



# Computational Issues in Data Assimilation for Operational NWP

*Andrew Lorenc.*

WWRP/THORPEX Workshop on 4D-Var and Ensemble Kalman Filter Inter-comparisons.  
Buenos Aires - Argentina, 10-13 November 2008

© Crown copyright Met Office



# Contents

This presentation covers the following areas

- What is needed for a world-class DA system for NWP.
- Handling a wide range of space- & time-scales.
- Vision for next decade DA for global NWP: 4D-Var - computational issues.
- 10~20 years??
- Some thoughts on convective scale DA.

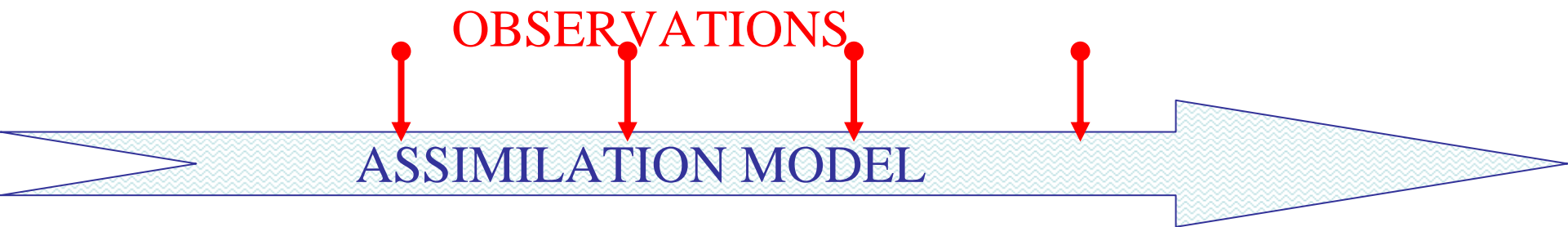


# ***Data Assimilation is the process of absorbing and incorporating observed information into a prognostic model.***

OED "assimilate, v. t. ... II: to absorb and incorporate."

This is normally done by integrating the model forward in time, adding observations.

- The model state summarises in an organised way the information from earlier observations.
- It is modified to incorporate new observations, by combining new & old information in a statistically optimal way.





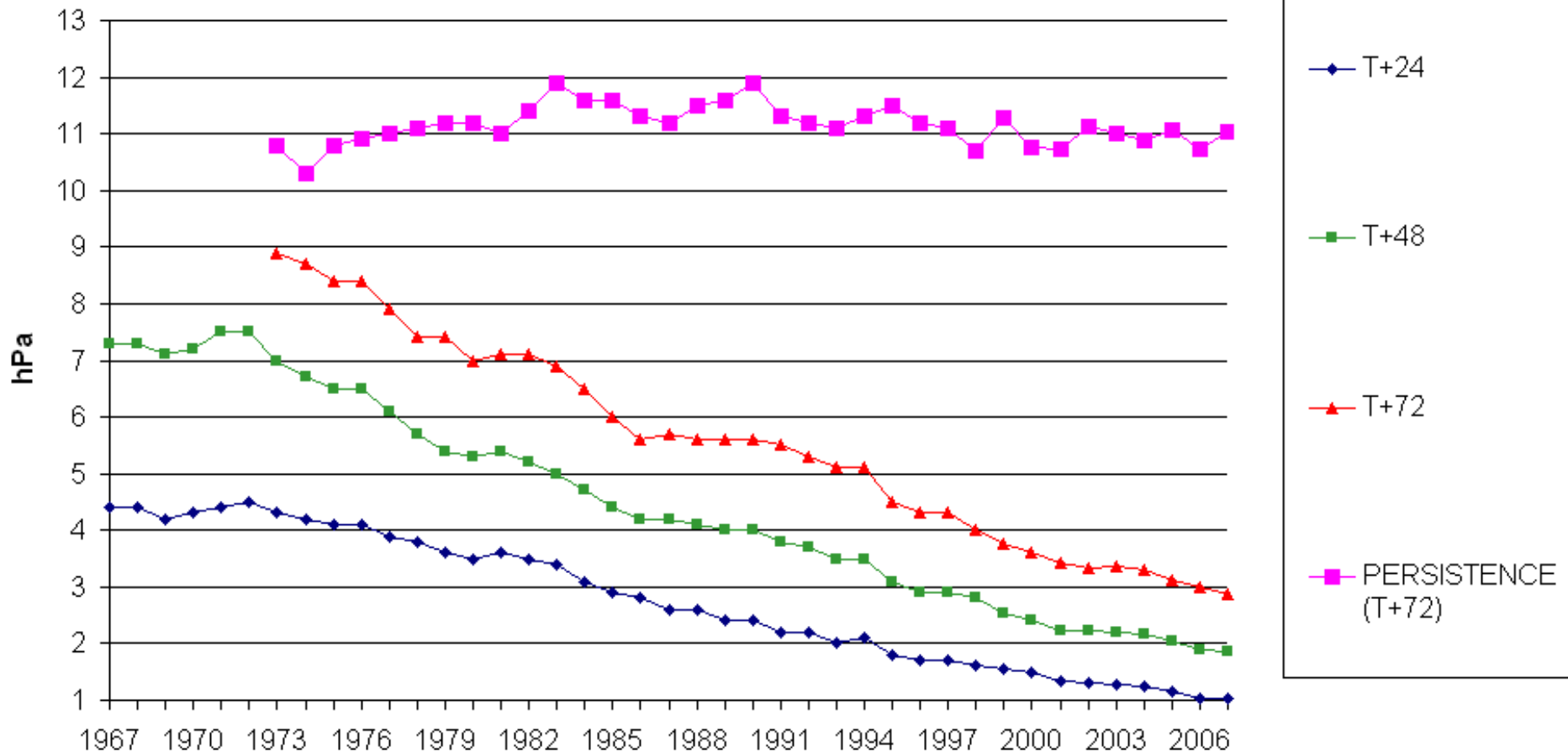
# Data Assimilation, to be good,

1. Needs a good NWP model:
  - to carry information from past observations to current time;
  - to diagnose unobserved quantities via physical modelling relationships.
2. Needs careful statistical-dynamical combination of information:
  - Forecasts are generally more informative than latest observations, yet all the information from each observation should be extracted.
  - Observation networks are incomplete. Information on unobserved variables must be inferred (e.g. from satellite radiances).
  - It is impossible to properly sample error distributions – physical insight is needed to give:
    - a good model of observational variances and biases.
    - a good model of the structure of forecast errors.
3. Advanced Data Assimilation methods also use models to predict the evolution of forecast errors.

# Performance Improvements

“Improved by about a day per decade”

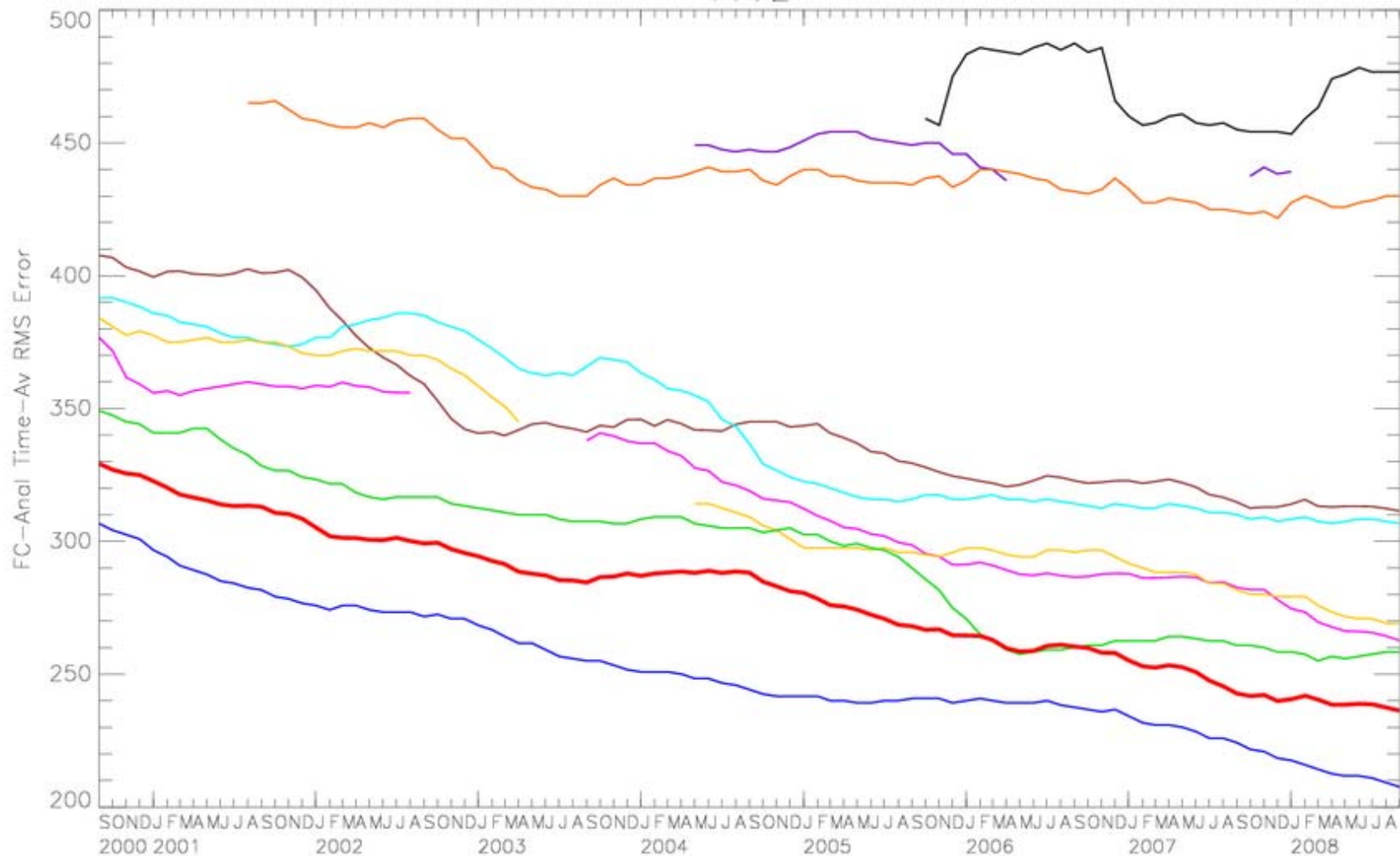
RMS surface pressure error over the NE Atlantic



Mean Sea Level Pressure (Pa): Analysis  
Northern Hemisphere  
12-Month Moving Average

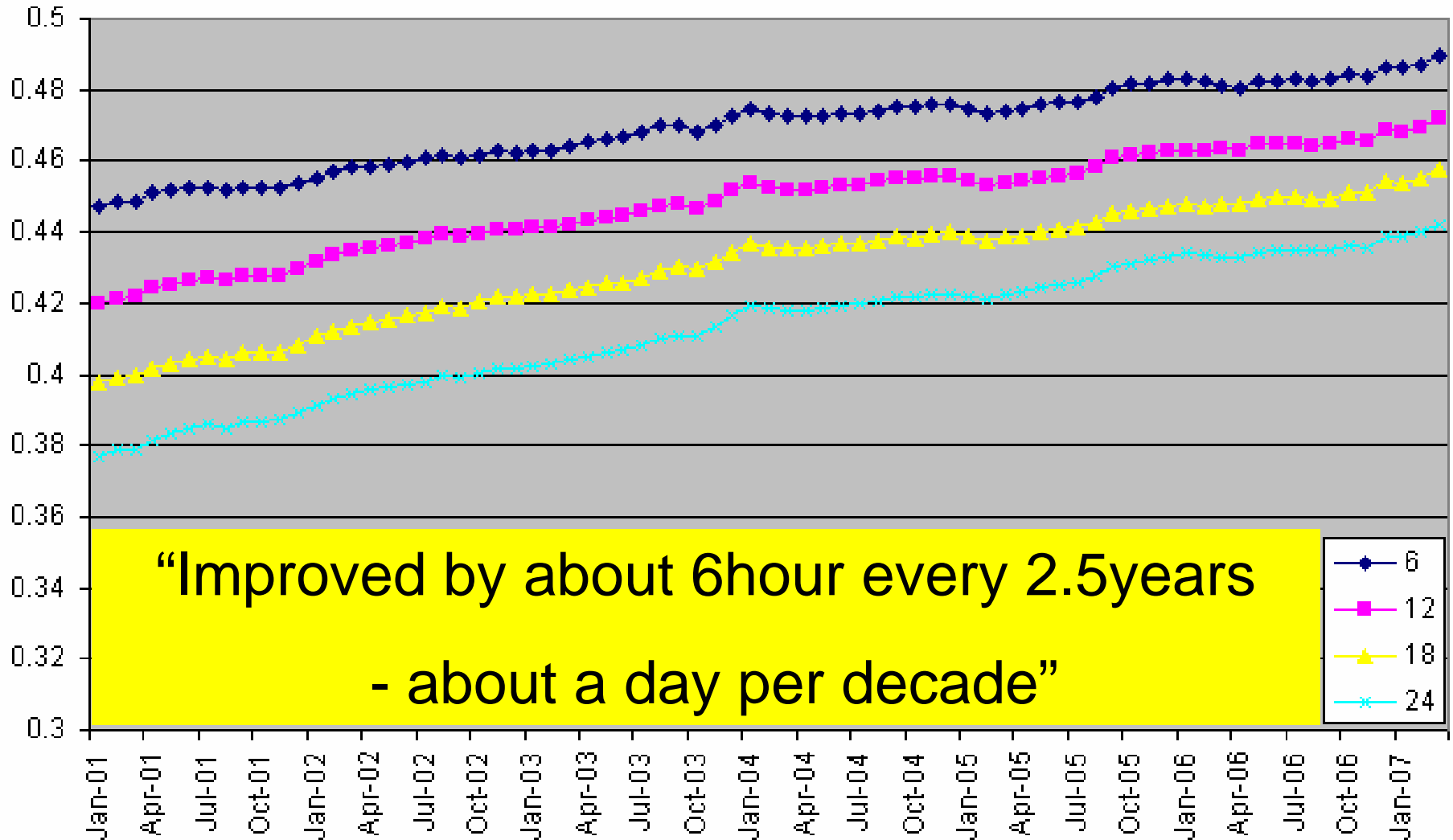
Cases: UK 12Z, ECMWF 12Z, USA MRF 00Z, France 12Z, Germany 12Z, Japan 12Z, Canada 00Z, Australia 12Z, India 00Z, Russia 12Z

T+72





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



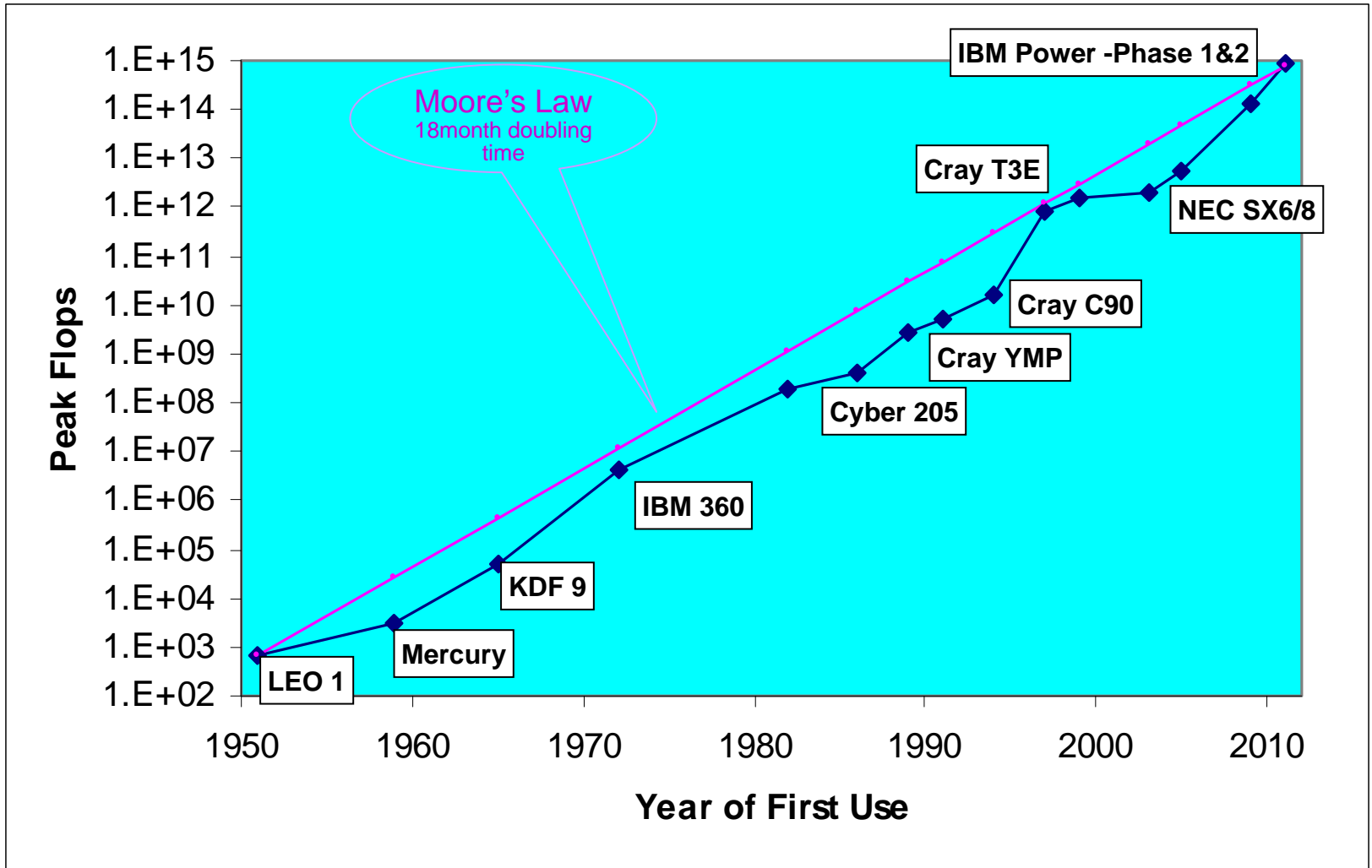


# Importance of forecast model

- A large part of the increase in ***assimilation*** accuracy comes from improvements to the model
- A large part of the increase in model accuracy comes from improvements in resolution
- The resolution has been limited by computer power, so the increase in skill is related to Moore's Law.
- Still true today – a much larger part of planned increases in computer power will be spent on increased resolution than on improved algorithms.



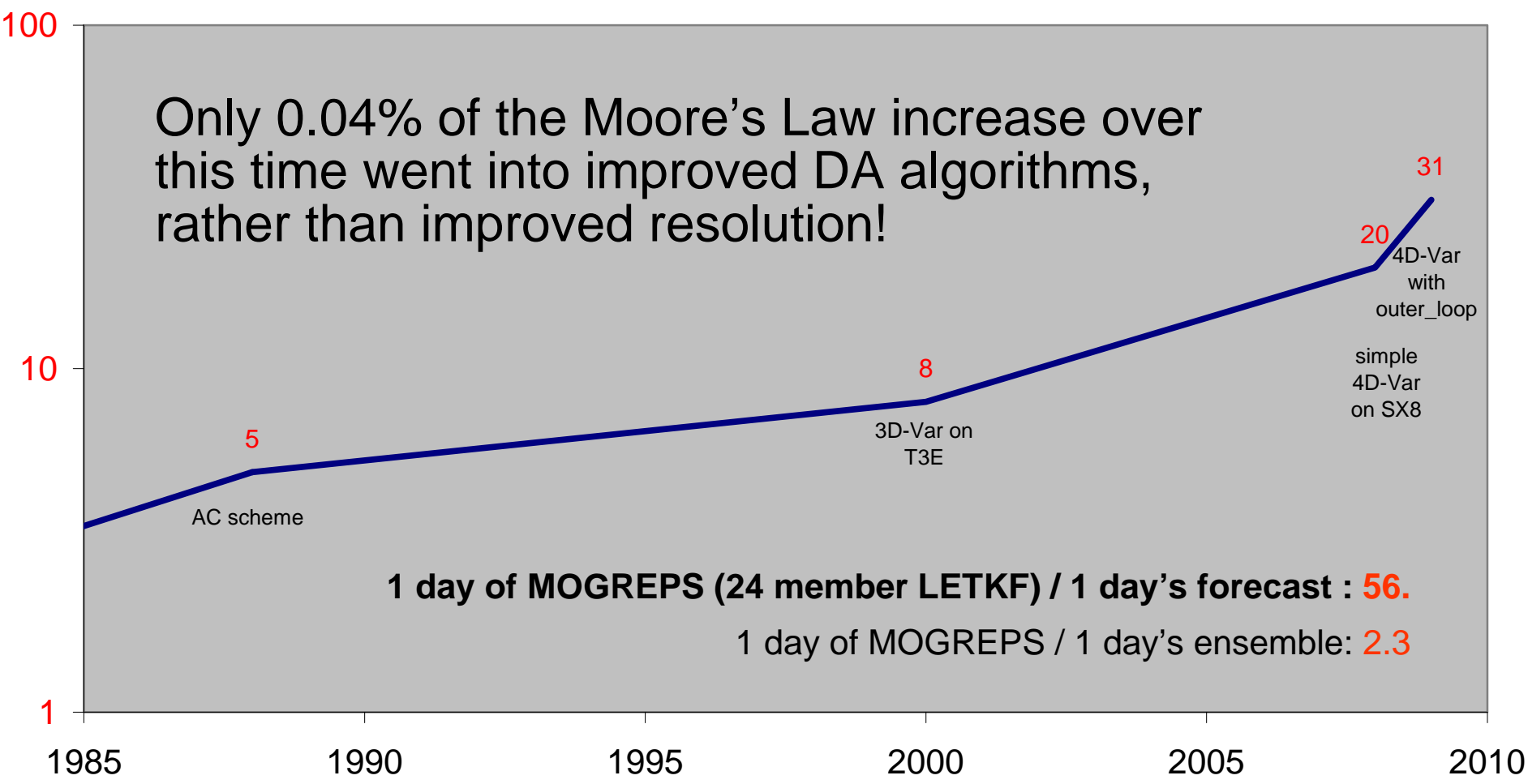
# 60 Years of Met Office Computers





# Ratio of global computer costs: 1 day's DA (total incl. FC) / 1 day's forecast.

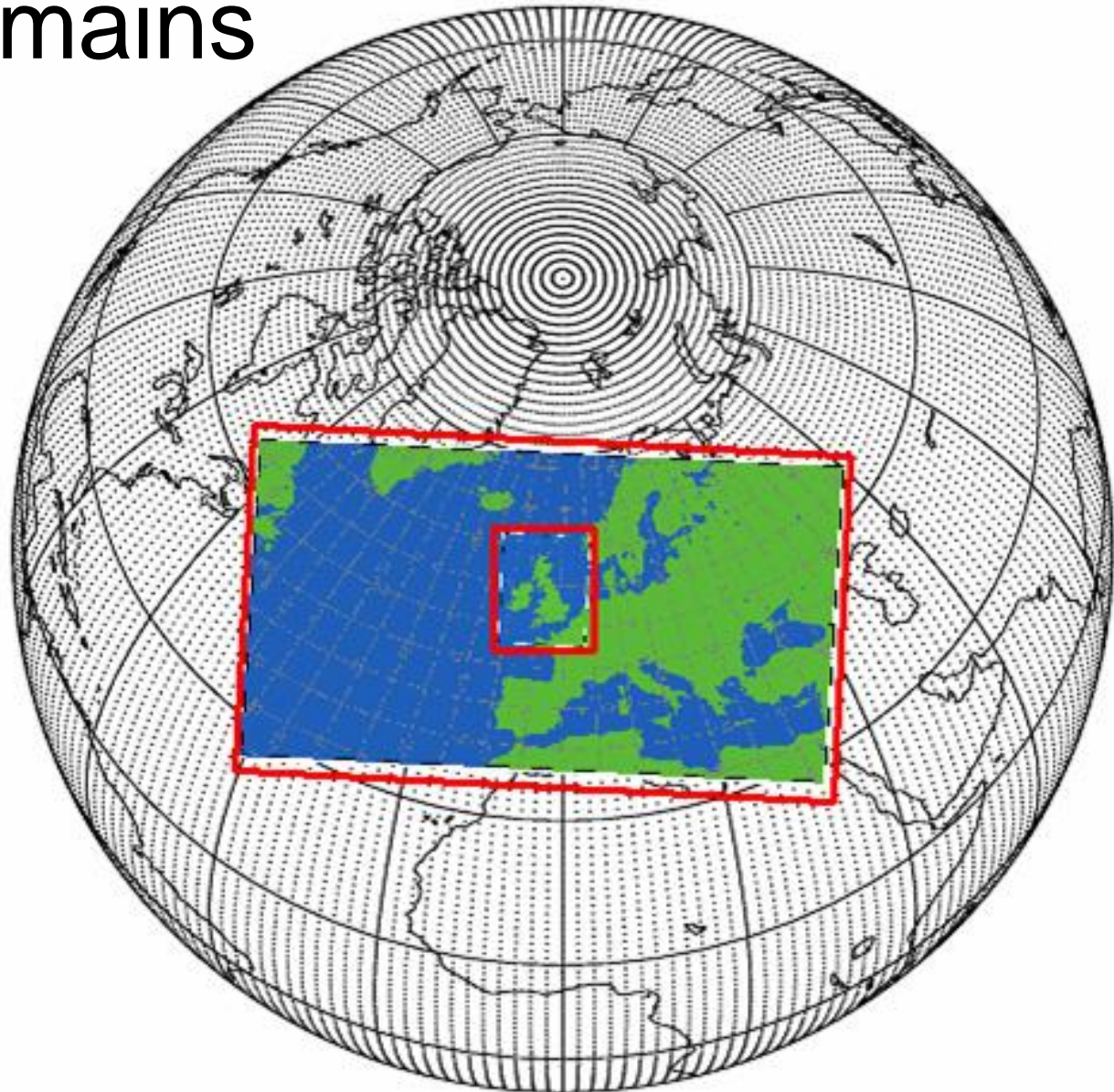
Only 0.04% of the Moore's Law increase over this time went into improved DA algorithms, rather than improved resolution!



**1 day of MOGREPS (24 member LETKF) / 1 day's forecast : 56.**

1 day of MOGREPS / 1 day's ensemble: 2.3

# Global, regional (NAE) and UK domains



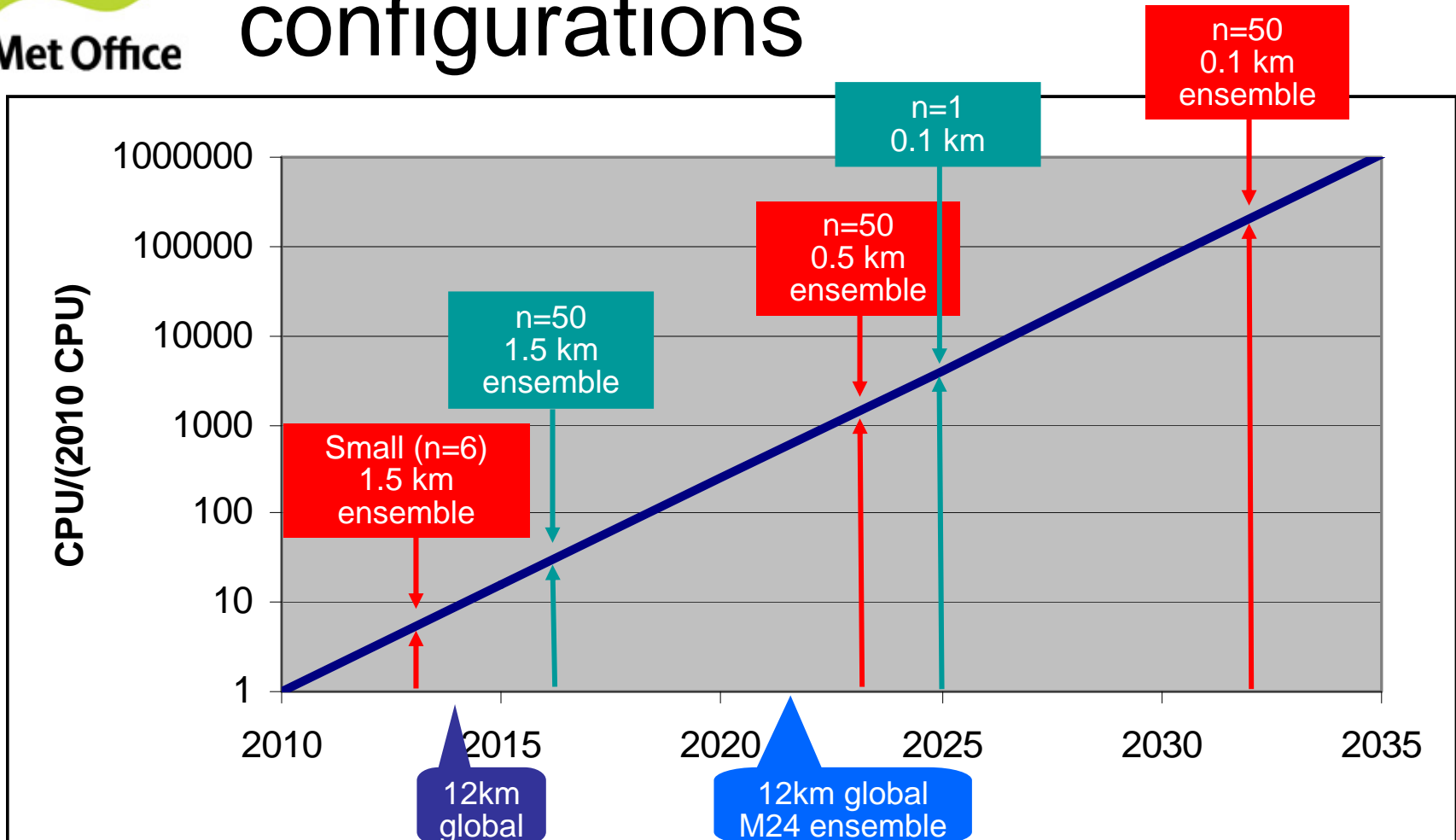


# Met Office plans on new computer ( $\times 6.5$ ) in 2009

- **Global 25km L70** model (was 40km L50)
  - Incremental 4D-Var
  - 60km 24m ETKF ensemble
- **Regional NAE 12km L70** model
  - Incremental 4D-Var (with outer-loop, cloud & ppn)
  - 16km 24m L70 ETKF ensemble (was 24km L38)
- **UK 1.5km** model (stretched) (was 4km)
  - 3D-Var + nudging of ppn & cloud
  - Ensemble driven by NAE perturbations (experimental)
  - Small domain 4D-Var RUC (experimental)



# Possible UK model configurations



Assumes Moore's Law continues with 15 month doubling time

(It probably won't!)

Pete Clark



# Data Assimilation, to be good,

## 1. Needs a good NWP model:

- to carry information from past observations to current time;
- to diagnose unobserved quantities via physical modelling relationships.

## 2. Needs statistical-dynamical combination of information:

- Forecasts are generally more informative than latest observations, yet all the information from each observation should be extracted.
- Observation networks are incomplete. Information on unobserved variables must be inferred (e.g. from satellite radiances).
- It is impossible to properly sample error distributions – physical insight is needed to give:
  - a good model of observational variances and biases.
  - a good model of the structure of forecast errors.

## 3. Advanced Data Assimilation methods also use models to predict the evolution of forecast errors.



Met Office

# Use of incomplete historical data to infer the present state of the atmosphere

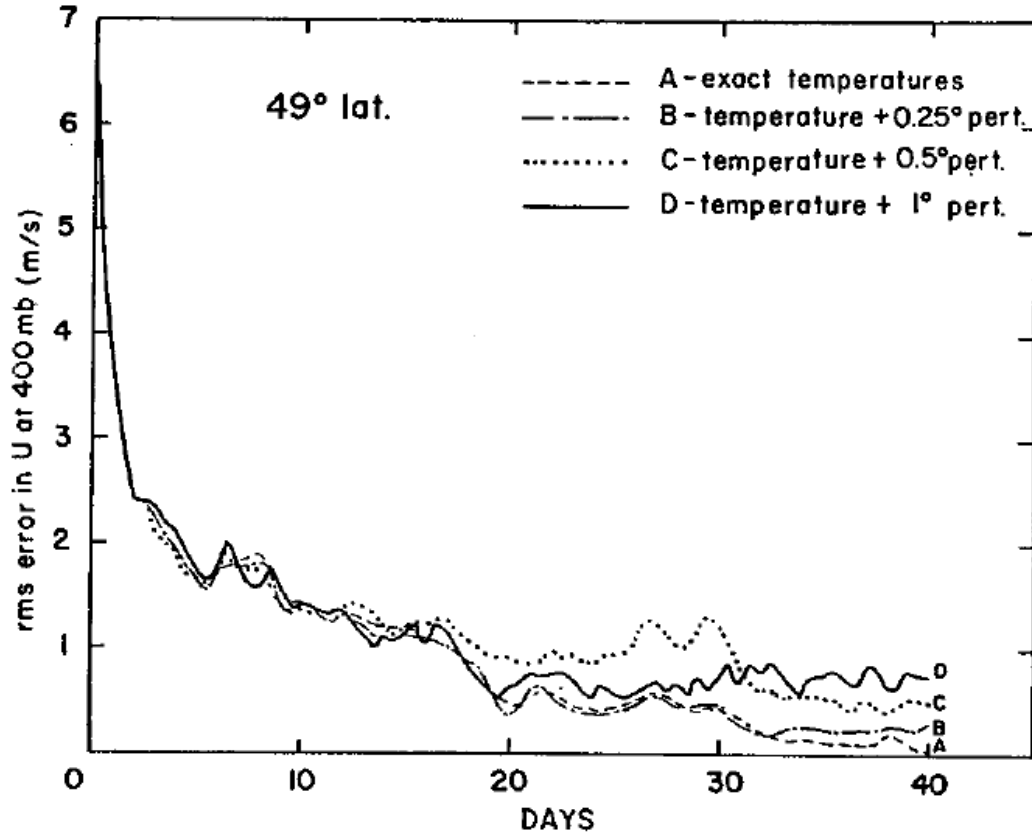


FIG. 1. The rms error in zonal wind ( $\text{m sec}^{-1}$ ) at 400 mb at  $49^\circ$  latitude, in cases where temperatures with random error perturbations of 0, 0.25, 0.5 and  $1^\circ\text{C}$  are inserted every 12 hr at all grid points.

**OSSE** using  
Mintz-Arakawa model :  
 $9^\circ \times 7^\circ \times 2$  levels.

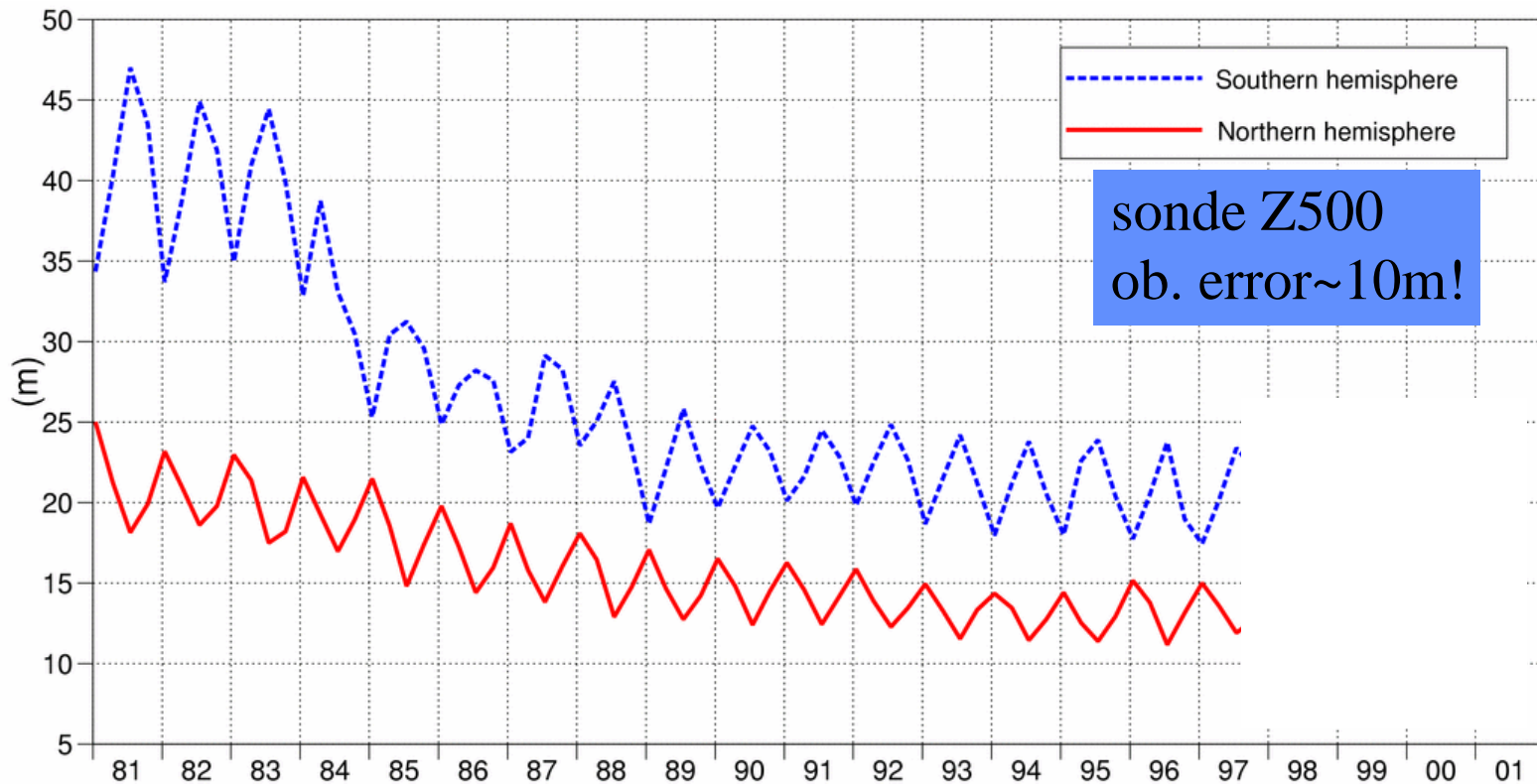
✓ *Satellite sounders could become a major part of the global OS.*

✗ *Direct insertion of satellite temperature retrievals is a viable DA method.*



Met Office

# Evolution of the r.m.s day -one 500hPa height forecast error 1981-2001



*Simmons & Hollingsworth, 2002*



**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.

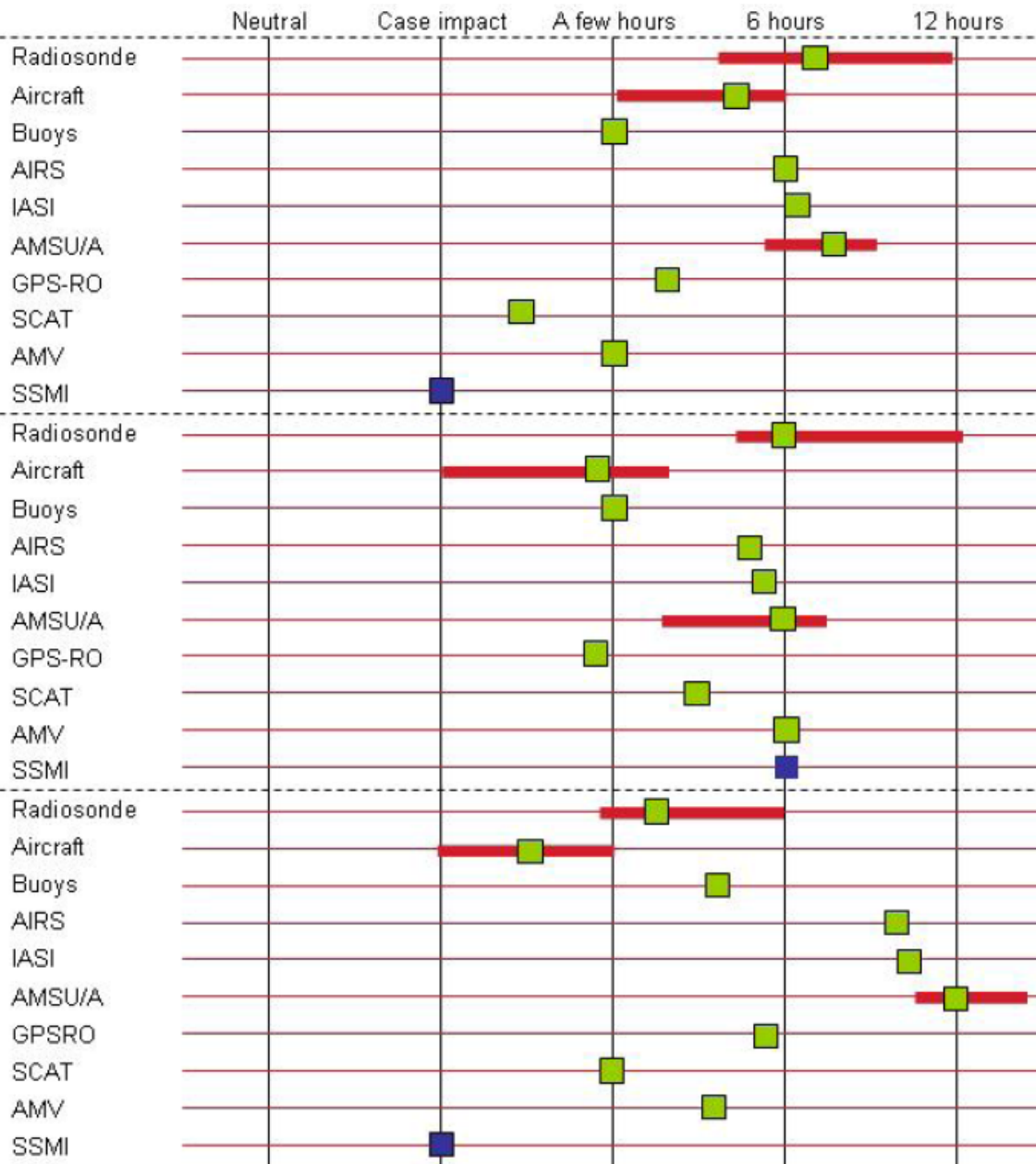
*Fourth WMO Workshop on the Impact of Various Observing Systems on NWP.*

*Geneva, Switzerland, 19-21 May 2008*

**Northern Hemisphere Extra-tropics**

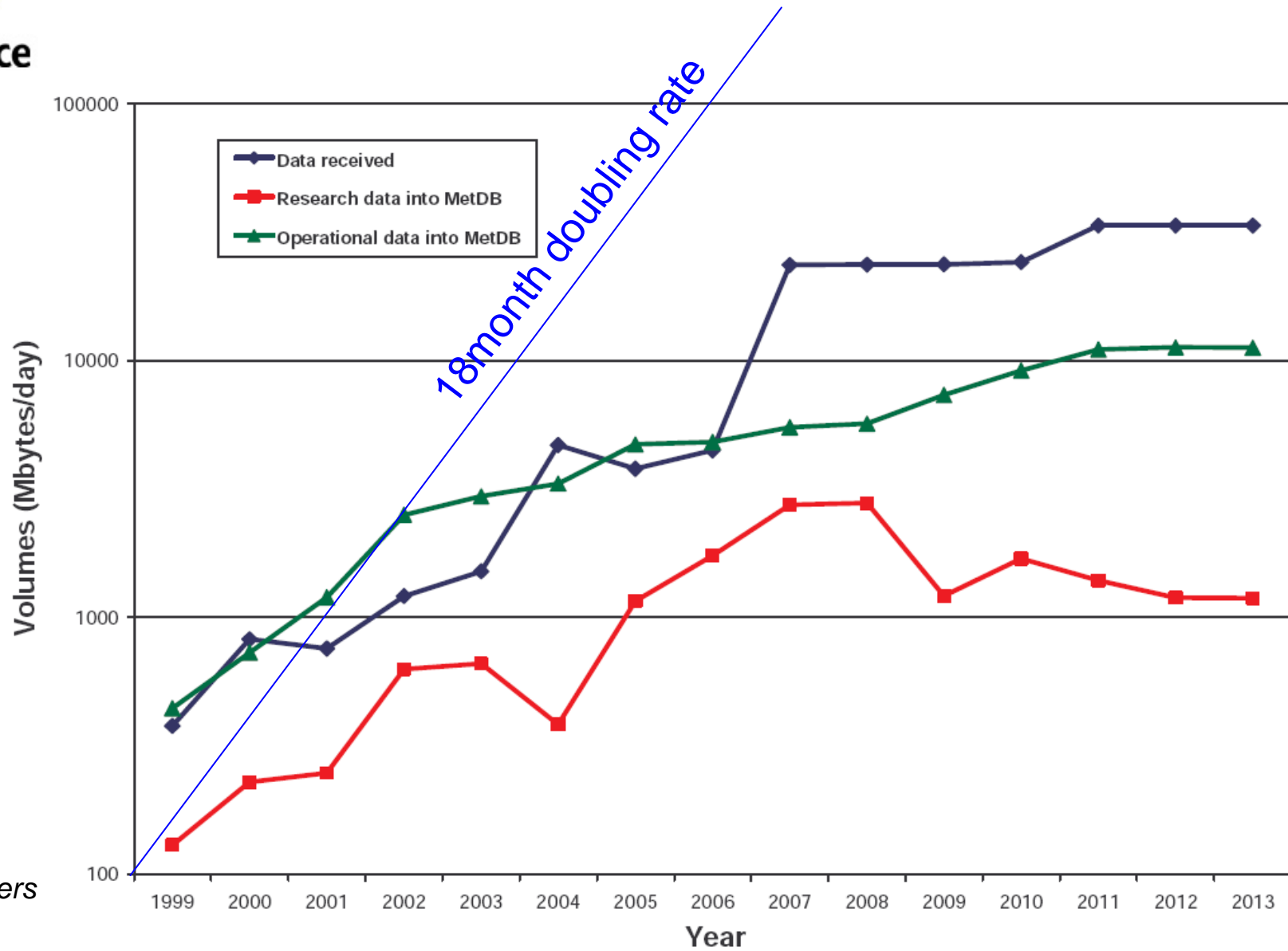
**Tropics**

**Southern Hemisphere Extra-tropics**





# Daily satellite data volumes



Roger Saunders



# Comments on observational data volumes and impacts

- All observation types are important – each worth about 1-2 years' typical improvements (i.e. enough to overtake other NWP centres.)
- These positive impacts required attention to bias correction & QC. Further improvements depend even more on this.
- Much remains to be done to use satellite data:
  - Full resolution,
  - Use of observations of cloud and precipitation.
- The flood of global satellite data has stopped growing as fast (well inside Moore's Law), ***but radar data are coming along for the convective scale.***



# Data Assimilation, to be good,

## 1. Needs a good NWP model:

- to carry information from past observations to current time;
- to diagnose unobserved quantities via physical modelling relationships.

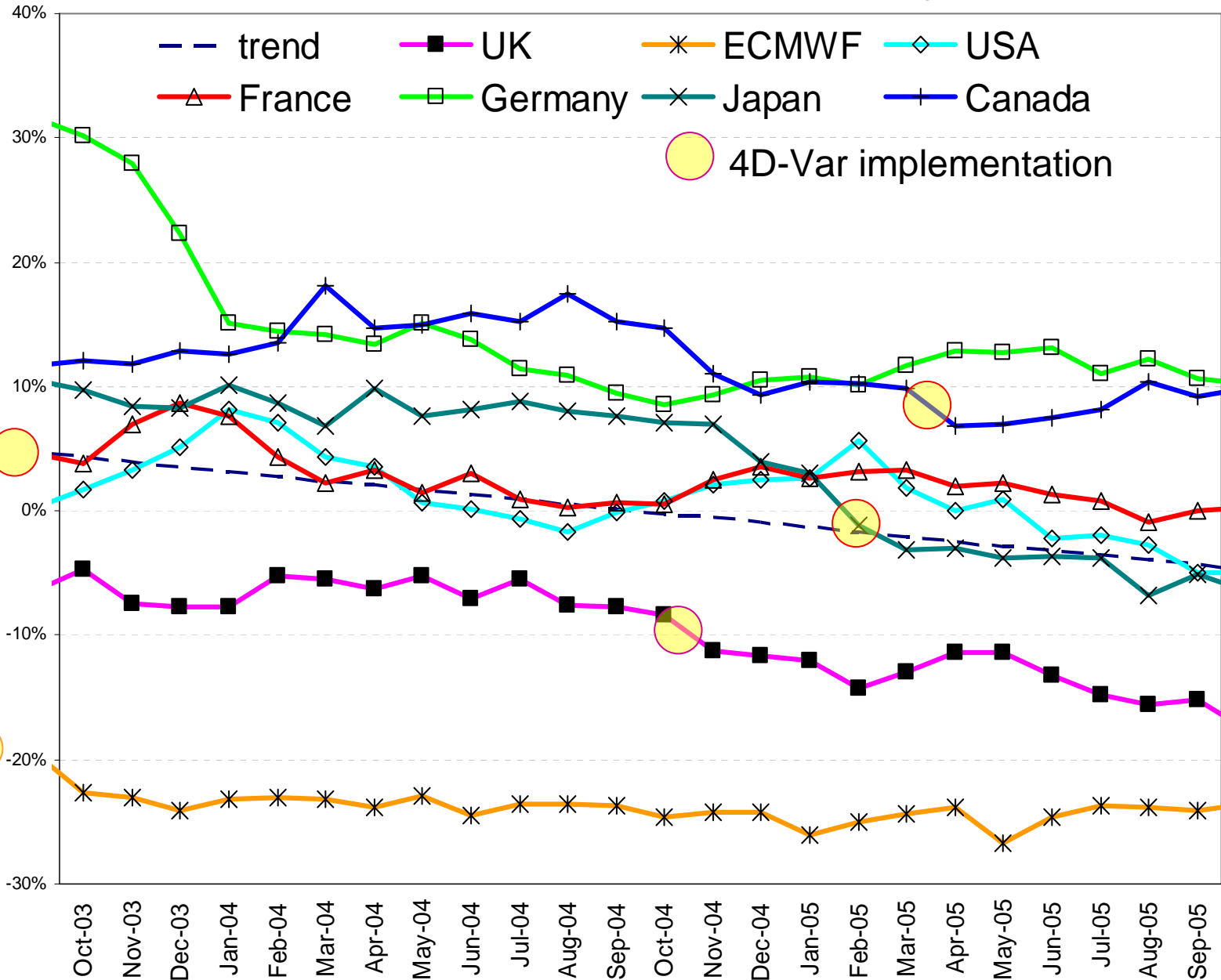
## 2. Needs statistical-dynamical combination of information:

- Forecasts are generally more informative than latest observations, yet all the information from each observation should be extracted.
- Observation networks are incomplete. Information on unobserved variables must be inferred (e.g. from satellite radiances).
- It is impossible to properly sample error distributions – physical insight is needed to give:
  - a good model of observational variances and biases.
  - a good model of the structure of forecast errors.

## 3. Advanced Data Assimilation methods also use models to predict the evolution of forecast errors.

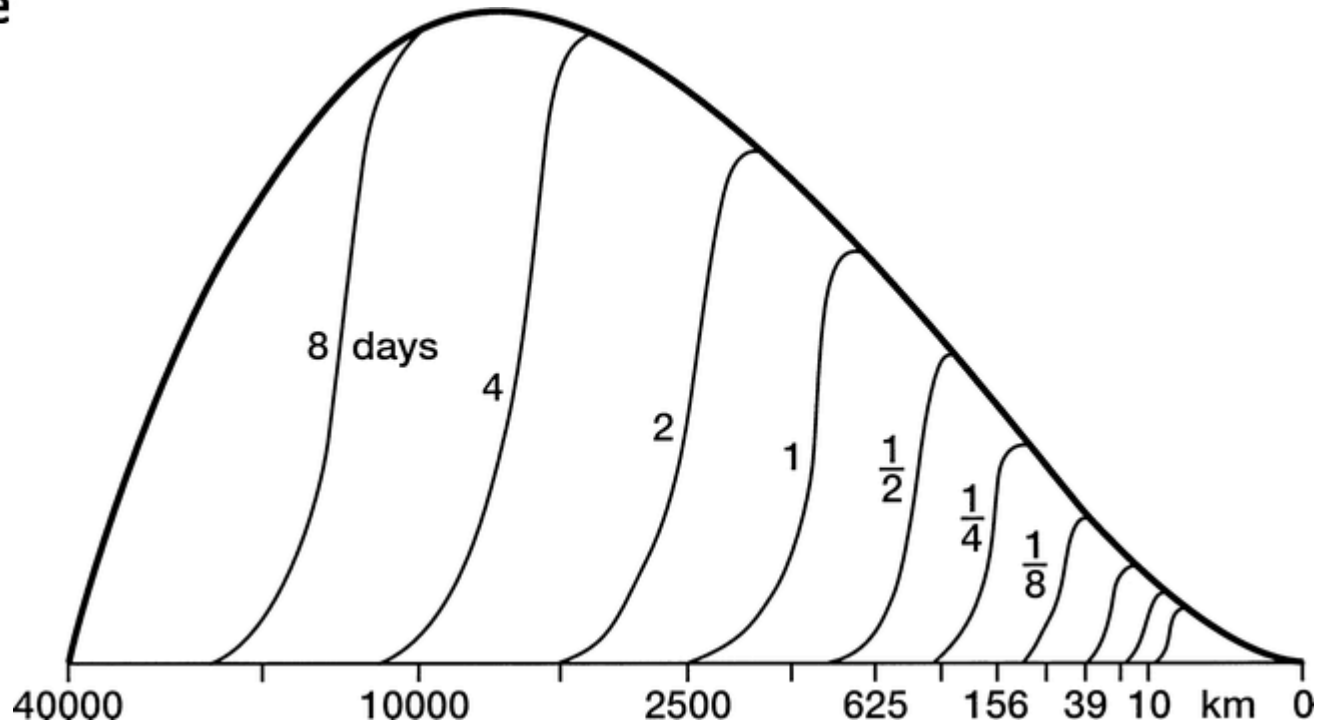


# RMS errors with mean intra-annual variability removed





# Time & Space Scales



Growth of errors initially confined to smallest scales, according to a theoretical model Lorenz (1984) . Horizontal scales are on the bottom, and the upper curve is the full atmospheric motion spectrum. (from Tribbia & Baumhefner 2004).



# Discussion of Scales

- What scales can we assimilate?
  - EKF 4D-Var & EnKF depend on linear approximation, so we can only handle scales which are well enough known for errors to behave linearly.
    - Globally, convective scales are not well known.
    - Locally, doppler radars can determine convective scales.
- What scales do we need to assimilate together?
  - With the current observation network, global DA uses information from about the past 5 days.
  - Global DA has continued to improve with model resolution (to ~25km).
  - Because of nonlinearity, scale separation is a poor approximation, especially concerning precipitation (e.g. fronts, convection). Nesting boundaries do not behave well!
  - In the UK, **most** forecast errors for convection are partly due to errors in large-scale.
  - Much of the information in imagery sequences comes from the perceived movement of small-scale features in the larger-scale flow.



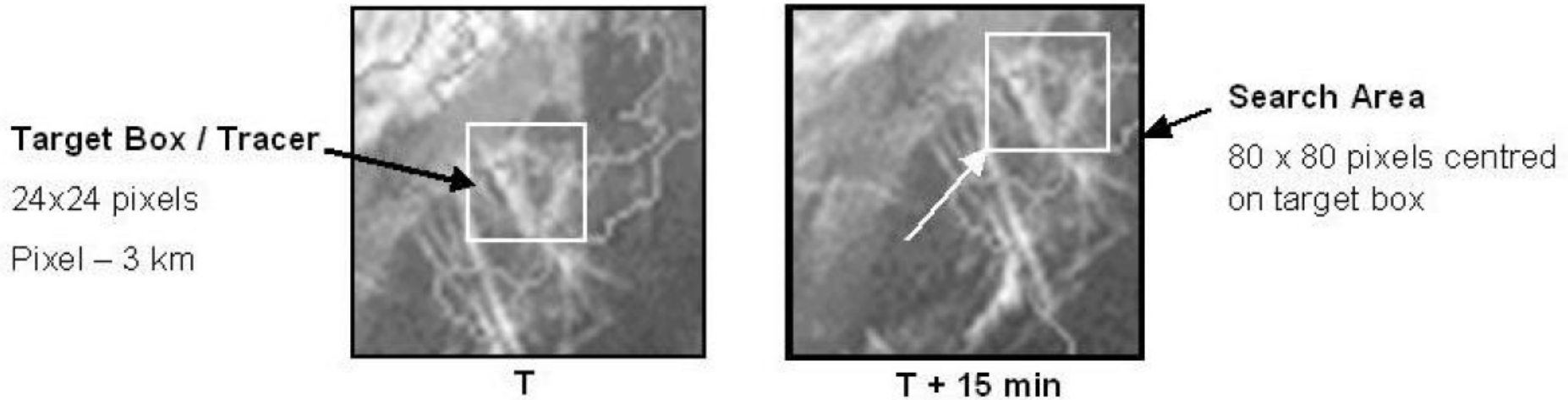
# Information content of imagery sequences

- Humans can make reasonable forecasts based on imagery alone (satellite or radar): information scarcely used in NWP.
- Time-sequences aid the interpretation of images.
- Some important information is multi-scale; details at high-resolution are used to recognise patterns whose larger-scale movements are significant.



# Current methods for assimilating imagery

- AMVs (aka cloud track winds) produced from the motion of patterns seen in  $\sim 32^2$  pixels.
- Satellite sounders give coarse-grained imagery repeated every  $\sim 6$  hours.
- Met Office recently implemented 4D-Var assimilation of cloud.
- Radar radial winds and reflectivity are assimilated in research EnKF and 4D-Var systems.



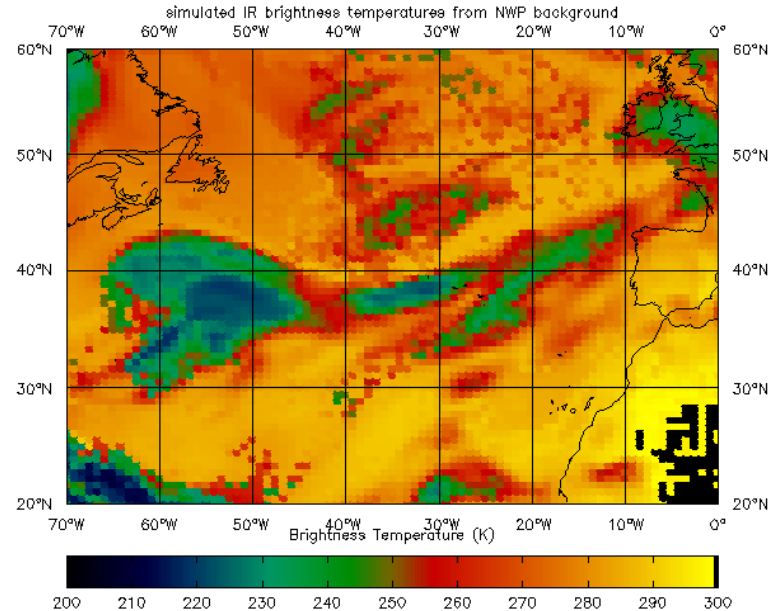
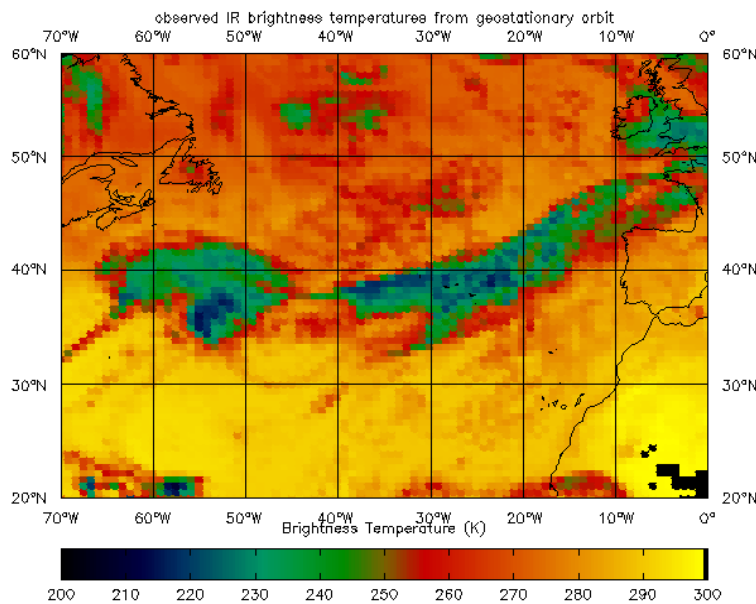
*Figure 1: An illustration of the AMV tracking step for Meteosat-9 IR AMVs. The location of the target in the later image is determined by best match of the individual pixel counts of the target with all possible locations of the target in the search area using cross-correlation in the Fourier domain. The wind vector is taken as the displacement between the locations of the target boxes in the two images.*

- I am not suggesting we could replace AMVs by 4DDA in the near future!
- However they provide an example of demonstrated useful information from imagery sequences, which a method should in principle be able to extract.
- 4DDA methods could, in theory, improve on current AMV techniques in allowing for development and dynamical coupling of features.



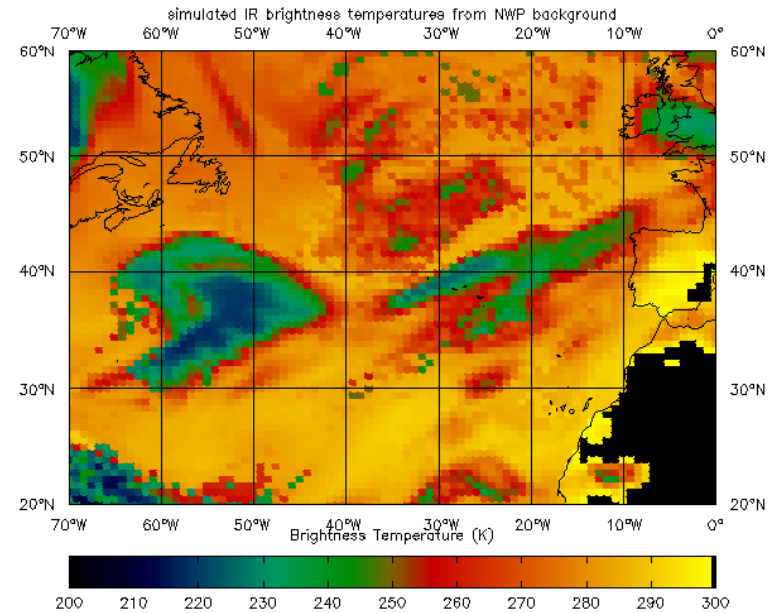
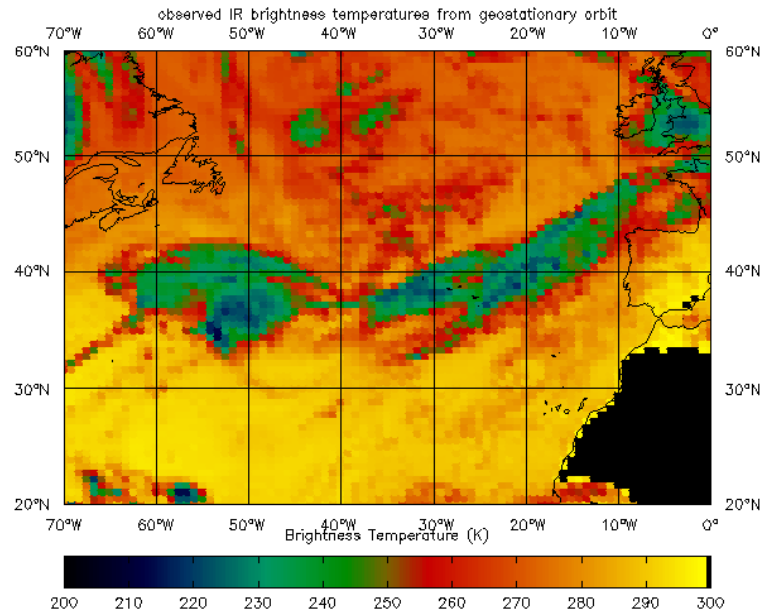
# Comparison of observed and modelled cloud

9Z 13-10-2002

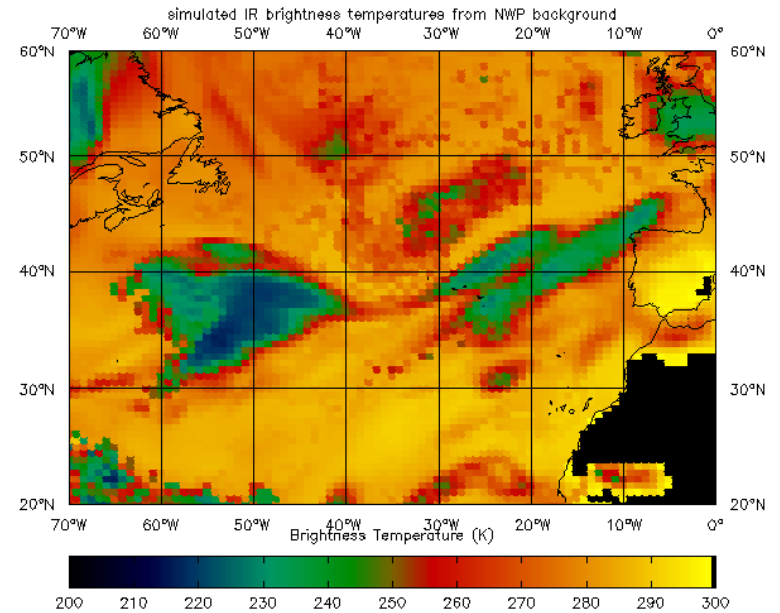
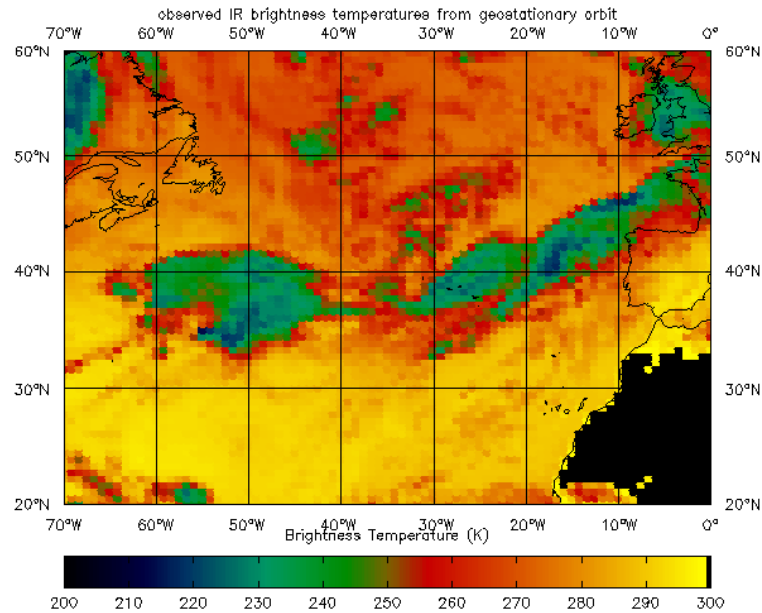


Samatha Pullen

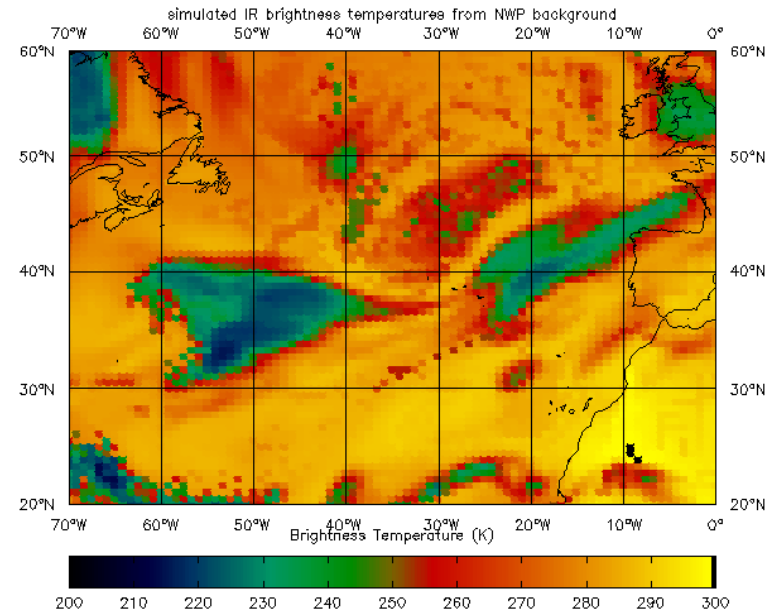
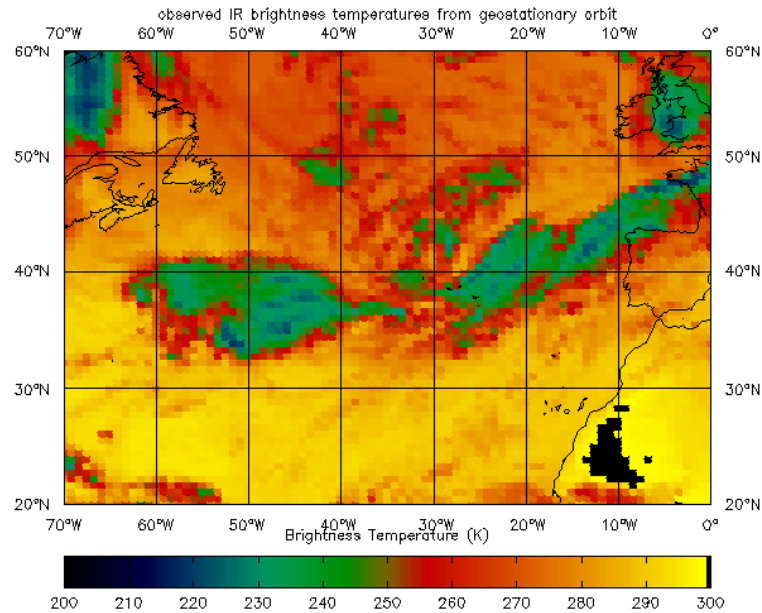
12Z 13-10-2002



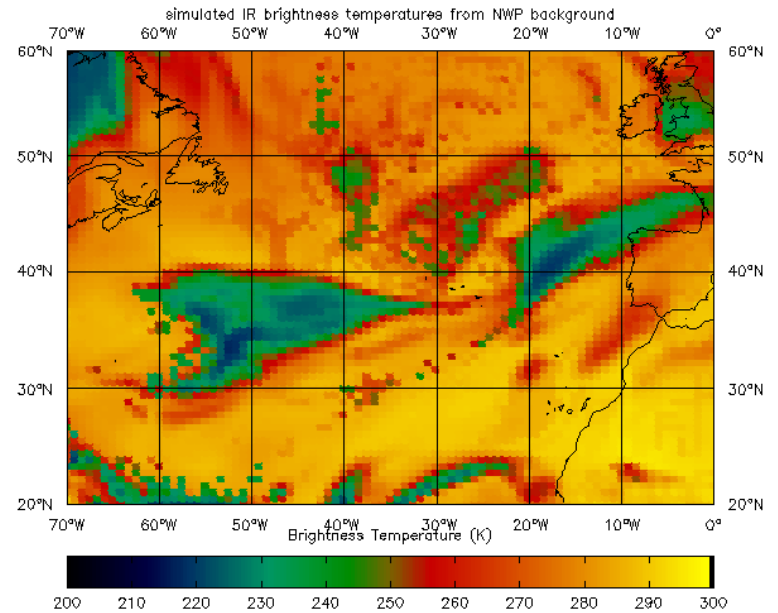
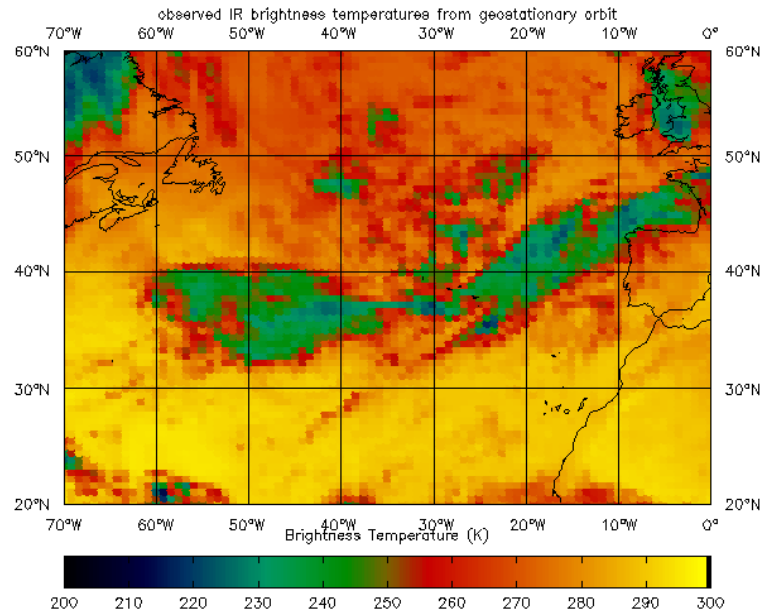
15Z 13-10-2002



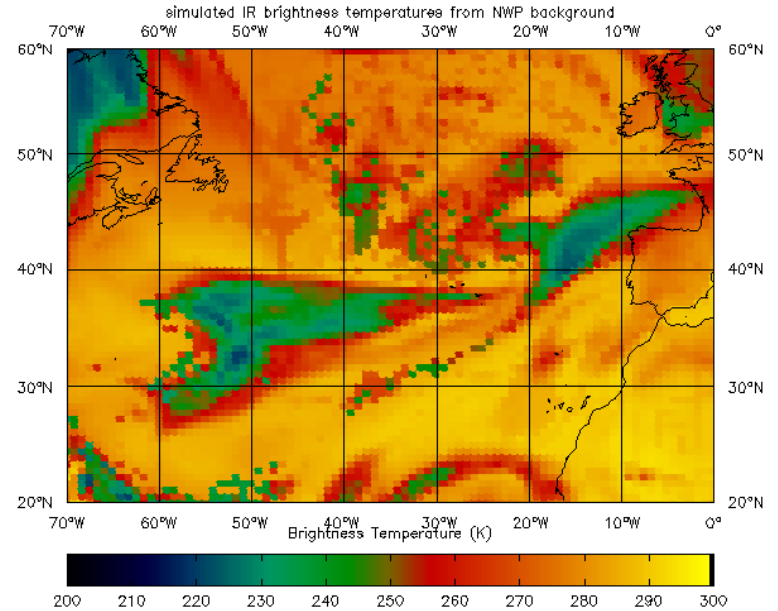
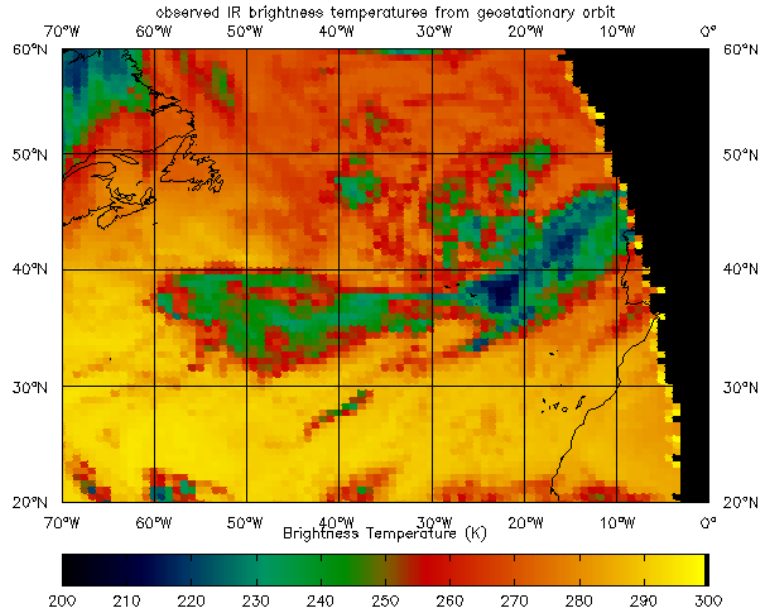
18Z 13-10-2002



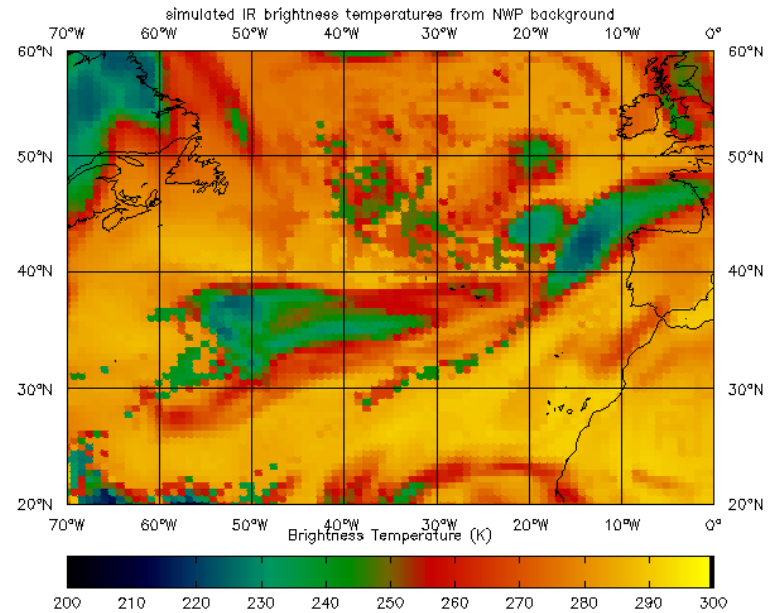
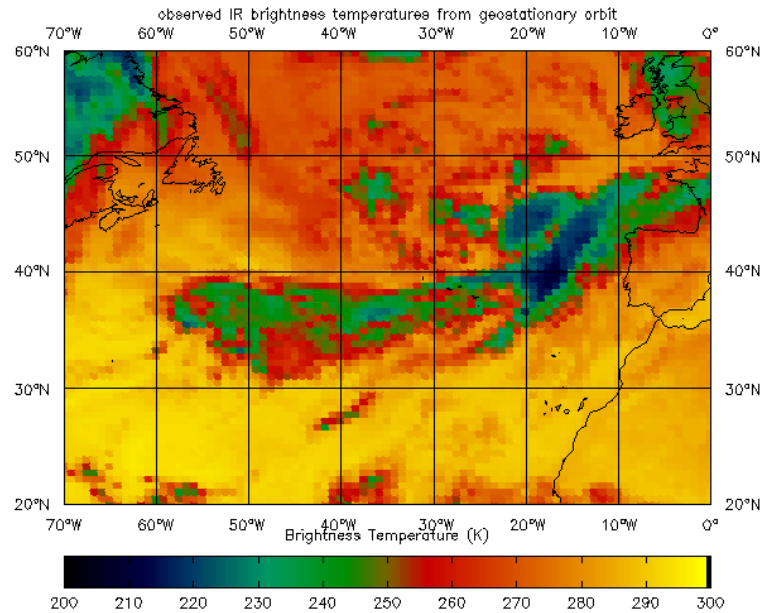
21Z 13-10-2002



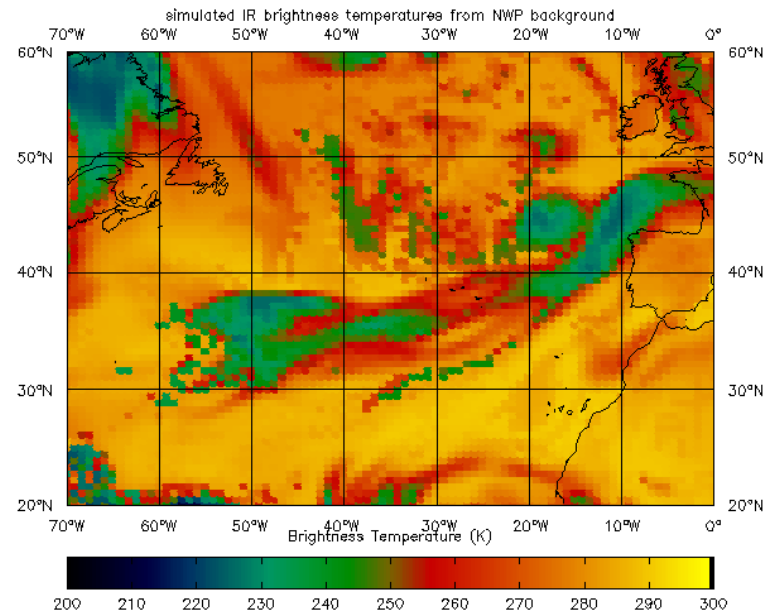
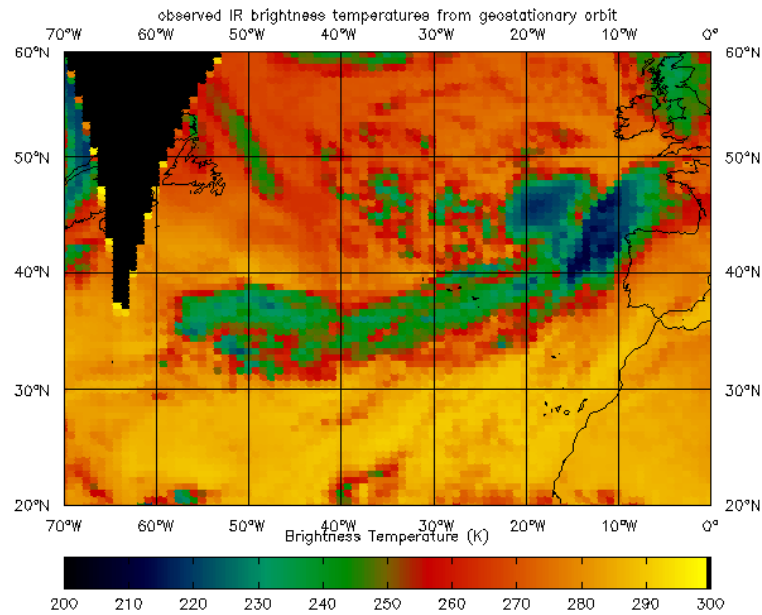
0Z 14-10-2002



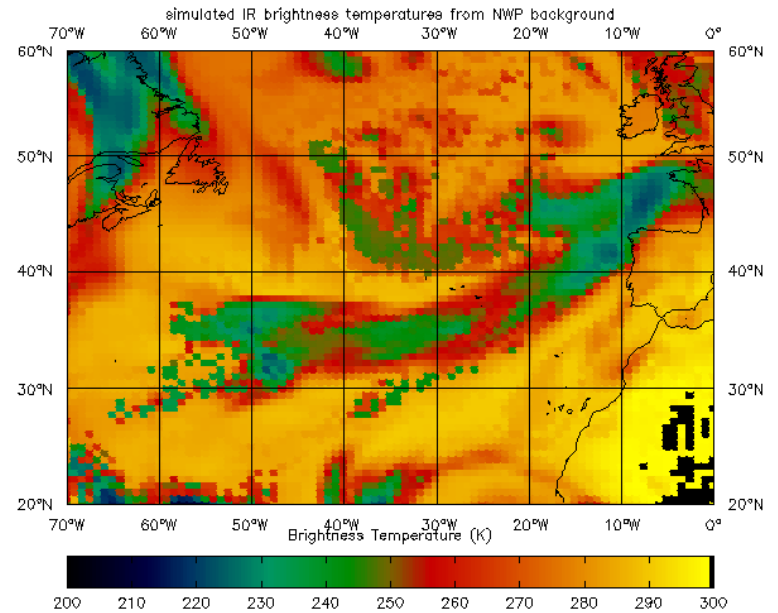
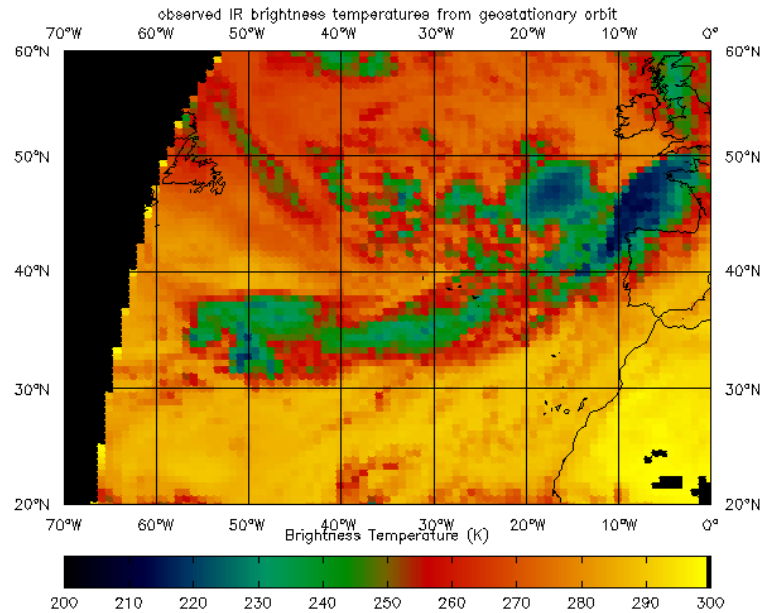
3Z 14-10-2002



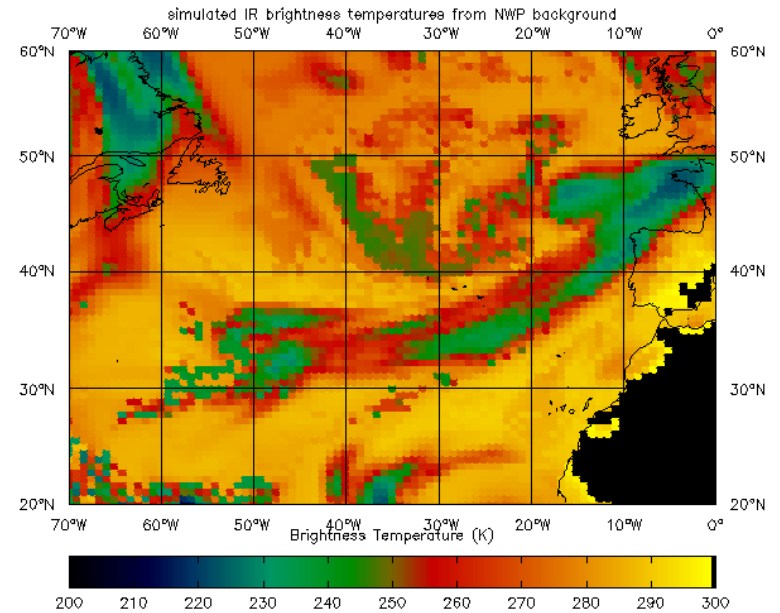
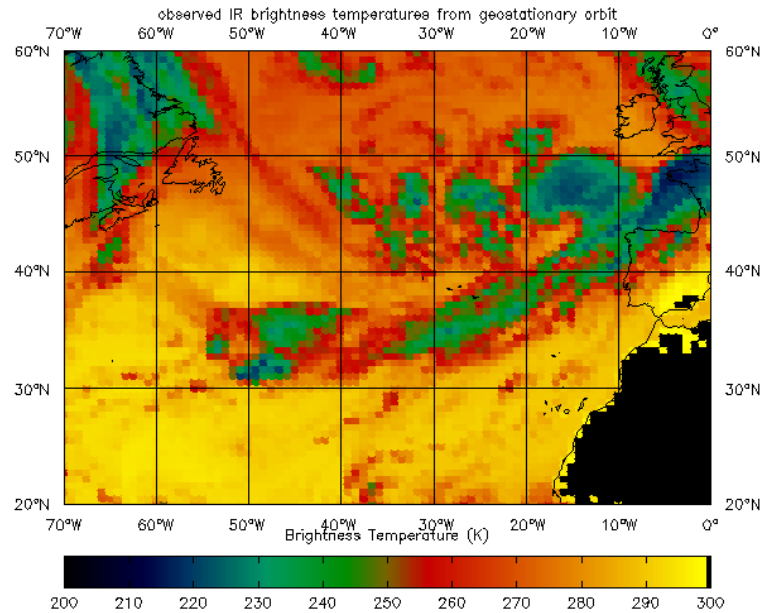
6Z 14-10-2002



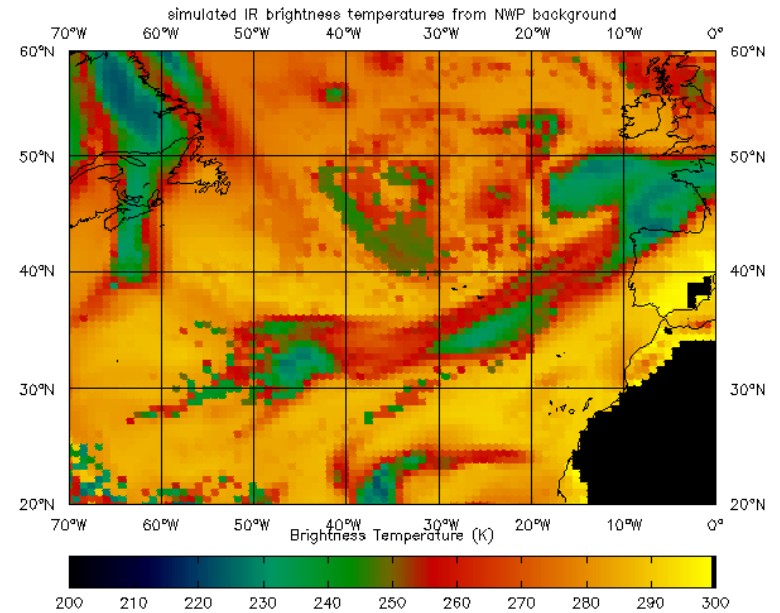
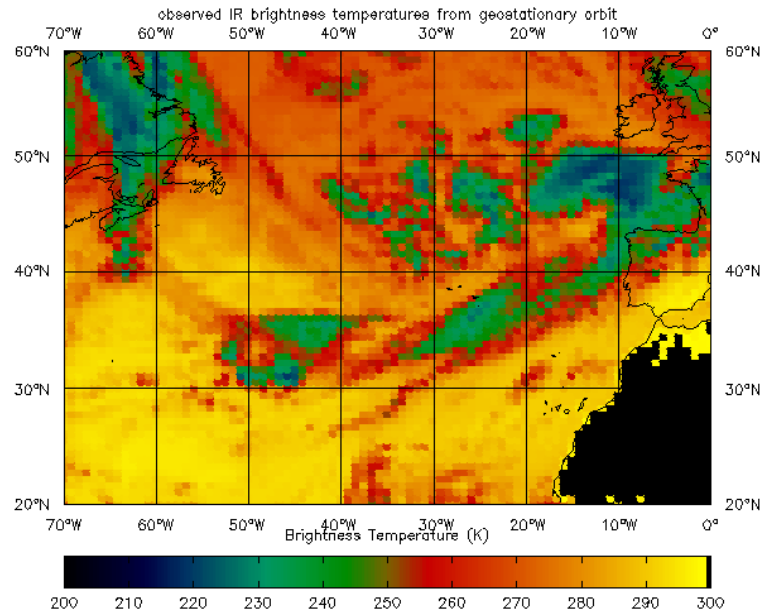
9Z 14-10-2002



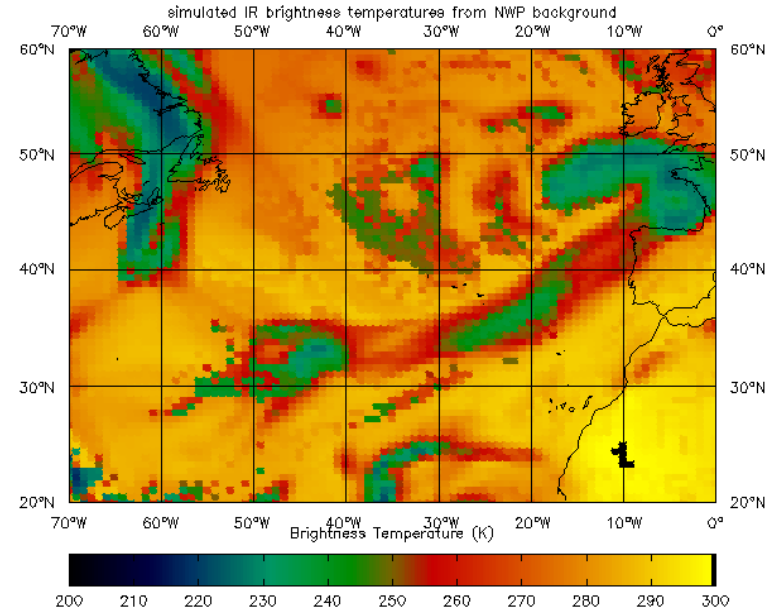
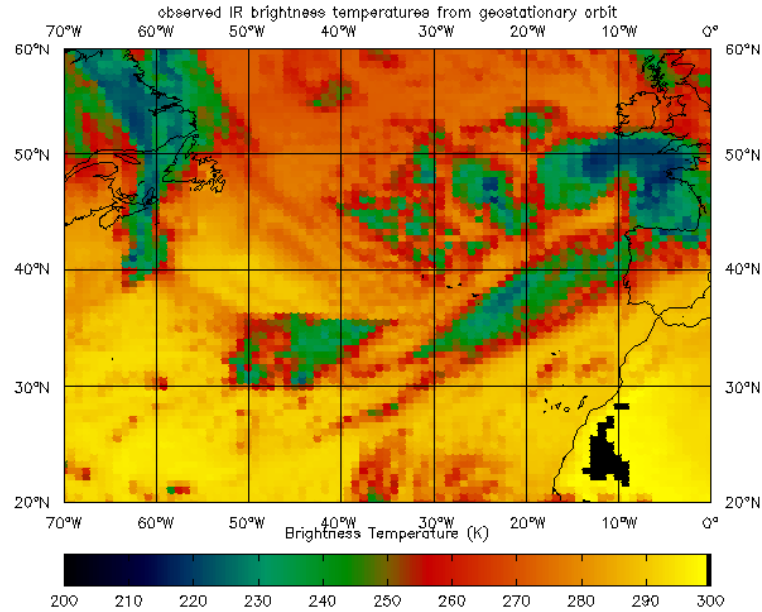
12Z 14-10-2002



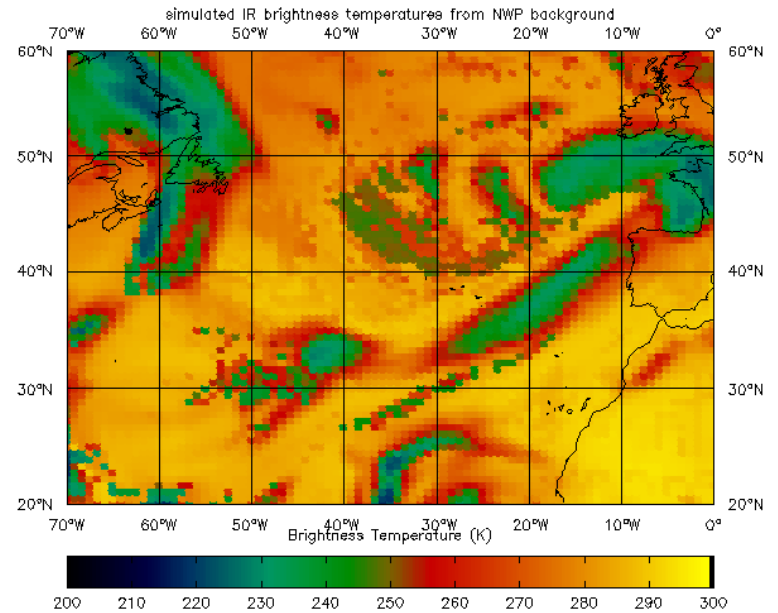
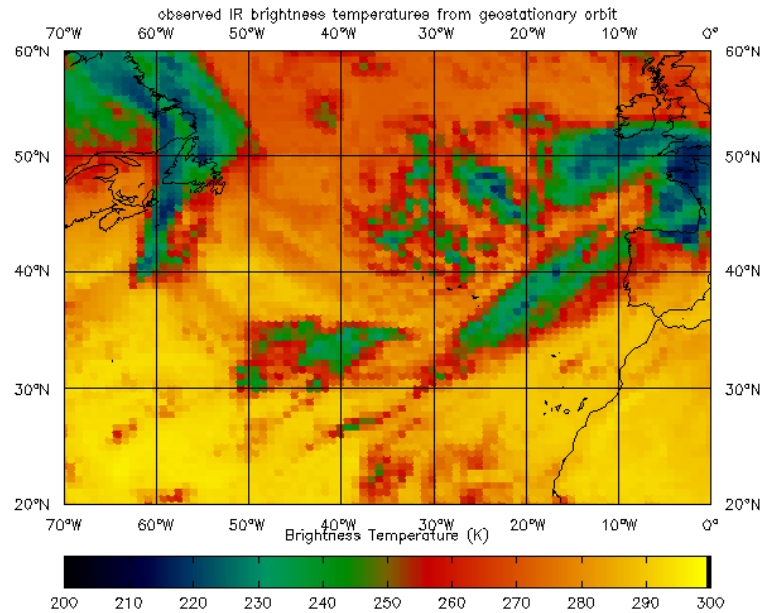
15Z 14-10-2002



18Z 14-10-2002



21Z 14-10-2002





# Equations for tracer advection

$$\frac{Dm}{Dt} = S$$

$$\frac{\partial m}{\partial t} + \nabla(\mathbf{u} m) = S$$

Determining  $\mathbf{u}$  &  $m$  simultaneously is a nonlinear problem.

$$\frac{\partial m'}{\partial t} + \mathbf{u} \bullet \nabla m' + \mathbf{u}' \bullet \nabla m = S'$$

In the linearised equations,

changes to the wind depend on the gradient of the linearisation state  $m$ ,  
biases in observations or model  $S'$  can change the wind.



Met Office

# Nonlinear 4D-Var

OPTIMAL NONLINEAR OBJECTIVE ANALYSIS

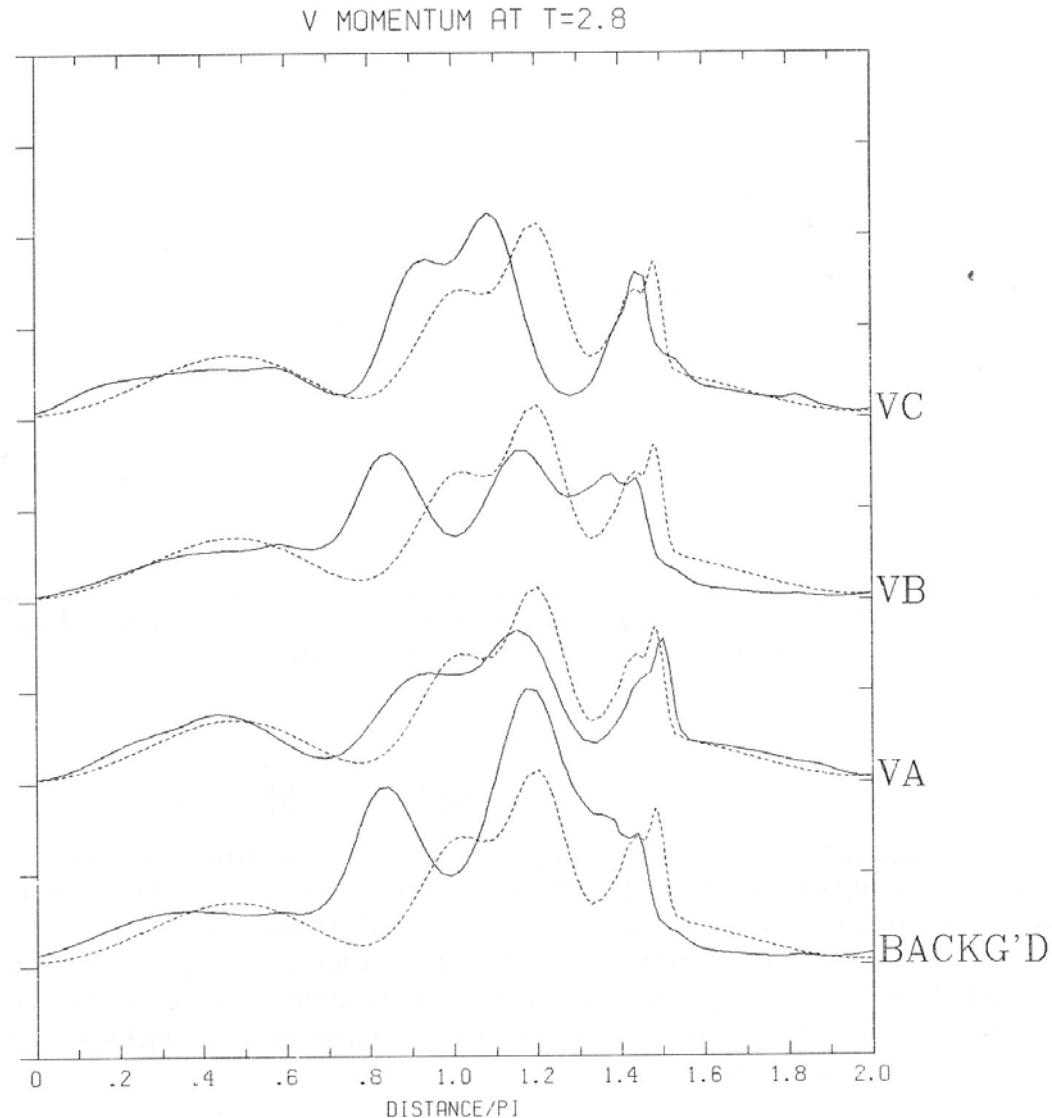
Lorenç 1988 showed that nonlinear 4D-Var of tracer obs at two times in a shallow water model improved forecast.

Cycled 3D-Var of tracer at two times

3D-Var of tracer at one time

4D-Var of tracer at two times

Forecast from background



# 4D-Var “retrieved” winds

T42L19, 24hr, adiabatic, not incremental, no  $J_b$

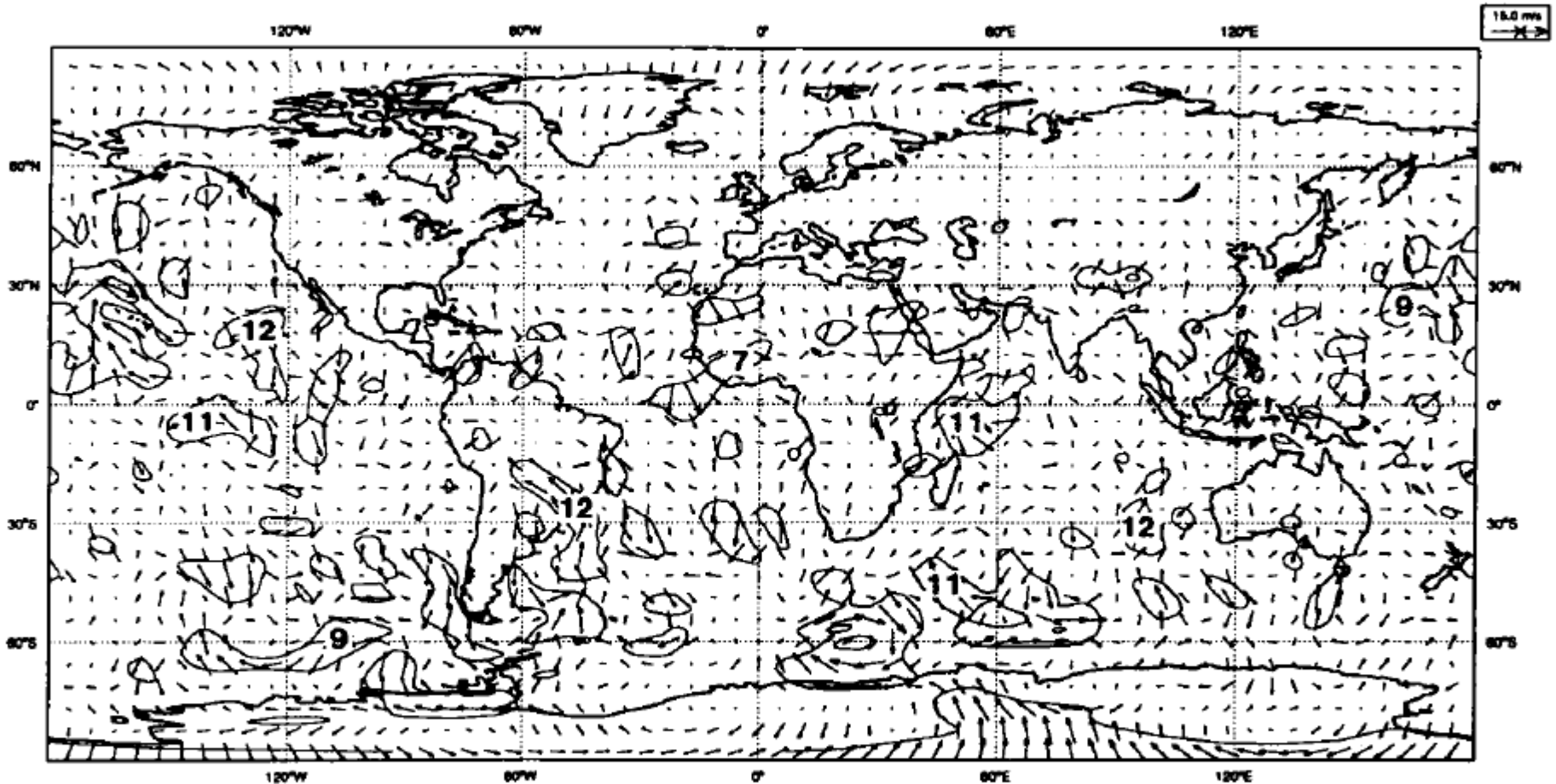


Figure 16. As Fig. 15, but for the wind field. Wind speed is contoured with an interval of 5 m s<sup>-1</sup>.



# 4D-Var “retrieved” winds

648

E. ANDERSSON *et al.* QJ 1994

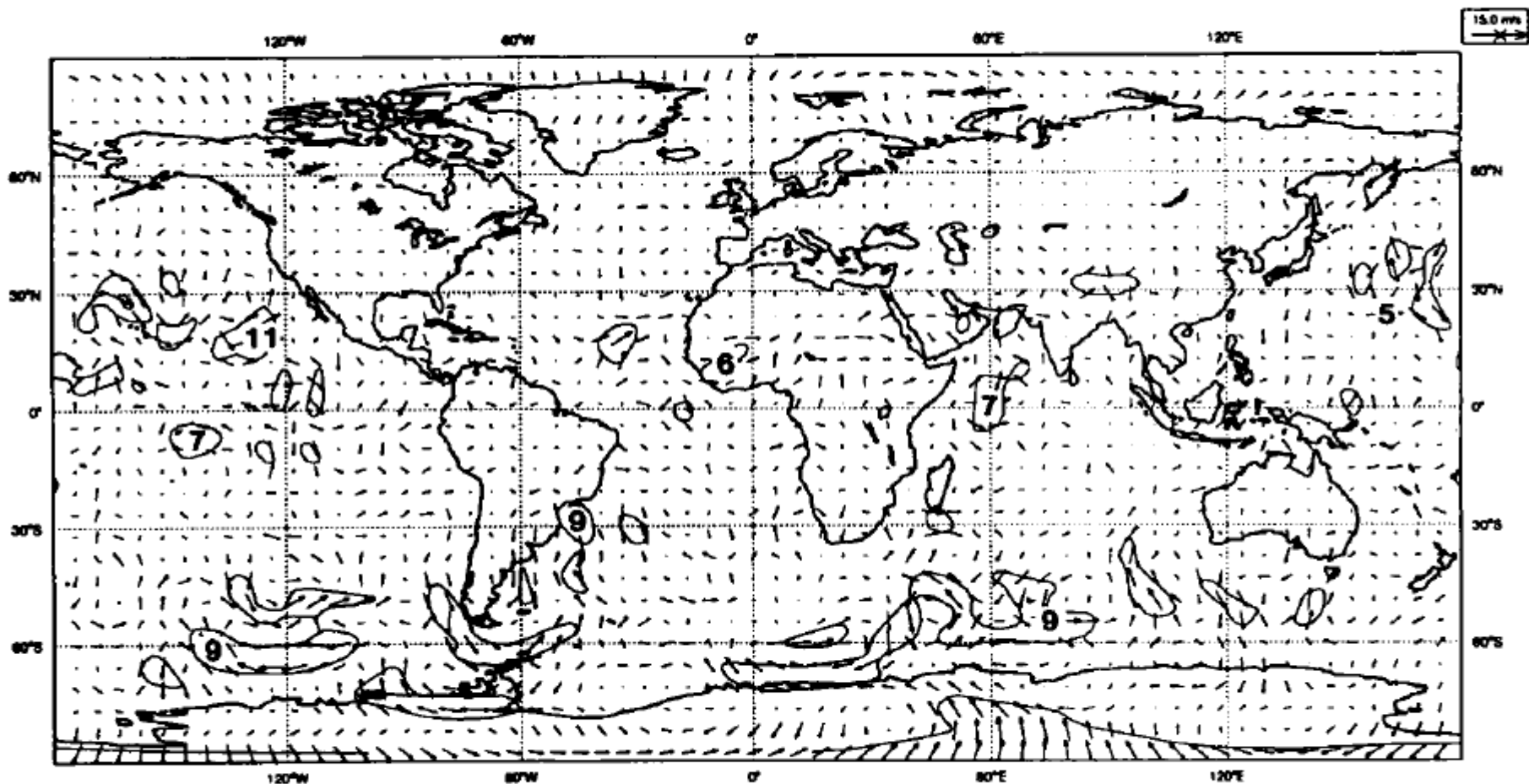


Figure 17. As Fig. 16 but humidity TOVS channels HIRS-11 and HIRS-12 have been excluded from the assimilation using TOVS radiances.



# Linearized Extended Kalman Filter

Daley (1995, 1996) studied linearized equations in EKF.

*Wind field can be recovered provided:*

- *sufficient structure in the constituent field,*
- *observations are frequent and accurate,*
- *data voids are small.*

i.e. filter estimated field must stay close enough to the truth for gradients to be accurate.



# Will linear incremental 4D-Var work? *Not very well!*

- Wind increments are calculated using gradients of the guess.
- In a long window (several ob times):
  - Cannot alter both the initial  $m$  (to fit early obs) and the wind  $\mathbf{u}$  which advects it (to fit late obs).
  - the guess is less likely to be accurate.
- In a short-window cycle (mimicking EKF):
  - $\mathbf{u}'$  will be derived from the movement of background  $m$  to fit observations.
  - But 4D-Var does not know in which areas background  $m$  is unreliable (due to past data voids) and may derive unreliable  $\mathbf{u}'$ .



# Multi-scale DA

- If displacement (between obs)  $\geq$  size of features (or if features have sharp edged, e.g. cloud/no cloud):
  - Multiple maxima in fit to obs are possible;
  - Linearisation fails if obs increments fall in regions with zero gradient;  
$$\frac{\partial m'}{\partial t} + \mathbf{u} \bullet \nabla m' + \mathbf{u}' \bullet \nabla m = S'$$
  - So we need a good guess at the displacement.
- Might obtain this from a preliminary iteration at reduced resolution (such that features are smoothed).
- This fits well with multiple outer-loop 4D-Var.



# Ensemble Kalman Filter

*Filter:* All info from past obs must be represented in the ensemble before assimilating the next batch.

*Kalman:* Batches of observations are added using the optimal linear algorithm based on covariances.

*Ensemble:* Covariances are sampled from ~100 model integrations for NWP applications, even though there are millions of degrees of freedom.



# All info must be in the ensemble

- All ensemble members must represent the detail in the observed image which will determine the fit to a later image.
- This detail has many degrees of freedom, e.g.  $24 \times 24$  pixels. This is more than the ensemble size.
- So cannot rely on ensemble covariances – must use severe localisation.
- But AMV experience shows it is best to get a single wind from pattern matching a  $24 \times 24$  area – wind correlations are much broader than the forced localisation.
- The localisation needed to fit an image is likely to damage the larger-scale multivariate relationships between image position and wind.



# linear algorithm based on covariances

- A single linear Kalman Update equation to increment ensemble mean estimate based on observed innovations.
- Plus a method (depending on the flavour of the Ensemble Kalman Filter), *afaik* always linear, to update the ensemble spread covariance.
- No method (analogous to multi-resolution outer-loop 4D-Var) for dealing with nonlinear penalty coming from high-resolution imagery.



# Vision – ideal Global DA for NWP, using quasi-linear methods

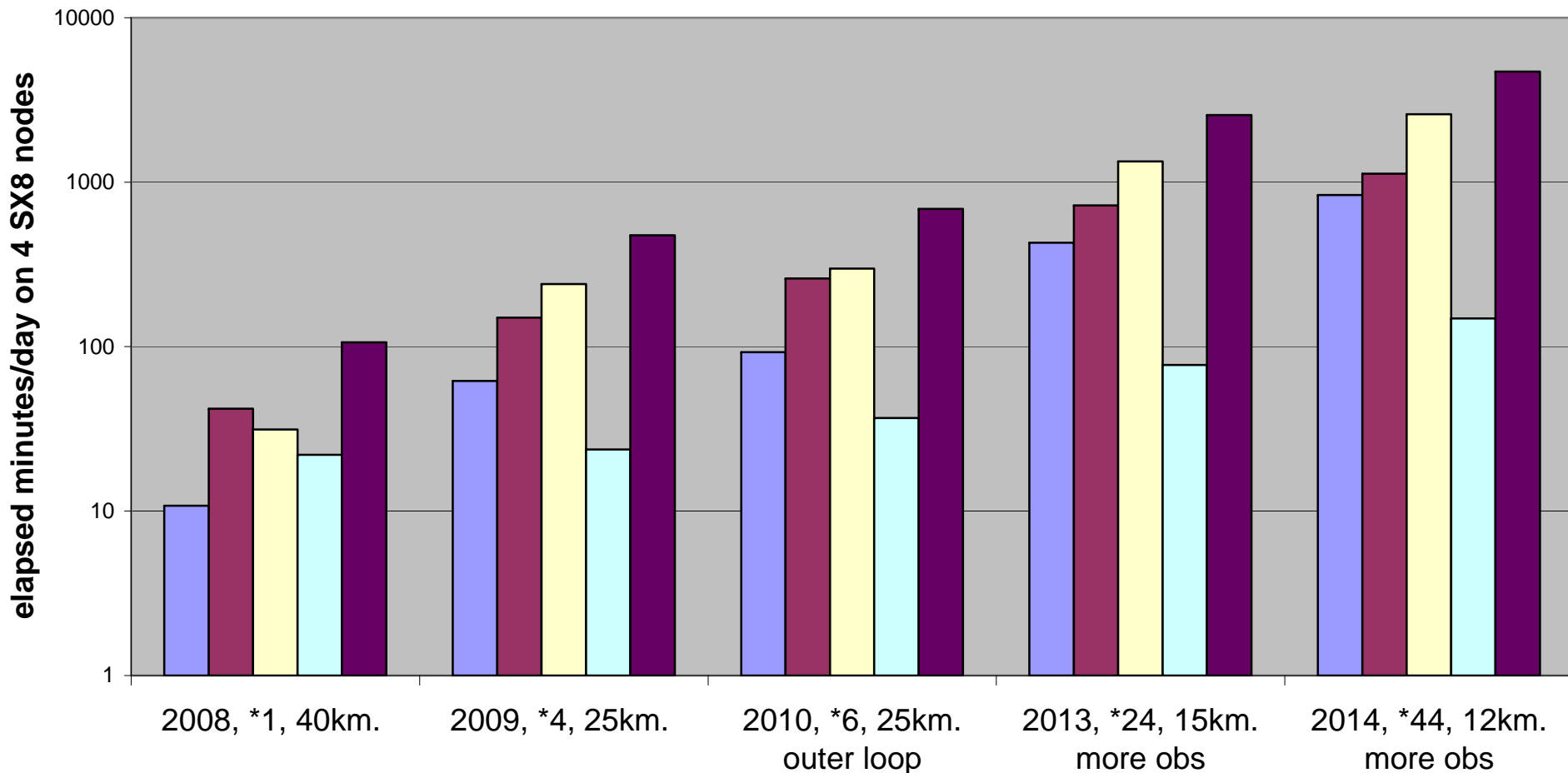
- “Best estimate” DA of “known” scales (~12km), using 4D-Var because of:
  - Desire to treat all scales together;
  - Desire to make best use of satellite obs e.g. by bias correction, using high-resolution.
- Hybrid ensemble to carry forward error information from past few days.
- May still be scope for nested regional systems to give more rapid running and higher resolution.

*N.B. This vision is good for perhaps a decade, while we are restricted to well known scales, so the KF theory of a “best estimate” + a covariance description of uncertainty is useful.*

# Extrapolated cost of vision

## Global 4D-Var DA cycle: computer costs

Model forecasts   reconfiguration & IO   4D-Var   OPS   total





# Discussion of global DA cost extrapolation

- 2008 figures are actuals: 40km full model, 120km linear model, single 50 iteration 4D-Var loop, 4 \* 6hr cycles.
- The years given are based on a Moore's Law extrapolation from our IBM in 2009. There may not be computers or funding for this.
- Figures do not allow for better methods other than outer-loop and \*4 observations. Improvements (**Var bias correction, Var QC, nonlinear cloud and precipitation obs, hybrid use of EnKF perturbations, ...**) may add a factor of ~4 and hence ~3 years.
- Risk is that the linear model will not "scale" to run more quickly on MPP. This is a theoretical bottleneck for 4D-Var since linear model runs are sequential. *Not very apparent on upgrade to IBM Power6 in 2009 (5.3 speed-up compared to 6.5 average).*
- After initial processing of observations in OPS, observation costs in 4D-Var are nearly negligible.



# Question re global NWP

- In 10~20 years we will be able to run global ensembles at resolutions such that the initial errors are non-Gaussian.
  - *If the ensemble mean is so smooth as to be significantly implausible as a real state, the errors are non-Gaussian.*
- Kalman Filter based methods (i.e. 4D-Var & EnKF) are not appropriate.
- [ Nonlinear initialisation / the model attractor / spin-up ] will be very important because of assimilation of imagery data and the desire for short-period precipitation forecasts.
- Models and observations will still be imperfect.
- Particle filters will be unaffordable.
- What will you do? (I will be retired 😊)



# Convective-scale DA for NWP

Some thoughts – no vision yet!

- Nested in global to use obs from wider space & time window
- Needs ~1km model for typical UK weather
- When to do DA:
  - In chaotic regimes, it is only feasible to do DA for convection in well observed regions, with radar.
  - In stable regimes, DA of high-resolution is worthwhile.
  - In some regimes, downscaling using high-resolution topography adds value, without additional DA.
- It took global-scale NWP 20 years to learn how to use satellite soundings well. How long will it take with radar?
- Operational practice is in its infancy. 3D-Var & “nudging” are common. Cannot afford to do 4D-Var or EnKF properly for UK for 10 years.
- Will the quasi-linear 4D-Var or EnKF methods work OK for operational convective-scale NWP, or are non-Gaussian methods really needed?



Met Office

# Summary

- What is needed for a world-class DA system for NWP?
  1. Good model
  2. Statistical/dynamical methods for extracting observed information
  3. 4D-Var or EnKF or similar
- Handling a wide range of space- & time-scales.
  1. Large-scales still uncertain, small scales increase accuracy.
  2. Want to represent and extract information from tracers.
- Vision for next decade DA for global NWP: **4D-Var**
  - Probably affordable up to 12km.
  - Couple to LETKF lower-resolution ensemble.
  - Allows simple implementation of better observation processing:
    - VarBC, nonlinear cloud & ppn, use of tracers, ...
- 10~20 years??
  - How do we initialise unknown detail?
- Some thoughts on convective scale DA.



# Repeat of discussion questions

- In 10~20 years we will be able to run global ensembles at resolutions such that the initial errors are non-Gaussian. What will you do?
- It took global-scale NWP 20 years to learn how to use satellite soundings well. How long will it take with radar?
- Will the quasi-linear 4D-Var or EnKF methods work OK for operational convective-scale NWP, or are non-Gaussian methods really needed?



# Questions and discussion