Direct methods for the inversion of limb scattering measurements by machine learning techniques.

D. Fussen, N. Baker, A. Berthelot, E. Dekemper, P. Demoulin , G. Franssens, P. Gramme, N. Mateshvili, K. Rose and S. Sotiriadis

Royal Belgian Institute for Space Aeronomy

ATMOS2024 @ Bologna

Outline

- Context and objectives
- Inverse problems
- Limb geometry
- Test bench with OMPS
- Conclusions

Context and objectives

- **<u>ALTIUS level 2 processor</u>** is well advanced for ozone retrievals.
  - it is based on standard methods
  - heavy hardware constraints : f.i. no GPU allowed -> no “direct chain” possible with our Radiative Transfer Model (RTM)
  - proxies and look-up tables and “L-M” algos
  - many instrumental side effects are possible and difficult to manage: straylight, convolutions,..etc

- Objective: **<u>to explore “direct methods”</u>** = combining two powerful numerical weapons
  - use of orthogonal function bases given by Principal Component Analysis (PCA)
  - nonlinear regression by Machine Learning (ML)

Inverse problems

- Probably the most frequent problems in experimental physics: the retrieved quantity results from one or several integrations of an unknown distribution
- Huge amount of references and methods: Bayesian optimal estimation, Philips-Twomey-Tikhonov regularization, constrained non-linear LS (L-M) for L2, L1,..norms, linear and log (Chahine) relaxation methods, Backus-Gilbert, Maximum Entropy Methods, ..etc

Inverse problem:
measure "**y**" , then compute '**x**'

$$\boxed{\mathbf{y} = \mathbf{F}(\mathbf{x}) + \epsilon}$$

- In the Bayesian approach, there is a prior knowledge of $\mathbf{x} = \mathbf{x}_a$ characterized by an associated covariance matrix $\mathbf{S}_a$ that will combine with the measurement error covariance matrix $\mathbf{S}_\epsilon$ to define the iterative update of the solution state vector as:

$$\mathbf{x}_{i+1} = \mathbf{x}_i + (\mathbf{S}_a^{-1} + \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i)^{-1} [\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} (\mathbf{y} - \mathbf{F}(\mathbf{x}_i)) - \mathbf{S}_a^{-1} (\mathbf{x}_i - \mathbf{x}_a)]$$

- In the least-squares Philips-Twomey-Tikhonov method (Twomey, 1977), the stabilization of the ill-posed problem is usually achieved by using a regularizing operator $\mathbf{H}$ embedded in the Levenberg-Marquardt algorithm (Marquardt, 1963):

$$\mathbf{x}_{i+1} = \mathbf{x}_i + (\mathbf{K}_i^T \mathbf{K}_i + \gamma_i \mathbf{H})^{-1} \mathbf{K}_i^T (\mathbf{y} - \mathbf{F}(\mathbf{x}_i))$$

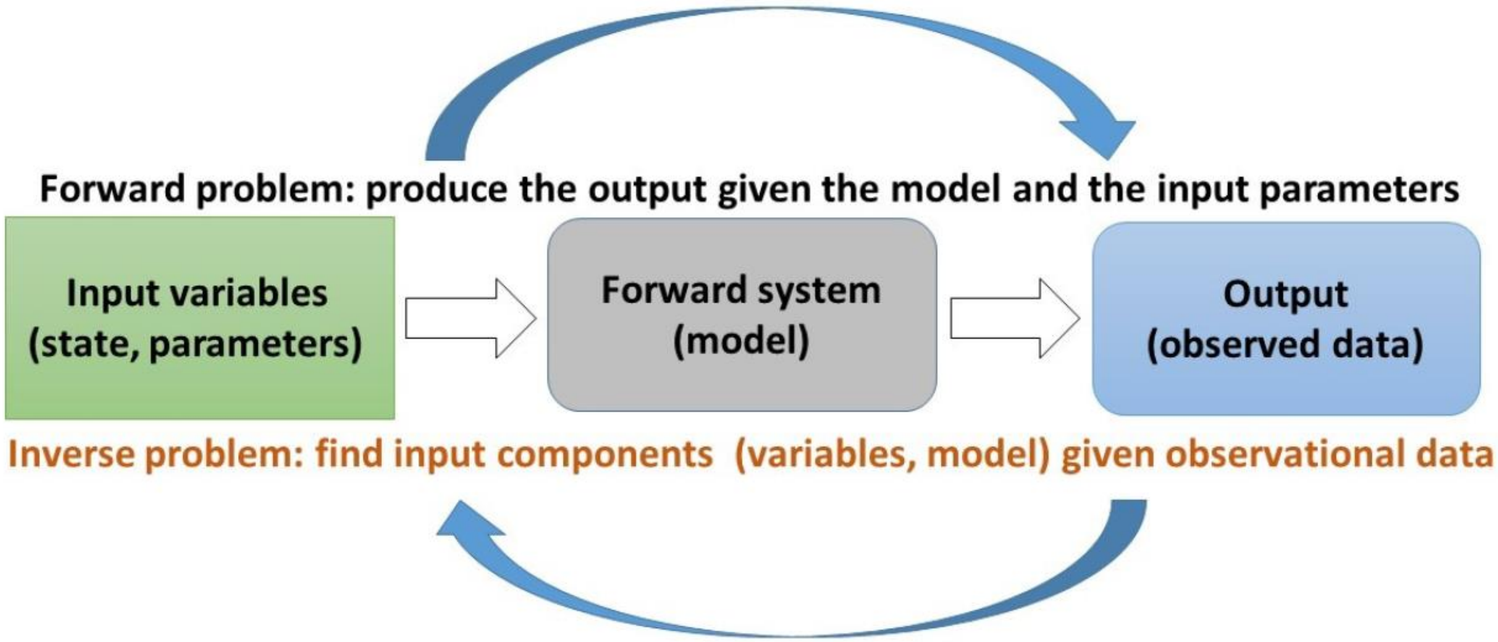
- Relaxation methods (Chahine, 1968; Twomey et al., 1977) are slow and need to be stopped when noise amplification starts to dominate but they do not require the computation of the Jacobian at every step:

$$\mathbf{x}_{i+1}^j = \mathbf{x}_i^j \frac{\mathbf{y}^j}{\mathbf{F}(\mathbf{x}_i)^j}$$

Trying direct inverse methods…

- WHEN a large number of observations has to be processed, the TOTAL computational cost may be considered. **How many calls to the forward model ?**
- Non-linearities in retrieval imply **iterative** schemes.
    - f.i. in ALTIUS, resolution is driving the number of forward model calls by the L-M algorithm.
    - even worse: all intermediate computations along the minimization path, including Jacobians, are lost.
- A large “training” set (**LTS**) is affordable at an equivalent computing load. Hereafter, “large” means 25 600 synthetic simulations by our Monte-Carlo radiative transfer code “SmartG”

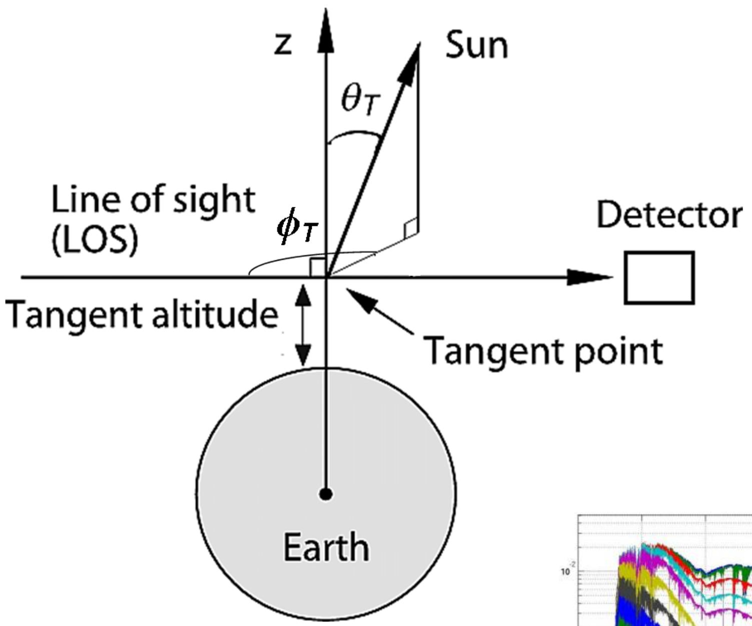
-> 2 weeks / 2 Tera photons shot / 0,3 % precision

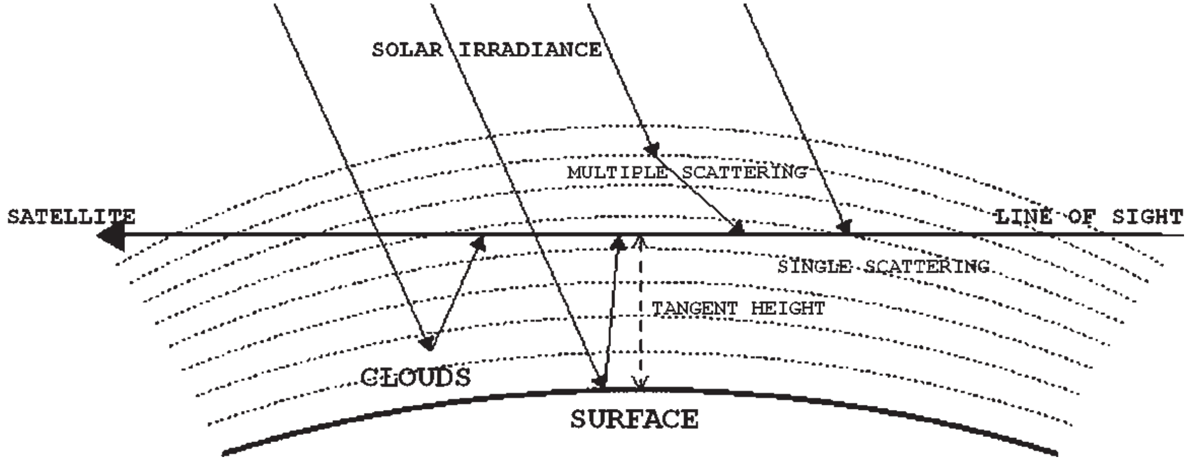


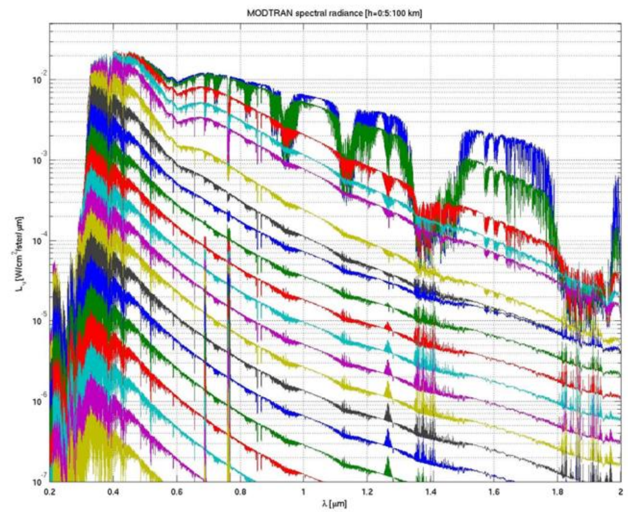
Direct inversion ?

- Generate a **LTS**
- Just **swap** green and blue boxes.
- Replace the forward model by a **black box**.
- For LTS, force the **black box** to predict the **green box** from the **blue box**.

Focus on solar limb scattering geometry



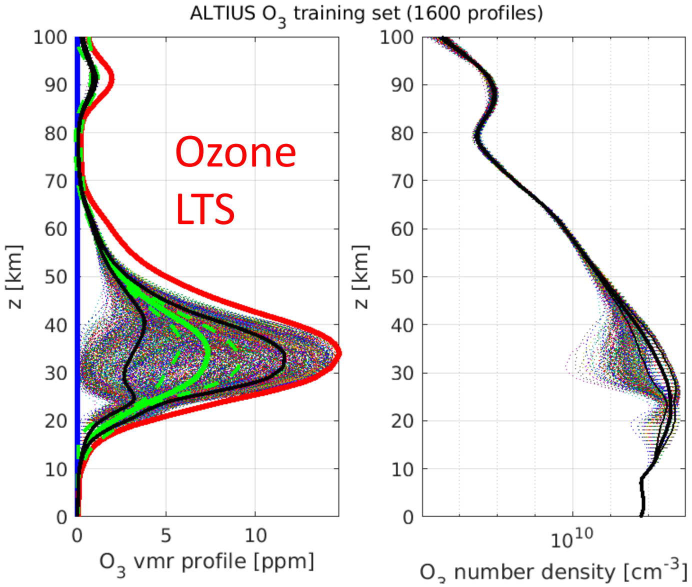


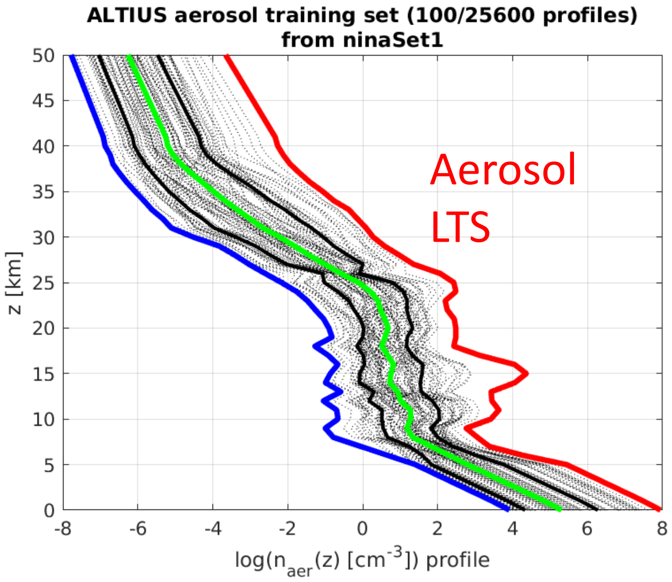


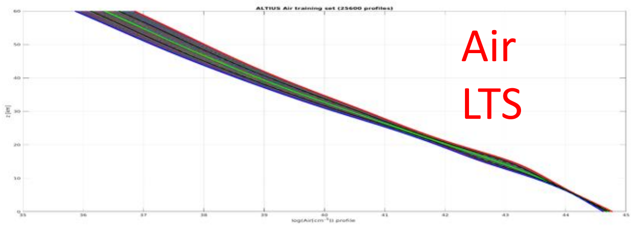
<u>Simplified atmosphere:</u>

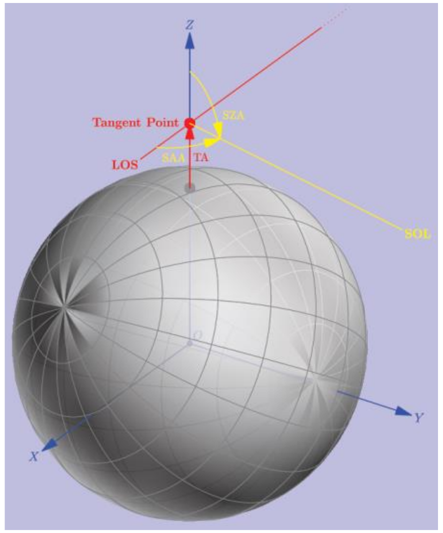
air + ozone + stratospheric aerosols + effective albedo + solar irradiance @ z=0:1:100 km

LTS Factory



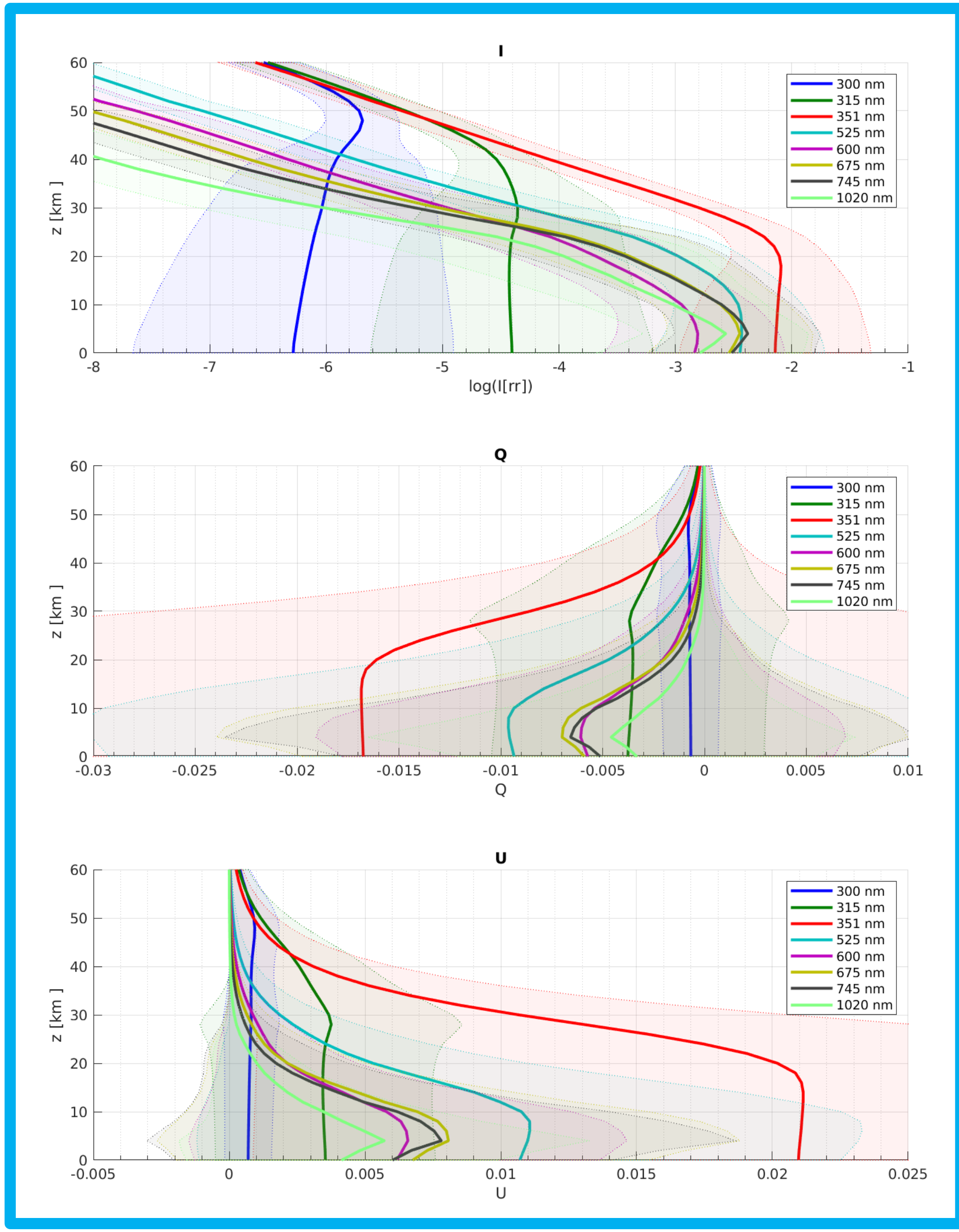






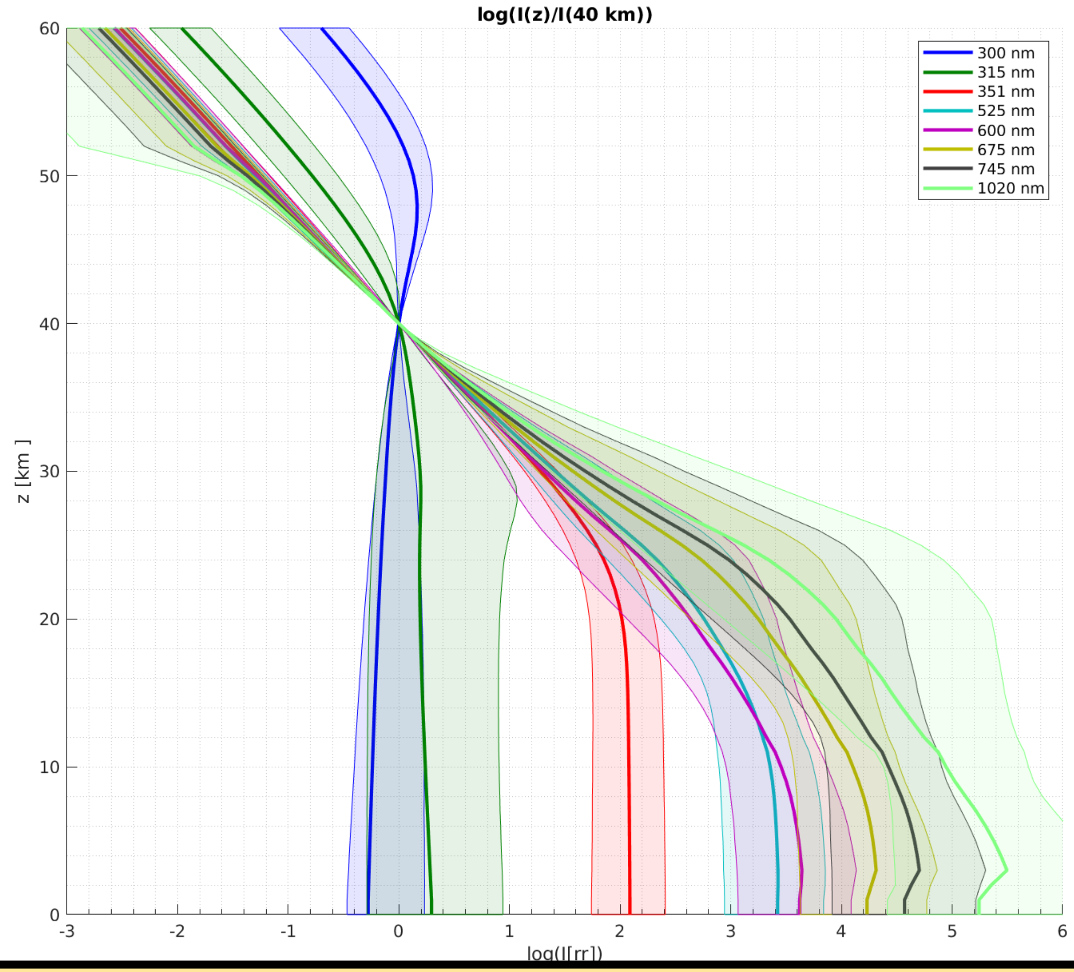
+ PSD, albedo
+ Solar angles
LTS

$$\lambda = 300, 315, 351, 525, 600, 675, 745, 1020\,\mathrm{nm}$$



Stokes vector profiles

ONNI: Ozone Neural Network Inversion

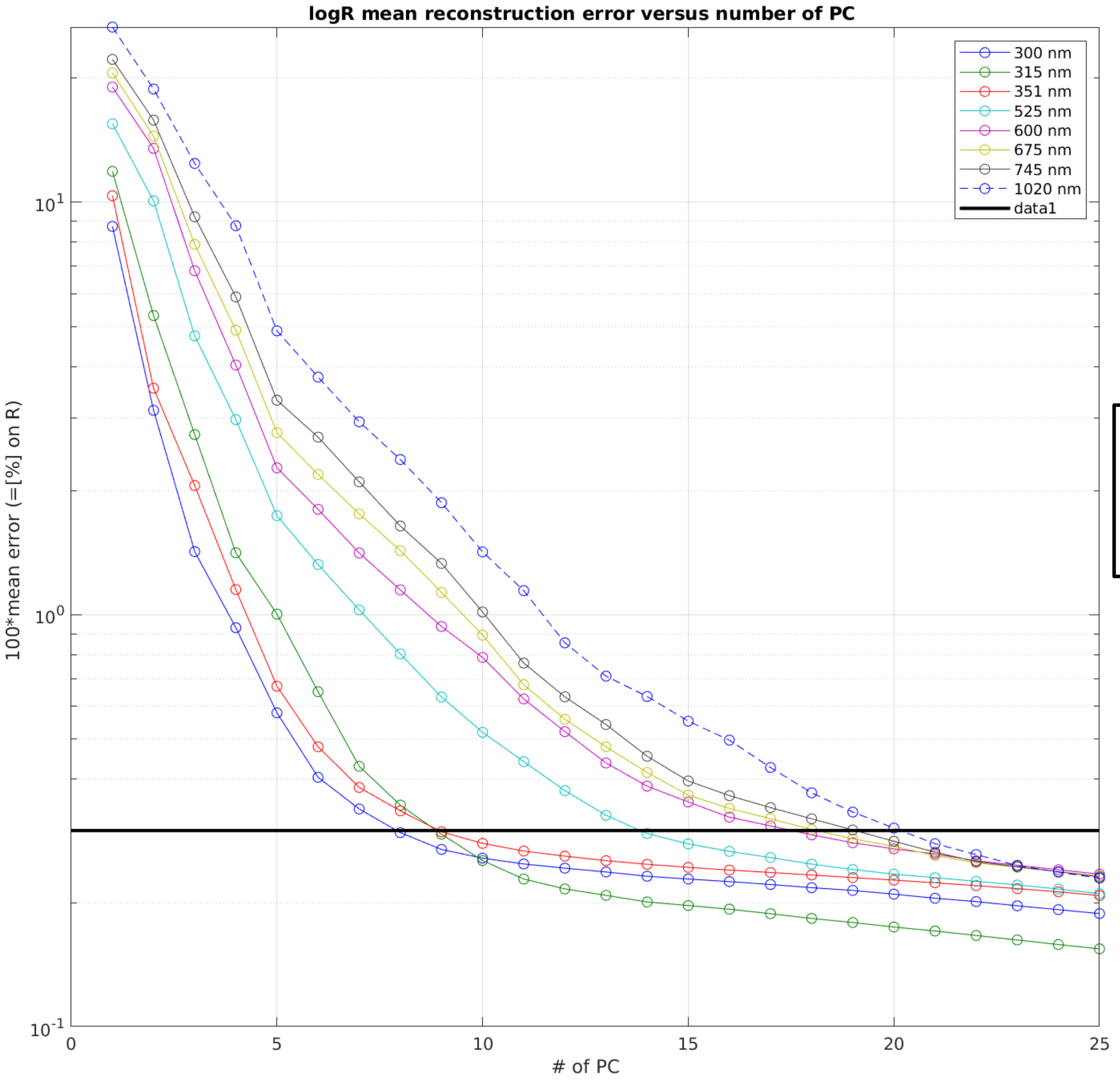


limited to unpolarized instrument for the time being

Key concept 1: what is the information content of a radiance profile ? -> LTS PCA !!!

Ch / λ [nm]	# PC
1 / 300	8
2 / 315	9
3 / 351	9
4 / 525	14
5 / 600	18
6 / 675	19
7 / 745	20
8 / 1020	21

@ 0.3 % accuracy level



Information content is reduced from 8*61=488 to 118

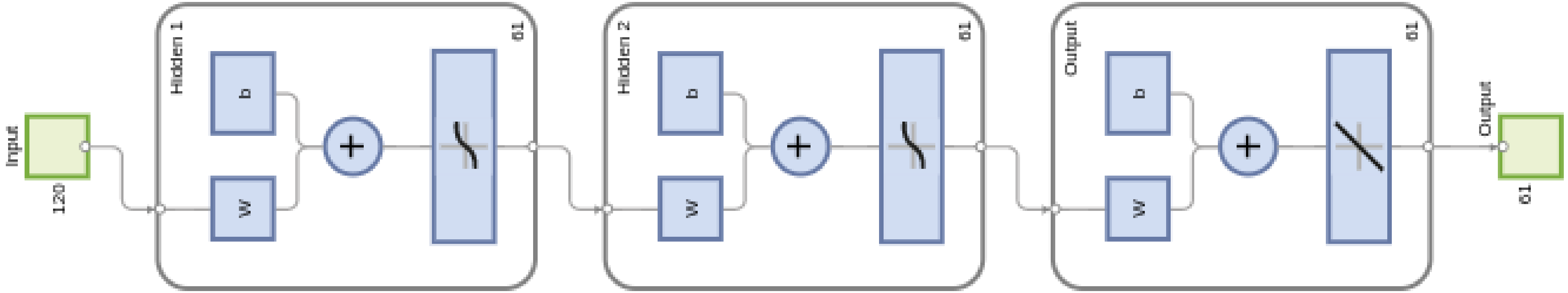
Ch / λ [nm]	# PC
1 / 300	8
2 / 315	9
3 / 351	9
4 / 525	14
5 / 600	18
6 / 675	19
7 / 745	20
8 / 1020	21

Define NRV (Normalized Radiance Vector)
a1,1
a1,2
⋮
a1,8
a2,1
⋮
a2,9
⋮
a8,1
⋮
a8,21
SZA
SAA

length(RV)=
91+2=93

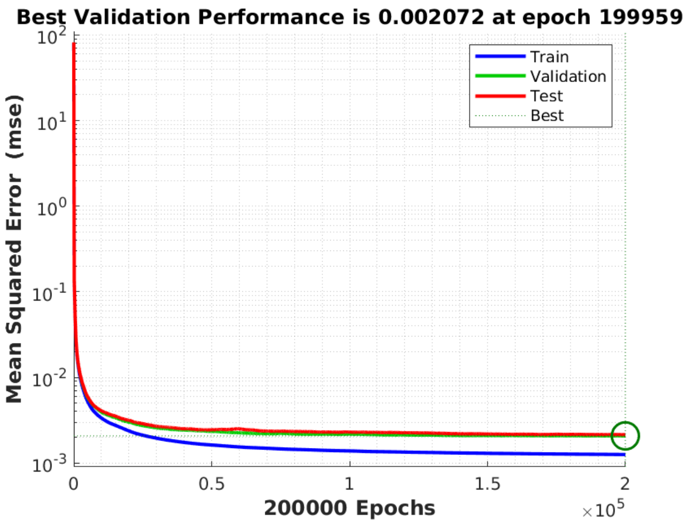
if all (8) channels
are considered

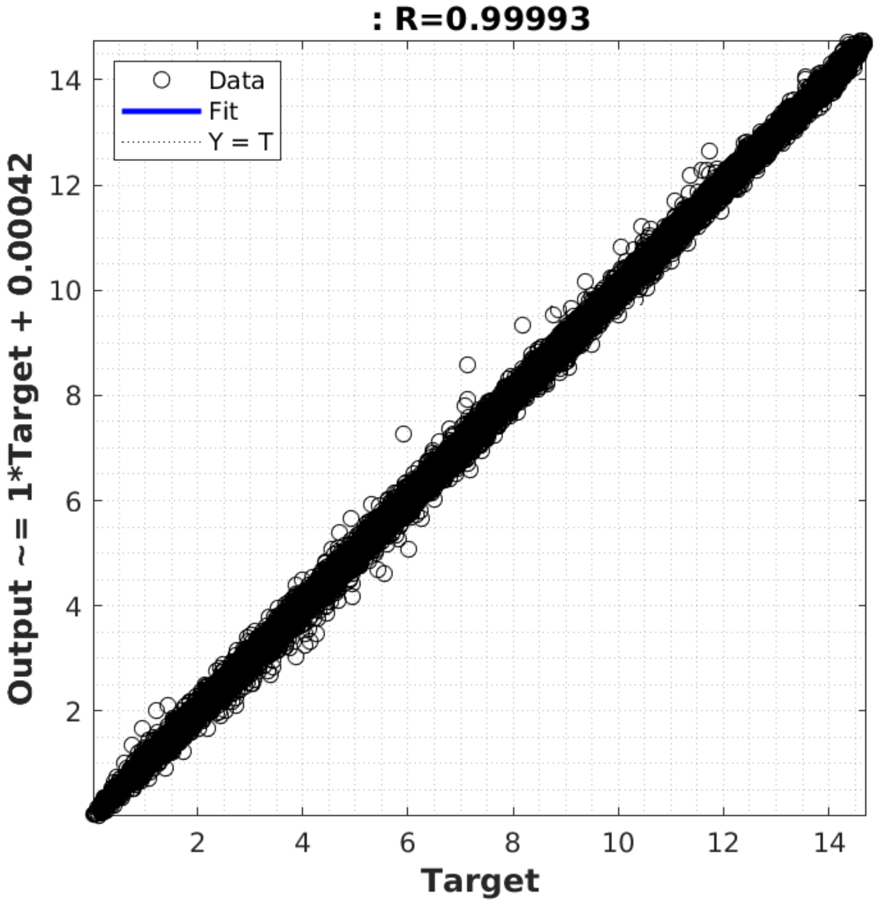
Key concept 2 : build an reverse mapping (118+2) NRV to ozone vmr profile (61)



How to map a large measurement vector of onto another large state vector by a nonlinear transformation ?

No LUT ! → Use an Artificial Neural Network

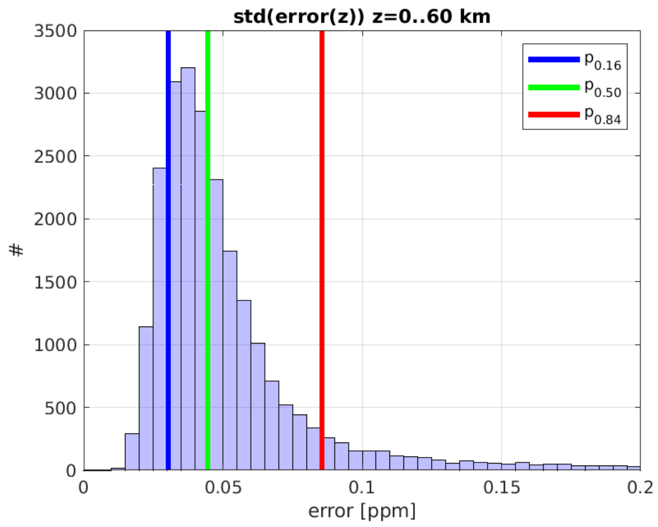


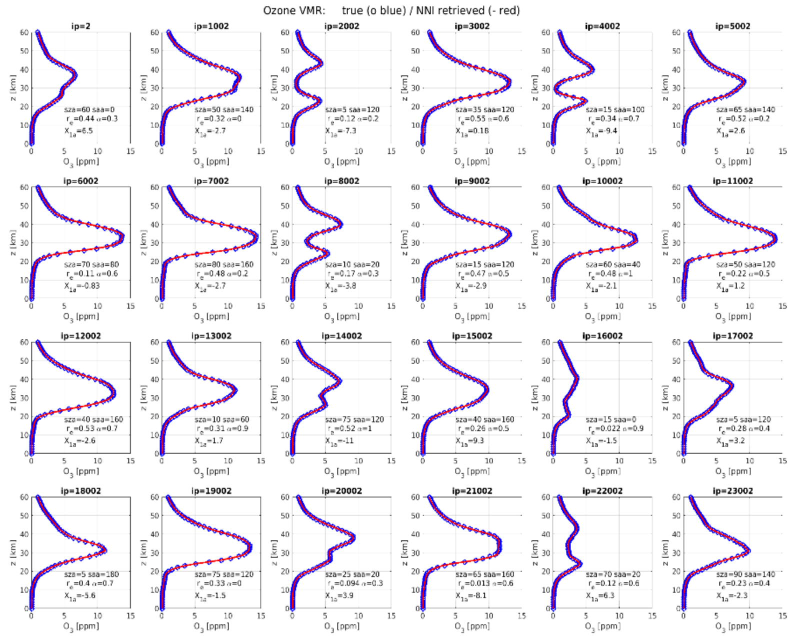


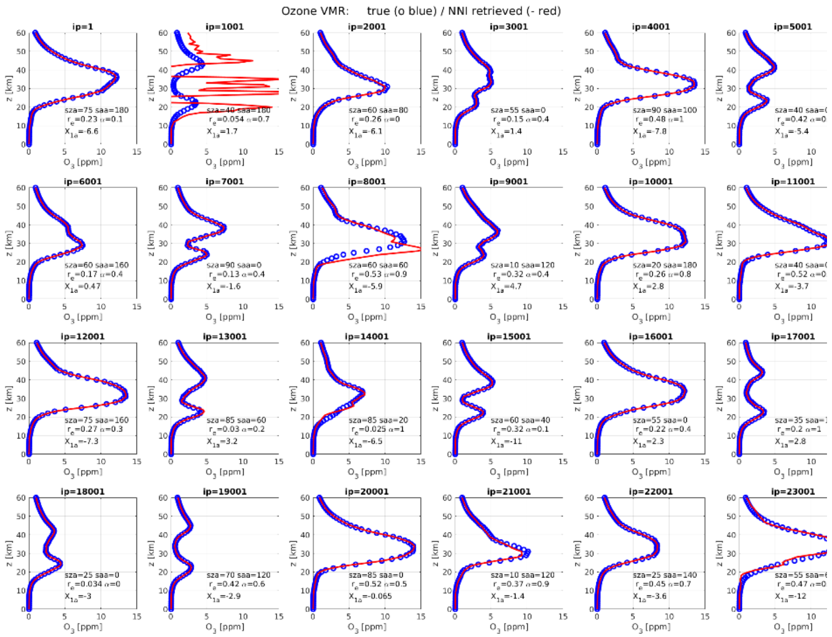
ANN topology is: 120 x 61 x 61 x 61
case subsets: training= 70 % / convergence= 15% / test=15 %
algo= Scaled Conjugated Gradients

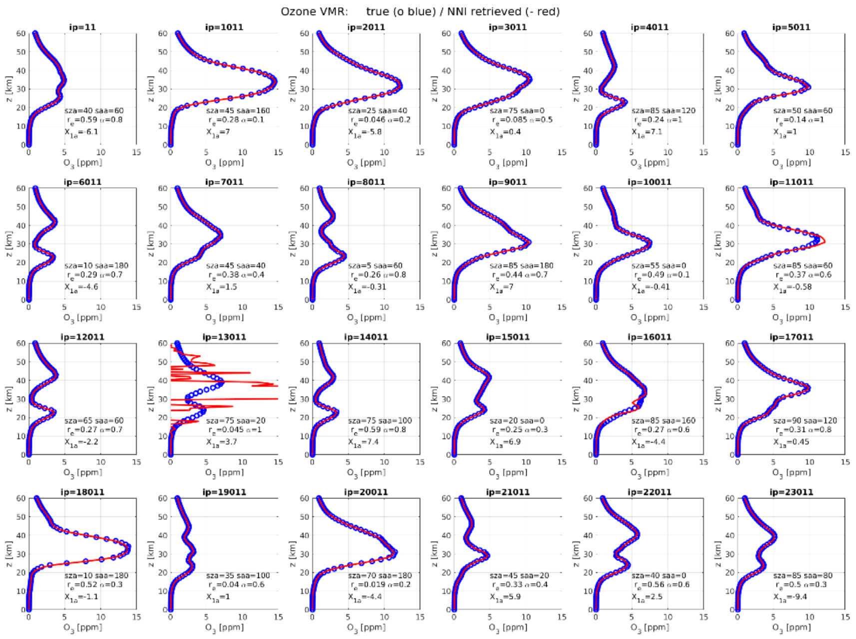
<u>Key concept 3</u> : reverse mapping by a shallow « deep » neural network

Validation wrt full data set (25 600 cases)







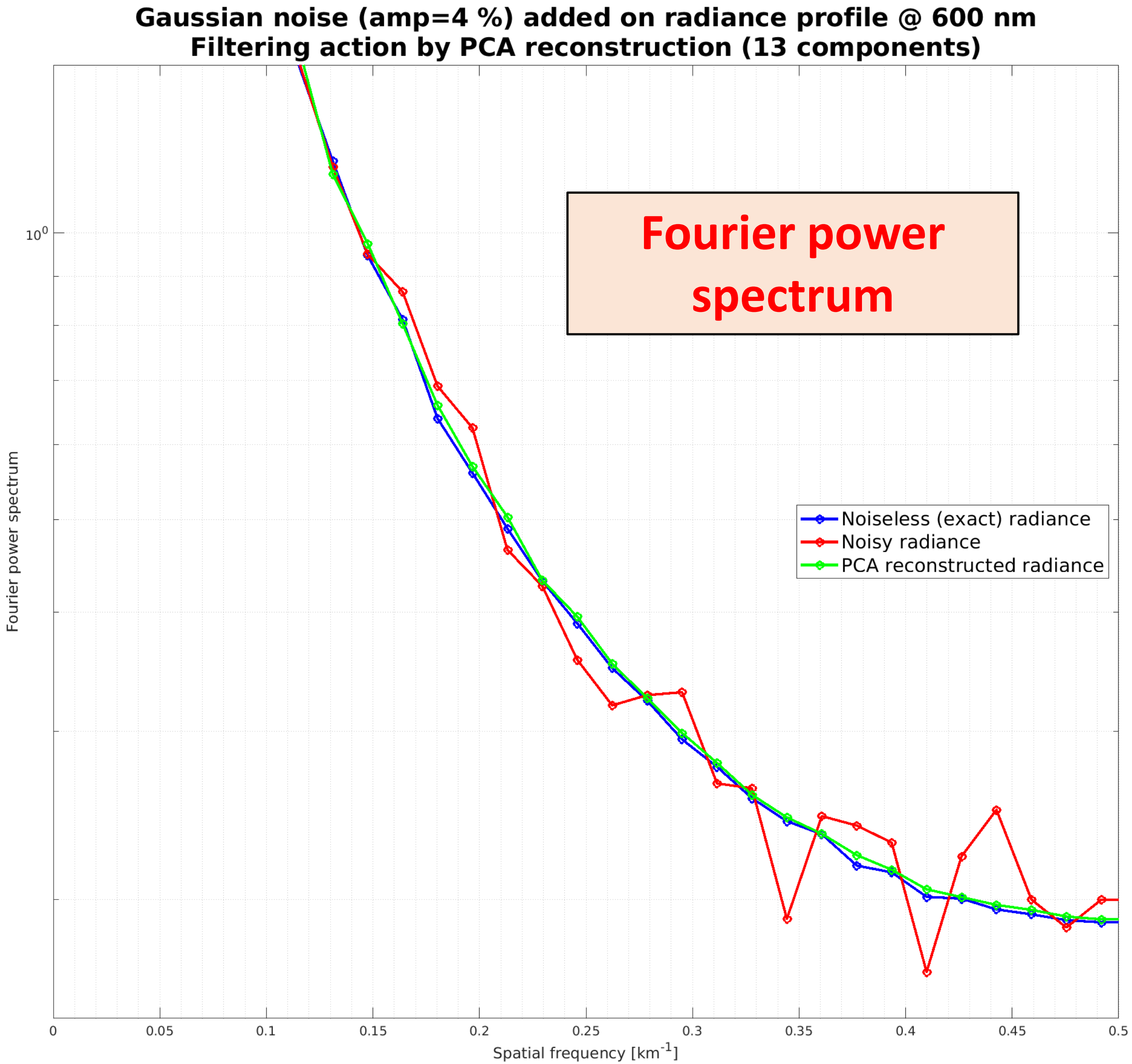


Mostly a fair agreement .. with a few hallucinations

The Magic of radiance PCA !

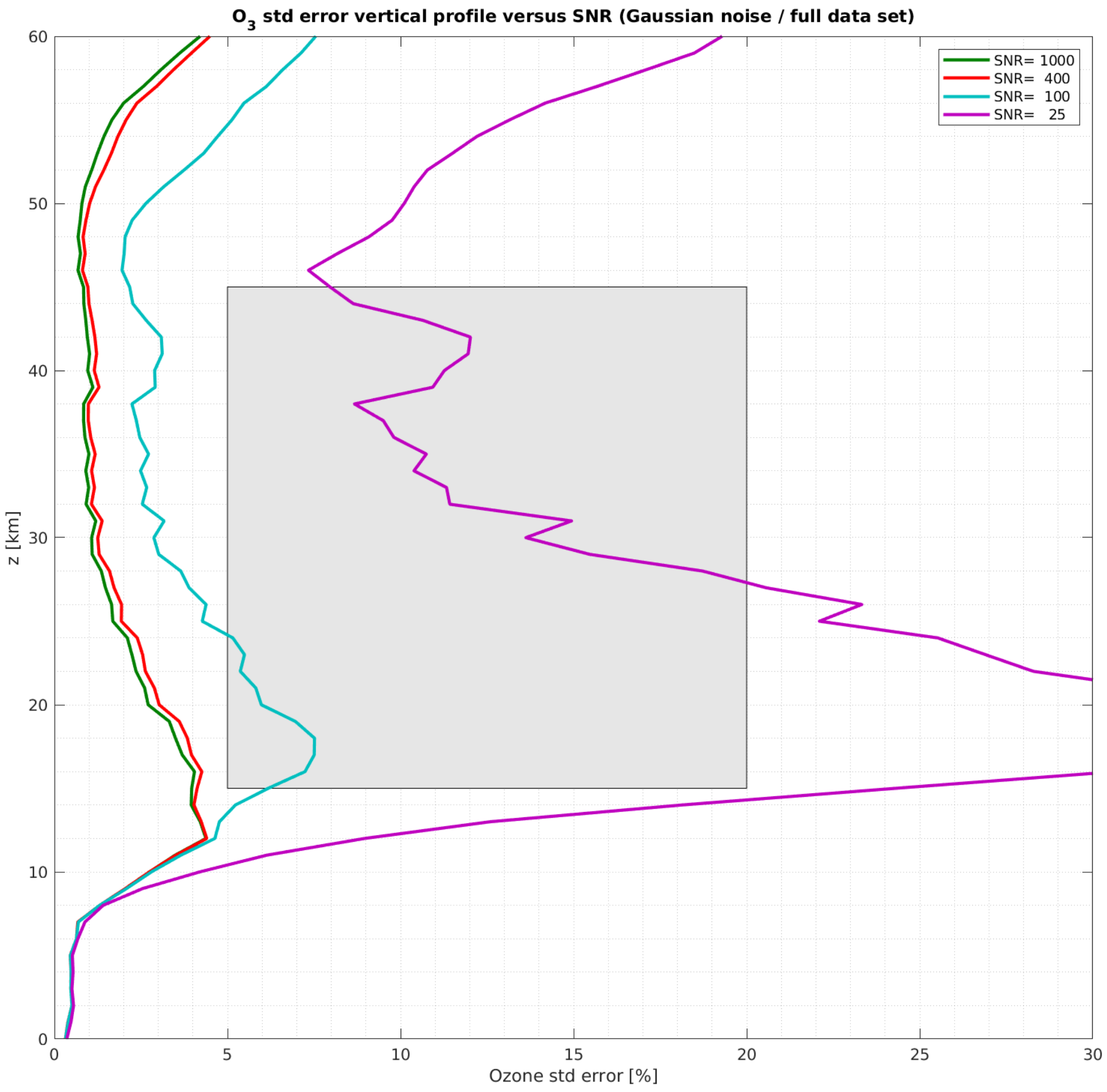
1. it tells you the **information content** of your observation.

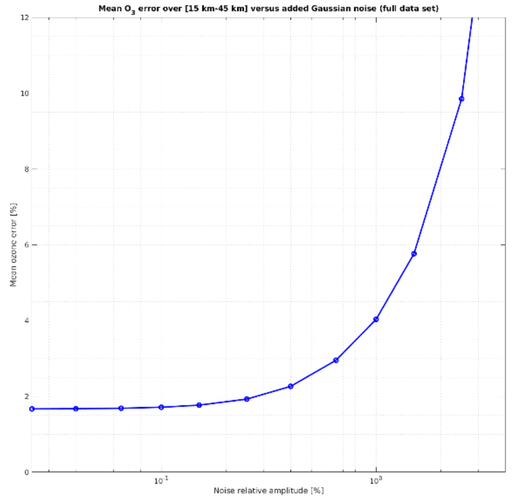
2. it is a powerful (and natural) **denoising filter** → no need to regularize the inversion process.

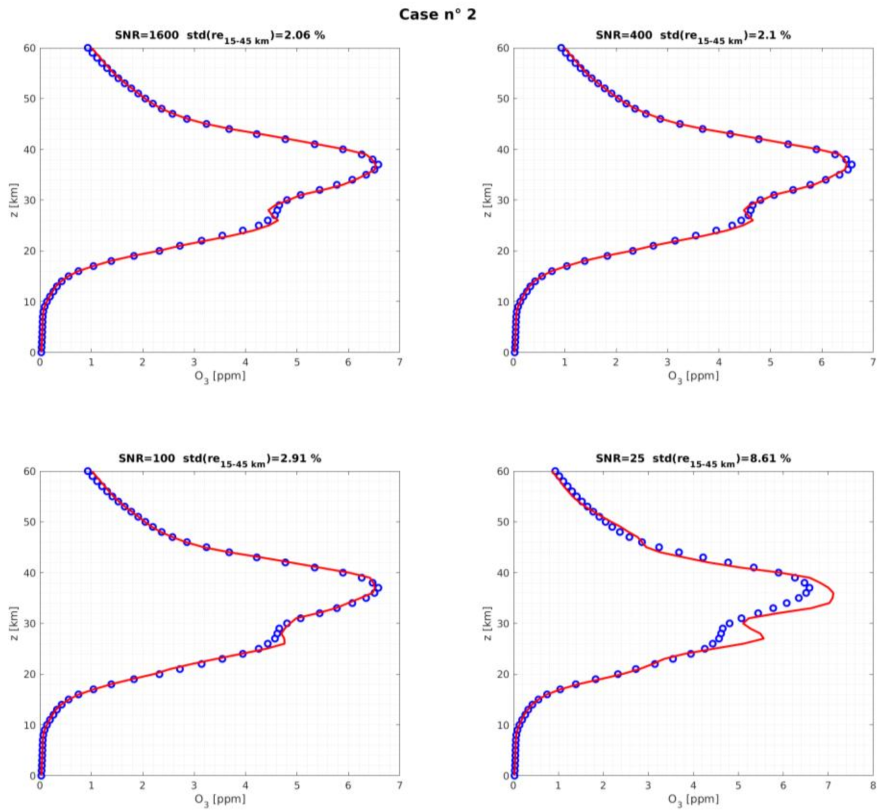


Check **robustness against noise.**

e.g. ALTIUS requirements=[5 %, 20 %] in the 15-45 km range

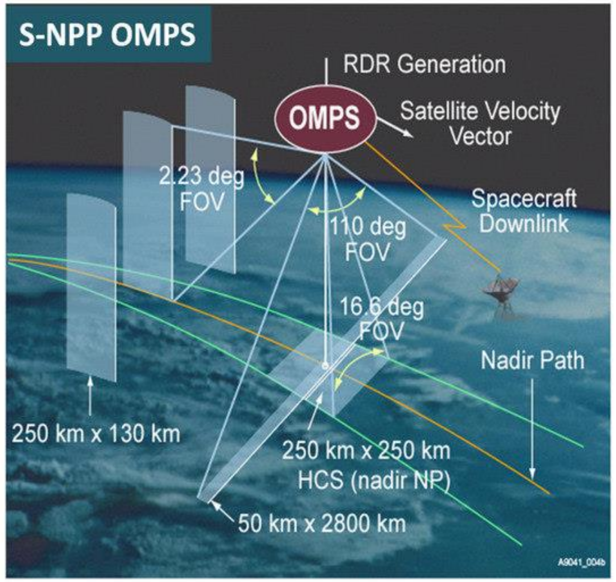


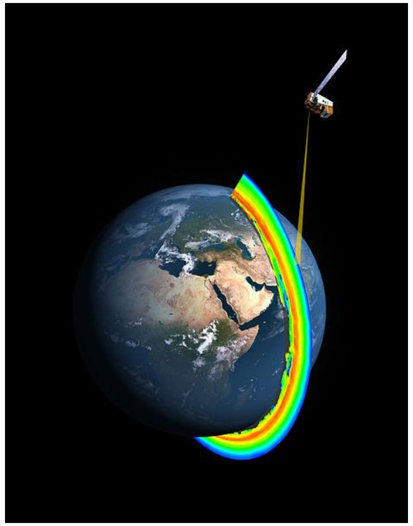




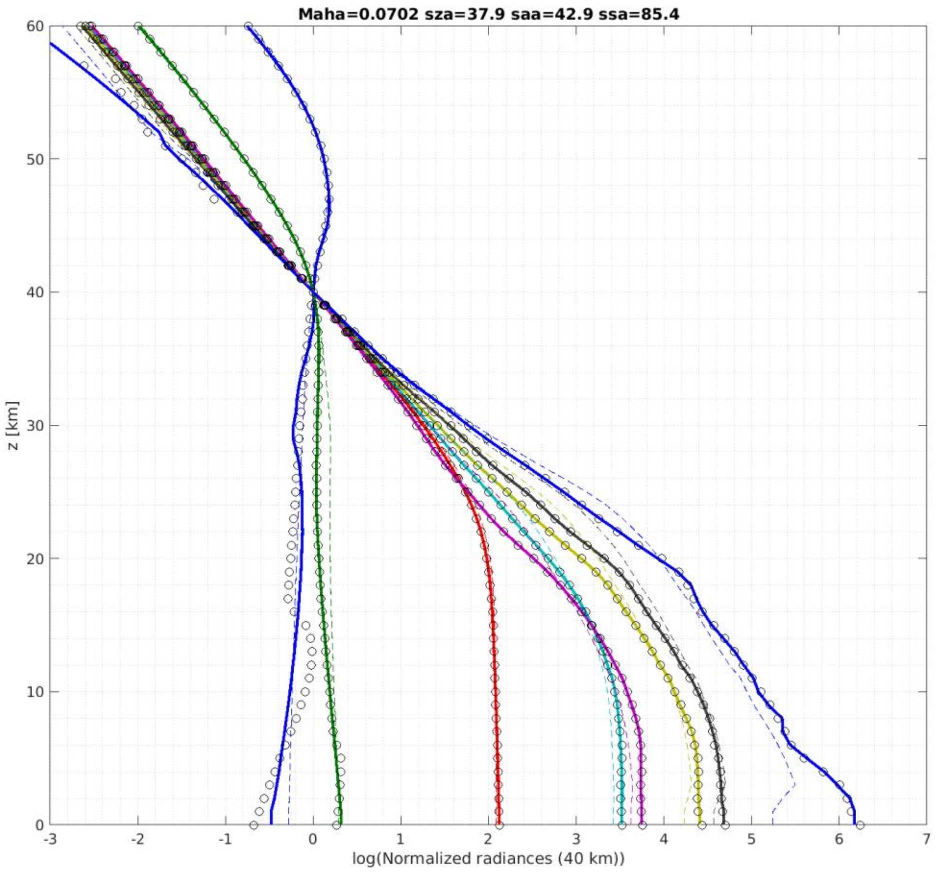
ONNI applied to OMPS data

- target is March 2016 (31 x 15 x 140 potential observations)
- processing of **RAW L1 data**. No instrumental functions, no cloud detection, no TGH correction, no straylight, no extensive error budget,..etc
- objective: to show that ONNI gives “reasonable” (or even “good”) results when applied to a real case with respect to other methods. Fine tuning is possible but out of scope.
- Two competitors:
  - NASA v2.6 (Kramarova et al. [2018])
  - BREMEN (Arosio et al. [2018])

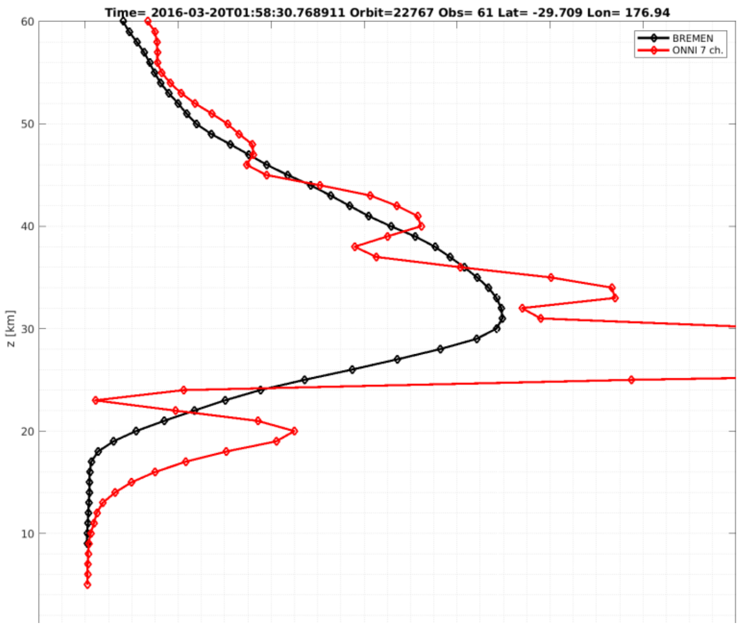




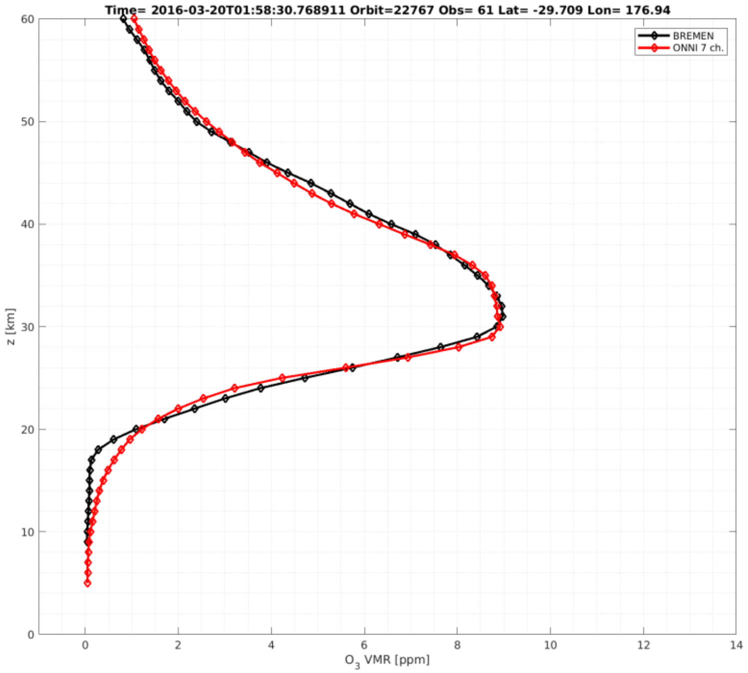
1020 nm OMPS channel is contaminated → ONNI was recomputed for 7 channels



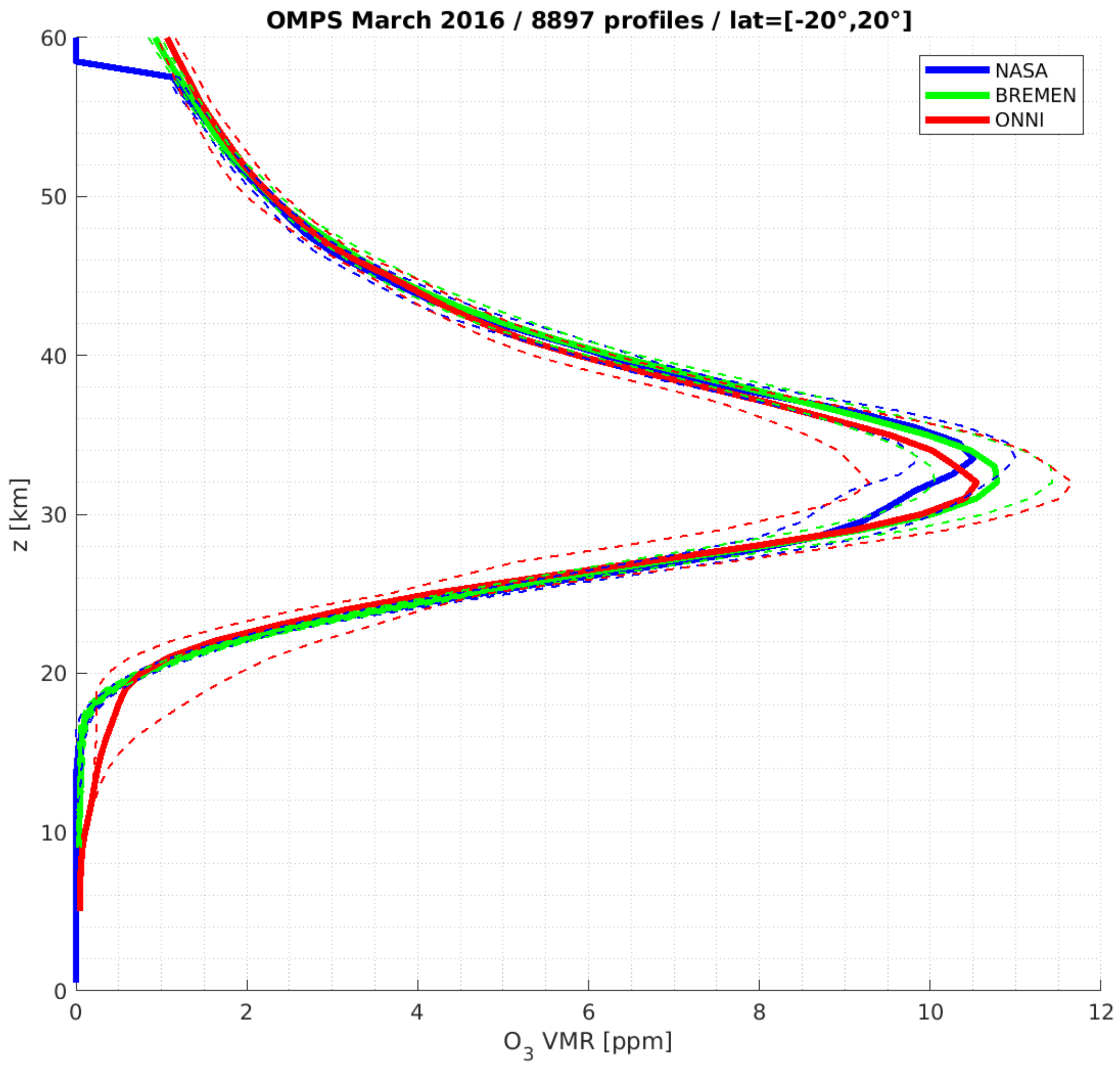
8 ch



7 ch

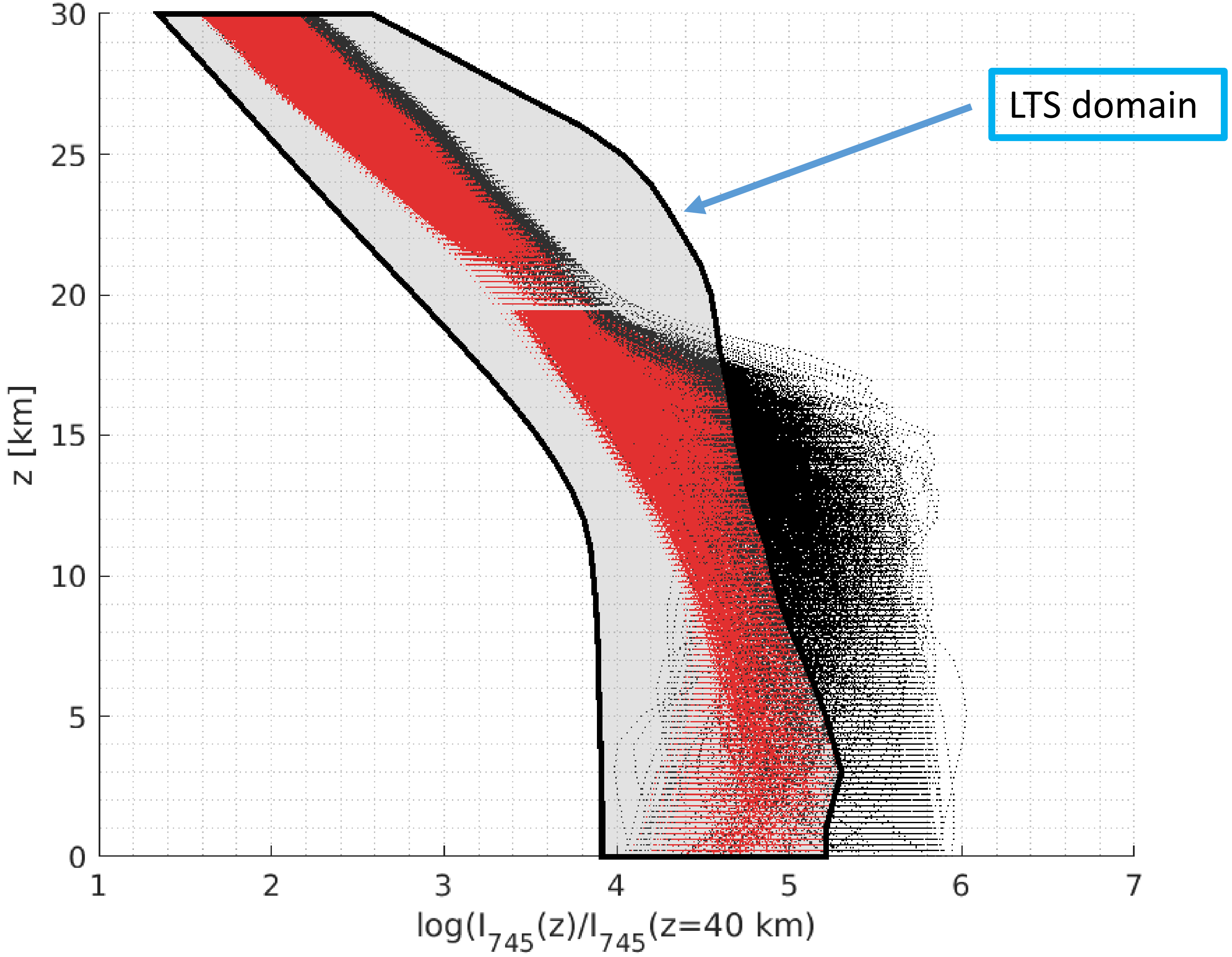


Intercomparison in tropical band (as Arosio et al)

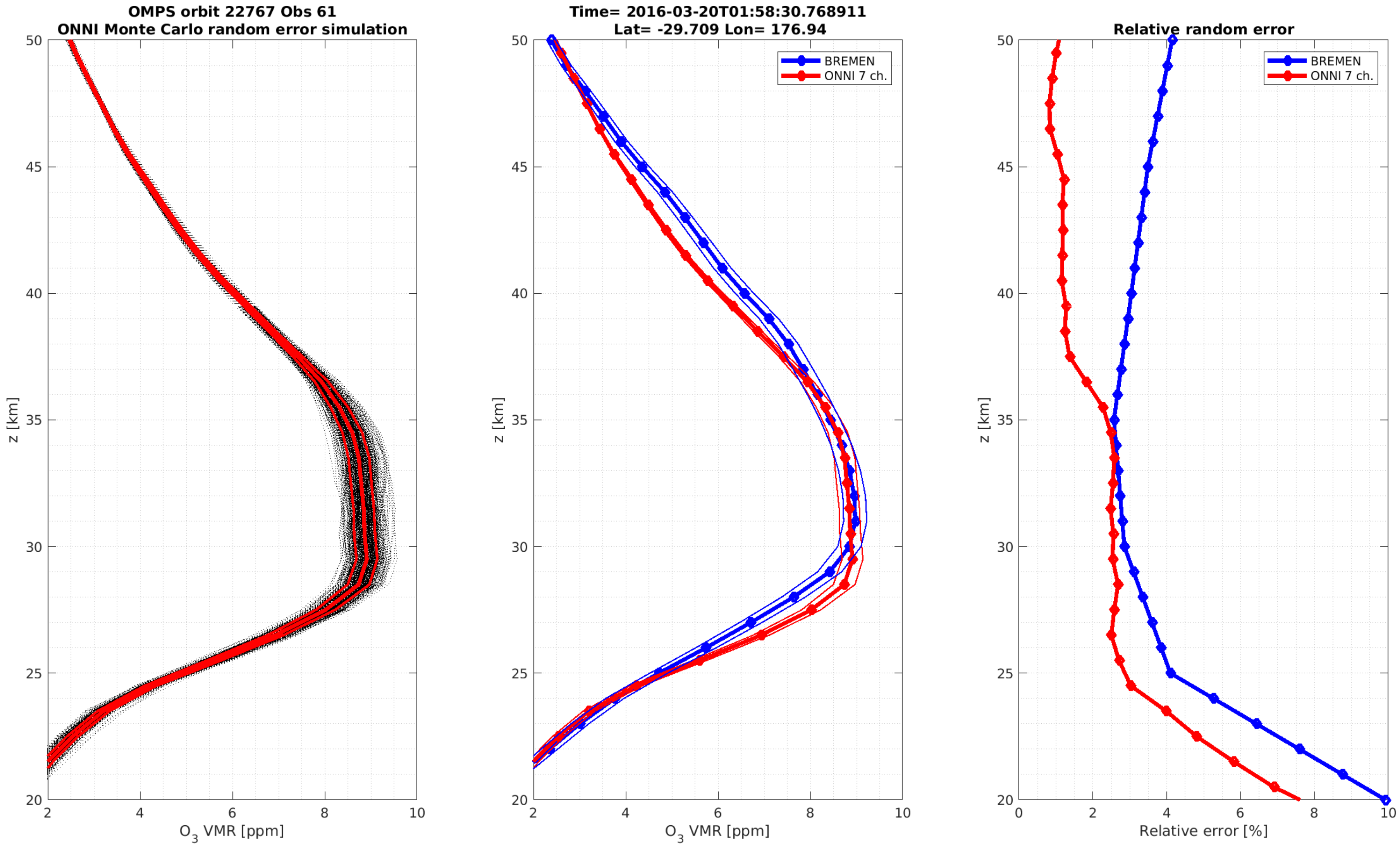


Nice but ONNI has a larger dispersion (more wavy profiles). Why ?

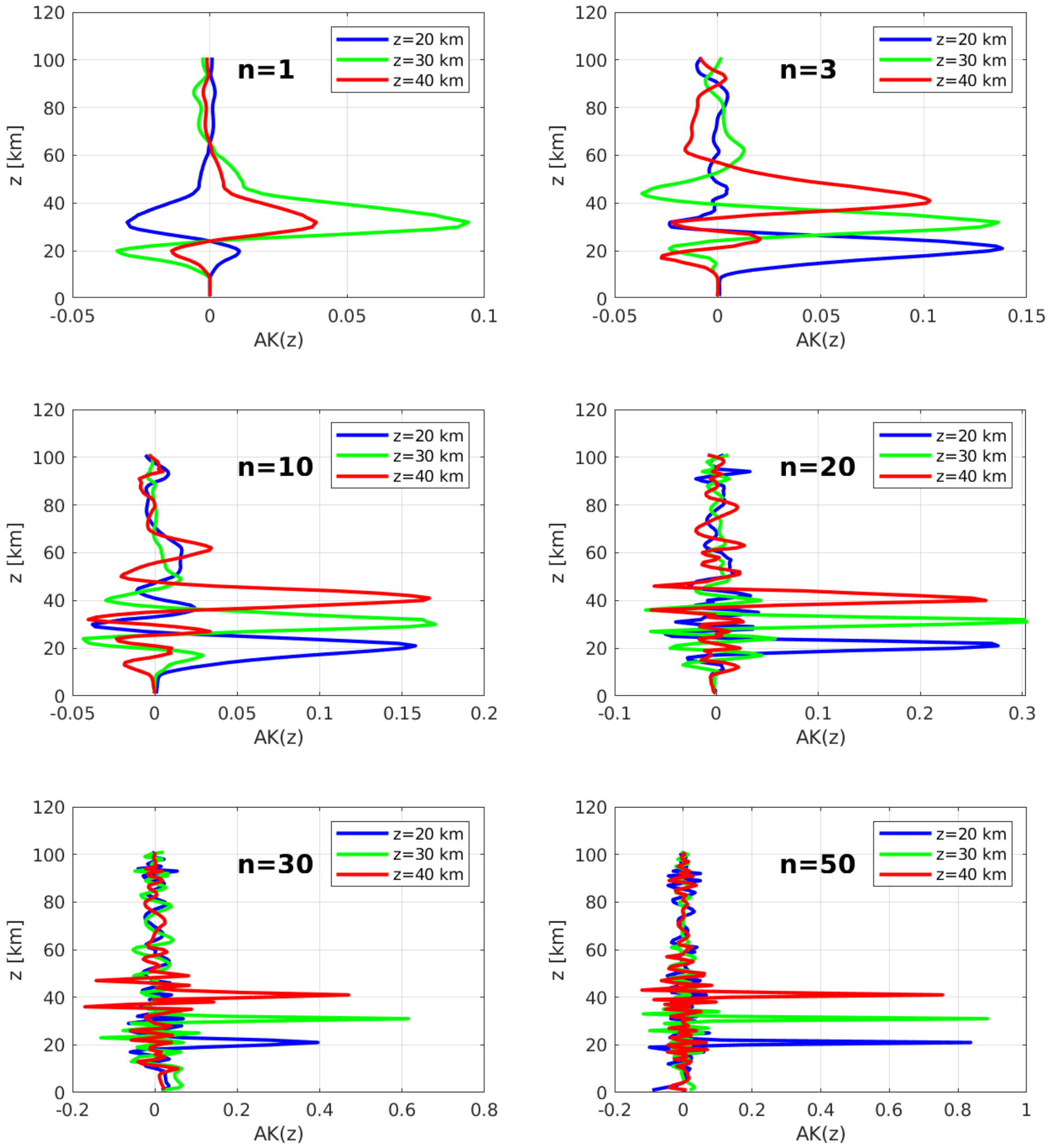
CLOUDS !

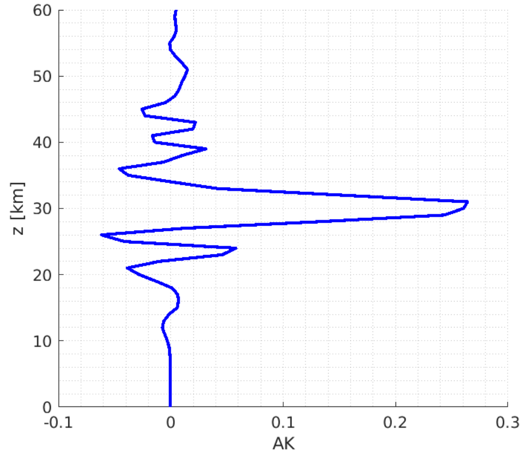


ONNI error computed with full Monte-Carlo simulations using logRad residuals



Profile representation by PCA defines the **user-defined** vertical resolution. It should match the experimental SNR level.





Summary of "Direct inversion methods" (potentially for all applications)

PRO

- main computing load moved to **synthetic LTS** generation.
- LTS allows for **PCA analysis** in measurement space and information compressing.
- orthogonality of PCA -> increase resolution up to the SNR.
- **reverse mapping** "measurement vector to state vector" is a nonlinear regression where **ANN** are considered to be superior to any method.
- trained **ANN are extremely cheap**: use brute force to derive error budgets and observation kernels
- **self-denoising:** no regularization needed

CON

- A **change in the forward** model triggers a re-generation of the LTS (except if these changes are parameterized and perturbative)
- ANN topology ? **heuristic**, no clear rules, trial and error. Number of neurons, hidden layers, activation functions. Optimal topology may depend on the solution...
- so far, nobody understands **what happens inside the box ?**

Spare slides

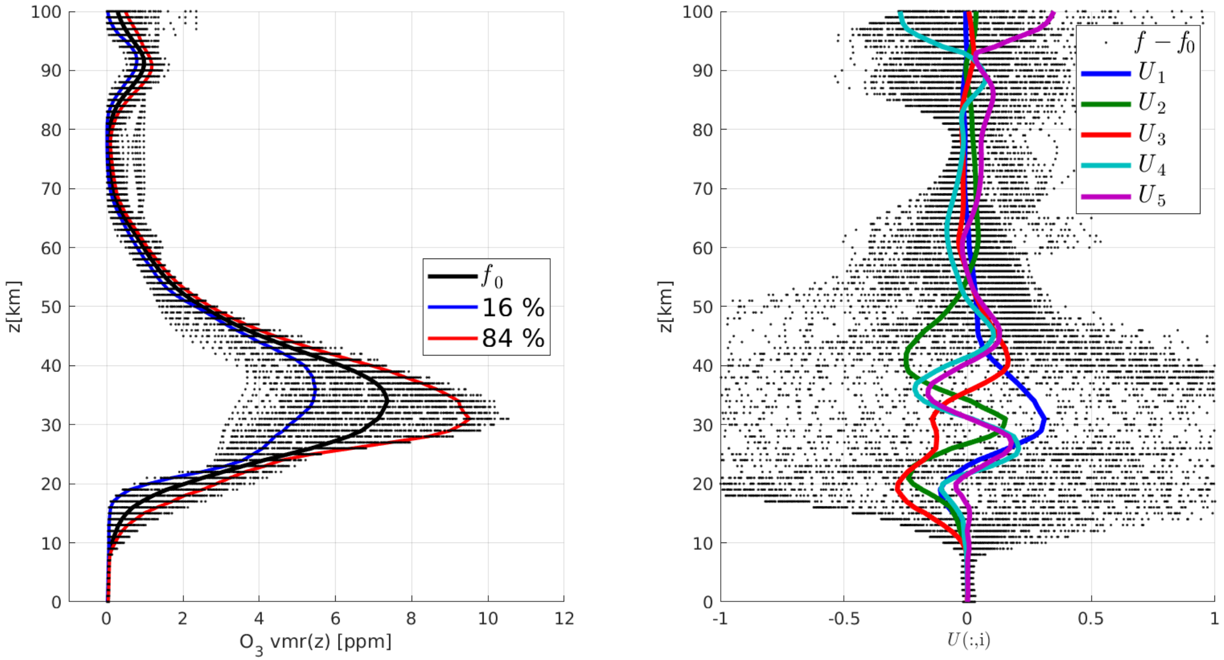
Some open questions:

- investigate hallucinations (2-3 % of cases)
- is full PCA (spatial+spectral) better to avoid spectral non-orthogonality ?
- instrumental parameters
  - direct parameterization to append to the measurement vector
  - LTS update from a subset and training from unperturbed ANN
- parallelization of L-M algorithm (Python ?)
- optimal LTS generation: which minimal training set of state vectors must be generated to represent the measurement space (GAN ? variational autoencoders ?)
- ...

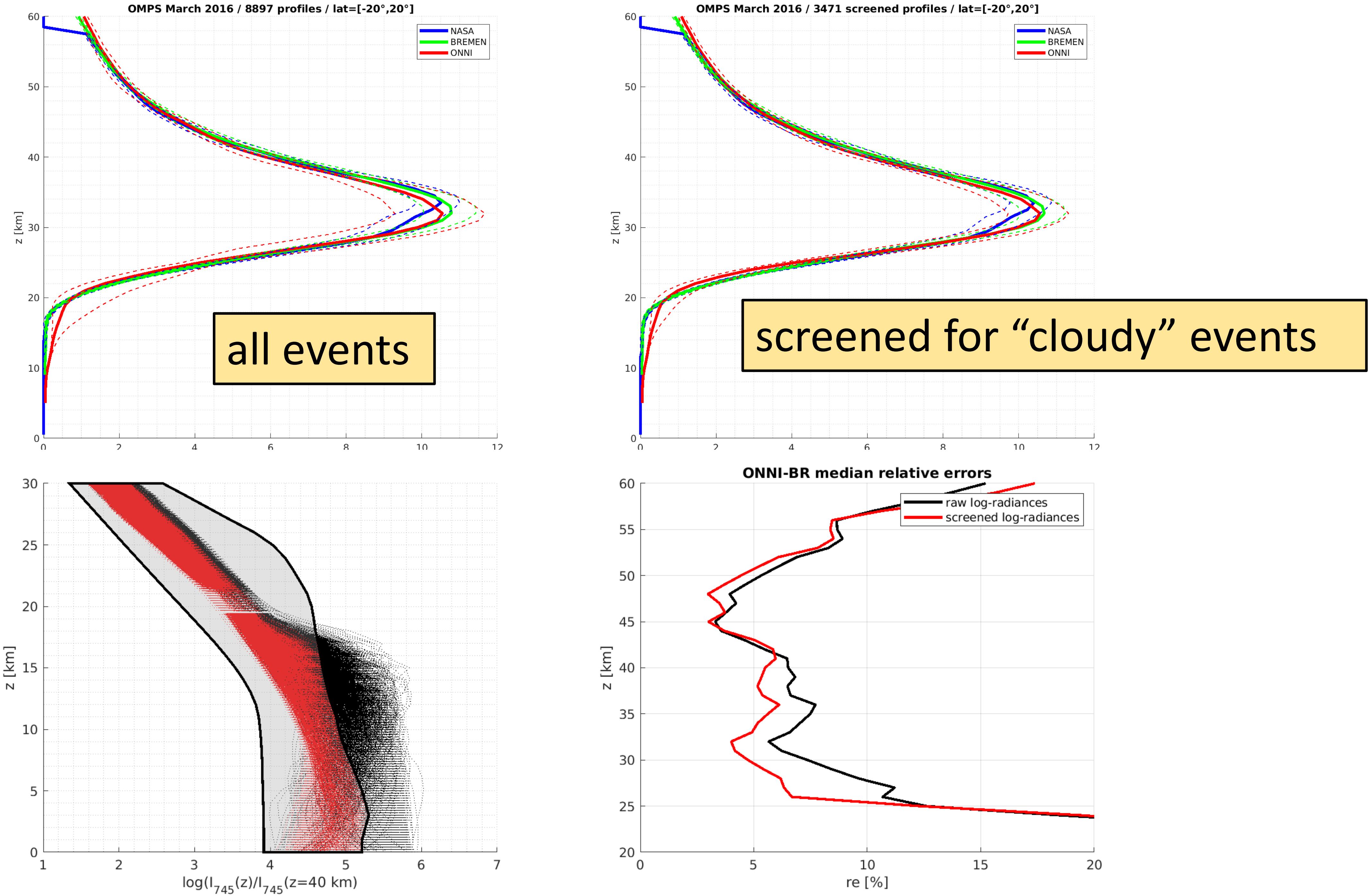
Reminder about PCA

Start from a climatology $\rightarrow$ compute the covariance matrix of the vertical profiles $\rightarrow$ SVD $\rightarrow$ use the eigenvectors as "principal components"

$$f(z) \simeq f_0(z) + \sum_{i=1}^{n} a_i \phi_i(z) \equiv \mathbf{f}_0 + \mathbf{a}(1:n) * \mathbf{U}(:,1:n)^T$$

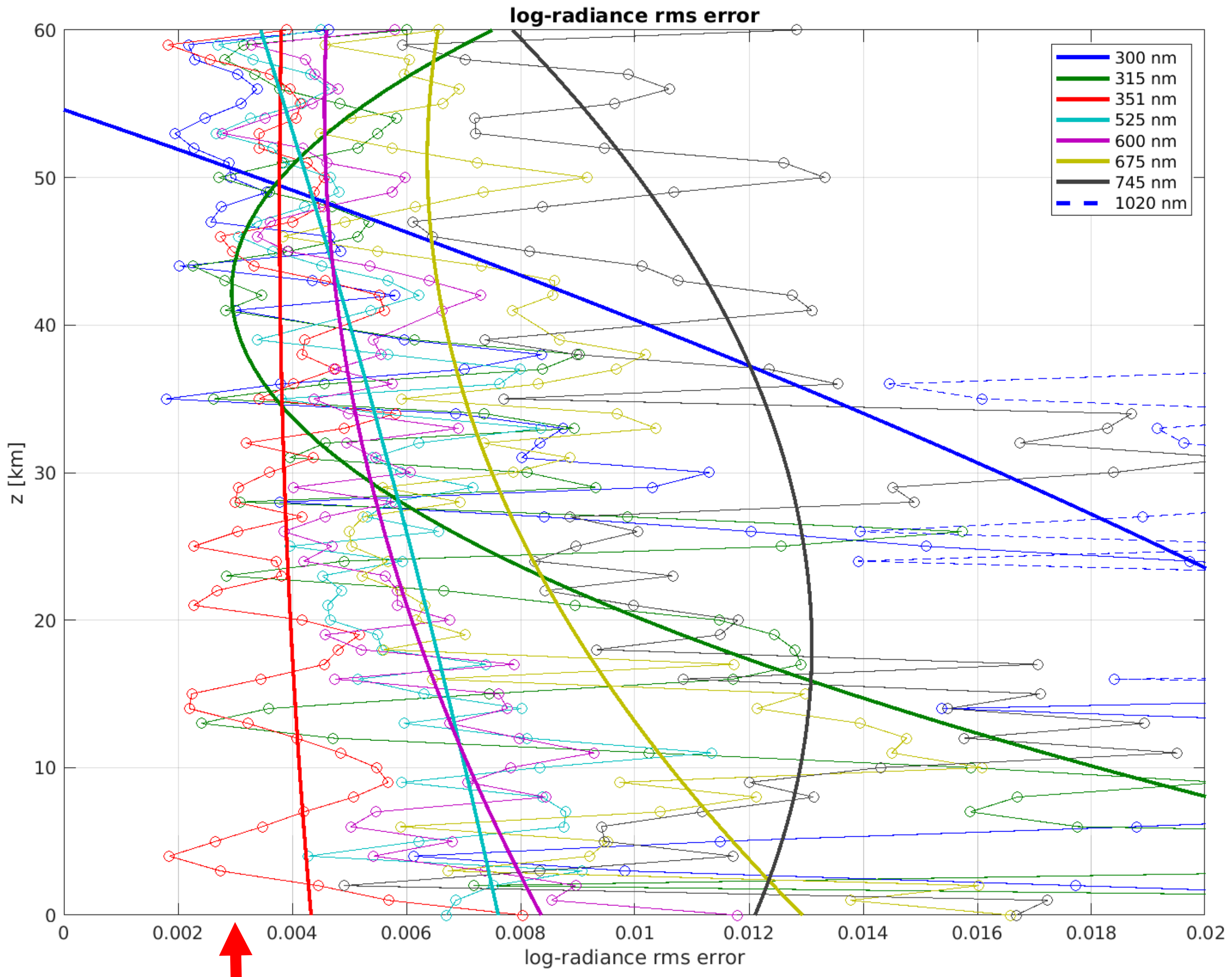


	RTM	Spectral range	normalization	regularization	Albedo, aerosols etc	CS	Algo
NASA	GSLS	1 triplet @ 600 nm [549-633 nm] for 12.5-35.5 km 3 doublets [302,312,322]/355 for 28.5-50.5 km	UV: 55.5 km VIS: 40.5 km	Covariance matrix (McP-Labow)	Albedo retrieved at 675 nm Retrieve ozone 1 km above cloud height (color method) Aerosol: independent retrieval (Loughman, [2017])	Bass &Paur	Opt.Estimation
BREMEN	SCIATRAN	285-302 nm 305-313 322-331 508-660	63.5 km 52.5 47.5 42.5 $\mathrm{Log}(y/y_*)-P_n$	1st order Tikhonov	Albedo: simul retrieval Reject TH's below threshold (color method) Aerosol: extinction at 869 nm + frozen log-norm and Mie	Serdyuchenko	Opt.Estimation
ONNI	SMART-G	300,315,351,525,600,675,745,(1020) nm	z= 40 km	PCA low-pass filter on logRad	NNI integrates the ozone signature wrt NRV	Serdyuchenko	Direct NNI



OMPS March 2016 / 8897 profiles / lat=[-20°,20°]
NASA
BREMEN
ONNI
z [km]
all events
OMPS March 2016 / 3471 screened profiles / lat=[-20°,20°]
NASA
BREMEN
ONNI
z [km]
screened for “cloudy” events
z [km]
log(I745(z)/I745(z=40 km)
ONNI-BR median relative errors
raw log-radiances
screened log-radiances
z [km]
re [%]

Stochastic error budget



theoretical limit=0,3%

Measurement noise is estimated from PCA residuals (same approach as Arosio et al. using inversion residuals)

The noise amplitude is altitude and wavelength dependent.

Summary of ONNI (specific to limb scattering)

- logRadiance **PCA representation coupled with a NN reverse mapping** is an efficient inversion method for ozone retrieval
- **self-consistency and robustness** have been verified
- **no regularization**. The PCA is a low-pass filter consistent with measurement SNR
- **error budget and AK's** are cheap to compute, even with full Monte-Carlo simulations.
- A **raw** application of ONNI to OMPS data shows **fair inter-comparison with NASA and BREMEN algos**