

Asteroid taxonomic classification in the *Gaia* photometric system (Research Note)

J. Warell and C.-I. Lagerkvist

Department of Astronomy and Space Physics, Uppsala University, Box 515, 751 20 Uppsala, Sweden
e-mail: johan.warell@astro.uu.se

Received 22 February 2006 / Accepted 13 March 2007

ABSTRACT

Aims. We evaluate the prospects of performing automatic taxonomic classification of asteroids in a proposed broad and medium band photometric system of *Gaia*.

Methods. The study is based on asteroid colors from the Eight-Color Asteroid Survey (ECAS) and CCD spectra from the Small Mainbelt Asteroid Spectral Survey II (SMASSII). The success of the *Gaia* photometric systems for taxonomic classification is evaluated using supervised classification techniques and mean taxonomic class spectra in the Tholen and Bus&Binzel taxonomic systems. Our supervised classification method is based on rms differences between individual asteroid spectra and taxonomic mean spectra and provides probability estimates of membership in all taxonomic classes.

Results. We find that both photometric systems of *Gaia* are able to discriminate between all of the twelve Tholen asteroid classes for noise-free data. The medium band system is able to discriminate between the majority of the 26 SMASSII asteroid classes in case of high quality photometric data. For both the Tholen and Bus & Binzel taxonomies we find that about 25% of the asteroids are spectrally more similar to another taxonomic class in a best-fit sense, though the differences within the three major complexes (C, S and X) are 1–10%.

Conclusions. Among the two main existing taxonomies, the *Gaia* photometric system is found to be best suited for the Bus & Binzel taxonomy. The medium band system is the preferred system for all but the faintest objects. The classification method employed here results in more concentrated taxonomic class domains in principal component space, and mean taxonomic spectra that are formed from less divergent class members, than the case for the nominal classification systems. It provides statistical probability estimates for class memberships and naturally reflects the fact that asteroid spectral shapes form a continuum in principal component space.

Key words. minor planets, asteroids – methods: data analysis – space vehicles: instruments

1. Introduction

Asteroid taxonomy has evolved with the development of more sophisticated and accurate photometric and spectroscopic observations. The method of principal component analysis (PCA) and a minimal tree algorithm based on 7-color spectra (ECAS, Eight Color Asteroid Survey) published by Zellner et al. (1985) were used by Tholen (1984) to define 12 taxonomic classes. This classification scheme set a standard which has since been followed and extended. With infrared data from IRAS, Barucci et al. (1987) were able to refine and extend the taxonomic system. They used G-mode multivariate statistical analysis based on a number of descriptive variables for the classification. The classifications of Tholen and Barucci are thoroughly described by Tholen & Barucci (1989). Tedesco et al. (1989) provide a comprehensive summary of different methods of asteroid classifications made to that date. With data from their spectroscopic survey SMASSII (Small Main belt Asteroid Spectral Survey II, Bus & Binzel 2002a), Bus & Binzel (2002b) primarily used PCA analysis to develop a feature-based taxonomy with 26 classes based on an expanded Tholen taxonomy reflecting the increased discriminative ability of CCD spectra.

The European Space Agency's *Gaia* mission due for launch in 2011 will have a fundamental impact on asteroid science: it will measure the masses of about 100, directly measure sizes of about 1000, provide data for derivation of spin properties and shapes for about 10 000, and improve the orbits provide spectra

for 300 000 asteroids to visual magnitude about 20 (Cellino et al. 2005). In preparation for the data return from *Gaia* we evaluate here the potential of the two proposed photometric systems for taxonomic classification of asteroids. We have considered two of the most extensive spectral surveys available, and the corresponding taxonomies derived therefrom: the Eight Color Asteroid Survey (ECAS; Zellner et al. 1985) and the SMASSII survey (Bus & Binzel 2002a). We perform supervised classification in the broad band (C1B) and medium band (C1M) photometric systems of *Gaia* (Jordi et al. 2006), which allows derivation of revised classifications in the Tholen (1984) and Bus & Binzel (2002b) taxonomic systems including probabilities.

2. Transformations to the *Gaia* photometric systems

The *Gaia* C1B and C1M photometric systems and data acquisition procedure is described in detail in Jordi et al. (2006). The broad passband system C1B is composed of five filters and the medium passband C1M system is composed of 14 filters with central wavelengths between 325 and 965 nm. It is possible that the proposed photometric systems for *Gaia* will not be the final choices, but the analysis and main results will still be applicable, particularly with respect to the classification method.

From the Small Bodies Node (PDS Asteroid Data Archive, <http://www.psi.edu/pds/>), 531 ECAS objects with

complete color information and 1332 full wavelength range spectra of numbered asteroids from SMASSII were selected for analysis. The transformation to the *Gaia* systems was performed by deriving a magnitude-scale spectrum from the ECAS colors, fitting the ECAS data with the average of a piecewise cubic hermite interpolating polynomial and cubic spline function, followed by interpolation to find magnitudes at the *Gaia* wavelengths, and adding a constant such that the 550 nm value was set to zero (as in the ECAS system). For SMASSII, the median intensity of the spectra from data within the *FWHM* range of the *Gaia* filters was calculated and normalized to unity at 550 nm (as for SMASSII spectra). Since the wavelength coverage of SMASSII is narrower than for C1M, it is not possible to calculate the transformed intensities for the four bluest and the reddest bands, and the number of bands used in the classification analysis was thus necessarily limited to 9 (we refer to this as the “cropped” C1M system).

3. Discriminative ability of the Tholen and Bus & Binzel taxonomies

As the full data set will not be possible to reduce rigorously for the first version of the *Gaia* object catalog, initial classification of asteroid spectra will be made as part of the automatic processing pipeline. These preliminary class assignments will thus have to be made on the basis of a previously defined taxonomic system using a fast and accurate algorithm.

We evaluated which of the Tholen and Bus & Binzel (hereafter termed B&B) taxonomies is the most discriminant in the *Gaia* systems, and whether the corresponding taxonomic classes are different enough to discriminate with the *Gaia* photometric systems. The root-mean-square (rms) differences between all possible taxonomic class pair combinations in each of the ECAS, C1M and C1B systems were calculated. The same was done for the B&B taxonomy in the SMASSII, C1M and C1B systems. We found that the C1M system is more discriminative than C1B for ECAS spectra in the Tholen taxonomy, but the nominal ECAS system is more discriminative than either the C1M or C1B systems. For the B&B taxonomy, the C1B system is more discriminative than either C1M or the nominal SMASSII system.

Considering the nominal spectra on which the Tholen and B&B taxonomies are based, Tholen’s system is thus more discriminative since it contains fewer classes and encompasses a greater spectral range. The derived discriminative capability of the B&B taxonomy in the cropped C1M system is significantly reduced to the true case as we do not have sufficient information to evaluate the B&B taxonomy for the full range of C1M wavelengths.

We can conclude that, in the *Gaia* C1B photometric system, the Tholen taxonomy is more discriminative than the B&B taxonomy, and would be the optimal choice for faint asteroids. The reverse is true for the C1M system, which would be the optimal choice for bright asteroids. However, in the *Gaia* data analysis, we choose to use the B&B taxonomy for classification, since the full taxonomy is possible to use for the brighter minor planets with high-quality spectra. We accept that the classification ambiguity will increase and that the corresponding probabilities will decrease for fainter asteroids.

4. Optimal method of taxonomic classification

One of the weaknesses of both the B&B and the Tholen taxonomies is the assignment of a specific taxonomic class to

an asteroid without recognition of the probability of membership in the given class. This is a natural result of the mode of classification.

We have tested a number of spectral parameterizations to identify an automatic supervised classification method suitable for pipeline processing, which produces classification results most similar to those obtained by Bus & Binzel (2002b), and which naturally provides probability estimates for each classification. The method reflects the natural continuity between the different taxonomic classes, expressed in terms of the probabilities of membership which may be high for more than one class.

The selected parametrization which forms the basis for the classification is an rms value Δ defined as

$$\Delta(j, k) = (1/N \sum_{i=1}^N I_{\text{diff}}^2)^{1/2}$$

where

$$I_{\text{diff}} = I_{\text{tax}}(i, k) - [(c - 1) + I_{\text{ast}}(i, j)]$$

is used in the comparison to magnitude-scale ECAS data, and

$$I_{\text{diff}} = I_{\text{tax}}(i, k) - c \times I_{\text{ast}}(i, j)$$

is used in the comparison to intensity-scale SMASSII data, as well as in the presented classification in the *Gaia* system (Table 1). Here, j is the asteroid, $k = 1, \dots, 26$ (12) is the number of B&B (Tholen) classes, I_{tax} and I_{ast} are the values of the mean taxonomic and asteroid spectra at the respective band wavelengths, $i = 1, 2, \dots, N$ is the number of spectral bands (5 for C1B and 9 for the cropped C1M), and c is a scaling factor optimized to find the best fit. This factor was found to have mean and standard deviation values of 0.9999 and 0.0239 for ECAS, and 1.0025 and 0.0093 for SMASSII.

To derive a probability estimate $p(j, k)$ for assignment of each taxonomic class, the ratio $R(j, k)$ of the Δ value of the asteroid j , relative to the standard deviation of the Δ values for all classes k , was calculated,

$$R(j, k) = \Delta(j, k) / \sigma([\Delta(j, 1), \Delta(j, 2), \dots, \Delta(j, k)]),$$

and used as input to the normal cumulative distribution function

$$p(j, k) = F(R(j, k) | \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{R(j, k)} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt.$$

Here, p is the probability that a single observation from a normal distribution with mean μ and standard deviation σ will fall in the interval $(-\infty, R]$. In the present case, $\mu = 0$.

The success rate $\rho(k)$ for each class k was then evaluated with respect to the classification of Tholen (1984) or Bus & Binzel (2002b), calculated as

$$\rho(k) = N_{\text{tax}}(k) / N$$

where N_{tax} is the number of asteroids for which the most probable taxonomic class k coincides with the Tholen or B&B class assignment for any given asteroid, and N is the total number of evaluated ECAS or SMASSII asteroids.

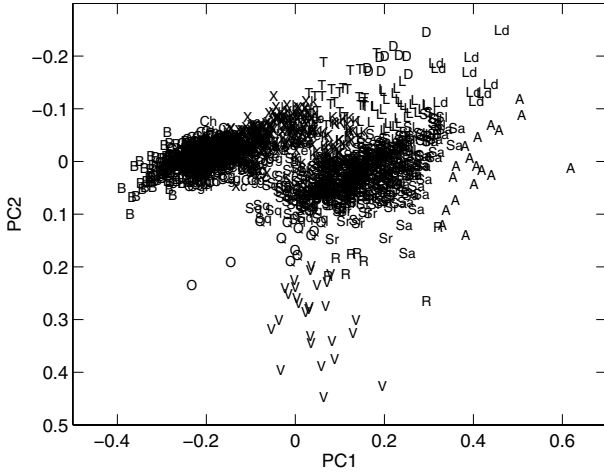


Fig. 1. The first two principal components of 1332 SMASSII asteroid spectra classified with the Δ parameter using our mean taxonomic spectra in the *Gaia* C1M system, derived from transformed Bus & Binzel (2002b) spectra. The axes have been oriented for easy comparison of relative cluster locations in Bus & Binzel.

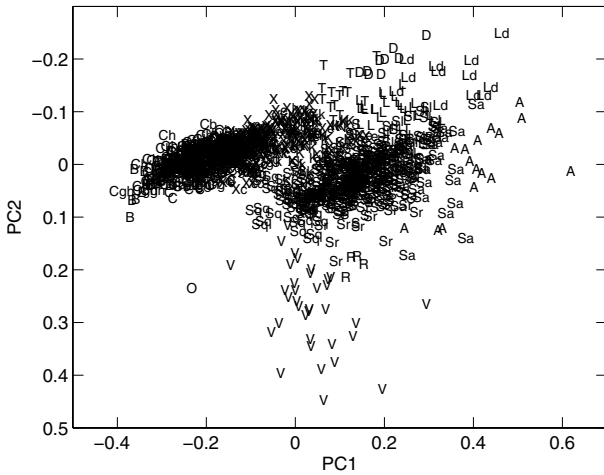


Fig. 2. As for Fig. 1, but for the nominal taxonomic classification of Bus & Binzel (2002b).

5. Results

For the C1M system, we find that the most probable classification is different from the nominal Tholen and B&B assignments for 17% and 39% of the asteroids, respectively. However, as many as 98% and 93%, respectively, of any of the three most probable classifications is the same as the nominal classification. This is a dramatic indication of the small spectral differences between many of the taxonomic classes as represented in the C1M system, particularly in the B&B taxonomy.

Differences in our classifications compared those derived by Tholen or Bus & Binzel are due to the different classification approaches used: Tholen cut the longest branches in a minimum spanning tree solution based on PCA of the ECAS colors, while Bus & Binzel employed a more complex scheme where the average spectral slope was removed before PCA was applied to quantify absorption features. In the latter case, the derived slope values and the two primary PC vectors were employed in the classification. Here, we employ the rms metric using taxonomic mean spectra as templates in the evaluation of each asteroid spectrum. It is thus primarily objects which are peripheral in their class domains that are reclassified.

Table 1. Statistics of supervised classification results in the Bus & Binzel (2002b) taxonomy for data in the SMASSII and *Gaia* C1M systems, employing parameter Δ . The data set encompasses 1332 asteroids in SMASSII on which our classification is performed. Column 2 gives the number of objects classified into a particular taxonomic class in the C1M system, Col. 3 the number of Col. 2 objects that are assigned to the same class by B&B, and Col. 4 the ratio of the two numbers. Column 5 gives the number of objects classified into a particular class by B&B and Col. 6 the ratio of Cols. 5 to 3. The last column gives the number of objects of a particular class in the C1M system to the number of (not necessarily the same) objects of the same class in the B&B classification.

Class	2	3	4	5	6	7
A	18	13	1.38	16	1.23	1.12
B	86	49	1.76	60	1.22	1.43
C	132	88	1.50	142	1.61	0.93
Cb	63	26	2.42	33	1.27	1.91
Cg	25	4	6.25	9	2.25	2.78
Cgh	52	4	13.00	15	3.75	3.47
Ch	63	54	1.17	137	2.54	0.46
D	9	8	1.12	9	1.12	1.00
K	45	18	2.50	31	1.72	1.45
L	34	23	1.48	34	1.48	1.00
Ld	11	9	1.22	12	1.33	0.92
O	2	1	2.00	1	1.00	2.00
Q	10	0	–	0	–	–
R	8	4	2.00	4	1.00	2.00
S	236	233	1.01	381	1.64	0.62
Sa	43	26	1.65	34	1.31	1.26
Sk	76	15	5.07	15	1.00	5.07
Sl	75	36	2.08	49	1.36	1.53
Sq	34	34	1.00	49	1.44	0.69
Sr	47	10	4.70	14	1.40	3.36
T	24	13	1.85	14	1.08	1.71
V	26	26	1.00	35	1.35	0.74
X	99	67	1.48	111	1.66	0.89
Xc	52	30	1.73	60	2.00	0.87
Xe	27	7	3.86	28	4.00	0.96
Xk	35	18	1.94	39	2.17	0.90

A robust check on the success of the taxonomic classification may be made from the distribution of asteroids in PC space. The first two PC scores for the distribution of B&B classifications derived in the C1M system are compared to the nominal classifications in Figs. 1 and 2. Our method on average produces smaller and better separated taxonomic class domains than either of the original B&B or Tholen classifications. Whether the classifications more accurately relate surface mineralogies or cosmochemical origin between the classes cannot presently be answered, but *Gaia* will provide information that will contribute to resolve this issue.

Statistics of the classification results for SMASSII and *Gaia* cropped C1M data in the Bus & Binzel (2002b) taxonomy are given for each class in Table 1. Here, we do not see any drastic differences in the total number of C-complex (C, Cb, Cg, Cgh, and Ch classes), S-complex (S, Sa, Sk, Sl, Sq, and Sr classes) or X-complex (X, Xc, Xe, and Xk classes) objects, though there are significant variations within the complexes. For instance, Cg, Cgh, Sk, Sr and Xe objects are several factors more numerous in our classification than in Bus & Binzel's, while Ch and S classes are drastically depleted. We note that among objects classified to a particular class in our study, the majority of the same objects are also found to be of the same class by Bus & Binzel, though drastic differences are found for the Cg, Cgh, Sk, Sr and Xe classes. Finally, while there are no Q objects according to

the B&B classification among the 1332 asteroids we selected for study, we find that 10 of these objects are most likely classified as such (the ten Q objects found by B&B were among the remaining 115 SMASSII objects not selected for study here).

We note that the probability of asteroid (4) Vesta belonging to the V class is low in our classification of SMASSII spectra in the cropped C1M system. Its location at $[-0.15, 0.18]$ in Figs. 1 and 2 is closest to the geometric center of the O-class. Our classification O68 indicates that the probability of O-class membership is 68%, followed by Q33, Sq20 and V13 (in the Tholen system, Vesta is the defining and only V class object in the ECAS sample and is thus classed as V100). This is consistent with the shape of Vestas spectrum: it is non-typical compared to average V-class members, with the NIR band being considerably shallower. However, the known number of O-type asteroids is very small (e.g., only four in the full SMASSII sample) and classifications will undoubtedly change with a greater sample and higher-precision extended-wavelength spectra.

6. Discussion

One may question the appropriateness of using the B&B classes at all rather than finding a completely new, more discriminative taxonomic system to apply to the *Gaia* systems. We have seen that the *Gaia* C1M system is sufficiently sampled to be superior for classification in the B&B system using the nominal SMASSII spectral range; thus the merits of the SMASSII feature-based classification system is exploitable in C1M as well. This is apparent in some particular classes, which are discriminative due to existence of limited-range shallow absorptions: the slope change at 400 nm for the majority of taxonomic classes, the 490 nm absorption in Xe, the subtle slope change at 550 nm for Sl, Sa and Sr, the broad 700 nm feature due to hydration in Ch and Cgh, and the slope change at 700–750 nm in Xe and Xk. Thus the only feature which is foreseen to possibly remain undetected in lower-quality *Gaia* C1M spectra is the 600 nm absorption band in the Xe and Sq classes, which is not particularly well sampled.

The *Gaia* C1B system does have the possibility to allow detection of the 700 nm hydration band, though the narrower 600 and 900 nm bands will likely remain undetected. In contrast to the SMASSII survey the UV range will be sampled and thus increase the discriminability, particularly for the C and X class complexes. The *Gaia* C1M system is superior with respect to the SMASSII system as it will sample both shorter and longer wavelengths and thus can better define the slopes and shapes of the UV and NIR absorption bands. Compared to ECAS, the wavelength coverage of C1M is nearly identical, though C1M is superior due to the greater number of bands and narrower passbands. In addition, the photometric precision and the internal consistency in photometric data between objects and over time will be drastically improved compared to ECAS and SMASSII due to the location of *Gaia* outside of Earth's atmosphere. Furthermore, the phase angles for all objects will be similar, minimizing the relative phase reddening effect.

Acknowledgements. Two anonymous reviewers are thanked for providing constructive comments which significantly improved the original manuscript. This work was supported by the Swedish National Space Board.

References

- Barucci, M. A., Capria, M. T., Coradini, A., & Fulchignoni, M. 1987, *Icarus*, 72, 304
- Bus, S. J., & Binzel, R. P. 2002a, *Icarus*, 158, 106
- Bus, S. J., & Binzel, R. P. 2002b, *Icarus*, 158, 146
- Cellino, A., et al. 2005, in *Proceedings of The Three-Dimensional Universe with Gaia* (ESA SP-576).
- Jordi, C., Høg, E., Brown, A. G. A., et al. 2006, *MNRAS*, 367, 290
- Tedesco, E. F., Williams, J. G., Matson, D. L., et al. 1989, *AJ*, 97, 580
- Tholen, D. J. 1984, *Asteroid taxonomy from cluster analysis of photometry*, Ph.D. Thesis (Tucson: University of Arizona)
- Tholen, D. J., & Barucci, M. A. 1989, in *Asteroids II*, Proceedings of the Conference, Tucson, AZ, Mar. 8–11, 1988 (Tucson: University of Arizona Press), 298
- Zellner, B., Tholen, D. J., & Tedesco, E. F. 1985, *Icarus* 61, 355