Recent Developments in Data Assimilation

Florence Rabier & Jean-Noël Thépaut

October 2011

Introduction

- This presentation is a digest of two major events this year:
  - THORPEX DAOS
  - ECMWF annual seminar
- We will concentrate on:
  - Progress in data assimilation methods
  - Observation Usage and Impact

- Data assimilation methods
- Observation related aspects
- Real data assimilation systems
- Efficient use of computer architectures

www.ecmwf.int

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:

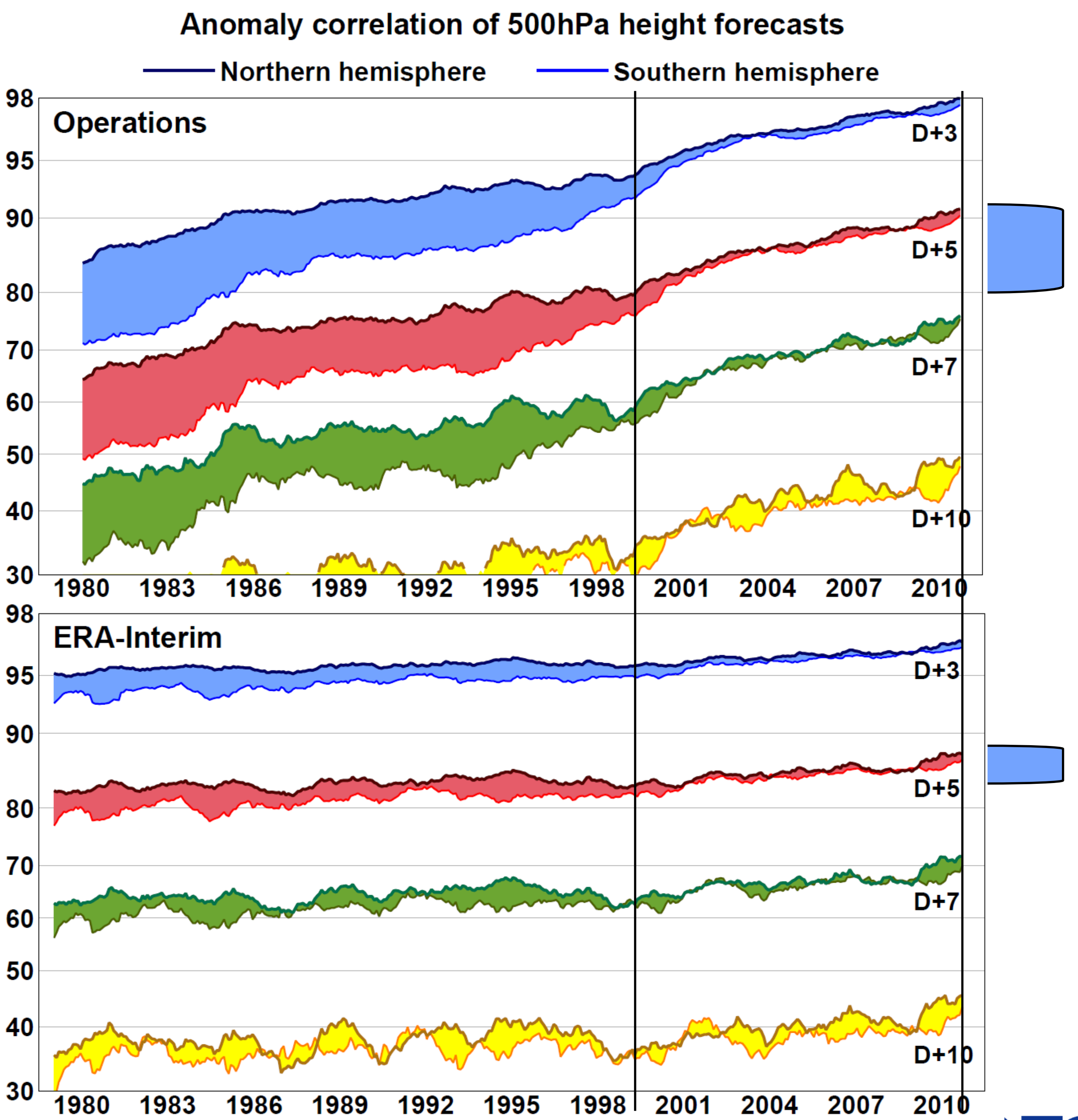
1. ***Model improvements, especially resolution.***

2. ***Careful use of forecast & observations, allowing for their information content and errors.*** *Achieved by variational assimilation e.g. of satellite radiances. (Simmons & Hollingsworth 2002)*

3. ***Advanced assimilation using forecast model: 4D-Var***

4. ***Better observations.***

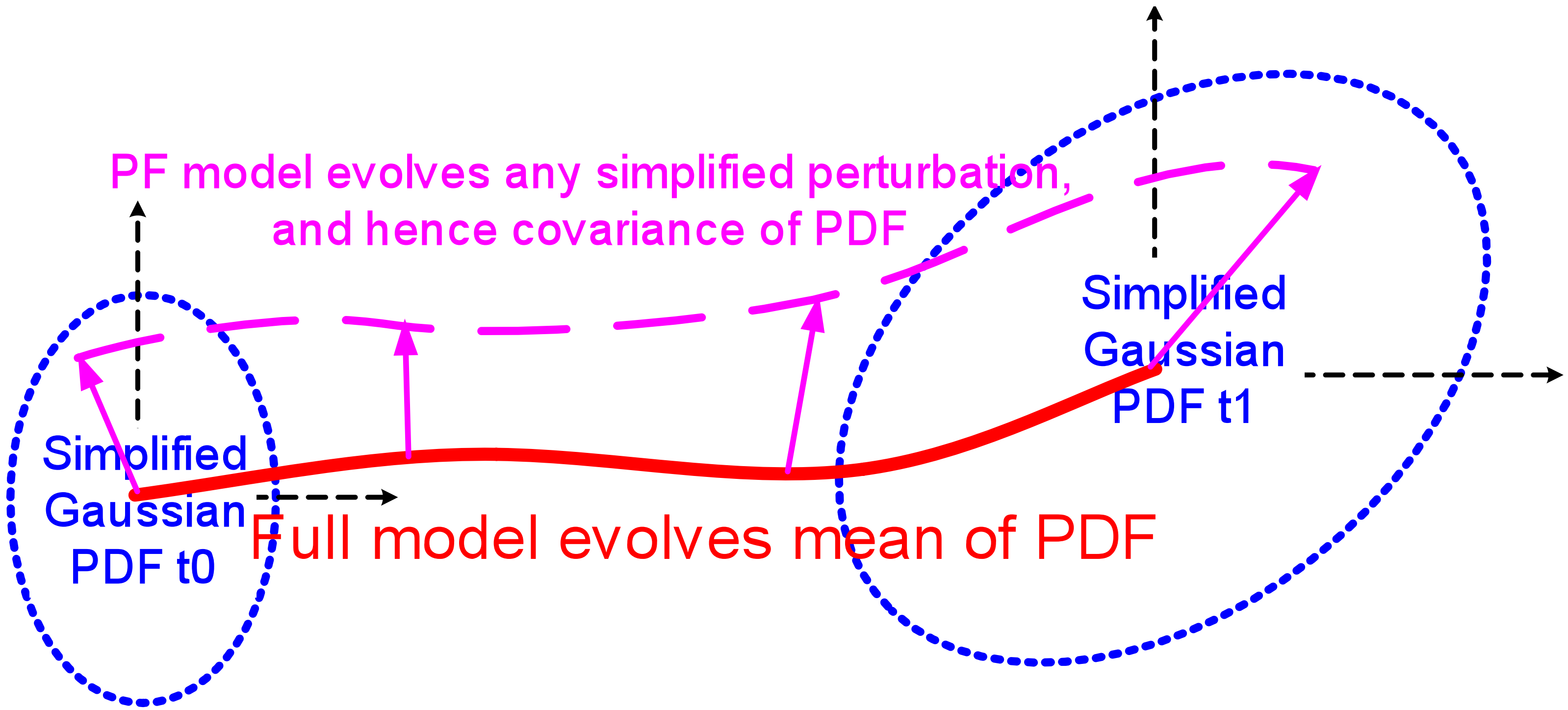
Andrew Lorenc



12% per decade

5% per decade

Statistical, incremental 4D-Var



optionally augmented by a model error correction term.

Andrew Lorenc

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.

Andrew Lorenc

Hybrid Var/EnKF - best of both worlds?

Features from EnKF	Features from VAR
Extra flow-dependence in $\mathbf{P}^b$	Localization done correctly (in model space)
More flexible treatment of model error (can be treated in ensemble)	Reduction in sampling error in time-lagged covariances (full rank evolution of $\mathbf{P}^b$ in assimilation window in 4DVar).
Automatic initialization of ensemble forecasts, propagation of covariance info from one cycle to the next.	Ease of adding extra constraints to cost function

--: covariance inflation,
covariance localization

--: scalability,
static B, maintenance cost

Jeff Whitaker

Hybrid methods

Hybrid method: Use flow-dependent state error estimates (from an EnKF/EDA system) in the deterministic 3/4D-Var analysis system:

1) **Integrate flow-dependent state error covariance information into the "static" variational analysis**
2) **Keep the full rank representation of B and its implicit evolution that 3/4D-Var provide inside the assimilation window**
3) **More robust than pure EnKF for limited ensemble sizes and large model errors**
4) **Allow for flow-dependent QC of observations**

Operational hybrid methods

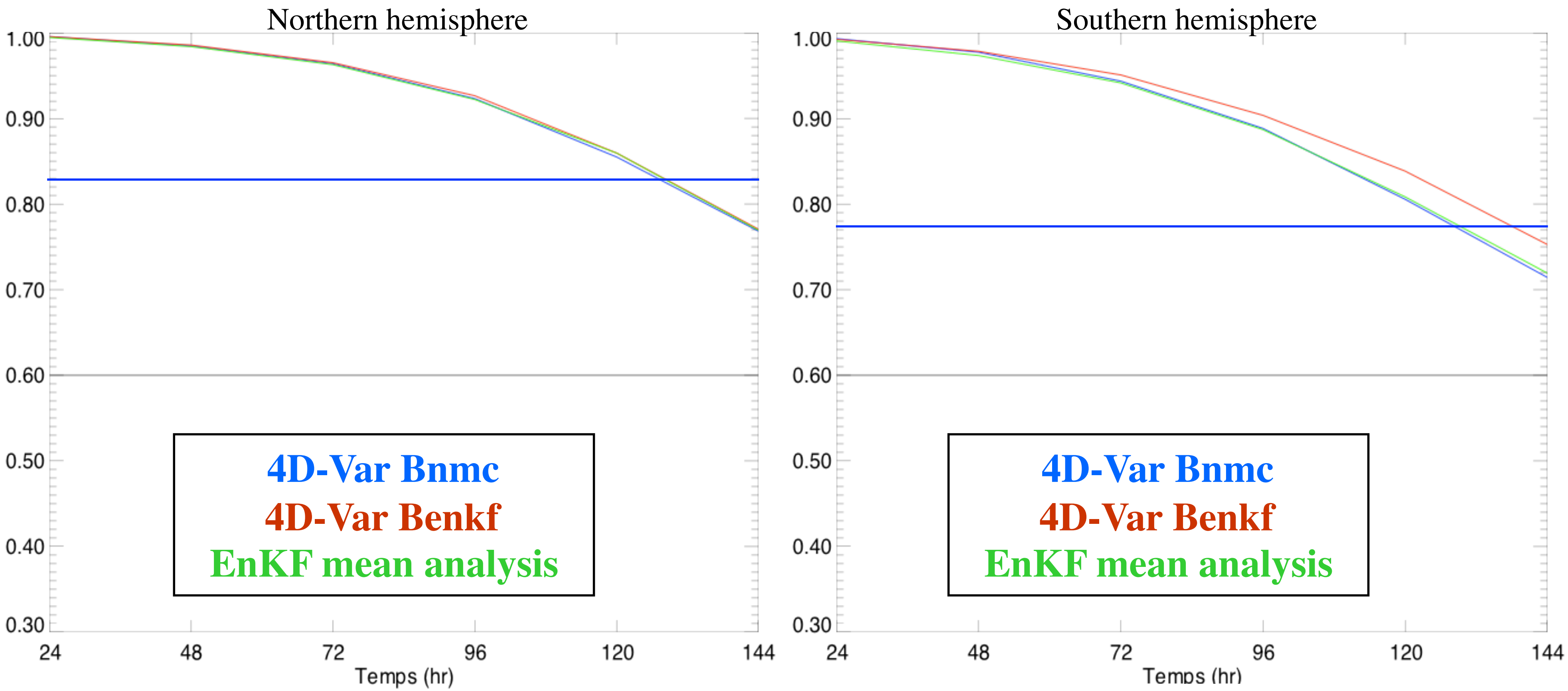
In operational use (or in an advanced testing), there are currently two
main approaches to doing an hybrid DA in a variational context:

1. **Alpha control variable method (Met Office, NCEP/GMAO, CMC)**
2. **Ensemble of Data Assimilations method (Météo-France, ECMWF)**

Comparison of 4D-Var/EnKF
(Buehner et al 2010)

Dale Barker

Verifying analyses from 4D-Var with Bnmc



Conclusion: Combined 4D-Var + EnKF covariances->~10hrs SH skill

ECMWF

Hybrid methods: α control variable

1. Alpha control variable method (Met Office, NCEP/GMAO)

Conceptually add a flow-dependent term to the climatological **B** matrix:

$$\mathbf{B} = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{P}_e \circ \mathbf{C}_{loc}$$

$\mathbf{B}_c$ is the static, climatological covariance

$\mathbf{P}_e \circ \mathbf{C}_{loc}$ is the localised ensemble covariance

In practice this is done through augmentation of control variable:

$$\delta \mathbf{x} = \beta_c \mathbf{B}_c^{1/2} \mathbf{v} + \beta_e \mathbf{X}' \circ \boldsymbol{\alpha}$$

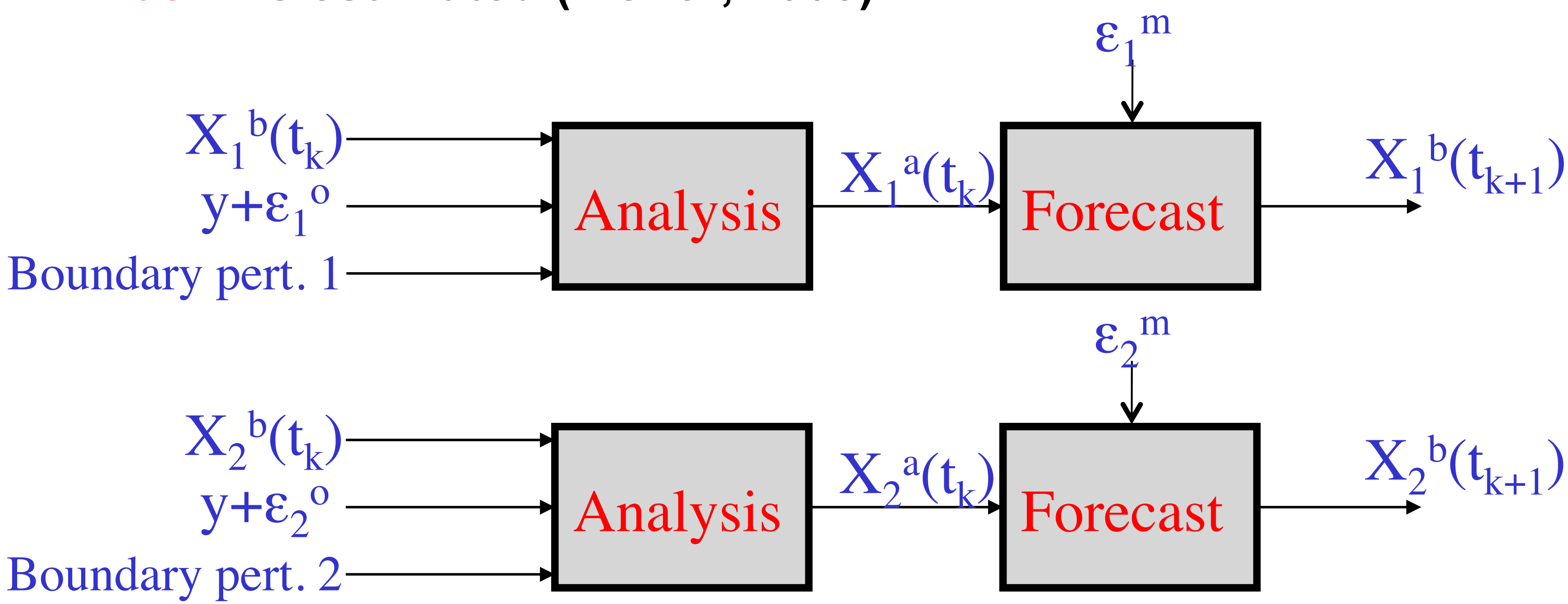
and introducing an additional term in the cost function:

$$J = \frac{1}{2} \boldsymbol{v}^T \boldsymbol{v} + \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{C}_{loc}^{-1} \boldsymbol{\alpha} + J_o + J_c$$

from: A.Clayton, MetOffice

Hybrid methods: EDA

2. **The** **Ensemble of Data Assimilations** **(EDA, Raynaud et al., 2010, Isaksen *et al.* 2010) can be considered a** **flow-dependent extension** **of the way the** ***climatological background error matrix*** **is estimated (Fisher, 2003).**



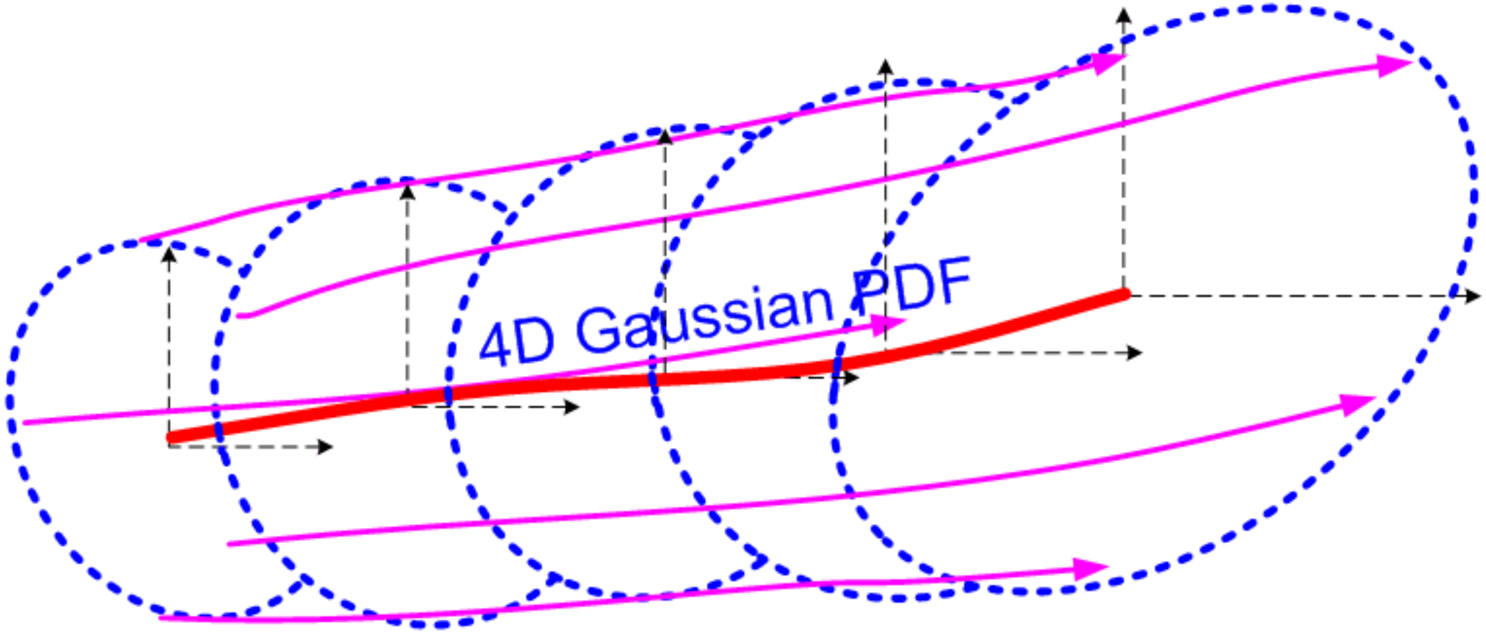
Scalability – exploiting massively parallel computers

- **4D-Var as usually implemented requires *sequential* running of a reduced resolution linear PF 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, but is difficult to precondition well.**
- **Ensemble DA methods run a similar number of model integrations in *parallel*. It is attractive to replace the iterated running of the PF model by precalculated ensemble trajectories: *4D−Ensemble-Var*. Other advantages of VAR can be retained.**

Andrew Lorenc

Incremental 4D-Ensemble-Var

Andrew Lorenc



Trajectories of perturbations from ensemble mean

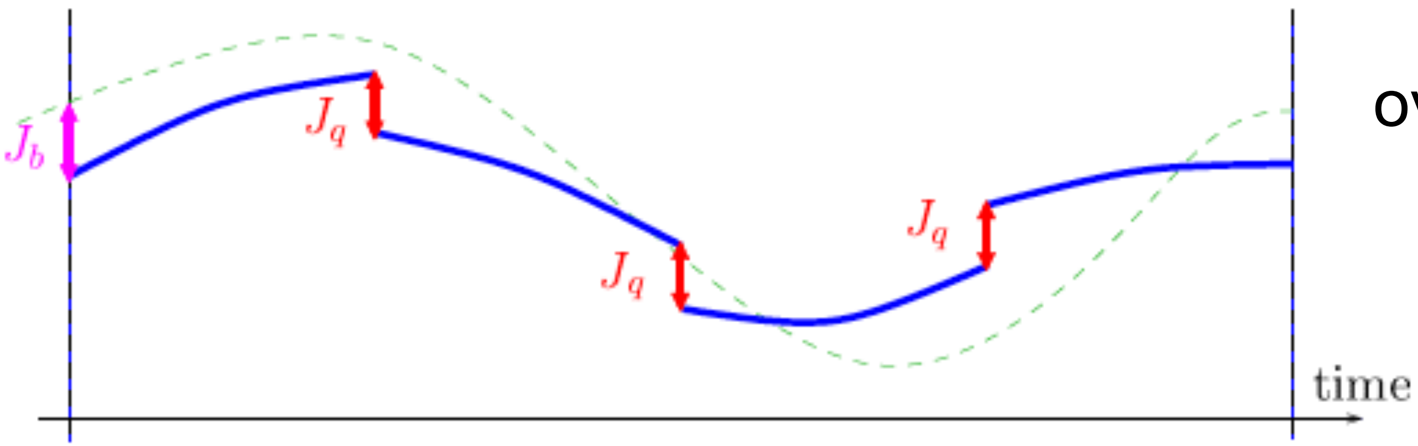
Full model evolves mean of PDF

Localised trajectories define 4D PDF of possible increments

4D analysis is a (localised) linear combination of nonlinear trajectories. It is not itself a trajectory.

Long window weak constraint 4D-Var

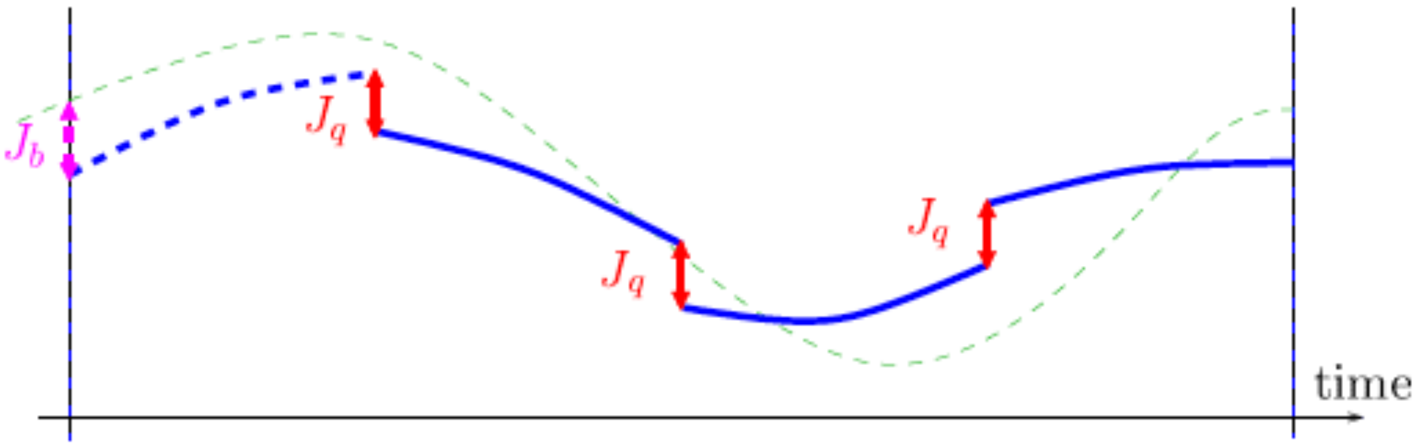
Suppose we extend the window by a few hours:



Parallelisation over sub-windows

We expect very little change in the the analysis for the first sub-window:

Mike Fisher



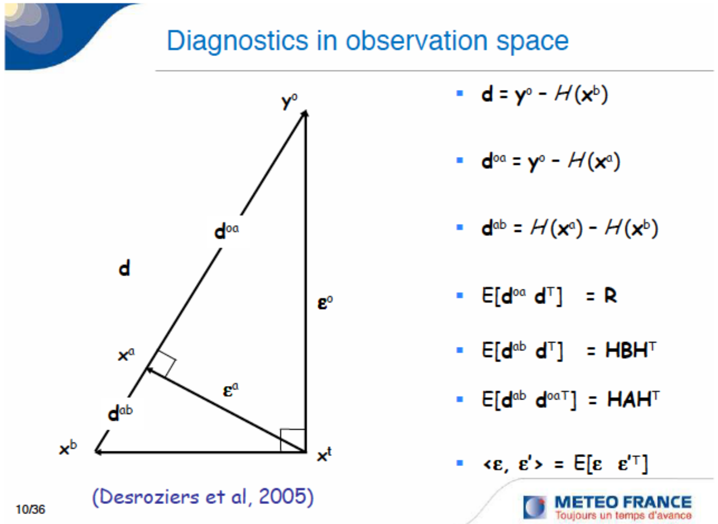
ECMWF

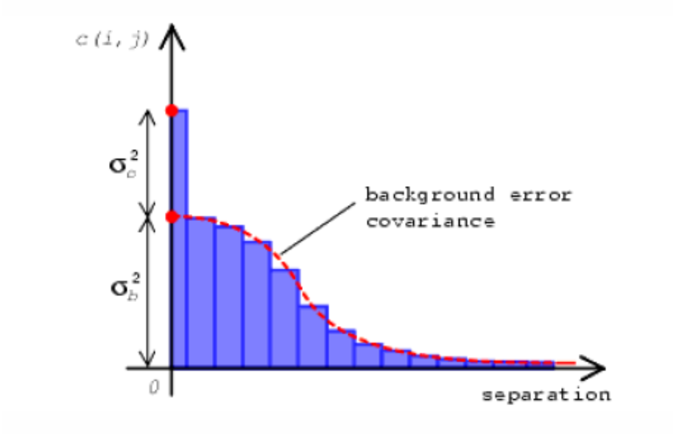
Other important aspects

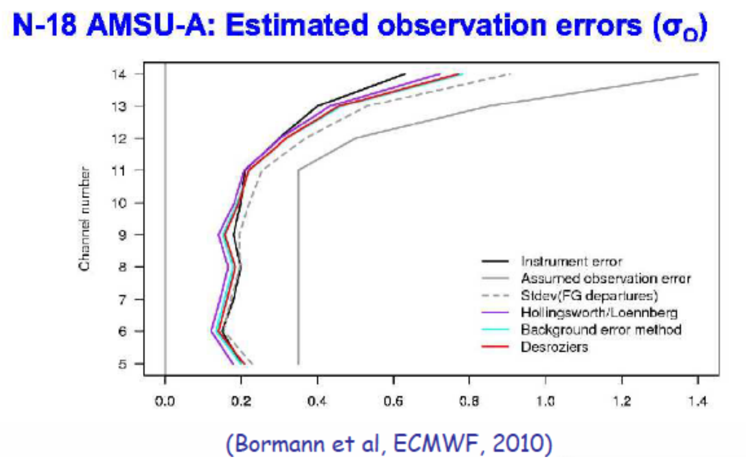
- **Diagnostics for specifying observation error covariances in the assimilation**

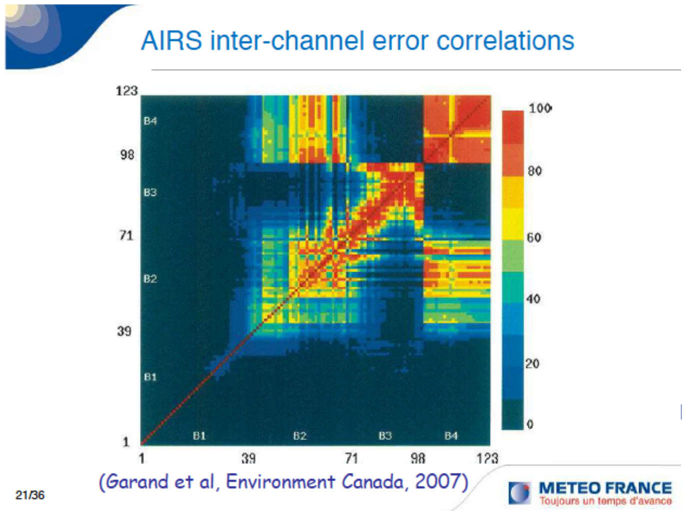
Gerald Desroziers

  - **Desroziers, Lonnberg & Hollingsworth, etc.**
  - **Effort in all centres to better characterize structure and amplitude**



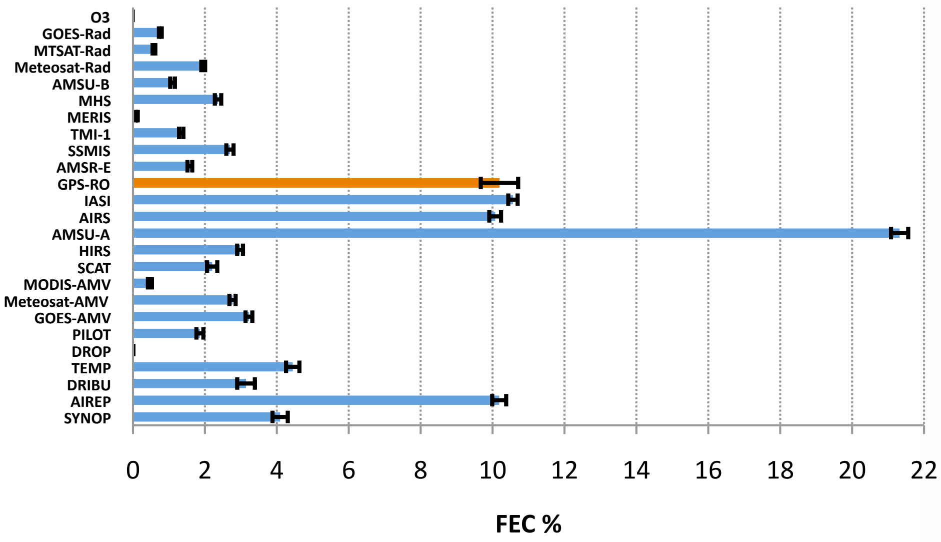






Other important aspects

- **Enhanced diagnostics of assimilation and forecast performance (obs, R, B)**



Carla Cardinali

- **The invisible world: pre- and post- processing in Data Assimilation**

Transforming the raw data
- Transforming into a different space
- Averaging the data
- Filtering the observations

Comparing model and observations
- Monitoring and choice of observations
- Bias correction
- Removing wrong data

Thinning the data
- Reducing data quantity and error correlation
- Choosing the most relevant local data
- Selective thinning depending on the flow

Filtering the analysis
- Initialisation methods
- Influence on the analysis

Florence Rabier

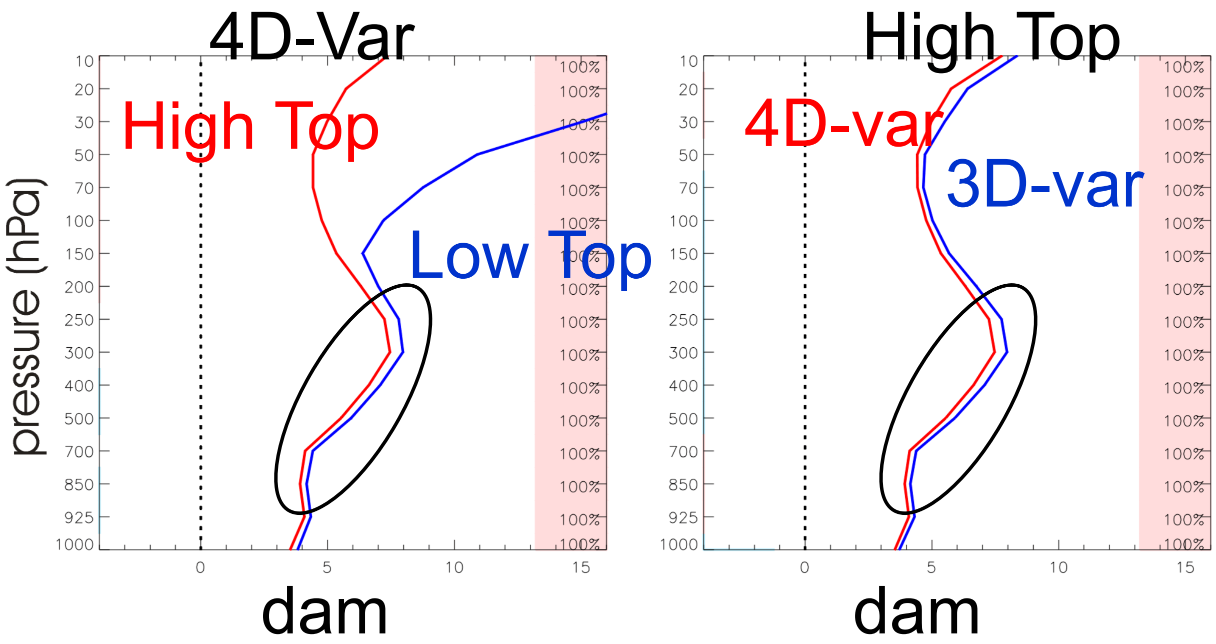
Other important aspects: DA in stratosphere

- What are the challenges in stratospheric and mesospheric data assimilation?
  - Separation of model and observation error biases
    - Add more low-bias obs with vertical structure information (more limb data needed)
  - Vertical spreading of information through covariances
    - Are background error covariances appropriately defined in the upper stratosphere given the poor vertical resolution provided by the observing system?
    - Ad hoc measures prevent spurious increments from contaminating mesosphere.
  - Lack of wind information in tropics
    - Without clear mass-wind balance, temperature information of limited use. Solution: new obs such as ADM or SWIFT? 4D-var and tracer assimilation?

Improving the stratosphere improves 5-day forecasts in the troposphere

On June 22, 2009 Canadian Meteorological Centre implemented operationally a global stratospheric model (0.1 hPa) for medium range weather forecasts

O-F(5 day) against NH sondes for GZ



A good stratosphere impacts troposphere forecasts as much as 4D-Var

Polavarapu et al (2011)

Winter

Dec. 20 – Jan. 26, 2006
(75 cases)

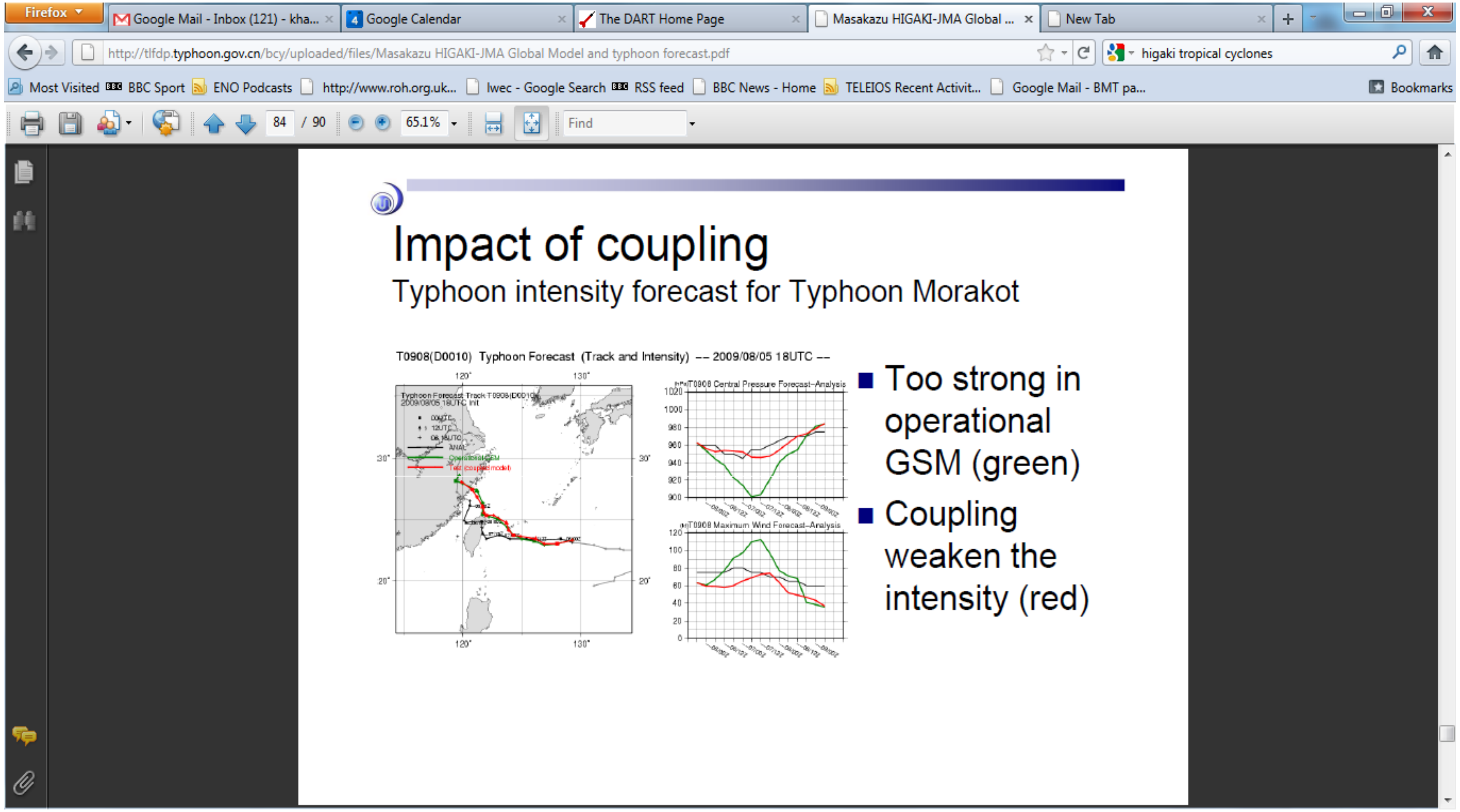
Other important aspects: Ocean Data assimilation

Summary

- Ocean DA is diverse and mature
- Many basic challenges still exist:
  - expansion of control vector (B?)
  - tracer assimilation
  - initialization shock & filtering
  - vertical projection of satellite obs
  - covariance models
  - biogeochemical data assimilation
  - model error
  - internal tides
  - quality control & bias correction
  - air-sea coupling at all scales
- Sub-mesoscale and deep ocean are poorly observed (and poorly constrained)

Andy Moore

Other important aspects: Ocean/atmosphere coupled data assimilation



Keith Haines

Other important aspects

- Regional aspects
  - High resolution data assimilation, hydrometeors
- Challenge of satellite data assimilation
- Assimilation of the hydrological cycle
- Ocean/atmosphere coupled data assimilation
- Nonlinear data assimilation
  - Particle filters, etc.