A&A 600, A115 (2017)
DOI: 10.1051/0004-6361/201630015

Astronomy & Astrophysics

Multivariate statistical analysis of OSIRIS/Rosetta spectrophotometric data of comet 67P/Churyumov-Gerasimenko

D. Perna[1], M. Fulchignoni[1], M. A. Barucci[1], S. Fornasier[1], C. Feller[1], J. D. P. Deshapriya[1], P. H. Hasselmann[1], H. Sierks[2], C. Barbieri[3], P. L. Lamy[4], R. Rodrigo[5,6], D. Koschny[7], H. Rickman[8,9], M. A'Hearn[10], J.-L. Bertaux[11], I. Bertini[3], G. Cremonese[12], V. Da Deppo[13], B. Davidsson[14], S. Debei[15], J. Deller[2], M. De Cecco[16], M. R. El-Maarry[17], M. Fulle[18], O. Groussin[19], P. J. Gutierrez[20], C. Güttler[2], M. Hofmann[2], S. F. Hviid[21], W.-H. Ip[22,23], L. Jorda[19], H. U. Keller[21,24], J. Knollenberg[21], R. Kramm[2], E. Kührt[21], M. Küppers[25], L. M. Lara[20], M. Lazzarin[3], J. J. Lopez Moreno[20], F. Marzari[3], G. Naletto[26,27,13], N. Oklay[2], N. Thomas[17], C. Tubiana[2], and J.-B. Vincent[2]

(Affiliations can be found after the references)

Received 4 November 2016 / Accepted 15 December 2016

ABSTRACT

Context. The ESA Rosetta mission explored comet 67P/Churyumov-Gerasimenko in 2014–2016, following its target before and after the perihelion passage on 13 August 2015. The NAC camera of the OSIRIS imaging system allowed to map the nucleus surface acquiring images with different filters in the visible wavelength range.

Aims. Here we study the spectrophotometric behaviour of the nucleus by a multivariate statistical analysis, aiming to distinguish homogeneous groups and to constrain the bulk composition.

Methods. We applied the G-mode clustering algorithm to 16 OSIRIS data cubes acquired on 5–6 August 2014 (mostly covering the northern hemisphere) and 2 May 2015 (mostly covering the southern hemisphere), selected to have complete coverage of the comet's surface with similar observing conditions.

Results. We found four similar homogeneous groups for each of the analysed cubes. The first group corresponds to the average spectrophotometric behaviour of the nucleus. The second (spectrally redder) and the third (spectrally bluer) groups are found in regions that were already found to deviate from the average terrain of the comet by previous studies. A fourth group (characterised by enhancements of the flux at 700–750 nm and 989 nm, possibly due to H_2O^+ and/or NH_2 emissions) seems connected with the cometary activity rather than with the bulk composition.

Conclusions. While our aim in this work was to study the spectrophotometric behaviour of the nucleus of 67P/Churyumov-Gerasimenko as a whole, we found that a follow-up application of the G-mode to smaller regions of the surface could be useful in particular to identify and study the temporal evolution of ice patches, as well as to constrain the composition and physical processes behind the emission of dust jets.

Key words. comets: individual: 67P/Churyumov-Gerasimenko – methods: data analysis – techniques: photometric

1. Introduction

The European Space Agency's Rosetta mission performed its exploration of comet 67P/Churyumov-Gerasimenko (67P/C-G hereafter) between the beginning of August 2014 and the end of September 2016. Besides the contribution of the in situ measurements provided by the Philae module immediately after its landing (Bibring et al. 2015), our understanding of the physical and compositional properties of the comet nucleus has been obtained mostly by analysis of data collected by the Optical, Spectroscopic and Infrared Remote Imaging System (OSIRIS; Keller et al. 2007) and the Visual IR Thermal Imaging Spectrometer (VIRTIS; Coradini et al. 2007) onboard the main spacecraft. Many works have been devoted to the study of morphological structures (Thomas et al 2015; Vincent et al. 2015; El-Maarry et al. 2015, 2016) and photometric and spectrophometric properties (Fornasier et al. 2015; Oklay et al. 2015, 2016; Barucci et al. 2016; Deshapriya et al. 2016; Feller et al. 2016). Sierks et al. (2015) presented the first analysis on the possible colour variability on the comet's surface, while Fornasier et al. (2015) were the first to study the spectrophotometric slopes over the nucleus surface of the northern hemisphere, and to associate them to three different groups of terrains (blue, moderately red and red). Filacchione et al. (2016) used pre-landing complete VIRTIS spectra (0.4–5 μm) to analyse the global surface composition. They pointed out some indicators of the spectral behaviour and defined three classes with different spectral properties (smooth and active areas, dust covered areas, and depressions).

In this work, we have applied a multivariate statistical analysis to the complete surface of the comet to distinguish homogeneous groups. This was done without grouping criteria a priori and to constrain the bulk composition, based on a spectrophotometric analysis of OSIRIS images.

2. Observations and data reduction

Here we present observations of 67P/C-G acquired with the narrow-angle camera (NAC) of OSIRIS on 5–6 August 2014 and 2 May 2015, for a complete coverage of the nucleus surface. These datasets were chosen as they have similar observing conditions to each other. The conditions are similar not only in terms of spatial resolution (~2.1–2.4 m/px) but also of phase angle coverage (~49°–62°), with the comet almost fulfilling the NAC field of view. A similar phase angle coverage is mandatory in the analysis, as an important phase reddening effect was observed on the comet's surface (Fornasier et al. 2015, 2016; Feller et al. 2016). The May 2015 data were taken at a closer heliocentric distance (1.73 au) than the Aug. 2014 ones (3.60 au),

however they are the first colour sequences acquired over a rotational period covering the southern hemisphere of the comet, that started to be visible from the Rosetta spacecraft only after the comet's equinox. The August 2014 observations mostly cover the equatorial-north regions of the nucleus. More information on the observing conditions can be found in Fornasier et al. (2015) and Fornasier et al. (2016), for the 2014 and 2015 data respectively.

The data were reduced using the OSIRIS standard pipeline up to level 3B, including correction for bias, flat field, geometric distortion, absolute flux calibration and conversion to radiance factor, following the data reduction steps described in Tubiana et al. (2015). We generated simulated images of the solar incidence, emission and phase angles, using the 3D shape models based on the stereophotogrammetric (SPG, Preusker et al. 2015) and stereo-photoclinometry (SPC, Jorda et al. 2016) analysis. For the spectrophotometric analysis we produce colour cubes of the surface by stacking registered and illumination-corrected images. For the illumination correction, we used a simple Lommel-Seeliger disk law,

$$D(i, e) = \frac{2\cos(i)}{\cos(i) + \cos(e)}, \quad (1)$$

dividing the images in radiance factor by the disk function D.

Illumination corrected images of a given observing sequence were co-registered using the F22 NAC filter (centred at 649.2 nm) as reference. A python script was developed to perform the co-registration based on an adaptation of the python code described by Van Der Walt et al. (2011). The co-registration produces some spurious data because of small differences in the viewing angle of pixels in the successive monochromatic images: this affects in particular those pixels which lie at abrupt transitions between illuminated and shadow regions. Hence we applied a numerical filter to eliminate from the following analysis the pixels which could not be superimposed correctly. Although May 2015 observations were performed with all the 11 filters of the NAC camera, we apply our analysis only to the seven filters that are common between the two datasets (filters F22, F23, F24, F27, F28, F41 and F71, ranging from 480.7 nm to 989.3 nm, cf. Table B.1).

3. Data analysis

3.1. Method

In order to analyse the heterogeneity of the bulk compositional behaviour of 67P/C-G, we applied the G-mode multivariate statistical analysis (Coradini et al. 1977; Gavrishin et al. 1992; Fulchignoni et al. 2000) to the data described above. The G-mode is a classification technique that allows the user to distinguish homogeneous groups within a population of $N_{\rm tot}$ objects described by a given number M of variables without a priori grouping criteria, taking into account the instrumental errors in measuring the variables and taking also into account the independence of the variables in characterising the groups. The complete classification procedure consists in the following phases: 1) the original multivariate population is collapsed into a univariate population through successive transformations. The new variable, g_j, is normalised to a quasi-Gaussian distribution with mean 0 and variance 1; 2) N' out of the $N_{\rm tot}$ elements are selected with a test of the hypothesis of appurtenance to the first group, that comes out through an automatic iterative selection process when a "starter" is defined within the analysed population; 3) the procedure is then iterated on the $N_{\rm tot} - N'$ objects left, in order to identify the second group, composed by $N'' \leq (N_{\rm tot} - N')$ objects. The iteration proceeds on the $N''' \leq N_{\rm tot} - (N' + N'')$ objects, and so on. We note that the in the G-mode technique: i) the instrumental error in measuring the variables is taken into account in the normalization of the g_j variables in each of the found homogeneous groups: if the variance of one variable describing the objects belonging to a given group is lower than the instrumental error, this is assumed as the minimum estimate of the "true" variance of the variable and the normalization and the selection procedures are repeated; ii) the g_j variable is scaled by a parameter R, inversely proportional to the sum of the squares of the coefficients of the correlation matrix of the M variables characterising the elements of the group under definition: the scaling takes into account the degree of dependence of the variables in each group and define the number of the "real" degrees of freedom of the group as $f = N \times M \times R$; iii) the selection criterion to attribute an object to a given group is based on the statistical inference rules, the only a priori choice in the decision process is the confidence level (defined as $1 - \alpha$, where α is called the error of first type and measures the probability to refuse the tested hypothesis, considered wrong when is true). The user selects the confidence level, which corresponds to a given critical value q of the variable g_j. An object belongs to the analysed group if $g_j < q$. The larger the q, the less detailed the classification is. We refer the reader to Appendix A for a more detailed description of the G-mode method.

In the image cubes analysed in the present work, the pixels (binned 2×2) are the objects and the reflectances (normalised to the F23 filter) are the variables. The instrumental errors are of about 2% for all the filters but F41 (3%) and F71 (4%), though these values basically have to be doubled because of uncertainties introduced by the shape model. We analysed a total of 16 cubes coming from observations taken on 5–6 August 2014 (seven cubes) and 2 May 2015 (nine cubes), which allowed us to obtain complete coverage of the comet's surface at very similar resolutions (~4.5 m/px after the image rebinning).

3.2. Results

We applied the G-mode analysis to each cube (each of them containing from $\sim 2 \times 10^5$ to $\sim 5 \times 10^5$ binned pixels) and, at a confidence level of 95% (corresponding to a q value of two), we found four similar homogeneous groups of pixels in each cube. Owing to the fact that the comet nucleus images are taken under different viewing angles, the total number of the analysed pixels in each image is different, as well as the number of pixels in each group. Figure 1 presents the results for the 16 analysed cubes, where different colours are associated to the four homogeneous groups we found. In Table B.1, the mean value, the 3σ deviation of each variable and the number of objects in each group found applying the G-mode to the 16 analysed images (cubes) are reported. The variables content of each of the four groups is very similar in all of the sixteen cubes: they differ at most of 1–2% for the reflectance values having a lower instrumental error (F24, F22, F27, F28) and of 2–5% for those with higher instrumental error (F41, F71). To describe the spectrophotometric behaviour of comet 67P/C-G, separately for 2014 and 2015 observations, we adopted the average of the normalised reflectance values $R_{i\lambda}$ of each group, weighted with the inverse of their variance $\sigma^2_{R_{i,\lambda}}$

$$R_\lambda = \frac{\sum_1^T p_i \cdot R_{i,\lambda}}{\sum_1^T p_i} \text{ where } p_i = \sigma^{-2}_{R_{i,\lambda}} \text{ and } \sigma_{R_\lambda} = \left(\sum_1^T p_i\right)^{-1/2}, \quad (2)$$

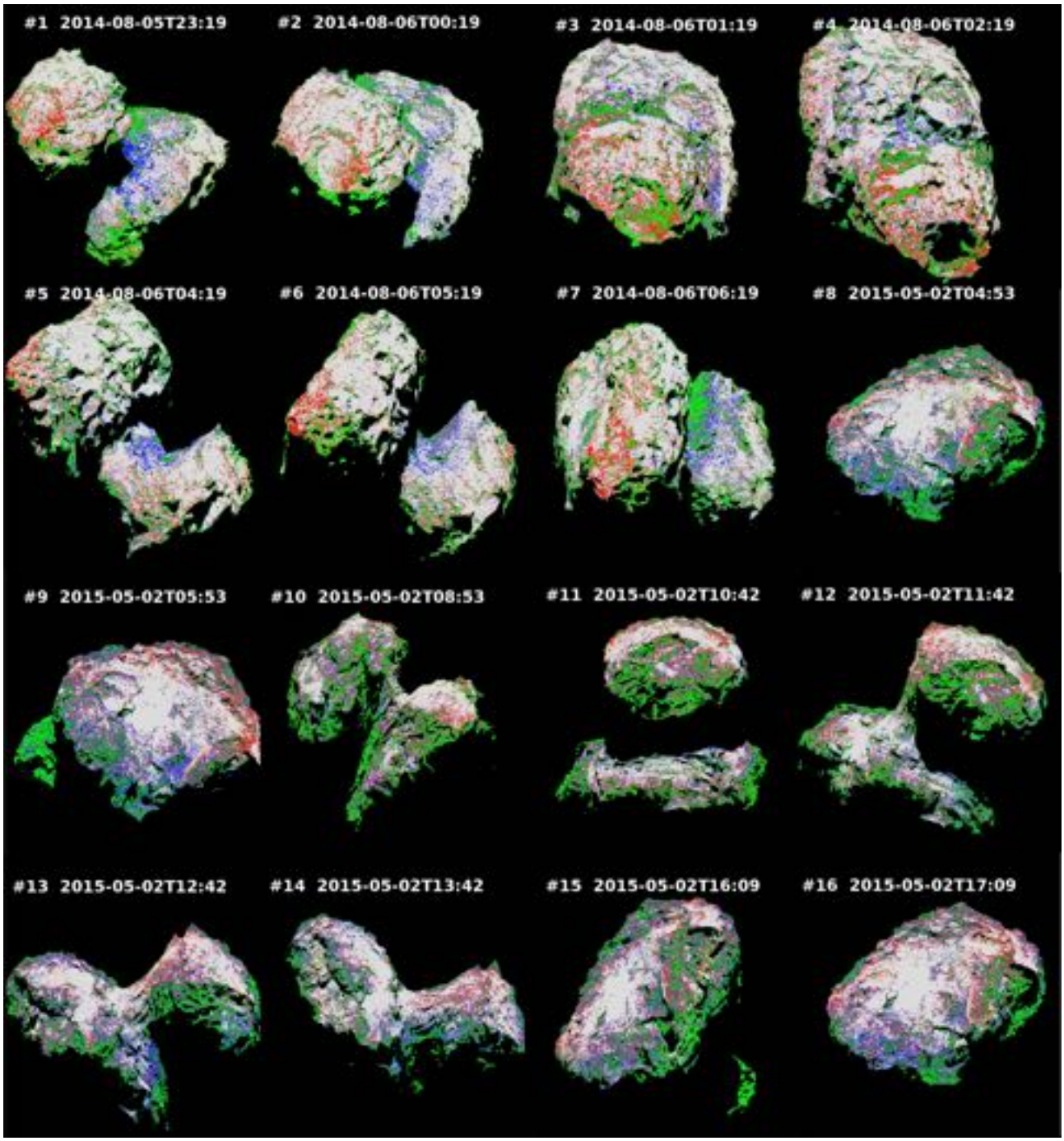

Fig. 1. Clustering results of the G-mode for the 16 analysed image cubes (grey = first group, red = second group, blue = third group, green = fourth group). The acquisition start date and time are reported for each cube.

where λ is the central wavelength of the NAC filters used in this analysis, while T (the number of cubes) is equal to seven for the August 2014 dataset and nine for the May 2015 dataset. These normalised reflectance mean values are reported in Table B.1 and represented in Figs. 2 and 3.

The general trends of all the groups are very similar in the two epochs. In particular the spectrophotometric behaviour of the first group is practically identical. This group is by far the most populous after our clustering analysis, and clearly represents the average terrain of the nucleus. In the next section we will discuss the physical meaning of the other groups, as well as the subtle variations presented by their spectra obtained in 2014 and 2015.

4. Discussion

Fornasier et al. (2015) have already studied the 5−6 August 2014 data by analysing the spectral slope in the 535−882 nm range, and proposed the existence of three groups of terrains: a "blue" group with slopes of 11−14%/(100 nm) in the considered wavelength range (and for the given phase angle of ∼50°), a "moderately red" group with slopes of 14−18%/(100 nm) and a "red" group with slopes >18%/(100 nm). They interpreted bluer regions as presenting a higher abundance of water ice at the (sub)surface. Conversely, a more spectrally red behaviour has been associated to higher abundances of organic material (Filacchione et al. 2016), and has been found mostly

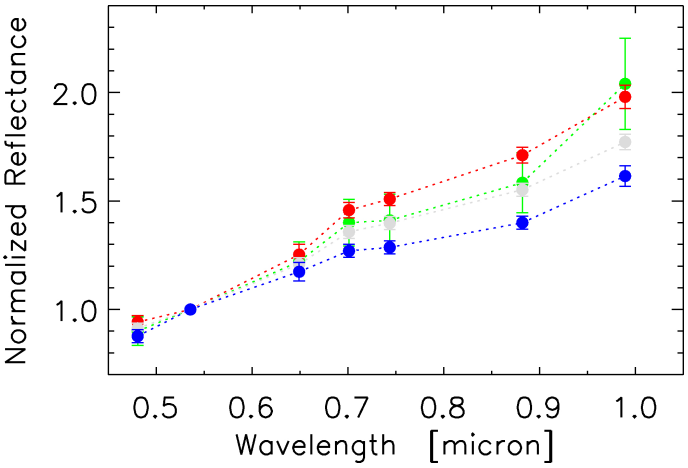

Fig. 2. Normalised mean reflectance of the four spectral groups identified by G-mode analysis for 5–6 August 2014 data.

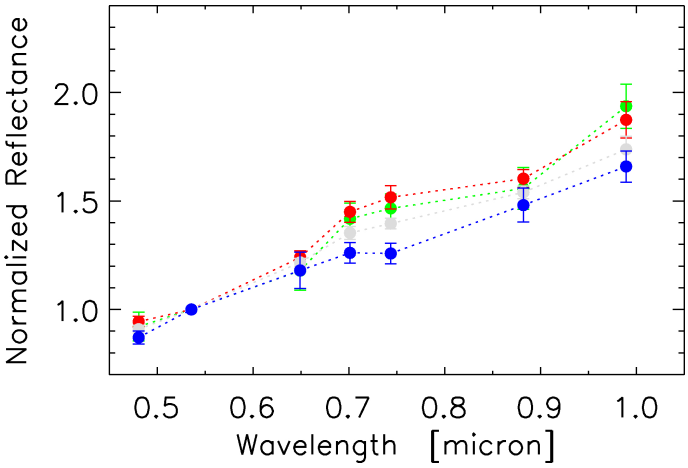

Fig. 3. Normalised mean reflectance of the four spectral groups identified by G-mode analysis for 2 May 2015 data.

on dusty regions (Fornasier et al. 2015). It is noteworthy that Fornasier et al. (2016) analysed the 2 May 2015 data in the same 535−882 nm spectral range, and found that the colour of the nucleus of 67P/C-G became globally less red compared to 2014 observations.

The first three out of the four spectrophotometric groups we identified with our G-mode analysis seem overall in very good agreement (cf. Table B.2) with the three types of terrain proposed by Fornasier et al. (2015). The first group represents the average terrain and corresponds to the globally distributed dark material discussed by Capaccioni et al. (2015). In agreement with Fornasier et al. (2015, 2016 – we refer the reader to these works for a more detailed analysis of the correspondences between surface colour and morphological regions of the 67P/C-G nucleus, which are defined in Thomas et al. 2015; and El-Maarry et al. 2015, 2016), regions like Apis, Maftet and the borders of the Hatmehit depression present a redder spectral slope (i.e. with a considerable abundance of the second group identified by the G-mode algorithm), while other regions like Hapi, Babi, Hathor and Bes look globally blue or neutral/blue (i.e. with an important presence of the third group coming out of our clustering).

The physical interpretation of the fourth group we identify is less obvious. We stress that this group is mostly characterised by an important surge of the reflectance at 989 nm compared to that at 882 nm. As Fornasier et al. (2015, 2016) analysed slopes only in the 535−882 nm range, our fourth group (which presents a behaviour similar to the average surface in this spectral range) basically falls in their "moderately red" terrains. The spectral behaviour of the fourth group deviates from that of the first one (i.e. the average terrain) also at about 700−750 nm, where a more pronounced flux excess seems to be present. The enhancements of the flux at 700−750 nm and 989 nm were already noticed by Fornasier et al. (2015), who interpreted them as due to cometary emissions (possibly H_2O^+ and/or NH_2). Such flux excesses were observed basically all over the nucleus, but in particular in shadowed areas, where the strong nucleus signal is absent. We note that, looking at its distribution on the image cubes (cf. Fig. 1), it appears that the fourth group identified by G-mode preferentially lies in regions close to shadows or in "bad" illumination conditions (e.g. due to a residual misalignment of the monochromatic images composing each cube). We stress that in the G-mode clustering, in order to diminish the noise and so to improve the robustness of our analysis, we eliminated those pixels with fluxes less than 10% and 25% of the nucleus average (values assumed as "high-pass" filters) for 2014 and 2015 data, respectively. Increasing such threshold from 10% to 25% was necessary due to the increased cometary activity in May 2015 with respect to August 2014. However, in Fig. 4 we present a detail of the cube #14 (2015-05-02T13.42.29) where the threshold was kept to 10%. One should notice that in this way pixels belonging to jets rather than to the surface are also kept for the G-mode analysis, and that these jets are dominated by the fourth group. The above suggests that the nature of such fourth group is related to a contamination from the cometary activity and not (only) to the spectrophotometric properties of the surface.

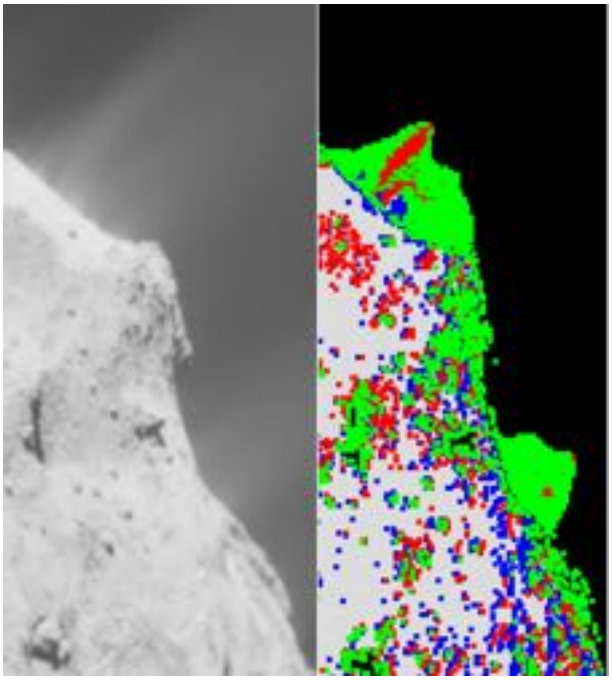

Fig. 4. Enlargement of cube #14 (2015-05-02T13.42.29): when pixels belonging to jets departing from the surface are kept in the G-mode clustering analysis, such pixels (as well as those close to shadows) are dominated by the fourth group, enhancing the possible interpretation of this group as related to the cometary activity.

Concerning the jets showed in Fig. 4, we also call attention to the presence of a "red tail" composed by pixels belonging to the second group identified by G-mode analysis. Following

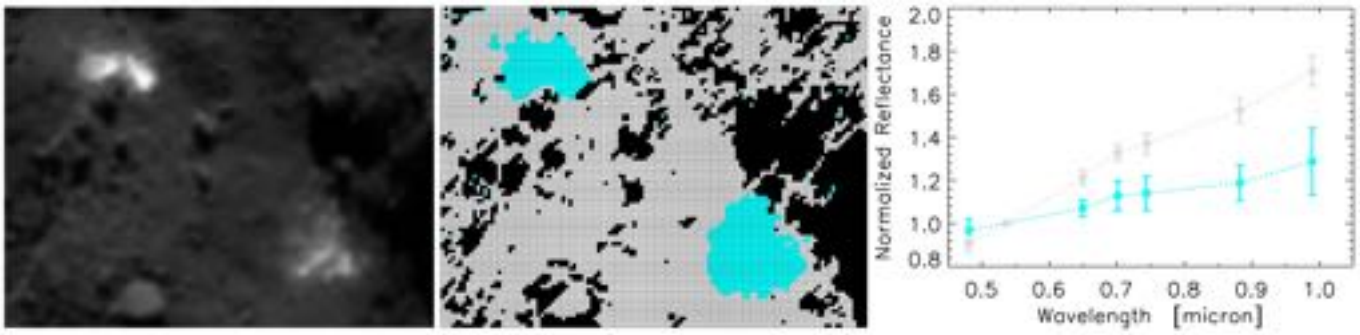

Fig. 5. G-mode automatic detection of icy patches on a selected region of cube #10 (2015-05-02T08.53.48). The reflectance spectra of the two identified spectral groups are reported in the *right panel*. We stress that the associated 3σ deviations are larger than those in Figs. 2 and 3, as here we are considering a small portion of a cube instead of a weighted average of several (7 and 9 cubes, respectively) much larger datasets.

what said above, these could correspond to a collimated jet of organic-rich dust. Hence our clustering technique offers also the possibility to investigate the structure, composition and dynamics (if several data cubes are available) of dust jets: we plan a future work specifically devoted to the study of this issue.

As stated above, the average colours of the four groups we identified are in agreement (within the error bars) between the 2014 and 2015 datasets (acquired when 67P/C-G was at heliocentric distances of 3.60 au and 1.73 au, respectively). Some colour changes are, however, evident. As noticed by Fornasier et al. (2016), the 535−882 nm spectral slope globally decreases in 2015 observations (cf. Figs. 2 and 3, and Table B.2), probably because the increased activity (67P/C-G reached its perihelion on 13 August 2015) removed important quantities of dust partially exposing the subsurface water ice. The reflectance at 989 nm does not follow this trend, but again this could be explained in terms of an increased contamination by gaseous emissions at these wavelengths. It is also interesting to note how the relative proportions of the groups identified by G-mode analysis changed between 2014 and 2015 observations: for 5−6 August 2014 data, the first group contains 70−80% of the total number of pixels in each cube, the second one 8−13%, the third 5−7% and the fourth 8−11%. For 2 May 2015 data, the first group contains 53−67% of the total number of pixels in each cube, the second one 12−17%, the third 11−14% and the fourth 9−20%. The third and fourth groups basically doubled their relative weights in the pixels populations: interpreting such variations in terms of the increased cometary activity, this means that more ice-rich regions are visible in the 2015 images of the nucleus of 67P/C-G, and that the contamination of the surface colours by the coma (dust and gas) became much more important.

5. Ice patches

In order to test the stability of the G-mode in analysing datasets of different size, we choose a small portion (120 × 90 pixels, Fig. 5) of the cube #10 (2015-05-02T08.53.48) where Fornasier et al. (2016) found two ice patches, that are too small (each of them contains just ∼0.15% of the total number of pixels in the cube) to be put in evidence by the analysis of the whole cube (in this case the ice patches were classified within the third group, still appropriate for relatively flat spectra when considering the uncertainties associated to our variables especially at the longer wavelengths). The G-mode analysis applied to this selected portion of the cube separated the pixels, at a confidence level of 99%, into two groups whose reflectance spectra are also shown in Fig. 5. The most populous group corresponds to the average spectrophotometric behaviour of the comet's surface (the first group discussed in the previous sections), while the second group identifies the two ice patches (with a flatter spectrophotometric behaviour than the third group identified from the analysis of the whole cube, as expected for exposed water ice features). This result indicates that the G-mode may be a powerful tool in finding small local variations in the spectrophotometry of the comet's surface, even if the scale of these variations is negligible in the global context (eventually subdividing the image cubes in smaller portions).

6. Conclusions

We applied the G-mode multivariate statistical analysis on 16 7-filter data cubes acquired (on 5−6 August 2014 and 2 May 2015, for a complete coverage of the nucleus with similar observing conditions) with the OSIRIS/NAC camera onboard the Rosetta mission, to study the spectrophotometric behaviour of the cometary nucleus of 67P/Churyumov-Gerasimenko.

Without a priori grouping criteria and at a confidence level of 95%, we identified four homogeneous groups, presenting a similar spectrophotometric behaviour for all of the 16 analysed data cubes.

The first three groups are in good agreement with the three types of terrain proposed by Fornasier et al. (2015) based on the analysis of spectral slope in the 535−882 nm range. The first group is by far the most populous after our clustering analysis, and corresponds to the average dark material of the surface, while the second and the third groups correspond to globally redder (organic-richer) and bluer (ice-richer) regions, respectively.

The nature and distribution of the fourth group seem correlated with the cometary activity. Such group is in fact characterised by an enhancement of the flux at 700−750 nm and 989 nm (possibly due to a spectral contamination by H_2O^+ and/or NH_2 emissions) and dominates low-illuminated areas where a strong signal from the nucleus is absent (including jets departing from the surface).

We plan to use the G-mode algorithm for a follow-up investigation of selected regions (to investigate the composition, morphology and temporal evolution of dust jets and/or ice patches) of OSIRIS data cubes, which could provide more detailed information about the active processes ongoing on comet 67P/C-G.

Acknowledgements. OSIRIS was built by a consortium of the Max-Planck-Institut für Sonnensystemforschung, Göttingen, Germany, CISAS–University of Padova, Italy, the Laboratoire d'Astrophysique de Marseille, France, the Instituto de Astrofísica de Andalucia, CSIC, Granada, Spain, the Research and Scientific Support Department of the European Space Agency, Noordwijk, The Netherlands, the Instituto Nacional de Técnica Aeroespacial, Madrid, Spain, the Universidad Politéchnica de Madrid, Spain, the Department of Physics and Astronomy of Uppsala University, Sweden, and the Institut für Datentechnik und Kommunikationsnetze der Technischen Universität Braunschweig, Germany. The support of the national funding agencies of Germany (DLR), France (CNES), Italy (ASI), Spain (MEC), Sweden (SNSB), and the ESA Technical Directorate is gratefully acknowledged. M.A. acknowledges NASA funding through Jet Propulsion Laboratory contract No. 1267923 and from the Akademie der Wissenschaften zu Göttingen. W.-H.Ip acknowledges the Ministry of Science and Technology, Taiwan (grant No. NSC 102-2112-M-008) and Macau University of Science and Technology (grant No. FDCT 017/2014/A1). We thank the ESA teams at European Space Astronomy Centre, European Space Operations Centre, and European Space Research and Technology Centre for their work in support of the Rosetta mission. Rosetta/OSIRIS data are available through the ESA's Planetary Science Archive (PSA, http://www.cosmos.esa.int/web/psa/rosetta).

References

Abramowitz, M. & Stegun I. A. 1972, Handbook of Mathematical Functions (New York: Dover)
Bagrov, A. A. 1978, Tr. Gigronet Nauch. Issl. Zentra, 44, 3
Barucci, M. A., Filacchione, G., Fornasier, S., et al. 2016, A&A, 595, A102
Bibring, J.-P., Taylor, M. G. G. T., Alexander, C., et al. 2015, Science, 349, 0671
Capaccioni, F., Coradini, A., Filacchione, G., et al. 2015, Science, 347, 628
Coradini, A., Fulchignoni, M., Fanucci, O., & Gavrishin, A.I. 1977, Computers and Geosciences, 3, 85
Coradini, A., Capaccioni, F., Drossart, P., et al. 2007, Space Sci. Rev., 128, 529
Deshapriya, J. D. P., Barucci, M. A., Fornasier, S., et al. 2016, MNRAS, 462, S274
El-Maarry, M. R., Thomas, N., Giacomini, L., et al. 2015, A&A, 583, A26
El-Maarry, M. R., Thomas, N., Gracia-Berna, A., et al. 2016, A&A, 593, A110
Feller, C. Fornasier, S., Hasselmann, P. H., et al. 2016, MNRAS, 462, S287
Filacchione, G., Capaccioni, F., Ciarniello, M., et al. 2016, Icarus, 274, 334
Fornasier, S., Hasselmann, P. H., Barucci, M. A., et al. 2015, A&A, 583, 30
Fornasier, S., Mottola, S., Keller, H. U., et al. 2016, Science, 354, 1566
Fulchignoni, M., Birlan, M., & Barucci, M. A. 2000, Icarus, 146, 204
Gavrishin, A. I., Coradini, A., & Cerroni, P. 1992, Earth, Moon and Planets, 59, 141
Jorda, L., Gaskell, R., Capanna, C., et al. 2016, Icarus, 277, 257
Keller, H. U., Barbieri, C., Lamy, P., et al. 2007, Space Sci. Rev., 128, 433
Oklay, N., Vincent, J.-B., Sierks, H., et al. 2015, A&A, 583, A45
Oklay, N., Vincent, J.-B., Fornasier, S., et al. 2016, A&A, 586, A80
Preusker, F., Scholten, F., Matz, K.-D., et al. 2015, A&A, 583, A33
Sierks, H., Barbieri, C., Lamy, P. L., et al. 2015, Science, 347, 1044
Thomas, N., Sierks, H., Barbieri, C., et al. 2015, Science, 347, 0440
Tubiana, C., Güttler, C., Kovacs, G., et al. 2015, A&A, 583, A46
Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Comp. Sci. Eng., 13, 22
Vincent, J.-B., Bodewits, D., Besse, S., et al. 2015, Nature, 523, 63

[1] LESIA – Observatoire de Paris, PSL Research University, CNRS, Sorbonne Universités, UPMC Univ. Paris 06, Univ. Paris Diderot, Sorbonne Paris Cité, 5 place Jules Janssen, 92195 Meudon, France
e-mail: davide.perna@obspm.fr

[2] Max-Planck-Institut für Sonnensystemforschung, Justus-von-Liebig-Weg, 3 37077 Göttingen, Germany
[3] University of Padova, Department of Physics and Astronomy, Vicolo dell'Osservatorio 3, 35122 Padova, Italy
[4] Laboratoire d'Astrophysique de Marseille, UMR 7326 CNRS & Université Aix-Marseille, 38 rue Frédéric Joliot-Curie, 13388 Marseille Cedex 13, France
[5] Centro de Astrobiologia (INTA-CSIC), 28850 Torrejon de Ardoz, Madrid, Spain
[6] International Space Science Institute, Hallerstrasse 6, 3012 Bern, Switzerland
[7] Scientific Support Office, European Space Research and Technology Centre/ESA, Keplerlaan 1, Postbus 299, 2201 AZ Noordwijk ZH, The Netherlands
[8] Department of Physics and Astronomy, Uppsala University, Box 516, 75120 Uppsala, Sweden
[9] PAS Space Reserch Center, Bartycka 18A, 00716 Warszawa, Poland
[10] University of Maryland, Department of Astronomy, College Park, MD 20742-2421, USA
[11] LATMOS, CNRS/UVSQ/IPSL, 11 boulevard d'Alembert, 78280 Guyancourt, France
[12] INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
[13] CNR-IFN UOS Padova LUXOR, via Trasea, 7, 35131 Padova , Italy
[14] Jet Propulsion Laboratory, M/S 183-301, 4800 Oak Grove Drive, Pasadena, CA 91109, USA
[15] Department of Mechanical Engineering University of Padova, via Venezia 1, 35131 Padova, Italy
[16] University of Trento, via Mesiano 77, 38100 Trento, Italy
[17] Physikalisches Institut der Universität Bern, Sidlerstr. 5, 3012 Bern, Switzerland
[18] INAF–Osservatorio Astronomico di Trieste, via Tiepolo 11, 34014 Trieste, Italy
[19] Aix-Marseille Université, CNRS, LAM (Laboratoire d'Astrophysique de Marseille), UMR 7326, 38 rue Frédéric Joliot-Curie, 13388 Marseille Cedex 13, France
[20] Instituto de Astrofisica de Andalucia (CSIC), c/ Glorieta de la Astronomia s/n, 18008 Granada, Spain
[21] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Planetenforschung, Asteroiden und Kometen, Rutherfordstrafle 2, 12489 Berlin, Germany
[22] National Central University, Graduate Institute of Astronomy, 300 Chung-Da Rd, 32054 Chung-Li, Taiwan
[23] Space Science Institute, Macau University of Science and Technology, Macau, PR China
[24] Institut für Geophysik und Extraterrestrische Physik (IGEP), Technische Universität Braunschweig, Mendelssohnstr. 3, 38106 Braunschweig, Germany
[25] Operations Department, European Space Astronomy Centre/ESA, PO Box 78, 28691 Villanueva de la Canada, Madrid, Spain
[26] University of Padova, Department of Information Engineering, via Gradenigo 6/B, 35131 Padova, Italy
[27] Center of Studies and Activities for Space (CISAS) "G. Colombo", University of Padova, via Venezia 15, 35131 Padova, Italy

Appendix A: The G-mode algorithm

The G-mode multivariate statistical analysis separates a sample of $N_{\rm tot}$ objects into J homogeneous groups containing N_l objects each ($N_{\rm tot} = \sum_{l=1}^{J} N_l$). Each object is described by M variables i ($i = 1, ..., M$). The data are arranged in a $N_{\rm tot} \times M$ matrix, the mean value ($\bar{x}_i$) and the variance (σ_i^2) for each variable, and the $M \times M$ correlation matrix of the variables are computed. Each object will be then represented by a new variable z_j^2

$$z_j^2 = \sum_{i=1}^{M} z_{ij}^2 = \sum_{i=1}^{M} \frac{\left(x_{ij}^2 - \bar{x}_i^2\right)}{\sigma_i^2}, \tag{A.1}$$

where x_{ij} is the ith variable of the jth sample. When x_{ij} are independent and normally distributed, z_j^2 follow a χ^2 distribution with M degrees of freedom. If x_{ij} are not independent, the dependence of the variables is represented by

$$R = \frac{M}{\sum_{k,m=1}^{M} r_{k,m}}, \tag{A.2}$$

where $r_{k,m}$ are the elements of the correlation matrix. The z_j^2 variable is then defined

$$z_j^2 = R_l \cdot \sum_{i=1}^{M} z_{ij}^2, \tag{A.3}$$

and it follows a χ^2 distribution with $f = N \times M \times R$ degrees of freedom (Bagrov 1978) that can be transformed (Abramowitz & Stegun 1972) in a standard normal distribution by the parameter

$$g_j = \sqrt{2z_j^2} - \sqrt{2f - 1}. \tag{A.4}$$

The identification of homogeneous groups of objects consists in an iterative procedure based on a test of the hypothesis of appurtenance of the jth object to a "zero group". The centre of this zero group (the starter of the procedure) is obtained as the sum of the three closer objects (those having the minimum value of z),

$$z_{p,q,t} = \min\left(\sum_{i=1}^{M}\left[\left(z_{pi} - z_{qi}\right)^2 + \left(z_{pi} - z_{ti}\right)^2 + \left(z_{qi} - z_{ti}\right)^2\right]\right), \tag{A.5}$$

where z_{pi}, z_{qi} and z_{ti} are the normalised values of the ith variable of the pth, qth and tth object, respectively. Where the three objects satisfying the above equation are found, the mean value and the variance of each variable are computed

$$\bar{x}_{j*} = \frac{1}{3} \cdot \sum_{j=1}^{3} x_{ij} \qquad \sigma_{i*}^2 = \frac{1}{2} \cdot \sum_{i=1}^{3} \left(x_{ij}^2 - \bar{x}_{i*}^2\right). \tag{A.6}$$

The values of z_j^2, f and g_j are recomputed substituting the values obtained in A.6 into into A.1–A.3, and the value given by A.4 is compared with a critical value q selected a priori. Thus, we test the hypothesis of appurtenance of a given object described by its g_j value to the group defined by A.6. The N' objects with $g_j < q$ are considered to belonging to the same homogeneous group. The values of the mean and the variance of these N' objects are computed and a new iteration is started using these values in the expressions A.1 to A.4. The iterations are stopped when N' and R' are not changing in two successive cycles. The same procedure is then applied to the remaining $N - N'$ objects and the grouping is continued until the number of objects left is less than three.

Appendix B: Tables

Table B.1. Results of the G-mode clustering.

# Cube (N_{cube})	Group (N_{group})	R_{F24} (481 nm)	R_{F22} (649 nm)	R_{F27} (701 nm)	R_{F28} (744 nm)	R_{F41} (882 nm)	R_{F71} (989 nm)
1 (304399)	1 (223943)	0.915 ± 0.048	1.211 ± 0.060	1.358 ± 0.060	1.399 ± 0.060	1.552 ± 0.066	1.772 ± 0.072
	2 (28944)	0.940 ± 0.066	1.252 ± 0.090	1.438 ± 0.066	1.487 ± 0.066	1.676 ± 0.072	1.926 ± 0.096
	3 (17949)	0.887 ± 0.060	1.170 ± 0.072	1.283 ± 0.048	1.310 ± 0.048	1.432 ± 0.054	1.626 ± 0.072
	4 (33563)	0.898 ± 0.132	1.219 ± 0.174	1.368 ± 0.210	1.394 ± 0.216	1.553 ± 0.276	1.910 ± 0.360
2 (327353)	1 (260985)	0.914 ± 0.054	1.210 ± 0.072	1.362 ± 0.078	1.403 ± 0.072	1.561 ± 0.078	1.789 ± 0.090
	2 (24256)	0.941 ± 0.072	1.249 ± 0.120	1.466 ± 0.078	1.516 ± 0.078	1.721 ± 0.084	2.004 ± 0.144
	3 (14923)	0.882 ± 0.066	1.175 ± 0.120	1.267 ± 0.072	1.280 ± 0.072	1.392 ± 0.078	1.607 ± 0.120
	4 (27189)	0.904 ± 0.156	1.218 ± 0.222	1.403 ± 0.306	1.418 ± 0.330	1.591 ± 0.378	2.021 ± 0.702
3 (410190)	1 (288080)	0.913 ± 0.054	1.207 ± 0.078	1.357 ± 0.078	1.396 ± 0.078	1.549 ± 0.084	1.776 ± 0.096
	2 (51525)	0.937 ± 0.084	1.239 ± 0.132	1.455 ± 0.102	1.507 ± 0.096	1.721 ± 0.132	2.004 ± 0.204
	3 (26582)	0.875 ± 0.072	1.195 ± 0.150	1.273 ± 0.102	1.275 ± 0.096	1.377 ± 0.114	1.615 ± 0.174
	4 (44003)	0.906 ± 0.168	1.229 ± 0.234	1.433 ± 0.300	1.454 ± 0.372	1.654 ± 0.528	2.055 ± 0.738
4 (475503)	1 (340051)	0.914 ± 0.060	1.204 ± 0.084	1.355 ± 0.078	1.390 ± 0.078	1.542 ± 0.084	1.747 ± 0.090
	2 (56984)	0.940 ± 0.090	1.247 ± 0.144	1.455 ± 0.102	1.502 ± 0.096	1.702 ± 0.108	1.984 ± 0.144
	3 (32162)	0.872 ± 0.084	1.185 ± 0.162	1.276 ± 0.102	1.277 ± 0.090	1.395 ± 0.108	1.637 ± 0.186
	4 (46306)	0.908 ± 0.192	1.229 ± 0.288	1.452 ± 0.324	1.460 ± 0.366	1.650 ± 0.402	2.191 ± 0.570
5 (358995)	1 (279686)	0.911 ± 0.060	1.206 ± 0.078	1.350 ± 0.078	1.388 ± 0.078	1.543 ± 0.084	1.755 ± 0.090
	2 (29128)	0.947 ± 0.090	1.264 ± 0.126	1.457 ± 0.096	1.501 ± 0.090	1.704 ± 0.108	1.957 ± 0.132
	3 (18925)	0.866 ± 0.090	1.156 ± 0.126	1.253 ± 0.090	1.270 ± 0.078	1.378 ± 0.102	1.629 ± 0.162
	4 (31256)	0.887 ± 0.204	1.206 ± 0.270	1.395 ± 0.312	1.380 ± 0.312	1.548 ± 0.390	2.142 ± 0.522
6 (310067)	1 (243749)	0.913 ± 0.060	1.210 ± 0.078	1.352 ± 0.084	1.393 ± 0.078	1.544 ± 0.090	1.764 ± 0.102
	2 (26422)	0.943 ± 0.090	1.262 ± 0.132	1.458 ± 0.096	1.511 ± 0.090	1.717 ± 0.102	2.002 ± 0.204
	3 (14300)	0.869 ± 0.084	1.170 ± 0.126	1.249 ± 0.084	1.266 ± 0.084	1.367 ± 0.102	1.593 ± 0.168
	4 (25596)	0.890 ± 0.198	1.215 ± 0.264	1.377 ± 0.330	1.390 ± 0.330	1.557 ± 0.372	2.078 ± 0.792
7 (387105)	1 (285185)	0.912 ± 0.066	1.215 ± 0.090	1.371 ± 0.090	1.416 ± 0.084	1.577 ± 0.096	1.812 ± 0.102
	2 (34612)	0.946 ± 0.078	1.262 ± 0.144	1.490 ± 0.090	1.552 ± 0.084	1.763 ± 0.102	2.045 ± 0.120
	3 (23684)	0.877 ± 0.078	1.190 ± 0.132	1.263 ± 0.084	1.274 ± 0.072	1.373 ± 0.090	1.588 ± 0.120
	4 (43624)	0.903 ± 0.198	1.226 ± 0.270	1.413 ± 0.360	1.430 ± 0.408	1.612 ± 0.462	2.081 ± 0.576
Average 5–6 August 2014	1	0.913 ± 0.024	1.209 ± 0.030	1.358 ± 0.030	1.398 ± 0.030	1.552 ± 0.030	1.772 ± 0.036
	2	0.942 ± 0.030	1.253 ± 0.048	1.458 ± 0.036	1.509 ± 0.030	1.711 ± 0.036	1.980 ± 0.054
	3	0.877 ± 0.030	1.174 ± 0.042	1.270 ± 0.030	1.286 ± 0.030	1.400 ± 0.030	1.615 ± 0.048
	4	0.900 ± 0.066	1.221 ± 0.090	1.400 ± 0.108	1.411 ± 0.120	1.584 ± 0.138	2.040 ± 0.210
8 (313322)	1 (195698)	0.907 ± 0.060	1.220 ± 0.066	1.364 ± 0.084	1.404 ± 0.072	1.552 ± 0.096	1.763 ± 0.180
	2 (38602)	0.932 ± 0.060	1.244 ± 0.066	1.441 ± 0.096	1.508 ± 0.114	1.613 ± 0.102	1.880 ± 0.222
	3 (37883)	0.880 ± 0.066	1.191 ± 0.264	1.293 ± 0.126	1.292 ± 0.108	1.513 ± 0.240	1.724 ± 0.216
	4 (41139)	0.920 ± 0.156	1.195 ± 0.324	1.425 ± 0.108	1.474 ± 0.144	1.586 ± 0.402	1.953 ± 0.222
9 (355663)	1 (225091)	0.907 ± 0.060	1.220 ± 0.078	1.362 ± 0.084	1.402 ± 0.072	1.553 ± 0.096	1.763 ± 0.180
	2 (48128)	0.941 ± 0.090	1.240 ± 0.084	1.456 ± 0.150	1.508 ± 0.156	1.622 ± 0.144	1.917 ± 0.288
	3 (43528)	0.870 ± 0.090	1.183 ± 0.264	1.271 ± 0.156	1.282 ± 0.138	1.487 ± 0.246	1.674 ± 0.234
	4 (38916)	0.917 ± 0.246	1.182 ± 0.342	1.401 ± 0.312	1.450 ± 0.192	1.553 ± 0.276	1.905 ± 0.324
10 (209223)	1 (110344)	0.912 ± 0.072	1.214 ± 0.096	1.342 ± 0.096	1.386 ± 0.108	1.526 ± 0.096	1.708 ± 0.210
	2 (27785)	0.958 ± 0.090	1.242 ± 0.090	1.462 ± 0.126	1.537 ± 0.192	1.582 ± 0.138	1.841 ± 0.270
	3 (30231)	0.870 ± 0.084	1.174 ± 0.276	1.223 ± 0.168	1.216 ± 0.156	1.458 ± 0.240	1.595 ± 0.198
	4 (40863)	0.920 ± 0.234	1.174 ± 0.246	1.391 ± 0.312	1.437 ± 0.228	1.523 ± 0.396	1.871 ± 0.342
11 (240094)	1 (131700)	0.911 ± 0.060	1.214 ± 0.078	1.343 ± 0.072	1.383 ± 0.102	1.527 ± 0.096	1.712 ± 0.210
	2 (32516)	0.958 ± 0.078	1.238 ± 0.078	1.466 ± 0.180	1.530 ± 0.210	1.598 ± 0.144	1.867 ± 0.270
	3 (33759)	0.863 ± 0.078	1.179 ± 0.306	1.221 ± 0.192	1.218 ± 0.180	1.460 ± 0.240	1.606 ± 0.216
	4 (42119)	0.928 ± 0.216	1.185 ± 0.270	1.412 ± 0.312	1.476 ± 0.240	1.549 ± 0.300	1.908 ± 0.354
12 (256335)	1 (143944)	0.910 ± 0.060	1.214 ± 0.078	1.344 ± 0.078	1.384 ± 0.096	1.531 ± 0.096	1.721 ± 0.180
	2 (32267)	0.952 ± 0.078	1.244 ± 0.072	1.453 ± 0.150	1.520 ± 0.168	1.597 ± 0.108	1.853 ± 0.222
	3 (36026)	0.866 ± 0.078	1.176 ± 0.228	1.236 ± 0.174	1.232 ± 0.156	1.474 ± 0.240	1.637 ± 0.210
	4 (44098)	0.926 ± 0.186	1.191 ± 0.252	1.413 ± 0.312	1.470 ± 0.204	1.559 ± 0.258	1.903 ± 0.288
13 (267910)	1 (160395)	0.909 ± 0.060	1.215 ± 0.060	1.349 ± 0.102	1.388 ± 0.102	1.535 ± 0.096	1.729 ± 0.192
	2 (35955)	0.952 ± 0.090	1.233 ± 0.090	1.461 ± 0.180	1.540 ± 0.222	1.599 ± 0.144	1.875 ± 0.282
	3 (36160)	0.868 ± 0.084	1.178 ± 0.228	1.239 ± 0.162	1.218 ± 0.156	1.472 ± 0.240	1.642 ± 0.222
	4 (35400)	0.925 ± 0.216	1.175 ± 0.252	1.432 ± 0.318	1.493 ± 0.252	1.566 ± 0.606	1.971 ± 0.366
14 (238292)	1 (159547)	0.909 ± 0.060	1.216 ± 0.078	1.350 ± 0.090	1.391 ± 0.084	1.537 ± 0.096	1.731 ± 0.180
	2 (27819)	0.941 ± 0.078	1.234 ± 0.090	1.440 ± 0.138	1.510 ± 0.156	1.590 ± 0.132	1.864 ± 0.264
	3 (26868)	0.876 ± 0.084	1.174 ± 0.222	1.261 ± 0.144	1.256 ± 0.138	1.477 ± 0.240	1.673 ± 0.228
	4 (24058)	0.919 ± 0.168	1.163 ± 0.240	1.402 ± 0.312	1.449 ± 0.198	1.539 ± 0.222	1.932 ± 0.282
15 (293354)	1 (162521)	0.908 ± 0.060	1.218 ± 0.084	1.357 ± 0.090	1.400 ± 0.090	1.544 ± 0.096	1.741 ± 0.180
	2 (49150)	0.947 ± 0.102	1.239 ± 0.096	1.453 ± 0.174	1.530 ± 0.192	1.614 ± 0.150	1.887 ± 0.276
	3 (37798)	0.861 ± 0.114	1.181 ± 0.246	1.263 ± 0.150	1.252 ± 0.138	1.486 ± 0.240	1.678 ± 0.204
	4 (43885)	0.921 ± 0.222	1.178 ± 0.306	1.416 ± 0.312	1.472 ± 0.216	1.579 ± 0.294	1.978 ± 0.324
16 (333309)	1 (220936)	0.908 ± 0.060	1.220 ± 0.072	1.364 ± 0.072	1.405 ± 0.072	1.552 ± 0.096	1.761 ± 0.180
	2 (46468)	0.933 ± 0.072	1.238 ± 0.090	1.444 ± 0.132	1.510 ± 0.156	1.614 ± 0.132	1.898 ± 0.276
	3 (35405)	0.878 ± 0.072	1.181 ± 0.270	1.286 ± 0.126	1.284 ± 0.114	1.504 ± 0.240	1.718 ± 0.222
	4 (30500)	0.917 ± 0.180	1.179 ± 0.246	1.421 ± 0.318	1.470 ± 0.168	1.580 ± 0.252	1.981 ± 0.270
Average 2 May 2015	1	0.909 ± 0.018	1.217 ± 0.018	1.353 ± 0.030	1.396 ± 0.024	1.540 ± 0.030	1.738 ± 0.060
	2	0.944 ± 0.024	1.240 ± 0.030	1.450 ± 0.048	1.517 ± 0.053	1.603 ± 0.042	1.874 ± 0.084
	3	0.871 ± 0.030	1.180 ± 0.084	1.261 ± 0.048	1.258 ± 0.048	1.481 ± 0.078	1.659 ± 0.072
	4	0.921 ± 0.066	1.179 ± 0.090	1.418 ± 0.072	1.466 ± 0.066	1.558 ± 0.096	1.937 ± 0.102

Table B.2. Spectral slopes in the 535−882 nm range.

G-mode group	5–6 August 2014 %/(100 nm)	2 May 2015 %/(100 nm)	Fornasier et al. (2015) %/(100 nm)
1	15.9	15.6	14-18 (“moderately red”)
2	20.5	17.4	>18 (“red”)
3	11.5	13.9	11–14 (“blue”)
4	16.8	16.1	