Measures of observation impact in non-Gaussian data assimilation

Alison Fowler* | Peter Jan van Leeuwen

University of Reading

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Abstract

Here we study the effect of relaxing the assumption of Gaussian errors on the impact of the observations on the analysis.

Introduction

It has traditionally been assumed in data assimilation (DA) that the error distributions of the prior and likelihood are Gaussian. However there are many different potential sources of **non-Gaussianity**, e.g. if the prior information has come from a highly non-linear forecast model. Approximating the error distributions as Gaussian can have a detrimental impact on the accuracy of the analysis . As such there has been much interest in developing non-Gaussian/non-linear DA methods for the Geosciences. Many of these methods are based on the direct application of **Bayes' theorem** to find the full **posterior distribution**, $p(\mathbf{x}|\mathbf{y})$:

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

In applying Bayes' theorem there is no restriction on the choice of **prior**, $p(\mathbf{x})$, or **likelihood**, $p(\mathbf{y}|\mathbf{x})$.

A measure of the impact observations are having in these new methods of DA can help to make the best use of the available data. This information can also allow us to understand the effect of approximating the error distributions as Gaussian.

Measures of observation impact

There are many measures of observation impact, here we introduce 3:

The analysis sensitivity to observations

This is given by

$$\mathbf{S} = \frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y},$$

where $\boldsymbol{\mu}_a$ is the mean of the posterior distribution, our **analysis**, $\boldsymbol{\mu}_y$ is the mean of the likelihood, our **observation**, and $\mathbf{H}$ is the linearised transform from observation to state space.

Relative entropy

This measures the **entropy** (uncertainty) of the posterior relative to the prior. It is sensitive to the change in the position and shape of the posterior.

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

Mutual information

Mutual information is a measure of the change in entropy when an observation is made. This is only sensitive to a change in the shape of the posterior.

$$MI = -\int p(\mathbf{x}) lnp(\mathbf{x}) d\mathbf{x} + \int p(\mathbf{y}) \int p(\mathbf{x}|\mathbf{y}) lnp(\mathbf{x}|\mathbf{y}) d\mathbf{x}d\mathbf{y}.$$

Note $MI = \int p(\mathbf{y}) RE d\mathbf{y}$

Observation impact in Gaussian DA

In Gaussian data assimilation explicit expressions can be given for each of the three measures of observation impact:

- $\mathbf{S}^G = \mathbf{HP_aH^TR^{-1}}$, where $\mathbf{P_a}[= (\mathbf{B^{-1}} + \mathbf{H^TR^{-1}H})^{-1}]$ and $\mathbf{R}$ are the analysis and observation error covariance matrices respectively (Cardinali et al. (2004)). $\mathbf{B}$ is the prior error covariance matrix.
- $MI^G = \frac{1}{2} ln|\mathbf{BP_a}^{-1}|$ (Rodgers (2000)).
  - This can also be written in terms of the eigenvalues of $\mathbf{S}^G$. $MI^G = -\frac{1}{2}\sum_{i=1}^{r} \ln(1-\lambda_i)$, where λ_i is the i^{th} eigenvalue of $\mathbf{S}^G$ and r is the rank of $\mathbf{S}^G$
- $RE^G = \frac{1}{2}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_x)^T\mathbf{B}^{-1}(\boldsymbol{\mu}_a - \boldsymbol{\mu}_x) + \frac{1}{2} ln|\mathbf{BP_a}^{-1}| + \frac{1}{2} tr(\mathbf{B^{-1}P_a}) - \frac{n}{2}$, where $\boldsymbol{\mu}_x$ is the mean of the prior and n is the size of state space.

Except for relative entropy these are all dependant solely on $\mathbf{H}$, $\mathbf{B}$ and $\mathbf{R}$. The first term of relative entropy is a quadratic function of the observation value, as $\boldsymbol{\mu}_a$ is linearly related to $\boldsymbol{\mu}_y$. However a study by Xu (2009) found that when defining the optimum radar scan configurement it did not matter which measure was used, suggesting RE^G, in this case, had a small dependence on $\boldsymbol{\mu}_y$ compared its dependence on the ratio $\mathbf{HBH^TR^{-1}}$.

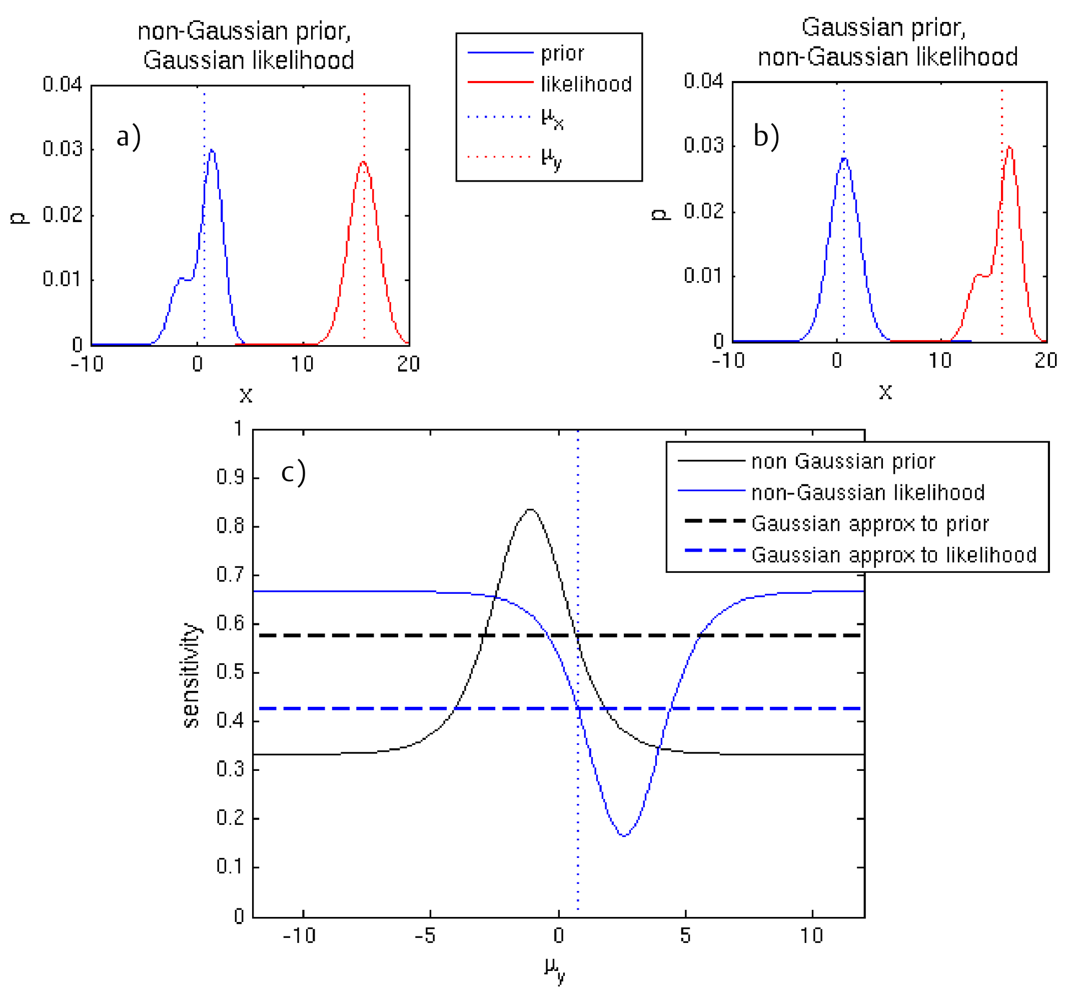

Figure 1. Sensitivity in non-Gaussian DA: A comparison between the sensitivity of the analysis to observations when I: the prior is non-Gaussian and when II: the likelihood is non-Gaussian.
a) The likelihood and prior in case I.
b) The likelihood and prior in case II.
c) The sensitivities in the two cases (solid lines) along with their estimates when the prior/likelihood is approximated by a Gaussian (dashed lines).

Sensitivity in non-Gaussian DA

The sensitivity of the analysis w.r.t to the observation can be calculated when either the likelihood is Gaussian and the prior is arbitrary or vice versa.

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

Arbitrary prior, Gaussian likelihood

In this case $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y}$ is known to be $-p(\mathbf{y}|\mathbf{x})(\boldsymbol{\mu}_y - \mathbf{H}(\mathbf{x}))^T\mathbf{R}^{-1}$ and so it is easy to prove

$$\mathbf{S}^{nGp(x)} = \frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = \mathbf{HP}_a\mathbf{H^TR^{-1}}. \quad (1)$$

This has the same form as in Gaussian DA but now $\mathbf{P}_a$ is a function of the observation value.

Gaussian prior, Arbitrary likelihood (but linear observation operator)

In this case $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y}$ is unknown however $\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \boldsymbol{\mu}_y} = -\frac{\partial p(\mathbf{y}|\mathbf{x})}{\partial \mathbf{x}}$, and so we can use integration by parts to find

$$\mathbf{S}^{nGp(y|x)} = \frac{\partial \mathbf{H}\boldsymbol{\mu}_a}{\partial \boldsymbol{\mu}_y} = \mathbf{I}_p - \mathbf{HP}_a\mathbf{B}^{-1}\mathbf{H}^T(\mathbf{HH}^T)^{-1}. \quad (2)$$

When the likelihood is Gaussian, and $\mathbf{P}_a = (\mathbf{B^{-1}} + \mathbf{H^TR^{-1}H})^{-1}$, this is also the same form as in Gaussian DA.

As a function of $\boldsymbol{\mu}_y$ there is an interesting symmetry between $\mathbf{S}^{nGp(x)}$ and $\mathbf{S}^{nGp(y|x)}$, as one is proportional to the analysis error covariance and the other is inversely proportional to the analysis error covariance. This is illustrated in Fig. 1.

Comparison of measures in non-Gaussian DA

In Fig 2. we compare the error in approximating the different measures of observation impact assuming Gaussian distributions as a function of $\boldsymbol{\mu}_y$.

Relative entropy is sensitive to both i) the change in the position of the posterior compared to the prior and ii) the change in its shape. The error in the change in the position when making Gaussian approximations will be zero when $\boldsymbol{\mu}_y = \boldsymbol{\mu}_x$ and proportional to the error in the sensitivity away from this. The error in the change of shape is inversely related to the error in $\mathbf{P}_a$. This part of the error in relative entropy is therefore inversely related to the sensitivity when the prior is non-Gaussian and directly related to the sensitivity when the likelihood is non-Gaussian. This explains the difference between Fig, 2a) (non-Gaussian prior) and Fig. 2b) (non-Gaussian likelihood). This also means there is a difference in the error in MI which is the expected value of relative entropy.

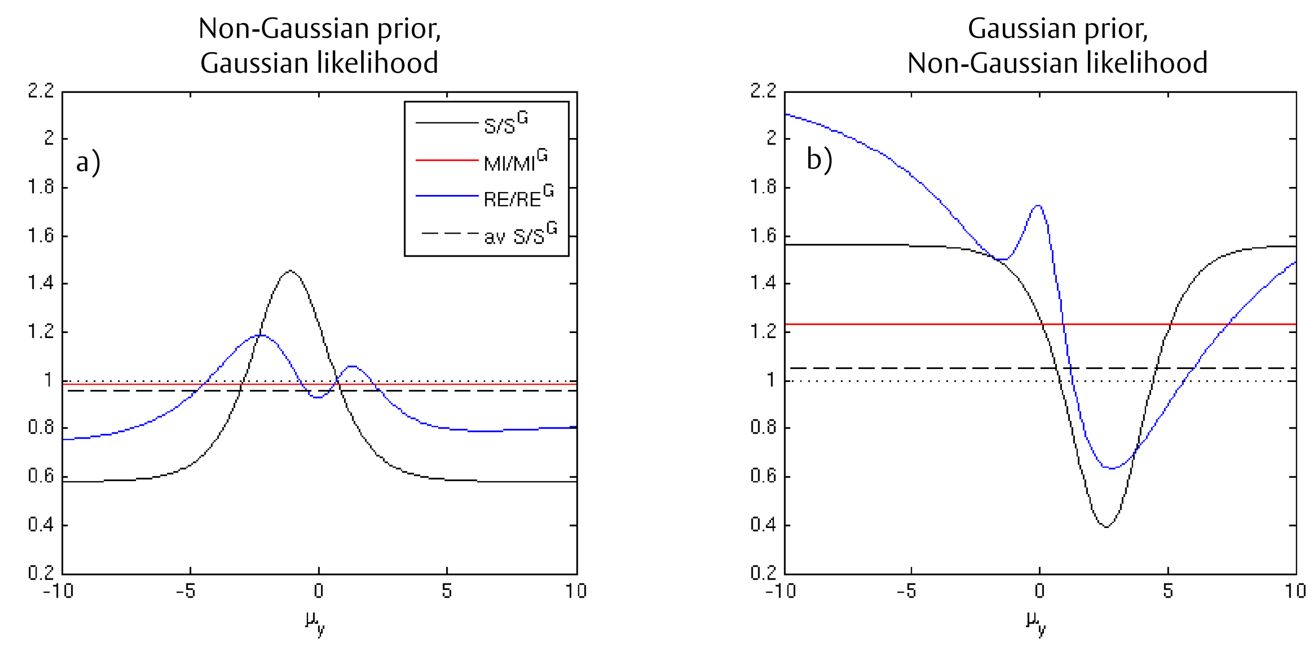

Figure 2. A comparison between the different measures of observation impact each normalised by their Gaussian approximation.
a) When the prior is non-Gaussian
b) When the likelihood is non-Gaussian.

Conclusions

- **For a given observation value, the error in observation impact when Gaussian distributions are assumed can be very different depending on how it is measured.**
- **When the prior is non-Gaussian the error in relative entropy has a smaller range of values as a function of the observation than the error in the sensitivity and vice versa when the likelihood is non-Gaussian. This can be understood by the different relationship the sensitivity has with the analysis error covariances in each of theses cases (see Eqns. (1) and (2)).**
- **Mutual information, which is the expected relative entropy, tends to be much closer to it's Gaussian approximation. However the error in the Gaussian approximation is again sensitive to the source of the non-Gaussianity.**

Contact information*

Department of Meteorology, University of Reading, Whiteknights, RG6 6AH.
Email: a.m.fowler@reading.ac.uk
Web: www.met.reading.ac.uk/~gj901587

References.

Observation impact in Gaussian DA

- Cardinali et al., 2004: Influence-matrix diagnostics of a data assimilation system, *Q. J. R. Met. Soc.*, **130**, 2767-2786.
- Rodgers, 2000: Inverse methods for atmospheric sounding.
- Xu, et al., 2009: Measuring information content from observations for data assimilation: connection between different measures and application to radar scan design. *Tellus*, **61A**, 144-153.

Observation impact in non-Gaussian DA

- Fowler and van Leeuwen, 2012: Measures of observation impact in non-Gaussian data assimilation. *Accepted by Tellus A.*
- Fowler and van Leeuwen, 2012: Measures of observation impact in data assimilation: the effect of a non-Gaussian likelihood. *In Preparation.*