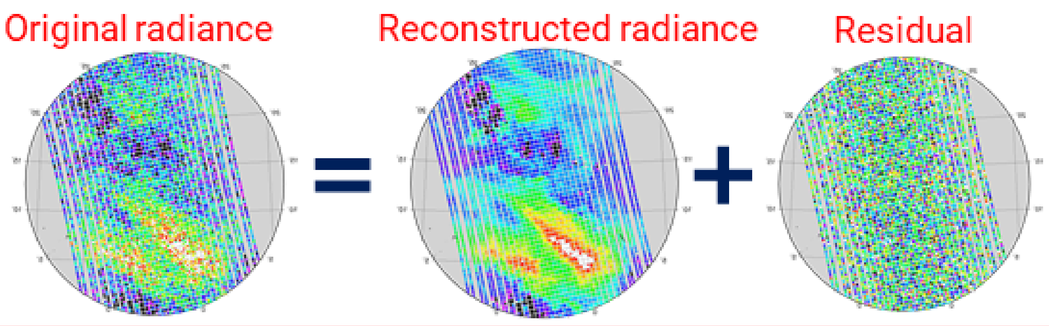
Hybrid Principal Component (PC) Compression

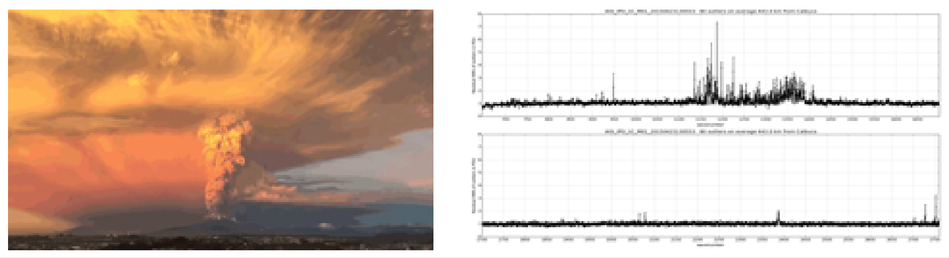
Tim Hultberg (tim.hultberg@eumetsat.int)
EUMETSAT

The Hybrid PC cartoon

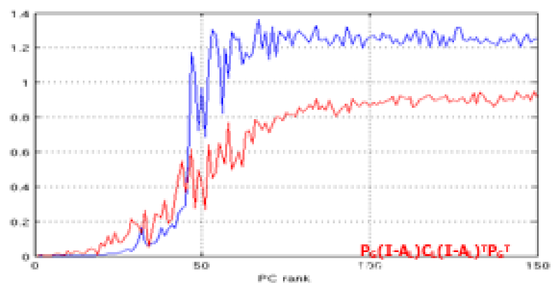
PC compression based on a big global set of past observations works excellent



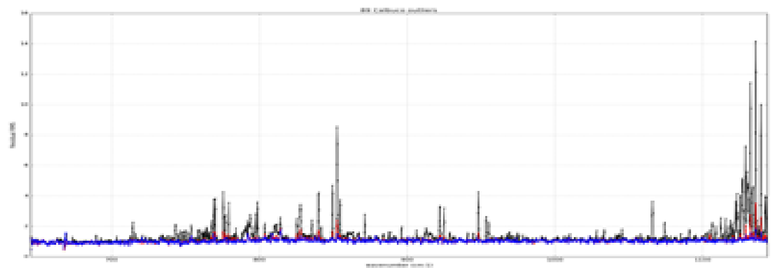
Only in very rare situations new spectral features orthogonal to the previously observed directions occur, which can not be represented well, but are flagged.



Eigenvectors computed for the local granule could solve this issue, but would retain more noise and less atmospheric signal

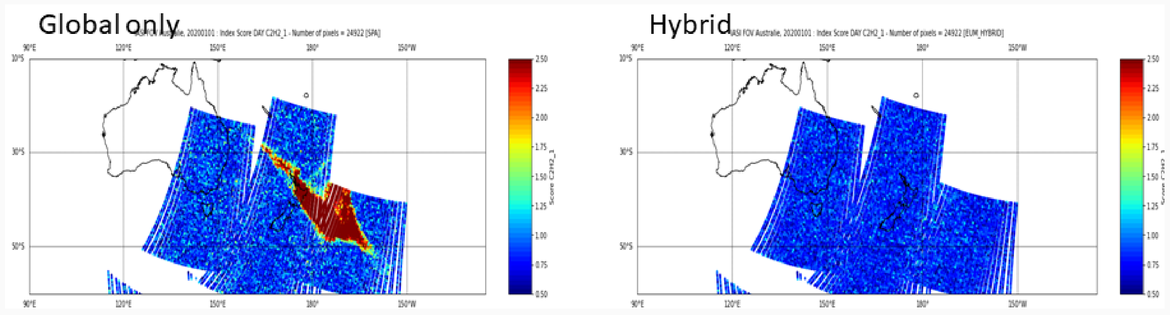


Instead we can supplement the global eigenvectors with a few local eigenvectors, to represent new signal



Capturing unusual features

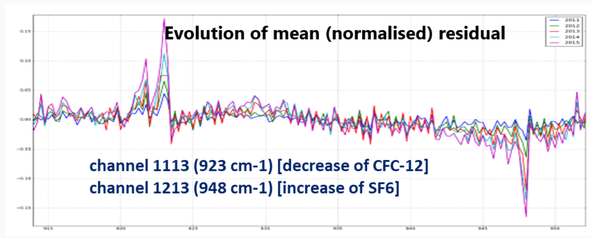
Uncommon gases or unusual amounts of common gases can result in spectral features, which are orthogonal to the global eigenvectors. Normally such situations affect several pixels within a granule with a correlated signal and can therefore be represented by the first eigenvectors based on the local residuals (after application of the global eigenvectors).

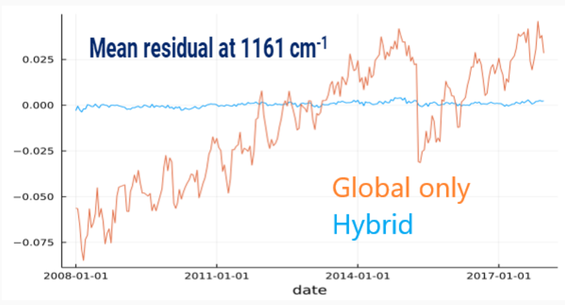


The signature (RMS of normalised residual in 728-732 cm^{-1} spectral region) of acetylene (C_2H_2) in the residuals after global only and hybrid PC compression.

Capturing long term trends

By construction, the residuals are essentially unbiased for the period used for the training set of the global eigenvectors. But as the mean atmospheric state slowly evolves it creates small but growing biases in the residuals.

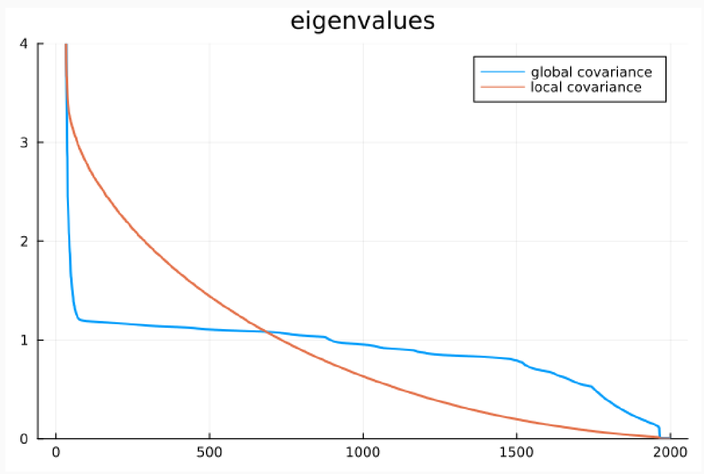




These biases are captured by the local eigenvectors, making hybrid PC suitable for climate purposes. Note, however that also biases caused by changes to the instrument, such as the switch to the redundant IMS of IASI-A on 2015.04.13 (the jump in the figure to the right) are captured.

Why not local only?

The problem with using eigenvectors computed from a single data granule is that the limited number of spectra is not sufficient to estimate the covariance well. This means that the random instrument noise influences the eigenvectors and the noise can not be separated from the signal as well as with global eigenvectors.



Local eigenvectors (only) retain a higher amount of noise and a lower amount of atmospheric signal than global eigenvectors and should be avoided.

Isolated outliers

PC compression works by exploiting correlated spectral features - an isolated spectral feature is indistinguishable from noise. Isolated outliers are rare but possible (sometimes they can be caused by a very localised surface emissivity feature). Usually an isolated outlier will have the highest residual RMS within its granule and we can dedicate a local PC direction to represent it (or essentially equivalently we can disseminate it with the local eigenvectors).

Nevertheless, it is conceivable that the spectral feature of the isolated outlier is so narrow (and weak) that the residual RMS is not increased sufficiently to make it the highest in the granule. Investigations are ongoing to devise a method for identifying such spectra for example by considering the RMS over smaller spectral regions or by using the 3-norm instead of the RMS. This identification is further challenged by the fact that the noise, and thereby the expected residual RMS, depends on the detector and the radiance intensity.

Another option which might help to avoid isolated outliers is to analyse past isolated outliers and dedicate a direction based on them in the global PC basis.

Conclusions

PC compression is a robust and proven method for representing hyperspectral satellite data. However, it can be challenging to faithfully capture:

- **Exceptional situations**
- **Atmospheric long term trends**

The hybrid approach mostly solves both problems. However, when the exceptional situation only affects a single (or very few) pixels within a granule it isn't captured by the local eigenvectors either.

More details about hybrid PC compression can be found in the original paper: "Local or global? How to choose the training set for principal component compression of hyperspectral satellite measurements: a hybrid approach.", Proc. SPIE 10423, Sensors, Systems, and Next-Generation Satellites XXI, 104231G (29 September 2017) by T. Hultberg, T. August and F. Lenti. The method has been further consolidated and evaluated in two parallel EUMETSAT studies conducted by ULB Squares and SPASCIA.

The hybrid Principal Component (PC) Compression approach for hyperspectral data supplements the usual global PCs with a few additional PCs specific to each individual granule. This captures atmospheric trends, which would otherwise be lost, and improves the ability to represent unusual spectral features. It is the baseline representation of MTG-IRS spectra for dissemination and will be introduced operationally for IASI soon (currently planned for 30/3 2023).