Data assimilation for weather forecasting – G.W. Inverarity

What is data assimilation?

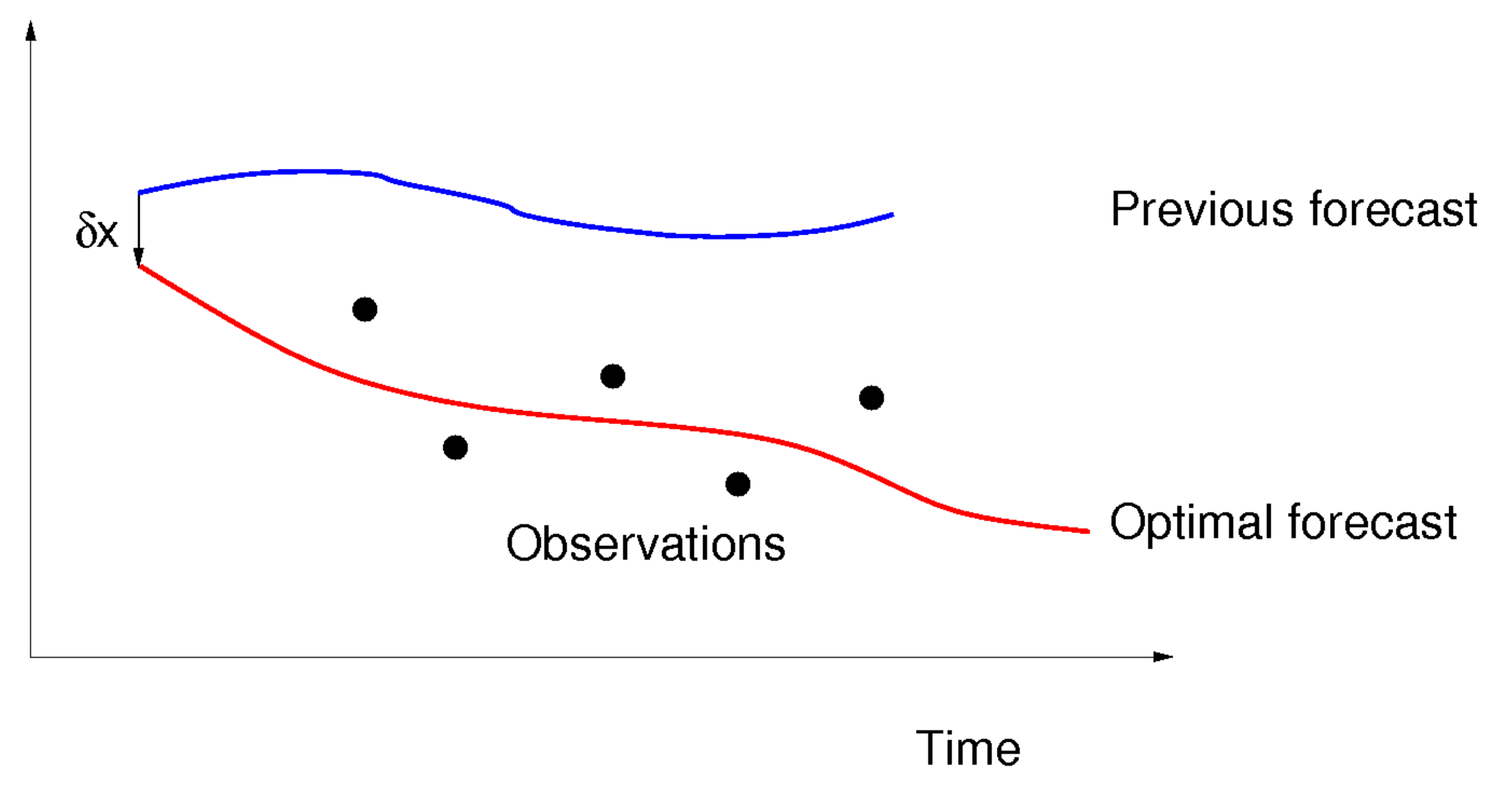
A weather forecast requires an estimate of the current state of the atmosphere to use as an initial condition for the forecast model. However, the number of points in the model grid is about ten times greater than the number of usable weather observations collected between forecasts. Furthermore, the observation locations are not distributed uniformly and the observations themselves are imperfect so simple interpolation does not provide the best estimate of the atmospheric state. Fortunately, we do not have to rely only on the observations as we have a gridded approximation of the atmospheric state in the form of the previous forecast valid over the period of interest. Like the observations, this so-called background state is imperfect. Data assimilation therefore seeks to merge the imperfect background state with the imperfect observations to produce an estimate of the atmospheric state.

Optimization

We adjust a previous forecast using recently collected observations spanning a time window (six hours for the global forecast) and minimize a nonlinear weighted least-squares cost function measuring the departure δ**x** of the adjusted forecast trajectory from the previous forecast at the start of the window $\mathbf{x}^b$ evolved by the forecast model $M_{i,0}$ to time index i and converted to a simulated observation using the observation operator H_i. Covariance matrices for the forecast error (**B**) and observation/ representation error ($\mathbf{R}_i$) are used in the Mahalanobis norm $||\mathbf{z}||_\mathbf{A}=(\mathbf{z}^\mathrm{T}\mathbf{A}^{-1}\mathbf{z})^{1/2}$.

$$J(\delta\mathbf{x}) = \frac{1}{2}\left\|\delta\mathbf{x}\right\|_\mathbf{B}^2 + \frac{1}{2}\sum_{i=1}^{N}\left\|\mathbf{y}_i - H_i(M_{i,0}(\mathbf{x}^\mathrm{b}+\delta\mathbf{x}))\right\|_{\mathbf{R}_i}^2$$

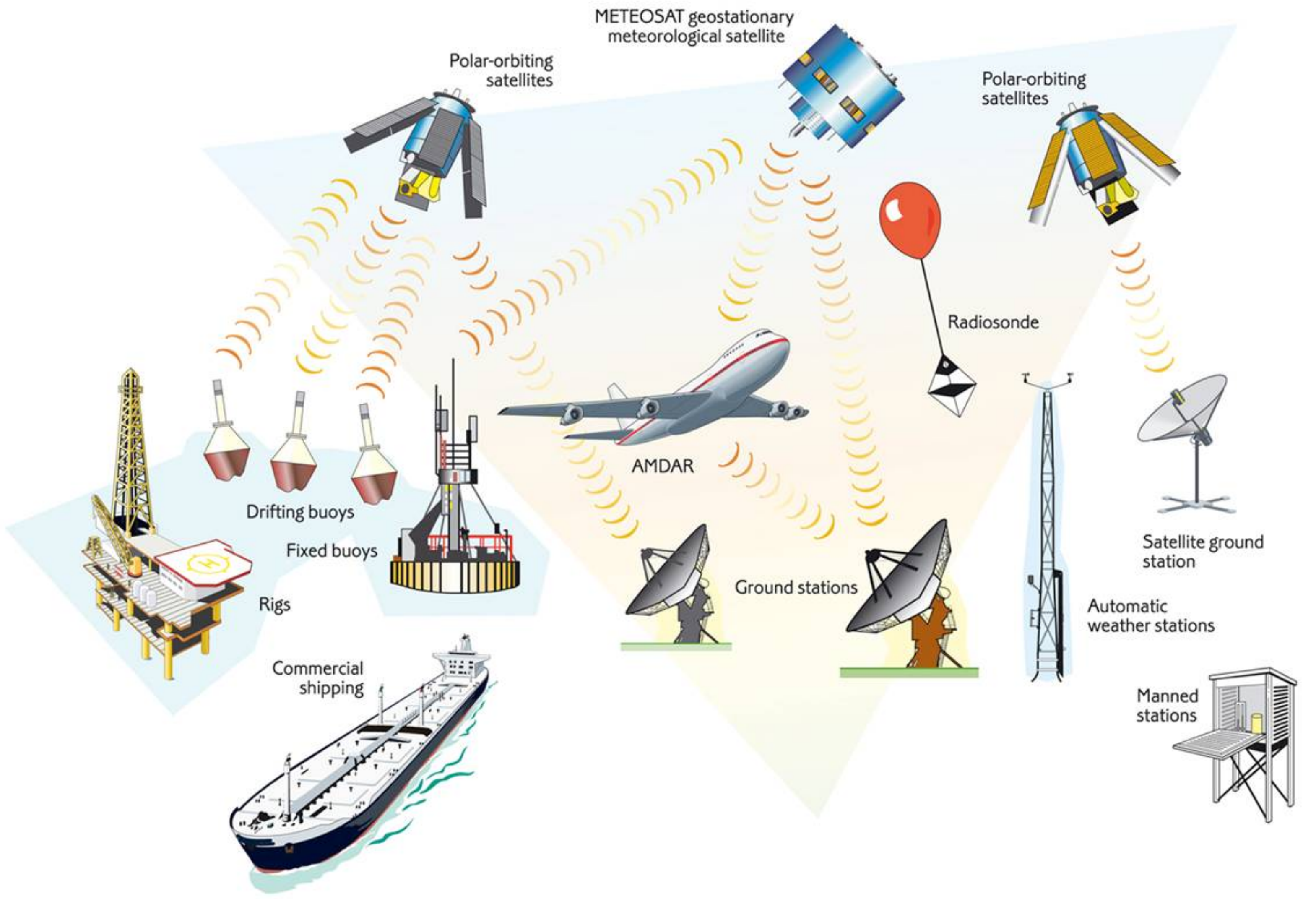
Unfortunately, δ**x** has order 10^9 elements so the **B** matrix has order 10^{18} entries – too many to be directly estimated or even stored. **B** is either modelled using a sequence of operators or estimated from a low-rank ensemble of forecasts representing the forecast error. Operationally we use a hybrid combination of both these approaches.



The need to produce timely forecasts for customers means that we have twenty minutes to minimize the cost function before the forecast step needs to start. A conjugate gradient method preconditioned with the leading eigenvectors of the cost function Hessian is applied to a sequence of quadratic approximations of the cost function, using 70 evaluations of the cost function and its gradient.

Observations

The global model assimilates order 10^6 observations every six hours. We use about a tenth of the available satellite observations, mainly due to difficulties in interpreting radiances in the presence of cloud and our knowledge of how different instrument channels are correlated.


METEOSAT geostationary meteorological satellite
Polar-orbiting satellites
Polar-orbiting satellites
Radiosonde
AMDAR
Drifting buoys
Fixed buoys
Rigs
Commercial shipping
Ground stations
Satellite ground station
Automatic weather stations
Manned stations

Current challenges

Having good covariance matrix estimates is critical to the success of our weighted least-squares algorithm. The number of satellite radiance observations that can be assimilated is limited by our ability to represent the correlations between different instrument channels. Meanwhile, most of the information in the optimal forecast comes from the previous forecast rather than the observations so a significant component of data assimilation research is directed at either modelling the **B** matrix as a steady-state representation of forecast error through the operators that transform to uncorrelated variables and impose homogeneity and isotropy assumptions or estimating the flow-dependent forecast error through a low-rank ensemble.

The above cost function assumes that the forecast model is perfect. Accounting for model error by allowing the cost function to depend on adjustments to the model trajectory throughout the observation window would help to separate model error from initial condition error but introduces the challenge of estimating the associated model-error covariance matrix.

It would be beneficial for our regional modelling, which uses lateral boundary conditions produced by the global model, if the global forecast/assimilation cycle could be run every hour with overlapping six hour observation windows.

Met Office FitzRoy Road, Exeter, Devon, EX1 3PB United Kingdom
Tel: 01392 885680 Fax: 01392 885681
Email: gordon.inverarity@metoffice.gov.uk