This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871149.

Machine Learning for automated mapping

Mario D'Amore*

German Aerospace Center (DLR)

Institute for Planetary Research (PF)

Planetary Spectroscopic Laboratories (PLL)

Rutherfordstrasse 2, 12489 Berlin

*mario.damore@dlr.de

europlanet 2024 Research Infrastructure

(Tentative) Plan

1) Brief introduction of the target (Mercury) and its exploration (presentation) : **skip if interested in data only.**

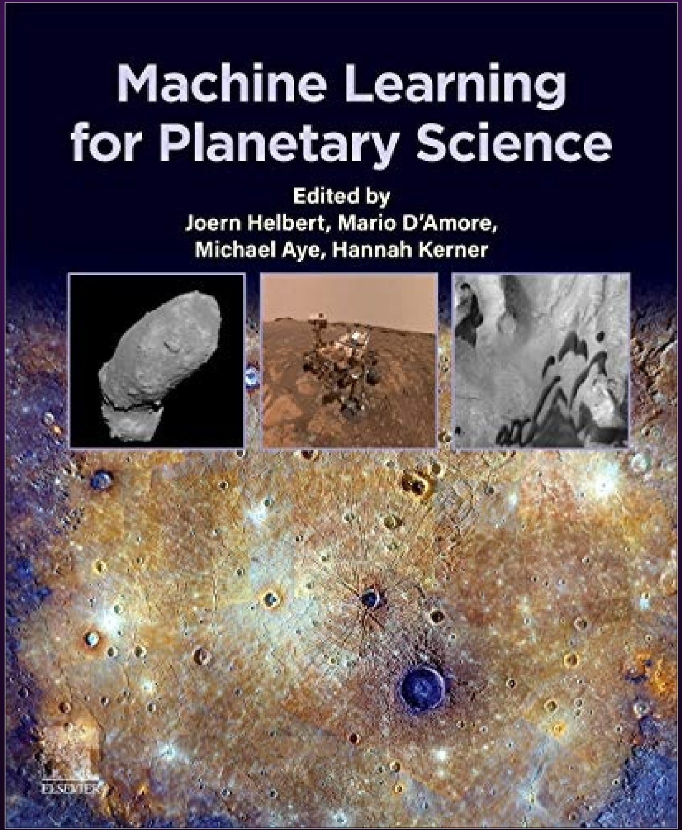
2) Explanation on how to get a datacube from hyperspectral sparse observation : not essential for the part 3, but useful.

3) Work on data : this will be a mixed presentation/notebook. We will produce the same images as in the chapter. You can run it on your own. Please open a github issue if you find some problem.

4) Comparison with other products.

All underlined text are hyperlinks, use them to explore further.

Main reference



This work is described in the chapter 7 of Elsevier book “Machine Learning for Planetary Science”, 2022.

D'Amore, Mario, and Sebastiano Padovan. "Automated surface mapping via unsupervised learning and classification of Mercury Visible–Near-Infrared reflectance spectra.".

There is some background material here not included in the book.

The code is being organised and documented thanks to Europlanet 2024 RI, funded b by the European Union's Horizon 2020 grant (No 871149)

Public Repository:

https://github.com/epn-ml/MESSENGER-Mercury-Surface-Cassification-Unsupervised_DLR/

GMAP

Geological Mapping

Table of contents

Jupyter Notebook – Needed tools/data

Tools: conda/miniconda/mamba/minimamba installed

Options :

1) run ***make create_environment*** in the root directory
2) run ***conda create --name PROJECT_NAME --file requirements.txt***
3) create an enviroment manually, activate it and run ***pip install -r requirements.txt***

Data : run the following in the root directory or download manually in ***output data/processed/***

curl https://zenodo.org/record/7433033/files/grid_2D_0_360_-90_%2B90_1deg_st_median_photom_iof_sp_2nm.png --output data/processed/grid_2D_-180_+180_-90_+90_1deg_st_median_photom_iof_sp_2nm.geojson.gz

Background – Mercury

Table 5.1 Physical and orbital Characteristics of Mercury	
Mean Distance from Earth	0.387
Period of Revolution about Earth	88 d
Period of Rotation	59 d
Inclination of Axis	28°
Equatorial Diameter	4,880
Mass (Earth = 1)	0.055
Volume (Earth = 1)	0.06
Density	5.44 g/cm3
Atmosphere (main components)	O, Na, K (thin)
Surface Temperature	100 to 700 K
Surface Pressure	0.37
Surface Gravity (Earth = 1)	0.01

Mercury is an extreme among terrestrial rocky planets: the smallest, the densest, the oldest surface, the largest daily variations in surface temperature, the least explored.

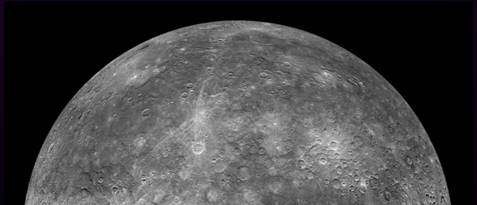
Mercury has the most eccentric orbit of all the planets : 0.21

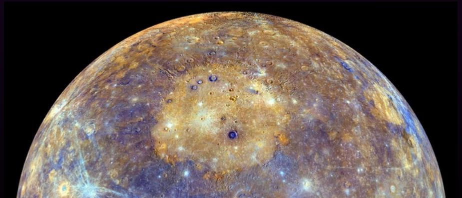
The surface is flexed by Sun induced tidal bulges are about 17 times stronger than the Moon's on Earth.

Mercury is locked in a 3:2 spin–orbit resonance.

This produces complex variations of the surface temperature.

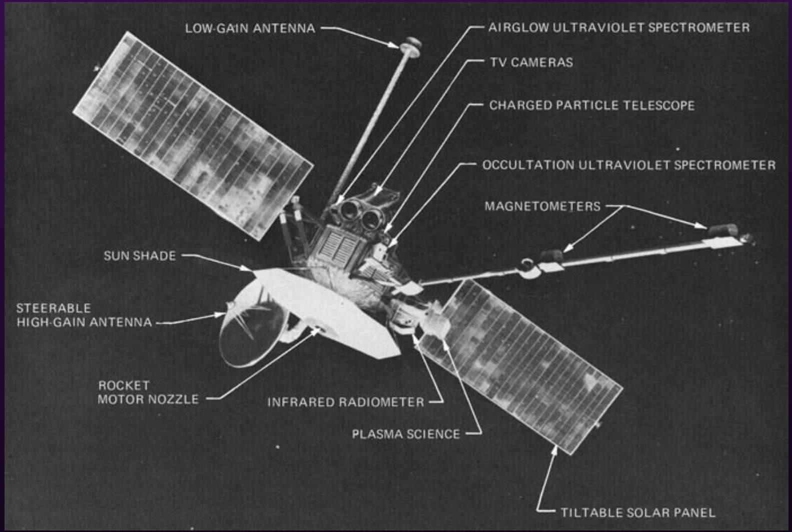
Understanding this "end member" among the terrestrial planets is crucial to developing a better understanding of how the planets in our Solar System formed and evolved.





Exploration far past : Mariner10

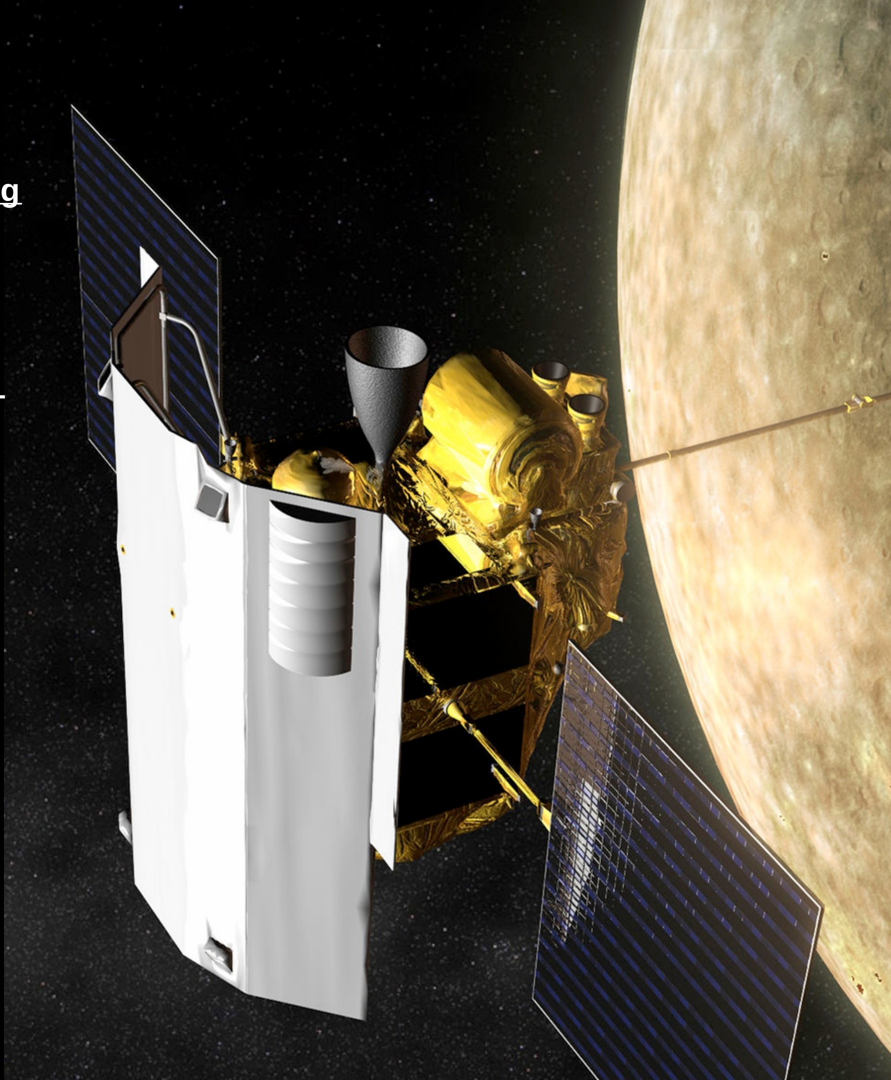
- First spacecraft to use gravitational slingshot
- 1 Venus and 3 Mercury flybys between 1974 and 1975
- Imaged ~45% of the surface
- Main findings:
  - heavily cratered surface (i.e. old) similar to the Moon
  - dominated by compressive tectonic structures
  - weak dipolar magnetic field of ~300 nT
  - thin atmosphere composed mainly by H, He and O



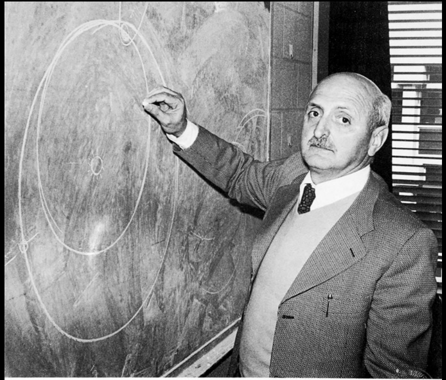
MESSENGER Mission

Mercury Surface, Space Environment,Geochemistry, and Ranging

- The second mission to reach Mercury.
- First spacecraft to orbit Mercury on March 18, 2011
- 1 Earth, 2 Venus and 3 Mercury flybys between 2005 and 2011
- Highly elliptic, near-polar orbit (200-15000 km)
- Instruments for imaging, atmospheric and surface composition, magnetic field and magnetosphere, topography and gravity
- Impacted the surface of Mercury on April 30, 2015, after low-altitude campaign



BepiColombo Mission



Giuseppe Colombo 1920 - 1984

https://sci.esa.int/web/bepicolombo/

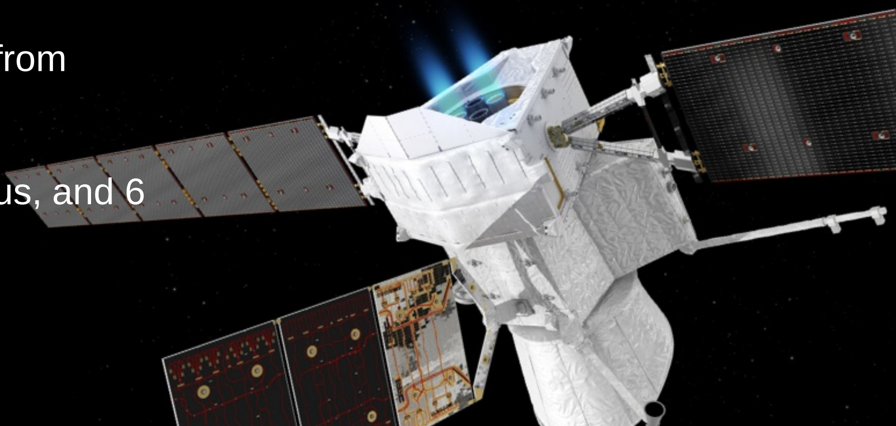
Joint mission of the European Space Agency (ESA) and the Japan Aerospace Exploration Agency (JAXA)

Two satellites launched together: the Mercury Planetary Orbiter (MPO) and Mio (Mercury Magnetospheric Orbiter, MMO).

Bepi will characterize Mercury magnetic field, magnetosphere, and both interior and surface structure.

It was launched on an Ariane 5 rocket on 20 October 2018 from Europe’s Spaceport in Kourou, French Guiana

Arrival planned for 5 December 2025, after a 1 Earth, 2Venus, and 6 Mercury flybys .



Work abstract

- **Hypothesis** : surface compositional information can be effectively derived from spectral reflectance measurements

- **Analysis** : we try to identify and characterize spectral units from all orbital observations made with MASCS during the primary mission applying :

  1) ICA for data decompostion/compression scikit decomposition

  2) Manifold learning (UMAP) to project the data in low dimensional (2D) space UMAP documentations

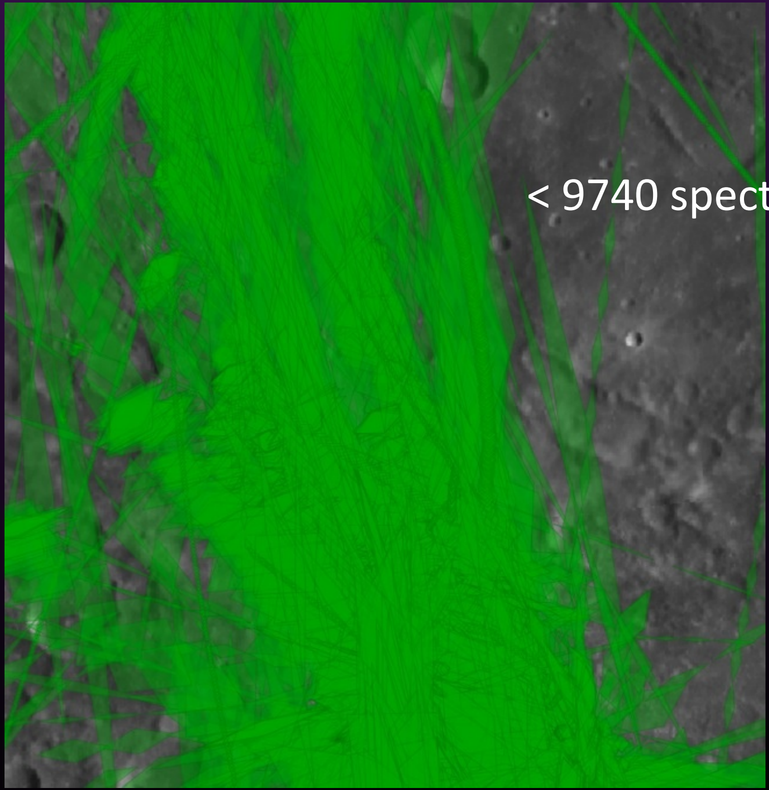
  3) hierarchical agglomerative clustering : scikit-learn clustering

  4) Comparison with other maps products from different sources.

Data extraction

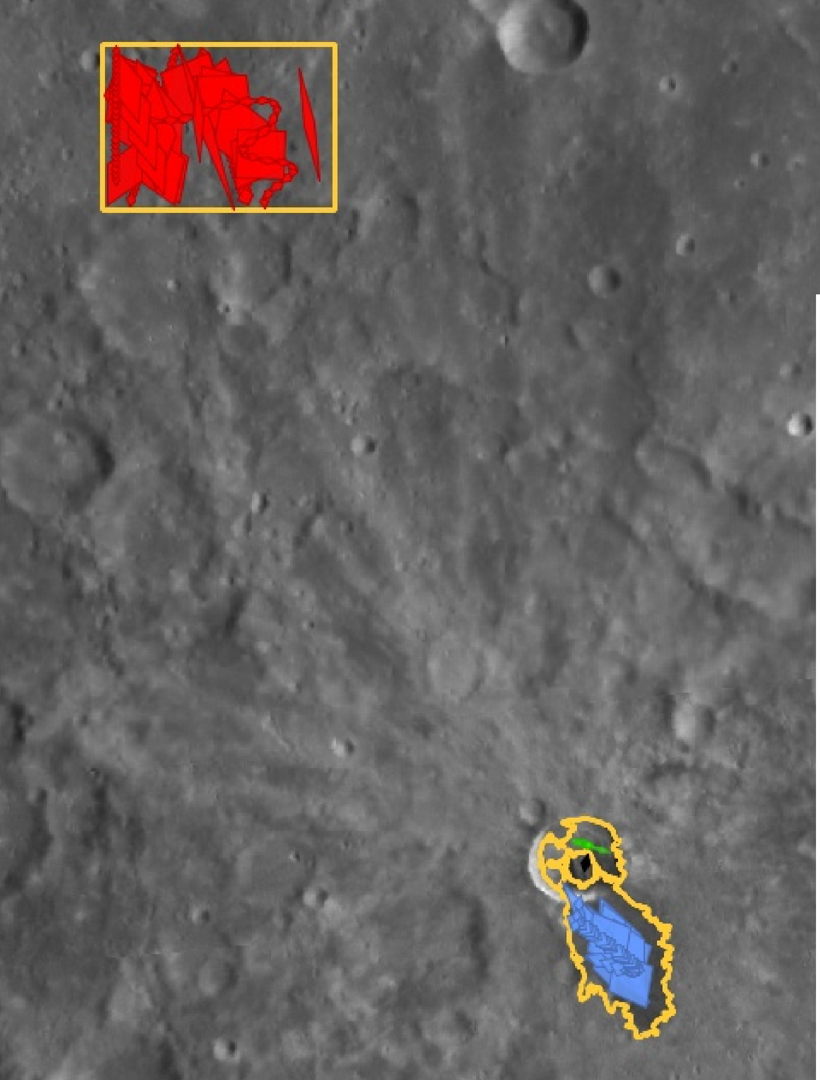
Area of Interest and Polygons definition

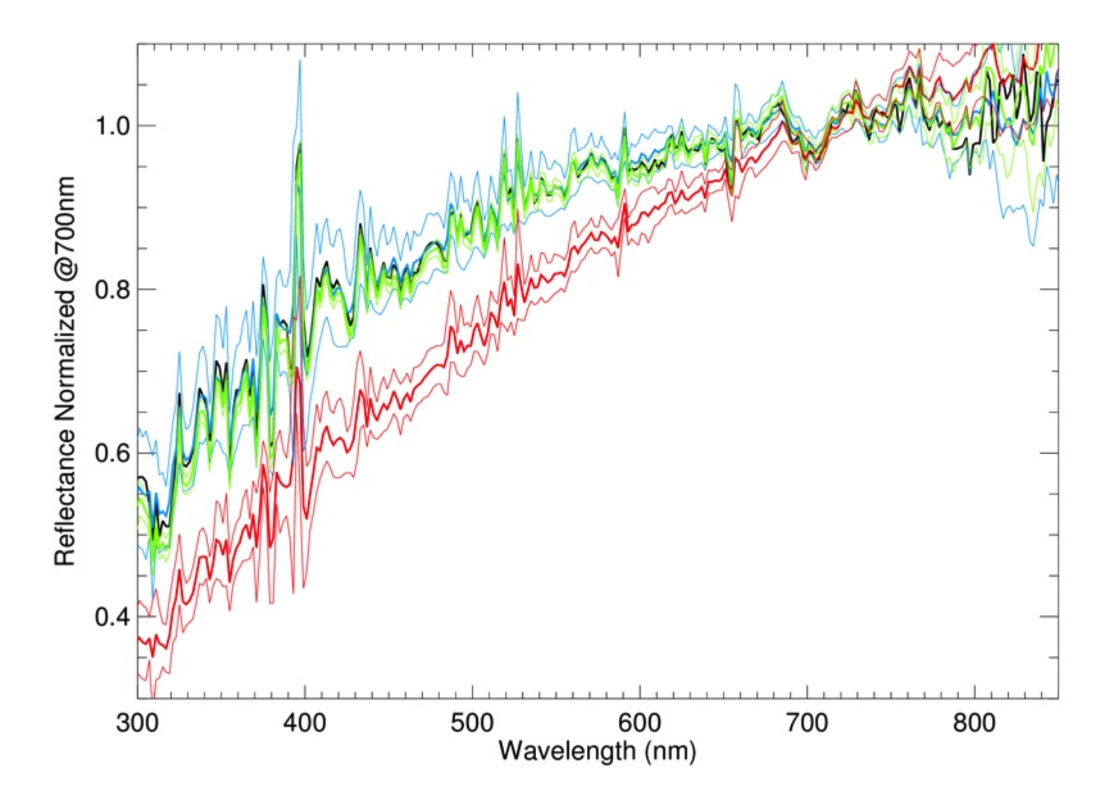
Spectra in the area



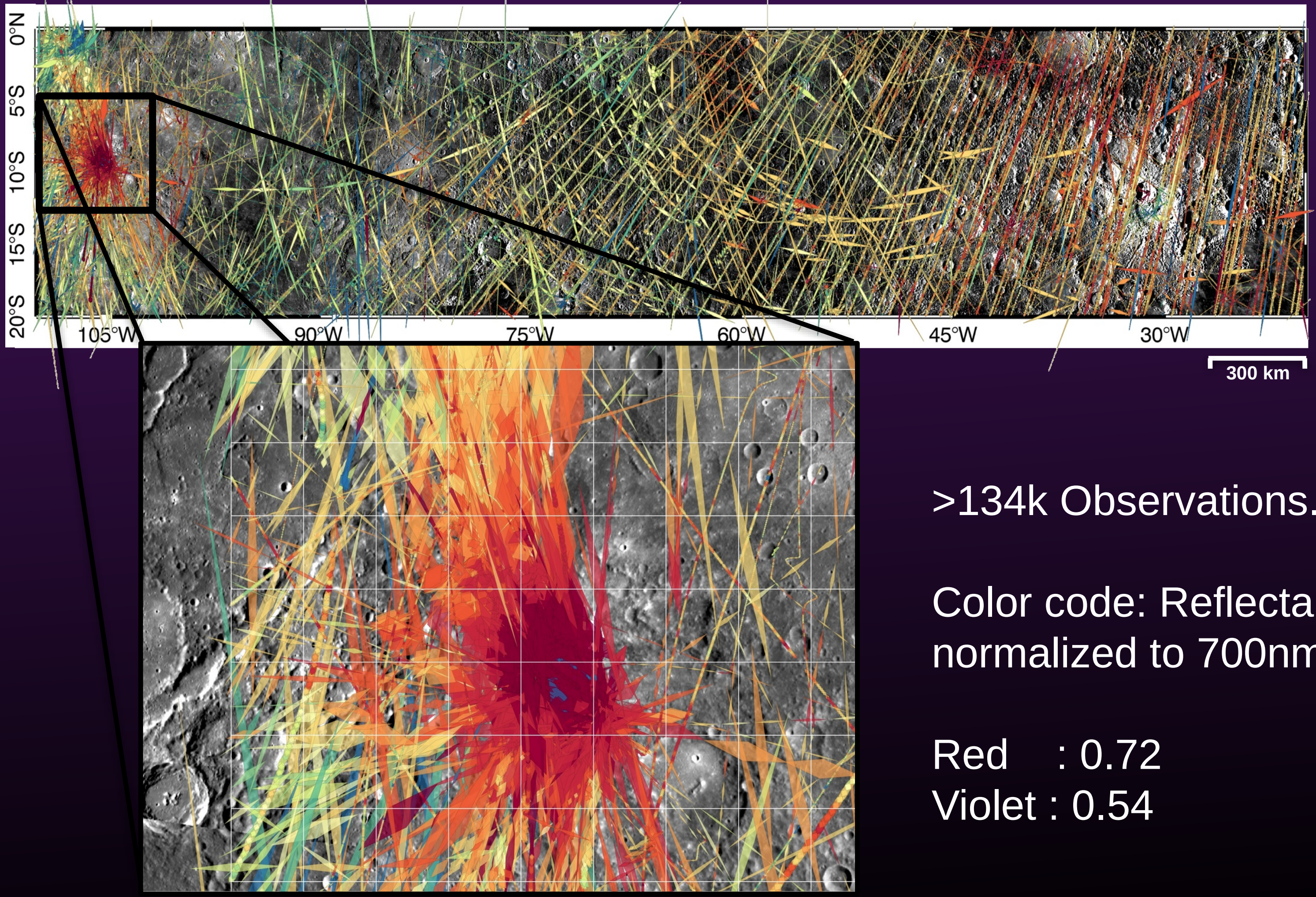
Data extraction

Polygons and measurements intersection





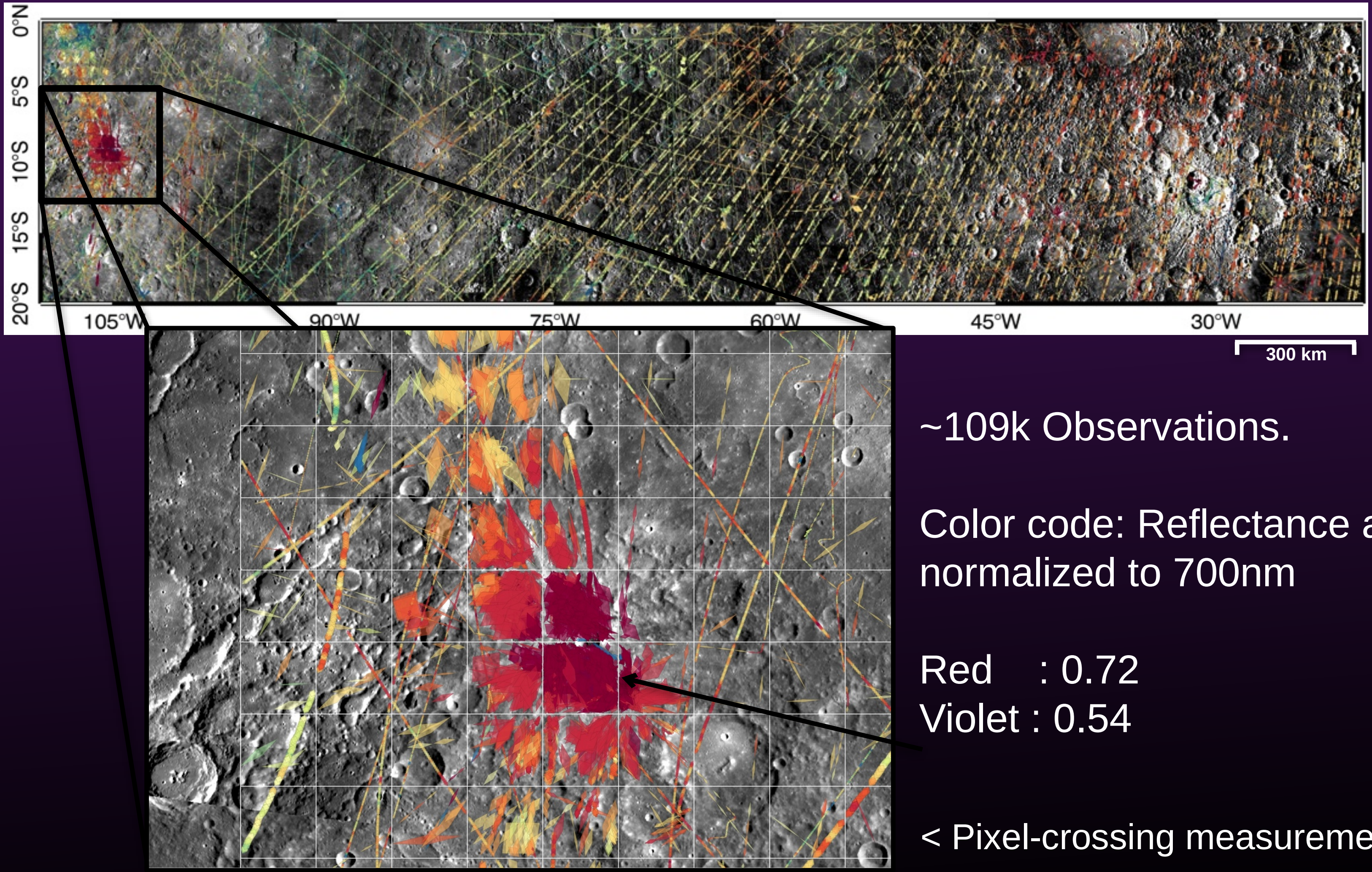
Data selection



>134k Observations.

Color code: Reflectance at 450nm normalized to 700nm

Red : 0.72
Violet : 0.54



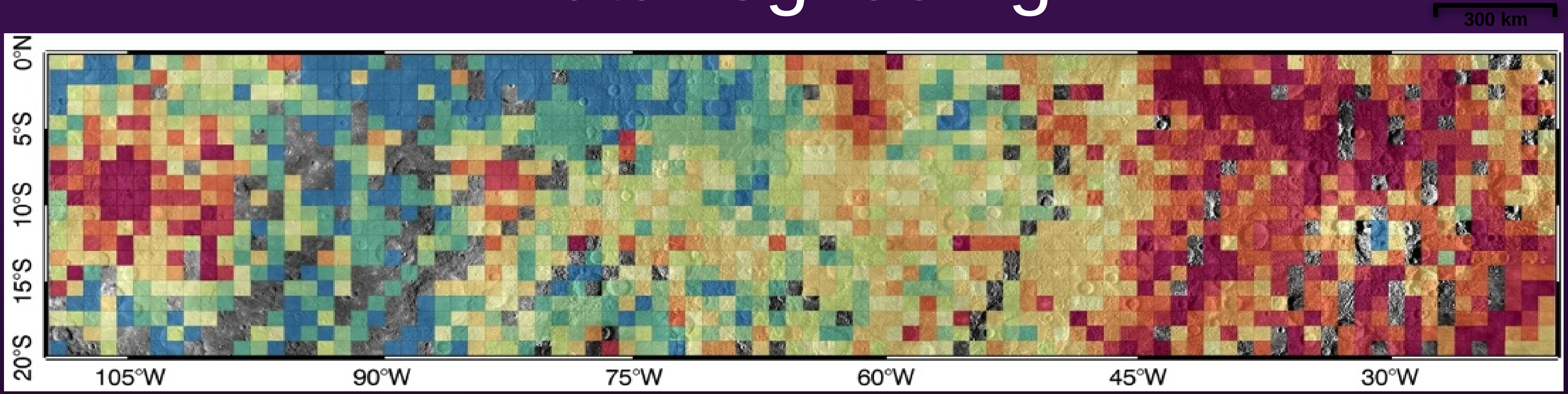
~109k Observations.

Color code: Reflectance at 450nm normalized to 700nm

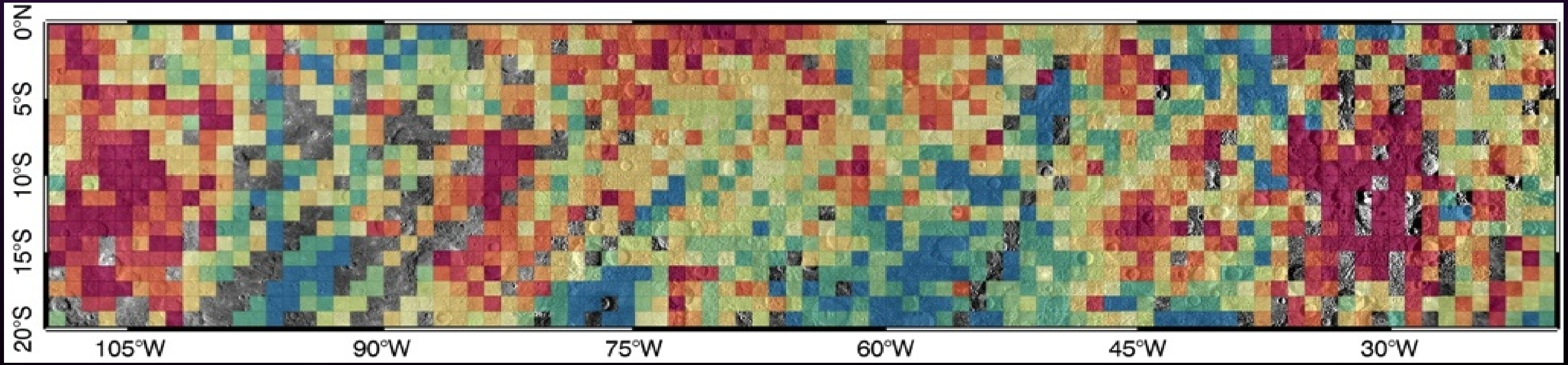
Red : 0.72
Violet : 0.54

< Pixel-crossing measurements filtered

Data regridding



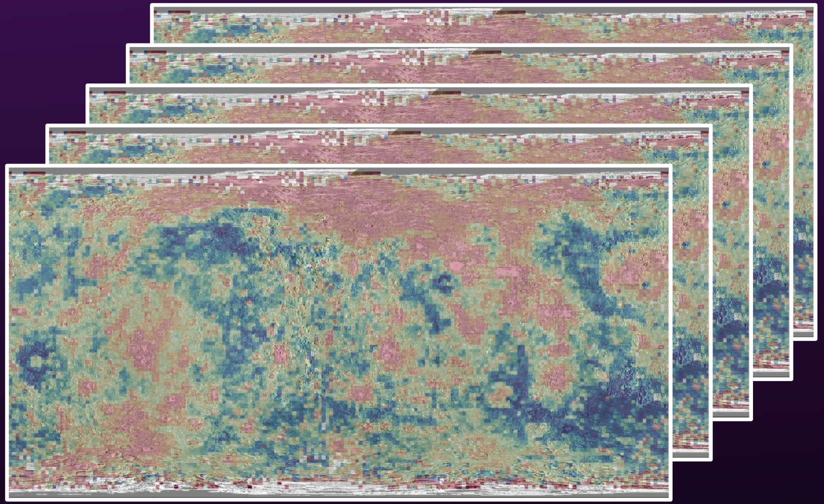
Reflectance at 450nm / 700nm – Blue : 0.54 / Red : 0.72



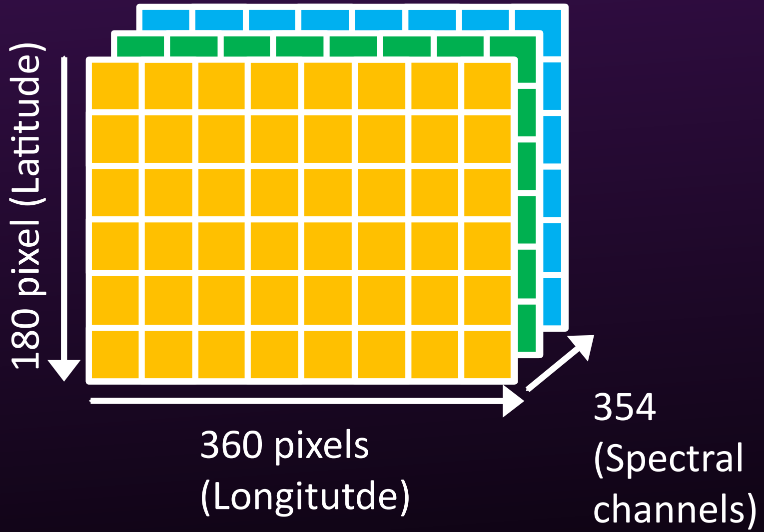
Reflectance at 700nm – Blue : 0.04 / Red : 0.09

MASCS/MESSENGER Data Cube

Data cube 3D



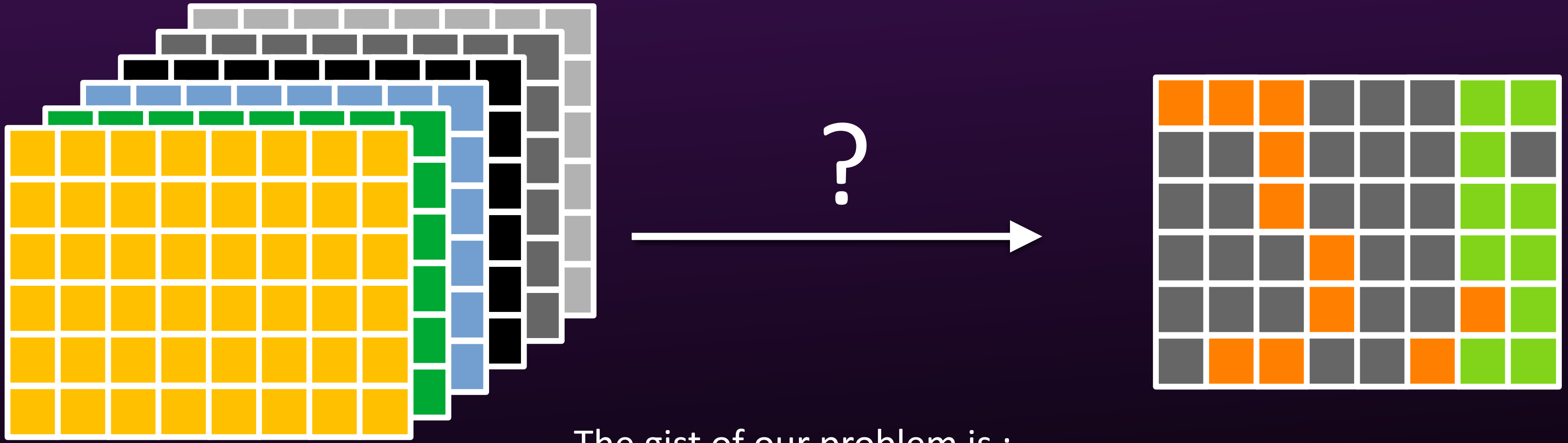
Data cube 3D schematic



Schematic workflow

Data cube 3D
dimensions[360,180,396]*

Classification map 2D
dimensions[360,180,1]



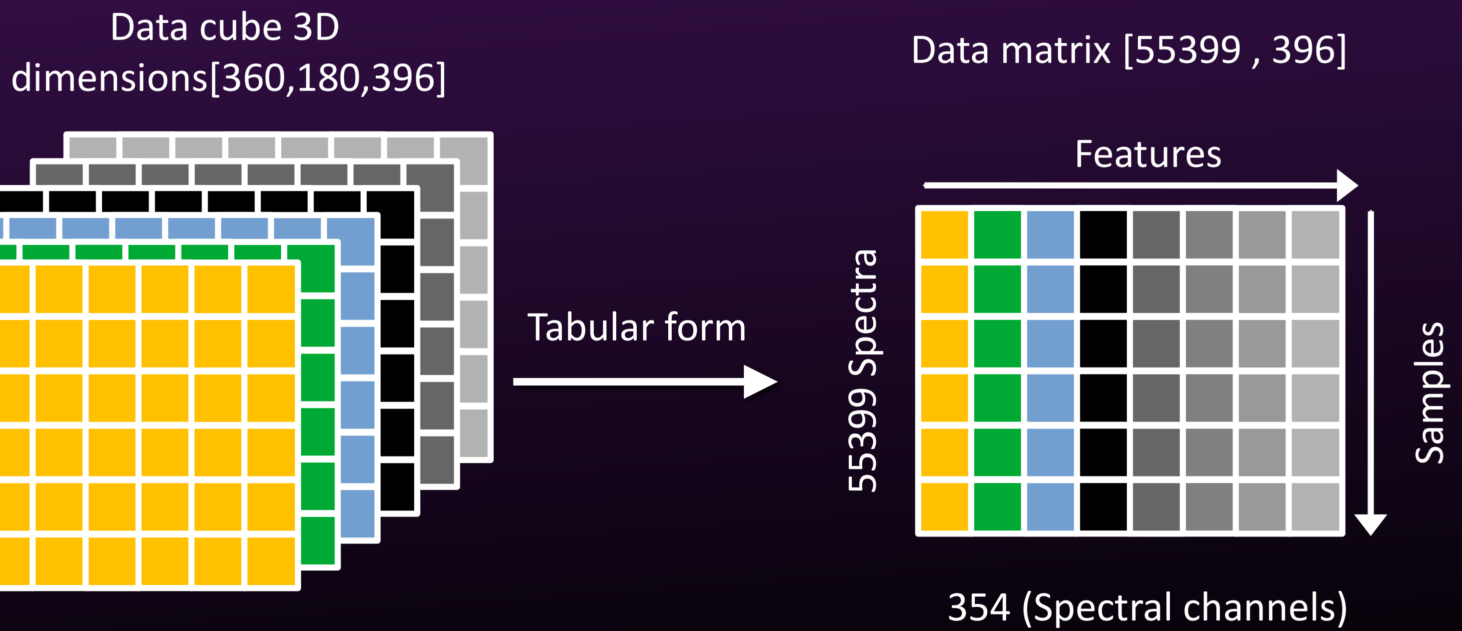
The gist of our problem is :

How to go from a [360,180,354] array to [360,180,1] one?

[360,180,1] is essentially a map containig the same values for the same units.

Schematic workflow

The instrument didn't covered all surface, so not all [360,180] pixels contain data.
Data quality filtering too gives a data matrix of 55399 spectra
instead of 360 × 180 = 64800



Schematic workflow

Data matrix

ICA/PCA
decompostion
scikit decomposition

manifold
learning
UMAP docs

hierarchical
agglomerative
clustering
scikit-learn clustering

[55399 , 354]

coefficiens
[55399 , 4]

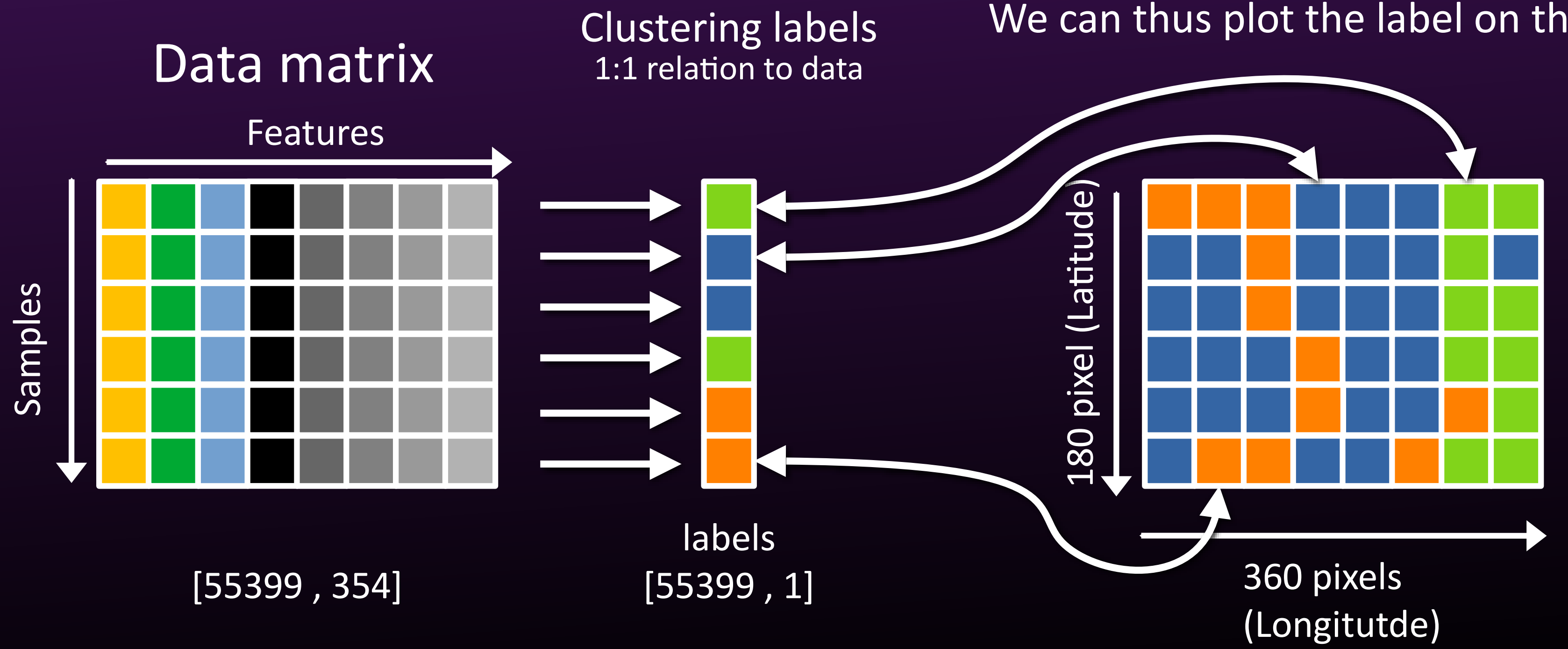
coefficiens
[55399 , 2]

labels
[55399 , 1]

Schematic workflow

Each label is in relation to each initial pixel and its geographical coordinates.

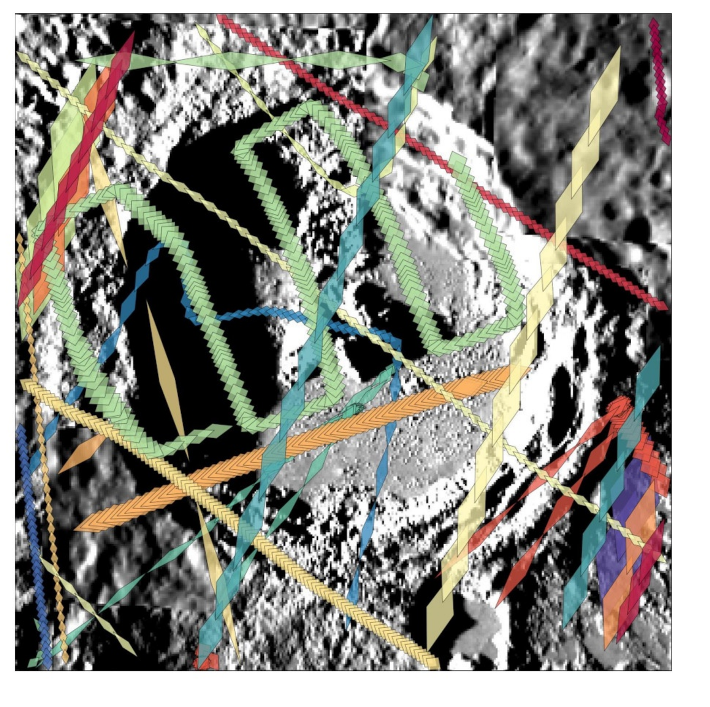
We can thus plot the label on the map.

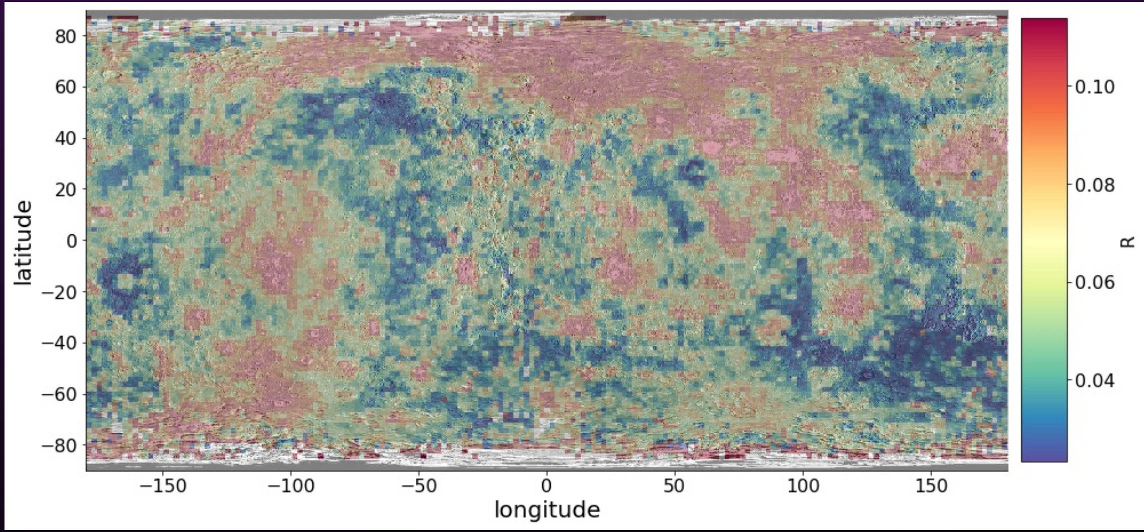


MASCS/MESSENGER Data

MASCS VIRS data in **Kuiper crater**, 62-km-diameter crater with a central peak (11.35 S 31.23 W)

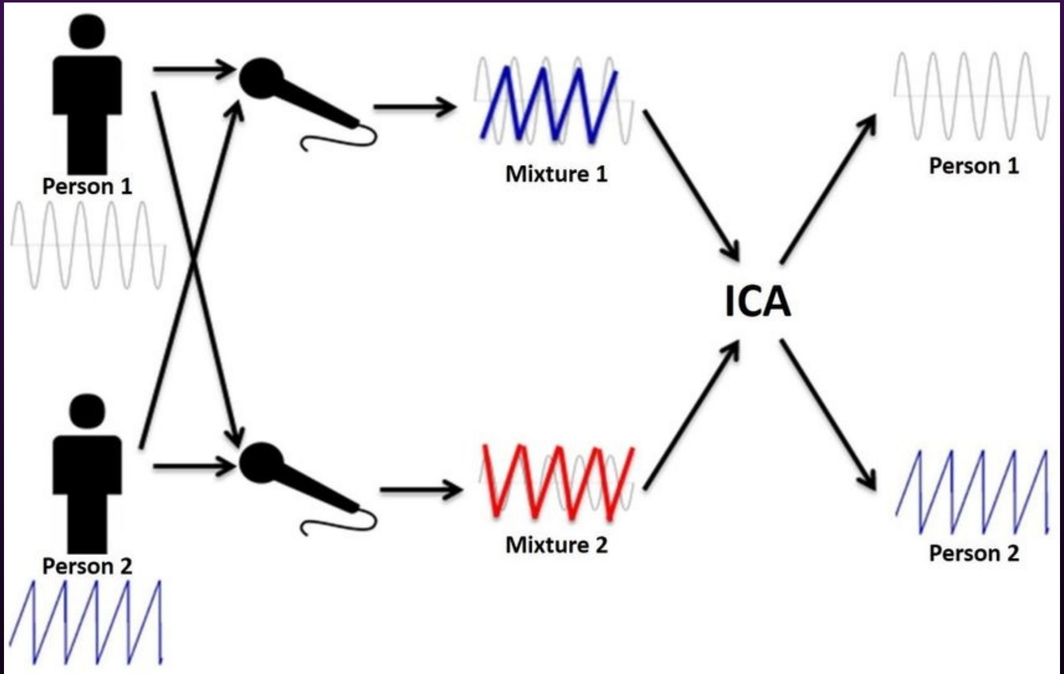
Kuiper shows the highest albedo of the planet and has a fresh ray system, suggesting that it is among the youngest craters of Mercury.





Global datacube.
Color is median **reflectance at 700 nm.**
Data dimension [55399, 354].

Independent component analysis (ICA)



ICA classical example : the cocktail-party problem.

The goal is to recover the individual signals P1 and P2 from the mixtures signals X1 and X2.

Credit : DOI:10.1190/int-2018-0109.1

Independent component analysis (ICA)

The data are decomposed via Independent component analysis (ICA), a signal processing technique for separating multivariate signal into additive sub components.

Data dimension from [55399,396] to [55399, 4].

ICA tries to find a linear static transformation W so

s = W * x

```
s [55399,   k] :
X [55399, 354] :
W [354    , k] :
```

- **s [55399, k]** : k maximally independent components
- **X [55399, 354]** : initial data
- **W [354 , k]** : mixing matrix for each component s(k)

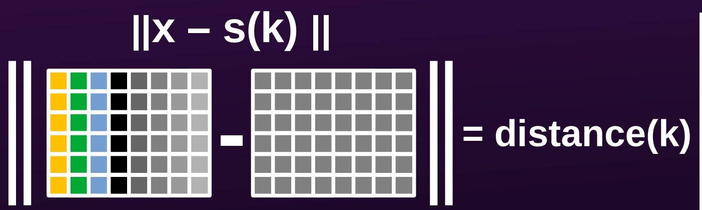
Assumptions : (1) at most one subcomponent is Gaussian , (2) the subcomponents are statistically independent from each other.

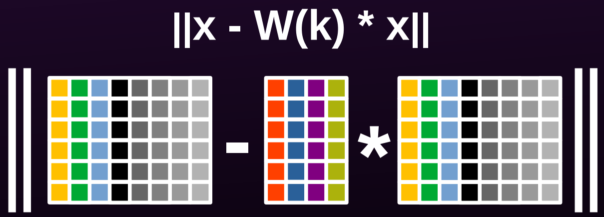
Independent component analysis (ICA)

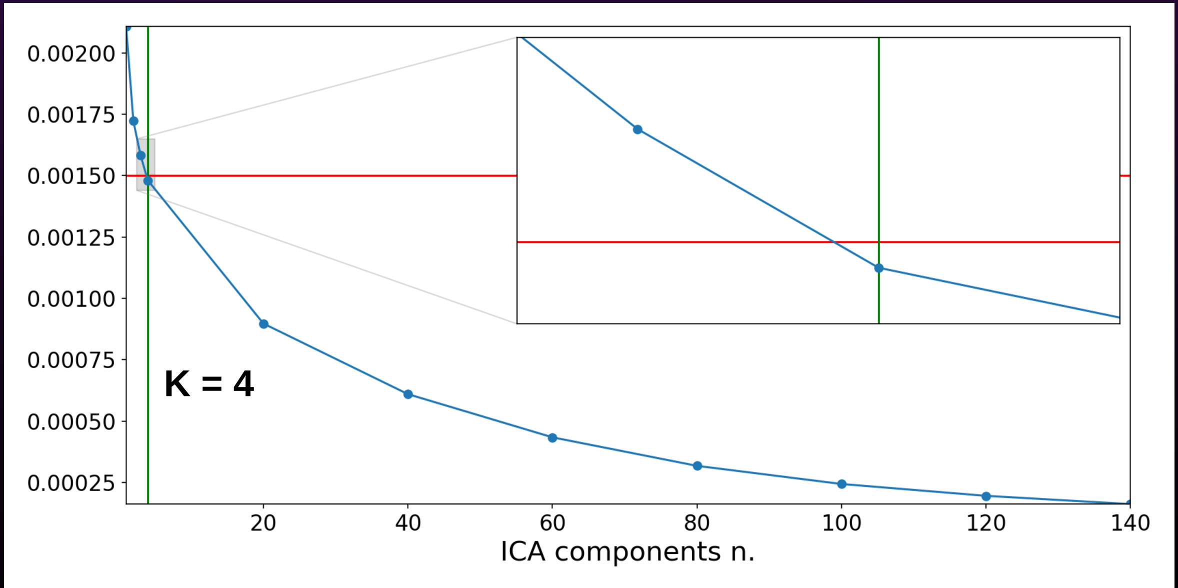
Determining the optimal number of components for accurate representation is the hardest part.

Increasing the number of k of independent components **s** the reconstruction error decreases.

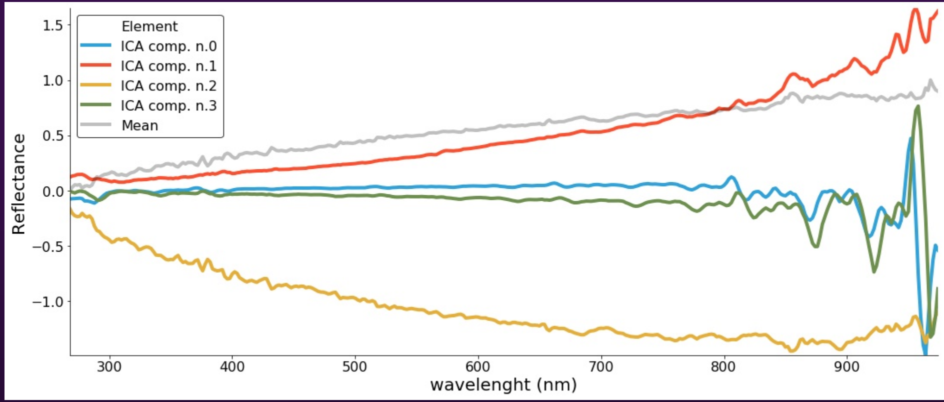
When reconstruction error ||x−s(k)||<0.0015 we reached the threshold that represents the noise in the data.





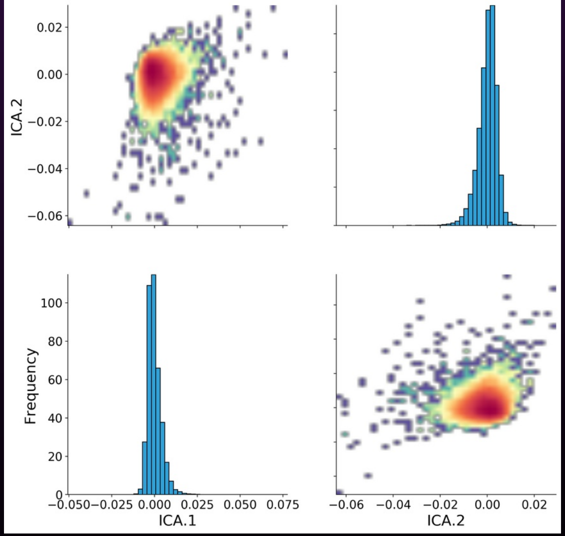


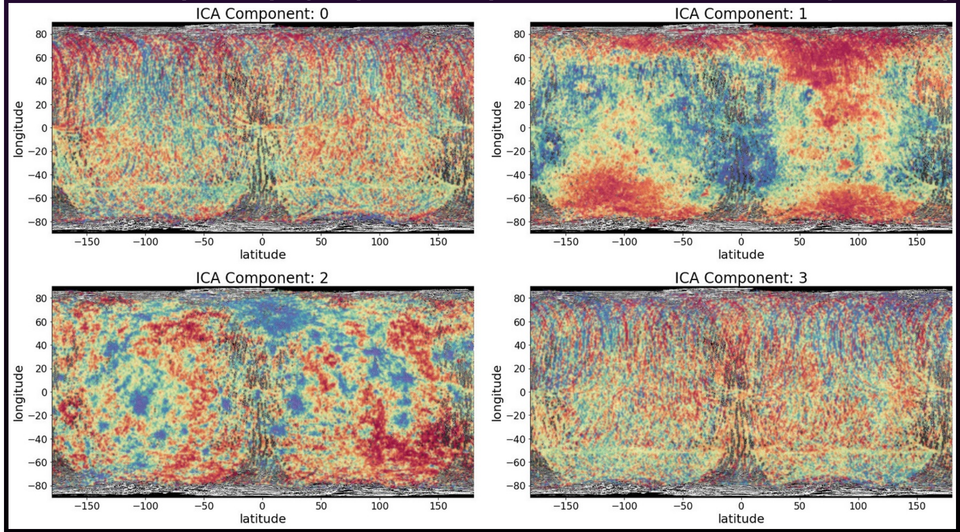
Independent component analysis (ICA)



< **s** [55399,4] maximally independent components

W [354,4] loading for each component $\mathbf{s}_i$
scatterplot (right) and spatial distribution (below)

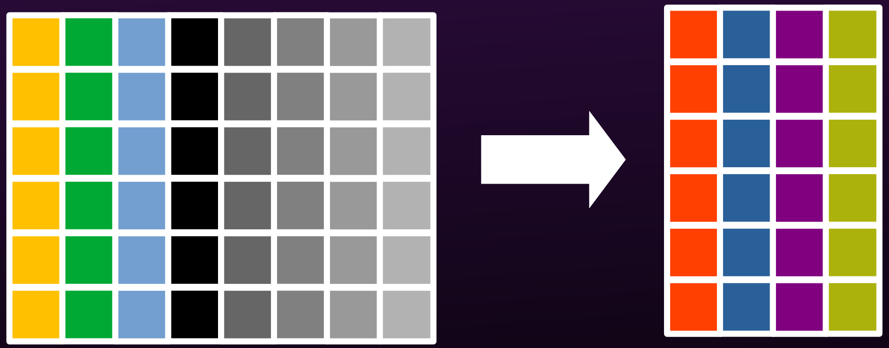




Schematic workflow

Data matrix

ICA/PCA
decompostion
scikit decomposition



[55399 , 396]

coefficiens
[55399 , 4]

Uniform Manifold Approximation and Projection (UMAP)

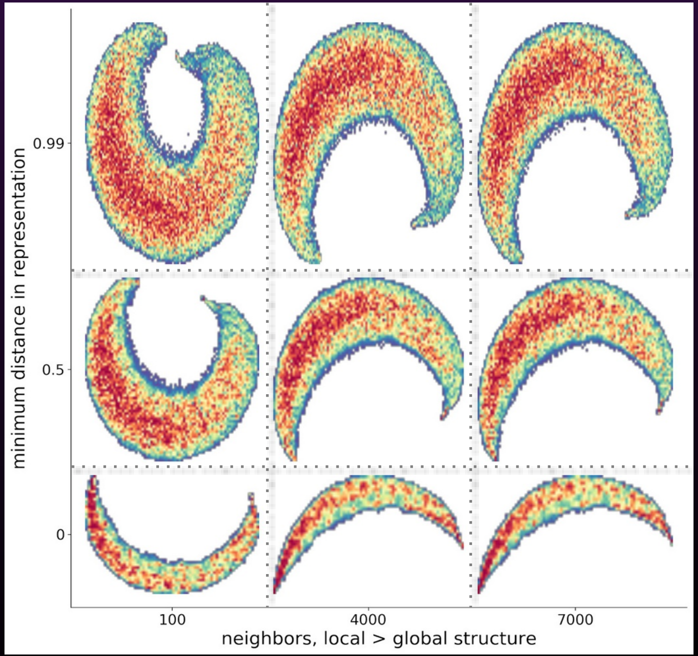
UMAP compress data dimension from [55399,4] to [55399, 2].

- UMAP uses local manifold approximations to construct a topological representation of the high dimensional data (high to low dimension)

- Given some low dimensional representation of the data, a similar process can be used to construct an equivalent topological representation in the other direction (low to high dimension).

- UMAP then searches for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological structure to the high dimensional one.

- hyperparameters
  - *n_neighbors* : sensitivity to local or global structures in the data by constraining the size of the local neighborhood. Values can go from 0 (local) to the size of the data (global).
  - *min_dist :* minimum distance apart that points are allowed to be in the low dimensional representation. Low values will result in more dense embeddings while larger values will result in more sparse embeddings.



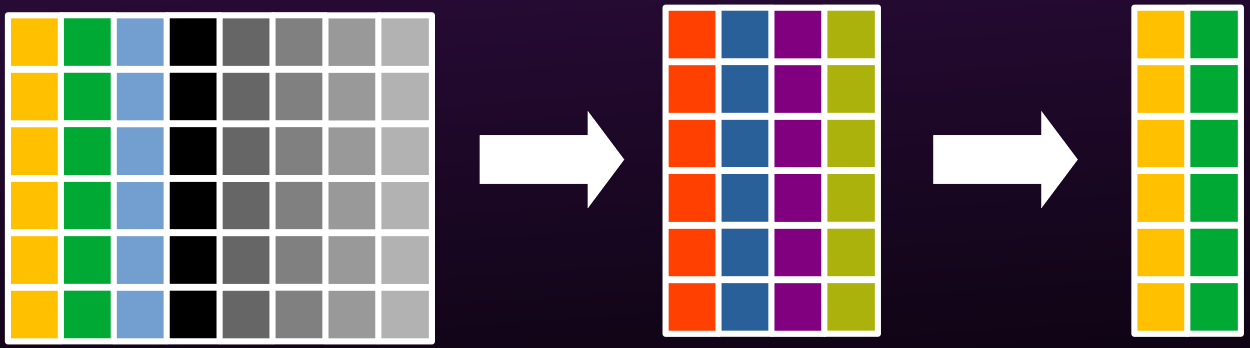
A rigorous mathematical background could be fund at umap-learn.readthedocs.io

Schematic workflow

Data matrix

ICA/PCA
decompostion
<u>scikit decomposition</u>

manifold
learning
<u>UMAP docs</u>



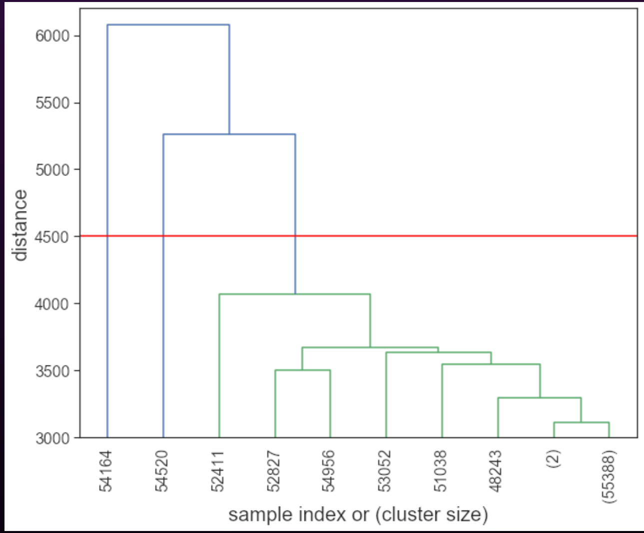
[55399 , 396]

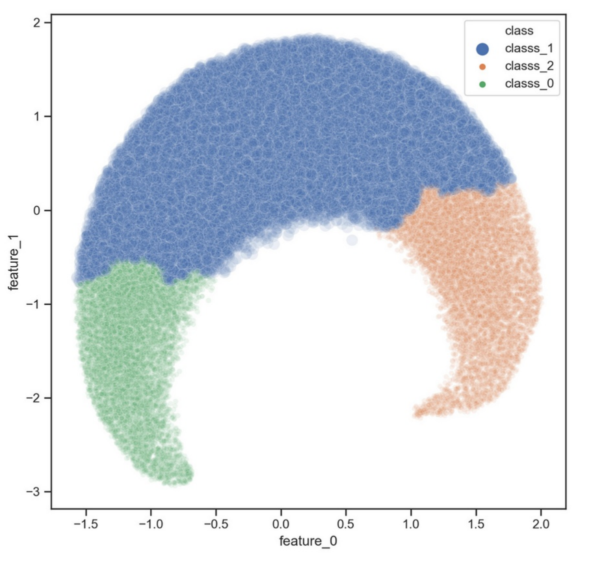
coefficiens
[55399 , 4]

coefficiens
[55399 , 2]

Unsupervised Classification : Hierarchical Clustering

- Hierarchical clustering (HC) algorithms build nested clusters by merging or splitting them successively.
- HC merges closer points together, irrespective of the final cluster class balance.
- The hierarchy of clusters is represented as a dendrogram or tree.
- The final tree has one unique cluster comprising all the samples at the root and clusters with only one sample at the leaves (bottom of t tree).
- The branch lengths represent the distance between the child clusters.
- Small gaps connect more similar clusters and big gaps connect more different clusters.
- The cluster distance is computed as the maximum Euclidean distance distances between all observations of the two clusters.





Schematic workflow

Data matrix

ICA/PCA
decompostion
scikit decomposition

manifold
learning
UMAP docs

hierarchical
agglomerative
clustering
scikit-learn clustering

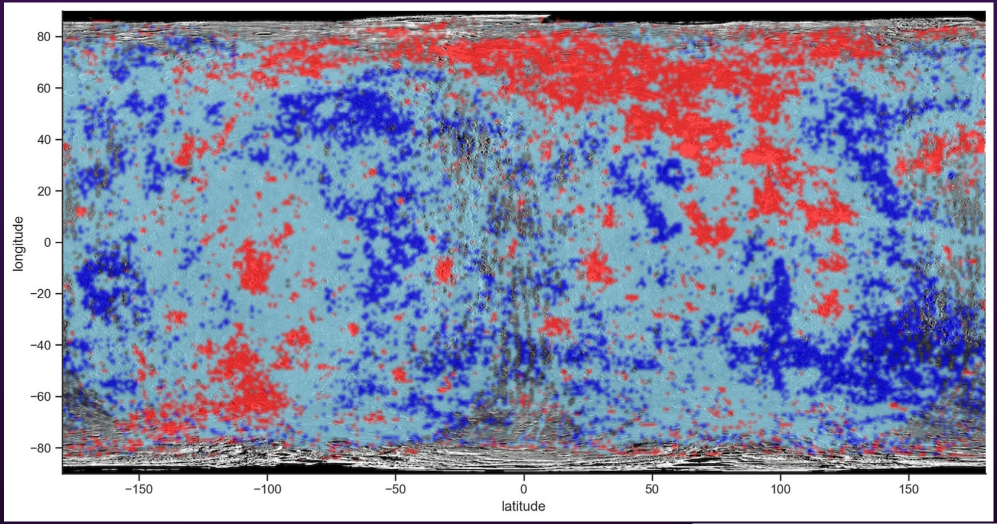
[55399 , 396]

coefficiens
[55399 , 4]

coefficiens
[55399 , 2]

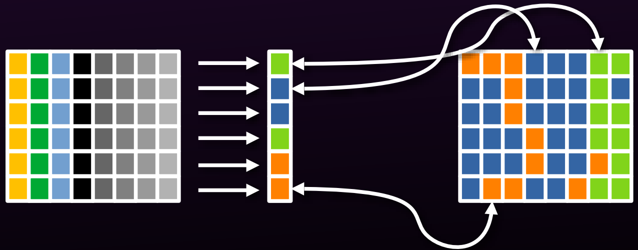
labels
[55399 , 1]

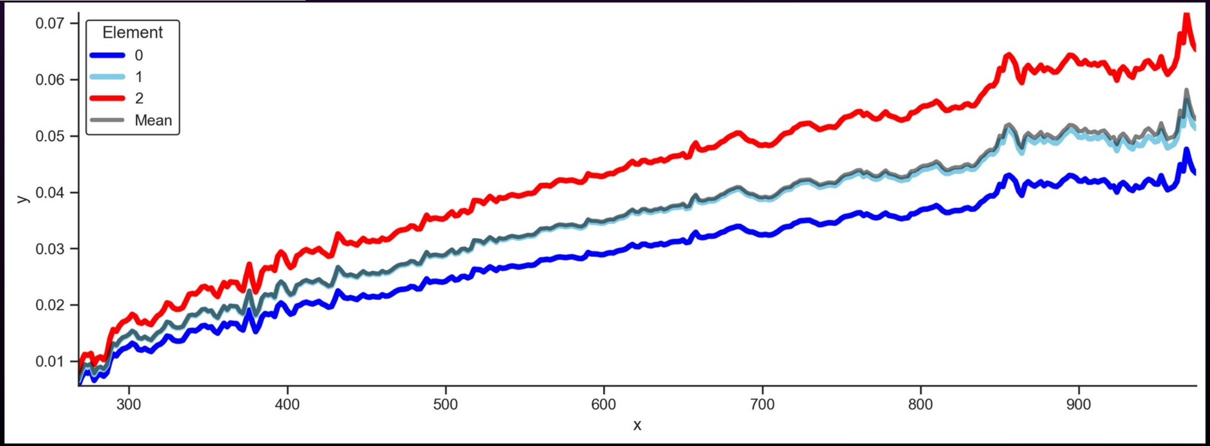
Unsupervised Classification : Hierarchical Clustering



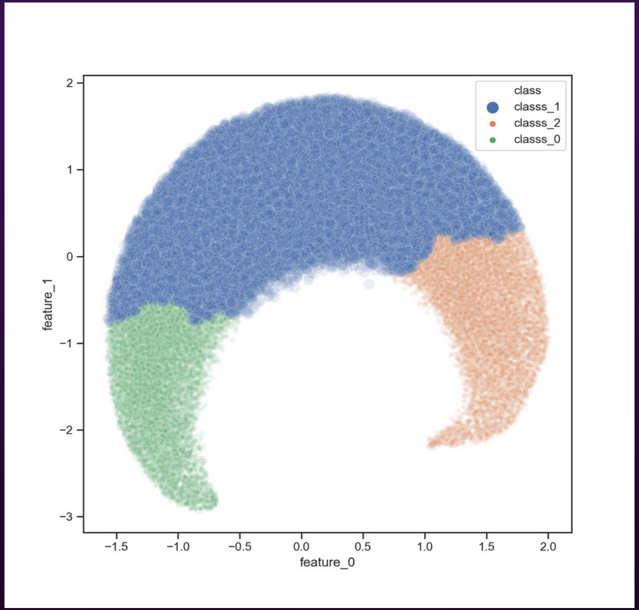
Surface unsupervised classification spatial distribution (**left**) and average classes spectra (**below**)

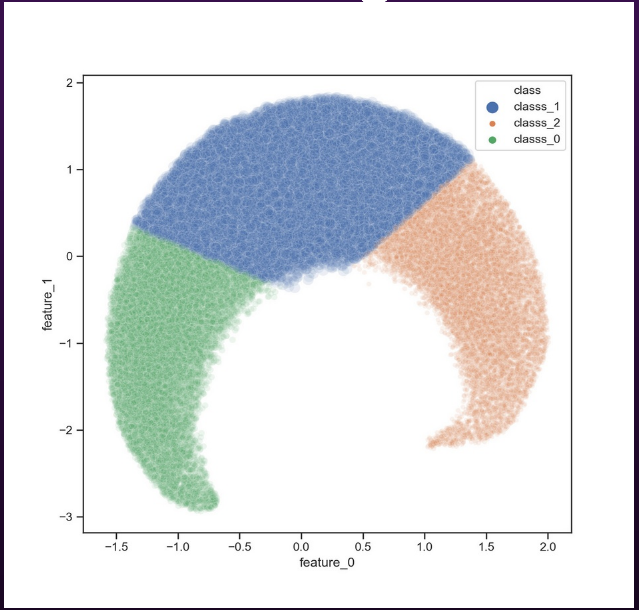
Each label is in relation to each initial pixel and its geographical coordinates.

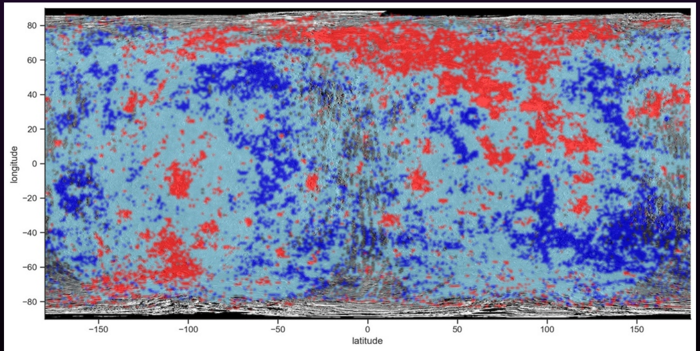


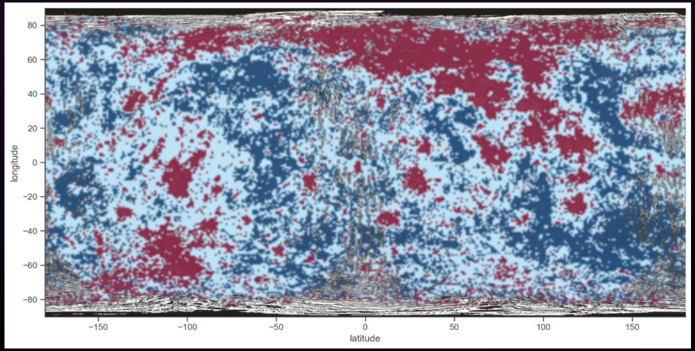


Hierarchical Clustering vs K-Means



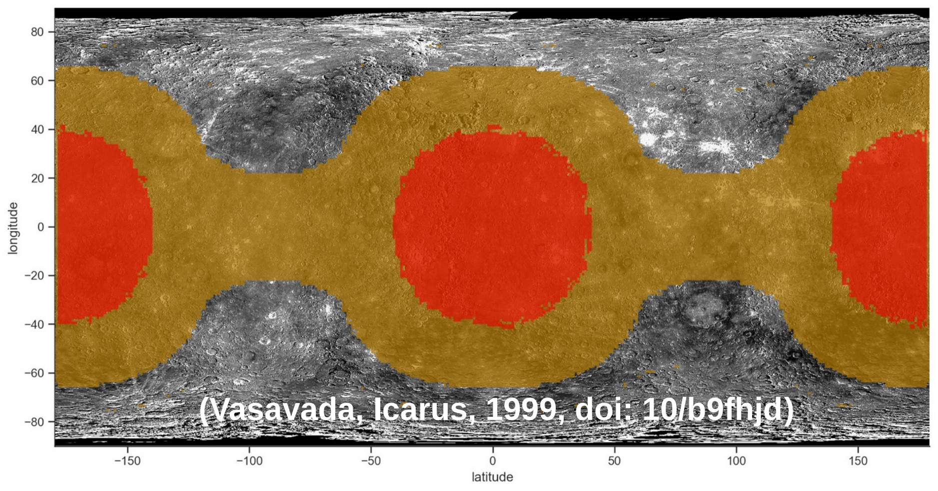




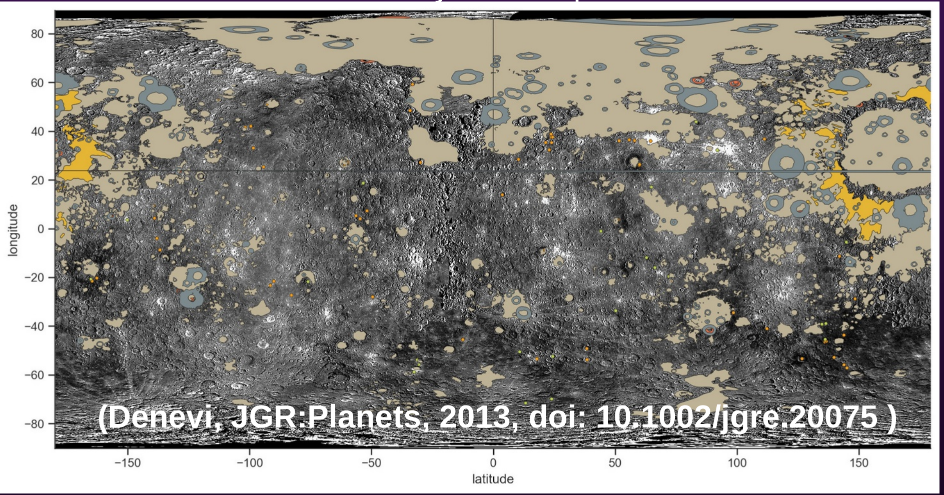


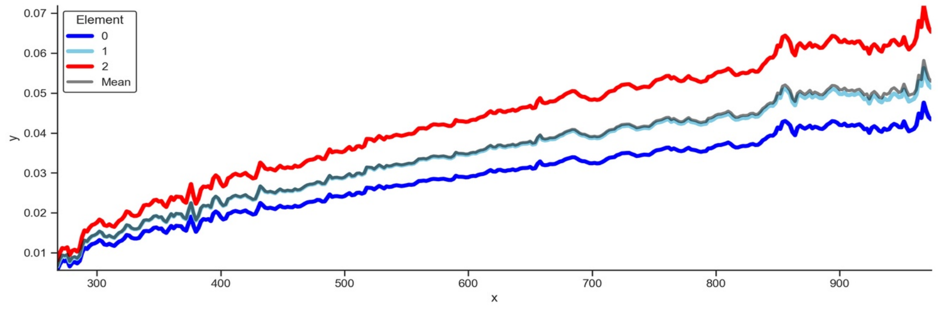
Comparison with other maps

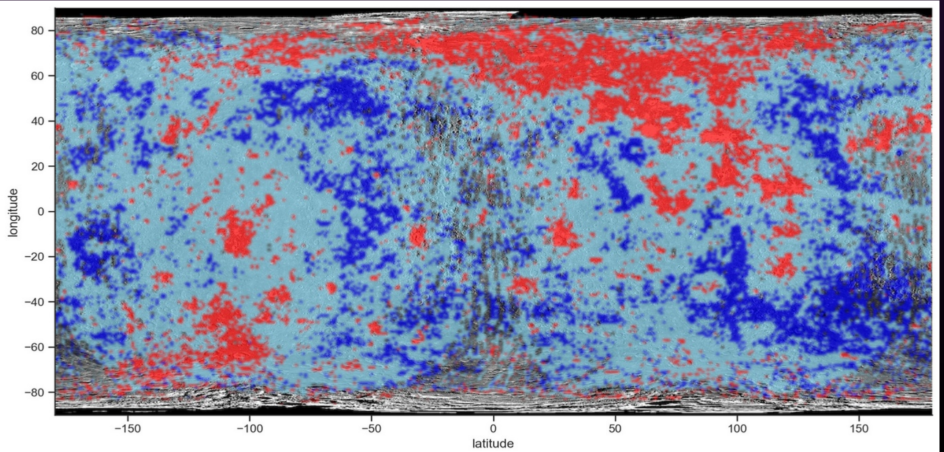
Maximum surface temperature: red > 690 K, brown > 550K



Mercury smooth plains

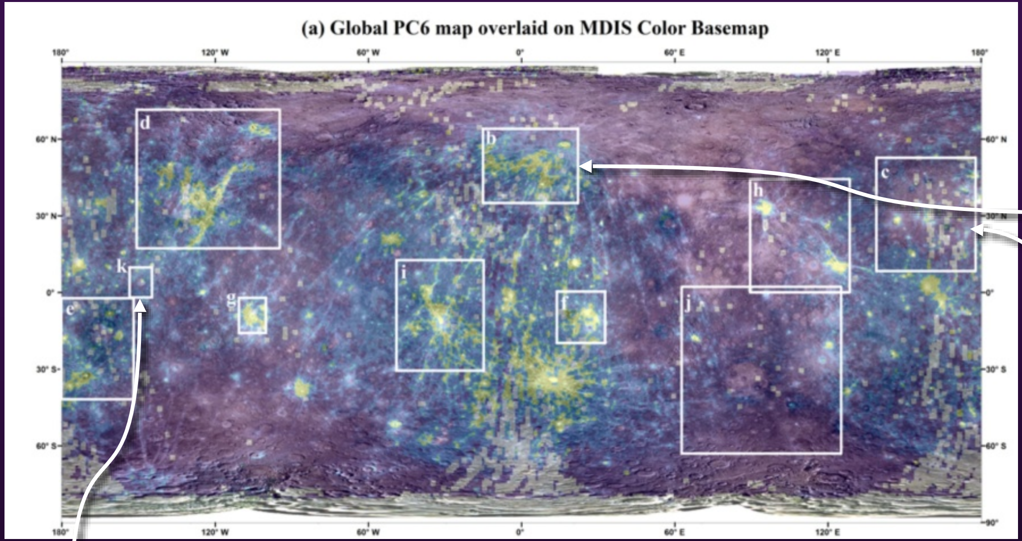


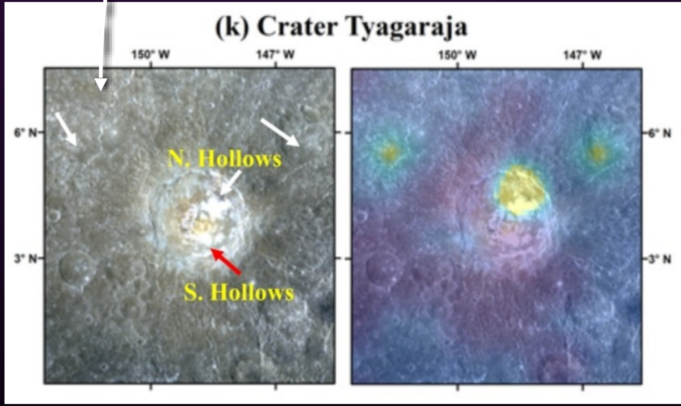


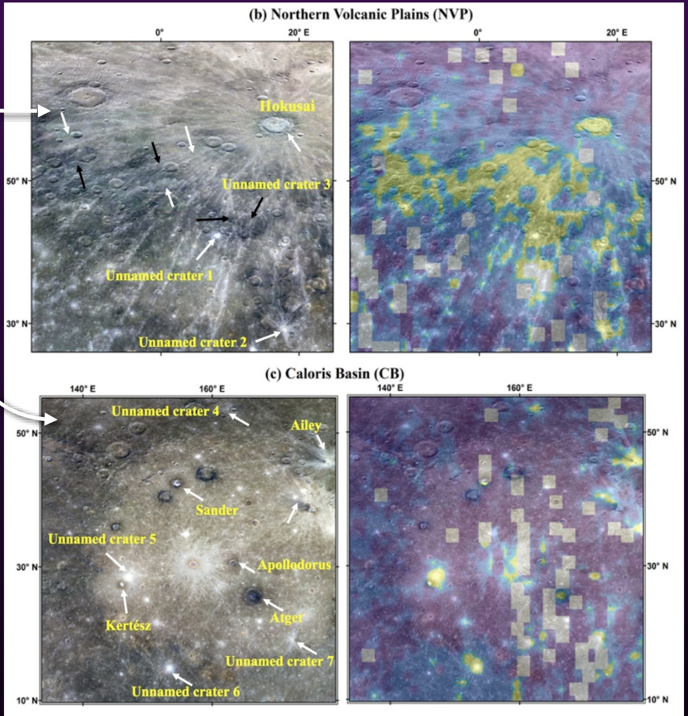


Surface unsupervised classification spatial distribution (**above**) and average classes speactra (**right**)

Other works







Varatharajan, I., D'Amore, M., Domingue, D., Helbert, J., Maturilli, A., 2020. Global Multivariate Spectral Analysis of Mercury and the Identification of Geochemical Terrains: Derived from the MASCS Spectrometer onboard NASA's MESSENGER Mission.Earth and Space Science Open Archive. https://doi.org/10.1002/essoar.10501760.1

Results

- The results indicate a **dichotomy in major units**, with one spectral unit in the polar regions and a spectrally distinct unit in equatorial areas.
- The spatial extent of the **polar unit** in the northern hemisphere shows a generally **good correlation with the northern volcanic plains** and with the regions of **enhanced potassium abundance** as mapped by the Gamma-Ray Spectrometer on MESSENGER.
- This **asymmetry** indicates that **peak surface temperature**, although potentially important, is **not the only factor** that contributes to the different spectral character of the surface units.
- We conclude that **compositional differences** also contribute to the spectral differences between the two major units on the surface of Mercury identified with this hierarchical clustering approach.