Measures of observation impact

Alison Fowler

Department of Meteorology, University of Reading.

NCEO Data Assimilation Course

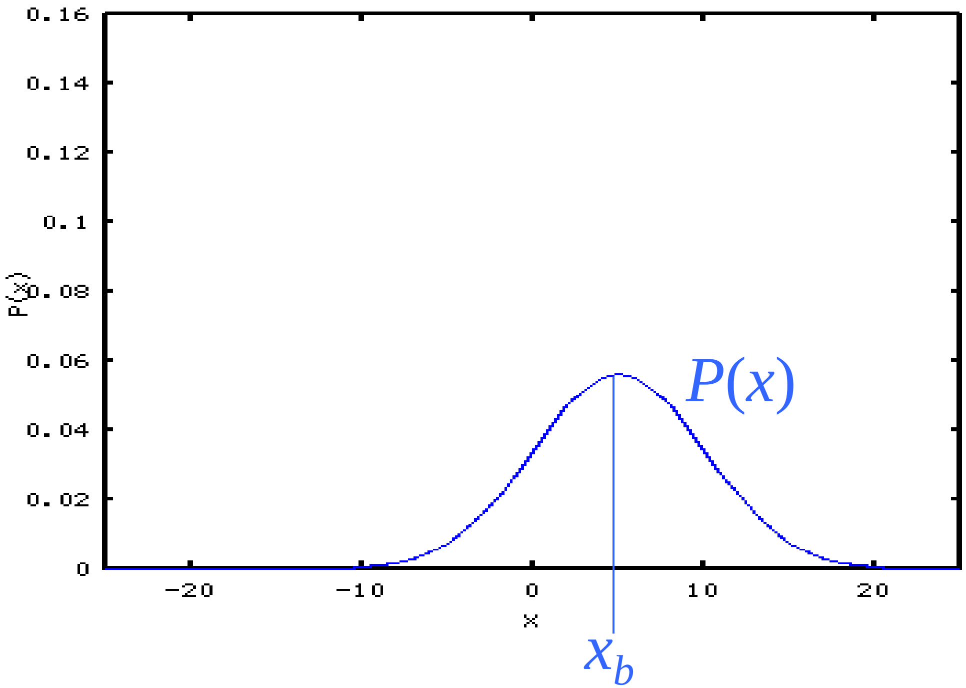
National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

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, $h(\mathbf{x})$, to give an analysis of the current state, $\mathbf{x}_a$.

A 1D Gaussian example:



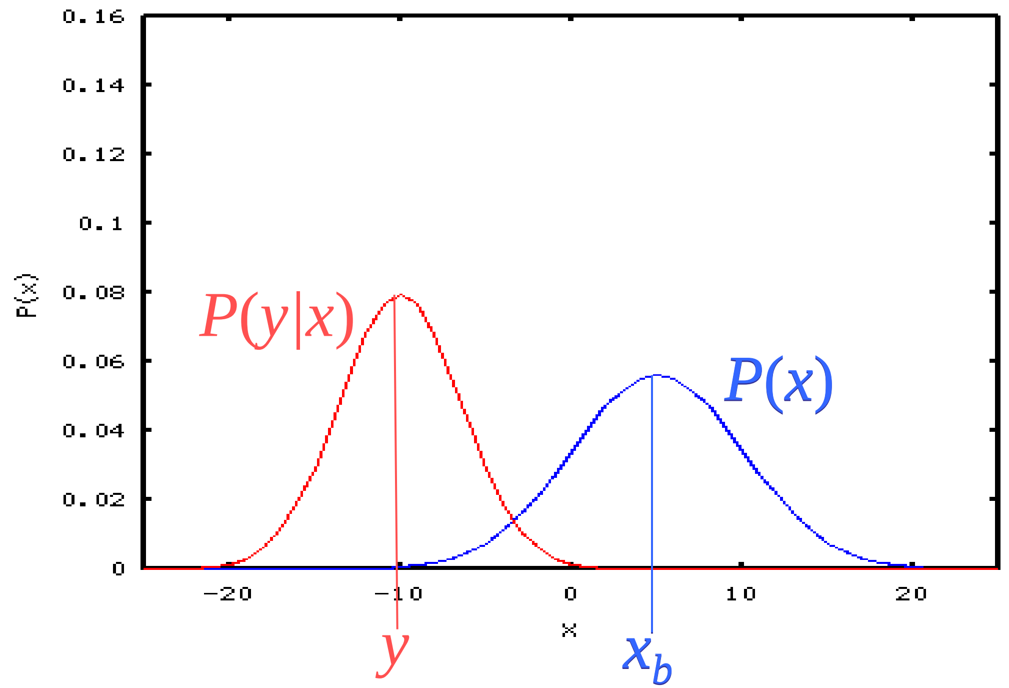
x_b:background

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, $h(\mathbf{x})$, to give an analysis of the current state, $\mathbf{x}_a$.

A 1D Gaussian example:



x_b:background, y:observation

Introduction

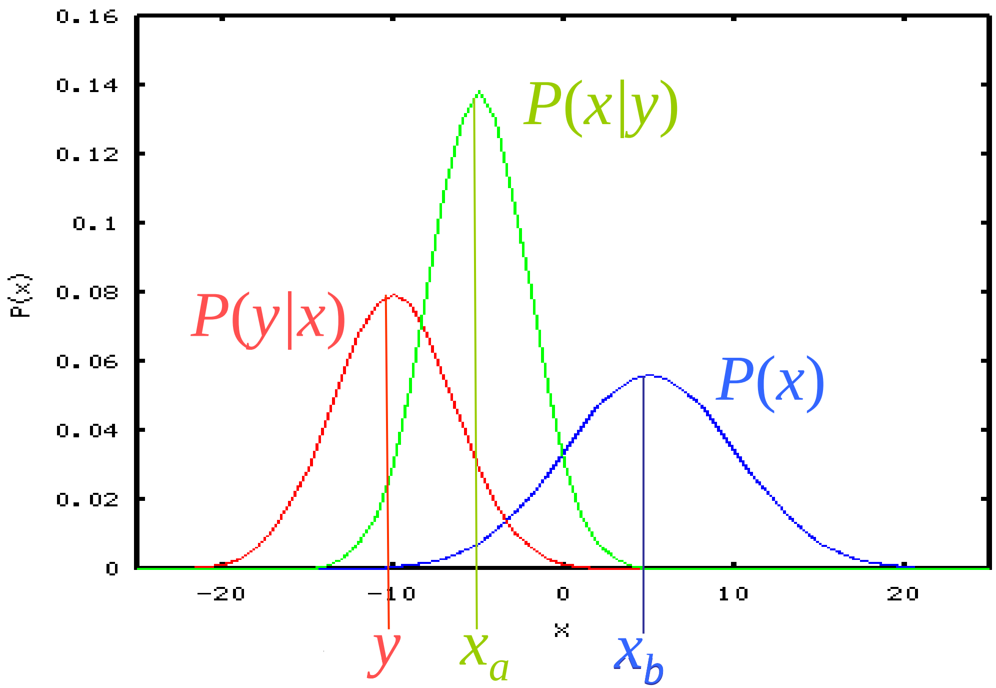
The general aim of data assimilation is to improve a forecast by reducing the error in the initial conditions.

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, $h(\mathbf{x})$, to give an analysis of the current state, $\mathbf{x}_a$.

A 1D Gaussian example:

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

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



x_a:analysis, x_b:background, y:observation, σ_b:background error st dev, σ_y:ob error st dev, **K**:Kalman gain

Introduction

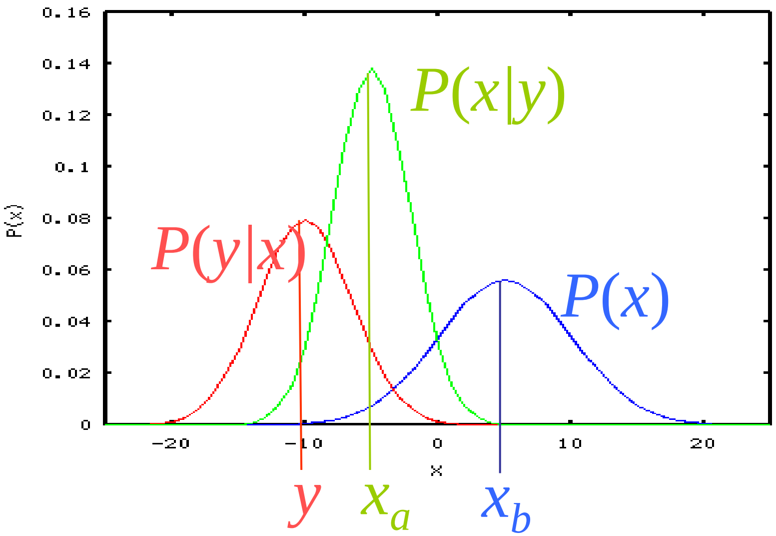
In the geosciences, observations tend to be very expensive and so it is important to monitor the impact they have in a data assimilation system.

A 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

Influence matrix

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



$$\mathbf{S} = \frac{\partial \mathbf{H}\mathbf{x}_a}{\partial \mathbf{y}} = \mathbf{K}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} \equiv \mathbf{R}^{-1}\mathbf{H}\mathbf{P}_a\mathbf{H}^{\mathrm{T}} \equiv \mathbf{R}^{-1}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1}\right)^{-1}\mathbf{H}^{\mathrm{T}}$$

This is a pxp matrix, where p is the number of observations.

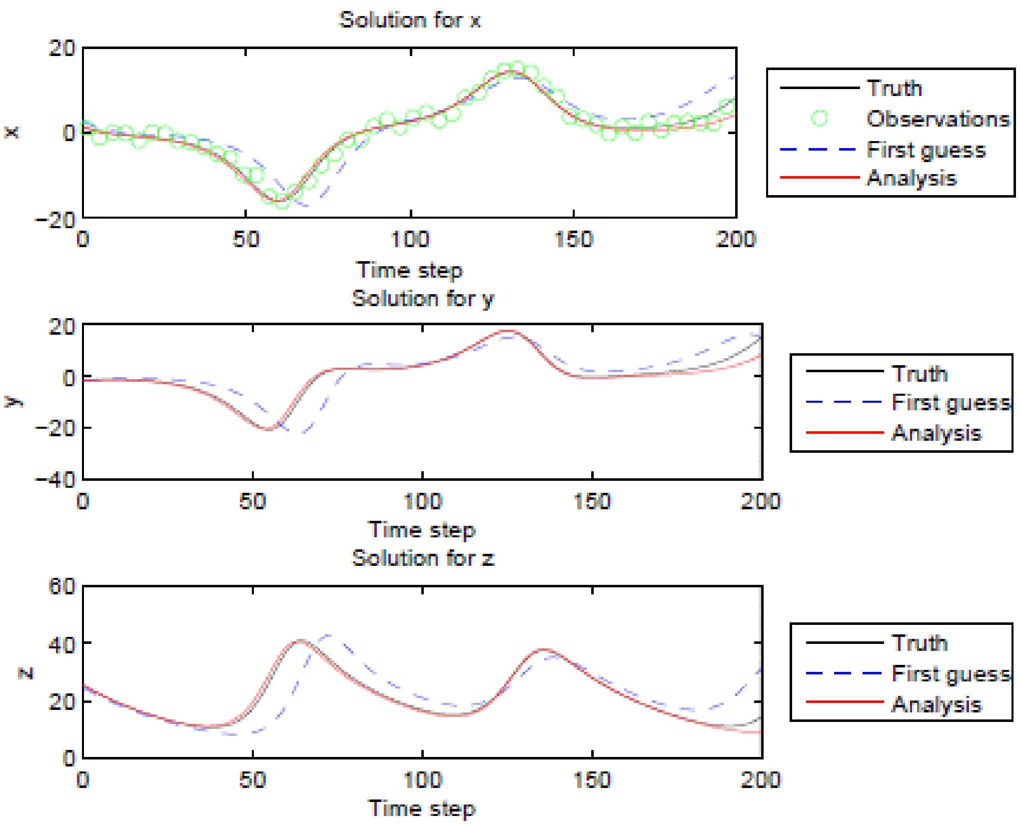
When $\mathbf{R}$ is diagonal the diagonal elements of $\mathbf{S}$ are bounded by 0 and 1.

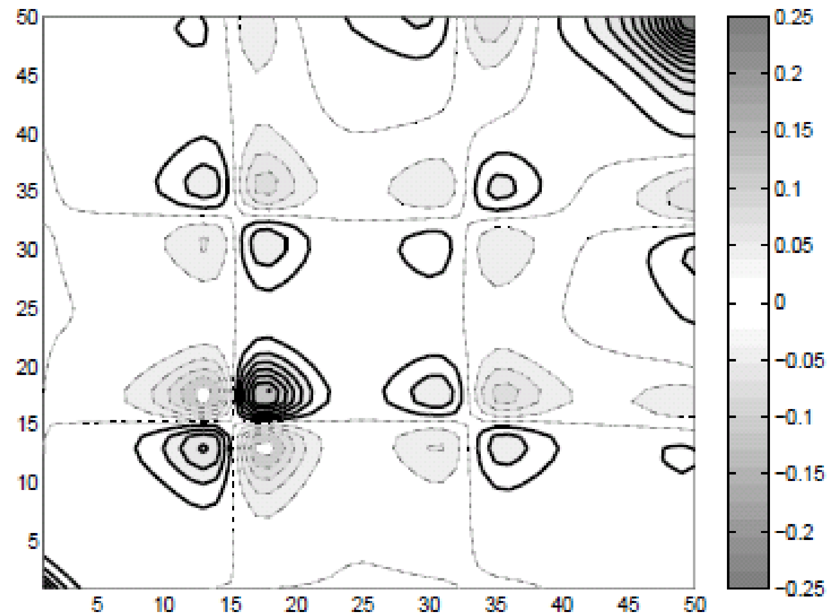
Allows the most influential observations or group of observations to be identified.

H:linearised ob operator, R:ob error covariances, B:background error covariances, P_a:analysis error covariances

Influence matrix- Lorenz '63 example

- $\mathbf{x}=(x,y,z)^T$ and $\mathbf{y}$ consists of observations of x made at every 4th time step.
- $\mathbf{B}= \sigma_b^2 \begin{pmatrix} 1 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 1 \end{pmatrix}$, where σ_b^2 and σ_y^2 (the observation error variance, identical at each time) are both equal to 1.

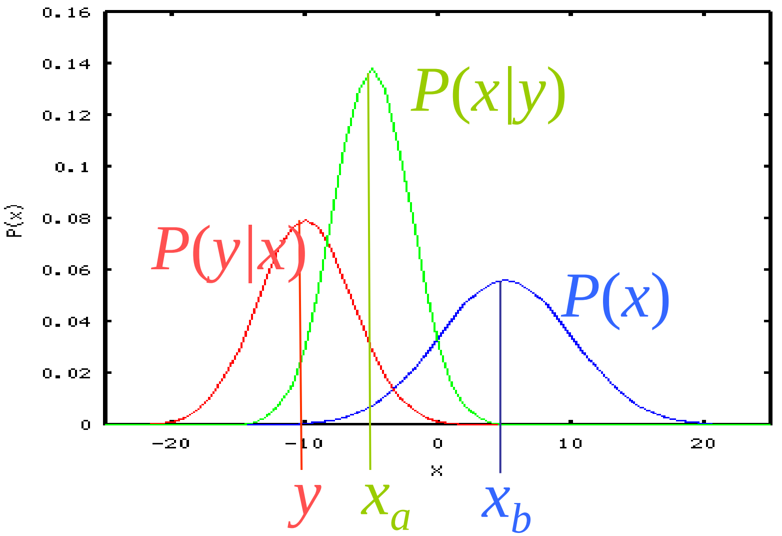




S: solid (dashed) contours give positive (negative) values

Degrees of Freedom for signal

The degrees of freedom for signal quantifies the amount of information coming from the observations as opposed to the background (e.g. Rodgers, 2000).



It can be quantified as

$$d_s = E\left\{ (\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_b) \right\} = trace(\mathbf{HK}) = trace(\mathbf{S})$$

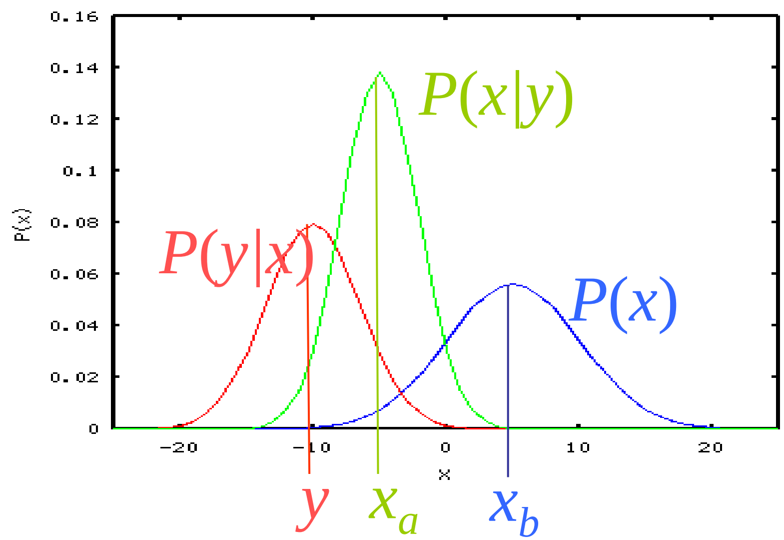
d_s is bounded by 0 and p.

d_s can be rewritten in terms of the eigenvalues of $\mathbf{S}$

$$d_s = \sum_{i=1}^{p} \lambda_i$$

Mutual Information

Mutual information quantifies the amount of information introduced by the observations. It is defined as the change in entropy between the prior and the posterior (e.g. Rodgers, 2000; Eyre, 1990, QJRMetSoc).



Entropy is a measure of the uncertainty and so for a Gaussian error distribution is simply related to error variance.

$$Entropy = H(\chi) = \int p(\chi) \ln p(\chi) d\chi = n \ln(2\pi e)^{1/2} + \frac{1}{2} \ln |\mathbf{C}_\chi|$$

The mutual information is therefore

$$MI = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{y}) = \frac{1}{2} \ln |\mathbf{B}\mathbf{P}_a^{-1}| = -\frac{1}{2} \ln \prod_{i=1}^{p} (1 - \lambda_i)$$

n:size of state space

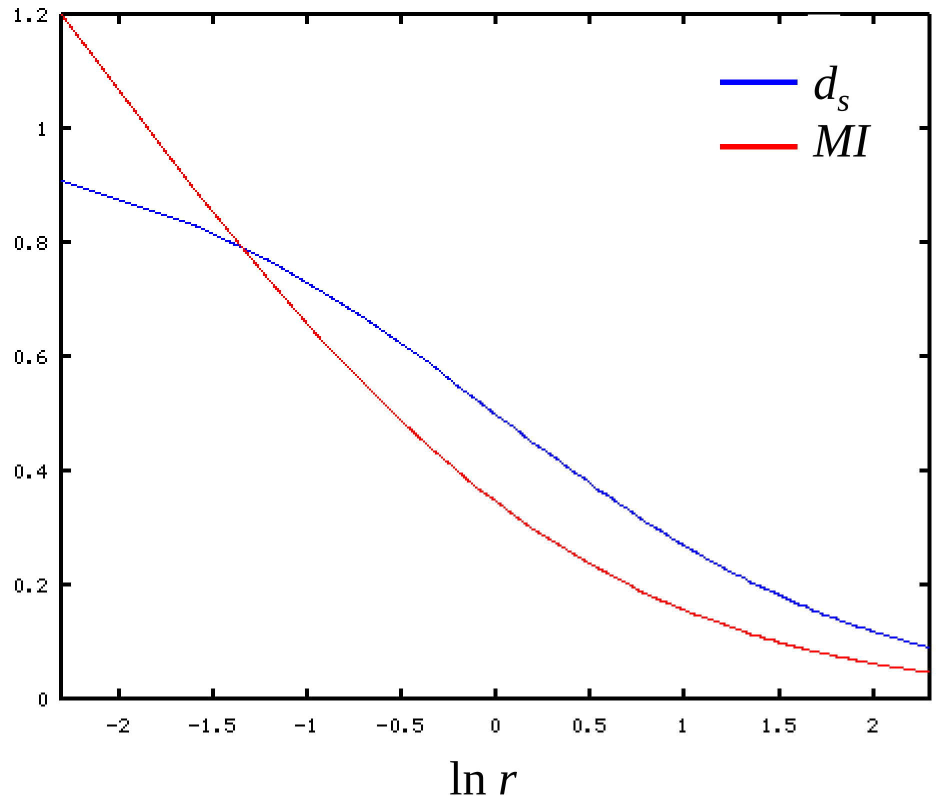
Comparison of d_s and MI - 1D example

Both d_s and MI can be expressed as the ration of the observation error standard deviation to the background error standard deviation r.

$$r = \frac{\sigma_y^2}{\sigma_b^2}$$

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

$$MI = \frac{1}{2}\ln\left(\frac{1+r}{r}\right)$$

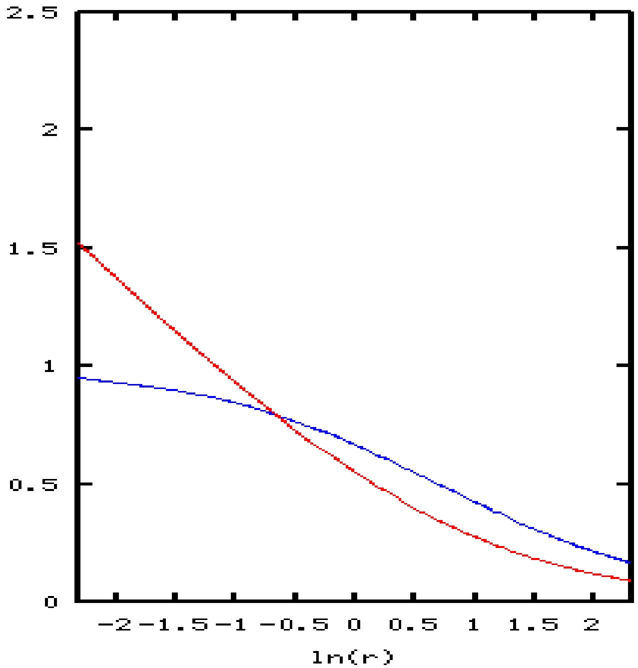


Comparison of d_s and MI - 2D example

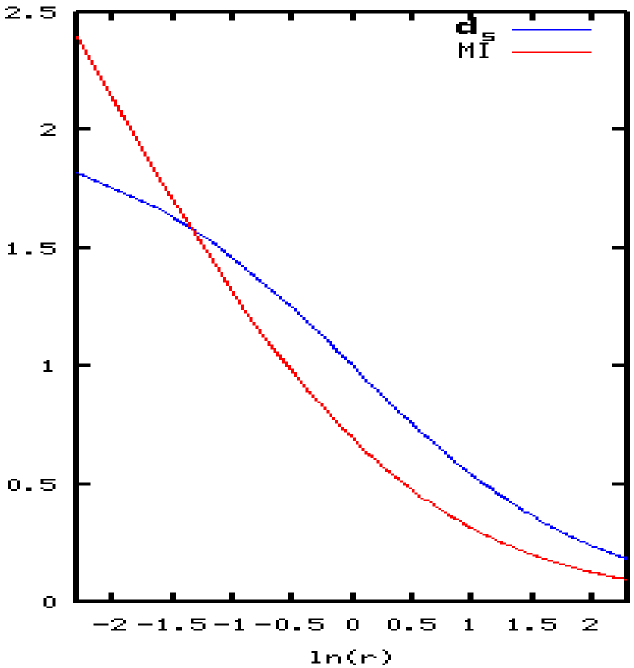
Let $\mathbf{B} = \sigma_b^2 \begin{pmatrix} 1 & \alpha \\ \alpha & 1 \end{pmatrix}$ and $\mathbf{R} = \sigma_y^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. $\mathbf{H}=\mathbf{I}_2$

Then $d_s = \dfrac{2(1+r-\alpha^2)}{(1+r)^2-\alpha^2}$ and $MI = \dfrac{1}{2}\ln\left(\dfrac{(1+r)^2-\alpha^2}{r^2}\right)$

Obs close together (α=1)



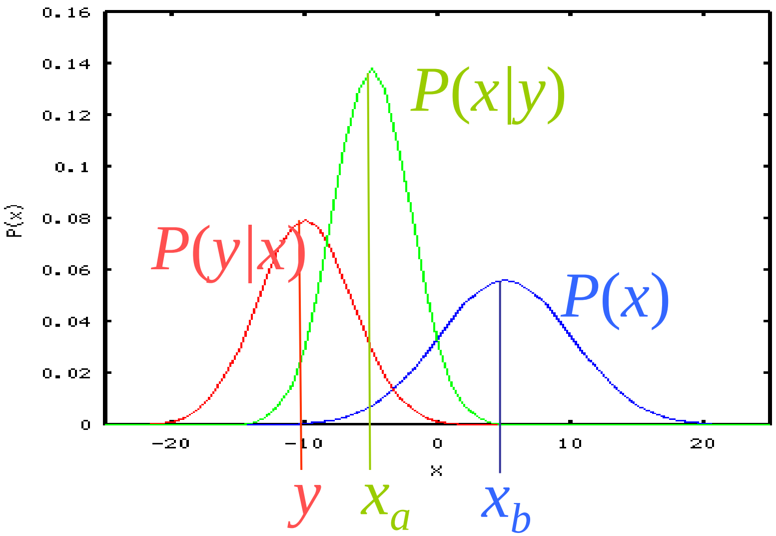
Obs far apart (α=0)



Relative Entropy

Relative entropy is a non-symmetric measure of the difference between the prior and posterior (Xu, 2006, Tellus).

$$RE = \int p(x\,|\,y) \ln \frac{p(x\,|\,y)}{p(x)} dx$$



Unlike the other measures relative entropy is sensitive not only to the change in the covariances but also the change in the mean.

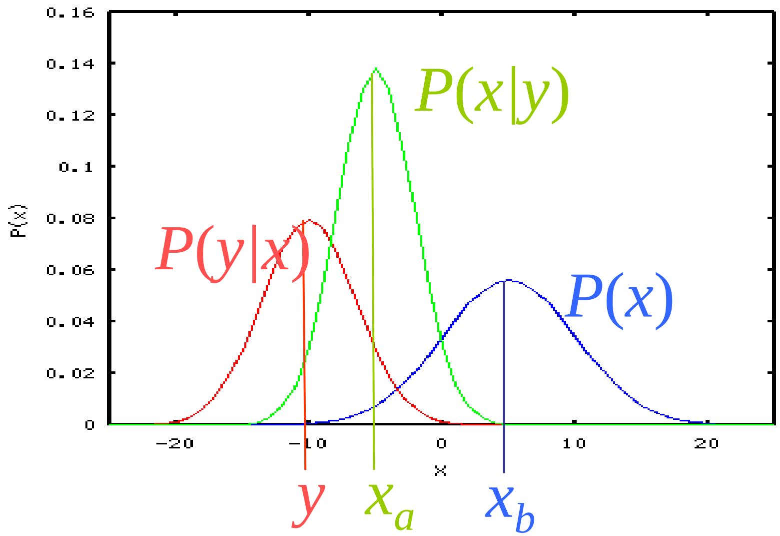
$$RE = \frac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_b) + \frac{1}{2}\left(\ln |\mathbf{B}\mathbf{P}_a^{-1}| + trace(\mathbf{B}^{-1}\mathbf{P}_a) - n \right)$$

n:size of state space

Relative Entropy

Relative entropy is a non-symmetric measure of the difference between the prior and posterior (Xu, 2006, Tellus).

$$RE = \int p(x\,|\,y) \ln \frac{p(x\,|\,y)}{p(x)} dx$$



Unlike the other measures relative entropy is sensitive not only to the change in the covariances but also the change in the mean.

$$RE = \frac{1}{2}(\mathbf{x}_a - \mathbf{x}_b)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_a - \mathbf{x}_b) + \frac{1}{2}\Big(\underbrace{\ln|\mathbf{B}\mathbf{P}_a^{-1}|}_{MI} + \underbrace{trace(\mathbf{B}^{-1}\mathbf{P}_a) - n}_{d_s}\Big)$$

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

n:size of state space

Comparison of the four measures

- The influence matrix gives a full pxp matrix whereas d_s and MI give a single value summary of this matrix.

- Only RE is sensitive to the value of the observation in addition to its error variance relative to that of the background.

- It is possible to show that MI, unlike d_s, is additive with successive observations.

- RE averaged over all possible observations gives MI.

- MI is conserved for a linear change of coordinates.

- RE is conserved under a general non-linear change of coordinates.

- A study of how RE, MI and d_s depend upon model error is given in Xu et al., 2009, Tellus.

Observation impact on the forecast

A measure of the observations impact on the analysis implicitly gives information about their impact on the forecast (Liu et al., 2009, QJRMetSoc).

However if you wish to compare different observations then it is important to take in to account their different dynamical roles.

I shall briefly introduce the following methods:

- OSSEs and data denial experiments
- Adjoint techniques
- Flow of entropy

OSSEs and data denial experiments

OSSEs (observation system simulation experiments) (e.g. Arnold and Dey (1986), Atlas (1997), Masutani et al. (2010)) and data denial experiments (e.g. Bouttier and Kelly (2001), Kelly et al. (2007)) compare a control forecast to a forecast which has had additional (simulated) observations assimilated 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)).

It is also 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 to the observations.

The error norm is given by $||\mathrm{x}_f - \mathrm{x}_t||_{\mathbf{C}}$, where x_f is the forecast resulting from the assimilation of observations and x_t is a verifying analysis (which must be chosen carefully). $\mathbf{C}$ is a matrix of energy weighting coefficients that represents dry total energy (Rabier et al. (1996)).

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) and Cardinali (2009) .

Flow of entropy (the evolution of uncertainty)

Calculate 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 a particle filter to evolve the posterior to the forecast time without any assumptions about the linearity of the model – only possible for small dimensions!

e.g. Time lagged mutual information (e.g. Kleeman, 2011, Entropy)

$$TLMI = H(x(t_f)) - H(x(t_f) \mid x(t_i))$$

t_f: forecast time, t_i: inital time

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.

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.

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.