Principal Component Analysis of IASI measurements for the detection of extreme events: methodology and application to the detection of rare spectral signatures

Pascal Prunet, Sarah Pipien, Claude Camy-Peyret, Olivier Lezeaux, Dominique Jolivet, Bruno Monsterleet, Adrien Vuvan, Anne Boynard, Cathy Clerbaux, Patrice Henry, et al.

Principal Component Analysis of IASI measurements for the detection of extreme events: methodology and case studies

SPASCIA

Pascal Prunet[1], Sarah Pipien[1], Claude Camy-Peyret[2], Olivier Lezeaux[1]
Dominique Jolivet[3], Bruno Monsterleet[3], Adrien Vu Van[4,1],
Anne Boynard[1,4], Cathy Clerbaux[4], Patrice Henry[5], Laura Le Barbier[5]

[1]SPASCIA, [2]IPSL, [3]HYGEOS, [4]LATMOS, [5]CNES

CNES contract : n°201128/00

OBJECTIVES

We have implemented and tested the feasibility and performances of a systematic and global processing of IASI L1 measurements based on the Principal Component Analysis (PCA) of the spectra, for the fast detection, identification and monitoring of atmospheric extreme events.

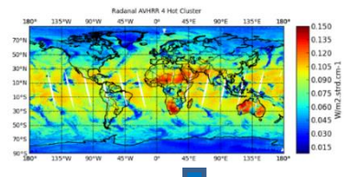
The detected extreme events are defined as data outliers with respect to the representative global variability of IASI spectra (excluding outliers), and thus the detection is driven by a specific metric of the rarity (infrequency) of the spectrum behaviour.

This approach called IASI-PCA and its metrics are described. A first evaluation on 1 year of global IASI data has been done. The ability of this processing to deal with both clear and cloudy data is illustrated, and the analyses of some specific cases are discussed.

1. Principle of the method and processing

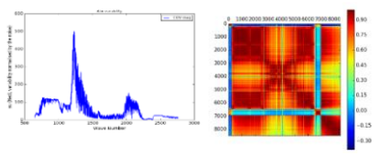
From a reference database of IFOVS
~120 000 spectra : global / 1 year

1



1. Build a global training set, i.e. a representative set of normalised IASI L1C spectra,
2. Compute the corresponding covariance matrix C (dimension of 8461 x 8461),
3. Eigenvector (E) decomposition of the covariance matrix C,
4. Choice of the optimum number of PC (troncature by thresholding eigenvalues as a function of the PC rank)

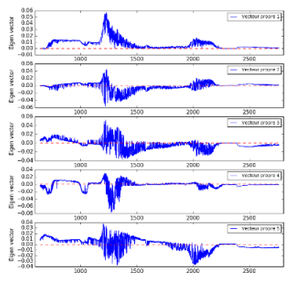
Statistics of IASI global variability …
Mean spectrum $\bar{y}$ and spectral variances/covariances



2

3

and eigenvectors database E

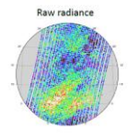


E, $\bar{y}$

4

5. Compute the principal components by projecting one spectrum (y) on the truncated eigenbasis,
6. Compute the spectral residual r (difference between original and reconstructed radiances, normalised by IASI instrumental noise N. Compute the reconstruction score R over m spectral channels to efficiently identify outliers.

Raw radiance

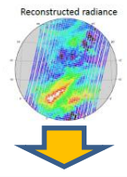


5

$$PCC = E^{*T} N^{-1} (y - \bar{y})$$

$$\tilde{y} = \bar{y} + NE^{*} PCC$$

Reconstructed radiance



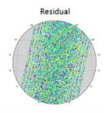
6

$$r = N^{-1}(y - \tilde{y})$$

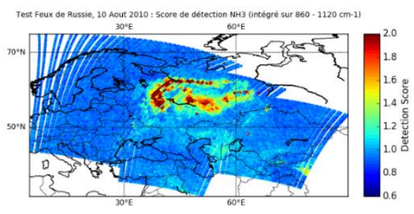
$$R = \sqrt{\frac{1}{m}\sum_{i=0}^{m} r_i^2}$$

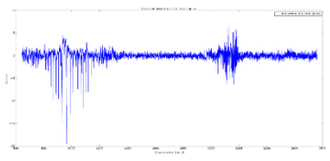
Here R is the global reconstruction score

Residual



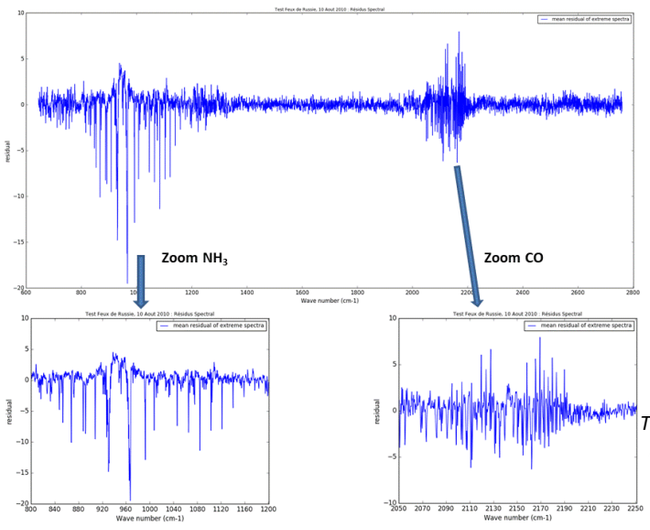
Detect outliers, map extreme events





2. Indicators

Extreme events are detected using **indicators** for which reconstruction scores are computed over reduced spectral intervals targeted *a priori* on species of interest

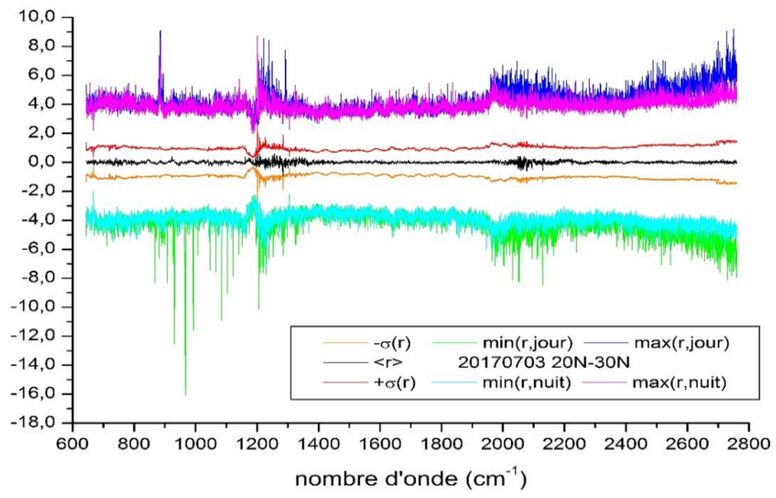


ind_name	wn_1	wn_2	thres_day	thres_night
SO2_1	1138.500	1148.000	1.537	1.535
SO2_2	1320.000	1338.000	1.439	1.340
SO2_3	1338.250	1355.000	1.360	1.318
SO2_4	1371.000	1371.750	1.832	1.848
SO2_5	1376.000	1376.750	1.672	1.695
SO2_6	1344.750	1345.250	1.709	1.680
N2O_1	1153.750	1163.500	1.641	1.606
N2O_2	2191.000	2199.250	1.647	1.489
CO2_1	758.000	766.750	1.673	1.633
CO2_2	788.000	793.000	1.702	1.690
CO2_3	968.250	975.250	1.557	1.541
CO2_4	2048.250	2057.500	1.998	1.694
CO2_5	2380.000	2400.000	1.558	1.369
CO2_6	2064.500	2065.500	2.235	2.143
CO2_7	2077.250	2078.250	2.258	2.243
NH3_1	961.000	971.000	1.670	1.460
NH3_2	925.000	935.000	1.670	1.460
CO_1	2143.000	2181.000	1.668	1.353
CO_2	2150.000	2159.250	1.805	1.560
CO_3	2178.500	2191.000	1.650	1.442

Table 1 : Extract of the molecule indicators list used to detect and characterize extreme events. They have been defined using known absorption features (spectroscopic database, HITRAN, 2012).

3. Spectral residual statistics

Extreme events are carefully analysed and interpreted by identifying anomalies (or outliers) in the **spectral residual statistics** i.e. cases for which the extremes fall outside a given "confidence intervals".

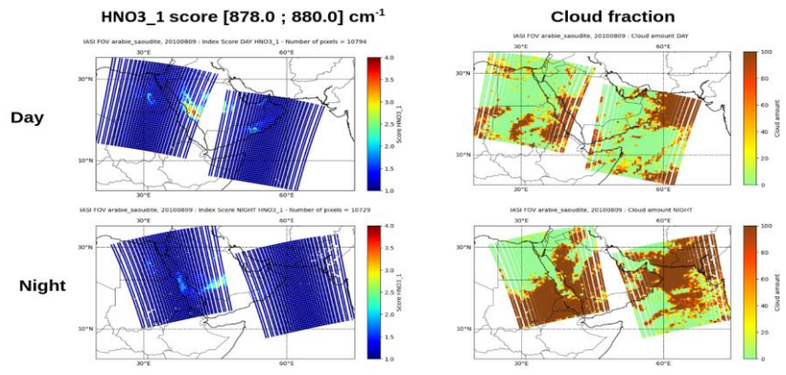


We determine the statistical characteristics of the reconstruction residuals of all the IFOVs of a certain type (day / night, clear / mixed / overcast) for a given period and geographical area (1 day and 1 latitude band) the average (black), the dispersion around the mean (red and orange), the two extremes (max in purple and blue, min in green and light blue) for all spectral channels.

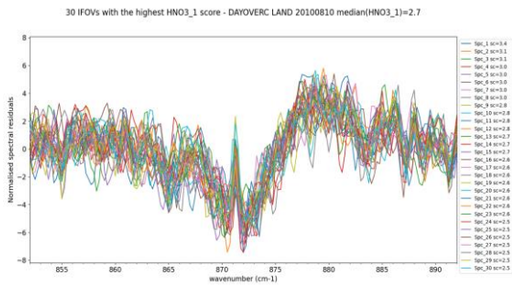
EXAMPLES OF RESULTS

Saudi Arabia outliers

Unknown extreme event is detected from dedicated indicators on 9th (figures below) and 10th of August 201



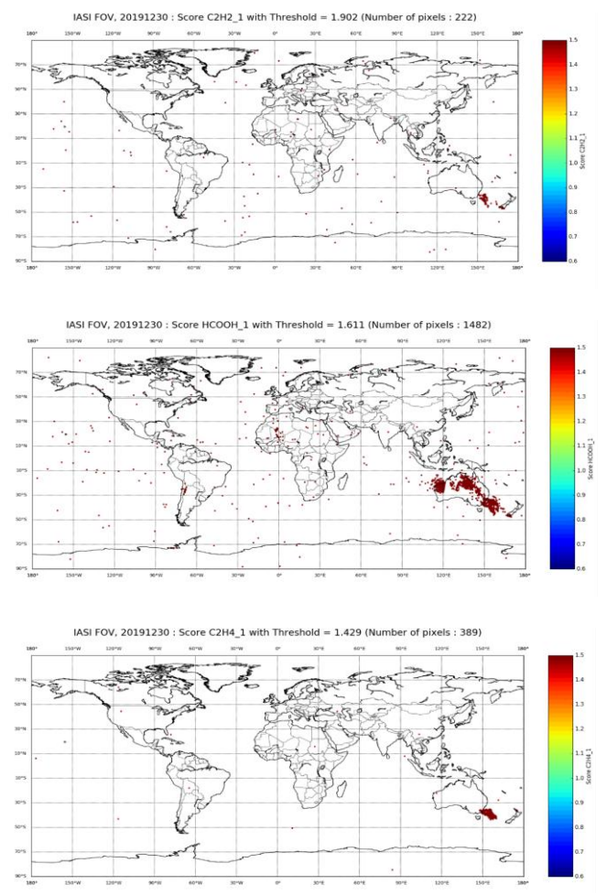
High scores are observed in different conditions (night/day, land/sea, clear/overcast). It allows to detect and follows the plume in clear and cloudy conditions. The spatio-temporal behaviour of this event is consistent with a gas or particle (desert dust) plume.



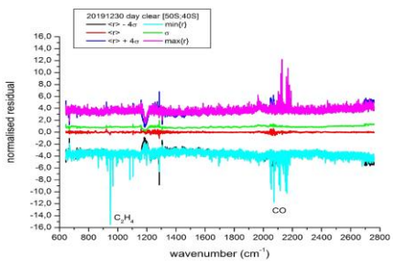
The outliers show a spectral signature in the same domain as the C_2H_4O gas (absorption band centered at 872 cm^{-1}). However, such a signature can also be due to particules such as calcite (Clarisse et al. 2013, Atmos. Chem. Phys., 13, 2195-2221). MODIS image and MERRA reanalysis confirm the presence of a dust event during the same period. Investigations on the physical origin of this event are ongoing.

Analysis of Australien fires from indicators and spectral residual statistics

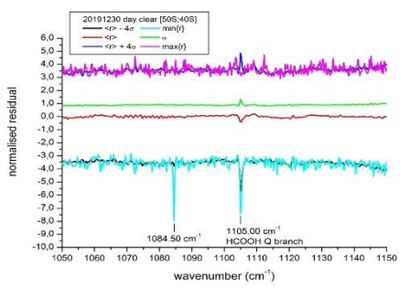
Predefined indicators allow the automatic detection, mapping and preliminary characterization of fire plumes. Maps below are examples of detection of C_2H_2, C_2H_4 and HCOOH high scores



Analysis of spectral residual statistics over the concerned days and latitude band allow to confirm the presence of detected gases, to identify additional species, and to highlight spectral signatures of interest to be further analysed. The precise analysis of these spectral residual statistics is very powerful in detecting and identifying anomalies.



This analysis makes it possible to immediately identify strongly marked peaks, indicating in particular the presence of outliers associated with the species C_2H_4, as well as with CO (which can be analyzed as specific for fires in a hot and convective atmosphere).



In this zoom over the domain [1050.00-1150.00] cm^{-1}, a signature of the Q branch of HCOOH is detected. Beyond these typical species of major fire event,s we identify unknown or unexpected spectral signatures, thus making ipossible to detect new species, or events not identified a priori

First conclusions: A processing method for hyperspectral atmospheric sounding measurements has been developed to detect and characterize "rare" atmospheric events. This treatment has been prototyped, evaluated and tested on IASI data. It was then applied for different datasets and objectives (see also posters from Adrien Vu Van and Sarah Pipien). At this point some interesting results have been obtained : ability of the approach to process clear as well as cloudy data; ability of the processing to work systematically on large amounts of data; implementation of different methods for the analysis of reconstruction to identify anomalies of interest and characterize the corresponding events. While the potential seems high, there is still a lot to do and to show.