

Automated surface mapping via unsupervised learning and classification of Mercury Visible–Near-Infrared reflectance spectra

Mario D’Amore¹, Sebastiano Padovan^{1,2,3}

¹*German Aerospace Center (DLR), Rutherfordstraße 2, 12489 Berlin, Germany*

²*EUMETSAT, Eumetsat Allee 1, 64295 Darmstadt, Germany*

³*WGS, Berliner Allee 47, 64295 Darmstadt, Germany*

Abstract

In this work we apply unsupervised learning techniques for dimensionality reduction and clustering to remote sensing hyperspectral Visible–Near Infrared (VNIR) reflectance spectra datasets of the planet Mercury obtained by the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) mission. This approach produces cluster maps, which group different regions of the surface based on the properties of their spectra as inferred during the learning process. While results depend on the choice of model parameters and available data, comparison to expert-generated geologic maps shows that some clusters correspond to expert-mapped classes such as smooth plains on Mercury. These automatically generated maps can serve as a starting point or comparison for traditional methods of creating geologic maps based on spectral patterns. The code and data used in this work is available as python jupyter notebook on the github public repository [MESSENGER-Mercury-Surface-Cassification-Unsupervised_DLR](#)¹ funded by the European Union’s Horizon 2020 grant No 871149.

Keywords: Semi-automated, surface, mapping, unsupervised learning, classification, Mercury, Visible–Near-Infrared, reflectance, spectra

¹ 1. Introduction

The sheer amount of data returned by scientific missions aimed at exploring the solar system and observing exoplanets in recent decades overwhelms classical methods to explore and discover important scientific aspects of the target body. As an example, the Mercury data return for Mariner 10 was less than 100 MB, while MESSENGER delivered about 23 TB. Future missions are expected to exceed this limit. In addition, there is a trend of increasing complexity in the data itself, e.g., going from the of Mariner-10 to the hyperspectral datasets expected from BepiColombo. This situation clearly indicates that some form of automated analysis would be beneficial, provided it is able to save time without a loss of the information content of the data.

Keeping the focus on hyperspectral remote sensing data, the typical approach for analysing this kind of data is to model the observed radiation with a forward radiative model (like Hapke, as in [Hamilton et al., 2005](#)) or attempt to reproduce the observed radiation by setting up relevant samples in a laboratory setting using chemical and/or geomorphological context information (e.g., [Helbert et al., 2013](#)). Complex forward models that are able to take into account the relevant physics are

¹https://github.com/epn-ml/MESSENGER-Mercury-Surface-Cassification-Unsupervised_DLR

15 typically computationally intensive and difficult to use to investigate the very large parameter space
16 covered by hyperspectral data. This consideration is even more relevant for laboratory investigations
17 : physical simulation needs the target to be physically fabricated, hence more and and more param-
18 eters means more experiments and more time. Models need computational power to be calculated in
19 a reasonable amount of time, but could be distributed on several machines to overcome this limita-
20 tion. This workaround is not effective for laboratory experiment, because most only few places meets
21 of the environment needed for space sample simulation, like high-vacuum, -temperature, -radiation
22 and so on.

23 Without a way to efficiently and rapidly explore large amounts of complex data, it is likely that
24 valuable information will be missed in large hyperspectral data sets.

25 Geological maps are the gold standard for remote planetary surface studies, but producing them
26 is an extremely time-consuming task. This process can suffer from user bias and typically only uses a
27 few data points (e.g., 3-channel images) to describe different units. For example, [Denevi et al. \(2009\)](#)
28 mapped the distribution and extent of major terrain types of Mercury using MESSENGER Mercury
29 Dual Imaging System (MDIS) camera observations of Mercury. While the camera has 11 spectral
30 bands, the maps typically used for the terrain differentiation are RGB, where 3 representative
31 spectral bands are mapped onto the three image color channels.

32 Geomorphological maps take in account additional features like surface roughness and crater
33 density as a proxy for the age, where the correlation between age and crater density are derived
34 from models (e.g., [Bland, 2003](#); [Kerr, 2006](#)). Automated techniques are becoming more common in
35 planetary science applications, as this books testifies, and the aim of this chapter is to illustrate how
36 to apply unsupervised learning techniques to remote sensing data. This approach requires minimal
37 user interaction and yields scientifically interesting products like classification maps that can be
38 directly compared with geomorphological maps and models. We present an analysis of spectral re-
39 flectance data of Mercury’s surface collected by the Mercury Atmospheric and Surface Composition
40 Spectrometer (MASCS) instrument during orbital observations of the NASA MESSENGER mission
41 between 2011 and 2015 ([McClintock and Lankton, 2007](#)). MASCS is a three sensor point spectrom-
42 eter with a spectral coverage from 200 nm to 1450 nm. After a brief overview of the instrument
43 and its significance for the investigation of Mercury (section 2), we will illustrate how we extract
44 and resample the data to a format useful for our ML application (section 3). Then we show how
45 to compress the data (section 4.1), how to project them to a lower number of dimensions (section
46 4.2), and finally, how to group “similar” data points together to discover salient spectral classes and
47 their distribution on the surface. We conclude in section 4.4 by providing a basic comparison of
48 the result of the discovered spectral class distribution with maps of the surface of Mercury obtained
49 using classical methods, in order to provide a first assessment of the machine learning techniques
50 presented here.

51 2. Mercury and the MASCS instrument

52 Surface mineralogy and composition are important indicators of the past of a planetary body,
53 since they provide hints about the processes that formed and altered the crust, which is largely the
54 result of the interior evolution. For example, the possibility of identifying specific mineral assemblage
55 like metamorphic rocks, which are known to form in specific pressure and temperature conditions,
56 would provide indications on the physical processes occurring in the subsurface that produced those
57 rocks and later transported the rocks to the surface (e.g., [Namur and Charlier, 2017](#)). Similarly,
58 observations of hydrated minerals can be interpreted as indicating the possible past presence of
59 water, as in the case of Mars ([Meslin et al., 2013](#)).

While some investigations have been published on Mercury’s surface mineralogy (e.g., E. Vander Kaaden et al., 2017; Namur and Charlier, 2017; Vilas et al., 2016; Sprague et al., 2009), its link to the endogenous (e.g., mantle convection) and exogenous (e.g., impacts) processes that operated during the history of the planet is still difficult to elucidate (e.g., Padovan et al., 2017). A relevant example is the geological features known as hollows, discovered on the surface of Mercury in MESSENGER data. Hollows are rimless depressions with flat floors, surrounded by halos of high-albedo material, and typically found in clusters (?). Given this evidence, their formation mechanism likely includes the loss of volatile material through one or more processes such as sublimation, space weathering, outgassing, or pyroclastic flow. Hollows are associated with a particular spectral signature in MESSENGER’s MDIS camera (Vilas et al., 2016), but a specific spectral signature in spectrometer data could not be identified due to the coarse spatial resolution of the spectrometer. Overall, the only clear inference based on VNIR spectra obtained by the MASCS instrument is that Mercury’s surface shows little variation, displaying no distinct spectral features except for the possible indication of sulfide mineralogy within the hollows (Vilas et al., 2016).

MASCS consists of a small Cassegrain telescope with an effective focal length of 257 mm and a 50-mm aperture that simultaneously feeds an UltraViolet and Visible Spectrometer (UVVS) and a Visible and InfraRed Spectrograph (VIRS) channel. VIRS is a fixed concave grating spectrograph with a focal length of 210 mm, equipped with a beam splitter that simultaneously disperses the light onto a 512-element sensor (VIS, 300–1050 nm) and a 256-element infrared sensor array (NIR, 850–1450 nm). Data obtained by MASCS covers almost the entire surface of Mercury. The spatial resolution is highly latitude dependent due to the very elliptical orbit of the spacecraft, but a reference value ~ 5 km. This low spatial resolution is a trade-off for higher spectral resolution and more spectral channels compared to the imaging instruments (i.e., the MDIS).

The NIR sensor is characterized by 3 – 5 times lower signal-to-noise ratios (SNRs) than the VIS detector and does not add significant information to the VIS sensor in our tests. NIR measurement can be linked and corrected to match corresponding VIS measurements following (However, see Besse et al., 2015, for a successful VIS/NIR cross correction.). The biggest obstacle is that the most accurate photometric corrections is only available for the VIS channel (see. Domingue et al., 2019a,b). We then analysed only data from VIS channel, that is enough for the sake of illustrating unsupervised learning techniques.

3. Data preparation

In what follows, we use “feature” to refer to a single spectral reflectance value in a given spectrum and refer to the spectrum as an “observation”. An observation is taken on a particular area of the surface of Mercury at a given time and thus with a given illumination condition. MASCS data are obtained as a collection of binary files following the PDS3 standard (NASA, 2009). We need to preprocess these to be represented as a matrix of shape $[N \times M]$ where M is the number of features and N is the number of observations before they can be used as input to typical machine learning algorithms. A spectrum can be represented as a vector, with the single spectral channels being its components, i.e., $\hat{x} = (x_1, \dots, x_{512})$, for the case where the spectrum \hat{x} has 512 components. The entire MASCS dataset comprises about 5 million spectra, each with 512 channels. The coverage is not homogeneous across Mercury’s surface, as illustrated in Figure 1a. In this section we describe how to create a homogeneous gridded representation of the observations across the entire surface so that we can apply our clustering algorithm to the entire surface, as shown in Figure 1b.

A typical workflow for doing a planetary science analysis starts with the definition of a number of Regions Of Interests (ROIs), followed by a search for all data points that fall within these ROIs. This approach facilitates the extraction of spectral signatures that are specific to user-defined geological

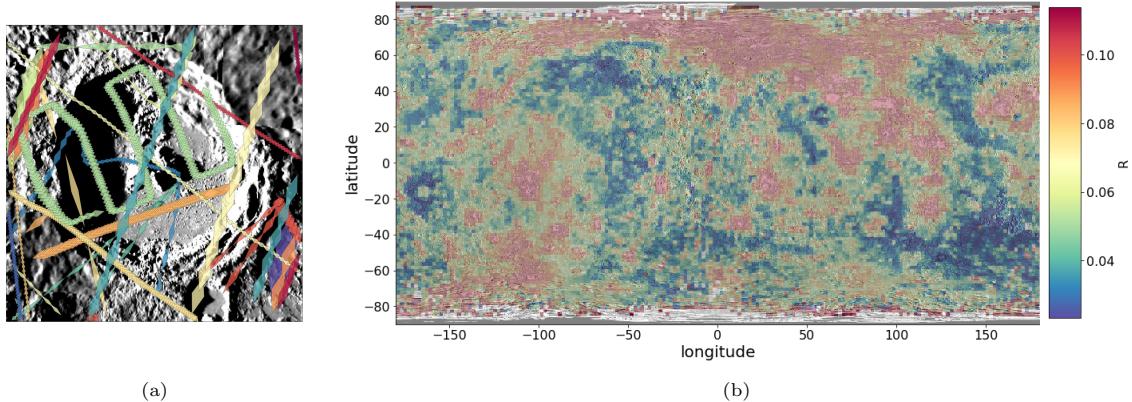


Figure 1: **a.** Example of MASCS VIRS observations in the region of the Kuiper crater, a 62-km-diameter crater with a central peak located at 11.35°S 31.23°W on Mercury. Kuiper is the brightest landmark on the planet and has a fresh ray system, suggesting that it is among the youngest craters of Mercury Neish et al. (2013); D’Incecco et al. (2015). Each polygon approximately represents a VIS and NIR observation (in reality the two channel are slightly misaligned). Different colors represent different spacecraft orbits. Note how different orbits show different polygon sizes, probably due to different observation geometry or spacecraft altitude. **b.** Square polygons are used as base pixels to construct a global datacube, using the median value of all the observations within each square as the final pixel reflectance. This plot shows reflectance values from the 700 nm band of VIS.

106 units in order to explore their properties in the context of different ROIs. For this chapter, we
 107 developed a workflow that starts by extracting the data from the original files using the GDAL²
 108 python binding. We then organized them in a PostgreSQL relational database³, with the PostGIS
 109 spatial extension⁴. The database is currently hosted at the DLR in Berlin.

110 First, the whole dataset of ~ 5 million spectra is resampled to a planet-wide rectangular grid of
 111 $1^\circ \times 1^\circ$ in the latitudinal band between $\pm 80^\circ$. The cell longitudinal size varies between ~ 40 km at
 112 the equator to a minimum of ~ 10 km at $\pm 80^\circ$. Thus, the area spanned by each grid cell depends on
 113 the latitude. However, the same is true for the acquisition process, where higher spatial resolution
 114 is reached near the equator and lower resolution at the poles. Next, we resampled the data in the
 115 spectral dimension to a common wavelength range⁵ from 260 nm to 1052 nm with a 4 nm resolution
 116 (2 nm spectral sampling), resulting in 396 spectral channels. This approach slightly oversamples the
 117 original 4.77 nm spectral resolution and removes some points from the original 200-1050 nm range.
 118 The resulting data matrix is expressed in tabular form, with each row representing a single grid cell
 119 or pixel on the surface. The elements of each row are the spectral reflectance values from the VIS
 120 instrument at 396 (resampled) wavelengths. The dataset now has dimension $[N \times M]$ where N is the
 121 number of grid cells ($360 \times 180 = 64,800$) and M is the number of spectral features (396). Due to

²We found a GDAL bug when reading 8 bytes real values that is solved for version $\geq 2.3.0$ after our report to the developer. See <https://trac.osgeo.org/gdal/wiki/Release/2.3.0-News> and use this version or higher when manipulating MASCS data.

³PostgreSQL is a relational database management that controls the creation, integrity, maintenance and use of a database

⁴PostGIS adds support for geographic objects in geographic information system and extends the database language with functions to create and manipulate geographic objects. PostGIS follows the Simple Features for SQL specification from the Open Geospatial Consortium (OGC).

⁵MASCS VIS data have different wavelength sampling and part of the global Mercury campaign had different spectral binning.

122 the incomplete coverage and data filtering, some grid cells are empty. After removing these empty
 123 cells, the size of the dataset is $[55399 \times 396]$. As an example, Figure 1b shows the distribution of
 124 the normalized reflectance at a fixed wavelength (700 nm).

125 We used a recent dataset that had large-scale photometric corrections (Domingue et al., 2019a,b)
 126 and thus was almost free from observation geometry effects. However, extreme geometry are still
 127 present and are typically associated with high noise and some residual instrumental effects. Based
 128 on our empirical tests, we filtered out observations with an emission/incidence angle $\geq 80^\circ$. We
 129 also calculated the median value per wavelength and per cell grid when constructing the global
 130 hyperspectral data cube and filtered out observations falling under the 2nd percentile and above
 131 99.9th percentile to clean some residual geometry effects. With this approach we create an effective
 132 noise filter while retaining enough observations to be able to analyse the entirety of the surface of
 133 the planet.

134 4. Learning from multivariate data

135 4.1. Dimensionality reduction: ICA

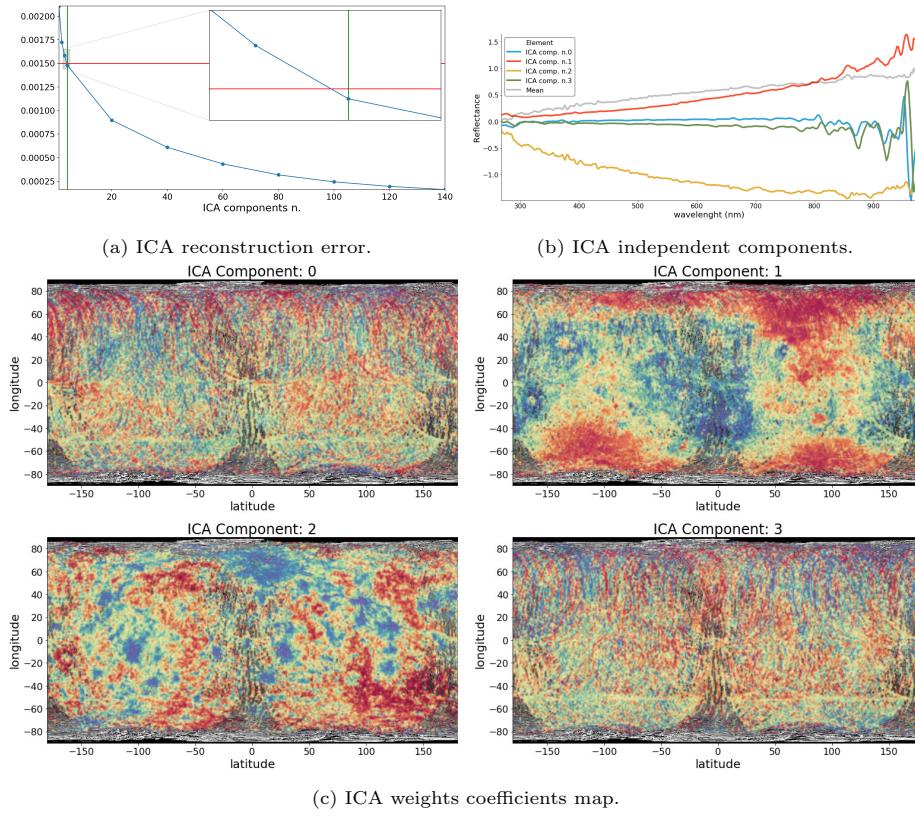


Figure 2

136 One way to discover salient patterns in a large set of multidimensional data is to perform dimen-
 137 sionality reduction. Two popular techniques for dimensionality reduction are Independent Compo-
 138 nent Analysis (ICA) and Principal Component Analysis (PCA).

139 PCA try to find a reduced-rank representation of the data and seeks for that best explains the
140 variability of the data.

141 ICA is a case of blind source separation techniques and assumes that the observations are non-
142 Gaussian signals and statistically independent from each other (Hastie et al., 2009; Hyvärinen and
143 Oja, 2000). ICA aims to find a set of new basis vectors as Principal Component Analysis, but with
144 different assumptions and significance of the new basis. ICA search a basis where each vector is an
145 independent component: if one think of a mix of superposed audio signals, the ICA basis will return
146 a vector for each independent signal, like in the Blind Signal Separation techniques. ICA helps to
147 find a representation of the data as independent sub-elements.

148 In practical terms, PCA helps to compress data and ICA helps to separate data.

149 The data cube obtained from previous section is then decomposed with ICA, trying to separate
150 multivariate signal into additive sub-components.

151 We represent the input data as $\hat{x} = (\hat{x}_1, \dots, \hat{x}_n)^T$, where \hat{x}_i is a vector of 396 spectral reflectance
152 values for a given observation. ICA tries to find a linear transformation W so that $\hat{s} = W\hat{x}$, where
153 $\hat{s} = (\hat{s}_1, \dots, \hat{s}_k)^T$ is a vector of maximally independent components and $k \leq n$. W is also called the
154 mixing matrix and has dimension $[f] = [(\text{data dimension}) \times (\text{number of ICA independent sources})]$.
155 The original sources \hat{s} can be recovered by multiplying the observed signals \hat{x} with the inverse of
156 the mixing matrix $W = A^{-1}$, also known as the unmixing matrix.

157 The goal is to find a number of independent components that can somehow be connected to
158 geologic patterns observed on the surface. There is no “true number” of independent components
159 based on the surface geology, so to choose k we use a reconstruction error threshold that represents
160 the noise in the data. This value is chosen from the residual standard deviation when calculating the
161 data median for each pixel in the surface $1^\circ \times 1^\circ$ grid (see 3) and represents the sub-pixel variation
162 of the data.

163 The independent vectors s are function of the number k of component and by increasing it
164 the reconstruction error decrease. Figure 2a shows how the reconstruction error decreases with
165 additional components and that the condition $\|x - s(n)\| < 0.0015$ is true for $n \geq 4$. The final result
166 is a compression of the data from $[55399 \times 396]$ to $[55399 \times 4]$.

167 In this application it is interesting to visualize the mixing weight matrix W , i.e., the linear op-
168 erator that maps independent sources to the data \hat{x} . The weight coefficient for each component s
169 represents how close each data point is to the corresponding independent component. Figure 2c
170 shows the weight coefficients maps, illustrating the spatial distribution of each component. Com-
171 ponents 1 and 2 show well-defined clusters enriched (red) and depleted (blue) in the corresponding
172 component. For components 0 and 3, the spatial distribution does not display any particular struc-
173 ture. This could be the result of some residual instrumental effects that introduce variance in the
174 data.

175 4.2. Manifold learning

176 In the previous section, we reduced the entire dataset to a representation where each pixel in
177 the surface grid was represented by 4 features computed using ICA instead of 396 spectral values
178 (Figure 2c). Before performing clustering on the dataset, we further reduced the dimensionality using
179 manifold learning techniques. While it is technically possible to use manifold learning techniques
180 to reduce the dimensionality without the ICA step, we applied ICA first because manifold learning
181 algorithms do not computationally perform well for high-dimensional data. This is why they are
182 usually used after a data compression step.

183 A number of supervised and unsupervised linear dimensionality reduction techniques exist, in-
184 cluding but not limited to: Isomap, Locally Linear Embedding (LLE), Hessian Eigenmapping, t-
185 distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and

186 Projection (UMAP). See, e.g., the following review papers and the references therein : [Tenenbaum et al. \(2000\)](#); [Roweis and Saul \(2000\)](#); [Donoho and Grimes \(2003\)](#); [van der Maaten and Hinton \(2008\)](#); [McInnes et al. \(2020\)](#). The linear algorithms most important aspect is that their results
187 are be easier to interpret than non-linear algorithms, but by definition they could miss important
188 non-linear structures in the data.
189

190 Manifold Learning methods are a class of non-linear dimensionality reduction techniques. They
191 are based on the idea that the intrinsic dimensionality of many data sets is actually much lower than
192 its actual dimensionality.
193

194 The linear dimensionality reduction techniques can be considered also dimensionality reduction
195 techniques, because this terms embraces a broader set of algorithms. If we look at ICA and the
196 Manifold Learning methods as a black box, they both compress input data. Their underlying
197 assumption, the steps to find the lower dimension representation and the insight we gain from them
198 are different, as we will see below.
199

200 The typical manifold learning problem is unsupervised: it learns the high-dimensional structure
201 of the data from the data itself, without the use of predetermined classes ([Lee and Verleysen, 2007](#)).
202

203 UMAP and t-SNE are popular methods for manifold learning. At its core, t-SNE projects high-
204 dimensional data in a space of lower dimensions. t-SNE initialization aims to model the probability
205 distribution that represents similarities between neighbors (data points). It starts with a small
206 random Gaussian distribution and for the first iterations the input probabilities are multiplied by
207 an integer factor to 'exaggerate' the distribution. t-SNE converts similarities between data points
208 to joint probabilities then tries to minimize the divergence between the joint probabilities of the
209 low-dimensional embedding and the high-dimensional data. Original data points that are similar
210 are modeled by nearby points in the low dimensional embedding. Dissimilar data points are modeled
211 by distant points with high probability in the embedded space.
212

213 Compared to other techniques, t-SNE is particularly sensitive to local structure and can reveal
214 the structure at many scales on a single map for data that lie in different manifolds or clusters.
215 While linear methods such as Isomap and LLE unfold a single continuous low dimensional manifold,
216 t-SNE will tend to extract clustered local groups of samples, since it focuses on the local structure
217 of the data. For a complete overview of t-SNE we refer the reader to [Roweis and Saul \(2000\)](#) and
218 the references therein.
219

220 While powerful, t-SNE is extremely computationally expensive and can take several hours on
221 million-sample data sets where PCA will take seconds to minutes. In addition, the downside of
222 using a random initialization is repeatability: even if you use a fixed seed, there is no guarantee
223 that the final results will be the same. This means that t-SNE can produce results with different
224 initialization on the same data, so a good practice is to run t-SNE for multiple random seeds for
225 the same dataset. Thus, we chose to apply UMAP ([McInnes et al., 2020](#)) since this method offers
226 a number of advantages over t-SNE, most notably increased speed with preservation of the global
227 data structure.
228

229 We will give here an high level description of UMAP, trying not to make the description too
230 complicated and not to fall into areas beyond this book. This can leave room for ambiguity and errors
231 that are only our fault due to too much simplification of a complex topic. A rigorous mathematical
232 background could be fund in [McInnes et al. \(2020\)](#) and nice visualisation in the https://umap-learn.readthedocs.io/en/latest/how_umap_works.html⁶.

233 UMAP uses local manifold approximations, represent those with a local fuzzy simplicial set
234 representations and uses their union to construct a topological representation of the high dimensional
235 space.
236

6https://umap-learn.readthedocs.io/en/latest/how_umap_works.html

231 data. A simplicial complex is a set composed of points, line segments, triangles, and their n-
232 dimensional counterparts. Fuzzy topology means that being in an open set is no longer a binary
233 yes or no, but instead a fuzzy value between zero and one. The probability of being in a ball
234 of a given radius for points will decay moving away from the center of the ball. Given some low
235 dimensional representation of the data, a similar process can be used to construct an equivalent
236 topological representation in the other direction (low to higher dimension). UMAP then searches
237 for a low dimensional projection of the data that has the closest possible equivalent fuzzy topological
238 structure to the high dimensional one.

239 UMAP follows the philosophy of t-SNE, but has a number of improvements such a different cost
240 function and the absence of normalization of high- and low-dimensional probabilities. Further details
241 can be found in the original paper, [McInnes et al. \(2020\)](#), and its reference python implementation,
242 [McInnes \(2018\)](#).

243 The two main hyperparameters controlling the representation learned by UMAP are *n_neighbors*
244 and *min_dist*⁷. With *n_neighbors* one can adjust the sensitivity to local or global structures in the
245 data by constraining the size of the local neighborhood. Low values correspond to a focus on
246 very local structures, possibly losing the big picture, while large values, corresponding to larger
247 neighborhoods, may miss the fine structure. The best value to choose depends on the particular
248 data structure and the desired range (local vs. global) the application is required to probe. Values
249 can go from 0 (local) to the size of the data (global). *min_dist* regulates how points are packed
250 together, providing the minimum distance apart that points are allowed to be in the low dimensional
251 representation. Low values will result in more dense embeddings while larger values will result in
252 more sparse embeddings.

253 Figure 3a shows the density pair plots of the four ICA weight coefficients. This shows that the
254 data manifold is a slightly asymmetric 4D cloud, with most of the data concentrated in the center.
255 Figure 3b shows the UMAP density plot on the ICA weight coefficients for various combinations
256 of the two hyperparameters described above. The projected data topology follow the 4D manifold,
257 with a central core of high density observations and a rarefied peripheral region. We chose to use
258 *n_neighbors* = 4000 and *min_dist* = 0.99. In our particular application because local structures are
259 not apparent at any level, so the only significant structure is the overall shape of the low dimension
260 representation and how where the data points lies relative to each other.

261 After using UMAP to further reduce the representation of the dataset compressed using ICA,
262 the final dimensionality of the dataset is [55399 × 2].

263 4.3. Cluster analysis

264 After reducing the dimensionality of the data, our final goal is to identify surface regions with
265 similar spectral properties that we can use to make geologic interpretations. As described in the
266 Introduction, this is because geologic surfaces are generally classified into a few classes which are
267 related with a geological taxonomy.

268 Cluster analysis comprises a series of techniques that attempt to group similar observations in
269 groups or clusters. The idea is that observations belonging to a cluster are more similar to each
270 other than to those in other clusters. Before clustering the data, we first standardized the data by
271 subtracting the mean and scaling them to unit variance. This is a standard and often necessary step
272 for any machine learning algorithm.

⁷The interactive tutorial "Understanding UMAP" give some insight in how the hyperparameters influence UMAP. See [Coenen and Pearce \(2019\)](#) and <https://pair-code.github.io/understanding-346umap>

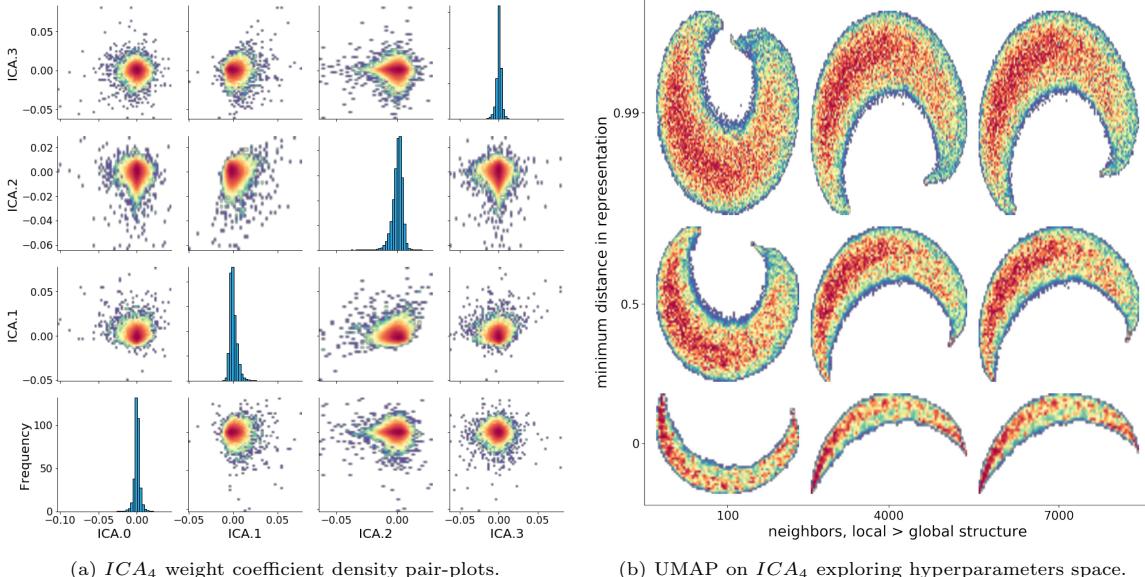


Figure 3

K-means clustering is a very popular clustering algorithm due to its computational efficiency and its many available implementations. K-means is a vector quantization technique, where an observation belongs to the cluster with the nearest mean (cluster center or cluster centroid), serving as a prototype of the cluster. K-means assumes that the variance of the distribution of each feature is spherical, that all features have similar variance, and that the prior probability for all k clusters is the same, i.e., each cluster has roughly equal numbers of observations.

Another family of clustering algorithms is known as hierarchical clustering algorithms, which build nested clusters by merging or splitting them successively. Hierarchical agglomerative clustering merges closer points together, irrespective of the final cluster class balance. The hierarchy of clusters is represented as a dendrogram as in Figure 4d. The final tree has one unique cluster comprising all the samples at the root and clusters with only one sample at the leaves (bottom of the tree). The branch lengths represent the distance between the child clusters. Small gaps connect more similar clusters and big gaps connect more different clusters. The cluster distance is computed as the maximum Euclidean distance distances between all observations of the two clusters.

We choose the agglomerative clustering algorithm because it could be easily adapt to use different distance metrics and directly produce the tree like representation of the whole clusterisation (the dendrogram). It is more computationally intensive than the K-mean, but the latter gives directly a way to judge the correct number of cluster.

We used the reconstruction error threshold to choose the value of k , i.e., the number of clusters. Interested readers could explore other methods like the silhouette score⁸ calculated using the mean intra-cluster distance and the mean nearest-cluster distance for each sample. Figures 4a and 4b show the resulting cluster assignments in the UMAP embedded space when using agglomerative clustering and K-means clustering with $k = 3$ clusters.

⁸see for example "Selecting the number of clusters with silhouette analysis on KMeans clustering" https://scikit-learn.org/stable/auto_examples/cluster/plot_kmeans_silhouette_analysis.html

The clustering methods we described are just a few of the available clustering techniques and methdos may have better results for different datasets (Rui Xu and Wunsch, 2005). For example, when dealing with data containing clusters of similar density interleaved with low density zones (like clusters of data in some dimension with added noise), Density Based Spatial Clustering of Applications with Noise (DBSCAN) may give better results. DBSCAN separates core samples of high density regions and expands clusters from them.

4.4. Conclusion

The result of our analysis is a set of cluster maps such as in Fig.5a. These maps can be directly compared with with models (i.e. surface maximum temperature, roughness or age) and with expert-generated maps that are based on different methods of analysis described in the Introduction. Fig.5b shows the finer clusters defined when splitting the data in 12 clusters. The structure is essentially the same as the 3-cluster case, indicating that this reduction does not appreciably affect the information content gained from the cluster map. The spatial distribution of the red cluster in the north compares favourably with the location of younger volcanic smooth plains plotted in Figure 5d. These plains are widespread on Mercury, but are more concentrated in the north and in the surroundings of the Caloris impact basin (30.5°N 189.8°W) and are most likely of volcanic origin. The smooth plains are younger than the Caloris basin, as indicated by the lower crater densities in those regions (Denevi et al., 2013). The basin floor is filled by geologically distinct plains implanted after the impact at the origin of the basin itself.

The result of our works is the division of the surface in three classes or clusters, that capture some inherent properties of Mercury in the VNIR range as measured by MASCS instrument.

The red clusters (fig.5a) are concentrated outside the hot poles, hinting at the possible role that thermal processing could play in modifying the spectra of the surface.

The representative spectra for those classes differs mostly for the spectral tilt, which is the ratio between longer (redder) and shorter (bluer) wavelengths, as shown in Figure 4c. It is important to stress that this is a general property of Mercury MASCS spectra. They do not show any significant absorption bands, thus it is difficult to match the spectra with known mineral assemblages. This is even more pronounced taking the average of a great number of spectra, washing out single spectral characteristic. This is not always the case, i.e. Bandfield et al. (2000) did found spectral classes on Mars and linked them to terrestrial basaltic and an andesitic composition, roughly related to the Martian dichotomy and older/younger surfaces.

In the case of Mercury, several factors are in play.

Mercury's 3:2 spin-orbit resonance with the Sun creates the characteristic surface temperature distribution shown in Figure 5c, which maps the maximum temperature reached at each point of the surface, based solely on solar irradiation. There are two hot poles around 0° and 180° .

Thermal processing is known to tilt spectra to higher reddening, i.e., steeper spectra (Maturilli et al., 2014).

Solar wind irradiation also concurs to reddens the spectra, even though the literature on this topic focused mostly on asteroids with high-Fe content, while Mercury's surface is extremely low in iron (Nittler et al., 2020). Mercury reflectance spectra are also extremely dark and amorphous carbon has been proposed as darkening agent (Peplowski et al., 2016).

To summarize, dimensionality reduction, manifold learning, and cluster analysis techniques are powerful tools to explore large sets of uncategorized planetary data. The workflow we presented could be easily extended to include additional instrument data, for example including data on the same geographical locations of chemical composition extracted from the X-Ray Spectrometer (XRS) (Nittler et al., 2011, 2020). While we focused only on reflectance data in this study, the technique of data fusion could further promote the discovery of scientifically-interesting patterns in the data,

thus helping to understand complex physical mechanisms that could not be revealed by data coming from a single instrument.

4.5. Acknowledgment

The code and data used in this work is available as python jupyter notebook on the github public repository [MESSENGER-Mercury-Surface-Cassification-Unsupervised_DLR⁹](#) funded under the European Union's Horizon 2020 research and innovation programme grant agreement No 871149. The Working Package "Machine Learning Solutions for Data Analysis and Exploitation in Planetary Science" aims to develop and implement an ambitious sustainable ML toolset, tailored for the needs of the planetary science community by analysing a representative set of scientific cases.

References

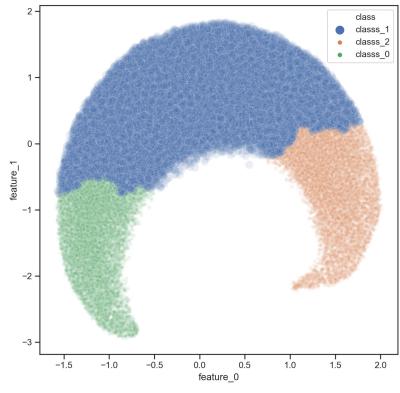
- Bandfield, J., Hamilton, V., and Christensen, P. A Global View of Martian Surface Compositions from MGS-TES. *Science*, 287(March):1626–1630, 2000. doi: 10/fjh6x2.
- Besse, S., Doressoundiram, A., and Benkhoff, J. Spectroscopic properties of explosive volcanism within the Caloris basin with MESSENGER observations. *Journal of Geophysical Research: Planets*, 120(12):2102–2117, December 2015. ISSN 21699097. doi: 10.1002/2015JE004819.
- Bland, P. Crater counting. *Astronomy & Geophysics*, 44(4):4.21–4.21, August 2003. ISSN 1366-8781. doi: 10/dsw66x.
- Coenen, A. and Pearce, A. Understanding UMAP. <https://pair-code.github.io/understanding-umap/>, 2019.
- Denevi, B. W., Robinson, M. S., Blewett, D. T., Domingue, D. L., Head III, J. W., McCoy, T. J., McNutt, R. L., Jr., Murchie, S. L., and Solomon, S. C. MESSENGER global color observations: Implications for the composition and evolution of Mercury's crust. In *Lunar and Planetary Science Conference*, pages 1–2, 2009.
- Denevi, B. W., Ernst, C. M., Meyer, H. M., Robinson, M. S., Murchie, S. L., Whitten, J. L., Head, J. W., Watters, T. R., Solomon, S. C., Ostrach, L. R., Chapman, C. R., Byrne, P. K., Klimczak, C., and Peplowski, P. N. The distribution and origin of smooth plains on Mercury. *Journal of Geophysical Research: Planets*, 118(5):891–907, May 2013. ISSN 21699097. doi: 10.1002/jgre.20075.
- D'Incecco, P., Helbert, J., D'Amore, M., Maturilli, A., Head, J. W. J., Klima, R. L. R., Izenberg, N. R. N., McClintock, W. E. W., Hiesinger, H., and Ferrari, S. Shallow crustal composition of Mercury as revealed by spectral properties and geological units of two impact craters. *Planetary and Space Science*, 119:250–263, 2015. ISSN 00320633. doi: 10.1016/j.pss.2015.10.007.
- Domingue, D. L., D'Amore, M., Ferrari, S., Helbert, J., and Izenberg, N. R. Analysis of the MESSENGER MASCS photometric targets part I: Photometric standardization for examining spectral variability across Mercury's surface. *Icarus*, 319:247–263, February 2019a. ISSN 0019-1035. doi: 10/gh3dp5.

⁹https://github.com/epn-ml/MESSENGER-Mercury-Surface-Cassification-Unsupervised_DLR

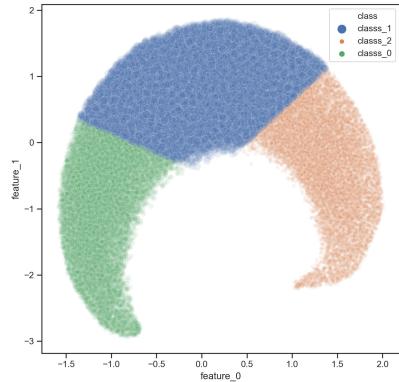
- 379 Domingue, D. L., D'Amore, M., Ferrari, S., Helbert, J., and Izenberg, N. R. Analysis of the MES-
380 SENGER MASCS photometric targets part II: Photometric variability between geomorphological
381 units. *Icarus*, 319:140–246, February 2019b. ISSN 0019-1035. doi: 10/gh3dp6.
- 382 Donoho, D. L. and Grimes, C. Hessian eigenmaps: Locally linear embedding techniques for high-
383 dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, May 2003.
384 ISSN 0027-8424, 1091-6490. doi: 10/cnjc4z.
- 385 E. Vander Kaaden, K., M. McCubbin, F., R. Nittler, L., N. Peplowski, P., Z. Weider, S., A. Frank,
386 E., and J. McCoy, T. Geochemistry, mineralogy, and petrology of boninitic and komatiitic rocks
387 on the mercurian surface: Insights into the mercurian mantle. *Icarus*, 285:155–168, March 2017.
388 ISSN 0019-1035. doi: 10/gg3j22.
- 389 Hamilton, V. E., McSween, H. Y., and Hapke, B. Mineralogy of Martian atmospheric dust inferred
390 from thermal infrared spectra of aerosols. *Journal of Geophysical Research*, 110(E12):1–11, 2005.
391 ISSN 0148-0227. doi: 10/bhsb7j.
- 392 Hastie, T., Tibshirani, R., and Friedman, J. *The Elements of Statistical Learning: Data Mining,
393 Inference, and Prediction, Second Edition*. Springer Series in Statistics. Springer-Verlag, New
394 York, second edition, 2009. ISBN 978-0-387-84857-0. doi: 10.1007/978-0-387-84858-7.
- 395 Helbert, J., Maturilli, A., D'Amore, M., and D'Amore, M. Visible and near-infrared reflectance
396 spectra of thermally processed synthetic sulfides as a potential analog for the hollow forming
397 materials on Mercury. *Earth and Planetary Science Letters*, 369-370:233–238, May 2013. ISSN
398 0012821X. doi: 10/gbddt9.
- 399 Hyvärinen, A. and Oja, E. Independent component analysis: Algorithms and applications. *Neural
400 Networks*, 13(4):411–430, June 2000. ISSN 0893-6080. doi: 10/cx35gq.
- 401 Kerr, R. A. Who Can Read the Martian Clock? *Science*, 312(5777):1132–1133, May 2006. ISSN
402 0036-8075, 1095-9203. doi: 10/b6v8tt.
- 403 Lee, J. A. and Verleysen, M. *Nonlinear Dimensionality Reduction*. Information Science and Statistics.
404 Springer-Verlag, New York, 2007. ISBN 978-0-387-39350-6. doi: 10.1007/978-0-387-39351-3.
- 405 Maturilli, A., Helbert, J., St. John, J., Head, J., Vaughan, W., D'Amore, M., Gottschalk, M., and
406 Ferrari, S. Komatiites as Mercury surface analogues: Spectral measurements at PEL. *Earth and
407 Planetary Science Letters*, 398, 2014. ISSN 0012821X. doi: 10/gg3jwp.
- 408 McClintock, W. E. and Lankton, M. R. The mercury atmospheric and surface composition spectrometer
409 for the MESSENGER mission. *Space Science Reviews*, 131(1-4):481–521, 2007. ISSN
410 00386308. doi: 10/btc6f6.
- 411 McInnes, L. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction —
412 umap 0.5 documentation. <https://umap-learn.readthedocs.io/en/latest/index.html>, 2018.
- 413 McInnes, L., Healy, J., and Melville, J. UMAP: Uniform Manifold Approximation and Projection
414 for Dimension Reduction. *arXiv:1802.03426 [cs, stat]*, September 2020.
- 415 Meslin, P.-Y., Gasnault, O., Forni, O., Schröder, S., Cousin, A., Berger, G., Clegg, S. M., Lasue, J.,
416 Maurice, S., Sautter, V., Mouélic, S. L., Wiens, R. C., Fabre, C., Goetz, W., Bish, D., Mangold,
417 N., Ehlmann, B., Lanza, N., Harri, A.-M., Anderson, R., Rampe, E., McConnochie, T. H., Pinet,
418 P., Blaney, D., Léveillé, R., Archer, D., Barracough, B., Bender, S., Blake, D., Blank, J. G.,

- 419 Bridges, N., Clark, B. C., DeFlores, L., Delapp, D., Dromart, G., Dyar, M. D., Fisk, M., Gondet,
 420 B., Grotzinger, J., Herkenhoff, K., Johnson, J., Lacour, J.-L., Langevin, Y., Leshin, L., Lewin,
 421 E., Madsen, M. B., Melikechi, N., Mezzacappa, A., Mischna, M. A., Moores, J. E., Newsom, H.,
 422 Ollila, A., Perez, R., Renno, N., Sirven, J.-B., Tokar, R., de la Torre, M., d'Uston, L., Vaniman,
 423 D., Yingst, A., and Team, M. S. Soil Diversity and Hydration as Observed by ChemCam at Gale
 424 Crater, Mars. *Science*, 341(6153), September 2013. ISSN 0036-8075, 1095-9203. doi: 10/f3sdqx.
- 425 Namur, O. and Charlier, B. Silicate mineralogy at the surface of Mercury. *Nature Geoscience*, 10
 426 (1):9–13, January 2017. ISSN 1752-0908. doi: 10/f9r3qp.
- 427 NASA. PDS: PDS3 Standards Reference. <https://pds.nasa.gov/datastandards/pds3/standards/>,
 428 2009.
- 429 Neish, C. D., Blewett, D. T., Harmon, J. K., Coman, E. I., Cahill, J. T., and Ernst, C. M. A
 430 comparison of rayed craters on the moon and mercury. *Journal of Geophysical Research E: Planets*,
 431 118(10):2247–2261, 2013. ISSN 01480227. doi: 10/gg3j2w.
- 432 Nittler, L. R., Starr, R. D., Weider, S. Z., McCoy, T. J., Boynton, W. V., Ebel, D. S., Ernst,
 433 C. M., Evans, L. G., Goldsten, J. O., Hamara, D. K., Lawrence, D. J., McNutt, R. L., Schlemm,
 434 C. E., Solomon, S. C., and Sprague, a. L. The Major-Element Composition of Mercury's Surface
 435 from MESSENGER X-ray Spectrometry. *Science*, 333(6051):1847–1850, September 2011. ISSN
 436 0036-8075. doi: 10/bps3b8.
- 437 Nittler, L. R., Frank, E. A., Weider, S. Z., Crapster-Pregont, E., Vorburger, A., Starr, R. D., and
 438 Solomon, S. C. Global major-element maps of mercury from four years of MESSENGER X-Ray
 439 Spectrometer observations. *Icarus*, page 113716, February 2020. ISSN 0019-1035. doi: 10/ggm5sv.
- 440 Padovan, S., Tosi, N., Plesa, A.-C., and Ruedas, T. Impact-induced changes in source depth and
 441 volume of magmatism on Mercury and their observational signatures. *Nature Communications*, 8
 442 (1):1945, December 2017. ISSN 2041-1723. doi: 10/gcn9p2.
- 443 Peplowski, P. N., Klima, R. L., Lawrence, D. J., Ernst, C. M., Denevi, B. W., Frank, E. A., Goldsten,
 444 J. O., Murchie, S. L., Nittler, L. R., and Solomon, S. C. Remote sensing evidence for an ancient
 445 carbon-bearing crust on Mercury. *Nature Geoscience*, 9(4):273–276, March 2016. ISSN 1752-0894.
 446 doi: 10.1038/ngeo2669.
- 447 Roweis, S. T. and Saul, L. K. Nonlinear Dimensionality Reduction by Locally Linear Embedding.
 448 *Science*, 290(5500):2323–2326, December 2000. ISSN 0036-8075, 1095-9203. doi: 10/cbws2r.
- 449 Rui Xu and Wunsch, D. Survey of clustering algorithms. *IEEE Transactions on Neural Networks*,
 450 16(3):645–678, May 2005. ISSN 1941-0093. doi: 10/c3pfgf.
- 451 Sprague, A., Donaldson Hanna, K., Kozlowski, R., Helbert, J., Maturilli, A., Warell, J., and Hora,
 452 J. Spectral emissivity measurements of Mercury's surface indicate Mg- and Ca-rich mineralogy,
 453 K-spar, Na-rich plagioclase, rutile, with possible perovskite, and garnet. *Planetary and Space
 454 Science*, 57(3):364–383, March 2009. ISSN 00320633. doi: 10/fbd9jq.
- 455 Tenenbaum, J. B., de Silva, V., and Langford, J. C. A Global Geometric Framework for Nonlinear
 456 Dimensionality Reduction. *Science*, 290(5500):2319–2323, December 2000. ISSN 0036-8075, 1095-
 457 9203. doi: 10/cz8wgk.
- 458 van der Maaten, L. and Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning
 459 Research*, 9(86):2579–2605, 2008. ISSN 1533-7928.

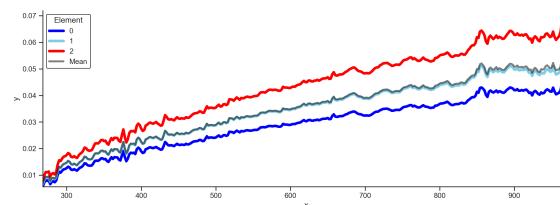
- ⁴⁶⁰ Vasavada, a. Near-Surface Temperatures on Mercury and the Moon and the Stability of Polar Ice
⁴⁶¹ Deposits. *Icarus*, 141(2):179–193, October 1999. ISSN 00191035. doi: 10/b9fhjd.
- ⁴⁶² Vilas, F., Domingue, D. L., Helbert, J., D’Amore, M., Maturilli, A., Klima, R. L., Stockstill-Cahill,
⁴⁶³ K. R., Murchie, S. L., Izenberg, N. R., Blewett, D. T., Vaughan, W. M., and Head, J. W.
⁴⁶⁴ Mineralogical indicators of Mercury’s hollows composition in MESSENGER color observations.
⁴⁶⁵ *Geophysical Research Letters*, 43(4):1450–1456, February 2016. ISSN 00948276. doi: 10/gg3j2v.



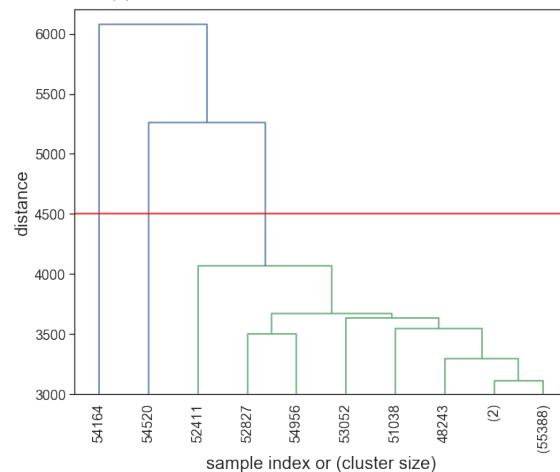
(a) Agglomerative Clustering labels on UMAP features.



(b) K-Means labels on UMAP features.

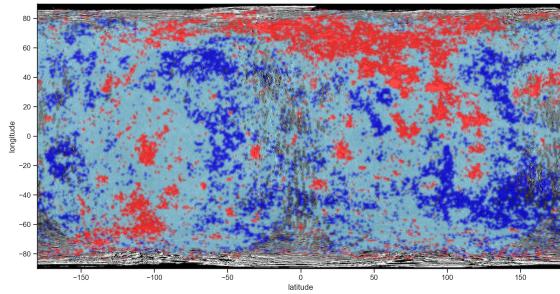


(c) Agglomerative Clustering centroids.

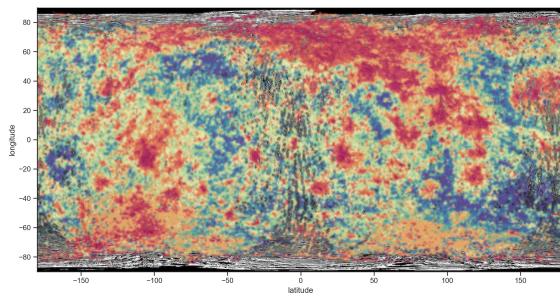


(d) Agglomerative Clustering dendrogram.

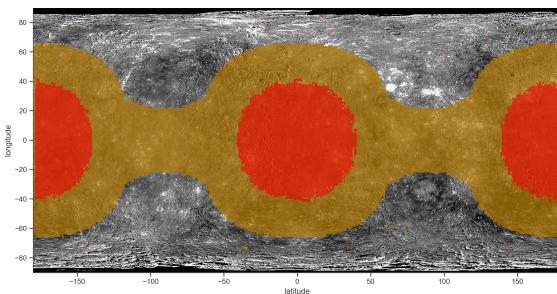
Figure 4



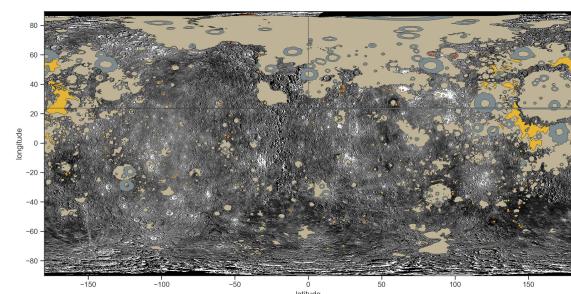
(a) Agglomerative Clustering 3 classes



(b) Agglomerative Clustering 12 Classes



(c) Maximum surface temperature: red > 690 K, brown > 550 K, adapted from [Vasavada \(1999\)](#)



(d) Mercury smooth plains, adapted from [Denevi et al. \(2013\)](#)

Figure 5