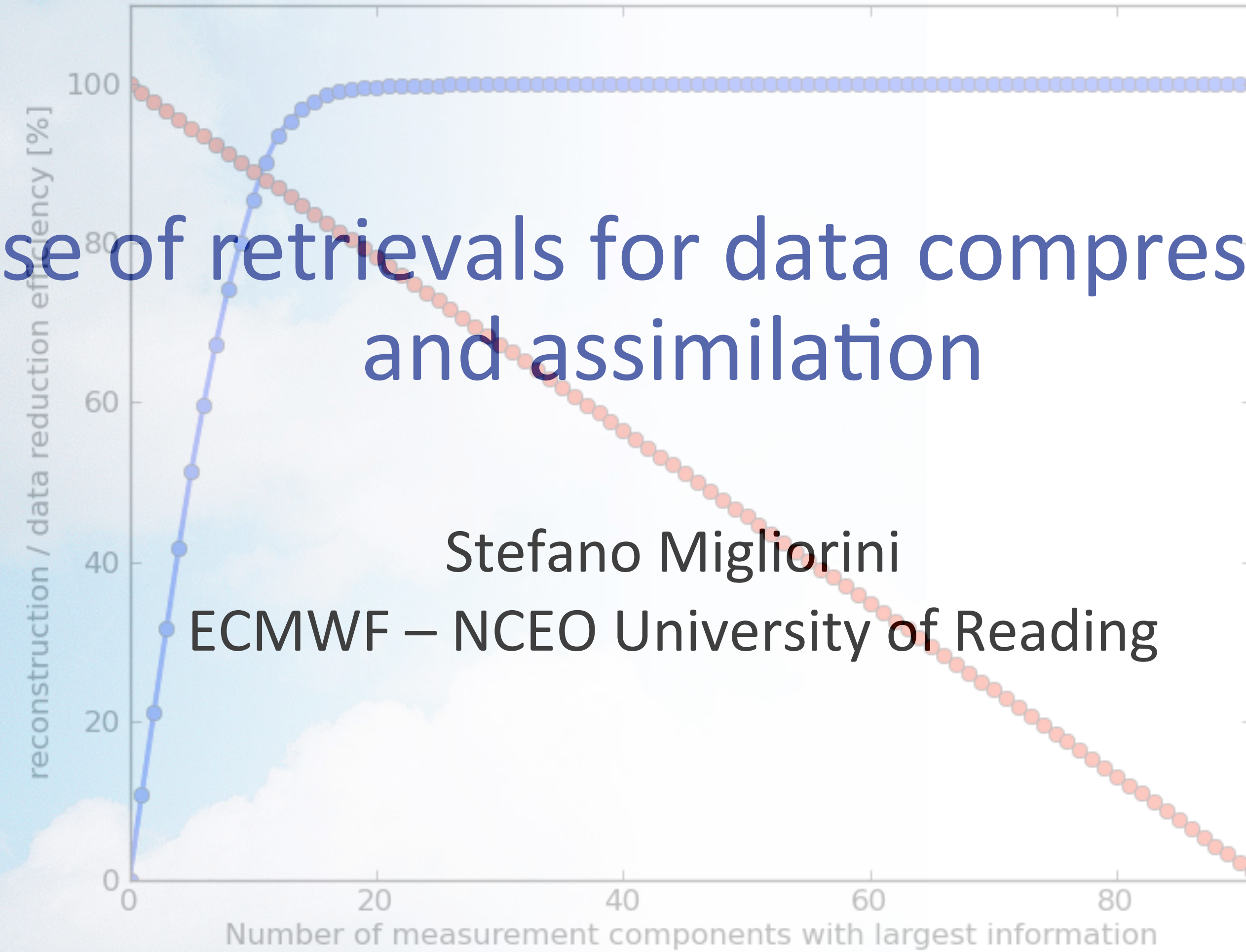
Use of retrievals for data compression and assimilation



Stefano Migliorini

ECMWF – NCEO University of Reading

Introduction

Some basic facts:

- Skilful weather forecasts require *realistic* models (with "small" model error) and accurate initial (and boundary) conditions for prediction
- Observations are needed to improve accuracy of initial conditions for prediction
- More sophisticated numerical schemes and physical parametrizations can improve the model's realism, which needs to be assessed through comparisons with *observations*
- The only possibility to reduce uncertainty at predictable scales, is to make appropriate *measurements*
- Observations of a given system (e.g., the atmosphere) over a period of time are an essential source of *information*
- This does not necessarily mean we need as many obs as possible, but that we have to make sure we extract from them as much information as possible

Issues

- Observability problem:

  Given a set of measurements **y**(t) over t ∈ [0, t] it is not possible to determine **x**(t) for all t ∈ [0,t]
- Ill posed problem: in general we cannot extract information about all *n* relevant components of the Earth system at any given time (*n* is infinite for continuous profiles!)
- Nonlinear inverse problem. Needs fully nonlinear methods or iterative linearizations (that may change the nature of the problem!)
- Quantitative estimate of information requires (statistical) knowledge of errors, which are hidden
- Measurements and models can be biased
- In essence, the statement of our problem is: how can we choose as least measurements as possible to provide estimate of the observable part of the state when we need to solve an ill-posed nonlinear inverse problem affected by errors with not entirely known statistical distributions?

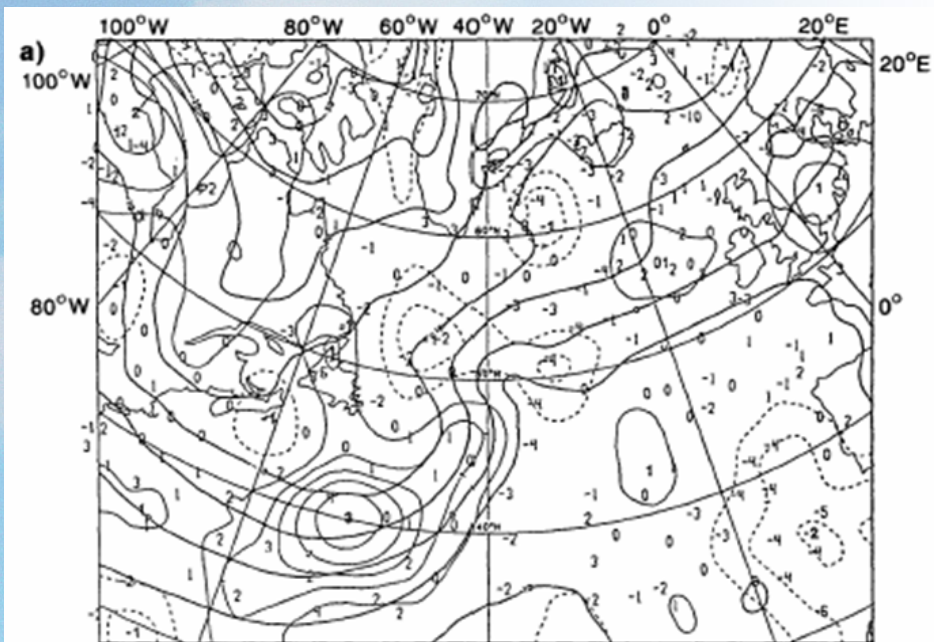
Transformation to retrieval space

- A possible way to proceed is in two steps:

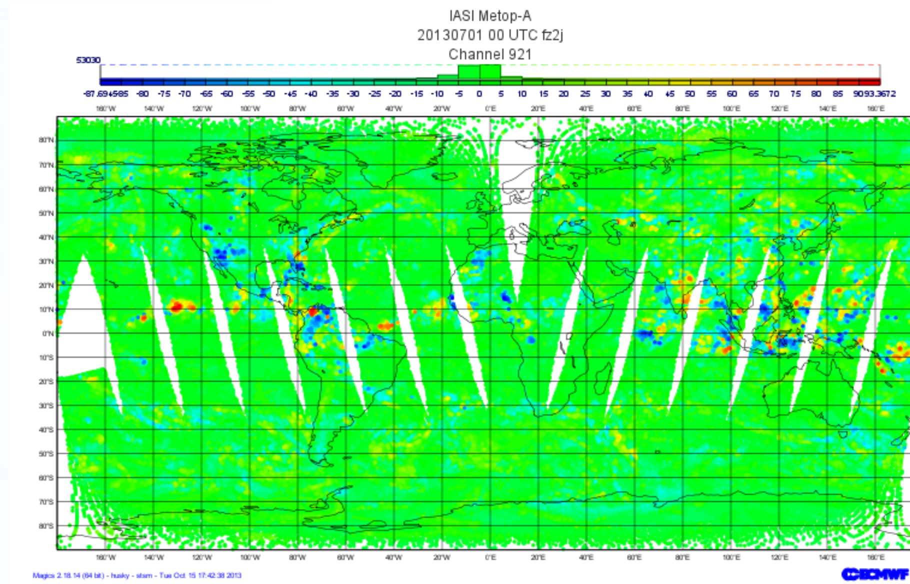
  first retrieve an estimate of the state using an inverse method and then assimilate the retrieval
- The late 1970s saw the first attempts to assimilate temperature retrievals from satellite sounders for numerical weather prediction (NWP).
- Initial results had a modest impact on forecast skill (best over oceans).
- In the 1980s, improvements in NWP models caused reduction of impact of satellite data.
- Problems due to background information contained in retrievals inconsistent with that used in data assimilation: bias and spurious correlations
- Early 1990s: variational assimilation for NWP. Observation operator can be nonlinear: assimilation of satellite radiances.

Comparing models and observations

- Retrieval increments and brightness temperature innovations



From Andersson et al (1991) and cited in Eyre (2007).
Analysis increments and background, 1000-700 hPa



Metop-A IASI innovations (Tb) 30 June 2013 21 UTC – 1 July 2013 09 UTC

Assimilation of satellite radiances

- Radiance assimilation has since proved to obtain excellent results, especially with passive remote sounders of temperature and humidity.
  Simple error structure and effective observation monitoring.
- Among current challenges:
  - assimilation of cloud-affected radiances in the infrared.
  - atmospheric composition sounding.
- Observation operator represents solution of radiative transfer eq. (compliant with NWP near-real-time requirements) and has to model characteristics of the instrument. Very high number of channels for high-res sounders.
- Recent interest (e.g., Joiner and Da Silva, 1998; Rodgers 2000, Migliorini et al., 2008, 2011) in efficient assimilation of reduced amount of sounding data: e.g. efficient assimilation of "transformed retrievals".

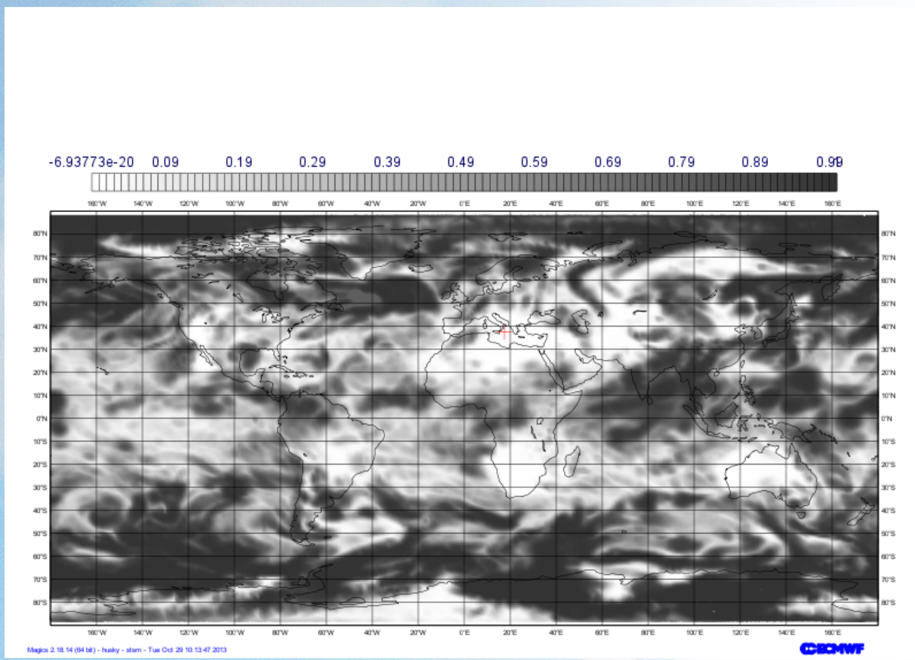
Assimilation of transformed retrievals

Could be viable assimilation options in some cases. Relevant questions:

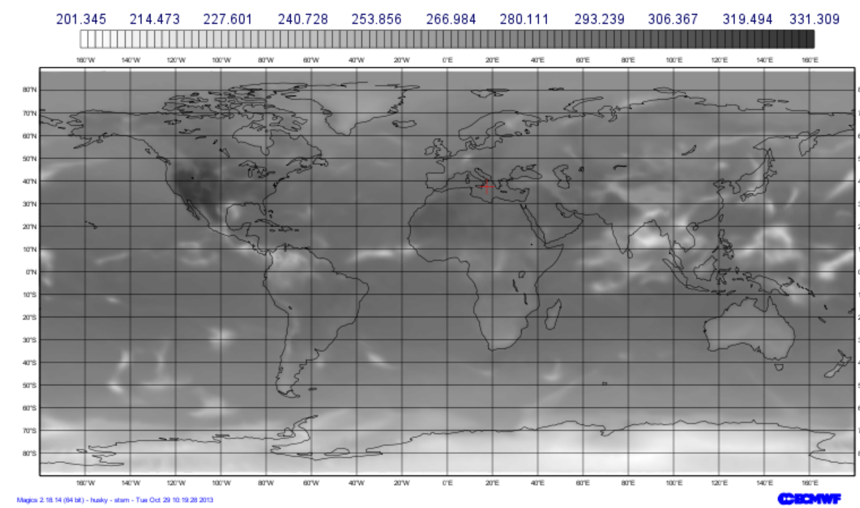
- Can we simplify the retrieval error structure?
- Is it possible to avoid introducing synoptic biases?
- If we first determine a retrieval from a set of radiances and (possibly) some prior information and then we assimilate an appropriately transformed retrieval, can we find the same estimate we would find when we directly assimilate the same set of radiances? Equivalence conjecture.
- Can we reduce considerably the measurements components we want to assimilate and preserve most of their information?
- Can we determine transformed retrievals that can be assimilated in any data assimilation system?
- Which and how many vertical levels for the transformed jacobians?

Transformed retrievals: numerical simulations

- Selection of a clear column over sea



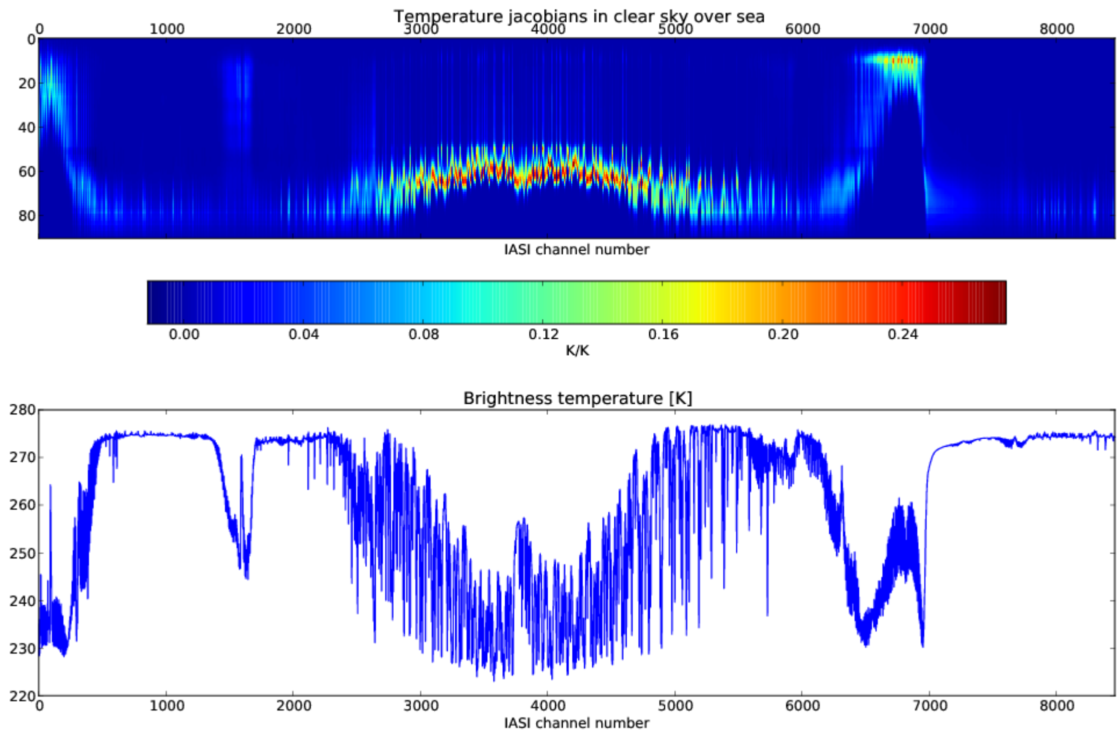
Cumulative cloud fraction



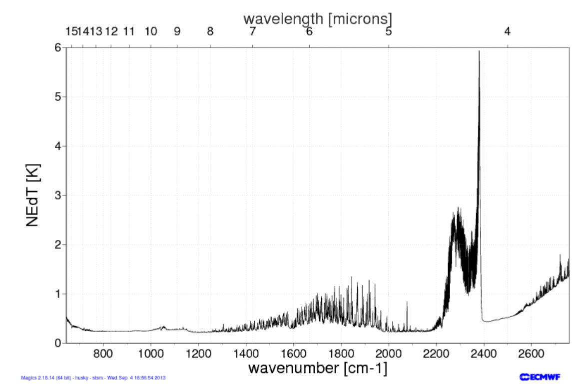
Tb at 921 (window channel) and clear sky column

IASI full spectrum simulations

IASI temperature jacobians in clear sky over sea (RTTOV 11)

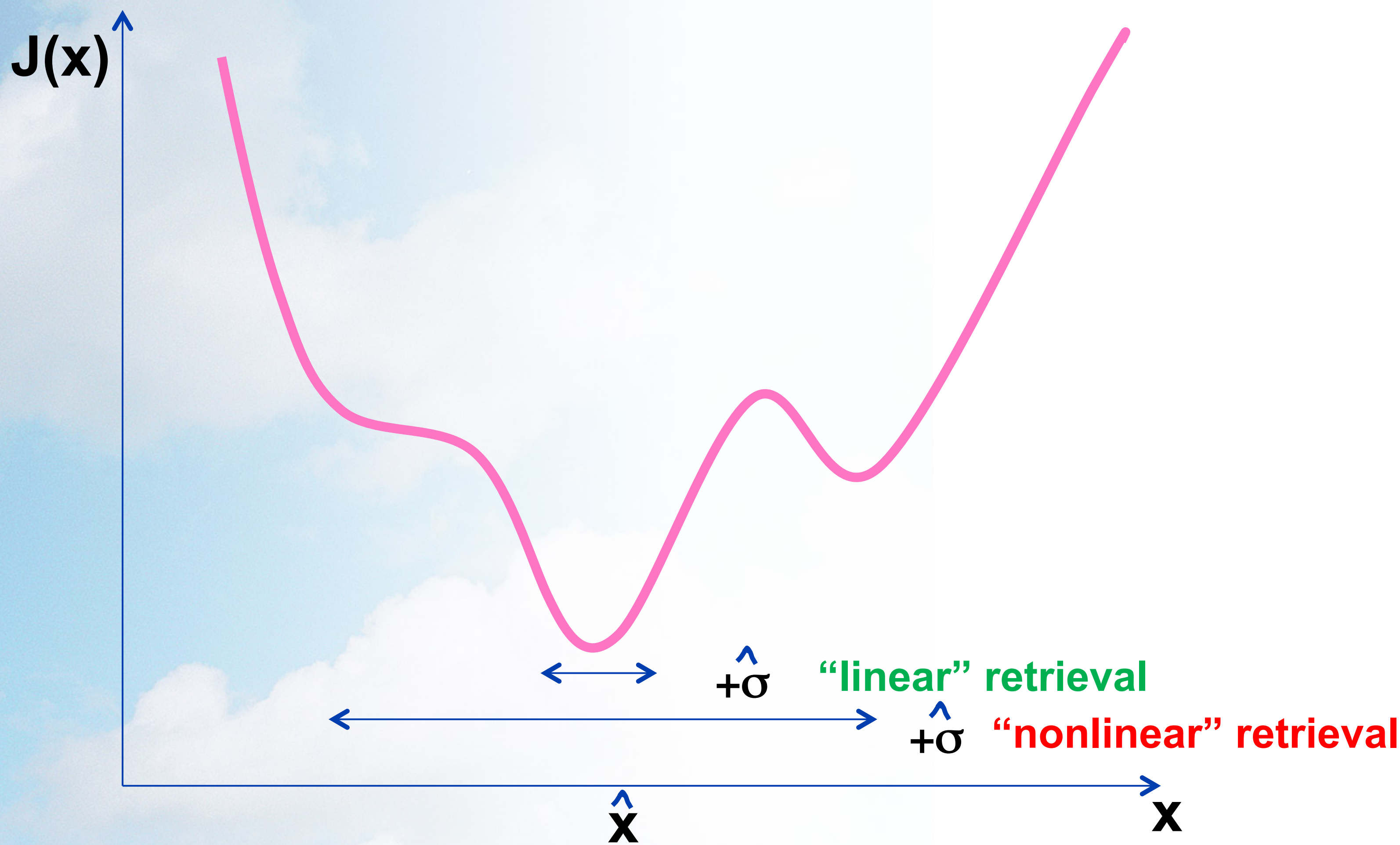


IASI observation error standard deviation



Moderate nonlinearity assumption

- We assume approximate linearity of observation operator
- Validity of assumption depends on degree of nonlin and prior uncertainty



Overdetermined or maximum-likelihood (ML) case

- When the number of measurements *m* is much larger than the dimension *n* of the subset of the state space that is relevant to predict the measurements ("the state"), a possible data compression strategy could be that of performing a *retrieval*.
- Perfectly viable solution if the inverse problem overdetermined, i.e., not ill-posed and linearizable as

$$\mathbf{H}' \equiv \mathbf{R}^{-1/2}\mathbf{H} \qquad \varepsilon' \equiv \mathbf{R}^{-1/2}\varepsilon$$

$$\mathbf{y}^o \cong \mathbf{H}\mathbf{x}^t + \varepsilon^o \qquad \mathrm{cov}(\varepsilon^o) = \mathbf{R} \qquad \mathbf{y}' \cong \mathbf{H}'\mathbf{x}^t + \varepsilon' \qquad \mathrm{cov}(\varepsilon') = \mathbf{I}$$

$$\hat{\mathbf{x}} = \left(\mathbf{H}'^T \mathbf{H}'\right)^{-1} \mathbf{H}'^T \mathbf{y}' \qquad \hat{\mathbf{x}} \cong \mathbf{x}^t + \left(\mathbf{H}'^T \mathbf{H}'\right)^{-1} \mathbf{H}'^T \varepsilon' \qquad \mathrm{cov}\left(\left(\mathbf{H}'^T \mathbf{H}'\right)^{-1} \mathbf{H}'^T \varepsilon'\right) = \left(\mathbf{H}'^T \mathbf{H}'\right)^{-1}$$

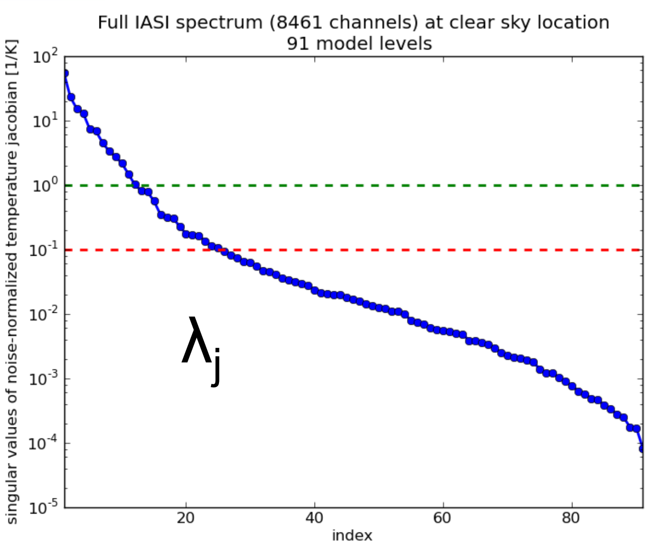
- The retrieval is a direct (lossless!) estimate of the state, and we have reduced the data by about an order of magnitude (from m=1000s to n=100s, e.g. for T)
- Sounders are in this case like "poor-quality radiosondes" (J. Eyre), with an *interesting* error characterization (e.g., vertical correlations) that may also provide undesirable effects (ill-conditioning)

Ill-conditioning and data reduction in the ML case

- Causes of ill-conditioning can be exposed through SVD of **H'** (m > n): $\mathbf{H}' = \mathbf{U}_n \Lambda_n \mathbf{V}_n^T$

$$\mathrm{cov}(\hat{\mathbf{x}} - \mathbf{x}^t) = \left(\mathbf{H}'^T \mathbf{H}'\right)^{-1} = \mathbf{V}_n \Lambda_n^{-2} \mathbf{V}_n^T$$

- Small errors in radiances providing very small constraint to the retrieval over a given region of the state space cause large local variations of the retrieval
- Ill-conditioning can be avoided if we consider new set of "measurements" given by a transformed retrieval with errors not too small
- Now natural way to reduce the data is to discard measurements that provide no or too small constraint to the estimate
- Maximum of n < m components matter as λ_j=0 for j > n



Transforming a retrieval for assimilation

- From m = 8461 to n = 91: only 1% of components provide estimate on 91 model levels (e.g., for temperature). Trivial error structure (cov(**y**')) is now unit (rank-n) matrix)

$$\mathbf{y}'_{\text{ret}} \equiv \Lambda_n \mathbf{V}_n^T \hat{\mathbf{x}} = \mathbf{U}_n^T \mathbf{y}' = \Lambda_n \mathbf{V}_n^T \mathbf{x}^t + \mathbf{U}_n^T \varepsilon' \equiv \mathbf{H}'_{\text{ret}} \mathbf{x}^t + \varepsilon''$$

- But assimilation of lossless estimate requires n + n² = n(n+1) components (8372 for full IASI spectrum!) as there is need to provide observation operator
- What happens if $\mathbf{y}'_{\text{ret}}$ is "assimilated" with $\mathbf{H}'_{\text{ret}}$ e.g. using ML method?

$$\hat{\mathbf{x}}_{\text{ret}} = \left(\mathbf{H}'_{\text{ret}}\right)^{-1} \mathbf{y}'_{\text{ret}} = \left(\mathbf{H}'_{\text{ret}}\right)^{-1} \Lambda_n \mathbf{V}_n^T \hat{\mathbf{x}} = \hat{\mathbf{x}}$$

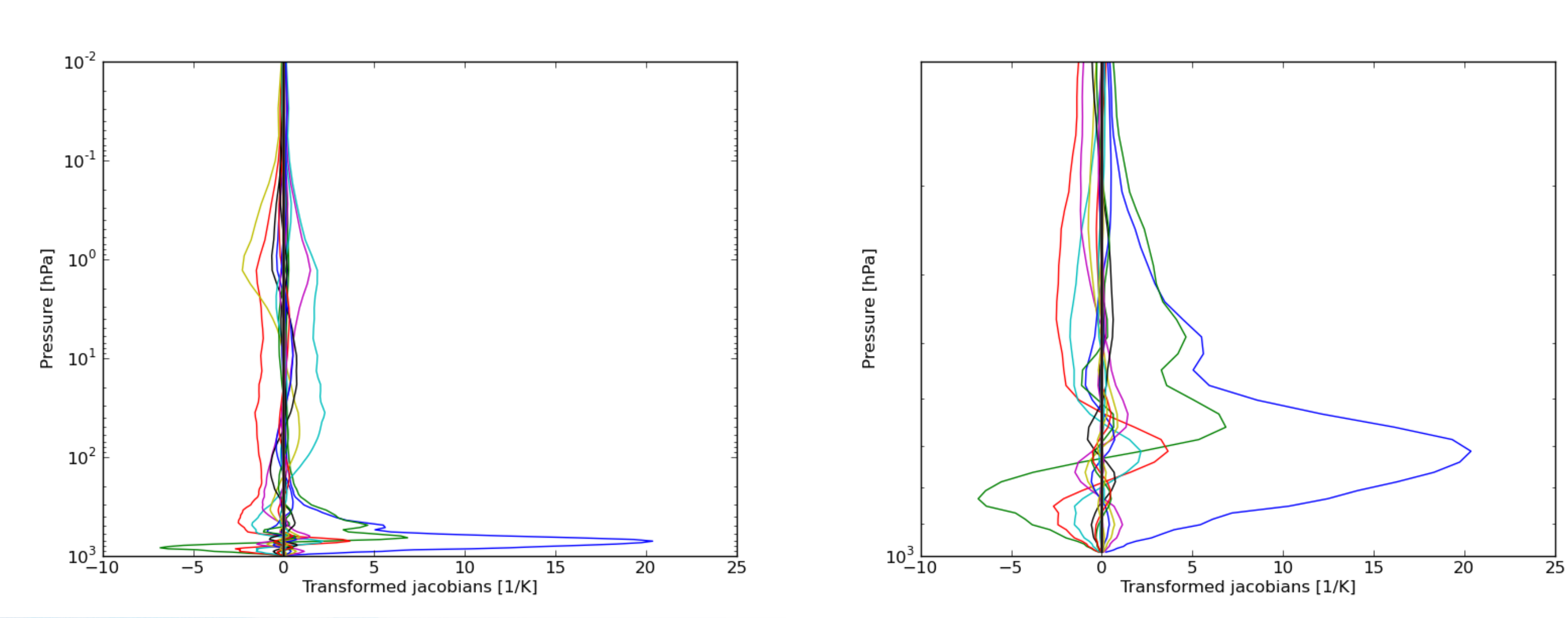
- Equivalence is achieved, but only when same inversion method is used both for retrieval and assimilation
- $\hat{\mathbf{x}}_{\text{ret}}$ is still a direct estimate of the state

$$\hat{\mathbf{x}}_{\text{ret}} = \mathbf{x}^t + \left(\mathbf{H}'_{\text{ret}}\right)^{-1} \varepsilon' = \mathbf{x}^t + \mathbf{V}_n \Lambda_n^{-1} \mathbf{U}_n^T \varepsilon'$$

- Estimate is still ill-conditioned if components with too small λ_i are retained

Transformed jacobians: ML case

- 91 rows of $\mathbf{H}'_{\mathrm{ret}} \equiv \Lambda_n \mathbf{V}_n^T$

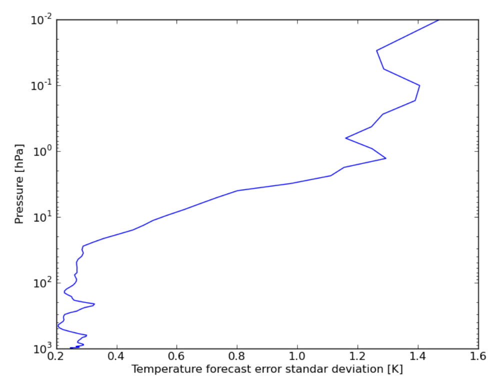


Ill-conditioning and data reduction in the ML case (Cont)

- To avoid numerical problems (ill-conditioning) we need to discard components of $\mathbf{y}'_{ret}$ corresponding to small singular values of $\mathbf{H}'_{ret}$
- Arbitrary data reduction threshold: by ordering components according to λ_j there is no guarantee to choose components with largest SNR

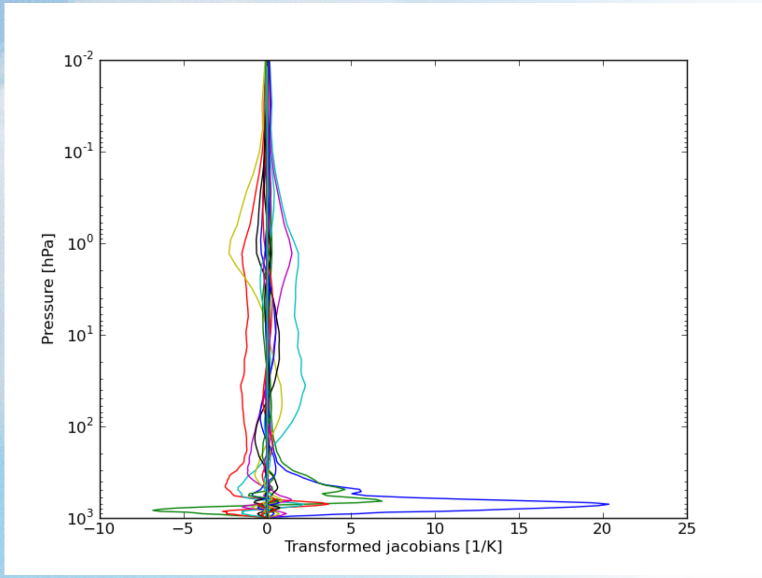
$$\mathrm{cov}(\mathbf{y}'_{\mathrm{ret}}) = \Lambda_n \mathbf{V}_n^T \mathbf{B} \mathbf{V}_n \Lambda_n + \mathbf{I}_n$$

- Average σ_b for T in tropospheric column: ~ 0.26 K (0.44 K overall)
- Ad hoc criterion: $\lambda_j \langle \sigma_b \rangle > \sim 1$ → 14 components with $\lambda_i > \max(\sigma_b) = 0.44$ K

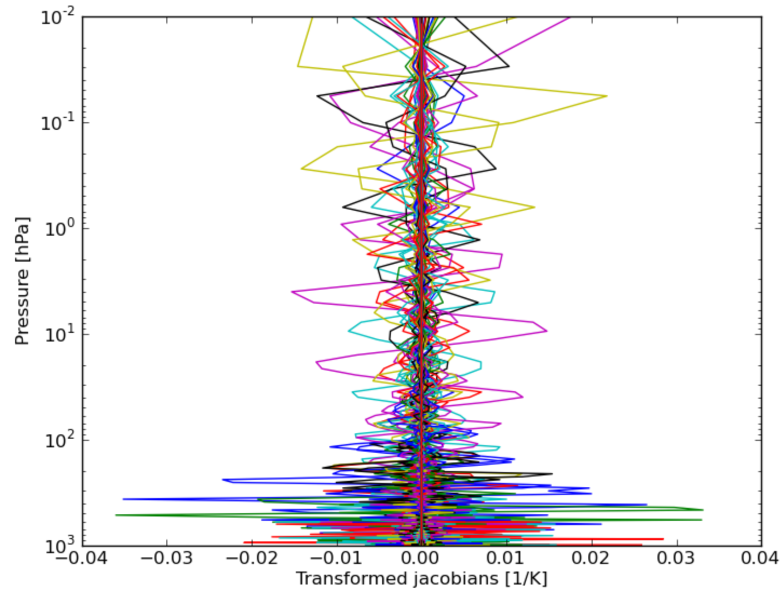


- Only 7 components with $\lambda_i > 4$, 12 with $\lambda_i > 1$ and 25 with $\lambda_i > 0.1$ (still only 0.2% data retention!)

Transformed ML jacobians: reduced set



Components with 14 largest sing val



“residual” components

Effects of lossy data reduction

- When using $\mathbf{y}'_{\text{ret}} \equiv \Lambda_r \mathbf{V}_r^T \hat{\mathbf{x}}$ with r < n measurement components how can I determine the n components of the retrieval in the "data assimilation system"?
- Underconstrained problem: n-r free variables (infinitely many solutions)
- In absence of prior information, we can choose the retrieval with minimum-length:

$$\hat{\mathbf{x}}_{\text{ret}} = \mathbf{H}'^{T}_{\text{ret}} \left(\mathbf{H}'_{\text{ret}} \mathbf{H}'^{T}_{\text{ret}} \right)^{-1} \mathbf{y}'_{\text{ret}}$$

- If we then assimilate r components of $\mathbf{y}'_{\text{ret}}$ with $\Lambda_r \mathbf{V}_r^T$

$$\hat{\mathbf{x}}_{\text{ret}} = \mathbf{H}'^{T}_{\text{ret}} \left(\mathbf{H}'_{\text{ret}} \mathbf{H}'^{T}_{\text{ret}} \right)^{-1} \mathbf{y}'_{\text{ret}} = \mathbf{V}_r \mathbf{V}_r^T \hat{\mathbf{x}} \neq \hat{\mathbf{x}}$$

- The minimum-length retrieval from r measurement components not equivalent to overdetermined ML retrieval due to averaging kernel $\mathbf{V}_r \mathbf{V}_r^T \neq \mathbf{I}_n$
- This means the retrieval is not anymore a direct estimate of the state:

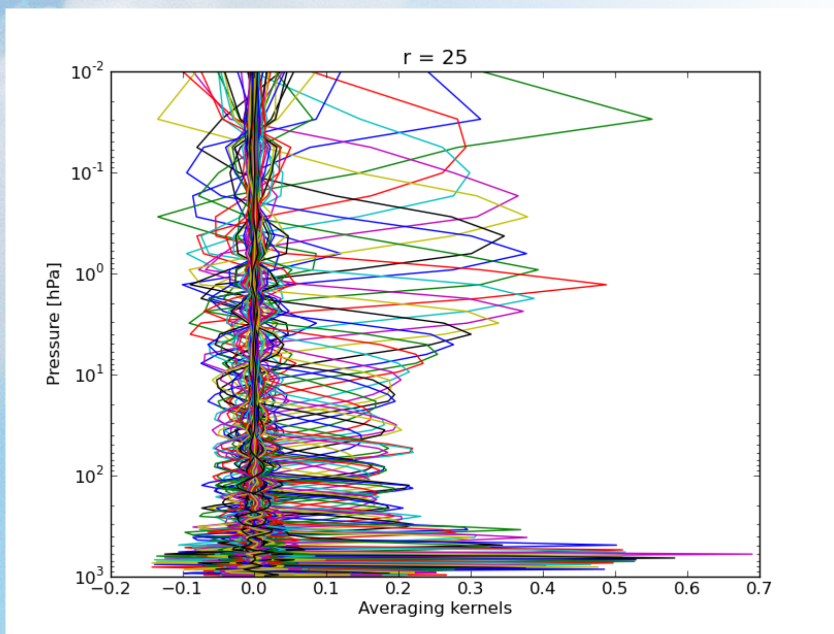
$$\hat{\mathbf{x}}_{\text{ret}} \cong \mathbf{V}_r \mathbf{V}_r \mathbf{x}^t + \mathbf{V}_r \Lambda_r^{-1} \mathbf{U}_r^T \varepsilon'$$

Reconstruction error

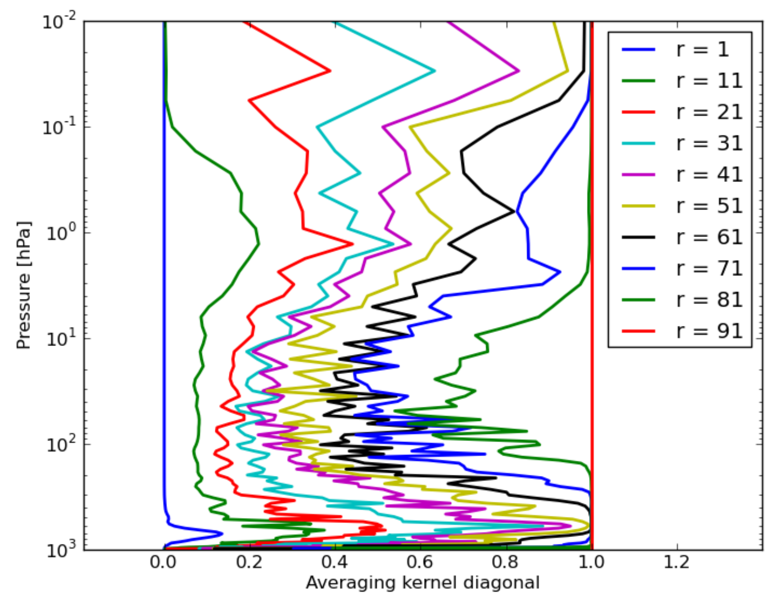
$$\hat{\mathbf{x}} - \hat{\mathbf{x}}_{\text{ret}} = (\mathbf{I}_n - \mathbf{V}_r\mathbf{V}_r^T)\hat{\mathbf{x}}$$

$$\left\|\hat{\mathbf{x}} - \hat{\mathbf{x}}_{\text{ret}}\right\|_2 = \left\|(\mathbf{I}_n - \mathbf{V}_r\mathbf{V}_r^T)\hat{\mathbf{x}}\right\|_2 \leq \left\|\mathbf{I}_n - \mathbf{V}_r\mathbf{V}_r^T\right\|_2 \left\|\hat{\mathbf{x}}\right\|_2$$

$$\mathbf{V}_r\mathbf{V}_r^T$$



$$diag(\mathbf{V}_r\mathbf{V}_r^T)$$



- Severe reconstruction error even when most components are retained
- But here comparison is with direct and likely ill-conditioned estimate

Assimilation of MAP retrievals with same prior information

- Now ill-conditioning kept under control through use of prior estimate

$$\hat{\mathbf{x}} = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}'\mathbf{x}^b) \qquad \mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}'\mathbf{B}\mathbf{H}'^T + \mathbf{I})^{-1}$$

$$\hat{\mathbf{x}} \cong \mathbf{x}^b + \mathbf{K}\mathbf{H}'(\mathbf{x}^t - \mathbf{x}^b) + \mathbf{K}\varepsilon^o$$

- As usual we assume that the observation operator for the retrieval is approximately linear around a neighbourhood of $\hat{\mathbf{x}}$ of radius comparable to the estimation error.
- The retrieval is then transformed using the matrix $\mathbf{U}_r$ of the left singular vectors of the signal-to-noise matrix **S**, with $\operatorname{rank}(\mathbf{S}) \equiv r \leq \min(m,n)$

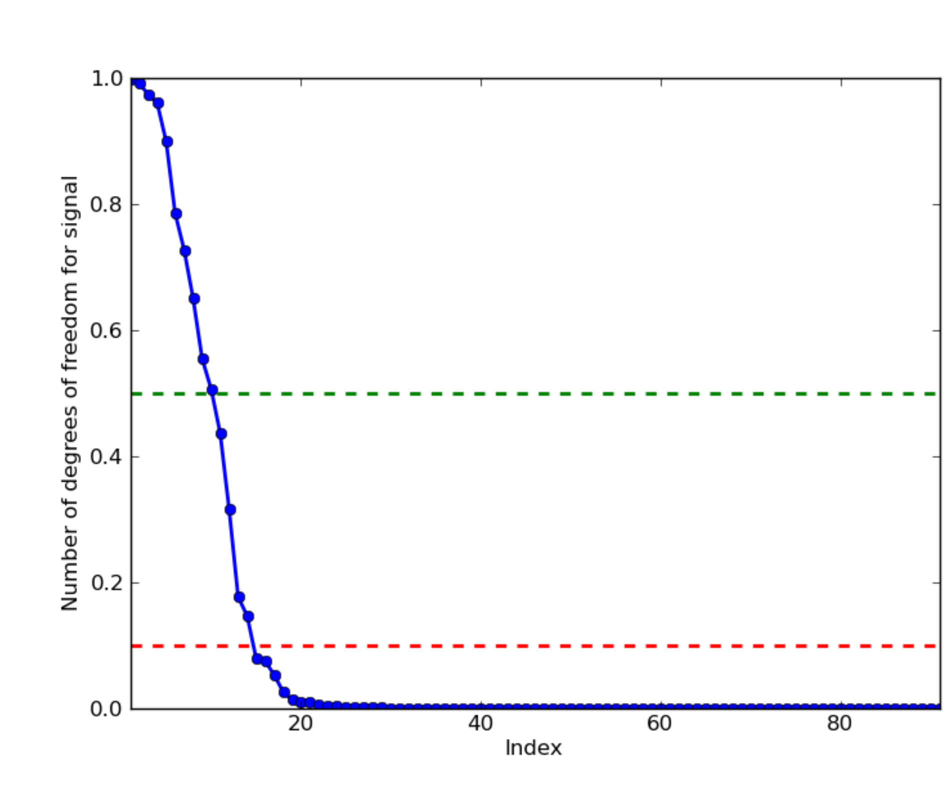
$$\mathbf{S} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{B}^{1/2} = \mathbf{H}'\mathbf{B}^{1/2} = \mathbf{U}\Lambda\mathbf{V}^T$$

$$\mathbf{y}_{ret} = \mathbf{U}_r^T\mathbf{R}^{-1/2}\mathbf{y}^o \qquad \mathbf{H}_{ret} = \mathbf{U}_r^T\mathbf{H}' = \Lambda\mathbf{V}^T\mathbf{B}^{-1/2}$$

- The transformed retrieval is then assimilated in a NWP model with the same prior estimate. It can be shown that the result is equivalent to assimilating the original radiances.

Data reduction strategy: degrees of freedom for signal

- We can calculate $d_{sj} = \lambda_j^2/(1+\lambda_j^2)$ degrees of freedom for signal (DFS)
- Only 10 comp with DSF > 0.5 ($\lambda_j > 1$) and 14 with DSF > 0.1

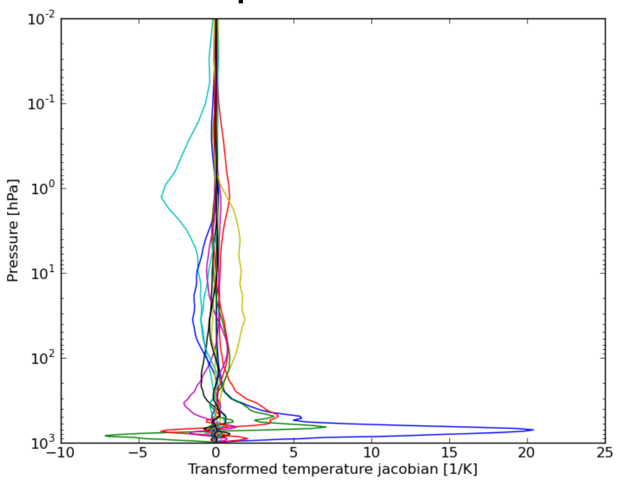


- Data reduction potential of 99.8% (DSF > 0.1) when only interested in T

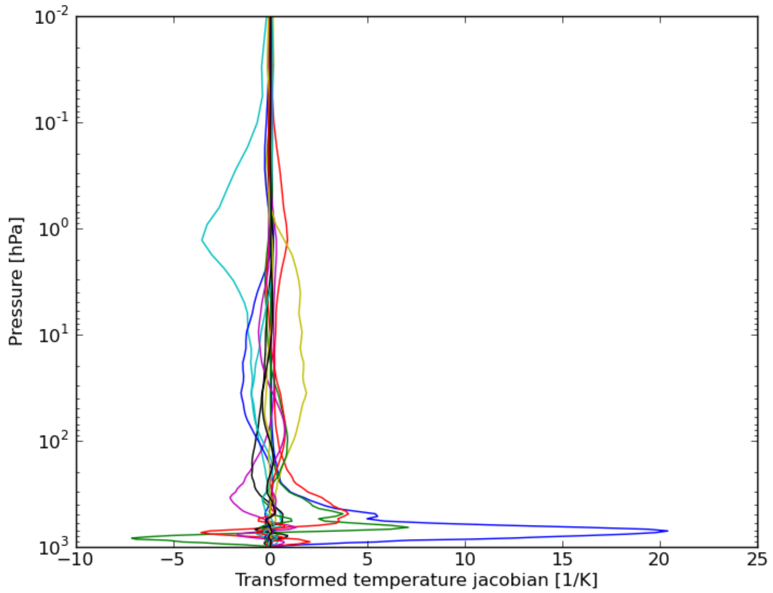
Transformed MAP jacobians

- Rows of $\mathbf{H}_{ret} = \mathbf{U}_r^T \mathbf{H}' = \Lambda \mathbf{V}^T \mathbf{B}^{-1/2}$

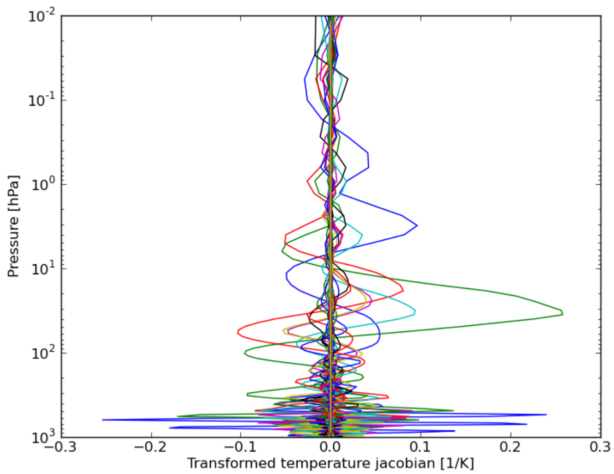
All components



First 14 most informative components



Least informative 91-14=77 components



Uncertainty reduction profiles: DFS weighting functions

$$\hat{\mathbf{x}} \cong \mathbf{x}^b + \mathbf{A}(\mathbf{x}^t - \mathbf{x}^b) + \mathbf{K}\varepsilon^o$$

$$\mathbf{A} = \mathbf{B}^{1/2}\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{B}^{-1/2} \qquad \mathbf{D} = diag\{d_{sj} = \lambda_j^2/(1+\lambda_j^2)\}$$

- Reduction of uncertainty can be quantified with number of DFS = d_s

$$d_s = tr(\mathbf{A}) = tr\left(\mathbf{I}_n - \mathbf{P}^a\mathbf{B}^{-1}\right) = \sum_{j=1}^{r} d_{sj} tr\left(\mathbf{B}^{1/2}\mathbf{v}_j\mathbf{v}_j^T\mathbf{B}^{-1/2}\right) = \sum_{j=1}^{r} d_{sj} tr\left(\mathbf{v}_j\mathbf{v}_j^T\right) = \sum_{j=1}^{r} d_{sj}\mathbf{v}_j^T\mathbf{v}_j$$

- Trivial identity, but we can now define the DFS weighting function for y_j" as

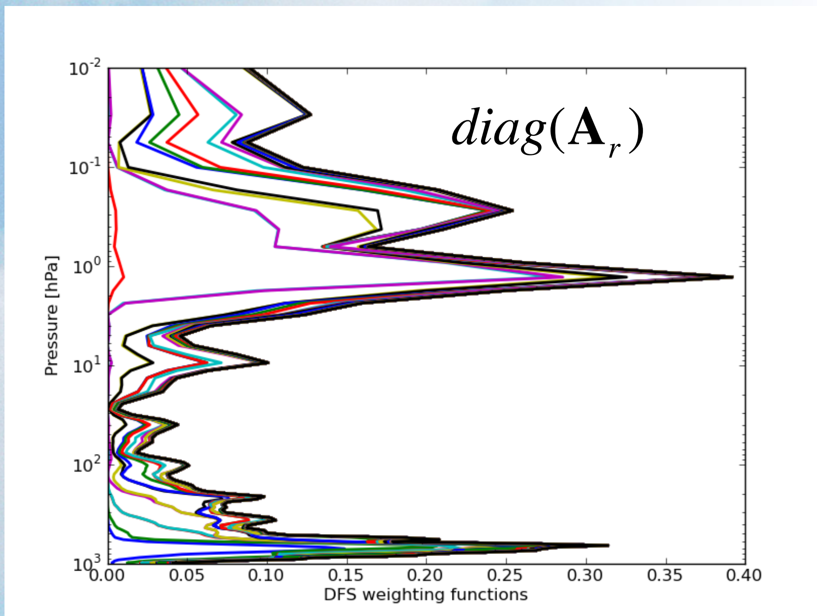
$$\mathbf{s}_j = \left(d_{sj}v_{1j}^2, \cdots, d_{sj}v_{nj}^2\right) = d_{sj}\mathbf{v}_j \circ \mathbf{v}_j$$

- And the cumulative DFS weighting function as

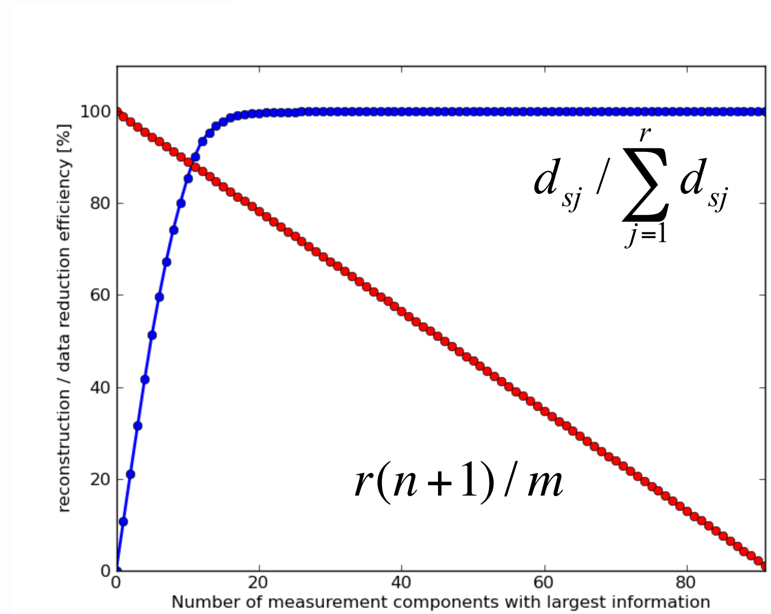
$$\mathbf{s} = \sum_{j=1}^{r}\mathbf{s}_j = \left(\sum_{j=1}^{r} d_{sj}v_{1j}^2, \cdots, \sum_{j=1}^{r} d_{sj}v_{nj}^2\right) = \sum_{j=1}^{r} d_{sj}\mathbf{v}_j \circ \mathbf{v}_j$$

Reconstruction efficiency in the MAP case

DFS weighting functions for T



Reconstruction / data reduction efficiency



- d_{sj} is the sum of elements of DFS weighting function $d_s \equiv \sum_{j=1}^{r} d_{sj} \cong 9.4$
- We just need up to 20 components to have almost complete information conservation
- 10 components with signal above $1\sigma_{noise}$ retain 85.5% information and have 1 - r(n+1)/m = 1 - 920/8461 = 89.1 % data reduction efficiency (11 or 12 components to maximize both)

Information sensitivity on choice of prior estimate

- First consider to retrieve with prior error cov **B** and to assimilate with $\mathbf{B}_{DAS}=\mathbf{B}$
- Information of original measurements **y'** given by the logarithm of det(cov(**y'**)) $= \det(\mathbf{H'BH'}^T + \mathbf{I}_m) = 1/2\sum_{j=1}^{r}\log(1+\lambda_j^2) \equiv h$
- In the DAS we have information

$$\frac{1}{2}\log(\det(\mathrm{cov}(\mathbf{y}''))) = \frac{1}{2}\log(\det(\Lambda_r \mathbf{V}_r^T \mathbf{B}^{-1/2}\mathbf{B}_{DAS}\mathbf{B}^{-1/2}\mathbf{V}_r\Lambda_r + \mathbf{I}_r)) = \frac{1}{2}\log(\det(\Lambda_r^2 + \mathbf{I}_r)) = 1/2\sum_{j=1}^{r}\log(1+\lambda_j^2)$$

- As expected, same information in retrieval and assimilation
- But if we retrieve with $\mathbf{B} \rightarrow \alpha^2\mathbf{B}$ with $\alpha^2<1$ and $\mathbf{B}_{DAS}=\mathbf{B}$, information both in retrieval system and DAS given by ($\lambda \rightarrow \alpha\lambda$)

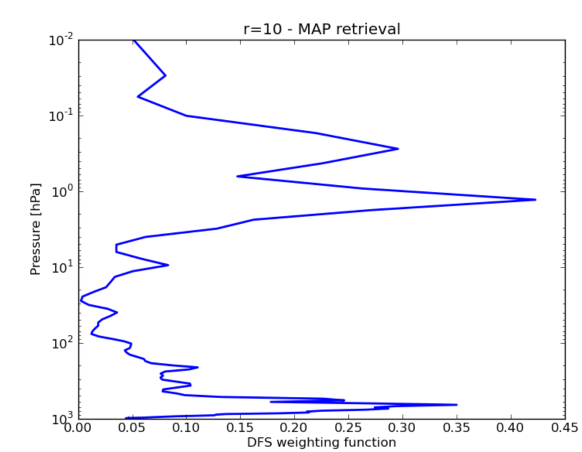
$$h' = 1/2\sum_{j=1}^{r}\log(1+\alpha^2\lambda_j^2) < h$$

- Information loss in retrieval system results in information loss also in DAS
- Safer to reduce data with climatological **B**. But too little constraint can cause stability problems and local-only minimization

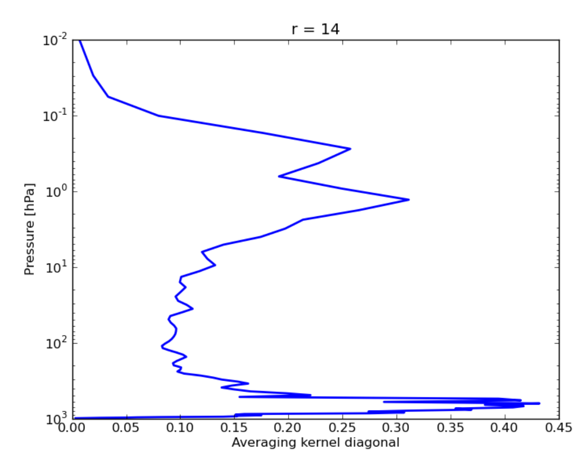
Data reduction and information

- MAP methods allow efficient information retention with as fewer measurement components as possible. See, e.g., mesosphere where ML is blind for same number of components (r=10) as MAP
- MAP gives more relative importance to upper stratosphere as σ_b is large
- ML needs more components to preserve same or less information: ML (MAP) needs 14 (10) components to retain signal above $1\sigma_{noise}$
- But ML may be appropriate when retrieval prior not well known

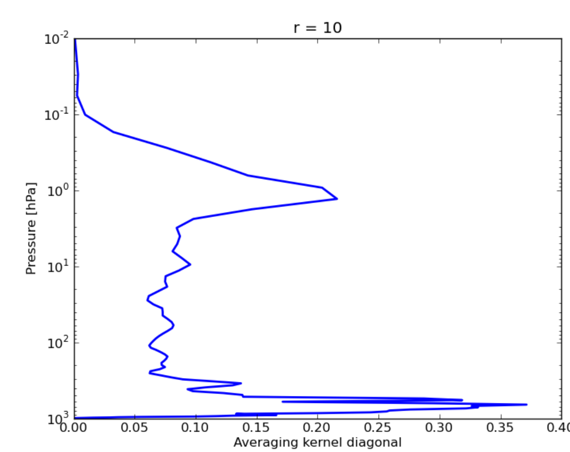
r=10 DFS weighting function (MAP)



r=14 DF weighting function (ML)



r=10 DF weighting function (ML)



Summary and conclusions

- Under certain conditions it is possible to transform a retrieval to achieve equivalence with radiance assimilation
- Required conditions are linearizability of observation operator and use of same prior information for both retrieval and assimilation
- When the number of measurements is significantly larger than the dimension of the state of the system, it is possible to reduce the data while retaining most of the information to estimate the state
- In ML case data reduction improves conditioning of estimate
- Retrievals that make use of a prior estimate (MAP retrievals) can be transformed in a way to retain more information per measurement component
- When prior estimate not well known it is best to determine MAP retrieval with less stringent prior as this allows more information from measurements to be used in the DAS, where a more reliable prior estimate is available
- Efficient data reduction and information retention strategy is best achieved when it is tailored to specific characteristics of the system where reduced dataset is to be used