Aspects of using observations in Data Assimilation

Lars Isaksen

DA Section, ECWMF

Acknowledgements to: Angela Benedetti, Antje Inness, Bruce Ingleby, David Tan, Erik Andersson, Mike Fisher and Tony McNally

Overview of Lecture

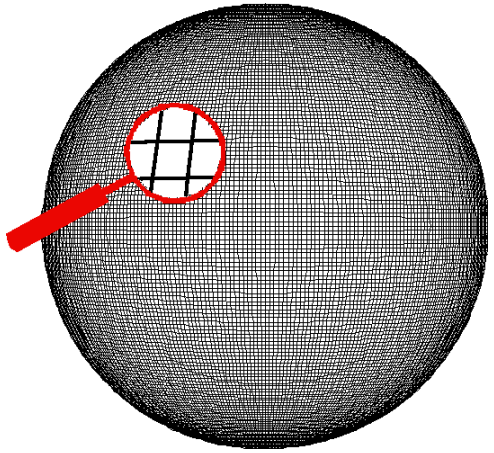
- Observations are essential for data assimilation
- Introduction to observation operators
- Different flavours of observation operators
- Jacobians (linearized operators)
- Why variational data assimilation is very flexible with respect to observation usage
- Summary

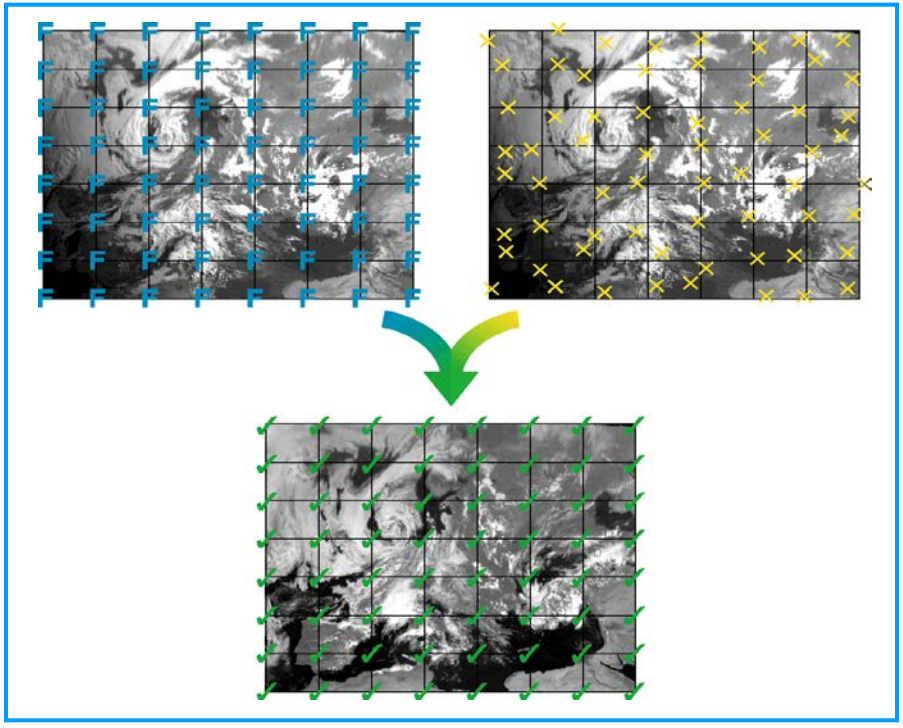
Data assimilation for weather prediction

The FORECAST is computed on a quasi-regular grid over the globe.

The meteorological OBSERVATIONS comes at any time from any location on the globe.

The computer model's prediction of the atmosphere is compared against the available observations, in near real time



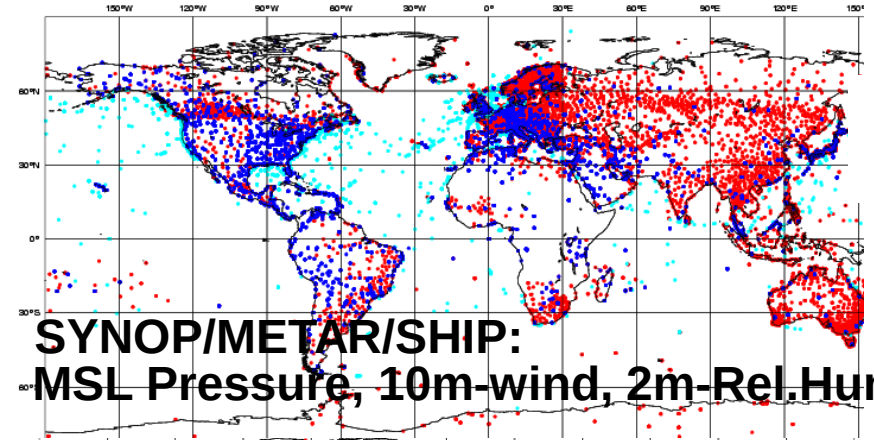


A short-range forecast provides an estimate of the atmosphere that is compared with the observations.

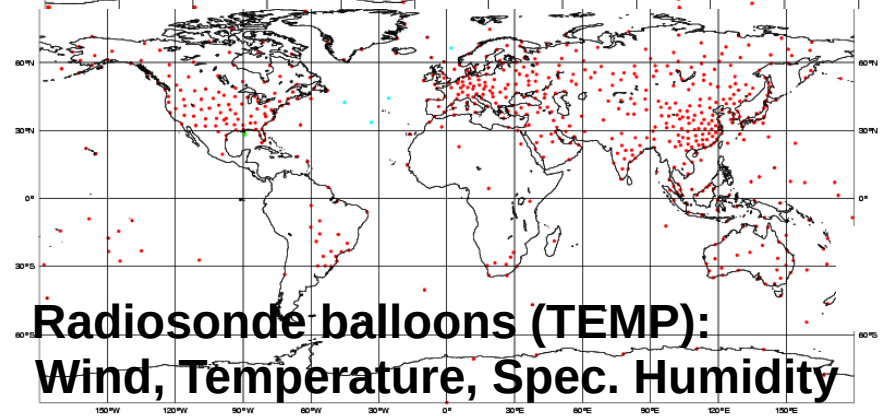
The two kinds of information are combined to form a corrected atmospheric state: the analysis.

Corrections are computed and applied twice per day at ECMWF. This process is called 'Data Assimilation'.

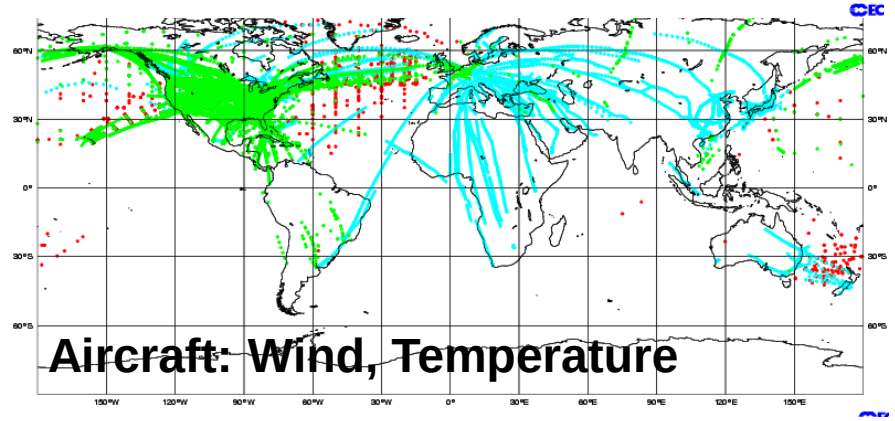
Conventional observations used by ECMWF's analysis



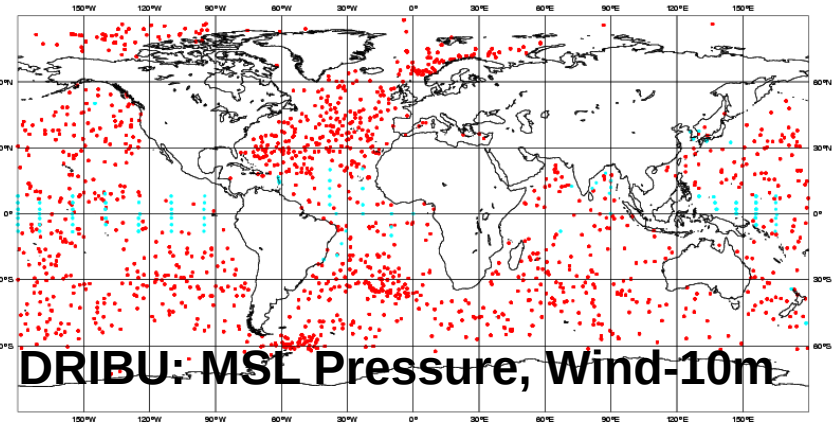
SYNOP/METAR/SHIP:
MSL Pressure, 10m-wind, 2m-Rel.Hum.



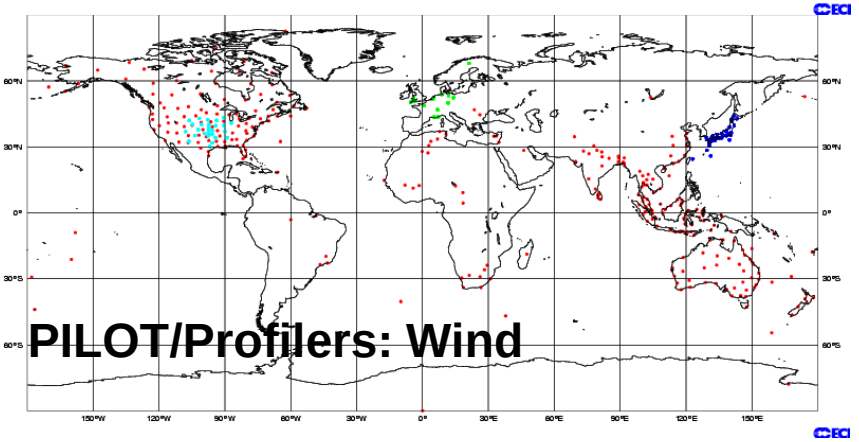
Radiosonde balloons (TEMP):
Wind, Temperature, Spec. Humidity



Aircraft: Wind, Temperature



DRIBU: MSL Pressure, Wind-10m



PILOT/Profilers: Wind

Note: Data assimilation only use a limited number of the observed variables - especially over land.

Quality control of observations is very important

Data extraction

- Check out duplicate reports
- Ship tracks check
- Hydrostatic check

Thinning

- Some data is not used to avoid over-sampling and correlated errors
- Departures and flags are still calculated for further assessment

Blacklisting

- Data skipped due to systematic bad performance or due to different considerations (e.g. data being assessed in passive mode)
- Departures and flags available for all data for further assessment

Model/4D-Var dependent QC

- First guess based rejections
- VarQC rejections

Used data → **Increments**

Analysis

More on observation QC in tomorrow's lecture by Elias Holm

From Mike Fisher's lecture this morning:

- We learned the basic concepts of data assimilation
- We learned the definition of the observation operator, H

I will now repeat a few of Mike's slides

Extension to Multiple Dimensions

- Now, let's turn our attention to the multi-dimensional case.
- Instead of a scalar prior estimate T_b, we now consider a vector $\mathbf{x}_b$.
- We can think of $\mathbf{x}_b$ as representing the entire state of a numerical model at some time.
- The elements of $\mathbf{x}_b$ might be grid-point values, spherical harmonic coefficients, etc., and some elements may represent temperatures, others wind components, etc.
- We refer to $\mathbf{x}_b$ as the background
- Similarly, we generalize the observation to a vector $\mathbf{y}$.
- $\mathbf{y}$ can contain a disparate collection of observations at different locations, and of different variables.

Extension to Multiple Dimensions

- The major difference between the simple scalar example and the multi-dimensional case is that there is no longer a one-to-one correspondence between the elements of the observation vector and those of the background vector.
- It is no longer trivial to compare observations and background.
- Observations are not necessarily located at model gridpoints
- The observed variables (e.g. radiances) may not correspond directly with any of the variables of the model.
- To overcome this problem, we must asume that our model is a more-or-less complete representation of reality, so that we can always determine "model equivalents" of the observations.

Extension to Multiple Dimensions

- We formalize this by assuming the existance of an observation operator, $\mathcal{H}$.
- Given a model-space vector, $\mathbf{x}$, the vector $\mathcal{H}(\mathbf{x})$ can be compared directly with $\mathbf{y}$, and represents the "model equivalent" of $\mathbf{y}$.
- For now, we will assume that $\mathcal{H}$ is perfect. I.e. it does not introduce any error, so that:

$$\mathcal{H}(\mathbf{x}^*) = \mathbf{y}^*$$

where $\mathbf{x}^*$ is the true state, and $\mathbf{y}^*$ contains the true values of the observed quantities.

Extension to Multiple Dimensions

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}\left(\mathbf{y} - \mathcal{H}(\mathbf{x}_b)\right)$$

- Remember that in the scalar case, we had

$$\begin{aligned} T_a &= \alpha T_o + (1-\alpha) T_b \\ &= T_b + \alpha(T_o - T_b) \end{aligned}$$

- We see that the matrix $\mathbf{K}$ plays a role equivalent to that of the coefficient α.
- $\mathbf{K}$ is called the gain matrix.
- It determines the weight given to the observations
- It handles the transformation of information defined in "observation space" to the space of model variables.

Comparing model and observations

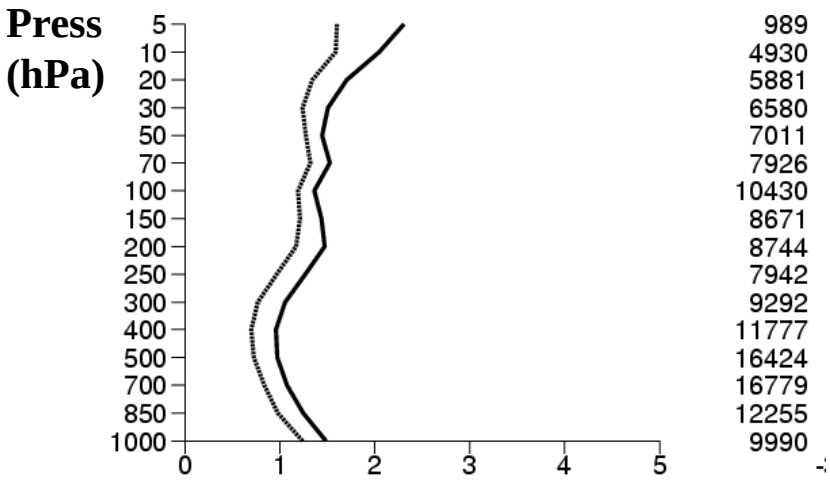
- The forecast model provides the background (or *prior*) information to the analysis
- Observation operators allow observations and model background to be compared (“O-B”)
- The differences are called departures or innovations
- They are central in providing observation information that corrects the background model fields
- These corrections, or increments, are added to the background to give the analysis (or *posterior estimate*)
- Observation operators also allow comparison of observations and the analysis (analysis departures “O-A”)

Example: Statistics of departures

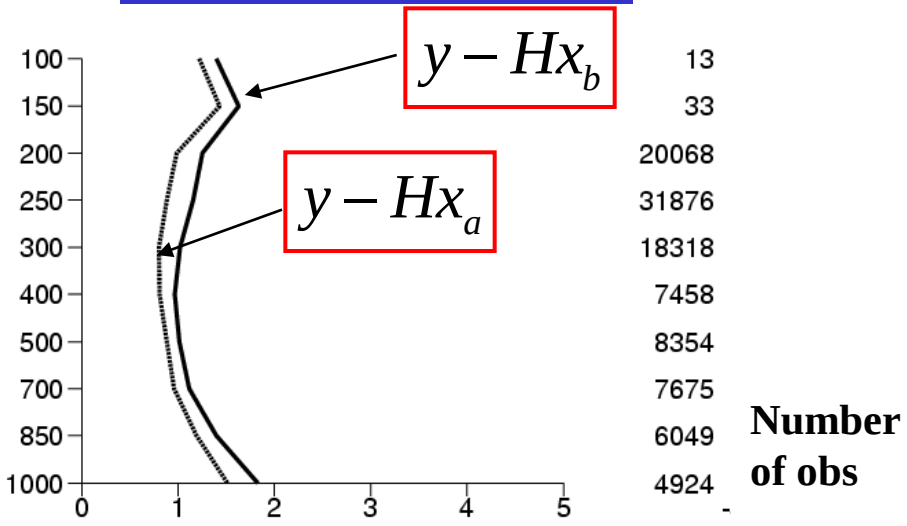
Background departures: $y - Hx_b$ (*O-B*)
Analysis departures: $y - Hx_a$ (*O-A*)

y = **observations**
x_a = **analysis state**
x_b = **background state**

Radiosonde temperature



Aircraft temperature



- The standard deviation of background departures for both radiosondes and aircraft is around 1-1.5 K in the mid-troposphere.
- The standard deviation of the analysis departures is here approximately 30% smaller – the analysis has “drawn” to the observations.

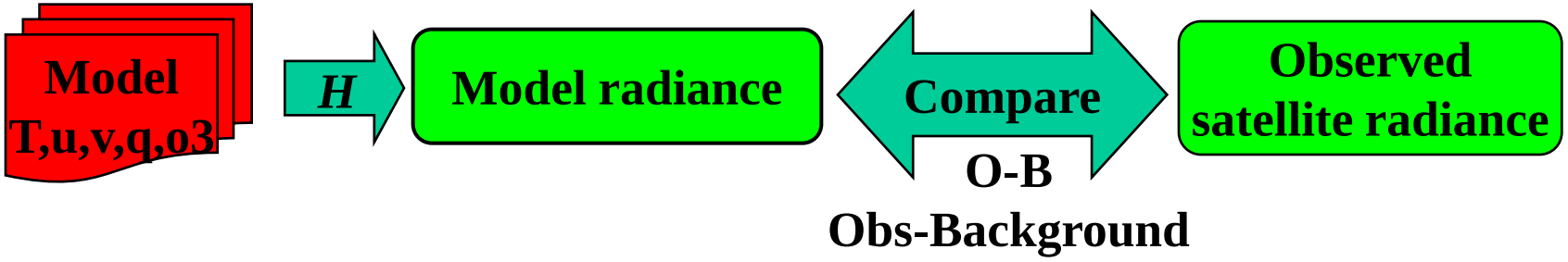
The observation operator "*H*"

How to compare model with observations? Satellites measure radiances/brightness temperatures/reflectivities, etc., NOT directly temperature, humidity and ozone.

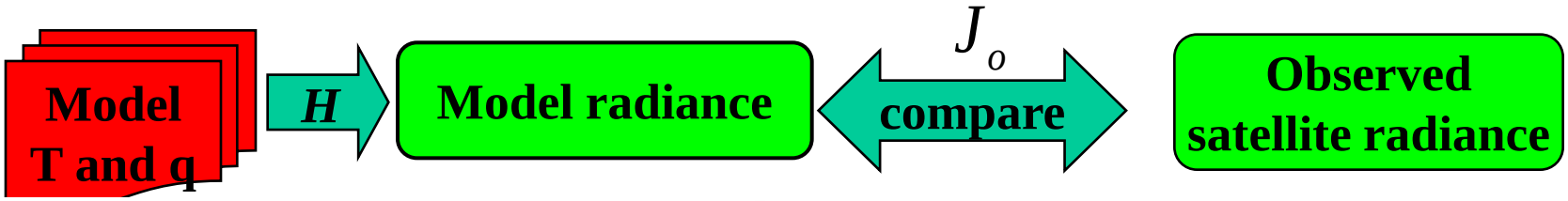
A model equivalent of the observation needs to be calculated to enable comparison in observation space (or related model-equivalent space).

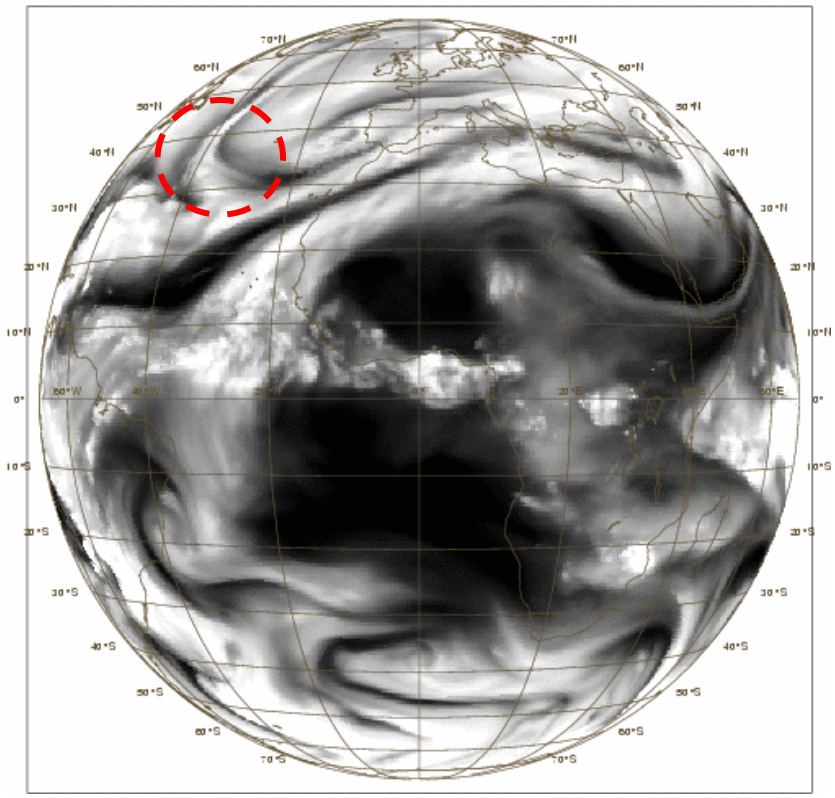
This is done with the 'observation operator', *H*.

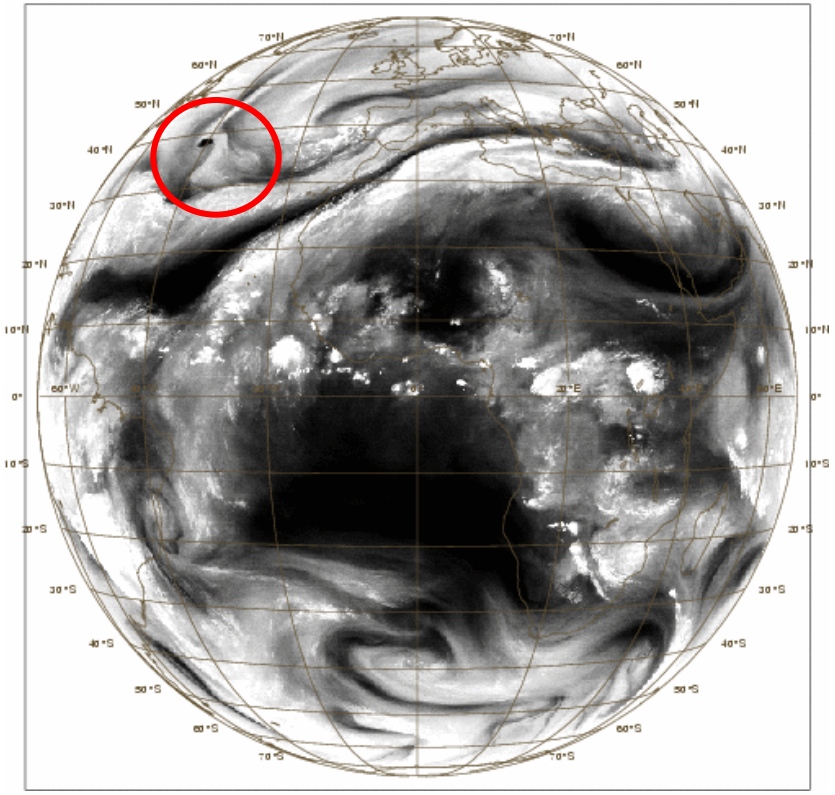
- *H* may be a simple interpolation from model grid to observation location for direct observations, for example, of winds or temperature from radiosondes
- *H* may possibly perform additional complex transformations of model variables to, for example, satellite radiances:



Forecast versus observed fields







Meteosat imagery – water vapour channel

"Usual" observation operators: Radiative transfer operator*

• Satellite instruments (active and passive) simply measure the **RADIANCE** *L* that reaches the top of the atmosphere at given frequency ν . The measured radiance is **related** to geophysical atmospheric variables (T,Q,O_3) by the **radiative transfer equation** (covered in detail in other lectures).

$$L(\nu) = \int_0^{\infty} B(\nu, T(z)) \left[\frac{d\tau(\nu)}{dz} \right] dz + \text{Surface emission} + \text{Surface reflection/ scattering} + \text{Cloud/rain contribution} + \ldots$$

• The observation operator for satellite measurements is in this case a convolution of the horizontal and vertical interpolation operators and the radiative transfer equation applied to the interpolated model state variables (for example, temperature, humidity) and solved via approximations which depend on the specific channel/frequency.

***** Details in presentations by T. McNally and at the SAT Training course next week!

“Unusual” observation operators: Aerosol Optical Depth

- The AOD operator is based on tabulated or parameterized optical properties of the N aerosol species that are modeled. These optical properties are then weighted by the mass of the particulate to provide the extinction coefficient and integrated vertically to give the total optical depth at a given wavelength:

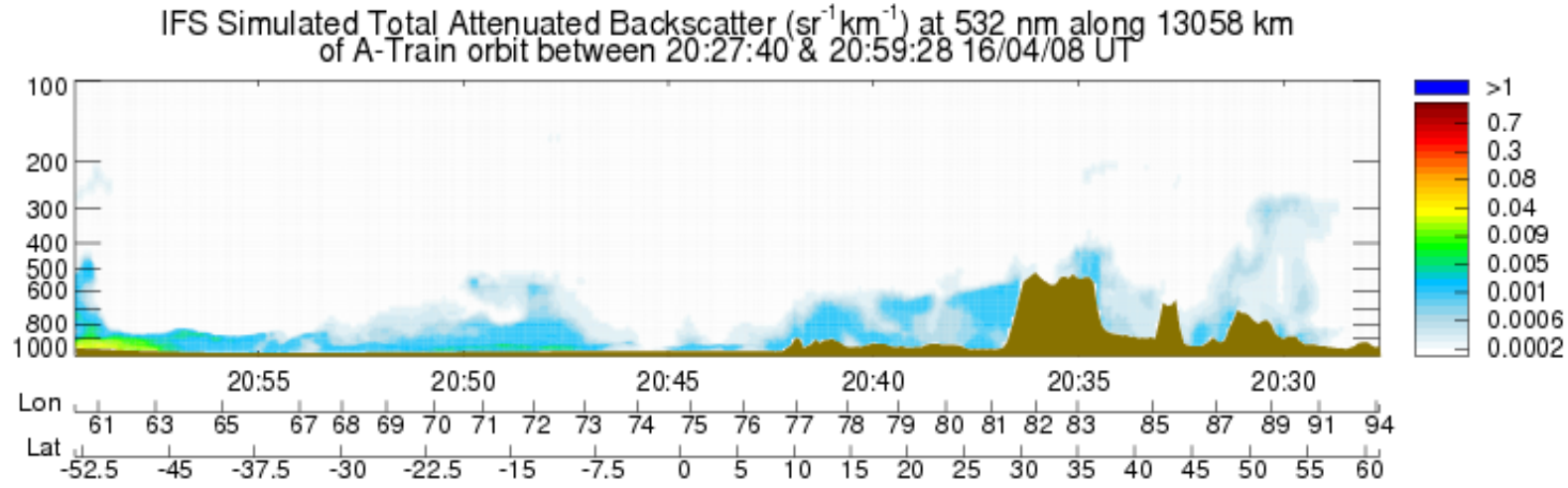
$$\tau = \int \sigma_{ext}(r, z, RH)dz$$

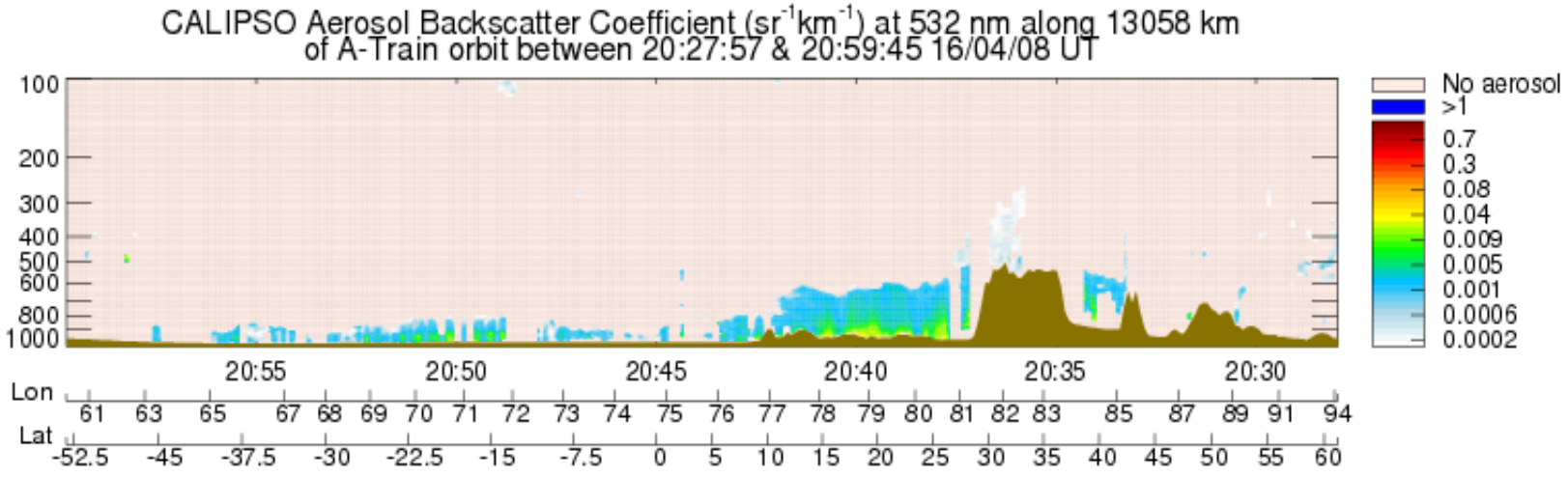
where $\sigma_{ext} = \sum_{i=1}^{N} \sigma_{ext,i}$ is the extinction coefficient which is a function of particle size and height (and relative humidity for some aerosol types).

- An aerosol optical depth operator is currently used operationally in the aerosol forecasts for the EU project “Monitoring of the Atmospheric Composition and Climate” (MACC) for assimilation of the Moderate Resolution Imaging Spectroradiometer (MODIS) Aerosol Optical Depths at 550 nm.

Forecast versus observed fields

CALIPSO satellite– lidar aerosol backscatter





Model aerosols

H

Model lidar backscatter

compare

Observed lidar backscatter

ECMWF

Direct observations versus retrievals

What about transforming observations closer to model space before assimilation?
It is possible to use “satellite retrievals” instead of “direct observations”.

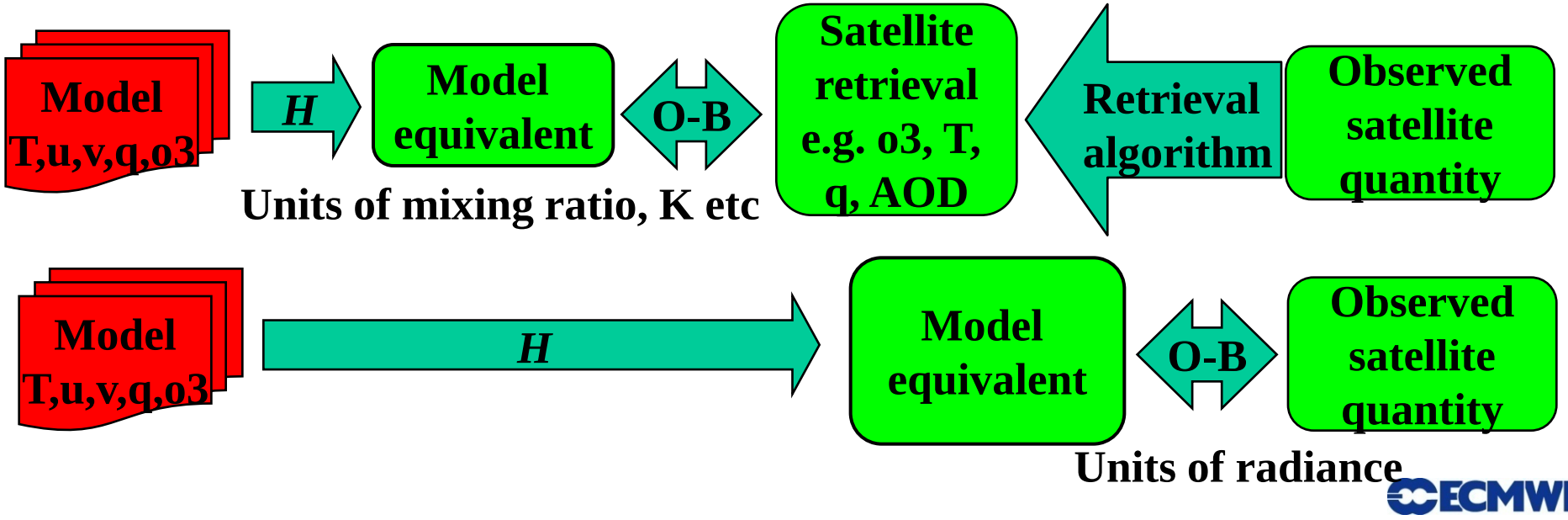
☺ *H* becomes simpler for a satellite retrieval product

☺ Related tangent linear & adjoint operators also simpler

☹ Retrieval assumptions made by data providers not always explicit/valid

☹ Use of NWP data in retrieval algorithms can introduce correlated errors

The choice is often application dependent!



Actual implementation of observation operators used in the ECMWF 4D Variational Data assimilation

The observation operator provides the link between the model variables and the observations (Lorenc 1986; Pailleux 1990).

The observation operator is typically implemented as a sequence of operators transforming the analysis control variable *x* into the equivalents of each observed quantity *y*, at observation locations.

This sequence of operators can be multi-variate (can depend on many variables) and may include:

- "Interpolation" from forecast time to observation time (in 4D-Var this is actually running the forecast model over the assimilation window)
- Horizontal and vertical interpolations
- Vertical integration
- If limb-geometry, also horizontal integration
- If radiances, radiative transfer computation
- Any other transformation to go from model space to observation space.

Actual implementation of observation operators used in the ECMWF 4D Variational Data assimilation

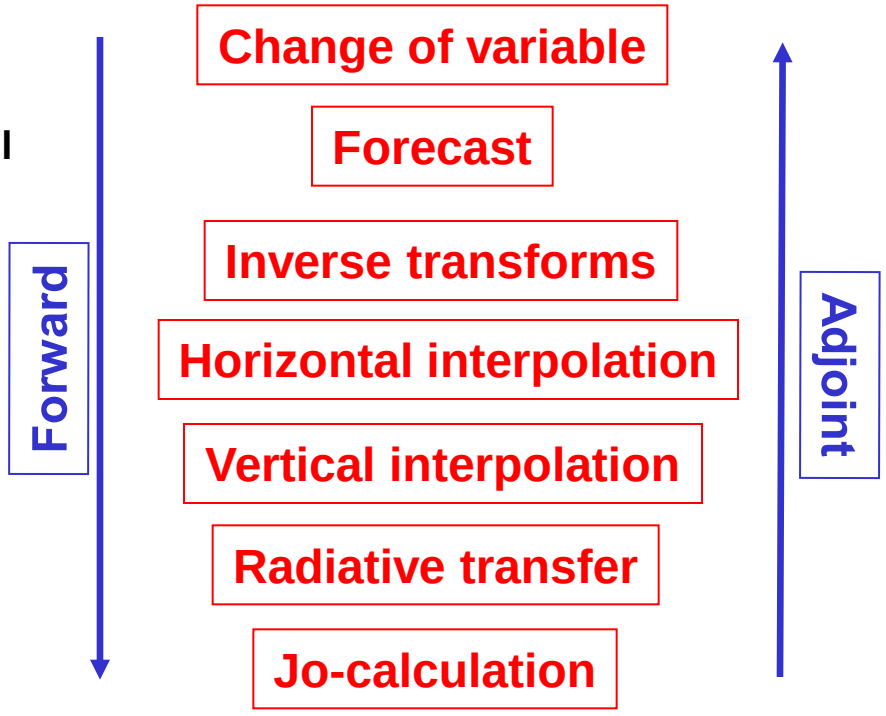
The observation operator is typically implemented as a sequence of operators transforming the analysis control variable *x* into the equivalents of each observed quantity *y*, at observation locations:

These first three steps are common for all data types:

- **The inverse *change of variable* converts from control variables to model variables**
- **The inverse *spectral transforms* put the model variables on the model's reduced Gaussian grid**
- **A 12-point bi-cubic *horizontal interpolation* gives vertical profiles of model variables at observation points**

Further steps (depend on the specific observations treated):

- ***Vertical interpolation* to the level of the observations. The vertical operations depend on the observed variable**
- ***Vertical integration* of, for example, the hydrostatic equation to obtain geopotential, and of the radiative transfer equation to obtain top of the atmosphere radiances.**



Upper-air observation operators

The following vertical interpolation techniques are employed:

Wind:	linear in ln(p) on full model levels
Geopotential:	as for wind, but the interpolated quantity is the deviation from the ICAO standard atmosphere
Humidity:	linear in p on full model levels
Temperature:	linear in p on full model levels
Ozone:	linear in p on full model levels

Interpolation of highly nonlinear fields

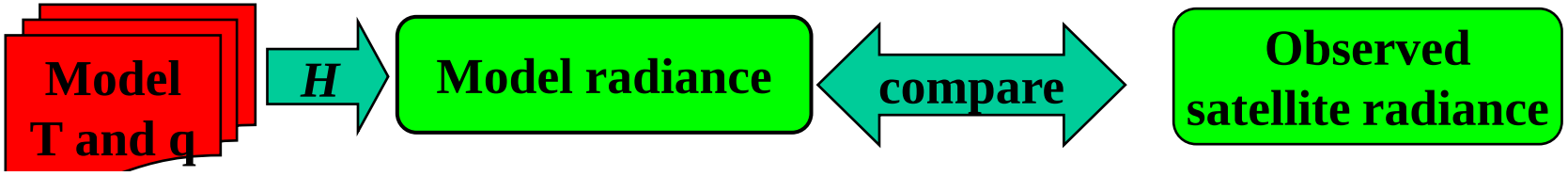
- The problem of a “correct” horizontal interpolation is especially felt when dealing with heterogeneous model fields such as precipitation and clouds.

- In the special case of current rain assimilation, model fields are not interpolated to the observation location. Instead, an average of observation values is compared with the model-equivalent at a model grid-point – observations are interpolated to model locations.

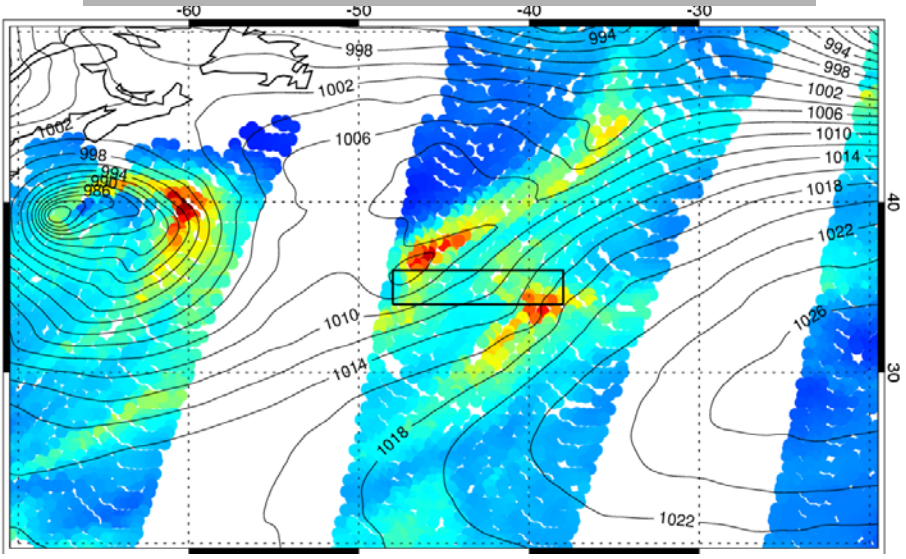
- Different choices can affect the departure statistics which in turn affect the observation error assigned to an observation, and hence the weight given to observations.

Observation operator for and use of rain affected microwave radiances – a difficult task

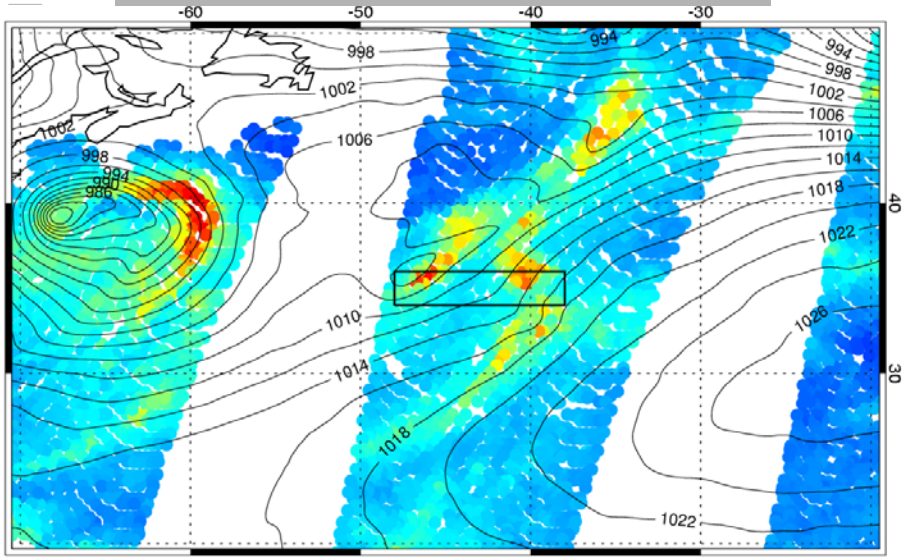
Main difficulties: inaccurate moist physics parameterizations (location/intensity), formulation of observation errors, bias correction, linearity assumptions



4D-Var first guess SSM/I ΔTb 19v-19h [K]



SSM/I observational ΔTb 19v-19h [K]



Representativeness Errors

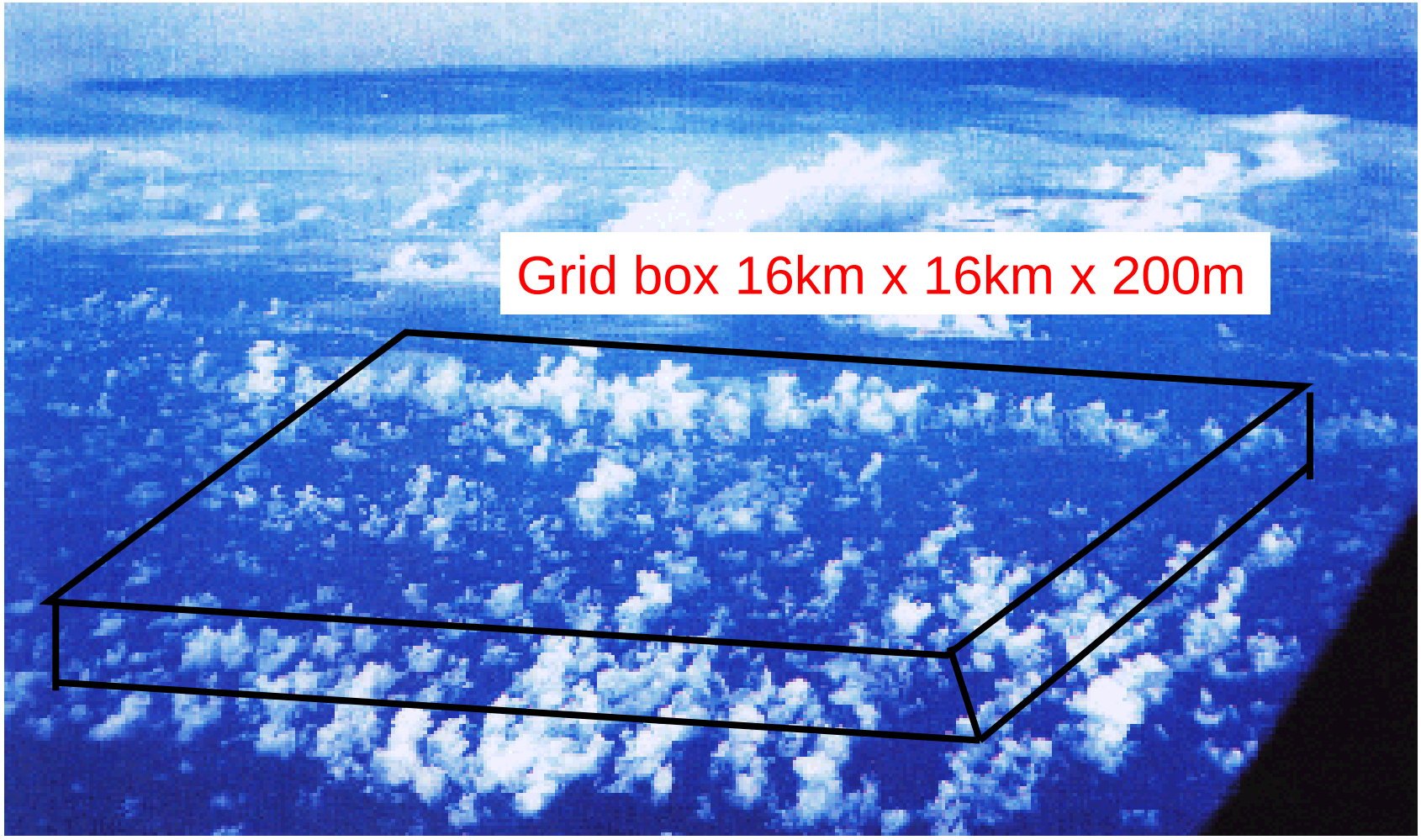
- Errors introduced by the interpolation operator, and also by the finite resolution of the model fields, are often accounted for by adding “representativeness” errors to the “instrument” error of the observations.

- For some data, e.g. radiosonde winds, the representativeness errors are the dominant contribution to the observation errors in the matrix R.

- In effect, we compare model and observations not in observation space, but in model-equivalent space. We ask what would the observed quantity be if we degraded the atmosphere down to model resolution?

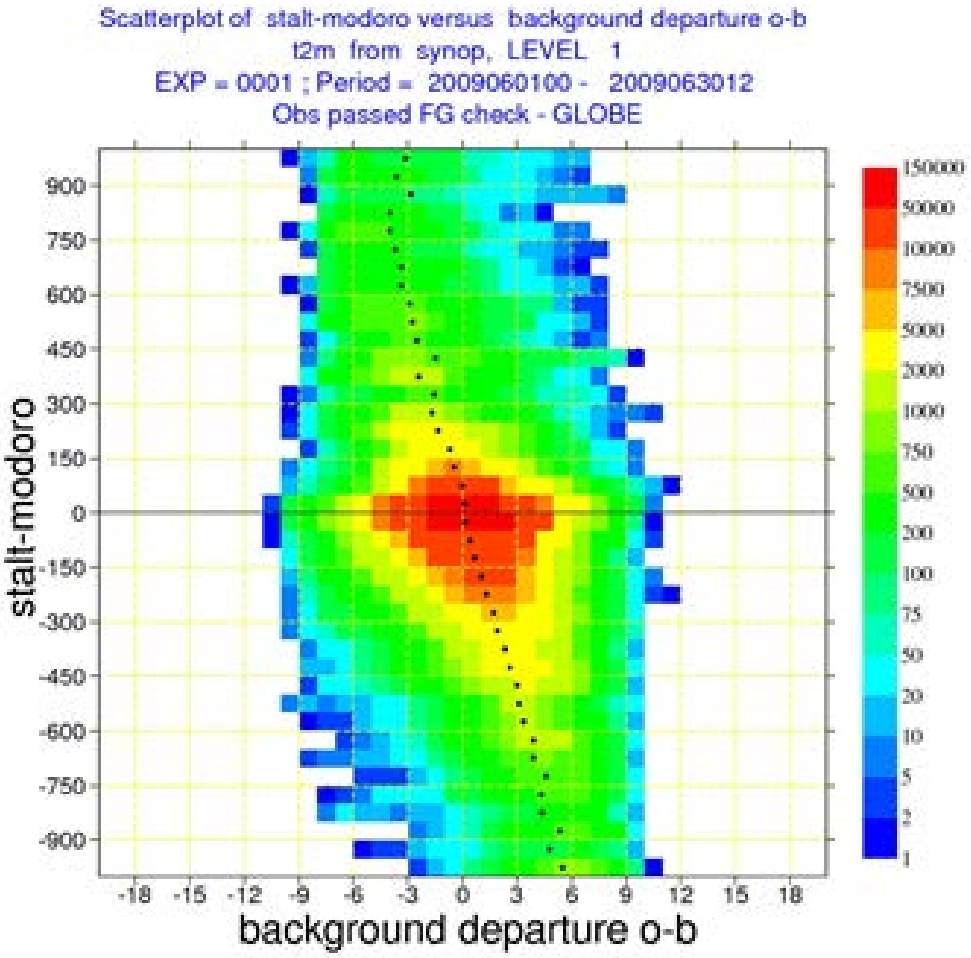
Representativeness Errors

Model grid box small, but still large compared to reality



Representativeness errors

At ECMWF 2 metre SYNOP temperatures are are not adjusted for differences between station height and model height. This shows up as a lapse rate error of close to 6.5 °/km.

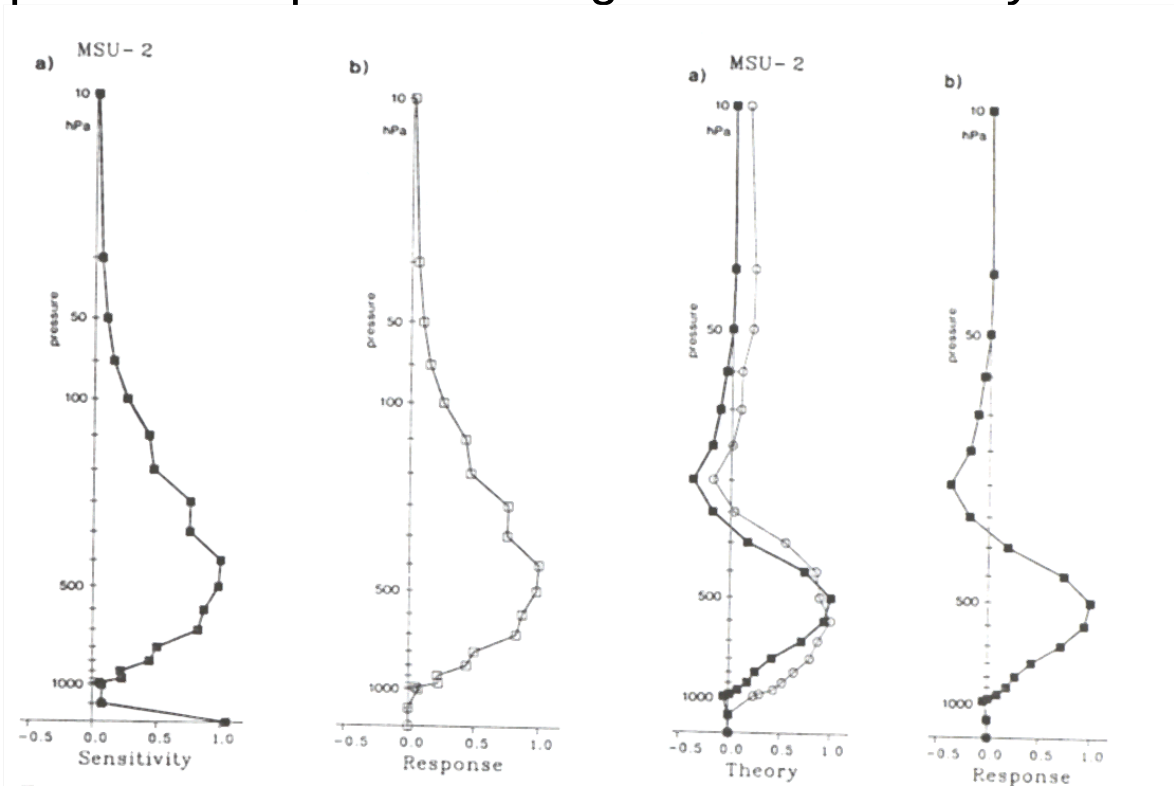


The Jacobian matrix

- The tangent linear of the observation operator H consists of the partial derivatives of *H* with respect to all of its inputs, expressing the variations of *H* in the vicinity of the background x_b.
- H is sometimes referred to as the Jacobian of *H*. If H varies significantly with x_b then *H* is non-linear.
- Needed for the incremental formulation of 4D-Var (adjoint too). You will more about this later this week.
- But also, for complex observation operators, study of the Jacobian highlights observation sensitivities to input model variables at specific points (quantified as information content, effectively which model components have been “observed”).
- In the case of radiances, for example, columns of H will express each channel’s ‘weighting function’ on the model geometry.

A Jacobian example: MSU-2

- The Jacobian for TOVS channel MSU-2 is shown in the leftmost diagram. It can be interpreted as a profile of weights for a vertically averaged temperature.
- The second diagram on the left shows the analysis increment in the special case of a diagonal B-matrix.
- The two panels to the right show theoretical and actual analysis increments, with a realistic B-matrix.



- Weighting functions and analysis increments are similar but not identical because they also include the background term (as per equation):

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

Summary: Variational data assimilation allows easy use of direct observations

- One of the advantages of variational data assimilation is that it allows the direct assimilation of radiances and other “indirect” observations (e.g. optical depth, reflectivities, lidar backscatter):
  - Physical (based on radiative transfer calculations)
  - Simultaneous (in T, q, o3 and ps …)
  - Uses an accurate short-range forecast as background information, and its error covariance as a constraint
- In 3D-Var:
  - Horizontal consistency is ensured
  - Other data types are used simultaneously
  - Mass-wind balance constraints are imposed
- In 4D-Var:
  - Consistency in time. Frequent data can be used.
  - The dynamics of the model acts as additional constraint

Summary: This flexibility is very important

The notion of 'observation operators' makes variational assimilation systems particularly flexible with respect to their use of observations.

- This has been shown to be of real importance for the assimilation of radiance data, for example. And it will be even more important in the coming years as a large variety of data from additional space-based observing systems become available - each with different characteristics.

- There is no need to convert the observed data to correspond to the model quantities. The retrieval is, instead, seen and integral part of the variational estimation problem.

Summary: Role of the observation operator *H*

Every observed quantity that can be related to the model variables in a meaningful way, can be used in 3D/4D-Var. The link to enable the comparison of model and observations is the observation operator *H*

- Accurately computing *H*(x) is important for deriving background departures and the assimilation's analysis increments. Assumptions/approximations within the observational operator need to be taken into account, including the interpolation error and finite model resolution (they contribute to the representativeness error).
- One observation can depend on several model quantities: Geop=*H*(T,Ps,q), Rad=*H*(T,q,o3,…). That means it is multi-variate.
- *H* may be non-linear. Some TOVS channels vary strongly non-linearly with q (humidity), for example. Its first derivative H (the Jacobian) then depends on the atmospheric state.

Moreover, when the observations are significantly influenced by geophysical quantities that are not analysed in the system (e.g. surface emissivity, skin temperature, clouds or precipitation) then difficulties may arise.

Summary:
Typical issues associated with observation usage

- How can we compare the model to the observations?
  - Can we implement a forward operator *H* ?
    - Do we have all the information needed by the forward model?
    - How about the tangent linear and adjoint operators?
    - Are the operators computationally affordable?
  - Would it be better to use a retrieval product made elsewhere?
    - What additional assumptions are being made? Are they valid?
    - Can the processing assumptions/parameters be controlled?
  - There is often a transition from retrievals to more raw products like radiances
    - This can take several years

- How much weight should we give to the observations and background?
  - What are the sources of error?
  - Are the errors random or systematic?

- Answers from previous experience and further research
- Even “old” observations need to be re-assessed (i.e. fundamental for reanalysis activities)