EPSC Abstracts
Vol. 12, EPSC2018-738, 2018
European Planetary Science Congress 2018

Semi-automated surface mapping via unsupervised classification Mercury'S Visible–Near-Infrared reflectance spectra

Mario D'Amore(1), Jörn Helbert(1), Alessandro Maturilli(1), Indhu Varatharajan (1).

(1) Institute for Planetary Research, DLR, Rutherfordstrasse 2, Berlin, Germany (mario.damore@dlr.de)

1. Abstract

The surface of Mercury has been mapped in the 400–1145 nm wavelength range by the Mercury Atmospheric and Surface Composition Spectrometer (MASCS) instrument during orbital observations by the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft. Under the hypothesis that surface compositional information can be efficiently derived from spectral reflectance measurements with the use of statistical techniques, we have conducted unsupervised hierarchical clustering analyses to identify and characterize spectral units from MASCS observations. The results display a dichotomy, with two spectrally distinct groups: polar and equatorial units (see Fig. 1). The spatial extent of the polar unit in the northern hemisphere generally correlates well with that of the northern volcanic plains [1]. We extended our analysis on the latest MESENGER data delivery to PDS including the new spectral photometric correction ([2] in review, extension of [3]) , finding result consistent with our previous analysis based on our custom photometric effect removal.

2. Methods

2.1 Data managing: PostgreSQL

The most recent version of our data analysis procedure uses PostgreSQL, a type of database management that controls the creation, integrity, maintenance and use of a database. It embeds a high-level query language, which greatly simplifies database organization as well as retrieval and presentation of database information. We set up a data pipeline using the to update automatically the MASCS data, read them from the NASA Planetary Data System format, regrid the data to a common grid length, and store all information in the database. All data are then readily available to any authorized user in our network. We are working on a library to access the data directly from within our analysis software, and some preliminary functions have been implemented. As an example, the calculation of a parameter representing the database takes a few seconds even for the full dataset of ~5 million entries, if exploiting pre-indexed columns. It is thus straightforward to create and analyze rapidly the data, as for example the distribution of normalized radiance at a fixed wavelength. The new methodology provides facilities for controlling data access, enforcing data integrity, managing concurrency control, and recovering the database after a failure and restoring it from backup files, as well as maintaining database security.

2.1 Data retrieval : PostGIS

We use PostGIS that adds support for geographic objects in geographic information system and extends the database language with functions to create and manipulate geographic objects. A typical application is the definition of a large number of regions of interest (ROIs) and the search for all data points falling within each ROI. This facility may be used to extract spectral signatures specific to user-defined geological units in a few seconds and to explore the properties of the data from the different ROIs. A typical search for data from areas defined by a simple ROI, such as regions of impact melt associated with a given crater, takes less than 1 second. More elaborated query requires more computational power and result in longer response time. A typical example is the search for a MASCS measurement where no point outside a ROI were taken, that could increase the response time to > 60 seconds. We resample the whole dataset of ~5 Milion spectra on a planetary fixed grid or extract the information from collection of planetary region of interest in few minutes, allowing to quickly analyze the spectral characteristic of Mercury. We successfully tested remote access to the database using through a GIS visualization system, creating data visualization on the surface of Mercury that layers camera data and real-time-queried MASCS data.

solar lines in the calibrated spectra. The resultant hyperspectral map was then visually inspected to search for anomalies that originated mainly in regions of low coverage or from high levels of spectral variation within a single pixel. Our approach consist of 1. a data cleaning step, to remove data artifact, 2. Principal Component Analysis (PCA) feature compression and 3. K-means clustering (see scikit-learn python implementation in [4]). We found the existence of two large and spectrally distinct regions, which we call the polar spectral unit (PSU) and the equatorial spectral unit (ESU) (Fig. 1). Further analysis indicates the presence of smaller sub-units that lie near the boundaries of these large regions and may be transitional areas of intermediate spectral character.

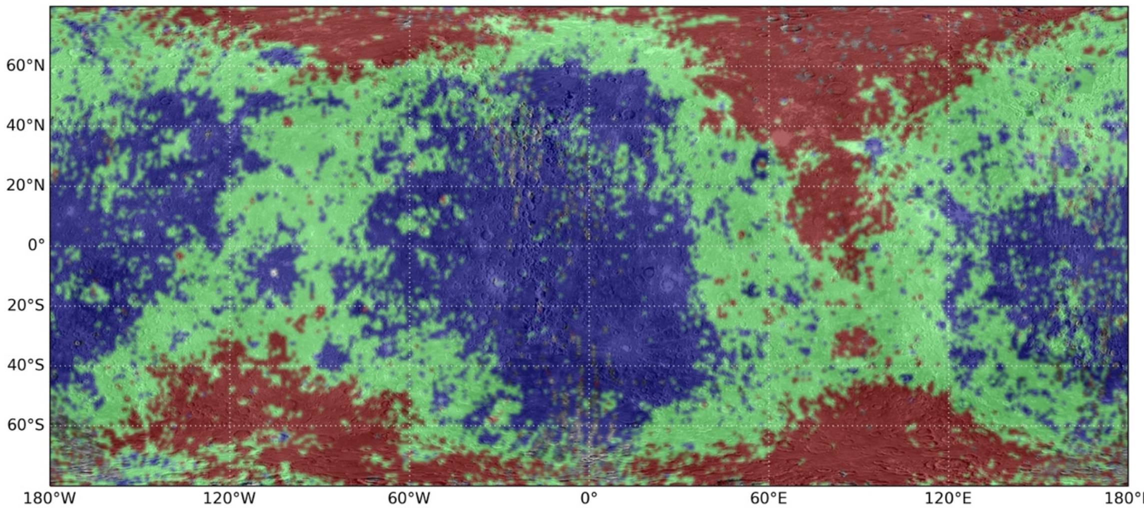

Figure 1. Spatial distribution of the unsupervised hierarchical clusters derived from normalized MASCS VIS spectra (overlaid on an MDIS base map). Red pixels indicate the polar spectral unit (PSU); blue pixels denote the equatorial spectral unit (ESU). Green is the intermediate transition region.

2.1 Machine Learning on multivariate data

We use a global hyperspectral data cube image of normalized MASCS visible (VIS) detector spectra, from the first Earth year of the orbital mission, to perform our unsupervised hierarchical clustering analysis. We reduced the grid resolution to 4°/pixel to improve the spatial coverage of the final map, but at the cost of increased sub-pixel variation. Data coverage varies from region to region, but global maps at 1°/pixel can be obtained with a high signal-to-noise ratio (SNR). With the absence of a formal global photometric correction for the MASCS data, we have corrected the dataset in an approximate fashion by normalizing all the spectra at 700 nm to account for large variations in observing geometry. We have excluded the most extreme observing geometries by limiting the incidence and emission angles to <85°, which means that latitudes poleward of 80° are excluded. For this analysis we used six spectral channels, each with a bandwidth of 10 nm, in order to focus on “interesting” spectral regions, e.g., bands at ~600 nm that are indicative of sulfides. By using wider spectral channels, we also avoided biases caused by artifacts in the spectra, e.g., the presence of solar lines in the calibrated spectra.

References

[1] Head, J. W. et al. (2011), Science, 333, 1853–1856.

[2] Domingue, D. L. et al, Icarus, in review.

[3] Domingue, D. L. et al, Icarus, 209, 2010, 101-124.

[4] Pedregosa, F., Journal of Machine Learning Research, 2013, 108-122