The Planetary Science Journal, 3:231 (13pp), 2022 October https://doi.org/10.3847/PSJ/ac8f43

Mapping Lunar Swirls with Machine Learning: The Application of Unsupervised and Supervised Image Classification Algorithms in Reiner Gamma and Mare Ingenii

Frank C. Chuang, Matthew D. Richardson, John R. Weirich, Amanda A. Sickafoose, and Deborah L. Domingue
Planetary Science Institute, 1700 E Fort Lowell Road, Suite 106, Tucson, AZ 85719 USA; chuang@psi.edu
Received 2022 May 20; revised 2022 August 5; accepted 2022 September 2; published 2022 October 13

Abstract

Lunar swirls are recognized as broad, bright albedo features in various regions of the Moon. These features are often separated by dark off-swirl lanes or terminate against the dark background, such as lunar maria. Prior mapping of swirls has been done primarily by albedo contrast, which is prone to subjectivity. Closer examination of on-swirl areas shows that they are not uniform, making the boundary between on- and off-swirl difficult to map with certainty. We have applied machine learning techniques to address these issues by identifying the number of swirl units and then mapping them based on actual reflectance, or I/F data. Using LROC NAC paired stereo images that are converted to I/F reflectance at a range of incidence angles, we applied both unsupervised K-means clustering and supervised Maximum Likelihood Classification algorithms to classify and map portions of lunar swirls in Reiner Gamma and Mare Ingenii. Results show that the classification maps are a reasonable match to the representative albedos for the two study regions. A third transitionary swirl unit, termed diffuse-swirl, is present in both the maps and the cumulative distribution plots of the reflectance values. Overall, we find that the use of both algorithms provides independent confirmation of both the number and location of these units and their interrelation. More importantly, the algorithms remove mapping subjectivity by using quantitative information. The data and the statistics generated from the maps also have value in future studies by placing limits for categorizing swirl units in different regions on the Moon.

Unified Astronomy Thesaurus concepts: Solar system planets (1260); Solar system terrestrial planets (797); The Moon (1692); Lunar science (972)

1. Introduction

Lunar swirls are large, distinct albedo features that are present within various regions on the Moon, both near- and farside (El-Baz 1972; Schultz & Srnka 1980; Bell & Hawke 1981; Blewett et al. 2007; Kramer et al. 2011a and the references therein). These features are defined as broad, bright, curvilinear areas separated by darker off-swirl lanes or terminating against the darker surrounding background. They are particularly distinct within dark lunar maria, but have also been observed on light-toned highlands terrain. Lunar swirls are generally associated with local crustal anomalies (Hood et al. 1979; Hood & Schubert 1980), though not all magnetic anomalies have swirls. The origin of the magnetic anomalies has been debated, and it is not clear if they were part of an ancient active dynamo (Fuller 1974; Fuller & Cisowski 1987; Garrick-Bethell et al. 2009). One of the early proposed formation processes for these features suggests that the magnetic anomalies shield the surface from the solar wind ions, thus reducing the surface darkness and making it appear brighter (Hood et al. 1979; Hood & Schubert 1980; Hood & Williams 1989). Other processes that have been proposed for the possible formation and (or) evolution of lunar swirls include dust levitation and transport through electrostatic charging (Garrick-Bethell et al. 2011; Pieters & Garrick-Bethell 2015), sorting and relocation of dark magnetic grains (Pieters & Garrick-Bethell 2015), and surface scouring by cometary impacts (Schultz & Srnka 1980; Pinet et al. 2000; Starukhina & Shkuratov 2004; Syal & Schultz 2015).

From past studies up to the present time, the identification and mapping of swirls have been done primarily through observation. Because the swirls are defined by albedo contrast in orbital images, one of the difficulties is identifying the distinct boundaries between the bright on-swirl and dark off-swirl regions (dark lanes and (or) background terrain). Moreover, visual inspection of the on-swirl regions shows that they are not uniform in albedo. There are less-bright areas internally within a swirl that are not as dark as the dark lanes or surrounding background. Such a unit may represent a possible transition region. As prior mapping was not based solely on data values, it is prone to subjectivity (Kramer et al. 2011a; Blewett et al. 2011; Domingue et al. 2022; Denevi et al. 2016; Domingue et al. 2021; Blewett et al. 2021). In this study, we have applied machine learning techniques to identify the number of swirl units and to map them based on surface reflectance data. This quantitative analysis provides measurable criteria and repeatable classification for the different types and numbers of swirl units, eliminating the subjective bias. We note that the term albedo hereafter refers to the reflectance in context with the reflectance values of the scene, which is taken under similar illumination and viewing geometries.

2. Background

Machine learning is a rapidly growing area of semi- to fully automated data analysis that is applied to very large data sets in terms of storage size and sheer volume (Sarker 2021). Such analyses have been used in business, health, industrial technology, cybersecurity, military, natural, and applied sciences, and many other fields. Machine learning is generally

one of four principal types: supervised, unsupervised, semi-supervised, and reinforcement (Sarker 2021; Kerner et al. 2022). Supervised follows a task-driven approach whereby labeled training data are fed into mathematical algorithms or models to predictively label and classify all remaining input data. Unsupervised is a data-driven approach that takes all input data as unlabeled and attempts to extract trends or groupings within the data set without prior user knowledge. Semisupervised is similar to supervised, with the exception that models are run on both labeled and unlabeled data. Reinforcement involves models based on a trial-and-error approach in an interactive environment with input from its actions and experiences. In this study, we focus on both supervised and unsupervised as they are simpler to apply, readily available, and not strictly model driven. We refer to the summary on machine learning by Sarker (2021) for further details on the other two types. Within the planetary sciences, the use of machine learning has grown in response to the exponential growth of planetary data from spacecraft missions over the last several decades, as well as Earth-based observatories and laboratories (Azari et al. 2021; Kerner et al. 2022).

One of the key features of supervised learning is classification, which involves predictive modeling. Classification has been used for many decades in the Earth remote-sensing community and is generally referred to as image classification (Richards & Jia 1999). One of the early uses of image classification was identifying various types of land from Earth-orbiting multispectral data such as from the United States LANDSAT program (Wulder et al. 2019 and references therein). Feature "class" training areas identified by a user, such as forest, agricultural, wetlands, water, or pavement, are fed into a predictive mathematical algorithm to quickly classify scenes with many hundreds of thousands to millions of pixels.

Unsupervised learning typically involves cluster analysis, and in the case of image data, it is a means by which pixels are assigned to classes without prior user knowledge about class types. The number of clustered classes and the location of class pixels then need to be evaluated by the user in relation to any known properties for each class. Unsupervised learning is often performed as a step prior to supervised classification. Some of the more common unsupervised clustering algorithms include K-means (Asada et al. 2010; Anderson & Bell 2013; Collier et al. 2020; Rammelkamp et al. 2021; Kerner et al. 2022), Density-Based Spatial Clustering of Applications with Noise, Gaussian Mixture Models, and Agglomerative Hierarchal Clustering (Sarker 2021).

In this study, we have applied both supervised and unsupervised algorithms to surface reflectance data for two regions with lunar swirls, one on the lunar nearside and one on the farside. We believe the use of both algorithms provides a more robust analysis and mapping of lunar swirls than a single algorithm alone. The use of both algorithms also provides a check to determine whether the mapping results are valid, particularly from unsupervised to supervised classification and for specific areas or features within a region of interest.

3. Data Products and Study Areas

The reflectance and corresponding topography data used in this study are derived from Lunar Reconnaissance Orbiter Camera (LROC) Narrow-Angle Camera (NAC) images, which have resolutions up to $\sim$0.5 m pixel^{-1}. The topography that is used to align the reflectance data was produced using stereophotoclinometry (SPC). SPC combines stereo positioning and photoclinometry to generate a digital terrain model (DTM) from LROC NAC stereo images (Gaskell et al. 2008; Palmer et al. 2016). The DTMs are generated from SPC maplets, which are fundamentally 99×99 pixel representations of the topography. Each maplet pixel contains the height of the surface, and the central pixel of the entire maplet is defined from the LROC NAC images by stereo positioning. After the central pixel is located, the remaining maplet pixels are defined using photoclinometry, where the brightness of each pixel is determined by the albedo and slope of the surface from the corresponding NAC images at different incidence angles with respect to the surface normal. Images with low overall incidence emphasize the albedo contribution to the brightness, while high overall incidence emphasizes the topographic contribution to the brightness. The height of the central pixel and the slopes of the remaining maplet pixels are then integrated to produce a topographic surface. A single SPC DTM product is generally a mosaic of thousands to tens of thousands of overlapping maplets. The uncertainty in the height or vertical position of a maplet pixel is about the same as the horizontal resolution of the DTM.

In this study, we examined swirl regions that contain both on- and off-swirl areas. In selecting the representative albedo images, full or maximum coverage of areas both on- and off-swirl were required. In almost all instances, the coverage involved the use of paired LROC NAC stereo images, i.e., from the left and right (L/R) cameras. L/R images of a given observation have nearly the same incidence angle because they are collected at the same ephemeris time with a small spatial overlap. The L/R images were combined together into a single image, and if the centers of two NAC pixels fell within a single DTM pixel, only the left image from the L/R pair was used. Portion(s) of any image with no data were treated as zero values. When sufficient NAC coverage was identified for the study regions, image pairs with low, moderate, and high incidence-angle values were chosen. Using a range of angles allows for a less-biased photometric analysis and mapping of the swirls. The raw DN values in each image were then converted to reflectance, or I/F, by applying a scaling factor of 3.051 850 947 599 72e^{-5} from the image header, which is the same for all the images. Last, the NAC images were aligned and sampled to the same resolution as the SPC DTM. Further details of this process are described in Domingue et al. (2018).

We have selected parts of two lunar swirl regions on which to apply supervised and unsupervised classification algorithms. The first region is within Reiner Gamma (center: 7.4° N, 301° E), the archetypical example of lunar swirls within basaltic maria, near the western edge of Oceanus Procellarum on the lunar nearside (Figure 1(a)). The study region is slightly south of the central eye of the swirl containing bright on-swirl areas separated by dark lanes. The on-swirl area gradually reduces in albedo from the center as it transitions to off-swirl. We refer to these transition areas as diffuse-swirl. The highlands located along the edge of Oceanus Procellarum and the greater western Procellarum basin (Wilhelms et al. 1987) are $\sim$180 km from this study region, which may consist of a mixture of highlands and mare material. Results from early spectral mixing models support this possibility (Bell & Hawke 1981). Highlands material predates the mare, but the volcanic infill may be thin, with highlands material not far below the surface. The western Procellarum basin is also a region with medium to high

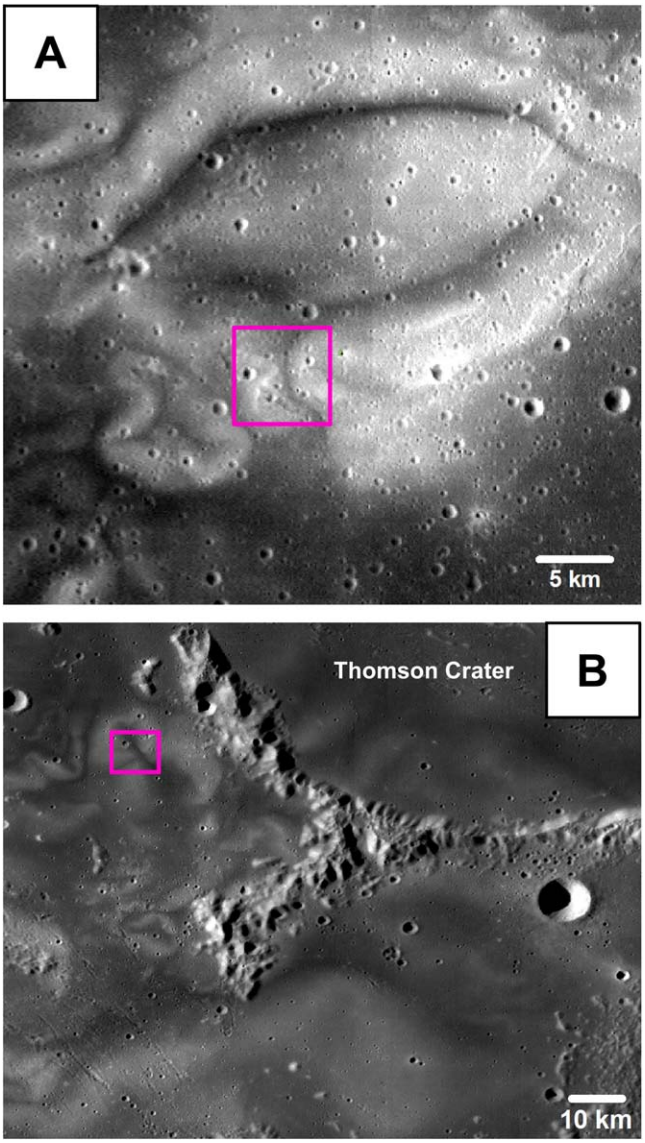

Figure 1. (A) Central portion of the Reiner Gamma swirls, with the study region highlighted in pink. The center of the entire swirl region is located at ~7.4° N, 301° E on the lunar nearside. (B) Portion of the swirls in Mare Ingenii, with the study region highlighted in pink. The center of the entire swirl region is located at ~33.24° S, 164.83° E on the lunar farside. The background image in both panels is the 100 m pixel^{-1} LROC WAC mosaic. North is up and west is to the left.

percentages of TiO_2 (>5 wt%) compared to older maria in other parts of the Moon, indicating shallow mantle depths and partial melting of late-stage cumulates in a magma ocean (Sato et al. 2017). Other spectral, photometric, and radar studies of Reiner Gamma from both lunar missions and Earth-based telescopic observations have been performed (Pinet et al. 2000; Kreslavsky & Shkuratov 2003; Campbell et al. 2006; Chevrel et al. 2006; Kaydash et al. 2009), and a summary of these can be found in Kramer et al. (2011a). For Reiner Gamma, a total of 29 LROC NAC images covered the study region. Two L/R pairs and another single image were used in the image classification algorithms (see the Data Processing section and Table 1).

The second region is within Pre-Nectarian Mare Ingenii (center: 33.2° S, 164.8° E) on the lunar farside (Wilhelms et al. 1987). This region has prominent examples of swirls within one of several nested impact craters that are infilled by dark basaltic mare (Figure 1(b); Kramer et al. 2011b). The study region is located in the eastern half of the largest crater, approximately eight kilometers from the W-SW rim of Thomson Crater. Similar to Reiner Gamma, the bright on-swirl areas are separated by dark off-swirl lanes with intermediate diffuse-swirl areas in between. The proximity of the Thomson Crater rim suggests possible mixing of impact ejecta and mare materials. From the WAC global mosaic, there is no indication of major surface modification or morphological features formed since the mare infill. Mare Ingenii is antipodal to the large Imbrium impact, which resulted in the formation of furrowed terrain along the basin margins (Schultz & Gault 1975; Stuart-Alexander 1978; Richmond et al. 2005) and the possible convergence of Imbrium impact ejecta deposits (Moore et al. 1974). Impact ejecta deposits from O'Day Crater superposes the mare along the western edge of Mare Ingenii, but the study region is well east of the farthest extent of the deposits. For Mare Ingenii, a total of 39 LROC NAC images covered the study region. Three L/R pairs were used in the image classification algorithms (see the Data Processing section and Table 1).

4. Data Processing and Classification Algorithms

We used Environmental Systems Research Institute ArcGIS for Desktop® 10.4 (ArcGIS) software to import and analyze the image data. ArcGIS is a commercial Geographic Information Systems package with a large and robust set of tools for image processing, mapping, 3D viewing, and geostatistical analysis of both raster- and vector-based data. Included with the tools are several image classification algorithms. In this study, we apply one of the most common supervised classification algorithms to the reflectance data, the maximum likelihood classification (MLC). In addition, we apply an unsupervised K-means clustering classification algorithm that is available through the open-source Python Sci-Kit Learn Library software (Pedregosa et al. 2011).

Reflectance data were produced using the Integrated Systems for Imaging Spectrometers (ISIS) software as 32 bit cube (.cub) files for each study region. The ISIS .cub file contains three bands of reflectance data (a representative albedo image), each at a different incidence angle. For Reiner Gamma, the incidence angles are 18°, 42°, and 66° (Figure 2(a)). For the Mare Ingenii study region, the incidence angles are 33°, 48°, and 63° (Figure 2(b)). Table 1 lists the individual NAC images used for each representative albedo image, as well as the emission and phase angles. After opening the cube file in ArcGIS, each band was saved as a 32-bit floating point band sequential (.bsq) format file and then warped using control points to closely match a 100 m pixel^{-1} LROC WAC global basemap in equidistant cylindrical map projection. Last, all three bands for each study region were then combined into a single three-band GeoTiff format (.tif) file for image classification.

In order to avoid areas with extreme reflectance due to positive- or negative-sloping features, we masked these by first manually tracing the margins of impact craters down to 50 m in diameter and most large features such as troughs, hills, and pits that are hundreds to thousands of meters across. In ArcGIS, both circular and elliptical drawing tools were used to trace these features to their maximum visible extent (i.e., outermost margins of impact crater rims, troughs, etc.) and were then

Table 1
LROC NAC Images Used and General Statistical Information for the Reflectance Data in the Reiner Gamma and Mare Ingenii Study Regions

Reflectance Data (Incidence Angle)	Incidence Angle[a]	Emission angle[a]	Phase Angle[a]	Max Value	mean ± standard Deviation	LROC NAC Image IDs
Reiner Gamma (18°)	17.83	11.82	29.27	0.113 89	0.070 26 ± 0.00672	M1149901166L, M1149901166R
Reiner Gamma (42°)	42.18	41.25	5.98	0.168 44	0.107 35 ± 0.00822	M1145205753L
Reiner Gamma (66°)	66.04	3.10	62.95	0.059 93	0.025 37 ± 0.00338	M1112203387L, M1112203387R
Mare Ingenii (33°)	32.94	13.95	36.49	0.106 93	0.051 51 ± 0.01442	M1163750779L, M1163750779R
Mare Ingenii (48°)	48.11	1.43	48.36	0.082 21	0.033 79 ± 0.01359	M105802305L, M105802305R
Mare Ingenii (63°)	62.90	1.43	63.18	0.066 72	0.020 31 ± 0.01113	M1128423456L, M1128423456R

Note.
[a] Average values for data with stereo image pairs.

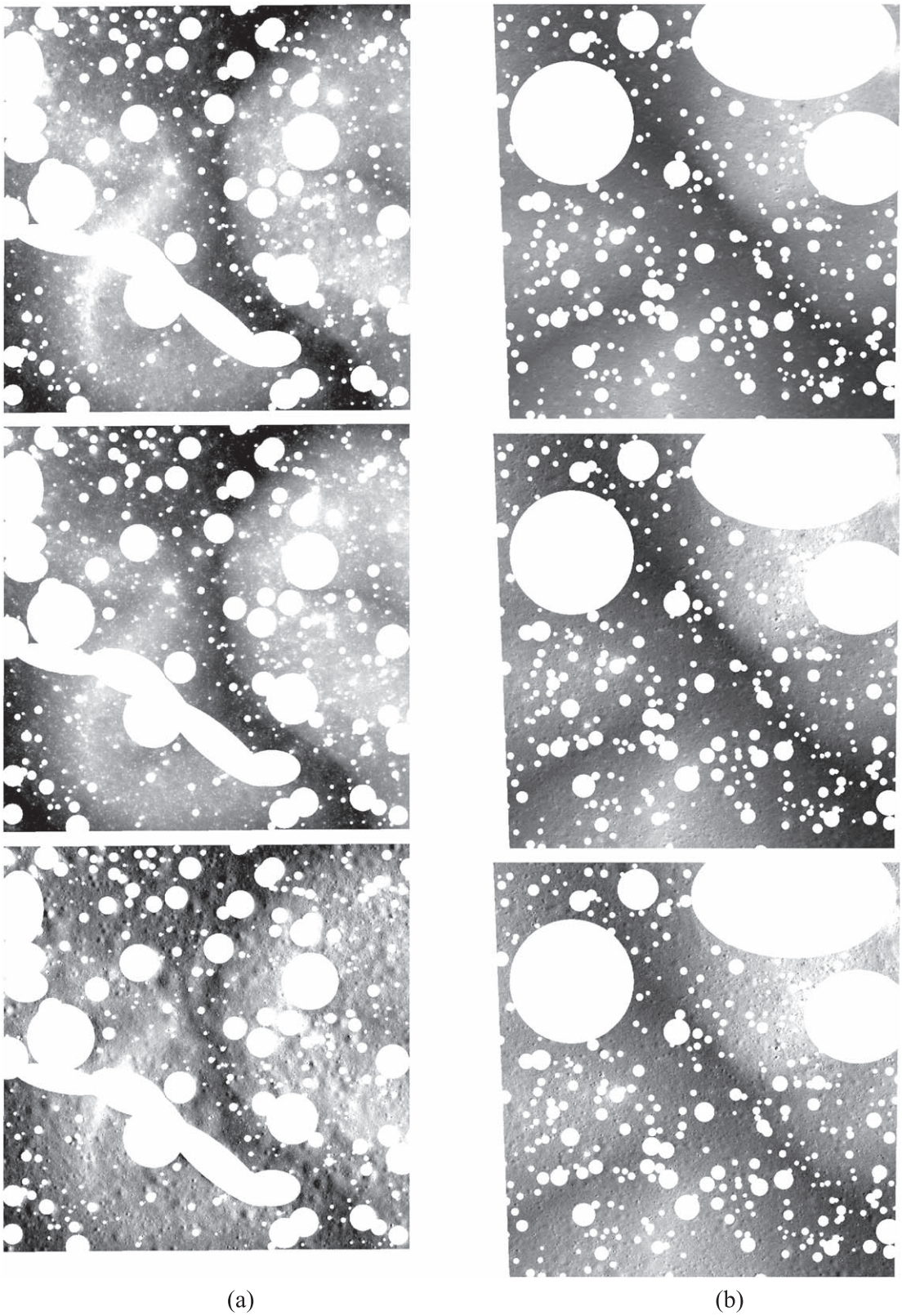

Figure 2. (A). Surface reflectance images of the Reiner Gamma study region at 18° incidence angle (top), 42° incidence angle (middle), and 66° incidence angle (bottom). Reflectance was derived by applying a conversion constant to raw DN values in the LROC NAC left and right stereo pairs. See text for details. White areas are impact craters and large geologic features that have been masked. Images have an applied contrast stretch of 2.5 standard deviations of the mean. These images are in the boxed area shown in Figure 1(a). (B). Surface reflectance images of the Mare Ingenii study region at 33° incidence angle (top), 48° incidence angle (middle), and 63° incidence angle (bottom). Reflectance was derived by applying a conversion constant to raw DN values in the LROC NAC left and right stereo pairs. See text for details. White areas are impact craters and large geologic features that have been masked. Images have an applied contrast stretch of 2.5 standard deviations of the mean. These images are in the boxed area shown in Figure 1(b).

saved as a polygon shapefile. We note that impact craters smaller than 50 m were not included as their relief and size do not produce significant brightness changes relative to the entire study region. Using the shapefile, the reflectance data within these features were nulled for each study region. As the K-means clustering algorithm is performed outside of ArcGIS, each of the crater and feature-masked reflectance images were exported as ASCII text files as 2D arrays with header information. Data within each text file, excluding any header information, were then extracted and reformatted using a Python-based script with the spatial and reflectance values for each pixel stored as records in a Pandas DataFrame (McKinney 2010; The Pandas Development Team 2020).

Basic statistical information including the means and standard deviations for the two regions with craters and features masked is listed in Table 1. The middle-incidence angle (42°) for Reiner Gamma has the highest absolute reflectance of the three incidence-angle images, which is expected given that it was acquired at a near-opposition geometry (phase angle close to zero). This low phase opposition is a well-known effect for the Moon (Gehrels et al. 1964; Oetking 1966) and causes increased brightness across the entire image. Despite the increased reflectance, we demonstrate below that this does not have a major effect on the image classification results.

In this study, the MLC method is available as part of the Spatial Analyst Tools extension package in ArcGIS. Because MLC is a supervised classification, training areas for each swirl class, hereafter referred to as units, were digitized based on the representative albedo image that provided the most inherent contrast, which is typically the lowest incidence angle (Figures 3(a) and (b)), along with some additional user contrast-stretching. Three swirl units were used: on-swirl, off-swirl, and diffuse-swirl. On-swirl areas typically have higher representative albedo, which generally defines the majority of the observed bright areas within a swirl. Off-swirl are generally areas with the lowest representative albedo and appear dark compared to on-swirl areas. Diffuse-swirl is a third unit that represents a transition from on- to off-swirl with an observed brightness that is between the two end-members. Selecting areas for diffuse-swirl is more difficult than for on- and off-swirl. However, we believe these transition zones are clearly present in defining a third unit. Three units are shown to be the optimum number from the K-means classification, and we discuss this further in this section. The total area and pixel counts for on-, off-, and diffuse-swirl training areas for both study regions are shown in Table 2.

Reflectance values within the MLC training areas for each unit were extracted and evaluated. Information including the mean, number of data points, and the variance-covariance matrix for each unit were calculated and stored in an ArcGIS signature file prior to applying the MLC algorithm. The matrix is one of the key parts of any multiband remote-sensing data analysis, where the scatter among the pixel positions in multispectral space is quantified relative to the expected mean position (Richards & Jia 1999). The degree of scatter from training-area pixel values forms the basis for unit clouds that are used to evaluate nontraining-area pixels, including their unit assignment and probability. Using the signature file and multiband reflectance images, output from the MLC algorithm consists of two products. The first product is a classification map, where every pixel in N-dimensional spectral space is assigned to a unit, and the second product is a probability map of how well each pixel is assigned to the unit. The classification map has the same number of units as those defined in the training data. Units can be weighted prior to running the algorithm, but none were used, and thus, all were treated equally. The output probability map is based on a nonweighted scale from a high of one to a low of 14. The highest rank can be thought of as a 99% probability of being assigned the correct unit, and the lowest rank as an 86% probability of being the

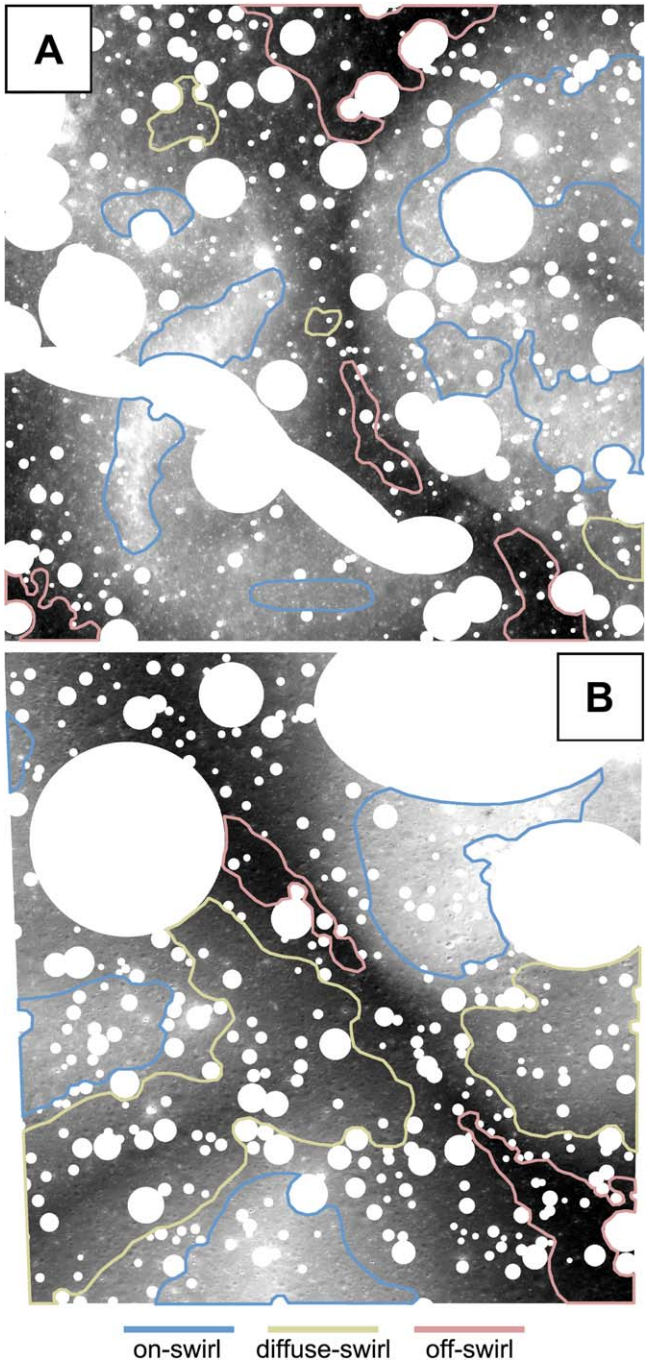

Figure 3. Locations of training areas in (A) Reiner Gamma and (B) Mare Ingenii for the three units of lunar swirls used in the MLC algorithm. On-swirl areas are outlined in blue, off-swirl areas in pink, and diffuse-swirl areas in tan. For details on the surface area and pixel counts of the training areas, see Table 2. The background image is the albedo at 18° and 48° incidence angles for Reiner Gamma and Mare Ingenii, respectively.

Table 2
Surface Area and Pixel Counts of MLC Algorithm Training Areas for On-, Off-, and Diffuse-swirl Units

Unit Type	Surface Area (m^2)	Pixel Count (% of Total Study Area)
Reiner Gamma on-swirl	6,849,994	942,174 (21.6%)
Reiner Gamma off-swirl	2,716,450	401,842 (9.2%)
Reiner Gamma diffuse-swirl	655,654	96,990 (2.2%)
Mare Ingenii on-swirl	8,184,677	1,210,751 (23.6%)
Mare Ingenii off-swirl	2,538,982	375,589 (7.3%)
Mare Ingenii diffuse-swirl	12,174,963	1,801,030 (35.1%)

correct unit. A complete treatment of the MLC probability function(s), decision rules, and the relation to the Bayes theorem is described in Richards & Jia (1999).

Application of the K-means algorithm is straightforward, fully automated, and only requires a defined number of cluster units (k). To execute the algorithm, the user passes two required arguments, the data to be used, including the portion of the DataFrame with only the reflectance values and the parameter k that sets the number of cluster units. The algorithm only uses the reflectance values as this is the parameter space in which the clustering will be evaluated. The algorithm initiates the process of identifying clusters by randomly selecting k number of cluster coordinates within the multidimensional parameter space, each dimension corresponding to a single reflectance band. This initial selection during the first iteration will serve as the first possible cluster centroids. The Euclidean distance between the first data point and each centroid location is then calculated, with the data point being assigned to its closest centroid. This process is applied to each data point in the entire data set. After each data point has been assigned, the location of each centroid with its current cluster points is updated and the data set is reevaluated based on the new centroid locations. This process is iterated until the positions of the centroids stabilize. The final product is a list of cluster assignments that is equivalent to the number and the order of records in the DataFrame. The assigned unit at each point location can then be re-output as a 2D array ASCII text file.

A variety of methods can be used to calculate the optimal number of clusters, but in this study, we use the common elbow method. This method first calculates the sum of the squared distances (SoSD) between each data point relative to its cluster centroid for all data points within a single cluster. The SoSD calculation is repeated for other clusters present within the parameter space. All SoSD calculations within each cluster are then summed to obtain the within-cluster sum of squares (WCSS). The elbow method plots the WCSS against k clusters being evaluated, and the optimal number occurs at an inflection elbow point when the WCSS no longer changes significantly at a given k and beyond. The selection of the elbow point is automated via the use of the Kneedle algorithm (Satopaa et al. 2011).

5. Results

For both study regions, the MLC classification maps are generally a good match to what is observed in the representative albedo images for all three units of on-, off-, and diffuse-swirl

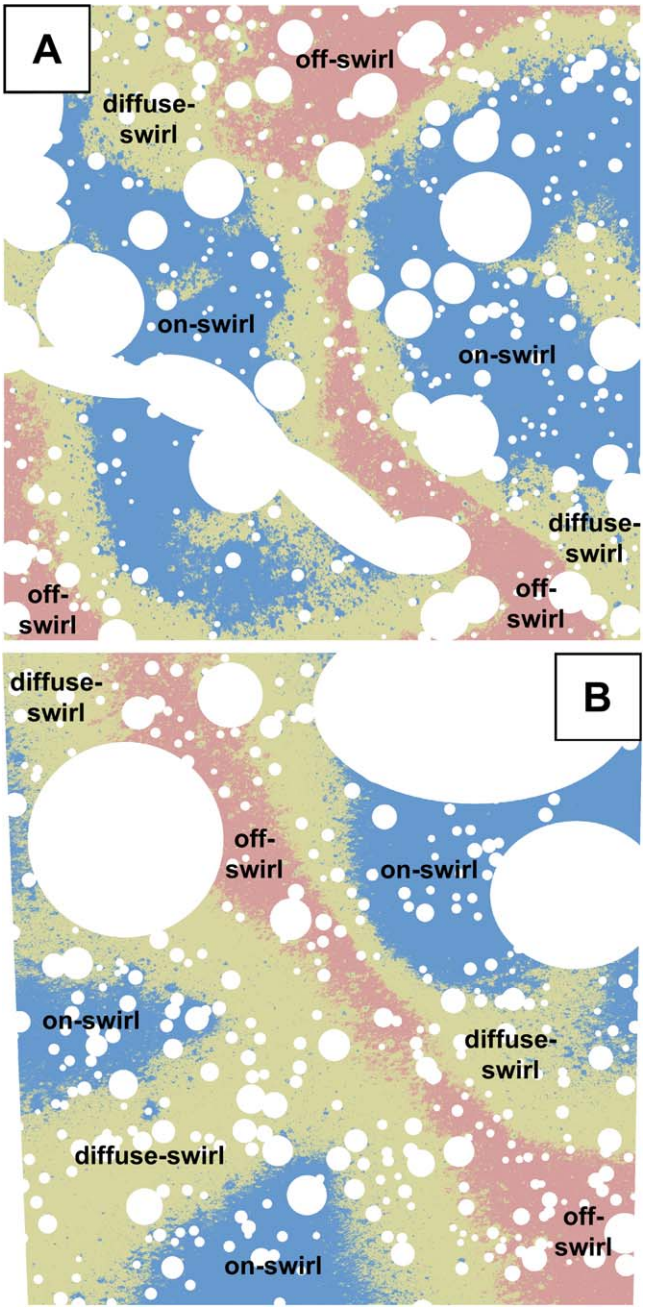

Figure 4. Classification maps using the MLC algorithm for the (A) Reiner Gamma and (B) Mare Ingenii study regions. Multiband albedo data including all three incidence angles were used in the algorithm for both study regions. On-swirl areas are shown in blue, off-swirl in pink, and diffuse-swirl in tan. For details on the surface area and pixel counts of each unit type, see Table 3.

(Figures 4(a) and (b)). On-swirl areas tend to be surrounded by diffuse-swirl, indicating the presence of a third unit that is distinct from the other two endmember units. Total areas, pixel counts, and reflectance statistics for each unit are listed in Table 3. From the probability maps, most areas with lower probability (<91%) are associated with anomalously high reflectance, typically areas with greater relief or features such as the edges of crater rims or ejecta deposits (Figures 5(a) and (b)). In other cases, low probabilities may be associated with areas that contain reflectance values that either do not fall within or are near the ends of a particular unit range from the training data. Despite the low probabilities, the algorithm still assigned

Table 3
Surface Area, Pixel Counts, and Reflectance Statistics for Both MLC and K-Means Classification Maps of On-, Off-, and Diffuse-swirl Units with Craters and Geologic Features Masked

Unit and Image Incidence Angle (Classification Type)	Surface Area (km^2)	Pixel Count (% of Total Study Area)	Reflectance Min[a]	Reflectance Max	Reflectance Median	Reflectance Mean	Reflectance std. dev.
			Reiner Gamma				
on-swirl 18° (MLC)	12.604	1,864,531 (42.7%)	0	0.113 89	0.075 99	0.076 56	0.003 12
on-swirl 42° (MLC)	12.604	1,864,531 (42.7%)	0	0.167 56	0.114 60	0.114 86	0.003 71
on-swirl 66° (MLC)	12.604	1,864,531 (42.7%)	0	0.059 93	0.027 61	0.027 69	0.002 79
off-swirl 18° (MLC)	5.744	849,655 (19.5%)	0.051 06	0.065 28	0.060 95	0.065 28	0.002 45
off-swirl 42° (MLC)	5.744	849,655 (19.5%)	0.085 31	0.106 50	0.096 06	0.106 50	0.003 12
off-swirl 66° (MLC)	5.744	849,655 (19.5%)	0.007 99	0.034 78	0.021 62	0.034 77	0.001 96
diffuse-swirl 18° (MLC)	11.159	1,650,780 (37.8%)	0.055 10	0.093 22	0.068 62	0.068 51	0.002 64
diffuse-swirl 42° (MLC)	11.159	1,650,780 (37.8%)	0	0.168 44	0.105 45	0.105 29	0.003 98
diffuse-swirl 66° (MLC)	11.159	1,650,780 (37.8%)	0.004 31	0.046 83	0.024 81	0.024 78	0.002 01
on-swirl 18° (K-Means)	9.689	1,433,365 (32.8%)	0.059 1	0.113 89	0.076 65	0.076 96	0.003 37
on-swirl 42° (K-Means)	9.689	1,433,365 (32.8%)	0.099 9	0.168 44	0.115 76	0.115 95	0.003 33
on-swirl 66° (K-Means)	9.689	1,433,365 (32.8%)	0.008 2	0.059 93	0.028 43	0.028 69	0.002 30
off-swirl 18° (K-Means)	8.415	1,244,836 (28.5%)	0	0.085 41	0.062 34	0.062 19	0.003 30
off-swirl 42° (K-Means)	8.415	1,244,836 (28.5%)	0	0.120 89	0.097 26	0.097 11	0.003 60
off-swirl 66° (K-Means)	8.415	1,244,836 (28.5%)	0	0.040 00	0.021 98	0.021 93	0.002 01
diffuse-swirl 18° (K-Means)	11.402	1,686,765 (38.7%)	0.053 98	0.098 31	0.070 83	0.070 87	0.003 30
diffuse-swirl 42° (K-Means)	11.402	1,686,765 (38.7%)	0.094 82	0.133 26	0.108 19	0.107 97	0.003 48
diffuse-swirl 66° (K-Means)	11.402	1,686,765 (38.7%)	0.003 69	0.036 55	0.025 32	0.025 21	0.001 89
			Mare Ingenii				
on-swirl 33° (MLC)	10.222	1,512,222 (29.4%)	0	0.097 82	0.063 07	0.063 97	0.004 54
on-swirl 48° (MLC)	10.222	1,512,222 (29.4%)	0	0.079 79	0.045 02	0.045 47	0.004 33
on-swirl 63° (MLC)	10.222	1,512,222 (29.4%)	0	0.066 72	0.030 54	0.030 70	0.004 13
off-swirl 33° (MLC)	5.542	819,795 (16.0%)	0.028 28	0.043 86	0.040 84	0.040 51	0.002 15
off-swirl 48° (MLC)	5.542	819,795 (16.0%)	0.017 36	0.034 29	0.028 46	0.028 29	0.001 84
off-swirl 63° (MLC)	5.542	819,795 (16.0%)	0.005 32	0.027 18	0.018 69	0.018 60	0.001 85
diffuse-swirl 33° (MLC)	18.997	2,805,751 (54.6%)	0	0.059 11	0.051 12	0.051 15	0.004 26
diffuse-swirl 48° (MLC)	18.997	2,805,751 (54.6%)	0.014 66	0.048 36	0.036 07	0.036 12	0.003 56
diffuse-swirl 63° (MLC)	18.997	2,805,751 (54.6%)	0.001 92	0.039 59	0.024 37	0.024 39	0.003 11
on-swirl 33° (K-Means)	8.141	1,204,354 (23.4%)	0.043 03	0.097 82	0.064 48	0.065 00	0.004 38
on-swirl 48° (K-Means)	8.141	1,204,354 (23.4%)	0.030 52	0.079 79	0.046 15	0.046 84	0.003 52
on-swirl 63° (K-Means)	8.141	1,204,354 (23.4%)	0.014 05	0.066 72	0.031 50	0.032 10	0.003 14
off-swirl 33° (K-Means)	12.252	1,812,455 (35.3%)	0	0.078 55	0.044 09	0.043 91	0.003 78
off-swirl 48° (K-Means)	12.252	1,812,455 (35.3%)	0	0.048 70	0.030 72	0.030 48	0.002 69
off-swirl 63° (K-Means)	12.252	1,812,455 (35.3%)	0	0.035 34	0.020 24	0.020 00	0.002 36

Table 3
(Continued)

Unit and Image Incidence Angle (Classification Type)	Surface Area (km^2)	Pixel Count (% of Total Study Area)	Reflectance Min[a]	Reflectance Max	Reflectance Median	Reflectance Mean	Reflectance std. dev.
diffuse-swirl 33° (K-Means)	14.338	2,120,959 (41.3%)	0.037 04	0.090 32	0.054 49	0.054 50	0.003 64
diffuse-swirl 48° (K-Means)	14.338	2,120,959 (41.3%)	0.019 61	0.054 45	0.038 51	0.038 49	0.002 50
diffuse-swirl 63° (K-Means)	14.338	2,120,959 (41.3%)	0.003 49	0.040 01	0.026 00	0.026 00	0.002 15

Note.
[a] Values represent null or no data and were not included in the statistical calculations.

the majority of these areas to the expected unit based on our observation of the representative albedo trends. The highest probabilities (>94%) generally occur within training areas and are most common to the central portion of output unit areas. Moderate probabilities (91%–94%) occur elsewhere and make up the bulk of most unit areas. The classification map using the K-means clustering algorithm is similar to the MLC map in terms of consistency between the major locations for on-, off-, and diffuse-swirl units (Figures 6(a) and (b)). The optimal number of units from K-means using the elbow method is three for both study regions (Figures 7(a) and (b)), lending support to the presence of a diffuse-swirl unit.

Pixel counts for the three units using the MLC algorithm show that both on-swirl and diffuse-swirl have similar coverage in Reiner Gamma, 42.7% and 37.8% respectively, while off-swirl has considerably less coverage (19.5%). Results using the K-means algorithm show roughly even coverage for the three units with 32.8% on-swirl, 28.5% off-swirl, and 38.7% diffuse-swirl. For Mare Ingenii, pixel counts using the MLC algorithm show that slightly over half (54.6%) of the study region is mapped as diffuse-swirl with 29.4% on-swirl and 16% off-swirl. The coverage using the K-means algorithm is 23.4% on-swirl, 35.3% off-swirl, and 41.3% diffuse-swirl. In comparing the maps for both MLC and K-Means, it is clear that more off-swirl is classified using the K-means algorithm. This is particularly evident for the main NW-SE off-swirl region through Mare Ingenii, which is significantly wider compared to that in the MLC map. Furthermore, a narrow tongue of off-swirl is classified near the SW corner of the Mare Ingenii region, which is almost exclusively diffuse-swirl in the MLC map (compare Figures 4(b) and 6(b)).

Statistically, reflectance values for both study regions follow two main trends. Based strictly on unit, on-swirl regions have the highest median and mean values, off-swirl regions have the lowest median and mean values, and diffuse-swirl regions have values that are between those of on- and off-swirl regions (see Table 3). The other trend is the relation of high- to low-reflectance values, where regardless of the unit, the highest values are associated with the lowest incidence and lowest values are associated with the highest incidence, an inverse relation. Both of these statistical trends are consistent with either MLC or K-means image classification. The lone exception occurs with the middle-incidence data for Reiner Gamma, which were taken near opposition, resulting in high-reflectance values.

5.1. Comparison of MLC and K-Means Results: Cumulative Distribution Plots

To better evaluate the similarities and differences between the MLC and K-means classification results, we show the distribution of reflectance values in the form of cumulative distribution function (CDF) plots at all incidence angles for both Reiner Gamma and Mare Ingenii. The plots show the proportion of all reflectance values that are lower than or greater than a value in the data set. The benefit of CDFs over histograms and other plotting styles is that all values are shown without binning. As expected, on-swirl areas have the highest reflectance values, followed by diffuse-swirl and off-swirl with lower values.

For Reiner Gamma, there is relatively even separation of reflectance ranges between on-, off-, and diffuse-swirl areas at each incidence angle using the MLC algorithm (Figures 8(a)–(c)). Using the midpoint of the CDF plot (50%) for 18° incidence-angle data as reference, the difference in reflectance is ∼0.007 between on- and diffuse-swirl as well as between diffuse- and off-swirl. For 42° incidence-angle data, the difference is ∼0.008 between on- and diffuse-swirl and ∼0.009 between diffuse- and off-swirl. At 66° incidence, the differences are smaller, but relatively even between on- and diffuse-swirl (∼0.002 5) and diffuse- and off-swirl (∼0.0033).

The separation of reflectance ranges for Reiner Gamma are slightly less even with the K-Means algorithm at low and middle-incidence angles (Figure 8). Here, diffuse-swirl values are shifted closer to on-swirl values. For example, at 18° low incidence, on- and diffuse-swirl have a difference of ∼0.005, but diffuse- and off-swirl have a ∼0.008 difference. At 42° middle incidence, the difference between on- and diffuse-swirl is ∼0.007 and ∼0.009 between diffuse- and off-swirl. Even with the minor differences here compared to the MLC results, the CDF plots show overall that there is clear separation of reflectance value ranges between three units for both the MLC and K-means algorithms.

For Mare Ingenii, the results are similar to those for Reiner Gamma, though it appears that the separation is more even among the three types with K-means rather than for MLC (Figure 9). At the midpoint of the 33° low-incidence data, the reflectance separation between on- and diffuse-swirl, as well as between diffuse- and off-swirl, are ∼0.01. At 48° middle incidence, the separations are ∼0.009 between on- and diffuse-swirl and ∼0.007 between diffuse- and off-swirl. At 63° high incidence, the separations for both are ∼0.006. Similar to Reiner Gamma, the CDF plots show a clear separation of the

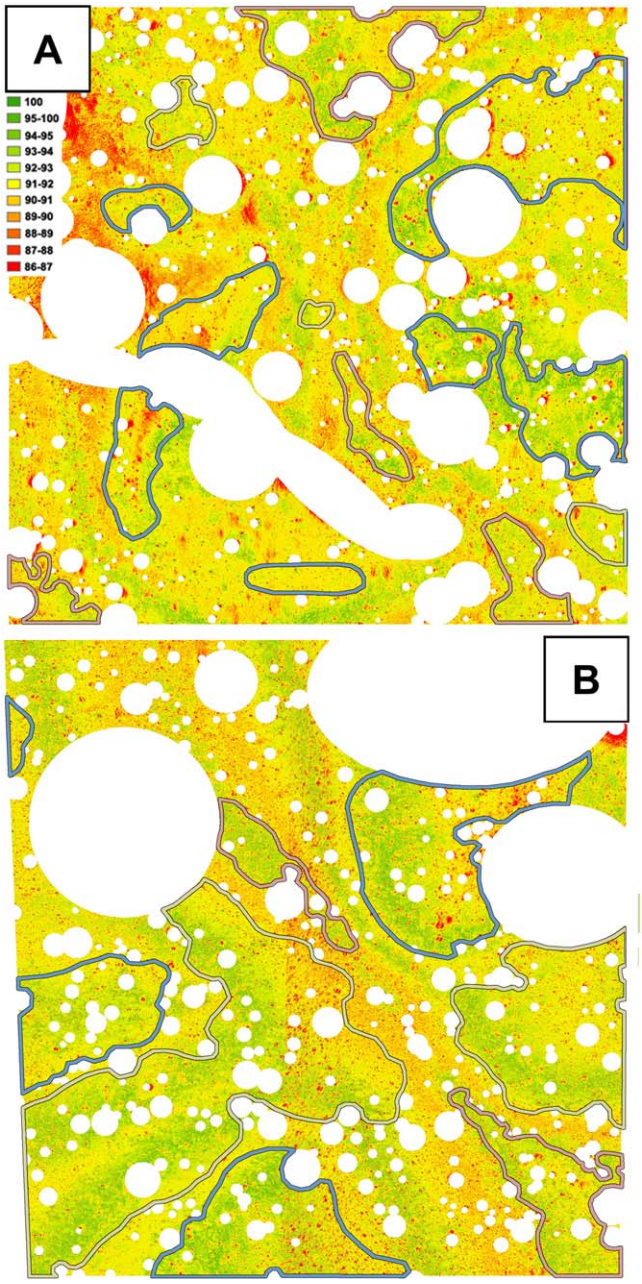

Figure 5. Probability maps from the MLC algorithm for the (A) Reiner Gamma and (B) Mare Ingenii study regions. Lower probabilities (<91%) are red to orange, moderate probabilities (91%–94%) are yellow to light green, and higher probabilities are light green to dark green (>94%). Lower probabilities are associated with anomalously high albedos, typically areas with greater relief or geologic features such as the edges of crater rims or ejecta deposits. For reference, the training areas for on-swirl (blue), off-swirl (pink), and diffuse-swirl (tan) are shown (see Figures 3(a) and (b)).

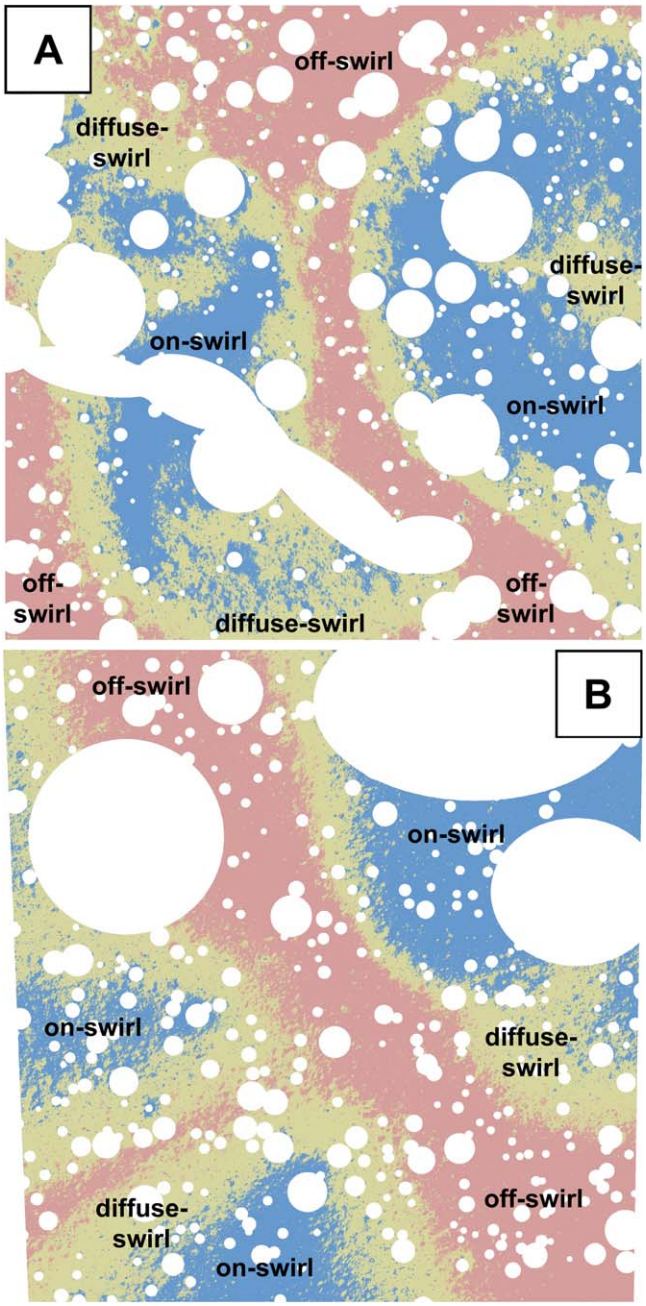

Figure 6. Classification maps using the K-means algorithm for the (A) Reiner Gamma and (B) Mare Ingenii study regions. Multiband albedo data including all three incidence angles were used in the algorithm for both study regions. On-swirl areas are shown in blue, off swirl in pink, and diffuse-swirl in tan. For details on the surface area and pixel counts of each unit type, see Table 3.

reflectance value ranges between the three units for both algorithms.

6. Discussion and Summary

Results from the supervised and unsupervised classifications clearly show that each unit has a distinct range of reflectance values, with overlap only at the extreme low and high ends of the range, typically the outermost 2%–3% (see Figures 8 and 9). Because of the separation, this confirms the presence of a third principal swirl unit, the diffuse-swirl. This unit has a range of values that are between those of on- and off-swirl, and it is typically located between the two units, indicative of a transitionary region. From the CDF curves for both study regions, low and middle-incidence data generally have larger separation between on-, diffuse-, and off-swirl than at high incidence, i.e., a wider range of representative reflectance values for each of the defined units. A possible explanation for the wider ranges may be the better representative albedo contrast at moderate- to high sun illumination than at low sun illumination, where the topography is enhanced by shadows, de-emphasizing the overall contrast.

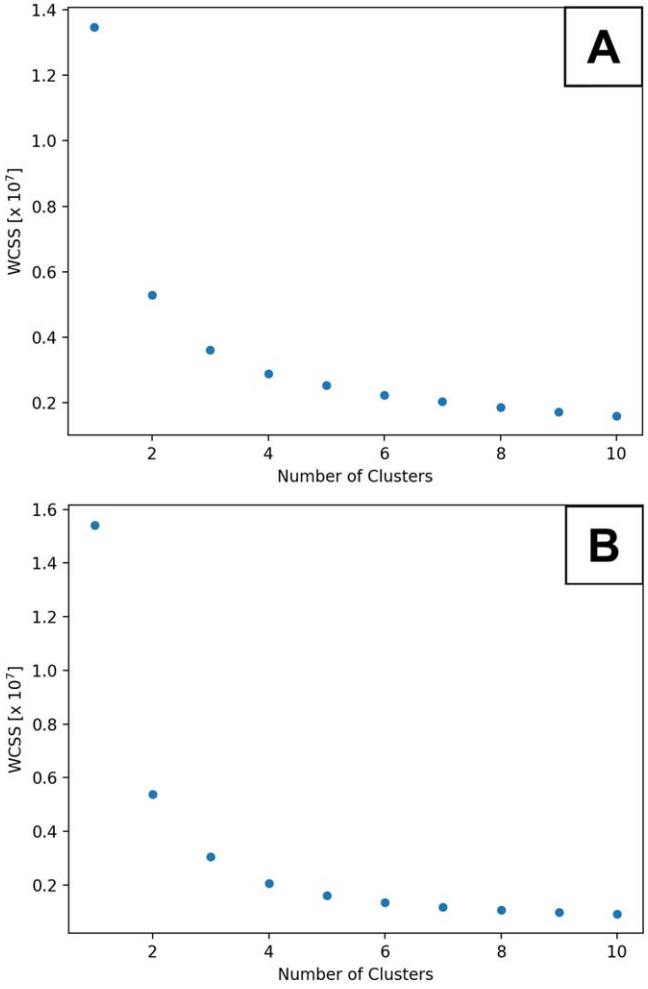

Figure 7. Elbow plots for (A) Reiner Gamma and (B) Mare Ingenii of the optimal number of k cluster units using the K-means classification algorithm. The total within-cluster sum of squared distances (WCSS) between data points and the centroid location is plotted against the k number of units being evaluated (see text for details). Here, the elbow occurs at $k = 3$, or three units, when the WCSS no longer changes significantly at a given k and beyond.

From the K-means and MLC classification maps, there is also some level of refinement in unit types from the unsupervised to supervised classification. This is reflected in the statistics (see Table 3), where the trend is from roughly even percentages of on-, off-, and diffuse-swirl for K-means to higher percentages of on- and diffuse-swirl for the MLC algorithm. While the percentages are likely due to the actual reflectance and their distribution in each study region, they may also be related to the differences between the two algorithms. K-means is an unsupervised classification with no user input and uses only the full data set values. Thus, it would be likely that an initial evaluation using K-means would return a more even distribution of pixels among the three units without prior knowledge. Conversely, the MLC with user-defined unit values from training areas would help better define such areas in the output map and would thus result in a less even distribution.

Overall, we believe the application of both supervised and unsupervised image classification algorithms for two regions with prominent lunar swirls shows that machine learning methods are effective. More specifically, they are effective in determining the number of units, mapping the entire region using quantitative information, and have the capability of using

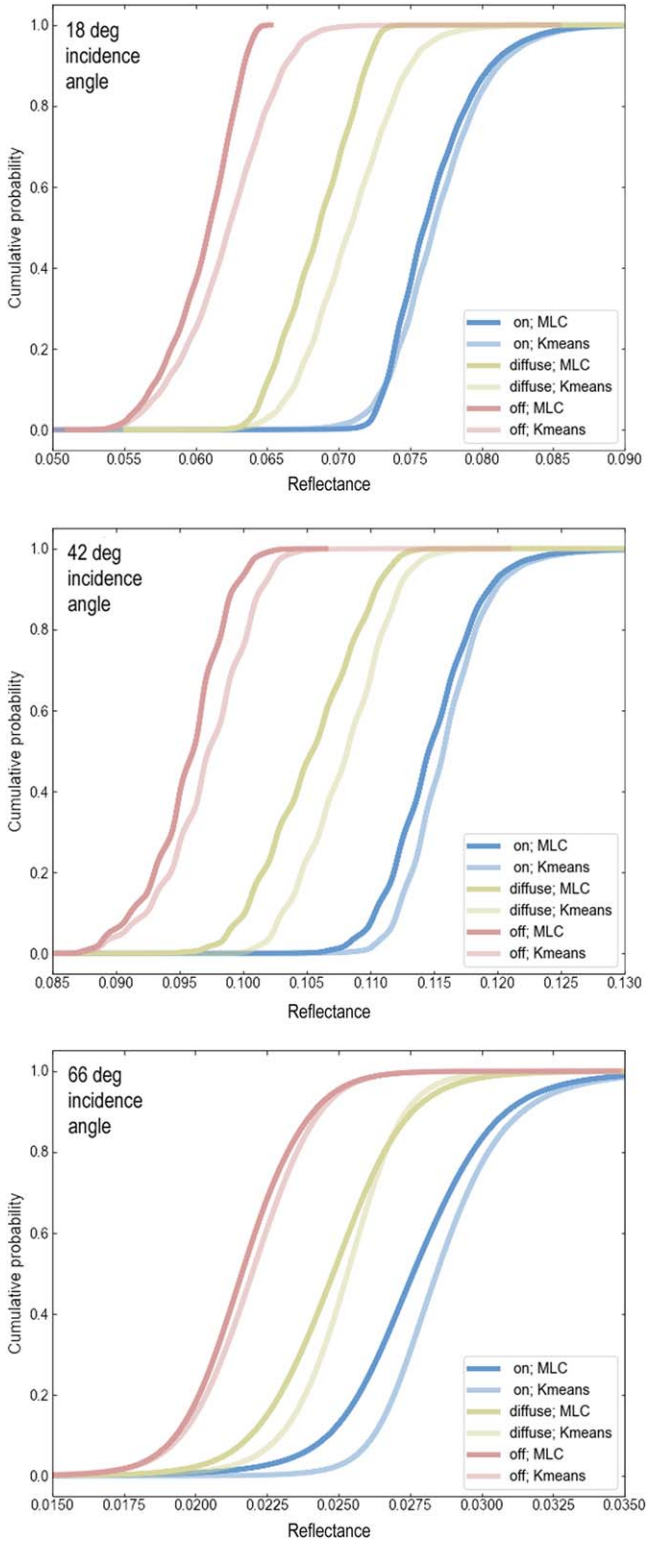

Figure 8. Cumulative distribution function plots of reflectance values for on- (blue), off- (pink), and diffuse-swirl (tan) classes in the Reiner Gamma study region at (top) 18°, (middle) 42°, and (bottom) 66° incidence angles. The spacings between the three units are relatively even for the MLC algorithm, whereas the diffuse-swirl values are shifted slightly closer to on-swirl values for the K-means algorithms at both 18° and 42° incidence angles. The distinct reflectance ranges for each unit indicate three distinct swirl units.

multiple data sets to produce classification maps comparable to the observed representative albedo differences. Unsupervised K-means classification provides a first-hand evaluation of

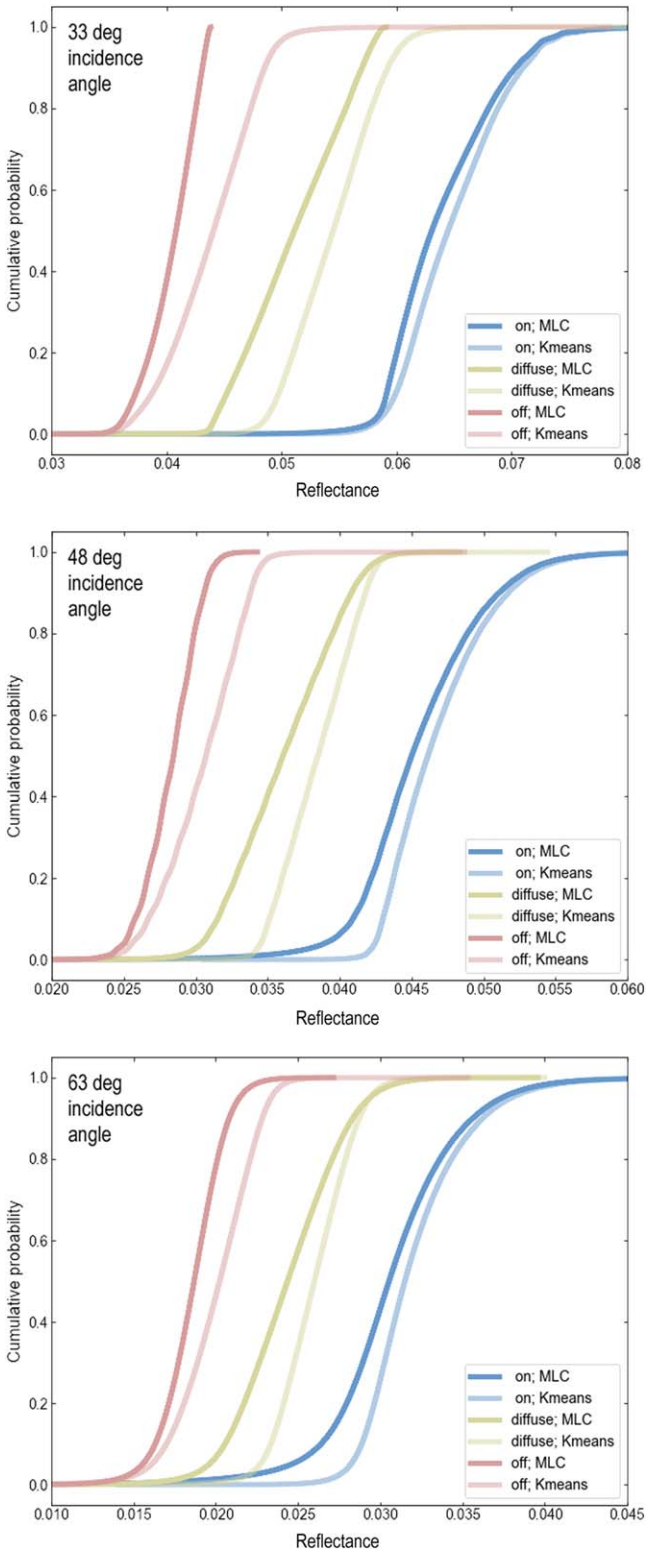

Figure 9. Cumulative distribution function plots of reflectance albedo values for on- (blue), off- (pink), and diffuse-swirl (tan) classes in the Mare Ingenii study region at (top) 33°, (middle) 48°, and (bottom) 63° incidence angles. The spacings between the three units are relatively even for the K-means algorithm. The distinct reflectance ranges for each unit indicate three distinct swirl units.

swirls that is unbiased, with the optimal number of units. The resulting map can then be used as a means to search for training areas within specific units for the MLC algorithm. With the training areas, the supervised classification allows for greater refinement using data values that are most representative for each unit. The resulting MLC and probability maps can then be used to further evaluate the mapped areas, particularly those near the transition point from one unit to another. The probability also provides some level of confidence in the classification, and the results for our two study regions show that even at low probabilities, the majority of these areas met our expectations based on the representative albedo images. The sequence of unsupervised followed by supervised classification, which has been traditionally applied by Earth remote-sensing scientists, is appropriate here for mapping lunar swirls.

While both algorithms used all three incidence-angle images as input, the map results show that classification tends to be dictated by the data with the better contrast, which are at lower incidence angles (i.e., higher Sun). Thus, the opposition effect that is present in the middle-incidence reflectance data for Reiner Gamma did not affect the assigned swirl units for the MLC and K-means algorithms.

The combination of both supervised and unsupervised algorithms produces robust mapping results that are likely more effective than just a single algorithm. They provide independent confirmation of the number of units, the location of these units, and the interrelation between these units. Thus, any proposed formation process or processes for lunar swirls need to account for these results. More importantly, however, this study shows that machine learning can (1) identify units using actual reflectance data, which is an improvement over observed representative albedo contrast and removes potential user subjectivity, and (2) better define units and their boundaries using a quantitative rather than qualitative approach. Moreover, the quantitative data obtained from the swirl units can be used in terms of a relative comparison to similar data in other lunar swirl locations assuming the incidence, emission, and phase angles can be held constant. Reflectance data and the statistics generated from them provide measurable criteria, which have value in placing limits for categorizing swirl units.

This study was supported under NASA Lunar Data Analysis Program (LDAP) grant 80NSSC17K0278 to D. Domingue and NASA's Solar System Exploration Research Virtual Institute (SSERVI) Toolbox for Research and Exploration (TREX) grant 80ARC017M0005. The statistics and CDF plots in this study were generated using open-source Pandas and NumPy array programming software for Linux (Harris et al. 2020). The I/F reflectance data, both masked and nonmasked, in raster GeoTiff (.tif) or space-delimited ASCII text format (.txt) may be obtained at Zenodo (http://www.zenodo.org) by searching under the lead author name. We thank Eric Palmer (Planetary Science Institute) for his insight on exploring quantitative methods for evaluating lunar swirls.

ORCID iDs

Frank C. Chuang https://orcid.org/0000-0001-8290-7930
Matthew D. Richardson https://orcid.org/0000-0002-9122-5082
John R. Weirich https://orcid.org/0000-0002-2830-1708
Amanda A. Sickafoose https://orcid.org/0000-0002-9468-7477
Deborah L. Domingue https://orcid.org/0000-0002-7594-4634

References

Anderson, R. B., & Bell, J. F. 2013, Icar, 223, 157
Asada, N., Hirata, N., Hirohide, N., et al. 2010, AdG, 19, 77
Azari, A., Biersteker, J. B., Dewey, R. M., et al. 2021, BAAS, 53, 128
Bell, J. F., & Hawke, B. R. 1981, LPSC, 12, 59
Blewett, D. T., Coman, E. I., Hawke, B. R., et al. 2011, JGRE, 116, E02002
Blewett, D. T., Denevi, B. W., Cahill, J. T. S., et al. 2021, Icar, 364, 114472
Blewett, D. T., Hughes, C. G., Hawke, B. R., et al. 2007, LPSC, 38, 1232
Campbell, B. A., Carter, L. M., Campbell, D. B., et al. 2006, LPSC, 37, 1717
Chevrel, S., Pinet, P. C., Jehl, A., et al. 2006, LPSC, 37, 1173
Collier, M. R., Gruesbeck, J. R., Connerney, J. E. P., et al. 2020, JGRE, 125, e06366
Denevi, B. W., Robinson, M. S., Boyd, A. K., et al. 2016, Icar, 273, 53
Domingue, D., Palmer, E., Gaskell, R., et al. 2018, Icar, 312, 61
Domingue, D., Weirich, J., Chuang, F., et al. 2021, GeoRL, 49, e95285
Domingue, D., Weirich, J., Chuang, F., et al. 2022, PSJ, submitted
El-Baz, F. 1972, Apollo 16 Preliminary Science Report NASA SP-315 NASA, 29–93, https://www.lpi.usra.edu/lunar/documents/NASA%20SP%20315.pdf
Fuller, M. 1974, RvGSP, 12, 23
Fuller, M., & Cisowski, S. M. 1987, Geomagnetism, 2, 307
Garrick-Bethell, I., Head, J. W., III, & Pieters, C. M. 2011, Icar, 212, 480
Garrick-Bethell, I., Weiss, B. P., Shuster, D. L., et al. 2009, Sci, 323, 356
Gaskell, R. W., Barnouin-Jha, O. S., Scheeres, D. J., et al. 2008, M&PS, 43, 1049
Gehrels, T., Coffeen, T., & Owings, D. 1964, AJ, 69, 826
Harris, C. R., Millman, K. J., & van der Walt, S. J. 2020, Natur, 585, 357
Hood, L. L., Coleman, P. J., & Wilhelms, D. E. 1979, Sci, 204, 53
Hood, L. L., & Schubert, G. 1980, Sci, 208, 49
Hood, L. L., & Williams, C. R. 1989, LPSC, 19, 99
Kaydash, V., Kreslavsky, M., Shkuratov, Y., et al. 2009, Icar, 202, 393
Kerner, H., Campbell, J., & Strickand, M. 2022, in Machine Learning for Planetary Science, ed. J. Helbert et al. (Amsterdam: Elsevier), 1
Kramer, G. Y., Besse, S., Dhingra, D., et al. 2011a, JGRE, 116, E00G18
Kramer, G. Y., Combe, J.-P., Harnett, E. M., et al. 2011b, JGRE, 116, E04008
Kreslavsky, M. A., & Shkuratov, Y. G. 2003, JGRE, 108, 5015
McKinney, W. 2010, in SciPy 2010, 56, https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf
Moore, H. J., Hodges, C. A., & Scott, D. H. 1974, LPSC, 5, 71
Oetking, P. 1966, JGR, 71, 2505
Palmer, E. E., Head, J. N., Gaskell, R. W., et al. 2016, E&SS, 3, 488
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, arXiv.1201.0490
Pieters, C. M., & Garrick-Bethell, I. 2015, LPSC, 46, 2120
Pinet, P. C., Shevchenko, V. V., Chevrel, S. D., et al. 2000, JGR, 105, 9457
Rammelkamp, K., Gasnault, O., Forni, O., et al. 2021, E&SS, 8, e01903
Richards, J. A., & Jia, X. 1999, Remote Sensing Digital Image Analysis (3rd ed.; New York: Springer)
Richmond, N. C., Hood, L. L., Mitchell, D. L., et al. 2005, JGR, 110, E05011
Sarker, I. 2021, Springer-Nature Computer Sci., 2, 160
Sato, H., Robinson, M. S., Lawrence, S. J., et al. 2017, Icar, 296, 216
Satopaa, V., Albrecht, J., Irwin, D., et al. 2011, in IEEE 31st Int. Conf. on Distributed Computing Systems Workshops (Piscataway, NJ: IEEE), 1
Schultz, P., & Srnka, L. 1980, Natur, 284, 22
Schultz, P. H., & Gault, D. E. 1975, Moon, 12, 159
Starukhina, L. V., & Shkuratov, Y. G. 2004, Icar, 167, 136
Stuart-Alexander, D. E. 1978, Geologic map of the central far side of the Moon, IMAP 1047, USGS, https://doi.org/10.3133/i1047
Syal, M. B., & Schultz, P. H. 2015, Icar, 257, 194
Wilhelms, D. E., McCauley, J. F., & Trask, N. J. 1987, The geologic history of the Moon, Professional Paper 1348, USGS, https://doi.org/10.3133/pp1348
Wulder, M. A., Loveland, T. R., Roy, D. P., et al. 2019, RSEnv, 225, 127