Hybrid PCs

"Action M4 A.9: To give information on consolidated strategy about how to capture rare and special event by next MAG (It could be by providing additional local PC or residuals) (Tim Hultberg)"

- ✓ Last IRS MAG meeting:
- The basics of PC compression
- Global PCs shown to yield better separation of signal and noise than local
- Hybrid PC approach to retain new features was introduced and illustrated with a single case

- ➢ Today:
- Further case studies with the hybrid PC approach
- Isolated outliers not well captured (exemplified with the Russian fires 2010 case)
- An attempt to solve this
- Trends in the mean residuals (for some channels)
- Recommendation to use 5 local PCs per band

Global or Local? ➔ Hybrid!
## A reminder slide…

PC compression with static eigenvectors based on a big <u>global</u> set of past observations <u>works excellent</u> and is <u>preferred by some Users</u>.
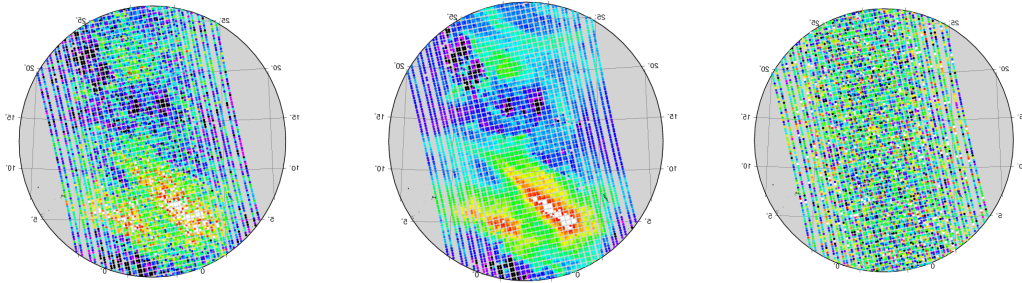

Only in <u>very rare situations</u> new spectral features orthogonal to the previously observed directions occur, which can not be represented well, <u>but are flagged</u>.
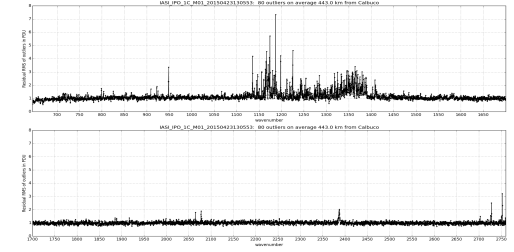
Eigenvectors computed for the <u>local</u> set of current observations could solve this issue, but would <u>retain more noise</u> and <u>less atmospheric signal</u>.

$$P_G(I-A_L)C_L(I-A_L)^T P_G^T$$
$$P_L(I-A_G)C_L(I-A_G)^T P_L^T$$
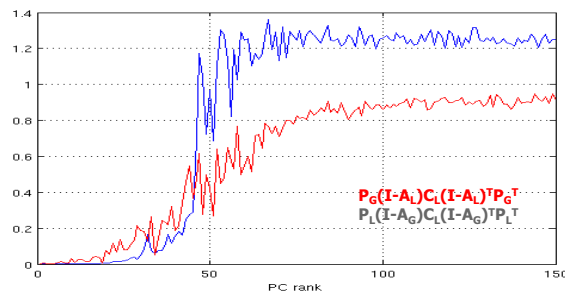
Instead we can <u>supplement the global</u> eigenvectors with a <u>few local</u> eigenvectors, to represent <u>new signal</u>.
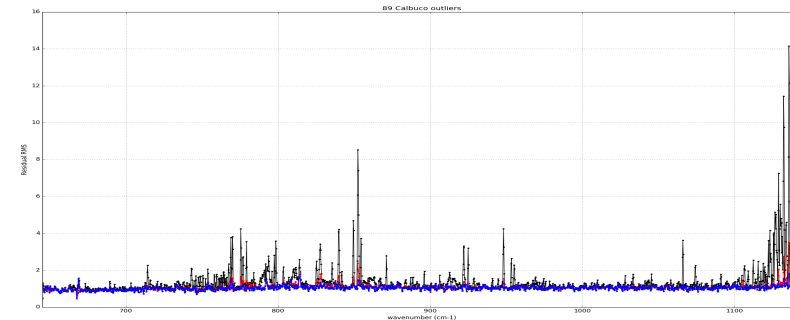
The hybrid strategy

Disseminate s_L local PCs and corresponding PC scores based on the residuals, after application of global PCs

For each individual granule:

Compute the s_L leading eigenvectors of $\frac{RR^T}{n}$ where R is the matrix of the n the noise normalized residuals using global PCs.

Disseminate these s_L local PCs and their corresponding scores in addition to the usual scores of the global PCs.

Including the local PC scores in the reconstruction of radiances makes them closer to the raw radiances but only very seldom adds information.

➔ It is fine to ignore the local PCs/PC scores: the atmospheric information is well captured in the global EVs
➔ It is fine to always include the local PCs/PC scores in the reconstruction: small increase of noise
➔ Their purpose is to avoid the very rare situations of atmospheric loss due to previously unobserved features

In the hybrid strategy we use the same fixed number s_L of local PCs for all spectra.
The alternative solution of dissemination of raw spectra for any outlier would correspond to $s_L = m - s_G$ for outliers and $s_L = 0$ for non-outliers.

Experiments using IASI data

Not using operational IASI PCs because
- Outliers from special events (volcanic eruptions and fires) have already been added
- Diagonal noise normalization - full matrix noise normalization is baseline for MTG-IRS

Eigenvectors based on 12 full days (2nd in each month) in 2016 of data from both satellites.

Applied on 4 case studies:
- ✓ Kasatoshi: 20080811 (4 days after begin of eruption)
- ✓ Calbuco: 20150424 (2 days after eruption)
- ✓ Russian wildfires: 20100801 – 20100809 (The 2010 Russian wildfires were several hundred wildfires that broke out across Russia, primarily in the west in summer 2010. They started burning in late July and lasted until early September 2010)
- ✓ A "normal" day: 20180515
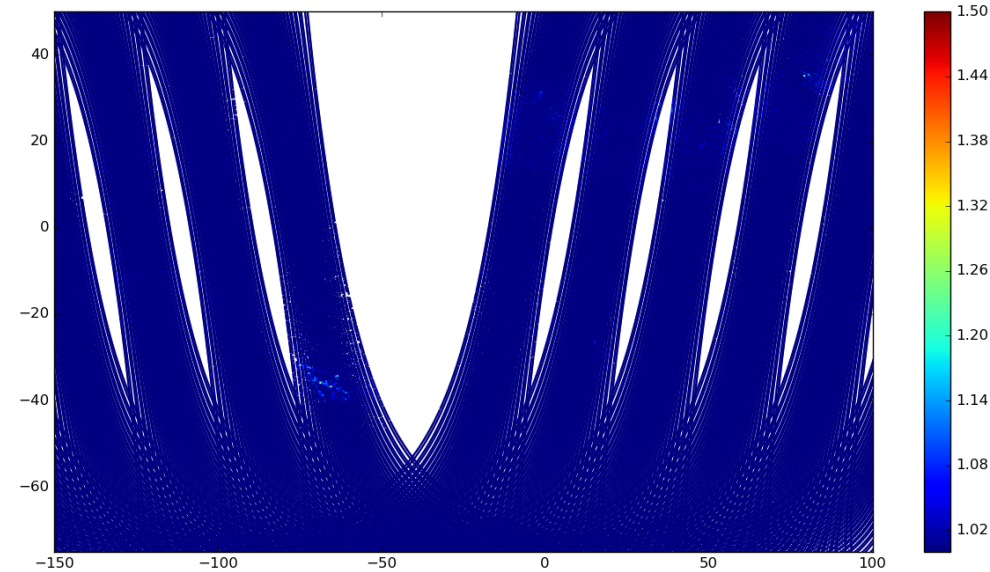
Kasatoshi   -   Band 2 reconstruction score 20080811 PM

0 local PCs
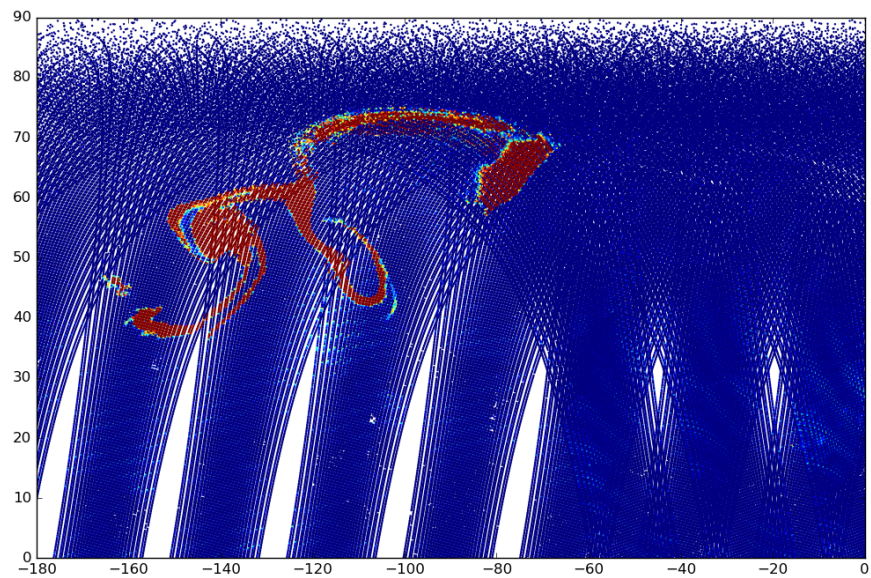
1 local PCs
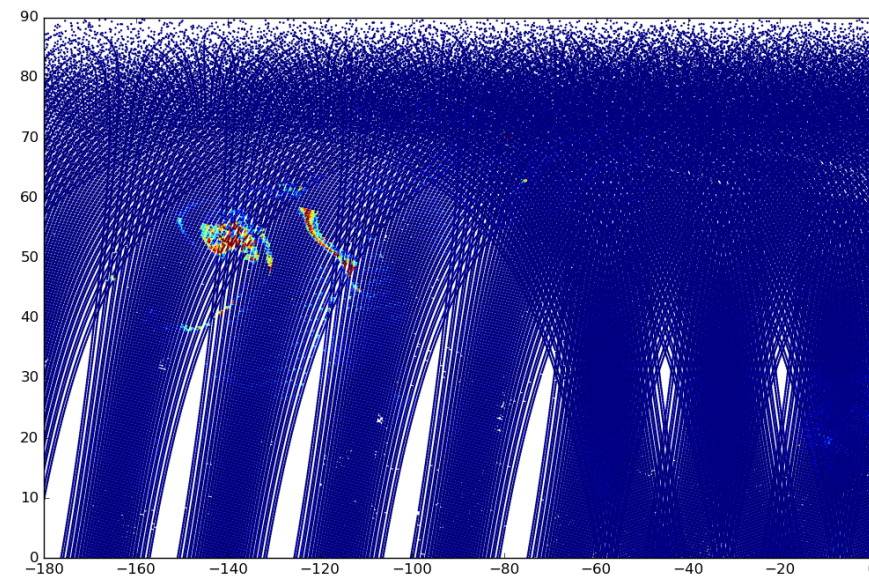
3 local PCs
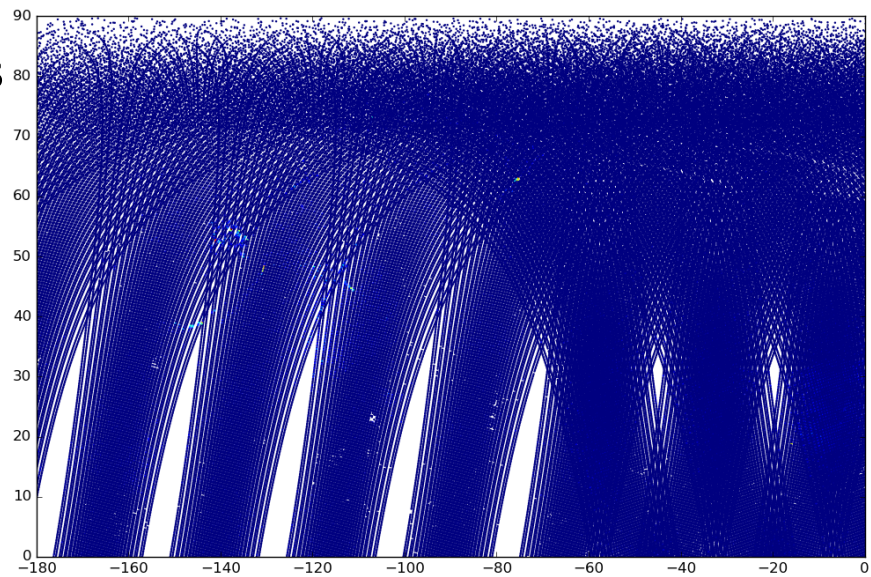
5 local PCs
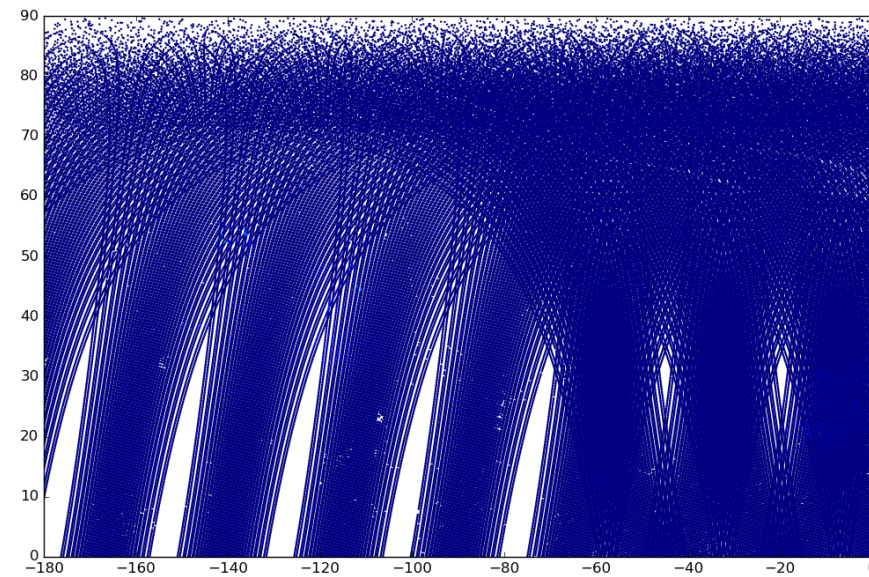
Calbuco - Band 2 reconstruction score 20150424 AM (Metop-A)
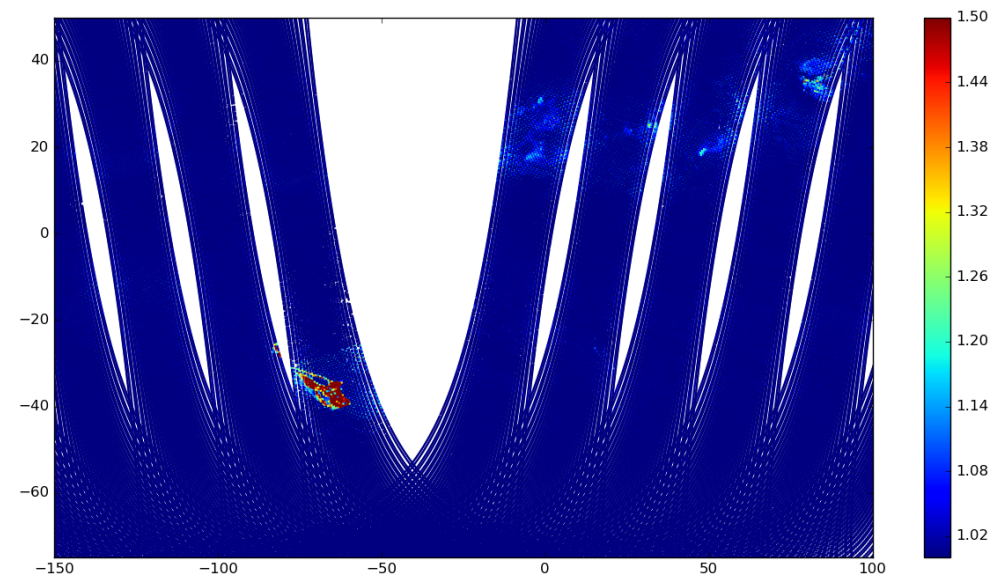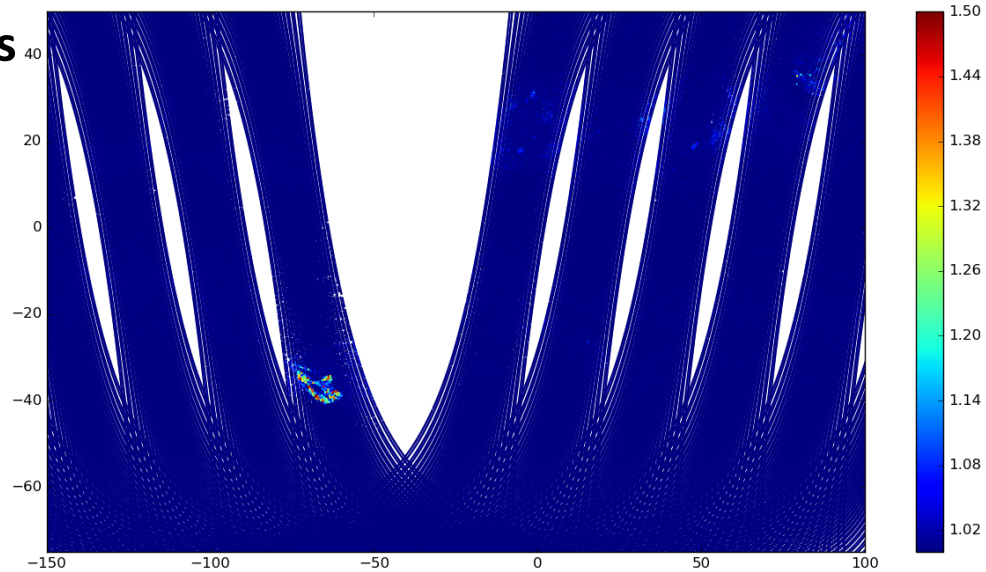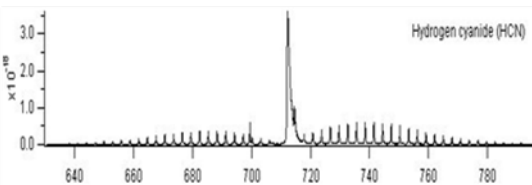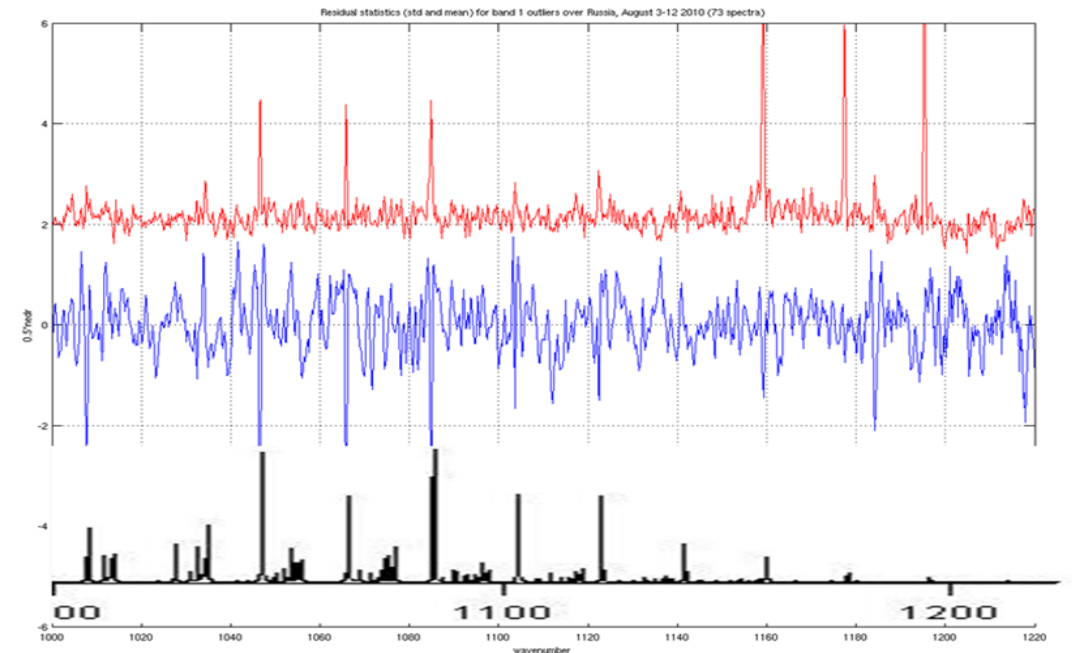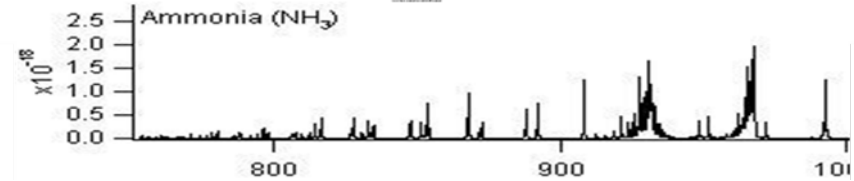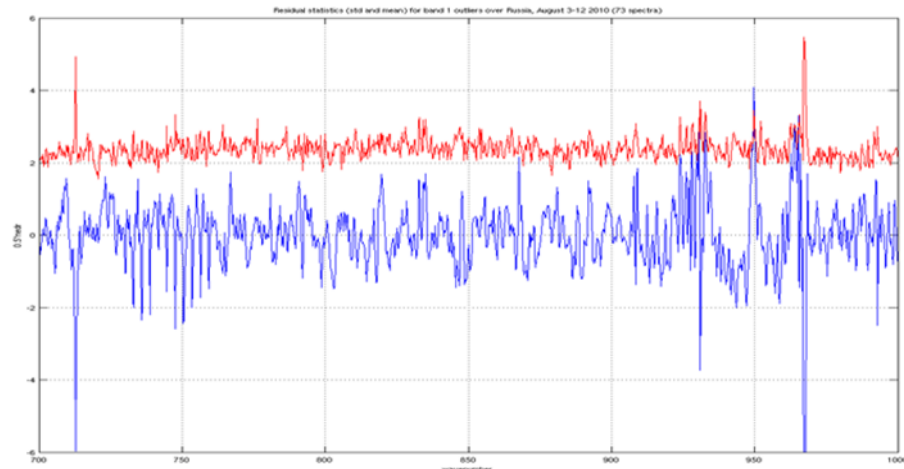
Russian wildfires summer 2010

Band 1 reconstruction score in a single PDU
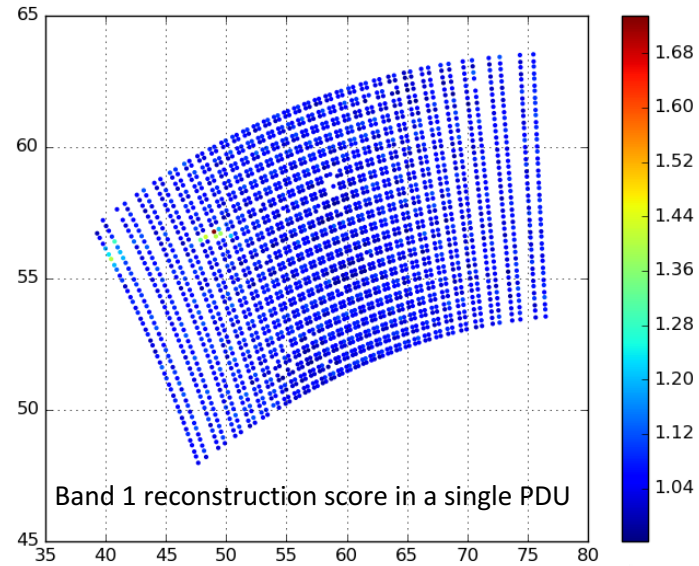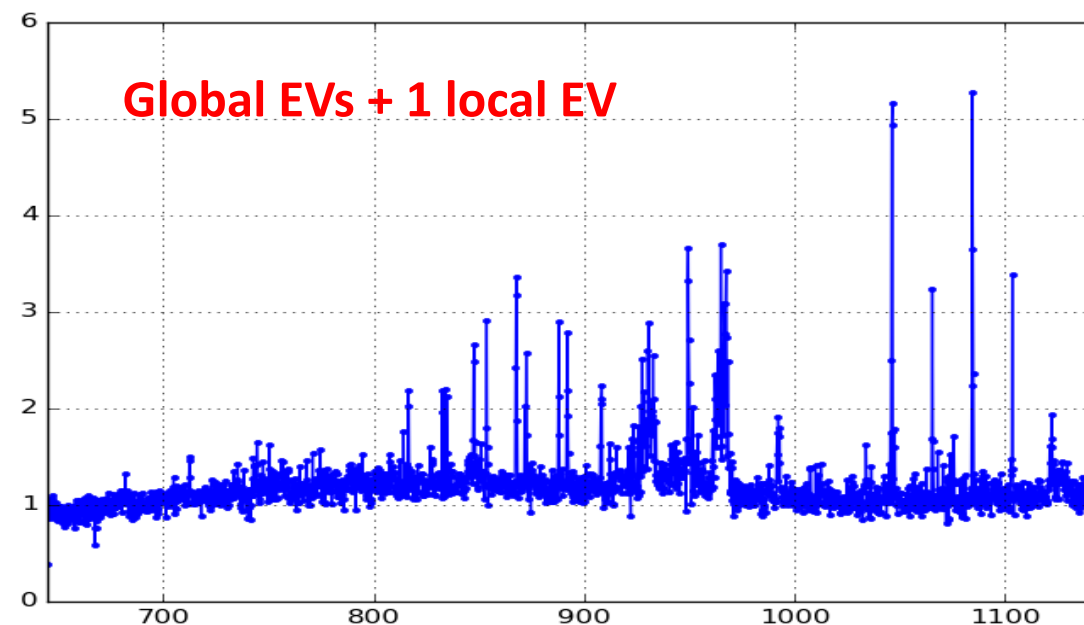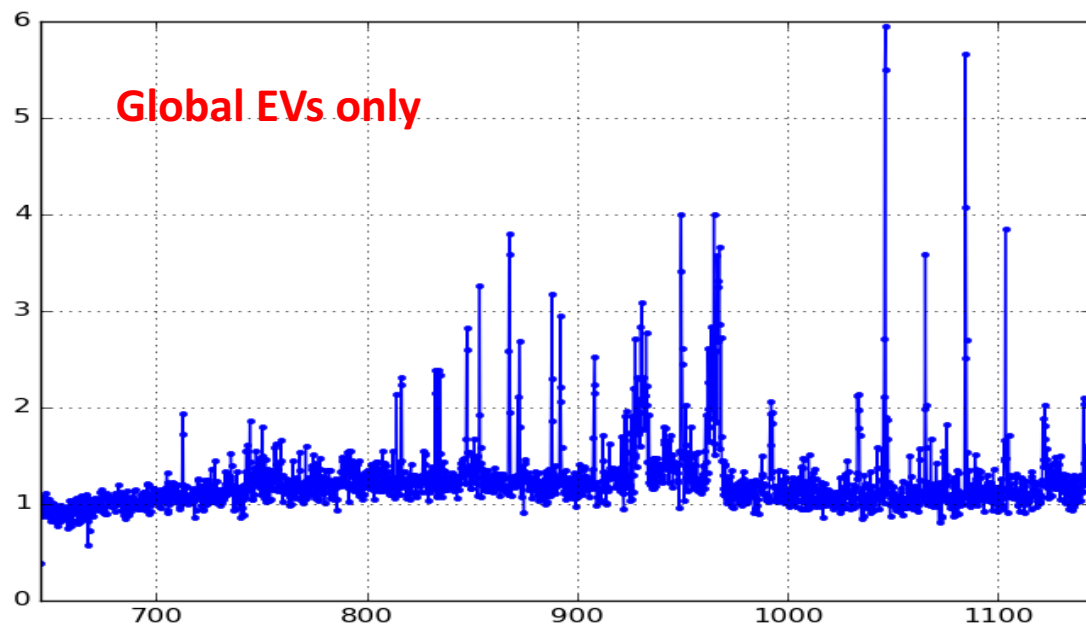
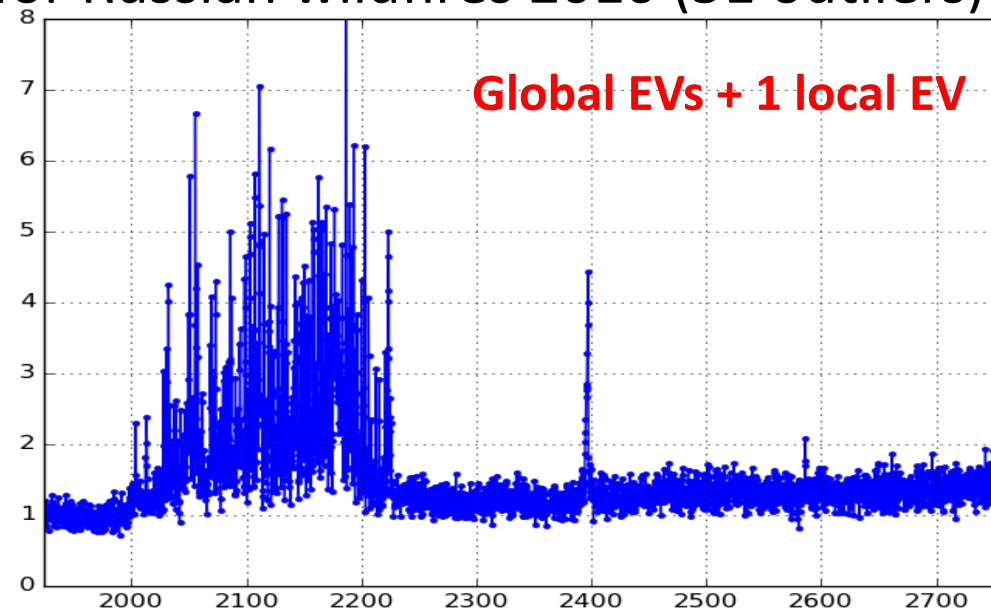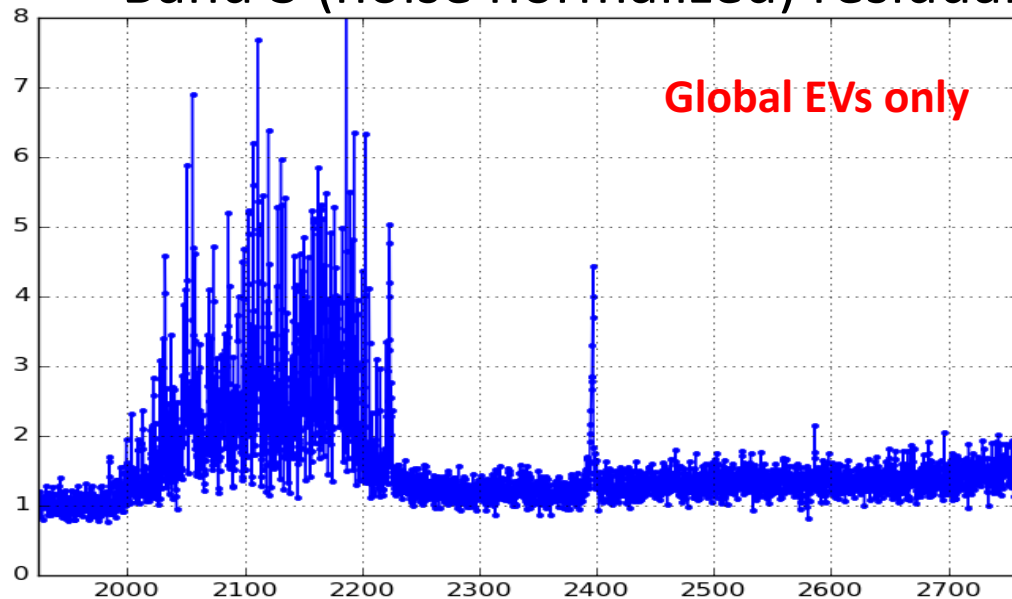73 Russian fire outlier spectra collected from 10 days were added to the training set for the operational IASI PCs

Hydrogen cyanide (HCN)

Ammonia (NH₃)

Band 1 (noise normalized) residual RMS for 96 Russian wildfires 2010 outliers

Band 3 (noise normalized) residual RMS for Russian wildfires 2010 (51 outliers)

Global EVs only

Global EVs + 1 local EV
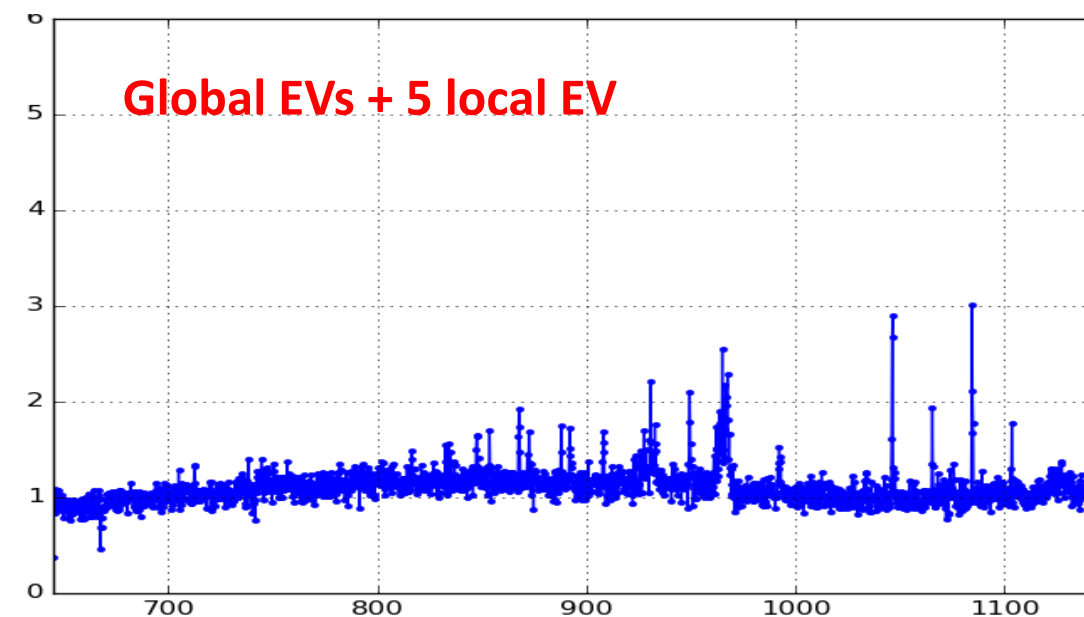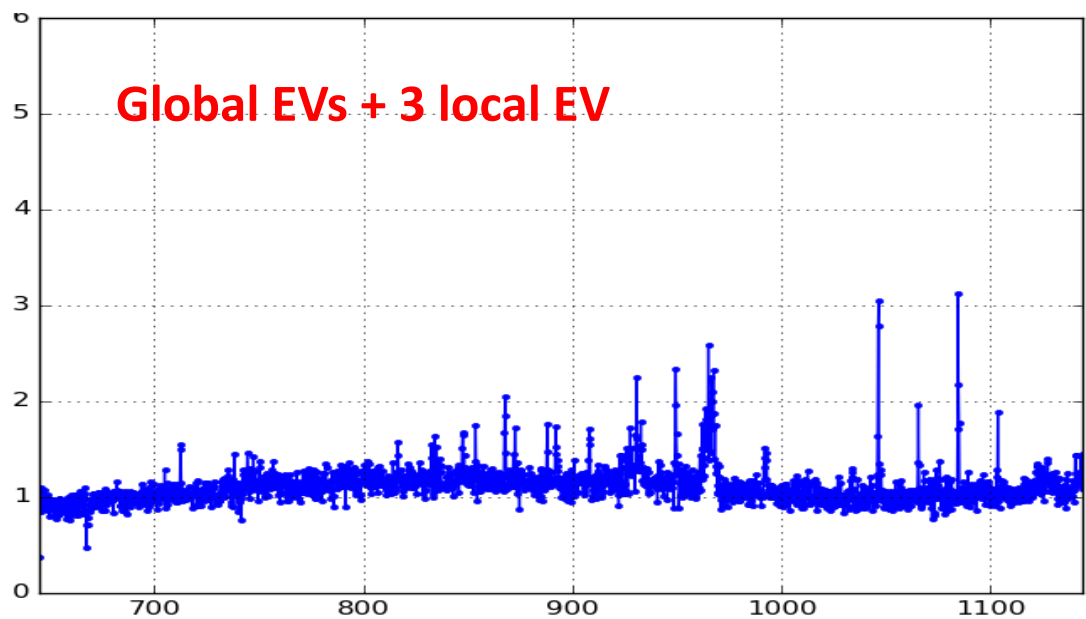
Limitation: The hybrid approach is not fully solving the problem of adequately representing outliers for dissemination when there is only very few (or just one) of them in a granule.

Global EVs + 3 local EV

Global EVs + 5 local EV
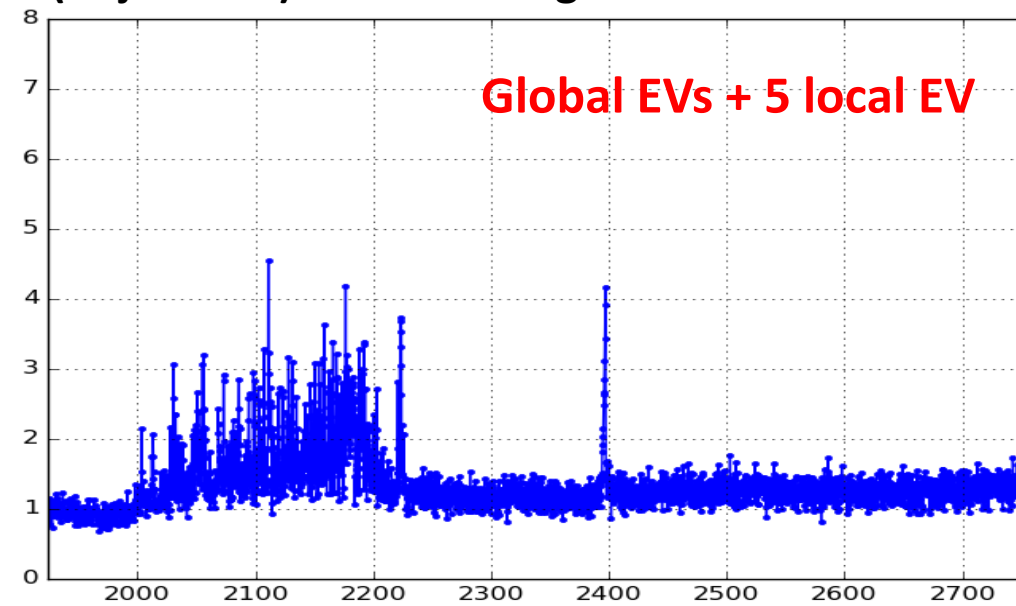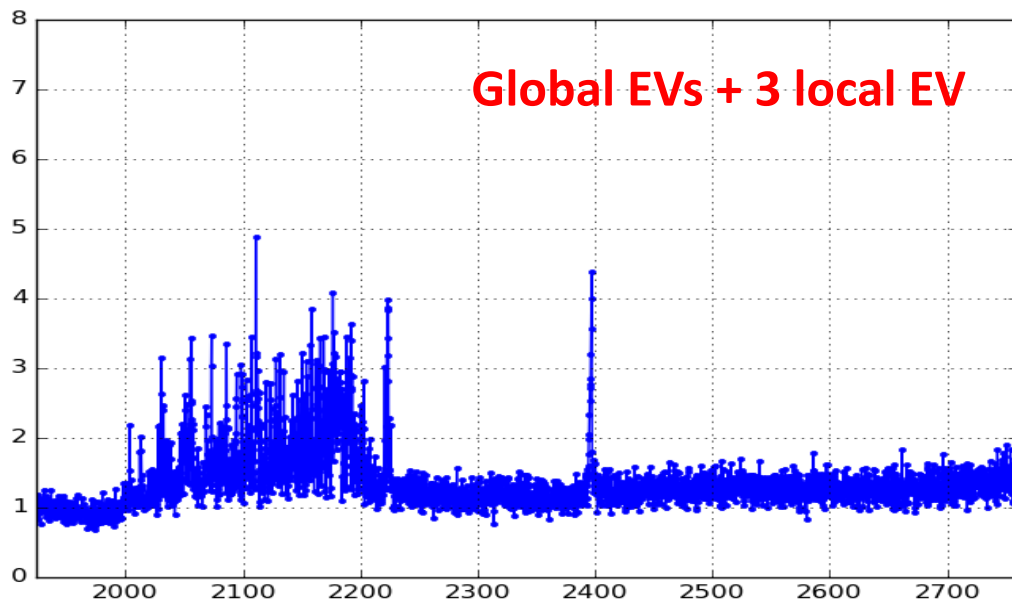
Matrix of (noise normalized) residuals (one PDU IASI B1)
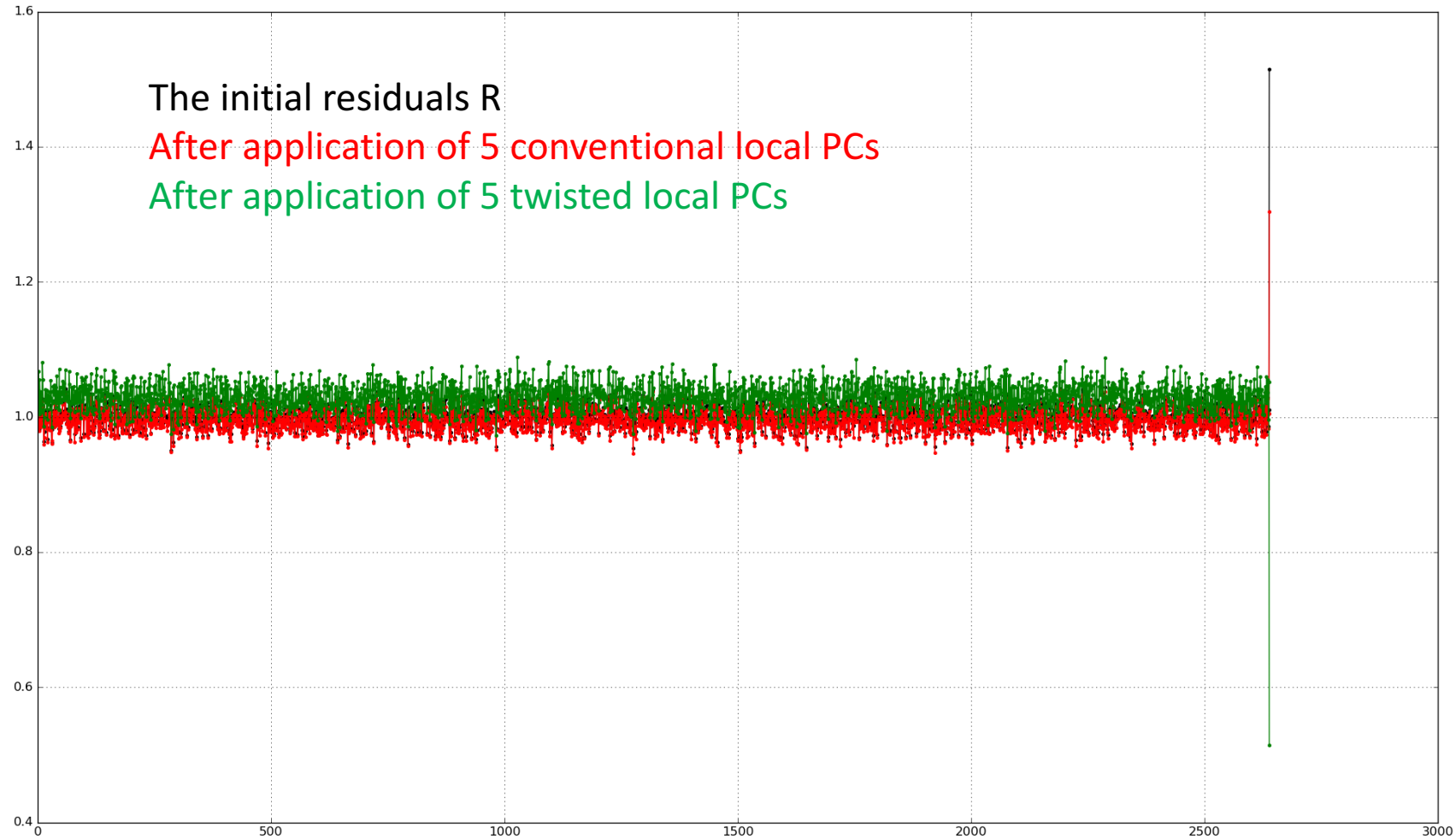
R
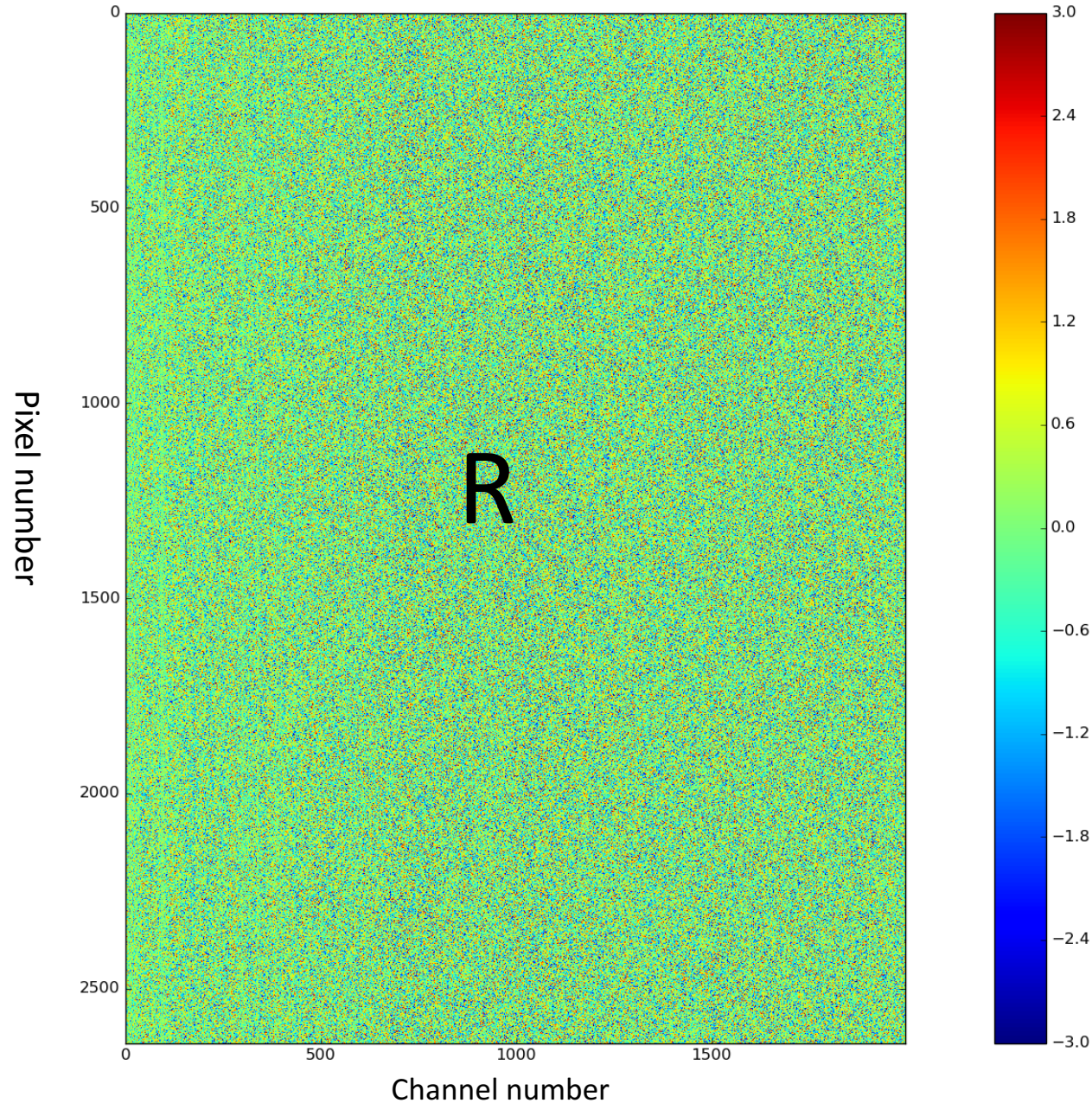
Local eigenvectors:

Low rank approximation of R

A similar (twisted) idea:

Low rank approximation of sign(R) * (R*R)

➔ more weight to high absolute values

To check that it works as intended, lets look at a **synthetic case** with m=1997, n=2640. We construct R with random elements (Gaussian,0,1) but add 5 to the first hundred channels of the last spectrum. (R[0:100, 2639] += 5)

Reconstruction score for each of the 2640 synthetic residuals

The initial residuals R
After application of 5 conventional local PCs
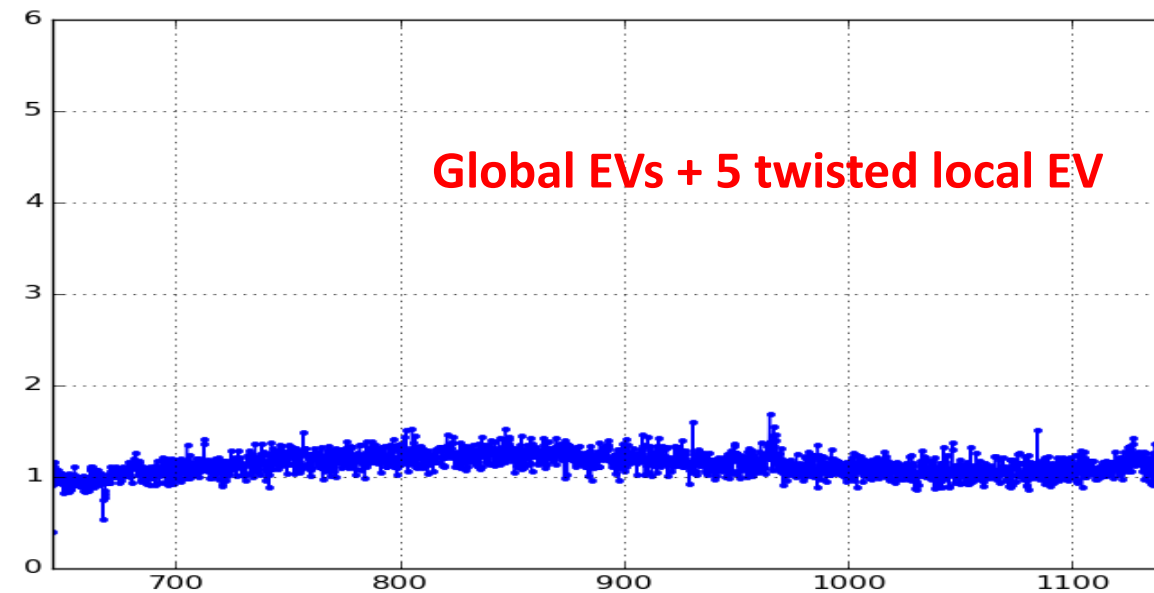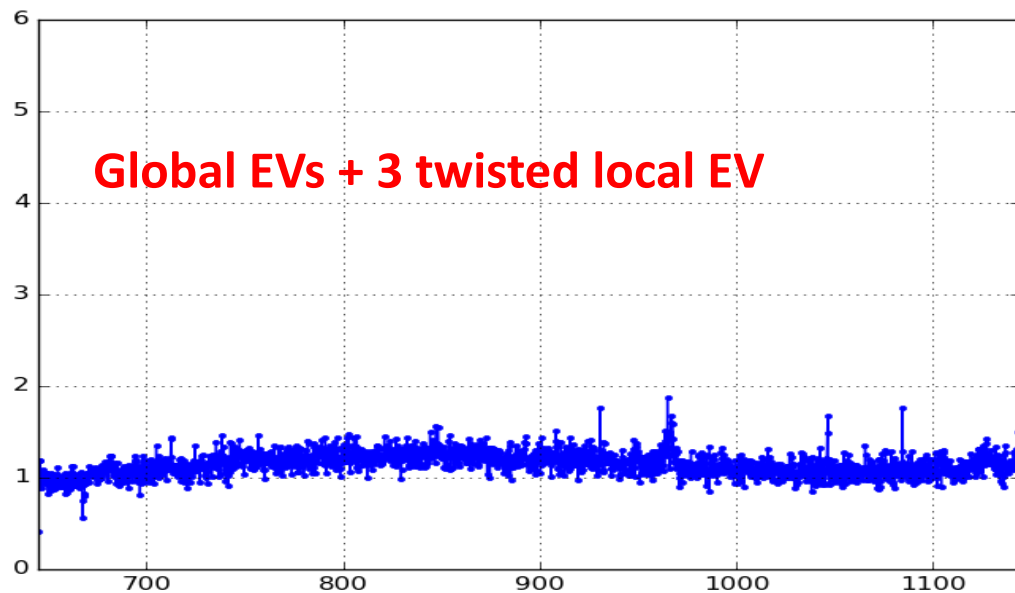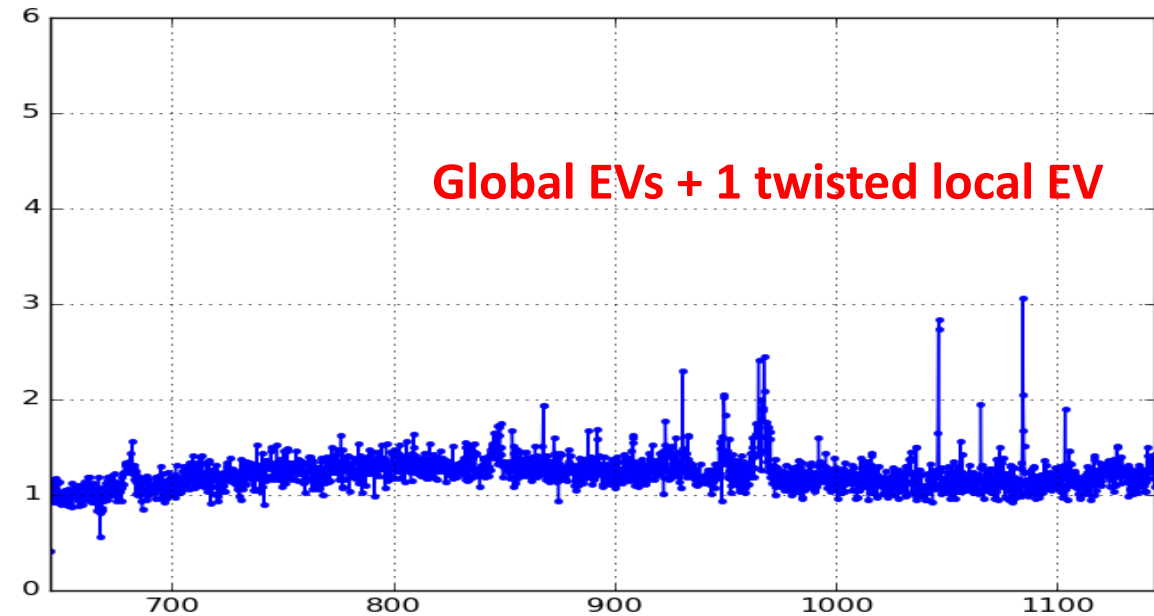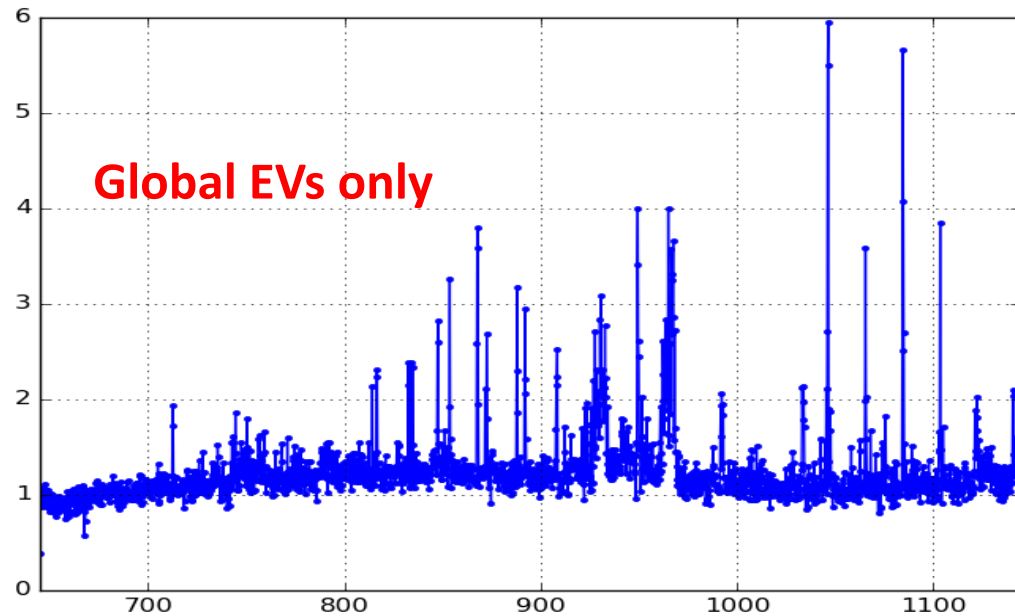After application of 5 twisted local PCs

We observe that the outlier has a very strong reduction of its reconstructions score at the cost of a slight general increase of reconstruction score.
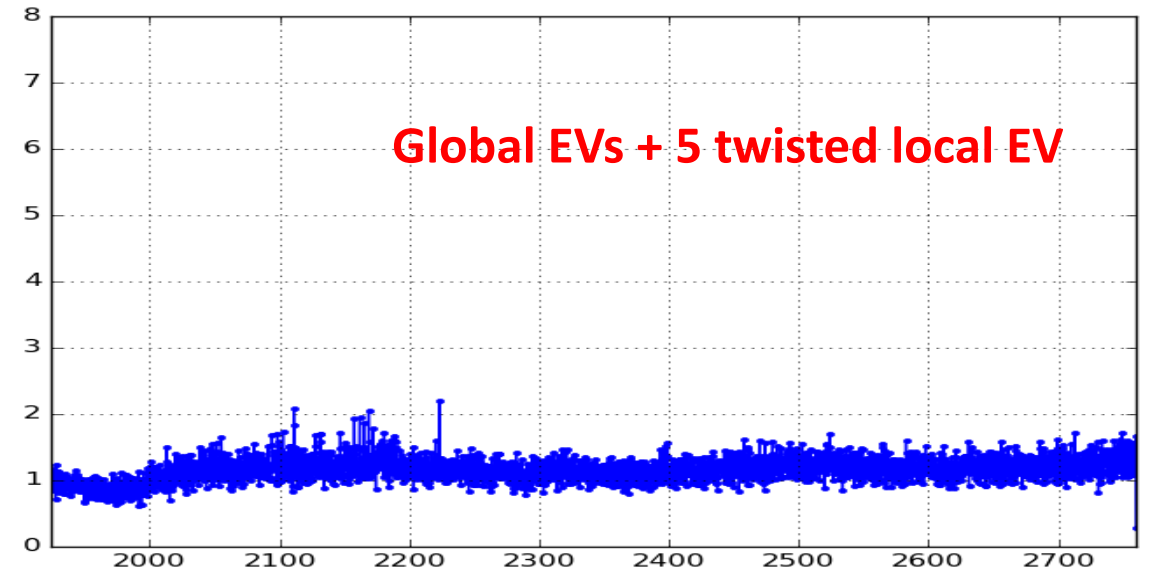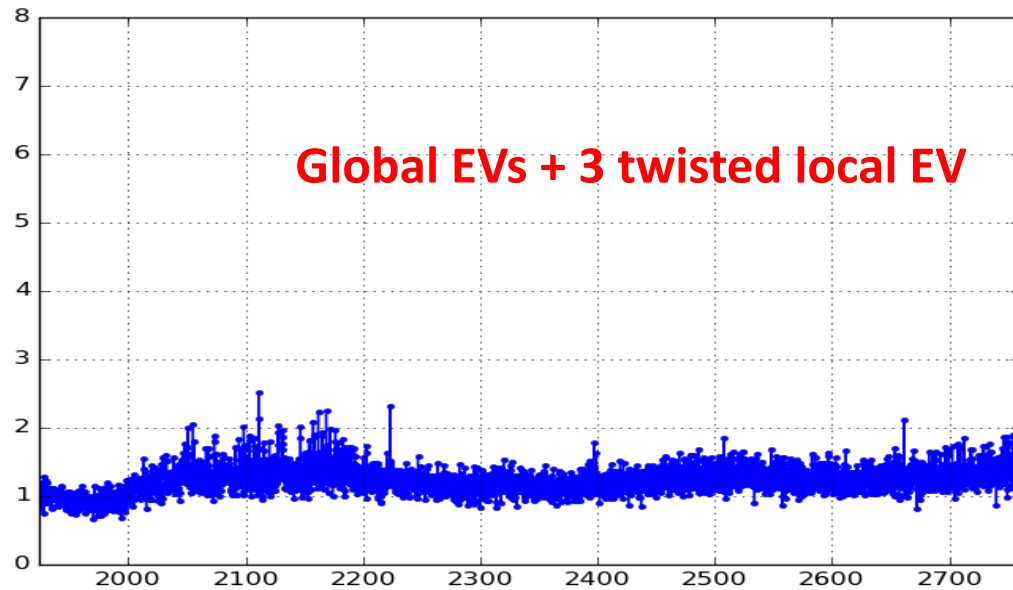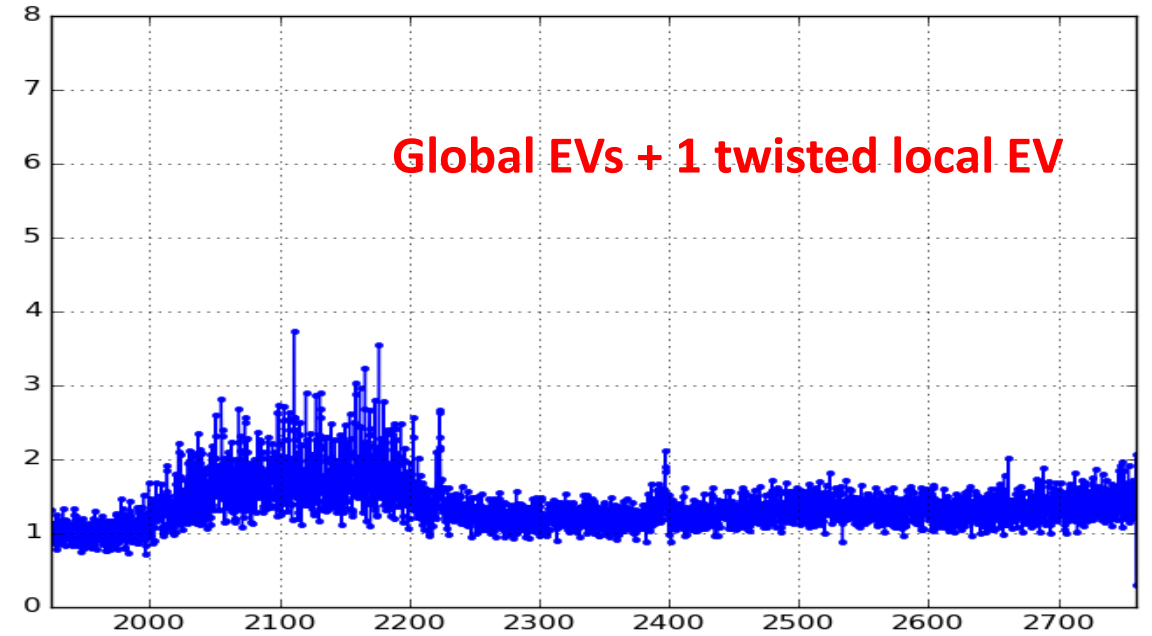
The figures below show the original synthetic outlier residual as well as the corresponding final residuals obtained with each of the two flavors of the hybrid method. The strongly increased ability to retain the unusual spectral features of the outlier with the twisted approach comes at the cost of also retaining more noise (less is left in the residual).
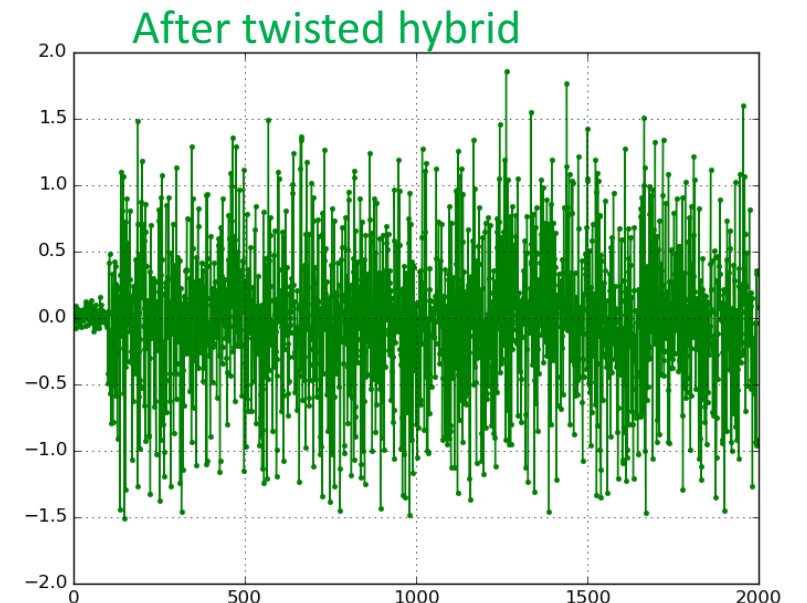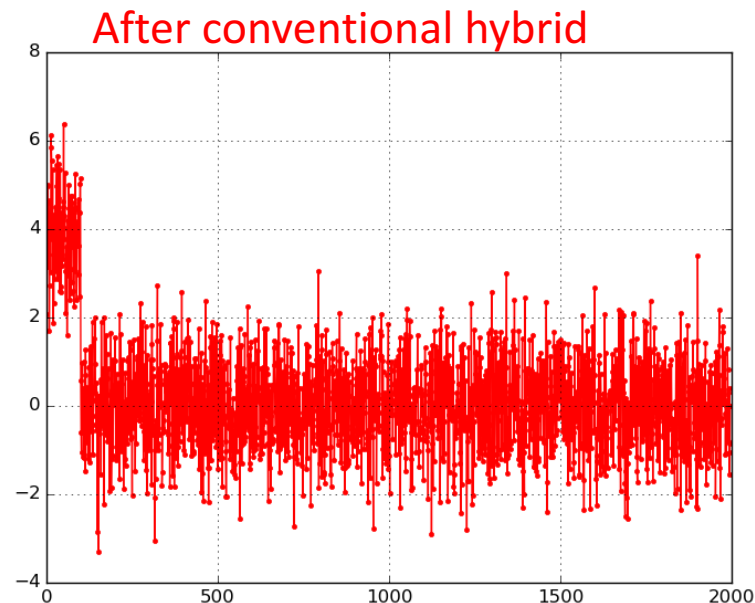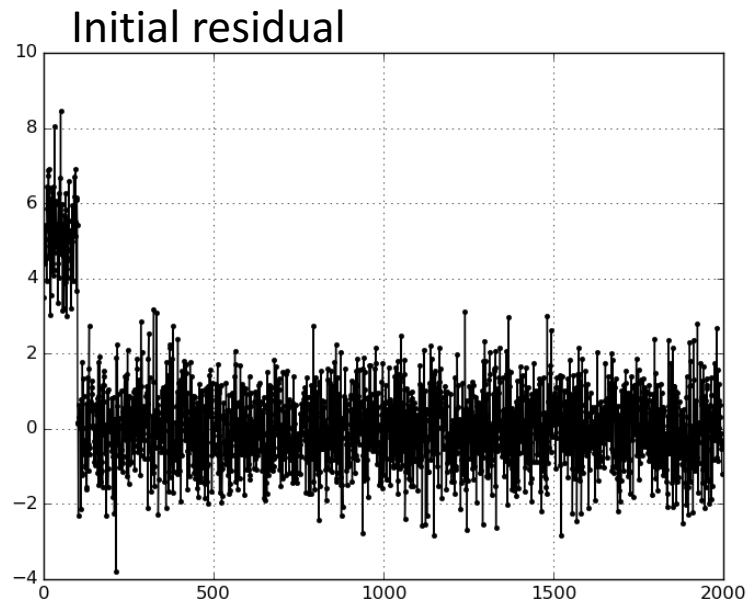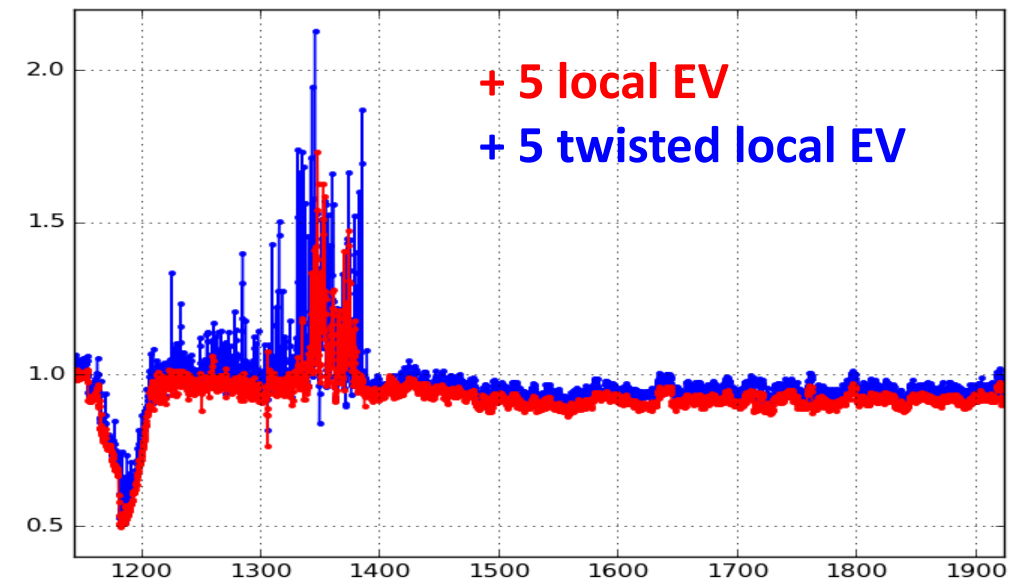
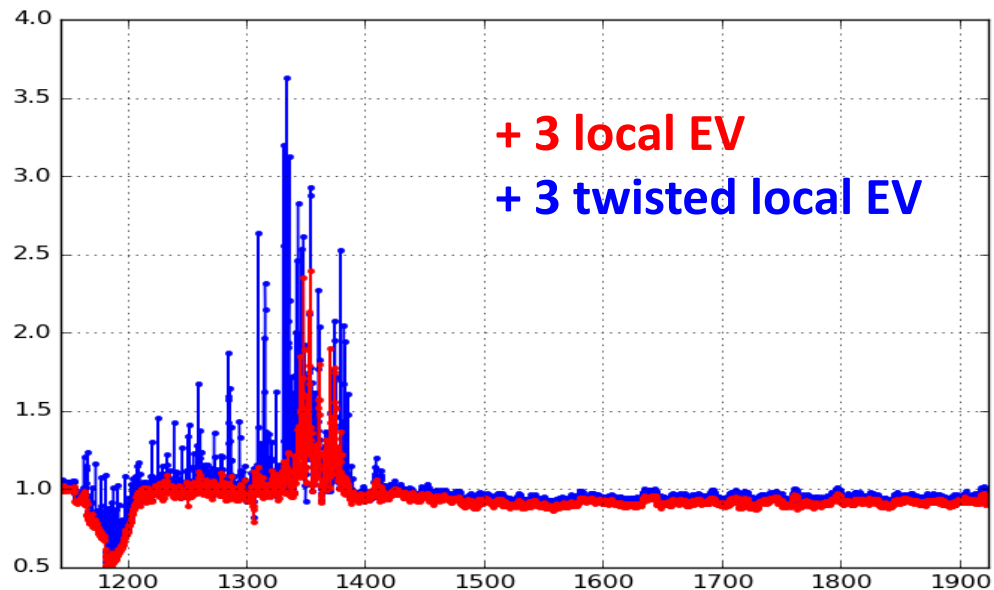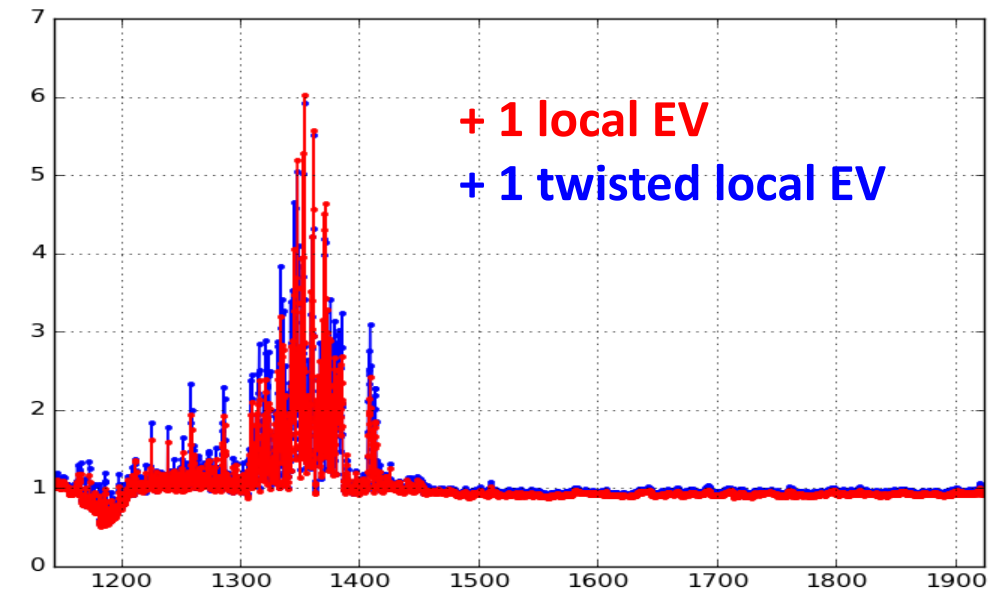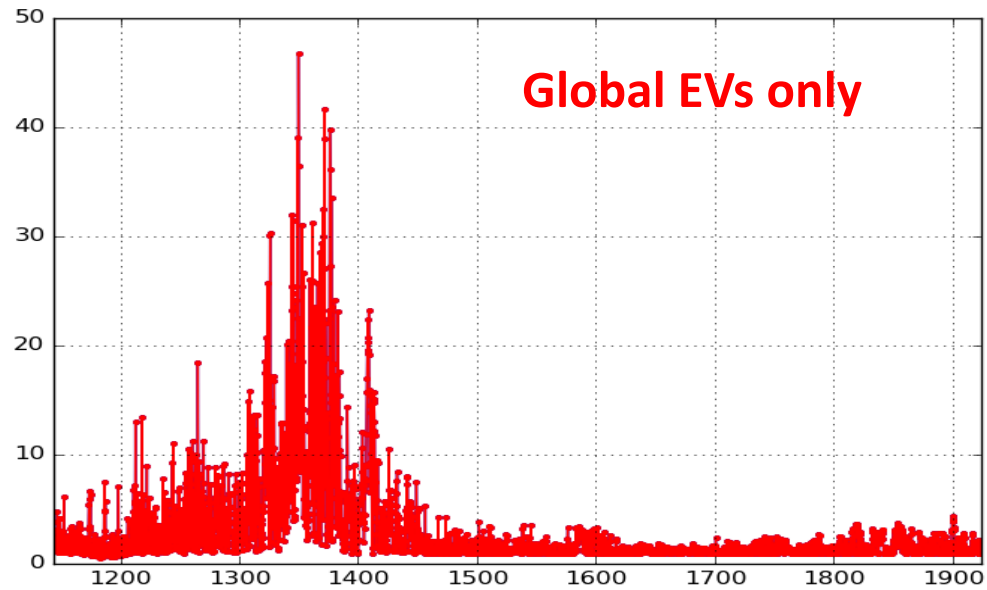# Band 1 (noise normalized) residual RMS for 96 Russian wildfires 2010 outliers

Band 3 (noise normalized) residual RMS for Russian wildfires 2010 (51 outliers)
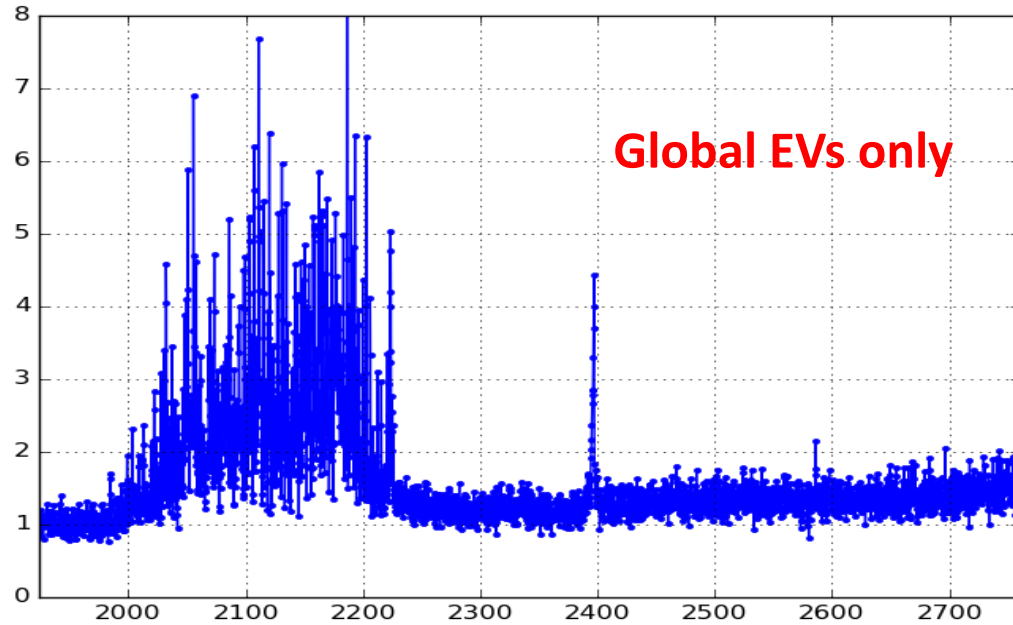
Kasatoshi outliers Band 2, 20080811 (twisted approach worse in this case ➔ don't use it)

Global EVs only

+ 1 local EV
+ 1 twisted local EV

+ 3 local EV
+ 3 twisted local EV

+ 5 local EV
+ 5 twisted local EV

Spectral residual RMS before and after application of local PCs (20180515 Metop-B)

Global PCs only

Global PCs + 5 local PCs
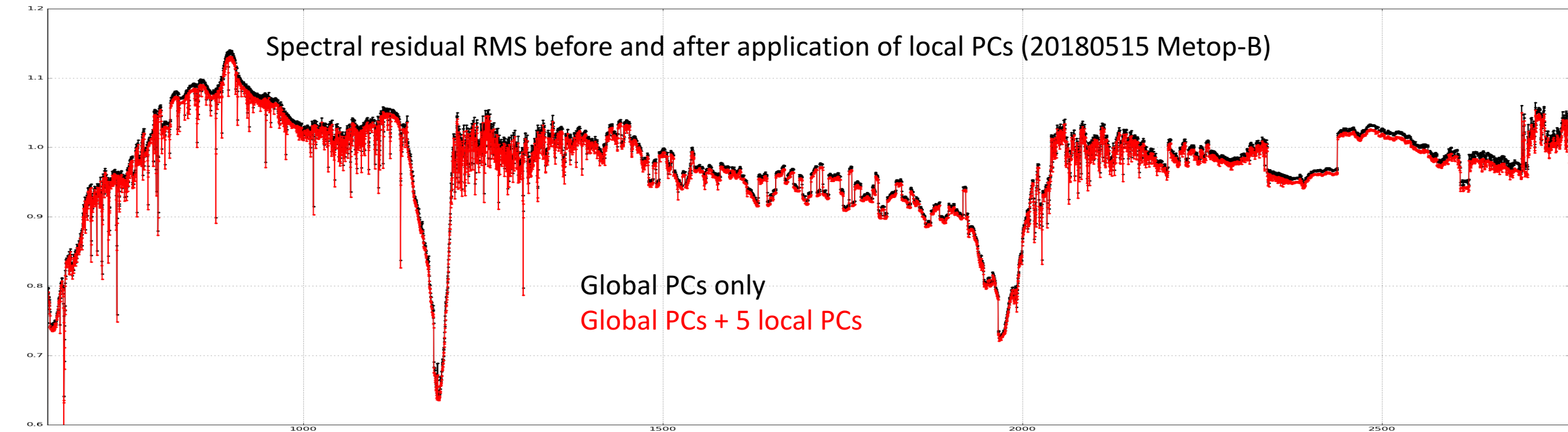
Impact of local PCs on Spectral residual RMS (20180515)

Metop-A

Metop-B

923 cm⁻¹   948 cm⁻¹
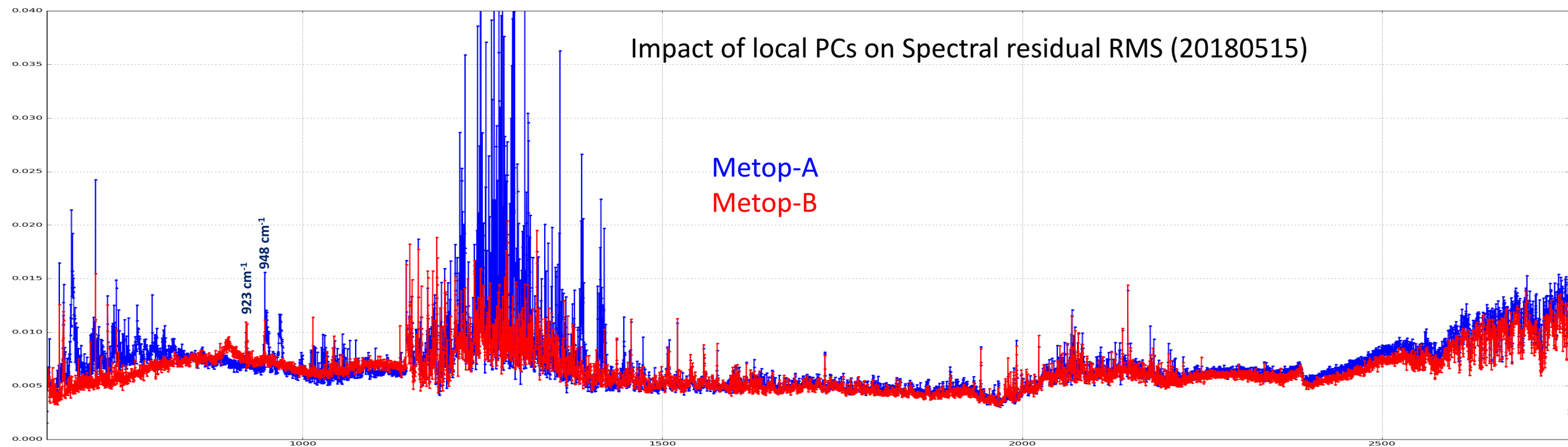
Mean of noise normalized residuals (first week of November 2017, both IASI-A and IASI-B)
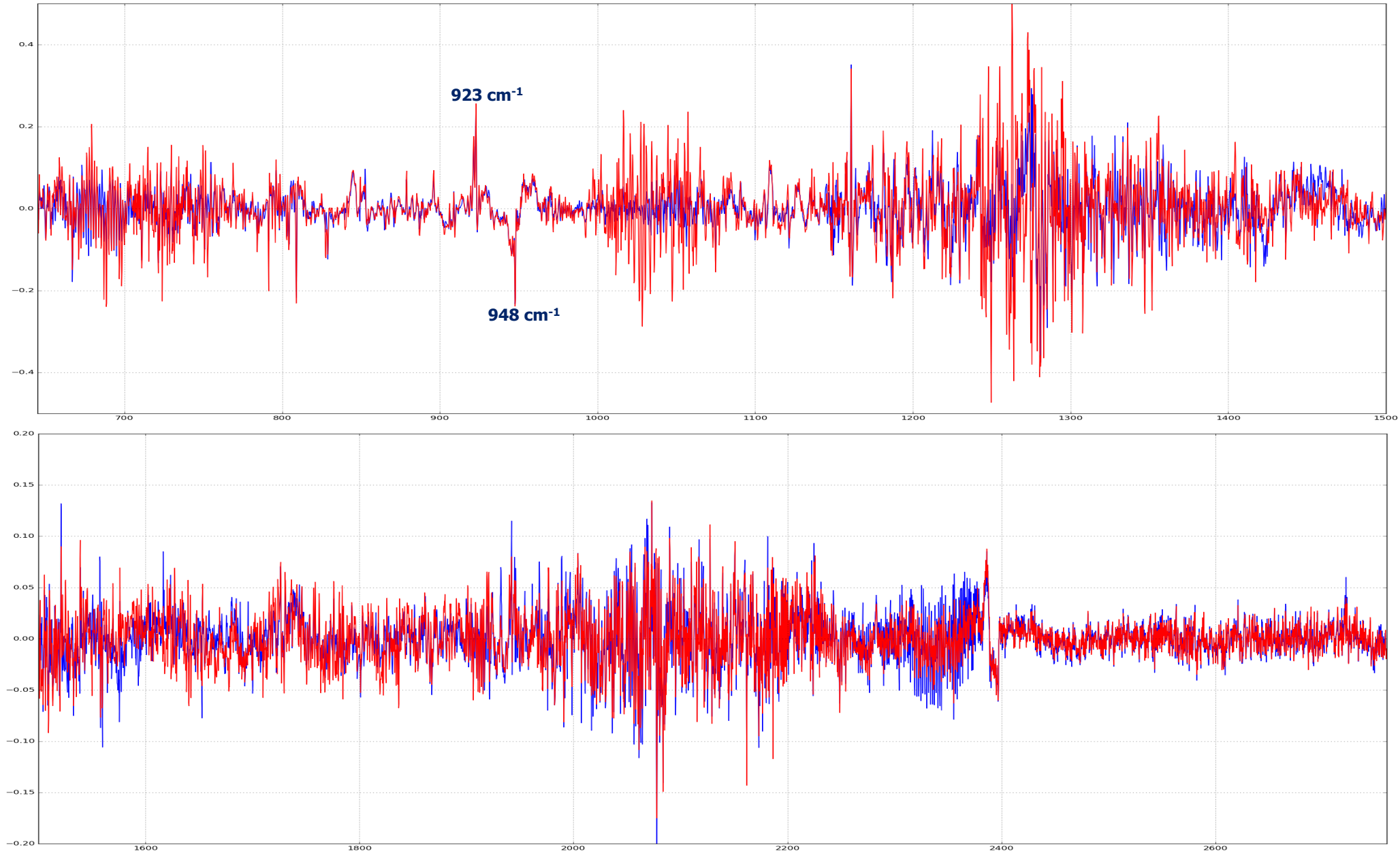
Evolution of mean residual
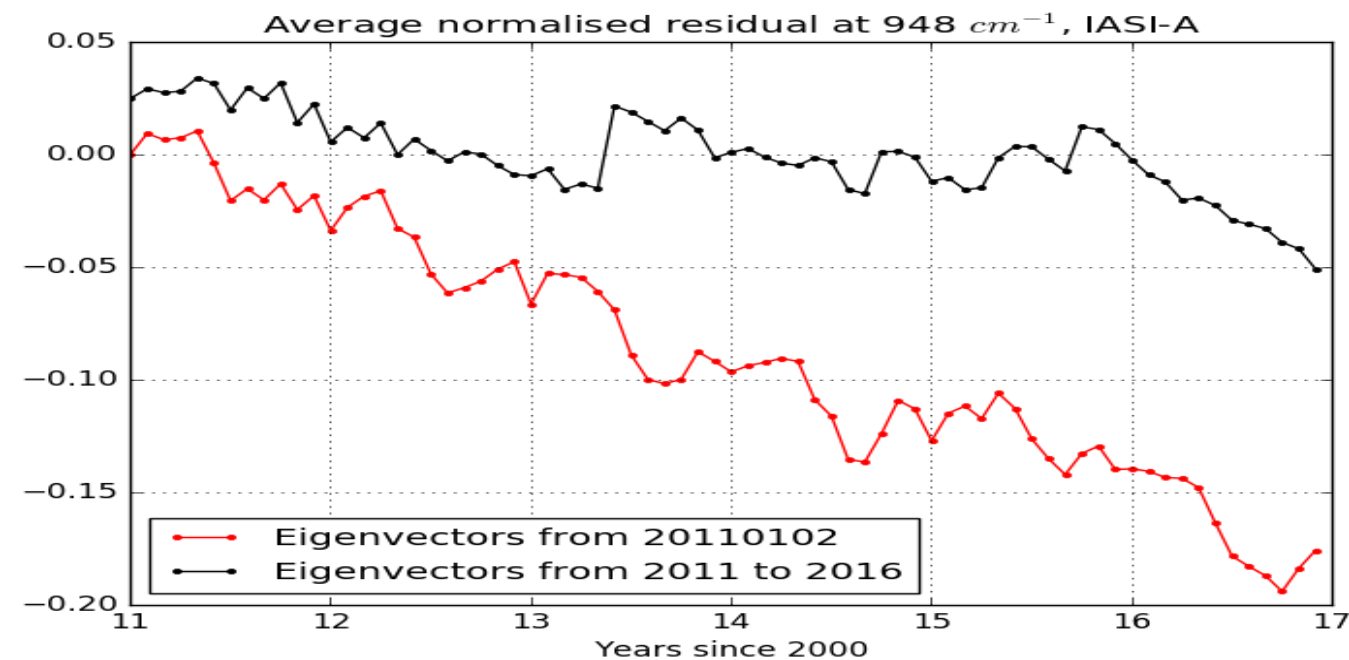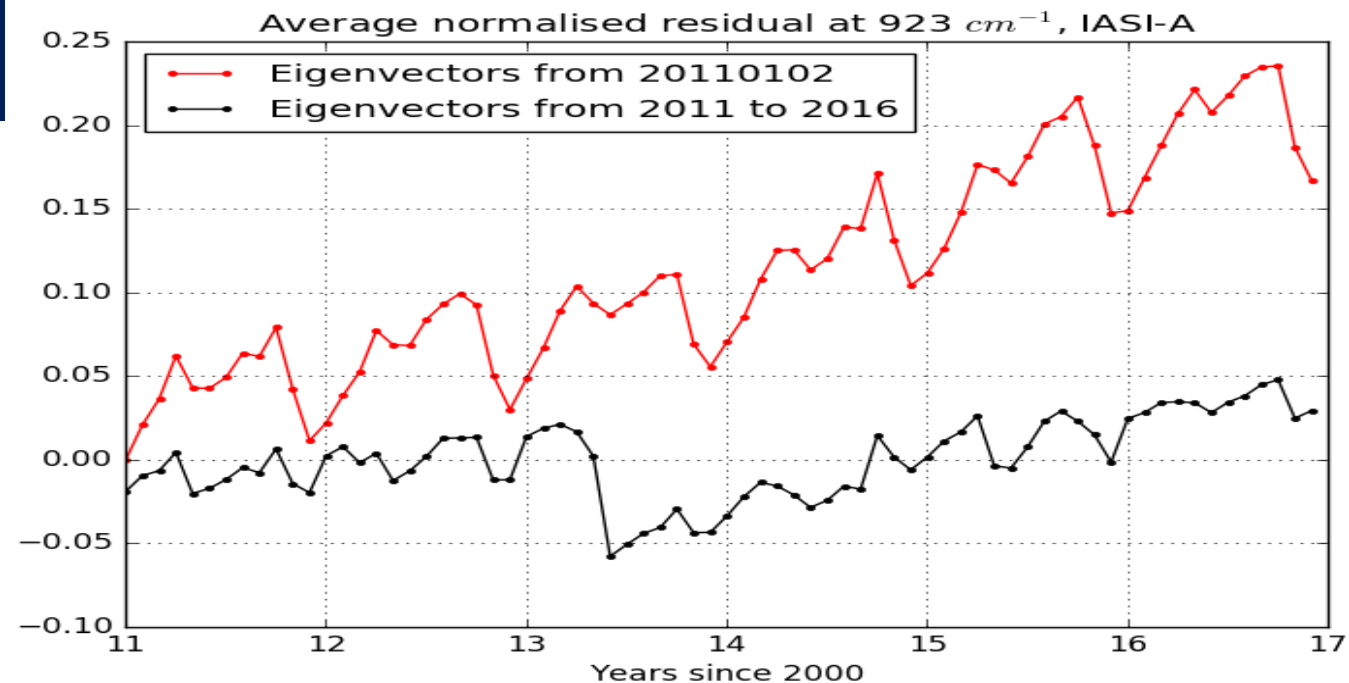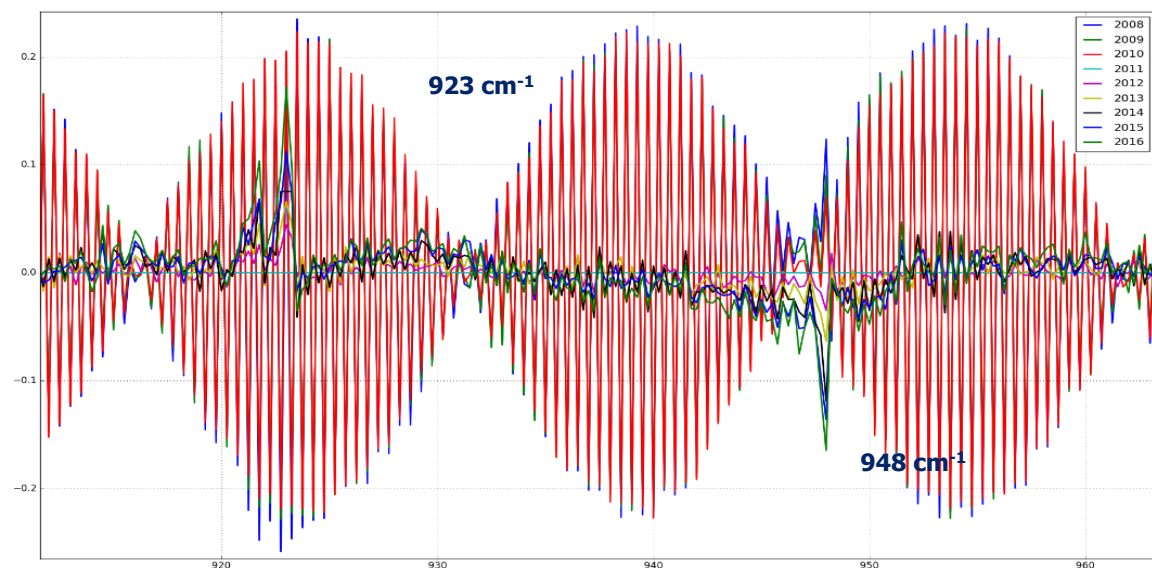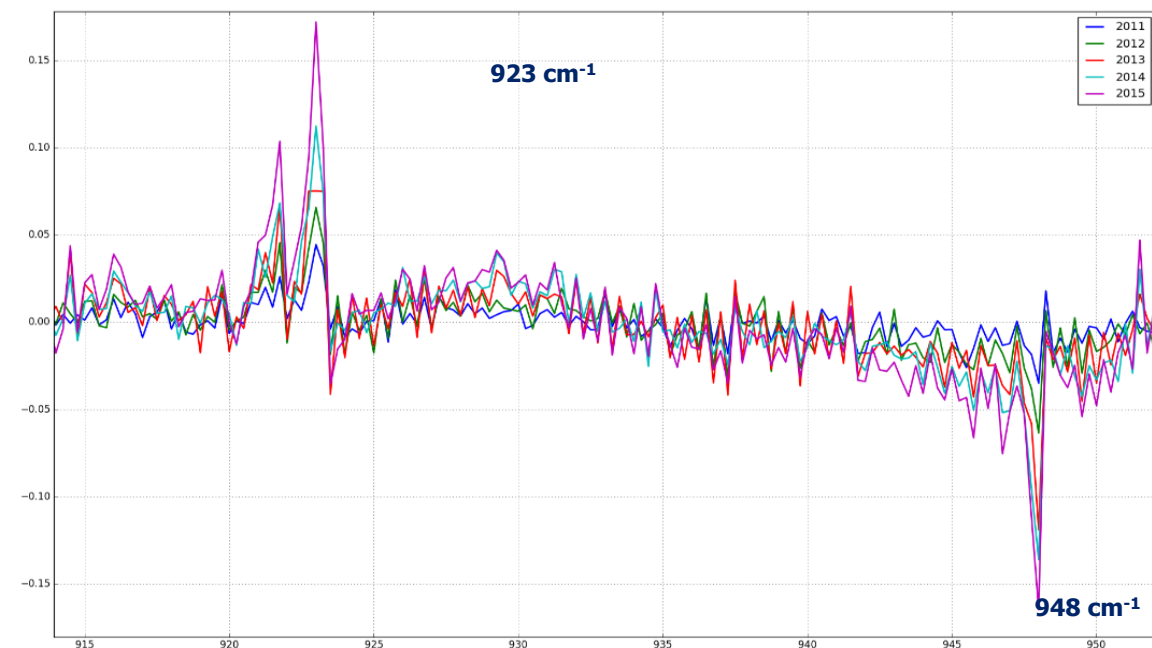
Hybrid approach
IRS PC product content
minor size overhead ➔ 0.6% (rough estimate)

25 600 pixels

300 global PCS

5 local PCS

5 local eigVec

Conclusions

The hybrid approach
- ✓ Meets User requests for stable PCs
- ✓ Offers possibility to represent new spectral features which are correlated within a granule
- ✓ Reduce the need of updating the global PCs:
    - ▪ mitigates need for *ad hoc* updates (new events, trends)
    - ▪ can be planned and announced well ahead to Users
- ✓ Limited additional space required (about 0.6%)
- ✓ Small additional computational effort needed *(few seconds per dwell)*

➔ suitable solution for a PC-based NRT dissemination

Limited skill for isolated outliers in a granule:
- ✓ Less probable with IRS than IASI *(small contiguous pixels vs sparse 12-40km footprints) (?)*
- ✓ Flagged
- ¿ NRT needs ?

Disseminate remaining PC scores in case of outliers ?
- ! Most outliers not caused by the atmosphere: higher instrument noise, spikes,…
- ! Risk impact on bandwidth (and irregular size of products)

Spare slides

Most outliers not caused by atmospheric signal

Undetected "spikes":
High-frequency disturbance of the interferogram, most often observed in the South Atlantic Anomaly. (Band 2 outliers)
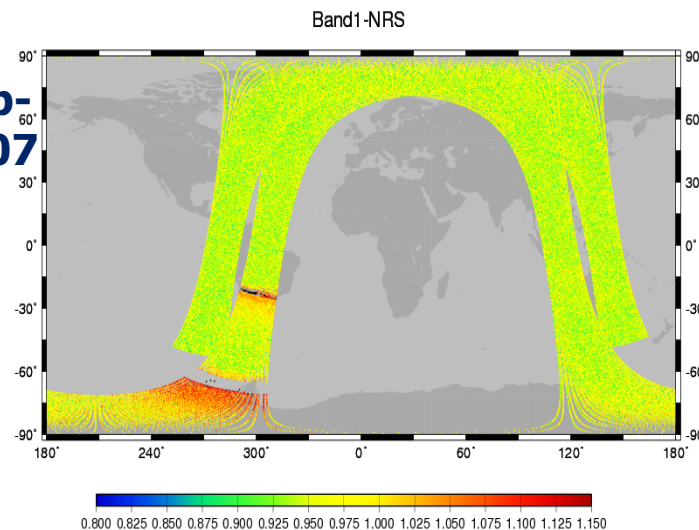

Jan/Feb/Mar 2008
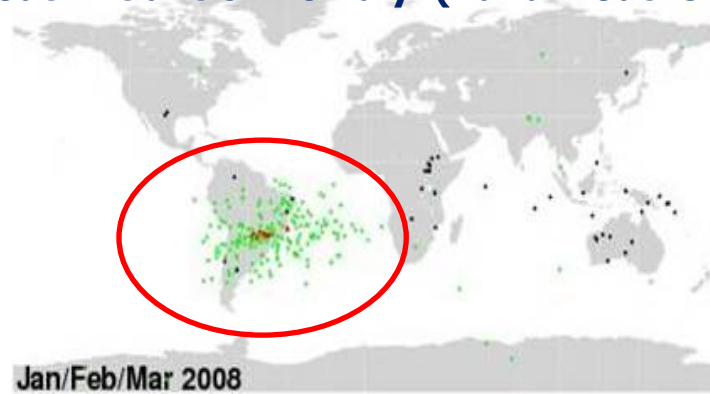
Back to normal operation after external calibration mode:
No history available for deriving filtered calibration coefficients.

Band 1, Reconstruction score after external calibration, 20080825
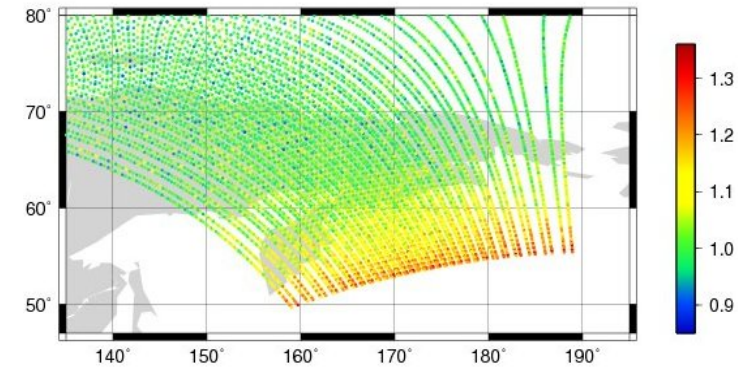
Anomaly related to Metop-B manoeuvre on 20130807 (Met-Office noticed a sudden increase in bias over Brazil when the manoeuvre occurred)

Band1-NRS

0.800 0.825 0.850 0.875 0.900 0.925 0.950 0.975 1.000 1.025 1.050 1.075 1.100 1.125 1.150

Evolution of normalised residual standard deviation (Band 1)

When the noise increase, so does the variance of the residuals
➔ Outliers threshold should be dynamic