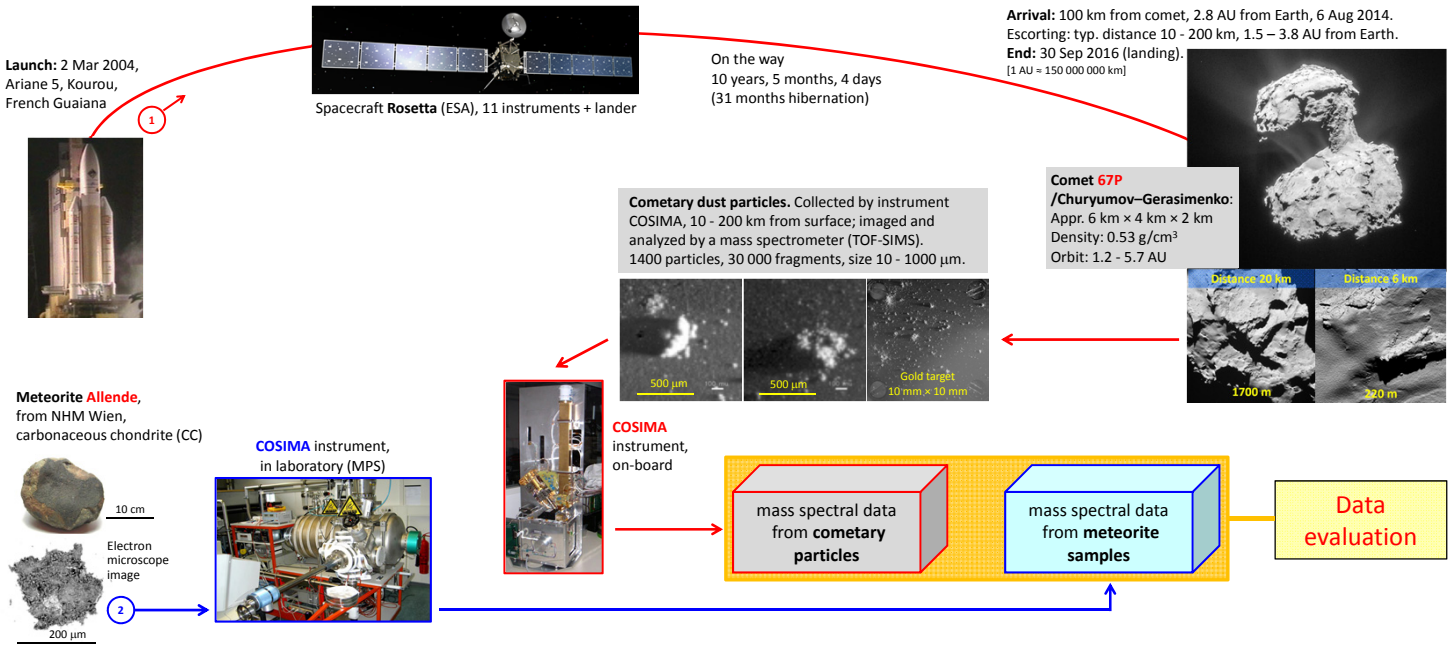
# CSTAT - Computational Statistics

Kurt Varmuza*, Irene Hoffmann, Peter Filzmoser

# C o M e C S Project

## Comet and Meteorite Materials - Studied by Chemometrics of Spectroscopic Data

Data from deep space (comet), and laboratory (meteorites) → Multivariate statistics Chemoinformatics → Information about chemical composition of samples




## Data selection

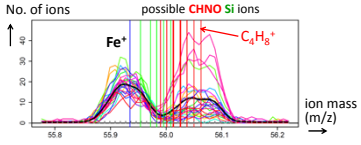
More than 30000 mass spectra measured in space; about 6000 spectra on meteorite samples.

Original spectrum: a vector with about 130000 numbers (ions counted in mass intervals).

Selection of spectra measured on particles or on background: by spectroscopic concepts, supported by multivariate statistics (e. g., orthogonal distance to robust PCA models, KNN approach, NMF).

Selection of variables (from relevant mass intervals) by exhaustive chemical formulae generation.

Example: 27 spectra around mass 56, measured on a cometary particle




## Exploration

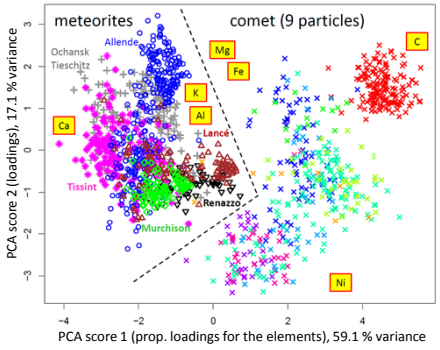
**Data**

$n$ = 1426 objects (cases, COSIMA spectra), 509 from nine cometary particles, 917 from seven meteorites (NHM Wien).

$m$ = 7 variables (counts of positive ions for C, Mg, Al, K, Ca, Fe, Ni), centered log-ratio transformed (compositional data).

**Method**

Principal component analysis (PCA); biplot with scores and loadings.




**Interpretation** (*preliminary, also considering other results*)

- Cometary particle surfaces are clearly different from meteorites.
- Cometary particles show higher carbon contents than the carbon-rich meteorites (CC: Allende, Lancé, Murchison, Renazzo).

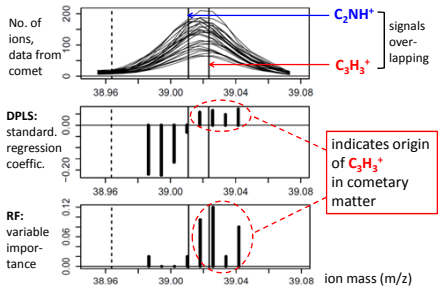
## Discrimination

**Data**

$n_1$ = 29 objects (cases, COSIMA spectra) from a cometary particle,
$n_2$ = 59 objects (cases, COSIMA spectra) from background
$m$ = 540 variables (counts of positive ions containing C, H, N, O)

**Methods and Aim**

Estimation of **variable importance** for discrimination between the two classes of objects by (1) discriminant partial-least squares classification (DPLS), and by (2) random forest classification (RF). Variables with high importance may indicate CHNO ions originating from **organic material in cometary matter**.




**Interpretation** (*preliminary, also considering other results*)

- Cometary material contains carbon from CH(NO)-substances with high molecular masses; typical ions are $C^+$, $CH^+$, $CH_2^+$, $CH_3^+$.
- Further (unsaturated) ions from organics are $C_2H_{2\text{-}3}^+$, $C_3H_{0\text{-}3}^+$, $C_4^+$.

**People (Austria)**

**Project coordination:** Kurt Varmuza
**TU Wien/105/CSTAT:** Peter Filzmoser, Irene Hoffmann, students
**TU Wien/164/Analytical Chemistry:** Günter Allmaier, Ernst Pittenauer, students
**BOKU Wien:** Notburga Gierlinger (IR microscopy)
**Naturhistorisches Museum Wien (NHM):** Franz Brandstätter, Ludovic Ferrière, Christian Köberl

**Supported by the Austrian Science Fund (FWF), project P 26871 - N20.**

**Collaborations**

**COSIMA team:** Principal investigator of the COSIMA project (within the Rosetta project of the European Space Agency, ESA) is Martin Hilchenbach, Max Planck Institute for Solar System Research (MPS), Göttingen; research groups from Germany, France, Finland, Sweden, and Austria.

**Details, references, PDFs:** **http://www.lcm.tuwien.ac.at/comecs/**



* kurt.varmuza@tuwien.ac.at; www.lcm.tuwien.ac.at/vk