Issues in 4D-Var for NWP

A brief presentation of research issues designed to stimulate discussion at Durham Symposium on Maths of DA. Aug 2011. *Andrew Lorenc*

Content

- NWP context
- **B** modelling assumptions
- Long-window 4D-Var
- Conditioning of long-window 4D−Var
- Long-window 4D-Var for a chaotic model
- Nonlinearity – benefitting from the attractor
- Fitting models of model error
- Scalability – exploiting massively parallel computers

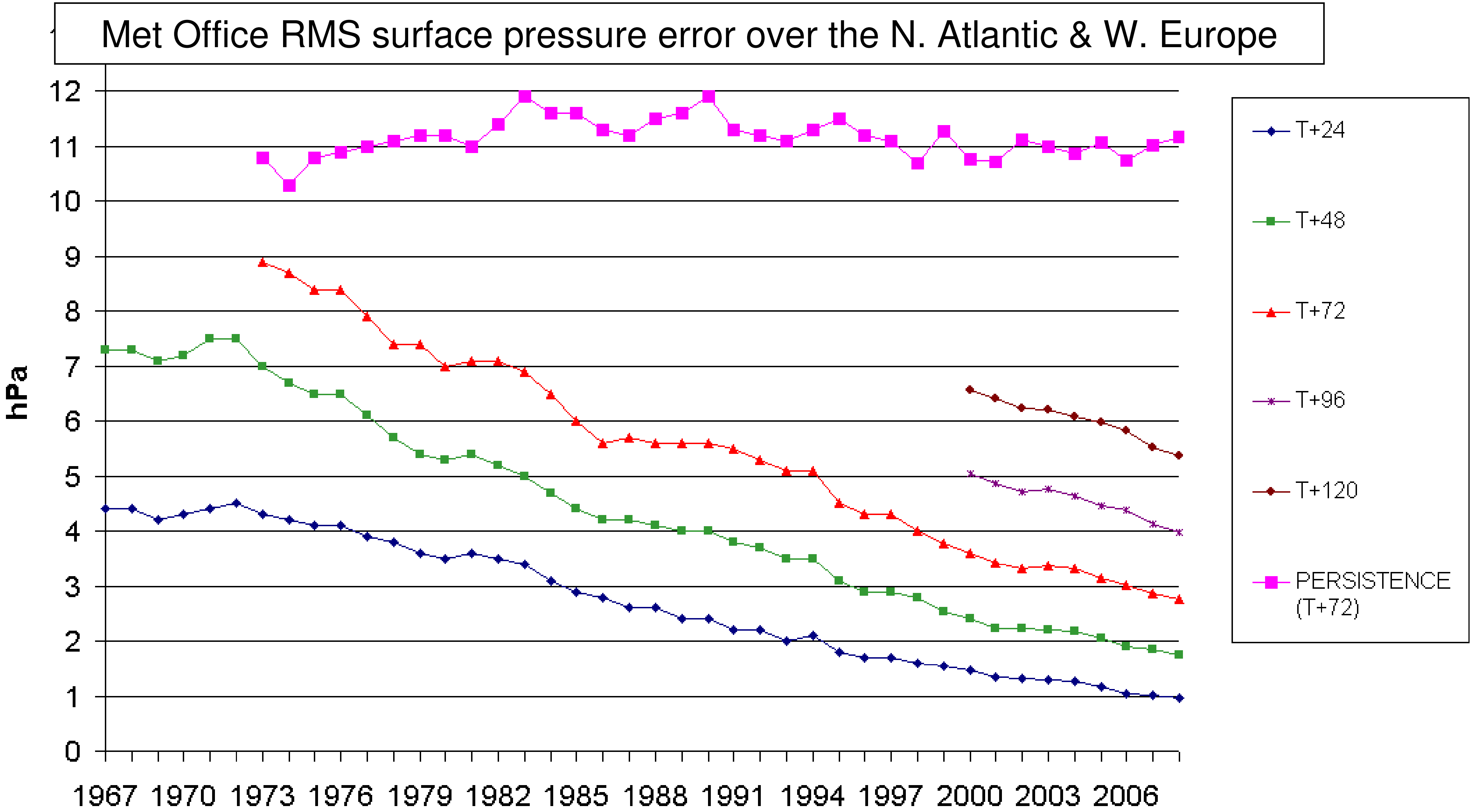
Historical Background:
What has been important for getting the best NWP forecast? *(over last 3 decades)*

NWP systems are improving by 1 day of predictive skill per decade. This has been due to:

1. *Model improvements, especially resolution.*
2. *Careful use of forecast & observations, allowing for their information content and errors.*
   Achieved by variational assimilation e.g. of satellite radiances.
3. *Advanced assimilation using forecast model: 4D-Var.*
4. *Better observations.*

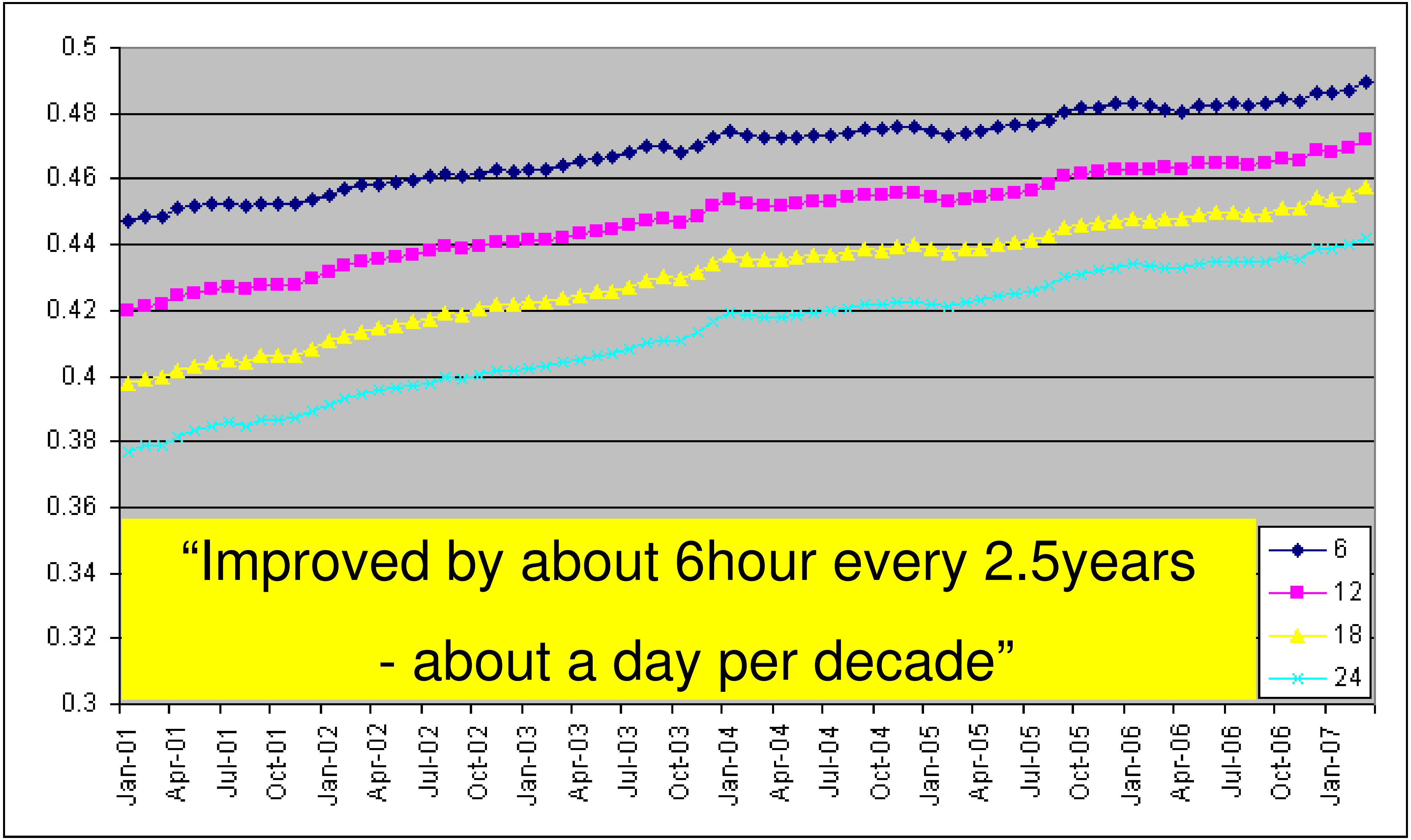
Performance Improvements

"Improved by about a day per decade"



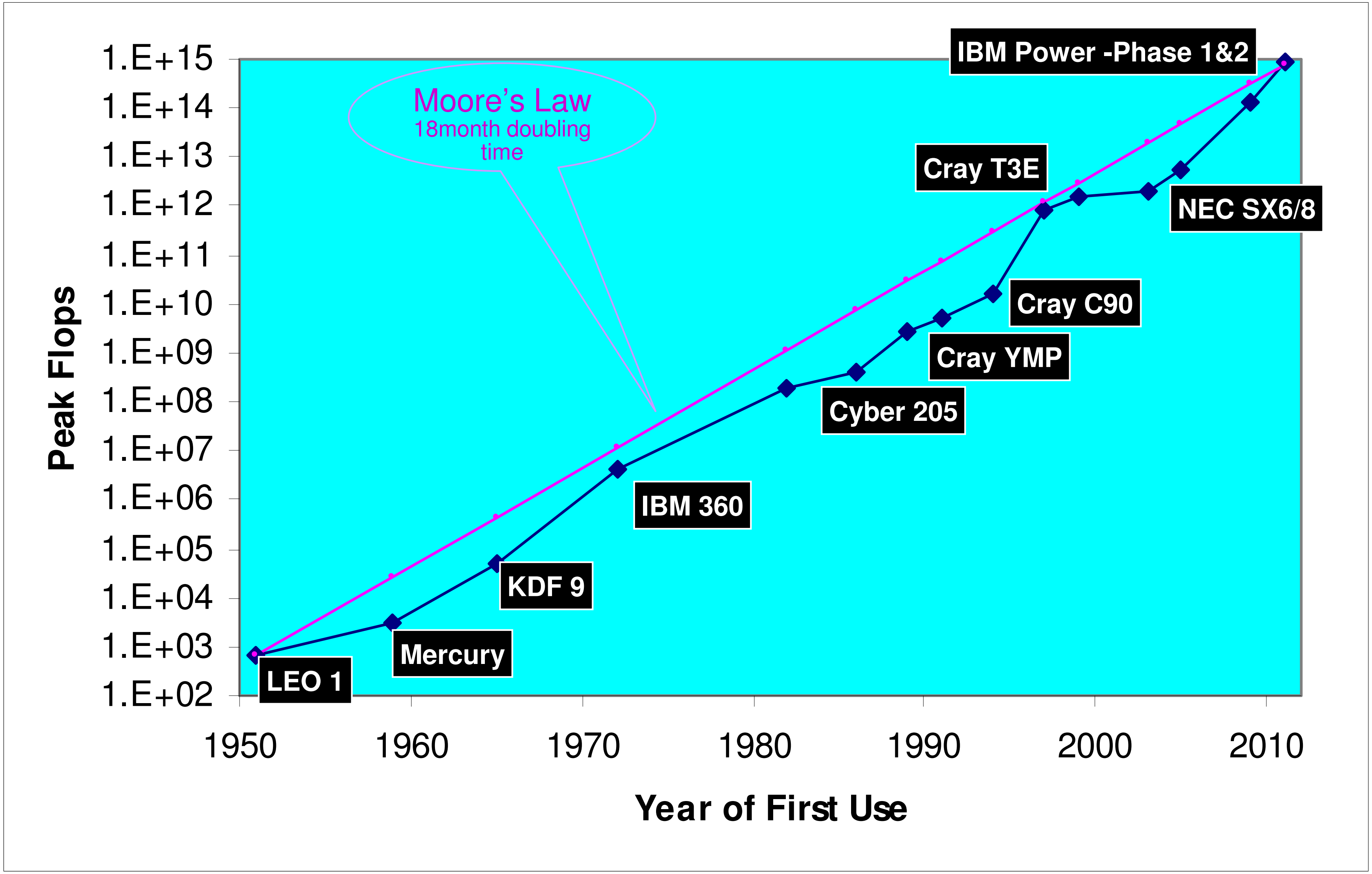
UK Index Improvement:
skill scores vs UK SYNOPS for T wind ppn cloud visibility

Met Office



60 Years of Met Office Computers

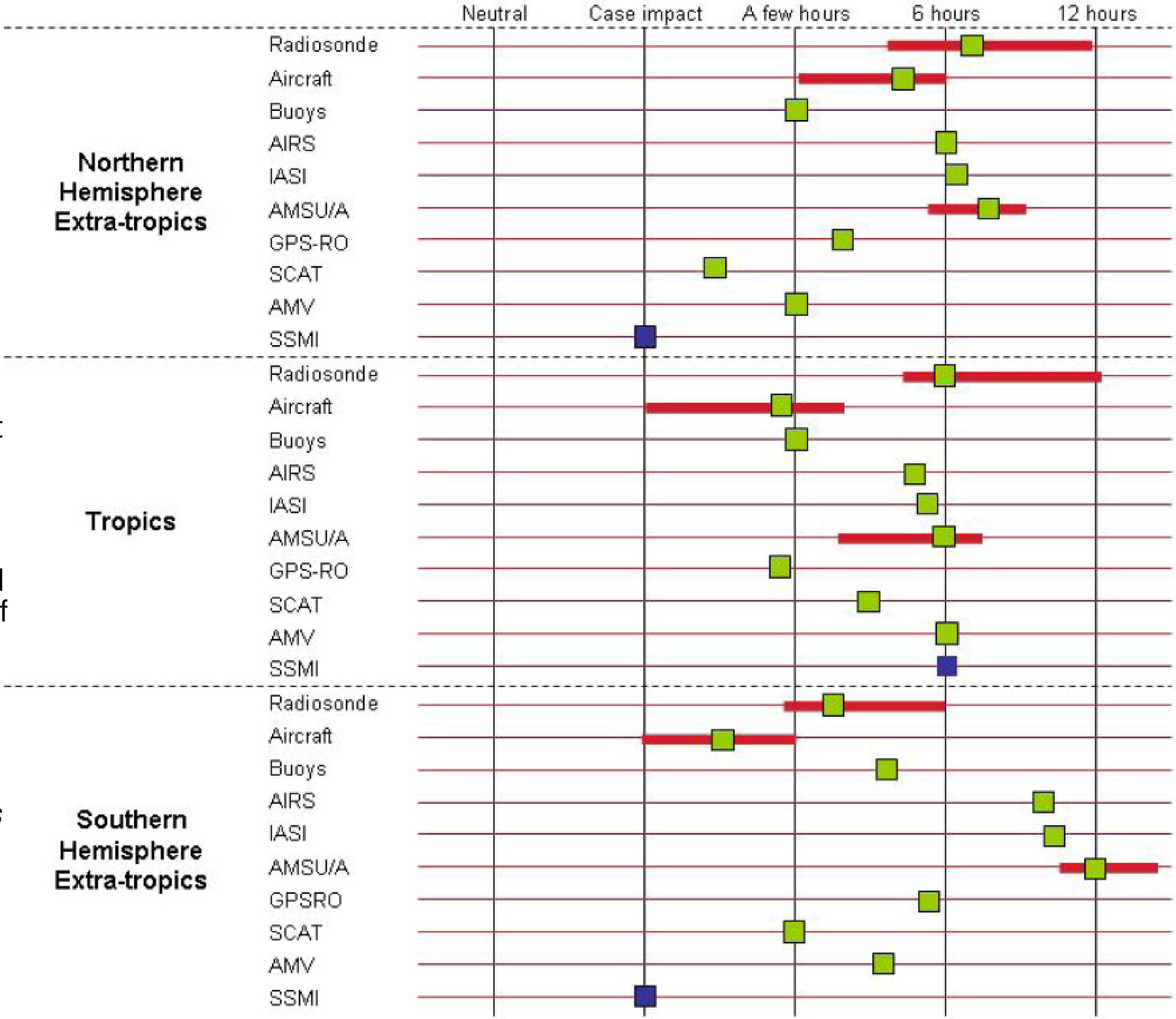
Met Office



Impact of different observing systems.

Current contributions of parts of the existing observing system to the large-scale forecast skill at short and medium-range. The green colour means the impact is mainly on the mass and wind field. The blue colour means the impact is mainly on humidity field. The contribution is primarily measured on large-scale upper-air fields. The red horizontal bars give an indication of the spread of results among the different impact studies so far available.

Fourth WMO Workshop on the Impact of Various Observing Systems on NWP.
Geneva, Switzerland, 19-21 May 2008



Background error (prior) covariance **B** modelling assumptions

*The first operational 3D multivariate statistical analysis method (Lorenc 1981) made the following assumptions about the **B** which characterizes background errors, **all of which are wrong!***

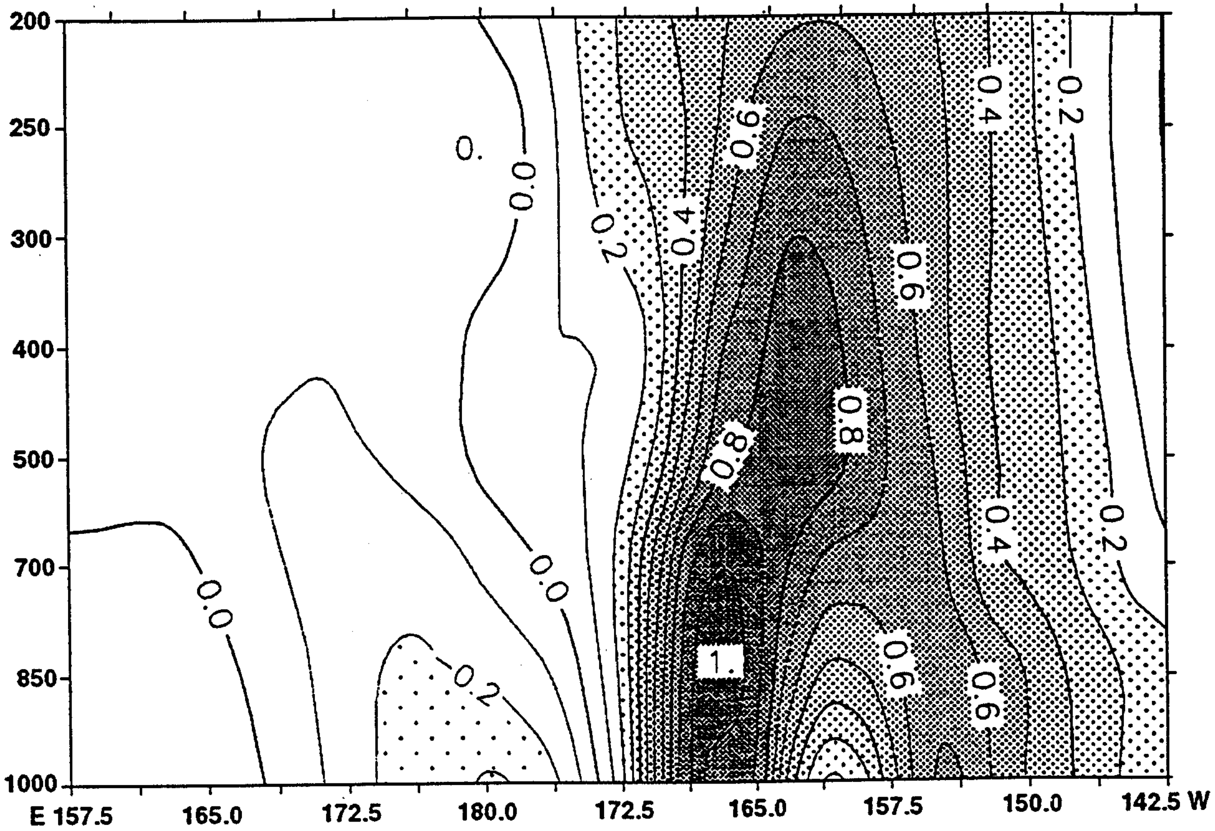
- Stationary – time & flow invariant
- Balanced – predefined multivariate relationships exist
- Homogeneous – same everywhere
- Isotropic – same in all directions
- 3D separable – horizontal correlation independent of vertical levels or structure & vice versa.

Since then many valiant attempts have been made to address them individually, but with limited success because of the errors remaining in the others. ***The most attractive ways of addressing them all are long-window 4D-Var or hybrid ensemble-VAR.***

3D Covariances dynamically generated by 4D-Var

If the time-period is long enough, the evolved 3D covariances also depend on the dynamics:

$$\mathbf{B}_{(x(t_n))} = \mathbf{M}_{n-1} \ldots \mathbf{M}_1 \mathbf{M}_0 \mathbf{B}_{(x(t_0))} \mathbf{M}_0^T \mathbf{M}_1^T \ldots \mathbf{M}_{n-1}^T$$



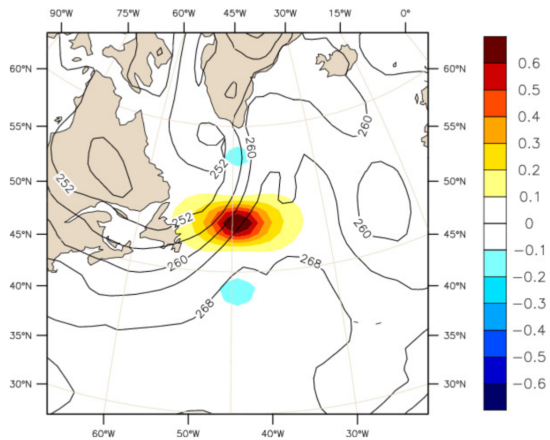
Cross-section of the 4D-Var structure function (using a 24 hour window).

Thépaut, Jean-Nöel, P. Courtier, G. Belaud and G Lemaître: 1996 "Dynamical structure functions in a four-dimensional variational assimilation: A case study" *Quart. J. Roy. Met. Soc.*, **122**, 535-561

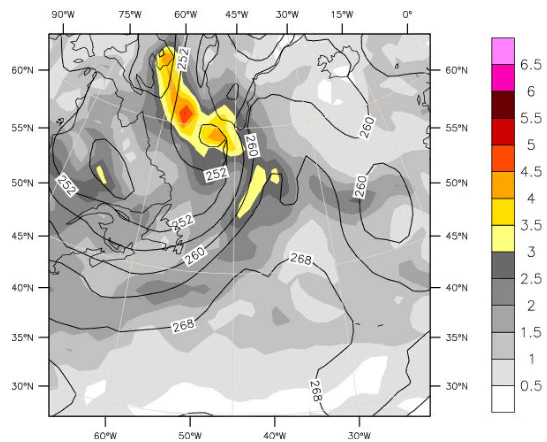
Single observation tests

u response to a single u observation at centre of window

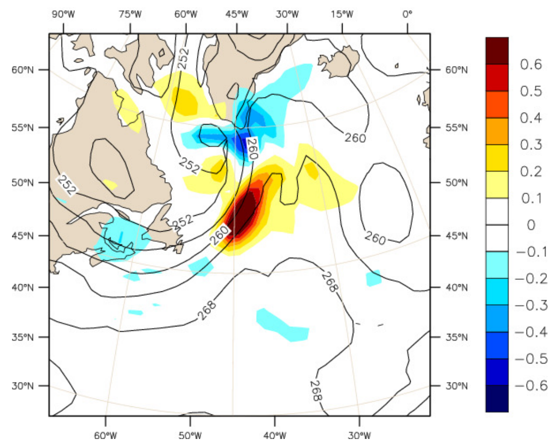
Horizontal



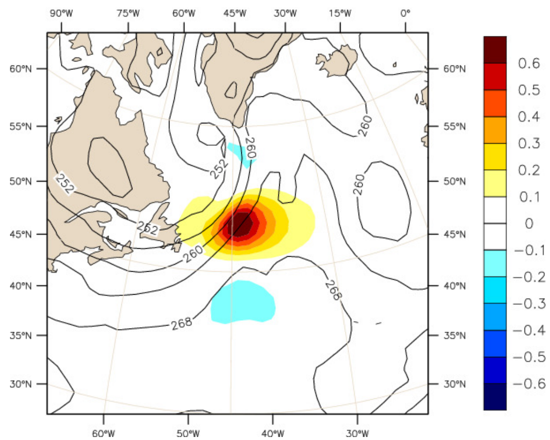
Standard 3D-Var



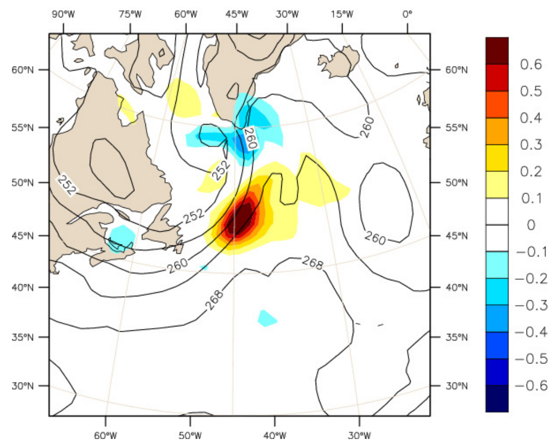
Ensemble RMS



Pure ensemble 3D-Var



Standard 4D-Var



50/50 hybrid 3D-Var

Adam Clayton

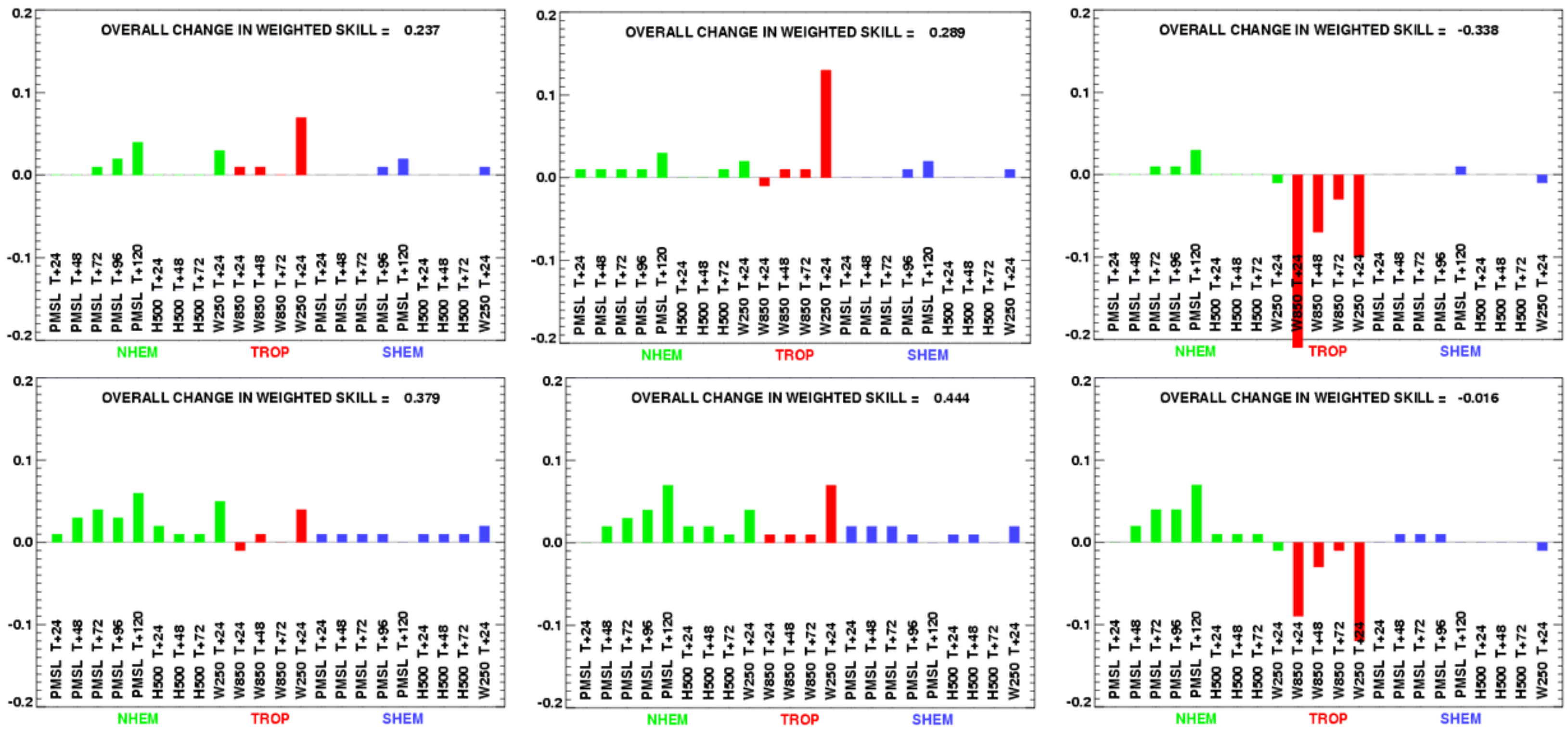
Pre-operational hybrid trials

Summary of skill scores

vs. obs | **vs. ECMWF analyses** | **vs. own analyses**

Dec uncoupled:

Jun coupled:



- Scores vs. ECMWF analyses more consistent with scores vs. obs
- When changing the character of the analysis, verification against own analyses is incestuous and misleading, so we are looking to change the NWP index
- (WMO CBS scores will remain flawed!)

Long-window 4D-Var

- Fisher et al (2005) showed that long-window 4D-Var, allowing for model error, can be equivalent to the Kalman Filter.
- Lorenc (2003) showed that the variable transform model for 3D covariances **B** in 3D-Var can be extended to 4D using Model Error Control Variables (MECV). $\mathbf{B} = \underline{\mathbf{U}}\mathbf{U}_{3D}\mathbf{I}\mathbf{U}_{3D}^{T}\underline{\mathbf{U}}^{T}$
- Tremolet (2006) showed that, alternatively, a 4D-state control variable (4DCV) might have *additional parallelism and relaxed linearity assumption*.
- The MECV and 4DCV methods are solving the same problem – at convergence they should have identical solutions. Different properties must be due to approximate solutions or incomplete convergence.

Conditioning of long-window 4D–Var

Lorenc (2003) showed that adding a MECV-transform

$$\underline{\mathbf{U}} = \begin{pmatrix} \mathbf{I} & 0 & \cdots & 0 \\ \mathbf{M}_0 & \mathbf{I} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{M}_{n-2}\cdots\mathbf{M}_0 & \mathbf{M}_{n-2}\cdots\mathbf{M}_1 & \cdots & \mathbf{I} \end{pmatrix}$$

to the variable transforms of 3D-Var, gives a diagonal 4D

$$\mathbf{B}_{(\underline{\mathbf{v}})} = \begin{pmatrix} \mathbf{P}^f(t_0) & 0 & \cdots & 0 \\ 0 & \mathbf{Q}_0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{Q}_{n-2} \end{pmatrix}.$$

with similar conditioning to 3D-Var: $\widehat{J''} = \mathbf{I} + \mathbf{B}^{\mathrm{T}/2}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{B}^{1/2}$.

Tremolet (2006) showed that the 4DCV method is less well conditioned, with Hessian that of

J_o plus the tri-diagonal matrix:

$$\begin{pmatrix} \mathbf{B}^{-1}+\mathbf{M}_1^{\mathrm{T}}\mathbf{Q}_1^{-1}\mathbf{M}_1 & -\mathbf{M}_1^{\mathrm{T}}\mathbf{Q}_1^{-1} & \cdots & 0 & 0 \\ -\mathbf{Q}_1^{-1}\mathbf{M}_1 & \mathbf{Q}_1^{-1}+\mathbf{M}_2^{\mathrm{T}}\mathbf{Q}_2^{-1}\mathbf{M}_2 & \cdots & 0 & 0 \\ 0 & -\mathbf{Q}_2^{-1}\mathbf{M}_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -\mathbf{M}_{m-1}^{\mathrm{T}}\mathbf{Q}_{m-1}^{-1} & 0 \\ 0 & 0 & \cdots & \mathbf{Q}_{m-1}^{-1}+\mathbf{M}_m^{\mathrm{T}}\mathbf{Q}_m^{-1}\mathbf{M}_m & -\mathbf{M}_m^{\mathrm{T}}\mathbf{Q}_m^{-1} \\ 0 & 0 & \cdots & -\mathbf{Q}_m^{-1}\mathbf{M}_m & \mathbf{Q}_m^{-1} \end{pmatrix}$$

Convergence of 4DCV long–window 4D–Var

- The time-dimension of this 4DCV method is treated like the space-dimensions in 3D-Var without a CV-transform.
- Lorenc (1997) showed that a trivial 2D-Var without a CV-transform can be disastrously ill-conditioned.
- While Tremolet (2006), Fisher (personal communication) and Payne (personal communication) say that methods for pre-conditioning exist, results have not been published (AFAIK).

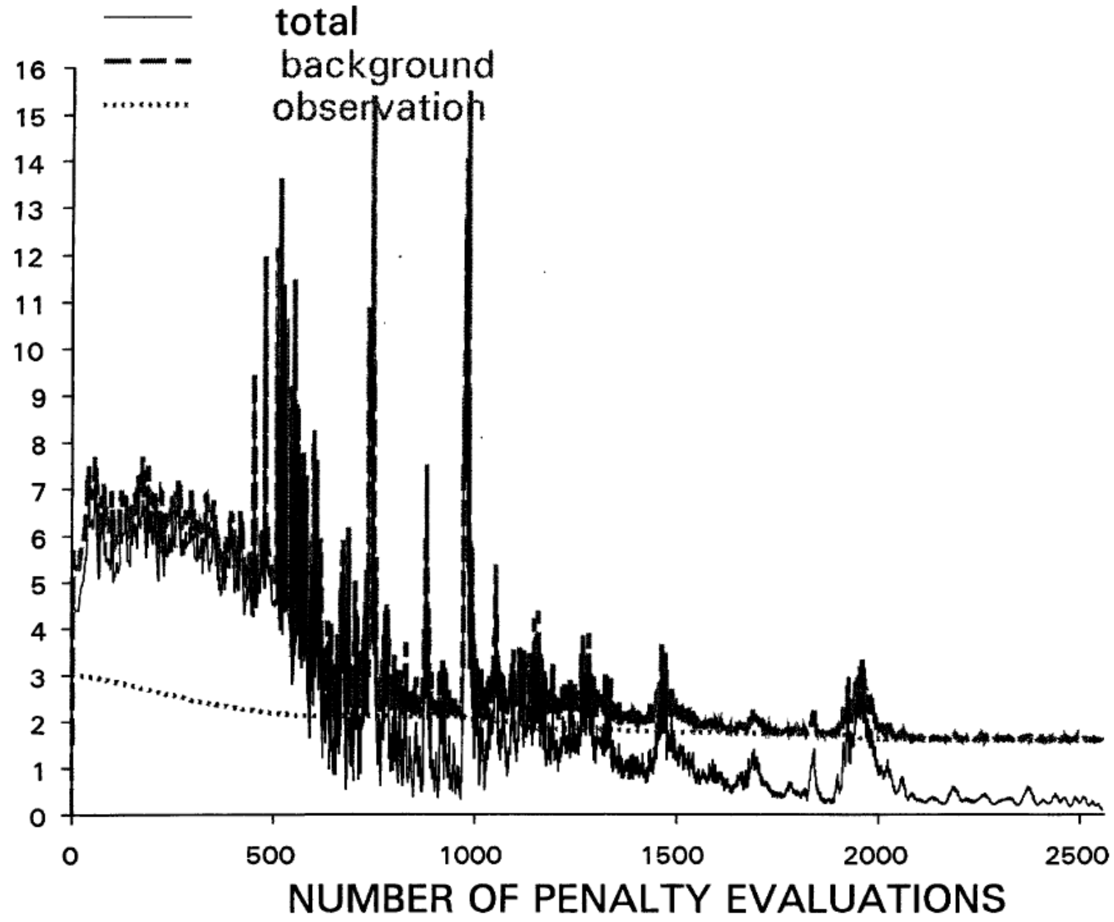


Fig. 4. Solid line — norm of gradient of J at each iteration during the minimisation (at convergence it should be zero). Dashed and dotted lines — norm of contribution to gradient from J^b and J^o terms.

Long-window 4D-Var for a chaotic model

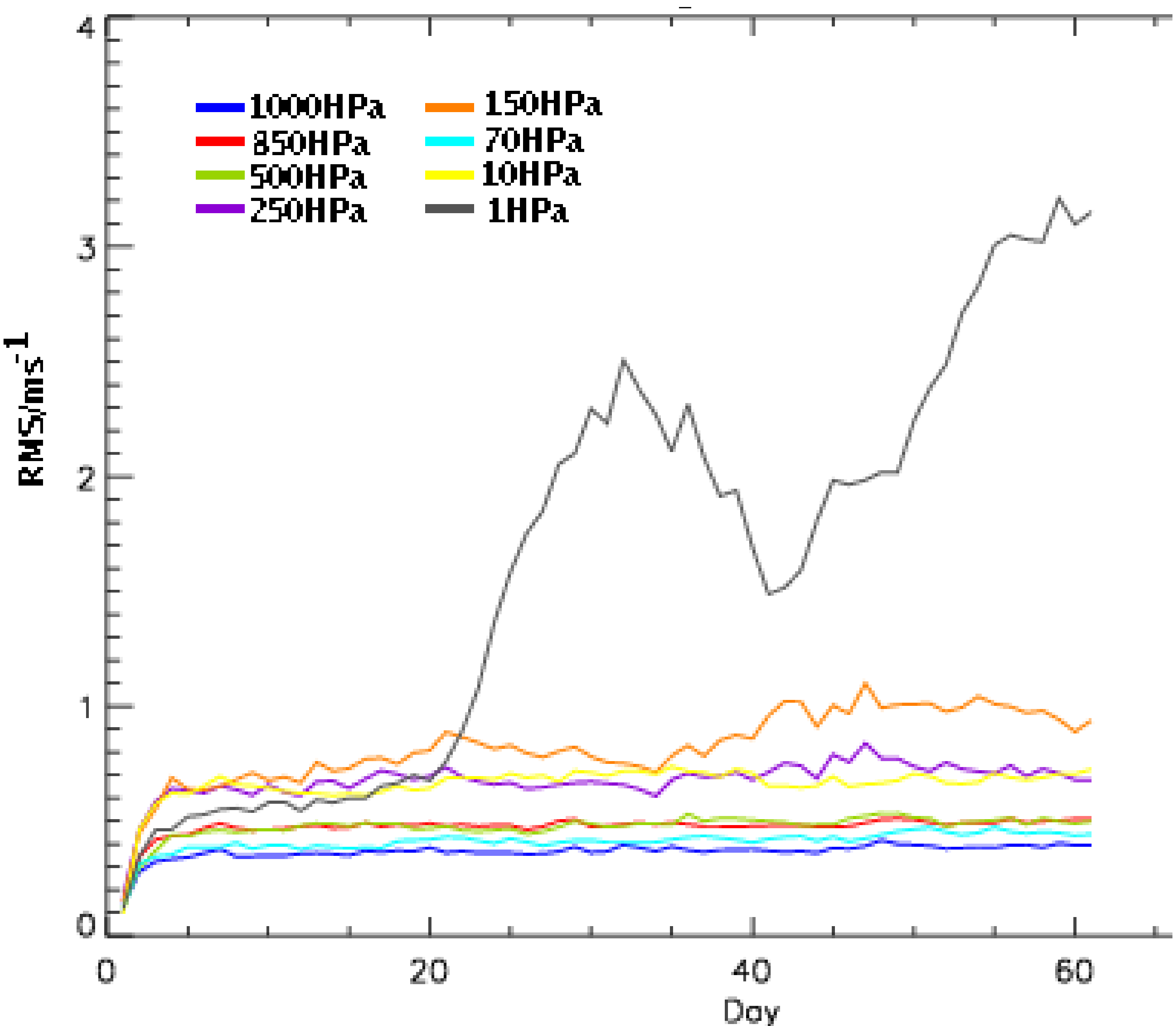
- Lorenc and Payne (2007) discussed 4D-Var for models with a ***wide range of scales*** – the small scales behave chaotically and cause problems. They suggested using a regularised linear model which filters poorly observed small scales (e.g. eddies in ocean DA, Hoteit *et al.* 2005).
- Abarbanel *et al.* (2010), approaching data assimilation as ***synchronised chaos***, say that there must be enough [observational] controls to move the positive conditional Lyapunov exponents on the synchronization manifold to negative values. (E.g. this was the case for the toy model used by Fisher *et al*. (2005)).
- Modern high-resolution global NWP models have regions (e.g. the middle atmosphere, Polavarapu *et al.* 2005) where neither approach is easy to apply.

Differences between identical NWP assimilations due to small initial perturbations.

Met Office

Global RMS differences between analysed u–fields at different levels.

Peter Jermey



Nonlinearity – benefitting from the attractor

- The atmospheric state is fundamentally governed by nonlinear effects, e.g. convective-radiative equilibrium, condensation cloud & precipitation. Nonlinear chaotic systems have an ***attractor manifold*** of states that occur in reality – far fewer than all possible states. This gives us recognisable weather systems and practical weather prediction!
- Usual minimum variance **"best" estimate is not on the attractor**.
- We can only afford to handle linear effects accurately in 4D–Var, nonlinearities are considered afterward as ***spin-up*** and ***initialisation***. So it cannot properly handle creating, removing or moving weather systems.
- The ***Particle Filter*** algorithm can do accurate nonlinear DA, but is prohibitively expensive for NWP. Van Leeuwen (2010) suggests adapting the particle filter by ***nudging*** the model towards the observations. The MECV version of 4D-Var would be an ideal way of doing the nudging, while ***PF*** ideas might better handle ***creating or removing weather systems***.
- 4DCV 4D-Var avoids long model runs, so should do less well on ***spin-up***.

Effect of 0% & 100% limits on RH

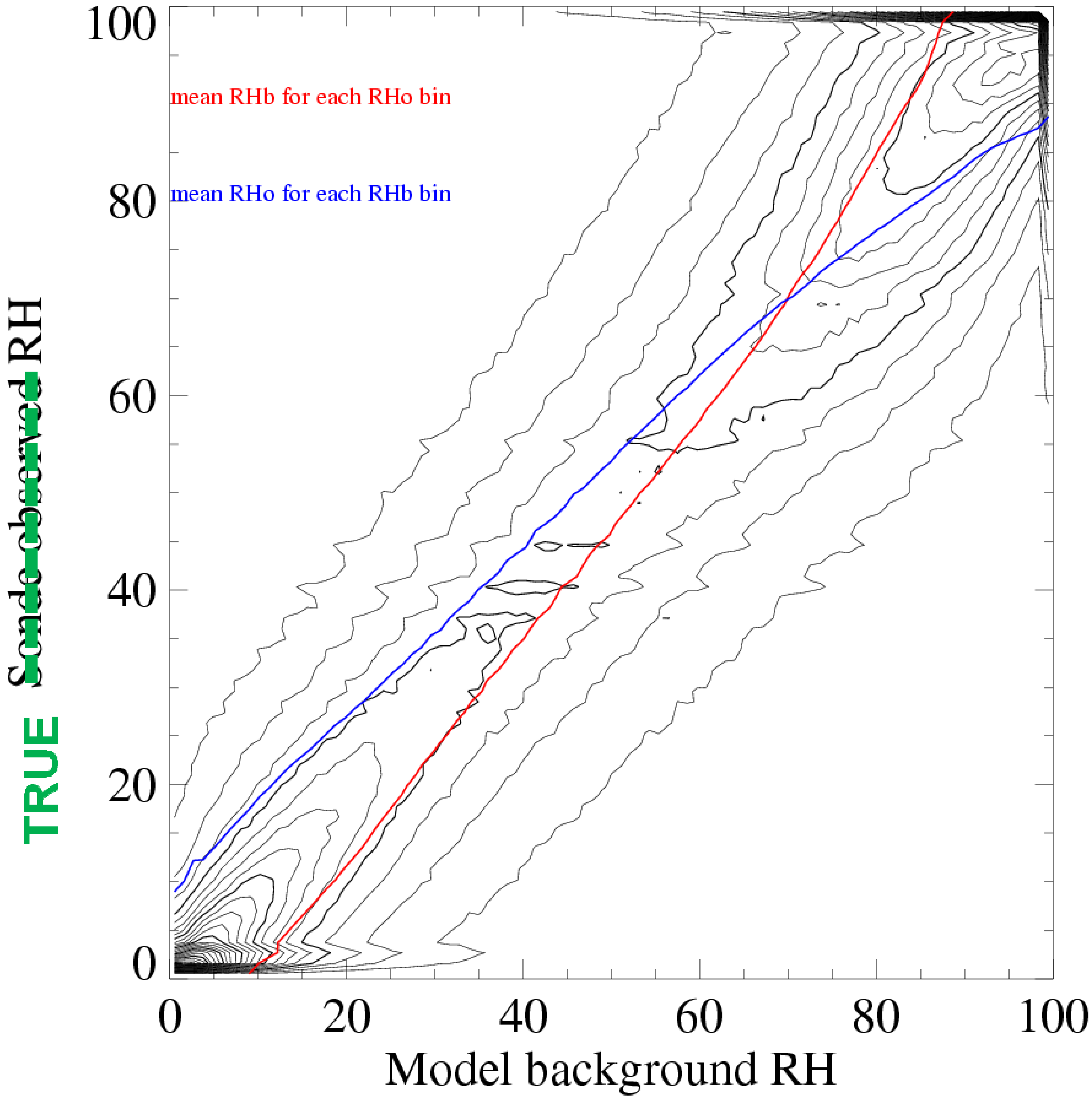
Met Office

$P(x|x_b)$ is biased, with mean given by blue line.

$\Rightarrow$ “best” estimate obtained by modifying x_b away from limits.

This would damage forecasts of cloud and rain!!

Diagram from Lorenc (2007)



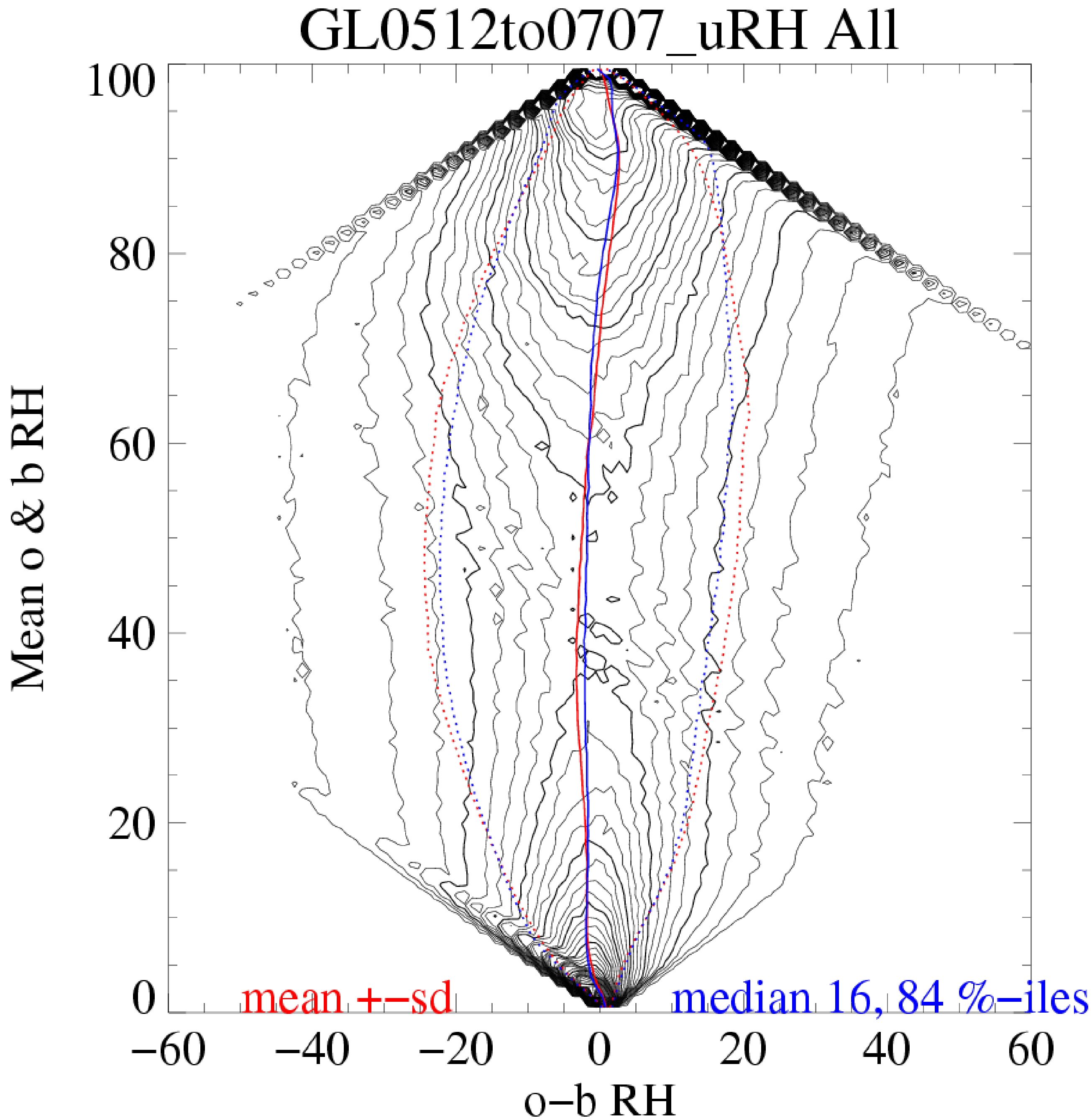
The principle of symmetry

Acknowledgement to Elias Holm.

Make all statistics used in DA symmetric functions of x_b and x_a.

e.g. Plot shows $Rh_o\text{-}RH_b \mid Rh_o\text{+}RH_b$ is unbiased.

This makes DA process implicit!



Fitting models of model error

Met Office

- The VAR approach encourages us to build ***physically based models of errors***: observational, representativity, background, observational bias, forecast model, …
  This approach is more likely to give a DA method applicable to a ***wide range of regimes***.

- We ***cannot uniquely fit*** these models to one set of o−b statistics. Nevertheless, as long as we have conceptual insight as to what should be common, we can use statistics from a range of regimes to estimate some of them (e.g. Hollingsworth and Lonnberg 1986, Desroziers *et al.* 2005, Dee 2005).

- Tremolet (2007) has started applying these ideas to model error, but much remains to be done.

Scalability – exploiting massively parallel computers

- 4D-Var as usually implemented requires ***sequential*** running of a reduced resolution linear model and its adjoint. It will be difficult to exploit computers with more (but not faster) processors to make 4D-Var run as fast at higher resolution.
- Improved current 4D-Var algorithms ***postpone*** the problem a few years, but it will probably return, hitting 4D-Var before the high-resolution forecast models.
- ***4DCV 4D-Var can be parallelised*** over each CV segment.
- Ensemble DA methods run a similar number of model integrations in ***parallel***. This is inherently less suited to finding the best analysis, but the time saved by easy parallelisation might be deployed to offset this (e.g. by a bigger ensemble or higher resolution). ***4D-Ensemble-Var*** is an attractive approach since other advantages of VAR can be retained.

Concluding remarks

Met Office

- 4D-Var is currently the best DA method for operational NWP. We can expect it to remain so for several years, so R&D to improve it is worthwhile.
- The ability of model-space variational methods to handle dense but "incomplete" indirect observations has been very beneficial. We do not want to abandon this!
- Eventually 4D-Var is likely to be replaced by ensemble methods which are better on parallel computers and for providing probabilistic forecasts.
- The success of the hybrid approach makes it a strong candidate to supplement 4D-Var (instead of very long windows), and perhaps eventually to replace 4D-Var when it is on future computers.

Questions and answers

References

Abarbanel, H. D. I., Kostuk, M. and Whartenby, W. (2010), Data assimilation with regularized nonlinear instabilities. *Quart. J. Roy. Met. Soc.*, **136**: 769–783. doi: 10.1002/qj.600

Dee, D. P. (2005), Bias and data assimilation. *Quart. J. Roy. Met. Soc.*, **131**: 3323–3343. doi: 10.1256/qj.05.137

Desroziers, G., L. Berre, B. Chapnik and P. Poli 2005: Diagnosis of observation, background and analysis error statistics in observation space. *Quart. J. Roy. Met. Soc.*, **131**, 3385-3396.

Fisher, M., M. Leutbecher and G. A. Kelly 2005: "On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation." *Quart. J. Roy. Met. Soc.*, **131**, 3235-3246.

Hollingsworth, A. and Lonnberg, P. 1986: " The statistical structure of short-range forecast error as determined from radiosonde data. Part I: The wind field." *Tellus*, **38A**, 111-136.

Hoteit, I., B. Cornuella, A. Kohl and D. Stammer, 2005: "Treating Strong Adjoint Sensitivities in Tropical Eddy-Permitting Variational Data Assimilation", *Quart. J. Roy. Met. Soc.*, **131**, 3659-3682.

Lorenc, A.C. 1981: "A global three-dimensional multivariate statistical analysis scheme." *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, Andrew C., 1997: "Development of an Operational Variational Assimilation Scheme." *J. Met. Soc. Japan*, Special issue "Data Assimilation in Meteorology and Oceanography: Theory and Practice." **75**, No. 1B, 339-346.

Lorenc, A. C., 2003: Modelling of error covariances by four-dimensional variational data assimilation. *Quart. J. Roy. Met. Soc.*, **129**, 3167-3182.

Lorenc, A. C., 2003b: The potential of the Ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Met. Soc.,* **129**, 3183-3203.

Lorenc, Andrew C. 2007: A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers. *Met Office FRTR 504.* Sept 2007. http://www.metoffice.gov.uk/research/nwp/publications/papers/technical_reports/

Lorenc, Andrew C. and Tim Payne 2007: "4D-Var and the Butterfly Effect: Statistical four-dimensional data assimilation for a wide range of scales." *Quart. J. Roy. Met. Soc.*, **133**, 607-614.

Thépaut, Jean-Nöel, P. Courtier, G. Belaud and G Lemaître: 1996 "Dynamical structure functions in a four-dimensional variational assimilation: A case study" *Quart. J. Roy. Met. Soc.*, **122**, 535-561

Tremolet, Y. (2006), Accounting for an imperfect model in 4D-Var. Quarterly Journal of the Royal Meteorological Society, 132: 2483–2504. doi: 10.1256/qj.05.224 & Corrigendum. Quarterly Journal of the Royal Meteorological Society, 132: 3127. doi: 10.1002/qj.2006132621114

Trémolet, Y. (2007), Model-error estimation in 4D-Var. Quarterly Journal of the Royal Meteorological Society, 133: 1267–1280. doi: 10.1002/qj.94

Polavarapu, S., T. G. Shepherd, Y. Rochon and S. Ren 2005: Some challenges of middle atmosphere data assimilation. *Quart. J. Roy. Met. Soc.*, **131**, 3513-3528.

van Leeuwen, Peter Jan 2010: "Nonlinear Data Assimilation in geosciences: an extremely efficient particle filter". *Quart. J. Roy. Met. Soc.*, 136: 1991–1999. doi: 10.1002/qj.699