Automatic Classification of Interplanetary Dust Particles

Jeremie Lasue[1,2], Tomasz F. Stepinski[1] and Samuel W. Bell[1,3], [1]Lunar and Planetary Institute, 3600 Bay Area Blvd., Houston TX 77058 USA (lasue@lpi.usra.edu) [2]Los Alamos National Laboratory ISR-1, Mail Stop D-466, Los Alamos, NM 87545 USA, and [3]Amherst College, Amherst, MA 01002 USA

We present an automatic classification of the stratospheric particles collected by NASA-JSC based on their X-ray energy-dispersive spectrometry (EDS) spectra. Agglomerative clustering is applied to the 467 particles constituting Volume 15 of the Cosmic Dust Catalog [1]. The Sammon's map algorithm is used to visualize relationships between the clusters; six clusters correspond to cosmic particles and ten clusters correspond to terrestrial contaminants. One-third of the particles have been relabeled with this new technique, and we point out 31 terrestrial contaminants particles that may be of cosmic origin.

Background

Mineralogical properties of comet 81P/Wild 2 samples returned by the Stardust mission show greater variations than previously thought suggesting the possibility that some particles, classified in the catalog as terrestrial contaminants, may be instead of cosmic origin. In particular, the origin of samples having peculiar compositions, including Na, Cr and K-rich particles [2, 3], needs to be reevaluated.

Methodology

The curation team at the JSC devised a preliminary particle origin classification system into four main origin types: C (cosmic), TCN (natural terrestrial contaminant), TCA (artificial terrestrial contaminant) and AOS (aluminum oxide sphere). Ambiguous classification is accompanied by a question mark [1].

In this work, our classification is based on the particles EDS spectra represented by multi-dimensional vectors. We use two complementary techniques to classify the particles. First, the agglomerative clustering is a bottom-up method of grouping multivariate data (N-dimensional vectors) into clusters. We use Euclidean distance between 1522-dimensional spectra vectors as a measure of similarity between spectra of any two particles. We complement the clustering by a data visualization technique to map the N-dimensional vector space into a 2-dimensional plane in a way that best preserves the inherent structure of the high-dimensional data. We used the Sammon's map technique [4], an iterative nonlinear map that transforms N-dimensional vector space into two dimensions in such a way as to minimize a cost function that reflects the relative differences between distances of items in high-dimensional space and distances between corresponding points in the 2-dimensional plane.

Results and Conclusions

Fig. 1 presents the Sammon's map and agglomerative clustering for all 467 spectra defined as vectors in 1522-dimensions. We determined that a set of 16 clusters best reflects the structure of the data as seen on the map. These clusters are identified by their color on the map. We assessed the purity of our classification with respect to the JSC preliminary classification represented in Fig. 1 by the shape of the data points (C=▲, TCN=×, TCA=+ and AOS=■). The map not only shows a division of particles into spectra-defined clusters, but also similarity of the clusters as expressed by the magnitudes of their spatial separations. The clusters grouping chondritic particles (clusters 1 and 2) are located in the middle of the map. The clusters containing Si-rich particles (3 to 5) and Mg-Si-righ particles (6 and 7) are located to the north-east of chondritic clusters 1 and 2. Clusters 8 to 12, containing the bulk of TCN and TCA particles are located to the west, south-west and north-west of the chondritic clusters. Clusters 13 and 14, grouping the Fe-Ni-S-rich and Fe-Ni-rich particles are located to the south of the chondritic clusters. Finally, clusters 15 and 16, representing AOS particles are located in the upper left corner of the map, to the far north-west of the chondritic clusters.

This coherent structure of different clusters suggests spatial separation of particles of different origins on the Sammon's map. Indeed, particles having different JSC labels are found in different regions of the map. The map reveals a layered structure with the layers predominantly extending in the south-west to north-east direction. The bottom layer corresponds to the cosmic particles, the subsequent layers corresponding to TCN, TCA, and AOS particles, respectively. Interestingly, a rather significant gap separates the cosmic particles group from the layers of terrestrial particles corresponding to the fact that all cosmic particles as a group are spectrally different than all terrestrial particles as a group.

Thus, our results quantifies the fact that particles of different origin have different compositions expressed by different spectra. We used this finding to re-label 155

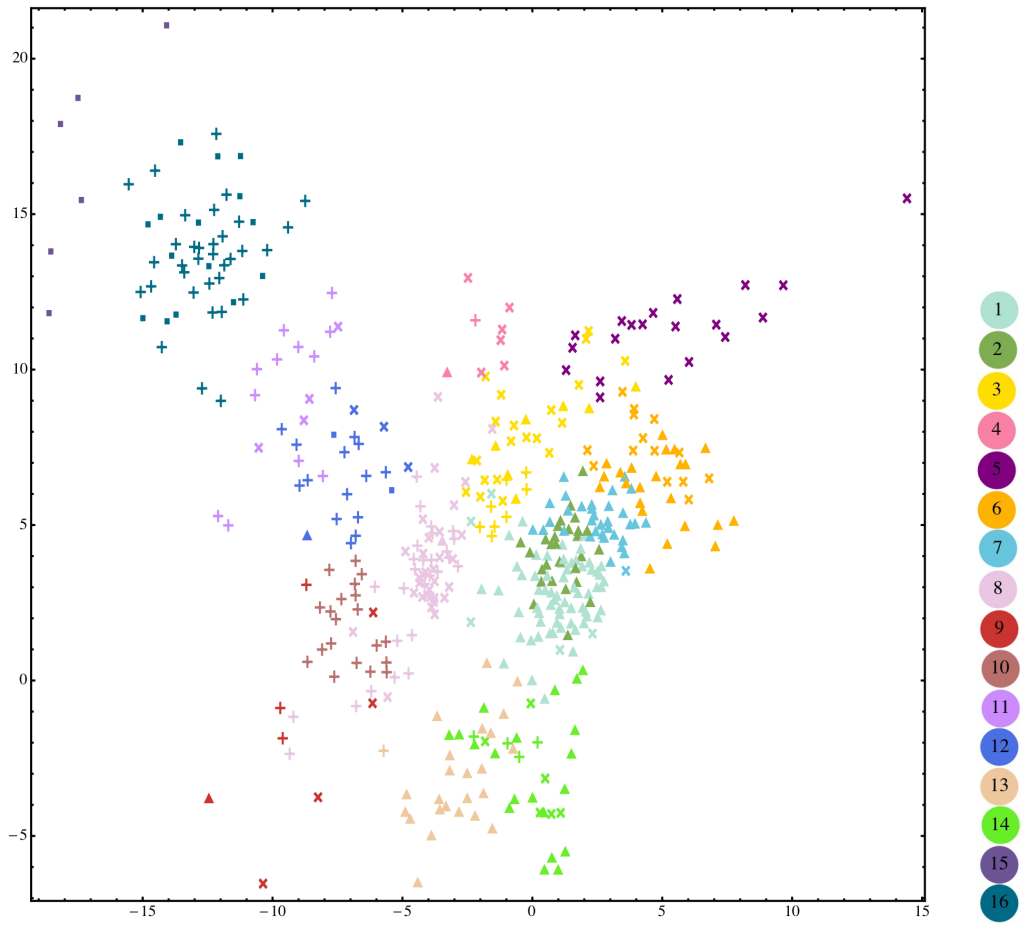

Figure 1: Sammon's map representation of the particles classifications. The color corresponds to the automated clustering of the particles. The JSC preliminary classification is represented by the shape of the data points: C=▲, TCN=×, TCA=+ and AOS=■.

particles in the Volume 15 of the catalog thus updating the catalog. Most importantly, 31 particles, previously cataloged as terrestrial contaminants, are labeled as cosmic by our classification. These particles should be further examined in order to establish their true origin.

Finally, it is possible to extend our approach to non-numerical data. Future research will incorporate, in addition to the spectra, petrographic, geometric, and texture properties of particles into their classification.

Acknowledgments

The authors are grateful to Mike Zolensky, Keiko Nakamura-Messenger, and Allan Treiman for fruitful discussions. S.W. Bell acknowledges support from the Lunar and Planetary Institute Summer Intern Program. This research was conducted at the Lunar and Planetary Institute, which is operated by the USRA under contract CAN-NCC5-679 with NASA.

References

[1] Warren J., Zolensky M. et al. 1997 *Cosmic Dust Catalog Volume 15* NASA-JSC.

[2] Joswiak D.J., Matrajt G. et al. 2007 *LPSC* vol. 38 of *Lunar and Planetary Inst. Technical Report* pp. 2142–+.

[3] Joswiak D.J., Brownlee D.E. et al. 2008 *LPSC* vol. 39 of *Lunar and Planetary Inst. Technical Report* pp. 2177–+.

[4] Sammon J.J.W. 1969 *IEEE Transactions on computers* vol. 39 of *Lunar and Planetary Inst. Technical Report* pp. 401–409.