# How does 4D-Var handle Nonlinearity and non-Gaussianity?

**Mike Fisher**

Acknowledgements: Christina Tavolato, Elias Holm, Lars Isaksen, Tavolato, Yannick Tremolet

# Outline of Talk

- **Non-Gaussian pdf's in the 4d-Var cost function**
  - **Variational quality control**
  - **Non-Gaussian background errors for humidity**
- **Can we use 4D-Var analysis windows that are longer than the timescale over which non-linear effects dominate?**
  - **Long-window, weak constraint 4D-Var**

# Non-Gaussian pdf's in the 4D-Var cost function

- **The 3D/4D-Var cost function is simply the log of the pdf:**

$$J(x) \propto -\log\left(p(x \mid y, x_b)\right) \propto -\log\left(p(y \mid x)\right) - \log\left(p(x_b \mid x)\right) - \ldots$$

- **Non-Gaussian pdf's of observation error and background error result in non-quadratic cost functions.**
- **In principle, this has the potential to produce multiple minima – and difficulties in minimization.**
- **In practice, there are many cases where the ability to specify non-Gaussian pdf's is very useful, and does not give rise to significant minimization problems.**
  - **Directionally-ambiguous scatterometer winds**
  - **Variational quality control**
  - **Bounded variables: humidity, trace gasses, rain rate, etc.**

# Variational quality control and robust estimation

- **Variational quality control has been used in the ECMWF analysis for the past 10 years.**
- **Observation errors are modelled as a combination of a Gaussian and a flat (boxcar) distribution:**

$$p(y \mid x) = (1 - P_G)N + P_G G, \quad \text{where } P_G = p(\text{gross error}), \text{ and:}$$

$$N = \frac{1}{\sigma_o \sqrt{2\pi}} \exp\left[ -\frac{1}{2} \left( \frac{y - H(x)}{\sigma_o} \right)^2 \right]$$

$$G = \frac{1}{D} \quad \text{if } \left| y - H(x) \right| < D/2, \quad \text{zero otherwise}$$

- **With this pdf, observations close to x are treated as if Gaussian, whereas those far from x are effectively ignored.**

# Variational quality control and robust estimation

- **An alternative treatment is the Huber metric:**

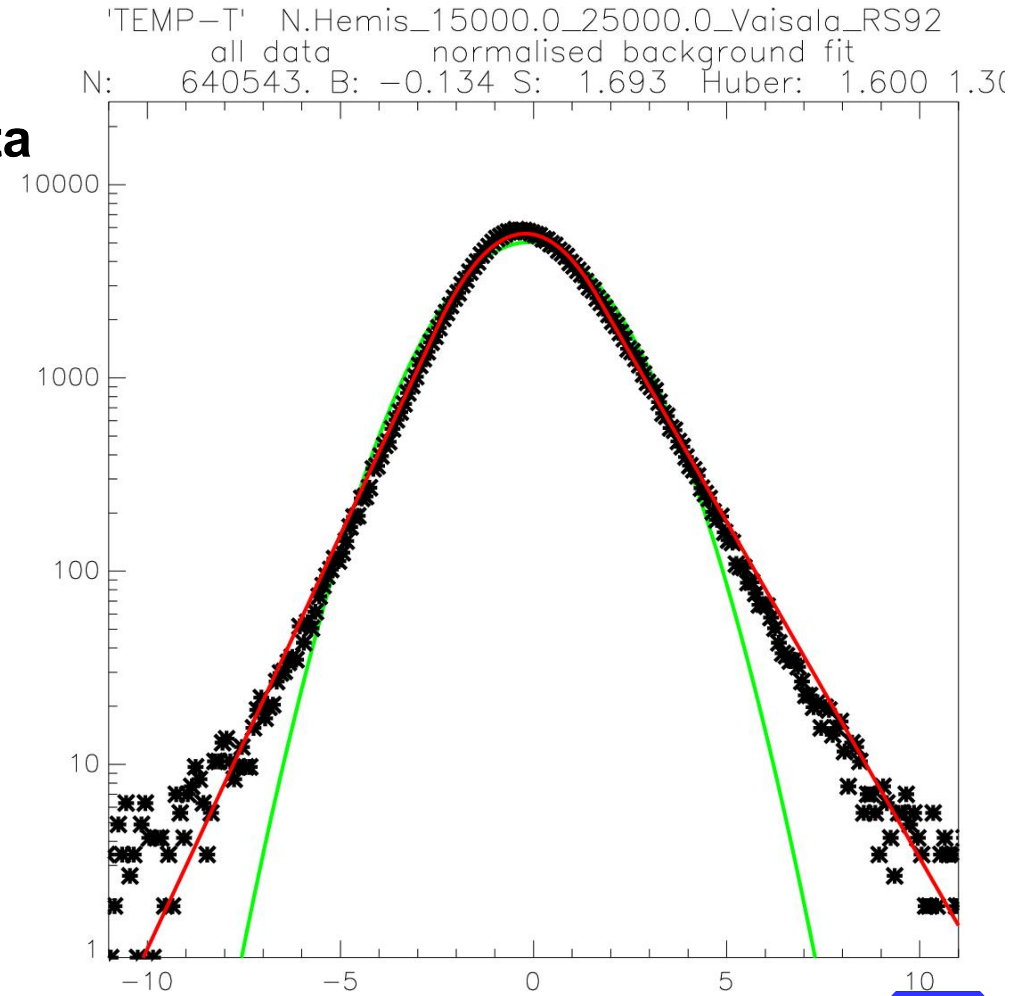
$$p(y|x) = \begin{cases} \frac{1}{\sigma_o\sqrt{2\pi}} \exp\left(\frac{a^2}{2} - |a\delta|\right) & \text{if } a < \delta \\ \frac{1}{\sigma_o\sqrt{2\pi}} \exp\left[-\frac{1}{2}\delta^2\right] & a \le \delta \le b \\ \frac{1}{\sigma_o\sqrt{2\pi}} \exp\left(\frac{b^2}{2} - |b\delta|\right) & \text{if } \delta > b \end{cases}$$

$$\text{where } \delta = \frac{y - H(x)}{\sigma_o}$$

- **Equivalent to $L_1$ metric far from x, $L_2$ metric close to x.**
- **With this pdf, observations far from x are given less weight than observations close to x, but can still influence the analysis.**
- **Many observations have errors that are well described by the Huber metric.**

# Variational quality control and robust estimation

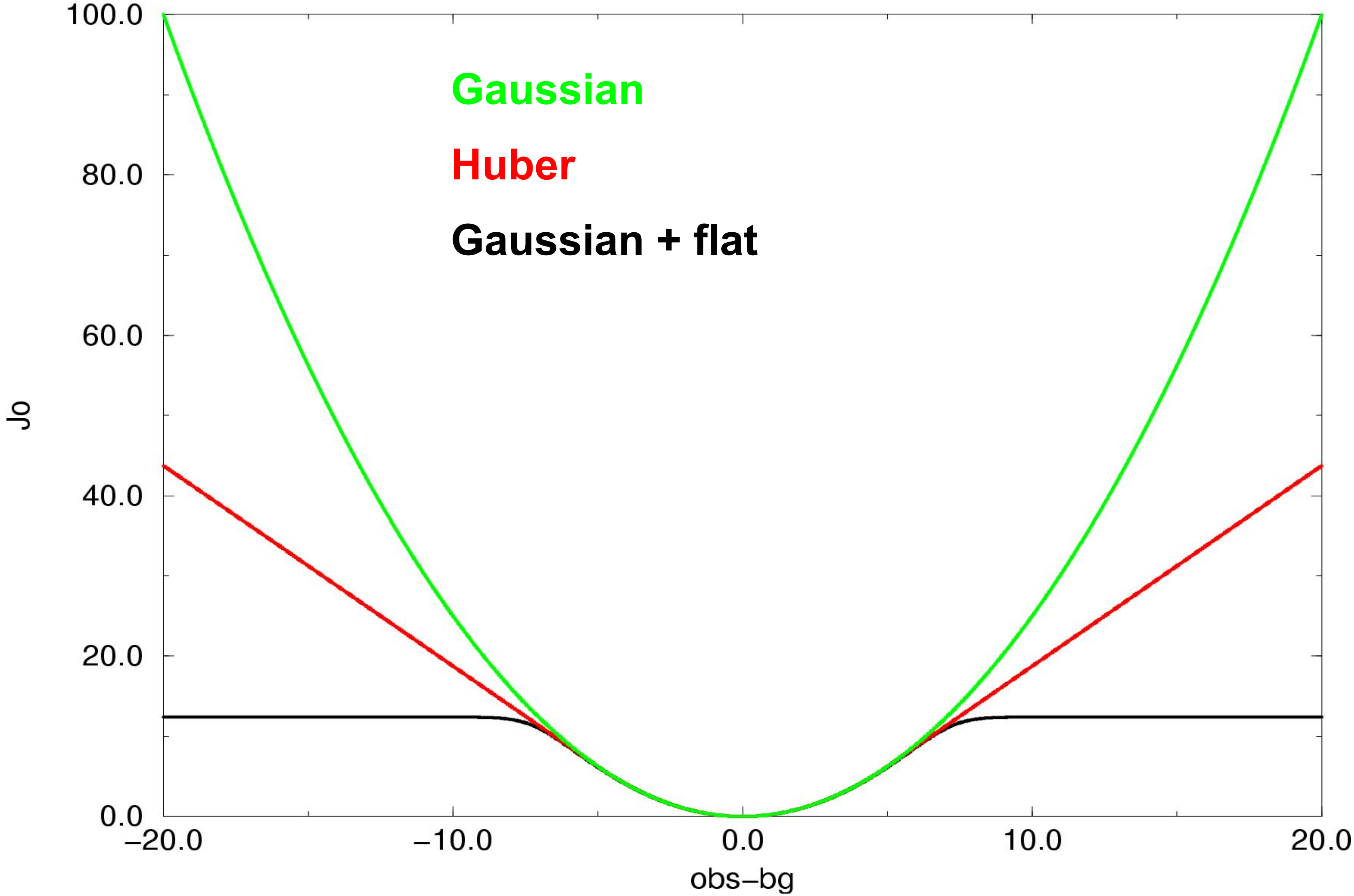
- **18 months of conventional data**
  - **-Feb 2006 – Sep 2007**
- **Normalised fit of PDF to data**
  - **- Best Gaussian fit**
  - **- Best Huber norm fit**






ECMWF

# Variational quality control and robust estimation



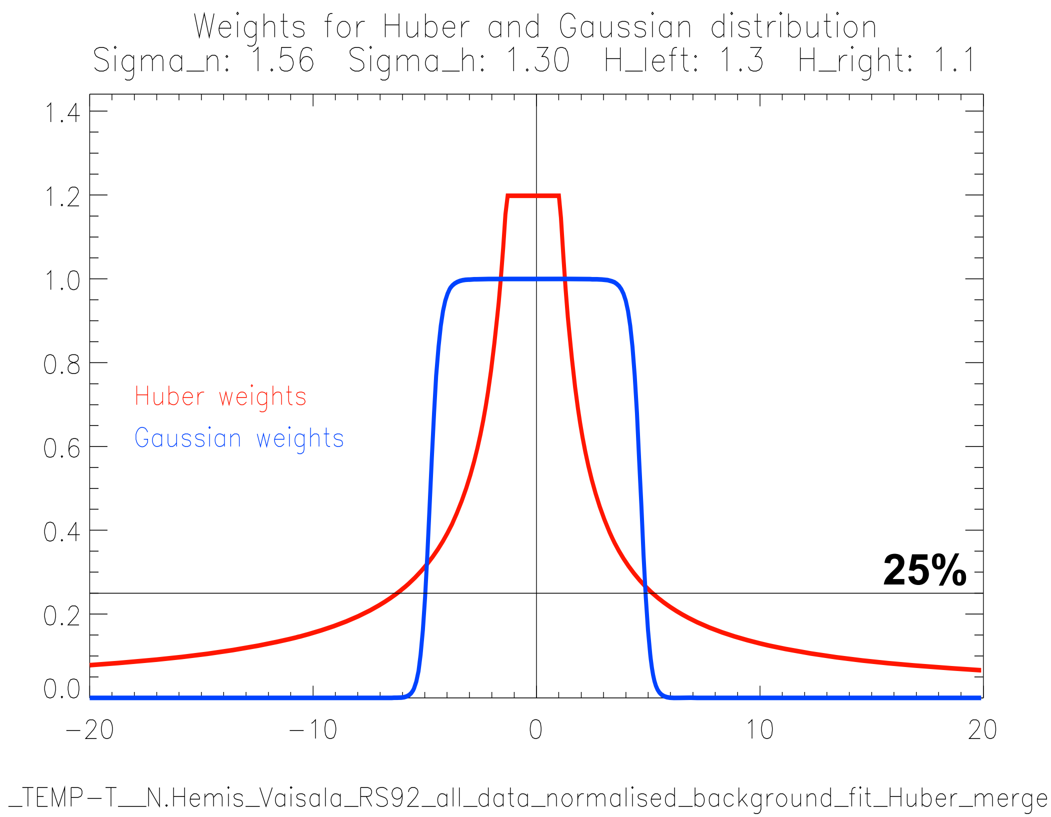



ECMWF

# Comparing optimal observation weights
## Huber-norm (red) vs. Gaussian+flat (blue)

- **More weight in the middle of the distribution**
  - **-$\sigma_o$ was retuned**
- **More weight on the edges of the distribution**
- **More influence of data with large departures**
  - **-Weights: 0 – 25%**

Weight relative to gaussian (no VarQC) case



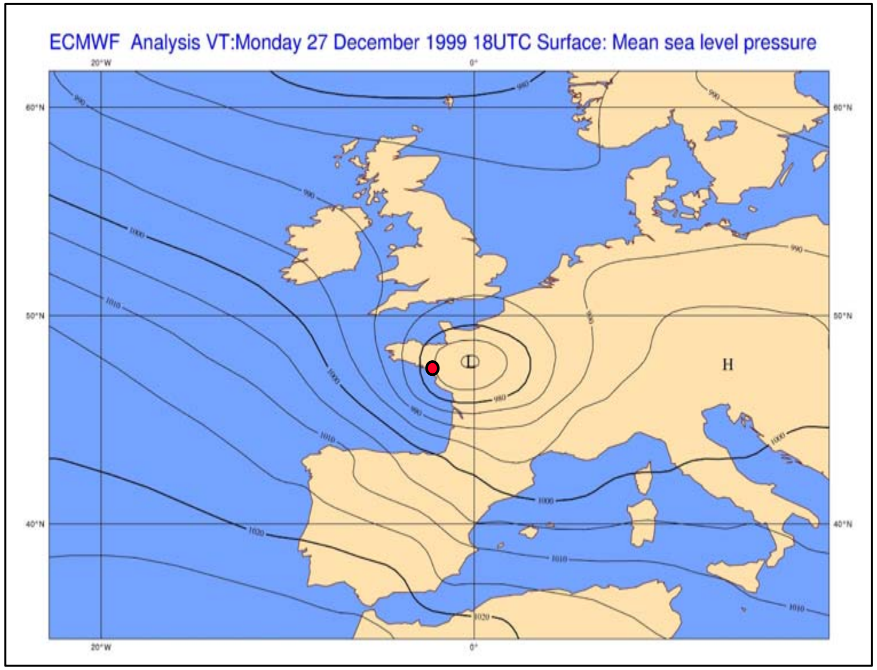



ECMWF

# Test Configuration

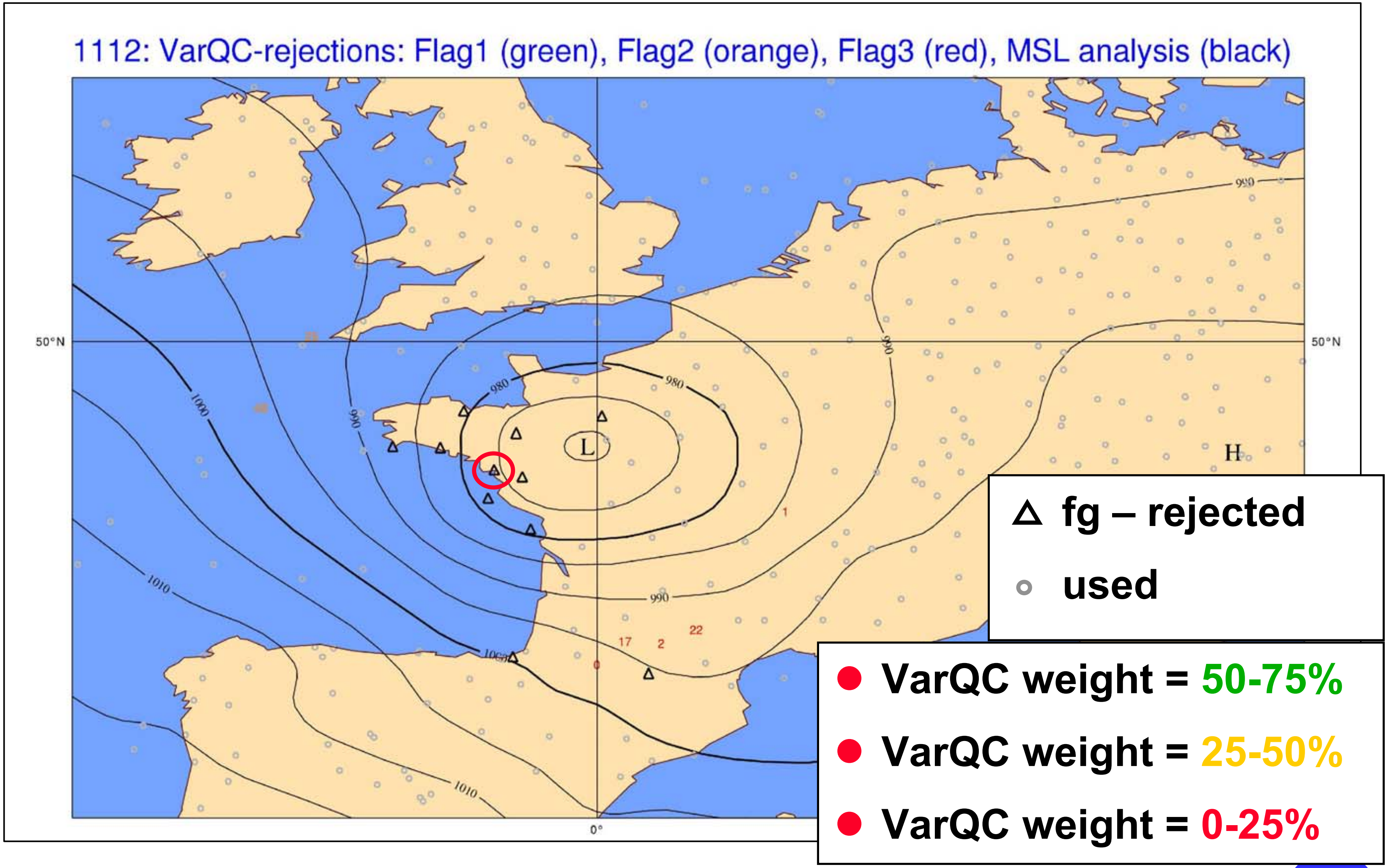
- **Huber norm parameters for**
  - **SYNOP, METAR, DRIBU: surface pressure, 10m wind**
  - **TEMP, AIREP: temperature, wind**
  - **PILOT: wind**
- **Relaxation of the fg-check**
  - **Relaxed first guess checks when Huber VarQC is done**
  - **First Guess rejection limit set to 20σ**
- **Retuning of the observation error**
  - **Smaller observation errors for Huber VarQC**

# French storm 27.12.1999

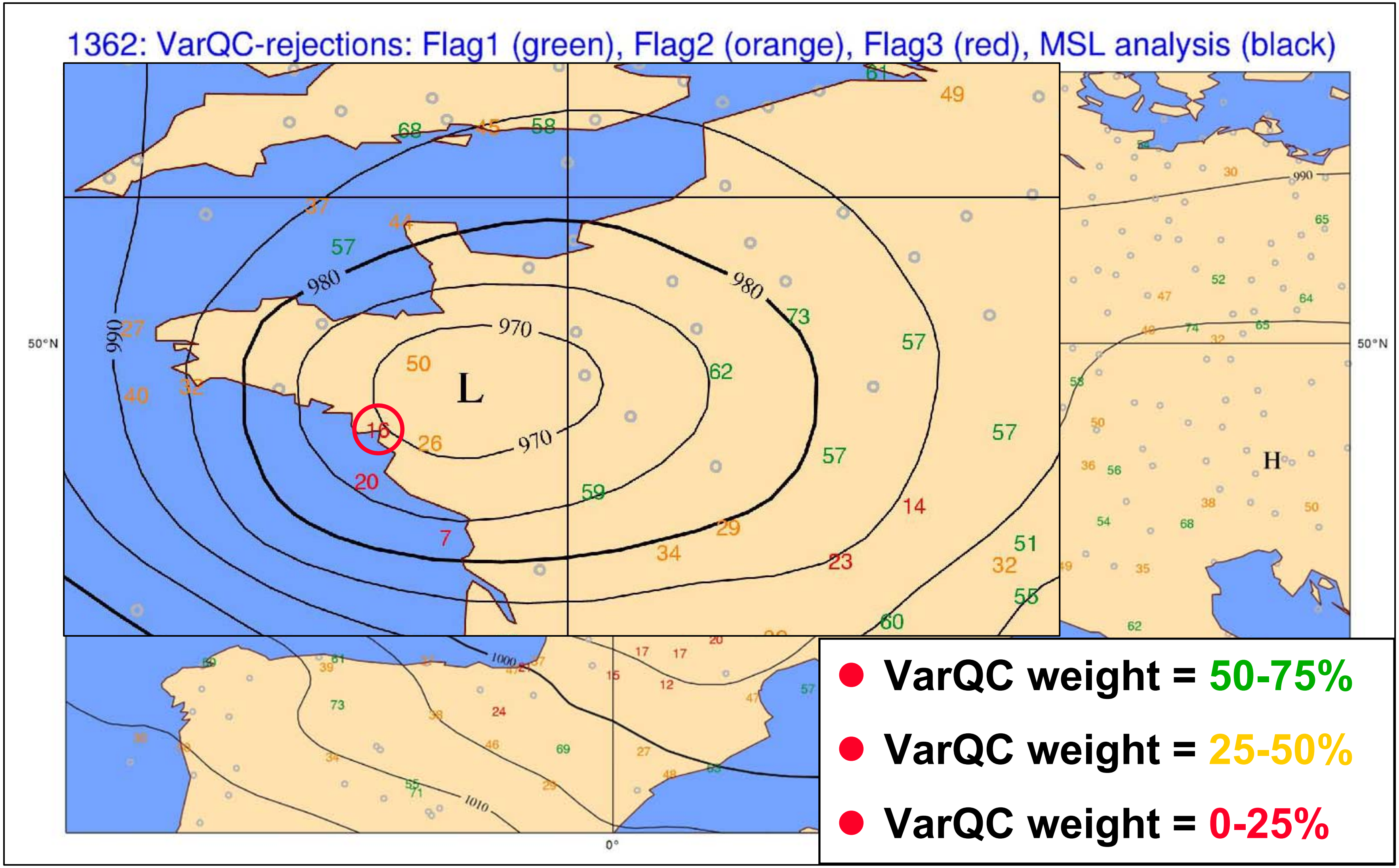



- **Surface pressure:**
  - **-Model (ERA interim T255): 970hPa**
  - **-Observations: 963.5hPa**
  - **-Observation are supported by neighbouring stations!**

# Data rejection and VarQC weights – Era interim 27.12.99 18UTC +60min

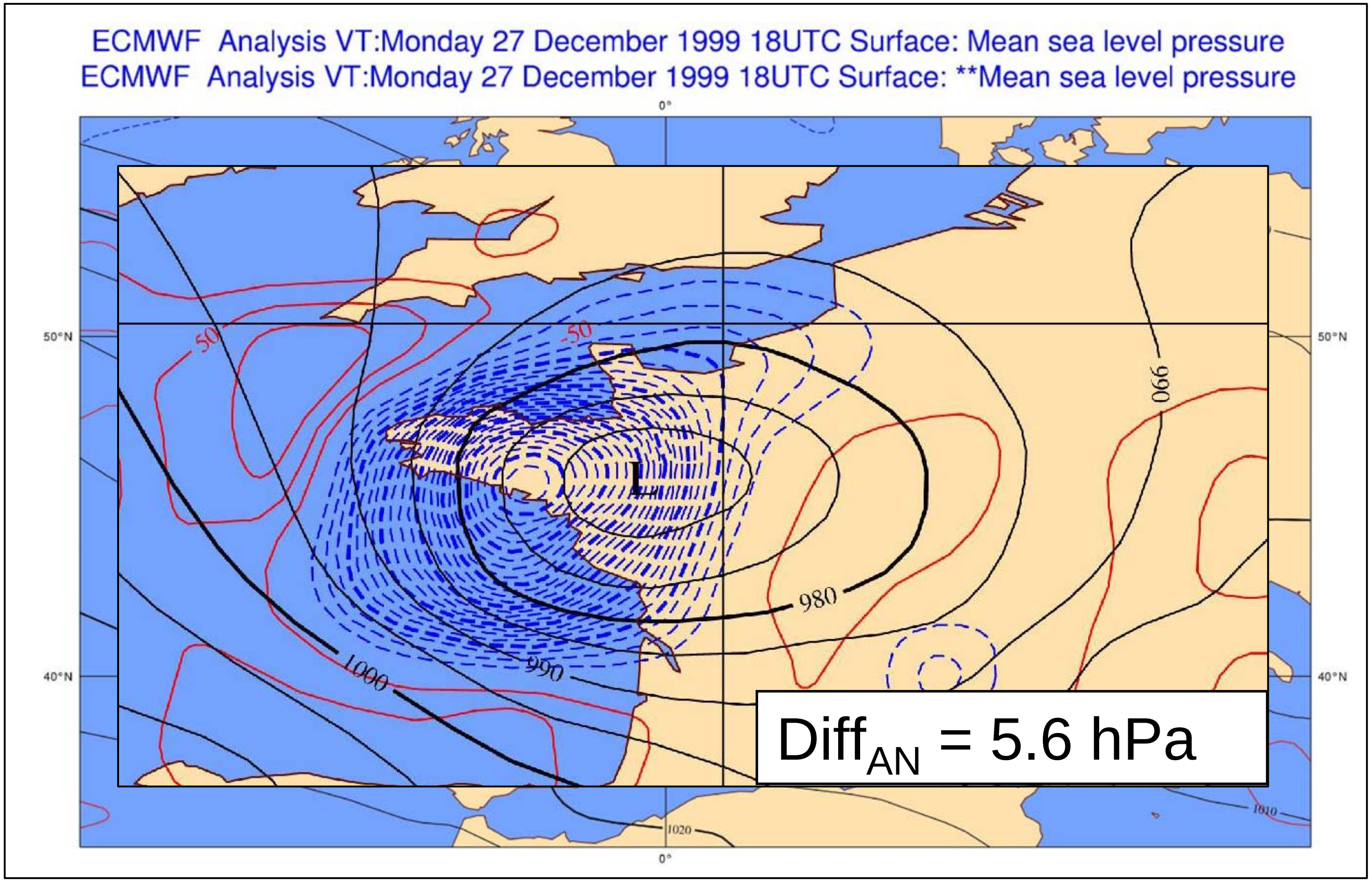


# Data rejection and VarQC weights – Huber exp.




- VarQC weight = 50-75%
- VarQC weight = 25-50%
- VarQC weight = 0-25%

# MSL Analysis differences: Huber – Era

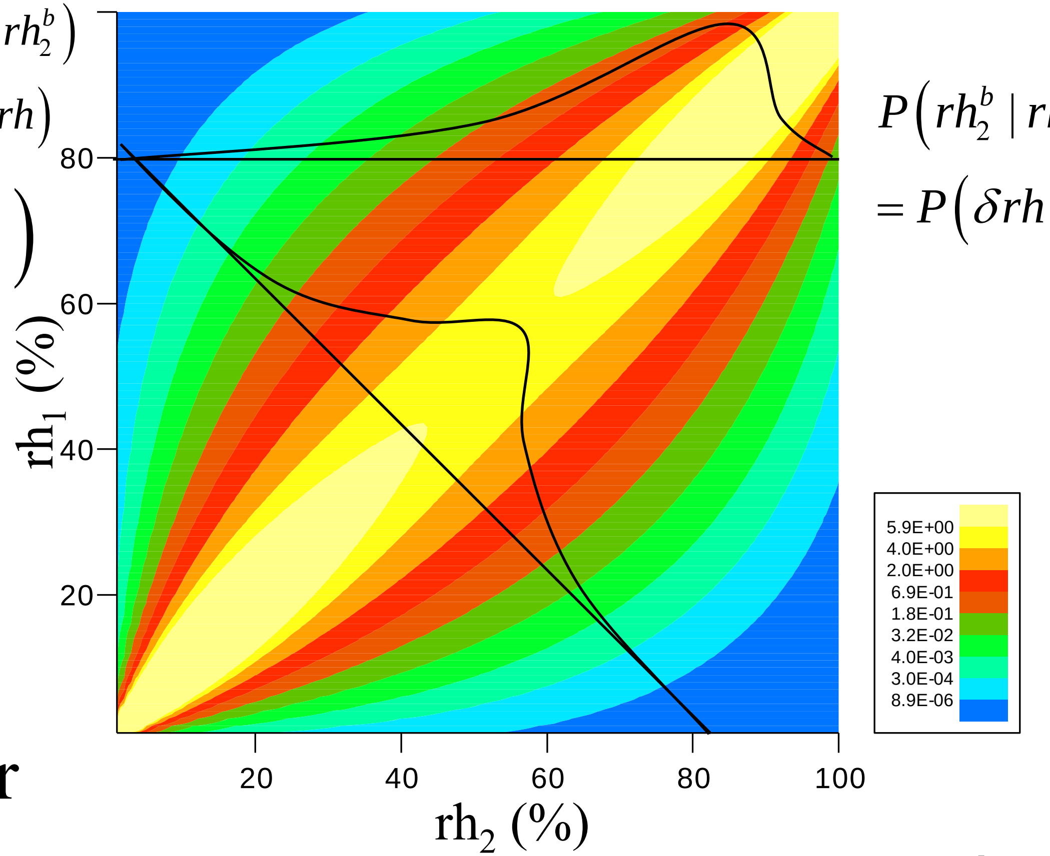



- **New min 968 hPa**
- **Low shifted towards the lowest surface pressure observations**

# Humidity control variable

$$P\left(rh_2^b - rh_1^b \mid rh_1^b + rh_2^b\right) = P\left(\delta rh \mid 2rh_1^b + \delta rh\right)$$

Joint pdf: $P\left(rh_1^b, rh_2^b\right)$ for two members of an ensemble of 4D-Var analyses. $\delta rh = rh_2^b - rh_1^b$ is representative of background error

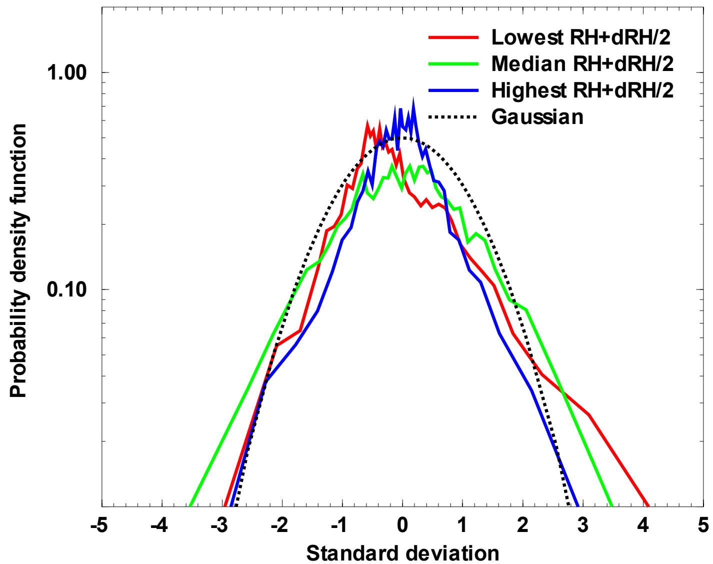


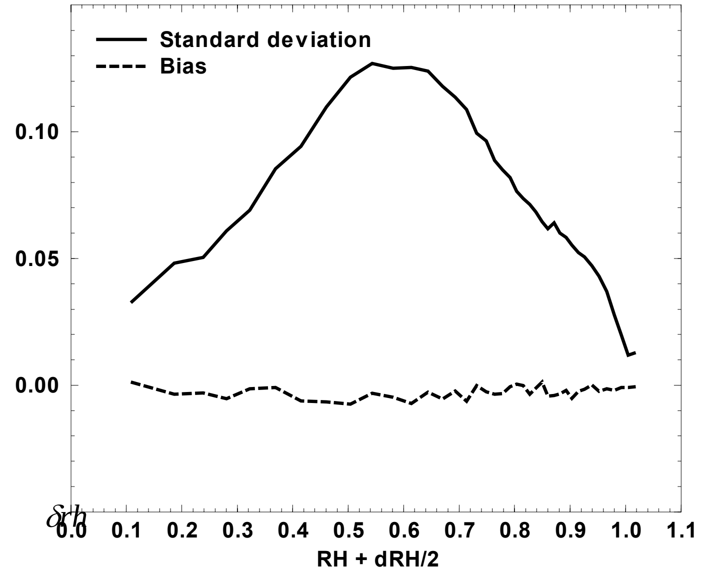

$$P\left(rh_2^b \mid rh_1^b\right) = P\left(\delta rh + rh_1^b \mid rh_1^b\right)$$

The pdf of background error is asymmetric when stratified by $rh^b$

The pdf of background error is symmetric when stratified by $rh^b + \frac{1}{2}\delta rh$

# Humidity control variable




The symmetric pdf $P\left(\delta rh \mid rh^b + \frac{1}{2}\delta rh\right)$ can be modelled by a normal distribution.

The variance changes with $rh^b + \frac{1}{2}\delta rh$ and the bias is zero.

A control variable with an approximately unit normal distribution is obtained by a nonlinear normalization:

$$\tilde{\delta rh} = \frac{\delta rh}{\sigma\left(rh^b + \frac{1}{2}\delta rh\right)}$$


ECMWF

# Humidity control variable

- **The background error cost function $J_b$ is now nonlinear.**

$$J_b = \left[ f\left(\delta rh\right)\right]^T B^{-1} f\left(\delta rh\right) \quad \text{where} \quad f\left(\delta rh\right) = \frac{\delta rh}{\sigma\left(rh^b + \frac{1}{2}\delta rh\right)}$$

- **Our implementation requires linear inner loops (so that we can use efficient, conjugate-gradient minimization).**
  - **Inner loops: use** $\widetilde{\delta rh} = \delta rh / \sigma\left(rh^b\right)$
  - **Outer loops: solve for** $\delta rh$ **from the nonlinear equation:**

$$\frac{\delta rh}{\sigma\left(rh^b + \frac{1}{2}\delta rh\right)} = \widetilde{\delta rh}$$

# What about Multiple Minima?

- **Example: strong-constraint 4D-Var for the Lorenz three-variable model:**

from: Roulstone, 1999

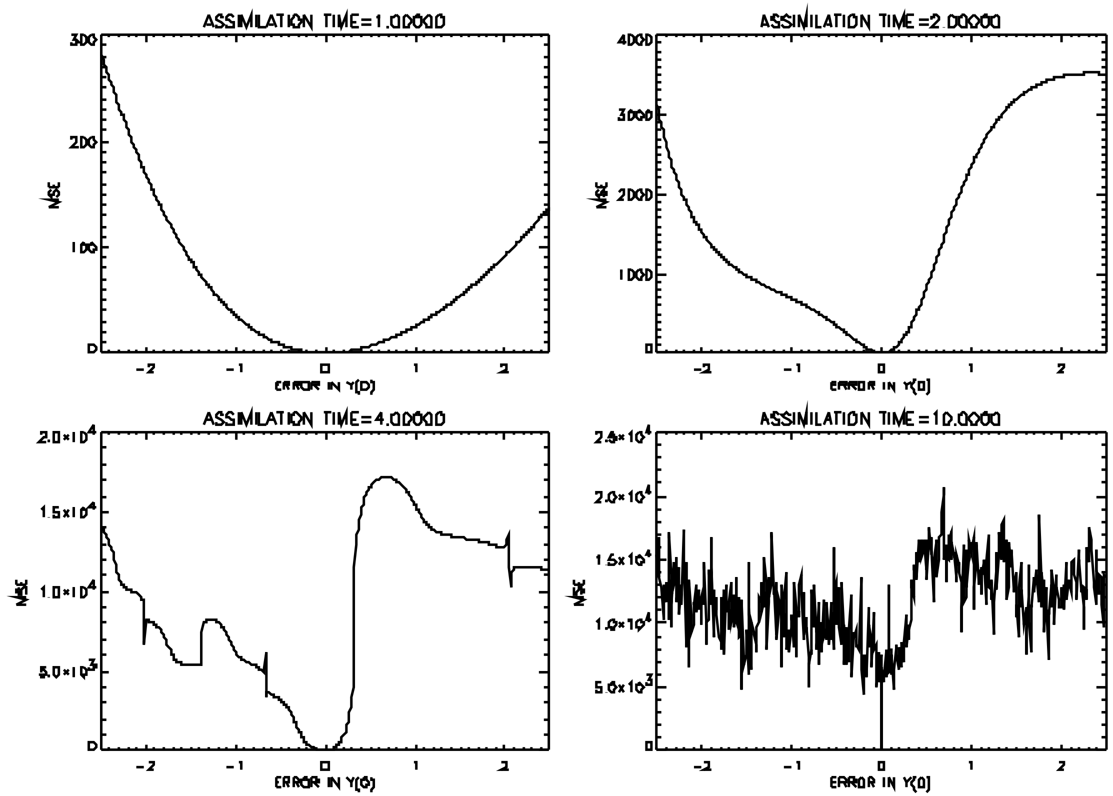



Figure 1: The MSE cost function in the Lorenz model as a function of error in the initial value of the Y coordinate. The function becomes increasingly pathological as the assimilation period is increased.


ECMWF

# What about Multiple Minima?

- **In strong-constraint 4D-Var, the control variable is $x_0$.**

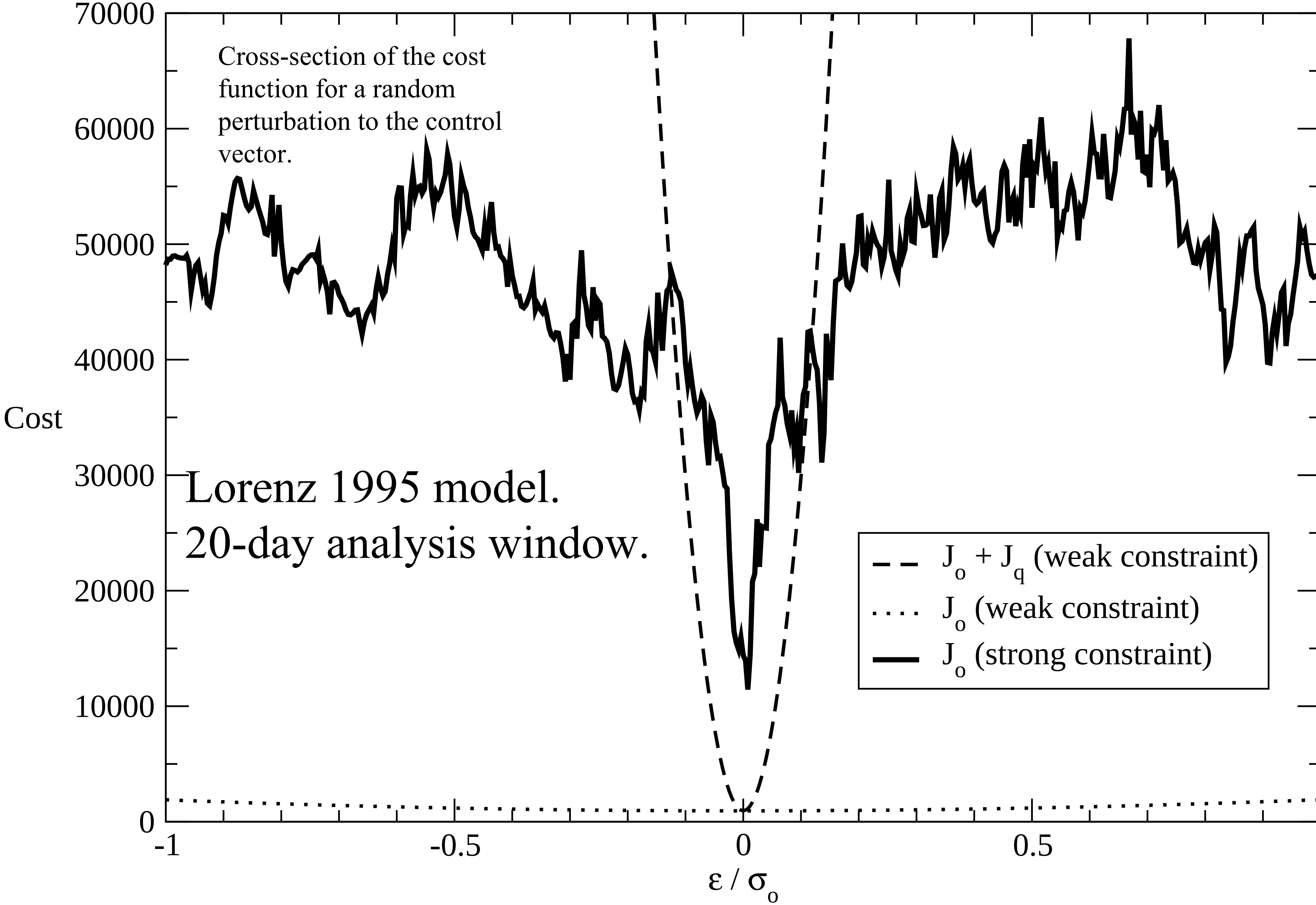
$$J_o = \sum_{k=0}^{K} \left[ y_k - H_k \left( M_{t_0 \to t_k} \left( x_0 \right) \right) \right]^T R^{-1} \left[ y_k - H_k \left( M_{t_0 \to t_k} \left( x_0 \right) \right) \right]$$

- **We rely on the model to propagate the state from initial time to observation time.**
- **For long windows, this results in a highly nonlinear $J_o$.**
- **In weak-constraint 4D-Var, the control variable is $(x_0, x_1, \ldots, x_K)$, and (for linear observation operators) $J_o$ is quadratic.**

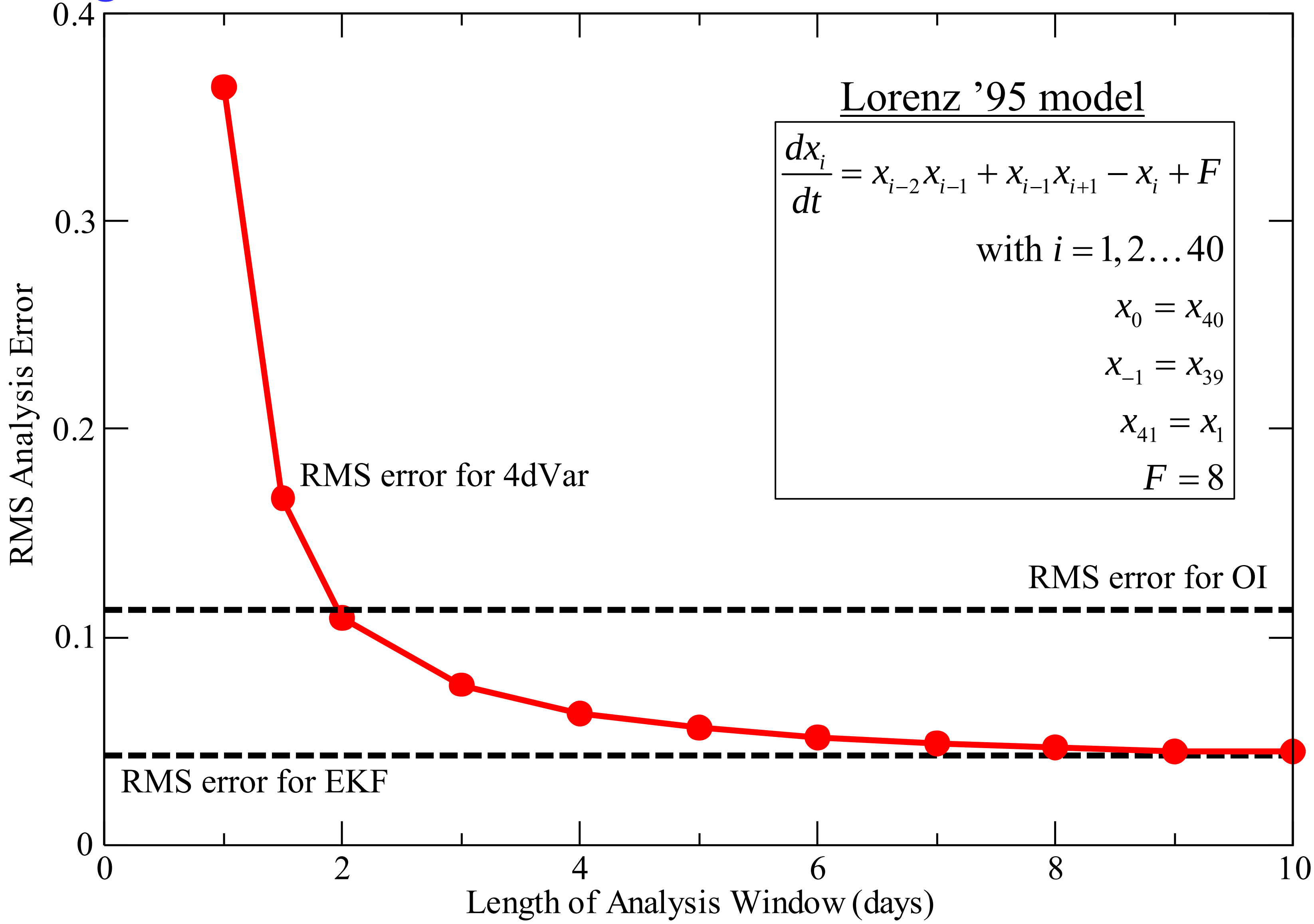
$$J_o = \sum_{k=0}^{K} \left[ y_k - H_k \left( x_k \right) \right]^T R^{-1} \left[ y_k - H_k \left( x_k \right) \right]$$

- **$J_q$ is close to quadratic if the TL approximation is accurate over the sub-interval $[t_{k-1}, t_k]$.**

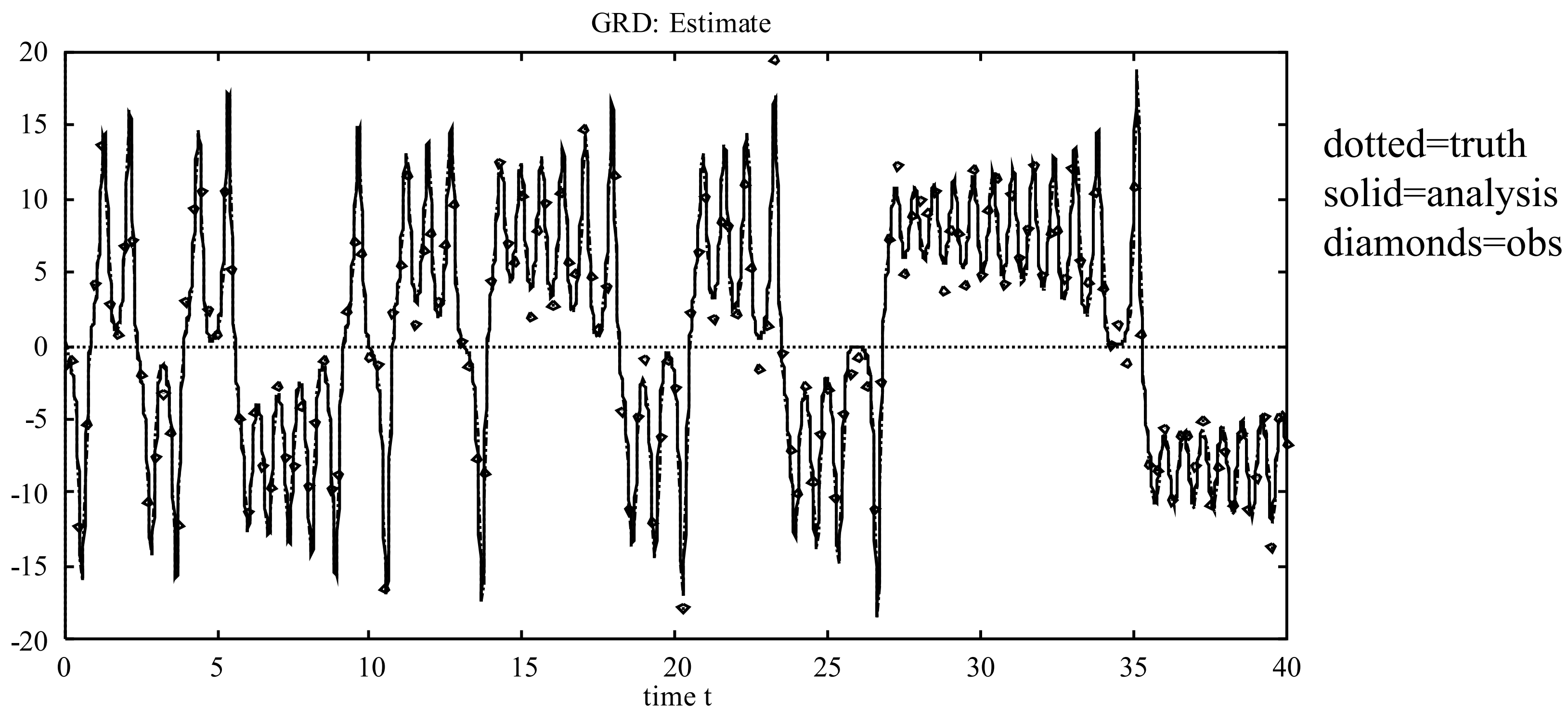
$$J_q = \sum_{k=1}^{K} \left[ x_k - M_{t_{k-1} \to t_k} \left( x_{k-1} \right) \right]^T R^{-1} \left[ x_k - M_{t_{k-1} \to t_k} \left( x_{k-1} \right) \right]$$


70000
60000
50000
40000
30000
20000
10000
0
Cost
-1
-0.5
0
0.5
1
ε / $\sigma_o$
Cross-section of the cost function for a random perturbation to the control vector.
Lorenz 1995 model.
20-day analysis window.
$J_o + J_q$ (weak constraint)
$J_o$ (weak constraint)
$J_o$ (strong constraint)

# Long-window, weak-constraint 4D-Var



# What about multiple minima?




- **From: Evensen (MWR 1997 pp1342-1354: Advanced Data Assimilation for Strongly Nonlinear Dynamics).**
  - **Weak constraint 4dVar for the Lorenz 3-variable system.**
  - **~50 orbits of the lobes of the attractor, and 15 lobe transitions.**

# What about multiple minima?

- The abstract from Evensen's 1997 paper is interesting:
  - This paper examines the properties of three advanced data assimilation methods when used with the highly nonlinear Lorenz equations. The ensemble Kalman filter is used for sequential data assimilation and the recently proposed ensemble smoother method and a gradient descent method* are used to minimize two different weak constraint formulations.
  - The problems associated with the use of an approximate tangent linear model when solving the Euler-Lagrange equations, or when the extended Kalman filter is used, are eliminated when using these methods. All three methods give reasonable consistent results with the data coverage and quality of measurements that are used here and overcome the traditional problems reported in many of the previous papers involving data assimilation with highly nonlinear dynamics.

*i.e. weak-constraint 4D-Var

# Weak Constraint 4D-Var in a QG model

- **The model:**
  - **Two-level quasi-geosptrophic model on a cyclic channel**

$$\frac{Dq_i}{Dt} = 0 \quad (\text{for } i = 1, 2)$$

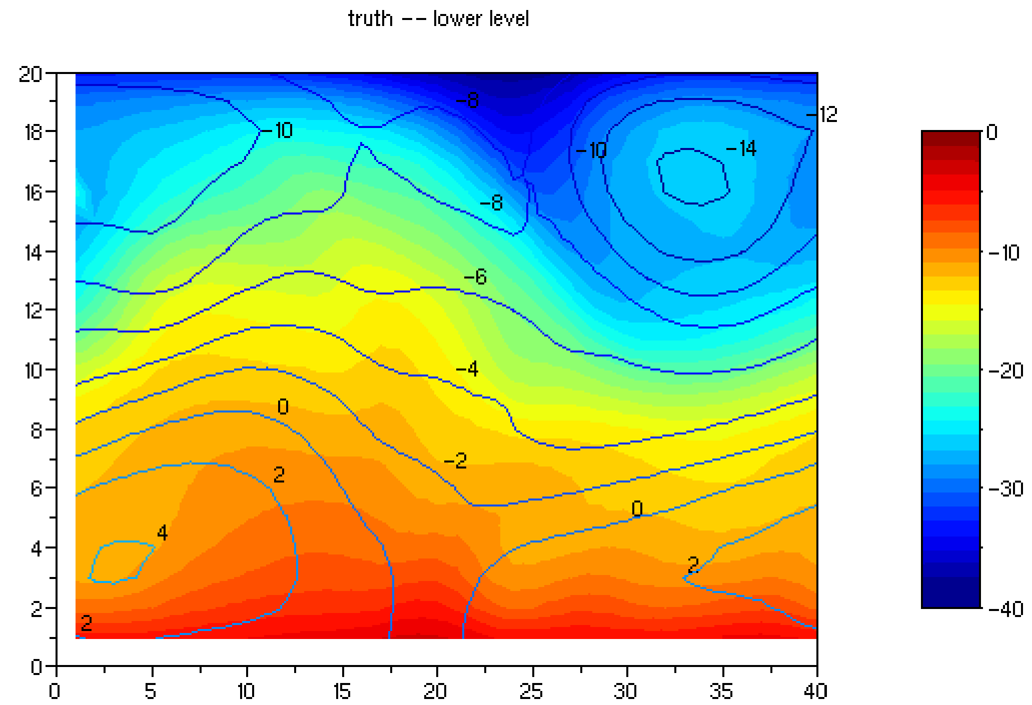
$$q_1 = \nabla^2 \psi_1 - F_1 \left( \psi_1 - \psi_2 \right) + \beta y$$

$$q_2 = \nabla^2 \psi_2 - F_2 (\psi_2 - \psi_1) + \beta y + R_s$$

  - **Solved on a 40×20 domain with Δx=Δy=300km**
  - **Layer depths $D_1$=6000m, $D_2$=4000m**
  - **$R_o$ = 0.1**
  - **Very simple numerics: first order semi-Lagrangian advection with cubic interpolation, and 5-point stencil for the Laplacian.**

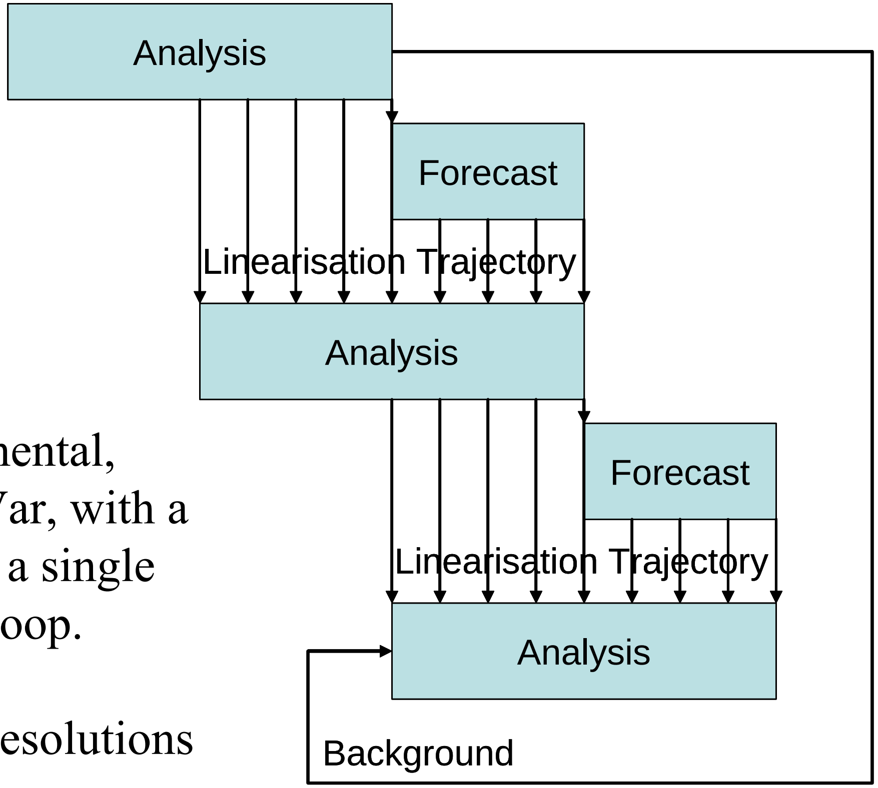
# Weak Constraint 4D-Var in a QG model

- **dt = 3600s**
- **dx = dy = 300km**
- **$f = 10^{-4}\ s^{-1}$**
- **$\beta = 1.5 \times 10^{-11}\ s^{-1}m^{-1}$**
- **D1 = 6000m**
- **D2 = 4000m**
- **Orography:**
  - **Gaussian hill**
  - **2000m high, 1000km wide at i=0, j=15**
- **Domain: 12000km × 6000km**
- **Perturbation doubling time is ~30 hours**



# Weak Constraint 4D-Var in a QG model

- **One analysis is produced every 6 hours, irrespective of window length:**

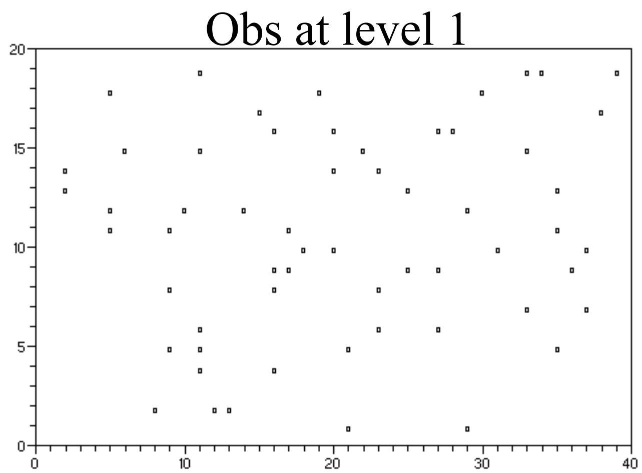


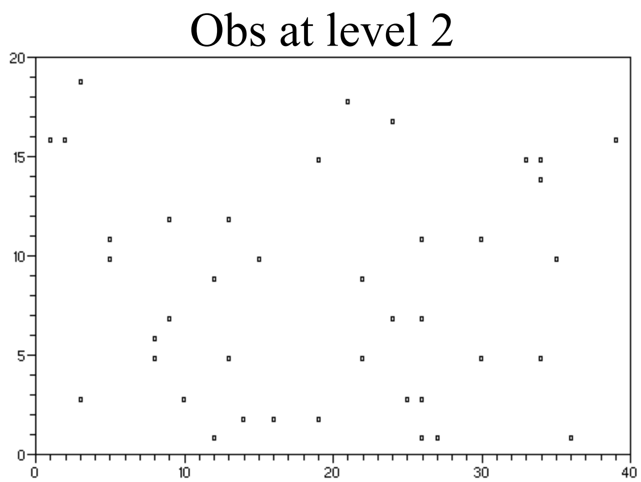

The analysis is incremental, weak-constraint 4D-Var, with a linear inner-loop, and a single iteration of the outer loop.

Inner and outer loop resolutions are identical.

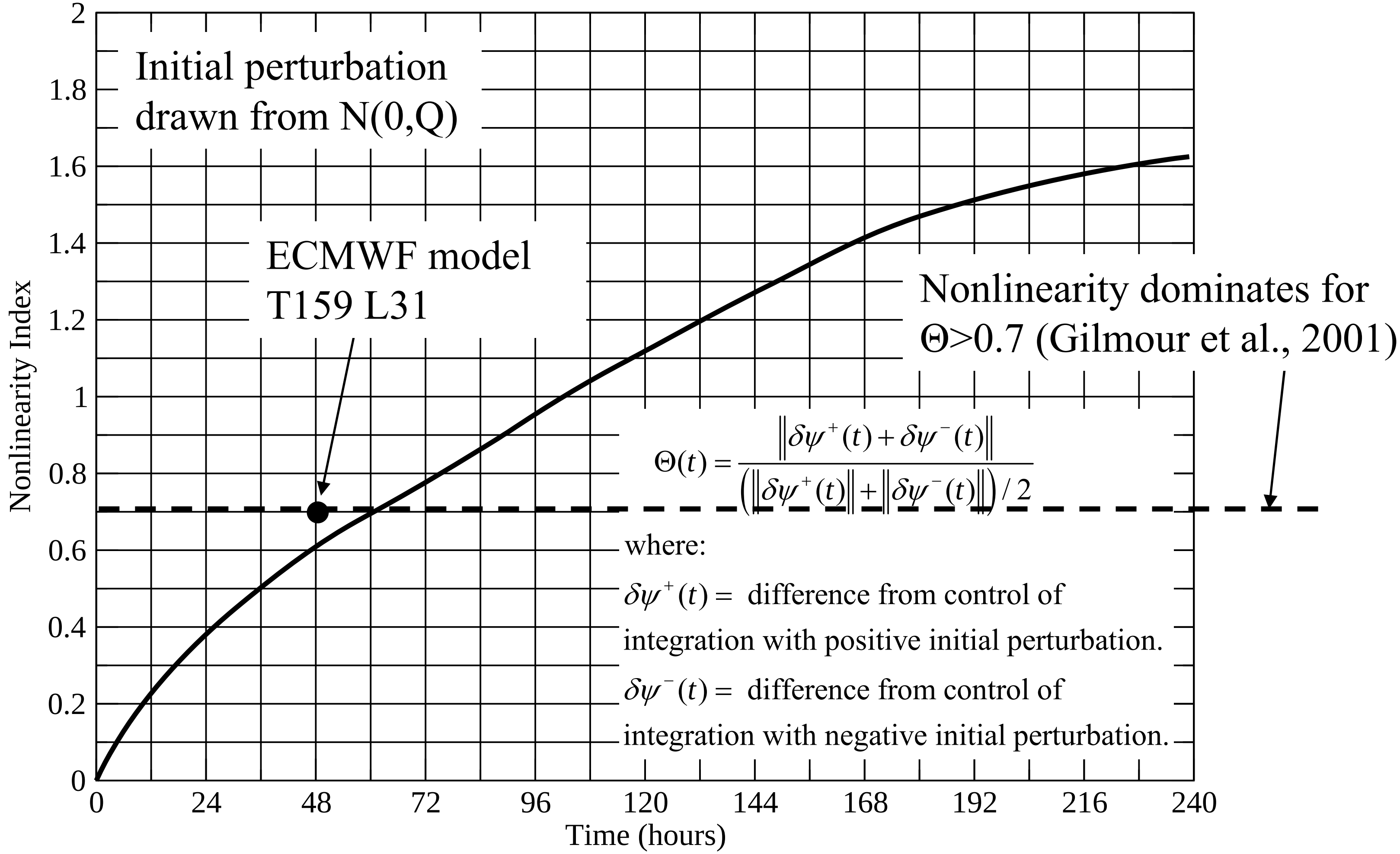
# Weak Constraint 4D-Var in a QG model

- **Observations:**
  - **100 observing points, randomly distributed between levels, and at randomly chosen gridpoints.**
  - **For each observing point, an observation of streamfunction is made every 3 hours.**
  - **Observation error: $\sigma_o$=1.0 (in units of non-dimensional streamfunction)**



W

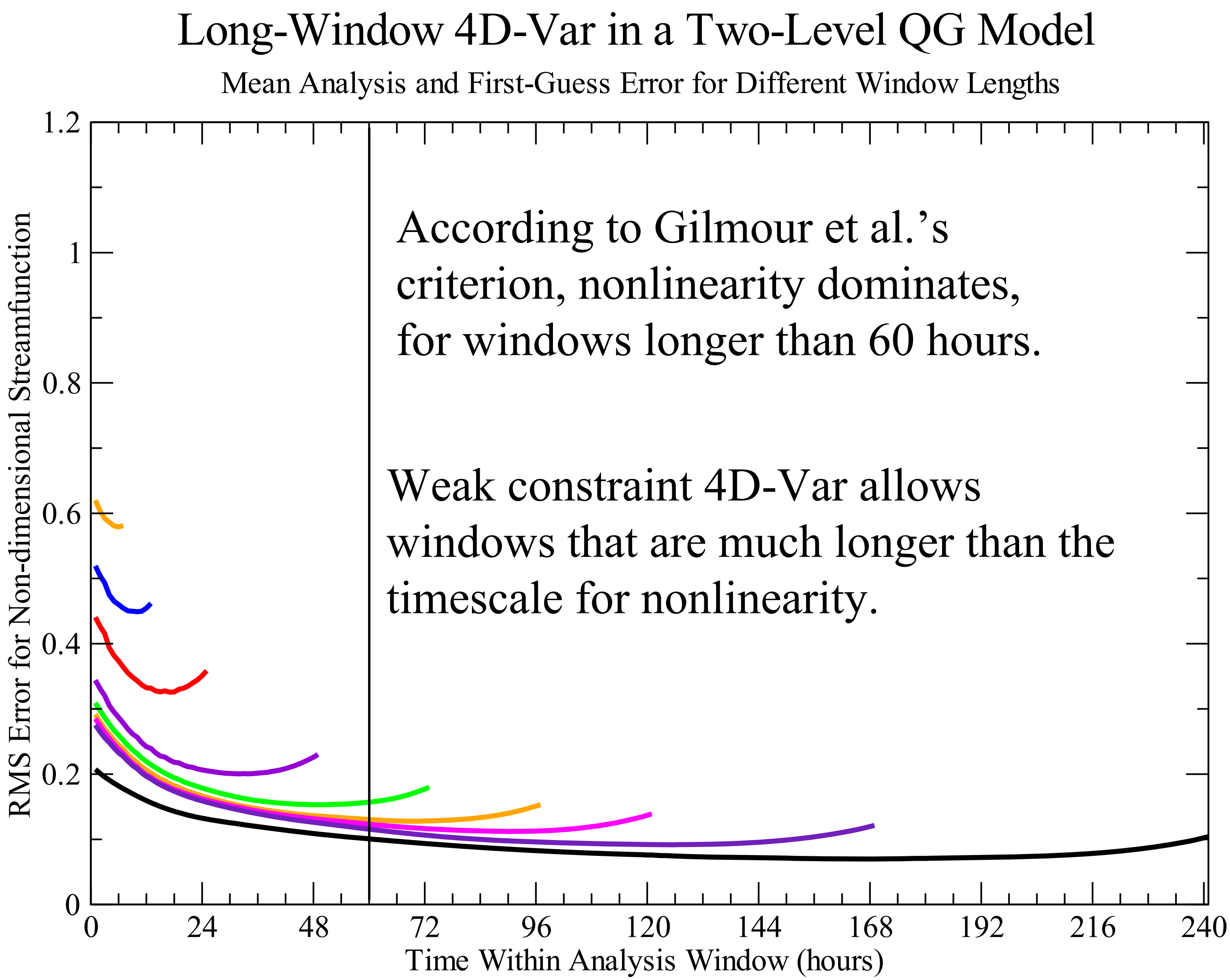


Initial perturbation drawn from N(0,Q)
ECMWF model T159 L31
Nonlinearity dominates for Θ>0.7 (Gilmour et al., 2001)
$\Theta(t) = \frac{\left\| \delta\psi^{+}(t) + \delta\psi^{-}(t) \right\|}{\left( \left\| \delta\psi^{+}(t) \right\| + \left\| \delta\psi^{-}(t) \right\| \right) / 2}$
where:
$\delta\psi^{+}(t) =$ difference from control of integration with positive initial perturbation.
$\delta\psi^{-}(t) =$ difference from control of integration with negative initial perturbation.
Nonlinearity Index
Time (hours)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
0
24
48
72
96
120
144
168
192
216
240



ECMWF

# Weak Constraint 4D-Var in a QG model



# Summary

- **The relationship: J=-log(pdf) makes it straightforward to include a wide range of non-Gaussian effects.**
  - **VarQC**
  - **Non-gaussian bakground errors for humidity,etc.**
  - **nonlinear balances**
  - **nonlinear observation operators (e.g. scatterometer)**
  - **etc.**
- **In weak-constraint 4D-Var, the tangent-linear approximation applies over <u>sub-windows</u>, not over the full analysis window.**
  - **The model appears in $J_q$ as** $M_{t_{k-1} \to t_k}$
- **Window lengths >> nonlinearity time scale are possible.**

# How does 4D-Var handle Nonlinearity and non-Gaussianity?

**Surprisingly Well!**

Thank you for your attention.

Acknowledgements: Christina Tavolato, Elias Holm, Lars Isaksen, Tavolato, Yannick Tremolet