Compositional Units on Mercury along MESSENGER Ground Tracks from Principal Component Analysis of Spectral Observations. Jörn Helbert[1], Mario. D'Amore[1], Alessandro Maturilli[1], Noam R. Izenberg[2], Ann L. Sprague[3], Gregory M. Holsclaw[4], James W. Head[5], William E. McClintock[4], David T. Blewett[2] and Sean C. Solomon[6], [1]Institute for Planetary Research, DLR, 12489 Berlin, Germany (joern.helbert@dlr.de); [2]Johns Hopkins University Applied Physics Laboratory, Laurel, MD 20723; [3]Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ 85721, USA; [4]Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO 80303, USA; [5]Department of Geological Sciences, Brown University, Providence, RI 02912, USA; [6]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA

Introduction: The MESSENGER spacecraft performed two flybys of Mercury in 2008, a third will follow in September 2009, and in 2011 MESSENGER will enter orbit. During both flybys the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) instrument obtained spectra of the surface along ground tracks that each cross much of the planet [1,2]. We have started analysis of the surface spectra using a principal component approach. The main goal of this analysis is to identify surface units along the MASCS ground tracks and characterize them.

Preprocessing: The data coming from the MASCS spectrometer were first checked visually. Because of a high level of noise, the region between 800 and 925 nm, the wavelength band where the two channels of MASCS overlap, was excluded from this analysis. The calibrated spectra were converted to reflectance, taking into account the solar irradiance at Mercury [3] and, because of large variations in the viewing geometry, correcting for limb brightening. We attempted a basic phase angle correction, extracting the phase curve from RObotic Lunar Observatory (ROLO) data at the Apollo 16 landing site and interpolating these data for the effective MASCS viewing geometry.

Data analysis: To retrieve and characterize the number and spectral shapes of the different components present in the dataset we apply an R-mode factor analysis, a well-established technique in remote sensing [4,5,6]. The factor analysis expresses the data in a new vectorial base, for which the data covariance is minimized. The identification of the different components and their abundance is accomplished by principal component analysis (PCA). The eigenvectors and eigenvalues of the covariance matrix are evaluated, and the covariance matrix is decomposed in the space generated by the eigenvectors. The eigenvectors corresponding to larger eigenvalues are associated with most of the information contained in the data. The smaller (secondary) eigenvalues are related to featureless eigenvectors that contribute very little to the data.

Finding the crossing point between principal and secondary eigenvalues is the primary task of PCA. We used the eigenvalue ratio [4], the reconstruction error, and visual inspection of spectra to control models. The spectra in the dataset are assembled in matrix form as $D = R \cdot C$, where D is the matrix of the data, R the matrix of reconstructing vectors, and C the matrix of relative concentration coefficients. The goal of PCA is to decompose D into two matrices; R will be composed of the eigenvectors calculated from D, equivalent to diagonalizing the D matrix. There is no unique solution to this problem, and it is a common situation in remote sensing to have more equations than unknowns, resulting in an underdetermined system. An estimation of the vectors needed to reconstruct the data given the noise is the essential step to solve the problem and to be able to converge to an accurate solution.

Because of the wide spectral range of the data, we choose to apply the analysis both to the entire range and also to each individual channel, to monitor potential differences in behavior between the visible (VIS) and the near-infrared (NIR) portions of the spectrum.

Application to the full MASCS dataset shows that in general seven eigenvectors are sufficient to reconstruct the data within the error. Even if there are small differences between the two channels, the eigenvectors

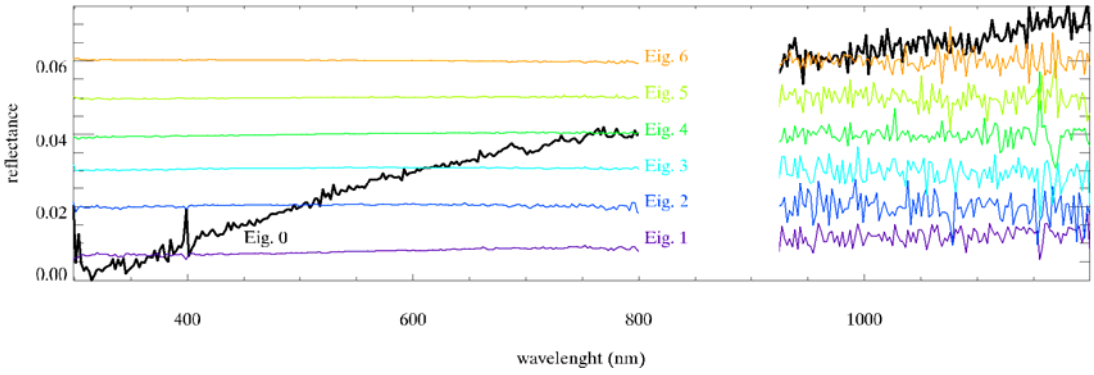

Fig. 1. Eigenvectors extracted for the entire spectral range of MASCS, each shifted by 0.01 in reflectance for clarity.

do not show strong differences. Fig. 1 shows the extracted eigenvectors for the entire spectral range. A comparison of the different channels indicates that the NIR portion is carrying significantly less information than the VIS portion. The first eigenvector always displays a strong reddish slope, compared with the others, and all eigenvectors show characteristic spectral signatures.

Each spectral eigenvector can be regarded as a representative of different spectral classes, changing in abundance along the track. The concentration coefficients in the C matrix indicate that spectral units show significant geographical variation. Moreover, the spectral unit variations show a strong correlation with surface units mapped by MESSENGER's Mercury Dual Imaging System (MDIS).

To characterize the spectral units at a more detailed level, we considered each observation as a collection of its C matrix coefficients, obtaining a 7-fold vectorial space where each point represents a single observation. Then we estimated the pairwise distance between each couple by a Chebyshev distance (or metric) algorithm. The estimated distance was used to compute the hierarchical clustering of the data points by a weighted centroid approach. In this approach the distance between two clusters is defined as the distance between the centroids of each cluster, and the centroid of a cluster is defined by the average position of all the sub-clusters, weighted by the number of objects in each sub-cluster.

Because of the high volume of data produced, we elaborated a visualization based on "clustering steps." At each process step the data points are clustered in the nearest cluster. After enough steps the closest points (closeness being defined by the adopted metric) are gathered together, leaving alone the farthest point. These points are away from the data cloud because they exhibit a rare combination of C matrix coefficients, and they must belong to exotic spectral units, relative to the observed surface.

Summary: By visualizing the "clustering steps" (Fig. 2) against the dataset geographical distribution we observe the presence of isolated spectral units. These spectral units show a strong correlation with surface units mapped by the MDIS imaging system. The next step is a detailed analysis of each identified unit. At the same time, we make use of the newly available high-temperature spectra from our Planetary Emissivity Laboratory [7] to progress toward the identification of the components of each unit.

References: [1] McClintock, W.E. et al. (2008) *Science*, *321*, 92-94; [2] Izenberg, N.R. et al. (2008) *Eos Trans. AGU*, 89(53), U11C-05; [3] Thuillier, G.M. et al. (2003) *Solar Physics*, *214*, 1-22; [4] Bandfield, J.L. et al. (2000) *JGR*, *105*, 9573-9588; [5] Ramsey, M.S. and Christensen, P.R. (1998) *JGR*, *103*, 577-596; [6] Smith, M. D. et al. (2000) *JGR*, *105*, 9589-9607; [7] Helbert, J. et al., *LPS*, *XL*, this meeting.

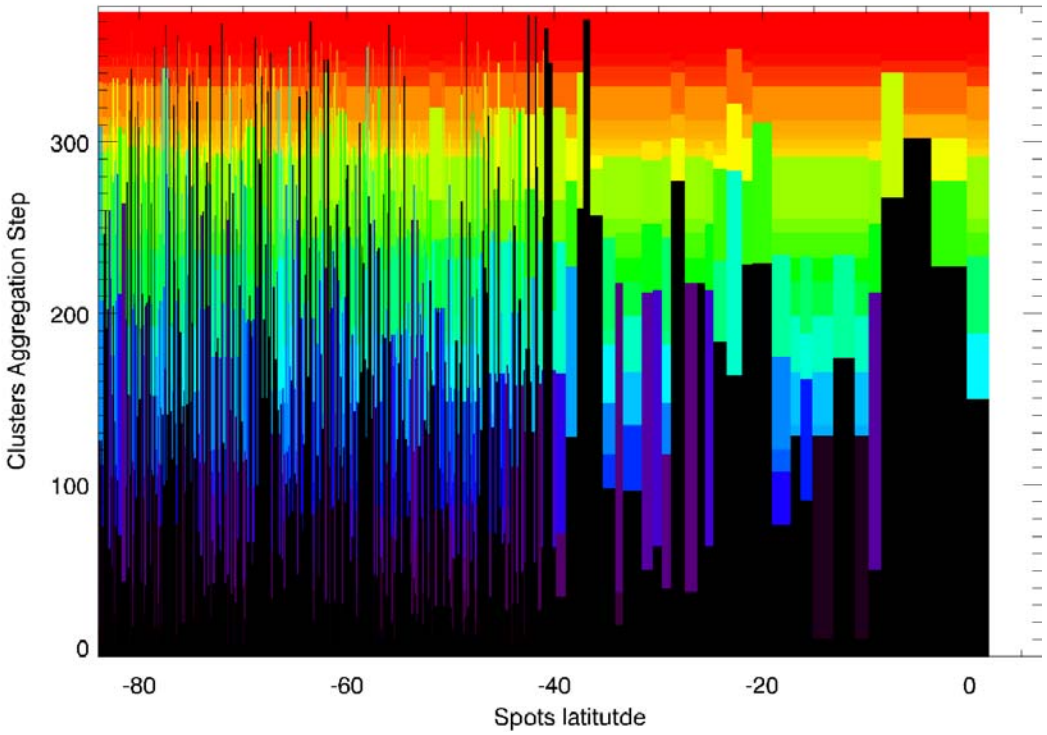

Fig. 2. Each pixel represents an observation. Starting at the bottom of the figures, at each step the algorithm gathers data in several clusters. The black color denotes all those data points not yet belonging to any cluster, while other color codes indicate that a point belongs to a specific cluster. At the starting point at the bottom of the figure, all points are isolated and therefore color-coded black. The clustering algorithm starts to group the closest points, resulting, on the top, in one big cluster. The long black lines reveal spectral units with exotic composition because they "resist" being categorized in any of the growing clusters.