COMET DUST COMPOSITION EXPLORED BY CHEMOMETRIC METHODS USING MASS SPECTRAL DATA FROM COSIMA/ROSETTA

Kurt Varmuza[1*], Donia Baklouti[2], Anais Bardyn[3,4], Hervé Cottin[3], Cécile Engrand[2], Peter Filzmoser[1], Nicolas Fray[3], Martin Hilchenbach[5], Irene Hoffmann[1], Jochen Kissel[5], Paola Modica[4], Johan Silén[6], Sandra Siljeström[7], Oliver Stenzel[5]

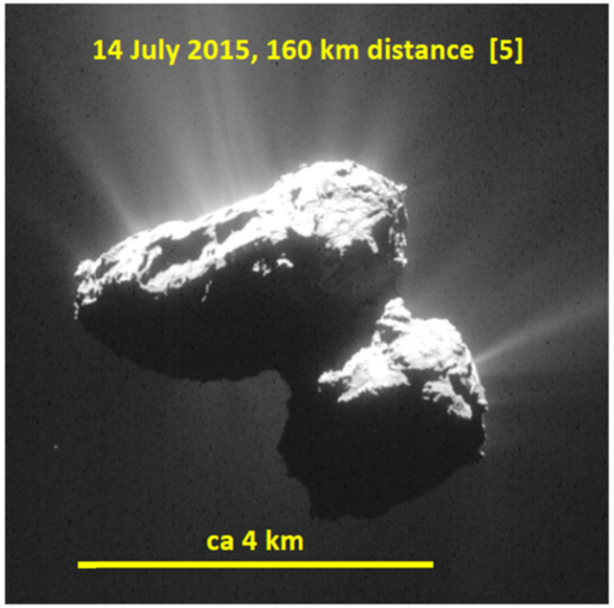
[1] Vienna University of Technology, Institute of Statistics and Mathematical Methods in Economics, Research Unit Computational Statistics, Vienna, **Austria**;
[2] Université Paris Sud, Orsay, **France**; [3] Université Paris-Est Créteil et Université Paris Diderot, Créteil, **France**;
[4] Université d'Orléans, Orléans, **France**; [5] Max Planck Institute for Solar System Research (MPS), Göttingen, **Germany**;
[6] Finnish Meteorological Institute, Helsinki, **Finland** ; [7] RISE Research Institutes of Sweden, Stockholm, **Sweden**

- **COSIMA** [1] was operated on-board of spacecraft **Rosetta**;
  - collected ca. 35,000 **cometary dust particles**;
  - imaged them (size determination and categorization) [2, 3];
  - measured ca 34,000 mass spectra [4] on ca. 400 particles.
- A set of selected **mass spectra** has been evaluated
  - by various chemometric methods,
  - to characterize the **homogeneity** and **composition** of the particles, and to search for **CHNO containing substances**.

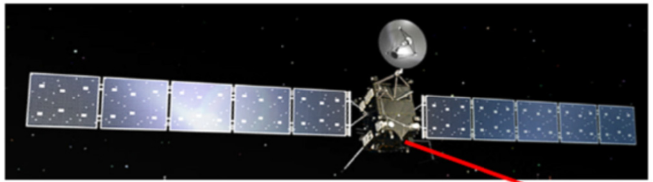
Poster at SSC15 – 15th Scandinavian Symposium on Chemometrics, 19 – 22 June 2017, Naantali, Finland
http://ssc15.metropolia.fi/

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

Version 170705



Comet 67P/Churyumov-Gerasimenko
Orbit: 1.2 – 5.7 AU from sun, 6.4 years;
rotation 12.76 h; density 0.4-0.5 g/cm^3;
albedo 5 % reflectance (very black).



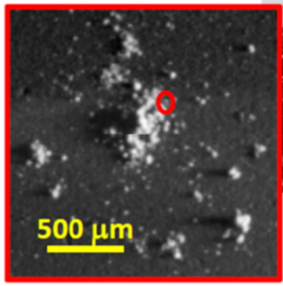
Rosetta spacecraft (ESA)
Total 3000 kg, incl. 100 kg lander *Philae*,
165 kg instruments, 1700 kg propellant
(methyl-hydrazine + N_2O_4). 3 m x 2 m x 2 m;
solar panels 32 m wide; dish antenna 2.2 m .

2 Mar 2004 launch
6 Aug 2014 arrival (100 km)

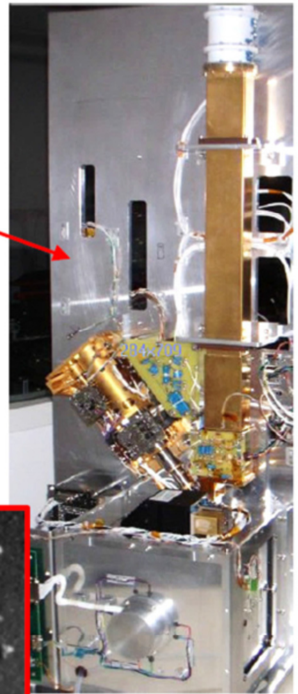
Escorting the comet,
typ. 10 – 200 km distance,
1.5 – 4.8 AU from Earth,
1.2 – 3.8 AU from sun.

12 Nov 2014 *Philae* landing

30 Sep 2016 controlled end
of mission by impact on
comet (2 m/s), switch-off.



Cometary particle(s)
Named *Kerttu*, **O** SIMS spot size ca 50 µm x 70 µm



Assembling the instrument
COSIMA [1] for operation
on-board of **Rosetta.**

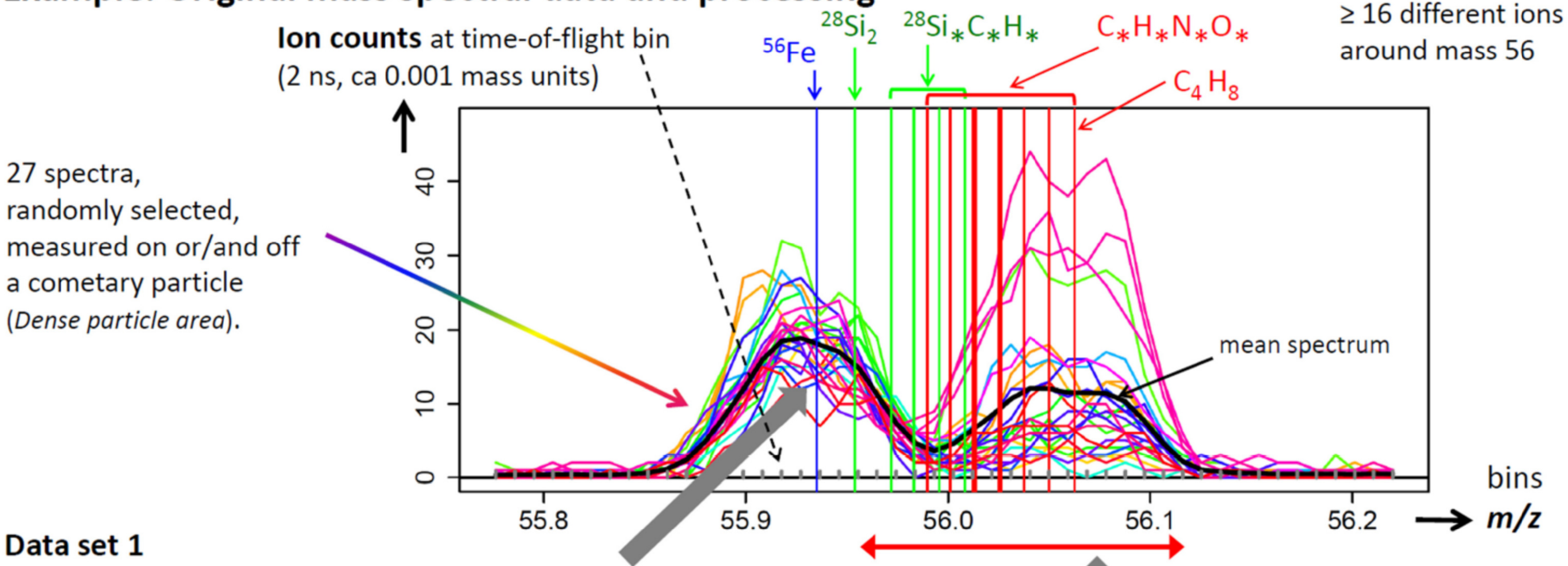
- ☐ Comet particle collection (targets 1 cm x 1 cm, typ. Au).
- ☐ Imaging (1024 x 1024 pixel, 14 µm resolution).
- ☐ **SIMS** measurements

Secondary ion mass spectrometry with TOF (time-of-flight) analyzer, single ion counting in 2 ns flight-time bins; mass resolution ca 1400 at *m/z* 100 (50%); mass range up to 6500 Dalton .

Typ. spectrum: 2.2 minutes,
42,000 bins for *m/z* 0 – 300.

20 kg, 20 W

Example: Original mass spectral data and processing



Data set 1

Peak fitting for 10 appropriate ions, C^+ (mass 12), CH^+ (13), CH_2^+ (14), CH_3^+ (15), Mg^+ (24), Al^+ (27), K^+ (39), Ca^+ (40), Fe^+ (56), Ni^+ (58), gives relative intensities (ion counts) for $m = 10$ variables. Not used (contaminations): Na^+, Si^+.

Data set 2

The many possible CHNO(Si) ions cannot be mass separated. *Raw data* from defined mass ranges are used as variables .

Homogeneity and composition of cometary particle surfaces

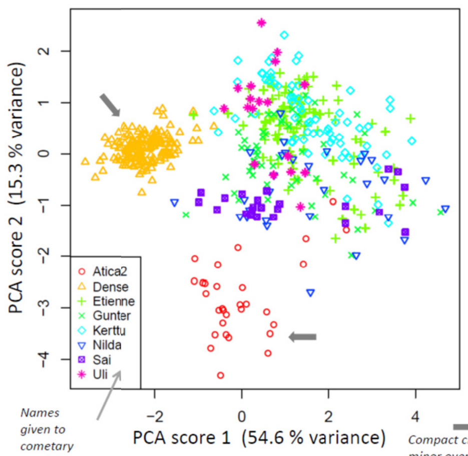
Data

- ❍ $n1$ = 516 spectra from 8 cometary particles.
- ❍ $n2$ = 674 reference spectra from 4 meteorites (❖). (CC; carbonaceous chondrites *Allende, Lancé, Murchison, Renazzo*), considered being similar to cometary material.
- ☐ m = 10 variables (**data set 1**), compositional data (CoDa); only relative values or ratios are relevant; processed by centered log-ratio transformation (CLR) [6] : x_{ij} (CLR) = log(x_{ij} /$\mathbf{G}(x_i)$) with $\mathbf{G}(x_i)$ for the geometric mean of all variables of observation (spectrum) *i;* with values x_{ij} = 0 replaced by artificial noise >0 (uniformly distributed).

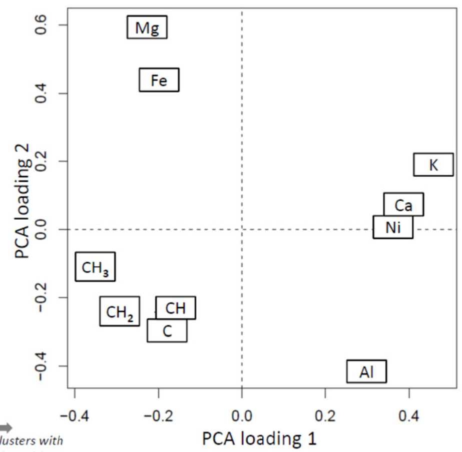
Methods

- ● PCA for visualization of the homogeneity or (diversity) of the cometary particle surfaces.
- ● KNN with rdCV (repeated double cross validation [7]) for a quantitative estimation of the separation of the spectra measured on the 8 cometary particles (and on meteorites).
- ● Correlation coefficient (Pearson) matrix for the 10 variables.
- ● Comparison of the mean spectra (normalized to sum 100) for cometary material and CC meteorites.
- ❖ Meteorite samples: COSIMA laboratory twin instrument.

PCA Comet

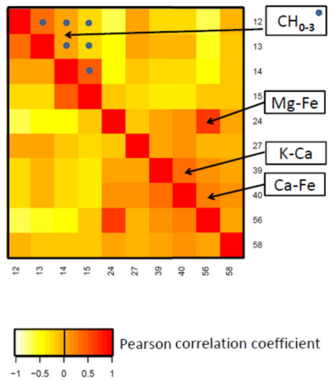


Names given to cometary particles.

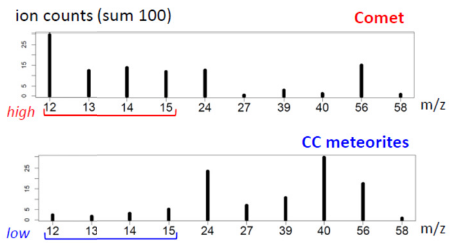


Compact clusters with minor overlap with another class also give high prediction in KNN.

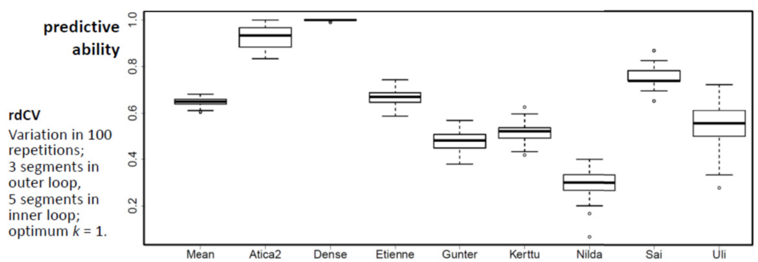
Variable correlation Comet



Mean spectra



KNN + rdCV Comet



Results

- Cometary particle surfaces show varying compositions (e. g., carbon-rich, Mg+Fe-rich, Mg+K+Fe-rich).
- Highest positive correlations appear between the carbon-containing ions, and for the pairs Mg-Fe, K-Ca, and Ca-Fe.
- Cometary materials have considerably higher carbon contents than carbonaceous meteorites - interesting for discussions of their *similar* age or origin.
- Cometary material is significantly different (based on the 10 used ion signals) than the CC meteorites, yielding e. g., 100% correct discrimination by KNN + rdCV..

Search for CHNO containing substances on cometary particle surfaces

Data

- ❍ $n1$ = 30 spectra measured at a cometary particle (*Sai*).
- ❍ $n2$ = 60 spectra measured at Au background.
- ☐ *m* = 665 variables (**data set 2**), ion counts for 665 mass bins, covering all 322 possible ion formulae with C, H, N, O in the mass range 12 to 72. Normalized to sum=100.
- ◆ The 2 object classes are well separated (PCA, KNN).

Methods

- ● Search for marker variables indicating ions from CHNO compounds in cometary material (significant higher intensities at cometary particle than at background).
- ■ Statistical t- and u-test (-log(*p*)).
- ■ Random forest (MDA, mean decreasing accuracy).
- ■ D-PLS (regression coefficients of linear discriminant variab

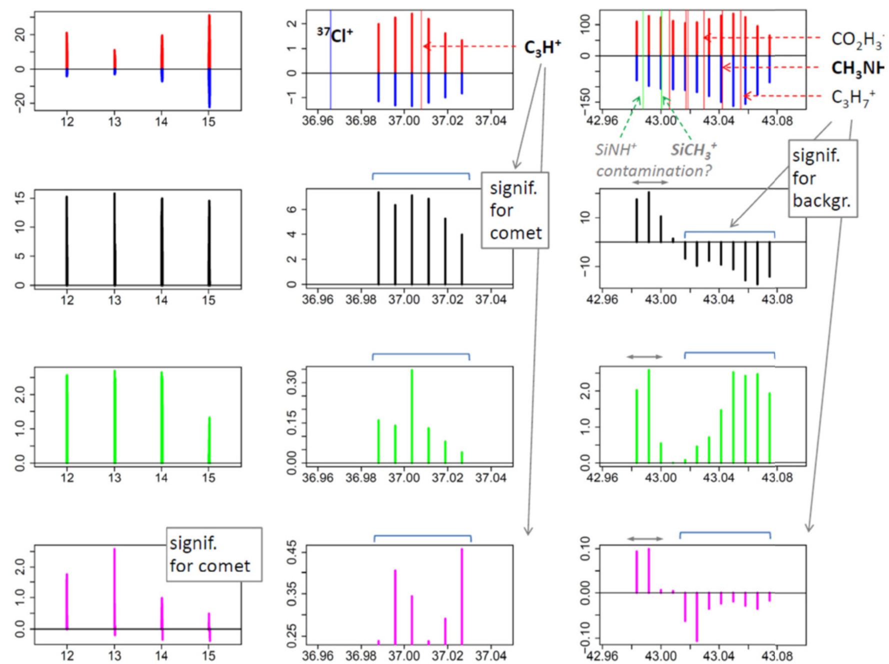
Examples — mass 12 – 15; mass 37; mass 43

Class means
comet, red;
background, blue;
ion counts, sum 100

t-test
***sgn* (-log(*p*))** ; *p* for H_0 ;
sgn = +1 if
$mean_{COMET} \geq mean_{BACKGR}$
else *sgn* = -1

Random forest (RF)
MDA,
Mean **D**ecreasing **A**ccuracy,
variable importance for
class discrimination

D-PLS
Standardized regression coefficients of discriminant variable;
5 PLS components optim.
by D-PLS-rdCV [7]



Tentative interpretation

☞ Ions, probably **from the cometary material**:

{ } not separable by mass; coded as **obvious** or *guess*

C^+; CH^+; {CH_2^+, *N^+*}; {**CH_3^+**, *NH^+*}; **$C_2H_2^+$**; {**$C_2H_3^+$**, *CNH^+*};

C_3^+; C_3H^+; {**$C_3H_2^+$**, *C_2N^+*}; {*CH_2CN^+, $C_3H_4^+$*}; *CO_2^+*; ***C_4^+***; ...

☞ Ions, probably from background:
saturated or lowly unsaturated CH ions:
e. g., $C_3H_{5\text{-}9}^+$, $C_4H_{7\text{-}9}^+$, $C_5H_{7\text{-}12}^+$

Results

- Cometary particle surfaces contain CH(NO) compounds.
- No distinct organic substance classes are evident from the data; a complex mixture of unsaturated organic compounds may be present.
- These results from multivariate data evaluation do not contradict the presence of high molecular weight structures [8].
- The applied methods for characterizing the variable importance are complementary. For the used data, t-test and u-test yield almost identical results, and appeared more user-friendly (in terms of interpretability) than RF or DPLS.

References

[1] Kissel J., et al.: Space Sci. Rev., **128**, 823 (2007)
[2] Langevin Y., et al.: Icarus, **271,** 76 (2016)
[3] Hornung K., et al.: Planetary and Space Science, **133**, 63 (2016)
[4] Hilchenbach M., et al.: The Astrophysical Journal Letters, **816**: L32 (2016)
[5] http://www.esa.int/spaceinimages/Images/2015/07/Comet_on_14_July_2015_NavCam
[6] Varmuza K., Filzmoser P.: Introduction to multivariate statistical analysis in chemometrics, CRC Press, Boca Raton, FL, USA (2009)
[7] Varmuza K., et al.: Chemom. Intell. Lab. Syst., **138**, 64 (2014)
[8] Fray N., et al.: Nature, **528**, 72 (2016)

This work is supported by the Austrian Science Fund (FWF), project P 26871 - N20.

Acknowledgments. COSIMA was built by a consortium led by the Max-Planck-Institut für Extraterrestrische Physik, Garching, Germany, in collaboration with the Laboratoire de Physique et Chimie de l'Environnement et de l'Espace, Orléans, France, the Institut d'Astrophysique Spatiale, CNRS/Université Paris Sud, Orsay, France, the Finnish Meteorological Institute, Helsinki, Finland, the Universität Wuppertal, Wuppertal, Germany, von Hoerner und Sulger GmbH, Schwetzingen, Germany, the Universität der Bundeswehr, Neubiberg, Germany, the Institut für Physik, Forschungszentrum Seibersdorf, Seibersdorf, Austria, the Institut für Weltraumforschung, Österreichische Akademie der Wissenschaften, Graz, Austria and is led by the Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany. The support of the national funding agencies of Germany (DLR, grant 50QP1302), France (CNES), Austria, Finland and the ESA Technical Directorate is gratefully acknowledged. The authors thank the other members of the COSIMA team for their contributions, and C. Koeberl, F. Brandstätter and L. Ferrière from the Vienna Natural History Museum for providing meteorite samples.

Abstract

COMET DUST COMPOSITION EXPLORED BY CHEMOMETRIC METHODS USING MASS SPECTRAL DATA FROM COSIMA/ROSETTA

Kurt Varmuza[1*], Donia Baklouti[2], Anais Bardyn[3,4], Hervé Cottin[3], Cécile Engrand[2], Peter Filzmoser[1], Nicolas Fray[3], Martin Hilchenbach[5], Irene Hoffmann[1], Jochen Kissel[5], Paola Modica[4], Johan Silén[6], Sandra Siljeström[7], Oliver Stenzel[5]

[1] Vienna University of Technology, Vienna, Austria; [2] Université Paris Sud, Orsay, France
[3] Université Paris Est Créteil et Université Paris Diderot, Créteil, France; [4] Université d'Orléans, Orléans, France
[5] Max-Planck-Institute for Solar System Research, Göttingen, Germany
[6] Finnish Meteorological Institute, Helsinki, Finland; [7] SP Technical Research Institute of Sweden, Borås, Sweden

* e-mail: kurt.varmuza@tuwien.ac.at; www.lcm.tuwien.ac.at/comecs/

The spacecraft Rosetta of the European Space Agency was launched in March 2004, arrived in August 2014 near the comet 67P/Churyumov-Gerasimenko, and escorted the comet until September 2016 [1]. The instrument COSIMA [2] onboard of Rosetta collected within these two years about 35,000 cometary dust particles from the inner coma of the comet [3]. The particles were imaged, the pictures sent to Earth, and classified [4]. Some hundred particles were analyzed onboard by COSIMA with secondary ion mass spectrometry (SIMS) using a time-of-flight (TOF) mass analyzer (mass resolution about 1400 at mass 100). More than 30,000 mass spectra (each with ion count data for 130,000 flight-time intervals of 2 ns, m/z up to 6000) were sent to Earth. Data analysis is challenging and can be supported by the application of chemometric methods.

SIMS data characterize the chemical composition of a surface, and contain information about elements and molecules, as well as isotope abundances. The main tasks in data preprocessing are [3]: Re-calibration of the mass scale; peak recognition and fitting; discrimination of spectra measured on cometary particles and on the metal background; elimination of signal contributions from contaminations. Multivariate data analyses are capable to support the spectra selection, for instance by one-class classification or by non-negative matrix factorization (NMF). Characterization of the mineral contents (based on element ion counts) is convenient by using KNN classification [5]. The search for ions that are characteristic for comet material is related to variable selection techniques as used in biomarker searches; a promising approach is the random forest method. The peak patterns in the mass range 12 to 15 indicate a high-molecular, carbon-rich material, made from crosslinked, highly diverse molecules [6]. The results are consistent with the assumption of a common origin of comets and carbonaceous (carbon-rich) meteorites - both built in the early time of the solar system.

Acknowledgments. Supported by the Austrian Science Fund (FWF), project P 26871-N20. Thanks to the members of the COSIMA team for their contributions.

[1] Schulz R.; et. al. Comet 67P/Churyumov-Gerasimenko sheds dust coat accumulated over the past four years. Nature **2015**, *518*, 216-218.
[2] Kissel J.; et al. COSIMA – High resolution time-of-flight secondary ion mass spectr. for the analysis of cometary dust particles onboard ROSETTA. Space Sci. Rev. **2007**, *128*, 823-867.
[3] Hilchenbach M.; et al. Comet 67P/Churyumov–Gerasimenko: Close-up on dust particle fragments. The Astro-physical Journal Letters **2016**, *816*, L32.
[4] Langevin Y.; et al. Typology of dust particles collected by the COSIMA mass spectrometer in the inner coma of 67P/Churyumov-Gerasimenko. Icarus **2016**, *271*, 76-97.
[5] Varmuza K.; et al. KNN classification - evaluated by repeated double cross validation: Recognition of minerals relevant for comet dust. Chemom. Intell. Lab. Syst. **2014**, *138*, 64-71.
[6] Fray N.; et al. High-molecular-weight organic matter in the particles of comet 67P/Churyumov-Gerasimenko. Nature **2016**, *528*, 72-74.