Dipartimento di Ingegneria e Fisica dell'Ambiente

Consiglio Nazionale delle Ricerche
Istituto di Metodologie per l'Analisi Ambientale

Neural Network Approach to the Inversion of High Spectral Resolution Observations for Temperature, Water Vapor and Ozone

- *DIFA-University of Basilicata (C. Serio and A. Carissimo)*
- *IMAA-CNR (G. Masiello, M. Viggiano, V. Cuomo)*
- *IAC-CNR (I. DeFeis)*
- *DET-University of Florence (A. Luchetta)*

Thanks to EUMETSAT (P. Schlüssel)

Outline

- Neural Network Methodology
- Neural Net vs. EOF Regression (IASI Retrieval exercise)
- Physical inversion methodology
- Physical inversion vs. EOF Regression (a few retrieval examples from IMG)

The Physical forward/inverse scheme for IASI

- φ-IASI is a software package intended for

- Generation of IASI synthetic spectra
- Inversion for geophysical parameters:

    1. temperature profile,
    2. water vapour profile,
    3. low vertical resolution profiles of O_3, CO, CH_4, N_2O.

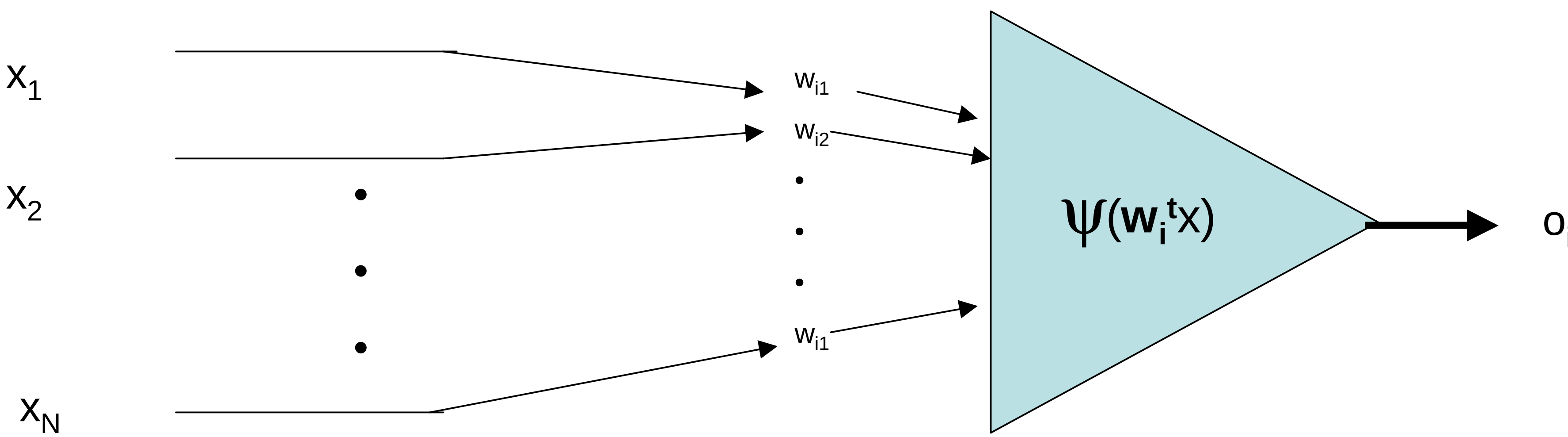
The φ-IASI family

- **σ-IASI**: forward model (available in Fortran with user's guide)
- **δ-IASI**: physical inverse scheme (available in Fortran with user's guide)
- **ν^2-IASI:** neural network inversion scheme (available in C++ with user's guide)
- **ε-IASI**: EOF based regression scheme (available in MATLAB, user's guide in progress).

Theoretical basis of Neural Network

K. Hornik, M. Stinchcombe, and H. White,
Multilayer feedforward networks are universal approximators
Neural Networks, **2**, 359-366, 1989

Basic Mathematical Structure of a generic i-th Neuron

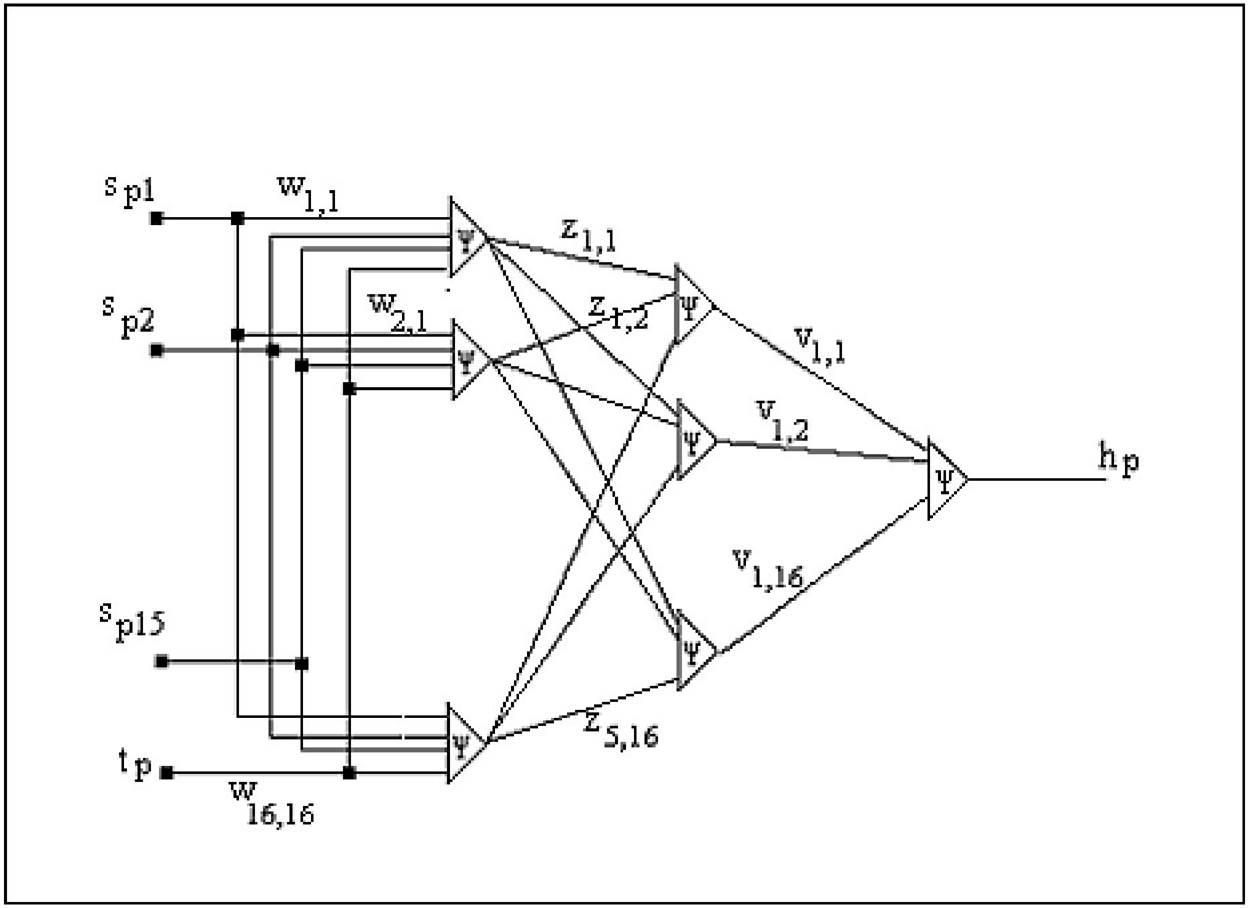


$$o_i = \psi(y) = \frac{2}{1+\exp(-\lambda y)} - 1; \quad \text{with } y = \mathbf{w}_i^t \mathbf{x} = w_{i1}x_1 + \cdots + w_{iN}x_N$$

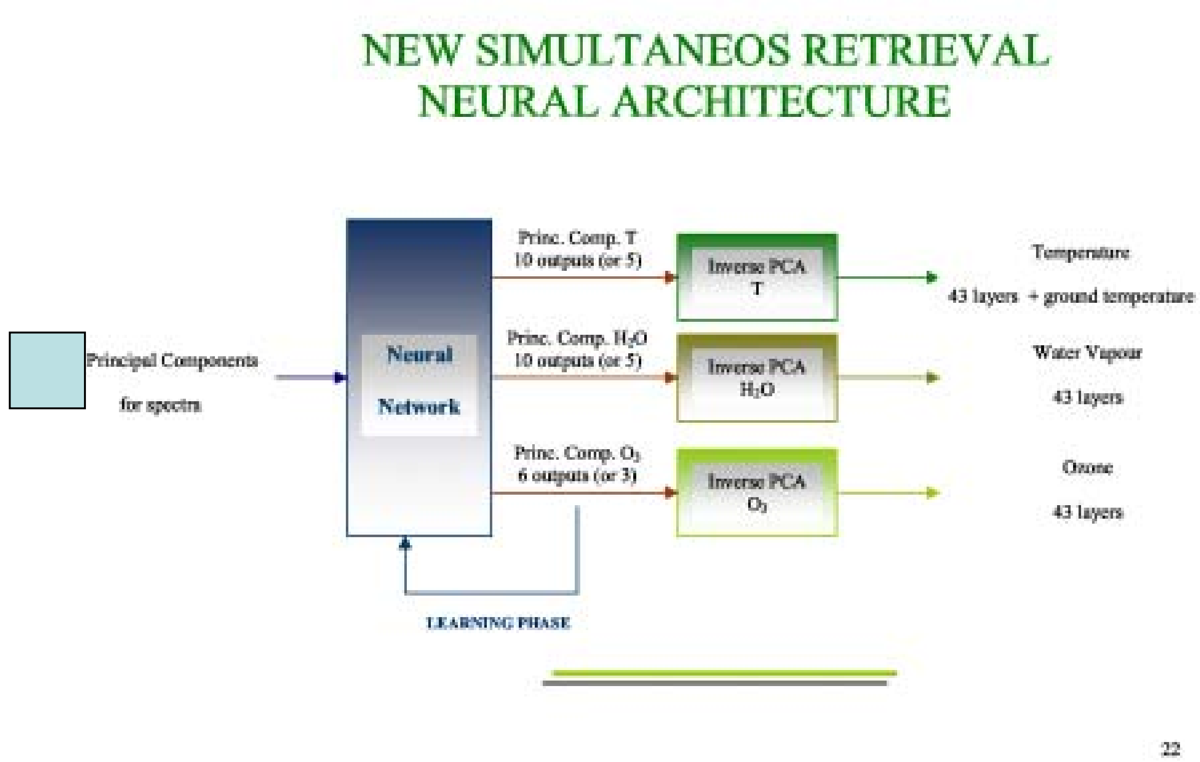
Cost Function to determine the weights

$$E = \frac{1}{2}(d_i - o_i)^2 = \frac{1}{2}(d_i - \psi(y))^2$$

Multilayer feed-forward Architecture



Simultaneous Architecture for (T, H_2O, O_3) retrieval



Inter-comparison exercise: NN vs. EOF regression

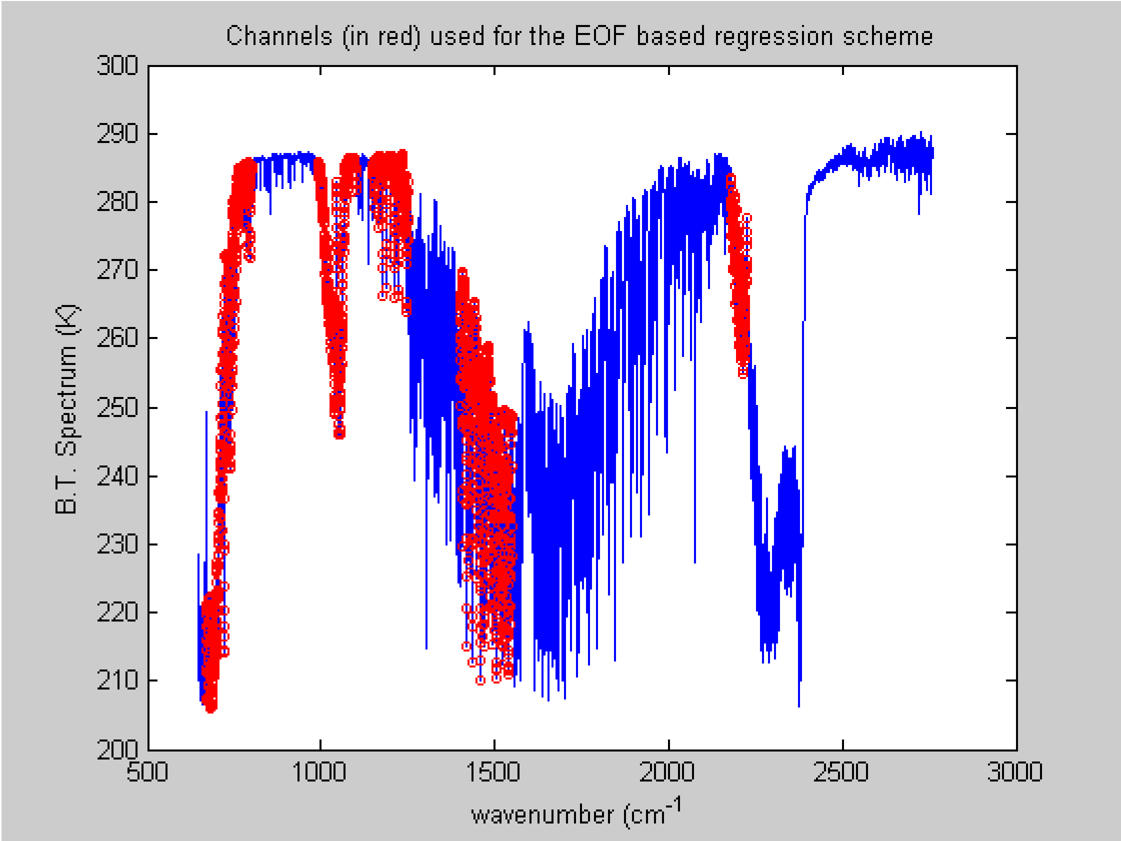
- **NEURAL NET**
- Input projected into an EOF basis: 50 PCs retained (Optimized)
- Output projected into EOF basis: 15 PC for T, 10 for H_2O, 15 for O_3

- **EOF regression**
- Input projected into an EOF basis: 200 PCs retained (Optimized)
- Output projected into EOF basis: 15 PC for T, 10 for H_2O, 15 for O_3

Rule of the comparison

- Compare the two schemes on a common basis:
  - same a-priori information (training data-set),
  - same inversion strategy: simultaneous
  - same quality of the observations

Definition of the spectral ranges



Training/Validation and Test Data Sets

Air Mass Type	Training/ Validation data set	Test data set
Tropical	595	221
Mid-Latitude Summer	305	43
Mid-Latitude Winter	388	155
High Latitude Summer	283	80
High Latitude Winter	740	164

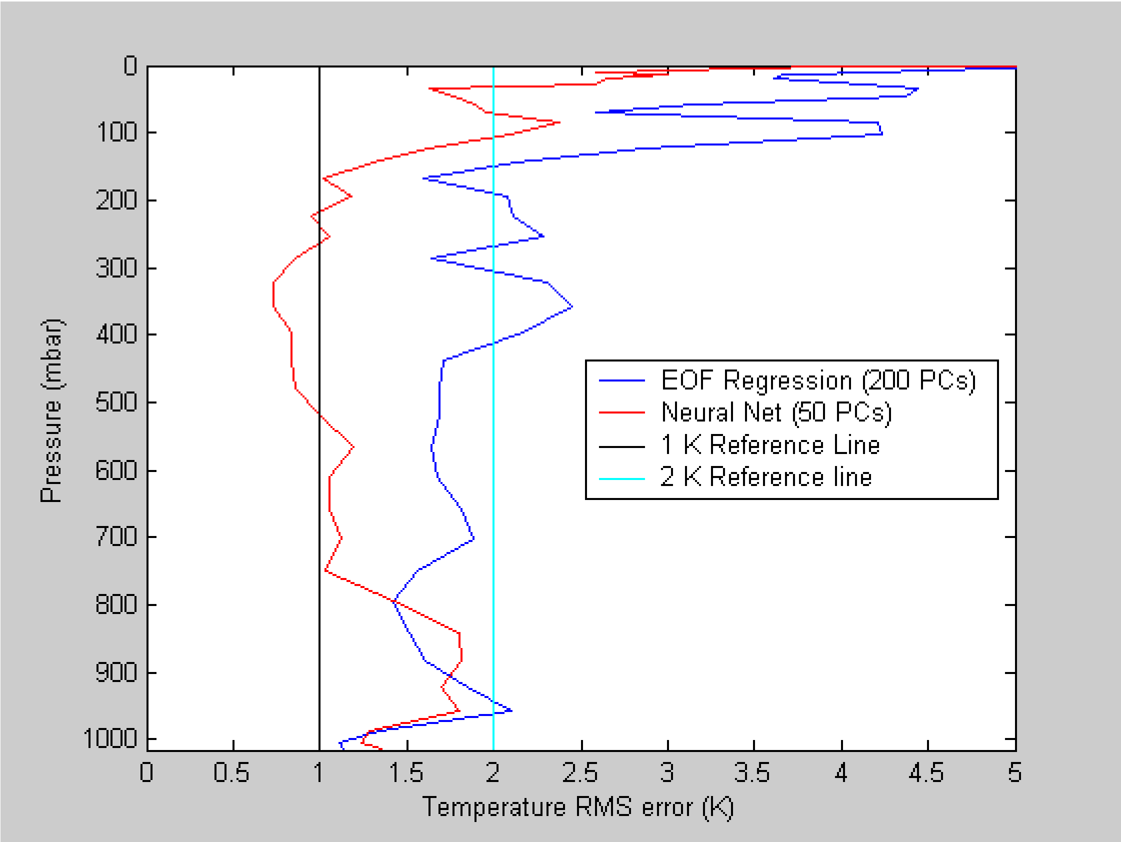
2311 TIGR profiles

(TIGR-3 data base)

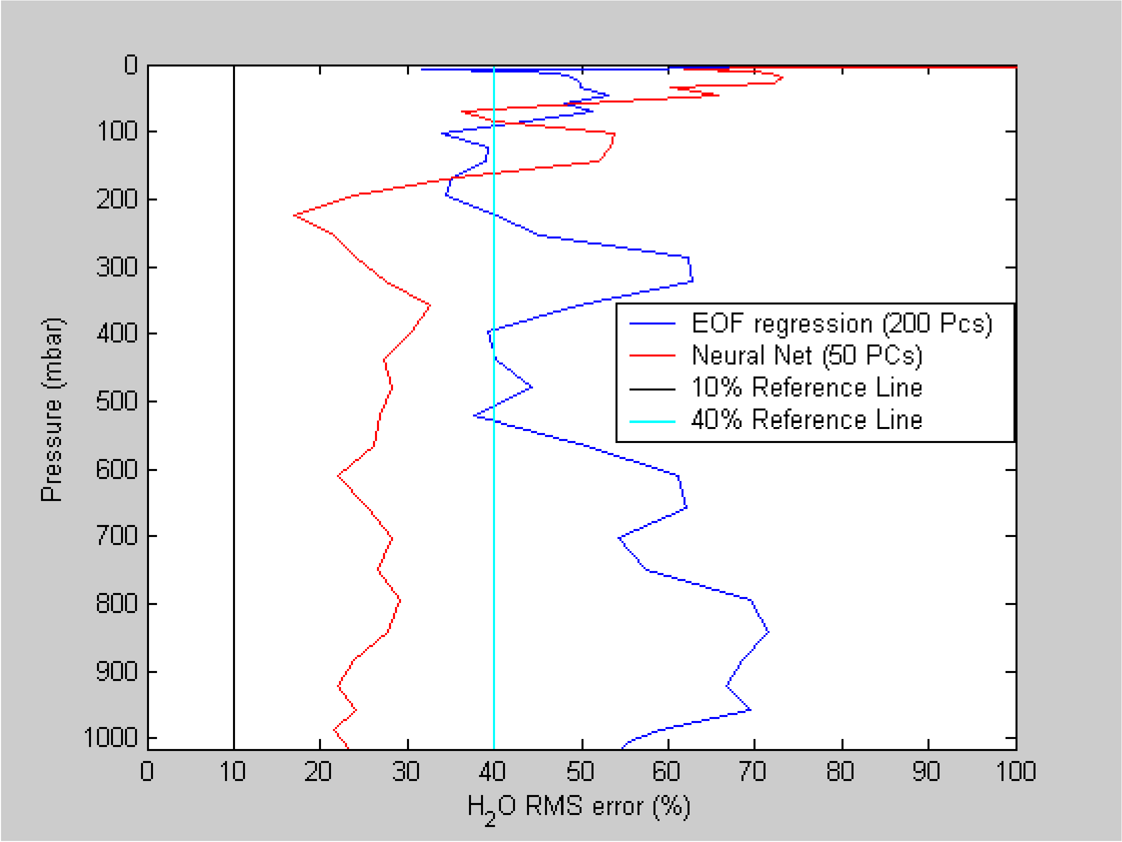
RIE Test

Data Set, EUMETSAT

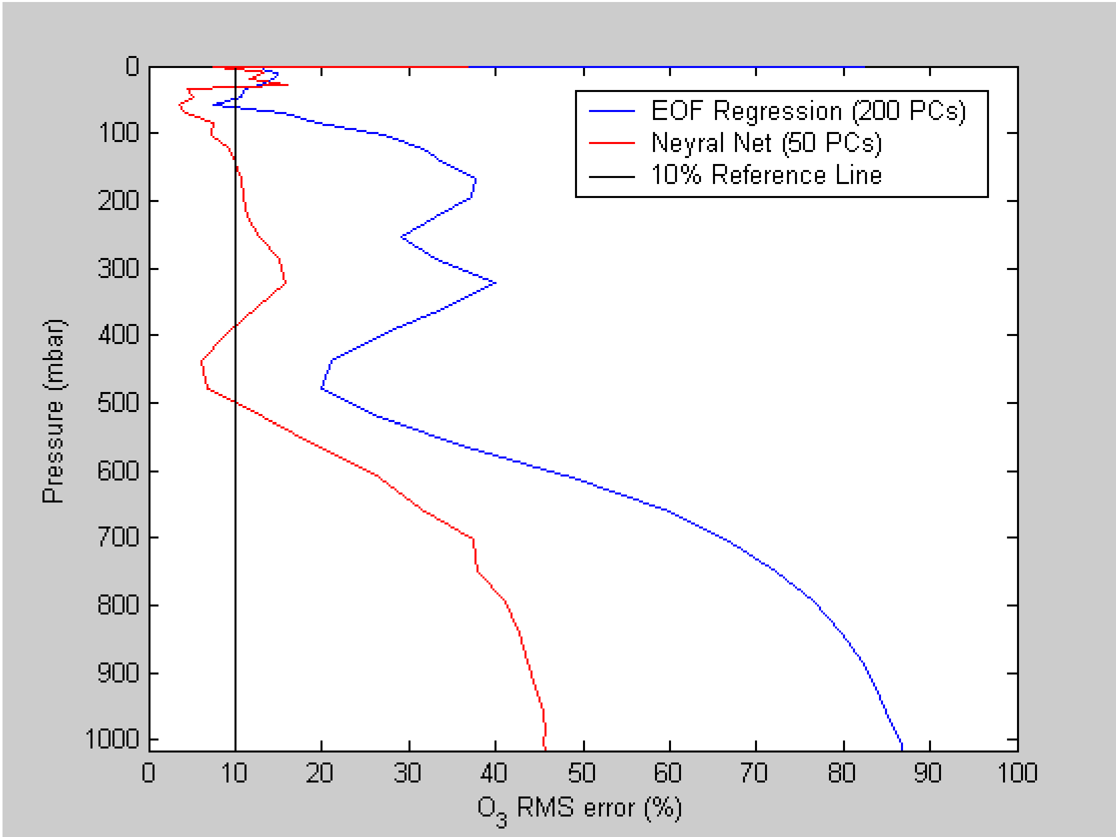
Tropical Air Mass
(RIE test data set)



Tropical Air Mass



Tropical Air Mass



Summary

- NN performs better than EOF Regression provided that they are evaluated on a common basis.
- NN is parsimonious with respect to EOF regression (50 PCs vs. 200 PCs)

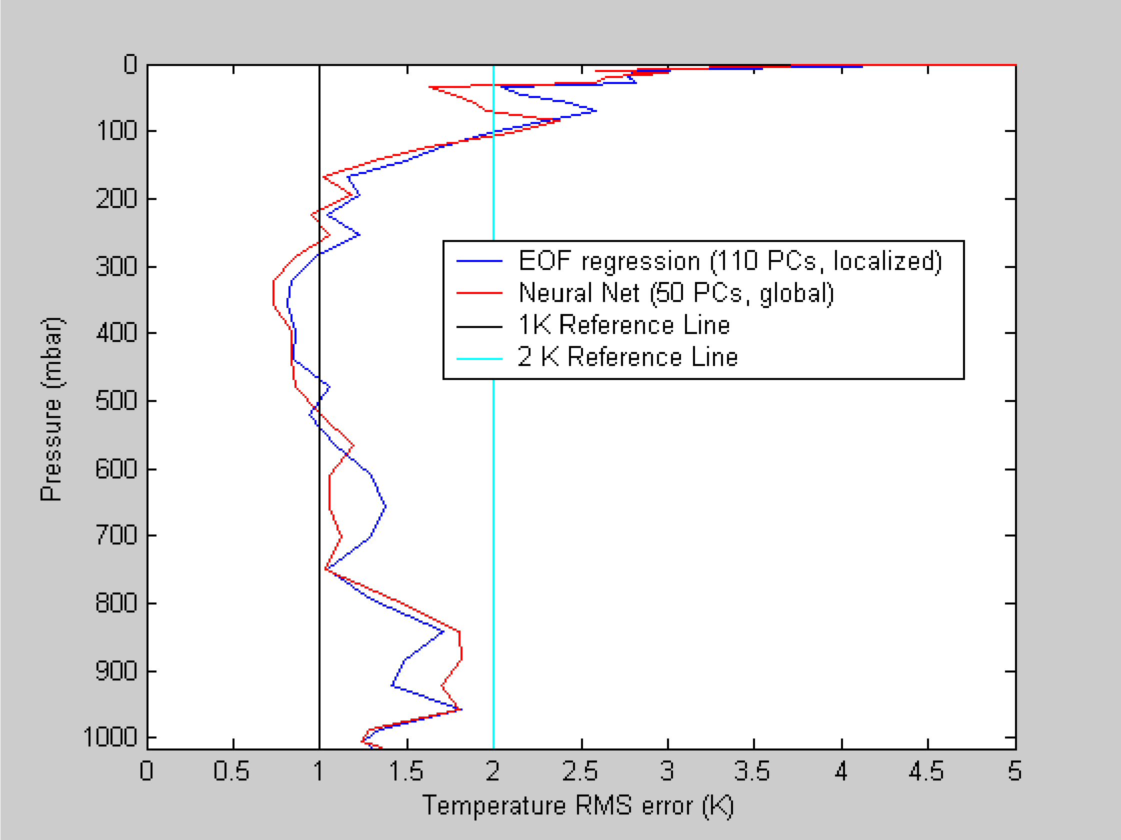
Optimization

Eof Regression

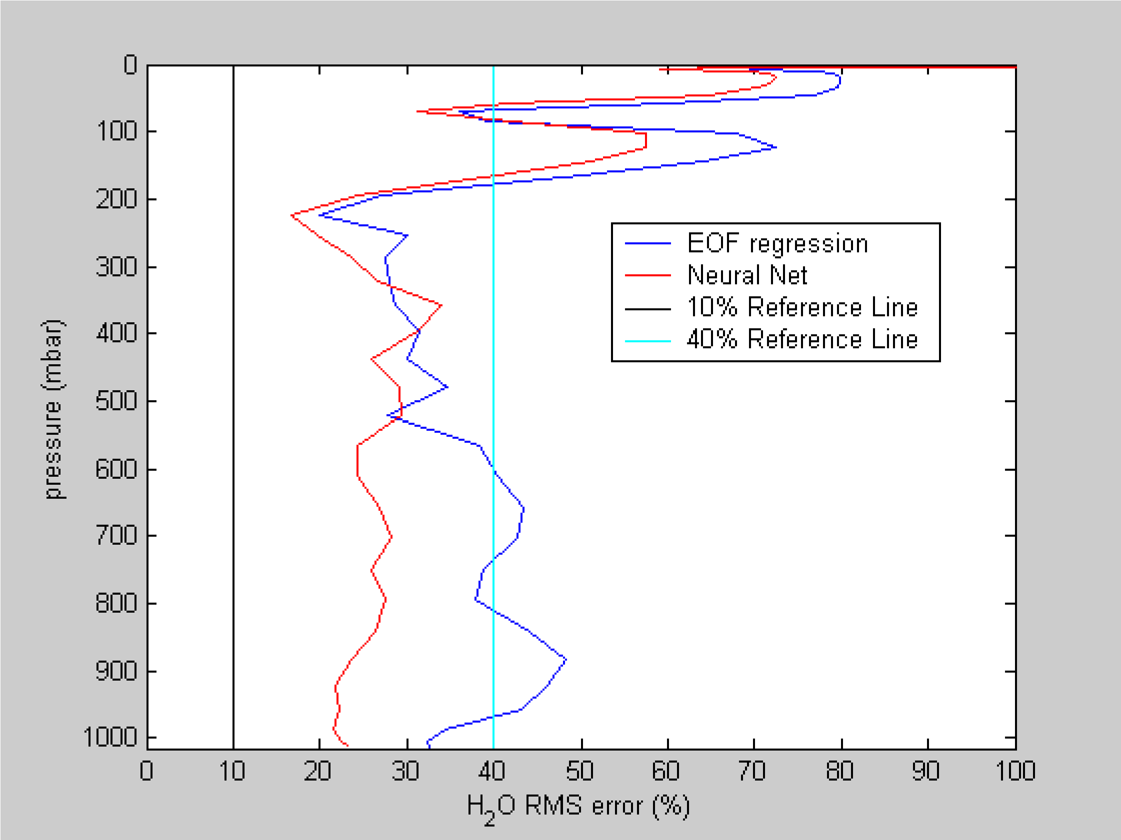
Localize training

(Tropics, Mid-Latitude, and so on)

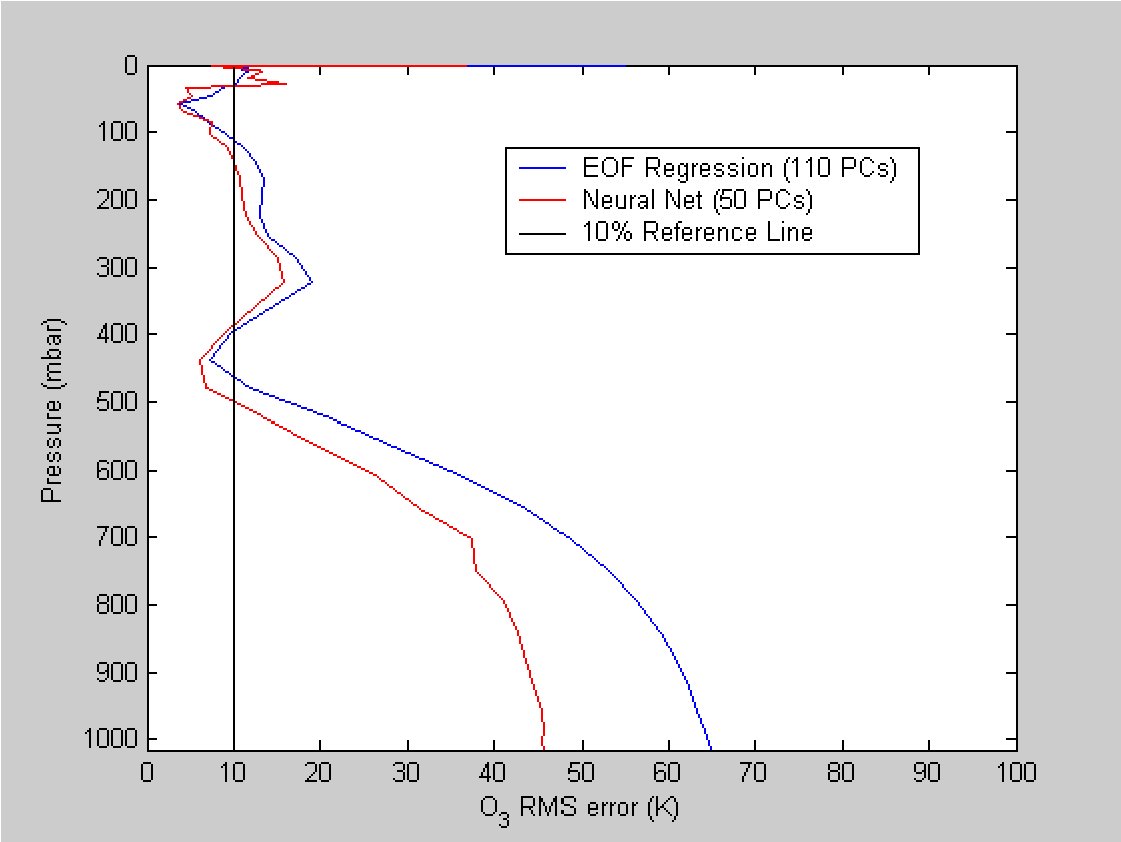
Tropical Air Mass



Tropical Air Mass



Tropical Air Mass



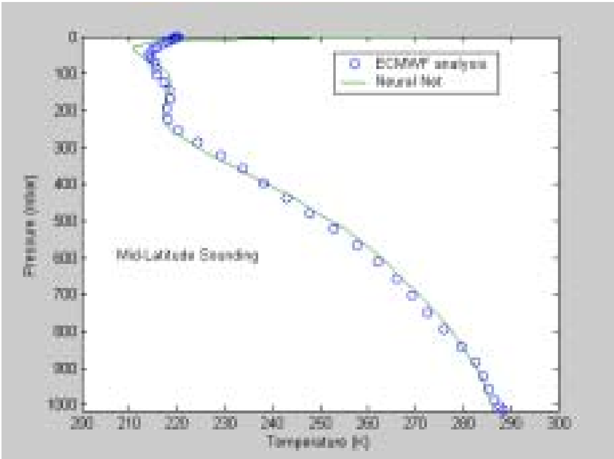
Summary

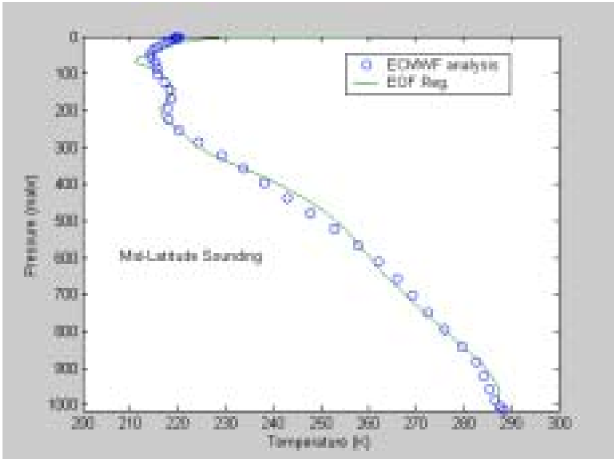
- EOF regression improves when properly localized
- Neural Net is expected to improve, as well. Results are not yet ready.

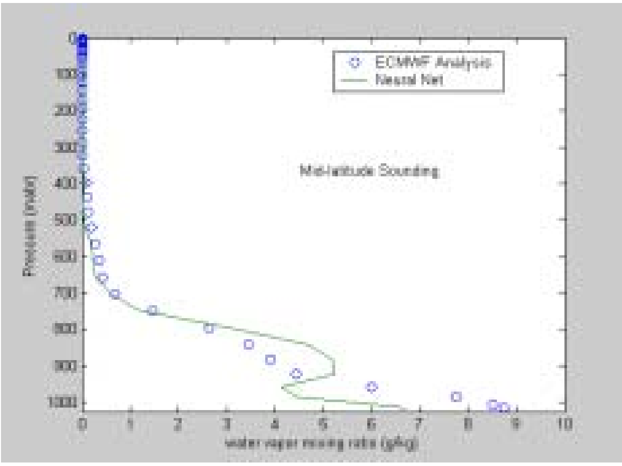
Dependency on the Training data set

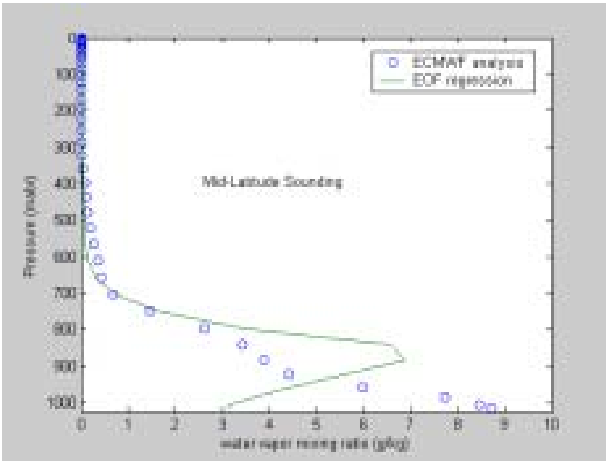
- One major concerns with both the schemes is their critical dependence on the training data set

Comparing N.N. to EOF Reg.
IMG Mid-Latitude Observation

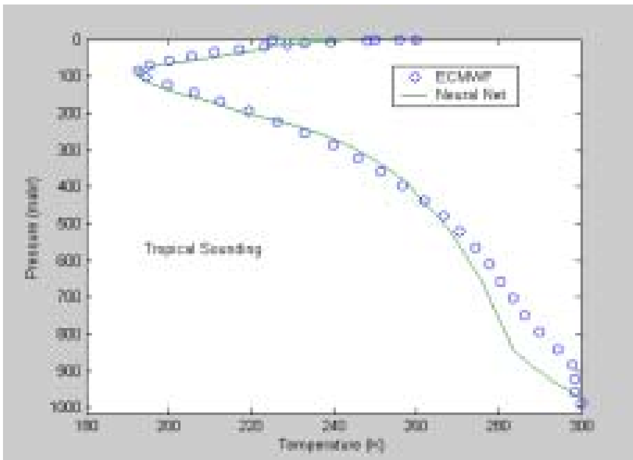


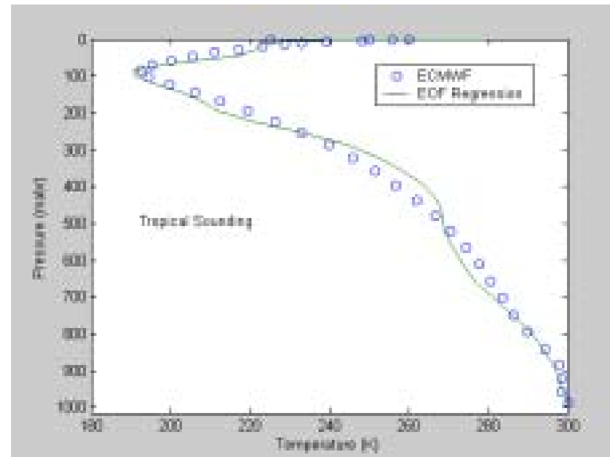


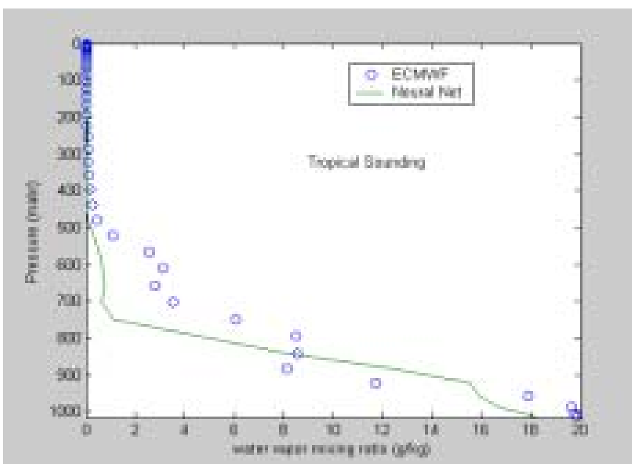


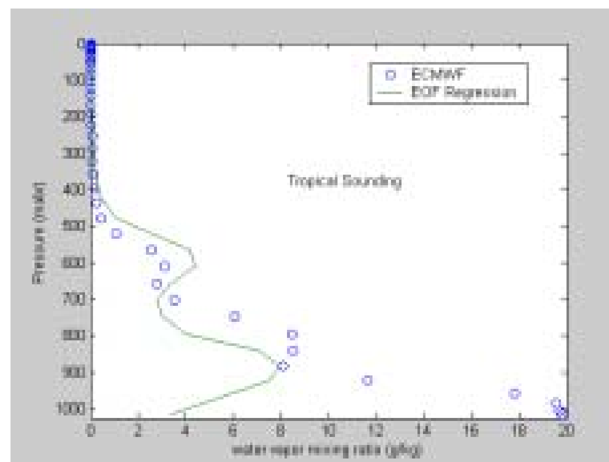


Comparing N.N. to EOF Reg.
IMG Tropical Observation







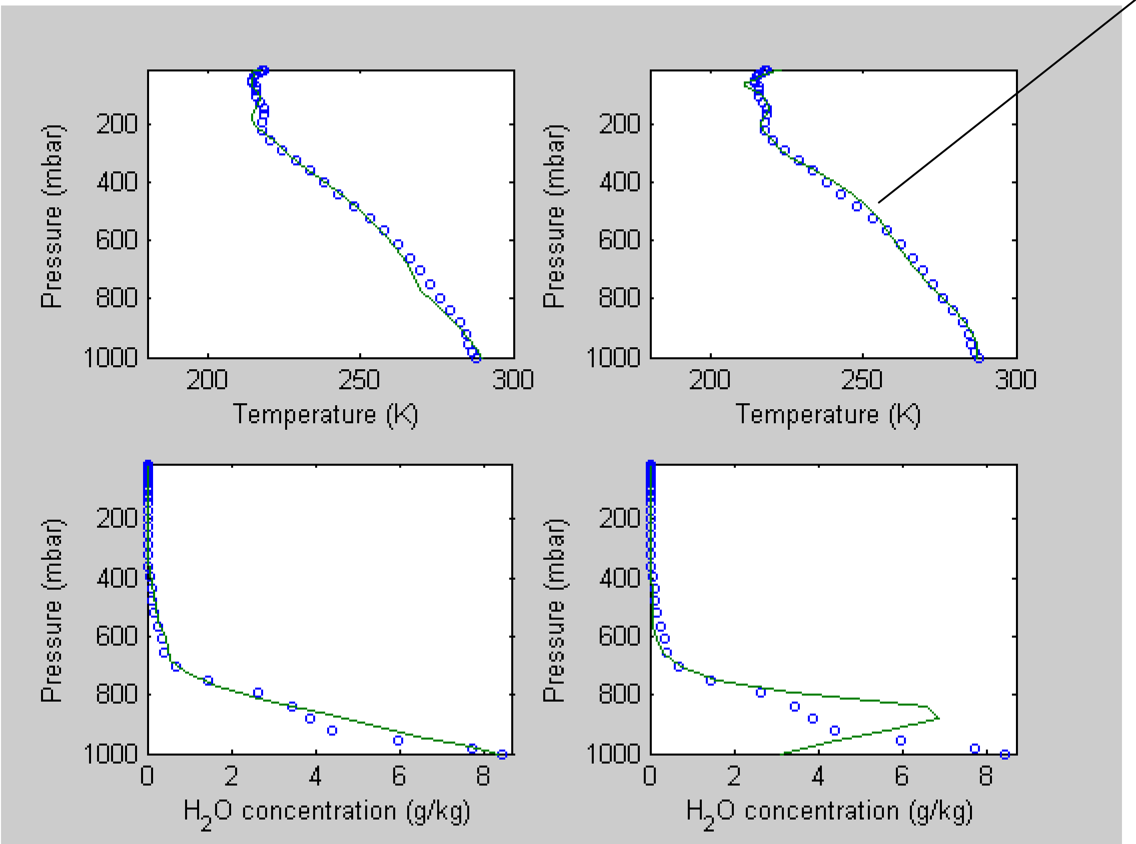


How to get rid of data-set dependency?

Physical Retrieval

(Initialized by Climatology)

EOF Regression



Circles: ecmwf;

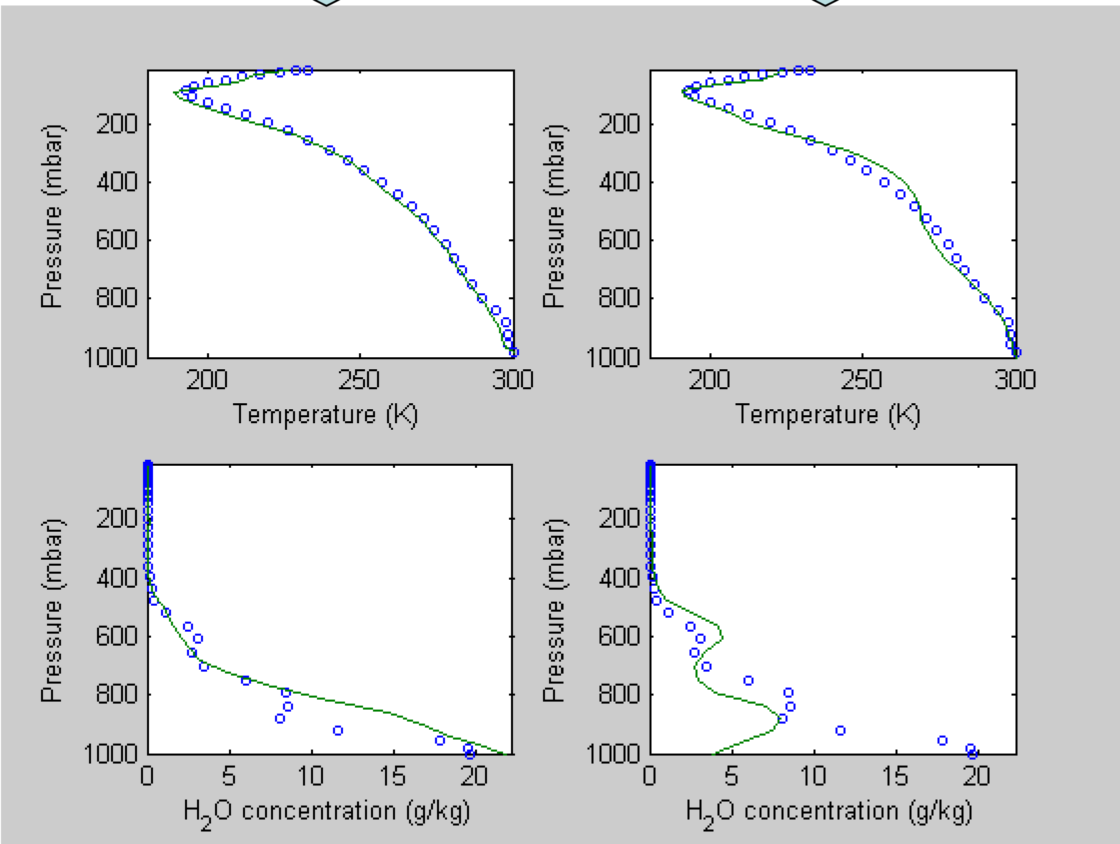
Line: retrieval

Introducing Physical Inversion

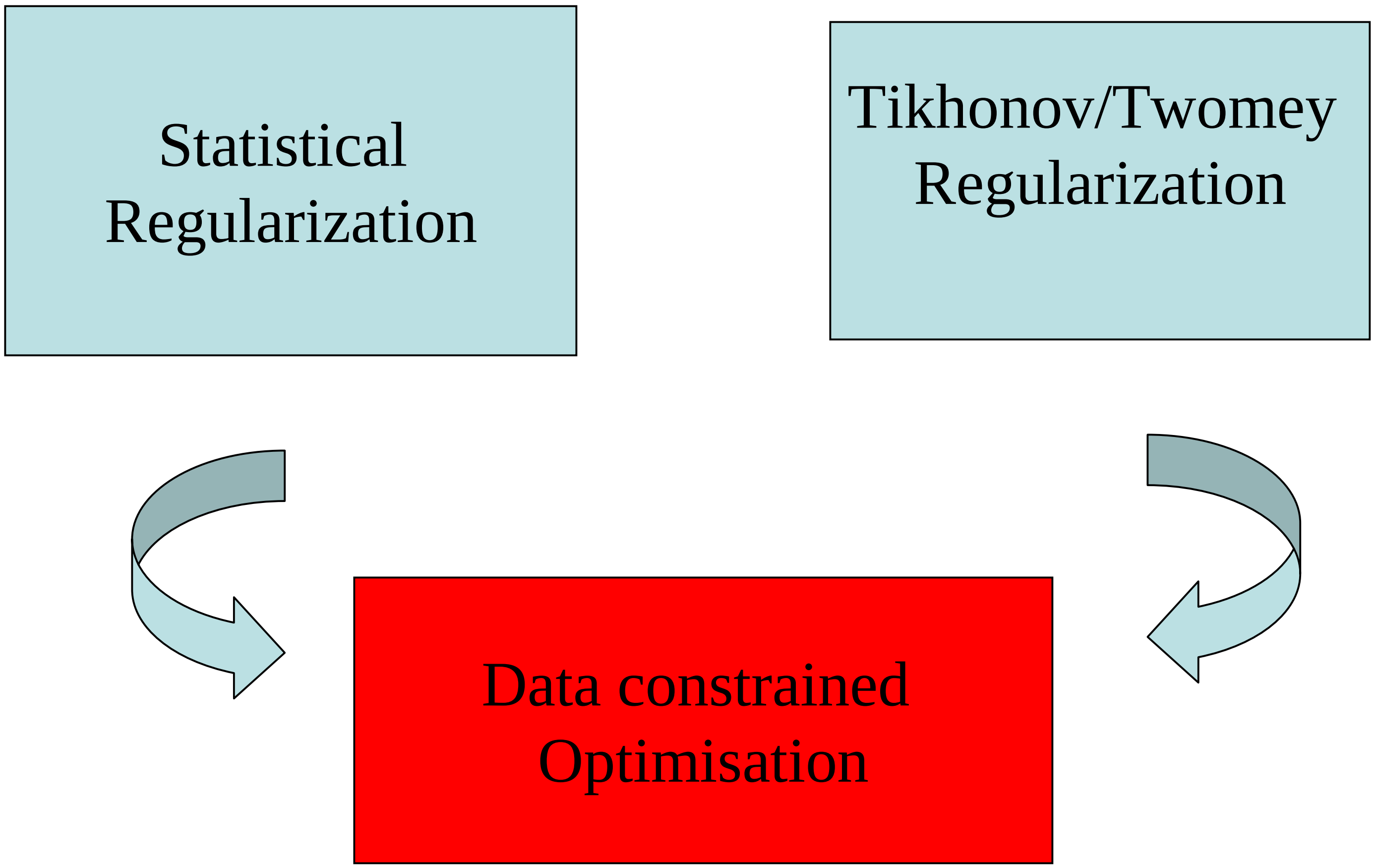
Physical Retrieval

(Initialized by Climatology)

EOF Regression



Introducing Physical Inversion


Statistical Regularization
Tikhonov/Twomey Regularization
Data constrained Optimisation

The subscript *g* stands for a suitable background atmospheric state!

$$(\hat{\mathbf{v}} - \mathbf{v}_g)^t \mathbf{L} (\hat{\mathbf{v}} - \mathbf{v}_g) \qquad \text{MIN!}$$

$$(\mathbf{y} - \mathbf{Kx})^{\mathrm{t}} \mathbf{S}^{-1} (\mathbf{y} - \mathbf{Kx}) \leq \chi_\alpha^2$$

S : Obs. Cov. Matrix

L : Smoothing Operator

About **L**

- Twomey's approach

$$\mathbf{L} \equiv \int_0^{+} \left| \frac{d^n \hat{x}}{dh^n} \right|^2 dh;$$

- n=0, 1, 2

- Rodgers' approach
- **L** is intrinsically a covariance operator, **B** in his notation

Physical Consistency of the L-norm

$$(\widehat{\mathbf{v}} - \mathbf{v}_{\mathbf{g}})^{t}\,\mathbf{L}(\widehat{\mathbf{v}} - \mathbf{v}_{\mathbf{g}})$$

Twomey's L is lacking dimensional consistency! it attempts to Sum unlike quantity, e,g, (K+g/kg)

Rodgers' **L** ensures dimensional consistency

$\mathbf{L}=\mathbf{B}^{-1}$

which makes the norm above dimensionless

Our Choice

- Since we are interested in simultaneous inversion which involves unlike quantities such as Temperature, water vapour concentration and so on we choose
  - $\mathbf{L=B^{-1}}$
- However, the methodology we are going to discuss still hold for any Twomey’s **L**

Finding the solution through Lagrange multiplier method

$$\hat{\mathbf{x}} = \hat{\mathbf{v}} - \mathbf{v}_g = (\gamma\mathbf{B}^{-1} + \mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})^{-1}\mathbf{K}^t\mathbf{S}^{-1}\mathbf{y}$$

$\gamma = 1$ gives the usual Statistical Regularization;

$$\mathbf{S}_v = (\gamma\mathbf{B}^{-1} + \mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})^{-1}(\gamma^2\mathbf{B}^{-1} + \mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})(\gamma\mathbf{B}^{-1} + \mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})^{-1};$$

$$\begin{cases} \mathbf{S}_v = (\mathbf{B}^{-1} + \mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})^{-1}; & \text{for } \gamma = 1 \\ \mathbf{S}_v = (\mathbf{K}^t\mathbf{S}^{-1}\mathbf{K})^{-1}; & \text{for } \gamma \to 0 \\ \mathbf{S}_v = \mathbf{B}; & \text{for } \gamma \to \infty \end{cases}$$

Uncovering the elemental constituent of regularization

$$\mathbf{B} = \mathbf{B}^{\frac{1}{2}}\mathbf{B}^{\frac{t}{2}}$$

$$(\gamma\mathbf{B}^{-\frac{t}{2}}\mathbf{B}^{-\frac{1}{2}} + \mathbf{J}^t\mathbf{J})\hat{\mathbf{x}} = \mathbf{J}^t\mathbf{z}; \text{ with } \mathbf{J} = \mathbf{S}^{-\frac{1}{2}}\mathbf{K}, \quad \mathbf{z} = \mathbf{S}^{-\frac{1}{2}}\mathbf{y};$$

$$\mathbf{B}^{-\frac{t}{2}}(\gamma\mathbf{I} + \mathbf{B}^{\frac{t}{2}}\mathbf{J}^t\mathbf{J}\mathbf{B}^{\frac{1}{2}})\mathbf{B}^{-\frac{1}{2}}\hat{\mathbf{x}} = \mathbf{J}^t\mathbf{z};$$

$$\mathbf{G} = \mathbf{J}\mathbf{B}^{\frac{1}{2}} \text{ and } \hat{\mathrm{u}} = \mathbf{B}^{-\frac{1}{2}}\hat{\mathbf{x}};$$

$$(\gamma\mathbf{I} + \mathbf{G}^t\mathbf{G})\hat{\mathbf{u}} = \mathbf{G}^t\mathbf{z}$$

Ridge Regression

Continued

- The same decomposition may obtained for Twomey regularization by putting
  - $\mathbf{L}=\mathbf{M}^t\mathbf{M}$
- **M** may be obtained by Cholesky decomposition for any symmetric full rank matrix **L**
- Twomey's **L** is typically singular. Nevertheless the above decomposition may be obtained by resorting to GSVD (Hansen, *SIAM Review*, Vol. 34, pp. 561, (1992))

Summary

- The RIDGE regression is the paradigm of any regularization method,
- The difference between the various methods is:
  - the way they normalize the Jacobian
  - the value they assign to the Lagrange multiplier

- **Levenberg-Marquardt**:

➢ γ is assigned alternatively a small or a large value, the Jacobian is not normalized, that is **L=I .**

- **Thikonov**

➢ γ is a free-parameter (chosen by trial and error), the Jacobian is normalized through a mathematical operator.

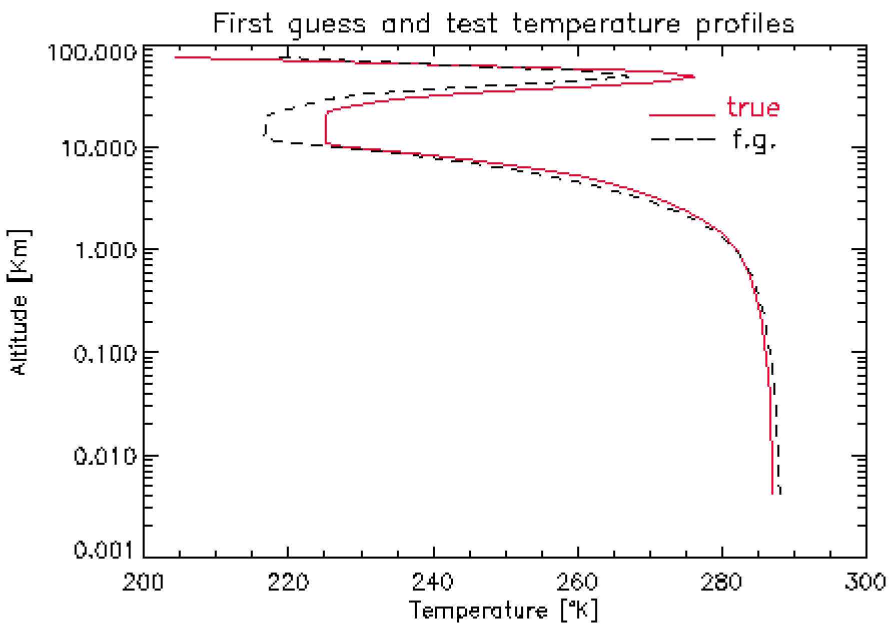
- **Rodgers**:

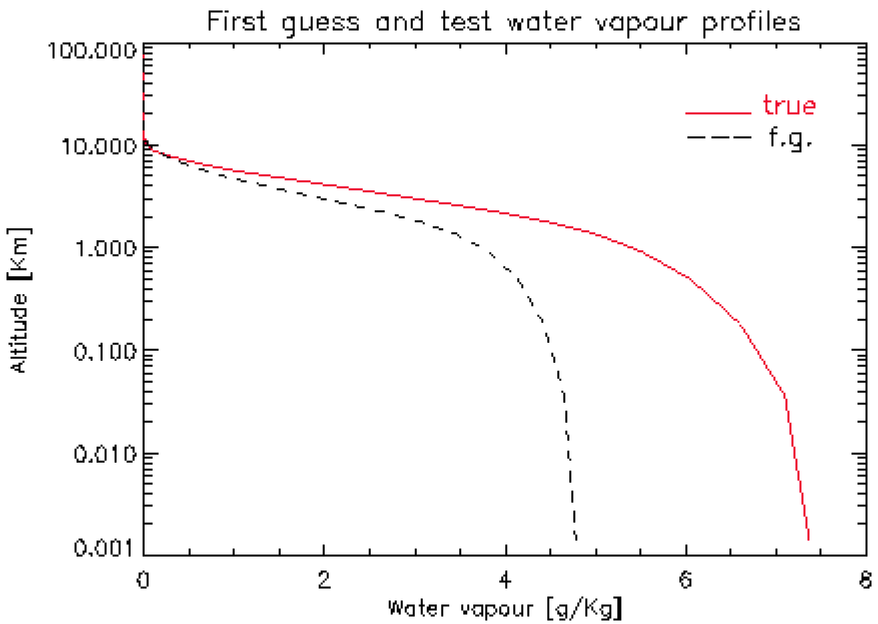
➢ γ =1, the Jacobian is normalized to the a-priori covariance matrix. It is the method which enables dimensional consistency.

- **Our Approach**

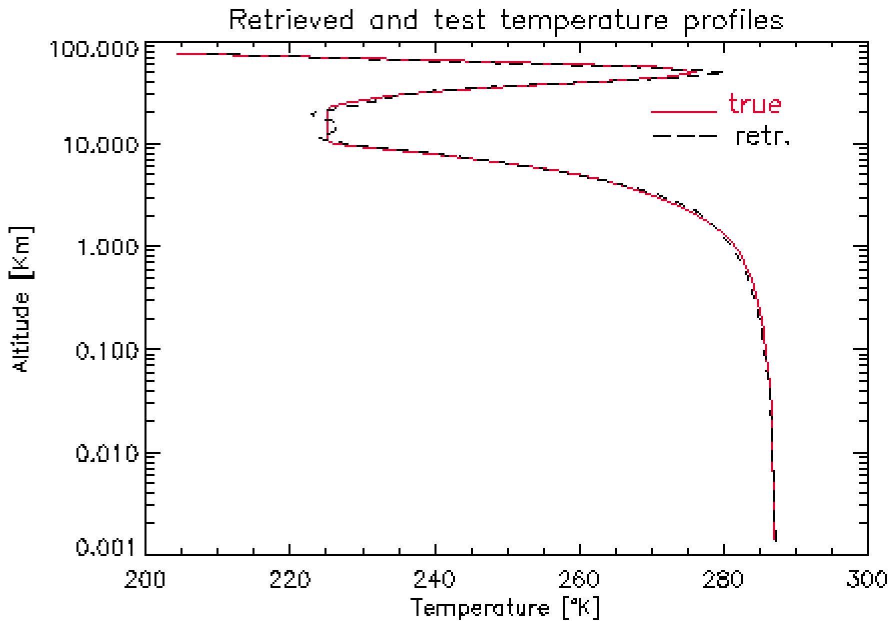
➢ Rodgers approach combined with an optimal choice of the γ parameter (L-curve criterion).

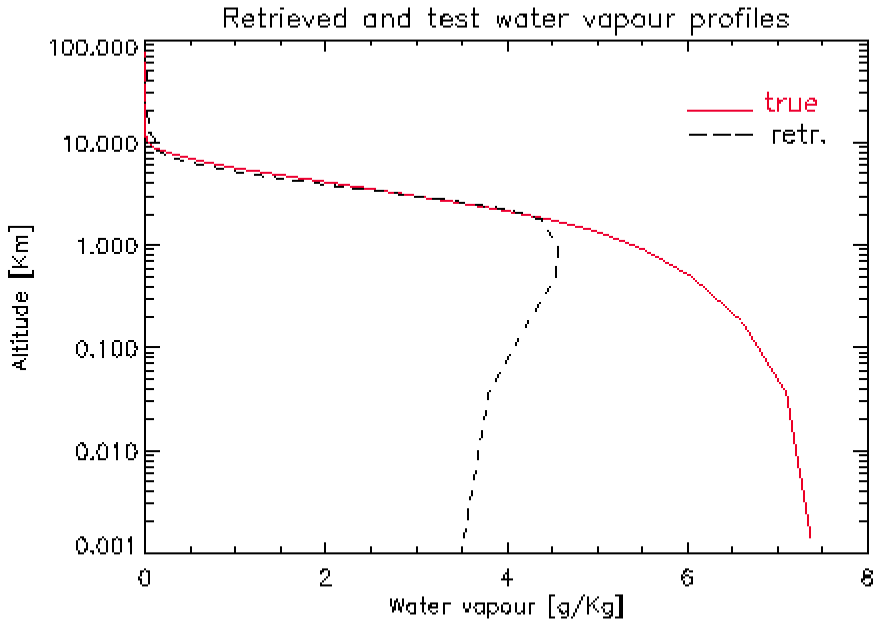
A simple numerical exercise



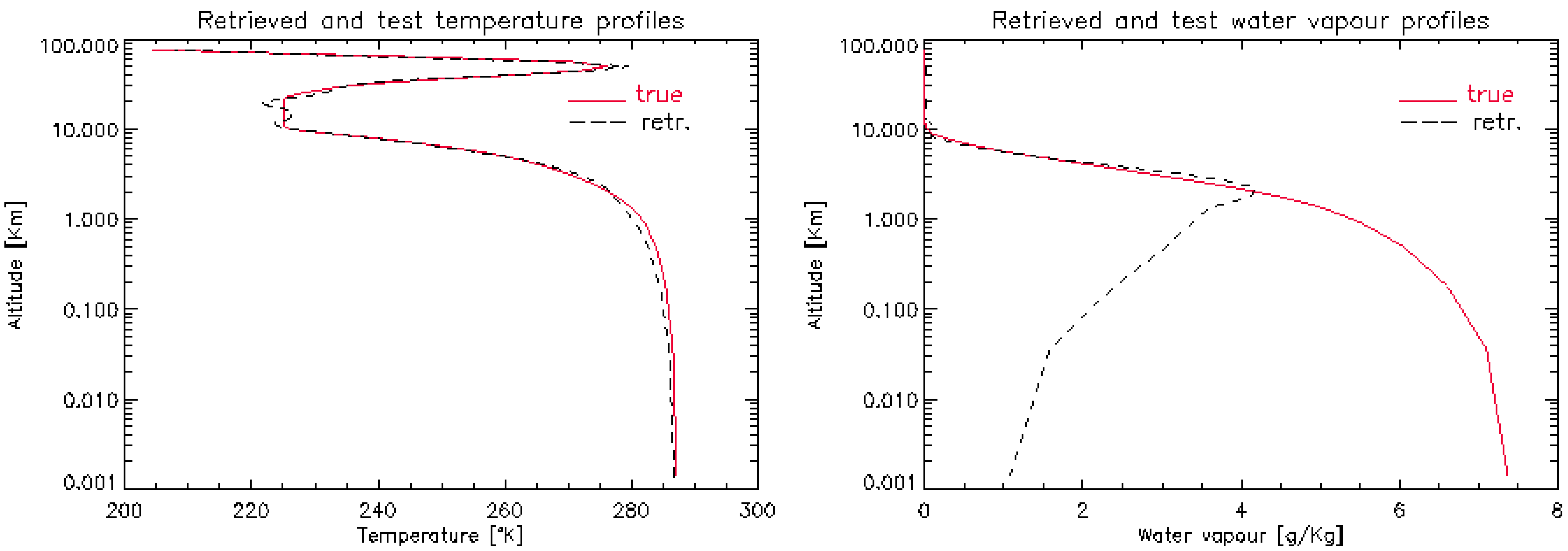


Statistical Regularization 1st Iteration

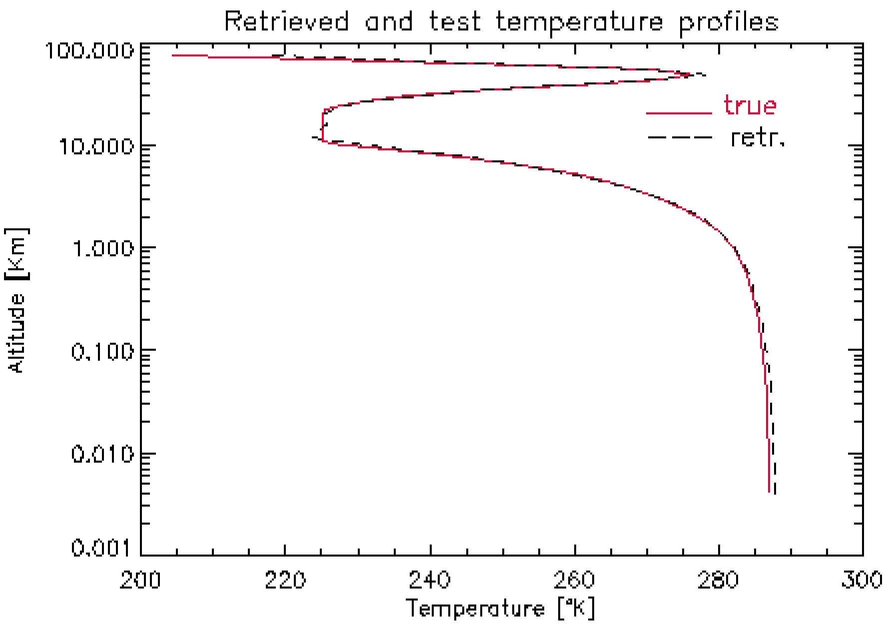


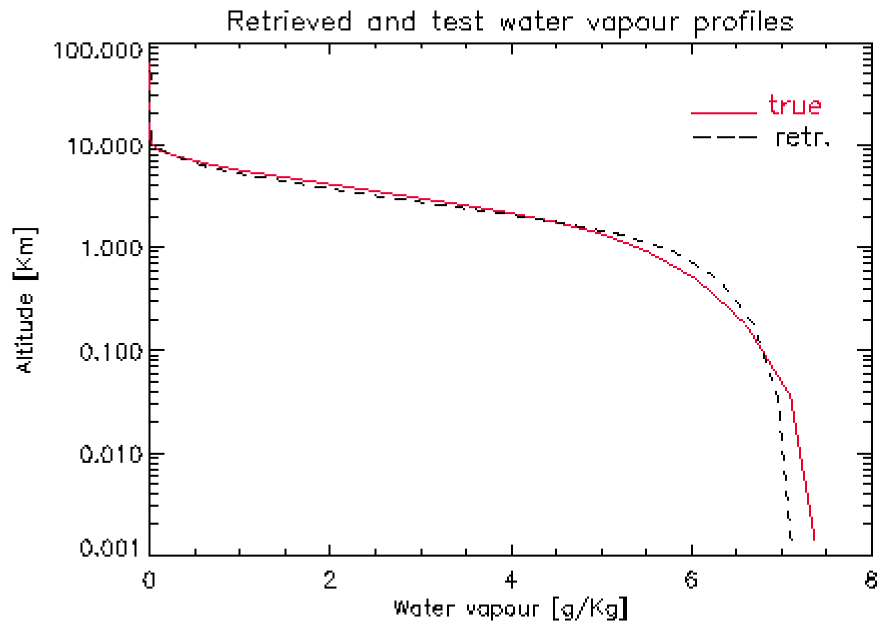


Statistical Regularization
2nd Iteration

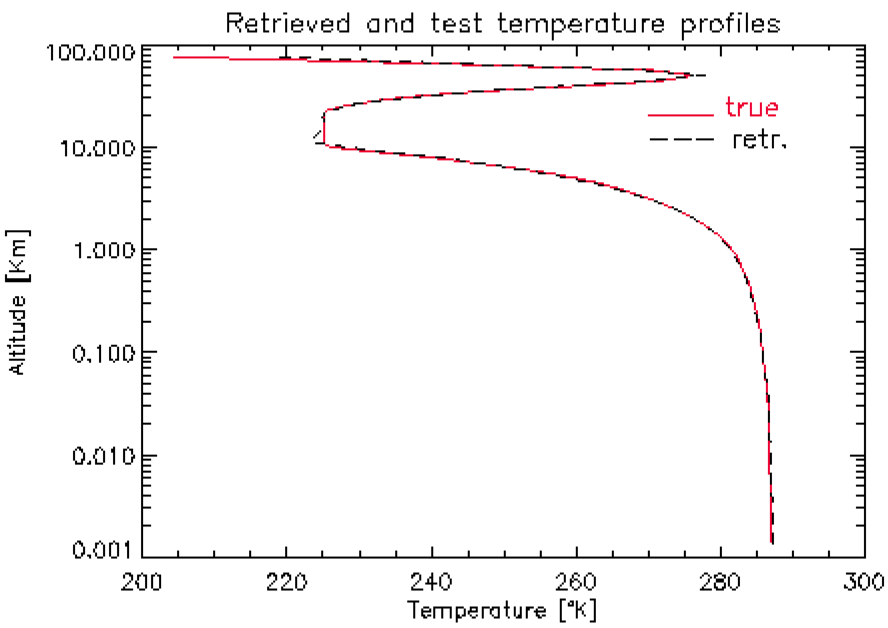


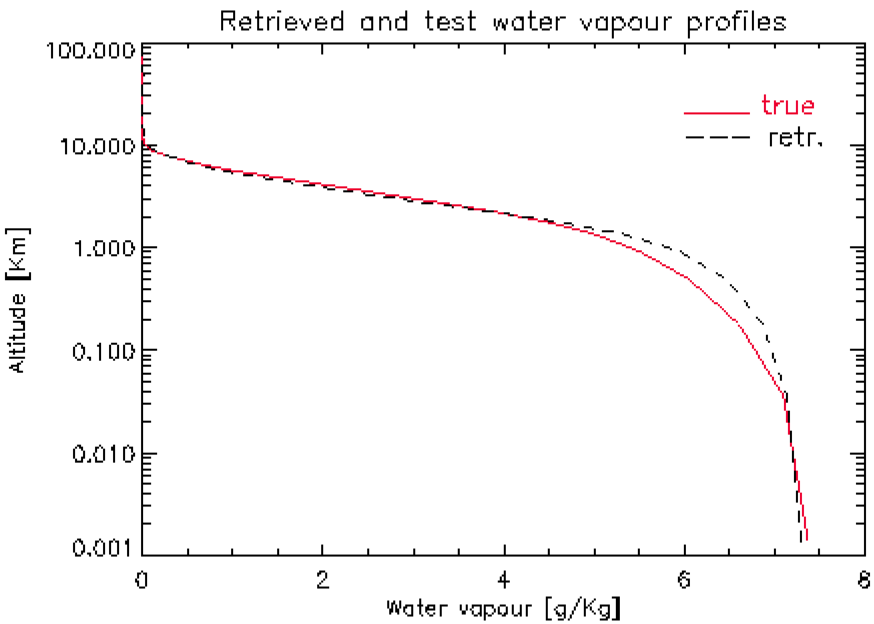
L-Curve, 1st Iteration





L-Curve, 2nd Iteration

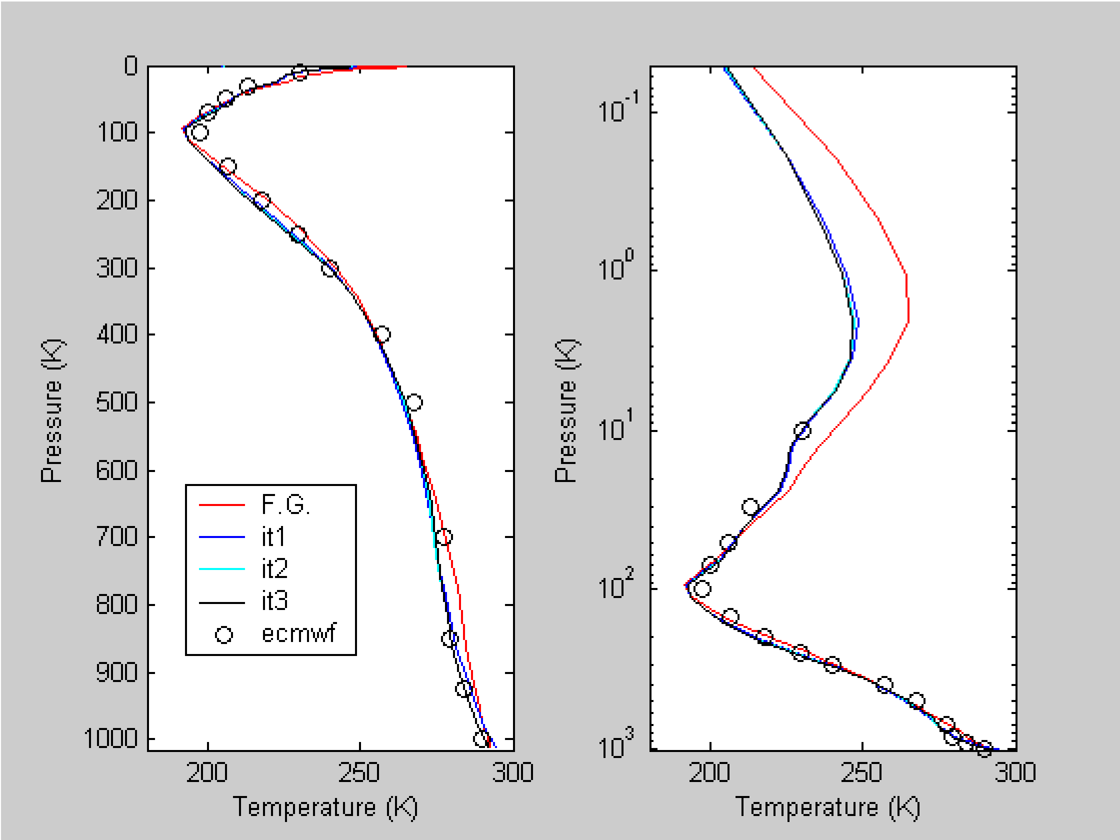




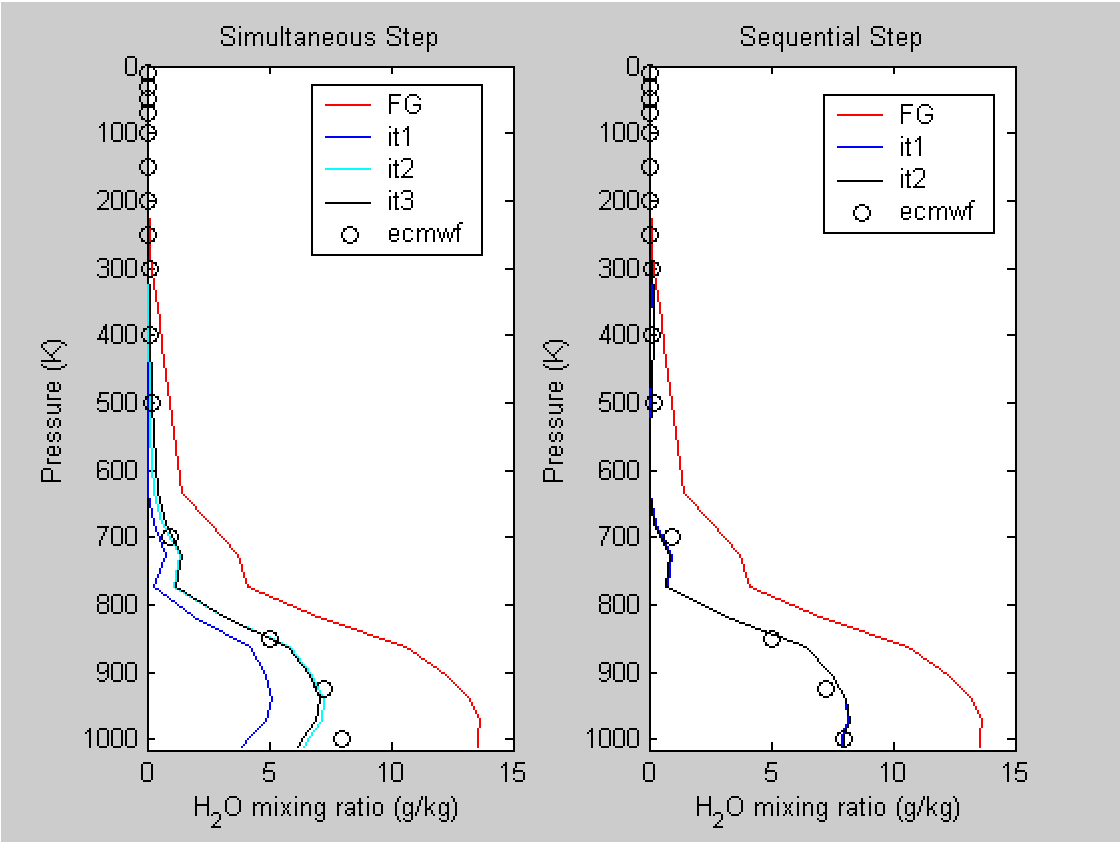
Convergence example based on an IMG real spectrum

- Inversion strategy:
  - 667 to 830 cm^{-1} simultaneous for (T,H_2O)
  - 1100 to 1600 cm^{-1} (super channels) sequential for H_2O alone
  - 1000 to 1080 cm^{-1} sequential for Ozone

Temperature

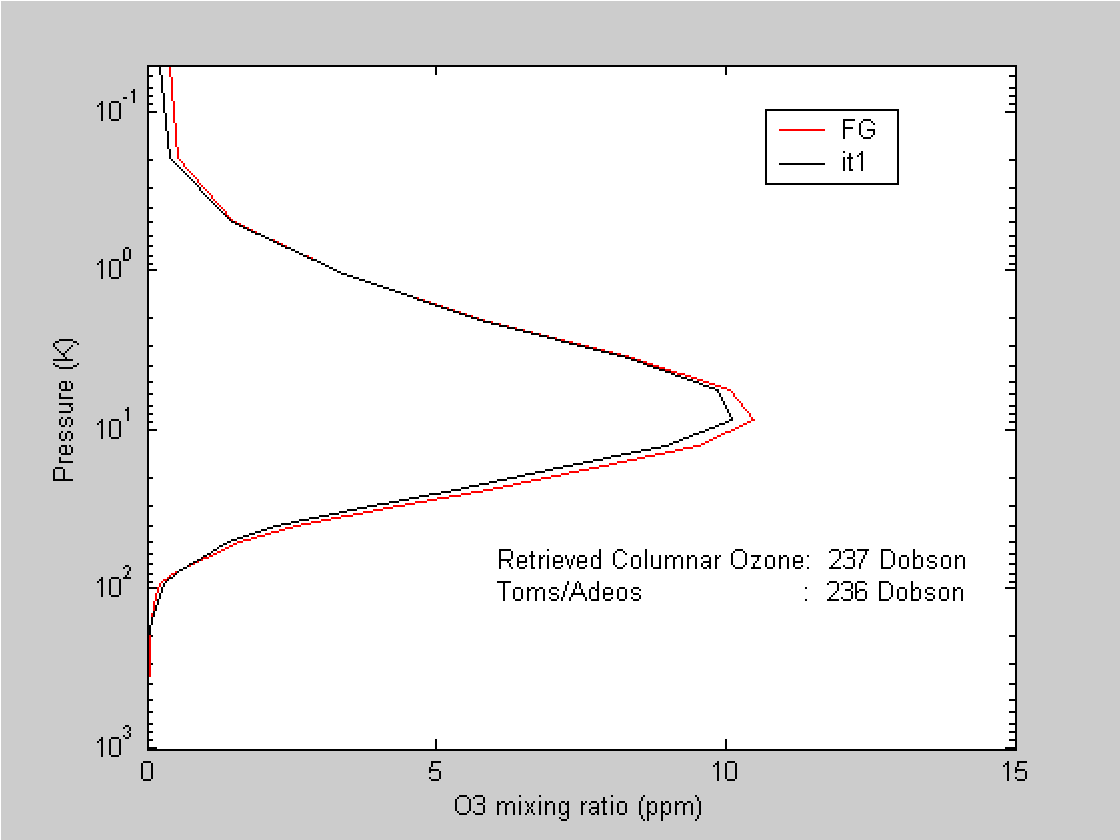


Water Vapor



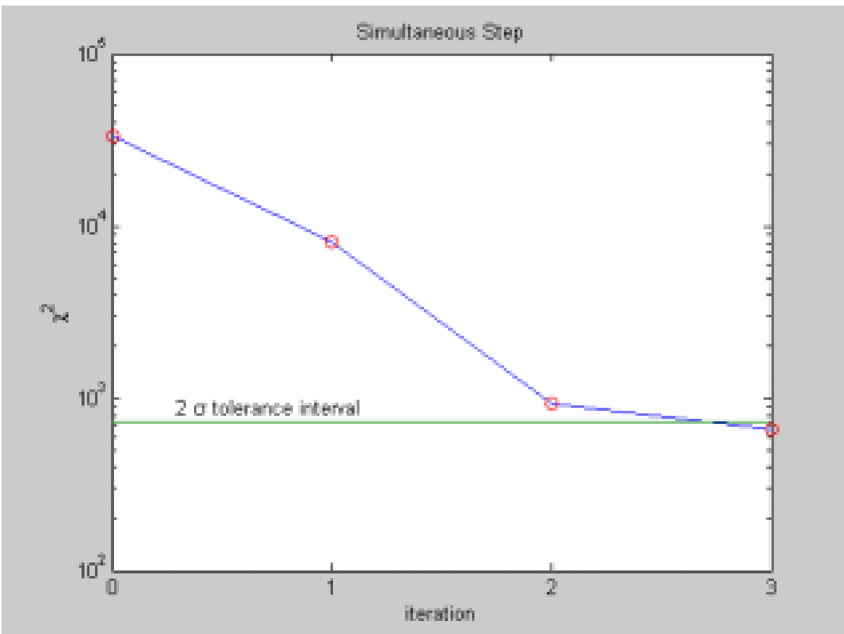
Simultaneous Step
Pressure (K)
0
100
200
300
400
500
600
700
800
900
1000
FG
it1
it2
it3
ecmwf
0
5
10
15
H2O mixing ratio (g/kg)
Sequential Step
Pressure (K)
FG
it1
it2
ecmwf
H2O mixing ratio (g/kg)

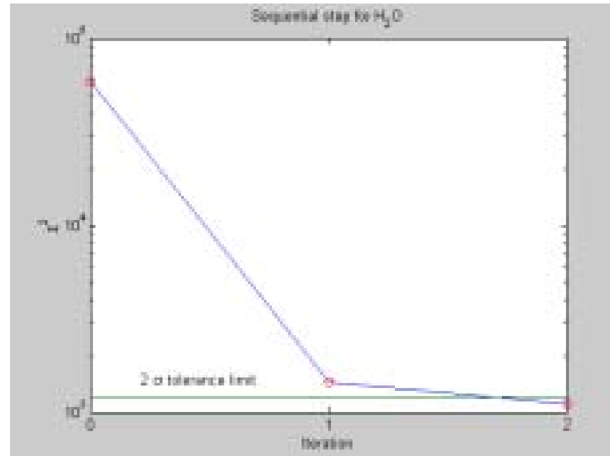
Ozone

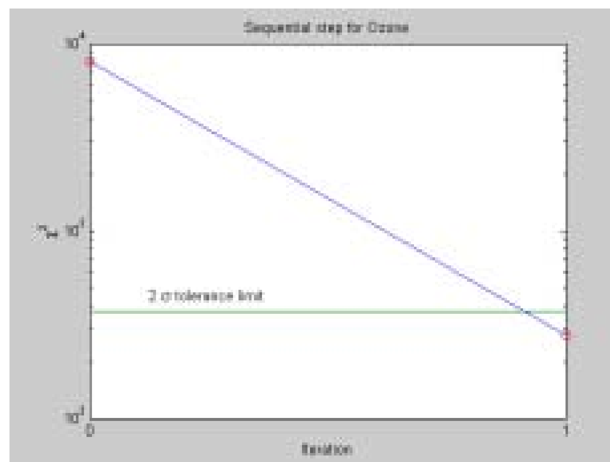


FG
it1
Retrieved Columnar Ozone: 237 Dobson
Toms/Adeos : 236 Dobson
Pressure (K)
O3 mixing ratio (ppm)
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
0
5
10
15

χ^2-constraint







Physical Inversion vs. EOF regression: Exercise based on Real Observations (IMG)

- **EOF regression:**
- Training data set: a set of profiles from ECMWF analyses

Physical Inversion

Based on Climatology

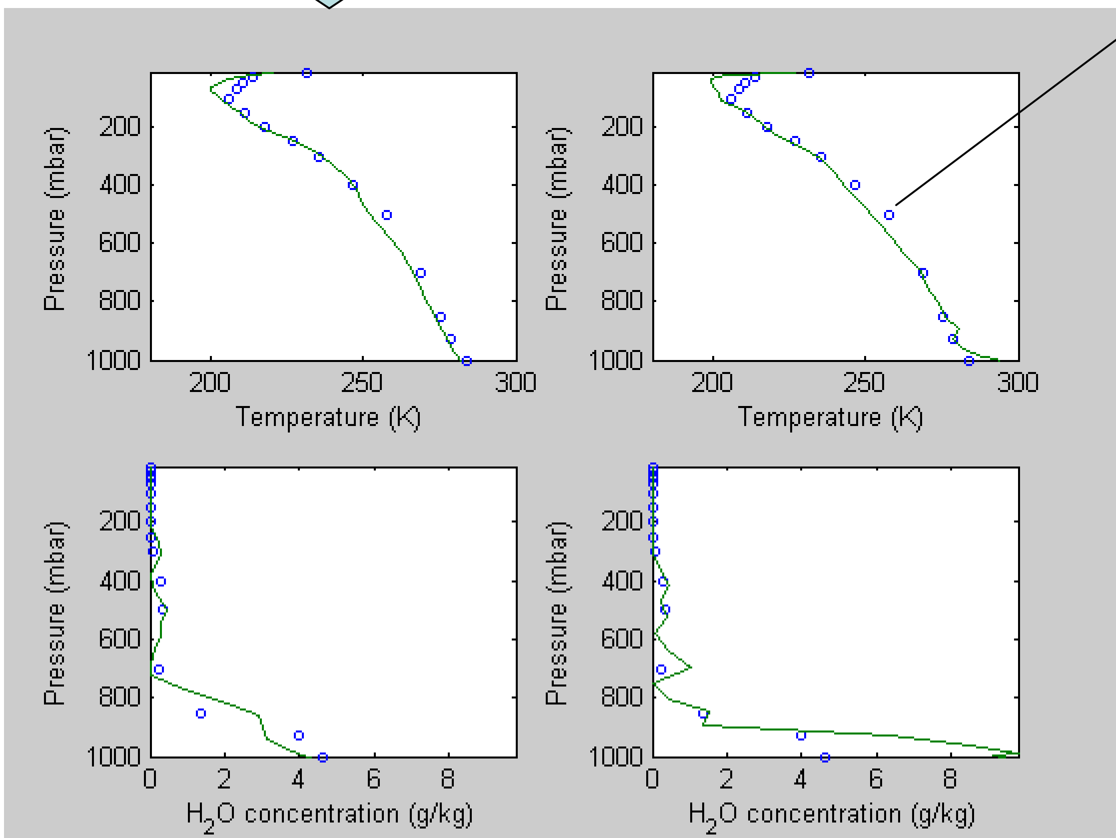
EOF

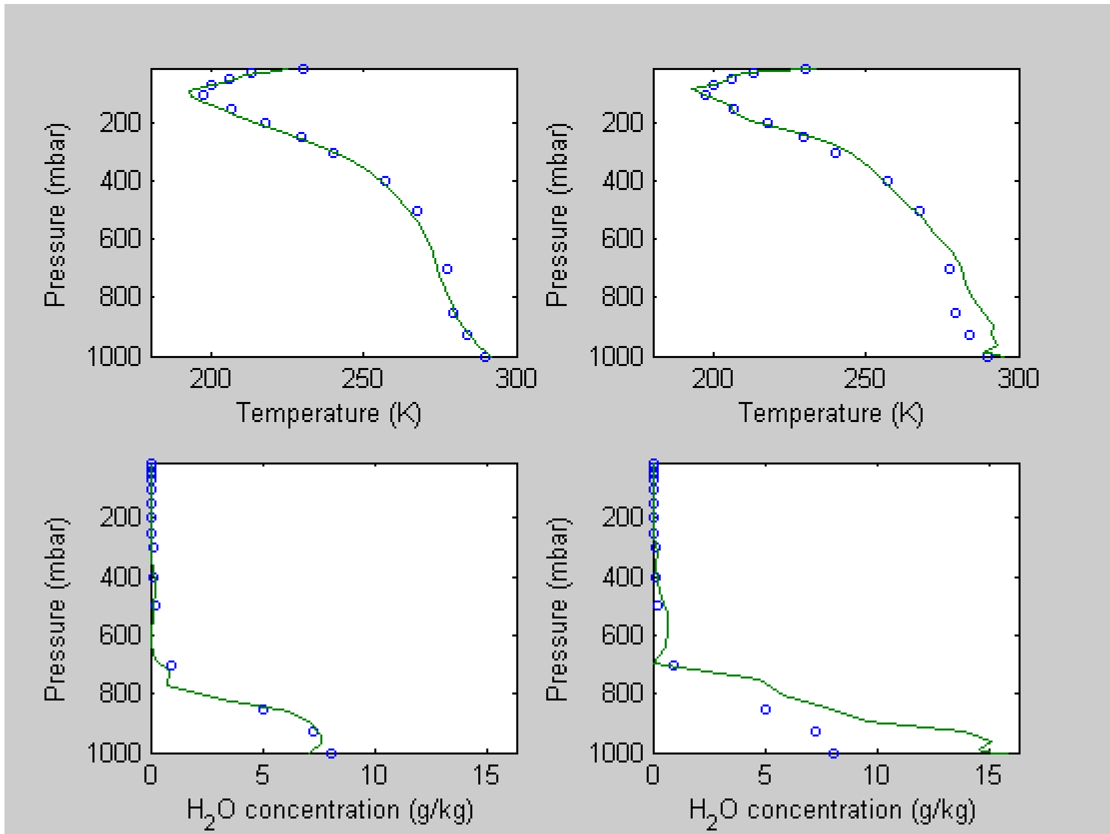
Regression

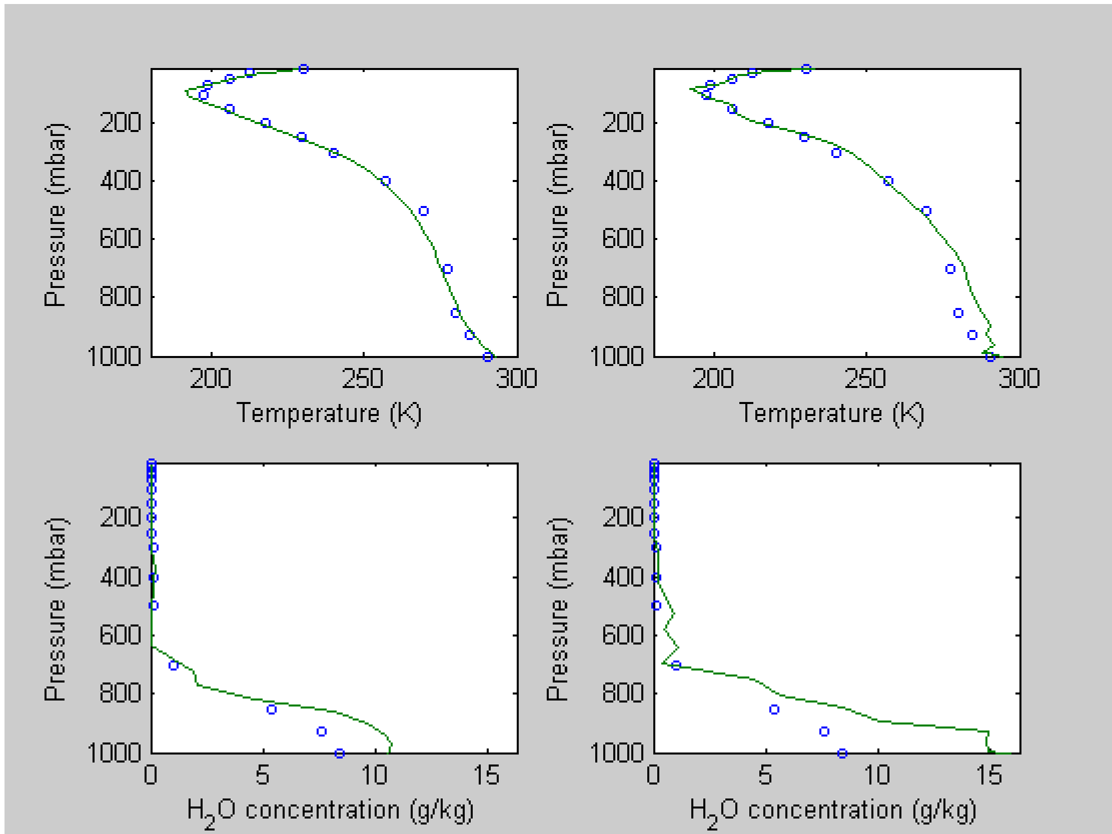
Circles:

ECMWF

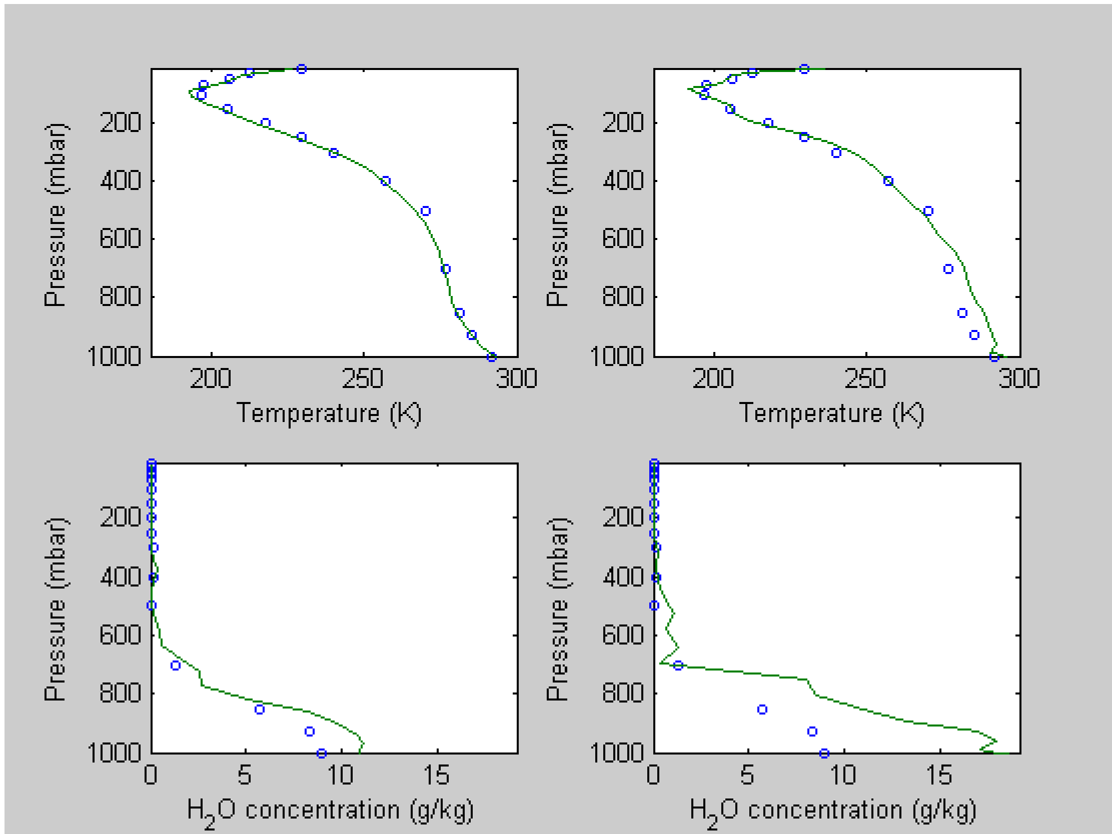
Analysis

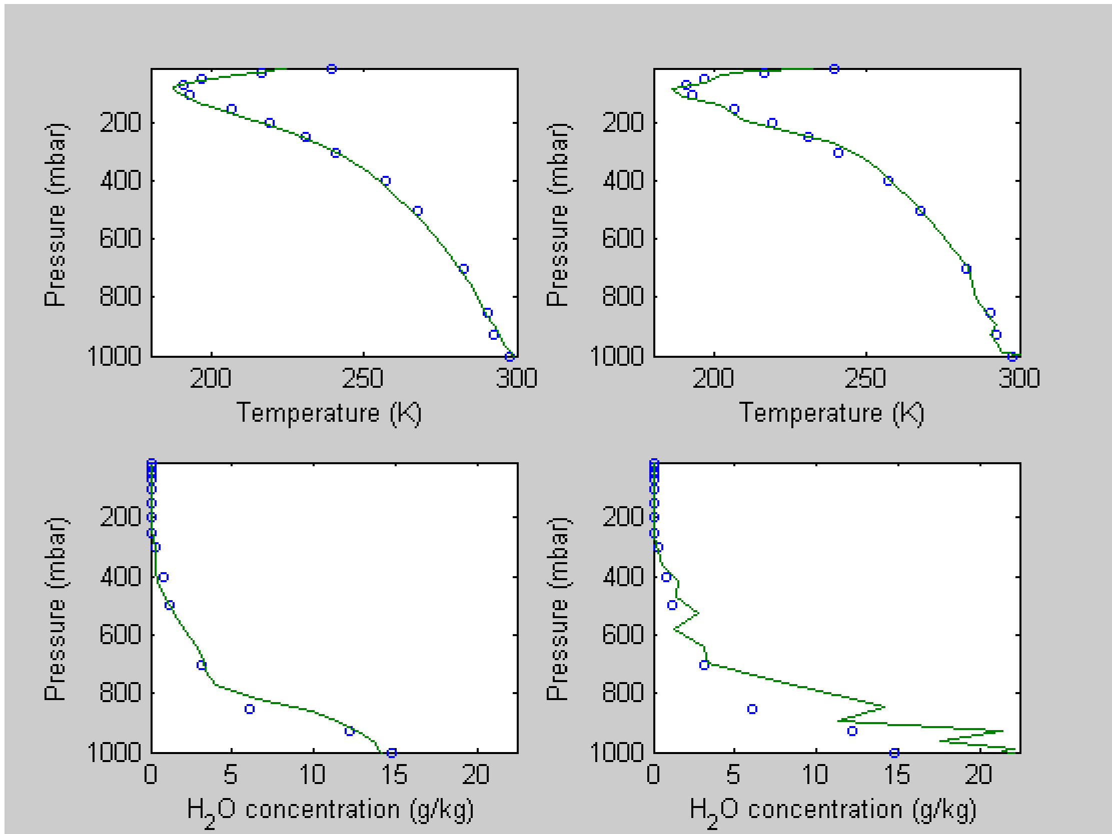


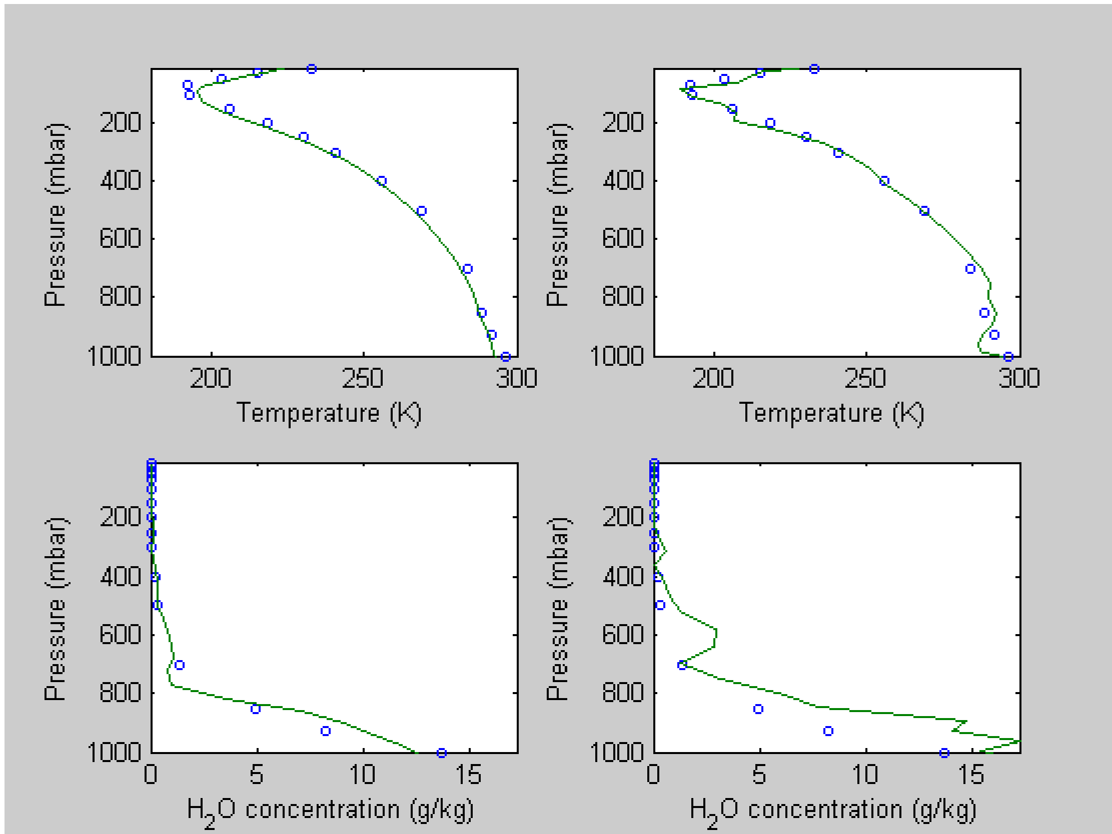

Pressure (mbar)
Temperature (K)
Pressure (mbar)
Temperature (K)
Pressure (mbar)
H2O concentration (g/kg)
Pressure (mbar)
H2O concentration (g/kg)

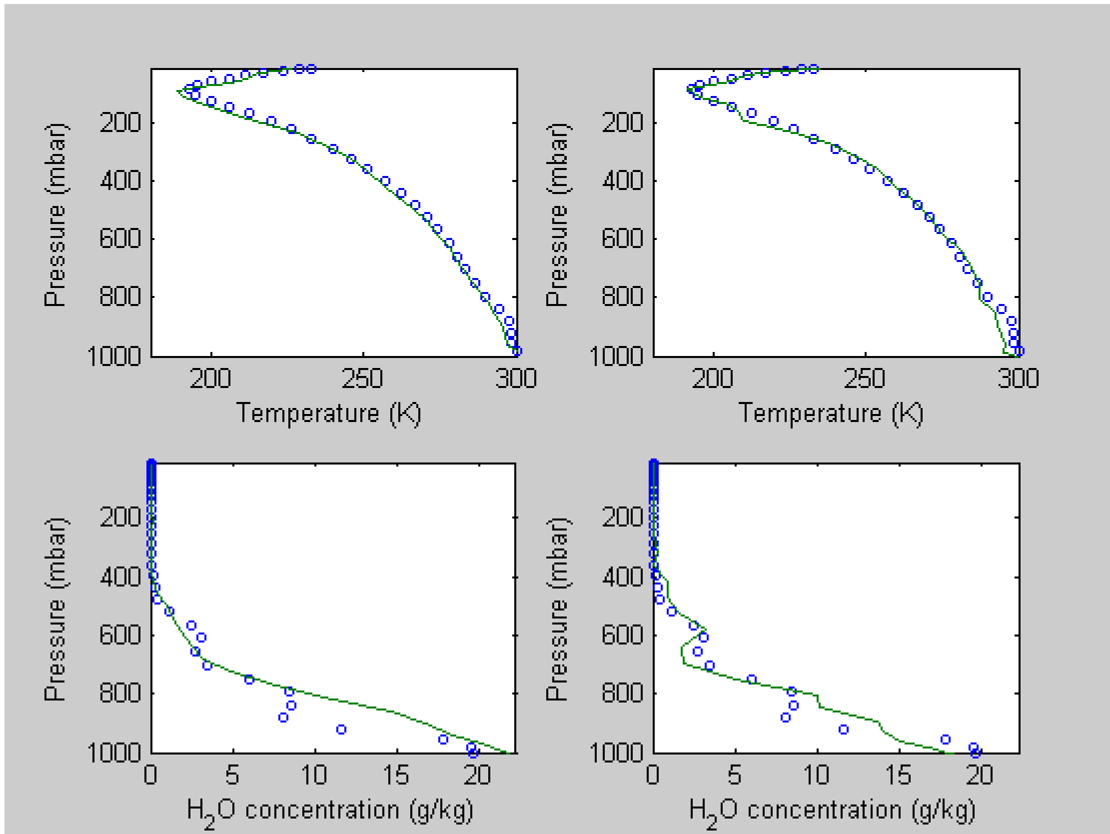


Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
H2O concentration (g/kg)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
H2O concentration (g/kg)

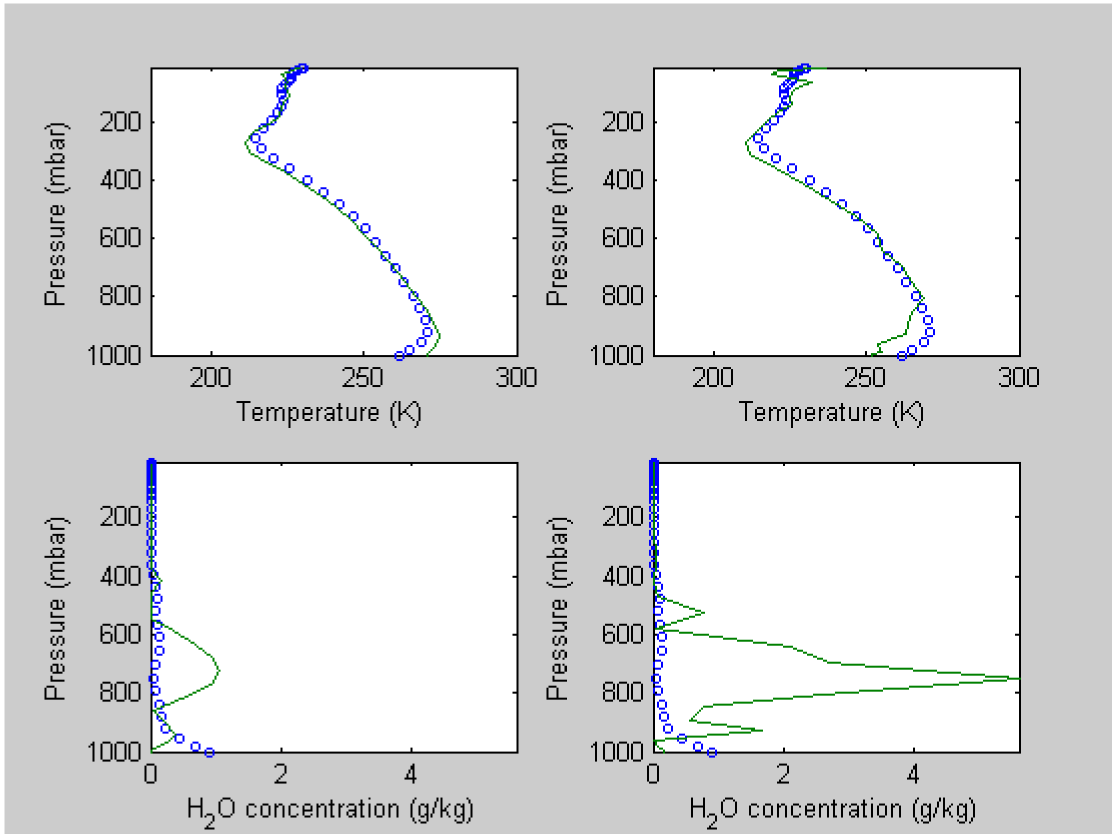

Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
H2O concentration (g/kg)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
H2O concentration (g/kg)


Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
20
H2O concentration (g/kg)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
20
H2O concentration (g/kg)

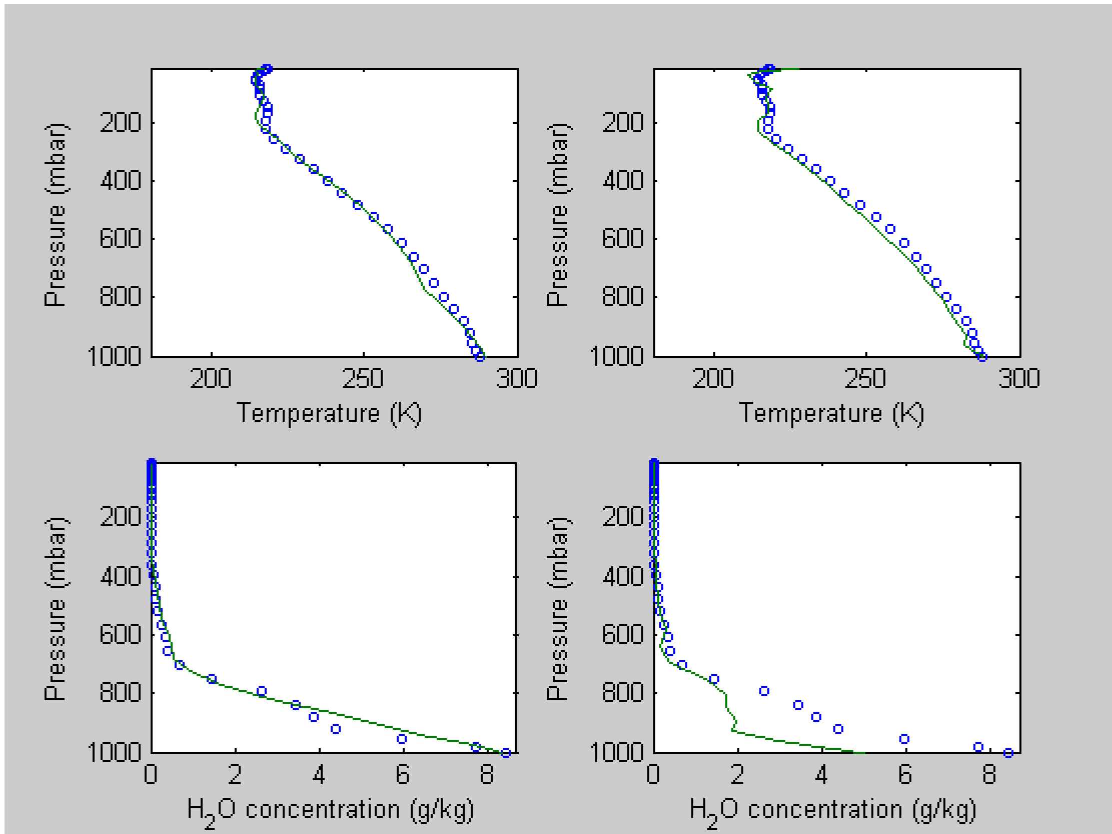

Pressure (mbar)
Temperature (K)
Pressure (mbar)
Temperature (K)
Pressure (mbar)
H2O concentration (g/kg)
Pressure (mbar)
H2O concentration (g/kg)



Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
20
H_2O concentration (g/kg)
Pressure (mbar)
200
400
600
800
1000
0
5
10
15
20
H_2O concentration (g/kg)



Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
H2O concentration (g/kg)
0
2
4



Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
200
250
300
Temperature (K)
Pressure (mbar)
200
400
600
800
1000
0
2
4
6
8
H_2O concentration (g/kg)
Pressure (mbar)
200
400
600
800
1000
0
2
4
6
8
H_2O concentration (g/kg)

Exercise for tomorrow

- Inter-compare
  – Physical Inversion
  – EOF Regression
  – Neural Net
- With NAST-I data (work supported by EUMETSAT)
- With our AERI-like BOMEM FTS (work supported by Italian Ministry for the research)

Research program to speed up physical inversion (next future)

Develop the RTE in EOF-basis

$$\mathbf{r} = \mathbf{K}_T \mathbf{x}_T + \mathbf{K}_w \mathbf{x}_w + \text{h.o.t.};$$

$$\mathbf{r} = \mathbf{R} - \mathbf{R}_o; \quad \mathbf{x}_T = \mathbf{T} - \mathbf{T}_o; \quad \mathbf{x}_w = \mathbf{w} - \mathbf{w}_o$$

$$\mathbf{E} = (\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_M); \quad \mathbf{C} = \frac{1}{M}\mathbf{E}^t\mathbf{E}; \quad \mathbf{S}_R = \text{diag}(\mathbf{C});$$

$$\mathbf{S}_R^{-\frac{1}{2}}\mathbf{r} = \mathbf{S}_R^{-\frac{1}{2}}\mathbf{K}_T\mathbf{S}_T^{\frac{1}{2}}\mathbf{S}_T^{-\frac{1}{2}}\mathbf{x}_T + \mathbf{S}_R^{-\frac{1}{2}}\mathbf{K}_w\mathbf{S}_w^{\frac{1}{2}}\mathbf{S}_w^{-\frac{1}{2}}\mathbf{x}_w$$

$$\begin{cases} \mathbf{y} = \mathbf{S}_R^{-\frac{1}{2}}\mathbf{r} \\ \mathbf{t} = \mathbf{S}_T^{-\frac{1}{2}}\mathbf{x}_T \\ \mathbf{z} = \mathbf{S}_w^{-\frac{1}{2}}\mathbf{x}_w \\ \mathbf{A}_T = \mathbf{S}_R^{-\frac{1}{2}}\mathbf{K}_T\mathbf{S}_T^{\frac{1}{2}}; \quad \mathbf{A}_w = \mathbf{S}_R^{-\frac{1}{2}}\mathbf{K}_w\mathbf{S}_w^{\frac{1}{2}} \end{cases}$$

$$\mathbf{y} = \mathbf{A}_T\mathbf{t} + \mathbf{A}_w\mathbf{z}$$

EOF decomposition of the linearized RTE

$$\mathbf{U}_R ; \quad \mathbf{U}_T ; \quad \mathbf{U}_w$$

$$\begin{cases} \mathbf{c}_y = \mathbf{U}_R^t \mathbf{y} \\ \mathbf{c}_t = \mathbf{U}_T^t \mathbf{t} \\ \mathbf{c}_w = \mathbf{U}_w^t \mathbf{z} \end{cases}$$

$$\mathbf{U}_R^t \mathbf{y} = \mathbf{U}_R^t \mathbf{A}_T \mathbf{U}_T \mathbf{U}_T^t \mathbf{t} + \mathbf{U}_R^t \mathbf{A}_w \mathbf{U}_w \mathbf{U}_w^t \mathbf{w}$$

$$\boxed{\mathbf{c}_y = \mathbf{G}_T \mathbf{c}_t + \mathbf{G}_w \mathbf{c}_w}$$

$$\begin{cases} \mathbf{G}_T = \mathbf{U}_R^t \mathbf{A}_T \mathbf{U}_T \\ \mathbf{G}_w = \mathbf{U}_R^t \mathbf{A}_w \mathbf{U}_w \end{cases}$$

Conclusions

- The inversion tools developed within the ISSWG activities by the DIFA-IMAA-IAC groups have been presented
- A comparison have been provided of the relative performance of the various methods (although more work is needed)
- A tentative list for now see at the top the
  1. Physical inversion (not suitable for operational end-uses)
  2. Neural Network (very fast, still complex to train, its dependence on the training data set has to be assessed)
  3. EOF Regression (appealing for its simplicity, the training needs to be localized, does not seem to provide reliable results for H_2O)