The Sensitivity of Analysis Errors to the Specification of Background Error Covariances

Mike Fisher

ECMWF

σ_b from an Ensemble of Analyses

- **ECMWF's specification of σ_b for dynamical variables is based on a cycling algorithm for vorticity errors:**
  1. **Estimate σ_a from leading eigenvectors of the Hessian**
  2. **Apply a simple error-growth model (Savijarvi, 1995):**

$$\frac{d\sigma}{dt} = (a + b\sigma)\left(1 - \frac{\sigma}{\sigma_\infty}\right)$$

- **There is no flow-dependence in this model**
  - **(a and b are constants, and σ_∞ is specified from climatology)**
- **Flow-dependence for temperature errors is introduced through the action of the balance operators:**
  - **Non-linear balance (linearised about the background)**
  - **Quasi-Geostrophic ω-equation (linearised about the background)**

σ_b from an Ensemble of Analyses

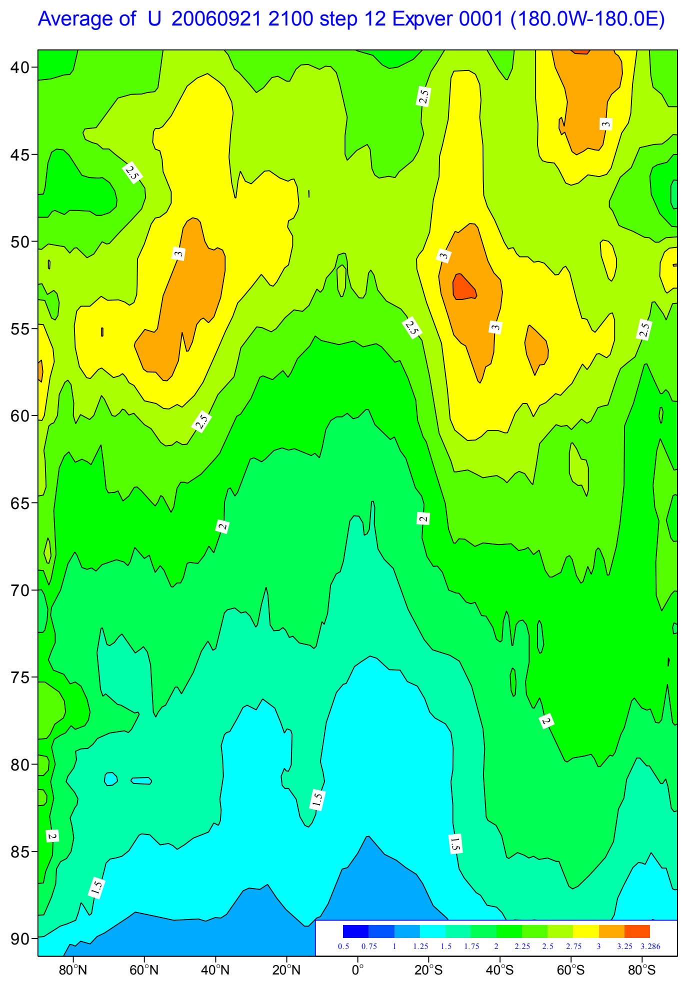
- In an attempt to produce more realistic, flow-dependent fields of σ_b, we have tried estimating σ_b from the spread of a small (10 member) ensemble of 4dVar analyses.
  - Each analysis member is cycled independently from the other members.
  - Observations are perturbed by adding Gaussian noise with the assumed characteristics of observation error.
  - Spatial correlation of observation error is taken into account for cloud-drift winds only.
  - Model error is taken into account through the use of a "stochastic backscatter" scheme.

- Ensemble spread is too small, so we estimate σ_b as:

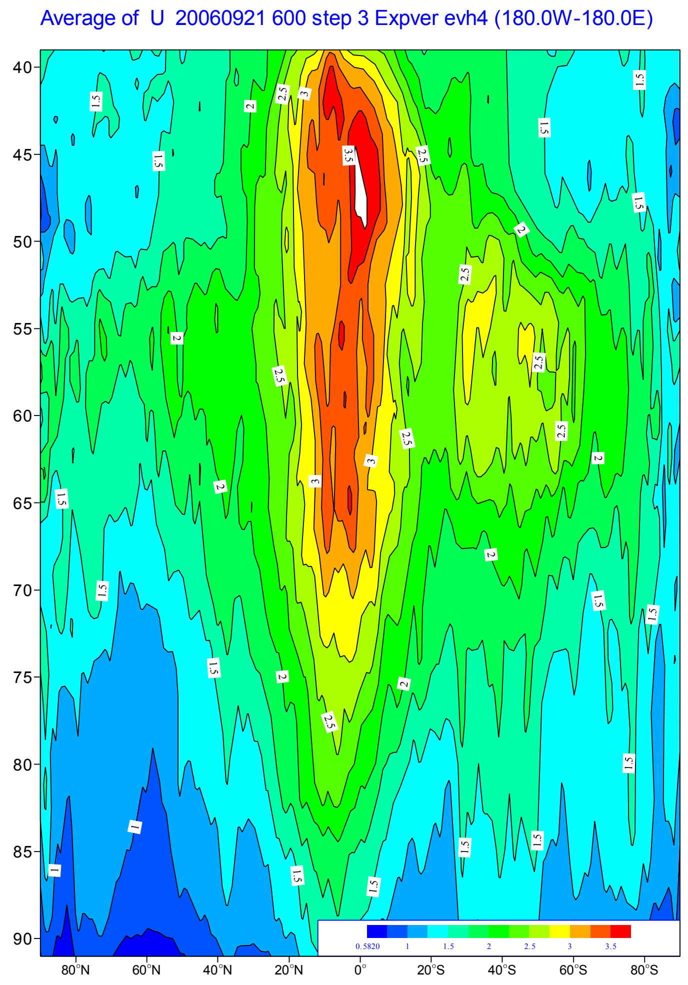
$$\sigma_b = 2 \times \text{Spread}$$

σ_b from an Ensemble of Analyses

U



Current estimate of σ_b



~100hPa

~200hPa

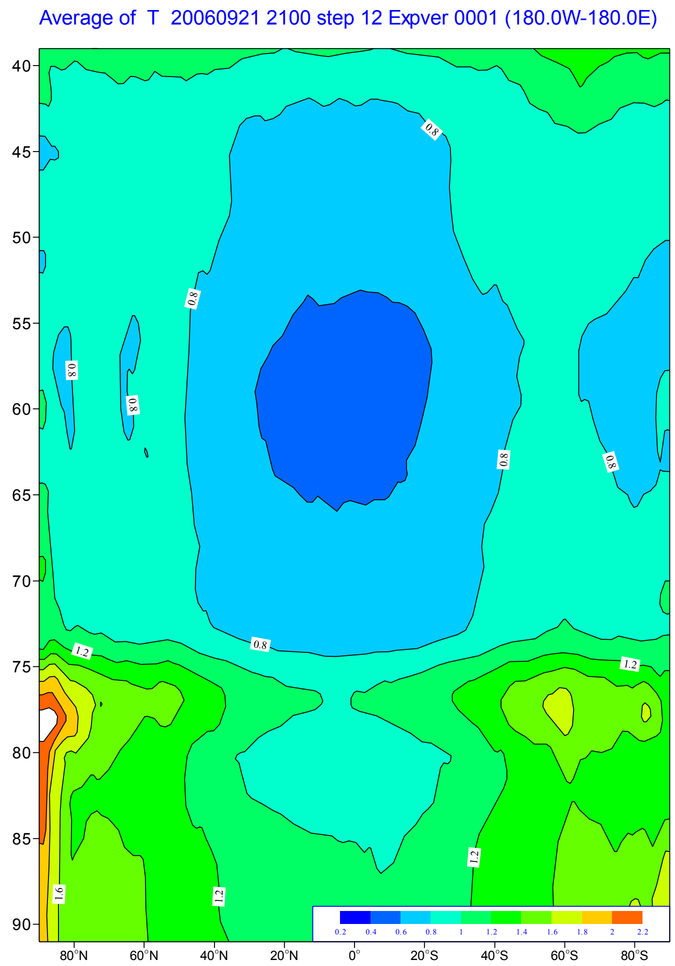
~500hPa

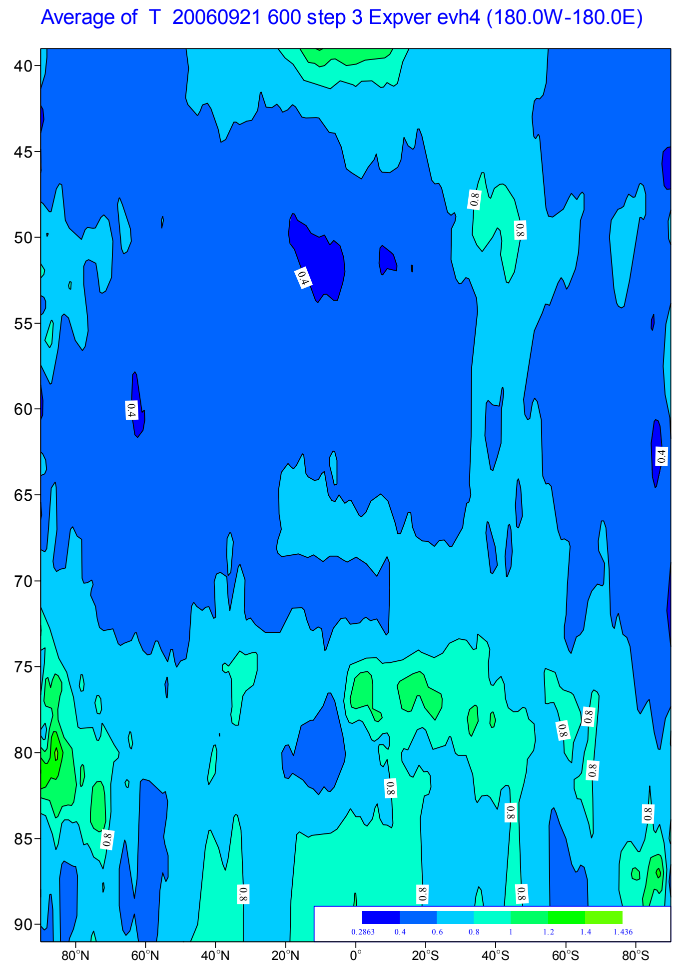
~850hPa

2 × Ensemble Spread

σ_b from an Ensemble of Analyses

T





~100hPa

~200hPa

~500hPa

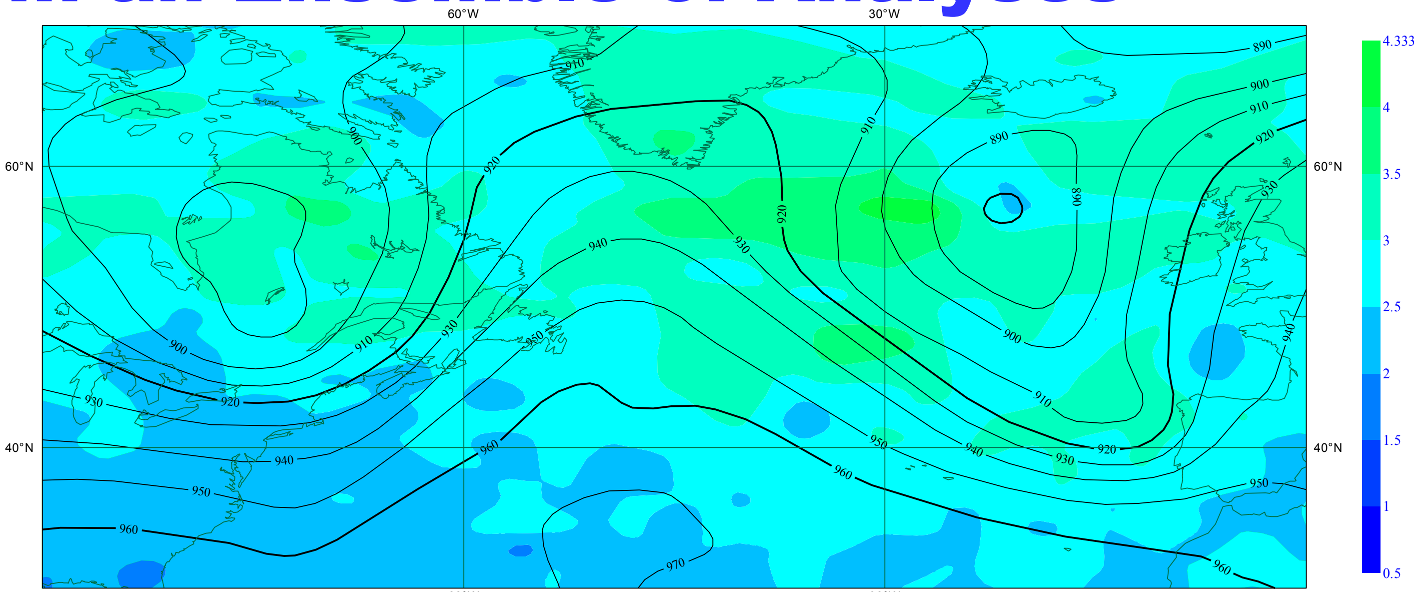
~850hPa

Current estimate of σ_b

$2 \times$ Ensemble Spread

σ_b from an Ensemble of Analyses

Current σ_b

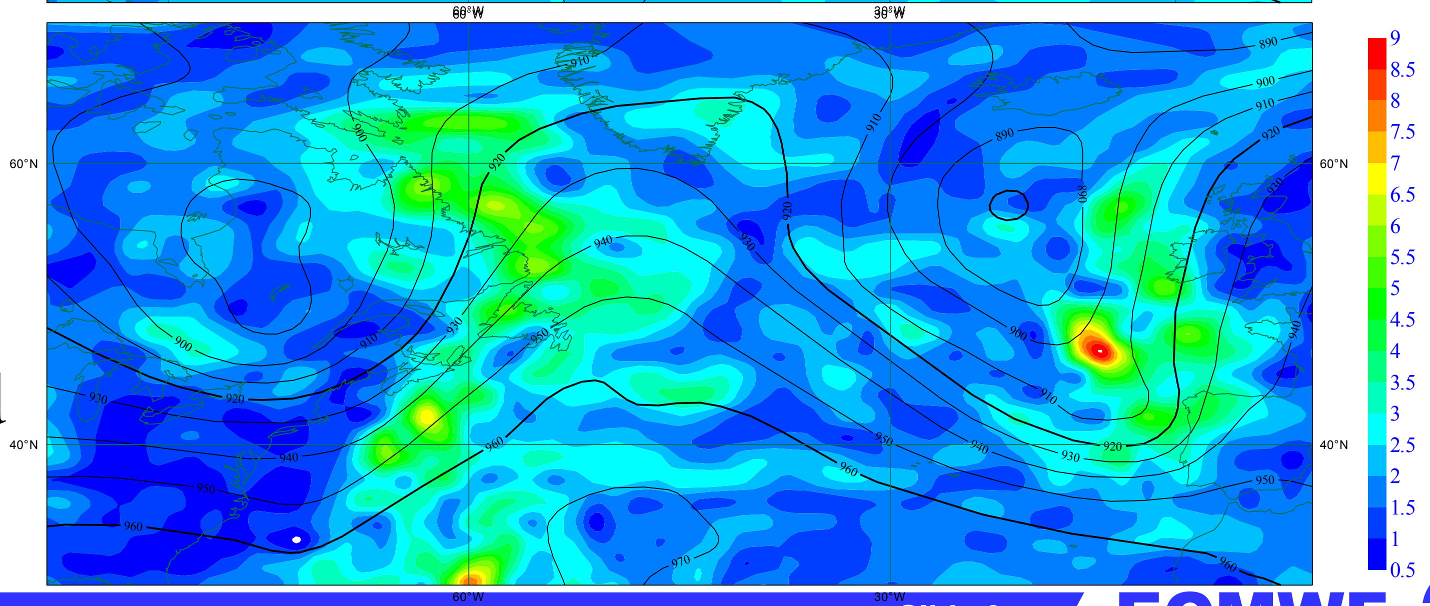


U

Contours: z300

Colours: σ_b

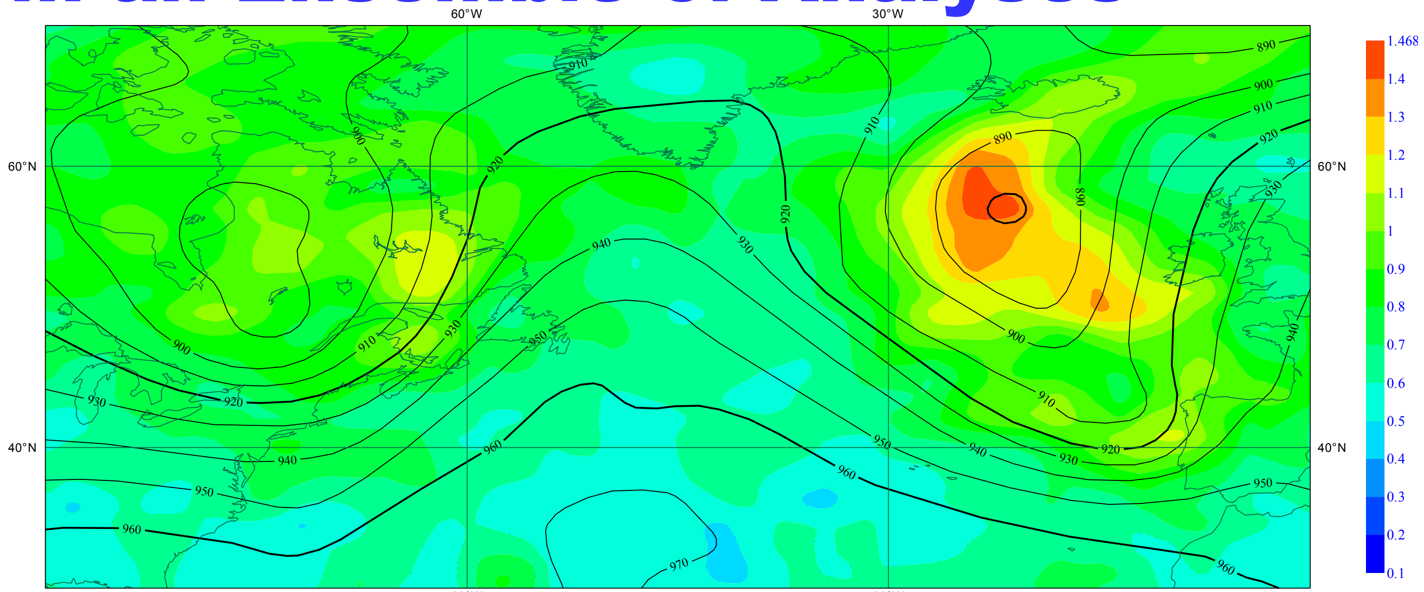
2 × Spread



ECMWF

σ_b from an Ensemble of Analyses

Current σ_b

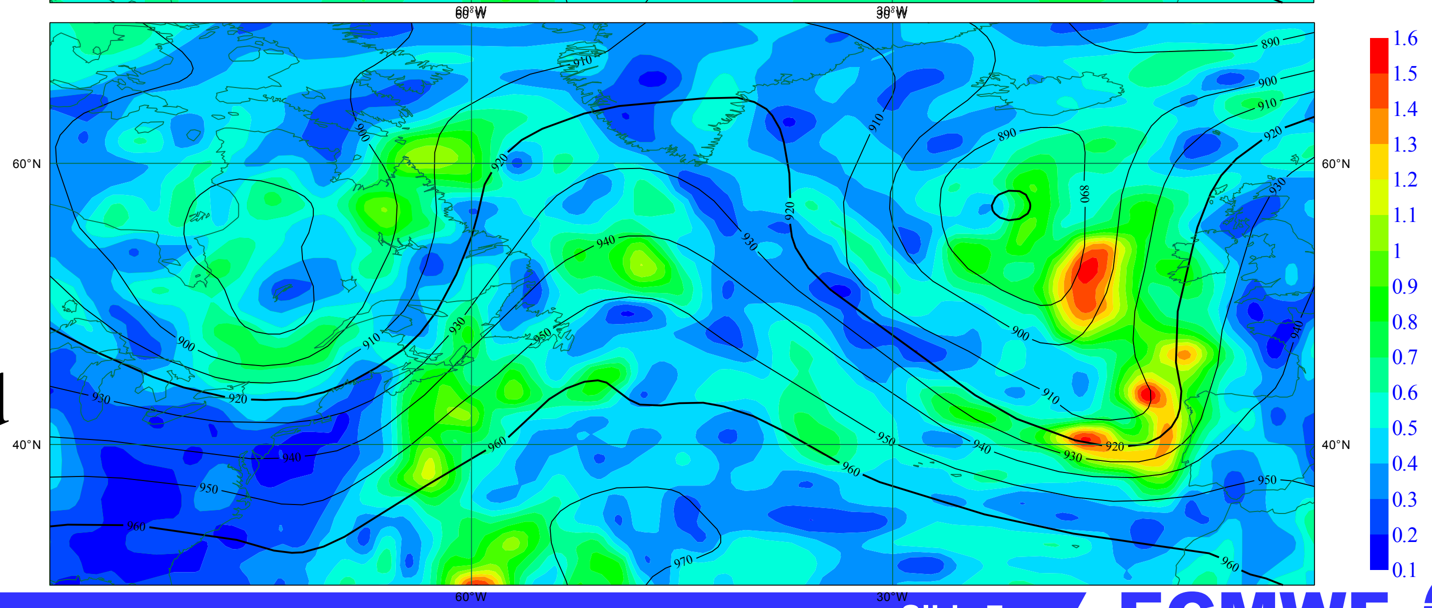


T

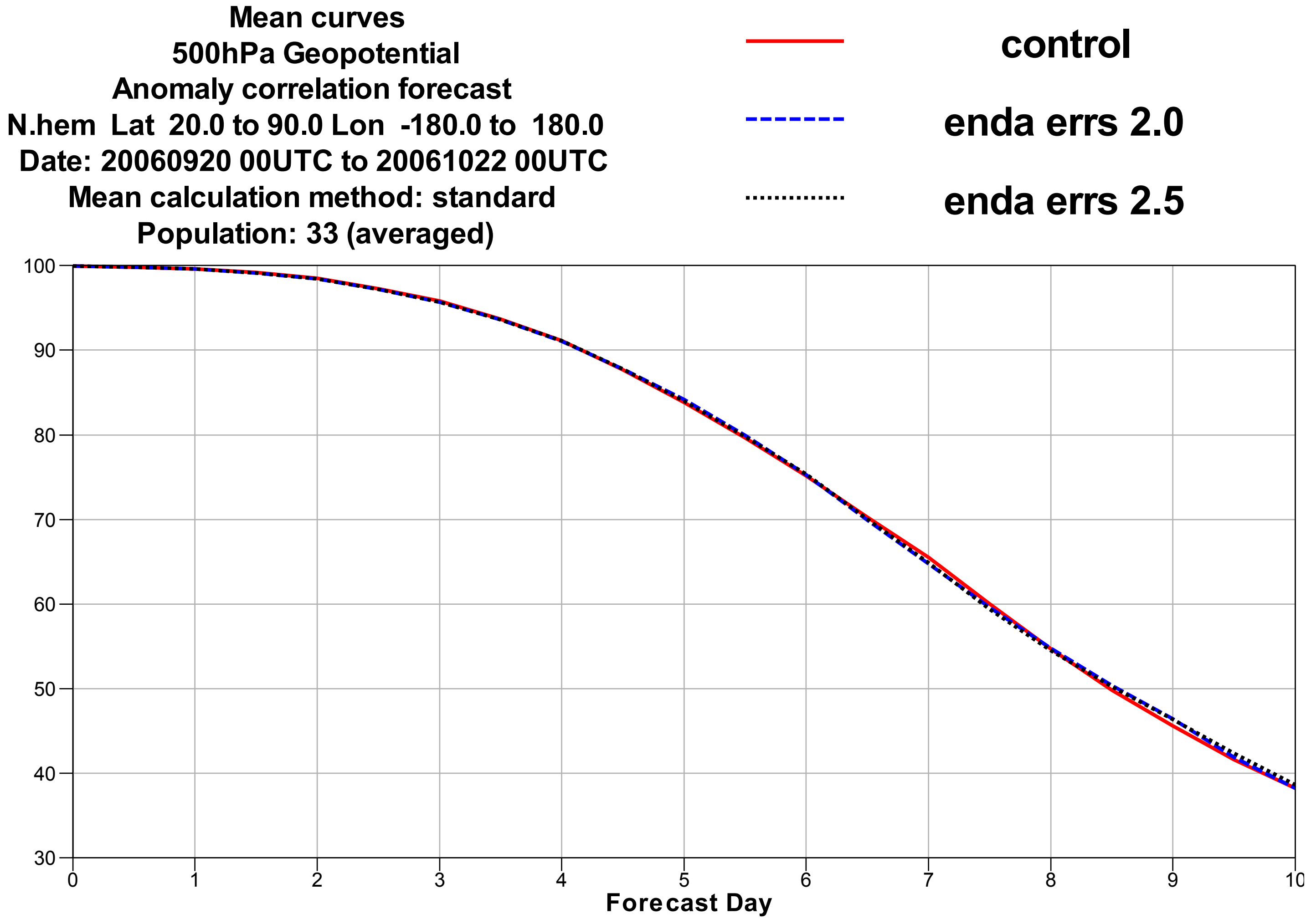
Contours: z300

Colours: σ_b

2 × Spread

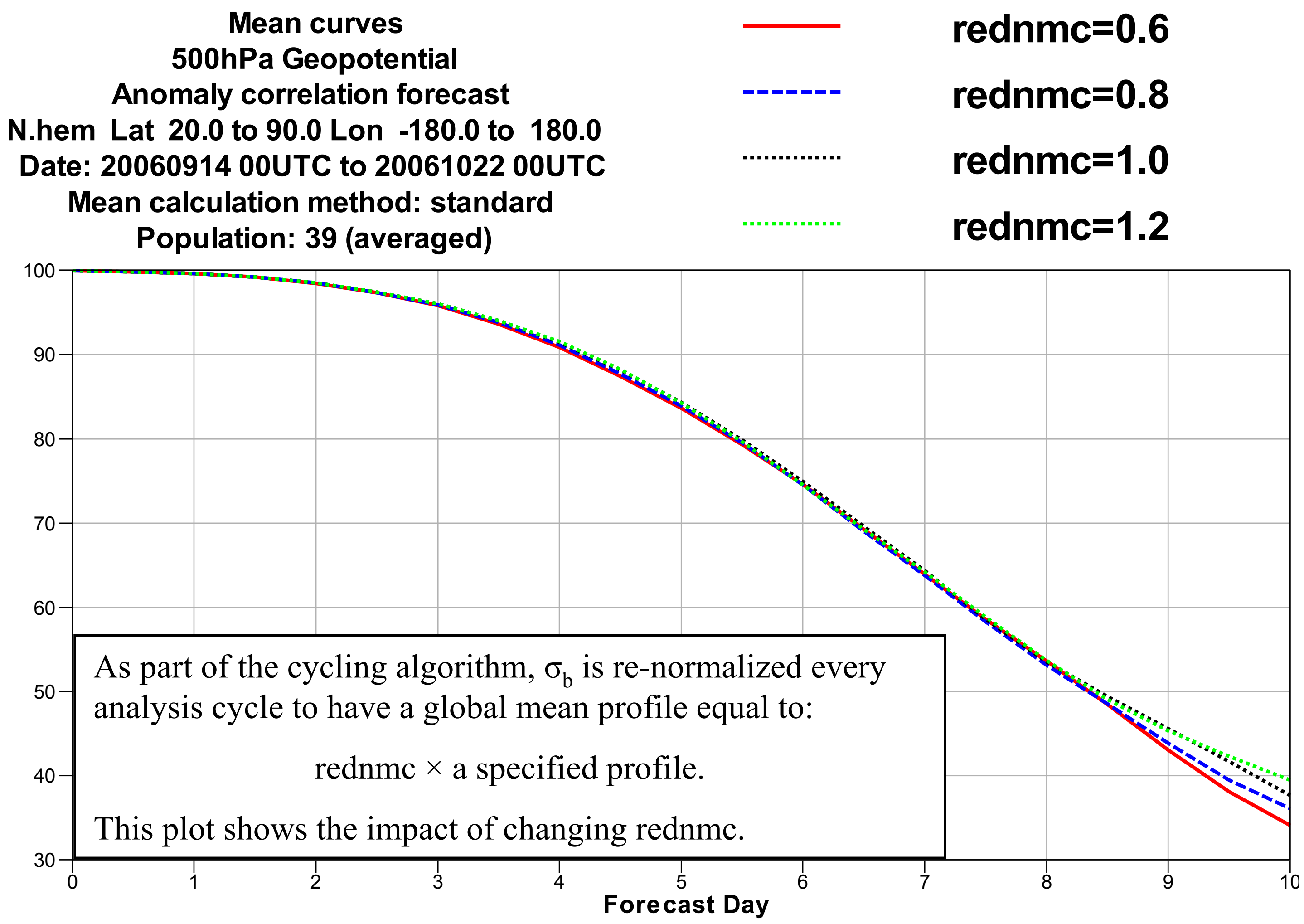


Impact of Flow-Dependent σ_b



ECMWF

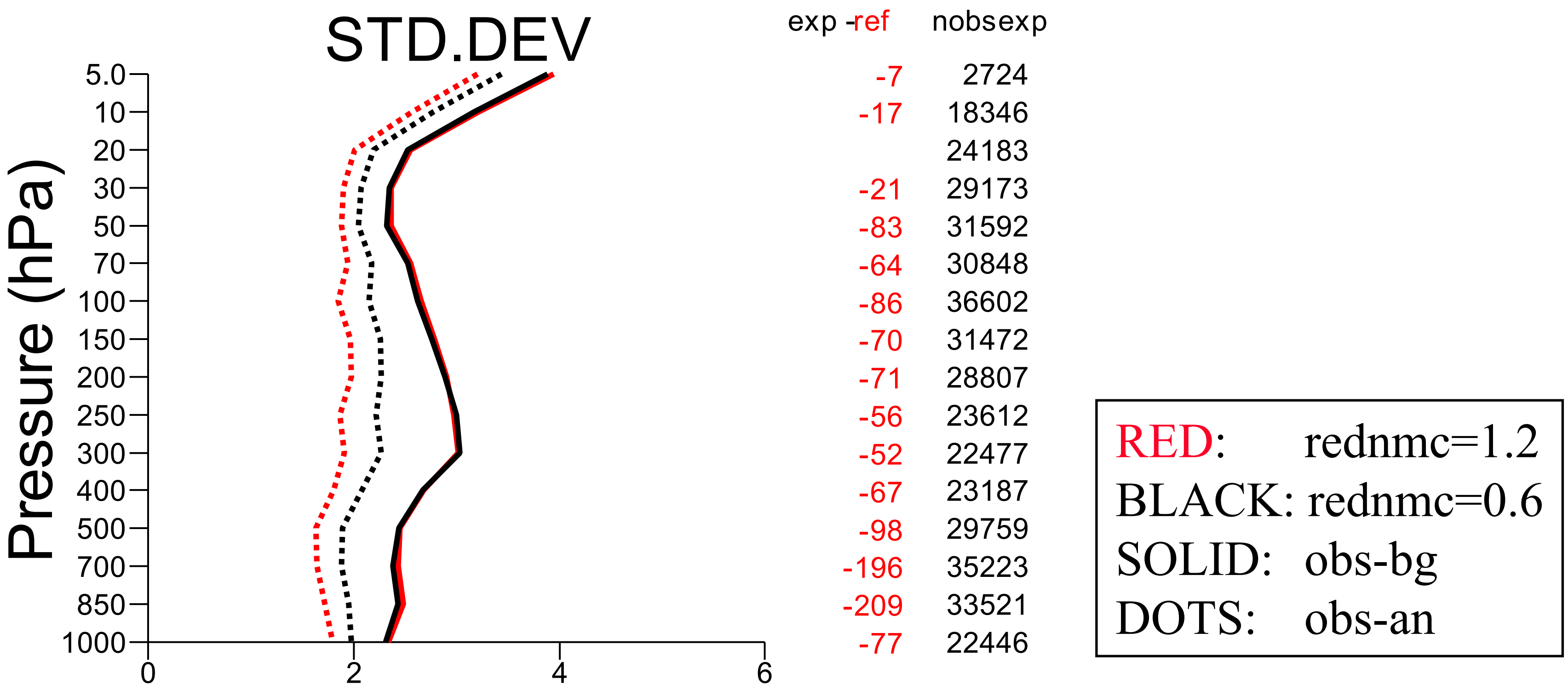
Impact of a Global Scaling of σ_b



ECMWF

Impact of a Global Scaling of σ_b

exp:ew7n /DCDA (black) v. evxw/DCDA 2006092100-2006100412(12)
TEMP-Uwind N.Hemis
used U



ECMWF

Impact of Changing σ_b

- What impact should we expect from changing σ_b?

- Consider a simple scalar analysis (one observation, one background value):

- Daley's book gives the following equation for analysis error in the case that σ_b is misspecified as $\tilde{\sigma}_b$:

$$\sigma_a = \frac{\sqrt{\tilde{\sigma}_b^4 \sigma_o^2 + \sigma_o^4 \sigma_b^2}}{\sigma_o^2 + \tilde{\sigma}_b^2}$$

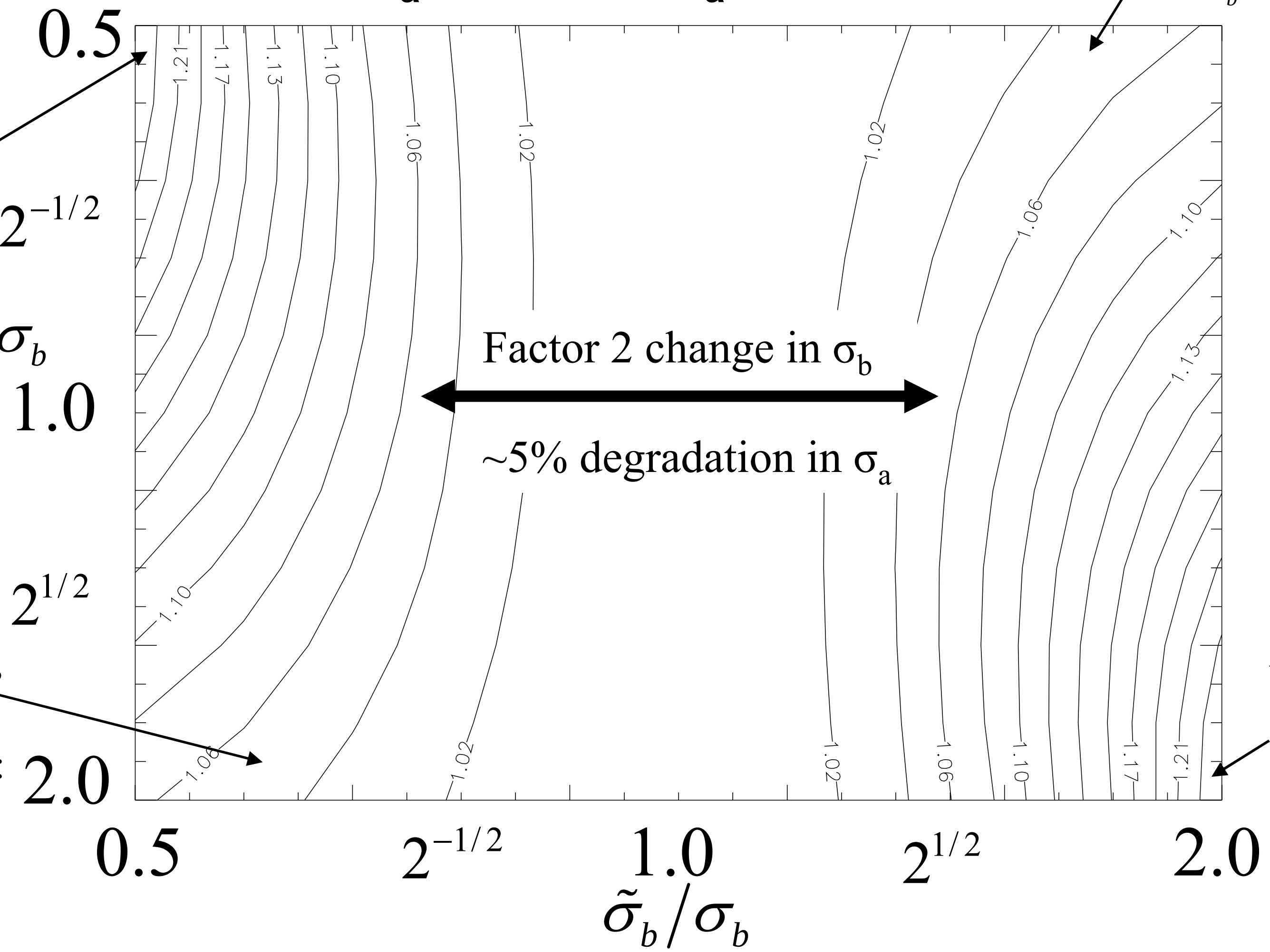
- We can plot this as a function of $\tilde{\sigma}_b / \sigma_b$ and σ_o / σ_b.

Impact of Changing σ_b

This corner: accurate observations are ignored

If $\sigma_o < \sigma_b$, it is best to err on large side: $\tilde{\sigma}_b > \sigma_b$

σ_a(actual) / σ_a(optimal)



If $\sigma_o > \sigma_b$, it is best to err on the small side: $\tilde{\sigma}_b < \sigma_b$

This corner: accurate background is ignored

Impact of Changing σ_b

- For a factor-of-two range of σ_b values, centred around the optimal value, there is only about 5% degradation in analysis error.
- What is the effect on forecast skill of a small increase in analysis error?
  - Analysis A: $\sigma_f(t) = a\,2^{t/T_D}$ where T_D = doubling time.
  - Analysis B: $\sigma_f(t) = b\,2^{t/T_D}$
  - Difference in time to reach the same forecast error: Δt

$$b\,2^{(t+\Delta t)/T_D} = a\,2^{t/T_D}$$

$$\Rightarrow 2^{\Delta t/T_D} = \frac{a}{b}$$

$$\Rightarrow \Delta t = T_D \ln\left(a/b\right)/\ln\left(2\right)$$

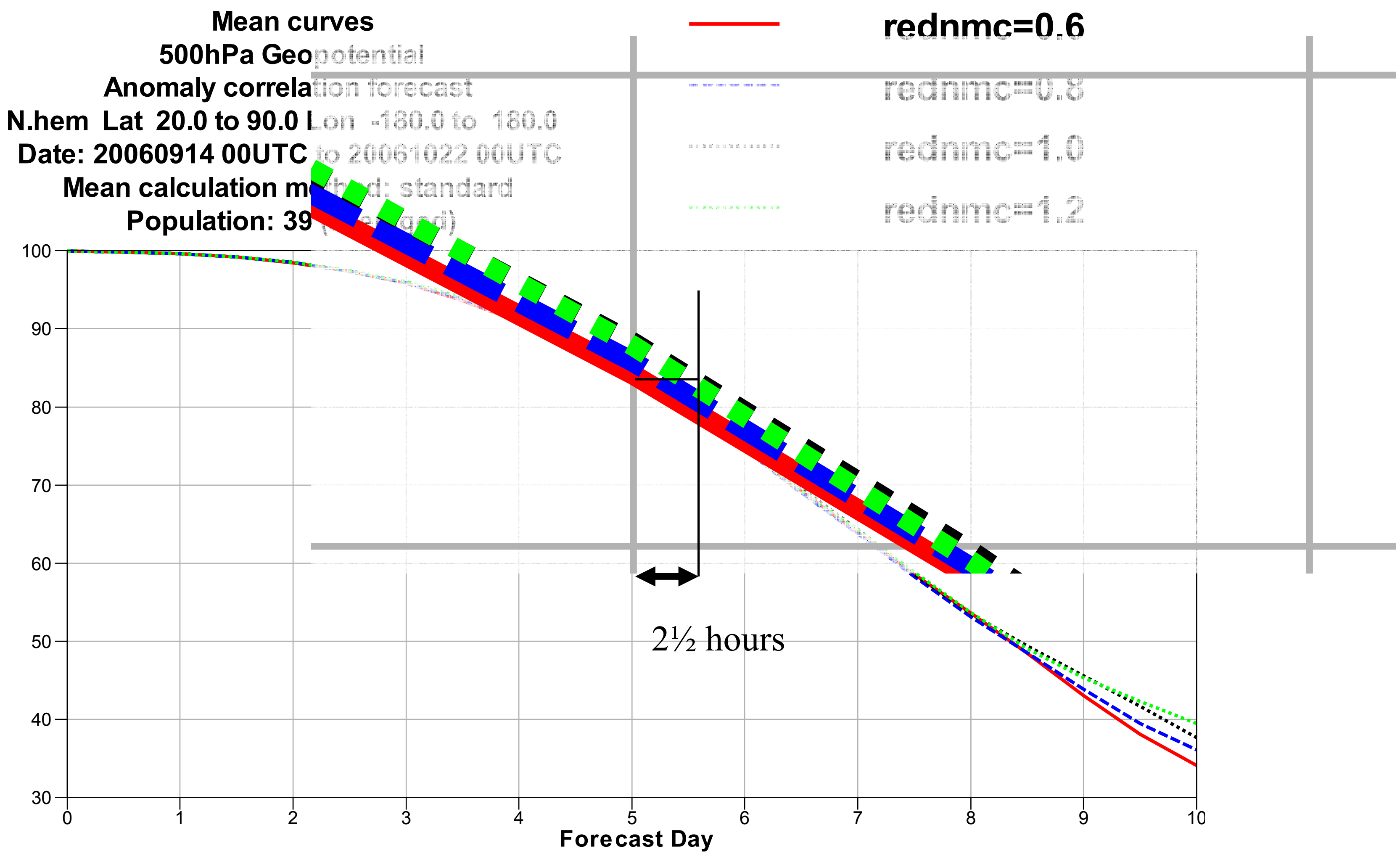
Impact of Changing σ_b

$$\Delta t = T_D \ln\left(a/b\right)/\ln\left(2\right)$$

- For T_D = 36 hours:

a/b	Δt (hours)
1.02	1.03
1.05	2.53
1.10	4.95
1.20	9.47

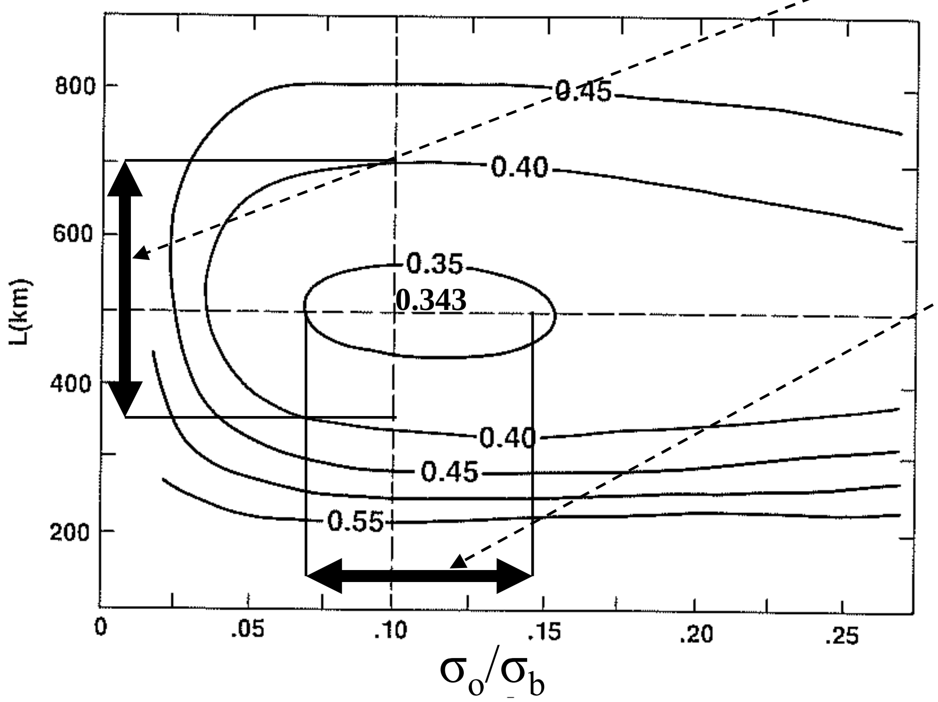
Impact of a Global Scaling of σ_b



ECMWF

Impact of Misspecified statistics

- **Daley's book shows this:**



4.14 Standard deviation of the normalized expected analysis error ε_A as the background error characteristic scale and normalized observation error ε_O are allowed to vary around their correct values, $L = 500$ km and $\varepsilon_O = 0.10$. (After Seaman, *Aus. Met. Mag.* 31: 225, 1983. AGPS Canberra, reproduced by permission of Commonwealth of Australia copyright.)

Analysis degradation is larger (~17%) for a factor of two range of L.

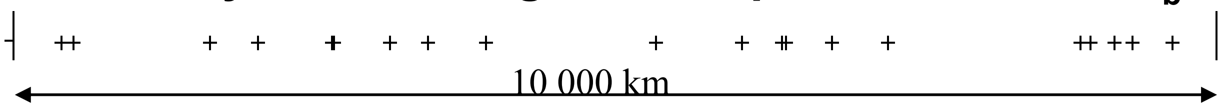
Analysis error remains <2% above its optimal value for a factor of two range of σ_o/σ_b.

NB: Assumes $\sigma_b = 10\times\sigma_o$. This is unrealistic for today's analysis systems, for which $\sigma_b \sim \sigma_o$.

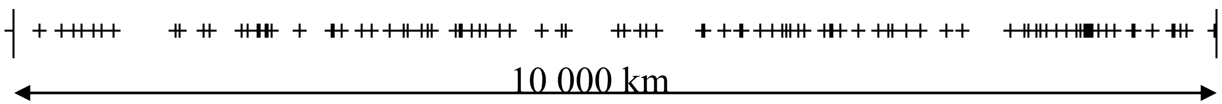
Impact of Misspecified statistics

- Daley's figure 4.14 (Seaman 1983) suggests little sensitivity to σ_b over a wide range of values, and a somewhat larger sensitivity to correlation length-scale.
- But, σ_b/σ_o=10 is unrealistic for today's analysis systems.

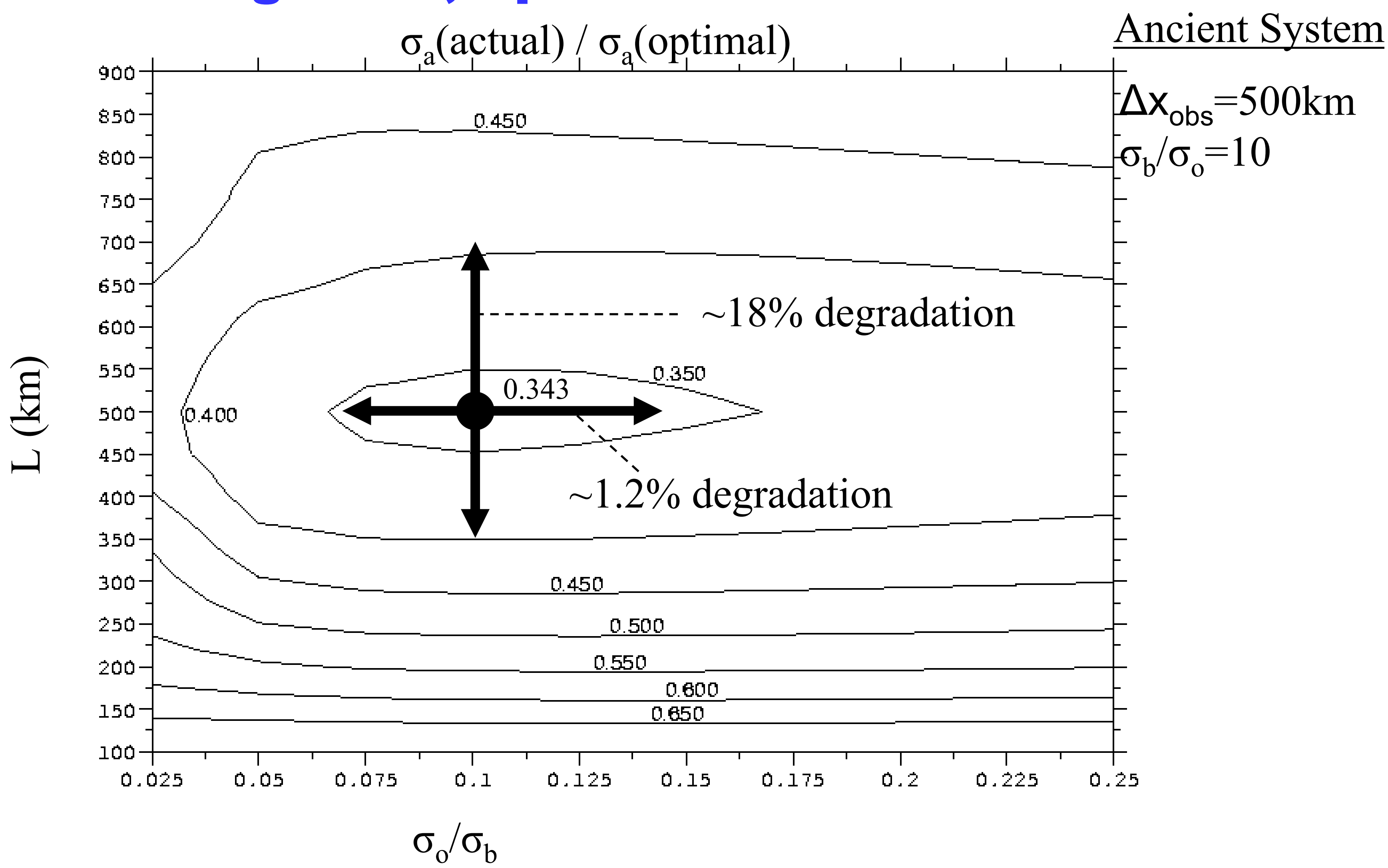
- We consider a simple, one-dimensional analysis:
  - Homogeneous Gaussian background error correlations L=500km.
  - Observations distributed randomly along the line
  - "Ancient" system: average obs. separation 500km, σ_b/σ_o=10



  - "Modern" system: average obs. separation 100km, σ_b/σ_o=1



Impact of Misspecified statistics: Poor Background, Sparse Observations



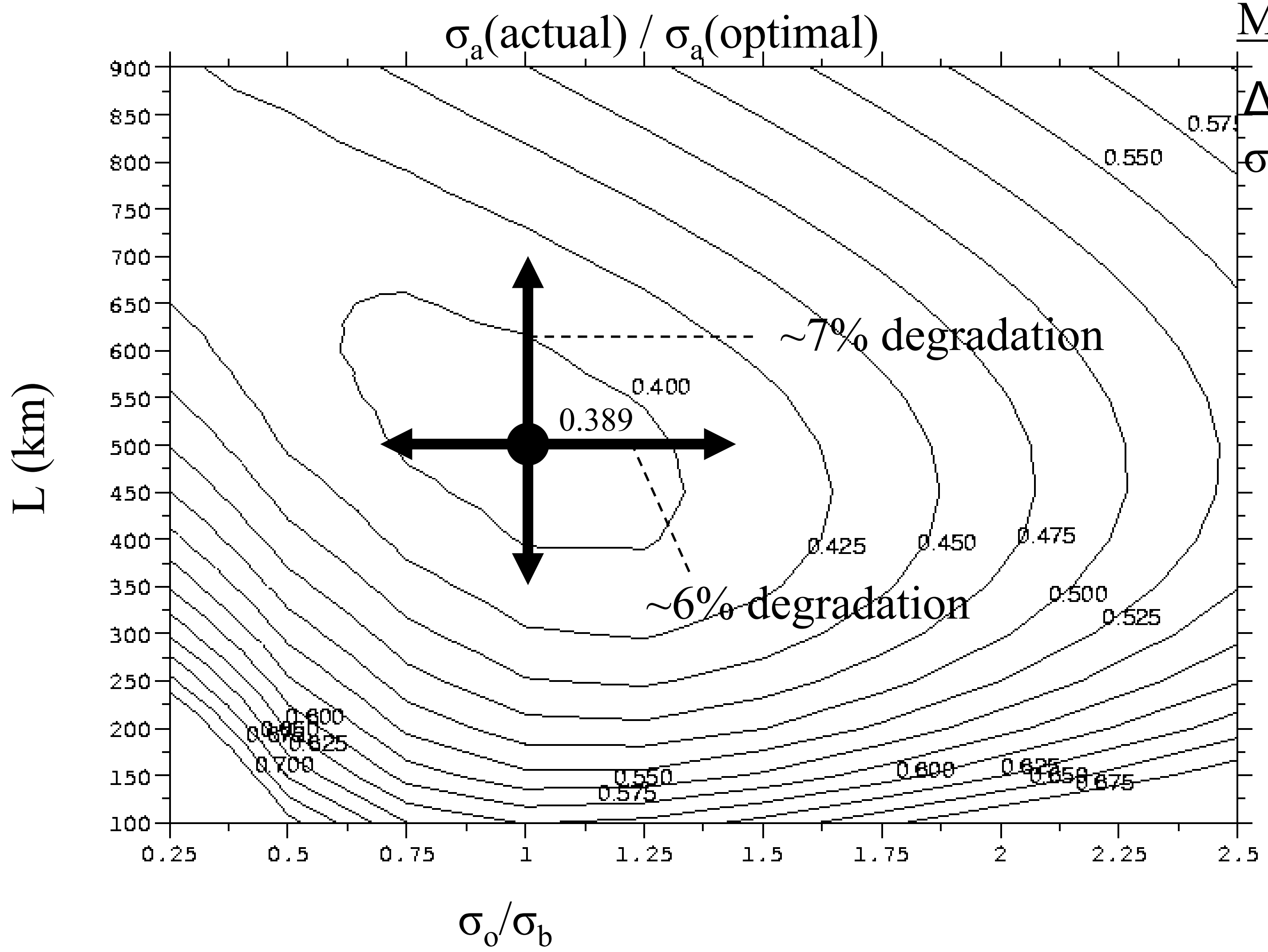
ECMWF

Impact of Misspecified statistics: Accurate Background, Dense Observations

Modern System

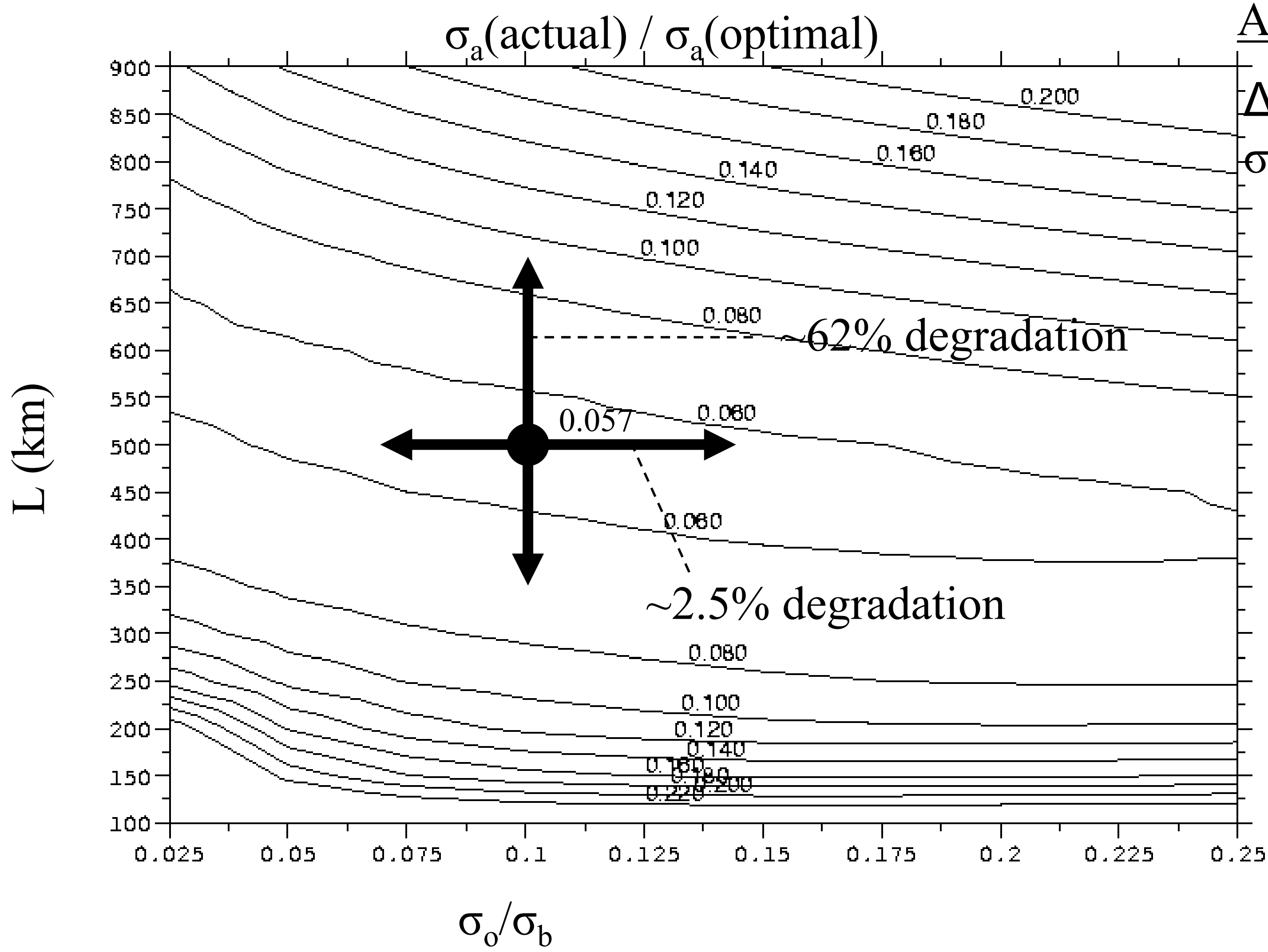
Δx_{obs}=100km

σ_b/σ_o=1



ECMWF

Impact of Misspecified statistics: Poor Background, Dense Observations



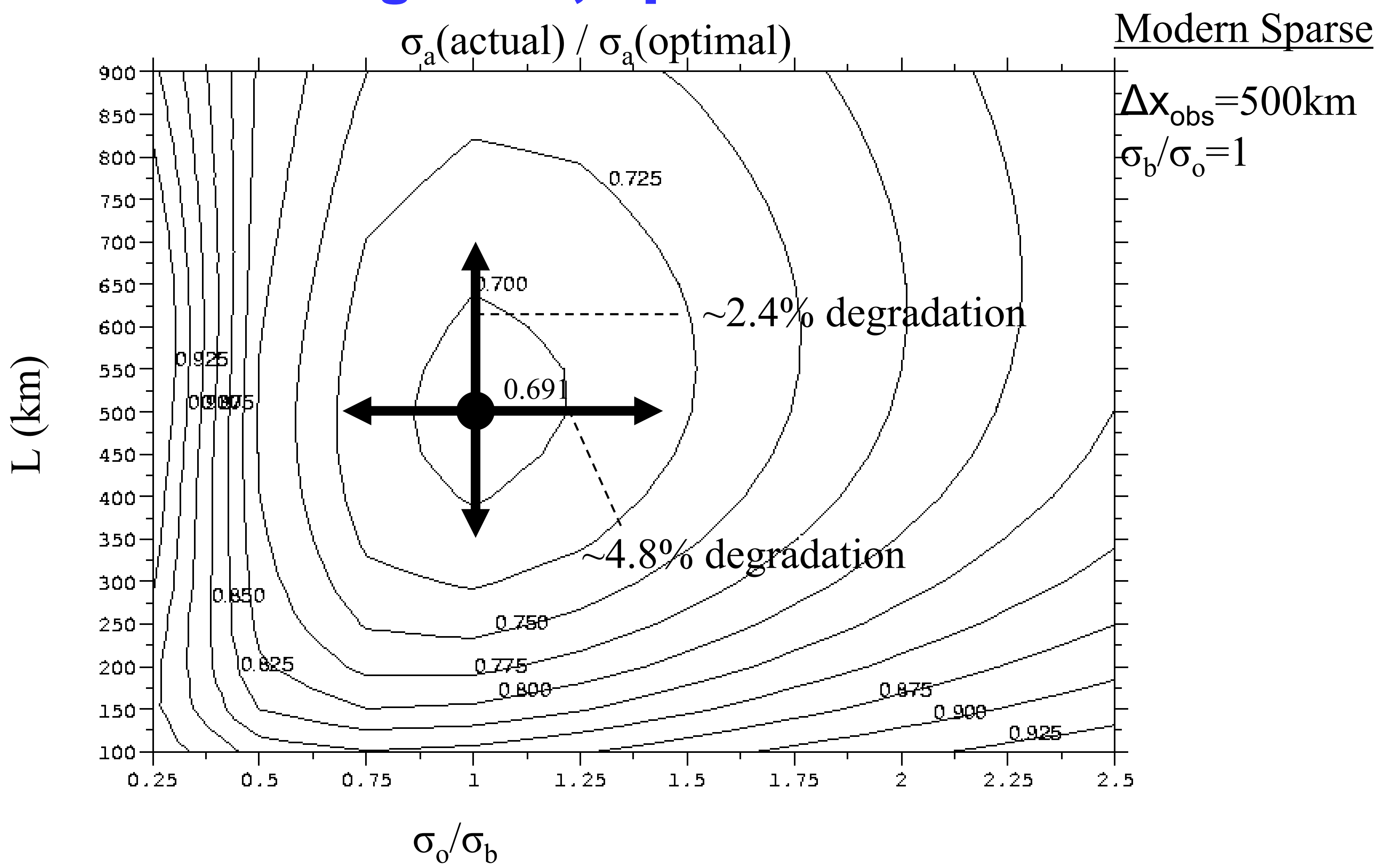
Ancient Dense

Δx_{obs}=100km

σ_b/σ_o=10

NB: σ_a/σ_b=0.057 is not a very realistic scenario

Impact of Misspecified statistics : Accurate Background, Sparse Observations



ECMWF

Impact of Misspecified statistics

	Degradation: misspecified σ_b	Degradation: misspecified length scale
Poor Backgound Sparse Observations	1.2%	18%
Poor Background Dense Observations	2.5%	62%
Accurate Background Sparse Observations	4.8%	2.4%
Accurate Background Dense Observations	6%	7%

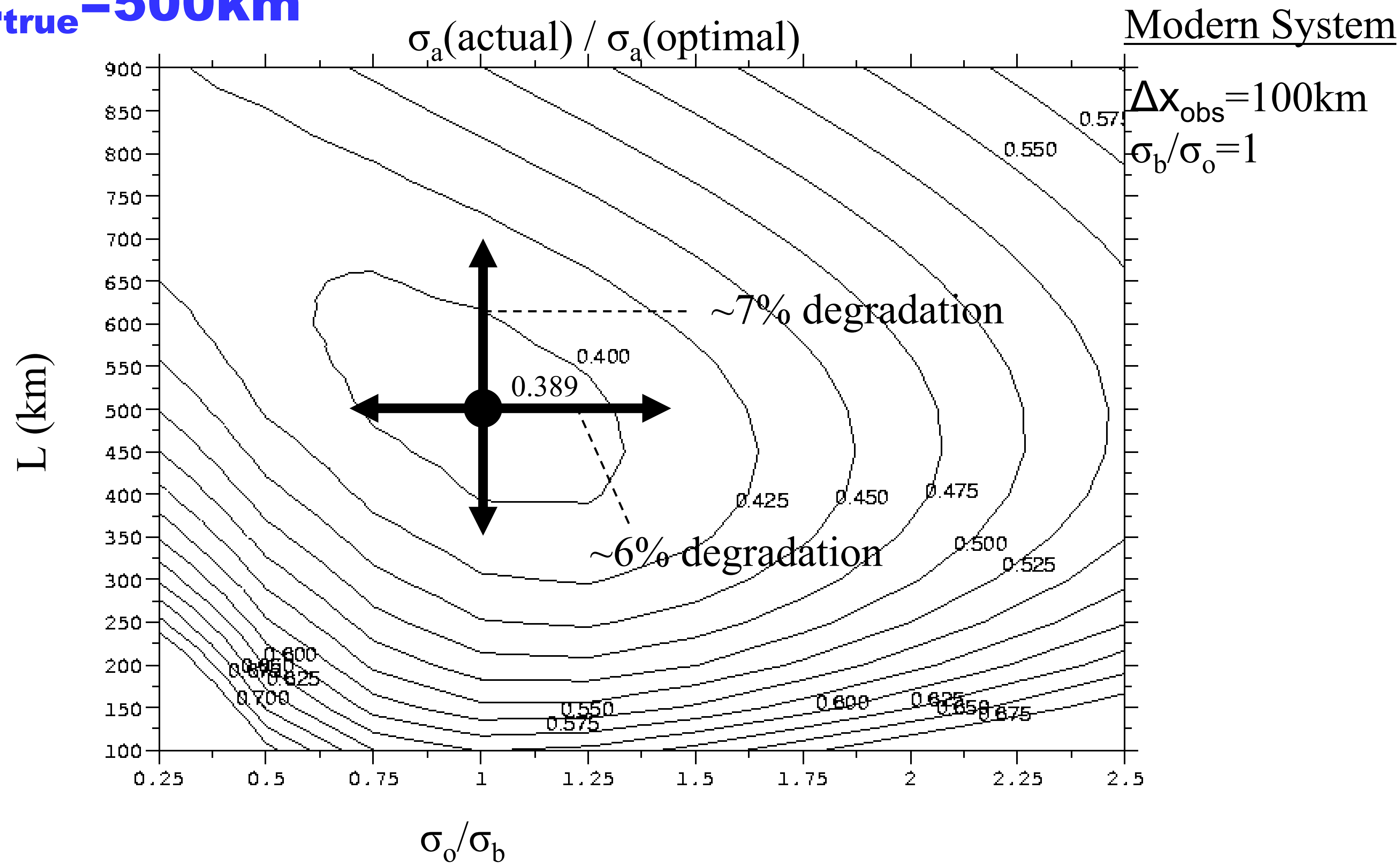
- Accurate Background => Low sensitivity to length scale, increased (but still small) sensitivity to σ_b.

Impact of Misspecified statistics

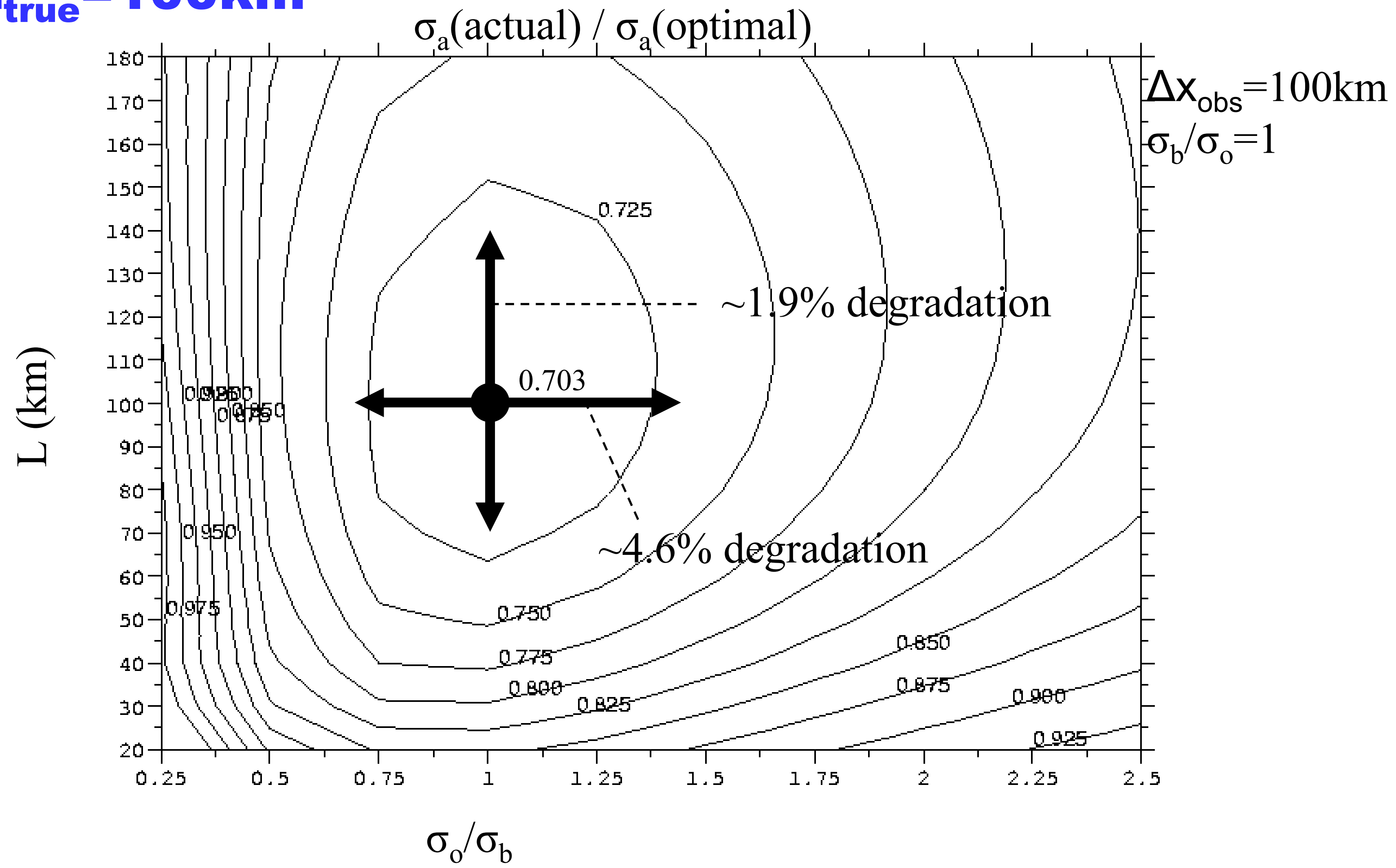
	Degradation: misspecified σ_b	Degradation: misspecified length scale
Poor Backgound Sparse Observations	1.2%	18%
Poor Background Dense Observations	2.5%	62%
Accurate Background Sparse Observations	4.8%	2.4%
Accurate Background Dense Observations	6%	7%

- Dense observations => increased sensitivity to both parameters.

Impact of Misspecified statistics: Accurate Background, Dense Observations, L_{true}=500km

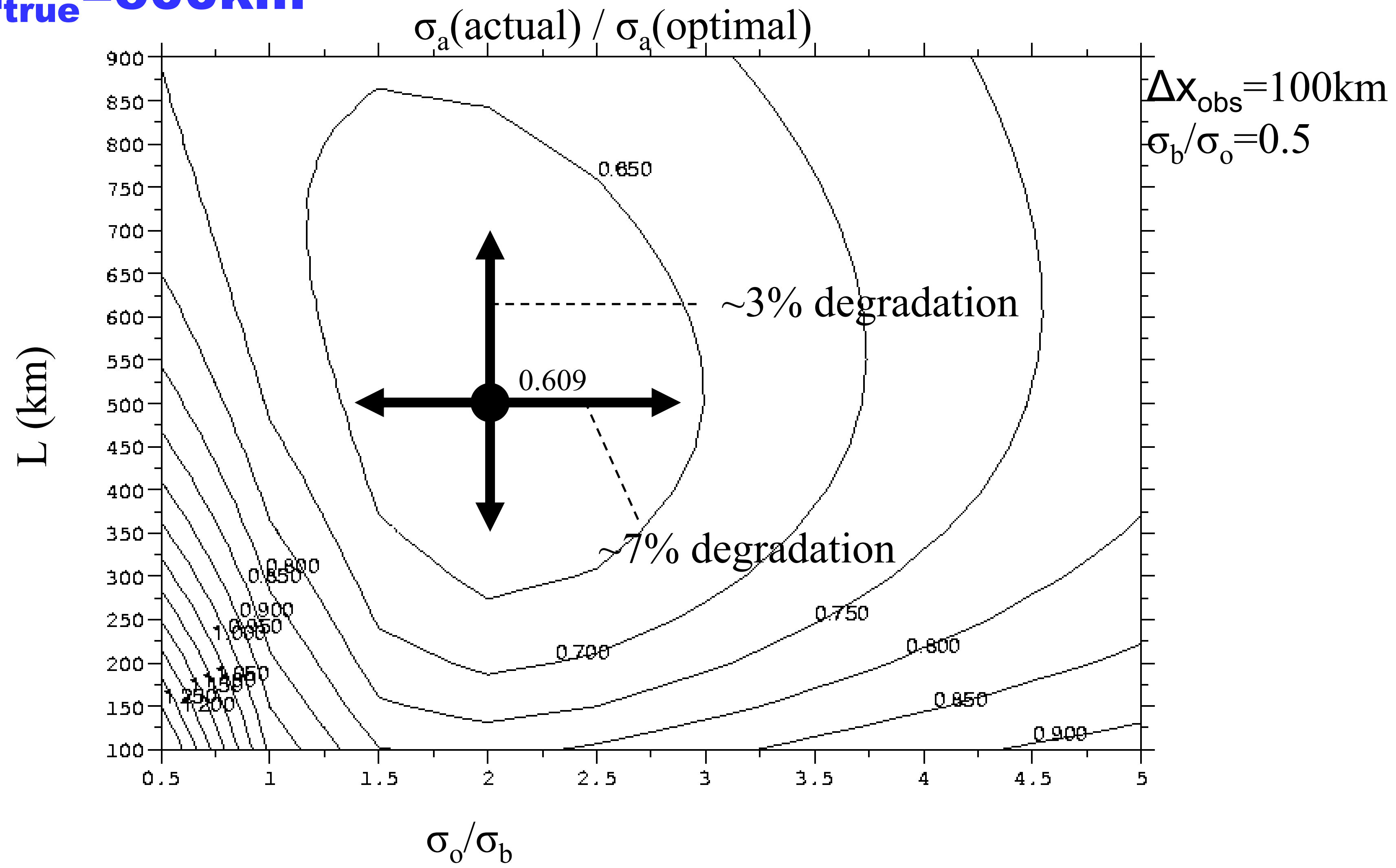


Impact of Misspecified statistics : Accurate Background, Dense Observations, L_{true}=100km



ECMWF

Impact of Misspecified statistics : Very Accurate Background, Dense Obs, L_{true}=500km



ECMWF

Conclusions (1)

- **Modern data assimilation systems are surprisingly insensitive to misspecification of background error standard deviation and length-scale.**
  - <7% analysis degradation over a factor-of-two range for both parameters
  - This translates to <3½ hours loss of forecast skill.
- **Of course, we should not be too dismissive of about such small changes in skill. Significant improvements in forecast skill have been achieved over the last several years largely through the accumulation of many small improvements of this order.**

- **When $\sigma_b < \sigma_o$, it is better to underestimate σ_b than to overestimate it.**

Conclusions (2)

- The ability to misspecify σ_b, without significant detriment to the quality of the analysis, may be useful:
  - Overestimating σ_b could improve the analysis of extreme events, without excessively degrading mean scores.
  - Using too-large σ_b in an ensemble of analyses may be a useful way to increase ensemble spread.
- Ultimately, an optimal analysis requires us to specify the optimal σ_b. Ensembles seem to be our best bet. But, we should not be surprised if improvements are not dramatic.
- There are plenty of other good arguments for running ensembles of analyses:
  - To provide perturbations for use in ensemble prediction
  - To provide good flow-dependent estimates of analysis quality
  - To improve quality-control decisions near tropical cyclones