

Principal Component Compression of IASI data : impact on trace gases information for atmospheric chemistry

*E.Dufour¹, T.Lalanne¹, P.Prunet¹, J.Chimot¹,
C.Clerbaux², L.Clarisse³, D. Hurtmans³, P.F.Coheur³,
M.Eremenko⁴, G.Dufour⁴, J.-M. Flaud⁴
T. Hultberg⁵
S. Payan⁶, C. Camy-Peyret⁶, C. Crevoisier⁷*

1: Noveltis, 2: LATMOS, 3: ULB, 4: LISA, 5:EUMETSAT, 6:LPMAA, 7:LMD

Study supported by CNES

Context /objectives

- IASI/MetOp allows the monitoring of global scale distributions of several atmospheric trace gases of interest for atmospheric chemistry and climate.
- After the launch of MetOp 2, the data volumes and fluxes to deliver in real time will require the set up of specific compression processes. EUMETSAT envisions a dedicated dissemination and archiving scheme for IASI L1c spectra involving Principal Components Compression (PCC).
- Considering the hypothesis of a real time distribution of IASI data exclusively in compressed form through PCC, this study aims at exploring the capability of EUMETSAT PCC processor to properly capture and retrieve the information of interest for atmospheric chemistry, pollution and climate applications (e.g., capture the signal associated with emissions or pollutant plumes).

Organization and partners

This is an ongoing study initiated by a group of scientific users of the IASI data, NOVELTIS and CNES. Analysis are based on specific IASI L1C datasets, selected by the scientists and covering different applications

	L1C dataset				
	Test 1	Test 2	Test 3	Test4	...
contact	LATMOS  ULB				
	C. Clerbaux, P.F. Coheur	G. Dufour	C. Crevoisier (LMD)	S. Payan, C. Camy-Peyret	...
volume	19 ASCII files	1 BUFR 3 minutes	480 BUFR 3 minutes (1 full day of data)	4 ASCII files	...
Number of spectra	19	2760 successive	About 1 300 000 (successive)	4 (successive)	...
Applications / interest	Atmospheric chemistry Pollutant emissions and plumes	Ozone fields Regional pollution	Spatial and temporal variability of greenhouse gaz (climate studies)	IASI validation product (IASI balloon campaign)	...

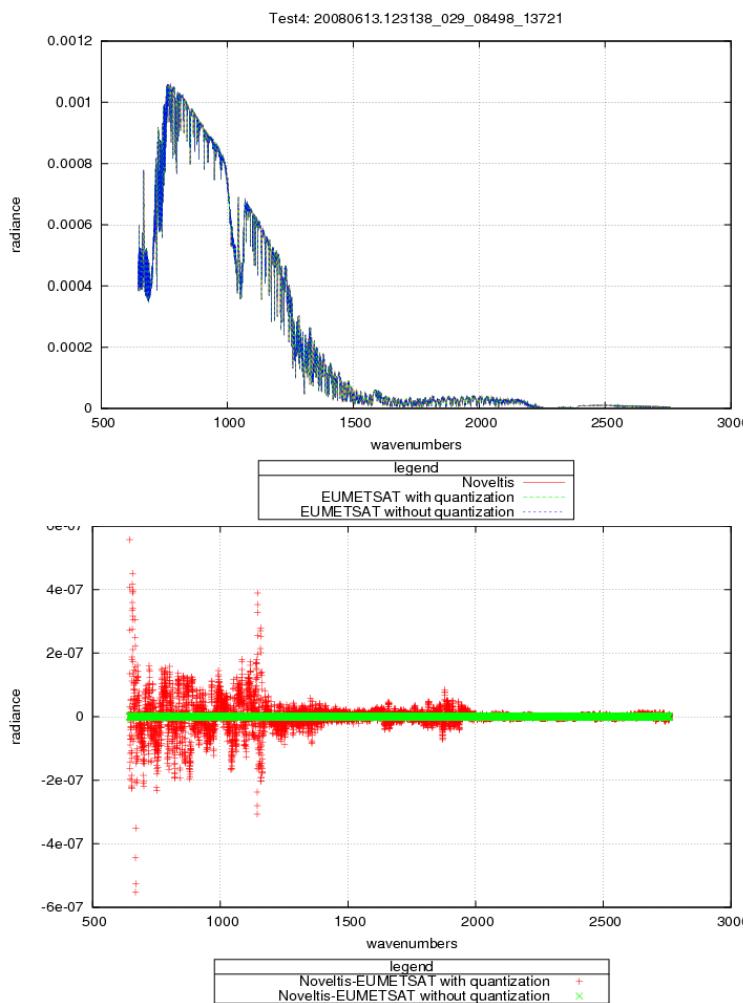
Collaborative work with EUMETSAT (Tim Hultberg) for expertise, validation and interactive evolution with respect to the PCC processor

Study objectives

- Implementation and validation of the PCC processor based on the up-to-date EUMETSAT specifications
- Processing of the identified test data
- Qualitative and quantitative signal loss analysis, at level 1 and level 2
- Identifying possible limitations (from the EUMETSAT PCC processor and/or from the users sides) and propose keys for enhancement

The IASI PCC processor

- PCC module developed at Noveltis as a copycat version of the EUMETSAT PCC processor
 - ◆ Exact reproduction of the impact of PCC as planned by EUMETSAT
 - ◆ Control of the user request
 - ◆ Coupling with level 2 information content analysis tools
 - ◆ Allow for exploring non standard parameterization of the compression processor and identifying possible enhancements
- Validated in collaboration with EUMETSAT



The IASI PCC processor

How does it work ?

- IASI PCC spectra relies on the transformation of the IASI multidimensional data from the highly-correlated spectral domain to one of much lower dimension described by a truncated set of eigenvectors of the IASI data covariance matrix.
- The reconstructed IASI radiance spectrum can be easily computed from the compressed representation

X (m channels x n spectra)	training set of IASI spectra
\bar{x} (m channels)	Mean of the training set X
N (m channels x m channels)	noise normalisation matrix
E (m channels x s components)	The s most significant eigenvectors of the covariance matrix of the noise normalized training set $N^{-1} X$

	Raw	Compression	Reconstruction
Radiance spectrum	x	$p = E^T N^{-1}(x - \bar{x})$	$\tilde{x} = N p + \bar{x}$
Noise covariance matrix	R	$E^T N^{-1} R N^{-1} E^T$	$NEE^T N^{-1} R N^{-1} E^T EN$

Reference :

- IASI PCC FAQ (Tim Hultberg, EUMETSAT)
- Tim Hultberg's poster (session 2) : IASI Principal Component Compression - First experience

The IASI PCC processor

● Main features of the EUMETSAT processor

- ◆ Applied to the apodised IASI radiance (L1C)
- ◆ The PCC is applied separately in each IASI band
- ◆ Current number of principal components in IASI band 1, 2 et 3 is respectively 90, 120, 80.
- ◆ Use of a diagonal noise normalization matrix (by noise-normalising the spectra prior to the application of the compression technique, the ability to fit the data is enhanced by avoiding giving too much weight to variance caused by noise)
- ◆ Eigenvectors based on one global training set of observed spectra composed of
 - ★ A common baseline of 74719 spectra
 - ★ A dataset of outlier spectra grown iteratively

Detection of the atmospheric signal loss in the reconstructed spectrum

- Comparison of data sets (Test 1, 2, 3, 4) with and without compression. Approach :
 - ◆ 1) Use of the noise-normalized residual rms (so called "reconstruction scores") as an indicator of a loss of atmospheric signal (Noveltis)
 - ◆ Ongoing comparison at level 1 (spectral residual) and level 2 (retrievals) from scientists
 - ★ Visual analysis of the individual spectral residual and qualitative detection of spectral structure of residual atmospheric signal emerging from the noise (LATMOS on Test 1)
 - ★ Retrieval tests with compressed data (LISA on test 2)
- Version of EUMETSAT PCC processor of August 2009 including 6664 outliers

Detection of the atmospheric signal loss in the reconstructed spectrum : reconstruction scores

Methodology : Detecting reconstruction error in the residuals

- PC compression is a lossy process in the sense that the reconstructed and original spectra are different.
- The difference can be explained both by a loss of the atmospheric signal (reconstruction error) and the part of random measurement noise removed after compression (denoising)

x_0	True radiance
ε_x	Associated measurement noise realisation
$x = x_0 + \varepsilon_x$	Measured radiance written as a sum of the true radiance and the noise
\tilde{x}_0	Reconstructed true radiance
\tilde{x}	Reconstructed measured radiance

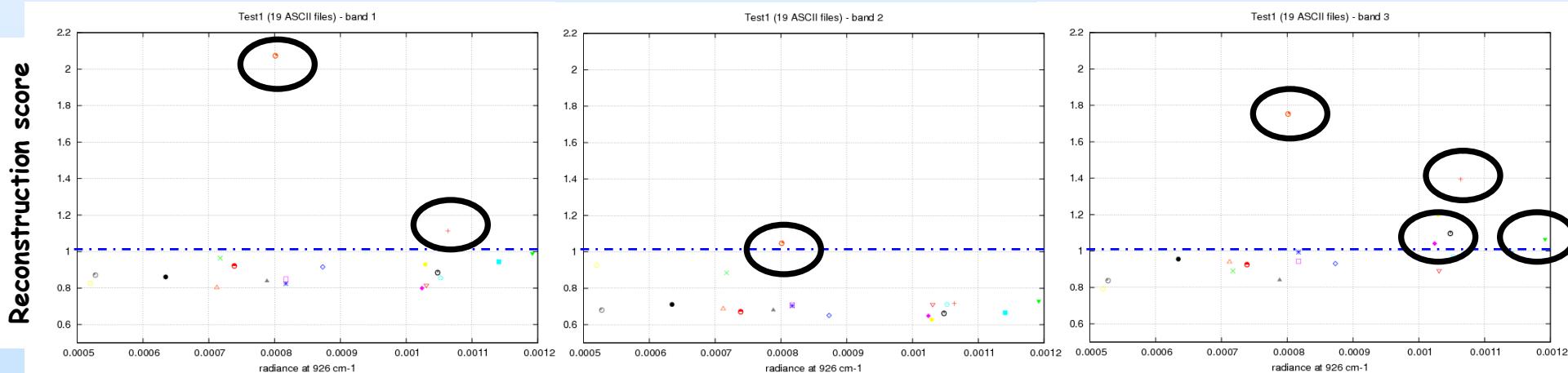
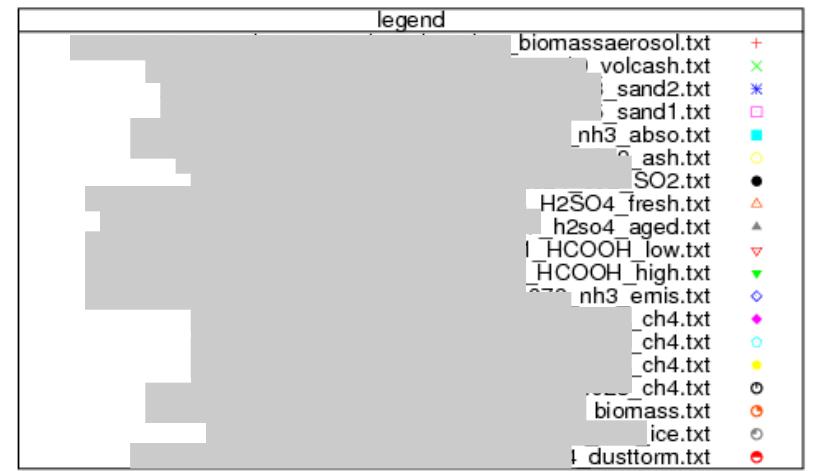
radiance
spectrum
residual

$$x - \tilde{x} = x_0 - \tilde{x}_0 + \varepsilon_x (I - NEE^T N^{-1})$$

reconstruction error part of the noise removed

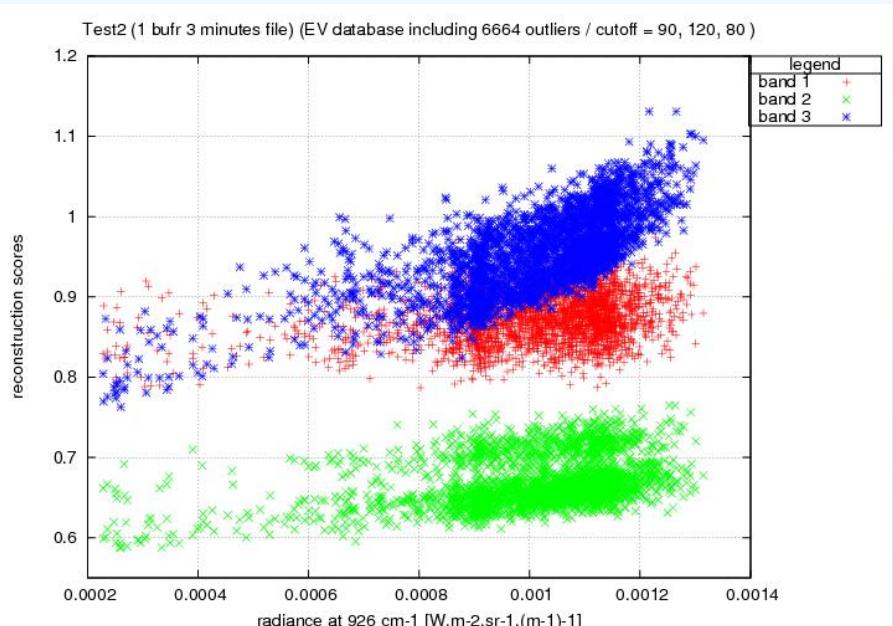
Detection of the atmospheric signal loss in the reconstructed spectrum : reconstruction score on Test 1

- If the noise figure used for normalization correspond to the actual noise in the measurements and there is no loss of atmospheric signal, the reconstruction score should be close to 1
- Values significantly higher than 1 translate a feature in the atmospheric signal that is not correctly reproduced

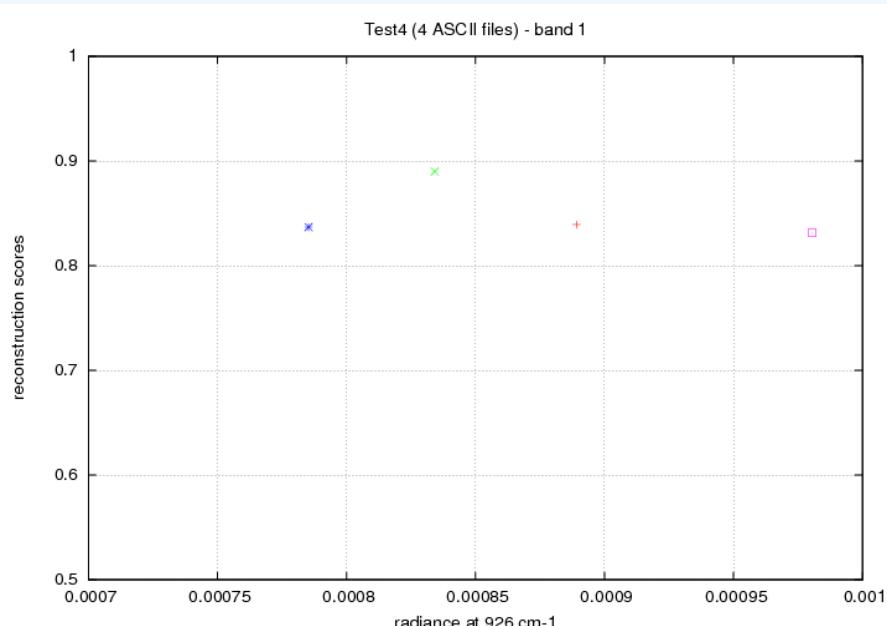


Detection of the atmospheric signal loss in the reconstructed spectrum : reconstruction score on Tests 2 and 4

- Test 2 data : ozone monitoring



- Test 4 data : IASI balloon campaign



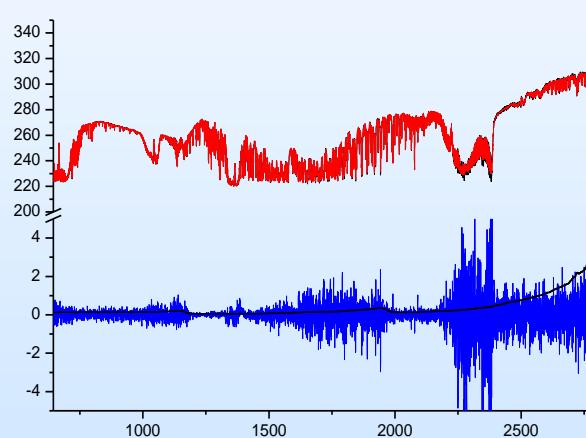
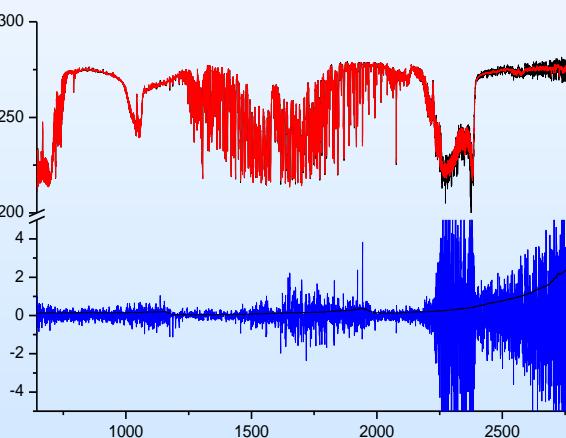
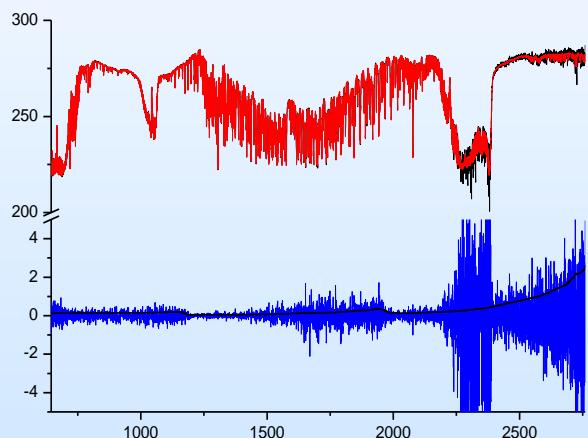
Preliminary results from scientists :

LATMOS spectra analysis at level 1 (Test 1 : selected spectra on specific pollution events)

LISA analysis at level 2 (Test 2 : ozone retrieval)

Detection of the atmospheric signal loss in the reconstructed spectrum : Test 1

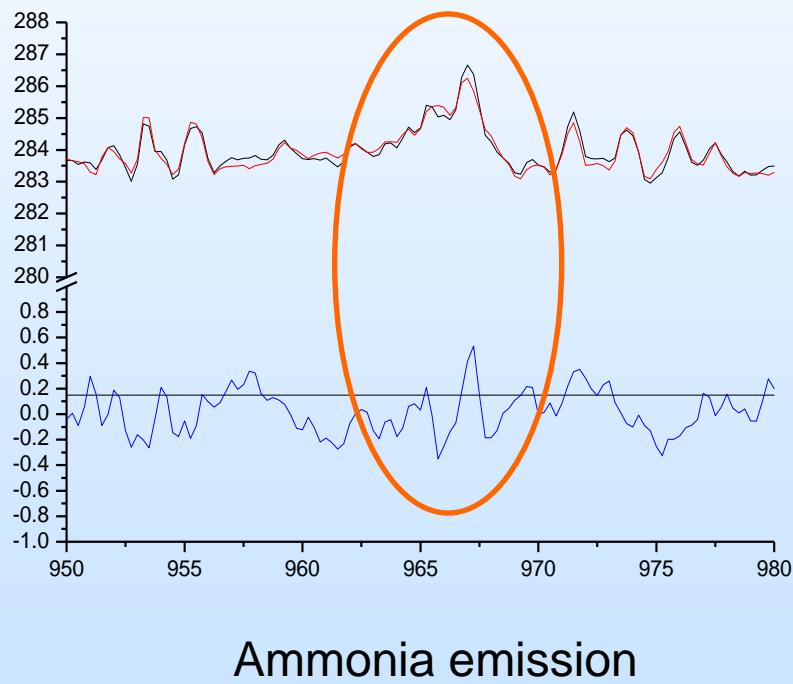
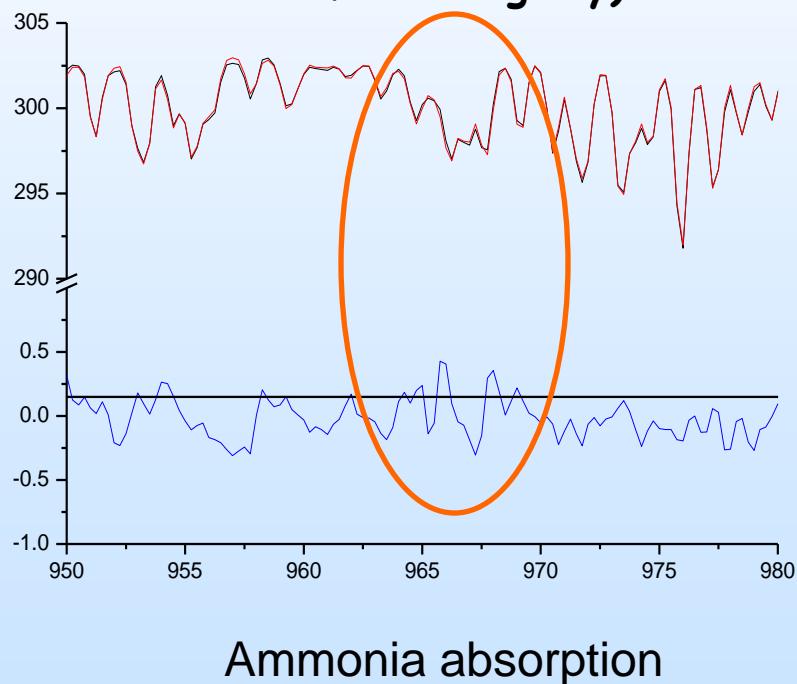
- Overall aerosol shape is captured well



Detection of the atmospheric signal loss in the reconstructed spectrum : Test 1

For large ammonia loadings

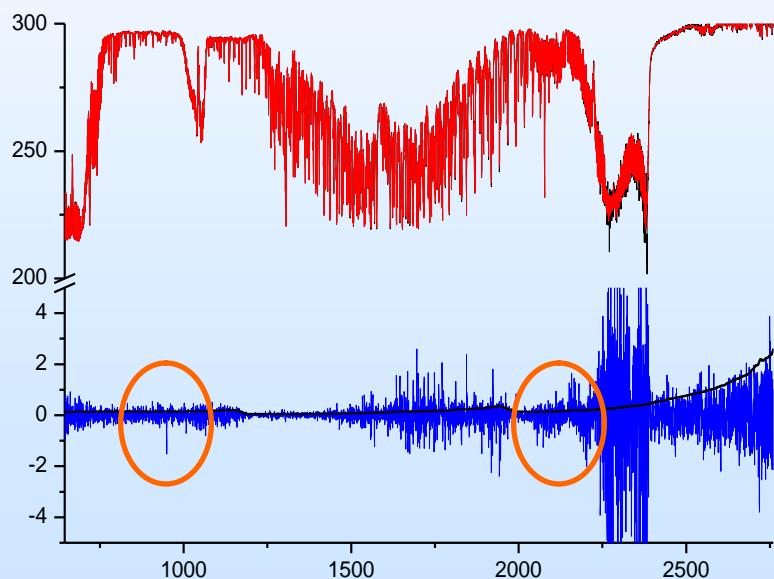
- ◆ Overall pretty good reconstruction of the signal
- ◆ Yet, the ammonia signal is not totally captured (loadings should be underestimated slightly)



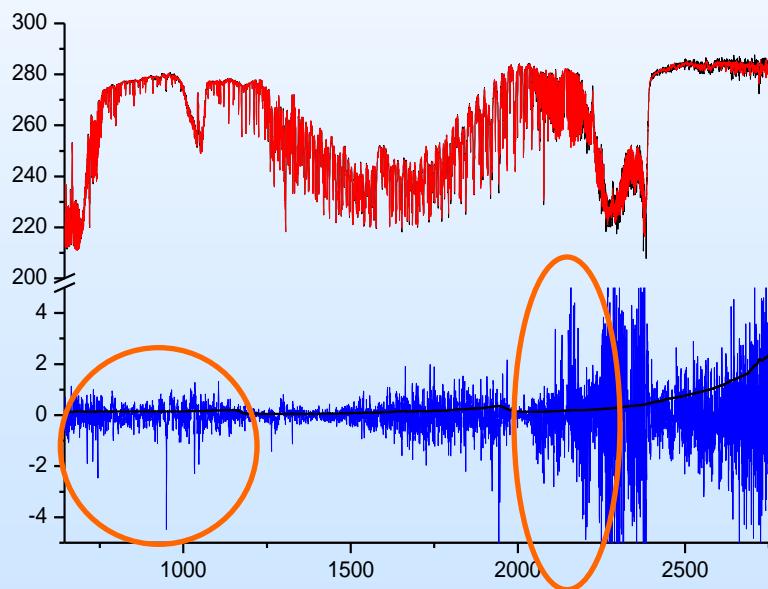
Detection of the atmospheric signal loss in the reconstructed spectrum : Test 1

For biomass burning events:

- ◆ More rare trace gases are not captured at all
- ◆ CO is poorly reproduced and should lead to large errors in the retrieval



Biomass spectrum 1

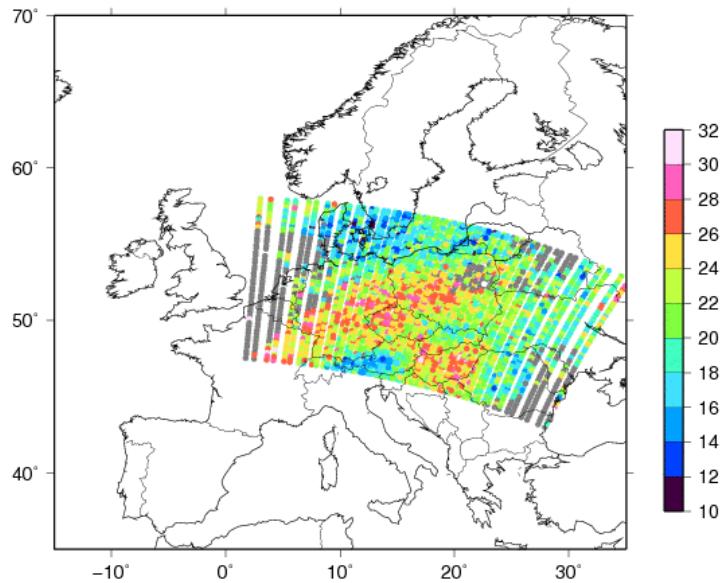


Biomass spectrum 2

Detection of the atmospheric signal loss in the reconstructed spectrum : Test 2

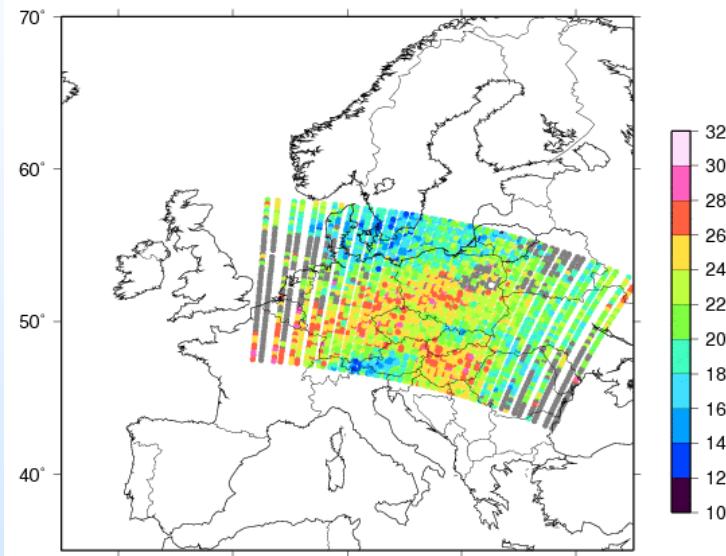
Retrieved from standard spectra

Retrieved from Standard spectra



Retrieved from PC-compressed spectra

Retrieved from PC-compressed spectra



IASI 20070715_AM
lower tropospheric ozone columns (0-6km)

Similar ozone structures in the two retrieved fields

NOV-3788-SL-8767

© Noveltis 2010

Ce document est la propriété de Noveltis et ne peut être reproduit ou communiqué sans son autorisation.

Preliminary conclusions

Work is ongoing, all points below shall be consolidated / discussed

- First elements of analysis suggest that :
 - ◆ PCC processor capture very well the signal of interest in many cases (good reconstruction scores, examples of aerosols, ozone, ammonia)
 - ◆ With the PCC version tested, in some cases, reconstructed spectra present missing information (underestimation of ammonia ?) or fail to capture signal (biomass burning)

Preliminary conclusions

Work is ongoing, all points below shall be consolidated / discussed

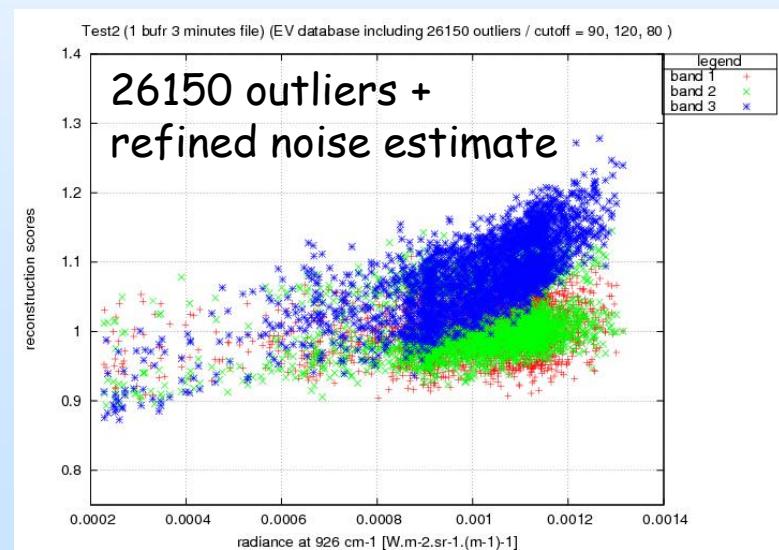
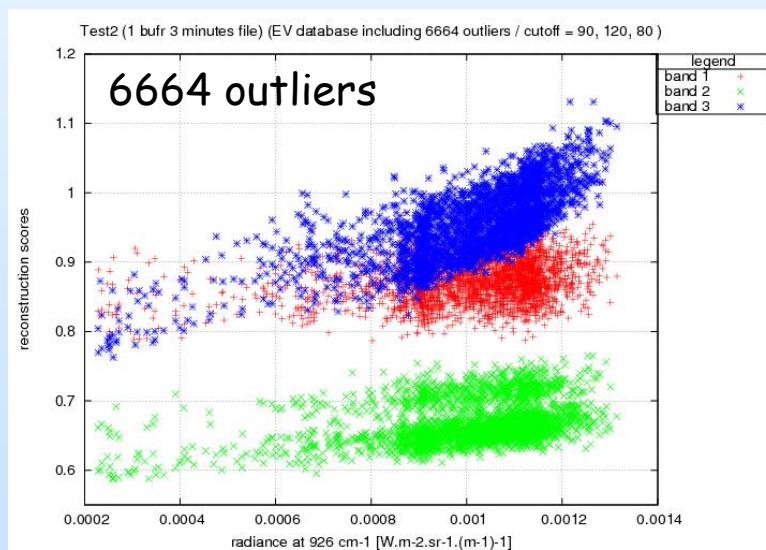
● Possible limitations :

- ◆ PCC processor : limited representativity of the training database (most probable reason, tests are ongoing), non optimal level of truncature (probably not, because of large work already done by EUMETSAT on this point, additional tests are ongoing in this study)
- ◆ Proper use of the compressed data by the users : availability of proper measurement noise, capability to deal with measurement error correlations, impact of error correlations on the retrievals

EUMETSAT approach : iterative evolution of the PCC performance

Context of rapid evolution of the EUMETSAT PCC processor

- ◆ 3 versions of the processor delivered by EUMETSAT since the beginning of the study in august 2009 due to
 - ★ Progress in the characterization of the normalization noise
 - ★ update of the training dataset (through the addition of so-called outlier spectra having residual rms in one of the 3 bands above a certain threshold value)
- ◆ Reconstructed spectra are regularly reprocessed at Noveltis to follow the PCC processor updates



Preliminary conclusions

Work is ongoing, all points below shall be consolidated / discussed

- Iterative EUMETSAT approach for improvement of the PCC processor, and great interactivity and reactivity of EUMETSAT in this study.
 - ◆ The training database is continuously improved, with consideration of the users needs
 - ◆ All outlier spectra identified by the users or detected by PCC monitoring are integrated to the database
 - ◆ PCC capability of reducing noise shall improve the use of data if users are able to properly deal with noise

But what about the IASI information elements present in the data but not identified at the moment (future events, climate trend, non-studied species ..) ? Is the risk to remove them from the data before their identification ?