

Manuscript prepared for Nonlin. Processes Geophys. Discuss.
with version 2014/09/16 7.15 Copernicus papers of the \LaTeX class copernicus.cls.
Date: 27 November 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 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.

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 dynamical control variables used in our data assimilation, namely vorticity and divergence, also show distinct non-Gaussian 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 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

$$P_a(x|y) \propto P_o(y|x)P_b(x), \quad (1)$$

where P_a , P_b and P_o respectively are the PDFs of analysis, background errors (*a priori* PDF), and observations errors. For high dimensional systems, 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) centres have relied on variational assimilation schemes that are Gaussian or corrections to a Gaussian analysis-based strategy.

The time integration of the model nonlinear dynamics 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 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 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 what would be the most probable state in Eq. (1) in the Gaussian case, 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 rely on a Monte-Carlo approach to document the spatial variations of non-Gaussianities of background and of analysis errors for a particular meteorological case, in the context of convective scale NWP. For this purpose, a large ensemble of perturbed cycled assimilations has been set up with the AROME-France¹ model. The perturbations simulate the evolution of the true background and analysis errors

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

(Houtekamer et al., 1996; Fisher, 2003; Berre et al., 2006). Local and spatially averaged diagnostics 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 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, while it is the Gaussianity of the global joint PDF that matters in Eq. (1), only univariate marginal PDF testing for NG are diagnosed in this paper. Spatial variations and average of local values may however give an insight of non Gaussian behaviours for the meteorological case treated here.

2.1 D'Agostino test

The D'Agostino test (hereafter K^2 test, D'agostino et al., 1990) is a univariate 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 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 more than the half of the ensemble is 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 \bar{x} as

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

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

with m_2 , m_3 , and m_4 the sample second, 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; Anscombe and Glynn, 1983). f_3 is defined as

$$A = G_3 \times \sqrt{\frac{(N_s + 1)(N_s + 3)}{6(N_s - 2)}}$$

$$B = 3 \frac{(N_s^2 + 27N_s - 70)(N_s + 1)(N_s + 3)}{(N_s - 2)(N_s + 5)(N_s + 7)(N_s + 9)}$$

$$C = \sqrt{2(B - 1)} - 1$$

$$D = \sqrt{C}$$

$$E = \frac{1}{\sqrt{\ln(D)}}$$

$$F = \frac{A}{\sqrt{\frac{2}{C-1}}}$$

$$f_3(G_3) = E \times \ln(F + \sqrt{F^2 + 1})$$

5

and f_4 is defined as

$$O = G_4 \times \frac{N_s(N_s + 1)}{(N_s - 1)(N_s - 2)(N_s - 3)} - 3 \frac{(N_s - 1)}{(N_s + 1)}$$

$$P = \frac{24N_s(N_s - 2)(N_s - 3)}{(N_s + 1)^2(N_s + 3)(N_s + 5)}$$

$$Q = \frac{(N_s - 2)(N_s - 3)}{(N_s + 1)(N_s - 1)\sqrt{P}} \times O$$

$$5 \quad R = \frac{6(N_s^2 - 5N_s + 2)}{(N_s + 7)(N_s + 9)} \sqrt{\frac{6(N_s + 3)(N_s + 5)}{N_s(N_s - 2)(N_s - 3)}}$$

$$S = 6 + \frac{8}{R} \left[\frac{2}{R} + \sqrt{1 + \frac{4}{R^2}} \right]$$

$$T = \frac{1 - \frac{2}{S}}{1 + Q\sqrt{\frac{2}{S-4}}}$$

$$10 \quad f_4(G_4) = \frac{1 - \frac{2}{9S} - T^{\frac{1}{3}}}{\sqrt{\frac{2}{9S}}}.$$

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.

15 $f_3(G_3)$ 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:

$$K^2 = f_3^2(G_3) + f_4^2(G_4). \quad (4)$$

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.

2.2 Evaluation

The efficiency of the K^2 test can be evaluated by measuring its probability of detection (POD) for the Gaussian 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 $\text{POD} = 1$.

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 $N_{\text{xp}} = 10^5$ experiments. For both tests, different ensemble sizes N_s are tested ($N_s = 20, 30, 40, 50, 60, 70, 80, 90, 100, 200$). 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 $\text{POD} > 0.8$) 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.

A review of other well-established tests for 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 ARPEGE². Assimilation steps are done every three hours with a 3D-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-

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

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, named Cevenol 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 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 3 h-forecasts of 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 left-tailed PDF of the background errors, meaning that many values are more humid than the ensemble mean.

In the troposphere, K^2 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 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 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 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 $f_3(G_3)$ is mainly negative (corresponding to left-tailed distributions) over the domain and has a larger contribution than transformed kurtosis $f_4(G_4)$ 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.

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.

5 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 .

10 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
15 simulated cloud water exceeds 0.1 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.

20 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
25 acting on nearly Gaussian pdfs instead of linear processes acting non-Gaussian pdfs.

It is interesting to notice that different 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 ranges, while the amplitude of the higher one increases. According to Fig. 6c that displays the time average 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, 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 ζ , the unbalanced divergence η_u , the unbalanced temperature and surface pressure $(T, P_s)_u$, 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, 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 an 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 indicates 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 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 in average a locally nearly Gaussian variable (see Fig.3a), its spatial differentiation 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 differentiation for at least a part of the NG displayed for the dynamical control variables.

While very similar, the horizontal structures of K^2 for 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 below the tropopause. From surface to 750 hPa, NG of q_u is equal to 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 local NG of a NWP system at convective scale. This diagnostic is computed from the univariate sample skewness and kurtosis from a 90-members ensemble. Even if checking local NG is not sufficient to describe deviations from the global Gaussian hypothesis for AROME-France background and analysis errors, spatial variations and average of local diagnostics give an insight of non-Gaussian behaviours for a case study characterized by a Cevenol event and an active cold front.

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

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, 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.

5 Despite this work is attributing non-Gaussian 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. 10 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.

15 This study uses an ensemble at convective scale that does not include model error either in the analysis 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 20 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 25 scientific advice and their careful readings that proved very useful to improve the manuscript.

References

- Anderson, E. and 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.
- 5 Anscombe, Francis J. and Glynn, William J.: Distribution of the kurtosis statistic b_2 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, 10 *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 Ștefănescu, S., and Belo Pereira, M.: The representation of the analysis effect in three error simulation techniques, *Tellus A*, 58, 196–209, 2006.
- 15 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.
- 20 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 G_1 , *Biometrika*, 57, 679–681, 25 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.
- 30 Ducrocq, V., Belamari, S., Boudevillain, B., Bousquet, O., Cocquerez, P., Doerenbecher, A., Drobinski, P., Flamant, C., Labatut, L., Lambert, D.,

- Nuret, M., Richard, E., Rousset, O., Testor, P., Arbogast, P., Ayrat, 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.
- 5 Ducrocq, V., Braud, I., Davolio, S., Ferretti, R., Flamant, C., Jansá, A., Kalthoff, N., Richard, E., Taupier-Letage, I., Ayrat, 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., Rousset, O., Saïd, 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.
- 10 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.
- 20 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.
- 25 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., Lefavre, L., Derome, J., Ritchie, H., and Mitchell, H. L.: A system simulation approach to ensemble prediction, *Mon. Weather Rev.*, 124, 1225–1242, 1996.
- 30 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.

5 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.

10 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.

15 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.

20 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.

25 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.

30 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.
- 5 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

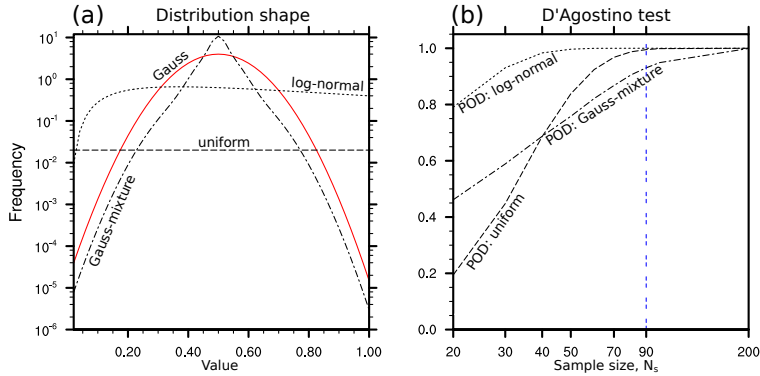


Figure 1. (a) Three non-Gaussian distributions on which POD have been estimated: uniform distribution, log-normal distribution, and Gaussian mixture (see text for description). (b) Probability of detection (POD) for K^2 test. POD are computed over $N_{xp} = 10^5$ one-dimensional experiments for different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

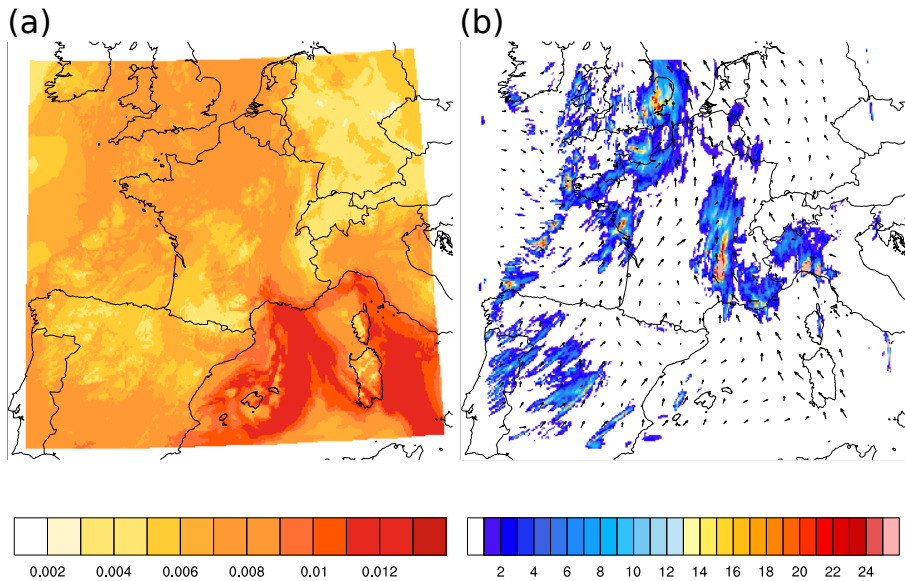


Figure 2. (a) Specific humidity (q , kg kg^{-1}) and (b) surface cumulative precipitation (mm h^{-1}) overlaid with winds vector, at model level 52 ($\approx 920 \text{ hPa}$). Maps are given for one member of AROME-France 3 h-forecasts ensemble, valid at 03:00 UTC the 4 November 2011.

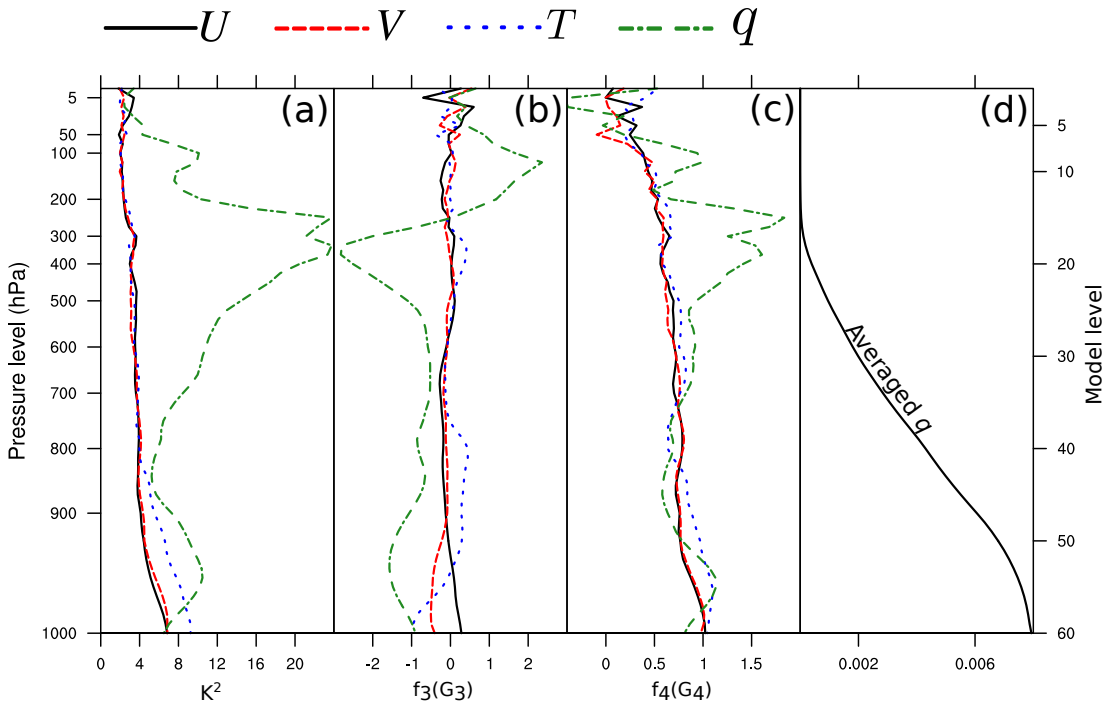


Figure 3. Vertical profiles of (a) K^2 , (b) transformed skewness $f_3(G_3)$, (c) transformed kurtosis $f_4(G_4)$, and (d) q (kg kg^{-1}) for one member of the ensemble. For each level, values are averaged over the horizontal domain. Profiles are computed from the 90-members ensemble of AROME-France 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 .

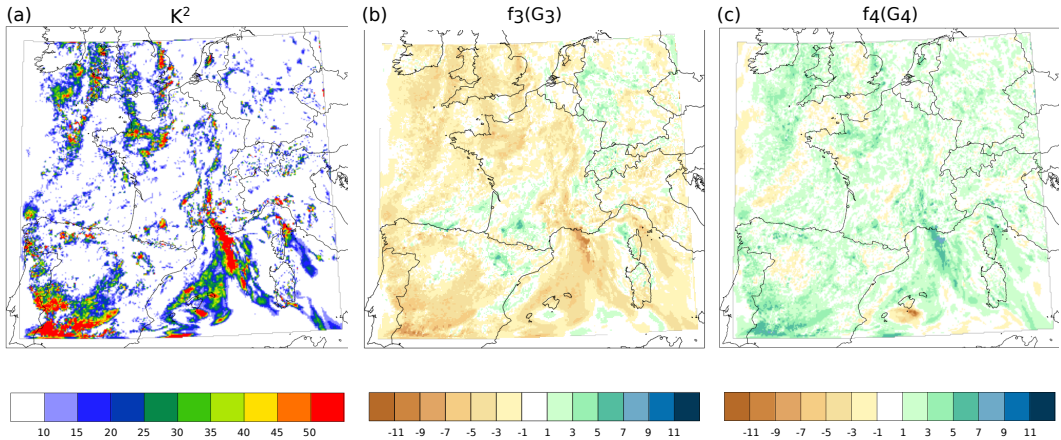


Figure 4. (a) K^2 , (b) transformed skewness $f_3(G_3)$, and (c) transformed kurtosis $f_4(G_4)$, for q at model level 52 (≈ 920 hPa), computed from the 90-members ensemble of AROME-France 3 h-forecasts valid at 03:00 UTC the 4 November 2011.

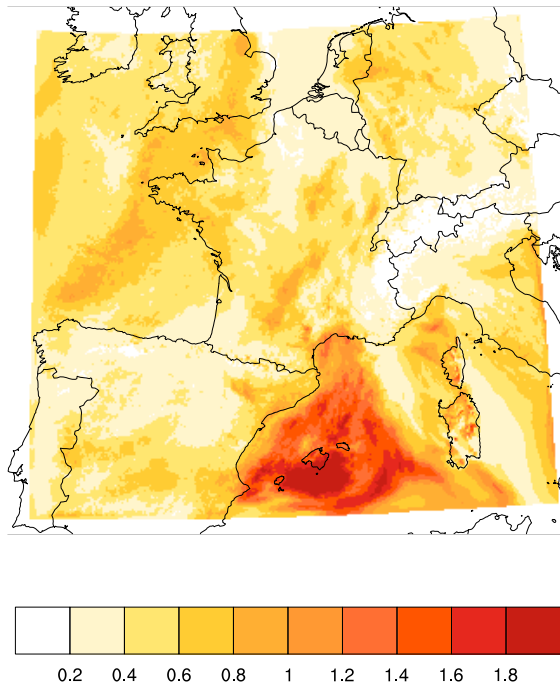


Figure 5. Background-error standard deviations of q (g kg^{-1}) for the model AROME-France, at model level 52 (≈ 920 hPa). Standard deviations are estimated from the 90-members ensemble of 3 h-forecasts, valid at 03:00 UTC the 4 November 2011.

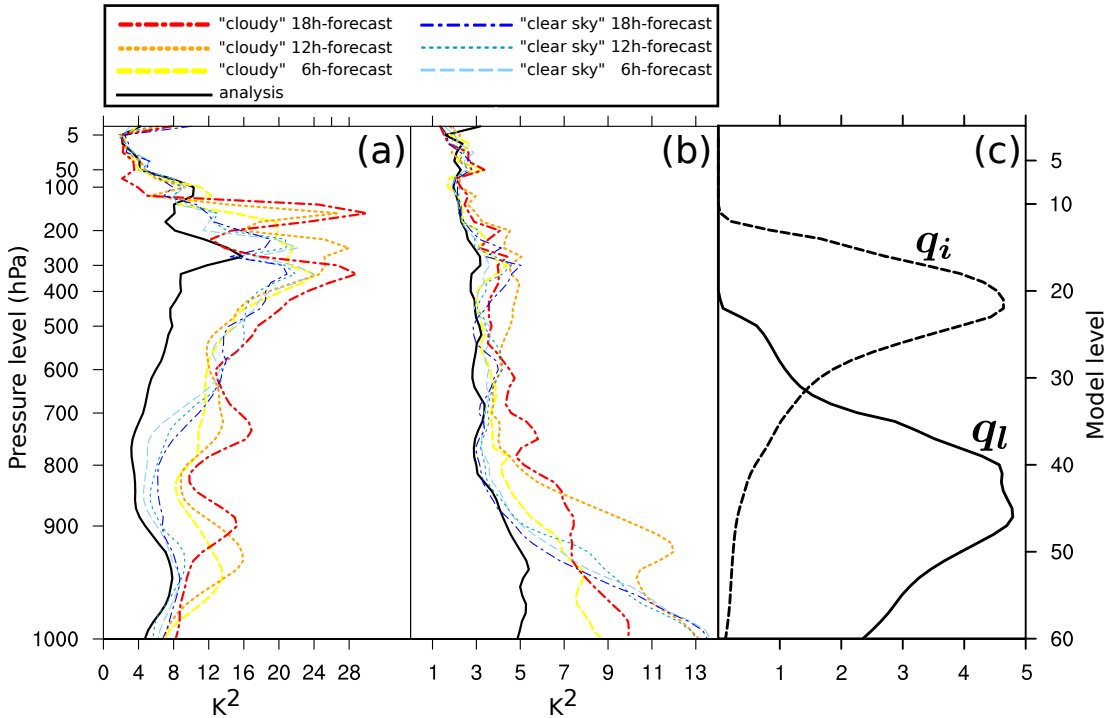


Figure 6. Time evolution of the vertical profiles of K^2 for (a) q and (b) T computed for (thick and hot colours) “cloudy” and (thin and cold colours) “clear sky” points (see text). (c) Vertical profiles of liquid cloud q_l (solid line) and ice cloud q_i (dashed line) contents (g kg^{-1}). The cloud 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 time profiles of K^2 are common for the two bins. Profiles have been computed from 00:00 to 18:00 UTC, every 6 h, using forecasts initialized with analysis states valid the 4 November 2011 00:00 UTC.

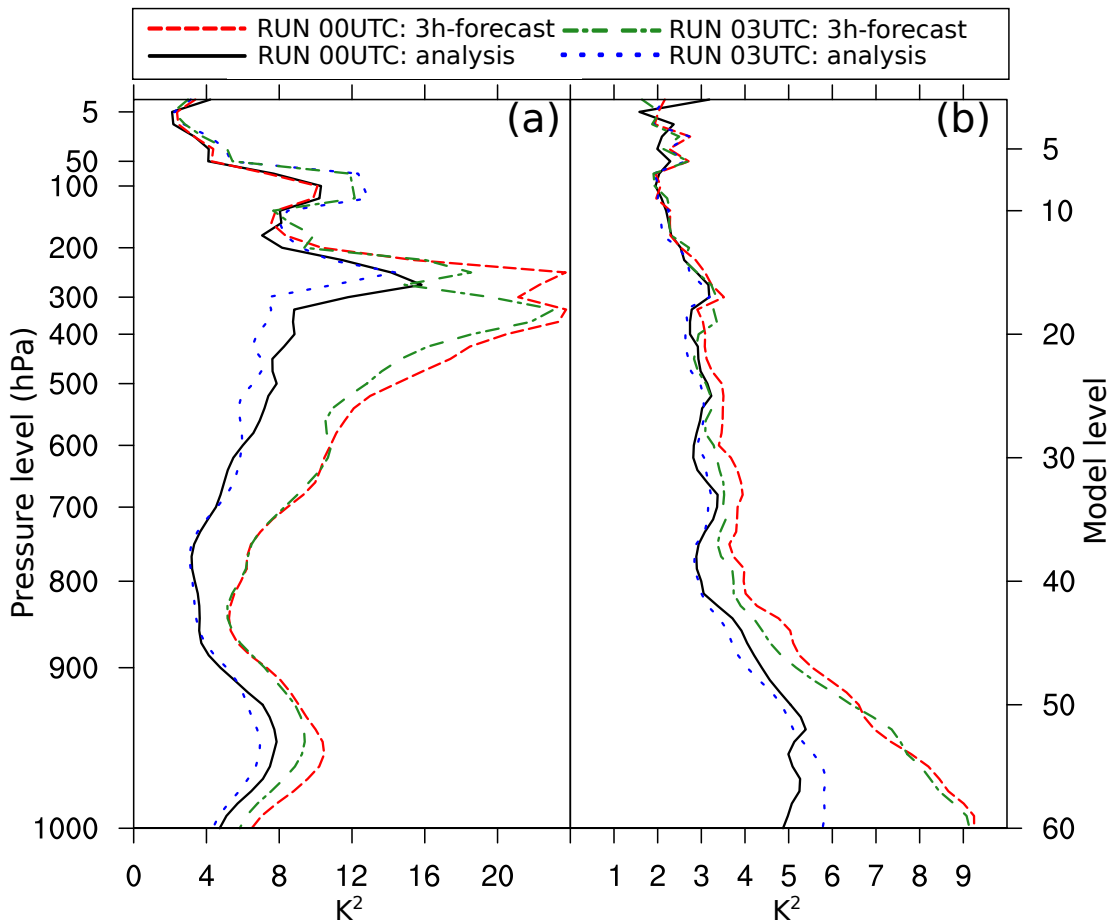


Figure 7. Vertical profiles of K^2 on background and analysis errors for (a) q and (b) T , for two successive cycled assimilation/3 h-forecast steps starting at 00:00 UTC the 4 November 2011.

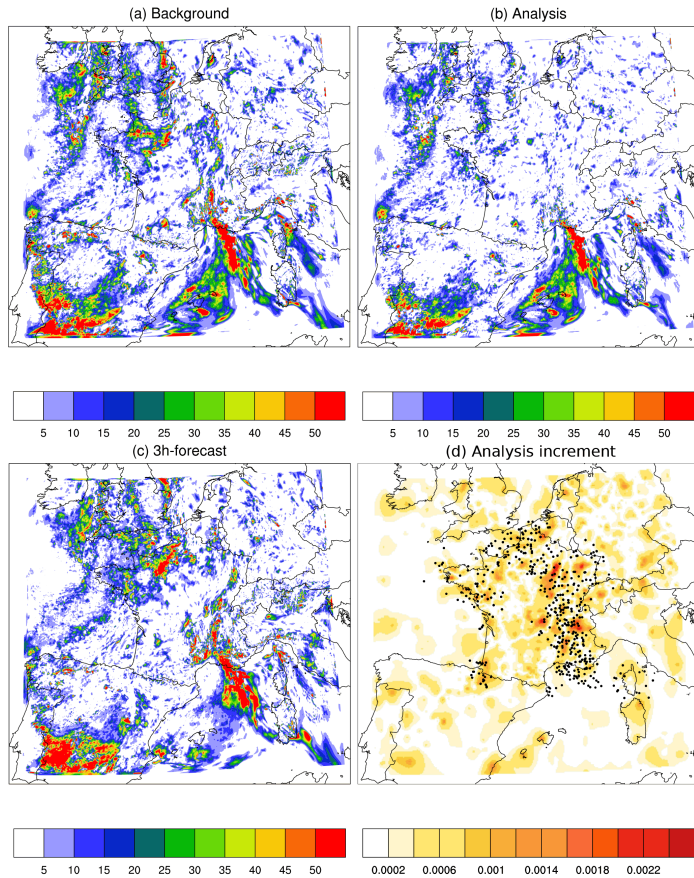


Figure 8. K^2 for q at level 52 (≈ 920 hPa) for (a) the background, (b) the analysis, and (c) the following 3 h-forecast starting at 03:00 UTC the 4 November 2011. (d) Corresponding analysis increment (kg kg^{-1}) with positions of radar precipitation observations assimilated.

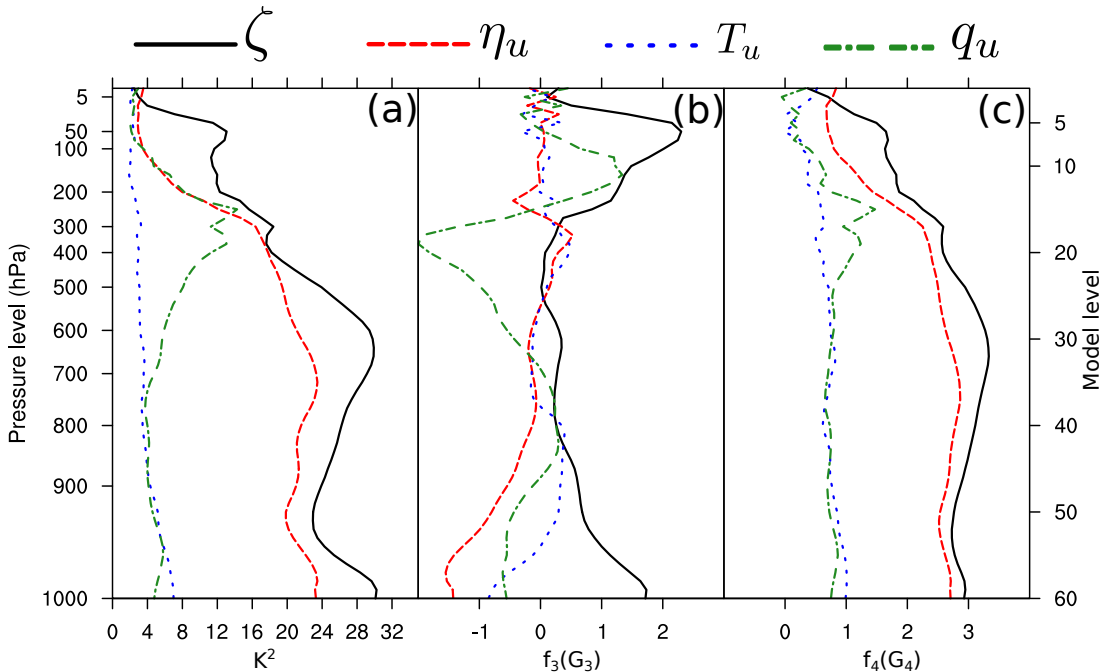


Figure 9. Vertical profiles of (a) K^2 , (b) transformed skewness $f_3(G_3)$, and (c) transformed kurtosis $f_4(G_4)$. For each level, values are averaged over the horizontal domain. Profiles are computed from the 90-members ensemble of AROME-France 3 h-forecasts valid at 03:00 UTC the 4 November 2011.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

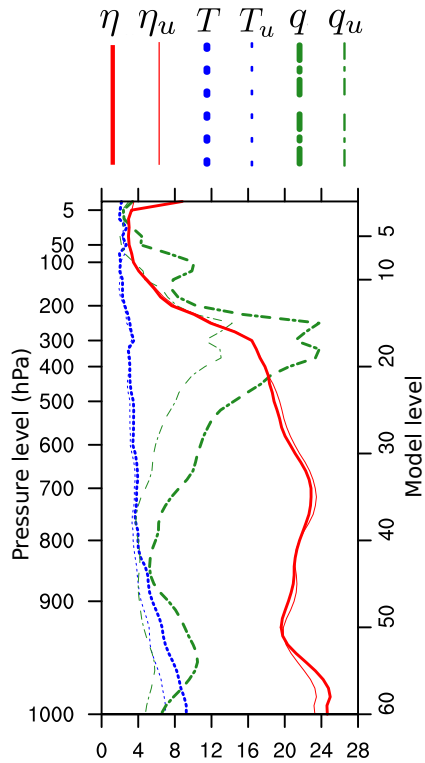


Figure 10. Comparison of K^2 vertical profiles of model variables (thick lines) and control variables (thin lines). Profiles are computed from 3 h-forecasts valid at 03:00 UTC the 4 November 2011.