Physica D 239 (2010) 1701–1717

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Diagnosis and impacts of non-Gaussianity of innovations in data assimilation

Carlos A. Pires [a,*], Olivier Talagrand [b], Marc Bocquet [c,d]

[a] Instituto Dom Luiz, University of Lisbon, Portugal
[b] Laboratoire de Météorologie Dynamique (LMD), École Normale Supérieure, Paris, France
[c] Université Paris-Est, CEREA, Joint Laboratory École des Ponts ParisTech and EDF R&D, Champs-sur-Marne, France
[d] INRIA, Paris-Rocquencourt Research Center, France

ARTICLE INFO

Article history:
Received 13 May 2009
Received in revised form
19 February 2010
Accepted 12 May 2010
Available online 21 May 2010
Communicated by H.A. Dijkstra

Keywords:
Non-Gaussianity
Data assimilation
Innovation statistics
Maximum entropy method

ABSTRACT

Most of the atmospheric and oceanic data assimilation (DA) schemes rely on the Best Linear Unbiased Estimator (BLUE), which is sub-optimal if errors of assimilated data are non-Gaussian, thus calling for a full Bayesian data assimilation. This paper contributes to the study of the non-Gaussianity of errors in the observational space. Possible sources of non-Gaussianity range from the inherent statistical skewness and positiveness of some physical observables (e.g. moisture, chemical species), the nonlinearity, both of the data assimilation models and of the observation operators among others. Deviations from Gaussianity can be justified from *a priori* hypotheses or inferred from statistical diagnostics of innovations (observation minus background), leading to consistency relationships between the error statistics. From samples of observations and backgrounds as well as their specified error variances, we evaluate some measures of the innovation non-Gaussianity, such as the skewness, kurtosis and negentropy. Under the assumption of additive errors and by relating statistical moments from both data errors and innovations, we identify potential sources of the innovation non-Gaussianity. These sources range from: (1) univariate error non-Gaussianity, (2), nonlinear correlations between errors, (3) spatio-temporal variability of error variances (heteroscedasticity) and (4) multiplicative noise. Observational and background errors are often assumed independent. This leads to variance-dependent bounds for the skewness and the kurtosis of errors. From innovation statistics, we assess the potential DA impact of some scenarios of non-Gaussian errors. This impact is measured through the mean square difference between the BLUE and the Minimum Variance Unbiased Estimator (MVUE), obtained with univariate observations and background estimates. In order to accomplish this, we compute maximum entropy probability density functions (pdfs) of the errors, constrained by the first four order moments. These pdfs are then used to compute the Bayesian posterior pdf and the MVUE. The referred impact is studied for a large range of statistical moments, being higher for skewed innovations and growing in average with the skewness of data errors, specially if the skewnesses have the same sign. An application has been performed to the quality-accepted ECMWF innovations of brightness temperatures of a set of High Resolution Infrared Sounder (HIRS) channels. In this context, the MVUE has led in some extreme cases to a potential reduction of 20%–60% of the posterior error variance as compared to the BLUE, specially for extreme values of the innovations.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Most of the operational atmospheric–oceanic data assimilation (DA) systems rely on the linear estimation theory which aims to compute the Best Linear Unbiased Estimator (BLUE) of the system state from all available data (observations and background) and the corresponding error covariance matrices. For some cases, errors of assimilated data (direct or indirect observations and *a priori* background estimates) have non-Gaussian probability density functions (pdfs). When data errors are non-Gaussian, the *a posteriori* error variance minimiser is the Minimum Variance Unbiased Estimator (MVUE), obtained from the mean of the Bayesian *a posteriori* pdf combining all data [1]. However, when the MVUE is restricted to the subspace of linear estimators, then it coincides with the BLUE, even when data errors are non-Gaussian. Before implementing a full Bayesian DA scheme, we shall evaluate from simple diagnostics, what is the potential impact of considering non-Gaussian data errors. This paper aims to contribute to the study and impact in data assimilation of the non-Gaussianity of errors in the observational space.

Sampling statistics and histograms of innovations (observations minus projected backgrounds), often reveal statistically significant

* Corresponding address: Faculdade de Ciências, Edifício C8, DEGGE, Campo Grande, 1749-016 Lisboa, Portugal. Tel.: +351 217500886; fax: +351 217500807.
E-mail address: capires@fc.ul.pt (C.A. Pires).

0167-2789/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2010.05.006

non-Gaussianity, thus proving that observation or background errors are separately non-Gaussian or even that errors are jointly non-Gaussian in the observational space. This in turn gives proof of the BLUE sub-optimality as far as the *a posteriori* error variance minimization is concerned. The assumption of Gaussian errors generally comes with the linearity of the observation operators and Gaussianity of errors in state variables, which is not true in general. In fact, assimilated observations can have instrumental and representativeness errors of non-Gaussian nature, particularly if they come from complex discontinuous and/or nonlinear observation operators such as in the cases of specific humidity, tracing and reactive chemical species (e.g. ozone) [2,3] cloud droplet concentrations, land-species concentrations, remote sensing data (e.g. satellite radiances, scatterometer and Doppler radar data).

Moreover, Gaussianity of errors implies a symmetric pdf around the mean, which is not valid for example in cases dealing with moisture, gaseous pollutants and aerosols. Extreme events like droughts and floods are also not well modeled by Gaussian pdfs. Another source of non-Gaussianity in observation errors is the presence of extreme gross errors [4]. The modeling of non-Gaussianity in geophysical data assimilation is thoroughly reviewed in [5]. Data assimilation schemes coping with non-Gaussian error pdfs such as the lognormal are now available [6–9]. When errors are non-Gaussian, the maximum likelihood posterior state lies at the absolute mode of the Bayesian posterior pdf or at the *Maximum A Posteriori* (MAP). This likelihood is normally expressed in terms of a cost function which can be decomposed into a quadratic term corresponding to Gaussian errors, with the minimum at the BLUE, into a term measuring the departure from Gaussianity and eventually into a term of weak constraining. Under strongly constrained variational DA, the cost function minimum, the MAP, is shifted from the BLUE in accordance with the weight of non-Gaussianity. This appears as a correcting term due to *a priori* ill-specified error pdfs or equivalently to a generalized parametric bias. The removal of parametric bias, as considered above, is automated by the Variational Bias Correction [10] by the addition of penalty terms to the cost function, weighting the deviation of parameters from certain *a priori* estimates. An alternative scheme of dealing with non-Gaussianity, though still not satisfactory, is to build nonlinear control variables by Gaussian anamorphosis, whose small departures are nearly Gaussian as presented by Holm [11] for moisture. The assumptions about pdfs of errors always depend on statistical hypotheses, which must be supported by independent sampling statistics.

Innovations are expressed as differences between observation errors and background errors in the observational space. Therefore, measures of non-Gaussianity of innovations, estimated from samples, can be directly related to the non-Gaussianity of errors, through their skewnesses and kurtoses. We aim to clarify how data (observations and background) errors can accommodate non-Gaussianity under certain hypothesis on its origin, either univariate or bivariate, and also under the hypothesis of observation and background error's independence. This is shown in Section 3 using samples of innovations, described in Section 2, of the brightness temperature (BT) measured at a number of certain HIRS (High resolution Infrared Sounder) channels carried by satellite NOAA-17. The samples are collected from the ECMWF data assimilation feedback files. Then, in Section 4, we establish statistical consistency relationships linking the non-Gaussianity of innovations and errors, particularly when errors are independent. In Section 5, we apply the Maximum Entropy (ME) method [12–14], constrained by innovations statistical moments up to the fourth order, and a few hypotheses about the errors, in order to find the least committing error pdfs, compatible with the available statistical information. This is performed on the observational space for univariate observations and background estimates. Finally in Section 6, we compute the Minimum Variance Unbiased Estimator (MVUE), obtained with the error pdfs determined by the ME method, and compare it with the BLUE in the observational space. The average difference between MVUE and BLUE measures the potential impact of taking into account non-Gaussianity of errors in the *a posteriori* analysis. This impact is measured for an extended range of possible values of the skewness and kurtosis of innovations and errors and for different scenarios of non-Gaussian origin. In Section 7, the method is applied to observations and background estimates of BTs. Discussion and conclusions come in Section 8 followed by an Appendix with the numerical implementation and properties of the pdfs obtained by the ME method.

2. Description of the data

The data on which the diagnostics are performed below are extracted from the database of ECMWF. It consists on quality-controlled observations of brightness temperature (BT) performed over the ocean by the High resolution Infrared Sounder (HIRS) carried by satellite NOAA-17 (http://www.ncdc.noaa.gov). The observations have been performed in 8 wavelength channels: 4, 5, 6, 7, 11, 12, 14 and 15, over 8 different days in the period October 2005–January 2006: 20051015, 20051030, 20051115, 20051201, 20051215, 20051230, 20060115 and 20060130 in yyyyMMdd notation. In the ECMWF assimilation system, both the observations and the background are assumed to be unbiased. The standard deviation σ_{so} of the observational error, as specified by the ECMWF assimilation system, is taken constant for each channel. Hereafter, the subscript *so* stands for 'specified for the observations'. The specified standard deviation σ_{sb} of the background error depends on the channel, pixel position and temporal instant [15,16]. The subscript *sb* stands for 'specified for the background'. The five channels: 5, 6, 7, 14 and 15, are sensitive to sea-ice, whose temperature is considered to be ill-predicted by the forecast model. The corresponding σ_{sb} is therefore taken larger when sea-ice is forecasted, which occurs mostly at latitudes poleward of 60°. Fig. 1, relative to channel 6, shows this fact through the scatter plot of pairs (σ_{sb}, North latitude), thus putting in evidence the clustering in the distribution of σ_{sb}. This has led to partitioning the data pairs (observation, background) for each of the five channels into two subsamples, meant to correspond respectively to 'ice-free' and 'ice-covered' surface conditions. In the case of Fig. 1 for channel 6, the partitioning is defined by the thick lines; a similar partitioning is used for the other ice-sensitive channels. Samples corresponding to 'ice-free' surface conditions are identified by the subscript s: 4s, 5s, 6s, 7s, 11s, 12s, 14s and 15s, where the leading number stands for channel. Data points under 'ice-free' surface conditions were previously 'thinned' to a mutual distance of at least 300 km. The samples for 'ice-covered' conditions, identified by the subscript i: 5i, 6i, 7i, 14i and 15i were not subject to any thinning in order not to drastically reduce the sample size. The main statistical features of the 13 samples are given in Table 1. For each sample, N denotes the size while $m(\sigma_{sb})$ and $\mathrm{std}(\sigma_{sb})$ denote, respectively, the mean and standard deviation of σ_{sb} estimated from data. It is seen that both $m(\sigma_{sb})$ and $\mathrm{std}(\sigma_{sb})$ are larger for the 'ice-covered' samples, comparatively to the 'ice-free' samples for each of the channels 5, 6, 7, 14 and 15.

3. Innovation statistics

We recall that the innovation corresponding to an observation $y = x + \varepsilon_o$, where x is the exact value and ε_o is the observational error, is defined as the difference:

$$d = y - x_b = \varepsilon_o - \varepsilon_b, \tag{1}$$

Table 1
Statistics of the background errors, observation errors and innovations for the samples 4s, 5s, 6s, 7s, 11s, 12s, 14s 15s for 'ice-free' conditions and samples 5i, 6i, 7i, 14i and 15i for 'ice-covered' conditions. Table lists: (a) sample size N; (b) Specified standard deviation σ_{so}, of the observation error, mean $m(\sigma_{sb})$ and standard deviation std(σ_{sb}) of the specified standard deviation of the background error, innovation bias b_d, standard deviation σ_d of the innovation (all in K); (c) Skewness s_d, kurtosis k_d, and approximated negentropy J_d of the innovation with the 99%-statistically significant values of s_d, k_d and J_d marked in bold; (d) Heteroscedasticity measures h_{1b}, h_{3b} and h_{4b}, of σ_{sb}; (e) Maximum variance fraction β_g of the Gaussian error; (f) Bounds of the tuning factors: f_{o1}, f_{b1}, f_{o2} and f_{b2} of the specified standard deviation of errors. See text for details concerning quantities in (c), (d), (e) and (f).

Sample	4s	5s	6s	7s	11s	12s	14s	15s	5i	6i	7i	14i	15i
N	5907	4419	4132	3580	4426	6741	4081	4234	4980	4306	2043	3237	3981
σ_{so}	0.6	0.6	0.6	0.75	1.0	2.0	0.5	0.6	0.6	0.6	0.75	0.5	0.6
$m(\sigma_{sb})$	0.083	0.126	0.243	0.432	0.904	1.040	0.354	0.207	0.378	1.031	2.502	1.442	0.618
std(σ_{sb})	0.013	0.027	0.058	0.132	0.252	0.425	0.135	0.050	0.059	0.186	0.419	0.256	0.111
b_d	−0.01	−0.08	−0.13	−0.06	−0.02	0.17	−0.07	−0.04	0.03	−0.05	0.03	0.03	−0.01
σ_d	0.184	0.221	0.343	0.422	1.011	1.305	0.321	0.274	0.261	0.439	0.502	0.448	0.374
s_d	**−0.70**	**−0.63**	**−0.25**	0.10	**0.38**	**0.27**	−0.01	**−0.22**	**−0.34**	**−0.21**	**−0.26**	**−0.14**	**−0.21**
k_d	**1.02**	**0.90**	**0.35**	**0.56**	**0.65**	**2.07**	0.14	**0.23**	−0.07	**−0.47**	**−0.24**	**−0.33**	**−0.28**
J_d	**0.063**	**0.050**	**0.008**	**0.007**	**0.021**	**0.095**	0.0004	**0.005**	**0.010**	**0.008**	**0.007**	**0.004**	**0.005**
h_{1b}	0.99	0.98	0.97	0.96	0.96	0.93	0.93	0.97	0.99	0.98	0.99	0.98	0.98
h_{3b}	1.04	1.08	1.09	1.19	1.11	1.23	1.29	1.10	1.03	1.04	1.04	1.04	1.05
h_{4b}	1.10	1.23	1.28	1.61	1.33	1.72	2.08	1.32	1.10	1.11	1.10	1.12	1.12
β_g	0.646	0.675	0.852	0.983	0.810	0.965	1.000	0.843	0.637	0.566	0.622	0.643	0.638
f_{o1}	0.248	0.304	0.528	0.557	0.910	0.641	0.643	0.419	0.347	0.551	0.528	0.719	0.499
f_{b1}	1.762	1.409	1.266	0.925	0.969	1.141	0.848	1.182	0.544	0.316	0.156	0.245	0.477
f_{o2}	0.184	0.211	0.220	0.074	0.441	0.123	0.014	0.181	0.262	0.482	0.412	0.536	0.376
f_{b2}	1.305	0.978	0.527	0.122	0.470	0.219	0.018	0.511	0.411	0.276	0.122	0.183	0.359

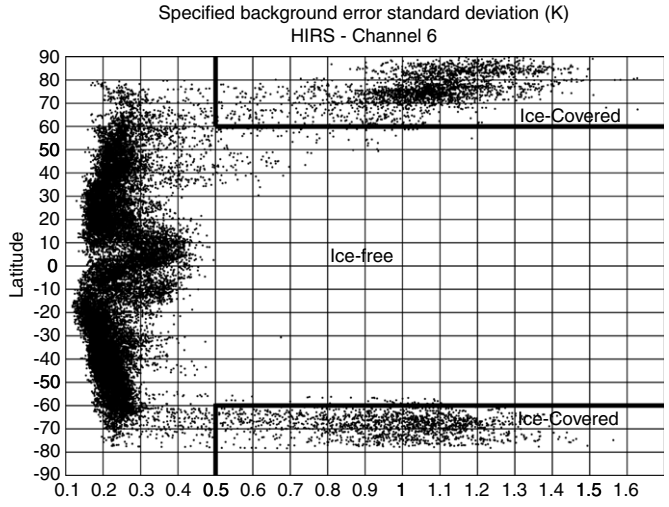


Fig. 1. Scatter plot of pairs: (specified background error standard deviation σ_{sb}, North latitude) for channel 6. Boxes delimit samples corresponding to 'Ice-free' and 'Ice-covered' conditions.

where $x_b = x + \varepsilon_b$ is the background estimate for x, and ε_b the background error. Under the commonly assumed hypothesis of additive noise, the innovation is the only combination of the data (observation and background) that is independent of the exact value x, and is as such the only objective source of information on the errors ε_o and ε_b. However, it clearly brings no information on either ε_o or ε_b individually.

3.1. Bias and variance

Rows 5 and 6 of Table 1 show the bias b_d and the standard deviation σ_d of the innovations, computed along each set. For all the 13 samples, the bias b_d is of the order of a few hundredths of a degree Kelvin, while the standard deviation σ_d is typically larger by one or two orders of magnitude. The biases are statistically significant, as shown by the fact that the ratio $\frac{|b_d|}{\sigma_d/\sqrt{N}}$ is significantly larger than 1, varying along the samples between 10 and 40.

In the ECMWF assimilation system, background and observation errors are assumed, as usually, to be mutually uncorrelated. The variance of the innovation should therefore be the sum of the variances of the two errors. This is not so in the present case. It can be seen from the Table 1 that the latter sum is larger than the innovation variance for all 13 subsets, by a large factor which varies between 1.84 (sample 11s) and as much as 27.72 (sample 7i). This is due in part to the assumed variance for observational error, which, with the only exception of sample 11s, is larger than the variance of the innovation. This corresponds to a purposeful inflation of the variance of the observational error in the ECMWF assimilation system. That inflation is meant to compensate for the fact that the spatial correlation of errors in satellite observations is ignored. For the 5 'ice-covered' samples, the specified background error variance exceeds largely the variance of the innovation.

3.2. Higher order statistical moments and non-Gaussianity

The skewness and kurtosis of innovations d, are defined as:

$$s_d \equiv \overline{(d'/\sigma_d)^3}, \tag{2a}$$

$$k_d \equiv \overline{(d'/\sigma_d)^4} - 3, \tag{2b}$$

where the bar and the prime stand respectively for the sampling mean and departure with respect to the mean. The skewness and kurtosis are respectively the third and fourth order *standardized* cumulants. The skewness can take all values in the interval $]-\infty, +\infty[$. When $s_d \neq 0$, the corresponding pdf is asymmetric with respect to its mean, with in general a more extended tail of extreme values on the side with the same sign as s_d. For a Gaussian pdf, $s_d = 0$ and $k_d = 0$, so that any statistically significant deviation from those values is a proof of non-Gaussianity. The probability distribution is said to be platykurtic (resp. leptokurtic) if the kurtosis k_d is negative (resp. positive). It thus has tails that are less populated (resp. more populated) than those of the Gaussian distribution with the same mean and variance. The kurtosis is very sensitive to the presence of outliers, which can significantly increase its value.

There is a lower bound for k_d as a function of s_d [17]. This is seen from the variable $r_d \equiv (d'/\sigma_d)^2 - s_d(d'/\sigma_d)$, which has sampling variance:

$$\eta_d \equiv \mathrm{var}\,(r_d) = k_d - s_d^2 + 2 \geq 0. \tag{3}$$

As a consequence, the kurtosis ranges in the interval $[-2, +\infty[$. The variable r_d is the second order polynomial of the *standardized* innovation $d^* \equiv d'/\sigma_d$, which has mean 1 and minimal variance.

On the other hand, we easily get any pair (s_d, k_d) satisfying (3) by using a pdf mixture of three weighted Dirac delta functions. The set of pairs (skewness, kurtosis) satisfying inequality (3) is hereafter called the SK-domain. The value of η_d for a Gaussian pdf equals 2, whereas it ranges in the 13 BT samples from 1.49 (sample 6i) to 3.99 (sample 12s).

There is another way of measuring non-Gaussianity through the concept of *relative entropy* [14]. The *relative entropy* of a pdf $\rho_1(x)$ with respect to another pdf $\rho_2(x)$, also called the Kullback–Leibler divergence between $\rho_1(x)$ and $\rho_2(x)$, is defined as:

$$KL(\rho_1, \rho_2) \equiv \int \rho_1(x) \ln \left[\frac{\rho_1(x)}{\rho_2(x)} \right] \mathrm{d}x. \tag{4}$$

It is equal to zero for $\rho_1 = \rho_2$ and is strictly positive for $\rho_1 \neq \rho_2$. It also possesses the fundamental property of being invariant in any one-to-one mapping $x \leftrightarrow \xi$. Despite not defining a distance in the sense of metric spaces (it is not symmetric), the relative entropy is a useful measure of the discrepancy between ρ_1 and ρ_2. The negentropy of a pdf ρ in the sense of information theory, is the relative entropy of ρ with respect to the Gaussian pdf ρ_G with the same mean and variance [14]. For small absolute values of skewness s_d and kurtosis k_d, the negentropy is approximated thanks to the Gram–Charlier expansion of ρ/ρ_G [18]:

$$KL(\rho, \rho_G) \approx \frac{1}{12} s_d^2 + \frac{1}{48} k_d^2 \equiv J_d. \tag{5}$$

The approximate negentropy J_d, expressed in terms of s_d and k_d, will be used as an integrated diagnostic of non-Gaussianity.

Rows 7 to 9 of Table 1 show the three quantities s_d, k_d and J_d for the 13 samples of innovations. For each sample, checks of deviation from Gaussianity have been performed on the basis of an ensemble of 1000 realizations, each with the same size N, of the Gaussian probability distribution with the same mean and variance of the sample. Values marked in bold in rows 7 to 9 of the Table 1 indicate non-Gaussianity at the corresponding 99% level of confidence. It is seen that only sample 14s can be considered Gaussian by those checks and 10 subsets out of 13 are non-Gaussian by all three checks. It can also be noted that the kurtosis is positive (leptokurtic) for the 8 'ice-free' samples. The 5 'ice-covered' samples are platykurtic (negative kurtosis) and negatively skewed. In order to show a graphical evidence of non-Gaussianity, we present in Fig. 2, the histogram of innovations for the sample 4s which has the largest absolute skewness $|s_d|$. In that figure, we superpose the Gaussian pdf fit and the non-Gaussian pdf fit obtained by the ME method, constrained by moments up to the fourth order (cf. Section 5). There is a clear difference between the two fits which shows that the sample is non-Gaussian.

It is interesting to elucidate the origin of those features as a result of the observations and background statistics, the quality control and other possible sources of non-Gaussianity. The elimination, through the quality control, of too large absolute values of innovations can lead to an extra reduction of innovation extremes and therefore leading to possible platykurtosis of innovations.

It is seen that there are significant inconsistencies between the *a priori* hypotheses that lie at the basis of the assimilation system on the one hand, and the observed statistics of the innovation on the other. Both the first and the second order statistical moments of the innovation are inconsistent with the *a priori* specified biases and variances of the observation and background errors. In addition, the innovation turns out to be significantly non-Gaussian, meaning that the assimilation system, even if optimal in the sense of the BLUE, does not achieve Bayesian estimation. These conclusions are not limited to the satellite observations considered above. Similar results (not shown) have been obtained in the ECMWF system with radiosonde observations. The only significant difference with the case of satellite observations is that the amplitude of the observational error does not seem to be overevaluated or at least not systematically.

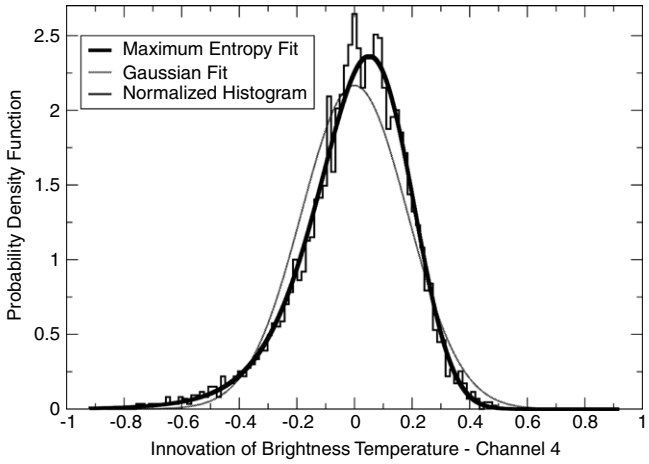


Fig. 2. Normalized histograms (in stairs), corresponding Gaussian pdf fit (grey solid line), and Maximum Entropy pdf fit (dark solid line) for the sample 4s of innovations.

4. Statistical consistency between innovations and errors

4.1. Mean and variance

The statistical moments of the innovations d can be expressed as combinations of single and joint moments of the errors ε_o, ε_b, leading to consistency relationships between the statistics of errors, which can be specified either by the assimilation scheme or through additional error hypotheses. The statistical moments can also be expressed in terms of cumulants [19,20], *i.e.* the Taylor coefficients of the cumulant generating function. The use of cumulants instead of moments will lead to consistency cumulant-relationships. In order to accomplish consistency, we need in general to perform appropriate tuning of the specified error statistics including in particular bias error correction and tuning of error variances [21,22].

Therefore, let us consider the statistically consistent biases b_{co} and b_{cb}, respectively of ε_o and ε_b, taken uniform over each sample. The subscript c in co and cb refers to statistical consistency. Then we introduce tuning factors f_o and f_b, respectively for observations and backgrounds, with both taken constant within each sample. These factors relate the consistent standard deviations $(\sigma_{co}, \sigma_{cb})$ of the unbiased errors ε'_o and ε'_b, to the specified standard deviations of errors $(\sigma_{so}, \sigma_{sb})$ as: $\sigma_{co} \equiv f_o \sigma_{so}$ and $\sigma_{cb} \equiv f_b \sigma_{sb}$. The values of σ_{sb} and σ_{cb} vary within samples meaning that background errors as well as innovations are heteroscedastic random variables. Heteroscedasticity is the property of a set of statistical distributions having different variances which can be either prescribed by hypothesis or made dependent in a certain way from external variables, (e.g. time in a stochastic process, geographical position). Therefore, the effect of the heteroscedasticity must be taken into account in the consistency relationships of the innovations. In order to isolate its effects, we have considered homoscedastic *i.e.* uniform variance variables as the *standardized* observation and background errors:

$$\varepsilon_o^* \equiv \frac{\varepsilon_o - b_{co}}{\sigma_{co}} \equiv \frac{\varepsilon'_o}{\sigma_{co}}, \tag{6a}$$

$$\varepsilon_b^* \equiv \frac{\varepsilon_b - b_{cb}}{\sigma_{cb}} \equiv \frac{\varepsilon'_b}{\sigma_{cb}}. \tag{6b}$$

We will make here the crucial hypothesis that the joint and marginal assumed pdfs of the *standardized* errors ε_o^* and ε_b^*, are independent of σ_{cb}, say independent from the observation pixel *i.e*, independent of the geographical position and time of the observation. This independence extends also, by hypothesis to

the biases b_{co}, b_{cb}, to the standard deviation of observation error σ_{co}, and to the tuning factors f_o, f_b, which, as already said, are considered constant within each data subset. As a consequence of the above hypothesis, the *standardized* errors ε_o^*, ε_b^*, are also independent of σ_{cb}. Using the above definitions and the hypothesis of additive noise, the innovations can be written as

$$d = \sigma_{co}\varepsilon_o^* - \sigma_{cb}\varepsilon_b^* + b_{co} - b_{cb}, \tag{7}$$

in which d is then written as a function of two random variables ε_o^*, ε_b^*, with varying tuned error standard deviations σ_{co}, and σ_{cb}. In the case of multiplicative errors, one simply needs to multiply those standard deviations by certain weighting x-dependent functions.

Under the hypothesis of additive errors, the average, the centered moments and the cumulants of d are then factorized in terms of assumed single and joint *standardized* error moments, as well as in terms of power averages of σ_{cb}. Therefore,the innovation bias reads as:

$$\overline{d} \equiv b_d = \sigma_{co}E\left[\varepsilon_o^*\right] - \overline{\sigma_{cb}}E\left[\varepsilon_b^*\right] + b_{co} - b_{cb} = b_{co} - b_{cb}, \tag{8}$$

where E is the expectation operator taken over the *standardized* errors and the bar operator runs through the sample. From that expression, we get the departure $d' = \sigma_{co}\varepsilon_o^* - \sigma_{cb}\varepsilon_b^*$, from which centered moments of the innovation are computed. Taking into account that $E[\varepsilon_b^{*2}] = E[\varepsilon_o^{*2}] = 1$, by definition, one writes the variance of innovation as

$$\overline{d'^2} \equiv \sigma_d^2 = \sigma_{co}^2 + \overline{\sigma_{cb}^2} - 2\sigma_{co}\overline{\sigma_{cb}}c_{ob}, \tag{9}$$

where $c_{ob} = \mathrm{E}[\varepsilon_o^*\varepsilon_b^*]$ is the observational-background error correlation. Under the usual hypothesis of uncorrelated errors, the innovation variance is simply the sum of the mean error variances:

$$\sigma_o^2 \equiv \sigma_{co}^2 = f_o^2\sigma_{so}^2 \equiv \beta_o\sigma_d^2, \tag{10a}$$

$$\sigma_b^2 \equiv \overline{\sigma_{cb}^2} = f_b^2\overline{\sigma_{sb}^2} \equiv \beta_b\sigma_d^2 \tag{10b}$$

where β_o and β_b are the explained variance fractions of the innovation variance. Uncorrelated errors lead to: $\beta_o + \beta_b = 1$.

4.2. Skewness and kurtosis

The formulae of the skewness s_d and kurtosis k_d of innovations are obtained after expanding the third and fourth order powers of d' and then averaging it. For that purpose we need to define the single error skewnesses:

$$s_{co} \equiv \mathrm{E}\left[\varepsilon_o^{*3}\right], \tag{11a}$$

$$s_{cb} \equiv \mathrm{E}\left[\varepsilon_b^{*3}\right], \tag{11b}$$

the single error kurtoses:

$$k_{co} \equiv \mathrm{E}\left[\varepsilon_o^{*4}\right] - 3, \tag{12a}$$

$$k_{cb} \equiv E\left[\varepsilon_b^{*4}\right] - 3, \tag{12b}$$

the joint *standardized* cumulants of order 3:

$$\kappa_{oob} \equiv \mathrm{E}\left[\varepsilon_o^*\varepsilon_o^*\varepsilon_b^*\right], \tag{13a}$$

$$\kappa_{bbo} \equiv \mathrm{E}\left[\varepsilon_b^*\varepsilon_b^*\varepsilon_o^*\right] \tag{13b}$$

and the joint *standardized* error cumulants of order 4:

$$\kappa_{ooob} \equiv \mathrm{E}\left[\varepsilon_o^*\varepsilon_o^*\varepsilon_o^*\varepsilon_b^*\right] - 3\mathrm{E}\left[\varepsilon_o^*\varepsilon_b^*\right], \tag{14a}$$

$$\kappa_{oobb} \equiv \mathrm{E}\left[\varepsilon_o^*\varepsilon_o^*\varepsilon_b^*\varepsilon_b^*\right] - 2\mathrm{E}\left[\varepsilon_o^*\varepsilon_b^*\right]^2 - 1, \tag{14b}$$

$$\kappa_{bbbo} \equiv \mathrm{E}\left[\varepsilon_b^*\varepsilon_b^*\varepsilon_b^*\varepsilon_o^*\right] - 3\mathrm{E}\left[\varepsilon_o^*\varepsilon_b^*\right]. \tag{14c}$$

If observation and background errors are assumed bivariate Gaussian, then all the above, single and joint cumulants will vanish. If errors are independent, then only the joint cumulants vanish. The joint cumulants appear also in correlations between nonlinear functions of errors.

The expression for s_d and k_d depends also on the average of powers of the tuned background error standard deviation σ_{cb}. We express those averages in a non-dimensional form as:

$$h_{nb} \equiv \frac{\overline{\sigma_{cb}^n}}{\left(\overline{\sigma_{cb}^2}\right)^{n/2}} = \frac{\overline{\sigma_{sb}^n}}{\left(\overline{\sigma_{sb}^2}\right)^{n/2}}, \quad n = 1, 2, 3, 4. \tag{15}$$

The h-terms h_{nb} are measures of the heteroscedasticity of background errors. Since the tuning factor f_b is assumed uniform, those terms are simply obtained from statistics of σ_{sb}. The h-terms are equal to 1 if the specified σ_{sb} does not vary along the sample, *i.e.* in homoscedastic conditions. By definition, $h_{2b} = 1$. In addition the inequalities: $h_{1b} \leq 1$, $h_{3b} \geq 1$ and $h_{4b} \geq 1$, hold by virtue of the Jensen inequality [23], applied to the convex power functions of the type: $g(u) = u^p$ ($u = \sigma_{cb}^2$, $p \geq 1$). Table 1 lists the h-terms: h_{1b}, h_{3b} and h_{4b} for the 13 samples. The h_{3b} terms in Table 1 range in the interval [1.03, 1.29] from which one sees a more marked heteroscedasticity in samples 7s, 11s, 12s, 14s, and 15s.

The absolute deviations of the h-terms h_{1b}, h_{3b} and h_{4b} from unity, are positively correlated with each other and vary in a consistent way among the 13 samples.

The skewness s_d of innovations then reads:

$$s_d = \left\{\beta_o^{3/2}s_{co} - \beta_b^{3/2}h_{3b}s_{cb}\right\} + \left\{3\beta_b\beta_o^{1/2}\kappa_{bbo} - 3\beta_o\beta_b^{1/2}h_{1b}\kappa_{oob}\right\}, \tag{16}$$

in terms of the skewnesses of errors s_{co}, s_{cb}, and of the third order joint error cumulants, weighted by powers of the variance fractions β_o and β_b and heteroscedasticity h-terms. Since there are no specified error skewnesses by the assimilation system, they remain unknown. In accordance with innovation definition, observation and background errors skew the innovations, towards respectively the same side and the opposite side. Innovations can be skewed even when errors are uncorrelated ($c_{ob} = 0$) and have Gaussian marginal pdfs with ($s_{co} = s_{cb} = 0$). That is due to nonlinear error correlations brought by third order joint error cumulants. Dividing Eq. (9) by σ_d^2, and using the definitions of β_o, β_b from (10a), (10b) one gets:

$$\beta_o + \beta_b - 2c_{ob}h_{1b}\beta_o^{1/2}\beta_b^{1/2} = 1. \tag{17}$$

Playing with the above equation, the kurtosis k_d of the innovation comes as:

$$k_d = k_{d1} + k_{d2} + k_{d3}, \tag{18a}$$

$$k_{d1} = \beta_o^2k_{co} + \beta_b^2h_{4b}k_{cb}, \tag{18b}$$

$$k_{d2} = 6\beta_o\beta_b\kappa_{oobb} - 4\beta_o^{3/2}\beta_b^{1/2}h_{1b}\kappa_{ooob} - 4\beta_o^{1/2}\beta_b^{3/2}h_{3b}\kappa_{bbbo}, \tag{18c}$$

$$k_{d3} = 3\beta_b^2\left(h_{4b} - 1\right) + 12c_{ob}\beta_o^{1/2}\beta_b^{1/2}\left[\beta_b\left(h_{1b} - h_{3b}\right) - \beta_o^{1/2}\beta_b^{1/2}c_{ob}\left(h_{1b} - 1\right)^2\right]. \tag{18d}$$

The term within k_{d1} in Eq. (18a) is the contribution from the kurtosis of individual errors. The second term k_{d2} is uniquely due to the fourth order joint cumulants and finally the third term k_{d3} is due to heteroscedasticity. The kurtosis k_d can be completely attributed to one of the above terms. This means that a non-vanishing k_d can occur for non-independent, uncorrelated errors with Gaussian marginal pdfs ($k_{co} = k_{cb} = c_{ob} = 0$). On the other hand, bivariate Gaussian independent errors in a

sample of heteroscedastic background errors produce a symmetric leptokurtic innovation pdf with $s_d = 0$ and $k_d = 3\beta_b^2(h_{4b} - 1) > 0$. This is well understood because the mixture of Gaussian random variables of different variances is non-Gaussian. For the case of multiplicative noises, all the errors are multiplied by x-dependent factors as it happens for heteroscedastic errors. Then both effects work similarly in a combined form onto the innovation moments.

4.3. Constraints under independent errors

We discuss here some necessary consequences, relative in particular to skewnesses and kurtoses, of the hypothesis that backgroung and observation errors are statistically independent under the framework of additive noises. The first property comes from the Cramèr theorem which states that a necessary and sufficient condition for a sum of independent variables to be Gaussian is that those variables are themselves Gaussian [24]. Therefore, if the innovation, taken as a difference of independent errors, is non-Gaussian, then at least one of the errors will be non-Gaussian. The averaged skewness s_b and averaged kurtosis k_b of the background error over a heteroscedastic sample are respectively given by:

$$s_b = h_{3b} s_{cb}, \tag{19a}$$

$$k_b = h_{4b} k_{cb} + 3\,(h_{4b} - 1)\,. \tag{19b}$$

The observational errors being homoscedastic, we obtain directly their mean skewness and kurtosis:

$$s_o = s_{co}, \tag{20a}$$

$$k_o = k_{co}. \tag{20b}$$

When errors are assumed statistically independent, the cross-cumulants vanish which allows writing the averaged skewness s_d and the averaged kurtosis k_d as linear combinations, respectively of the error skewnesses: (s_o, s_b) and of the error kurtosis (k_o, k_b):

$$s_d = \beta_o^{3/2} s_o - \beta_b^{3/2} s_b, \tag{21a}$$

$$k_d = \beta_o^2 k_o + \beta_b^2 k_b. \tag{21b}$$

The corresponding weights are proportional to powers of error variance fractions β_o and β_b suggesting that non-Gaussianity of innovations tends to be explained by the non-Gaussian error with the largest variance. Under independent errors, the value of η_d in (3) has a positive lower bound resulting from the positiveness of the following variance:

$$\begin{aligned} \operatorname{var}\left(r_d + \alpha \frac{\varepsilon_o' \varepsilon_b'}{\sigma_d^2}\right) &= \eta_d + \beta_o \beta_b \alpha\,(\alpha - 4) \\ &\geq \eta_d - 4\beta_o\beta_b \geq 0, \quad \forall \alpha, \end{aligned} \tag{22}$$

with r_d meaning the same as in Eq. (3). This defines a necessary condition for error independence for fixed variance fractions of errors. In all the 13 analysed samples, $\eta_d \geq 1$, which does not impose bounds on β_o, β_b since $\max(4\beta_o\beta_b) = 1$.

Eqs. (21a), (21b) can be solved for s_o and k_o, or s_b and k_b. Then applying inequalities of type (3) to the observation and background errors, *i.e.* $k_o - s_o^2 + 2 \geq 0$ and $k_b - s_b^2 + 2 \geq 0$, one gets upper and lower bounds for the kurtoses of the errors as follows:

$$s_b^2 - 2 \leq k_b \leq \beta_b^{-2}\left(k_d + 2\beta_o^2\right) - \beta_b\beta_o^{-1}\left(s_b + \beta_b^{-3/2} s_d\right)^2, \tag{23a}$$

$$s_o^2 - 2 \leq k_o \leq \beta_o^{-2}\left(k_d + 2\beta_b^2\right) - \beta_o\beta_b^{-1}\left(s_o - \beta_o^{-3/2} s_d\right)^2. \tag{23b}$$

The upper limit of k_b in (23a) is obtained by expanding the inequality $-\beta_b^{-2}\beta_o^2\left(k_o - s_o^2 + 2\right) \leq 0$ in which we substitute $k_o = \beta_o^{-2}\left(k_d - \beta_b^2 k_b\right)$ and $s_o = \beta_b^{3/2}\beta_o^{-3/2}\left(s_b + \beta_b^{-3/2} s_d\right)$ from (21a) and (21b). An analogous manipulation, starting by solving (21a), (21b) for s_b and k_b, leads to the upper bound in (23b). This is obtained from the inequality $-\beta_o^{-2}\beta_b^2\left(k_b - s_b^2 + 2\right) \leq 0$ with $k_b = \beta_b^{-2}\left(k_d - \beta_o^2 k_o\right)$ and $s_b = \beta_o^{3/2}\beta_b^{-3/2}\left(s_o - \beta_o^{-3/2} s_d\right)$.

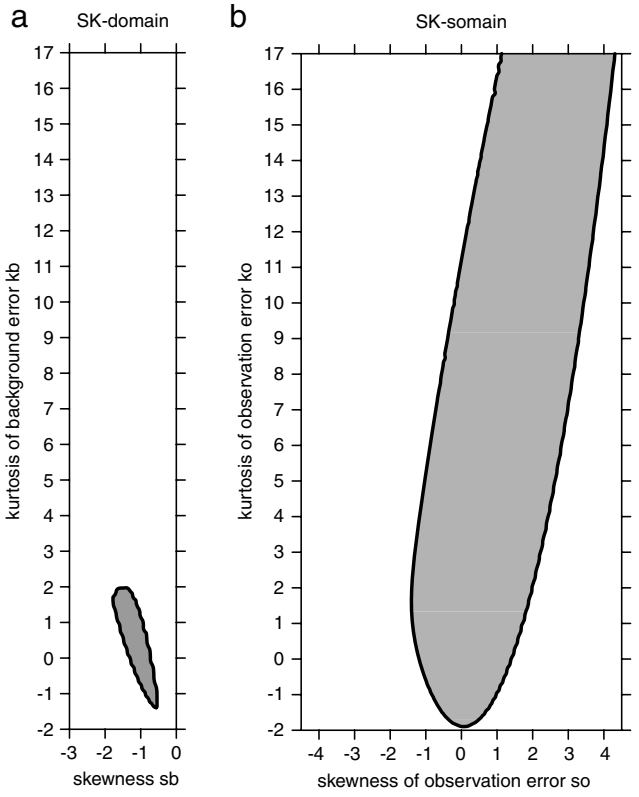


Fig. 3. Sets in the SK-domain of errors which are compatible with the condition $s_d = 1$, $k_d = 1$ and error variance fractions: $\beta_o = 0.25$, $\beta_b = 0.75$, respectively for background errors (a) and observation errors (b). Sets are marked in grey.

Inequalities (23a) and (23b), for given values of β_o, β_b, s_d and k_d, define each, two parabolas bounding in turn admissible sub-domains for the couples (s_o, k_o) and (s_b, k_b) in their respective SK domains. It is easily verified that inequality (22) defines a necessary and a sufficient conditions for those sub domains not to be void.

The largest of the two sets of pairs (s_o, k_o) and (s_b, k_b) corresponds to the error with smallest variance fraction (β_o or β_b). This is illustrated in the example of Fig. 3 which shows, for $\beta_o = 0.25$, $\beta_b = 0.75$, $s_d = 1$ and $k_d = 1$ the admissible sub domains for (s_b, k_b) and (s_o, k_o) (panels (a) and (b) respectively). It is seen in particular that the point $(s_b = 0, k_b = 0)$ lies outside of the admissible sub domain in panel 3(a). This suggests that, when the innovation is not Gaussian, the errors with largest variances tend to be forbidden to be Gaussian. When heteroscedasticity is present, the skewnesses and kurtoses must be substituted by (19a), (19b), (20a), (20b) in the inequalities (23a), (23b).

We have analysed the conditions for the full attribution of non-Gaussianity to one of the errors, keeping the remaining one to be Gaussian. The necessary condition for the existence of a single non-Gaussian error of variance β is given by the third order polynomial inequality:

$$R(\beta) \equiv \beta^3 + \beta\frac{k_d}{2} - \frac{s_d^2}{2} \geq 0; \quad \beta \in [0, 1]\,. \tag{24}$$

Let us prove (24), by considereing for example that the background error is non-Gaussian and the observation error is Gaussian *i.e.* $k_o = s_o = 0$. Then we have $\beta = \beta_b$, $k_b = \beta^{-2} k_d$ and $s_b = -\beta^{-3/2} s_d$. Introducing those values into $\frac{\beta^3}{2}\left(k_b - s_b^2 + 2\right) \geq 0$

leads directly to (24). The variance fraction of the Gaussian error is $\beta' = 1 - \beta$ which satisfies to an equivalent inequality to (24):

$$P(\beta') \equiv \beta'^3 - 3\beta'^2 + \beta'\left(3 + \frac{k_d}{2}\right) - \frac{\eta_d}{2} = -R(\beta) \leq 0; \quad \beta' \in [0, 1]. \tag{25}$$

The polynomial $P(\beta')$ has a single root β_g in the interval $[0, 1]$ which is a function of (s_d, k_d). Since $P(0) = -\eta_d/2 \leq 0$ and $P(1) = s_d^2 \geq 0$, one concludes that the maximum variance fraction of the Gaussian error is $\beta_g(s_d, k_d)$ and the required minimum variance of the non-Gaussian error is $1 - \beta_g(s_d, k_d)$. In other words, (25) is equivalent to: $\beta' \leq \beta_g(s_d, k_d)$ or $\beta \geq 1 - \beta_g(s_d, k_d)$. The contour field of $\beta_g(s_d, k_d)$ appears in Fig. 4. The value of $\beta_g(s_d, k_d)$ is symmetrical with respect to s_d and decreases when the limiting curve $\eta_d = 0$ is approached thus requiring increasing variance to the non-Gaussian error with the variance fraction $1 - \beta_g$. In practice, comparison of β_g with specified consistent values of β_o and β_b makes it possible to state whether non-Gaussianity of innovations can originate exclusively from one of the errors (scenario of single source of non-Gaussianity) or whether on the contrary it must originate from both observation and background errors (scenario of multiple source of non-Gaussianity). These two scenarios will be explored farther in the paper. The reverse condition of (25) states that if β' is the variance fraction of a certain error and $P(\beta') > 0$, then either this error is non-Gaussian or the errors are Gaussian but not independent.

The error fractions β_o, β_b depend on tuning factors f_o, f_b, according to (10a) and (10b). This means that the partition of the non-Gaussianity of innovations among the errors is not independent from the problem of tuning of the error variances. Bounds are therefore imposed on tuning factors in the case of single attribution of non-Gaussianity and independent errors. Let us consider the quotient between the specified error variances and the innovation variance

$$\beta_{so} = \frac{\sigma_{so}^2}{\sigma_d^2}, \qquad \beta_{sb} = \frac{\overline{\sigma_{sb}^2}}{\sigma_d^2} = \frac{m(\sigma_{sb})^2 + \mathrm{std}(\sigma_{sb})^2}{\sigma_d^2}. \tag{26}$$

Then, the necessary conditions for Gaussian observation errors or equivalently the necessary condition for non-Gaussianity coming exclusively from background errors is:

$$f_o \leq \left(\beta_g/\beta_{so}\right)^{1/2} \equiv f_{o1}, \tag{27a}$$

$$\beta_{sb}^{-1/2} \geq f_b \geq \left[\left(1 - \beta_g\right)/\beta_{sb}\right]^{1/2} \equiv f_{b2}. \tag{27b}$$

On the other hand, the necessary conditions for Gaussian background errors or equivalently the necessary condition for non-Gaussianity coming exclusively from observation errors is:

$$f_b \leq \left(\beta_g/\beta_{sb}\right)^{1/2} \equiv f_{b1}, \tag{28a}$$

$$\beta_{so}^{-1/2} \geq f_o \geq \left[\left(1 - \beta_g\right)/\beta_{so}\right]^{1/2} \equiv f_{o2}. \tag{28b}$$

The values of β_g and the limiting tuning factors $f_{o1}, f_{o2,}, f_{b1}, f_{b2}$ are listed in Table 1 for the 13 studied samples. For example, under the hypothesis of Gaussian observation errors, many samples require a maximum tuning f_{o1} of the order of 0.5. This accounts for the need of a strong deflation of the specified variance of the observation error. We also notice the necessary inflation of background error variance for sample 4s at which $f_{b2} = 1.305$. Under the hypothesis of Gaussian background errors, we notice in particular the need for a strong deflation of the specified background error variances in the 5 'ice-covered' samples, as it is corroborated by the values f_{b1} lying in the interval [0.16, 0.54].

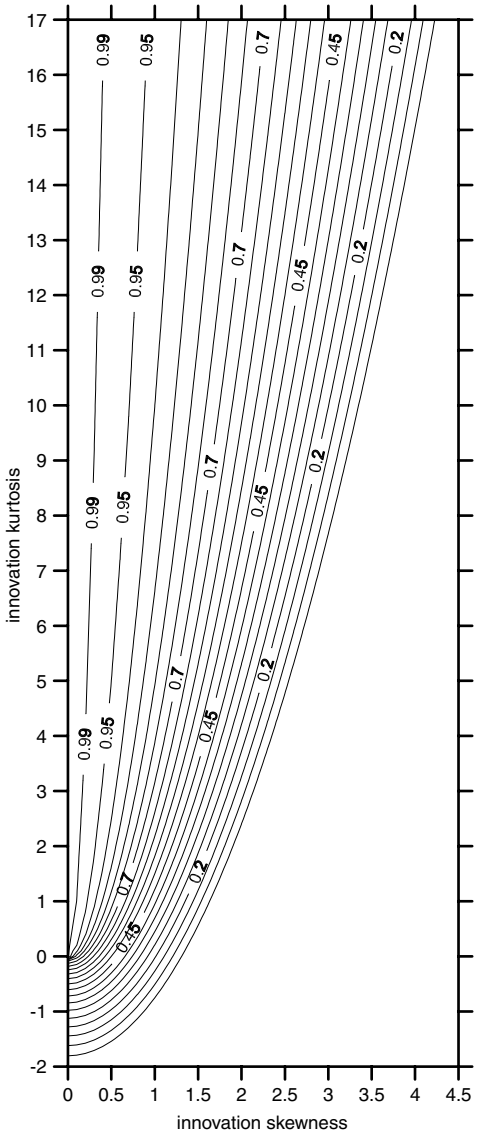


Fig. 4. Contour map of the maximum variance fraction β_g of a Gaussian error, as function of the innovation skewness s_d and kurtosis k_d. The contour map is symmetric for $s_d < 0$.

5. Error pdfs obtained by the maximum entropy method

5.1. General formulation

The motivation of this section is to build non-Gaussian univariate pdfs of the errors in the observational space which are compatible with prescribed values of the bias, variance, skewness and kurtosis, both of the errors and of the innovations. The interest of formulating a probabilistic model or a pdf of the errors is the possibility of computing the Bayesian estimate of the projected system state from the pair (y, x_b) and to compare it with the basic BLUE emanating from the same *a priori* data. This, as shown in the next section will give an indication of the impact of taking into account the non-Gaussianity of errors in the posterior estimation. However, there is a large range of possible error pdfs satisfying the imposed moments. In order to resolve that undeterminacy, we have used the Maximum Entropy (ME) method. The least committing method of finding the joint pdf $\rho_{ob}(\varepsilon_o, \varepsilon_b)$ of the errors $(\varepsilon_o, \varepsilon_b)$ is the Maximum Entropy method [12–14], which looks for the maximum of the differential joint Shannon entropy [25] of the errors $H(\varepsilon_o, \varepsilon_b)$, subject to a certain set of constraints in terms of

expectations. Applications of the ME method in data assimilation and inversion problems can also be found in [26–29].

Now, let us explain how to formulate the ME problem in our context. A given sample of quadruplets $(y, x_b, \sigma_{co}, \sigma_{cb})$, of possible varying error standard deviations, *i.e*, subjected to heteroscedasticity, as it is case for BTs, corresponds to a sample of error pairs $(\varepsilon_o, \varepsilon_b)$. Then we assume that all the error pairs are independent from each other. In the case of BT samples, this is a reasonable hypothesis due the spatial and temporal data thinning, thus minimizing the problem of spatial and temporal error correlations. Therefore, the sampling mean Shannon entropy is written as:

$$\begin{aligned}\overline{H(\varepsilon_o, \varepsilon_b)} &\equiv \overline{\mathrm{E}\left[-\ln \rho_{ob}(\varepsilon_o, \varepsilon_b)\right]} \\ &= \overline{\ln(\sigma_{co})} + \overline{\ln(\sigma_{cb})} + H\left(\varepsilon_o^*, \varepsilon_b^*\right). \end{aligned} \tag{29}$$

where the bar is the sample mean as previously defined and $H(\varepsilon_o^*, \varepsilon_b^*)$ is the Shannon entropy of the *standardized* errors. In view of the assumed independence of errors, the joint entropy reduces to the sum of individual error entropies:

$$\begin{aligned} H\left(\varepsilon_o^*, \varepsilon_b^*\right) &= H\left(\varepsilon_o^*\right) + H\left(\varepsilon_b^*\right) \\ &= E\left[-\ln\left(\varepsilon_b^*\right)\right] + E\left[-\ln\left(\varepsilon_o^*\right)\right], \end{aligned} \tag{30}$$

where $\rho_{o*}\left(\varepsilon_o^*\right)$, and $\rho_{b*}\left(\varepsilon_b^*\right)$ are the pdfs of ε_o^* and ε_b^* respectively.

The problem of entropy maximization reduces to the search of $H_M(\varepsilon_o^*, \varepsilon_b^*) \equiv \max[H(\varepsilon_o^*, \varepsilon_b^*)]$, over the pdfs satisfying a set Ω of N_c strong constraints in terms of expectations of functions $g_1, \ldots, g_{Nc}$ of the *standardized* errors:

$$\Omega = \left\{\mathrm{E}\left[g_k\left(\varepsilon_o^*, \varepsilon_b^*\right)\right] = c_k;\ k = 1, \ldots, N_c\right\}. \tag{31}$$

In our case, the constraints correspond to moments, both of the errors and of the innovations. Due to error independence, the constraint functions are additive separable functions of the background error and of the observational error in the form of monomials. Then we write:

$$g_k\left(\varepsilon_o^*, \varepsilon_b^*\right) = g_{k,o}\left(\varepsilon_o^*\right) + g_{k,b}\left(\varepsilon_b^*\right). \tag{32}$$

For a well-posed ME problem, the maximum entropy H_M is finite and it is found at the global minimum of a globally convex function Γ, of N_c Lagrange multipliers (LMs) $\lambda_1, \ldots, \lambda_{Nc}$, one for each constraint, and which reads:

$$\begin{aligned}\Gamma\left(\lambda_1 \cdots \lambda_{N_c}\right) \equiv{}& \ln \int \exp\left(\sum_{k=1}^{N_c}\left(\lambda_k g_{k,o}\left(\varepsilon_o^*\right)\right)\right) \mathrm{d}\varepsilon_o^* \\ &+ \ln \int \exp\left(\sum_{k=1}^{N_c}\left(\lambda_k g_{k,b}\left(\varepsilon_b^*\right)\right)\right) \mathrm{d}\varepsilon_b^* \\ &- \sum_{k=1}^{N_c} \lambda_k c_k. \end{aligned} \tag{33}$$

The LM values at the minimum of Γ are respectively: $\hat{\lambda}_1, \ldots, \hat{\lambda}_{Nc}$. In order to assure the convergence of integrals and the well-posedness of the ME problem, the largest order of the imposed statistical moments must be an even number if the pdf support is $]-\infty, +\infty[$. This condition holds in our case since the maximum moment order is four, corresponding to kurtosis.

Finally, the individual Maximum Entropy pdfs of the *standardized* errors, ME-pdfs for short, belong to the exponential family [30] and reads:

$$\rho_{t^*}\left(\varepsilon_t^*\right) \equiv \frac{\exp\left(\sum_{k=1}^{N_c}\left(\hat{\lambda}_k g_{k,t}\left(\varepsilon_t^*\right)\right)\right)}{\int \exp\left(\sum_{k=1}^{N_c}\left(\hat{\lambda}_k g_{k,t}(u)\right)\right) \mathrm{d}u}, \quad t = o, b. \tag{34}$$

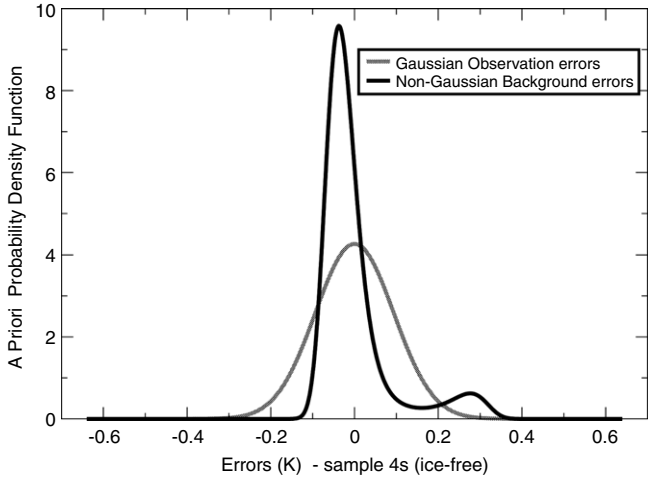


Fig. 5. *A priori* pdf of Gaussian observation errors (grey line) and non-Gaussian background errors (dark line), for sample 4s, found by the Maximum Entropy method, fixing the error variance partition: $\beta_b = 0.47$, $\beta_o = 0.53$ and $s_b = 2.18$, $k_b = 4.65$.

In general, the lower the number of constraints N_c, the larger is the value of H_M. When the set Ω includes only error moments of order one and two, then the ME-pdf is Gaussian. For the standard Gaussian the entropy reaches the value $H_G = 0.5\ln(2\pi e) \approx 1.41$ nats where nat is the natural logarithmic unit of Shannon entropy and information with the conversion: 1nat $= \log_2(e)$ bit $\approx$ 1.44 bit.

5.2. Maximum entropy errors constrained by innovation statistics

Two kinds of maximum entropy problems are formulated bellow. The first one is the so-called problem of 'prescribed error non-Gaussianity'. In the first problem, each observation and background *standardized* errors satisfy the $N_c = 4$ constraints:

$$\begin{aligned}&\mathrm{E}\left[\varepsilon_t^*\right] = 0; \qquad \mathrm{E}\left[(\varepsilon_t^*)^2\right] = 1; \qquad \mathrm{E}\left[(\varepsilon_t^*)^3\right] = s_{ct}; \\ &\mathrm{E}\left[(\varepsilon_t^*)^4\right] = k_{ct} + 3; \quad t = o, b, \end{aligned} \tag{35}$$

with $s_{co}, s_{cb}, k_{co}, k_{cb}$, being defined by Eqs. (11a), (11b), (12a), (12b), and satisfying innovation constraints expressed by Eqs. (16), (18a). Then, two ME problems are solved, one for the background error and another for the observation error. A particular case is that in which all non-Gaussianity is fully attributed to one of the errors *i.e.* the case of single source of non-Gaussianity. The ME-pdf is obtained from Eq. (34) after applying the corresponding LMs in the exponential and computing the integral giving the normalization factor.

We now present an example of non-Gaussian ME-pdf. This is derived from the most strongly skewed sample of innovations (4s), where $s_d = -0.70$, $k_d = 1.02$ and the heteroscedasticity is quite small. The corresponding ME-pdf fit of the innovations is shown in Fig. 2. The chosen error variance partition is: $\beta_b = 0.47$, $\beta_o = 0.53$, corresponding to tuning factors $f_b = 1.50$ and $f_o = 0.22$. Then we assume that non-Gaussianity is coming exclusively from background errors, thus leading to $s_b = 2.18$, $k_b = 4.65$, $s_o = 0$, $k_o = 0$. The non-Gaussian ME-pdf of the unbiased tuned background error, shown in Fig. 5, presents a clear positively skewed tail, bimodality and an enhanced maximum near the origin due to leptokurtosis. The Gaussian pdf of the observation error is also shown.

The (s, k) or (skewness, kurtosis) dependence of the maximum entropy H_M and of the four Lagrange multipliers λ_p $(p = 1, \ldots, 4)$, corresponding to the powers p in Eq. (35) is discussed by [17] as well as in the Appendix.

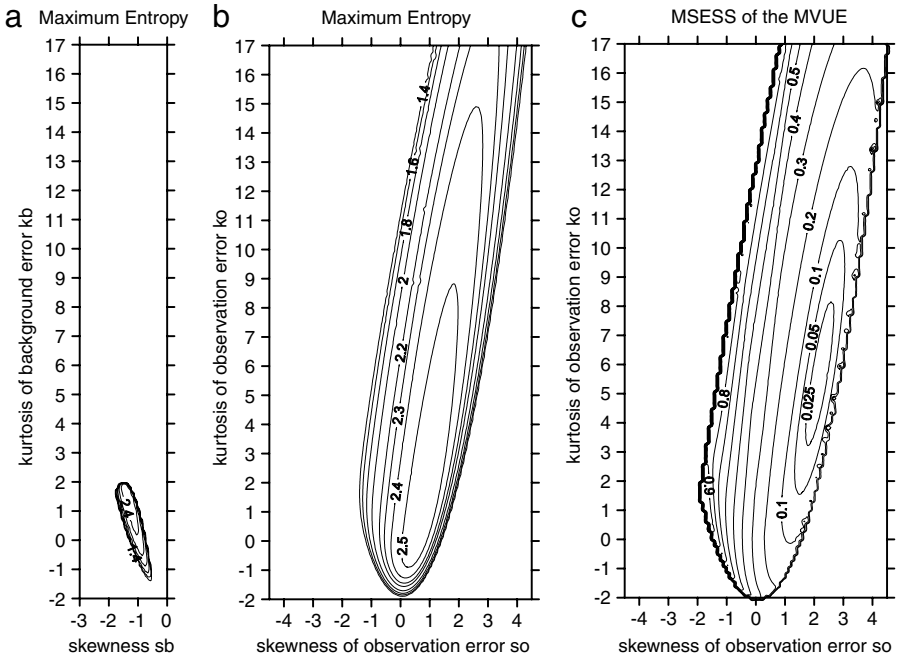


Fig. 6. (a) Contour field of the joint maximum entropy (in nats) of background and observation errors for the allowable domain set in the (s_b, k_b) plane; (b) same as (a) in the (s_o, k_o) plane; (c) MSESS of the MVUE in the (s_o, k_o) plane. In all experiments we set $s_d = 1$, $k_d = 1$ and error variance fractions: $\beta_o = 0.25$, $\beta_b = 0.75$.

The second kind of ME problem, hereafter called 'Maximum entropy sharing of non-Gaussianity', correspond to the case in which the skewnesses and kurtoses of errors are not anchored to prescribed values, though they still satisfy the innovation constraints given by Eqs. (21a), (21b). In this case we get a total of $N_c = 6$ constraints terms in the *standardized* errors:

$$\mathrm{E}\left[\varepsilon_t^*\right] = 0; \qquad \mathrm{E}\left[(\varepsilon_t^*)^2\right] = 1, \quad \text{with } t = o, b, \tag{36a}$$

$$\mathrm{E}\left[\varepsilon_o^{*3}\beta_o^{3/2} - \varepsilon_b^{*3}h_{3b}\beta_b^{3/2}\right] = s_d, \tag{36b}$$

$$\mathrm{E}\left[\varepsilon_o^{*4}\beta_o^2 + \varepsilon_b^{*4}h_{4b}\beta_b^2\right] = k_d + 3\left(\beta_o^2 + \beta_b^2\right), \tag{36c}$$

to which correspond 6 Lagrange multipliers: λ_{o1}, λ_{o2}, λ_{b1}, λ_{b2}, λ_{d1}, λ_{d2} with the first and second pair of LMs, corresponding to the mean and variance constraints, respectively for the observation and background errors. The last two LMs correspond to constraints related to s_d and k_d respectively. The function Γ to be minimized reads:

$$\begin{aligned}
&\Gamma\left(\lambda_{o1}, \lambda_{o2}, \lambda_{b1}, \lambda_{b2}, \lambda_{d3}, \lambda_{d4}\right) \\
&= \sum_{t=b,o} \ln\left[\int \exp\left(\lambda_{t1}\varepsilon + \lambda_{t2}\varepsilon^2\right) \mathrm{d}\varepsilon\right] \\
&\quad + \ln\left\{\int \exp\left[\lambda_{d3}\beta_o^{3/2}\varepsilon^3 + \lambda_{d4}\beta_o^2\varepsilon^4\right] \mathrm{d}\varepsilon\right\} \\
&\quad + \ln\left\{\int \exp\left[-\lambda_{d3}\beta_b^{3/2}h_{3b}\varepsilon^3 + \lambda_{d4}\beta_b^2 h_{4b}\varepsilon^4\right] \mathrm{d}\varepsilon\right\} \\
&\quad - \left\{\lambda_{b2} + \lambda_{o2} + s_d\lambda_{d3} + \lambda_{d4}\left[k_d + 3\left(\beta_o^2 + \beta_b^2\right)\right]\right\}.
\end{aligned} \tag{37}$$

The corresponding ME-pdfs of the *standardized* errors are written in the condensed equation:

$$\rho_{t^*}(\varepsilon) = \frac{\exp\left(Q(\varepsilon)\right)}{\int \exp\left(Q(u)\right) \mathrm{d}u} \quad \text{with} \tag{38a}$$

$$Q(\varepsilon) = \hat{\lambda}_{t1}\varepsilon + \hat{\lambda}_{t2}\varepsilon^2 + \hat{\lambda}_{d4}\beta_t^2 h_{4t}\varepsilon^4 \pm \hat{\lambda}_{d3}\beta_t^{3/2}h_{3t}\varepsilon^3. \tag{38b}$$

Positive and negative signs in Eq. (38b) apply respectively to the observation error ($t = o$) and to the background error ($t = b$) while using $h_{3o} = h_{4o} = 1$ in the equation of the pdf of ε_o^*. The LMs are those minimizing the function in Eq. (37).

For background homoscedastic errors, we have: $h_{3b} = h_{4b} = 1$ and $(s_b, k_b) = (s_{cb}, k_{cb})$. In this particular case, the ME solution is uniquely defined by the triplet (s_d, k_d, β_o). As consequence of the anti-symmetric roles of errors, the corresponding skewnesses s_o and s_b are anti-symmetrical with respect to s_d and invariant in the mapping: $(s_o, \beta_o) \leftrightarrow (-s_b, 1 - \beta_o)$. Furthermore, k_o and k_b are symmetrical with respect to s_d and invariant for the mapping: $(k_o, \beta_o) \leftrightarrow (k_b, 1 - \beta_o)$. The pairs (s_o, s_b) and (k_o, k_b) are approximately given by minimizing the approximate error statistical *negentropies*, defined as in Eq. (5) and subject to constraints given by Eqs. (21a), (21b), thus leading to:

$$(s_o, s_b) \approx \frac{s_d}{\beta_o^3 + \beta_b^3}\left(\beta_o^{3/2}, -\beta_b^{3/2}\right), \tag{39a}$$

$$(k_o, k_b) \approx \frac{k_d}{\beta_o^4 + \beta_b^4}\left(\beta_o^2, \beta_b^2\right). \tag{39b}$$

As a consequence, when one of the errors is dominant, *i.e.* $\max(\beta_o, \beta_b)$ is close to unity, the non-Gaussianity of innovations through (s_d, k_d) tends to be explained by the error with the largest variance, whereas the remaining one is close to Gaussianity.

We will refer again to the example of Fig. 3(a), (b), corresponding to $(s_d = 1, k_d = 1)$ and $\beta_o = 0.25$, $\beta_b = 0.75$, in order to show where the ME solution lies in the allowable SK domains of the observation and background errors. The contours of the sum of maximum individual entropies: $H_M = H_M(\varepsilon_o^*) + H_M(\varepsilon_b^*)$, are drawn as function, both of (s_b, k_b) (Fig. 6(a)) and of (s_o, k_o) (Fig. 6(b)). The field of H_M has a minimum of ∼1.4 nats at the edge of the allowable SK domains. The maximum of $H_M \sim 2.55$ nats, lies at the interior of the domain, at $(s_o, k_o) \sim (0.65, 0.71)$ and $(s_b, k_b) \sim (-1.41, 1.70)$, where both ME-pdfs are not very far from Gaussian at which $H_M = 2 \times H_G \approx 2.82$ nats. The Eqs. (39a), (39b), give the close estimation $(s_o, k_o) \sim (0.29, 0.19)$ and $(s_b, k_b) \sim (-1.48, 1.75)$.

All of the above ME problems are numerically solved following a modified version of the algorithm presented in [31], also used in [32]. Numerical details are given in the Appendix.

6. Minimum variance and best linear unbiased estimators

6.1. Generic properties

The *Minimum Variance Unbiased Estimator* (MVUE) x_M, determined from an observation y and a background estimate $x_b = y - d$ is equal to the conditional mean

$$x_M = E\left[x|x_b, y\right] = x + \varepsilon_M, \tag{40}$$

where ε_M is the posterior error. x_M can be expressed in terms of the correction Δ_{Mb} to be added to the unbiased background estimate in the form:

$$\begin{aligned}\Delta_{Mb} &\equiv x_M - (x_b - b_{cb}) = \varepsilon_M - (\varepsilon_b - b_{cb})\\ &= -E\left[\varepsilon_b - b_{cb}|x_b, y\right]\\ &= -E\left[\varepsilon_b - b_{cb}|d, y\right].\end{aligned} \tag{41}$$

Therefore, the posterior error ε_M is regarded as the deviation of the unbiased background error with respect to the innovation conditional mean. Expressing ε_b in terms of the *standardized* error and assuming that it is independent from the observation y for a given innovation, one obtains:

$$\Delta_{Mb} = -\sigma_{cb} E\left[\varepsilon_b^*|d, y\right] = -\sigma_{cb} E\left[\varepsilon_b^*|d\right]. \tag{42}$$

Since d is a known function of the errors (7), we can evaluate Δ_{Mb} from the conditional mean of ε_b^* by using the joint specified pdf of ε_b^* and ε_o^* and the Bayes formula for pdfs. Then, using the independence of errors, we get:

$$\rho_{b*}\left(\varepsilon_b^*|d\right) = \frac{\rho_{b*}\left(\varepsilon_b^*\right)\rho_{o*}\left[\left(d - b_d + \sigma_{cb}\varepsilon_b^*\right)\sigma_{co}^{-1}\right]}{\int \rho_{b*}(u)\,\rho_{o*}\left[(d - b_d + \sigma_{cb}u)\,\sigma_{co}^{-1}\right]\mathrm{d}u}. \tag{43}$$

The normalization integral in the denominator of Eq. (43) is proportional to the theoretical innovation pdf $\rho_d(d|\sigma_{cb})$, inferred from pdfs of errors and computed at the particular value of d and σ_{cb}. The empirical estimate of $\rho_d(d|\sigma_{cb})$ can be obtained from the innovations samples and its difference from the theoretical pdf is a measure of the goodness of the specified pdfs of the errors. Smaller is this difference, more consistent are the specified error pdfs with the innovations.

Using Eq. (43), the correction Δ_{Mb} will be given by the formula of the mean:

$$\Delta_{Mb} = -\sigma_{cb}\int u\rho_{b^*}(u|d)\,\mathrm{d}u. \tag{44}$$

On the other hand, the pdf $\rho_M(\varepsilon_M)$ of the posterior error ε_M is given by a scaled transform of the pdf (43) as:

$$\rho_M(\varepsilon_M) = \sigma_{cb}^{-1}\rho_{b^*}\left(\varepsilon_{b^*} = \frac{\varepsilon_M - \Delta_{Mb}}{\sigma_{cb}}\right). \tag{45}$$

This pdf is completely determined by pdfs of errors thus providing a way of computing centered moments and cumulants of the posterior error such as the variance, skewness and kurtosis, as well as the Shannon entropy, say the posterior entropy of the error. An alternative estimator, not used here, to the MVUE is the *Maximum A Posteriori* (MAP) corresponding to the absolute mode of the posterior pdf, and given by $x_b - b_{cb} - \sigma_{cb}\arg\max\rho_{b*}(\varepsilon_b^*|d)$.

The mean square of the MVUE error is given by the conditional variance of ε_b computed at the innovation value d, *i.e.* $\mathrm{E}[\varepsilon_M^2|d] = \mathrm{var}[\varepsilon_b|d]$. For particular extreme innovations, say larger than two standard deviations and certain non-Gaussian error pdfs (e.g. exponential pdf laws), that posterior variance can be larger than the smallest of the error *a priori* variances *i.e.* the $\min(\sigma_b^2, \sigma_o^2)$ [1]. The same inequality holds for the posterior Shannon entropy of the error [33]. In spite of that, the overall mean square of the *a posteriori* error, say for a large sample of innovations values d, shall be smaller than the *a priori* mean square error: $\min(\sigma_b^2, \sigma_o^2)$. In fact, this is due to the 'law of total variance' in the form:

$$\begin{aligned}\mathrm{E}_d\left(E\left(\varepsilon_M^2|d\right)\right) \equiv \sigma_M^2 &= \mathrm{E}_d\left[\mathrm{E}\left((\varepsilon_b - b_{cb})^2|d\right)\right]\\ &= \sigma_b^2 - \mathrm{E}_d\left[\mathrm{E}\left(\varepsilon_b|d\right)^2\right] \le \min(\sigma_b^2, \sigma_o^2),\end{aligned} \tag{46}$$

where the expectation E_d is the average running through a large sample of innovations.

The mean Shannon entropy of posterior errors $H(\varepsilon_M) = E_d[H(\varepsilon_b|d]$ also decreases with respect to that of the prior errors *i.e.* $H(\varepsilon_M) \le \max(H(\varepsilon_o), H(\varepsilon_b))$, due the general property of the conditional entropy: $E_b[H(a|b)] \le H(a)$ for any pair (a, b) of random variables.

The prior entropy of errors also relates to the posterior error entropy in the form:

$$\begin{aligned}H(\varepsilon_b, \varepsilon_o) &= H(\varepsilon_b, d) = H(\varepsilon_b|d) + H(d)\\ &= H(\varepsilon_M) + H(d),\end{aligned} \tag{47}$$

thanks to the conditional entropy relationship and the volume preserving linear mapping $(\varepsilon_b, \varepsilon_o) \leftrightarrow (\varepsilon_b, d)$. Therefore, maximization of the entropy of prior errors lead to maximization of the mean entropy of posterior errors, under a fixed innovation pdf and its entropy. Furthermore, by using the *standardized* posterior error: $\varepsilon_M^* = \varepsilon_M/\sigma_M$, we get the decomposition

$$H(\varepsilon_M) = \ln\sigma_M + H(\varepsilon_M^*), \tag{48}$$

into an increasing σ_M term and a non-Gaussianity dependent term. This shows that the ME posterior state, maximizing $H(\varepsilon_M)$, is not necessarily that which maximizes variance.

When errors are Gaussian, the MVUE is the classical Best Linear Unbiased Estimator, [34] given by:

$$x_L = x_b - b_{cb} + \Delta_{Lb} = x + \varepsilon_L; \quad \varepsilon_L = \varepsilon_b - b_{cb} + \Delta_{Lb}, \tag{49a}$$

$$\Delta_{Lb} \equiv \frac{\sigma_{cb}^2}{\sigma_{cb}^2 + \sigma_{co}^2}\left(y - x_b - b_d\right), \tag{49b}$$

where Δ_{Lb} is the linear correction to be added to the unbiased background and ε_L is the BLUE error. Even when errors are non-Gaussian, the BLUE still gives the optimal linear estimator of the state x from the observation y and the background x_b. The impact in the *a posteriori* estimate of taking into account non-Gaussianity of errors is given by the difference between the BLUE and MVUE:

$$\Delta_{ML} \equiv x_M - x_L = \Delta_{Mb} - \Delta_{Lb} = \varepsilon_M - \varepsilon_L, \tag{50}$$

also given by the difference between the BLUE and MVUE errors. The mean square error of the BLUE, over a large set of innovations, admits the decomposition:

$$\mathrm{E}_d\left[\varepsilon_L^2\right] \equiv \sigma_L^2 = \sigma_d^2\beta_o\beta_b = \sigma_M^2 + \mathrm{E}_d\left[\Delta_{ML}^2\right] \ge \sigma_M^2, \tag{51}$$

which shows that a self-consistent assimilation scheme based on the non-Gaussian MVUE has in fact minimal variance thus performing better than the BLUE.

6.2. Average skill in the SK-domain

In this section we assess the performance of the MVUE over the innovation SK-domain. In order to evaluate the overall relative skill of the MVUE, compared to the BLUE, we compute the non-dimensional Mean Square Error Skill Score (MSESS), taken with respect to the BLUE, also often used in forecasting evaluation [35]:

$$\mathrm{MSESS} \equiv \frac{\mathrm{E}_d\left(\Delta_{ML}^2\right)}{\sigma_L^2} = 1 - \frac{\sigma_M^2}{\sigma_L^2} \in [0, 1]. \tag{52}$$

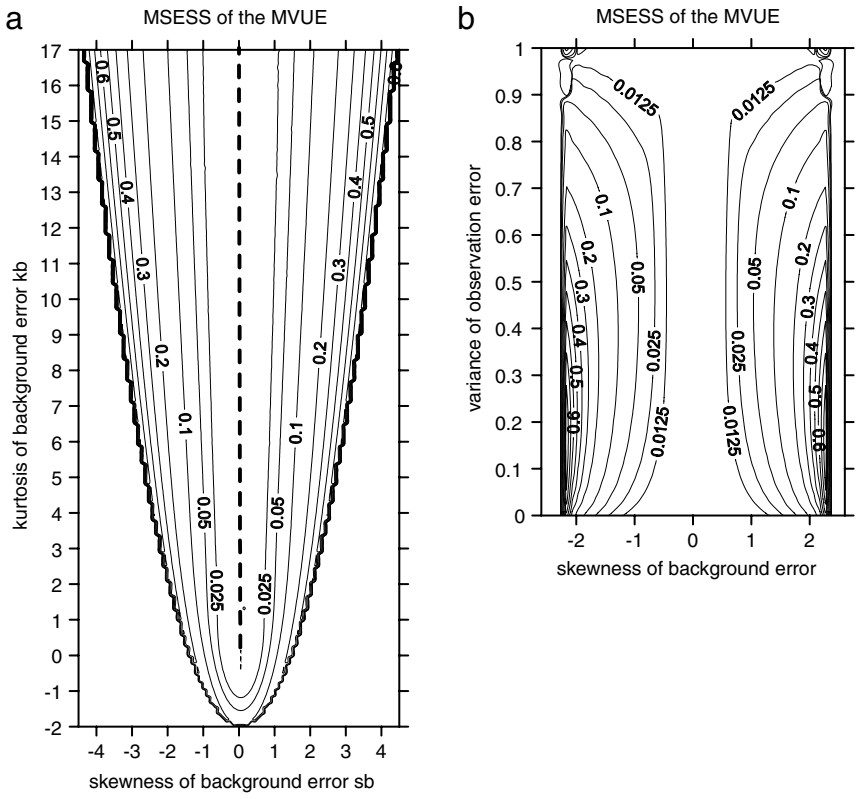


Fig. 7. Contour field of the MSESS of the MVUE obtained from ME-pdfs with a Gaussian error: ($s_o = 0$, $k_o = 0$) and a non-Gaussian error with skewness s_b, and kurtosis k_b. (a) Cross section along the plan (s_b, k_b) for a fixed variance fraction $s_o = 0.5$. The semi-straight line corresponding to un-skewed leptokurtosis is marked as dashed. (b) Contour field of the MSESS in the cross section along s_b and varying β_o from 0 up to 1 for fixed $k_b = 3$.

This is independent from the innovation variance and ranges between 0, for Gaussian errors and 1 when the MVUE is exact, *i.e.* null *a posteriori* error, holding if one of the error variances vanishes. The MSESS depends on the error variance partition: β_o, $\beta_b = 1 - \beta_o$ and also on the assumed pdfs of the *standardized* errors. The MSESS can also be seen as a measure of the BLUE sub-optimality in presence of non-Gaussian errors.

Next, we study how MSESS changes with the fixed constraints for the three scenarios of non-Gaussianity partition among errors, described in Section 5.2. The full assessment is performed numerically since no analytical formula exists for the MSESS dependence on the imposed error moments. In Section 7, the MSESS is evaluated for the analysed samples of brightness temperatures.

6.2.1. Single source of non-Gaussianity

Here, we compute the MSESS, (52) when only one of the errors is non-Gaussian. Without loss of generality we put all the non-Gaussianity into the background error of variance β_b thus leading to $(s_b, k_b) = (-s_d/\beta_b^{3/2}, k_d/\beta_b^2)$. Let us now study how the MSESS depends on s_b, k_b. The MSESS contours for the case of equal error variances: $\beta_o = \beta_b = 0.5$ are shown in Fig. 7(a) for $|s_b| < 4.5$ and $-2 < k_b < 17$. The MSESS increases monotonically with $|s_b|$, for any k_b, from the zero value, at the semi-straight line: $s_b = 0$, $k_b \geq 0$ (dashed in Fig. 7(a)), corresponding to leptokurtic non-Gaussian errors, up to a maximum value near the limit curve: $\eta_b \equiv k_b - s_b^2 + 2 = 0$. There, the non-Gaussian error pdf is tightly constrained, leading to a singular behaviour of the posterior pdf and to a concentration of MSESS contour lines. The MSESS grows with the absolute value $|k_b|$ for platykurtic non-Gaussian errors ($k_b < 0$), contrary to the case of leptokurtic errors ($k_b > 0$) where MSESS slightly decreases with kurtosis and fixed skewness. Similar qualitative MSESS behaviour holds for different error variance partitions.

Still keeping the same scenario, we study how MSESS changes with the Gaussian error variance β_o and the skewness s_b by fixing the non-Gaussian background kurtosis. The corresponding MSESS contours are shown in Fig. 7(b) for the particular case $k_b = 3$. There, the MSESS field is symmetrical with respect to s_b, which occurs in general when one of the errors is un-skewed *i.e.* $s_o = 0$. As expected, the MSESS tends to zero, *i.e*, one finds the exact estimation, when β_o tends to 0 or 1, corresponding respectively to exact observation and exact background. Then, for a given s_b, the maximum MSESS lies at a particular value of β_o, close to 0.5, corresponding to the error variance partition for which the *a posteriori* BLUE variance is maximum. The value of the non-Gaussian error variance β_b at the MSESS maximum, slightly increases as far as the absolute value of the error skewness increases. When the largest error variance comes from the Gaussian error (high β_o), the BLUE is rather satisfactory as it is seen from the low MSESS values (Fig. 7(b)). On the other hand, if the dominant error is non-Gaussian (low β_o), the MVUE is much better than the BLUE, especially at high values of $|s_b|$.

Now, fixing k_d and s_d and increasing β_o up to the allowable maximum $\beta_g(s_d, k_d)$ leads to increasing values of the skewness and kurtosis of the non-Gaussian error and also to an increasing MSESS (Fig. 8(a)–(c) for $k_d = 1, 5, 13$ respectively). This means that the smaller is the variance of the error to which non-Gaussianity is attributed, the larger is the difference between the MVUE and BLUE. If the non-Gaussian error variance β_b is much lower β_o much closer to 1, then k_b can assume extremely high values with remarked effects of the finiteness of the ME-pdfs supports (e.g. $k_d = 5$ and $\beta_b = 0.1$ leads to $k_b = 503$). This effect happens inside hill-shaped regions bounded by contour inflections in Fig. 8(a)–(c).

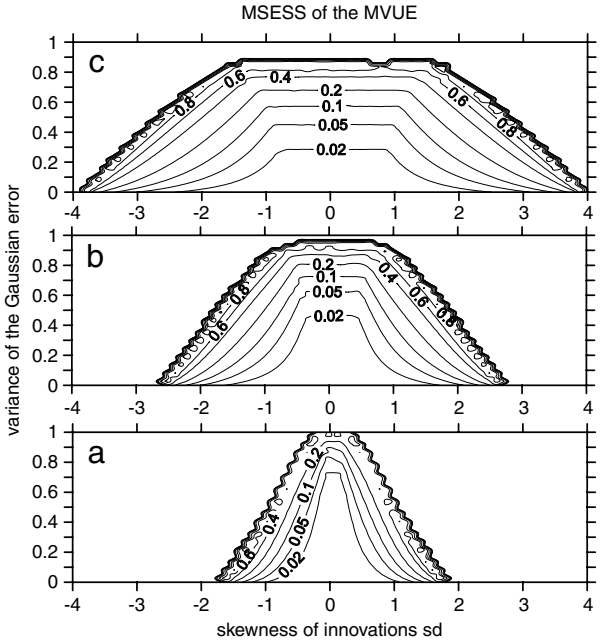


Fig. 8. Contour field of MSESS of the MVUE under the scenario of single source of non-Gaussianity in the plane determined by the variance of the Gaussian error β_o versus innovation skewness s_d (a) case for fixed innovation kurtosis $k_d = 1$, (b) same as (a) for $k_d = 5$, (c) same as (a) for $k_d = 13$. The cases for which $|s_b| > 5$ or $|k_b| > 17$ are blanked, appearing inside the hill-shaped regions bounded by contour inflections in (a), (b) and (c).

6.2.2. Prescribed non-Gaussian errors

In a second experiment, we study the case of two non-Gaussian errors. In order to mimic that situation, we compute MSESS by fixing equal error variances: $\beta_o = \beta_b = 0.5$. For the observation error we fix $(s_o, k_o) = (1.5, 3)$ and let varying (s_b, k_b) of the background error. Corresponding MSESS contours are shown in Fig. 9(a), homologous to Fig. 7(a) (fixed Gaussian error). By comparing those figures, the symmetry around s_b is broken. The effect of non-Gaussianity of errors in the *a posteriori* estimate is stronger (MSESS higher) when the skewnesses of errors have the same sign. Conversely the MSESS is lower when the error skewnesses tend to equilibrate or have opposite skewness signs (left side of Fig. 9(a)), as compared to the case of equal signs (right side of Fig. 9(a)). The minimum of MSESS (∼0), is found at $s_b = -s_o = -1.5$ and equal error kurtosis: $k_b = k_o = 3$. This means that both the MVUE and BLUE are quite similar despite the fact that the *a posteriori* pdf is strongly non-Gaussian. Indeed, when prior skewnesses have opposite signs, the posterior tends to be more symmetric leading to more equal MVUE and BLUE. The MVUE and the BLUE are still the same for any symmetric *a posteriori* pdf. Fig. 9(b) is homologous to Fig. 7(b) for a fixed $k_b = 3$. MSESS increases when one approaches the limits $s_b = \pm(k_b + 2)^{1/2}$. For s_b and s_o of the same sign (right side of Fig. 9(b)), the MSESS grows regularly up to ∼0.8, whereas in the left side (opposite signs) it hardly reaches 0.2 or at much 0.5 for high values of $|s_b|$. When errors are skewed towards the same side, *i.e.* for fixed $s_b > 0$ (right side of Fig. 9(b)), the MSESS is maximum for an error variance partition such that $\beta_o - \beta_b$ is a monotonous increasing function of $|s_o - s_b|$. In other words, high MSESS is favoured by high (small) error's skewness weighted by high (small) variances. On the other hand, if s_b, s_o have opposite signs (left side of Fig. 9(b)), high MSESS holds when the error of largest variance has the largest absolute skewness. Equal error variances under opposite error skewnesses tend to produce small differences between the MVUE and BLUE.

Although errors have finite non-null variance, the effect of error non-Gaussianity can have effects ranging from nearly exactness of the MVUE *i.e.*MSESS = 1 up to non-existence of differences

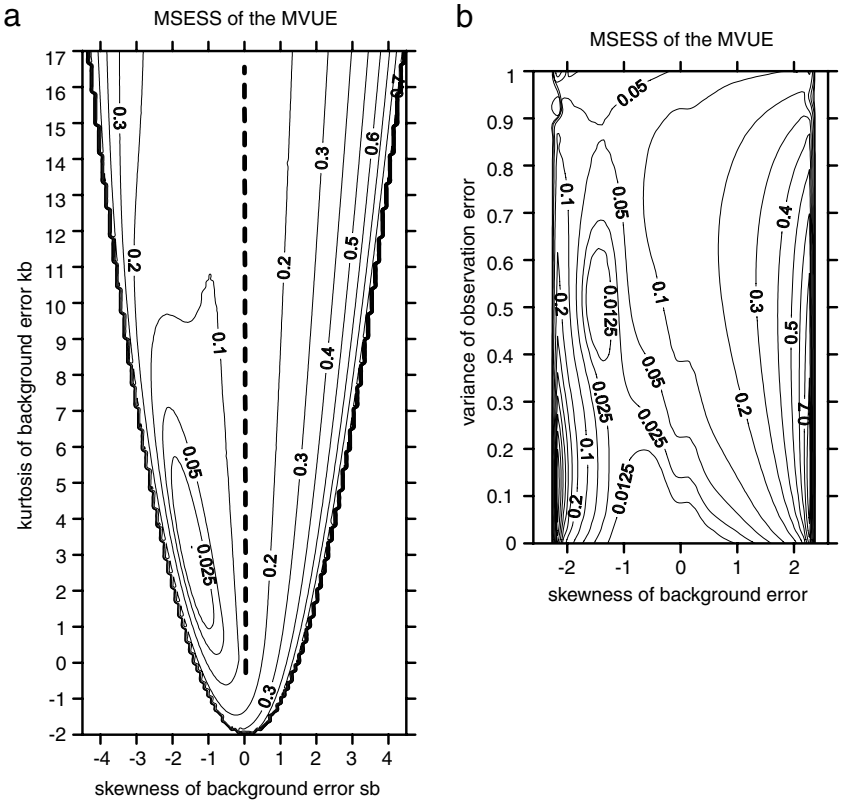


Fig. 9. (a) The same as Fig. 7(a) fixing skewness and kurtosis of observation errors to: $s_o = 1.5$ and $k_o = 3$ and error variance fraction $\beta_o = 0.5$ (b) The same as Fig. 7(b) with the same values: $s_o = 1.5$ and $k_o = 3$.

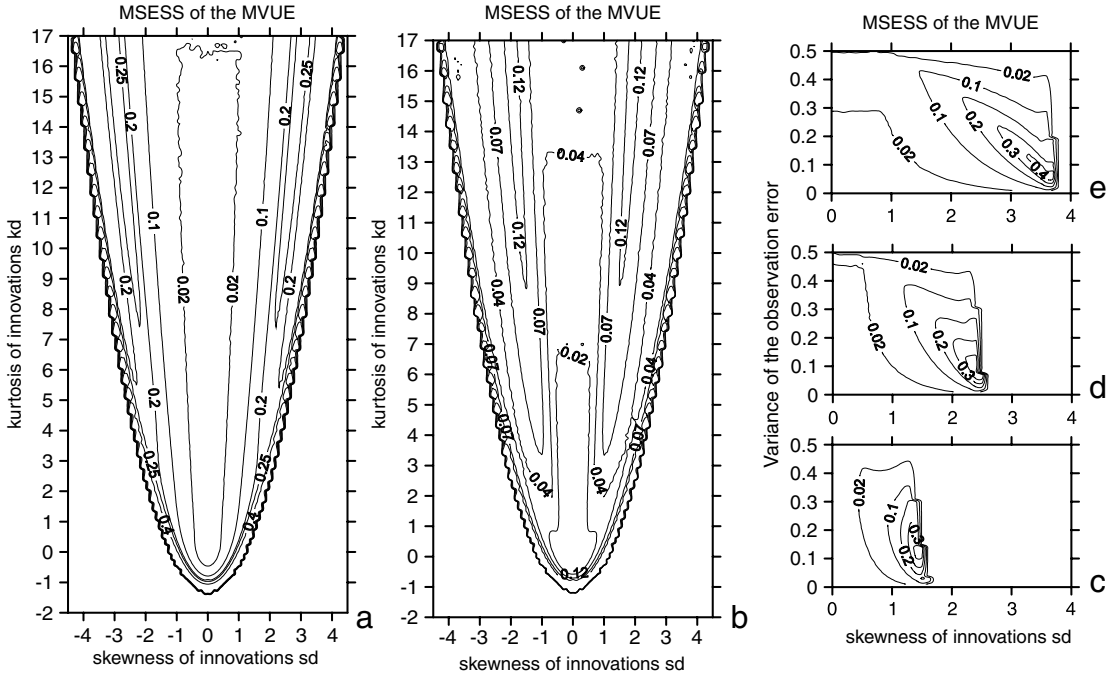


Fig. 10. Contour field of the MSESS of the MVUE for the scenario of *Maximum entropy sharing of non-Gaussianity*. (a) For fixed $\beta_o = 0.25$ and varying (s_d, k_d), (b) Similar to (a) with $\beta_o = 0.40$, (c) For fixed $k_d = 1$ and varying (β_o, s_d), (d) Similar to (c) with $k_d = 5$; (e) Similar to (c) with $k_d = 13$.

between the MVUE estimator and the BLUE (MSESS $= 0$). This is illustrated in MSESS contouring of Fig. 6(c) in the allowable SK-domain of observation errors for the case ($s_d = 1$, $k_d = 1$, $\beta_o = 0.25$). The MSESS field ranges between the maximum ∼0.98, lying near the left edge of the permitted SK-domain. The minimum of MSESS, ∼0.01 lies at $(s_o, k_o) \sim (2.37, 5.94)$, close to the right edge of the set.

6.2.3. Maximum entropy sharing of non-Gaussianity

Here, we assess how the MSESS depends on both innovation non-Gaussianity through (s_d, k_d) and the error variance partition (β_o, β_b) under the scenario of 'maximum entropy sharing of non-Gaussianity'. From symmetry considerations, the MSESS is symmetric with respect to s_d and invariant for the mapping $\beta_o \leftrightarrow 1 - \beta_o$, due to the symmetrical roles of errors.

Now, let us study the case of fixed error variances (β_o, β_b) and fixed innovation statistics. In this case, σ_L is fixed (51) and thus the minimum of MSESS lies at maximum of σ_M, which is replaced from the solution maximizing $H(\varepsilon_M)$ as seen from the decomposition (48). This fact is clear for the case discussed in Fig. 6(c) for $s_d = 1$, $k_d = 1$, $\beta_o = 0.25$, $\beta_b = 0.75$. In fact, at the ME solution, $H_M \sim 2.55$ nats, $(s_o, k_o) \sim (0.65, 0.71)$, and the MSESS is ∼0.14. At the point of a Gaussian observation error, $H_M \sim 2.45$ nats, $(s_o, k_o) = (0, 0)$, and the MSESS is ∼0.25.

Contours of MSESS over the (s_d, k_d) plane are shown in Fig. 10(a), (b), for $\beta_o = 0.25$ and 0.4 respectively. We notice, in the above figures, that k_d has a lower bound given by Eq. (22) in which $k_d \geq 4\beta_o(1 - \beta_o) + s_d^2 - 2$. By setting high enough values of β_o and k_d, the MSESS increases from nearly zero at $s_d = 0$, where the skewnesses of errors equilibrate themselves, and then reaching a maximum at a specific intermediate $|s_d|$ value and then decreases towards the SK-domain edge (Fig. 10(a), (b)). In Fig. 10(c)–(e), we plot the MSESS as a function of $|s_d|$ and β_o, respectively for $k_d = 1, 5, 13$. We see that in the case of leptokurtic innovations ($k_d > 0$) and relatively low skewness values, say $|s_d| \lesssim 1$, the MVUE and BLUE do not differ much, according to the values MSESS $\lesssim 0.01$. The highest MSESS values hold when one of the error variances β_o or β_b range between 0.1 and 0.3, and the innovation skewness is sufficiently high.

Fig. 11(a)–(e) show how the s_o and k_o change for different values of s_d (in abscissas) and β_o (in ordinates) for the same previous values of k_d. Fig. 11(a), (c), (e) refer to s_o and Fig. 11(b), (d), (f) refer to k_o respectively for $k_d = 1, 5, 13$. From those figures, one verifies that the ME solution tends to concentrate the skewness and kurtosis of the innovations onto the error with largest variance.

7. Average skill for brightness temperature innovation samples

One of the aims of this article is to assess the potential average gains from using the MVUE, using priors inferred from the Maximum Entropy method, as an alternative to the BLUE. This is applied here to the assimilation of brightness temperatures extracted from HIRS channels. In order to measure that, we use an adimensional score (larger when it is better), based on the mean square of the correction (MVUE minus BLUE): $\Delta_{ML} = x_M - x_L$, computed for different classes of *standardized* innovation: $d^* = (d - b_d)/\sigma_d$ over the innovation samples. The score is obtained from conditional averages over three classes of the *standardized* innovations according to its size as:

$$\begin{aligned} &\mathrm{MSC}_j = \sigma_d^{-2}\overline{\Delta_{ML}^2 \,|\, |d^*| \in \mathrm{I}_j}, \\ &\mathrm{I}_a = [0, 1[; \qquad \mathrm{I}_b = [1, 2[; \qquad \mathrm{I}_c = [2, \infty[\end{aligned} \tag{53}$$

where MSC_j $(j = a, b, c)$ stands for 'mean Score over classes', where MSC_a, MSC_b, and MSC_c, are computed for successively higher values of $|d^*|$, within one, two and over two standard deviations, say the intervals I_a, I_b and I_c. In the present study the analysed innovation samples are not very far from Gaussianity. In fact, the pairs (s_d, k_d) from Table 1 are quite close to point (0, 0) and, therefore, from Fig. 7(a) a small MSESS is expected as compared to the cases of extreme non-Gaussianity shown in the previous sections. Therefore, the frequencies in each interval are nearly those of the standard Gaussian, say $p_a \sim 68\%$, $p_b \sim 26\%$ and $p_c \sim 6\%$, respectively for I_a, I_b and I_c, where the mean value of $|d^*|$ is close to the corresponding Gaussian values: 0.4, 1.5 and 2.5. The scores MSC_j $(j = a, b, c)$ are measures of the squared MVUE minus BLUE, scaled by the innovation variance. The average MSC

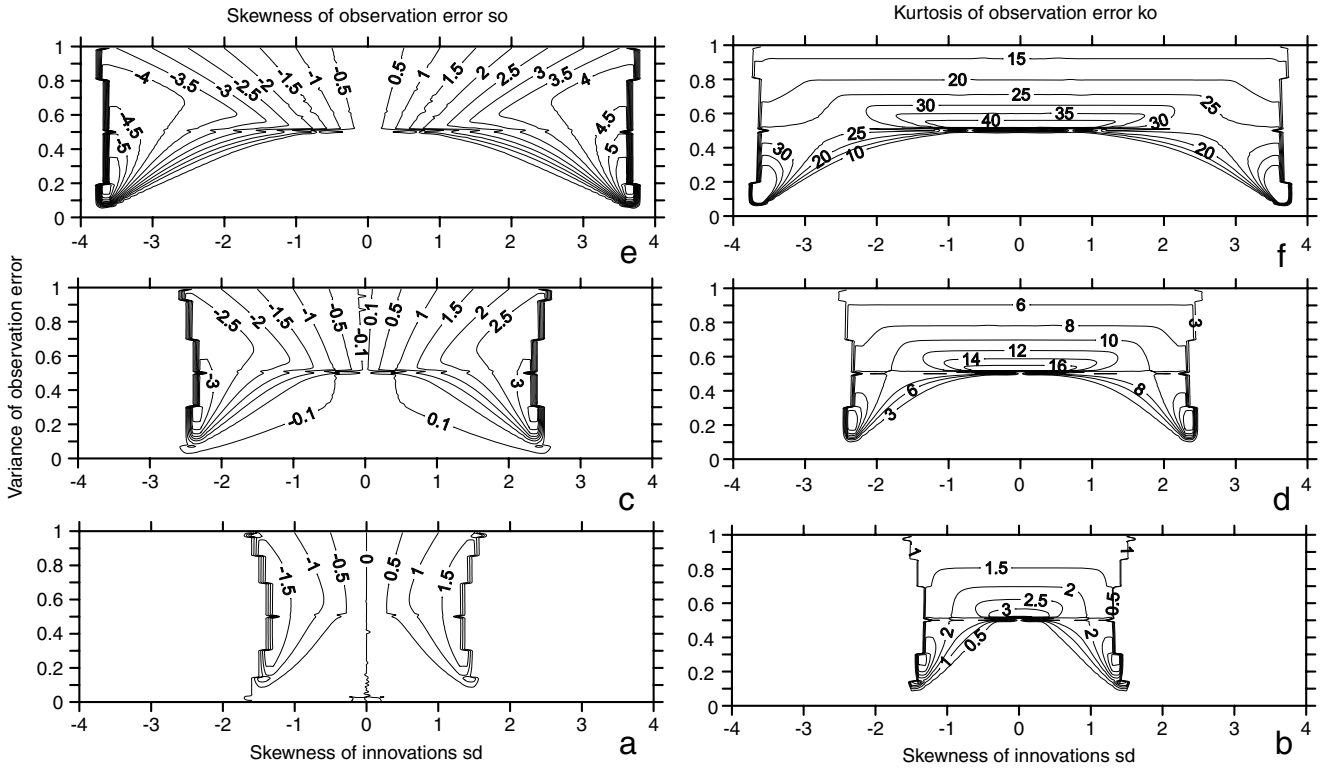


Fig. 11. Contour field of s_o (a), (c), (e) and k_o (b), (d), (f) for the scenario of *Maximum entropy sharing of non-Gaussianity* as function of the innovation skewness s_d and the error variance fraction β_o for fixed values of the innovation kurtosi (k_d). (a), (b) for $k_d = 1$. (c), (d) for $k_d = 5$. (e), (f) for $k_d = 13$.

runs over a large set of innovations and then it is pertinent to equal sample averages to innovation expectations as:

$$\mathrm{MSC} = \sum_{j=a,b,c} p_j \mathrm{MSC}_j = \frac{\mathrm{E}_d\left[\Delta_{ML}^2\right]}{\sigma_d^2}. \tag{54}$$

Using (51) and (52), one finds the relationship between the MSC and the MSESS:

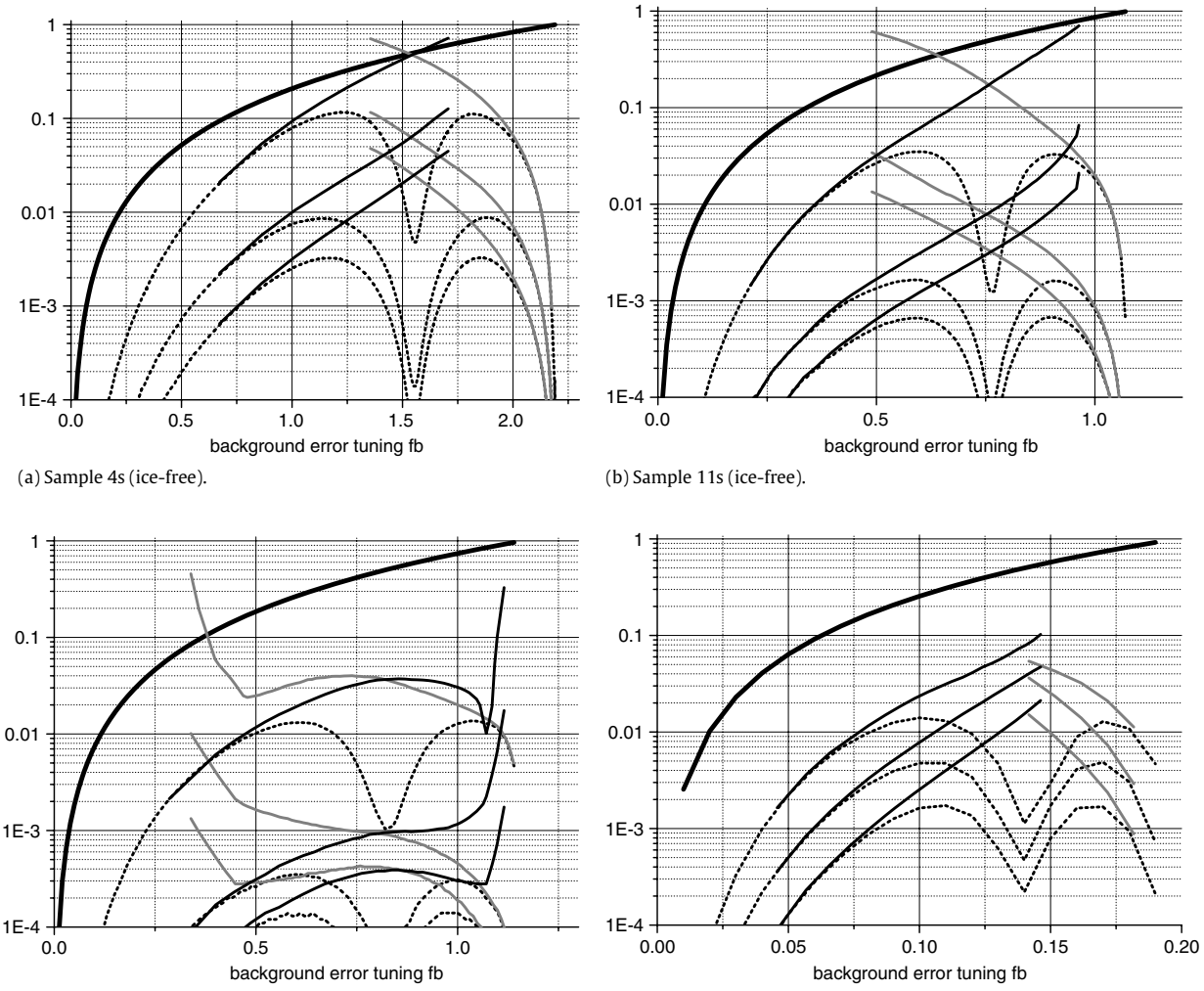
$$\mathrm{MSESS} = \mathrm{MSC}\frac{\sigma_d^2}{\sigma_L^2} = \frac{\mathrm{MSC}}{\beta_o \beta_b}, \tag{55}$$

which makes both scores directly comparable. The three scores MSC_j ($j = a, b, c$) are computed in a quite exhaustive way for 12 brightness temperature innovation samples, where non-Gaussianity is statistically significant. The sample 14s has been excluded due to its non-significant non-Gaussianity. For each sample, we have considered the three following scenarios: (1) non-Gaussianity coming exclusively from observation errors; (2) non-Gaussianity coming exclusively from background errors and (3) 'maximum entropy sharing of non-Gaussianity'. For each scenario, we compute the scores for all possible values of the background error variance fraction β_b by applying the appropriate tunings f_o and f_b, within the allowed intervals.

For four representative samples (4s, 11s, 12s and 7i), we present logarithmic graphics of the scores defined by Eq. (53) in Fig. 12(a)–(d), as functions of f_b, or equivalently as functions of the variance fraction β_b of the background error. The scores are computed for the above referred scenarios: 1, 2 and 3 and the three classes a, b and c each. The tuning values range within estimated bounds given by Eqs. (27) and (28) for the scenarios 1 and 2. All in all, there are 9 score graphics per sample in each Fig. 12(a)–(d). The root mean square of the differences between the BLUE and MVUE are obtained by taking the root square of MSC_j ($j = a, b, c$) and then multiplying it by σ_d from Table 1.

Some conclusions, common to every one of the 12 samples can be drawn, even for the 8 samples whose graphics are not shown. Firstly, for fixed error variances (β_o, β_b) and all the three scenarios of non-Gaussianity (1, 2 or 3), one verifies that the partial scores MSC_j ($j = a, b, c$) tend be higher for the most extreme values of the absolute innovation $|d|$, *i.e.* $\mathrm{MSC}_a < \mathrm{MSC}_b < \mathrm{MSC}_c$. For the scenarios of single source of non-Gaussianity (scenarios 1 and 2) and several samples, the partial values of MSC reaches values up to 0.1–0.6 for the extreme class I_c, *i.e.* when $|d|$ differs more than two standard deviations, corresponding to ∼6% of the cases. This is particularly evident in some 'ice-free' samples namely: 4s (Fig. 12(a)), 5s, 6s, 11s (Fig. 12(b)), and 15s in which MSC_c, corresponding to the most extreme innovation values, go beyond the threshold of 10%, or a mean correction for non-Gaussianity Δ_{ML} (MVUE minus BLUE) of about ∼0.31 σ_d, for a quite large range of possible background error variances, thus showing that, independently of the correct tunings, the relative performance of the MVUE is meaningful. Furthermore, in most of those cases, the variance of the non-Gaussian errors does not need to reach extremely low variances corresponding to unlikely high values of skewness and kurtosis. For the 'ice-covered' samples, the impact of non-Gaussianity in the MVUE is not very high as shown for the sample 7i (Fig. 12(d)).

Secondly, in the medium classes I_a and I_b, the correction Δ_{ML} is quite small for the majority of consistent combinations of error skewnesses and kurtoses. For each scenario and class, the behaviour of MSC_j ($j = a, b, c$) as a function of the variance of errors follows closely the behaviour described in Section 6.2 in terms of MSESS. In fact, in scenarios 1 and 2 (single source of non-Gaussianity), MSC increases as far as the variance of the non-Gaussian error decreases, as suggested from Fig. 8(a) to 8(c). There are quite local exceptions to this behaviour in samples 7s and 12s (Fig. 12(c)), for extreme values of the tunings. In the scenario 3, the partial values of the MSC present the lowest values, except in samples 4s and 5s where it reaches values close to 0.1. In that



(a) Sample 4s (ice-free).

(b) Sample 11s (ice-free).

(c) Sample 12s (ice-free).

(d) Sample 7i (ice-covered).

Fig. 12. Common to all figures (a) to (d): Solid thick line: Value of β_b with the tuned mean variance of background errors. Three sets of curves with the scores MSC_j ($j = a, b, c$) and non-Gaussianity source scenarios: (1) Gaussian background errors: black thin solid lines; (2) Gaussian observation errors: grey thin solid lines and (3) *Maximum entropy sharing of non-Gaussianity*: black dotted lines. The values of the partial scores MSC_a, MSC_b, and MSC_c for each scenario of non-Gaussianity and the same tuning factor f_b obey in all cases to: $\mathrm{MSC}_a < \mathrm{MSC}_b < \mathrm{MSC}_c$. (a) to (d) correspond respectively to samples 4s, 11s, 12s and 7i. The graphics are in logarithmic scale, values of the partial scores below 10^{-4} are not plotted. See text for details.

scenario, the partial scores reach local maxima at intermediate values of the error variance fraction, one near 30% and the other near 70%, as already suggested in Fig. 10(c) to 10(e) showing the MSESS dependence on the error variances.

Finally, we notice that only certain tuning intervals are compatible with the hypothesis of single source of non-Gaussianity. For example, in sets 4s (Fig. 12(a)), 5s, 7s and all 'ice-covered' samples (e.g. sample 7i, in Fig. 12(d)), the absence of background error tuning, *i.e.* $f_b = 1$ do not match with an exclusive non-Gaussian background error, since the upper bound $\beta_{sb}^{-1/2}$ for f_b (Eq. (27b)) is lower than 1. This shows that tuning of error variances and the attribution of non-Gaussianity to errors must be integrated in a more general problem.

In order to see how the MVUE and the BLUE correct the background estimator, we have considered the most skewed sample of innovations (4s) ($s_d = -0.70, k_d = 1.02$), for the case with error pdfs presented in Fig. 5, *i.e.* $\beta_b = 0.47$, $\beta_o = 0.53$, a Gaussian observation error and a non-Gaussian background error. In Fig. 13, we plot the corrections added to the background estimate due to the BLUE (Δ_{Lb}) and due to the MVUE (Δ_{Mb}). The BLUE correction is linear in terms of the innovation d, with a weight given according to Eq. (49b). The spread of points in both quantities is due to varying σ_{sb} along the set. From the definition of innovation $d = \varepsilon_o - \varepsilon_b$, positive (negative) values of d are favoured by negative (positive) values of ε_b, which in our case are quite reduced (enhanced) due to the positive skewness of ε_b. Therefore, as expected, for negative extreme values of the innovation the

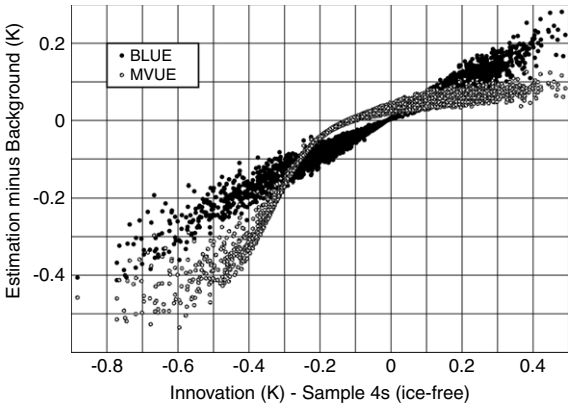


Fig. 13. Scatter plot of the background corrections: Δ_{Lb} (dark spots) and Δ_{Mb} (grey spots), respectively for the BLUE and the MBUE as function of the innovation for the sample 4s, fixing the error variance partition: $\beta_b = 0.47$, $\beta_o = 0.53$.

MVUE-correction: Δ_{Mb} is quite enhanced, whereas for positive d a reduced correction is expected as shown in Fig. 13. In general, under the scenario of single source of non-Gaussianity, the MVUE-correction to the background tend to be enhanced with respect to the BLUE correction for extreme innovations d such that $\mathrm{sgn}(d) = \mathrm{sgn}(\sigma_d)$. For a certain value d, a relatively marginal correction holds when: $\mathrm{sgn}(d) = -\mathrm{sgn}(\sigma_d)$.

8. Conclusions and further work

Most data assimilation (DA) schemes rely on the Best Linear Unbiased Estimator (BLUE) combining observations and background information. However, the *a posteriori* validation of DA algorithms, through the diagnostics both of the innovations (observations minus background) and of the 'data minus analysis', often reveal error biases and inconsistencies in the specified error variances as well as the consequent sub-optimality of the DA schemes. These *a posteriori* diagnostics are extended here to some non-Gaussianity measures of the innovations such as the skewness, kurtosis and negentropy, *i.e*, the difference in terms of Shannon entropy with respect to the Gaussian distribution with the same first and second statistical moments. The presence of statistically significant non-Gaussianity in innovations points out the fact that errors of the assimilated data are non-Gaussian. Therefore, as far as the minimization of the analysis error variance is concerned, the BLUE is not optimal thus calling for the application of the Minimum Variance Unbiased Estimator (MVUE), obtained from the Bayesian data assimilation, using in general non-Gaussian pdfs for errors.

In this paper we present consistency relationships for the third and fourth order moments of innovations, relating it to skewnesses and kurtoses of the errors, coming both from observations and background estimates in the observational space. These diagnostics are obtained from samples of observations and backgrounds of assimilated data as well as from the corresponding specified variance errors, as specified operationally in a certain DA scheme. Significant estimates of non-Gaussianity are obtained from data assimilation ECMWF files of brightness temperature (BT), obtained from HIRS channels. In the paper, we address possible sources of the non-Gaussianity of innovations as: (a) the marginal non-Gaussianity of observation and background errors, (b) the nonlinear error correlations and (c) the heteroscedasticity of errors, *i.e*, the variability of error variances and (d) multiplicative noise. The usually assumed observation-background error independence in DA, imposes the existence of error variance-dependent bounds for the skewness and the kurtosis of errors. As consequence the *a posteriori* tuning of error variances also depends on how non-Gaussian are the errors.

As an application of the innovation diagnostics, we present a measure of the potential impact of the non-Gaussian errors in data assimilation. This is obtained through the mean square difference between the BLUE and the Minimum Variance Unbiased Estimator (MVUE), emanted from the Bayesian filter in the observational space, obtained with single observations and background estimates. This is accomplished by specifying non-Gaussian probability density functions (pdfs) of errors. These pdfs are obtained by the maximum entropy method, constrained by statistical moments up to the fourth order. This is because the ME method provides the least committing pdfs when they are constrained by a certain set of constraints.

The impact of non-Gaussian errors is studied for a large range of moments, being higher for skewed extreme innovations and growing in average with the absolute value of the skewnesses of data errors, especially if those skewnesses have the same sign. The impact is larger for the higher absolute values of innovations, typically larger than two standard deviations. The impacts are assessed for two scenarios, one prescribing the skewness and kurtosis of errors and the second one in which only the non-Gaussianity of innovations is imposed. The square average of the difference (MVUE minus BLUE) is smaller for the second scenario because error pdfs are allowed to be closer to Gaussian ones. All of those conclusions are corroborated in practice in certain samples of BTs.

The above diagnostics only give an indication of the potential impacts of non-Gaussian errors in a Bayesian–ME-based DA scheme. However those diagnostics have the advantage of being obtained from *a posteriori* DA files, without the need of running data assimilation models. An alternative, much intensive assessment of the effect of non-Gaussian errors is the systematic comparison of analyses obtained from a BLUE-based DA scheme and a 'frozen' Bayesian-based DA scheme, which is obviously out of the scope of the paper.

The plans for further work are to develop the approach for the multivariate innovation vector taking into account spatial error correlations. For this case, we have to estimate third and fourth order tensors corresponding respectively to multivariate skewnesses and multivariate kurtosis as presented in [37]. However, there must have a tradeoff between the number of statistics and the sample size. This can probably be solved by considering a reduced number of non-Gaussian variables or by considering integrated diagnostics of non-Gaussianity. The effect of correcting for non-Gaussian errors in DA can be the addition to the BLUE-based cost functions of penalty terms weighting non-Gaussianity.

Acknowledgements

This research was developed at IDL and LMD with support from the project: 'Développements méthodolo-giques pour l'assimilation des observations: Validation *a posteriori*, diagnostics d'optimalité et de Gaussianité' from the Programme National: 'Les Enveloppes Fluides et l'Environnement'. Thanks are due to Elias Holm for providing ECMWF feedback data assimilation files, Pedro Viterbo, Rui Perdigão, Peter Jan van Leeuwen and Sandra Pires for their comments and help.

Appendix. Maximum entropy pdfs constrained by skewness and kurtosis

The ME-pdf of a random variable ε (e.g. error) is closely obtained by solving the ME problem in a finite interval $D = [-L, L]$ such that $L \gg |c_k|^{(1/k)}$ where $\{c_k\}$ is the set of constrained k-order moments with $max(k)$ being an even number in order to get a well-posed ME problem. The value of L is made large enough in order the maximum entropy H_M and the Lagrange multipliers (LMs) does not practically change with increasing L. By using the Leibnitz derivation rule and fixing the expected values, it is easy to obtain the derivative of the ME H_M with respect to L:

$$\frac{\mathrm{d}H_M(\varepsilon)}{\mathrm{d}L} = \rho(L) + \rho(-L), \tag{A.1}$$

where ρ is the ME-pdf. In the present paper, since the maximum order moment is 4, the logarithm $\ln \rho$ at the interval bounds is scaled as $O(-L^4)$, decreasing when L increases. That exponent, 4, is here the maximum even power of the constraining moments. Due to the finite L and for pdfs with kurtosis k and skewness s, near the limit $k \approx s^2 - 2$, a small hump in the pdf tail will accommodate the assumed innovation constraints. Given the above scaling and the range of moments, we can extend the pdf support D sufficiently further so as to get negligible bound effects, both on the entropy and on the ME-pdfs. Furthermore, in order to get integrands of order $\exp(O(1))$, during the optimization process, we solve the ME problem for the scaled variable ε/L in the interval $[-1, 1]$ by

taking the appropriate scaled constraints. Afterwards, we apply the scaling entropy relationship:

$$H_M(\varepsilon) = H_M(\varepsilon/L) + \log(L), \tag{A.2}$$

where $H_M(\varepsilon/L)$ is the entropy of the scaled variable. The LMs also depend on the interval amplitude L, but stabilizing for L high enough. The derivatives of the LMs with respect to L are obtained from the L-derivatives of the expectation constraints as:

$$\frac{d\lambda_k}{dL} = \sum_{i=1}^{N_c=4} \left(M^{-1}\right)_{k,i} \left[\rho(L)\left(L^i - c_i\right) + \rho(-L)\left\{(-L)^i - c_i\right\}\right]; \quad k = 1, \ldots, N_c = 4, \tag{A.3}$$

where M is a square matrix of order N_c whose elements $M_{k,i}$ are the covariances between the powers ε^k and ε^i as determined by the ME-pdf for L. The elements $M^{-1}_{k,i}$ in (A.3) are the (k, i) elements of the inverse matrix of M. The minimization problem is solved by the Quasi-Newton method using the M1QN3 routine [36], starting at $\lambda_k = 0$ for all k. The integrals giving the functional Γ to minimize and its λ-derivatives are approximated by the Gauss quadrature rule with N_p weighting factors in the interval $[-1, 1]$. In order to get full resolution during the minimization, and to avoid NAN (Not-a-Number) and INF (Infinite) numbers in computation, we subtract the polynomials in the arguments of exponentials by the correspondent maximum in D. Finally, the functional Γ is multiplied by a sufficient high factor F_a in order to emphasize the gradient.

In the paper, we have considered centered, unitary variance errors ε with skewness $s \in [0, 4.5]$ and kurtosis $k \in [-2, 17]$. For each pair (s, k), the LMs$\lambda_1, \ldots, \lambda_4$ and the normalization factor λ_0 of the ME-pdf have been computed for which some generic properties can be found in [38]. Recall that for a standard Gaussian variable $(s, k) = (0, 0)$, $\lambda_0 = \sqrt{2\pi}$, $\lambda_1 = \lambda_3 = \lambda_4 = 0$, $\lambda_2 = -0.5$ at which the entropy reaches the maximum $H_G = 0.5\ln(2\pi e) \approx 1.41$ nats. The absolute deviations from the corresponding Gaussian values, both for the LMs and the H_M tend to increase as far as the limit curve $k = s^2 - 2$ is reached [31]. For the range of constraint moments, several experiments have led to the reasonable values: $L = 20$, $N_p = 500$, $F_a = 1000$, for which convergence is obtained nearly at 60 iterates with an accuracy of 10^{-6} of the gradient of Γ. Larger values of L require larger values of N_p.

References

[1] O. Talagrand, Bayesian estimation. Optimal interpolation. Statistical linear estimation, in: R. Swinbank, V. Shutyaev, W. Lahoz (Eds.), Data Assimilation for the Earth System, Kluwer Academic Publishers, 2002, pp. 21–35.
[2] Q. Errera, D. Fonteyn, Four-dimensional variational chemical assimilation of crista stratospheric measurements, J. Geophys. Res. 106 (2001) 12253.
[3] H. Elbern, H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, J. Geophys. Res. 106 (2001) 3569.
[4] N.B. Ingleby, A.C. Lorenc, Bayesian quality control using multivariate normal distributions, Q. J. R. Meteorol. Soc. 119 (1993) 1195.
[5] M. Bocquet, C.A. Pires, L. Wu, Beyond Gaussian statistical modeling in geophysical data assimilation, Mon. Weather Rev. (2010), in press (doi:10.1175/2010MWR3164.1).
[6] S.J. Fletcher, M. Zupanski, A hybrid normal and lognormal distribution for data assimilation, Atmos. Sci. Lett. 7 (2006) 43.
[7] S.J. Fletcher, M. Zupanski, A data assimilation method for log-normally distributed observational errors, Q. J. R. Meteorol. Soc. 132 (2006) 2505.
[8] M. Bocquet, High resolution reconstruction of a tracer dispersion event, Q. J. R. Meteorol. Soc. 133 (2007) 1013.
[9] S. Fletcher, Mixed Gaussian-lognormal four-dimensional data assimilation, Tellus 62 (3) (2010) 266.
[10] D.P. Dee, Bias and data assimilation, Q. J. R. Meteorol. Soc. 131 (2005) 3323.
[11] E. Holm, Revision of the ECMWF humidity analysis: construction of a Gaussian control variable, in: Proceedings of the ECMWF/GEWEX Workshop on Humidity Analysis, 2002, pp. 1–6.
[12] E.T. Jaynes, Information statistics and statistical mechanics I, Phys. Rev. 106 (1957) 620.
[13] E.T. Jaynes, Information statistics and statistical mechanics II, Phys. Rev. 108 (1957) 171.
[14] Thomas M. Cover, A. Thomas, Elements of Information Theory, in: Wiley Series on Communications, 1991.
[15] J. Derber, F. Bouttier, A reformulation of the background error covariance in the ECMWF global data assimilation system, Tellus 51 (1999) 195.
[16] E. Anderson, M. Fisher, R. Munro, A. McNally, Diagnosis of background errors for radiances and other observable quantities in the variational data assimilation scheme and the explanation of a case of poor convergence, ECMWF Technical Memorandum, 1999, p. 296.
[17] X. Wu, Calculation of maximum entropy densities with application to income distribution, J. Econometrics 115 (2003) 347.
[18] P. Comon, Independent component analysis, a new concept? Signal Process. 36 (1994) 287.
[19] J.F. Kenney, E.S. Keeping, Cumulants and the cumulant-generating function, additive property of cumulants, and Sheppard's correction, in: Mathematics of Statistics, Van Nostrand, 1951, pp. 77–82.
[20] G. Maurice, A. Stuart, J.K. Ord, A. O'Hagan, Kendall's Advanced Theory of Statistics, 6th ed., Edward Arnold, 1999.
[21] G. Desroziers, S. Ivanov, Diagnosis and adaptative tuning of information error parameters in variational assimilation, Q. J. R. Meteorol. Soc. 127 (2001) 1433.
[22] B. Chapnik, G. Desroziers, F. Rabier, O. Talagrand, Properties and first application of an error-statistics tuning method in variational assimilation, Q. J. R. Meteorol. Soc. 130 (2004) 2253.
[23] V. Rudin, Real and Complex Analysis, McGraw-Hill, 1987.
[24] G.M. Clarke, D. Cooke, A Basic Course in Statistics, Oxford University Press, New York, USA, 1991.
[25] C.E. Shannon, The mathematical theory of communication, Bell Syst. Tech. J. 27 (1948) 379.
[26] B. Urban, Humidity profile retrieval using a maximum entropy principle, Mon. Weather Rev. 123 (1995) 1185.
[27] M. Bocquet, Reconstruction of an atmospheric tracer source using the principle of maximum entropy I, Q. J. R. Meteorol. Soc. 131 (2005) 2191.
[28] M. Bocquet, Reconstruction of an atmospheric tracer source using the principle of maximum entropy II, Q. J. R. Meteorol. Soc. 131 (2005) 2209.
[29] M. Bocquet, Inverse modelling of atmospheric tracers: non-Gaussian methods and second-order sensitivity analysis, Nonlinear Process. Geophys. 15 (2005) 127.
[30] O. Barndorff-Nielsen, Information and Exponential Families in Statistical Theory, in: Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Ltd., 1978.
[31] M. Rockinger, E. Jondeau, Entropy densities with an application to autoregressive conditional skewness and kurtosis, J. Econometrics 106 (2002) 119.
[32] C. Pires, R. Perdigão, Non-gaussianity and asymmetry of the winter monthly precipitation estimation from NAO, Mon. Weather Rev. 135 (2007) 430. doi:10.1175/MWR3407.1.
[33] C. Pires, Aplicação da entropia estatística à assimilação de dados, in: 3º Simpósio de Meteorologia e Geofísica da APMG, Associação Portuguesa de Meteorologia e Geofísica, 2003, p. 220.
[34] A. Gelb, Applied Optimal Estimation. The Analytical Sciences Corporation, MIT Press, 1974.
[35] M. Déqué, Continuous variables, in: I. Jollife, D. Stephenson (Eds.), A Practitioner's Guide in Atmospheric Science, John Wiley and Sons, England, 2003, pp. 97–119.
[36] J.C. Gilbert, C. Le Maréchal, Some numerical experiments with variable storage quasi-Newton algorithms, Math. Program. 45 (1989) 407.
[37] I. Takafumi, On a measure of multivariate skewness and a test for multivariate normality, Ann. Inst. Statist. Math. 34 (1982) 531.
[38] D. Ormoneit, H. White, An efficient algorithm to compute maximum entropy densities, Econometric Rev. 18 (1999) 127.