AIRS L2 Product Error Characterization: Propagation of Internal Error Estimates

Eric Maddy, Chris Barnet

04/27/2011

Background/Motivation

- AIRS V5+ retrieval files include QC error estimates.
  - Error estimates are derived using a statistical regression between AIRS quality parameters and differences between ECMWF and AIRS L2 products.
  - QC error estimates are derived at the end of the retrieval and are therefore not optimally used within the L2 software/algorithms.
- The current algorithm lacks the ability to accurately calculate and propagate internal error estimates-important as these are used to stabilize retrievals
  - Discussed by George (context of cloud-clearing) yesterday.
  - Initially shown by Barnet (2004, 2005…), detailed analysis of CCRs in 2007-8 timeframe.
  - Ensemble error estimates (static first guess error) and null estimates (retrieval error floor) files include no information on case-dependent magnitude of errors or vertical or inter-product correlations - need revisiting
- Averaging kernels and error estimates are inter-related.
  - Our ability to accurately characterize our L2 retrieval's dependence on the FG (a priori) state is directly dependent on our ability to characterize errors within the algorithm.
- NOAA is funded through ROSES 2009 to enable propagation of errors within the algorithm. Work should lead to:
  - More accurate products-needs to be demonstrated.
  - Better product quality assessment- better averaging kernels and error bars .

Objectives

1) Provide characterization of end-to-end AIRS retrieval system errors.
   - Requires knowledge of input errors (either regression, climatology, NN, NWP?) as well as optimization of retrieval algorithm constraints.
   - Requires estimation of CCR errors – not trivial.
   - Work with modeling groups to best define product output formats (error covariance, averaging kernels, etc.).
2) Improve spatio-temporal stability and climate quality of the AIRS retrieval with respect to trace gas initialization (CO_2, CO, CH_4, etc.)
   - Update and maintain existing trace gas climatologies.
   - For CO_2 we currently are using a linear trend – test other approaches.
3) Modify down-stream algorithms (e.g., O_3, CO, CH_4) to take advantage of 1) and 2)

ERROR ESTIMATES – THE EASIER PART

Averaging Kernels and Error Estimates

- Averaging kernels represent the response of a retrieval to some perturbation in the true atmospheric state and also the dependence of the retrieval on first guess assumptions.
- For a given iteration, we can linearize our retrieval equation to enable an estimation of the bias in the retrieved state.

$$\hat{x} - x \quad = \quad \underset{Smoothing}{(\mathbf{A} - \mathbf{I})(x - x_a)} \quad + \quad \underset{Interference}{\mathbf{G}\mathbf{K}_b(b - b_a)} \quad + \quad \underset{Instrument\,/\,ForwardModel}{\mathbf{G}\varepsilon}$$

Averaging Kernel

Retrieval Operator/Gain Matrix

Jacobians due to Background Parameters

Averaging Kernels and Error Estimates

- Averaging kernels represent the response of a retrieval to some perturbation in the true atmospheric state and also the dependence of the retrieval on first guess assumptions.
- For a given iteration, we can linearize our retrieval equation to enable an estimation of the bias in the retrieved state.

$$\underset{\text{Bias in our retrieval}}{\hat{x} - x} = \underbrace{(\mathbf{A}-\mathbf{I})(x - x_a)}_{Smoothing} + \underbrace{\mathbf{G}\mathbf{K}_b(b - b_a)}_{Interference} + \underbrace{\mathbf{G}\varepsilon}_{Instrument\,/\,ForwardModel}$$

Bias in our retrieval

Bias in our FG

Bias in things held fixed in a given step

Can our algorithm propagate a linear error estimate?

- Following 2 slides show examples of predicted propagated error in CLEAR simulation using *ad-hoc Markov* covariance matrices:

$$S_{i,j} = \sigma_i \sigma_j \exp\left(-\frac{|i-j|}{h}\right)$$

- Given initial biases in the state (*e.g.* from regression, climatology, NN, etc.), background parameters (*i.e*, things held fixed in a given step) and instrument and forward model noise estimates can we analyze the propagation of bias through our algorithm

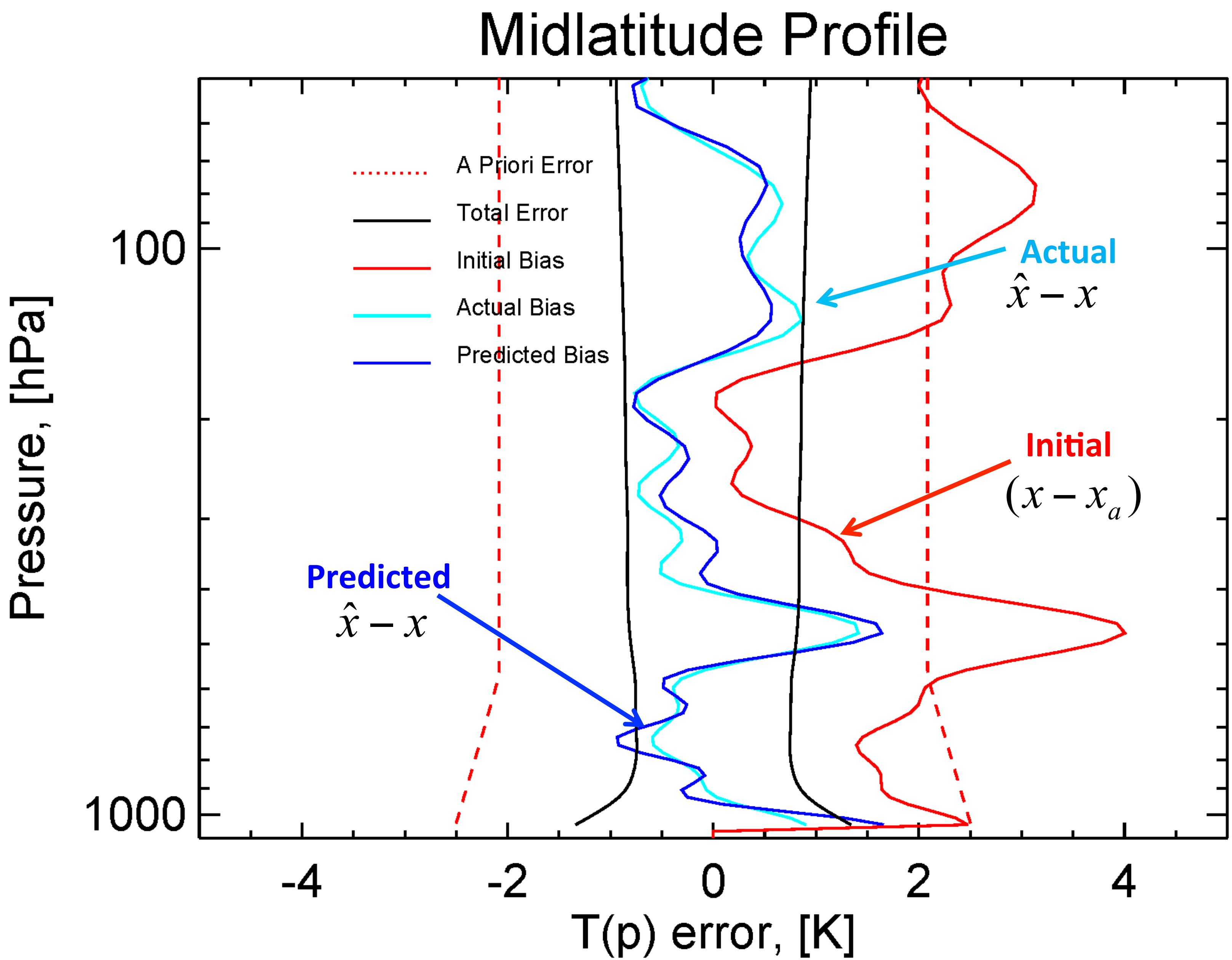
$$\hat{x} - x = \underset{Smoothing}{(\mathbf{A}-\mathbf{I})(x - x_a)} + \underset{Interference}{\mathbf{G}\mathbf{K}_b(b - b_a)} + \underset{Instrument\,/\,ForwardModel}{\mathbf{G}\varepsilon}$$

Bias in our retrieval

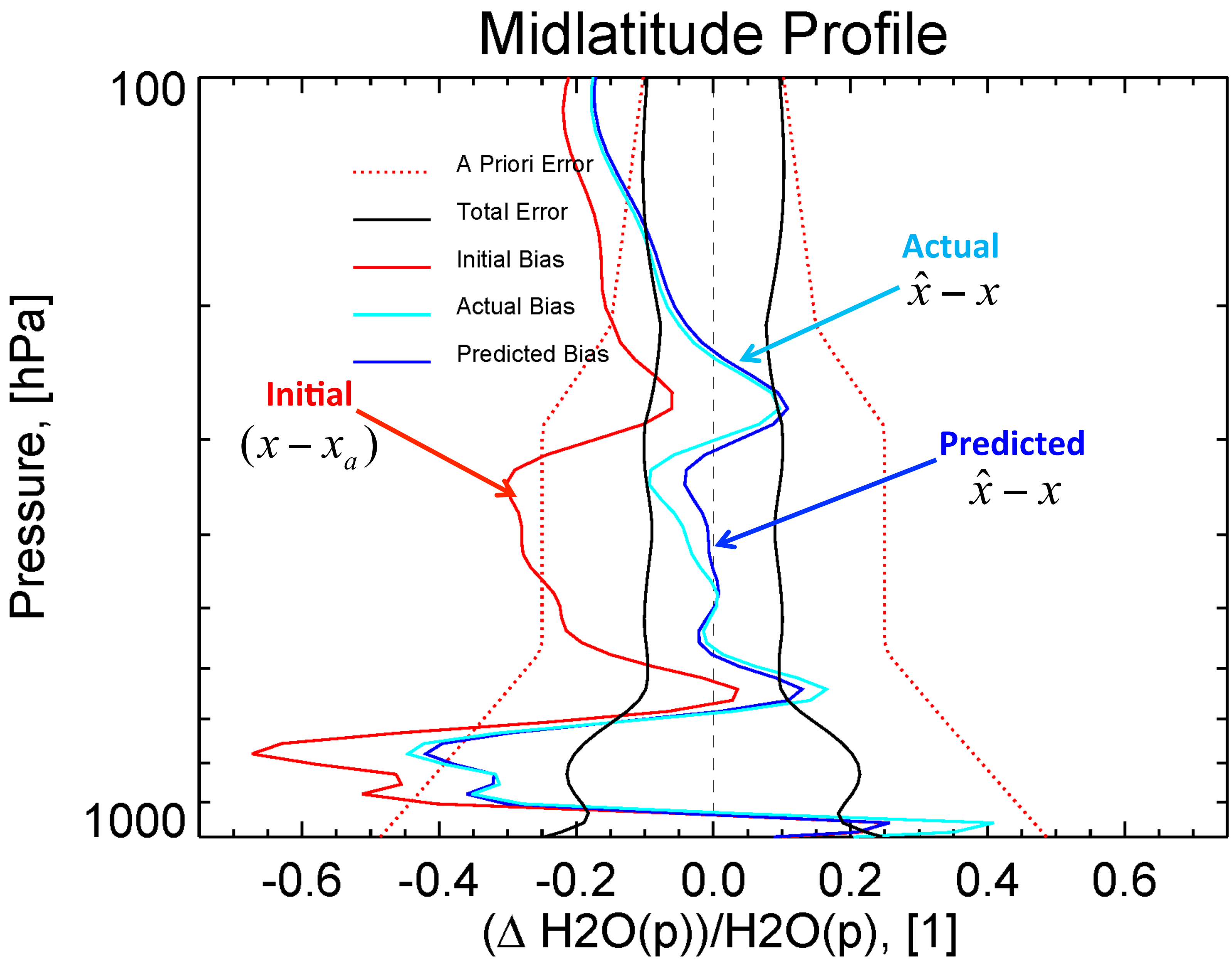
Bias in our FG

Bias in things held fixed in a given step

Example estimation of T(p) retrieval bias for a clear-sky scene



Example estimation of H_2O(p) retrieval bias for a clear-sky scene



Discussion

- Previous example showed that we are able to estimate the bias in a simulated clear-sky temperature and water retrieval using straightforward error propagation techniques.
- Results are promising but:
  - initial error is based on random perturbations eigenvectors of S_{ij}. Estimated smoothing error is well constrained by assumed *a priori* covariance. Won't be true for a system that uses regressions.
  - clouding-clearing errors that are more difficult to predict and propagate and are not included in this analysis.
- In practice we cannot know the magnitude and sign of the initial state bias, background bias, and bias contributions due to instrument and forward model uncertainties.
- The best that we can do is use a statistical estimate of the uncertainty (covariance) in the various inputs to the algorithm.

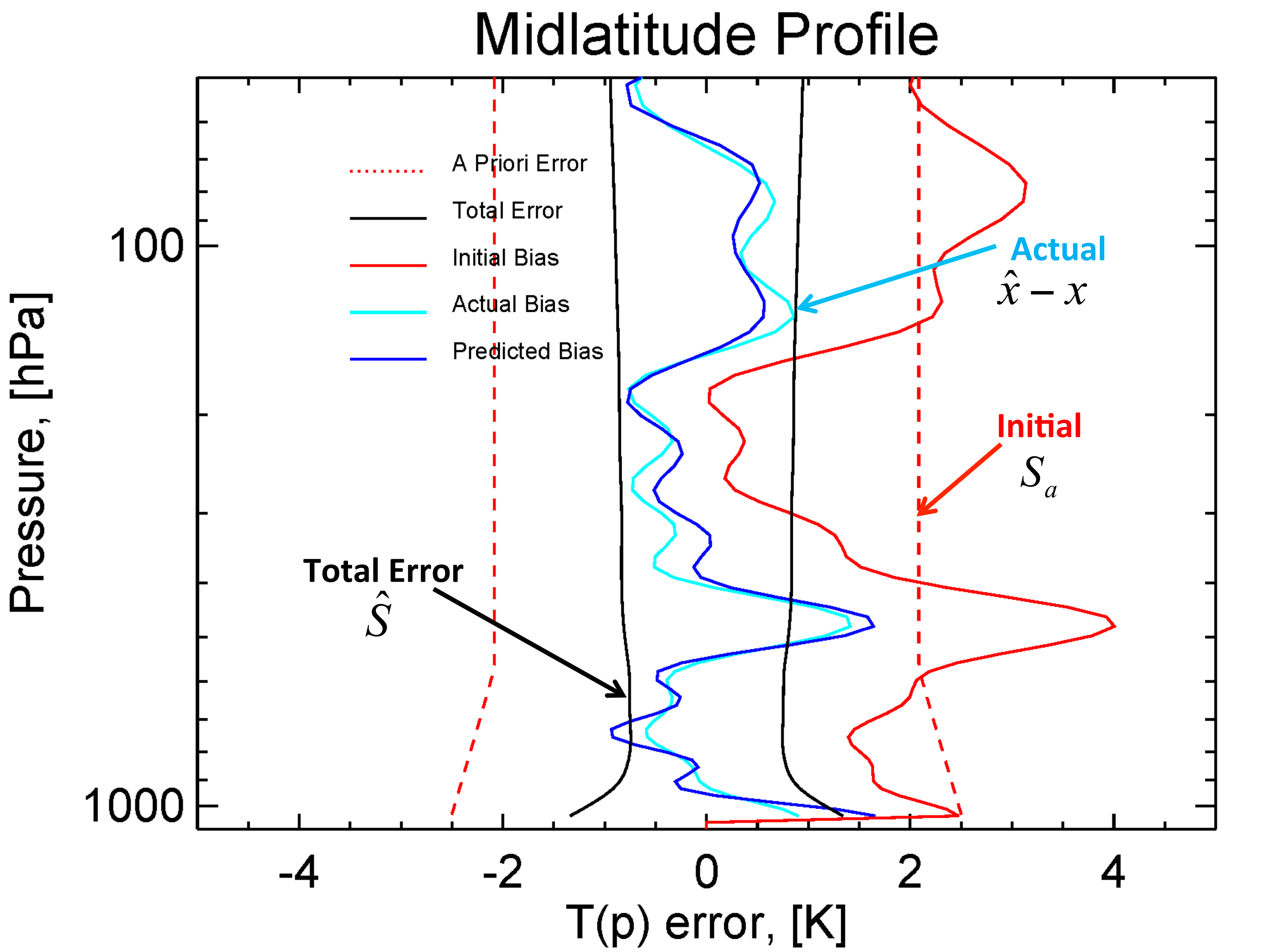
$$\hat{\mathbf{S}} = \underbrace{(\mathbf{A}-\mathbf{I})\mathbf{S}_a(\mathbf{A}-\mathbf{I})^T}_{Smoothing} + \underbrace{\mathbf{G}\mathbf{K}_b\mathbf{S}_b(\mathbf{G}\mathbf{K}_b)^T}_{Interference} + \underbrace{\mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}^T}_{Instrument\,/\,ForwardModel}$$

What we need to propagate an error estimate (clear-sky)

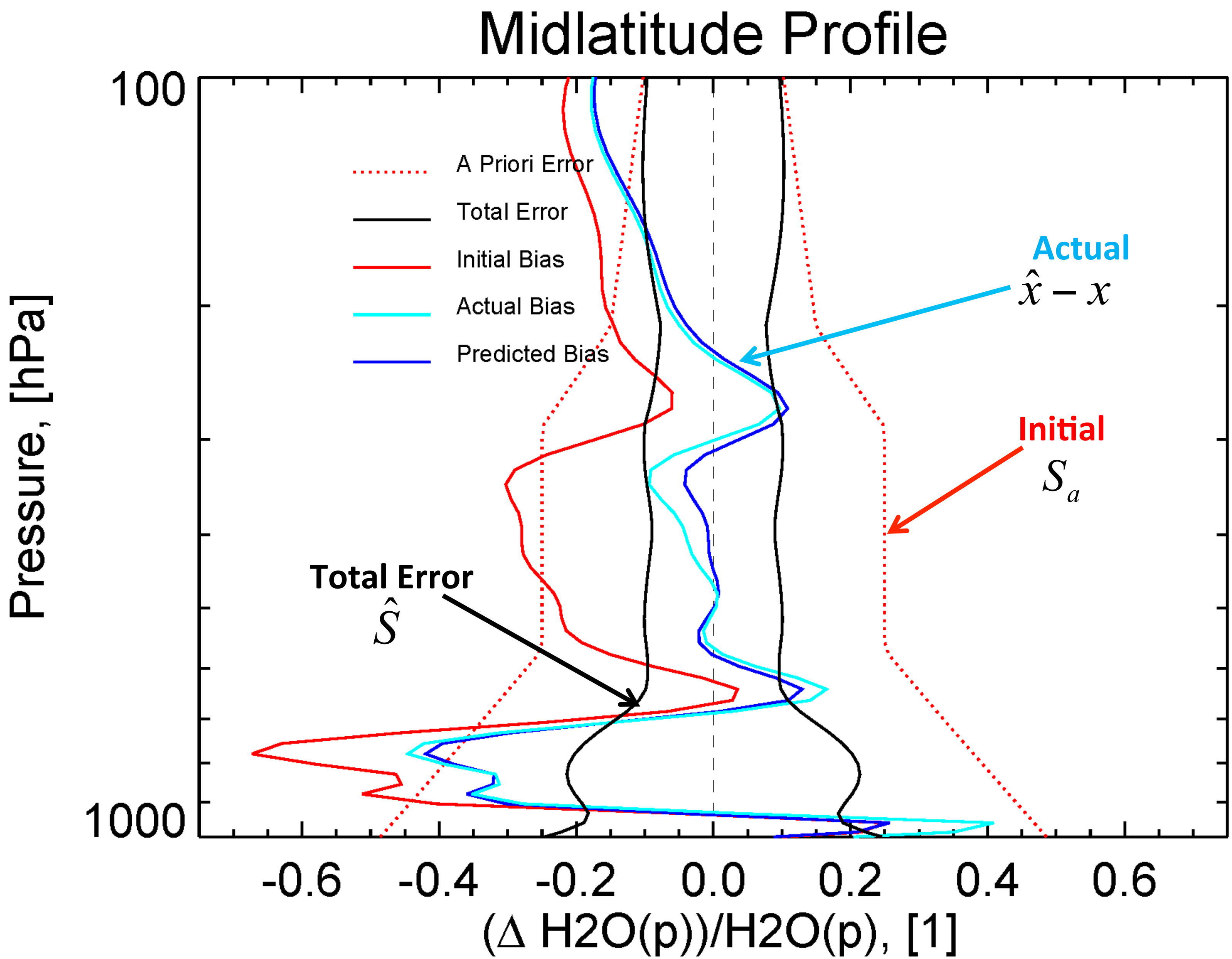
- $\mathbf{S}_a$ - initial startup state (regression, climatology, SCCNN, NWP?) covariance
- $\mathbf{S}_b$– Things held fixed covariance (*e.g.*, $\delta H_2O(p)\,(\delta H_2O(p))^T$, $\delta O_3(p)\,(\delta O_3(p))^T$,… in T(p))
- $\mathbf{K}_b$ – derivatives with respect to things held fixed.
- $\mathbf{S}_\varepsilon$– estimate of instrument error

$$\hat{\mathrm{S}} = \underbrace{(\mathbf{A}-\mathbf{I})\mathbf{S}_a(\mathbf{A}-\mathbf{I})^T}_{Smoothing} + \underbrace{\mathbf{G}\mathbf{K}_b\mathbf{S}_b(\mathbf{G}\mathbf{K}_b)^T}_{Interference} + \underbrace{\mathbf{G}\mathbf{S}_\varepsilon\mathbf{G}^T}_{Instrument\,/\,Forward Model}$$

Example estimation of error estimates for a clear-sky scene



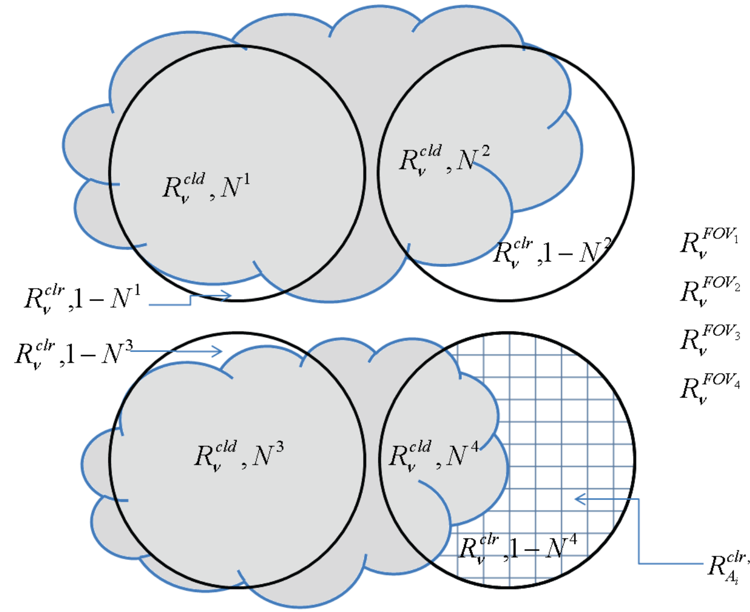
Example estimation of $H_2O(p)$ retrieval bias for a clear-sky scene



ERROR ESTIMATES – THE HARDER PART

Discussion of the Estimation of Uncertainty in our Cloud Corrections

- Previous examples show that for most cases initial errors within several K and 80% H_2O we should be able to propagate errors through our algorithm.
- Cloud corrections can be many 10's of K and therefore the uncertainties in cloud-cleared radiances can be much larger than those used in our simple example.
- Collocated imager (MODIS) measurements are an obvious choice for quality controlling and/or characterizing errors in cloud-cleared radiances.



$$R_\nu^{FOV_1} = R_\nu^{clr}(1-N^1) + R_\nu^{cld} N^1$$
$$R_\nu^{FOV_2} = R_\nu^{clr}(1-N^2) + R_\nu^{cld} N^2$$
$$R_\nu^{FOV_3} = R_\nu^{clr}(1-N^3) + R_\nu^{cld} N^3$$
$$R_\nu^{FOV_4} = R_\nu^{clr}(1-N^4) + R_\nu^{cld} N^4$$

$$R_{A_i}^{clr} = \sum_l IPSF_l R_{A_i}^{clr,l}$$
$$R_{A_i}^{FOV_j} = \sum_\nu SRF_{i,\nu} R_\nu^{FOV_j}$$

$$\longrightarrow \quad \eta(j,k) = \frac{\sum_i \frac{1}{\sigma_i}[R_{A_i}^{clr} - R_{A_i}^{FOV_j}][R_{A_i}^{FOV_j} - R_{A_i}^{FOV_k}]}{\sum_i \frac{1}{\sigma_i}[R_{A_i}^{FOV_j} - R_{A_i}^{FOV_k}]^2}$$

Above diagram is for IASI + AVHRR
We can use similar methodologies for AIRS + MODIS

Use of climatological covariance as a cloud-clearing constraint

- George, Evan and I have shown in the past that over the ocean (tropics especially) the surface leaving radiance is well constrained by the climatological surface temperatures
  - We can use this property as a constraint to the cloud-cleared radiances
  - Questions about what do we do over land? At higher latitudes?

Work Plan (over the next year)

- Work with JPL on the estimation (calculation) of system startup biases and error covariance matrices (may already exist for climatology system).
  - Replace static ensemble error estimates and retrieval null estimates with more realistic error covariances (lat,lon,time,etc.).
- Install and provide a demonstration of the error propagation routines in the NOAA offline system
  - NOAA gridded datasets, NOAA radiosondes, etc. used to characterize and “validate” outputs.
- Test and demonstrate our ability to estimate and propagate cloud-clearing uncertainties
  - use collocated MODIS data to QC and/or
  - use collocated MODIS data directly (as the clear estimate) estimate our cloud-clearing extrapolation parameters (eta’s).

THE END. THANKS.

Error Estimates: Partitioning of Terms

- Linear error estimate for a regularized retrieval is partitioned into three terms
  - Smoothing term (finite width of kernel functions and correlation on first guess or *a priori* errors)
  - Systematic terms (errors due to interference of other geophysical parameters, cloud-clearing, *etc.*)
  - Random term (instrument noise)
- *A-priori* covariance can be derived from a validation ensemble – such as operational sondes, aircraft, etc.
- Smoothing & Random components are the easiest to predict. The systematic errors (*esp.* those due to cloud-clearing) are the most difficult.