Principal Component Analysis (PCA) for detection and classification of aerosols and clouds using IASI/MetOp sounder

Justin Rusalem, Hervé Herbin, Philippe Dubuisson

Ph.D student, Laboratoire d'Optique Atmosphérique, Université de Lille 1, F-59655, Villeneuve d'Ascq, France

Introduction

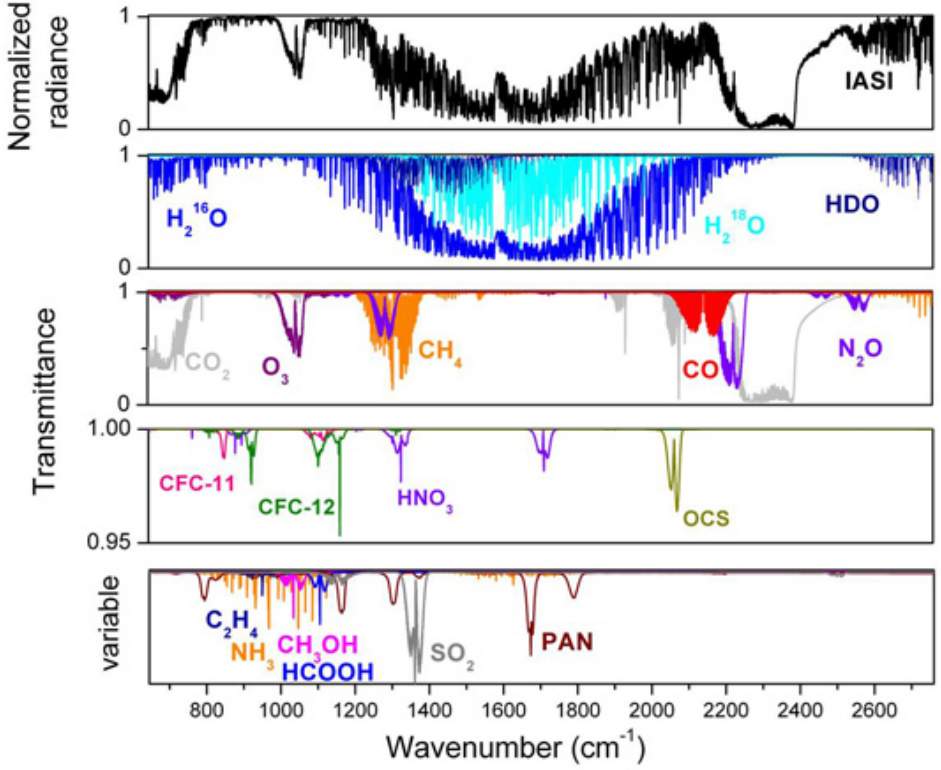


Figure: From [Clerbaux et al (2009)], normalized radiance (Top). Absorption signature of different gases

Goals of the IASI missions (2006-):

- Meteorology: Temperature profiles (<1K), H_2O profiles (<10%).
- Tropospheric chemistry: Study of 24 tropospheric gases and short lives species [Clarisse et al. (2011)]

Give a global spatial distribution at high spectral resolution to support operational meteorology, environnemental forecasting and global climate monitoring

Sensitivity to particules

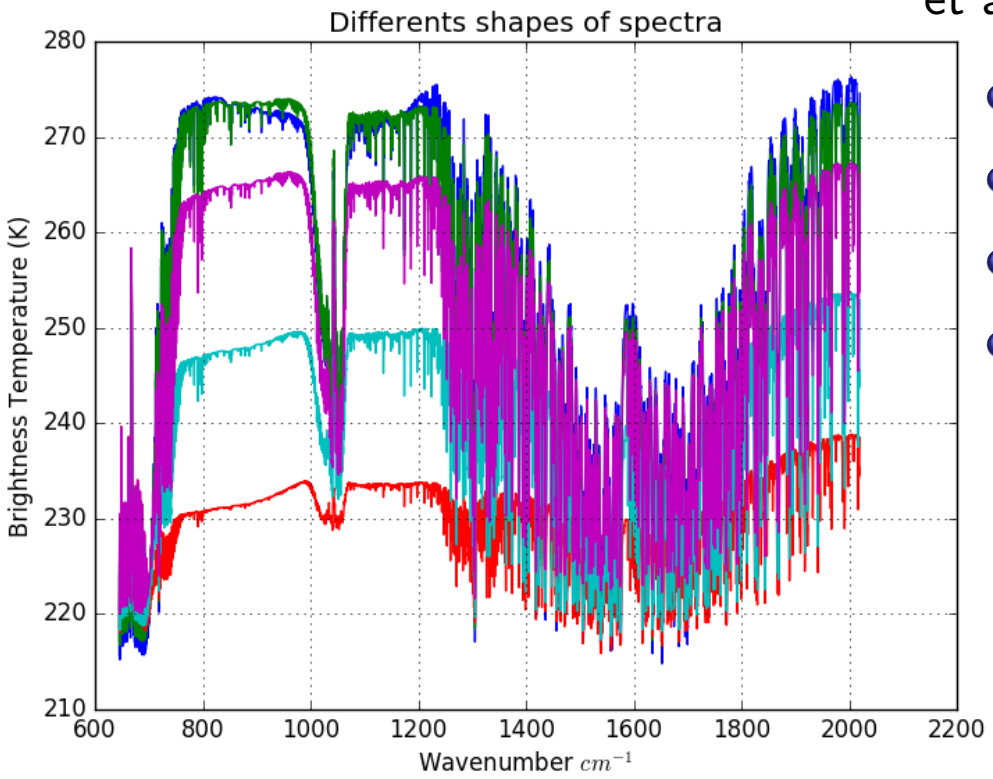


Figure: Specific signatures of :Aerosols, Clear sky, Water clouds,Mix clouds and Ice clouds

We can furthermore access [Clarisse et al. (2012)], Herbin et al. (2013)]:

- Particules size distributions,
- Concentration,
- Altitude,
- Layer thickness

$\hookrightarrow$ **Need a retrieval algorithm, and some preliminary information...**

Refractive index

Link between optical properties and micro-physical properties

$$m = n + i\kappa$$

Figure: Water

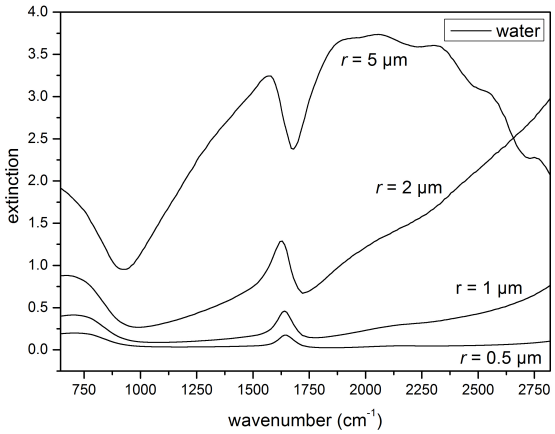


Figure: Dust

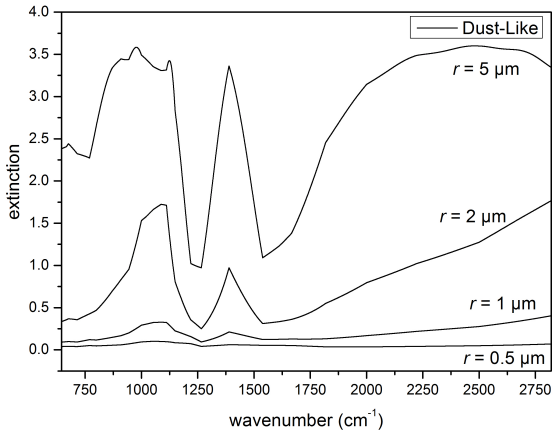
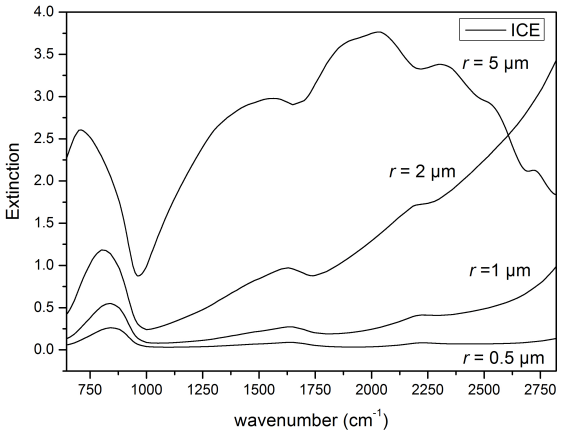


Figure: Ice



Spectral distribution of extinction in function of geometric radius (r) for different types of particules.

To access this a priori information, we need to have an information about the "nature" (type) of each IASI pixels .

Situation and Problems

Thus, we have to take in account, for assimilation of data in meteorological centers (ECMWF, Meteo France...) :

- Number of daily data : $\simeq 10^6$
- Lenght of spectra : 8640 channels
  $\Longrightarrow$ **Typically, a split-window (DBTs) characterization time is $\simeq$ 20h for a global scene.**
  To caracterize pixels in 5 categories: Clear sky (Cc), Aerosols (Ae), Ice clouds (IceC), Water clouds (WatC) and Mix clouds (MixC).

$\hookrightarrow$ We investigate a method based on the Principal Component Analysis...

Generalities

This approach is used in:

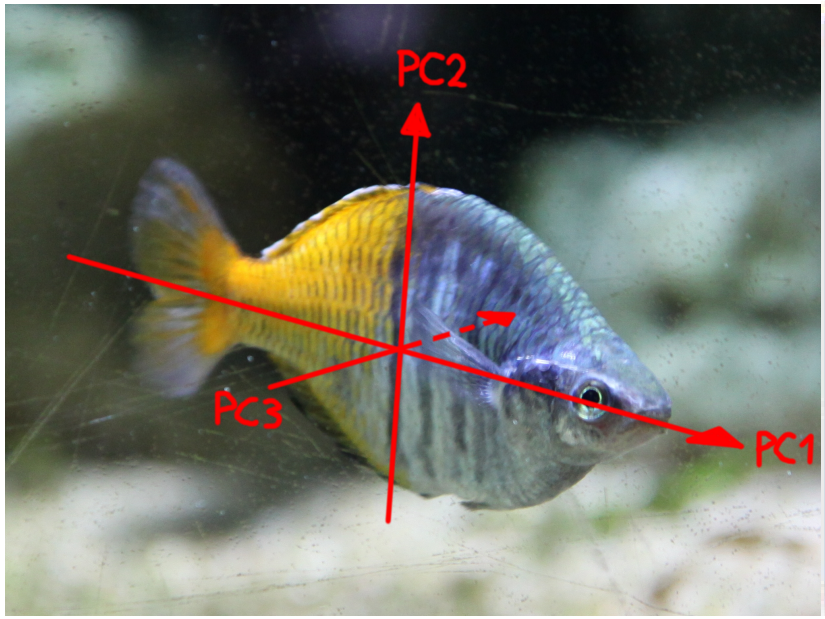
- Data Mining-BIGData, econometry, informatics, statistics,mathematics, chemistry, biology, etc.
- Remote sensing (very few referenced studies):
  Eumetsat -->
  - Datas storage (onboard)
  - Datas transfers
  - ground segments

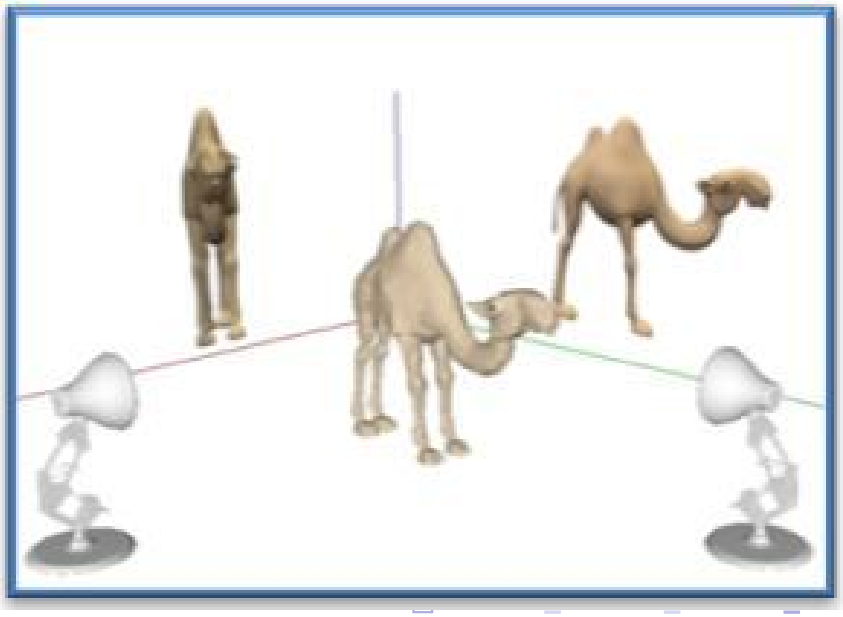
  Atmosphere
  - Noise reducing and spectrum analysis [Atkinson et al. (2010)]

What PCA is ?

PCA allows:

- a geometrical representation (thus in a three-dimensional Euclidean space) of a data group whose number of variables is large.
- to reduce the number of variables.
- to conserve a maximum of the information.
- and to "remove" interdependencies between variables (data de-correlation)





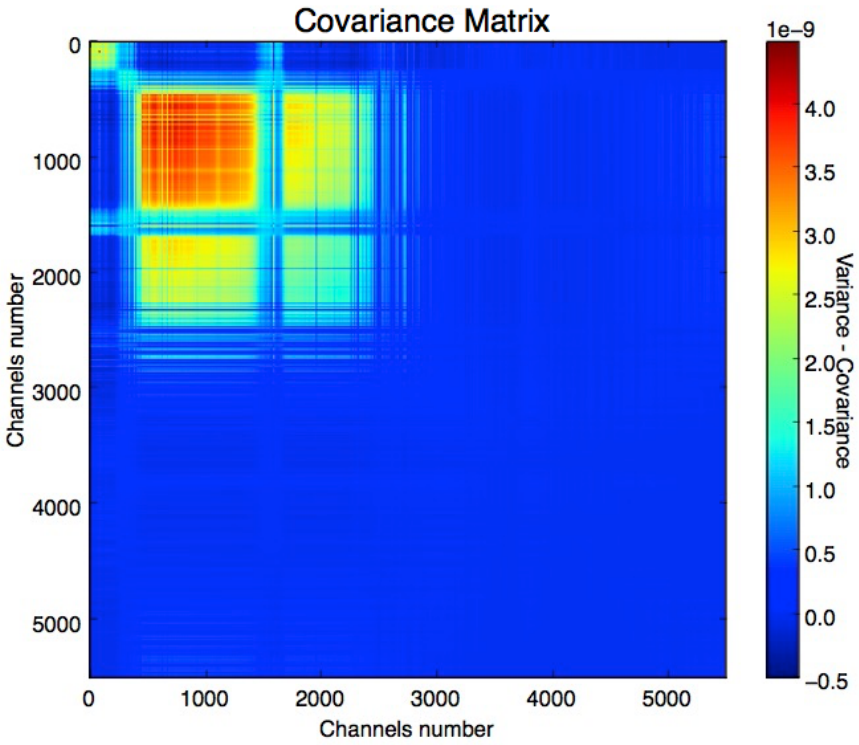
Methodology – Mathematical development

- Place a set of N observations: IASI pixels of P observables (variables): IASI channels into a matrix $\mathcal{A}^P{}_N$

$$\mathcal{A}^P{}_N = \begin{pmatrix} Spectrum\ \#\ 1 \\ \dots \\ Spectrum\ \#\ N \end{pmatrix} = \begin{pmatrix} \leftarrow & Number\ of\ channels & \rightarrow \\ \vdots & & \\ \downarrow & Number\ of\ spectra & \end{pmatrix}$$

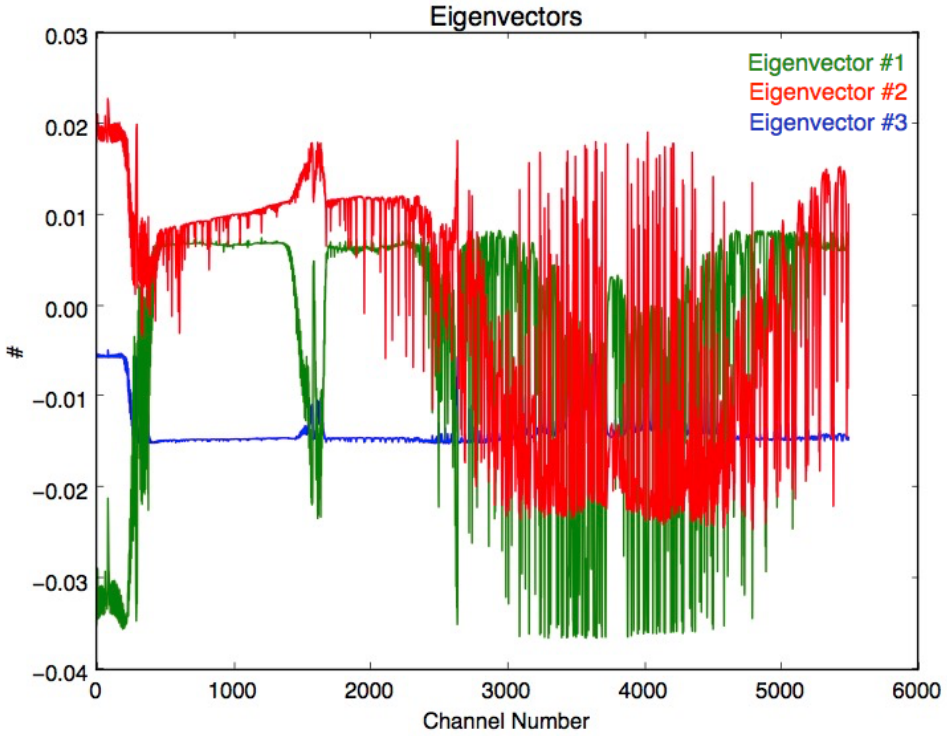
Methodology – Mathematical development

- Place a set of N observations: IASI pixels of P observables (variables): IASI channels into a matrix $\mathcal{A}^P_N$
- Build a variance-covariance matrix



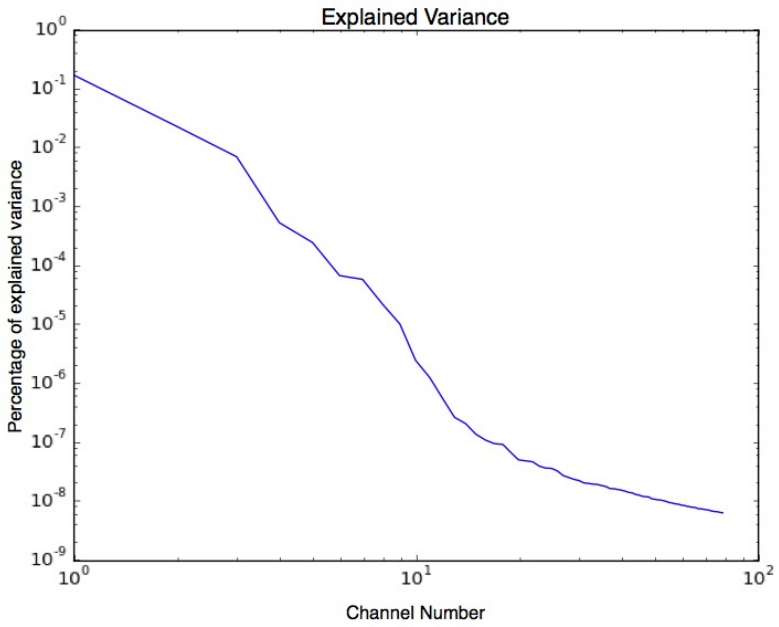
Methodology – Mathematical development

- Place a set of N observations: IASI pixels of P observables (variables): IASI channels into a matrix $\mathcal{A}^P_N$
- Build a variance-covariance matrix
- Take m eigenvectors and eigenvalues (orthogonalisation).



Methodology – Mathematical development

- Place a set of N observations: IASI pixels of P observables (variables): IASI channels into a matrix $\mathcal{A}^P_N$
- Build a variance-covariance matrix
- Take m eigenvectors and eigenvalues (orthogonalisation).
- Projection of the observations on the subspace (in order of variances so in order of eigenvalues) created by eigenvectors.

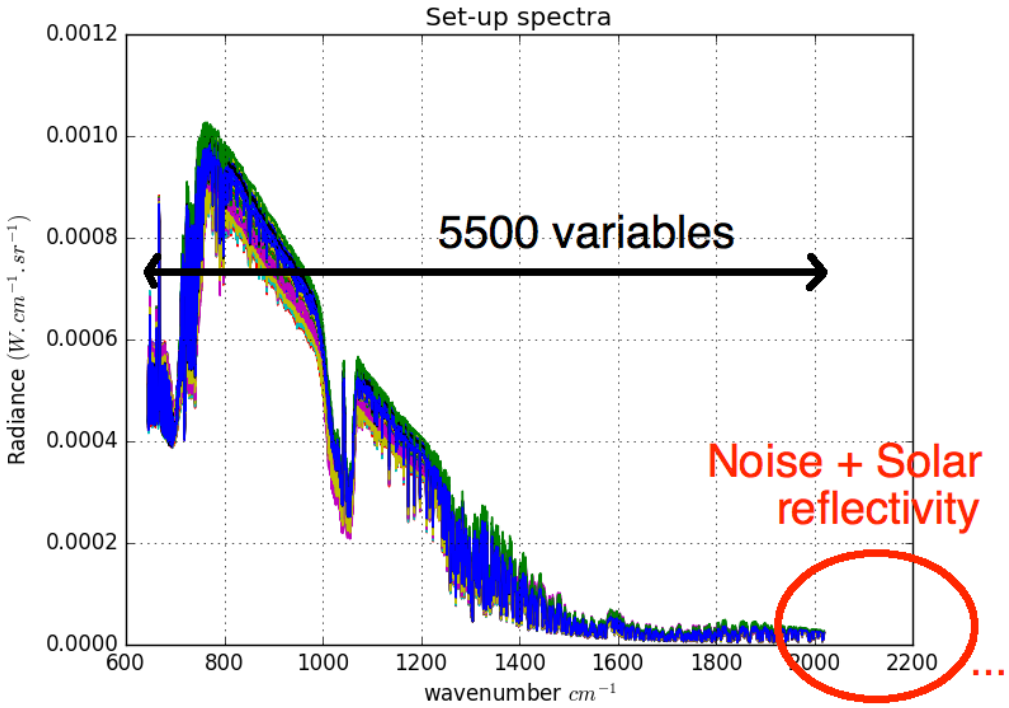


Methodology – Mathematical development

- Place a set of N observations: IASI pixels of P observables (variables): IASI channels into a matrix $\mathcal{A}^P_N$
- Build a variance-covariance matrix
- Take m eigenvectors and eigenvalues (orthogonalisation).
- Projection of the observations on the subspace (in order of variances so in order of eigenvalues) created by eigenvectors.
- Visualisations on 2D/3D-Euclidians subspaces.

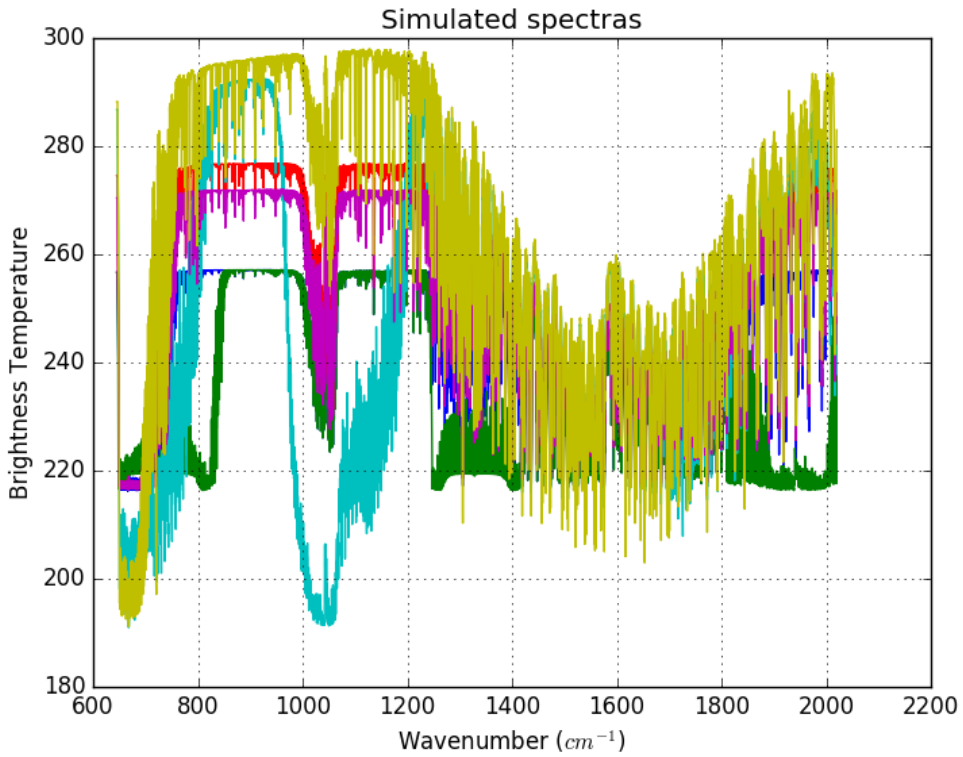
How to build a PCA for atmospheric data? A "Clear-sky" PCA

An IASI observation is composed of 8640 variables (**channels**).
For our works, only 5500 channels are taken.



How to build a PCA for atmospheric data? A "Clear-sky" PCA

To compute an usefull and "unbiased" PCA, we need to process:



1. Simulation of n clear-sky spectra with respect of vertical profiles by a Line-by-line radiative transfert code ([Dubuisson et al. (JQRST, 1996)], [Villerot (2015): ARAHMIS]).
   $\surd$ 4 latitudes bands
   $[+60; +90]$, $[+30; +60]$, $[-30; +30]$, $[-60; -30]$
   $\surd$ All seasons

How to build a PCA for atmospheric data? A "Clear-sky" PCA

To compute an usefull and "unbiased" PCA, we need to process:

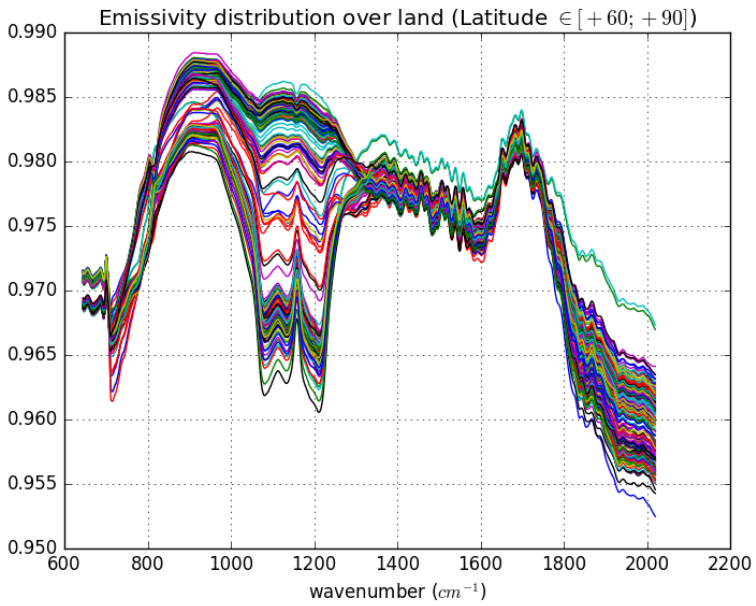
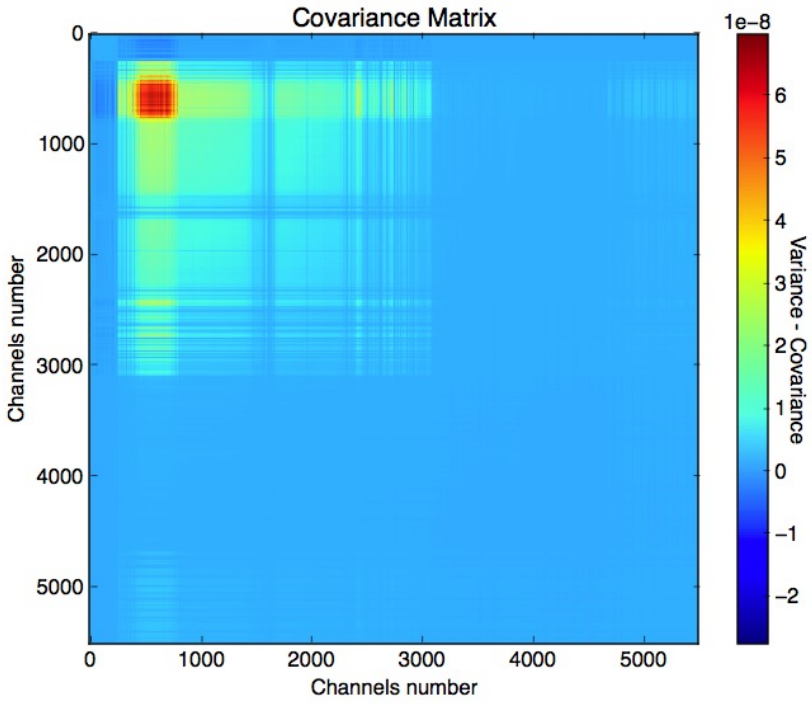


Figure: From Daniel K. Zhou (NASA) 2011-2015

1. Simulation of n clear-sky spectra with respect of vertical profiles by a Line-by-line radiative transfert code .
   - $\surd$ 4 latitudes bands
   - $\surd$ All seasons
2. Evaluation of emissivities (From *D. Zhou [2011-2015]*) of N pixel uniformly distributed in:
   - $\surd$ Land and sea
   - $\surd$ Global coverage

How to build a PCA for atmospheric data? A "Clear-sky" PCA

To compute an usefull and "unbiased" PCA, we need to process:

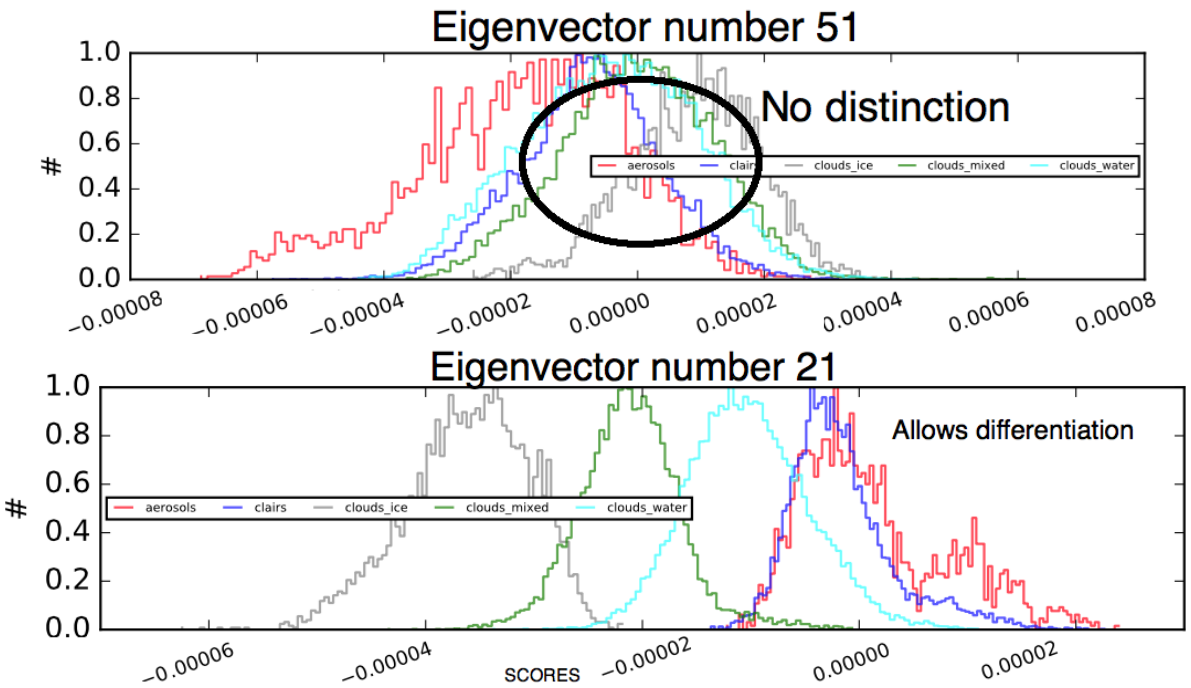


1. Simulation of n clear-sky spectra with respect of vertical profiles by a Line-by-line radiative transfert code .
   - √ 4 latitudes bands
   - √ All seasons
2. Evaluation of emissivities (From *D. Zhou [2011-2015]*) of N pixel uniformly distributed in:
   - √ Land and sea
   - √ Global coverage
3. Creation by latitudes of the covariance matrix, eigenvectors, and the mean clear sky spectrum.

Projections

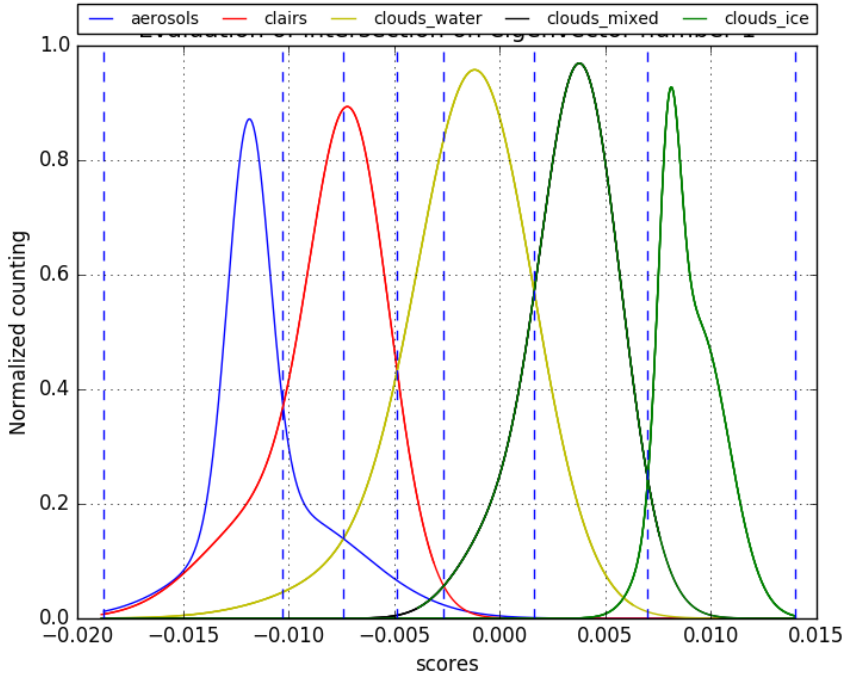
$$p = E^T(y_i - \bar{y})$$

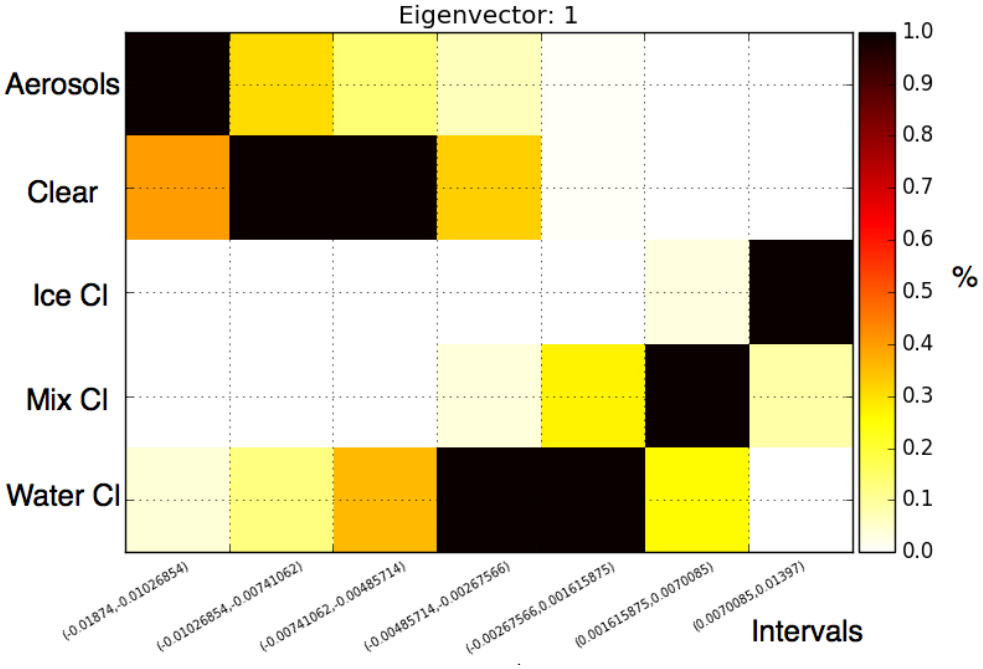
where E is the first 150 eigenvectors based on clear sky covariance matrix, y_i are the measured spectra and $\bar{y}$ the mean clear sky spectrum.
$\hookrightarrow$ **A-priori on data types given by a MODIS split-Window treatment.**



Projections show different distributions nearly Gaussian.

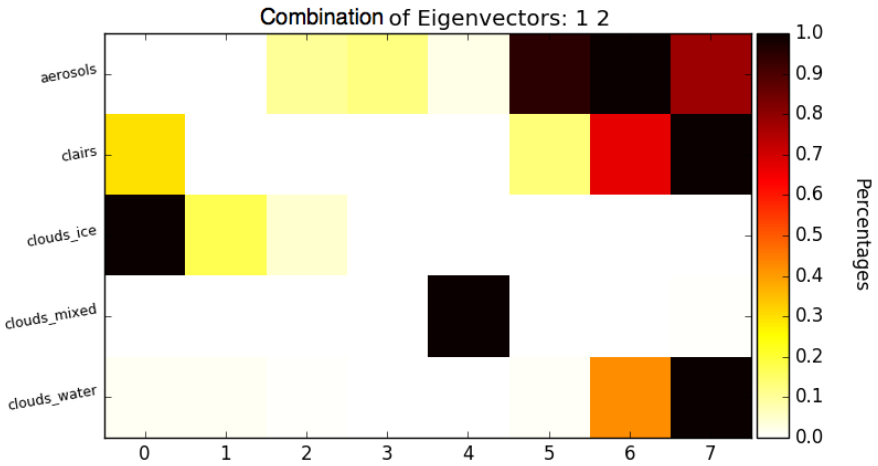
Gaussian algorithm: Baye's theorem

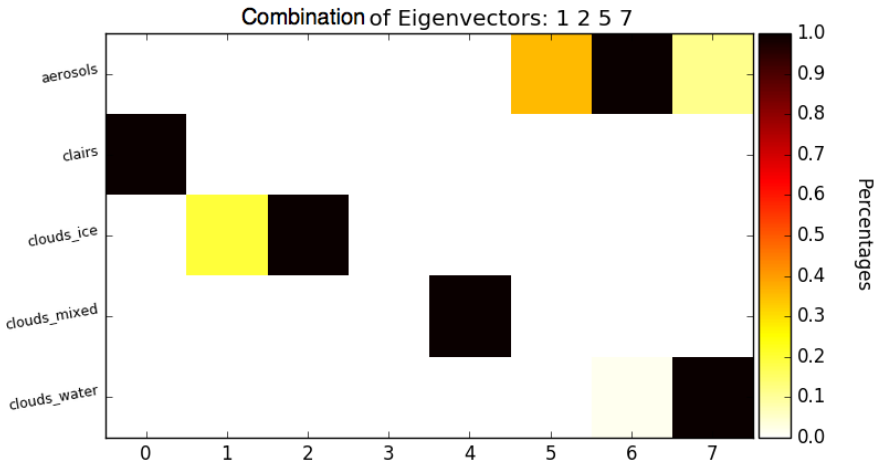


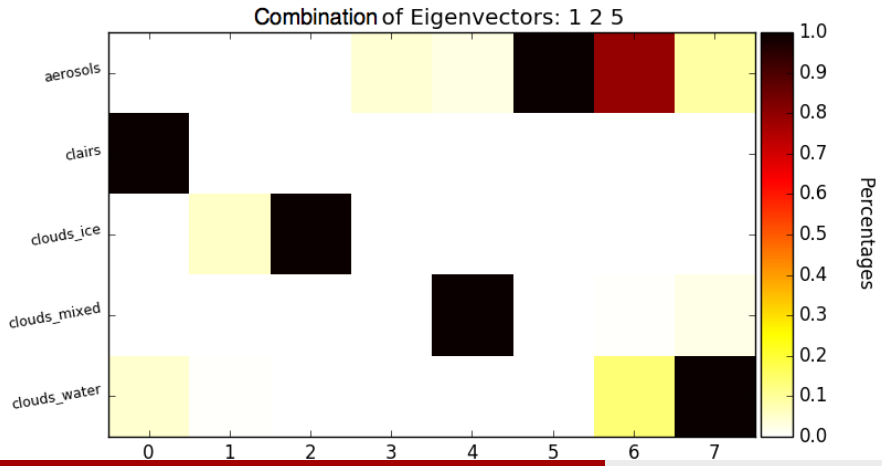


- Build intervals via intersection of all Gaussian distributions (i.e for each type)
- Evaluation (weighted integrals) of the percentage of "presence".
- Visualisation on "Baye's Matrices"
- Selection can be made on intervals with the higher % of presence.

Combinations of probability of presence



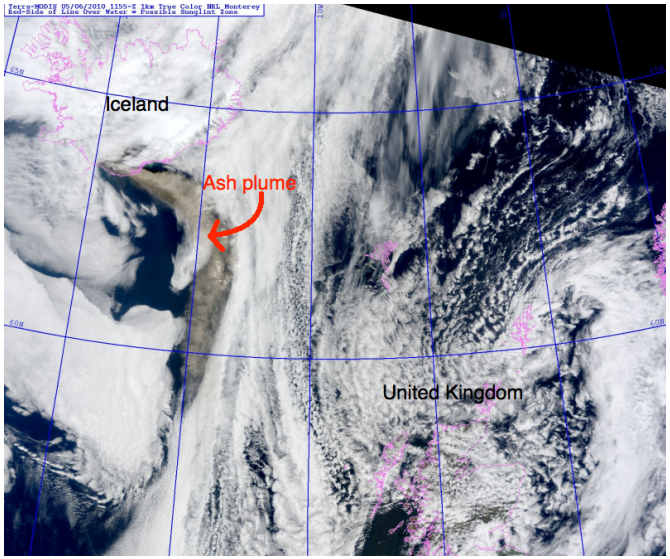


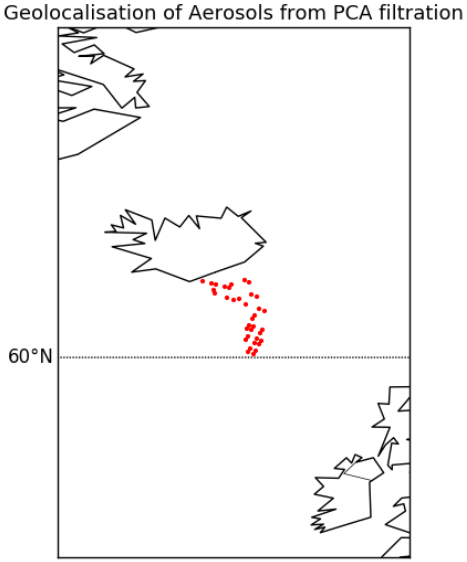


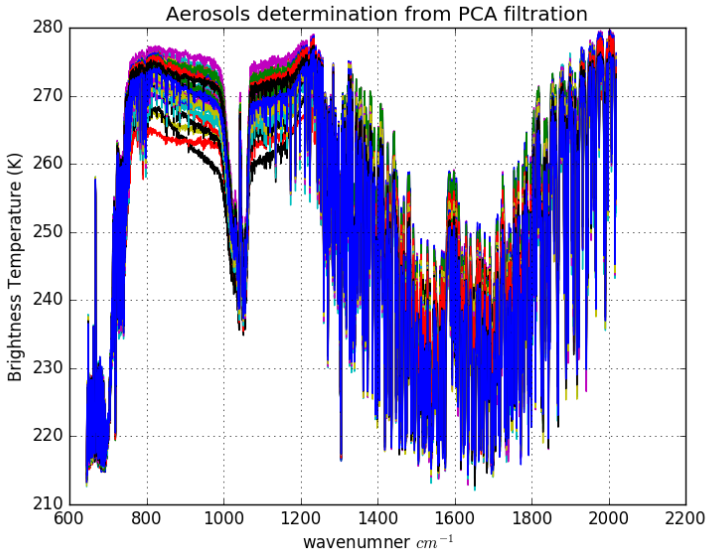
- take the best of all combinaisons.
- 4/5 eigenvectors only are needed
- Improve retrieval rate

Volcanic eruption: Eyjafjoll May, 06th 2010

Geographical repartition of Aerosols pixels from the eruption.

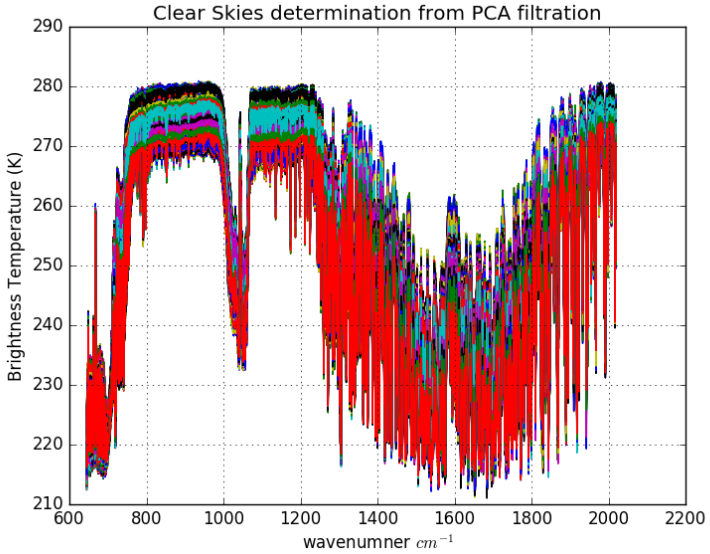


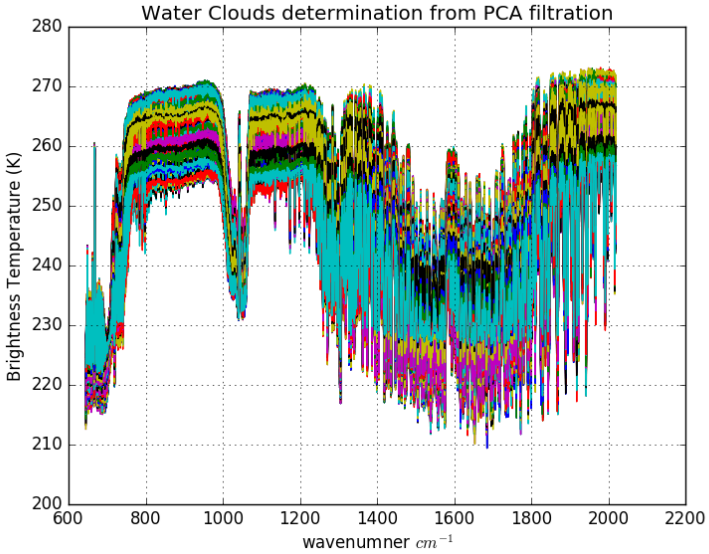


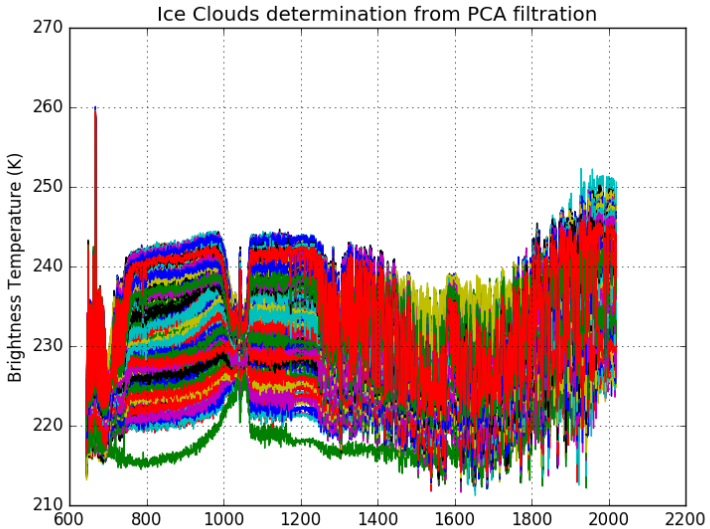


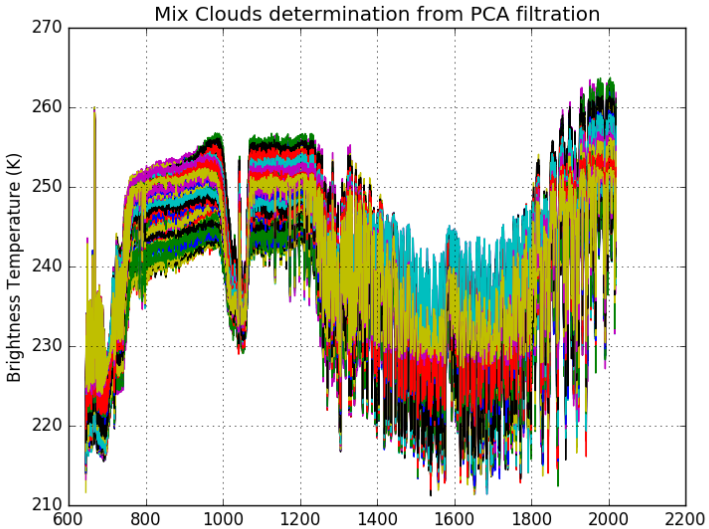
Figures show quasi-identical shape of the plume.
Coherent spectral signature with aerosols.

Volcanic eruption: Eyjafjoll May, 06th 2010









Volcanic eruption: Eyjafjoll May, 06th 2010

- PCA allows a distinction of 97%. Almost 3% are not taken.
- Criteria are based from the global PCA.
- Time processing $<$ 10 secondes for this scene (5000 spectra)

Comparison of aerosols: Dust and Ash

PCA allows to distinguish the nature of aerosols.

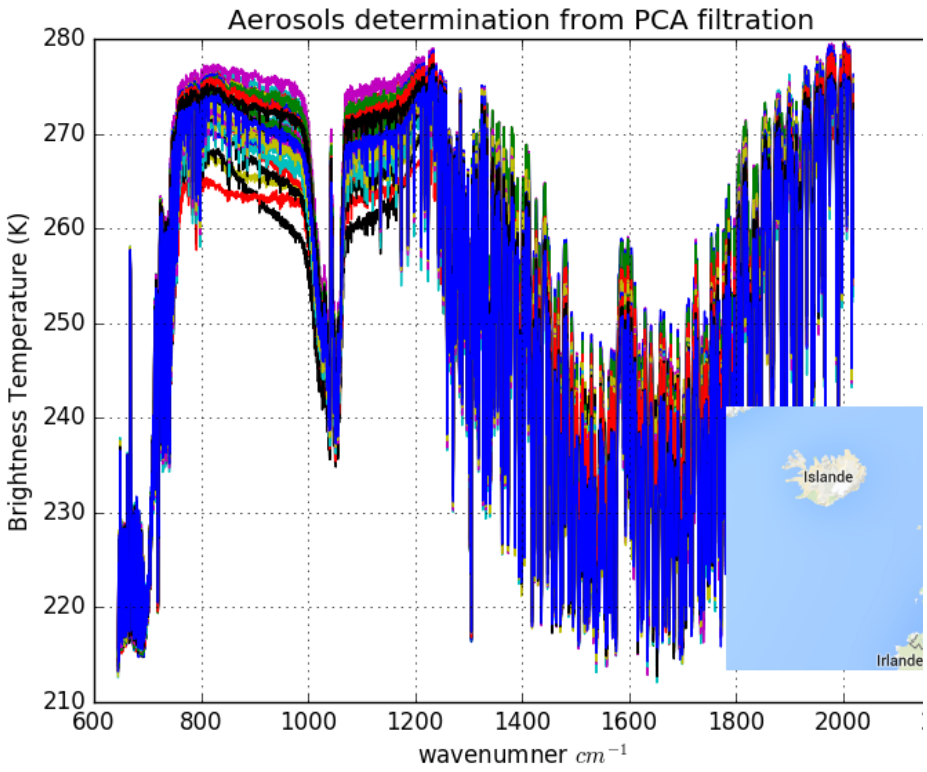


Figure: Spectral distribution of aerosol from Eyjafjöll

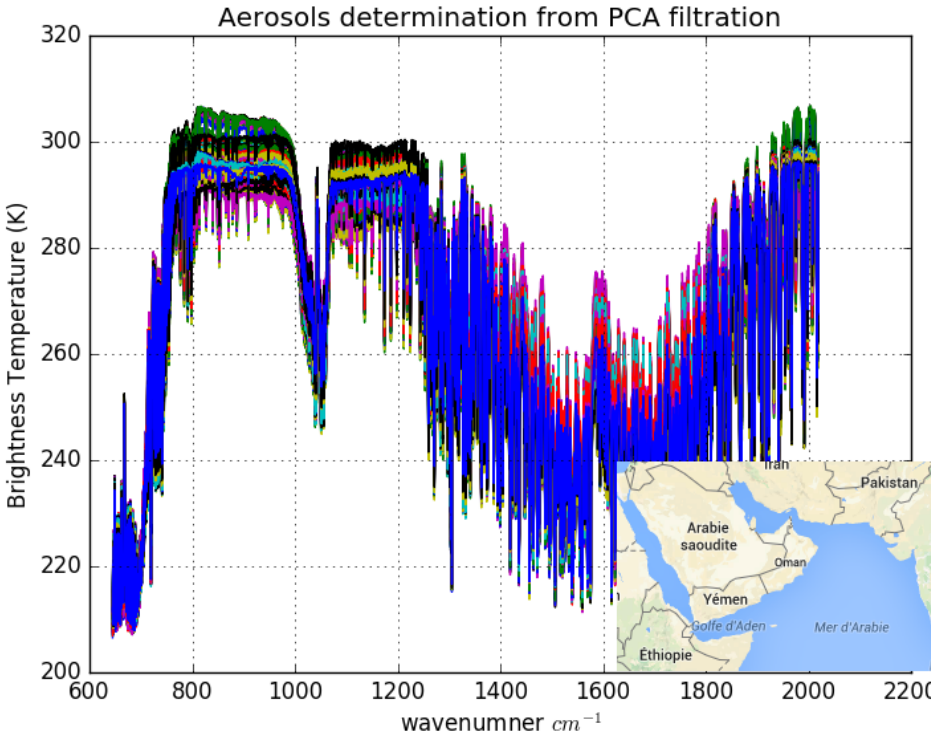


Figure: Spectral distribution of aerosol from Saoudia

It shows to be independent of the surface emissivity.

A multiple aerosols emission : April, 24th 2015

Aerosols treatment at global scale

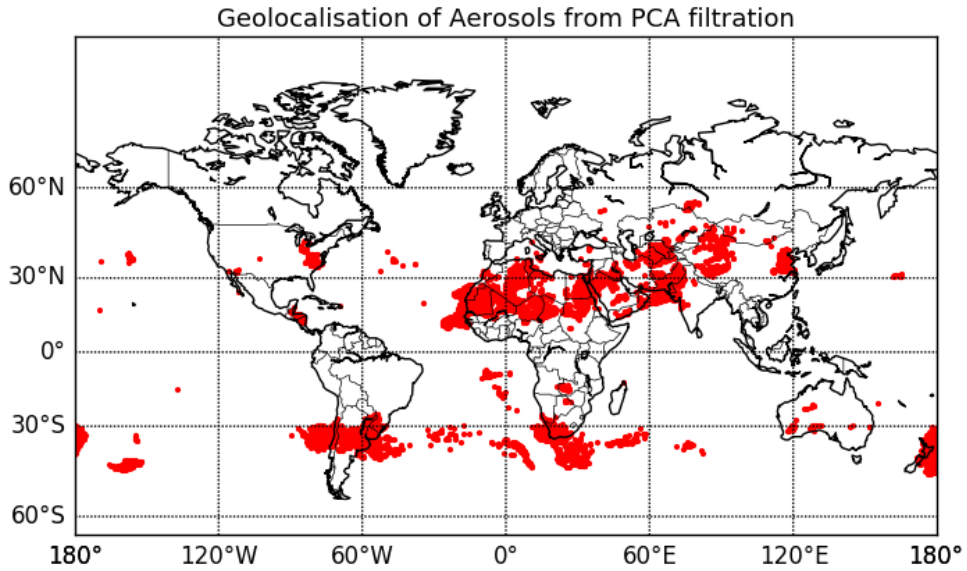


Figure: Aerosols distribution at global scale

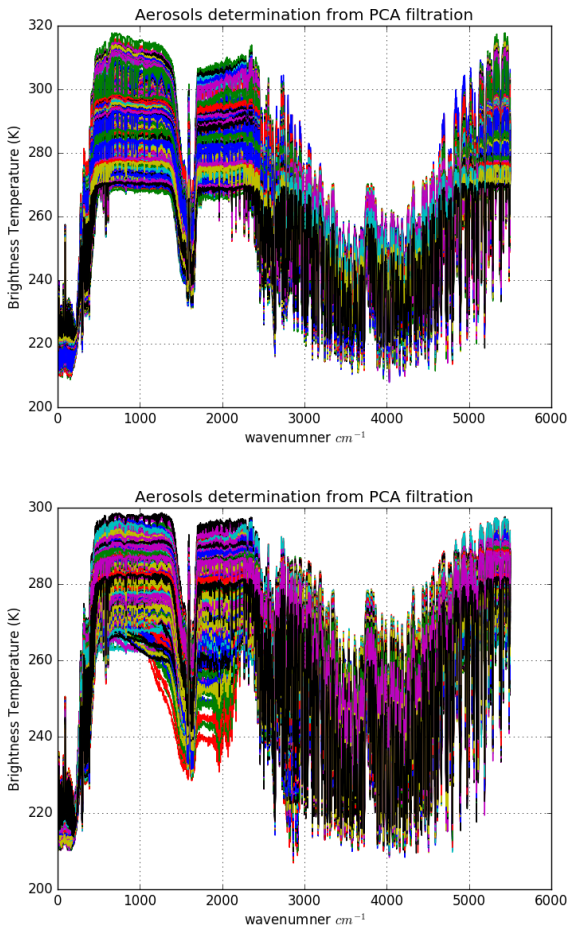


Figure: Spectral signature of aerosols on latitudes $[+30, +60]$ (top) and $[-60, -30]$ (down)

A multiple aerosols emission : April, 24th 2015

Aerosols treatment at global scale

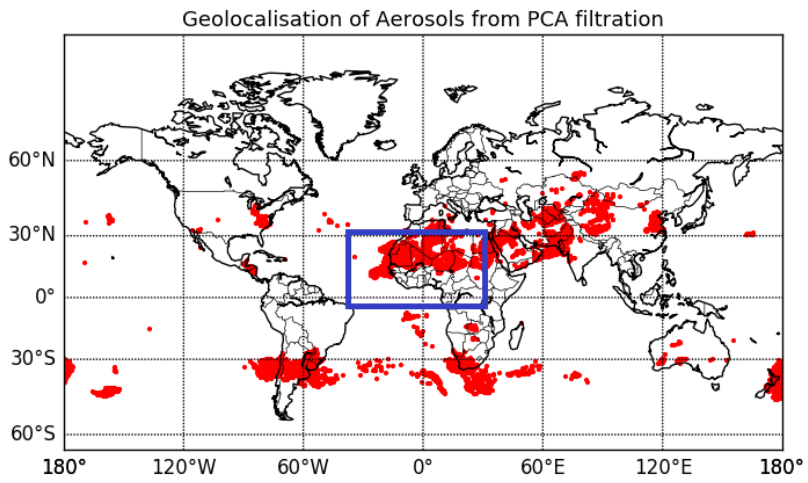
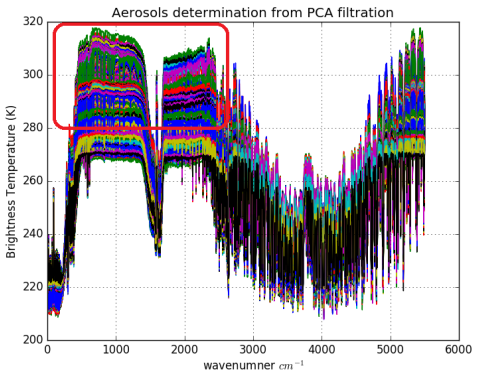


Figure: Aerosols distribution at global scale



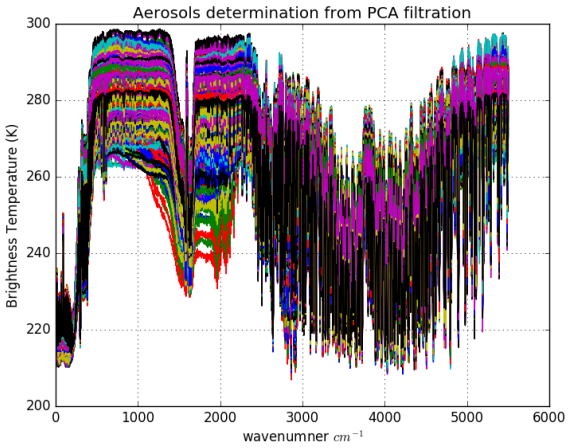


Figure: Spectral signature of aerosols on latitudes $[+30, +60]$ (top) and $[-60, -30]$ (down)

A multiple aerosols emission : April, 24th 2015

Aerosols treatment at global scale

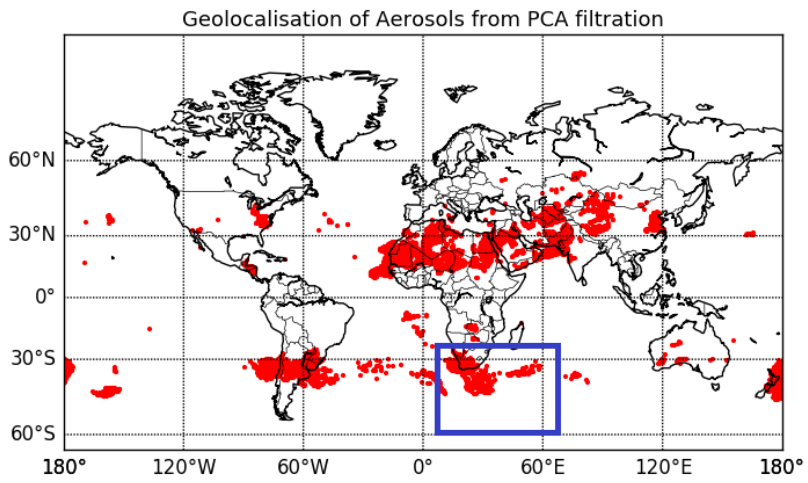


Figure: Aerosols distribution at global scale

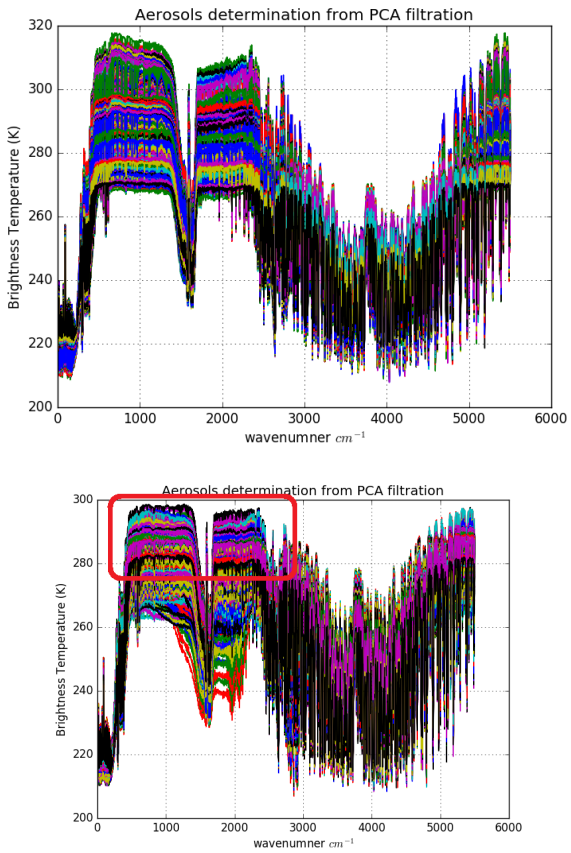


Figure: Spectral signature of aerosols on latitudes $[+30, +60]$ (top) and $[-60, -30]$ (down)

A multiple aerosols emission : April, 24th 2015

Aerosols treatment at global scale

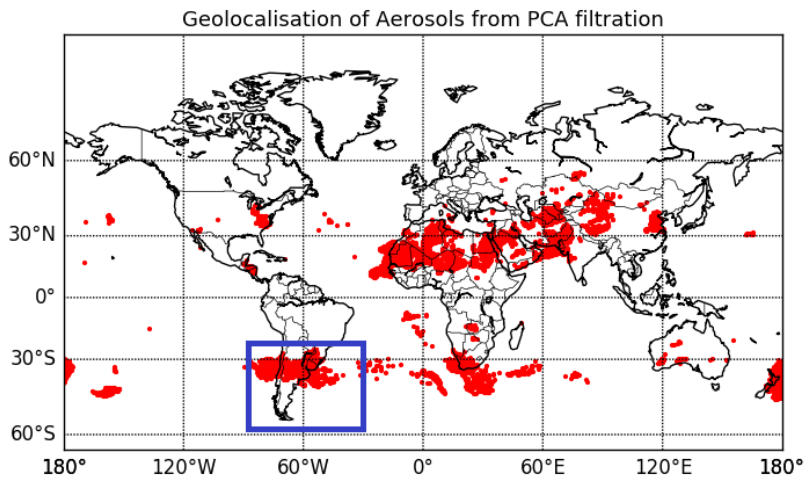


Figure: Aerosols distribution at global scale

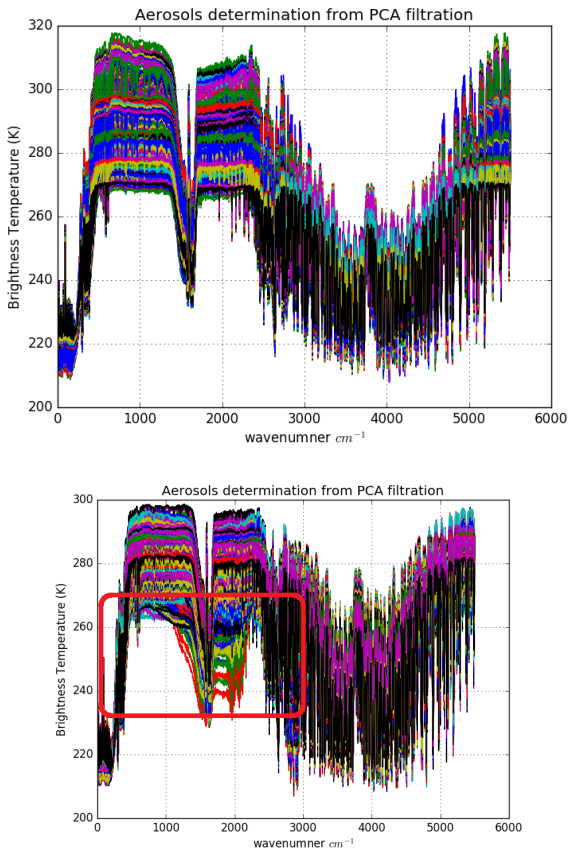


Figure: Spectral signature of aerosols on latitudes $[+30, +60]$ (top) and $[-60, -30]$ (down)

A multiple aerosols emission : April, 24th 2015

Ice clouds treatment at global scale

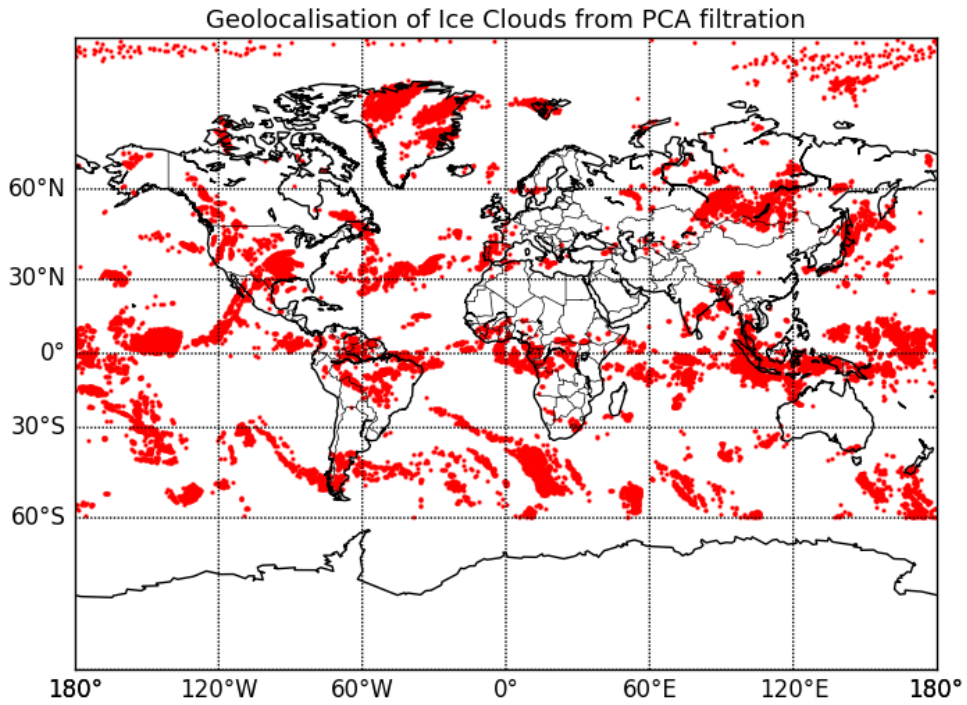


Figure: Ice clouds distribution at global scale

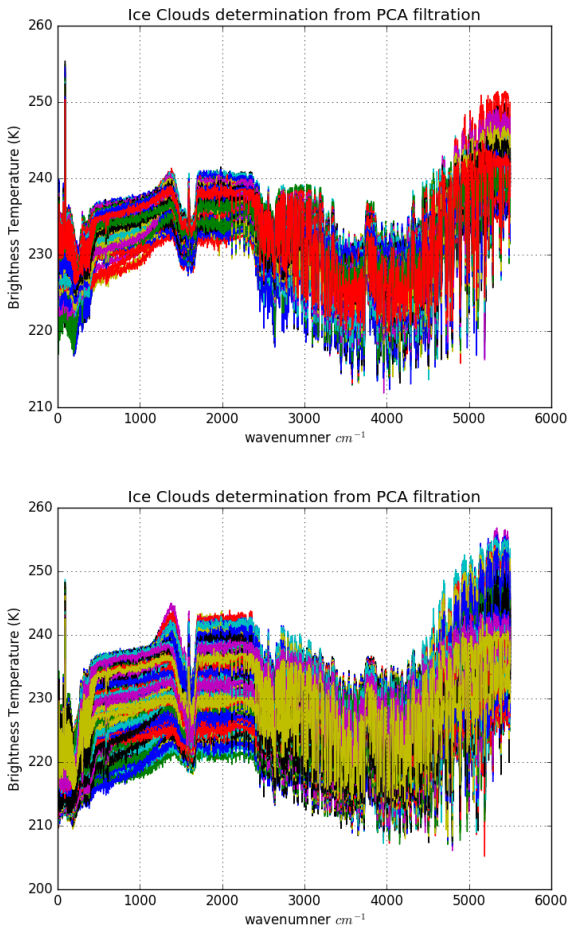


Figure: Spectral signature of ice clouds on latitudes $[+60, +90]$ (top) and $[-60, -30]$ (down)

Conclusions :

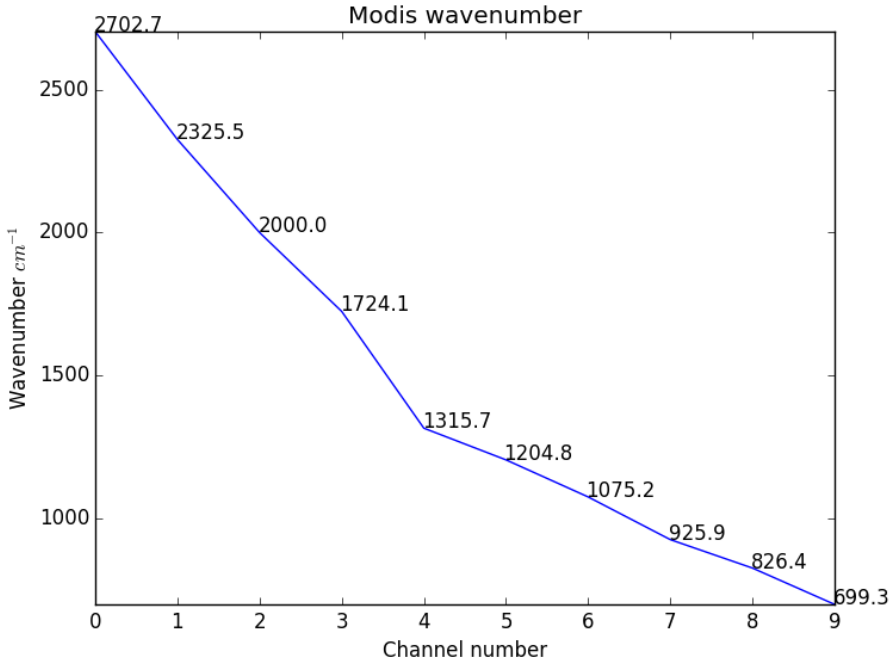
- The PCA is a powerfull tool to pre-characterize observations in scattering.
- Very few false-detections ($< 5\%$) and low time processing (< 10 min) for a treatment of all the spectra recorded during 24h at global sclae (10^6 spectra).
- Suitable for forward models and assimilation processes.

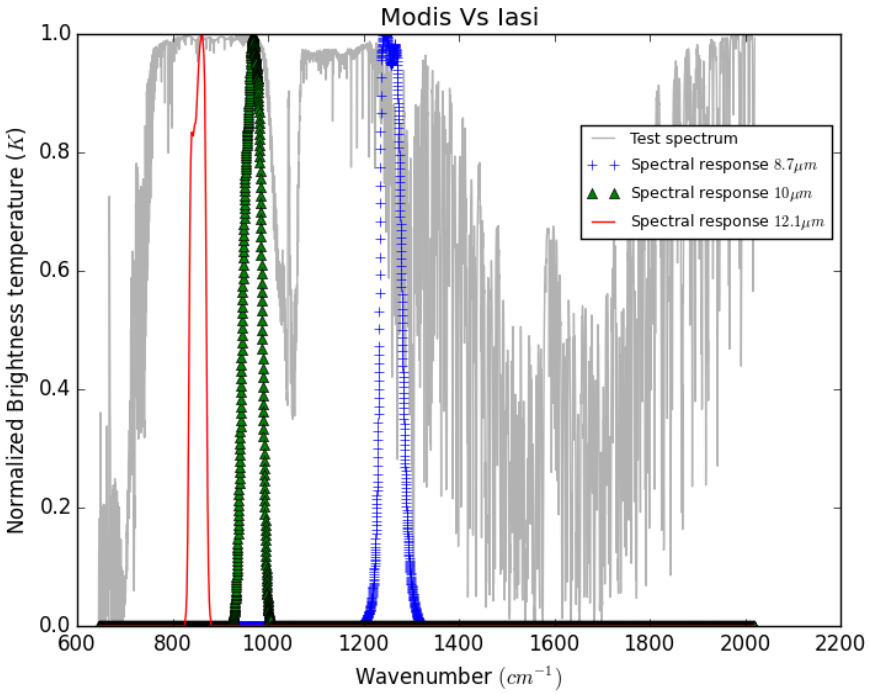
Short-time perspectives:

- Go further with a cross-comparison with ECMWF center.
- Investigation on types of aerosols (PCA aerosols), trace gases and short live species.

THANKS FOR YOUR ATTENTION!

Appendices : Modis split-window Algorithm





- Selection of 3 spectral bands $(8.7, 10, 12\mu m)$
- weighting of "Emissivity-corrected spectra" by Modis spectral responses.
- Selection of types by analysing Difference Brightness Temperatures ($\Delta(10, 8.7)$, $\Delta(12, 10)$, $\Delta(12, 8.7)$)

Appendices : Why "clear" PCA ??

A clear spectrum depends of :

- Temperature
- Pressure
- surface emissivity

A non-clear (aerosols or clouds) spectrum depends of:

- Temperature
- Pressure
- surface emissivity
- particule size distribution
- granulometry
- altitude, ...

With a good estimation of emissivity, we can obtain the quasi-total clear sky variability.