Atmospheric Data Quality Control and Its Impact on Data Assimilation

JCSDA Summer Colloquium on Data Assimilation

Andrew Lorenc, Stevenson, Washington. July 7 – 17, 2009

Content

1. Bayes' Theorem for discrete events: "There is a gross error"
    - Operational Met Office Bayesian QC.
2. Non-Gaussian PDFs and non-quadratic penalty functions
3. Importance of prior statistics
4. Interaction of QC between observations
    - Non-Gaussian DA
    - L1 & Huber norms
5. What answer do we want?
6. Observation Monitoring
7. (Designing linear PF model near nonlinearities)

Bayes Theorem – revision from lecture 2

Gaussian PDFs

Non-Gaussian observational errors - Quality Control

Thanks to Jim Purser who introduced me to the Bayesian approach.

Bayes' Theorem for Discrete Events

$A \quad B$	events
$P(A)$	probability of A occurring, or knowledge about A's past occurrence
$P(A \cap B)$	probability that A and B both occur,
$P(A \mid B)$	conditional probability of A given B

We have two ways of expressing $P(A \cap B)$:

$$P(A \cap B) = P(B)\,P(A \mid B) = P(A)\,P(B \mid A)$$

$\Rightarrow$ Bayes' Theorem: $$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)}$$

Can calculate $P(B)$ from: $$P(B) = P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A})$$

Quality Control example: Bayesian Dice

Discrete Bayes Theorem Applied to Gross Observational Errors

I have two dice. One is weighted towards throwing sixes. I have performed some experiments with them, and have the prior statistics that:

for the weighted (W) die, $P(6|W) = 58/60$

for the good (G) die, $P(6|G) = 10/60$

I choose one at random: $P(W) = P(G) = ½ = 50\%$

I throw this die, and it shows a six. Now:-

$$P(6) = P(6|W)\ P(W) + P(6|G)\ P(G)$$
$$= 58/60 \quad 1/2 \quad + \quad 10/60 \quad 1/2$$

We can now apply Bayes' Theorem:

$$P(G|6) = P(6|G)\ P(G)\ /\ P(6)$$
$$= 10/60 \quad 1/2 \quad / \ 34/60 \ = \ 5/34 \ = 15\%$$
$$P(W|6) = P(6|W)\ P(W)\ /\ P(6)$$
$$= 58/60 \quad 1/2 \quad / \ 34/60 = 29/34 = 85\%$$

Simple model for PDF of observations with errors

Assume that a small fraction of the observations are corrupted, and hence worthless. The others have Gaussian errors.

For each observation we have:

$$p\left(y^o|x\right) = p\left(y^o|G \cap x\right)P(G) + p\left(y^o|\overline{G} \cap x\right)P\left(\overline{G}\right)$$

G is the event "there is a gross error" and $\overline{G}$ means *not* G.

$$p\left(y^o|\overline{G} \cap x\right) = N\left(y^o|H(x), \mathrm{E}+\mathrm{F}\right)$$

$$p\left(y^o|G \cap x\right) = \begin{cases} k & \text{over the range of plausible values} \\ 0 & \text{elsewhere} \end{cases}$$

Applying this model

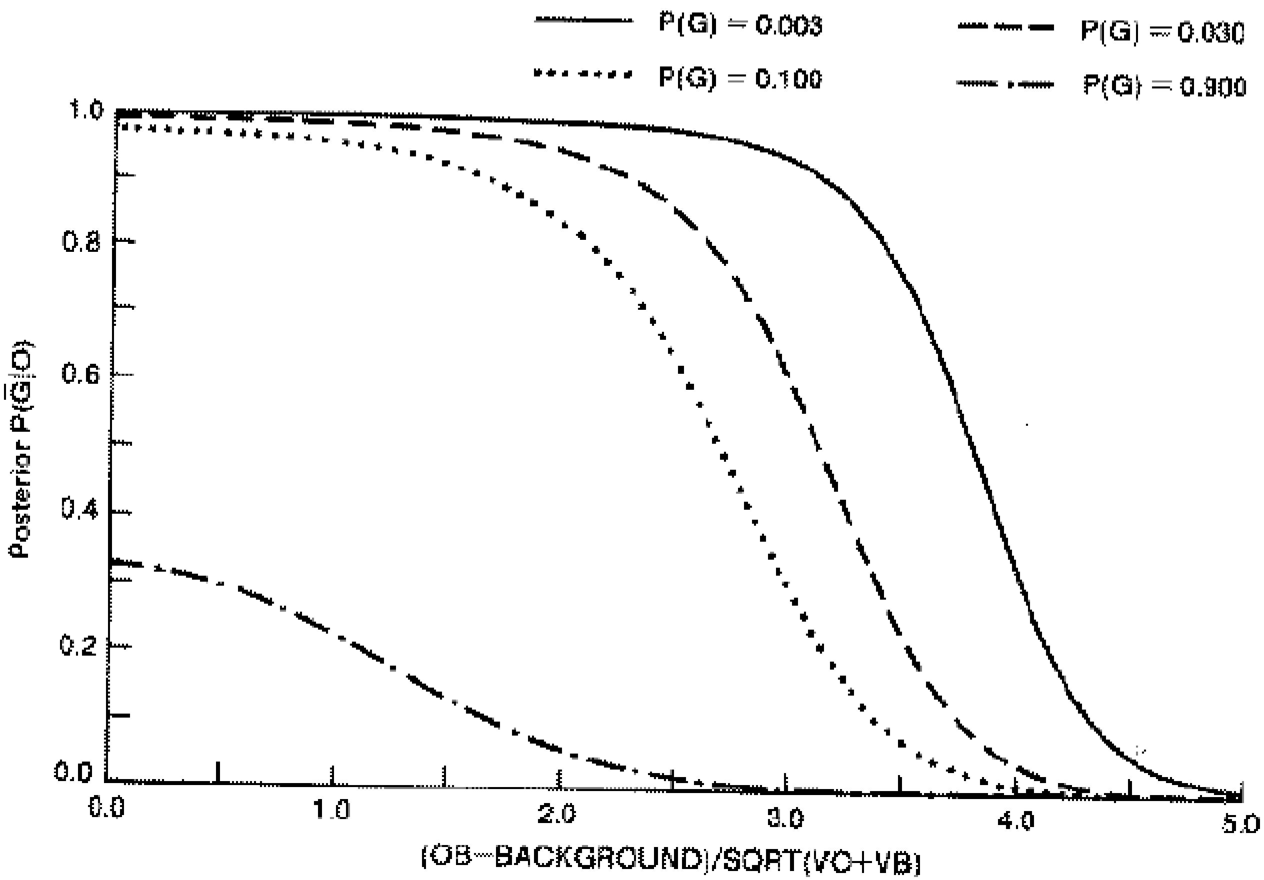
- Can simply apply Bayes Theorem to the discrete event G

$$P(\mathrm{G} \mid y^o) = \frac{P(y^o \mid \mathrm{G}) P(\mathrm{G})}{P(y^o)}$$

$$= \frac{k\, P(\mathrm{G})}{k\, P(\mathrm{G}) + N\left(y^o \middle| H\left(x^b\right), \mathbf{R} + \mathbf{HBH}^T\right) P\left(\overline{\mathrm{G}}\right)}$$

Lorenc, A.C. and Hammon, O., 1988: "Objective quality control of observations using Bayesian methods. Theory, and a practical implementation." *Quart. J. Roy. Met. Soc.*, **114**, 515-543

Posterior probability that an observation is "correct", as a function of its deviation from the background forecast



Posterior probability of an observation not having a gross error, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Lorenc and Hammon, 1988

Met Office Bayesian DA

1. Prior experience sets P(G).
2. Background check (just described) modifies P(G).
3. Allows for independent gross error affecting whole report as well as in each reported element:

$$P(\overline{G_p} \cap \overline{G_s} | O_s \cap O_u \cap O_T) = \frac{P(O_s \cap O_u \cap O_T | \overline{G_p} \cap \overline{G_s}) P(\overline{G_p} \cap \overline{G_s})}{P(O_s \cap O_u \cap O_T)}$$

4. Sequential comparison of “buddies” modifies P(G):

$$P(G_1 | O_1 \cap O_2) = P(G_1 | O_1) \frac{P(O_1) P(O_2)}{P(O_1 \cap O_2)}$$

5. Finally, reject if P(G)>0.5

Lorenc & Hammon, 1988

Applying this model

- Can simply apply Bayes Theorem to the discrete event G

$$P\left(G|y^o\right) = \frac{p\left(y^o|G\right)P(G)}{p\left(y^o\right)}$$

$$= \frac{k\,P(G)}{k\,P(G) + N\left(y^o|H\left(x^b\right), R+HBH^T\right)P\left(\overline{G}\right)}$$

Lorenc, A.C. and Hammon, O., 1988: "Objective quality control of observations using Bayesian methods. Theory, and a practical implementation." *Quart. J. Roy. Met. Soc.*, **114**, 515-543

- Or we can use the non-Gaussian PDF directly

$$p\left(y^o|x\right) = p\left(y^o|G \cap x\right)P(G) + p\left(y^o|\overline{G} \cap x\right)P\left(\overline{G}\right)$$

Ingleby, N.B., and Lorenc, A.C. 1993: "Bayesian quality control using multivariate normal distributions". *Quart. J. Roy. Met. Soc.,* **119**, 1195-1225
Andersson, Erik and Jarvinen, Heikki. 1999: "Variational Quality Control" *Quart. J. Roy. Met. Soc*., **125**, 697-722

Bayes theorem in continuous form,

to estimate a value x given an observation y^o

$$p(x \mid y^o) = \frac{p(y^o \mid x)p(x)}{p(y^o)}$$

$p(x \mid y^o)$ is the posterior distribution,

$p(x)$ is the prior distribution,

$p(y^o \mid x)$ is the likelihood function for x

Can get $p(y^o)$ by integrating over all x: $p(y^o) = \int p(y^o \mid x)p(x)dx$

Assume Gaussian pdfs

Prior is Gaussian with mean x^b, variance V_b : $x \sim N(x^b, V_b)$

$$p(x) = \left(2\pi V_b\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x - x^b)^2}{V_b}\right)$$

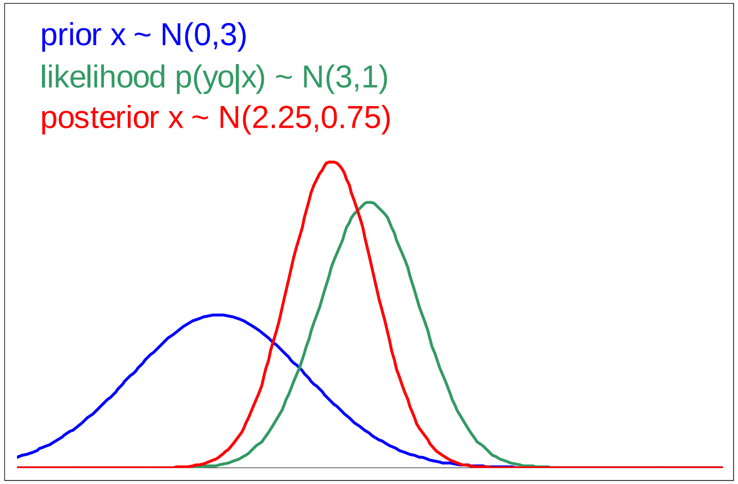
Ob y^o, Gaussian about true value x variance V_o : $y^o \sim N(x, V_o)$

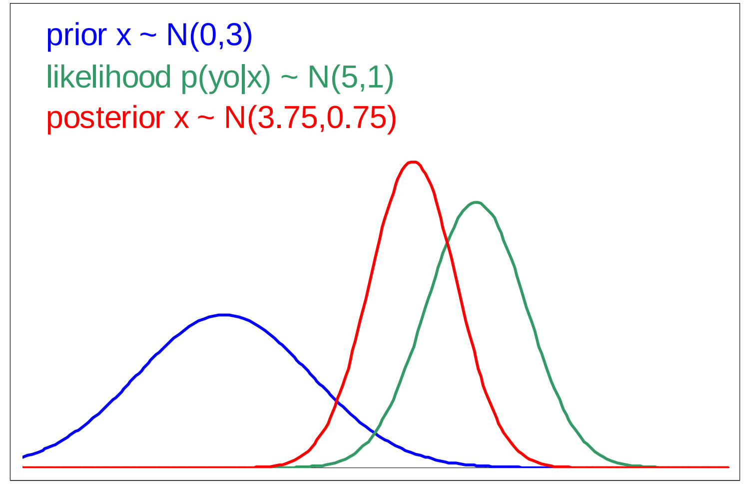
$$p(y^o | x) = \left(2\pi V_o\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(y^o - x)^2}{V_o}\right)$$

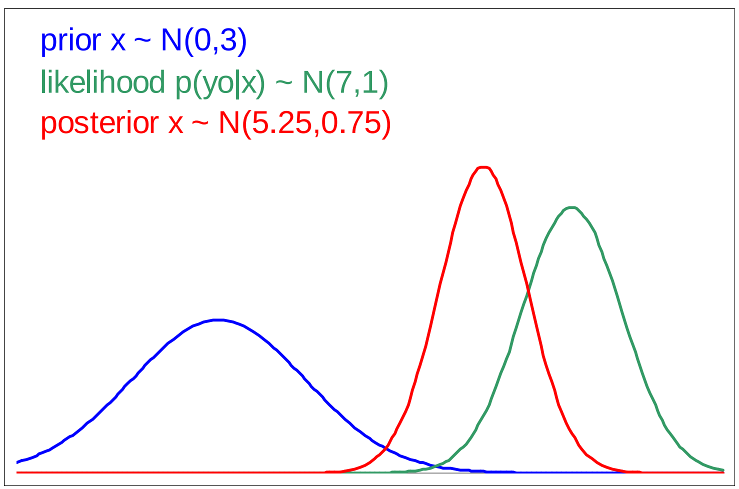
Substituting gives a Gaussian posterior: $x \sim N\left(x^a, V_a\right)$

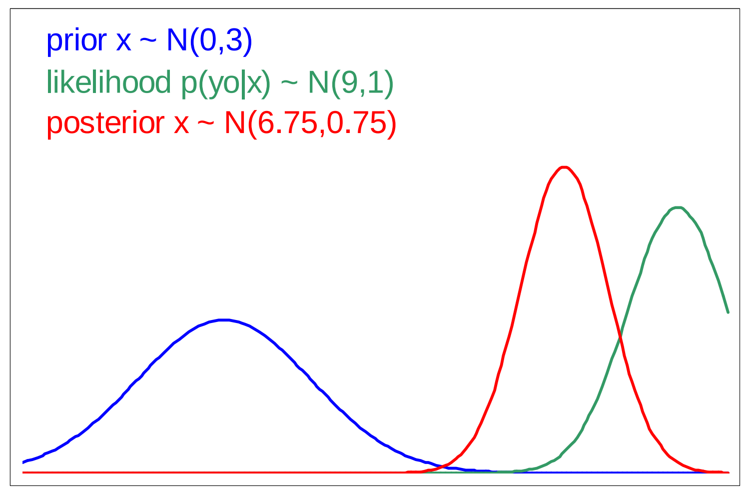
$$p(x) = \left(2\pi V_a\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x - x^a)^2}{V_a}\right)$$

Combination of Gaussian prior & observation
- Gaussian posterior,
- weights independent of values.

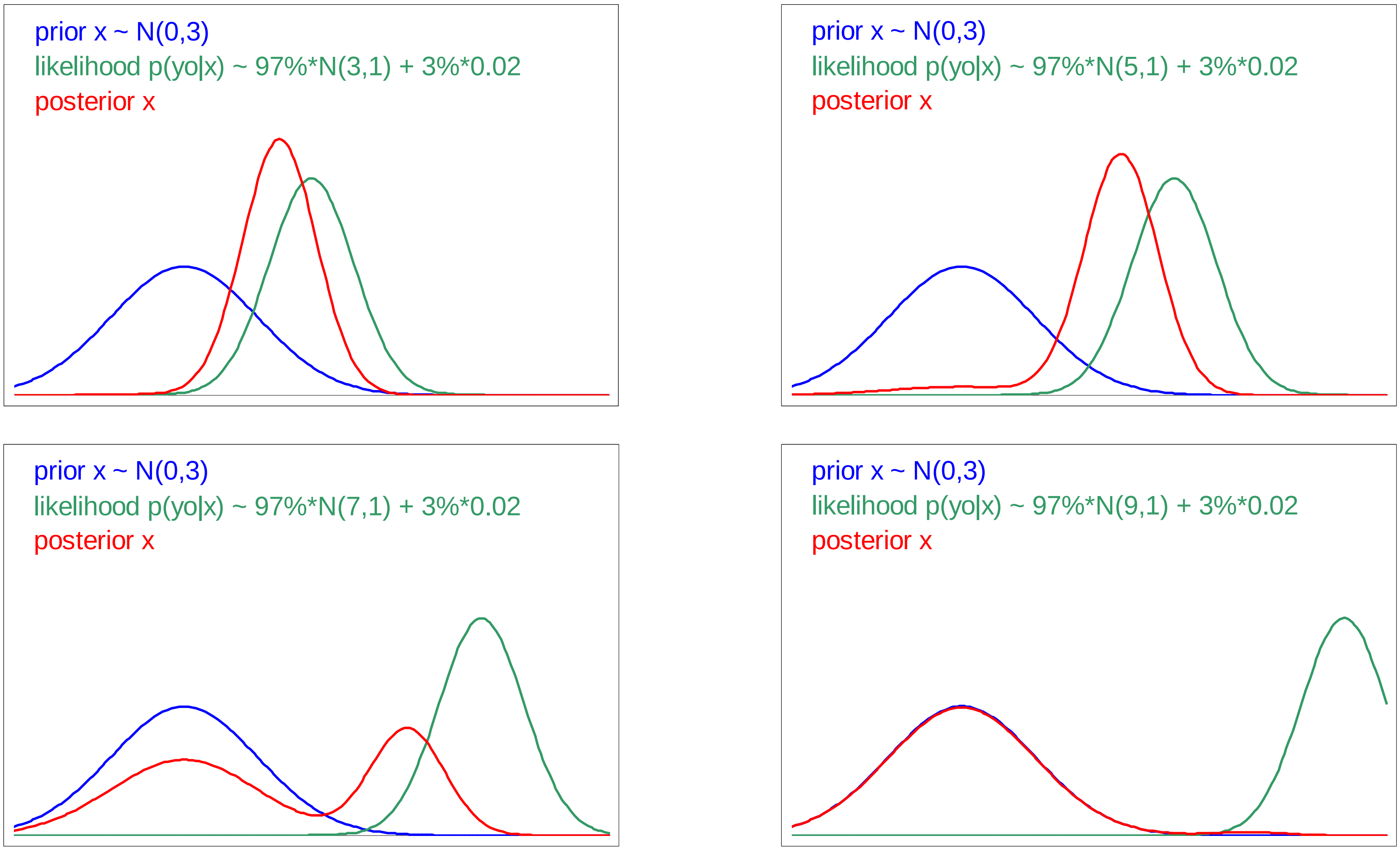








Gaussian prior combined with observation with gross errors - extreme obs are rejected.

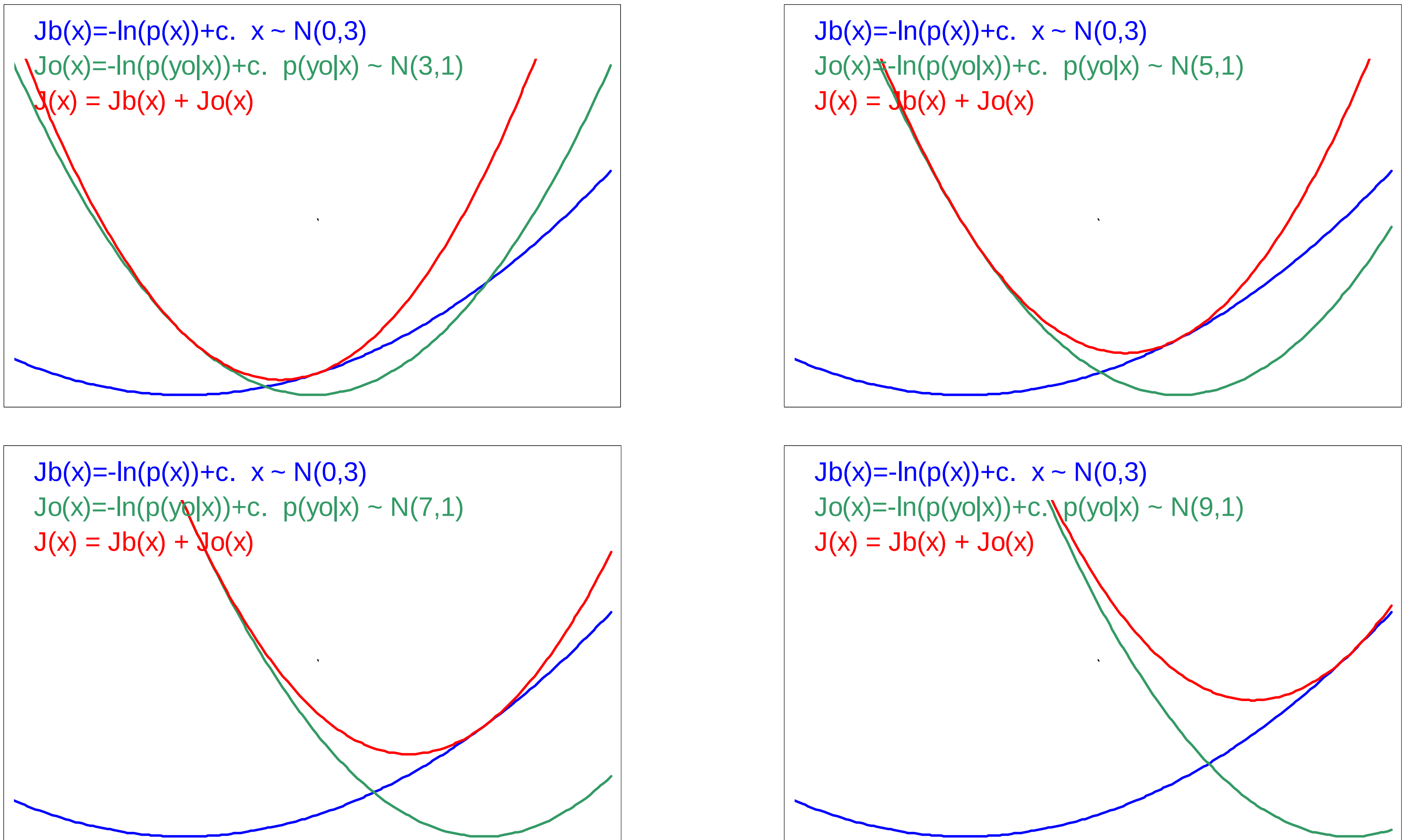


Js

Variational Penalty Functions

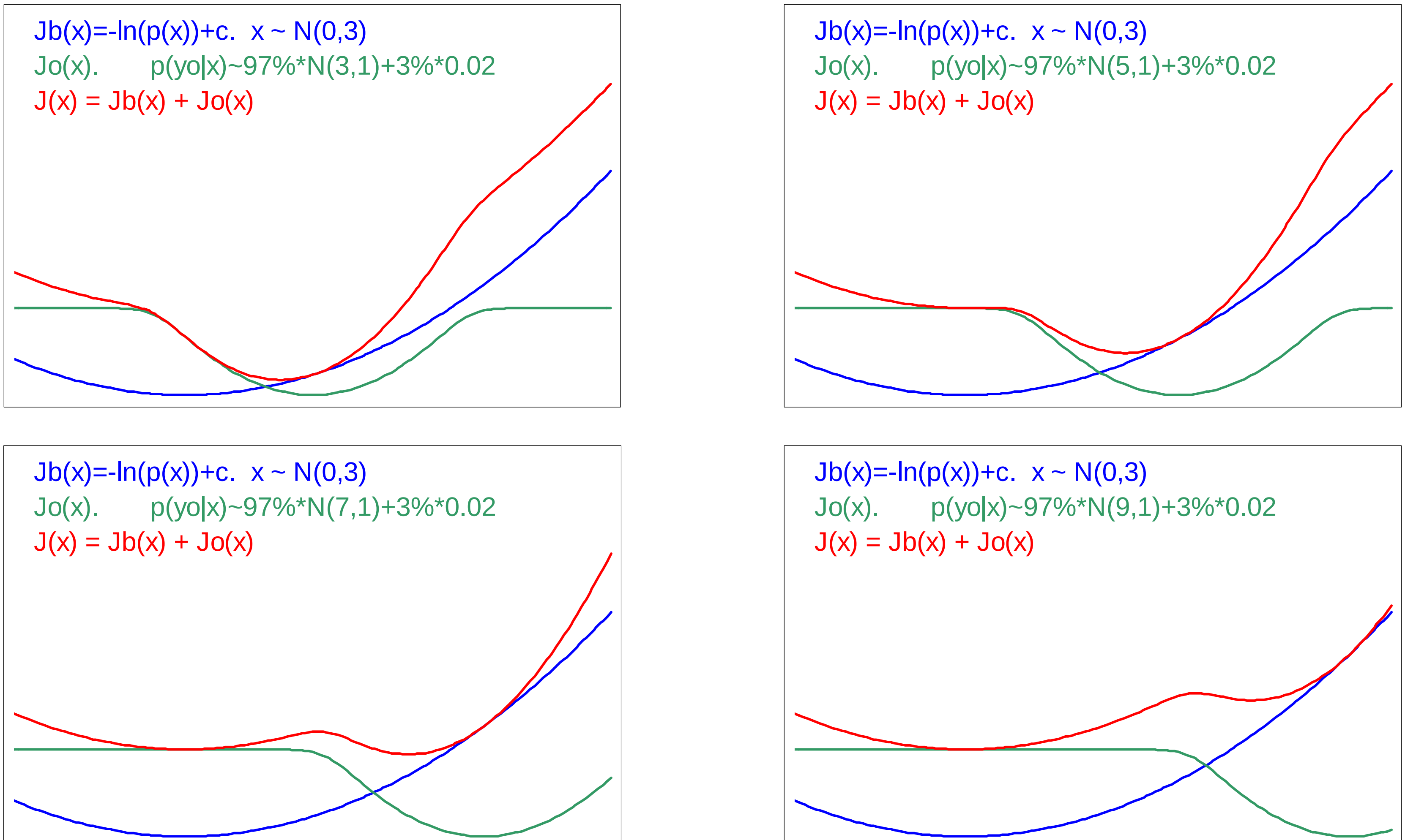
- Finding the most probable posterior value involves maximising a product [of Gaussians]
- By taking $-ln$ of the posterior PDF, we can instead minimise a sum [of quadratics]
- This is often called the "Penalty Function" J
- Additive constants can be ignored

Penalty functions: $J(x) = -\ln(p(x))+c$
p Gaussian $\Rightarrow$ J quadratic



PDFs

Penalty functions: $J(x)$ = -ln($p(x)$)+c
p non-Gaussian ⇒ J non-quadratic



PDFs

Importance of prior statistics on observation errors

- The Bayesian approach does more than assess the probability of the observed value being correct; it needs also prior estimates of the characteristics of erroneous observations.
- To avoid more accurate (e.g. weather ship) observations being more likely to be rejected, we need to allow for them being more reliable too.
- Testing for different sources of error can lead to logically complex, often built on speculative foundations – these are usually only worthwhile if extra information is available (e.g. radiosonde internal consistency, cloud affected radiances, track check for ships & aircraft, …).

Importance of prior statistics on background

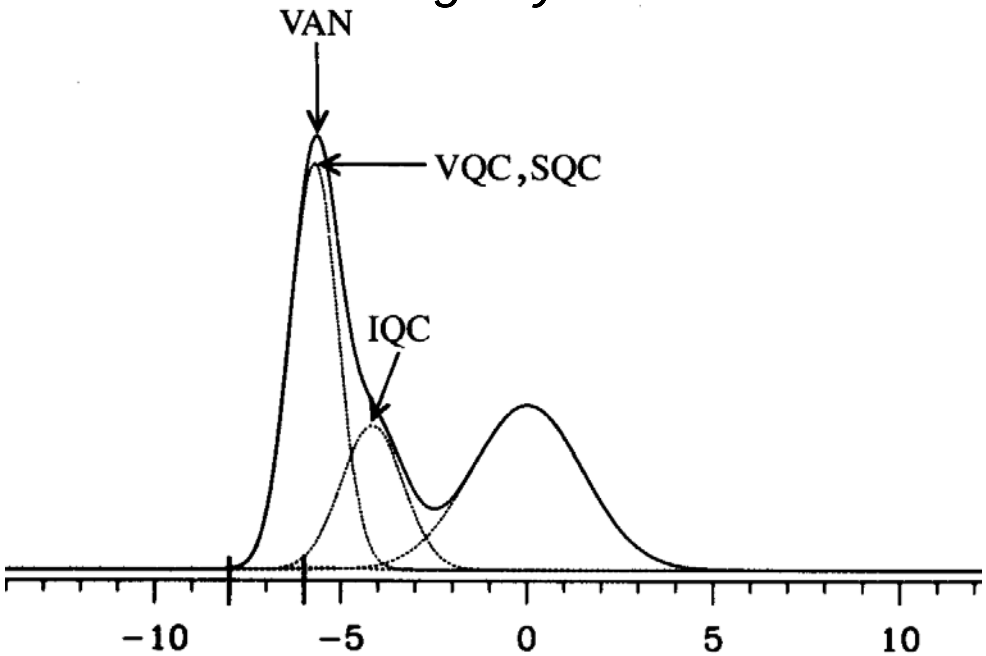
- Usually the most important information source for checking observations is the background forecast. Assuming this has Gaussian errors gives a rather sharp rejection criterion.
- Actual background error variances are flow dependent.
  If we do not allow for this, good observations can be erroneously rejected in high-impact weather events.
- Can estimate flow dependent background error variances
  - from simple pattern recognition & regression (Parrett 1992);
  - from EnKF;
  - from misfit of all observations in an area (Dee *et al.* 2001).

QC of “buddy” observations interacts.

- Supportive “buddies” may retain correct observations which would be rejected individually.
- But an observation with a gross error can throw suspicion on nearby good observations.
- It is perhaps better to calculate posterior probability of every combination of reject/accept (Ingleby & Lorenc 1993). But this is impracticable for more than a few observations.
- Preliminary checks, plus re-doing QC based on preliminary decisions, can work (Lorenc 1981, Lorenc & Hammon 1988).
- It is attractive to combine this with other DA iterations, especially in variational methods.

Complexity of ideal buddy check for only 2 observations. *Ingleby & Lorenc 1993.*

	G_2	$\overline{G}_2$	$G_2 \cup \overline{G}_2$
G_1	.393	.191	.584
$\overline{G}_1$	.003	.413	.416
$G_1 \cup \overline{G}_1$	.396	.604	1.0



- *N* observation give 2^N combinations, shown for {-6, -8}, with P(G)=0.04, as thin lines on the figure and entries in the table.
- Accepting both observations is best Simultaneous QC result, ***but sequential one-by-one re-checking cannot find this.***
- Individual QC accepts #1 and rejects #2 – not a likely combination.
- Most probable value, goal of Variational QC, is near SQC, ***but simple descent algorithm cannot find this.***

Non-Gaussian DA.

The need for QC of observations depends on the use being made of them.

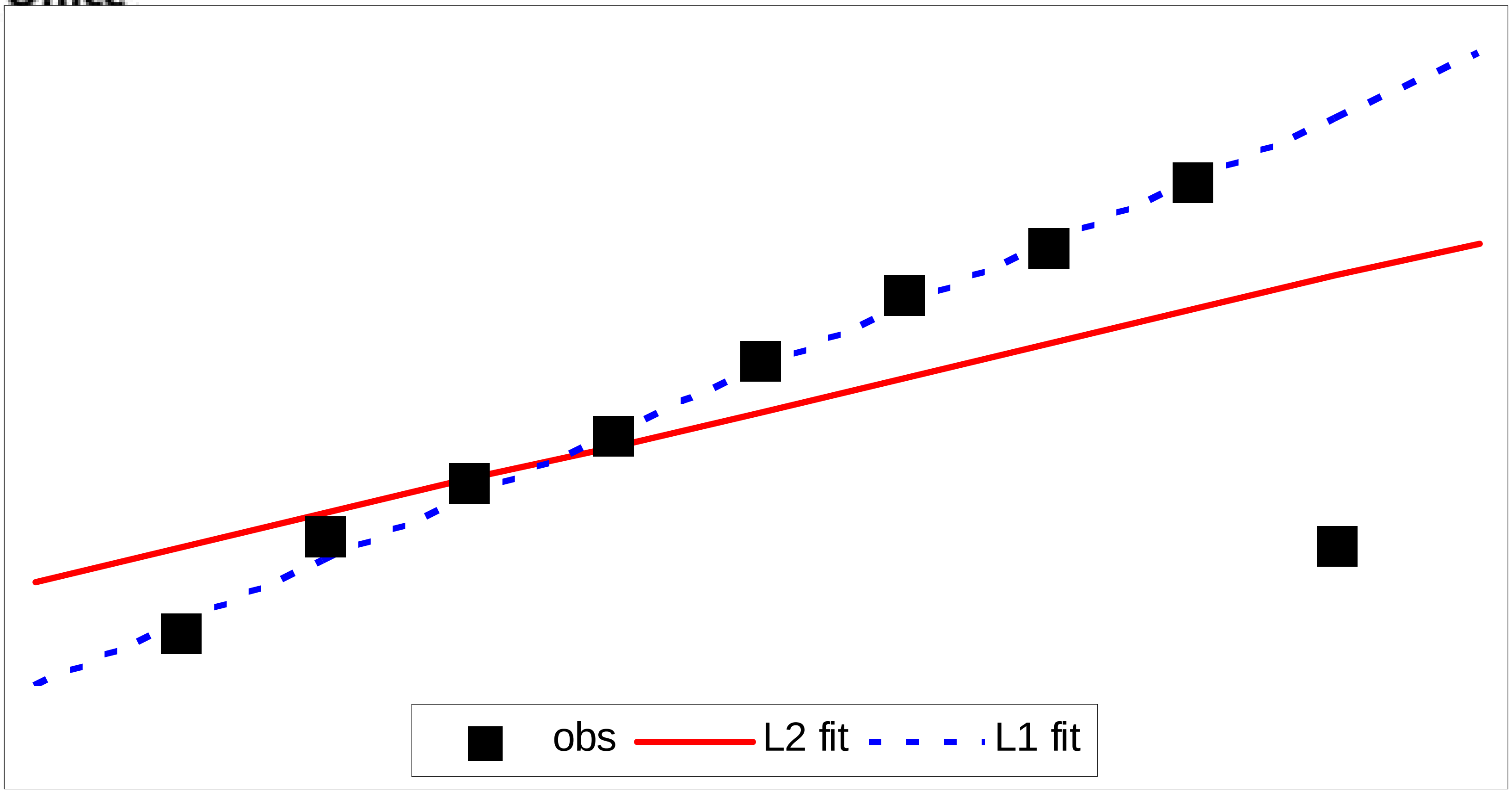
- All observations can be plotted for human analysis – an experienced forecaster will allow for errors while drawing up weather chart.
- In contrast, “Objective” verification statistics are very sensitive to QC of verifying observations.
- Gaussian-based (least-squares) analysis methods have fairly clear requirements to reject observations which do not fit the Gaussian hypothesis.
- Non-Gaussian methods can allow directly for error properties, making a separate QC step unnecessary.

Accounting for a range of error characteristics – not just the possibility of gross error.

- Situation dependent observational errors.
- Situation dependent bias.
- Observations a nonlinear function of model variables.
- Surface wind speed.
- Scatterometer wind aliases.
- Cloud and precipitation.

Variational methods can be developed to address all of these simultaneously as part of the core variational algorithm. Otherwise, we develop specific stand-alone algorithms.

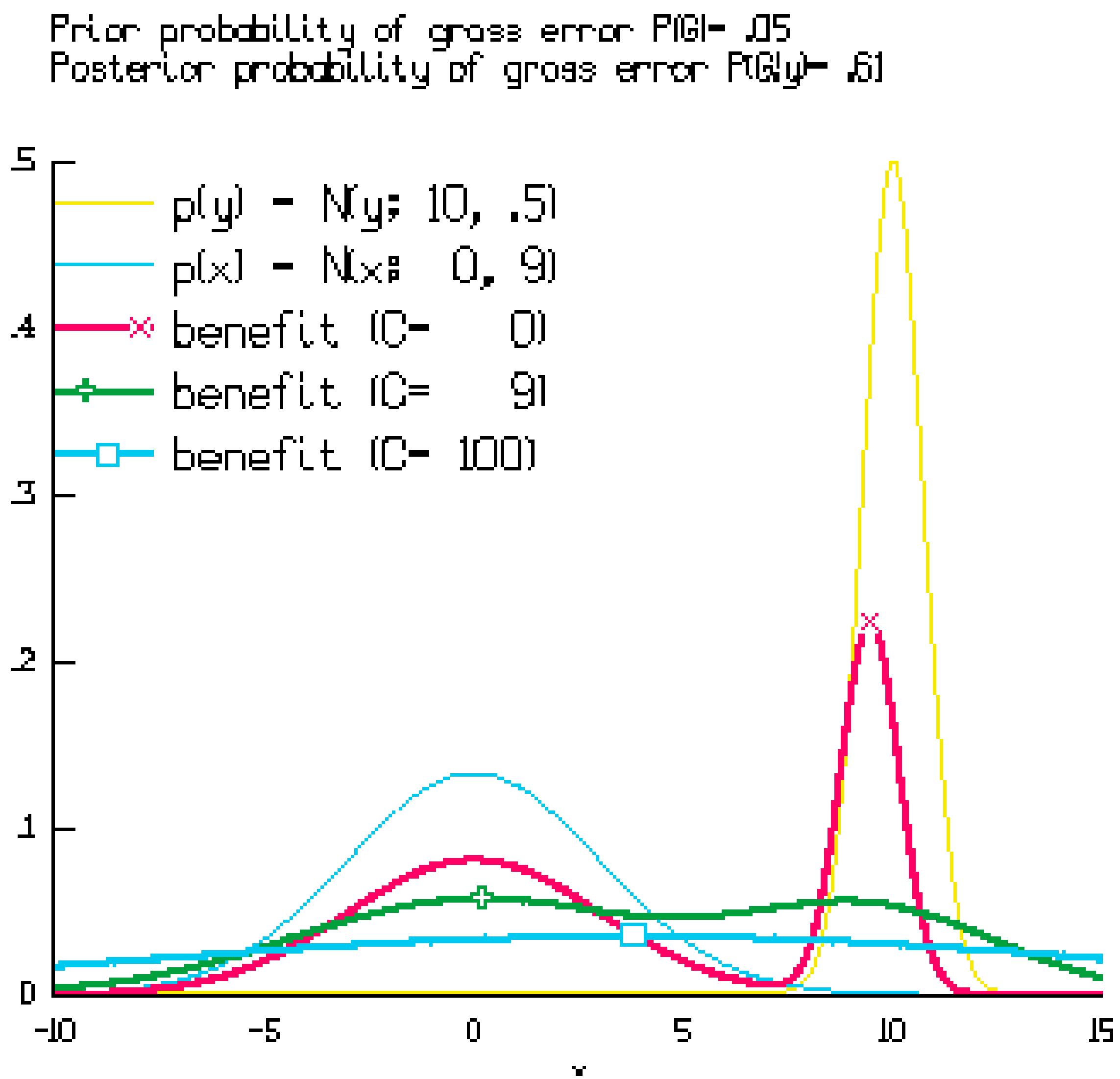
L1 norm is robust to outliers



Best fit straight lines to data including a gross error: solid line using a quadratic (L2) norm, dotted line using a mean absolute (L1) norm. (Based on Tarantola 1987).

What answer do we want?

- If we can afford to represent it (eg by an ensemble) then the full posterior PDF carries all the information.
- Normally we can only cope with a single “best” estimate.
- Can only find best w.r.t. a “loss function” specifying how much we lose for a wrong estimate. Lorenc (2002) suggested using a Gaussian-shaped function:
  - a very narrow “delta-function” finds the mode *x*,;
  - a very broad (quadratic) function finds the mean □;
  - can argue for a width similar to the background PDF’s; this finds the mean of the largest peak **+**.



Lorenc, A. C., 2002: Atmospheric Data Assimilation and Quality Control. *Ocean Forecasting*, eds Pinardi & Woods.

Coping with multiply minima –

finding best combination of inter-related decisions.

- Do a preliminary “pass” [with larger scale] before performing QC.
- Start with inflated observational errors, decreasing to correct values during the iteration (Dharssi *et al*. 1992)
- Only switch on variational QC after N (~20) iterations.
- Rely mainly on QC v background forecast (which is assumed to have Gaussian errors).
- Design iterative “buddy check” algorithm to start with easiest decisions.
- Change to robust DA algorithm (e.g. Huber norm).

Models for Observation Error PDFs

- Gaussian + "null" prior.

  Simple. Assumes that erroneous observations do not add to prior knowledge. Used in most of my QC work. Problems normalising – does not integrate to 1. Gives multiple minima.

- Gaussian + wider Gaussian.

  Similar in practice to above. Normalised. Erroneous obs make small change to prior. Gives multiple minima.

- Back to back exponentials.

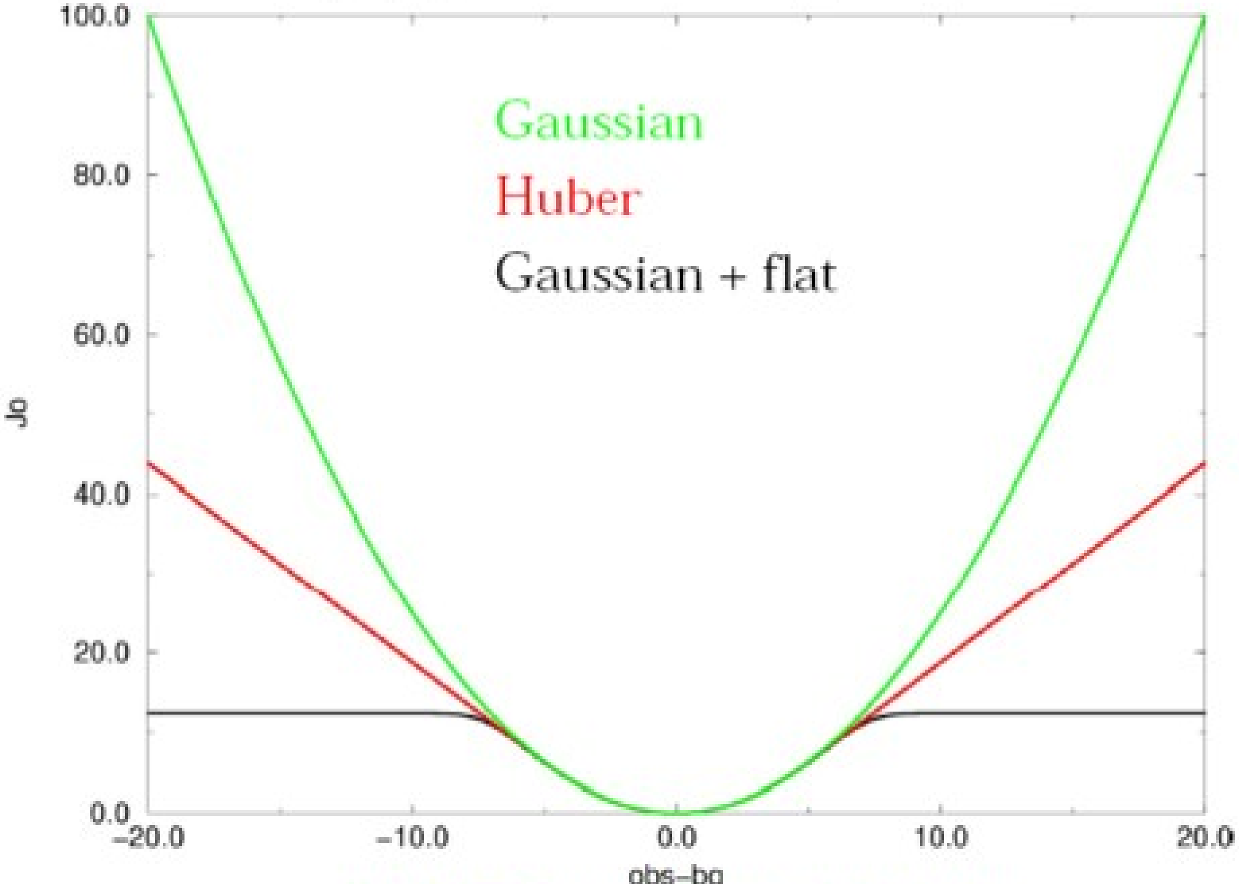
  Implied by use of L1 norm (Tarantola 1987). Finds median – "pull" of all observations is equal. Not differentiable at origin, so difficult to minimise. Robust.

- Huber Norm.

  Huber (1973), Guitton and Symes (2003). Used at ECMWF. L2 for small deviations, L1 for large. Finds consensus average - "pull" of observations limited, rather than increasing indefinitely with misfit. Robust.

The Huber-norm – a compromise between the l_2 and l_1 norms

$$p^H = \begin{cases} x^2/2 & \text{if } |x| <= k, \\ k|x| - k^2/2 & \text{if } |x| > k, \end{cases}$$



DA/SAT Training Course, April 2008

Erik Andersson

Different “best” values from the posterior PDF, plotted against ob-background.

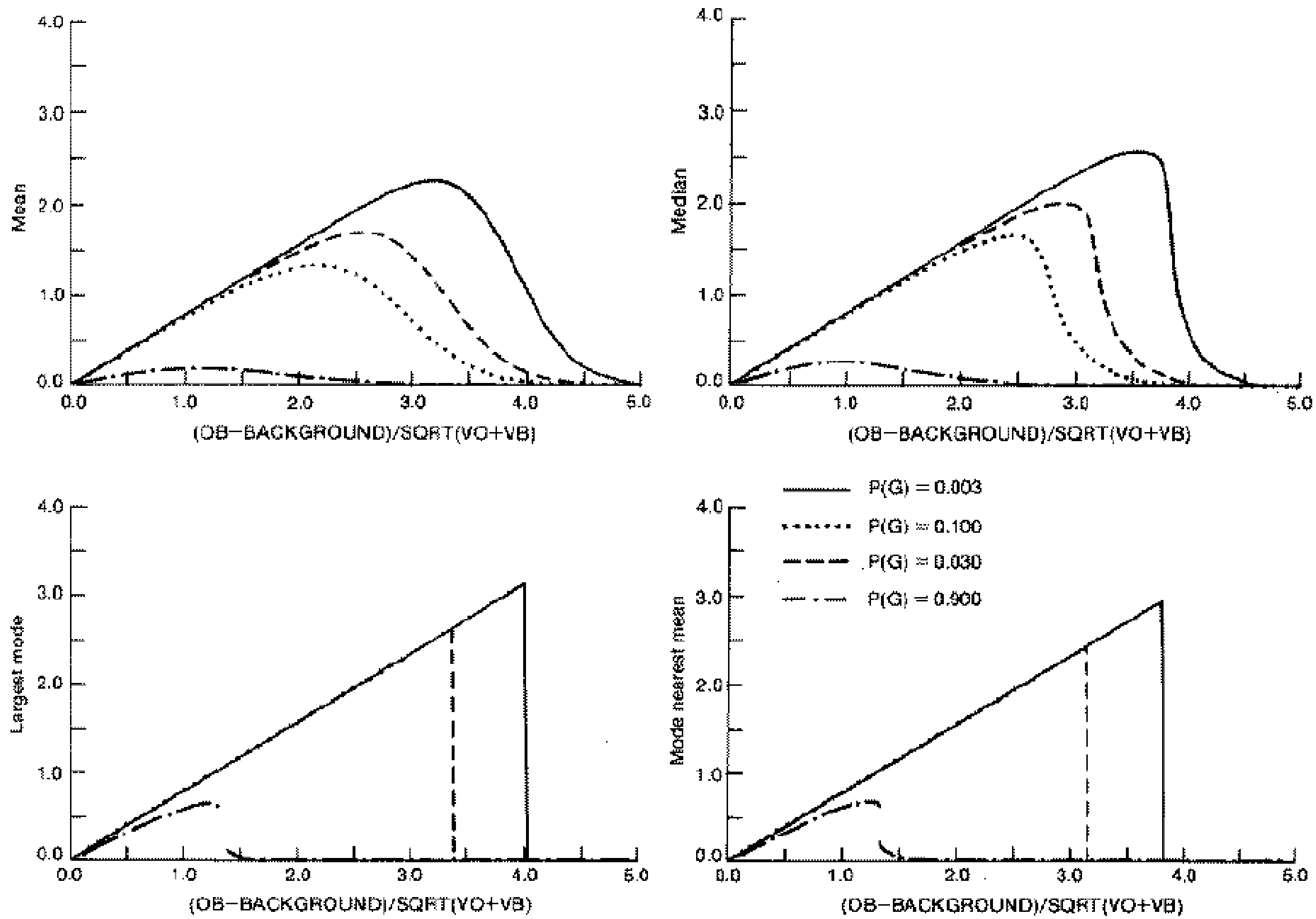
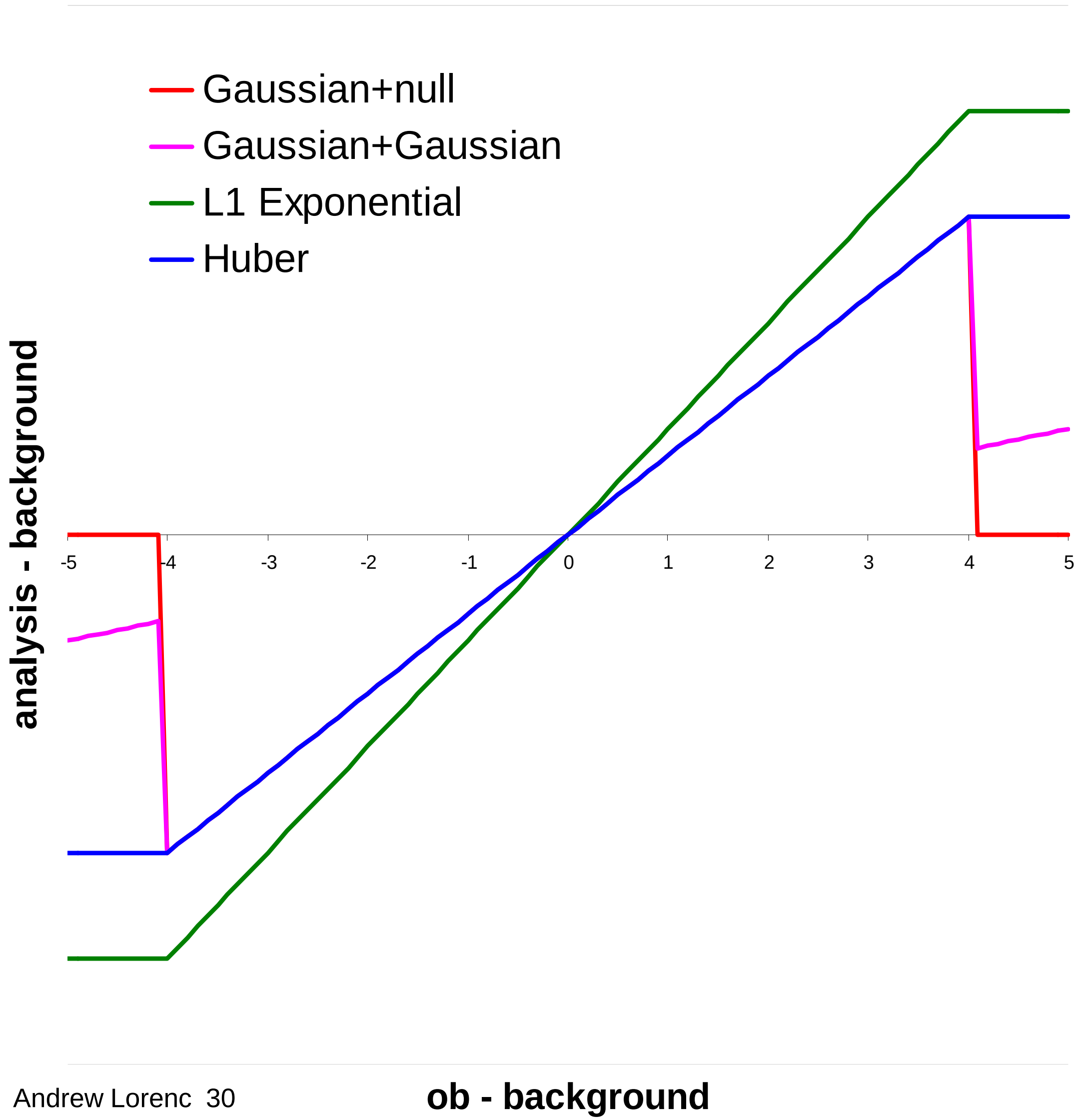


Figure 3. Mean, median, largest mode, and mode nearest the mean, of the analysis distribution, plotted against normalized observed minus background value, for various prior probabilities of gross error.

Lorenc and Hammon, 1988

Minimum of single observation penalty function for various observation norms



Observation Monitoring

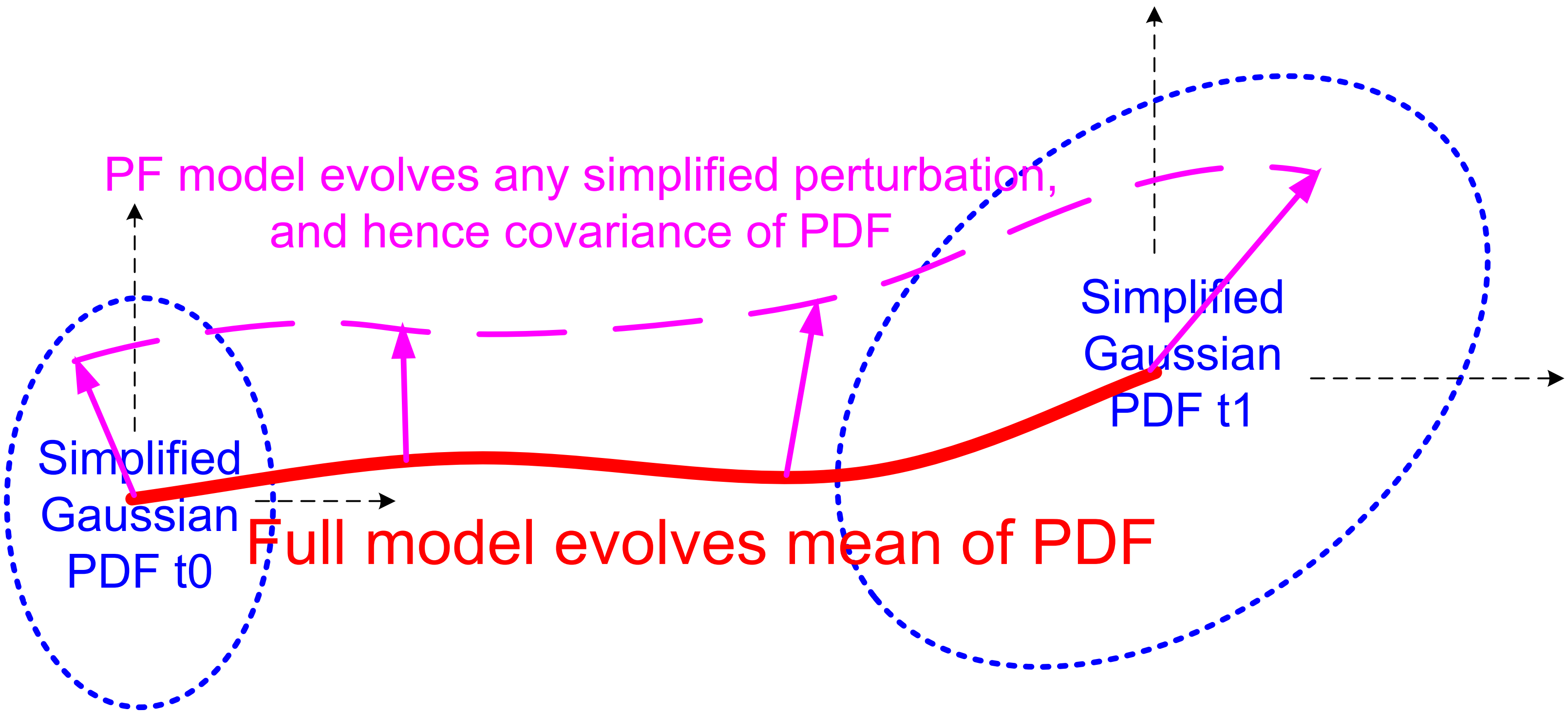
- *It is surprisingly difficult to demonstrate consistent significant impact from QC of individual observations!*
- “Monitoring” is probably more important – collection of statistics on the performance of observing and processing systems, detection of systems not performing as expected, & feedback so the deficiency is corrected. Needs:
  - a comprehensive database of basic and processed observed values, independent estimates of the same quantities, and parameters affecting the processing (recently enhanced by “adjoint sensitivities”);
  - software for categorising, sorting, and analysing the database;
  - effort to try categorisations and look for "unexpected" behaviour;
  - communications, willpower and persistence, to get errors from stages out of your direct control rectified.

Met Office

Aside: Nonlinearities in Outer-loop statistical 4D-Var

Thanks to Jim Purser who introduced me to the Bayesian approach.

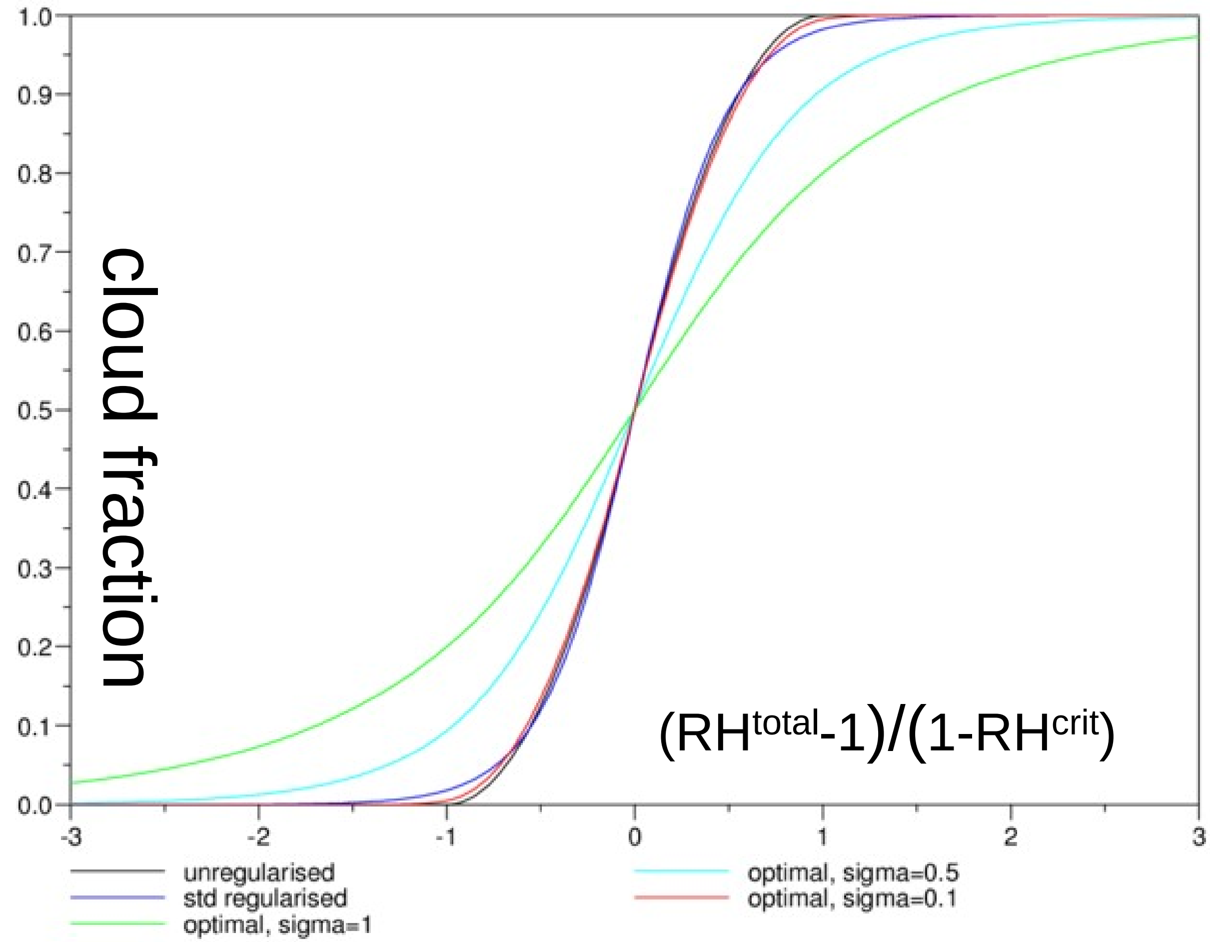
Statistical, incremental 4D-Var



Statistical 4D-Var approximates entire PDF by a Gaussian.

Perturbation Forecast model for Incremental 4D-Var

- Minimise: $I = E\left\{ \left\| \underline{M}\left(\mathbf{x}+\delta\mathbf{x}\right) - \underline{\bar{M}}\left(\mathbf{x}\right) - \underline{\tilde{\mathbf{M}}}\left(\mathbf{x}\right)\delta\mathbf{x} \right\| \right\}$
- Designed to give best fit for finite perturbations
- Not Tangent-Linear
- Filters unpredictable scales and rounds IF tests
- Requires physical insight – not just automatic differentiation

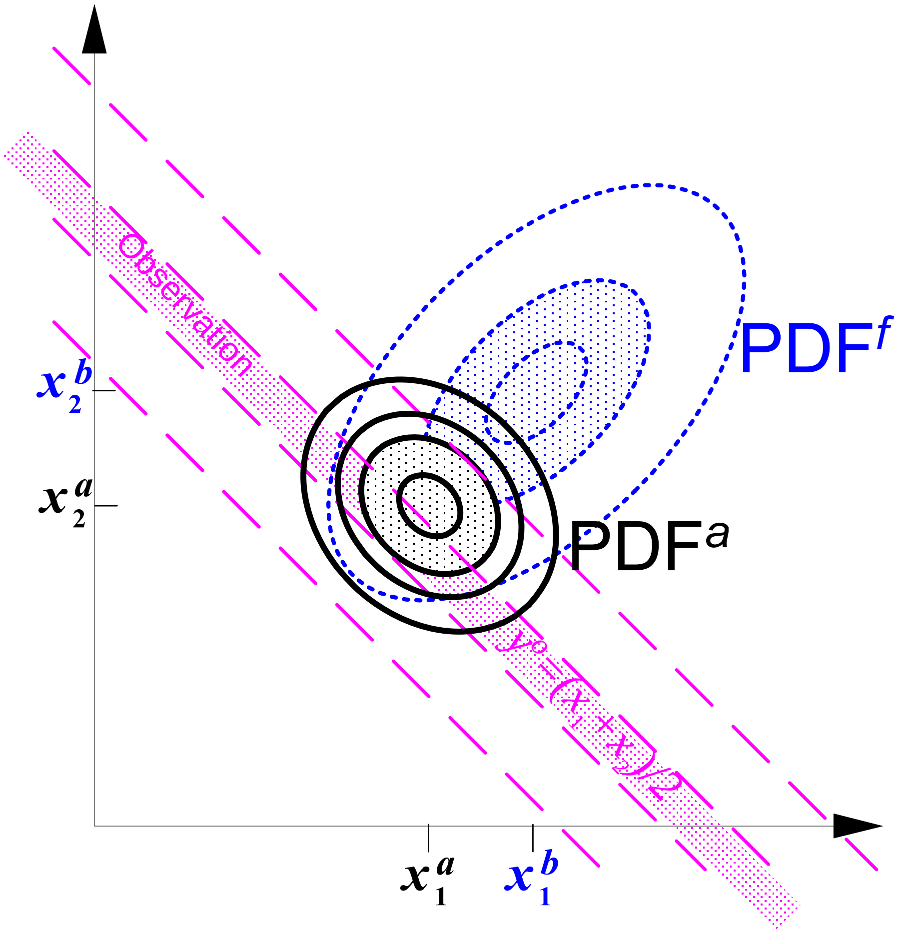


Tim Payne

What spread to assume in regularisation?

- If guess=background, need to approximate whole of PDF^f

- In final outer-loop, only need to approximate PDF^a



Questions and answers

Content

1. Bayes' Theorem for discrete events: "There is a gross error"
   - Operational Met Office Bayesian QC.
2. Non-Gaussian PDFs and non-quadratic penalty functions
3. Importance of prior statistics
4. Interaction of QC between observations
   - Non-Gaussian DA
   - L1 & Huber norms
5. What answer do we want?
6. Observation Monitoring
7. (Designing linear PF model near nonlinearities)