Measuring the impact of observations

Alison Fowler

School of Mathematical and Physical Sciences,
University of Reading.

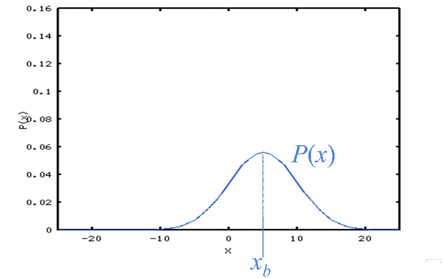
Data assimilation and visualisation in environmental sciences.
15-19 September 2014

Introduction

The general aim of data assimilation is to improve a forecast by reducing the error in the initial conditions.
Observations, $\mathbf{y}$, and a-priori data, $\mathbf{x}_b$, are combined utilising a statistical description of their respective errors and a description of the relationship between state and observation space, $h(\mathbf{x})$. The resulting analysis of the current state, $\mathbf{x}_a$, then becomes the initial conditions for the next forecast.

Introduction

scalar Gaussian example

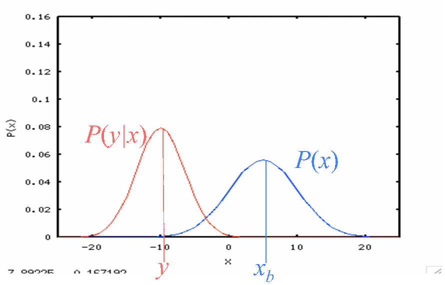


Notation

$\mathbf{x}_b$: background

Introduction

scalar Gaussian example

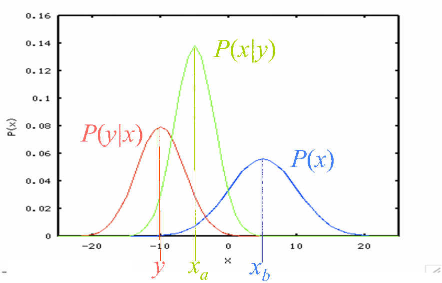


Notation

$\mathbf{x}_b$: background $\mathbf{y}$: observation

Introduction

scalar Gaussian example

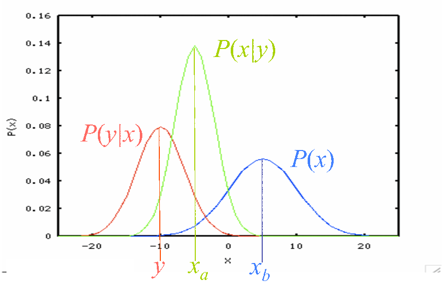


$$x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2+\sigma_y^2}(y - x_b)$$

$$\sigma_a^2 = \frac{\sigma_b^2\sigma_y^2}{\sigma_b^2+\sigma_y^2}$$

Introduction

scalar Gaussian example



$$x_a = x_b + \frac{\sigma_b^2}{\sigma_b^2+\sigma_y^2}(y - x_b)$$
$$\sigma_a^2 = \frac{\sigma_b^2\sigma_y^2}{\sigma_b^2+\sigma_y^2}$$

Multiple dimensions

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left(\mathbf{y} - h(\mathbf{x}_b)\right) \quad (1)$$

$$\mathbf{P}_a = \left(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}$$
$$\mathbf{K} = \mathbf{P}_a\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}$$

Notation

$\mathbf{x}_b$: background	$\mathbf{y}$: observation	$\mathbf{x}_a$: analysis
$\mathbf{B}$ or σ_b^2: $\mathbf{x}_b$ error variance	$\mathbf{R}$ or σ_y^2: $\mathbf{y}$ error variance	$\mathbf{P}_a$ or σ_a^2: $\mathbf{x}_a$ error variance
$\mathbf{K}$: Kalman gain	$h(\mathbf{x})$:observation operator	$\mathbf{H}$: Linearised $h(\mathbf{x})$

Introduction

Monitoring the impact of the observations is particularly important in the geosciences where observations tend to be very expensive
An objective measure of the impact of the observations may be used for:

- the assessment of the data assimilation scheme
- the design of new observing systems
- defining targeted observations
- data thinning

Introduction

Measures of observation impact can broadly be split into two type

- those measuring the impact on the analysis
- those measuring the impact on the forecast

It can be expected that observations which have a large impact on the analysis will also have a large impact on the forecast. However care must be taken in comparing observations with different dynamical types due to the different roles they have in the forecast.

Influence matrix

The influence matrix measures the sensitivity of the analysis in observation space to the observations (Cardinali et al., 2004).

$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}} \tag{2}$$

This is a $p \times p$ matrix, where p is the number of observations, allowing for the most influential observations or groups of observations to be identified.
From the expression for the Gaussian analysis given by (1), we can rewrite (2) as

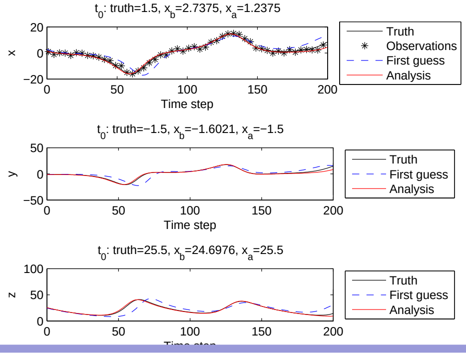
$$\mathbf{S} = \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} \equiv \mathbf{R}^{-1}\mathbf{H}\mathbf{P}_a\mathbf{H}^{\mathrm{T}}$$

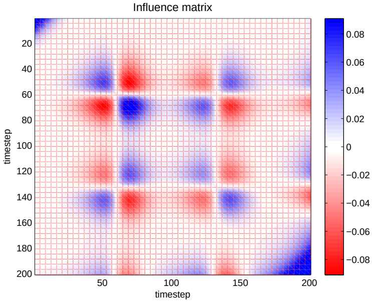
When $\mathbf{R}$ is diagonal the diagonal elements of $\mathbf{S}$ are bounded by 0 and 1.

Influence matrix

Lorenz 1963 identical twin experiment

The state is given by $\mathbf{x} = (x, y, z)^{\mathrm{T}}$ and observations are made of x at every forth time step.
$\mathbf{B} = \begin{pmatrix} 1 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 1 \end{pmatrix}$ and the observation error variance is 1 at each time step and uncorrelated in time.

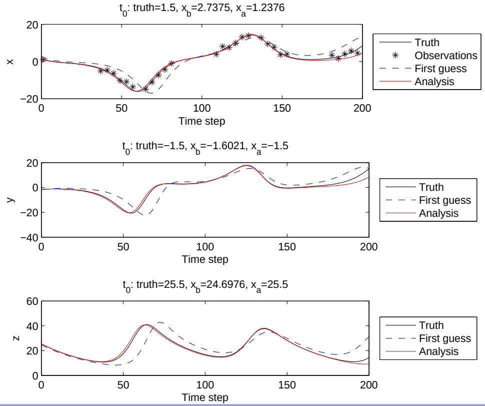




Influence matrix

Lorenz 1963 identical twin experiment

Can use the influence matrix to reduce the number of observations in the assimilation. In this case the observations have been reduced by 60% by summing up the magnitude of the values in each column of S and taking the 30 observations corresponing to the greatest influence on the analysis.



Degrees of freedom for signal

The degrees of freedom in data assimilation is given by the expectation of the cost function evaluated at the analysis. This can be shown to equal the number of observations, p.

The degrees of freedom may be split into two parts; that measuring the signal, d_s, and that measuring the noise, d_n. $d_s + d_n = p$.

$$d_s = E[(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_b)] \tag{3}$$

From the expression for the Gaussian analysis given by (1), we can rewrite (3) as

$$d_s = trace(\mathbf{HK}) \equiv trace(\mathbf{S})$$

Hence d_s is bounded by 0 and p.

Can look at subset of observations by calculating $d_{s(i)} = trace(\mathbf{S}_i)$. $d_{s(i)}$ is then bounded by 0 and p_i.

Degrees of freedom for signal

The degrees of freedom in data assimilation is given by the expectation of the cost function evaluated at the analysis. This can be shown to equal the number of observations, p.

The degrees of freedom may be split into two parts; that measuring the signal, d_s, and that measuring the noise, d_n. $d_s + d_n = p$.

$$d_s = E[(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b)] \tag{3}$$

From the expression for the Gaussian analysis given by (1), we can rewrite (3) as

$$d_s = trace(\mathbf{HK}) \equiv trace(\mathbf{S})$$

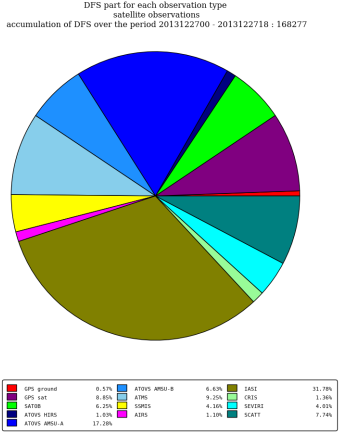
Hence d_s is bounded by 0 and p.

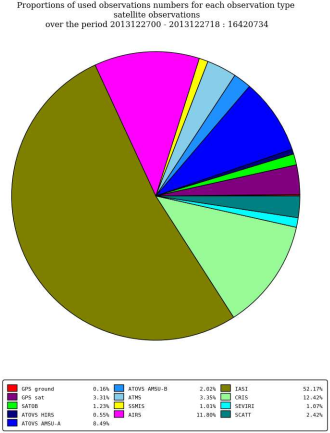
Can look at subset of observations by calculating $d_{s(i)} = trace(\mathbf{S}_i)$. $d_{s(i)}$ is then bounded by 0 and p_i.

For the Lorenz 1963 example shown previously, $d_s = 2.327$ when the full 50 observations were assimilated and $d_s = 1.992$ when the reduced observations were assimilated (a reduction of 86%).

Degrees of freedom for signal

Statistics from Meteo France assimilation system





Entropy measures

$\mathbf{S}$ and d_s both aim to quantify the effect of the observations on the analysis alone. This may give a limited view of the impact the observation is having.

Mutual information and *relative entropy* are two measures of observation impact which quantify the effect of the observation on the change in *entropy*. Entropy is a measure of the uncertainty associated with a variable. Entropy for a single random variable is defined as

$$H(\mathbf{x}) = \int P(\mathbf{x}) \ln P(\mathbf{x}) \mathrm{d}\mathbf{x}.$$

For a Gaussian distribution the entropy is given by

$$H(\mathbf{x}) = n \ln(2\pi e)^{1/2} + \frac{1}{2} \ln |\mathbf{C}_x|,$$

where n is the size of the vector $\mathbf{x}$ and $|\mathbf{C}_x|$ is the determinant of it's error covariance matrix.

Mutual information

Mutual information (also known as Shannon information content in Gaussian data assimilation) is given by the reduction in entropy after the observations have been assimilated.

$$MI = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \tag{4}$$

For Gaussian error statistics (4) becomes

$$MI = \frac{1}{2} \ln |\mathbf{BP}_a^{-1}| \equiv -\frac{1}{2} \sum_{i=1}^{p} \ln(1 - \lambda_i),$$

where λ_i are the eigenvalues of the influence matrix.

Mutual information

Mutual information (also known as Shannon information content in Gaussian data assimilation) is given by the reduction in entropy after the observations have been assimilated.

$$MI = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) \tag{4}$$

For Gaussian error statistics (4) becomes

$$MI = \frac{1}{2}\ln|\mathbf{BP}_a^{-1}| \equiv -\frac{1}{2}\sum_{i=1}^{p}\ln(1-\lambda_i),$$

where λ_i are the eigenvalues of the influence matrix.
For the Lorenz 1963 example shown previously, $MI = 5.9952$ when the full 50 observations were assimilated and $MI = 4.0910$ when the reduced observations were assimilated (a reduction of 68%).

Comparison of Mutual information and degrees of freedom for signal

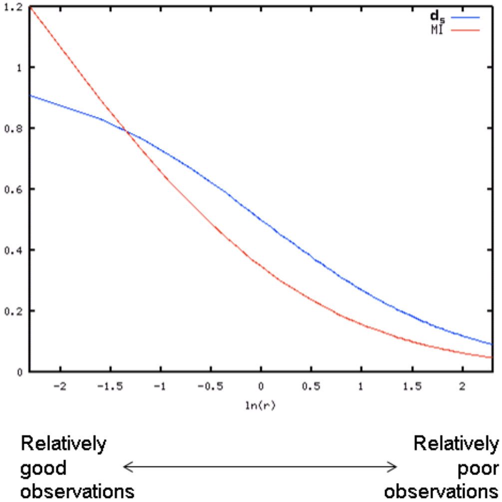
scalar Gaussian example

Let $r = \frac{\sigma_y^2}{\sigma_b^2}$.
Then

$$d_s = \frac{1}{1+r}$$

and

$$MI = \frac{1}{2} \ln \frac{1+r}{r}$$



Relative entropy

Relative entropy is a non-symmetric measure of the difference between the prior and posterior pdf.

$$RE = \int P(\mathbf{x}|\mathbf{y}) \ln \frac{P(\mathbf{x}|\mathbf{y})}{P(\mathbf{x})} d\mathbf{x} d\mathbf{y} \tag{5}$$

For Gaussian error statistics (5) becomes

$$\begin{aligned} RE &= \tfrac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) \quad + \quad \tfrac{1}{2}(\ln|\mathbf{B}\mathbf{P}_a^{-1}| \quad + \quad trace(\mathbf{B}^{-1}\mathbf{P}_a) - n), \\ &= \tfrac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) \quad + \quad MI \quad - \quad \tfrac{1}{2} d_s \end{aligned}$$

Unlike the other measures RE depends on the value of the analysis increment as well as the error covariances. It therefore cannot be calculated before an observation has been made.
From the definition of d_s it is clear that RE averaged over observation space is MI.

Relative entropy

Relative entropy is a non-symmetric measure of the difference between the prior and posterior pdf.

$$RE = \int P(\mathbf{x}|\mathbf{y}) \ln \frac{P(\mathbf{x}|\mathbf{y})}{P(\mathbf{x})} d\mathbf{x} d\mathbf{y} \tag{5}$$

For Gaussian error statistics (5) becomes

$$\begin{aligned} RE &= \tfrac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) + \tfrac{1}{2}(\ln|\mathbf{B}\mathbf{P}_a^{-1}| + trace(\mathbf{B}^{-1}\mathbf{P}_a) - n), \\ &= \tfrac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_a - \mathbf{x}_b) + MI - \tfrac{1}{2}d_s \end{aligned}$$

Unlike the other measures RE depends on the value of the analysis increment as well as the error covariances. It therefore cannot be calculated before an observation has been made.

From the definition of d_s it is clear that RE averaged over observation space is MI.

For the Lorenz 1963 example shown previously, $RE = 6.8172$ when the full 50 observations were assimilated and $RE = 5.0801$ when the reduced observations were assimilated (a reduction of 75%).

Comparison of measures

Each measure introduced provides complementary information. The choice of measure to use depend on the intended application.

- Understanding impact of individual observation types: The influence matrix is the only measure to provide information about an individual observations impact on the whole of the analysis. Degrees of freedom may be used to compare predefined subsets of observations. Mutual information has the benefit of being additive with successive observations.
- Data thinning: Degrees of freedom for signal and mutual information have both been used for channel selection of satellite instruments. This is often performed offline.
- Assessment of data assimilation scheme: Degrees of freedom for signal can be used to check the assumptions made in 4D-Var as the $E(J_b) = trace(\mathbf{S})$ *iff* these assumptions are correct. To calculate $E(J_b)$ an ensemble of 4d-Var is necessary (see Desroziers et al., 2009).

Relative entropy is unique in that it depends on the observation value therefore it is sensitive to extreme observations. It could therefore be useful for routine monitoring, identifying situations where observations are having an unusually large influence on analysis.

Observation impact on the forecast

Often the ultimate aim of data assimilation is to improve the forecast.
However, there is not always a clear relationship between observation impact on the analysis and observation impact on the forecast.
For applications such as defining targeted observations it is particularly important to take into to account the forecast dynamics and the synoptic situation.

OSSEs and data denial experiments

OSSEs (observation system simulation experiments) (e.g. Masutani et al. (2010)) and *data denial experiments* (e.g. Kelly et al. (2007)) compare a control forecast to a forecast which has had additional (simulated) observations or fewer observations assimilated.

The difference between these forecasts gives an indication of the impact of the observations on a variety of measures.

Caution is needed in careful validation and calibration when large amounts of data are added or removed from a system which may have been optimally tuned for the original set of observations (Gelaro and Zhu (2009)).

These experiments are very expensive and so only impact of a large subset of observations can be looked at one time.

Adjoint techniques

The adjoint technique as proposed by Langland and Baker (2004) approximates the sensitivity of a scalar forecast error norm, C, to the observations.

$$\begin{aligned} \frac{\partial C}{\partial \mathbf{y}} &= \frac{\partial \mathbf{x}_a}{\partial \mathbf{y}} \frac{\partial C}{\partial \mathbf{x}_a} \\ &= \mathbf{K}^{\mathrm{T}} \frac{\partial C}{\partial \mathbf{x}_a} \end{aligned} \tag{6}$$

To calculate the sensitivities with respect to the analysis need the adjoint of the forecast model, $\mathbf{M}^{\mathrm{T}}$,

$$\frac{\partial C}{\partial \mathbf{x}_a} = \mathbf{M}^{\mathrm{T}} \frac{\partial C}{\partial \mathbf{x}_f}, \tag{7}$$

where $\partial \mathbf{x}_f$ is the forecast field at the time of validation.
Like the influence matrix this approach allows the impact of individual observations or subsets of observations to be computed simultaneously making it advantageous over the data denial experiments.
Subject to accuracy of linearised model- so can only look at the sensitivity of a short-term forecast.
A comparison of the adjoint and data denial techniques is performed by Gelaro and Zhu (2009).

Flow of entropy

Another quantity of interest is how the entropy at the forecast time depends on the uncertainty in the initial conditions (which we know from *MI* depends upon the observations).
For linear systems the evolved posterior remains Gaussian.
For non-linear systems could use an ensemble technique to evolve the posterior to the forecast time without any assumptions about the linearity of the model. However due to sampling noise this would only be feasible for a small dimensional problem.
e.g. Time lagged mutual information (e.g. Kleeman, 2011)

$$TLMI = H(\mathbf{x}(t_f)) - H(\mathbf{x}(t_f)|\mathbf{x}(t_0))$$

Summary

Many different complimentary measures of observation impact are available. In quantifying the impact of the observations it is important to consider

- the application (e.g. monitoring of observations, maximising efficiency of the assimilation, design of new instruments...)
- and subsequently which aspect of the system you are most interested in (e.g. the analysis or forecast, the spread in their errors...).

You could even get creative and design your own measure of observation impact tailored to your problem.

Discussion

What happens when the assumptions you have made are wrong?

When calculating the information content of potential new observations it is important to understand what simplifying assumptions have been made and how accurate they are.

For example neglecting the error correlations of the observations will effect the way the observations are assimilated and the conclusions that can be made about their information content (e.g. Stewart et al. 2013).

Neglecting sources of error could also have a dramatic effect on the observation impact, e.g. not allowing for model error would make the observations appear to be inconsistent with the model trajectory.

Discussion

non-Gaussian data assimilation

Recent development of advance data assimilation techniques mean that it is no longer necessary to assume Gaussian statistics. However the break down of the linear theory which methods such as variational data assimilation rely on also means that the way observation impact is calculated needs to be readdressed. Measures such as the analysis sensitivity may no longer be meaningful. Measures based on the change in entropy are a more natural choice although become more expensive to calculate as can no longer express them analytically (Fowler and van Leeuwen, 2013).

Conclusions

A measure of observation impact can be a powerful tool for maximising the efficiency of and monitoring the data assimilation process. When choosing which measure to use, thought needs to be given to both the application and aspect of the assimilation you are most interested in. Care most also be taken in interpretting the measure of observation impact given any simplifying assumptions made.

Useful references

Observation impact on the analysis

Influence matrix

- Cardinali et al., 2004. Influence-matrix diagnostic of a data assimilation system. Q. J. R. Meteorol. Soc., 130, pp 2767-2786.

Degrees of freedom for signal

- Rodgers, 2000. Inverse methods for atmospheric Sounding. World Scientific Publishing
- Eyre, 1990. The information content of data from satellite sounding systems: A simulation Study. Q. J. R. Meteorol. Soc., 116, pp401–434.

Entropy measures

- Xu, 2006. Measuring information content from observations for data assimilation: relative entropy versus Shannon entropy difference. Tellus, 59A, pp198–209.
- Majda et al., 2002. A mathematical framework for quantifying predictability through relative entropy . Meth. Appl. Analysis, 9, pp425-444.

Useful references

Observation impact on the forecast

OSSEs and data denial experiments

- Kelly et al., 2007. The value of observations. I: Data denial experiments for the Atlantic and the Pacific. Q. J. R. Meteorol. Soc., 133, pp1803-1815.
- Masutani et al., 2010. Observing system simulation experiments at the National Centers for Environmental Prediction. J. Geophys. Res.,155, D07101.
- Gelaro and Zhu, 2009. Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. Tellus, 61A:179-193.

Adjoint techniques

- Langland and Baker, 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. Tellus, 55A, pp189-201.

Entropy measures

- Kleeman, 2011. Information theory and dynamical system predictability. Entropy, 13, pp 612-649.