I must first apologize to the authors for the long delay in sending this review.

The paper presents results that basically deserve publication in my mind. At the same time, I consider a number of clarifications, and also some additional tests, are necessary before the paper can be accepted. My main comments are as follows.

1. There is some confusion in the paper about the roles of linearity and Gaussianity in assimilation. The abstract reads *Two common derivations respectively lead to the Kalman filter and to variational approaches. They rely on either assumptions of linearity or assumptions of Gaussianity of the probability density functions of both observation and background errors.* Maybe I am mistaken on the authors' intentions, but these sentences mean in effect that the hypotheses of linearity (leading to Kalman filter) and Gaussianity (leading to variational assimilation) are mutually exclusive. They are not. Both Kalman filter and variational assimilation are based on the same linear assumptions (and both are empirically extended to weakly nonlinear situations). Under these linear assumptions, they are only two different algorithms that solve the same problem. In addition, they both achieve Bayesian estimation in the case when the errors affecting the data are Gaussian. More precisely

P. 1063, ll. 10-11. …, *up to now operational Numerical Weather Prediction (NWP) has relied on assimilation schemes that are Gaussian* …. The authors do not say which assimilation schemes they have in mind, but I presume they mean schemes of the general 'Kalman' form

$$\boldsymbol{x}^a = \boldsymbol{x}^b + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b) \tag{1}$$

where $\boldsymbol{x}^b$ and $\boldsymbol{x}^a$ are respectively the background and the analysis, $\boldsymbol{y}$ is the observation, $\boldsymbol{H}$ the corresponding (linear) observation operator, the difference $\boldsymbol{d} \equiv \boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}^b$ being the innovation vector. $\boldsymbol{K}$ is the gain matrix which, in the context of least variance estimation, is defined as $\boldsymbol{K} \equiv \boldsymbol{C}_{zd}\, \boldsymbol{C}_{dd}^{-1}$, where $\boldsymbol{C}_{zd}$ is the cross-covariance matrix of the background error $\boldsymbol{z} \equiv \boldsymbol{x} - \boldsymbol{x}^b$ with the innovation, and is $\boldsymbol{C}_{dd}$ is the covariance matrix of the innovation itself**.**

I stress there is nothing necessarily 'Gaussian' in Eq. (1) above. That equation can be obtained as defining the *Best Linear Unbiased Estimator* (*BLUE*) of $\boldsymbol{x}$ from $\boldsymbol{x}^b$ and $\boldsymbol{y}$, independently of any Gaussian hypothesis. It can also be obtained, also independently of any Gaussian hypothesis, on a principle of maximum entropy. Linearity, on the other hand, is always necessary. Gausianity is only a 'plus' which, if it comes in addition to linearity, ensures Bayesianity of the estimation.

The authors write (p. 1065, ll. 2-3, *efforts* [to] *be made to improve linear assumptions* … Well, if Gaussianity is obtained at the expense of linearity, this may result in a degradation of the accuracy of the final estimate.

P. 1064, ll. 7-8. *It* [the 4D-Var algorithm] *solves for the most probable state* […] *by minimizing a non-quadratic cost-function* …. If there are nonlinearities and the cost-function is non-quadratic, it is very unlikely that minimizing it will lead to the most probable state. Actually, that is guaranteed only in the linear and Gaussian case.

Please revise all parts ot the paper relative to the basic principles of assimilation and to the questions of linearity, Gaussianity and bayesianity. It must be clear in particular that, among the hypotheses to be made for Kalman filtering and variational assimilation, linearity must come before Gaussianity.

2. The significance of the D'Agostino test, and the interpretation to be given to the results it produces, must be clarified.

I mention first that formulæ (2) and (3) for the skewness G_3 and the kurtosis G_4 are not exact. The denominator in the expression for the variance should be N_S-1, and similar corrections are to be made in the expressions for the third- and fourth-order moments.

More importantly, the fundamental purpose of the test is the following. For given ensemble size N_S and exact Gaussianity, by how much can one expect G_3 and G_4 to deviate from their Gaussian values 0 and 3 ? The authors define transformed skewness and kurtosis $f_3(G_3)$ and $f_4(G_4)$ through formulæ whose significance is obscure (and which would be in my opinion more appropriately put in an appendix than in the main text of the paper). The transformed $f_3(G_3)$ and $f_4(G_4)$ are said to be standard Gaussian (*i.e.* with expectation 0 and variance 1) if the original variable is Gaussian. For which values of N_S is that statement true (it cannot be for any N_S, in view for instance of a term N_S -3 in several of the formulæ leading to the definition of $f_4(G_4)$) ?

The next step is to test the Gaussianity of the transformed $f_3(G_3)$ and $f_4(G_4)$. But what is then the interest of making the test on $f_3(G_3)$ and $f_4(G_4)$ rather than on the raw G_3 and G_4 ? Is it that a possible non-Gaussianity will show up more clearly on the former ? Is so, say it clearly. In any case, explain.

The authors then introduce the parameter K^2 of which they write (p. 1068, ll. 9-10) that it follows an approximate χ^2 distribution with two degrees of freedom. Well, if $f_3(G_3)$ and $f_4(G_4)$ are independent standard Gaussians, K^2 will follow an exact χ^2 distribution with two degrees of freedom (with expectation 2 and variance 4). Is it because G_3 and G_4 are not independent in the first place that the distribution cannot be expected to be an exact χ^2 ?

It is not clear how the values obtained for $f_3(G_3)$, $f_4(G_4)$ and K^2 must be interpreted. The authors write (p. 1069, last sentence) *describing the values of K^2 has the advantage to prevent the results from depending on the chosen confidence level*. Which confidence level are you referring to ? A level similar to the one given (p. 1068, l. 11) for N_S = 100 ? But that does not say how to interpret the values obtained for K^2. One could expect that a χ^2 mean value of 2 for K^2, with a variance of 4, could be interpreted as proof of Gaussianity. And you mention a value of 2.7 (p. 1072, l. 11) as indicating Gaussianity. But Fig. 3a shows values, at all levels and for all variables except q, which are about 4, which seems to indicate significant deviation from Gaussianity. Nevertheless, you write in the conclusion (p. 1076, l. 14) *Deviation from Gaussianity for U, V , and T only appears in the boundary layer*. All that is confusing.

A similar remark applies to the parameters $f_3(G_3)$ and $f_4(G_4)$, of which it is not clearly said (except for the large values of $f_3(G_3)$) how they must be interpreted. For instance, how the fact that the values of $f_4(G_4)$ are positive in Fig. 3c must be interpreted ($f_4(G_4)$ clearly does not have the standard Gaussian distribution to be expected if the basic variables are Gaussian) ?

All those aspects must be clarified. In particular, explain in what it is better to use the parameters $f_3(G_3)$ and $f_4(G_4)$ (and K^2) rather than the raw diagnostics G_3 and G_4. And explain better how the values found for $f_3(G_3)$, $f_4(G_4)$ and K^2 must be interpreted (see also comment 4 below).

3. Subsection 4.2.1 and associated Fig. 9. You present diagnostics for control variables, and particularly vorticity and divergence and for a 3-hour forecast. You have shown previously that, for other variables, the analysis ensembles are more Gaussian than the forecast ensembles. I suggest you also present diagnostics for the analysed control variables.

4. Subsection 4.2.1. You write on the basis if Fig.9 that the vorticity, unlike the wind components, is strongly non-Gaussian. This is what comparison of Figs 3 and 9 may suggest, but the vorticity is a linear function of the wind components, and cannot be as such be less Gaussian than those components. This requires clarification.

5. Concerning also Fig. 9, you write that the unbalanced divergence η_u, like vorticity, is strongly

non-Gaussian, while the variables T_u and q_u display much more Gaussian profiles. Well, according to the caption of Fig. 9, it is T_u which, in addition to vorticity, shows large values of K^2, while η_u shows smaller values. Is there an error in the caption, or what ?

And, speaking of vorticity, you use the Greek letter ξ (pronounced *xi*) to denote it. The usual notation is ζ (pronounced *zeta*). I suggest you follow the established practice.

6. P. 1073, ll. 9-11. *For q, NG is mainly found in "cloudy" areas,* [...] *with two peaks around 900 and 700 hPa*. According to Fig. 6a, there is a much more marked peak in the layer 100-300 hPa.

7. Abstract, ll. 18-19, *The mass control variables used in our data assimilation, namely vorticity and divergence*. Well, vorticity and divergence are not mass variables (check for other possible similar mistakes elsewhere in the paper)

8. P. 1066, ll. 3-4, *Positive (negative) values are associated with a mode of the PDF smaller (larger) than its mean*. This statement may not be true of the mode of the distribution (which can be arbitrarily modified with infinitesimal change to the distribution), but is true of its median.

9. And there are erroneous statements concerning the relationship between skewness and tails pp. 1071, l. 13, and 1072, l. 1.

10. P. 1068, l. 11, what is *unilateral testing* ?

11. P. 1073 and 1076, ll. 11 and 19, *forecast ~~terms~~ ranges*

12. P. 1064, l. 15, *Laroche and Pierre, 1998*. Do you mean *Laroche and Gauthier* ?

13. P. 1064, l. 20 (and elsewhere). The proper spelling is *Järvinen* (with a diaeresis)

14. P. 1077, ll. 7-8, ... *does not include model error, ~~n~~either in the analysis ~~n~~or in the forecast steps* (what you write is analogous to writing in French *Je n'ai pas vu personne*)

15. P. 1076, l. 3, *below* the tropopause