



THÈSE DE DOCTORAT

Taxonomie des astéroïdes: une synthèse probabiliste de la spectrométrie et de l'albédo à partir d'observations complètes et partielles

MAX MAHLKE

Laboratoire J.-L. Lagrange | Observatoire de la Côte d'Azur

Présentée en vue de l'obtention du grade de docteur en
Sciences de la Planète et de l'Univers
de l'Université Côte d'Azur

Dirigée par Benoit Carry
Soutenue le 14 September 2022

Devant le jury composé de
Benoit Carry, *Astronome-Adjoint, Observatoire de la Côte d'Azur, Nice, France*
Catherine Lantz, *Astronome-Adjoint, Université Paris Saclay, Paris, France*
Guy Libourel, *Professeur, Université Côte d'Azur, Nice, France*
Patrick Michel, *Directeur de Recherches, Observatoire de la Côte d'Azur, Nice, France*
Philippe Rousselot, *Professeur, Université de Franche-Comté, Besançon, France*
Maria Cristina De Sanctis, *Primo Ricercatore, Istituto Nazionale di Astrofisica, Rome, Italy*
Cristina A. Thomas, *Assistant Professor, Northern Arizona University, Arizona, USA*
Pierre Vernazza, *Chercheur CNRS, Laboratoire d'Astrophysique de Marseille, Marseille, France*

Taxonomie des astéroïdes: une synthèse probabiliste de la spectrométrie et de l'albédo à partir d'observations complètes et partielles

Asteroid Taxonomy: A Probabilistic Synthesis of Spectrometry and Albedo from Complete and Partial Observations

RAPPORTEURS

Maria Cristina De Sanctis

Primo Ricercatore, Istituto Nazionale di Astrofisica, Rome, Italy

Pierre Vernazza

Chercheur CNRS, Laboratoire d'Astrophysique de Marseille, Marseille, France

EXAMINATEURS

Benoit Carry

Astronome-Adjoint, Observatoire de la Côte d'Azur, Nice, France

Cateline Lantz

Astronome-Adjoint, Université Paris Saclay, Paris, France

Guy Libourel

Professeur, Université Côte d'Azur, Nice, France

Patrick Michel

Directeur de Recherches, Observatoire de la Côte d'Azur, Nice, France

Philippe Rousselot

Professeur, Université de Franche-Comté, Besançon, France

Cristina A. Thomas

Assistant Professor, Northern Arizona University, Arizona, USA

Résumé Les petits corps du système solaire sont les témoins de l'histoire de sa formation. La diversité de leur composition est représentative de la matière présente initialement dans le disque protoplanétaire, altérée par des centaines de millions d'années d'évolution collisionnelle, de métamorphisme thermique, d'altération aqueuse et d'irradiation par l'environnement spatial. L'histoire du système solaire est donc fossilisée dans les petits corps, et l'étude de leurs propriétés d'ensemble permet de mieux comprendre les processus à grande échelle qui ont façonné notre système planétaire. Les astéroïdes, petits corps entre le Soleil et Jupiter, sont les résidus de planétésimaux qui se sont formés dans le système solaire interne ou y ont été implanté depuis la région externe via des processus dynamiques.

La taxonomie des astéroïdes est l'outil utilisé pour décrire les différents groupes de composition et les tendances observées dans la population d'astéroïdes. Dans ce travail, je présente une nouvelle itération de la taxonomie des astéroïdes. En faisant progresser la méthode établie à l'aide d'une approche d'apprentissage automatique non supervisée, j'ai mis en œuvre trois améliorations clés par rapport à la méthode existante. Tout d'abord, la classification peut être appliquée aux observations de spectres de réflectance dans les longueurs d'onde visibles, proche infrarouge ou à la fois visibles et proche infrarouge, fournissant des classes taxonomiques cohérentes dans les trois cas. Ensuite, l'affectation des classes repose sur une méthode probabiliste plutôt qu'absolue, ce qui permet d'affiner la sélection des populations en se basant sur la probabilité que chaque population appartienne à une classe donnée. Enfin, l'albédo visuel est réintroduit en tant qu'observable dans cette classification, résolvant la dégénérescence spectrale et l'ambiguïté autour de la minéralogie des astéroïdes présentant des spectres de réflectance sans bandes caractéristiques.

La méthode est appliquée dans le cadre de la plus grande étude de composition des astéroïdes à ce jour, basée sur les spectres de réflectance et l'albédo visuel de 5906 observations d'astéroïdes. 2983 observations sont utilisées pour définir les classes de la taxonomie. La nouvelle taxonomie combine les systèmes précédemment établis de Tholen (1984), Bus et Binzel (2002a), et DeMeo et al. (2009); avec les observations et études minéralogiques de la population d'astéroïdes effectuées au cours des dernières décennies. Le cœur de la taxonomie est composé de trois complexes, les complexes d'astéroïdes carbonés C et silicatés S bien établis, ainsi que le nouveau complexe M, qui héberge une variété d'astéroïdes métallosilicatés et qui remplace le complexe X. Une nouvelle classe Z est introduite pour caractériser les astéroïdes extrêmement rouges dispersés au sein de la ceinture principale.

La taxonomie est appliquée dans le cadre de l'appariement des spectres d'astéroïdes avec ceux de météorites, tout en tenant compte des propriétés inconnues des surfaces planétaires. La connexion entre les chondrites CV/CO et les astéroïdes de type K et L est étudiée ainsi que le lien entre les chondrites ordinaires équilibrées ou non et les astéroïdes de type S. Je présente en outre deux outils d'analyse des données d'astéroïdes que j'ai développés. `rocks` simplifie l'exploration et l'acquisition de paramètres d'astéroïdes, tandis que `classy` permet la classification taxonomique des observations d'astéroïdes dans le schéma taxonomique présenté ici.

Mots-Clés astéroïdes, taxonomie, petits corps

Asteroid Taxonomy: A Probabilistic Synthesis of Spectrometry and Albedo from Complete and Partial Observations

Abstract The minor bodies of the Solar System are witnesses of the formation history of our Solar System. Their compositional diversity represents the material present in the protoplanetary disk altered through hundreds of millions of years of collisional evolution, thermal metamorphism, aqueous alteration, and irradiation from the space environment. The history of the Solar System is therefore fossilised in the minor bodies and the study of their ensemble properties provides insights into large-scale processes that shaped our planetary system. Asteroids, the minor bodies between the Sun and Jupiter, are the remnants of planetesimals which either formed in the inner Solar System or were implanted into it from the outer region via dynamical processes.

The asteroid taxonomy is a language used to describe the compositional groups and trends observed among the asteroid population. In this work, I present a new iteration of the asteroid taxonomy. By advancing the established method with an unsupervised machine learning approach, I implement three key improvements over the previous method. First, the classification can be applied to observations of visible-, near-infrared, or visible-near-infrared reflectance spectra, providing consistent taxonomic classes in all three cases. Second, the class assignments are probabilistic rather than absolute, enabling fine selections of populations based on their probabilities to belong to any class. Third, the visual albedo is reintroduced as classification observable, resolving the spectral degeneracy and mineralogical ambiguity of the featureless reflectance spectra.

The method is applied in the largest compositional study of asteroids based on reflectance spectra and visual albedo to date with 5906 observations, 2983 of which are used to derive the class scheme of the taxonomy. The new taxonomy combines the previously established frameworks of Tholen (1984), Bus and Binzel (2002a), and DeMeo et al. (2009) with the observational and mineralogical insights into the asteroid population over the past decade. The core of the taxonomy are three complexes, the well-established carbonaceous C and silicaceous S complexes as well as the new M complex, which hosts a variety of metallic-silicaceous compositions and replaces the X complex. A new class Z is introduced for extremely-red asteroids which are found throughout the Main Belt.

The method is applied to the spectral matching of asteroids and meteorites while accounting for the unknown properties of planetary surfaces. The connection between CV/CO chondrites to K- and L-type asteroids is studied as well as the link between equilibrated and unequilibrated ordinary chondrites and S-types. I further present two tools for analysis of asteroid data I developed. `rocks` simplifies the exploration and acquisition of asteroid parameters, while `classify` enables the taxonomic classification of asteroid observations in the presented taxonomic scheme.

Keywords asteroids, taxonomy, minor bodies

PUBLICATIONS

I present the list of publications which I either first-authored or contributed to before and during my PhD. A selection of figures and ideas presented in this manuscript have been published in these works.

- I **Mahlke, M.**, Carry, B., Mattei, P.-A., *Asteroid Taxonomy from Cluster Analysis of Spectrometry and Albedo*, [A&A 665](#), 2022
- II Eschrig, J., Bonal, L., **Mahlke, M.**, et al., *Investigating S-type asteroid surfaces through reflectance spectra of ordinary chondrites*, [Icarus 381](#), 2022
- III Kruk, S., Martín, P. G., Popescu, M., et al. (incl. **Mahlke, M.**), *Hubble Asteroid Hunter: I. Identifying asteroid trails in Hubble Space Telescope images*, [A&A 661](#), 2022
- IV Racero, E., Giordano, F., Carry, B. et al. (incl. **Mahlke, M.**), *ESASky SSOSS: Solar System Object Search Service and the case of Psyche*, [A&A 659](#), 2022
- V **Mahlke, M.**, Carry, B., Denneau, L., *Asteroid phase curves from ATLAS dual-band photometry*, [Icarus 354](#), 2021
- VI Morate, D., Marcio Carvano, J., Alvarez-Candal, A., et al. (incl. **Mahlke, M.**), *J-PLUS: A first glimpse at the spectrophotometry of asteroids*, [A&A 655](#), 2021
- VII Cortés-Contreras, M., Bouy, H., Solano, E., et al. (incl. **Mahlke, M.**), *The Gran Telescopio Canarias OSIRIS broad-band first data release*, [MNRAS 491](#), 2020
- VIII **Mahlke, M.**, Solano, E., Bouy, H., et al., *The ssos pipeline: Identification of Solar System objects in astronomical images*, [A&C 28](#), 2019
- IX Cortés-Contreras, M., Jiménez-Esteban, F. M., **Mahlke, M.**, et al., *Identification of asteroids using the Virtual Observatory: the WFCAM Transit Survey*, [MNRAS 490](#), 2019
- X **Mahlke, M.**, Bouy, H., Altieri, B., et al., *Mining the Kilo-Degree Survey for solar system objects*, [A&A 610](#), 2018

CONTENTS

I	Minor Bodies of the Solar System	1
1	Populations	3
1.1	Outer Solar System	4
1.2	Inner Solar System	6
2	Observables	13
2.1	Reflectance and Emission of Asteroid Surfaces	14
2.2	Spectroscopy	16
2.3	Photometry	17
2.4	Albedo	18
2.5	Phase Curve Coefficients	19
2.6	Honorary Mentions	23
3	Meteorites	25
3.1	Terminology	26
3.2	Compositional Constituents	27
3.3	Classification	31
3.4	Interplanetary Dust Particles	34
3.5	Spectral Ambiguity in Matching Asteroids and Meteorites	34
4	Taxonomy	37
4.1	Principles and Design	37
4.2	Complexes of Asteroids	39
II	Advancing the Method	47
5	Mathematical Foundation	49
5.1	Terminology	49
5.2	Dimensionality Reduction	50
5.3	Clustering	56
5.4	Dimensionality Reduction and Clustering with Missing Data	58
6	Asteroid Data and How to Manage it	65
6.1	SsODnet: A Unified Access Point for Asteroid Data	66
6.2	rocks: Quick Access to Asteroid Data	68
7	Asteroid Composition from Non-Targeted Phase Curves	73
7.1	Observations and Parameter Inference	74
7.2	Uncertainties	77
7.3	Compositional Information	82

III	A New Asteroid Taxonomy	95
8	Motivation	97
8.1	Practical Considerations	97
8.2	Compositional Considerations	98
9	Observables	101
9.1	Selection	101
9.2	Preprocessing	103
9.3	Final Dataset	106
10	Clustering	111
10.1	Application of Model to Observations	111
10.2	Latent Space	114
10.3	From Clusters to Classes	116
10.4	Data Availability	125
11	Classes	127
11.1	C-Complex: B, C, Ch, P	127
11.2	M-Complex: K, L, M	133
11.3	S-Complex: S, Q	138
12	Classification in Practice	145
12.1	Classification Tool: <code>classy</code>	145
12.2	Probabilistic Classifications	146
12.3	Degeneracies	147
12.4	Comparison to Previous Taxonomies	151
IV	Asteroids and Meteorites	155
13	Compositional Distribution of Asteroids	157
13.1	Primitive Material	159
13.2	Hydration	161
13.3	Metals and Silicates	161
13.4	Trends in Near-Earth Space	163
14	Spectral Connection of Asteroids and Meteorites	165
14.1	A Primer on Spectral Matching	166
14.2	CV-, CO Chondrites and K-, L-Types	167
14.3	Equilibrated-, Unequilibrated OC and S-Types	177
V	Conclusion	183
	Conclusion and Outlook	185
	Bibliography	189
	Acknowledgements	213

LIST OF FIGURES

Figure 1.1	Distribution of small bodies in outer Solar System.	5
Figure 1.2	Distribution of small bodies in inner Solar System.	7
Figure 1.3	Dynamical structure of Main Belt.	8
Figure 2.1	Black body spectra, atmospheric transmission, and spectrum of (3) <i>Juno</i> in the visible-near-infrared.	15
Figure 2.2	Maximum observable solar phase angle α for different asteroid populations.	20
Figure 2.3	Phase curve of (20) <i>Massalia</i> from Gehrels (1956).	21
Figure 2.4	G_1 , G_2 parameter space related to photometric slope, opposition effect, and taxonomy.	22
Figure 3.1	Reflectance spectra of silicate minerals.	28
Figure 3.2	Reflectance spectra of native elements, oxides, and CAI.	29
Figure 3.3	Classification of meteorites.	32
Figure 4.1	Visualisation of asteroid taxonomy over time.	44
Figure 5.1	Representation of simulated dataset in subspace of data-space.	51
Figure 5.2	Dimensionality reduction via principal component analysis.	53
Figure 5.3	Dimensionality reduction and clustering via principal component analysis and finite-mixture model.	55
Figure 5.4	Principal components and scores from DeMeo et al. (2009).	56
Figure 5.5	Dimensionality reduction via factor analysis.	60
Figure 5.6	Local and global dimensionality reduction of simulated dataset.	62
Figure 5.7	Dimensionality reduction and clustering via mixture of common factor analysers.	63
Figure 7.1	Phase curve of (442) <i>Eichsfeldia</i> observed by ATLAS.	75
Figure 7.2	Phase curve of (250) <i>Bettina</i> at different apparitions.	78
Figure 7.3	Magnitude dispersion due to rotation and aspect angle from DAMIT shape models.	79
Figure 7.4	Effect of insufficient coverage of opposition effect on G_1 , G_2 .	82
Figure 7.5	Absolute magnitudes derived with H , G_1 , G_2 and H , G_{12}^* models.	83
Figure 7.6	Phase curve coefficients derived with H , G_1 , G_2 and H , G_{12}^* models.	84
Figure 7.7	G_1 , G_2 for 15 taxonomic classes in cyan and orange.	88
Figure 7.8	G_1 , G_2 for taxonomic classes Ad, Bk, Ds, and K1.	90
Figure 7.9	Composition of asteroid families from phase curves.	93
Figure 9.1	Overview of preprocessing steps of observations.	103
Figure 9.2	Observations visualised as a missingness-matrix.	105

Figure 9.3	Distribution of 2125 individual asteroids in observations over absolute magnitude and orbital class. 107
Figure 10.1	Overview of clustering- and classification steps. 112
Figure 10.2	Latent components derived from observations using MCFA model. 113
Figure 10.3	Latent scores of observations in first and second latent dimension. 114
Figure 10.4	Latent scores of observations in third and fourth latent dimension. 115
Figure 10.5	Division of 2983 reflectance spectra into 50 Gaussian clusters. 117
Figure 10.6	Division of 2324 visual albedos into 50 Gaussian clusters. 118
Figure 10.7	Distribution of visual albedos in clusters associated to X-complex. 121
Figure 10.8	Spectra presenting e-, h-, and k-features. 123
Figure 11.1	Reflectance spectra and visual albedos of classes in C-complex. 128
Figure 11.2	Distribution of C-complex and endmember classes in latent space. 130
Figure 11.3	Reflectance spectra and visual albedos of classes in M-complex. 134
Figure 11.4	Distribution of M-complex and endmember classes in latent space. 136
Figure 11.5	Distribution of e- and k-features in latent space. 138
Figure 11.6	Reflectance spectra and visual albedos of classes in S-complex. 139
Figure 11.7	Distribution of S-complex and endmember classes in latent space. 140
Figure 11.8	Distribution of S-complex endmembers in latent space. 142
Figure 12.1	Classification of (325) <i>Heidelberg</i> with partial and complete observations. 148
Figure 12.2	Confusion matrix of classifications. 149
Figure 12.3	Confusion matrix of classifications for visible-only observations. 150
Figure 12.4	Confusion matrix of classifications from complete versus visible-only observations. 151
Figure 12.5	Confusion matrix of classifications for near-infrared-only observations. 152
Figure 12.6	Confusion matrix of classifications with literature values. 154
Figure 13.1	Compositional distribution of minor bodies in inner Solar System. 158
Figure 13.2	Distribution of taxonomic classes over inner Solar System. 159
Figure 13.3	Orbital distribution of D and Z-types. 160
Figure 13.4	Compositional distribution of asteroids in families. 162

Figure 13.5	Fraction of S and C near-Earth objects against absolute magnitude. 164
Figure 14.1	Spectral matching in data- versus latent space. 166
Figure 14.2	Spectra of CK, CO, and CV chondrites from Eschrig et al. (2021) and RELAB. 169
Figure 14.3	Spectra of K- and L-types and Barbarians. 170
Figure 14.4	Spectral variability in CV and CO chondrites. 171
Figure 14.5	Evaluation of space-weathering model through irradiation experiment. 172
Figure 14.6	Matching of CV, CO chondrites and K-, L-types and Barbarians. 175
Figure 14.7	Example matches identified between CV, CO chondrites and K-, L-, and M-types. 176
Figure 14.8	CV, CO chondrites and K, L-type asteroids in latent space. 177
Figure 14.9	Reflectance spectra of ordinary chondrites from Eschrig et al. (2022) and S-types. 179
Figure 14.10	Principal component analysis of ordinary chondrites and S-types. 180
Figure 14.11	Evaluation of Gaussian mixture model of S-type principal scores. 181

LIST OF TABLES

Table 3.1	References of proposed asteroid-meteorite matches. 36
Table 4.1	References of asteroid taxonomy flowchart. 45
Table 7.1	Effect of magnitude dispersion and opposition effect sampling on phase curve parameters. 81
Table 7.2	Mapping of literature classifications to simplified class scheme. 86
Table 7.3	Distribution of G_1 , G_2 for 15 taxonomic classes. 89
Table 7.4	Distribution of G_1 , G_2 for asteroid families. 91
Table 8.1	Comparison of taxonomy with previous taxonomies. 100
Table 9.1	References of reflectance spectra. 108
Table 9.2	References of visual albedos, absolute magnitudes, and diameters. 109
Table 10.1	Cluster-to-class decision-tree. 120
Table 10.2	Parameters for automated feature recognition fits. 124
Table 10.3	Number of observations and asteroids per taxonomic and orbital class. 126
Table 11.1	Description of taxonomic classes. 144
Table 12.1	Evolution of taxonomic scheme from Tholen (1984) to Bus-DeMeo to this work. 153

ACRONYMS

zMASS	Two Micron All Sky Survey
AA	Aqueous Alteration
ATLAS	Asteroid Terrestrial-impact Last Alert System
BIC	Bayesian Information Criterion
CC	Carbonaceous Chondrite
CAI	calcium-aluminium-rich inclusion
CDS	Centre de Données astronomiques de Strasbourg
DAMIT	Database of Asteroid Models from Inversion Techniques
EOC	Equilibrated Ordinary Chondrite
ECAS	Eight-Color Asteroid Survey
EKB	Edgeworth-Kuiper Belt
EKBO	Edgeworth-Kuiper Belt Object
ESA	European Space Agency
ESO	European Southern Observatory
IRAS	Infrared Astronomical Satellite
FA	Factor Analysis
GMM	Gaussian Mixture Model
IAU	International Astronomical Union
JAXA	Japan Aerospace Exploration Agency
JT	Jovian Trojan
KDE	Kernel Density Estimation
MBA	Main Belt Asteroid
LSST	Legacy Survey of Space and Time
ICL	Integrated Completed Likelihood
IDP	Interplanetary Dust Particle
IMB	Inner Main Belt
IMCCE	Institut de mécanique céleste et de calcul des éphémérides
JWST	James Webb Space Telescope
MC	Mars-Crosser
MCFA	Mixture of Common Factor Analysers
MCMC	Markov chain Monte Carlo
MIR	mid-infrared
MMB	Middle Main Belt
MMR	mean-motion resonance

MPC	Minor Planet Center
NASA	National Aeronautics and Space Administration
NIR	near-infrared
NEO	Near-Earth Object
OC	Ordinary Chondrite
OMB	Outer Main Belt
PC	principal component
PCA	Principal Component Analysis
PDS	Planetary Data System
PPCA	Probabilistic Principal Component Analysis
PT	Petrologic Type
RMS	root mean square
SDO	Scattered Disk Object
SDSS	Sloan Digital Sky Survey
SMASS	Small Main-Belt Asteroid Spectroscopic Survey
SsODNet	Solar system Open Database Network
TM	Thermal Metamorphism
TNO	Trans-Neptunian Object
UOC	Unequilibrated Ordinary Chondrite
UV	ultra-violet
VisNIR	visible-near-infrared
VISTA	Visible and Infrared Survey Telescope for Astronomy
WISE	Wide-field Infrared Survey Explorer

Part I

Minor Bodies of the Solar System

POPULATIONS

1.1	Outer Solar System	4	
1.1.1	Trans-Neptunian Objects		4
1.1.2	Centaurs	6	
1.2	Inner Solar System	6	
1.2.1	Jovian Trojans	6	
1.2.2	Main Belt Asteroids		7
1.2.3	Mars-Crossers	9	
1.2.4	Near-Earth Objects		10
1.2.5	Interior to Venus	10	

The Solar System is made up of a class G star, eight planets, several hundreds of natural satellites, and millions of small bodies. Small bodies include comets, irregular satellites of giant planets, minor bodies beyond Jupiter, and asteroids. This work focuses on the latter population.

The first asteroid added to the census was (1) *Ceres*, discovered in 1801 by Giuseppe Piazzi in Palermo, Italy. More than 200 years later, in July 2022, there are 1 217 458 small bodies listed in the *Asteroid Orbital Elements Database*¹ of the Lowell Observatory (Bowell 1996). 616 690 (51 %) have been observed sufficiently to compute their orbit around the Sun within a given uncertainty, after which they first receive a number and finally a name.

There are three main reservoirs of minor bodies in the Solar System: the Main Belt in the inner-, the Edgeworth-Kuiper Belt (EKB) in the outer-, and the Oort cloud at the edge of the Solar System. More than 99 % of minor bodies discovered to date are in the Main Belt, nevertheless, the EKB is expected to host about 100-1000 times the number of Main Belt Asteroids (MBAs) in any given size range (Bottke et al. 2020; Morbidelli et al. 2021). The Oort Cloud is thought to be the source region of the long-period comets (Levison and Dones 2007). The large distance to the objects past Neptune pose an observational challenge which is taken on by the new generation of telescopes, such as NASA's *JWST* and the *Simonyi Survey Telescope* at the Vera C. Rubin Observatory. Meanwhile, in this work, I focus on the objects that have already been discovered and, in part, characterised.

This chapter provides an overview of the populations of minor bodies in the Solar System. While the following chapters focus exclusively on the asteroids, it is important to have a global view of our planetary system in mind as no population of minor bodies exists in isolation. As we will see, a steady migration of objects is taking place over astronomically short lifetimes. By reviewing the interchange between the reservoirs, we can in the following better understand compositional trends within and between populations. The focus is thus placed on the dynamics of the populations, in particular for the asteroids, while the compositional aspects are treated in later chapters.

It may be nine planets (Batygin and Brown 2016).

A nice example of dynamical evolution affecting composition is the transition from amorphous to crystalline ices in objects from the outer Solar System as they enter the inner part (Klinger 1981).

¹ <https://asteroid.lowell.edu/main/astorb/>

1.1 OUTER SOLAR SYSTEM

The outer Solar System is the realm of the giant planets, namely Jupiter (\J), Saturn (\S), Uranus (\U), and Neptune (\Ne). It extends outwards from the orbit of Jupiter and includes the EKB, as illustrated in Fig. 1.1. The outer edge is a matter of definition, placed at 100 au or beyond depending on which branch of astronomer or astrophysicist is tasked to identify it.

The giant planets shape the outer Solar System through their gravitational force, both in the past during their formation and migration as well as in the present. Among other dynamical effects, they create volumes in the three-dimensional orbital space of bodies in the Solar system (spanned by the semi-major axis a , the eccentricity e , and the inclination i) where the orbits of smaller bodies are either enhanced in stability or instability. Small bodies which acquire orbital elements within these volumes thus are either fixated on this orbit by the giant planet or they are quickly removed from it via gravitational perturbation. These volumes are referred to as mean-motion resonances (MMRs) and occur when two objects in orbit around a third, central body have orbital periods which are integer ratios of one another. The resulting perturbations of the orbits of the resonant objects are highly complex and specific to each resonance (Morbidelli 2002). For minor bodies, the relevant MMRs occur with the planets, and it is the Goliath which dictates the motion of the David. MMRs are denoted as $p : q$, where p is the number of orbital periods for the inner and q the one for the outer object.

Minor body populations are often defined based on the confines that the MMRs around them dictate, as well as on the closest distance to the Sun that they reach on their orbit, their perihelion q , given by $a(1 - e)$, and the furthest distance to the Sun, their aphelion Q , given by $a(1 + e)$. I now briefly discuss these populations, going from the outside inwards.

1.1.1 *Trans-Neptunian Objects*

Trans-Neptunian Objects (TNOs) include minor bodies with semi-major axes larger than that of Neptune, $a > a_{\Ne}$. Their distribution in $a - e$ is shown in Fig. 1.1. While it was initially assumed that the bodies beyond Neptune would form a low-inclination, low-eccentricity extension of the minor bodies interior to its orbit (Edgeworth 1943; Kuiper 1951), a surprisingly large amount of structure is readily apparent in the orbital distribution. Specifically, we identify three populations here: a wide band of increasing semi-major axis and eccentricity hosting among others (136199) *Eris*, which is referred to as *Scattered Disk*, several populations of constant semi-major axis caught in MMRs with Neptune such as the Plutinos around (134340) *Pluto*, referred to as resonant objects, and an accumulation of objects around (15760) *Albion* at $(a, e) = (44 \text{ au}, 0.17)$. The latter population looks most like the one predicted before the discovery of TNOs and is therefore named the EKB.

The structure of the TNOs can simplistically yet intuitively be explained by gravitational interaction between Neptune and the disk of planetesimals around it during the early stage of the Solar System. As Neptune forms and migrates outwards, it scatters the planetesimals along its path (Fernández and

(15760) *Albion* was the first TNO discovered since (134340) *Pluto* in 1992, marking the beginning of *Pluto*'s demotion which was finished by the discovery of (136199) *Eris* in 2006.

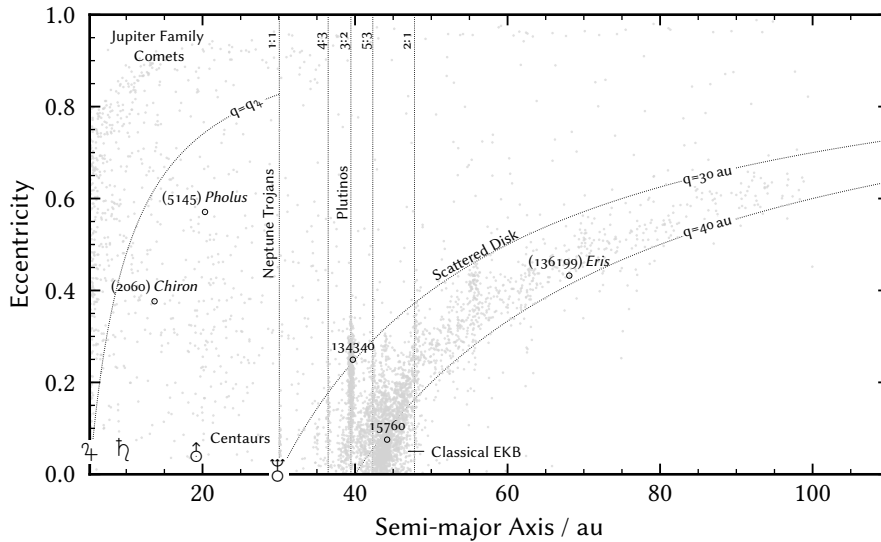


Figure 1.1: The distribution of minor bodies in the outer Solar System. The filled, grey circles give the osculating semi-major axes and eccentricities of minor bodies in the *Asteroid Orbital Elements Database*. The planets of the Solar System are marked by their respective symbols. Dotted lines indicated mean-motion resonances and perihelia distances which are commonly used to define different populations. Minor bodies mentioned in the text are highlighted by open, black circles. Those indicated only by their number are (15760) *Albion* and (134340) *Pluto*. The outer Solar System continues further outwards, the area here is focused for legibility.

Ip 1984; Nesvorný and Vokrouhlický 2016), giving rise to the Scattered Disk Objects (SDOs). The planetary encounter changes the orbital parameters of these planetesimals while keeping their perihelia q conserved, as indicated in Fig. 1.1. SDOs interact gravitationally with Neptune at perihelion which excites their orbits further, pushing them ever closer to the outer edge of the Solar System.

The early migration of Neptune further leads to a migration of its MMRs. The stabilising resonances may capture planetesimals and bind them to the migrating planet (Malhotra 1993; Gladman et al. 2012). Four external resonances of Neptune are given in Fig. 1.1, in addition to in 1:1 resonance with it, referred to as *Neptune Trojans*. These objects are leading or trailing the planet in its L_4 and L_5 Lagrange points of the Sun-Neptune system. The offset in their orbital phase with respect to the one of Neptune stabilises their orbits.

Finally, there are non-resonant objects located around (15760) *Albion* between the 3:2 and 2:1 MMRs. They constitute the “classical belt” of minor bodies predicted by Edgeworth (1943) and Kuiper (1951) as remnant of the planetary formation process. These objects appear to never have been significantly disturbed by the migration of Neptune, as inferred from the large amount of wide binary objects in this population (Parker and Kavelaars 2010), apart from a population which might have been implanted by the giant planet’s push outwards (Fraser et al. 2017). Hence, they may have formed

in-situ, and the sharp edge of the EKB around 47 au might mark the outer extent of the protoplanetary disk of the early Solar System.

As they encode the history of giant planet migration, the dynamical distribution of objects in the 20 au beyond Neptune are of high interest for planetary formation models. In terms of their composition, TNOs appear to be at least bi-modal, suggesting that planetesimals that formed in different locations are mixed together (Gulbis et al. 2006; Fraser et al. 2017). The exact distribution in terms of composition is not relevant for the remainder of this work and I will broadly describe all TNOs as being rich in volatile elements and organic molecules due to their formation in the outskirts of the protoplanetary disk, beyond the lines where water and carbon-monoxide may be found in liquid or solid form.

1.1.2 Centaurs

Centaurs are minor bodies with orbits between those of Jupiter and Neptune, $q > a_J$, $a < a_{\text{♃}}$ (Gladman et al. 2008). The first Centaur to be discovered was (2060) *Chiron* in 1977, 15 years before the discovery of the first TNO and the second Centaur, (5145) *Pholus*.

Centaurs may be regarded as the low- q equivalent of the SDOs, i. e. they are objects of the classical EKB scattered inwards by interaction with Neptune (Duncan and Levison 1997). As a consequence, we expect the same pristine material on their surfaces. Further source regions are resonant TNOs in the 3:2 and 2:1 MMRs (Morbidelli 1997; Tiscareno and Malhotra 2009). The dynamical lifetime of this population is in the order of a few to hundreds of Myr, after which they drift into the population of Jupiter Family Comets or into the outer Solar System (Tiscareno and Malhotra 2003), and in either case, outside the scope of this work.

An exception is (5145)Pholus who makes an appearance in Chapter 11.

1.2 INNER SOLAR SYSTEM

The inner Solar System hosts the rocky planets Mercury (♿), Venus (♀), Earth (♁), and Mars (♂). It extends from the Sun up to Jupiter. Between the giant planet and Mars lies the Main Belt, the largest reservoir of asteroids. Additional minor planets can be found around the rocky planets in orbital space.

The minor bodies in the inner Solar System are the main subject of this thesis. While Jupiter is counted to the outer Solar System, I include its Trojan population in this section as they are also treated in the later chapters. The distribution of minor bodies in the inner Solar System in $a - e$ is shown in Fig. 1.2.

1.2.1 Jovian Trojans

As the Neptunian Trojans, the Jovian ones occupy the L_4 - and L_5 Lagrange points of their neighbouring planet, i.e. they lead or trail the orbit of the planet by 60 deg in orbital phase in stable points of the Sun-Jupiter gravitational system (Emery et al. 2015). They are named exclusively after heroes of the

Both camps make use of espionage: (624)Hektor hides itself in L_4 for Troy while (617)Patroclus resides in L_5 for the Greek.

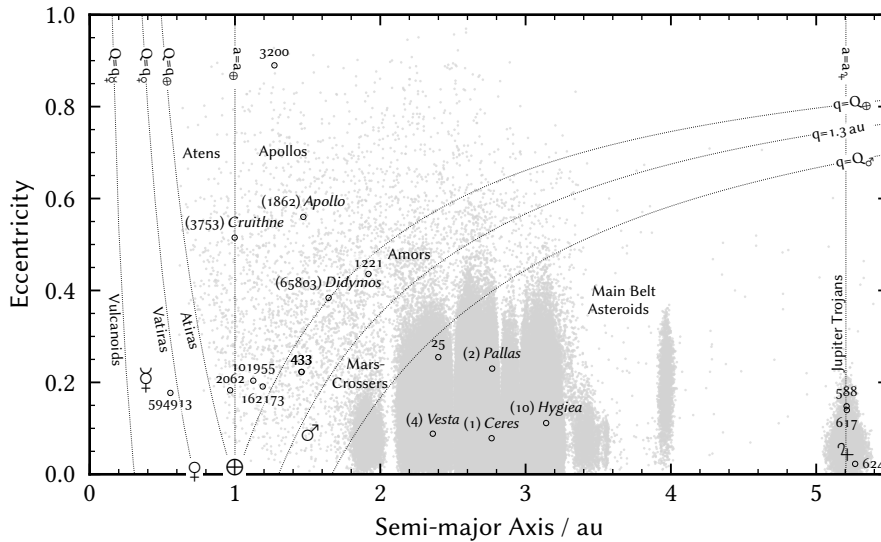


Figure 1.2: As Fig. 1.1, giving the distribution of minor bodies in the inner Solar System. Asteroids indicated only by their number are: (25) *Phocaea*, (433) *Eros*, (588) *Achilles*, (617) *Patroclus*, (624) *Hektor*, (1221) *Amor*, (2062) *Aten*, (3200) *Phaethon*, (101955) *Bennu*, (162173) *Ryugu*, and (594913) *'Aylo'chaxnim*.

Trojan War, which explains why populations co-orbital to the planets are referred to as Trojans. In the Jovian system, Trojans in the L_4 point are named after heroes of the Greek side while those in L_5 are named after champions of the Trojan side.

The first three Jovian Trojans (JTs) were discovered in 1906 and 1907, (588) *Achilles*, (617) *Patroclus*, and (624) *Hektor* (Nicholson 1961). There are more than 12 000 JTs known as of June 2022, and the number of objects with a diameter $D > 100$ m is estimated to be within a few tens of percent of the Main Belt population based on their size-frequency distribution (Wong and Brown 2015).

The Jovian Trojans are remarkably uniform in their composition (Barucci et al. 2002; Marsset et al. 2014). Although they are in close proximity to the inner Solar System, in the following chapters, we will see that JTs are compositionally closer to objects of the outer Solar System (Dotto et al. 2008). Both points are further discussed in Chapters 4 and 11. This supports results by Morbidelli et al. (2005), who proposes that Trojans originally formed in the region of the solar nebula which now hosts the EKB and were later scattered during the giant planet migration and captured by Jupiter.

1.2.2 Main Belt Asteroids

The Main Belt is located between Mars and Jupiter, spanning the range of 1.8 au–3.5 au. It hosts the largest asteroid in the Solar System, (1) *Ceres*, visited by the National Aeronautics and Space Administration (NASA) mission *Dawn*, which measured a mean diameter of 939.4 km for the dwarf planet (Russell

Trojans are not exclusive to planets: two moons of Saturn, Tethys and Dione, are escorted by other, gravitationally-locked moons on their orbits around Saturn.

Dwarf planets are bodies which orbit the Sun and are in hydrostatic equilibrium yet do not dominate their orbits.

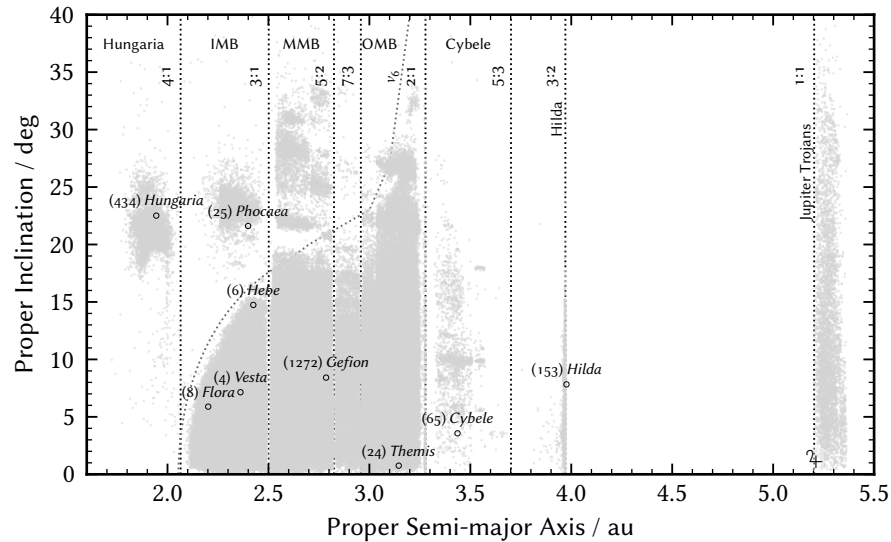


Figure 1.3: The dynamical structure of the Main Asteroid Belt and the Jovian Trojans. The distribution of asteroids is indicated using their proper orbital elements from the *Asteroids Dynamic Site*. Mean motion resonances and the secular resonance ν_6 are indicated via the dotted lines.

et al. 2016). About 200 asteroids are above the 100 km-threshold, while observational survey efforts are narrowing in on a completion-boundary at 1 km size (Hendler and Malhotra 2020). Models of the size frequency distribution of asteroids based on the sizes of craters observed on the space mission targets such as (4) *Vesta*, (101955) *Bennu*, and (162173) *Ryugu* estimate the number of asteroids with diameters above 100 m at 10^8 (Bottke et al. 2020). Figure 1.3 shows the distribution of almost 600 000 MBAs in proper semi-major axis (a_p) and proper inclination (i_p). The proper orbital elements of an object are free of the quasi-periodic perturbations introduced by the gravitational influence of its environment (Milani and Knežević 1994), making them constant over astronomical time-scales. The elements shown in Fig. 1.3 are extracted from the *Asteroids Dynamic Site*² (Knežević and Milani 2003). Hence, only the long-term gravitational interaction between the planets and the minor bodies in the Main Belt is observable in these elements. In particular, several MMRs can be seen to carve trenches in the distribution of asteroids, referred to as the Kirkwood Gaps. They are used to divide the Main Belt into smaller populations, namely the Inner Main Belt (IMB) (between the 4:1 and 3:1 resonance), Middle Main Belt (MMB) (3:1 and 5:2), and Outer Main Belt (OMB) (5:2 and 2:1). Three more populations just beside the Main Belt are highlighted in Fig. 1.3: the Hungarias, a high-inclination population inside the IMB, the Cybeles, just outside the OMB and delimited by the 5:3 resonance, and the Hildas, which are located in the stabilising 3:2 resonance with Jupiter. We look at these populations more in depth in Chapter 4.

Apart from the use for nomenclature, the Kirkwood Gaps are of high relevance to the study of the dynamical Main Belt. The Main Belt is the

The region between the 5:2 and 7:3 resonances is sometimes referred to as the “primitive Main Belt” as its resident asteroids are on generally less excited orbits than the other populations in the Main Belt.

² <https://newton.spacedys.com/astdys/>

source reservoir for the Near-Earth Objects (NEOs) and meteorites, therefore, understanding which objects in the Main Belt are delivered into the near-Earth space is important. The transport mechanism revolves around resonances such as the MMRs (Granvik et al. 2017; Granvik and Brown 2018). After asteroids approach the resonances in orbital space and eventually get captured, their eccentricities and inclinations are increased, i. e. the orbits are excited. This eventually leads them onto planet-crossing orbits (Greenberg and Nolan 1989). The most efficient resonances for this delivery of material into the near-Earth space are the ν_6 secular resonance and the 3:1, 5:2, and 2:1 MMRs (Morbidelli et al. 2002). Secular resonances ν_i affect asteroids whose mean rate of perihelion precession is equal to that of the i th planet. The ν_6 has a particularly strong influence on the Main Belt as is apparent in that the distribution of MBAs, which follow resonance outline, refer to Fig. 1.3. The ν_6 is approximately independent of the eccentricity (Morbidelli et al. 2002), which makes it a powerful driver of excitation for objects captured in it. It is a gateway for MBAs into the near-Earth space, though most asteroids captured within eventually fall into the Sun (Farinella et al. 1994). The median lifetime of objects in the ν_6 is only 2 Myr (Gladman et al. 1997).

Finally, we turn our eye to more structure apparent in the distribution of MBAs in Fig. 1.3. There are regions in the orbital parameter space where asteroids cluster, such as the region around (434) *Hungaria* in $(a_p, i_p) = (2.0, 23)$. These “condensations” (Hirayama 1918) of asteroids are asteroid families, groups of objects which share a region in orbital parameter space and likely a common history, e. g. as remnants of a collision of a previous generation of asteroids. Figure 1.3 gives the positions of the supposed parent bodies of some of the largest asteroid families, the largest being the family of (4) *Vesta* called the “Vestoids” (Binzel and Xu 1993). Families are relevant for the compositional analysis as we may observe different layers of the disrupted parent body and investigate the effects of on-going exposure of the surfaces of asteroids of different sizes to the space environment (Novakovic et al. 2022).

1.2.3 Mars-Crossers

Mars-Crossers (MCs) are asteroids with perihelia q smaller than the aphelion Q_{\oplus} of Mars (1.67 au) but larger than 1.3 au. There are more than 24 000 known MCs to date. The expected number of MC with $D > 100$ m is about 10^5 following the estimate in Carry (2018). They are a steady-state population, meaning that the rate of asteroids entering the region from source regions is equal to the rate of asteroids leaving the region (Morbidelli et al. 2002).

The dynamical evolution of MCs highly depends on their exact orbital parameters. Via numerical integration, Michel et al. (2000) show that about 20 % of the population is dynamically stable, i. e. they do not suffer an encounter with Mars within 100 Myr. Of the remaining 80 %, about two thirds reach Earth-crossing orbits within 10^7 yr due to random-walks in their semi-major axis triggered by planetary encounters. These orbits largely lead to the final fate of encountering the Sun. MCs close to the ν_6 resonance are in particular

The 1.3 au boundary was arbitrarily chosen based on the orbital frequency distribution of known NEOs at the time (Shoemaker and Helin 1978). The martian perihelion is at 1.38 au.

vulnerable to orbital disturbances. A small percentage of MCs are further ejected into the outer Solar System (Włodarczyk 2020).

1.2.4 Near-Earth Objects

NEOs are a dynamic steady-state population of minor bodies within the Earth's orbital vicinity. Four groups are recognised based on their relationship to Earth's orbit: Amors ($Q_{\oplus} \leq q \leq 1.3$ au), which do not cross the Earth's path but may graze it from the outside, Apollos ($a > 1$ au, $q < Q_{\oplus}$) and Atens ($a < 1$ au, $Q > q_{\oplus}$), which cross the Earth's orbit, and Atiras ($q_{\oplus} < Q < q_{\oplus}$) whose orbits are entirely contained within the orbit of Earth, refer to Fig. 1.2. The number of known NEOs across all groups in June 2022 is close to 30 000. Apollos and Amors take the lion's share of the population, while only a few tens of Atiras have been discovered. Population models estimate a number of NEOs with $D > 100$ m in the order of 10^4 (Granvik et al. 2016). Given the proximity to Earth, this size-range is largely completed in the observational surveys.

NEOs form a steady-state population with dynamical lifetimes of 10^6 yr (Morbidelli et al. 2002). The population is sourced primarily by Main Belt asteroids and comets and depleted by planetary encounters in the relatively crowded region of the terrestrial planets (Binzel et al. 2015).

Any impactor on Earth originating from the Main Belt has necessarily transitioned through the near-Earth space. As such, the compositional study of NEOs is of high interest, in particular in terms of surface alterations induced by the solar irradiation, as discussed in Sect. 3.5.1. Their proximity further allows to study them in much greater detail than the more distant minor bodies, including radar observations, which are not possible for objects farther away due to the rapidly decreasing intensity of the radar echo with distance (Ostro et al. 2002).

Furthermore, NEOs are the most accessible targets for space and sample-return missions due to the low required change of velocity required to rendezvous with them, in particular co-orbital objects like (3753) *Cruithne* (Abell et al. 2015). In Fig. 1.2, the positions of all asteroids which have been visited by spacecraft including a surface touchdown is given, namely those of (433) *Eros*, (25143) *Itokawa*, (101955) *Bennu*, and (162173) *Ryugu*.

1.2.5 Interior to Venus

Minor planets orbiting on interior orbits to that of Earth are challenging to discover due to the close angular distance to the Sun, making observations only feasible during the relatively bright twilight hours or during total solar eclipses (Perrine 1902). Numerical modeling of the NEO population in preparation for the launch of the Canadian *NEOSSat* mission revealed the possibility of injecting MBA into orbits contained entirely within the orbit of Venus, $q_{\oplus} < Q < q_{\oplus}$ (Greenstreet et al. 2012). The authors coined the term "Vatiras" for the at the time theoretical population of tens of objects. In 2020, the *Zwicky Transit Facility* reported the first observations of a Vatira asteroid, 2020 AV2, now officially designated as (594913) 'Aylo'chaxnim (Bolin

The groups among the NEOs are named after the first asteroid identified within the respective group - except for the Amors, where (433)Eros was skipped. This was later reconciled by making (433)Eros the first asteroid to be visited by a spacecraft, NASA's NEAR Shoemaker.

After September 26, 2022, (65803)Didymos may be added to this list if we count NASA's DART projectile as touchdown.

'Aylo'chaxnim means "daughter of Venus" in the language of the Palomar Tribe. Maybe the Vatiras should be referred to as 'Aylo'chaxnims (or the simpler Aylos) following the NEO scheme?

et al. 2020). Spectroscopic observations (discussed in [Chapter 2](#)) show that (594913) *Aylo'chaxnim* has a basaltic composition (Popescu et al. 2020). Its predicted fate within the next million years is to transition into the Atira population (Greenstreet 2020).

The Vulcanoid asteroids, a population within the orbit of Mercury, $Q < q_{\oplus}$, remain hypothetical. A search by Schumacher and Gay (2001) using data from the ESA *SOHO* mission placed an upper limit of 60 km on the diameter of possibly undetected Vulcanoids. Granvik et al. (2016) show that objects on such low-perihelion orbits suffer from catastrophic destruction, likely due to thermal disintegration (Delbó et al. 2014) or rapid spin-up to the point of destruction caused by the solar irradiation.

The Vulcanoids take their name from the planet "Vulcan", hypothesised to reside between Mercury and the Sun in the late 19th century.

OBSERVABLES

2.1	Reflectance and Emission of Asteroid Surfaces	14
2.2	Spectroscopy	16
2.3	Photometry	17
2.4	Albedo	18
2.5	Phase Curve Coefficients	19
2.6	Honorary Mentions	23
2.6.1	UV and MIR Spectroscopy	23
2.6.2	Polarimetry	23
2.6.3	Radar	24

The study of the composition of the minor bodies in the Solar System aims to derive fundamental properties like their mineralogy and the structure of their surface (Gaffey and McCord 1979). The most detailed description of these parameters is acquired in laboratories on Earth, where mineral abundances, petrographic parameters, and isotopic anomalies can be precisely quantified (Brearley et al. 1998; Grady and Wright 2006). However, the asteroids in the Main Belt are distant from the terrestrial laboratories and thus evade a detailed description. One solution to this issue are sample return missions which visit asteroids and return samples from their surface to Earth for detailed analysis (Yoshikawa et al. 2015). These missions give spectacular insights, however, the required investment both in time and money makes the description of the entire Main Belt via sample return infeasible.

Alternatively, we can study the pieces of the asteroids which reach Earth on their own volition: the meteorites. Meteorites give us precise insights into the early Solar System (Weisberg et al. 2006; Krot et al. 2014), yet the vision they provide of the Main Belt is a blurred scenery. Meteorites are letters without return address (Greenwood et al. 2020). We do not know which asteroid sent which meteorite, so we can only tentatively infer on the properties of individual asteroids based on meteorite analogues. Meteorites are the subjects of [Chapter 3](#).

Hence, we remain with the method which has been used to study celestial sources for millennia: the passive observation of the light that asteroids reflect and emit (Chapman and Gaffey 1979). The underlying principle is that, as light is reflected by the asteroid, a compositional signature is imprinted in its properties due to the interaction with the surface layer of the body. The way a surface returns the incident light is determined by its surface properties such as the composition, grain size, the porosity, and the roughness (Belskaya et al. 2015; Li et al. 2015; Reddy et al. 2015).

This chapter describes remote-sensing observations with a focus on observables with compositional information, specifically those which are important for the following chapters. I go in-depth in the description of asteroid phase curves in [Sect. 2.5](#) as the compositional analysis of phase curve coefficients resulted in the first publication during this PhD (Mahlke et al. 2021) and is

the topic of [Chapter 7](#). I conclude the chapter with a mention of observables which are compositionally relevant but have not made it into the asteroid taxonomy (yet).

2.1 REFLECTANCE AND EMISSION OF ASTEROID SURFACES

Remote-sensing observations of planetary bodies record the radiation which is reflected or emitted by the target. The primary source of the incident radiation on the body is the Sun, while emission occurs in form of thermal radiation depending on the bodies temperature. Radiation in the visible- and near-infrared regime is commonly referred to as light.

We can approximate both the solar irradiation and the thermal emission of asteroids using the black body spectrum. A black body is a theoretical perfect absorber of incident radiation, emitting the absorbed energy in form of thermal emission. The emission spectrum $B(\lambda, T)$ is calculated using the Planck function (Lebofsky et al. 1986)

$$B(\lambda, T) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{\exp(hc/\lambda kT) - 1}, \quad (2.1)$$

where λ is the wavelength of the emission, T is the temperature of the body, h is the Planck constant, c is the vacuum speed of light, and k is the Boltzmann constant. In the upper part of [Fig. 2.1](#), the black body spectrum of a body at $T = 5800$ K is shown, approximating the emission spectrum of the Sun (Goody and Yung 1989). For the majority of asteroids, the thermal emission dominates the solar reflection around the mid-infrared region, well outside the focus we place here on the visible and near-infrared. However, for NEOs with dark surfaces on orbits with low perihelia, the thermal emission may be non-negligible in the $2 \mu\text{m}$ -region (Marchi et al. 2009), as shown in the upper part of [Fig. 2.1](#) by means of the black body spectra at $T = \{1000, 750, 500\}$ K. These “thermal tails” are then modelled and accounted for (e. g. Rivkin et al. 2005).

The reflectance spectrum of a solid body is shaped by its surface properties. The deposited energy excites electron transitions in the surface layer, absorbing light at characteristic wavelengths, which are recorded using spectroscopy (Adams 1975). The quantum mechanics at play help us understand the composition of the macroscopic subjects of our studies. We look more at these fingerprints in [Chapter 3](#).

Before we record the reflected light on Earth, there are two more effects altering its appearance. First, the Sun is not emitting equally at all wavelengths, leading to a biased spectral appearance of minor planets which reflect its irradiation that has to be corrected for. This is achieved by observing a solar-like star and dividing the observed reflectance of the asteroid by the reflectance of the solar analogue. Second, the light we observe passes through the Earth’s atmosphere, where absorption and scattering take place as a function of wavelength. Molecules in the atmosphere such as bound carbon dioxide and water are absorbing at given energies, diminishing the incoming flux in these wavelength regions. These wavelength regions are best observed using space-borne instruments. The atmospheric transmission

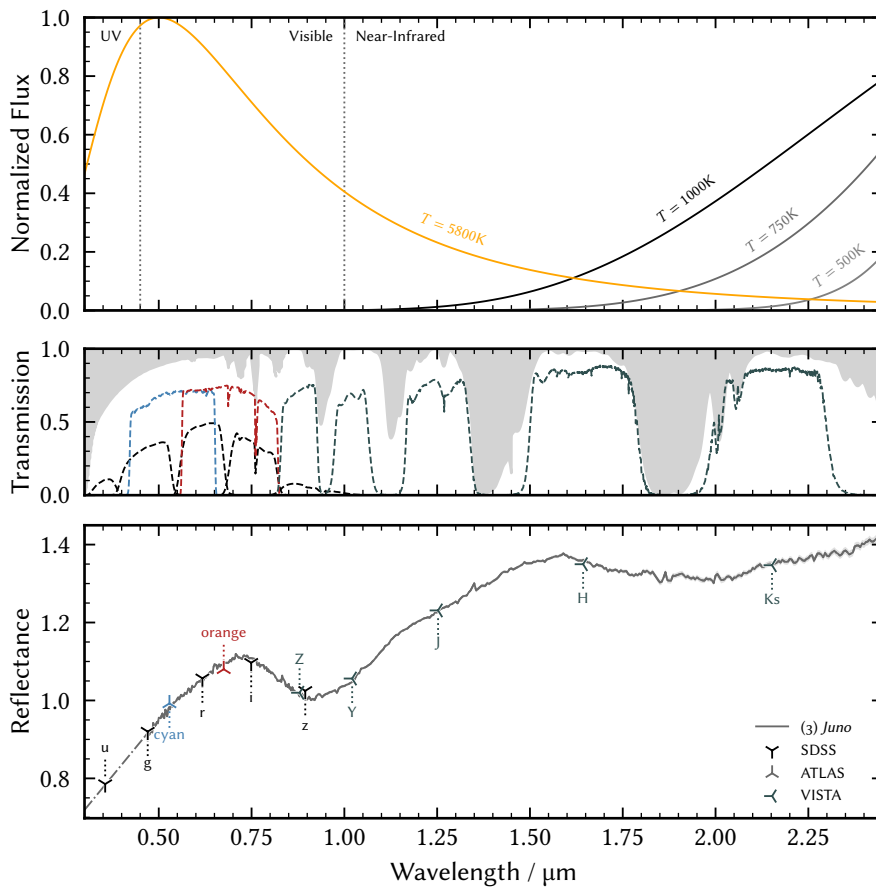


Figure 2.1: The black body spectra at different temperatures, atmospheric transmission and photometric filters, and the spectrum of (3) *Juno* visible-near-infrared wavelength region. *Top*: The normalised flux of a black body at 5800 K (orange) approximates the solar irradiation spectrum. The thermal emission of near-Earth objects in extreme cases is represented by the black body spectra at 1000 K (black), 750 K (dimgrey), and 500 K (grey). The dotted vertical lines give commonly used transition regions between the ultra-violet (UV), visible, and near-infrared wavelength ranges. *Middle*: The transmission of Earth's atmosphere is shown in white. The transmission of photometric filters used by the *Sloan Digital Sky Survey* (black), *ATLAS* observatory (red, blue), and the *VISTA* survey (grey) are given by the dashed lines. *Bottom*: The visible-near-infrared spectrum of (3) *Juno* is shown by the grey line. The spectrum has been extrapolated into the UV region (dash-dotted line). The convolution of the spectrum with the photometric filters in the middle plot give the apparent magnitude values marked by respective filters name. The atmospheric transmission data is from *ATRAN* (Atmospheric TRANsmission, Lord (1992)) (<https://atran.arc.nasa.gov/cgi-bin/atran/atran.cgi>), the photometric filters are from the SVO Filter Profile Service (<http://svo2.cab.inta-csic.es/theory/fps/>), and the spectrum of (3) *Juno* from *SMASS* (<https://smass.mit.edu>).

is shown in the middle part of Fig. 2.1. Recorded reflectance spectra of asteroids are corrected for the atmospheric absorption using theoretical models of the absorption and light-scattering behaviours of gases in the atmosphere, adapted to the observational circumstances such as the airmass Lord (1992) and Bus et al. (2002).

2.2 SPECTROSCOPY

In spectroscopic observations, the incoming light of the asteroid is dispersed via a prism or grating to record the intensity per wavelength interval. This dispersment reveals the characteristic energy intervals at which the surface layers of the minor bodies absorb photons. An example of such a reflectance spectrum is shown in the lower part of Fig. 2.1 for asteroid (3) *Juno*. An overall increase in the reflectance towards the near-infrared (NIR) is visible, as well as two broad absorption bands (“features”) around 1 μm and 2 μm .

Spectra are the compositionally most informative observables of asteroids. A visible-near-infrared (VisNIR) spectrum enables us to infer on the target’s mineralogy. However, there are several pitfalls when recording and processing the VisNIR spectrum of a minor planet. For the observational side, refer the reader to the PhD thesis of Popescu (2012), as I exclusively dealt with asteroid spectra after they had been observed and preprocessed.

When it comes to processing the data products, a major source of uncertainty stems from the fact that the visible- and the NIR parts generally cannot be observed using the same instrument as different types of detectors are required to record the incoming light in these wavelengths. Typically recorded are the ranges between 0.4 μm –1 μm using a charge-coupled device Xu (1994) and the 0.8 μm –2.5 μm using arrays of indium antimonide (InSb) (Rayner et al. 2003). The overlapping wavelength regions are used to attach the spectral parts to one another. For featureless spectra, this is generally straight-forward, while the presence of features around 0.9 μm can complicate the matter and lead to several possible merging solutions. A subjective choice has to be made of how and where to merge the spectra, which affects important parameters such as the slope and even the presence of absence of features.

Furthermore, as we generally do not know the size of the illuminated surface area, there is no measure of the absolute flux of the reflectance spectrum, and we record the flux relative to some normalisation wavelength. Asteroid spectra are usually merged to unity at 0.55 μm for visible-only or VisNIR and 1.25 μm for NIR-only, aiming to normalise at a wavelength which is not affected by absorption bands. In Chapter 9, we see that normalising spectra becomes difficult as soon as we combine data which does not cover the same wavelength range.

An additional challenge arises from the phase-angle dependence of the reflectance spectra of asteroids. The solar phase angle α is the angle between the Sun and Earth as seen from the object, refer to Fig. 2.2. It is the complimentary angle to the solar elongation. Observations of asteroids at increasing phase angles lead to an increase in the observed spectral slopes and changes in the depths of absorption features (Gradie et al. 1980; Gradie and Veverka 1986;

A nice visualisation of this subjective process is given in Figure 3 of Clark et al. (2009).

Sanchez et al. 2012). This alteration of the spectral appearance is generally not corrected for in the post-processing of observations. I discuss this effect more in Chapter 7.

Acquiring asteroid spectra requires a bright planetary surface due to the dispersion of the incoming brightness. Hence, asteroids which are close to Earth or large in size are preferentially observed. By decreasing the resolving power of the observational setup, we open up the observable space to more asteroids, outlined in the next section.

2.3 PHOTOMETRY

In photometric observations, the incoming light of the asteroid is passed through a filter with a characteristic transmission behaviour. This allows to get an accurate value of the brightness of an asteroid in a range of wavelengths. As the light is not diffused via a prism or grating, the intensity remains higher and much fainter targets can be observed. The design of the filters is specific to the instruments and shaped by the Earth's atmosphere. The middle part of Fig. 2.1 shows the filters employed in three wide-field imaging surveys, the Sloan Digital Sky Survey (SDSS), the Asteroid Terrestrial-impact Last Alert System (ATLAS) observatory, and the Visible and Infrared Survey Telescope for Astronomy (VISTA) survey by the European Southern Observatory (ESO).

We may derive the absolute flux F originating from a minor planet and relate it to the brightness of other sources in the image, e. g. stellar sources. The brightness of minor planets is commonly expressed in terms of their apparent magnitudes m , given by

$$m - m_{\text{ref}} = -2.5 \log_{10} \frac{F}{F_{\text{ref}}}, \quad (2.2)$$

where F is the brightness of the minor body and F_{ref} and m_{ref} are the brightness and apparent magnitude of a reference source.

Depending on the brightness of the minor planet and the observational capabilities of the telescope, magnitudes with a high signal-to-noise ratio may be recorded within the order of seconds to minutes. This enables us to characterise the temporal evolution of the brightness by observing the same target in the same filter multiple times in quick succession, which is referred to as the lightcurve of the object. The lightcurve shows periodic variability on different timescales due to the minor planet's rotation and the viewing geometry within the Sun - Earth - asteroid system. These aspects affect the reflectance spectra as well, yet, while they are a nuisance to be accounted for in spectroscopy, in photometry, we make use of the temporal dimension to infer on global properties of the asteroid. An example are the detailed models of the shape derived from the inverse-problem presented by the lightcurves (e. g. Kaasalainen 2001).

If two magnitudes in different bands are acquired within the order of a few minutes, their difference encodes the colour of the object. Colours are used for the compositional analysis of asteroids (e. g. Sergeev and Carry 2021). However, minor planets are generally not the main scientific objective of wide-

The best place to get the filter transmission data for the major surveys and telescopes is the [Filter Profile Service](#) of the Spanish Virtual Observatory (Rodrigo et al. 2012; Rodrigo and Solano 2020).

A great place to get shape models of asteroids is the [DAMIT](#) website.

field imaging surveys, and the quality of the compositional characterisation using the apparent magnitudes depends largely on the cadence and filter selection of the survey. Taxonomic classification based on the SDSS and VISTA colours are presented in DeMeo and Carry (2013) and Popescu et al. (2018a). The results show that major taxonomic complexes (discussed in Chapter 4) are distinguishable but class assignments based on photometric colours are less reliable.

Photometric observations of wide-field imaging surveys are ideal for the discovery and astrometric characterisation of asteroids. I have developed a pipeline for the automatic recovery of asteroids in wide-field images (Mahlke et al. 2019) which I apply daily in a fully automatic fashion to the images of the J-VAR survey, a time-of-opportunity survey at the *Observatorio Astrofísico de Javalambre* for non-photometric observing conditions. Originally developed to discover and extract minor bodies in images of the Kilo-Degree Survey at ESO (Mahlke et al. 2018), I have further applied it to the recovery of asteroids in the first data release of the J-PLUS survey at the *Observatorio Astrofísico de Javalambre* (Mahlke et al. 2019), the WFCAM Transit Survey at the *United Kingdom Infrared Telescope* (Cortés-Contreras et al. 2019), and in images of the first data release of the *Gran Telescopio Canarias* OSIRIS instrument (Cortés-Contreras et al. 2020).

To relate the magnitudes of minor planets to their physical size, we compute the reduced magnitude V ,

$$V = m - 5 \log(r\Delta), \quad (2.3)$$

where we account for the distances between the objects and the Sun (r) as well as the objects and the observer (Δ). Observing the reduced magnitude of a minor planet at opposition, i. e. when the Sun, Earth, and minor planet are aligned in a straight line, averaged over one period of rotation of the body yields its absolute magnitude H . It enables to compute colours from non-simultaneous observations in different bands as well as, in combination with the observation of the diameter, the asteroid's surface albedo diameter, as shown in the next section.

2.4 ALBEDO

The albedo p of a surface quantifies how much light it absorbs or reflects. It ranges typically between 0 and 1, where dark surfaces have an albedo close to 0 as they preferentially absorb light while bright surfaces have albedos around unity as they reflect most incoming irradiation. We thus see intuitively that the surface albedo depends on the properties of the surface, in particular, the mineralogy.

There are different definitions of the albedo. In this work, I treat in particular the geometric albedo, given by the brightness of a body at zero degree phase angle relative to a theoretical, perfectly reflecting disk of equal cross-section (Lester et al. 1979). This theoretical disk is a diffuse scatterer, hence, the geometric albedo may be larger than 1 if the physical body has a preferred direction of reflectance. If only light in the visible wavelength region is regarded, the albedo is referred to as the visual albedo p_V .

The compositional information provided by the surface albedo may be considered orthogonal to the one provided by spectroscopy. Asteroids which have similar spectra in the VisNIR may exhibit different albedos, allowing for a more detailed classification. Wright et al. (2016) show that the distribution of visual albedos of all minor planets may be well described by a superposition of two Rayleigh distributions, one with a “dark” peak around 0.034 and one with a bright peak around 0.151. As such, the albedo allows to place minor planets into one of two or three taxonomic complexes, outlined further in Chapter 4. An exception is the taxonomic class E, which is unique in that its members have $p_V > 0.5$, much larger than the ensemble of asteroids.

The albedo of a minor planet can be derived using several observational methods, such as radiometry, polarimetry, and photometry. For explanations of the former two, I refer to Chapman et al. (1975), who employ albedos derived from both methods to derive what I consider to be the first asteroid taxonomy. The photometric method is of more interest here as most visual albedos which are used in the following analysis are derived in this manner using observations of minor planets by space satellites, notably NASA’s Infrared Astronomical Satellite (IRAS) (Tedesco et al. 2002), Wide-field Infrared Survey Explorer (WISE) (Grav et al. 2011; Mainzer et al. 2011; Masiero et al. 2011), and *Spitzer* (Trilling et al. 2016) and Japan Aerospace Exploration Agency (JAXA)’s Akari (Usui et al. 2011). The method relies on the measurement of the thermal emission of asteroids at infrared wavelengths. Given the distance to the observer and the respective phase angle, this emission can be converted into a total emitted flux, which yields the objects effective diameter using thermal-emission models such as the *Standard Thermal Model* (Lebofsky et al. 1986) or the *Near-Earth Asteroid Thermal Model* (Harris 1999). We may then compute the visual albedo using the absolute magnitude H and the diameter D of an asteroid as (Harris and Lagerros 2002)

$$D = \frac{1329}{\sqrt{p_V}} 10^{-0.2H}. \quad (2.4)$$

2.5 PHASE CURVE COEFFICIENTS

The brightness of celestial bodies as observed from Earth depends on their phase angle α . Fig. 2.2 depicts the maximum observable phase angles for a NEO, an MC, and a Main Belt asteroid. Measuring the reduced magnitude V of an asteroid at different phase angles in the same photometric filter records its *phase curve* $V(\alpha)$. Figure 2.3 shows the phase curve of (20) *Massalia* as observed by Gehrels (1956).

In first order, the brightness of an asteroid decreases with increasing phase angle as less of its illuminated side is visible from Earth. Two smaller-order effects shape the phase curve in detail, imprinting compositional information on it as they depend on the surface parameters such as the regolith grain size and porosity (Li et al. 2015). The *opposition effect* describes a non-linear surge in brightness at phase angles $\alpha \lesssim 5$ deg caused by coherent backscattering of the incident light (Muinonen et al. 2010b). The degree of *mutual shadowing* of the surface regolith dictates the slope of the phase curve at larger phase

For once, the darker asteroids are easier to characterise in an observable as they emit more thermal radiation.

The phases of Venus famously enabled Galileo Galilei to establish the heliocentric system of Aristarchus of Samos.

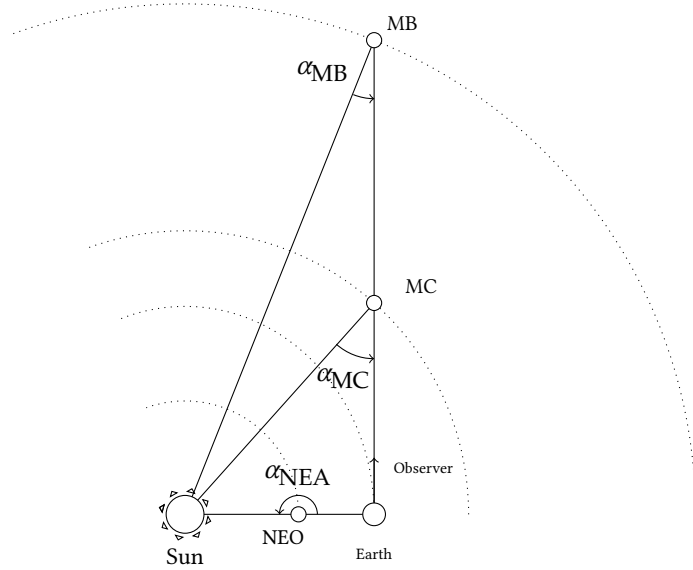


Figure 2.2: The maximum observable solar phase angle α is visualised for different asteroid populations: a near-Earth object (NEO), a Mars-Crosser (MC), and a Main Belt asteroid (MB). With increasing distance to Earth, the maximum observable α decreases, assuming the asteroid does not pass between the Earth and the Sun. α is at a maximum when the object is in quadrature, indicated by the “Observer” arrow.

angles (Muinonen et al. 2002). Dark surfaces reflect less light, therefore the apparent brightness decreases faster with increasing phase angle and the phase curve becomes steeper. The exact surface processes are complex and have been intensively studied and modeled by Hapke (1981), Shkuratov et al. (1999), and Muinonen et al. (2012).

Recording the phase curve of a minor planet requires multiple observational campaigns at different phase angles. To account for the rotational magnitude variation, the target has to be observed for a full rotational period at each visit, typically in the order of several hours (Warner et al. 2009). As such, accurate phase curves are available for only a few hundred objects due to the large amount of required telescope time. Examples of campaigns dedicated for phase curves can be found in Shevchenko et al. (1997), Shevchenko (2002), Shevchenko et al. (2008), and Shevchenko et al. (2016).

As evolution of the H, G model for photometric phase curves (Bowell et al. 1989), Muinonen et al. (2010a) derived a three-parameter model for the phase curve of atmosphere-less bodies,

$$V(\alpha) = H - 2.5 \log_{10} [G_1 \Phi_1(\alpha) + G_2 \Phi_2(\alpha) + (1 - G_1 - G_2) \Phi_3(\alpha)], \quad (2.5)$$

where the Φ_i are basis functions describing the shape of the phase curve at $\alpha > 7.5$ deg (subscripts 1 and 2) and the opposition effect (subscript 3). In the following, I refer to G_1, G_2 as *phase curve coefficients*, while G_1, G_2 in combination with the absolute magnitudes are the *phase curve parameters*.

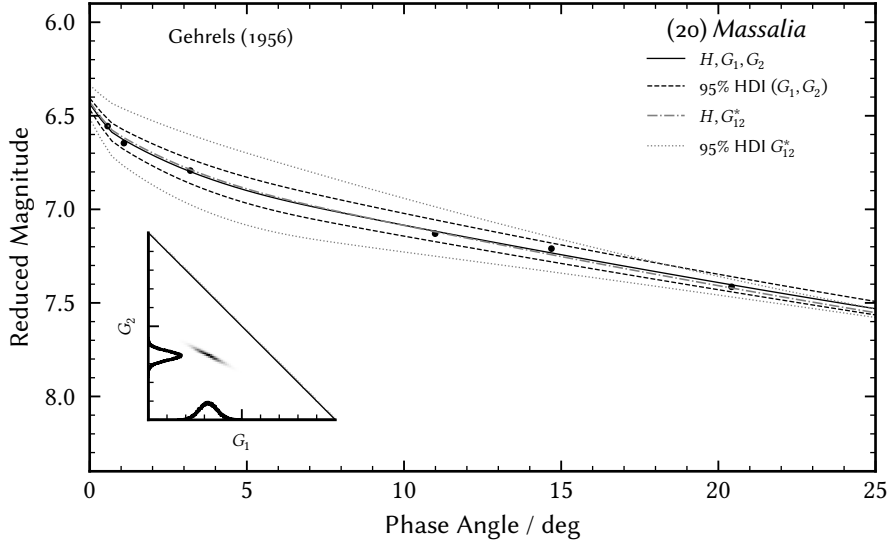


Figure 2.3: The phase curve of (20) *Massalia* as observed by Gehrels (1956) (black dots). The lines give the model fits and error intervals of the H, G_1, G_2 (black) and H, G_{12}^* (grey) models by Muinonen et al. (2010a) and Penttilä et al. (2016). The inset axis shows the samples of the G_1, G_2 model parameters acquired using a Markov Chain Monte Carlo simulation. The fitting method is outlined in Chapter 7. Published in Mahlke et al. (2021).

By linear combination of G_1 and G_2 , a two-parameter function can further be defined in this model for observations where the three-parameter function may lead to overfitting (i. e. for sparsely-sampled phase curves). This H, G_{12}^* -model is later refined observationally by Penttilä et al. (2016) to the H, G_{12}^* -model, where

$$(G_1, G_2) = \begin{pmatrix} 0 \\ 0.53513350 \end{pmatrix} + G_{12}^* \begin{pmatrix} 0.84293649 \\ -0.53513350 \end{pmatrix}. \quad (2.6)$$

Considering the physical constraint that planetary surfaces must get fainter as they are observed at larger phase angles, the G_1, G_2 -space may be confined to

$$G_1, G_2 \geq 0, \quad (2.7a)$$

$$1 - G_1 - G_2 \geq 0. \quad (2.7b)$$

As mentioned above, the compositional information of the minor planet's surface is contained in the photometric slope k , in particular in the phase angle range between 0 deg and 7.5 deg, and the size of the opposition effect $\zeta - 1$. These quantities can be expressed in terms of G_1, G_2 following Muinonen et al. (2010a) as

$$k = -\frac{1}{5\pi} \frac{30G_1 + 9G_2}{G_1 + G_2}, \quad (2.8)$$

and the size of the opposition effect $\zeta - 1$ as

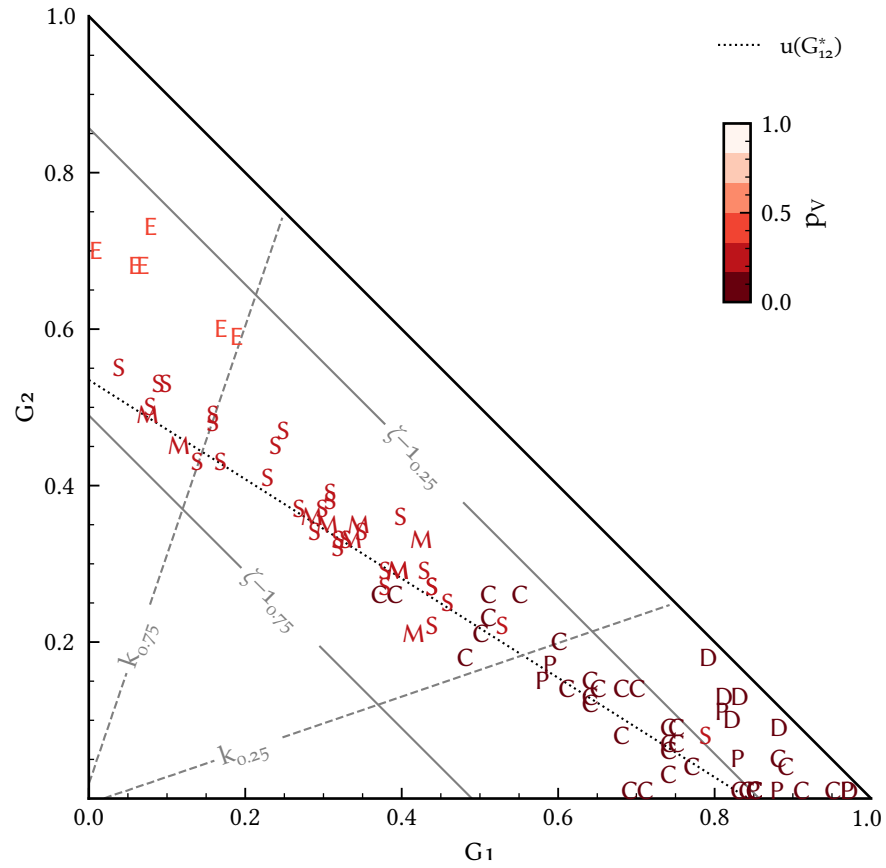


Figure 2.4: The G_1 , G_2 parameter space related to the photometric slope, opposition effect, and taxonomic classes. The dashed lines show 25- and 75-percentiles of the photometric slope k . As the slope is negative, the phase curve becomes steeper towards the smaller percentiles. The solid lines give the size of the opposition effect $\zeta - 1$, again for the 25- and 75-percentiles. Further shown are the coefficients of targeted phase curves presented in Shevchenko et al. (2016), where the position of the G_1 , G_2 -parameters is indicated by the taxonomic class of the respective asteroid.

$$\zeta - 1 = \frac{1 - G_1 - G_2}{G_1 + G_2}, \quad (2.9)$$

where k is in units of mag/rad and $\zeta - 1$ gives the additive contribution of the opposition effect to the absolute magnitude in units of mag. In Figure 2.4, I summarise the G_1 , G_2 -space and the results of Shevchenko et al. (2016), using taxonomic classes of asteroids which are outlined in Chapter 4.

Belskaya and Shevchenko (2000) show that k has a negative correlation with the albedo, i. e. darker surfaces exhibit steeper phase curves, while $\zeta - 1$ peaks for asteroids with moderate albedos and diminishes for both asteroids with large or small albedos. The correlation between the shape of the phase curve and the albedo reminds of the relationship between colours and spectra: one is the more accessible but lower resolution version of the other.

2.6 HONORARY MENTIONS

There are additional ways to study asteroids which yield compositional information. I did not consider them for the new taxonomy in this work, yet I believe that they could provide valuable additional information to the ones used in this study if they are recorded for a sufficiently large sample of asteroids.

2.6.1 *UV and MIR Spectroscopy*

The extension of the spectroscopically studied wavelength ranges into the ultra-violet (UV) could facilitate the differentiation of asteroids which are feature-less in the VisNIR regime. UV spectra of 28 asteroids obtained by the European Space Agency (ESA) *Internal Ultraviolet Explorer* satellite show intriguing differences in slope and feature-presence between asteroids commonly associated to the same taxonomic classes (Butterworth and Meadows 1985). The separation of feature-less asteroids in the Eight-Color Asteroid Survey (ECAS) colors as studied by Tholen (1984) (refer to Chapter 4) is a further indication that the UV is compositionally informative. A downside of the UV range is the decrease in the Solar emission spectrum and an increase in the telluric absorption (refer to Fig. 2.1), which means that only the brightest asteroids are reliably observed in this regime from Earth. The recent third data release by the ESA *Gaia* satellite (Galluccio et al. 2022; Tanga et al. 2022; Vallenari et al. 2022) contains a few data points in the UV area, allowing to estimate general trends in this region for more than 60 000 asteroids.

Further of interest is the 3 μm -region in the mid-infrared (MIR), where water ice and hydroxyl leave their absorption fingerprints. The presence of water on planetary surfaces narrows down their formation history and has implications for the search of life outside the Earth. As we will see in the next chapter, the visible wavelength region contains two absorption bands around the 0.7 μm which indicate the presence of hydroxyl (OH), yet comparative studies of the visible- and 3 μm -regions of individual asteroids show that this band is only present in about half of the cases where water is observed in the 3 μm -region (Rivkin et al. 2015b).

2.6.2 *Polarimetry*

As the incident light gets reflected of the asteroid surface, it gets polarised through the interaction with the surface regolith. The polarisation may be decomposed into two parts, one where the vector of polarisation is orthogonal to the layer of reflectance and one where it is parallel to the layer. The *polarisation degree* P_r of the light is a combination of these components and indicates whether the parallel or the orthogonal part is dominating. It is negative when the parallel component is superior, which is the case for asteroid surfaces observed at small phase angles. As the asteroid is observed at larger phase angles, however, the polarisation degree becomes positive at different phase angles, depending on the surface parameters. The phase angle of the transition between negative and positive polarisation degree

is referred to as the inversion angle of the negative branch of polarisation (Belskaya et al. 2015).

The inversion angle and other characteristic parameters of the polarisation curve have been shown to be different from the ensemble of asteroids for taxonomic class with fewer asteroids, which are difficult to identify in both spectra and visual albedo, namely the F-class, which requires UV information to be distinguished from the B-class, and the enigmatic L-class (Belskaya et al. 2015; Devogèle et al. 2018). As such, the addition of polarimetric curve parameters to the taxonomic space would be highly valuable.

I discuss the polarimetric behaviour of some asteroids and the L class more in-depth in [Chapter 14](#).

2.6.3 Radar

Radar observations of minor planets are unique in the sense that the source of the reflected irradiation is not the Sun but a transmitter on Earth. The inverse-square law has to be applied twice during these observations, meaning that the intensity of the radar echo scales inversely proportional to the distance between the observer and the minor planet to the power of four. As such, radar observations are preferentially done for NEOs, where they yield fascinating insights into the surface structure, shape, and multiplicity of objects (Benner et al. 2015). An example is the analysis of (216) *Kleopatra* by Shepard et al. (2018), who show that the asteroid is a dumbbell-shaped contact binary with an intriguing dichotomy in its radar albedo, being radar-bright at the poles and radar-dark around the equator.

A high surface radar albedo is a smoking-gun observation for metal on or near the surface of asteroids (Shepard et al. 2015). Metal-rich asteroids have only few if any features in their reflectance spectra and the visual albedo is degenerate with other classes. Hence, a clear characteristic of these supposed cores of early planetesimals would be of high values. Furthermore, the radar echo is polarised as it is reflected on the surface. In the so-called *circular polarisation ratio*, the E- and V-types are readily distinguished from other classes (Benner et al. 2008). While V-types are typically the first group to identify using spectroscopy, this could be beneficial for the identification of E-types when lacking visual albedo information as they are spectrally degenerate with some M- and L-types.

METEORITES

3.1	Terminology	26
3.2	Compositional Constituents	27
3.2.1	Minerals	27
3.2.2	Building Blocks	29
3.2.3	Alteration and Metamorphism	30
3.3	Classification	31
3.3.1	Chondrites	31
3.3.1.1	Ordinary Chondrites	31
3.3.1.2	Carbonaceous Chondrites	33
3.3.1.3	Enstatite Chondrites	33
3.3.2	Achondrites	34
3.4	Interplanetary Dust Particles	34
3.5	Spectral Ambiguity in Matching Asteroids and Meteorites	34
3.5.1	Space- and Terrestrial Weathering	35
3.5.2	Grain Size	35
3.5.3	Temperature	36

In the previous two chapters, I outlined where asteroids are located and how we can decipher their chemical composition using remote-sensing observations. This chapter describes the compositions we may expect to observe of asteroids. To this end, I make use of the terrestrial meteorite collection.

Meteorites are fragments of celestial bodies on Earth (Chladni 1794; Olbers 1803). They enable us to study in-depth the mineralogy and petrology of their parent bodies in space. More than 50 000 meteorites have been found on Earth so far, making up the terrestrial meteorite collection.¹ Meteorites give fascinating insights into the early Solar System, from the precise age of the Solar System (Amelin et al. 2010), the dichotomy of the material in the protoplanetary disk (Warren 2011) and the chief role of Jupiter in maintaining it (Kruijjer et al. 2017), to the timing of meteorite delivery to Earth via the timeline told by exposure to cosmic rays (Herzog 2007).

However, when studying the composition of asteroids through the meteorites, there are important caveats to consider. For one, we do not know which parent sent which meteorite. As such, we may only infer on the properties of bodies which look like the meteorite at hand, without the final degree of certainty. The identification of the asteroid parent body of meteorites is one of the fundamental goals of planetary science and I briefly outline the challenges at the end of this chapter and possible solutions to them in Chapter 14. Furthermore, there is a strong selection bias in terms of which asteroids become meteorites, first concerning the transition from the Main Belt to near-Earth space (refer to Chapter 1) and later at the Earth's entry

Olbers considered that meteorites are predominantly of lunar origin (“moon volcanoes”) yet raised the issue that this would require an immense mass loss of the Moon to explain the flux of meteorites on Earth (Olbers 1803).

¹ The number of meteorites per class and fall/find statistics given throughout this chapter are extracted from the *Meteoritical Bulletin Database* at <https://www.lpi.usra.edu/meteor/>.

level. Meteorites need a certain mass and cohesive strength to survive the friction and heat applied as they traverse the terrestrial atmosphere. Those that do pass the trial are altered on their surface by the journey through the atmosphere and the abrupt end upon impact on the ground, acquiring a fusion crust and getting deformed by the shock pressure (Burbine 2016).

As such, there is the interesting contrast that we may study meteorites with the microscope at incredible resolution yet when we look up to see where they came from our vision is blurred. In this chapter, I describe the diversity and spectroscopy of meteorites. This is an extensive topic and to cover it briefly hardly does it justice. I do frequent simplifications and focus on the minerals which I encountered most, and refer the reader to Brearley et al. (1998), Weisberg et al. (2006), Krot et al. (2014), Burbine (2016), and Cloutis et al. (2018) for in-depth introductions to the topics covered in this chapter. I conclude with aspects which have to be considered when aiming to match asteroids and meteorites via spectroscopy.

3.1 TERMINOLOGY

Entering the world of meteorites (and even the one of asteroids) as a physicist by training, I frequently encountered geological vocabulary that was foreign to me. In the following, I briefly outline the definitions of common terminology which may not be obvious.

When a geological body gets heated to the point of melting, *differentiation* occurs. The *siderophile* (iron-loving, Goldschmidt (1937)) elements dissolve in iron and sink to the core of the body due to their high density, while the lighter *lithophile* (stone-loving) elements form the crust and the mantle in form of igneous (i. e. magmatic) rock. The body is then said to be *differentiated* (e. g. Scheinberg et al. 2015).

Another distinction of chemical elements is between the *volatile* and *refractory* elements. The former have condensation temperatures below ~ 1000 K while the latter condensate at temperatures above ~ 1500 K. This distinction is in particular applied when discussing the elements present in different regions of the protoplanetary disk. The heat of the young Sun depleted the inner Solar System of volatile elements, leaving the refractory metals and silicate minerals as primary constituents of the planetesimals forming in this region. The presence of volatile elements in meteoritic bodies is thus a smoking-gun observation for formation in the outskirts of the Solar System (Krot et al. 2015).

And, to conclude, the simple distinction of meteorites as *falls* if their fall was observed and can be dated and as *finds* if this is not the case. This distinction serves to communicate the potential alteration of the meteorite due to prolonged exposure to the Earth's environment (Hughes 1981).

3.2 COMPOSITIONAL CONSTITUENTS

3.2.1 Minerals

The compositional building blocks of minor planets leave a characteristic fingerprint in the spectroscopic observations. The description of these fingerprints by means of crystal field theory is the foundation to relate certain absorption features to a transition in the crystal structure of a given mineral (Adams 1975). The different building blocks of meteorites are made up of over 300 different minerals (Jones et al. 2005; Cloutis et al. 2018), yet, we may describe the majority of the variance observed in asteroid reflectance spectra with just a few, optically-dominant minerals (Burbine 2016).

Silicate minerals are the most abundant minerals in meteorites (Brearley et al. 1998). Their defining constituent is the SiO_4 anion. The most common minerals in meteorites include the silicates olivine and pyroxene. These minerals are of high practical importance for the study of asteroids as most asteroid spectra are well explained by some combination of these two. Olivine has the chemical formula $(\text{Fe,Mg})_2\text{SiO}_4$, meaning that the SiO_4 anion is accompanied by either an Fe_2 or Mg_2 cation. Assemblages of olivine show different ratios of iron or magnesium cations, and the VisNIR reflectance spectra of the two endmembers, the iron-rich fayalite (Fe_2SiO_4) and the magnesium-rich forsterite (Mg_2SiO_4), are given in the upper part of Fig. 3.1. The signature feature of olivine are three superimposed absorption bands around $1\ \mu\text{m}$ due to Fe^{2+} crystal field transitions (Gaffey 1974). We see that the bands of fayalite are at slightly larger wavelengths than the ones for forsterite. The Fe_2 cation is larger than the Mg_2 cation, hence the electronic potential in the fayalite system is smaller and less energy is absorbed.

Pyroxene is another group of silicate minerals which is ubiquitous in asteroid spectra. While the olivine group has two, the pyroxenes have three cation substitutions with which we concern ourselves here, $(\text{Fe,Mg,Ca})_2\text{SiO}_6$. This leads to a large variety of possible pyroxene compositions and spectral appearance, of which the common denominator are two prominent absorption features around $0.9\ \mu\text{m}$ and $1.9\ \mu\text{m}$ (Gaffey 1974). The spectra of the iron-rich endmember ferrosilite (Fe_2SiO_6), the magnesium-rich endmember enstatite (Mg_2SiO_6), and high-calcium pyroxene augite ($20\% < \text{Ca} < 45\%$) are shown in the middle part of Fig. 3.1.

A further common siliceous component of meteorites are feldspars, though they are spectrally largely featureless and thus less relevant for remote-sensing (Cloutis et al. 2018).

Finally, there are the important phyllosilicates which contain water (H_2O) or hydroxyl (OH). These hydrated silicates indicate that the body has come into contact with water. In the lower part of Fig. 3.1, the spectra of saponite and serpentine are shown. In asteroid spectra, we typically only observe the absorption features around $0.7\ \mu\text{m}$ (Cloutis et al. 2018).

Besides siliceous elements, some members of other mineral groups are commonly identified in asteroids. Carbon-rich meteorites show large contents of graphite (Brearley et al. 1998), commonly identified in iron-rich and siliceous meteorites (Jones et al. 2005). Compared to terrestrial rocks, mete-

Two great places to get mineral and meteorite spectra are spectral library of the United States Geological Survey and the Reflectance Experiment Laboratory (RELAB) of Brown University (Pieters 1983).

Millimeter-sized diamonds “of little commercial value” have further been identified in some meteorites. (Foote 1891).

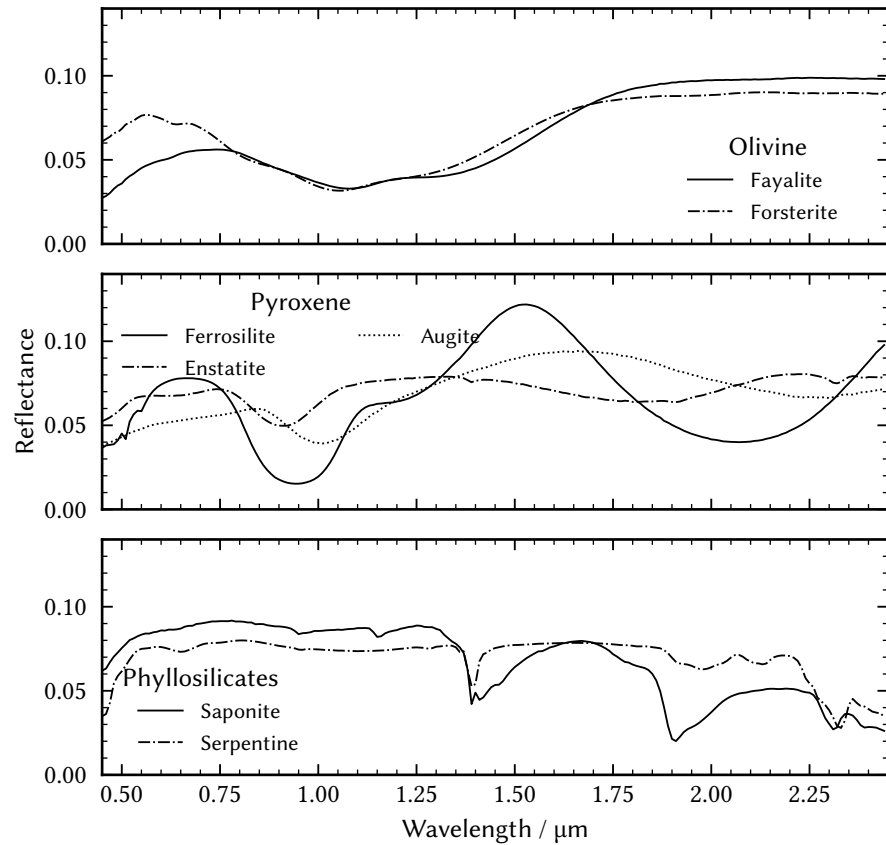


Figure 3.1: Reflectance spectra of silicate minerals commonly found in meteorites. All spectra use the same wavelength sampling pattern and have been normalised to unity in their L_2 -norm for comparability (refer to Chapter 9). The spectra of fayalite (C1DD98) and ferrosilite (C1DL61A) are from the RELAB database (Pieters 1983), the remaining spectra are from the spectral library of the United States Geological Survey at <https://crustal.usgs.gov/speclab/>.

orites are enriched in nickel as nickel is siderophile and thus predominantly ended up in Earth's core during the differentiation. The spectra of graphite and taenite is given in the upper part of Fig. 3.2 as well. Both the spectra of graphite and taenite have no discernible absorption features in the VisNIR. However, carbon- or iron-rich asteroids get characteristic slope behaviours from these minerals.

Magnetite is an oxidation product of iron, indicating that the parent body has gotten in contact with liquid water. It has a characteristic, wide 1.3 μm -band shown in the middle part of Fig. 3.2. Finally, aluminous spinel is a magnesium iron oxide commonly found in calcium-aluminium-rich inclusion (CAI). The importance of detecting hydration and CAI in asteroids is outlined below and in parts in Chapter 14.

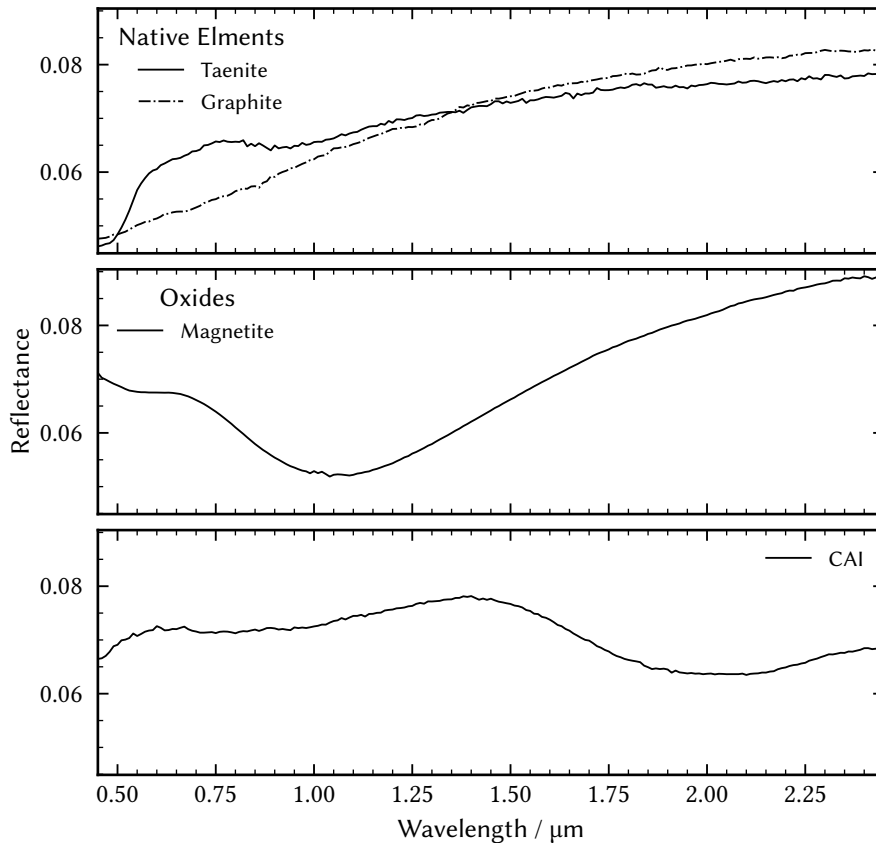


Figure 3.2: Spectra of common minerals in meteorites and the powder of two calcium-aluminium-rich inclusion extracted from CV chondrite Allende. The spectra of taenite (C1MB16), graphite (CAGP05), and the calcium-aluminium-rich inclusion (CACA00) are from the *RELAB* database (Pieters 1983), the spectrum of magnetite is from the spectral library of the *United States Geological Survey* at <https://crustal.usgs.gov/speclab/>.

3.2.2 Building Blocks

Most meteorites are in first order an agglomerate of three basic components (Krot et al. 2014), each of which is thought to have formed separately in the protoplanetary disk (Ebel 2006; Fedkin and Grossman 2006): chondrules, refractory inclusions, and matrix material. The former two are embedded in the matrix. Iron-nickel alloys are further frequent constituents of meteorites.

Chondrules are millimeter-sized grains of varying composition and texture (Jones et al. 2005). They are grouped into two types, type I being depleted in FeO and volatile elements while type II is rich in FeO and has solar composition. The exact formation scenario is an open issue due to numerous constraints (Sears 2004; Ciesla 2005; Hewins et al. 2005), yet the general consensus is that chondrules formed by first rapid heating followed by rapid cooling of grainy material in the protoplanetary disk. The study of chondrules and their mineralogy thus allows inference on the precursor material in the protoplanetary disk as well as events of heating.

Refractory inclusions are the oldest solids in the Solar System with an age of 4567.3 ± 0.16 Myr (Krot 2019). They contain the most refractory elements and are built from the first condensates in the hot protoplanetary disk irradiated by the Sun in its early evolutionary stage (Ebel 2006). Refractory inclusions broadly fall into two categories, CAI and amoeboid olivine aggregates. The most common CAIs are rich in aluminous spinel, while amoeboid olivine aggregates are assemblages including forsterite and iron-nickel (Komatsu et al. 2001). A uniform formation history is unlikely for CAI due to their textural diversity. Most are thought to have crystallised after complete or partial melting events (Beckett et al. 2006). They might be precursor material of chondrules (Russell et al. 2000; Krot and Keil 2002). Refractory inclusions are of importance for cosmochemistry as they date the condensation of the first solids, the time zero of the solar system. The spectrum of a sample of powdered CAIs from the Allende meteorite is shown in the bottom part of Fig. 3.2. The most pronounced feature is at $2 \mu\text{m}$ and associated to aluminous spinel (Sunshine et al. 2008).

The matrix is the “background material” of the meteorite, made of highly comminuted, i. e. fine-grained material with a variety of mineralogies. It is of particular interest when studying the events that altered the composition of the meteorite’s parent body after its formation, such as events of heating or hydration (Zolensky et al. 1993).

3.2.3 Alteration and Metamorphism

After their formation, the history of the parent bodies of meteorites frequently included events of heating or hydration, triggering metamorphism of their minerals and building blocks (Krot et al. 2006). These events affect their textural and chemical appearance today, thus, we may conclude on the history of the body by looking for signatures of certain minerals.

Thermal Metamorphism (TM) of parent bodies is triggered by heat, which is for example generated by the decay of radioactive nuclei (Huss et al. 2006). Most meteorites show tracers of TM to varying degrees of strength, with some meteorites having been heated sufficiently to melt either partially or completely (Weisberg et al. 2006). These tracers include the equilibration of minerals, meaning that the minerals become more homogeneous. An example is the enrichment of iron-poor forsteritic olivine in iron from the iron-rich matrix minerals (Cloutis et al. 2018). The altered crystal structure of the olivine is reflected in deeper absorption bands and a shift of the band minima towards longer wavelengths. We regard this effect in some asteroids further in Chapter 14.

Aqueous Alteration (AA) occurs when the parent body comes into contact with liquid water, e. g. by melting of previously accreted water (Brearley 2006). Liquid water triggering AA appears to have been widespread in the early Solar System and affected most types of meteorites (e. g. Zolensky et al. 1999). The hydration triggers a metamorphism of some minerals into others. Anhydrous silicates such as olivine and pyroxene transform into phyllosilicates while iron oxidises to form magnetite. Minerals such as serpentine (bottom row of Fig. 3.1), and magnetite (middle row of Fig. 3.2) may thus allow us to

conclude that the asteroid or its parent body has been hydrated at some point throughout its history.

Melting of water ice requires some form of heat source, which shows that TM and AA are not strictly separate processes (Huss et al. 2006).

3.3 CLASSIFICATION

Meteorites are classified into classes, groups, and subgroups based on their mineralogical, petrographic, and isotopic properties (Weisberg et al. 2006). The most basic distinction is between the chondrites and the achondrites. The former have primitive, solar-like compositions except for the volatile elements and are associated to undifferentiated parent bodies (Sears and Dodd 1988). The latter consist of igneous rocks like basalts and are associated to differentiated asteroids and larger celestial bodies like terrestrial planets and the Moon. Most chondrites contain chondrules while achondrites do not due to melting events. An overview of the classification and connections to the meteorites is given in Fig. 3.3.

3.3.1 Chondrites

Among the chondrites, there are three classes, Carbonaceous Chondrites (CCs), Ordinary Chondrites (OCs), and enstatite chondrites, which are further divided into groups. Each group is supposed to come from a single, unique asteroid parent body. The different groups of chondrites are primarily distinguished by their SiO_2/MgO ratios (Van Schmus and Wood 1967). Meteorites in each group are further differentiated based on their Petrologic Type (PT). The PT is a number between 1-7 used to indicate the degree of TM (PT of 3,1-7) or AA (PT of 1-2), with 3.0 representing pristine material (Van Schmus and Wood 1967).

Two smaller groups of the chondrites are the Kakangari and Rumuruti chondrites. Kakangari chondrites are similar to OCs, however, they differ significantly in the O-isotopic composition (Weisberg et al. 2006). Rumuruti chondrites are technically a grouplet as they have fewer than five members.

3.3.1.1 Ordinary Chondrites

OCs are divided based on their total iron abundance and the ratio of metallic iron to total iron (Fe^0 / Fe) (Weisberg et al. 2006) into the groups H, L, and LL. H chondrites have a high total iron fraction, while L chondrites have a low total iron fraction, and LL have both low total iron and low total metal. OCs are the most common meteorite by far.

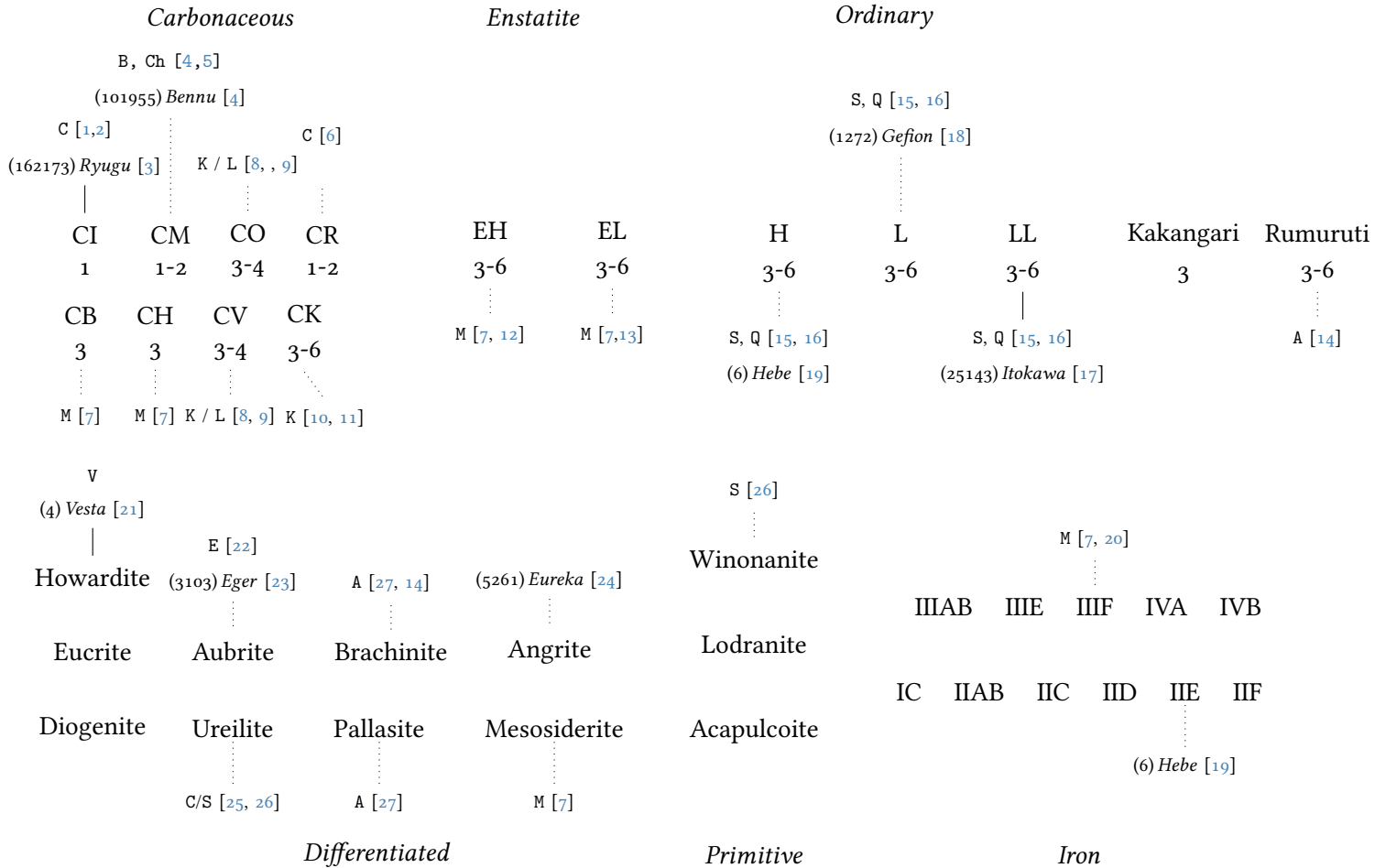
OCs are primarily composed of olivine, pyroxene, and metal, with a large fraction of chondrules while CAI are rare. They have a PT of 3 or higher, though hints of AA have been reported for some OCs (Zolensky et al. 1999; Grossman et al. 2000).

5 ungrouped meteorites of similar mineralogies, chemistries, and oxygen isotopes are needed to form a meteorite group (Weisberg et al. 2006).

"Ordinary" is a rather unfortunate name as it understates their value for planetary science. It relates to the ubiquity of OCs among meteorites.

The meteorite which impacted over the city of Chelyabinsk in 2013 is an LL chondrite linked to the Flora- or Baptistina families (Reddy et al. 2014).

CHONDRITES



ACHONDRITES

Figure 3.3: The classification scheme of meteorites following Weisberg et al. (2006). The numbers below the group names of chondrites give the observed range of Petrologic Type in the members of the group, following Krot et al. (2014). Dashed lines connect meteorite groups to asteroid classes introduced in Chapter 4, following the overview given in Burbine (2016). The references given in brackets are given in Table 3.1.

3.3.1.2 Carbonaceous Chondrites

The class of CCs is divided into eight groups based on bulk composition, oxygen isotopic ratios, elemental ratios (Van Schmus and Wood 1967; Weisberg et al. 2006). Some classes form larger clans of meteorites if they show compositional similarities (Weisberg et al. 2006). CCs generally have PT between 1 and 3.

The CCs groups are named using a double-letter scheme, the first letter being the C for carbonaceous and the second being the first letter of a prominent member of the group (with one exception), which is mostly the first discovered one.

CI (*Ivuna*) have solar composition and no chondrules. They consist almost entirely of matrix material and lack CAIs or chondrules (Brearley et al. 1998). CI chondrites are of high interest as their elemental abundances closely match the ones of the solar photosphere apart from the volatile elements (Anders and Grevesse 1989). This indicates that they represent closely the material of the solar nebula, and they are used as reference frame for the abundances in other chondrites. Unfortunately, they are extremely rare meteorites, with only 9 entries of CI chondrites in the *Meteoritical Bulletin Database*.

CM (*Mighei*) chondrites are abundant in phyllosilicates (Brearley et al. 1998). They consist mostly of matrix material, with some chondrules and fewer CAIs. They are the most abundant class of CCs. The CO (*Ornans*) chondrites show similarity in the oxygen isotopic properties to the CM, hence the two build a clan. CO chondrites have PT between 3-3.7, while most CM chondrites are of PT 2 (Weisberg et al. 2006). CO chondrites further have the largest percentage of CAIs with respect to their volume, given at 13 vol% by Brearley et al. (1998).

Another clan consists of the CK (*Karoonda*) and CV (*Vigarano*) chondrites. Both show the largest chondrules in terms of diameters found in chondrites at around 1 mm in diameter (Weisberg et al. 2006). Like CO chondrites, CV chondrites have a large amount of CAIs with 10 vol% (Brearley et al. 1998). CK chondrites are the only CCs which have PT of larger than 4, spanning the range between 3-6, while all CV chondrites are of PT 3 (Krot et al. 2014).

Finally, the CR (*Renazzo*), CH (*high metal*), and CB (*Bencubbin*) chondrites form the CR clan. They are enriched in metal with respect to the other classes of CCs though the CB chondrites have by far the largest abundance of metal at around 60 vol% (Weisberg et al. 2006; Krot et al. 2014).

3.3.1.3 Enstatite Chondrites

Enstatite chondrites primarily consist of enstatite pyroxene and iron metal. They are further divided into EH chondrites (*high iron*) and EL chondrites (*low iron*) (Weisberg et al. 2006). Both CAI and matrix material are rare in enstatite chondrites (Krot et al. 2014). Their silicate components are poor in iron (Weisberg et al. 2006). Enstatite chondrites are the most reduced chondrites and thought to best represent the precursor material of Earth in terms of composition (Javoy et al. 2010).

As for OCs, the carbonaceous is an unfortunate historical designation as many CCs contain only little carbon.

The Chicxulub impactor responsible for the demise of the dinosaurs likely was a CM chondrite (Manning et al. 2022).

3.3.2 Achondrites

Achondrites are igneous rocks that come from differentiated parent bodies. They may be broadly divided into irons and stony-irons. For historical reasons, achondrites are less structured and groups of achondrites are no longer supposed to originate from the same asteroid parent body, apart of the groups of iron meteorites (Scott 1972). I do not mention all groups of achondrites, instead focusing on the ones that reappear in the latter parts of the thesis and referring the reader to (Weisberg et al. 2006) or Krot et al. (2014) for a full overview.

Iron meteorites consist primarily of the iron-nickel alloys kamacite and taenite. They are divided based on their nickel, gallium, germanium, and iridium content into 14 groups, using a combination of roman numerals (I-IV) followed by one or two capital letters (Goldstein et al. 2009). IIIAB is the largest group of iron meteorites.

The howardite, eucrite, and diogenite groups build the HED clan of achondrites. They are linked to (4) *Vesta* (McCord et al. 1970; Mittlefehldt 2015). They make up almost three quarters of all achondritic meteorites.

Aubrites are achondrites consisting between 75%–98% of iron-free enstatite (Watters and Prinz 1979). Their precursor material is likely the same as of enstatite chondrites. The lack of transition metals gives them a white colour, equivalent to a high albedo (Burbine 2016).

Further of interest for the remained of the thesis are angrites, olivine-rich meteorites thought represent the crust of a disrupted parent body (Krot et al. 2014), meosiderites, mixtures of iron and silicates, pallasites, objectively beautiful inclusion of olivine into a metal, and brachinites, which are olivine-rich yet appear only partially differentiated (Krot et al. 2014).

3.4 INTERPLANETARY DUST PARTICLES

It is worth briefly mentioning Interplanetary Dust Particles (IDPs) here as they have been proposed to represent material from a large fraction of asteroids (Vernazza et al. 2015). IDPs are micrometer-sized grains that orbit the Sun and may become meteoritic material due to solar wind or light deviating their orbits (Bradley 2014). They are divided into chondritic and non-chondritic particles based on their mineralogy and porous and smooth particles based on textures. Chondritic IDPs have similar bulk compositions to CI and CM chondrites. Further subdivisions based on the dominant constituent mineral (pyroxene, olivine, or phyllosilicates) are recognised (MacKinnon and Rietmeijer 1987).

3.5 SPECTRAL AMBIGUITY IN MATCHING ASTEROIDS AND METEORITES

Given the detailed knowledge of meteoritic mineralogy and petrology, the process of forging links between asteroids and meteorites focuses on the comparison of their reflectance spectra (Gaffey 1993; Binzel 1995; DeMeo et al. 2022). However, reflectance spectra of rocks are more than the sums of their

parts: the resulting spectrum is a complex function of the optical properties of the materials and their relative abundances (Gaffey and McCord 1979). In addition, the spectra of asteroids are a function of mineralogy, surface properties such as grain size, and secondary effects such as space weathering (Cloutis et al. 2018).

Here, I discuss effects which have a strong impact on the spectral appearance of asteroids and meteorites and need to be established before definitive matches can be claimed. Some points mentioned here are further regarded in Chapter 14.

3.5.1 *Space- and Terrestrial Weathering*

As asteroids are atmosphere-less planetary bodies, their surfaces are subject to space weathering (Zeller and Ronca 1967). This term encompasses surface alterations triggered by the exposure to the space environment, e. g. impact of micrometeorites or bombardment with heavy ions through the solar irradiation or cosmic rays. This effect was demonstrated to affect lunar samples by Adams and McCord (1971).

The spectral alteration of the surface depends on its mineralogy. Siliceous asteroids generally redden with increasing degree of space weathering, while the albedo diminishes and absorption features get smaller (Chapman 1996; Brunetto et al. 2006; Chrbolková et al. 2021). For carbonaceous asteroids, the response is much less certain. Statistical treatment of asteroid families shows both an increase and a decrease in the spectral slope (Lantz et al. 2018; Thomas et al. 2021), confirmed by laboratory experiments of different CCs (Lantz et al. 2017).

Meteorites exposed to Earth's environment suffer from terrestrial weathering. Iron oxides start to form as soon as the meteorite gets in contact with the water vapor in the Earth's atmosphere (Salisbury and Hunt 1969; Salisbury and Hunt 1974). This decreases the overall reflectivity of the spectral region below 0.7 μm and increases the steepness of the spectrum in the same region (Gooding 1982). The 3 μm region is further affected by the hydration due to liquid water.

3.5.2 *Grain Size*

The spectral appearance of meteorites depends on the sample preparation, in particular, the grain size of the powder. Using UVB colours of lunar, martian, and meteoritic samples, Hapke (1971) pointed out the alteration of the visible spectral region due to varying grain size. For CCs, variations in grain size are shown to affect the spectral slope in (Johnson and Fanale 1973). Cloutis et al. (1986) illustrate that grain size on siliceous asteroids affects the albedo, yet not the position or strength of the absorption features. The spectral appearance is dominated by the smallest particles (Mustard and Hays 1997).

The open issue is the grain size of the particles on the surfaces of asteroids. The consensus is that diameter and average regolith grain size are anti-correlated, based on observations of thermal properties (Delbó et al. 2007), reflectance spectroscopy in the mid-infrared (Vernazza et al. 2010) and VisNIR

Vernazza et al. (2016), and asteroid rendez-vous mission to (4) *Vesta* (Russell et al. 2012), (21) *Lutetia* (Sierks et al. 2011), and (25143) *Itokawa* (Miyamoto et al. 2007). This is intuitively explained by the increasing gravitational potential with asteroid diameter, making it easier to retain fine grains on the surface, as well as the larger surface age for larger asteroids leading to a higher fraction of produced fine-grained regolith due to thermal fatigue (Delbó et al. 2014).

3.5.3 Temperature

Meteorites are studied at room-temperature, while asteroids are generally at lower temperatures. The reflectance spectra of minerals depend on the temperature. Bands widths decrease and the band centers move to shorter wavelengths (Singer and Roush 1985; Reddy et al. 2015).

Table 3.1: References for proposed asteroid-meteorite matches shown in Fig. 3.3.

[1] Johnson and Fanale (1973)	[10] Clark et al. (2009)	[19] Gaffey and Gilbert (1998)
[2] Cloutis et al. (2011)	[11] Cloutis et al. (2012a)	[20] Cloutis et al. (1990)
[3] Yada et al. (2022)	[12] Chapman and Salisbury (1973)	[21] McCord et al. (1970)
[4] Clark et al. (2011)	[13] Gaffey and McCord (1979)	[22] Zellner (1975)
[5] Rivkin (2012)	[14] Sunshine et al. (2007)	[23] Gaffey et al. (1992)
[6] Hiroi et al. (1996)	[15] Gaffey et al. (1993a)	[24] Rivkin et al. (2007)
[7] Shepard et al. (2010)	[16] Binzel et al. (2004c)	[25] Jenniskens et al. (2009)
[8] Bell et al. (1988)	[17] Nakamura et al. (2011)	[26] Gaffey et al. (1993b)
[9] Sunshine et al. (2008)	[18] Nesvorný et al. (2009)	[27] Cruikshank and Hartmann (1984)

TAXONOMY

4.1	Principles and Design	37
4.2	Complexes of Asteroids	39
4.2.1	C-Complex	39
4.2.2	S-Complex	40
4.2.3	X-Complex	41
4.2.4	Endmembers	42

The systematic categorisation of objects into groups based on a metric of similarity is referred to as *taxonomy*. Taxonomies are ubiquitous in science, classifying diverse objects such as plants (Candolle 1813), languages (Ferguson 1959), or stars (Morgan et al. 1943). Another example is the meteorite taxonomy we encountered in Chapter 3. The reason that taxonomies are omnipresent is that they simplify communication. By designing and agreeing upon a common terminology, we can express complex behaviour in populations with simple terms.

The asteroid taxonomy has been in development for almost 50 years, since the technological advancement of the photoelectric tubes allowed to characterise at first photometric colours and later spectra of an increasing number of minor planets (McCord et al. 1970; Gehrels 1979). It was quickly established that asteroid classes should be described by a single letter which may hint towards a mineralogical interpretation, such as the carbonaceous C, the siliceous S, the *metallic* M, and the *enstatite* E (Zellner 1973; Chapman et al. 1975; Zellner and Gradie 1976). Changes in the class definitions, assignments of asteroids, and new letter additions quickly lead to what Bobrovnikoff et al. (1983) referred to as the “alphabet soup”.

In this chapter, I conclude the introductory part of this thesis by outlining the recipe and main ingredients of the asteroid taxonomy. I outline the basic principles that are underlying taxonomies in general and the asteroid taxonomy in particular before confronting these principles with the reality of a practical design. I then describe the basic classes and complexes of the asteroid classification which are found in the major taxonomies. The chapter concludes with an overview of the classification over time in Fig. 4.1.

4.1 PRINCIPLES AND DESIGN

Taxonomy is a framework of communication, a tool to simplify understanding between researchers in a field. If we understand taxonomy as a language, we can intuitively come up with some principles for the design of an effective taxonomy.

First, we want to use simple terms to describe the groups of populations. These simple terms may invoke associations to interpretations, yet, this should be used sparsely. The asteroid taxonomy achieves this using mostly

single letters to indicate classes of the asteroids. Less elegant is the meteorite taxonomy in this regard, where the terms “ordinary” and “carbonaceous” chondrites have strong connotations on the nature of the objects. Weisberg et al. (2006) thus proposes to use single letters like O and C chondrites akin to the asteroid system.

Furthermore, we should construct a language with a large enough vocabulary to describe the maximum compositional entropy among our groups, while at the same time keeping it concise and easy to understand. This is usually achieved by subclassing, i. e. building a decision-tree of classes. An example is the subclassing in the stellar taxonomy, where the first letter indicates the temperature of a star in a broad range, the following number in a finer range, and a Roman numeral to describe the luminosity of a star (Morgan et al. 1943). This way, we can describe the Sun in the context of all G-class stars or as one of few G2V stars. The asteroid lacks an elegant subclassing scheme like this.

Finally, just as languages, we have to consider that taxonomies evolve. Given new data or observational features, the description of classes may have to be updated to facilitate communicating these new insights. When revising a taxonomy to update the vocabulary, we want to remain as closely as possible to the existing scheme. Significant changes to the vocabulary will lead to confusing and delay or even prevent the adoption of the revision.

For the asteroid taxonomy in particular, one of the core concepts is that its metric of similarity is built solely around the observed properties of asteroids. Any mineralogical interpretation of the classes is secondary to their design and not the purpose of the taxonomy. This guideline may strike as hindering to conversation, yet the difficulty in designing a class based on mineralogical interpretation lies in the fact that these interpretations may be plain wrong. The classes are thus based on observed characteristics, in the hope that similar characteristics represent similar mineralogies. Class assignment does not impose or substitute a mineralogical analysis (Binzel et al. 2015).

Now that the principles are laid out, let us regard the practical side of designing and revising the asteroid taxonomy. First discrepancies with the principles arise when we seek to continue the existing class scheme while simultaneously deriving objective, data-driven classes. This is in part due to the fact that previous taxonomies such as Tholen (1984) used different observables to derive the taxonomy, for example, the albedo. Later schemes like the Bus-DeMeo system did not have this information, and classes based on the visual albedo had to be dropped. Particularly for these classes, the Tholen (1984) remains prevalent in the community because of its better compositional vocabulary. In addition, there may be classes which the community uses to a great extent in its communication, yet, which may no longer arise from the observables. Here, again, continuity and objectivity have to be weighed against each other. Weisberg et al. (2006) highlighted this for the meteorite classification, where the chondrites and achondrites follow different classification patterns, yet “historical entrenchment” (Weisberg et al. 2006) would make the adaptation of a unified system for these classes difficult.

The next issue arises when aiming to cover the maximum compositional entropy with the asteroid taxonomy. It follows that the classes should be

The core principle of the asteroid taxonomy appears to be that it is designed by a PhD student, refer to Chapman (1972), Tholen (1984), Bus (1999), DeMeo (2010), and the latter chapters of this work.

derived based on the observables of as many members of the population as possible to ensure that variations are entailed in the metric of similarity. We therefore aim to maximise the sample size upon which we base the taxonomy. At the same time, we want to use the most compositionally informative observables of our asteroids. As outlined in [Chapter 2](#), these are the reflectance spectra. However, reflectance spectra can only be acquired for the brightest asteroids. We thus have to decide whether we go for a large number of asteroids considered in the class design or the use of the compositionally most-informative features. I will highlight in several places of this work that I believe the latter to be the better decision.

Finally, there is the limitation that letters are boxes while nature is a continuum. I can describe asteroids as S or C, yet, I am unable to communicate any variation within these populations. Even the introduction of subclasses will in the end only move this problem to a smaller scale. Thus, taxonomy is a first-order estimation, and objects on the class boundaries may be insufficiently described.

In conclusion, when designing a taxonomy, we have to weigh several principles against each other. This process is subjective and may lead to disagreement within the community. A lesson I learned late during this work but still in time: if the taxonomy is a tool for the community, the community should have a say in its design. It should not be designed in isolation, even if this isolation is imposed by a pandemic.

4.2 COMPLEXES OF ASTEROIDS

I briefly introduce the main classes of current taxonomic systems, focusing on the two most commonly used, the Tholen system by Tholen (1984) and the Bus-DeMeo system by Bus and Binzel (2002a) and DeMeo et al. (2009). In lieu of a detailed review, I visualise the history of the alphabet soup at the end of the chapter in [Fig. 4.1](#) and refer to the reader to the stated references for points of interest as well as to Burbine (2014, 2016) for more extensive overviews. Meteorite links of the classes are shown in [Fig. 3.3](#).

4.2.1 *C-Complex*

The C-complex of asteroids encompasses asteroids with primitive compositions. The spectra are indicative of organics on the surface and a history of AA through the presence of phyllosilicates and magnetite. Dark surfaces are characteristic for the carbonaceous asteroids. The asteroids in the C-complex are dominating in the OMB and beyond, in particular the regions of the Hildas, Cybeles, and JT. Nevertheless, C-types may also be found in the IMB. Asteroids of the C-complex are generally linked to their chondritic namesakes, the Carbonaceous Chondrite. However, the general lack of features in the spectra make associations tentative. Vernazza et al. (2015) find good agreement between C-complex material and material which is not represented in the terrestrial meteorite collection, the IDPs.

C The C-class was among the first two classes introduced to the taxonomy by Zellner (1973) and Chapman et al. (1975). They have feature-less spectra with a slightly concave shape in the NIR. They make up the most mass in the Main Belt, in particular due to (1) *Ceres* being a C-type asteroid (DeMeo and Carry 2013). Johnson and Fanale (1973) and Cloutis et al. (2011) link CI chondrites to C-class asteroids. Jenniskens et al. (2009) compares them to ureilites using the recovered meteorite of asteroid 2008 TC₃, observed within approximately one day before impact. Hiroi et al. (1996) adds CR chondrites to the list of potential analogue material.

B The B-types were introduced by Tholen (1984), separated from the C-types by generally larger visual albedos (“bright”-C (Tholen 1984)) and a blue slope longwards of 0.7 μm . The common mnemonic used today is “blue”. There are classes which are defined by the UV behaviour, F and G, which have largely disappeared due to the sparsity of the UV data. Nevertheless, the *Gaia* DR₃ and observational efforts presented in (Tatsumi et al. 2022) may offer to reintroduce this distinction, in particular in combination with polarimetric data and albedo information.

Ch The Ch-class differs from the C-class in the presence of the 0.7 μm -feature indicating phyllosilicates (Vilas and Gaffey 1989; Bus and Binzel 2002a). Ch are linked tentatively to CM chondrites in different works (Burbine 1998; Rivkin 2012), while I see the most convincing links appear in Vernazza et al. (2016).

D The D-types were introduced in particular for the compositionally unique and homogeneous Jovian Trojan population in Zellner and Bowell (1977) as T-types but later renamed to RD for “reddish dark” (Degewij and Houten 1979), where the leading letter was dropped in Gradie and Tedesco (1982) likely to fit the at the time single-letter scheme. The ungrouped C2 chondrite *Tagish Lake* has been suggested as meteorite match for D-types (Hiroi et al. 2001). Vernazza et al. (2015) links the population to chondritic IDPs.

4.2.2 S-Complex

The siliceous S-complex members are spectrally dominated by assemblages of olivine and pyroxene and most prevalent in the IMB. Several large families exist, some of them close to resonances and suspected to deliver asteroid material to the near-Earth space and to the meteorite population, predominantly OCs and select groups of achondrites.

S The S-class was among the first two classes of asteroids to be distinguished in Chapman et al. (1975), together with the C-class. The siliceous composition is readily apparent in the spectra, which are generally dominated by pyroxene with varying contributions of olivine (Gaffey et al. 1993a). The largest asteroid families contain mostly S-types, among them the (8) *Flora* family in the IMB and the (1272) *Gefion* family in MMB (Nesvorný 2015)

which are both close to MMRs and thus suspects for the delivery of meteorites to Earth. The JAXA *Hayabusa* mission confirmed that (25143) *Itokawa* is an LL chondrite (Nakamura et al. 2011) as previously hypothesised by Binzel et al. (2004b). An interesting conundrum is that most NEOs look mineralogically like LL chondrites, yet they are by far the fewest OCs to arrive on Earth (Binzel et al. 2015). Vernazza et al. (2008) propose preferential delivery of smaller asteroids from the IMB and, in particular, the *Flora* family at the origin of this bias based on compositional and dynamical arguments. Gaffey et al. (1993a) link S-types to all groups of OCs as well as to winonanite, mesosiderites, and ureilites. Nesvorný et al. (2009) propose (1272) *Gefion* as parent body of L chondrites, while Gaffey and Gilbert (1998) links (6) *Hebe* to H chondrites and IIE iron meteorites, though the link to the ordinary chondrites is rejected by Marsset et al. (2017) based on a shape model of the asteroid, which does not show sufficient depressions in comparison to the volume of H chondrites.

Q The Q-types were introduced in the Tholen (1984) for the NEO (1862) *Apollo*. The reflectance spectra of Q-types resemble S-types in terms of features yet they are much bluer, similar to OCs. Binzel et al. (2004c) thus link Q-types to H, L, LL chondrites. Q-types are particularly prominent among the NEOs, and Binzel et al. (2004c) relate the blue appearance to a surface refreshment induced by the frequent planetary encounters in the near-Earth space. Following this hypothesis, S-types are weathered Q-types.

O The O-class was introduced in Binzel et al. (1993) for the suspected “ordinary-chondritic” (3628) *Boznemcova*. It is spectrally similar to Q- and S-types, except for bowl-shaped 1 μm -feature of the asteroid (DeMeo et al. 2009). Only one more asteroid, (7472) *Kumakiri*, has since been observed with a similar band appearance.

R The R-class was introduced early for the unusual asteroid (349) *Dembowska* (Bowell et al. 1978), which had the reddest UBV colours in the cited observational campaign. The A-class introduced below was later split off the R-class, and while the former has been growing steadily, the R-class remains basically a single-object class.

4.2.3 X-Complex

With the X-complex, Tholen (1984) unified the previously introduced M- and E-types (metallic and enstatite, Zellner and Gradie (1976)) and the pseudo-M alias P-types, introduced in Zellner et al. (1981) as X-types and rebranded in Gradie and Tedesco (1982). In the observables of the ECAS survey, all three types are spectrally degenerate, yet the distinct nature was clear as the albedos were vastly different. While the taxonomy in Tholen (1984) still employed the visual albedo for class distinction, the Bus-DeMeo system does not utilise it. Instead, the X-complex was divided into subclasses based on the behaviour in the NIR of the objects, e. g. the presence of features or the slope.

Members of the X-complex are found throughout the Main Belt and are linked to various chondritic and achondritic meteorite groups, as further outlined below.

E E-types were introduced early in the evolution of the asteroid taxonomy (Zellner and Gradie 1976). The authors had acquired polarimetric data and found that both (44) *Nysa* and (64) *Angelina* stood out in terms of their polarimetric albedo. The class designation is chosen as the authors suspected to have found the match of the enstatite achondrites, i. e. the aubrites. The large albedo of the E-types remains their hallmark feature. E are predominantly found among the Hungarian population and family. Zellner (1975) suggest aubrites as match for (44) *Nysa*, the only known E-type at the time. E-type (3103) *Eger* was linked directly to aubrite achondrites by Gaffey et al. (1992) due to its near-Earth orbit and the overall rarity of E-types.

M M-types were introduced together with the E-types in Zellner and Gradie (1976). The population around (16) *Psyche* and (21) *Lutetia* was then thought to represent the exposed metallic cores of disrupted planetesimals and was linked to iron meteorites and enstatite chondrites (Zellner and Gradie 1976). While this interpretation may still hold true for (16) *Psyche*, the interpretation of the nature of (21) *Lutetia* and other M-types has become less certain. In the generally linear, red spectra of M-types, silicate features around 0.9 μm and 1.9 μm have been observed (Shepard et al. 2010; Neeley et al. 2014). Density estimates of M-types further require large values of porosities in the asteroids to be commensurable with a pure metal nature (Carry 2012). While families of M-types were long thought to be missing entirely, Brož et al. (2022) recently identified a family around (22) *Kalliope*. Spectrally, some M-type asteroids match the iron meteorites (Cloutis et al. 1990) while others are linked to EH chondrites by Chapman and Salisbury (1973) and Shepard et al. (2010). Gaffey and McCord (1979) and Shepard et al. (2010) further link M-types to EL chondrites, and Shepard et al. (2010) adds mesosiderites, CB and CH chondrites to the list of analogue candidates.

(16) *Psyche* is the target of NASA mission carrying the same name, expected to launch in 2023 or 2024.

P The class of P-type asteroids was introduced in Gradie and Tedesco (1982) for asteroids which are spectrally similar to M-types yet have visual albedos below 7%. The “pseudo-M” (Gradie and Tedesco 1982) have red and featureless spectra indicative of complex organics on the surface. They are common among Cybele and Hilda asteroids. As far as I can tell, no meteoritic analogue material has been proposed, which may be due to the fact that P-types are placed in the X-complex, distinct from their most likely compositional companions, the C-complex asteroids. Vernazza et al. (2015) associate the class to chondritic IDPs, highlighting the affinity to the C-complex.

4.2.4 Endmembers

These classes do not fit quite into the three complexes and are often denoted as “endmembers”.

V (4) *Vesta* was the first asteroid to be observed spectroscopically (McCord et al. 1970). For years, it was unique in terms of its spectral appearance, which shows strong features of pyroxene. The first time the V-class was defined was in the Tholen-system, and it has been a constant ever since. A family of “Vestoids” has been discovered by Binzel and Xu (1993) and is the largest family of asteroids following Nesvorný (2015). Its members bridge the distance between (4) *Vesta* and the 3:1 resonance in the Main Belt. These suspected remnants of large-scale impacts on (4) *Vesta* make up the HED clan of achondrites (McCord et al. 1970; Mittlefehldt 2015).

A A-type asteroids are easily recognised by the strong olivine features around 1 μm and an extremely red VisNIR spectrum (DeMeo et al. 2019), likely due to the effect of space weathering on olivine (Chrbolková et al. 2021). A-types are assumed to be the mantle-material of disrupted parent bodies due to their strong olivine nature (e. g. DeMeo et al. 2015). However, given that the mantle of a differentiated body should make up a considerable fraction of it, only few objects are classified as A-types (DeMeo et al. 2009; DeMeo and Carry 2013), and no A-type family is known. Burbine et al. (1996) suggest that olivine-dominated objects in the break-up of the parent bodies were “battered to bits”. Links to A-types have been proposed for brachinites (Cruikshank and Hartmann 1984; Sunshine et al. 2007), pallasites (Cruikshank and Hartmann 1984), and to rumuruti chondrites (Sunshine et al. 2007). Rivkin et al. (2007) further linked (5261) *Eureka*, a Martian Trojan with a spectral appearance between a S- and a A-type, to aegirites.

K The K-class was introduced independently by Tedesco et al. (1989) and Tholen and Barucci (1989) based on observations of Bell (1988) of members of the (221) *Eos* family. K-type asteroids have a strong forsteritic olivine imprint at 1 μm yet are overall much more neutral to even blue-sloped (Mothé-Diniz et al. 2008), making them appear intermediate to the C- and S-complex. Most members of the K-class are associated to the (221) *Eos* family and to CV, CO (Bell et al. 1988), and CK chondrites (Clark et al. 2009; Cloutis et al. 2012a).

L The definition of L-types first arose from observations of (387) *Aquitania* and (980) *Anacostia* as part of the 52-color asteroid survey (Bell et al. 1988). Burbine et al. (1992) noted that the two asteroids were anomalous in that they do not present a 1 μm - but only a 2 μm -band. Sunshine et al. (2008) attributed this band to aluminous spinel found in CAIs. L-type asteroids are thus associated to CV and CO chondrites, which have the largest volume percentage of CAIs among the meteorites (Sunshine et al. 2008). Further noteworthy about L-types is the high fraction of members which depict the largest inversion angles of the negative polarization branch (Devogèle et al. 2018), refer to Chapter 2. Families with a large amount of L-types are the (729) *Watsonia*, (2085) *Henan*, and (1400) *Tirela* families (Mothé-Diniz and Nesvorný 2008; Brož et al. 2013; Cellino et al. 2014). Apart from the taxonomic analysis in the later chapters, K- and L-types are one focus of Chapter 14.

(4)Vesta was the fourth “planet” discovered within 6 years in 1807. Its discoverer, German astronomer H. Olbers, had previously discovered (2)Pallas, which may be the reason why he chose to title the publication to announce the discovery of (4)Vesta simply “Yet another planet.” (Olbers 1807).

The letter K was chosen as it lies in the middle between C and S in the alphabet.

Bus (1999) proposed the letter L to “stress the apparent spectral continuum” between the members of the K and L types. Prominent members of either class, such as (42)Isis, (599)Luisa, and (397)Vienna, have switched class assignments at some point.

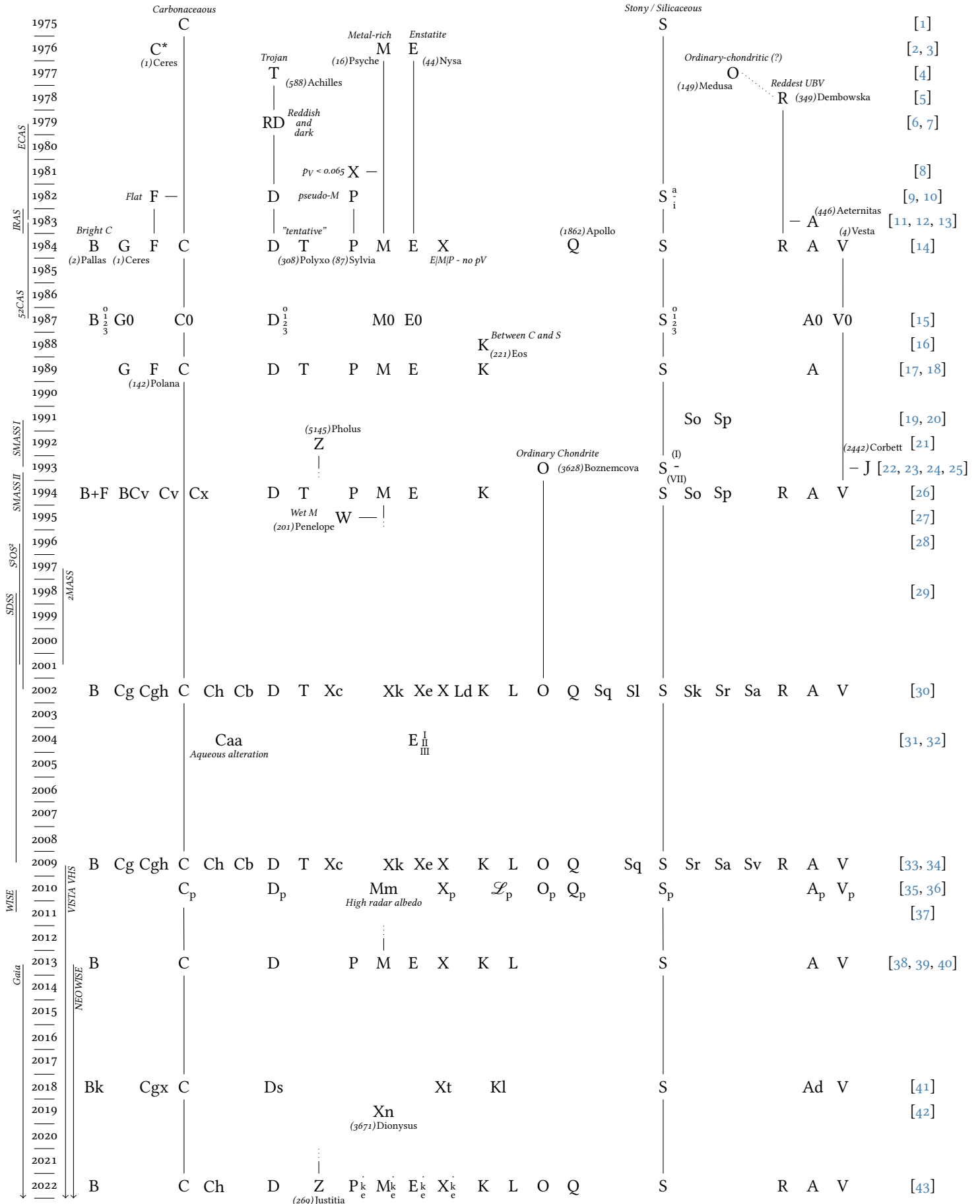


Figure 4.1: Bobrovnikoff’s Alphabet Soup. The evolution of the letter designations of minor planets over time is shown. The key to the bibliographic references is given in Table 4.1. After Tholen and Barucci (1989).

Table 4.1: References of the taxonomic classes and surveys shown in Fig. 4.1.

[1] Chapman et al. (1975)	[16] Bell (1988)	[31] Lazzaro et al. (2004)
[2] Zellner and Gradie (1976)	[17] Tedesco et al. (1989)	[32] Gaffey and Kelley (2004)
[3] Chapman (1976)	[18] Tholen and Barucci (1989)	[33] DeMeo et al. (2009)
[4] Zellner and Bowell (1977)	[19] Howell et al. (1991)	[34] Popescu et al. (2016)
[5] Bowell et al. (1978)	[20] Xu et al. (1995)	[35] Carvano et al. (2010)
[6] Degewij and Houten (1979)	[21] Mueller et al. (1992)	[36] Shepard et al. (2010)
[7] Zellner et al. (1985)	[22] Binzel and Xu (1993)	[37] Masiero et al. (2011)
[8] Zellner et al. (1981)	[23] Binzel et al. (1993)	[38] DeMeo and Carry (2013)
[9] Gradie and Tedesco (1982)	[24] Gaffey et al. (1993a)	[39] Tanga et al. (2008)
[10] Feierberg et al. (1982)	[25] Bus and Binzel (2002a)	[40] Masiero et al. (2017)
[11] Matson et al. (1986)	[26] Howell et al. (1994)	[41] Popescu et al. (2018a)
[12] Veeder et al. (1983)	[27] Rivkin (1995)	[42] Binzel et al. (2019)
[13] Bell et al. (1988)	[28] Lazzaro et al. (2007)	[43] Mahlke et al. (2022)
[14] Tholen (1984)	[29] Sergeev and Carry (2021)	
[15] Barucci et al. (1987)	[30] Bus and Binzel (2002a)	

Part II

Advancing the Method

5.1	Terminology	49	
5.2	Dimensionality Reduction	50	
5.2.1	Dichotomy of Dimensionality	51	
5.2.2	Principal Component Analysis	52	
5.2.3	Interpretation of Principal Space	54	
5.2.4	Feature Weighting	54	
5.3	Clustering	56	
5.3.1	Manual	57	
5.3.2	Finite Mixture Models	57	
5.4	Dimensionality Reduction and Clustering with Missing Data	58	
5.4.1	Imputation	59	
5.4.2	Latent Generative Variable Models	59	
5.4.3	Global and Local Dimensionality Reduction	61	
5.4.4	Mixture of Common Factor Analysers	63	
5.4.5	Model Availability	63	

The derivation of a taxonomy falls into the realm of unsupervised learning. We seek to identify patterns and populations in our data using an algorithm which does not rely on a priori information such as labels (Bishop 2006). The two primary steps are *dimensionality reduction* and *clustering*.

This chapter focuses on the mathematical aspect of deriving the asteroid taxonomy. Since the work by Tholen (1984), the major taxonomies have used the same methodology to identify the asteroid classes, which I outline in the following. I then demonstrate how we can evolve the approach overall to accomplish three improvements, namely

- I. the ability to classify observations with missing data,
- II. the computation of probabilistic classifications,
- III. and the combination of different observables in the same classification scheme.

The motivation for these improvements is given in [Chapter 8](#). I make references to the types data encountered in this work to facilitate the understanding of the mathematical concepts.

5.1 TERMINOLOGY

Before diving into the mathematical concepts, I introduce some of the terminology used in the following while building a simulated dataset used to illustrate concepts throughout this chapter.

The dataset consists of N observations \mathbf{X} of a p -dimensional variable, hence $\mathbf{X} \in \mathbb{R}^p$. We refer to \mathbb{R}^p as the *data-space*. The p dimensions might be spanned

The common terms for observations and observables in the machine learning literature are samples and features.

Random covariance matrices can be generated in python using the `make_spd_matrix` function of the `sklearn.datasets` module.

by different observables. For example, the albedo and the reflectance spectra of asteroids are two observables, where the former contributes one dimension of the p dimensions while the latter contributes as many dimensions as there are wavelength bins.

The N observations are drawn from multivariate Gaussian distributions. I infuse some structure into the artificial dataset by building it from four components. These components are the populations (e. g. taxonomic classes, exoplanet populations) which I later want to identify and describe in the analysis. Each component has a different number of samples, mean, and standard deviation, yet, they are all drawn from Gaussian distributions.

The simulated dataset has a total of 1000 samples, where the i th component \mathbf{X}_i has $i \times 100$ samples. I arbitrarily set p to 17 and draw the mean $\boldsymbol{\mu}_i$ value for the i th component from a uniform distribution between $[0, 5)$. The covariance matrices Σ_i are randomly generated symmetric, positive-definite $p \times p$ matrices. In summary,

$$\begin{aligned} \mathbf{X}_i &\sim \mathcal{N}(\boldsymbol{\mu}_i, \Sigma_i) & i \in \{1, 2, 3, 4\}, \\ \boldsymbol{\mu}_{i,j} &\sim \mathcal{U}[0, 5) & j \in \{1, 2, \dots, p\}, \\ \mathbf{X} &= \sum_{i=1}^4 \mathbf{X}_i. \end{aligned} \tag{5.1}$$

The high-dimensional dataset is difficult to visualise in its entirety. Instead, \mathbf{X} is shown in a subspace of the data-space in Fig. 5.1. The components \mathbf{X}_i are shown in different symbols and colours, however, at no stage do I provide this information to the algorithms in the following as it represents the unknown structure of interest.

Furthermore, note that this dataset does not include any errors. Observational uncertainties are typically not accounted for directly in machine learning algorithms. Instead, a common strategy is to bootstrap the dataset, i. e. duplicating it while replacing each value with a draw from a Gaussian distribution centred on the value with a standard deviation equal to its uncertainty. The machine learning algorithm is then applied to all duplicated incarnations of the input data and the variance in the results reflects the propagated uncertainties in the dataset (Tresp et al. 1993).

5.2 DIMENSIONALITY REDUCTION

Dimensionality reduction serves to reduce the dimensionality p of the observations to a smaller value q while retaining as much information in the data as possible. This serves several purposes, such as to overcome the *curse of dimensionality*, as we see in the next section, or to effectively visualise high-dimensional data, which is challenging as we saw in the previous section (Verbeek 2004; Bishop 2006).

The information contained in a dataset is quantified via its covariance matrix Σ . We may understand this intuitively by considering the case of two highly-correlated observables. Observing one allows to infer the value of the other with high accuracy, therefore, the subsequent observation of the second observable provides us with less information than if the observables were

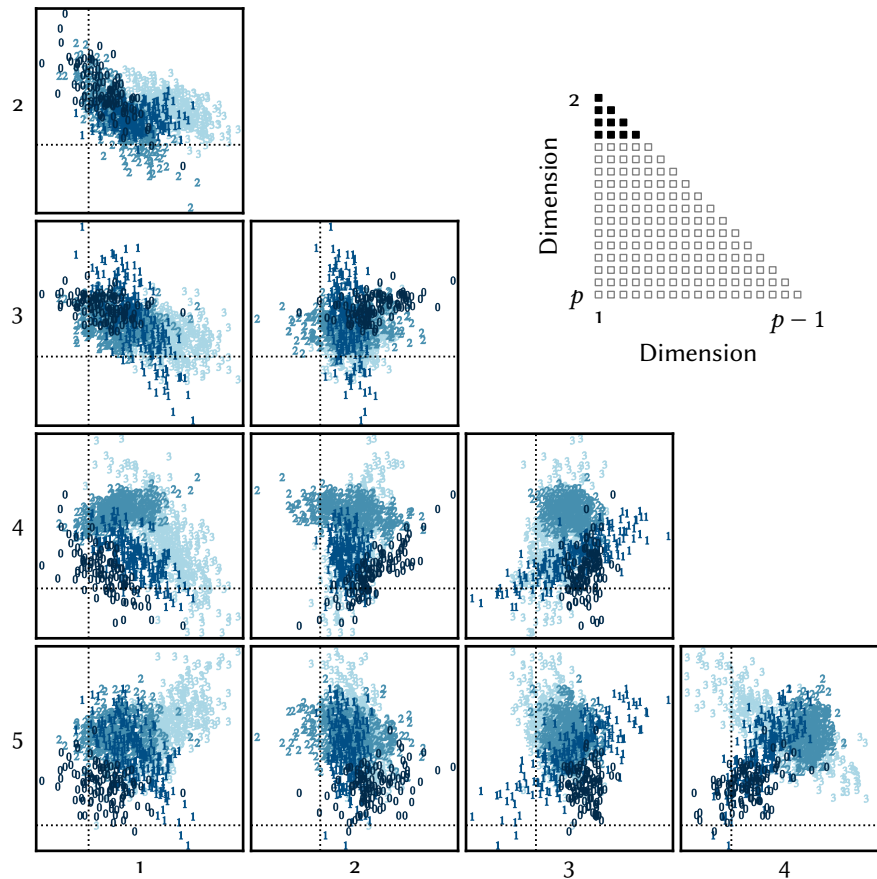


Figure 5.1: A subspace of the data-space \mathbb{R}^p containing the simulated data X . The data is split into the four simulated components X_i , differentiated by the markers and shades of blue. Dotted lines mark $x = y = 0$. Displaying high-dimensional space is commonly done using pairplots of different dimensions. The upper right graphic shows the subset of pairplots displayed here, with black squares representing dimension-pairs which are shown and grey ones those which are left out.

not correlated. An example from the world of asteroids is the relationship between reflectance spectra and colours outlined in [Chapter 2](#).

5.2.1 Dichotomy of Dimensionality

High-dimensionality is frequently seen as a nuisance in datasets. There are practical and conceptual issues that arise with a large number of dimensions which are encompassed in the literature under the term *curse of dimensionality* (Bellman 1966). For the intriguing conceptual issues, I refer to the examples given in Bishop (2006) and Bouveyron et al. (2019).

A practical consequence of high dimensionality is that the generalisation of patterns in the observations requires that they cover an appreciable fraction of the parameter space. Hence, as the number of dimensions increases, the number of required observations to generalise increases as well, and it

increases exponentially (Bishop 2006). In fields like asteroid spectroscopy, where we are far away from the realm of *big data* generally associated to machine learning applications, dimensionality reduction is thus a must-do in the analysis.

However, while Bellman (1966) speaks of dimensionality as a “curse”, Bouveyron et al. (2019) point out that there is a frequent “blessing” in high-dimensional space as well. The blessing is that high-dimensional space is mostly empty. We can thus reduce its dimensionality without loss of information. An illustrative example from the literature revolves around a set of images of handwritten digits. The images have, for example, pixel dimensions of 128×128 , spanning a 128^2 -dimensional space. If the data in this space was truly random, it would require an enormous amount of samples to describe its variability. However, because the data has an inherent structure, given by the representation of the digits of 0 to 9 in two dimensions, the variability is much smaller and the high-dimensional parameter space is mostly empty (Bishop 2006; Domingos 2012).

We may thus reduce the dimensionality of our observations under the assumption that there is inherent structure limiting its variability.

5.2.2 Principal Component Analysis

Principal Component Analysis (PCA) serves to identify the linear combinations of the data dimensions along which the projected data retains the most information. Since the work by Tholen (1984), PCA has been the standard method for dimensionality reduction when deriving asteroid taxonomies.

Early works such as Chapman et al. (1975) and Bowell et al. (1978) derived linear combinations of the observable dimensions themselves, e. g. to characterise the slope of a spectrum in the visible.

The definition of the PCA algorithm can be approached from two perspectives: the *minimum-error* formulation as given by Pearson (1901) and the *maximum-variance* one as derived by Hotelling (1933). Both approaches are illustrated in Fig. 5.2. The left hand side shows a two-dimensional dataset. It is apparent that the data dimensions are highly correlated and that we may reduce the dimensionality without significant loss of information. The minimum-error formulation of PCA searches for the principal component (PC) Φ onto which the data points are projected while minimising the projection cost, i. e. the squared distance between the data points and Φ . This is akin to the χ -square minimisation fitting routine.

The maximum-variance formulation is illustrated on the right-hand side of Fig. 5.2, where the previously two-dimensional data set is shown in the one-dimensional space. This space is referred to as *principal* or *reduced space*. The values of the projected data \mathbf{X} in this principal space are the *principal scores* \mathbf{Z} , given by

$$\Phi^T \mathbf{X} = \mathbf{Z}. \quad (5.2)$$

In the maximum-variance formulation of PCA, we construct Φ in the manner that maximises the variance $\Sigma_{\mathbf{Z}}$ of the principal scores. Following Eq. (5.2), the mean $\mu_{\mathbf{Z}}$ and variance $\Sigma_{\mathbf{Z}}$ are given by

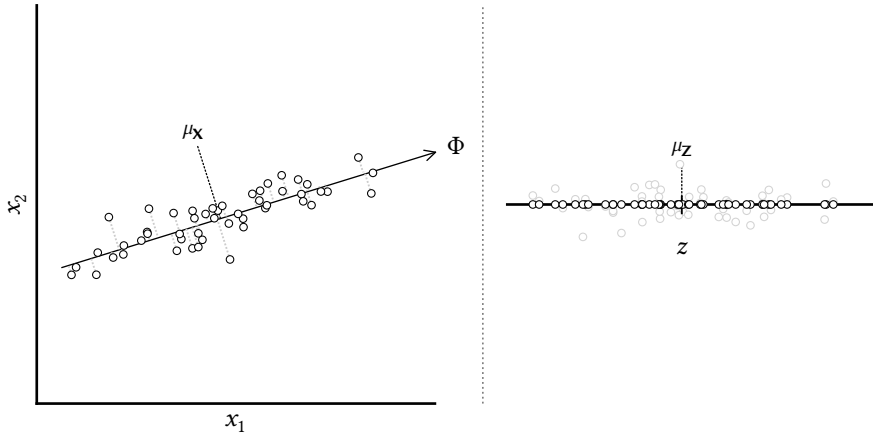


Figure 5.2: Dimensionality reduction via Principal Component Analysis. *Left:* The two-dimensional data points (circles) display about 90% of their variance along the principal component Φ . The principal space is found by minimising the projection cost (grey dotted lines) of the data onto the component. *Right:* The one-dimensional data Z (black circles) retains 90% of the variance of original data (grey circles) in the principal space. The original data is shown for comparison purposes only, the principal space is one-dimensional in this example.

$$\begin{aligned}\Phi^T \mu_X &= \mu_Z, \\ \Phi^T \Sigma_X \Phi &= \Sigma_Z.\end{aligned}\tag{5.3}$$

μ_X is commonly set to 0 by subtracting the mean from the data, giving $\mu_Z = 0$. To maximise Σ_Z , we thus search for Φ which maximises $\Phi^T \Sigma \Phi$. We constrain Φ to be a unit vector to rule out solutions where the norm of Φ increases towards infinity, i. e. $\Phi^T \Phi = 1$. Therefore, the second part of Eq. (5.3) can be rewritten using a Lagrange multiplier λ as

$$\Phi^T \Sigma \Phi + \lambda(1 - \Phi^T \Phi) = \Sigma_Z\tag{5.4}$$

We now compute the derivative of Eq. (5.4) with respect to Φ and equate it to zero, giving

$$\Sigma \Phi = \lambda \Phi\tag{5.5}$$

Readers with a keen eye for linear algebra may recognise that this is the eigendecomposition of Σ , meaning that Φ and λ are given by its eigenvectors and eigenvalues. Hence, PCA leads to a diagonalisation of the data covariance matrix Σ_X . This derivation can be extended to any number of dimensions. In the case where the principal space has more than one dimension, the PCs Φ are sorted in descending order of their corresponding eigenvalues λ , as λ represents the amount of retained variance. The elements of Φ are the *principal loadings*.

5.2.3 Interpretation of Principal Space

An important result of PCA and related dimensionality reduction algorithms are the principal components Φ and scores Z . They enable us to view and interpret the data in a compact yet informative manner. An intuitive way to understand the meaning of the components and scores is the vector product given in Eq. (5.2), as I outline in the following.

The PCs Φ are linear combinations of the data dimensions. Their loadings of the components tell us which dimensions are the most informative. The upper part of Fig. 5.3 shows the first two PCs of the simulated data X . Large absolute values of the loadings mean that the corresponding data dimension has a large impact on the principal score corresponding to the PC. Negative loadings decrease the principal scores, positive loadings increase it. Dimension 12 is thus the most informative dimension of the simulated data X as it has the largest absolute loading in the first PC.

The explained variances of the two components, i. e. the relative values of the corresponding eigenvalues, are 32.7 % and 21.1 %. Using PCA, we thus reduced the dimensionality q from 17 to 2 while retaining about 50 % of the variance. We may increase the number of dimensions of the principal space by including further PCs in the analysis. Note that the loadings of the PCs are not affected by the choice of the dimensionality of the principal space; we are selecting a subset of the eigenvectors of the diagonalised covariance matrix Σ_X .

The lower part of Fig. 5.3 shows the projection of the simulated data X into the principal space spanned by the first two PCs, i. e. the principal scores Z . Each two-dimensional data point represents one of the 17-dimensional observations in X . Note that X has been demeaned prior to the projection, hence, a score of 0 represents the projection of μ_X .

A practical example of PCA is shown in Fig. 5.4. In the upper part, the first four PCs derived from the sample of 371 VisNIR spectra used in DeMeo et al. (2009) are shown. They resemble the reflectance spectra of siliceous minerals shown in Chapter 3. Intuitively, this makes sense: the variance in reflectance spectra of asteroids is caused by the varying abundance of minerals, which imprint their features into the spectra we observe. For visual emphasis, vertical lines mark the band centres of two randomly chosen olivine (dotted) and pyroxene (dashed) minerals from Hiroi and Takeda (1992), which align well with the minima of the PCs. The lower part of Fig. 5.4 shows the principal scores of the reflectance spectra in the first two reduced dimensions. Note that DeMeo et al. (2009) subtracted the slope of the spectra prior to the PCA.

5.2.4 Feature Weighting

As variance relates to information and PCA maximises the retained variance, the variance of the observations X along its dimensions determines the result of the projection. If the variance of one dimension is much larger than that of any other, it dominates the principal scores Z . It is therefore common practice to *standardise* the variance of the data dimensions to ensure that

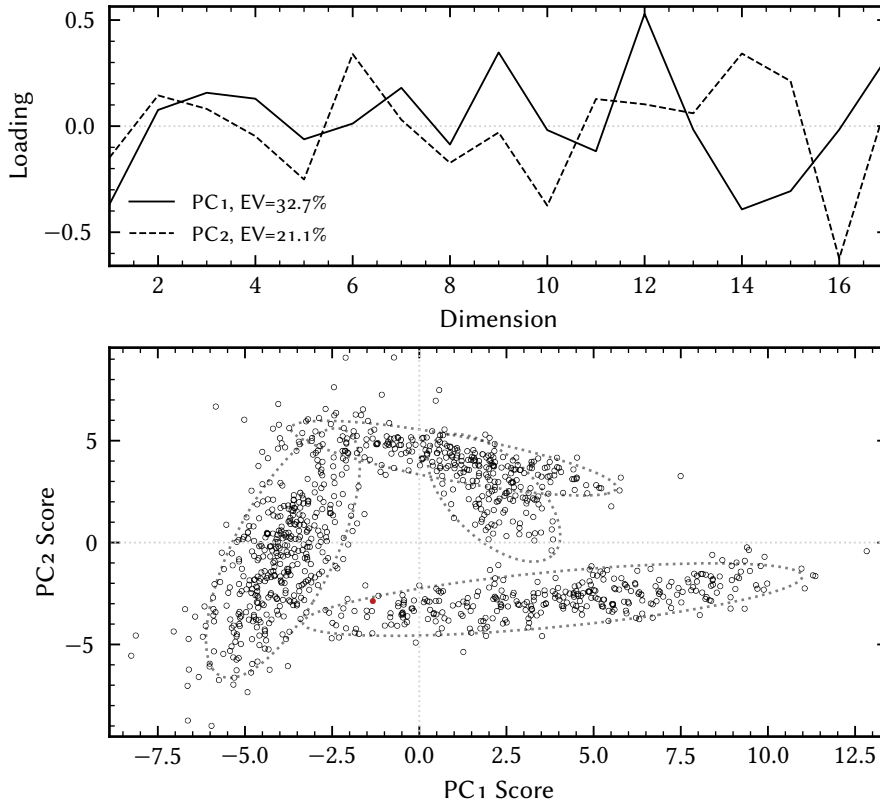


Figure 5.3: The principal components and scores of the simulated dataset. *Top*: The first two principal components derived from the data are given as solid and dashed line. The explained variance (EV) of each component is indicated in the legend. *Bottom*: The principal scores of the data projected along the first two principal components are shown as circles. A four-component Gaussian Mixture Model is fit to the scores. The covariances of the components are marked by dotted lines. One observation is not assigned to the correct population X_i by the model and marked as filled, red circle.

each dimension is equally weighted in the projection. Standardisation means to transform the values x in each column of X to zero mean and unit variance (Bishop 2006),

$$x' = \frac{x - \mu_x}{\sigma_x}. \quad (5.6)$$

This makes sense if all data dimensions are derived from the same observable, e. g. from reflectance spectra, and thus should have equal weights. However, if we add a second observable, e. g. the visual albedo, the standardisation means that the visual albedo has as much weight as a single wavelength bin of the spectra. Tholen (1984) highlights this issue and thus demotes the visual albedo to a secondary classification feature subsequent to the PCA. Nevertheless, we may include the albedo into the PCA while increasing its weight relative to the spectral bins by means of its variance. If we increase the variance of the visual albedo while standardising the spectral bins, its relative

The effect of a missing standardisation is seen in the PCs of Bus and Binzel (2002a), which are step functions around the normalisation wavelength 0.55 μm .

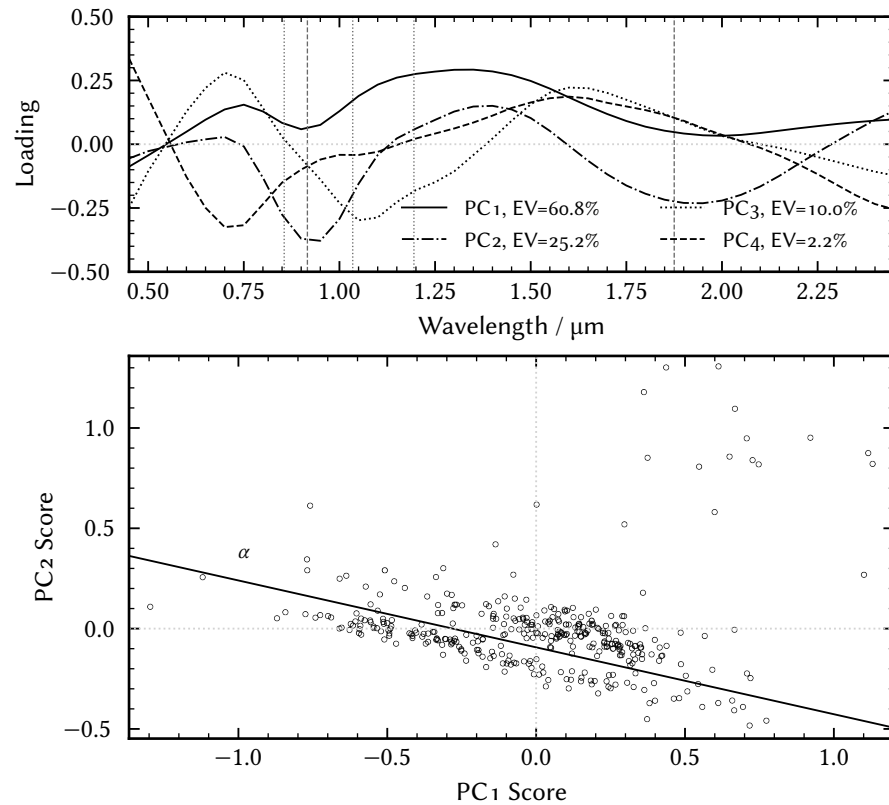


Figure 5.4: The principal components and scores from DeMeo et al. (2009). *Top*: The first four principal components derived from the slope-removed sample of 371 visible-near-infrared reflectance spectra. The explained variance (EV) of each component is given in the legend. The vertical lines indicate the band centre positions of samples of olivine (dotted) and pyroxene (specifically bronzite, dashed) minerals from Hiroi and Takeda (1992). *Bottom*: The principal scores of the 371 reflectance spectra in the first and second principal dimension. The α -line marks the “grand divide” (DeMeo et al. 2009) and is used to manually cluster the principal scores into taxonomic classes. Data from DeMeo et al. (2009).

importance is increased, and we achieve goal III of the method outlined in the introduction of this chapter. This is further explained and made use of in [Chapter 10](#).

5.3 CLUSTERING

Dimensionality reduction enables clustering of high-dimensional data in a compact, reduced space. Clustering aims to identify regions of high-density in the data. I discuss two approaches in the following, manual and model-based clustering.

The overview of clustering methods of the sklearn python package is highly recommended.

5.3.1 Manual

Manual clustering relies on our biological neural networks to identify patterns in the data. It has been the method of choice for the majority of asteroid taxonomies (e. g. Chapman et al. 1975; Zellner and Howell 1977; Tholen 1984). The upside is a high degree of freedom in the definition of clusters, i. e. taxonomic classes. The data may be carved into as many clusters as seen fit and the cluster borders can be positioned to include or exclude any observation. Asteroids with unique reflectance spectra like (349) *Dembowska* can be put into their own classes, often with the expectation that yet unobserved asteroids may show similar spectral behaviour.

The downside of manual clustering is the large degree of subjectivity. If a taxonomy should be objective and data-driven, as outlined in Chapter 4, then the number of clusters and their properties should fulfil these criteria as well.

An example of the manual clustering from DeMeo et al. (2009) is shown in the lower part of Fig. 5.4. The α -line marks the “grand divide”, the separation in the principal space between asteroids which depict a 2 μm -feature and those that do not. DeMeo et al. (2009) heuristically define this line and use parallel and orthogonal lines to divide the populations on either side into classes.

5.3.2 Finite Mixture Models

Finite mixture models are an objective approach to clustering. We aim to describe the distribution of observations using a model of their probability distribution. Specifically, I here regard models where the probability distribution is composed of a given number G of additive *mixture components* (i. e. clusters) which themselves may follow different probability distributions. For example, the components X_i of the simulated dataset follow different Gaussian distributions, refer to Eq. (5.1). The general form of these models is described as (Bouveyron et al. 2019)

$$p(x_i) = \sum_{g=1}^G \tau_g f_g(x_i|\theta_g), \quad (5.7)$$

where $p(x_i)$ is the probability distribution of the observation x_i , τ_g the probability that x_i is generated by the g th component, and $f_g(x_i|\theta_g)$ is the probability density of the g th component given its parameters θ_g . τ_g is referred to as *responsibility* and is constrained by $\tau_g > 0$ and $\sum_g \tau_g = 1$. In this work, I use multivariate Gaussian distributions $\mathcal{N}(x_i|\theta_g)$ as probability densities f_g , meaning that we work with a subclass of finite mixture models referred to as Gaussian Mixture Model (GMM) (Bouveyron et al. 2019). The parameters θ_g are the mean μ_g and standard deviation Σ_g of the Gaussian distribution. They are fit to the observations by maximising the likelihood of the observed data given the model parameters using different parameter estimation strategies, refer to Bishop (2006) and Bouveyron et al. (2019).

An example application is shown in the lower part of Fig. 5.3. A four-component GMM is fit to the principal scores of the simulated data in the

two-dimensional principal space. The structures of the covariances of the fitted components are represented by dotted lines. Each observation x_i is assigned to the component whose τ_g is maximum, refer to Eq. (5.7). As we know the different populations of the simulated data, we can visualise the success of this model. If the observation is not assigned to the correct component X_i , it is coloured in black, which is the case for one of the 1000 observations, around (-2, -2). Even though the two-dimensional principal space only retains 50 % of the variance of \mathbf{X} , we are thus able to recover the populations in the data.

The GMM could also have been fit to the simulated data without prior PCA. However, in this case, we encounter the curse of dimensionality again. The GMM model has a number N_v of free model parameters given by (Bouveyron et al. 2019)

$$N_v = G - 1 + Gp + G(p(p + 1)/2), \quad (5.8)$$

where the first term ($G - 1$) arises from the responsibility τ_g , the second (Gp) from the means and the third ($G(p(p + 1)/2)$) from the covariances of the components. Plugging in $G = 4$ and $p = 17$ yields $N_v = 683$. In the case of the simulated data with 1000 observations, we thus could fit the GMM model with four components. However, if we consider the dimensionality of VisNIR reflectance spectra of around 40-50 dimensions, we quickly get several thousands of free parameters versus a much smaller number of observations. Dimensionality reduction is thus necessary in our application for the asteroid taxonomy. The number of model components G and the dimensionality p are *hyperparameters* of finite mixture models as they are defined prior to the model application.

Using a finite mixture model, we achieve goal II of the taxonomy, as we compute the probability of each observation to belong to any class, given by the responsibility τ . The advantage of clustering using a GMM is thus the probabilistic classification and the decrease in subjectivity of the class definitions. The disadvantage, however, is the loss in flexibility, which especially complicates the definition of niche classes like O and R.

5.4 DIMENSIONALITY REDUCTION AND CLUSTERING WITH MISSING DATA

Missing data refers to gaps in the observations. Compared to a VisNIR spectrum, a visible-only spectrum has a missing NIR part. Missing data is ubiquitous, especially in astronomy, where most experiments are passive observations of light reaching Earth and factors like the target brightness or the weather may render observations incomplete.

In reflectance spectroscopy, as outlined in Chapter 2, the visible- and NIR parts are generally observed independently. It is more common to have either one or the other than to have the full spectrum, highlighted further in Chapter 9.

PCA does not allow for missing data in the analysis as the covariance matrix of a matrix with missing elements is not defined. Thus, all previous taxonomies relied on complete observations, which in turn means that only

complete observations can be classified. However, intuitively, a partial spectrum allows us to quantify the composition to a degree which is certainly larger than no classification at all. Therefore, we should derive a scheme which allows for taxonomic classification with missing data, expressed as goal I in the introduction of this chapter.

In the following, I show conceptually how we can evolve the PCA approach of the taxonomy to allow for missing data in the classification. First, however, I briefly address a common method to circumvent missing data, which is the imputation of the observations prior to the analysis.

5.4.1 *Imputation*

Imputation refers to the replacement of missing data by best-estimate values (Little and Rubin 2019). A common way is mean imputation: each missing value is filled in via the mean value of the entire dataset, the row, or the column. Other imputation methods include imputation via linear regression or multiple imputation by chained equations (Buuren and Groothuis-Oudshoorn 2011). All methods essentially try to infer the missing values based on the observed ones.

I tried different methods of imputing the reflectance spectra of asteroids, using the information from VisNIR spectra to infer on the properties of visible- or NIR-only spectra. Successful imputation would have allowed me to make use of a wide variety of dimensionality reduction and clustering algorithms, which generally require a complete dataset. However, the imputed observations were never fully satisfactory: artificial offsets or unrealistic spectral slopes were common occurrences as the imputation algorithm was not aware of the different populations of asteroids in the data. In essence, the mean of the data or any random subset of the data most probably resembles an S-type asteroid as it is the most frequent type of reflectance spectrum in the observations, hence, the imputation was biased towards this spectral shape. Instead, in the following, I use a model which only makes use of the observed parts of a spectrum, effectively ignoring the missing parts.

5.4.2 *Latent Generative Variable Models*

To allow for the presence of missing data in our analysis, we make an assumption. The assumption is that there exists a variable z in the principal space which *causes* the observations we make in the data space. To give an example, we may consider a rock made of different minerals, including an olivine and a pyroxene mineral. We can then express the composition of the rock as a two-dimensional function of the fraction of olivine and pyroxene plus the fraction of the remaining minerals, which is derived from the bounding condition that the fractions add to unity. This two-dimensional space cannot be observed directly, it is hidden or *latent*. The fractional abundances of olivine f_{ol} and pyroxene f_{px} are the *latent variables* which *generate* the observations in the data space. We then observe this rock via near-infrared reflectance spectroscopy in a high-dimensional data-space. Using the $2\ \mu\text{m}$ -band, we measure f_{px} . How can we derive f_{ol} when the visible information is missing?

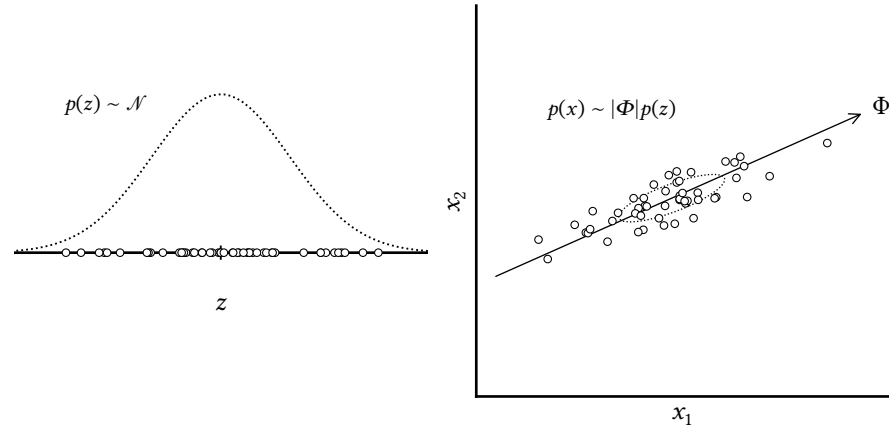


Figure 5.5: The dimensionality reduction of a two-dimensional dataset via factor analysis. *Left*: The one-dimensional latent scores Z (circles) are distributed following a Gaussian distribution (dotted line). *Right*: The two-dimensional data (circles) follows a probability distribution (dotted line) which is given by the tensor product of the latent component Φ (arrow) and the probability distribution $p(z)$ of the latent scores Z .

The way to continue the analysis with missing data is to make an assumption on the probability distribution $p(f_{ol} | f_{px}, f_{rest})$. A first assumption may be that olivine and pyroxene are the only minerals in the rock, i. e. $f_{rest} = 0$. We can then determine f_{ol} from f_{px} and, given that we know the reflectance spectra of both minerals, we can infer the visible part of the spectrum. I emphasise this condition: knowledge of the VisNIR reflectance spectra of olivine and pyroxene, either a priori or from other, complete observations in the analysis is required. Missing data cannot be imputed out of nowhere, the information needs to be present in the data.

This is, in essence, the way we account for missing data. This approach is referred to as a *latent generative variable model* (Bouveyron et al. 2019). The p -dimensional observations \mathbf{X} are modelled as high-dimensional projection of the latent scores \mathbf{Z} , which themselves are N realisations of a q -dimensional underlying latent variable z_i . A general model can be expressed by (Tipping and Bishop 1999a)

$$\mathbf{X} = \mathbf{f}(\mathbf{Z}, \Phi) + \epsilon, \quad (5.9)$$

where f is a function of the *latent scores* \mathbf{Z} and the *latent components* Φ and ϵ describes variance induced by a noise-process independent of \mathbf{Z} (e. g. observational uncertainties). Given the distributions of \mathbf{Z} and ϵ , values of \mathbf{X} can be generated. A common assumption for $p(z)$ is, once again, the Gaussian distribution $\mathcal{N}(\mu, \Sigma)$.

PCA can be formulated probabilistically by employing a latent variable model. Here, we assume that the q -dimensional \mathbf{Z} are latent variables generating \mathbf{X} and that the features of Z are uncorrelated, i. e. the covariance of Z is diagonal. In Probabilistic Principal Component Analysis (PPCA) (Tipping and Bishop 1999b), the observations \mathbf{X} are modelled as

As we are now working with a latent model, we switch the terminology from principal to latent. The interpretation of the components and scores, however, does not change.

$$\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Phi}\mathbf{Z} + \mathbf{D}, \quad (5.10)$$

where $\boldsymbol{\mu}$ is a p -dimensional vector containing the mean values of \mathbf{Y} along the data dimensions, $\boldsymbol{\Phi}$ is a $p \times q$ matrix containing the basis vectors of the latent subspace in terms of the p dimensions, and \mathbf{D} is an isotropic Gaussian noise matrix $\mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, meaning that the unique variance of each data dimension is identical. The latent variables \mathbf{Z} follow a normal distribution with zero mean and unit covariance, $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. PPCA is identical to PCA in the case where the columns of $\boldsymbol{\Phi}$ correspond to the q eigenvectors of $\Sigma_{\mathbf{X}}$ with the largest eigenvalues λ , and σ^2 is the mean variance given by the remaining $p - q$ eigenvalues which are left out. PPCA allows for missing values as the model parameters are estimated using a maximum likelihood approach.

By redefining the noise-process matrix \mathbf{D} in Equation (5.10) as diagonal rather than isotropic, $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$, we go from PPCA to Factor Analysis (FA) (Spearman 1904; Bartholomew 1984). FA is thus a more general approach than PPCA as it does not require all data dimensions (e. g. the wavelength bins of a spectrum or the visual albedo) to have the same uncertainty. A subtle consequence is that the latent components $\boldsymbol{\Phi}$ in general are no longer identical to the eigenvectors of $\Sigma_{\mathbf{X}}$. Furthermore, unlike the PCs in PCA, the latent components do not follow a hierarchy given by corresponding eigenvalues, and the latent components may differ depending on the dimensionality of the latent space. The principle of FA is shown in Fig. 5.5.

FA solves goal I of the new taxonomy by allowing for missing data. Using training algorithms like expectation-maximisation (Rubin and Thayer 1982) or stochastic gradient-descent (e. g. Bottou et al. 2018), the latent components $\boldsymbol{\Phi}$ are derived from the observed parts of an observation only.

Many of these caveats are circumvented in Chapter 10 by initialising the Factor Analysis model using PPCA.

5.4.3 Global and Local Dimensionality Reduction

We may thus replace PCA by FA and then apply a GMM in the obtained latent space to identify clusters within observations with missing data and achieve all goals we set out. However, there is one further improvement we can achieve.

Dimensionality reduction like PCA finds the rotation of the input data to maximise the variance of the whole dataset projected into the reduced space. However, this may not be the best projection to identify different populations within the data. An illustration is shown in Fig. 5.6. The column-wise variance of the entire simulated dataset (refer to Eq. (5.1)) is shown in the colour-coded upper part of the plot. The first two principal components of the dataset are shown below and reflect this variance in terms of their loadings. The first component captures the variance in the most variable dimension, while the second one describes part of the remaining variance. This is *global* dimensionality reduction. The lower part of the figure shows the dimensionality reduction for each of the components X_i of the dataset separately. We see that the variance of the populations themselves in the different data dimensions is different from the variance of the entire dataset and, therefore, each population is best projected by a different principal component, shown in the lowest part of Fig. 5.6.

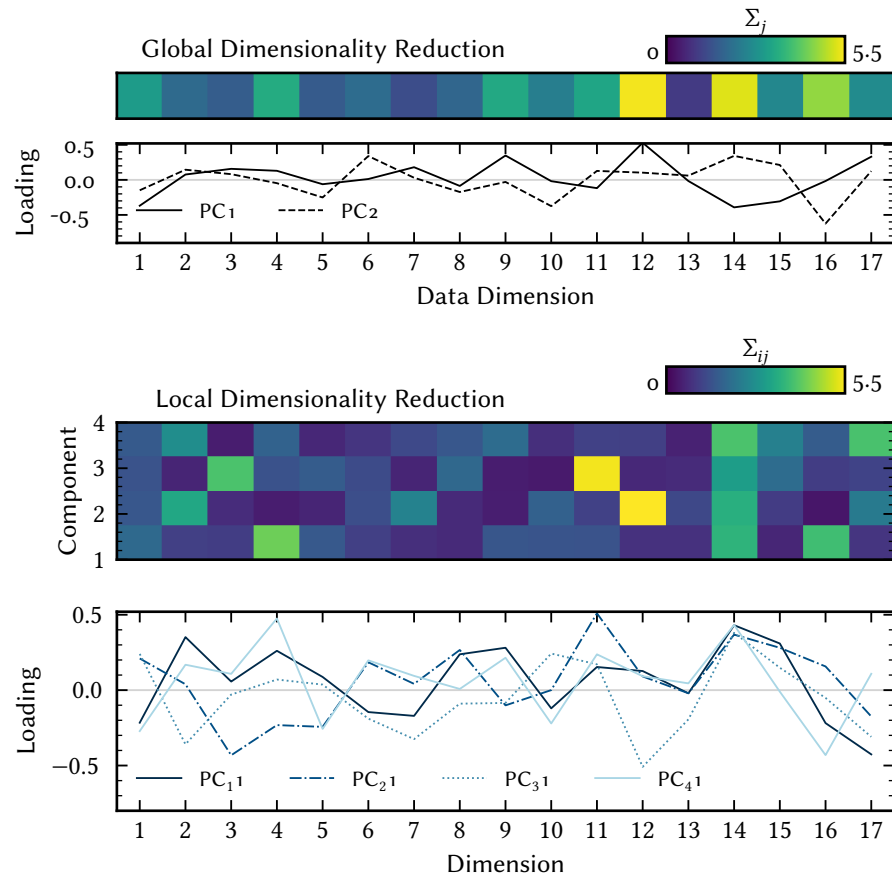


Figure 5.6: Local versus global dimensionality reduction of the simulated dataset. *Top*: Global dimensionality reduction regards the variance per observable along the entire dataset. The principal components represent this variance in their loadings: large absolute loadings represent a large variance of the dataset in the respective dimension. *Bottom*: Local dimensionality reduction identifies populations in the data and optimises the principal components for each population. The upper part shows the variance along each dimension for each of the four populations X_i of the simulated dataset.

This *local* dimensionality reduction is the actual focus, to identify the projections where each population in the data shows the largest variance. To draw an analogy: A dataset of reflectance spectra of A- and V-type asteroids will show large variability in terms of olivine and pyroxene contents. However, each class of asteroids is best represented by variability along only one of these two minerals. If we want to maximise the resolution in latent space for both populations, it is best to identify and separate during the projection. We achieve this by modelling each population separately, as outlined in the following.

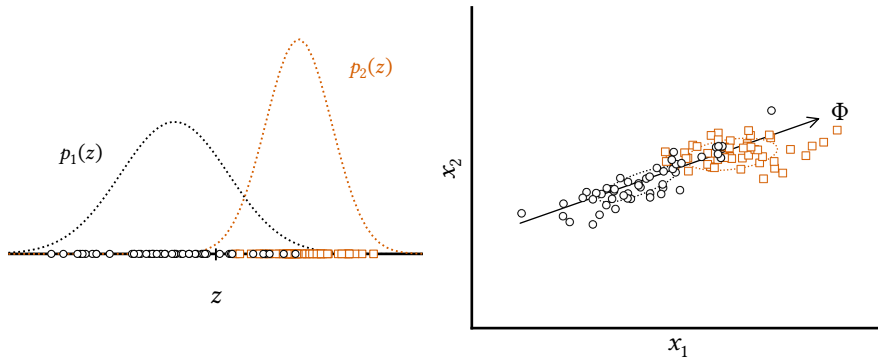


Figure 5.7: The dimensionality reduction and concurrent clustering of a two-dimensional dataset is achieved via Mixture of Common Factor Analysers. *Left:* The one-dimensional latent scores Z (circles) are distributed following two Gaussian distributions (dotted lines). *Right:* The two-dimensional data (circles and squares) follows two probability distributions (dotted lines) which are given by the tensor product of the latent component Φ (arrow) and the probability distribution $p_1(z)$ and $p_2(z)$ of the latent scores Z .

5.4.4 Mixture of Common Factor Analysers

Mixture of Common Factor Analysers (MCFA) is a combination of the latent generative data models and the finite-mixture models discussed above (Baek et al. 2010). The assumption here is that the observation is caused by one of several latent variables, all of which occupy the same latent space spanned by Φ ,

$$\begin{aligned}
 p(\mathbf{X}) &= \sum_{g=1}^G \tau_g f(\mathbf{X} | \boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g), \\
 \boldsymbol{\mu}_g &= \Phi \boldsymbol{\xi}_g, \\
 \boldsymbol{\Sigma}_g &= \Phi \boldsymbol{\Omega}_g \Phi^T + \mathbf{D},
 \end{aligned} \tag{5.11}$$

where $\boldsymbol{\xi}_g$ and $\boldsymbol{\Sigma}_g$ are the mean and covariances of the cluster components. MCFA thus achieves dimensionality reduction and clustering simultaneously. During the training, the observations are projected into the q latent dimensions and divided into G Gaussian clusters while maximising the probability of the data given the model parameters. The model is illustrated in Fig. 5.7.

5.4.5 Model Availability

There are two MCFA implementations that I am aware of, one by Baek et al. (2010) for the R programming language¹ and one by Casey et al. (2019) for

An application of MCFA in the field of stellar physics is presented in (Casey et al. 2019). The authors include an extensive discussion of the model with nice illustrations and toy models.

¹ <https://github.com/suren-rathnayake/EMMIXmfa>

the python programming language². I have tried out both implementations, however, I ultimately decided to implement one myself for python as the one from Baek is written in the R programming language, making its implementation into an analysis pipeline written in python difficult, while the one from Casey applies mean imputation to account for missing data in the dataset, which is not applicable in our case as outlined in [Sect. 5.4.1](#).

The implementation is available on Github³ and based on the tensorflow python package (Abadi et al. 2015). It uses a stochastic-gradient descent for training, as further outlined in [Chapter 10](#).

² <https://github.com/andycasey/mcfa>

³ <https://github.com/maxmahlke/mcfa>

6.1	SsODnet: A Unified Access Point for Asteroid Data	66
6.1.1	Name Resolution	67
6.1.2	All Estimates of Asteroid Parameter	68
6.1.3	Best Estimate of Asteroid Parameter	68
6.2	rocks: Quick Access to Asteroid Data	68
6.2.1	Names, Numbers, Designations	69
6.2.2	Taxonomic Diversity	70
6.2.3	Compilation of Visual Albedos	71

The amount of asteroid observations and data products is increasing exponentially. In each of the last three decades, the order of magnitude of known asteroids has increased by one (DeMeo et al. 2015). With the next-generation wide-field imaging surveys like the Legacy Survey of Space and Time (LSST) (Ivezic et al. 2019) and the J-PAS (Benitez et al. 2014) on the horizon, this mark might be reached for the fourth decade in a row. The catalogue of asteroid diameters and albedos based on mid-infrared observations from the IRAS satellite from 1992 contained 2228 entries (Matson et al. 1986; Tedesco et al. 2002). Since 2010, the WISE/NEOWISE surveys have produced the same parameters for over 150 000 objects (e. g. Grav et al. 2011; Mainzer et al. 2011; Masiero et al. 2011). Until June 13, 2022, as far as I am aware, I had created the largest collection of asteroid reflectance spectra in a single location. Then, the third data release of ESA’s *Gaia* provided 60 518 visible spectra to the asteroid community (Galluccio et al. 2022; Tanga et al. 2022; Vallenari et al. 2022).

This wealth in observations has clear benefits. For one, it enables me to derive an asteroid taxonomy without requiring a single observation from my side. This work is of course not the first to rely fully on archival data, which has been exploited for large-scale studies like the compositional structure of the Main Belt (DeMeo and Carry 2014), the identification of asteroid families (Masiero et al. 2013), or the effect of space weathering on different taxonomic complexes (Thomas et al. 2012, 2021).

At the same time, the riches of the archives provide new challenges to researchers in order to fully exploit the potential. The visual inspection of each single measurement is no longer feasible and statistical treatment is in the focus, such as I presented in Chapter 5. Furthermore, tools are required which facilitate the quick exploration of and access to the data in the archives and which tie in with the subsequent analysis to enable the automated tracking of the source of each data point for full reproducibility.

This thesis project was intensive on the data acquisition and processing side. In the following, I focus on some practical aspects of the data analysis. The chapter begins with the presentation of the unified access point SsODNet for asteroid data developed in the framework of the Virtual Observatory at the Institut de mécanique céleste et de calcul des éphémérides (IMCCE) in Paris to

which I contributed first as alpha-tester and later as developer. I then present the access point `rocks` which I developed during the thesis, highlighting how it serves to solve common and challenging tasks I encountered.

6.1 SSODNET: A UNIFIED ACCESS POINT FOR ASTEROID DATA

The collection and aggregation of asteroid data can be a tedious task. In general, various online databases have to be sourced for different parameters such as the proper elements from *AstDyS*¹ and colours from SDSS (Ivezić et al. 2002; Sergeyev and Carry 2021) or VISTA (Popescu et al. 2016), and each database has its own conventions on the data format, units, or asteroid designations. Online data aggregation platforms like NASA’s Planetary Data System (PDS)² and Centre de Données astronomiques de Strasbourg (CDS)³ facilitate this by representing single access points with uniform conventions, however, not all data is on these platforms, including a considerable amount which is only tabulated in articles.

When analysing data, the acquisition step should take as little time as possible so we can move on to the relevant parts. From this premise, we can derive the need for a homogeneous, unified database of asteroid data. Several efforts towards this goal are on-going, such as the Jet Propulsion Laboratory Small Bodies Database⁴, the Lowell Observatory Minor Planet Services⁵, or the Minor Planet Physical Properties Catalogue⁶ of the Observatoire de la Côte d’Azur. These services offer access to the data via their web interfaces, including the possibility to query and visualise data. However, while these interfaces offer nice ways to present data, they are inefficient when it comes to retrieving a lot of data repeatedly with different selection criteria. Instead, I prefer sending data queries programmatically. When I started the work on the taxonomy in 2019, neither of the services offered a programmatic access via HTTP requests.

Therefore, I first started using the IMCCE’s Solar system Open Database Network (SsODNet) in its early development stage and later joined the development efforts by contributing a python client for the service which I called `rocks`. SsODNet⁷ is a web service provided by the IMCCE in Paris (Berthier et al (incl. Mahlke), in prep.), who also operate the asteroid search and identification service SkyBoT⁸ and the ephemeris generator Miriade⁹ (Berthier et al. 2006, 2008). At its core, SsODNet aims to provide the best estimates of asteroid parameters, including dynamical parameters like the proper elements and family membership as well as physical parameters like the diameter, the thermal inertia, and the taxonomic class. This task is divided into four services,

Since December 2021, the Jet Propulsion Laboratory offers access to its Small Bodies Database via HTTP queries.

For full disclosure, I mention that my thesis supervisor Benoit Carry is one of the two developers of SsODNet.

¹ <https://newton.spacedys.com/astdys/>

² <https://pds.nasa.gov/>

³ <https://cdsweb.u-strasbg.fr/>

⁴ <https://ssd.jpl.nasa.gov>

⁵ <https://asteroid.lowell.edu>

⁶ <https://mp3c.oca.eu/>

⁷ <https://ssp.imcce.fr/webservices/ssodnet/>

⁸ <https://ssp.imcce.fr/webservices/skybot/>

⁹ <https://vo.imcce.fr/webservices/miriade/>

- asteroid identification via `quaero`,
- a compilation of asteroid parameter values from the literature name `dataCloud`,
- a best-estimate parameter summary for each asteroid referred to as `ssoCard`,
- and a table of best-estimate parameters for all asteroids named the `ssoBFT`.

Each service can be accessed independently from the others either via a web interface or via HTTP requests. We look at each step in more detail below.

6.1.1 Name Resolution

Before we characterise a moving target, we have to be sure of its identity so that we may look it up in the archives and communicate our results. The `quaero` service of SsODNet identifies asteroids, dwarf planets, comets, exoplanets, space probes and space debris based on a user-provided identification input. Once a week, the list of possible identifications is compiled from different sources: the Minor Planet Center (MPC) (Marsden 1980), IMCCE's CometPro Database (Rocher and Cavelier 1996), the Extrasolar Planets Encyclopaedia¹⁰, and CelesTrak¹¹, including over 5 000 000 names and designations. The identification input may be the official name of the object or one of many recognised aliases. The latter part is the crucial aspect as I describe now while focusing on the issue of identifying asteroids.

There are several ways to refer to an individual asteroid. After its discovery, which requires observations over two nights typically within a week from one another, it receives a *provisional designation* made up of the year of discovery, the month, and half week, and a number¹². This has occurred for 1 217 458 asteroids so far based on the orbital elements database¹³ of the Lowell Observatory.

It may be the case that the newly discovered and designated asteroid has been discovered and designated before, however, these observations could not be linked to each other right away due to uncertainties when computing the asteroid's orbit. It then has two or more provisional designations, of which one is selected as the official one.

Once the asteroid has been observed at several oppositions and its orbital uncertainty is sufficiently low,¹⁴ it receives an integer number in order of its discovery, starting with (1) *Ceres* and currently at (616690) 2016 YY12. A numbered asteroid may then receive a name which is proposed by its discoverer and accepted or rejected by the International Astronomical Union (IAU) *Working Group for Small Bodies Nomenclature*.

As the way to identify an individual asteroid evolves over time with the certainty of its orbital solution, it may appear under different synonyms in

A unique case is (134340) Pluto, which received a large number not reflecting its early discovery date due to the initial classification as planet.

(1723) Klemola has 39 aliases.

¹⁰ <http://www.exoplanet.eu/>

¹¹ <https://celestrek.com/>

¹² <https://minorplanetcenter.net/iau/info/Astrometry.html>

¹³ <https://asteroid.lowell.edu/main/astorb/>

¹⁴ <https://minorplanetcenter.net/iau/info/UValue.html>

database: with one or more provisional designations, its number, or its name. Therefore, it is required to ensure that we have the current official name or designation when working with archival data.

6.1.2 *All Estimates of Asteroid Parameter*

The issue of asteroid data aggregation is outlined above. The `datacloud` service of SsODNet is a compilation of published measurements and estimations of numerous asteroid parameters including their uncertainties. The data is compiled from about 3000 bibliographic sources so far, and every value in the database is linked to a complete bibliographic record. Further metadata includes, among others, the method used to derive the value (e. g. spectroscopy or photometry for taxonomic classifications) and the wavelength range of the observation when applicable. As such, the data is presented with the necessary information to select the best estimates and to reject dubious ones.

6.1.3 *Best Estimate of Asteroid Parameter*

Producing a large collection of asteroid parameters quickly gives rise to the question of how to identify or compute the most probable value. In most cases, this question should have an objective answer, and as such, it can be done on the side of SsODNet rather than on the user-side. The `ssoCard` is the collection of the best estimates of parameters for a given asteroid. The decision-trees to identify the best among many estimates are outlined in Berthier et al. (in prep.) and were revised by the experts in the corresponding measurement techniques. Typically, the decision-trees are based on the parameter observation or estimation method, which are ranked by accuracy. If several parameters of the highest-ranking available method are observed, a weighted average is computed. Some parameters, however, require unique decision-trees, such as for the choice of taxonomic classification, where the wavelength range and the used taxonomic scheme further contribute to the selection.

Finally, the BFT service is a tabulation of `ssoCards`. In its current version in July 2022, it has over 1 200 000 rows, each representing an individual asteroid, and over 560 columns, each representing a parameter. It serves to quickly select asteroids based on parameter values, e. g. all members of a given family or all asteroids within a certain size range. 18 % of the cells of the BFT are filled as most asteroids are sparsely characterised.

6.2 `rocks`: QUICK ACCESS TO ASTEROID DATA

As web service, SsODNet offers HTTP access to each of the services, returning the responses in JSON format. Therefore, to access the data in any preferred format, all that is required is a way to generate HTTP requests and a way to parse the JSON response. My format of choice is a python interface called `rocks`, which I started to develop for SsODNet in December 2019. `rocks` is a programmatic access point to the SsODNet services built in the python language. It provides both a command line interface for quick data exploration

and a python interface for scripted data analysis and retrieval. It launches asynchronous HTTP queries to the SsODNet service and stores the responses in a cache directory on the user's computer. In simpler terms, rocks is fast.

6.2.1 Names, Numbers, Designations

There are different ways that asteroids are identified in databases. Common is a numeric column with asteroid numbers and an alphabetical column with names and designations mixed. Sometimes, there is a single column with mixed types, e. g. names and provisional designations or numbers and provisional designations. Another common scheme is to provide the packed designation format used by the MPC.¹⁵

The `rocks.identify` function addresses the issue of converting the identifiers to official names or designations. It takes a single argument which may be of different types. Both numerical and alphabetical inputs are supported, either provided one-by-one or in a list of many. The function returns the asteroid name and number for each identifier it is passed.

```
>>> import rocks
>>> rocks.identify("bennu")
("Bennu", 101955)
>>> rocks.identify(3)
("Juno", 3)
>>> rocks.identify(["cristinathomas", 7561, "1999rr219", "
    Vernazza", "1981 EE45", "J92G05P", "rousselet"])
[("Cristinathomas", 8063),
 ("Patrickmichel", 7561),
 ("Lantz", 11799),
 ("Vernazza", 20607),
 ("Libourel", 10281),
 ("Mariacristina", 17899),
 (None, nan)]
```

The >>> syntax indicates that this code is to be run in an interactive python session. The lines without preceding >>> are results from the code execution.

If, for some reason, credit was not given where credit is due and the identifier is not recognised, the returned name is a None value. For unnumbered asteroids, a nan ("not-a-number") value is returned in place of the number. Note the use of the packed designation format and outdated provisional designations with different formatting.

The command line version of `rocks.identify` is the `id` command. I find it most convenient when reading articles and stumbling over numbers which I cannot relate to a name or vice versa.

```
$ rocks id 8467
(8467) Benoitcarry
```

I can also request all recognised aliases of an object.

```
$ rocks aliases 1723
(1723) Klemola, aka
['1913 UF', '1915 CD', '1916 KB', '1929 WE1', '1931 FB',
 '1936 DJ', '1936 FX', '1941 BF', '1942 HK', '1947 GG',
 '1949 UZ', '1955 XT', '1957 DE', '1957 DF', '1958 LB',
 '1959 PC', '1964 PF', '1965 TC', '1968 KV', '2001723',
 'J13U00F', 'J15C00D', 'J16K00B', 'J29W01E', 'J31F00B',
```

The \$ syntax indicates that this code is to be run in a shell session. The lines without preceding \$ are results from the code execution.

¹⁵ <https://minorplanetcenter.net/iau/info/PackedDes.html>

```
'J36D00J', 'J36F00X', 'J41B00F', 'J42H00K', 'J47G00G',
'J49U00Z', 'J55X00T', 'J57D00E', 'J57D00F', 'J58L00B',
'J59P00C', 'J64P00F', 'J65T00C', 'J68K00V']
```

Name resolution is typically the first step in any of my analyses. I therefore invested time in increasing the resolution speed. A local index file stored on the user's computer handles the resolution of names, numbers, and provisional designations. Only if the local resolution was unsuccessful does rocks query the quaero backend. In practice, I find that most look-ups can be resolved locally. Name resolution with rocks is currently restricted to asteroids and dwarf planets as they were the focus of this project.

I conclude with a warning about the intuitive idea about using asteroid names as filenames when storing analysis results. Asteroid names are not restricted to characters of the Latin alphabet, which may lead to unexpected results especially in cases like the asteroid below, which includes several special shell characters.

```
$ rocks id 229762
(229762) G!kun||'hondima
```

Instead, either the number or the provisional designation should be used. For the latter, the space between the year and the letter-number block should be removed prior to saving.

6.2.2 Taxonomic Diversity

Unsurprisingly, I found myself frequently looking up taxonomic classification of asteroids in the past years. The current best-estimate classification is stored in the ssoCard of SsODNet. Using rocks, any parameter of any asteroid is only one command line query away.

```
$ rocks taxonomy 70
Ch
```

More informative, however, are the collections of classifications stored in the datacloud. By specifying that we are interested in the taxonomies rather than the taxonomy, we get an overview of the classifications of individual asteroids.

```
$ rocks taxonomies luisa
+-----+-----+-----+-----+-----+
| class | method | range | scheme | shortbib |
+-----+-----+-----+-----+-----+
| S      | Phot   | VIS   | Tholen | Tholen+1989 |
| T      | Spec   | VIS   | Bus    | Binzel+2001 |
| K      | Spec   | VIS   | Bus    | Bus&Binzel+2002 |
| L      | Spec   | VISNIR | Bus-DeMeo | Clark+2009 |
| L      | Spec   | VISNIR | Bus-DeMeo | DeMeo+2009 |
| L      | Spec   | NIR   | Bus-DeMeo | Devogèle+2018 |
| L      | Phot   | VIS   | Bus-DeMeo | Sergeyev&Carry2021 |
+-----+-----+-----+-----+-----+
```

We see that (599) *Luisa* has a taxonomic history typical of L-types: start off as anomalous S, classified as K based on a visible spectrum until the 2 μm -band information in the NIR led to a classification as L.

rocks offers a convenient access to all parameters stored in the ssoCard of an asteroid via the Rock class. Once again, the notation is simple yet powerful: parameters are accessed via the dot-notation typical for python syntax, giving a code which is easy to understand.

```
# import required packages
from collections import Counter
import rocks

taxonomies = []

# for each of the first 1000 asteroids
for number in range(1, 1001):

    # get the ssoCard
    asteroid = rocks.Rock(number)
    # get this asteroid's taxonomic class
    taxonomic_class = asteroid.taxonomy.class_
    # store the taxonomic class in a list
    taxonomies.append(taxonomic_class)

# finally, print the distribution of the taxonomic classes
print(Counter(taxonomies))
```

Listing 6.1: Taxonomic survey of first 1000 asteroids with rocks.

The python script in Listing 6.1 prints the following result when executed, highlighting the observational bias for S-types and against C-types:

```
Counter({'S': 233, 'X': 147, 'C': 123, 'Ch': 106, 'Xk': 43, 'Cb': 31, 'B': 29, 'L': 29, 'Sl': 27, 'Xc': 26, 'T': 25, 'D': 23, 'K': 23, 'Sw': 18, 'Xe': 14, '': 13, 'Cgh': 11, 'A': 9, 'M': 9, 'Sq': 8, 'Ld': 7, 'Sk': 7, 'Cg': 5, 'V': 4, 'P': 4, 'Sr': 3, 'Kl': 3, 'Sqw': 2, 'Sa': 2, 'XC': 2, 'STD': 1, 'XFC': 1, 'O': 1, 'BU': 1, 'ST': 1, 'R': 1, 'DCX': 1, 'F': 1, 'Cgx': 1, 'PD': 1, 'TD': 1, 'SD': 1, 'CB': 1, 'Ds': 1})
```

Data retrieved from either the ssoCard or the datacloud service is cached locally on the user's computer to increase the speed of repeated queries.

6.2.3 Compilation of Visual Albedos

The visual albedo plays a major role in this taxonomy. However, in terms of data processing, the required effort to compile the collection of albedos was minimal thanks to the accessibility of SsODNet in combination with the scripting opportunities offered by rocks.

As outlined further in Chapter 9, I created a ranked list of visual albedos for each asteroid. The best estimate of the albedo is derived from the best estimates of the absolute magnitude H and the diameter D using the relation given in Eq. (2.4). The second best was the weighted average of the highest ranked entries in the datacloud. The remaining estimates are the datacloud entries in order of ranking.

Using rocks, this otherwise significant endeavour of data acquisition, merging, and processing is achieved with a few lines of python.

Note that this is not the most efficient way to query the data of many asteroids as it does not make use of the asynchronous queries rocks offers. The rocks.rocks function should be used here.

The visual albedo given in the ssoCard is computed based on H and D except if in-situ space measurements are available, as for (1)Ceres and (4)Vesta.

```

# import required packages
import numpy as np
import rocks

# example list of asteroids
asteroids = [2, 20, 200]

# ranked list of methods
methods = ['SPACE', 'Polarimetry', 'ADAM', 'KOALA', 'SAGE',
           'Radar', 'LC+Occ', 'LC+IM', 'LC+TPM', 'TPM',
           'TE-IM', 'TE-Occ', 'IM', 'Occ', 'IM-PSF',
           'Interferometry', 'NEATM', 'NESTM', 'STM', 'FRM']

for asteroid in asteroids:

    # get ssoCard and datacloud entries of albedos
    asteroid = rocks.Rock(asteroid, datacloud='albedos')

    # I: albedo based on H and D
    albedo_1 = asteroid.albedo.value

    # II: albedo based on weighted average of best datacloud
    entries
    albedo_2, _ = asteroid.albedos.weighted_average('albedo')

    # III - ...: albedos in datacloud ranked by method
    albedo_3_x = []

    for method in methods:
        for _, entry in asteroid.albedos.iterrows():
            if entry.method == method and entry.albedo != 0:
                albedo_3_x.append(entry.albedo)

    print(f'The best-estimates of visual albedos for ({
asteroid.number}) {asteroid.name} are')
    print(f'I: {albedo_1:.3f}')
    print(f'II: {albedo_2:.3f}')
    print(f'III - ...: {albedo_3_x}')

```

Listing 6.2: Compilation of visual albedos with rocks

which prints:

```

The best-estimates of visual albedos for (2) Pallas are
I: 0.153
II: 0.146
III - ...: [0.15, 0.145, 0.126, 0.142, 0.142, 0.117, 0.09]
The best-estimates of visual albedos for (20) Massalia are
I: 0.189
II: 0.252
III - ...: [0.283, 0.258, 0.241, 0.227, 0.187]
The best-estimates of visual albedos for (200) Dynamene are
I: 0.043
II: 0.049
III - ...: [0.04, 0.04, 0.042, 0.041, 0.054, 0.05, 0.051,
0.038, 0.04, 0.049, 0.04, 0.048, 0.043]

```

The compilation of all visual albedos required to derive the asteroid taxonomy presented in this work is outlined in [Chapter 9](#).

ASTEROID COMPOSITION FROM NON-TARGETED
 PHASE CURVES

7.1	Observations and Parameter Inference	74	
7.1.1	Observations by ATLAS	75	
7.1.2	Bayesian Parameter Inference	76	
7.2	Uncertainties	77	
7.2.1	Observational Uncertainties	77	
7.2.2	Rotational and Aspect Angle	78	
7.2.3	Effect on Phase Curve Parameters	80	
7.2.4	Coverage of Opposition Effect	80	
7.3	Compositional Information	82	
7.3.1	Absolute Magnitudes	83	
7.3.2	Phase Curve Coefficients	85	
7.3.3	Distribution of Taxonomic Complexes	86	
7.3.4	Wavelength-Dependency of Phase Curve Coefficients		87
7.3.5	Identification of Interlopers	91	
7.3.6	Distribution of G_1 , G_2 in Families	92	

In [Chapter 2](#), I outlined the compositional information that phase curves of asteroids contain using the observational campaigns by [Shevchenko et al. \(1997\)](#), [Shevchenko \(2002\)](#), [Shevchenko et al. \(2008\)](#), and [Shevchenko et al. \(2016\)](#). While reflectance spectra allow for a more detailed class assignment, phase curve parameters provide two ways to constrain the taxonomic complex of asteroids, the phase curve coefficients G_1 , G_2 (refer to [Fig. 2.4](#)) and the absolute magnitude H , which, in combination with an absolute magnitude acquired in a different photometric filter, provides an absolute colour of the asteroid.

Phase curves are calculated from the apparent magnitudes of asteroids, which are observed in great number by several wide-field imaging surveys, such as SDSS and the ESO VISTA surveys ([Ivezić et al. 2001](#); [Popescu et al. 2016](#); [Sergeyev and Carry 2021](#)). While asteroid taxonomies have already been derived based on the colour information that these surveys provide ([Carvano et al. 2010](#); [DeMeo and Carry 2013](#); [Popescu et al. 2018a](#)), we could consider exploiting the compositional information of the phase curves constructed by these surveys as well. This is in particular the case for surveys where the time between observations in different bands is larger than several minutes, rendering the computation of colours from the acquired apparent magnitudes uncertain or even useless due to the asteroid's rotation ([Warner et al. 2009](#)). In this case, absolute magnitudes are the only manner to determine asteroid colours from the data. A significant case where this limitation takes place is the upcoming LSST survey, expected to discover millions of asteroids, however, with a cadence that likely does not allow for colour computation based on apparent magnitudes ([Jones et al. 2009](#); [Ivezić et al. 2019](#)).

However, the compositional information has only been shown to be present in the observationally expensive *targeted* phase curves, i. e. observations of the apparent magnitudes of asteroids at phase angles which adequately sample the phase curve and which involve the correction for lightcurve modulations arising from the 3D shape of asteroids. For *non-targeted* phase curves, i. e. observations of the apparent magnitudes of asteroids acquired for example by wide-field imaging surveys which are not optimised to sample the phase curve sufficiently and which are not corrected for the influence of 3D shape, the compositional content is questionable. Oszkiewicz et al. (2011) investigated the dependence of non-targeted phase curve coefficients G_1 , G_2 on the taxonomic class of the target asteroid and the resulting mean distance of classes in the G_1 , G_2 space was small in comparison to the mean uncertainty on the phase coefficients.

Nevertheless, given the motivation for the new taxonomy to include as many asteroids and observables as possible, as well as some caveats of the results of Oszkiewicz et al. (2011) outlined below, I investigated the question of whether non-targeted phase curves carry sufficient compositional information to be used in taxonomic scheme. In this chapter I outline the compilation of a catalogue of phase-curve coefficients I computed based on observations by ATLAS. I further present a study on the compositional information contained in non-targeted phase curves. The method and results are published in Mahlke et al. (2021) and the text present here follow this publication.

7.1 OBSERVATIONS AND PARAMETER INFERENCE

Non-targeted phase curve parameters have been computed from observations in the MPC observations database¹ by Oszkiewicz et al. (2011). However, there are some caveats with respect to the provided catalogue² of phase curve parameters. First, the authors combined apparent magnitudes acquired in different photometric filters after converting them to V -band magnitudes using mean colour indices. As noted by the authors, this introduces systematic photometric errors as the correct photometric conversion between bands depends on the asteroids' reflectance spectra. In addition, the spectral reddening with increasing phase angle is non-linear (Gradie and Veverka 1986; Sanchez et al. 2012), hence, different conversion factors are required at different phase angles. The use of uniform conversion factors thus blurs the compositional information in the dataset prior to the analysis. Second, Oszkiewicz et al. (2011) did not restrict the H , G_1 , G_2 model fits to consider only physical solutions, i. e. those where asteroids get fainter with increasing phase angle. 52 % of the more than 500 000 asteroids in the catalogue have non-physical phase curve coefficients. Third, since the publication of the catalogue, the number of known asteroids has almost doubled, while the number of observations has increased likely by an order of magnitude thanks to the multiple dedicated asteroid survey efforts.

Scripted access to the MPC database is provided by the astroquery python package.

¹ https://www.minorplanetcenter.net/db_search

² <https://wiki.helsinki.fi/display/PSR/Asteroid+absolute+magnitude+and+slope>

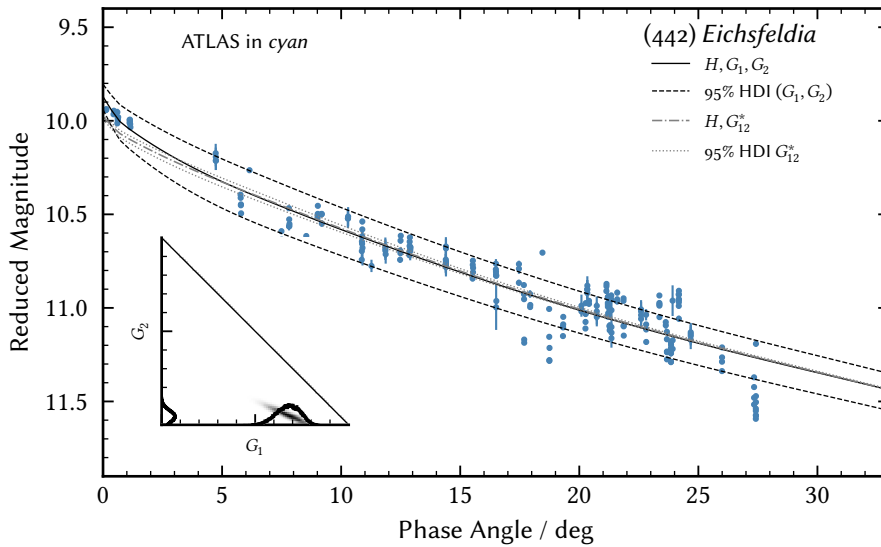


Figure 7.1: As Fig. 2.3, showing the phase curve of (442) *Eichsfeldia* (blue) observed by the ATLAS observatory. Published in Mahlke et al. (2021).

Instead, I searched the database of the MPC for the largest contributors of the past years, which include mainly US-based observatories dedicated to the detection of NEOs such as the *Lincoln Near-Earth Asteroid Research* (LINEAR, Elowitz et al. (1999)), the *Catalina Sky Survey* (Drake et al. 2009), the *Panoramic Survey Telescope and Rapid Response System* (Pan-STARRS, Hodapp et al. (2004)), and the ATLAS observatory (Tonry et al. 2018). I looked specifically for observatories which (1) reported a large amount of observations, (2) observed at small and large phase angles, and (3) observed in several bands. The second criterion is a particular challenge as observatories aiming at the discovery of NEO observe preferentially objects at quadrature, i. e. at the largest possible phase angle. I finally chose to compute a consistent set of phase curves parameters from observations of the ATLAS observatory as they fulfilled the three criteria best. For observations from Pan-STARRS, Vereš et al. (2015) have computed phase curve coefficients in the H, G system by Bowell et al. (1989).

7.1.1 Observations by ATLAS

The ATLAS³ observatory is a NASA funded facility located in Hawaii, United States. Since 2017, two independent 0.5 m telescopes survey the sky for NEOs on impact trajectory with Earth a field-of-view of 30 deg² (Tonry et al. 2018). The primary observation bands are *cyan* (420 nm–650 nm) and *orange* (560 nm–820 nm), shown in the lower part of Fig. 2.1.

Members of ATLAS provided me with observations acquired between 2015 and 2018. Using the MPC database, I added observations reported by ATLAS to the MPC in 2019, extending the catalogue to a total of 34 million observations of 180 025 individual asteroids, where 179 719 asteroids have

Between 2015 and 2017, ATLAS was observing with a single telescope.

³ <https://atlas.fallingstar.com>

been observed in both *cyan* and *orange*. For each observation, I added the ephemerides using the IMCCE's *Miriade* software.⁴ An example phase curve of (442) *Eichsfeldia* is given in Fig. 7.1.

7.1.2 Bayesian Parameter Inference

Fitting photometric phase curve models to reduced magnitudes is a challenging task, especially in the case of non-targeted phase curves (Karttunen and Bowell 1989; Kaasalainen et al. 2003). Instead of a least-squares regression approach as used e. g. by Oszkiewicz et al. (2011), I model the single-band phase curves of each asteroid with the H , G_1 , G_2 and H , G_{12}^* models introduced in Chapter 2 using Bayesian parameter inference. The use of Bayesian parameter inference in combination with Markov chain Monte Carlo (MCMC) simulations is computationally expensive but enables to inspect the posterior probability distribution of the parameters, which in turn provides detailed insights into the compositional significance of the inferred phase coefficients.

The first step is to define prior probability distributions for each inferred parameter. I use a weakly-informative normal distribution as prior for the absolute magnitudes,

$$p(H), p(H_{12}) = \mathcal{N}(\mu = 10, \sigma = 100), \quad (7.1)$$

where H is the absolute magnitude estimated with the H , G_1 , G_2 model and H_{12} the one estimated with the H , G_{12}^* model. As outlined further in Mahlke et al. (2021), using this weakly-informative prior instead of a more informative one based on the absolute magnitude distribution of objects in the Main Belt does not affect the results significantly.

The phase curve coefficients are modelled with priors of uniform probability distributions between 0 and 1,

$$p(G_1), p(G_2), p(G_{12}^*) = \mathcal{U}[0, 1], \quad (7.2)$$

following the constraints given in Eq. (2.7) in Chapter 2. This choice in the priors does not necessitate that $1 - G_1 - G_2 < 0$ but simplifies the computational implementation. To apply the missing constraint, I remove MCMC samples violating the inequality after the sampling process. Finally, I define the likelihood function that the apparent magnitudes \mathbf{m}_α observed at a phase angle α arise from a normal distribution with a mean of \mathbf{m}_α and a standard deviation of σ_{m_α} , i.e. the uncertainties of the observed apparent magnitudes

$$p(\mathbf{m}_\alpha | \Theta) = \mathcal{N}(\mu = \mathbf{m}_\alpha, \sigma = \sigma_{m_\alpha}), \quad (7.3)$$

where Θ are the inferred parameters of the photometric models.

The posterior probability distribution $p(\theta | \mathbf{m})$ is estimated using MCMC sampling implemented in *python* with the *pymc3* package (Salvatier et al. 2016) for the Bayesian parameter inference and the *sbpy* package (Mommert et al. 2019) for the implementation of the photometric phase curve models. For

⁴ <https://vo.imcce.fr/webservices/miriade/>

each phase curve, I draw 480 000 MCMC samples of the parameters' posterior distributions. The best-fit values and uncertainties of the parameters are given as the mean value and the 95 % highest density interval of the posterior distributions respectively.

7.2 UNCERTAINTIES

Before evaluating the results of the Bayesian parameter inference for the full sample of asteroid phase curves, it is worthwhile considering the sources of the uncertainties on the apparent magnitudes and how they affect the inferred parameter values.

The main sources of uncertainties of an observed apparent magnitude m at a given phase angle α can be divided into different components given by the photometric system and the 3D-shape of the asteroid, namely

$$\sigma \propto \sqrt{\sigma_{\text{PHOT}}^2 + \sigma_{\text{PREC}}^2 + \sigma_{\text{SYS}}^2 + \sigma_{\text{ROT}}^2 + \sigma_{\text{APP}}^2}. \quad (7.4)$$

Here, σ_{PHOT} is the photometric uncertainty of a given apparent magnitude, σ_{PREC} is the rounding precision error introduced when magnitudes are recorded with a limited number decimal places in a database, σ_{SYS} is the systematic uncertainty introduced by varying observational parameters (e. g. different photometric filters or standard stars), σ_{ROT} is the uncertainty introduced by the unknown rotational phase of the observed asteroid, and σ_{APP} is the uncertainty added by the unknown offset due to the current aspect angle under which the asteroid is observed. This list is non-exhaustive yet contains the dominating sources of uncertainties. Each term effects the dispersion of the reduced magnitudes per phase angle bin as shown in Fig. 7.1. In the following, I first estimate the order of magnitude for each uncertainty and then investigate their influence on the phase curve parameters.

7.2.1 Observational Uncertainties

The photometric uncertainty σ_{PHOT} of observations is caused for example by random photon noise. In the database of ATLAS observations, the mean σ_{PHOT} is 0.14 mag with a standard deviation of 0.08 mag. The uncertainty tends to be larger for fainter targets. The LSST survey aims at a mean σ_{PHOT} of 0.01 mag for targets with an apparent magnitude of 21 in the r band, two magnitudes fainter than the limiting magnitude of ATLAS (Tonry et al. 2018; Ivezić et al. 2019).

The precision uncertainty σ_{PREC} is a minor contributor to the total uncertainty given in Eq. (7.4). However, it could be completely negligible if coarse rounding was avoided. As it stands, observations reported to the MPC have to be truncated either to the 0.1 or 0.01 magnitude level. Both magnitude levels are above the actual capability of large observatories. The MPC has been developing a new data processing system which should remove this forced truncation (Chesley et al. 2017).

The systematic uncertainty σ_{SYS} affects mainly phase curves where apparent magnitudes from different observatories are combined, though observatories with changes in their calibration pipelines may also introduce

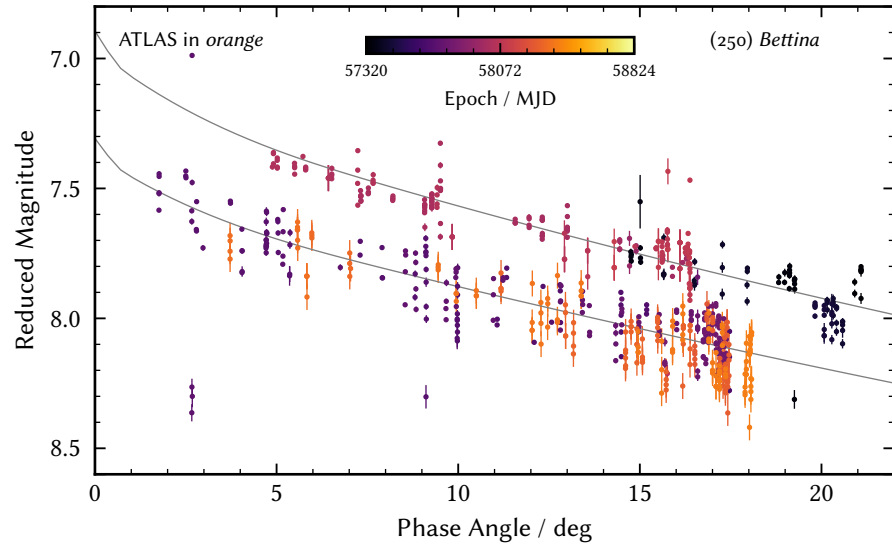


Figure 7.2: The phase curve of (250) *Bettina* as observed by ATLAS in orange. The observations are colour-coded by their epoch, highlighting the four different apparitions that were captured. The triaxial ellipsoid ratios of (250) *Bettina* are 1.4:1:1 (Viikinkoski et al. 2017). The grey lines show the H , G_1 , G_2 -model fits to the apparitions, split into pairs of two. Published in Mahlke et al. (2021).

it into their data. This is discussed in-depth in Oszkiewicz et al. (2011). To illustrate σ_{SYS} , I make use of the fact that Pan-STARRS and SDSS share the g , r , i , and z filters, though the transmission curves differ slightly. I compute the difference in the apparent magnitudes that the surveys would observe based solely on the difference in their filter design for the 24 taxonomic class templates provided in DeMeo et al. (2009). The average σ_{SYS} are 0.09 mag, 0.01 mag, 0.02 mag, and 0.08 mag for g , r , i , and z respectively. σ_{SYS} may thus be larger than σ_{PHOT} , showing the difficulty in combining phase curve observations from different sources. At the same time, as outlined further below, phase curve estimation benefits from a densely-sample phase curve. As such, there is a trade-off scenario when combining data from different observatories.

7.2.2 Rotational and Aspect Angle

In addition to the observational uncertainties, there are the uncertainties on the apparent magnitudes introduced by the unknown rotational phase and aspect angle of the asteroid (Alvarez-Candal et al. 2022). The rotation adds a periodic modulation to the lightcurve, generally with a period in the order of a few hours (Warner et al. 2009). The aspect angle of asteroids may change between different apparitions, leading to an unknown offset in the apparent magnitude. These uncertainties are the main difference between targeted and non-targeted observations, as they are accounted for in the former scenario either via the acquire light curves (e. g. Shevchenko et al. 1997).

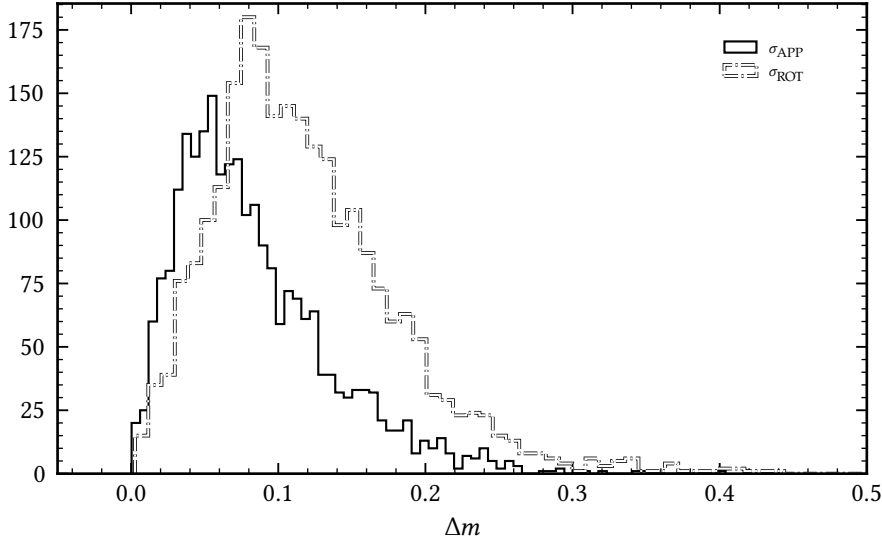


Figure 7.3: The amplitude of magnitude dispersion due to the rotation of the asteroids (black) and due to the change in aspect angle between different apparitions (white). The modulations are computed based on asteroid shape models from the DAMIT database Durech et al. (2010). Published in Mahlke et al. (2021).

The effect of both effects is visible in the phase curve of (250) *Bettina* shown in Fig. 7.2. The observations include four different apparitions of the asteroid, indicated via the colour coding. The offset in the apparent magnitude between the first and third and the second and fourth apparitions is readily apparent and visualised by fitting observations from these apparitions pairs separately with the H , G_1 , G_2 model.

To quantify the order of magnitude of σ_{ROT} and σ_{APP} , I use shape models of 2407 asteroids in the Database of Asteroid Models from Inversion Techniques (DAMIT) database (Durech et al. 2010) to approximate each asteroid as a triaxial, smooth ellipsoid, where the axes dimensions a , b , and c are extracted from the shape models and $a > b > c$. The modulation of the apparent magnitude due to the spin and apparition is then given by

$$m = -2.5 \log \left(\pi abc \cdot \left(\left(\frac{\cos \beta \cos \lambda}{b} \right)^2 + \left(\frac{\cos \beta \sin \lambda}{a} \right)^2 + \left(\frac{\sin \beta}{c} \right)^2 \right)^{0.5} \right), \quad (7.5)$$

where λ , β are the longitude and latitude of the subobserver point (Surdej and Surdej 1978; Ostro and Connolly 1984). Next, for each asteroid, I simulate lightcurves covering a full rotation around its spin axis for every ten days over an entire orbit around the Sun to probe a variety of Sun-asteroid-observer geometries. For each lightcurve, the unknown rotational state of the asteroid during observations described with σ_{ROT} is estimated as the root-mean-square difference of all apparent magnitudes to the mean apparent magnitude of the

lightcurve. The offset σ_{APP} due to the change in aspect angle is given by the difference of the mean apparent magnitudes of the lightcurve at hand to the one of a reference lightcurve where the observer is within the equatorial plane of the asteroid. The distributions of σ_{ROT} and σ_{PHOT} are given in Fig. 7.3. The unknown rotational state gives a median uncertainty of 0.11 mag, slightly larger than the 0.07 mag as median value of the apparition offset. Both effects are significant contributors to the overall uncertainty budget. As for σ_{SYS} , σ_{APP} may be completely avoided by not combining asteroid observations acquired at different apparitions, however, the trade-off between uncertainty in apparent magnitude and phase curve coverage occurs here as well.

7.2.3 Effect on Phase Curve Parameters

I now explore how the uncertainties discussed above affect the inference of the phase curve parameters H , G_1 , and G_2 . To this end, I simulate 100 non-targeted phase curves of (20) *Massalia*, based on the one observed by Gehrels (1956) and shown in Fig. 2.3. The phase curves have the same number of samples (6) at the same phase angles, however, I randomly draw the apparent magnitudes from normal distributions with means at the observed apparent magnitudes from Gehrels (1956) and standard deviations given by the propagated uncertainties given in Eq. (7.4), where $\sigma_{PHOT}=0.1$ mag, $\sigma_{PREC}=0.1/\sqrt{12}$ mag, $\sigma_{SYS}=0.05$ mag, $\sigma_{ROT}=0.11$ mag, and $\sigma_{APP}=0.07$ mag. I then fit each phase curve using the Bayesian parameter inference and compare the resulting parameters to the ones of the original, targeted observation of (20) *Massalia*. The median and standard deviations of the H , G_1 , G_2 model parameters of the simulated phase curves and the original one are given in the first row of Table 7.1. For H and G_1 , we see considerable systematic offsets, while all three parameters show standard deviations of about 0.1. This may be considered as order of magnitude on the dispersion in the phase curve parameters due to the apparent magnitude uncertainties inherent in non-targeted phase curves. The final degree of alteration of the phase curve further depends on its sampling pattern, as explored in the next part.

7.2.4 Coverage of Opposition Effect

Phase curves are sufficiently described by the photometric slope and the size of the opposition effect. While the former is observable over a wide range of phase angles and generally well described, the latter takes place in a narrow phase angle region, $\alpha \leq 5$ deg. As mentioned above, many observations are biased towards large phase angles due to observations at quadrature. Therefore, I explore how insufficient coverage of the opposition effect affects the inferred phase curve parameters.

First, I select the asteroid phase curves from ATLAS which (1) have a minimum observed phase angle of $\alpha_{min} \leq 1$ deg, (2) have at least 50 samples, $N \geq 50$, and (3) whose target asteroids have shape models in the DAMIT database. 917 phase curves of 720 individual asteroids remain after this selection. The apparent magnitudes observed by ATLAS are then corrected for the rotation- and apparition-induced dispersions using the DAMIT shape

Table 7.1: Effect of magnitude dispersion and insufficient opposition effect coverage on phase curve parameters. The columns give the mean (μ) and standard deviation (σ) of the difference in the phase curve parameters of the altered to the original observations, refer to the text. Published in Mahlke et al. (2021).

	μ_{Δ_H}	σ_{Δ_H}	$\mu_{\Delta_{G_1}}$	$\sigma_{\Delta_{G_1}}$	$\mu_{\Delta_{G_2}}$	$\sigma_{\Delta_{G_2}}$
σ	0.04	0.11	0.07	0.08	-0.01	0.10
1 deg	0.00	0.07	-0.00	0.04	0.01	0.01
2 deg	-0.00	0.12	-0.01	0.07	0.01	0.02
3 deg	-0.01	0.15	-0.02	0.09	0.02	0.03
4 deg	-0.00	0.18	-0.03	0.09	0.03	0.03
5 deg	0.02	0.21	-0.03	0.10	0.05	0.05

models as well as a provided lightcurve generation software.⁵ In the following simulation, this correction should increase the influence of the phase angle sampling pattern over the magnitude dispersion studied in the previous part.

The mean α_{\min} of the 912 phase curves is 0.6 deg. I now compute the H , G_1 , G_2 model parameters for the complete phase curves followed by truncating the phase curve at phase angles below $\{1, 2, 3, 4, 5\}$ and repeating the phase curve fit. The change in the inferred phase curve parameters should now reflect the importance of sufficient opposition effect coverage. The results of the simulation are given in Table 7.1, using the same notation as for the simulation results in the previous part. A large dispersion of the resulting H , G_1 , and G_2 parameters is apparent, including systematic offsets for the phase curve coefficients G_1 and G_2 . Compared to the effect of the magnitude dispersion of non-targeted phase curve, the lack of the opposition effect coverage affects the derived absolute magnitude in particular, with uncertainties of up to 0.2 mag.

The simulated changes in the phase curve coefficients is displayed in Fig. 7.4. Of the 720 asteroids in the sample, 112 are C- and 218 are S-types. A 2D Kernel Density Estimation (KDE) with a Gaussian kernel is fitted to the G_1 , G_2 parameters of the complete phase curves. White contours give the 1-level of the fitted probability distribution, i. e. 68% of the summed probabilities is contained within the contours. The solid black (dash dotted, dotted) contour gives the 1-level of the fitted 2D KDE corresponding to the G_1 , G_2 parameters of the phase curves truncated at 1 deg (3 deg, 5 deg). The figure shows that, as the opposition effect coverage diminishes, so does the compositional information decrease.

This results is relevant in particular for large scale surveys such as the LSST, where absolute magnitudes are the only way to acquire colours of asteroids. A sufficient amount of observations of minor bodies at opposition

⁵ https://astro.troja.mff.cuni.cz/projects/damit/pages/software_download

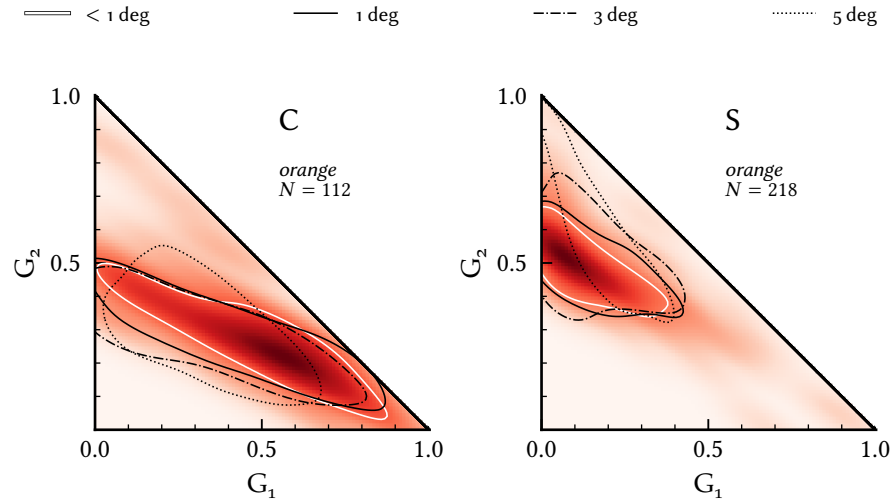


Figure 7.4: Effects of insufficient phase curve coverage on the phase curve coefficients G_1 , G_2 . The red distributions show the 2D Kernel Density Estimation distributions fitted to the G_1 , G_2 parameters of complete phase curves of C- and S-type asteroids. The white contours give the 1-level of kernel density estimators. The solid black (dash dotted, dotted) contour gives the 1-level of the fitted probability densities corresponding to the G_1 , G_2 parameters of the phase curves truncated at 1 deg (3 deg, 5 deg). Published in Mählke et al. (2021).

is required to achieve this, which has to be accounted for in the observation planning.

7.3 COMPOSITIONAL INFORMATION

Following the discussion of the observations, Bayesian parameter inference, and uncertainty budget, I now discuss the compositional information contained in the parameters of non-targeted phase curves. Given the results on the phase curve parameter uncertainties from the previous section, in particular the effect of insufficient opposition effect coverage, I select the sample from the original 180 025 phase curves which should a priori contain compositional information. I select phase curves which have at least one observation below phase angles up to 3 deg ($\alpha_{\min} \leq 3$ deg), which rejects about one third of the original sample. I further require at least 50 observations, $N \geq 50$, and a maximum observed phase angle α_{\max} of at least 10 deg. A total of 127 012 phase curves of 94 777 individual asteroids, with 36 441 phase curves observed in *cyan* and 90 571 observed in *orange*.

These are fitted with the H , G_1 , G_2 and H , G_{12}^* models as described in Sect. 7.1.2. The two-parameter model H , G_{12}^* is less flexible than the three-parameter model H , G_1 , G_2 , and Muinonen et al. (2010b) motivate it as predictive model for sparsely-sampled phase curves. The resulting phase curve parameters are available at the CDS.⁶

⁶ <http://cdsarc.u-strasbg.fr/viz-bin/qcat?VII/288>

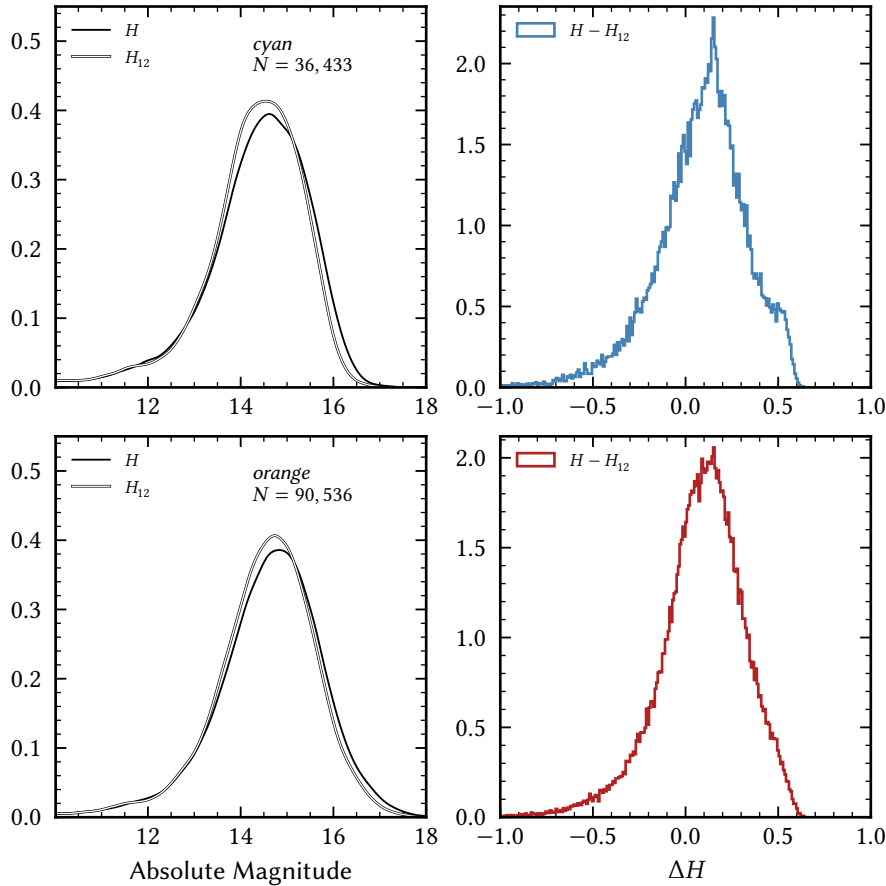


Figure 7.5: *Left*: The distribution of the absolute magnitudes H (black) and H_{12} (white) derived from phase curves of 94 777 asteroids observed by ATLAS in *cyan* (top) and *orange* (bottom) using the H, G_1, G_2 - and the H, G_{12}^* model. Magnitudes below 10 (0.8% of the sample) are not shown for readability. *Right*: The difference of the absolute magnitudes derived with the two photometric models for phase curves acquired in *cyan* (top) and *orange* (bottom). Published in Mahlke et al. (2021).

7.3.1 Absolute Magnitudes

The left hand side of Fig. 7.5 shows the distribution of the absolute magnitudes H and H_{12} of the asteroids in *cyan* and *orange* inferred from the phase curves. The four distributions peak around magnitude 15 and have a slightly larger tail towards smaller magnitudes.

The compositional information in the absolute magnitudes of phase curves depends on the photometric filter that the observations were acquired in. Erasmus et al. (2020) studies the information in the *cyan-orange* colours and finds that it is sufficient to distinguish between members of the C- and S complexes. Other combinations of filter systems such as SDSS and VISTA have been shown to contain enough compositional information to distinguish among the taxonomic complexes and some endmember classes (Carvano et al. 2010; DeMeo and Carry 2013; Popescu et al. 2018a). We may thus conclude

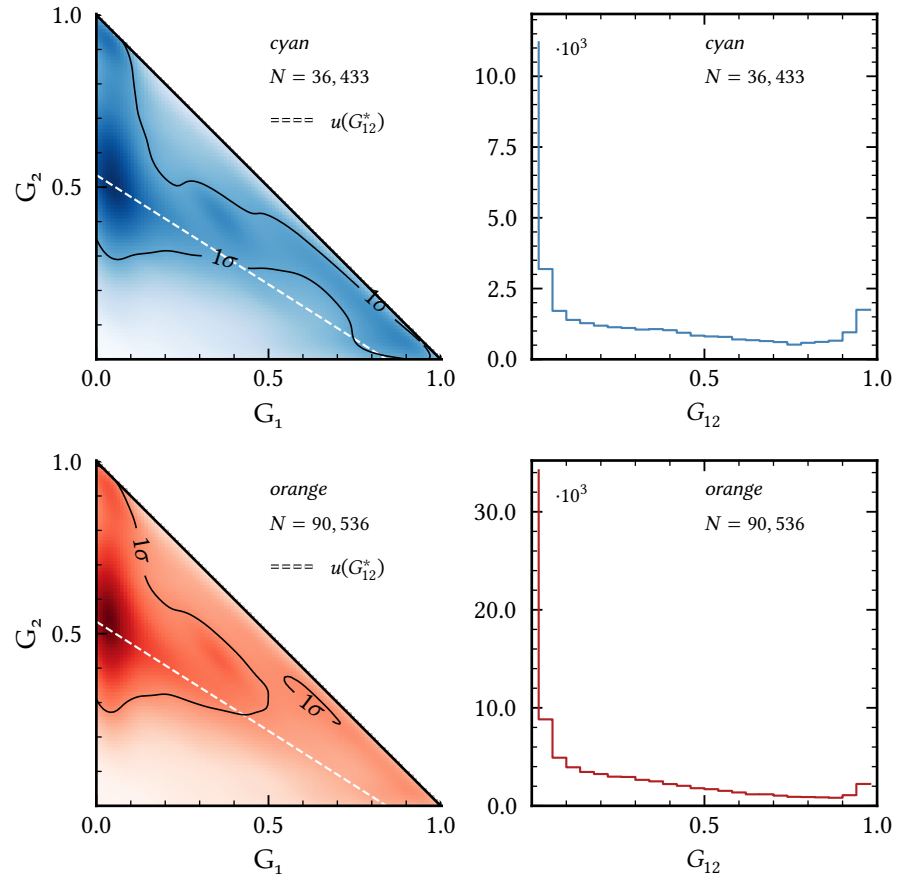


Figure 7.6: *Left*: The 2 D-KDE distribution fitted to the G_1 , G_2 parameters of the phase curves observed in *cyan* (top) and *orange* (bottom). The black contours outline the 1σ -levels of the KDE distributions. *Right*: The histogram of the G_{12}^* -parameter derived from the same sample of phase curves, aligned in the same order. Published in Mahlke et al. (2021).

that, assuming the absolute magnitude is accurate, this phase curve parameter is compositionally informative.

An open question remains: Does the model selection affect the compositional conclusion? The distribution of the differences $H - H_{12}$ is shown on the right hand side of Fig. 7.5 for the phase curves acquired in *cyan* (top) and those acquired in *orange* (bottom). Both distributions peak around 0.1 mag and have tails which include differences up to 1 mag. Given the sample-selection criteria, the fitted phase curves are not sparsely sampled, and as such, I expect the H , G_1 , G_2 model to be a better predictor of the absolute magnitude than the more limited H , G_{12}^* model. Accepting H as ground truth, I can quantify the effect of the model selection by comparing the difference in absolute magnitudes to the differences of taxonomic classes in a given colour space, for which I choose the common SDSS colours. I compute this difference by convolving each the Bus-DeMeo class template spectra⁷ with the filter transmission curves of the SDSS system (Rodrigo et al. 2012; Rodrigo and

⁷ <http://smass.mit.edu/busdemeoclass.html>

Solano 2020). The average colour difference between the classes is 0.03 mag ($u - g$), -0.03 mag ($g - r$), -0.04 mag ($r - i$), and 0.02 mag ($i - z$). These mean differences are significantly smaller than the average difference in absolute magnitudes acquired due to the model selection. I thus conclude that the H, G_{12}^* model applied to sparse phase curves may suffice as estimator for the absolute magnitude, however, it is not sufficient for taxonomic classification based on colours.

7.3.2 Phase Curve Coefficients

Results of the study of phase curve coefficients such as presented in Bel-skaya and Shevchenko (2000) and Penttilä et al. (2016) show that the G_1, G_2 parameters describing the shape of the phase curve contain compositional information. The left hand side of Fig. 7.6 shows the 2D KDE fitted to the distribution of the G_1, G_2 parameters of the phase curves acquired in *cyan* (top) and in *orange* (bottom) using a Gaussian kernel. Black contours give the 1-level of the fitted probability distribution. The white line gives the projection of the G_{12}^* parameter into the G_1, G_2 parameter space, refer to Eq. (2.6).

Comparing the distribution of G_1, G_2 parameters for the 127 012 phase curve fits with the results of Shevchenko et al. (2016) shown in Fig. 2.4, we can conclude that most asteroids in the sample show coefficients typical of S-type asteroids. This is in line with the observational bias for S-types in wide-field imaging surveys as they have higher albedos, are on average closer to Earth and more numerous than the C-type asteroids (DeMeo and Carry 2013). In the *cyan* distribution, the bias is less pronounced, and we see the 1-distribution extend towards the low-albedo, large-photometric-slope region of the G_1, G_2 parameter space. In both *cyan* and *orange*, the G_1, G_2 parameters occupy the same regions as the results of Shevchenko et al. (2016) do for the targeted phase curves.

The majority of phase curve coefficients in both *cyan* and *orange* are located above the line defining the G_{12}^* parameter. Again assuming that the H, G_1, G_2 model better describes the phase curves as they are densely sampled, I conclude that the H, G_{12}^* model cannot accurately describe phase curves where the G_1, G_2 parameters are not located on the G_{12}^* definition line. In particular, we see a tendency to overestimate the opposition effect by the H, G_{12}^* model, as the opposition effect increases non-linearly towards the origin of the G_1, G_2 parameter space, refer to Fig. 2.4, and most phase curve coefficients lie above the G_{12}^* line. This explains the large tail towards negative differences of the $H - H_{12}$ distributions shown on the right hand side of Fig. 7.5.

The distributions of the G_{12}^* parameters of all phase curve fits are shown on the right hand side of Fig. 7.6. They are biased towards small G_{12}^* values in both *cyan* and *orange*, with 42 % and 50 % of the samples having values below 0.1 respectively. A smaller bias towards $G_{12}^* = 1$ is further visible in both distributions. The tendency towards the edges of the physically valid values shows that most phase curves are not well-described with the H, G_{12}^* model. As for the absolute magnitude, I therefore conclude that the G_1, G_2

Table 7.2: Mapping of various taxonomic classifications found in the literature into a simplified class scheme. The classes are sorted in increasing value of the mean visual albedo. Published in Mahlke et al. (2021).

Class	Σ	N	\bar{p}_V	σ_{p_V}
P, PC, PD, X, XC, XD, XL, Xc, Xe, Xk, Xt	→ P	593	0.05	0.02
D, DP	→ D	425	0.06	0.02
Cgh, Ch	→ Ch	266	0.06	0.06
B, F, FC	→ B	523	0.08	0.06
C, CB, CD, CF, CG, CL, CO, Cb, Cg, Cgx, Co	→ C	3,670	0.09	0.09
T	→ T	62	0.12	0.06
M, X, XD, XL, Xc, Xe, Xk, Xt	→ M	660	0.15	0.05
K	→ K	586	0.18	0.09
L, LQ, Ld	→ L	776	0.19	0.09
O	→ O	5	0.21	0.10
S, SQ, SV, Sa, Sk, Sl, Sp, Sq, Sqw, Sr, Srw, Sv, Sw	→ S	8,875	0.26	0.08
A, AQ	→ A	69	0.28	0.09
Q, QO, QV	→ Q	185	0.28	0.11
V, Vw	→ V	1,412	0.36	0.11
E, X, XD, Xc, Xe, Xn, Xt	→ E	46	0.46	0.16
X, XD, XL, Xe, Xk, Xt	→ X	202	-	-

coefficients contain compositional information while the G_{12}^* coefficients do not.

7.3.3 Distribution of Taxonomic Complexes

To further highlight the compositional information of the G_1 , G_2 parameters, I compute the distribution of the parameters for asteroids assigned to a given taxonomic class in the literature, making use of rocks as shown in Chapter 6. I further increase the required number of phase curve samples from $N = 50$ to $N = 125$ to ensure that the opposition effect is well-described by the phase curves. This decreases the number of phase curves by about half, from 127 012 to 61 184.

The taxonomic classifications and visual albedos of the asteroids are retrieved from various reference sources using the SsODNet service⁸ of the IMCCE, which is further explained Chapter 9. The classifications follow various taxonomic schemes. I define a mapping of all classifications to 15 different taxonomic classes, using the information of the visual albedo to divide the X-complex into E, M, and P. The mapping is shown in Table 7.2. The majority of

⁸ <https://ssp.imcce.fr/webservices/ssodnet/>

the previous classifications follow the Bus-DeMeo taxonomy, however, some classifications following the Tholen (1984) or Tedesco et al. (1989) schemes are also in the sample. In particular, there are several cases where two or more classes are merged into the final designation, a common method to indicate uncertainty about the class assignment prior to the flexibility introduced by the Bus-DeMeo scheme.

Most classes are straight-forward to map to one of the core classes introduced in Chapter 4. The X-complex is resolved using the visual albedo information if present, mapping asteroids with $p_V \leq 0.075$ to P, $0.075 < p_V < 0.3$ to M, and $p_V \geq 0.3$ to E. O-types are excluded due to the small number of classifications in the sample, while ambiguous assignments such as DS or CQ are rejected due to the high compositional uncertainty of these classifications.

Figure 7.7 shows the distribution of the G_1 , G_2 coefficients in cyan and orange for the 15 taxonomic classes, described using 2D KDEs with Gaussian Kernels. The black contours give the 1 level of the KDEs. The classes are sorted in increasing value of the mean visual albedo (refer to Table 7.2), which makes it readily apparent that the G_1 , G_2 distributions correlate well with the albedo. Again, we observe that the distributions for each class resemble those of the targeted phase curve results in Shevchenko et al. (2016). The high-albedo classes such as E, V, A, and S occupy regions of low photometric slope and large opposition effect, while low-albedo classes such as P and D show large photometric slopes and low opposition effects. The intermediate region is sparsely populated by members of the M and T classes. I summarise the distributions in Table 7.3 in terms of the geometric centre and area of the 1σ -probability distributions as measure for the location and dispersion of the class members in G_1 , G_2 space. The table further gives the photometric slope and size of the opposition effect based on the geometric centre for each class in both orange and cyan.

The distributions of the phase curve coefficients show in general significant dispersions, with the worst cases being C and D. I attribute parts of this dispersion to the uncertainties on the phase curve coefficients discussed in Sect. 7.2. Another, possible larger contributing factor is the number of misclassified asteroids in this dataset. Most classifications are based on SDSS photometry which, as discussed in Carvano et al. (2010), is ambiguous for a number of asteroids. Furthermore, the separation of the X-complex relies on the visual albedo, which on average carries uncertainties of around 17.5% (Masiero et al. 2018). This hypothesis is supported by the small dispersion of the Ch-class in G_1 , G_2 space, as assigning this classification is generally done more carefully than the much larger C-class.

Nevertheless, Fig. 7.7 shows that G_1 , G_2 parameters of non-targeted phase curves carry compositional information and, in particular, classes with large differences in albedo are readily distinguishable from one another.

7.3.4 Wavelength-Dependency of Phase Curve Coefficients

The wavelength-dependence of phase curve coefficients is a well-known phenomenon (Gradie and Veverka 1986) and is the underlying reason for the non-linear reddening of reflectance spectra with increasing phase an-

Technically, G_1 , G_2 correlate with the albedo in the same photometric filter as the phase curve observations. Considering the visual albedo is observed at $0.55\ \mu\text{m}$, this is more accurate for cyan than for orange, as shown in Fig. 2.1.

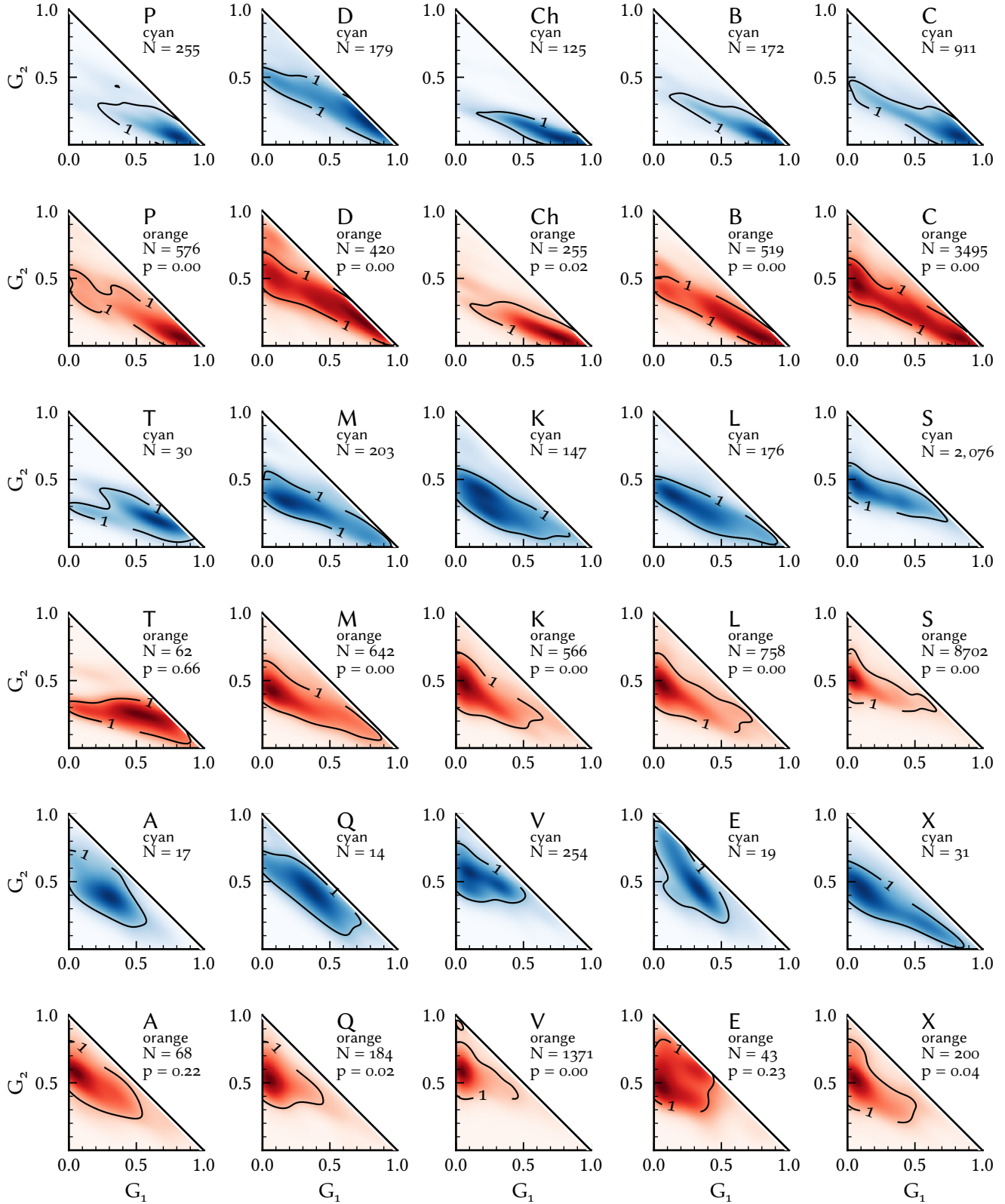


Figure 7.7: The distributions of the G_1 , G_2 phase coefficients for 15 taxonomic classes are described using 2D KDEs with Gaussian kernels, colour-coded for the cyan and orange photometric filters of the ATLAS observatory. The black contours give the 1 level of the KDE distributions. The number N of phase curves per class and filter are given in addition to the two-sample 2D Kolmogorov-Smirnov p -values, computed between the distributions in cyan and orange for each class. The classes are sorted in increasing value of the mean visual albedo. Published in Mahlke et al. (2021).

Table 7.3: Summary of the distribution of phase curve coefficients G_1, G_2 for 15 taxonomic classes. For each class, the number N of phase curves in *cyan* (subscript c) and *orange* (subscript o) is given, followed by the geometric centres C and areas A of the 1σ -outlines of the fitted 2D probability distributions. The areas are multiplied by a factor of 1000 for notation purposes. From the geometric centres, the photometric slope k and the size of the opposition effect $\zeta - 1$ is computed. The p -value of the two-sample 2D Kolmogorov-Smirnov statistic is given in the last column. Published in Mahlke et al. (2021).

Σ	N_c	N_o	C_c	C_o	A_c	A_o	k_c	k_o	$\zeta - 1_c$	$\zeta - 1_o$	p
P	255	576	(0.80, 0.05)	(0.83, 0.06)	4.0	6.4	-1.82	-1.81	0.16	0.12	0.00
D	179	419	(0.77, 0.17)	(0.72, 0.20)	8.5	10.6	-1.67	-1.62	0.06	0.09	0.00
Ch	125	255	(0.77, 0.05)	(0.76, 0.07)	4.1	5.2	-1.84	-1.80	0.22	0.21	0.02
B	172	519	(0.82, 0.06)	(0.77, 0.08)	4.5	8.0	-1.82	-1.79	0.14	0.17	0.00
C	965	3,609	(0.82, 0.06)	(0.83, 0.06)	6.2	5.0	-1.81	-1.82	0.13	0.13	0.00
T	30	62	(0.65, 0.19)	(0.53, 0.24)	6.3	7.5	-1.61	-1.49	0.18	0.29	0.66
M	203	642	(0.19, 0.34)	(0.07, 0.42)	9.0	7.5	-1.05	-0.77	0.92	1.02	0.00
K	147	566	(0.18, 0.40)	(0.06, 0.48)	8.6	6.6	-0.99	-0.72	0.71	0.87	0.00
L	176	758	(0.16, 0.37)	(0.06, 0.47)	9.0	6.7	-0.96	-0.73	0.89	0.89	0.00
S	2,076	8,702	(0.08, 0.46)	(0.04, 0.51)	6.4	3.5	-0.76	-0.67	0.87	0.81	0.00
A	17	68	(0.30, 0.39)	(0.05, 0.57)	7.5	6.2	-1.16	-0.68	0.46	0.60	0.22
Q	14	184	(0.36, 0.44)	(0.05, 0.52)	9.2	4.6	-1.18	-0.70	0.25	0.74	0.02
V	254	1,371	(0.10, 0.56)	(0.04, 0.58)	6.5	3.2	-0.78	-0.67	0.50	0.60	0.00
E	19	43	(0.33, 0.45)	(0.06, 0.48)	8.0	8.8	-1.14	-0.73	0.29	0.86	0.23
X	31	200	(0.11, 0.45)	(0.06, 0.52)	9.0	5.6	-0.83	-0.70	0.81	0.73	0.04

gle (Sanchez et al. 2012). This spectral reddening for example affects the reflectance spectra from the third data release of the ESA *Gaia* satellite, some of which are acquired at phase angles above 30 deg (Cellino et al. 2020; Galuccio et al. 2022). However, using phase curves, we can compute the degree of spectral reddening between the wavelengths at which the phase curves were acquired. The spectral slope in units of %/100 nm is given by

$$S_S = \frac{f_o - f_c}{\lambda_o - \lambda_c} \cdot 10^4, \quad (7.6)$$

where f_c and f_o are the observed reflectances in *cyan* (subscript c) and *orange* (subscript o), and $\lambda_c = 518$ nm and $\lambda_o = 663$ nm are the effective wavelengths of the filters. Using the Pogson scale, we can then express the spectral slope as

$$S_S = \frac{f_c(10^{-0.4(m_o - m_c)} - 1)}{\lambda_o - \lambda_c} \cdot 10^4. \quad (7.7)$$

Normalising the reflectance at λ_c gives $f_c = 1$ and the remaining variable is the difference $m_o - m_c$, which we derive using the phase curves $m_c(\alpha)$ and $m_o(\alpha)$

$$\Delta m = m_o(\alpha, H_o, G_{1,o}, G_{2,o}) - m_c(\alpha, H_c, G_{1,c}, G_{2,c}). \quad (7.8)$$

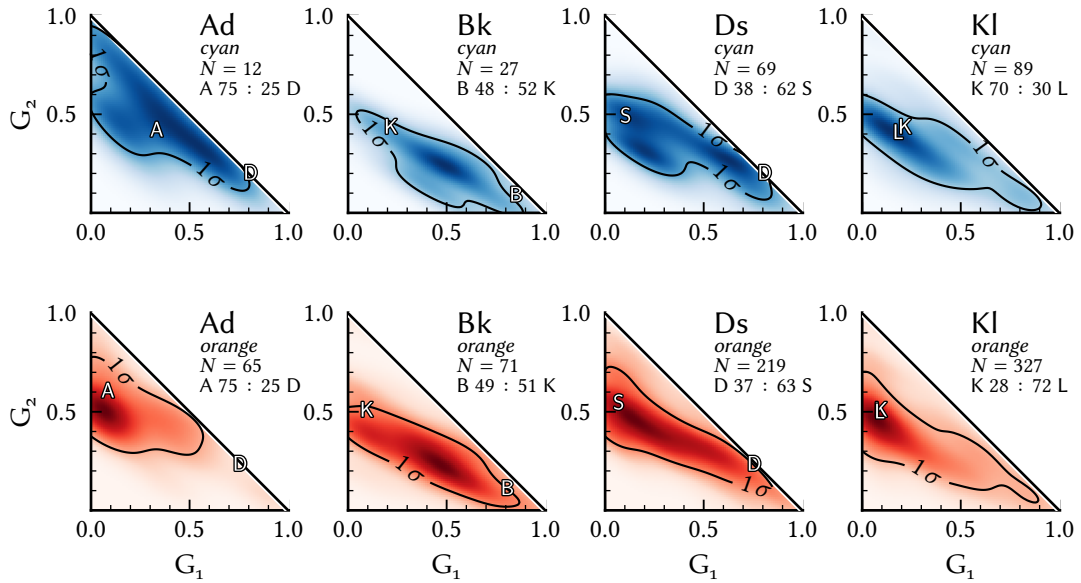


Figure 7.8: As Fig. 7.7, but giving the distribution in G_1 , G_2 of the taxonomic classes Ad, Bk, Ds, and Kl from Popescu et al. (2018a). The number N of observations per class is given as well as their derived ratio in percent between the core classes that the degenerate classes comprise. The geometric centres of the class distributions in G_1 , G_2 are indicated by their respective letters, refer to Table 7.3.

Carvano and Davalos (2015) show that the spectral reddening with phase angle is different for the taxonomic complexes. In particular, objects which present a siliceous $1\ \mu\text{m}$ -band are more prone to spectral reddening. The ATLAS dataset of dual-band observations for a large number of asteroids is a prime case to further investigate the complex dependency. However, one caveat of the observations is that the *cyan* and *orange* filters partially overlap, thus, the signal of a possible wavelength-dependency is reduced.

For each taxonomic class, I use the Kolmogorov-Smirnov p -value to describe the similarity between the G_1 , G_2 distributions in *cyan* and *orange*. The p -value quantifies the probability that the underlying population of two sampled distributions is identical (Peacock 1983). The results are given in Fig. 7.7 and in Table 7.3. Most values are below 0.1, indicating the that the distributions do not sample the same population, i. e. the G_1 , G_2 distributions are wavelength-dependent. The three classes where p is larger than 0.1 (A, E, and T) are three of the four smallest samples. Visual comparison of the G_1 , G_2 distributions in *cyan* and *orange* of these classes in Fig. 7.7 suggests that they may differ, yet the p -value does not capture it, likely due to the small sample size.

Table 7.4: Summary of the G_1 , G_2 distribution observed in asteroid families. Given are the number N of family members with phase curve coefficients, the geometric centers C of the G_1 , G_2 1σ probability-level contour, and the area A of the 1σ -contour as observed by ATLAS in *cyan* (subscript c) and *orange* (subscript o). The area A is multiplied by 1,000 for notation purposes. Further given are the taxonomic classifications of the families and their references.

Family	N_c	N_o	C_c	C_o	A_c	A_o	Class	Reference
(4) <i>Vesta</i>	229	1647	(0.07, 0.50)	(0.04, 0.55)	162	106	V	Zappalà et al. (1990)
(5) <i>Astraea</i>	59	524	(0.11, 0.48)	(0.07, 0.48)	197	156	S	Huaman et al. (2017)
(10) <i>Hygiea</i>	101	473	(0.75, 0.11)	(0.08, 0.44)	191	184	C	Carruba (2013)
(15) <i>Eunomia</i>	383	1647	(0.11, 0.43)	(0.06, 0.49)	183	170	S	Nathues (2010)
(24) <i>Themis</i>	528	1218	(0.80, 0.05)	(0.73, 0.08)	96	151	C	Mothé-Diniz et al. (2005)
(93) <i>Minerva</i>	114	539	(0.08, 0.49)	(0.07, 0.49)	159	170	S	Mothé-Diniz et al. (2005)
(135) <i>Hertha</i>	264	1777	(0.36, 0.34)	(0.07, 0.49)	172	131	S	Dykhuis and Greenberg (2015)
(158) <i>Koronis</i>	502	1333	(0.06, 0.46)	(0.03, 0.52)	122	71	S	Tholen (1984)
(170) <i>Maria</i>	100	472	(0.19, 0.41)	(0.05, 0.47)	228	169	S	Zappalà et al. (1997)
(221) <i>Eos</i>	697	2732	(0.13, 0.36)	(0.04, 0.44)	174	134	K	Masiero et al. (2014)

7.3.5 Identification of Interlopers

In Chapter 2, I mention that the compositional information in phase curve coefficients and visual albedos is orthogonal to the one acquired from spectroscopy and photometry. This means that we may use information acquired from the former to resolve degeneracies occurring in the latter. An example of these degeneracies are the Ad, Bk, Ds, and Kl classes presented in Popescu et al. (2018a). These designations are based on NIR colours from the VISTA survey. Popescu et al. (2018b) shows that pairs of classes are degenerate in this observable, and fuse the respective class letters into new classes, e. g. class Ad contains members of the classes A and D.

Using the G_1 , G_2 information of the asteroids classified in Popescu et al. (2018a) and present in the ATLAS sample, I refine their taxonomic classification by computing the distance in G_1 , G_2 of each phase curve to the geometric centres of the respective degenerate classes (refer to Table 7.3), assigning each asteroid to the class it is closest. The distribution in G_1 , G_2 of the asteroids for each class and the derived ratio of the core classes in percent are given in Fig. 7.8. It is apparent that this method is most effective for classes with large differences in the albedo, such as A and D. On the other hand, the separation of the centres of K and L in G_1 , G_2 inspires little hope that the phase curve information is able to resolve this degeneracy.

A more appropriate statistical treatment may account for the actual probability distributions derived for each class instead of the strongly reduced information of the geometric centre. Nevertheless, this underlines how non-targeted phase curve aid in the compositional characterisation of asteroids.

7.3.6 Distribution of G_1 , G_2 in Families

Compositional information for a large number of asteroids are of particular interest for the analysis of asteroid families (Mothé-Diniz et al. 2005; Thomas et al. 2012; Masiero et al. 2015, e. g.), where ensemble properties like the taxonomic class are used to identify family interlopers, i. e. spurious asteroids assigned to the family based on proximity in orbital space. Compositional information can further reveal large scale trends within the family such as the presence of subfamilies, as is the case for the *Themis* family (*Beagle* subfamily, Mothé-Diniz et al. (2005)) or the *Koronis* family (*Karin* subfamily, Vernazza et al. (2006)).

From the sample of phase curves with more than 125 observations used to derive the distribution of the taxonomic complexes in G_1 , G_2 above, I select the asteroids assigned to families which have at least 500 members observed in the sample of phase curves. This is the case for ten families, whose G_1 , G_2 distribution I summarise in Table 7.4 and shown in Fig. 7.9 in proper elements space, colour-coded using the G_1 , G_2 values of the family members. G_1 , G_2 corresponding to high-albedo classes are red, those corresponding to low-albedo classes are blue. Readily apparent are the uniformity of the *Vesta* and *Themis* family members based on the visual impression in Fig. 7.9. Note that the *Beagle* subfamily is a B-type family (Fornasier et al. 2016) while *Themis* is a C-type family (Mothé-Diniz et al. 2005), hence, the uniformity in G_1 , G_2 is expected even when accounting for the subfamily. More heterogeneous are the *Eos* and *Hertha* family, the latter of which is part of the *Nysa-Polana* family complex (Dykhuis and Greenberg 2015).

Figure 7.9 shows that compositional outliers in families may be recognised in G_1 , G_2 space. Outliers in both the *Vesta* and *Themis* families stand out from the background population. However, referring back to Fig. 7.7, we have to account for the dispersion of the G_1 , G_2 coefficients of each taxonomic class, independent of whether it is inherent to the class or spurious dispersion due to the uncertainties of the phase curve parameters. Therefore, I argue that the phase curve parameters from non-targeted phase curves are best regarded as an ensemble property to quantify the composition of a large number of asteroids, while the parameters of single asteroid are less informative with respect to its composition.

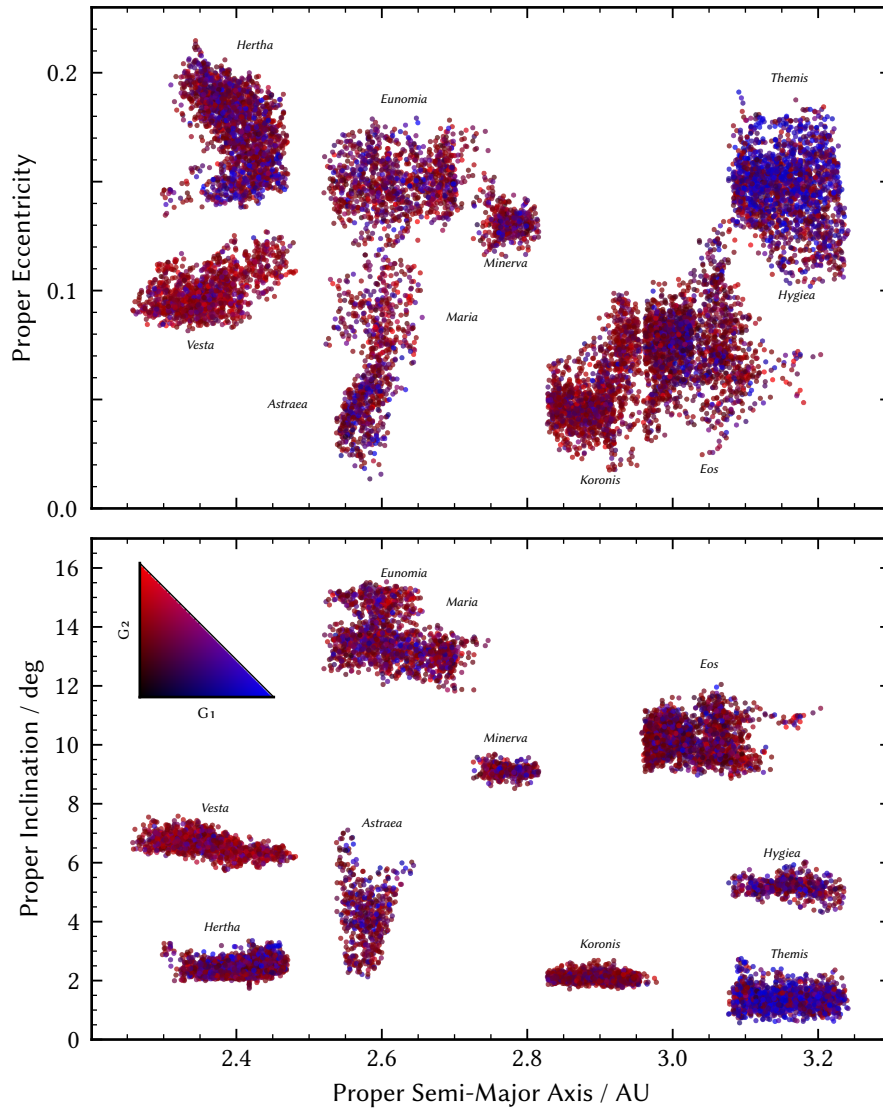


Figure 7.9: The composition of asteroid families based on phase curve coefficients G_1 , G_2 . Each asteroid with a G_1 , G_2 measurement is colour-coded by its corresponding values using the colour-mapping shown in the top left corner of the lower part.

Part III

A New Asteroid Taxonomy

MOTIVATION

8.1	Practical Considerations	97
8.2	Compositional Considerations	98

The introduction of a new classification scheme requires convincing motivation through promising advancement and improvement to overcome the inertia within the community to stick with established system. The current state-of-the-art, the Bus-DeMeo system (Bus and Binzel 2002a; DeMeo et al. 2009), is the cumulation of 50 years of taxonomic classification of asteroids. It is at least the fourth major taxonomic scheme and captures by far the largest compositional entropy among the asteroid population compared to its predecessors.

In this chapter, I outline the motivation to revise the taxonomic scheme. I separate the motivation into practical considerations, which include aspects on the observables we use to classify asteroids, as well as compositional considerations, which regard the design of the classes in light of recent observational studies. A comparison of the resulting taxonomy with previous taxonomies is given in Table 8.1.

Yes, just like in XKCD #927.

8.1 PRACTICAL CONSIDERATIONS

The aspects in this part of the motivation derive primarily from the methodology of deriving the taxonomy which has been followed by major taxonomies since Tholen (1984). As outlined more in-depth in Chapter 5, the mathematical approach used to derive the taxonomy limits the resulting classification scheme to be only applicable to complete observations. The Bus-DeMeo system, built on VisNIR spectra, can thus only classify asteroids which have a complete VisNIR spectrum observed. However, intuitively, any part of the spectrum that is observed provides information about the asteroid's composition. The visible- or NIR parts on their own may be sufficient for a good estimation of the final VisNIR classification (Bus and Binzel 2002a; Popescu et al. 2018a), further discussed in Chapter 12.

In addition, the necessity of completeness dictates that adding observables to the classification space always decreases the number of asteroids which can be classified in the new classification scheme. If I extend the classification to account for both the VisNIR spectra and the visual albedo using this mathematical approach, I can only classify the intersection of the asteroids which have both observables observed. This strict limit of the sample size led Tholen (1984) to use the visual albedo only as a secondary classification step and Bus and Binzel (2002a) to remove it altogether. Yet, adding observables like the visual albedo increases the compositional entropy that can be captured in the taxonomy, leading to a taxonomy with a larger vocabulary.

Both issues outlined above highlight the potential improvement of the taxonomy that lies in employing a method which accounts for any partial information that is present about the asteroid. Mathematically speaking, the method has to account for the *missing data*, as explained in [Chapter 5](#).

The second practical improvement I propose for the current taxonomy concerns the definition of class boundaries. The design of the first classes in Chapman et al. (1975) and Zellner and Gradie (1976) was done for a few tens of objects, and hard boundaries in terms of observable parameters were defined to divide the asteroids into classes. As we observed more asteroids, the space between these classes became populated with the continuum that is present between almost all types of asteroids (Bus 1999). The boundaries are subjective, both in their absolute values and in their total amount. Yet, if we strive to derive a observable-driven taxonomy, we should seek to minimise subjectivity as much as possible. Defining the class boundaries may thus benefit from a proper statistical treatment, referred to as *clustering* and further outlined in [Chapter 5](#).

8.2 COMPOSITIONAL CONSIDERATIONS

Since the last revision of the asteroid taxonomy by DeMeo et al. (2009), the mineralogical interpretations of the three complexes presented in [Chapter 4](#) have advanced thanks to focused observational efforts. It is thus opportune to align the language of the taxonomy again with these findings.

The classes of Tholen's C-complex (B,C, F, and G, here referred to as BCFG-types) have long been interpreted as alteration sequence of material similar to CI and CM chondrites (Hiroi et al. 1993, 1996; Clark et al. 2010; León et al. 2012). The spectral variability among the BCFG-types is here explained by different degrees of Aqueous Alteration and Thermal Metamorphism, derived from the matching spectral appearance of experimentally altered meteorite samples and the asteroids. Vernazza et al. (2015) pointed out that, even after accounting for sample biases, these metamorphosed meteorites are rare in comparison to the mass fraction that the associated classes represent within the Main Belt. Furthermore, the authors mentioned that the measured densities between the BCFG-types and the supposed CM-like Ch-class (Burbine 1998; Fornasier et al. 1999; Rivkin 2012) defined in Bus and Binzel (2002a) differ within a factor of 1.3-3, which is at odds with a common meteoritic analogue material, as well as discrepancies in the mid-infrared spectral behaviour.

Instead, Vernazza et al. (2015) raised the possibility that the classes B, C, D, and P consist of IDPs, a material which is not at all represented in the meteorite collection as it does not survive the Earth's atmosphere in significant quantities. Via radiative transfer models, they are able to construct convincing matches between mixtures of constituents of IDPs and spectral observations of members of the asteroid classes. The Ch presenting the 0.7 μm are convincingly matched to CM chondrites in a different study by Vernazza et al. (2016), including a large observational campaign of Ch-types, confirming previous tentative links between the meteorites and their parent bodies (Burbine 1998; Fornasier et al. 1999; Rivkin 2012).

Either hypothesis on the nature of the C-complex material highlights the continuity between the members. It is thus counter-intuitive to have classes based on differences in UV- or NIR slope, such as C and C_b defined in the Bus-DeMeo system, in particular as the surface grain size may fully explain variations in the slope (Vernazza et al. 2016). We might seek a simplification of the class scheme in this complex, supposing that it arises from the data. Furthermore, Vernazza et al. (2015) highlighted that the P-class belongs to the C-complex. The primitive nature of P-types was never in doubt (Gradie and Tedesco 1982), yet the principles of the observable-based taxonomy detained the P into the degenerate X-complex. We may overcome this issue using the visual albedo.

An open issue I see for the C-complex lies in the relatively low correlation between the 0.7 μm -band of the phyllosilicates which is commonly used as proxy of the 3 μm -band, which represents the direct observation of hydroxyl or water on the surface of asteroids. Rivkin et al. (2015b) summarises previous observational results and concludes that the 0.7 μm -band is only present in 50 % of the cases where the 3 μm -band is present). Thus, the C_h class is an incomplete description of the hydrated C-complex. The only way to resolve this is to include the mid-infrared region around 3 μm into the taxonomy, which is not done in this work due to the sparsity of these observations. NASA's James Webb Space Telescope (JWST) may add significantly to recent observational efforts in this area (Usui et al. 2011; Takir and Emery 2012; Takir and Emery 2022).

The diversity of mineralogies in the X-complex was at first due to the spectral degeneracy of the E, M, and P classes in the ECAS colours, utilised in the Tholen (1984) system, and later due to the removal of the visual albedo as differentiator in the Bus-DeMeo system. Both systems are prevalent in recent literature (e. g. Neeley et al. 2014; Shepard et al. 2015). The advantage of Tholen's system using the visual albedo has been discussed sufficiently at this point, while the feature-based differentiation of the X-complex introduced in Bus and Binzel (2002a) and DeMeo et al. (2009) and amended in Binzel et al. (2019) contains important information as features are the most reliable indicators of the mineralogy, as further discussed in Chapter 10. Here, the taxonomy could benefit from a combination of both systems, made possible by the re-introduction of the visual albedo into the observables space.

Detailed characterisations of some M- and X-types in a variety of observables has so far been unsuccessful in identifying a common denominator for these objects (Clark et al. 2004; Shepard et al. 2008c, 2010; Neeley et al. 2014; Shepard et al. 2015). The spectral appearance in terms of silicate features, hydration in the 3 μm -region, and the NIR-behaviour are as unpredictable as the densities and radar albedos among these objects. For a taxonomy built around the visual albedo and VisNIR spectroscopy, there is thus little hope that mineralogical trends may be identified. I believe that the taxonomy here should guide the conversation by allowing to highlight spectral features, yet the similarity of these objects should not be hidden by intensive subclassing.

The S-complex is the best-understood taxonomic complex. The *Hayabusa* sample return from (25143) *Itokawa* confirmed the previous association of S-types and OCs in general and (25143) *Itokawa* and LL-chondrites in particular

Table 8.1: Comparison of taxonomy presented in this work with previous taxonomies. PCA and MCFA refer to dimensionality reduction methods discussed in [Chapter 5](#). “Obs.” abbreviates “observations”, “Dim.” abbreviates “dimensionality”. The number of classes derived in this work is motivated in [Chapter 10](#).

	OBS.	OBSERVABLES	DIM. REDUCTION	CLUSTERING	CLASSES
Chapman et al. (1975)	110	Visible Spectrophotometry, Albedo, Polarimetry	Manual	Manual	2
Tholen (1984)	405	Visible Photometry: Seven Colours, (Albedo)	PCA	Manual	14
Tedesco et al. (1989)	357	Visible Photometry: Two Colours, Albedo	-	Manual	11
Bus and Binzel (2002a)	1447	Visible Spectra	PCA	Manual	26
DeMeo et al. (2009)	371	VisNIR Spectra	PCA	Manual	25
Mahlke et al. (2022)	2983	VisNIR Spectra, Albedo		MCFA	17

(Binzel et al. [2004b](#); Nakamura et al. [2011](#)). The effects of space-weathering on siliceous surfaces has been extensively investigated experimentally (e. g. Brunetto et al. [2006](#); Chrbolková et al. [2021](#)). We may thus seek to align the class definition with the mineralogical interpretation based on our understanding of its affect on the spectral appearance. In particular, classes in the S-complex should align primarily with variations in the olivine and pyroxene mineral abundances and only secondarily in slope, as the latter is affected by space-weathering.

Given our understanding of the S-complex, it may seem intuitive to align the asteroid- to the meteorite classification here. As given in [Chapter 3](#), there are three groups of OCs, H, L, and LL. However, they are defined based on ratios of metallic to oxidised iron abundances, which may not be sufficiently reflected in the spectral appearance to allow for differentiation via remote sensing. Indeed, Vernazza et al. ([2015](#)) showed that most asteroids spectra, fall into two compositional groups, H-like and LL-like, and are similar to unequilibrated (PT is 3) OC. We regard this result more in [Chapter 14](#). In either case, as for the other complexes, the S-complex may benefit from a reduction in the number of subclasses with respect to the Bus-DeMeo system, in particular those based on variations in the slope.

OBSERVABLES

9.1	Selection	101
9.2	Preprocessing	103
9.2.1	Reflectance Spectra	103
9.2.2	Visual Albedo	105
9.3	Final Dataset	106
9.3.1	Data Availability	107
9.3.2	Bibliographic Record	108

The choice of the observables used to classify asteroids shapes the resulting taxonomic scheme like no other decision during the process. Initially, I intended to include four different observables in this taxonomic scheme, presented in [Chapter 2](#): reflectance spectra, colours, visual albedo, and phase curve coefficients. The idea behind this choice was that the compositional information is learned from the reflectance spectra and visual albedo of a few thousand asteroids, while the link to the colours and phase curve coefficients allows to classify a much larger number of asteroids.

In the following, I motivate the decision to classify based on reflectance spectra and visual albedo “only”. I then describe the data accumulation and preprocessing steps done to produce the final dataset used to derive the taxonomy. [Figure 9.1](#) gives an overview of these steps, [Tables 9.1](#) and [9.2](#) give the bibliographic sources of the reflectance spectra, visual albedos, absolute magnitudes, and diameters used in the data compilation process. The preprocessing steps can be applied to observations using the `classy` tool presented in [Chapter 12](#). Several ideas and results presented here are treated in [Mahlke et al. \(2022\)](#).

9.1 SELECTION

The inclusion of reflectance spectra and visual albedos as observables in this taxonomy requires little motivation: the former is the compositionally most informative observable that can be acquired via remote-sensing while the value of the latter has been demonstrated in earlier works ([Chapman et al. 1975](#); [Tholen 1984](#)). At the same time, the compositional information contained in colours ([DeMeo and Carry 2013](#); [Popescu et al. 2018a](#); [Sergeyev and Carry 2021](#)) is comparable if not greater to that of the albedo and, as shown in [Mahlke et al. \(2021\)](#) and [Chapter 7](#), the phase curves coefficients G_1 , G_2 contain compositional information even if they are acquired using non-targeted observations and lack the correction of magnitude modulations induced by the 3D shape of the target. Colours and phase curve coefficients are more accessible observables than reflectance spectra and albedos, meaning that a significantly larger number of asteroids may be classified if they are included in the taxonomic scheme.

I thus included colours from SDSS (visible), VISTA, and Two Micron All Sky Survey (2MASS) (NIR) as well as the phase curve coefficients G_1 , G_2 in *cyan* and *orange* from the catalogue presented in Mahlke et al. (2021) and Chapter 7 in the data. After applying upper limits on the uncertainties of the colours, the time difference between the measurement of the apparent magnitudes, the final database of colours included 260 423 entries of 122 254 individual asteroids, where each entry contained one or more colours either from SDSS, VISTA, or 2MASS. For G_1 , G_2 , I selected all phase curves based on at least 125 observations with a minimum solar phase angle below 3 deg and a maximum solar phase angle above 10 deg, selected based on the discussions presented in Chapter 7. A total of 61 163 phase curve coefficients of 51 673 individual asteroids remained from the initial dataset of 127 012 phase curves, 10 327 acquired in *cyan* and 50 836 in *orange*. The number of colours and phase curve coefficients exceeded the number of reflectance spectra and visual albedos (presented below) by factors of 3-35, where the latter number is the ratio of colours to reflectance spectra.

As outlined in Chapter 5, the machine learning approach that taxonomies are based on aims to generalise patterns it identifies in the input data. The more a pattern is present in the data, the more it will be emphasised in the resulting model. Due to the large colour observations in the initial dataset, the patterns in the colour information dominated the separation of the observations in the reduced data projection. These patterns, however, are far less compositionally informative than the ones contained in the reflectance spectra. A similar situation arose with the visual albedo and the phase curve coefficients, which are correlated to a certain degree as described in Chapter 2.

The first solution that comes to mind is to reduce the amount of colours and phase curve coefficients included in the model. After truncating the dataset to equal ratios of the observables, however, the high variance of the reflectance spectra and visual albedos compared to the lower-resolution observables leads to a model which neglects the information in the latter.

There is certainly a middle ground to this issue, where the right ratio of low- and high-resolution observables leads to an equilibrated taxonomic model. However, I came to the conclusion that adding low-resolution data like asteroid colours to a taxonomy which already contains the highly correlated and more informative reflectance spectra is a conceptual mistake. The taxonomic classes should be derived on the most informative data only, as the compositional information in the colours is already contained in the reflectance spectra. Again, a similar conclusion is made for the visual albedo and phase curve coefficients.

Therefore, I built the taxonomy based on the reflectance spectra and visual albedos. Nevertheless, the correlation between colours and spectra as well as albedos and phase curve coefficients makes me suspect that it is worth exploring a method to classify asteroids with colour or phase curve information based on a taxonomic scheme derived on the other observables.

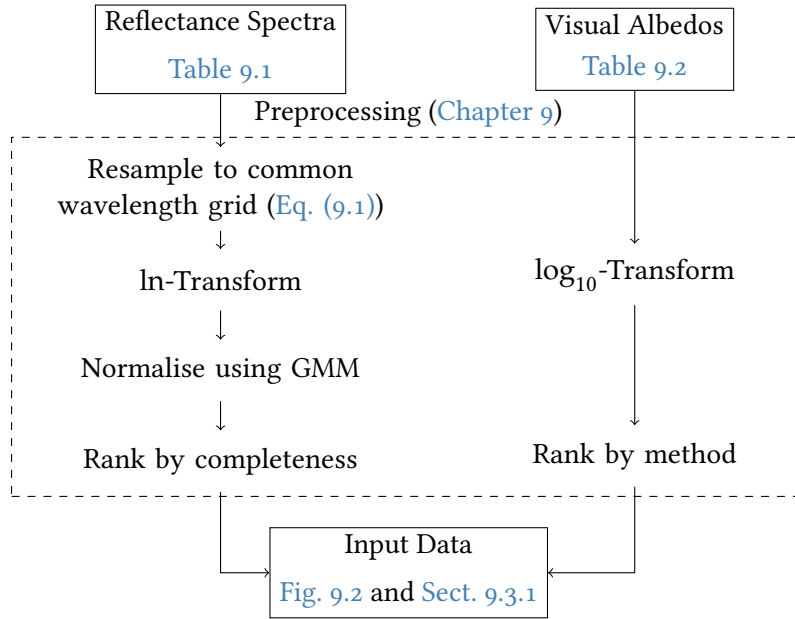


Figure 9.1: The preprocessing steps of the reflectance spectra and visual albedos before merging them into a single dataset used to derive the asteroid taxonomy.

9.2 PREPROCESSING

With the observables chosen, I now outline the data accumulation and preprocessing steps.

9.2.1 Reflectance Spectra

The reflectance spectra used in this work are a collection from numerous sources, refer to Table 9.1 at the end of this chapter. The collection is built from online repositories such as Small Main-Belt Asteroid Spectroscopic Survey (SMASS) and PDS, from spectra presented in the literature, and by contacting observers directly. The initial collection contains more than 7500 spectra. The major contributors are SMASS (Xu et al. 1995) and MITHNEOS (Binzel et al. 2019; Marsset et al. 2022). After rejection of duplicates and low-quality data as well as limiting the number of spectra per individual asteroid to five at most, 5906 spectra of 4526 individual asteroid remained. Over 50 % of the spectra are visible-only data, refer to Fig. 9.2.

The spectra in the collection are heterogeneous in numerous aspects, such as their sampling pattern and the observed wavelength region. For the following analysis, however, a homogeneous sampling pattern is required. This pattern should be fine enough to accurately describe absorption feature parameters like the band centres and widths while, at the same time, limit the dimensionality of the dataset as much as possible due to the curse of dimensionality outlined in Chapter 5. I follow DeMeo et al. (2009) in the design of the sampling pattern, though I increase the density of the sampling

A unified database of spectral observations of asteroids including relevant metadata like the epoch of observation would be a worthy effort. The SMASS website is a good example though it is limited to data from the SMASS and MITHNEOS surveys and lacks a programmatic access point.

in the visible to better describe the band positions of the silicate minerals around $0.7\ \mu\text{m}$, $0.9\ \mu\text{m}$, and $1\ \mu\text{m}$. The sampling pattern is given by

$$\lambda_S \in \{0.45, 0.475, 0.50, \dots, 1.0, 1.025, 1.05, 1.10, 1.15 \dots, 2.40, 2.45\} \mu\text{m}, \quad (9.1)$$

totaling 53 wavelength bins. As outlined in [Chapter 5](#), each of these wavelength bins reflects one data dimension in the following analysis.

To resample the ensemble of reflectance spectra to this pattern, I smooth the original spectra using a Savitzky-Golay filter (Savitzky and Golay 1964) and interpolate the result. The filter has two parameters, the size of the filter window and the degree of the polynomial fitted to the points in the filter window. However, the visible and the NIR parts of asteroid spectra often have different sampling patterns and noise behaviour even after they are merged. After trying different automated ways to determine these parameters given in the literature (e. g. Vivó-Truyols and Schoenmakers 2006), I finally set them manually for each spectrum based on the visual agreement of the smoothed spectrum and the original data. I further remove reflectance values which are obvious outliers, affected by detector drop-offs, and the occasional thermal tail (e. g. Rivkin et al. 2005).

After cleaning, smoothing, and resampling the reflectance spectra, the next step is to normalise them. This turned out to be a rather challenging. Typically, reflectance spectra are normalised to unity at a common wavelength, $0.55\ \mu\text{m}$ for visible or VisNIR data, $1.25\ \mu\text{m}$ for NIR-only data. However, my dataset contains a mixture of observed wavelength ranges, and as shown in [Chapter 5](#), the mean values of each data dimension affect the MCFA model, hence, I cannot use two different normalisation wavelengths.

The problem is complicated by the fact that there is no objectively right way to normalise reflectance spectra if the absolute reflectance is unknown. The “best” alternative normalisation depends on the use case: do I want to compare features in the visible or the NIR? Here the use case is the subsequent analysis via dimensionality reduction and clustering. After trial and error of numerous normalisation methods, I concluded that normalising spectra to a unit L_2 -norm achieves satisfying dimensionality reduction in the case of complete observations. However, for observations with missing data, L_2 normalisation is not appropriate due to the varying number of data points.

The underlying idea to find the most appropriate normalisation of the reflectance spectra is that the MCFA model identifies Gaussian clusters in the dimensions of the input data. The distribution of the reflectance values of all spectra over the individual wavelength bins is determined by the normalisation, and therefore, we seek to normalise the spectra by maximising their Gaussianity. This is achieved by fitting a GMM with a given number of components to the reflectance spectra, where the free model parameter is the normalisation constant of each spectrum. The natural logarithm is applied to the spectra prior to this model fit to convert the constant from a multiplicative to an additive one. The number of model components is set to 30 after trying out the normalisation procedure with different values.

Said mathematically, I assume that each spectrum can be written as $\alpha\mathbf{x}$, where $\alpha \in \mathbb{R}$ is a normalisation constant depending on the considered spec-

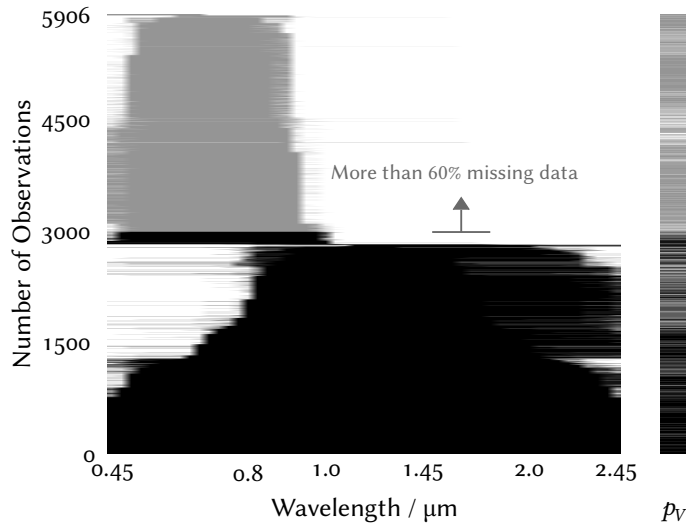


Figure 9.2: The aggregated observations consisting of reflectance spectra and visual albedos visualised as a missingness-matrix. The x -axis shows the data dimensions of the observables, the reflectance spectra and visual albedos. Each row represents one observation. The matrix cells are white if the corresponding data dimension of the observation is empty and filled if the value is observed. Black cells highlight the visible-near-infrared sample which is used to derive the asteroid taxonomy. Grey cells show the visible-only sample which is not used to derive the taxonomy but classified in a later step.

trum, while x is a spectrum to normalise. Further assuming that x follows a mixture of k log-normal distributions with diagonal covariances, all parameters of the models can be estimated from an incomplete data set via an expectation-maximisation algorithm (Dempster et al. 1977). The normalisation algorithm is subject of a publication in preparation (Mattei et al. (incl. Mahlke), in prep.).

To normalise spectra which are not included in this original dataset, e.g. to classify new observations in this taxonomic scheme, the mean normalisation constant of the five closest neighbours of the new spectrum in Euclidean distance is taken.

9.2.2 Visual Albedo

The visual albedos used in this work were compiled for the IMCCE SsODNet service presented in Chapter 6. The main contributors in this compilation are IRAS (Matson et al. 1986; Tedesco et al. 2002), WISE (Masiero et al. 2011), AKARI (Usui et al. 2011), and Spitzer (Trilling et al. 2016) satellites.

There are different ways to identify the best estimate of the visual albedo of an asteroid. As described in Chapter 6, I use the SsODNet client rocks to first compute the visual albedo based on the asteroid's absolute magnitude and diameter using Eq. (2.4). As most catalogued asteroid albedos are derived in this manner and the absolute magnitude estimates get more accurate over time, this method should get the most reliable albedo estimate. Second, for

each asteroid, I take the weighted average mean of all available albedo estimates based on a ranking of the estimation-methods presented in [Chapter 6](#) and [Berthier et al. \(incl. Mahlke, in prep.\)](#). Finally, I add all visual albedos in the SsODNet data cloud of the asteroids of which I have reflectance spectra. The bibliographic sources of this data are given in [9.2](#). This way, I collect 4704 albedo measurements of 3543 individual asteroids.

As for the reflectance spectra, a Gaussian distribution of the visual albedos ingested into the MCFA model is preferred for the model training. [Wright et al. \(2016\)](#) show that the distribution of albedos follows a double Rayleigh distribution, where one represents the bright asteroids like S-types and the other represents the darker asteroids like C-types. Transforming the visual albedos using a logarithm of base 10 yields a double Gaussian distribution.

9.3 FINAL DATASET

With the observables in place, I now create the observations used to train the MCFA model and to derive the taxonomy. The first step is to merge the reflectance spectra and visual albedos of the individual asteroids into observations. The overarching goal is to create as many complete observations as possible and to merge the best spectra with the best visual albedos.

For each asteroid, I rank the available reflectance spectra by completeness and the available visual albedos by the computation method as outlined above. I then match the highest ranking values of each observable to build an observation and continue until there are no more reflectance spectra of this asteroid available. All observations require some coverage of the reflectance spectra as the visual albedo alone is not sufficiently compositionally-informative.

The complete dataset after merging the observables is shown in [Fig. 9.2](#) in form of a missingness-matrix. The columns represent the data dimensions, including the 53 wavelength bins and the one dimension from the visual albedo p_V . The rows show the merged observations. Matrix cells which are white represent missing data. There are 718 complete rows, i. e. VisNIR spectra including visual albedos. Another 43 observations have VisNIR spectra but no visual albedo.

In [Fig. 9.2](#), observations with more than 60 % missing wavelength bins are shown in grey. Almost all of them have visible-only spectra and most of them include a visual albedo. They make up almost 50 % of the sample. When training the MCFA model with the entire dataset of 5906 observations, a similar phenomenon occurred as discussed with the colours and the spectra above: the visible-only spectra dominated the clustering due to their limited variance when projected into the reduced space. The model was essentially learning to separate a visible-only spectrum from other observations, as, compared to VisNIR observations, visible-only spectra contain significantly less compositional information.

As above, I therefore decide to separate the spectra into a sample which I use to train the model and derive the taxonomy with, denoted the *VisNIR sample* and shown in black in [Fig. 9.2](#), and the sample of observations including mainly visible-only spectra which is later classified in the new class scheme, denoted the *visible-only sample* and shown in grey in [Fig. 9.2](#).

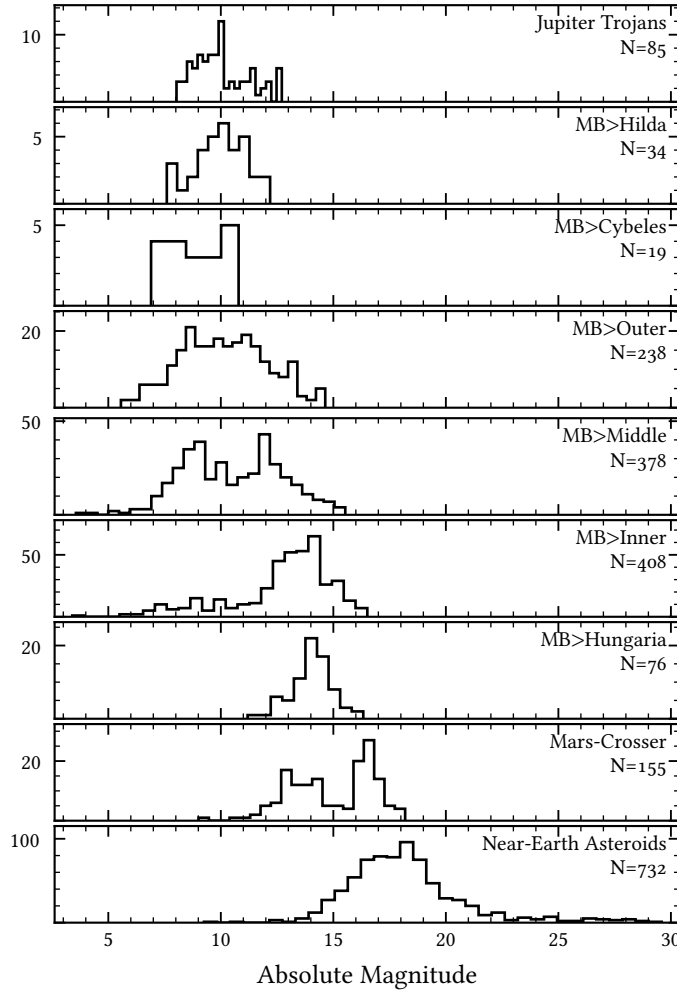


Figure 9.3: Distribution of 2125 individual asteroids in the VisNIR sample over absolute magnitude and orbital class. The number N of asteroids per orbital class is given below the respective orbital class. The bin size of the histogram varies with N . MB stands for main belt.

The VisNIR sample consists of 2983 observations of 2125 individual asteroids. Their distribution in absolute magnitude and orbital space is shown in Fig. 9.3. 1576 asteroids have a single observation in the sample, the remaining ones are represented between two and five times.

9.3.1 Data Availability

The preprocessed VisNIR- and visible-only samples shown in Fig. 9.2 are available at the CDS¹ and in the GitHub repository² of the `classy` tool presented in Chapter 12.

¹ <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/665/A26>

² <https://github.com/maxmahlke/classy>

9.3.2 *Bibliographic Record*

Being able to revise a taxonomy with an order of magnitude increase in the observations without contributing a single observation from my side is a testimony of the observational dedication of the community. I thank the authors and observers who have published their spectra in articles, online, or shared them when contacted individually. The bibliographic references of the reflectance spectra are given in [Table 9.1](#), the ones of the visual albedos, absolute magnitudes, and diameters in [Table 9.2](#).

Table 9.1: References of reflectance spectra used in this work.

Xu [1994](#); Xu et al. [1995](#); Bus [1999](#); Burbine [2000](#); Binzel [2001](#); Binzel et al. [2001](#); Bus and Binzel [2002a,b](#); Emery and Brown [2003](#); Rayner et al. [2003](#); Bendjoya et al. [2004](#); Binzel et al. [2004a,b,c](#); Birlan et al. [2004](#); Clark et al. [2004](#); Duffard et al. [2004](#); Lazzarin et al. [2004](#); Marchi et al. [2004](#); Rivkin et al. [2004](#); Lazzarin et al. [2005](#); Marchi et al. [2005](#); Vernazza et al. [2005](#); Alvarez-Candal et al. [2006](#); Birlan et al. [2006](#); Vernazza et al. [2006](#); Vilas et al. [2006](#); Birlan et al. [2007](#); Fornasier et al. [2007](#); Lazzaro et al. [2007](#); Nedelcu et al. [2007](#); Sunshine et al. [2007](#); Yang and Jewitt [2007](#); Ockert-Bell et al. [2008](#); Shepard et al. [2008b](#); Sunshine et al. [2008](#); Binzel et al. [2009](#); Burbine et al. [2009](#); Clark et al. [2009](#); Duffard and Roig [2009](#); Moskovitz et al. [2009](#); Willman et al. [2009](#); Fieber-Beyer [2010](#); León et al. [2010](#); Moskovitz et al. [2010](#); Ockert-Bell et al. [2010](#); Reddy [2010](#); Birlan et al. [2011](#); De Sanctis et al. [2011a](#); De Sanctis et al. [2011b](#); Emery et al. [2011](#); Fieber-Beyer and Gaffey [2011](#); Fieber-Beyer et al. [2011](#); Fornasier et al. [2011](#); Hardersen et al. [2011](#); León et al. [2011](#); Ostrowski et al. [2011](#); Popescu et al. [2011](#); Reddy et al. [2011](#); Yang and Jewitt [2011](#); Fieber-Beyer et al. [2012](#); Gietzen et al. [2012](#); Popescu et al. [2012](#); Jasmim et al. [2013](#); Kasuga et al. [2013](#); Sanchez et al. [2013](#); Birlan et al. [2014](#); Fieber-Beyer and Gaffey [2014](#); Fornasier et al. [2014](#); Hardersen et al. [2014](#); Kuroda et al. [2014](#); Marsset et al. [2014](#); Neeley et al. [2014](#); Polishook et al. [2014](#); Popescu et al. [2014](#); Sanchez et al. [2014](#); Vernazza et al. [2014](#); Fieber-Beyer and Gaffey [2015](#); Hardersen et al. [2015](#); Kasuga et al. [2015](#); Landsman et al. [2015](#); Fornasier et al. [2016](#); Pinilla-Alonso et al. [2016](#); Reddy and Sanchez [2016](#); Vernazza et al. [2016](#); Borisov et al. [2017](#); Lucas et al. [2017](#); Migliorini et al. [2017](#); Reddy and Sanchez [2017](#); Wong et al. [2017](#); Barucci et al. [2018](#); Borisov et al. [2018](#); De Prá et al. [2018](#); Devogèle et al. [2018](#); Hardersen et al. [2018](#); Hasegawa et al. [2018](#); Ieva et al. [2018](#); Licandro et al. [2018](#); Migliorini et al. [2018](#); Perna et al. [2018](#); Reddy et al. [2018](#); Devogèle et al. [2019](#); Lucas et al. [2019](#); Moskovitz et al. [2019](#); Popescu et al. [2019](#); Matlovič et al. [2020](#); Oszkiewicz et al. [2020](#); Yang et al. [2020](#); Arredondo et al. [2021](#); Gartrelle et al. [2021](#); Hasegawa et al. [2021a](#); Pinilla-Alonso et al. [2021](#); Marsset et al. [2022](#)

Table 9.2: References of visual albedos, absolute magnitudes, and diameters used in this work.

Bowell et al. 1994; Helfenstein et al. 1994; Thomas et al. 1994; Helfenstein et al. 1996; Thomas et al. 1996; Clark et al. 1999; Thomas et al. 1999; Thomas 2000; Veverka et al. 2000; Benner 2002; Delbó et al. 2003; Müller and Blommaert 2004; Fujiwara et al. 2006; Magri et al. 2007; Drummond and Christou 2008; Shepard et al. 2008a,b; Delbó and Tanga 2009; Keller et al. 2010; Ryan and Woodward 2010; Trilling et al. 2010; Grav et al. 2011; Mainzer et al. 2011; Masiero et al. 2011; Matter et al. 2011; Mueller et al. 2011; Sierks et al. 2011; Usui et al. 2011; Grav et al. 2012a,b; Jorda et al. 2012; Mainzer et al. 2012; Marchis et al. 2012; Masiero et al. 2012; Pravec et al. 2012; Russell et al. 2012; Alí-Lagoa et al. 2013; Huang et al. 2013; Li et al. 2013; Matter et al. 2013; Rozitis et al. 2013; Berthier et al. 2014; Mainzer et al. 2014a,b; Masiero et al. 2014; Müller et al. 2014; Rozitis and Green 2014; Becker et al. 2015; Hanuš et al. 2015; Koren et al. 2015; Nugent et al. 2015; Ryan et al. 2015; Alí-Lagoa et al. 2016; Dong-fang et al. 2016; Hanuš et al. 2016; Li et al. 2016; Licandro et al. 2016; Nugent et al. 2016; Russell et al. 2016; Trilling et al. 2016; Alí-Lagoa and Delbó 2017; Hanuš et al. 2017; Masiero et al. 2017; Viikinkoski et al. 2017; Yu et al. 2017; Alí-Lagoa et al. 2018; Drummond et al. 2018; Hanuš et al. 2018; Tatsumi et al. 2018; Herald et al. 2019; Masiero et al. 2019, 2020a,b; Chavez et al. 2021; Jiang and Ji 2021; Masiero et al. 2021; Vernazza et al. 2021; Hung et al. 2022

10.1	Application of Model to Observations	111
10.1.1	Hyperparameters	111
10.1.2	Initialisation and Training	112
10.2	Latent Space	114
10.2.1	Latent Components	114
10.2.2	Latent Scores	115
10.2.3	Clusters	116
10.3	From Clusters to Classes	116
10.3.1	Class Continuity	116
10.3.2	Probabilistic Approach	119
10.3.3	Dissolving the X-complex	121
10.3.4	Feature Flags	122
10.3.5	Feature Parametrisation	123
10.3.6	Class per Asteroid	124
10.4	Data Availability	125

This chapter describes the combination of the input data outlined in [Chapter 9](#) with the MCFA model introduced in [Chapter 5](#) to derive the taxonomy. I give an interpretation of the resulting latent space and conclude with the combination of the 50 Gaussian clusters to build 17 taxonomic classes. The model training steps are outline in [Fig. 10.1](#). The final number of samples and asteroids per taxonomical and orbital class is given in [Table 10.3](#).

10.1 APPLICATION OF MODEL TO OBSERVATIONS

In the following, I outline how the model hyperparameters are chosen and the model parameters is initialised to achieve an appropriate weighting of the visual albedo with respect to the reflectance spectra.

10.1.1 *Hyperparameters*

The MCFA model has two hyperparameters, the number of latent dimensions q that the data is reduced to and the number of mixture components G which are used to cluster the data in the reduced space. Larger parameter values lead to a model which retains more variance of the data and identifies patterns on smaller scales, while at the same time increasing the number of free model parameters and the required amount of observations to constrain them. This trade-off can be evaluated objectively using model-selection criteria like the Bayesian Information Criterion or the Integrated Completed Likelihood (Bouveyron et al. 2019). These criteria choose the model which maximises the likelihood of the data given the model parameters. However, the model assumes that the classes follow Gaussian distributions in the latent space.

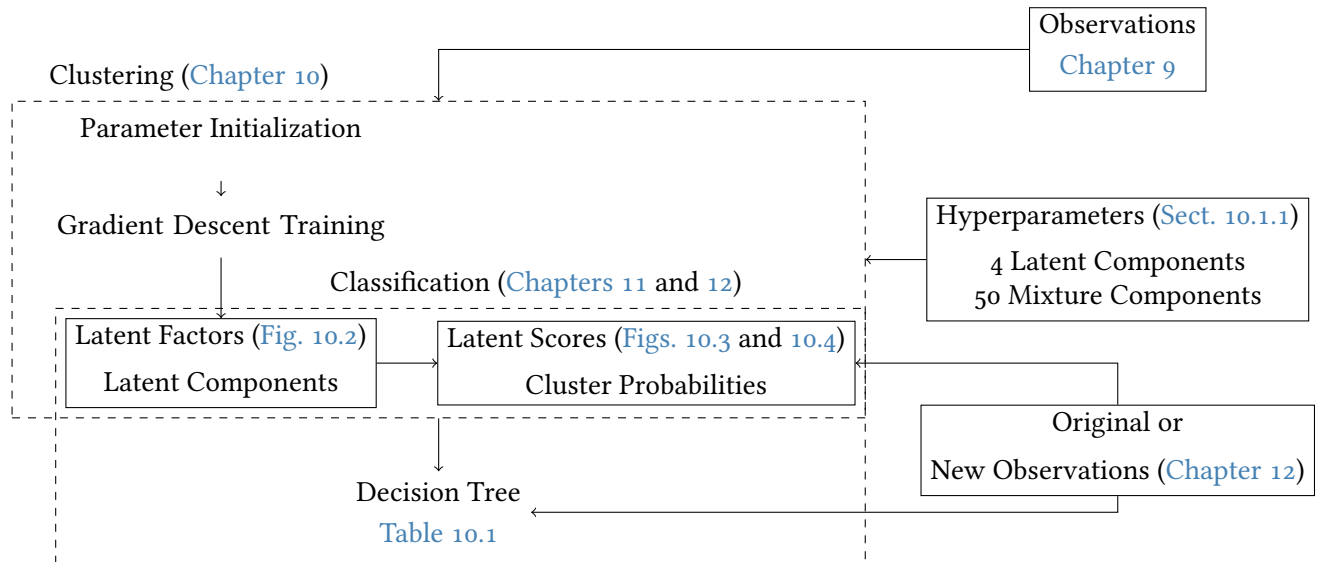


Figure 10.1: Overview of the clustering- and classification steps to derive the asteroid taxonomy. Published in Mahlke et al. (2022).

This is not necessarily true, hence, the most likely model may not represent the best taxonomy that we can derive.

Therefore, I choose a trial-and-error approach, fitting the model numerous times with different hyperparameters and evaluating the resulting clustering of the observations in the latent space. This process takes a considerable amount of time as issues with the clustering (e. g. insufficient separation or numbers of clusters within a taxonomic complex to derive appropriate asteroid classes) are often subtle and reveal themselves at a later stage when designing the cluster-to-class decision-tree as outlined below.

The hyperparameters I finally choose are $q = 4$ and $H = 50$. I note that clustering in five latent dimensions gave equally acceptable results and four dimensions are chosen as the model has a fewer free parameters. The large number of 50 mixture components initially does not mean that I expect 50 classes. Instead, this accounts for the assumption of Gaussianity: I do not expect the classes to be necessarily Gaussian in the latent space. However, a non-Gaussian cluster may be well approximated with a superposition of several Gaussian clusters. I thus anticipate the reduction of the 50 clusters into a smaller number of classes.

10.1.2 Initialisation and Training

Before the MCFA model is trained on the observations using a stochastic gradient descent algorithm, its parameters need to be initialised. These parameters are the sum of the parameters of the latent space and the mixture components (refer to Chapter 5) and their initial values define the global starting position of the stochastic gradient-descent training. As such, the initialisation has a large impact on the final result, which I exploit to solve the issue of the weighting of the albedo as follows.

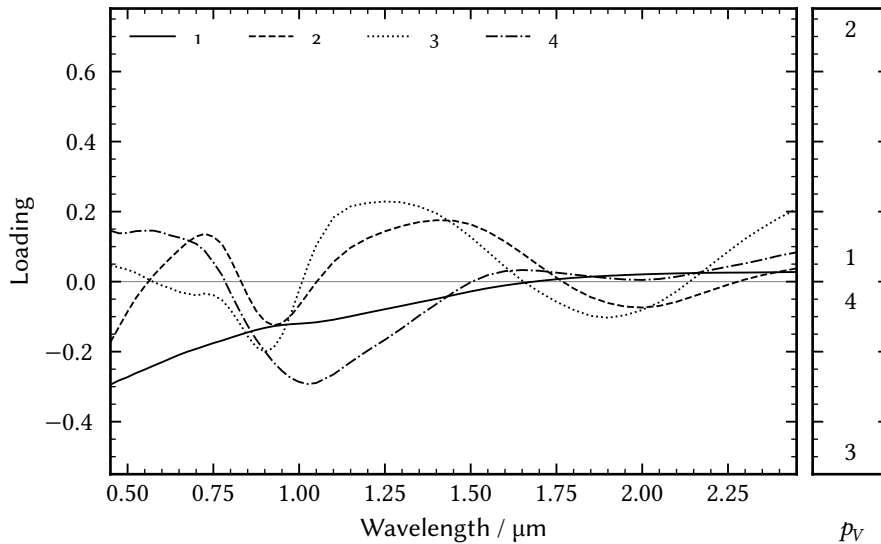


Figure 10.2: The latent components derived from the input observations using the Mixture of Common Factor Analysers model. *Left*: The loadings of the four latent components with respect to the data dimensions corresponding to the reflectance spectra. *Right*: The loadings of the four latent components with respect to the data dimension corresponding to the visual albedo. Note that both parts use the same scale for the y-axis. Published in Mahlke et al. (2022).

As described in [Chapter 5](#), PCA is variant to feature-scaling, i. e. the principal components depend on the relative variance of the data dimensions to one another. The more variance the data has in a particular dimension, the larger is the respective loading of the principal component. On the other hand, FA is invariant to feature-scaling, and the dimensionality reduction of the MCFA model is based on FA. I therefore initialise the latent space of the MCFA model using PPCA. The variance of the visual albedos in the observations (\log_{10} -transformed as described in [Chapter 9](#)) is several times larger than that of the normalised and \ln -transformed reflectance values in a given wavelength bin. Therefore, the initial latent components place a large emphasis on the visual albedo. As the model is trained, this bias is reduced with each training step, making the albedo weighting a function of the number of training epochs. I then interrupt the training at different epochs and inspect the resulting clustering, judging in particular the separation of the high-albedo E-types from the medium-albedo M-types. This is a subjective yet effective method to ensure that the single data dimension of the visual albedo is correctly weighted with respect to the 53 dimensions of the reflectance spectra.

The parameters of the mixture components are initialised by fitting a GMM with 50 components to the principal scores of the initial PPCA fit and assigning each observation to its most probable cluster.

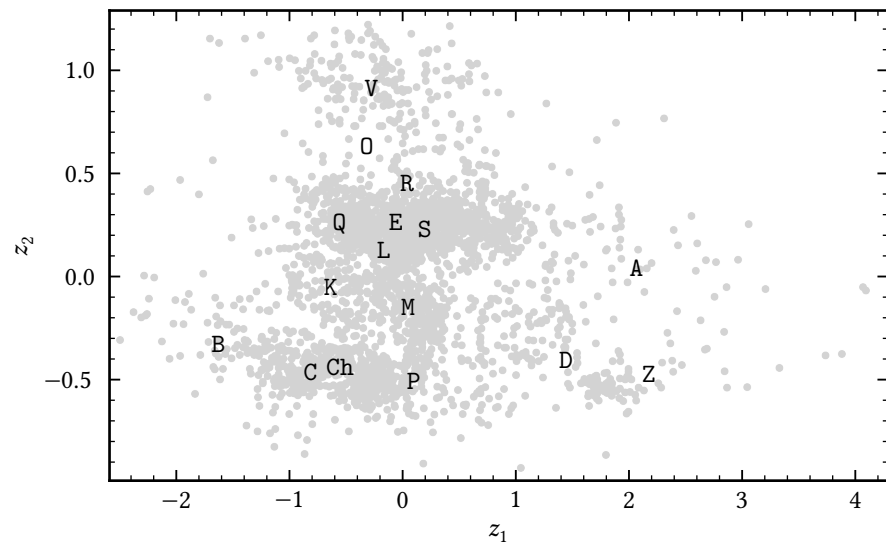


Figure 10.3: The latent scores of the input observations in the first and second latent dimension. Each grey circle corresponds to one observation. The mean latent scores of all observations assigned to the classes defined in this taxonomy are indicated by the respective class letter. For readability, the mean score of the C-class has been shifted by -0.1 in z_1 . Published in Mahlke et al. (2022).

10.2 LATENT SPACE

I now interpret the latent space resulting from the MCFA model trained on the input observations. I make use of the concepts outlined in [Chapter 5](#).

10.2.1 Latent Components

The resulting latent components are shown in [Fig. 10.2](#). As emphasised in [Chapter 5](#), they do not have a priori information, but are purely shaped by the covariance of the input data. The spectral loadings on the left hand side of [Fig. 10.2](#) show strong resemblance to the spectra of some of the minerals discussed in [Chapter 3](#). I note that the spectral components reflect the variance of the \ln -transformed spectra. The first latent component describes a featureless, red component. The second and third latent components both resemble the reflectance spectra of pyroxene assemblages, with the third one depicting deeper bands at slightly shorter wavelengths. The latter further shows a blue slope in the visible and a slight feature around $0.7 \mu\text{m}$. The fourth component has the characteristic shape of the olivine $1 \mu\text{m}$ -feature, though the indication of a $2 \mu\text{m}$ -feature and the neutral-to-blue slope in the visible weaken the resemblance.

On the left hand side of [Fig. 10.2](#), we see the loading per component corresponding to the visual albedo. The loadings of the first and fourth dimension have negligible absolute values compared to the ones of the second and third dimensions. Asteroids with large visual albedos will see an increases

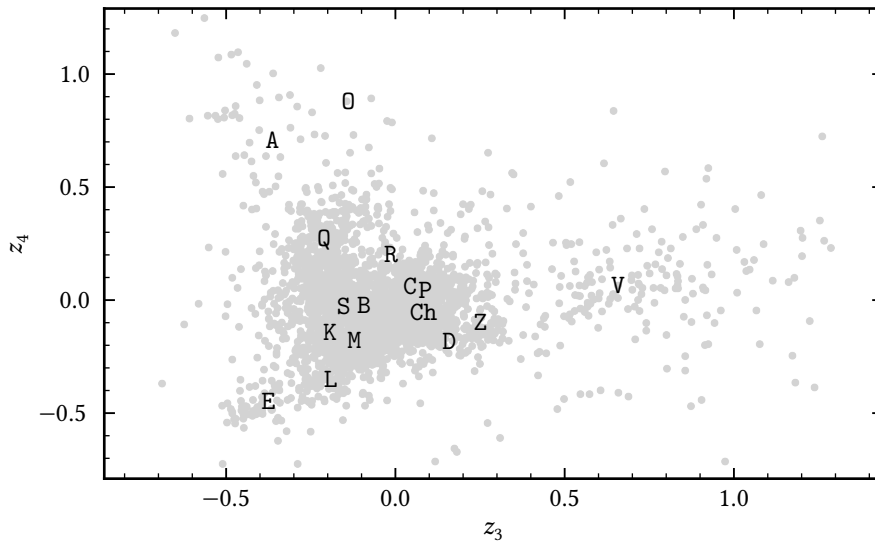


Figure 10.4: As Fig. 10.3, but giving the latent scores in the third and fourth latent dimension. For readability, the mean score of the S-class has been shifted by -0.02 in z_3 and the one of the P-class by 0.04 in z_4 . Published in Mahlke et al. (2022).

in their z_2 -score and a decrease in their z_3 -score. Nevertheless, as we are projecting both the spectrum and the visual albedo into the latent space simultaneously, the scores also depend on the spectrum. However, the latent loadings corresponding to the albedo are much larger than the individual spectral bins by design, as described in the previous section.

10.2.2 Latent Scores

The latent scores of the observations projected into the first and second latent dimension are shown in Fig. 10.3. The mean score of all observations assigned to a given class is indicated by the respective class letter. The classes are derived below and further explained in Chapter 11, yet I make use of them here to facilitate the interpretation of the latent scores. The red slope of the first latent component means that asteroids with overall red reflectance spectra have larger z_1 -scores than those with overall blue spectra. The blue B-types and the red D-types indicate this in Fig. 10.3. Furthermore, the second latent component resembles a pyroxene spectrum, hence, the V-types have the largest z_2 -scores while featureless spectra as found in the C-complex have low z_2 -scores. This appears to be contradicted by the position of the generally featureless E-types among the S-complex. The reason for this seeming misplacement is the linear combination of the dimensions of the reflectance spectra and the visual albedo into the latent components. The visual albedo contributes significantly to the second latent dimensions, as apparent from its large loading in Fig. 10.2. Therefore, though the E-types do not have pyroxene-like features, they have large z_2 -scores.

A similar effect is visible in the z_3 -scores corresponding to the third latent dimension shown in Fig. 10.4. The spectral loadings of the latent component resemble a pyroxene spectrum, hence, the V-types have the largest scores. However, the feature-rich S-types have lower z_3 -scores than the classes of the C-complex due to the negative loading of the visual albedo, refer to Fig. 10.2. The E-types have the lowest z_3 -scores. A- and O-types have the largest z_4 -scores due to the olivine-like $1\ \mu\text{m}$ -feature in combination with the shallow $2\ \mu\text{m}$ -feature of the fourth latent component. The former yields large vector products with reflectance spectra of the olivine-rich A-types as well as for classes with wide $1\ \mu\text{m}$ -feature, such as O- and Q-types.

10.2.3 Clusters

While the MCFA model is trained and the latent space takes shape, the input observations are divided into 50 clusters. Besides their Gaussianity, the clusters are not constrained in terms of their covariance matrix, yielding a variety of orientations, volumes, and number of members per cluster. The clusters could be illustrated by means of their two-dimensional projections into the latent scores in Figs. 10.3 and 10.4, however, due to their large number, these makes for rather messy figures. Instead, the reflectance spectra of observations assigned to each cluster and the distribution of their visual albedos is shown for each cluster in Figs. 10.5 and 10.6 respectively.

10.3 FROM CLUSTERS TO CLASSES

At this stage, each observation has a probability to belong to any of the 50 clusters and is assigned to the cluster with the highest probability. I now design the cluster-to-class decision-tree used to merge observations in different clusters into taxonomic classes, or to split observations in one cluster into different taxonomic classes if the Gaussian clustering does not describe compositional trends. The resulting decision-tree is summarised in Table 10.1.

10.3.1 Class Continuity

As outlined in Chapter 4, the class definition should primarily be objective and data-driven. However, continuity between the resulting scheme and the two most similar predecessors, the Tholen (1984) scheme and the Bus-DeMeo system (Bus and Binzel 2002a; DeMeo et al. 2009), as well as mineralogical and meteoritic interpretations are accounted for as well. These influences and the classes themselves are outlined in detail in Chapter 11.

Differences to the existing taxonomic schemes arise primarily due to two changes: the reintroduction of the albedo enabling me to dissolve the X-complex and the fundamental difference between manual and visual clustering. In particular classes which represent continuous trends within the population of asteroids and niche classes like O and R are difficult to recreate in a probabilistic approach.

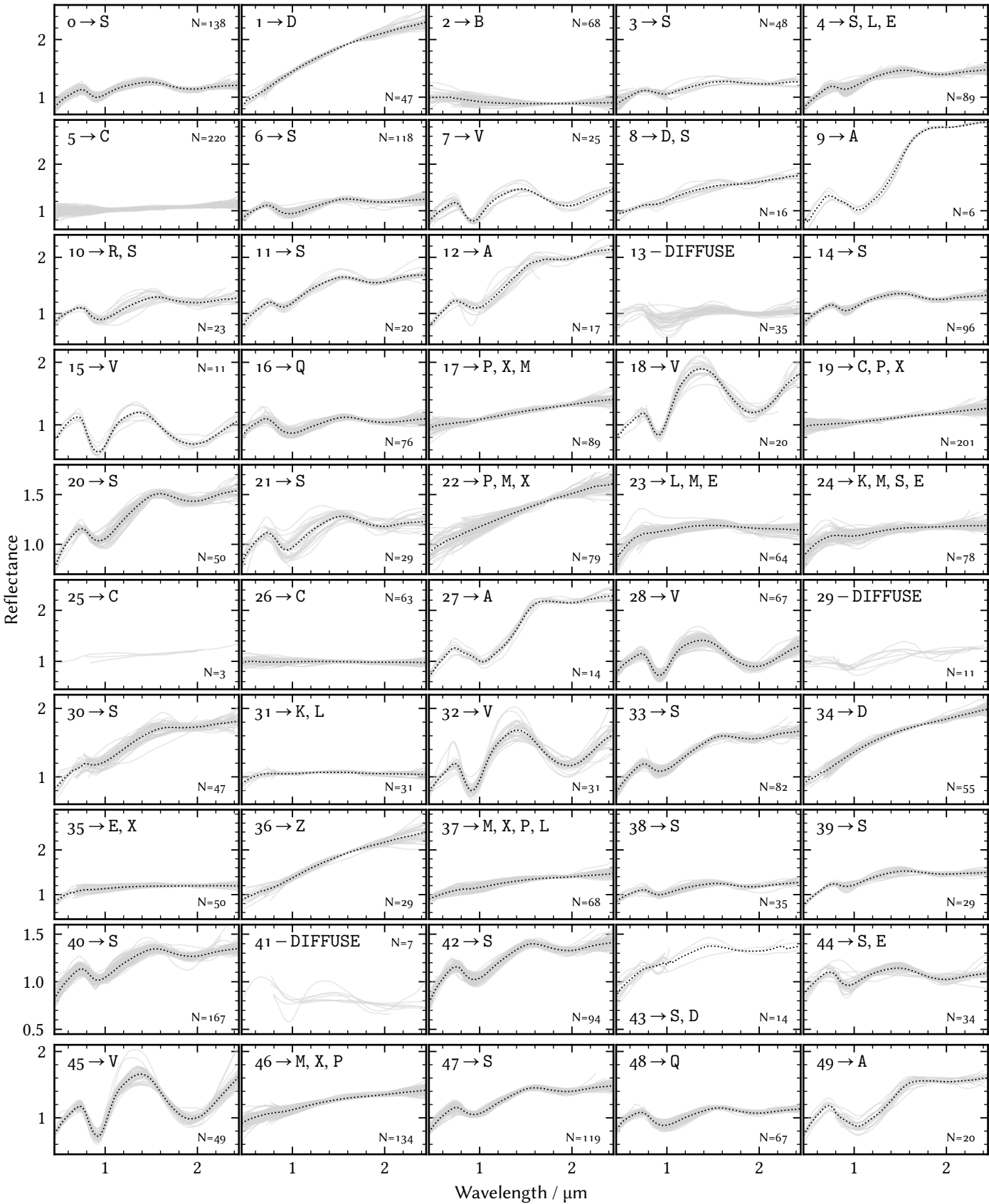


Figure 10.5: Overview of the 2983 reflectance spectra assigned to one of 50 Gaussian clusters. The number N of spectra and the asteroid classes to which the cluster contributes is given for each cluster. Classes with fewer than three observations contributes from the cluster are not given except for cluster 25, which only has three members. The classes are sorted by the total number of observations the cluster contributes. The dotted line gives the mean spectrum per cluster except for the diffuse clusters and cluster 25. The mean spectra are normalised to unity at $0.55 \mu\text{m}$. Note the change in y-axis limits per row. Published in Mählke et al. (2022).

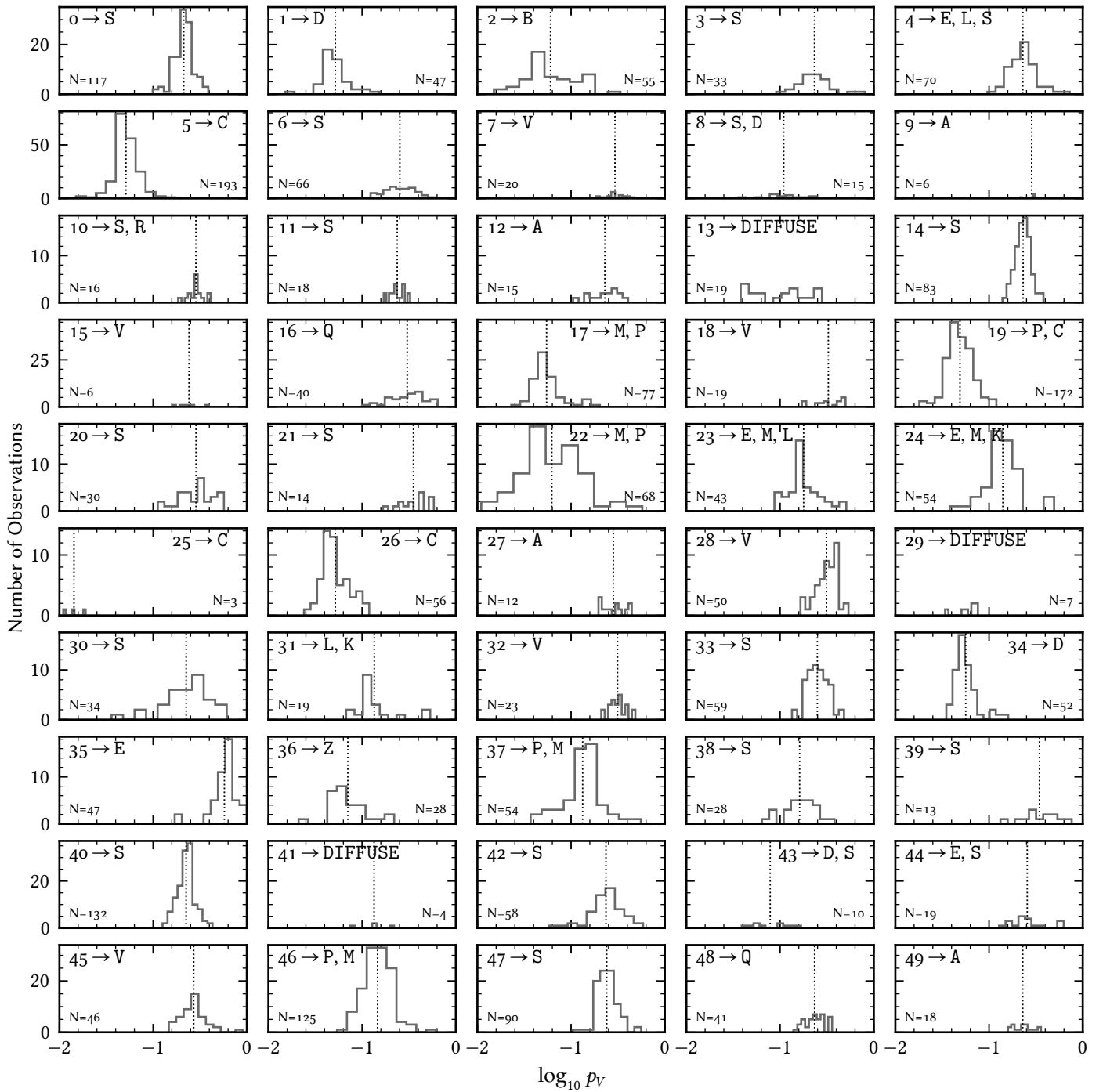


Figure 10.6: As Fig. 10.5, but showing the distribution of the 2324 visual albedos over the 50 Gaussian clusters. Published in Mahlke et al. (2022).

The baseline of the taxonomic scheme is the division between the C and the S complex, as it has been since Zellner (1973) and Chapman et al. (1975). A smaller population with fainter features than S-types yet stronger ones than C-types occupies the space between these complexes in DeMeo et al. (2009). This population has grown considerably thanks to dedicated observational efforts (Ockert-Bell et al. 2010; Neeley et al. 2014; Shepard et al. 2015), and I acknowledge it as third complex of this taxonomy, denoted as M-complex after its most populous class.

The taxonomic constants A- and V-types are readily identified. More challenging is the separation of the Q- from the S-types due to their continuous transition in slope behaviour. The same holds for the subclasses Sa, Sq, Sr, and Sv of the Bus-DeMeo system. The S-complex is the most populated complex in this taxonomy and is divided into multiple clusters. Labelling all of them with letters denoting different subclasses would lead to a large degree of confusion. Instead, as outlined in Chapter 8, I aim to provide a mean to mineralogically interpret of the variability of the complex in Chapter 11.

10.3.2 Probabilistic Approach

Most cluster occupy a narrow volume in the latent space and contain observations which show little variation in terms of their reflectance spectra or visual albedos. For example, clusters 0, 3, and 6 are quickly identified to contain S-type asteroids based on their reflectance spectra, refer to Fig. 10.5. 33 of the 50 clusters can be mapped in a straight-forward manner to a single asteroid class in this manner. For all observations in the data used to derive the taxonomy and for those which are later classified in this scheme, the probability to belong to any of these 33 clusters is directly added to the probability to belong to the associated class. For example, the probability of a given observation to belong to cluster 0, 3, or 6 all add to the overall probability of being an S-type. These mappings are shown in the overview figures Figs. 10.5 and 10.6 and in the summary Table 10.1.

Less straight-forward are clusters which can be summarised in two different scenarios. The first are clusters which encompass Gaussian trends in the input observations which are typically divided into separate asteroid classes. One example are C- and P-types in cluster 19, both of which are featureless and neutral- to red-sloped. Once mapped into the latent space, their reflectance spectra are a continuous population with predominant variance along z_1 . Here, the principle of deriving an objective taxonomy clashes with the need for class continuity. Putting the P-type prototypes (65) *Cybele* and (87) *Sylvia* to the C-types would have been justified based on their spectral appearance yet meant a significant cut in the class continuity. These type of clusters are common as there are often continuous trends between defined classes, e. g. the K- and L-types in cluster 31, or for clusters of the X complex, which are discussed below. For these clusters, a secondary decision step is required, for which I mostly use GMMs in a subspace of the latent space, fitted to a number of components equal to the number of classes that may result from the cluster. These steps are discussed for each class in Chapter 11. For any observation classified in the scheme, the probability to belong to this

Table 10.1: The cluster-to-class decision-tree. For each observation, the probability to belong to any of the clusters in the upper part is converted entirely to the respective class given on the right hand side. The probability to belong to any of the clusters in the lower part is split over different classes based on the decision given in the middle column. GMM refers to a Gaussian Mixture Model with a number of mixture components equal to the number of possible outcome classes, fitted to the latent scores given in the parentheses. p_x refers to the probability to belong to class or cluster x , evaluated in the latent scores given in the parentheses. The feature recognition in the bottom part is the final step of the decision-tree. Adapted from Mahlke et al. (2022).

Cluster		Class
0, 3, 6, 11, 14, 20, 21, 30, 33, 38, 39, 40, 42, 47	→	S
1, 34	→	D
2	→	B
5, 25, 26	→	C
7, 15, 18, 28, 32, 45	→	V
9, 12, 27, 49	→	A
16, 48	→	Q
36	→	Z
4	$P_{23}(z_3, z_4)/P_{40}(z_3, z_4)$	L, S
8, 43	$GMM(z_2, z_4)$	D, S
10	$GMM(z_1, z_2)$	R, S
13	$GMM(z_2, z_4)$	C, O, Q
17, 22, 35, 37, 46	$P_E(p_V)/P_M(p_V)/P_P(p_V)$	E, M, P, X
19	$GMM(z_1, z_4)$	C, P
23	$GMM(z_1, z_4)$	L, M
24	$GMM(z_2, z_3)$	K, M
29	$GMM(z_1, z_2)$	A, B, C, D, M, P, S, Q, V
31	$GMM(z_3, z_4)$	K, L
37	$GMM(z_2, z_4)$	L, M
41	$GMM(z_1, z_2)$	B, V
44	$P_E(p_V)/P_M(p_V)$	E, S
Class is B, C, P, or X and h-feature is present		Ch
Class is E, M, P, or X and e-/k-feature is present		+e/+k

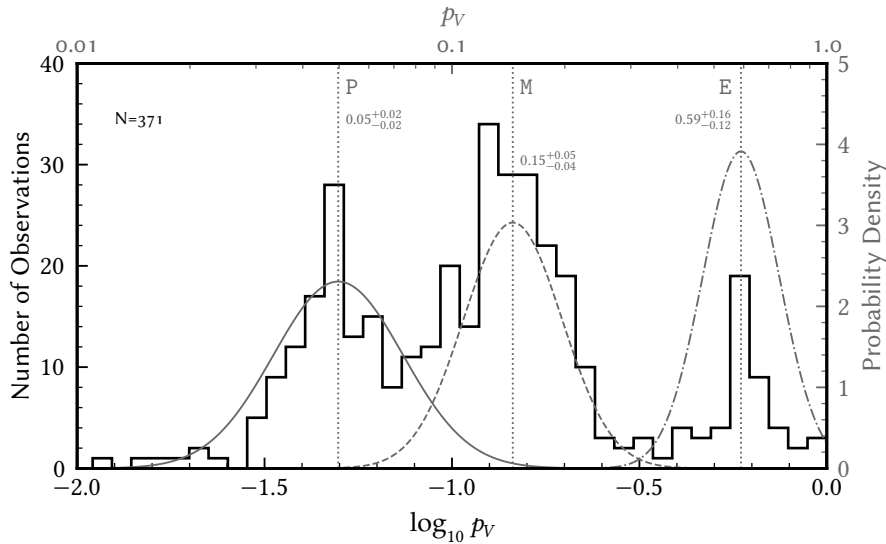


Figure 10.7: The distribution of visual albedo in clusters associated to X-complex. The spectral degeneracy of the E-, M-, and P-classes is resolved by fitting a three-component Gaussian Mixture Model (grey lines) to the joint albedo distribution (black) of clusters 17, 22, 35, 37, and 46, consisting of 371 observations in total. The vertical dotted lines give the mean visual albedo of the fitted mixture components and are labelled by the assigned class designations P, M, E, in order of increasing visual albedo. The mean and standard deviation of each mixture component is given below the class letter. I note that these values do not represent the final visual albedo distributions per class as observations from other clusters may be assigned to these classes. Published in Mahlke et al. (2022).

cluster is split using the secondary decision step into over the probabilities to belong to the possible outcome classes.

The second type of diverse clusters are typically low-number and large-volume clusters. They capture the low-density noise around the Gaussian core clusters in the latent space and thus contain different reflectance and albedo characteristics. These few, often dubious samples are best described (i. e. they have the highest likelihood) by a Gaussian with a large covariance. There is three of the diffuse cluster, 13, 19, and 29, and they are resolved like the other type of diverse cluster. However, the classifications are flagged in the output using the DIFFUSE flag, and the observations assigned to them should always undergo visual scrutiny as they might contain spectra of dubious quality.

10.3.3 Dissolving the X-complex

Untangling the degeneracy of the X-complex into its constituents E, M, and P is one of the main goals of this revision of the asteroid taxonomy. There are several cluster which, based on the distribution of their reflectance spectra, are associated to the X-complex, namely clusters 17, 22, 35, 37, and 46, refer to Fig. 10.5.

To resolve the X-complex, I first fit the distribution of visual albedos of all observations in the aforementioned clusters using a three-component GMM. The result is shown in Fig. 10.7 and has the familiar three-peaks structure. Using the fitted components, I construct the secondary decision-tree steps to divide the probability to belong to either of the X-complex structures into the probability to belong to E, M, or P using the visual albedo of the input observation. The letter X is retained for observations in these clusters without visual albedo information, following Tholen (1984).

10.3.4 Feature Flags

The Bus-DeMeo taxonomy is a *feature-based* taxonomy, meaning that the authors placed a large emphasis on the presence and characteristics of features to define the taxonomic classes. Four classes in particular are largely based on the presence of different spectral features (Bus and Binzel 2002a; DeMeo et al. 2009; Binzel et al. 2019):

- Ch The class members present a feature at $0.7\ \mu\text{m}$ associated to phyllosilicates and a history of aqueous alteration (Vilas and Gaffey 1989). An example is shown for (13) *Egeria* in Fig. 10.8.
- Xk The class members present a broad and shallow feature around $0.8\ \mu\text{m}$ to $1.0\ \mu\text{m}$. It may appear in conjunction with a weak feature around $1.9\ \mu\text{m}$, the latter though is never present without the former (Neeley et al. 2014). An example is given with (110) *Lydia* in Fig. 10.8. Hardersen et al. (2005) argue that these features indicate magnesium-iron silicates.
- Xe The class members present a narrow feature around $0.5\ \mu\text{m}$, as shown for (64) *Angelina* in Fig. 10.8. It is associated to oldhamite (Watters and Prinz 1979) in aubrites and is found in E-type asteroids (Gaffey and Kelley 2004). Shestopalov et al. (2010) further suggest titanium-bearing pyroxene as origin of the band.
- Xn The class members present a narrow feature around $0.9\ \mu\text{m}$. The most prominent example is present in the spectrum of the archetype of the E class (44) *Nysa*, refer to Fig. 10.8 and Binzel et al. (2019). Gaffey et al. (1989) relate it to low-iron enstatite pyroxene.

In the following, I refer to the features by the letter denoting the subclass, i. e. the e-, h-, k-, and n-feature.

The presence of features in the reflectance spectra of the generally featureless C- or X-complex members demands a distinction in the class assignment. I thus carry over this notation from the Bus-DeMeo system with two modifications.

First, I do not differentiate between the k- and the n-feature. This follows a practical issue of reliably differentiating the features. After parametrisation of the $0.9\ \mu\text{m}$ features in the reflectance spectra of C- and X-complex members, I did not identify two distinct groups. I do not rule out the possibility of different mineralogies, however, making a distinction with two letters here may communicate a certainty in the classification and mineralogical

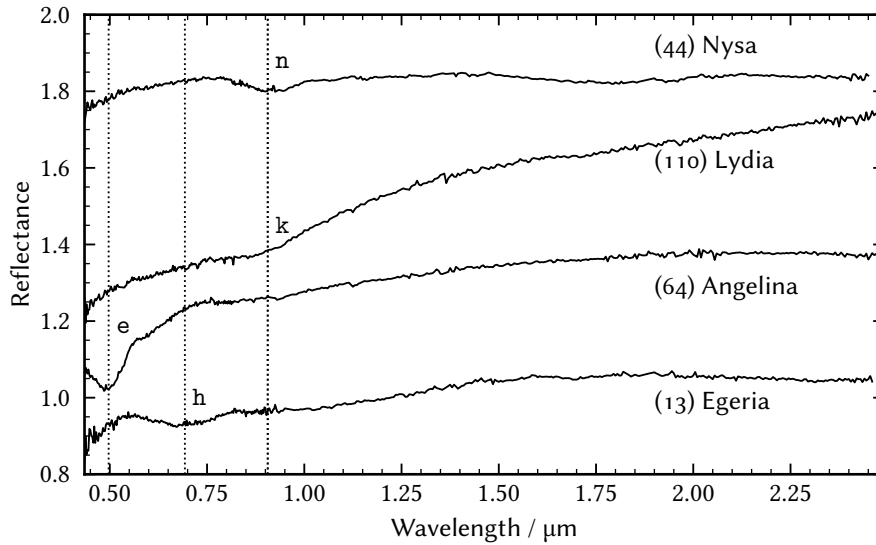


Figure 10.8: Example spectra depicting the e-, h-, and k-features. The vertical dotted lines indicate the mean band centers (e: 0.50 μm , h: 0.69 μm , k: 0.91 μm) from a set of spectra where the features were visually identified. Further shown is the n-feature defined in Binzel et al. (2019) in the spectrum of (44) *Nysa*. Adapted from Mahlke et al. (2022). Data from SMASS.

interpretation which is not given in the spectra. Thus, I denote both the k- and the n-feature as k.

Second, the e- and, in particular, the k-feature appear in reflectance spectra with a variety of slopes and overall appearances, further shown in Chapter 11. It is therefore not appropriate to define a single class based on the presence of the features following the Xe and Xk definitions. Instead, I propose the use of the e- and k-letters as “feature flags”, to be appended to the major class designation of the respective object, which are the X-complex members E, M, and P (and X in case of lacking albedo information). Furthermore, the features can appear together. Class designations such as Me, Eek, Pk are therefore possible.

Consistent with the previous taxonomy, I keep the Ch-class as unique and proper class. The h-feature appears consistently only in spectra which are associated to the C-complex.

The feature recognition is the last step of the classification decision-tree, refer to Table 10.1. An algorithm for the feature detection and parametrisation is presented in Chapter 12.

10.3.5 Feature Parametrisation

The visual identification of absorption features is the most reliable one. However, as we move towards larger number of reflectance spectra to be taxonomically classified, an automated feature detection and parametrisation routine may be preferred over the manual approach.

Table 10.2: The parameters for the automated feature recognition procedure are given for each feature.

Feature	Center / μm	Lower Limit / μm	Upper Limit / μm
e	0.50 ± 0.01	0.450	0.539
h	0.69 ± 0.01	0.549	0.834
k	0.91 ± 0.02	0.758	1.060

The identification and parametrisation of the e, h, and k features is achieved in a semi-automated manner. For each feature, I visually identify a set of spectra which present it and measure the centre wavelengths using a polynomial fit of degree four, following Fornasier et al. (1999). I further record the upper and lower wavelength of the fitted region which visually gives the best fit of the feature. The mean parameters of each feature are given in Table 10.2.

For each feature, I determine the band depth with respect to the spectral continuum and the uncertainties of the reflectance values using the polynomial fit. The uncertainties are defined as the standard deviation of the residuals of the polynomial fit to the spectrum. The fit is computed over the mean upper and lower limits derived from the training set of the spectra. The feature is considered to be present if the derived band centre is within three standard deviations of the mean of the band centres from the training set and if the ratio of the band depth to the uncertainties (the signal-to-noise ratio) is larger than one.

This automatic fitting procedure is executed for the e- and k-feature for spectra assigned to classes of the former X-complex: E, M, P, and X. The h-feature is fit for members of the C-complex (B, C, P, refer to Chapter 11) as well as X-types as they may be P-types. In practice, I find that it results in many false-positives as the signal-to-noise threshold is set rather low. However, increasing the threshold rejects a lot of true-positives. I therefore propose a semi-automated approach where the classification tool presented in Chapter 12 identifies the features automatically and the user is prompted for visual confirmation.

Both the e- and k-feature may be artificial. The k-feature is particularly vulnerable to being artificial, as it lies in the transition of the visible and the NIR spectral region and may be due to stitching artifacts (Clark et al. 2009). (Bus and Binzel 2002b) point out that some SMASS spectra have artificial features around $0.515 \mu\text{m}$ – $0.535 \mu\text{m}$. The e-feature should only be considered as present if the central wavelength is below this region, around $0.50 \mu\text{m}$.

10.3.6 Class per Asteroid

A unique aspect of this taxonomy which has not been addressed so far is the fact that some asteroids are represented up to five times with different reflectance spectra and visual albedos in the input observations used to train the MCFA model. The variability of the spectral appearance of asteroids between observations is noted by Carvano et al. (2010) in their SDSS colours,

and visual inspection confirms this behaviour for the reflectance spectra at hand, both in terms of slope and even in terms of features.

A total of 549 of the 2125 asteroids have more than one observation, opening up the possibility that they have multiple, distinct classifications. This variability might be valuable information assuming that it is compositional in nature (e. g. variability in degree of space weathering on (832) *Karin*, Sasaki et al. (2004), though challenged by (Chapman et al. 2007)) rather than instrumental (e. g. the effect of selecting different standard star, Marsset et al. (2020)), however, I also recognise the convenience of a catalogue with a single taxonomic classification per asteroid. I provide this classification by resolving the case of multiple, distinct classifications in the following manner. For each classification of an individual asteroid, I sum the class probabilities weighted by the fraction of the completeness of the observation. If the albedo is observed, the observation gets an additional weight corresponding to 25 data dimensions. This means that a visible-only spectrum plus albedo has approximately equal weight as a VisNIR spectrum without albedo. The final class of the asteroid is the most probable class resulting from the weighted sum. If any observation carries any of the e-, h-, or k-feature flags, the final class is altered accordingly.

The number of observations and asteroids per taxonomic class and their orbital distributions are given in Table 10.3.

10.4 DATA AVAILABILITY

The class and cluster probabilities per observation and the final classifications per observation and per asteroid are available in the CDS¹ and GitHub² repositories. The mean and standard deviation of the reflectance spectra and visual albedos for each class (“class templates”) are further made available.

¹ <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/665/A26>

² <https://github.com/maxmahlke/classy>

Table 10.3: Number of observations (Obs.) and asteroids (Ast.) per taxonomic class and per orbital class. The fraction in percent of asteroids per class in this work is compared to the one in DeMeo et al. (2009). The fraction in the DM09 column does not add up to 100 % due to the removal of the T class in this work. The orbital classes use the following abbreviations: NEO - Near-Earth Object, MC - Mars-Crosser, H - Hungarians, IMB - Inner Main Belt, MMB - Middle Main Belt, OMB - Outer Main Belt, Cyb - Cybeles, JT - Jovian Trojan. Published in Mahlke et al. (2022).

Class	Obs.	Ast.	Fraction		Orbital Class								
			Here	DM09	NEO	MC	H	IMB	MMB	OMB	Cyb	Hildas	JT
A	57	32	1.5	1.6	2	3	2	7	10	8	-	-	-
B	68	45	2.1	1.1	15	4	1	12	5	8	-	-	-
C	299	221	10.4	7.3	69	8	2	89	72	79	2	2	5
Ch	144	107	5.0	4.8	9	2	-	20	47	26	2	-	1
D	119	82	3.9	4.3	6	1	-	1	4	5	5	16	44
E	65	46	2.2	-	7	4	27	4	3	1	-	-	-
K	59	42	2.0	4.3	21	2	-	5	2	12	-	-	-
L	76	58	2.7	5.9	20	4	3	4	22	3	-	-	2
M	252	142	6.7	-	29	7	2	17	47	28	-	2	10
O	4	2	0.1	0.3	-	-	-	-	1	1	-	-	-
P	195	135	6.4	-	14	6	1	11	26	36	12	12	17
Q	158	107	5.0	2.2	89	5	-	7	4	2	-	-	-
R	15	10	0.5	0.3	7	-	-	2	-	1	-	-	-
S	1188	898	42.3	53.8	404	101	35	140	172	45	-	1	-
V	206	142	6.7	4.6	28	2	-	104	4	4	-	-	-
X	50	33	1.6	8.6	20	8	2	1	-	2	-	-	-
Z	28	23	1.1	-	1	-	1	4	6	3	-	1	7
Σ	2983	2125	100	98.9	741	157	76	428	425	264	21	34	86

CLASSES

11.1	C-Complex: B, C, Ch, P	127
11.1.1	B-Types	129
11.1.2	C-Types	129
11.1.3	Ch-Types	131
11.1.4	P-Types	131
11.1.5	Endmembers: D, Z	132
	11.1.5.1 D-Types	132
	11.1.5.2 Z-Types	132
11.2	M-Complex: K, L, M	133
11.2.1	K-Types	133
11.2.2	L-Types	135
11.2.3	M-Types	136
11.2.4	Endmembers: E	137
11.3	S-Complex: S, Q	138
11.3.1	S-Types	139
11.3.2	Q-Types	141
11.3.3	Endmembers: A, O, R, V	141
	11.3.3.1 A-Types	142
	11.3.3.2 O-Types	142
	11.3.3.3 R-Types	143
	11.3.3.4 V-Types	143

In this chapter, I describe the 17 classes defined in the taxonomy, divided into three complexes: C, M, and S. I focus on their their properties in data- and latent space and refer to [Chapter 4](#) for mineralogical and meteoritic interpretations found in the literature. I further highlight differences in the definitions with respect to the previous taxonomies, in particular, the Tholen (1984) and the Bus-DeMeo systems (Bus and Binzel 2002a; DeMeo et al. 2009). The class discussions conclude with explicit mentions of the cluster-to-class decision-tree branches leading to their assignment, of which an overview is given in [Table 10.1](#). [Table 11.1](#) provides a single-page overview of all classes, including their main properties and prototype spectra.

The X-types are not discussed separately as they are E-, M-, or P-types lacking visual albedo information. The class descriptions largely follow the text presented in Mahlke et al. (2022).

11.1 C-COMPLEX: B, C, CH, P

The spectral appearance of the C-complex members is diverse in terms of the slope, ranging from overall blue to neutral to extremely red. They are generally featureless except for signs of aqueous alteration in form of the h-feature at 0.7 μm attributed to phyllosilicates and a wide 1.3 μm -feature

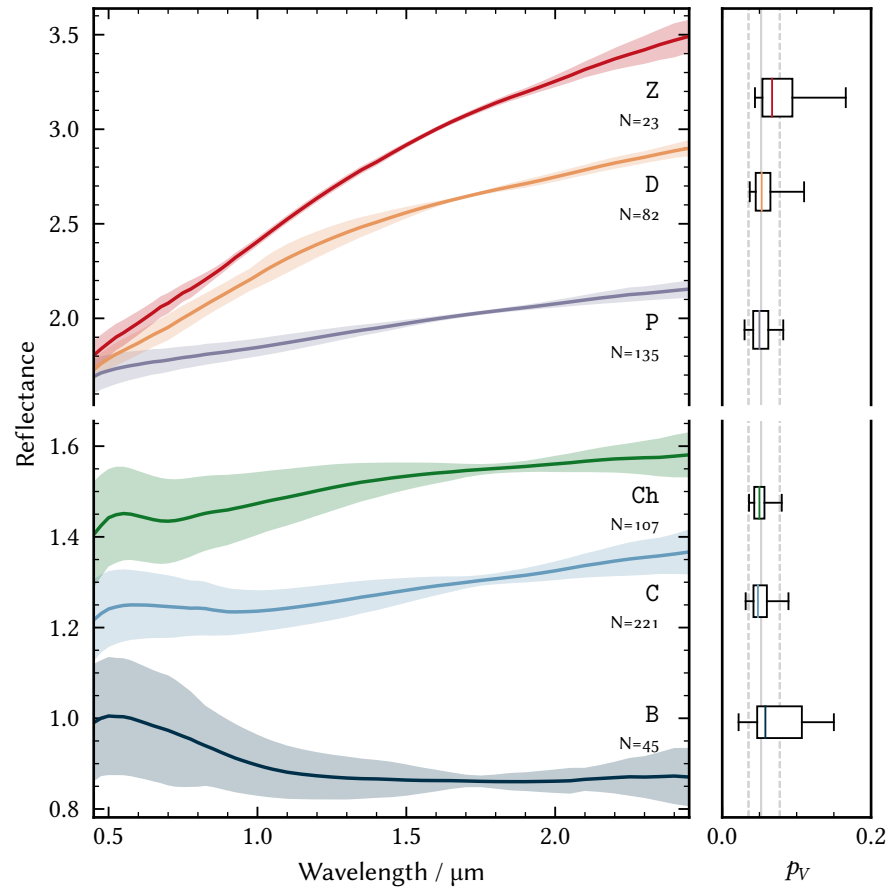


Figure 11.1: Distribution of observables of classes and endmembers of the C complex
Left: The mean (solid) and standard deviation (shaded) of the reflectance spectra are shown for each class, shifted along the y-axis for comparability. Note the change in the y-scale between Ch and P. The number N of asteroids in each class is given below the respective letter. *Right:* The median (solid), lower and upper quartiles (box) and the 5th and 95th percentiles (whiskers) of the distribution of visual albedos per class are given for each class. The grey lines give the mean (solid) and standard deviation (dashed) of the whole complex ($0.05^{+0.03}_{-0.02}$).

indicative of magnetite, refer to [Chapter 3](#). The $1.3\ \mu\text{m}$ band may lead to an overall concave shape, though convex spectra are more common.

Based on the distribution of the asteroid observations in the latent space as outlined in [Chapter 10](#), I define six classes for the C-complex: B, C, Ch, P, D, and Z. The distribution of spectra and visual albedo for all samples assigned to each class is shown in [Fig. 11.1](#). The most notable differences to the previous systems is the introduction of the Z-type and the placement of the P-class into the C-complex. The slope-based classes Cb, Cg, and Cgh from the Bus-DeMeo system are dropped due to the continuous trend in slopes observed in the data.

The distribution of the classes is shown in a subspace of the latent space in [Fig. 11.2](#). The apparent diagonal gaps between the C- and Ch-types in the lower part of [Fig. 11.2](#) are an unfortunate artefact of the spectral normalisation (refer to [Chapter 9](#)) which organises spectra in clusters prior to the dimensionality

reduction. A large number of asteroids have observations on either side of the gaps.

11.1.1 B-Types

Figure 11.1 shows that both interpretations of the mnemonic of the B-types are compatible with the class definition in this taxonomy: they have a blue spectral slope longwards of $0.5\ \mu\text{m}$ – $0.6\ \mu\text{m}$ and visual albedos above the average of the C-complex, reaching up to 10 % and higher. However, there are also dark B-types, which are recognised as F-types in Tholen (1984) based on their flat UV spectral appearance and lower albedos than the B-types. Indeed, in the upper part of Fig. 11.2, we see a trend of decreasing albedo with increasing slope in the B-types based on their z_1 and z_2 scores, i. e. bluer B-types are brighter. However, as the UV region is not accounted for in the taxonomy, I do not attempt to split the B-types.

Figure 11.2 further shows that B-types exhibit a large dispersion compared to the other complex members. The continuity between B and C noted by Tholen (1984) is also apparent in this work, in particular the lower end of the distribution in z_1 may also represent C-type asteroids. I note that the *Themis*-like asteroids with a neutral-to-red slope in the NIR described in Clark et al. (2010) and León et al. (2012) as B-types are C-types here, in agreement with their classification in the Bus-DeMeo system (see subpanel (h) in Fig. 11.2).

Prototypes of the B class are the archetype (2) *Pallas*, NEO (3200) *Phaethon*, and *Pallas*-family member (531) *Zerlina*, highlighted in the subpanels (a) and (f) in Fig. 11.2. A total of 45 asteroids (2.1 %) are classified as B-types, refer to Table 10.3. The B-class consists of a single cluster (2).

11.1.2 C-Types

The reflectance spectra of the carbonaceous C-types are neutral to slightly red-sloped and generally featureless except for a broad feature around $1.3\ \mu\text{m}$ as shown in Fig. 11.1. The albedo distribution is narrow and centred around 5 %. As such, the dispersion of the class members in z_1 - z_2 shown in the upper part of Fig. 11.2 is largely given along z_1 . In the lower part of the same figure, a positive correlation between the scores in z_1 and z_4 is observable. The latter is proportional to the concavity of the reflectance spectra, indicating that C-types on average get more concave as they get redder. Apart from the artificial separation of the class members in the z_1 - z_4 projection, no clear distinctions appear and I refrain from defining analogues to the classes Cb, Cg, and Cgh in the Bus-DeMeo system.

The class archetypes (1) *Ceres* and (10) *Hygiea* are highlighted in subpanel (b) of Fig. 11.2. In subpanel (h), I highlight (24) *Themis*, (45) *Eugenia*, and (52) *Europa*, which, according to Vernazza et al. (2015), have an IDP-like composition.

A total of 221 asteroids (10.4 %) are classified as C-types in this taxonomy. C-types are present in three clusters (5, 19, 26), including the two largest of the 50 clusters in the model.

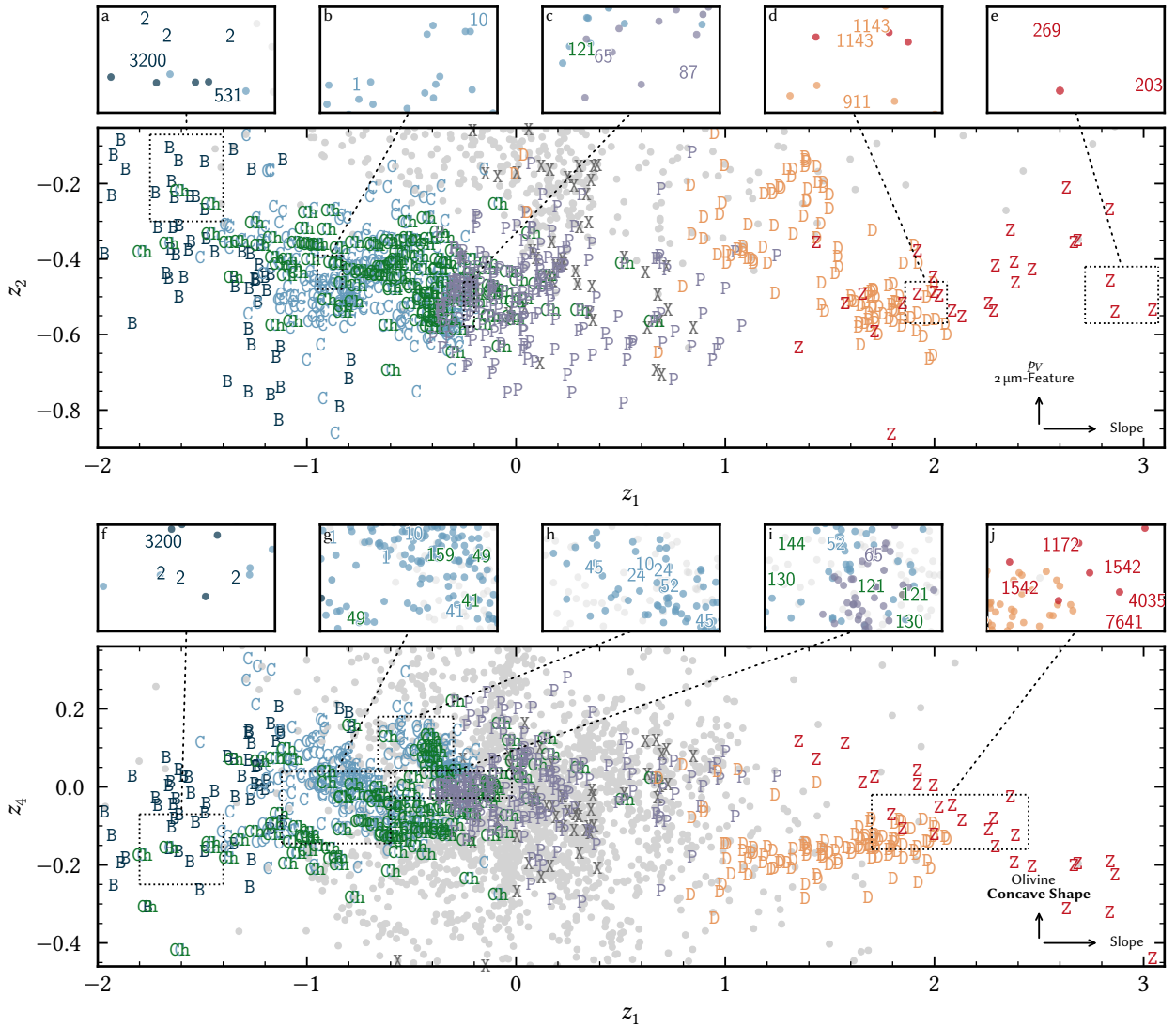


Figure 11.2: Distribution of C-complex classes and endmembers in latent space. The observations assigned to each class are indicated with the respective class letter. The latent scores of observations outside these classes are shown in grey. Some outliers in z_2 and z_4 are not shown for readability. The subpanels highlight asteroids-of-interest by replacing the class letter with the respective asteroid's number. If more than one spectrum of an asteroid is in the observations, its number may appear more than once.

Apart from the prominent C-types (45) *Eugenia* and (52) *Europa*, cluster 19 contains prominent the P-types (65) *Cybele* and (87) *Sylvia*, shown in sub-panel (c) in Fig. 11.2. The cluster overall resembles the Cb-class in terms of the spectral appearance. I thus divide this cluster between C and P using a two-component GMM in z_1 - z_4 . Post-clustering steps like this are a concession to class continuity and I generally avoid them. However, though the data associates (65) *Cybele* and (87) *Sylvia* to the C-types, both are generally accepted as P-types in the literature (Tholen and Barucci 1989; DeMeo et al. 2009; Fornasier et al. 2014).

11.1.3 Ch-Types

In comparison to C-types, members of the Ch class exhibit the 0.7 μm -feature and a generally more convex spectral shape. The distribution of visual albedos is similar for both classes, (Fig. 11.1). As outlined in Chapter 10, the Ch class is the only continued feature-based class from the Bus-DeMeo system, and the narrow distribution of albedos and reflectance spectra of the class members shown in Fig. 11.1 confirms this decision.

In latent space, Ch-types are found throughout the C-complex, with an apparent concentration on the region where C- and P-types overlap. In z_4 , Ch-types have show lower scores than C-types, corresponding to a flat rather than a concave spectra. In subpanels (g) and (i) of Fig. 11.2, I highlight the positions of (41) *Daphne*, (49) *Pales*, (121) *Hermione* (144) *Vibilia*, and (159) *Aemilia*, all of which are associated to CM chondrite Vernazza et al. (2015). (130) *Elektra* is further linked to these objects based on its density (Carry 2012; Yang et al. 2016; Hanuš et al. 2017).

The 0.7 μm h-feature is present in at least one spectrum of 107 asteroids (5.0%). Ch-types are present in clusters 2, 5, 17, 19, and 26. The assignment requires the identification of the 0.7 μm h-feature. Within the C-complex, 20.4% of samples present the h-feature. The actual number is likely higher as 12.1% of samples in the C-complex are NIR-only spectra, such as the of (41) *Daphne* indicated as C-type in subpanel (g) of Fig. 11.2. Rivkin (2012) estimated that about one third of all C-complex asteroids are hydrated.

11.1.4 P-Types

This is the first definition of the P-types in terms of their VisNIR reflectance spectra in combination with their visual albedo. The spectra are generally red and featureless. The visual albedos are comparable to those of C-types and Ch-types those they extend to lower values, as apparent in Fig. 11.1.

The distribution of the P-types in the latent space shown in Fig. 11.2 continues the trend of continuity among the C-complex members. A diffuse branch of P-types further extends towards the steeply-red spectral region of the D-types.

As members of the former X-complex, the e- and k-features introduced in Chapter 10 are searched for in reflectance spectra of P-type asteroids. I find that no asteroid assigned to P exhibits either feature. Three samples assigned to P show the e-feature, yet they are of asteroids which are later

assigned to the M class: (4660) *Nereus* and (5645) *1990 SP*. On the other hand, as new member of the C-complex, the spectra are further investigated for the presence of the h-feature, which is present in 19.2 % of samples in the P class.

A total of 135 asteroids (6.4 %) are classified as P-types in this work. The class P consists primarily of clusters 17, 19, and 22, where cluster 19 contains member of class C and is resolved using the two-component GMM described above. Clusters 17 and 22 contain M-type asteroids and are both part of the decision-tree branch which resolves the X-complex, outlined in [Chapter 10](#).

11.1.5 Endmembers: D, Z

Although some P-types bridge the region between the core of the C-complex and the D- and Z-types ([Fig. 11.2](#)), I refer to the latter two as endmember classes due to their large separation from B, C, Ch, and most P.

11.1.5.1 D-Types

The D-types exhibit red, featureless spectra with an overall convex shape, refer to [Fig. 11.1](#). They are redder than the P-types and in previous taxonomies were the reddest featureless asteroids. Most D-types are located beyond the OMB among the JT population and, to a minor extent, among the Cybeles and Hildas (DeMeo and Carry [2013, 2014](#)). The orbital distribution is further regarded in [Chapter 13](#).

The remarkably homogeneous nature of the D in reflectance spectra and albedos is well established (DeMeo and Carry [2013](#); Emery et al. [2015](#)) and it is apparent in their latent-space distribution in [Fig. 11.2](#). In subpanel (d) of the figure, I highlight the position of the class prototypes (911) *Agamemnon* and (1143) *Odysseus*. In the z_1 - z_2 projection, D-types are split into a red cluster and an even redder one. However, I could not find systematic differences in the visual albedos or spectral appearance, and I therefore interpret this split as an artefact of the normalisation. (2246) *Bowell* and (2674) *Pandarus* are present in both clusters.

A total of 82 asteroids (3.9 %) are classified as D-types in this work. Two clusters contain the majority of D-types, the homogeneous main cluster 1 and a more diffuse cluster 34. The latter contains interlopers of classes P and M. Furthermore, there are two small clusters of D- and S-types. 16 VisNIR spectra of D-types and strongly-sloped S-types make up cluster 8 and are separated using a two-component GMM in z_2 - z_4 , where the feature-rich S-types have higher scores in z_2 . Cluster 43 contains 14 visible-only spectra of D-types and S-types, which I separate in the same way as in cluster 8.

11.1.5.2 Z-Types

Throughout the numerous trial-and-error approaches to the clustering, one of the few constants was cluster with a small number of featureless and extremely red asteroids, extending from the homogeneous D-types towards larger z_1 -values. These objects build the new class Z. While in z_1 they form a continuous branch of the D-types, the lower part of [Fig. 11.2](#) shows that

the covariance structure in z_1 - z_4 is noticeably distinct between the classes. Z-types show a clear trend towards a more convex shape with increasing slope. The class prototypes are (203) *Pompeja* and (269) *Justitia*, shown in subpanel (e) of Fig. 11.2. The distribution of visual albedos is skewed towards larger values for Z in comparison to D, refer to Fig. 11.1. The final evidence in support of a new class is the orbital distribution of these objects, which, unlike D-types, are located almost entirely within the Main Belt. This distribution and previous identifications of D-type material in the Main Belt is discussed in to Chapter 13.

A total of 23 asteroids (1.1 %) are classified as Z-types in this work. They are grouped exclusively in cluster 36. A certain number of D-type interlopers is likely present in this class, indicated by the overlap in the latent space. In subpanel (j) of Fig. 11.2, I highlight the Trojan asteroids (1172) *Aneas*, (1542) *Schalen*, (4035) *Thestor*, and (7641) *Cteatus*, which spectrally match D-types.

The class letter Z is proposed following Mueller et al. (1992). The class prototype could thus be (5145)Pholus, however, as Centaur, it is not included in this taxonomy.

11.2 M-COMPLEX: K, L, M

Asteroids of the M-complex are located between the C and S complexes in terms of their reflectance spectra and visual albedos. The spectra are neutral to red-sloped and show weak features around 1 μm or 2 μm , as shown in Fig. 11.3. In albedo, K, L, and M show values between 10 %–20 %, while E-types present values generally above 50 %, which is the most distinct identification feature aside from the spectral appearance of V and A discussed further below.

The M-complex is a new taxonomic complex yet may be regarded as the successor of the X-complex. In comparison to the latter, as defined in Tholen (1984), the P class is removed and K and L are added. The complex constituents are thus K, L, M, and E. Even before describing these classes in more detail, I mention that the division of the M-complex was the most challenging part of this taxonomy and did not lead to a fully satisfactory conclusion. I do not identify a clear cut between K, L, or M, and I rely on information from the literature for all the classes to assign class boundaries which aid in the discussion. The spectral appearance in particular of E-, K-, and L-types is degenerate, and M-types with neutral NIR slope (X_c in the Bus-DeMeo system) with the k-feature may be classified as K (refer to Fig. 11.4). Apart from E-types, the classes do not differ in their visual albedos. These difficulties are not new, especially member of K and L are frequently reclassified (Shepard et al. 2008b; León et al. 2012; Binzel et al. 2019). However, this work shows that they are not restricted to single objects but present on the large scale as well.

11.2.1 K-Types

K-types have a neutrally sloped spectrum with a deep 1 μm -band and visual albedos around 13 %, refer to Fig. 11.3. In comparison to the remaining M-complex classes, they have low values in z_1 and large values in z_4 due to their reflectance spectra, shown in Fig. 11.1.

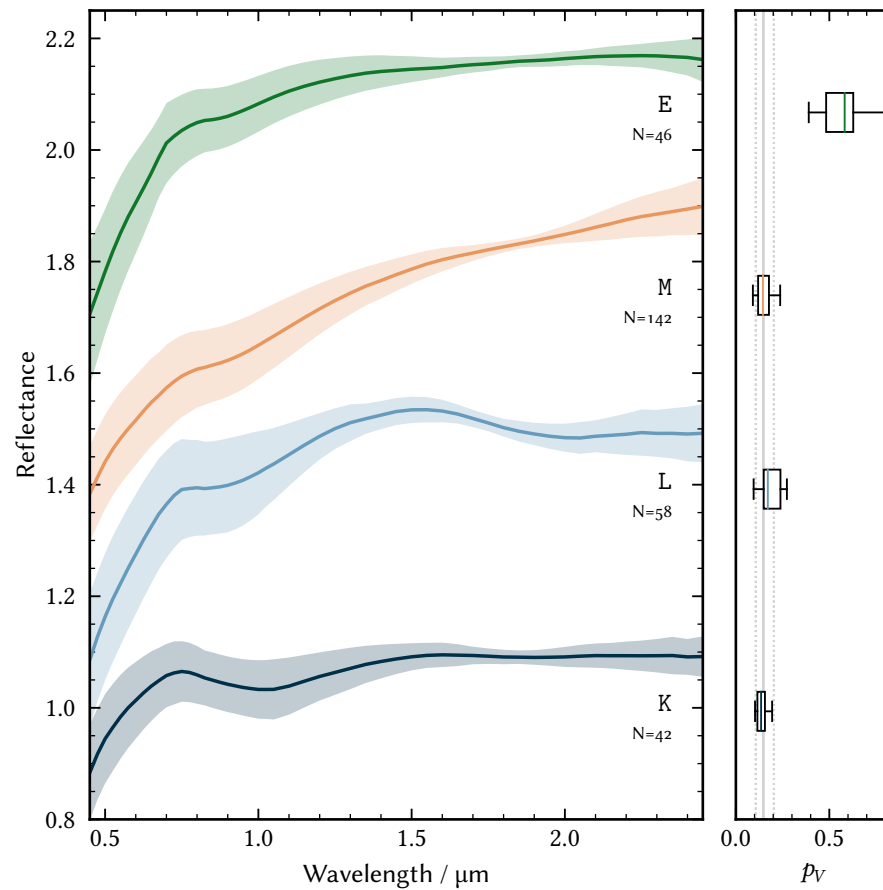


Figure 11.3: As Fig. 11.1, but showing the data space properties of classes of the M-complex. The E-types are excluded from the computation of the mean and standard deviation of the visual albedo distribution of the complex ($0.15^{+0.06}_{-0.04}$).

The K-types are associated largely to the (221) *Eos* family. Its members present deep 1 μm -bands in comparison to K-types outside the family (Clark et al. 2009). This is apparent in subpanel (h) of Fig. 11.4 for the family members (221) *Eos* and (661) *Cloelia* in comparison to (402) *Chloe* and (1545) *Therhoe*. The z_1 - z_2 distribution of K-types overlaps significantly with that of M-types. The transition between K- and M is finally drawn to include the Eos-family members (579) *Sidonia* and (653) *Berenike* in the K-class.

A further, less intuitive degeneracy is given between K- and B-types. Some K-types, notably (3028) *Zhangguoxi* and class archetype (221) *Eos*, have blue NIR slopes, as visible in their low z_1 -scores in subpanels (a) and (b) of Fig. 11.4. NIR-only spectra of these objects lacking the 1 μm -band information may therefore be confused with the high-albedo members of the B class and vice versa. (2100) *Ra-Shalom*, shown in subpanel (a) of Fig. 11.4, is classified as K-type in this work based on two NIR spectra. In the literature, both B (León et al. 2012; Binzel et al. 2019) and K (Shepard et al. 2008b) classifications are given for (2100) *Ra-Shalom*. The latter by Shepard et al. (2008b) is supported by the high thermal inertia of the asteroid. The spectral degeneracy of B and

Early on, the Eos family was known to be homogeneous, yet as there was no class between C and S, the members were classified as C or S based on subtle variations, erroneously implying a large diversity (Bell 1988).

K in NIR spectra is further commented on by Clark et al. (2009) and Popescu et al. (2018a). In Chapter 7, I show that the phase curve coefficients G_1 , G_2 resolve the degeneracy.

A total of 42 asteroids (2.0%) are classified as K-types in this work. K-types are present in cluster 24, which they share with M-types with neutral NIR slopes, and cluster 31, which contains NIR-only observations of B-types as well as L-types. Cluster 24 is resolved into K- and M-types using a two-component GMM fit to the cluster distribution in z_2 - z_3 , separating K-types based on the deep 1 μ m-band. Cluster 31 is split into K and L members only, recognising that the degeneracy of the NIR-only observations of B-types cannot be resolved with the observables in this taxonomy. The cluster members are split based on their probability to belong to cluster 23 (L) or 24 (K) in z_2 - z_3 .

11.2.2 L-Types

The mean spectral appearance of L-type asteroids is characterised by a broad 2 μ m-feature paired with the absence of a 1 μ m-band and a sharp transition into a steep slope towards the UV region around 0.7 μ m, refer to Fig. 11.3. However, there is a considerable heterogeneity in the spectra of L-types, both in terms of slopes and the presence and appearance of features. The visual albedo distribution is centred around 18%. The nature of the 2 μ m-feature and its implications are discussed in detail in Chapter 14.

The spectral diversity makes it difficult to reliably identify the L-class in this taxonomy. I therefore focus on the distribution of class members given in the literature (Sunshine et al. 2008; DeMeo et al. 2009; Devogèle et al. 2018), which cluster most in z_2 - z_4 (lower part of Fig. 11.4) due to the 2 μ m-band in combination with an absent or weak 1 μ m-feature, which separates them from K, M, and S-types. The degenerate E-types in this projection are easily distinguished if visual albedo information is present. Subpanels (i) and (j) of Fig. 11.4 show the L-prototypes (234) *Barbara*, (599) *Luisa*, and (824) *Anastasia*.

In the first latent dimension, a bimodality in terms of the slope of L-types identified in z_2 - z_4 is apparent. Comparing the spectra associated to both groups show that, unlike in similar cases for the C-complex, this bimodality is not caused by the normalisation but appears to be a property of L-types. Noteworthy is the appearance of (599) *Luisa* in both the blue and the red clusters, as shown in subpanels (b) and (d) of Fig. 11.4.

A total of 58 asteroids (2.7%) are classified as L-types in this work. Most L-types are present in clusters 4 and 23, though both clusters are populated by members from other classes as well. For cluster 4, I separate L- from S-types based on a two-component GMM fitted to the distribution of the members of clusters 23 and cluster 40 in z_3 - z_4 . For cluster 23, I split L- from M-types based on a two-component GMM in z_1 - z_4 . A small fraction of L-types are in cluster 37, which consists predominantly of M-types. The L-types are identified using a two-component GMM in z_2 - z_4 .

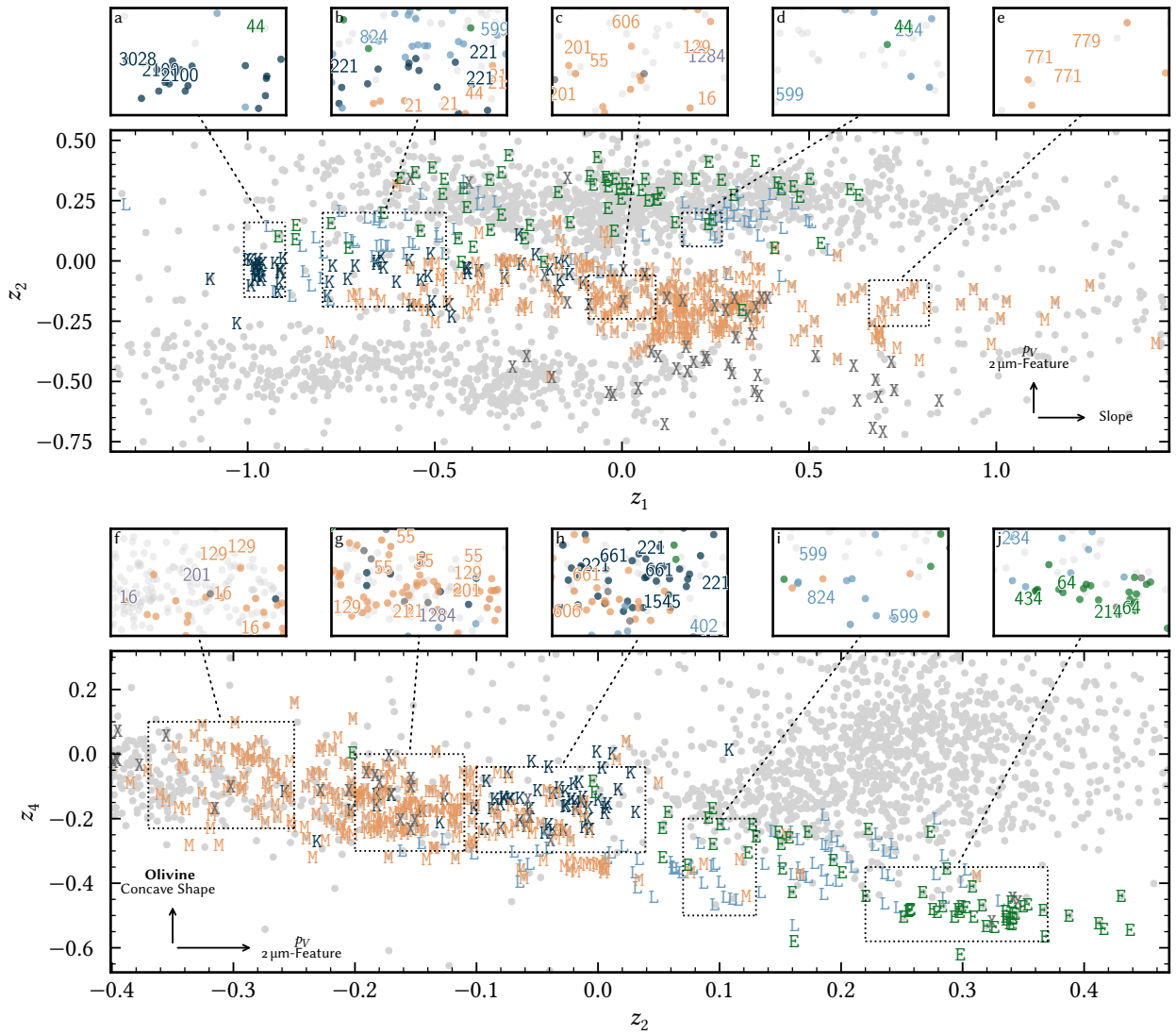


Figure 11.4: As Fig. 11.2, but showing the distribution of the classes of the M-complex and its endmember class E.

11.2.3 M-Types

The M-types were the third class defined, yet unlike its predecessor C and S, the class definition and interpretation have evolved and changed considerably over time. The reason is that the M-types are enigmatic, as discuss in Chapter 4. Their reflectance spectra of M-types are generally linearly red in the visible and NIR, as is the case for (771) *Libera* and (779) *Nina* shown in subpanel (e) of Fig. 11.4. However, some M-types show an overall convex shape with a neutral or even blue slope in the NIR, as is the case for (21) *Lutetia* in four out of five observations in this work. Objects like (21) *Lutetia*, highlighted in subpanels (b) and (g) of Fig. 11.4, resemble the Xc class from the Bus-DeMeo taxonomy. The visual albedos of M-types are in the range of 10%–20%, shown in Fig. 11.3.

The k-feature indicative of silicates is likely more present than absent in the reflectance spectra of M-types, as 40.9 % of observations present the feature, while for 30.2 % the corresponding wavelength region is not observed. In Fig. 11.5, the distribution of observations presenting the e- or k-feature is shown in the first two latent dimensions. The distribution of the k-feature is concentrated among M-types like (16) *Psyche*, (55) *Pandora*, (129) *Antigone*, and (201) *Penelope* further show the k-feature in one or several observations, and like (16) *Psyche* are highlighted in subpanels (c), (f), and (g) in Figure 11.4.

The distribution of M-types in latent space and the results acquired in the observational studies mentioned in Chapter 4 suggest to me that there are at least two populations of M-types, the chondritic population, of which (21) *Lutetia* may be the archetype, and the metallic population, of which (16) *Psyche* is the prototype (Vernazza et al. 2011; Viikinkoski et al. 2017). However, with the observables used in this taxonomy, I cannot identify a reliable distinction between populations of the diverse M-types. To not increase the entropy of the taxonomy in a false direction, I refrain here from dividing the M-class here. I give some thoughts on the continuation of the taxonomy with respect to the M-types in Part V.

A total of 142 asteroids (6.7 %) are classified as M-types in this work. M-types are predominantly found in clusters 22, 37, and 46, with smaller contributions from clusters 17 and 35. These clusters make up the X-complex and the members are split into E, M, and P as described in Chapter 10. Additional M-types are recovered from clusters 23 and 24, as described in the descriptions of the classes K and L.

11.2.4 Endmembers: E

The definition of the E-types is straight-forward, especially in comparison to the other classes in the M-complex. While they are spectrally similar to L- and some M-types with a neutral to red NIR slope, the visual albedo is the standout feature with a mean of 57 %, shown in Fig. 11.3. This high value leads to high z_2 - and low z_3 -scores in the latent space.

As L- and M-type asteroids, E-types depict a large variability of the reflectance spectra of individual asteroids. Furthermore, in the case no visual albedo information is present, E-types are degenerate with large parts of the M-complex. An example of both is the class archetype (44) *Nysa*, indicated in subpanels (a), (b), and (d) of Fig. 11.4. The observations in subpanels (a) and (d) have an albedo observation associated and are correctly classified as E-type, however, the third observation in subpanel (b) lacks an albedo value and is classified as an M-type.

Most E-types from Tholen (1984) are classified as Xe in the Bus-DeMeo system based on the e-feature at 0.5 μm . This feature is much rarer than the k-feature, as shown in Fig. 11.5. Thirteen samples in the M-complex present the feature, while 65.4 % of observations lack the corresponding wavelength region. Of the 13 samples, 4 are classified as E-type. Considering the relative sizes of the M- and E-class, the latter are hence more likely to exhibit the feature, yet, I do not observe a clustering of the e-feature. The prototype for this feature is (64) *Angelina*, while the E-class archetype (434) *Hungaria* does

Since its classification as E-type in Tholen and Barucci (1989), (214) Aschera has been classified as X, B, Cgh, and C in different works (Lazzaro et al. 2004; DeMeo et al. 2009; León et al. 2012). With a visual albedo above 50 %, (214) Aschera is here classified as Ek-type and concludes its spin through the alphabet soup.

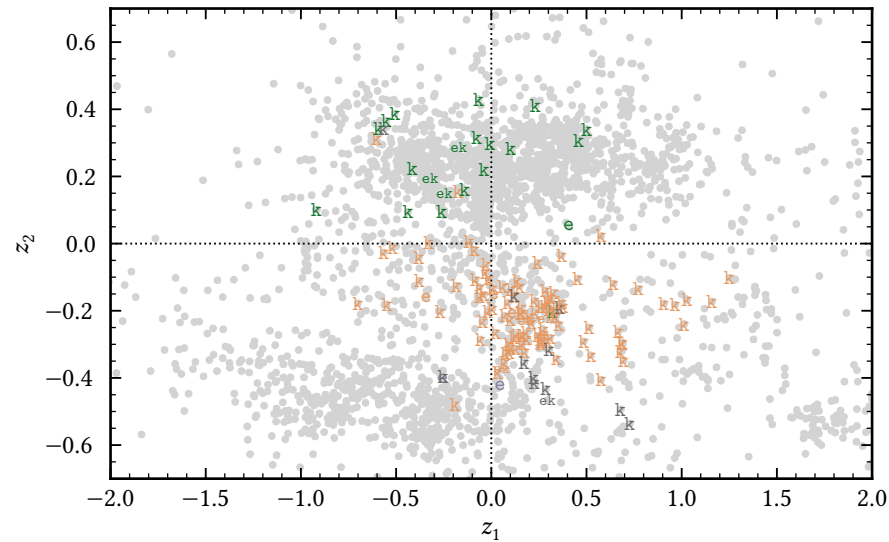


Figure 11.5: Distribution of observations which present the e- and k-feature in the first two latent dimensions, colour-coded by the class they are assigned to: green - E, orange - M, purple - P, grey - X. A smaller font size is used if the observation presents both e and k.

not present it. The k-feature is present in 30.8 % of E-type samples, while 36.9 % of observations lack the corresponding wavelength region. Observations of (64) *Angelina*, (214) *Aschera*, and (434) *Hungaria* are highlighted in subpanel (j) of Fig. 11.4.

A total of 46 asteroids (2.2 %) are classified as E-types in this work. Most E-types are located in cluster 35, though other clusters of the M-complex may also contain single samples of E-types, which are identified and assigned to the E-class based on the albedo distributions of E, M, and P outlined in Chapter 10. E-types further appear in cluster 44 among S-types, from which they are separated based on a two-component GMM fitted to the albedo distribution of the cluster.

11.3 S-COMPLEX: S, Q

The S-complex is by far the largest complex in terms of number of individual asteroids, both in this work and in previous taxonomies. This can be attributed to observational biases such as the numeric dominance of the S-types in the IMB and near-Earth space (DeMeo and Carry 2013, 2014; Binzel et al. 2019) and the high average albedo of more than 20 % compared to the 5 % of the C-complex.

The S-complex is made up of the core classes S, Q and the endmembers A, O, R, and V. Therefore all S-complex classes defined in the Tholen (1984) system remain in this taxonomy. Compared to the Bus-DeMeo system, I remove the subclasses of the S-class, as motivated in the following.

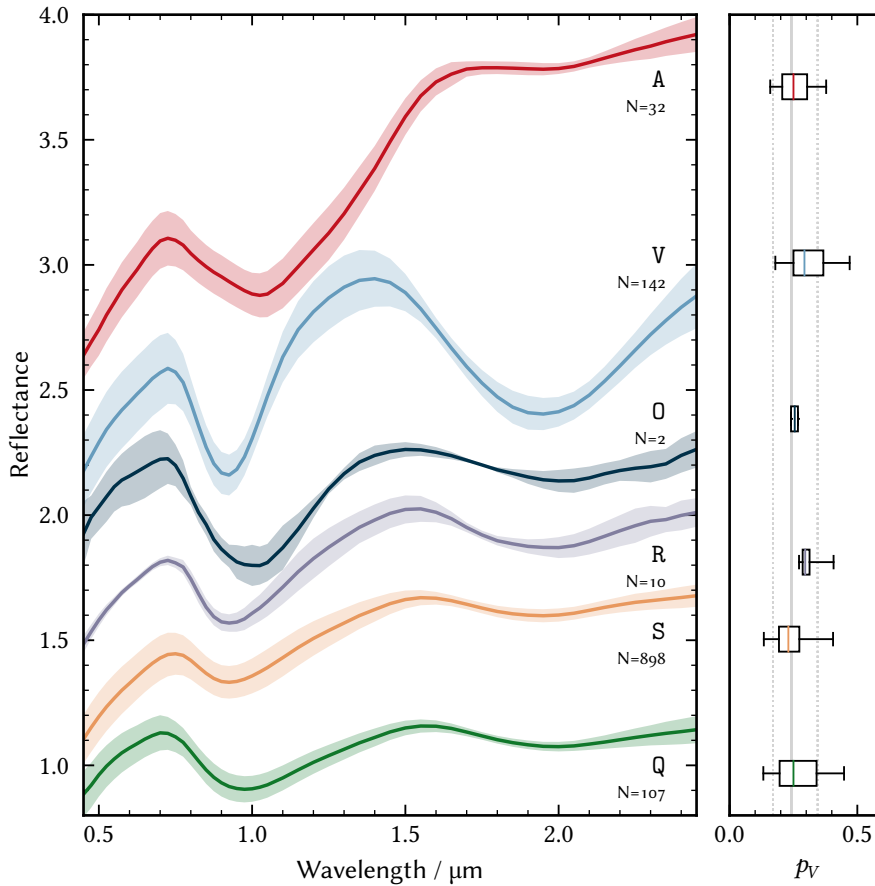


Figure 11.6: As Fig. 11.1, but showing the data space properties of classes of the S-complex. The visual albedo distribution of the S-complex is $0.15^{+0.06}_{-0.04}$.

11.3.1 S-Types

The S-types are spectrally dominated by pyroxene and smaller amounts of olivine, giving the characteristic $1\ \mu\text{m}$ and $2\ \mu\text{m}$ spectral features, refer to Fig. 11.6. The large number of S-types makes the homogeneity both in their reflectance space and visual albedo shown in Fig. 11.6 even more remarkable. Marginal variations in the spectral appearance, largely related to band depths and positions of the $1\ \mu\text{m}$ and $2\ \mu\text{m}$ features are well explained by different pyroxene and olivine ratios as well as trends of thermal alteration in ordinary chondrites (Vernazza et al. 2014; Eschrig et al. 2022).

While C class has been divided into subclasses early on (Gradie and Tedesco 1982; Tholen 1984), the S class remained a single entity until Bus and Binzel (2002a) as the siliceous surfaces are particularly subject to variability in slope and band structure induced by phase-angle effects (Sanchez et al. 2012) and space weathering (Strazzulla et al. 2005), as noted by Tholen (1984). The Bus-DeMeo system accounts for these effects by subtracting the spectral slope before classification. I instead rely on the interpretation of the latent components to serve as vectors in the compositional analysis of the S-types given in Chapter 10.

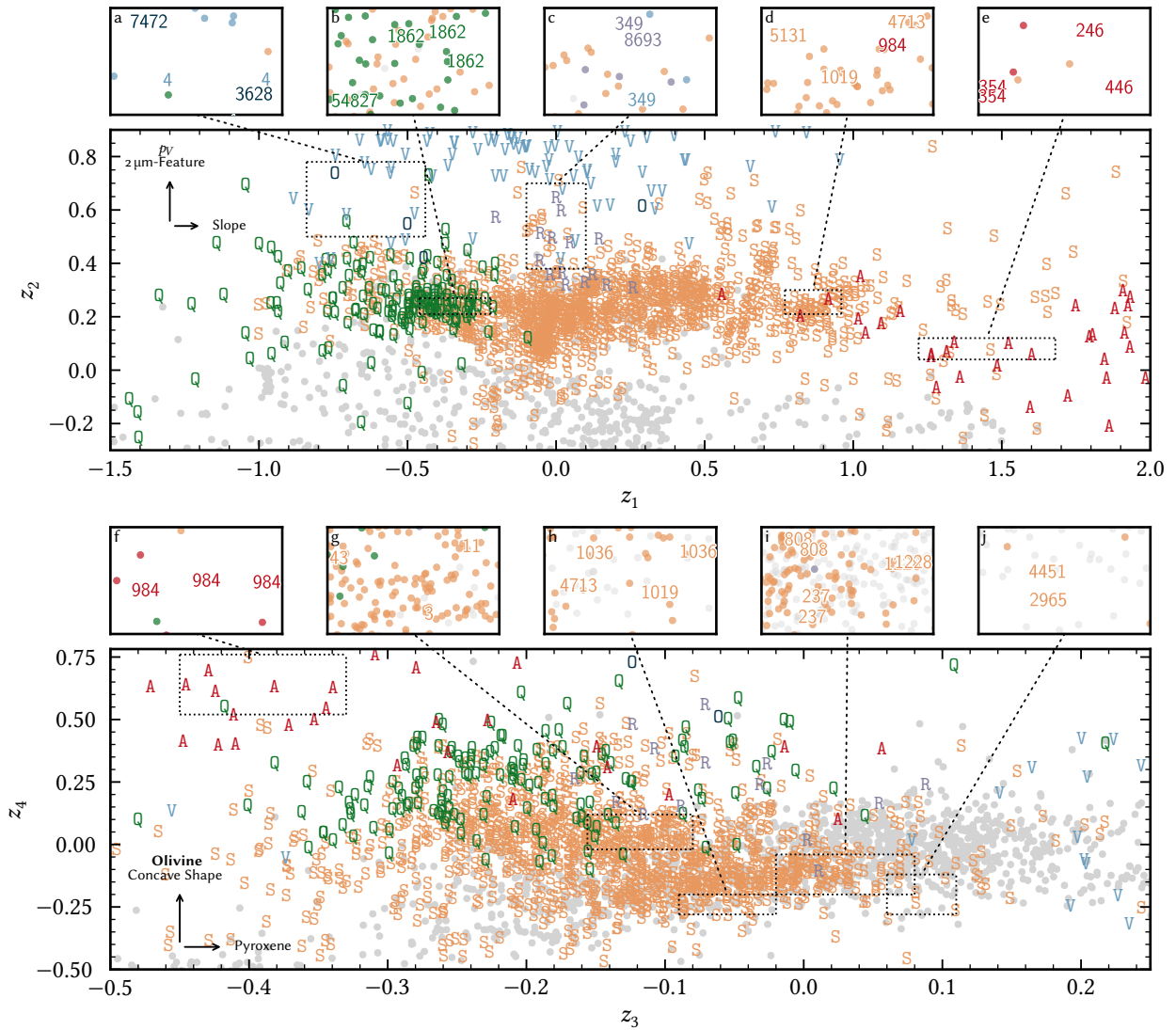


Figure 11.7: As Fig. 11.2, but showing the distribution of the classes of the S-complex and two of its endmembers, O and R. The distributions of A- and V-types are truncated for readability and shown in Fig. 11.8 instead.

The second and third latent components in Fig. 10.2 both resemble pyroxene members as this mineral dominates the spectral appearance of the S-class, in addition to the large contribution in terms of variance provided by the V-types. The first component resembles the slope, hence the vector of space weathering within the S-complex can be approximated with it (e. g. Brunetto et al. 2006). S-types with the *w*-suffix for “weathered” in the Bus-DeMeo system have higher z_1 scores than their class siblings with fresh surfaces. The degeneracy between a weathered S-type and an olivine-rich S-type (Sa in the Bus-DeMeo system), which is redder by mineralogy rather than by surface alteration, is resolved in the third and fourth latent component, which separates the pyroxene-olivine composition of objects.

For example, in subpanel (d) in Fig. 11.7 I show the Bus-DeMeo Sa-types (984) *Gretia* and (5131) *1990 BG* and the Sw-types (1019) *Strackea* and (4713) *Steel*.

The subpanel (h) shows that both Sw-types have below average olivine components, indicating that the red surface is indeed due to weathering; also shown in this subpanel is the S-type (1036) *Ganymed*. (984) *Gretia* is classified as A-type in this study due to its high z_4 score (see subpanel (f) in Fig. 11.7).

The Bus-DeMeo system further recognises Sq-, Sr-, and Sv-types in addition to the regular S-types. The prototypes given in DeMeo et al. (2009) for these subclasses are highlighted respectively in subpanels (g) ((3) *Juno*, (11) *Parthenope*, (43) *Ariadne*), (i) ((237) *Coelestina*, (808) *Merxia*, (1228) *Scabiosa*), and (j) ((2965) *Surikov*, (4451) *Grieve*) in Fig. 11.7. The continuous distribution between the main S-complex and the subclasses confirms our decision to not subdivide the S-class.

A total of 898 asteroids (42.3%) are classified as S-types in this work. Numerous clusters build the class: 0, 3, 6, 11, 14, 20, 21, 30, 33, 38, 39, 40, 42, and 47. Clusters 4, 8, 10, 43, and 44 contain members from other classes, which are separated based on GMMs as described in the respective class descriptions (L, D, R, D, and E, in order of the clusters).

11.3.2 Q-Types

Q-types resemble S-types but depict wider $1\ \mu\text{m}$ features and bluer slopes, refer to Fig. 11.6. The distribution of visual albedos is similar to the one of S-types as well, with a mean at 24%. In latent space, Q-types make up the blue end of the S-complex with low z_1 -scores. They further separate marginally from S-types in z_4 due to their strong $1\ \mu\text{m}$ -band, as apparent in Fig. 11.7. Class archetype (1862) *Apollo* and class member (54827) *Kurpfalz* are highlighted in subpanel (b).

I considered merging the Q-class into the S-class as it represents the overall continuity in the S-complex and the sample returned from (25143) *Itokawa* shows the inherent link of S-types and Q-types Binzel et al. (2004a) and Nakamura et al. (2011). However, similar to the case of the Z class, the orbital distribution of the Q-types (discussed in Chapter 13) and the useful distinction of Q as S-type with a refreshed surfaces for the NEO community convinced me to keep the class.

A total of 107 asteroids (5.0%) are classified as Q-types in this work, 83.2% of which are NEOs, considerably more than the average of 34.4% over all asteroids in the input data. Q-types are present in clusters 16 and 48 as well as in the diffuse cluster 13 whose division is outlined in Sect. 11.3.3.2.

11.3.3 Endmembers: A, O, R, V

I count the A and V class as endmembers of the S-complex due to their unique spectral appearances. O and R could also be considered core classes of the complex in terms of their observables, however, I place them among the endmembers because rarity: both are by far the smallest classes in terms of the number of asteroids. The distribution of A, O, R, and V in z_3 - z_4 is shown in Fig. 11.8.

Admittedly, the “endmember” nomenclature is somewhat arbitrary.

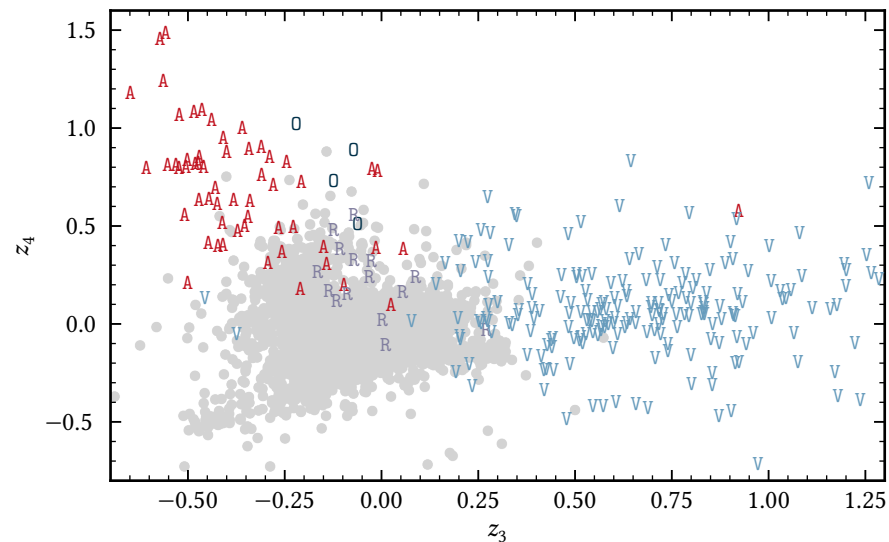


Figure 11.8: Distribution of the endmember classes of the S-complex in the third and fourth latent dimension.

11.3.3.1 A-Types

A-type asteroids have red spectra with deep $1\ \mu\text{m}$ -features representing an olivine-rich composition. The visual albedo is typically between 20%–30%, as shown in Fig. 11.6. In latent space, the spectral appearance gives A-types both the largest scores z_4 of all asteroids (refer to Fig. 11.8), and only Z-types show on average larger z_1 scores. Nevertheless, all three spectra of A-type and MC asteroid (1951) *Lick* are the reddest in the ensemble of observations based on the z_1 scores. The class prototypes (246) *Asporina*, (354) *Eleonora*, and (446) *Aeternitas* are indicated in subpanel (a) of Fig. 11.7.

A total of 32 asteroids (1.5%) are classified as A-types in this work. They are present in clusters 9, 12, 27, and 49.

11.3.3.2 O-Types

O-types have a characteristic wide (“bowl-shaped” DeMeo et al. (2009)) $1\ \mu\text{m}$ -feature and a weak $2\ \mu\text{m}$ -feature, shown in Fig. 11.6. In latent space, they have medium z_3 and large z_4 scores compared to S- and Q-types, placing them just above the core members of the complex (Fig. 11.7).

However, the O class consists of only two objects, the class archetype (3628) *Boznemcova* and (7472) *Kumakiri*, both indicated in subpanel (a) of Fig. 11.7. As such, the continuation of the O class in this taxonomy is not compatible with the overall approach of data-driven clustering. However, in the end, the unique feature and position of (3628) *Boznemcova* in latent space convinced me to keep this niche class.

Two asteroids (0.1%) are classified as O-types in this work. The O-class is difficult to define based on the clusters. It is derived from a three-component GMM of the already diffuse cluster 13, which is split into C, O, and Q. Any

assignment of the O-class should be preceded by visual scrutiny and direct comparison of the candidate spectrum to the one of (3628) *Boznemcova*.

11.3.3.3 R-Types

R-types show 1 μm and 2 μm features which are slightly deeper and wider than for S-types, refer to Fig. 11.6. The latent scores shown in Fig. 11.7 show that R-types have similar scores to S in z_1 and to Q in z_3 and z_4 .

Like O-types, the R-types are defined around a single asteroid, archetype (349) *Dembowska*. However, unlike for the former class, there are other asteroids which resemble the class archetype, like (8693) *Matsuki* shown in subpanel (c) of Fig. 11.7. Nevertheless, the definition of niche classes like these goes against the Gaussian clustering approach used in this taxonomy and is done purely for class-continuity considerations.

A total of 10 asteroids (0.5 %) are classified as R-types in this work. The class is derived from cluster 10 in a two-component GMM fit in z_1 - z_2 , where objects with larger z_2 scores are assigned to the R-class.

11.3.3.4 V-Types

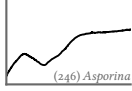
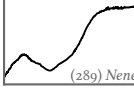
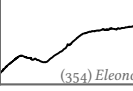


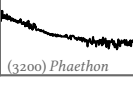
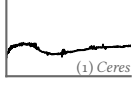
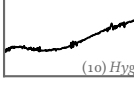
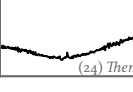
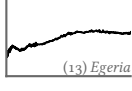
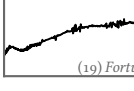
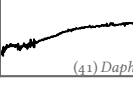
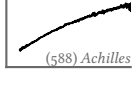
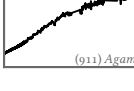
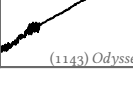
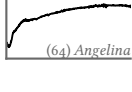
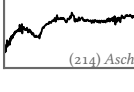
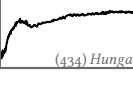
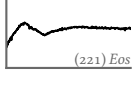
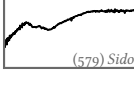
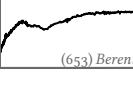

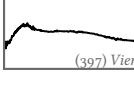
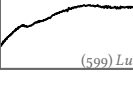
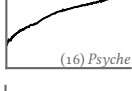
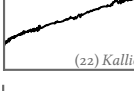
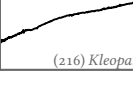
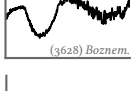

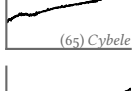

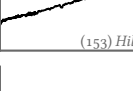
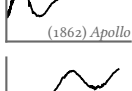
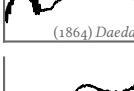
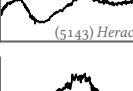
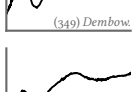
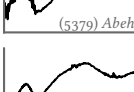
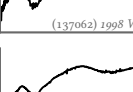
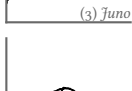
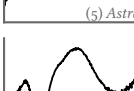
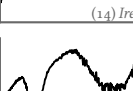
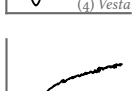

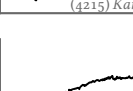
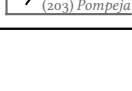
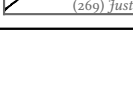
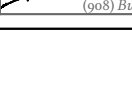
V-types have deep features at 1 μm and 2 μm indicative of a pyroxene-rich composition. The visual albedos distribution has a large standard deviation with a mean value around 30 %, refer to Fig. 11.6. The spectral properties make them easy to recognise in data as well as in latent space, where V-types present both the largest z_2 and z_3 scores. Some members of the V population are shown Fig. 11.7, including class archetype (4) *Vesta* in subpanel (a). An overview of the whole class is given in Fig. 11.8.

Due to their deep features, V-types are responsible for a large part of the variance in the input data. As such, dimensionality reduction methods such as PCA and MCFA will place a lot of emphasis on them. This makes V-types easy to recognise (such as in both the latent score plots in Figs. 10.3 and 10.4), but also suppresses the resolution of the C-, M-, and S-complex.

A total of 142 asteroids (6.7 %) are classified as V-types in this work. V-types make up clusters 7, 15, 18, 28, 32, and 45. V-types with a blue slope in the NIR further compose the diffuse cluster 41 with B-types. I did not identify a subpopulation within the V-types as it had been suggested by Binzel and Xu (1993).

An interesting experiment could be a two-step classification, where the first step identifies and removes A- and V-types. This could increase the resolution of other classes in the lower-dimensional space.

Table 11.1: Description of taxonomic classes defined in the Mahlke et al. (2022) taxonomy. Given are the spectral appearance, visual albedo distribution giving the mean value and the lower and upper standard deviation, and spectral prototypes of the 17 classes defined in this taxonomy excluding the X-types.

Class	Spectrum	Albedo	Prototypes		
A	Broad and deep absorption feature at 1 μm , strong red slope in the near-infrared.	0.25 $^{+0.09}$ $^{-0.07}$			
B	Neutral- to blue slope in the visible, blue slope in the near-infrared.	0.06 $^{+0.05}$ $^{-0.03}$			
C	Red visible slope with a possible broad feature around 1 μm and a red near-infrared slope. The classes/spectrum might have an overall concave shape.	0.05 $^{+0.02}$ $^{-0.01}$			
Ch	Absorption feature at 0.7 μm . The near-infrared slope is red while the overall shape might be convex.	0.05 $^{+0.02}$ $^{-0.01}$			
D	Featureless with steep red slope with a possible convex shape longward of 1.5 μm .	0.06 $^{+0.03}$ $^{-0.02}$			
E	Strong red slope in the visible with a feature around 0.9 μm of varying depth and a neutral near-infrared continuation.	0.57 $^{+0.15}$ $^{-0.12}$			
K	Strong red slope in the visible with a broad feature around 1 μm followed by a blue- to neutral near-infrared slope.	0.13 $^{+0.04}$ $^{-0.03}$			
L	Variable appearance apart from a red visible slope. A small feature around 1 μm and a possible one at 2 μm . The near-infrared slope is blue or red.	0.18 $^{+0.07}$ $^{-0.05}$			
M	Linear red slope with possible faint features around 0.9 μm and 1.9 μm . Might show convex shape in the near-infrared.	0.14 $^{+0.05}$ $^{-0.04}$			
O	Broad, bowl-shaped 1 μm absorption feature and a weaker feature at 2 μm .	0.26 $^{+0.02}$ $^{-0.02}$			
P	Linear red slope and generally featureless. Less red than D-types.	0.05 $^{+0.02}$ $^{-0.01}$			
Q	Broad absorption at 1 μm and a shallow feature at 2 μm . An overall blue slope in the near-infrared.	0.24 $^{+0.12}$ $^{-0.08}$			
R	Strong feature at 1 μm and a feature at 2 μm . The latter feature is shallower than in V-types.	0.30 $^{+0.05}$ $^{-0.04}$			
S	Moderate features around 1 μm and 2 μm and a neutral- to red near-infrared slope.	0.24 $^{+0.10}$ $^{-0.07}$			
V	Deep absorption features at 1 μm and 2 μm . The former is much narrower than the latter.	0.29 $^{+0.11}$ $^{-0.08}$			
Z	Extremely red slope, redder than the D-types. Featureless but may exhibit concave shape in the near-infrared.	0.07 $^{+0.04}$ $^{-0.03}$			

12.1	Classification Tool: <code>classy</code>	145
12.2	Probabilistic Classifications	146
12.3	Degeneracies	147
12.3.1	Visible-near-infrared	147
12.3.2	Visible-only	149
12.3.3	NIR-only	150
12.4	Comparison to Previous Taxonomies	151

This chapter treats the practical aspects of the classification. The classification tool `classy` is introduced which facilitates the taxonomic classification of asteroid observations. I highlight the utility of probabilistic classifications and use it to quantify degeneracies in the class scheme. The chapter concludes with a comparison of this taxonomy to previous schemes.

12.1 CLASSIFICATION TOOL: `classy`

Providing a tool to conveniently apply a taxonomic scheme to new observations might be as important as the design of the class scheme itself. Unlike previous taxonomic schemes which could be applied with the knowledge of principal components and a straight-forward decision-tree based on the principal scores, the method used in this work requires more preprocessing steps and the parameters of numerous models trained during the derivation, such as the ones of the MCFA model or the several GMMs required for the cluster-to-class decision-tree. This complicates the matter of making the classification accessible to anyone.

Observations can be classified in the Bus-DeMeo scheme using a web interface.¹ I am more familiar with writing tools for the command-line and python packages, hence, I here present `classy`. `classy` facilitates the classification of asteroid observations in this taxonomic scheme. It is written in python and distributed as open-source software² with documentation and tutorials³. The software source further contains all data necessary to recreate the analyses presented in this work. While it is certainly challenging to examine the entire taxonomy by means of the source code, in principle, it is possible.

`classy` provides the full scope of necessary steps to go from an observed reflectance spectrum including a possible visual albedo to the final vector of probabilistic classifications. This includes the preprocessing outlined in Fig. 9.1 including the feature parametrisation, the clustering as shown in Fig. 10.1, and the cluster-to-class decision-tree presented in Table 10.1. Inter-

I recognise that command-line interfaces are not everyone's favourite. A graphical user interface for `classy` is in development.

If no visual albedo is present, `classy` offers to retrieve a literature value using rocks, refer to Chapter 6.

¹ smass.mit.edu/busdemeoclass.html

² <https://github.com/maxmahlke/classy>

³ <https://classy.readthedocs.io/>

active dialogues facilitate the smoothing and resampling of the reflectance spectra as well as the semi-automated feature recognition.

For now, `classy` is primarily a command-line tool. The development is ongoing and the syntax may change, nevertheless, I give here a brief application example.⁴ The example data is the `sp291`⁵ observation run of the MITHNEOS survey consisting of six NIR-only spectra, one of which has been extended into the visible region using data from SMASS. These observations were made after the model training described in [Chapter 10](#) and as such represent new data to the classifier. There are six files, one for each spectrum. For now, `classy` expects the input to be transposed, i. e. the wavelength information is given in the columns and each spectrum represents a row in the input data. After joining the six observations in this manner in a file called `sp291.csv`, they can be preprocessed for the classification using the simple syntax

```
$ classy preprocess sp291.csv
INFO      Stored preprocessed data to sp291_preprocessed.csv
```

This command launches interactive plots of each spectrum to facilitate the smoothing and resampling as well as the semi-automated feature recognition described in [Chapter 10](#). The spectra are then normalised and the result is stored to a new file with the same filename except for an added `_preprocessed` suffix. The classification is achieved in a similar manner, by calling the corresponding command on the preprocessed data:

```
$ classy classify sp291_preprocessed.csv
INFO      Looks like we got 2 S, 1 Ee, 1 B, 1 X, 1 Q
INFO      Stored classified data to sp291_classified.csv
```

As the example shows, we get an immediate feedback on the classification results. The classified data is stored in a new filename with the same filename as the input data except for an added `_classified` suffix.

I intent to keep this process as simple as possible. As shown in the example, the only user interaction is the correct formatting of the input data and the following input for the smoothing of the spectra and the feature recognition.

Feedback and suggestions are always welcome on the project's [GitHub page](#).

12.2 PROBABILISTIC CLASSIFICATIONS

The probabilistic classification of asteroids is one of the three main improvements over previous taxonomies. Depending on the point of view, it either quantifies the uncertainty of the classification or the compositional mixture of the asteroid. An example is shown in [Fig. 12.1](#) for (325) *Heidelberg*, where the left hand side depicts the observation, i. e. the reflectance spectrum and visual albedo, while the resulting class probabilities from this observation are shown on the right hand side. From the top down, I provide an increasing amount of observational data to the classification tool `classy`. For the first row, only the albedo of 8.8% is given to the classifier. The class probabilities show that this information is most compatible with the classes of the C-complex, although M and S get significant probabilities. This is a remnant of the model training, as the probability prior at the start of the classification

⁴ This example is presented in more detail in the online documentation at <https://classy.readthedocs.io/en/latest/tutorial.html>.

⁵ <http://smass.mit.edu/catalog.php>

is the ratio of the classes of the observations that were used to train the model. As given in Table 10.3, M- (8.4 %) and S-types (39.8 %) were the most numerous.

In the second row of Fig. 12.1, I pass the visual region of the spectrum including the visual albedo to `classy`. The picture is much clearer at this point, three classes share the probability almost evenly among each other: D, M, and P. For all three classes, both albedo and reflectance spectrum fit to a certain degree. The visual albedo is on the low end for M yet on the high one for D and P, refer to Table 11.1. The linear and featureless spectrum does not resolve the ambiguity. The picture remains the same when adding more data in the second and third row. In the end, (325) *Heidelberg* is classified as M-type, yet both D and P have about 30 % probability each.

This example shows the benefit of probabilistic classifications. If we think of taxonomy as a language and of the classes D, M, and P as *red*, *green*, and *blue*, probabilistic classifications now allows us to call (325) *Heidelberg* *white*. Furthermore, the example is a warning that singular measurement may be misleading. The reason that (325) *Heidelberg* is ambiguous here is due to the albedo of around 9 %. A quick exploration with `rocks` ensures us of the M classification, as most reported albedos cross the 10 % threshold:

```
$ rocks albedos heidelberg
+-----+-----+-----+
|   | albedo | method | shortbib |
+-----+-----+-----+
| 1 | 0.107  | STM    | Tedesco+2002a |
| 2 | 0.138  | STM    | Ryan+2010      |
| 3 | 0.085  | NEATM  | Ryan+2010      |
| 4 | 0.105  | NEATM  | Usui+2011      |
| 5 | 0.161  | NEATM  | Masiero+2012   |
| 6 | 0.104  | NEATM  | Alí-Lagoa+2018 |
| 7 | 0.096  | NEATM  | Alí-Lagoa+2018 |
| 8 | 0.083  | NEATM  | Alí-Lagoa+2018 |
+-----+-----+-----+
```

Passing an empty observation actually returns this prior, which somehow makes sense: if I have no information, I can assume it is an S-type with ~ 40 % probability.

12.3 DEGENERACIES

The probabilistic classifications enable us to quantify degeneracies of classes when different regions of the spectrum are observed. These degeneracies are inspected in this part.

12.3.1 Visible-near-infrared

Using the classifications from the VisNIR spectra and visual albedos used to train the machine learning model, I quantify the class confusion, i. e. the average probability of each class to be classified as any other class. The result is shown in Fig. 12.2. The matrix diagonal corresponds to the average certainty of the classification of a given class. While not shown in Fig. 12.2, these values are typically above 80 % and classes like A and V have values close to 100 %, meaning that a V classification is almost always certain due to the unique properties of the class. Less certain are K, L, and M with probabilities around 80 %. Pairs of classes which are partly degenerate are B and C, where

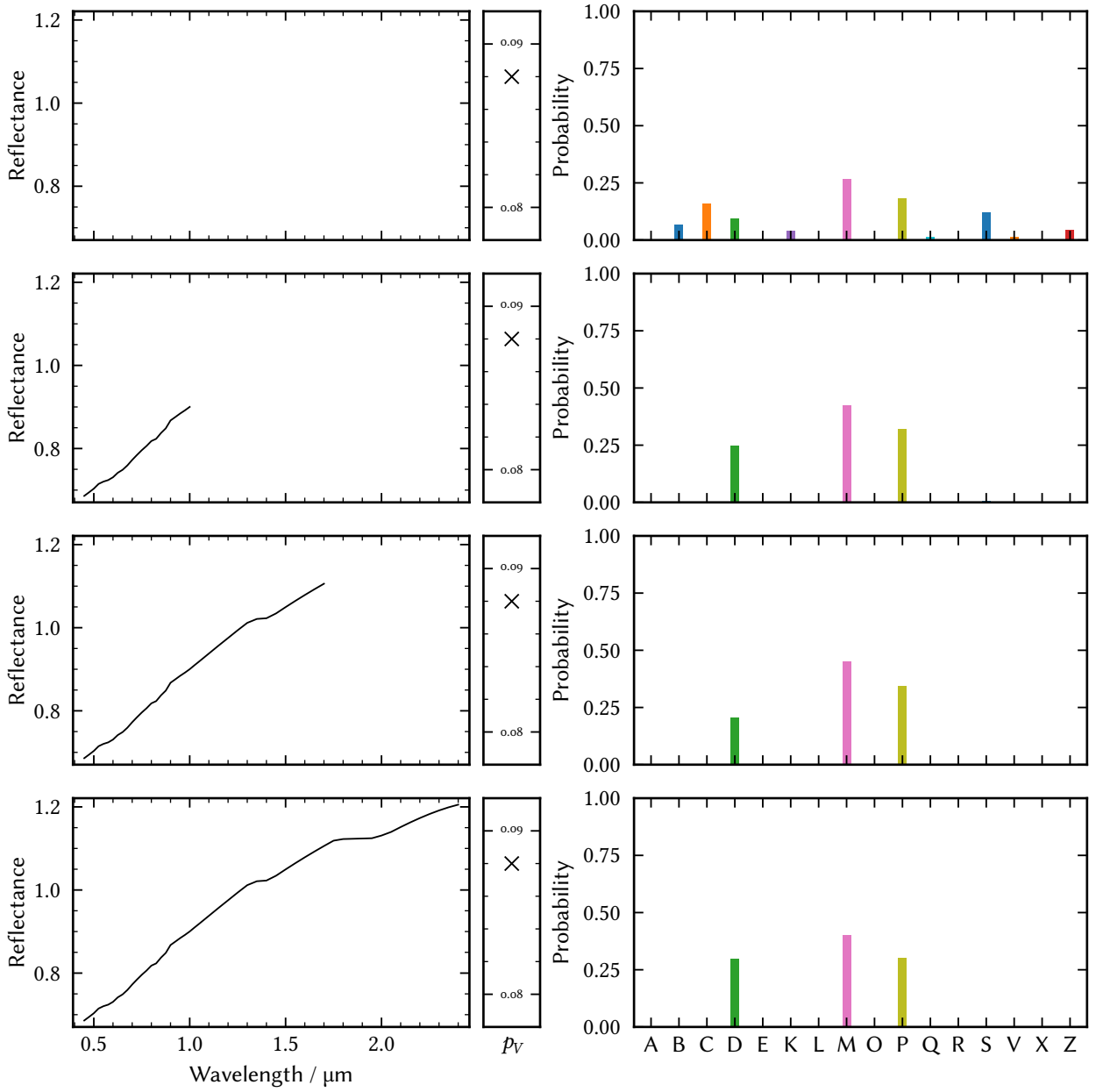


Figure 12.1: Classification of (325) Heidelberg. Ch does not have a class probability assigned as it is based on the detection of the h-feature in members of the C-complex.

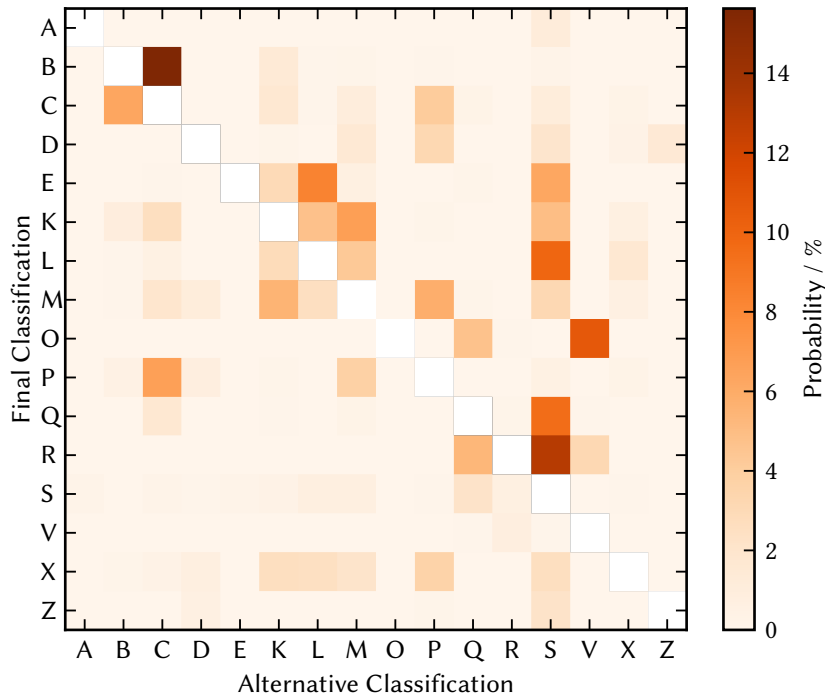


Figure 12.2: Confusion matrix between the classes defined in this taxonomy in the visible-near-infrared and albedo input space. For each class in the taxonomic scheme, we give the average probability of its samples to be classified as any other class based on the complete dataset. Note that the Ch-class is missing as it relies on the detection of the $0.7\ \mu\text{m}$ h-feature and does not have a class probability associated. For better readability, the main matrix diagonal corresponding to the equal-class cases is left empty. These values are generally above 80 % and lowest for K, L, M, and R.

classifications of the former on average have about 15 % probability to be a C-type, and R and S. Nevertheless, we see that the combination of VisNIR and visual albedo generally gives classifications with low uncertainties, and cases like the one of (325) *Heidelberg* presented above are the exception.

12.3.2 Visible-only

Visible-only spectra are the most common type of spectra, in particular after the third data release of ESA's *Gaia* satellite. In Fig. 12.3, I repeat the calculation of the confusion matrix as above for the sample of 2923 visible-only spectra including 81.4 % visual albedos which were not included in the training observations, refer to Chapter 9. The overall result is similar yet shifted by about 15 %–20 % towards larger uncertainties. The matrix diagonal values now span values between 63 %–91 %. This result is expected, visible-only spectra contain significant compositional information, especially in combination with the visual albedo, yet the missing NIR information takes its toll. To quantify which classes are most affected by the missing NIR information, I cross-match the asteroids in both the VisNIR and the visible-

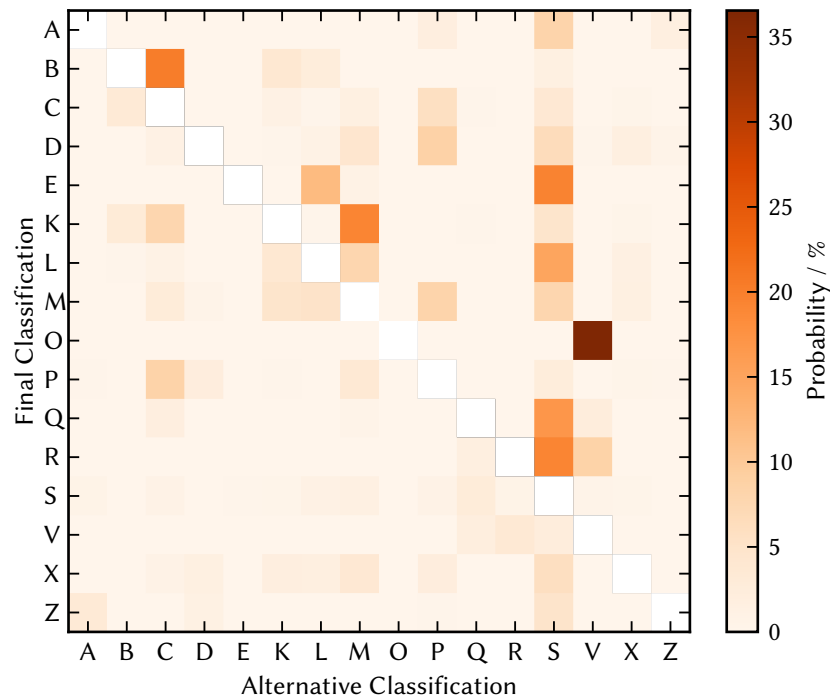


Figure 12.3: As Fig. 12.2, using the dataset of 2923 visible-only spectra with 81.4 % albedo observations. Note the different colourbar scale compared to Fig. 12.2. The main matrix diagonal values are between 63 %–91 % and lowest for K, L, M, and O.

only samples and compare the final classifications of asteroids that are in both dataset. This comparison is shown in Fig. 12.4 for 328 observations of 267 individual asteroids. Here, we expect the matrix diagonal to be the most populated, however, note that the number of samples per row changes significantly: S (160), C (117), and P (60) make up 89 % of the sample. In addition, asteroids classified as B, C, or P in the VisNIR dataset may miss the h-band information, while in the visible-only dataset, they are classified as Ch based on the $0.7\ \mu\text{m}$ -feature. Understanding these caveats shows that most classifications are consistent between the datasets. As above, we see the most confusion in members of the M-complex.

12.3.3 NIR-only

As there is no dataset of NIR-only spectra which were not used in the model training, I instead take the VisNIR sample, truncate all spectra below $0.8\ \mu\text{m}$ and remove the visual albedo information. I then classify these spectra using `classy` and repeat the computation of the class confusion matrix as above. The result is shown in Fig. 12.5. The matrix diagonals span values from 55 %–99 %, with the highest uncertainty among M and P types, which derives from the missing visual albedo information to differentiate the classes. Most noteworthy in Fig. 12.5 is that classifications of the large classes C and S become more probable for the observations in the smaller classes, which is

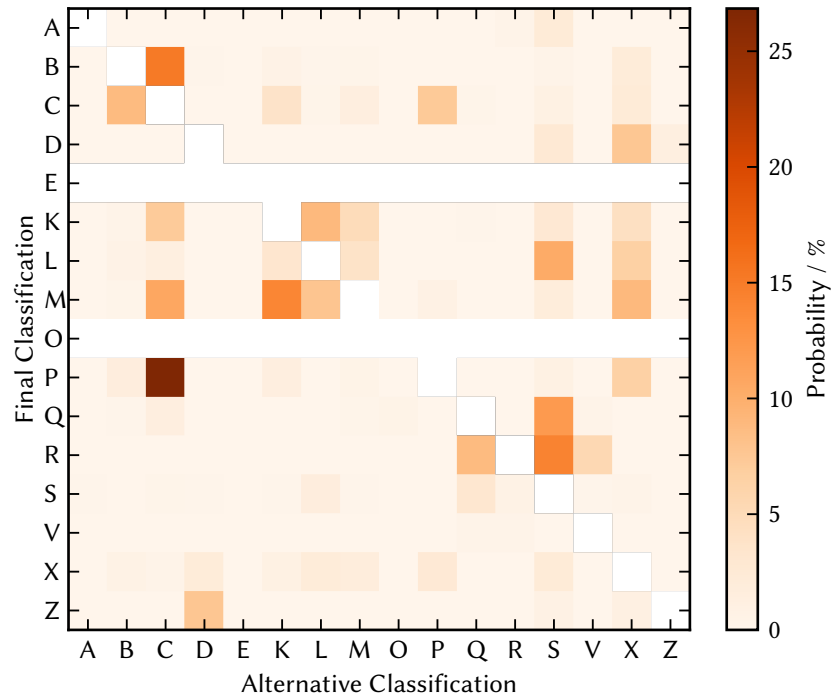


Figure 12.5: As Fig. 12.2, using the dataset of 2813 NIR-only spectra without albedo information. Note the different colourbar scale compared to Fig. 12.2. No observation in this sample is classified as E or O. The main matrix diagonal values are between 55 %–99 % and lowest for M and P.

The T-class is not present in this taxonomy and the feature characteristic of the Xn has been grouped into the k-feature. The evolution of the X-complex between the taxonomies is unclear as the visual albedo is not taken into account in the Bus-DeMeo system. No analogues for K and L were defined in Tholen (1984).

Figure 12.6 shows a comparison of the classifications in this work to those found in the literature on a per-asteroid basis. The literature classifications are extracted from SsODNet using rocks and mapped into the scheme of this taxonomy following Table 12.1. To map the X-complex from Bus-DeMeo to this scheme, I convert the previous X-, Xe-, Xk-, and Xn-classifications to E, M, P, or X depending on the visual albedo of the asteroid.

Overall, the continuity of class assignments is high and the previous classification is the most likely current classification for asteroids of all taxonomic classes except for O, R, and the new class Z. The difficulty of defining O and R in this scheme are outlined in Chapter 11, hence, this result is neither unexpected nor an issue.

Table 12.1: Evolution of taxonomic scheme from Tholen (1984) to Bus-DeMeo (Bus and Binzel 2002a; DeMeo et al. 2009) to this work. Published in Mahlke et al. (2022).

Tholen		Bus-DeMeo		This Work
B	→	B	→	B
F	↗			
G	→	Cg	↘	
	→	Cgh	↘	
C	→	C	→	C
	→	Ch	→	Ch
	→	Cb	↗	
D	→	D	→	D
			→	Z
P	...	Xc	...	P
M	...	Xk	...	M
X	...	X	...	X
E	...	Xe	...	E
	...	Xn		
T	→	T		
		K	→	K
		L	→	L
Q	→	Q	→	Q
		Sq		
	↗	Sr	↘	
S	→	S	→	S
	↘	Sa	↗	
		Sv		
O	→	O	→	O
R	→	R	→	R
A	→	A	→	A
V	→	V	→	V

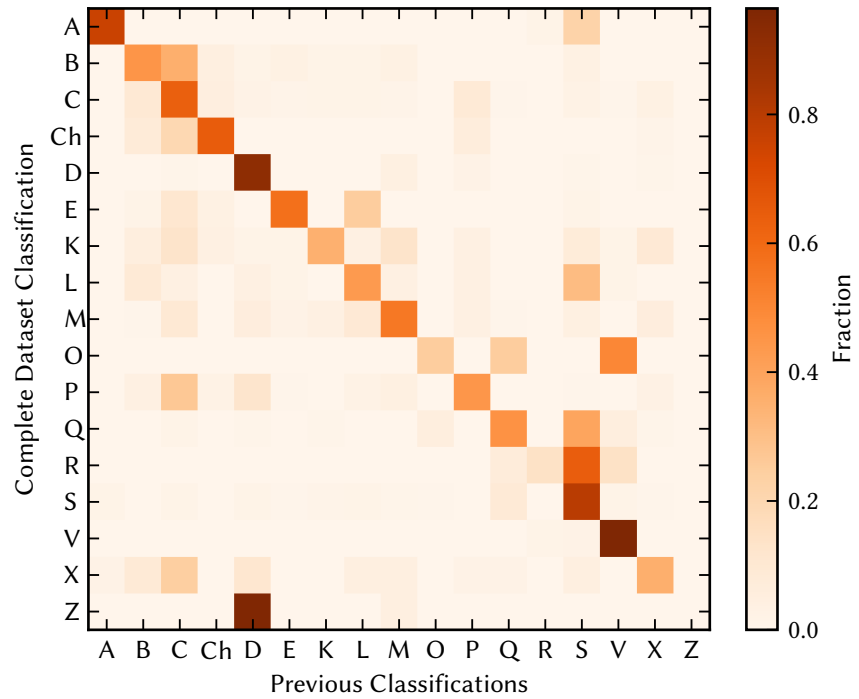


Figure 12.6: The confusion matrix of asteroid classifications in this work compared to those found in the literature. This comparison includes 2676 samples of 1852 individual asteroids. The literature classifications were mapped into this taxonomy scheme following [Table 12.1](#). Note that the number of samples differs between the rows.

Part IV

Asteroids and Meteorites

13.1	Primitive Material	159
13.2	Hydration	161
13.3	Metals and Silicates	161
13.4	Trends in Near-Earth Space	163

It is time to reap the benefits of the labour presented in this work. The dynamical distribution of asteroids in the Solar System is discussed and displayed in [Chapter 4](#). I now add a new dimension of information in the form of the compositional distribution derived in the previous chapters. One caveat to keep in mind is that the sample of asteroids with reflectance spectra is biased in many ways (Binzel et al. 2019; Marsset et al. 2022), including target properties (towards bright and large asteroids; McCord et al. (1970), Gradie and Tedesco (1982), and Zellner et al. (1985)) as well as populations (towards asteroids in families and NEOs; e. g. Vernazza et al. (2006), Willman et al. (2008), Fieber-Beyer et al. (2011), De Prá et al. (2018), Devogèle et al. (2019), and Arredondo et al. (2021)). Conclusions based on a sample like this are therefore pending confirmation by further statistical analyses, which I do not have time for in this work as France (rightfully) applies a strict upper limit of three years on the duration of a PhD project.

The distribution of asteroids in the Main Belt is shown in the upper part [Fig. 13.1](#) by means of their proper semi-major axis and proper inclination. Resonant regions and dynamical populations are marked as described in [Chapter 1](#). Superimposed are the taxonomic classifications of 3136 individual Main Belt asteroids classified in this work, indicated by the respective class letters. The lower part shows the distribution of all 4526 classified asteroids in their osculating semi-major axes versus eccentricities, with focus on the composition of NEOs and MCs is highlighted. The borders of the different near-Earth populations are indicated as defined in [Chapter 1](#).

The global trends in [Fig. 13.1](#) are familiar: S-types dominate the IMB and MMB, while the OMB and beyond are dominated by members of the C-complex (Chapman et al. 1975; DeMeo and Carry 2014). The JTs are predominantly D- and P-type asteroids (Gradie and Tedesco 1982; DeMeo and Carry 2014). For C- and S-types, we see the aforementioned bias towards observations of family members. The near-Earth space is dominated by S-types, though Q-types are more common at lower perihelia (DeMeo et al. 2014b). I discuss these aspects in more detail in the following. While [Fig. 13.1](#) is a visually appealing summary of three years of work, [Fig. 13.2](#) gives part of the information in a more readable format. Shown are the probability distributions of the classes separated by the three complexes over their semi-major axes. The probability distributions are derived by fitting a one-dimensional KDE with a cosine kernel to the distribution in semi-major axes of each class. Note that the distributions are normalised to one, hence, the classes are compared in relative abundances rather than absolute ones.

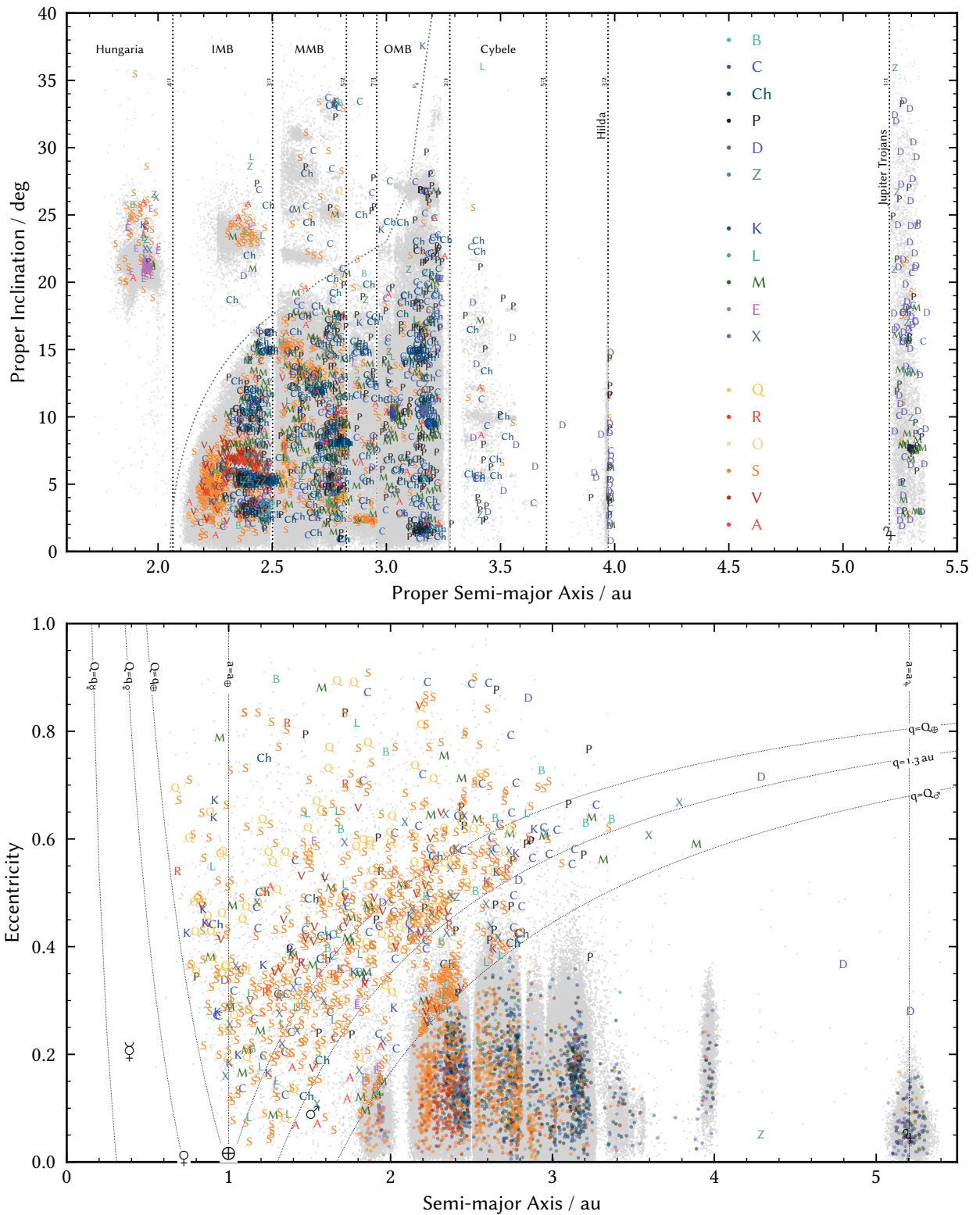


Figure 13.1: The compositional distribution of asteroids superimposed on the dynamical one. *Top*: The compositional distribution of Main Belt Asteroids studied in this work is indicated by the respective class letter of each asteroid in proper-elements space. The mean motion resonances and the secular resonance ν_6 are indicated by dotted lines, refer to [Chapter 1](#). *Bottom*: The compositional distribution of all asteroids in the inner Solar System is shown in osculating-elements space. The orbital elements of the planets are given by their respective symbols, refer to [Chapter 1](#).

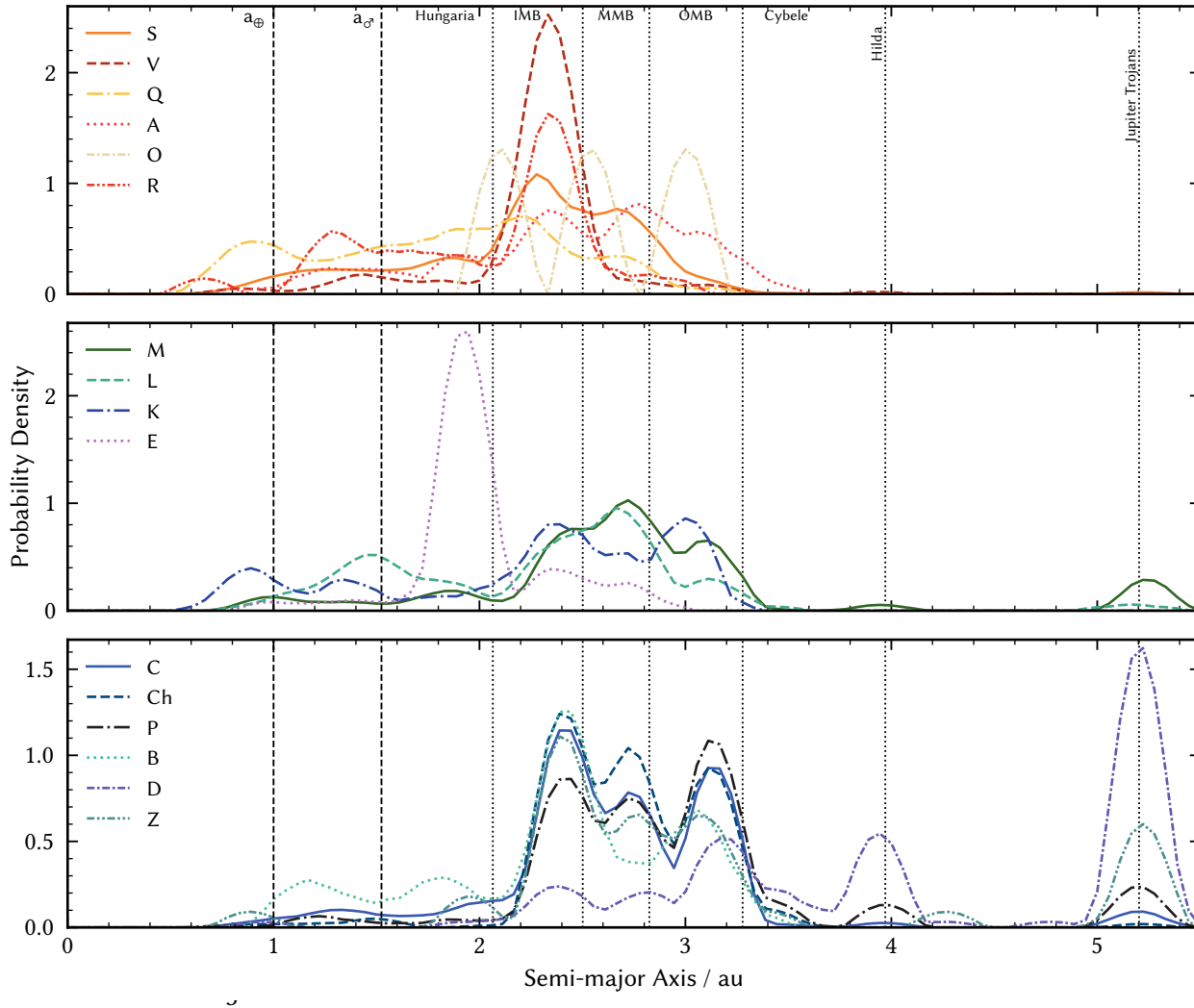


Figure 13.2: The probability distributions of the taxonomic classes of the S (*top*), M (*middle*), and C (*bottom*) complexes are shown by means of a one-dimensional kernel density estimator fits to the semi-major axes of all asteroids belonging to the respective class.

13.1 PRIMITIVE MATERIAL

As stated in [Chapter 4](#), the Z-class is made up of asteroids with extremely-red reflectance spectra. Previously, its members were assigned to the class D, which is associated to JT. Nevertheless, D-types in the Main Belt have previously been identified, such as (3283) *Skorina*, (15112) *Arlenewolfe*, and (17906) 1999 *FG32*, which had been recognised in SDSS observations (e. g. Carvano et al. 2010) and described in a follow-up study by DeMeo et al. (2014a), who further identified (908) *Buda* as a similar object. Marsset et al. (2022) even note a statistical excess of D-type asteroids among the NEO population compared to the predicted influx from the Main Belt source regions.

In this work, I show that there is a population of asteroids primarily in the Main Belt which are consistently redder than the D-types in the JT

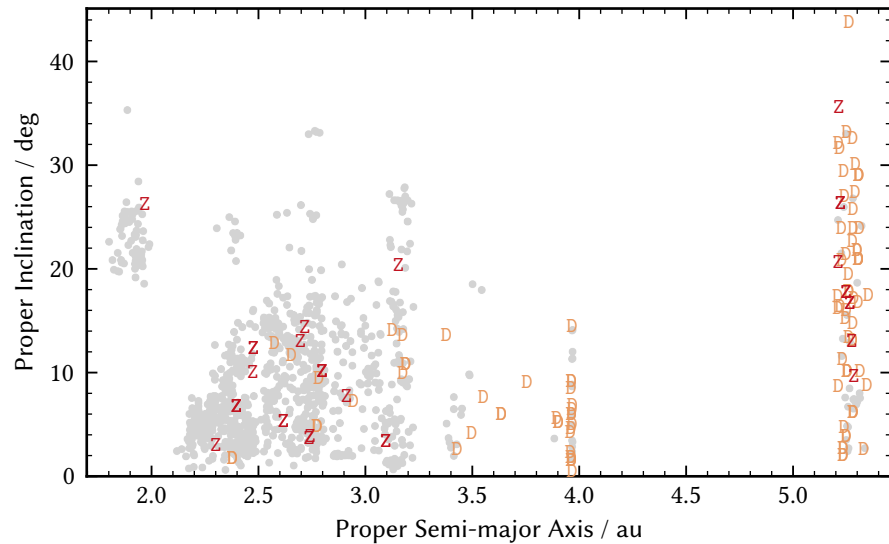


Figure 13.3: The orbital distribution of D and Z-types identified in this work.

region, refer to [Chapter 11](#). I therefore assign them to the new class Z. The orbital distributions of D- and Z-types are shown in [Fig. 13.3](#). The majority of the former are located in the JT and Hilda regions, while the latter are scattered primarily throughout the Main Belt. There is no clear-cut separation, however, as seen in [Chapter 4](#), the compositional transition between D and Z is continuous as well.

A question worth considering: Should the proper elements be included in a taxonomic scheme?

What is the compositional interpretation of the Z-types? Hasegawa et al. (2021b) suggest that the now prototypes of the Z-class (203) *Pompeja* and (269) *Justitia* are TNOs implanted into the Main Belt, owing their extremely red appearance to complex surface organics. The implantation of material from the trans-Neptunian region into the Main Belt has previously been demonstrated by Levison et al. (2009) in the framework of the Nice model of the formation of the Solar System (Gomes et al. 2005; Morbidelli et al. 2005; Tsiganis et al. 2005). This process would explain the scattered distribution of the Z-class members over the Main Belt. An alternative scenario is that Z-types are weathered D-types as they are on average closer to the Sun and the compositional distribution between both is continuous as mentioned above. I propose methods to investigate either scenario in the conclusions and outlook part of this work in [Part V](#).

Further apparent in the bottom part of [Fig. 13.2](#) is that P-types as identified in this work are mostly located in the Main Belt, particularly in the OMB. Historically, they have been associated to the D-types in the Jovian region, yet DeMeo and Carry (2014) already showed a significant fraction of P-types in the Main Belt. In addition, as seen in [Chapter 4](#), they are closer to B- and C-types than to D-types in terms of their observables. Vernazza et al. (2021) suggest that P-types may be the shell of partly-differentiated primitive asteroids as shown for (87) *Sylvia* (Carry et al. 2021), where the core material may be C-like.

13.2 HYDRATION

This brings us to the in-parts hydrated side of the C-complex. The fraction of hydrated to non-hydrated C-type asteroids in this work is between 39 %–48 %, where the first number is the ratio of Ch asteroids to the sum of B-, C-, and P-types, while the latter is an upper limit assuming that all 12.4 % of the C-complex samples in the VisNIR sample which do not have NIR information are hydrated. These numbers are between previous estimates of about one third of Ch- to C-types by Rivkin (2012) and 50 % by Fornasier et al. (2014).

Vernazza et al. (2016) suggest that variations in the spectra of Ch-types may be explained by variations in the surface regolith grain size, indicating that the class is homogeneous over different size ranges. This is consistent with high-resolution imaging observations of Ch-type (41) *Daphne* (Carry et al. 2019) and literature density values for 75 other Ch-types (Carry 2012). Bland and Travis (2017) argue that the parent bodies of Ch asteroids formed as “mud balls” rather than rocky bodies, leading to homogeneous bodies which can explain the homogeneity of the CM-like material. This in turn suggests that, as the parent bodies are disrupted, the resulting families should be homogeneous in terms of hydrated or anhydrous C-complex members as well.

The right hand side of Fig. 13.4 shows the taxonomically classified samples of the *Themis*-, *Hertha*-, and *Erigone*-families over their proper orbital elements. The distribution of all family members extracted from SsODNet is given as grey background population. In the upper left hand side of each figure, I indicate the fraction of C-type asteroids versus Ch-types and those that do not have NIR information, marked as C*. Assuming that 24 % of C-type asteroids which show a 3.0 μm band due to hydration do *not* show the 0.7 μm proxy band from Rivkin et al. (2015a), the *Erigone* family might be fully hydrated and therefore associated to CM chondrites (Vernazza et al. 2016). The *Hertha* family, on the other hand, does not show any sign of hydration. For *Themis*, we see a fraction of C and Ch which is not compatible with complete hydration or lack therefore using the Rivkin et al. (2015a) assumption.

B-types are observed among the *Nysa-Polana* family complex which includes the *Hertha*-family (a_p : 2.49 au, i_p : 2.67 deg, Nesvorny (2015)) and in the *Themis*-subfamily *Beagle* (3.14 au, 1.09 deg, Fornasier et al. (2016)). A curious feature in the lower part of Fig. 13.2 is that the B-types are the only C-complex member with a decreasing probability distribution in the MMB. However, as noted in the introduction, this might be a sample bias due to the aforementioned families.

13.3 METALS AND SILICATES

The middle row of Fig. 13.2 gives the distributions of members of the M-complex respectively. As is tradition, E-types are predominantly found in the *Hungaria*-family (1.94 au, 20.9 deg) (Gradie and Tedesco 1982; Bottke et al. 2012). A third of K-types are located in the *Eos*-family (3.01 au, 10 deg) (Bell 1988) while the rest are scattered over the IMB and near-Earth space. M and L do not show any apparent clustering in our sample. Four members of the

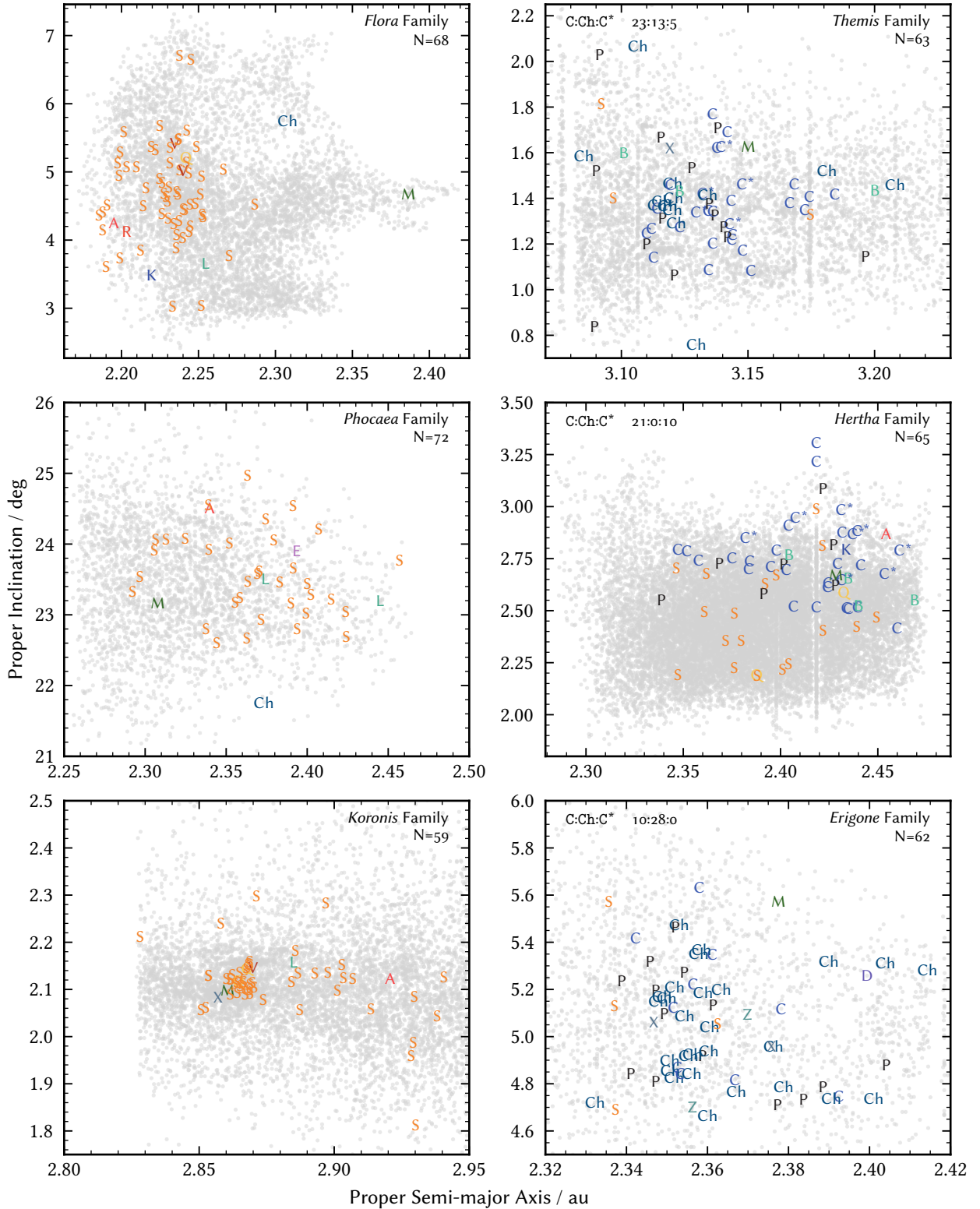


Figure 13.4: Compositional distribution of asteroids in large S- and C-type families against their proper elements. The number N of classified asteroids per families is given in each panel. For C-type families, the ratio of C to Ch and C* (NIR-only C, thus either C or Ch) is given.

proposed *Barbarian* (associated to L-types, (Devogèle et al. 2018)) family *Watsonia* (2.76 au, 17.4 au) are in our sample, neither of which is classified as L-type (Cellino et al. 2014). (729) *Watsonia* itself is classified as Mk and further addressed in Chapter 14. Two out of four members of the (1400) *Tirela* L-type family are L-types in this classification (Mothé-Diniz and Nesvorný 2008). Curious is a bump in the probability distribution of L-types around the semi-major axis of Mars. About two thirds of the NEOs L-types are within 0.4 au of the martian semi-major axis. Some M-types are present among the JT and Hilda populations (refer to Fig. 13.2).

The top row of Fig. 13.2 gives the distributions of members of the S-complex. V-types are almost entirely concentrated around (4) *Vesta* (2.36 au, 6.4 deg) in the IMB (Binzel and Xu 1993) though there are V-types present in the MMB and even the OMB (Carruba et al. 2014; Oszkiewicz et al. 2017; Ieva et al. 2018). Concerning the few-member classes, the distribution of the R-types mimics that of V, and I dare not speak of a *distribution* in the case of the three members of the O-types, where (3628) *Boznemcova* and (7472) *Kumakiri* are joined by (52366) 1993 *FN39* based on a visible-only spectrum.

More interesting are the distributions of A and S. A-types are spread all over the entire Main Belt, including significant fractions in the OMB. This is in line with results from DeMeo et al. (2019) who find no statistically significant clustering of A-types in the Main Belt. The scarcity of the differentiated asteroids is known as the “missing-mantle” problem (DeMeo et al. 2015). If we look for remnants of differentiation, we may thus turn our eyes to families of possibly differentiated parent bodies. Oszkiewicz et al. (2015) suggest that the parent body of the *Flora* family may have been differentiated based on a fraction of V-type asteroids which reside dynamically stable within the family, suggesting that they are parts of the collisional history of the family. The right hand side of Fig. 13.4 shows the *Flora*-, *Phocaea*-, and *Koronis* families, the three S-type families with the most members classified in this taxonomy. Neither family shows any significant number of differentiated, i. e. A, M, or V interlopers. However, the small number of family members available here prevents any statistically sound conclusion with respect to the fraction of differentiated material in the family. In addition, the picture of the *Flora* family in Fig. 13.4 is in agreement with the taxonomic distribution Oszkiewicz et al. (2015) derived based on photometric data of 4225 family members. An open analysis step is to correlate the distribution of M- and S-types from this work after accounting for the observational biases of S-types in families, supported by the extension of the taxonomic scheme to include photometric colours.

13.4 TRENDS IN NEAR-EARTH SPACE

Q-types are present in the Main Belt yet the majority is found in the near-Earth space, refer to Figs. 13.1 and 13.2. They are particularly dominant on the excited orbits with large orbital eccentricities, in agreement with the planetary-encounter scenario leading to a rejuvenation of the asteroidal surface and the following change in spectral appearance (Binzel et al. 2004a). As DeMeo et al. (2014b) and Devogèle et al. (2019), I observe that Q-types dominate the low-perihelion regions where these planetary encounters are

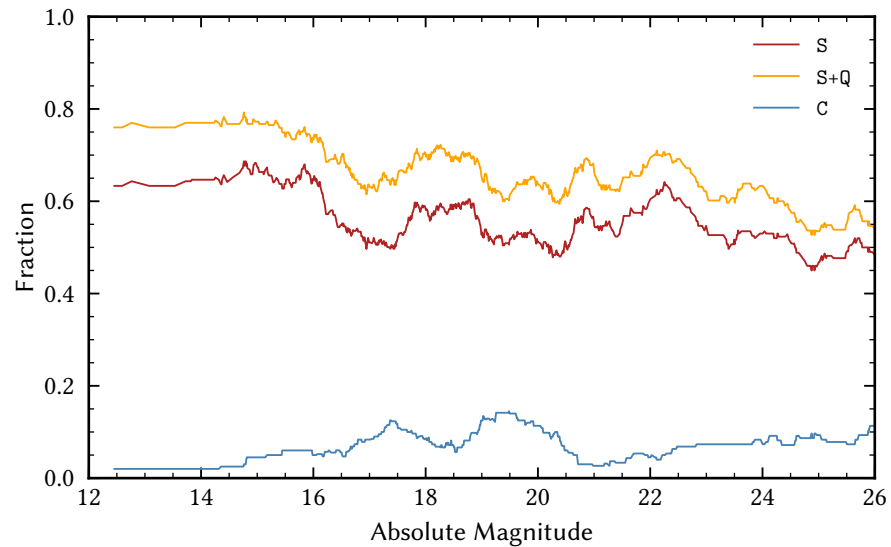


Figure 13.5: The fraction of S, C, and S and Q among near-Earth objects is shown as a running-average.

most likely. A peak is visible in the distribution of A-types around the semi-major axis of Mars. This is also in agreement with previous results showing that the olivine-rich Martian Trojans like (5261) *Eureka* are impact ejecta from the basaltic martian surface (Christou et al. 2017; Polishook et al. 2017).

As most reflectance spectra of NEOs are from large surveys such as MITH-NEOS (Binzel et al. 2019; Marsset et al. 2022), MANOS (Devogèle et al. 2019), and NEO-SHIELD (Harris et al. 2013), the agreement at first look between the results of the surveys and the distributions derived here are not surprising, including for example the fraction of Q to S for low perihelia or the increase in number of A-types around Mars. One aspect the surveys reported different results on is the fraction of S-types among NEOs with increasing absolute magnitude, where Binzel et al. (2019) report a near-constant rate of S at 60 % while Devogèle et al. (2019) observe a drop of 20 % between magnitude 16 and 26. I show the same fraction as a running-average in Fig. 13.5 for S, C, and S plus Q. The fraction of S-types observed here decreases from around 60 % to 50 %, hence, the combination of the survey data does not resolve the disagreement.

SPECTRAL CONNECTION OF ASTEROIDS AND METEORITES

14.1	A Primer on Spectral Matching	166
14.2	CV-, CO Chondrites and K-, L-Types	167
14.3	Equilibrated-, Unequilibrated OC and S-Types	177

The connection between meteorites and their asteroidal parent bodies is one of the fundamental goals of the compositional analysis of asteroids. Evident matches allow to infer on the asteroid properties in great detail through to the laboratory analysis of their counterparts.

In [Chapter 3](#), I outlined the most firmly established matches of asteroids and meteorite groups, which are (25143) *Itokawa* and the LL6 OC chondrites (Nakamura et al. [2011](#)), (162173) *Ryugu* and the CI chondrites (Yada et al. [2022](#)), as well the HED meteorites and (4) *Vesta* (McCord et al. [1970](#); Kelley et al. [2003](#)). The first two were matched by sample return missions while the third has overwhelming observational evidence in combination with the uniqueness of (4) *Vesta* among the large asteroids. The returned samples of (101955) *Bennu* are expected to match CM chondrites (Clark et al. [2011](#)). General consensus exists for asteroids with unique appearance in the spectral or albedo space, where the supposed parent body and the corresponding meteorite groups are similar to each other yet different to the rest of the populations: E-types and Aubrites, A-types and Brachinites, Pallasites, and R chondrites (DeMeo et al. [2022](#)).

Going from the general matching of classes and groups to the bodies on an individual level, Greenwood et al. ([2020](#)) estimated the number of parent bodies of the terrestrial meteorite collection between 95 and 148 based on differences in mineral abundances and isotopic anomalies. In [Chapter 1](#) and [Chapter 3](#), I outline the dynamical and compositional approach to establish asteroid-meteorite connections. I further highlight the uncertainty in the interpretation of asteroid spectra due to the unknown surface properties such as the degree of space weathering and the regolith grain size. In this chapter, I outline an approach of establishing asteroid and meteorite connections using spectroscopy while considering and accounting for the secondary effects. I focus on the proposed matches of the CV and CO chondrites with the K- and L-types and the Barbarians (Bell [1988](#); Sunshine et al. [2008](#); Devogèle et al. [2018](#)), as well as on the interpretation of S-types asteroids as remnants of the core or mantle of their parent bodies (Vernazza et al. [2014](#)). The analysis presented in [Sect. 14.2](#) is in preparation for a publication, while the results in [Sect. 14.3](#) have been presented in Eschrig et al. ([2022](#)). For the meteoritic perspective of the following analyses, the reader is referred to the PhD thesis of J. Eschrig ([2022](#)). I discuss a possibility to continue this work in the outlook in [Part V](#).

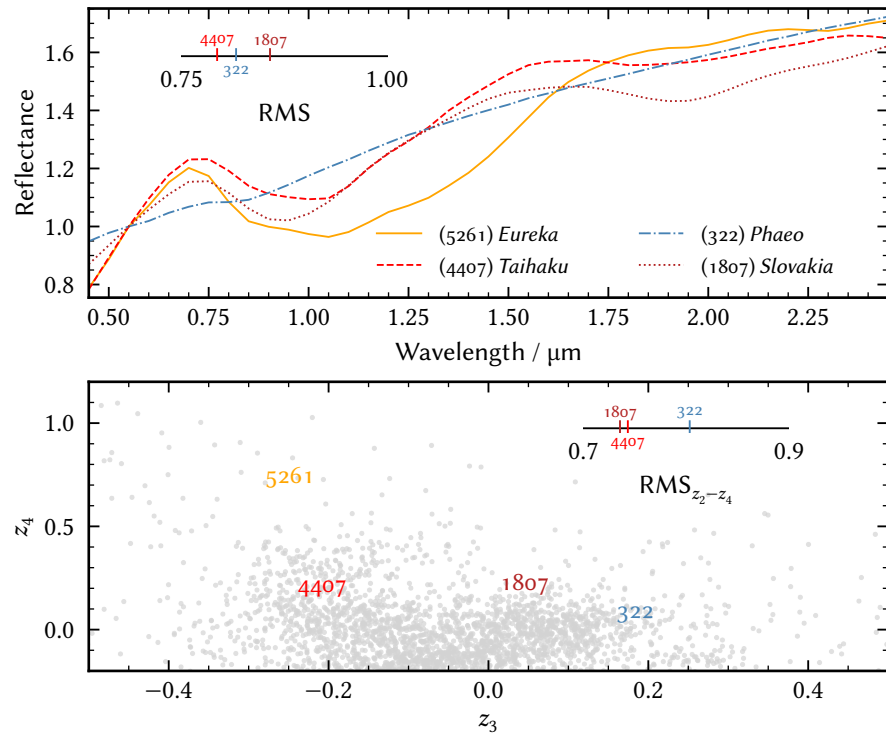


Figure 14.1: Spectral matching in data- versus latent space. *Top*: The spectra of S-types (1807) *Slovakia* (dash-dotted) and (4407) *Taihaku* (dashed) and D-type (322) *Phaeo* (dotted) are compared to the one of A-type (5261) *Eureka* (solid) using the root mean square of their differences. *Bottom*: The four spectra are shown projected into the latent space. The root mean square computation includes the second to fourth latent dimensions of the taxonomy. Data from DeMeo et al. (2009).

14.1 A PRIMER ON SPECTRAL MATCHING

A common metric for the similarity of two spectra X and Y is the root mean square (RMS) of their differences,

$$\text{rms} = \sqrt{\frac{1}{N} \sum_i^N (x_i - y_i)^2}. \quad (14.1)$$

This is equivalent to their Euclidean distance divided by the number N of wavelength bins. It has the advantage of being fast and easy to compute, yet there is a pitfall associated with this method. To illustrate this, I depict four spectra in the upper part of Fig. 14.1: A-type (5261) *Eureka*, S-types (1807) *Slovakia* and (4407) *Taihaku*, and D-type (322) *Phaeo*. If I visually determined the closest match to (5261) *Eureka*, the two S-types are the only contenders due to their strong olivine imprint. Yet, as indicated, all three spectra have similar RMS with respect to (5261) *Eureka*. The slope of (322) *Phaeo* matches the one of the A-type well, and over the whole observed range, they have a similar Euclidean distance as the two S-types have to (5261) *Eureka*.

A more suitable metric here should focus on the features and place less emphasis on the general shape, as done in the feature-based Bus-DeMeo taxonomy. The attentive reader of the previous chapters may guess where this is going: if the spectral matching occurs in the latent space rather than the data space, we can place the emphasis by combining different latent dimensions. At first, after projecting the spectra into the latent space, the situation has worsened as (322) *Phaeo* now has the smallest RMS with respect to (5261) *Eureka*. This is because, as we project the data, we place more emphasis on the slope as it dominates the covariance structure of the data. However, we may now compute the RMS while disregarding the first latent dimension. The result of the computation is shown in the bottom part of Fig. 14.1. (322) *Phaeo* has the largest RMS if we calculate it only over z_2 - z_4 . The figure further shows that if we disregarded another dimension, z_2 , we may even differentiate the slightly different pyroxene abundances between (1807) *Slovakia* and (4407) *Taihaku*, with the latter being the closest match to (5261) *Eureka*.

This simple example shows that the latent space enables us to place emphasis on certain features or the slope much better than we could do in the data space. More advanced metrics may even place different weights on each latent component, for example, if we aim to account for space weathering and its varying effects on pyroxene and olivine at different time scales, as discussed in Chapter 3.

14.2 CV-, CO CHONDRITES AND K-, L-TYPES

Since the first observations of the *Eos* family in the NIR as part of the 52 color asteroid survey and the following definition of the K-class, this class is linked to the CV3 and CO3 chondrites. Bell (1988) highlighted the spectral resemblance in terms of the absorption feature at 1 μm , the neutral and feature-less NIR spectral region, and their IRAS albedos around 10 %. A few years later, Burbine et al. (1992) linked the “anomalous” S-types (387) *Aquitania* and (980) *Anacostia* to CV3 and CO3 chondrites, noting that both asteroids lack the 1 μm -band while showing a prominent 2 μm -band, a configuration which is not observed in assemblages of olivine- and pyroxene or other common meteoritic minerals. The authors identify the aluminium-magnesium oxide spinel (MgAl_2O_4) as possible source of the 2 μm -band. This mineral is found in CAIs, of which the CV3 and CO3 chondrites show the largest abundance in terms of percentage of volume (Brearley et al. 1998). Burbine et al. (2001) further suggested the CO3 chondrite Warrenton as spectral analogue of the K-types (221) *Eos* and (653) *Berenike*, while (599) *Luisa* is linked to CV3 chondrite Mokoia.

At this point, there are two connections between CV3 and CO3 chondrites and the asteroid population, specifically, members of the K- and L-class as we refer to them since Bus and Binzel (2002a). The spectral appearance of (221) *Eos* as described by Bell (1988) is noticeably different to the one of (387) *Aquitania* and (980) *Anacostia* given by Burbine et al. (1992), yet both works are able to identify spectral similarities between the asteroids and unnamed meteorites within the group. The reason is that, just as we saw for

(980) *Anacostia* has come full circle as, in this taxonomy, it is considered to be an “anomalous” L-type and it is assigned into the S-class again. I suspect that the spectrum of (980) *Anacostia* used here suffers from the artificial 0.9 μm -feature due to the joining of the visible- and the NIR spectral parts, and it is in reality an L-type.

the K- and L-types in [Chapter 11](#), the spectral appearance of the CV3 and CO3 chondrites is variable yet continuous, as shown in [Figs. 14.2](#) and [14.3](#). Based on chemical and isotopic abundances, CV and CO chondrites may have belonged to a single parent body (Greenwood et al. 2010).

In [Chapter 3](#), I pointed out the cosmochemically important role that CAIs play in the reconstruction of the early history of the Solar System. The discovery of a class of asteroids which appears to have a large abundance of this material is therefore of high interest, and in a follow-up analysis to Burbine et al. (1992), Sunshine et al. (2008) concluded that both subjects of the former study, (387) *Aquitania* and (980) *Anacostia*, as well as (234) *Barbara* are even more enriched in CAIs than the CV3 and CO3 chondrites in the terrestrial meteorite collection. By modeling the asteroid spectra in terms of the spectra of different types of CAIs found in the CV3 chondrite Allende in combination with its matrix material, the authors identify CAI abundances of 20%–40%, considerably more than the abundance in CO3 (13%) and CV3 (10%) chondrites (Brearley et al. 1998). This implies that the L-type asteroids formed earlier compared to the other asteroids and in a special location or time in the protoplanetary disk (Sunshine et al. 2008), making them ideal probes of the early Solar System chemistry.

Another piece of the puzzle is observed by Cellino et al. (2006). (234) *Barbara* is reported to have an anomalously large inversion angle of the negative polarisation branch, refer to [Chapter 2](#). This behaviour is later observed in more objects, among them both (387) *Aquitania* and (980) *Anacostia* (Gil-Hutton et al. 2008; Masiero and Cellino 2009), and the term *Barbarians* is coined for this group of objects. The physical nature of this polarisation behaviour is unclear, suspected are unique properties of the regolith such as a large degree of concavities (Devogèle et al. 2017). Frattin et al. (2019) showed in the laboratory that CV and CO chondrites exhibit the same behaviour, further cementing the link between CV, CO and K-, L-types. Further noteworthy is the abundance of asteroids with long rotation periods among the Barbarians (Devogèle et al. 2017).

To summarise, we have two classes of asteroids which are linked to two classes of meteorites. Among the L-class, there are objects which depict an unusually large inversion angle of the negative polarisation branch, and these objects are presumed to be enriched in the important CAIs. Devogèle et al. (2018) combined the previous work with extensive observational efforts, employing both spectroscopy as well as polarimetry. The observational results confirm the link between the spectral appearance of L-types and the polarimetric behaviour, leading Devogèle et al. (2018) to conclude that all L-types are Barbarians and vice versa. Modeling the 2 μm -band using spectra of two CV3 meteorites, Allende and Y-86751, the enrichment of CAIs in the asteroids found by Sunshine et al. (2008) is confirmed and even superseded with up to 51% of CAIs for (3844) *Lujiayi*.

With all the pieces of the puzzle put together, I will now take a closer look at the resulting motive. Certainly, the claim of asteroids with large abundances of CAIs is the most important one and requires strong observational evidence. While Sunshine et al. (2008) based it on the spectral appearance of three asteroids, Devogèle et al. (2018) extended it to 28 asteroids and confirms

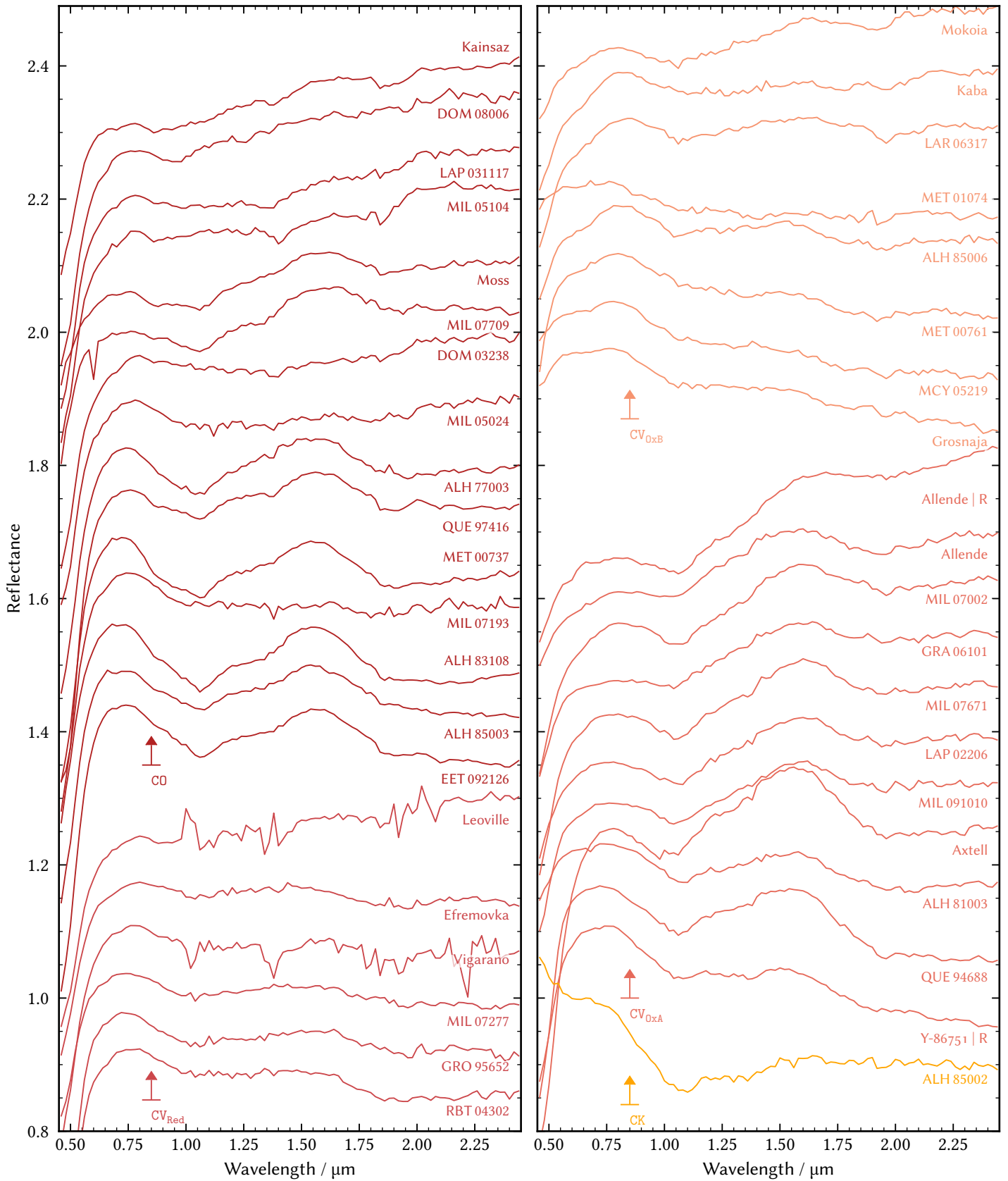


Figure 14.2: The spectra of the CK, CO, and CV chondrites used in this study. All but three spectra are from Eschrig et al. (2021). The spectrum of CK ALH 85002 is unpublished (Data: J. Eschrig). One spectrum of Allende and the spectrum of Y-86751 are from the RELAB database (IDs mptxh009 and c1mt71 respectively), indicated by the *R* behind their names.

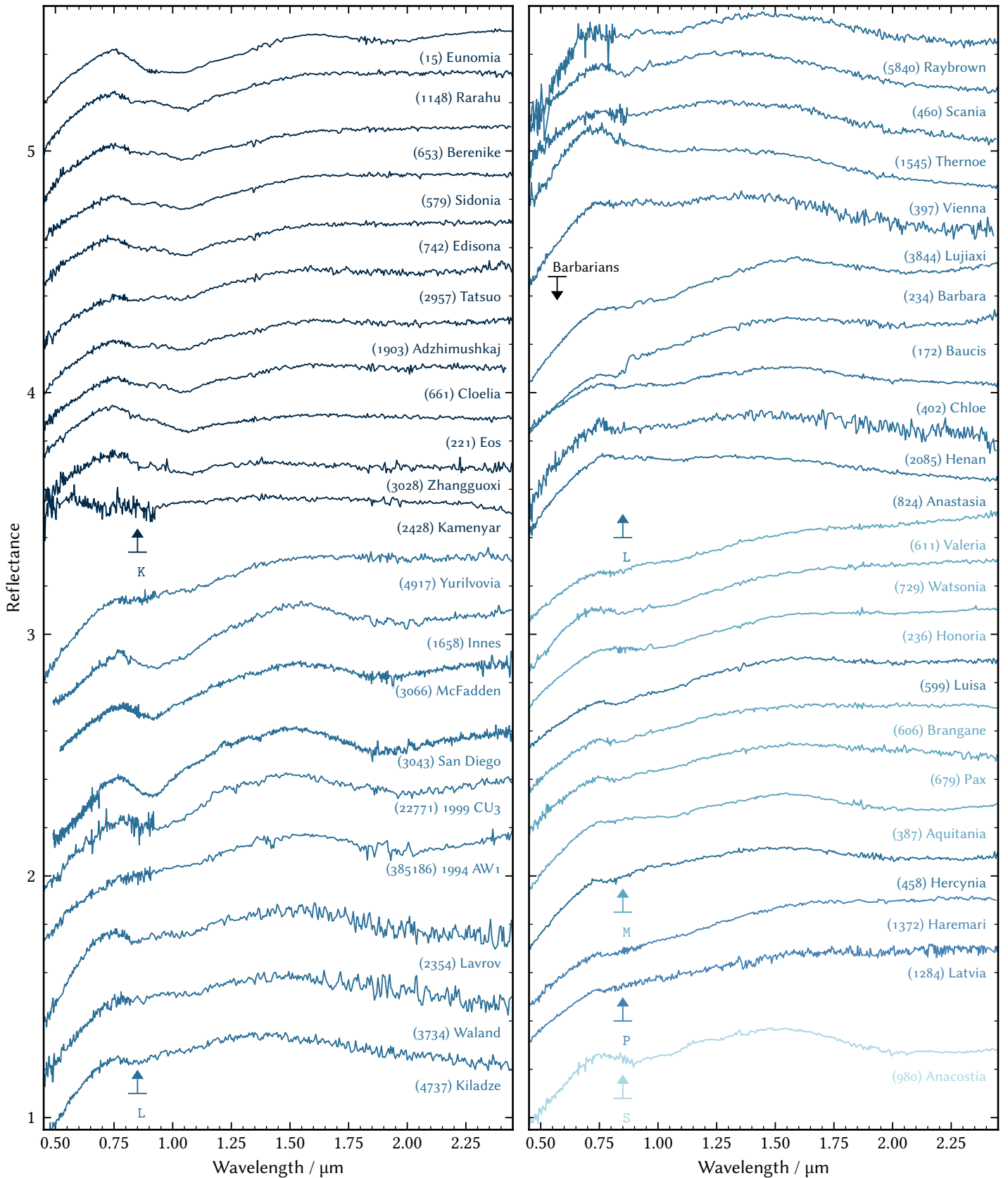


Figure 14.3: Spectra of K- and L-types and confirmed Barbarians. The corresponding names and numbers are always given below the spectrum. The Barbarians were selected based on the polarimetric behaviour given in Devogèle et al. (2018).

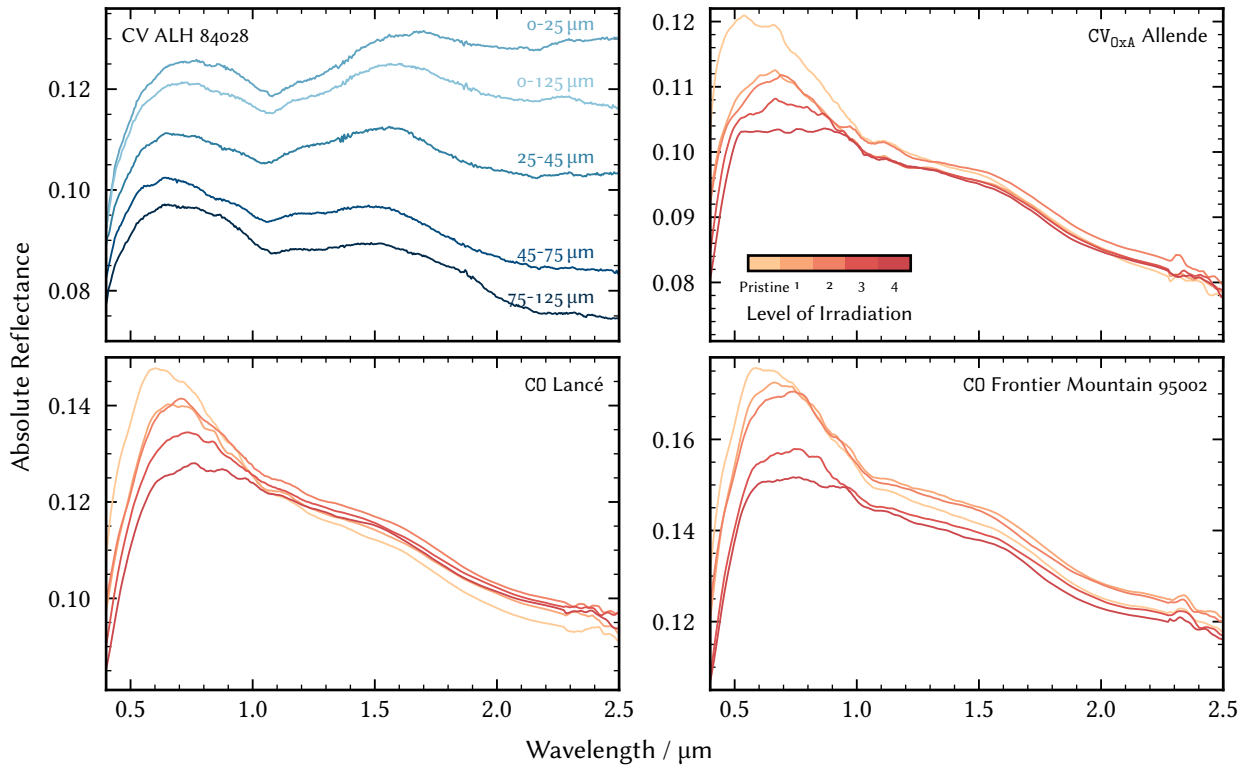


Figure 14.4: Spectral variability in CV and CO chondrites with grain size and space weathering. *Top Left:* The spectrum of powdered CV chondrite ALH 84028 depending on the grain size range in the powder. Data from Cloutis et al. (2012b). *Other:* The spectra of CV Allende (top right), CO Lancé (bottom left), and CO Frontier Mountain 95002 (bottom right) before and at different stages of irradiation with Helium ions. Data from Lantz et al. (2017).

the enrichment for almost all of them. However, in their modeling, both use only one (Allende, Sunshine et al. (2008)) or two (Allende and Y-86751, Devogèle et al. (2018)) chondrites as comparison spectra. Combining the spectral components of these two meteorites lead to the enrichment in CAI. Yet, in Devogèle et al. (2018), we see a drastic difference in the percentage of CAI in the models using Y-86751 compared to the ones using Allende, differences of up to 35%. A question arises: can we explain the spectral appearance of L-types without the need to invoke enrichment in CAI by considering different CO3 or CV3 analogues?

Eschrig et al. (2021) presented visible-near-infrared spectra of 15 CO3 and 23 CV3 chondrites, in addition to a so far unpublished CK chondrite (Data: J. Eschrig). All spectra were acquired in a consistent experimental setting at the *Institut de Planétologie et d'Astrophysique* in Grenoble, France. The CV chondrites are characterised in detail in Bonal et al. (2020), who show systematic differences between the subgroups built around the oxidation of the samples (CV_{Red} - reduced; CV_{OxA} - oxidised, Allende-like; CV_{OxB} - oxidised, Bali-like). The variability between these groups and both CO3 and CK3 spectra is apparent in Fig. 14.2: spectra of the OxA subgroup show more prominent 1 μm bands, while both oxidised subgroups show more variability in their spectral slope than the reduced subgroup. Finally, I add the spectra of the bulk

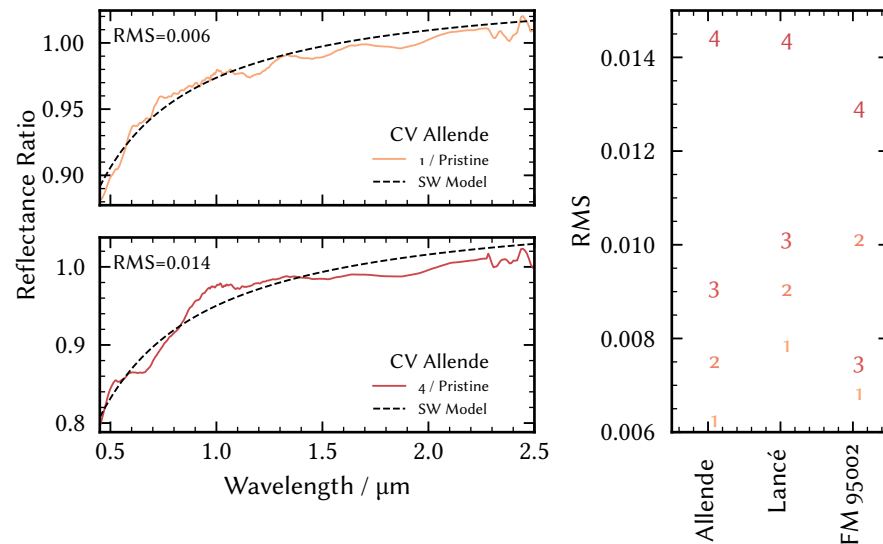


Figure 14.5: Evaluation of exponential space-weathering model through irradiation experiment. *Left*: The ratio of the spectra of CV Allende after irradiation of degree 1 (top) and degree 4 (bottom) to the pristine spectral appearance (solid) is fitted with the exponential space-weathering model (dashed). *Right*: The root mean square difference between the model fits and all ratio-spectra computed for CV Allende and COs Lancé and Frontier Mountain 95002. The average root mean square is 0.0095. Data from Lantz et al. (2017).

Allende and Y-86751 samples used in Sunshine et al. (2008) and Devogèle et al. (2018).

In Fig. 14.3, I display the spectra of confirmed Barbarians following Devogèle et al. (2018) in addition to spectra of asteroids classified as either K or L in the taxonomy presented in this work. I only consider VisNIR spectra with a high signal-to-noise based on visual inspection. The spectra of (3734) *Waland*, (2354) *Lavrov*, and (2085) *Henan* were on the edge but I believe the overall structure of the 2 μm is recognisable. Notably, the subjects of Burbine et al. (1992), Burbine et al. (2001), and Sunshine et al. (2008) are included in this sample: (221) *Eos*, (234) *Barbara*, (387) *Aquitania*, (599) *Luisa*, (653) *Berenike*, and (980) *Anacostia*. Note that the Barbarians contain objects classified as M, P, and S. This is a first discrepancy to the results of Devogèle et al. (2018), who concluded that all L-types defined following the Bus-DeMeo systems are Barbarians and vice versa. Considering the spectra shown in Fig. 14.3, in particular those of (1372) *Haremari* and (1284) *Latvia*, this does not hold for our sample. This again underlines the distinction between taxonomy as a guideline and a mineralogical analysis as a necessary follow-up for interpretation given in Chapter 4.

I now seek to identify potential matches between the meteorites depicted in Fig. 14.2 and the asteroids shown in Fig. 14.3. Sunshine et al. (2008) and Devogèle et al. (2018) derived the anomalously high abundance of CAI on the Barbarian asteroids using radiative transfer models. I do not intend to derive endmember abundances in the asteroid spectra. Instead, I want to stress the

I deem visual inspection of the signal-to-noise valid here as I have spent three years training my personal neural network for just this task.

The description of the radiative transfer modeling procedure and sources of the endmember spectra and optical properties given in Devogèle et al. (2018) is great to learn about these steps.

need to account for the meteoritic variability in these models. As such, I use the simple assumption that asteroids are space-weathered meteorites. If I can identify a meteorite which is spectrally similar to an asteroid in this sample after accounting for space-weathering, I may establish the CV, CO chondrite and K-, L-type connection without invoking an abundance of CAI which is not represented in the meteorite collection.

To model the space-weathering, again, one might evoke radiative transfer modeling. I choose the experimental approach, making use of the data presented in Lantz et al. (2018) on irradiation experiments with CCs. One CV (Allende) and two CO (Lancé and Frontier Mountain 950002) chondrites formed part of this experiment. The spectra of the pristine samples and the samples after four different levels of irradiation with Helium ions are shown in Fig. 14.4. The overall blue slope at wavelengths larger than 0.6 μm is accounted to the sample preparation as pellets rather than as powder (Lantz et al. 2017). All samples get redder and darker with increasing irradiation.

Further shown in Fig. 14.4 is the spectral appearance of powdered samples with different ranges of particle grain sizes of CV ALH 84028 as presented in Cloutis et al. (2012b). The spectral appearance of powders is dominated by the smallest particle sizes (Mustard and Hays 1997), which is apparent in the similarity of the samples containing grain sizes between 0 μm –125 μm and 0 μm –25 μm . The spectra get bluer with an increasing lower limit of the grain size, while the band structure appears to be less affected. Unfortunately, the grain size structure on asteroid surfaces is not well constrained apart from the general idea that larger asteroids should be able to retain smaller grains owing to their increased gravitational strength (Delbó et al. 2007). In Eschrig et al. (2022), we showed that the variation of the depth of the 1 μm -band in S-type asteroids is consistent with this trend of smaller grain sizes for larger asteroids. Nevertheless, the minimum grain size is clearly a factor to account for in the spectral appearance of asteroids and meteorites.

Getting back to modeling the effect of space weathering, the irradiated spectra of the three chondrites shown in Fig. 14.4 can be used to assess the spectral changes. As discussed in Sect. 3.5.1, a common approximation for the modulation (R_{SW}) of a spectrum due to space weathering is an exponential model (Brunetto et al. 2006),

$$R_{\text{SW}} = ke^{c/\lambda}, \quad (14.2)$$

where k and c are free scale parameters and λ is the wavelength at which the reflectance is evaluated. The model therefore assumes that the asteroid spectrum is equivalent to the meteorite spectrum times an exponential scale. To assess its suitability, I divide the spectra acquired by Lantz et al. (2017) of the irradiated surfaces of the three meteorites by the spectra of the pristine surfaces and compare the ratio to the model. On the left hand side of Fig. 14.5, I show the example of the surfaces of Allende after brief (level 1) and prolonged (level 4) irradiation divided by its pristine appearance and the fitted exponential model. There are some residuals due to the band structure of Allende yet the overall trend is well matched by the exponential, in particular for the less-irradiated surfaces. The right hand side of Fig. 14.5 gives the RMS of the model fit to all combinations of the pristine and irradiate surfaces

of the three chondrites Allende, Lancé, and Frontier Mountain 95002. The mean RMS of the fits is 0.0095. Therefore, if we identify matches between asteroids and meteorites with RMS value around 0.01, the difference might sufficiently be explained by the space-weathering model approach without invoking discrepancies in mineral abundances.

I repeat the fitting procedure for the asteroid and meteorite spectra, dividing the former by the latter and fitting the ratio with the exponential model in Eq. (14.2). Terrestrial weathering affects in particular the visible spectral region of meteorites due to the formation of iron oxides (Salisbury and Hunt 1969; Salisbury and Hunt 1974), thus, the wavelengths below $0.7\ \mu\text{m}$ are excluded from the fit. The RMS of all pairs of spectra are given in Fig. 14.6. The best match for each asteroid is shown by the filled white circles in Fig. 14.6, while the second-best match is indicated with an open circle. Six meteorites match well with six or more asteroids. The colour-coded values show that the K-types and non-L Barbarians are best matched by different CV and CO chondrites, while L-types (Barbarians and non-Barbarians) are best matched by three CV_{0xA} chondrites: Axtell, QUE 94688, and Y-86751. The RMS for these fits is within the RMS of the space weathering model mismatch. Thus, we identify meteorite analogues for L-types which do not require an anomalously large CAI abundance. The best matches for some Barbarians and K-types are shown in Fig. 14.7.

For K-types, the Allende spectrum from RELAB as well as the only CK chondrite, ALH 85002, are generally good matches. However, the spectrum of Allende is redder than the asteroid spectra, meaning that the space weathering is required to decrease the spectral slope, i. e. $c > 0$ is required in Eq. (14.2). In fact, of the lowest RMS matches among K-types and meteorites, we observe equal tendency for spectral reddening and spectral blue-ing. For CC, space weathering may lead to both a reddening and a blue-ing of the surface (Lantz et al. 2017; Thomas et al. 2021), hence, we do not reject these solutions even though the CV and CO chondrites in Fig. 14.5 do not show this behaviour. Excluding meteorite matches which require K-types to get bluer with increasing space weathering based on the three experimental results shown in Fig. 14.5 might lead to the same narrow conclusion as for the CAI content of L-types. A larger sample of CV and CO chondrites should be analysed in irradiation experiments before making this definite claim, especially as the initial, pristine spectra are noticeably bluer than the meteorite spectra in Fig. 14.2 due to the pellet-structure of the samples in Lantz et al. (2017).

I further emphasise the value of a large number of asteroid and meteorite reflectance spectra when investigating possible links. This value is apparent in the different RMS values that the two spectra of Allende show in Fig. 14.6. Using either spectrum, we may conclude that K-types are matched or not matched by Allende.

The question remains as to why some meteorites match a large number of asteroids so well. The answer lies in the insufficient space weathering model, which selects matches where the space weathering is exponential. This is likely not the case for the majority of asteroids, and the ones where the model fortuitously matches the closest are preferentially selected. This does not

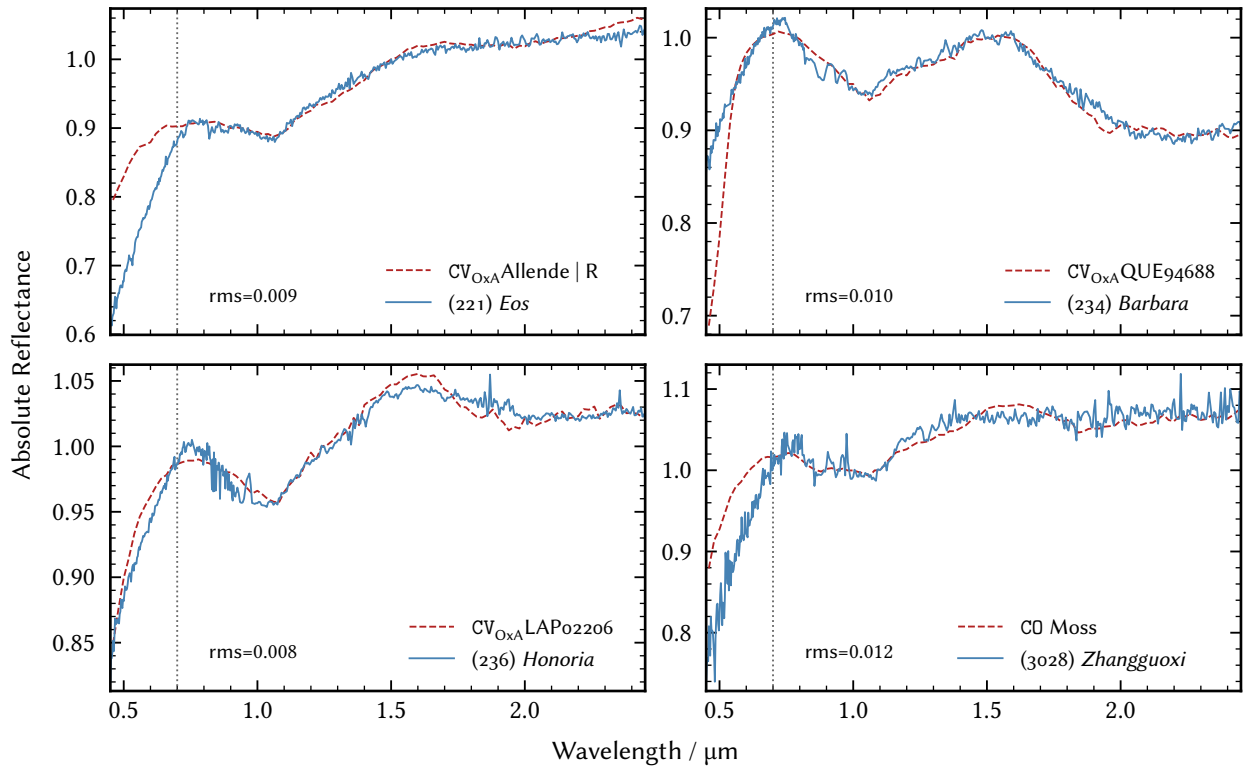


Figure 14.7: Example matches identified between CV, CO chondrites and K-, L-, and M-types. The asteroid spectra (solid) are deweathered using an exponential model fit to the ratio of the asteroid to the meteorite (dashed) spectra. The root mean square is calculated for wavelengths larger than $0.7 \mu\text{m}$, indicated by the vertical line.

spectra as possible to avoid narrow conclusions. Furthermore, not all L-types are Barbarians and not all Barbarians are L-types.

Finally, we might ask ourselves what could be a model-independent way to identify possible matches. Given experimental data of space weathering, I believe that we can improve upon existing methods by searching for matches in latent space, as outlined in Sect. 14.1. By mapping the experimental results into the latent space, we obtain a description of the space weathering without setting an absolute scale. A proof-of-concept is shown in Fig. 14.8, where the vectors of grain size and space weathering are computed from the data presented in Fig. 14.4. I use the difference between the spectra of $25 \mu\text{m}$ – $45 \mu\text{m}$ and $75 \mu\text{m}$ – $125 \mu\text{m}$ as vector of increasing grain size, whereas for the space weathering, I give the mean change in latent space between the pristine and the fully irradiated (level 4) spectra of the CV and CO chondrites.

Figure 14.8 shows that the asteroids and meteorites have a comparable variability in the latent space. In general, asteroids lie in the trajectory of the space weathering from the meteorites. The vectors of grain size and space weathering are opposite in terms of slope and orthogonal in terms of the strength of features, indicating that these effects might be separated if we have more samples of irradiated spectra with different sample sizes. Regarding the distribution of the K-types, we see in the upper right that they generally have larger slopes (indicated by an increasing z_1 -score) than

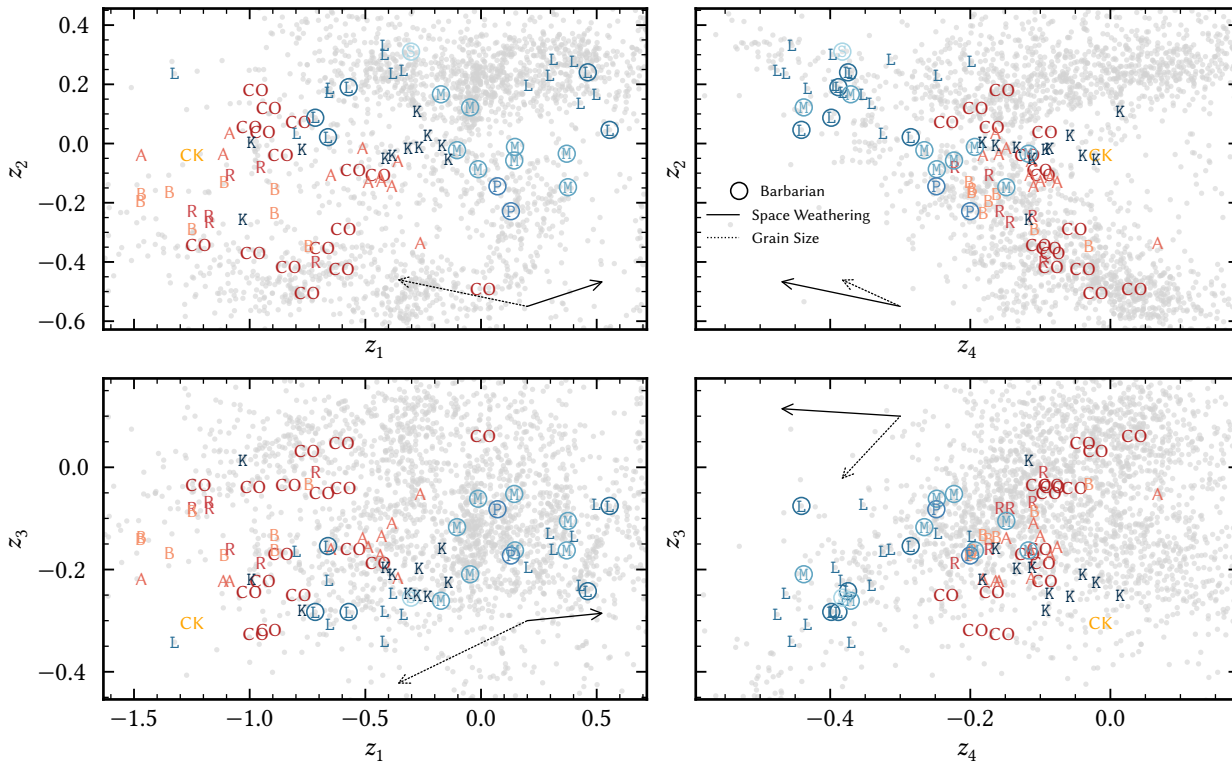


Figure 14.8: Distribution of CV and CO chondrites and K- and L-types asteroids in the latent space of the taxonomy presented in Mahlke et al. (2022). The positions of asteroids and meteorites are given by their respective colour-coded class letters. Asteroids which are Barbarians have circles around their class letters. CV chondrites are represented by their subgroups: 0xA - A, 0xB - B, reduced - R. The two vectors in each part show the directions of increasing space weathering and grain size found experimentally based on the spectra given in Fig. 14.4. Dotted, grey lines indicate the asteroids and meteorites of the four matches shown in Fig. 14.7.

the meteorites, indicating that they redden with space weathering. In the lower right part, the strength of the olivine $1\ \mu\text{m}$ is encoded in the z_4 -score. We observe here that the K-types have equal or larger z_4 -scores than most meteorites. Considering the experimental result that the band depth of olivine features decreases with increasing irradiation (Chrbolková et al. 2021), this rules out the vast majority of CV and CO chondrites as matches for most K-types, leaving CK ALH 85002 and the RELAB spectrum of CV Allende as only potential matches. As the latter requires spectral blue-ing with space weathering, the complete picture points towards the CK chondrites as best matches for K-types. However, as said above, there is only one CK chondrite in the sample and no space weathering results are presented for this group. The final conclusion thus requires more data on the CK chondrites.

14.3 EQUILIBRATED-, UNEQUILIBRATED OC AND S-TYPES

The link between S-type asteroids and OC meteorites is undisputed, as outlined in Chapters 3 and 4. Open questions, however, are the finer details of

this connection, such as: How many parent bodies are sampled by the OC collection? What is the ratio of H-, L-, and LL-material among S-type asteroids in the Main Belt? How many planetary surfaces show signs of significant thermal metamorphism? The answers to these questions have implications on the mineralogical interpretation of observations of individual S-types as well as on the formation of the ensemble of planetesimals.

There are strong biases in the terrestrial OC collection with respect to their mineralogy and thermal metamorphism. Unequilibrated Ordinary Chondrites (UOCs), i. e. OC of PT 3–3.9, make up 6 % of all OCs following the inventory of the Meteoritical Bulletin Database.¹ Equilibrated Ordinary Chondrite (EOC) of PT 4 – 7 are therefore far more common, indicating that the parent bodies of OC in general experienced thermal metamorphism after or during their formation. Furthermore, as mentioned in [Chapter 3](#), the ratio of H : L : LL OCs is in a ratio of 3:3:1.

For S-type asteroids, reflectance spectroscopy of NEOs suggests that the majority are LL-like (Vernazza et al. 2008; León et al. 2010; Thomas and Binzel 2010; Dunn et al. 2013). Vernazza et al. (2014) identify two compositional groups among S-type asteroids based on the ratio of olivine to low-calcium pyroxene, asteroids like (6) *Hebe*, which resemble equilibrated H chondrites, and (8) *Flora*-like objects, which have an LL-like mineralogy. The implications of these interpretation are vast, from the homogeneity of OC-like planetesimals to their formation time and duration in the early Solar System (Vernazza et al. 2014). Hence, they are worth iterating over.

Eschrig et al. (2022) present reflectance spectra of 39 EOCs and 41 UOCs, shown in [Fig. 14.9](#). The meteorite samples were characterised mineralogically and petrographically to confirm or update the assigned OC group and PT from the literature (Bonal et al. 2016). The consistent sample preparation (e. g. grind time and -mechanism to create powdered sample) and characterisation (e. g. measurement under vacuum and at 80 °C, measurement geometry) produce a dataset of high value as these aspects vary significantly among sample spectra available on platforms such as RELAB. The importance of a consistent dataset is outlined in Eschrig et al. (2022).

As presented in Eschrig et al. (2022), I make use of this dataset to investigate whether a mineralogical or petrographic trend is apparent in the S-type asteroids. To this end, I use 321 VisNIR spectra of 242 individual asteroids classified as S-types in this work. The spectra of the 331 asteroids are shown in the bottom part of [Fig. 14.9](#). Visual comparison of the asteroid and meteorite spectra in [Fig. 14.9](#) show that (1) the visible part of the meteorite spectra is affected by terrestrial weathering, hence, I truncate all spectra below 0.65 µm the following analyses, (2) the EOCs depict a large variability specifically in their 1 µm-band in comparison to both UOCs and S-types, and (3) S-type asteroids are homogeneous and, in comparison to the meteorites, much redder (not the different y-scales between the asteroid and meteorite parts in [Fig. 14.9](#)).

By now, we are well aware that the reflectance spectra of asteroids are altered by unknown degrees of space weathering. Furthermore, the unknown distribution of grain sizes on the planetary surfaces affect their spectral inter-

In the analysis presented in Eschrig et al. (2022), I used the reflectance spectra of asteroids which are classified as S-types in the literature rather than in this taxonomy. However, as we are dealing with VisNIR spectra, the differences of these samples are marginal.

¹ <https://www.lpi.usra.edu/meteor/metbull.php>

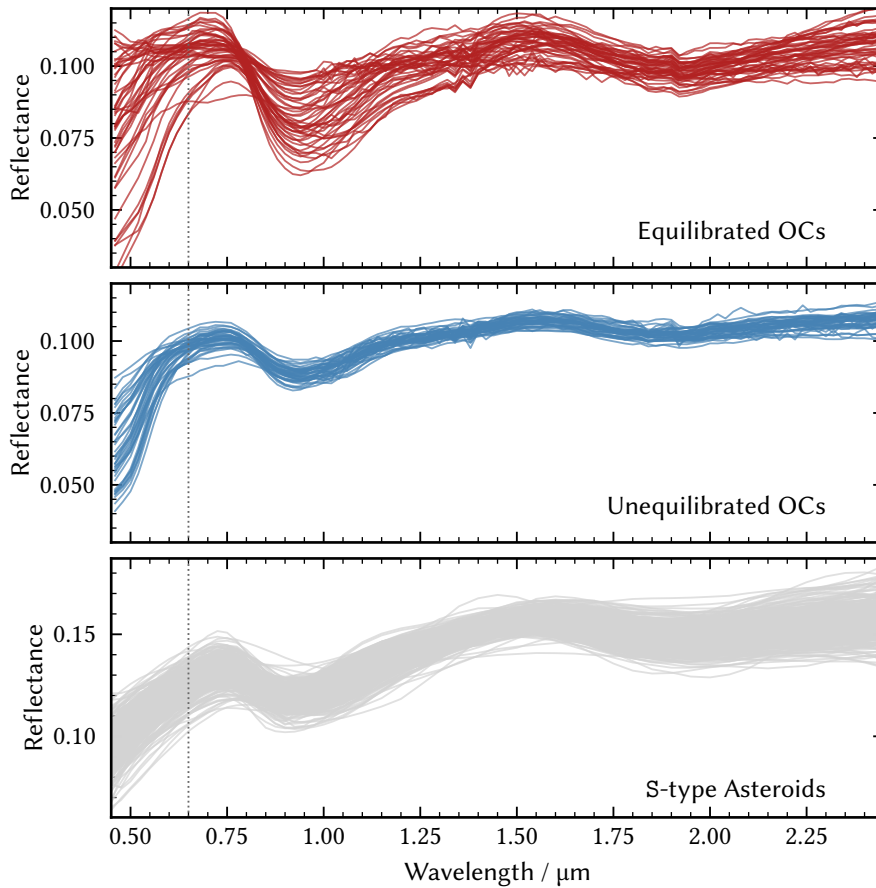


Figure 14.9: The reflectance spectra of equilibrated (*top*) and unequilibrated (*middle*) ordinary chondrites from Eschrig et al. (2022) and S-type asteroids (*bottom*) as classified in this work. The vertical dotted lines give the limit at $0.65\ \mu\text{m}$ below which the spectra are truncated due to the effect of terrestrial weathering of meteorites on the visible region.

pretation. As in the previous section, I propose to express these alterations in terms of vectors in the reduced space rather than using absolute values derived e. g. from radiative transfer modelling.

I therefore compute the PCA of the S-type reflectance spectra truncated below $0.65\ \mu\text{m}$. The result is shown in Fig. 14.10, where the distribution of their principal scores is shown as grey circles. As the spectra are demeaned prior to the projection, the origin of the reduced space represents the mean S-type spectrum. The homogeneity of S-types is apparent: their distribution appears Gaussian in all three latent scores shown in the top left and bottom panels of Fig. 14.10. The top right panel of Fig. 14.10 shows the PCs and their explained variances in percent. The first component resembles a slope with an olivine-like $1\ \mu\text{m}$ structure, while the second and third most resemble pyroxene and olivine spectral components.

Further shown in Fig. 14.10 are the principal scores of the OCs spectra projected into the principal space. Two expected effects are visible: the OCs scores are offset from the origin due to the space-weathered appearance of

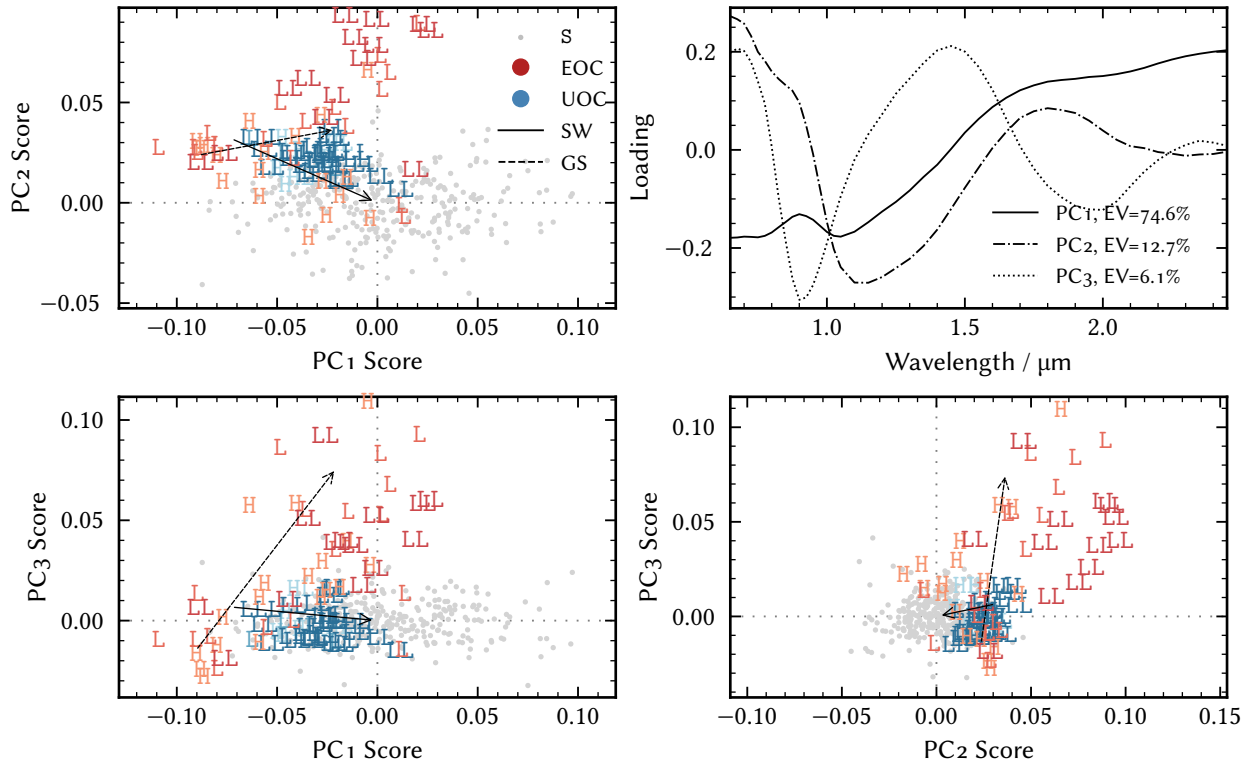


Figure 14.10: Principal component analysis of equilibrated and unequilibrated ordinary chondrites and S-type asteroids. *Top right*: The principal components derived from the S-type spectra shown in Fig. 14.9. *Other*: The principal scores of the S-types (circles) and the OC chondrites are given, where the latter are colour-coded by their state of equilibration and indicated using the letters of their respective chondrite group. Vectors of increasing space weathering (SW) and grain size (GS) are derived as explained in the text. Adapted from Eschrig et al. (2022).

the asteroids and the UOCs scores show less variance than the ones of the EOCs, as expected from the visual comparison of the reflectance spectra in Fig. 14.9.

A vector representing increasing degrees of space weathering is drawn by computing the mean S-type and deweathering it using the exponential model from Brunetto et al. (2006). The vector connects the mean, original S-type spectrum, i.e. the origin of the reduced space, to the score of the deweathered mean spectrum projected into the reduced space. Note that I assume here that space weathering affects all S-type surfaces in the same manner, described by the direction of the space weathering vector. The norm of the vector is arbitrary as the degree of space weathering is not uniform for S-type asteroids due to different surface ages (Thomas et al. 2012). The vector aligns visually with the distance between asteroids and meteorites, as expected.

To evaluate the effect of grain size, the powdered samples of six EOC and one UOC are ground further using a laboratory ball mill for first 5 min and second 10 min, recording a VisNIR reflectance spectrum for each sample after each additional step (e.g. Eschrig et al. 2022). The vector for increasing grain size is based on the average scores of the six EOCs which were additionally

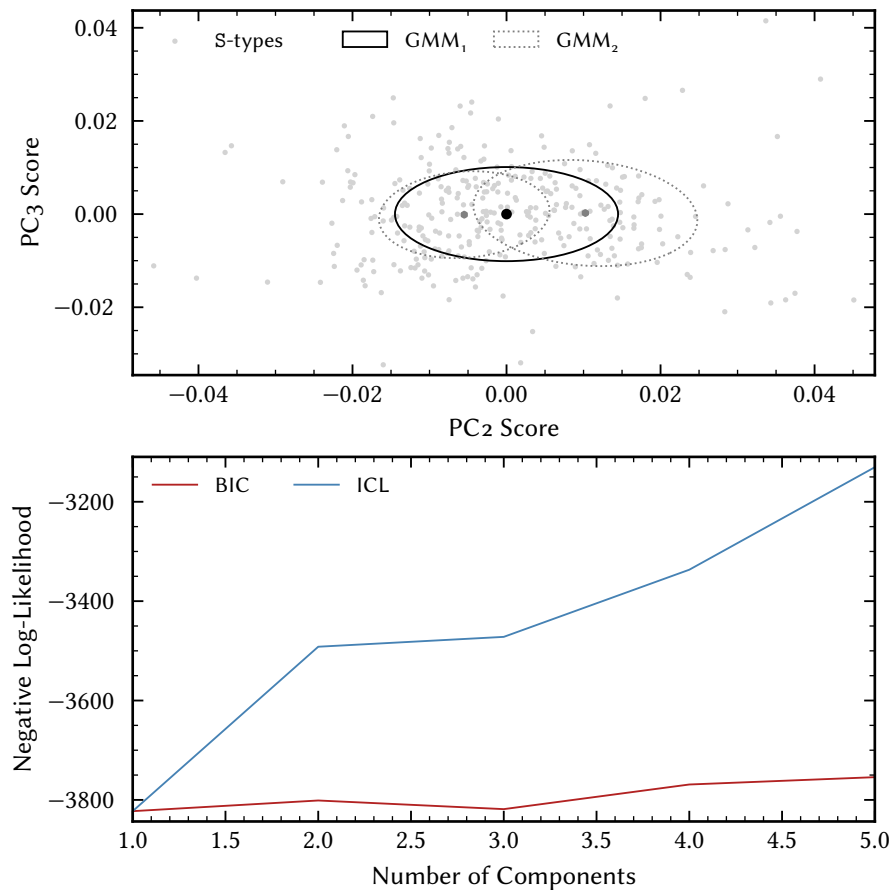


Figure 14.11: Investigation of modality of S-type principal scores. *Top*: A unimodal (black, solid) and a bimodal (grey, dotted) Gaussian Mixture Model are fit to the distribution of S-type asteroids (circles) in principal space. *Bottom*: The Bayesian Information Criterion (BIC) and Integrated Completed Likelihood (ICL) model selection criteria for Gaussian mixture models with one to five components. Lower values indicate a higher likelihood of the data given the model.

ground, where the base of the vector represents the average score of the fine-grained samples and the head the average score of the coarser grains. In all three principal scores, the vector suggests that smaller grain sizes for EOCs would lead to a better overlap. This is in agreement with the visual discrepancy between EOCs and S-types in Fig. 14.9, as, with decreasing grain size, silicate absorption features get shallower (Mustard and Hays 1997; Sultana et al. 2021). I only show the grain size trend for EOCs as a similar measurement is only available for one UOC.

Figure 14.10 shows that matching the distributions of UOCs and S-types is trivial and the offset is well explained by space weathering. However, the variance of the UOCs distribution is smaller than that of the S-types. For EOCs, there are some satisfactory matches with S-types when only considering space weathering. However, to fully match the EOC and S-type distributions, the former require smaller grain sizes than in the sample presented

here. A dependency of the OC scores on their group is visible in $PC1$ and $PC3$ for UOC, where H-like OC have lower $PC1$ and larger $PC3$ scores. For EOC, trends are apparent in all three principal scores, suggesting that the mineralogy of equilibrated S-type surfaces may be reliably established via remote-sensing.

Any comparison with results from Vernazza et al. (2014) has limited interpretability due to the different grain sizes of the meteorite samples. The UOCs samples presented in Vernazza et al. (2014) were sieved, which removes larger grains. They will therefore show smaller features than the samples presented in Eschrig et al. (2022). Instead, I investigate the bimodality of the S-types proposed in Vernazza et al. (2014), making use of the three times larger sample of asteroids here. 77 out of the 83 S-types in the Vernazza et al. (2014) study are present in this dataset. To test the bimodality, I fit the principal scores of the S-types in the second and third latent dimensions with GMMs. These dimensions resemble pyroxene and olivine spectra (upper right part of Fig. 14.10) and are reduced in terms of the slope, which is encoded in the first principal component. The model fit is shown in Fig. 14.11. The upper part shows the principal scores of the S-types as circles and the fitted Gaussian components of a GMM with one mixture component (black, solid) and one with two mixture components (grey, dotted). Both models visually describe the distribution of S-types adequately. To evaluate the model fits, I compute the BIC and ICL parameters of the data for GMMs models with one to five mixture components. Both parameters serve to select the best model for the data (Bouveyron et al. 2019). The BIC parameter computes the likelihood of the data given the model parameters and applies a penalty for the number of free model parameters. The ICL adds a further penalty on the similarity of the resulting mixture components, therefore always selecting a model with fewer mixture components than the BIC. The BIC and ICL values for the five GMMs fits to the principal scores are shown in the lower part of Fig. 14.11. I vertically displace the BIC scores to match the first ICL score for comparability. As only the relative values among each selection criterion are of interest, this displacement does not affect the result. Lower values indicate a larger likelihood of the data given the model. We see that both criteria favour a single-component model, i. e. suggest that the data is unimodal. The BIC shows little difference between a uni-, bi-, or trimodal distribution while the stricter ICL criterion clearly favours the unimodal distribution. The upper part of Fig. 14.11 shows the fits of a uni- and a bimodal GMM to the principal scores of S-types. The two components of the bimodal fit overlap significantly, leading to a large penalty in the ICL score.

To summarise, the results of the PCA thus suggest that S-types largely resemble UOCs, both in the band structure as well as in the overall variability of each population. This in turn suggests that the terrestrial collection of OCs is subject to extreme biases. However, the large dependence of EOC spectra on the grain size indicates that planetary surfaces with finer grains than the powdered samples used here are EOC-like material. I do not identify a bimodality among S-types.

Part V

Conclusion

CONCLUSION AND OUTLOOK

Over the past three years and 185 pages, I have dedicated myself to revising and improving the asteroid taxonomy. I motivated several improvements over the existing method and showed how they may be realised in [Chapter 5](#). The development of asteroid science towards a data-driven research field and the high value of a unified access point for observationally derived parameters is shown in [Chapter 6](#). The analysis of asteroid phase curves from ATLAS shows that the full potential of non-targeted observational campaigns is not yet fully exploited.

I further discuss in detail the importance of the selection of observables to build a taxonomy and the large amount of effort required to build a homogeneous dataset out of a heterogeneous foundation of reflectance spectra in [Chapter 9](#). The probabilistic approach to the taxonomy is discussed in [Chapter 10](#), giving rise to the 17 classes of asteroids presented in [Chapter 11](#). Preliminary analyses of the compositional distribution of the 4526 individual asteroids taxonomically classified here and the application of the classification method to the problem of establishing connections between asteroids and meteorites based on reflectance spectroscopy are given in [Chapters 13](#) and [14](#).

While this point in time marks the end of my doctoral research, it does not conclude my ambition to pursue open issues and points of interest which arose during this time. I give some thoughts on how to continue below.

COMPOSITIONAL ANALYSIS OF POPULATIONS

To ensure that the taxonomy is built on a *rock*-solid foundation, I reiterated over every step of the taxonomy multiple times. This came at the cost of the following, objectively more interesting compositional analyses of the results.

The Ambiguity of the M-Complex I have outlined in detail the conundrum that the M-class poses when comparing observations of its members in almost any two pairs of observables. The continuous and partly degenerate appearance of the K and L types adds to the confusion.

A global study of K, L, and M in more observables, in particular those highlighting metal content such as the radar albedo (Clark et al. 2004; Ockert-Bell et al. 2010; Shepard et al. 2015) and the polarimetric parameters for the mysterious *Barbarian* population (Cellino et al. 2006, 2014; Devogèle et al. 2018) within this complex is key to unravelling this mystery. Further clues may be hidden in *Watsonia*-family (Cellino et al. 2014), as family of Barbarians with different spectral appearances, and the newly-discovered *Kalliope*-family (Brož et al. 2022), as family of the second largest M-type and the first family dynamically associated to a M-type (Davis et al. 1999).

A (possibly subjective) noticed rise in the interest of ATLAS observations for phase curves at this years Europlanet Science Congress makes me believe that this message was noted.

The BCP Continuum This work included the first definition of the P-class since Tholen (1984) using the most complete observational record of these objects and still, I found it quite challenging to differentiate the core of the canonical P population from the C types. A similar confusion between what constitutes a B- or a C-type is apparent in the literature (DeMeo et al. 2009; Clark et al. 2010; León et al. 2012), and it may have been more fortune than observational evidence which provided me with a clear separation of B and C in the clustering analysis of this taxonomy.

Understanding the ambivalent influence of space weathering on these objects is a prerequisite to disentangle their distributions and understand the transitions (Thomas et al. 2021). In combination with laboratory, theoretical, and observational studies on the effect of varying grain sizes (Cloutis et al. 2011; Gundlach and Blum 2013; Vernazza et al. 2016), we may then better identify what changes in the reflectance spectra are of mineralogical nature and what changes may just be phenomenological, inherent to the compositional class itself. A proof-of-concept is shown in Chapter 14 where the vectors of space weathering derived from laboratory experiments (Lantz et al. 2017) help interpret trends among consistent samples of meteorite reflectance spectra (Eschrig et al. 2021).

The Nature of Z-types What is the nature of the Z-types? Are they implanted TNOs (Hasegawa et al. 2021b), is their red colour caused by spectral weathering, or is there another explanation for their dispersion over the Main Belt?

The first, short-term step is to identify more Z-types and thereby open the door for statistical analyses. The search for more Z-types in the photometric archives of the SDSS and VISTA surveys or in the third data release of *Gaia* (Popescu et al. 2016; Sergeev and Carry 2021; Galluccio et al. 2022).

One possibility to investigate a trans-Neptunian origin is to search for binaries among the Z-types. Fraser et al. (2017) show that Edgeworth-Kuiper Belt Objects (EKBOs) form as similar-sized binaries, while Nesvorný et al. (2018) used the survival of the binary system of (617) *Patroclus* during its migration from the outer to the inner Solar System to time the planetary migration. Identification of a large fraction of binaries among the Z population may thus support the implantation hypothesis. To investigate the space weathering hypothesis, the diameter and semi-major axis distributions of Z and D should be compared, in addition to the space weathering analyses of the material in the C-complex mentioned above.

MATCHING ASTEROIDS AND METEORITES

Both latent space analyses of CV/C0 versus K/L and EOC/UOC versus S presented in Chapter 14 are proof-of-concepts, highlighting that grain size and space weathering have large effects on the matching and need to be accounted for in a variable manner rather than in absolute values. The clear downside of the approach for now is that I do not quantify the degrees of space weathering or the grain sizes on asteroid surfaces via the latent space analyses, which is admittedly unsatisfactory.

One remedy I propose is to assume that the probability distribution of meteorites in the latent space is the ground truth that we aim to model using the probability distribution of the asteroids. To match the distributions, we add parameters of grain size and space weathering which displace and transform the asteroid distribution to best match the meteorite one. This could be an objective parametrisation of both grain size and space weathering influences. More detailed analyses may then make the grain size factors a function of asteroid diameter or thermal inertia, and the space weathering a function of families membership.

EXTENSION OF THE TAXONOMY

If we extrapolate past evolutions into the future, in about 10-15 years, a PhD student will be tasked to revise the asteroid taxonomy. This is necessary as taxonomy is a language which has to evolve with the new findings and conclusions in the field.

The MCFA method presented in [Chapter 5](#) and applied here has set up of the field of taxonomy for the next, hopefully data-intensive decades. Indeed, the machine learning approach thrives with increasing data, and more observations to define the projection and clusters allow to resolve the latent space in finer detail, revealing more compositional trends.

The way to improve the taxonomy, as I have mentioned throughout this work, lies in the expansion of the observables. Each add observable adds compositional resolution. The UV information provided by *Gaia* (Galluccio et al. 2022) and in the upcoming *Javalambre-Physics of the Accelerated Universe Astrophysical Survey* at the Observatorio Astrofísico de Javalambre resolves degeneracies in the C-complex (Tholen 1984; Tatsumi et al. 2022). Observations in the 3 μm region (Usui et al. 2011; Takir and Emery 2012) reveal hydration from the C-complex to the M-types (Jones et al. 1990; Rivkin 2012; Landsman et al. 2015; Shepard et al. 2015). Polarimetric parameters are the only apparent way to reliably identify *Barbarians* (Devogèle et al. 2018), and in combination with thermal inertia measurements (Delbó and Tanga 2009), they provide important information on the regolith properties (Delbó and Harris 2002; Shkuratov et al. 2007). The radar albedo quantifies the abundance of metal in the surface layer (Shepard et al. 2008c; Ockert-Bell et al. 2010). In addition, the on-going MITHNEOS (Binzel et al. 2019; Marsset et al. 2022) and the upcoming SPHEREx (Ivezić et al. 2022) surveys will continue to expand the number of reflectance spectra in the VisNIR region.

These observables can provide significant additional compositional resolution to the taxonomy. Deriving a method to then apply this classification on the basis of observations in the lower resolution observables such as photometry (Popescu et al. 2016; Sergeyev and Carry 2021) and phase curve coefficients (Oszkiewicz et al. 2011; Mahlke et al. 2021, and [Chapter 7](#)) to hundreds of thousands of asteroids supported by observations of LSST would put us within reach of maybe the ultimate goal of taxonomy: a consistent classification of asteroids based on any compositionally-informative observation of asteroids. From my own experience, it is not an easy feat, but if taxonomy is a language, then just imagine the conversation we could have.

BIBLIOGRAPHY

- Abadi, M. et al. (2015). TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from [tensorflow.org](https://www.tensorflow.org).
- Abell, P. A., Barbee, B. W., Chodas, P. W., Kawaguchi, J., Landis, R. R., Mazanek, D. D., and Michel, P. (2015). Human Exploration of Near-Earth Asteroids. In: *Asteroids IV*, 855–880.
- Adams, J. B. (1975). Interpretation of visible and near-infrared diffuse reflectance spectra of pyroxenes and other rock-forming minerals. In: *Infrared and Raman Spectroscopy of Lunar and Terrestrial Minerals*, 91–116.
- Adams, J. B. and McCord, T. B. (1971). *Science* 171 (3971), 567–571.
- Alí-Lagoa, V., Licandro, J., Gil-Hutton, R., Cañada-Assandri, M., Delbó, M., León, J. de, Campins, H., Pinilla-Alonso, N., Kelley, M. S. P., and Hanuš, J. (2016). *Astronomy & Astrophysics* 591, A14.
- Alí-Lagoa, V. and Delbó, M. (2017). *Astronomy & Astrophysics* 603, A55.
- Alí-Lagoa, V., León, J. de, Licandro, J., Delbó, M., Campins, H., Pinilla-Alonso, N., and Kelley, M. S. (2013). *Astronomy & Astrophysics* 554, A71.
- Alí-Lagoa, V., Müller, T. G., Usui, F., and Hasegawa, S. (2018). *Astronomy & Astrophysics* 612, A85.
- Alvarez-Candal, A., Benavidez, P. G., Campo Bagatin, A., and Santana-Ros, T. (2022). *Astronomy & Astrophysics* 657, A80.
- Alvarez-Candal, A., Duffard, R., Lazzaro, D., and Michtchenko, T. (2006). *Astronomy & Astrophysics* 459 (3), 969–976.
- Amelin, Y., Kaltenbach, A., Iizuka, T., Stirling, C. H., Ireland, T. R., Petaev, M., and Jacobsen, S. B. (2010). *Earth and Planetary Science Letters* 300 (3-4), 343–350.
- Anders, E. and Grevesse, N. (1989). *Geochimica et Cosmochimica Acta* 53 (1), 197–214.
- Arredondo, A., Campins, H., Pinilla-Alonso, N., León, J. de, Lorenzi, V., and Morate, D. (2021). *Icarus* 358, 114210.
- Baek, J., McLachlan, G. J., and Flack, L. K. (2010). *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (7), 1298–1309.
- Bartholomew, D. J. (1984). *Biometrika* 71.2, 221–232.
- Barucci, M. A., Capria, M. T., Coradini, A., and Fulchignoni, M. (1987). *Icarus* 72 (2), 304–324.
- Barucci, M. A., Cruikshank, D. P., Mottola, S., and Lazzarin, M. (2002). Physical Properties of Trojan and Centaur Asteroids. In: *Asteroids III*, 273–287.
- Barucci, M. A., Perna, D., Popescu, M., Fornasier, S., Doressoundiram, A., Lantz, C., Merlin, F., Fulchignoni, M., Dotto, E., and Kanuchova, S. (2018). *Monthly Notices of the Royal Astronomical Society* 476 (4), 4481–4487.
- Batygin, K. and Brown, M. E. (2016). *The Astronomical Journal* 151 (2), 22.
- Becker, T. M. et al. (2015). *Icarus* 248, 499–515.
- Beckett, J. R., Connolly, H. C., and Ebel, D. S. (2006). Chemical Processes in Igneous Calcium-Aluminum-rich Inclusions: A Mostly CMAS View of Melting and Crystallization. In: *Meteorites and the Early Solar System II*. Ed. by D. S. Lauretta and H. Y. McSween, 399.

- Bell, J. F. (1988). *Meteoritics* 23, 256–257.
- Bell, J. F., Owensby, P. D., Hawke, B. R., and Gaffey, M. J. (1988). The 52-Color Asteroid Survey: Final Results and Interpretation. In: *Lunar and Planetary Science Conference*. 19. Lunar and Planetary Science Conference, 57.
- Bellman, R. (1966). *Science* 153.3731, 34–37.
- Belskaya, I. and Shevchenko, V. (2000). *Icarus* 147 (1), 94–105.
- Belskaya, I., Cellino, A., Gil-Hutton, R., Muinonen, K., and Shkuratov, Y. (2015). Asteroid Polarimetry. In: *Asteroids IV*.
- Bendjoya, P., Cellino, A., Di Martino, M., and Saba, L. (2004). *Icarus* 168 (2), 374–384.
- Benitez, N. et al. (2014). eprint: 1403.5237v1.
- Benner, L. A. M., Busch, M. W., Giorgini, J. D., Taylor, P. A., and Margot, J.-L. (2015). Radar Observations of Near-Earth and Main-Belt Asteroids. In: *Asteroids IV*.
- Benner, L. A. M. (2002). *Icarus* 158 (2), 379–388.
- Benner, L. A. M., Ostro, S. J., Magri, C., Nolan, M. C., Howell, E. S., Giorgini, J. D., Jurgens, R. F., Margot, J.-L., Taylor, P. A., Busch, M. W., and Shepard, M. K. (2008). *Icarus* 198 (2), 294–304.
- Berthier, J., Hestroffer, D., Carry, B., Durech, J., Tanga, P., Delbo, M., and Vachier, F. (2008). A Service of Position and Physical Ephemerides Computation Dedicated to the Small Bodies of the Solar System. In: *Asteroids, Comets, Meteors 2008*. 1405, 8374, 8374.
- Berthier, J., Vachier, F., Marchis, F., Durech, J., and Carry, B. (2014). *Icarus* 239, 118–130.
- Berthier, J., Vachier, F., Thuillot, W., Fernique, P., Ochsenbein, F., Genova, F., Lainey, V., and Arlot, J. E. (2006). SkyBoT, a new VO service to identify Solar System objects. In: *Astronomical Data Analysis Software and Systems XV*. 351. Astronomical Society of the Pacific Conference Series, 367.
- Binzel, R. P. (1995). *Meteoritics* 30.5, 486.
- Binzel, R. P. (2001). *Icarus* 151 (2), 139–149.
- Binzel, R. P., Reddy, V., and Dunn, T. L. (2015). The Near-Earth Object Population: Connections to Comets, Main-Belt Asteroids, and Meteorites. In: *Asteroids IV*, 243–256.
- Binzel, R. P., Rivkin, A. S., Bus, S. J., Sunshine, J. M., and Burbine, T. H. (2001). *Meteoritics and Planetary Science Supplement* 36, A20.
- Binzel, R. P. et al. (2019). *Icarus* 324, 41–76.
- Binzel, R. P., Birlan, M., Bus, S. J., Harris, A. W., Rivkin, A. S., and Fornasier, S. (2004a). *Planetary and Space Science* 52 (4), 291–296.
- Binzel, R. P., Perozzi, E., Rivkin, A. S., Rossi, A., Harris, A. W., Bus, S. J., Valsecchi, G. B., and Slivan, S. M. (2004b). *Meteoritics & Planetary Science* 39 (3), 351–366.
- Binzel, R. P., Rivkin, A. S., Stuart, J., Harris, A. W., Bus, S. J., and Burbine, T. H. (2004c). *Icarus* 170 (2), 259–294.
- Binzel, R. P., Rivkin, A. S., Thomas, C. A., Vernazza, P., Burbine, T. H., DeMeo, F. E., Bus, S. J., Tokunaga, A. T., and Birlan, M. (2009). *Icarus* 200 (2), 480–485.
- Binzel, R. P. and Xu, S. (1993). *Science* 260 (5105), 186–191.

- Binzel, R. P., Xu, S., Bus, S. J., Skrutskie, M. F., Meyer, M. R., Knezek, P., and Barker, E. S. (1993). *Science* 262 (5139), 1541–1543.
- Birlan, M., Nedelcu, D. A., Descamps, P., Berthier, J., Marchis, F., Merouane, S., and Popescu, M. (2011). *Monthly Notices of the Royal Astronomical Society* 415 (1), 587–595.
- Birlan, M., Nedelcu, D. A., Popescu, M., Vernazza, P., Colas, F., and Kryszczyńska, A. (2014). *Monthly Notices of the Royal Astronomical Society* 437 (1), 176–184.
- Birlan, M., Vernazza, P., Fulchignoni, M., Barucci, M. A., Descamps, P., Binzel, R. P., and Bus, S. J. (2006). *Astronomy & Astrophysics* 454 (2), 677–681.
- Birlan, M., Vernazza, P., and Nedelcu, D. A. (2007). *Astronomy & Astrophysics* 475 (2), 747–754.
- Birlan, M., Barucci, M. A., Vernazza, P., Fulchignoni, M., Binzel, R. P., Bus, S. J., Belskaya, I., and Fornasier, S. (2004). *New Astronomy* 9 (5), 343–351.
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning.
- Bland, P. A. and Travis, B. J. (2017). *Science Advances* 3 (7).
- Bobrovnikoff, N., McCord, T., Chapman, C., Tedesco, E., Tholen, D., Zellner, B., Veeder, G., Matson, D., Kowal, C., Gradie, J., and Gaffey, M. (1983). *Sky and Telescope* 65, 326.
- Bolin, B., Masci, F. J., and Ye, Q.-Z. (2020). MPEC 2020-A99: 2020 AV2. Minor Planet Electronic Circulars. M.P.E.C. 2020-A99.
- Bonal, L., Gattacceca, J., Garenne, A., Eschrig, J., Rochette, P., and Krämer Ruggiu, L. (2020). *Geochimica et Cosmochimica Acta* 276, 363–383.
- Bonal, L., Quirico, E., Flandinet, L., and Montagnac, G. (2016). *Geochimica et Cosmochimica Acta* 189, 312–337.
- Borisov, G., Christou, A., Bagnulo, S., Cellino, A., Kwiatkowski, T., and Dell’Oro, A. (2017). *Monthly Notices of the Royal Astronomical Society* 466 (1), 489–495.
- Borisov, G., Christou, A. A., Colas, F., Bagnulo, S., Cellino, A., and Dell’Oro, A. (2018). *Astronomy & Astrophysics* 618, A178.
- Bottke, W. F., Vokrouhlický, D., Ballouz, R.-L., Barnouin, O. S., Connolly, H. C., Elder, C., Marchi, S., McCoy, T. J., Michel, P., Nolan, M. C., Rizk, B., Scheeres, D. J., Schwartz, S. R., Walsh, K. J., and Laretta, D. S. (2020). *The Astronomical Journal* 160 (1), 14.
- Bottke, W. F., Vokrouhlický, D., Minton, D., Nesvorný, D., Morbidelli, A., Brasser, R., Simonson, B., and Levison, H. F. (2012). *Nature* 485 (7396), 78–81.
- Bottou, L., Curtis, F. E., and Nocedal, J. (2018). *SIAM Review* 60.2, 223–311.
- Bouveyron, C., Celeux, G., Murphy, T. B., and Raftery, A. E. (2019). Model-Based Clustering and Classification for Data Science: With Applications in R. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge.
- Bowell, E. (1996). VizieR Online Data Catalog.
- Bowell, E., Chapman, C. R., Gradie, J. C., Morrison, D., and Zellner, B. (1978). *Icarus* 35.3, 313–335.
- Bowell, E., Hapke, B., Domingue, D., Lumme, K., Peltoniemi, J., and Harris, A. W. (1989). Application of photometric models to asteroids. In: *Asteroids II*, 524–556.

- Bowell, E., Muinonen, K., and Wasserman, L. H. (1994). A Public-Domain Asteroid Orbit Data Base. In: *Asteroids, Comets, Meteors 1993*. 160, 477–481.
- Bradley, J. P. (2014). Early Solar Nebula Grains - Interplanetary Dust Particles. In: *Meteorites and Cosmochemical Processes*. 1, 287–308.
- Brearley, A. J., Jones, R. H., and Papike, J. J. (1998). Chondritic meteorites. In: *Planetary Materials*. 36. Reviews in Mineralogy, C1.
- Brearley, A. J. (2006). The Action of Water. In: *Meteorites and the Early Solar System II*. Ed. by D. S. Lauretta and H. Y. McSween, 584.
- Brož, M., Ferrais, M., Vernazza, P., Ševeček, P., and Jutzi, M. (2022). eprint: [2205.15736v1](https://arxiv.org/abs/2205.15736v1).
- Brož, M., Morbidelli, A., Bottke, W. F., Rozehnal, J., Vokrouhlický, D., and Nesvorný, D. (2013). *Astronomy & Astrophysics* 551, A117.
- Brunetto, R., Vernazza, P., Marchi, S., Birlan, M., Fulchignoni, M., Orofino, V., and Strazzulla, G. (2006). *Icarus* 184 (2), 327–337.
- Burbine, T. H., Binzel, R. P., Bus, S. J., and Clark, B. E. (2001). *M&PS* 36.2, 245–253.
- Burbine, T. H. (1998). *Meteoritics & Planetary Science* 33 (2), 253–258.
- Burbine, T. H. (2000). Forging asteroid-meteorite relationships through reflectance spectroscopy. PhD Thesis. Massachusetts Institute of Technology.
- Burbine, T. H. (2014). Asteroids. In: *Planets, Asteroids, Comets and The Solar System*. Ed. by A. M. Davis. 2, 365–415.
- Burbine, T. H. (2016). Asteroids.
- Burbine, T. H., Buchanan, P. C., Dolkar, T., and Binzel, R. P. (2009). *Meteoritics & Planetary Science* 44 (9), 1331–1341.
- Burbine, T. H., Gaffey, M. J., and Bell, J. F. (1992). *Meteoritics* 27 (4), 424–434.
- Burbine, T. H., Meibom, A., and Binzel, R. P. (1996). *Meteoritics & Planetary Science* 31 (5), 607–620.
- Bus, S. J., Vilas, F., and Barucci, M. A. (2002). Visible-Wavelength Spectroscopy of Asteroids. In: *Asteroids III*, 169–182.
- Bus, S. J. and Binzel, R. P. (2002a). *Icarus* 158 (1), 146–177.
- Bus, S. J. and Binzel, R. P. (2002b). *Icarus* 158 (1), 106–145.
- Bus, S. J. (1999). Compositional structure in the asteroid belt: Results of a spectroscopic survey. PhD Thesis. Massachusetts Institute of Technology.
- Butterworth, P. and Meadows, A. (1985). *Icarus* 62 (2), 305–318.
- Buuren, S. v. and Groothuis-Oudshoorn, K. (2011). *Journal of Statistical Software* 45 (3).
- Candolle, A. P. d. (1813). Exposition des principes de la classification naturelle et de l'art de décrire et d'étudier les végétaux. In: *Théorie élémentaire de la botanique*.
- Carruba, V. (2013). *Monthly Notices of the Royal Astronomical Society* 431 (4), 3557–3569.
- Carruba, V., Huaman, M. E., Domingos, R. C., Santos, C. R. D., and Souami, D. (2014). *Monthly Notices of the Royal Astronomical Society* 439 (3), 3168–3179.
- Carry, B. (2012). *Planetary and Space Science* 73 (1), 98–118.
- Carry, B. (2018). *Astronomy & Astrophysics* 609, A113.
- Carry, B., Vernazza, P., Vachier, F., Neveu, M., Berthier, J., Hanus, J., Ferrais, M., Jorda, L., and al, N. (2021). *Astronomy & Astrophysics*.

- Carry, B. et al. (2019). *Astronomy & Astrophysics* 623, A132.
- Carvano, J. M. and Davalos, J. A. G. (2015). *A&A* 580, A98.
- Carvano, J. M., Hasselmann, P. H., Lazzaro, D., and Mothé-Diniz, T. (2010). *Astronomy and Astrophysics* 510, A43.
- Casey, A. R., Lattanzio, J. C., Aletti, A., Dowe, D. L., Bland-Hawthorn, J., Buder, S., Lewis, G. F., Martell, S. L., Nordlander, T., Simpson, J. D., Sharma, S., and Zucker, D. B. (2019). *The Astrophysical Journal* 887 (1), 73.
- Cellino, A., Bagnulo, S., Tanga, P., Novaković, B., and Delbó, M. (2014). *Monthly Notices of the Royal Astronomical Society: Letters* 439 (1), L75–L79.
- Cellino, A., Belskaya, I., Bendjoya, P., Di Martino, M., Gil-Hutton, R., Muinonen, K., and Tedesco, E. (2006). *Icarus* 180 (2), 565–567.
- Cellino, A., Bendjoya, P., Delbó, M., Galluccio, L., Gayon-Markt, J., Tanga, P., and Tedesco, E. F. (2020). *Astronomy & Astrophysics* 642, A80.
- Chapman, C. R., Enke, B., Merline, W., Tamblyn, P., Nesvorný, D., Young, E., and Olkin, C. (2007). *Icarus* 191 (1), 323–329.
- Chapman, C. R. and Gaffey, M. J. (1979). Reflectance spectra for 277 asteroids. In: *Asteroids*, 655–687.
- Chapman, C. R. (1972). Surface properties of asteroids. PhD Thesis. Massachusetts Institute of Technology.
- Chapman, C. R. (1976). *Geochimica et Cosmochimica Acta* 40 (7), 701–719.
- Chapman, C. R. (1996). *Meteoritics & Planetary Science* 31 (6), 699–725.
- Chapman, C. R., Morrison, D., and Zellner, B. (1975). *Icarus* 25 (1), 104–130.
- Chapman, C. R. and Salisbury, J. W. (1973). *Icarus* 19 (4), 507–522.
- Chavez, C. F., Müller, T. G., Marshall, J. P., Horner, J., Drass, H., and Carter, B. (2021). *Monthly Notices of the Royal Astronomical Society* 502 (4), 4981–4992.
- Chesley, S. R., Hockney, G. M., and Holman, M. J. (2017). Introducing ADES: A New IAU Astrometry Data Exchange Standard. In: *AAS/Division for Planetary Sciences Meeting Abstracts*. 49. AAS/Division for Planetary Sciences Meeting Abstracts, 112.14.
- Chladni, E. (1794). Ueber den Ursprung der von Pallas gefundenen und anderer ihr ähnlicher Eisenmassen.
- Chrbolková, K., Brunetto, R., Ďurech, J., Kohout, T., Mizohata, K., Malý, P., Dědič, V., Lantz, C., Penttilä, A., Trojánek, F., and Maturilli, A. (2021). eprint: 2108.00870v1.
- Christou, A. A., Borisov, G., Dell’Oro, A., Cellino, A., and Bagnulo, S. (2017). *Icarus* 293, 243–258.
- Ciesla, F. J. (2005). Chondrule-forming Processes - An Overview. In: *Chondrites and the Protoplanetary Disk*. 341. Astronomical Society of the Pacific Conference Series, 811.
- Clark, B. E., Bus, S. J., Rivkin, A. S., Shepard, M. K., and Shah, S. (2004). *The Astronomical Journal* 128 (6), 3070–3081.
- Clark, B. E., Veverka, J., Helfenstein, P., Thomas, P., Bell, J., Harch, A., Robinson, M., Murchie, S., McFadden, L., and Chapman, C. (1999). *Icarus* 140 (1), 53–65.
- Clark, B. E., Binzel, R. P., Howell, E. S., Cloutis, E. A., Ockert-Bell, M., Christensen, P., Barucci, M. A., DeMeo, F., Lauretta, D. S., Connolly, H., Soderberg,

- A., Hergenrother, C., Lim, L., Emery, J., and Mueller, M. (2011). *Icarus* 216 (2), 462–475.
- Clark, B. E., Ockert-Bell, M. E., Cloutis, E. A., Nesvorný, D., Mothé-Diniz, T., and Bus, S. J. (2009). *Icarus* 202 (1), 119–133.
- Clark, B. E., Ziffer, J., Nesvorný, D., Campins, H., Rivkin, A. S., Hiroi, T., Barucci, M. A., Fulchignoni, M., Binzel, R. P., Fornasier, S., DeMeo, F., Ockert-Bell, M. E., Licandro, J., and Mothé-Diniz, T. (2010). *Journal of Geophysical Research (Planets)* 115.E6, E06005.
- Cloutis, E. A., Hudon, P., Hiroi, T., and Gaffey, M. J. (2012a). *Icarus* 221 (2), 911–924.
- Cloutis, E. A., Hudon, P., Hiroi, T., Gaffey, M. J., and Mann, P. (2012b). *Icarus* 220 (2), 466–486.
- Cloutis, E., Hiroi, T., Gaffey, M., Alexander, C., and Mann, P. (2011). *Icarus* 212 (1), 180–209.
- Cloutis, E. A., Gaffey, M. J., Jackowski, T. L., and Reed, K. L. (1986). *Journal of Geophysical Research* 91 (B11), 11641.
- Cloutis, E. A., Gaffey, M. J., Smith, D. G. W., and Lambert, R. S. J. (1990). *Journal of Geophysical Research* 95 (B6), 8323.
- Cloutis, E. A., Izawa, M. R., and Beck, P. (2018). Reflectance Spectroscopy of Chondrites. In: *Primitive Meteorites and Asteroids*, 273–343.
- Cortés-Contreras, M., Bouy, H., Solano, E., Mahlke, M., Jiménez-Esteban, F., Alacid, J. M., and Rodrigo, C. (2020). *Monthly Notices of the Royal Astronomical Society* 491 (1), 129–152.
- Cortés-Contreras, M., Jiménez-Esteban, F. M., Mahlke, M., Solano, E., Ďurech, J., Barceló Forteza, S., Rodrigo, C., Velasco, A., and Carry, B. (2019). *Monthly Notices of the Royal Astronomical Society* 490 (3), 3046–3060.
- Cruikshank, D. P. and Hartmann, W. K. (1984). *Science* 223 (4633), 281–283.
- Davis, D. R., Farinella, P., and Marzari, F. (1999). *Icarus* 137 (1), 140–151.
- De Prá, M., Pinilla-Alonso, N., Carvano, J., Licandro, J., Campins, H., Mothé-Diniz, T., De León, J., and Alí-Lagoa, V. (2018). *Icarus* 311, 35–51.
- De Sanctis, M. C., Migliorini, A., Luzia Jasmin, F., Lazzaro, D., Filacchione, G., Marchi, S., Ammannito, E., and Capria, M. T. (2011a). *Astronomy & Astrophysics* 533, A77.
- De Sanctis, M. C., Ammannito, E., Migliorini, A., Lazzaro, D., Capria, M. T., and McFadden, L. (2011b). *Monthly Notices of the Royal Astronomical Society* 412 (4), 2318–2332.
- Degewij, J. and Houten, C. J. van (1979). Distant asteroids and outer Jovian satellites. In: *Asteroids*. Ed. by T. Gehrels and M. S. Matthews, 417–435.
- Delbó, M., dell’Oro, A., Harris, A. W., Mottola, S., and Mueller, M. (2007). *Icarus* 190 (1), 236–249.
- Delbó, M. and Harris, A. W. (2002). *Meteoritics & Planetary Science* 37 (12), 1929–1936.
- Delbó, M., Harris, A. W., Binzel, R. P., Pravec, P., and Davies, J. K. (2003). *Icarus* 166 (1), 116–130.
- Delbó, M., Libourel, G., Wilkerson, J., Murdoch, N., Michel, P., Ramesh, K. T., Ganino, C., Verati, C., and Marchi, S. (2014). *Nature* 508 (7495), 233–236.
- Delbó, M. and Tanga, P. (2009). *Planetary and Space Science* 57 (2), 259–265.
- DeMeo, F. E. (2010).

- DeMeo, F. E. and Carry, B. (2013). *Icarus* 226 (1), 723–741.
- DeMeo, F. E. and Carry, B. (2014). *Nature* 505.7485, 629–634.
- DeMeo, F. E., Alexander, C., Walsh, K., Chapman, C., and Binzel, R. (2015). *Asteroids IV*.
- DeMeo, F. E., Binzel, R. P., Carry, B., Polishook, D., and Moskovitz, N. A. (2014a). *Icarus* 229, 392–399.
- DeMeo, F. E., Binzel, R. P., and Lockhart, M. (2014b). *Icarus* 227, 112–122.
- DeMeo, F. E., Binzel, R. P., Slivan, S. M., and Bus, S. J. (2009). *Icarus* 202 (1), 160–180.
- DeMeo, F. E., Burt, B. J., Marsset, M., Polishook, D., Burbine, T. H., Carry, B., Binzel, R. P., Vernazza, P., Reddy, V., Tang, M., Thomas, C. A., Rivkin, A. S., Moskovitz, N. A., Slivan, S. M., and Bus, S. J. (2022). *Icarus* 380, 114971.
- DeMeo, F. E., Polishook, D., Carry, B., Burt, B. J., Hsieh, H. H., Binzel, R. P., Moskovitz, N. A., and Burbine, T. H. (2019). *Icarus* 322, 13–30.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). *Journal of the Royal Statistical Society. Series B (Methodological)* 39.1, 1–38.
- Devogèle, M. et al. (2017). *Astronomy & Astrophysics* 607, A119.
- Devogèle, M. et al. (2018). *Icarus* 304, 31–57.
- Devogèle, M. et al. (2019). *The Astronomical Journal* 158 (5), 196.
- Domingos, P. (2012). *Communications of the ACM* 55 (10), 78–87.
- Dong-fang, Z., Peng, L., Wei, Z., Chang-ning, H., Hong-wei, Z., and Xiang-long, T. (2016). *Chinese Astronomy and Astrophysics* 40 (4), 555–568.
- Dotto, E., Emery, J. P., Barucci, M. A., Morbidelli, A., and Cruikshank, D. P. (2008). De Troianis: The Trojans in the Planetary System. In: *The Solar System Beyond Neptune*, 383.
- Drake, A. J., Djorgovski, S. G., Mahabal, A., Beshore, E., Larson, S., Graham, M. J., Williams, R., Christensen, E., Catelan, M., Boattini, A., Gibbs, A., Hill, R., and Kowalski, R. (2009). *The Astrophysical Journal* 696 (1), 870–884.
- Drummond, J. D. and Christou, J. (2008). *Icarus* 197 (2), 480–496.
- Drummond, J. D., Merline, W. J., Carry, B., Conrad, A., Reddy, V., Tamblyn, P., Chapman, C. R., Enke, B. L., Pater, I. d., Kleer, K. d., Christou, J., and Dumas, C. (2018). *Icarus* 305, 174–185.
- Duffard, R. and Roig, F. (2009). *Planetary and Space Science* 57 (2), 229–234.
- Duffard, R., Lazzaro, D., Licandro, J., De Sanctis, M. C., Capria, M. T., and Carvano, J. M. (2004). *Icarus* 171 (1), 120–132.
- Duncan, M. J. and Levison, H. F. (1997). *Science* 276 (5319), 1670–1672.
- Dunn, T. L., Burbine, T. H., Bottke, W. F., and Clark, J. P. (2013). *Icarus* 222 (1), 273–282.
- Durech, J., Sidorin, V., and Kaasalainen, M. (2010). *A&A* 513, A46.
- Dykhuis, M. J. and Greenberg, R. (2015). *Icarus* 252, 199–211.
- Ebel, D. S. (2006). Condensation of Rocky Material in Astrophysical Environments. In: *Meteorites and the Early Solar System II*, 253.
- Edgeworth, K. E. (1943). *Journal of the British Astronomical Association* 53, 181–188.
- Elowitz, R. M., Stokes, G. H., Bezpalko, M., Blythe, M. S., Evans, J. B., Pearce, E. C., Sayer, R. W., Shelly, F. C., and Viggh, H. E. M. (1999). A Progress Report on the Lincoln Near Earth Asteroid Research Project. In: *Ameri-*

- can Astronomical Society Meeting Abstracts*. 195. American Astronomical Society Meeting Abstracts, 108.01.
- Emery, J. P. and Brown, R. H. (2003). *Icarus* 164 (1), 104–121.
- Emery, J. P., Burr, D. M., and Cruikshank, D. P. (2011). *The Astronomical Journal* 141 (1), 25.
- Emery, J. P., Marzari, F., Morbidelli, A., French, L. M., and Grav, T. (2015). The Complex History of Trojan Asteroids. In: *Asteroids IV*.
- Erasmus, N., Navarro-Meza, S., McNeill, A., Trilling, D. E., Sickafoose, A. A., Denneau, L., Flewelling, H., Heinze, A., and Tonry, J. L. (2020). *The Astrophysical Journal Supplement Series* 247.1, 13.
- Eschrig, J., Bonal, L., Beck, P., and Prestgard, T. (2021). *Icarus* 354, 114034.
- Eschrig, J., Bonal, L., Mahlke, M., Carry, B., Beck, P., and Gattacceca, J. (2022). *Icarus* 381, 115012.
- Farinella, P., Froeschlé, C., Froeschlé, C., Gonczi, R., Hahn, G., Morbidelli, A., and Valsecchi, G. B. (1994). *Nature* 371 (6495), 314–317.
- Fedkin, A. V. and Grossman, L. (2006). The Fayalite Content of Chondritic Olivine: Obstacle to Understanding the Condensation of Rocky Material. In: *Meteorites and the Early Solar System II*. Ed. by D. S. Lauretta and H. Y. McSween, 279.
- Feierberg, M. A., Larson, H. P., and Chapman, C. R. (1982). *The Astrophysical Journal* 257, 361.
- Ferguson, C. A. (1959). *Word* 15 (2), 325–340.
- Fernández, J. and Ip, W.-H. (1984). *Icarus* 58 (1), 109–120.
- Fieber-Beyer, S. K. (2010). Mineralogical characterizations of asteroids in/near the 3:1 Kirkwood gap. PhD Thesis. University of North Dakota.
- Fieber-Beyer, S. K. and Gaffey, M. J. (2011). *Icarus* 214 (2), 645–651.
- Fieber-Beyer, S. K. and Gaffey, M. J. (2014). *Icarus* 229, 99–108.
- Fieber-Beyer, S. K. and Gaffey, M. J. (2015). *Icarus* 257, 113–125.
- Fieber-Beyer, S. K., Gaffey, M. J., Hardersen, P. S., and Reddy, V. (2012). *Icarus* 221 (2), 593–602.
- Fieber-Beyer, S. K., Gaffey, M. J., Kelley, M. S., Reddy, V., Reynolds, C. M., and Hicks, T. (2011). *Icarus* 213 (2), 524–537.
- Foote, A. E. (1891). *American Journal of Science* s3-42 (251), 413–417.
- Fornasier, S., Dotto, E., Hainaut, O., Marzari, F., Boehnhardt, H., Deluise, F., and Barucci, M. (2007). *Icarus* 190 (2), 622–642.
- Fornasier, S., Clark, B. E., Migliorini, A., and Ockert-Bell, M. (2011). NASA Planetary Data System, EAR-A-I1092-2-MSPECTRA-V1.0.
- Fornasier, S., Lantz, C., Barucci, M., and Lazzarin, M. (2014). *Icarus* 233, 163–178.
- Fornasier, S., Lantz, C., Perna, D., Campins, H., Barucci, M., and Nesvorný, D. (2016). *Icarus* 269, 1–14.
- Fornasier, S., Lazzarin, M., Barbieri, C., and Barucci, M. A. (1999). *Astronomy and Astrophysics Supplement Series* 135 (1), 65–73.
- Fraser, W. C. et al. (2017). *Nature Astronomy* 1 (4).
- Frattin, E., Muñoz, O., Moreno, F., Nava, J., Escobar-Cerezo, J., Gomez Martin, J. C., Guirado, D., Cellino, A., Coll, P., Raulin, F., Bertini, I., Cremonese, G., Lazzarin, M., Naletto, G., and La Forgia, F. (2019). *Monthly Notices of the Royal Astronomical Society* 484 (2), 2198–2211.

- Fujiwara, A. et al. (2006). *Science* 312 (5778), 1330–1334.
- Gaffey, M. J. and Kelley, M. S. (2004). Mineralogical Variations Among High Albedo E-Type Asteroids: Implications for Asteroid Igneous Processes. In: *Lunar and Planetary Science Conference*. Lunar and Planetary Science Conference, 1812.
- Gaffey, M. J. and McCord, T. B. (1979). Mineralogical and petrological characterizations of asteroid surface materials. In: *Asteroids*, 688–723.
- Gaffey, M. J. (1993). *Science* 260 (5105), 167–168.
- Gaffey, M. J., Bell, J. F., Brown, R., Burbine, T. H., Piatek, J. L., Reed, K. L., and Chaky, D. A. (1993a). *Icarus* 106 (2), 573–602.
- Gaffey, M. J., Bell, J. F., and Cruikshank, D. P. (1989). Reflectance spectroscopy and asteroid surface mineralogy. In: *Asteroids II*, 98–127.
- Gaffey, M. J., Burbine, T. H., and Binzel, R. P. (1993b). *Meteoritics* 28 (2), 161–187.
- Gaffey, M. J. and Gilbert, S. L. (1998). *Meteoritics & Planetary Science* 33 (6), 1281–1295.
- Gaffey, M. J., Reed, K. L., and Kelley, M. S. (1992). *Icarus* 100 (1), 95–109.
- Gaffey, M. J. (1974). A systematic study of the spectral reflectivity characteristics of the meteorite classes with applications to the interpretation of asteroid spectra for mineralogical and petrological information. PhD Thesis. Massachusetts Institute of Technology.
- Galluccio, L., Delbó, M., De Angeli, F., Pauwels, T., Tanga, P., Mignard, F., Cellino, A., Brown, A., Muinonen, K., Penttilä, A., and Jordan, S. (2022). *Astronomy & Astrophysics*.
- Gartelle, G. M., Hardersen, P. S., Izawa, M. R. M., and Nowinski, M. C. (2021). *NASA Planetary Data System*, 6.
- Gehrels, T. (1979). The asteroids: history, survey, techniques, and future work. In: *Asteroids*. Ed. by T. Gehrels and M. S. Matthews, 3–24.
- Gehrels, T. (1956). *The Astrophysical Journal* 123.
- Gietzen, K. M., Lacy, C. H. S., Ostrowski, D. R., and Sears, D. W. G. (2012). *Meteoritics & Planetary Science* 47 (11), 1789–1808.
- Gil-Hutton, R., Mesa, V., Cellino, A., Bendjoya, P., Peñaloza, L., and Lovos, F. (2008). *Astronomy & Astrophysics* 482 (1), 309–314.
- Gladman, B., Lawler, S. M., Petit, J.-M., Kavelaars, J., Jones, R. L., Parker, J. W., Van Laerhoven, C., Nicholson, P., Rousselot, P., Bieryla, A., and Ashby, M. L. N. (2012). *The Astronomical Journal* 144 (1), 23.
- Gladman, B., Marsden, B. G., and Vanlaerhoven, C. (2008). Nomenclature in the Outer Solar System. In: *The Solar System Beyond Neptune*, 43.
- Gladman, B. J., Migliorini, F., Morbidelli, A., Zappalà, V., Michel, P., Cellino, A., Froeschlé, C., Levison, H. F., Bailey, M., and Duncan, M. (1997). *Science* 277 (5323), 197–201.
- Goldschmidt, V. M. (1937). *J. Chem. Soc.* 0 (0), 655–673.
- Goldstein, J., Scott, E., and Chabot, N. (2009). *Geochemistry* 69 (4), 293–325.
- Gomes, R., Levison, H. F., Tsiganis, K., and Morbidelli, A. (2005). *Nature* 435 (7041), 466–469.
- Gooding, J. L. (1982). Lunar and Planetary Science Conference Proceedings 12, 1105–1122.
- Goody, R. M. and Yung, Y. L. (1989). Atmospheric Radiation.

- Gradie, J. and Tedesco, E. (1982). *Science* 216.4553, 1405–1407.
- Gradie, J. and Veverka, J. (1986). *Icarus* 66 (3), 455–467.
- Gradie, J., Veverka, J., and Buratti, B. (1980). Lunar and Planetary Science Conference Proceedings 1, 799–815.
- Grady, M. M. and Wright, I. (2006). Types of extraterrestrial material available for study. In: *Meteorites and the Early Solar System II*, 3.
- Granvik, M. and Brown, P. (2018). *Icarus* 311, 271–287.
- Granvik, M., Morbidelli, A., Jedicke, R., Bolin, B., Bottke, W. F., Beshore, E., Vokrouhlický, D., Delbò, M., and Michel, P. (2016). *Nature* 530 (7590), 303–306.
- Granvik, M., Morbidelli, A., Vokrouhlický, D., Bottke, W. F., Nesvorný, D., and Jedicke, R. (2017). *Astronomy & Astrophysics* 598, A52.
- Grav, T., Mainzer, A. K., Bauer, J. M., Masiero, J. R., and Nugent, C. R. (2012a). *The Astrophysical Journal* 759 (1), 49.
- Grav, T. et al. (2011). *The Astrophysical Journal* 742 (1), 40.
- Grav, T. et al. (2012b). *The Astrophysical Journal* 744 (2), 197.
- Greenberg, R. and Nolan, M. C. (1989). Delivery of asteroids and meteorites to the inner solar system. In: *Asteroids II*, 778–801.
- Greenstreet, S. (2020). *Monthly Notices of the Royal Astronomical Society: Letters* 493 (1), L129–L131.
- Greenstreet, S., Ngo, H., and Gladman, B. (2012). *Icarus* 217 (1), 355–366.
- Greenwood, R., Franchi, I., Kearsley, A., and Alard, O. (2010). *Geochimica et Cosmochimica Acta* 74 (5), 1684–1705.
- Greenwood, R. C., Burbine, T. H., and Franchi, I. A. (2020). *Geochimica et Cosmochimica Acta* 277, 377–406.
- Grossman, J. N., Alexander, C. M. O., Wang, J., and Brearley, A. J. (2000). *Meteoritics & Planetary Science* 35 (3), 467–486.
- Gulbis, A., Elliot, J., and Kane, J. (2006). *Icarus* 183 (1), 168–178.
- Gundlach, B. and Blum, J. (2013). *Icarus* 223 (1), 479–492.
- Hanuš, J., Delbo', M., Ďurech, J., and Alí-Lagoa, V. (2015). *Icarus* 256, 101–116.
- Hanuš, J., Viikinkoski, M., Marchis, F., Ďurech, J., Kaasalainen, M., Delbó, M., Herald, D., Frappa, E., Hayamizu, T., Kerr, S., Preston, S., Timerson, B., Dunham, D., and Talbot, J. (2017). *Astronomy & Astrophysics* 601, A114.
- Hanuš, J., Vokrouhlický, D., Delbo', M., Farnocchia, D., Polishook, D., Pravec, P., Hornoch, K., Kučáková, H., Kušnirák, P., Stephens, R., and Warner, B. (2018). *Astronomy & Astrophysics* 620, L8.
- Hanuš, J. et al. (2016). *Astronomy & Astrophysics* 592, A34.
- Hapke, B. (1971). Inferences from Optical Properties Concerning the Surface Texture and Composition of Asteroids. In: *NASA Special Publication*. 267, 67.
- Hapke, B. (1981). *Journal of Geophysical Research: Solid Earth* 86 (B4), 3039–3054.
- Hardersen, P., Gaffey, M., and Abell, P. (2005). *Icarus* 175 (1), 141–158.
- Hardersen, P. S., Cloutis, E. A., Reddy, V., Mothé-Diniz, T., and Emery, J. P. (2011). *Meteoritics & Planetary Science* 46 (12), 1910–1938.
- Hardersen, P. S., Reddy, V., Cloutis, E., Nowinski, M., Dievendorf, M., Genet, R. M., Becker, S., and Roberts, R. (2018). *The Astronomical Journal* 156 (1), 11.

- Hardersen, P. S., Reddy, V., and Roberts, R. (2015). *The Astrophysical Journal Supplement Series* 221 (1), 19.
- Hardersen, P. S., Reddy, V., Roberts, R., and Mainzer, A. (2014). *Icarus* 242, 269–282.
- Harris, A. (1999). *Icarus* 142 (2), 464–475.
- Harris, A. W., Barucci, M. A., Cano, J. L., Fitzsimmons, A., Fulchignoni, M., Green, S. F., Hestroffer, D., Lappas, V., Lork, W., Michel, P., Morrison, D., Payson, D., and Schäfer, F. (2013). *Acta Astronautica* 90 (1), 80–84.
- Harris, A. W. and Lagerros, J. S. V. (2002). Asteroids in the Thermal Infrared. In: *Asteroids III*, 205–218.
- Hasegawa, S., Kasuga, T., Usui, F., and Kuroda, D. (2021a). *Publications of the Astronomical Society of Japan* 73 (1), 240–255.
- Hasegawa, S., Marsset, M., DeMeo, F. E., Bus, S. J., Geem, J., Ishiguro, M., Im, M., Kuroda, D., and Vernazza, P. (2021b). *The Astrophysical Journal Letters* 916.1, L6.
- Hasegawa, S. et al. (2018). *Publications of the Astronomical Society of Japan* 70 (6).
- Helpenstein, P., Veverka, J., Thomas, P., Simonelli, D., Klaasen, K., Johnson, T., Fanale, F., Granahan, J., McEwen, A., Belton, M., and Chapman, C. (1996). *Icarus* 120 (1), 48–65.
- Helpenstein, P. et al. (1994). *Icarus* 107 (1), 37–60.
- Hendler, N. P. and Malhotra, R. (2020). *The Planetary Science Journal* 1 (3), 75.
- Herald, D., Frappa, E., Gault, D., Hayamizu, T., Kerr, S., Moore, J., and Giachini, B. (2019). *NASA Planetary Data System*, 3.
- Herzog, G. (2007). Cosmic-Ray Exposure Ages of Meteorites. In: *Treatise on Geochemistry*, 1–36.
- Hewins, R. H., Connolly, H. C., Lofgren, G. E. J., and Libourel, G. (2005). Experimental Constraints on Chondrule Formation. In: *Chondrites and the Protoplanetary Disk*. 341. Astronomical Society of the Pacific Conference Series, 286.
- Hirayama, K. (1918). *The Astronomical Journal* 31, 185.
- Hiroi, T., Zolensky, E. M., and Pieters, M. C. (2001). The Tagish Lake meteorite: First sample from D asteroids. In: *Antarctic Meteorites XXVI*. Ed. by P. Bagnall. 26, 29–31.
- Hiroi, T., Pieters, C. M., Zolensky, M. E., and Lipschutz, M. E. (1993). *Science* 261 (5124), 1016–1018.
- Hiroi, T. and Takeda, H. (1992). *Physics and Chemistry of Minerals* 19 (4).
- Hiroi, T., Zolensky, M. E., Pieters, C. M., and Lipschutz, M. E. (1996). *Meteoritics & Planetary Science* 31 (3), 321–327.
- Hodapp, K. W. et al. (2004). *Astronomische Nachrichten* 325 (6–8), 636–642.
- Hotelling, H. (1933). *Journal of educational psychology* 24.6, 417.
- Howell, E. S., Merényi, E., and Lebofsky, L. A. (1991). *Meteoritics* 26, 347.
- Howell, E. S., Merényi, E., and Lebofsky, L. A. (1994). *J. Geophys. Res.* 99.E5, 10847–10866.
- Huaman, M., Carruba, V., Domingos, R. C., and Aljbaae, S. (2017). *Monthly Notices of the Royal Astronomical Society* 468 (4), 4982–4991.
- Huang, J. et al. (2013). *Scientific Reports* 3 (1).

- Hughes, D. W. (1981). *Meteoritics* 16 (3), 269–281.
- Hung, D., Hanuš, J., Masiero, J. R., and Tholen, D. J. (2022). *The Planetary Science Journal* 3 (3), 56.
- Huss, G. R., Rubin, A. E., and Grossman, J. N. (2006). Thermal Metamorphism in Chondrites. In: *Meteorites and the Early Solar System II*, 567.
- Ieva, S., Dotto, E., Lazzaro, D., Fulvio, D., Perna, D., Mazzotta-Epifani, E., Medeiros, H., and Fulchignoni, M. (2018). *Monthly Notices of the Royal Astronomical Society* 479 (2), 2607–2614.
- Ivezić, Ž., Juric, M., Lupton, R. H., Tabachnik, S., and Quinn, T. (2002). eprint: [astro-ph/0208099v1](https://arxiv.org/abs/astro-ph/0208099v1).
- Ivezić, Ž., Kahn, S. M., Tyson, J. A., Abel, B., Acosta, E., Allsman, R., Alonso, D., AlSayyad, Y., Anderson, S. F., Andrew, J., and, e. al. et (2019). *ApJ* 873, 111.
- Ivezić, Ž., Ivezić, V., Moeyens, J., Lisse, C. M., Bus, S. J., Jones, L., Crill, B. P., Doré, O., and Emery, J. P. (2022). *Icarus* 371, 114696.
- Ivezić, Ž. et al. (2001). *The Astronomical Journal* 122 (5), 2749–2784.
- Jasmim, F. L., Lazzaro, D., Carvano, J. M. F., Mothé-Diniz, T., and Hasselmann, P. H. (2013). *Astronomy & Astrophysics* 552, A85.
- Javoy, M., Kaminski, E., Guyot, F., Andrault, D., Sanloup, C., Moreira, M., Labrosse, S., Jambon, A., Agrinier, P., Davaille, A., and Jaupart, C. (2010). *Earth and Planetary Science Letters* 293 (3-4), 259–268.
- Jenniskens, P. et al. (2009). *Nature* 458 (7237), 485–488.
- Jiang, H. and Ji, J. (2021). *The Astronomical Journal* 162 (2), 40.
- Johnson, T. V. and Fanale, F. P. (1973). *Journal of Geophysical Research* 78 (35), 8507–8518.
- Jones, R. H., Grossman, J. N., and Rubin, A. E. (2005). Chemical, Mineralogical and Isotopic Properties of Chondrules: Clues to Their Origin. In: *Chondrites and the Protoplanetary Disk*. 341. Astronomical Society of the Pacific Conference Series, 251.
- Jones, R. L., Chesley, S. R., Connolly, A. J., Harris, A. W., Ivezić, Z., Knežević, Z., Kubica, J., Milani, A., and Trilling, D. E. (2009). *Earth, Moon, and Planets* 105 (2-4), 101–105.
- Jones, T. D., Lebofsky, L. A., Lewis, J. S., and Marley, M. S. (1990). *Icarus* 88.1, 172–192.
- Jorda, L., Lamy, P., Gaskell, R., Kaasalainen, M., Groussin, O., Besse, S., and Faury, G. (2012). *Icarus* 221 (2), 1089–1100.
- Kaasalainen, M. (2001). *Icarus* 153 (1), 24–36.
- Kaasalainen, S., Piironen, J., Kaasalainen, M., Harris, A., Muinonen, K., and Cellino, A. (2003). *Icarus* 161 (1), 34–46.
- Karttunen, H. and Bowell, E. (1989). *A&A* 208.1-2, 320–326.
- Kasuga, T., Usui, F., Ootsubo, T., Hasegawa, S., and Kuroda, D. (2013). *The Astronomical Journal* 146 (1), 1.
- Kasuga, T., Usui, F., Shirahata, M., Kuroda, D., Ootsubo, T., Okamura, N., and Hasegawa, S. (2015). *The Astronomical Journal* 149 (2), 37.
- Keller, H. U. et al. (2010). *Science* 327 (5962), 190–193.
- Kelley, M. S., Vilas, F., Gaffey, M. J., and Abell, P. A. (2003). *Icarus* 165 (1), 215–218.
- Klinger, J. (1981). *Icarus* 47 (3), 320–324.

- Knežević, Z. and Milani, A. (2003). *Astronomy & Astrophysics* 403 (3), 1165–1173.
- Komatsu, M., Krot, A. N., Petaev, M. I., Ulyanov, A. A., Keil, K., and Miyamoto, M. (2001). *Meteoritics & Planetary Science* 36 (5), 629–641.
- Koren, S. C., Wright, E. L., and Mainzer, A. (2015). *Icarus* 258, 82–91.
- Krot, A. N., Hutcheon, I. D., Brearley, A. J., Pravdivtseva, O. V., Petaev, M. I., and Hohenberg, C. M. (2006). Timescales and Settings for Alteration of Chondritic Meteorites. In: *Meteorites and the Early Solar System II*, 525.
- Krot, A. N., Keil, K., Scott, E. R. D., Goodrich, C. A., and Weisberg, M. K. (2014). Classification of Meteorites and Their Genetic Relationships. In: *Meteorites and Cosmochemical Processes*. 1, 1–63.
- Krot, A. N., Nagashima, K., Alexander, C. M. O., Ciesla, F. J., Fujiya, W., and Bonal, L. (2015). Sources of Water and Aqueous Activity on the Chondrite Parent Asteroids. In: *Asteroids IV*, 635–660.
- Krot, A. N. (2019). *Meteoritics & Planetary Science* 54 (8), 1647–1691.
- Krot, A. N. and Keil, K. (2002). *Meteoritics & Planetary Science* 37 (1), 91–111.
- Kruijer, T. S., Burkhardt, C., Budde, G., and Kleine, T. (2017). *Proceedings of the National Academy of Sciences*, 201704461.
- Kuiper, G. P. (1951). On the Origin of the Solar System. In: *50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics*, 357.
- Kuroda, D., Ishiguro, M., Takato, N., Hasegawa, S., Abe, M., Tsuda, Y., Sugita, S., Usui, F., Hattori, T., Iwata, I., Imanishi, M., Terada, H., Choi, Y.-J., Watanabe, S.-i., and Yoshikawa, M. (2014). *Publications of the Astronomical Society of Japan* 66 (3), 51–51.
- Landsman, Z. A., Campins, H., Pinilla-Alonso, N., Hanuš, J., and Lorenzi, V. (2015). *Icarus* 252, 186–198.
- Lantz, C., Binzel, R., and DeMeo, F. (2018). *Icarus* 302, 10–17.
- Lantz, C., Brunetto, R., Barucci, M., Fornasier, S., Baklouti, D., Bourçois, J., and Godard, M. (2017). *Icarus* 285, 43–57.
- Lazzarin, M., Marchi, S., Barucci, M., Di Martino, M., and Barbieri, C. (2004). *Icarus* 169 (2), 373–384.
- Lazzarin, M., Marchi, S., Magrin, S., and Licandro, J. (2005). *Monthly Notices of the Royal Astronomical Society* 359 (4), 1575–1582.
- Lazzaro, D., Angeli, C., Carvano, J., Mothe-Diniz, T., Duffard, R., and Florczak, M. (2004). *Icarus* 172 (1), 179–220.
- Lazzaro, D., Angeli, C. A., Carvano, J. M., Mothe-Diniz, T., Duffard, R., and Florczak, M. (2007). NASA Planetary Data System, EAR-A-I0052-8-S3OS2–V1.0.
- Lebofsky, L. A., Sykes, M. V., Tedesco, E. F., Veeder, G. J., Matson, D. L., Brown, R. H., Gradie, J. C., Feierberg, M. A., and Rudy, R. J. (1986). *Icarus* 68 (2), 239–251.
- León, J. de, Licandro, J., Serra-Ricart, M., Pinilla-Alonso, N., and Campins, H. (2010). *Astronomy and Astrophysics* 517, A23.
- León, J. de, Mothé-Diniz, T., Licandro, J., Pinilla-Alonso, N., and Campins, H. (2011). *Astronomy & Astrophysics* 530, L12.
- León, J. de, Pinilla-Alonso, N., Campins, H., Licandro, J., and Marzo, G. (2012). *Icarus* 218 (1), 196–206.

- Lester, T. P., McCall, M. L., and Tatum, J. B. (1979). *Journal of the Royal Astronomical Society of Canada* 73, 233.
- Levison, H. F., Bottke, W. F., Gounelle, M., Morbidelli, A., Nesvorný, D., and Tsiganis, K. (2009). *Nature* 460 (7253), 364–366.
- Levison, H. F. and Dones, L. (2007). Comet Populations and Cometary Dynamics. In: *Encyclopedia of the Solar System*, 575–588.
- Li, J.-Y., Helfenstein, P., Buratti, B., Takir, D., and Clark, B. E. (2015). Asteroid Photometry. In: *Asteroids IV*, 129–150.
- Li, J.-Y., Le Corre, L., Schröder, S. E., Reddy, V., Denevi, B. W., Buratti, B. J., Mottola, S., Hoffmann, M., Gutierrez-Marques, P., Nathues, A., Russell, C. T., and Raymond, C. A. (2013). *Icarus* 226 (2), 1252–1274.
- Li, J.-Y. et al. (2016). *The Astrophysical Journal* 817 (2), L22.
- Licandro, J., Alí-Lagoa, V., Tancredi, G., and Fernández, Y. (2016). *Astronomy & Astrophysics* 585, A9.
- Licandro, J., Popescu, M., León, J. de, Morate, D., Vaduvescu, O., De Prá, M., and Ali-Laoga, V. (2018). *Astronomy & Astrophysics* 618, A170.
- Little, R. and Rubin, D. (2019). *Wiley Series in Probability and Statistics*.
- Lord, S. D. (1992). A new software tool for computing Earth's atmospheric transmission of near- and far-infrared radiation. NASA Technical Memorandum 103957.
- Lucas, M. P., Emery, J. P., Hiroi, T., and McSween, H. Y. (2019). *Meteoritics & Planetary Science* 54 (1), 157–180.
- Lucas, M. P., Emery, J. P., Pinilla-Alonso, N., Lindsay, S. S., and Lorenzi, V. (2017). *Icarus* 291, 268–287.
- MacKinnon, I. D. R. and Rietmeijer, F. J. M. (1987). *Reviews of Geophysics* 25 (7), 1527.
- Magri, C., Nolan, M. C., Ostro, S. J., and Giorgini, J. D. (2007). *Icarus* 186 (1), 126–151.
- Mahlke, M., Bouy, H., Altieri, B., Verdoes Kleijn, G., Carry, B., Bertin, E., Jong, J. T. A. de, Kuijken, K., McFarland, J., and Valentijn, E. (2018). *Astronomy & Astrophysics* 610.
- Mahlke, M., Carry, B., and Mattei, P.-A. (2022). *Astronomy & Astrophysics* 665, A26.
- Mahlke, M., Solano, E., Bouy, H., Carry, B., Verdoes Kleijn, G., and Bertin, E. (2019). *Astronomy and Computing* 28.
- Mahlke, M., Carry, B., and Denneau, L. (2021). *Icarus* 354, 114094.
- Mainzer, A., Bauer, J., Grav, T., Masiero, J., Cutri, R. M., Wright, E., Nugent, C. R., Stevenson, R., Clyne, E., Cukrov, G., and Masci, F. (2014a). *The Astrophysical Journal* 784 (2), 110.
- Mainzer, A., Grav, T., Masiero, J., Bauer, J., Cutri, R. M., McMillan, R. S., Nugent, C. R., Tholen, D., Walker, R., and Wright, E. L. (2012). *The Astrophysical Journal* 760 (1), L12.
- Mainzer, A., Grav, T., Masiero, J., Hand, E., Bauer, J., Tholen, D., McMillan, R. S., Spahr, T., Cutri, R. M., Wright, E., Watkins, J., Mo, W., and Maleszewski, C. (2011). *The Astrophysical Journal* 741.2, 90.
- Mainzer, A. et al. (2014b). *The Astrophysical Journal* 792 (1), 30.
- Malhotra, R. (1993). *Nature* 365 (6449), 819–821.

- Manning, P., DePalma, R., Murrasse, F., Burnham, D., Gurche, L., Klingler, J., Larson, P., Beasley, T., Oleinik, A., Geraki, T., Ignatyev, K., Egerton, V., and Wogelius, R. (2022). The identity of the Chicxulub impactor: KPg ejecta-encapsulated meteoric fragments.
- Marchi, S., Lazzarin, M., Paolicchi, P., and Magrin, S. (2005). *Icarus* 175 (1), 170–174.
- Marchi, S., Delbó, M., Morbidelli, A., Paolicchi, P., and Lazzarin, M. (2009). *Monthly Notices of the Royal Astronomical Society* 400 (1), 147–153.
- Marchi, S., Lazzarin, M., and Magrin, S. (2004). *Astronomy & Astrophysics* 420 (1), L5–L8.
- Marchis, F. et al. (2012). *Icarus* 221 (2), 1130–1161.
- Marsden, B. G. (1980). *Celestial Mechanics* 22 (1), 63–71.
- Marsset, M., Vernazza, P., Gourgéot, F., Dumas, C., Birlan, M., Lamy, P., and Binzel, R. P. (2014). *Astronomy & Astrophysics* 568, L7.
- Marsset, M. et al. (2017). *Astronomy & Astrophysics* 604, A64.
- Marsset, M., DeMeo, F. E., Binzel, R. P., Bus, S. J., Burbine, T. H., Burt, B., Moskovitz, N., Polishook, D., Rivkin, A. S., Slivan, S. M., and Thomas, C. (2020). *The Astrophysical Journal Supplement Series* 247 (2), 73.
- Marsset, M., DeMeo, F. E., Burt, B., Polishook, D., Binzel, R. P., Granvik, M., Vernazza, P., Carry, B., Bus, S. J., Slivan, S. M., Thomas, C. A., Moskovitz, N. A., and Rivkin, A. S. (2022). *The Astronomical Journal* 163 (4), 165.
- Masiero, J. R., DeMeo, F. E., Kasuga, T., and Parker, A. H. (2015). Asteroid Family Physical Properties. In: *Asteroids IV*.
- Masiero, J. and Cellino, A. (2009). *Icarus* 199 (2), 333–337.
- Masiero, J. R., Grav, T., Mainzer, A. K., Nugent, C. R., Bauer, J. M., Stevenson, R., and Sonnett, S. (2014). *The Astrophysical Journal* 791.2, 121.
- Masiero, J. R., Mainzer, A. K., Bauer, J. M., Cutri, R. M., Grav, T., Kramer, E., Pittichová, J., Sonnett, S., and Wright, E. L. (2020a). *The Planetary Science Journal* 1 (1), 5.
- Masiero, J. R., Mainzer, A. K., Bauer, J. M., Cutri, R. M., Grav, T., Kramer, E., Pittichová, J., and Wright, E. L. (2021). *The Planetary Science Journal* 2 (4), 162.
- Masiero, J. R., Mainzer, A. K., Bauer, J. M., Grav, T., Nugent, C. R., and Stevenson, R. (2013). *The Astrophysical Journal* 770 (1), 7.
- Masiero, J. R., Mainzer, A. K., Grav, T., Bauer, J. M., Cutri, R. M., Nugent, C., and Cabrera, M. S. (2012). *The Astrophysical Journal* 759 (1), L8.
- Masiero, J. R., Mainzer, A. K., and Wright, E. L. (2018). *The Astronomical Journal* 156.2, 62.
- Masiero, J. R., Nugent, C., Mainzer, A. K., Wright, E. L., Bauer, J. M., Cutri, R. M., Grav, T., Kramer, E., and Sonnett, S. (2017). *The Astronomical Journal* 154 (4), 168.
- Masiero, J. R., Smith, P., Teodoro, L. D., Mainzer, A. K., Cutri, R. M., Grav, T., and Wright, E. L. (2020b). *The Planetary Science Journal* 1 (1), 9.
- Masiero, J. R., Wright, E. L., and Mainzer, A. K. (2019). *The Astronomical Journal* 158 (3), 97.
- Masiero, J. R. et al. (2011). *The Astrophysical Journal* 741.2, 68.
- Matlovič, P., Leon, J. de, Medeiros, H., Popescu, M., Rizos, J. L., and Mansour, J.-A. (2020). *Astronomy & Astrophysics* 643, A107.

- Matson, D., Veeder, G., Tedesco, E., Lebofsky, L., and Walker, R. (1986). *Advances in Space Research* 6 (7), 47–56.
- Matter, A., Delbo, M., Carry, B., and Ligorì, S. (2013). *Icarus* 226 (1), 419–427.
- Matter, A., Delbo, M., Ligorì, S., Crouzet, N., and Tanga, P. (2011). *Icarus* 215 (1), 47–56.
- McCord, T. B., Adams, J. B., and Johnson, T. V. (1970). *Science* 168 (3938), 1445–1447.
- Michel, P., Migliorini, F., Morbidelli, A., and Zappalà, V. (2000). *Icarus* 145 (2), 332–347.
- Migliorini, A., De Sanctis, M. C., Lazzaro, D., and Ammannito, E. (2018). *Monthly Notices of the Royal Astronomical Society* 475 (1), 353–358.
- Migliorini, A., De Sanctis, M. C., Lazzaro, D., and Ammannito, E. (2017). *Monthly Notices of the Royal Astronomical Society* 464 (2), 1718–1726.
- Milani, A. and Knežević, Z. (1994). *Icarus* 107 (2), 219–254.
- Mittlefehldt, D. W. (2015). *Geochemistry* 75 (2), 155–183.
- Miyamoto, H., Yano, H., Scheeres, D. J., Abe, S., Barnouin-Jha, O., Cheng, A. F., Demura, H., Gaskell, R. W., Hirata, N., Ishiguro, M., Michikami, T., Nakamura, A. M., Nakamura, R., Saito, J., and Sasaki, S. (2007). *Science* 316 (5827), 1011–1014.
- Mommert, M., Kelley, M., De Val-Borro, M., Li, J.-Y., Guzman, G., Sipőcz, B., Āurech, J., Granvik, M., Grundy, W., Moskovitz, N., Penttilä, A., and Samarasinha, N. (2019). *Journal of Open Source Software* 4 (38), 1426.
- Morbidelli, A., Bottke W. F., J., Froeschlé, C., and Michel, P. (2002). Origin and Evolution of Near-Earth Objects. In: *Asteroids III*, 409–422.
- Morbidelli, A., Levison, H. F., Tsiganis, K., and Gomes, R. (2005). *Nature* 435 (7041), 462–465.
- Morbidelli, A., Nesvorný, D., Bottke, W., and Marchi, S. (2021). *Icarus* 356, 114256.
- Morbidelli, A. (1997). *Icarus* 127 (1), 1–12.
- Morbidelli, A. (2002). Modern celestial mechanics : aspects of solar system dynamics.
- Morgan, W. W., Keenan, P. C., and Kellman, E. (1943). An atlas of stellar spectra, with an outline of spectral classification.
- Moskovitz, N., Jedicke, R., and Willman, M. (2009). NASA Planetary Data System, EAR-A-I0039/I0576-4-SDSSMOCVSPEC-V1.0.
- Moskovitz, N. A., Willman, M., Burbine, T. H., Binzel, R. P., and Bus, S. J. (2010). *Icarus* 208 (2), 773–788.
- Moskovitz, N. A. et al. (2019). *Icarus* 333, 165–176.
- Mothé-Diniz, T., Carvano, J., Bus, S., Duffard, R., and Burbine, T. (2008). *Icarus* 195 (1), 277–294.
- Mothé-Diniz, T. and Nesvorný, D. (2008). *Astronomy & Astrophysics* 492 (2), 593–598.
- Mothé-Diniz, T., Roig, F., and Carvano, J. (2005). *Icarus* 174 (1), 54–80.
- Mueller, B. E., Tholen, D. J., Hartmann, W. K., and Cruikshank, D. P. (1992). *Icarus* 97 (1), 150–154.
- Mueller, M. et al. (2011). *The Astronomical Journal* 141 (4), 109.
- Muinonen, K., Mishchenko, M. I., Dlugach, J. M., Zubko, E., Penttilä, A., and Videen, G. (2012). *The Astrophysical Journal* 760 (2).

- Muironen, K., Piironen, J., Kaasalainen, S., and Cellino, A. (2002). *Memorie della Societa Astronomica Italiana* 73.3, 716.
- Muironen, K., Belskaya, I. N., Cellino, A., Delbò, M., Levasseur-Regourd, A.-C., Penttilä, A., and Tedesco, E. F. (2010a). *Icarus* 209 (2), 542–555.
- Muironen, K., Tyynela, J., Zubko, E., and Videen, G. (2010b). Coherent backscattering in planetary regoliths. In: *Light Scattering Rev.* 5, 477–518.
- Müller, T. G. and Blommaert, J. A. D. L. (2004). *Astronomy & Astrophysics* 418 (1), 347–356.
- Müller, T. G., Kiss, C., Scheirich, P., Pravec, P., O'Rourke, L., Vilenius, E., and Altieri, B. (2014). *Astronomy & Astrophysics* 566, A22.
- Mustard, J. and Hays, J. (1997). *Icarus* 125 (1), 145–163.
- Nakamura, T. et al. (2011). *Science* 333 (6046), 1113–1116.
- Nathues, A. (2010). *Icarus* 208 (1), 252–275.
- Nedelcu, D. A., Birlan, M., Vernazza, P., Binzel, R. P., Fulchignoni, M., and Barucci, M. A. (2007). *Astronomy & Astrophysics* 473 (3), L33–L36.
- Neeley, J., Clark, B., Ockert-Bell, M., Shepard, M., Conklin, J., Cloutis, E., Fornasier, S., and Bus, S. (2014). *Icarus* 238, 37–50.
- Nesvorný, D. (2015). NASA Planetary Data System, EAR-A-VARGBDDET-5-NESVORNYFAM-V3.0.
- Nesvorný, D. and Vokrouhlický, D. (2016). *The Astrophysical Journal* 825 (2), 94.
- Nesvorný, D., Vokrouhlický, D., Bottke, W. F., and Levison, H. F. (2018). *Nature Astronomy* 2 (11), 878–882.
- Nesvorný, D., Vokrouhlický, D., Morbidelli, A., and Bottke, W. F. (2009). *Icarus* 200 (2), 698–701.
- Nicholson, S. B. (1961). Leaflet of the Astronomical Society of the Pacific 8.381, 239.
- Novakovic, B., Vokrouhlický, D., Spoto, F., and Nesvorný, D. (2022). eprint: 2205.06340v1.
- Nugent, C. R., Mainzer, A., Bauer, J., Cutri, R. M., Kramer, E. A., Grav, T., Masiero, J., Sonnett, S., and Wright, E. L. (2016). *The Astronomical Journal* 152 (3), 63.
- Nugent, C. R., Mainzer, A., Masiero, J., Bauer, J., Cutri, R. M., Grav, T., Kramer, E., Sonnett, S., Stevenson, R., and Wright, E. L. (2015). *ApJ* 814.2, 117.
- Ockert-Bell, M. E., Clark, B. E., Shepard, M. K., Isaacs, R. A., Cloutis, E. A., Fornasier, S., and Bus, S. J. (2010). *Icarus* 210 (2), 674–692.
- Ockert-Bell, M. E., Clark, B. E., Shepard, M. K., Rivkin, A. S., Binzel, R. P., Thomas, C. A., DeMeo, F. E., Bus, S., and Shah, S. (2008). *Icarus* 195 (1), 206–219.
- Olbers, N. (1803). *Annalen der Physik* 14 (5), 38–45.
- Olbers, N. (1807). *Annalen der Physik* 25 (3), 344–344.
- Ostro, S. J., Hudson, R. S., Benner, L. A. M., Giorgini, J. D., Magri, C., Margot, J. L., and Nolan, M. C. (2002). Asteroid Radar Astronomy. In: *Asteroids III*, 151–168.
- Ostro, S. J. and Connelly, R. (1984). *Icarus* 57 (3), 443–463.
- Ostrowski, D. R., Lacy, C. H., Gietzen, K. M., and Sears, D. W. (2011). *Icarus* 212 (2), 682–696.

- Oszkiewicz, D., Muinonen, K., Howell, E., Trilling, D., Penttilä, A., Pieniluoma, T., Wasserman, L., and Enga, M.-T. (2011). *Journal of Quantitative Spectroscopy and Radiative Transfer* 112 (11), 1919–1929.
- Oszkiewicz, D., Kankiewicz, P., Włodarczyk, I., and Kryszczyńska, A. (2015). *Astronomy & Astrophysics* 584, A18.
- Oszkiewicz, D. et al. (2020). *Astronomy & Astrophysics* 643, A117.
- Oszkiewicz, D. A., Skiff, B. A., Moskovitz, N., Kankiewicz, P., Marciniak, A., Licandro, J., Galiazzi, M. A., and Zeilinger, W. W. (2017). *Astronomy & Astrophysics* 599, A107.
- Parker, A. H. and Kavelaars, J. J. (2010). *The Astrophysical Journal* 722 (2), L204–L208.
- Peacock, J. A. (1983). *Monthly Notices of the Royal Astronomical Society* 202 (3), 615–627.
- Pearson, K. (1901). *The London, Edinburgh, and Dublin philosophical magazine and journal of science* 2.11, 559–572.
- Penttilä, A., Shevchenko, V., Wilkman, O., and Muinonen, K. (2016). *Planetary and Space Science* 123, 117–125.
- Perna, D., Barucci, M., Fulchignoni, M., Popescu, M., Belskaya, I., Fornasier, S., Doressoundiram, A., Lantz, C., and Merlin, F. (2018). *Planetary and Space Science* 157, 82–95.
- Perrine, C. D. (1902). *Lick Observatory Bulletins* 1, 183.
- Pieters, C. M. (1983). *Journal of Geophysical Research: Solid Earth* 88 (B11), 9534–9544.
- Pinilla-Alonso, N., De Prá, M., León, J. de, Morate, D., Lorenzi, V., Arredondo, A., Campins, H., Licandro, J., Delbó, M., Cabrera-Lavers, A., Walsh, K., DeMeo, F., and Sarid, G. (2021). *NASA Planetary Data System*, 8.
- Pinilla-Alonso, N., León, J. de, Walsh, K., Campins, H., Lorenzi, V., Delbó, M., DeMeo, F., Licandro, J., Landsman, Z., Lucas, M., Alí-Lagoa, V., and Burt, B. (2016). *Icarus* 274, 231–248.
- Polishook, D., Jacobson, S. A., Morbidelli, A., and Aharonson, O. (2017). *Nature Astronomy* 1 (8).
- Polishook, D., Moskovitz, N., Binzel, R. P., DeMeo, F. E., Vokrouhlický, D., Žižka, J., and Oszkiewicz, D. (2014). *Icarus* 233, 9–26.
- Popescu, M. (2012). *Techniques for Asteroid Spectroscopy*. PhD Thesis.
- Popescu, M., León, J. de, Fuente Marcos, C. de la, Vaduvescu, O., Fuente Marcos, R. de la, Licandro, J., Pinter, V., Tatsumi, E., Zamora, O., Fariña, C., and Curelaru, L. (2020). *Monthly Notices of the Royal Astronomical Society* 496 (3), 3572–3581.
- Popescu, M., Birlan, M., Binzel, R., Vernazza, P., Barucci, A., Nedelcu, D. A., DeMeo, F., and Fulchignoni, M. (2011). *Astronomy & Astrophysics* 535, A15.
- Popescu, M., Birlan, M., and Nedelcu, D. A. (2012). *Astronomy & Astrophysics* 544, A130.
- Popescu, M., Birlan, M., Nedelcu, D. A., Vaubaillon, J., and Cristescu, C. P. (2014). *Astronomy & Astrophysics* 572, A106.
- Popescu, M., Licandro, J., Carvano, J. M., Stoicescu, R., León, J. de, Morate, D., Boacă, I. L., and Cristescu, C. P. (2018a). *Astronomy & Astrophysics* 617, A12.

- Popescu, M., Licandro, J., Morate, D., León, J. de, Nedelcu, D. A., Rebolo, R., McMahon, R. G., Gonzalez-Solares, E., and Irwin, M. (2016). *A&A* 591, A115.
- Popescu, M. et al. (2019). *Astronomy & Astrophysics* 627, A124.
- Popescu, M., Perna, D., Barucci, M. A., Fornasier, S., Doressoundiram, A., Lantz, C., Merlin, F., Belskaya, I. N., and Fulchignoni, M. (2018b). *Monthly Notices of the Royal Astronomical Society* 477 (2), 2786–2795.
- Pravec, P., Harris, A. W., Kušnirák, P., Galád, A., and Hornoch, K. (2012). *Icarus* 221 (1), 365–387.
- Rayner, J. T., Toomey, D. W., Onaka, P. M., Denault, A. J., Stahlberger, W. E., Vacca, W. D., Cushing, M. C., and Wang, S. (2003). *Publications of the Astronomical Society of the Pacific* 115 (805), 362–382.
- Reddy, V. (2010). NASA Planetary Data System, EAR-A-10046-5-REDDYSPEC-V1.0.
- Reddy, V., Dunn, T. L., Thomas, C. A., Moskovitz, N. A., and Burbine, T. H. (2015). Mineralogy and Surface Composition of Asteroids. In: *Asteroids IV*.
- Reddy, V. and Sanchez, J. A. (2016). NASA Planetary Data System, EAR-A-10046-3-REDDYMSPEC-V1.0.
- Reddy, V. and Sanchez, J. A. (2017). NASA Planetary Data System.
- Reddy, V., Carvano, J. M., Lazzaro, D., Michtchenko, T. A., Gaffey, M. J., Kelley, M. S., Mothé-Diniz, T., Alvarez-Candal, A., Moskovitz, N. A., Cloutis, E. A., and Ryan, E. L. (2011). *Icarus* 216 (1), 184–197.
- Reddy, V., Sanchez, J. A., Bottke, W. F., Cloutis, E. A., Izawa, M. R., O'Brien, D. P., Mann, P., Cuddy, M., Le Corre, L., Gaffey, M. J., and Fujihara, G. (2014). *Icarus* 237, 116–130.
- Reddy, V., Sanchez, J. A., Furfaro, R., Binzel, R. P., Burbine, T. H., Le Corre, L., Hardersen, P. S., Bottke, W. F., and Brozovic, M. (2018). *The Astronomical Journal* 155 (3), 140.
- Rivkin, A. S., Binzel, R. P., and Bus, S. J. (2005). *Icarus* 175 (1), 175–180.
- Rivkin, A. S., Binzel, R. P., Sunshine, J., Bus, S. J., Burbine, T. H., and Saxena, A. (2004). *Icarus* 172 (2), 408–414.
- Rivkin, A. S., Campins, H., Emery, J. P., Howell, E. S., Licandro, J., Takir, D., and Vilas, F. (2015a). Astronomical Observations of Volatiles on Asteroids. In: *Asteroids IV*.
- Rivkin, A. S. (1995). *Icarus* 117 (1), 90–100.
- Rivkin, A. S. (2012). *Icarus* 221 (2), 744–752.
- Rivkin, A. S., Thomas, C. A., Howell, E. S., and Emery, J. P. (2015b). *The Astronomical Journal* 150 (6), 198.
- Rivkin, A. S., Trilling, D. E., Thomas, C. A., DeMeo, F., Spahr, T. B., and Binzel, R. P. (2007). *Icarus* 192 (2), 434–441.
- Rocher, P. and Cavelier, C. (1996). Production d'elements orbitaux de cometes sur PC. In: *Dynamics, Ephemerides, and Astrometry of the Solar System*. 172, 357.
- Rodrigo, C. and Solano, E. (2020). The SVO Filter Profile Service. In: *XIV.0 Scientific Meeting (virtual) of the Spanish Astronomical Society*, 182, 182.
- Rodrigo, C., Solano, E., and Bayo, A. (2012).
- Rozitis, B., Duddy, S. R., Green, S. F., and Lowry, S. C. (2013). *Astronomy & Astrophysics* 555, A20.

- Rozitis, B. and Green, S. F. (2014). *Astronomy & Astrophysics* 568, A43.
- Rubin, D. B. and Thayer, D. T. (1982). *Psychometrika* 47 (1), 69–76.
- Russell, C. T. et al. (2012). *Science* 336 (6082), 684–686.
- Russell, C. T. et al. (2016). *Science* 353 (6303), 1008–1010.
- Russell, S. S., MacPherson, G. J., Leshin, L. A., and McKeegan, K. D. (2000). *Earth and Planetary Science Letters* 184 (1), 57–74.
- Ryan, E. L., Mizuno, D. R., Shenoy, S. S., Woodward, C. E., Carey, S. J., Noriega-Crespo, A., Kraemer, K. E., and Price, S. D. (2015). *Astronomy & Astrophysics* 578, A42.
- Ryan, E. L. and Woodward, C. E. (2010). *The Astronomical Journal* 140 (4), 933–943.
- Salisbury, J. W. and Hunt, G. R. (1969). *Nature* 222 (5189), 132–136.
- Salisbury, J. W. and Hunt, G. R. (1974). *Journal of Geophysical Research* 79 (29), 4439–4441.
- Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). *PeerJ Computer Science* 2, e55.
- Sanchez, J. A., Michelsen, R., Reddy, V., and Nathues, A. (2013). *Icarus* 225 (1), 131–140.
- Sanchez, J. A., Reddy, V., Kelley, M. S., Cloutis, E. A., Bottke, W. F., Nesvorný, D., Lucas, M. P., Hardersen, P. S., Gaffey, M. J., Abell, P. A., and Corre, L. L. (2014). *Icarus* 228, 288–300.
- Sanchez, J. A., Reddy, V., Nathues, A., Cloutis, E. A., Mann, P., and Hiesinger, H. (2012). *Icarus* 220.1, 36–50.
- Sasaki, T., Sasaki, S., Watanabe, J.-i., Sekiguchi, T., Yoshida, F., Kawakita, H., Fuse, T., Takato, N., Dermawan, B., and Ito, T. (2004). *The Astrophysical Journal* 615 (2), L161–L164.
- Savitzky, A. and Golay, M. J. E. (1964). *Analytical Chemistry* 36, 1627–1639.
- Scheinberg, A., Fu, R. R., Elkins-Tanton, L. T., and Weiss, B. P. (2015). Asteroid Differentiation: Melting and Large-Scale Structure. In: *Asteroids IV*, 533–552.
- Schumacher, G. and Gay, J. (2001). *Astronomy & Astrophysics* 368 (3), 1108–1114.
- Scott, E. (1972). *Geochimica et Cosmochimica Acta* 36 (11), 1205–1236.
- Sears, D. W. G. (2004). The Origin of Chondrules and Chondrites.
- Sears, D. W. G. and Dodd, R. T. (1988). Overview and classification of meteorites. In: *Meteorites and the Early Solar System*, 3–31.
- Sergeyev, A. V. and Carry, B. (2021). *Astronomy & Astrophysics* 652, A59.
- Shepard, M. K., Clark, B. E., Ockert-Bell, M., Nolan, M. C., Howell, E. S., Magri, C., Giorgini, J. D., Benner, L. A., Ostro, S. J., Harris, A. W., Warner, B. D., Stephens, R. D., and Mueller, M. (2010). *Icarus* 208 (1), 221–237.
- Shepard, M. K., Kressler, K. M., Clark, B. E., Ockert-Bell, M. E., Nolan, M. C., Howell, E. S., Magri, C., Giorgini, J. D., Benner, L. A., and Ostro, S. J. (2008a). *Icarus* 195 (1), 220–225.
- Shepard, M. K., Timerson, B., Scheeres, D. J., Benner, L. A., Giorgini, J. D., Howell, E. S., Magri, C., Nolan, M. C., Springmann, A., Taylor, P. A., and Virkki, A. (2018). *Icarus* 311, 197–209.
- Shepard, M. K. et al. (2008b). *Icarus* 193 (1), 20–38.
- Shepard, M. K. et al. (2008c). *Icarus* 195 (1), 184–205.

- Shepard, M. K. et al. (2015). *Icarus* 245, 38–55.
- Shestopalov, D., Golubeva, L., McFadden, L., Fornasier, S., and Taran, M. (2010). *Planetary and Space Science* 58 (10), 1400–1403.
- Shevchenko, V. (2002). *Icarus* 155 (2), 365–374.
- Shevchenko, V., Chiorny, V., Gaftonyuk, N., Krugly, Y., Belskaya, I., Tereschenko, I., and Velichko, F. (2008). *Icarus* 196 (2), 601–611.
- Shevchenko, V., Belskaya, I., Chiorny, V., Piironen, J., Erikson, A., Neukum, G., and Mohamed, R. (1997). *Planetary and Space Science* 45 (12), 1615–1623.
- Shevchenko, V. G., Belskaya, I. N., Muinonen, K., Penttilä, A., Krugly, Y. N., Velichko, F. P., Chiorny, V. G., Slyusarev, I. G., Gaftonyuk, N. M., and Tereschenko, I. A. (2016). *Planetary and Space Science* 123, 101–116.
- Shkuratov, Y., Starukhina, L., Hoffmann, H., and Arnold, G. (1999). *Icarus* 137 (2), 235–246.
- Shkuratov, Y., Bondarenko, S., Kaydash, V., Videen, G., Muñoz, O., and Volten, H. (2007). *Journal of Quantitative Spectroscopy and Radiative Transfer* 106.1. IX Conference on Electromagnetic and Light Scattering by Non-Spherical Particles, 487–508.
- Shoemaker, E. M. and Helin, E. F. (1978). Earth-approaching asteroids: Populations, origin, and compositional types. In: *NASA Conference Publication*.
- Sierks, H. et al. (2011). *Science* 334 (6055), 487–490.
- Singer, R. B. and Roush, T. L. (1985). *Journal of Geophysical Research* 90 (B14), 12434.
- Spearman, C. (1904). *The American Journal of Psychology* 15.2, 201–292.
- Strazzulla, G., Dotto, E., Binzel, R., Brunetto, R., Barucci, M., Blanco, A., and Orofino, V. (2005). *Icarus* 174 (1), 31–35.
- Sultana, R., Poch, O., Beck, P., Schmitt, B., and Quirico, E. (2021). *Icarus* 357, 114141.
- Sunshine, J. M., Connolly, H. C., McCoy, T. J., Bus, S. J., and La Croix, L. M. (2008). *Science* 320 (5875), 514–517.
- Sunshine, J. M., Bus, S. J., Corrigan, C. M., McCoy, T. J., and Burbine, T. H. (2007). *Meteoritics & Planetary Science* 42 (2), 155–170.
- Surdej, A. and Surdej, J. (1978). *A&A* 66.1-2, 31–36.
- Takir, D. and Emery, J. P. (2022). A Large 3-Micron Spectroscopic Survey of Mid-Outer Main Belt Asteroids. In: *LPI Contributions*. 2678, 1481.
- Takir, D. and Emery, J. P. (2012). *Icarus* 219 (2), 641–654.
- Tanga, P. et al. (2022). *Astronomy & Astrophysics*. eprint: 2206.05561.
- Tanga, P., Hestroffer, D., Delbò, M., Frouard, J., Mouret, S., and Thuillot, W. (2008). *Planetary and Space Science* 56 (14), 1812–1818.
- Tatsumi, E., Domingue, D., Hirata, N., Kitazato, K., Vilas, F., Lederer, S., Weissman, P. R., Lowry, S. C., and Sugita, S. (2018). *Icarus* 311, 175–196.
- Tatsumi, E., Tinaut-Ruano, F., León, J. de, and Licandro, J. (2022). eprint: 2205.13917v1.
- Tedesco, E. F., Williams, J. G., Matson, D. L., Weeder, G. J., Gradie, J. C., and Lebofsky, L. A. (1989). *The Astronomical Journal* 97, 580.
- Tedesco, E. F., Noah, P. V., Noah, M., and Price, S. D. (2002). *The Astronomical Journal* 123 (2), 1056–1085.
- Tholen, D. J. and Barucci, M. A. (1989). Asteroid taxonomy. In: *Asteroids II*, 298–315.

- Tholen, D. J. (1984). Asteroid Taxonomy from Cluster Analysis of Photometry. PhD Thesis. University of Arizona, Tucson.
- Thomas, C. A. and Binzel, R. P. (2010). *Icarus* 205 (2), 419–429.
- Thomas, C. A., Trilling, D. E., and Rivkin, A. S. (2012). *Icarus* 219 (1), 505–507.
- Thomas, C. A., Trilling, D. E., Rivkin, A. S., and Linder, T. (2021). *The Astronomical Journal* 161 (3), 99.
- Thomas, P. (2000). *Icarus* 145 (2), 348–350.
- Thomas, P., Belton, M., Carcich, B., Chapman, C., Davies, M., Sullivan, R., and Veverka, J. (1996). *Icarus* 120 (1), 20–32.
- Thomas, P., Veverka, J., Bell, J., Clark, B., Carcich, B., Joseph, J., Robinson, M., McFadden, L., Malin, M., Chapman, C., Merline, W., and Murchie, S. (1999). *Icarus* 140 (1), 17–27.
- Thomas, P., Veverka, J., Simonelli, D., Helfenstein, P., Carcich, B., Belton, M., Davies, M., and Chapman, C. (1994). *Icarus* 107 (1), 23–36.
- Tipping, M. E. and Bishop, C. M. (1999a). Neural computation 11.2, 443–482.
- Tipping, M. E. and Bishop, C. M. (1999b). Journal of the Royal Statistical Society: Series B (Statistical Methodology) 61.3, 611–622.
- Tiscareno, M. S. and Malhotra, R. (2003). *The Astronomical Journal* 126 (6), 3122–3131.
- Tiscareno, M. S. and Malhotra, R. (2009). *The Astronomical Journal* 138 (3), 827–837.
- Tonry, J. L., Denneau, L., Heinze, A. N., Stalder, B., Smith, K. W., Smartt, S. J., Stubbs, C. W., Weiland, H. J., and Rest, A. (2018). *Publications of the Astronomical Society of the Pacific* 130 (988).
- Tresp, V., Ahmad, S., and Neuneier, R. (1993). Training Neural Networks with Deficient Data. In: *Advances in Neural Information Processing Systems*. 6.
- Trilling, D. E. et al. (2010). *The Astronomical Journal* 140 (3), 770–784.
- Trilling, D. E., Mommert, M., Hora, J., Chesley, S., Emery, J., Fazio, G., Harris, A., Mueller, M., and Smith, H. (2016). *The Astronomical Journal* 152 (6), 172.
- Tsiganis, K., Gomes, R., Morbidelli, A., and Levison, H. F. (2005). *Nature* 435 (7041), 459–461.
- Usui, F., Kuroda, D., Müller, T. G., Hasegawa, S., Ishiguro, M., Ootsubo, T., Ishihara, D., Kataza, H., Takita, S., Oyabu, S., Ueno, M., Matsuhara, H., and Onaka, T. (2011). *Publications of the Astronomical Society of Japan* 63 (5), 1117–1138.
- Vallenari, A., Brown, A., Prusti, T., and None, N. (2022). *Astronomy & Astrophysics*.
- Van Schmus, W. and Wood, J. (1967). *Geochimica et Cosmochimica Acta* 31 (5), 747–765.
- Veeder, G., Matson, D., and Tedesco, E. (1983). *Icarus* 55 (1), 177–180.
- Verbeek, J. (2004). Mixture models for clustering and dimension reduction. PhD Thesis. Universiteit van Amsterdam.
- Vereš, P. et al. (2015). *Icarus* 261, 34–47.
- Vernazza, P., Binzel, R. P., Thomas, C. A., DeMeo, F. E., Bus, S. J., Rivkin, A. S., and Tokunaga, A. T. (2008). *Nature* 454 (7206), 858–860.

- Vernazza, P., Birlan, M., Rossi, A., Dotto, E., Nesvorný, D., Brunetto, R., Fornasier, S., Fulchignoni, M., and Renner, S. (2006). *Astronomy & Astrophysics* 460 (3), 945–951.
- Vernazza, P., Carry, B., Emery, J., Hora, J., Cruikshank, D., Binzel, R., Jackson, J., Helbert, J., and Maturilli, A. (2010). *Icarus* 207 (2), 800–809.
- Vernazza, P., Lamy, P., Groussin, O., Hiroi, T., Jorda, L., King, P., Izawa, M., Marchis, F., Birlan, M., and Brunetto, R. (2011). *Icarus* 216 (2), 650–659.
- Vernazza, P., Marsset, M., Beck, P., Binzel, R. P., Birlan, M., Brunetto, R., Demeo, F. E., Djouadi, Z., Dumas, C., Merouane, S., Mousis, O., and Zanda, B. (2015). *The Astrophysical Journal* 806.2, 204.
- Vernazza, P., Marsset, M., Beck, P., Binzel, R. P., Birlan, M., Cloutis, E. A., DeMeo, F. E., Dumas, C., and Hiroi, T. (2016). *The Astronomical Journal* 152 (3), 54.
- Vernazza, P., Mothé-Diniz, T., Barucci, M. A., Birlan, M., Carvano, J. M., Strazzulla, G., Fulchignoni, M., and Migliorini, A. (2005). *Astronomy & Astrophysics* 436 (3), 1113–1121.
- Vernazza, P., Zanda, B., Binzel, R. P., Hiroi, T., DeMeo, F. E., Birlan, M., Hewins, R., Ricci, L., Barge, P., and Lockhart, M. (2014). *The Astrophysical Journal* 791 (2), 120.
- Vernazza, P. et al. (2021). *Astronomy & Astrophysics* 654, A56.
- Veverka, J. et al. (2000). *Science* 289 (5487), 2088–2097.
- Viikinkoski, M., Hanuš, J., Kaasalainen, M., Marchis, F., and Ďurech, J. (2017). *Astronomy & Astrophysics* 607, A117.
- Vilas, F., Smith, B. A., McFadden, L. A., Gaffey, M. J., Larson, S. M., Hatch, E. C., and Jarvis, K. S. (2006). NASA Planetary Data System, EAR-A-3-RDR-VILAS-ASTEROID-SPECTRA-V1.1.
- Vilas, F. and Gaffey, M. J. (1989). *Science* 246 (4931), 790–792.
- Vivó-Truyols, G. and Schoenmakers, P. J. (2006). *Analytical Chemistry* 78 (13), 4598–4608.
- Warner, B. D., Harris, A. W., and Pravec, P. (2009). *Icarus* 202 (1), 134–146.
- Warren, P. H. (2011). *Earth and Planetary Science Letters* 311 (1-2), 93–100.
- Watters, T. R. and Prinz, M. (1979). Lunar and Planetary Science Conference Proceedings 1, 1073–1093.
- Weisberg, M. K., McCoy, T. J., and Krot, A. N. (2006). Systematics and evaluation of meteorite classification. In: *Meteorites and the Early Solar System II*. 19.
- Willman, M., Jedicke, R., and Moskovitz, N. (2009). NASA Planetary Data System, EAR-A-I0039-4-IANNINISPEC-V1.0.
- Willman, M., Jedicke, R., Nesvorný, D., Moskovitz, N., Ivezić, Ž., and Fevig, R. (2008). *Icarus* 195 (2), 663–673.
- Włodarczyk, I. (2020). *Monthly Notices of the Royal Astronomical Society* 500 (3), 3569–3578.
- Wong, I. and Brown, M. E. (2015). *The Astronomical Journal* 150 (6), 174.
- Wong, I., Brown, M. E., and Emery, J. P. (2017). *The Astronomical Journal* 154 (3), 104.
- Wright, E. L., Mainzer, A., Masiero, J., Grav, T., and Bauer, J. (2016). *The Astronomical Journal* 152 (4), 79.

- Xu, S. (1994). CCD photometry and spectroscopy of small main-belt asteroids. PhD Thesis. Massachusetts Institute of Technology.
- Xu, S., Binzel, R. P., Burbine, T. H., and Bus, S. J. (1995). *Icarus* 115 (1), 1–35.
- Yada, T. et al. (2022). *Nature Astronomy* 6 (2), 214–220.
- Yang, B., Hanuš, J., Brož, M., Chrenko, O., Willman, M., Ševeček, P., Masiero, J., and Kaluna, H. (2020). *Astronomy & Astrophysics* 643, A38.
- Yang, B., Wahhaj, Z., Beauvalet, L., Marchis, F., Dumas, C., Marsset, M., Nielsen, E. L., and Vachier, F. (2016). *The Astrophysical Journal* 820 (2), L35.
- Yang, B. and Jewitt, D. (2007). *The Astronomical Journal* 134 (1), 223–228.
- Yang, B. and Jewitt, D. (2011). *The Astronomical Journal* 141 (3), 95.
- Yoshikawa, M., Kawaguchi, J., Fujiwara, A., and Tsuchiyama, A. (2015). Hayabusa Sample Return Mission. In: *Asteroids IV*, 397–418.
- Yu, L.-L., Ji, J., and Ip, W.-H. (2017). *Research in Astronomy and Astrophysics* 17 (7), 070.
- Zappalà, V., Cellino, A., Barucci, A. M., Fulchignoni, M., and Lupishko, D. F. (1990). *A&A* 231.2, 548–560.
- Zappalà, V., Cellino, A., Di Martino, M., Migliorini, F., and Paolicchi, P. (1997). *Icarus* 129 (1), 1–20.
- Zeller, E. J. and Ronca, L. B. (1967). *Icarus* 7 (1-3), 372–379.
- Zellner, B. (1973). Polarimetric Albedos of Asteroids. In: *Bulletin of the American Astronomical Society*. 5, 388.
- Zellner, B. (1975). *The Astrophysical Journal* 198, L45.
- Zellner, B. and Bowell, E. (1977). Asteroid compositional types and their distributions. In: *IAU Colloq. 39: Comets, Asteroids, Meteorites: Interrelations, Evolution and Origins*, 185–195.
- Zellner, B. and Gradie, J. (1976). *AJ* 81, 262–280.
- Zellner, B., Tedesco, E. F., and Tholen, D. J. (1981). Highlights from the Eight-Color Asteroid Survey. In: *Bulletin of the American Astronomical Society*. 13, 717.
- Zellner, B., Tholen, D., and Tedesco, E. (1985). *Icarus* 61 (3), 355–416.
- Zolensky, M., Barrett, R., and Browning, L. (1993). *Geochimica et Cosmochimica Acta* 57 (13), 3123–3148.
- Zolensky, M. E., Bodnar, R. J., Gibson, E. K., Nyquist, L. E., Reese, Y., Shih, C.-Y., and Wiesmann, H. (1999). *Science* 285 (5432), 1377–1379.

ACKNOWLEDGEMENTS

In the beginning there was nothing, which exploded.

– Terry Pratchett

If these are the first lines of the text that you encounter, I inform you that I followed the left-to-right writing system and that the manuscript begins on the other end of the document. If you just finished your read-through (or deliberately skipped ahead), then welcome, take a rest. It's a lot to take in. I assure you that this amount of work within three years is not possible without the support of many people. This section is dedicated to all of them, and some of them are mentioned below.

First and foremost, I want to express my gratitude to Benoit Carry. Your style of supervision is not only *visionary* and *transformational*, but most of all attentive and guiding. I had the freedom to try out my own ideas and your support if they didn't work out. And either way, we'd go for celebratory rounds to *Ketje*. I could not have asked for a better advisor. Although you could know a little bit more about football. And I could know a little bit more about music.

Further invaluable to my PhD research was Pierre-Alexandre Mattei, who first introduced me to the clustering approach and then provided practical support whenever the mathematical side got the best of me. The patience and knowledge with which you replied to my many questions is much appreciated.

A fortuitous misinterpretation of the rules of the *école doctorale* concerning the PhD jury composition resulted in a large number of members who kindly agreed to listen to and judge my work and to give their seal of approval on my independence as researcher. I'm grateful to Cateline Lantz, Guy Li-bourel, Patrick Michel, Philippe Rousselot, Maria Cristina De Sanctis, Cristina A. Thomas, and Pierre Vernazza. I highlight the roles of Maria Cristina De Sanctis and Pierre Vernazza, who served as *rapporteurs* and as such studied the manuscript in detail to compose valuable reports.

On a smaller yet continuous scale, Emanuele Tassi and Aurélien Crida were part of my *comité de suivi de thèse* and provided helpful feedback at different stages throughout my PhD.

I thank the *Observatoire de la Côte d'Azur* for not only hosting this PhD but also for financing 50 %-ish of it. The remaining funding was provided by the *European Space Agency*, and I thank in particular Michael Küppers and Bruno Altieri for their support in securing the partnership and throughout the PhD.

I further thank the administrative and science staff at *Laboratoire Lagrange*, and in particular the TOP team, with whom I had the pleasure to discuss weekly the most diverse topics of planetary science. It is this "big picture"-approach to planetary science which I cherished the most and which

influenced the manuscript in many of its parts, and on even more pages which I had to leave out to not trail too far off the argumentative path.

Peu importe où j'irai à l'avenir, je sais qu'aucun endroit ne répondra aux normes élevées de nourriture et de convivialité que la cantine de l'observatoire s'est fixées. Je remercie *le chef* pour les discussions sur le football, la politique, la vie et pour m'avoir appris quelques subtilités importantes de la langue française. Peu importe où j'irai dans l'avenir, j'ai toujours hâte de revenir chez *Khaled*.

Although we spent many weeks in *télétravail*, I've had the pleasure to share the office with many people over the years. I thank in particular Marjorie Galinier and Alexey Sergeev for fruitful discussions and reviewing figures while I wrote the manuscript. Marjorie ensured the correctness of the French versions of abstract and title, any complaints in this regard may be directed to her.

Camille Granier and Nicolas Müller started this journey at the same time as I did and it did not take much more to make us allies. I thank them for the diversion and support during PhD and pandemic, and for tolerating my tremendous sense of humour for so long. In the following years, the number of PhD student in the PHC grew exponentially and I appreciate the distractions and discussions it provided without giving names as I worry to leave someone out. I'm making an exception for Saburo Howard, who never hesitated to provide a ride up to mountain on rainy days or to the occasional football match in Villefranche.

Throughout my academic career, my family became less and less certain of what area of research I am currently active in. Nevertheless, I could always be certain of their support for my choice to continue the academic path, showing that support and understanding do not follow Heisenberg's principle.

Finally, I appreciate the never-ending support of my work acquaintance and later roommate Julia Maia. I wrote significant fractions of this thesis during my summer "holidays" in Brazil, and I'm looking forward to many more visits, and this time, the thesis is actually finished.

Malheureusement, la nature de l'astéroïde B612 reste un mystère.

COLOPHON

The view expressed in this publication can in no way be taken to reflect the official opinion of the European Space Agency.

This document was typeset using the typographical look-and-feel classicthesis developed by André Miede and Ivo Pletikosić.

Final Version as of October 19, 2022.