Treatment and Impact of Specification of Errors in Data Assimilation

Sue Ballard,NetFAM/COST workshop April 2006

Z. Li, M. Dixon, S.Swarbrick, O.Stiller and H. Lean

Met Office, JCMM, Reading University

Richard Renshaw and Data Assimilation and Observation Section

Met Office, Exeter

Contents

The ultimate problem in meteorology

Initial value problem

Deterministic (?)

Observations -> analysis -> forecast

Now accept that we need to allow for uncertainty and
can gain information
ensembles, probability, distributions

Allow for error in observations
error in forecast model

Sources of information

Observations – direct and indirect
eg screen temperature, mast wind, aircraft temperature
v satellite radiance, radar doppler radial wind, reflectivity

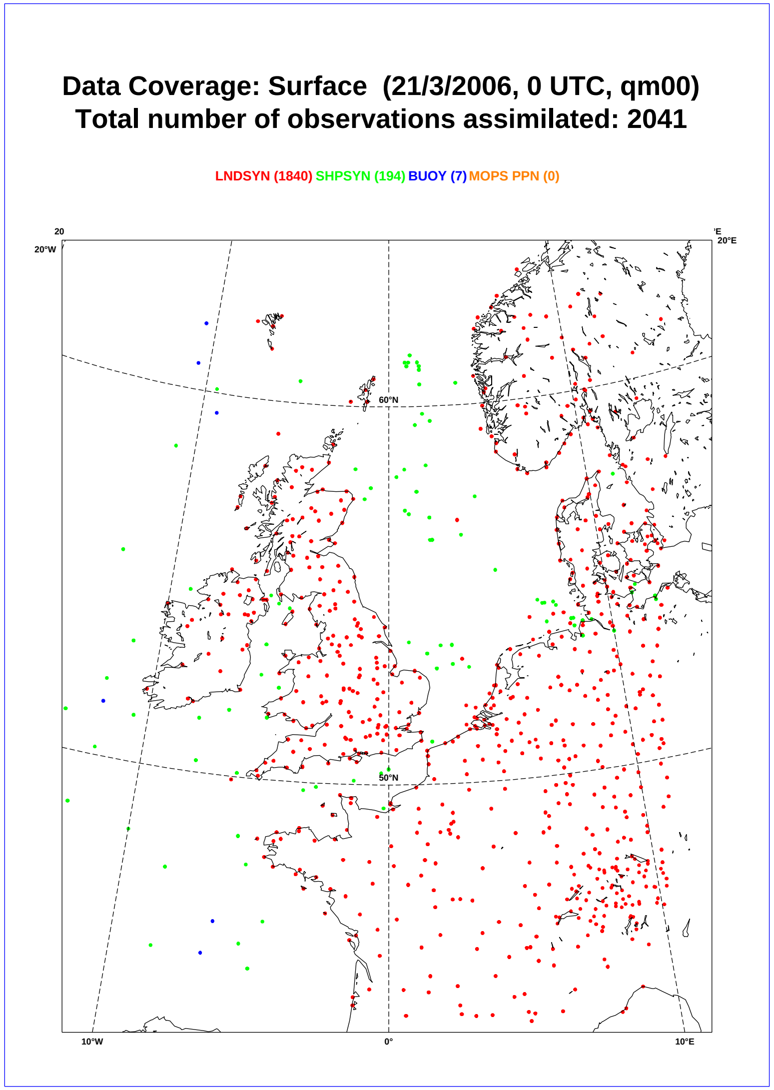
No uniform network, incomplete information

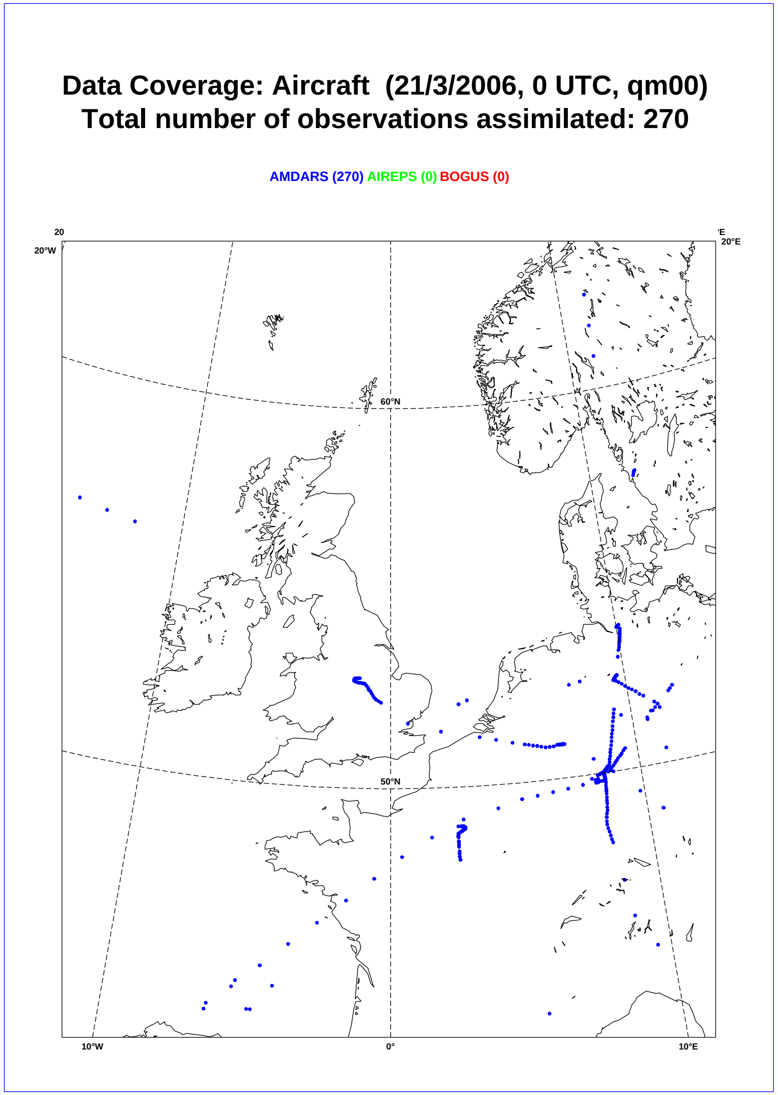
However information is advected and evolves nonlinearly

Previous forecast contains information from previous observations

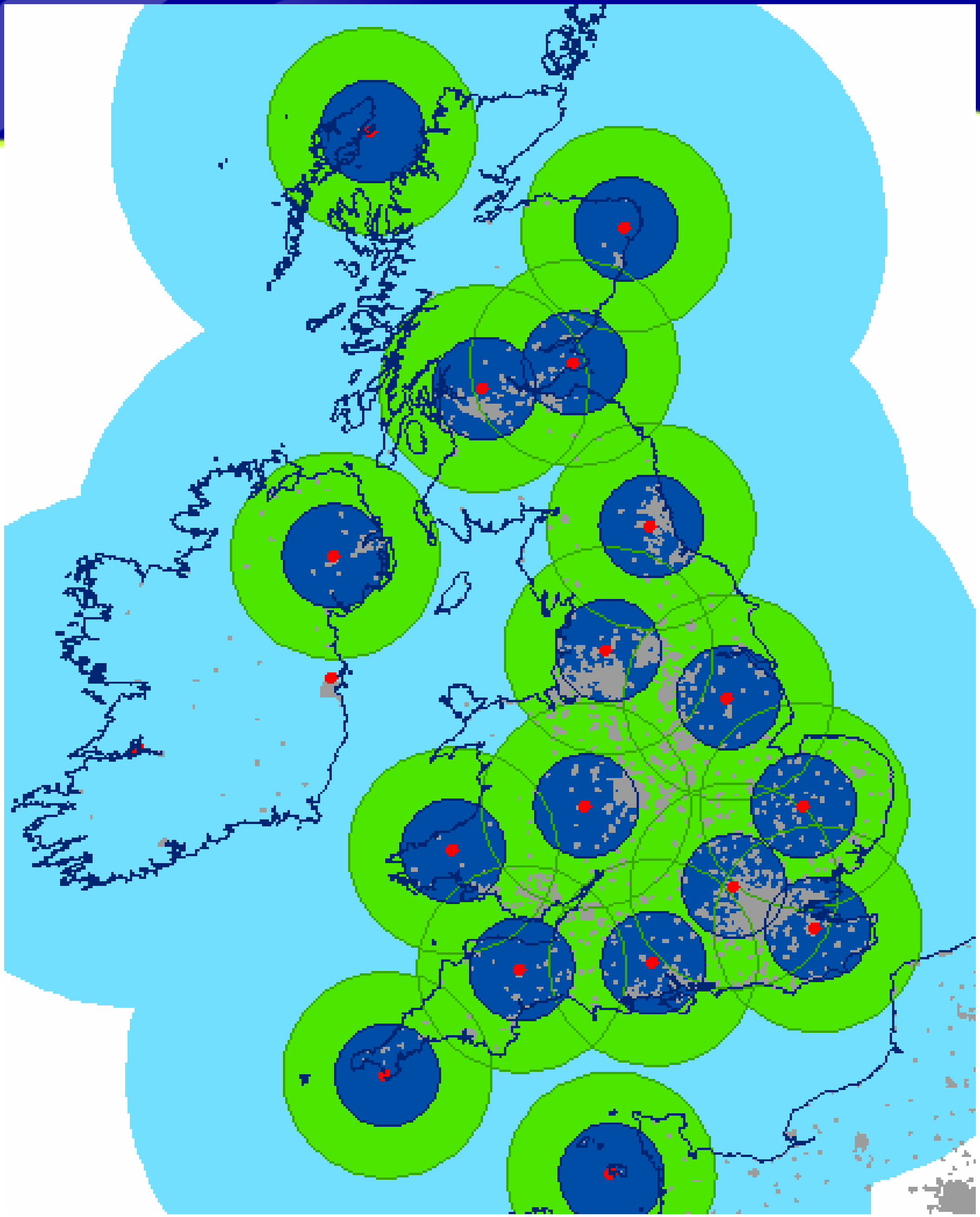
Therefore combine latest forecast with latest observations

Sources of information





Radar Network Coverage – 2006

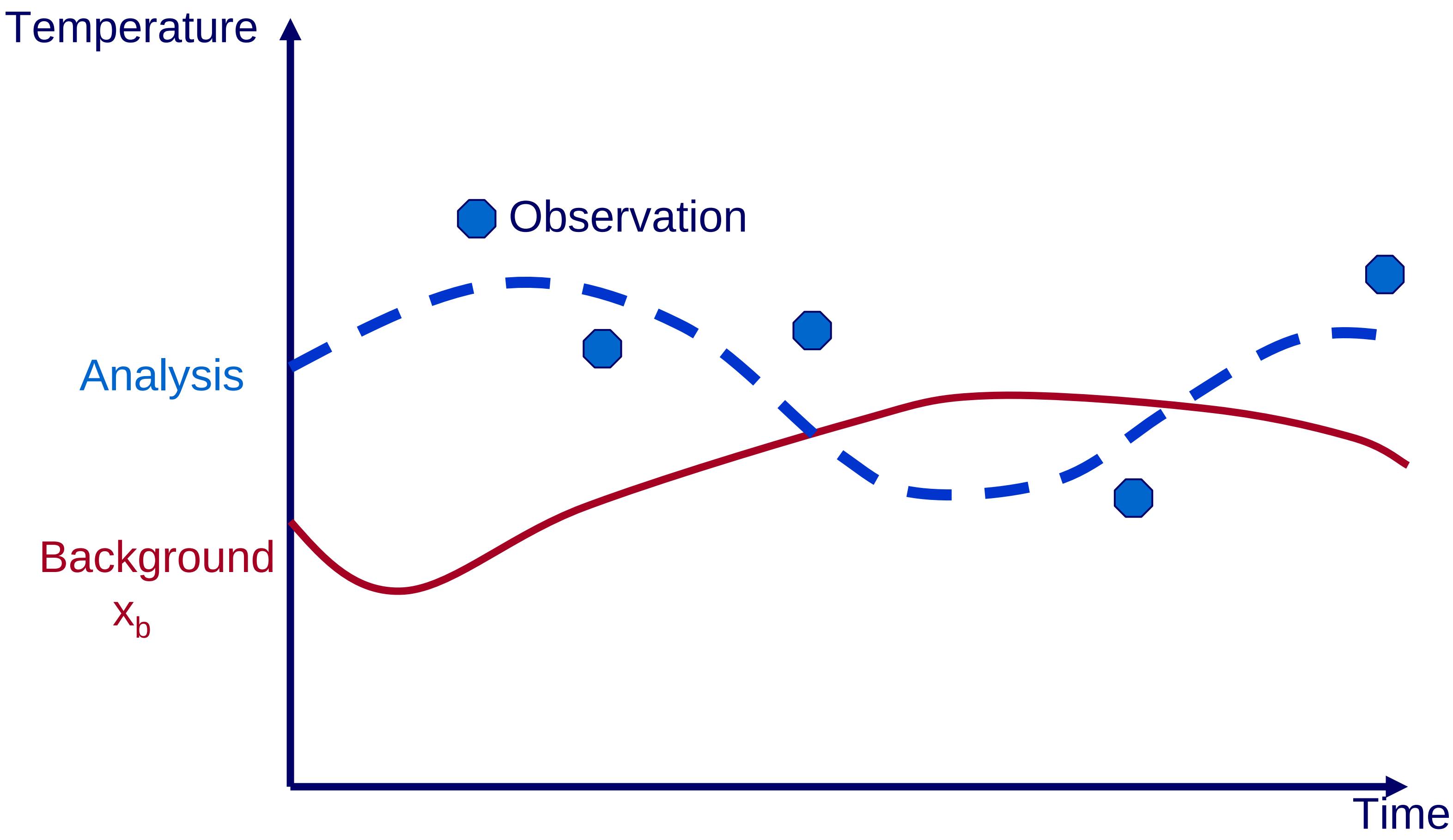


1km resolution

2km

5km

Four-dimensional variational assimilation (4D-Var)



Key Ideas from Gauss:

Gauss, 1809: *Theoria Motus Corporum Coelestium*

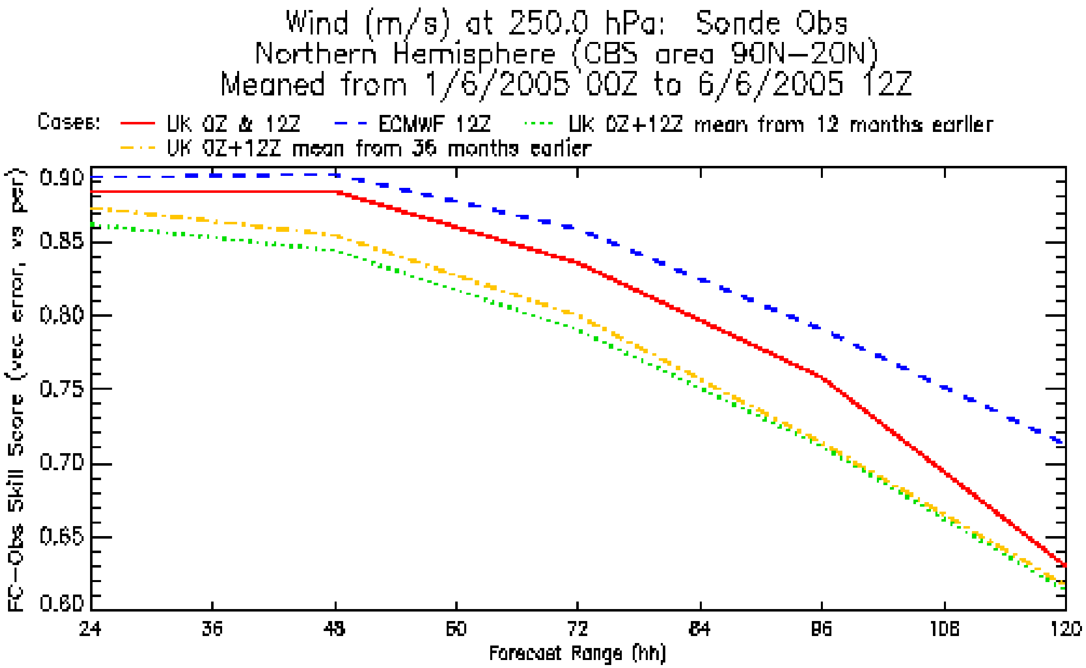
-1823: *Theoria combinationis observationum erroribus minimis obnoxiae*

- all models and observations are approximate
- the resulting analysis will also be approximate
- the observations must be combined in some optimal fashion
- it is better to have enough observations to over-determine the problem
- the model is used to provide a preliminary estimate
- the final estimate should fit the observations within their (presumed) observational error

Note: $O(10^6)$ obs

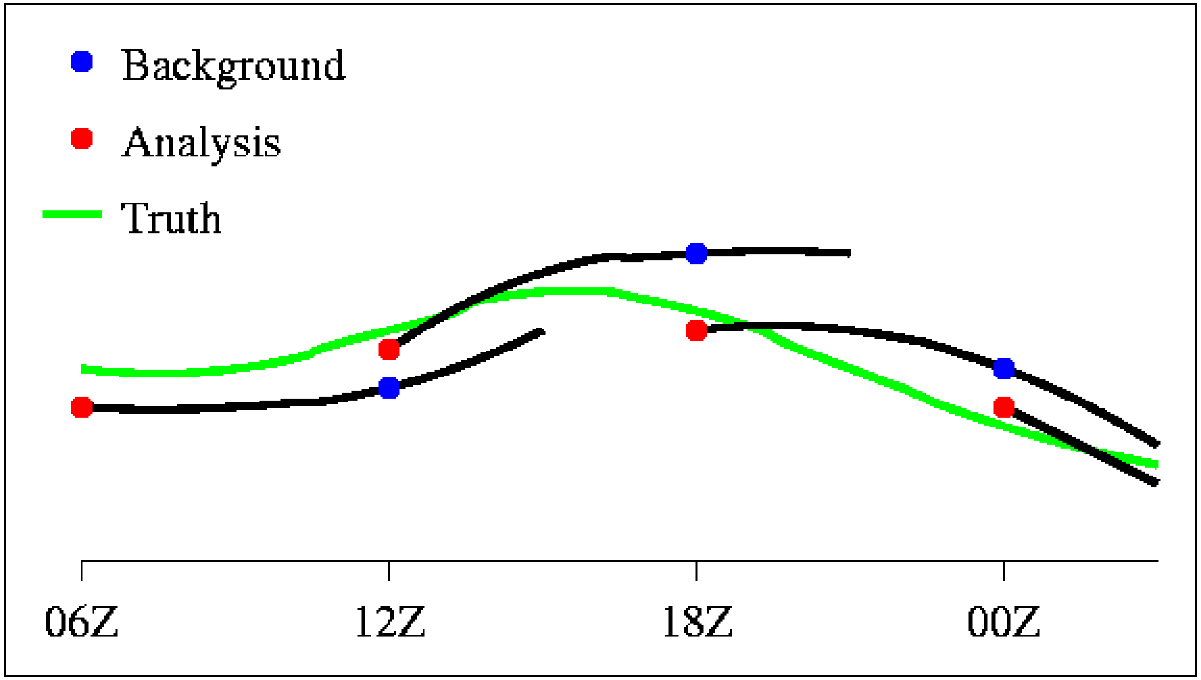
$O(10^7)$ model variables * grid-points

The background has information…

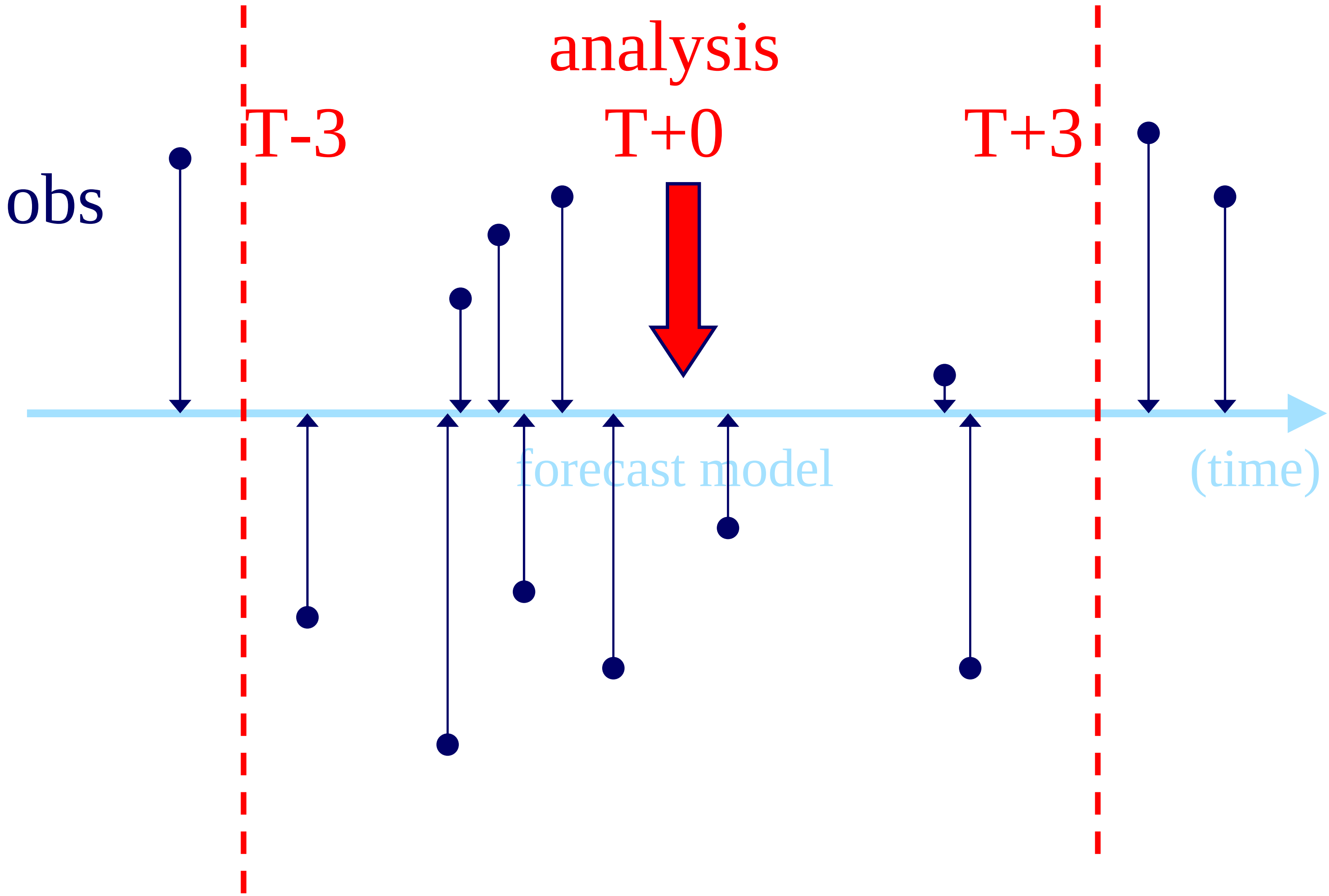


What is data assimilation?

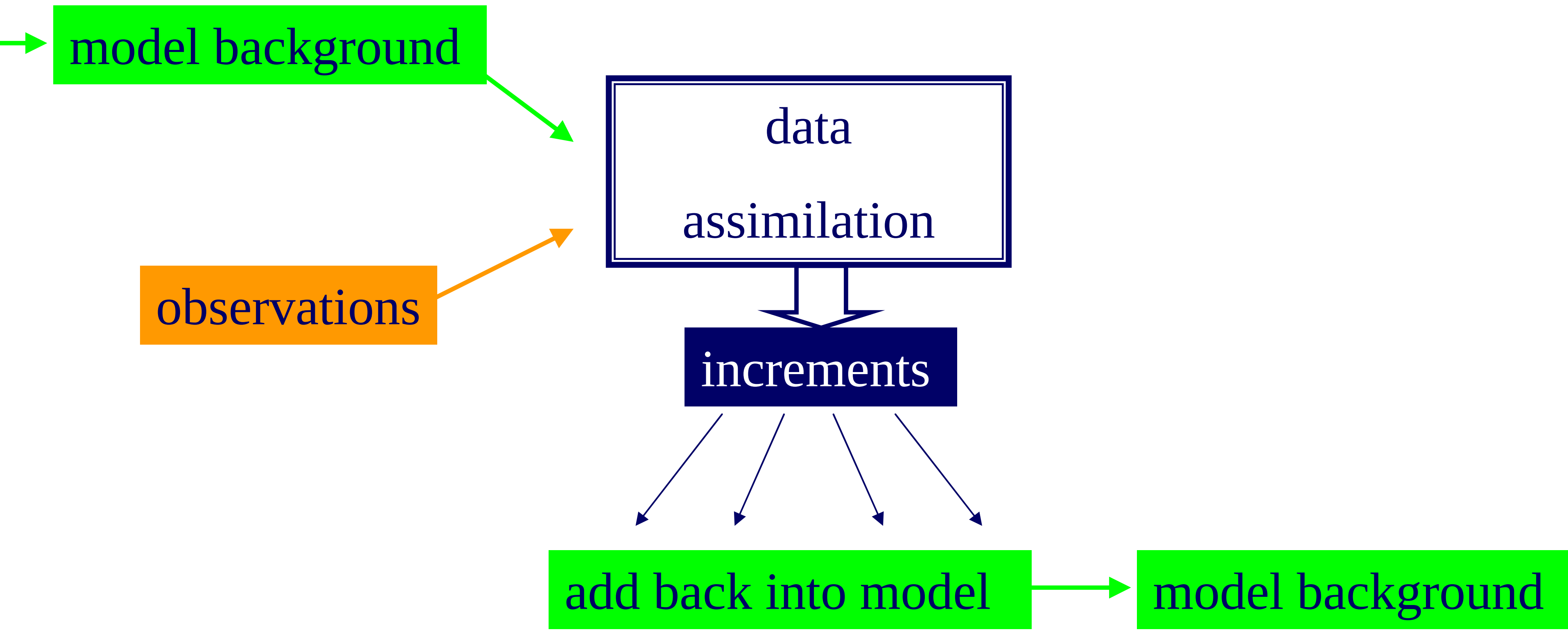
- The best and most powerful analysis systems are obtained by incorporating numerical models into analysis algorithms.
- The model encapsulates our understanding of the physical laws, and can be used to propagate observational information forwards in time.
- **Data assimilation**: the production of regular analyses for, and in conjunction with, a forecast model.
- Typically, an *intermittent* data assimilation cycle is used:



Intermittent Data Assimilation



Data Assimilation



The Analysis Problem

Maximise $P(x|y,x^b)$ where $y = H(x) + e^o$ (observations) $x^b = x + e^b$ (background)

In case of Gaussian errors in the first guess and observations minimise

$$J(\mathbf{x}) \quad \frac{1}{2} \quad \mathbf{x} \quad \mathbf{x}^b \quad {}^T \mathbf{B} \; {}^1 \quad \mathbf{x} \quad \mathbf{x}^b \qquad \frac{1}{2} \quad H(\mathbf{x}) \quad \mathbf{y}^o \quad {}^T \mathbf{R} \; {}^1 \quad H(\mathbf{x}) \quad \mathbf{y}^o$$

Where $B = \langle e^b (e^b)^T \rangle$ background error covariances

$R = \langle e^o (e^o)^T \rangle$ observation error covariances

- Assumptions: Gaussian errors, no bias, no correlations between background and observation errors
- Quality control needed to make errors more Gaussian
- The evaluation of H(x) may involve time integration
- Data assimilation requires cycling in time (and ideally updating of B)
- The background can be thought of as a special set of observations

Met Office incremental 3D-Var

- To make VAR more practical, we can approximate the penalty J in terms of increments $\delta\mathbf{w}$ to a simplified model, and linearise about a guess state $\mathbf{x}^g$:

$$J(\mathbf{x}) = \frac{1}{2}\left[\mathbf{x}-\mathbf{x}^b\right]^T \mathbf{B}^{-1}\left[\mathbf{x}-\mathbf{x}^b\right] + \frac{1}{2}\left[H(\mathbf{x})-\mathbf{y}^o\right]^T \mathbf{R}^{-1}\left[H(\mathbf{x})-\mathbf{y}^o\right]$$

$$\downarrow$$

$$J(\delta\mathbf{w}) = \frac{1}{2}\left[\delta\mathbf{w}-\delta\mathbf{w}^g\right]^T \mathbf{B}_{(\mathbf{w})}^{-1}\left[\delta\mathbf{w}-\delta\mathbf{w}^g\right] + \frac{1}{2}\left[\mathbf{G}\,\delta\mathbf{w}-\mathbf{d}\right]^T \mathbf{R}^{-1}\left[\mathbf{G}\,\delta\mathbf{w}-\mathbf{d}\right]$$

$$\delta\mathbf{w}^g = S\left(\mathbf{x}^g-\mathbf{x}^b\right)$$

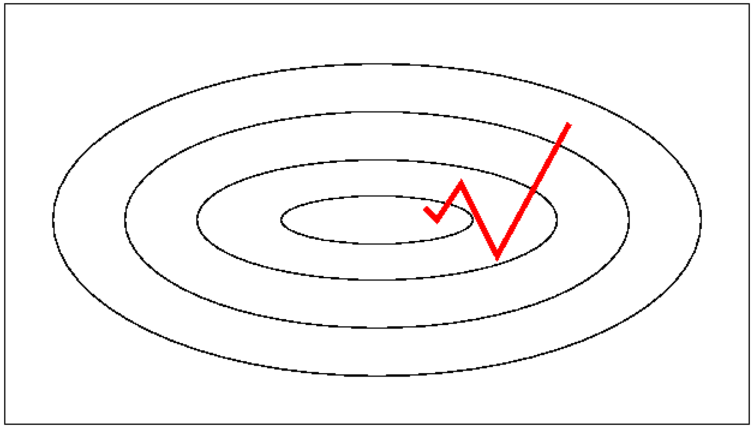
$$\mathbf{d} = \mathbf{y}^o - H(\mathbf{x}^g)$$

$$\mathbf{G} = \mathbf{H}S^{-I}$$

$$\mathbf{x} = \mathbf{x}^g + S^{-I}\,\delta\mathbf{w}$$

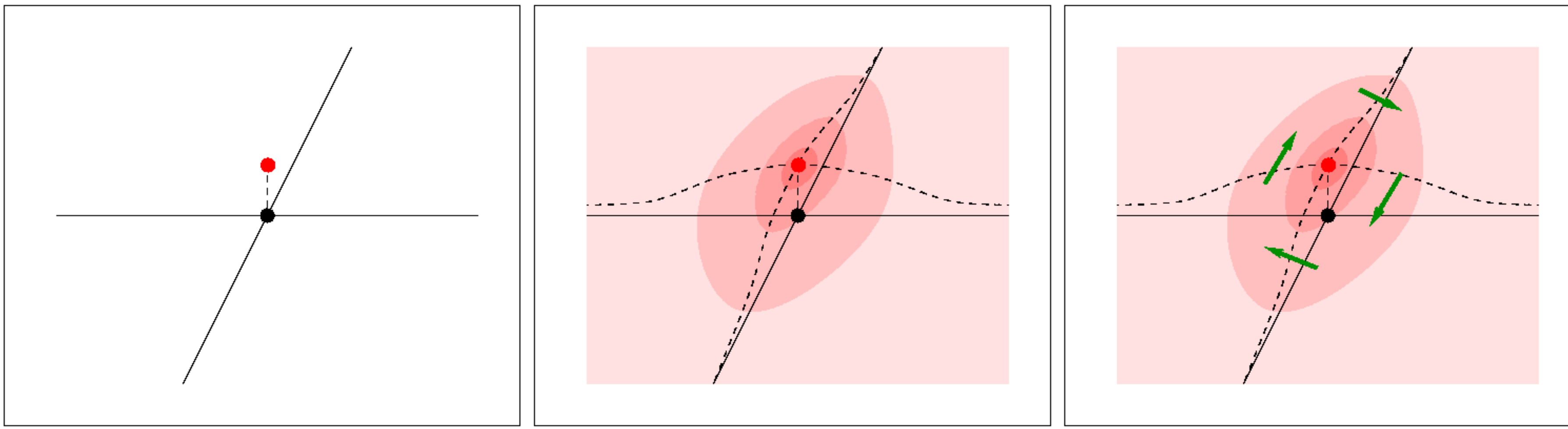
Variational data assimilation – inverse problem

- VAR is an inverse problem: we don't `interpolate' observations onto the model grid, we **vary** the model state until we find that which is most compatible with the data, as defined by J.
- We can assimilate variables not directly related to the model variables, as long as we can write a reasonably accurate observation operator.
- For example, we can use satellite radiance measurements directly.
- To minimise J, we use an iterative **descent algorithm**. On each iteration, the algorithm needs J and its gradient wrt **x**.



Variational data assimilation – background error

- The background error covariance matrix **B** describes the error variance for each model variable, and the **correlations** between errors in different model variables; ie. how information from observations should be spread:



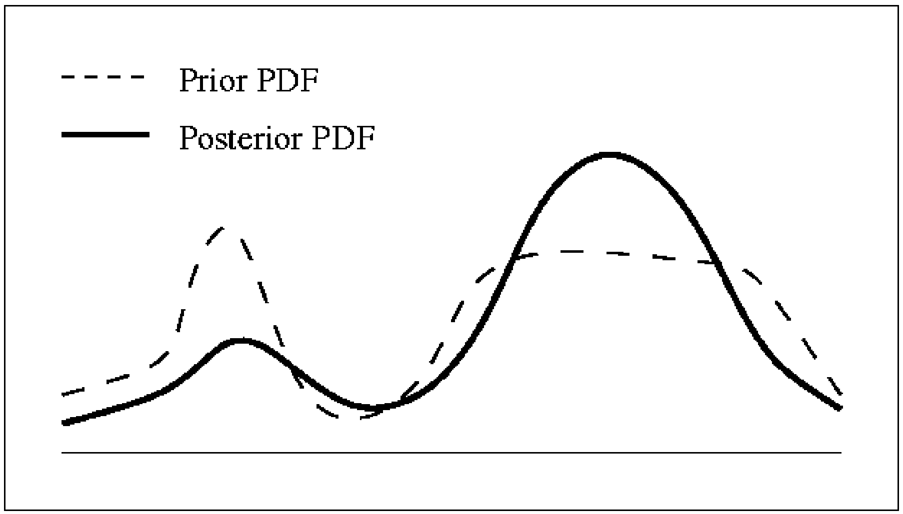
- Incorporating better approximations of the `true' background error covariance matrix is perhaps THE most important theoretical challenge in data assimilation.

Variational data assimilation – control variables

- **B** is NxN, where N is the number of (simplified) model variables.
- In order to make **B** manageable, we need to make some assumptions.
- Transform to variables that are approximately uncorrelated:
  - Transform to **streamfunction**, **velocity potential**, **unbalanced pressure** and **rh**.
  - Project onto vertical modes.
  - Project onto horizontal spectral modes.
- **B** is then diagonal, and easily dealt with.

Variational data assimilation

- If the statistics of the background and observation errors were known, we could in theory use Bayes' Theorem to deduce the PDF $p(\mathbf{x} \quad \mathbf{x}_t)$ and choose an appropriate analysis:



- Statistics are not known, and in general could not be represented anyway.
- Need to make some assumptions.

Variational data assimilation – modelling B

Generation of Background Errors - Not known so have to be estimated eg

NMC method

Statistics are gathered by comparing pairs of forecasts valid at the same time (eg. T+48 - T+24)

The statistics are **climatological**, and approximately **homogeneous** and **isotropic**

can extend to vary latitudinally or use defined horizontal correlations

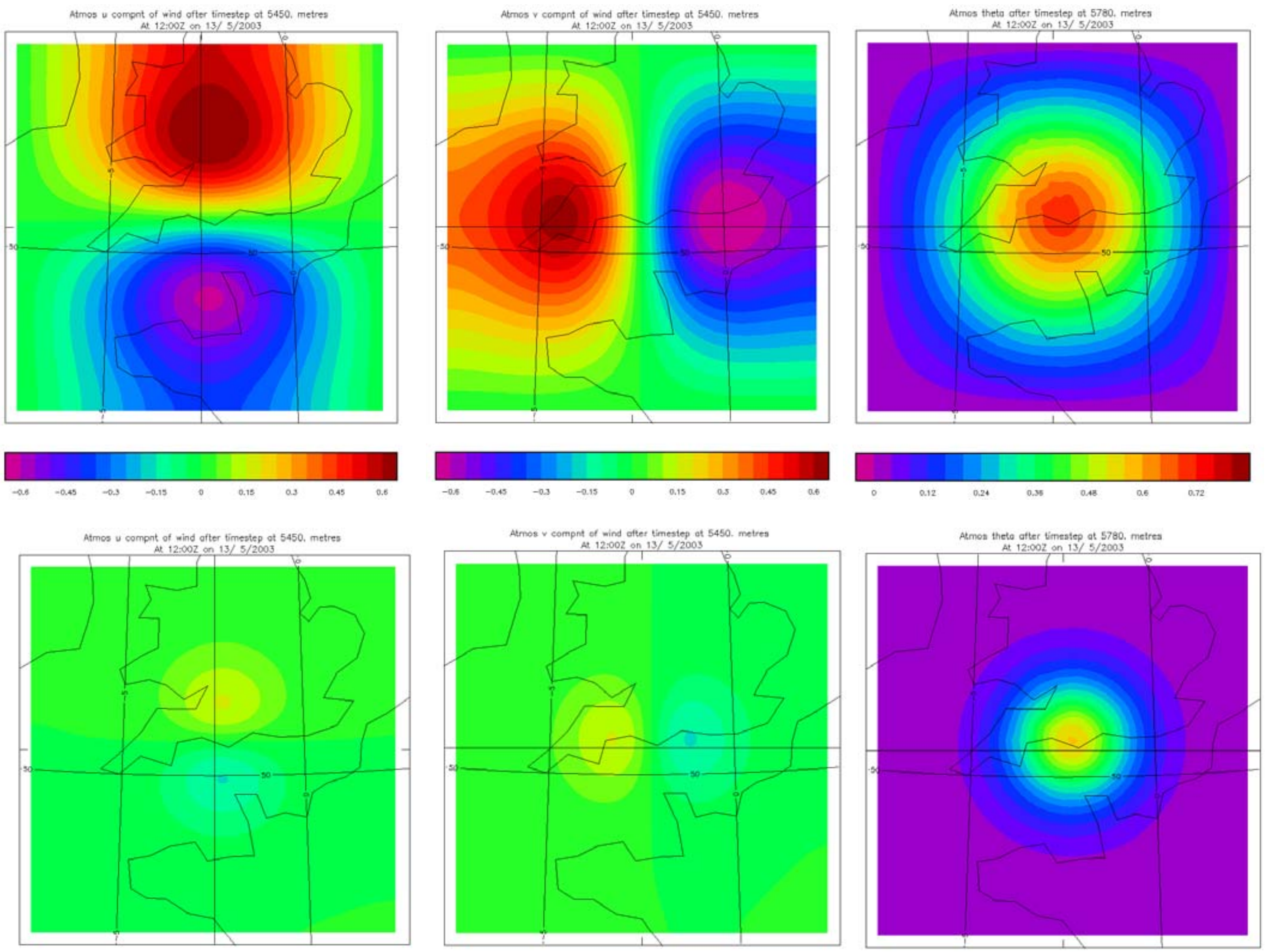
Idealised horizontal and vertical correlations

Ensembles of forecasts valid at forecast range of background

Analyses and Forecasts are very sensitive to specification of background errors variances, correlations and lengthscales

Spreading information – background errors

u v theta



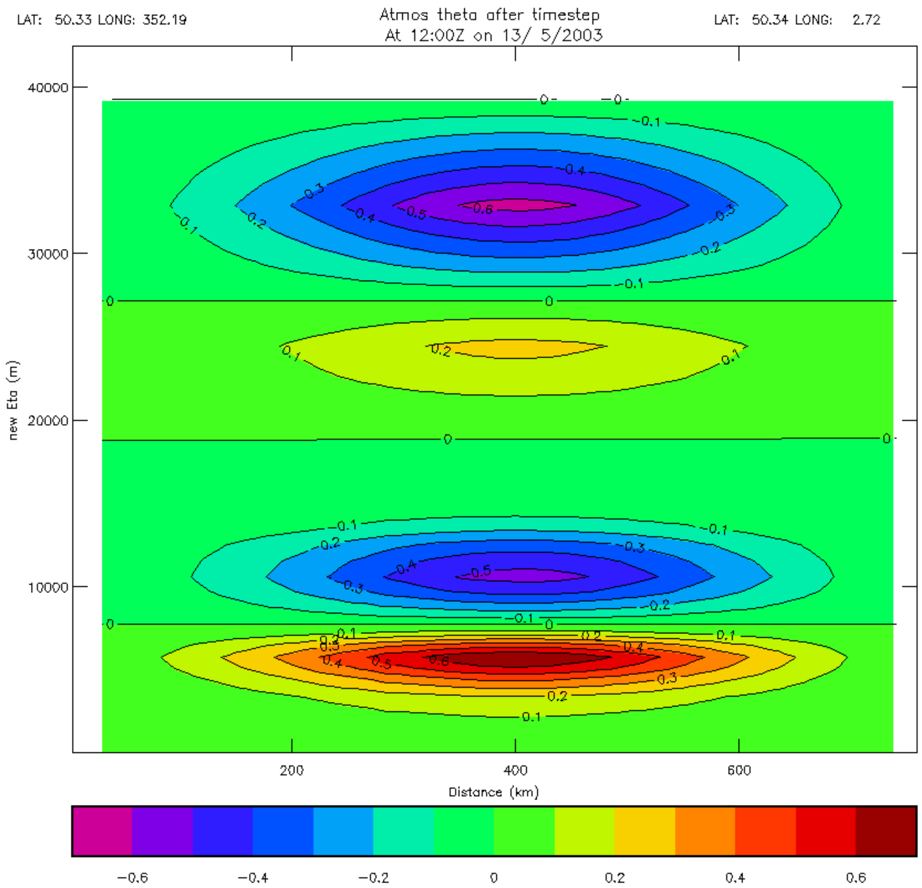
Derived for 12km model
Used in 4km

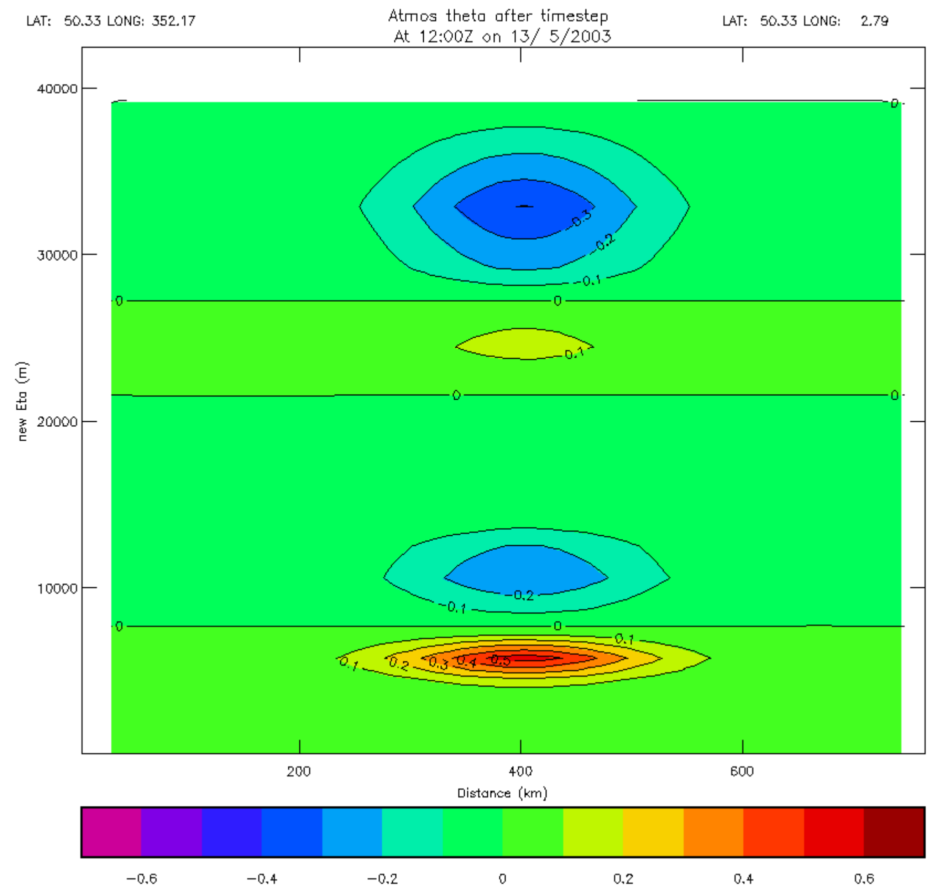
5.78 km
Approx
500hPa

½ lengthscale

Spreading information – background errors theta increment at 5.78km

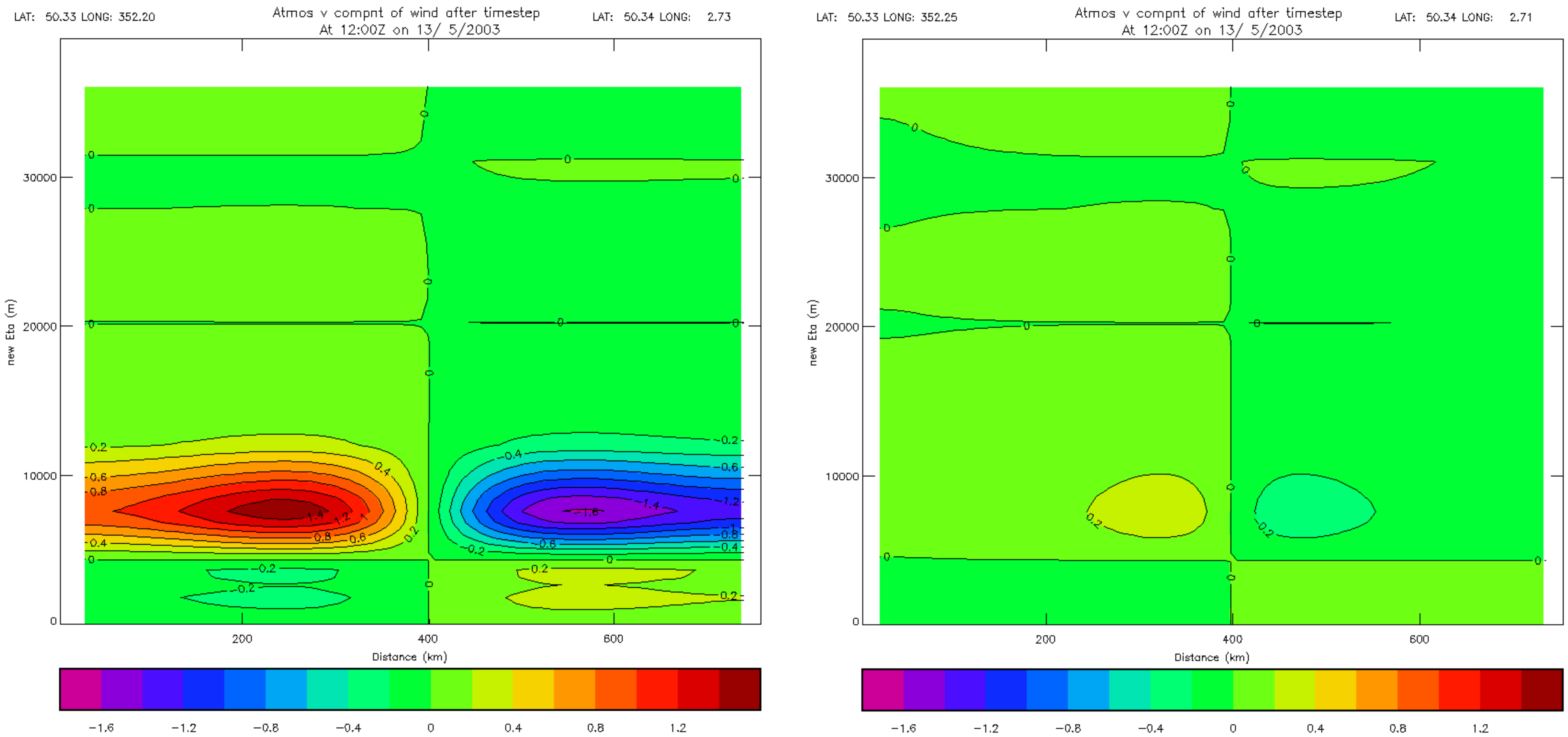
Derived for 12km model theta ½ lengthscale





Spreading information – background errors v increment at 5.78km

Derived for 12km model v ½ lengthscale

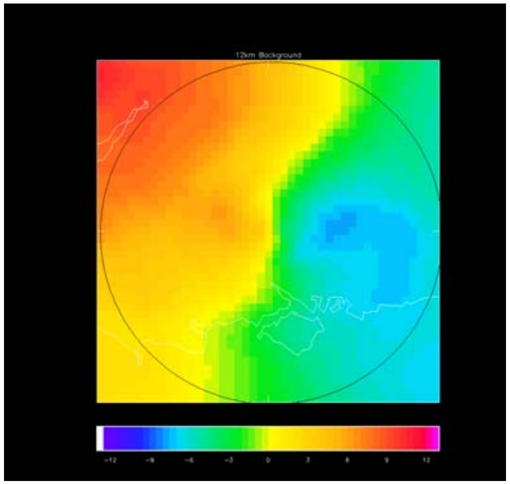


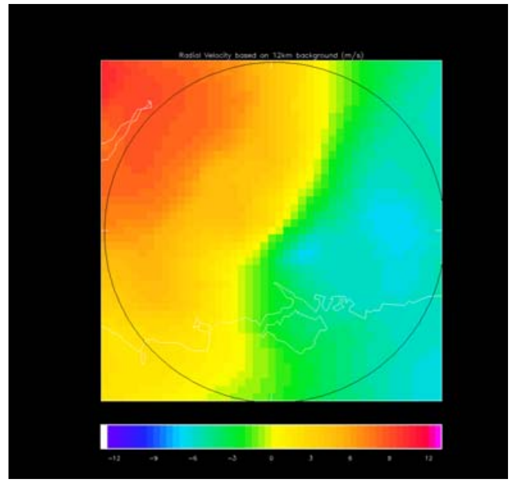
Impact of S-band radial radar wind data - radial wind on 1deg scan elevation

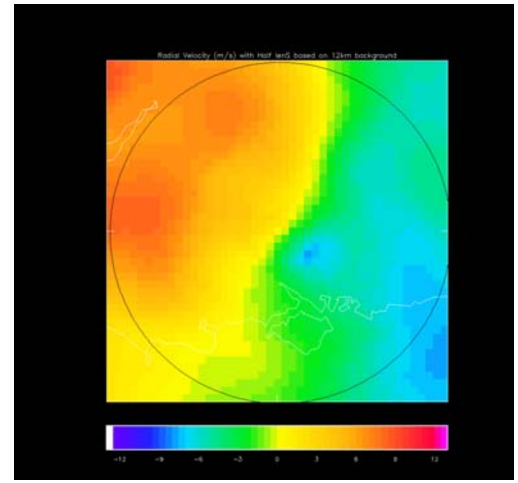
Analysis

½ Length scale

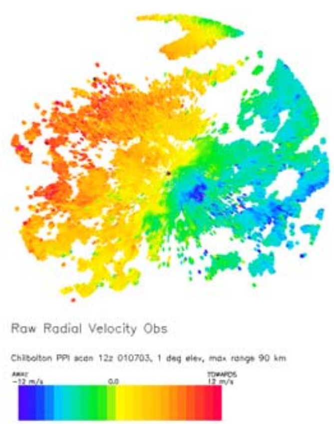
12km Back-ground

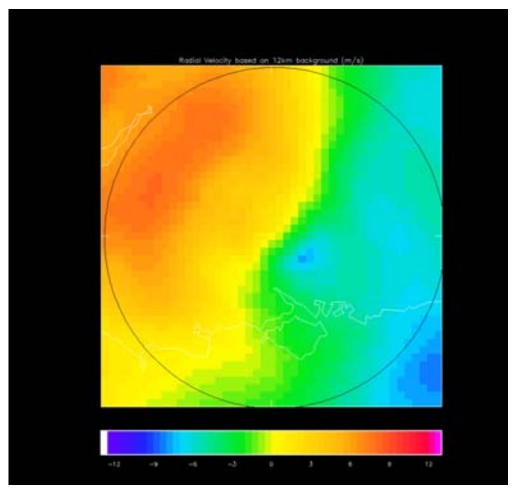


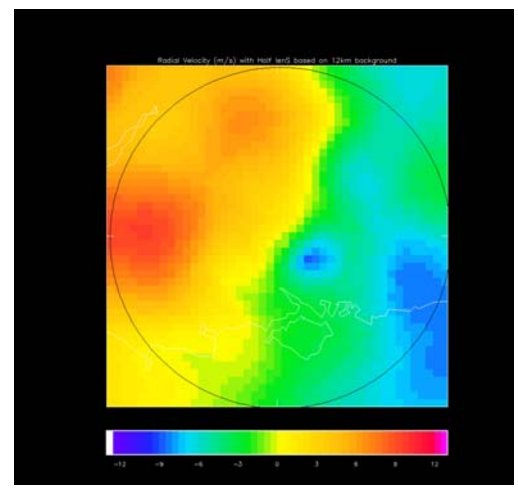




Super-obbed Radar Doppler wind







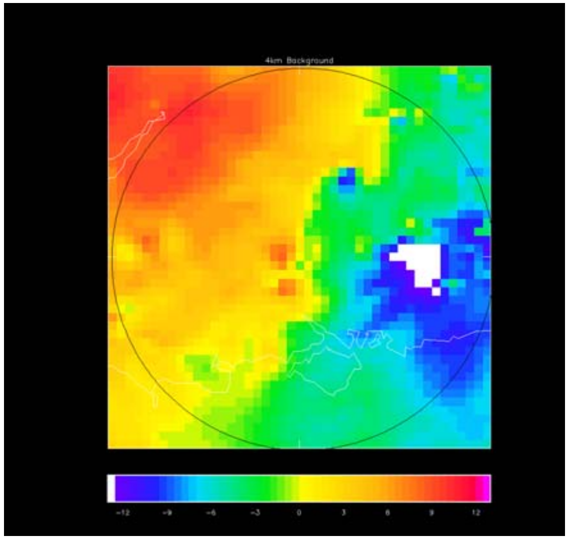
Reduced background wt

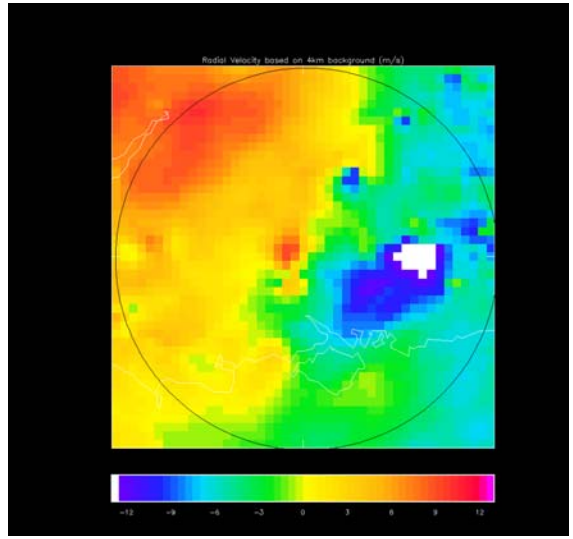
Impact of S-band radial radar wind data - radial wind on scan elevation

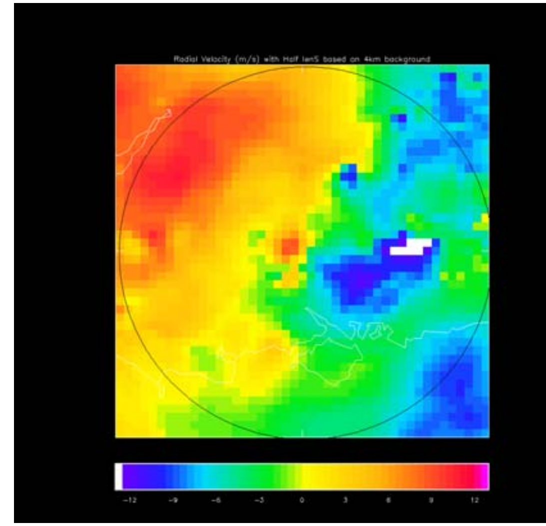
Analysis

½ Length scale

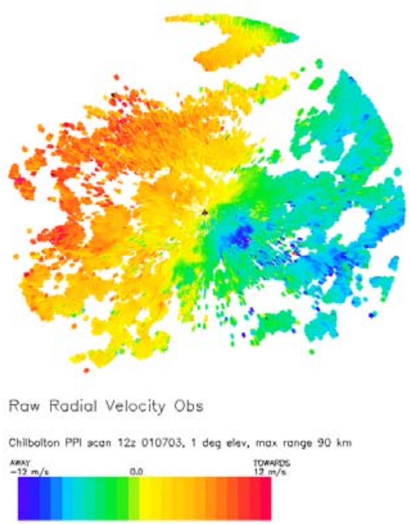
4km Background

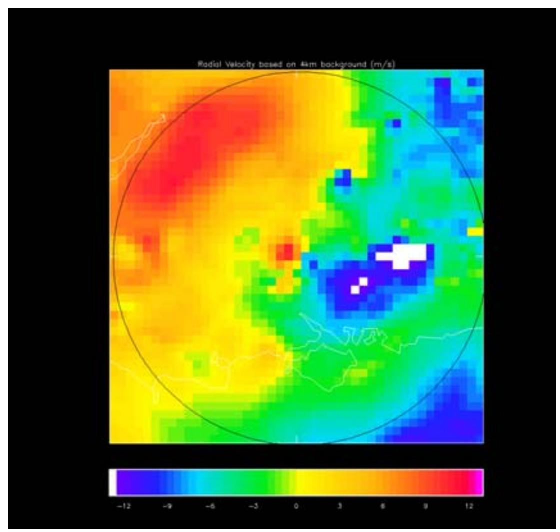


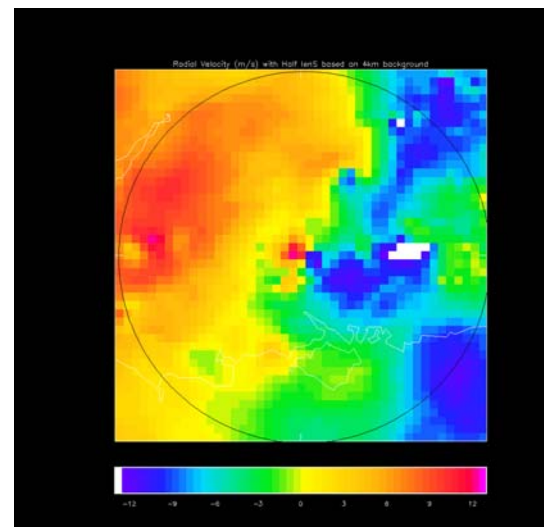




Super-obbed Radar Doppler wind



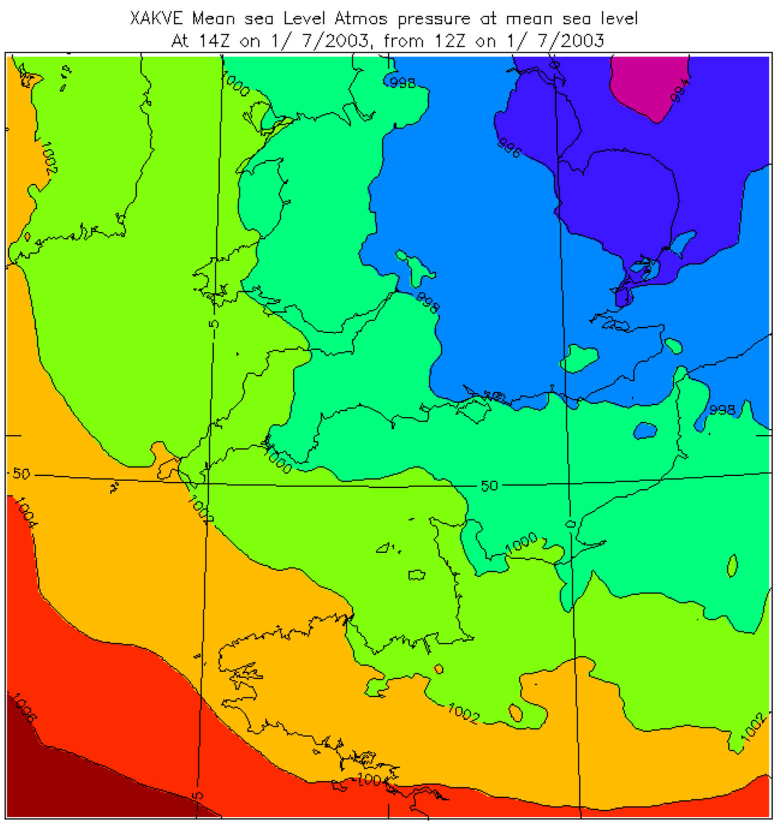




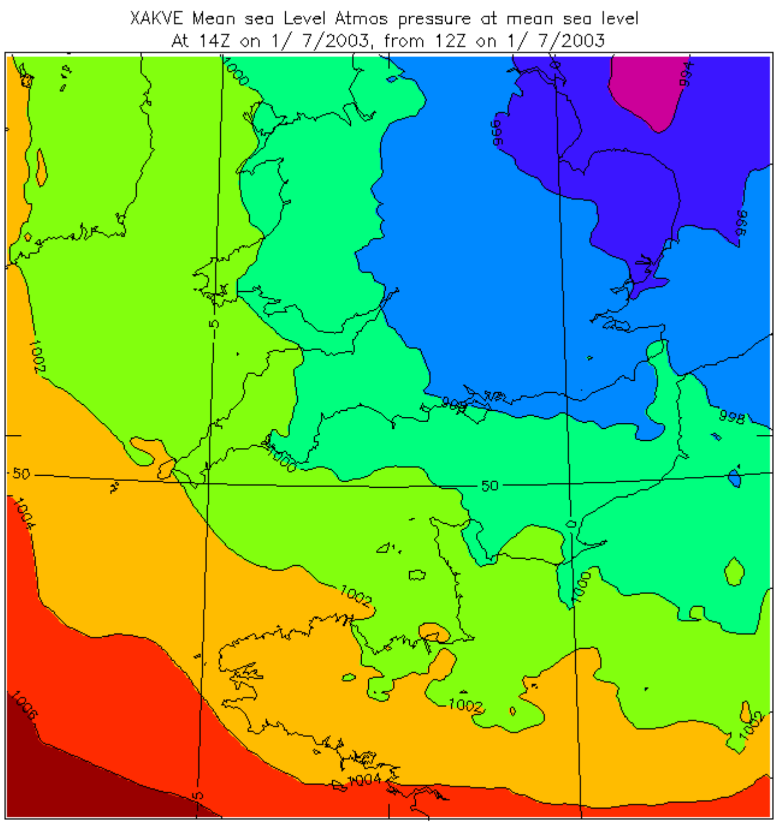
Reduced background wt

Impact of S-band radial radar wind data - T+2 surface pressure

Background forecast

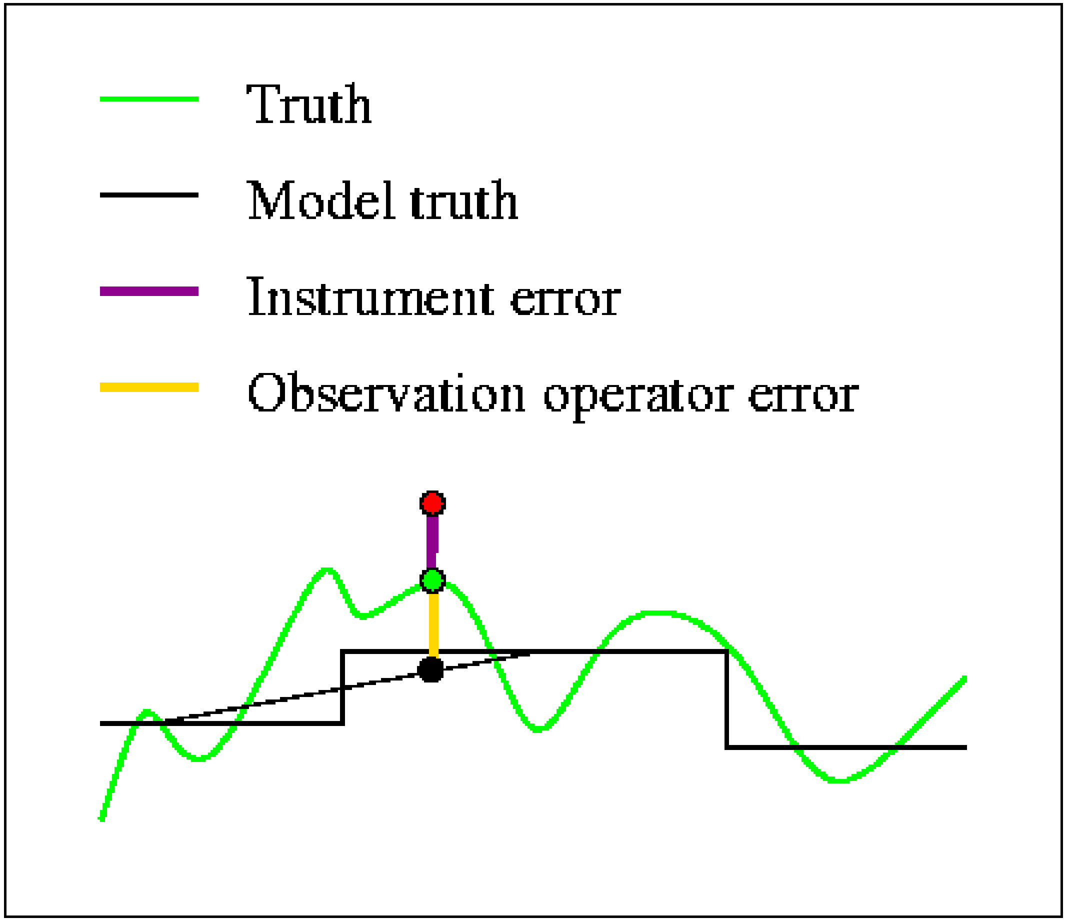


Including radar radial doppler winds



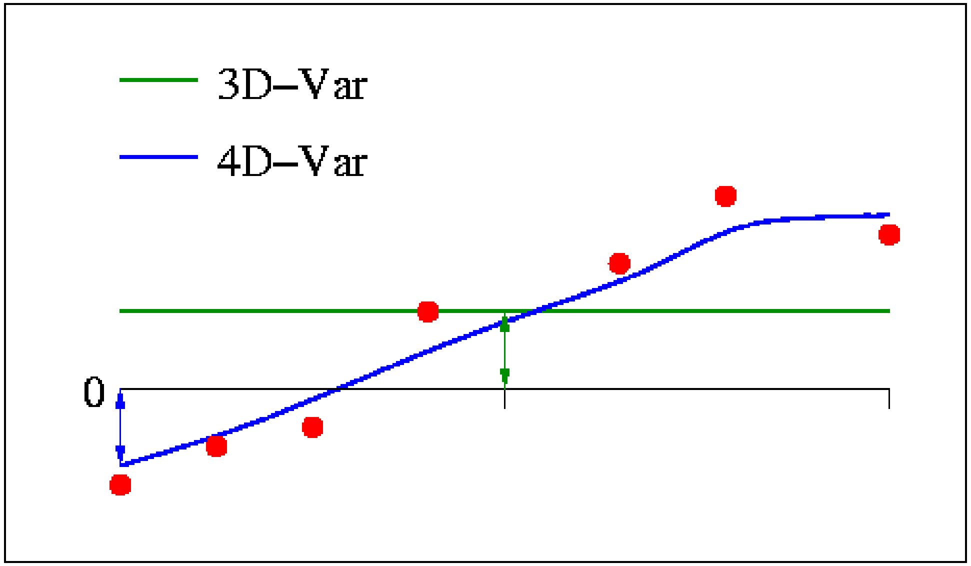
Variational data assimilation - observation error

- **R** describes the observation error, which combines **instrument error** and errors in the observation operator H (**representativeness error**):



4D-Var – observation operator

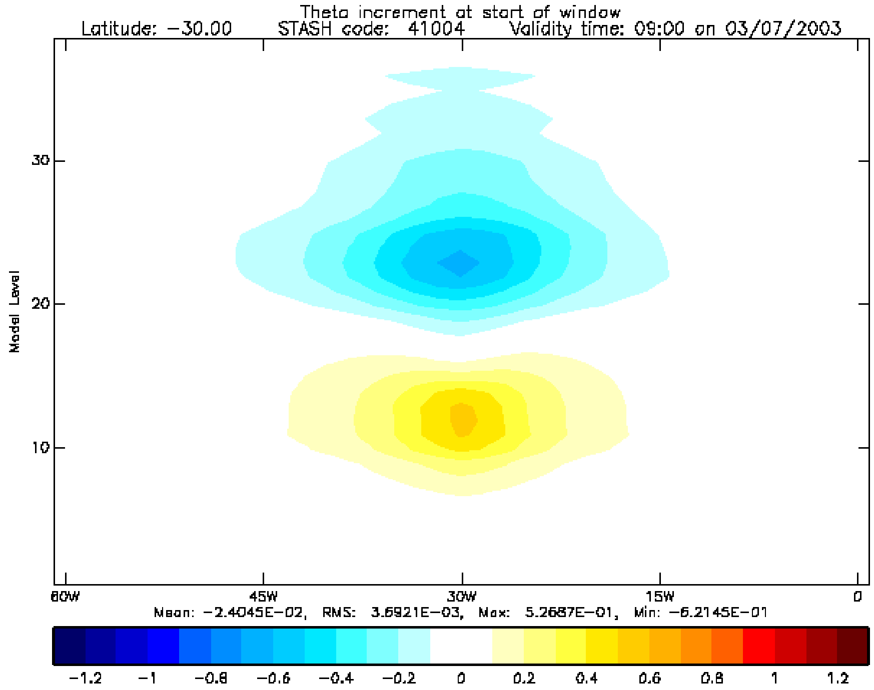
- In 4D-Var, the observation operator **H** includes model forecasts to the observation times:



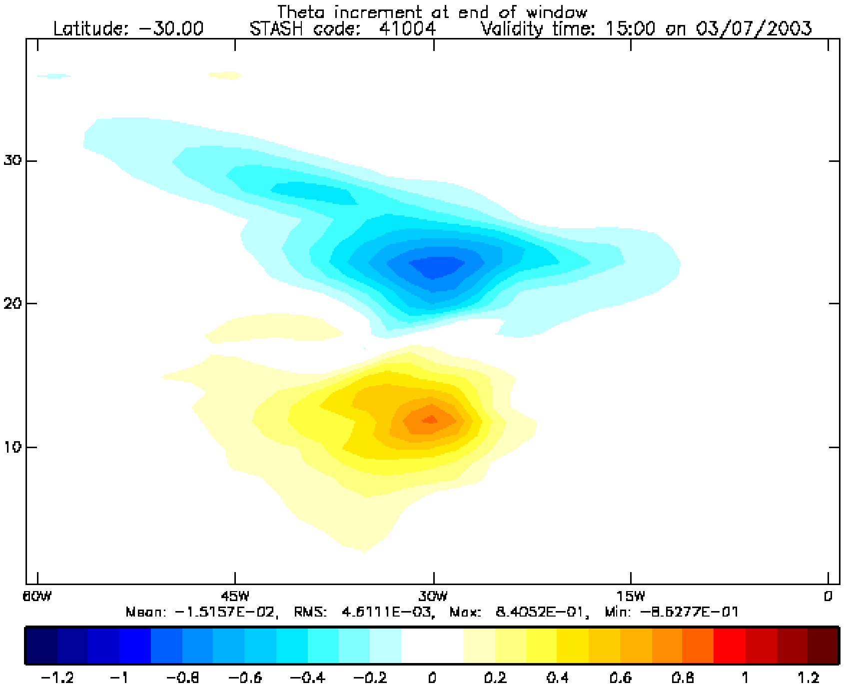
- Initial “background” forecast with nonlinear model
- 3D-Var either one value at centre of time window or forecast at time of observation and analyse increments assuming valid at analysis time FGAT
- Forecasts are updated with a simplified linear model; one forecast plus one **adjoint** integration per iteration. (Usually run about 50 iterations.)

4D-Var – background error

- In 4D-Var, the background error covariance matrix **B** is implicitly evolved by the linear model:



Pseudo pressure ob at beginning of window



Pseudo pressure ob at end of window

- This imposes some degree of dynamical consistency on the increments, and is probably a key advantage of 4D-Var.

Observations: Bias Correction and Variational Quality Control

- The occurrance of observations with gross errors can be handled
  - either in a prior Bayesian quality control, rejecting those not likely to come from the assumed Gaussian distribution characterising "good" observations
  - or by altering the variational penalty function to have a similar effect.
- In either method it is important to have a reliable estimate of the background error variance for that case, otherwise we can end up rejecting just those observations which show up a significant error in the background state.
- Theory assumes unbiased data

Observations: Quality Control

Conventional observations - Checks for:

- Physically plausible
- Position (e.g. ships over land)
- Track (movement since last report)
- Buddy checking (against neighbours)
- Model background O-B comparison
- Rejection lists from regular monitoring (O-B, O-A)

Satellite bias correction

Both background and observation can be biased

- Correct for O-B difference bias
- Recalculate each month
- Vary strongly with scan position and latitude
- Scan – difference from nadir
- Airmass – linear regression
- If not possible reject data

$$y_{cor} \quad y_{raw} \quad c_{scan} \quad c_1 y_1 \quad .. \quad c_n y_n \quad c_{air}$$

Satellite quality control

- Bad raw radiance T
- O-B threshold test
- Bad scan position
- Masked out
- Error in RTTOV
- Failed to converge
- High altitude
- Bad retrieved brightness temperature

Observations: Radar Errors

Is radar data biased? What does bias depend on?

Need to bias correct the data and quality control it either externally or using variational quality control eg as at ECMWF

How do we specify error variances for radar data?

- Precipitation rates
- Reflectivity
- Doppler radial winds
- Refractivity

Do we have an absolute measure?

However in fact the errors need to be relative to errors for background and other observation types (tuning)

Errors correlated as in satellite scans

Observations: Radar Radial Wind Errors

Sources of error

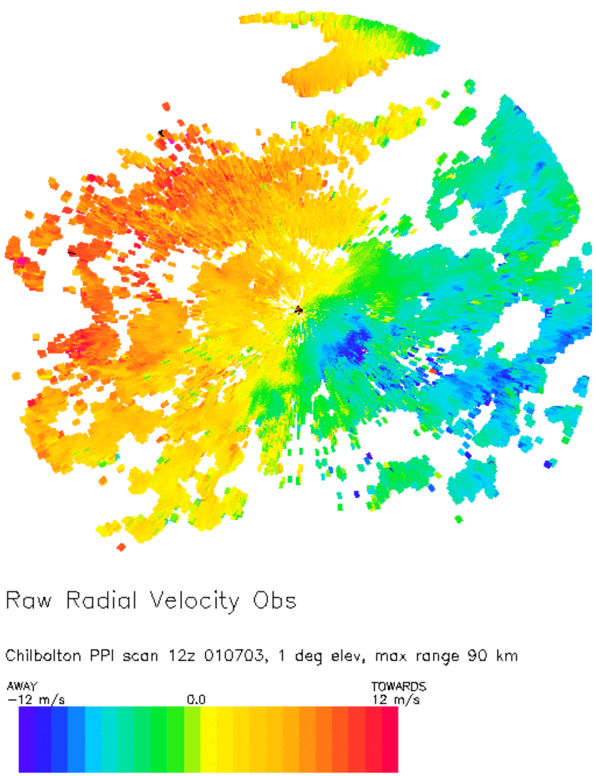
Ground clutter, anomalous propagation, sea clutter, velocity folding, noise

Velocity gradient in pulse volume

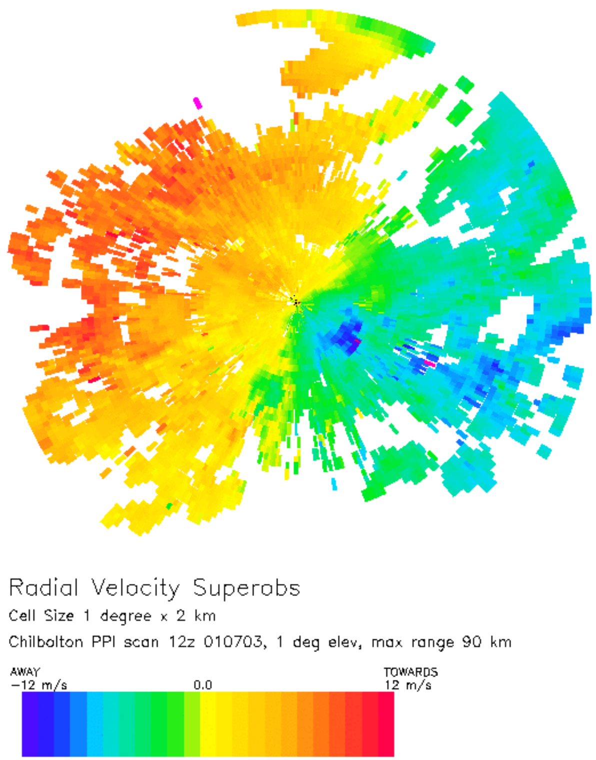
?

Superobbing of radar doppler winds

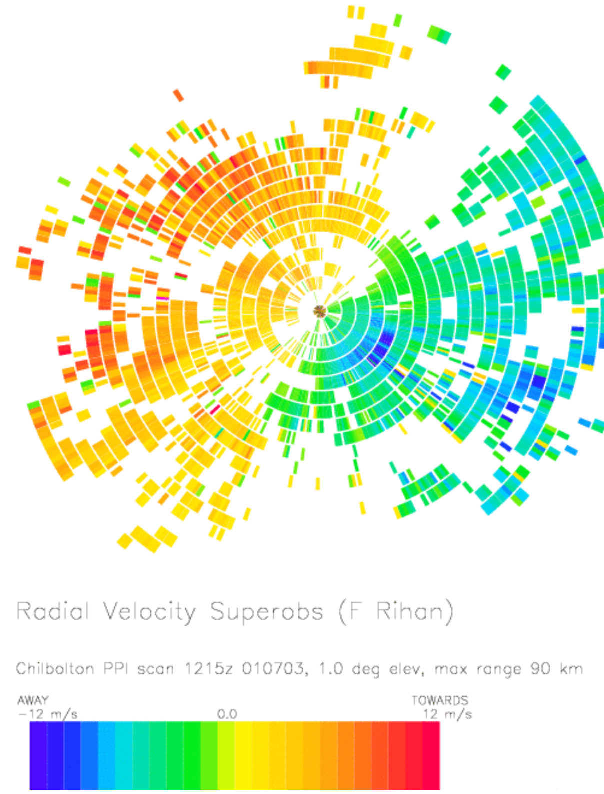
Raw data



Superobbed data
Difference from background method



Superobbed data
Remapping method

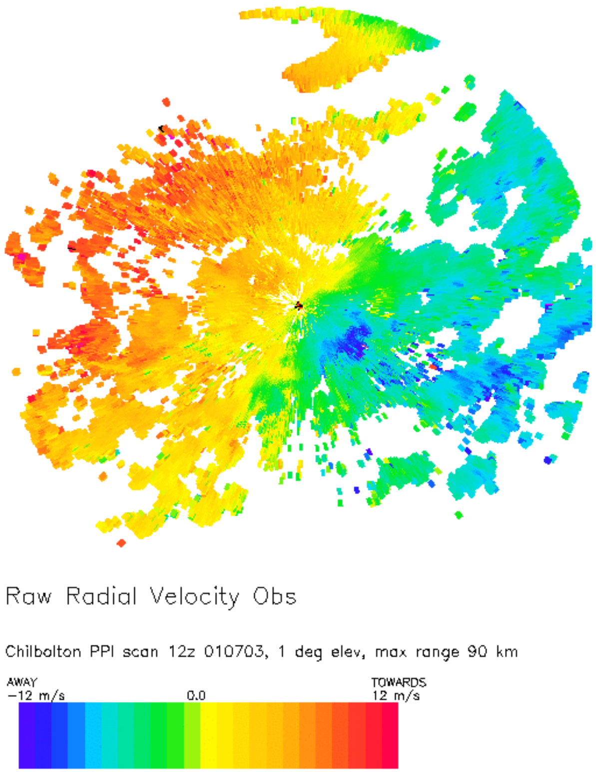


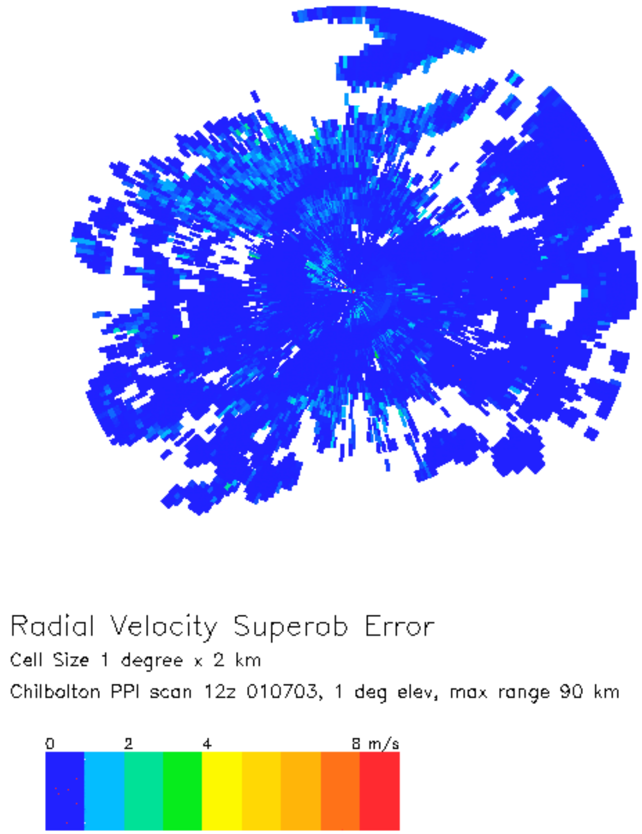
Errors of Superobbed radar doppler winds

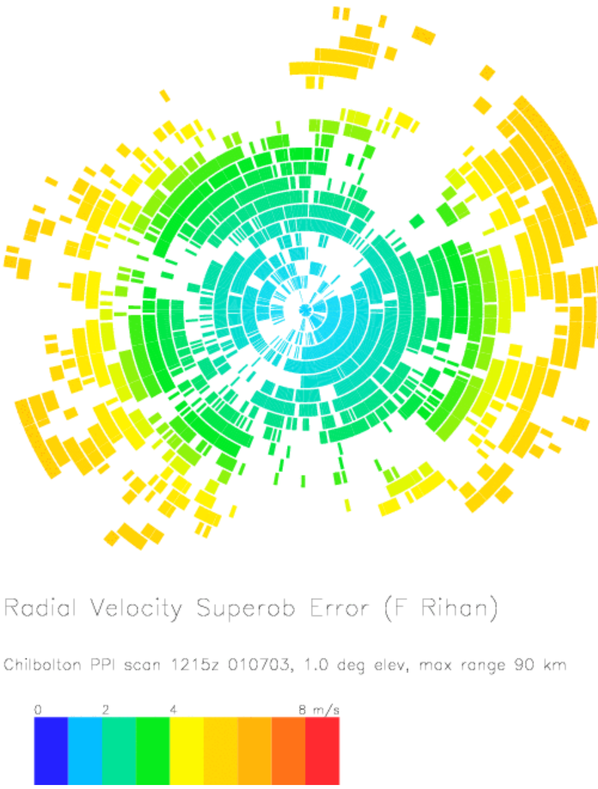
Raw data

Superobbed data
Difference from background
Method – observation errors

Superobbed data
Remapping method
Total observation error







Observations: Summary

- Need sophisticated quality control
- Need specification of error as variance/standard deviation in units of observation
- Ideally need to allow for correlation of errors
- Need to thin data or perform super-obbing

High resolution data assimilation

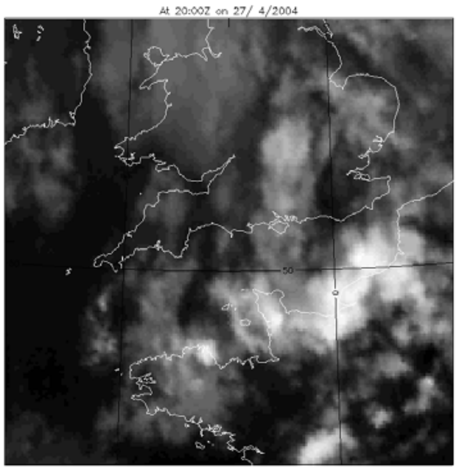
4km 3D-VAR with continuous cycles

With or without moisture and Latent Heat Nudging (LHN) using AC scheme

(referred to as MOPS data – moisture observation processing system)

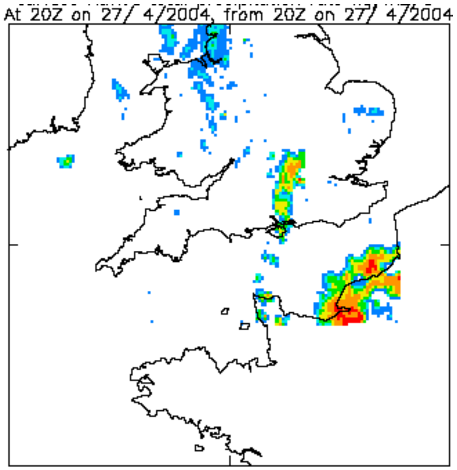
- IAU – increments output from 3D-Var and fixed over time window
- AC scheme – increments depend on latest model fields so vary with timestep through weighting factor and model evolution/impact of data

Moisture Observation Preprocessing



Satellite data

Surface reports



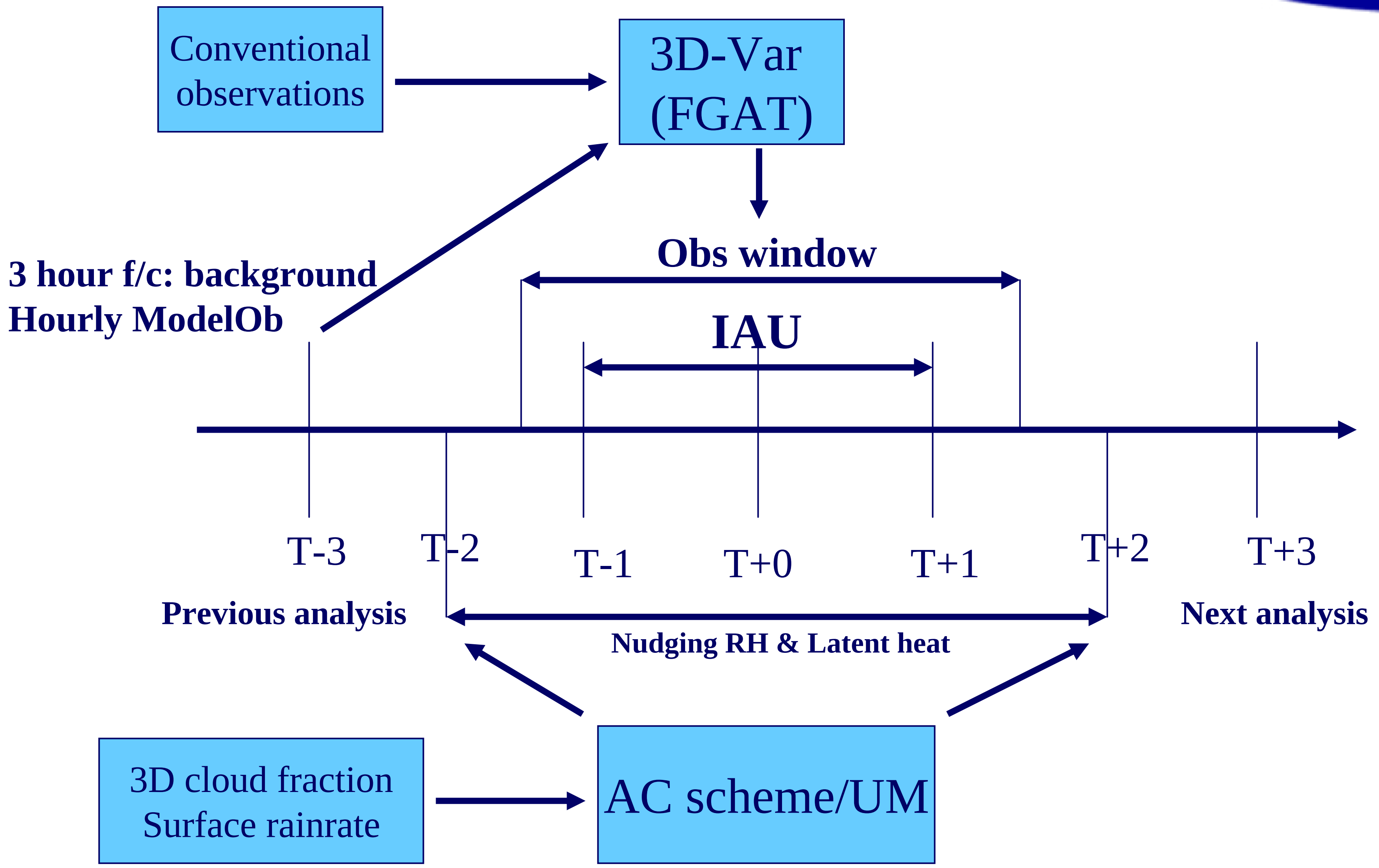
Radar data

Precipitation
5km
smoothed to
15km
Hourly
Testing 15min

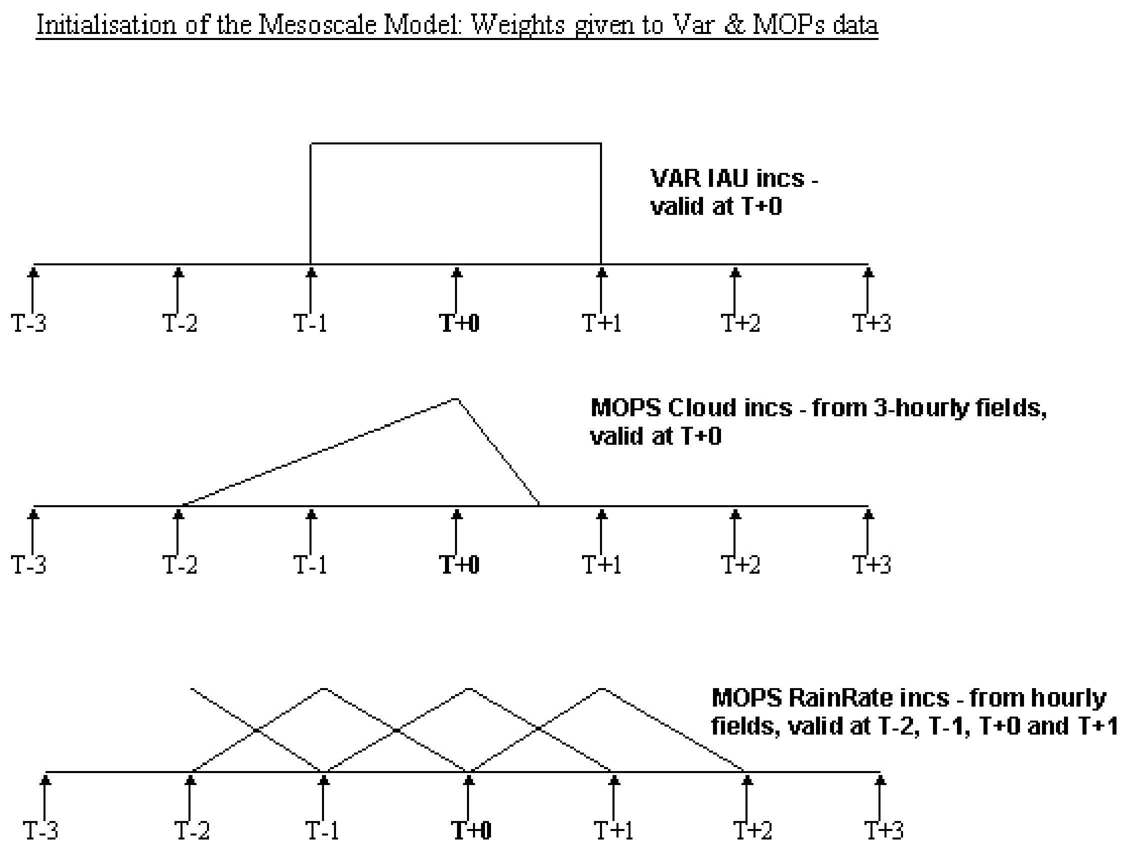
3D Cloud fraction → **3D Relative humidity** → **Nudge model state**

- Resolution: 15km, 3 hours

  (Testing 1 hour)


Conventional observations
3D-Var (FGAT)
Obs window
3 hour f/c: background
Hourly ModelOb
IAU
T-3
T-2
T-1
T+0
T+1
T+2
T+3
Previous analysis
Next analysis
Nudging RH & Latent heat
3D cloud fraction
Surface rainrate
AC scheme/UM

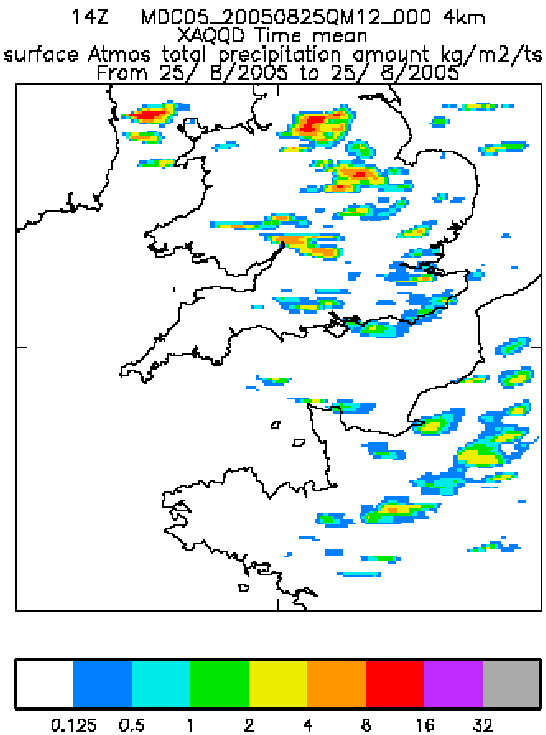
Period over which observations and analysis increments are nudged into Unified Model



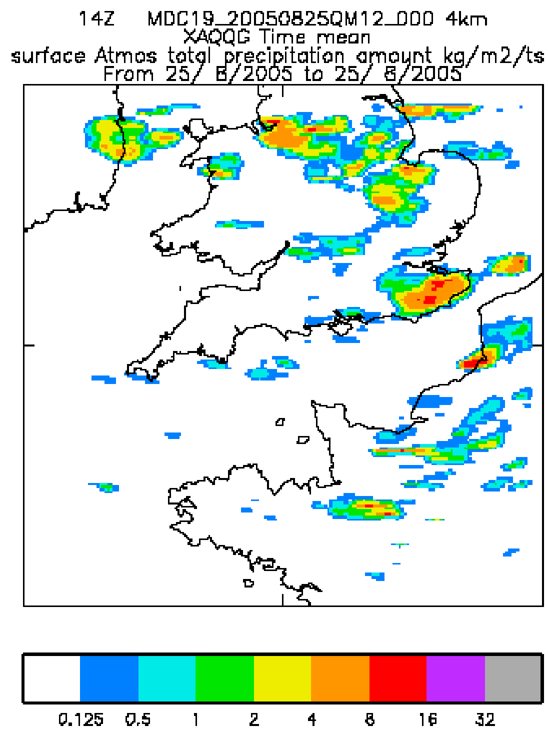
Impact of cloud and precipitation data

14UTC 25 August 2005 – CSIP IOP 18

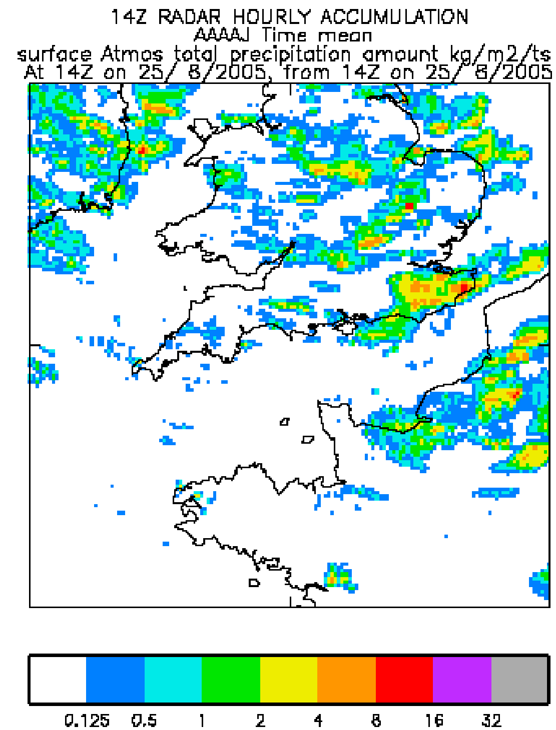
T+2 forecast
No MOPS data



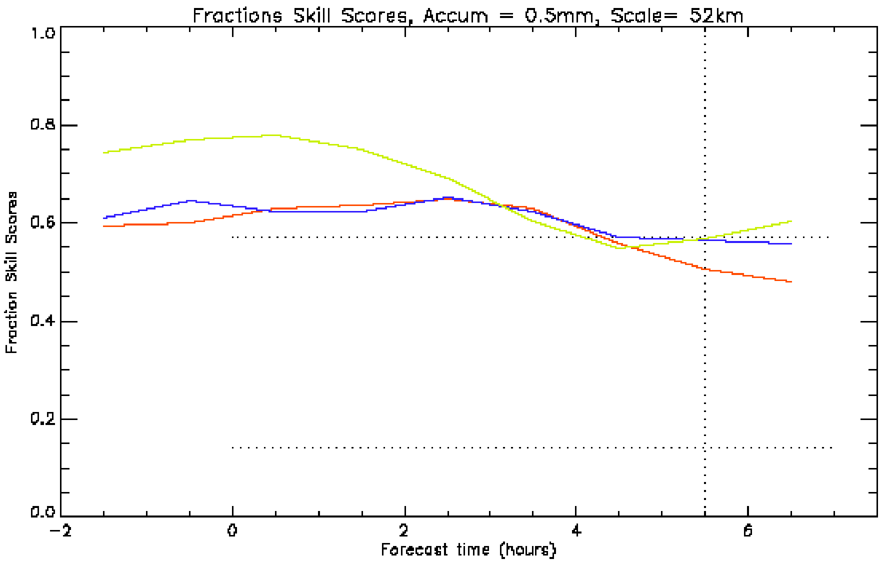
T+2 forecast
15min precip and hrly cloud



Radar
1 hour accumulation

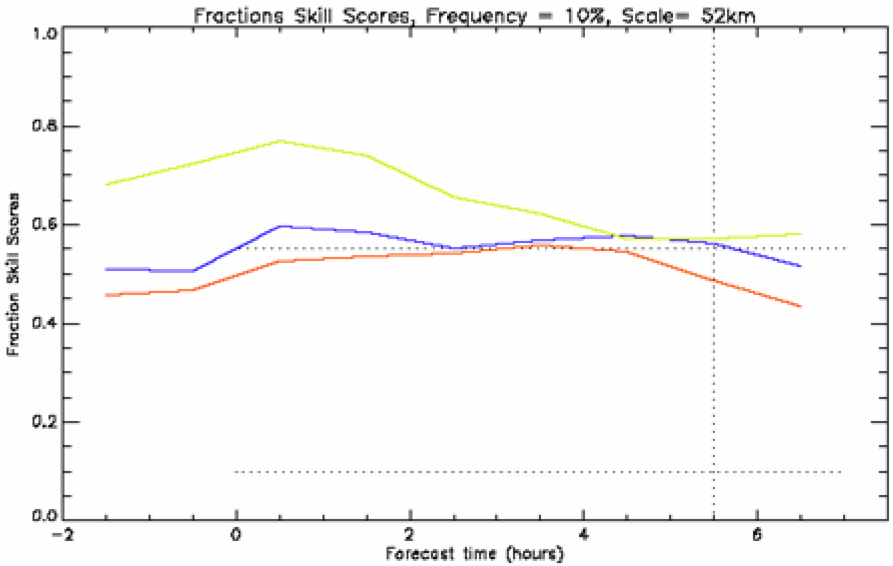


Impact of cloud and precipitation data



Skillscore for accumulation > 0.5mm

Skillscore for top 10% accumulations

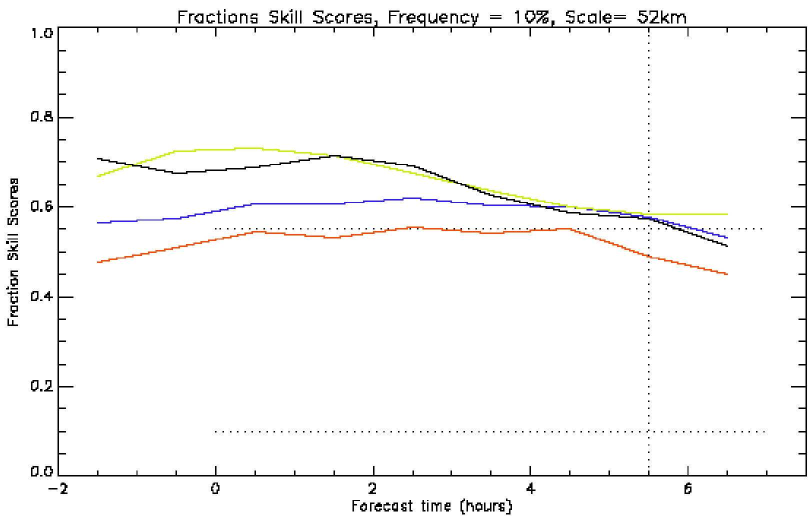


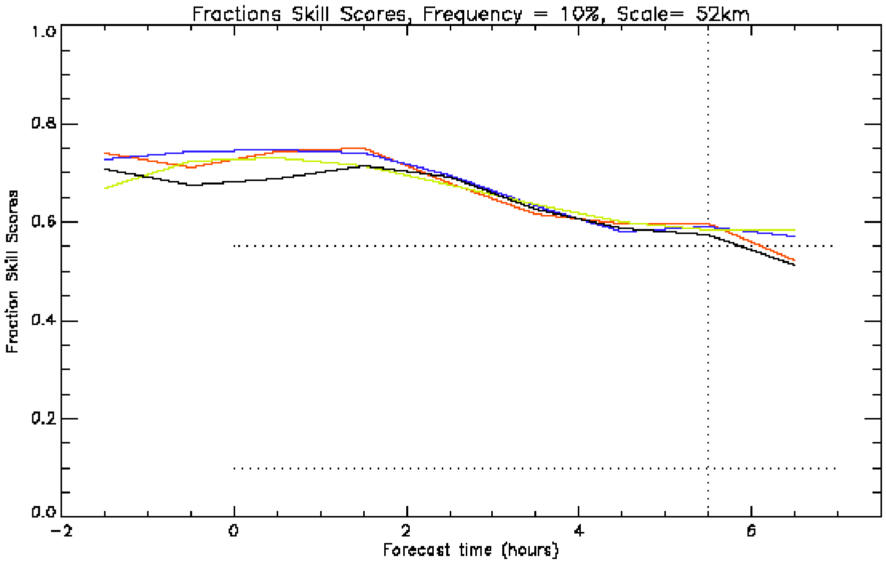
____ 15min precip, hrly cloud reduced filtering

____ hourly precip, 3 hr cloud full filtering

____ no precip and cloud

Impact of cloud and precipitation data
5 cases





Skillscore for top 10% accumulations

____cloud = 3hrs; rain = 1hr; no filter
____cloud = 1hr; rain = 15min; no filter
____ cloud=3hrs; rain=1hr
full filtering fwhm=42km
____ no precip and cloud

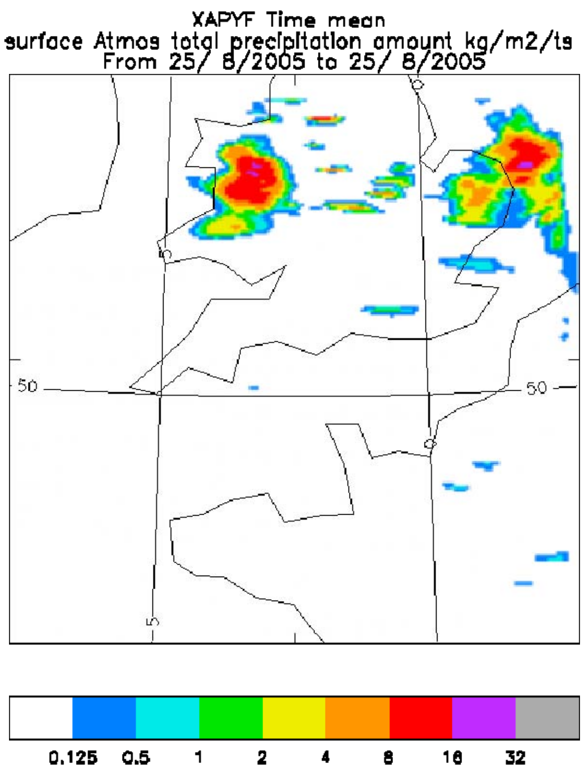
Skillscore for top 10% accumulations

____cloud = 3hrs; rain = 1hr; no filter
____cloud = 1hr; rain = 15min; no filter
____ cloud=3hrs; rain=1hr; no filter
diagnostic rain
____ cloud=1hrs; rain=15min; no filter
no subcloud LHN

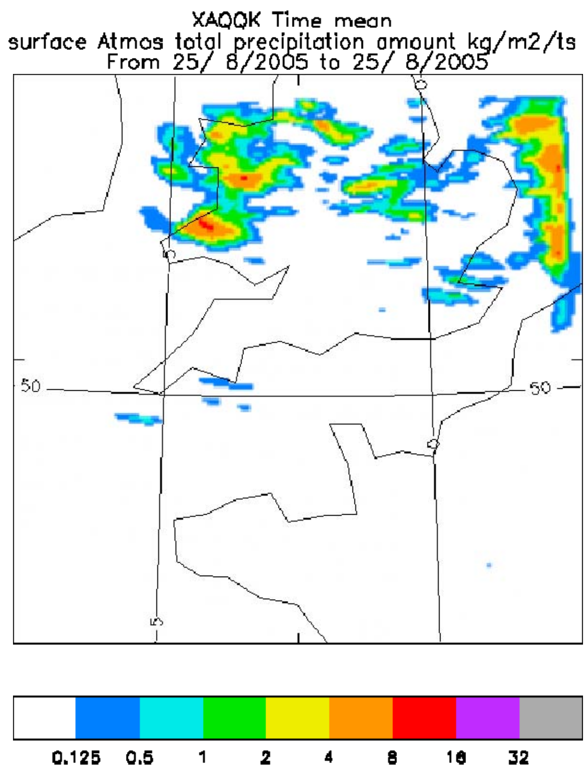
Impact of cloud and precipitation data

17UTC 25 August 2005 – CSIP IOP 18

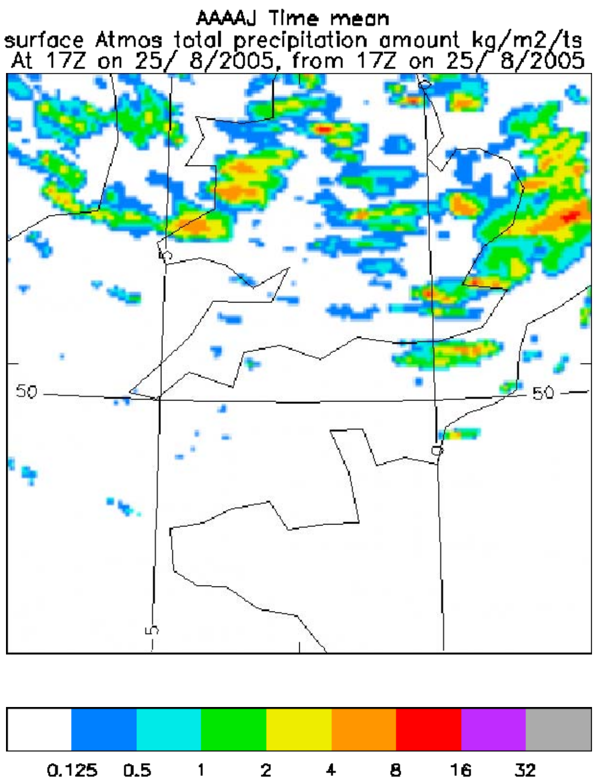
T+2 forecast
42km filter
cloud=3hr rain=1hr



T+2 forecast
1km filter(no filter)
Cloud=3hr rain=1hr



Radar
1 hour accumulation



Conclusions

- Analysis is very sensitive to specification of background errors and observation errors
- Need to bias correct and quality control data
- Specification of errors tends to be a matter of tuning to find best overall forecast skill score
- Need careful balance of errors and data quantity for different data types to get best forecast

Future directions

1. Dealing with model error. Statistical characteristics of model error are difficult to determine. However, relatively simple models may bring benefits.
2. Incorporating `errors of the day'. Ie. making the background error covariances more synoptically dependent. Also tying them to boundary layer depth.
3. Ensemble techniques – enables calculation of background error distribution/covariances – time evolving and allows for observation and forecast uncertainty
4. Variational quality control and bias correction
5. Specification of bias correction, quality control, thinning/superobbing, errors for radar doppler radar winds, reflectivity and refractivity data
6. Allow for correlated observation errors

Questions & Answers