TRAINING
COURSE

EUMETSAT/ ECMWF NWP-SAF satellite data assimilation

NWP SAF
Numerical Weather Prediction

ECMWF

ECMWF/EUMETSAT NWP-SAF Satellite data assimilation Training Course

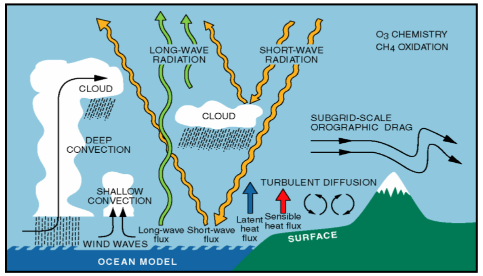
Data assimilation algorithms and key elements

What is Data Assimilation ?

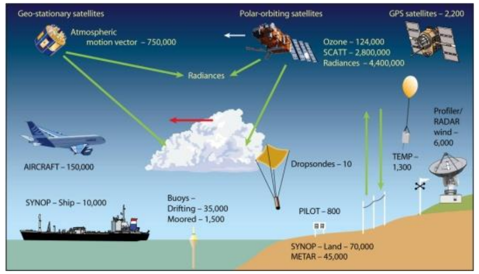
- Models give a complete description of the atmospheric, but **errors grow rapidly** in time
- Observations provide an **incomplete description** of the atmospheric state, but bring up to date information
- Data assimilation **combines** these two sources of information to produce an optimal (best) estimate of the atmospheric state
- This state (the *analysis*) is used as **initial conditions** for extended forecasts.

The assimilation system:

- Model



- Observations



- Assimilation algorithm

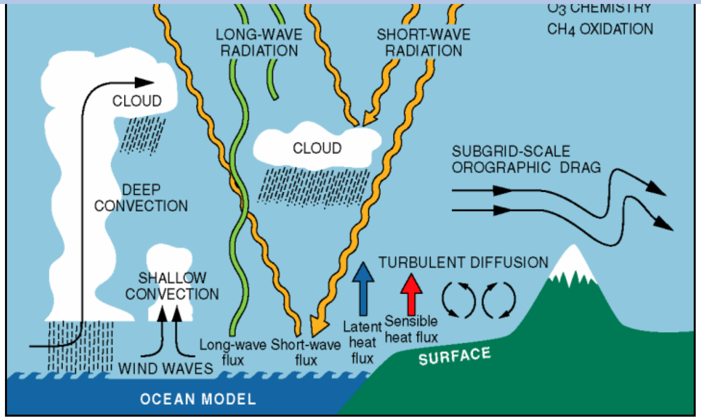
$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

The forecast model

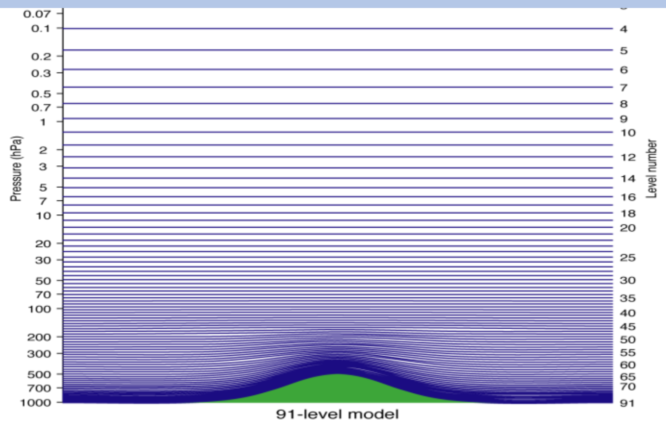
$$X_{t=0} \longrightarrow X_{t=t}$$

The forecast model

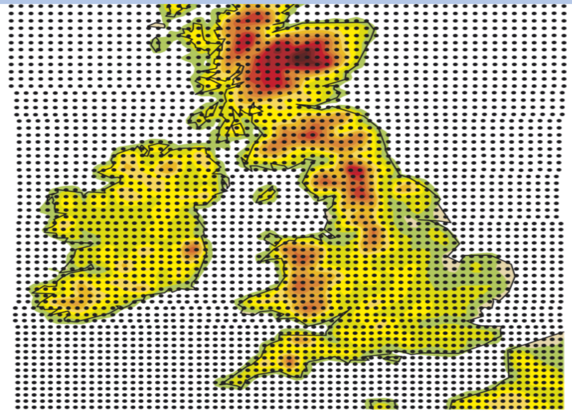
Physical and dynamical processes updated every 10 minutes



91 (**137**) vertical levels from the surface to 0.01hPa (approx: 80Km)

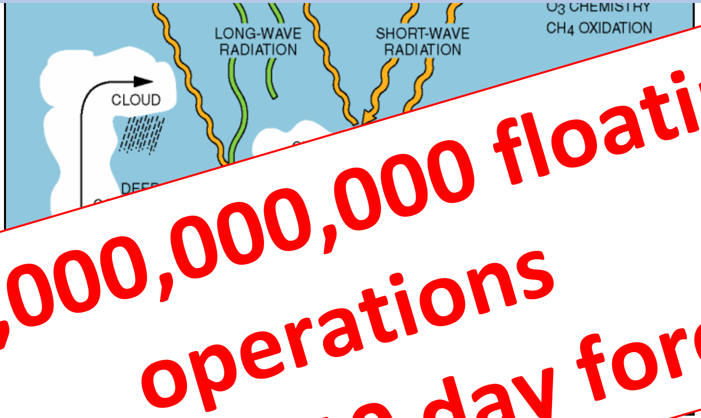


Global TCO1279 spectral resolution (9km grid point spacing)



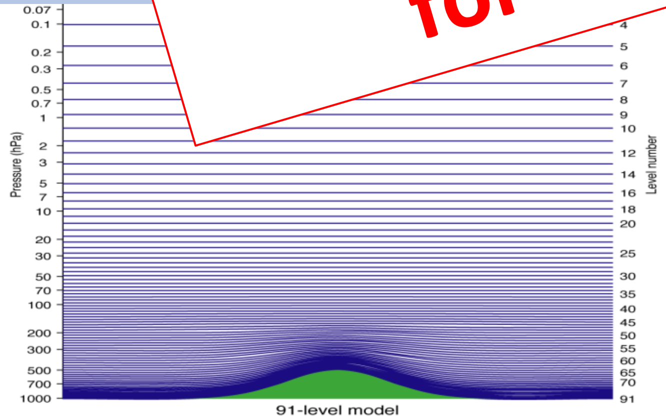
The forecast model

Physical and dynamical processes updated every 10 minutes

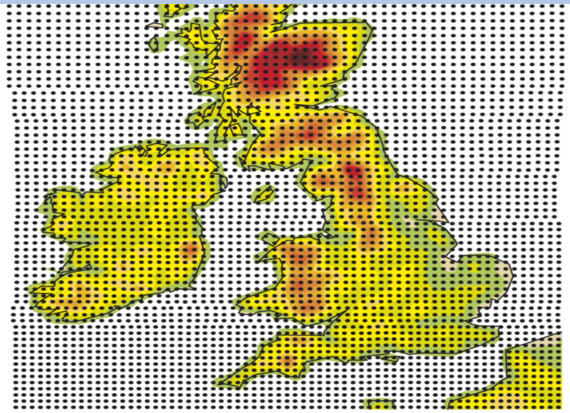


6,300,000,000,000,000 floating point operations for a single 10 day forecast

91 (**137**) v…
to 0.01hPa



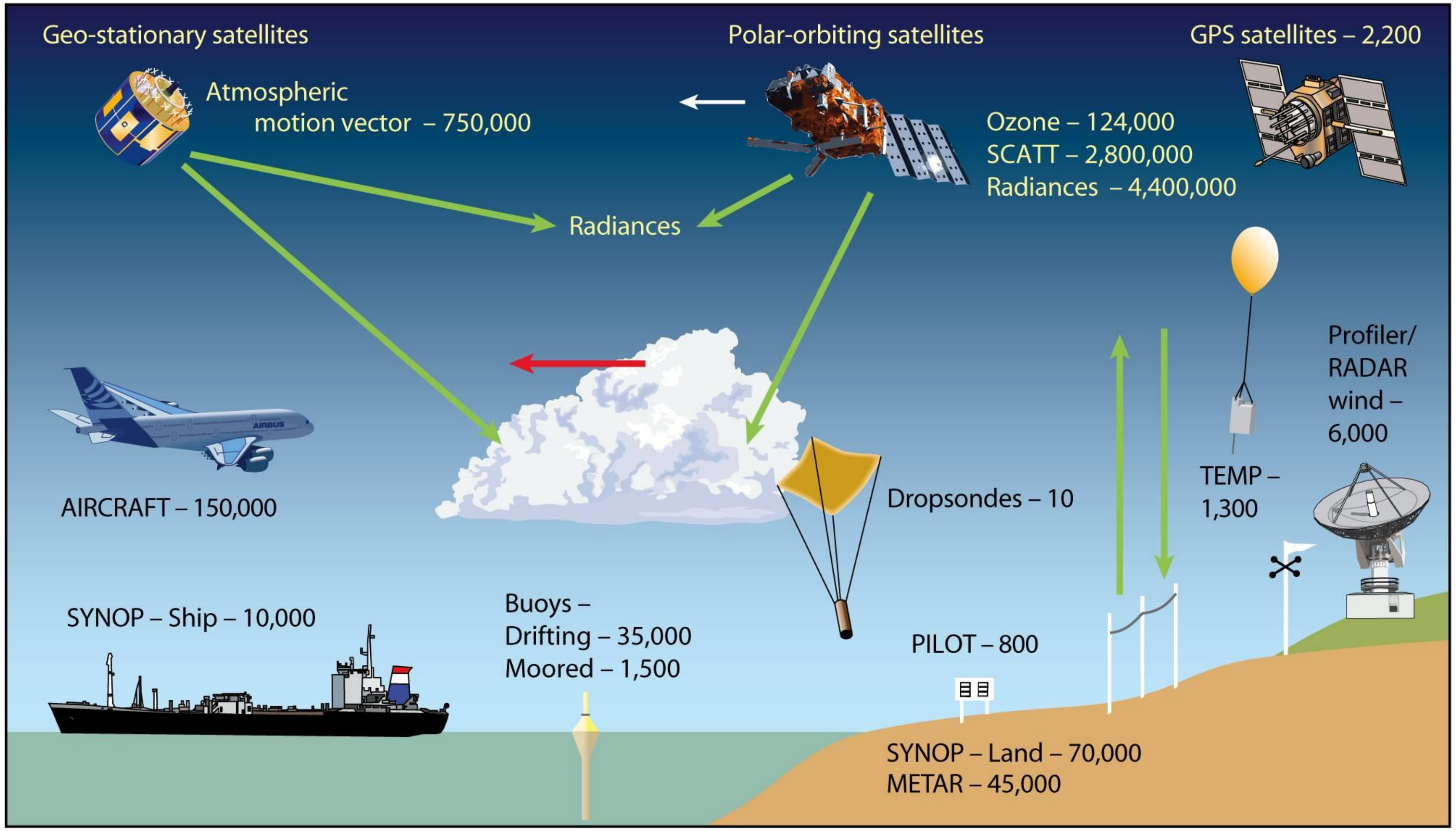
Global TCO1279 spectral resolution (9km grid point spacing)



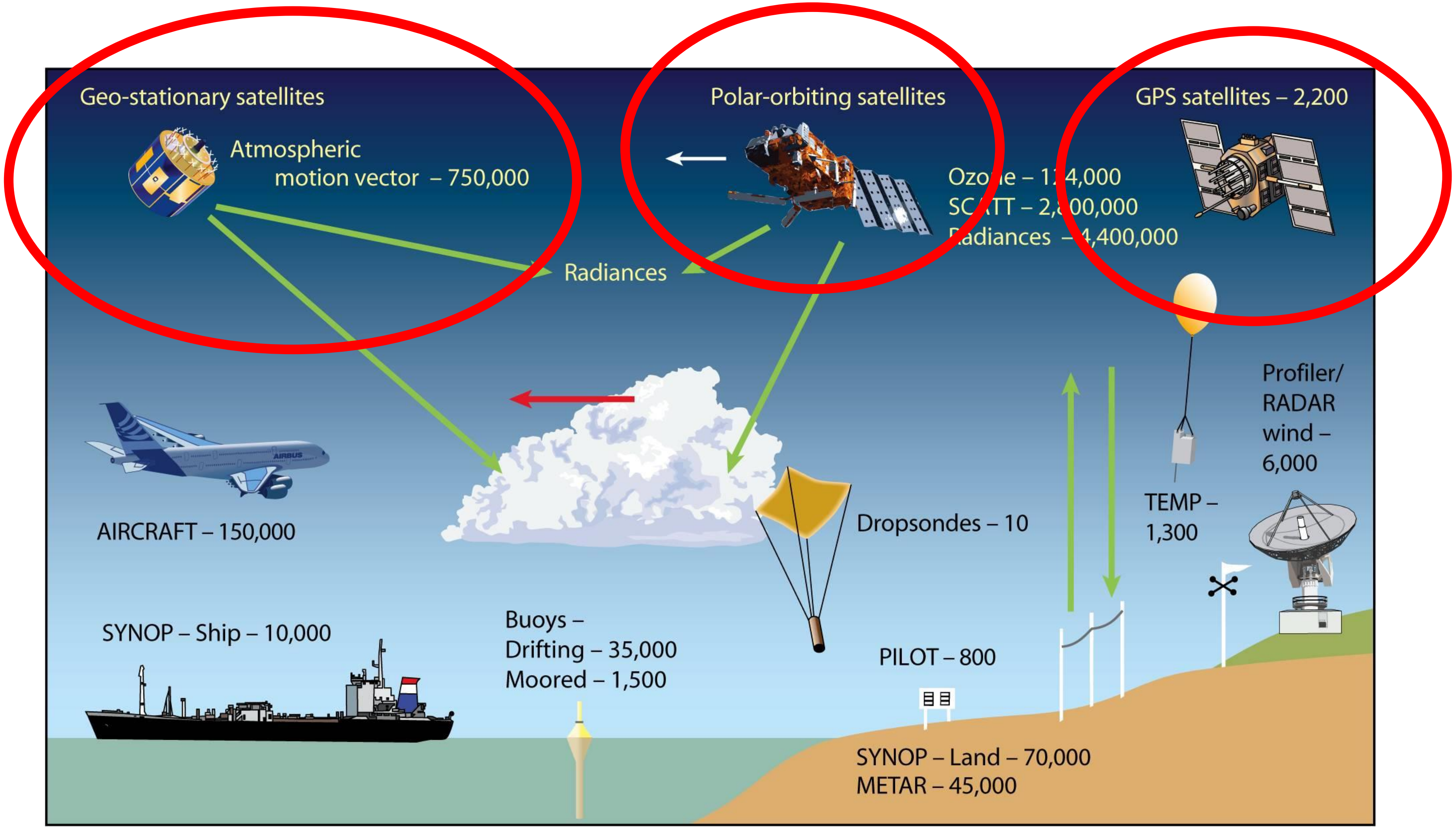
The Observations

$$\boxed{Y_{obs}}$$

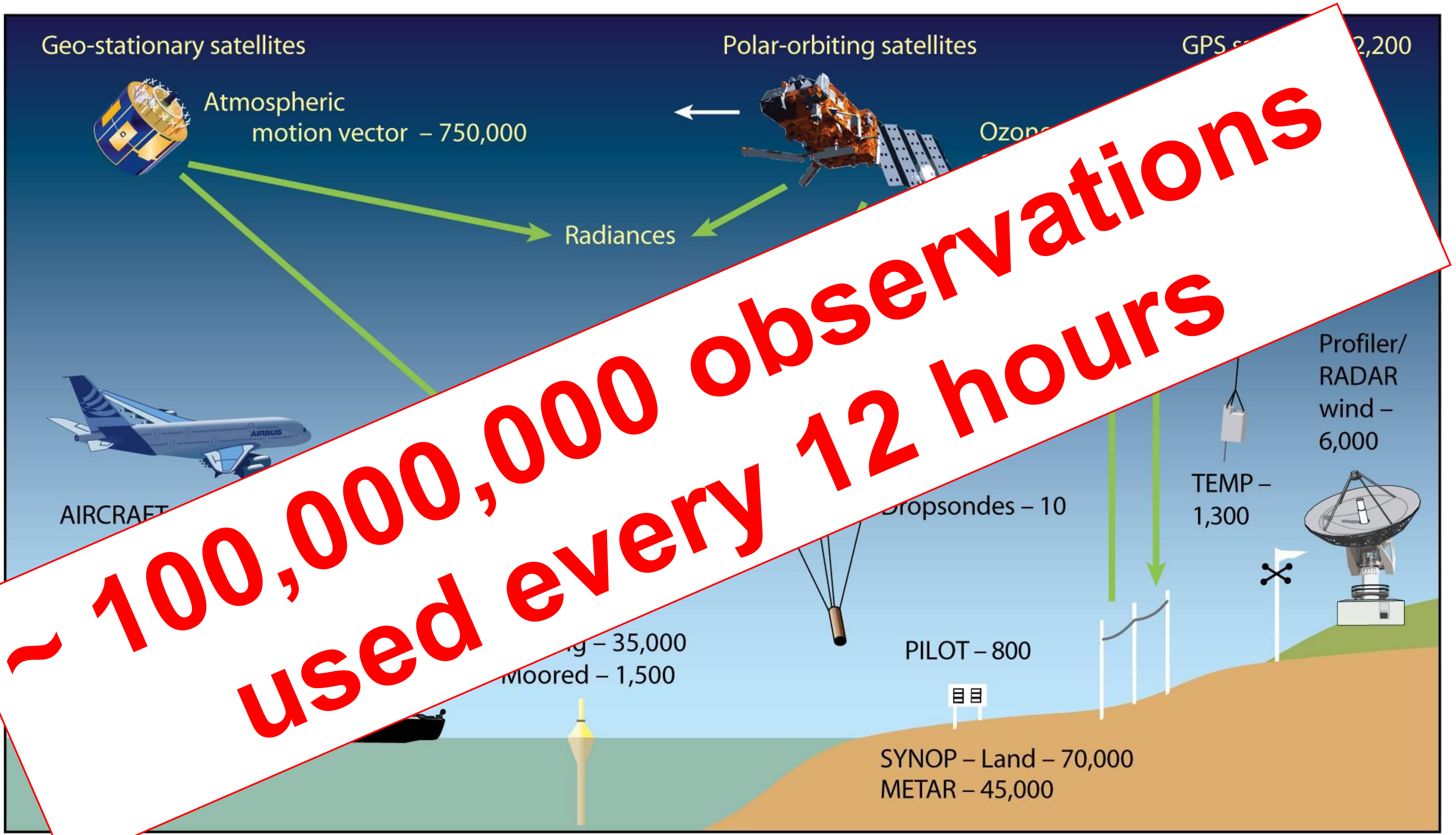
Operational Global Observing Network



Operational Global Observing Network



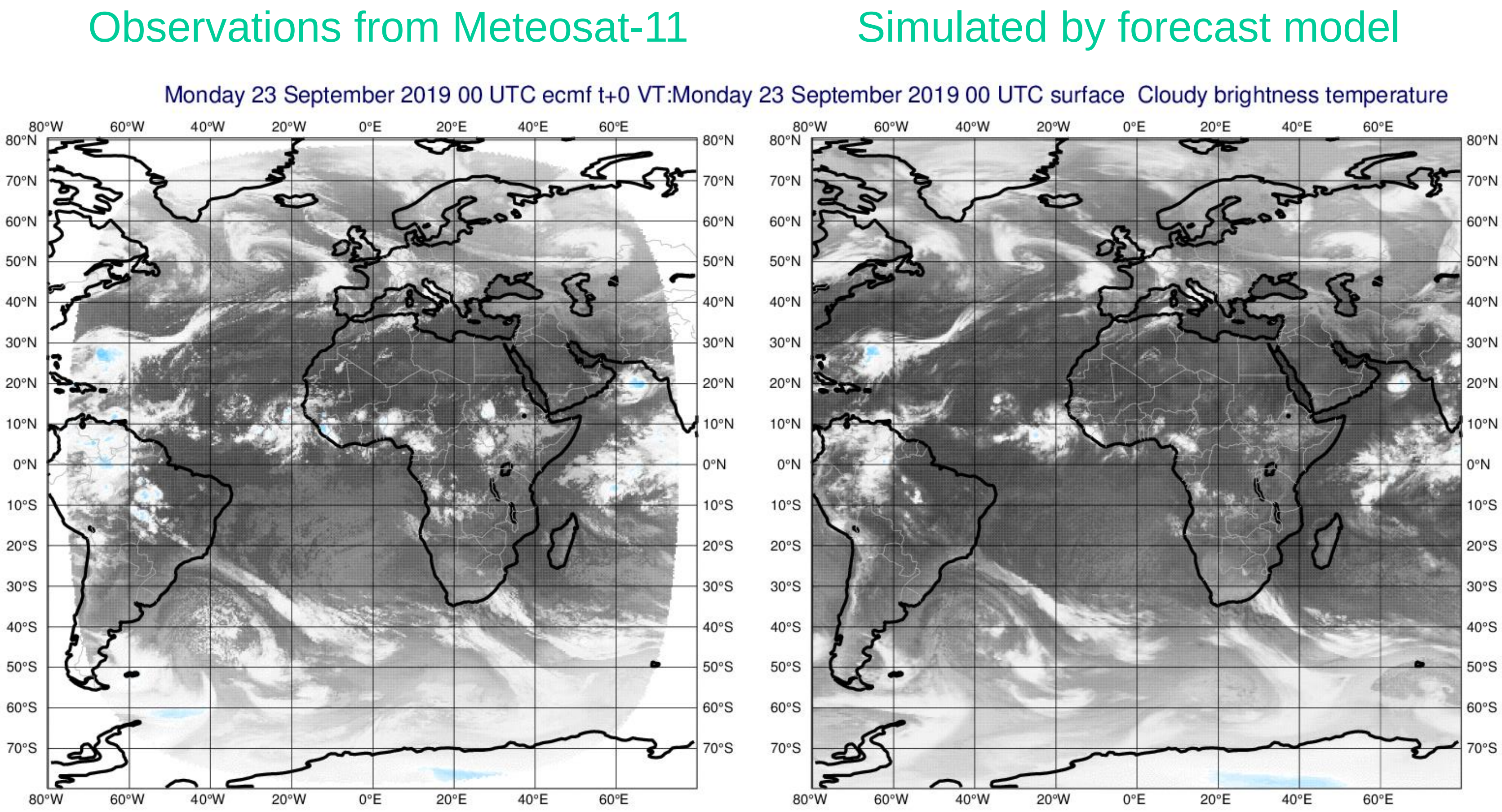
Operational Global Observing Network



~ 100,000,000 observations used every 12 hours

The assimilation algorithm

Comparing OBS with model in OBS space



Modern radiative transfer can simulate atmospheric radiation very accurately …so why do these diverge ?

Combining information

- At ECMWF we employ **variational data assimilation** methods

- These are based upon the **maximum likelihood combination** of observations and background information

- It can be shown that the most probable state of the atmosphere given a background $\mathbf{X_b}$ and some observations **Y** is that which minimises a **cost or penalty function J**

- The solution obtained is **optimal** in that it fits the prior (or background) information and measured radiances **respecting the uncertainty in both.**

The cost function *J (X)*

model state

background error covariance

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) +$$

$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

observations

observation* error covariance

observation operator (maps the model state to the observation space)

The cost function components (J_b)

$$J(x) = \boxed{(x - x_b)^T \mathbf{B}^{-1} (x - x_b)} +$$
$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

Fit of the solution to the background estimate of the atmospheric state weighted inversely by the background error covariance **B**

The cost function components (J_o)

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) +$$
$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

Fit of the solution to the observations weighted inversely by the measurement error covariance **R** (observation error + error in observation operator **H**)

…a helpful linear analogue …

It can be shown that the state that minimizes the cost function is equivalent to a linear **correction** of the background using the observations:

$$x_a = x_b + [\mathbf{HB}]^T [\mathbf{HBH}^T + \mathbf{R}]^{-1} (y - \mathbf{H} x_b)$$

correction term

…and the **improvement** can be quantified in terms of the key parameters of the assimilation…(i.e. **B**, **R**, **H**)

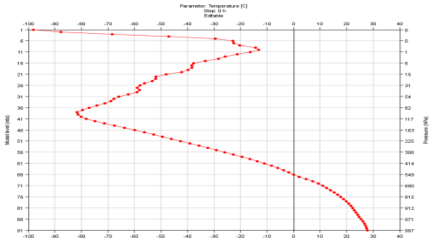
$$S_a = B - [\mathbf{HB}]^T [\mathbf{HBH}^T + \mathbf{R}]^{-1} HB$$

improvement term

Various implementations of the assimilation algorithm

- 1D-Var
- 3D-Var
- 4D-Var

One dimensional variational analysis (1D-Var)



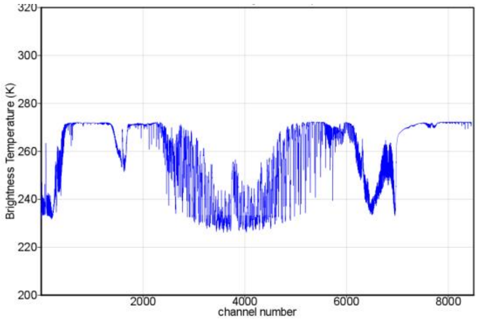
1D model state profile

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) +$$

$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

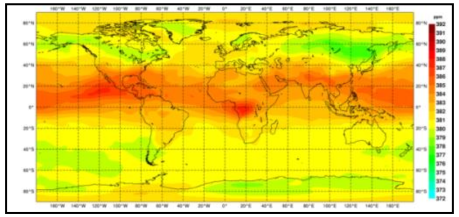
vector of measured radiances at one location

observation Operator
= radiative transfer model



$$L(\nu) = \int_0^\infty B(\nu, T(z)) \left[\frac{d\tau(\nu)}{dz} \right] dz$$

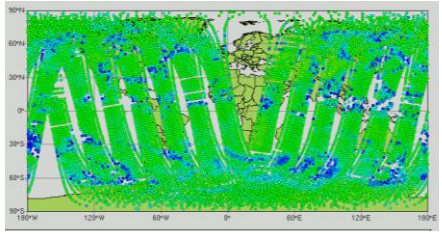
Three dimensional variational analysis (3D-Var)



3D model state

$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) + (y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

global vector of measured radiances

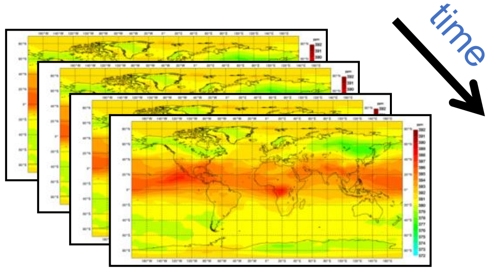


observation operator
= spatial interpolation +
radiative transfer model

$$L(\nu) = \int_0^{\infty} B(\nu, T(z)) \left[\frac{d\tau(\nu)}{dz} \right] dz$$

Four dimensional variational analysis (4D-Var)

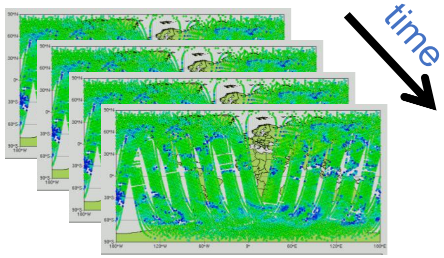
4D model state



$$J(x) = (x - x_b)^T \mathbf{B}^{-1} (x - x_b) +$$

$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$

global time windows of measured radiances



observation operator
= spatial interpolation + forecast model
radiative transfer model

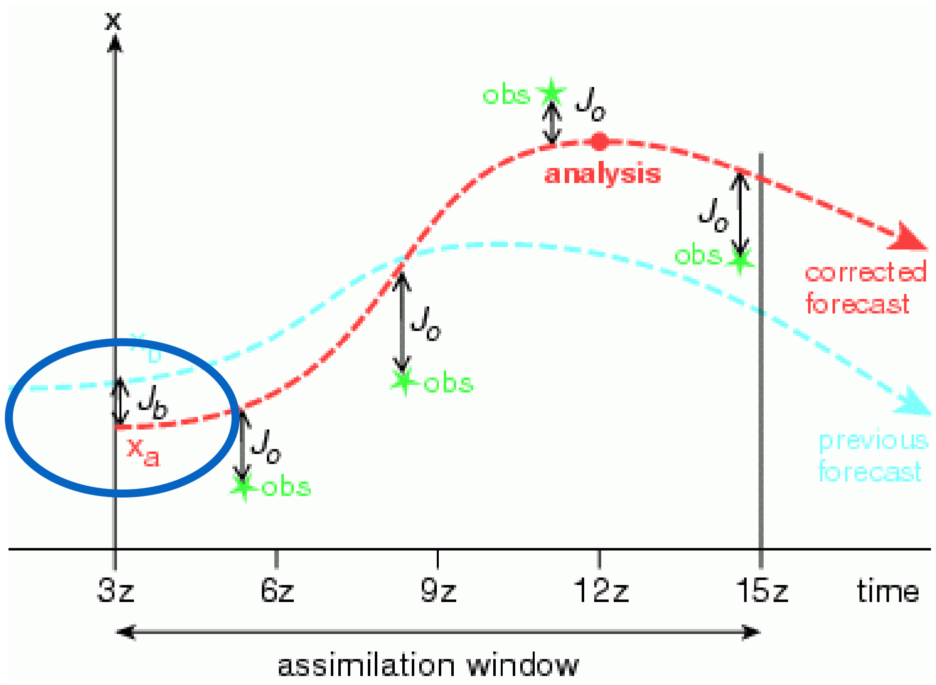
$$L(\nu) = \int_0^{\infty} B(\nu, T(z)) \left[\frac{d\tau(\nu)}{dz} \right] dz$$

$X_{t=0}$ → $X_{t=t}$

The 4D-Var Algorithm J_b

$$J(x) = \boxed{(x - x_b)^T \mathbf{B}^{-1} (x - x_b)} +$$

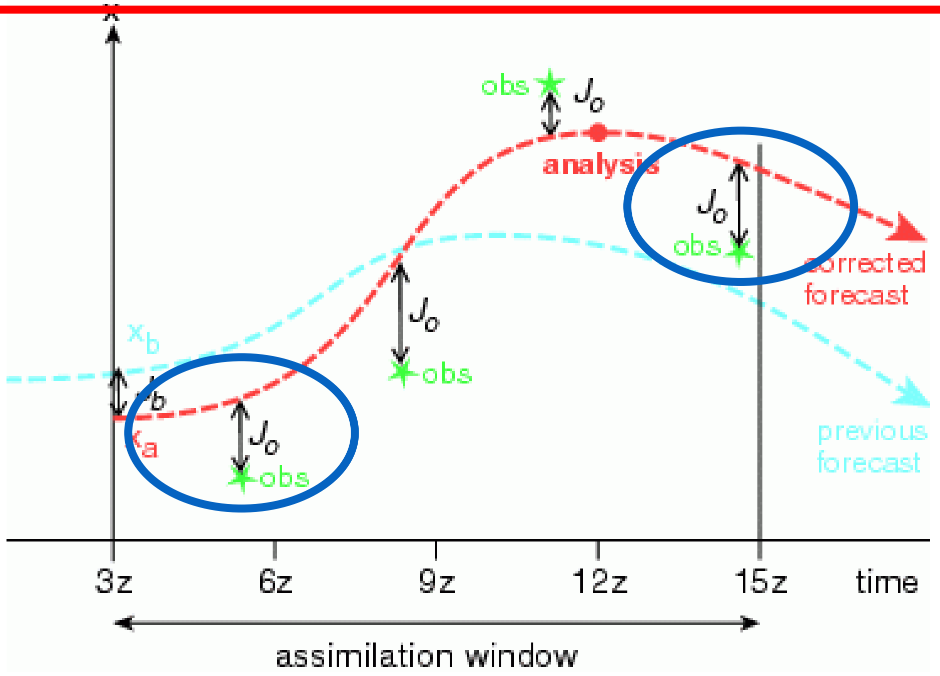
$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1} (y - \mathrm{H}[x])$$



The 4D-Var Algorithm J_o

$$J(x) = (x - x_b)^T \mathbf{B}^{-1}(x - x_b) +$$

$$(y - \mathrm{H}[x])^T \mathbf{R}^{-1}(y - \mathrm{H}[x])$$



The key elements of a satellite data assimilation system

Key elements of a data assimilation system

- **observation operator**
- **background errors**
- **observation errors**
- **bias correction**
- **data selection and quality control**

Key elements of a data assimilation system

- **observation operator**
- background errors
- observation errors
- bias correction
- data selection and quality control

Observation operator

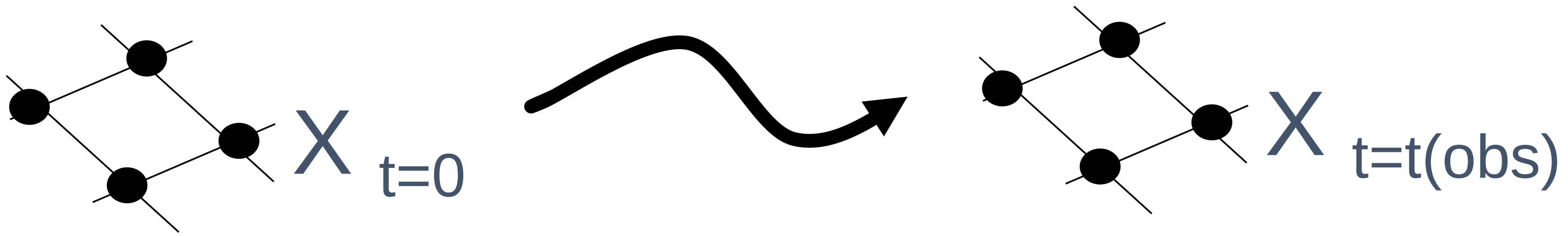
- The observation operator must map the model state at beginning of the assimilation window (t=0) to the observation time and location.

- In the **direct assimilation of radiance observations**, the observation operator must incorporate an additional step to compute radiances from the model state variables (radiative transfer model RTTOV).

- This means that radiance observations are significantly more computationally expensive than conventional observations (e.g. radiosonde temperature data)

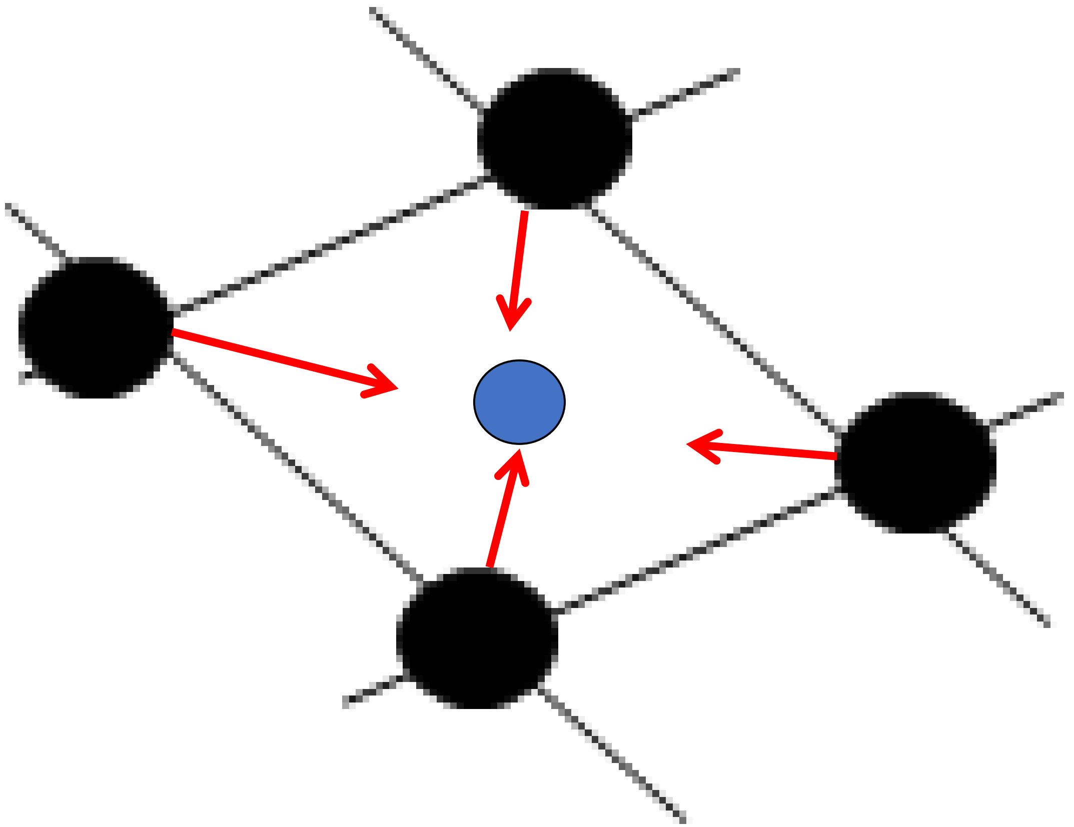
Observation operator

1) Time evolution of forecast model field to OBS time



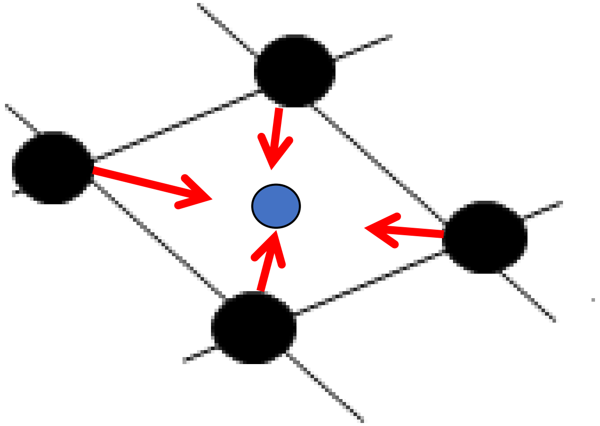
Observation operator

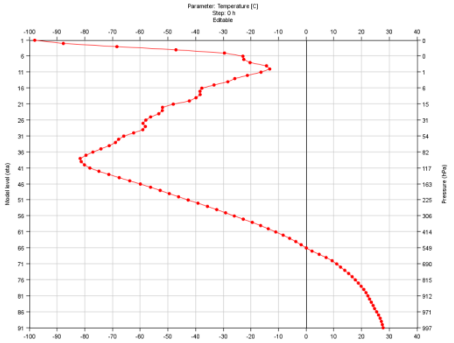
2) Spatial interpolation of model grid to OBS location



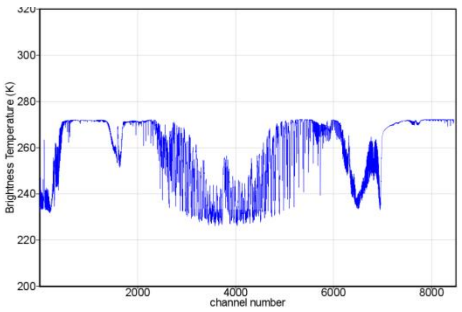
Observation operator

3) Radiative transfer calculation from model state at that location to radiances at that location





RTTOV



Observation operator (RT component)

- The RT model should produce an accurate **simulation of the satellite radiance from the model state**, based upon the best knowledge of the instrument characteristics and up to date spectroscopic information.

- However, the model must be **fast enough** to process huge quantities of data in near real time (thus line-by-line models are not suitable)

- In addition, the **adjoint and tangent linear** versions of the RT model are required by the algorithm that minimises the cost function

- Ideally the **same** RT model should be used for **all satellite** sensors being assimilated

Key elements of a data assimilation system

- observation operator
- **background errors**
- observation errors
- bias correction
- data selection and quality control

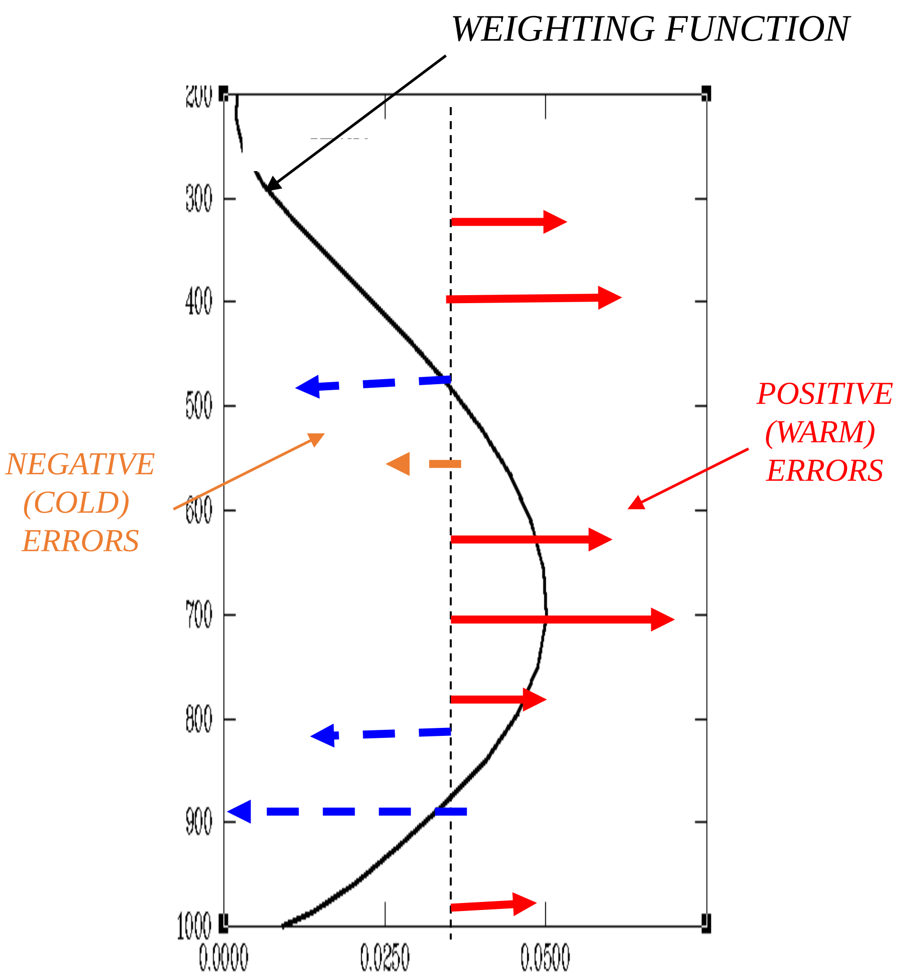
Background errors (and vertical resolution)

- The matrix B must accurately describe errors in the background estimate of the atmospheric state. It determines the weight given to the background information.

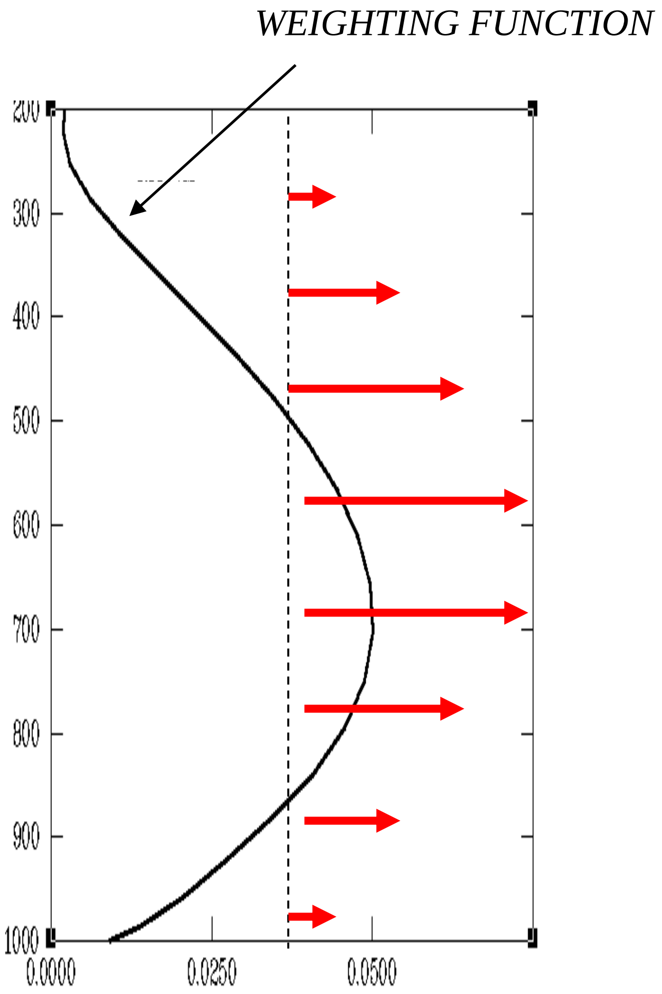
- A very important aspect for the assimilation of near-nadir viewing satellite radiances are the **vertical correlations** that describe how background errors are distributed in the vertical (sometimes called structure functions)

- These are important because satellite radiances have very **limited vertical resolution** (previous lecture)

Background errors (and vertical resolution)



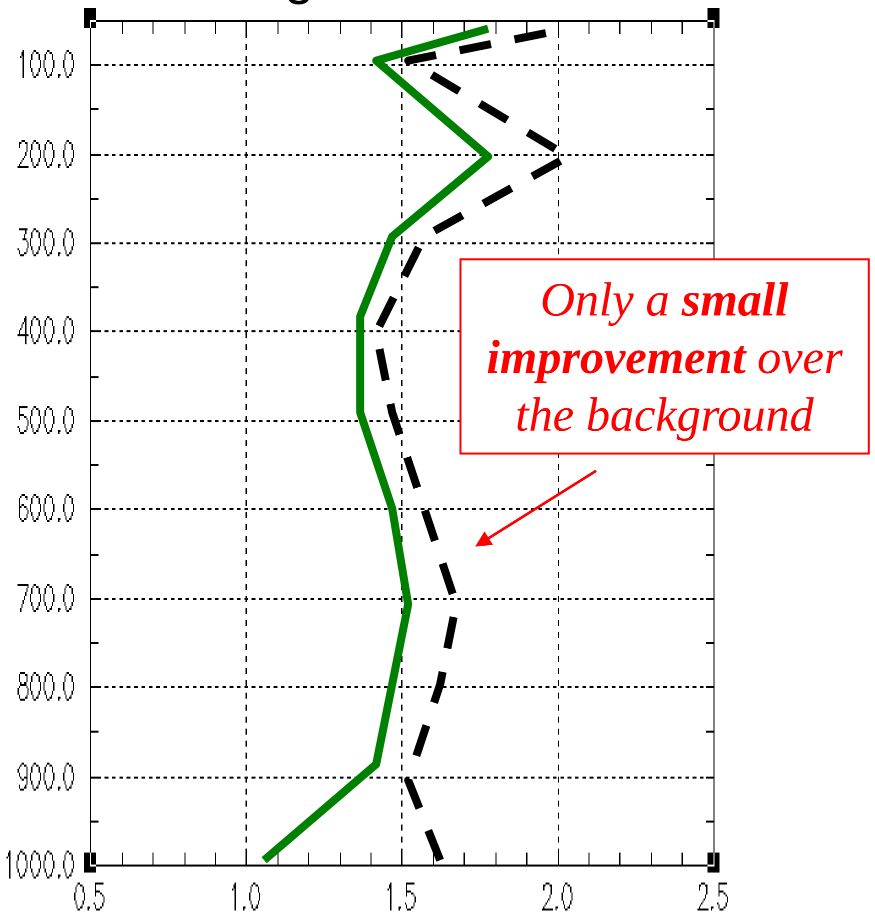
"Difficult" to correct



"Easy" to correct

Background errors (and vertical resolution)

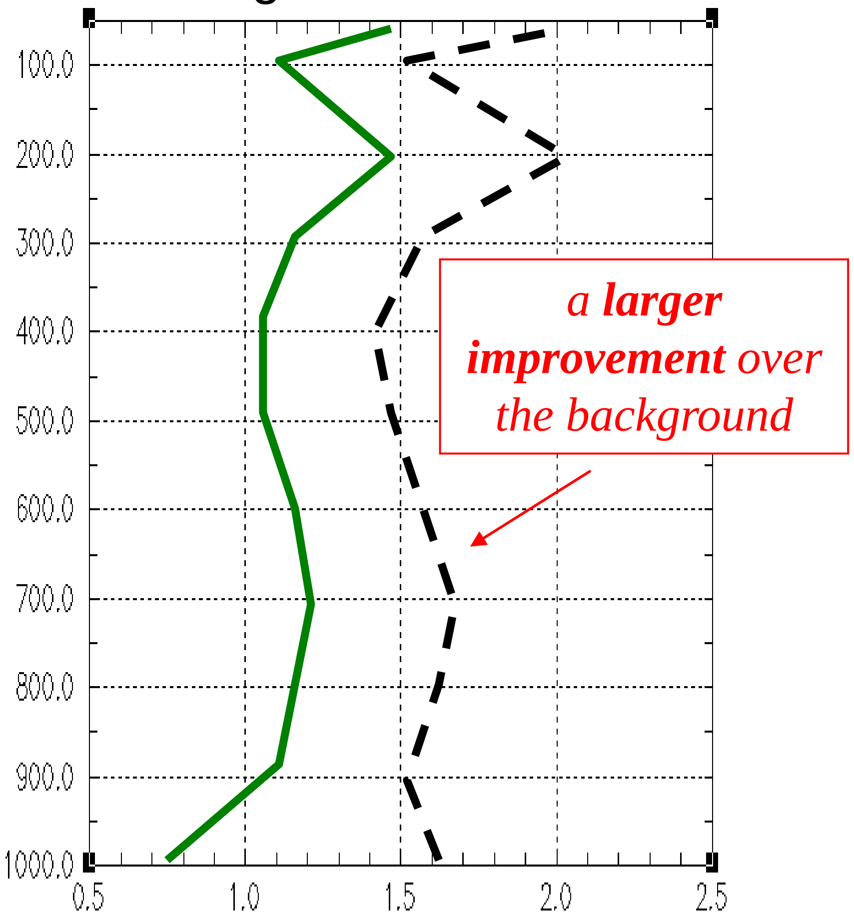
Sharp / anti-correlated background errors



Error standard deviation (K)

S_a

Broad / deep correlated background error

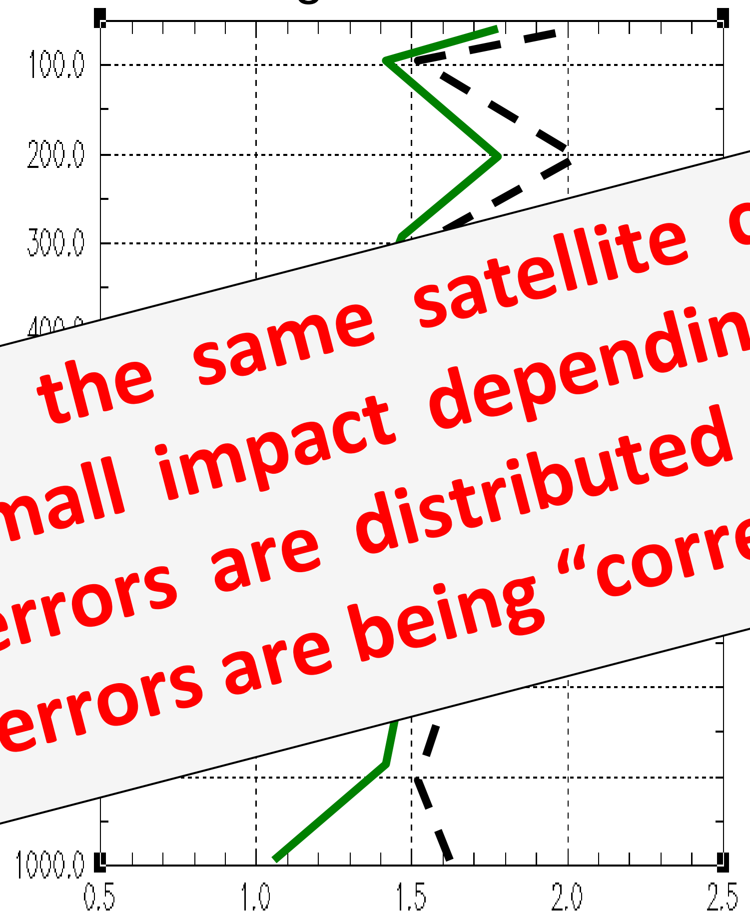


Error standard deviation (K)

S_a

Background errors (and vertical resolution)

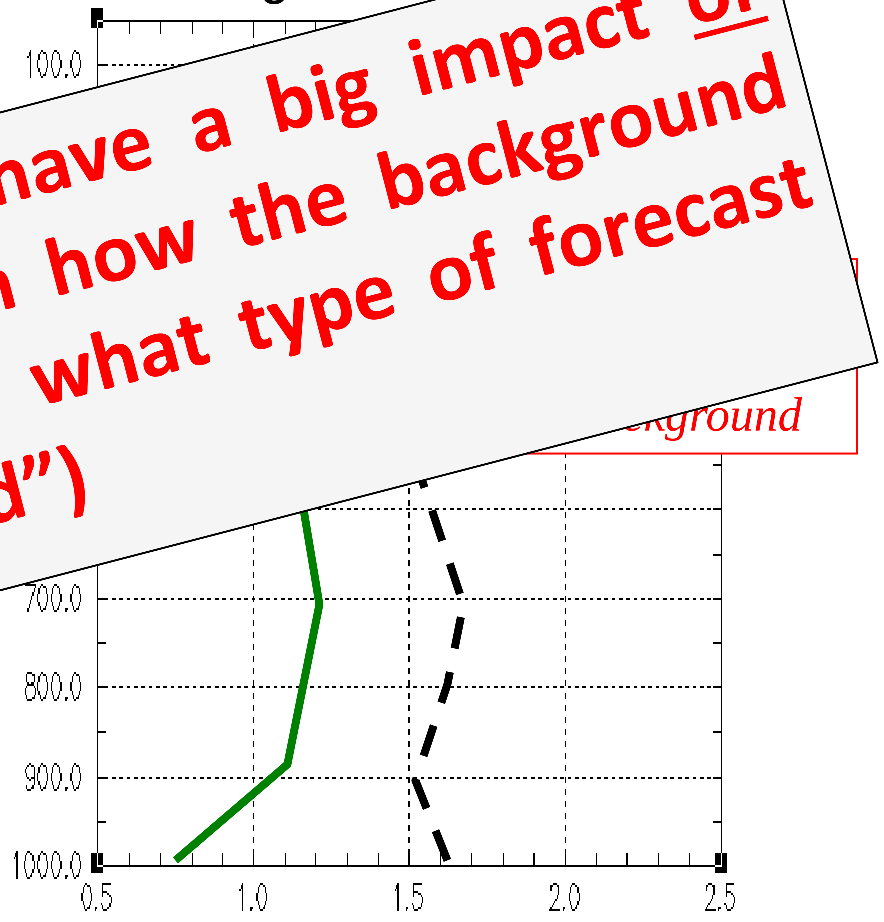
Sharp / anti-correlated background errors



Error standard deviation (K)

S_a

Broad / deep correlated background errors



Error standard deviation (K)

S_a

So the same satellite can have a big impact <u>or</u> small impact depending on how the background errors are distributed (i.e. what type of forecast errors are being "corrected")

..lecture later this week on background errors…

Key elements of a data assimilation system

- observation operator
- background errors
- **observation errors**
- bias correction
- data selection and quality control

Observation errors:

- These determine the weight we give to the radiance observations. The observation error must account for random uncertainties in the observation operator (e.g. RT model), errors in data screening (e.g. residual clouds) and errors of representativeness (e.g. scale mismatch).

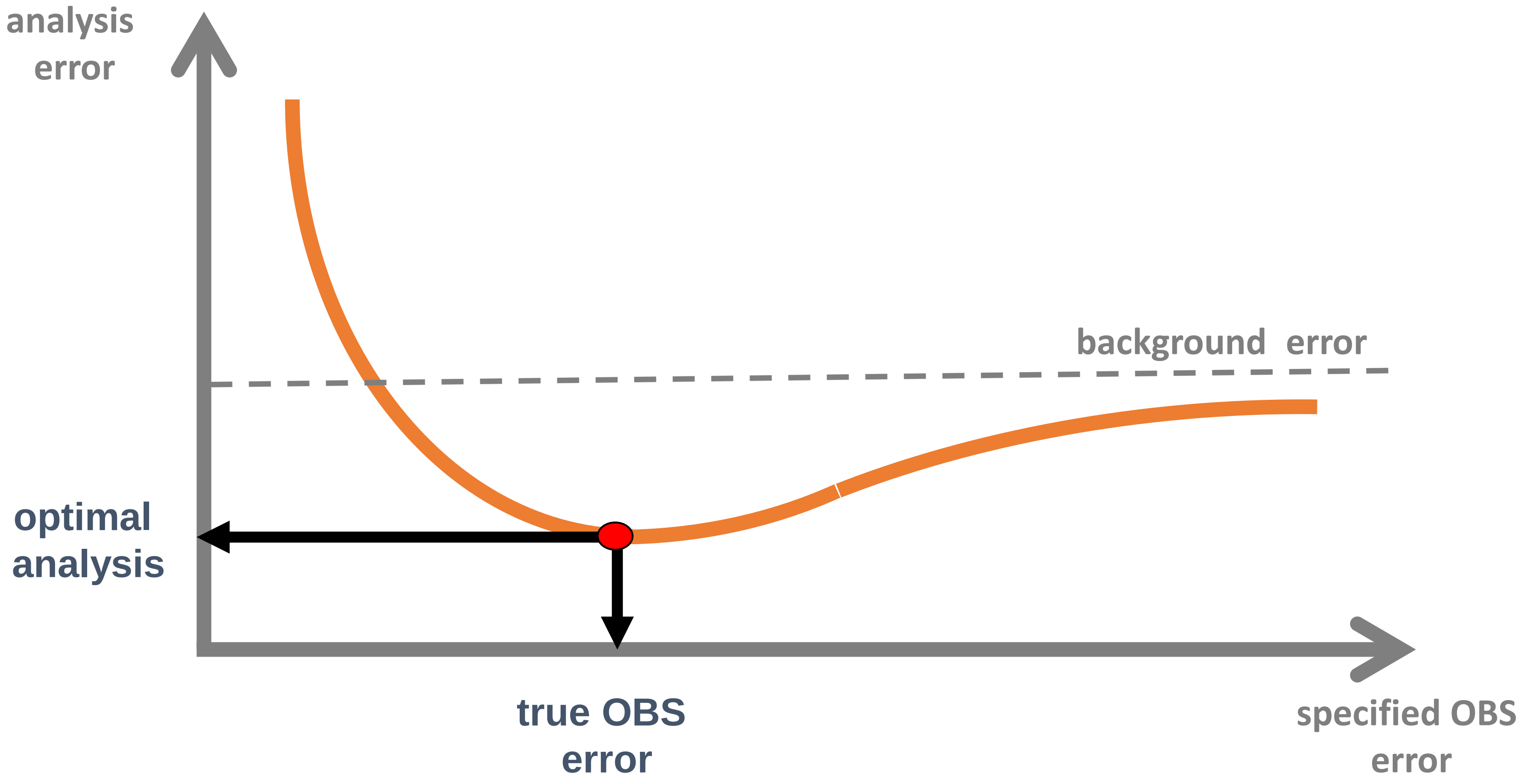
- It is important to model both the magnitude of errors (diagonals of R) and any inter-channel correlations

- Lecture this week by Niels Bormann!

- Wrongly specified observation errors can lead to an analysis with **larger errors than the background**!

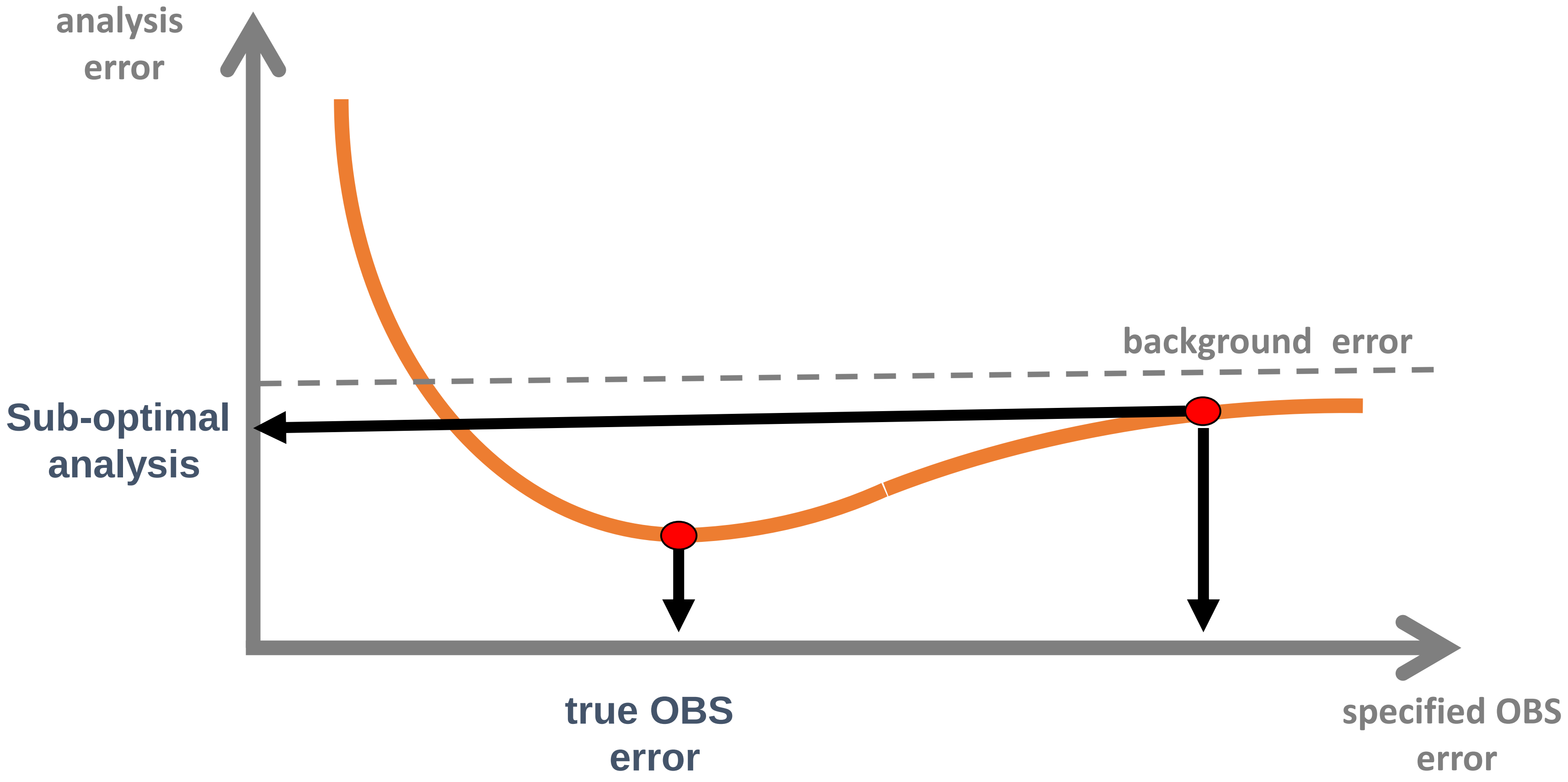
Observation errors:

- Specifying the correct observation error produces an optimal analysis with minimum error.



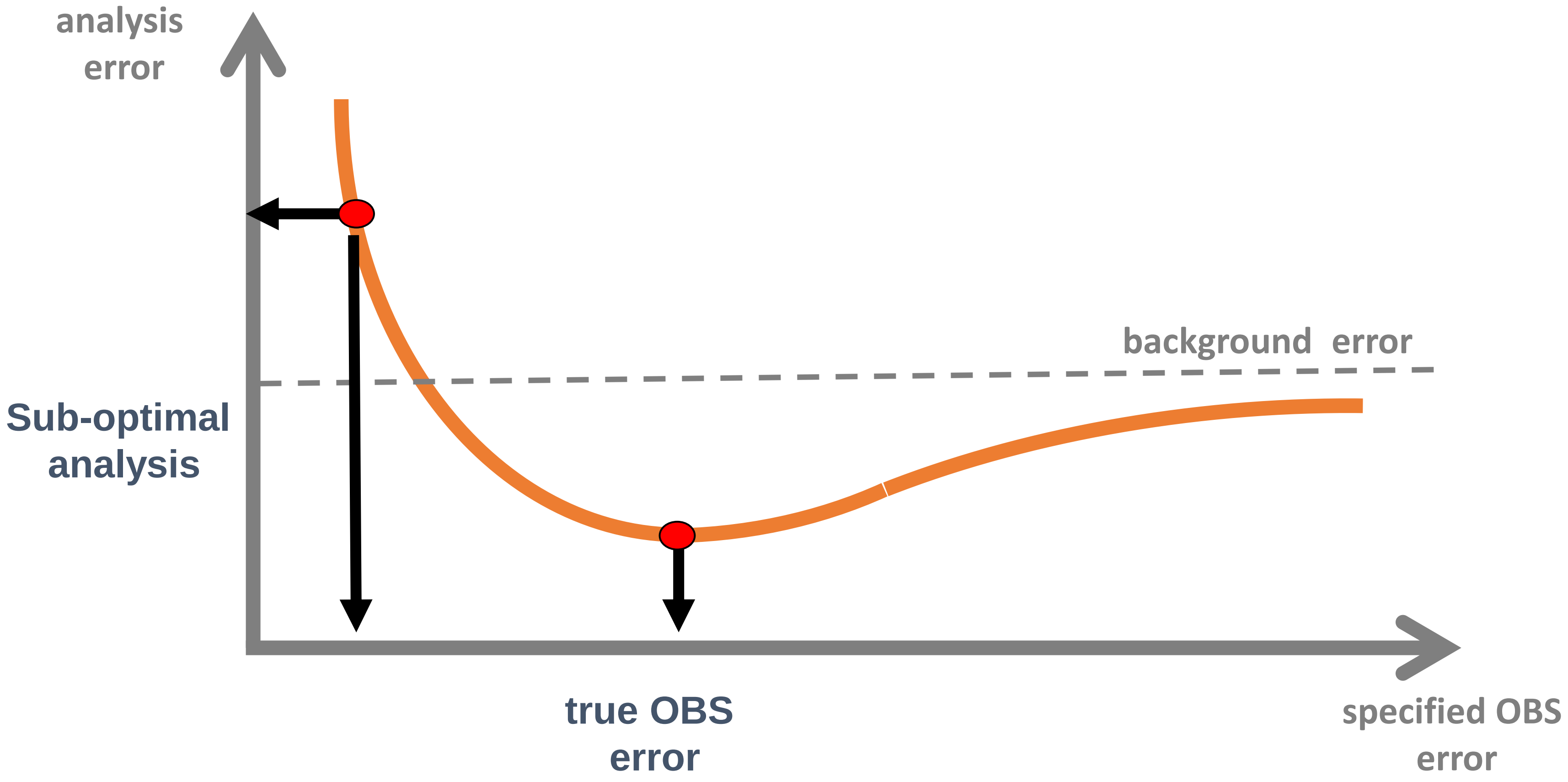
Observation errors:

- Over-estimating the OBS error degrades the analysis, but the result will not be worse than the background.



Observation errors:

- Under-estimating the OBS error degrades the analysis, and **the result can be worse than the background**!



...lecture later this week on observation errors…

Key elements of a data assimilation system

- observation operator
- background errors
- observation errors
- **bias correction**
- data selection and quality control

Bias correction:

Systematic errors must be removed otherwise biases will propagate in to the analysis (causing **global damage** in the case of satellites!). A bias in the radiances is defined as:

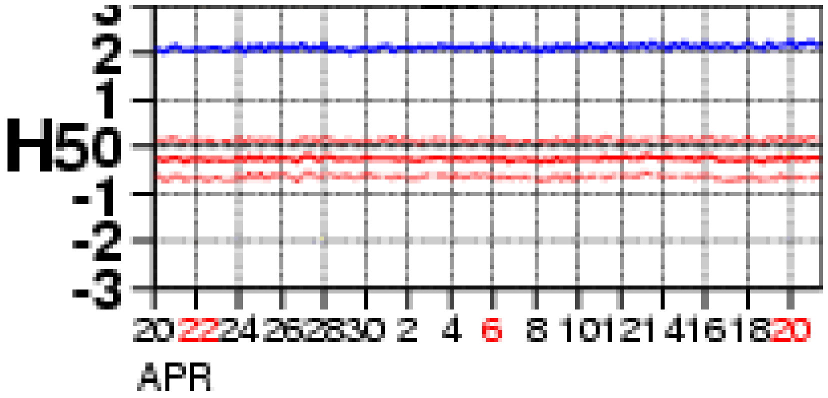
$$\textit{bias} = \textit{mean}\ [\ Y_{obs} - H(X_{true})\ ]$$

Sources of systematic error in radiance assimilation include:

- instrument error (scanning or calibration)
- radiative transfer error (spectroscopy or RT model)
- cloud / rain / aerosol screening errors

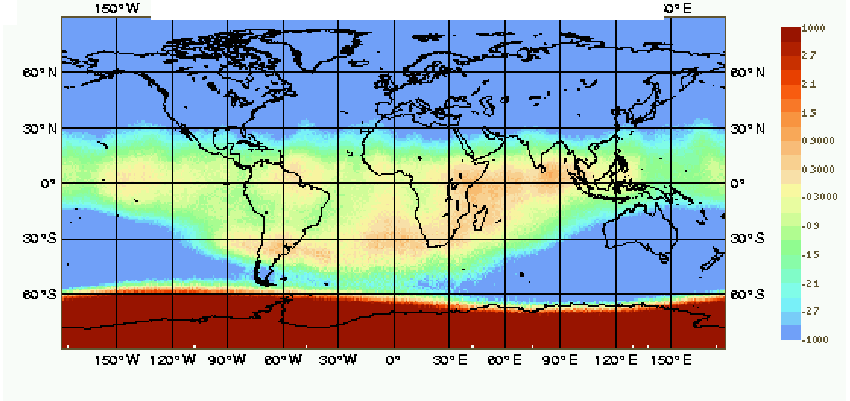
Bias correction:

HIRS channel 5



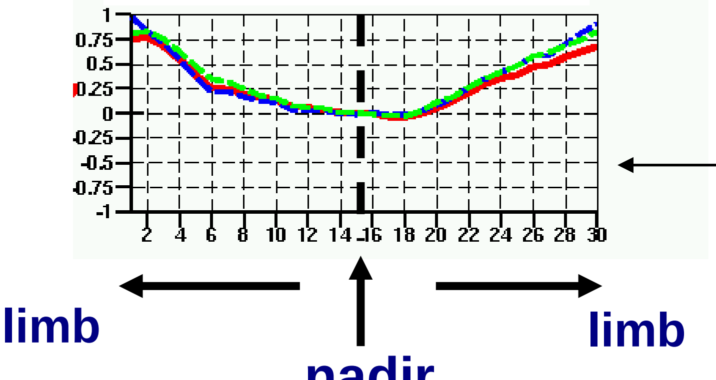
simple flat offset biases that are constant in time

AMSU-A channel 14



biases that vary depending on location or air-mass

AMSU-A channel 7



biases that vary depending on the Scan position of the satellite instrument

Bias correction:

But sometimes **NWP systematic errors** can make it difficult to diagnose and correct observation biases

What we would like to quantify is:

$$\mathit{Bias} = \mathit{mean}\,[\,Y_{obs} - H(X_{true})\,]$$

But in practice all we can monitor is :

$$\mathit{Bias} = \mathit{mean}\,[\,Y_{obs} - H(X_{b/a})\,]$$

..lecture later this week on systematic errors…

Key elements of a data assimilation system

- observation operator
- background errors
- observation errors
- bias correction
- **data selection and quality control**

Data selection and quality control (QC):

The primary purpose of this is to ensure that the observations entering the analysis are consistent with the assumptions in the observations error covariance (**R**) and the observation operator (**H**).

Primary examples include the following:

- Rejecting bad data with **gross error** (not described by **R**)
- Rejecting data affected by **clouds** if **H** is a clear sky RT
- Thinning data if no **correlation** is assumed (in **R**)
- Always **blacklisting** data where we do not trust our QC!

Data selection and quality control (QC):

Often checks are performed using the forecast background as a reference. That is an observations is rejected if the departure from the background exceeds a threshold T_{QC}:

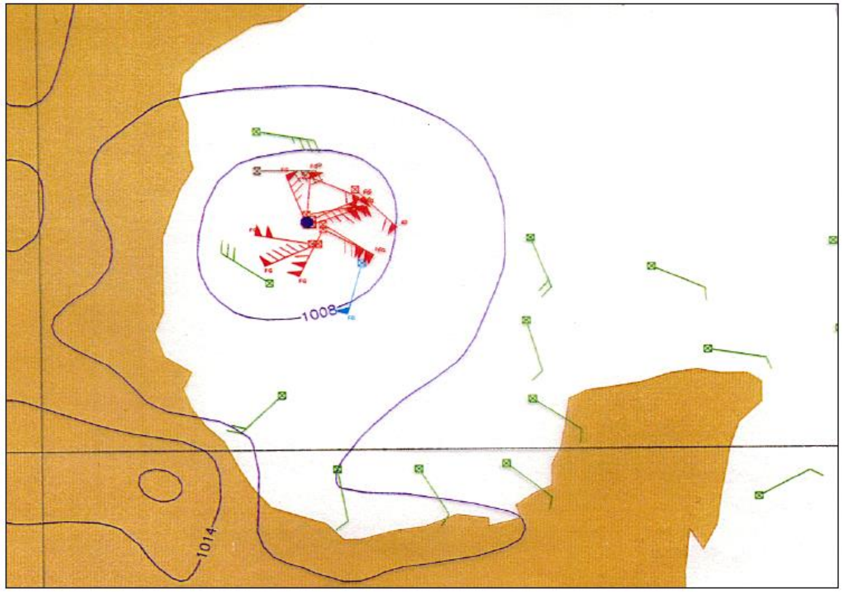
$$Y_{obs} - H(X_{true}) > T_{QC}$$

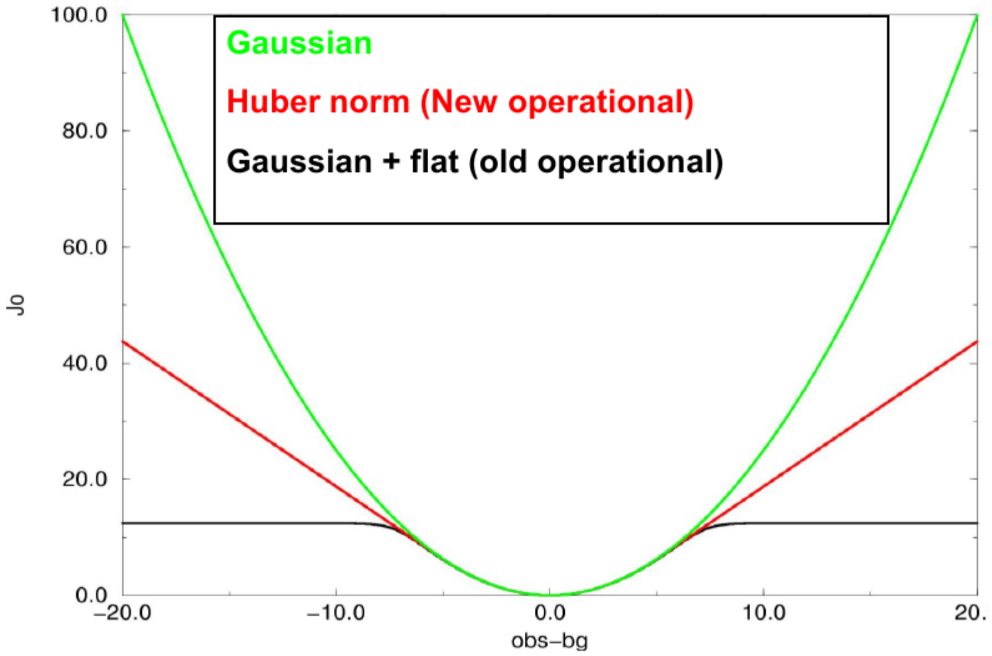
But sometimes large errors in the background can lead to:

- False rejection of a good observation
- Missed rejection of a bad observation

Data selection and quality control:

- False rejection of a **good** observation



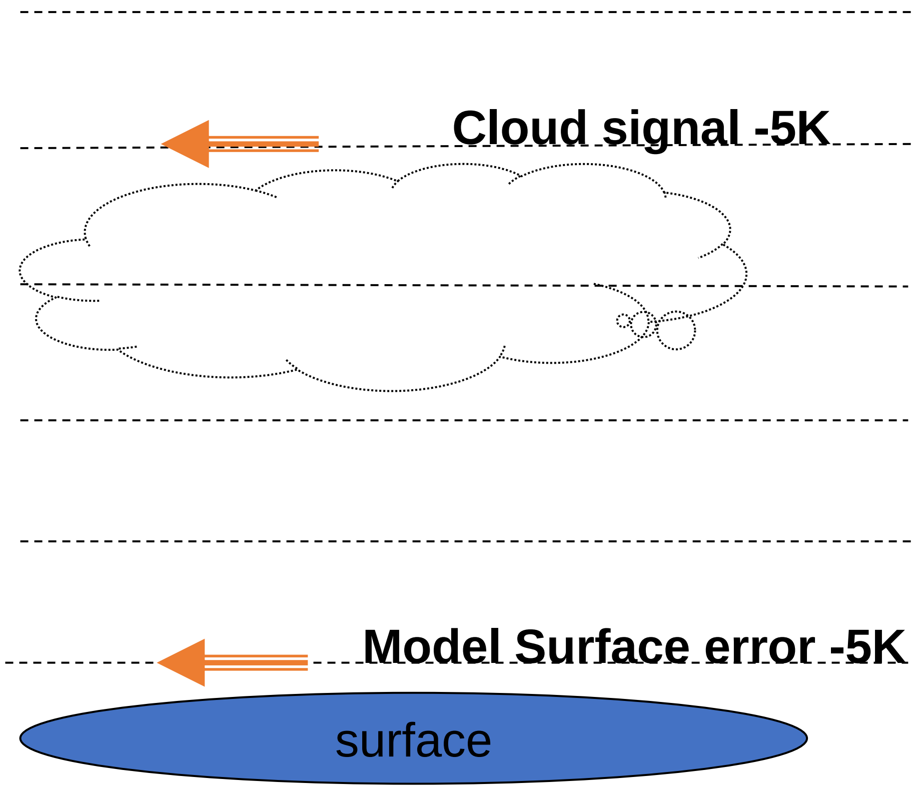


The **numerous** failing observations are good, but a bad background is causing them to be rejected. We **need** these observations to improve the analysis !

Instead of rejecting, we give the observations a lower weight so **collectively** they can influence and improve the analysis. In this framework a single bad observation would do no damage.

Data selection and quality control:

- Missed rejection of a **bad** observation



The radiance are contaminated by cloud (**cold 5K**) compared to the clear sky value.

But our computation of the clear sky value from the background is also **cold by 5K** due to an error in the surface skin temperature.

Thus our checking (against the background) sees no reason to reject the observation and is it **passed**!

Summary

- observation operator
  (complex and expensive for radiances)
- background errors
  (important due to limited vertical resolution)
- observation errors
  (must be specified correctly)
- bias correction
  (global impact of small bias)
- data selection and quality control
  (false alarms and missed rejections)

Questions ?