Reply to referee comments (R. Bannister)

We would like to thank R. Bannister for his thorough review and for useful suggestions that should help to improve the manuscript. Here below are the responses to his comments. The comments of the reviewer are in italic. Note that sometimes, a reviewer's comment is split to allow the inclusion of a part of our response.

General comments

This paper is essentially a review of the spectral method as used to model efficiently horizontal correlations of background errors data assimilation. It applies the method to a tracer problem (with real data) of stratospheric ozone data assimilation. The spectral method to model homogeneous and isotropic background error correlations is well established in data assimilation, but this paper is a reasonably comprehensive collection of useful references for the practitioner of data assimilation. It is accessible to mathematicians and non- mathematicians alike. While most of the contents of this paper are not new, I think that it will be a very useful and well-cited source.

Specific comments

This is a list of issues that the authors might like to think about or do. I do not suggest that all need to be done to make the paper publishable.

1. *In the assimilation experiments with the single observation, is the observation always exactly on a grid point in both the LL and GG cases?*

   The observation is located at the same model grid point in the LL and GG experiments. However, it seems that the grid definition is not very clear (see also your technical comment 4 and our response). To clarify, note that the LL and GG experiments have the same model grid which is equally spaced. In the LL experiments, the spectral transform operates form the spectral space to the model grid. In the GG experiments, the spectral transform operates from the spectral space to the Gaussian grid and a mapping operation from the Gaussian grid to the model grid is included. This will be clarified.

   It would also be interesting to see how the results change as the observation location is increased in latitude to see how the interpolation errors introduced by the G-operator change.

   Here below are shown the scores of several "1-obs" assimilation experiments where the observation is placed at several latitudes, namely: the Equator (as in the original manuscript), 40°N and 80°N. We see that the results confirm those found at the Equator which have been shown in the ACPD version of the paper. This table will be included in the revised version of the paper and the text will be updated accordingly.

		Equator		40°N		80°N	
	Expected	LL	GG	LL	GG	LL	GG
J	1.0	0.5029	0.5208	0.5020	0.5200	0.5013	0.5166
$H(\mathbf{x})$	1.1	1.1015	1.0966	1.101	1.0971	1.0999	1.0967
$L_{\mathrm{h}}^{\mathrm{lat}}$	600	576	622	576	619	581	622
$L_{\mathrm{h}}^{\mathrm{lon}}$	600	575	574	576	575	601	607
L_{v}	3	2.88	2.87	2.88	2.87	2.9	2.9
$\mathrm{RMS}_{\mathrm{lat}}$	0	3.0E−5	2.6E−5	2.9E−5	2.0E−5	2.3E−5	2.0E−3
$\mathrm{RMS}_{\mathrm{lon}}$	0	3.1E−5	1.2E−4	3.7E−5	1.3E−4	7.3E−5	4.0E−4
$\mathrm{RMS}_{\mathrm{lev}}$	0	5.4E−5	2.0E−4	5.2E−5	1.7E−4	4.7E−5	1.6E−3

2. *The paper states (P. 16782, line 372) that the reason for the difference in the fitted correlation lengthscales to the true lengthscale is not known. Would any light be shed on this problem by repeating the experiment with different grid resolutions? If the resolution is difficult to control in this system, then a similar experiment could be performed by changing the correlation lengthscale instead, where the* L_h/(grid length) *would be the important quantity.*

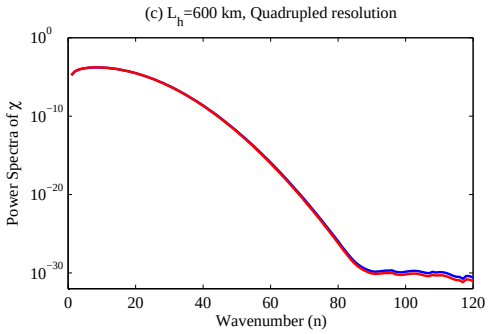
   Such kinds of experiences have been done for example by doubling the number of latitudes and longitudes (quadruple resolution). No significant changes have been observed against the experiment presented in the paper. The question thus remains open.

3. *The discussion of fig 2 (P. 16782, line 26 - P. 16783, line 8), there appears to be an underlying assumption that the higher the value of the power, the better, as more information is being provided by the observations.*

   With this figure, we rather try to see how the mapping from the Gaussian grid to the model grid degrades the analyses obtained using the GG grid, where the analyses obtained with the LL grid are taken as reference. Fig. 2 tells us that working with the Gaussian grid and a mapping operator provides analyses where horizontal correlations are overestimated (underestimated) at low (high) frequencies. This will be clarified in the paper.

   I would say that there is an optimal size of the power spectrum - over and under this value would represent a suboptimality of the assimilation. I would suggest (either for this paper or for future work) that the authors attempt some analytic work (or at least some numerical work on a very high resolution grid which introduces negligible finite-size errors) to derive an optimal spectrum.

   Such numerical experiments have been performed with a horizontal resolution of $0.75^\circ \times 0.75^\circ$. For GG and LL experiments, we found that no additional information is provided for wavenumber below $\sim$85 and $\sim$180 for, respectively, $L_h = 600$ km (see the correlation spectra in the figure below) and $L_h = 300$ km (not shown). In this case, the model resolution could have been reduced to around $2^\circ \times 2^\circ$ and $1^\circ \times 1^\circ$ for, respectively, $L_h = 600$ and $L_h = 300$. In real case assimilation where the horizontal correlation coefficients are calibrated, this test could be used to estimate the optimal model resolution. This will be discussed in the paper (and the figure below will be displayed).



4. *The authors may wish to interpret the improved impact of the CORREL results over the DIAG results in figs 3-6 in the following way. The structure functions in CORREL are broader than in DIAG, but (presumably) have the same amplitude as the deltafunction- shaped structure functions in DIAG. (What I mean by structure functions is the following: structure functions are like correlation functions, but relate to covariance instead or correlation.) This means that each observation in CORREL is responsible for giving rise to larger total analysis increments when integrated over a region. When observations are biased with respect to the background (as is hinted to here for MIPAS, P. 16785, line 14), this means that the effect of neighbouring observations don't act to "cancel each other out" over a region, and so the effect is enhanced. It would be interesting to see how the results compare when the MIPAS observations are bias corrected. Also would an increased Σ in Eq. (36) have the same effect? An experiment comparing DIAG (with an increased Σ) with CALLIB (with the normal Σ) would be interesting to see. The increased Σ value would be chosen such that the area under each structure function is the same for CALLIB and DIAG.*

   If we correctly undestand this comment, the reviewer suggests us (1) to correct the bias of MIPAS in the UTLS and (2) to calibrate the background error standard deviation. To comment (2), several recent assimilation experiments of MIPAS using a $\mathbf{B}$ calibrated by the NMC method do not have shown any significant improvements in the OmF statistics (like those exhibited in Fig. 3). Currently, it is not clear for us why the system does not benefit from a calibrated $\mathbf{B}$ while it is a crutial task in meteorological assimilation.

   Regarding (1), no effort have been done to try to remove the bias in the MIPAS data in the UTLS. The reason is that such a kind of modification is always tricky to implement (and we have no experience in bias correction). For example, how do we define the boundary between the equatorial region and the mid-latitudes where the bias in the MIPAS data is different? Moreover, the signal-to-noise ratio of the MIPAS spectra decreases in the UTLS where the ozone concentration is relatively small. So, in addition to the bias revealed by the validation effort, the MIPAS data are also much more noisy in the UTLS. This means that correcting the bias of the data could probably reduce the bias

in the OmF but probably not the standard deviation in the OmF.

So, we suggest the following minor changes in the manuscript:

- Replace "Thus, it would ..." on P16785L15-16 by "Removing the bias in the MIPAS data could help to improve the analyses in this region. Nevertheless, this task has been postponed to future studies."
- Replace "..., nothing tells us that the correlations have a Gaussian shape ..." on P16785L17-18 by "..., nothing tells us that the background error standard deviation is 30% of the first guess field, that the correlations have a Gaussian shape ..."

5. *The first part of appendix A4 is very useful, but I think that it could be improved. It attempts to show that if error covariances on the sphere are homogeneous and isotropic then the error covariance between the spectral modes is diagonal. In going from (A18) to (A19) a result is used from an earlier part of the paper Eq. (21) which is valid when one of the points (to compute covariance with) is at the pole. As it stands this is therefore not a proof of homogeneity, unless it can be shown that (21) is valid for any two points.*

   It should be said that Eq. (21) is valid for any pair of points, even if none of the points is located at the pole. This is due to the fact that homogeneous and isotropic correlations over the sphere are invariant with rotation. This is clarified by replacing P16774L8-11 by:

   "Let us come back to Eq. (19). As homogeneous (and isotropic) correlations over the sphere are invariant with rotation, let us suppose that one of the two points is at the North Pole. Then θ is the co-latitude angle, i.e. $\theta \equiv \frac{\pi}{2} - \phi$. Consequently, we have $\cos\theta = \mu$. In this configuration, the correlations between the two points are independent of the longitude. "

   An easier way may be to show that homogeneity and isotropy follow from the imposition of a diagonal covariance matrix in spectral space. This may be done as follows. Equation (A.17) is the structure function between points Ω and Ω' (thinking of Ω' as fixed and the structure function being a function of Ω):

$$\langle \epsilon(\Omega)[\epsilon(\Omega')]^* \rangle = \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{n'=0}^{N} \sum_{m'=n'}^{n'} \langle \epsilon_n^m [\epsilon_{n'}^{m'}]^* \rangle Y_n^m(\Omega)[Y_{n'}^{m'}(\Omega')]^*. \qquad (1)$$

   When the covariance matrix in spectral space is diagonal, then (A.21) holds giving:

$$\langle \epsilon(\Omega)[\epsilon(\Omega')]^* \rangle = \sum_{n=0}^{N} \sum_{m=-n}^{n} \sum_{n'=0}^{N} \sum_{m'=n'}^{n'} b_n \delta_n^{n'} \delta_m^{m'} Y_n^m(\Omega)[Y_{n'}^{m'}(\Omega')]^*, \qquad (2)$$

$$= \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n Y_n^m(\Omega)[Y_{n'}^{m'}(\Omega')]^*. \qquad (3)$$

The addition theorem (A.15) then results in:

$$\langle \epsilon(\Omega)[\epsilon(\Omega')]^* \rangle = \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n (N_n^0)^2 P_n(\cos\theta), \tag{4}$$

where θ is the angular separation between Ω and Ω'. This is a general result and does not rely on (21).

We are not sure that this way is correct. By imposing that the covariance matrix is diagonal, we find that $\langle \epsilon_n^m [\epsilon_{n'}^{m'}]^* \rangle = b_n^m \delta_n^{n'} \delta_m^{m'}$ (note the b_n^m instead of the b_n written in (5)). So, no modification has been implemented in App. A4.

6. *In table 1, the expected value of J is quoted as 1. In think that the value should be 1/2 for a single observation (half of the number of observations). This is a statistical expectation and not a precise result expected every time. This is a chi-squared statistic and so a little needs to be done to assess which of the results (i.e. LL or GG) is best (or if they are distinguishable). The reason for the empirical values close to 1, rather than 1/2, may be an omission of the 1/2 in the definition of the cost function that has been coded-up.*

   The expected value of J is indeed 1/2 when the background and observational errors are optimaly setup. This would have meant to use a value of $\sqrt{0.02}$ ($\sim$0.14) instead of 0.1 for these errors. The experiments have been reprocessed with this new value. We will also add the following sentence on P16781, L13:"... error standard deviation of $\sqrt{0.02}$. With this configuration, the value of J after assimilation is expected to be 1/2."

Technical corrections

I have found a number of minor points with the paper, but each is very easy to correct.

1. *P. 16764, line 9: "interpolating on" change to "interpolating to".*

   This will be corrected.

2. *2. P. 16764, line 22: The paper says that two approaches have been developed to represent spatial correlations in variational data assimilation (spectral and diffusion operators). A third approach is also known, namely recursive filters (see e.g. Purser et al., 2003, MWR 131, 1524-1535).*

   The work of Purser et al. will be mentioned in the introduction as follows (after P16765L22): "A third approach in variational data assimilation which does not compute explicitly the error covariance is the recursive filter (Purser et al., 2003). Like the diffusion operator approach, the recursive filter approach attemps to compute Gaussian correlations. The method consists in evaluating the effect of a Gaussian correlation model on a state vector, by applying a sequence of 1D finite difference operators in different directions on the state vector. Repeated applications of these finite difference operators in carefully chosen directions can lead to approximate the smoothing effect of Gaussian

homogeneous and isotropic correlations. Although positive-definitness can be obtained, the scheme is approximated and computationally complex."

3. *P. 16765, line 9: "done" change to "made".*

   This will be corrected.

4. *P. 16766, line 9: The paper states that the inverse transform is not needed in variational data assimilation. This is true if a calibration step is not completed (see point x below) and if the 'guess' (or reference) state is the same as the background state. The calibration step requires a population of forecast errors to be known in spectral space (from model space) so the vertical covariance matrices and spectral coefficients of the horizontal spectra can be determined. If the 'guess' state, x_g is not the same as the background then the difference between x_b and x_g needs to be put in control variable space (call this χ_b), where the right hand side of (31) becomes $(\chi - \chi_b)^T(\chi - \chi_b)$. The inverse transform is need to do this.*

   We agree with the reviewer on the fact that the inverse transform is necessary for the calibration of $\mathbf{B}$. However, the inverse should not be necessarily the "exact" inverse such that the inverse method presented in App. A2 can be used. Another option would also be to interpolate the error fields to the Gaussian grid before operating the inverse transform in order to estimate the coefficients of the correlation matrix. This is clarified in Sect. 4 (and not in the introduction as in that section, we refer to the exact inverse transform), after P16780L5:" ... complete review. It is important to note that the calibration method will require an inverse of the operation $\mathbf{S}$ in order to estimate the correlation spectra b_n. If one uses an equally spaced model grid, the exact inverse of $\mathbf{S}$ is not necessary and the method described in App. A2 can be used. Even more simpler would be to interpolate the error fields to the Gaussian grid before the inversion."

5. *After Eq. (1): Please define how μ is related to latitude.*

   μ is already defined in the introduction of Sect. 2 (P16767 L7). Do we also need to redefined μ after Eq. (1)?

6. *In connection with Eqs. (8) and (10): In practice, a fast Fourier transform will be done instead of literal evaluations of the summations shown. This warrants a mention in the paper.*

   The use of fast Fourier transform will be mentioned.

7. *P. 16769, line 8: Here and at many other points in the paper the "spectral grid" referred to seems to be a lat/long space that is used immediately before/after performing the spectral transform. This is not in spectral space and so I am puzzled why it is called a spectral grid. In my mind the spectral grid is the set of n and m values, but then I might not know what the convention is for this.*

We understand the confusion rised by the term "spectral grid". In our mind, the set of n and m values are represented by the "spectral coefficients", not by the spectral grid. For us, the term of grid is applied to any discretisation of the physical space. As the spectral space is by definition discrete, the termonology of grid is then not applicable in that space.

In the paper, the "spectral grid", which belong to the physical space, is defined as the target grid of the spectral operator (see P16770 L5-6). In the paper, two types of grid have been used: the equally spaced grid and the Gaussian grid. Depending that the spectral operator acts to one of those grids, the spectral grid is either the equally spaced grid or the Gaussian grid.

In order to clarify this point in the paper, the term "spectral grid" will be replaced by "target grid of the spectral transform" or simply by the "target grid".

8. *Before Eq. (11): I think that it would aid some readers if the Gaussian quadrature formula is given in general form (ie for an arbitrary integrand), which will lead to the derivation of (11).*

   The general form of the Gaussian quadrature will be given.

9. *P. 16770, line 19: Replace "allows" with "allows one" or "allows us".*

   This will be corrected.

10. *P. 16771, line 6: Here it states that "Since $\psi(\lambda, \mu)$ is real we have $S^* = S^T$". Although $\psi(\lambda, \mu)$ is real, this does not mean that S is a real-valued operator. Indeed (8) shows that this operator is complex.*

    This will be clarified.

11. *Eqs. (16) and (17): These are adjoint equations and so they no longer relate to ψ fields. The notation that I am familiar with is to add hats, ie $\hat{\psi}$, to show that they are 'adjoint variables' (i.e. derivatives with respect to ψ, i.e. $\hat{\psi} = \partial J / \partial \psi$).*

    The adjoint notation $\hat{\psi}$ will be used in the paper.

12. *P. 16772, line 22: Replace "allows" with "allows one" or "allows us".*

    This will be corrected.

13. *Eq. (19), and all equations that use expectation. The second item should be subject to a * operator to indicate transpose and complex conjugate. This is not necessary when the second element is real (as in (19)), but is a good habit to get into. It is necessary in, eq. (22) and many of the equations in appendix A.*

    This will be corrected.

14. *Eq. (30): This should read $B = LL^T$ (or more strictly $B = LL^*$).*

    This will be corrected.

15. *Eq. (35): The derivative of H should be with respect to x and not to χ.*

    This will be corrected.

16. *P. 16777, line 10: If not converged, the algorithm should go back to step 2, not step 3.*

    This will be corrected.

17. *P. 16778, line 13: Replace "allows" with "allows one" or "allows us".*

    This will be corrected.

18. *Eq. (39) and comment that follows it: Λ is a symmetric matrix, but $\Lambda^{1/2}$ need not be. Such matrices have an infinite number of valid square-roots. Some of them are symmetric, in which case $\Lambda^{1/2} = \Lambda^{1/2*}$. Unless the authors are restricting the square- root to a symmetric case then (39) holds, otherwise a * should be added to $\Lambda^{1/2}$.*

    $\Lambda^{1/2}$ is replaced by $\Lambda^{1/2*}$ in Eq. (39).

19. *P. 16779, line 11: Replace "allow" with "allow one" or "allow us".*

    This will be corrected.

20. *P. 16779, line 16: Replace "never" with "not yet".*

    This will be corrected.

21. *P. 16779, line 25: Replace "matrix" with "matrices".*

    This will be corrected.

22. *P. 16782, lines 10-11: Since the lines on fig 1 are so close to each other, why not plot differences?*

    Here below is a figure which includes the differences and which will replace the Fig. 1 of the paper.

23. *P. 16786, line 7: Replace "in" with "of".*

    This will be corrected.

24. *P. 16787, lines 2-3 (two occurrences): Replace "allow" with "allow one" or "allow us".*

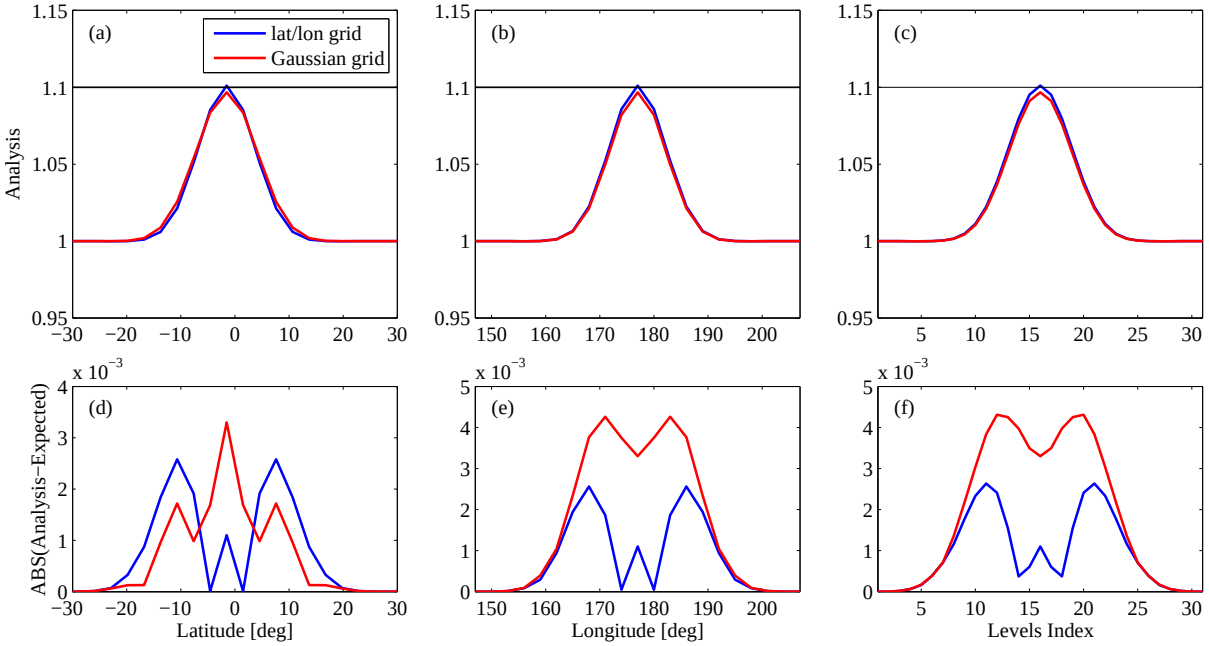
    This will be corrected.

25. *P. 16787, line 8: Replace "in" with "of".*

    This will be corrected.

26. *P. 16787, line 9: Replace "method" with "methods".*

    This will be corrected.



27. *P. 16788, line 8: Replace "value" with "values".*

    This will be corrected.

28. *P. 16789, line 11: Replace "introduced" with "introduce".*

    This will be corrected.

29. *Eq. (A14): I think that the μ and μ' in this equation should be ϕ and ϕ' respectively.*

    This will be corrected.

30. *Eq. (A20): The m summation index should be m".*

    This will be corrected.

I hope that these comments are useful to the authors. R.N.Bannister.

Yes, thank you very much.