Unsupervised Machine Learning Algorithms to Detect CO2 Clouds on Mars

Vincent CAILLÉ [1]

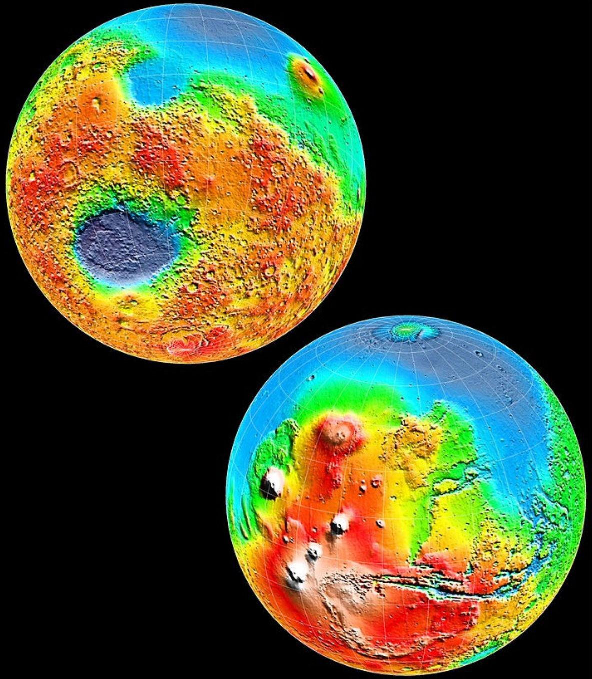
PhD Supervisors : Anni Määttänen [1] and Aymeric Spiga [2]

[1] LATMOS
[2] LMD

MOLA (Mars Orbiter Laser Altimeter)

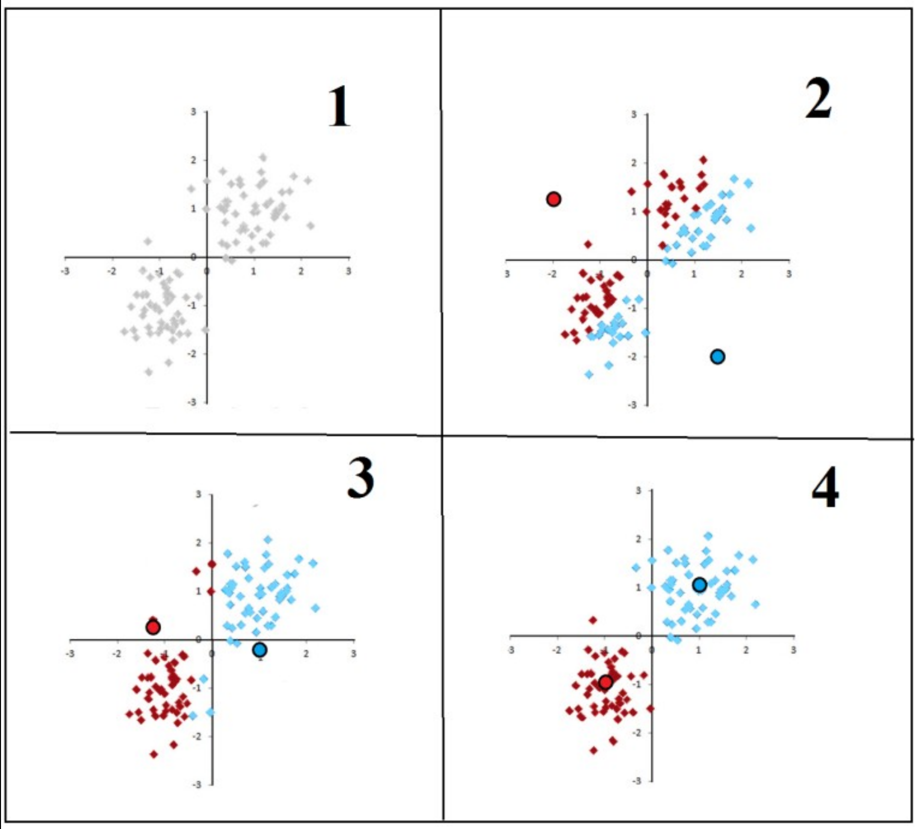
Artist concept of the MGS *(NASA)*

- Aboard Mars Global Surveyor, in Mars orbit from 1997 to 2006
- Altimeter : characterize Mars surface topography with infrared laser pulses toward Mars
- Data from MOLA until 2001,
  → huge set of data
- More sensitive than expected with a good vertical resolution allowed detection of non surface features, like clouds



MOLA topographic map of Mars hemispheres (*Science* cover 1999)

Unsupervised machine learning : k-mean method



Steps of the K-mean clustering algorithm
(cf Kalmar et al., 2012)

- Each observation point belongs to the cluster with the closest centroid (cluster mean)
- Move clusters centroids to minimize within-cluster variances

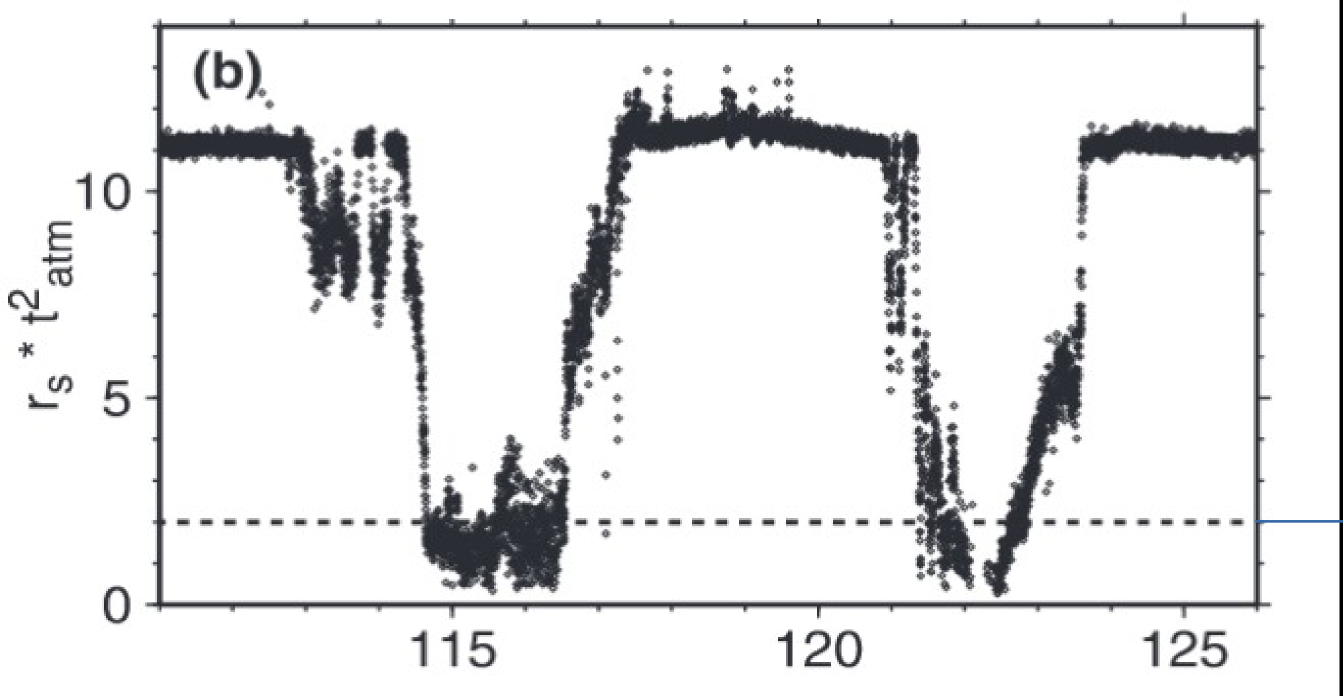
We have to determine :
- best parameters to separate clouds, noise and continuum
- best number of clusters k

Could k-mean method make a cluster of clouds from MOLA data ?

Choice of parameters to run our method with

$$E_{\mathrm{rec}} = E_{\mathrm{trans}} t_r A_r \frac{r_s}{\pi} \cdot \frac{t_{\mathrm{atm}}^2}{z_{\mathrm{Mars}}^2},$$

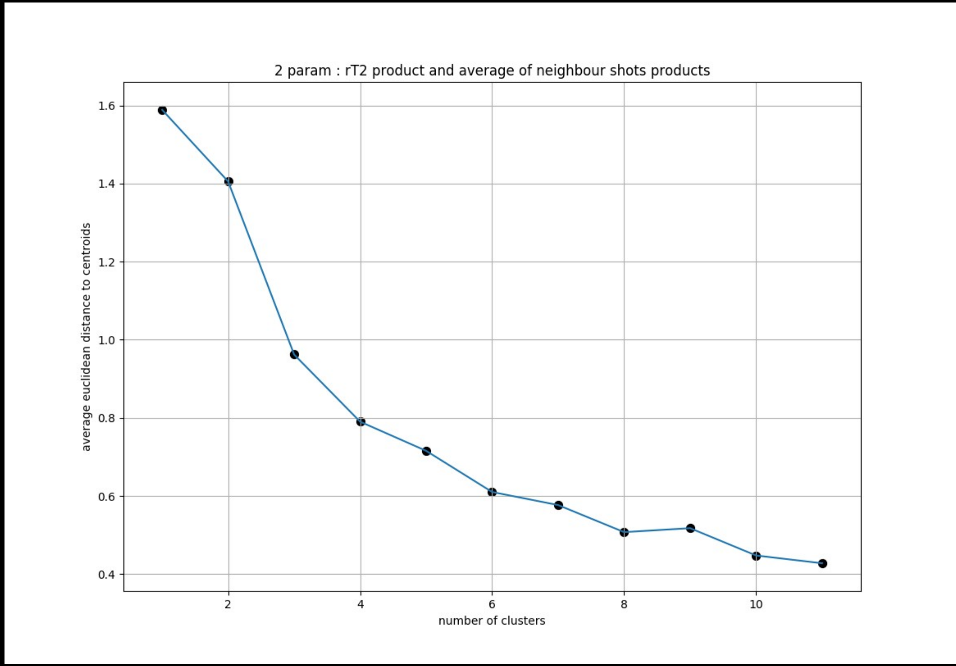
Product of surface reflectivity and two-way atmosphere transmissivity (normalized « return energy »)



Clouds example in MOLA data, rT² product against time
Cf Neumann el al., 2003

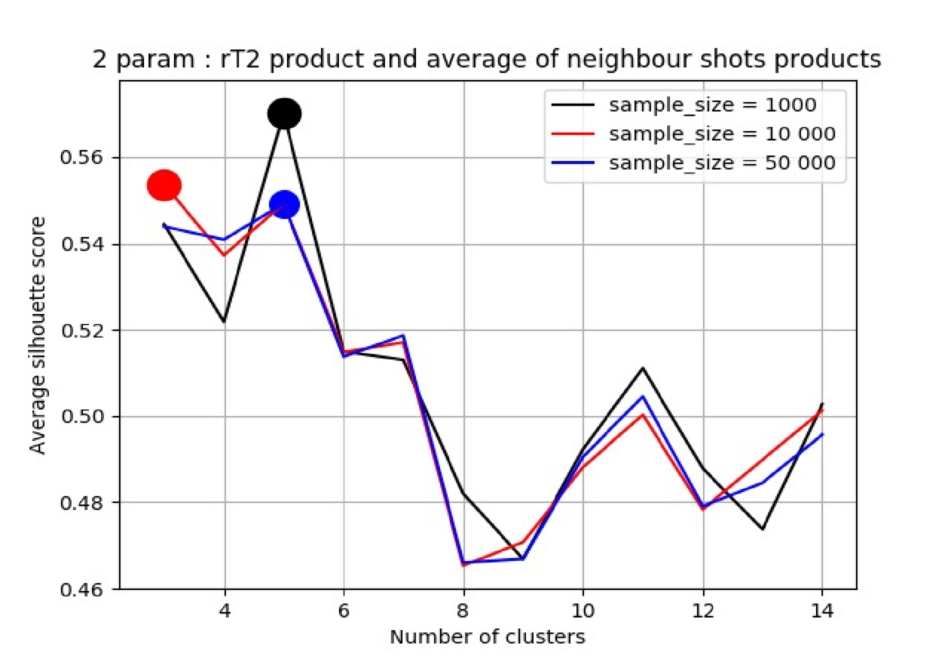
Problem : strict and limiting detection criteria

How to optimize the number of clusters k ?



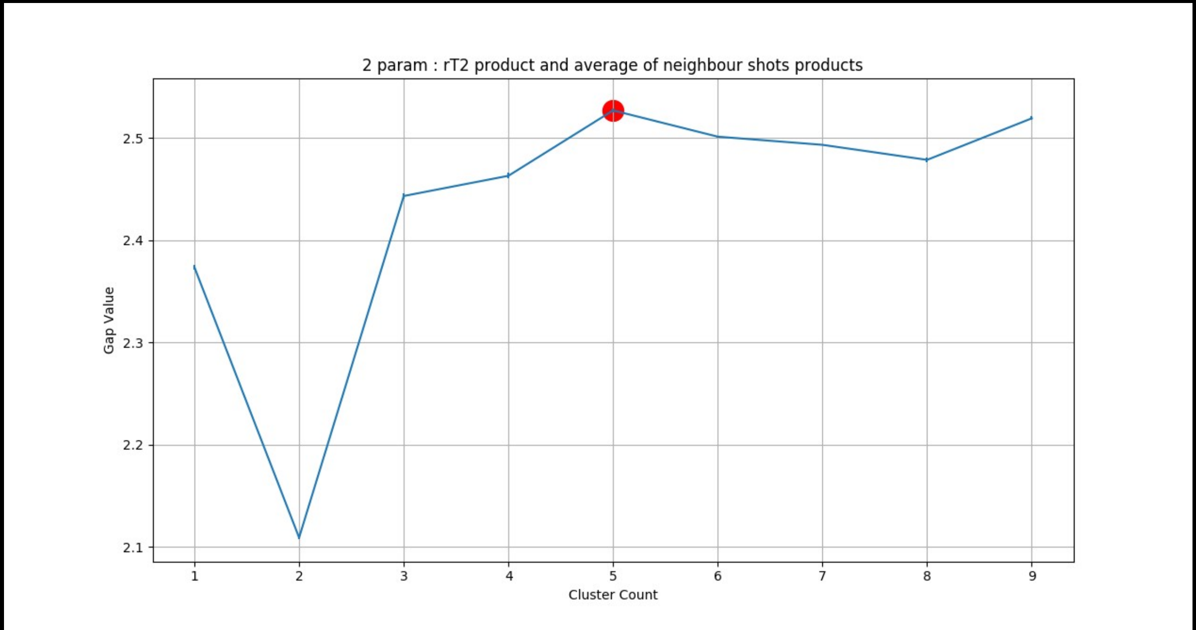
Elbow Method

Computes the total intra-cluster variation through total within-cluster sum of square



Average Silhouette Method (Kaufman et al., 1990)

Computes the « silhouette » score from distance of which point to its cluster and closest one



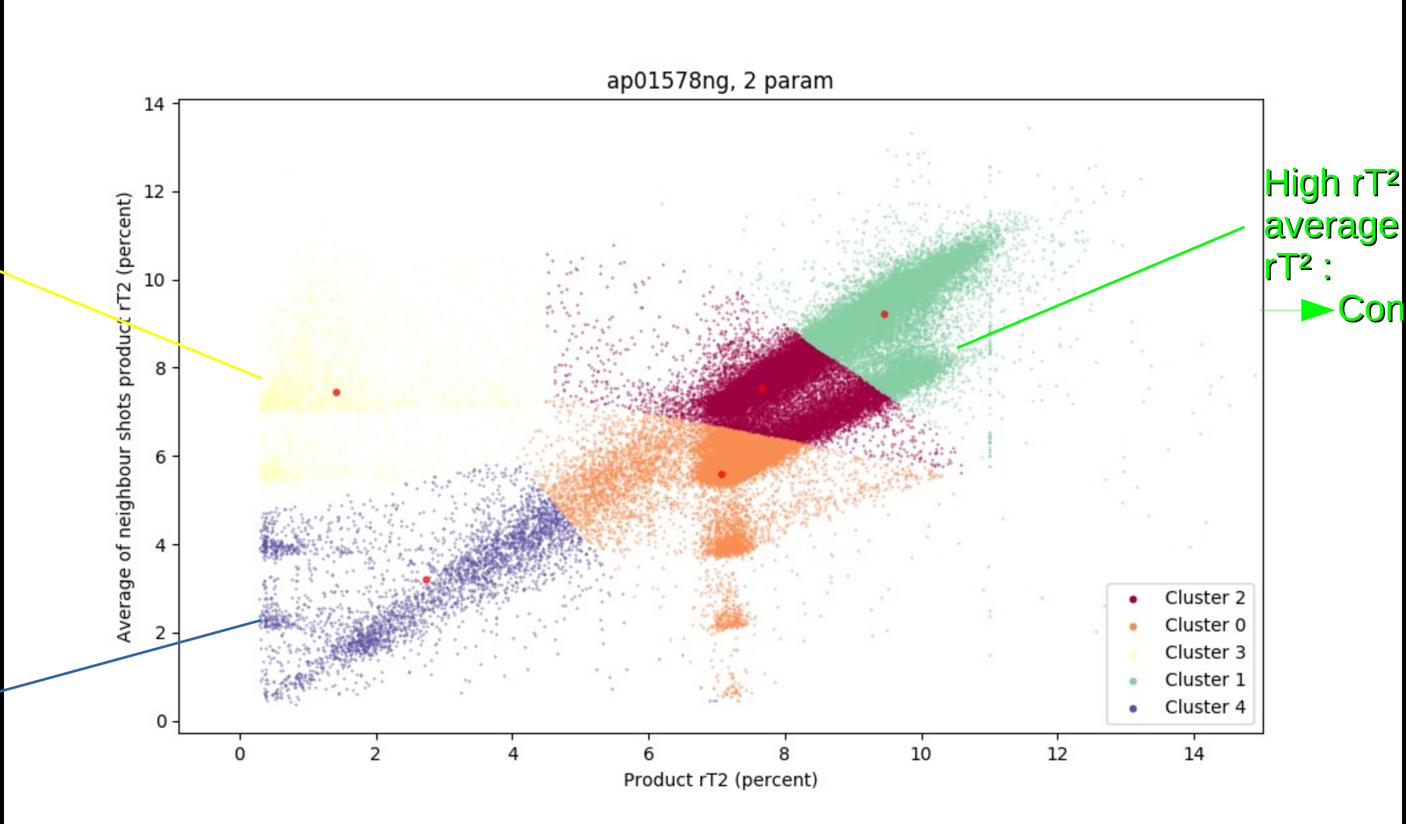
Gap Statistic Method
(cf Tibshirani et al., 2001)

Computes how far the clustering structure is from the random uniform distribution of points

Three independent methods lead to the same number of clusters k = 5

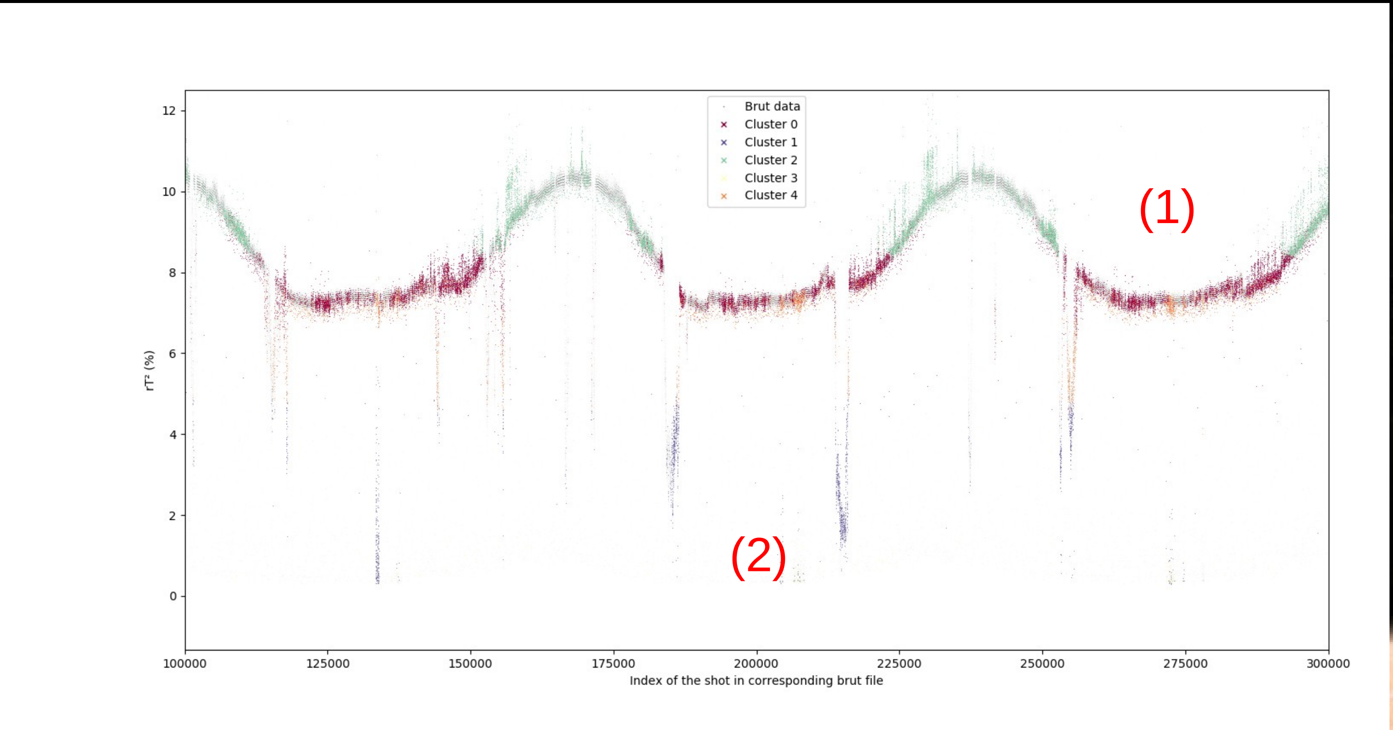
Cluster distribution for a single file of laser returns

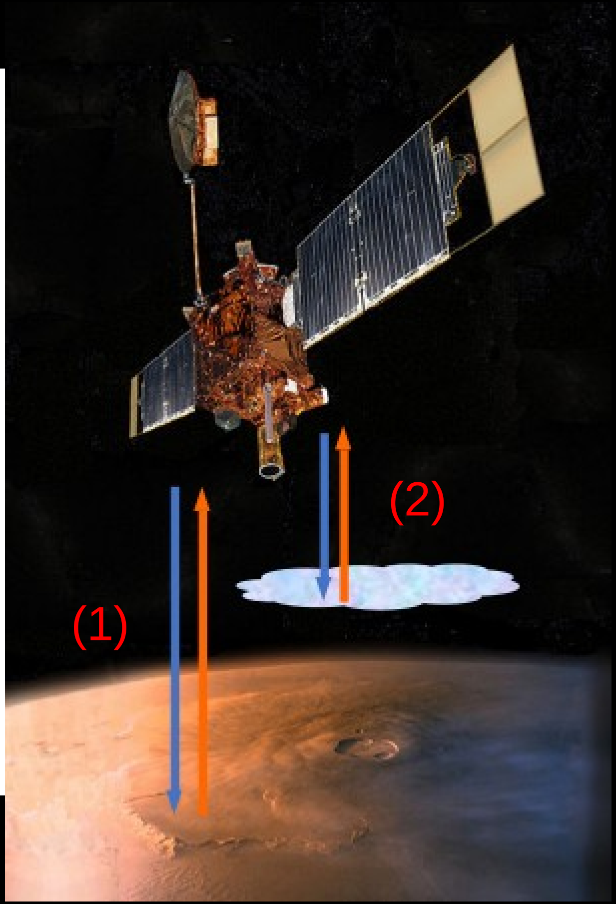
Low rT^2 but high average neighbour rT^2 : → noise ?



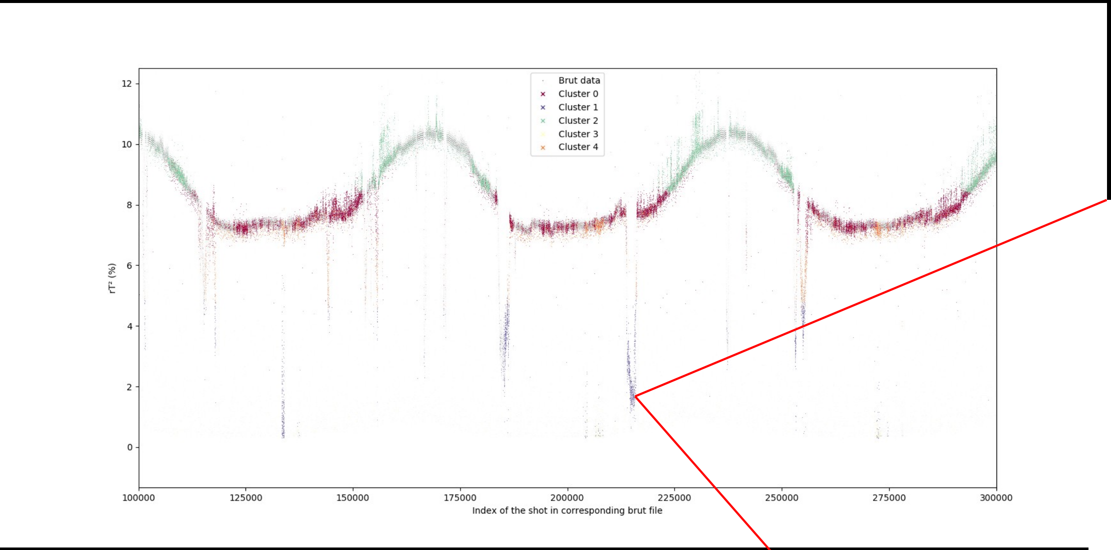
High rT^2 and high average neighbour rT^2 : → Continuum ?

Low rT^2 and low average neighbour rT^2 : → clouds ?

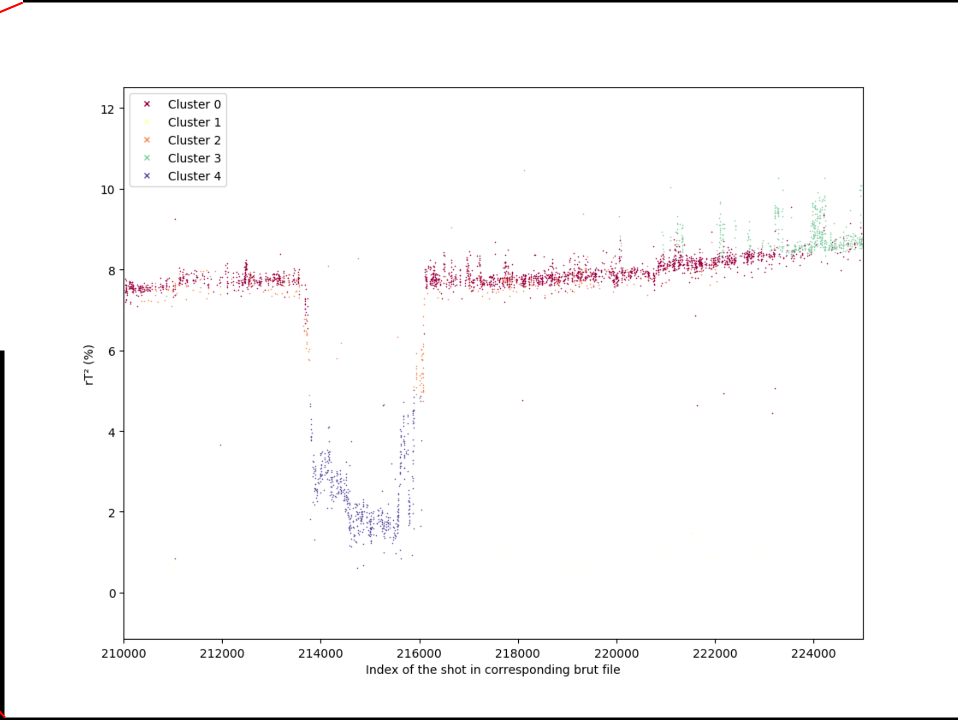




Clusters distribution in rT^2 against « time » plot



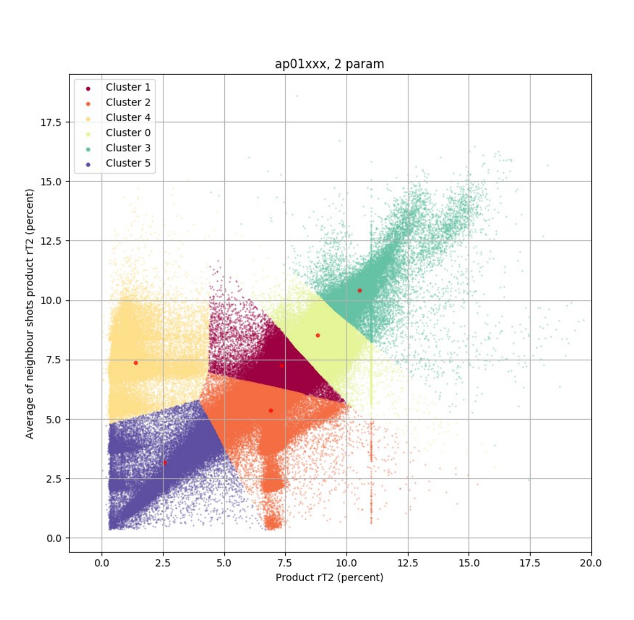
Cloud example

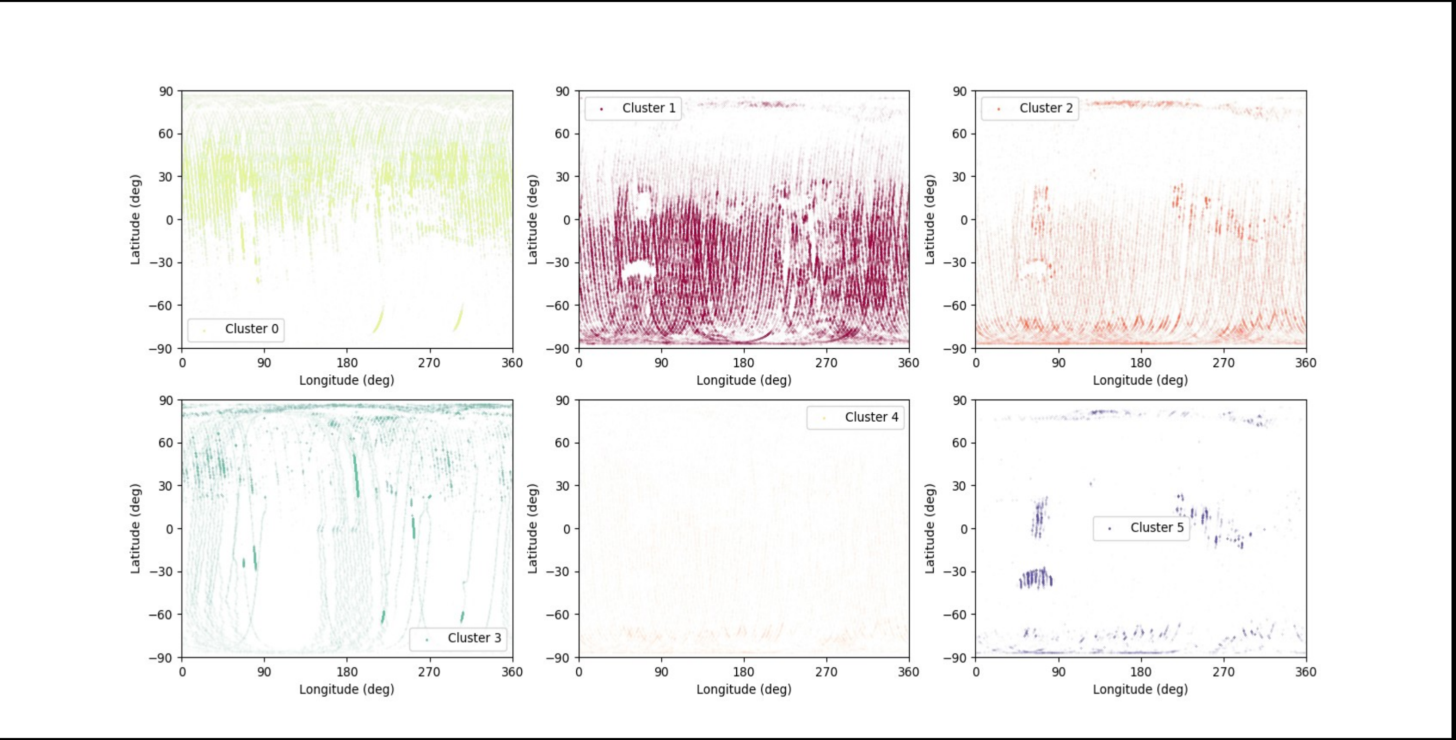


Clusters distribution in rT^2 against « time » plot

One of our clusters seems to represent clouds !

Clouds map ?





Geographical clusters distribution for ≈10 % of the whole data set

Conclusions :

- We’ve been able to find the best parameters for our clustering algorithms.
- Clustering method enabled us to distinguish continuum / noise and clouds laser returns.
- Unsupervised machine learning allows less stringent criteria than previous studies, leading to more cloud detections (and eventually a new type of cloud ?)

Perspectives :

- Plotting latitude against solar longitude for clouds returns to study the seasonal variability of clouds
- Work only with cloud returns to cluster different kinds of clouds (absorptive, reflective, CO_2, H_2O …)
- Use those observations for comparing with microphysical model results.