Status of the Community Hyper-Spectral Infrared Microwave Earth Retrieval Algorithm (CHIMERA): algorithms improvements, validation and applications

Antonia Gambacorta, Chris Barnet, Mitch Goldberg
NASA Sounder Science Team Meeting
Wednesday, October 14, 2015

Outline of this talk

- **Part I**: Introduction to the CHIMERA project
  - Algorithm history, scientific basis, the path forward
- **Part II**: Recent algorithm development
  - A priori, error estimates
  - MW-only retrieval module
  - MW+IR retrieval module
- **Part III**: Applications and Ongoing Validation

CHIMERA: something old, something borrowed, something new…

- **Something new**. CHIMERA (Community Hyperspectral Infrared and Microwave Earth's Retrieval Algorithm) is a NASA ROSES funded project to develop a climate quality retrieval algorithm for hyperspectral sounders
- **Something borrowed**. Uses the NOAA Unique CrIS ATMS Processing System (NUCAPS) as an algorithm baseline
- **Something old**. Is a legacy algorithm of the AIRS Science Team code.

What's unique in the AST algorithm?

- **Designed to use all available sounding instruments.**
  - Climatological startup. Only ancillary information used is surface pressure from GFS model
  - Microwave radiances used in microwave-only physical retrieval, "allsky" regression solution, "cloud cleared" regression and downstream physical T(p) and q(p) steps.
- **Utilizes the high-information content of the hyper-spectral infrared – both radiances and physics.**
  - Sequential physical algorithm allows for a robust and stable system with minimal prior information
  - Utilizes forward model derivatives to help constrain the solution
  - Error from previous steps are mapped into an error estimate from interfering parameters
  - All channels used in linear regression first guesses.
- **Utilizes cloud clearing**
  - Goal is to sound as close to the surface as possible
  - Sacrifices spatial resolution to achieve global coverage: no clear sky biases
  - Allows graceful degradation with decreased information content
  - Avoids ad hoc switches between clear sky only and cloudy sky single FOV algorithm

What's unique about NUCAPS?

- **NOAA operational algorithm heritage of the AST code**, with additional unique components
- **Designed, from the beginning, to be product-centric** rather than sensor-centric (NPP Science Team priority recommendation)
  - AIRS/AMSU, IASI/AMSU/MHS, and CrIS/ATMS are processed with literally the same NUCAPS code.
  - Extremely fast compared to other approaches (1 CPU for CrIS/ATMS)
  - Same underlying spectroscopy (as best as we could do)
  - Instrument agnostic: specific items are file-driven, not hardwire
  - Code is backward and forward (as much as possible) compatible.
  - Retrieval components are programmable via namelists (can quickly compare retrieval enhancements and/or methodologies).
  - Operational code is a "filtered" version of the science code.
  - Capable of processing CrIS full-resolution spectra (Gambacorta 2013 IEEE GRSL);
- **Uses an open framework** (NPP Science Team priority recommendation)
  - other researchers can link other algorithms for the core products and new algorithms for ancillary products (e.g., cloud microphysical products, trace gases, etc.).
- **Could add new products**
  - Ammonia (pending project, PI is Juying Warner), Formic Acid (HCOOH), and Peroxyacetyl Nitrate (PAN)

What will be new and unique about CHIMERA?

- Definition of a climate quality algorithm:
  - A retrieval algorithm that can be characterized by explicitly evaluating the functional form of the relationship between the retrieved profile, the true atmosphere, and the various error sources.

- Three necessary modifications to the existing AST/NUCAPS code:
  - Modify each retrieval step to have a formal a-priori state and associated covariance
  - Propagate the full 2D error covariance from step to step
  - Output error covariance (T, q) or averaging kernel (trace gases)

Employ a formal a-priori

- The inversion solution minimizes the cost function J

$$J = \left(R_n^{obs} - R_n\left(X_j^{i-1}\right)\right)^T \cdot N_{n,n}^{-1} \cdot \left(R_n^{obs} - R_n\left(X_j^{i-1}\right)\right) + \left(X_j^{i-1} - X_j^A\right)^T \cdot C_{j,j}^{-1} \cdot \left(X_j^{i-1} - X_j^A\right)$$

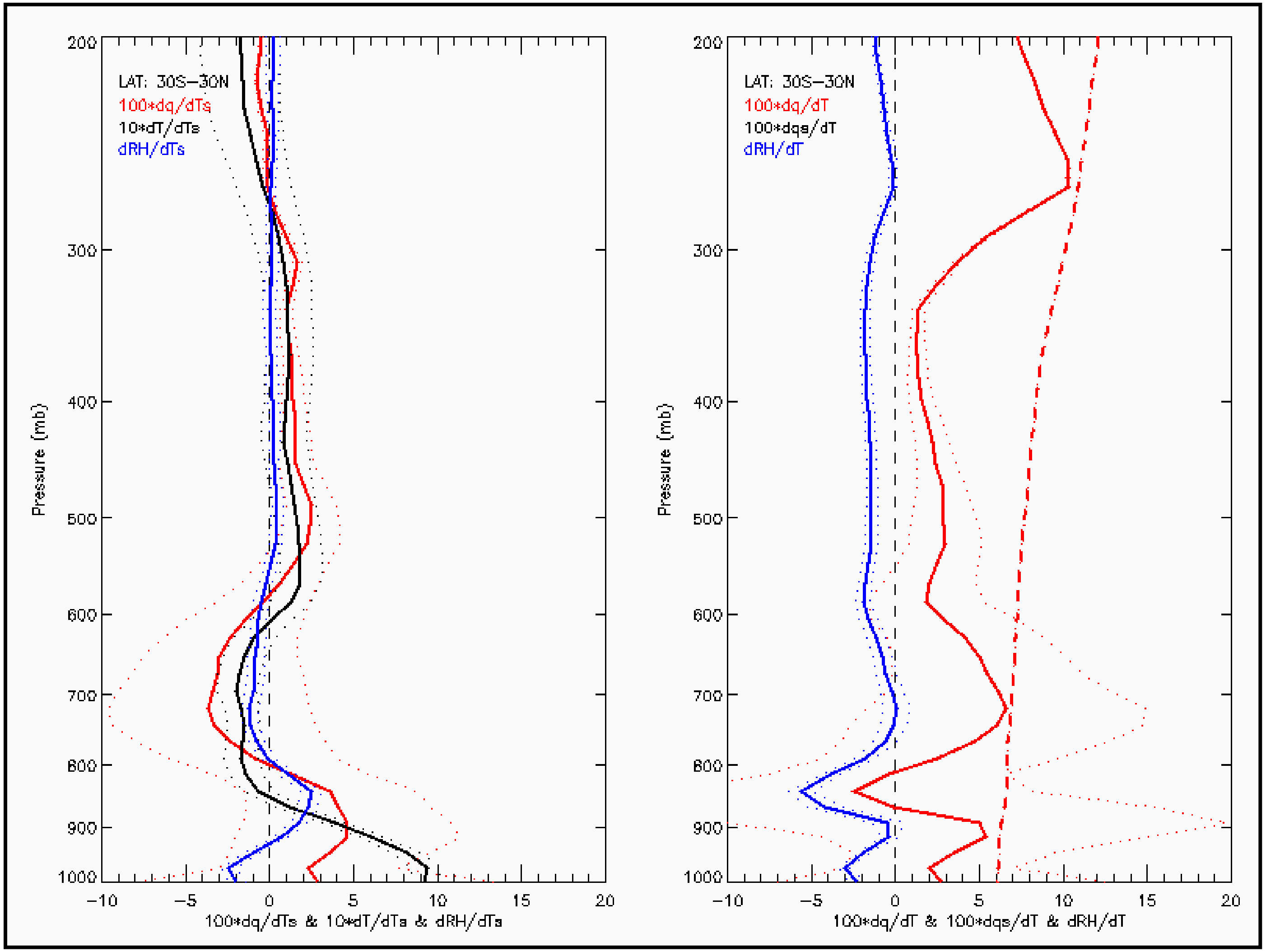
- Departure from a priori can be seen as a weighted mean of measurement and a priori contributions

$$X_j^i = X_j^A + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} + C_{j,j}^{-1}\right]^{-1} \cdot K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left[R_n^{obs} - R_n(X^{i-1}) + K_{n,j} \cdot \left(X_j^{i-1} - X_j^A\right)\right]$$

- AST approach pivots around the previous iteration; employs a regularization and a removal of background term from residuals

$$X_j^i = X_j^{i-1} + \left[K_{j,n}^T \cdot N_{n,n}^{-1} \cdot K_{n,j} + H_{j,j}\right]^{-1} \cdot K_{j,n}^T \cdot N_{n,n}^{-1} \cdot \left[R_n^{obs} - R_n(X^{i-1}) - \Psi_n^{i-1}\right]$$

Drawing a conclusion on the tropical water vapor feedback



Slide from AGU Fall Meeting 2008, Gambacorta et al., GRL, 2008

The problem of a formal a priori

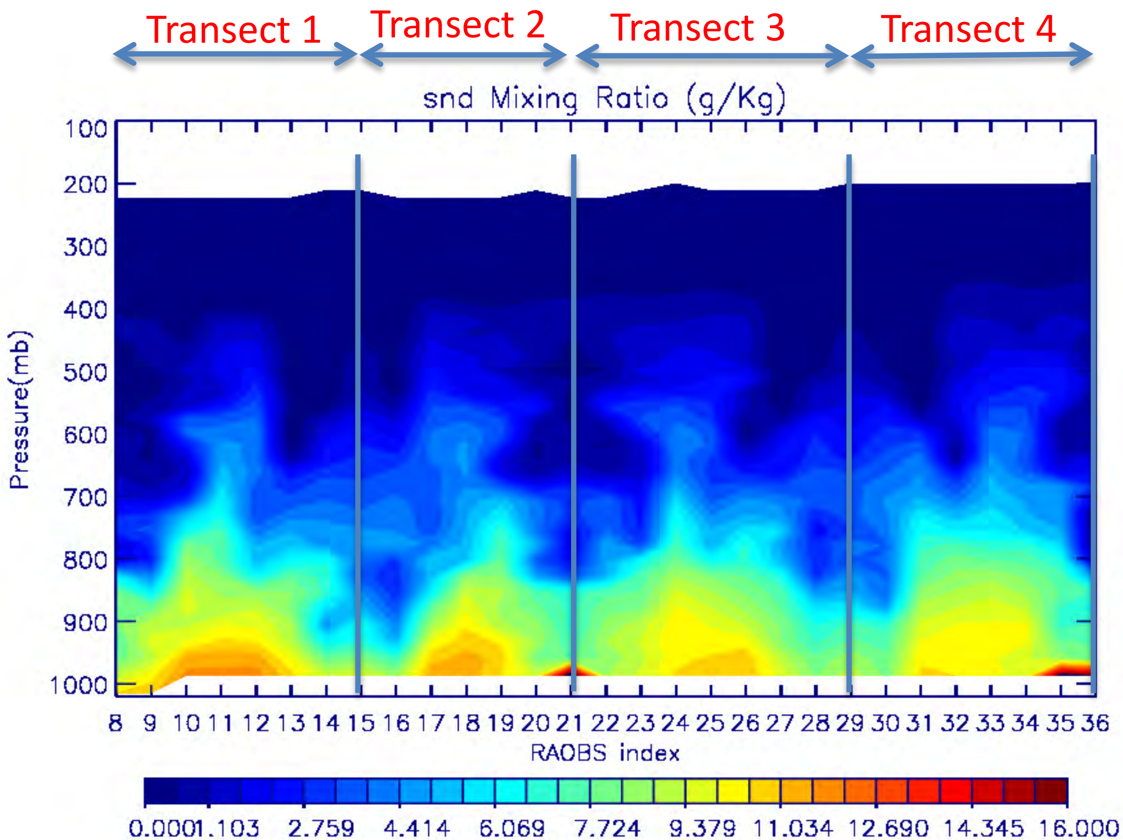
Problem: the lack of a formal a priori in the AST algorithm is mathematically sub-optimal.

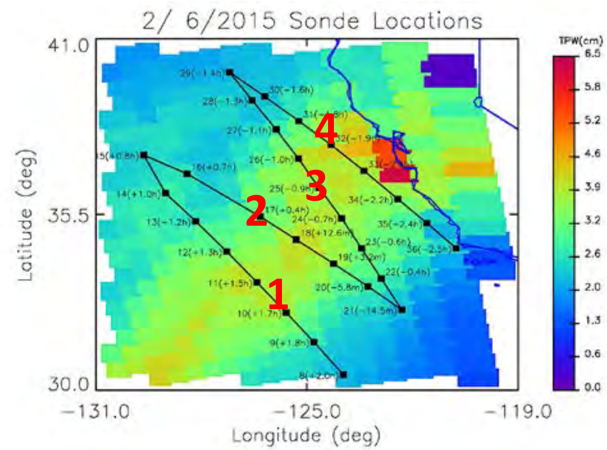
- The AST solution does not equate the minimization of the cost function
- The *a- priori* represents our knowledge of the atmospheric state before the measurements are made and should not be replaced by something that is the result of using the measurements (Rodgers, 2000). Pivoting around the previous iteration prevents from partitioning sources of errors in the retrieval
- Using a statistical operator as first guess can introduce sub-resolved structures in the retrieval

We have started investigating three possible climate *a-priori:*

- 1) climatology built from a decade of ECMWF (this has already been constructed by the AIRS science team and will be tested)
- 2) ERA-interim; NCEP reanalysis; MERRA.
- 3) microwave-only retrieval. For CrIS/ATMS this has the potential to be an exceptional *a-priori*. For AIRS/AMSU it is unlikely that the AMSU information content is sufficient.

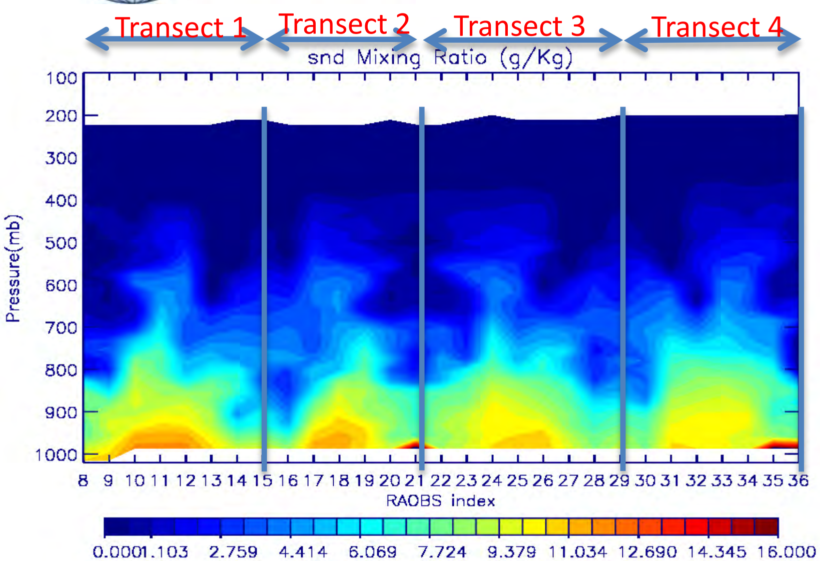
Radiosonde measurements from CalWater 2015 February 6th test case



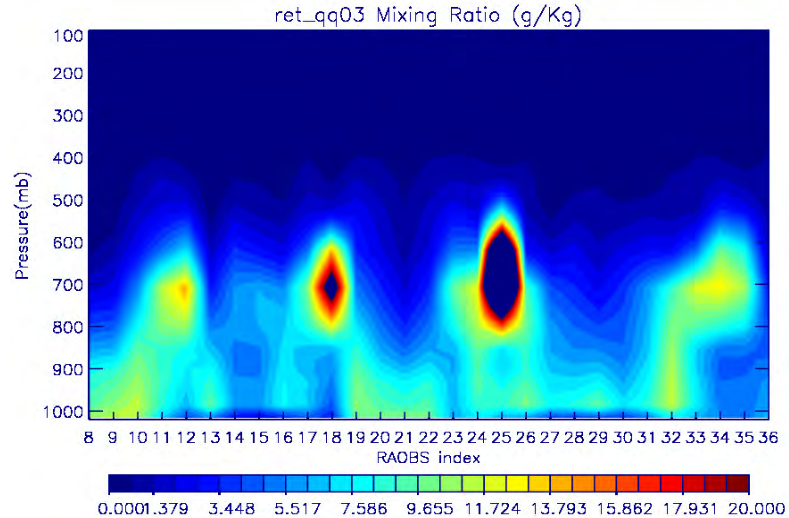


- ~ 4 hours flight with 4 transects across the river capturing pre, in and post river environment as the river quickly approaches the US West coast
- Good spatial and temporal matching with NPP (drop sonde location 19 is ~ 3.2 minutes ahead of over pass

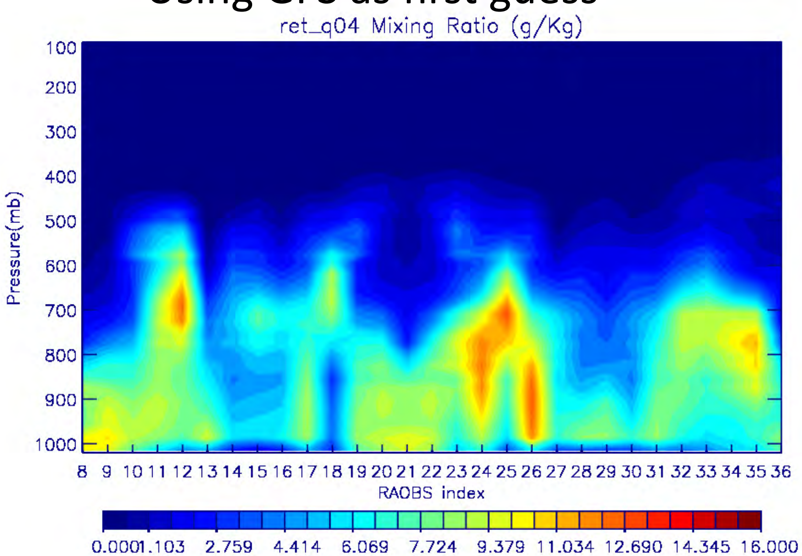
Towards new a-priori choices (work in progress)



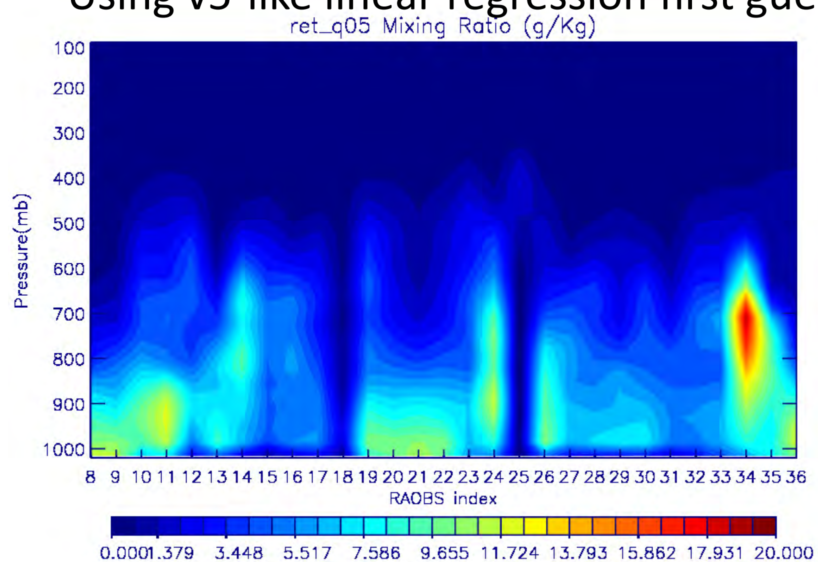
Using the MW-only retrieval as first guess



Using GFS as first guess



Using v5-like linear regression first guess



Correct propagation of retrieval errors in the sequential approach

- Solving the linearized radiative transfer equation requires a *weighted* least square minimization to properly take into account the non linear sources of errors in the measurements.

$$R_{OBS} - R_{CALC} = K \cdot \Delta X + \varepsilon$$

$$\mathbf{X}_j^i = \mathbf{X}_j^A + \left[\mathbf{K}_{j,n}^T \cdot \mathbf{N}_{n,n}^{-1} \cdot \mathbf{K}_{n,j} + \mathbf{C}_{j,j}^{-1}\right]^{-1} \cdot \mathbf{K}_{j,n}^T \cdot \mathbf{N}_{n,n}^{-1} \cdot \left[\mathbf{R}_n^{obs} - \mathbf{R}_n(\mathbf{X}^{i-1}) + \mathbf{K}_{n,j} \cdot \left(\mathbf{X}_j^{i-1} - \mathbf{X}_j^A\right)\right]$$

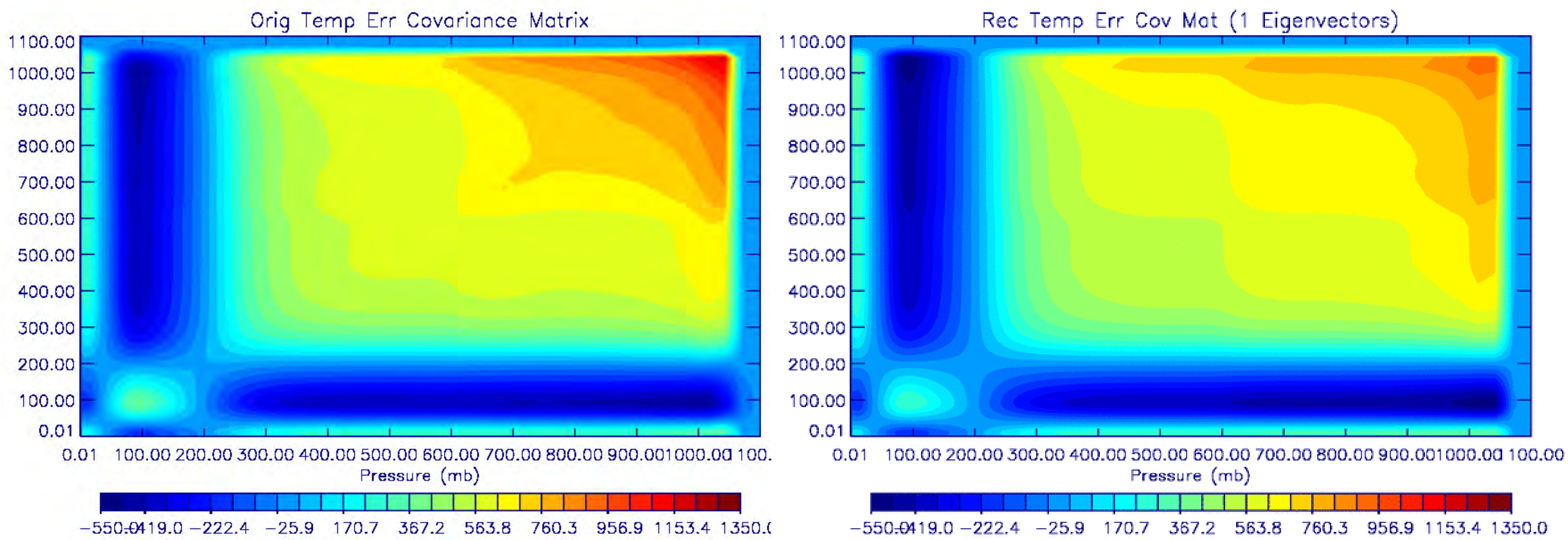
- Solving the linearized radiative transfer equation *in steps*, implies proper computation and propagation of retrieval error estimates of all the unsolved *g* retrieved species kept constant at a given step.

$$N_{n,n} = \Delta nedn_{n,n} + \Delta CCR_{n,n} + \sum_g K_{n,L} \delta X_L^g \delta X_L^{g^T} K_{L,n}^T$$

Correct propagation of retrieval errors in the sequential approach

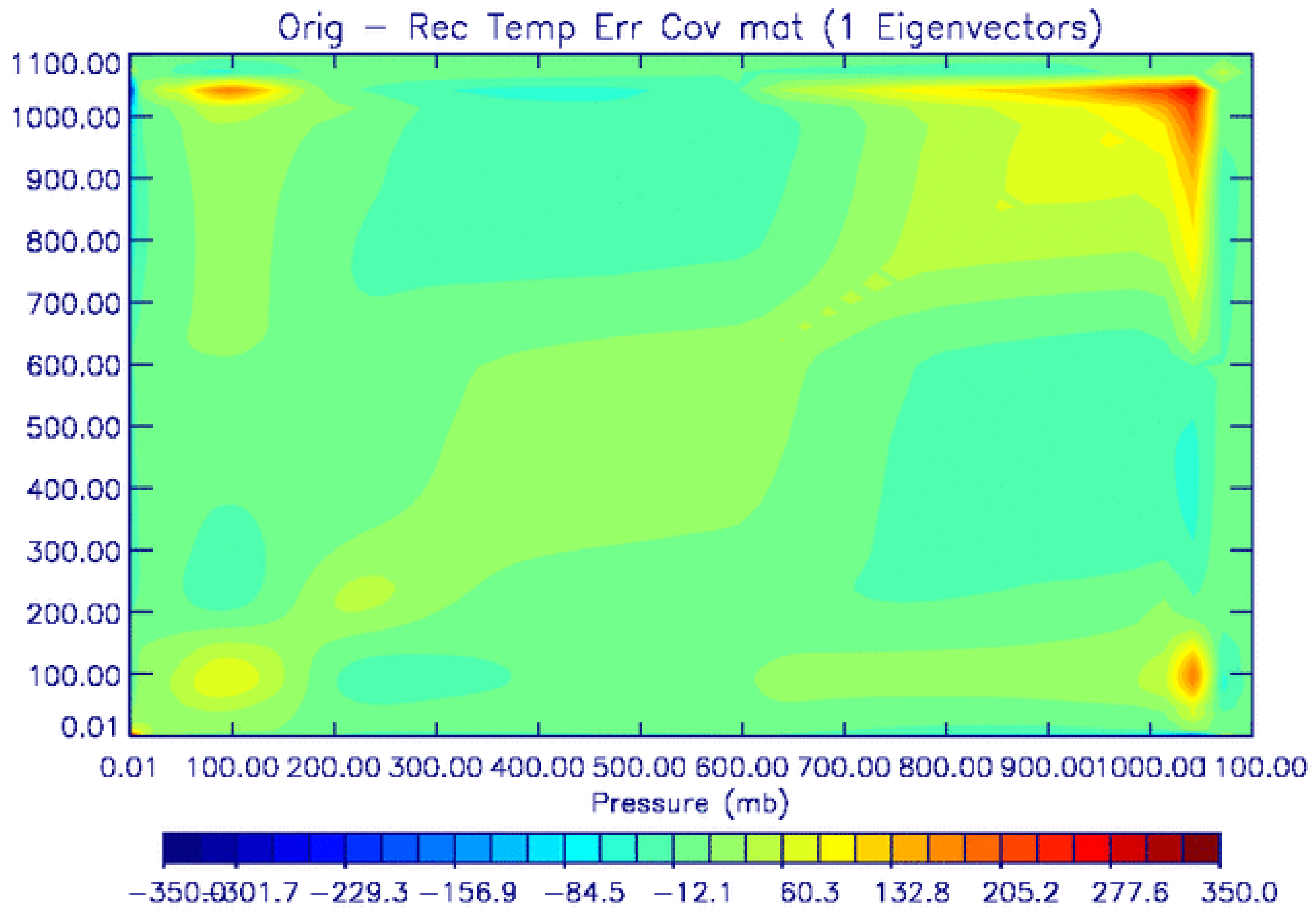
- We currently only map the diagonal component of the retrieval error covariance into down-stream steps.
  - It has been shown that there is a robust way to pass the full retrieval error covariance from one step to the next (Chris Barnet Mar. 23 2007 AIRS meeting, session 6, and Eric Maddy, AIRS meeting, Apr. 27, 2011 session 6). This is by compressing the retrieval error estimate covariance matrix and only propagating the significant eigenvalues and eigenvectors to the next step. Then reconstruct the retrieval error covariance and use it to compute the measurement error covariance.
  - We have recently investigated the number of significant pieces of information needed to do this. See next slides.

Original vs reconstructed temperature error covariance



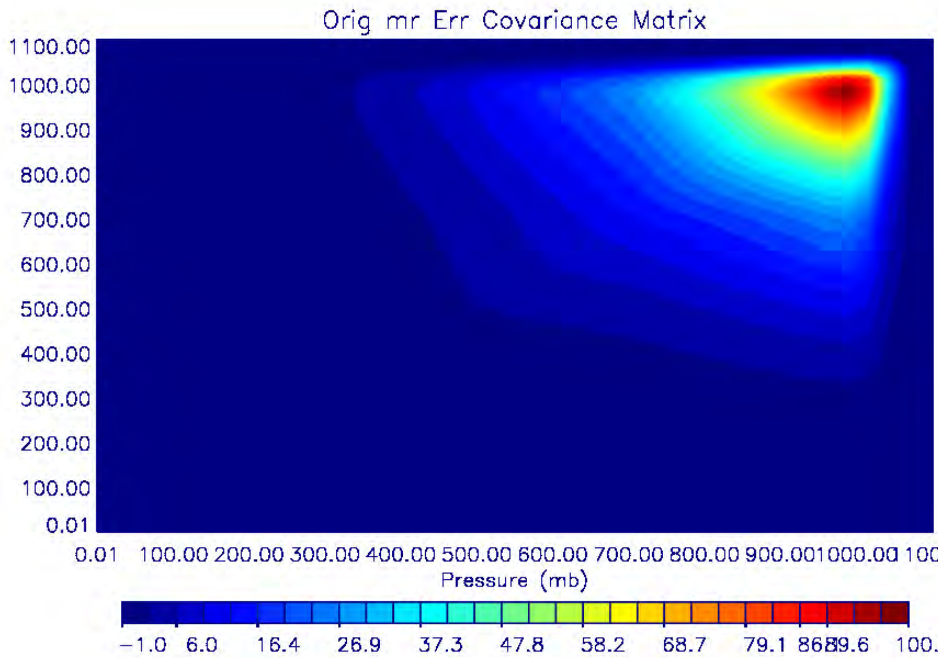
- Global Focus day: 2013/04/15/
- Temperature error:
  - Temp err δT_L = Retrieved Temperature – ECMWF Temperature
- Covariance matrices:
  - (Temp err - <Temp err>)##(Temp err - <Temp err>)

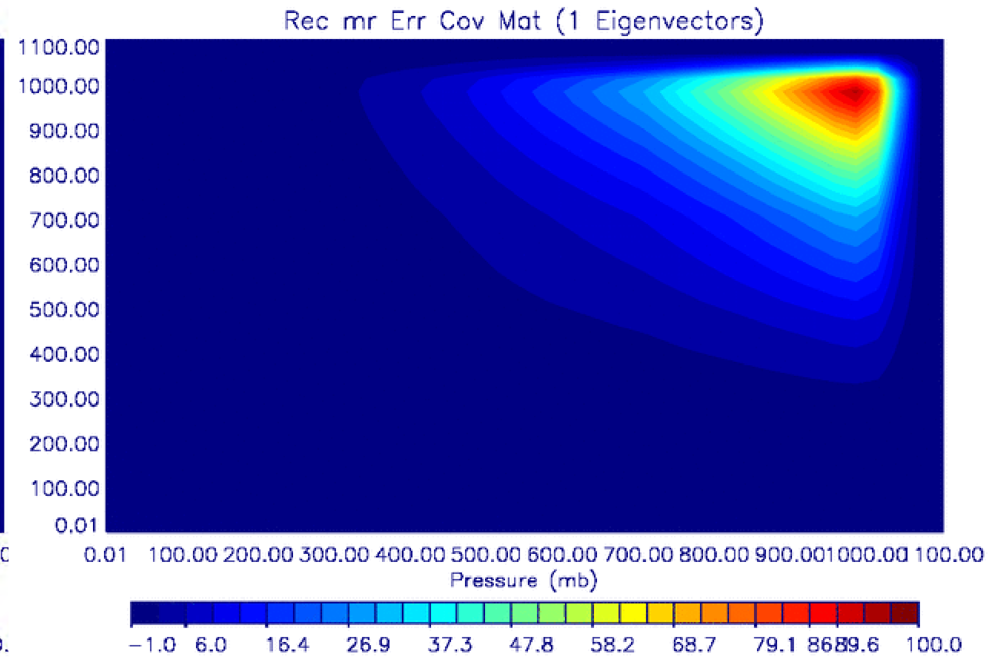
Original – Reconstructed Temperature Error Covariance Matrix



6 Eigenvectors are enough to fully reconstruct the original temperature error covariance

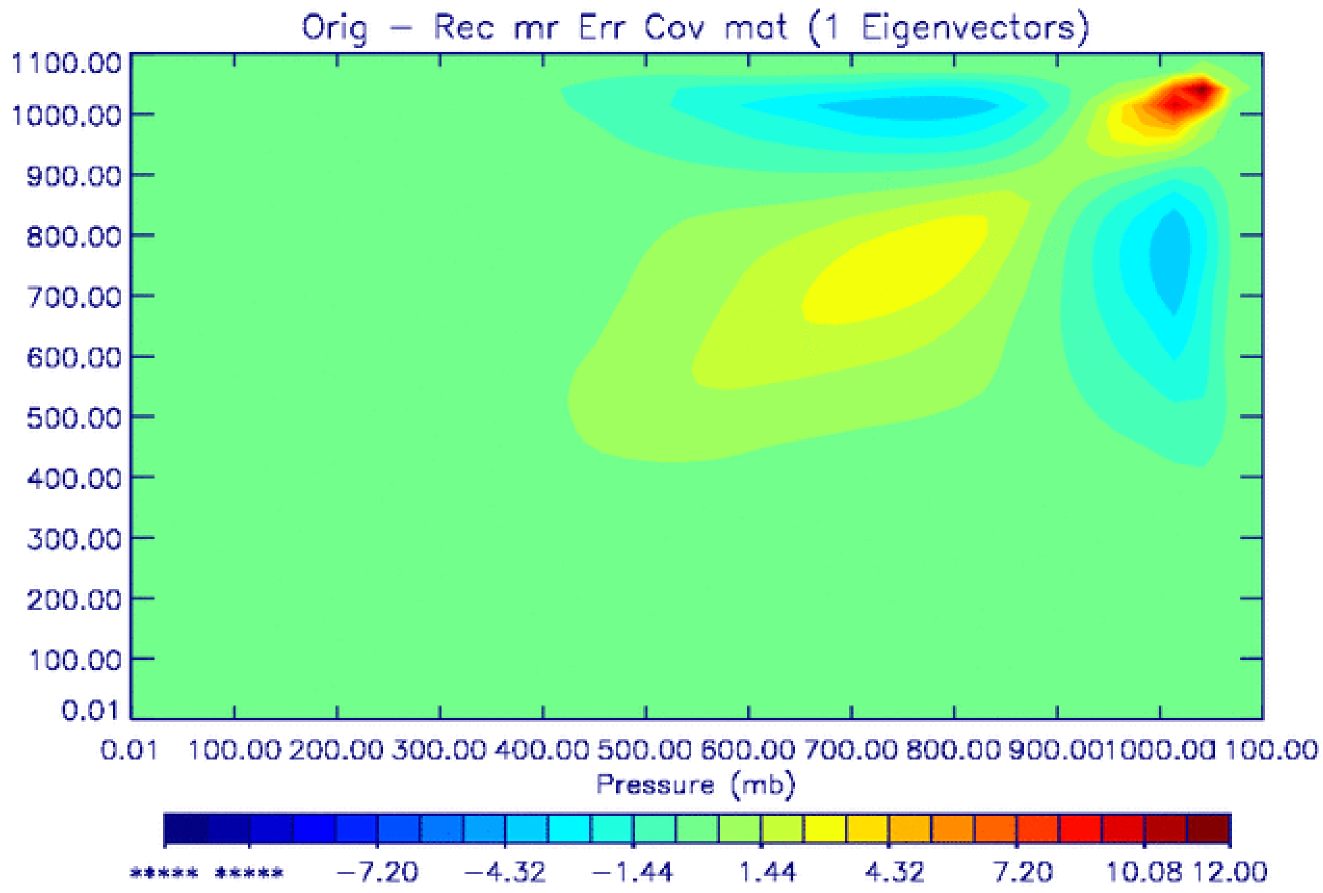
Original vs Reconstructed Water vapor Error Covariance Matrix





- Focus day: 2013/04/15/
- Water vapor mix ratio (kg/kg) error:
  mr err δq_L = Ret mr – ECMWF mr
- Covariance matrices:
  (mr err - <mr err>)##(mr err - <mr err>)

Original - Reconstructed Water vapor Error Covariance Matrix



6 Eigenvectors are enough to fully reconstruct the original water vapor error covariance

Formal computation of error estimates

- The retrieval error covariance can be partitioned into measurement and *a-priori* covariance components: $\delta X \delta X^T = [K^T N^{-1} K + C^{-1}]^{-1}$
- Once you have a formal a-priori and a full representation of the measurement error covariance, you can compute formal error estimates
- Error estimates are essential to provide the level of confidence on any climate application performed with this data set.

Error estimates and QC products

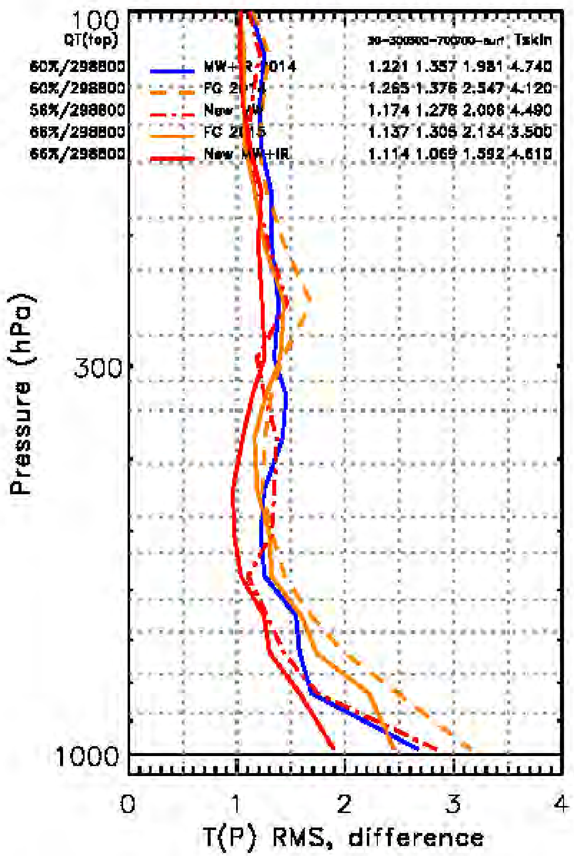
- Current AST algorithm uses empirical error estimates derived from a linear regression between ECMWF and internal quality indicators (retrieval residuals, cloud clearing noise amplification, etc.)
- NUCAPS QCs use the physical approach of the AST version 4: residuals, degrees of freedom, comparison between MW-only and MW+IR retrieval, MW residuals using the MW+IR retrieval. No error estimates due to the lack of a formal a-priori.
- CHIMERA will output the same QC parameters as in NUCAPS in addition to the error estimates (for temperature, moisture, surface, and cloud products) and averaging kernels for the trace gases.

Recent Algorithm Enhancements - MW+IR Retrieval

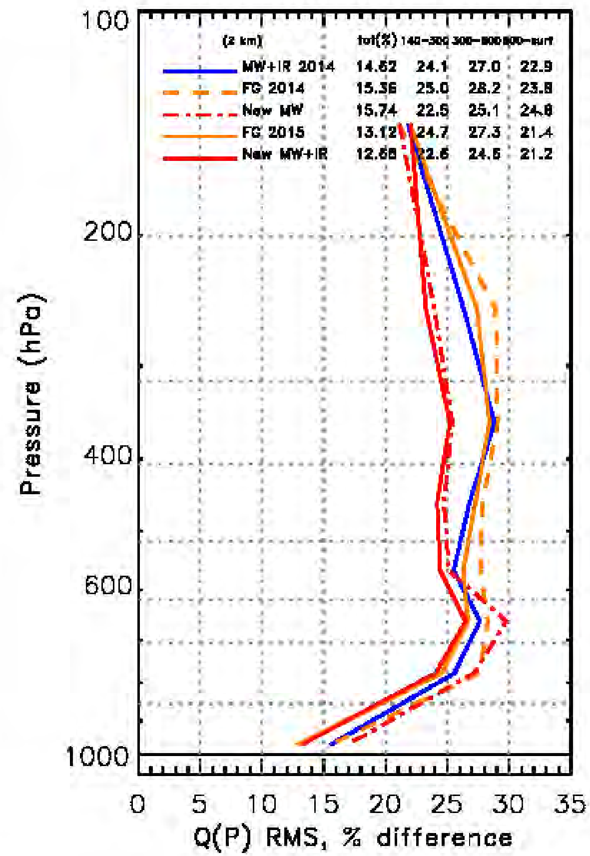
MW+IR Module

- 2014 MW+IR System
- OLD FG (dash orange)
- NEW FG (solid orange)
- New MW-Only System (dot-dash red)
  - New RTA error and bias tuning
  - New nedt
- New first guess (solid red)

Temperature



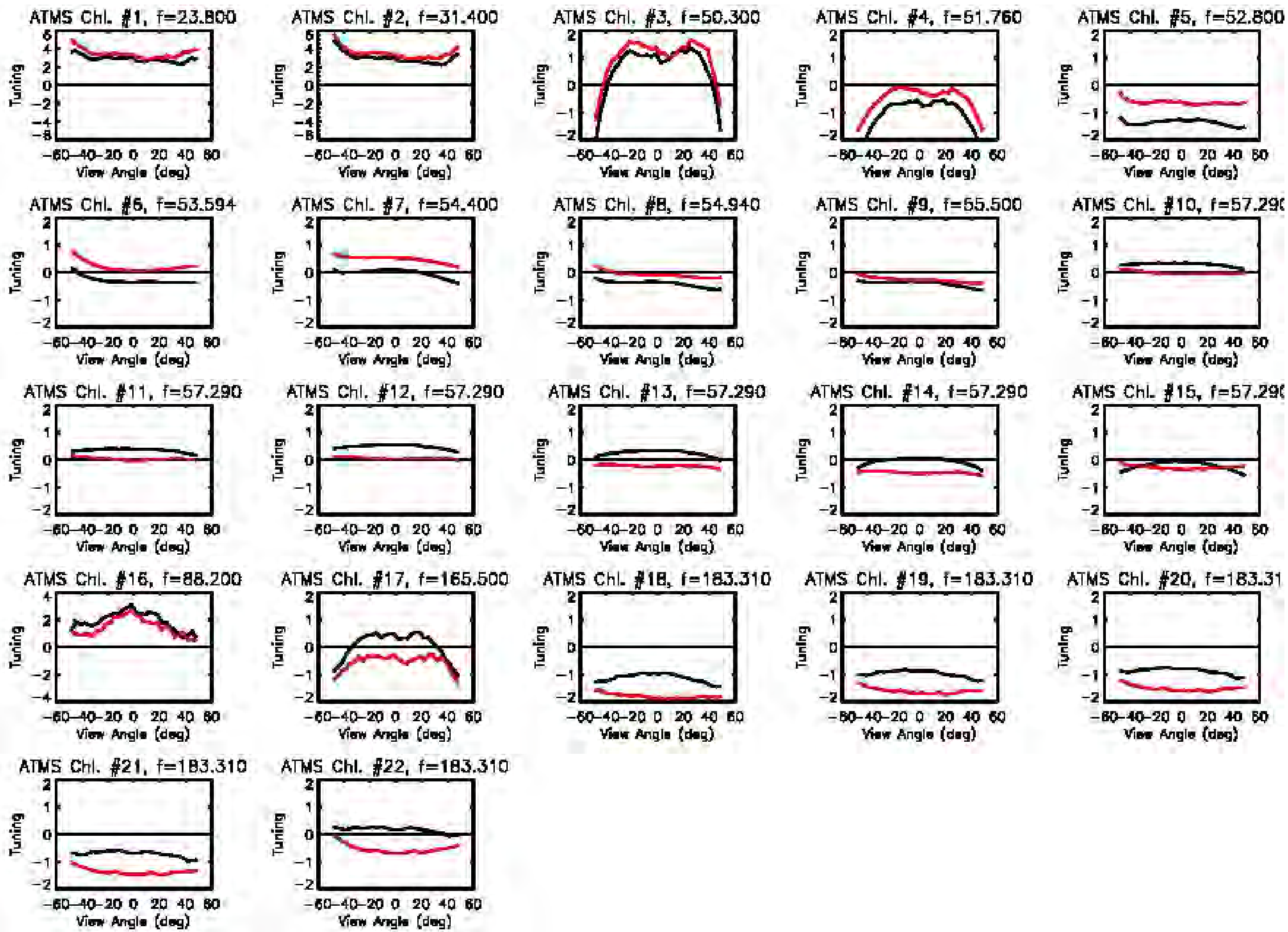
Water vapor



183GHz bias problem

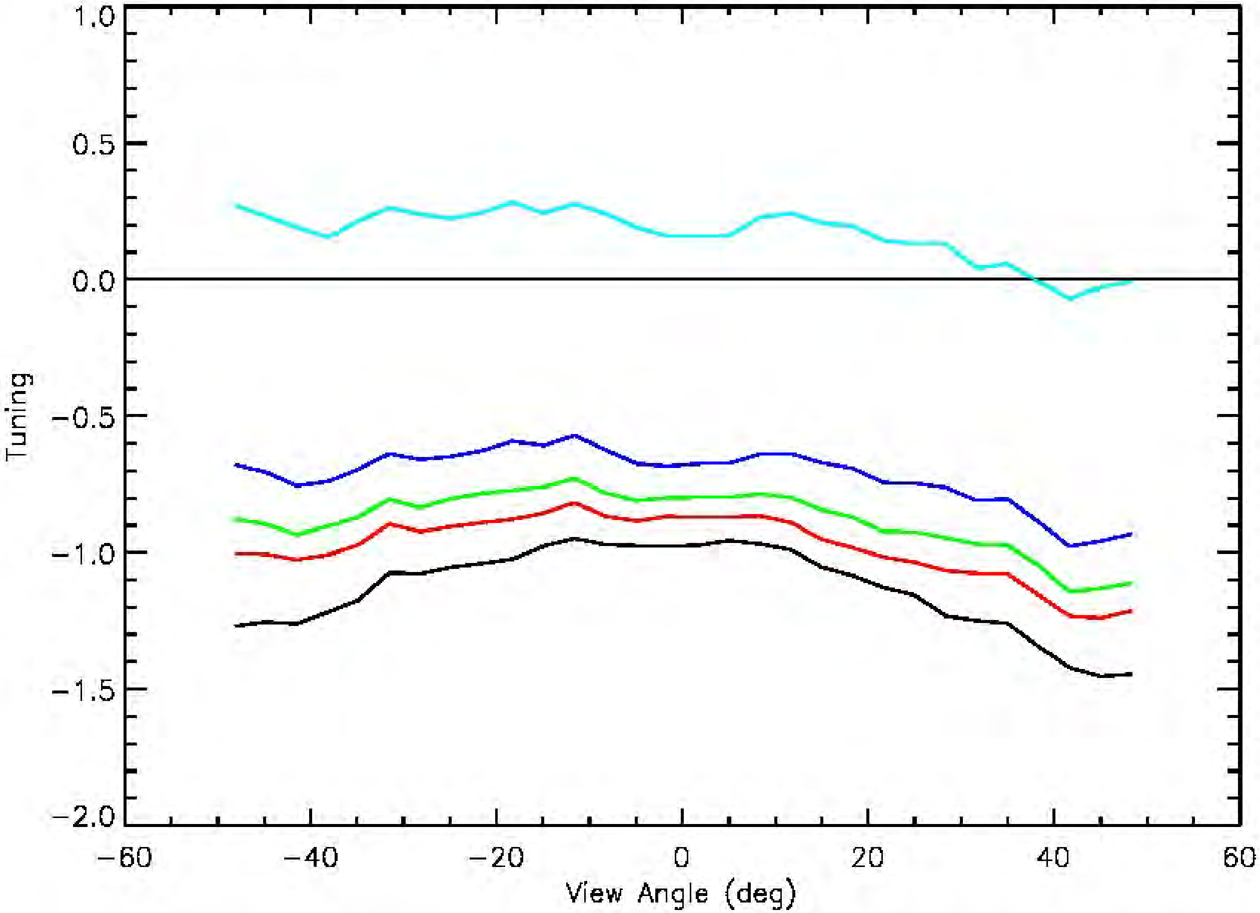
- Ongoing discussion on the sensitivity peak height dependent bias in the 183GHz band
  - OBS-CALC bias computation is observed to increase with lower peaking 183GHz channels
  - Problem is observed across all current forward models and MW instruments (AMSU, SAPHIR, ATMS)
  - Problem is observed on both ATMS TDR and SDR files (next 2 slides)
  - 29-30 June 2015: a dedicated workshop to study the issue. Executive summary available upon request.
  - Possible sources: surface, precipitation contamination, water vapor continuum.
  - We are in contact with Phil Rosenkranz who has an updated forward model with improved water vapor transmittance.

ATMS tuning
TDR (black) & SDR (red)



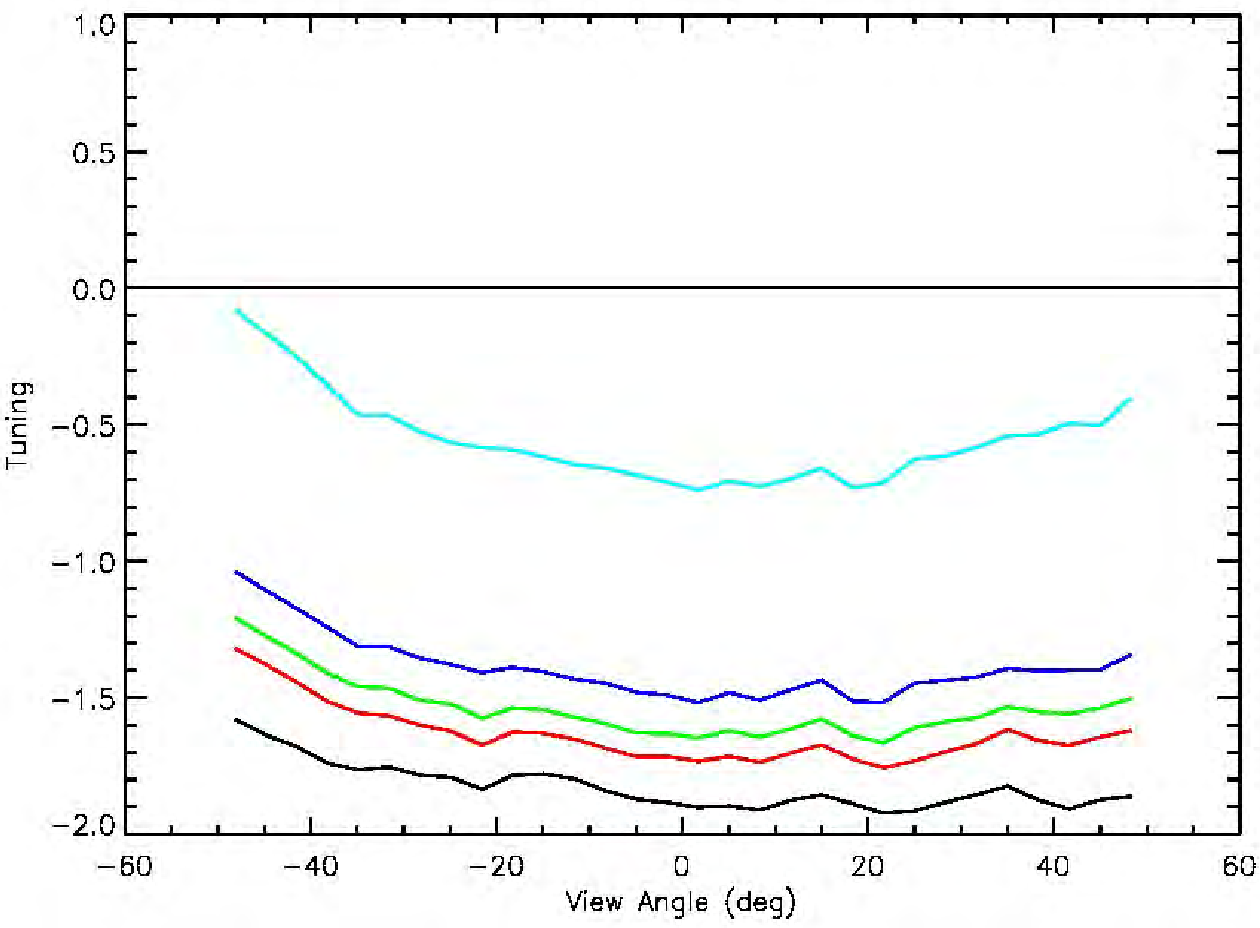
183 GHz bias (OBS-CALC): TDR cases

ATMS Chn. 22
ATMS chn. 21
ATMS chn. 20
ATMS chn. 19
ATMS chn. 18



183 GHz bias (OBS-CALC): SDR cases

ATMS Chn. 22
ATMS chn. 21
ATMS chn. 20
ATMS chn. 19
ATMS chn. 18



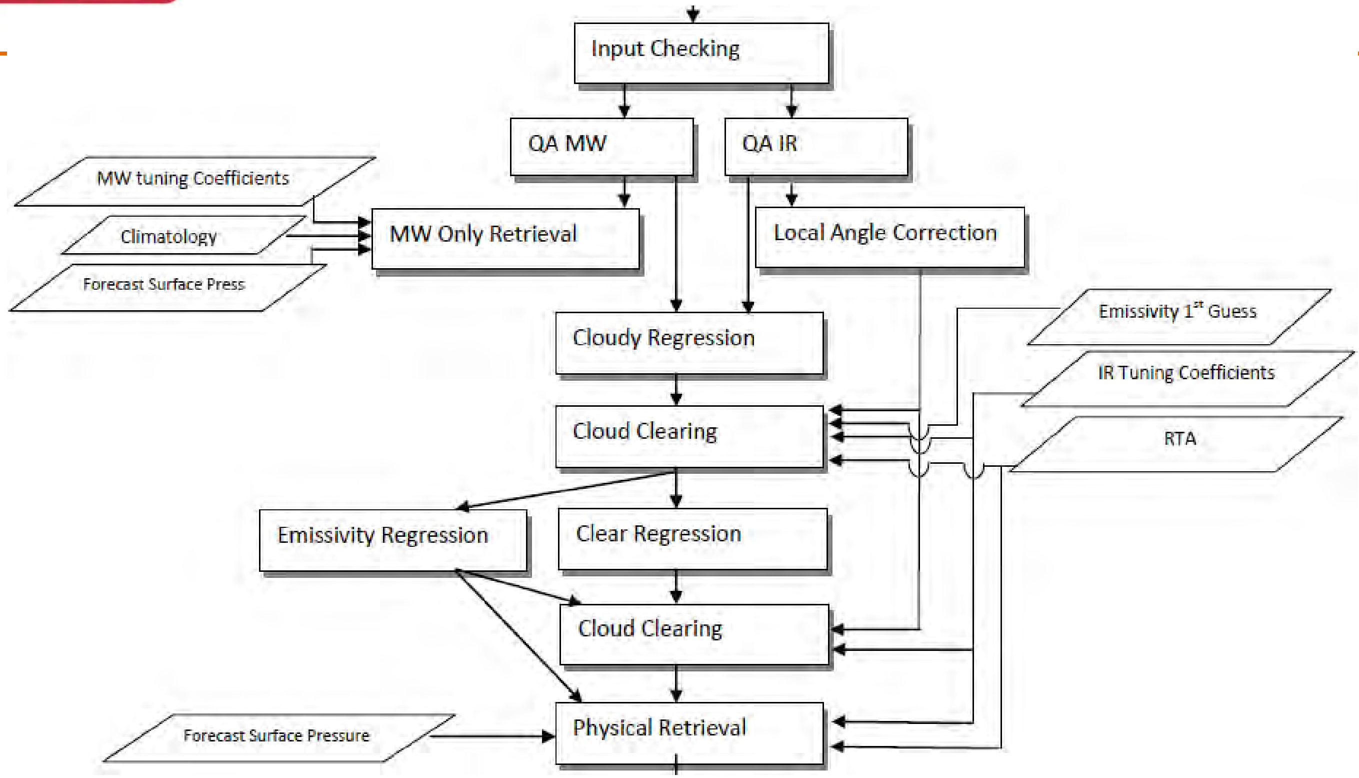
Summary

- CHIMERA: progress are being made to achieve a stable climate quality hyperspectral retrieval algorithm
- On going international projects to inter-compare and validate existing algorithms:
  - 183GHz working group (workshop executive summary to be submitted for peer review).
  - GEWEX water vapor assessment (workshop in Madison November 4-5, 2015)
  - Eumetsat heperspectral algorithm comparison (workshop in Potenza, Italy, late November 2015).

… Keeping ourselves busy

Back up slides

The AIRS Science Team Algorithm



- **I. A microwave retrieval module** which computes Temperature, water vapor and cloud liquid water (Rosenkranz, 2000)
- **II. A fast eigenvector regression** retrieval that is trained against ECMWF and CrIS all sky radiances which computes temperature and water vapor (Goldberg et al., 2003)
- **III. A cloud clearing module** (Chahine, 1974)
- **IV. A second fast eigenvector regression** retrieval that is trained against ECMWF analysis and CrIS cloud cleared radiances
- **V. The final infrared physical retrieval** based on a regularized iterated least square minimization: temperature, water vapor, trace gases (O3, CO, CH4, CO2, SO2, HNO3, N2O) (Susskind, Barnet, Blaisdell, 2003)