Dear Pr. Talagrand,

Thank you for those recommendations. Please find below, our answers to your comments and the associated changes added to the manuscript. Most of those responses refer to the ones we made for the three referees. Also, the three answers to each referee, and a latexdiff version of the manuscript have been joined below.

Best regards,

Raphaël Legrand, Yann Michel and Thibaut Montmerle

(Editor comments are written in black, and authors answers are in blue)

Comments from Editor→[...] I first mention that you have not in my opinion responded properly to the request I made before your paper was published in NPGD. That request was that you mention explicitly the 'Gaussian' values for the three criteria ($f_3(G_3)$, $f_4(G_4)$, and K^2) you use for estimating the Gaussianity of the ensembles. I understand the Gaussian value for $f_3(G_3)$ is 0, which is in agreement with your discussion of the results you obtain for that quantity. Concerning $f_4(G_4)$, you imply (l. 127) that it must have expectation 0 (otherwise the quantity K^2 would not follow a χ^2-distribution). But $f_4(G_4)$ takes in Fig. 3 values that are close to 1, and that you seem to consider as proof of Gaussianity. As for K^2 , if it follows a χ^2-distribution with two degrees of freedom (l. 127 again), it must have expectation 2. However, it almost always (if not always) takes larger values in your results, and you seem in particular to consider that the value 4 indicates Gaussianity. The inconsistency (which may result from nothing more than a change of origin or rescaling) must be explained.
Author's response→ The 'Gaussian' values for the three criteria ($f_3(G_3)$, $f_4(G_4)$, and K^2) are respectively $f_3(G_3) = 0$, $f_4(G_4) = 0$, and $K^2 = 2$. You are right to say that $f_4(G_4)$ takes in Fig.3 values that are between 0 and 1, and we state that this corresponds to nearly Gaussian behaviour. Indeed, using finite sampling with ensemble size $N_s > 20$ (Thode, 2002), $f_3(G_3)$ and $f_4(G_4)$ could be both assumed to follow a Gaussian law with a zero mean and a unity variance. In this case, K^2 follows approximately a χ^2 distribution with two degrees of freedom. Confidence intervals at 95% are then given by $f_3(G_3) \in [-1.96; 1.96]$, $f_4(G_4) \in [-1.96; 1.96]$, and $K^2 \in [0; 5.991]$. Thus, because G_3 and G_4 are uncorrelated but not independent, K^2 does not follow an exact χ^2 distribution, and confidence interval is slightly different. Using a right-tailed unilateral testing at 95% for $N_s = 100$, the critical value of K^2 is 6.271 instead of 5.991.(for the exact χ^2 distribution).

Now, many figures (including Fig. 3) present vertical profiles of horizontally averaged $f_3(G_3)$, $f_4(G_4)$, and K^2. Because quantities are horizontally averaged, they cannot be compared directly to the bounds above. Nevertheless, we shall consider that values below (above) those bounds in horizontal average indicate dominance of Gaussianity (non-Gaussianity), or simply state that some variable exhibit more (less) Gaussian behaviour when its averaged K^2 is lower (higher) than for another one.
Author's changes in manuscript→Additional comments have been added in section 2.1 as : "When testing a Gaussian distribution, asymptotic values for the three criteria ($f_3(G_3)$, $f_4(G_4)$, and K^2) are respectively $f_3(G_3) = 0$, $f_4(G_4) = 0$, and $K^2 = 2$. Using finite sampling with ensemble size $N_s > 20$ (Thode, 2002), $f_3(G_3)$ and $f_4(G_4)$ could be both assumed to follow a Gaussian law with a zero mean and a unity variance. In this case, K^2 follows approximately

a χ^2 distribution with two degrees of freedom. Confidence intervals at 95% are then given by $f_3(G_3) \in [-1.96; 1.96]$, $f_4(G_4) \in [-1.96; 1.96]$, and $K^2 \in [0; 5.991]$. Because G_3 and G_4 are uncorrelated but not independent, K^2 does not follow an exact χ^2 distribution, and confidence interval is slightly different. Using a right-tailed unilateral testing at 95% for $N_s = 100$, the critical value of K^2 is 6.271 instead of 5.991."

Comments from Editor→[...] [Referee 3] asks you in particular to include additional results in your paper, in the form of diagnostics of the Gaussianity of the control variables (subsection 4.2.1, his comment 3). He also questions the usefulness of the D'Agostino test, and of checking the Gaussianity of the quantities $f_3(G_3)$ and $f_4(G_4)$, rather than of the raw diagnostics G^3 and G^4 (his comment 2). It seems to me that you have actually already answered that point, by saying that the quantities $f_3(G_3)$ and $f_4(G_4)$, and the resulting K^2, are more appropriate for small ensembles (ll. 106-108). But it may be useful to be more specific, by saying for which values of the ensemble size N the D'Agostino test can be useful (referee 3's comment 2).
Author's response→Those questions are answered in our responses to Referee 3.
Author's changes in manuscript→In section 2.1, interpretations of $f_3(G_3)$ and $f_4(G_4)$ values are added, and it is now explicitly mentioned that D'Agostino is useful for samples larger than 20 members. Fig.1 has been changed to be in accordance with this threshold. Moreover, as regards NG of control variables, comment has been added in section 4.2.1.

Comments from Editor→Both referees 2 and 3 observe that vorticity and divergence, although they are linear functions of the wind components, are much less Gaussian than the latter (their comments 8, and 4-5 respectively). Referee 2 mentions heteroscedasticity (spatial variation of the variance of wind components) as a possible explanation. Please discuss this point.[...]
Author's response→This point is answered in our response to Referee 1.
Author's changes in manuscript→cf response to Referee 1, a comment on section 4.2.1 is added on this subject.

Références

Thode, H. C. : Testing for Normality, in : Vol. 164 of Statistics : Textbooks and Monographs, Marcel Dekker, New York, 2002.

REFEREE 1 :
(Reviewer comments are written in black, and authors answers are in blue)

Specific comments :

Comments from Referee→(1) p.1072, L.5-9 : what is the implication of small K^2 in the regions of large ensemble forecast variance ? Could you elaborate in more detail ?
Author's response→ Having small K^2 values implies that according to our diagnostics, background errors are sampled from a Gaussian distribution, which is consistent with the hypothesis of Gaussianity made in our data assimilation system. Additionally, in presence of large background error variances, the background is assumed to be less reliable and more weight should be put on observations in that area during the analysis.
Author's changes in manuscript→ None.

Comments from Referee→(2) p.1075 (section 4.2.1) : Could you further elaborate on vorticity and divergence control variables and the possible reasons for their non-Gaussian behaviour. Is it possible that this is related to their definition as second derivatives of stream function and velocity potential, both commonly used as Gaussian ? Would this suggest it may be better to use stream function and velocity potential as control variables in order to stay within the Gaussian framework ?
Author's response→The first or second order derivation use linear operators. So theoretically the derivative of an initially Gaussian distribution, is also Gaussian. We suspect however that derivation of a nearly-Gaussian process may indeed increase its NG. Furthermore another possible source of NG has been highlighted by the second reviewer : the heteroscedasticity (spatial variability of the variance) of the wind fields.
To go further on this topic, NG diagnostics have been computed for the temperature T, which is a nearly Gaussian field (cf Fig. 3a), for the temperature normalized by its standard deviation $\frac{T}{\sigma_T}$, and for their respective first-order spatial derivatives ($\frac{\partial T}{\partial x}$ and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$). Results are gathered in the attached new figure (Fig.1 of the comment) giving vertical profiles of K^2 for T (solid black line), $\frac{T}{\sigma_T}$ (dashed red line), $\frac{\partial T}{\partial x}$ (dotted blue line), and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$ (dot-and-dash green line). Profiles have been computed from a 90-members ensemble of 3h-forecasts valid the 4^{th} of November 2011 at 03 :00. Three conclusions arise from this experiment :

- the profiles of T and $\frac{T}{\sigma_T}$ are almost equal (differences smaller than 0.01). This supports the fact that NG diagnostics for a particular parameter do not depend on its variance.
- the large increase of K^2 between fields (normalized or not) and their derivative seems to support the fact that derivation of a nearly Gaussian variable increases its NG. Moreover, despite the use of T instead of one of the wind components, the order of magnitude of the NG is close to the one found in Fig.9(a) and Fig.10 for the vorticity and the divergence, which support the attribution to derivation for at least a part of their NG.
- differences of NG between $\frac{\partial T}{\partial x}$ and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$ enable to know more about the impact of heteroscedasticity. It seems that the homogenization (all variances set to 1 with the normalization) yields a systematic decrease of NG for every model levels yet very small compare to the increase implied by the spatial derivation.

Those conclusions are in accordance with the large NG of ζ and η which are defined as spatial first order derivatives of the largely Gaussian wind fields.
Considering those results, the couple stream function/velocity potential should display more

Gaussian behaviour than ζ/η, which would make them good candidates for being the dynamical control variable. (as in the Met Office assimilation system for instance). Future work is however still needed to confirm this point.
Author's changes in manuscript→ Added in section 4.2.1 : "To go further on this topic, NG diagnostics have been computed for the spatial first-order derivative of T. While T is a nearly Gaussian variable (see Fig.3a), its spatial derivation largely increases the NG (not shown), up to the order of magnitude found in Fig.9a and Fig.10 for ζ and η. This supports the attribution to derivation for at least a part of the NG displayed for the dynamical control variables.".

Technical corrections :

Comments from Referee→(3) p.1063, L.13 : Delete "Of course", start sentence with "In general ...". Substitute "will lead" by "could lead", unless there is a reference stating that. In that case include the reference.
Author's response→We agree with this correction
Author's changes in manuscript→ Changed in the introduction : "The time integration of the model nonlinear dynamics leads inevitably to non-Gaussian forecast errors (Bocquet et al., 2010).".

Comments from Referee→(4) Figs.3, 6, 8, 9, and 10 : It is difficult to distinguish between the dotted and dashed lines. Would it be possible to recreate this figure with more distinct lines ? Also, please include in the figure caption the description of lines (e.g., dashed, dotted, full,. . .).
Author's response→We agree with this correction
Author's changes in manuscript→Figs.3, 6, 8, 9, and 10 have been recreated with clearer line settings and colors. Moreover Fig6(c) has been simplified. Lines descriptions are retrieved in legends on top of each figure, in order to keep the captions as synthetic as possible.

Références

Bocquet, M., Pires, C. A., and Wu, L. : Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, 2010.

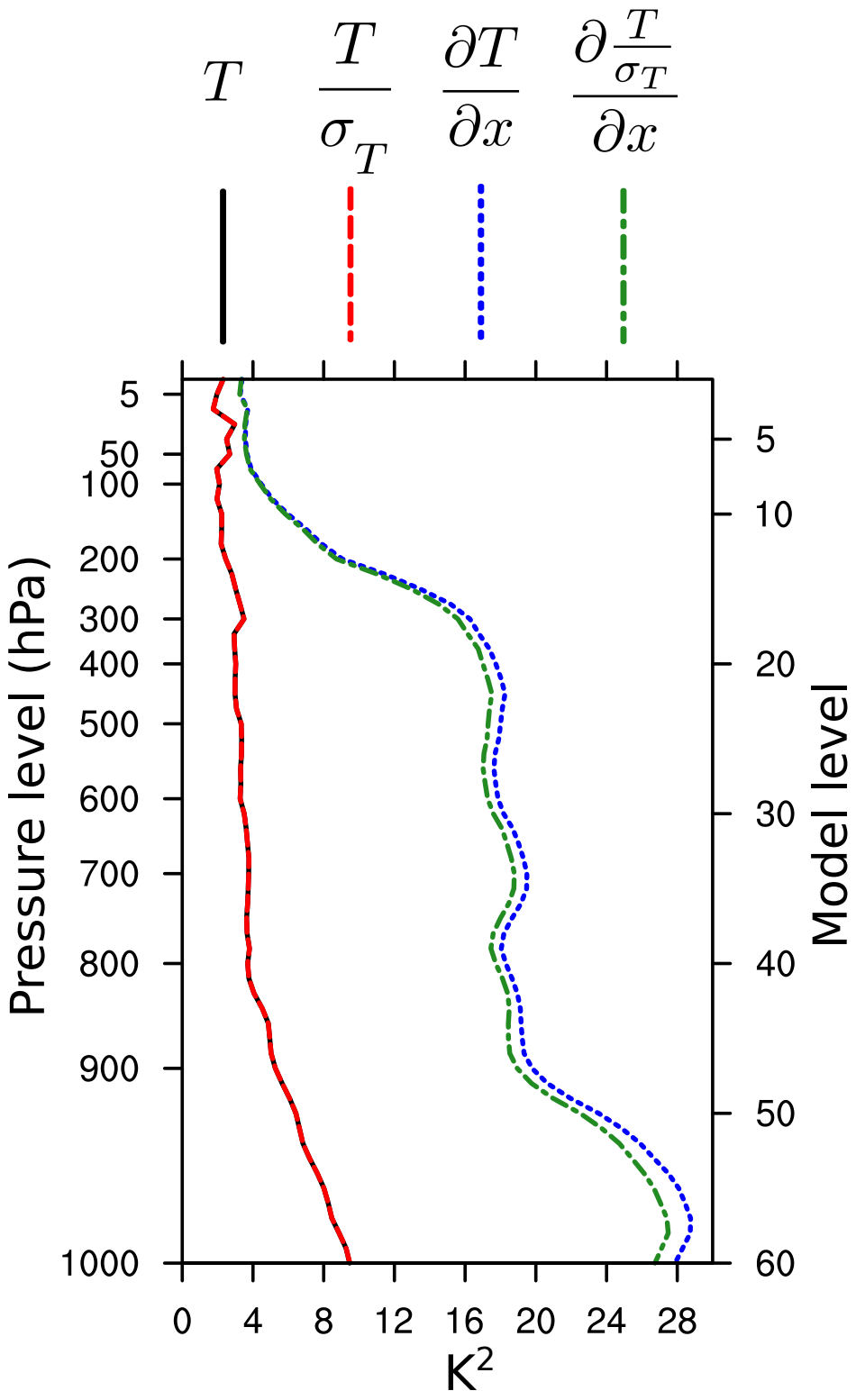

FIGURE 1 – Vertical profiles of K^2 for T (solid black line), $\frac{T}{\sigma_T}$ (dashed red line), $\frac{\partial T}{\partial x}$ (dotted blue line), and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$ (dot-and-dash green line) (see text). For each level, values are averaged over the horizontal domain. Profiles have been computed from a 90-members ensemble of 3h-forecasts valid the 4^{th} of November 2011 at 03 :00.

REFEREE 2 :
(Reviewer comments are written in black, and authors answers are in blue)

Comments from Referee→(1) The transformed estimator $f_4(G_4)$ for kurtosis was originally presented by Anscombe and Glynn(1983), after that of d'Agostino (1970) test for skewness. The omnibus test (4) in pg. 1068 was presented by D'Agostino et al. (1990). See refer- ences bellow and add to manuscript. Anscombe, F.J. ; Glynn, William J. (1983). "Distribution of the kurtosis statistic b2 for normal statistics". Biometrika 70 (1) : 227–234. doi :10.1093/biomet/70.1.227. JSTOR 2335960 D'Agostino, Ralph B. ; Albert Belanger ; Ralph B. D'Agostino, Jr (1990). "A suggestion for using powerful and informative tests of normality" (PDF). The American Statistician 44 (4) : 316–321.doi :10.2307/2684359. JSTOR 2684359.
Author's response→We agree with this point.
Author's changes in manuscript→ The two references (Anscombe and Glynn, 1983; D'agostino et al., 1990) have been added, and the K^2 is now rightly attributed to D'agostino et al. (1990) instead of D'Agostino (1970).

Comments from Referee→(2) Beyond K^2 (Eq. 4), other diagnostics of NG have been used on assimilation error and innovation diagnostics, like those relying cumulant-based expansions of the negentropy or the Kullback-Leibler divergence with respect to the fitting Gaussian pdf (Pires et al. 2010). Add the reference : Pires, C.A., O. Talagrand, M. Bocquet, 2010. Diagnosis and Impacts of non- Gaussianity of Innovations in Data Assimilation. Physica D. Nonlinear Phenomena, Vol. 239, (17), 1701-1717. doi :10.1016/j.physd.2010.05.006
Author's response→We agree with this point.
Author's changes in manuscript→References to kullback (1959), and Pires et al. (2010) have been added.

Comments from Referee→(3)- Fig 3b and 3c present profiles of the horizontal averages of f_3 and f_4 for humidity. However local values of f_3 and f_4 may exhibit quite larger and extreme values (see Figs 4b and 4c) than their horizontal averages. Figs 3b and 3c do not give an idea of the NG range over the area. A figure with profiles showing the range interval (e.g. 5-95%) of f_3 and f_4 would be useful. It would be more consistent with Fig 3a presenting spatial averages of the local K^2 values.
Author's response→Thank you for this remark. The description of the spatial variability of the NG diagnostics was indeed insufficient, and discussion of ranges for K^2, $f_3(G_3)$ and $f_4(G_4)$ has been added in the section 3.3 of the manuscript. This description is based on results displayed in Fig.1 below, which is not included in the text since many similarities have been found between parameters and materials that are already present in the article.
Author's changes in manuscript→ Added in section 3.3 : "The range, defined as the difference between the 95th and the 5th percentiles, could be used to describe roughly the horizontal spatial variability for each vertical level. Vertical profiles of ranges of K^2, $f_3(G_3)$, and $f_4(G_4)$ (not shown) have, all three, large similarities between each other, and with the shapes of K^2 profiles displayed in Fig3(a). It includes in particular two maxima in the boundary layer and in high troposphere for q and larger values towards the surface for T. Ranges are much larger for the four variables (approximately four times as large) than the respective mean values given in Fig.3, implying a large spatial variability for the three NG diagnostics. An example of the horizontal structures of NG is given for q in the boundary layer by Fig.4. They have large similarities with the meteorological coherent structures, as the Southerly convergent flow over South of France and the active cold front aloft North-West of France are associated with high values of K^2."

Comments from Referee→(4) - Pg 1071, line 15. Please be more rigorous not using the 'inversely proportional' attribute. For instance : larger values of K^2 generally occur for small values of q.
Author's response→We agree with this point.
Author's changes in manuscript→It has been corrected as : "...K^2 is increasing while the q mean content, displayed in Fig.3(d), is largely decreasing.

Comments from Referee→(5)- Pg 1072, line 1. Negative skewness - left-tailed distributions. Please correct
Author's response→We agree with this point.
Author's changes in manuscript→It has been corrected

Comments from Referee→(6)- Pg. 1073. Relationships between NG and physical processes must be analysed with care. Which makes you to link diabatic processes to NG ? NG can come from non-linear processes acting on Gaussian pdfs ; linear processes acting on NG pdfs or both. Add a short justification.
Author's response→We agree that this study does not demonstrate that NG is caused by diabatic processes. It is true that NG may come from non-linear processes acting on Gaussian pdfs ; linear processes acting on non-Gaussian pdfs or both. Yet we have shown that analysis errors have a more Gaussian behaviour in our system (e.g Fig.7 and Fig.8), such that NG in the background may rather come from non-linear processes acting on a nearly-Gaussian pdf. Also, we point out diabatic processes as a reasonable explanation based on Fig.6 that shows increased NG in cloudy conditions compared to clear sky ones.

A more extensive study of relations between physical processes and NG would require running ensembles with simplified physics, where some processes are turned off. This may be attempted in the future.
Author's changes in manuscript→Changes have been made in section 3.4 : "During the 6 first hours of forecasts, NG quickly increases [...] Those results support that NG in the background may rather come from non-linear processes acting on nearly Gaussian pdfs instead of linear processes acting non-Gaussian pdfs." and then "For T and q, diabatic processes are good candidates to produce NG because of intrinsic thresholds in cloud physics (e.g. moisture saturation) and non-linear processes like turbulence on cloud-top."

Comments from Referee→(7) - Sec. 4.1 highlights the drastic reduction of NG of the analysis compared to that of background (Fig8a,b), specially over regions of dense radar observations. This can only be due to the hypothesis of Gaussianity of observation errors (e.g. radar) which is for the moment the better hypothesis to use. Comment that.
Author's response→The analysis increment is equal to a gain matrix times an innovation vector (observations minus background), i.e it is a linear function of the innovation in model space. Thus the Gaussianity of the analysis increment mainly depends on the Gaussianity of the innovations. Practically, innovations are close to Gaussianity thanks to a rough selection applied beforehand to the observations, allowing to remove outliers.

As regards radar data, a 1D+3D Var approach is used operationally for AROME (Caumont et al., 2010). It consists in retrieving profiles of relative humidity (RH) from observed reflectivities at first, and to consider such profiles as pseudo-observations in the 3D-Var (Wattrelot et al., 2014). Only small innovations of RH are kept in the process, insuring the Gaussianity of the corresponding innovations.

Author's changes in manuscript→Additional comment added in section 4.1 : "The analysis increment being a linear function of the innovation vector in model space (observation minus background), its Gaussianity is insured by a rough selection applied beforehand to the observations, allowing to remove outliers (e.g. for radar data, Caumont et al., 2010; Wattrelot et al., 2014)."

Comments from Referee→(8) - Pg 1075. Despite the fact that vorticity and divergence are linear operators of the quasi Gaussian zonal and meridional wind fields, mostly of the NG comes from heteroscedasticidy (spatial variability of the wind variance). Refer this aspect
Author's response→This remark is similar to that made by reviewer 1 (and 2) in his second point. In order to study relative impact on NG of heteroscedasticity and spatial derivatives, NG diagnostics have been computed for the temperature T, which is a nearly Gaussian field (cf Fig. 3a), for the temperature normalized by its standard deviation $\frac{T}{\sigma_T}$, and for their respective first-order spatial derivatives ($\frac{\partial T}{\partial x}$ and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$). Results are shown and explained in the answer to reviewer 1.
Author's changes in manuscript→Same as for reviewer 1.

Références

Anscombe, Francis J. and Glynn, William J. : Distribution of the kurtosis statistic b2 for normal samples, Biometrika, 70 (1), 227–234, 1983.

Bocquet, M., Pires, C. A., and Wu, L. : Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, 2010.

D'Agostino, R. B. : Transformation to normality of the null distribution of G1, Biometrika, 57, 679–681, 1970.

D'agostino, Ralph B. and Belanger, Albert and D'Agostino Jr, Ralph B. : A suggestion for using powerful and informative tests of normality, The American Statistician, 44 (4), 316–321, 1990.

Kullback, S. : Information theory and statistics, Wiley, 395pp, 1959.

Pires, Carlos A and Talagrand, Olivier and Bocquet, Marc : Diagnosis and impacts of non-Gaussianity of innovations in data assimilation, Physica D : Nonlinear Phenomena, 239 (17),1701–1717, 2010.

Caumont O., Ducrocq V., Wattrelot E., Jaubert G., and PRADIER-VABRE S. : 1D+ 3DVar assimilation of radar reflectivity data : A proof of concept, Tellus A, 62 (2), 173—187, 2010

Wattrelot, E. and Caumont, O. and Mahfouf, J.-F. : Operational implementation of the 1D+ 3D-Var assimilation method of radar reflectivity data in the AROME model, Monthly Weather Review, 142 (5), 1852—1873, 2014

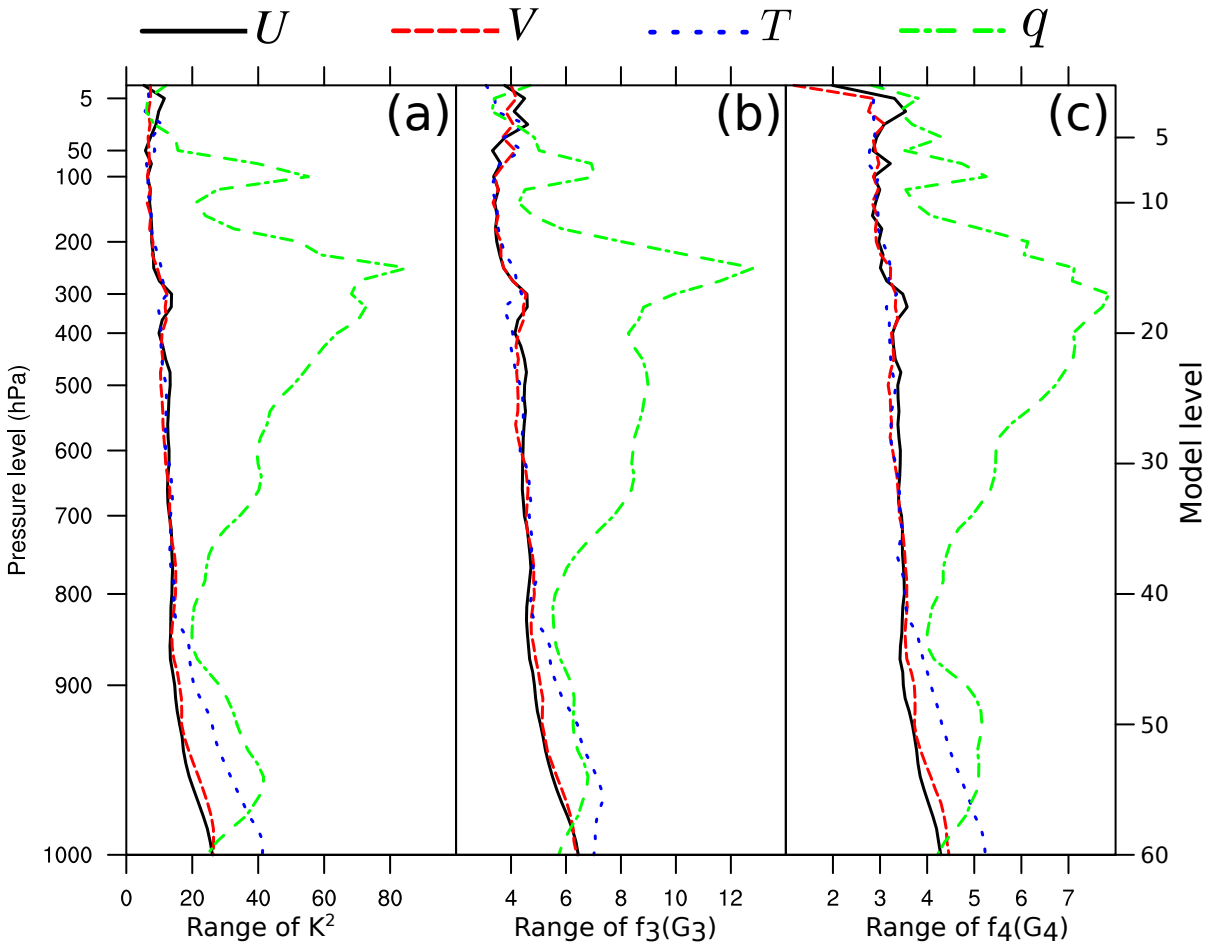

FIGURE 1 – Vertical profiles of ranges (see text) of **(a)** K^2, **(b)** transformed skewness $f_3(G_3)$, **(c)** transformed kurtosis $f_4(G_4)$. Profiles are computed from the 90-members ensemble of AROME-France 3h-forecasts valid at 03 :00 UTC the 4 November 2011. Profiles are given for four model variables : U, V, T, and q.

REFEREE 3 :
(Reviewer comments are written in black, and authors answers are in blue)

Comments from Referee→(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 non-linear 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

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_b) \tag{1}$$

where $\mathbf{x}_b$ and $\mathbf{x}_a$ are respectively the background and the analysis, $\mathbf{y}$ is the observation, $\mathbf{H}$ the corresponding (linear) observation operator, the difference $\mathbf{d} \equiv \mathbf{y} - \mathbf{H}\mathbf{x}_b$ being the innovation vector. $\mathbf{K}$ is the gain matrix which, in the context of least variance estimation, is defined as $\mathbf{K} \equiv \mathbf{C}_{zd}\mathbf{C}_{dd}^{-1}$, where $\mathbf{C}_{zd}$ is the cross-covariance matrix of the background error $\mathbf{z} \equiv \mathbf{x} - \mathbf{x}_b$ with the innovation, and is $\mathbf{C}_{dd}$ 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 $\mathbf{x}$ from $\mathbf{x}_b$ and $\mathbf{y}$, 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. Gaussianity 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 non-linearities 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 of 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.
Author's response→ There are two common derivations of the assimilation problem in the literature. The first one actually matches your derivation. It derives the Kalman gain as the best linear unbiased estimate, in the sense of a minimum variance estimate. Then it is possible to derive variational data assimilation as a minimization problem solving for this Kalman gain. The Gaussian assumption then is not necessary but ensure Bayesianity of the assimilation.

The second derivation takes maybe an other step : it starts from the derivation of a maximum likelihood problem using Bayes rules. This is the approach presented in (e.g.) Lorenc (1986), Bannister (2008), Bocquet et al. (2010), and Fisher et al. (2011). The Gaussian hypothesis is done next, leading to a minimization problem of a non-linear cost function $\mathcal{J}$ that is given for instance for a 3D-Var as :

$$2\mathcal{J} = (\mathbf{x} - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - \mathcal{H}(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y} - \mathcal{H}(\mathbf{x})) \tag{2}$$

with $\mathbf{B}$ and $\mathbf{R}$ the covariance matrix of background and observations errors. Non-linearities arise from the observation operators $\mathcal{H}$ (in addition to the direct model operator for the 4D-Var). As mentioned by Bocquet et al. (2010) about the minimization of $\mathcal{J}$, instead of using stochastic optimization methods which are intractable for NWP applications, a remedy is to use a succession of quadratic optimization problem with simplified and linearised operators. Fisher et al. (2011) explain the several linearisation as a way to resolve the minimisation problem with "a range of efficient methods". So, with this second derivation, the Gaussianity (or correction of Gaussianity) is seen as the only tractable choice and appears very soon in the $\mathcal{J}$ designing. Then, linearisation is seen as an additional technical assumption leading to better efficiency in the minimization process.

The equivalence between those two possible derivations may be obtain using some kind of EnKF approaches.
Author's changes in manuscript→In the abstract : "In numerical weather prediction, the problem of estimating initial conditions with a variational approach is usually based on a Bayesian framework associated with a Gaussianity assumption of the probability density functions of both observations and background errors. In practice, Gaussianity of errors is tied to linearity, in the sense that a nonlinear model will yield non-Gaussian probability density functions. In this context, standard methods relying on Gaussian assumption may perform poorly.", and at the beginning of the introduction : "In data assimilation, the analysis step may be seen as finding a maximum likelihood of the probability density functions (PDF) of the state x given the available observations y and a background state (usually a short range forecast). Usual Bayesian formulation yields (Kalnay, 2003)".

Comments from Referee→(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.
Author's response→ We don't think that there is a mistake here. According to (e.g.) Thode (2002) p45-46, G_3 and G_4 are using sample moments given as

$$m_k = \frac{1}{N_s} \sum_{i=1}^{N_s} (\mathbf{x}_i - \bar{\mathbf{x}})^k \tag{3}$$

Such definition of the second sample moment is used to compute G_3 and G_4, not the sample (unbiased) variance.
Author's changes in manuscript→We agree with this point.

Comments from Referee→More importantly, the fundamental purpose of the test is the following. For a 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 ?
Author's response→ According to Kendall and Stuart (1977) in case of exact normality G_3 is a zero mean random variable with a variance of

$$\sigma^2(G_3) = \frac{6(N_s-2)}{(N_s+1)(N_s+3)} \tag{4}$$

Asymptotically (N_s large enough), G_3 tends to be normally distributed with a zero mean and a variance of $6/N_s$. In our case for $N_s = 90$, G_3 is not exactly normally distributed and, with a bilateral testing at level 95%, the normality hypothesis is accepted when $-0.494 < G_3 < 0.494$ (Table B5, Thode, 2002).

As regards G_4, in case of normality G_4 is asymptotically (N_s large enough) normally distributed with a mean of 3 and a variance of $24/N_s$. For finite ensemble size, mean and variance are given as

$$E[G_4] = \frac{3(N_s-1)}{(N_s+1)} \tag{5}$$

$$\sigma^2(G_4) = \frac{24N_s(N_s-2)(N_s-3)}{(N_s+1)^2(N_s+3)(N_s+5)} \tag{6}$$

In our case for $N_s = 90$, G_4 is not exactly normally distributed and, with a bilateral testing at level 95%, the normality hypothesis is accepted when $2.24 < G_4 < 4.09$ (Table B6, Thode, 2002).

Those bounds are not mentioned in the text since, G_3 and G_4 are not used in this study. As justified below, only $f_3(G_3)$ and $f_4(G_4)$ are shown and analysed.
Author's changes in manuscript→None

Comments from Referee→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 in several of the formulæ leading to the definition of $f_4(G_4)$) ?
Author's response→ If the original variable is Gaussian, the normality of the transformed skewness and kurtosis is valid respectively for any values of $N_s > 8$ and $N_s > 20$ (resp. p48 and p52, Thode, 2002). Thus, the $N_s - 3$ coefficient that we found in P definition is actually coming from the variance of G_4, which is used in Q as a normalization coefficient.
Author's changes in manuscript→In section 2.1 : "For a Gaussian PDF and N_s higher than 20 (Thode, 2002), $f_3(G_3)$ and $f_4(G_4)$ could be both assumed to follow a Gaussian law with a zero mean and a unity variance.". Fig.1 has been changed to be in accordance with this threshold.

Comments from Referee→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.
Author's response→We see three main reasons of using $f_3(G_3)$ and $f_4(G_4)$ instead of G_3 and G_4. The first reason is that for $N_s = o(10^3)$, the asymptotic behaviour of G_3 and G_4 is not reached (D'Agostino, 1970; Anscombe and Glynn, 1983). So in order to simplify hypothesis testing, a transformation is needed to transform them as normal random variables. The second reason is

that their values weakly depend on N_s. This make possible to compare several studies using different ensemble sizes. The third reason is that $f_3(G_3)$ and $f_4(G_4)$ are both normally distributed. So the role of each of them in a possible deviation from Gaussianity could be compared, and they could be used to build an omnibus test of normality as K^2.
Author's changes in manuscript→None

Comments from Referee→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 ?
Author's response→You are right, K^2 is not distributed with an exact χ^2 since G_3 and G_4 are uncorrelated but not independent (p54, Thode, 2002). An other reason is that normality behaviour of G_3 and G_4 is only asymptotic. For those two reasons it is possible to correct critical values of the K^2 test (chapter *Moment* $(\sqrt{b_1}, b_2)$ *techniques*, D'Agostino and Stephens, 1986). For instance, with $N_s = 100$ the critical value is $K^2 = 6.271$ and not 5.991 as for an exact χ^2 distribution.
Author's changes in manuscript→None since it is already mentioned that K^2 is only "approximately" following a χ^2 distribution.

Comments from Referee→It is not clear how the values obtained for $f_3(G_3)$, $f_4(G_4)$ and $\mathbf{K}^2$ 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 $N_s = 100$?
Author's response→ When using hypothesis testing, conclusion of the test is always associated with a confidence level α (usually $1 - \alpha = 95\%$). Critical values X_c of the test are defined according to this level as (for a right-unilateral testing)

$$P(X > X_c) = \alpha \tag{7}$$

Instead of K^2, we could have shown binary result giving "this point is Gaussian or not". But since the critical value is depending on the confidence level of the test, results would have been different when using different confidence level. Moreover we want to see where the NG is the largest, and see structures. That's why we choose to show raw values of K^2.
Author's changes in manuscript→None

Comments from Referee→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.
Author's response→ As it is stated in the text, with unilateral testing at level 95%, the Gaussian hypothesis is rejected for $K^2 > 6.271$, so K^2 value around 4 are small enough to accept the Gaussian hypothesis of the sample tested. Moreover, discussion on K^2 values allow us to compare quantitatively Gaussian behaviour between variables.
Author's changes in manuscript→None

Comments from Referee→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) ?
Author's response→ Positive values of $f_4(G_4)$ means that distribution tails are heavier than Gaussian distribution, and also a bigger modal peak. Negative values of $f_4(G_4)$ show lighter tails and smaller modal peak.

Despite negative values appear in Fig.5, you are right noticing that $f_4(G_4)$ are in a large part positive. But it doesn't mean that $f_4(G_4)$ does not follow a standard Gaussian distribution. Indeed this conclusion needs the spatial ergodicity assumption which is not straightforward to us (since $f_3(G_3)$ and $f_4(G_4)$ distribution may depend on meteorological situation i.e spatial inhomogeneity). In this study we would simply test the normality of $f_4(G_4)$ with an hypothesis testing (see if $f_4(G_4)$ is larger than a critical value). An other way to test the normality of $f_4(G_4)$, would be to look at the distribution of an ensemble of $f_4(G_4)$ for each grid point. But this is costly since it needs an ensemble of ensembles.
Author's changes in manuscript→Added in section 2.1 : "While positive (negative) values of $f_3(G_3)$ point out distributions with a median smaller (higher) than the mean and with a longer right (left) tail, positive (negative) values of $f_4(G_4)$ mean that distribution tails are heavier (lighter) than Gaussian distribution's, with also a bigger (smaller) modal peak."

Comments from Referee→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).
Author's response→ We hope that previous answers are clarifying the use and interpretation of $f_3(G_3)$, $f_4(G_4)$, and K^2.
Author's changes in manuscript→None

Comments from Referee→(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.
Author's response→ To be consistent with the Fig.7, diagnostics of NG for vorticity ζ and total divergence η have been computed before and after the assimilation step of the 4^{th} November 2011 at 03 :00. Results are shown below in Fig.1 of this comment. While NG for levels higher than 900hPa are almost unchanged, the averaged K^2 of ζ and η is systematically lower for the analysis than the background state in the boundary layer. However the order of magnitude of the decrease is much smaller than for T and q, thus the dynamical variables ζ and η remain by far much more non-Gaussian than T and q.(Those conclusions have been added in the text).
Author's changes in manuscript→added in section 4.2.1 : "As for Fig.7, diagnostics of NG for vorticity ζ and total divergence η have been computed before and after the assimilation step (not shown). While NG of levels higher than 900hPa are almost unchanged, the averaged K^2 of ζ and η is systematically lower for the analysis than the background state in the boundary layer. However the order of magnitude of the decrease is much smaller than for T and q, and the dynamical variables ζ and η remain by far much more non-Gaussian.

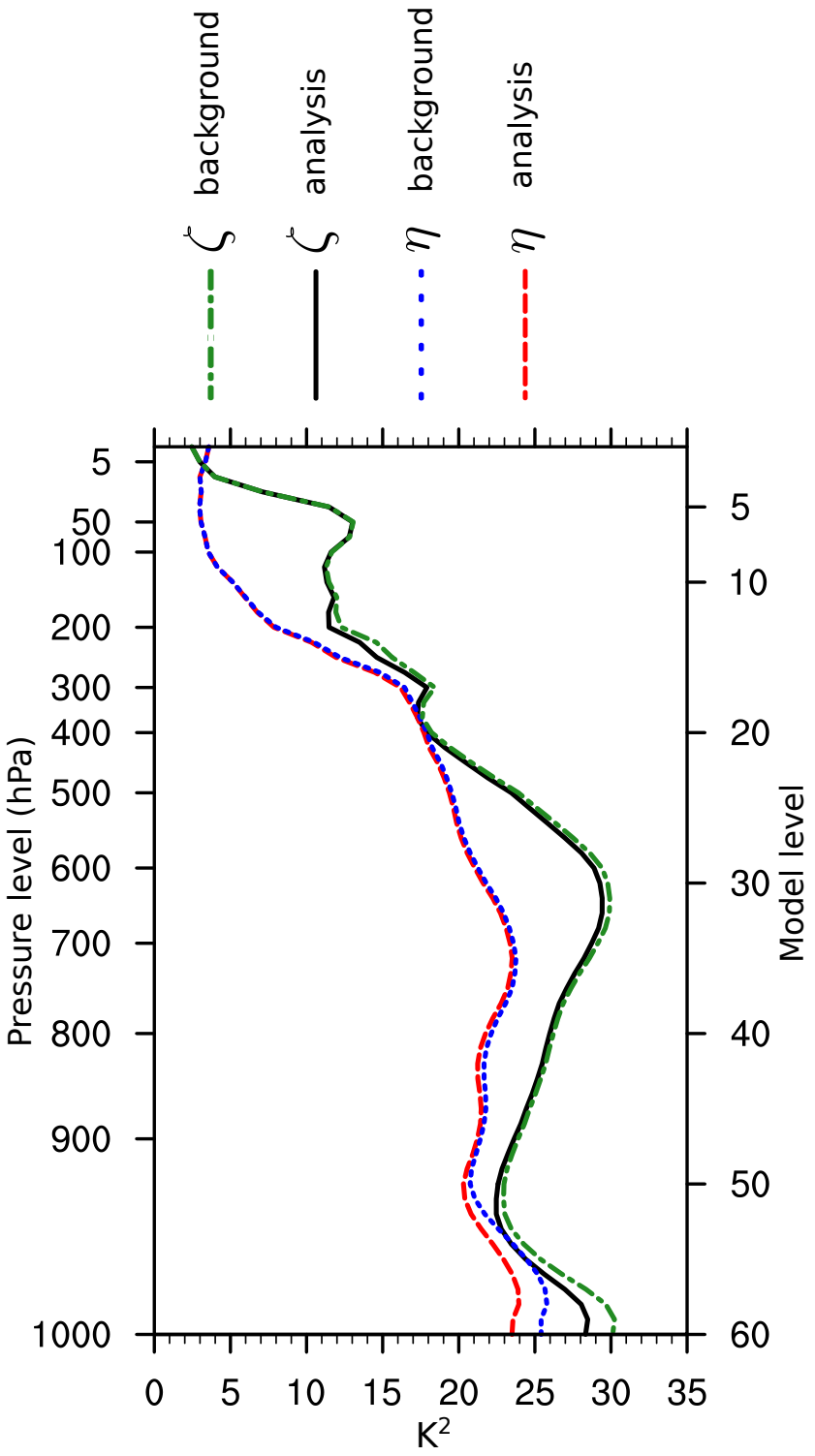

FIGURE 1 – Vertical profiles of Vorticity ζ and total Divergence η before ("background")and after ("analysis") the assimilation process. Results are computed from the ensemble of 90 background and analysis states valid the 4^{th} November 2011 at 03 :00.

Comments from Referee→(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.
Author's response→This remark is similar to that made by reviewer 1 (and 2) in his second point. In order to study relative impact on NG of heteroscedasticity and spatial derivatives, NG diagnostics have been computed for the temperature T, which is a nearly Gaussian field (cf Fig. 3a), for the temperature normalized by its standard deviation $\frac{T}{\sigma_T}$, and for their respective first-order spatial derivatives ($\frac{\partial T}{\partial x}$ and $\frac{\partial \frac{T}{\sigma_T}}{\partial x}$). Results are shown and explained in the answer to reviewer 1. A comment has also been added in the manuscript.
Author's changes in manuscript→Same as for reviewer 1.

Comments from Referee→(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 ?
Author's response→There was indeed an error in the caption. T was inverted with η_u. This has been corrected.
Author's changes in manuscript→Caption of Fig.9 corrected.

Comments from Referee→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.
Author's response→We agree with this point.
Author's changes in manuscript→ This has been corrected in the manuscript.

Comments from Referee→(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.
Author's response→ As noted in section 3.2, largest NG for q in high troposphere appear where q is almost non-existent. As it is stated in the text, those large values of NG have then to be taken with caution.
Author's changes in manuscript→None

Comments from Referee→(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)
Author's response→We agree with this point.
Author's changes in manuscript→This has been corrected in the manuscript as "dynamical control variables".

Comments from Referee→(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.
Author's response→We agree with this point.
Author's changes in manuscript→ This has been corrected in the manuscript.

Comments from Referee→(9) And there are erroneous statements concerning the relationship between skewness and tails pp. 1071, l. 13, and 1072, l. 1.
Author's response→We agree with this point.
Author's changes in manuscript→ This has been corrected in the manuscript.

Comments from Referee→(10) P. 1068, l. 11, what is unilateral testing ?
Author's response→ When testing an hypothesis (e.g.) H_0 : "$X_{observed}$ is sampled from a Gaussian random variable", with a confidence level α the test is right-tailed unilateral if H_0 is rejected when $P(X > X_{observed}) < \alpha$.
Author's changes in manuscript→ The "right-tailed" adjective has been added in the manuscript.

Comments from Referee→(11) P. 1073 and 1076, ll. 11 and 19, forecast ~~terms~~ ranges
Author's response→We agree with this point.
Author's changes in manuscript→ This has been changed

Comments from Referee→(12) P. 1064, l. 15, Laroche and Pierre, 1998. Do you mean Laroche and Gauthier ?
Author's response→We agree with this point.
Author's changes in manuscript→ This has been changed

Comments from Referee→(13) P. 1064, l. 20 (and elsewhere). The proper spelling is Järvinen (with a diaeresis)
Author's response→We agree with this point.
Author's changes in manuscript→ This has been changed

Comments from Referee→(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)
Author's response→We agree with this point.
Author's changes in manuscript→ This has been changed

Comments from Referee→(15) P. 1076, l. 3, *below* the tropopause
Author's response→We agree with this point.
Author's changes in manuscript→ This has been changed

Références

Anscombe, Francis J. and Glynn, William J. : Distribution of the kurtosis statistic b2 for normal samples, Biometrika, 70 (1), 227–234, 1983.

Bannister, Ross N. :A review of forecast error covariance statistics in atmospheric variational data assimilation. I : Characteristics and measurements of forecast error covariances, Quarterly Journal of the Royal Meteorological Society, 134 (637),1951–1970,2008.

Bocquet, M., Pires, C. A., and Wu, L. : Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, 2010.

D'Agostino, R. B. : Transformation to normality of the null distribution of G1, Biometrika, 57, 679–681, 1970.

D'Agostino, Ralph B. and Stephens, Michael A. : Goodness-of-fit techniques, Marcel Dekker, 1986.

Fisher, M. and Tremolet, Y. and Auvinen, H. and Tan, D. and Poli, P. :Weak-constraint and long-window 4D-Var, ECMWF Techn. Rep, 655, 2011.

Holm, E., Andersson, E., Beljaars, A., Lopez, P., Mahfouf, J.-F., Simmons, A., and Thépaut, J.-N. : Assimilation and modelling of the hydrological cycle : ECMWF's status and plans, European Centre for Medium-Range Weather Forecasts (ECMWF), Technical Memorandum, 2002.

Kalnay, E. : Atmospheric modeling, data assimilation, and predictability, Cambridge University Press, 2003.

Kendall, M. and Stuart, A. : The advanced theory of statistics. Vol. 1 : Distribution theory, 4th ed., 1977.

Lawson, W. G. and Hansen, J. A. : Alignment error models and ensemble-based data assimilation, Mon. Weather Rev., 133, 1687–1709, 2005.

Lorenc, A. C. : Analysis methods for numerical weather prediction, Q. J. Roy. Meteorol. Soc., 112, 1177–1194, 1986.

Ravela, S., Emanuel, K., and McLaughlin, D. : Data assimilation by field alignment, Physica D, 230, 127–145, 2007.

Thode, Henry C. :Testing for normality, volume 164, Marcel Dekker, CRC Press, 2002.

Trémolet, Yannick : Incremental 4D-Var convergence study, Tellus A, 59 (5), 706–718, 2007.

Manuscript prepared for Nonlin. Processes Geophys. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the LaTeX class copernicus.cls.
Date: 16 October 2015

Diagnosing non-Gaussianity of forecast and analysis errors in a convective scale model

R. Legrand, Y. Michel, and T. Montmerle

Centre National de Recherches Météorologiques (CNRM), Toulouse, France

Correspondence to: R. Legrand (raphael.legrand@meteo.fr)

Abstract

In numerical weather prediction, the problem of estimating initial conditions with a variational approach is usually based on a ~~Bayesian framework . Two common derivations respectively lead to the Kalman filter and to variational approaches. They rely on either assumptions of linearity or assumptions of Gaussianity~~ Bayesian framework associated with a Gaussianity assumption of the probability density functions of both ~~observation~~ observations and background errors. In practice, ~~linearity and~~ Gaussianity of errors ~~are tied to one another~~is tied to linearity, in the sense that a nonlinear model will yield non-Gaussian probability density functions~~, and that standard methods~~ . In this context, standard methods relying on Gaussian assumption may perform poorly~~in the context of non-Gaussian probability density functions~~.

This study aims to describe some aspects of non-Gaussianity of forecast and analysis errors in a convective scale model using a Monte-Carlo approach based on an ensemble of data assimilations. For this purpose, an ensemble of 90 members of cycled perturbed assimilations has been run over a highly precipitating case of interest. Non-Gaussianity is measured using the K^2-statistics from the D'Agostino test, which is related to the sum of the squares of univariate skewness and kurtosis.

Results confirm that specific humidity is the least Gaussian variable according to that measure, and also that non-Gaussianity is generally more pronounced in the boundary layer and in cloudy areas. The ~~mass~~ dynamical control variables used in our data assimilation, namely vorticity and divergence, also show distinct non-Gaussian ~~behavior~~behaviour. It is shown that while non-Gaussianity increases with forecast lead time, it is efficiently reduced by the data assimilation step especially in areas well covered by observations. Our findings may have implication for the choice of the control variables.

1 Introduction

In data assimilation, the analysis step ~~aims to find~~ may be seen as finding a maximum likelihood of the probability density functions (PDF) of the state x ~~from~~ given the available observations y and a ~~priori knowledge given by the PDF of the background (Kalnay, 2003) . The usual~~ background state (usually a short range forecast). Usual Bayesian formulation yields (Kalnay, 2003)

$$P_{\mathsf{a}}(x|y) \propto P_{\mathsf{o}}(y|x)P_{\mathsf{b}}(x), \tag{1}$$

where $P_{\mathsf{a}}, P_{\mathsf{b}}$ and P_{o} respectively are the PDFs of ~~the analysis, of the background (usually a short range forecast)and of the observations , which must be specified~~analysis, background errors (*a priori* PDF), and observations errors. For high dimensional systems, ~~the~~ to specify those PDFs as multivariate Gaussian is a natural choice for variables that may approximately verify the central limit theorem (Bocquet et al., 2010). Thus, up to now most of operational Numerical Weather Prediction (NWP) ~~has relied on data~~ centres have relied on variational assimilation schemes that are Gaussian or corrections to a Gaussian analysis-based strategy.

~~Of course in general the~~ The time integration of the model nonlinear dynamics ~~will lead~~ leads inevitably to non-Gaussian forecast errors (Bocquet et al., 2010) . For instance, the highly nonlinear processes involved in clouds and precipitation are known to give non-Gaussian background errors (Auligné et al., 2011). Some authors have reported on displacement errors of meteorological features that turn into non-Gaussian background errors (Lawson and Hansen, 2005). Keeping the Gaussian formalism in this case may yield unrealistic analyses that are distorted (Ravela et al., 2007).

In NWP, the analysis of humidity may be the most problematic with respect to non-Gaussianity (NG). This is due to the condensation effects near saturation and the intrinsic positivity of humidity. The choice of the control variable for humidity is a long-standing debate (Dee and da Silva, 2003). Specific humidity exhibits NG but is rather weakly correlated (in average) to other variables. Relative humidity has been found to be more

Gaussian but has stronger cross-covariances with temperature that are state-dependent and difficult to model. It still has skewed distribution near condensation or in dry conditions. The solution adopted in several operational ~~centers~~ centres is to use a normalized relative humidity variable. The normalization factor is the standard deviation of the relative humidity error, stratified according to the ~~analyzed~~ analysed relative humidity itself. The asymmetries in PDFs are also accounted for through a nonlinear transformation. This scheme has been implemented through several variants both in global (Holm et al., 2002; Ingleby et al., 2013) and in limited area models (Gustafsson et al., 2011).

The 4D-Var algorithm commonly used in NWP (e.g. Rabier et al., 2000) has some ability to handle nonlinearities. It solves for the most probable state in Eq. (1) by minimizing a non-quadratic cost function with nonlinearities in the model and in the observation operator mapping the model state to the observation space. The approach, known in the community as incremental 4D-Var (Courtier et al., 1994), is based on a form of truncated Gauss–Newton iterations. The problem is solved by minimizing a succession of inner-loop quadratic optimization problems with increasing horizontal resolutions, in which the model is simplified and linearised around the state adjusted by the previous outer-loop iteration (Laroche and Gauthier, 1998).

The PDF of observation errors is also non-Gaussian in general. In NWP, quality-control are performed to exclude observations that are outliers compared to the model and using statistical knowledge (Lorenc, 1986). Unfortunately, this can be erroneous and a more flexible framework has been introduced for instance by Anderson and Järvinen (1999). It explicitly computes the probability of gross error for each observation, given the preliminary analysis from the outer loops. The weight of each observation is smoothly decreased with increased likelihood for gross error. More recently, this scheme has been replaced by the use of a Huber norm (Tavolato and Isaksen, 2014). The NG of observation errors is out of the scope of this paper.

The main goal of this paper is to document the non-Gaussianities of background and of analysis errors in the context of convective scale NWP. For this purpose, a large

ensemble of perturbed cycled assimilations has been set up with the AROME-France[1] model. The perturbations simulate the evolution of the true background and analysis errors (Houtekamer et al., 1996; Fisher, 2003; Berre et al., 2006). The diagnosis of NG may help to find out for which variables and/or in which areas efforts could be made to improve Gaussian assumptions in the assimilation algorithm, or to help designing advanced data assimilation schemes taking into account displacement errors for instance (Ravela et al., 2007).

The paper is organized as follows: Sect. 2 presents the univariate D'Agostino test for NG ~~(D'Agostino, 1970)~~ (D'agostino et al., 1990) and evaluates its efficiency on some specified PDFs. Section 3 describes the ensemble from which the NG is diagnosed. This ensemble is composed of assimilations and forecasts performed by the AROME-France model for a highly precipitating event over the Mediterranean sea, of interest for the HyMeX campaign (Ducrocq et al., 2013). Results of the NG diagnostics are then documented. After an overview for model prognostic variables, time evolution of NG is discussed. The dependence of NG to physical nonlinear processes is then described by making use of geographical masks based on cloud contents. In Sect. 4, the impact of the data assimilation process on NG is studied by comparing diagnostics performed on both background and analysis errors, and by computing diagnostics in the control space of the minimization. Conclusions are given in Sect. 5.

2 An index of non-Gaussianity

In NWP, dimensions of the state and observation vectors, including satellite and radar, are huge (respectively around 10^8 and 10^5 in AROME-France). As mentioned in Bocquet et al. (2010) only the simpler statistical tests of Gaussianity are tractable for such high dimensional problems. Therefore, we will rely on simple univariate tests for NG.

[1] Application de la Recherche à l'Opérationnel à Méso-Echelle (Seity et al., 2011).

2.1 D'Agostino test

The D'Agostino test ~~(hereafter K^2 test, D'Agostino, 1970)~~ (hereafter K^2 test, D'agostino et al., 19 a statistical test where the deviation from Gaussianity is detected from the PDF's skewness and kurtosis. The skewness is a measure of the asymmetry of the PDF about its mean. Positive (negative) values are associated with a ~~mode~~ median of the PDF smaller (larger) than its mean and with a large right (left)-tail. For instance, a negative skewness for specific humidity at some point indicates that ~~at least a part~~ more than the half of the ensemble is ~~much dryer~~ more humid than the mean value of the ensemble. The kurtosis measures the peakedness of the distribution (Thode, 2002). A PDF with larger tails and a narrow modal peak has a large kurtosis.

The theoretical skewness and kurtosis are respectively estimated over an ensemble by the sample third (G_3) and fourth (G_4) standardized moments. They are defined given a sample $x_{i=1..N_s}$ of size N_s and its sample mean $\overline{x}$ as

$$G_3 = \frac{m_3}{m_2^{\frac{3}{2}}} = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s}(x_i-\overline{x})^3}{\left[\frac{1}{N_s}\sum_{i=1}^{N_s}(x_i-\overline{x})^2\right]^{\frac{3}{2}}} \quad (2)$$

$$G_4 = \frac{m_4}{m_2^2} = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s}(x_i-\overline{x})^4}{\left[\frac{1}{N_s}\sum_{i=1}^{N_s}(x_i-\overline{x})^2\right]^2} \quad (3)$$

with m_2, m_3, and m_4 the sample second~~(variance)~~, third, and fourth order moments. These quantities estimate the theoretical skewness and kurtosis of the distribution. For a Gaussian PDF, skewness is zero and kurtosis equals 3. Thus, the sample skewness and kurtosis defined above could be used to detect deviation from Gaussianity, yet their convergence to normality with ensemble size is slow. As reported in Tables 3.1 and 3.2 of Thode (2002), the normality is reached with sufficient accuracy typically for ensemble sizes of the order of ~ 5000. For smaller ensemble sizes (more suitable to NWP), it has been suggested to transform these quantities into $f_3(G_3)$ and $f_4(G_4)$ respectively, in order to remedy this situation ~~(D'Agostino, 1970)~~ (D'Agostino, 1970; Anscombe and Glynn, 1983) . f_3 is defined

as

$$
\begin{aligned}
A &= G_3 \times \sqrt{\frac{(N_s+1)(N_s+3)}{6(N_s-2)}} \\
B &= 3\frac{\left(N_s^2+27N_s-70\right)(N_s+1)(N_s+3)}{(N_s-2)(N_s+5)(N_s+7)(N_s+9)} \\
C &= \sqrt{2(B-1)}-1 \\
D &= \sqrt{C} \\
E &= \frac{1}{\sqrt{\ln(D)}} \\
F &= \frac{A}{\sqrt{\frac{2}{C-1}}} \\
f_3(G_3) &= E \times \ln(F+\sqrt{F^2+1})
\end{aligned}
$$

and f_4 is defined as

$$
\begin{aligned}
O &= G_4 \times \frac{N_s(N_s+1)}{(N_s-1)(N_s-2)(N_s-3)} - 3\frac{(N_s-1)}{(N_s+1)} \\
P &= \frac{24N_s(N_s-2)(N_s-3)}{(N_s+1)^2(N_s+3)(N_s+5)} \\
Q &= \frac{(N_s-2)(N_s-3)}{(N_s+1)(N_s-1)\sqrt{P}} \times O \\
R &= \frac{6(N_s^2-5N_s+2)}{(N_s+7)(N_s+9)}\sqrt{\frac{6(N_s+3)(N_s+5)}{N_s(N_s-2)(N_s-3)}} \\
S &= 6+\frac{8}{R}\left[\frac{2}{R}+\sqrt{1+\frac{4}{R^2}}\right] \\
T &= \frac{1-\frac{2}{S}}{1+Q\sqrt{\frac{2}{S-4}}}
\end{aligned}
$$

$$
f_4(G_4) = \frac{1-\frac{2}{9S}-T^{\frac{1}{3}}}{\sqrt{\frac{2}{9S}}}.
$$

~~The~~ While positive (negative) values of $f_3(G_3)$ point out distributions with a median smaller (higher) than the mean and with a longer right (left) tail, positive (negative) values of $f_4(G_4)$ mean that distribution tails are heavier (lighter) than Gaussian distribution's, with also a bigger (smaller) modal peak.

$f_3(G_3)$ and $f_4(G_4)$ statistics are then combined to produce an omnibus test K^2, able to detect deviations from normality due to either skewness or kurtosis:

$$
K^2 = f_3^2(G_3) + f_4^2(G_4). \tag{4}
$$

~~For a Gaussian PDF,~~ When testing a Gaussian distribution, asymptotic values for the three criteria ($f_3(G_3)$, $f_4(G_4)$, and K^2) are respectively $f_3(G_3) = 0$, $f_4(G_4) = 0$, and $K^2 = 2$. Using finite sampling with ensemble size $N_s > 20$ (Thode, 2002), $f_3(G_3)$ and $f_4(G_4)$ ~~both~~ could be both assumed to follow a Gaussian law with a zero mean and a unity variance. In this case, K^2 follows approximately a χ^2 distribution with two degrees of freedom. ~~This property may be used to evaluate the null hypothesis H_0: "the sample is from a Gaussian PDF". Using unilateral testing~~Confidence intervals at 95% are then given by $f_3(G_3) \in [-1.96; 1.96]$, ~~$K^2 > 6.271$ (for $N_s = 100$) indicates a reject of Gaussianity at the~~ $f_4(G_4) \in [-1.96; 1.96]$, and $K^2 \in [0; 5.991]$. Because G_3 and G_4 are uncorrelated but not independent, K^2 does not follow an exact χ^2 distribution, and confidence interval is slightly different. Using a right-tailed unilateral testing at 95~~confidence level~~% for $N_s = 100$, the critical value of K^2 is 6.271 instead of 5.991.

2.2 Evaluation

The efficiency of the K^2 test can be evaluated by measuring its probability of detection (POD) for the hypothesis H_0. For a sample known to be from a non-Gaussian PDF, the POD gives the probability that the test accurately rejects H_0. The best result is POD = 1.

POD of K^2 test is estimated from N_{xp} independent experiments. For each experiment, K^2 is computed from N_s elements sampled from a known distribution. Depending on the K^2 value, H_0 is accepted or rejected. When the known distribution is non-Gaussian, POD is given by the frequency of H_0 rejections over the N_{xp} experiments.

The POD is estimated for three non-Gaussian distributions: uniform, log-normal, and a Gaussian mixture. The Gaussian mixture is defined through its PDF as $P(x) = w_1 P_1(x) + w_2 P_2(x) + w_3 P_3(x)$ with P_1, P_2, and P_3, three Gaussian distributions with zero mean and respectively 0.1, 0.05, 0.02 as chosen standard deviation. The chosen weights are given by $(w_1, w_2, w_3) = (0.2, 0.5, 0.3)$. The representation of the shapes of these three distributions is given in Fig. 1a, alongside the Gaussian distribution.

POD are estimated over $N_{xp} = 10^5$ experiments. For both tests, different ensemble sizes N_s are tested ($N_s =$ ~~10,~~ 20, 30, 40, 50, 60, 70, 80, 90, 100, 200). Results of this ideal

case are shown in Fig. 1b. The log-normal distribution is the easiest one to discriminate from the Gaussian distribution, yielding the highest POD that reach almost one as soon as the ensemble size is above forty. For the two others, non-Gaussian distributions (uniform and Gaussian-mixture) K^2 test is only correctly discriminating from Gaussianity (with POD > 0.8) when $N_s > 70$. For $N_s = 90$, which corresponds to the ensemble size for the real dataset composed of AROME-France forecasts (see Sect. 3), POD values are over 0.9 for all three non-Gaussian distributions. In conclusion, the K^2 test is able to correctly discriminate NG for the ensemble size considered in this paper.

~~There are some~~ A review of other well-established tests for ~~univariate Gaussianity~~ Gaussianity are presented in Bocquet et al. (2010) , such as the measure of entropy (kullback (1959) , used in geophysics by Pires et al. (2010)), or the univariate Anderson–Darling goodness-of-fit test (Anderson and Darling, 1954). The latter has been also tested in the same framework and the performances proved to be very similar to the ones of the K^2 test. When comparing the results, obtained over the ensemble (Sect. 3), these two tests also give very similar results. e.g. they indicate the same areas of NG over $\approx 90\,\%$ of the domain. But, measuring skewness and kurtosis may be more informative and may be of interest for some assimilation schemes that account for skewness (Hodyss, 2012). Also, describing the values of K^2 has the advantage to prevent the results from depending on the chosen confidence level.

3 Diagnosis of the non-Gaussianity of AROME forecast errors

3.1 An AROME-France ensemble for a high-precipitating case

AROME-France is an operational non-hydrostatic model covering France with a 2.5 km horizontal resolution at the time of the experiments. Its lateral boundary conditions are given by the global model ARPEGE[2]. Assimilation steps are done every three hours with a 3D-

[2]Action de Recherche Petite Échelle Grande Échelle (Pailleux et al., 2000).

Var scheme and make use of a comprehensive set of observations such as conventional, satellite or Doppler radar data (see Seity et al. (2011) for more details).

The simulation of background and analysis errors is achieved by using a Monte-Carlo sampling, called an Ensemble Data Assimilation (EDA) in the context of NWP. A 90-members EDA is first run for the global model (AEARP, Berre and Desroziers, 2010). Each EDA member is based on a 4D-Var cycled assimilation which uses perturbed observations and a perturbed background, in order to simulate the error evolution (Berre et al., 2006). Observation perturbations are constructed as random draws of the specified observation error covariance matrix, and background perturbations result from the forecast evolution of previous analysis perturbations and from their inflation at the end of each forecast (Raynaud et al., 2012). This global ensemble provides perturbed boundary conditions to an ensemble of perturbed 3D-Vars for AROME-France, as described in Ménétrier et al. (2014). True background errors are then approximated by the deviations of the perturbed backgrounds from the ensemble mean. A few cycles (typically four) are necessary to reach a regime where the spread of the ensemble is representative of the true error spread; these cycles are discarded from the diagnostics presented below.

The case of interest is the 4 November 2011 between 00:00 and 06:00 UTC. A strong Southerly convergent flow occurs at low levels over Southern France (Fig. 2). Warm and moist air from the Mediterranean sea is advected over land, which triggers deep convection. Those high intensity events are studied by the HyMeX research program (Ducrocq et al., 2014). Associated precipitations are visible all along the Rhone valley, with local maxima exceeding $25\,\mathrm{mm\,h^{-1}}$. Also, associated with a low pressure area over the North-East Atlantic (not shown), a cold active front extending from the bay of Biscay to the eastern Britannic coast, is sweeping North-West of France with locally strong precipitations.

3.2 Vertical profiles of NG

The vertical profiles of quantities related to NG are shown in Fig. 3 for different variables, namely zonal (U) and meridian (V) winds, temperature (T) and specific humidity (q). On average, except near the surface, q is the variable that shows the largest deviation from

Gaussianity, confirming results obtained at the global scale (Holm et al., 2002). From 850 to 350 hPa, q is indeed characterized with an increase of the deviation from Gaussianity. As shown in Fig. 3b, this NG is partly explained by negative values of the skewness, highlighting a ~~right-tailed~~ left-tailed PDF of the background errors, meaning that many values are more humid than the ensemble mean.

In the troposphere, ~~the increase of~~ K^2 ~~seems inversely proportional to~~ is increasing while the q mean content, displayed in Fig. 3d, is largely decreasing. Values at higher levels, where q is almost nonexistent, may however be taken with caution. Below 850 hPa, K^2 is peaking around 960 hPa. Above 850 hPa, the wind components and T remain close to Gaussianity. Below however, all variables have significant deviation from Gaussianity, especially T for which high values of K^2 are found at ground level, making of it the less Gaussian variable in the boundary layer.

3.3 Horizontal structures of NG

The ~~horizontal~~ range, defined as the difference between the 95^{th} and the 5^{th} percentiles, could be used to describe roughly the horizontal spatial variability for each vertical level. Vertical profiles of ranges of K^2, $f_3(G_3)$, and $f_4(G_4)$ (not shown) have, all three, large similarities between each other, and with the shapes of K^2 profiles displayed in Fig. 3a. It includes in particular two maxima in the boundary layer and in high troposphere for q and larger values towards the surface for T. Ranges are much larger for the four variables (approximately four times as large) than the respective mean values given in Fig. 3, implying a large spatial variability for the three NG diagnostics. An example of the horizontal structures of NG ~~are shown~~ is given for q in the boundary layer by Fig. 4. They have large similarities with the meteorological coherent structures, as the Southerly convergent flow over South of France and the active cold front aloft North-West of France are associated with high values of K^2.

Supporting the conclusion drawn from Fig. 3, transformed skewness $f_3(G3)$ is mainly negative (corresponding to ~~right-tailed~~ left-tailed distributions) over the domain and

has a larger contribution than transformed kurtosis $f_4(G4)$ in large K^2 values. Over Mediterranean sea, the skewness represents on average 70 % of K^2.

It may be interesting to compare NG with the variance of the ensemble, as K^2 is defined from standard third and fourth standardized moment avoiding any scale effects. As displayed in Fig. 5, the variance does not coincide with overall NG, even if it happens that Gaussian areas may coincide with regions of low variance. However, the area diagnosed with the highest values of variance, located South of the Balearic islands, is associated with low values of K^2.

NG of the surface pressure is not shown in this study since, according to our diagnostics, it is a mainly Gaussian variable (averaged K^2 around 2.7). High values of K^2 appears around the cold front and the convergence area but they are very localized and of smaller amplitude compared to the other model variables.

3.4 Time evolution of non-Gaussianity

For each member of the ensemble, 18 h-forecasts have been run from the analyses performed at 00:00 UTC, the 04 November 2011. This allows to diagnose NG every 6 h during the first 18 h of integration. The corresponding vertical profiles are shown in Fig. 6 for the two most non-Gaussian variables according to Sect. 3.2: q and T.

In order to get insights into the processes that may be involved in NG development, the diagnostics have been separately computed for cloudy and for clear sky areas, following a similar approach to that of Montmerle and Berre (2010) and Michel et al. (2011), in which precipitating masks have been used. Grid points over the domain are separated in two bins: "cloudy" or "clear sky" points. "Cloudy" bin defines grid points whose vertically integrated simulated cloud water exceeds $0.1\,\mathrm{g\,kg^{-1}}$ for a majority of ensemble members (i.e. more than 45 members for the 90-members ensemble). The other points are classified as "clear sky". The percentage of "clear sky" points being three to five times larger (not shown) than the detected "cloudy points", similarities between "clear sky" profiles, and profiles averaged over the whole domain (as plotted in Fig. 3) are apparent.

During the 6 first hours of forecasts, NG quickly increases. For q, all tropospheric model levels are affected. For T, starting from a fairly Gaussian profile, increase of NG is mainly affecting the boundary layer and higher levels remain close to Gaussianity. During the following 12 h (from 6 to 18 h-forecast), changes of NG are smaller for both variables. Those results support that NG in the background may rather come from non-linear processes acting on nearly Gaussian pdfs instead of linear processes acting non-Gaussian pdfs.

It is interesting to notice that different ~~behaviors~~ behaviours can be found for diagnostics computed over "cloudy" and "clear sky" areas. For q, NG is mainly found in "cloudy" areas, where K^2 quickly reaches values above 8, with two peaks around 900 and 700 hPa. The altitude of the lower peak rises with forecast ~~terms~~ranges, while the amplitude of the higher one increases. According to Fig. 6c that displays the time evolution of the mean cloud contents, this evolution of NG in cloudy areas is likely due to nonlinear processes such as the vertical displacement error of cloud base and top within the ensemble and possibly the diabatic processes. In surface layers, K^2 for T quickly increases especially for clear air areas where turbulent and radiative processes occur. After 12 h, NG is more spread vertically within clouds, probably because of diabatic processes. For T and q, diabatic processes are good candidates to produce NG because of intrinsic thresholds in cloud physics (e.g. moisture saturation) and non-linear processes like turbulence on cloud-top.

For the wind components, ~~behaviors~~ behaviours close to T have been found, but with smaller amplitude (not shown): NG increases mainly in the boundary layer in "clear sky" areas and may be due to nonlinear turbulent processes.

4 Non-Gaussianity in the data assimilation process

Based on comparisons of NG diagnostics between successive background and analysis errors, this section focuses on the evolution of NG through cycled 3D-Var assimilations. Analysis errors will be treated for both model and control variables. The link between assimilated observations and NG reduction will be shown.

4.1 Overview

An overview of the NG evolution during the analysis process is given in Fig. 7 that shows averaged K^2 profiles for the analysis and the background errors computed for two consecutive assimilation/3 h-forecast steps. Comparable results are found for the two cycles, confirming the increase of NG during the model integration, and highlighting the substantial reduction of NG during the assimilation process, especially for levels where NG grows quickly. Values of K^2 are indeed brought back to much more Gaussian values, even in the lower levels for both q and T, and in higher troposphere for q.

Geographical variations of NG are illustrated in Fig. 8. As in Fig. 7, the NG of the background and of the following 3 h-forecast are similar. The largest decreases of NG between background and analysis error match areas with a large analysis increment, in particular where radar data are assimilated (Fig. 8d). The analysis increment being a linear function of the innovation vector in model space (observation minus background), its Gaussianity is insured by a rough selection applied beforehand to the observations, allowing to remove outliers (e.g. for radar data, Caumont et al., 2010; Wattrelot et al., 2014) . Some NG areas remain though, especially in areas where the background is less constrained by observations (e.g. above Spain and above the sea). However, most areas where NG has been reduced thanks to the data assimilation process recover their NG nature after 3 h of model integration.

4.2 Non-Gaussianity in control space

Previous results are documenting the NG of four model prognostic variables: U, V, T and q. As it is detailed in Brousseau et al. (2011), the assimilation scheme in AROME-France is based on a 3D-Var whose control variables are the vorticity $\xi\zeta$, the unbalanced divergence η_{u}, the unbalanced temperature and surface pressure $(T, P_{\mathrm{s}})_{\mathrm{u}}$, and the unbalanced specific humidity q_{u}. These control variables are linked to the model variables following the multivariate formalism of Berre (2000), which is based on the decomposition of the background error covariance matrix in spatial operators and balance transforms. Since the

minimization is performed in the control space, NG diagnostics have also been computed for these control variables.

4.2.1 Overview

Vertical profiles of NG for control variables are presented in Fig. 9. Unlike the zonal and meridian winds, ~~ξ~~ ζ and η_u are strongly non-Gaussian over the whole troposphere, whereas T_u and q_u display much more Gaussian profiles.

Negative values of $f_3(G_3)$ below 800 hPa for η_u (Fig. 9b) denote a larger spread of the distribution below the mean, probably due to the occurrence of low level convergence. At mid-troposphere, error distributions of all four variables are near symmetric. Except for q_u, distributions in tropospheric levels remain symmetric and the K^2 index is mainly explained by the kurtosis (Fig. 9c).

Those results agree with one of the conclusion of Ménétrier et al. (2015). These authors describe and algorithm to find the optimal truncation dedicated to sample covariances filtering. This algorithm has two variants. The first one assumes Gaussian PDF for the background perturbations while the second one does not. Their study indicate that, at convective scale, the Gaussian variant is accurate for T_u and q_u, but the more general non-Gaussian variant has to be used for ~~ξ~~ ζ and η_u, which are significantly non-Gaussian variables in agreement with our study. To go further on this topic, NG diagnostics have been computed for the spatial first-order derivative of T. While T is a nearly Gaussian variable (see Fig.3a), its spatial derivation largely increases the NG (not shown), up to the order of magnitude found in Fig.9a and Fig.10 for ζ and η. This supports the attribution to derivation for at least a part of the NG displayed for the dynamical control variables.

While very similar, the horizontal structures of K^2 for ~~ξ~~ ζ and η_u are noisier compared to the other variables, with very small scale and intense signals (not shown). Maps mostly follow the land–sea mask, with high values of K^2 over sea, and low values over land. Aloft, NG follows meteorological active structures (cold front and Cevenol event).

As for Fig.7, diagnostics of NG for vorticity ζ and total divergence η have been computed before and after the assimilation step (not shown). While NG of levels higher than 900hPa

are almost unchanged, the averaged K^2 of ζ and η is systematically lower for the analysis than the background state in the boundary layer. However the order of magnitude of the decrease is much smaller than for T and q, and the dynamical variables ζ and η remain by far much more non-Gaussian.

4.2.2 Non-Gaussianity in the multivariate transform

To go further in the discussion on Gaussianity of the control variables, this section is comparing the K^2 values for total and unbalanced variables.

According to Fig. 10, the debalancing process is not really affecting the NG for the divergence, except at lower levels where K^2 is slightly decreasing while keeping large values. K^2 values remain two to three times larger for the divergence (total or unbalanced) than for T and q from the surface to the mid-troposphere. On the contrary, NG decreases significantly for T and q during the debalancing process. Changes mainly appear in boundary layer for T. For q, changes appear for every model levels especially in the boundary layer and ~~under~~ below the tropopause. From surface to 750 hPa, NG of q_u is equal or smaller than the NG of T_u.

5 Conclusions

It is suggested to use the K^2 value from the D'Agostino test for diagnosing non-Gaussianity. This diagnostic is computed from the univariate sample skewness and kurtosis from an ensemble. This may allow to describe deviations from the Gaussian hypothesis for AROME-France background and analysis errors, as illustrated here on a case study characterized by a Cevenol event and an active cold front with a 90-members ensemble.

According to our diagnostic, among model variables, q has the largest deviation from Gaussianity, with a maximum of amplitude near the tropopause and in the boundary layer. Deviation from Gaussianity for U, V, and T only appears in the boundary layer. With an heterogeneous diagnostic, NG has been separately diagnosed for "cloudy" points and

"clear sky" points. For q, cloud covering leads to higher NG, especially at the bottom and at the top of the cloud layer. In "clear sky" situations, surface processes are expected to enlarge K^2 for T, in a larger manner than for "cloudy" points. Studying time evolution through forecast ~~terms~~ranges, NG is mainly increasing during the 6 first hours. The 3D-Var assimilation appears to efficiently reduce the growing NG of the forecast, especially in well-observed areas. Finally, among control variables of the assimilation, ~~ξ~~ ζ and η_u deviate from Gaussianity in a larger manner than T_u and q_u, which are much more Gaussian than their balanced counterparts.

Despite this work is attributing non-Gaussian ~~behaviors~~ behaviours to well-known nonlinear processes, such as the microphysical or boundary layer processes, it is not precisely addressing the cause of NG. However two important questions on variational data assimilation are highlighted. First, regarding control variables of the assimilation, according to our diagnostic, the most non-Gaussian variables are the vorticity and the divergence. Yet, main efforts have been put on "Gaussianisation" of specific humidity (e.g. Holm et al., 2002) but the discussion may also be focused on vorticity and divergence, either with a "Gaussianisation" of those variables or with a discussion on the possibility to use other dynamical variables. Second, with the cloud mask approach, cloud layers have been associated with high values of NG.

This study uses an ensemble at convective scale that does not include model error ~~neither~~ either in the analysis ~~nor~~ or in the forecast steps. It is possible that conclusions would be different if stochastic noise drawn explicitly from a Gaussian is added to the model states during the forecasts, as stated by Lawson and Hansen (2004). Also, this study is actually a part of a work focused on the correction of displacement errors. Since displacement errors are identified to cause NG (Lawson and Hansen, 2005), diagnostics of NG may be used to evaluate improvements in the current amplitude error correction step (3D-Var) brought by a displacement error correction (Ravela et al., 2007). This will be examined in a future work.

Acknowledgements. This work has been supported by the French Agence Nationale de la Recherche (ANR) via the IODA-MED Grant ANR-11-BS56-0005 and by the MISTRALS/HyMeX

program. The authors thank Benjamin Ménétrier, Gérald Desroziers, and Loïk Berre for their scientific advice and their careful readings that proved very useful to improve the manuscript.

References

Anderson, E. and ~~Jarvinen~~Järvinen, H.: Variational quality control, Q. J. Roy. Meteorol. Soc., 125, 697–722, doi:10.1002/qj.49712555416, 1999.

Anderson, T. W. and Darling, D. A.: A test of goodness of fit, J. Am. Stat. Assoc., 49, 765–769, 1954.

Anscombe, Francis J. and Glynn, William J.: Distribution of the kurtosis statistic b2 for normal samples, Biometrika, 70 (1), 227–234, 1983.

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, 2011.

Berre, L.: Estimation of synoptic and mesoscale forecast error covariances in a limited-area model, Mon. Weather Rev., 128, 644–667, 2000.

Berre, L. and Desroziers, G.: Filtering of background error variances and correlations by local spatial averaging: a review, Mon. Weather Rev., 138, 3693–3720, 2010.

Berre, L., Ecaterina Ştefănescu, S., and Belo Pereira, M.: The representation of the analysis effect in three error simulation techniques, Tellus A, 58, 196–209, 2006.

Bocquet, M., Pires, C. A., and Wu, L.: Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev., 138, 2997–3023, 2010.

Brousseau, P., Berre, L., Bouttier, F., and Desroziers, G.: Background-error covariances for a convective-scale data-assimilation system: AROME–France 3D-Var, Q. J. Roy. Meteorol. Soc., 137, 409–422, 2011.

Caumont O., Ducrocq V., Wattrelot E., Jaubert G., and PRADIER-VABRE S.: 1D+ 3DVar assimilation of radar reflectivity data: A proof of concept, Tellus A, 62 (2), 173–187, 2010

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteorol. Soc., 120, 1367–1387, 1994.

D'Agostino, R. B.: Transformation to normality of the null distribution of G1, Biometrika, 57, 679–681, 1970.

D'agostino, Ralph B. and Belanger, Albert and D'Agostino Jr, Ralph B.: A suggestion for using powerful and informative tests of normality, The American Statistician, 44 (4), 316–321, 1990.

Dee, D. P. and da Silva, A. M.: The choice of variable for atmospheric moisture analysis, Mon. Weather Rev., 131, 155–171, 2003.

Ducrocq, V., Belamari, S., Boudevillain, B., Bousquet, O., Cocquerez, P., Doerenbecher, A., Drobinski, P., Flamant, C., Labatut, L., Lambert, D.,
Nuret, M., Richard, E., Roussot, O., Testor, P., Arbogast, P., Ayral, P.-A., Van Baelen, J., Basdevant, C., Boichard, J.-L., Bourras, D., Bouvier, C., Bouin, M.-N., Bock, O., Braud,I., Champollion, C., Coppola, L., Coquillat, S., Defer, E., Delanoë, J., Delrieu, G., Didon-Lescot, J.-F., Durand, P., Estournel, C., Fourrié, N., Garrouste, O., Giordani, H., Le Coz, J., Michel, Y., Nuissier, O., Roberts, G., Saïd, F., Schwarzenboeck, A., Sellegri, K., Taupier-Letage, I., Vandervaere J.-P.: HyMeX, les campagnes de mesures: focus sur les événements extrêmes en Méditerranée, Société météorologique de France, Paris, France, La Météorologie, 80, 37–47, 2013.

Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansá, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayral, P.-A., Belamari, S., Berne, A., Borga, M., Boudevillain, B., Bock, O., Boichard, J.-L., Bouin, M.-N., Bousquet, O., Bouvier, C., Chiggiato, J., Cimini, D., Corsmeier, U., Coppola, L., Cocquerez, P., Defer, E., Delanoë, J., Di Girolamo, P., Doerenbecher, A., Drobinski, P., Dufournet, Y., Fourrié, N., Gourley, J. J., Labatut, L., Lambert, D., Le Coz, J., Marzano, F. S., Molinié, G., Montani, A., Nord, G., Nuret, M., Ramage, K., Rison, B., Roussot, O., Said, F., Schwarzenboeck, A., Testor, P., Van-Baelen, J., Vincendon, B., Aran, M., and Tamayo, J.: HyMeX-SOP1, the field campaign dedicated to heavy precipitation and flash flooding in the northwestern Mediterranean, B. Am. Meteorol, Soc., 95, 1083–1100, doi:10.1175/BAMS-D-12-00244.1, 2014.

Fisher, M.: Background error covariance modelling, in: Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean, Seminar on Recent Development in Data Assimilation for Atmosphere and Ocean, ECMWF, 45–63, 2003.

Gustafsson, N., Thorsteinsson, S., Stengel, M., and Holm, E.: Use of a nonlinear pseudo-relative humidity variable in a multivariate formulation of moisture analysis, Q. J. Roy. Meteorol. Soc., 137, 1004–1018, doi:10.1002/qj.813, 2011.

Hodyss, D.: Accounting for skewness in ensemble data assimilation, Mon. Weather Rev., 140, 2346–2358, 2012.

Holm, E., Andersson, E., Beljaars, A., Lopez, P., Mahfouf, J.-F., Simmons, A., and Thépaut, J.-N.: Assimilation and modelling of the hydrological cycle: ECMWF's status and plans, European Centre for Medium-Range Weather Forecasts (ECMWF), Technical Memorandum, 2002.

Houtekamer, P., Lefaivre, L., Derome, J., Ritchie, H., and Mitchell, H. L.: A system simulation approach to ensemble prediction, Mon. Weather Rev., 124, 1225–1242, 1996.

Ingleby, N., Lorenc, A., Ngan, K., Rawlins, F., and Jackson, D.: Improved variational analyses using a nonlinear humidity control variable, Q. J. Roy. Meteorol. Soc., 139, 1875–1887, 2013.

Kalnay, E.: Atmospheric modeling, data assimilation, and predictability, Cambridge University Press, 2003.

Kullback, S.: Information theory and statistics, Wiley, 395pp, 1959.

Laroche S. and Gauthier, P.: A validation of the incremental formulation of 4D variational data assimilation in a nonlinear barotropic flow, Tellus A, 50, 557–572, 1998.

Lawson, W. G. and Hansen, J. A.: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth, Mon. Weather Rev., 132, 1966–1981, 2004.

Lawson, W. G. and Hansen, J. A.: Alignment error models and ensemble-based data assimilation, Mon. Weather Rev., 133, 1687–1709, 2005.

Lorenc, A. C.: Analysis methods for numerical weather prediction, Q. J. Roy. Meteorol. Soc., 112, 1177–1194, 1986.

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. Meteorol. Soc., 140, 2050–2061, 2014.

Ménétrier, B., Montmerle, T., Michel, Y., and Berre, L.: Linear filtering of sample covariances for ensemble-based data assimilation, Part I: optimality criteria and application to variance filtering and covariance localization, Mon. Weather Rev., 143, 1622–1643, 2015.

Michel, Y., Auligné, T., and Montmerle, T.: Heterogeneous convective-scale background error covariances with the inclusion of hydrometeor variables, Mon. Weather Rev., 139, 2994–3015, 2011.

Montmerle, T. and Berre, L.: Diagnosis and formulation of heterogeneous background-error covariances at the mesoscale, Q. J. Roy. Meteorol. Soc., 136, 1408–1420, 2010.

Pailleux, J., Geleyn, J.-F., and Legrand, E.: La prévision numérique du temps avec les modèles ARPÈGE et ALADIN-Bilan et perspectives, La Météorologie, 2000.

Pires, C. A., Talagrand, O. and Bocquet, M.: Diagnosis and impacts of non-Gaussianity of innovations in data assimilation, Physica D: Nonlinear Phenomena, 239 (17),1701–1717, 2010.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics, Q. J. Roy. Meteorol. Soc., 126, 1143–1170, 2000.

Ravela, S., Emanuel, K., and McLaughlin, D.: Data assimilation by field alignment, Physica D, 230, 127–145, 2007.

Raynaud, L., Berre, L., and Desroziers, G.: Accounting for model error in the Météo-France ensemble data assimilation system, Q. J. Roy. Meteorol. Soc., 138, 249–262, 2012.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, Mon. Weather Rev., 139, 976–991, 2011.

Tavolato, C. and Isaksen, L.: On the use of a Huber norm for observation quality control in the ECMWF 4D-Var, Q. J. Roy. Meteorol. Soc., doi:10.1002/qj.2440, online first, 2014.

Thode, H. C.: Testing for Normality, in: Vol. 164 of Statistics: Textbooks and Monographs, Marcel Dekker, New York, 2002.

Wattrelot, E. and Caumont, O. and Mahfouf, J.-F.: Operational implementation of the 1D+ 3D-Var assimilation method of radar reflectivity data in the AROME model, Monthly Weather Review, 142 (5), 1852–1873, 2014

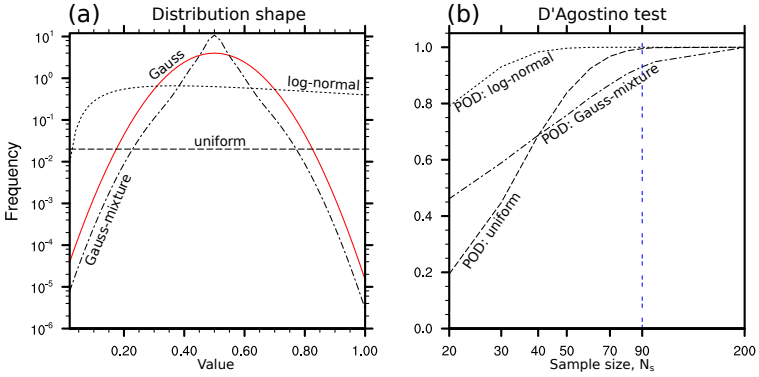

Figure 1. (a) Three non-Gaussian distributions on which POD have been estimated: uniform distribution, log-normal distribution, and Gaussian mixture (see text for description). **(b)** Probability of detection (POD) for K^2 test. POD are computed over $N_{\text{xp}} = 10^5$ one-dimensional experiments for different sample sizes N_{s}.

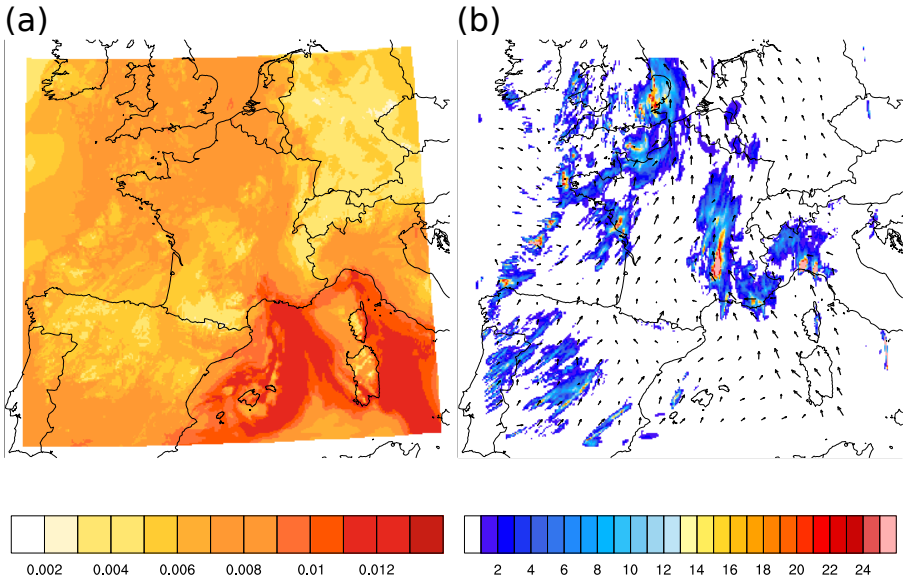

Figure 2. (a) Specific humidity (q, $\mathrm{kg\,kg^{-1}}$) and **(b)** surface cumulative precipitation ($\mathrm{mm\,h^{-1}}$) overlaid with winds vector, at model level 52 ($\approx 920\,\mathrm{hPa}$). Maps are given for one member of AROME-France 3 h-forecasts ensemble, valid at 03:00 UTC the 4 November 2011.

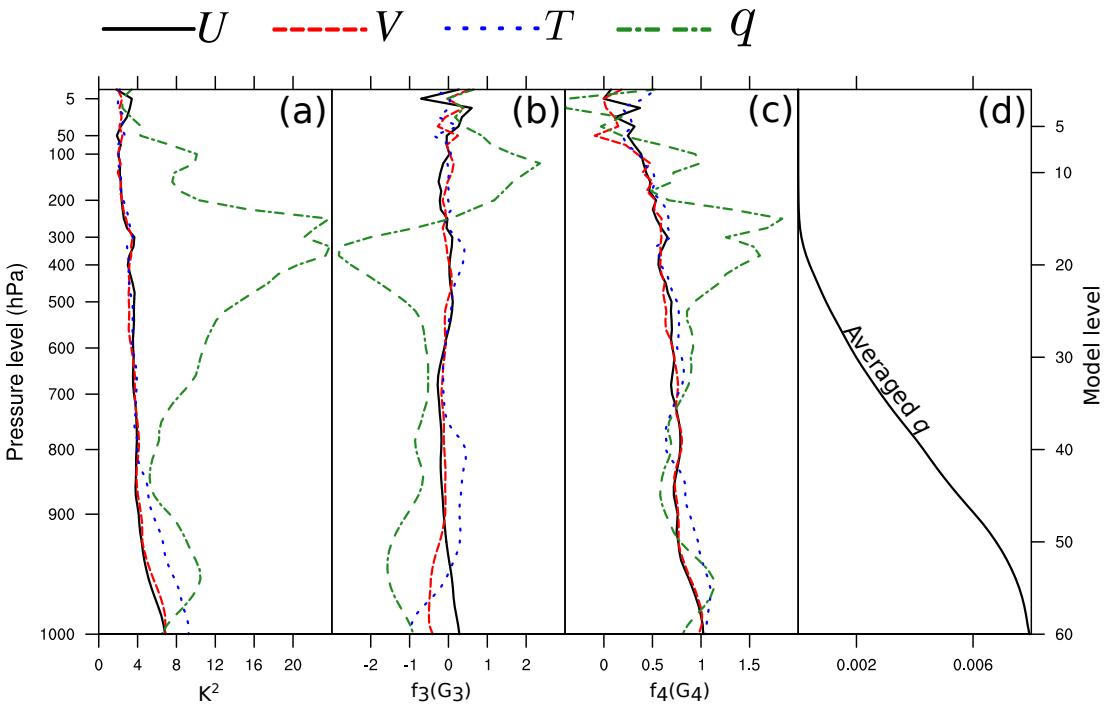

Figure 3. Vertical profiles of **(a)** K^2, **(b)** transformed skewness $f_3(G_3)$, **(c)** transformed kurtosis $f_4(G_4)$, and **(d)** q (kg kg^{-1}) for one member of the ensemble. For each level, values are averaged over the horizontal domain. Profiles are computed from the 90-members ensemble of AROME-France 3 h-forecasts valid at 03:00 UTC the 4 November 2011. Profiles in **(a)**, **(b)**, and **(c)** are given for four model variables: U, V, T, and q.

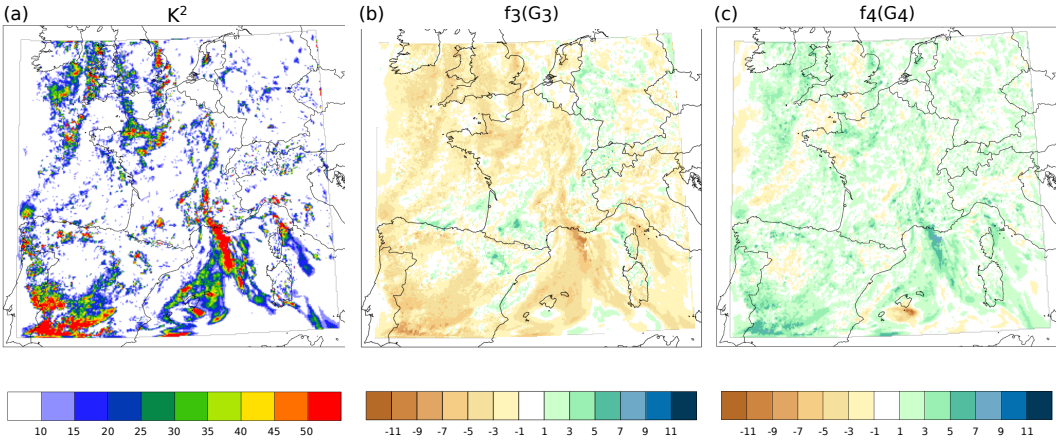

Figure 4. (a) K^2, **(b)** transformed skewness $f_3(G_3)$, and **(c)** transformed kurtosis $f_4(G_4)$, for q at model level 52 ($\approx$ 920 hPa), computed from the 90-members ensemble of AROME-France 3 h-forecasts valid at 03:00 UTC the 4 November 2011.

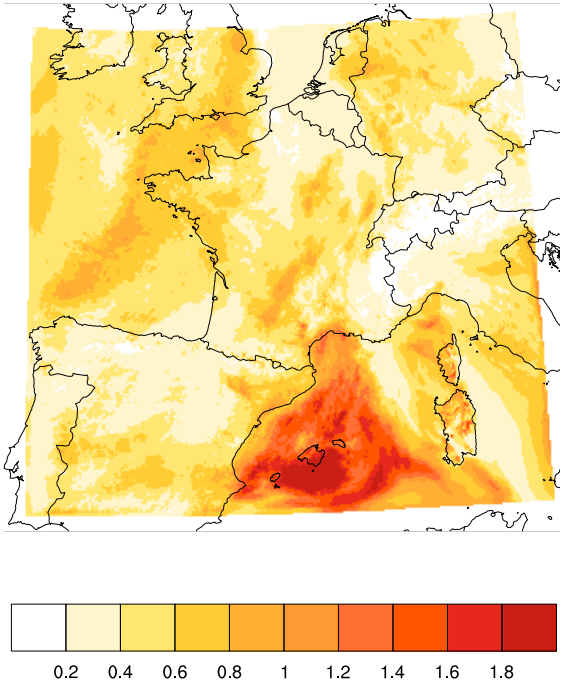

Figure 5. Background-error standard deviations of q ($\mathrm{g\,kg^{-1}}$) for the model AROME-France, at model level 52 ($\approx 920\,\mathrm{hPa}$). Standard deviations are estimated from the 90-members ensemble of 3 h-forecasts, valid at 03:00 UTC the 4 November 2011.

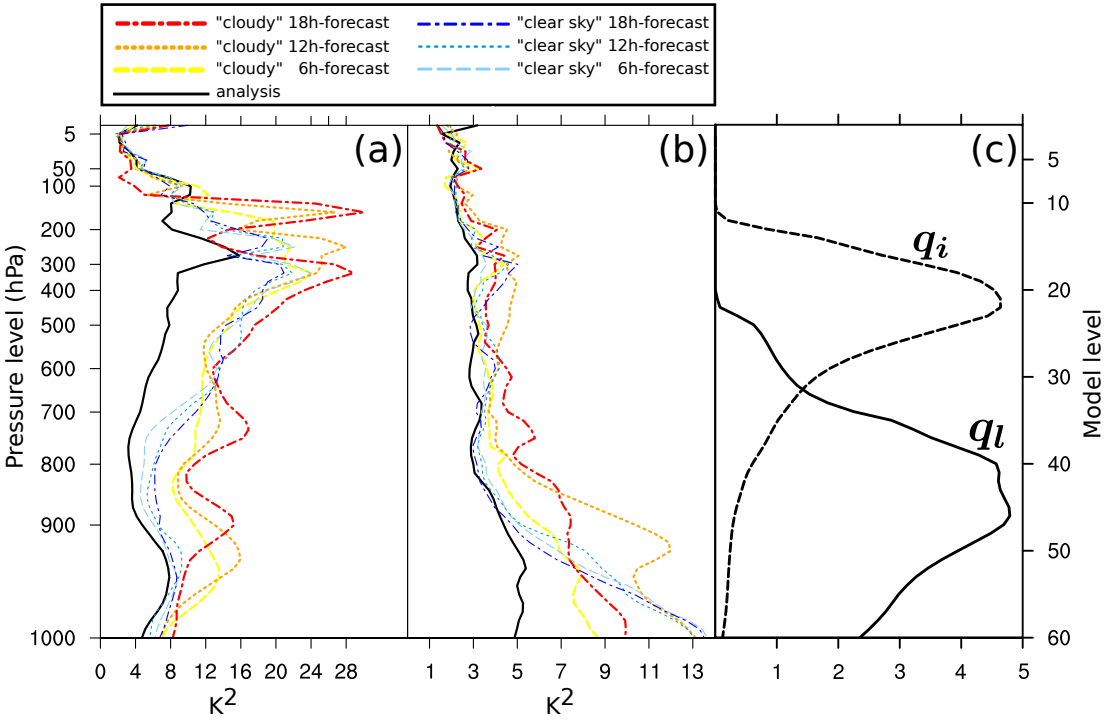

Figure 6. Time evolution (from 00:00 to 18:00 UTC, every 6 h) of the vertical profiles of K^2 for **(a)** q and **(b)** T computed for (thick and hot colours) "cloudy" and (thin and cold colours) "clear sky" points (see text). **(c)** Vertical profiles of ~~averaged~~ liquid cloud q_l (solid line) and ice cloud q_i (dashed line) contents (g kg^{-1}). The cloud contents are averaged over the domain and over times from 06:00 to 18:00 UTC, every 6 h. Initial cloud water profile is null because the hydrometeors are not cycled. Consequently the initial profiles of K^2 are common for the two ~~regions~~bins. Profiles have been computed using forecasts ~~from~~ initialized the 4 November 2011 00:00 UTC.

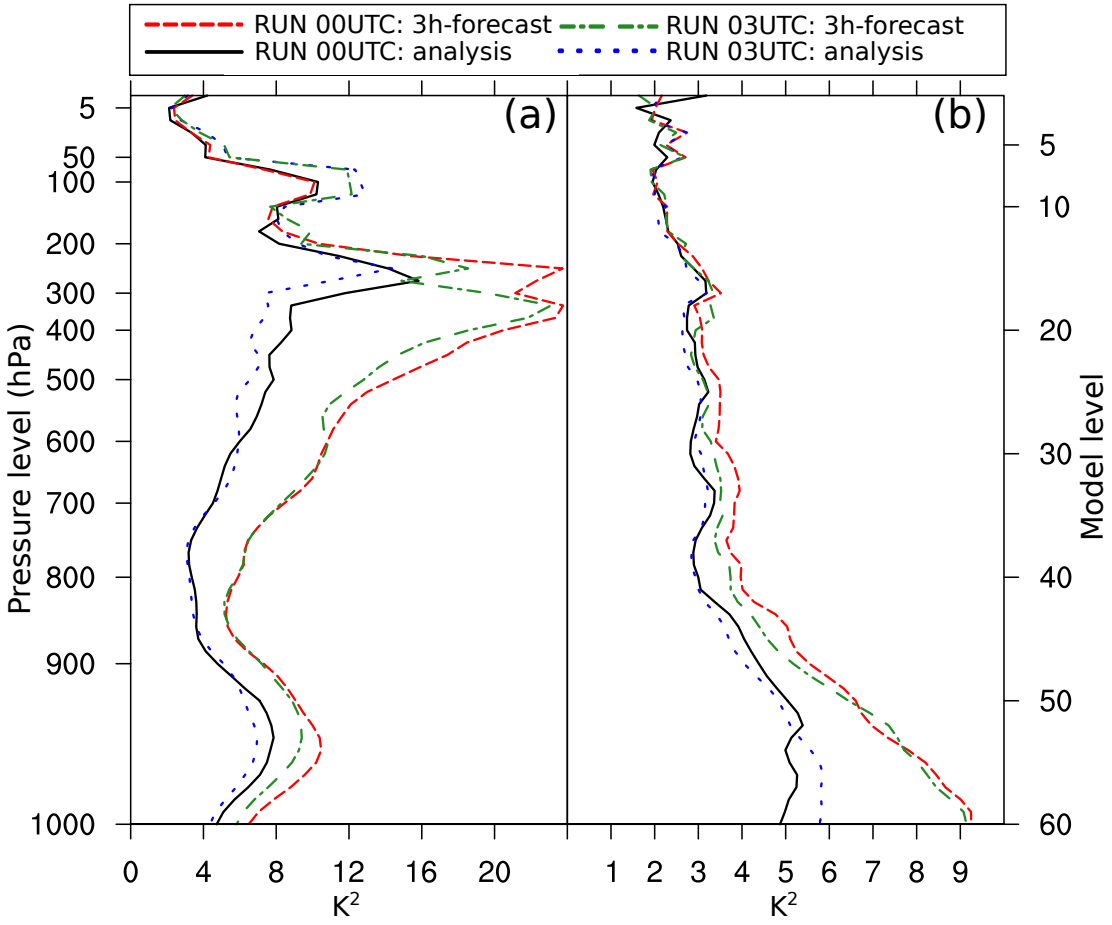

Figure 7. Vertical profiles of K^2 on background and analysis errors for **(a)** q and **(b)** T, for two successive cycled assimilation/3 h-forecast steps starting at 00:00 UTC the 4 November 2011.

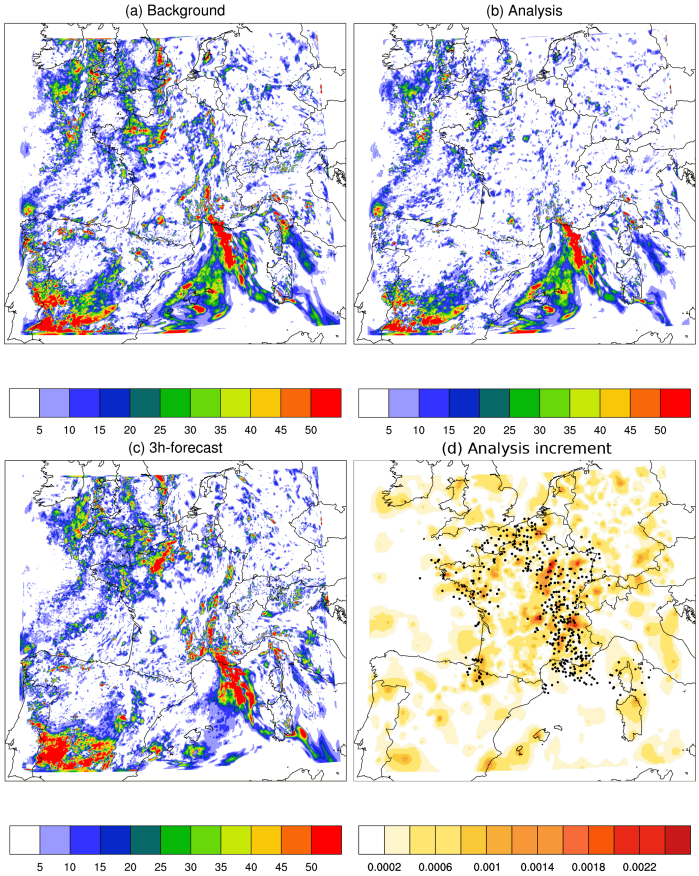

Figure 8. K^2 for q at level 52 ($\approx 920\,\mathrm{hPa}$) for **(a)** the background, **(b)** the analysis, and **(c)** the following 3 h-forecast starting at 03:00 UTC the 4 November 2011. **(d)** Corresponding analysis increment ($\mathrm{kg\,kg^{-1}}$) with positions of radar precipitation observations assimilated.

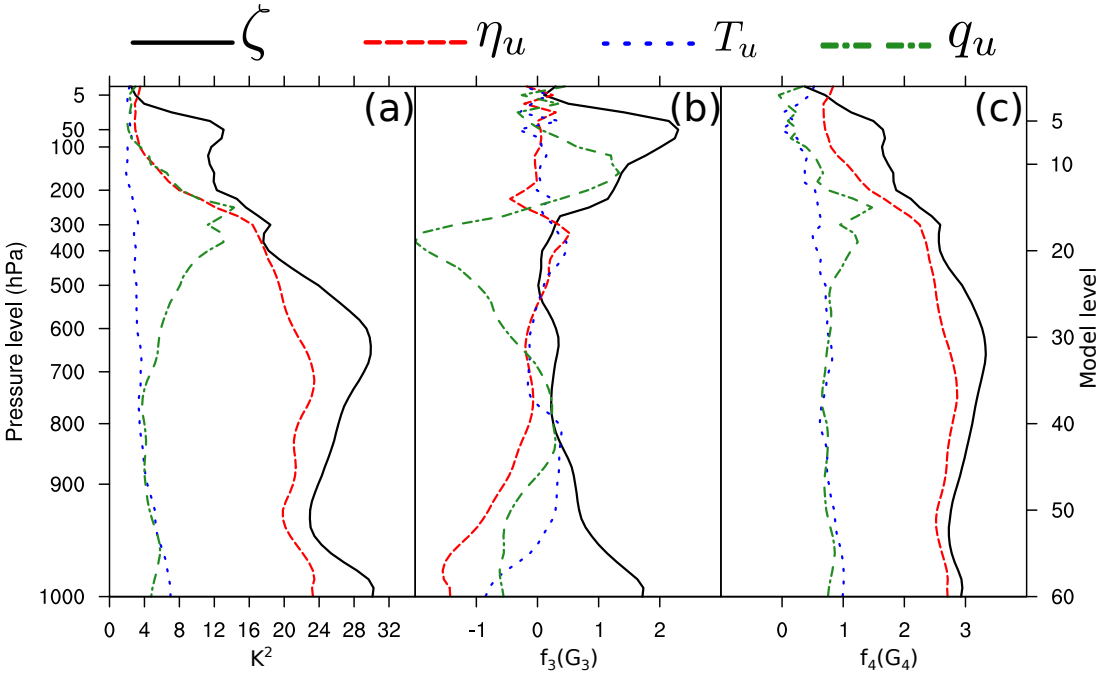

Figure 9. Vertical profiles of **(a)** K^2, **(b)** transformed skewness $f_3(G_3)$, and **(c)** transformed kurtosis $f_4(G_4)$. For each level, values are averaged over the horizontal domain. Profiles are computed from the 90-members ensemble of AROME-France 3 h-forecasts valid at 03:00 UTC the 4 November 2011. Profiles in **(a)**, **(b)**, and **(c)** are given for four control variables: ζ, η_u, T_u, and q_u.

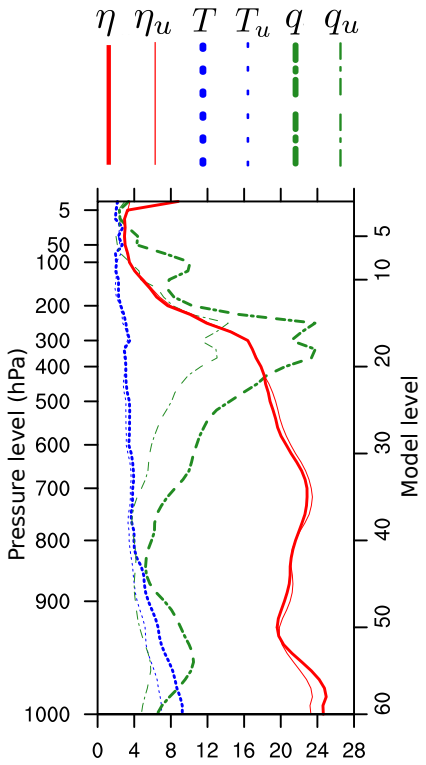

Figure 10. Comparison of K^2 vertical profiles of model variables (thick lines) and control variables (thin lines). Profiles are computed from 3 h-forecasts valid at 03:00 UTC the 4 November 2011.