# 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 $\delta x$
- 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 $\delta t_u$
- 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 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 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):"*

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*

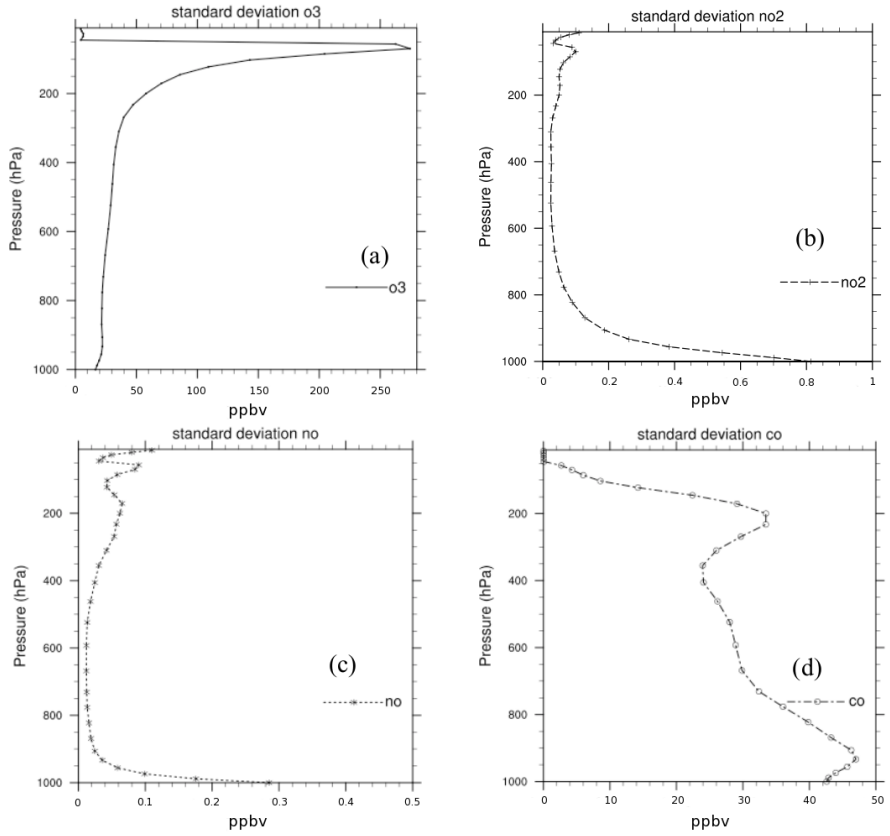
Figure 1: the last word is statistics.
*Corrected*

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.*