Information Content: Verticality

Eric Maddy, Chris Barnet

Outline

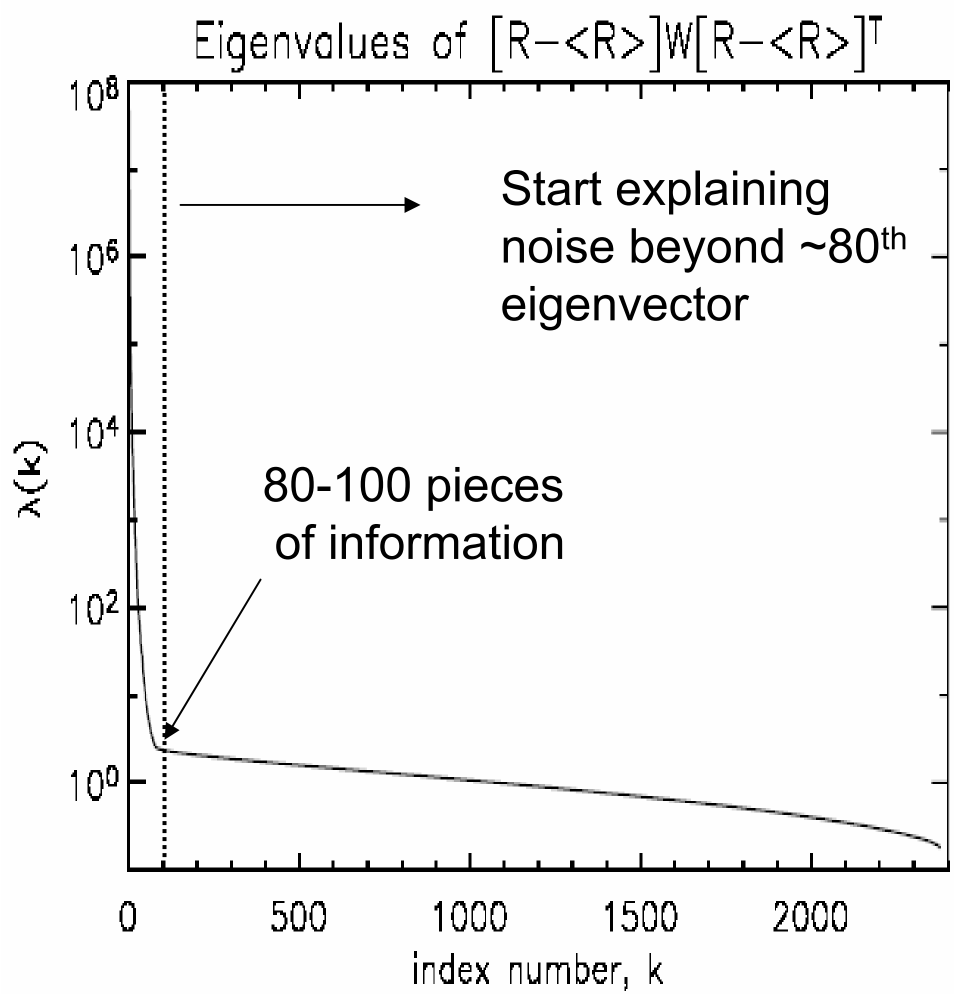
- What is information content?
  - Retrieval Theory 101
  - Overview of verticality
- Uses of Averaging Kernels
  - Assessment of retrieval function spacing
  - Retrieval Resolution
  - Statistics using Averaging Kernels
- Discussion of results

What is information content?

- Number of independent degrees of freedom
- For AIRS RTA we require:
  - 100 T(p), q(p), O3(p), CO(p), CH4(p), CO2(p), Tsurf, emissivity, etc.
  - So that the radiance computed is accurate to ~ noise level of AIRS (0.2K)
- We cannot retrieve this many independent pieces of information.

Q: How do we know how many quantities we have the right to solve for? How do we represent these quantities such that they are as independent as possible?

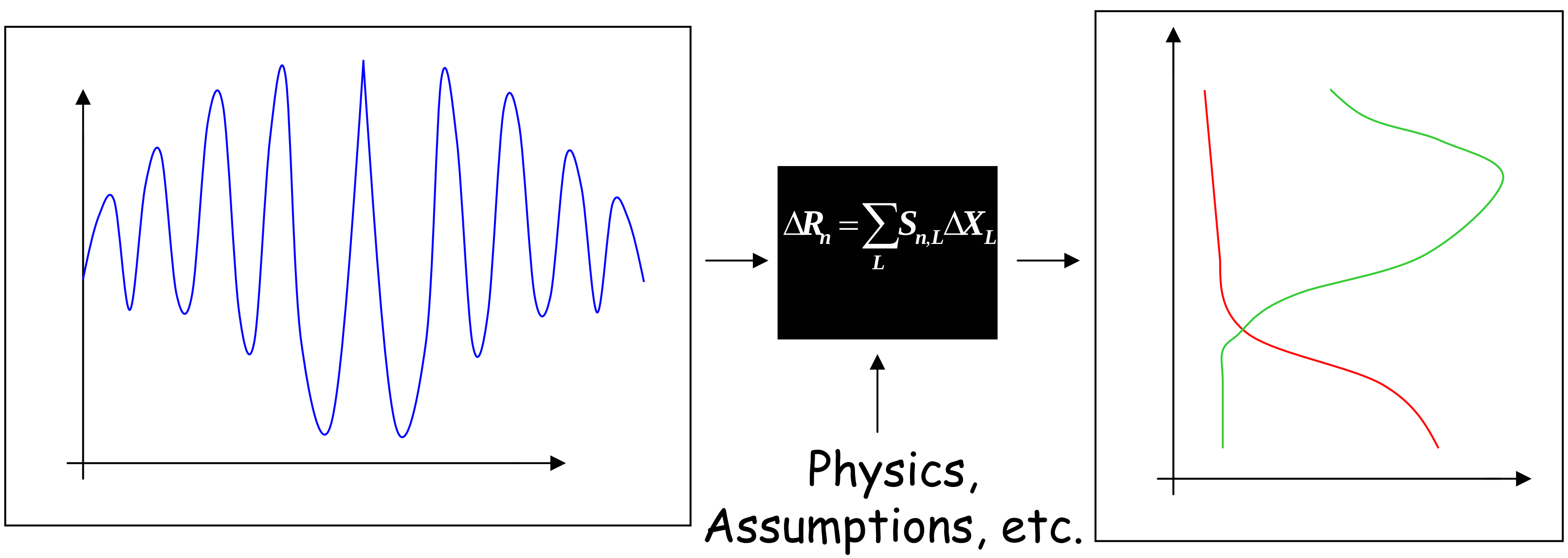
A: Averaging Kernels



rs_notes.pdf, Barnet.

Retrieval Theory 101

- In simple terms:
  - Estimation of atmospheric parameters (*e.g.* temperature profile, water vapor amount, *etc*.) from remotely sensed measurements.



Retrieval Theory 101

- In mathematical terms:
  - The simple treatment of the problem starts from the unconstrained solution to a linear problem:

Linear Taylor expansion of our measurement

$$\boldsymbol{R}_n = \boldsymbol{R}_n^0 + \frac{\partial \boldsymbol{R}_n}{\partial \boldsymbol{X}_L}(\boldsymbol{X}_L - \boldsymbol{X}_L^0) \quad ; \quad \boldsymbol{S}_{n,L} \equiv \frac{\partial \boldsymbol{R}_n}{\partial \boldsymbol{X}_L}$$

Derivatives of our measurement wrt parameters being solved for.

We can write this in matrix form:

$$\Delta \boldsymbol{R}_n = \boldsymbol{S}_{n,L} \Delta \boldsymbol{X}_L$$

$$\boxed{\Delta \boldsymbol{X}_L = [\boldsymbol{S}_{L,n}^T \boldsymbol{S}_{n,L}]^{-1} \boldsymbol{S}_{L,n}^T \Delta \boldsymbol{R}_n}$$

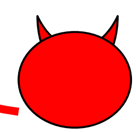
Retrieval Theory 101

- In mathematical terms:
  - The simple treatment of the problem starts from the unconstrained least squares solution to a linear problem:

$$\Delta \boldsymbol{X}_L = [\boldsymbol{S}_{L,n}^T \boldsymbol{S}_{n,L}]^{-1} \boldsymbol{S}_{L,n}^T \Delta \boldsymbol{R}_n$$

However, due to measurement noise and the fact that in most cases the problem is underdetermined (limited number of independent pieces of info.), we are required to stabilize the inverse of the covariance of our measurement:

$$\Delta \boldsymbol{X}_L = [\boldsymbol{S}_{L,n}^T \boldsymbol{S}_{n,L} + \boldsymbol{H}_{L,L}]^{-1} \boldsymbol{S}_{L,n}^T \Delta \boldsymbol{R}_n$$



in the details

Averaging Kernels 101

Regularized solution to the linear problem:

$$\Delta \boldsymbol{X}_{L} = [\boldsymbol{S}_{L,n}^{T}\boldsymbol{S}_{n,L} + \boldsymbol{H}_{L,L}]^{-1}\boldsymbol{S}_{L,n}^{T}\Delta\boldsymbol{R}_{n}$$

Where as before we define:

$$\Delta\boldsymbol{R}_{n} = \frac{\partial\boldsymbol{R}_{n}}{\partial\boldsymbol{X}_{L}}(\boldsymbol{X}_{L} - \boldsymbol{X}_{L}^{0}) \quad ; \quad \boldsymbol{S}_{n,L} \equiv \frac{\partial\boldsymbol{R}_{n}}{\partial\boldsymbol{X}_{L}}$$

Combining the two equations above yields:

$$\tilde{\boldsymbol{X}}_{L} = \boldsymbol{X}_{L'}^{0} + [\boldsymbol{S}_{L,n}^{T}\boldsymbol{S}_{n,L'} + \boldsymbol{H}_{L,L'}]^{-1}\boldsymbol{S}_{L,n}^{T}\boldsymbol{S}_{n,L'}(\boldsymbol{X}_{L'} - \boldsymbol{X}_{L'}^{0})$$

Averaging Kernels 101

Combining the two equations above yields:

$$\tilde{X}_L = X_{L'}^0 + [S_{L,n}^T S_{n,L'} + H_{L,L'}]^{-1} S_{L,n}^T S_{n,L'} (X_{L'} - X_{L'}^0)$$

We can write this more compactly as:

$$= X_{L'}^0 + A_{L,L'}(X_{L'} - X_{L'}^0); \quad A_{L,L'} = \frac{\partial \tilde{X}_L}{\partial X_{L'}}$$

Averaging Kernel

$$\text{as } H_{L,L'} \to 0, \; A_{L,L'} \to I_{L,L'}, \; \tilde{X}_L \to X_{L'}$$

$$\text{as } H_{L,L'} \to \infty, \; A_{L,L'} \to 0_{L,L'}, \; \tilde{X}_L \to X_{L'}^0$$

The magnitude of $H_{L,L'}$ restricts the solution to "stick" to X_L^0 or move to X_L

The magnitude of $A_{L,L'}$ relates how much of X_L^0 is in our solution. -> fraction of the radiances we believed

What is verticality? Replay the theory in words

- Averaging kernels are a linear representation of the vertical weighting of retrievals.
  - Related to the amount of information determined from the radiances and how much is due to the first guess [Rodgers, 1976].
    - To some degree avoids aliasing comparisons of in situ measurements vs. retrievals due to incorrect first guesses.
    - Enables assessment of where vertically we have information.
  - Related to the vertical resolution of retrievals [Backus and Gilbert, 1969; Rodgers, 1976; Purser and Huang, 1993]
  - Required by modelers to properly use AIRS trace gas products.
  - Enables assessment of retrieval skill on a case by case basis.
- In the IDEAL case (no damping): A = I : the identity matrix

What is verticality?

For the AIRS science team algorithm:

- Trapezoidal functions and damping introduce correlations (off-diagonal terms).
  - Damping in general causes the vertical weighting to fall short of the ideal case (i.e. diagonal terms of $A(k,k) < 1$)
  - The sum along any row (or column) can be though of as the fraction of information determined from the radiances.
    - Only as good as our internal error estimates as damping is determined from them.
    - Warning : Sum can be > 1 -- Implies that the fraction we believe any one function in the parameter space can be greater than 100%.

Very important

Verticality (cont.)

For the AIRS science team algorithm:

- Iteration (esp. background term)/stepwise retrieval complicate interpretation
  - Regression FG has it's own averaging kernel. If retrieval is damped to the regression the internal averaging kernel (from phys. ret.) loses interpretation.
  - There is a cross-talk between averaging kernels that is usually not addressed properly. The temperature retrieval believes a fraction of the radiances so that the averaging kernel for products other than temperature does not exactly relate to the amount of the radiances believed.
  - Non-linearity (I won't go into this too much here) is not properly handled by the linear analysis shown in previous slides.

- Vertical weighting is strictly defined on the retrieval grid, not the RTA grid.
  - Any estimate of resolution based on the internal averaging kernels is limited by the resolution of our retrieval functions.
  - Transformations between retrieval functions and AIRS layers exist; however they assume that we can "upsample" derivatives without loss of accuracy.

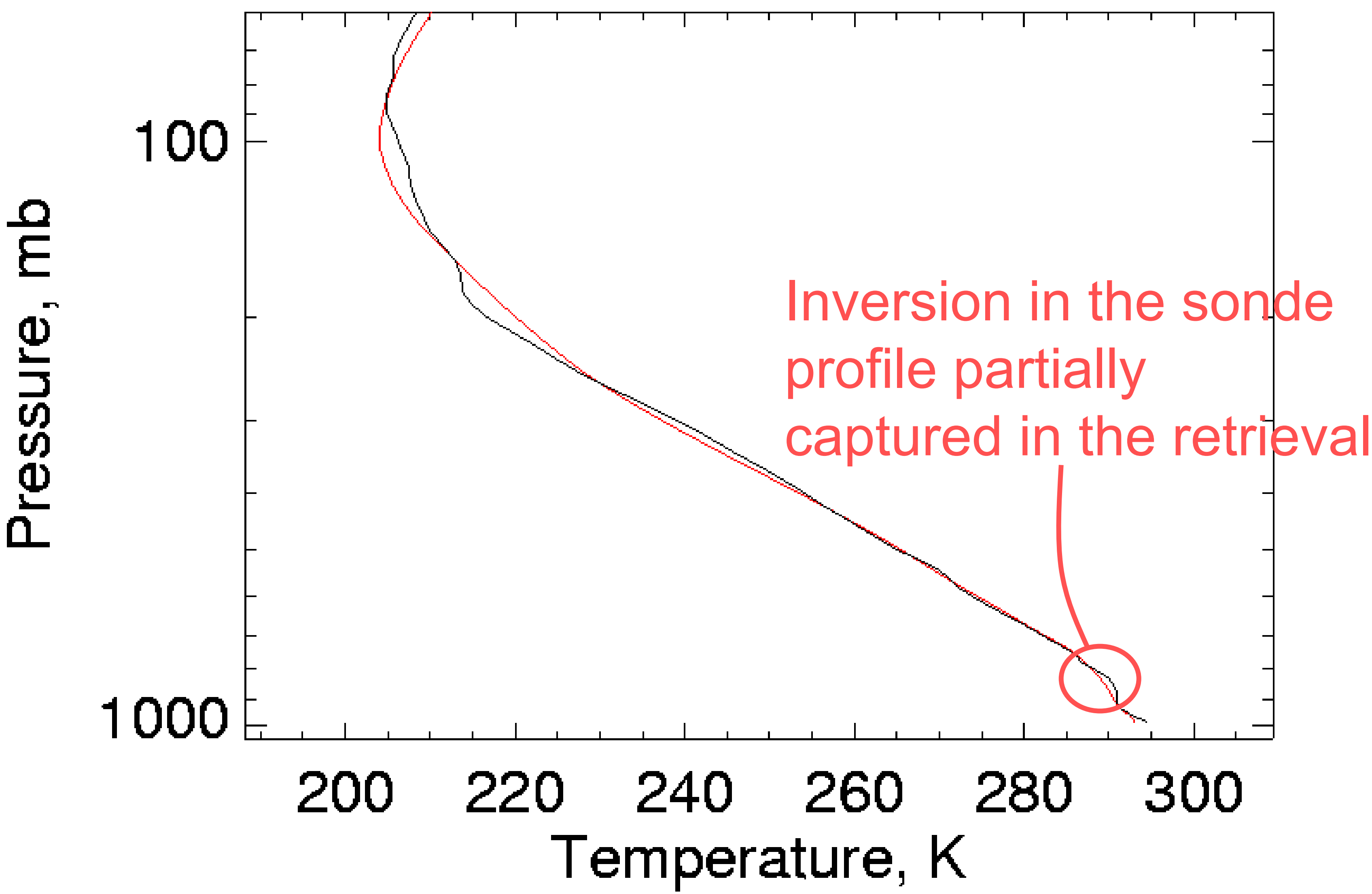
Examples of Vertical Weighting Functions

- Show information content for one sonde (RS-90 dataset):
  - 1st T(p) retrieval
  - q(p) retrieval
  - O_3(p) retrieval
  - 2nd T(p) retrieval

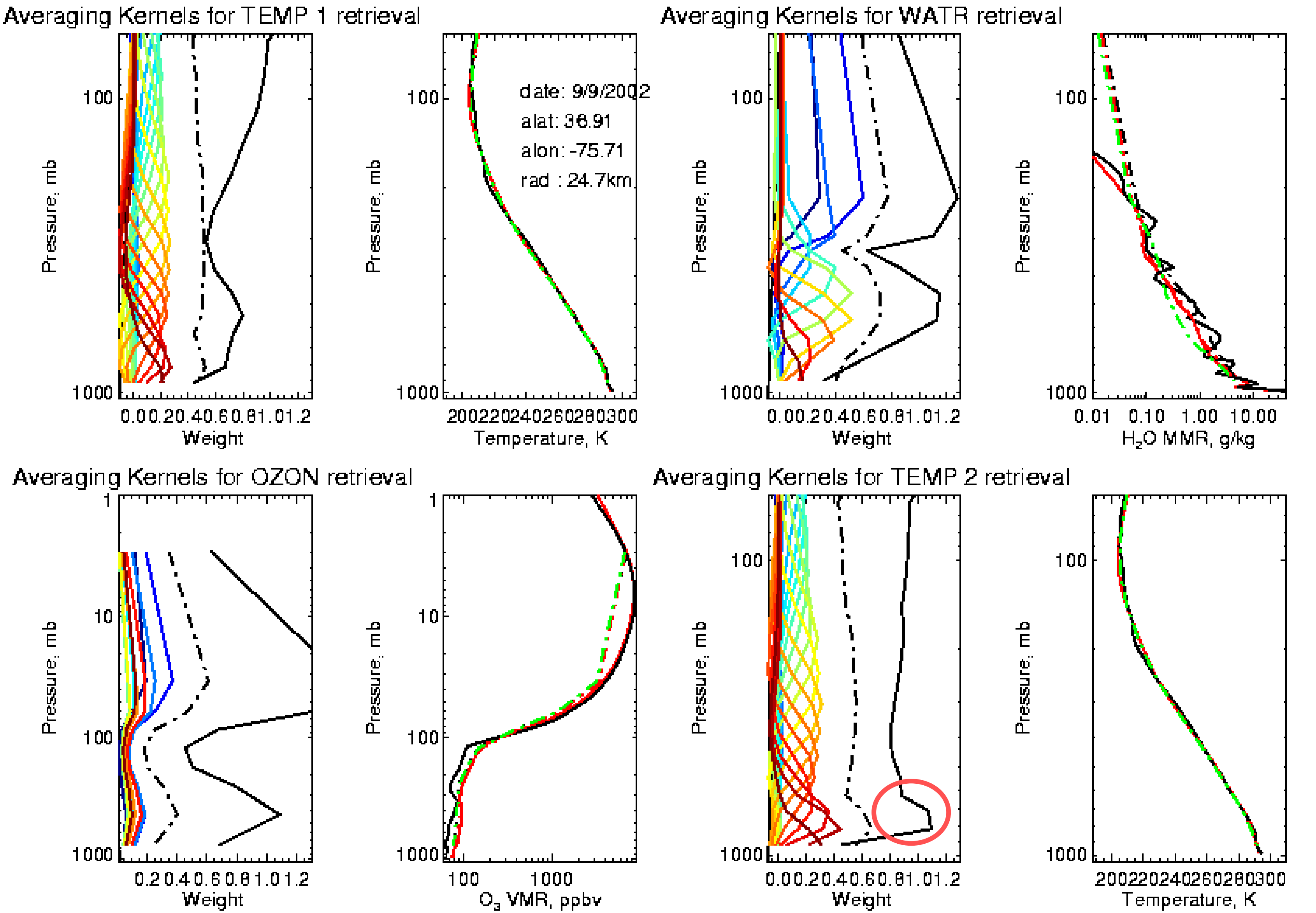
Focus on T(p) for now

- Between 1st and 2nd T(p) retrieval, O3 and q(p) retrievals will change internal error estimates.
  - 6 Water channels added to 2nd T(p) retrieval

Red: Retrieval, Black: Sonde



A matrix reflects increase of information



Brute Force Averaging Kernels

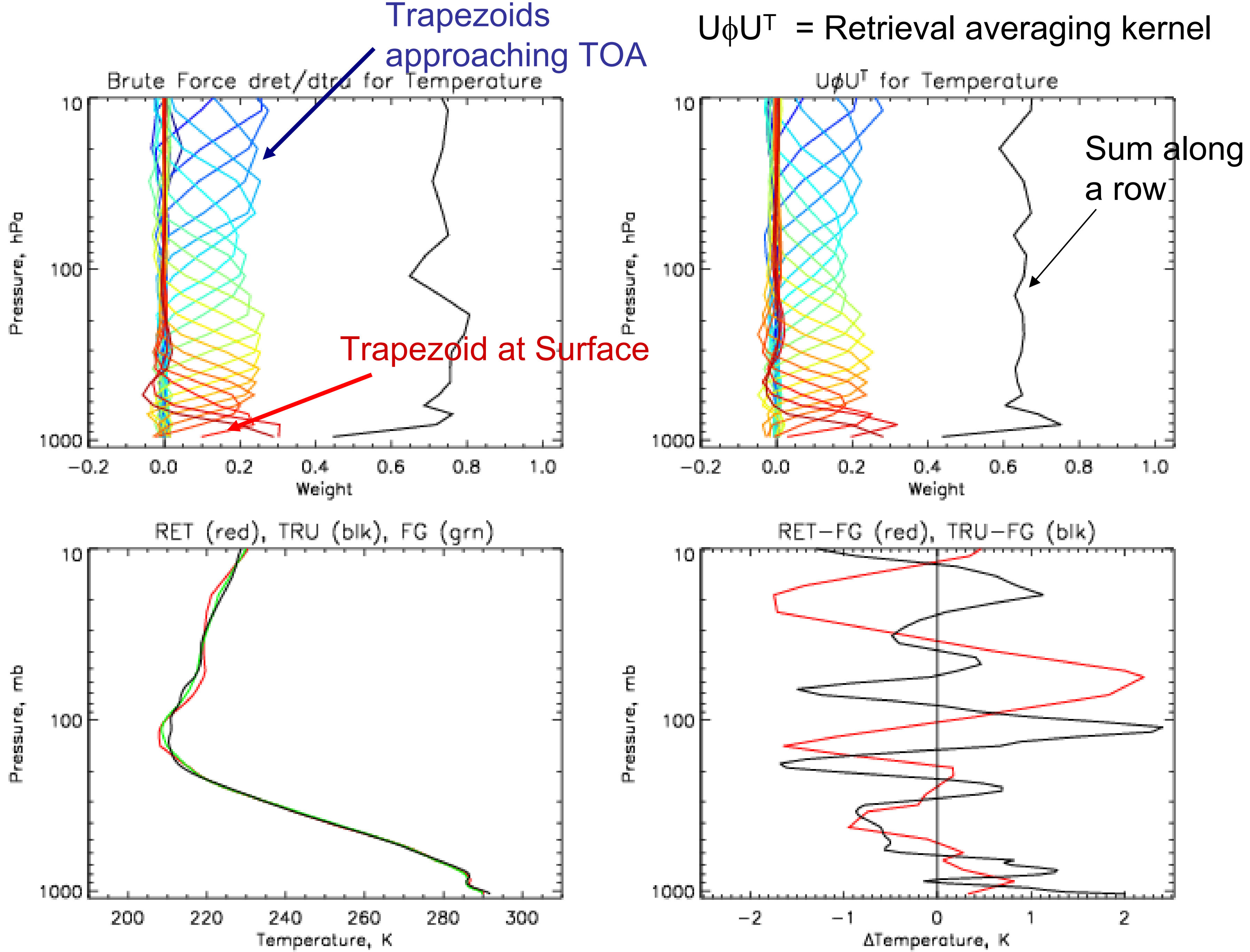
- Use small number of cases (any L2file) to assess impact of iteration (background term) and magnitude of damping on interpretation of vertical weighting.

  How well does the theory predict the behavior of the retrieval?

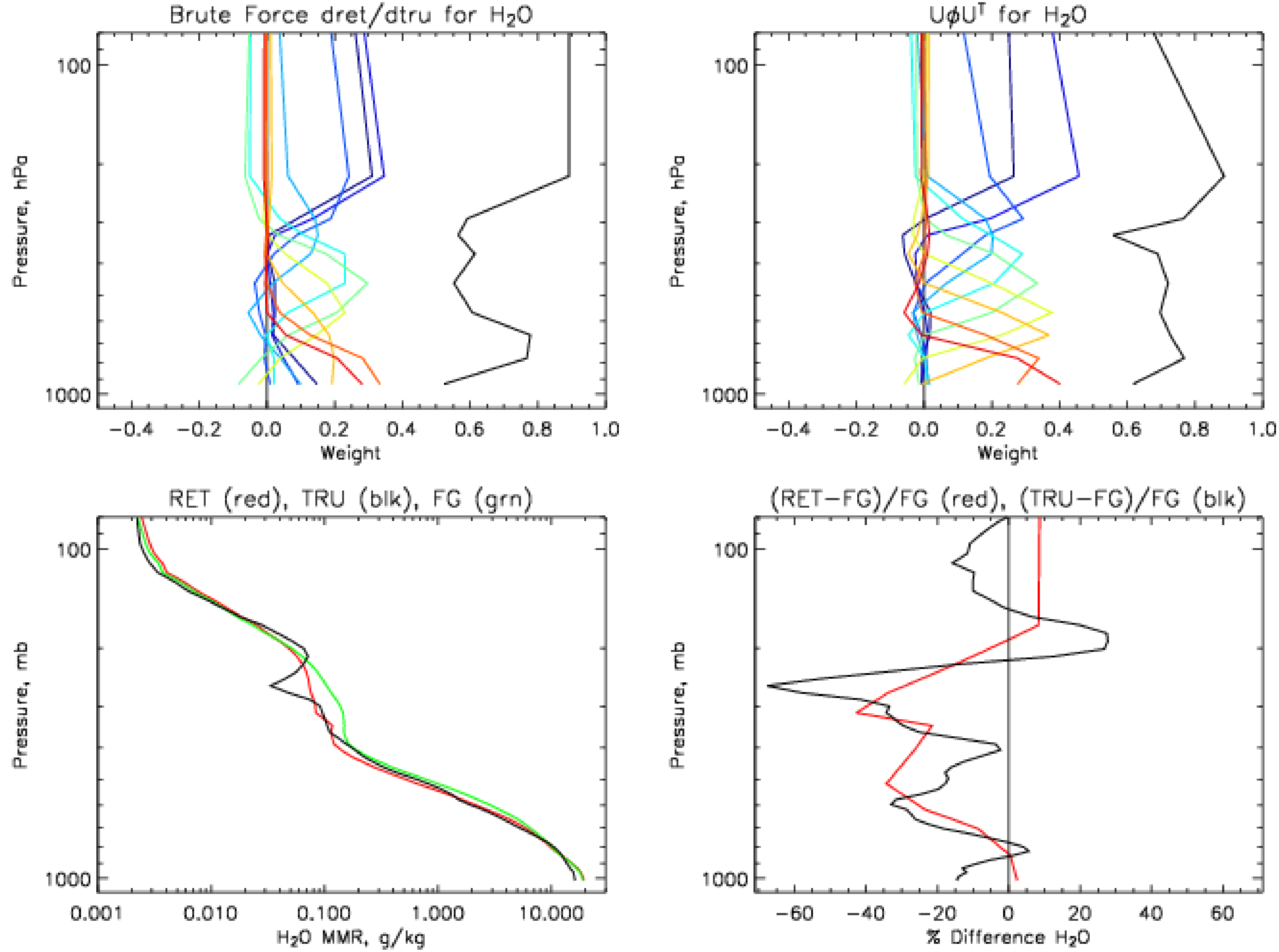
Procedure:

1. Calculate spectra for the dataset
2. Calculate spectra for the dataset using perturbations on all retrieval layers (or any layering scheme) separately.
3. Use offline retrieval system on both control and perturbed simulated spectra.
   - Calculate averaging kernel
     – $A \approx \Delta ret/\Delta tru$ = (ret(perturbed)-ret(control))/(perturbation)
4. Compare calculations of "brute force" averaging kernels to those obtained from the retrieval methodology.
   – Each were normalized in the following.

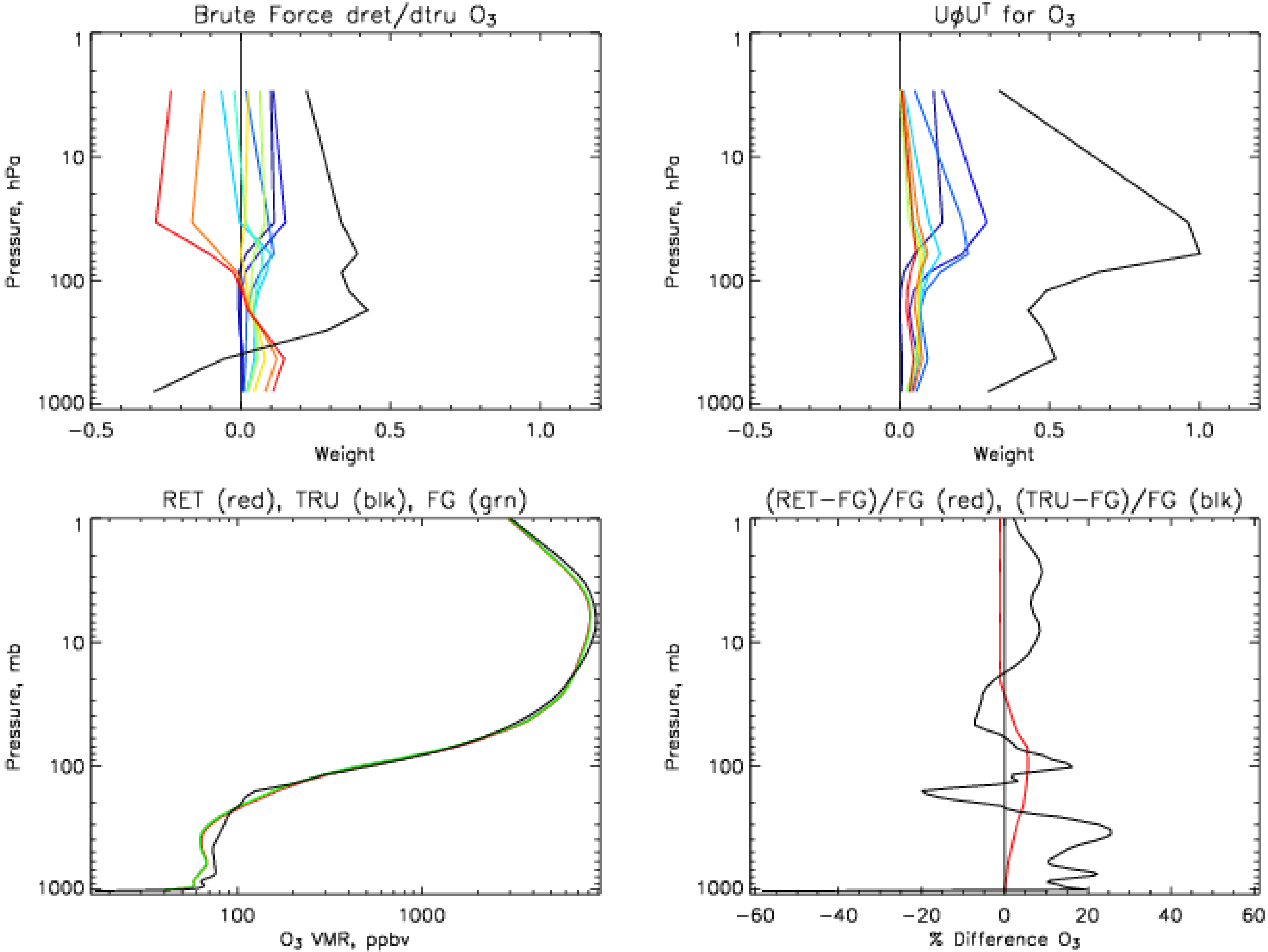
UφU^T = Retrieval averaging kernel

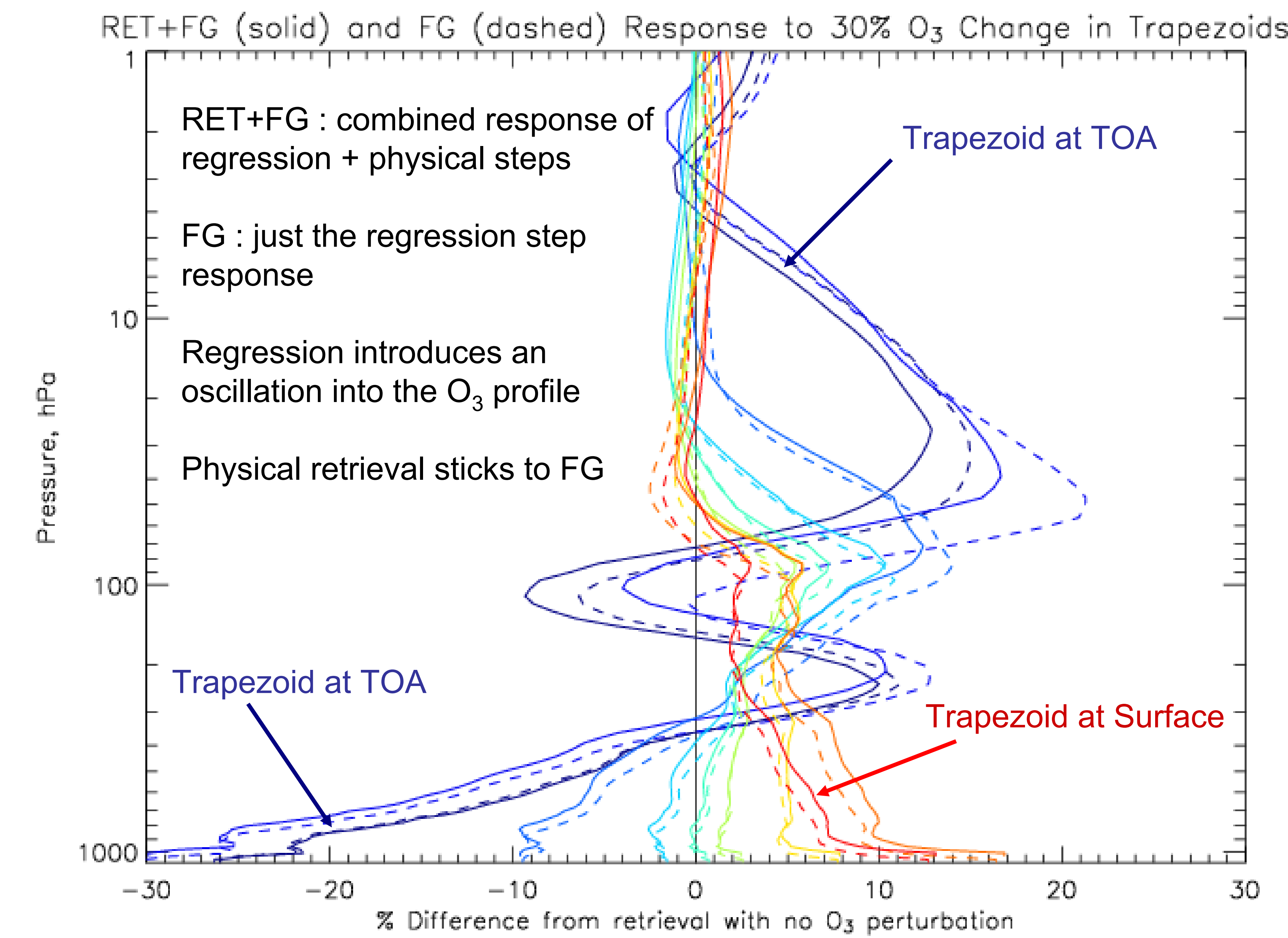


$U\phi U^T$ = Retrieval averaging kernel



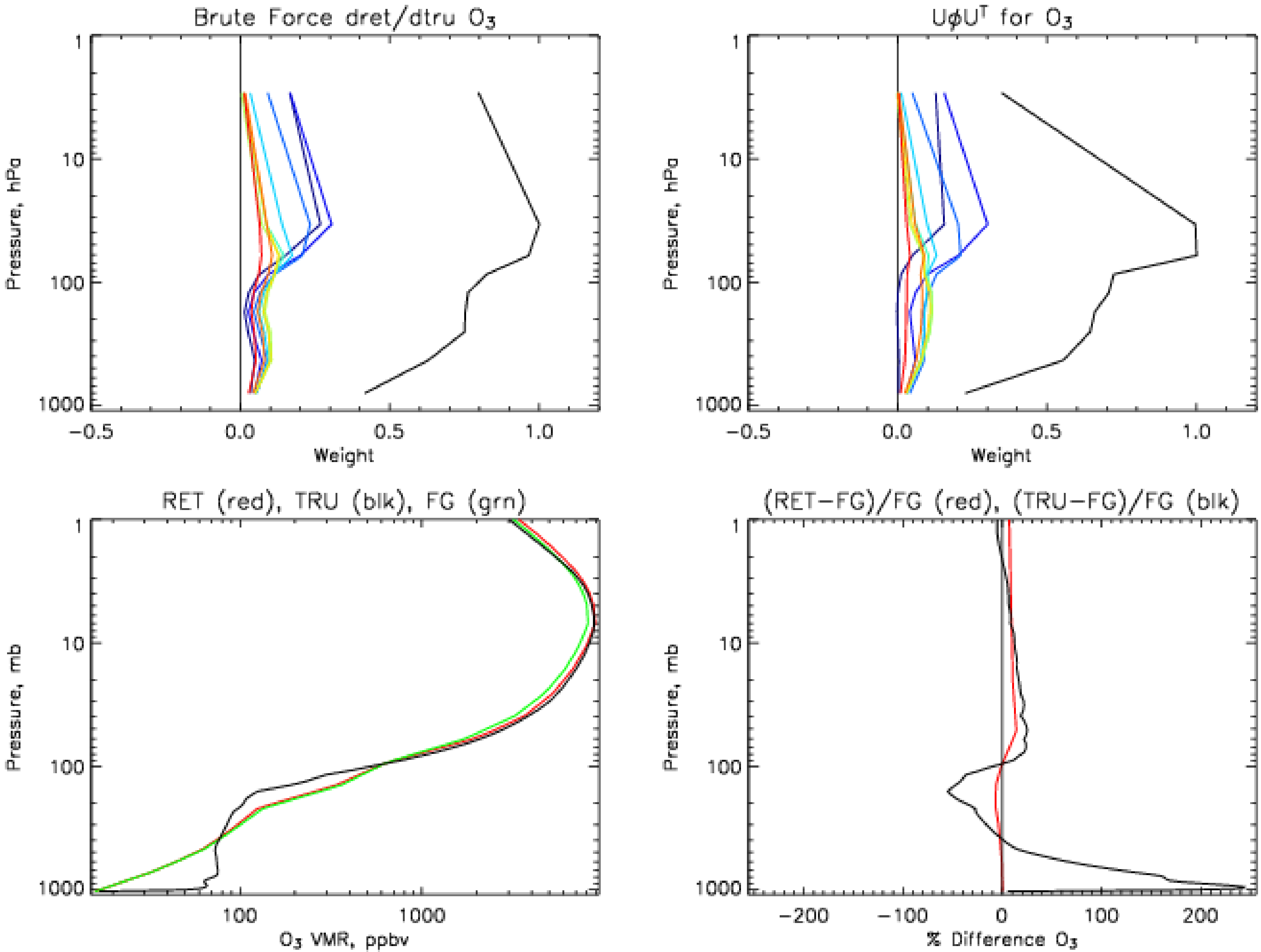
$U\phi U^T$ = Retrieval averaging kernel



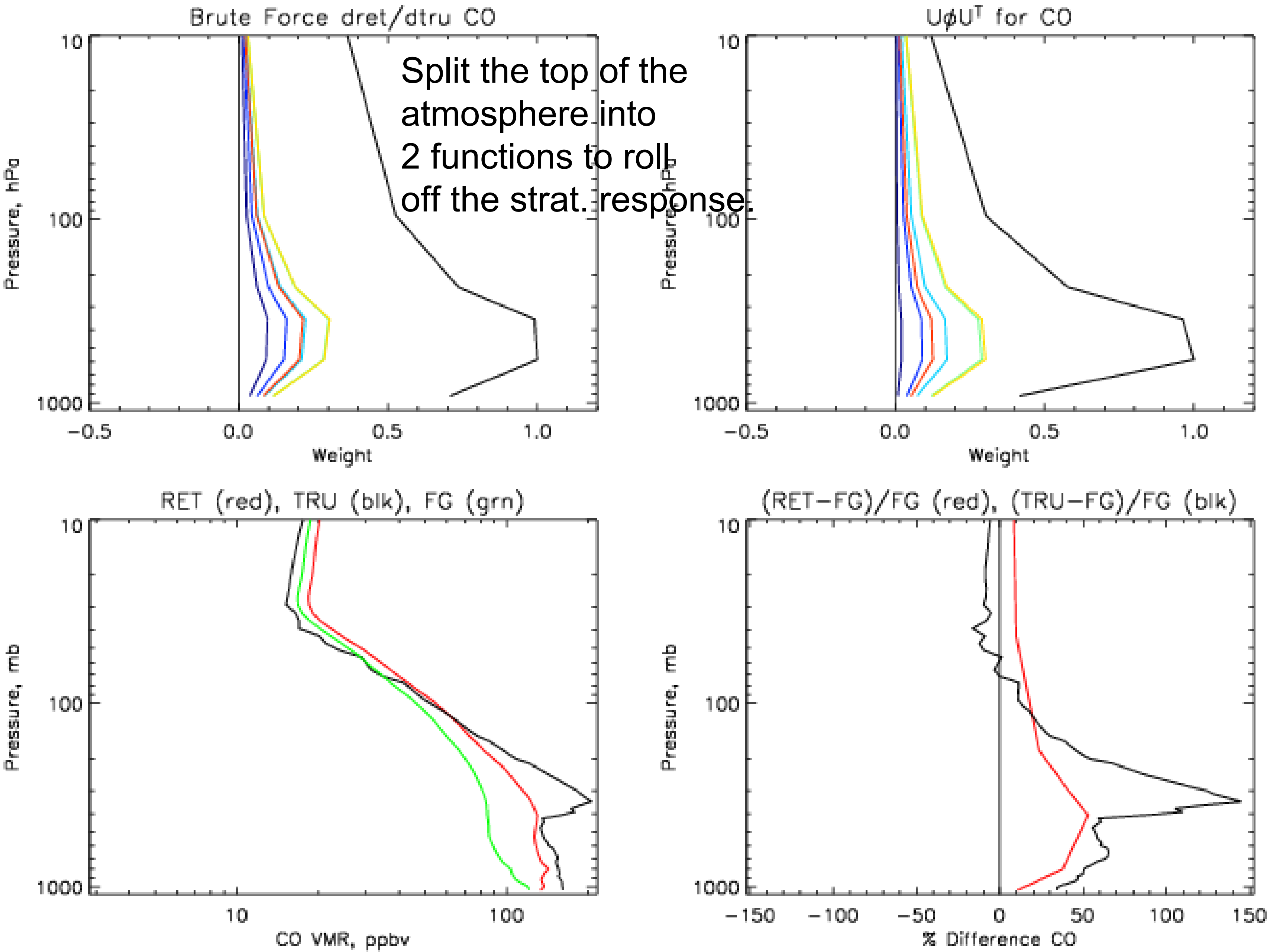


RET+FG (solid) and FG (dashed) Response to 30% O3 Change in Trapezoids
RET+FG : combined response of regression + physical steps
FG : just the regression step response
Regression introduces an oscillation into the O3 profile
Physical retrieval sticks to FG
Trapezoid at TOA
Trapezoid at TOA
Trapezoid at Surface
Pressure, hPa
1
10
100
1000
−30
−20
−10
0
10
20
30
% Difference from retrieval with no O3 perturbation

$U\phi U^T$ = Retrieval averaging kernel

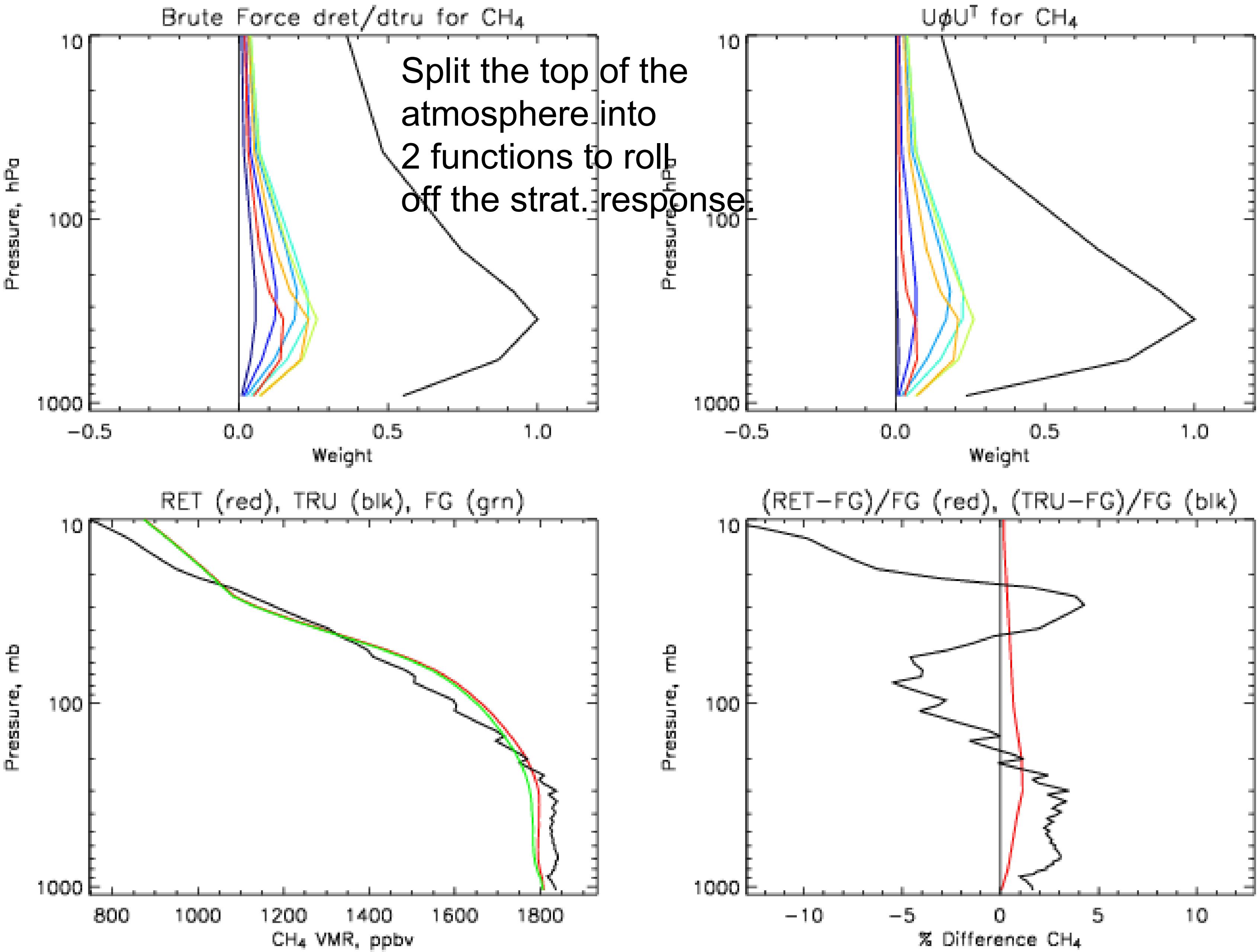


$U\phi U^T$ = Retrieval averaging kernel



$U\phi U^T$ = Retrieval averaging kernel

Split the top of the atmosphere into
2 functions to roll
off the strat. response.



Resolution estimates from brute-force averaging kernels

- Vertical resolution of any retrieval is related to the width of the kernel functions and hence averaging kernels.
  - Only averages of the true atmospheric state can be obtained by inversion
    - 1st principle of Backus-Gilbert method (*e.g.* layer to level conversion 2 pt running mean)
    - Due to the fact that kernel functions are NOT δ-functions
- We can use statistics of the brute-force averaging kernels to estimate the resolution of the retrieval in different regimes (e.g. f(lat, lapse rate, cloudiness, spectral resolution, *etc.*))
  - In the following I'll show only Polar, Midlatitude, Tropical comparisons.

Calculation of resolution

- Perturb the atmosphere in fine layers (much finer than retrieval functions)
- Calculate the 1st moment of $A^2(z,z')$:

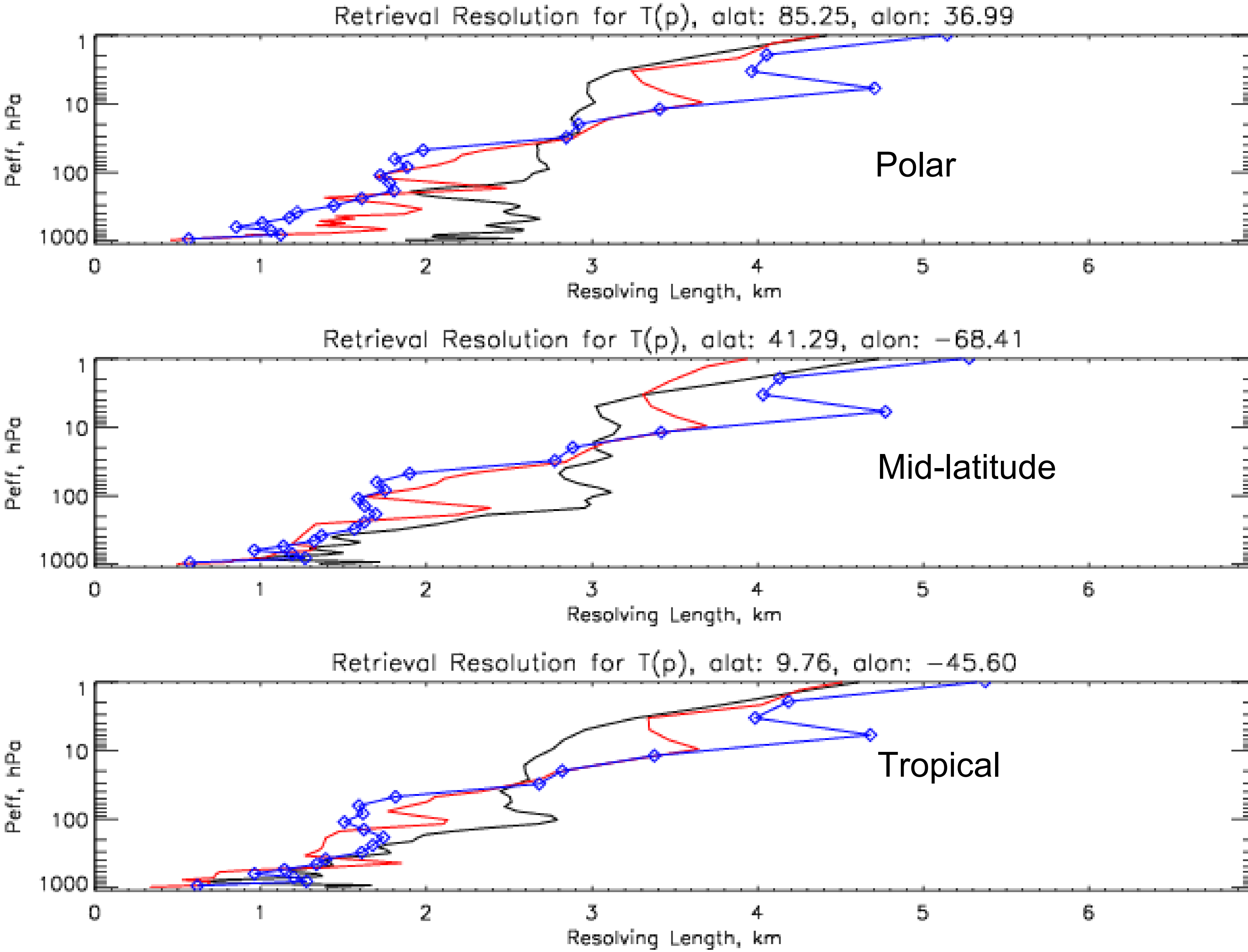
$$\boldsymbol{c}(\mathbf{z}) = \frac{\int_{z=0}^{\infty} \mathbf{z}' \cdot \mathbf{A}^2(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'}{\int_{z'=0}^{\infty} \mathbf{A}^2(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'}$$

Center or average value of $A^2(z,z')$

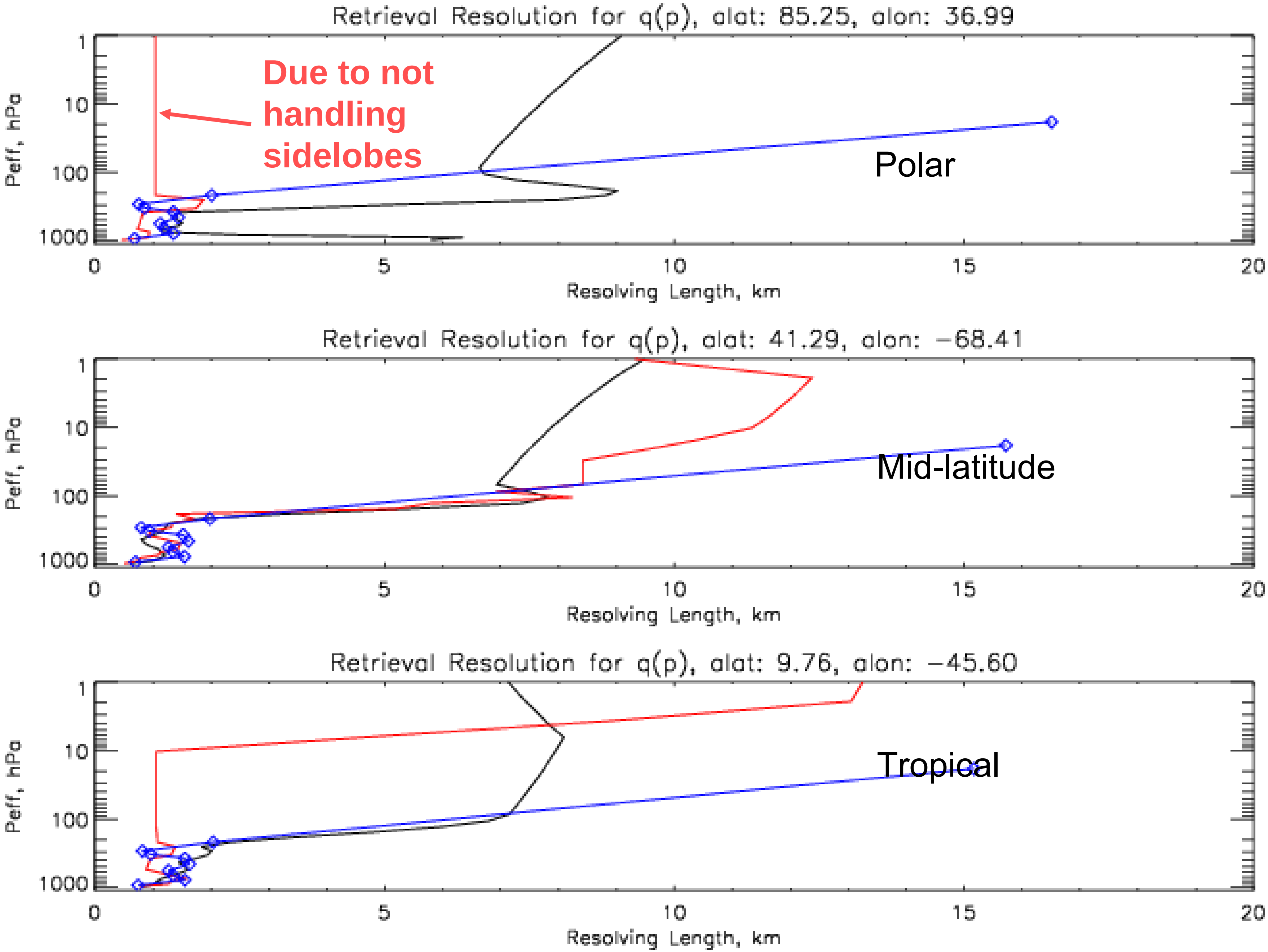
- Calculate 2nd moment centered moment of $A^2(z,z')$ about c(z):

$$\boldsymbol{l}(\mathbf{z}) = \sqrt{\frac{\int_{z=0}^{\infty} (\mathbf{z}'-\boldsymbol{c}(\mathbf{z}))^2 \cdot \mathbf{A}^2(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'}{\int_{z'=0}^{\infty} \mathbf{A}^2(\mathbf{z},\mathbf{z}')\mathrm{d}\mathbf{z}'}}$$

Std. Deviation of $A^2(z,z')$ about its central value or first moment-> resolution of the averaging kernel



Blue: RET functions, Red: HWHM $A^2_{L,L'}$, Black: $\sigma(A^2_{L,L'})$



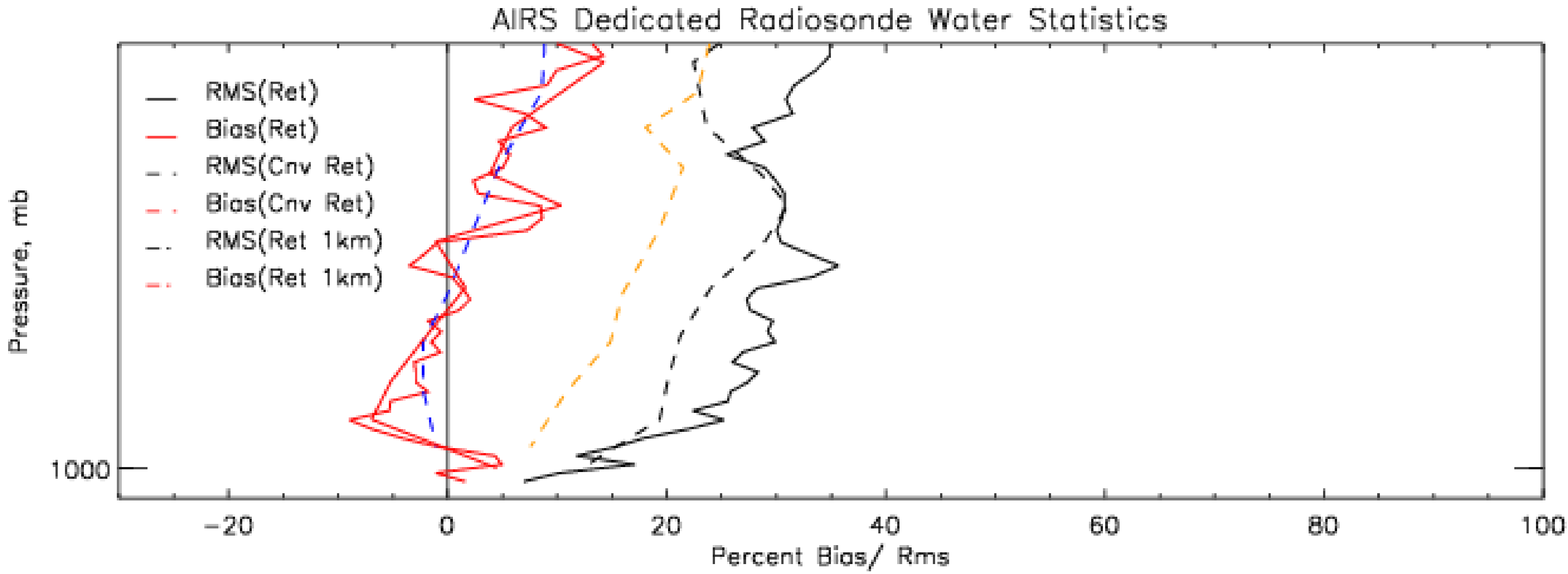
Blue: RET functions, Red: HWHM $A^2_{L,L'}$, Black: $\sigma(A^2_{L,L'})$

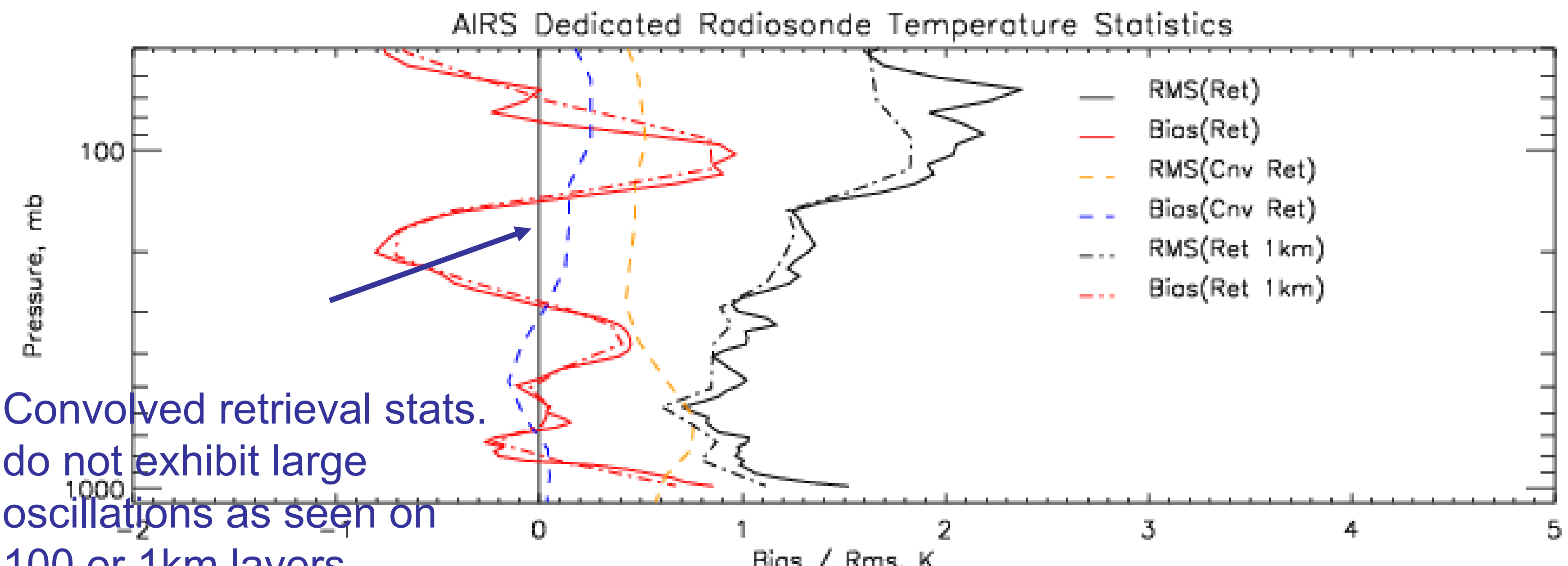
Examples of Statistics with Vertical Weighting Functions

- The information content of AIRS spectra is highly scene dependent (*e.g*. clear *vs.* cloudy, tropical *vs*. polar, ocean *vs*. land, *etc.*). Therefore, the vertical resolution and accuracy of any given retrieval is a function of scene.

- It makes sense to use an estimate of the information content on a case-by-case basis for comparisons of retrievals to correlative measurements.
  - Use the averaging kernel to convolve the correlative measurement such the this profile is more comparable to what the retrieval would “see” given that profile.

- The followings stats are for T(p) because it is the most mature AIRS product.
  - GSFC v4.2 emulation.
  - The statistics shown are for the RS-90 dataset currently being used for tuning. No clouds!
  - Enable assessment of the skill of the retrieval with and without regard for the internal estimate of the skill of the retrieval. How good are our internal CCR error estimates? Clear cases need not worry about them.





Convolved retrieval stats. do not exhibit large oscillations as seen on 100 or 1km layers

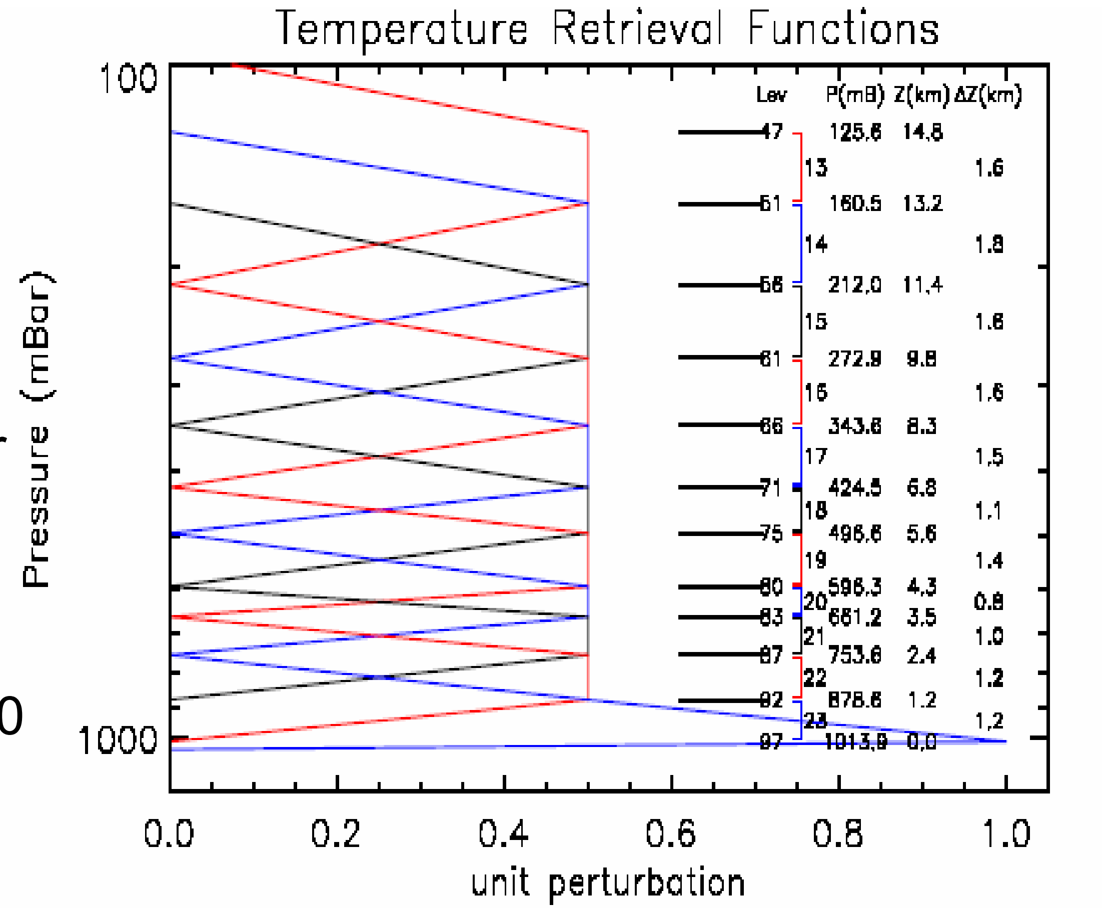
Discussion

- Averaging Kernels provide useful diagnostic capabilities.
  - Increased vertical resolution when inversions are detected.
- Brute force averaging kernels
  - Reveal whether retrieval functions are truly optimized.
  - Enable evaluation of retrieval resolution in different atmospheric conditions.
  - Enable assessment of the retrieval on a case-by-case basis using it’s own estimate of skill.

Questions?

Background

- AIRS Science team algorithm uses coarse layer perturbations or trapezoids to calculate derivatives:
  - Calculation of finite difference derivatives requires 2 calls to RTA per derivative – time consuming!
  - We cannot independently solve for 100 layer values per retrieved quantity – ~80 eigenvectors can explain most of the variance in the spectra.



rs_notes.pdf, Barnet.