COMPOSITIONAL UNITS ON MERCURY FROM PRINCIPAL COMPONENT ANALYSIS OF MESSENGER REFLECTANCE SPECTRA. Mario D'Amore[1], Jörn Helbert[1], Alessandro Maturilli[1], Noam R. Izenberg[2], Ann L. Sprague[3], Gregory M. Holsclaw[4], James W. Head[5], William E. McClintock[4], and Sean C. Solomon[6], [1]Institute for Planetary Research, DLR, 12489 Berlin, Germany (Mario.damore@dlr.de); [2]Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA; [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, Rhode Island 02912 USA; [6]Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington, DC 20015, USA.

Introduction: The MESSENGER spacecraft has completed three flybys of Mercury and in 2011 MESSENGER will enter into orbit about the innermost planet. During the first two flybys, the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) obtained spectra of the surface along instrument ground tracks, each of which traversed much of the planet [1,2]. We are analyzing surface spectra with a principal component approach and unsupervised classification techniques. The main goal of this analysis is to identify and characterize surface units along the MASCS ground tracks.

Data analysis: The data coming from the MASCS spectrometer [8] were first checked visually. Because of a high level of noise, the region between 800 and 925 nm wavelength, where the two channels of MASCS overlap, was excluded from this analysis. The calibrated spectra were converted to reflectance, taking into account the solar radiance at Mercury [3] without any other correction.

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-models goodness. The spectra in the dataset are assembled in matrix form as $\mathbf{D} = \mathbf{R} \cdot \mathbf{C}$, 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 spectral differences between the two channels, the eigenvectors resulting from VIS analysis do not display pronounced differences from those of the combined VIS+NIR analysis, indicating that the NIR portion is carrying significantly less information than the VIS portion of the spectra.

The first eigenvector always displays a strong positive or "red" slope with increasing wavelength, and all

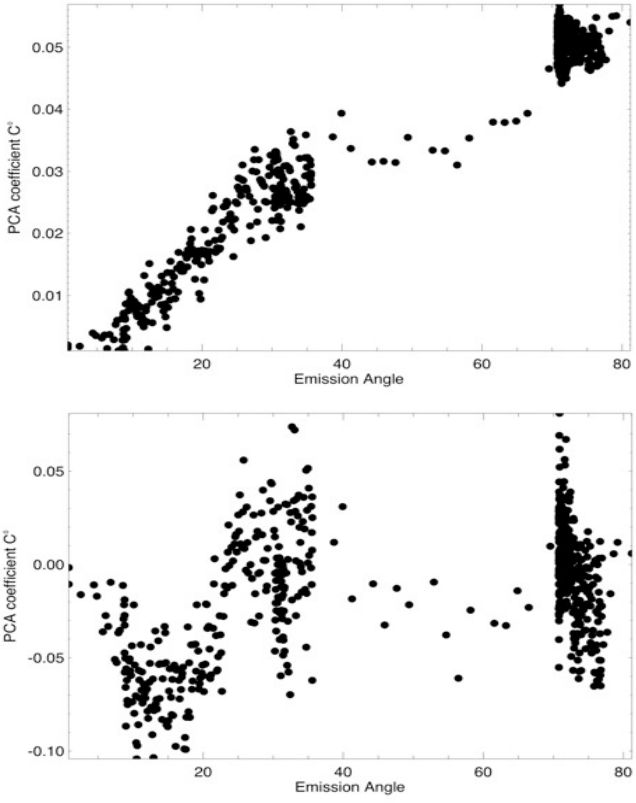

Fig 1. First eigenvector concentration coefficients versus emission angle, before (top) and after (bottom) decorrelation.

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 components of the data in eigenvector space, or 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 color images obtained with MESSENGER's Mercury Dual Imaging System (MDIS). Some concentration coefficients display a clear correlation when plotted versus observational parameter. The strongest feature can be observed in the first component, indicating that the corresponding eigenvector describes the most of the variance linked to changes in observation geometry. We apply a decorrelation technique (Mahalanobis transformation) to completely remove dependence on observation angle in the retrieved concentration coefficients, since we do not photometrically correct the data (Fig. 1).

To characterize the spectral units and to expose hidden geographical correlations between units, we map each observation into its C matrix coefficients, obtaining a 7-fold vectorial space where each point represents a single observation. We then 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, so they belong to uncommon spectral units.

To compare this result qualitatively with the MDIS data, we calculate for each MDIS field of view an average pixel for the 11 color channels of the wide-angle camera. Using that observation as input to the routines, we again obtained several clusters strongly correlated with surface features.

The Rudaki plains area in particular shows a clear correspondence between the MASCS units and the MDIS units (Fig. 2).

The next step will include a detailed analysis of each identified unit. At the same time, we will make use of the newly available high-temperature spectra from our Planetary Emissivity Laboratory to progress toward the identification of the compositional components of each unit. These applications give us confidence in the ability of these techniques to extract physical properties of surface material [7].

Summary: By visualizing the clustering steps in map coordinates we can identify the presence of isolated spectral units distinguishable in the MASCS surface reflectance observations. These spectral units show a strong correlation with surface units mapped by the MDIS imaging system, in particular in the Rudaki plains area in the example shown here. The use of the upcoming high-temperature spectra from our Planetary Emissivity Laboratory [7] will allow to decipher the components of each unit defined in this work.

References: [1] McClintock, W.E. et al. (2008) *Science, 321*, 92-94; [2] Izenberg, N.R. et al. (2008) *Eos Trans. AGU, 89*(53), Fall Meeting suppl., 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. (2010), *LPS, 41th*, this meeting; [8] Helbert, J. et al. (2010), *LPS, 41th*, this meeting.

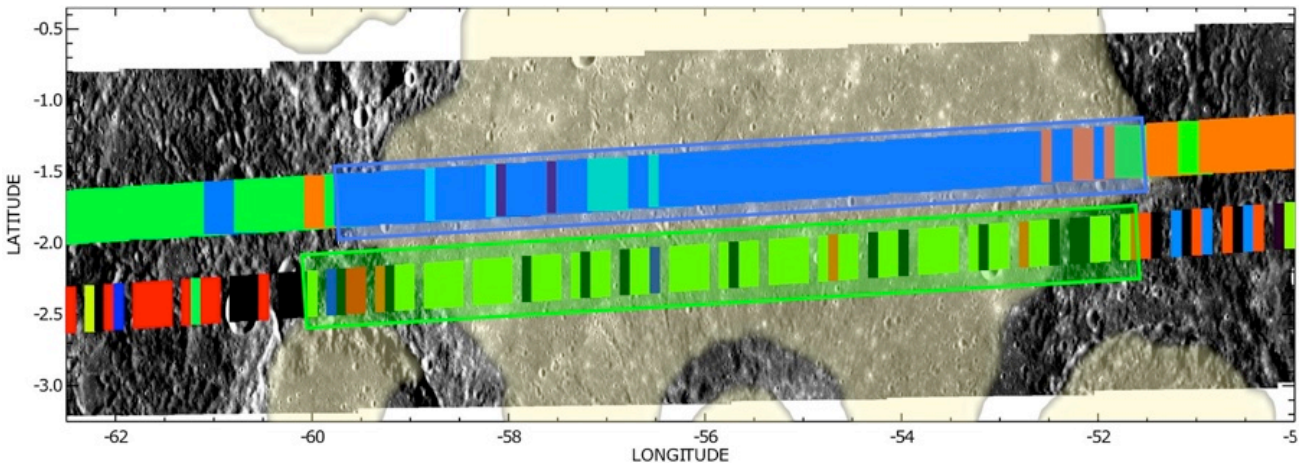

Fig 2. Spectral units defined by clusters. The top strip represents the MASCS data, with the overlaid blue box defining the Rudaki plains. The bottom strip represents the MDIS color images, shifted of about half a degree south for clarity.