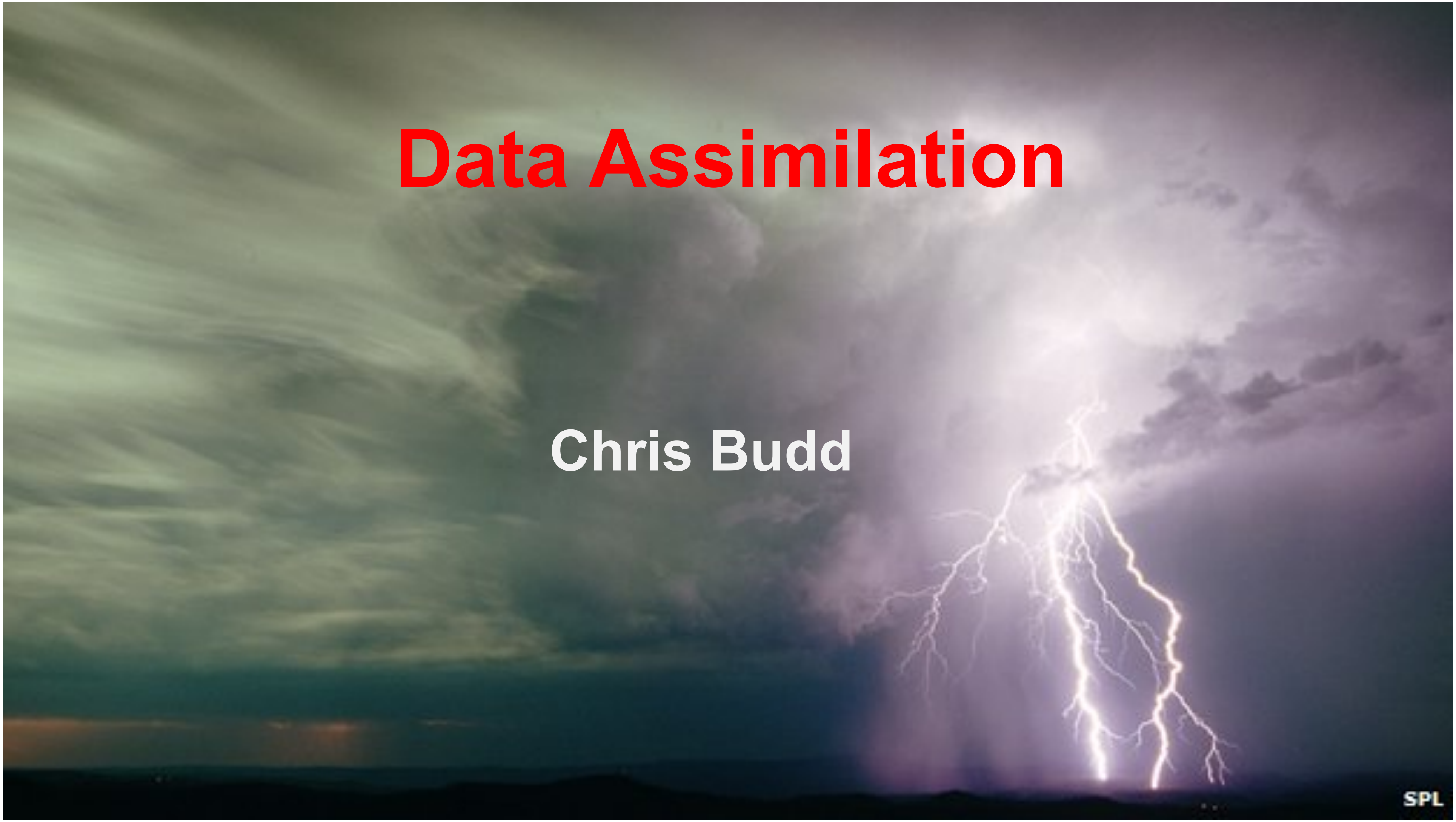
Data Assimilation

Chris Budd

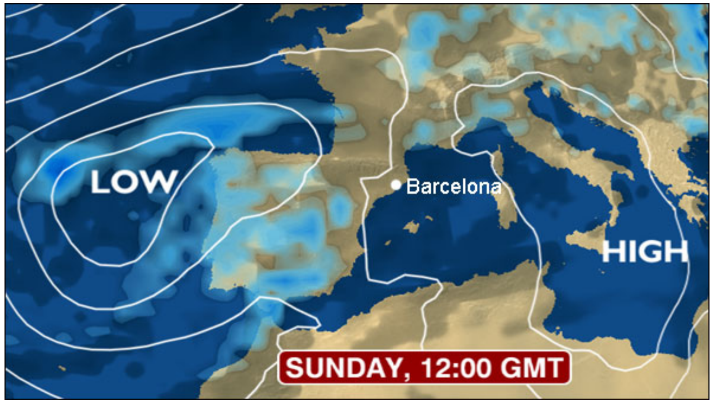


Institute for Mathematical Innovation

Met Office

UNIVERSITY OF BATH

Understanding and forecasting the weather is essential to the future of planet earth and maths place a central role in doing this



Accurate weather forecasting is a mixture of

- Careful modelling of the complex physics of the ocean and atmosphere
- Accurate computations on these models
- Systematic collection of data
- A fusion of data and computation

Data assimilation is the optimal way of combining a complex model with uncertain data

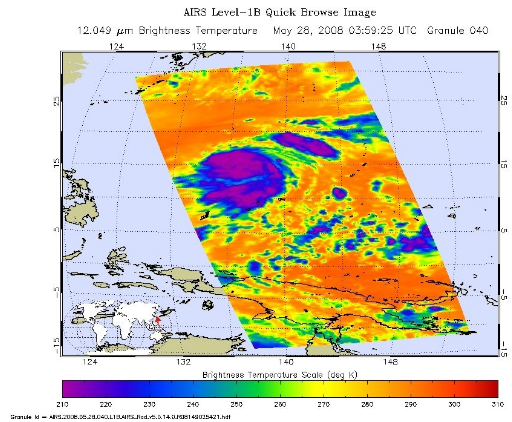
Basic Idea of Data Assimilation

True state of a system is x_t

A calculation (eg. NWP) gives a predicted state x_b
with an estimate of the error

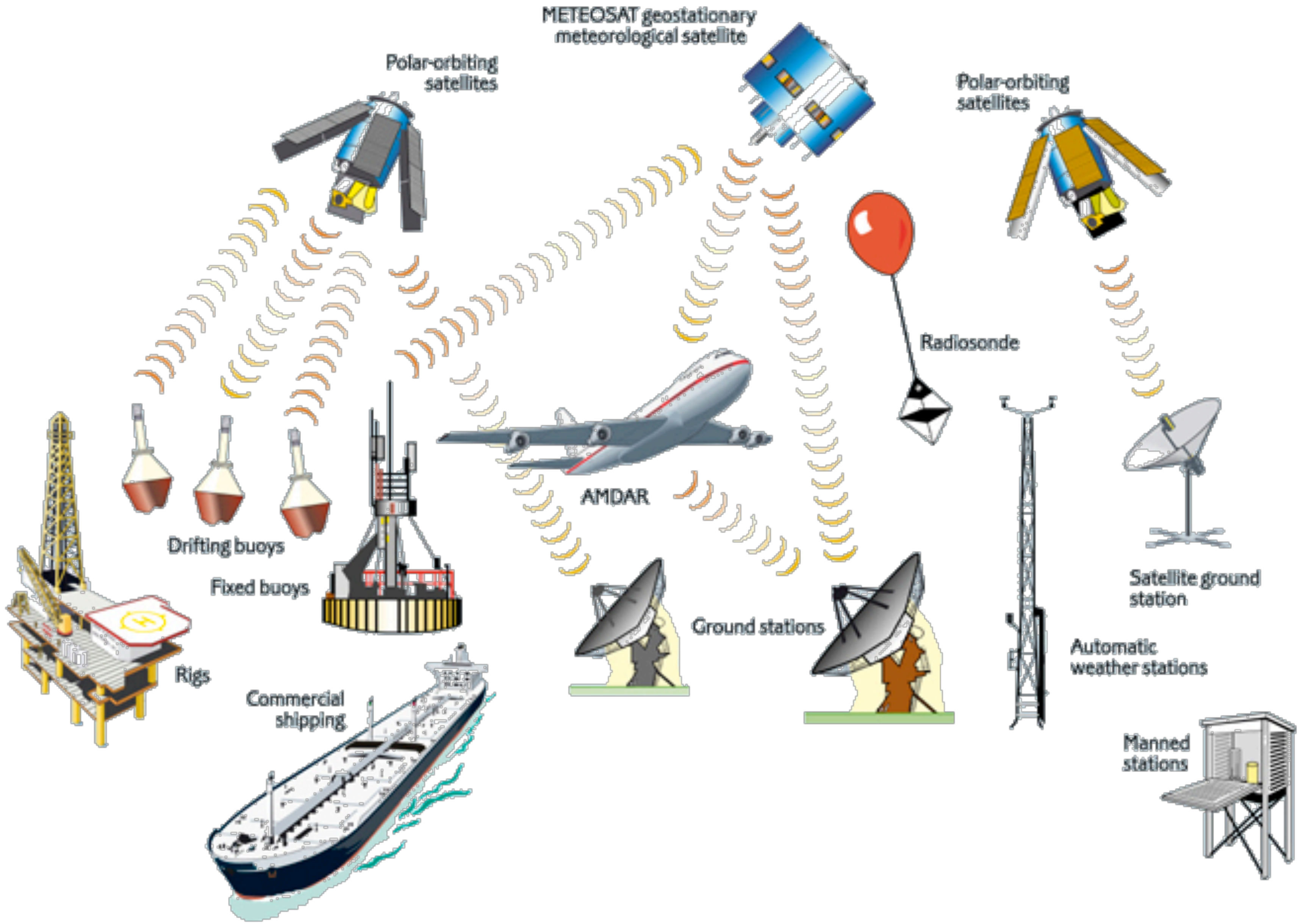
Make a series of observations y of some function $H(x_t)$ of the true state

Eg. Limited set of temperature measurements with error



Now combine the prediction with the observations

Data: Sources of observation



Both the prediction and the data have errors.

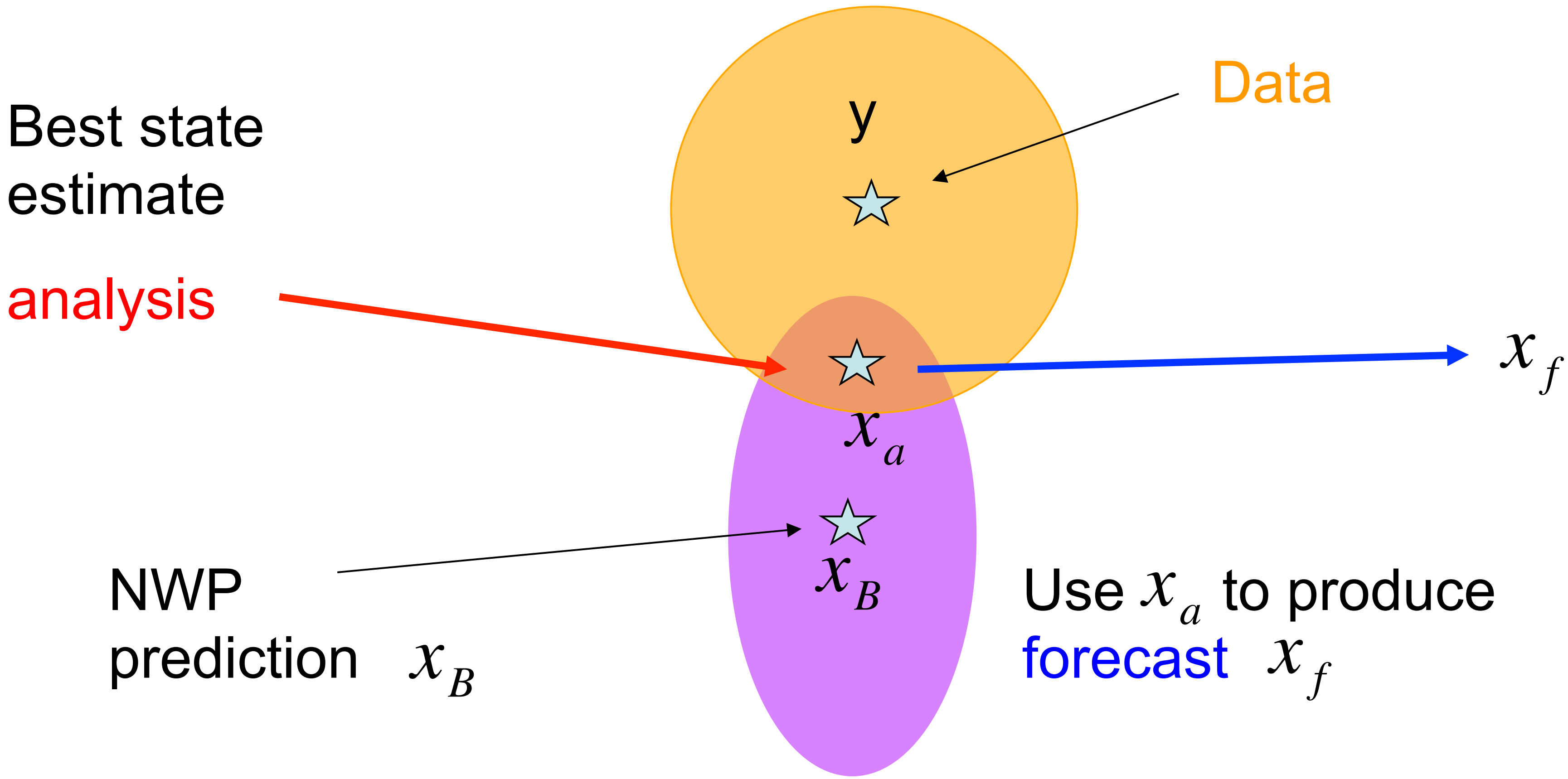
Can we optimally estimate the system state which is consistent with both the prediction and the data and estimate the resulting error?

NOTE: In weather prediction we have approximately

10^9 degrees of freedom

10^6 data points

So significantly underdetermined problem



Assume initially:

1. Errors are unbiased Gaussian variables
2. Data and prediction errors are uncorrelated
3. H(x) is a linear operator

Assumptions about the error

x_B

Data error: Gaussian, Covariance R

Background prediction error: Gaussian, Covariance B

Maximum likelihood of data y given truth x is

$$M = P(x|y)/P(x) = e^{-J(x)}$$

$$\boxed{J(x_a) = \frac{1}{2}\left(x_a - x_b\right)^T B^{-1}\left(x_a - x_b\right) + \frac{1}{2}(Hx_a - y)^T R^{-1}(Hx_a - y)}$$

BLUE: Find x_a which maximises M

So x_a minimises J

Implementation 1.

If R and B are known then the best estimate of the analysis is

$$x_a = x_b + K(y - Hx_b), \quad K = BH^T(R + HBH^T)^{-1}$$

Covariance of the analysis error

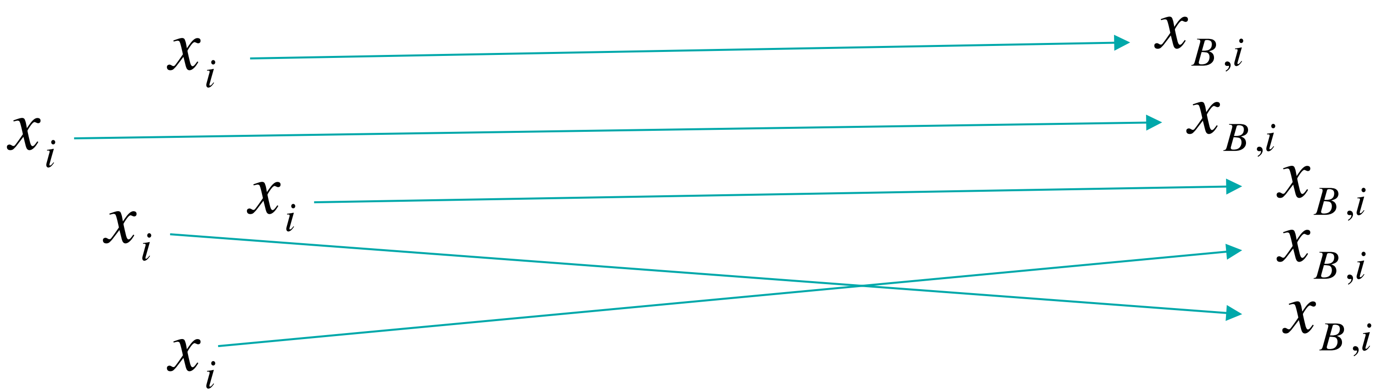
$$A = KRK^T + (I - KH)B(I - KH)^T$$

Kalman filter: Continuously updates the forecast and its error given the incoming data.

Ensemble Kalman Filter EnKF

This is a widely used Monte Carlo method that uses an ensemble of forecasts to estimate the terms in the Kalman filter

Idea: Take a large number of initial states x_i and estimate the resulting background states $x_{B,i}$



Estimate $\bar{x} = \frac{1}{N}\sum x_{b,i},\qquad B = \frac{1}{N-1}\sum \left(x_{B,i} - \bar{x}\right)\left(x_{B,i} - \bar{x}\right)^T$

Implementation 2:

Minimise the functional

$$J(x_a) = \frac{1}{2}\left(x_a - x_b\right)^T B^{-1}\left(x_a - x_b\right) + \frac{1}{2}(Hx_a - y)^T R^{-1}(Hx_a - y)$$

This is implemented as 3D-VAR (since 1999 in the Met Office)

x_B : Background, derived from 6 hour NWP forecast

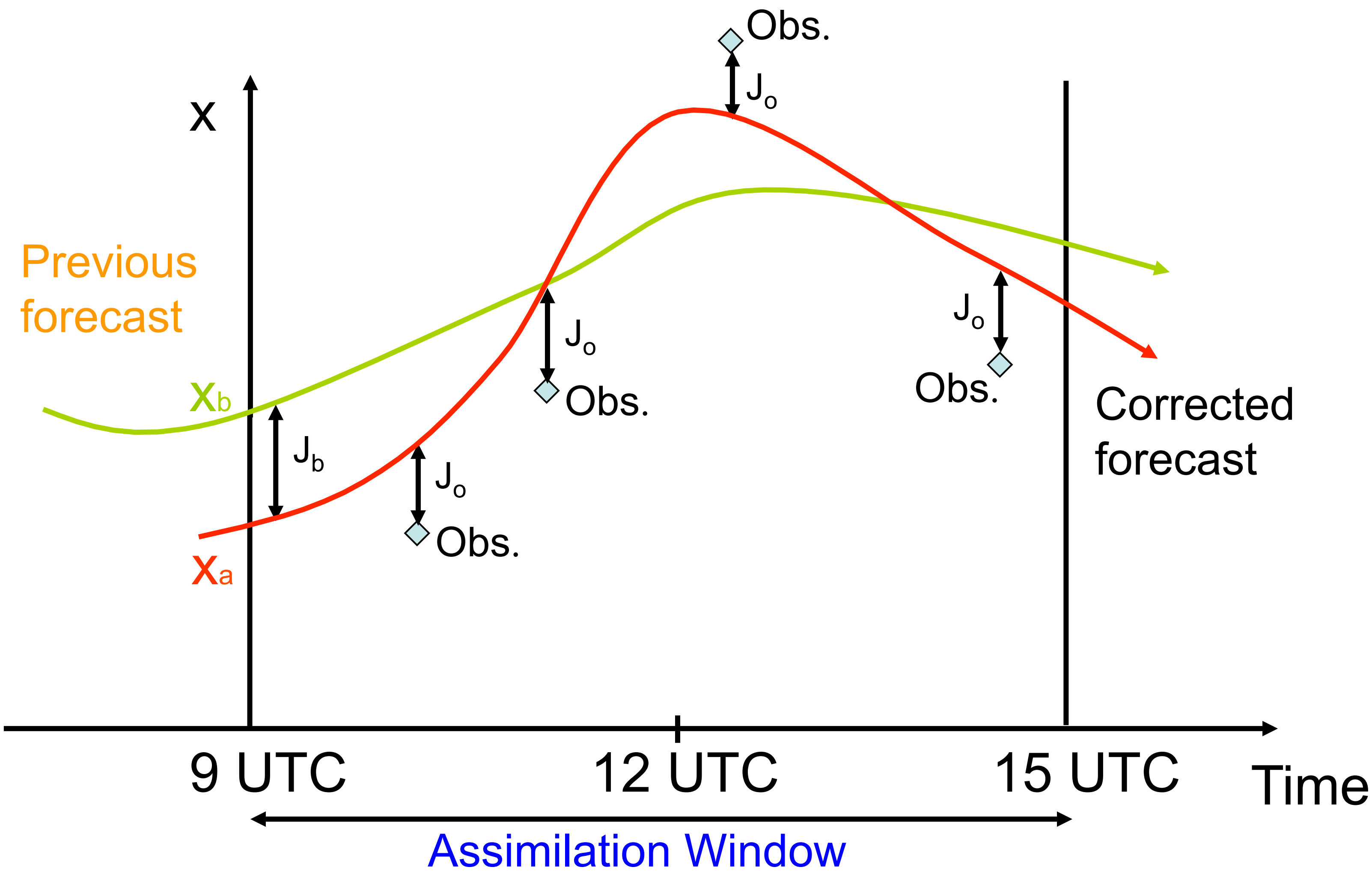
x_a : Analysis

x_f : NWP forecast using x_a as initial data

Implementation 3.

4D VAR … Preferred variational method

Use window of several observations (over 6 hours)



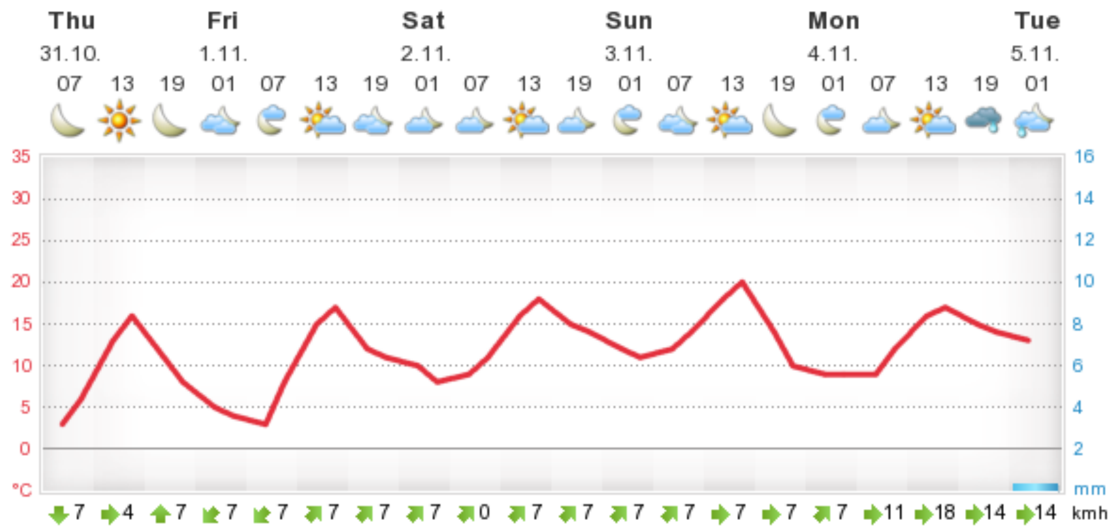
4D-VAR idea: Evolutionary model M (nonlinear)

Unknown initial state x_0

Times $t = t_0, t_1, t_2, \cdots$ Over a time window

Leads to state estimates $x_1, x_2, \ldots$

Data yi over window



Find x_0 so that the

estimates fit the data

Smoothing

Minimise

$$J(x_0) = \frac{1}{2}\left(x_0 - x_b\right)^T B^{-1}\left(x_0 - x_b\right) + \frac{1}{2}\sum_{i=0}^{N}(Hx_i - y_i)^T R^{-1}(Hx_i - y_i)$$

Subject to the strong model constraint

$$x_{i+1} = M_i(x_i)$$

At present assume perfect model, but can also deal with certain types of model error (both random and systematic) by using a weak constraint instead

Usually solved by introducing Lagrange multipliers

$$J(x_a) = \frac{1}{2}\left(x_0 - x_b\right)^T B^{-1}\left(x_0 - x_b\right) + \frac{1}{2}\sum_{i=0}^{N}(Hx_i - y_i)^T R^{-1}(Hx_i - y_i)$$
$$+\sum_{i=0}^{N-1}\lambda_i(x_{i+1} - M\, x_i)$$

And solving the adjoint problems

$$0 = \nabla J_i = H^T R^{-1}(Hx_i - y_i) + \lambda_{i-1} - \lambda_i\, M'_i\,(x_i)$$
$$0 = \nabla J_0 = B^{-1}(x_0 - x_b) + H^T R^{-1}(Hx_0 - y_0) - \lambda_0\, M'_0\,(x_0)$$

Estimation of the background and covariance errors

Good estimates of the covariance matrices R and B are important to the effectiveness of 3D-VAR

1. To get the physics correct
2. To avoid spurious correlations between parameters
3. To give well conditioned systems

NOTE: B is a very large matrix, difficult to store and very difficult to update. Impractical to calculate using the Fokker-Plank equation

Build meteorology into the calculation of B through Control Variable Transformations (CVTs)

IDEA: Choose more ‘natural’ physical variables χ which have uncorrelated errors so that the transformed covariance matrix is block diagonal or even the identity

Set $$\boxed{\delta x = U\chi = U_p U_v U_h \chi, \qquad B = UU^T}$$

Reduces the complexity of the system AND gives better conditioning for the linear systems

U_p^{-1} Separates physical parameters into uncorrelated ones eg. temperature, wind, balanced and unbalanced

U_v^{-1} Reduces vertical correlations by projecting onto empirical orthogonal vertical modes

U_h^{-1} Reduces horizontal correlations by projecting onto spherical harmonics

Effective, but errors arise due to lack of resolution of physical features leading to spurious correlations

[Cullen]

Eg. Problems with stable boundary and inversion layers and assimilating radiosonde data

Poor resolution leads to inaccurate predictions of fog and ice


© AFP/GETTY IMAGES

Solution one

Increase global resolution

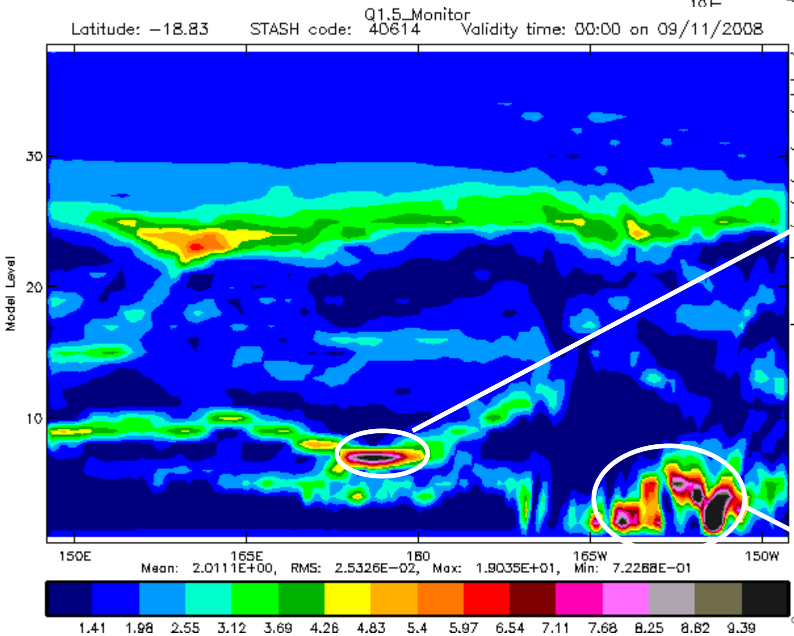
VERY EXPENSIVE!!!

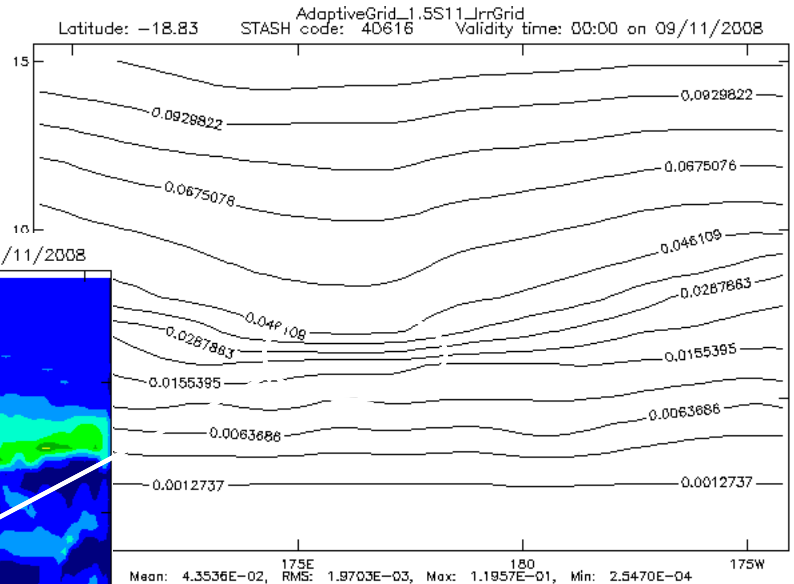
Solution two

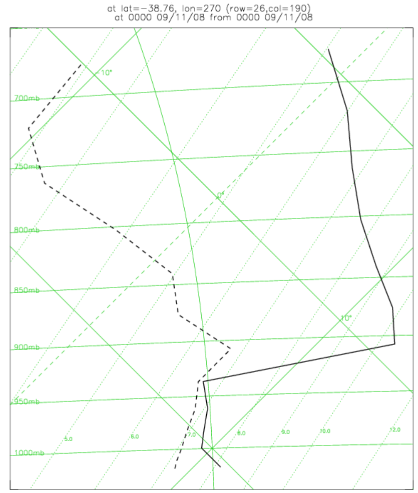
locally redistribute the computational mesh to resolve the features

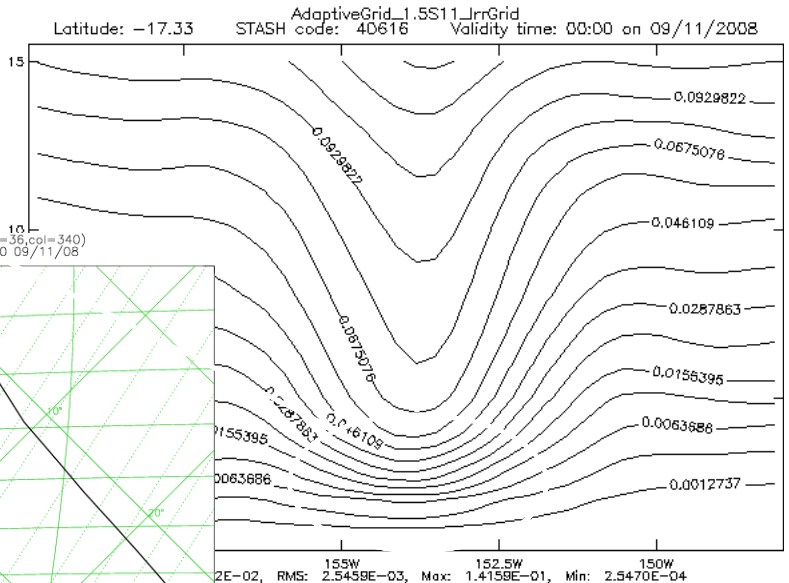
Cheap and effective! [Piccolo, Cullen, B,Browne, Walsh]

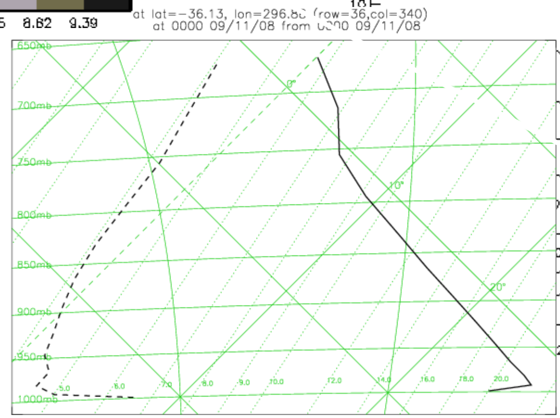
Monitor function and the Adaptive Grid









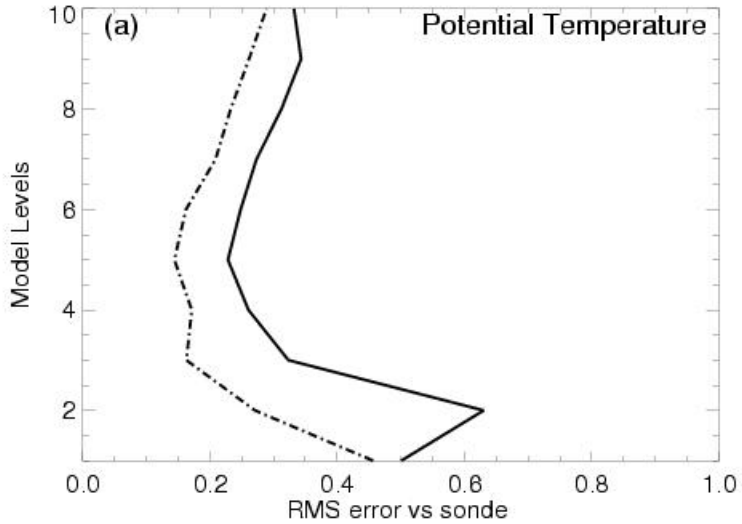


Piccolo&Cullen

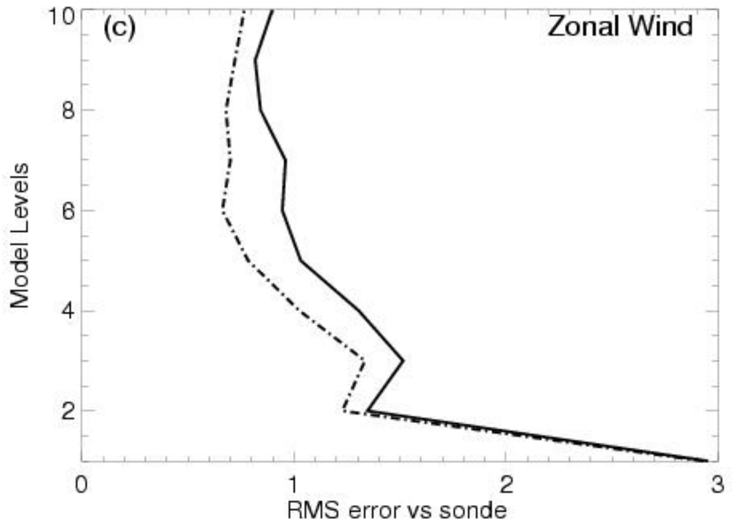
QJR Met Soc 2011

RMS error: Analysis - Observations

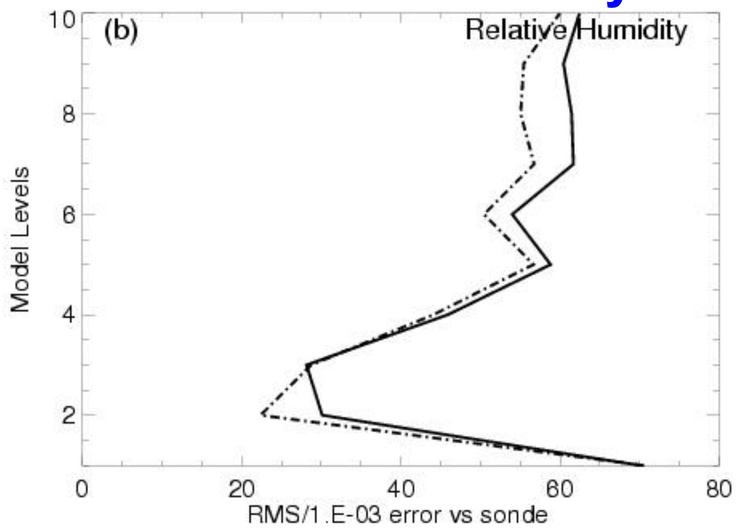
theta



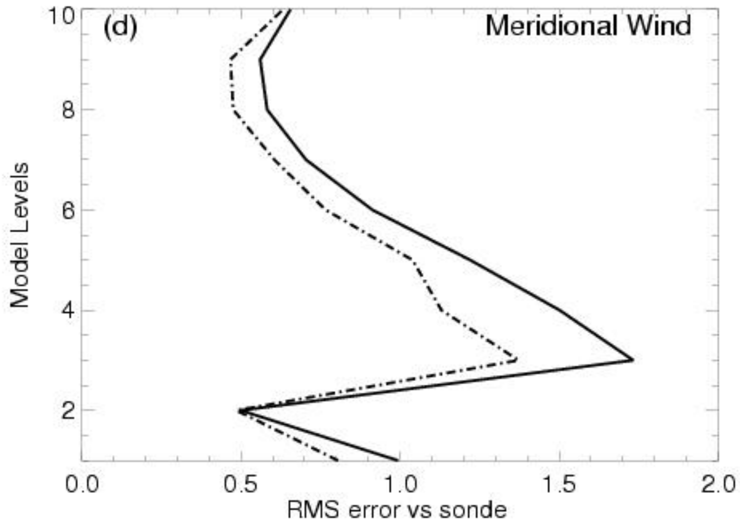
zonal wind



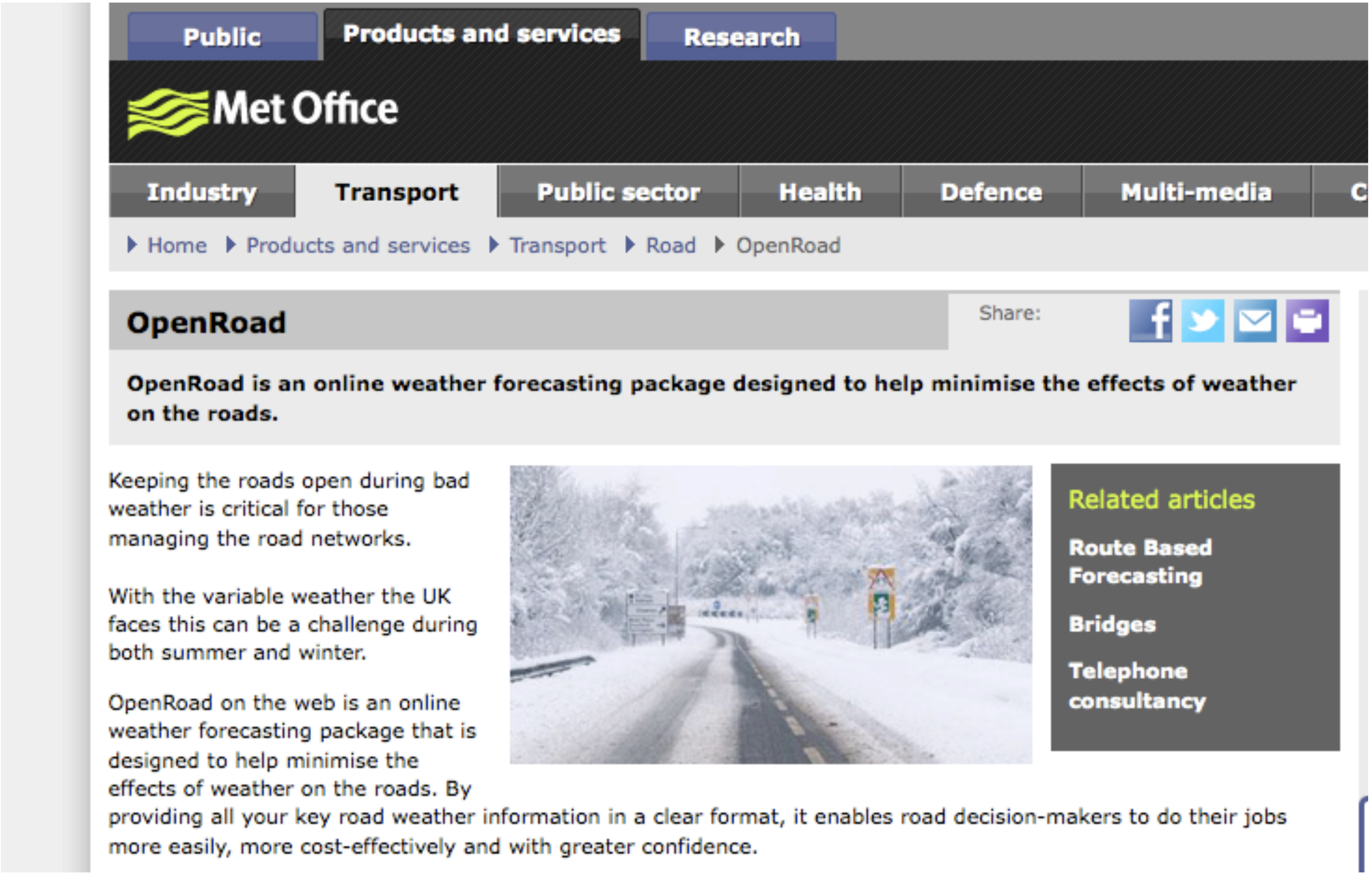
relative humidity



meridional wind



Used together with Met Office Open Road software to advise councils on road gritting over Christmas



Public | Products and services | Research

Met Office

Industry | Transport | Public sector | Health | Defence | Multi-media | C

Home ▸ Products and services ▸ Transport ▸ Road ▸ OpenRoad

OpenRoad

Share:

OpenRoad is an online weather forecasting package designed to help minimise the effects of weather on the roads.

Keeping the roads open during bad weather is critical for those managing the road networks.

With the variable weather the UK faces this can be a challenge during both summer and winter.

OpenRoad on the web is an online weather forecasting package that is designed to help minimise the effects of weather on the roads. By providing all your key road weather information in a clear format, it enables road decision-makers to do their jobs more easily, more cost-effectively and with greater confidence.

Related articles

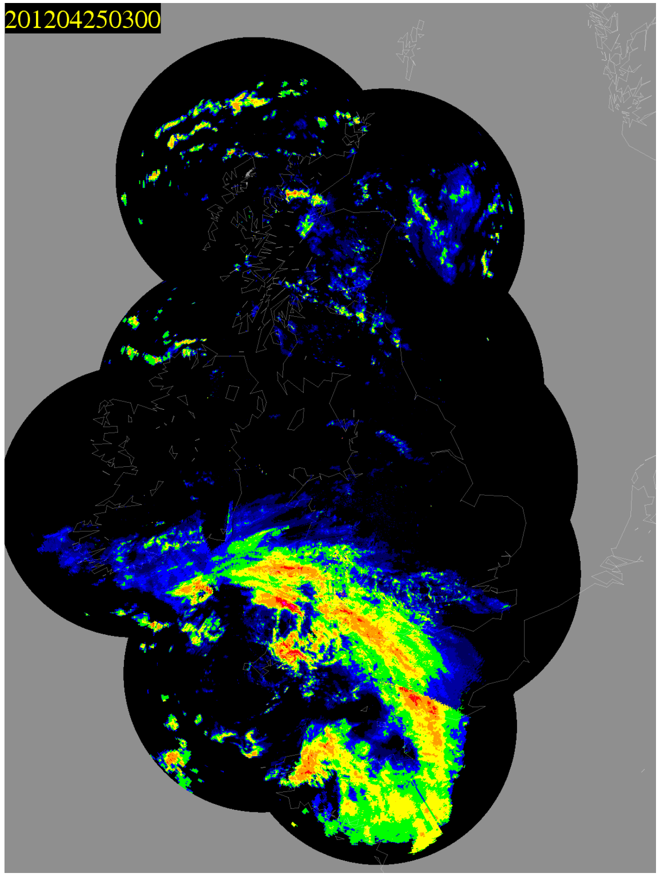
Route Based Forecasting

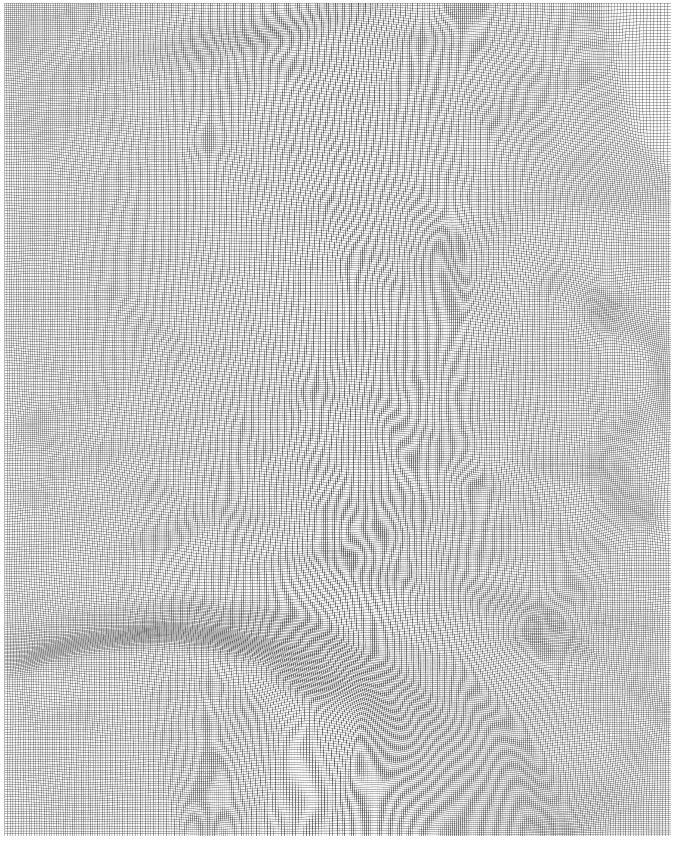
Bridges

Telephone consultancy

Adaptive mesh implemented operationally since November 2010.

Now extending it to a fully three dimensional implementation using optimal transport methods [B,Browne,McRae,Piccolo]





Dealing with nonlinearity

Lot of research into finding a compromise between dealing with the high dimensionality and nonlinearity in the system

Better use of appropriate (eg. Lagrangian) data

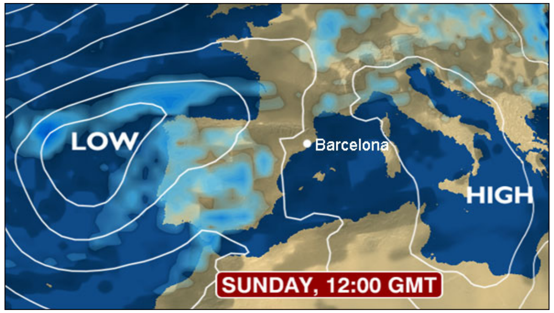
Tuning method to data [Jones, Stuart, Apte]

Use of particle filters

and MCMC methods [Peter Van Leeuwan]

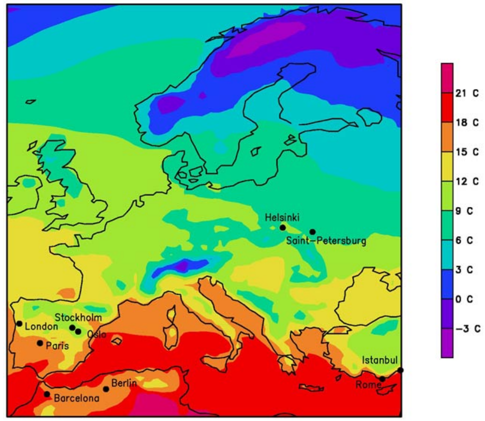
Conclusions

Data assimilation is an optimal way of merging models with data



Useful for model tuning, validation, evaluation, uncertainty quantification and reduction

Very effective in meteorology



Can be significantly improved with adaptivity and OT