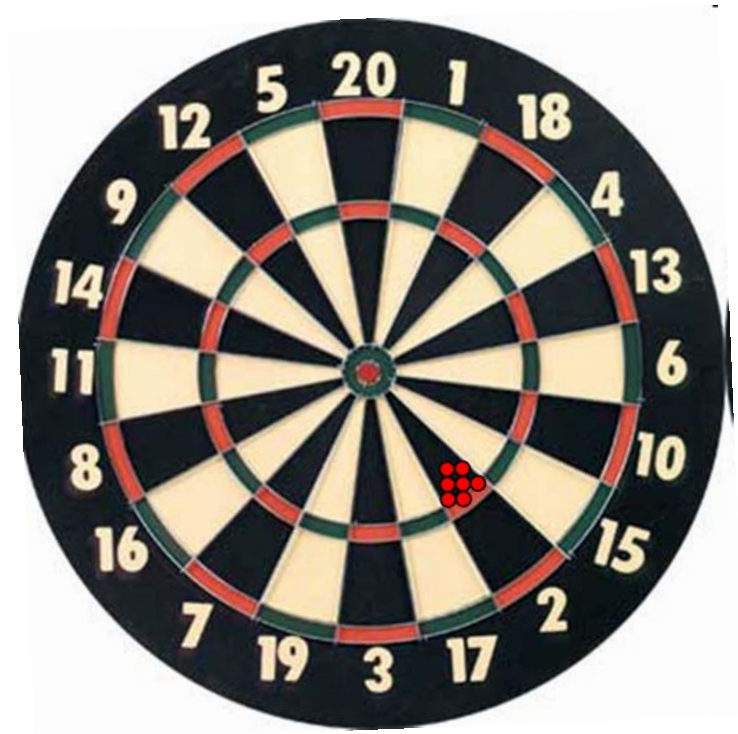
Appropriate usage of errors in the comparison of satellite retrieval

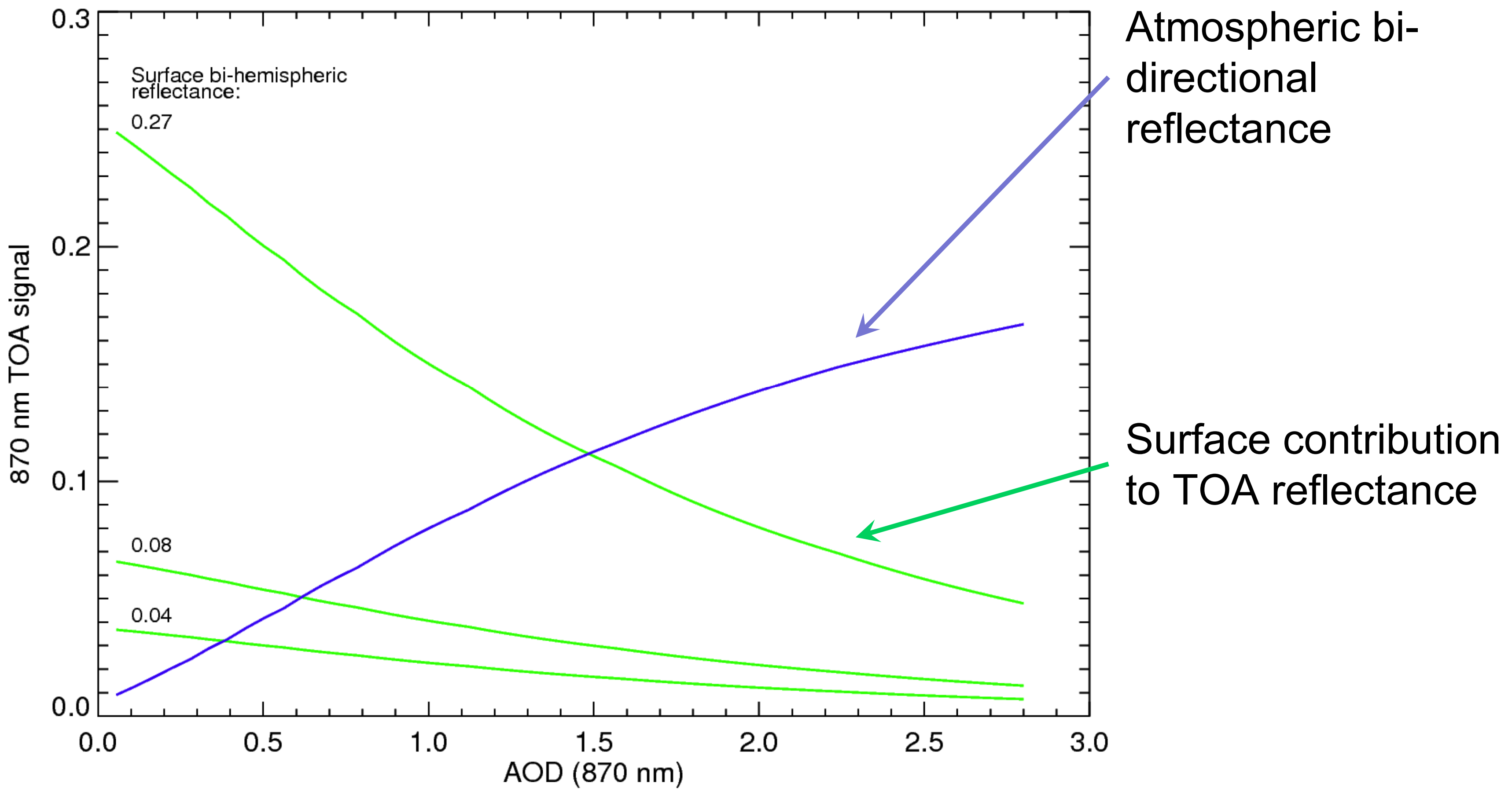


Gareth Thomas for Don Grainger

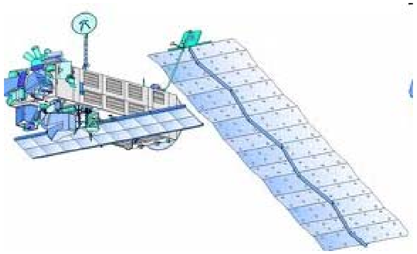
Earth Observation Data Group

Pixel-by-pixel errors?

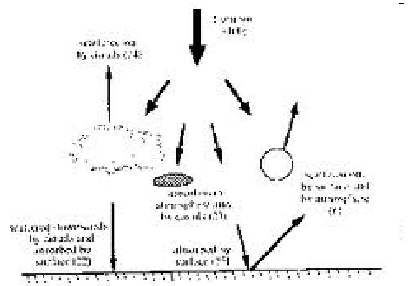
Both the atmospheric and surface state effect the precision of retrievals of either



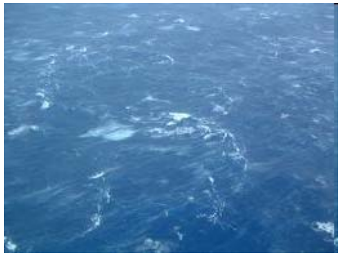
Formalisation



Measurement with
known error



Physics –
radiative transfer



What we want to know
e.g. sea surface temperature

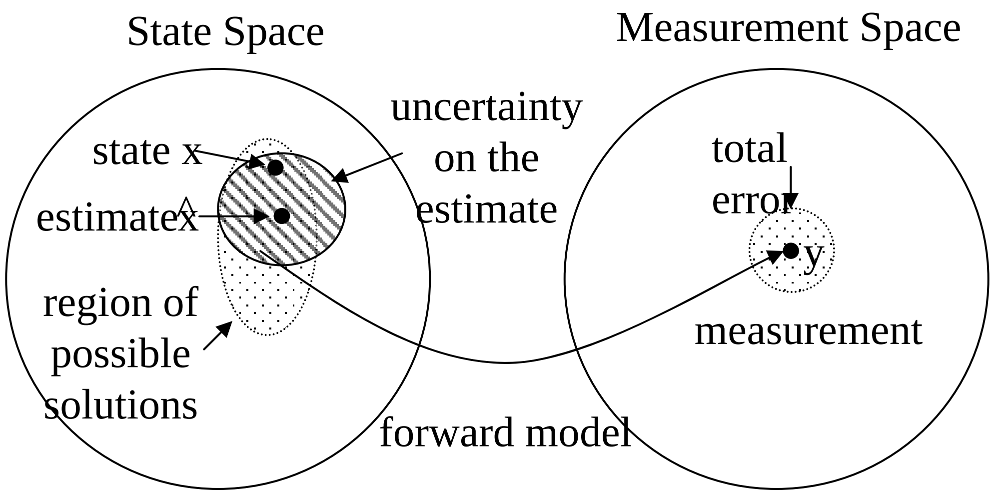
Measurement	Forward Model	State	Error Covariance
$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$	$f(\mathbf{x}, b)$	$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$	$\begin{bmatrix} \langle\epsilon_1\epsilon_1\rangle & \langle\epsilon_1\epsilon_2\rangle & \cdots & \langle\epsilon_1\epsilon_n\rangle \\ \langle\epsilon_2\epsilon_1\rangle & \langle\epsilon_2\epsilon_2\rangle & \cdots & \langle\epsilon_2\epsilon_n\rangle \\ \vdots & \vdots & & \vdots \\ \langle\epsilon_n\epsilon_1\rangle & \langle\epsilon_n\epsilon_2\rangle & \cdots & \langle\epsilon_n\epsilon_n\rangle \end{bmatrix}$ (usually diagonal)

$$\boxed{\mathbf{y} = \mathbf{K}\mathbf{x}}$$

$\mathbf{K}$ is the weighting function matrix

The Optimal Estimate

$$\mathbf{y} \quad = \quad \mathbf{Kx} \qquad \Rightarrow \qquad \mathbf{x} = \mathbf{Gy} \qquad \text{BUT} \qquad \mathbf{G} \neq \mathbf{K}^{-1}$$



The optimal estimate

$$\hat{\mathbf{x}} \quad = \quad \mathbf{Gy} = (\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\epsilon}^{-1}\mathbf{K} + \mathbf{S}_a)^{-1}(\mathbf{K}^{\mathrm{T}}\mathbf{S}_{\epsilon}^{-1}\mathbf{y} + \mathbf{S}_a\mathbf{x}_a).$$

where

$\mathbf{S}_{\epsilon}$ is the error covariance matrix

$\mathbf{x}_a$ is the apriori values of the state

$\mathbf{S}_a$ is the a priori error covariance matrix

The solution

- For the general non-linear problem we find the solution through iteration

The optimal estimate solution, $\hat{\mathbf{x}}$, minimises a joint cost function χ^2 defined by

$$\chi^2 = (\mathbf{y} - F(\hat{\mathbf{x}}))^{\mathrm{T}} \mathbf{S}_\epsilon^{-1} (\mathbf{y} - F(\hat{\mathbf{x}})) + (\hat{\mathbf{x}} - \mathbf{x}_a)^{\mathrm{T}} \mathbf{S}_a^{-1} (\hat{\mathbf{x}} - \mathbf{x}_a).$$

You get

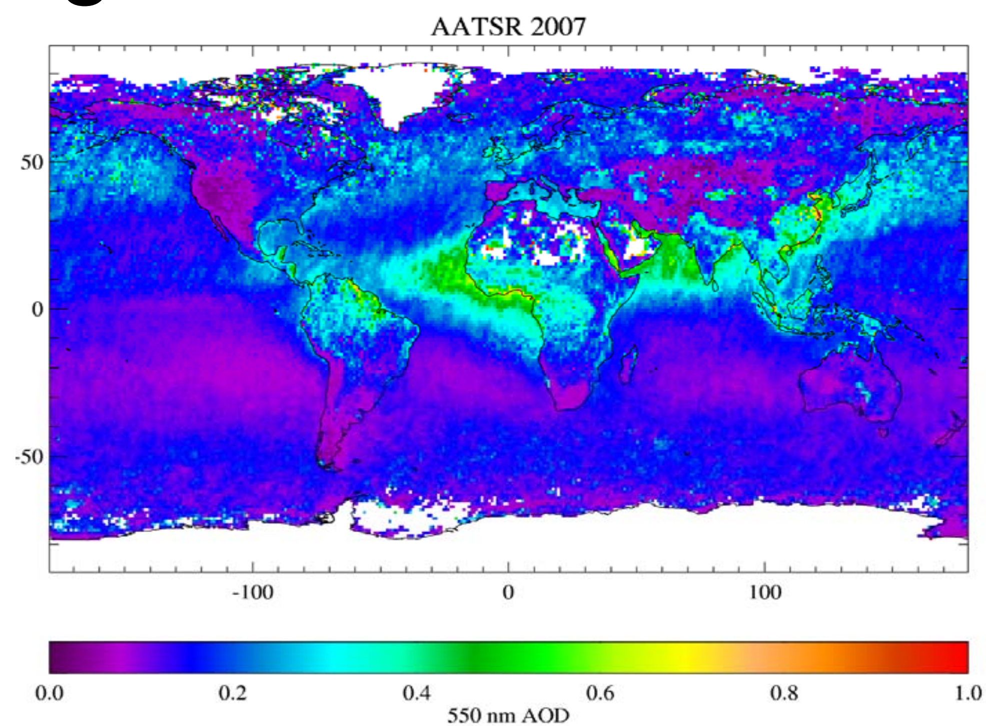
- quality control – how well the measurements have been fitted
- uncertainty – the measurement & forward model error propagated into state space

$$\hat{\mathbf{S}}^{-1} = (\mathbf{K}^T \mathbf{S}_\epsilon^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1}$$

Oxford-RAL Aerosol & Cloud

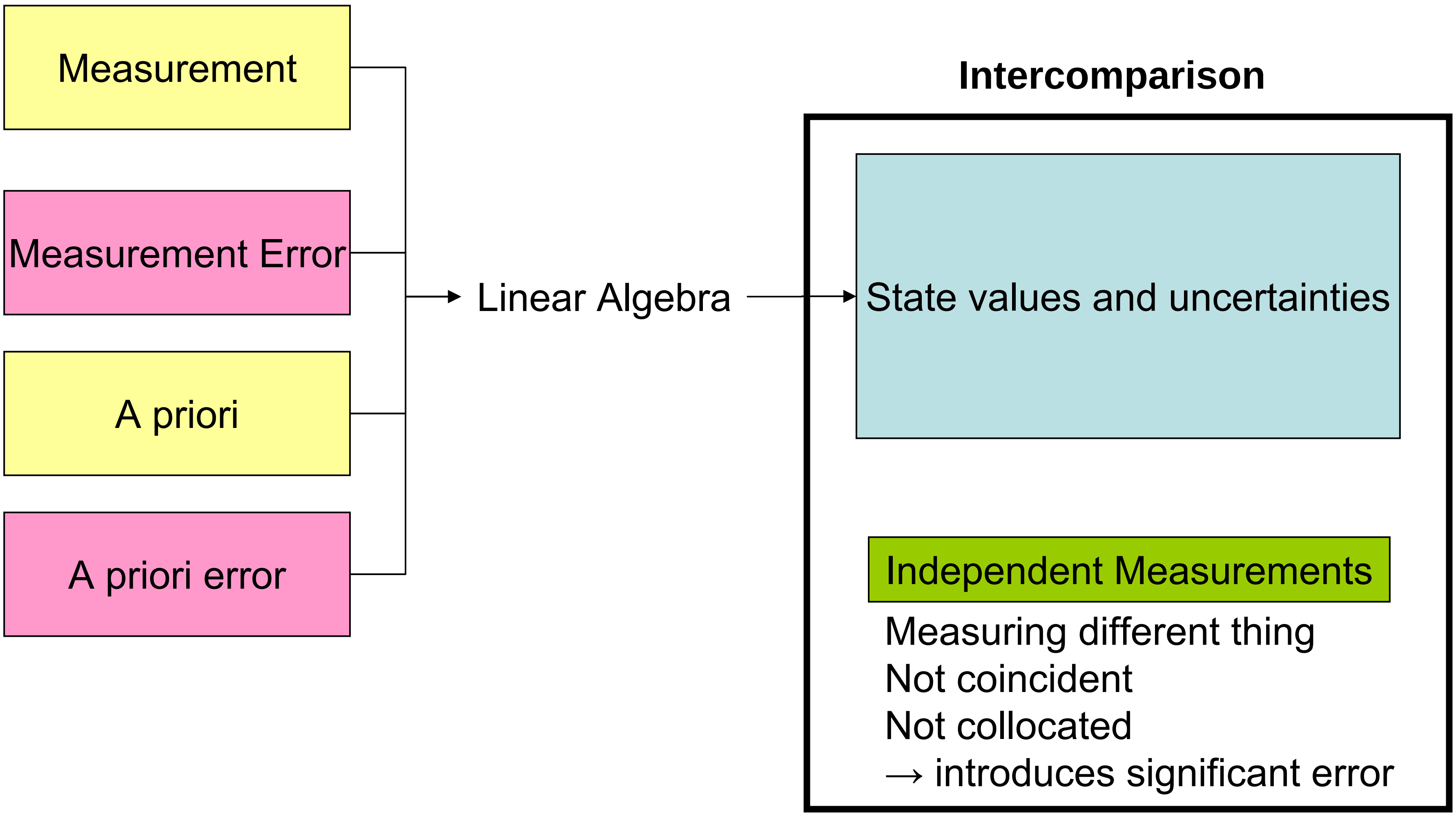
Optimal estimation aerosol and cloud retrievals

- GlobAEROSOL
- GRAPE (cloud & aerosol)
- ESA Climate Change Initiative:
  - Cloud_cci
  - Aerosol_cci



GlobAEROSOL annual mean AOD from ORAC-AATSR

Comparing with other data



Error Components

$$\hat{\mathbf{x}} - \mathbf{x} \quad = \quad (\mathbf{A} - \mathbf{I}_n)(\mathbf{x} - \mathbf{x}_a) + \mathbf{G}_y \mathbf{K}_b (\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{G}_y \Delta \mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \mathbf{G}_y \boldsymbol{\epsilon}.$$

- The first term, $(\mathbf{A} - \mathbf{I}_n)(\mathbf{x} - \mathbf{x}_a)$, is known as the smoothing error. This is the error due to the lack of sensitivity of the observing system to the individual parameters of the state vector. This term will be zero if on average $\mathbf{x} = \mathbf{x}_a$, i.e. the set of potential $\mathbf{x}$ is unbiased with respect to the a priori.

- The second term, $\mathbf{G}_y \mathbf{K}_b (\mathbf{b} - \hat{\mathbf{b}})$, is known as the model parameter error. Typically $\mathbf{S}_b$ is a diagonal matrix with the elements of the diagonal being the uncertainties in the model parameters.

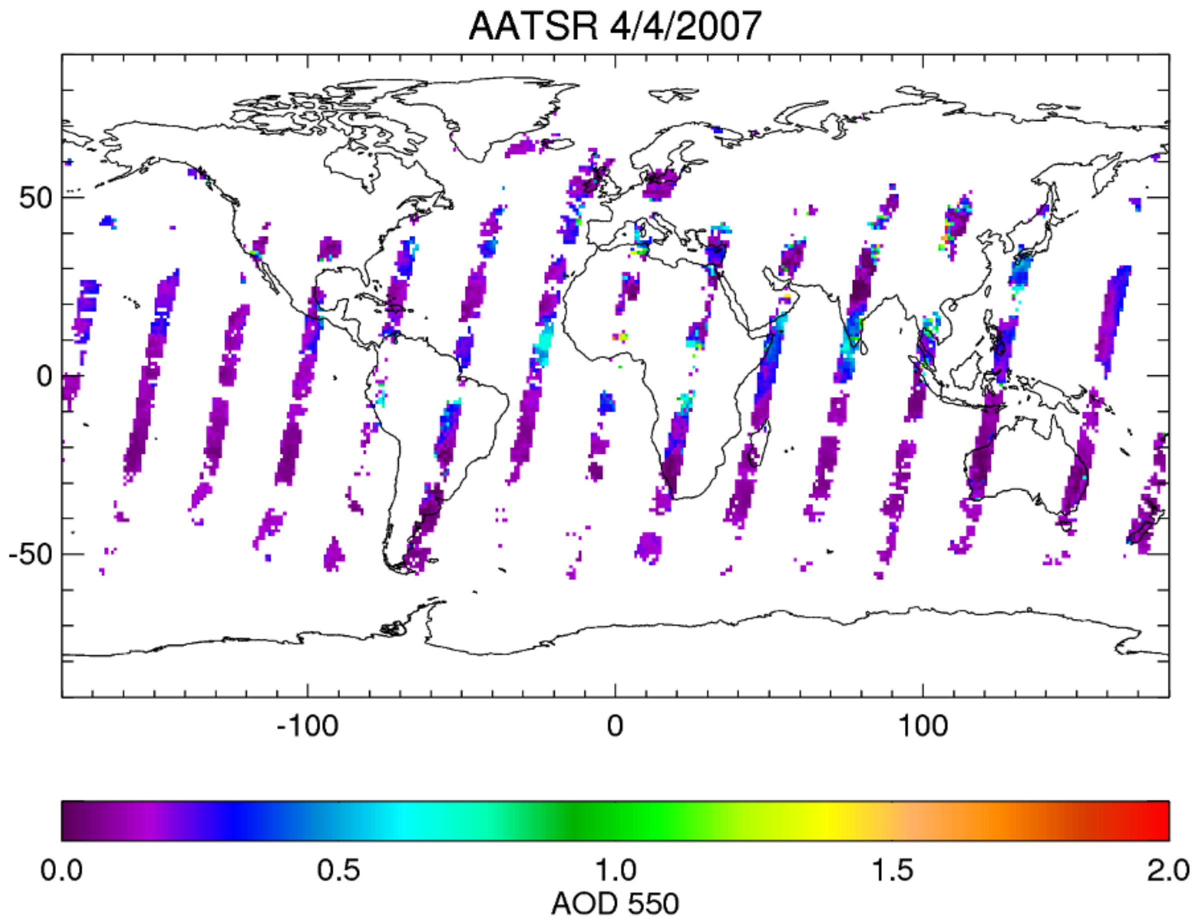
Error Components

$$\hat{\mathbf{x}} - \mathbf{x} \quad = \quad (\mathbf{A} - \mathbf{I}_n)(\mathbf{x} - \mathbf{x}_a) + \mathbf{G}_y\mathbf{K}_b(\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{G}_y\Delta\mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{b}') + \mathbf{G}_y\boldsymbol{\epsilon}.$$

- The third term, $\mathbf{G}_y\Delta\mathbf{f}(\mathbf{x}, \mathbf{b}, \mathbf{b}')$, is known as the forward model error. If the forward model is based on a mathematical approximation then the forward model error is calculated as the typical difference between the approximation and the more exact model. In other cases knowledge of the true physics may be so poor as to make estimates of the forward model error little more than an educated guess.

- The final term, $\mathbf{G}_y\boldsymbol{\epsilon}$, is known as the retrieval noise. It can be interpreted as the measurement noise projected into state space and its covariance is represented by $\mathbf{G}_y\mathbf{S}_y\mathbf{G}_y^{\mathrm{T}}$.

Sampling



Satellite products generally provide sparse coverage

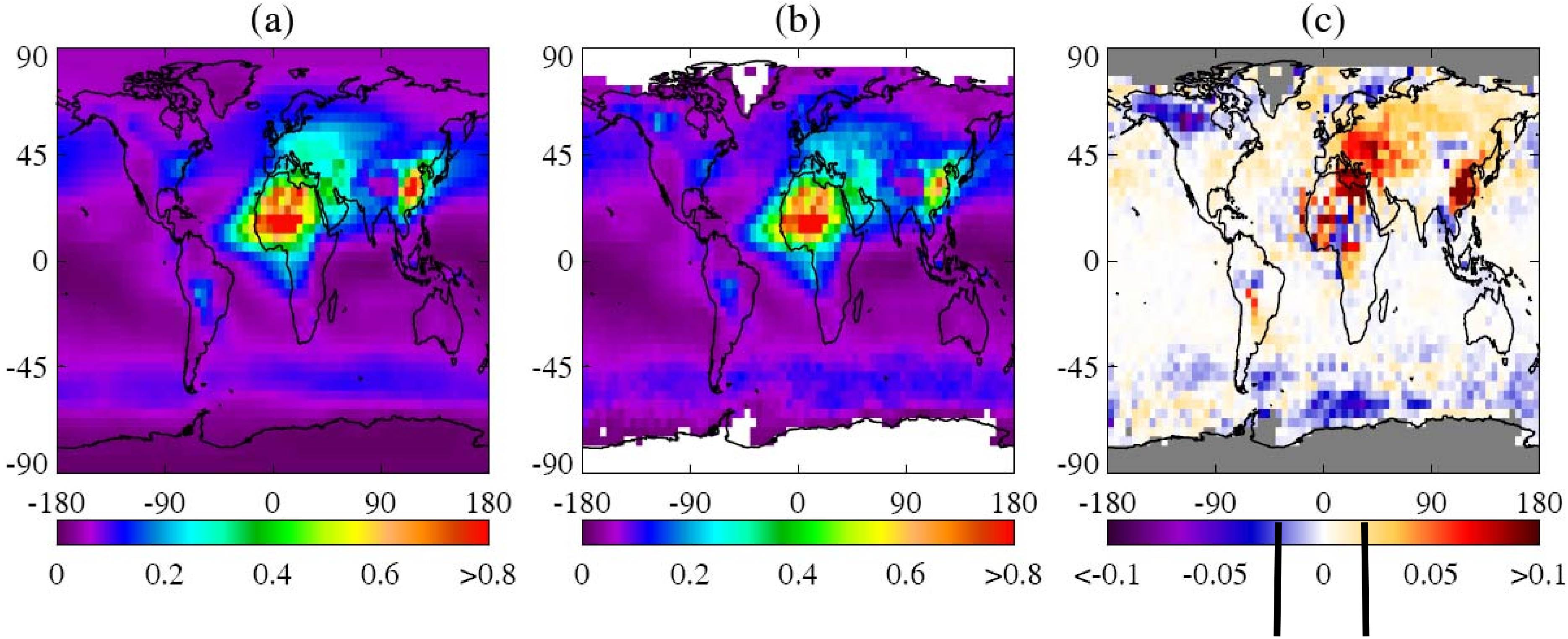
Even wide swath instruments are still limited by cloud cover

Aerosol loading doesn't generally follow anything like a Gaussian distribution

Thus, sampling is important!

Neglecting sampling…

All Model Values | Model Coincident with Observations | Difference



Requested Accuracy = 0.02

Fig. 1. Annual mean 550 nm AOD field from GEOS-Chem, generated for all data in (a) and by averaging only those days with coincident AATSR observations in (b). The difference between the two ('all data' - 'any GlobAerosol data') is shown in (c).

Sayer et al: Some implications of sampling choices on comparisons between satellite and model aerosol optical depth fields, ACPD, 10, 17789-17814, 2010.

Conclusion

- OE pixel-by-pixel errors give you a measure of how well constrained each retrieval is
- Validation gives you the accuracy of a product overall
- Monthly mean (level 3) are not, in general, suitable for comparison:
  - Don’t validate at level 3 unless you take into account the spatial correlation of errors
- Better to validate at L2 by “flying” an observation operator over model output