07 AM
Deleted: removing the balance parts through
titi toto 1/9/15 4:07 AM
Formatted ... [64]
titi toto 1/9/15 4:07 AM
Deleted: cross-correlations entries
titi toto 1/9/15 4:07 AM
Formatted ... [65]

- The U_v matrix, called vertical transform, defines the vertical auto-correlations for each of the u control variables. It is modeled by either homogeneous Empirical Orthogonal Functions (EOFs) or application of a recursive iterative filter.

- The U_h matrix, called horizontal transform, defines the horizontal auto-correlations for the u control variables. It is modeled through successive applications of recursive filters (Purser et al., 2003a and 2003b).

Wu et al. (2002), Barker et al. (2004), and Michel and Auligné (2010) explain in more detail the methods used to construct these operators.

3 Five stages to generate the background error covariance statistics (GEN_BE code version 2.0)

The general structure of the GEN_BE code version 2.0 has been designed to split the input, output, and algorithms in independent stages. The five steps, from stage 0 to 4, that model a background error covariance matrix, become independent of the choice of control variables and model input, which allows for more flexibility (Fig. 1). Stage 0 estimates the perturbations of the control variables based on variables coming from a Numerical Weather Prediction (NWP) model forecast. Stage 1 removes the mean of these perturbations and define the applied binning. Stage 2 defines the balance operator ($\mathbf{U_p}$) by estimating covariance errors between the control variables using linear regressions. Stage 3 determines the **S** operator by estimating the standard deviation that weighs the analysis increment for a given variable. It also computes the necessary parameters to spread out the information vertically ($\mathbf{U_v}$) in data assimilation processes. Stage 4 computes the horizontal length scale parameter used by the recursive filter to model correlated error on a two dimensional plane ($\mathbf{U_h}$). Technical details are presented in three Appendices. Appendix A describes the new features of the codes and should help to compute and to implement new modeling of **B**. Appendix B presents the namelist options and Appendix C explains how to compile and run the code.

Figures shown in the following Sect. were obtained from a numerical experiment with the Advanced Research WRF (WRF-ARW, called WRF hereafter) model involving an ensemble of 50 members (D-ensemble) over the CONtiguous United States (CONUS) domain at 15 km resolution (Res. 15 km Fig. 2). Figure 3 shows the Pressure (hPa) against vertical model levels. Each member, is a six hour forecast valid at 12:00z on 3 June 2012. The Ensemble

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted:), which are affordable approximations of horizontal diffusion.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: - The matrix U_v defines the vertical auto-correlations for each of the u control variables. It is modeled by either homogeneous Empirical Orthogonal Functions (EOFs) or the applications of recursive filter.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English

titi toto 1/9/15 4:07 AM
Deleted: <#>Motivation for an updated background error covariance model
The addition of new control variables in data assimilation systems requires the estimation of the error variances for each field, the calculation of the regression coefficients to derive uncorrelated (and unbalanced) control variables, and the estimation of the parameters to model vertical and horizontal correlated errors.
The structure of the GEN_BE code version 2.0 has been designed to perform those operations efficiently, to gather different methods to model B and make to additional developments easier. The new version of the code allows modeling a real time configuration of B like NCEP does using five control variables (CV5, e.g. Sect.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, German

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Deleted:), as well as, diagnosing and implementing a new model of B. The set of control variables has been expanded to include hydrometeors (CV9, e.g. Sect.

Adjustement Kalman Filter (EAKF), coming from the community system Data Assimilation Research Testbed (DART, Anderson et al. 2009), was used by Romine et al. (2014) to generate the analysis ensemble. Table 2, shown in Sect. 4, contains detailed setup information of this data assimilation experiment.

3.1 Sampling and binning (stage 0 and stage 1)

Since the background error covariance matrix is a statistical entity, samples of model forecasts are required to estimate the associated variances and correlations. Traditionally, two distinct techniques are used and available in stage 0 to compute the perturbations:

- Differences between two forecasts valid at the same time but initiated at different dates (time lagged forecast, e.g. 24 hour minus 12 hour forecasts), can be used to represent a sample of model background errors. This is an *ad hoc* technique, called the NMC (named for the National Meteorological Center) method (Parish and Derber, 1992), which has been widely used in operational centers where large databases of historical forecasts are available.

- Background error statistics can be evaluated from an ensemble of perturbations valid at the same time (Fisher, 2003; Pereira and Berre, 2006). This method tends to be more accurate because it better represents the background error of the day, rather than a climatological error, as with the NMC method. However, more computational resources are required to run an ensemble simulation and it may not provide automatically the optimum B for a particular system (Fisher 2003).

Pereira and Berre (2006) highlight the consequences of the evaluation of perturbations using the NMC method versus an ensemble approach (called ensemble of the day, D-ensemble). The authors point out that the NMC method tends to underestimate the background errors in data-sparse areas (when the forecast comes from cycling analysis). They show that correlation length scales, as described by Daley (1991), are smaller in D-ensemble methods compared to NMC. Table B1 summarizes the general options to compute these raw perturbations.

Since the number of sample of perturbations can be limited, a strategy to model a static error covariance over an entire domain and filter the sampling noise is used. The statistics are spatially averaged by gathering grid points with similar characteristics. The different options available for this technique, referred as binning, are described in Table B2, and can be setup in the namelist input file (Table B3). The simplest way to compute statistics for a domain can be done by vertical levels (*bin_type*=5). Moreover, such formulation of B, which allows

Formatted ... [66]
titi toto 1/9/15 4:07 AM
Deleted:) in order to assimilate cloudy ra ... [67]
Unknown
Field Code Changed ... [68]
titi toto 1/9/15 4:07 AM
Formatted ... [69]
titi toto 1/9/15 4:07 AM
Formatted ... [70]
titi toto 1/9/15 4:07 AM
Formatted ... [71]
titi toto 1/9/15 4:07 AM
Formatted ... [72]
titi toto 1/9/15 4:07 AM
Deleted: -
titi toto 1/9/15 4:07 AM
Formatted ... [73]
titi toto 1/9/15 4:07 AM
Deleted: -
titi toto 1/9/15 4:07 AM
Formatted ... [74]
titi toto 1/9/15 4:07 AM
Formatted ... [75]
titi toto 1/9/15 4:07 AM
Deleted: previous forecasts
titi toto 1/9/15 4:07 AM
Formatted ... [76]
titi toto 1/9/15 4:07 AM
Formatted ... [77]
titi toto 1/9/15 4:07 AM
Deleted:)
titi toto 1/9/15 4:07 AM
Formatted ... [78]
titi toto 1/9/15 4:07 AM
Deleted: Table 1
titi toto 1/9/15 4:07 AM
Formatted ... [79]
titi toto 1/9/15 4:07 AM
Deleted: different
titi toto 1/9/15 4:07 AM
Formatted ... [80]
titi toto 1/9/15 4:07 AM
Deleted: Some kind of spatial averaging n ... [81]
titi toto 1/9/15 4:07 AM
Formatted ... [82]
titi toto 1/9/15 4:07 AM
Deleted: available are
titi toto 1/9/15 4:07 AM
Formatted ... [83]
titi toto 1/9/15 4:07 AM
Deleted: and to reduce
titi toto 1/9/15 4:07 AM
Formatted ... [84]
titi toto 1/9/15 4:07 AM
Deleted: dimensional of statistical output ... [85]
titi toto 1/9/15 4:07 AM
Formatted ... [86]
titi toto 1/9/15 4:07 AM
Deleted: the Table 2
titi toto 1/9/15 4:07 AM
Formatted ... [87]
titi toto 1/9/15 4:07 AM
Deleted: Table 3). Their goals are to add ... [88]
titi toto 1/9/15 4:07 AM
Formatted ... [89]

modeling of homogeneous and istropic covariance, may be inadequate to specify natural phenomena. Other binning option can be applied to the different transform **U**$_p$, **U**$_v$, **U**$_h$ and **S** to have a heterogeneous formulation of B. For example, options *bin_type*=1, 2, 3, 4 compute statistics across the zonally averaged ensemble perturbations, to create a latitude-dependent correlation function, usually used for large and global domains where latitude flow dependency occurs (Wu et al., 2002). For example, the statistics of hydrometeors, as cloud liquid water, which are characterized by a high spatial and temporal variability can be skewed (Michel et al., 2011) if, at a given grid point, only few members of the D-ensemble indicate the presence of clouds. For that reason, it may be preferable to use a cloud mask in the hydrometeor cloud calculations, which is referred as “geographical binning“. Montmerle and Berre (2010) and Michel et al. (2011) show improvements using rain mask (option 7) with the vorticity and divergence control variables to characterize convection events.

For this reason, the GEN_BE code has been modified to facilitate the introduction of new binning options for specific applications (see Appendix B). Stage 1 removes the mean of the perturbations and defines the binning which is an important component in the model of **B** as it is applied in the following stages, especially in stage 2 for the balance operator.

3.2 Balance through linear regressions (stage2)

Analysis increment for one variable may impact an another if they have correlated errors. The simplest way to model these multivariate error cross-covariances is to use linear regressions that mimic physical balance between variables. First, regression coefficient between variables can be estimated by solving Eq. (6) following the example of the regression of the temperature (t) by the stream function (psi):

$$\alpha_{psi,t}(b,k,l)\bullet VAR_{psi}(b,k)=COVAR_{psi,t}(b,k,l) \tag{6}$$

Where $\alpha_{psi,t}$ is the regression coefficient estimated, $COVAR_{psi,t}(b,k,l)$ represents the vertical cross-covariance between t and psi averaged over the vertical level k, l for the given binning class index b, and $VAR_{psi}(b,k)$ is the variance.

In practice, the regression coefficient can be directly calculated as the ratio of the inverted variance with the covariance or by performing a Cholesky decomposition (see Appendix B for more details). Then, linear regressions are performed to derive uncorrelated (i.e. unbalanced) perturbations by removing the balanced part from other perturbation

titi toto 1/9/15 4:07 AM
Formatted ... [90]

titi toto 1/9/15 4:07 AM
Deleted: statitics

titi toto 1/9/15 4:07 AM
Formatted ... [91]

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted ... [92]

titi toto 1/9/15 4:07 AM
Deleted: . Morever, binning can become t ... [93]

titi toto 1/9/15 4:07 AM
Formatted ... [94]

titi toto 1/9/15 4:07 AM
Formatted ... [95]

titi toto 1/9/15 4:07 AM
Deleted: . All the algorithms

titi toto 1/9/15 4:07 AM
Formatted ... [96]

titi toto 1/9/15 4:07 AM
Deleted: different stages from 1 to 4 do n ... [97]

titi toto 1/9/15 4:07 AM
Formatted ... [98]

titi toto 1/9/15 4:07 AM
Deleted: option used. Stage 1 creates the ... [99]

titi toto 1/9/15 4:07 AM
Formatted ... [100]

titi toto 1/9/15 4:07 AM
Deleted: encapsulates this informat ... [101]

titi toto 1/9/15 4:07 AM
Formatted ... [102]

titi toto 1/9/15 4:07 AM
Deleted: different programs. This ... [103]

titi toto 1/9/15 4:07 AM
Formatted ... [104]

titi toto 1/9/15 4:07 AM
Deleted: introduction of a new binn ... [105]

titi toto 1/9/15 4:07 AM
Formatted ... [106]

titi toto 1/9/15 4:07 AM
Deleted: developer has to introduce

titi toto 1/9/15 4:07 AM
Formatted ... [107]

titi toto 1/9/15 4:07 AM
Deleted: method to update

titi toto 1/9/15 4:07 AM
Formatted ... [108]

titi toto 1/9/15 4:07 AM
Deleted: mask of binning in the rou ... [109]

titi toto 1/9/15 4:07 AM
Formatted ... [110]

titi toto 1/9/15 4:07 AM
Deleted: module io_input.f90. The binni ... [111]

variables. Eq. (7) shows how the unbalanced part of the t perturbation (δt_u) is deduced by substrating its full perturbation (δt) to its balanced part coming from psi:

$$\delta t_u(i,j,k)=\delta t(i,j,k)-\sum_{l=1}^{N_k}\alpha_{psi,t}(b,k,l)\delta psi(i,j,l) \qquad (7)$$

where *b* is the index of the binning class according to the triplet indexes of the grid point position *(i,j,k)*. N_k is the total number of vertical model levels.

Note, that in variational data assimilation process, balance operator $\mathbf{U_p}$ is applied to the variable themselves. It models correlations between variables and allows to transform the B matrix as a block diagonal in the control (uncorrelated) space. The GEN_BE code version 2.0, has been developed to allow the use of a broad set of control variables (shown in Table 1**) and to allow the definition of the $\mathbf{U_p}$ transform in a namelist input file. For example, Table B4 presents how to define the balance transform that involves five control variables** (CV5) as it can be used in the GSI system developed at NCEP for analyses operational purpose (Kleist et al., 2009). The parameters *covar* equals 1 means the unbalanced part of the velocity potential (chi_u), the temperature (t_u), and the pressure surface (ps_u) are calculated by substracting their balanced part coming from the stream function (psi). Benchmark results of pseudo temperature test involving different modeling of **B** and the same $\mathbf{U_p}$ transform (CV5) are shown Sect 4.

Futhermore, Bannister (2008b) described the $\mathbf{U_p}$ transform used in different operational centers with special emphasis on the definition of the balance operator for humidity. To determine a balance operator, diagnostics of vertical cross-covariance or vertical cross-correlation are helpful to analyze the relationship between variables and can also be done through stage 2. For example, Fig. 4 shows the cross-correlation between humidity and temperature for all atmosphere conditions (mixing dry and wet conditions). The errors are mostly anti-correlated, and specific humidity (Fig. 4a) has weaker correlated errors with respect to temperature than relative humidity (Fig. 4b). Moreover, the errors between specific humidity and temperature become highly correlated close to saturation (Holm et al., 2002; Ménétrier and Montmerle, 2011). At saturation, these statistics likely rely on processes of condensation and precipitation when the released latent heat flux warms the atmosphere (Holm et al., 2002). These characteristics highlight how binning that differentiates background statistics in the presence of clouds can be important according to the choice of

Formatted ... [112]
titi toto 1/9/15 4:07 AM
Deleted: regression coefficients bet ... [113]
titi toto 1/9/15 4:07 AM
Deleted: regression coefficient as a ... [114]
titi toto 1/9/15 4:07 AM
Formatted ... [115]
titi toto 1/9/15 4:07 AM
Formatted ... [116]
titi toto 1/9/15 4:07 AM
Deleted: uses a Choleski decomposition ... [117]
titi toto 1/9/15 4:07 AM
Formatted ... [118]
titi toto 1/9/15 4:07 AM
Deleted: specify and differentiate statisti ... [119]
titi toto 1/9/15 4:07 AM
Formatted ... [120]
titi toto 1/9/15 4:07 AM
Deleted: balanced part for each other
titi toto 1/9/15 4:07 AM
Formatted ... [121]
titi toto 1/9/15 4:07 AM
Deleted: as shown in Sect. 3.2.2. This pa ... [122]
titi toto 1/9/15 4:07 AM
Formatted ... [123]
titi toto 1/9/15 4:07 AM
Deleted: a
titi toto 1/9/15 4:07 AM
Deleted: bloc
titi toto 1/9/15 4:07 AM
Formatted ... [124]
titi toto 1/9/15 4:07 AM
Formatted ... [125]
titi toto 1/9/15 4:07 AM
Deleted: structure of the code, and speci ... [126]
titi toto 1/9/15 4:07 AM
Formatted ... [127]
titi toto 1/9/15 4:07 AM
Deleted: the goals is to have
titi toto 1/9/15 4:07 AM
Formatted ... [128]
titi toto 1/9/15 4:07 AM
Deleted: code flexible enough to diagnos ... [129]
titi toto 1/9/15 4:07 AM
Formatted ... [130]
titi toto 1/9/15 4:07 AM
Deleted: different
titi toto 1/9/15 4:07 AM
Formatted ... [131]
titi toto 1/9/15 4:07 AM
Deleted: ,
titi toto 1/9/15 4:07 AM
Formatted ... [132]
Unknown
Field Code Changed ... [133]
titi toto 1/9/15 4:07 AM
Deleted: , that models specific covarianc ... [134]
titi toto 1/9/15 4:07 AM
Formatted ... [135]
titi toto 1/9/15 4:07 AM
Deleted: Adding new control variables o ... [136]
titi toto 1/9/15 4:07 AM
Formatted ... [137]
titi toto 1/9/15 4:07 AM
Deleted: . ... [138]
titi toto 1/9/15 4:07 AM
Formatted ... [139]
titi toto 1/9/15 4:07 AM
... [140]
titi toto 1/9/15 4:07 AM
Formatted ... [141]
titi toto 1/9/15 4:07 AM

control variables. Thus, various studies have been dedicated to better estimate the background error of humidity in cloudy areas (Carron and Fillon, 2010; Montmerle et Berre 2010; Ménétrier and Montmerle 2011). Carron and Fillon (2010) use the specific humidity (qs) and show benefit to characterize heterogeneous formulation of **B** defined for dry and precipitation areas. For a winter test-case where stratiform-type precipitation is predominant, they explain that geostrophic imbalance in precipitation areas, can be characterized by the linear balance operator between the stream function and the mass fields (t and ps). Montmerle and Berre (2010) show potential improvements at convective scale by using a rainy mask in a multivariate approach for specific humidity that involves vorticity, divergence, temperature and surface pressure variables. While Ménétrier and Montmerle (2011) show the benefit of balancing the specific humidity only with the mass fields (t and ps) for fog data assimilation purposes. Dynamical variables such as vorticity and divergence are not included in the balance humidity operator since they do not drive fog formation processes.

Finally, result of an experiment that include hydrometeors and its correlated errors with humidity (CV9) are presented Sect. 5.1 and defined by the namelist input file Table B5.

3.3 Estimation of the vertical correlation and the variance (stage3)

After calculating the vertical auto-covariance matrix (VACM), two techniques are currently available in stage 3 to compute the parameters useful to model the mean vertical auto-correlation transform ($\mathbf{U}_v$). The first method diagonalizes the VACM, performing an EOF decomposition (i.e. computing eigenvectors and eigenvalues). The variable is re-written in this new base for each EOF. Stage 4 will later evaluate a length scale for each EOF mode. The vertical transform occurs with the change of base EOF-physical space and the variances are represented by the eigenvalues. The second method estimates, a vertical length scale from the vertical auto-correlation matrix directly in the physical space, to propagate the increment via recursive filters. The diagnostic of the vertical length scale (L_v) comes from Daley's formula (1991, p110) for a one dimension homogeneous and isotropic case:

$$L_v = \sqrt{\frac{1}{\frac{\partial^2 \rho(0)}{\partial^2 z}}} \qquad (8a)$$

with $\rho(0)$ the correlation taken at the origin.

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Heading 2

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, 12 pt, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: ,

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, 12 pt, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: (aka. EOFs)

titi toto 1/9/15 4:07 AM
Formatted ... [142]

titi toto 1/9/15 4:07 AM
Formatted ... [143]

titi toto 1/9/15 4:07 AM
Deleted: the

titi toto 1/9/15 4:07 AM
Formatted ... [144]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English

titi toto 1/9/15 4:07 AM
Deleted: ... [145]

Unknown
Formatted: Font:Times New Roman, 12

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted ... [146]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, German

titi toto 1/9/15 4:07 AM
Formatted ... [147]

Unknown
Formatted ... [148]

titi toto 1/9/15 4:07 AM
Formatted ... [149]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: Doing a Taylor Development o ... [150]

Unknown
Formatted ... [151]

$\rho(\delta x)$

Approximating Eq (8a) with finite difference to the second order derivatives of $\rho(\delta z)$ and assuming ρ *is* symmetric around the origin results in:

$$L_{vp} = \frac{\delta z}{\sqrt{2[1-\rho(\delta z)]}} \quad (8b)$$

where L_{vp} represents the vertical length scale approximate by a parabolic function.

If the correlation is approximated at the origin by a Gaussian function as follows:

$$\rho(\delta z) = \exp(-\frac{\delta z}{2L_{vg}^2}) \quad (9a)$$

the length scale L_{vg} can be written:

$$L_{vg} = \frac{\delta z}{\sqrt{-2\ln\rho(\delta z)}} \quad (9b)$$

Pannekoucke et al. (2008) studied the sensitivity of sampling errors of these formulae and shows that the Gaussian and the parabolic approximation give similar results. Furthermore, the vertical length scale can be computed uniform by vertical model level or binned. Table B6 in Appendix B contains description of the namelist option to define the vertical length scale in stage 3 and the horizontal length scale in stage 4.

3.4 Estimation of the horizontal correlation (stage 4)

Horizontal auto-correlations can be computed for each control variable at each grid point. Figure 5 shows a diagnostic of correlation for a few selected points of the WRF computational domain around 500 m above the ground (model level 5). The stream function (5a) and velocity potential control variables have larger and more isotropic spatial correlations while the temperature (5b) and the humidity (5c) control variables show smaller and anisotropic correlations at different locations. The radius of the area where the correlation overpasses 0.9 is within a range of 100 km to 400 km for stream function while this radius reaches its maximum around 100 km for temperature and humidity. Hydrometeors mixing ratio show even more local structures due to their sparse location on the horizontal and the vertical (5d).

Formatted ... [152]
titi toto 1/9/15 4:07 AM
Formatted ... [153]
titi toto 1/9/15 4:07 AM
Deleted: subsisting in Eq. (4), it
Unknown
Formatted ... [154]
titi toto 1/9/15 4:07 AM
Formatted ... [155]
titi toto 1/9/15 4:07 AM
Deleted: ... [156]
Unknown
Field Code Changed ... [157]
titi toto 1/9/15 4:07 AM
Formatted ... [158]
titi toto 1/9/15 4:07 AM
Deleted: with the
titi toto 1/9/15 4:07 AM
Formatted ... [159]
titi toto 1/9/15 4:07 AM
Deleted: approximation.
titi toto 1/9/15 4:07 AM
Formatted ... [160]
titi toto 1/9/15 4:07 AM
can be
titi toto 1/9/15 4:07 AM
Formatted ... [161]
titi toto 1/9/15 4:07 AM
following,
titi toto 1/9/15 4:07 AM
Formatted ... [162]
titi toto 1/9/15 4:07 AM
Deleted:
Unknown
Formatted ... [163]
titi toto 1/9/15 4:07 AM
Formatted ... [164]
titi toto 1/9/15 4:07 AM
Deleted: expression for the Gaussian ... [165]
titi toto 1/9/15 4:07 AM
Formatted ... [166]
titi toto 1/9/15 4:07 AM
Deleted: deduced from Eq. (5):
titi toto 1/9/15 4:07 AM
Formatted ... [167]
titi toto 1/9/15 4:07 AM
Deleted: ... [168]
Unknown
Field Code Changed ... [169]
titi toto 1/9/15 4:07 AM
Formatted ... [170]
titi toto 1/9/15 4:07 AM
Formatted ... [171]
titi toto 1/9/15 4:07 AM
Deleted: close
titi toto 1/9/15 4:07 AM
Formatted ... [172]
titi toto 1/9/15 4:07 AM
Deleted:
titi toto 1/9/15 4:07 AM
Deleted: be handled by inhomogeneous (... [173]
titi toto 1/9/15 4:07 AM
Formatted ... [174]
titi toto 1/9/15 4:07 AM
... [175]
titi toto 1/9/15 4:07 AM
Formatted ... [176]
titi toto 1/9/15 4:07 AM
Formatted ... [177]
titi toto 1/9/15 4:07 AM
Formatted ... [178]
titi toto 1/9/15 4:07 AM

In stage 4, we estimate horizontal length scales averaged by vertical level or EOF mode for a field analysis in a 2-D plane. It represents the radius of influence, calculated in grid point space, around the position of an observation and is an input parameter for recursive filters to spread out horizontally the increment (**U**$_h$). The different options available, as described below, are also contained in Table B6.

The first method (*ls_method=1*) employs a distribution function to fit the correlation for a 2-D field by vertical level or by EOF mode as explained in Sect 3.3. If a Gaussian function is chosen, the length scale is determined by solving Eq. (10a):

$$\rho(r) = \exp(-\frac{r^2}{2L}) \qquad (10a)$$

where ρ(r) is the correlation calculated for a distance r between two grid points.

If a second order autoregressive (SOAR) correlation function is used, the length scale *L* is determined by solving Eq. (10b):

$$\rho(r) = (1+\frac{r}{L})\bullet\exp(-\frac{r^2}{L}) \qquad (10b)$$

However, as this procedure is both computationally expensive and prone to sampling errors, a second option (*ls_method=2*) based on the ratio of the variance of a field (*φ*) and the variance of its laplacian, has been added:

$$L = \left(\frac{8\bullet Variance(\varphi)}{Variance(\nabla^2\varphi)}\right)^{1/4} \qquad (11)$$

Eq. (11) was used by Wu et al. (2002) and is similar to the diagnosis of Pereira and Berre (2006), which was analyzed in Pannekoucke et al. (2008).

The horizontal length scale can be uniformally calculated over a vertical model level or can be statistically binned. Homogeneous recursive filters are able to handle a unique length scale defined by model vertical level, or EOF mode. Inhomogeneous recursive filters (Purser et al., 2003b), as implemented in GSI, are able to handle heterogeneous length scale. In this case, the increment is spread out with a length scale according to the bin class of each grid point. Moreover, spatial filtering to smooth the length scale may be required because of recursive filters normalization issues (Michel and Auligné 2010).

Formatted ... [179]
titi toto 1/9/15 4:07 AM
Deleted: ... [180]
titi toto 1/9/15 4:07 AM
Formatted ... [181]
titi toto 1/9/15 4:07 AM
Formatted ... [182]
titi toto 1/9/15 4:07 AM
Deleted: EOF mode or by
titi toto 1/9/15 4:07 AM
Formatted ... [183]
titi toto 1/9/15 4:07 AM
Deleted: .
titi toto 1/9/15 4:07 AM
Formatted ... [184]
Unknown
Field Code Changed ... [185]
titi toto 1/9/15 4:07 AM
Deleted: the parameter *horizfunc='gauss'*
titi toto 1/9/15 4:07 AM
Formatted ... [186]
titi toto 1/9/15 4:07 AM
Deleted: selected
titi toto 1/9/15 4:07 AM
Formatted ... [187]
titi toto 1/9/15 4:07 AM
Deleted: *L*
titi toto 1/9/15 4:07 AM
Deleted: the
titi toto 1/9/15 4:07 AM
Formatted ... [188]
titi toto 1/9/15 4:07 AM
Deleted: 7
titi toto 1/9/15 4:07 AM
Formatted ... [189]
titi toto 1/9/15 4:07 AM
Formatted ... [190]
titi toto 1/9/15 4:07 AM
Deleted: $\rho(r) = \rho(0)\bullet\exp(-\frac{r^2}{8}$ ... [191]
titi toto 1/9/15 4:07 AM
Formatted ... [192]
titi toto 1/9/15 4:07 AM
Deleted: the parameter *horizfunc='SOAI* ... [193]
titi toto 1/9/15 4:07 AM
Formatted ... [194]
titi toto 1/9/15 4:07 AM
Deleted: modeled by fitting
titi toto 1/9/15 4:07 AM
Formatted ... [195]
titi toto 1/9/15 4:07 AM
Deleted: pseudo correlation of a field thr ... [196]
titi toto 1/9/15 4:07 AM
Formatted ... [197]
titi toto 1/9/15 4:07 AM
Deleted: the
titi toto 1/9/15 4:07 AM
Formatted ... [198]
titi toto 1/9/15 4:07 AM
Deleted: 8
titi toto 1/9/15 4:07 AM
Formatted ... [199]
titi toto 1/9/15 4:07 AM
Deleted: $\rho(r) = \rho(0)\bullet(1+\frac{r}{L})\bullet\exp(-$ ... [200]
titi toto 1/9/15 4:07 AM
Formatted ... [201]
titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM
Formatted ... [202]
Unknown
Field Code Changed ... [203]
titi toto 1/9/15 4:07 AM

4 Comparison of different modelling of B for two data assimilation systems

We present a benchmark of different modeling of B perfomed on the GSI and WRFDA data assimilation platforms. Both systems can handle the set of five control variables (CV5) and their balance operator ($\mathbf{U}_p$) defined Table B4. By default, the GSI system allows using a **B** matrix statisitics ($\mathbf{B}_{nam}$), pre-computed over an enlarged CONUS domain, using the NMC method and NAM (North American Mesoscale) forecasts. $\mathbf{B}_{nam}$ is used with GSI (Wu, 2005) to produce daily forecasts with NDAS (NAM Data Assimilation System; Rogers et al., 2009). Based on the D-Ensemble dataset coming from the DART experiment (i.e. Sect 3. and Romine et al. 2014), we present in Sect. 4.1 the parameters that define the vertical transform $\mathbf{U}_v$ by using EOF decomposition for WRFDA ($\mathbf{B}_{eof}$) and by using recursive filters for GSI ($\mathbf{B}_{rcf}$). Table 2, gathers the general setup that leads to the modeling of these three **B** matrices ($\mathbf{B}_{eof}$, $\mathbf{B}_{rcf}$ and $\mathbf{B}_{nam}$) and additional information about the used datasets. The physics of the model can be found in Romine et al. (2014), Rogers et al. (2009). Sect. 4.2 compares the results of a pseudo single observation test experiment using $\mathbf{B}_{eof}$, $\mathbf{B}_{rcf}$ and $\mathbf{B}_{nam}$ on the WRFDA and GSI data assimilation system.

4.1 Statistics of the background error covariance matrix for different transforms

4.1.1 Decomposition by EOF and length scale

If the EOF decomposition is used, the eigenvectors model the vertical transform ($\mathbf{U}_v$) and the associated eigenvalues represent the variance. The length scale is estimated in the EOF space and represents the horizontal transform ($\mathbf{U}_h$). In the data assimilation process, the eigenvalues weight the analysis increment and the recursive filter first spreads out the information in the EOF space according to length scale value. Then, the transformation from EOF mode to physical space spreads out the information vertically. The first five eigenvectors are shown in Fig. 6 for the control variables (CV5) and Fig. 7 shows the associated eigenvalues. 99% of the variance of the stream function and the velocity potential are represented by the first ten and twenty modes respectively, while more than 30 modes are useful for temperature and relative humidity. Also, the EOF decomposition allows optionally some filtering as the largest

titi toto 1/9/15 4:07 AM
Formatted ... [204]

titi toto 1/9/15 4:07 AM
Deleted: The real time configuration

titi toto 1/9/15 4:07 AM
Deleted: used at NCEP, that includes

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted ... [205]

titi toto 1/9/15 4:07 AM
Deleted: covariance errors (

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, German

titi toto 1/9/15 4:07 AM
Deleted:), can be used in both GSI and WRFDA data assimilation systems. In the following, we present first the different parameters that define the vertical transform U_v by using EOF decomposition for WRFDA (B_{eof}) and by using recursive filter for GSI (B_{rcf}). Finally, the results of data assimilation obtained with B_{eof} and B_{rcf}, determined for the CONUS domain at 15 km of resolution, are compared with the background error B_{nam} that operates in real time on GSI on the rapid refresh domain. B_{nam} statistics are based on NAM forecast at 0.1 degree of resolution and using the NMC method.

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted ... [206]

titi toto 1/9/15 4:07 AM
Formatted ... [207]

titi toto 1/9/15 4:07 AM
Formatted ... [208]

titi toto 1/9/15 4:07 AM
Deleted: square root of the

titi toto 1/9/15 4:07 AM
Formatted ... [209]

titi toto 1/9/15 4:07 AM
Deleted: Fig. 6

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: Fig. 7

titi toto 1/9/15 4:07 AM
Formatted ... [210]

titi toto 1/9/15 4:07 AM
Deleted: velocity

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

variances (i.e. eigen values) are associated with the first EOFs, the latest EOFs may be not taken into account if they mostly represent vertical noise in the system.

The horizontal length scales, estimated by Eq. (11), are presented in Figure 8. The stream function and the velocity potential have the largest length scale value reaching 600 km (39 grid points) for the first EOF mode. While, the unbalanced temperature length scale has a strong variation for the three first EOFs passing approximately from 135 km to 30 km (9 to 2 grid points) and from there, slightly decreases from 30 km to reach 15 km (2 to 1 point grid) for the last EOF mode. Relative humidity length scale remains small, decreasing from approximately 30 km to 15 km as a function of the EOF mode. The unbalanced temperature and the relative humidity have a relatively small length scale, which means that they have more local features represented by a small radius of influence. Thus, the analysis increment from these variables will remain closer to the observation. As the horizontal length scale is associated to EOF mode and not directly related to a vertical model level and futher discussions on the association of length scale with physiscal event may be difficult.

4.1.2 Horizontal and vertical length scales defined in physical space

The horizontal correlation is modeled by the application of recursive filters based on the estimation of the horizontal length scale solving Eq. (11), applied at every vertical model level for each variable, as shown in Fig. 9. The horizontal length scales diagnosed for each control variable by vertical level (Fig. 9) or by EOF mode (Fig. 8) have the same range of value. The length scales of the stream function and the velocity potential control variables have the largest values above 150 km (10 grid points) for all the vertical model levels, while the length scales of temperature and relative humidity remain in a range of 30 km to 60 km (1 to 2 grid points) below 200 hPa level. Temperature and humidity, which have more local structures, are modeled with smaller length scales. Globally, the horizontal length scales of different variables increase from the bottom to the top of the model as they represent larger scale events. Direct comparison of these statistics with the $\mathbf{B}_{nam}$ horizontal length scale is difficult as they are performed with different methods, models, configurations, and physical options (i.e. Table 2). However, it can be noted than the horizontal length scale was approximately twice as small than those for $\mathbf{B}_{nam}$ (Wu 2005) performed by using the NMC method. Usually, sharper correlations are found in the D-ensemble compared to the NMC method (Fisher, 2003; Pereira and Berre, 2006). Furthermore, a factor contributing to this

Formatted ... [211]
titi toto 1/9/15 4:07 AM
Deleted: 6
titi toto 1/9/15 4:07 AM
Formatted ... [212]
titi toto 1/9/15 4:07 AM
Deleted: velocity
titi toto 1/9/15 4:07 AM
Formatted ... [213]
titi toto 1/9/15 4:07 AM
Deleted: , i.e. close to 600 km
titi toto 1/9/15 4:07 AM
Formatted ... [214]
titi toto 1/9/15 4:07 AM
Deleted: EOF
titi toto 1/9/15 4:07 AM
Formatted ... [215]
titi toto 1/9/15 4:07 AM
Deleted: is
titi toto 1/9/15 4:07 AM
Formatted ... [216]
titi toto 1/9/15 4:07 AM
Deleted: more monotonically
titi toto 1/9/15 4:07 AM
Formatted ... [217]
titi toto 1/9/15 4:07 AM
Formatted ... [218]
titi toto 1/9/15 4:07 AM
Deleted: 6
titi toto 1/9/15 4:07 AM
Formatted ... [219]
titi toto 1/9/15 4:07 AM
Deleted: by
titi toto 1/9/15 4:07 AM
Formatted ... [220]
titi toto 1/9/15 4:07 AM
Deleted: Fig. 9
titi toto 1/9/15 4:07 AM
Formatted ... [221]
titi toto 1/9/15 4:07 AM
Deleted: diagnostic of
titi toto 1/9/15 4:07 AM
Formatted ... [222]
titi toto 1/9/15 4:07 AM
Deleted: scale shows similar behavior to ... [223]
titi toto 1/9/15 4:07 AM
Formatted ... [224]
titi toto 1/9/15 4:07 AM
Deleted: modes
titi toto 1/9/15 4:07 AM
Formatted ... [225]
titi toto 1/9/15 4:07 AM
Deleted: velocity
titi toto 1/9/15 4:07 AM
Formatted ... [226]
titi toto 1/9/15 4:07 AM
Deleted: one
titi toto 1/9/15 4:07 AM
Formatted ... [227]
titi toto 1/9/15 4:07 AM
Deleted: two
titi toto 1/9/15 4:07 AM
Formatted ... [228]
titi toto 1/9/15 4:07 AM
Deleted: under
titi toto 1/9/15 4:07 AM
Formatted ... [229]
titi toto 1/9/15 4:07 AM
titi toto 1/9/15 4:07 AM
Formatted ... [230]
titi toto 1/9/15 4:07 AM

difference may arise from the fact that we are comparing statistics from forecasts of different lengths.

The vertical correlation is modeled by the application of recursive filters based on the estimation of the vertical length scale coming from Eq. (8b). The stream function and the velocity potential in Fig. 10 that represent large scale horizontal flow have a bigger vertical length scale than those of temperature and humidity. The vertical gradients of temperature and humidity can vary strongly locally, decreasing the vertical correlation.

4.2 Pseudo single observation test on WRFDA and GSI data assimilation systems

The single pseudo-observation is a powerful way to provide a benchmark as it allows visualizing the increment of an isolated observation and its impact on other variables. Thus, the following are pseudo observation tests of temperature with an innovation of 1 Kelvin and an obervation error of 1 Kelvin using different modeling of **B** ($\mathbf{B}_{eof}$, $\mathbf{B}_{rcf}$ and $\mathbf{B}_{nam}$). The position of the pseudo-observation is arbitrarily taken at the center of the domain and at 500 hPa pressure level. The series of plots (Figs 11-13) represent horizontal and vertical slices of the resulting increment for temperature and wind components.

As expected, the horizontal cross-section at the 500 hPa level for temperature shows an isotropic response to the innovation of 1 Kelvin. The maxima of intensity simulated depend on the standard deviation (diagonal matrix S) value coming from the **B** matrix.

On one hand, the operator ($\mathbf{U}_v$) employs EOF decomposition, the J_b term of the cost function is weighted by the standard deviation coming from the square root of the eigenvalues of $\mathbf{B}_{eof}$. On the other hand, $\mathbf{U}_v$ is modeled by the estimation of a length scale and the recursive filters applied on the vertical ($\mathbf{B}_{rcf}$), the analysis is weighted by the standard deviation directly averaged on the vertical mesh grid. The increments of temperature are close for the three different tests and the increment from $\mathbf{B}_{nam}$ is slightly larger than that of $\mathbf{B}_{rcf}$ and $\mathbf{B}_{eof}$. In the case of $\mathbf{B}_{nam}$, recursive filters spread out the information in a larger area over a horizontal plane due to its larger length scales.

For the vertical cross-section (*XZ*), vertical increments coming from $\mathbf{B}_{rcf}$ and $\mathbf{B}_{eof}$ spread out in the same range of altitude (~ between the 800 hPa and 450 hPa pressure levels). Based on the same D-ensemble datasets, the U_v operator using EOF decompostion and recursives filters gives similar results on different platforms, as expected. Moreover, the temperature increment

Formatted ... [231]

titi toto 1/9/15 4:07 AM
Deleted: Fig. 10

titi toto 1/9/15 4:07 AM
Formatted ... [232]

titi toto 1/9/15 4:07 AM
Deleted: lowering down

titi toto 1/9/15 4:07 AM
Formatted ... [233]

titi toto 1/9/15 4:07 AM
Formatted ... [234]

titi toto 1/9/15 4:07 AM
Deleted: make

titi toto 1/9/15 4:07 AM
Formatted ... [235]

titi toto 1/9/15 4:07 AM
Deleted: a series of plots for a

titi toto 1/9/15 4:07 AM
Formatted ... [236]

titi toto 1/9/15 4:07 AM
Deleted: test

titi toto 1/9/15 4:07 AM
Formatted ... [237]

titi toto 1/9/15 4:07 AM
Deleted: the observation

titi toto 1/9/15 4:07 AM
Formatted ... [238]

titi toto 1/9/15 4:07 AM
Deleted: slice done

titi toto 1/9/15 4:07 AM
Formatted ... [239]

titi toto 1/9/15 4:07 AM
Deleted: the

titi toto 1/9/15 4:07 AM
Formatted ... [240]

titi toto 1/9/15 4:07 AM
Formatted ... [241]

titi toto 1/9/15 4:07 AM
Deleted: pseudo observation

titi toto 1/9/15 4:07 AM
Formatted ... [242]

titi toto 1/9/15 4:07 AM
Deleted: maximum

titi toto 1/9/15 4:07 AM
Formatted ... [243]

titi toto 1/9/15 4:07 AM
Deleted: with this pseudo observation depends

titi toto 1/9/15 4:07 AM
Formatted ... [244]

titi toto 1/9/15 4:07 AM
Deleted: variance

titi toto 1/9/15 4:07 AM
Formatted ... [245]

titi toto 1/9/15 4:07 AM
Deleted: When

titi toto 1/9/15 4:07 AM
Formatted ... [246]

titi toto 1/9/15 4:07 AM
Deleted: employed the

titi toto 1/9/15 4:07 AM
Formatted ... [247]

titi toto 1/9/15 4:07 AM
Deleted: computed

titi toto 1/9/15 4:07 AM
Formatted ... [248]

titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM
Formatted ... [249]

titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM

from $\mathbf{B}_{rcf}$ spreads out even more along the vertical compared to the $\mathbf{B}_{nam}$ experiment on the GSI system. This discrepancy can be associated with the computed vertical length scales from two different datasets. The length scales diagnosed over a D-ensemble are larger in this case for $\mathbf{B}_{rcf}$ than the one averaged over a long period of time in the NMC method (60 perturbations selected over a year). Also, statistics of $\mathbf{B}_{nam}$ are perfomed over an Eta grid of 60 vertical levels of WRF-NMM while the statistics of $\mathbf{B}_{rcf}$ and $\mathbf{B}_{eof}$ come from WRF defined on a hybrid-sigma grid of 39 vertical levels. Thus, the raw statistics of $\mathbf{B}_{nam}$ are interpolated on the WRF vertical grid in GSI before performing 3D-VAR data assimilation. Furthermore, differences in the definition of the physics of the model and the assimilated data may be contributing factors.

Finally, the multivariate approach, defined by CV5, induces increments in the wind components. The horizontal cross-section *(XY)* plotted for *U* and *V* showed dipole lobes, which can be explained by the geostrophic balance adjustment that the linear cross-covariances statistics reproduce. The vertical cross-section *(XZ)* follows the isocontour of 0 m s^{-1} for *U* while some differences can be observed on the slices of *V* for the $\mathbf{B}_{eof}$, $\mathbf{B}_{rcf}$, and $\mathbf{B}_{nam}$ experiments. A larger spread of the V increment along pressure levels is observed for $\mathbf{B}_{eof}$ and $\mathbf{B}_{rcf}$ compared to experiment of $\mathbf{B}_{nam}$.

These ensemble based background error $\mathbf{B}_{eof}$ and $\mathbf{B}_{rcf}$ covariance matrices potentially have more skill in estimating error statistics related to the present meteorological event and using the same model configuration.

Formatted ... [250]
titi toto 1/9/15 4:07 AM
Deleted: model these B matrices. By con ... [251]
titi toto 1/9/15 4:07 AM
Deleted: more climatological as they are ... [252]
titi toto 1/9/15 4:07 AM
Formatted ... [253]
titi toto 1/9/15 4:07 AM
Formatted ... [254]
titi toto 1/9/15 4:07 AM
Deleted: and they are interpolated on the mesh
titi toto 1/9/15 4:07 AM
Deleted: domain of our test case during t ... [255]
titi toto 1/9/15 4:07 AM
Formatted ... [256]
titi toto 1/9/15 4:07 AM
Formatted ... [257]
titi toto 1/9/15 4:07 AM
Deleted: coming from an ensemble
titi toto 1/9/15 4:07 AM
Formatted ... [258]
titi toto 1/9/15 4:07 AM
Deleted: day directly related to the mete ... [259]
titi toto 1/9/15 4:07 AM
Formatted ... [260]
titi toto 1/9/15 4:07 AM
Formatted ... [261]
titi toto 1/9/15 4:07 AM
Deleted: adjustement
titi toto 1/9/15 4:07 AM
Formatted ... [262]
titi toto 1/9/15 4:07 AM
Deleted: plan
titi toto 1/9/15 4:07 AM
Formatted ... [263]
titi toto 1/9/15 4:07 AM
Deleted: /
titi toto 1/9/15 4:07 AM
Formatted ... [264]
titi toto 1/9/15 4:07 AM
Deleted: more complex structures
titi toto 1/9/15 4:07 AM
Formatted ... [265]
titi toto 1/9/15 4:07 AM
Deleted: for
titi toto 1/9/15 4:07 AM
Formatted ... [266]
titi toto 1/9/15 4:07 AM
Deleted: . ... [267]
titi toto 1/9/15 4:07 AM
Formatted ... [268]
titi toto 1/9/15 4:07 AM
Deleted: GSI system, and B_{eof} under WR ... [269]
titi toto 1/9/15 4:07 AM
Formatted ... [270]
titi toto 1/9/15 4:07 AM
Deleted: been specifically determined.
titi toto 1/9/15 4:07 AM
Formatted ... [271]
titi toto 1/9/15 4:07 AM
Deleted:

5 Cloud and chemistry variational data assimilation

5.1 Generation of a multivariate background error covariance for hydrometeors

Code modifications have been done in the WRFDA code to add a multivariate balance operator for the hydrometeor variables: cloud liquid water mixing ratio (q_{cloud}), rain mixing ratio (q_{rain}), ice mixing ratio (q_{ice}), snow mixing ratio (q_{snow}), so that the WRFDA minimization is now performed over nine 3-D fields instead of the five previously included. The main scientific issue in this task is to define a proper **B** matrix and particularly, the cross-correlation terms that will ensure that the analysis of the hydrometeors is multivariate i.e. the observed and unobserved model fields are modified simultaneously and consistently during the analysis. The question of the estimation of the forecast error covariance matrix is the focus of this section. Figure 3 provides the conversion from vertical model level to pressure level.

5.1.1 Definition of the Balance operator for hydrometeors (CV9)

The $\mathbf{U}_{\text{p}}$ transform CV5 (defined Table B4) is modified in the WRFDA code to include a multivariate analysis for humidity and hydrometeors (Eq. 12a-c). In a first approach, relative humidity (rh) is balanced in Eq. (12a) with the mass fields of unbalanced temperature (t_{u}), unbalanced surface pressure (ps_{u}) and does not include dynamic variables such as the stream function (psi) and unbalanced velocity potential (chi_{u}):

$$\text{rh}_{\text{u}}(i,j,k)=\text{rh}(i,j,k)-\sum_{l=1}^{N_k}\alpha_{\text{rh},t_u}(b,k,l)t_u(i,j,l)-\alpha_{\text{rh,ps}_{\text{u}}}(b,k)\text{ps}_{\text{u}}(i,j) \qquad (12a)$$

The statitics coming from GEN_BE v2.0 code, i.e. regression coefficients and unbalance part of the variable can be estimated only by modifying the namelist file input. In this case, the line covar5 of Table B5 that describes the covariances between the fifth control variable, (relative humidity), with the third control variables t_{u} and the fourth ps_{u} is: *covar5 = 0, 0, 1, 1, 0, 0, 0, 0, 0, 0*. In the meantime, the control variables are expanded to include the mixing ratios of cloud water condensate (q_{cloud}), rain (q_{rain}), ice (q_{ice}) and snow (q_{snow}). The hydrometeors q_{cloud} and q_{ice} are balanced with respect to relative humidity as their presence or absence is directly related. The regression coefficients can be computed without any

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Heading 2, Indent: Left: 0", Hanging: 0.4"

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: Modifications code

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: to in WRFDA

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: (Table 7),

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: 14

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Formatted: Font:Times New Roman, Not Bold, Font color: Text 1, Lowered by 14 pt

assumptions (Figs. 14a-b), or filtered to take into account the perturbations that represent the transition of a non-cloudy to a cloudy area only (Figs 14c-d). This latter choice is made to intensify the statistical relationship of the statistical balance to be able to remove misplaced clouds, or to create clouds. However, we may want to localized this balance around a given vertical model level. For this reason, the line *covar6 = 0, 0, 0, 0, 1, 0, 0, 0, 0, 0* represented by Eq. (12b) can be replaced by the line *covar6 = 0, 0, 0, 0, 2, 0, 0, 0, 0, 0* represented by the Eq. (12c). In this case, only the diagonal terms of the regression coefficient are calculated and the increment is spread out by the recursive filters.

$$q_{\mathrm{cloud}_\mathrm{u}}(i,j,k) = q_{\mathrm{cloud}}(i,j,k) - \sum_{l=1}^{N_k} \alpha_{q_{\mathrm{cloud}_\mathrm{u}},\mathrm{rh}_\mathrm{u}}(b,k,l)\mathrm{rh}_\mathrm{u}(i,j,l) \qquad (12b)$$

$$q_{\mathrm{cloud}_\mathrm{u}}(i,j,k) = q_{\mathrm{cloud}}(i,j,k) - \alpha_{q_{\mathrm{cloud}_\mathrm{u}},\mathrm{rh}_\mathrm{u}}(b,k)\mathrm{rh}_\mathrm{u}(i,j,k) \qquad (12c)$$

Similar balance is applied to q_{ice}. q_{rain} and q_{snow} are defined univariate. Table B5 summarizes the definition of this balance operator called CV9

5.1.2 Statistics of the background error covariance matrix for hydrometeors.

The vertical and horizontal transforms retained are the recursive filters making the interpretation of the length scale parameter easier as they are directly associated to a vertical model level. The four main hydrometeors have been added in this study, as they could be useful for data assimilation in remote sensing such as satellite cloudy radiances and radar reflectivity.

The horizontal length scale values of the different hydrometeors shown in Fig. 15a are smaller in comparison of other control variables (less than 30 km, 2 grid points). Significant values of length scale, that overpass 15 km (1 grid point), are related to the presence of hydrometeors: it occurs below 150 hPa pressure level for q_{ice} and q_{snow} and below 400 hPa pressure level for q_{cloud} and q_{ice}. The maximum of q_{cloud} length scale, located approximately at 950 hPa, can be associated to the presence of low maritime clouds above the Pacific ocean noted by the high standard deviation in Figs 18a and b. In the lower levels of the model, the length scale of q_{ice} vanishes as expected.

Formatted ... [272]
Unknown
Formatted ... [273]
titi toto 1/9/15 4:07 AM
Formatted ... [274]
titi toto 1/9/15 4:07 AM
Formatted ... [275]
titi toto 1/9/15 4:07 AM
Formatted ... [276]
titi toto 1/9/15 4:07 AM
Deleted: analysis
titi toto 1/9/15 4:07 AM
Formatted ... [277]
titi toto 1/9/15 4:07 AM
Formatted ... [278]
titi toto 1/9/15 4:07 AM
Deleted: Fig. 15a do not overpass two gr ... [279]
titi toto 1/9/15 4:07 AM
Formatted ... [280]
titi toto 1/9/15 4:07 AM
Deleted: than that
titi toto 1/9/15 4:07 AM
Formatted ... [281]
titi toto 1/9/15 4:07 AM
Deleted: one
titi toto 1/9/15 4:07 AM
Formatted ... [282]
titi toto 1/9/15 4:07 AM
Deleted: (15km
titi toto 1/9/15 4:07 AM
Formatted ... [283]
titi toto 1/9/15 4:07 AM
Deleted: under
titi toto 1/9/15 4:07 AM
Formatted ... [284]
titi toto 1/9/15 4:07 AM
Deleted: under
titi toto 1/9/15 4:07 AM
Formatted ... [285]
titi toto 1/9/15 4:07 AM
Deleted: around
titi toto 1/9/15 4:07 AM
Formatted ... [286]
titi toto 1/9/15 4:07 AM
Deleted: remarkable
titi toto 1/9/15 4:07 AM
Formatted ... [287]
titi toto 1/9/15 4:07 AM
Deleted: heigh
titi toto 1/9/15 4:07 AM
Formatted ... [288]
titi toto 1/9/15 4:07 AM
Deleted: Fig. 18a-
titi toto 1/9/15 4:07 AM
Formatted ... [289]

The vertical correlation maxima of the precipitating hydrometeors are higher compared to that of cloud water, or cloud ice hydrometeors as they can drop freely through multiple levels (Fig. 16a). The vertical length scale of q_{rain} increases regularly from around 500 hPa until reaching a maximum at the ground. As the length scale increases fast after 800 hPa, where the highest density of the lower levels occurs, an arbitrary cut-off equal to one third of the total vertical grid point value is applied in order to avoid spreading out increment information outside the area of potential presence of rain with the recursive filter. The length scale of q_{snow} has two local maxima. The first one happens where the precipitating hydrometeors have the highest density at around 400 hPa. A steep increase occurs from 950 hPa until reaching the highest value close to the ground. The high rate of presence of snow mixing ratio equal to zero at these low levels tends to artificially enforce vertical correlation as well.

5.1.3 Example of a pseudo single observation of cloud mixing ratio in a multivariate approach.

To verify that our analysis is multivariate, we conducted a series of tests in which pseudo observations of hydrometeors were assimilated into WRFDA and the corresponding analysis increment was plotted. Figure 17 shows the analysis response for the q_{cloud} and q_{vapor} model variables when three simulated observations of cloud liquid water are assimilated. One obervation is taken over the Pacific ocean, a second one over Texas and the last one in Canada.

The intensity of the increment can be weighted by the 1-D variance or by the 3-D variance (S operator) coming from the ensemble. The 1-D variance, displayed in Fig. 18a, gives a general information by vertical level and binning type without any assumption of horizontal location. It is most of the time used when the perturbations come from the NMC method or when the variance is not diagnosed at the analysis time. In our test case, the increment is modulated by the 3-D variance computed from a 6-hour ensemble forecast with 50 members. The cloudy area coming from the background of the different members is represented by a high value of variance in Fig 18b while low variance takes place in the dry area. The increment is most likely greater than 10^{-3} g/kg where the variability of cloud presence exists (Fig. 17). The strongest increment occurs over the Pacific Ocean for higher q_{cloud} standard deviation. A minimum value would likely need to be set to retain possibility of increments in the dry area.

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: Fig. 16

titi toto 1/9/15 4:07 AM
Deleted: (level 18)

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Heading 3, Indent: Left: 0", First line: 0", Tabs: 0.75", List tab

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: To verify that our analysis is multivariate, we conducted a series of tests in which pseudo observations of hydrometeors were assimilated into WRFDA and the corresponding analysis increment was plotted. Fig. 17 shows the analysis response for the q_{cloud} and q_{vapor} model variables when three simulated observations of cloud liquid water are assimilated. One obervation is taken over the Pacific ocean, a second one over Texas and the last one in Canada.
The intensity of the increment can be weighted by the 1-D variance or by the 3-D variance (S operator) coming from the ensemble. The 1-D variance, displayed in Fig. 18a, gives a general information by vertical level and binning type without any assumption of horizontal location. It is most of the time used when the perturbations come from the NMC method or when the variance is not diagnosed for the analysis time. In our test case, the increment is modulated by the 3-D variance computed from a 6-hour ensemble forecast with 50 members. The cloudy area coming from the background of the different members is represented by a high value of variance in Fig. 18b while low variance takes place in the dry area. The increment is most likely important where the variability of cloud presence exists, as over the Pacific Ocean (Fig. 17). A minimum value would probably need to be set to retain the possibility of increments in the dry area.
The covariance between the mixing ratio of cloud water condensate and relative humidity, described in Sect. 3.2.2, can reinforce the possibility to add clouds in the dry area or to remove clouds in the cloudy background area. The univariate version of the balance for hydrometeors is beneficial at the analysis time as it allows including increments of hydrometeors directly at the analysis time. The multivariate balance is present to help to propagate the q_{cloud} increment in the forecast by adding q_{vapor}. The increments of temperature, due to the multivariate balance between humidity and temperature, are not significant.
The determination of the balance of humidity and hydrometeors is a difficult task as it involv ... [290]

The covariance between mixing ratio of cloud water condensate and relative humidity, described in Sect. 5.1.1 can reinforce the ability of adding clouds in the dry area or removing clouds in the cloudy area. The univariate version of the balance operator for hydrometeors may be beneficial at the analysis time as hydrometeors can be directly assimilated. The multivariate balance is present to help to propagate the q_{cloud} increment in the forecast by balancing it with a q_{vapor} increment.

The determination of the balance of humidity and hydrometeors is a difficult task as it involves the microphysical processes of meteorological NWP models and different local phenomena. The use of local covariances coming from the D-ensemble may help to balance those high sensible variables. Furthermore, operational centers, such as Météo-France with the Application of Research to Operations at Mesoscale system (AROME, Seity et al., 2011) and the Met Office with the Met Office Global and Regional Ensemble Prediction System (MOGREPS, Bowler et al., 2008; Migliorini et al., 2011), already use ensemble forecasts at high resolution to more accurately characterize specific meteorological events, such as precipitation and convection. Nowadays, their ensemble size remains small (often less then 10 members) because the cost of CPU (Central Processing Unit) time is still elevated. Studies have been dedicated to evaluate the sampling errors in the ensemble method and in the parameters, such as correlation length scales, that usually model the background errors (Pannekoucke et al., 2008; Ménétrier et al., 2014). When the ensemble size is small, methods that combine general statistics of the background errors and local balance are found to perform better (Hamill and Snyder, 2000). Figures 15a, b and 16a, b, that display horizontal and vertical length scales parameters respectively, for the hydrometeors in regards of the number of members, show stable results.

5.2 Background Error for Chemical Species

As a proof of concept, this last section shows the direct applicability of the GEN_BE v2.0 code as a diagnostic tool for other topics than meteorology. In recent decades, a large number of studies that investigate chemical data assimilation have been conducted. Some of the first studies on stratospheric and tropospheric chemistry data assimilation were performed roughly two decades ago (e.g. Austin, 1992; Fisher and Lary, 1995; and Elbern et al., 1997). During the last two decades, efforts have been made in order to improve atmospheric chemical modeling and data assimilation scheme performances.

The well characterization of the background error covariance matrix **B** in chemistry is a very important aspect of a successful data assimilation system. During the last few years, different studies have used different techniques to characterize the **B** matrix. Barré et al. (2013) and Emili et al. (2014) estimated a quasi-constant **B** based on the Ménard and Chang (2000) and Desroziers et al. (2005) a posteriori statistics, for tropospheric and stratospheric ozone data assimialtion. Since the latter studies put their interests on large-scale events (global scale chemical assimilation and synoptic events) data assimilation perform reasonably well with those first order **B** matrix estimation. Depending on the region of the atmosphere that is analyzed **B** needs to be updated at different timescales. Massart et al. (2012) showed the importance of using a monthly **B** matrix ensemble estimate for stratospheric ozone data assimilation purposes. For surface ozone assimilation Jaumouillé et al. (2012) and Gaubert et al. (2014) showed that an hourly ensemble estimate of **B** that represent diurnal variations of model errors improves the data assimilation skills. The last few years, studies on aerosol data assimilation within WRF-Chem (Pagowski et al., 2010, 2014, Schwartz et al., 2012) showed the importance of having a detailed estimation of the **B** matrix.

Statistics were analyzed in detail to ensure that **B** reproduced relevant correlation structures during data assimilation process. Since data assimilation of chemical species is more recent than for meteorology, the GEN_BE code version 2.0 may be useful to test new definitions of background error covariance matrices and to allow its usage on different platforms. Several chemical trace gases such as CO (Carbon Monoxide), NO_x (Nitrogen Oxides) and O_3 (Ozone) but also dust, sea salt, particulate matter (PM) have been already included as new possible control variables in the GEN_BE code. Results for CO, NO_x and O_3 are shown next. The statistics are estimated using 20 members over the CONUS domain. Each member comes from a 12h forecast of WRF-CHEM (WRF model coupled with Chemistry, Grell et al., 2005), valid at 12:00z on 14 June 2008, at 36 km of horizontal resolution and 33 vertical levels. The lateral boundary conditions coming from MOZART (Model for OZone And Related chemical Tracers, Emmons et al., 2010) and factors coming from MEGAN (Model of Emissions of Gases and Aerosols from Nature, Guenther et al., 2006) are perturbed using a pseudo-normal random noise. In order to avoid unphysical or negative values of concentration and emissions and keep ensemble mean boundary conditions values close to the original values, we then perturb the boundary conditions (emissions and boundary conditions) by using a standard deviation (sigma) of 25% of the original boundary condition value and we limit the perturbation to be no more than 3 sigma (i.e. 75%).

Figure 19 present the standard deviations for the chemical species of interest. Standard deviation of the background error is directly related to the species concentrations. Most of the ozone variability takes place in the middle atmosphere (stratosphere) on the ozone layer around 100 hPa (Fig. 19a). Concerning NO_x concentration fluctuates as well in Figs 19b and 19c, due to photochemistry in the stratosphere and in the troposphere. Because the NO_x are also emitted from the ground with a short lifetime, a strong peak of standard deviation is observed. Carbon monoxide (Fig. 19d), which is also emitted at the surface and has relatively long life time (1-2 months), show significant standard deviation values in all the troposphere with a maximum in the boundary layer.

Figure 20 displays the calculated horizontal chemical length scales. Ozone show horizontal length scales are around 100 km in the troposphere and around 125 km in the stratosphere. Pagaowski et al., 2010, used a NMC method and found that ozone horizontal length scale are around 100 km (150 km) in the troposphere (in the stratosphere). Concerning NO_2, GEN_BE v2.0 evaluates the tropospheric horizontal length scale between 70 km and 90 km. This range of values is consistent with the values found by Silver et al., 2013 that uses the NMC method. Horizontal length scales increase in the upper troposphere mostly due to the strong circulation (jets) and then advection of trace gases that increase the horizontal correlations.

Concerning the vertical correlations (Fig. 21), all the 4 species diagnosed, present a maximum close to the surface where they are emitted or secondarily produced for ozone. Then, they sharply decrease between 1000 hPa and 850 hPa. This strong decrease is not fully understood and need further work to be conclusive. A first hypothesis to explain this strong decrease would be caused by the reactions with the other short-lived species emission perturbations and create strong correlation in the lowest model levels. Another factor explaining this decrease would be the strong increase of first model levels layer thickness close to the surface. Vertical correlation then also decrease around 800 hPa due to weaker vertical mixing above the planetary boundary layer height and creates a decrease of correlations with the lower levels. Above 850 hPa, which is around the top of the boundary layer, the evolution of the vertical length scale decreases slowly from approximately 2 to 1 grid point. Then in the free troposphere, vertical diffusion of possible data assimilation increments will be less significant than in the boundary layer. Compared to Pagowski et al., 2010, the ozone vertical length scale profile present the same behavior. Strong vertival correlation close to the surface, followed by a strong decrease to the levels directly above and then a lower values in the upper levels of the boundary layer.

Here we have shown that the GEN_BE v2.0 code is able to model a **B** matrix for chemical variables with features that are associated with physical processes i.e. ozone layer, tracer lifetime, emissions and planetary boundary layer mixing. The diagnostics of simple statistics of the background for chemical species are straight forward with the GEN_BE code version 2.0. Moreover, data assimilation of chemistry components remains a challenge because of the uncertainties of various parameters that predict chemical processes as emission factors, deposition velocity and (photochemical) reaction constant. For these reasons, the analysis may fit the observation even if data assimilation does not involve the origin of the mismatch. Hybrid and ensemble methods may help to diagnose complex covariance structures in future work. In this paper, the chemical **B** matrix generated by GEN_BE v2.0 has not been extensively diagnosed. More investigations such as, the balance between chemical species, standard deviation and correlation length time and space variability could be investigated in further studies by the atmospheric chemistry modeling community using GEN_BE v2.0.

6 Summary and discussions

While variational methods have been successfully used in operational centers for a long time, the estimation of background errors needs to be continuously improved to assimilate new observations and to provide more accurate statistics. The GEN_BE v2.0 code has been developed to investigate and model univariate or multivariate covariance errors from control variables defined by a user as an input. It gathers some methods and options that can be easily applied to different model inputs and used on different data assimilation platforms by extending its former capabilities. The flexibility of the framework of the GEN_BE V2.0 code should help the diagnostics of correlated errors and the implementation of new background error modeling.

This document describes first the different stages and transforms that lead to the modeling of the background error covariance matrix **B** by performing benchmark tests and showing examples that use these new functionalities based on WRF and WRF-CHEM forecasts. Parameters such as length scales, eigenvectors, eigenvalues, standard deviation and linear regression coefficients were first estimated for the control variables (CV5) described in Kleist et al. (2009) for the GSI system developed at NCEP.

Second, the GEN_BE v2.0 code has been validated through multivariate single observation tests of temperature using three different modeling of **B** ($\mathbf{B}_{eof}$, $\mathbf{B}_{ref}$, and $\mathbf{B}_{nam}$) and on two different platforms. Based on the first dataset, D-ensemble, the single observation test performed with $\mathbf{B}_{eof}$ ($\mathbf{U}_v$, EOF decompostion) in WRFDA shows similar results than the single observation test of temperature performed with $\mathbf{B}_{ref}$ ($\mathbf{U}_v$ recursive filters) in GSI. The increments were spread out in a larger area along the vertical than those coming from the test using the $\mathbf{B}_{nam}$ statitiics calculated with the NMC method on a different vertical grid. While, the horizontal increments were spread out in a larger area using $\mathbf{B}_{nam}$.

Third, the GEN_BE code has been used to perform the statistics over an extended set of control variables that include mixing ratio of hydrometeors (CV9) for multivariate cloud data assimilation purpose. As clouds have an intermittent presence, the 3-D variance coming from an ensemble of the day gives a spatial envelope useful to weight the analysis relatively to the observation and the background confidence. The hydrometeors of cloud and ice condensate water are also balanced with humidity to be potentially able to create or remove misplaced clouds. The regression coefficients calculated, can be conserved for a next cycle analysis as they are averaged by bins or recalculated as they are not so expansive with regard to CPU time. In this paper, a pseudo observation test of cloud mixing ratio was performed using WRFDA and the next step is to test cloudy radiance data assimilation. Finally, statistics of background are estimated for chemical species such as carbon monoxide (CO), nitrogen oxides (NO_x) and ozone (O_3) coming from an ensemble of forecasts of WRF-CHEM, discussed and compared with existant studies. It has been shown that the statistics diagnosed are related to physical and chemical processes.

In these previous examples, GEN_BE code version 2.0 can handle input datasets coming from WRF, a model defined on a C-Arakawa grid, and the background error statistic outputs are computed on unstaggered A-Arakawa grid. Within minor modifications, the code would be able to handle other horizontal grids. Also, statitics could easily be done on models with different vertical grid definition. If we consider performing the background errors statistics on an unstructured grid, the structure of the code can remain the same but few mathematical operators, such as differential and laplacian, and estimation of the distance between two grid points, would need to be re-defined according to the grid. In fact, the $\mathbf{U_p}$ transform needs to be performed in the unstructured grid according to the user's choice of control variables. $\mathbf{U_v}$ transform will remain identical and $\mathbf{U_h}$ transform would be modified according to the mathematical operators. Another option would be to interpolate first the input dataset on a regular grid according to the data assimilation system used and then compute the statistics. Thus, implementation of models with different grid can be done in the GEN_BE v2.0 code based on its general framework and may be completed by adding new diagnostics.

The current trend is to model a more complex background error, expanding the control variables and correlated errors and using techniques to achieve more heterogeneity and anisotrpy. The geographical binning and the 3-D variance available in the GEN_BE v2.0 code can be utilized with new data assimilation algorithms. For example, hybrid data assimilation that combines variational and ensemble methods may be helpful especially by adding flow dependence in the estimation of the background error and to reduce the ensemble size due to CPU time constraints (Hamill and Snyder, 2000). Wang et al. (2008a, 2008b) performed a study using a hybrid 3DVAR-ETKF (Ensemble Transform Kalman Filter) technique that combines static (modeled) error and ensemble error covariances. Better results were obtained over North America at a coarse resolution (200 km) especially in data-sparse areas compared to those performed solely with 3DVAR. The extended control variable technique (Lorenc 2003) allows blending flow dependent errors with static covariance errors. Bannister et al. (2011b) investigated the benefit of a convection permitting prediction system ensemble (24 members) at a finer scale (i.e. 1.5 km of resolution) for nowcasting purposes based on MOGREPS (Migliorini et al. 2011a). Even though, the authors show how general balances that drive synoptic flow, in particular geostrophic balance, can diminish in convective situations at small scales, they highlight the necessity for a data assimilation system to better represent both the large scale and mesoscale components of the flow. In addition, Ménétrier et al. (2014) studied heterogeneous flow dependent background error covariances at a convective scale and showed that a small ensemble (6 members from AROME) contains relevant information with sampling noise, which can be reduced through filtering. Finally, the GEN_BE code may be a tool to diagnose inhomogeneous 3-D localization parameters in ensemble methods. The GEN_BE v2.0 code has been tested in atmospheric science but the flexibility of the code may be useful in other geophysical applications.

Appendix A: FORTRAN code and input/output description.

New FORTRAN modules have been developed to generalize the calculation of the error covariance matrix from different input models and for new control variables. Table A1 contains a complete list of these modules and their contents. All the algorithms from stage 1 to stage 4 are now independent of the choice of control variables and driven by a unique namelist file, called namelist.input, and read by the FORTRAN module configure.f90. Flexibility has been added for future experiments. Only few modifications are needed in stage 0 to add new control variables. The FORTRAN module io_input_models.f90 converts the standard variables from a given model to the analysis variables. The interface is already made with the WRF model. Only the FORTRAN module io_input_model.f90 needs to be updated to implement new model input and to run the different stages. The NetCDF format has been chosen to improve robustness and flexibility in the input and output of the different stages as shown in Table A2. The final NetCDF output file be.nc contains all the information needed for a variational data assimilation system, as shown in Table A3. Several converters from NetCDF format to binary have been developed to ensure backward compatibility to another data assimilation system. A binary file be.dat can be generated for the WRFDA application using the program gen_be_diags.f90 and a binary file be_gsi.dat can be created for GSI using the converter gen_be_nc2gsi.f90.

Appendix B: Description of the namelist options.

The “namelist.input” file drives the different stages 0 to 4 contains four different sections.

The namelist section "*&gen_be_info*", described in Table B1, defines the options to compute perturbations in stages 0 and 1 from input forecast model (e.g. WRF). Also, the data assimilation system can specified.

Table B2, presents eight binning available option and Table B3 explains how to set up the namelist section *"&gen_be_bin"*. In the GEN_BE code version 2.0, all the information that defines a binning option are encapsulated in the type bins_type. Since the algorithms of the different stages from 1 to 4 do not make any specific assumption on the binning option used, the implementation of new option is simplified as it needs to be defined just once in the *da_create_bins* FORTRAN routine of the module *io_input.f90*. In case of implementing geographical mask, developers have to introduce the method to update the mask in the *update_dynamical_mask* routine. All information related to binning is contained in the NetCDF file bin.nc created in stage 1.

The $\mathbf{U_p}$ transform is defined in section *"&gen_be_cv"* where the used control variables and balance operator are set up. Table B4 presents the CV5 control variable currently used in the GSI system (Kleist et al. 2009). In this example, the use of the relative humidity (rh) (line covar5) allows to performed statistics in GEN_BE for the normalized relative humidity described by Holm (2002) and implemented in GSI. Furthermore, when the regression coefficients are computed for a GSI regional application, a Cholesky decomposition is used and additional filtering is applied to the regression coefficients between stream function and temperature, and between stream function and pressure surface. This part of the code coming from the NCEP is flagged with *use_cholesky* variable in the gen_be_stage2.F FORTRAN program, and the called subroutines are contained in the *io_output_applications.f90* Fortran module. Table B5 shows the $\mathbf{U_p}$ transform, called CV9, which includes hydrometeors in a multivariate approach.

Table B6 contains the namelist section *"&gen_be_lenscale"* to diagnose parameters of the $\mathbf{U_v}$ and $\mathbf{U_h}$ transforms for stage 3 and stage 4 respectively. The vertical transform $\mathbf{U_v}$ can be performed by estimating a vertical length scale by model levels (*data_on_level=true*) or by a EOF decomposition (*data_on_level=false*). By default, statitics are binned with the same option defined section *"&gen_be_bin"* of the namelist.input file. Otherwise, the statistics are averaged by vertical level if the flag *global_bin* is true (which is equivalent to the defininition of *bin_type=5*).

Appendix C: Installation, compilation, set up and visualization.

The GEN_BE code version 2.0 is a standalone package that can be installed on different UNIX/LINUX systems. It has been tested with the Intel FORTRAN compiler, the Portland Group FORTRAN compiler, and the GNU FORTRAN compiler. It requires compilation of NetCDF libraries. First, a configuration file needs to be created using the command *configure* in the main directory of the code. Then, the compilation, is launched by the command *compile gen_be*. Once successfuly completed, the executables are created in the src directory.

Korn-shell scripts available in the scripts directory allow to setup the experiment. The wrapper script, named gen_be_wrapper.ksh, sets up some global variables and launches the main script gen_be.ksh. The user needs to setup most of the other options that determine the way to model the **B** matrix in the namelist.template file. The gen_be.ksh script fills out the initial date and the final dates, the frequency of date available (interval) coming from the global variables setup in the wrapper script and in the gen_be_set_defaults.ksh script, and generates a namelist.input file in the working directory during the first stage. The namelist.input file contains four main parts presented in Appendix B. Each stage can then be run successively by setting the environmental variable RUN_GEN_BE_STAGE [0,1,2,3,4] to true in the gen_be_set_defaults.ksh script. The output of the stages 0, 1, 2, 3 and the be.nc file can be easily visualized with existing tools (Ncview, NCL, Python, MatLab).

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

Acknowledgements

Funding for this work was provided by the U.S. Air Force Weather Agency. The authors benefited from numerous discussions with Yann Michel. Glen Romine thanked for providing the ensemble over the CONUS domain. Syed Rizvi is thanked for discussions concerning the previous version of the code.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: and Jerome Barré are

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: conus

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

References

Anderson, J., Hoar ,T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano A.: The data assimilation research testbed: A community facility, B. Am. Meteorol. Soc., 90, 1283-1296., doi: http://dx.doi.org/10.1175/2009BAMS2618.1, 2009.

Auligné, T., Lorenc A., Michel, Y., Montmerle, T., Jones, A., Hu, M. and Dudhia, J.: Toward a New Cloud Analysis and Prediction System, B. Am. Meteorol. Soc., 92, 207-210, doi: http://dx.doi.org/10.1175/2010BAMS2978.1, 2011.

Austin, J.: Toward the 4-dimensional assimilation of stratospheric chemical-constituents, J. Geophys. Res., 97, 2569–2588, 1992.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characterisitics and measurements of forecast error covariances, Q. J. Roy. Meteor. Soc. 134: 1951-1970, doi: 10.1002/qj.339, 2008a.

Bannister, R. N.: A review of forecast error covariance statistics in atmospheric variational data assimilation. II: Modelling the forecast error statistics, Q. J. Roy. Meteor. Soc. 134: 1971-1996, doi: 10.1002/qj.340, 2008b.

Bannister, R., Migliorini, S. and Dixon, M: Ensemble prediction with a convection-permitting model for nowcasting. Part II: Forecast error statistics. Tellus 63A, 497–51, DOI: 10.1111/j.1600-0870.2010.00500.x, 2011.

Barker, D. M., Huang, W., Guo, Y. R and Xiao, Q. N.: A Three-Dimensional (3DVAR) data assimilation system for use with MM5: implementation and initial results, Mon. Weather Rev. 132:897-914, 2004.

Barker, D. M., Huang, X. Y., Liu, Z., Auligné, T., Zhang, X., Rugg, S., Ajjaji, R. , Bourgeois, A., Bray, J., Chen, Y., Demirtas, M., Guo, Y. R., Henderson T., Huang, W., H. Lin, C., Michalakes, J., Rizvi S., and Zhang, X. The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA, B. Am. Meteorol. Soc, 93, 831–843, doi: http://dx.doi.org/10.1175/BAMS-D-11-00167.1, 2012.

Barré, J., Peuch, V.-H., Lahoz, W., A., Attié, J.-L., Josse, B., Piacentini, A., Eremenko, M., Dufour, G., Nedelec, P., von Clarmann, T. and El Amraoui, L.: Combined data assimilation of ozone tropospheric columns and stratospheric profiles in a high-resolution CTM, Q. J. Roy. Meteorol. Soc., 140: 966–981, doi: 10.1002/qj.2176, 2014.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted ... [291]

titi toto 1/9/15 4:07 AM
Formatted ... [292]

titi toto 1/9/15 4:07 AM
Formatted ... [293]

titi toto 1/9/15 4:07 AM
Formatted ... [294]

titi toto 1/9/15 4:07 AM
Formatted ... [295]

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted ... [296]

titi toto 1/9/15 4:07 AM
Deleted: Benedetti,

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: and Fisher

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .: Background error statistics for aerosols,

titi toto 1/9/15 4:07 AM
Formatted ... [297]

titi toto 1/9/15 4:07 AM
Deleted: . 133: 391-405,

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: :

titi toto 1/9/15 4:07 AM
Formatted: Font:(Default) Times New Roman, Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: 37, 2007.

Bowler, N. E., Arribas, A., Mylne, K. R., Robertson, K. B. and Beare, S. E: The MOGREPS short-range ensemble prediction system. Q. J. Roy. Meteorol. Soc. 134, 703–722, DOI: 10.1002/qj.234, 2008.

Caron, J. F. and Fillion L.: An Examination of Background Error Correlations between Mass and Rotational Wind over Precipitation Regions. Mon. Weather Rev., 138 (2), 563–578, doi: http://dx.doi.org/10.1175/2009MWR2998.1, 2010.

Courtier P., Thépaut, J. N. and Hollingsworth A., A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteor. Soc. (1994), 120, pp.1367-1387, 1994.

Daley, R.: Atmospheric Data Analysis. Cambridge Univeristy Press, 1991.

Davies, T., Cullen, M., J., P., Malcolm, A., Mawson, M., Staniforth., A., White, A. and Wood, N.:A new dynamical core for the Met Office's global and regional modelling of the atmosphere. Q. J. R. Meteorol. Soc. 131: 1759–1782, doi: 10.1256/qj.04.101, 2005.

Desroziers, G., Berre, L., Chapnik, B. and Poli, P. (2005), Diagnosis of observation, background and analysis-error statistics in observation space. Q. J. Roy. Meteorol. Soc., 131: 3385–3396. doi: 10.1256/qj.05.108

Elbern, H., Schimdt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys Res. Rev., 102, 15 967–15 985, 1997.

Emili, E., Barret, B., Massart, S., Le Flochmoen, E., Piacentini, A., El Amraoui, L., Pannekoucke, O., and Cariolle, D.: Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical transport model, Atmos. Chem. Phys., 14, 177-198, doi:10.5194/acp-14-177-2014, 2014.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., Orlando, J., Tie, X., Tyndall, G., Wiedinmyer, C., Baughcum, S. L., and Kloster, S.: Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4), Geosci. Model Dev., 3, 43-67, doi:10.5194/gmd-3-43-2010, 2010.

Fisher, M., 2003: Background error covariance modelling. Proceedings of the ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean, 45-63, 8-12 September 2003.

titi toto 1/9/15 4:07 AM
Formatted: Font:Times, 10 pt, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Space Before: 0 pt

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: Elbel

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Fisher, M. and Lary, D. J.: Lagrangian four-dimensional variational data assimilation of chemical species, Q. J. Roy. Meteor. Soc., 121, 1681–1704, 1995.

Hamill, T. M. and Snyder, C.: A Hybrid Ensemble Kalman Filter–3D Variational Analysis Scheme, Mon. Weather Rev., 128, 2905–2919, doi: http://dx.doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2, 2000.

Holm, E., Andersson, E., Beljaars, A., Lopez, P., Mahfouf, J-F., Simmons, A.J. and Thepaut, J-N. Assimilation and Modelling of the Hydrological Cycle: ECMWF's Status and Plans, Technical Memoranda 383, 2002.

Gaubert, B., Coman, a., Foret, G., Meleux, F., Ung, A., Rouil, L., Ionescu, A., Candau, Y., and Beekmann, M.: Regional scale ozone data assimilation using an ensemble Kalman filter and the CHIMERE chemical transport model. Geoscientific Model Development, 7(1), 283–302. doi:10.5194/gmd-7-283-2014, 2014.

Grell, G.A., Dudhia J., and Stauffer D.: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). NCAR Technical Note NCAR/TN-398+STR, DOI: 10.5065/D60Z716B, 1994.

Grell, G. A., Peckham, S. E., Schmitz, R., McKeen, S. A., Frost, G. J., Skamarock, W., and Eder, B.: Fully-coupled online chemistry within the WRF model, Atmos. Environ., 39, 6957–6975, 2005.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I., and Geron, C.: Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6, 3181-3210, doi:10.5194/acp-6-3181-2006, 2006.

Jaumouillé, E., Massart, S., Piacentini, A., Cariolle, D., & Peuch, V.-H.: Impact of a time-dependent background error covariance matrix on air quality analysis. Geoscientific Model Development, 5(5), 1075–1090. doi:10.5194/gmd-5-1075-2012, 2012.

Klesit, D., T., Parrish, D., F., Derber, J., C., Treadon, R., Wu, W. S. and Lord S.: Introduction of the GSI into the NCEP Global Data Assimilation System, Mon. Weather Rev., 24, 1691–1705, http://journals.ametsoc.org/doi/pdf/10.1175/2009WAF2222201.1, 2009.

Lorenc, A., C.: The potential of the ensemble Kalman Filter for NWP-A comparison with 4-D VAR. Q. J. R. Meteor. Soc. 595: 3183-3203, DOI: 10.1256/qj.02.132, 2003.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Massart, S., Piacentini, A., and Pannekoucke, O.: Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses, Q. J. Roy. Meteor. Soc., 138, 889–905, doi: 10.1002/qj.971, 2012.

Ménard R, Chang LP. 2000. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part II: χ2-validated results and analysis of variance and correlation dynamics. Mon. Weather Rev. 128: 2672–2686.

Ménétrier, B., and Montmerle, T.: Heterogeneous background-error covariances for the analysis and forecast of fog events, Q. J. Roy. Meteor. Soc., 137, 2004–2013, doi: 10.1002/qj.802 , 2011.

Ménétrier, B., Montmerle, T., Berre, L. and Michel Y.: Estimation and diagnosis of heterogeneous flow-dependent background-error covariances at the convective scale using either large or small ensembles, Q. J. Roy. Meteor. Soc., 140, 2050–2061, DOI:10.1002/qj.2267, 2014.

Michel Y. and Auligné T.: Inhomogeneous Background Error Modeling Over Antarctica. Mon. Weather Rev., 138, 6, pp. 2229–2252, doi: http://dx.doi.org/10.1175/2009MWR3139.1, 2010.

Michel Y., Auligné T. and Montmerle T.: Heterogeneous convective-scale Background Error Covariances with the inclusion of hydrometeor variables. Mon. Weather Rev., 139, 9, 2994-3015, doi: http://dx.doi.org/10.1175/2011MWR3632.1, 2011.

Migliorini, S., Dixon, M., A., G., Bannister, R., N., and Ballard, S., P.: Ensemble prediction for nowcasting with a convection-permitting model – Part I: description of the system and the impact of radar- derived surface precipitation rates. Tellus 63A, 468–496, DOI:10.1111/j.1600- 0870.2010.00503.x, 2011.

Montmerle, T., and Berre, L.: Diagnosis and formulation of heterogeneous background error covariances at mesoscale, Q. J. Roy. Meteor. Soc. 136:1408-1420, doi: 10.1002/qj.655, 2010.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D.: Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting – Chemistry model and Grid-point

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Statistical Interpolation, Q. J. Roy. Meteorol. Soc., 136, 2013–2024, doi:10.1002/qj.700, 2010.

Pagowski, M., Liu, Z., Grell, G. A., Hu, M., Lin, H.-C., and Schwartz, C. S.: Implementation of aerosol assimilation in Gridpoint Statistical Interpolation (v. 3.2) and WRF-Chem (v. 3.4.1), Geosci. Model Dev., 7, 1621-1627, doi:10.5194/gmd-7-1621-2014, 2014.

Pannekoucke, O., Berre, L., and Desroziers, G.: Background-error correlation length-scale estimates and their sampling statistics, Q. J. R. Meteor. Soc. 134: 497-508, doi: 10.1002/qj.212, 2008.

Parrish, D. F., and J. C. Derber: The National Meteorological Center's Spectral Statistical-interpolation Analysis System. Mon. Weather Rev., 120, 1747-1763, 1992.

Pereira, M. B., and Berre, L.: The Use of an Ensemble Approach to Study the Background Error Covariances in a Global NWP Model, Mon. Weather Rev., 134, 2466–2489, doi: http://dx.doi.org/10.1175/MWR3189.1, 2006.

Purser, R., J., Wu, W., S., Parrish, D., F., and Roberts, N., M.: Numerical aspects of the application of recursive filters to variational statistical analysis, Part I: Spatially homogeneous and isotropic Gaussian covariances. Mon. Weather Rev., 131, 1524–1535, doi: http://dx.doi.org/10.1175//1520-0493(2003)131<1524:NAOTAO>2.0.CO;2, 2003a.

Purser, R., J., Wu, W., S., Parrish D., F., and Roberts, N., M.: Numerical Aspects of the Application of Recursive Filters to Variational Statistical Analysis, Part II: Spatially Inhomogeneous and Anisotropic General Covariances. Mon. Weather Rev., 131, 1536–1548, , doi: http://dx.doi.org/10.1175//2543.1, 2003b.

Rogers, E., DiMego, G., Black, T., Ek, M., Ferrier, B., Gayno, G., Janjic, Z., Lin, Y., Pyle, M., Wong, V., Wu, W. S., and Carley, J.: The NCEP North American Mesoscale Modeling System: Recent changes and future plans, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction

, https://ams.confex.com/ams/23WAF19NWP/techprogram/paper_154114.htm, 2009

Romine, G., S., Schwartz C., S., Berner J., Fossell, R., K., Snyder C., Anderson J. and Weisman M., L.: Representing forecast error in a convection-permitting ensemble system, Mon. Weather Rev., doi: http://dx.doi.org/10.1175/MWR-D-14-00100.1, 2014.

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: , 2003b.
Romine G., Weisman M., Manning K., Wang W., Schwartz C., Anderson J. and Snyder C.: The Use of WRF-DART Analyses for 3 km Explicit Convective Forecasts in Support of the 2012 DC3 Field Program. 13th WRF Users Workshop, Boulder, CO, June 26-29 2012, WS2012/5.2, 2012.
Wu, W., S., Purser, R., J., and Parrish, D., F.: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances, Mon. Weather Rev., 130, 2905–2916, doi:

Unknown
Field Code Changed

titi toto 1/9/15 4:07 AM
Deleted: http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2

titi toto 1/9/15 4:07 AM
Deleted: , 2002.

Page Break

Table 1. General information defining the experiment in the namelist iput file (&gen_be_info part).
&gen_be_info ... [298]

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V., 2011: The AROME-France Convective-Scale Operational Model. Mon. Weather Rev., 139, 976–991, doi: http://dx.doi.org/10.1175/2010MWR3425.1, 2011.

Skamarock, W., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D., Duda, M. G., Huang, X. Y., and Wang, W., 2008: A Description of the Advanced Research WRF Version 3. NCAR Technical Note NCAR/TN-475+STR, doi: 10.5065/D68S4MVH, 2008.

Schwartz, C. S., Liu, Z., Lin, H. C., and McKeen, S. A.: Simulta- neous three-dimensional variational assimilation of surface fine particulate matter and MODIS aerosol optical depth, J. Geophys. Res., 117, D13202, doi:10.1029/2011JD017383, 2012.

Silver, J. D., Brandt, J., Hvidberg, M., Frydendall, J., and Christensen, J. H.: Assimilation of OMI NO2 retrievals into the limited-area chemistry-transport model DEHM (V2009.0) with a 3-D OI algorithm, Geosci. Model Dev., 6, 1-16, doi:10.5194/gmd-6-1-2013, 2013.

Wang, X., D., M., Barker, C. Snyder, Hamill, T., M.: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Mon. Wea. Rev*., 136, 5116-5131, doi: http://dx.doi.org/10.1175/2008MWR2444.1, 2008.

Wang, X., D., M., Barker, C. Snyder, Hamill, T., M.: A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part II: real observation experiments. *Mon. Wea. Rev*., 136, 5132-5147. doi: http://dx.doi.org/10.1175/2008MWR2445.1, 2008.

Wu, W., S.: Background error for NCEP's GSI analysis in regional mode, Proc 4th WMO International Symposium on Assimilations of Observations in Meteorology and Oceanography, Prague, Czech Republic, 2005.

Wu, W., S., Purser, R., J., and Parrish, D., F.: Three-Dimensional Variational Analysis with Spatially Inhomogeneous Covariances, Mon. Weather Rev., 130, 2905–2916, doi: http://dx.doi.org/10.1175/1520-0493(2002)130<2905:TDVAWS>2.0.CO;2, 2002.

Table 1: Description of the control variables available for the meteorology.

Nomenclature of the control variables	Description
psi	Stream function (ψ)
chi	Velocity potential (χ)
vor	Vorticity
div	Divergence
u	Horizontal wind component in x direction
v	Horizontal wind component in the y direction
t	Temperature
ps	Surface pressure
rh	Relative humidity
qs	Specific humidity
q_{cloud}	Cloud water mixing ratio
q_{rain}	Rain water mixing ratio
q_{ice}	Ice mixing ratio
q_{snow}	Snow mixing ratio
sst	Sea Surface Temperature

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted ... [299]

titi toto 1/9/15 4:07 AM
Formatted ... [300]

titi toto 1/9/15 4:07 AM
Formatted ... [301]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted ... [302]

titi toto 1/9/15 4:07 AM
Formatted ... [303]

titi toto 1/9/15 4:07 AM
Formatted ... [304]

titi toto 1/9/15 4:07 AM
Formatted ... [305]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted:

Table 2: Description of the setup of the background error matrix modelling diagnosed over the CONUS Domain. **B**$_{eof}$ and **B**$_{rcf}$ are diagnosed using GEN_BE code version 2.0 and the D-Ensemble method while **B**$_{nam}$ is performed by NCEP using the NMC method.

	B Modeling	
	B$_{eof}$ and **B**$_{rcf}$	**B**$_{nam}$
Model configuration	WRF model, resolution 15 km, 39 verticals levels on sigma hybrid grid	WRF-NMM model, resolution 12 km, 60 verticals levels on eta grid
Data assimilation setup	DART, EAKF with adaptative covariance inflation, cycling period of 6 h, Perturbated Boundary conditions from GFS, assimilation of conventional and cloud track winds observations	NDAS-GSI system, cycling period of 3 h, boundary conditions from GFS, assimilation of conventional and satellite radiances (clear-sky) observations
Method to compute the perturbations	D-ensemble method applied to 50 pertubations coming from 6 h forecasts of the different members of the ensemble	NMC method applied to 60 perturbations taken over a year, coming from time-lagged forecasts of 12 h and 24 h valid at the same time
B transforms	CV5 control variables **B**$_{rcf}$: **U**$_h$ and **U**$_v$ transforms defined by recursives filters **B**$_{eof}$: **U**$_v$ transform defined by EOF decomposition Statistics of **B**$_{rcf}$ et **B**$_{eof}$ averaged by vertical level.	CV5 control variables **B**$_{nam}$: **U**$_h$ and **U**$_v$ transforms defined by recursives filters Statistics **B**$_{nam}$ binned by latitude band of 1 degree

Deleted: . Description of the options ava ... [306]
titi toto 1/9/15 4:07 AM
Formatted ... [307]
titi toto 1/9/15 4:07 AM
Deleted: The statistics can be computed ... [308]
titi toto 1/9/15 4:07 AM
Formatted ... [309]
titi toto 1/9/15 4:07 AM
Formatted ... [310]
titi toto 1/9/15 4:07 AM
Deleted: 'true'
titi toto 1/9/15 4:07 AM
Formatted ... [311]
titi toto 1/9/15 4:07 AM
Formatted ... [312]
titi toto 1/9/15 4:07 AM
Formatted Table ... [313]
titi toto 1/9/15 4:07 AM
Deleted: vert_ls_method
titi toto 1/9/15 4:07 AM
Formatted ... [314]
titi toto 1/9/15 4:07 AM
Formatted ... [315]
titi toto 1/9/15 4:07 AM
Deleted: 1, 2
titi toto 1/9/15 4:07 AM
Deleted: Estimate the vertical length sca ... [316]
titi toto 1/9/15 4:07 AM
Formatted ... [317]
titi toto 1/9/15 4:07 AM
Formatted ... [318]
titi toto 1/9/15 4:07 AM
Deleted: ls_method
titi toto 1/9/15 4:07 AM
Formatted ... [319]
titi toto 1/9/15 4:07 AM
Formatted ... [320]
titi toto 1/9/15 4:07 AM
Deleted: 1, 2
titi toto 1/9/15 4:07 AM
Deleted: Estimate length scale (stage 3) ... [321]
titi toto 1/9/15 4:07 AM
Formatted ... [322]
titi toto 1/9/15 4:07 AM
Formatted ... [323]
titi toto 1/9/15 4:07 AM
Deleted: use_med_ls
titi toto 1/9/15 4:07 AM
Formatted ... [324]
titi toto 1/9/15 4:07 AM
Deleted: 'true'
titi toto 1/9/15 4:07 AM
Deleted: Estimate the length using the m ... [325]
titi toto 1/9/15 4:07 AM
Formatted ... [326]
titi toto 1/9/15 4:07 AM
Formatted ... [327]
titi toto 1/9/15 4:07 AM
Formatted ... [328]
titi toto 1/9/15 4:07 AM
Deleted: stride
titi toto 1/9/15 4:07 AM
Formatted ... [329]
titi toto 1/9/15 4:07 AM
Deleted: 1
titi toto 1/9/15 4:07 AM
Formatted ... [330]
titi toto 1/9/15 4:07 AM
Formatted ... [331]
titi toto 1/9/15 4:07 AM
Formatted ... [332]
titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, 12 pt, Font color: Text 1, German

Table

A1. FORTRAN code description of the GEN_BE v2.0 framework.

FORTRAN modules	Comments
variables_types.f90	It defines, declares and allocates new types as state_type, mesh_type, bin_type, state_matrix. Some basics operations as addition substraction, calculation of variance, covariance are available.
configure.f90	It reads the namelist.input file and initialize the variables
io_input_models.f90	It reads input standard variables from a model define by the user and convert them into control variables. If the user needs to introduce new input model, only this module needs to be updated to read and transform the data.
io_input.f90	It reads NetCDF input data and initialize new types
io_output.f90	It writes NetCDF output format for all new types
io_output_applications.f90	It writes output for different application needs

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, 12 pt, Font color: Text 1, German

titi toto 1/9/15 4:07 AM
Deleted: 6. Information related to the control variables and their covariance errors in the namelist input file (&gen_be_cv part, example CV5).
&gen_be_cv ... [333]

titi toto 1/9/15 4:07 AM
Formatted ... [334]

titi toto 1/9/15 4:07 AM
Formatted ... [335]

titi toto 1/9/15 4:07 AM
Formatted ... [336]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted ... [337]

titi toto 1/9/15 4:07 AM
Formatted ... [338]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted ... [339]

titi toto 1/9/15 4:07 AM
Formatted ... [340]

titi toto 1/9/15 4:07 AM
Formatted ... [341]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Times New Roman, 12 pt, Font color: Text 1

Table A2. Input and output of the different components of the GEN_BE v2.0 code.

Programs	Input	output	comments
gen_be_stage0.F	Various models (ex: WRF)	pert.ccyymmddhh	It contains the perturbations for all the control variables defined in the namelist
		mesh_grid.nc All_mesh.grid.nc	It contains all the static data as latitude array, longitude array, map factors
		mask.ccyymmddhh	This file exists only with the option dynamical_mask which is activated with bin_type=7 or bin_type=8
		standard_variable.txt control_variable.txt	It contains the list of the control variables in ASCII format
gen_be_stage1.F	pert.ccyymmddhh	var. ccyymmddhh	The input file is split per variables
		bins.nc	All the information related to the binning options are included in this file
gen_be_stage2.F	var. ccyymmddhh	gen_be_stage2_regcoeff.nc	All the regression coefficients are included in this file
		var(_u) ccyymmddhh	If a linear regression is applied to the current variable to remove its balanced part, an unbalanced output variable is written under this nomenclature
gen_be_stage3.F	var(_u). ccyymmddhh	gen_be_stage3_vert_lenscale.var(_u).nc	It contains the vertical length scale parameter for the full or unbalanced part of the variable
		gen_be_stage3_varce.var(_u).nc	Variance 3 dimensions by grid point
		gen_be_stage3_vert_varce(_u).nc	Binned vertical variance
		var(_u).ccyymmddhh.ennn.kkk	Intermediate binary files split by vertical level
gen_be_stage4.F	var(_u).ccyymmddhh.ennn.kkk	sl_print.blll.qcloud	Intermediate ASCII file format that contain the horizontal length scale
gen_be_diags.F	Results of the precedents stages from 2 to 4	be.nc	Final netcdf file that contains all the information to model B
gen_be_nc2gsi.F	be.nc	be_gsi_little_endian.gcv be_gsi_big_endian.gcv	Binary format directly readable by GSI

Formatted ... [342]
titi toto 1/9/15 4:07 AM
Formatted ... [343]
titi toto 1/9/15 4:07 AM
Deleted: B2
titi toto 1/9/15 4:07 AM
Formatted ... [344]
titi toto 1/9/15 4:07 AM
Formatted ... [345]
titi toto 1/9/15 4:07 AM
Formatted ... [346]
titi toto 1/9/15 4:07 AM
Formatted ... [347]
titi toto 1/9/15 4:07 AM
Formatted ... [348]
titi toto 1/9/15 4:07 AM
Formatted ... [349]
titi toto 1/9/15 4:07 AM
Formatted ... [350]
titi toto 1/9/15 4:07 AM
Formatted ... [351]
titi toto 1/9/15 4:07 AM
Deleted: exist
titi toto 1/9/15 4:07 AM
Formatted ... [352]
titi toto 1/9/15 4:07 AM
Formatted ... [353]
titi toto 1/9/15 4:07 AM
Formatted ... [354]
titi toto 1/9/15 4:07 AM
Formatted ... [355]
titi toto 1/9/15 4:07 AM
Formatted ... [356]
titi toto 1/9/15 4:07 AM
Deleted: ted
titi toto 1/9/15 4:07 AM
Formatted ... [357]
titi toto 1/9/15 4:07 AM
Formatted ... [358]
titi toto 1/9/15 4:07 AM
Formatted ... [359]
titi toto 1/9/15 4:07 AM
Formatted ... [360]
titi toto 1/9/15 4:07 AM
Formatted ... [361]
titi toto 1/9/15 4:07 AM
Formatted ... [362]
titi toto 1/9/15 4:07 AM
Formatted ... [363]
titi toto 1/9/15 4:07 AM
Formatted ... [364]
titi toto 1/9/15 4:07 AM
Formatted ... [365]
titi toto 1/9/15 4:07 AM
Formatted ... [366]
titi toto 1/9/15 4:07 AM
Formatted ... [367]
titi toto 1/9/15 4:07 AM
Formatted ... [368]
titi toto 1/9/15 4:07 AM
Deleted: ted
titi toto 1/9/15 4:07 AM
Formatted ... [369]
titi toto 1/9/15 4:07 AM
Formatted ... [370]
titi toto 1/9/15 4:07 AM

titi toto 1/9/15 4:07 AM
Formatted ... [371]
titi toto 1/9/15 4:07 AM
Formatted ... [372]
titi toto 1/9/15 4:07 AM

Table A3. Content of the final output file be.nc (NetCDF format) of the GEN_BE v2.0 code.

Name of the field	Description
Fields defined by control variable name (e.g. cv1)	
Lenscale_cv1	Horizontal length scale in EOFs space or physical space
vert_lenscale_cv1	Vertical length scale available only if the flag data_on_levels is true and the control variable number 1 is 3D.
vert_variance_cv1	Vertical variance of the control variable number 1 per bin
eigen_value_cv1	Eigenvalue of the control variable number 1 only available if the flag data_on_levels is false
eigen_vector_cv1	Eigenvector of the control variable number 1 only available if the flag data_on_levels is false
varce_cv1	Variance 3D
Regression coefficients	
list_regcoeff	Complete list of the regression coefficients used in the balance constraint.
regcoeff_cv1_cv2	Example of regression coefficient between the control variable 1 and 2. It can be 1D, 2D or 3D
vert_autocov_cv1	Vertical autocovariance of the control variable number 1
Binning parameters	
bin_type	Bin_type option selected
bin2d	Binning field 2D array
bins	Binning field 3D array

Formatted ... [373]

titi toto 1/9/15 4:07 AM
Deleted: B3

titi toto 1/9/15 4:07 AM
Formatted ... [374]

titi toto 1/9/15 4:07 AM
Formatted ... [375]

titi toto 1/9/15 4:07 AM
Formatted ... [376]

titi toto 1/9/15 4:07 AM
Formatted ... [377]

titi toto 1/9/15 4:07 AM
Formatted ... [378]

titi toto 1/9/15 4:07 AM
Formatted ... [379]

titi toto 1/9/15 4:07 AM
Deleted: len

titi toto 1/9/15 4:07 AM
Formatted ... [380]

titi toto 1/9/15 4:07 AM
Formatted ... [381]

titi toto 1/9/15 4:07 AM
Formatted ... [382]

titi toto 1/9/15 4:07 AM
Deleted: len

titi toto 1/9/15 4:07 AM
Formatted ... [383]

titi toto 1/9/15 4:07 AM
Formatted ... [384]

titi toto 1/9/15 4:07 AM
Formatted ... [385]

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted ... [386]

titi toto 1/9/15 4:07 AM
Formatted ... [387]

titi toto 1/9/15 4:07 AM
Deleted:

titi toto 1/9/15 4:07 AM
Formatted ... [388]

titi toto 1/9/15 4:07 AM
Formatted ... [389]

titi toto 1/9/15 4:07 AM
Formatted ... [390]

titi toto 1/9/15 4:07 AM
Formatted ... [391]

titi toto 1/9/15 4:07 AM
Formatted ... [392]

titi toto 1/9/15 4:07 AM
Formatted ... [393]

titi toto 1/9/15 4:07 AM
Formatted ... [394]

titi toto 1/9/15 4:07 AM
Formatted ... [395]

titi toto 1/9/15 4:07 AM
Formatted ... [396]

titi toto 1/9/15 4:07 AM
Formatted ... [397]

titi toto 1/9/15 4:07 AM
Formatted ... [398]

titi toto 1/9/15 4:07 AM
Formatted ... [399]

titi toto 1/9/15 4:07 AM
Formatted ... [400]

titi toto 1/9/15 4:07 AM
Formatted ... [401]

Table B1. General information defining the experiment in the namelist iput file (&gen_be_info part).

&gen_be_info	Namelist options	Description
model	‘WRF’	Set up the acronym for the model input allows GEN_BE to read different input model in the stage 0.
application	‘WRFDA’	‘WRFDA’ and ‘GSI’ interface have been developed and tested.
be_method	‘ENS’ or ‘NMC’	Compute perturbations from an ensemble or from different time lagged forecast.
ne	Number of members	If NMC method ne=1.
cut	0, 0, 0, 0, 0, 0,	Allow to subset an area of a domain, defined in grid points. imin, imax, jmin, jmax, kmin, kmax.
use_mean_ens	‘false’	If be_method=‘ENS’ is selected, the perturbation can be calculated from the mean of all the members or from 2 different members.
start_date	,_START_DATE_’	Initial date, format ccyymmddhh.
end_date	,_END_DATE_’	Final date, format ccyymmddhh.
interval	‘hh’	Frequency of the historical date data available, defined in hour (useful for the NMC method only).

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, German

titi toto 1/9/15 4:07 AM

Deleted:
Fig. 1. WRF domain over the conus area at the resolution of 15 km. Based on this configuration, the 50 members coming from a 6h forecast (DART, experiment DC3) are used to generate background error stsatistics.

Fig.

Unknown
Formatted: Font:Not Bold

Unknown
Formatted: Font:Not Bold

Table B2. Description of the binning options.

Bin_type	Description
0	Binning by grid point.
1	Binning by vertical level along the x direction point of the model.
2	Binning by vertical heights and by latitude num_bins_lat. The parameters binwidth_lat and binwidth_hgt define the width that splits the bins.
3	Binning by vertical model level and latitude dependent. The parameters lat_min, lat_max are computed from the model input data and the parameter binwidth_lat is defined in the namelist.input file.
4	Binning by vertical model level and along the y direction.
5	Binning on vertical model level including all the horizontal point.
6	Average over all points.
7	Binning rain/no-rain by vertical model level and based on thresholds in the model background (Michel and al., 2011.).

Table B3. Parameters defining the binning options of the namelist input file (&gen_be_bin part).

&gen_be_bin	Namelist options	Description
bin_type	0-7	Bin type option
lat_min, lat_max		Minimum and Maximum of latitude defined in degree. Used if bin_type = 2
binwidth_lat	5.0	Width of the bins defines by latitude in degree Used if bin_type = 2, 3, 4
hgt_min	1000.0	Used if bin_type = 2 (height, meter)
binwidth_hgt	2000.0	Width of bins defines by height in meter Used if bin_type = 2 (meter)

Table B4. Information related to the control variables and their covariance errors in the namelist input file (&gen_be_cv part, example CV5). At present, the parameter covar can take three values: 0, 1, and 2, meaning "no regression", "full regression" and "diagonal only".

&gen_be_cv	Namelist options	Description
nb_cv	5,	Number of control variables
cv_list	'psi','chi','*t*','ps','rh',	Variables used for the analysis
fft_method	1,2	Conversion of u and v to psi and chi 1=Cosine, 2=Sine transform
covar1	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,	First variable does not have covariance
covar2	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Covariance of variable 1 (psi) and variable 2 (chi)
covar3	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Covariance of variable 1 (psi) with variable 3 (t)
covar4	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Covariance of variable 1 (psi) with variable 3 (ps)
covar5	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Relative humidity univariate
covar6	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,	Other possible variable
use_chol_reg	.false.	by default, compute the regression coefficient as a ratio of covariance by variance. If true, use a cholesky decomposition (specific to GSI, CV5).

Table B5. Information related to the control variables and their covariance errors in the namelist input file (&gen_be_cv part, example CV9, definition of multivariate humidity and hydrometeors error covariance matrix).

&gen_be_cv	Namelist Options
nb_cv	9,
cv_list	'psi','chi','t','ps','rh',' q_{cloud}',' q_{ice}',' q_{rain}',' q_{snow}',
covar1	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
covar2	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
covar3	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
covar4	1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
covar5	0, 0, 1, 1, 0, 0, 0, 0, 0, 0,
covar6	0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
covar7	0, 0, 0, 0, 2, 0, 0, 0, 0, 0,
covar8	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
covar9	0, 0, 0, 0, 0, 0, 0, 0, 0, 0,

Table B6. Description of the options available in the namelist input file (&gen_be_lenscale part) to diagnose length scale parameter.

&gen_be_lenscale	Namelist options	Description
data_on_levels	'true'	The statistics can be computed by vertical model level (GSI) or by EOF mode (WRFDA) in stage 3
vert_ls_method	1, 2	Estimate the vertical length scale (stage 3) Option 1: parabolic approximation formula Option 2: gaussian approximation formula
ls_method	1, 2	Estimate horizontal length scale (stage 4) See Sect. 3.4 for more details.
use_med_ls	'false'	Estimate the length using the median value or not.
stride	1	Subset of point to speed up the stage 4
n_smth_ls	2	Number of point to smooth the length scale
use_global_bin	'false'	The statistics can be binned (use_global_bin=false) or not in stages 3 and 4. Only inhomogeneous recursive filters can handle binned length scale.

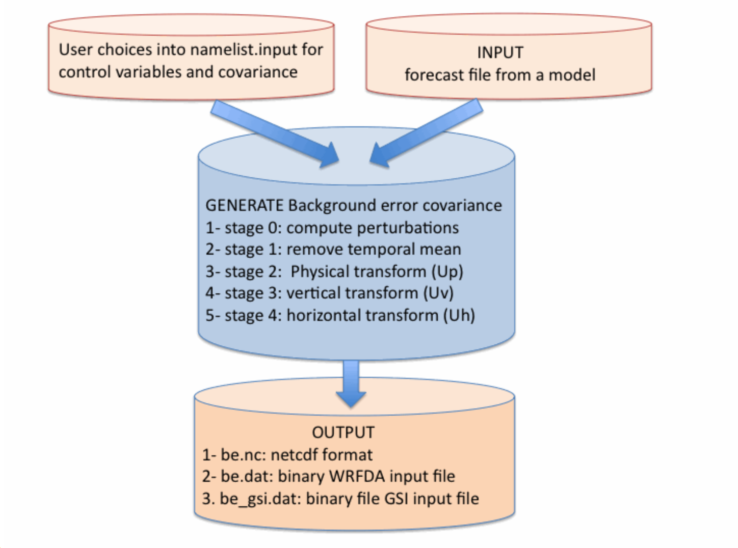


Figure 1. General structure of the code to generate a background error covariance matrix. The input and output are represented by the orange boxes and the five main stages that lead to model **B** are in blue.

Unknown
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

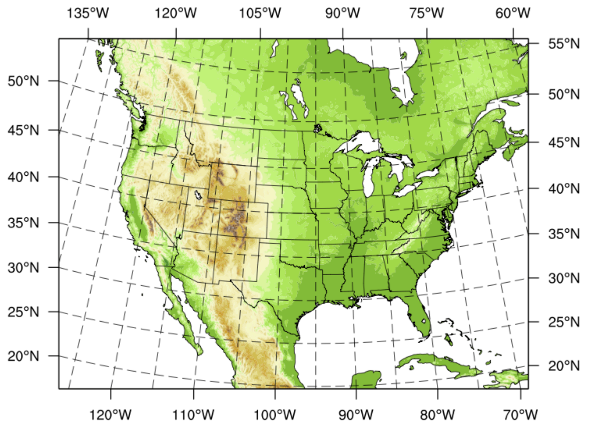
titi toto 1/9/15 4:07 AM
Deleted: is

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp><sp><sp><sp><sp><sp> ... [402]

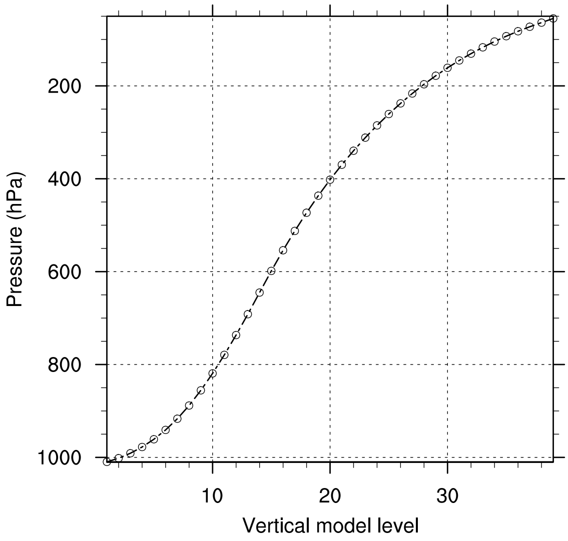


Unknown
Formatted: Font:Not Bold, Font color: Text 1

Figure 2. WRF domain over the conus area at the resolution of 15 km. Based on this configuration, the 50 members coming from a 6h forecast (DART experiment) are used to generate background error statistics.

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: . Horizontal autocorrelation performed at the center of each square grid over vertical model level 5, around 950 hPa, for the control variables (a) psi, (b) t, (c) rh, and (d) q_{cloud}. Larger correlations are oberved for psi compared to t and rh. q_{cloud} has the smallest correlation sparecly distrubuted.



Unknown
Formatted: Font:Not Bold, Font color: Text 1

Figure 3. Plot of Pressure (hPa) against vertical model levels (WRF, Res. 15 km).

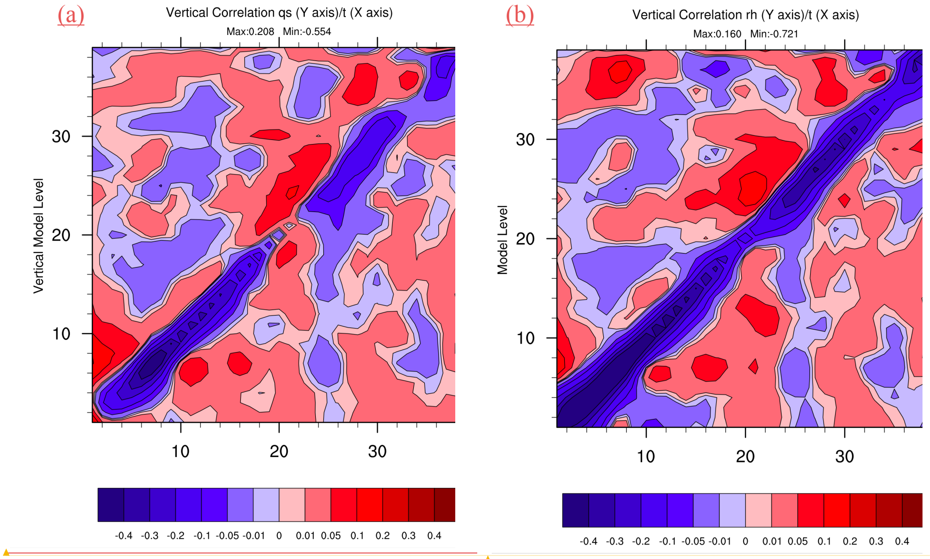


Figure 4. **(a)** Vertical cross-correlation between temperature (t) and specific humidity (qs), **(b)** vertical cross-correlation between temperature (t) and relative humidity (rh); (WRF, Res. 15 km, D-ensemble).

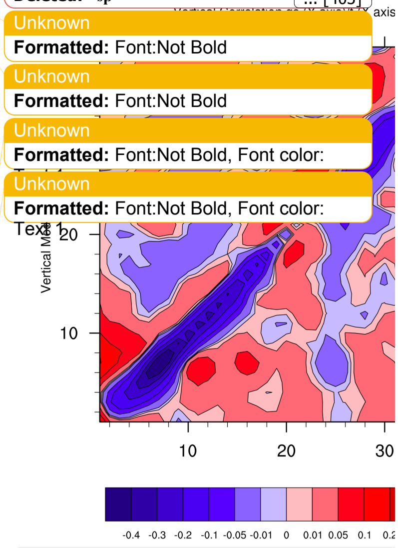
titi toto 1/9/15 4:07 AM
Deleted: <sp> ... [403]

Unknown
Formatted: Font:Not Bold

Unknown
Formatted: Font:Not Bold

Unknown
Formatted: Font:Not Bold, Font color:

Unknown
Formatted: Font:Not Bold, Font color: Text 1



titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: . ... [404]

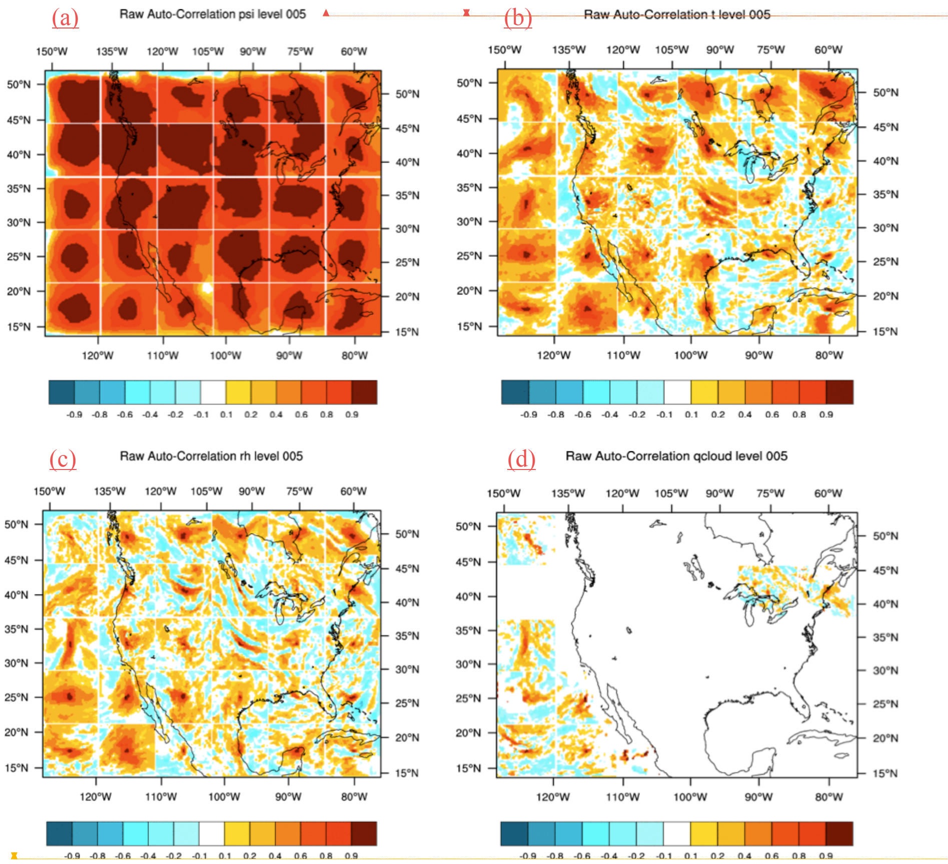


Figure 5. Horizontal autocorrelation performed at the center of each square grid over vertical model level 5, around 950 hPa, for the control variables **(a)** stream function (psi), **(b)** temperature (t), **(c)** relative humidity (rh), and **(d)** cloud mixing ratio (q_{cloud}). Larger correlations are observed for stream function compared to temperature and relative humidity. Cloud mixing ratio has the smallest correlation due to sparce location of hydrometeors (WRF, Res. 15 km, D-ensemble).

titi toto 1/9/15 4:07 AM
Deleted: <sp><sp> ... [405]

Unknown
Formatted: Font:Not Bold

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp> ... [406]

Unknown
Formatted: Font:Not Bold

titi toto 1/9/15 4:07 AM
Formatted Table

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: ... [407]

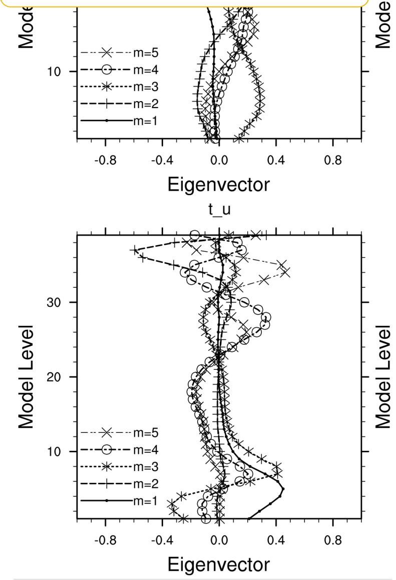
Unknown
Formatted: Font:Not Bold

Unknown
Formatted: Font:Not Bold

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: . (a) Raw vertical cross-correlations between q_{cloud} and rh, (b) filtered vertical cross-correlations between q_{cloud} and rh, (c) raw v ... [408]

Unknown
Formatted: Font:Not Bold



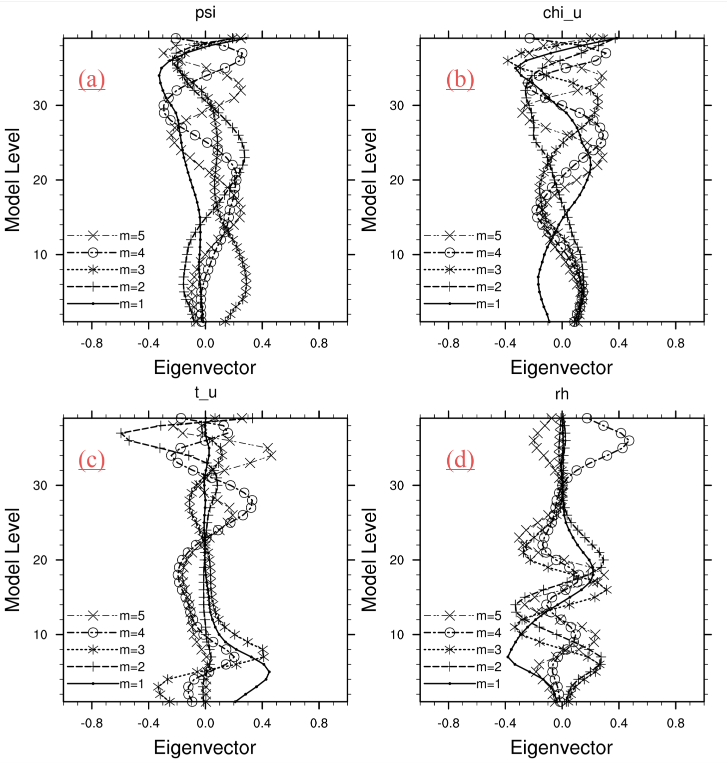


Figure 6. Representation of the first five eigenvectors resulting from the EOF decomposition of the vertical autocovariance matrix, eigenvectors of **(a)** psi, **(b)** chi$_u$, **(c)** t_u, and **(d)** rh. The eigenvectors are parameters that define the vertical transform (**U**$_v$); (WRF, Res. 15 km, D-ensemble, EOFs).

Unknown
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: of

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Italic, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

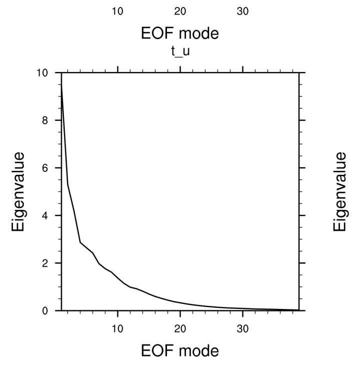
titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp><sp> ... [409]

Unknown
Formatted: Font:Not Bold



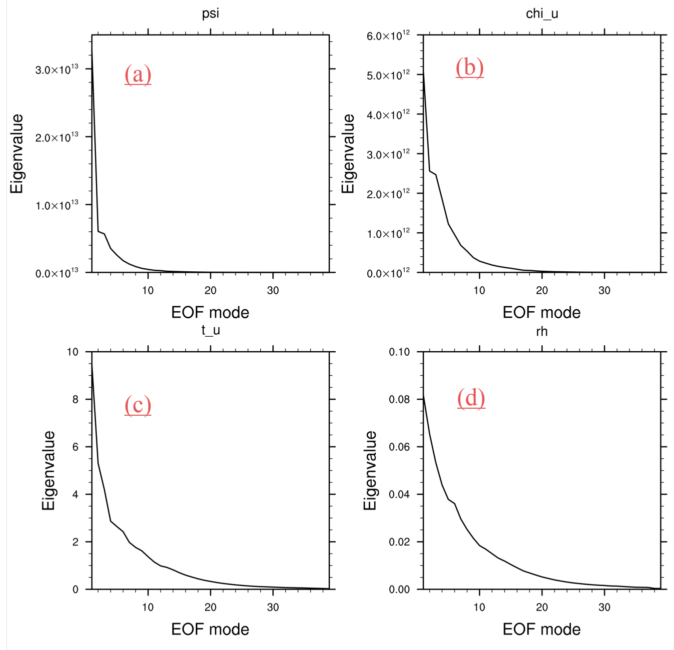


Figure 7. Eigenvalues computed by EOF mode for **(a)** psi, **(b)** chi$_u$, **(c)** t_u and **(d)** rh. They represent the variance of the control variables (WRF, Res. 15 km, D-ensemble, EOFs).

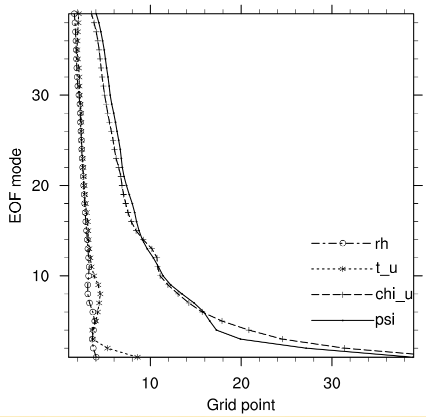


Figure 8. Length scales defined in grid point through EOF mode for CV5. The analysis control variables representating the dynamical variables, psi and chi$_u$, have longer length scales than t_u, and rh (WRF, Res. 15 km, D-ensemble, EOFs).

Formatted ... [410]

titi toto 1/9/15 4:07 AM
Formatted ... [411]

titi toto 1/9/15 4:07 AM
Formatted ... [412]

titi toto 1/9/15 4:07 AM
Formatted ... [413]

titi toto 1/9/15 4:07 AM
Formatted ... [414]

titi toto 1/9/15 4:07 AM
Formatted ... [415]

titi toto 1/9/15 4:07 AM
Formatted ... [416]

titi toto 1/9/15 4:07 AM
Formatted ... [417]

titi toto 1/9/15 4:07 AM
Formatted ... [418]

titi toto 1/9/15 4:07 AM
Formatted ... [419]

titi toto 1/9/15 4:07 AM
Formatted ... [420]

titi toto 1/9/15 4:07 AM
Formatted ... [421]

titi toto 1/9/15 4:07 AM
Deleted: former

titi toto 1/9/15 4:07 AM
Formatted ... [422]

titi toto 1/9/15 4:07 AM
Formatted ... [423]

titi toto 1/9/15 4:07 AM
Formatted ... [424]

Unknown
Formatted ... [425]

titi toto 1/9/15 4:07 AM
Formatted ... [426]

titi toto 1/9/15 4:07 AM
Formatted ... [427]

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted ... [428]

titi toto 1/9/15 4:07 AM
Deleted: larger

titi toto 1/9/15 4:07 AM
Formatted ... [429]

titi toto 1/9/15 4:07 AM
Formatted ... [430]

titi toto 1/9/15 4:07 AM
Formatted ... [431]

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted ... [432]

titi toto 1/9/15 4:07 AM
Formatted ... [433]

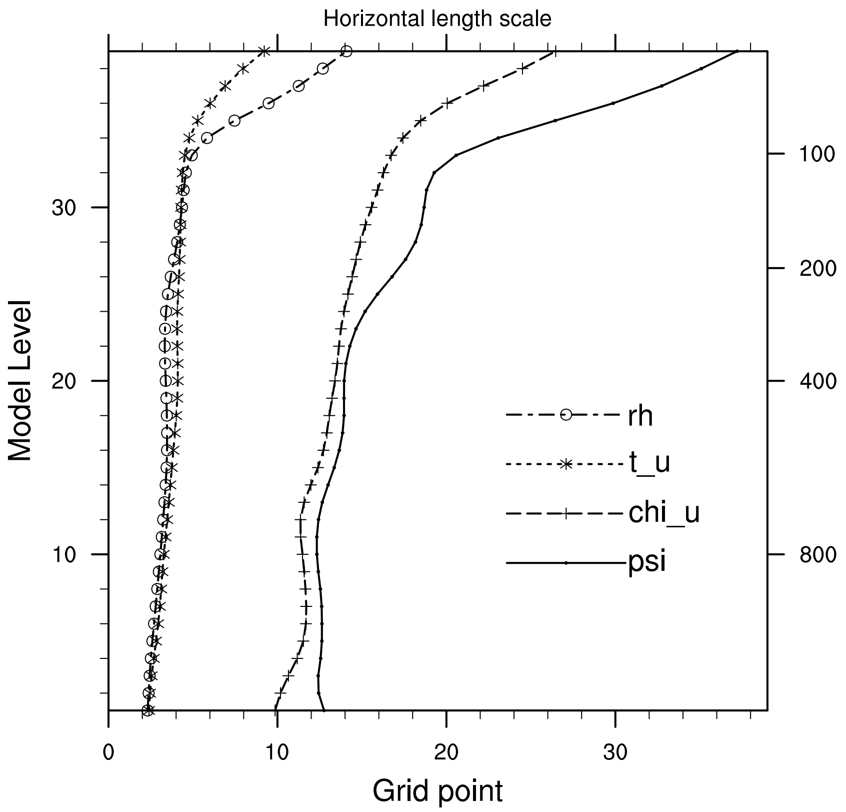


Figure 9. Horizontal length scales for CV5. t_u and rh, which have more local structures, are modeled by shorter length scales (WRF, Res. 15 km, D-ensemble, RFs).

Unknown
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Italic, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: smaller

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

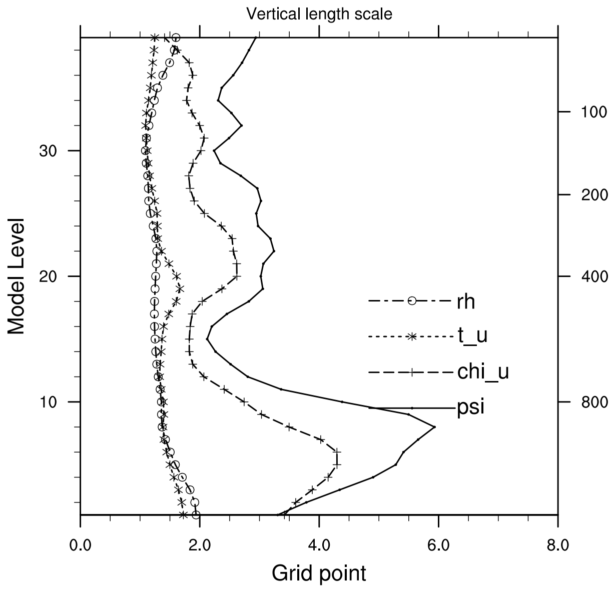


Figure 10. Vertical length scale for CV5 (WRF, Res. 15 km, D-ensemble, RFs).

Unknown
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

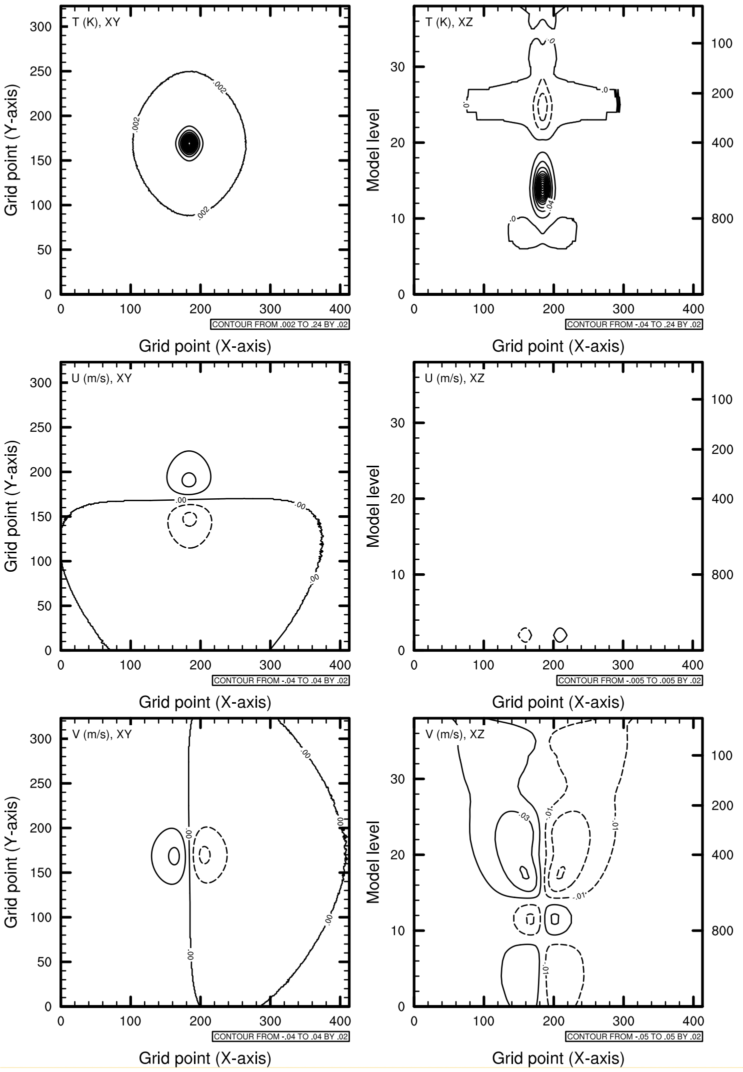


Figure 11. Pseudo observation test of temperature (innovation of +1 K) from the WRFDA application. The three plots on the left panel show, from top to bottom, horizontal cross-section (XY) of t (K), U and V wind component (m s^{-1}) respectively. Then, the right panel shows the corresponding cross-section (XZ) of the former variables ($\mathbf{B}_{eof}$: WRF Res. 15 km, D-ensemble, EOFs).

Unknown
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Deleted: +1K) under

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: using a background error covariance matrix diagnostic from GEN_BE v2.0 with EOF decomposition (U_v).

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

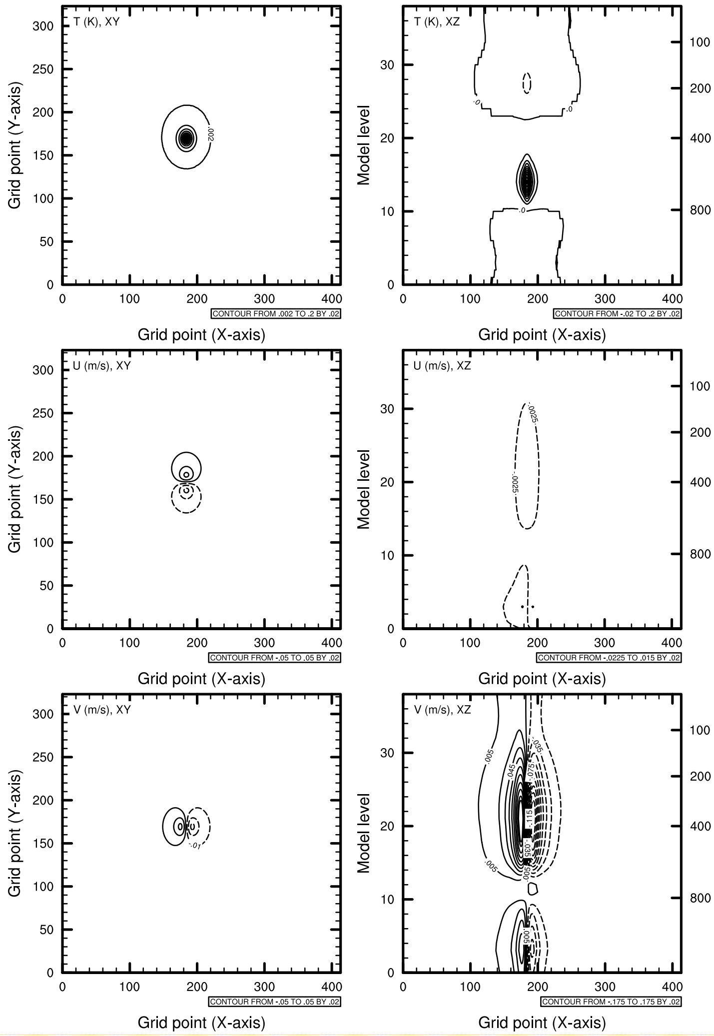


Figure 12. Pseudo observation test of temperature (innovation of +1 K) from the GSI application. The three plots on the left panel show, from top to bottom, horizontal cross-section *(XY)* of t (K), U and V wind component (ms^{-1}) respectively. Then, the right panel shows the corresponding cross-section (XZ) of the former variables. (**B**$_{ref}$: WRF Res. 15 km, D-ensemble, RFs).

Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Deleted: +1K) using recursive filters (U$_v$) under

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted ... [434]

titi toto 1/9/15 4:07 AM
Deleted: a background error covariance matrix diagnostic from GEN_BE v2.0

titi toto 1/9/15 4:07 AM
Formatted ... [435]

titi toto 1/9/15 4:07 AM
Deleted:).

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: ... [436]

Unknown
Formatted: Font:Not Bold

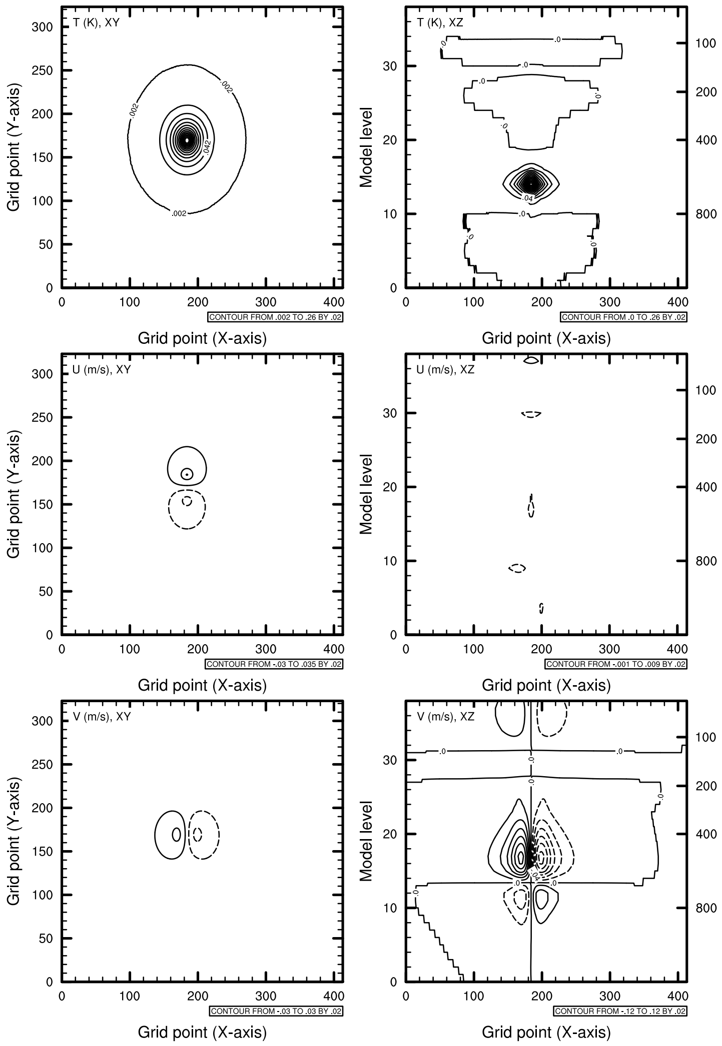


Figure 13. Pseudo observation test of temperature (innovation of +1 K) from the GSI application. The three plots on the left panel show, from top to bottom, horizontal cross-section (XY) of t (K), U and V wind component (ms^{-1}) respectively. Then, the right panel

Unknown
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: +1K) using recursive filters under

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

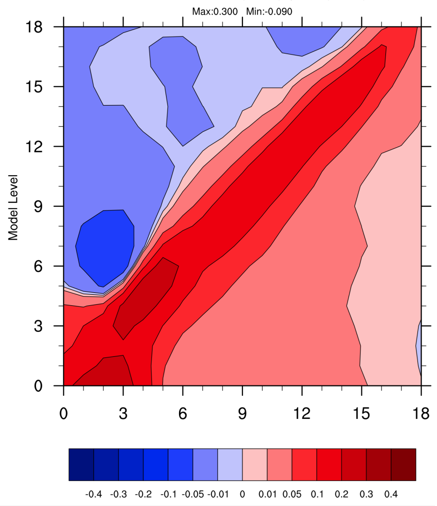
titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

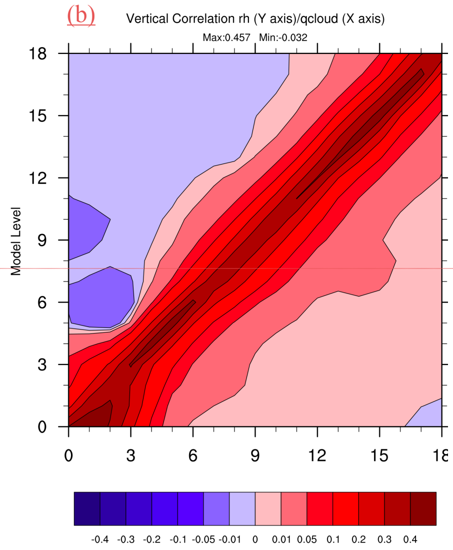
titi toto 1/9/15 4:07 AM
Deleted: the NAM background error covariance

shows the corresponding cross-section (XZ) of the former variables. (**B**$_{nam}$:WRF-NMM Res. 12 km, NMC, RFs).

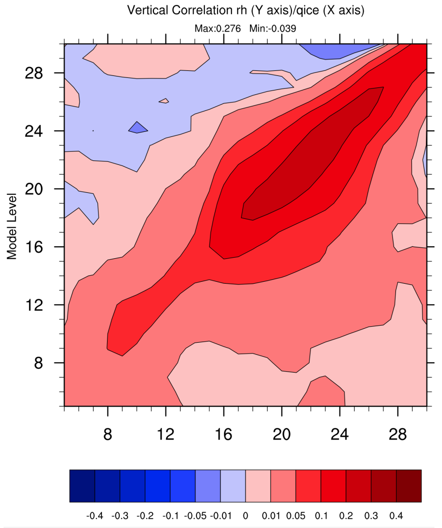
(a)



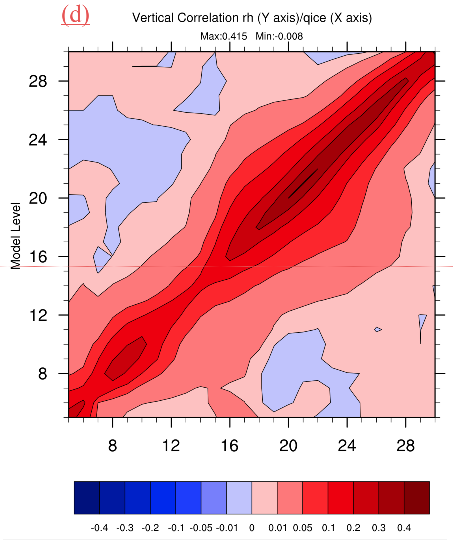
(b)



(c)



(d)



Formatted ... [437]

titi toto 1/9/15 4:07 AM

Deleted: ... [438]

Unknown

Formatted: Font:Not Bold

titi toto 1/9/15 4:07 AM

Deleted: <sp> ... [439]

titi toto 1/9/15 4:07 AM

Deleted: <sp> ... [440]

titi toto 1/9/15 4:07 AM

Formatted Table

Unknown

Formatted ... [441]

titi toto 1/9/15 4:07 AM

Formatted: Font color: Text 1

Unknown

Formatted ... [442]

titi toto 1/9/15 4:07 AM

Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM

Deleted:

titi toto 1/9/15 4:07 AM

Formatted Table

Unknown

Formatted ... [443]

titi toto 1/9/15 4:07 AM

Formatted: Font color: Text 1

Unknown

Formatted: Font:Not Bold

Unknown

Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM

Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM

... [444]

Figure 14. **(a)** Raw vertical cross-correlations between cloud mixing ratio (q_{cloud}) and relative humidity (rh), **(b)** filtered vertical cross-correlations between q_{cloud} and rh, **(c)** raw vertical cross-correlations between ice mixing ratio (q_{ice}) and rh, **(d)** filtered vertical cross-correlations between q_{ice} and rh. Taking into account the perturbations coming from the transition of a cloudy to a non-cloudy area only when reaching the threshold mixing ratio of 10^{-6} kg kg^{-1}, intensifies the vertical correlation (WRF, Res. 15 km, D-ensemble).

titi toto 1/9/15 4:07 AM

Deleted: .

titi toto 1/9/15 4:07 AM

Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM

Deleted: . Vertical length scale for the hydrometeors using (a) 50 members and (b) using 5 members (b) show a similar evolution along the model levels.

Page Break

<sp><sp>

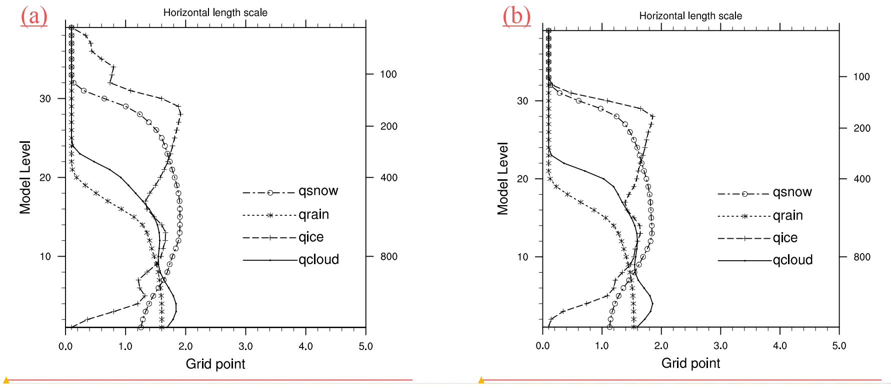
Fig.

titi toto 1/9/15 4:07 AM

Formatted: Font:Not Bold, Font color:

Unknown

Formatted: Font:Not Bold

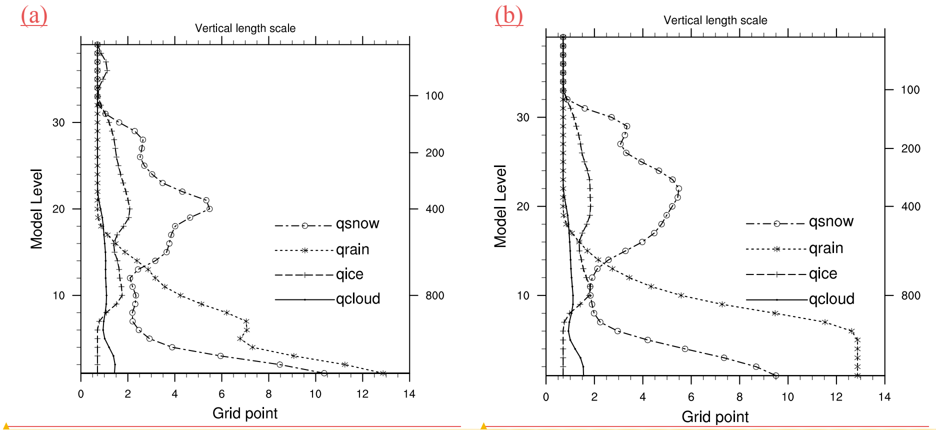


Unknown
Formatted: Font:Not Bold, Font color: Text 1

Unknown
Formatted: Font:Not Bold, Font color: Text 1

Figure 15. “Horizontal length scale for the hydrometeors using **(a)** 50 members and **(b)** using 5 members. The plots show similar characteristics regardless to the ensemble members (WRF, Res. 15 km, D-ensemble).

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1



Unknown
Formatted: Font:Not Bold, Font color: Text 1

Unknown
Formatted: Font:Not Bold, Font color: Text 1

Figure 16. Vertical length scale for the hydrometeors using (a) 50 members and (b) using 5 members. The plots show similar characteristics regardless to the ensemble members (WRF, Res. 15 km, D-ensemble).

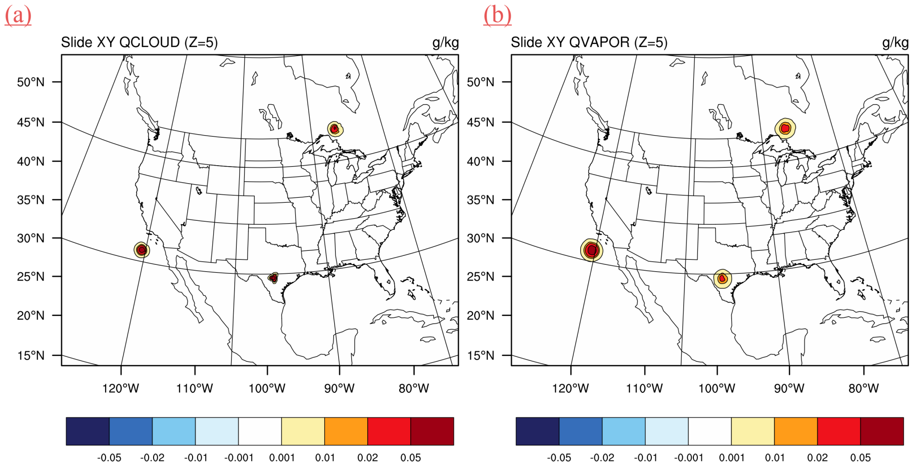


Figure 17. **(a)** Horizontal slide (vertical model level 5) of a pseudo observation test of cloud water condensate (Innovation and observation error of 0.1 g kg^{-1}) in a multivariate approach using the 3-D variance, **(b)** as a consequence there is a positive increment on qvapor (WRF, Res. 15 km, D-ensemble, RFs).

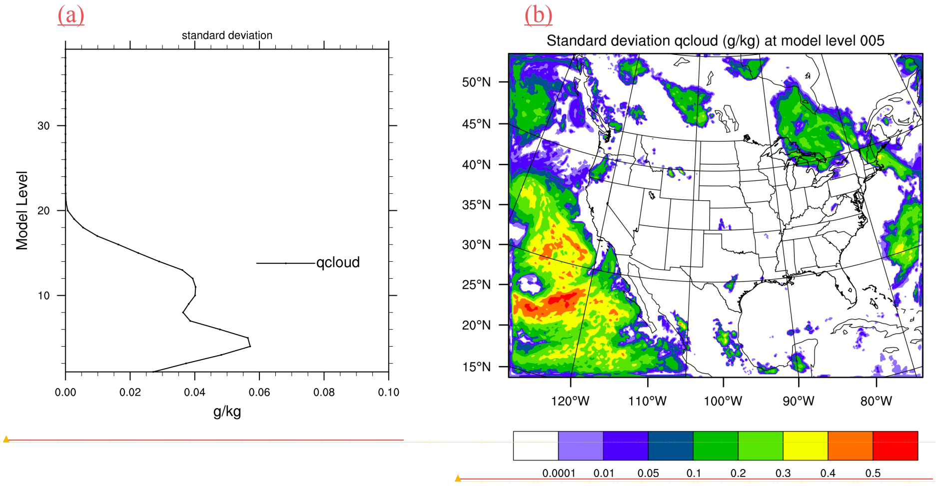


Figure 18. **(a)** Profile of standard deviation of liquid water condensate mixing ratio (q_{cloud} in g kg^{-1}) averaged along the vertical and **(b)** horizontal cross-section of standard deviation of

Unknown
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: 1g/

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp><sp> ... [445]

Unknown
Formatted: Font:10 pt, Not Bold

Unknown
Formatted: Font:10 pt, Not Bold

Unknown
Formatted ... [446]

Unknown
Formatted ... [447]

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: Standard

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: : (a

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

q_{cloud} at the vertical model level 5 (950 hPa). Both plots indicate the presence of low maritime clouds noted by high standard deviation (WRF, Res. 15 km, D-ensemble).

titi toto 1/9/15 4:07 AM
Deleted: and (b) profile of the average along the vertical

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1

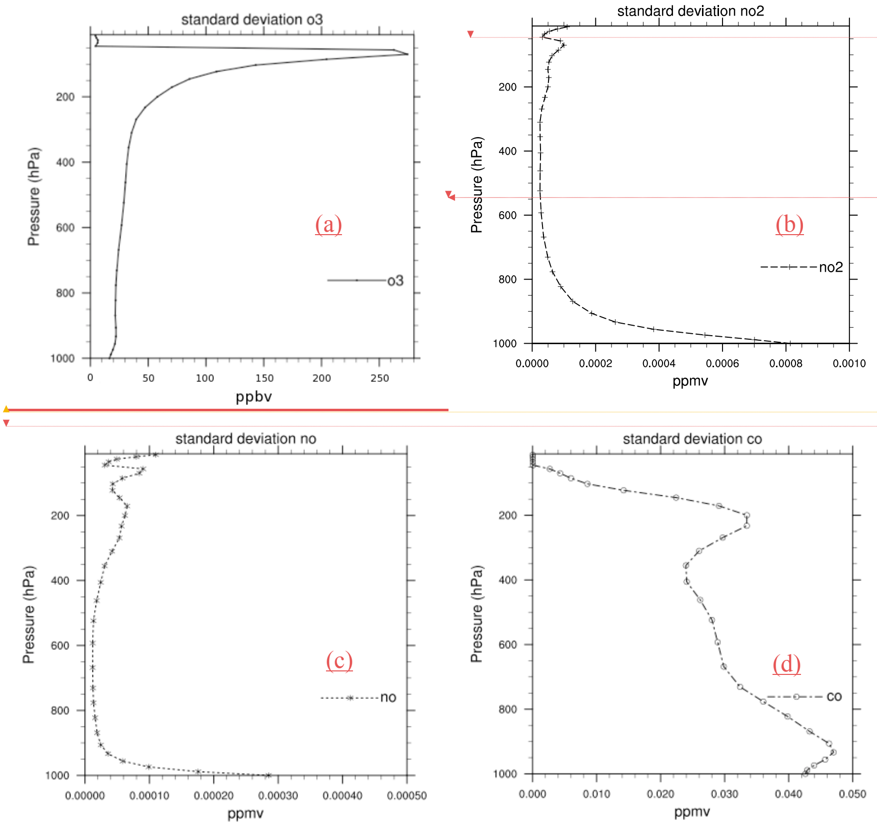


Figure 19. Vertical standard deviation in ppmv of (a) O3, (b) NO2, (c) NO, and (d) CO (WRF-CHEM, Res. 36 km, D-ensemble).

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp>

titi toto 1/9/15 4:07 AM
Formatted Table

Unknown
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Deleted: <sp><sp>

titi toto 1/9/15 4:07 AM
Deleted: <sp> ... [448]

titi toto 1/9/15 4:07 AM
Deleted: . A1.

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, German

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

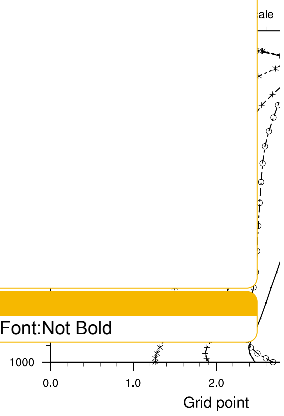
titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, German

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM



Deleted: Fig. A2.

Unknown
Formatted: Font:Not Bold

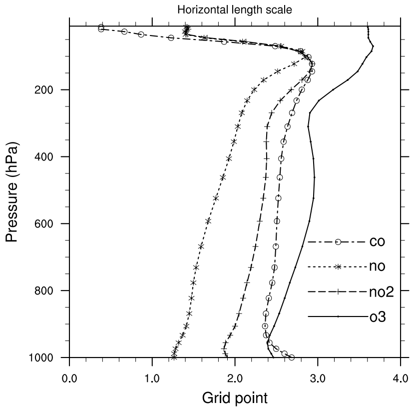


Figure 20. Horizontal length scale of O3, NO2, NO, and CO (WRF-CHEM, Res. 36 km, D-ensemble).

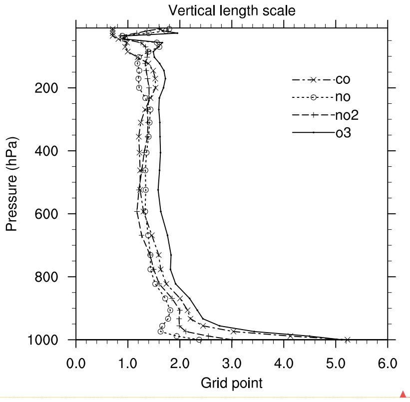


Figure 21. Vertical length scale of O3, NO2, NO, and CO (WRF-CHEM, Res. 36 km, D-ensemble).

Unknown
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, English (US)

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, Not Superscript/ Subscript

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, English (US)

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, English (US)

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, Not Superscript/ Subscript

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English

Unknown
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, English (US)

titi toto 1/9/15 4:07 AM
Deleted: . A3.

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color: Text 1, Not Superscript/ Subscript

titi toto 1/9/15 4:07 AM
Formatted ... [449]

titi toto 1/9/15 4:07 AM
Formatted: Font:Not Bold, Font color:

titi toto 1/9/15 4:07 AM
Formatted ... [450]

titi toto 1/9/15 4:07 AM
Formatted ... [451]

titi toto 1/9/15 4:07 AM
Deleted: .

titi toto 1/9/15 4:07 AM
Formatted ... [452]

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1, English

titi toto 1/9/15 4:07 AM
Formatted: Font color: Text 1