



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

Carlos Pires, Olivier Talagrand, Marc Bocquet

► **To cite this version:**

Carlos Pires, Olivier Talagrand, Marc Bocquet. Diagnosis and impacts of non-Gaussianity of innovations in data assimilation. *Physica D: Nonlinear Phenomena*, Elsevier, 2010, 239, <10.1016/j.physd.2010.05.006>. <inria-00538662>

HAL Id: inria-00538662

<https://hal.inria.fr/inria-00538662>

Submitted on 23 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modelling non-Gaussianity of background and observational errors by the Maximum Entropy method

Carlos A. Pires^{*,a}, Olivier Talagrand^b, Marc Bocquet^{c,d}

^a*Instituto Dom Luis, University of Lisbon, Portugal*

^b*Laboratoire de Météorologie Dynamique (LMD), École Normale Supérieure, Paris, France*

^c*Université Paris-Est, CEREAS, Joint Laboratory École des Ponts ParisTech and EDF R&D, Champs-sur-Marne, France*

^d*INRIA, Paris-Rocquencourt Research Center, France*

Abstract

The Best Linear Unbiased Estimator (BLUE) has widely been used in atmospheric and oceanic data assimilation (DA). However, when the errors from data (observations and background forecasts) have non-Gaussian probability density functions (pdfs), the BLUE differs from the absolute Minimum Variance Unbiased Estimator (MVUE), minimizing the mean square a posteriori error. The non-Gaussianity of errors can be due to the inherent statistical skewness and positiveness of some physical observables (e.g. moisture, chemical species) or because of the nonlinearity of the data assimilation models and observation operators acting on Gaussian errors. Non-Gaussianity of assimilated data errors can be justified from a priori hypotheses or inferred from statistical diagnostics of innovations (observation minus background). Following this rationale, we compute measures of innovation non-Gaussianity, namely its skewness and kurtosis, relating it to: a) the non-Gaussianity of the individual error themselves, b) the correlation between nonlinear functions of errors, and c) the heteroscedasticity of errors within diagnostic samples. Those relationships impose bounds for skewness and kurtosis of errors which are critically dependent on the error variances, thus leading to a necessary tuning of error variances in order to accomplish consistency with innovations. We evaluate the sub-optimality of the BLUE as compared to the MVUE, in terms of excess of error variance, under the presence of non-Gaussian errors. The error pdfs are obtained by the maximum entropy method constrained by error moments up to fourth order, from which the Bayesian probability density function and the MVUE are computed. The impact is higher for skewed extreme innovations and grows in average with the skewness of data errors, especially if those skewnesses have the same sign. 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% error variance as compared to the BLUE.

Key words: Non-Gaussianity, Data Assimilation, Innovation Statistics, Maximum Entropy Method

PACS: 92.60.-e, 93.65.+e, 93.85.+e, 92.90.+x,

1. Introduction

Most of the operational atmospheric-oceanic data assimilation systems rely on linear estimation theory that aim at 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]. Nevertheless, the MVUE restricted to the subspace of linear estimators coincides with the BLUE, even when data errors are Non-Gaussian.

Sampling statistics and histograms of innovations (*observation minus projected background*) often reveal statistically significant 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, especially 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 probability density function (pdf) around the mean, which is not valid for example in cases dealing with moisture, gaseous pollutants and aerosols. There are also extreme events like droughts and floods, which are not accommodated by a Gaussian pdf. Another source of non-Gaussianity in observation errors is the presence of extreme gross errors [4]. Data assimilation schemes coping with non-Gaussian error pdfs such as the lognormal are now available [5, 6, 7]. When errors are non-Gaussian, one can still formulate a cost-function with the minimum verifying at the maximum of the *a posteriori* Bayesian pdf, also designated by the *Maximum A Posteriori* (MAP). The referred cost-function can be decomposed into a quadratic term corresponding to Gaussian errors, with the minimum at the BLUE, and an additional term measuring the departure from Gaussianity. The cost function minimum, the MAP, is shifted from the BLUE in accordance to 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 (VarBC) [8] by the addition of

*Corresponding author
Faculdade de Ciências, Edifício C8, DEGGE, Campo Grande, 1749-016 Lisboa, Portugal, Tel. +351 217500886, fax. +351 217500807

Email address: capires@fc.ul.pt (Carlos A. Pires)

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 [9] for moisture. The assumptions about errors pdfs always depend on statistical hypothesis, which may or not be verifiable on samples independent from those where assimilation is performed.

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. The present article aims to diagnose certain measures of non-Gaussianity of innovations, namely its skewness and kurtosis. It aims to clarify how data (observations and background) errors can accommodate non-Gaussianity under certain hypothesis on its origin, either single or multiple, and also under the hypothesis of error observation and background error independence. This is shown in section 3 using samples of innovations, described in section 2, of the brightness temperature 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 study the statistical consistency between innovations and errors. In section 5, we apply the Maximum Entropy (ME) method[10, 11, 12], constrained by innovations statistical moments, up to the fourth-order, and a few hypothesis about the errors, in order to find the least committing error pdfs, compatible with the available statistical information. 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. The average difference between MVUE and BLUE measures the 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 for different scenarios of non-Gaussian origin. Discussion and conclusions come in section 7 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 is 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. The main statistical features of the data errors are given in Table 1. In the ECMWF assimilation system, both the observations and the background are assumed to be unbiased. The specified standard deviation σ_{so} of observational error is constant for each channel. The

standard deviation σ_{sb} of the background error as specified by the ECMWF assimilation scheme depends on the channel, pixel position and temporal instant [13, 14]. 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 background error standard deviation is therefore taken to be larger in the conditions when sea-ice is forecast to be present, preferentially at latitudes poleward of 60° . Fig. 1, relative to channel 6, shows this fact through the scatter plot of pairs $(\sigma_{sb}, \text{North latitude})$, 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 referred above 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 as in Table reftable1, 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. For each of the 13 presented data samples, N denotes the size of the working samples while $m(\sigma_{sb})$ and $\text{std}(\sigma_{sb})$ denote, respectively, the mean and standard deviation of σ_{sb} estimated from data. It is seen that both $m(\sigma_{sb})$ and $\text{std}(\sigma_{sb})$ 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:

$$d = y - x_b = \varepsilon_o - \varepsilon_b, \quad (1)$$

where $x_b = x + \varepsilon_b$ is the background estimate for x , and ε_b the corresponding error. 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 Kelvin degree, 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 $|b_d|/(\sigma_d/\sqrt{N})$ is significantly larger than 1; it varies between 10 and 40.

In the ECMWF assimilation system, background and observation errors are assumed, as is usual, 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. The excess of total error variance is also due to the assumed variance for the background error in the 5 ‘ice-covered’ samples, which is significantly larger than the variance of the innovation.

3.2. Higher-order Statistical Moments and non-Gaussianity

The classical skewness and kurtosis of innovations defined as:

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

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

are respectively the third and fourth-order standardized cumulants of innovations, where the bar and the prime stand respectively for (channel+surface condition) sampling mean and departure with respect to the mean. 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* (*leptokurtic*) if the kurtosis k_d is negative (positive). It thus has tails that are less populated (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 [15]. The kurtosis ranges in the interval $[-2, +\infty[$. This is seen from the variable $r_d \equiv (d'/\sigma_d)^2 - s_d(d'/\sigma_d)$, which has sampling variance:

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

The variable r_d is the second order polynomial of the standardized innovation $d^* \equiv d'/\sigma_d$, which has unitary quadratic coefficient and has the minimum 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-Deltas. 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 in the studied samples, η_d ranges from 1.49 (sample 6i) to 3.99 (sample 12s).

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:

$$KL(\rho_1, \rho_2) \equiv \int \rho_1(x) \ln \left(\frac{\rho_1(x)}{\rho_2(x)} \right) dx. \quad (4)$$

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 expectation and variance [12]. 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 [16]:

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

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 expectation 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 graphic evidence of non-Gaussianity, we present in Fig. 2, the histogram of innovations for the sample 4s which is the more skewed one or having the largest $|s_d|$. 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. There is a clear difference between the two fits. It is interesting to elucidate the origin of those features as 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 high 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 of 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 as systematically).

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 $\varepsilon_o, \varepsilon_b$, leading to consistency relationships between the statistics of errors, which can be specified either by the assimilation scheme or as additional error hypotheses. The statistical moments can also be expressed in terms of cumulants [17, 18], thus leading to related consistency cumulant-relationships. In order to accomplish consistency, we need in general to perform some tuning to the specified error statistics leading in particular to bias error correction and tuning of error variances [19, 20]. 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. Then we introduce tuning factors (f_o, f_b) taken constant within each innovation sample. These factors relate the consistent standard deviations of the unbiased errors ε'_o and ε'_b , noted $(\sigma_{co}, \sigma_{cb})$, 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 own heteroscedasticity, whose presence has to be taken into account in the consistency relationships. In order to isolate its effect, we have considered the standardized observation and background errors:

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

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

In this paper, we will make the fundamental hypothesis that the joint and marginal assumed pdfs of the standardized errors ε_o^* and ε_b^* , as well as the biases b_{co}, b_{cb} , the standard deviation of observation error σ_{co} , and the tuning factors f_o, f_b , are independent both from the observation pixel i.e. its geographical position and from the observation time, being considered uniform along each analysed set. As consequence of the above hypothesis, the standardized errors $\varepsilon_o^*, \varepsilon_b^*$, are also independent of σ_{cb} . Using the above definitions, the innovations can be written as:

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

in which d is a function of two random variables: $\varepsilon_o^*, \varepsilon_b^*$, and the background-error standard deviation σ_{cb} , running through all the possible tuned values σ_{cb} , produced by the data assimilation system. Under the above hypotheses, the average, the centred 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 then comes as:

$$\bar{d} \equiv b_d = \sigma_{co}E(\varepsilon_o^*) - \overline{\sigma_{cb}}E(\varepsilon_b^*) + b_{co} - b_{cb} = b_{co} - b_{cb}, \quad (8)$$

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 centred 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:

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

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

$$\sigma_o^2 \equiv \sigma_{co}^2 \equiv \beta_o\sigma_d^2; \quad (10a)$$

$$\sigma_b^2 \equiv \overline{\sigma_{cb}^2} \equiv \beta_b\sigma_d^2. \quad (10b)$$

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:

$$s_{co} \equiv E(\varepsilon_o^{*3}); \quad (11a)$$

$$s_{cb} \equiv E(\varepsilon_b^{*3}), \quad (11b)$$

the single error kurtosis:

$$k_{co} \equiv E(\varepsilon_o^{*4}) - 3; \quad (12a)$$

$$k_{cb} \equiv E(\varepsilon_b^{*4}) - 3, \quad (12b)$$

the joint cumulants of order 3:

$$\kappa_{oob} \equiv E(\varepsilon_o^*\varepsilon_o^*\varepsilon_b^*); \quad (13a)$$

$$\kappa_{bbo} \equiv E(\varepsilon_b^*\varepsilon_b^*\varepsilon_o^*) \quad (13b)$$

and the joint standardized error cumulants of order 4:

$$\kappa_{ooob} \equiv E(\varepsilon_o^*\varepsilon_o^*\varepsilon_o^*\varepsilon_b^*) - 3E(\varepsilon_o^*\varepsilon_b^*); \quad (14a)$$

$$\kappa_{oobb} \equiv \text{E}(\varepsilon_o^* \varepsilon_o^* \varepsilon_b^* \varepsilon_b^*) - 2\text{E}(\varepsilon_o^* \varepsilon_b^*)^2 - 1; \quad (14b)$$

$$\kappa_{bbbo} \equiv \text{E}(\varepsilon_b^* \varepsilon_b^* \varepsilon_b^* \varepsilon_o^*) - 3\text{E}(\varepsilon_o^* \varepsilon_b^*). \quad (14c)$$

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 writing of 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:

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

The h-terms h_{nb} work as 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, $h_{2b} = 1$. We still have the inequalities: $h_{1b} \leq 1$, $h_{3b} \geq 1$ and $h_{4b} \geq 1$, thanks to the Jensen inequality [21], 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 deviations of h_{1b} and h_{4b} from unity follows that of h_{3b} among the available samples.

The skewness s_d of innovations is then written as:

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

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. We verify that 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 ($s_{co} = s_{cb} = 0$). That is due to non-linear error correlations brought by third-order joint error cumulants. Dividing Equation (9) by σ_d^2 , one gets:

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

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

$$k_d = k_{d1} + k_{d2} + k_{d3}; \quad (18a)$$

$$k_{d1} = \beta_o^2 k_{co} + \beta_b^2 h_{4b} k_{cb}; \quad (18b)$$

$$k_{d2} = 6\beta_o \beta_b \kappa_{oobb} - 4\beta_o^{3/2} \beta_b^{1/2} h_{1b} \kappa_{oob} - 4\beta_o^{1/2} \beta_b^{3/2} h_{3b} \kappa_{bbbo}; \quad (18c)$$

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

The term within k_{d1} in Equation (18a) is the contribution from the kurtosis of individual errors. The second term k_{d2} is uniquely due to the fourth-order standardized error 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.

4.3. Constraints under independent errors

The statistical error independency is a common hypothesis taken in data assimilation. Let us present here some of the constraints of error skewnesses and kurtosis imposed by error independency. The averaged skewness s_b and averaged kurtosis k_b of the background error over a heteroscedastic sample are respectively given by:

$$s_b = h_{3b}s_{cb}; \quad (19a)$$

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

Since observational errors are homoscedastic, we get directly the mean skewness and the mean kurtosis:

$$s_o = s_{co}; \quad (20a)$$

$$k_o = k_{co}. \quad (20b)$$

When errors are assumed statistically independent, the skewness s_d and the kurtosis k_d are given as linear combinations, respectively of the error skewnesses: (s_o, s_b) and of the error kurtosis (k_o, k_b) :

$$s_d = \beta_o^{3/2}s_o - \beta_b^{3/2}s_b; \quad (21a)$$

$$k_d = \beta_o^2k_o + \beta_b^2k_b. \quad (21b)$$

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 stricter lower bound provided by the positiveness of the following variance:

$$\text{var} \left(r_d + \alpha \frac{\varepsilon'_o \varepsilon'_b}{\sigma_d^2} \right) = \quad (22)$$

$$\eta_d + \beta_o\beta_b\alpha(\alpha - 4) \geq \eta_d - 4\beta_o\beta_b \geq 0, \quad \forall \alpha,$$

with r_d meaning the same as in Equation (3). This sets a necessary condition of error's independency for fixed variance fractions of errors. In all the 13 analysed samples, $\eta_d \geq 1$, and thus not imposing bounds to β_o, β_b since $\max(4\beta_o\beta_b) = 1$. From the Equations (21a), (21b) of s_d and k_d we can invert s_o and k_o , and/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 $k_b - s_b^2 + 2 \geq 0$ with the substituted s_o, s_b, k_o and k_b leads to the following bounds:

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

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

Therefore, giving consistent error variance fractions β_o, β_b and innovation statistics (s_d, k_d) , there are bounded sets for the skewness and kurtosis of errors (k_o, k_b) delimited by two parabolas of opposite concavity in the SK-domains of (s_o, k_o) and (s_b, k_b) , *i.e.* those domains delimited by inequalities (23a), (23b). We easily verify that the necessary and sufficient condition for the non-vanishing of the set of solutions sets is given by the inequality (22). The biggest of the two sets of allowable pairs (s_o, k_o) and (s_b, k_b) corresponds to the error of smallest variance fraction (β_o or β_b). This is illustrated in the example of Fig. 3 by delimiting the possible domains of (s_b, k_b) (Fig. 3a) and (s_o, k_o) (Fig. 3b), compatible with conditions: $\beta_o = 0.25, \beta_b = 0.75, s_d = 1$ and $k_d = 1$. In particular, under the above conditions, the background error, the one with largest variance, is not allowed to be Gaussian, since the pair $(s_b = 0, k_b = 0)$ is excluded from the allowed set as seen in Fig. 3a. This suggests that errors with largest variances are forbidden to be Gaussian as long as consistence with non-Gaussianity of innovations is verified. When heteroscedasticity is present, the skewnesses and kurtoses must be substituted by (19a), (19b), (20a), (20b) in the inequalities (23a), (23b).

We have analyzed the conditions for the full attribution of non-Gaussianity to one of the errors, keeping the remaining one to be Gaussian. Playing with inequalities (23a), (23b), and taking $(s_o = 0, k_o = 0)$ and/or $(s_b = 0, k_b = 0)$, we prove that the necessary condition for the existence of a Gaussian error of variance fraction β (i.e β_o and/or β_b) is given by the following third-order polynomial inequality:

$$P(\beta) \equiv \beta^3 - 3\beta^2 + \beta \left(3 + \frac{k_d}{2} \right) - \frac{\eta_d}{2} \leq 0; \quad \beta \in [0, 1]. \quad (24)$$

The variance fraction of the non-Gaussian error is $1 - \beta$. The polynomial $P(\beta)$ has a single root β_g in the interval $[0, 1]$ which is 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, (24) is equivalent to: $\beta \leq \beta_g(s_d, k_d)$. 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, given specified consistent values β_o and β_b and comparing

them to β_g , it is possible to state whether non-Gaussianity of innovations can exclusively come from one of the errors (scenario of *Single Source of non-Gaussianity*) in alternative to its sharing among errors (scenario of *Multiple Source of non-Gaussianity*). These two scenarios will be explored in sections 5 and 6 for the analysed samples. The reverse condition of 24 states that if β is the variance fraction of a certain error and $P(\beta) > 0$, then that error cannot be Gaussian or the errors are not independent.

The consistent tuned error fractions β_o, β_b depend on tuning factors f_o, f_b . This means that the partition of innovation's non-Gaussianity among errors is not independent from the problem of variance error tuning. Therefore bounds are imposed to tuning factors in the case of single attribution of non-Gaussianity. Let β_{so} (β_{sb}) be the quotient between the mean of specified observation (background) error variances and the innovation variance. Then the necessary conditions for Gaussian observation errors or equivalently the necessary condition for non-Gaussianity coming exclusively from background errors is:

$$f_o \leq (\beta_g/\beta_{so})^{1/2} \equiv f_{o1}; \quad (25a)$$

$$\beta_{sb}^{-1/2} \geq f_b \geq [(1 - \beta_g)/\beta_{sb}]^{1/2} \equiv f_{b2} \quad (25b)$$

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

$$f_b \leq (\beta_g/\beta_{sb})^{1/2} \equiv f_{b1}; \quad (26a)$$

$$\beta_{so}^{-1/2} \geq f_o \geq [(1 - \beta_g)/\beta_{so}]^{1/2} \equiv f_{o2} \quad (26b)$$

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, that requires, for most of the samples, a maximum tuning f_{o1} of the order of 0.5, thus strongly deflating 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 $[0.16, 0.54]$.

5. Error pdfs obtained by the Maximum Entropy Method

5.1. General Formulation

The application of the Maximum Entropy (ME) method is quite useful in data assimilation and inversion problems when pdfs of errors are non-Gaussian. Some applications of the ME in this context can be found in [22, 23, 24, 25]. In this section we write down explicit pdfs of errors that are consistent both with: a) the

innovation statistical moments of orders one up to four and: b) the assumed marginal moments of errors. Let us consider the joint pdf ρ_{ob} of errors in the observation space and $\rho_{o^*b^*}$ the corresponding joint pdf of the standardized errors introduced by Equations (6a), (6b). Those pdfs are related by:

$$\rho_{ob}(\varepsilon_o, \varepsilon_b) = \frac{\rho_{o^*b^*}(\varepsilon_o^*, \varepsilon_b^*)}{\sigma_{co}\sigma_{cb}}. \quad (27)$$

The least committing method of finding $\rho_{ob}(\varepsilon_o, \varepsilon_b)$ is the Maximum Entropy (ME) method [10, 11, 12], looking for the maximum of the differential joint Shannon entropy [26] of errors $H(\varepsilon_o, \varepsilon_b)$, subject to a certain set of constraints. Each set of innovations, presented before, correspond to samples of realizations of the error pairs $(\varepsilon_o, \varepsilon_b)$. Then the pdf of errors is estimated by maximizing the entropy over the each set. Assuming that all those pairs are independent from each other due to the data thinning performed previously and also that standardized errors have the same pdf over the sample, we have the sampling mean Shannon entropy written as:

$$\begin{aligned} \overline{H(\varepsilon_o, \varepsilon_b)} &\equiv \overline{\mathbb{E}[-\ln \rho_{ob}(\varepsilon_o, \varepsilon_b)]} = \\ &\ln(\sigma_{co}) + \overline{\ln(\sigma_{cb})} + H(\varepsilon_o^*, \varepsilon_b^*). \end{aligned} \quad (28)$$

where $H(\varepsilon_o^*, \varepsilon_b^*)$ is the Shannon entropy of the standardized errors. 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 f_1, \dots, f_{N_c} of the standardized errors:

$$\Omega = \{\mathbb{E}[f_k(\varepsilon_o^*, \varepsilon_b^*)] = c_k ; k = 1, \dots, N_c\}. \quad (29)$$

When H_M is finite, it is found at the global minimum of a function Γ , globally convex in terms of the pdf, of N_c Lagrange multipliers (LMs) $\lambda_1, \dots, \lambda_{N_c}$, spanning the minimization control space:

$$H_M(\varepsilon_o^*, \varepsilon_b^*) = \min_{\lambda_1, \dots, \lambda_{N_c}} \Gamma(\lambda_1, \dots, \lambda_{N_c}) \equiv \Gamma(\hat{\lambda}_1, \dots, \hat{\lambda}_{N_c}), \quad (30)$$

with:

$$\begin{aligned} \Gamma(\lambda_1 \dots \lambda_{N_c}) &\equiv \ln Z - \sum_{k=1}^{N_c} \lambda_k c_k ; Z(\lambda_1 \dots \lambda_{N_c}) \equiv \\ &\int \int \exp\left(\sum_{k=1}^{N_c} (\lambda_k f_k(\varepsilon_o^*, \varepsilon_b^*))\right) d\varepsilon_o^* d\varepsilon_b^*, \end{aligned} \quad (31)$$

where Z is the partition or normalization function. The pdf of maximum entropy (ME-pdf) belongs to the exponential family [27] and reads:

$$\rho_{o^*b^*}(\varepsilon_o^*, \varepsilon_b^*) \equiv \exp\left(\sum_{k=1}^{N_c} (\hat{\lambda}_k f_k(\varepsilon_o^*, \varepsilon_b^*))\right) / Z(\hat{\lambda}_1, \dots, \hat{\lambda}_{N_c}), \quad (32)$$

where we have used the Lagrange multipliers found at the maximum entropy solution or minimum of Γ . In general, fewer is the number N_c of constraints, 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. In our applications below, the $f_k(k = 1, \dots, N_c)$ functions are polynomials of the errors, corresponding to error and innovation moments.

5.2. Maximum-Entropy errors constrained by Innovation statistics

We will restrict our study to the case of statistically independent errors. In this case the joint entropy of standardized errors is the sum of individual entropies: $H(\varepsilon_o^*) + H(\varepsilon_b^*)$ in Equation 28. In our case, the set of constraint functions restrict to sums of individual observation and background error monomials, thus leading to separable single error integrals in the formulation of the log-partition function $\ln(Z)$. Two kinds of maximum-entropy problems are formulated bellow. The first one is the so-called problem of '*Prescribed error non-Gaussianity*'. Here, each observation and background standardized errors satisfy the $N_c = 4$ constraints:

$$\begin{aligned} \text{E}[\varepsilon_t^*] &= 0; \text{E}[(\varepsilon_t^*)^2] = 1; \text{E}[(\varepsilon_t^*)^3] = s_{ct}; \\ \text{E}[(\varepsilon_t^*)^4] &= k_{ct} + 3; t = o, b, \end{aligned} \quad (33)$$

with $s_{co}, s_{cb}, k_{co}, k_{cb}$, being defined by Equations (11a), (11b), (12a), (12b), and satisfying innovation constraints given by Equations (16), (18a). 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 (s, k) or (skewness, kurtosis) dependence of the maximum entropy H_M and of the four Lagrange multipliers (LM) λ_p ($p=1, \dots, 4$), corresponding to the powers p in Equation 33 is discussed by [15] as well in the appendix.

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 Equations (21a), (21b). In this case we get a total of $N_c = 6$ constraints terms in the standardized errors:

$$\text{E}[\varepsilon_t^*] = 0; \text{E}[(\varepsilon_t^*)^2] = 1, \text{ with } t = o, b; \quad (34a)$$

$$\text{E}\left[\varepsilon_o^{*3}\beta_o^{3/2} - \varepsilon_b^{*3}h_{3b}\beta_b^{3/2}\right] = s_d; \quad (34b)$$

$$\text{E}\left[\varepsilon_o^{*4}\beta_o^2 + \varepsilon_b^{*4}h_{4b}\beta_b^2\right] = k_d + 3(\beta_o^2 + \beta_b^2), \quad (34c)$$

to which correspond 6 Lagrange multipliers: $\lambda_{o1}, \lambda_{o2}, \lambda_{b1}, \lambda_{b2}, \lambda_{d1}, \lambda_{d2}$ 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:

$$\begin{aligned}
\Gamma(\lambda_{o1}, \lambda_{o2}, \lambda_{b1}, \lambda_{b2}, \lambda_{d3}, \lambda_{d4}) = & \\
\sum_{t=b,o} \ln \left[\int \exp(\lambda_{t1}\varepsilon + \lambda_{t2}\varepsilon^2) d\varepsilon \right] + & \\
\ln \left\{ \int \exp \left[\lambda_{d3}\beta_o^{3/2}\varepsilon^3 + \lambda_{d4}\beta_o^2\varepsilon^4 \right] d\varepsilon \right\} + & \quad (35) \\
\ln \left\{ \int \exp \left[-\lambda_{d3}\beta_b^{3/2}h_{3b}\varepsilon^3 + \lambda_{d4}\beta_b^2h_{4b}\varepsilon^4 \right] d\varepsilon \right\} - & \\
\left\{ \lambda_{b2} + \lambda_{o2} + s_d\lambda_{d3} + \lambda_{d4} \left[k_d + 3(\beta_o^2 + \beta_b^2) \right] \right\}. &
\end{aligned}$$

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

$$\rho_{t^*}(\varepsilon) = \frac{\exp(Q(\varepsilon))}{\int \exp(Q(u)) du} \quad (36a)$$

$$Q(\varepsilon) = \hat{\lambda}_{t1}\varepsilon + \hat{\lambda}_{t2}\varepsilon^2 + \hat{\lambda}_{d4}\beta_t^2h_{4t}\varepsilon^4 \pm \hat{\lambda}_{d3}\beta_t^{3/2}h_{3t}\varepsilon^3 \quad (36b)$$

Positive and negative signs in (36b) 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^* . The LMs are those minimizing the function (35).

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 Equation (5) and subject to constraints given by Equations (21a), (21b), thus leading to:

$$(s_o, s_b) \approx \frac{s_d}{\beta_o^3 + \beta_b^3} \left(\beta_o^{3/2}, -\beta_b^{3/2} \right); \quad (37a)$$

$$(k_o, k_b) \approx \frac{k_d}{\beta_o^4 + \beta_b^4} \left(\beta_o^2, \beta_b^2 \right). \quad (37b)$$

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 Figs. 3a, 3b, 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. 3a) and of (s_o, k_o) (Fig. 3b). The field of H_M has a minimum of ~ 1.4 nats at the edge of the allowable SK domains. Nat is the natural logarithmic unit of Shannon entropy and

information with the conversion $1 \text{ nat} = \log_2 e \text{ bit} \approx 1.44 \text{ bit}$. The maximum of $H_M \sim 2.55 \text{ nats}$, lies at the interior of the domain, at $(s_o, k_o) \sim (0.65, 0.71)$ and $(s_b, k_b) \sim (-1.41, 1.70)$ whereas Equations (37a),(37b), 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 [28], also used in [29]. The support of ME-pdfs is the finite interval $[-20, 20]$, which seems satisfactory in our applications. Numerical details are given in the appendix.

6. Minimum Variance and Best Linear Unbiased Estimators

6.1. Generic Properties

When observation and background errors are statistically independent, as previously assumed, then the *Minimum Variance Unbiased Estimator* (MVUE) x_M , issued from an observation y and a background estimate $x_b = y - d$ is given through the convolution of the marginal error pdfs in the following form:

$$x_M = x_b - b_{cb} + \Delta_{Mb} = x + \varepsilon_M ; \quad \varepsilon_M = \varepsilon_b + \Delta_{Mb}; \quad (38a)$$

$$\begin{aligned} \Delta_{Mb} \equiv -\text{E}(\varepsilon_b|d) = \\ -\sigma_{cb} \frac{\int u \rho_{b*}(u) \rho_{o*} [(d - b_d + u\sigma_{cb}) \sigma_{co}^{-1}] du}{\int \rho_{b*}(u) \rho_{o*} [(d - b_d + u\sigma_{cb}) \sigma_{co}^{-1}] du}, \end{aligned} \quad (38b)$$

where the ME-pdfs of errors have been used and x is the true state value. An alternate estimator to the MVUE is the *Maximum A Posteriori* (MAP) corresponding to the absolute mode of the *a posteriori* pdf. The error of the MVUE is $\varepsilon_M = x - x_M$ and Δ_{Mb} is the correction to be added to the unbiased background in order to obtain MVUE. We notice that Δ_{Mb} is the symmetric of the conditional mean of ε_b for a given innovation value d . The normalization integral in denominator of Equation (38b) is proportional to the estimated conditional-innovation theoretical pdf $\rho_d(d|\sigma_{cb})$, inferred from ME-pdfs of errors and computed at the particular value of d and σ_{cb} . Then the innovation pdf over a certain set must be approached by the mean of $\rho_d(d|\sigma_{cb})$ running over all the innovations of that set. This pdf is comparable to the empirical raw innovation pdf, thus providing another consistency test beyond those given by the skewness and kurtosis, as discussed in Section 3. However, the mean theoretical pdf does not necessarily coincide with the ME-pdf constrained by innovation statistics only.

The mean square of the MVUE error is given by the conditional variance of ε_b computed at the innovation value d , *i.e.* $\text{E}(\varepsilon_M^2|d) = \text{var}(\varepsilon_b|d)$. 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 conditional *a posteriori* Shannon entropy [30]. 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 *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:

$$\begin{aligned} \mathbb{E}_d(\varepsilon_M^2) &\equiv \sigma_M^2 = \mathbb{E}_d[\mathbb{E}(\varepsilon_M^2|d)] = \\ \sigma_b^2 - \mathbb{E}_d[\mathbb{E}(\varepsilon_b|d)^2] &\leq \min(\sigma_b^2, \sigma_o^2), \end{aligned} \quad (39)$$

where the expectation \mathbb{E}_d is the average running through a large sample of innovations. When errors are Gaussian, the MVUE is the classical Best Linear Unbiased Estimator (BLUE), [31] given by:

$$x_L = x_b - b_{cb} + \Delta_{Lb} = x + \varepsilon_L; \varepsilon_L = \varepsilon_L + \Delta_{Lb}; \quad (40a)$$

$$\Delta_{Lb} \equiv \frac{\sigma_{cb}^2}{\sigma_{cb}^2 + \sigma_{co}^2} (y - x_b - b_d), \quad (40b)$$

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:

$$\Delta_{ML} \equiv x_M - x_L = \Delta_{Mb} - \Delta_{Lb} = \varepsilon_M - \varepsilon_L, \quad (41)$$

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:

$$\mathbb{E}_d(\varepsilon_L^2) \equiv \sigma_L^2 = \sigma_d^2 \beta_o \beta_b = \sigma_M^2 + \mathbb{E}_d(\Delta_{ML}^2) \geq \sigma_M^2, \quad (42)$$

which shows that a self-consistent assimilation scheme based on the non-Gaussian MVUE must in general perform better than that based on 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 [32]:

$$\text{MSESS} \equiv \frac{\mathbb{E}_d(\Delta_{ML}^2)}{\sigma_L^2} = 1 - \frac{\sigma_M^2}{\sigma_L^2} \in [0, 1]. \quad (43)$$

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: $\beta_o, \beta_b = 1 - \beta_o$ 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. Finally, values of the MSESS are assessed for the analysed samples of brightness temperatures.

6.2.1. Single-source of non-Gaussianity

We assess the MSESS, defined by Equation (43) when only one of the errors is non-Gaussian. Without loss of generality we will attribute all non-Gaussianity to the background error of variance β_b for which: $(s_b, k_b) = (-s_d/\beta_b^{3/2}, k_d/\beta_b^2)$. Conversely, the observation error has variance $\beta_o = 1 - \beta_b$. The MSESS contours for the case of equal error variances: $\beta_o = \beta_b = 0.5$ are shown in Fig. 5a 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. 5a), 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. 5b 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 (exact estimation) when β_o tends to 0 or 1, corresponding respectively to exact observation and exacto 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 Gaussian error variance β_o (the Non-Gaussian error variance β_b) at the MSESS maximum, slightly decreases (increases) as far as the error absolute 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. 5b). 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 (Figs. 6a-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 Figs. 6a-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. 7a, homologous to Fig. 5a (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 compensate or have opposite skewness signs (left side of Fig. 7a), as compared to the case of equal signs (right side of the Fig. 7a). 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 corresponding *a posteriori* pdfs can be completely different. The Fig. 7b is homologous to Fig. 5b for a fixed $k_b = 3$. There, MSESS increase when one approaches the limits $s_b = \pm(k_b + 2)^{1/2}$. For s_b, s_o of the same sign (right side of Fig. 7b), 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. 7b), 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. 7b), 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 between the MVUE estimator and the BLUE (MSESS=0). This is illustrated in MSESS contouring of Fig. 3c 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. The ME-solution, maximizing entropy, is not necessarily the one minimizing MSESS as it is clear for the case discussed in Fig. 3c. 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 error, $H_M \sim 2.45$ nats, $(s_o, k_o) = (0, 0)$, and the MSESS is ~ 0.25 . Nevertheless, the value of MSESS is not in general a monotonous decreasing function of H_M .

Contours of MSESS over the (s_d, k_d) plane are shown in Figs. 8a, 8b, for $\beta_o = 0.25$ and 0.4 respectively. We notice, in the above figures, that k_d has a lower bound given by Equation (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 compensate themselves, and then reaching a maximum at a specific intermediate $|s_d|$ value and then decreases towards the SK domain edge (Figs. 8a,8b). In Figs. 8c-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 $\text{MSESS} \lesssim 0.01$. 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.

The Figures 9a-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 . Figs. 9a,c,e refer to s_o and Figs. 9b,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.

6.3. 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, 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:

$$\begin{aligned} \text{MSC}_j &= \sigma_d^{-2} \overline{\Delta_{ML}^2 | d^* \in I_j}; \\ I_a &= [0, 1[; I_b = [1, 2[; I_c =]2, \infty[. \end{aligned} \tag{44}$$

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. In the present study the analyzed 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. 5a a small MSESS is expected as compared to the cases of extreme non-Gaussianity shown in previous sections. Therefore, the relative frequencies p_a , p_b and p_c of the classes I_a , I_b and I_c defined, in (44), must be close to those given by the standard Gaussian, *i.e.* $p_a \sim 68\%$, $p_b \sim 26\%$ and $p_c \sim 6\%$, corresponding to conditional averages of $|d^*|$: $q_a \sim 0.4$, $q_b \sim 1.5$ and $q_c \sim 2.5$ respectively for classes I_a , I_b and I_c . The MSC_j ($j = a, b, c$) is measured in terms of the innovation variance. Through (43), one shows that the average MSC and the MSESS values are given as:

$$\text{MSC} = \sum_{j=a,b,c} p_j \text{MSC}_j; \tag{45a}$$

$$\text{MSESS} = \frac{\text{MSC}}{\beta_o \beta_b}. \tag{45b}$$

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 the referred 12 samples, we present logarithmic graphics of the scores defined by Equation 44 in the Figures Aa-l, 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 (25), (26) for the scenarios 1 and 2. All in all, there are 9 score graphics per sample in each of Figure Aa-l. 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. 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 to be higher for the most extreme values of the absolute innovation $|d|$, *i.e.* $MSC_a < MSC_b < MSC_c$. For the scenarios of *Single Source of non-Gaussianity* (scenarios 1 and 2) and certain 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 $\sim 6\%$ of the cases. This is particularly evident in some ‘ice-free’ samples namely: 4s, 5s, 6s, 11s, 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 $\sim 0.31\sigma_d$, 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.

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 Figs. 6a to 6c. There are quite local exceptions in samples 7s and 12s. In the scenario 3 (*Maximum entropy sharing of non-Gaussianity*), 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 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 Figs. 8c to 8e 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, 5s, 7s and all ‘ice-covered’ samples, the absence of background error tuning, *i.e.* $f_b = 1$ do not match with an exclusive non-Gaussian background error. 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 a case for the most skewed sample of innovations (4s), where $s_d = -0.70$, $k_d = 1.02$ and the heteroscedasticity is quite small (Cf. Fig. 2). The error variance partition is: $\beta_b = 0.47$, $\beta_o = 0.53$, corresponding to tuning factors $f_b = 1.50$ and $f_o = 0.22$. We assume non-Gaussianity coming exclusively from background errors, hence leading to $s_b = 2.18$, $k_b = 4.65$, $s_o = 0$, $k_o = 0$. The corresponding ME-pdfs of ε_b and ε_o (Gaussian) are plotted in Figure 11(a), from which the positively skewed tail of background errors is quite clear. In Figure 11(b), 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 Equation (40b). 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 MVUE-correction: Δ_{Mb} is quite enhanced, whereas for positive d a reduced correction is expected as shown in Fig. 11(b). 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 $\text{sgn}(d) = \text{sgn}(\sigma_d)$. For a certain value d , a relatively marginal correction holds when: $\text{sgn}(d) = -\text{sgn}(\sigma_d)$.

7. Discussion and Conclusions

Histograms of innovations for certain assimilated observables (e.g. brightness temperature obtained from HIRS channels) reveal statistically significant non-Gaussianity, as far as skewness and kurtosis are concerned. Non-Gaussianity of innovations can be due to: marginal non-Gaussian observation or background errors, non-linear error correlations and ultimately to heteroscedasticity of errors within the sample used to estimate innovations statistics. The origin of non-Gaussianity can only be determined by independently tested hypothesis that would suggest the contributions of the instruments, observation representativeness, models and assimilation algorithms. When non-Gaussianity of innovations essentially comes from one of the errors, then its variance must be sufficiently high in order to accommodate it. This imposes some bounds on the error variance tuning. In some cases, it leads to incompatibility between absence of tuning and specific

sources of non-Gaussianity, *i.e.* variance tuning is compulsory while taking into account non-Gaussianity in the data assimilation schemes.

The least committing method of taking into account non-Gaussianity in the error's pdfs is the Maximum Entropy method. The resulting pdfs are then used to get the Minimum Variance Unbiased Estimator (MVUE) of the system state in the observational space. We analyse which conditions favour the largest increment of performance when compared to the Best Linear Unbiased Estimator (BLUE), vastly used in data assimilation. This relative performance is measured here by the MSESS (Mean Square Error Skill Score), ranging between 0 (no added value) up to 1 (total correction). The MSESS is particularly large, *i.e.*, the role of non-Gaussianity is high, when errors from a priori estimates (observations, backgrounds etc.), are skewed and in majority towards the same side. Compensation of error skewnesses degrades the performance of non-Gaussian data assimilation schemes. Moreover, in what concerns data errors and their non-Gaussianity impact: 'Large skewness put into a small variance' has a larger impact than 'Small skewness put into a large variance'. In the present paper, we have applied the Maximum Entropy method to compute error pdfs that are consistent with innovation statistics of the brightness temperature, obtained from HIRS channels. The largest difference between the MVUE estimator and the BLUE estimator holds at largest absolute innovations, typically larger than two standard deviations σ_d of innovations and reaching 10% – 60% of σ_d . This can eventually be applied in operational data assimilation schemes. The first step would be the computation of analysis obtained by the MVUE in the observation space which in turn would be re-assimilated as corrected unbiased data and also with the effect of non-Gaussianity of errors taken into account.

The implementation of non-Gaussian error models into data assimilation is still in its birth. Many aspects still require new ideas such as how to generalize non-Gaussian error models to the whole innovation vector. There, models for multivariate skewnesses and kurtoses [33] have to be taken into account and also the effect of spatial error correlations. Another goal is how to state error's non-Gaussianity in the spectral domain, often used by numerical forecasting models and how to compute it from the ensemble methods along with the background error covariance matrices and cumulant tensors of order three and four. Another possible future extension is the addition of penalty terms to data assimilation cost-functions, weighting the deviation of error's pdfs from Gaussianity. A systematic similar study to innovations of other observable types is in mind.

8. Acknowledgements

This research was developed at IDL and LMD with support from the project: 'Développements méthodologiques 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 and Rui Perdigão for their comments and help.

A. Appendix - Maximum entropy pdfs constrained by skewness and kurtosis

In order to study the change of the Lagrange multipliers (LMs) and their ME-pdfs, we have solved the ME problem in the SK-domain for skewness $s \in [0, 4.5]$ and kurtosis $k \in [-2, 17]$ with steps of 0.05 in each parameter, using the pdf support set $D = [-L, L] = [-20, 20]$ for numerical computations. In Figs. 12a-e we plot the contour lines, respectively of the Lagrange multipliers (LMs): $\lambda_1, \dots, \lambda_4$ and of the Shannon entropy, as function of the imposed skewness and kurtosis. The LMs correspond to successive moments ($p = 1, 2, 3, 4$). In Fig. 12f, we show the ME-pdfs for some typical pairs (s, k) . The λ -fields are smoothly varying over the SK-domain except when they approach the boundary curve $k = s^2 - 2$ where both the absolute values of the LMs and the negentropy tend to increase. There, the ME-pdfs exhibit very sharp picks. For symmetry reasons, the odd order λ -fields are anti-symmetric with respect to the axis $s = 0$ and the even order ones are symmetric. There are singular (s, k) points in which the contour lines of certain λ -fields tend to merge, as for example $(s, k) = (0, 1)$ for λ_1 and λ_3 and $(s, k) \sim (1.5, 0)$ for λ_2 . The ME field, in Fig. 12e, gets the maximum $H_G = 0.5 \ln(2\pi e)$ nats at the Gaussian point $(s, k) = (0, 0)$, decreasing from it as far the parabolic limit curve $k = s^2 - 2$ is approached, both in the skewness and the kurtosis axes. When the support half-extent $L \rightarrow \infty$, there are no ME-pdfs with $s = 0$ and $k > 0$ [34]. As consequence of finite L , $\lambda_2 \sim -0.5$ and $\lambda_1 \sim \lambda_3 \sim \lambda_4 \sim 0$ at the semi-straight line $(s = 0, k > 0)$ as seen in Figs. 12a-e. In spite of that, there are leptokurtic unskewed exponential pdfs ($s = 0, k > 0$) of the type $\rho(\varepsilon) = C \exp(-|\varepsilon|^\alpha)$, $\alpha \geq 2$, not satisfying the ME requirement. In fact, at $s = 0$, the negentropy approximation of a ME-pdfs, given in Equation (5), is only a good approximation for $k < 0$.

We have plotted in Fig. 12f, the ME-pdfs at four typical points in the SK domain: G(Gaussian): $(s, k) = (0, 0)$; SM(Skewed-Mesokurtic): $(s, k) = (1, 1)$; SP(Skewed-Platykurtic): $(s, k) = (1, -1)$; SL(Skewed Leptokurtic): $(s, k) = (3.5, 12)$. The pdf of SM is very close to Gaussian as well as those with small s and high kurtosis k , since H_M is very close to H_G (Fig. 12e). The pdf of SP is platykurtic with two modes and lower density near the origin. The pdf of SL is spiky near the origin with an extended tail of positive extreme anomalies and a small hump accommodating the positive skewness. For illustration purposes, we add in Fig. 2, both the non-Gaussian ME-pdf fit and the Gaussian pdf fit of the innovation histogram for the sample 4s.

Asymptotically converging solutions to the ME-pdfs are obtained by solving the ME problem in finite interval sets $D = [-L, L]$ of increasing length L up to stabilizing both the ME and the LMs. 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 :

$$\frac{dH_M(\varepsilon)}{dL} = \rho(L) + \rho(-L), \quad (46)$$

where ρ is the ME-pdf. The logarithm $\ln \rho$ at the interval bounds is scaled as $O(-L^4)$, decreasing as far as 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 scaling above 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:

$$H_M(\varepsilon) = H_M(\varepsilon/L) + \log(L), \quad (47)$$

where $H_M(\varepsilon/L)$ is the entropy of the scaled error. The LMs also depend on the interval amplitude L , but stabilizing for L high enough. After differentiating the expectation constraints for fixed moments c_k of order $k = 1, 2, 3, 4 = N_c$, we get:

$$\begin{aligned} \frac{d\lambda_k}{dL} &= \sum_{i=1}^{N_c=4} (M^{-1})_{k,i} \\ &\times \left[\rho(L) (L^i - c_i) + \rho(-L) \left\{ (-L)^i - c_i \right\} \right]; \quad k = 1, \dots, 4, \end{aligned} \quad (48)$$

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 $M_{k,i}^{-1}$ in (48) 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 [35], 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. Considering 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: V. S. R. Swinbank, 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) 12253–12265.
- [3] H. Elbern, H. Schmidt, Ozone episode analysis by four-dimensional variational chemistry data assimilation, *J. Geophys. Res.* 106 (2001) 3569–3590.
- [4] N. B. Ingleby, A. C. Lorenc, Bayesian quality control using multivariate normal distributions, *Quart. J. Roy. Meteor. Soc.* 119 (1993) 1195–1225.
- [5] S. J. Fletcher, M. Zupanski, A hybrid normal and lognormal distribution for data assimilation, *Atmos. Sci. Lett.* 7 (2006) 43–46.
- [6] S. J. Fletcher, M. Zupanski, A data assimilation method for log-normally distributed observational errors, *Quart. J. Roy. Meteor. Soc.* 132 (2006) 2505–2519.
- [7] M. Bocquet, High resolution reconstruction of a tracer dispersion event: application to etex, *Quart. J. Roy. Meteor. Soc.* 133 (2007) 1013–1026.
- [8] D. P. Dee, Bias and data assimilation, *Quart. J. Roy. Meteor. Soc.* 131 (2005) 3323–3343.
- [9] E. Holm, Revision of the ecmwf humidity analysis: Construction of a gaussian control variable, in: *Proceedings of the ECMWF/GEWEX Workshop on Humidity Analysis, ECMWF/GEWEX, Reading, 2002*, pp. 1–6.
- [10] E. T. Jaynes, Information statistics and statistical mechanics i, *Phys. Rev.* 106 (1957) 620–.
- [11] E. T. Jaynes, Information statistics and statistical mechanics ii, *Phys. Rev.* 108 (1957) 171–.
- [12] T. M. Cover, A. Thomas, *Elements of Information Theory*, Wiley Series on Communications, 1991.
- [13] J. Derber, F. Bouttier, A reformulation of the background error covariance in the ecmwf global data assimilation system, *Tellus A* 51 (1999) 195–222.
- [14] 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, *Technical memorandum - 296, ECMWF, [Available from ECMWF.]* (1999).
- [15] X. Wu, Calculation of maximum entropy densities with application to income distribution, *Journal of Econometrics* 115 (2003) 347–354.
- [16] P. Comon, Independent component analysis, a new concept?, *Signal Processing* 36 (1994) 287–314.
- [17] J. F. Kenney, E. S. Keeping, Cumulants and the cumulant-generating function, additive property of cumulants, and sheppard’s correction, in: J. F. Kenney (Ed.), *Mathematics of Statistics*, Van Nostrand, 1951, pp. 77–82.
- [18] G. Maurice, A. Stuart, J. K. Ord, A. O’Hagan, *Kendall’s Advanced Theory of Statistics*, 6th edition, Edward Arnold, 1999.
- [19] G. Desroziers, S. Ivanov, Diagnosis and adaptative tuning of information error parameters in variational assimilation, *Quart. J. Roy. Meteor. Soc.* 127 (2001) 1433–1452.
- [20] B. Chapnik, G. Desroziers, F. Rabier, O. Talagrand, Properties and first application of an error-statistics tuning method in variational assimilation, *Quart. J. Roy. Meteor. Soc.* 130 (2004) 2253–2275.
- [21] V. Rudin, *Real and Complex Analysis*, McGraw-Hill, 1987.
- [22] B. Urban, Humidity profile retrieval using a maximum entropy principle, *Mon. Wea. Rev.* 123 (1995) 1185–1196.
- [23] M. Bocquet, Reconstruction of an atmospheric tracer source using the principle of maximum entropy I, *Quart. J. Roy. Meteor. Soc.* 131 (2005) 2191–2208.
- [24] M. Bocquet, Reconstruction of an atmospheric tracer source using the principle of maximum entropy II, *Quart. J. Roy. Meteor. Soc.* 131 (2005) 2209–2223.
- [25] M. Bocquet, Inverse modelling of atmospheric tracers: non-gaussian methods and second-order sensitivity analysis, *Nonlin. Processes Geophys.* 15 (2005) 127–143.
- [26] C. E. Shannon, The mathematical theory of communication, *Bell Syst. Techn. Journal* 27 (1948) 379–423.

- [27] O. Barndorff-Nielsen, Information and exponential families in statistical theory, Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons Ltd., 1978.
- [28] M. Rockinger, E. Jondeau, Entropy densities with an application to autoregressive conditional skewness and kurtosis, *Journal of Econometrics* 106 (2002) 119–142.
- [29] C. Pires, R. Perdigão, Non-gaussianity and asymmetry of the winter monthly precipitation estimation from nao, *Mon. Wea. Rev.* 135 (2007) 430–448. doi:10.1175/MWR3407.1.
- [30] 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), APMG, Lisboa, 2003, pp. 220–225.
- [31] A. Gelb, Applied Optimal Estimation. The Analytical Sciences Corporation, MIT Press, 1974.
- [32] M. Déqué, Continuous variables, chapter 5, in: I. Jolliffe, D. Stephenson (Eds.), *A Practitioner’s Guide in Atmospheric Science*, John Wiley and Sons, England, 2003, pp. 97–119.
- [33] I. Takafumi, On a measure of multivariate skewness and a test for multivariate normality, *Annals of the Institute of Statistical Mathematics* 34 (1982) 531–541. doi:10.1007/BF02481051.
- [34] D. Ormoneit, H. White, An efficient algorithm to compute maximum entropy densities, *Econometric Reviews* 18 (1999) 127–140.
- [35] J. C. Gilbert, C. L. Maréchal, Some numerical experiments with variable storage quasi-newton algorithms, *Math. Program.* 45 (1989) 407–435.

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 $\text{std}(\sigma_{sb})$ 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_o	0.6	0.6	0.6	0.75	1.	2.	0.5	0.6	0.6	0.6	0.75	0.5	0.6
$m(\sigma_b)$	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
$\text{std}(\sigma_{sb})$	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

Specified background error standard deviation (K)

HIRS - Channel 6

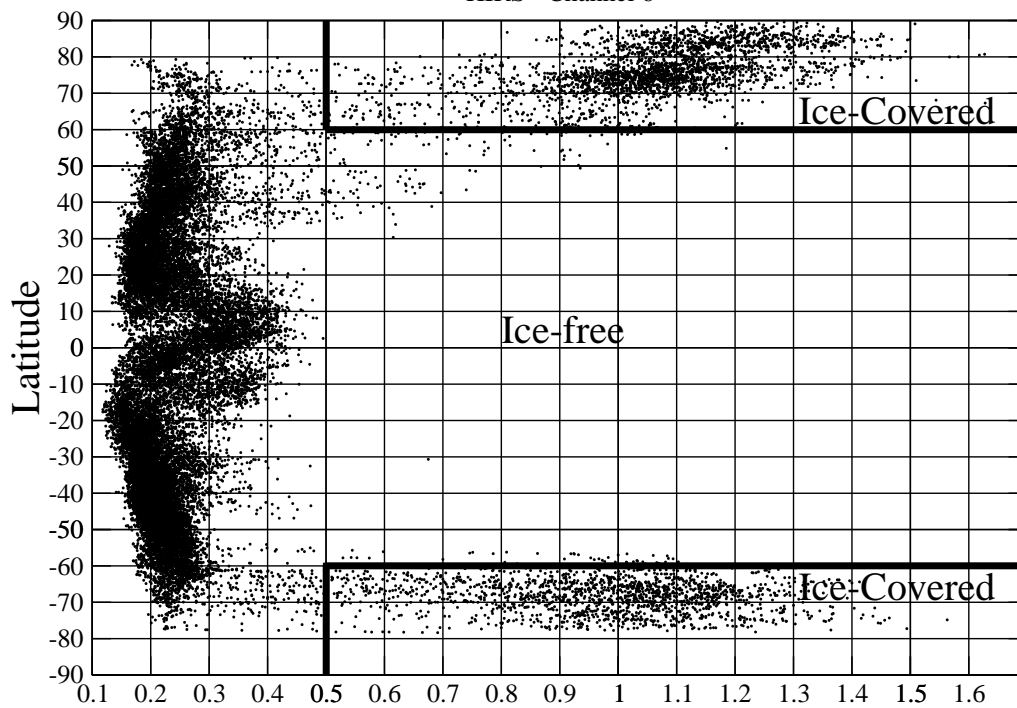


Figure 1: Scatter plot of pairs: (specified background error standard deviation σ_{sb} , North latitude) for channel 6. Boxes delimit samples corresponding to 'Ice-free' and 'Ice-covered' conditions.

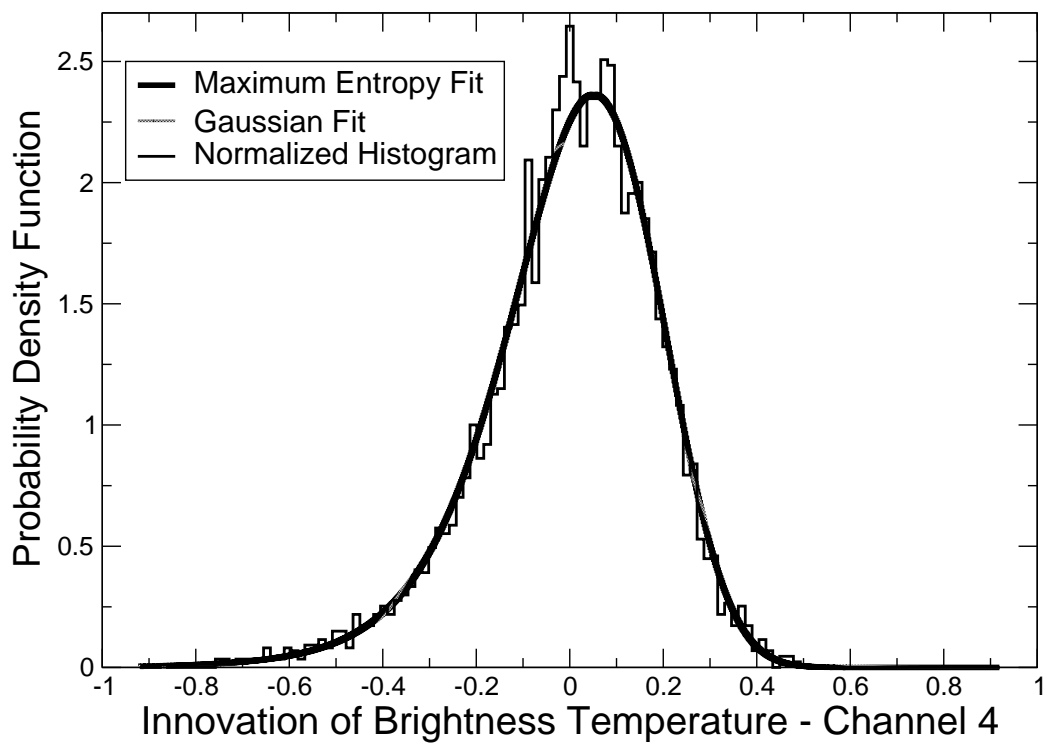


Figure 2: Normalized histograms (in stairs), corresponding Gaussian pdf fit (grey solid line), and Maximum Entropy pdf fit (dark solid line) for the sample 4s of innovations.

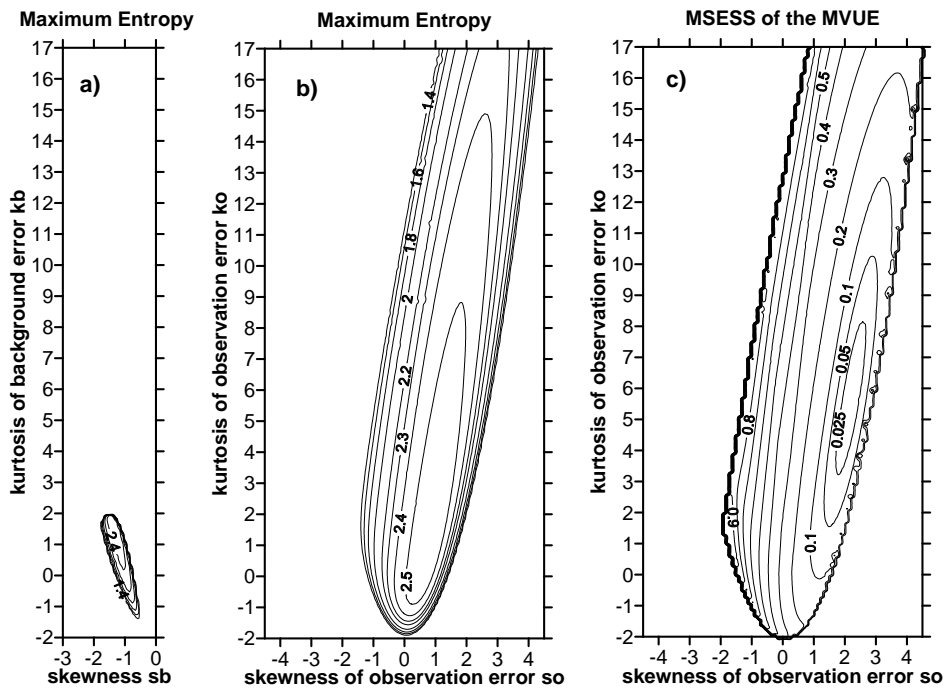


Figure 3: (a) Contour field of the joint maximum entropy (in nats) of background and observation errors for the allowable domain set in the (s_b, k_b) plane; (b) same as (a) in the (s_o, k_o) plane; (c) MSESS of the MVUE in the (s_o, k_o) plane. In all experiments we set $s_d = 1$, $k_d = 1$ and error variance fractions: $\beta_o = 0.25$, $\beta_b = 0.75$.

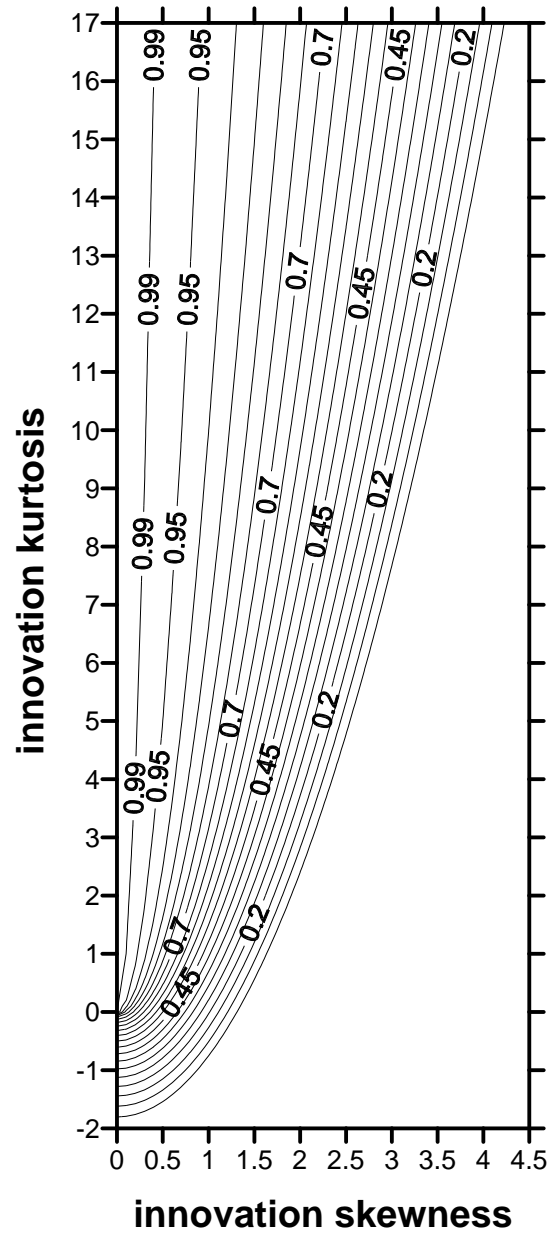


Figure 4: Contour map of the maximum variance fraction β_g of a Gaussian error, as function of the innovation skewness s_d and kurtosis k_d . The contour map is symmetric for $s_d < 0$

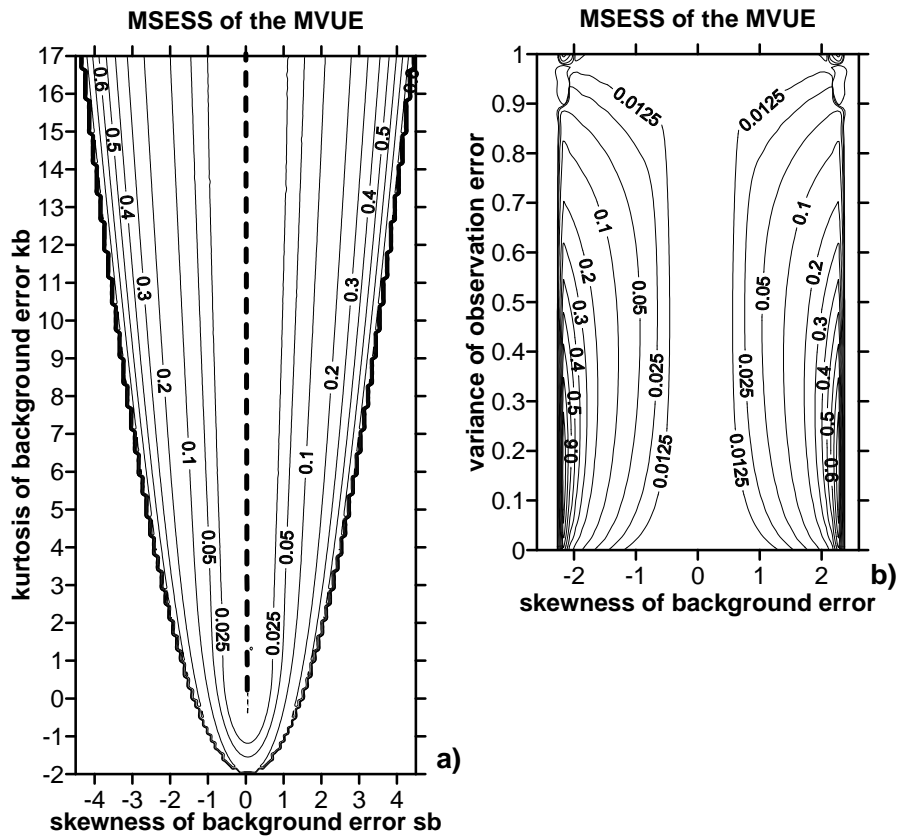


Figure 5: Contour field of the MSESS of the MVUE obtained from ME-pdfs with a Gaussian error: ($s_o = 0$, $k_o = 0$) and a non-Gaussian error with skewness s_b , and kurtosis k_b . a) Cross section along the plan (s_b, k_b) for a fixed variance fraction $s_o = 0.5$. The semi-straight line corresponding 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 $k_b = 3$.

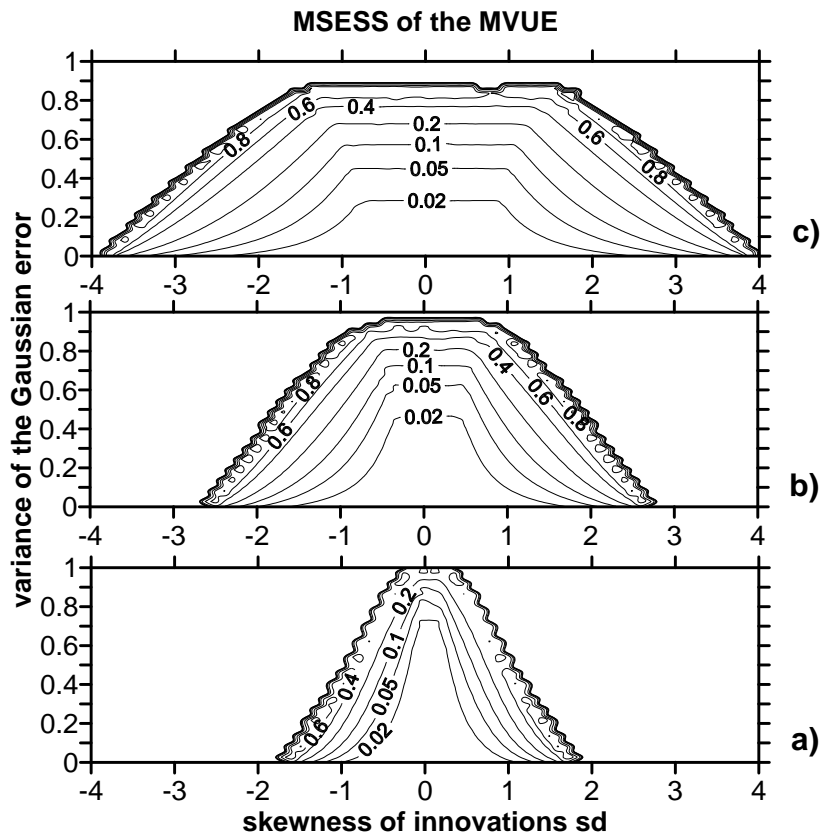


Figure 6: Contour field of $MSESS$ of the MVUE under the scenario of *Single Source of non-Gaussianity* in the plane determined by the variance of the Gaussian error β_o versus innovation skewness s_d (a) case for fixed innovation kurtosis $k_d = 1$, (b) same as (a) for $k_d = 5$, (c) same as (a) for $k_d = 13$. 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).

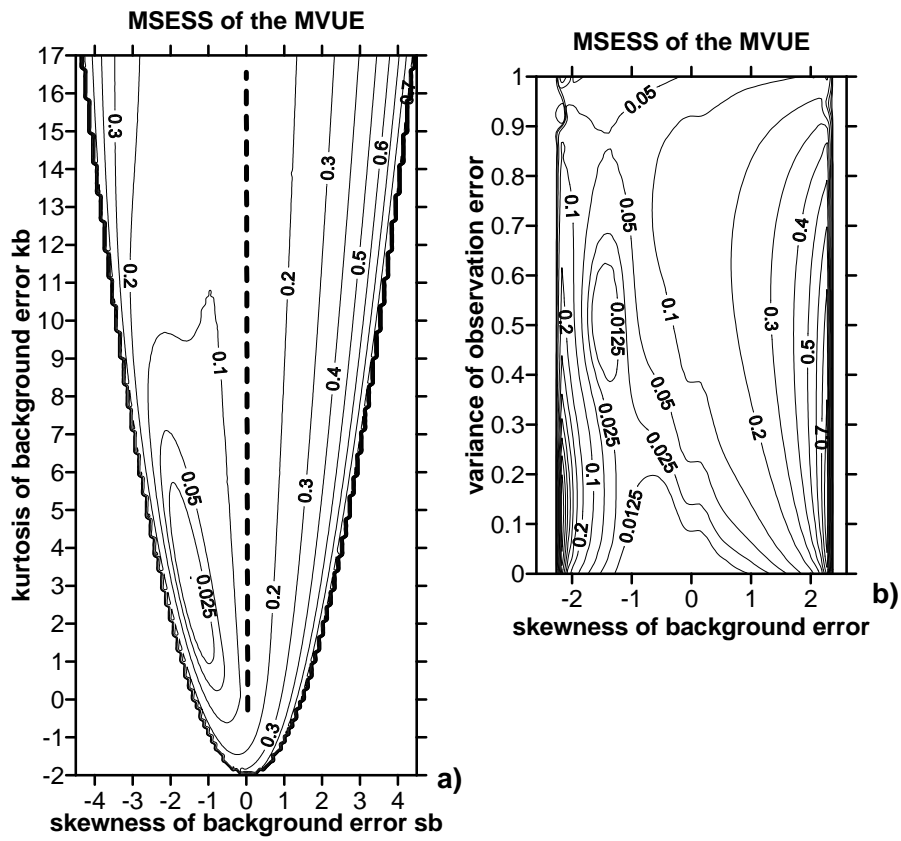


Figure 7: (a) The same as Figure 5a fixing skewness and kurtosis of observation errors to: $s_o = 1.5$ and $k_o = 3$ and error variance fraction $\beta_o = 0.5$ (b) The same as Figure 5b with the same values: $s_o = 1.5$ and $k_o = 3$

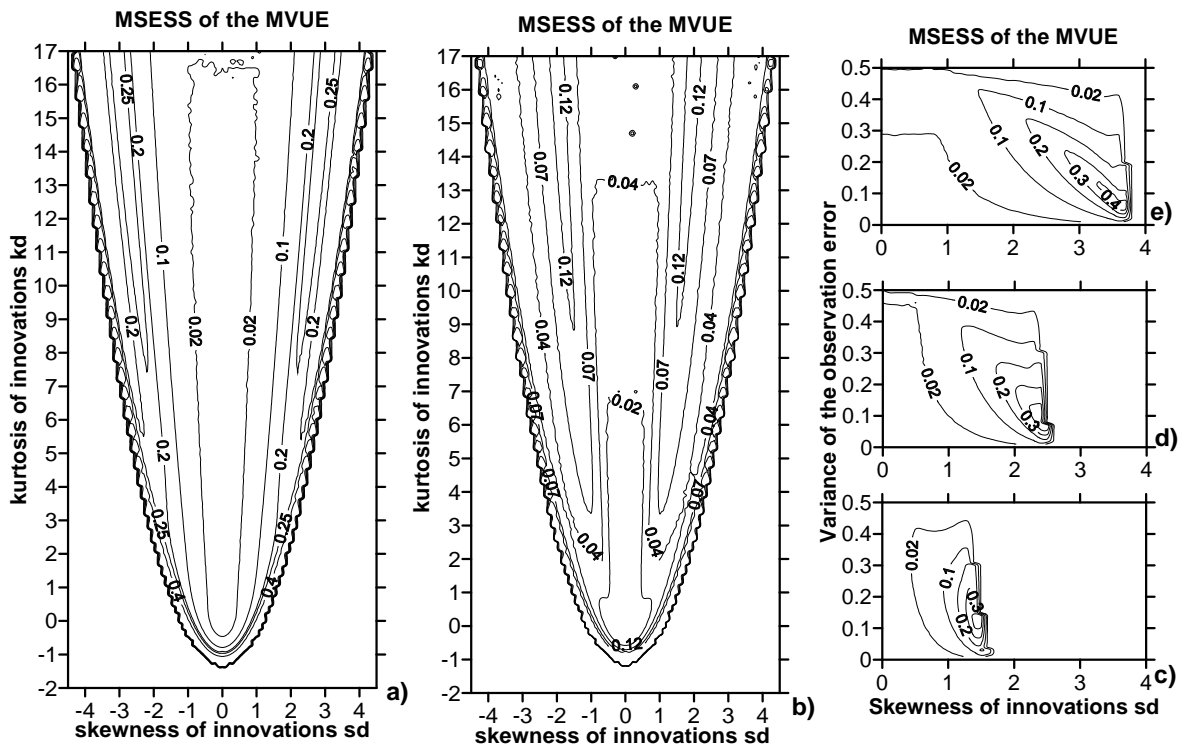


Figure 8: Contour field of the *MSESS* of the MVUE for the scenario of *Maximum entropy sharing of non-Gaussianity*. (a) For fixed $\beta_o = 0.25$ and varying (s_d, k_d) , (b) Similar to (a) with $\beta_o = 0.40$, (c) For fixed $k_d = 1$ and varying (β_o, s_d) , (d) Similar to (c) with $k_d = 5$; (e) Similar to (c) with $k_d = 13$.

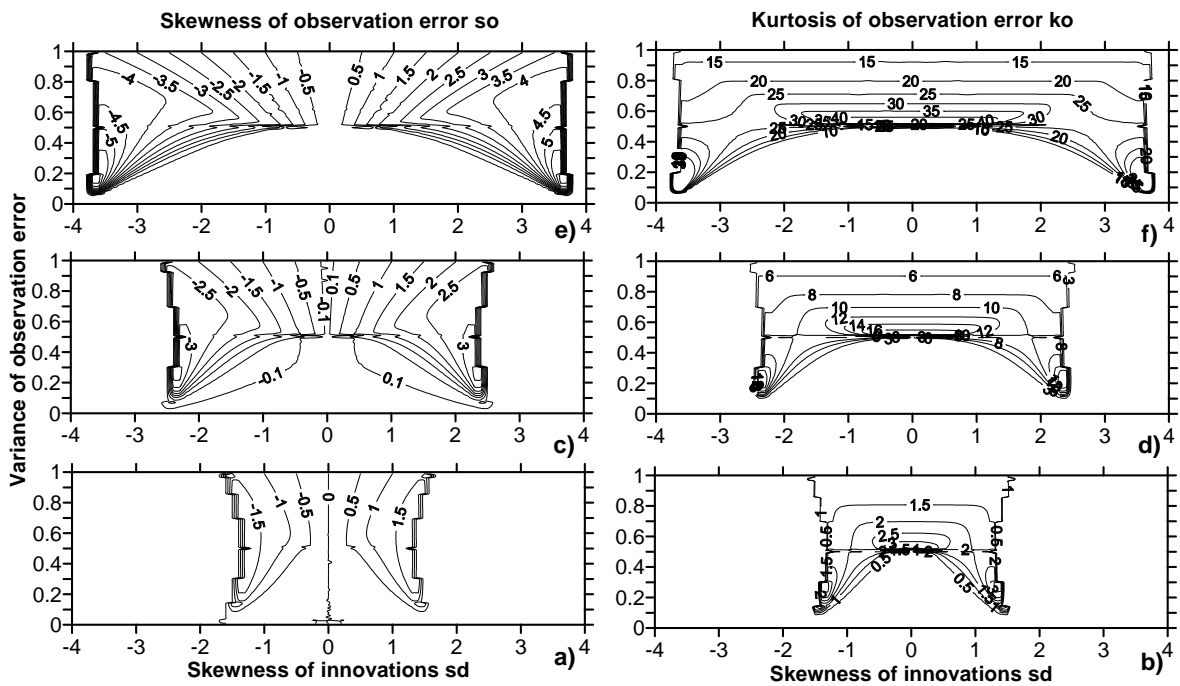


Figure 9: Contour field of s_o (Figures a,c,e) and k_o (Figures b,d,f) for the scenario of *Maximum entropy sharing of non-Gaussianity* as function of the innovation skewness s_d and the error variance fraction β_o for fixed values of the innovation kurtosi (k_d). Figures a, b for $k_d = 1$. Figures c, d for $k_d = 5$. Figures e, f for $k_d = 13$.

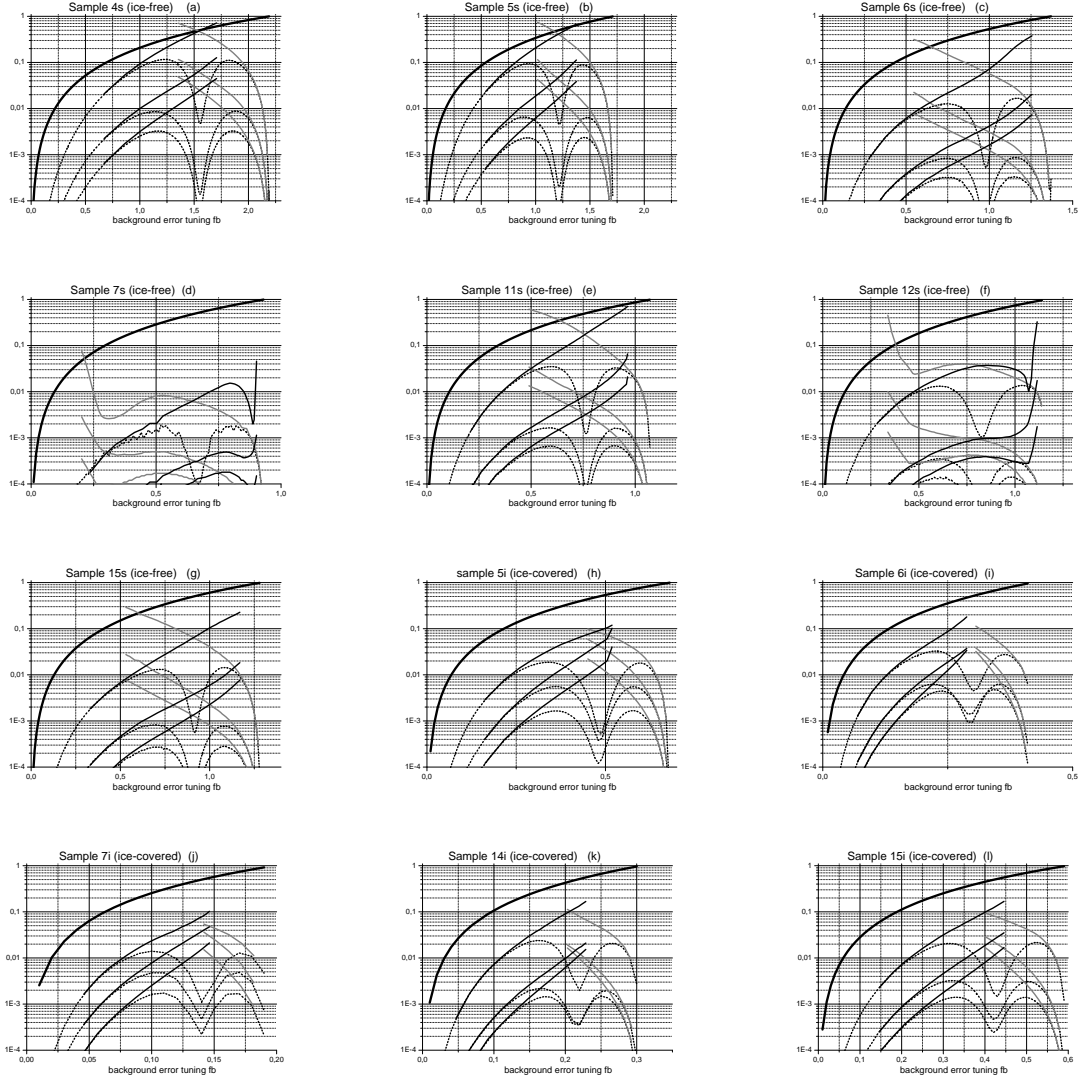


Figure 10: Common to all figures a) to l): Solid thick line: Value of β_b with the tuned mean variance of background errors. Three sets of curves with the scores MSC_j ($j = a, b, c$) and non-Gaussianity source scenarios: (1) Gaussian background errors: black thin solid lines; (2) Gaussian observation errors: gray 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: $MSC_a < MSC_b < MSC_c$. Figures a) to l) correspond respectively to samples 4s, 5s, 6s, 7s, 11s, 12s, 15s, 5i, 6i, 7i, 14i and 15i. The graphics are in logarithmic scale, values of the partial scores below 10^{-4} are not plotted. See text for details.

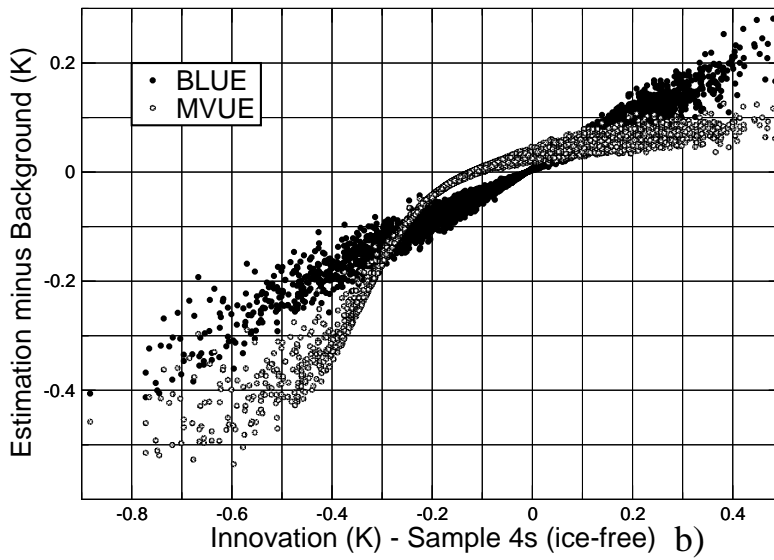
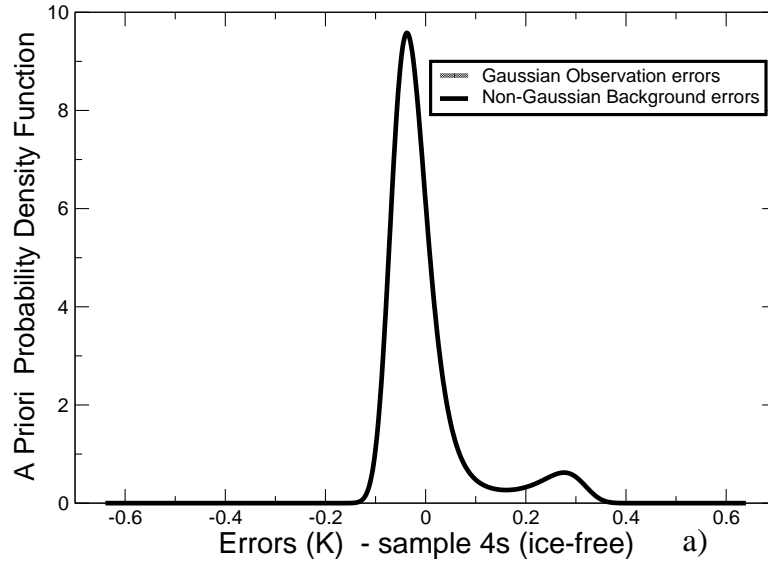


Figure 11: *A priori* pdf of Gaussian observation errors (grey line) and non-Gaussian background errors (dark line), for sample 4s, found by the Maximum Entropy method, fixing the error variance partition: $\beta_b = 0.47$, $\beta_o = 0.53$. b) Scatter plot of the background corrections: Δ_{Lb} (dark spots) and Δ_{Mb} (grey spots), respectively for the BLUE and the MBUE as function of the innovation and the same error variance partition as in a).

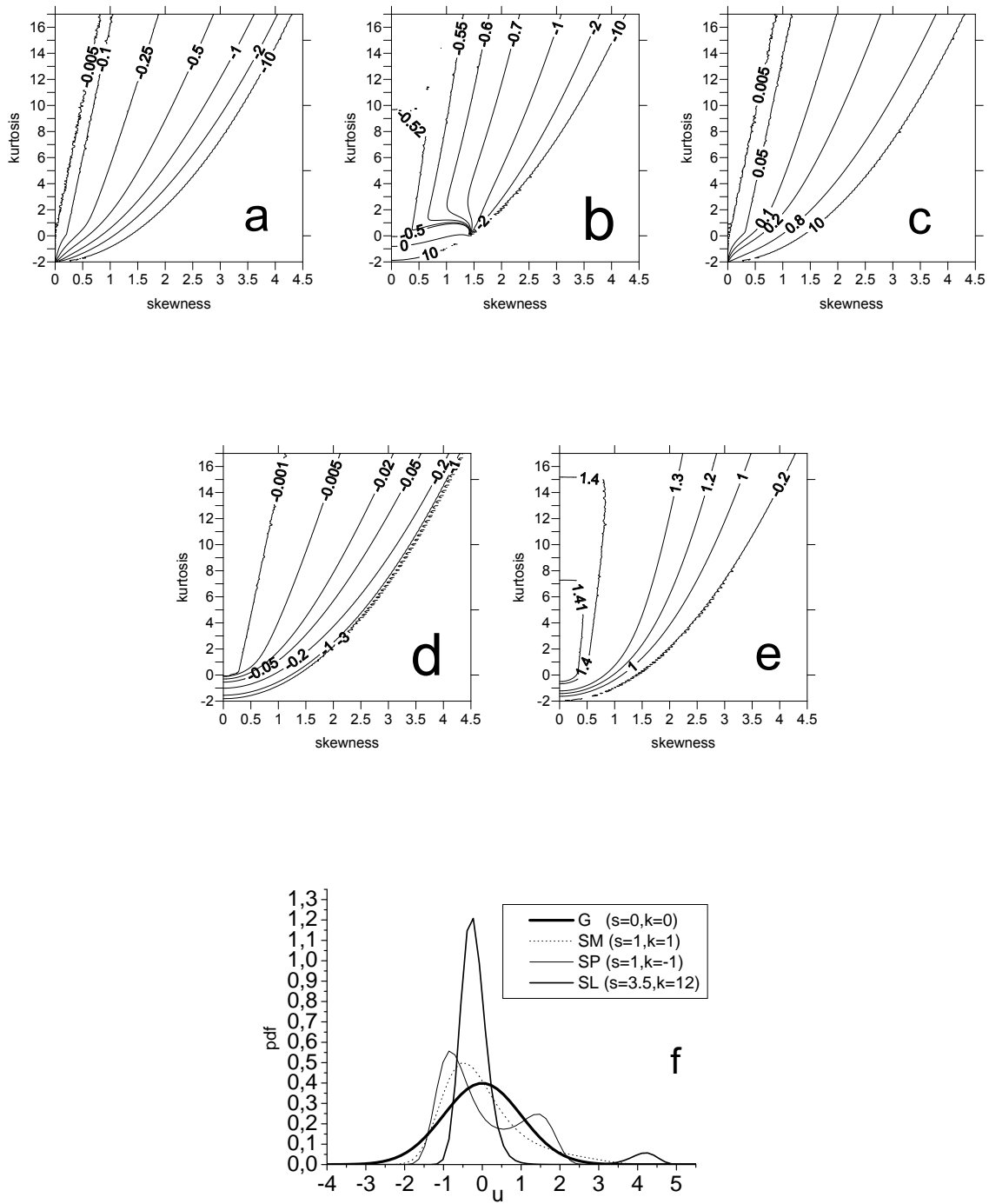


Figure 12: Lagrange Multipliers of the standardized ME-pdfs as function of its skewness s and kurtosis k : a) λ_1 ; b) λ_2 ; c) λ_3 ; d) λ_4 . e) Maximum entropy (nats) of the ME-pdfs in the SK domain, f) Standardized ME-pdfs for points G, SM, SP and SL. See text for details.