DeMeo Taxonomy: Categorization of Asteroids in the Near-Infrared

By

Francesca E. DeMeo

B.S., Earth, Atmospheric and Planetary Sciences (2006) B.S., Physics (2006)

Massachusetts Institute of Technology

Submitted to the Department of Earth, Atmospheric and Planetary Sciences in Partial Fulfillment of the Requirements for the Degree of Master of Science in Earth, Atmospheric, and Planetary Sciences

at the

Massachusetts Institute of Technology

June 2007

C 2007 Massachusetts Institute of Technology. **All** rights reserved.

Table of Contents

 $\mathcal{L}^{\text{max}}_{\text{max}}$

 $\sim 10^{-1}$

 \sim

DeMeo Taxonomy: Categorization of Asteroids in the Near-Infrared By Francesca E. DeMeo

Submitted to the Department of Earth, Atmospheric and Planetary Sciences

June 2007

In Partial Fulfillment of the Requirements for the Degree of Master of Science in Earth, Atmospheric, and Planetary Sciences

Abstract

This work presents the DeMeo taxonomy, an asteroid taxonomy with 24 classes based on Principal Component Analysis of spectral data over the visible and near-infrared wavelengths, specifically the 0.45 to 2.45 micron range. Principal Component Analysis was used by both Tholen (1984) and Bus (1999) to create taxonomies on visible data. There is no pre-existing taxonomic system for the entire suite of asteroid characteristics because only in the current decade has spectral data collection become available in the near-infrared for asteroids down to relatively faint $(V=17)$ limiting magnitudes. (Rayner et al. 2003) With a larger data range, which includes important absorption features at one and two microns suggesting the presence of minerals, there is a need for an extended system to encompass this range of information. In this work we explain the process of creating the taxonomy, the method for finding an object's taxonomic class under this system, and present spectral types for the 365 objects that were used to create the system.

Thesis Supervisor: Richard P. Binzel Title: Professor of Planetary Science

Acknowledgements

I am grateful for the help and advice from many colleagues and students. I thank Bobby Bus for all his guidance and the expertise he provided having created a previous taxonomic system. Thanks to Pierre Vernazza for all the mineralogical background and input he contributed which helped me understand the distinctions between spectra. Thanks to Cristina Thomas for always being available to help me with any problems or questions and for always taking the tough night shift during observing runs. I thank Shaye Storm for being a great UROP student and for helping me reduce and organize the data for this work. I would also like to thank many others who have taught me much and helped me along the way. Thanks Andy Rivkin, Mike Person, Steve Slivan, Jim Elliot, Jack Wisdom, Elisabeth Adams, Mark Avnet, and Susan Kern. **I** thank my family for everything.

I also owe huge thanks to my advisor Professor Rick Binzel who has guided me through not just my Masters thesis, but also my senior thesis. Thank you Rick for all the opportunities you have provided for me, the chance to visit the IRTF on Mauna Kea and Magellan in Chile, to travel to and present at conferences, to research in France and meet colleagues. Thank you for always taking the time to help and answer questions, for going out of your way to make sure you're available to help, for always being patient, and for always being accommodating for the projects **I** undertake and plans **I** make outside of my research. And **by** the way, that is a sweet Earth.

1. Introduction

Classification is important to all branches of sciences. As soon as a large amount of data are available, it is useful to group it into categories with similar characteristics for organization, comparison, and better comprehension.

The first asteroid taxonomy was created **by** Chapman et al. **(1975).** The taxonomy has grown and evolved over time to account for more and better data. Asteroid scientists started **by** characterizing asteroids with a few colors, and then many colors as technology improved. Some taxonomies incorporated albedo, and eventually they used spectra. Taxonomies were created **by** Tholen (1984) and most recently **by** Bus **(1999). All** current asteroid taxonomies are based on the visible wavelength range approximately 0.4 to **0.9** microns because data were most available over those wavelengths. Recently, with the creation of the SpeX instrument on the IRTF a growing library has been created of near-infrared spectral data. Gaffey et al. **(1993)** created an S-complex taxonomy based on near-infrared data, but there is no pre-existing taxonomic system for the entire suite of asteroid characteristics. From Figures 1 and 2 we can see examples of the need for a new taxonomy because of the divergence in the near-infrared of a single well-constrained visible class. **A** classification system that extends to the near-infrared is also important because significant mineral absorption features are present at one and two microns suggesting presence of pyroxene and olivine. Even though the taxonomy presented in this work is not connected to mineralogy it is hoped that future work would allow mineralogic interpretation of taxonomic classes.

Figure **1:** Bus Taxonomy A-Types. All spectra behave similarly in the visible and are designated A-Types in the Bus system, but have very different behaviors in the nearinfrared.

Figure 2: Bus Taxonomy L-Types. This plot shows L-types under the Bus classification system. Notice how well they match in the visible range, but diverge in the near-infrared. Some have stronger 1 micron absorptions, some have high slopes or even negative slopes after one micron.

This work presents the DeMeo taxonomy, an asteroid taxonomy based on Principal Component Analysis of spectral data over the visible and near-infrared wavelengths, specifically the 0.45 to 2.45 micron range. The aim of this new taxonomy is to have classification parameters for the entire wavelength range, but not to create an entirely new system. Principal Component Analysis (PCA) is a technique for reducing the dimensionality of a dataset to contain most of the information in a few principal components. It is a set of linear transformations changing the coordinate system according to the greatest variance. The greatest variance is along the dimension described by Principal Component 1 and the second greatest variance is Principal Component 2. PCA is used widely in other applications such as data compression, but was not applied to the field of asteroid taxonomy until Tholen (1984).

When creating this taxonomy there were seven possible outcomes for changes to each taxonomic class under the Bus system. Two classes could be indistinguishable over the entire wavelength range and merge into one class. One class could be degenerate and branch off into two (or more) separate classes. A class may need to be created if spectra exist that do not fit within the past framework. We found that one new class needed to be created, the Sv class, comprised of two objects that represented an intermediate class between S and V. The name of a class may need to be changed if the current notation seems misleading. Some classes may need to be redefined. For example the names Sr and Sa were conserved but the objects in those classes were largely redefined. There is the possibility that one class may correlate directly as near-infrared data are included and no change in that class occurs. Finally, it is possible that a class may remain the same, but a notation is added. In this taxonomy a "w" notation was used to indicate high slope and is explained in Section 4. Figure 3 is a diagram of the possible outcomes of extending the classes. See Appendix A for the progression from the Bus to DeMeo taxonomy.

Figure 3: Possible outcomes of extending taxonomic classification to the near-infrared. 1. Two classes from the Bus system could merge to one class. 2. One class could diverge into two distinct classes. 3. A class that did not previously exist may be created. 4. A previously existing class may be relabeled with a different name. *5.* A class may keep its label but its definition may change. 6. One class could correlate directly from the Bus to DeMeo taxonomies. 7. An additional notation to may be added within a class.

2. Observations and Data Reduction

Near-infrared spectral measurements from 0.8 to 2.5 microns were obtained using the SpeX instrument on the 3 meter NASA Infrared Telescope Facility (IRTF) located on Mauna Kea, Hawaii. As described in DeMeo and Binzel (2007), when possible, objects and standard stars were observed near the meridian to minimize their differences in airmass. Frames were taken so that the object was alternated between two different positions (usually noted as the 'A' and 'B' positions) on a 0.8 x 15 arcsecond slit aligned north-south. Solar-type standard stars were observed to divide the solar spectrum from

the asteroid spectra, with the result being reflectance spectra. Our primary solar analog standard stars were 16 Cyg B and Hyades 64. Additional solar analog stars were utilized around the sky, having comparable spectral characteristics. Two to three sets of eight images per set were taken for each object, with individual image exposures typically being 120 seconds. The total integration time for each of these objects therefore ranged from 30 to 120 minutes.

Reduction was performed using a combination of routines within the Image Reduction and Analysis Facility (IRAF), provided by the National Optical Astronomy Observatories (NOAO) (Tody 1993), and Interactive Data Language (IDL). We use a software tool called "autospex" to streamline reduction procedures outlined by S. J. Bus (personal communication). Autospex writes macros containing a set of IRAF (or IDL) command files that are then executed by the image processing package. Autospex procedures operate on a single night at a time, with the opportunity for the user to inspect and verify the results at each stage. Briefly, autospex writes macros that: trim the images down to their useful area, create a 'bad pixel map' from flat field images, flat field correct all images, perform the sky subtraction between AB image pairs, register the spectra in both the wavelength and spatial dimensions, co-add the spectra for individual objects, extract the 2-D spectra from co-added images, and then apply the final wavelength calibration. Using IDL, an absorption coefficient is determined for each object and star pair that best minimizes atmospheric water absorption effects for that pair. This coefficient correction is most important near 1.4 and 2.0 microns, locations of major absorption bands due to telluric H_2O . Because we operate in survey mode with our observing time heavily weighted toward new objects rather than on standard stars, our

telluric corrections are not perfect. Users of these data therefore must exercise caution in basing significant scientific conclusions on any unusual spectral features at these telluric wavelengths, which may be the result of imperfect correction. The last IDL step averages all the object and standard pairs to create the final reduced spectrum for each object.

All visible wavelength spectra (usually 0.4 to 0.9 microns) were taken from the Small Main Belt Asteroid Spectroscopic Survey (SMASS II) data set (Bus & Binzel 2002). Our sample was comprised of 365 objects with both visible and near-IR data. For a table of observations see Appendix B.

3. Background on Classification

3.1 Thoughts on Classification

Classification systems tend to emerge as soon as there are enough data to form groups. It is a tool used to organize information making it easier to compare, find holes or missing information, and makes it easier to communicate about data features.

Data on asteroids come from three sources: spacecraft missions, meteorites, and telescope observations. Spacecraft missions provide the most detailed information about asteroids, however, they only reach a limited number of targets. Meteorites provide a great opportunity for detailed study in a laboratory, but their links to specific asteroids and even general asteroid classes are nonexistent or weak. The only known connection between a meteorite and its parent asteroid are the HED meteorites to Vesta (Consolmagno & Drake 1977, Binzel & Xu 1993) Telescope observations provide data including brightness, albedo, and spectral reflectance on a large number of asteroids making it the best basis for classification of many objects.

The purpose of this classification system is to characterize the features present in the spectrum. These features may represent mineralogy, grain size, weathering, or other properties of or effects on an asteroid, however the taxonomy is not meant to suggest or confirm anything more than the features visible in the spectrum.

3.2 Previous Classification Systems

Asteroid taxonomy has evolved significantly over the past few decades, as more and higher quality data have become available. Starting in the 1950's UBV photometry observations enabled the separation of asteroids into two groups noted by Wood and Kuiper (1963) and Chapman et al. (1971). Using spectrophotometry and albedo measurements Chapman et al. (1975) created the first letter-based asteroid taxonomy. "C" represented carbonaceous, "S" was for stony, and "U" for unusual objects. Tholen Taxonomy was the most widely used when created with eight classes of spectral types, plus classes for three unusual objects. The Barucci Taxonomy, meant to expand on the Tholen taxonomy was based on G-mode analysis. (Tholen & Barucci 1989, Barucci et al. 1987). The most recent taxonomy was created by Bus using 1189 objects. It is feature-based taxonomy also using PCA (Bus 1999, Bus & Binzel 2002). The Bus system further separated the C, S, and X complexes.

Only in the current decade has spectral data collection become available in the near-infrared for asteroids down to relatively faint (V= **17)** limiting magnitudes. (Rayner et al. 2003) With a larger data range, which includes important absorption features, there is a need for an extended system to encompass this range of information. We chose not to invent an entirely new system, but instead strove to stay as consistent as possible with past visible taxonomies and notation.

3.3 Process To New Classification

3.3.1 PCA Preparation

Bus chose three prototype asteroids for each of his taxonomic classes that best exemplify the characteristics of the class. The process of creating a new taxonomy began with the evaluation of the Bus prototypes over the entire wavelength range. Some of these prototypes had to be rearranged and redesignated. Asteroid 150, for example, was a Cb under the Bus system, but looks much more like a C over the entire range. Asteroid 24 was aB and is now a C. 42 was an L and is now a K. 606 was a K and was redesignated to L. 1904 and 5111 were R-types and are now V. Even though there were many changes, most were within subclasses or closely related classes. This exercise was mainly to observe how each class behaved in the near-infrared and which classes were more or less well-defined. Because the new system has many more features present in the near-infrared than the visible it is natural that some objects that looked similar in the visible diverge in the near-infrared.

Next, to become more familiar with the data, we assigned a taxonomic type or complex by visually assessing spectral features and plotting it against prototypes. This was useful because when plotting objects in PCA space we could compare how objects grouped when labeled under the Bus system with what we believed they looked like in the visible and near-infrared range.

13

We then plotted all near-infrared data in a "splinefit" program which creates a list of data points that follow the features of the spectrum. This step smoothes the spectra, creating a good "best fit" curve to the data. This reduces the risk of noise or missing data points influencing the **PCA.** We sampled the region 0.45 to 2.45 microns and recorded points at increments of **0.05** resulting in 41 channels.

The "splined" data were then normalized at **0.55** microns and the slope was removed from the data and recorded. With our normalized data we simply took a linear regression line of each set of data. Because this is the most prominent feature of the spectra we can remove it from the data before performing principal component analysis thus making **PCA** more sensitive to other features, and therefore more effective. We divide all the data points **by** the average slope. The remaining data are a horizontal spectrum with residuals (including absorption features) above and below the horizontal. We now have data with slope removed having 41 channels normalized at **0.55** microns. Because each spectrum has a value of 1 at **0.55,** that channel provides no new information and so was removed from the data set to make **PCA** more effective. We therefore input 40 channels per object into **PCA**

To verify that we could remove the slope before running **PCA,** we ran **PCA** on the data without the slope removed. The first principal component **(PC1)** of **PCA** before removing slope was compared to the slope we remove directly from the splined data. In Figure 4, a plot of **PC1** versus slope, it is clear that the two are linearly correlated, thus **PC1** corresponds to slope and we are justified in removing it first.

Figure 4: Plot of PCi versus Slope. PCI is the first principal component from the data set that did not have the slope removed. It is clear from the plot that **PCI** and slope are linearly related and it is thus safe to remove the slope before running PCA to make PCA more sensitive to other features.

Using MATLAB, we performed Principal Component Analysis (PCA) on the splined files with slopes removed. PCA is a method of reducing the dimensionality of a data set. PCA involves coordinate transformations to minimize the variance. The first transformation rotates the data to maximize variance along the first axis, known as Principal Component 1' (PC1'), the second axis is the second Principal Component (PC2'). The first few principal components contain the majority of the information. For a more thorough explanation of PCA and why it is useful for asteroid taxonomy refer to Tholen (1984) and Bus (1999).

3.3.2 What Principal Components Tell

By plotting the principal components against one another we can see how different groups separate. PCA was used to create the Tholen and Bus taxonomies. We chose to use covariance instead of correlation for the PCA. Bus (1999) suggested the use of the covariance matrix of unscaled variances because they may be useful to distinguish spectra. Principal Component Analysis tells information about specific parts of the spectrum. It is difficult to determine what exact feature each principal component refers to because so many features exist over the visible to near-infrared wavelength. However, by looking at the eigenvectors we can determine over which wavelength range each principal component is most sensitive. Figure 5 is a plot of sensitivity of the eigenvectors throughout the range. This figure shows that PC4' and PC5' are more sensitive to features at the ends of the spectrum, while the others are more sensitive to features in the middle. See Appendix C for the eigenvectors. Wavelengths with higher magnitudes dominate that principal component using positive or negative weights to separate in opposite directions within that principal component space.

Figure 5: Sensitivity of Principal Components over wavelength range. High positive or negative values indicate that a principal component is more sensitive to information over that region. This figure shows that PC4' and PC5' are more sensitive to features at the ends of the spectrum, while the others are more sensitive to features in the middle.

3.3.3 Choosing the right number of Principal Components

Bus used slope plus three principal components to characterize his visible data. Because data in the extended wavelength range have many more features we use slope plus 5 principal components. We look at the amount of variance contained within each principal component to determine how many to use. Greater variance means more information is contained within that principal component.

Because 40 channels were put into PCA, 40 principal components were the output. Since PCA puts the most information in the first dimensions, and decreases with

each successive dimension, we only want to look at some of the principal components. To decide how many principal components we want to use in our analysis we must look at the variance contained within each. Table 1 shows percentage variance contained within each principal component. Note that slope is not included in this calculation, so this percentage represents the variance left over after the slope was removed. PC10' through PC40' contribute essentially no new information so are therefore not included in the table. Upon inspection we see that most of the variance exists in the first six principal components. Figure 6 shows a scree plot of the variance. Scree is rubble rock at the bottom of a mountain. On scree plot the point where the plot levels off after the drop determines at which principal component to stop. (Jackson 1993) We see the chart level off around PC5' and PC6'. We decided that since 99.2% of the variance was contained within the first 5 principal components, they were sufficient. It is evident that PC5' does not seem to contain a significant amount of information, however, we found it useful for classifying C-complex objects. By running PCA with the slope included in the data we found that slope accounts for 88.4% of all the variance within the data. All the other principal components combined account for 11.6%. Slope plus the first five PC' account for 99.9% of the variance. Table 2 shows the variances accounted for by the slope plus the first five principal components.

Principal	Variance
Component	(%)
PC1'	63.1
PC ₂ '	24.3
PC3'	8.9
PC4'	2.2
PC5'	0.6
PC6'	0.3
PC7'	0.2
PC8'	$0.1\,$
PC9'	$0.1\,$
PC10'-PC40'	0.1
Total:	100.0

Table 1 Variance Contained within Each Principal Component (not including slope) We find **99.2%** of the variance is accounted for **by** using **PC1'** through **PC5'**

Table 2 Variance Contained within Slope and Principal Components

PC	Variance (%)
Slope	88.4
PC1'	7.3
PC ₂ '	2.8
PC3'	1.0
PC4'	0.3
PC5'	0.1
Total:	99.9

Figure 6: Scree Plot. This plot shows the amount of variance contained in each principal component. The further down in principal component space, the less information is contained in each principal component. PCI' through PC5' contain most of the variance in the data set. Notice that PC5' lies just at the bottom of the downward slope indicating it is a good place to stop using components. (Jackson 1993)

3.3.4 **A** Note **on Notation: What is the difference between PC1' and PC1?**

If the original, normalized and splined data without the slope removed were analyzed with PCA the first transformation would be PC1. PC1 corresponds to slope. However, we remove the slope and re-perform PCA. In this case, we denote the slopefree PCA results as PC1', PC2', etc. We use the "'" notation so as to recall that PC1' is closely analogous to PC2 in the original (no slope removed) data set.

4. The Taxonomy

The most striking feature plotted from the principal components is the large gap in PCI' versus PC2' space seen in Figures 7 and 8. It appears that this clear boundary distinguishes between spectra that have a two micron absorption band and those that don't. This gap is well represented by the line:

$$
PC1' = -3.00PC2' - 0.28 \qquad (line \alpha)
$$
 (1)

All objects below and to the left of this line have no two micron absorption features and include all featureless (C, X complex) spectra. Objects plotted to the right of and above the line have two micron absorption features, and farther right signifies stronger absorption. Moving right in the orthogonal direction to line α also seems to signify narrowing 1 micron bands. For example, V-types plot furthest to the right. The only classes that cross this gap are the A and Sa classes. Interestingly the K-class, long considered as an intermediate between S and C, falls most squarely in the gap. Figures 7 and 8 plot all objects in PC1' and PC2' space. Figure 7 shows the objects plotted with original Bus labels, while Figure 8 shows the new DeMeo labels. An enlargement of the gap present at line α is shown in Figure 9.

Figure **7:** Plot of **PC1'** versus PC2' for all objects. These objects are plotted with their original labels under the Bus system. Objects that do not have Bus designations are labeled as a ".".

Figure 8: Plot of PC2' versus PCI'. All objects plotted are labeled with their DeMeo taxonomy classification. Notice the large gap between the S-complex and the C- and Xcomplexes. Line α separates objects with and without 2 μ m absorption bands. The direction orthogonal to line α indicates deeper 2 μ m and narrower 1 μ m absorption bands.

Figure **9:** Enlargement of gap in PCI' and PC2' space. This figure shows the clear distinction between the S-Complex and end members with $2 \mu m$ absorption bands and objects that are featureless or have only $1 \mu m$ absorption bands.

In defining classes for the taxonomy we created a flowchart (Appendix D) containing steps to find the right combination of principal components ranges that result in the correct class for each object. The order of the flow chart is significant because some classes overlap in certain principal components but can be separated in others. For all the new labels to the 365 objects in the dataset see Appendix E. We start by separating the A and Sa classes because they cross over the large gap in PC1' and PC2' space. The DeMeo Sa class was relabeled from the Bus system because the two Sa objects in this system were both Sr-types in the Bus system. Since these objects prove to be intermediate between S and A change the name of these two Sr-types. This is step 1 in the flow chart. Refer to Appendix C for the complete chart. Figure **10** shows the progression from S to Sa to A type.

S, Sa, A Types

Figure 10: Plot of S, Sa, and A. There is a clear progression from S with a shallow one micron band and low slope to A with a deep one micron band and high slope. Sr seem to to resemble A-type one micron band absorption very well, but are much less red than Atypes.

Step two starts by separating all objects by the divide (line α) in PC1' versus PC2' space. Step two continues to create boundaries for objects with a two micron band and step three addresses featureless objects (the C- and X-complexes) as well as K and L classes which have no significant two micron band.

We started by looking at the end member classes in PCA space since they separate more clearly making them the easiest to define. In Figure 10 you can see the data of Figure 8 with lines separating S-complex and end member classes. The V class lies to the right of the line parallel to line α noted as equation 2:

$$
PC1' = -3.0PC2' + 1.5
$$
 (line δ) (2)

The R class lies to the left of line **8** but right of equation 3 and above equation 4:

$$
PC1' = -3.0PC2' + 1.0 \t (line \gamma)
$$
 (3)

$$
PC1' = 1/3PC2' - 0.4 \t (line \eta) \t (4)
$$

To see the lines labeled in PCA space refer to Figure 11. Note that there is only one Rtype object in our sample, 349 Demboska, for which the class was created. Also, there are only two O-type objects among our data. Even though the class is separated in the flow chart, more data on R-type and O-type objects may change the region boundaries significantly. To see the differences between the end member classes V, R, Q, and O see Figure 12.

Figure 11: Plot of PC2' vs. PC1'. This plot shows the lines that divide the objects with and without 2 micron absorption bands, as well as the V and R class boundaries.

Figure 12: Plot of V, O, Q, R Types. This plot shows typical spectra in each end member class. Note O has a very wide 1 micron band and V has a very narrow band. The V-types with the deepest 2 micron bands plot farthest from line α .

The S-complex was by far the most difficult to define. Most classes seemed to blend together or scatter randomly in all combinations of PCA components. Sa and **Si** (named under the Bus taxonomy) were entirely indistinguishable so the S1 class was combined. Sa objects were most easily distinguishable not by absorption features, but by their greater slope (caused by slope increases in the 1 to 1.5 micron range.) The S, Sq, and Sk objects were initially impossible to define clearly because the boundaries seemed to blur and overlap. We performed a new principal component analysis to help guide us.

Because the main difference between the S-classes appears to be the width of the 1 micron absorption band we used the wavelength range 0.8 to 1.35 microns and did PCA on only S-type objects. Sa types were scattered randomly supporting that their main difference is slope. From this new PCA (sPCA) we found that sPC1 and sPC2 best separated the classes. We relabeled objects based on sPCA and plotted the newly changed labels back in the original PCA.

Once we used sPCA to guide our labels, we could use PC1' and PC2' to create definitions within our original principal components. We continued to use boundaries parallel and perpendicular to line α . Each class has its own box in PC1' and PC2' space. Refer to Figure 13 to see the S-complex boxes labeled in PCA space. The S class was right of line α (equation 1), to the left of equation 5 and above equation 6.

$$
PC1' = -3.0PC2' + 0.35 \qquad \text{(line } \beta\text{)} \tag{5}
$$

$$
PC1' = 1/3PC2' - 0.10
$$
 (line ζ) (6)

Objects that reside below line **ý** (equation 6) appear similar to Q-types, but with more shallow absorption bands. These are Sq-types transitioning between S and Q. They lie right of line α , left of line β , below line ζ , which is the S-type boundary, but above line η , which is the boundary for Q-types. Sr-types transition between S and R classes. They reside between lines β and γ , and between lines η and ϵ given in equation 7.

$$
PC1' = 1/3PC2' + 0.55 \qquad \text{(line } \varepsilon) \tag{7}
$$

The objects between β and γ , but above line ε were unique from Sr because they exhibited very narrow 1 μ m absorption bands. These two objects appear to transition between S and V classes. They are not included in the Bus dataset, and Bus (1999) did not report any cases of these objects. They are given the label Sv. The Sk type was nonunique and was absorbed into the other S classes

Figure 13: PC2' versus PC1' plotted for the S-complex. Each class has its unique box bounding it. All lines are perpendicular or parallel to line α .

Sa and **Sl** (labeled under the Bus system) objects were indistinguishable when extending to the near-infrared. Their features were also indistinguishable from other Scomplex objects, and it seems that their slope was the major distinction. Because the Sa and **Sl** designations did not indicate that these objects were intermediate classes between

S and A or L we felt that the notation was misleading. We removed the Sa and S1 designations given by Bus and replaced them by designations of S, Sq, Sr, and Sv based on PC1' and PC2' which represented absorption features. We felt, however, that even though slope did not merit a class distinction, it was worth noting in the taxonomy.

The S, Sq, and Sr classes all had very widely varying slopes. High slopes are indicative reddening by space weathering (Clark et al. 2002). Even though slope has no mineralogical significance we thought it fitting to distinguish between objects that had undergone significant space weathering from those who had not. We made a cutoff, albeit an arbitrary one, at Slope = *0.25* dividing "high slope" objects from other objects. These objects are not relabeled in a class of their own. Instead the S, Sk, and Sq objects with high slopes get a notation of "w" added to their name to indicate a "weathered" object. The high slope S objects are labeled Sw, Skw, and Sqw. There were also two Vtype objects and two Q-types with slopes greater than 0.25, which were labeled Vw and Qw. To see the difference between a low and high slope object, S and Sw, refer to Figure 14.

Figure 14: Plot of S and Sw. The absorption features for both are very similar. Slope is the most significant distinction between the two.

The choice of 0.25 for the slope cutoff was somewhat arbitrary. When plotting Bus labeled S, Sa, and S1 objects, there was a mixing around the 0.23 to 0.27 slope range. We wanted to keep the "w" notation more selective, but also didn't want to set the cutoff too high where objects with unusual slope features (such as deeper UV dropoffs) were preferentially selected rather than focusing on the significant slope range between one and two microns for the S-Complex. See Figure 15 for the plot of Slope versus PC1' that shows the line separating weathered objects from regular objects.

Figure 15: Plot of Slope versus PCI' for the S-Complex. All objects with slopes greater than 0.25, regardless of their subclass (S, Sq, Sk), get a "w" notation indicating weathering. Note the two V and Q class objects labeled "Vw" and "Qw." Dembowska, as the only R member, does not get a w.

Step three focused on objects below or to the left of line α (equation 1). We again start by removing end members. D and T classes are easily separated by their high slopes and PCI', PC2', and PC3' values. Next, we separate out L objects based on the PC2' versus PC1' plot. When an object fits in the L component space it is necessary to check for Xe type objects. Xe is a class defined in the Bus system that has a distinct hook at 0.49 microns. By visually inspecting the spectrum, one can notice the presence of a

feature at 0.49 and an absence of a slight feature around 1 micron and label the object an Xe instead of an L. Refer to Figure 16 to see Slope versus PC1' which shows how D, L, and L are fairly distinct. The K types can then be distinguished clearly in PC2' and PC3' space. We found that all Ld type objects under the Bus system could be fit in either the L or D classes when near-infrared data were added. Since it is not necessary for distinguishing classes, it was removed from the DeMeo system. Figure 17 shows how the K class is distinct.

Figure 16: Plot of Slope v. PC1'. Note how D, T, and L are fairly distinct in this component space. K separates better in PC2' and PC3' space.

Figure 17: Plot of PC3' v. PC2'. The K types are distinct in this component space.

This brings us to the X and C complexes. The B-types are easily distinguished by their negative slope as well as negative PC1' and PC4' values. The X-class can then be easily removed based on high slope values between 0.2 and 0.425. At this point Xk objects may be in X-class PC' space. When visually inspecting the spectrum if there is a feature present at 1 micron, the object is designated an Xk. Figure 18 shows a plot of the C- and X- complexes.

Figure 18: Plot of PC2' v. PCI' for the C- and X-Complex. These components help separate the two complexes. Slope will also differentiate the two.

The Ch objects are defined well in PC1' and PC4' space, but a check must be done to see if a strong 0.7 micron feature exists, making the object a Cgh. The C and Cb types separate in PCi', PC4', and PC5'. See Figure 19 and 20 to see the C-complex plotted in PCA space. After this the Cgh, Ch, Xk, Xc, and Xe types don't all separate cleanly in component space and are heavily dependent on visually detecting features.

Figure **19:** Plot of PC4' v. **PC5'** for C-Complex. The classes are easily distinguishable in this component space.

Figure 20: Plot of PC5' v. PC1'. A second plot demonstrating the separation among Ctypes.

We currently do not have a method for classifying spectra using only near-infrared data because putting reflectance values into the eigenvectors from this work would be meaningless since the entire range doesn't exist. In future work we hope to include a method for using near-infrared only data to find a taxonomic classification. However note: certain classes evolve unchanged from the Bus taxonomy that are based on features measured at visible wavelengths. Assignment to these classes (Cg, Cgh, Xc, Xe, Xk) therefore requires visible wavelength data.

5. Conclusion

A new taxonomy with was created using Principal Component Analysis and visible features to characterize visible and near-infrared wavelength spectral reflectance values. The system, based on the Bus visible taxonomy from Bus (1999), has 24 classes compared to 26 in the Bus system. The changes in classes are summarized in Appendix A. We used all of the seven possible changes from the previous system to the new highlighted in Figure 3. We eliminated three classes: Ld, S1, and Sk. All the S subclasses (Sa, S1, Sk, Sq, Sr) had objects that merged back into the S class, although many Sq objects remained Sq and two Sr objects were relabeled Sa. A new intermediate class, the Sv class, was created to branch the S and V classes. High-sloped S, Sa, Sq, Sr, V and Q objects were given a "w" notation to indicate weathering. Many of the classes that lie left of line α are either featureless or exhibit only small features at visible wavelengths identified by Bus (1999). It is still necessary to use these visible features to distinguish the classes because there are no other distinguishing features at near-infrared wavelengths. 365 objects were given types based on the DeMeo taxonomic system which was created using 6 dimensions including Slope and PCI' through PC5' of Principal Component Analysis.

References

- Barucci, M.A, Capria, M.T., Coradini, A., Fulchignoni, M. (1987). Classification of asteroids using G-mode analysis. Icarus 72, 304-324.
- Binzel, R.P., Xu, S. (1993). Chips off of asteroid 4 Vesta: Evidence for the parent body of basaltic achondrite meteorites. Science 260, 186-191.
- Binzel, R.P., Rivkin, A.S., Stuart, J.S., Harris, A.W., Bus, S.J., Burbine, T.H., (2004). Observed spectral properties of near-Earth objects: results for population distribution, source regions, and space weathering processes. Icarus 170, 259-294.
- Bowell, E., Chapman, C.R., Gradie, J.C., Morrison, D., Zellner, B. (1978). Taxonomy of Asteroids. Icarus 35, 313-335.
- Bus, S.J., (1999). *Compositional structure in the asteroid belt: results of a spectroscopic survey.* Doctoral thesis. Massachusetts Institute of Technology, Cambridge, Massachusetts.
- Bus, S.J., Binzel, R.P., (2002). Phase II of the Small Main-Belt Asteroid Spectroscopic Survey. Icarus 152, 106-145.
- Chapman, C.R., Johnson, T.V. (1971). A review of spectrophotometric study of asteroids. Physical Studies of Minor Planets (T. Gehrels, Ed.), 51-65. NASA SP-267.
- Chapman, C.R., Morrison, D., Zellner, B., (1975). Surface properties of asteroids: A synthesis of polarimetry, radiometry, and spectrophotometry. Icarus **25,** 104-130.
- Clark, B.E., Hapke, B., Pieters, C., Britt, D. (2002) Asteroid Space Weathering and Regolith Evolution. In *Asteroids III,* (W.F. Bottke, A. Cellino, P. Paolicchi, R.P. Binzel eds.), Univ. Arizona Press, 585-599.
- Consulmagno, G.J., Drake, M.J. (1977). Composition and evolution of the eucrite parent body Evidence from rare earth elements. Geochimica et cosmochimica Acta. Vol 41 Isssue 9, 1271-1282
- DeMeo, F.E., Binzel R.P (2007). Comets in the near-Earth Object Population. Icarus. Submitted for publication.
- Gaffey, M.J., Bell, J.F., Brown, J.R., Burbine, T.H, Platek, J.L., Reed, K.L., Chaky, D.A., (1993). Mineralogic Variations within the S-Type Asteroid Class. Icarus 106, 573-602.
- Jackson, D.A. (1993) Stopping rules in Principal Component Analysis: A comparison of heuristical and statistical approaches. Ecology 74 2202-2214.
- Rayner, J.T., Toomey, D.W., Onaka, P.M., Denault, A.J., Stahlberger, W.E., Vacca, W.E., Cushing, M.C., Wang, S. (2003) SpeX: A medium-resolution 0.8-5.5 micron spectrograph and imager for the NASA Infrared Telescope Facility. Astron. Soc. of the Pacific 115 362-382.
- Tholen, D.J. (1984) *Asteroid taxonomy from cluster analysis of photometry*. Doctoral Thesis, University of Arizona.
- Tholen, D.J., Barucci, M.A. (1989) Asteroid taxonomy. In *Asteroids II* (R.P.Binzel, T.Gehrels, and M.S. Matthews, Eds.), 298-315. Univ. of Arizona Press.
- Tody, D. (1993). IRAF in the Nineties. In Astronomical *Data Analysis Software and Systems II* (R. J. V. Brissenden, and J. Barnes, Eds.), pp 173. Astron. Soc. of the Pacific, San Francisco.
- Wood, J.H. Kuiper, G.P (1963) Photometric studies of asteroids. Astrophys. J. 137, 1279-1285.
- Vernazza, P., Birlan, M., Rossi, A., Dotto, E., Nesvorny, D., Brunetto, R., Fornasier, S., Fulchignoni, M., Renner, S. (2006) Physical characterization of the Karin Family. AA 460, 945-951

Appendix A: Bus-DeMeo class changes

w notation does not denote a distinct class.

Appendix B: Observation List

*Data from Vernazza et al. **(2006)**

 $\hat{\boldsymbol{\epsilon}}$

Appendix C: Principal Component Eigenvectors

 \sim

Appendix D: Flow chart. It's as easy as 1, 2, 3...

Checks* for Cg, Cgh, Xc, Xe, Xk Cg: Strong **UV** absorption feature before **0.55 atm Cgh: A Cg with a broad, shallow absorption feature at 0.7** μ **m.** Xc: Red and featureless with slight concave-down curvature. Xe: Concave-up absorption feature before 0.55 μ m. Xk: Red with a flat section around $0.75 \mu m$.

*These spectral features all exist in the visible and were identified in the Bus system (Bus **1999)**

Equations

m

Appendix E: New Labels to Data

