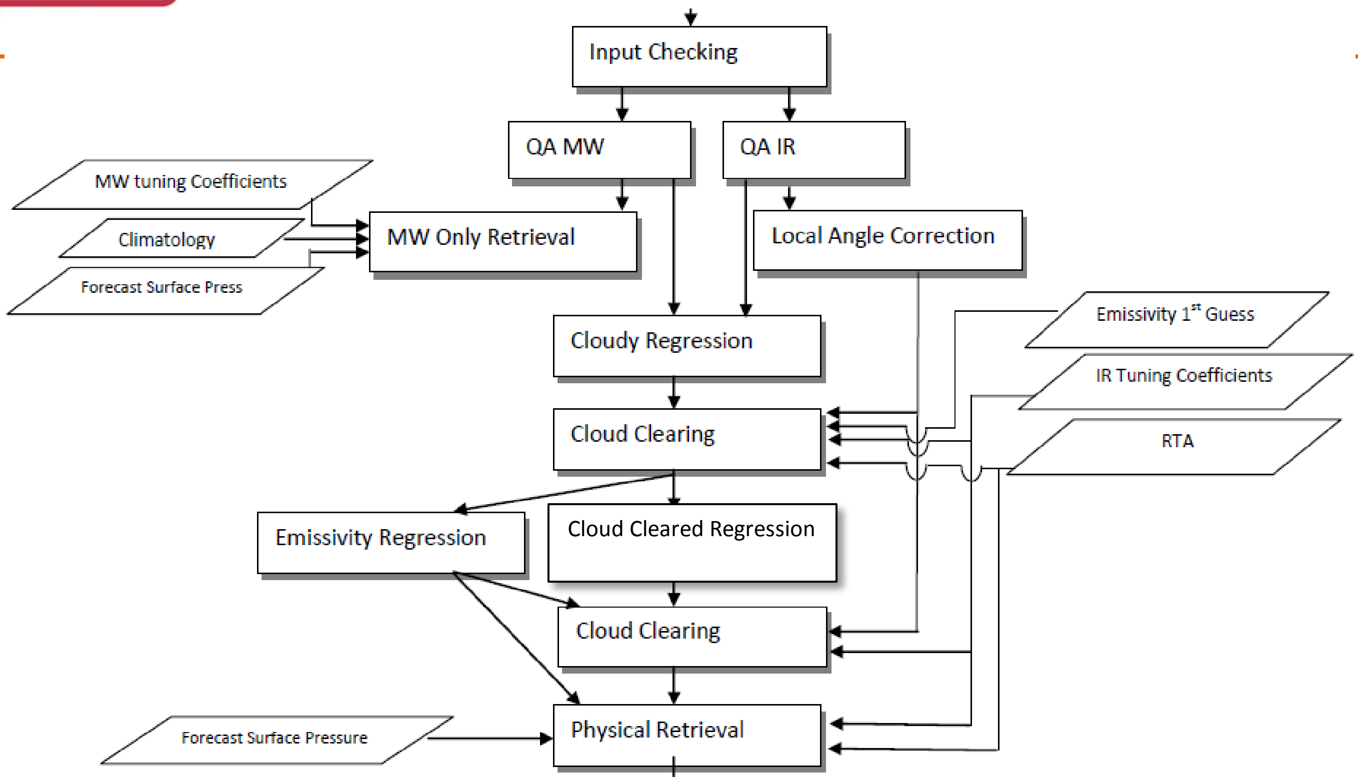
Status of the NOAA Unique CrIS ATMS Processing System (NUCAPS): algorithm development and lessons learned from recent field campaigns

Antonia Gambacorta [1]
Chris Barnet [1]
Mitch Goldberg [2]

1. Science and Technology Corporation, STC
2. NOAA JPSS Project Scientist

The NUCAPS 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. This is used as first guess for the physical retrieval.
- **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)

What’s unique about NUCAPS?

- **NOAA operational algorithm heritage of the AIRS Science Team (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.
- **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, Formic Acid (HCOOH), and Peroxyacetyl Nitrate (PAN)

What’s unique about NUCAPS?

- **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. Visible radiances can be used to characterize sub-pixel inhomogeneity.
- **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**
  - Sacrifices spatial resolution to achieve global coverage: no clear sky biases
  - Avoids ad hoc switches between clear sky only and cloudy sky single FOV algorithm
  - Goal is to sound as close to the surface as possible: essential to measure temperature inversions, stability indexes under fast developing storms.

NUCAPS ultimate goal: a climate quality retrieval algorithm

- **Why do we need a climate quality retrieval algorithm?**
  - An independent environmental data record to test Global Circulation Models (GCM) and understand current discrepancies.
  - An independent data record to study atmospheric variability and feedbacks.
- **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.

$$
\begin{aligned}
X_{ret} - x_{tru} = &\ \text{forward model parameter error} \\
&+ \text{forward model error} \\
&+ \text{retrieval noise} \\
&+ \text{smoothing error}
\end{aligned}
$$

Focus of this talk: how do we achieve climate quality?

- **The solution of the linearized inversion problem using measurement and a priori Gaussian statistics:**

Sensitivity matrix

Atmospheric state

Radiance measurement

$$y = Kx + \varepsilon$$

Error term

$$x = (K^T S_y^{-1} K + S_a^{-1})^{-1} (K^T S_y^{-1} y + S_a^{-1} x_a)$$

Measurement error covariance

A priori error covariance

- **Focus of this talk**
  - How do we maximize linearity?
  - How do we correctly build the measurement and a priori error covariance?
  - How do we produce a formal error estimate?

How do we maximize linearity?

- NUCAPS approach: treat non linear sources of errors as off-diagonal terms of the 2D measurement error covariance:

$$S_y = \Delta nedn + \Delta CCR + \sum_g K \delta X \delta X^T K^T$$

Instrument noise

Cloud Clearing Noise

Error from atmospheric species not being solved

- All sky retrievals vs cloud clearing
  - Removes the difficulty of separating clouds from temperature and water vapor, typical of simultaneous cloudy retrievals
  - Error introduced by cloud clearing is built into the measurement error covariance matrix
  - Simple concept: a small number of parameters can remove cloud contamination from thousands of channels.
  - Does not require a model of clouds and is not sensitive to cloud spectral structure
  - Works with complex cloud systems (multiple level of different cloud types).

How do we maximize linearity?

- **Sequential OE** (solve each state variable separately) **vs simultaneous OE** (solve all parameters simultaneously) **approach**

$$R_n^{obs} - R_n(\vec{X}) \simeq K^1_{n,i} \cdot \Delta\vec{T}_i + e_n$$

- Careful analysis of the physical spectrum will show that many components are physically separable (spectral derivatives are unique).
- Select channels within each step with large K and small e_n
- This makes solution more stable with respect to the simultaneous OE approach.
- State matrices are small and covariance matrices of the channel subsets are quite small. This has significant implications for operational execution time.

$$\begin{aligned} e_n &= K^2_{n,i} \cdot \delta\vec{q}_i \\ &+ K^3_{n,i} \cdot \delta\vec{O}_i \\ &+ K^4_{n,i} \cdot \delta\vec{C}O_i \\ &+ \ldots + \epsilon_n \end{aligned}$$

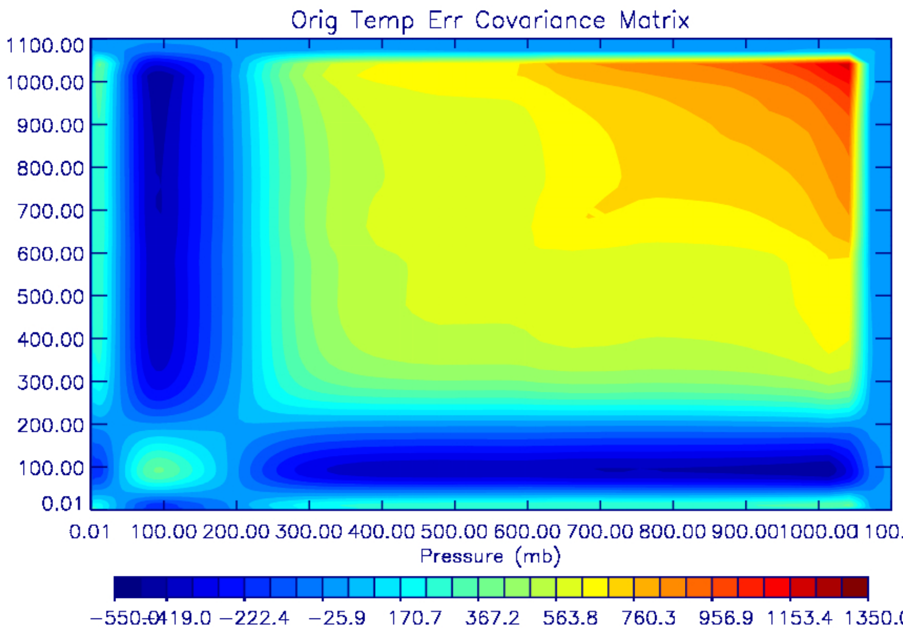
Simultaneous vs sequential OE approach

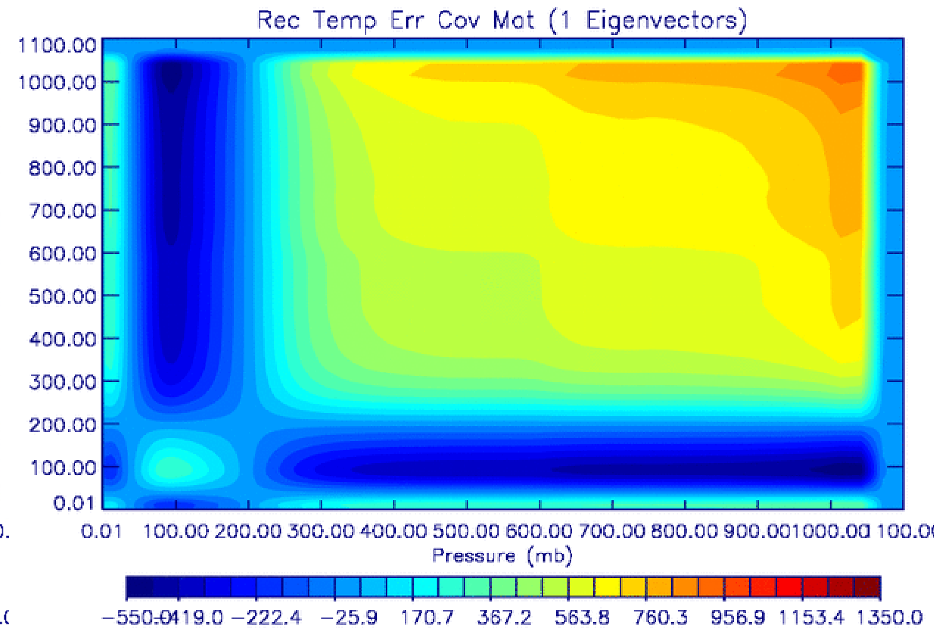
Simultaneous OE	Sequential OE
Solve all parameters simultaneously	Solve each state variable (*e.g.*, T(p)), separately.
Error covariance includes only instrument model.	Error covariance is computed for all *relevant* state variables that are held fixed in a given step. Retrieval error covariance is propagated between steps.
Each parameter is derived from all channels used (*e.g.*, can derive T(p) from CO2, H2O, O3, CO, … lines).	Each parameter is derived from the best channels for that parameter (*e.g.,* derive T(p) from CO2 lines, q(p) from H2O lines, etc.)
A-priori must be rather close to solution, since state variable interactions can de-stabilize the solution.	*A-priori* can be simple for hyperspectral.
This method has large state matrices (all parameters) and covariance matrices (all channels used). Inversion of these large matrices is computationally expensive.	State matrices are small (largest is 25 T(p) parameters) and covariance matrices of the channels subsets are quite small. Very fast algorithm. Encourages using more channels.

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 2D retrieval error covariance from one step to the next (Chris Barnet Mar. 23 2007 AIRS meeting, and Eric Maddy, AIRS meeting, Apr. 27, 2011). 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.

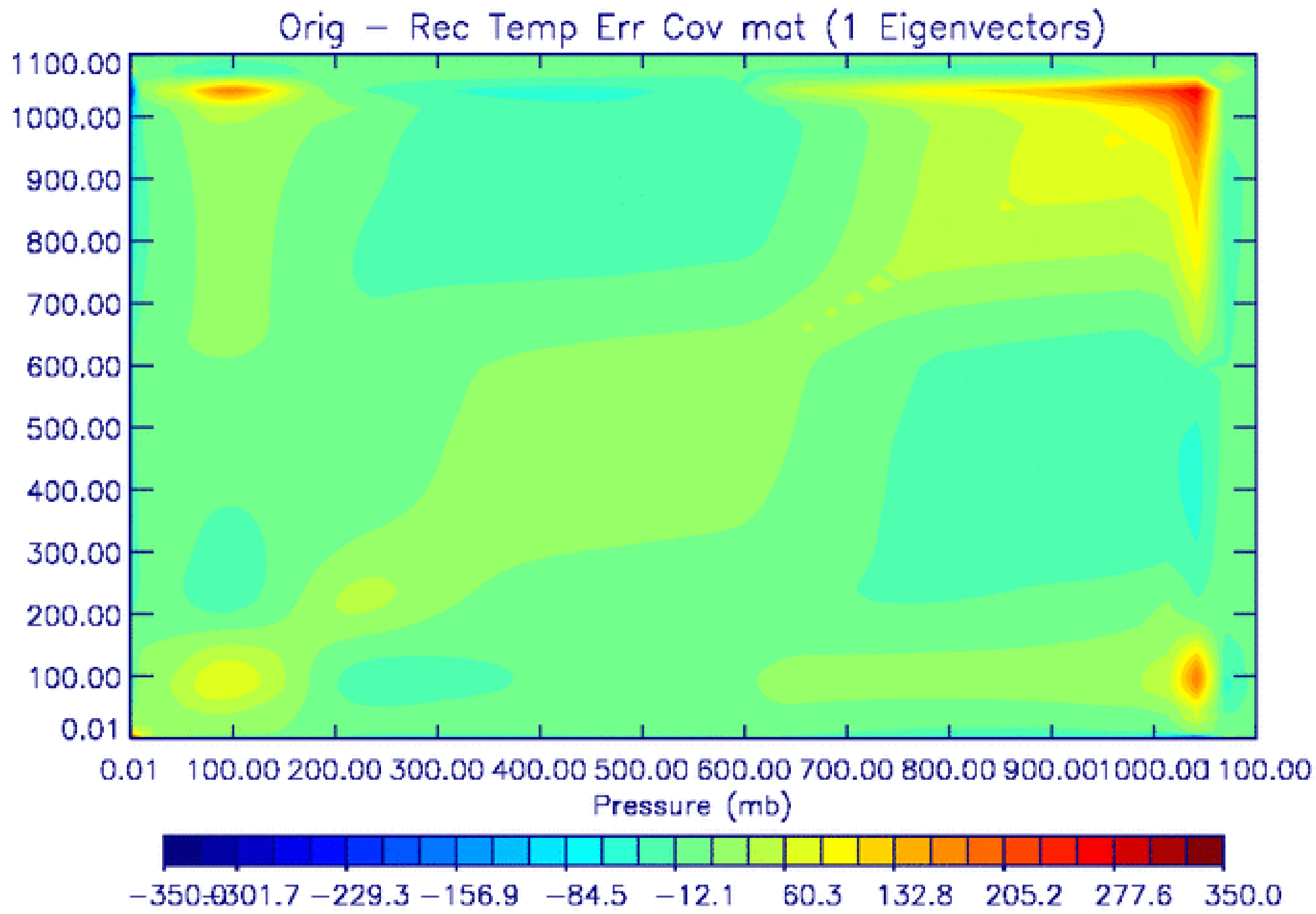
Original vs reconstructed temperature error covariance





- Global Focus day: 2013/04/15/
- Temperature error:
  - Temp err δT_L = Retrieved Temperature – “truth” 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

Correct choice of a priori and first guess

- NUCAPS is currently using a statistical operator (linear regression) as a priori
- Pro's and Con's Of Statistical Regression Retrievals

Pro's	Con's
Does not require a radiative transfer model for training or application.	Training requires a large number of co-located "truth" scenes.
Application of eigenvector & regression coefficients is VERY fast and for hyper-spectral instruments it is very accurate.	The regression operator does not provide any diagnostics or physical interpretation of the answer it provides. It can introduce sub-resolved structures in the retrieval
Since real radiances are used the regression implicitly handles many instrument calibration (e.g., spectral offsets) issues. This is a huge advantage early in a mission.	The regression answer builds in correlations between geophysical parameters. For example, retrieved O_3 in biomass regions might really be a *measurement* of CO with a statistical correlation between CO and O_3.
Since clouds are identified as unique eigenvectors, a properly trained regression tends to "see through" clouds.	Very difficult to assess errors in a regression retrieval without the use of a physical interpretation.

Correct choice of a priori and first guess

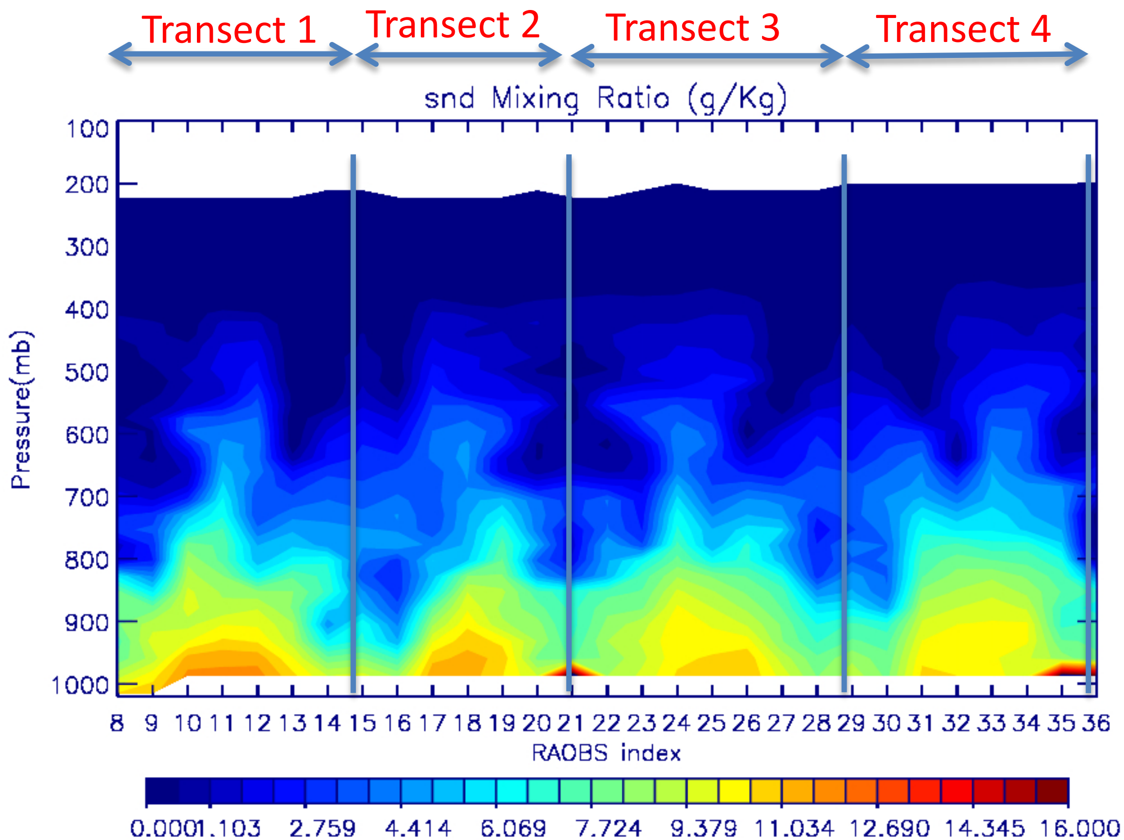
- Aforementioned expression of retrieval solution is derived on the assumption that both first guess and a priori have a Gaussian statistics. Gaussian statistics in a priori and first guess needs to be verified (on going work).
- We have started investigating three possible *a-priori:*

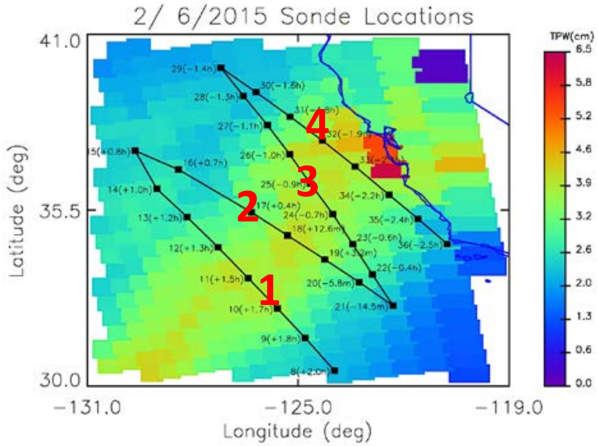
1) climatology built from a decade of ECMWF (this has already been constructed by the AIRS science team and will be tested)

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 and IASI/AMSU/MHS it is unlikely that the AMSU information content is sufficient.

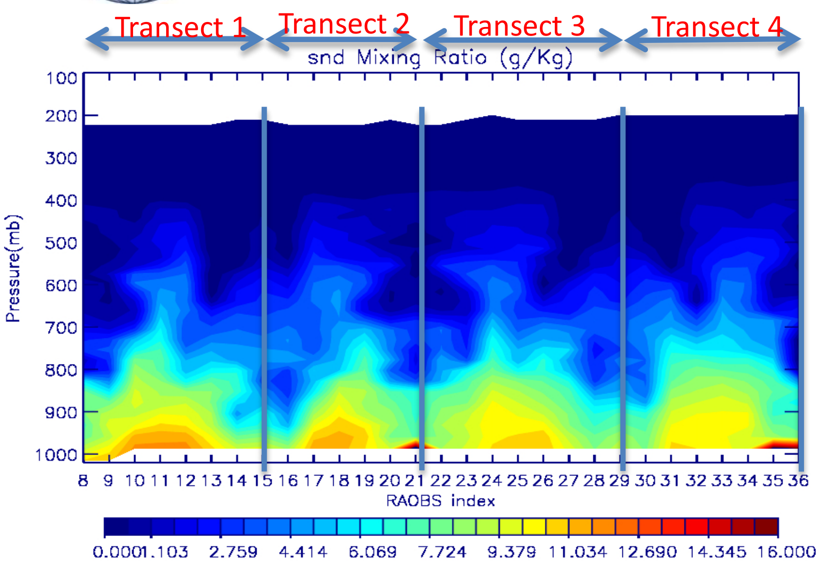
Radiosonde measurements from CalWater 2015 February 6th test case



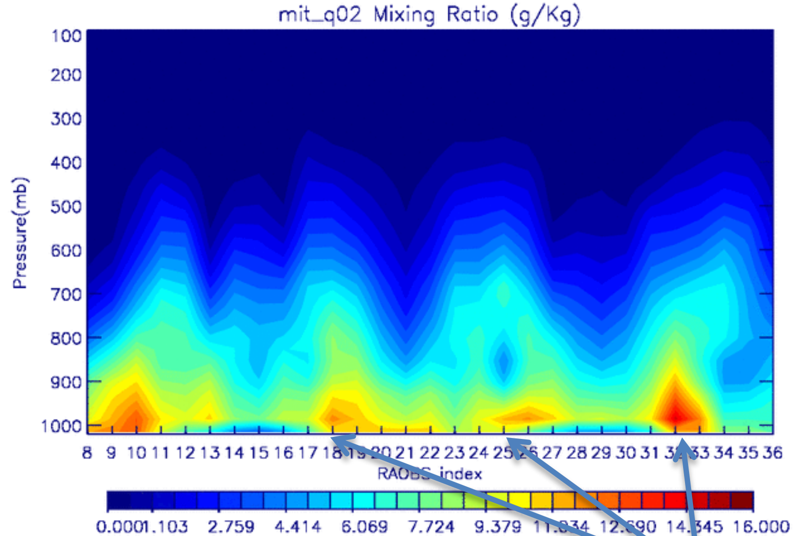


- ~ 4 hours flight with 4 transects across the river capturing pre, in and post river environment as the river quickly approaches the US West coast
- Good spatial and temporal matching with NPP (drop sonde location 19 is ~ 3.2 minutes ahead of over pass

Towards new a-priori choices

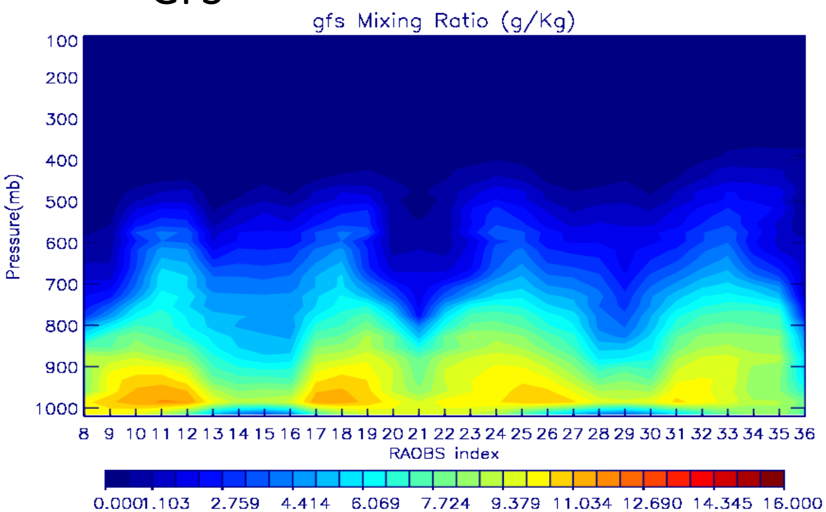


MW-only retrieval

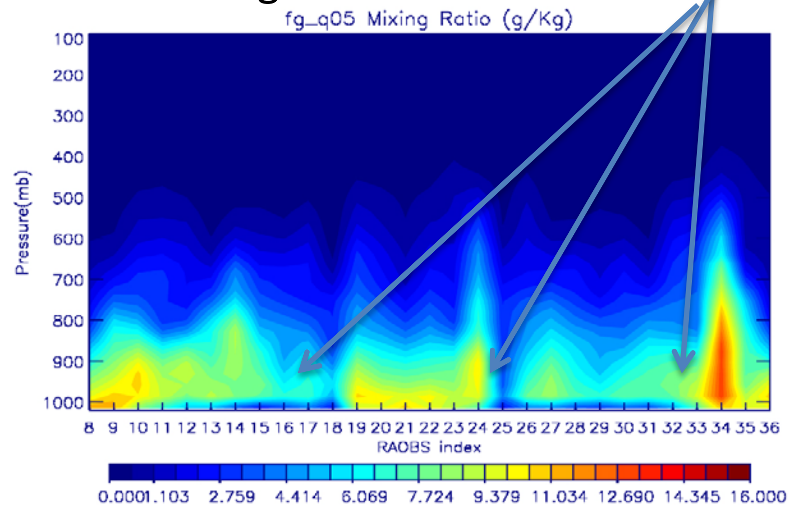


Precipitating cases

GFS



Linear Regression



Put it all together: ongoing work and future development

- Once you have a formal a-priori (ongoing work) and a full representation of the measurement error covariance (ongoing work), you can compute formal (*i.e. mathematical, non ad-hoc*) error estimates of the retrieval products.

$$\delta x \delta x^T = (K^T S_y^{-1} K + S_a^{-1})^{-1}$$

- **Future development**: we will add error estimates in addition to current QC parameters in the NUCAPS output package.
- Error estimates are essential to provide the level of confidence on any climate application.
- In essence, computing formal error estimates represents the definition of a climate quality algorithm.
- "It can be argued that a retrieval method without an error analysis and characterization is of little value" (Rodgers, 2000).

Back up slides

Quality Control

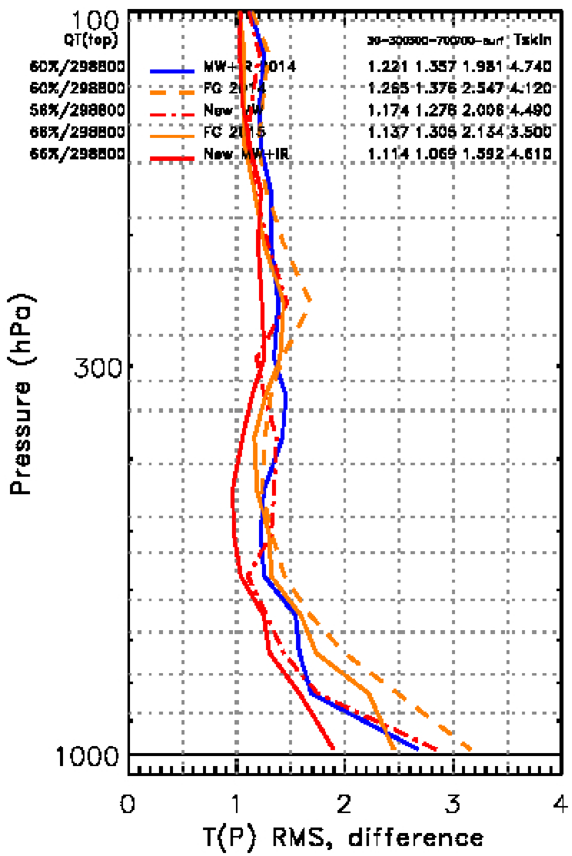
- NUCAPS QCs use the physical approach of the AST version 4: residuals, degrees of freedom, comparison between MW-only and MW+IR retrieval, MW residuals using the MW+IR retrieval. No error estimates are being distributed yet.
- Future development: we will output the same QC parameters as in current NUCAPS 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

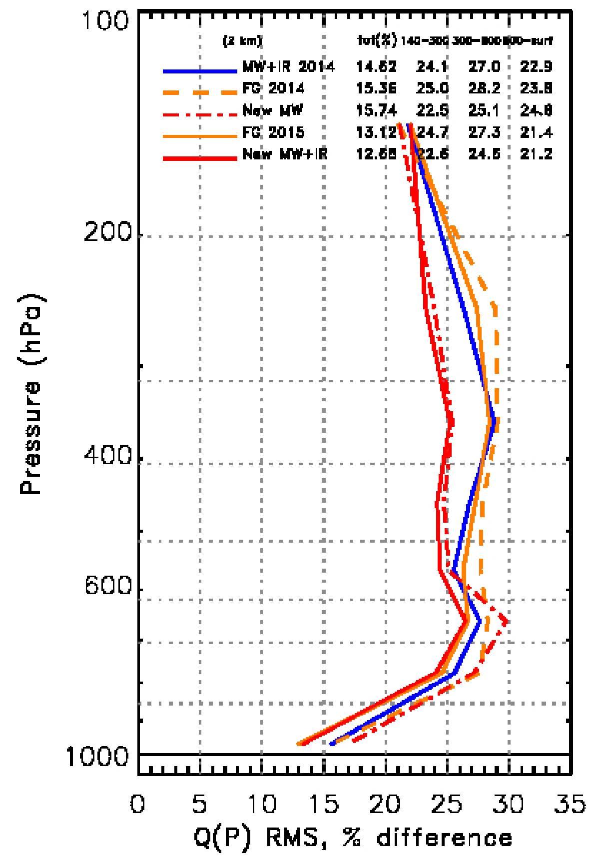
MW+IR Module

- 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



The inversion problem

- Bayesian approach: we have some prior understanding or expectation about some quantity and want to update this knowledge in light of a measurement.
- Imperfect prior knowledge can be described as a probability density function (P(x)) over the state space.
- A measurement, also imperfect due to experimental error, can be quantified as a pdf, P(y), in measurement space.
- The inversion problem: *how does the measurement pdf maps into state space and combines with prior knowledge?*
- *Baye's theorem:*

A posteriori pdf of the state vector

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$

A priori pdf of the state vector

The inversion problem as a linear problem with Gaussian statistics

- Linearization of the retrieval equation: the forward model *F* is linear in *x*

$$y = F(x) + \varepsilon = Kx + \varepsilon$$

- *Gaussian statistics for the error in the real measurement of y, if the vector state was x:*

$$P(y\,|\,x) \propto \exp\{-1/2(y-Kx)^T S_y^{-1}(y-Kx)\}$$

$$Sij = (yi - yi)(yj - yj)$$

- *Gaussian statistics of the a priori knowledge of the vector state* x_a*:*

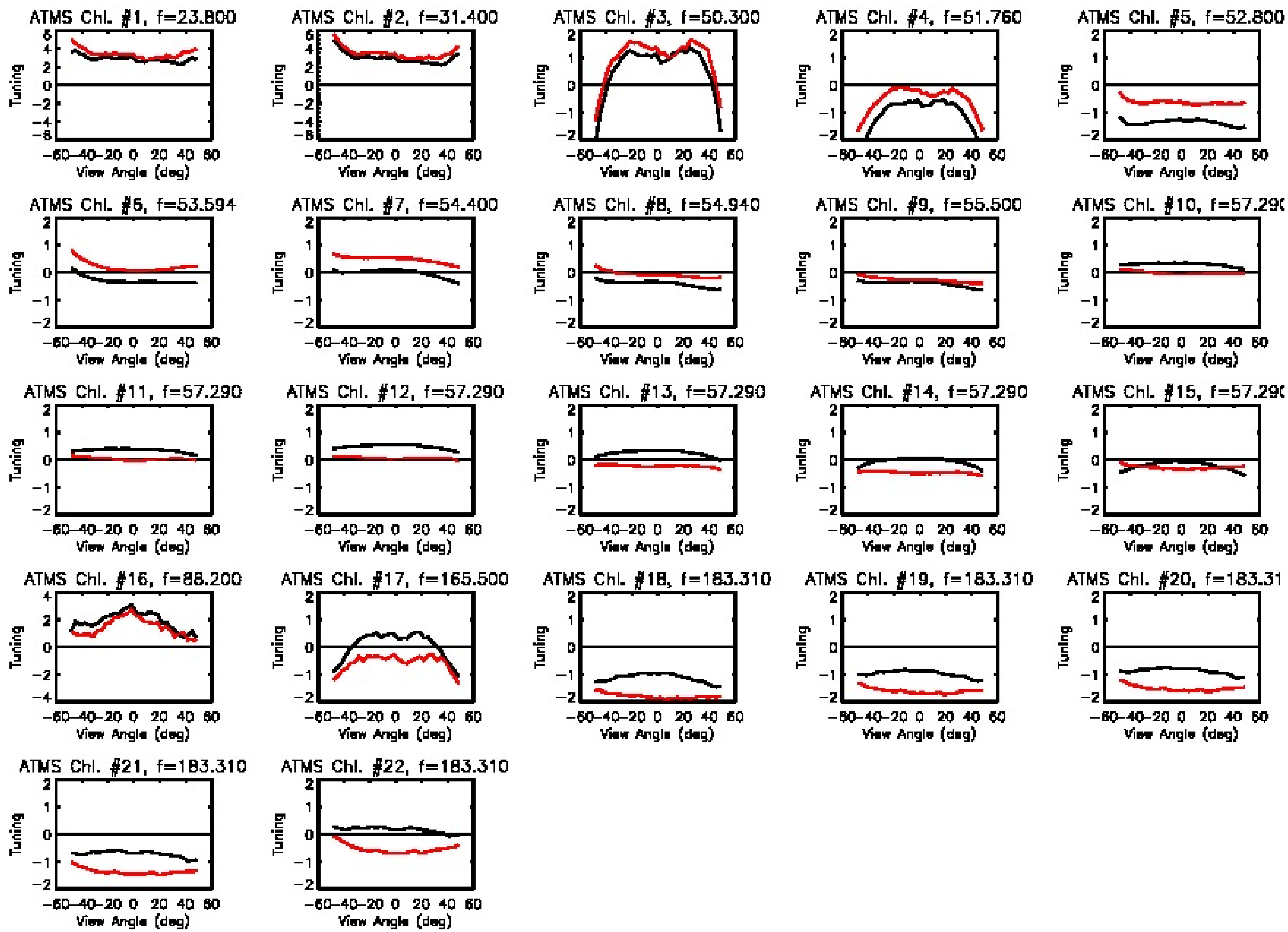
$$P(x) \propto \{-1/2(x-x_a)^T S_a^{-1}(x-xa)\}$$

$$Sa = (xi - xi)(xj - xj)$$

183GHz bias problem

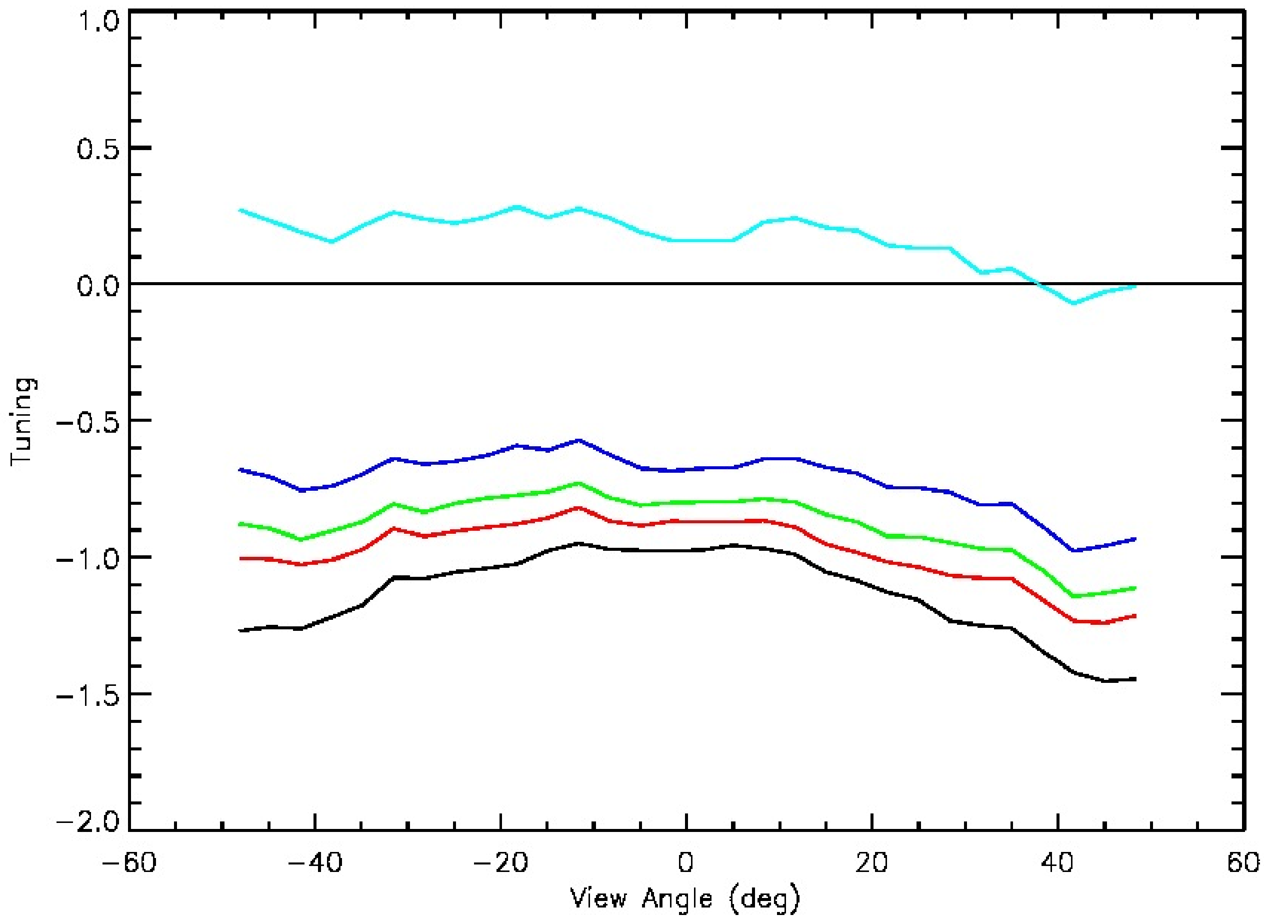
- Ongoing discussion on the sensitivity peak height dependent bias in the 183GHz band
  - OBS-CALC bias computation is observed to increase with lower peaking 183GHz channels
  - Problem is observed across all current forward models and MW instruments (AMSU, SAPHIR, ATMS)
  - Problem is observed on both 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)



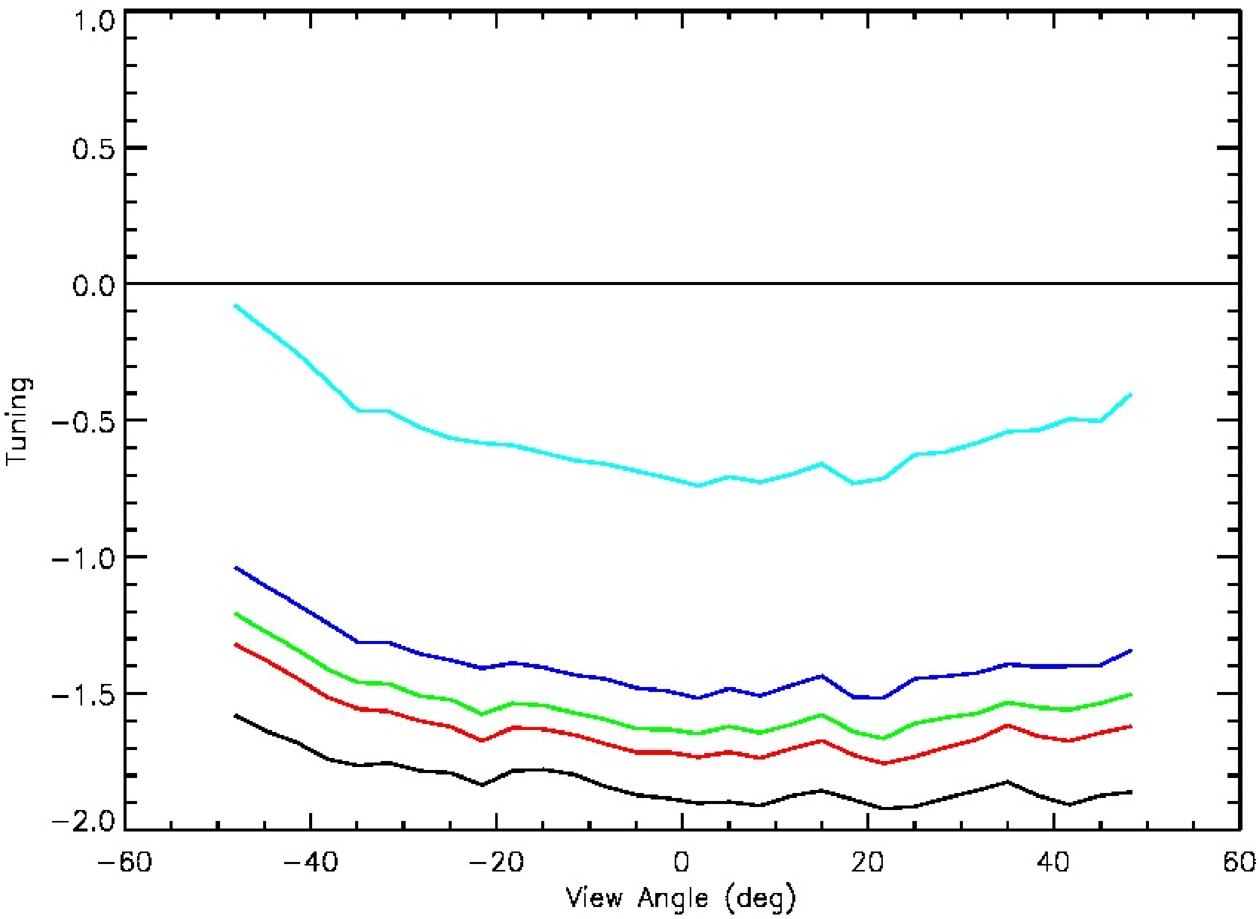
183 GHz bias (OBS-CALC): TDR cases

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



183 GHz bias (OBS-CALC): SDR cases

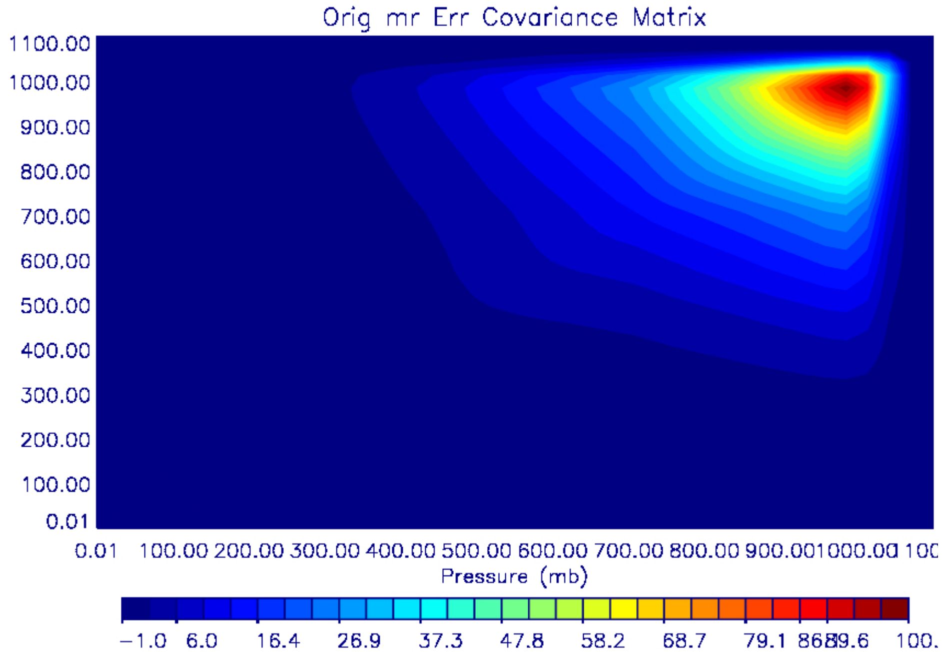
ATMS Chn. 22
ATMS chn. 21
ATMS chn. 20
ATMS chn. 19
ATMS chn. 18

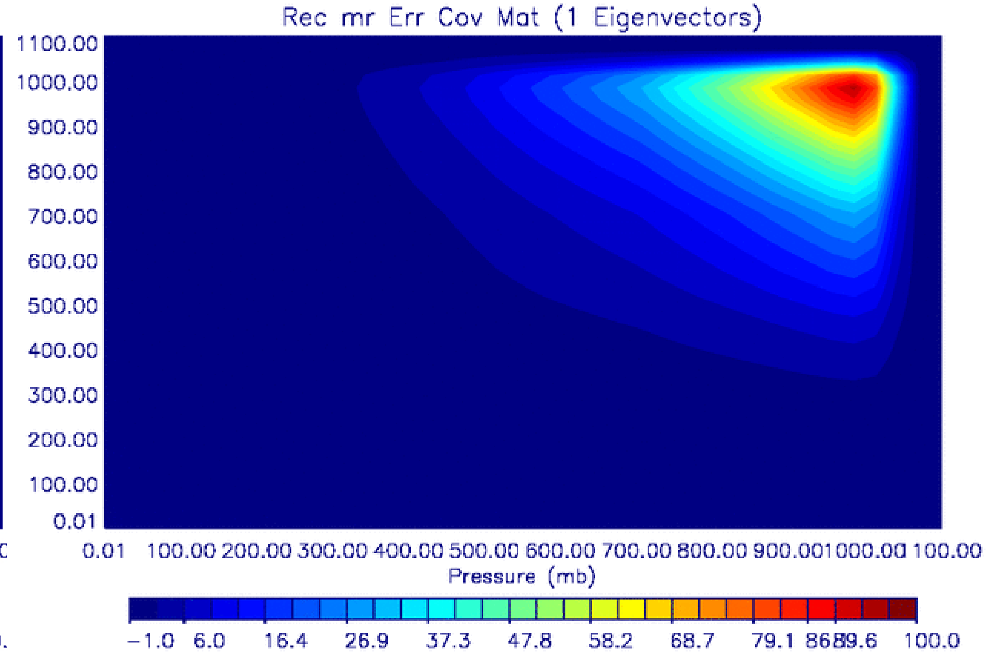


Summary

- 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). Validation using best estimates of atmospheric state.
  - Eumetsat heperspectral algorithm comparison (workshop in Potenza, Italy, late November 2015).

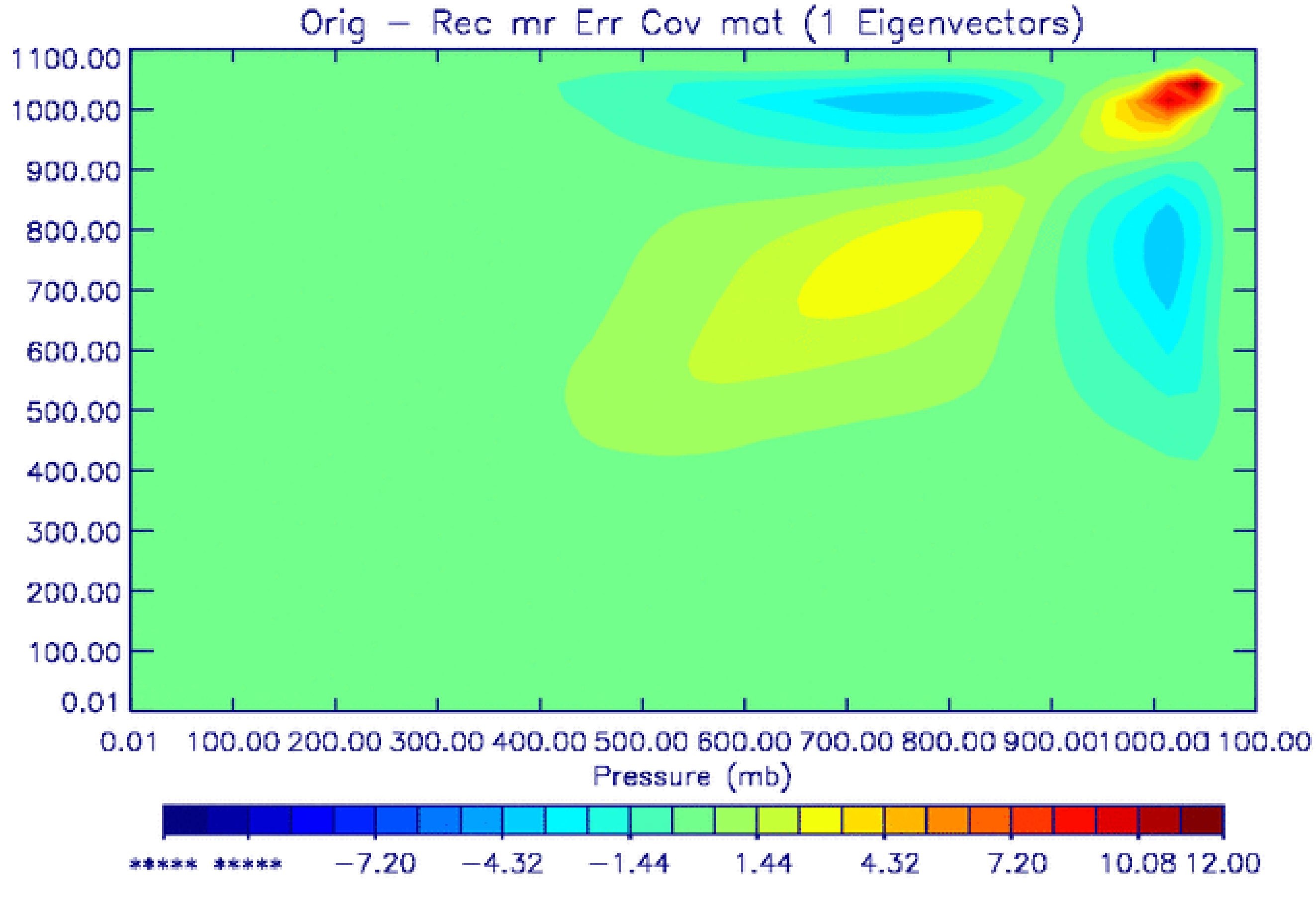
Original vs Reconstructed Water vapor Error Covariance Matrix





- Focus day: 2013/04/15/
- Water vapor mix ratio (kg/kg) error:
  mr err δq_L = Ret mr – "truth" 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