Tellus SERIES A DYNAMIC METEOROLOGY AND OCEANOGRAPHY
PUBLISHED BY THE INTERNATIONAL METEOROLOGICAL INSTITUTE IN STOCKHOLM

Observation impact in data assimilation: the effect of non-Gaussian observation error

By ALISON FOWLER[1]* and PETER JAN VAN LEEUWEN[2], [1]*Department of Mathematics, University of Reading, Reading, UK*; [2]*Department of Meteorology, University of Reading, Reading, UK*

(Manuscript received 6 November 2012; in final form 1 May 2013)

ABSTRACT

Data assimilation methods which avoid the assumption of Gaussian error statistics are being developed for geoscience applications. We investigate how the relaxation of the Gaussian assumption affects the impact observations have within the assimilation process. The effect of non-Gaussian observation error (described by the likelihood) is compared to previously published work studying the effect of a non-Gaussian prior. The observation impact is measured in three ways: the sensitivity of the analysis to the observations, the mutual information, and the relative entropy. These three measures have all been studied in the case of Gaussian data assimilation and, in this case, have a known analytical form. It is shown that the analysis sensitivity can also be derived analytically when at least one of the prior or likelihood is Gaussian. This derivation shows an interesting asymmetry in the relationship between analysis sensitivity and analysis error covariance when the two different sources of non-Gaussian structure are considered (likelihood vs. prior). This is illustrated for a simple scalar case and used to infer the effect of the non-Gaussian structure on mutual information and relative entropy, which are more natural choices of metric in non-Gaussian data assimilation. It is concluded that approximating non-Gaussian error distributions as Gaussian can give significantly erroneous estimates of observation impact. The degree of the error depends not only on the nature of the non-Gaussian structure, but also on the metric used to measure the observation impact and the source of the non-Gaussian structure.

Keywords: mutual information, relative entropy, sensitivity

1. Introduction

In assimilating observations with a model, the assumptions made about the distribution of the observation errors are very important. This can be seen objectively by measuring the impact the observations have on updating the estimate of the true state, as given by the data assimilation scheme.

Many data assimilation (DA) schemes are derivable from Bayes' theorem, which gives the updated estimate of the true state in terms of a probability distribution, $p(\mathbf{x}|\mathbf{y})$.

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} \tag{1}$$

In the literature, the probability distributions $p(\mathbf{y}|\mathbf{x})$, $p(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$ are known as the likelihood, prior and posterior, respectively. $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$ must be known or approximated in order to calculate the posterior distribution, while $p(\mathbf{y})$ is generally treated as a normalisation factor as it is independent of $\mathbf{x}$. The mode of the posterior distribution is then the most likely state given all available information and the mean is the minimum variance estimate of the state.

This paper aims to give insight into how the structure of the given distributions, $p(\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$, affect the impact the observations have on the posterior, $p(\mathbf{x}|\mathbf{y})$. It is known from previous studies that non-Gaussian statistics change the way observations are used in data assimilation (e.g. Bocquet, 2008). This paper presents analytical results to explain this change in observation impact. We begin by presenting the case of Gaussian statistics.

1.1. Gaussian statistics

An often useful approximation for $p(\mathbf{y}|\mathbf{x})$ and $p(\mathbf{x})$ is that they are Gaussian distributions, this allows the distributions to be fully characterised by a mean and covariance. The mean of $p(\mathbf{y}|\mathbf{x})$ is the value of the observations, $\mathbf{y}$, measuring the true state and the mean of $p(\mathbf{x})$ is our prior estimate of the true state, $\mathbf{x}_b$. The covariances represent the errors in these two estimates of the truth and are given by

*Corresponding author.
email: a.m.fowler@reading.ac.uk

Citation: Tellus A 2013, **65**, 20035, http://dx.doi.org/10.3402/tellusa.v65i0.20035

R and **B** for the observations and prior estimate, respectively. In the case when the observations and state are represented in different spaces, it is necessary to transform the likelihood into state space in order to apply Bayes' theorem. However, if the transform is linear the likelihood continues to be Gaussian in the observed subspace of the state.

In assuming the likelihood and prior (and subsequently posterior) are Gaussian the DA problem is greatly simplified. As such, these assumptions have been used in the development of operational DA schemes for use in numerical weather prediction (NWP). For example, the Gaussian assumption has been used in the development of variational techniques such as 4D-Var used at the Met Office and ECMWF (Rabier et al., 2000; Rawlins et al., 2007), and Kalman Filter techniques such as the ensemble Kalman filter used at Environment Canada (Houtekamer and Mitchell, 1998). In these operational settings, a measure of the impact of observations has been used for

- Improved efficiency of the assimilation by removing observations with a comparatively small impact, e.g. Peckham (1974); Rabier et al. (2002); Rodgers (1996).
- Highlighting erroneous observations or assumed statistics, e.g. Desroziers et al. (2009).
- Improving the accuracy of the analysis by adding observations which should theoretically have a high impact. For example, by defining targeted observations (Palmer et al., 1998) or the design of new observing systems (e.g. Wahba, 1985; Eyre, 1990).

In this work we will concentrate on three measures of observation impact: the sensitivity of the analysis (to be defined) to the observations; mutual information; and relative entropy. Below, these three measures are briefly introduced and interpreted for Gaussian error statistics. For a more in depth study of these measures see relevant chapters within the following books: Cover and Thomas (1991); Rodgers (2000); and Bishop (2006).

1.1.1. The sensitivity of the analysis to the observations. In Gaussian data assimilation the mode and mean of the posterior distribution are the same and unambiguously define the analysis. In this case the analysis, $\mathbf{x}_a$, is a linear function of the observations and prior estimate:

$$\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \mathbf{K}(\mathbf{y} - \mathbf{H}\mathbf{x}_\mathrm{b}), \tag{2}$$

where **K** is known as the Kalman gain and is a function of **B**, **R** and **H**. **H** is the observation operator, a (linear) map from state to observation space (See Kalnay, 2003 for an introduction to Gaussian data assimilation.)

The sensitivity of the analysis to the observations has an obvious interpretation in terms of observation impact (Cardinali et al., 2004). It is defined as:

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_\mathrm{a}}{\partial \mathbf{y}}. \tag{3}$$

This is a $m \times m$ matrix where m is the size of the observation space.

From eq. (2) we can see that $\mathbf{S}^\mathrm{G}$ (superscript G refers to the Gaussian assumption) is simply

$$\mathbf{S}^\mathrm{G} = \mathbf{H}\mathbf{K}. \tag{4}$$

The Kalman gain can be written in many different forms including $\mathbf{K} = \mathbf{P}_\mathrm{a}^\mathrm{G}\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}$ where $\mathbf{P}_\mathrm{a}^\mathrm{G}$ is the analysis error covariance matrix given by

$$\mathbf{P}_\mathrm{a}^\mathrm{G} = (\mathbf{H}^\mathrm{T}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}. \tag{5}$$

Therefore, the sensitivity is inversely proportional to **R** and proportional to $\mathbf{P}_\mathrm{a}^\mathrm{G}$. Hence it can be concluded from eqs. (4) and (5), that the analysis has greatest sensitivity to independent observations with the smallest error variance which provide information about the region of state space with the largest prior error.

The diagonal elements of **S** give the self-sensitivities and the off-diagonal elements give the cross-sensitivities. The trace of $\mathbf{S}^\mathrm{G}$ can be shown to give the *degrees of freedom for signal*, d_s, that is $d_\mathrm{s} = \sum_i \lambda_i$, where λ_i is the i^th eigenvalue of **HK** (Rodgers, 2000).

The analysis sensitivity has proven to be a useful diagnostic of the data assimilation system (Cardinali et al., 2004). It is possible to approximate eq. (4) during each analysis cycle giving a valuable tool for assessing the changing influence of observations and monitoring the validity of the error statistics.

1.1.2. Mutual information. Mutual information is the change in entropy (uncertainty) when the observations are assimilated (Cover and Thomas, 1991). It is given in terms of the prior, $p(\mathbf{x})$, and posterior, $p(\mathbf{x}|\mathbf{y})$, distributions:

$$MI = \int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} - \int p(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y}) \ln p(\mathbf{x}|\mathbf{y}) d\mathbf{x} d\mathbf{y}. \tag{6}$$

For Gaussian statistics it is unsurprisingly a function of the analysis and background error covariance matrices, $\mathbf{P}_\mathrm{a}^\mathrm{G}$ and **B**. In this case it is given by

$$MI^\mathrm{G} = \frac{1}{2} \ln |\mathbf{B}(\mathbf{P}_\mathrm{a}^\mathrm{G})^{-1}| \tag{7}$$

(Rodgers, 2000), where |*| represents the determinant. As with degrees of freedom for signal, mutual information can be written in terms of the eigenvalues of the sensitivity

matrix: $MI = -\frac{1}{2}\sum \ln |1-\lambda_i|$. Therefore, the observations which have the greatest contribution to mutual information should also be the observations which the analysis is most sensitive to.

Mutual information has been used in many studies of new observing systems. Eyre (1990) demonstrated its benefits over measuring the change in error variances alone as this measure incorporates information about the change in the covariances too [see eq. (7)].

1.1.3. Relative entropy. Relative entropy measures the relative uncertainty of the posterior compared to the prior (Cover and Thomas, 1991).

$$RE = \int p(\mathbf{x}|\mathbf{y}) \ln \frac{p(\mathbf{x}|\mathbf{y})}{p(\mathbf{x})} d\mathbf{x}. \tag{8}$$

For Gaussian statistics it is given by

$$RE^{\mathrm{G}} = \frac{1}{2}(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}}) + MI - \frac{1}{2}d_{\mathrm{s}} \tag{9}$$

(see Bishop, 2006). This is the only measure that depends on the value of the analysis, $\mathrm{x_a}$, and so is sensitive not only to how the covariance of the analysis error is affected by the observations but also how the observations affect the actual value of the analysis. Therefore the observations which result in the greatest relative entropy do not necessarily give the largest mutual information or analysis sensitivity if the signal term in eq. (9), $\frac{1}{2}(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{b}})$, is dominant. However, in a study of the measures, d_{s}, *MI* and *RE* by Xu et al. (2009), it was found that for defining an optimal radar scan configuration the result had little dependence on which of these measures were used.

Relative entropy has received little attention in operational DA due to its dependence on the observation values. However, the shift in the posterior away from the prior, not just the reduction in uncertainty, is clearly an important aspect of observation impact. For this reason, this measure has been included in our current study.

1.2. Non-Gaussian statistics

Although Gaussian data assimilation has proven to be a powerful tool, in some cases a Gaussian distribution gives a poor description of the error distributions. For example, it is found that describing the observation minus background differences (known as innovations) as a Gaussian distribution often underestimates the probability of extreme innovations (the tails are not fat enough) and so the associated observations are assumed to be unlikely rather than providing valuable information about extreme events, and so removed in quality control. Following on from Ingleby and Lorenc (1993), non-Gaussian likelihoods such as the Huber function (Huber, 1973), are being used in operational quality control to make better use of the available observations. Another study of innovation statistics performed by Pires et al. (2010) found significant deviations from Gaussian distributions in the case of quality controlled observations of brightness temperature from the High Resolution Infrared Sounder (HIRS). Pires et al. (2010) concluded that incorrectly assuming Gaussian statistics can have a large impact on the resulting estimate of the state. In this case, the magnitude of the effect on the estimate of the state was seen to depend upon the size of the innovation and the non-Gaussian structure in the likelihood relative to that in the prior.

From eq. (1), we see that there are no restrictions on our choice of $p(\mathbf{y}|\mathbf{x})$ or $p(\mathbf{x})$ when calculating the posterior and the more accurately these distributions are defined the more accurate our posterior's representation of our knowledge of the state will be. This simple fact has led to a recent surge in research into non-Gaussian DA methods which are applicable to the geosciences, see van Leeuwen (2009) and Bocquet et al. (2010) for a review of a range of possible techniques.

This work follows on from Fowler and van Leeuwen (2012), in which the effect of a non-Gaussian prior on the impact of observations was studied when the likelihood was restricted to a Gaussian distribution. The main conclusions from that paper are summarised below.

In Fowler and van Leeuwen (2012) a Gaussian mixture was used to describe the prior distribution to allow for a wide range of non-Gaussian distributions. It was shown that, in the scalar case, the sensitivity of the analysis to observations was still given by the analysis error variance divided by the observation error variance as in the Gaussian case [see eq. (4)]. However, the sensitivity could become a strong function of the observation value because the analysis error variance is no longer independent of the observation value. Therefore, when the prior and observation error distributions are fixed but the position of the likelihood, given by the observation value, is allowed to change, the analysis is most sensitive to observations which also result in the largest analysis error variance. This is not a desirable property for a measure of observation impact. Fowler and van Leeuwen (2012) concluded that comparing the analysis sensitivity to the sensitivity when a Gaussian prior was assumed eq. (4), showed that the Gaussian assumption could lead to both a large overestimation and a large underestimation depending on the value of the observation relative to the background and the structure of the prior.

The error in the Gaussian approximation to relative entropy was also seen to give a large range of errors as a function of the observation value relative to the background. However, for a particular realisation of the observation value, the errors in the Gaussian approximation to

relative entropy and the Gaussian approximation to the sensitivity did not necessarily agree in sign or magnitude. This highlighted the fact that care is needed when making conclusions about the influence of a non-Gaussian prior on the observation impact.

Mutual information is independent of the realisation of the observation error [as seen in eq. (6)] and so as a measure of the influence of a non-Gaussian prior it provides a more consistent result. Mutual information was also seen in Fowler and van Leeuwen (2012) to be affected a relatively small amount by a non-Gaussian prior.

To summarise: allowing for non-Gaussian prior statistics has a significant effect on the observation impact. The choice of metric is more important than in the Gaussian case as the consistency between the different measures breaks down.

Within this current paper we shall compare these previous findings to the case when it is the likelihood that is non-Gaussian. A non-Gaussian prior or likelihood may result from the properties of the state variable. For example, if the variable has physical bounds then we know, a priori, that the probability of the variable being outside of these bounds is zero which is inconsistent with the infinite support of the Gaussian distribution. This is a particular issue when the variable is close to these bounds. The non-Gaussian prior may also result from a non-linear forecast providing our prior estimate of the state. In this case a wide variety of non-Gaussian distributions are possible and techniques such as the particle filter (van Leeuwen, 2009) allow for the non-linear model to implicitly give the prior distribution. A non-Gaussian likelihood may similarly result from a non-linear map between the observation and state space. However, within this paper we shall only give consideration to linear observation operators. Non-linear observation operators greatly complicate the problem, as not only do they create non-Gaussian likelihoods out of Gaussian observation errors, the structure of the non-Gaussian PDF depends on the value of the observation.

The observation error in this case is defined as:

$$\epsilon = \mathbf{y} - \mathbf{H}\mathbf{x}_{\text{truth}}$$

Possible contributing factors to ϵ include:

- Random error associated with the measurement.
- Random errors in the observation operator, for example due to missing processes or linearisation. This is often distinguished from representivity error which deals with the additional error source due to the observations sampling scales which the model is unable to resolve (Janjić and Cohn, 2006).
- Systematic errors are also possible, which may have synoptic, seasonal or diurnal signals some of which can be corrected. There are also gross errors which need to be identified and rejected by quality control (e.g. Gandin et al., 1993; Qin et al., 2010; Dunn et al., 2012).

These sources of random error could all potentially lead to a non-Gaussian structure in the observation errors. Errors associated with the observation operator will in general be state dependent (Janjić and Cohn, 2006). For this reason, within this paper, we will focus on the case of perfect linear observation operators so that the non-Gaussian structure is a characteristic of the instrument error or pre-processing of the observations before they are assimilated. It is assumed that this error source is independent of the state.

In analysing the impact of the non-Gaussian likelihood (rather than a non-Gaussian prior) we shall follow a similar methodology to that in Fowler and van Leeuwen (2012). We shall first derive some general results for the sensitivity of the analysis to the observations. We will then look at a scalar example when the likelihood is described by a Gaussian mixture with two components each with identical variances, GM_2. This will allow for direct comparison to the results in Fowler and van Leeuwen (2012). Finally we will look at the case when the measurement error is described by a Huber function which cannot be described well by the GM_2 distribution.

2. The effect of non-Gaussian statistics on the analysis sensitivity

In non-Gaussian data assimilation the analysis must be explicitly defined. In this work we define the analysis as the posterior mean giving the minimum variance estimate of the state rather than the mode which can be ill-defined when the posterior is multi-modal. In extreme bimodal cases this does lead to the possibility of the analysis having low probability.

The sensitivity of the analysis to the observations can be calculated analytically when either the prior or likelihood is Gaussian, see appendix A. It can be shown that in the case of an arbitrary likelihood and Gaussian prior that the sensitivity is given by

$$\mathbf{S}^{\text{nG}p(y|x)} = \frac{\partial \mathbf{H}\boldsymbol{\mu}_{\text{a}}}{\partial \mathbf{y}} = \mathbf{I}_m - \mathbf{H}\mathbf{P}_{\text{a}}\mathbf{B}^{-1}\mathbf{H}^{\text{T}}(\mathbf{H}\mathbf{H}^{\text{T}})^{-1} \tag{10}$$

where $\boldsymbol{\mu}_{\text{a}}$ is the analysis (mean of the posterior). When the likelihood is Gaussian, $\mathbf{P}_{\text{a}} = \mathbf{P}_{\text{a}}^{\text{G}} = (\mathbf{B}^{-1} + \mathbf{H}^{\text{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}$ and the expression given in eq. (10) is equal to $\mathbf{HK}$ [see Section 1.1.1, eq. (4)]. However, a non-Gaussian likelihood means that $\mathbf{P}_{\text{a}}$ becomes a function of the observation value and hence the sensitivity is also a function of the observation value.

From eq. (10) it is seen that $\mathbf{S}^{\text{nG}p(y|x)}$ increases as $\mathbf{P}_{\text{a}}$ decreases and so for a fixed prior and likelihood structure,

the realisation of $\mathbf{y}$ for which the analysis has maximum sensitivity also gives the smallest analysis error covariance. In other words, the analysis is most sensitive to observations which improve its accuracy.

This is in contrast to when the prior is non-Gaussian and the likelihood is Gaussian. In this case the sensitivity is given by (see Appendix A)

$$\mathbf{S}^{\mathrm{nG}p(x)} = \frac{\partial \mathbf{H}\boldsymbol{\mu}_{\mathrm{a}}}{\partial \mathbf{y}} = \mathbf{H}\mathbf{P}_{\mathrm{a}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{11}$$

This has the same form as eq. (4) in Section 1.1.1. In this case, it is seen that $\mathbf{S}^{\mathrm{nG}p(x)}$ is proportional to the analysis error covariance. Therefore for a fixed prior and likelihood structure, the analysis is most sensitive to observations which reduce its accuracy.

The analysis error covariance is an important aspect in DA in which it is desirable to find a minimum value. Therefore, from eqs. (10) and (11), we can conclude that the influence of a non-Gaussian likelihood on the analysis sensitivity is of a fundamentally different nature to the influence of a non-Gaussian prior. This is demonstrated for a simple scalar example in the next section.

3. A simple example

For comparison to the results in Fowler and van Leeuwen (2012) in which the effect of a skewed and bimodal prior on the observation impact was studied we shall look at the scalar case when the likelihood can be described by a Gaussian mixture with two components each with identical variance, GM_2.

$$\begin{aligned} p(y|x) =& ((2\pi)\sigma^2)^{-\frac{1}{2}}\left(w\exp\left\{-\frac{(y+\nu_1-x)^2}{2\sigma^2}\right\}\right. \\ &\left. + (1-w)\exp\left\{-\frac{(y+\nu_2-x)^2}{2\sigma^2}\right\}\right). \end{aligned} \tag{12}$$

In this example it is assumed that we have direct observations of the state, x, and so the observation operator, $\mathbf{H}$, is simply the identity. From eq. (12) we see that as a function of x, the means of the Gaussian components are $\mu_1 = y + \nu_1$ and $\mu_2 = y + \nu_2$. To ensure that eq. (12) is non-biased, i.e. $\int (y-x)p(y|x)dx = 0$, we have the constraint $w\nu_1 + (1-w)\nu_2 = 0$, effectively making our observation, y, the mean of the likelihood. This could be restrictive, particularly in the case of a strongly bimodal likelihood when the observation would have a low probability. However, as long as the observation value is chosen to be the likelihood mean plus a constant, the analysis sensitivity presented below remains unchanged.

The likelihood in eq. (12) is described by four parameters: the relative weight of the Gaussian components, w, the means of the Gaussian components, μ_1 and μ_2, and the variance of the Gaussian components, σ^2. These four parameters give rise to a large variety of non-Gaussian distributions, this can be seen from expressions for the skewness and kurtosis.

The variance of the likelihood, $p(\mathrm{y}|\mathrm{x})$ as a function of x, is:

$$\sigma_y^2 = \sigma^2 + w(1-w)(\mu_1-\mu_2)^2.$$

Using this expression for the variance we can give the following expression for the skewness of $p(\mathrm{y}|\mathrm{x})$:

$$\psi_y^3 = \frac{\int (x-\mu_y)^3 p(y|x)dx}{\sigma_y^3} = \frac{w(1-w)(1-2w)(\mu_1-\mu_2)^3}{\sigma_y^3}.$$

And the kurtosis of $p(\mathrm{y}|\mathrm{x})$:

$$\begin{aligned} \kappa_y^4 &= \frac{\int (x-\mu_y)^4 p(y|x)dx}{\sigma_y^4} - 3 \\ &= \frac{(\mu_1-\mu_2)^4 w(1-w)(1-6w+6w^2)}{\sigma_y^4} \end{aligned}$$

The values of skewness and kurtosis are plotted in Fig. 1 as a function of w and μ_2–μ_1 when $\sigma^2 = 1$. It is clear that for a fixed value of μ_2–μ_1, the skewness increases as $|w$–$0.5|$ increases until it reaches a maximum (indicated by the dotted line in Fig. 1), then the skewness sharply returns to zero as $|w$–$0.5|$ approaches 0.5 and GM_2 returns to a Gaussian distribution. When the weights are equal ($w = 0.5$) only negative values of kurtosis are possible, increasing as μ_2–μ_1 increases. This results in a likelihood with a flatter peak than the Gaussian distribution becoming bimodal as $w(1$–$w)(\mu_1$–$\mu_2)^2$ exceeds σ^2. Positive values of kurtosis are possible when the distribution is also highly skewed. Note that a Gaussian distribution has zero kurtosis.

Non-Gaussian structure such as skewness could result from bounds on the observed variable or as a result of non-linear pre-processing. In such a case it should be possible to construct a model of the errors associated with the measurement, such as the GM_2 distribution introduced in this section, through comparison to other observations and prior information for which we have a good estimate of the errors. It is difficult to think of a situation when the likelihood may be strongly bimodal without accounting for a non-linear observation operator. For example, an ambiguity in the observation could result in a bimodal error such as in the use of a scatterometer to measure wind direction from waves (Martin, 2004). Modelling $y = |x|$ results in a likelihood with a GM_2 distribution with equal weights, if the error on the observation, y, is Gaussian. In this case the means of the two Gaussian components would be $\mu_1 = y$ and $\mu_2 = -y$. However, the general results provided in Section 2 do not hold. As such in the following

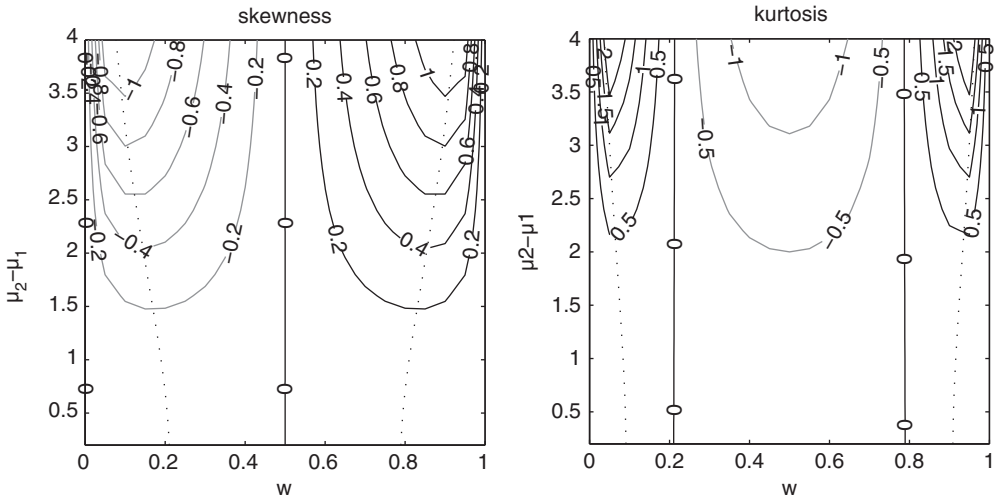

Fig. 1. Skewness (left) and kurtosis (right) of the GM2 distribution as a function of w and $\mu_2-\mu_1$ when $\sigma^2=1$.

analysis the emphasis is not on the extreme bimodal case but the smaller deviations from a Gaussian distribution that the GM_2 distribution allows.

When the prior is given by a Gaussian ($p(x)=N(\mu_x^G, k\sigma^2)$), the posterior will also be given by a GM_2 distribution [refer to Bayes' theorem, eq. (1)] with updated parameters. These updated parameters are given by

$$\tilde{w}=\frac{we^{-a_1}}{we^{-a_1}+(1-w)e^{-a_2}}, \tag{13}$$

where $a_i=((\mu_i-\mu_x^G)^2)/(2(1+k)\sigma^2)$.

$$\tilde{\mu}_i=\frac{k\mu_i+\mu_x^G}{1+k},$$

for $i=1,2$.

$$\tilde{\sigma}^2=\frac{k\sigma^2}{1+k}.$$

Note that these have the same form as in Fowler and van Leeuwen (2012) due to the symmetry of Bayes' theorem.

Given this expression for the posterior distribution we can calculate its mean as: $\mu_a=\tilde{w}\tilde{\mu_1}+(1-\tilde{w})\tilde{\mu_2}$. The sensitivity of the posterior mean to the observations can then be expressed in terms of the parameters of the likelihood distribution as:

$$S^{\mathrm{GM}_2p(y|x)}=\frac{\partial\mu_a}{\partial y}=\frac{k}{k+1}-\frac{kw(1-w)(\mu_1-\mu_2)^2e^{-a_1-a_2}}{(1+k)^2\sigma^2(we^{-a_1}+(1-w)e^{-a_2})^2}. \tag{14}$$

This expression has a striking resemblance to the sensitivity in the case of the non-Gaussian prior, $S^{\mathrm{GM}_2p(x)}$. In eq. (13) of Fowler and van Leeuwen (2012), when the prior has the same form as the likelihood described in eq. (12), $S^{\mathrm{GM}_2p(x)}$ was shown to be

$$S^{\mathrm{GM}_2p(x)}=\frac{1}{\kappa+1}+\frac{\kappa w(1-w)(\mu_1-\mu_2)^2e^{-\alpha_1-\alpha_2}}{(1+\kappa)^2\sigma^2(we^{-\alpha_1}+(1-w)e^{-\alpha_2})^2}. \tag{15}$$

In this case $p(y|x)=N(y,\kappa\sigma^2)$ and $\alpha_i=((y-\mu_i)^2)/(2(1+\kappa)\sigma^2)$. Note the distinction between k and κ; these parameters are used to define the ratio of the Gaussian variances to the Gaussian component variance for the non-Gaussian likelihood case and the non-Gaussian prior case, respectively.

An example of the setup of this simple scalar example is shown in Fig. 2. In the left-hand panel the non-Gaussian prior case which was the focus of Fowler and van Leeuwen (2012) is illustrated and in the right-hand panel the non-Gaussian likelihood case is illustrated. The values of k and κ have been chosen such that σ_y^2/σ_x^2 is fixed in the two setups, where σ_y^2 is the variance of the likelihood (either Gaussian or not) and σ_x^2 is the variance of the prior (which also may or may not be Gaussian).

We see in eq. (14) that the sensitivity tends towards an *upper* bound of $\frac{k}{k+1}$ as the likelihood becomes Gaussian ($\mu_1-\mu_2$ tends to zero) with *no lower bound* when the likelihood has two distinct modes (i.e. $(\mu_1-\mu_2)^2/\sigma^2$ is large). In contrast, in eq. (15), the sensitivity tends towards a *lower* bound of $\frac{1}{\kappa+1}$ as the prior becomes Gaussian with *no upper bound* when the prior has two distinct modes. The lack of lower and upper bounds for the non-Gaussian likelihood case and non-Gaussian prior case, respectively, has an important consequence for the analysis error variance. From the relationship between the sensitivity and posterior variance given in Section 2 we can conclude that when $S^{\mathrm{GM}_2p(y|x)}<0$ that $\sigma_a^2>\sigma_x^2$ and similarly when $S^{\mathrm{GM}_2p(x)}>1$ that $\sigma_a^2>\sigma_y^2$. In theory this should never be the case for

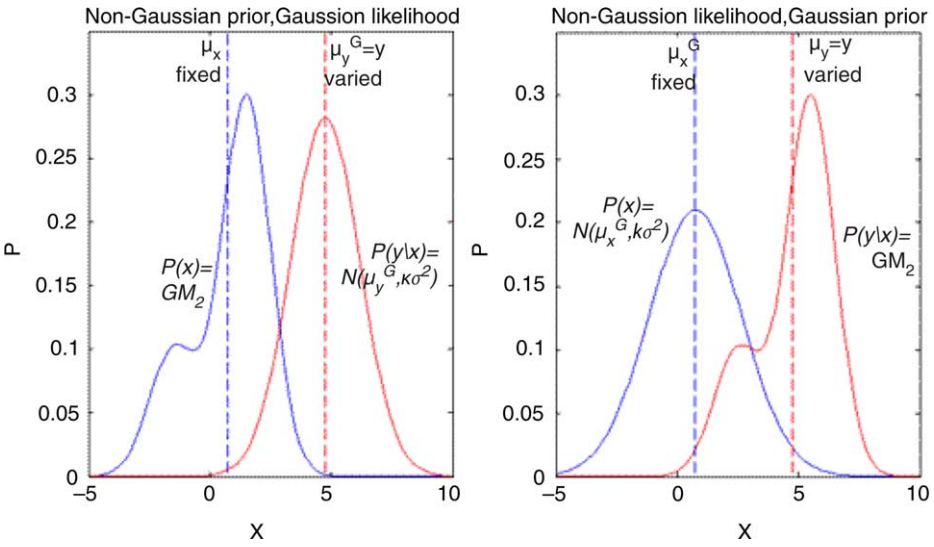

Fig. 2. Schematic of experimental setup in section 3. Left hand panel: Non-Gaussian prior and Gaussian likelihood as in Fowler and van Leeuwen (2012). Right hand panel: Non-Gaussian likelihood and Gaussian prior, which is the focus of this paper. In each case the non-Gaussian parameters are given by $w = 0.25$, $\sigma^2 = 1$, $|\mu_1 - \mu_2| = 3$. The variance of the Gaussian distributions are chosen such that in each case $\sigma_y^2/\sigma_x^2 = \frac{32}{43}$, for agreement with Fowler and van Leeuwen (2012), giving $k = 1849/512$ and $\kappa = 2$.

purely Gaussian data assimilation. The potential for the analysis error variance to be greater than the observation and prior error variances was similarly demonstrated for the case of errors following an exponential law in Talagrand (2003) and when a particle filter is used to assimilate observations with a non-linear model of the Agulhas Current in van Leeuwen (2003).

As a function of the innovation, $d = y - \mu_x$, the shape of $S^{\mathrm{GM}_2 p(y|x)}$ is similar to $S^{\mathrm{GM}_2 p(x)}$ although inverted. In each case the sensitivity is a symmetrical function of d about a central value. In the non-Gaussian likelihood case, this value of d, d_l, marks a minimum value of $S^{\mathrm{GM}_2 p(y|x)}$. In the non-Gaussian prior case, this value of d, d_p, marks a maximum value of $S^{\mathrm{GM}_2 p(x)}$. In general $d_\mathrm{l} \neq d_\mathrm{p}$ unless all parameters are identical with $w = 1/2$. Away from d_l and d_p the sensitivity tends to $\frac{k}{k+1}$ for the non-Gaussian likelihood case and $\frac{1}{\kappa+1}$ for the non-Gaussian prior case, as could be expected from the symmetry between k and $1/\kappa$. When the parameters describing the non-Gaussian distribution are the same, the relative speed at which the sensitivity asymptotes to $\frac{k}{k+1}$ or $\frac{1}{\kappa+1}$ depends on the values of k and κ, respectively, if $k = \kappa$ then it is the same.

An example of this is shown in Fig. 3, where $\sigma^2 = 1$, $w = 0.25$, $\mu_1 - \mu_2 = 3$ for both the non-Gaussian likelihood and non-Gaussian prior. $\kappa = 2$ for comparison to the example in Fowler and van Leeuwen (2012) and k is chosen to be 1849/512 so that in each case the Gaussian approximation to the sensitivity is the same, that is $S^G = \sigma_x^2(\sigma_x^2 + \sigma_y^2)^{-1}$ stays constant even though σ_x^2 and σ_y^2 are not identical in the two cases, in fact the error variances for the prior and likelihood are larger in the non-Gaussian likelihood case than in the non-Gaussian prior case (see Fig. 2). From eq. (12) and (13) $k > \kappa$ implies that $S^{\mathrm{GM}_2 p(y|x)}$ is a broader function of d (thin blue line) than $S^{\mathrm{GM}_2 p(x)}$ (thin black line) and there is less variance in the sensitivity. This illustrates, that unlike the Gaussian case, the sensitivity is now dependent on the actual values of the error variances rather than just their ratio. The Gaussian approximation to the sensitivity, S^G, is given by the bold dashed line.

As $|d|$ increases $\tilde{w}$ tends to 0/1 in both cases, in effect rejecting one of the modes. In other words the posterior asymptotes to a Gaussian with variance given by $\tilde{\sigma}^2 = \frac{k\sigma^2}{1+k}$ in the case of a non-Gaussian likelihood and $\tilde{\sigma}^2 = \frac{\kappa\sigma^2}{1+\kappa}$ in the case of a non-Gaussian prior. This explains why the sensitivity, which is given by eq. (10) in the non-Gaussian likelihood case and by eq. (11) in the non-Gaussian prior case, tends to a non-zero constant value as $|d|$ increases. It also explains (with some extra thought) why in the non-Gaussian likelihood case this constant sensitivity is greater than the Gaussian approximation and vice versa when the prior is non-Gaussian.

The value of d, which results in a peak value of $S^{\mathrm{GM}_2 p(x)}$, was given in Fowler and van Leeuwen (2012) in terms of the parameters of the non-Gaussian prior.

$$d_\mathrm{p} = \frac{1}{2(\mu_1 - \mu_2)}\left[\mu_1^2 - \mu_2^2 - 2(1+\kappa)\sigma^2 \ln\left(\frac{w}{(1-w)}\right)\right] - \mu_x. \tag{16}$$

We may similarly find an expression for d_l, which results in a minimum value of $S^{\mathrm{GM}_2 p(y|x)}$, when the likelihood is

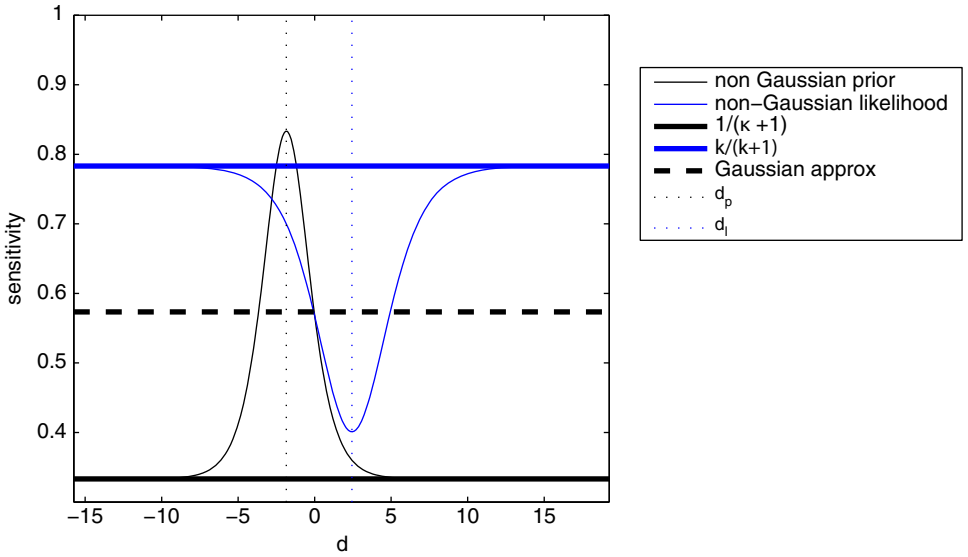

Fig. 3. Comparison of $S = \frac{\partial \mu_a}{\partial y}$ as a function of d when the likelihood is non-Gaussian (prior is Gaussian) (thin blue line) and when the prior is non-Gaussian (likelihood is Gaussian) (thin black line). In each case the non-Gaussian distribution is a two component Gaussian mixture with identical variances with parameter values as in Fig. 2. The variance of the Gaussian distributions, also given in Fig. 2, are chosen such that the Gaussian estimate to the sensitivity is the same in each case (bold, dashed line). Also marked on is $\frac{k}{k+1}$ (bold blue line) and $\frac{1}{\kappa+1}$ (bold black line), and d_p (black dashed line) and d_l (blue dashed line).

non-Gaussian. From eq. (10) we know that as the sensitivity increases the analysis error variance decreases. Therefore when the posterior is of the same form as eq. (12), the posterior variance is at a maximum, and hence the sensitivity is at a minimum, when the posterior weights are equal, i.e. the posterior is symmetric. This insight allows us to find the observation value for which the analysis has least sensitivity by finding the d which satisfies $\tilde{w} = 0.5$.

$$d_{\mathrm{l}} = \frac{1}{2(\mu_2 - \mu_1)}\left[(\mu_2 - \mu_1)^2(1-2w) + 2\sigma^2(1+k)\ln\frac{1-w}{w}\right] \tag{17}$$

Note that in deriving eq. (17) we have made use of the fact that for $d = y - \mu_x^{\mathrm{G}} = \mu_2 - w(\mu_2 - \mu_1) - \mu_x^{\mathrm{G}}$ the terms μ_x^{G}, w and μ_1–μ_2 are considered to be fixed. Therefore we only need to find an expression for μ_2 which satisfies $\tilde{w} = 0.5$. This can then be substituted back into the expression for d.

When the weights are equal in the non-Gaussian prior (i.e. $w = \frac{1}{2}$) $d_{\mathrm{p}} = 0$. Similarly when $w = \frac{1}{2}$ in the non-Gaussian likelihood, $d_{\mathrm{l}} = 0$. Therefore, when the prior is symmetric but with negative kurtosis the analysis is *most* sensitive when the mean of the (Gaussian) likelihood is equal to the mean of the prior. Conversely when the likelihood is symmetric but with negative kurtosis the analysis is *least* sensitive when the mean of the likelihood is equal to the mean of the (Gaussian) prior.

The results illustrated by the example of Fig. 3 can be shown for a range of non-Gaussian distributions described by eq. (12). In Fig. 4, contour plots of $S^{\mathrm{GM}_2 p(y|x)}/S^{\mathrm{G}}$ (left column) and $S^{\mathrm{GM}_2 p(x)}/S^{\mathrm{G}}$ (right column) are given as a function of d and μ_2–μ_1 (top row) and w (bottom row). In all cases k and κ are varied such that S^{G} remains the same value as in Fig. 3. In practice this means that as the variance of the non-Gaussian distribution is increased as a result of the parameters describing the distribution changing, the variance of the Gaussian distribution is similarly increased. Indicated in Fig. 4 is the increasing negative kurtosis as $\mu_2 - \mu_1$ increases and the increasing skewness as $|w - 0.5|$ increases, see Fig. 1 for comparison.

As expected the error in the Gaussian approximation to the sensitivity becomes larger as the skewness and kurtosis of the likelihood/prior increases. The magnitude of the error in the Gaussian approximation to the sensitivity is larger when the prior is non-Gaussian, because in this case $S^{\mathrm{G}} > 0.5$ leading to a narrower function of sensitivity than when the likelihood is non-Gaussian, as explained previously. If $S^{\mathrm{G}} < 0.5$ then the error in the Gaussian approximation to the sensitivity would be larger when the likelihood is non-Gaussian. The gradient in the maximum/minimum sensitivity as a function of w (see Figs. 2c and 2d) can be derived from eqs. (16) and (17).

3.1. *Comparison of sensitivity to other measures of observation impact*

The focus of Fowler and van Leeuwen (2012) was to compare the effect of a non-Gaussian prior on different measures of observation impact. It was seen that like the

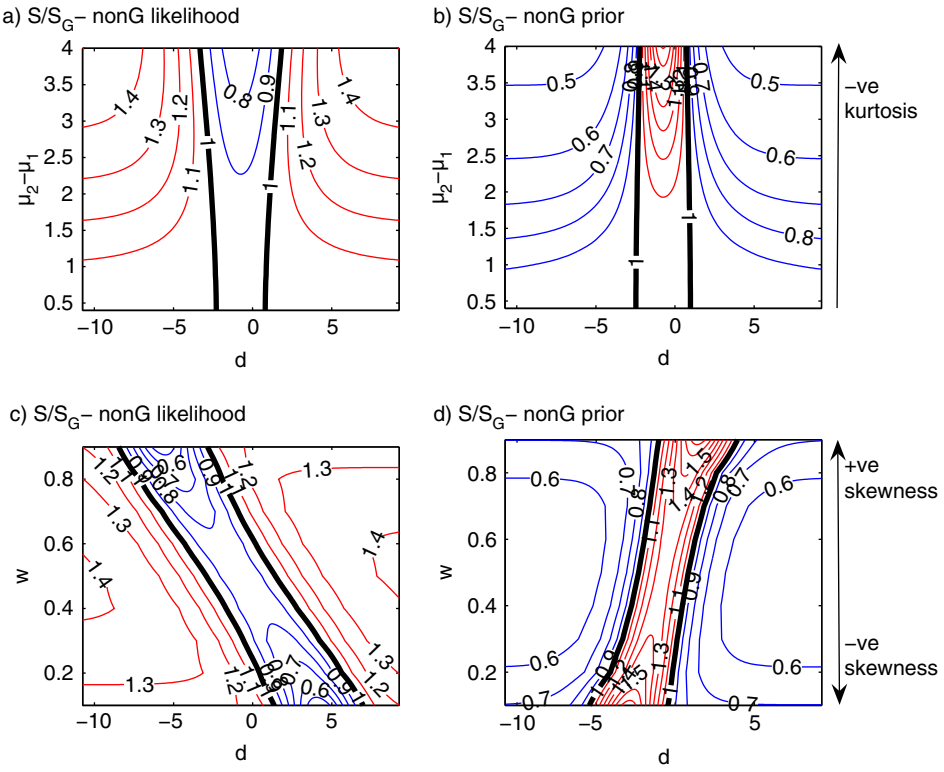

Fig. 4. Comparison of S normalised by its Gaussian approximation when the likelihood is non-Gaussian (prior is Gaussian) (a, c) and when the prior is non-Gaussian (likelihood is Gaussian) (b, d). In the top two panels the effect of separating the Gaussian components (y-axis) is given when the weights are equal. In the bottom two panels the effect of varying the weights of the Gaussian components (y-axis) is given when $|\mu_1 - \mu_2| = 3$. In all cases $\sigma^2 = 1$ and σ_y^2/σ_x^2 is kept constant at $\frac{32}{43}$ by varying the values of k and κ.

sensitivity, the error in the Gaussian approximation to relative entropy was a strong function of the innovation. The strong dependence of both the error in the sensitivity and the error in the relative entropy on the innovation means that there is no consensus as to the effect of a non-Gaussian prior on the observation impact for a given observation value. A similar conclusion can be arrived at when the likelihood is non-Gaussian by comparing Figs. 4a and 4c to 5a and 5c, in which fields of S/S^{G} and RE/RE^{G} have been plotted, respectively.

Relative entropy [see eq. (8)] is loosely related to the sensitivity in two ways:

(1) As seen when relative entropy was first introduced, relative entropy is dependent on the shift of the posterior distribution away from the prior. The shift of the posterior distribution away from the prior, given by $\mu_a - \mu_x$, is proportional to the sensitivity of μ_a to y due to the following relationship:

$$\frac{\partial \mathbf{H}\mu_a}{\partial \mathbf{y}} + \frac{\partial \mathbf{H}\mu_a}{\partial \mathbf{H}\mu_x} = \mathbf{I}_m, \tag{18}$$

where $\mathbf{I}_m$ is an identity matrix of size m (see Appendix A).

As a function of d, the error in the Gaussian approximation to the shift in the posterior away from the prior will be smallest when $\mu_a \approx \mu_x$. This is only approximate because, unlike in the purely Gaussian case, $y = \mu_x$ does not necessarily imply that $\mu_a = \mu_x$ In Fowler and van Leeuwen (2012) this was wrongly assumed to be true. However, as seen in the example given in Fig. 3, the sensitivities are almost equal to the Gaussian approximation of the sensitivity at $d = 0$. Therefore when $y = \mu_x$, μ_a is very close to μ_x when either the prior or likelihood is non-Gaussian. Away from this the Gaussian approximation will underestimate the shift when it underestimates the sensitivity and similarly overestimate the shift when it overestimates the sensitivity.

(2) Relative entropy also measures the reduction in the uncertainty between the prior and posterior. This is strongly linked to the posterior variance: the larger the posterior variance the smaller the reduction in uncertainty. The sensitivity's relationship to the posterior error variance is given by eqs. (10) and (11). Therefore when the likelihood is non-Gaussian the reduction in uncertainty is *overestimated* when the sensitivity is overestimated and when the prior is

non-Gaussian the reduction in uncertainty is *underestimated* when the sensitivity is overestimated.

These two comments explain why in Fowler and van Leeuwen (2012) it was found that the error in relative entropy was generally of a smaller magnitude than the error in sensitivity as the two processes above cancel to some degree. It also explains why in this case, when it is the likelihood that it is non-Gaussian, that the error in relative entropy is generally of a greater magnitude than the error in sensitivity as the two processes above reinforce each other to some degree. This can be seen by comparing Figs. 4 and 5.

These two comments also explain the asymmetry in the error in relative entropy as a function of d when $w \neq \frac{1}{2}$ [see Figs. 5c and 5d]. When $w \neq \frac{1}{2}$ the minimum in error in the shift of the posterior at $d \approx 0$ does not coincide with the maximum (minimum) in the reduction in the posterior variance at $d = d_{\mathrm{p(I)}}$.

Because of the large variability in the sensitivity and relative entropy as a function of observation value it is useful to look at their averaged values, $\int p(y)Sdy$ and $\int p(y)REdy$. The latter is known as mutual information, a measure of the change in entropy when an observation is assimilated (see Section 1.1.1 and Cover and Thomas, 1991).

On average the Gaussian approximation to the non-Gaussian likelihood underestimates the observation impact [see Figs. 6a and 6c]. This is because the Gaussian estimate of the likelihood underestimates the structure and hence the information in the likelihood. This is analogous to the non-Gaussian prior case presented in Fowler and van Leeuwen (2012) where on average the Gaussian approximation to the non-Gaussian prior overestimated the observation impact due to it underestimating the structure in the prior (see Figs. 6b and 6d).

As expected from mutual information's relation to relative entropy and consequently relative entropy's relation to the sensitivity, the error in the Gaussian approximation to MI is greater than the error in the Gaussian approximation to $\int p(y)Sdy$ when the true likelihood is non-Gaussian and vice versa when it is the prior that is non-Gaussian.

A summary of some of the key differences between observation impact when the likelihood and prior are non-Gaussian, as discussed in this Section are given in Table 1.

In this section we have studied the observation impact when a non-Gaussian distribution as described by a two-component Gaussian mixture with identical variances, given by eq. (10), is introduced. This has allowed us to understand how the source of non-Gaussian structure affects the different measures of observation impact when

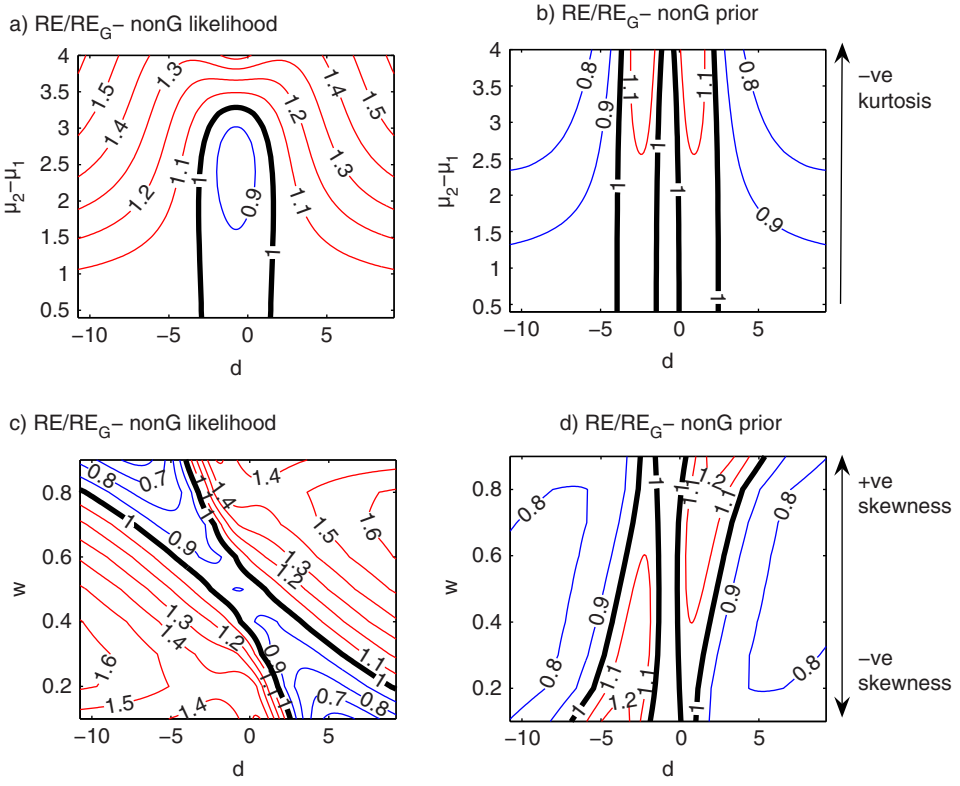

Fig. 5. As in Fig. 4 but for *RE*.

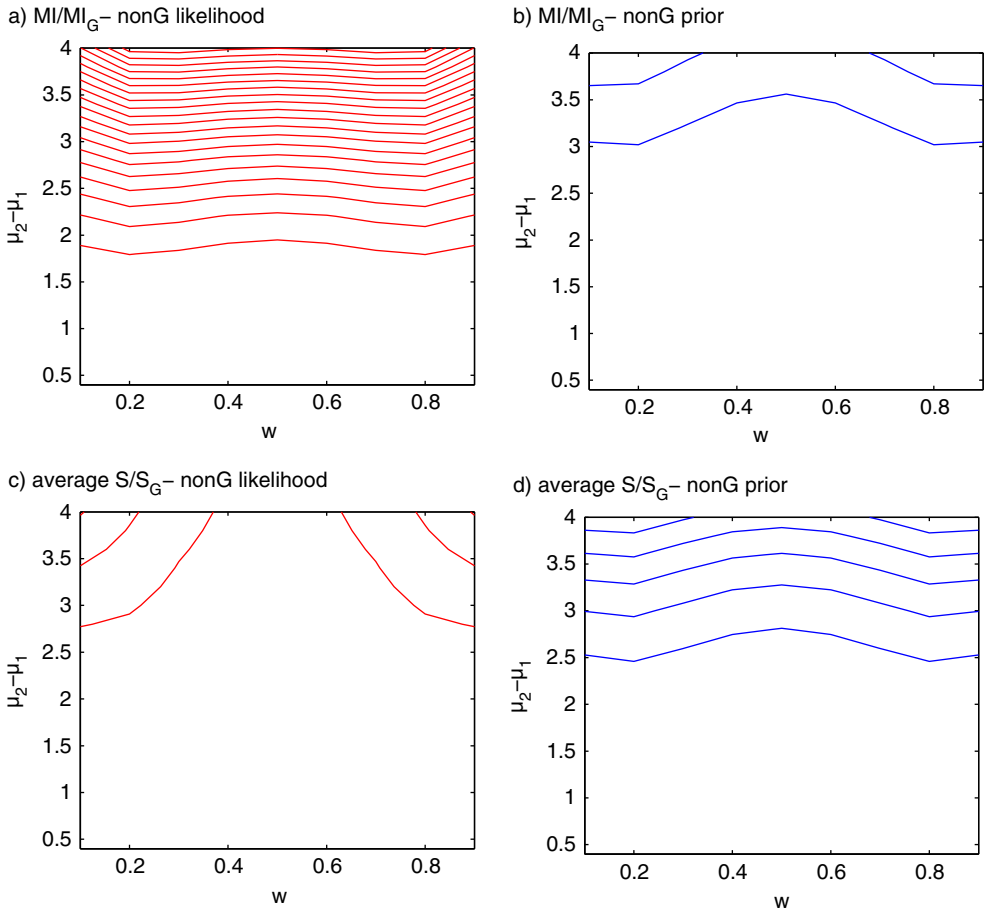

Fig. 6. Top: Comparison of *MI* normalised by its Gaussian approximation as a function of w (x axis) and μ_2–μ_1 (y axis) when (a) the likelihood is non-Gaussian (prior is Gaussian) and (b) when the prior is non-Gaussian (likelihood is Gaussian). Bottom: Comparison of $\int p(y)Sdy$ normalised by its Gaussian approximation as a function of w (x axis) and $\mu_2 - \mu_1$ (y axis) when (c) the likelihood is non-Gaussian (prior is Gaussian) and (d) when the prior is non-Gaussian (likelihood is Gaussian). In all cases $\sigma^2 = 1$ and σ_y^2/σ_b^2 is kept constant at $\frac{32}{43}$ by varying the values of k and κ. Red contours indicate values above 1 and blue contours indicate values below one. The contours are separated by increments of 0.02.

the distributions are skewed or have non-zero kurtosis. At the ECMWF, a mixed Gaussian and exponential distribution, known as a Huber function, has recently been introduced to model the observation error for some in-situ measurements (Tavolato and Isaksen, 2009/2010) during quality control. In the next section we will give a brief overview of the observation impact in this specific case.

4. The Huber function

The Huber function has been shown to give a good fit to the observation minus background differences seen in temperature and wind data from sondes, wind profilers, aircrafts, and ships (Tavolato and Isaksen, 2009/2010). From non-Gaussian observation minus background diagnostics it is difficult to derive the observations error structure alone (Pires et al., 2010). However, due to the difficulty in designing a data assimilation scheme around non-Gaussian prior errors, it is a pragmatic choice to assign the non-Gaussian errors to the observations only. The Huber function is described by the following

$$p(y|x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{a^2}{2} - |a\delta|) \ \textit{if} \ \delta < a \\ \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}\delta^2) \ \textit{if} \ a \le \delta \le b \\ \frac{1}{\sigma\sqrt{2\pi}} \exp(\frac{b^2}{2} - |b\delta|) \ \textit{if} \ \delta > b \end{cases}, \qquad (19)$$

where $\delta = \frac{y - H(x)}{\sigma} = d/\sigma$. The distribution is therefore characterised by the following four parameters: y, the observation value, σ^2; the variance of the Gaussian part of the distribution; and the parameters a and b which define the region and the extent of the exponential tails. Therefore as a and b are increased the Huber function relaxes back to a Gaussian distribution.

Table 1. Comparison of a non-Gaussian likelihood's and non-Gaussian prior's effect on the observation impact

non-Gaussian likelihood/Gaussian prior	non-Gaussian prior/Gaussian likelihood
$\frac{\partial \mu_a}{\partial \mu_y} = 1 - \sigma_a^2/\sigma_x^2$ in the scalar case.	$\frac{\partial \mu_a}{\partial \mu_y} = \sigma_a^2/\sigma_y^2$ in the scalar case
Sensitivity is bounded by $-\infty$ and 1.	Sensitivity is bounded by 0 and ∞.
As a function of innovation the peak in sensitivity coincides with a *minimum* in analysis error covariance.	As a function of innovation the peak in sensitivity coincides with a *maximum* in analysis error covariance.
Variability (as a function of d) in the error of the Gaussian approximation to sensitivity is *smaller* than the error in the relative entropy.	Variability (as a function of d) in the error of the Gaussian approximation to sensitivity is *larger* than the error in the relative entropy.
On average Gaussian approximation *underestimates* observation impact.	On average Gaussian approximation *overestimates* observation impact.
The error in the Gaussian approximation to the average sensitivity is *larger* than the error in the Gaussian approximation to mutual information (the average relative entropy).	The error in the Gaussian approximation to the average sensitivity is *smaller* than the error in the Gaussian approximation to mutual information.

The Huber function (Huber, 1973), results in a mixture of the l_2 norm traditionally used in variational data assimilation when the residual, δ, is small (analogous to a Gaussian distribution) and l_1 norm when the residual is large. Compared to a Gaussian with the same standard deviation this distribution is more peaked and has fatter tails. As such this is poorly represented by the GM_2 distribution. In particular the Huber norm leads to distributions with positive kurtosis values, while the GM_2 distribution can only give negative kurtosis values for a symmetric distribution. As with the GM_2 distribution it is possible to model skewed distributions with the Huber function when $|a| \neq |b|$.

Despite the differences between the Huber function and GM_2, the same general conclusions already made about observation impact can be applied:

(1) The sensitivity can be a strong function of the innovation:
This is illustrated in Fig. 7. In this example $a = -0.5$, $b = 1$ and $\sigma^2 = 2$. It is seen that the analysis sensitivity reduces to zero as the observed value gets further from the prior ($|d|$ increases), clear evidence that the Huber function robustly ensures that useful observations contribute to the analysis whilst observations inconsistent with the prior have no impact. This is in contrast to when the likelihood is assumed to be Gaussian and the sensitivity is constant (dashed line). From eq. (8) we can conclude that the peak in sensitivity close to high prior probability coincides with a minimum in the analysis error variance and as $|d|$ increases the analysis error variance tends towards that of the background.

(2) The error in the relative entropy assuming a Gaussian likelihood is of a greater magnitude than the error in the sensitivity:
This is illustrated in Fig. 8. The error in the relative entropy is also asymmetric unlike the error in the sensitivity which is symmetric. This was explained in Section 3.

(3) On average the observation impact is underestimated when a Gaussian likelihood is assumed:
This is also illustrated in Fig. 8. As was seen in the previous section, the Gaussian approximation to mutual information (red) is much poorer that the Gaussian approximation to the averaged sensitivity (black dashed line).

5. Conclusions and discussion

This work has followed on from the work of Fowler and van Leeuwen (2012), in which the effect of a non-Gaussian prior on observation impact was studied. Here we have compared this to the effect of a non-Gaussian likelihood (non-Gaussian observation error).

There has been much recent research activity in developing non-Gaussian data assimilation methods which are applicable to the Geosciences. It is assumed that by providing a more detailed and accurate description of the error distributions that the information provided by the observations and models will be used in a more optimal way. The aim of this work has been to understand how moving away from the Gaussian assumptions traditionally made in data assimilation will affect the impact that observations have. This analytical study differs from previous studies of observation impact in non-Gaussian DA, such as Bocquet (2008) and Kramer et al. (2012), in which particular case studies were considered.

In Gaussian data assimilation it is known that the impact of observations on the analysis, as measured by the analysis sensitivity to observations and mutual information, can be understood by studying the ratio of $\mathbf{HBH}^{\mathrm{T}}$ to $\mathbf{R}$. To use

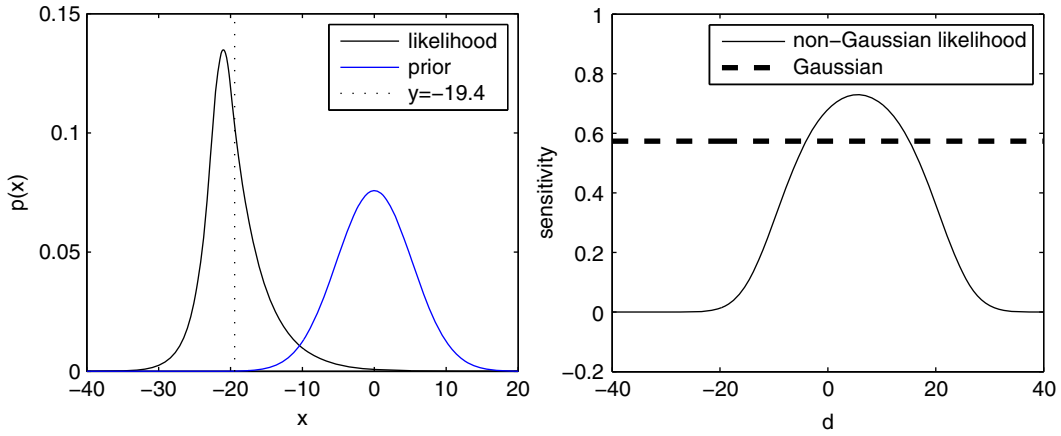

Fig. 7. Left: Huber likelihood (black) $a = -0.5$, $b = 1$ and $\sigma^2 = 2$, when $\mu_y = -19.4$. Gaussian prior (blue), $\mu_x = 0$, variance of prior chosen such that for the variance of the full likelihood (σ_y^2 not σ^2) σ_y^2/σ_x^2 is 32/43. Right: The sensitivity of the analysis to the observation.

relative entropy to measure the observation impact, it is also necessary to know the values of the observation and the prior estimate of the state. When the assumption of Gaussian statistics are relaxed we have shown that the impact of the observations on the analysis becomes much more complicated and a deeper understanding of the metric used to measure the impact as well as the source of the non-Gaussian structure is necessary.

We have shown that there exists an interesting asymmetry in the relationship of the analysis sensitivity to the analysis error covariance between the two sources of non-Gaussian structure. This means that relaxing the assumption of a Gaussian likelihood has a very different effect on observation impact, as given by this metric, than relaxing the assumption of a Gaussian prior. The sensitivity's dependence upon the analysis error variance only also

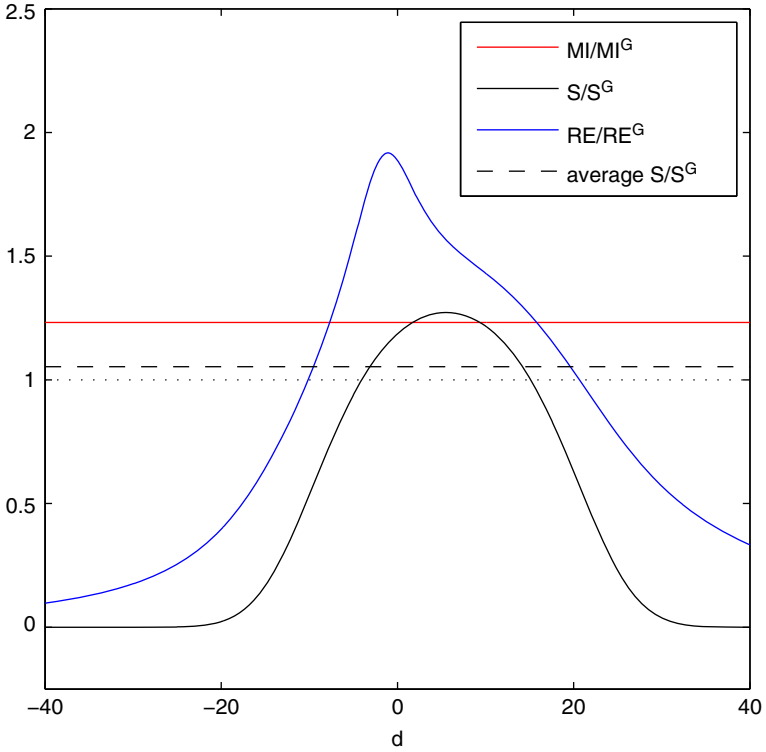

Fig. 8. Mutual information (red), sensitivity (black), relative entropy (blue) and the average sensitivity (black dashed) all normalised by their Gaussian approximations plotted as a function of d.

means that it does not measure the full influence of the observations and may erroneously indicate a degradation of the analysis due to the assimilation of the observations. From this we can conclude that the sensitivity of the analysis to the observations is no longer a useful measure of observation impact when non-Gaussian errors are considered. However the fact that it is possible to derive analytically its relationship with the analysis error covariance, has helped to give us insight into the different effects the two sources of non-Gaussian structure have on relative entropy and mutual information. These measures are much more suitable in the case of non-Gaussian error statistics because they take into account the effect of the observations on the full posterior distribution whilst still having a clear physical interpretation. Mutual information has the added benefit that it is independent of the observation value, and so provides a consistent measure of the observation impact as an experiment is repeated. It is also always positive by construction and so will always measure an improvement in our estimate of the state when observations are assimilated. However, as a consequence mutual information is more difficult to measure in non-Gaussian data assimilation because it involves averaging over observation space.

We have illustrated these findings in the case when the non-Gaussian distribution is modelled by a two component Gaussian mixture. This has allowed for an analytical study of the effect of increasing the skewness and bimodality on observation impact, and has helped to emphasise the differing effect of the source of the non-Gaussian structure on observation impact. The key conclusions from this analytical study have been shown to be applicable to other non-Gaussian distributions such as the Huber function.

The work presented here has been restricted to the case when the map between observation and state space is linear. However, there are many observation types which are not linearly related to state variables, for example, satellite radiances are a non-linear function of temperature, humidity, etc., throughout the depth of the atmosphere. In this case, even if the observation error were Gaussian, a non-linear observation operator would result in a non-Gaussian likelihood in state space. The results shown in this paper are not directly applicable to this source of non-Gaussianity, as the structure of the likelihood function in state space is now dependent on the observation value. This makes an analytical study much more difficult as shown in appendix A.3. A study of the effect of a non-linear observation operator is left for future work.

6. Acknowledgements

This work has been funded by the National Centre for Earth Observation (NCEO) part of the Natural Environment Research Council (NERC), project code H5043718, and forms part of the 'ESA Advanced Data Assimilation Methods' project, contract number ESRIN 4000105001/11/I-LG. We would also like to thank Carlos Pires and two anonymous reviewers for their valuable comments.

7. Appendix

A.1. The sensitivity of the analysis mean to the likelihood mean

Bayes' theorem states that the probability of the state, $\mathbf{x}$, given information $\mathbf{y}$ can be derived from the prior probability of $\mathbf{x}$ and the likelihood.

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})}. \tag{20}$$

where $p(\mathbf{y}) = \int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}$.
The analysis can be given by the mean of the posterior,

$$\boldsymbol{\mu}_a = \int \mathbf{x}p(\mathbf{x}|\mathbf{y})d\mathbf{x}. \tag{21}$$

Substituting eqs. (20) into (21) we see that the analysis is only dependent on the observation value through the likelihood, $p(\mathbf{y}|\mathbf{x})$. The sensitivity of the analysis in observation space to the observation value, $\mathbf{y} = \boldsymbol{\mu}_y(+const)$, is then given by

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = \frac{\int \mathbf{H}\mathbf{x}p(\mathbf{x})\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y}d\mathbf{x}}{\int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}} - \mathbf{H}\boldsymbol{\mu}_a\frac{\int p(\mathbf{x})\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y}d\mathbf{x}}{\int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}}. \tag{22}$$

Recall that $\mathbf{H}$ is the (linear) operator which transforms a vector from state to observation space.
It is also of interest to look at the sensitivity of the analysis to the mean of the prior, $\boldsymbol{\mu}_x$, in observation space. In this case it is only the prior, $p(\mathbf{x})$, in eq. (20) which is sensitive to $\boldsymbol{\mu}_x$ and so the sensitivity of the analysis to the mean of the prior is given by

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x} = \frac{\int \mathbf{H}\mathbf{x}p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}d\mathbf{x}}{\int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}} - \mathbf{H}\boldsymbol{\mu}_a\frac{\int p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}d\mathbf{x}}{\int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}}. \tag{23}$$

In the following subsections we will show that it is possible to evaluate these sensitivities when either the prior or likelihood are Gaussian.

A.2. Non-Gaussian prior, Gaussian likelihood

Let $p(\mathbf{x})$ be arbitrary and $p(\mathbf{y}|\mathbf{x})$ be Gaussian with mean $\boldsymbol{\mu}_y$and error covariance $\mathbf{R}$.

$$p(\mathbf{y}|\mathbf{x}) = ((2\pi)^m|\mathbf{R}|)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_y - \mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{\mu}_y - \mathbf{H}\mathbf{x})\right]. \tag{24}$$

Here m is the size of observation space. Therefore

$$\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y} = -p(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_y - \mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}. \tag{25}$$

A.2.1. Analysis sensitivity to observations. Equation (25) can be substituted into eq. (22) to give

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = -\mathbf{H}\boldsymbol{\mu}_a\,\boldsymbol{\mu}_y^{\mathrm{T}}\mathbf{R}^{-1} + \int \mathbf{H}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}p(\mathbf{x}|\mathbf{y})d\mathbf{x} + \mathbf{H}\boldsymbol{\mu}_a\boldsymbol{\mu}_y^{\mathrm{T}}\mathbf{R}^{-1} - \mathbf{H}\boldsymbol{\mu}_a\boldsymbol{\mu}_a^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{26}$$

Note that $\int \mathbf{x}\mathbf{x}^{\mathrm{T}}p(\mathbf{x}|\mathbf{y})d\mathbf{x} - \boldsymbol{\mu}_a\boldsymbol{\mu}_a^{\mathrm{T}}$ is the analysis error covariance matrix, $\mathbf{P}_a$. eq. (26) therefore simplifies to

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = \mathbf{H}\mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{27}$$

A.2.2. Analysis sensitivity to background. In calculating the sensitivity with respect to the background (the mean of the prior), in this case, we do not have access to $\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}$. However we do know $\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x} = -\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\mathbf{x}}$ as the change in the probability caused by perturbing the value of $\mathbf{x}$ is the same as perturbing $\boldsymbol{\mu}_x$ by the same magnitude but in the opposite direction. Therefore we can utilise integration by parts, that is, $\int \mathbf{u}\frac{\partial \mathbf{v}}{\partial \mathbf{x}}d\mathbf{x} = \mathbf{u}\mathbf{v} - \int \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{v}d\mathbf{x}$.

First evaluate the first term of eq. (23):

Let $\mathbf{u} = \mathbf{H}\mathbf{x}p(\mathbf{y}|\mathbf{x})$and $\frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\mathbf{x}}$. Therefore $\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{H}p(\mathbf{y}|\mathbf{x})(\mathbf{I}_n + \mathbf{x}(\boldsymbol{\mu}_y - \mathbf{H}\mathbf{x})^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})$. $\mathbf{v}$ can be found by noting $\frac{\partial p(\mathbf{x})}{\partial \mathbf{x}}\mathbf{H}^{\mathrm{T}} = \frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\mathbf{x}}\frac{\partial \mathbf{H}\mathbf{x}}{\partial \mathbf{x}}\mathbf{H}^{\mathrm{T}}$, it then follows that $\mathbf{v} = p(\mathbf{x})\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}$.

$$\begin{aligned}\int \mathbf{H}\mathbf{x}p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}d\mathbf{x} &= -\int \mathbf{H}\mathbf{x}p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\mathbf{x}}d\mathbf{x} \\ &= -\left[p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{H}\mathbf{x}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\right]_{-\infty}^{\infty} \\ &+ \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{H}(\mathbf{I}_n + \mathbf{x}(\boldsymbol{\mu}_y - \mathbf{H}(\mathbf{x})^{\mathrm{T}})\mathbf{R}^{-1}\mathbf{H})\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}d\mathbf{x}.\end{aligned} \tag{28}$$

The first term of eq. (23) is then

$$\mathbf{I}_m + \mathbf{H}\boldsymbol{\mu}_a\boldsymbol{\mu}_y^{\mathrm{T}}\mathbf{R}^{-1} - \mathbf{H}\int p(\mathbf{x}|\mathbf{y})\mathbf{x}\mathbf{x}^{\mathrm{T}}d\mathbf{x}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{29}$$

Likewise we may evaluate the second term of eq. (23):

Let $u = p(\mathbf{y}|\mathbf{x})$ and $\frac{\partial \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}$ again. Therefore $\frac{\partial u}{\partial \mathbf{x}} = (\boldsymbol{\mu}_y - \mathbf{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}p(\mathbf{y}|\mathbf{x})$ and v is unchanged.

$$\begin{aligned}\int p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x}d\mathbf{x} &= -\int p(\mathbf{y}|\mathbf{x})\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\mathbf{x}}d\mathbf{x} \\ &= -\left[p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}\right]_{-\infty}^{\infty} \\ &+ \int (\boldsymbol{\mu}_y - \mathbf{H}(\mathbf{x}))^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}d\mathbf{x}.\end{aligned} \tag{30}$$

The second term of eq. (23) is then

$$-\mathbf{H}\boldsymbol{\mu}_a\boldsymbol{\mu}_y^{\mathrm{T}}\mathbf{R}^{-1} + \mathbf{H}\boldsymbol{\mu}_a\boldsymbol{\mu}_a^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}. \tag{31}$$

It then follows that

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x} = \mathbf{I}_m - \mathbf{H}\mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1} \tag{32}$$

A.3. Non-Gaussian likelihood, Gaussian prior

Let $p(\mathbf{y}|\mathbf{x})$ be arbitrary and $p(\mathbf{x})$ be Gaussian with mean $\boldsymbol{\mu}_x$and error covariance $\mathbf{B}$.

$$p(\mathbf{x}) = ((2\pi)^p|\mathbf{B}|)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\boldsymbol{\mu}_x - \mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{\mu}_x - \mathbf{x})\right]. \tag{33}$$

Therefore

$$\frac{\partial p(\mathbf{x})}{\partial \mathbf{H}\boldsymbol{\mu}_x} = -p(\mathbf{x})(\boldsymbol{\mu}_x - \mathbf{x})^{\mathrm{T}}\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}. \tag{34}$$

A.3.1. Analysis sensitivity to observations. The derivation of $\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y}$ is analogous to the derivation of $\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x}$ in the previous section, as $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y}$ in this case is unknown and so we must make use of integration by parts.

It can be shown that in this case

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = \mathbf{I}_m - \mathbf{H}\mathbf{P}_a\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1} \tag{35}$$

A.3.2. Analysis sensitivity to background. The derivation of $\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x}$ is analogous to the derivations of $\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y}$ in the previous section as $\frac{\partial p(\mathbf{x})}{\partial \boldsymbol{\mu}_x}$ in this case is known.

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x} = \mathbf{H}\mathbf{P}_a\mathbf{B}^{-1}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{H}^{\mathrm{T}})^{-1}. \tag{36}$$

From this we can conclude that when either the prior or likelihood is Gaussian it is always the case that:

$$\frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} + \frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \mathbf{H}\boldsymbol{\mu}_x} = \mathbf{I}_m \tag{37}$$

A.4. Non-linear observation operator

Following a similar methodology as in Appendix A.2 and A.3, the analysis sensitivity to the observations when the

observation operator is non-linear, represented by $h(\mathbf{x})$, can be shown to be

$$\frac{\partial h(\boldsymbol{\mu}_a)}{\partial \boldsymbol{\mu}_y} = \mathbf{H}\left(\int \mathbf{x}(h(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} - \boldsymbol{\mu}_a \int (h(\mathbf{x}))^{\mathrm{T}} \mathbf{R}^{-1} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \right), \tag{38}$$

where $\mathbf{H}$ is the observation operator linearised about the analysis. In this expression it has been assumed that the likelihood in observation space is Gaussian, i.e. $\mathbf{y} \sim N(\boldsymbol{\mu}_y, \mathbf{R})$ but no assumptions about the prior have been necessary. When $h(\mathbf{x})$ is non-linear there is no longer a clear relationship between the sensitivity and the analysis error covariance matrix.

References

Bishop, C. 2006. *Pattern Recognition and Machine Learning*. Springer Science + Business Media, LLC, New York.

Bocquet, M. 2008. Inverse modelling of atmospheric tracers: non-gaussian methods and second-order sensitivity analysis. *Nonlinear Process. Geophys.* **15**, 127–143.

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

Cardinali, C., Pezzulli, S. and Andersson, E. 2004. Influence-matrix diagnostics of a data assimilation system. *Q. J. Roy. Meteorol. Soc.* **130**, 2767–2786. DOI: 10.1256/qj.03.205.

Cover, T. M. and Thomas, J. A. 1991. *Elements of Information Theory (Wiley Series in Telecommunications)*. John Wiley, New York.

Desroziers, G., Berre, L., Chabot, V. and Chapnik, B. 2009. A posteriori diagnostics in an ensemble of perturbed analyses. *Mon. Wea. Rev.* **137**, 3420–3436.

Dunn, R. H. J., Willet, K. M., Thorne, P. W., Woolley, E. V., Durre, I. and co-authors. 2012. HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Clim. Past.* **8**, 1649–1679.

Eyre, J. E. 1990. The information content of data from satellite sounding systems: a simulation study. *Q. J. Roy. Meteorol. Soc.* **116**, 401–434. DOI: 551.501.7:551.507.362.2.

Fowler, A. M. and van Leeuwen, P. J. 2012. Measures of observation impact in non-gaussian data assimilation. *Tellus.* **64**, 17192.

Gandin, L. S., Monrone, L. L. and Collins, W. G. 1993. Two years of operational comprehensive hydrostatic quality control at the national meteorological center. *Weather. Forecast.* **57**, 57–72.

Houtekamer, P. L. and Mitchell, H. L. 1998. Data assimilation using an ensemble kalman filter technique. *Mon. Wea. Rev.* **126**, 798–811.

Huber, P. J. 1973. Robust regression: asymptotics, conjectures, and monte carlo. *Ann. Stat.* **1**, 799–821.

Ingleby, B. and Lorenc, A. 1993. Bayesian quality control using multivariate normal distributions. *Q. J. Roy. Meteorol. Soc.* **119**, 1195–1225.

Janjić, T. and Cohn, S. 2006. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev.* **134**, 2900–2915.

Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, New York.

Kramer, W., Dijkstra, H. A., Pierini, S. and van Leeuwen, P. J. 2012. Measuring the impact of observations on the predictability of the kuroshio extension in a shallow-water model. *J. Phys. Oceanogr.* **42**, 3–17.

Martin, S. 2004. *An Introduction to Ocean Remote Sensing*. Cambridge University Press, Cambridge pp.

Palmer, T. N., Gelaro, R., Barkmeijer, J. and Buizza, R. 1998. Singular vectors, metric, and adaptive observations. *J. Atmos. Sci.* **55**, 633–653.

Peckham, G. 1974. The information content of remote measurements of the atmospheric temperature by satellite ir radiometry and optimum radiometer configurations. *Q. J. Roy. Meteorol. Soc.* **100**, 406–419.

Pires, C. A., Talagrand, O. and Bocquet, M. 2010. Diagnosis and impacts of non-gaussianity of innovations in data assimilation. *Pysica. D.* **239**, 1701–1717.

Qin, Z.-K., Zou, X., Li, G. and Ma, X.-L. 2010. Quality control of surface station temperature data with non-gaussian observation-minus-background distributions. *J. Geophys. Res.* **115**. DOI: 10.1029/2009JD013695.

Rabier, F., Fourrié, N., Chafai, D. and Prunet, P. 2002. Channel selection methods for infrared atmospheric sounding interferometer radiances. *Q. J. Roy. Meteorol. Soc.* **128**, 1011–1027.

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

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D. and co-authors. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Q. J. Roy. Meteorol. Soc.* **133**, 347–362.

Rodgers, C. D. 1996. Information content and optimisation of high spectral resolution measurements. *Proc. SPIE.* **2830**, 136–147.

Rodgers, C. D. 2000. *Inverse Methods for Atmospheric Sounding*. World Scientific, Singapore pp.

Talagrand, O. 2003. *Bayesian Estimation. Optimal Interpolation. Statistical Linear Estimation in Data Assimilation for the Earth System, Volume 26 of NATO Science Series IV Earth and Environmental Sciences*. Springer, Netherlands.

Tavolato, C. and Isaksen, L. Huber norm quality control in the IFS. ECMWF Newsletter No. 122, 2009/2010.

van Leeuwen, P. J. 2003. A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.* **131**, 2071–2084.

van Leeuwen, P. J. 2009. Particle filtering in geophysical systems. *Mon. Wea. Rev.* **137**, 4089–4114.

Wahba, G. 1985. Design criteria and eigensequence plots for satellite-computed tomography. *J. Atmos. Oceanic. Technol.* **2**, 125–132.

Xu, Q., Wei, L. and Healy, S. 2009. Measuring information content from observations for data assimilation: connection between different measures and application to radar scan design. *Tellus.* **61A**, 144–153.