Revisiting Atmospheric Features of Mars Orbiter Laser Altimeter Data Using Machine Learning Algorithms

Vincent Caillé, Anni Määttänen, Aymeric Spiga, Lola Falletti

JGR Planets

RESEARCH ARTICLE
10.1029/2022JE007384

Revisiting Atmospheric Features of Mars Orbiter Laser Altimeter Data Using Machine Learning Algorithms

Vincent Caillé[1], Anni Määttänen[1], Aymeric Spiga[2], and Lola Falletti[1]

[1]LATMOS/IPSL, Sorbonne Université, UVSQ Paris-Saclay, CNRS, Paris, France, [2]Laboratoire de Météorologie Dynamique/Institut Pierre Simon Laplace (LMD/IPSL), Sorbonne Université, Centre National de la Recherche Scientifique (CNRS), École Polytechnique, École Normale Supérieure (ENS), Paris, France

Key Points:

- We reanalyze the Mars Orbiter Laser Altimeter data set with clustering methods and retrieve a new, large atmospheric structure data set
- Comparing the data set with other observations allows us to provide a global view of atmospheric structures
- We observe the development of the aphelion cloud belt, the polar hoods, regional dust storms, and clouds over topographic features

Correspondence to:
V. Caillé,
vincent.caille@latmos.ipsl.fr

Citation:
Caillé, V., Määttänen, A., Spiga, A., & Falletti, L. (2023). Revisiting atmospheric features of Mars Orbiter Laser Altimeter data using machine learning algorithms. *Journal of Geophysical Research: Planets*, *128*, e2022JE007384. https://doi.org/10.1029/2022JE007384

Received 15 MAY 2022
Accepted 20 DEC 2022

Abstract The Mars Orbiter Laser Altimeter (MOLA) instrument has been drawing a map of Mars' topography between September 1997 and June 2001. It has also been able to observe clouds during the mission duration, providing data for the low Martian atmosphere for nearly 1.5 Mars years. The Mars Global Surveyor, which carried MOLA, also carried two other instruments that also observed clouds during the same time period (the Mars Orbiter Camera and the Thermal Emission Spectrometer). Combining observations from these three data sets could provide a complete recap of most atmospheric structures during MY24 and MY25. However, previous studies of MOLA data set often had to use stringent detection criteria. Using machine learning clustering methods, we end up finding way more atmospheric returns. Our results are presented in the form of an atmospheric features catalog that we then use to compare MOLA observations with Mars Orbiter Camera and Thermal Emission Spectrometer results, but also with more recent missions. We study the development of recurrent phenomenon in the Martian atmosphere, like the aphelion cloud belt or the south polar hood, but also spontaneous events such as regional dust storms. Methods could be tuned even more finely by using more complex clustering methods or deep learning algorithms to clearly distinguish atmospheric structures.

Plain Language Summary The Mars Orbiter Laser Altimeter (MOLA) instrument has been emitting laser pulses toward the Martian surface. Time of flight of the laser before returning to the instrument was originally used to estimate the altitude of Mars' surface, but the sensibility of the detector was good enough to detect clouds' signatures coming from the atmosphere. We propose that studying the MOLA data set using machine learning methods that gather similar laser returns into groups can enable the formation of a cluster made of atmospheric features, distinguishing them from noise and surface returns. These features are then grouped into clouds or dust structures and compared with other mission results that also observed the Martian atmosphere between 1997 and 2001. This paints a picture of many phenomena in the low Martian atmosphere, their seasonal and interannual variability and their varying intensity.

1. Introduction

Mars' atmosphere is made of a diversity of aerosols and clouds like water ice, CO_2 ice, and dust and strongly varies seasonally. Understanding of their characteristics, composition, extent, and spatial and temporal distribution have gradually improved over the last decades thanks to the diversity of Martian missions. First confirmations of water clouds composition were made by thermal infrared spectroscopy using the Mariner 9 Orbiter (Curran et al., 1973) that also highlighted large-scale cloud structures and behavior in the Martian atmosphere (Anderson & Leovy, 1978). Water climatology obtained by the two Viking orbiters' color imaging (Farmer et al., 1977) helped understanding how water clouds were correlated with topographic features (French et al., 1981). Infrared spectroscopy observations by the Mariner 6 and 7 Orbiters (Herr & Pimentel, 1970) had suggested the existence of CO_2 ice cloud formation, but no direct observations were available at that time. Since then, CO_2 Martian ice clouds have been observed by a diversity of methods like imaging, stellar, and solar occultations. Mars Global Surveyor (MGS) and Mars Pathfinder missions in the late 90s were precursors to missions that provided a wealth of data about Mars' atmosphere and clouds. For example, all of MGS Thermal Emission Spectrometer (TES) (Christensen et al., 2001), MGS Mars Orbiter Camera (MOC) (Malin & Edgett, 2001), and MGS Mars Orbiter Laser Altimeter (MOLA) (Smith et al., 2001) have allowed the observation of Martian clouds. TES showed the evolution of the aphelion cloud belt (Hale et al., 2011), gave dust and water ice aerosol optical depth, as well as water vapor column density during nearly 3 Martian Years, and provided interannual variability (Smith, 2004). These observations also allowed to distinguish between spatial variation of daytime and nighttime clouds and

the seasonal dependence of cloud formations (Pankine et al., 2013). MOC highlighted clear seasonal patterns in clouds formation and suggested different kinds of clouds (Wang & Ingersoll, 2002), MOLA gave the first observation of CO_2 ice clouds and snowfall (Ivanov & Muhleman, 2001; Neumann et al., 2003). Subsequent missions like Mars Climate Sounder (Kleinbohl et al., 2009; McCleese et al., 2007) or Mars Express (Wilson & Chicarro, 2004) enhanced our understanding of cloud formation, evolution, and impact on the Martian atmosphere. A recent review of Martian cloud observation history can be found in Clancy et al. (2017). Still, the number of missions is limited, and getting as many cloud observations as possible from them is very important to fulfill our understanding of involved processes.

MOLA's first goal was to determine Mars' topography, roughness, and albedo at 1,064-nm wavelength. Results from a similar instrument, the Shuttle Laser Altimeter, proved the capacity of the altimeter to detect clouds in Earth's atmosphere (Bufton et al., 1995; Garvin et al., 1998) and gave hope for MOLA detecting Martian clouds despite not knowing if they would be dense enough to trigger detection. In addition to allowing clouds detection, the MOLA instrument also enabled distinguishing reflective and absorptive clouds (Neumann et al., 2003) and how their formation is linked to low atmosphere properties and dynamics. Because previous studies were numerically restrained by the size of the MOLA data set, we suggest that a reanalysis with modern methods could furnish more cloud and dust observations.

The use of machine learning methods in planetary science has become more and more common in the last years. It can be beneficial at every stage of a mission (Azari et al., 2021). Deep learning, for example, became an important part of the Transiting Exoplanet Survey Satellite processing by doing a pre-identification of interesting candidates among the monthly ∼1 million light curves (Yu et al., 2019). Machine learning has led to a new way of looking at data sets and can easily be implemented on planetary science or geoscience data sets (Azari et al., 2020; Karpatne et al., 2019). Classification may be the most intuitive way to use machine learning, easing the distinction of specific features among large data set (Wagstaff et al., 2018). In case of incompletely defined features, unsupervised machine learning can help distinguishing different classes without fixing an arbitrary stringent limit (Ni & Benson, 2020).

We propose here that unsupervised machine learning methods, and especially clustering algorithms, would allow finding more cloud observations in the MOLA data set than previous studies. Validation of the results could then be done through comparison with other cloud observations. Such a comparison, especially with data sets from the same time period, such as TES and MOC on MGS, could lead to deeper understanding of cloud formation and evolution during the observed 2 Martian Years. Our results can also be compared with cloud observations from more recent missions.

2. Method and Data

2.1. Mars Orbiter Laser Altimeter Data

The MOLA (Zuber et al., 1992) was an instrument aboard the MGS whose first objective was to characterize very precisely Mars' surface topography (Smith et al., 2001). After reaching its mapping orbit at the end of aerobraking phase, it provided almost 1.5 Martian Years of data. MOLA provided data from 28 February 1999, L_s 103° of MY24, until 30 June 2001, L_s 187° of MY25. MOLA emission was interrupted three times during the mission. Around L_s 126° of MY24, the obstruction of the MGS high-gain antenna prevented MOLA mapping and forced emission to be turned off for 2 weeks. MOLA was also turned off during the solar conjunction phase, from 1 June 2000 (right after L_s 360° of MY24) to 1 August 2000 (L_s 30° of MY25). The last one week interruption occurred at L_s 154° of MY25 to resolve a spacecraft anomaly. MOLA includes a 1,064 μm-wavelength laser sending ∼8 ns short pulses at a 10 Hz rate. Pulses are emitted in the nadir direction forming ∼168 m footprints separated by approximately 300 m. MOLA measures the time of flight of each laser pulse between the instrument and the planet surface. When a pulse is emitted, a time interval unit (TIU) is triggered. The reflected laser is received on a 50-cm diameter parabolic mirror and directed to a silicon avalanche photodiode detector. If returned energy is above a controlled TIU threshold, it will stop the count of clock cycles, giving the time of flight of the relative pulse. Both emitted and received energy are saved for each pulse. Received signal is then amplified and goes through four parallel channels. Each channel is made of its own five-pole Bessel filter and comparator. Channels are hereafter referred by their number, 1–4, respectively, corresponding to a time constant of 20, 60, 80, and 540 n. These correspond to target dispersions of 3, 9, 27, and 81 m. Saved time of flight is the one from the triggered filter with the lowest impulse response pulse width. Channels give information about

Table 1
Distribution of Mars Orbiter Laser Altimeter (MOLA) Atmospheric Structure Among the 12 Orbital Phases of MOLA Data Set

Orbital phase	First return L_s(°)	Total returns	Atmospheric returns (% of total returns)	MAS number
ap01	103.59	6 120 716	152 668 (~2.49%)	1 375
ap10	107.72	44 772 367	952 115 (~2.13%)	7 947
ap11	146.42	69 062 588	3 989 110 (~5.78%)	21 740
ap12	190.97	65 619 212	9 518 336 (~14.51%)	34 839
ap13	241.17	68 248 559	6 954 178 (~10.19%)	42 858
ap14	291.97	62 127 844	5 728 149 (~9.22%)	30 006
ap15	338.98	32 588 064[a]	3 963 538 (~12.16%)	17 324
ap16	29.46	49 626 170[a]	2 726 280 (~5.49%)	20 169
ap17	56.58	64 214 511	3 573 247 (~5.56%)	22 658
ap18	92.50	64 260 115	2 976 179 (~4.63%)	20 937
ap19	130.15	59 884 115	5 492 900 (~9.17%)	28 525
ap20	172.30[b]	21 328 664	2 545 177 (~11.93%)	14 024
Total		607 852 905	48 571 877 (~7.99%)	261 862

[a]Partially contain solar conjunction period where no pulses were emitted. [b]Last return L_s is 187.20°.

the altitude dispersion of the layer in which the laser signal has been backscattered. Accuracy on pulse width and energy measures are ~5% (Abshire et al., 2000). The MOLA range z_{mars}, the distance between the instrument and the spot at Mars' surface, is calculated from the time of flight t_{flight} and the speed of light in vacuum c with the following equation (Smith et al., 2001):

$$z_{mars} = \frac{ct_{flight}}{2} \quad (1)$$

All information about a single pulse emission (like target longitude and latitude, transmitted energy), as well as the characteristics of its detection (like solar longitude of detection, triggered channel, backscattered energy, and MOLA range) make for one observation and will thereafter be called a "return." Because Mars' atmosphere is very thin, the delay it induces in the signal is considered negligible since it would modify the MOLA range only by a few centimeters (Smith et al., 2001). The MOLA range is given with an accuracy varying between 37.5 cm for smooth features and ~10 m for 30° slopes (Gardner, 1992). More technical information about MOLA elements can be found in Smith et al. (2001). The complete list of all the observational parameters that are available for each return can be found in the Precision Experiment Data Record (PEDR) repository (Ford et al., 1998). In order to reduce the impact of solar photons, thresholds are modified to track Mars' surface in particular. This results in a limitation in MOLA capacity in detecting clouds higher than ~10 km in the atmosphere for most of the cases. However, MOLA is also able to categorize a return as a cloud. It saves a histogram of ranges for every 2-s frame and can deduce it is following a cloud layer when it differs too much from a surface range histogram. When most of the returns from a MOLA packet of 14 s were from a cloud layer, MOLA can switch to an acquisition mode allowing detection of features up to 40 km in the atmosphere. Nevertheless, when going in acquisition mode, MOLA resets its detection threshold to a level that induces a very high background noise. This leads to massive false positive cloud detection. Therefore, acquisition mode returns are excluded from most studies, including ours.

Intuitively, it seems like triggered channel and MOLA range could be parameters that would allow distinguishing cloud returns among the MOLA data set. Initially, time constants of the four filters were determined to maximize the probability of detecting varying slopes on Mars' surface. Channel 1–4 were, respectively, designed for surface with slopes of 1.7°, 5°, 15°, and 39°. This notably allows for the detection of features for which variation of height within footprint is about 3, 9, 27, or 81 m, what we call target dispersion. Returned waveform is an indicator of the roughness of the targeted footprint. While channel 1 was supposed to track smooth surface, channel 4 was thought to detect eventual cloud signatures (Table 2 of Ivanov and Muhleman (2001)). However, clouds have been detected in every channels. Triggered channel gives information about the optical thickness of the detected atmospheric structures. Channel 1 detection are optically thick within only 3 m, while the ones in channel 4 are diffuse and are optically thick within 80 m. It could be expected that atmospheric structures with different composition (water ice, CO_2 ice, and dust) would have different reflectivity properties and could be differentiated by the most triggered channel but we will show that correlations are hard to find once laser returns are grouped among structures. In addition, this variety among atmospheric structures is the reason the MOLA range, which is supposed to be an indicator of the footprint height, cannot be used to distinguish atmospheric returns. Indeed, Neumann et al. (2003) showed that there are absorptive clouds signatures in MOLA data set. For this kind of cloud, the laser signal is only attenuated by its passage through the cloud. Reflection may still occur at the surface so the time of flight will not give a specific cloud altitude.

MOLA data are provided through daily MOLA PEDR (Smith et al., 1999) that contains width pulse and energy for each laser returns. Geometric information about each return is also given, such as the local time, the solar zenith angle, and the localization of targeted surface spot. MOLA provided 607 852 905 returns divided into 767 files, divided between 12 different orbital phases. Each orbital phase contains data for a range of approximately 40°–50° of solar longitude, and the starting time of each of them is given later in Table 1. Since orbital phases were made so that they gather around the same number of returns, it will later ease the statistical analysis of our

results. The MOLA PEDR data are our starting point in this study. No filtering was made on the data set and all returns were used.

2.2. *K*-Means Method

Neumann et al. (2003), when performing their cloud analysis on such a large amount of data, used stringent detection criteria on backscattered energy to distinguish cloud laser returns from surface or noise returns. This may have caused some potential cloud returns to be discarded. Unsupervised machine learning algorithms are adapted to numerically analyze the MOLA data set to highlight a certain class of returns without exactly knowing all of its characteristics. In particular, clustering algorithms are especially designed to divide the data set into several groups of similar returns. The *K*-means method is a good first approach because it is certain to converge (Selim & Ismail, 1984) in a manageable time (Arthur & Vassilvitskii, 2006; Har-Peled & Sadri, 2004). Its results can also later serve as initialization for more complex methods. The *K*-means method aims at partitioning the data set into k clusters with each data point belonging to only one cluster. The method iterates several times through the process described below to minimize a specific function depending on the chosen algorithm. The only hyperparameters of the method, that is, the factors that need to be chosen for the method to run, are the number of clusters k, the n distinguishing variables, and eventually fixing the number of steps.

In the n-dimensional space, each laser return is a point with n coordinates, corresponding to every clustering variables, and each cluster is defined by its centroid, that is, the mean value of every variables for this specific cluster. One algorithm step is made of the succession of two distinct phases: assignment and update. During the assignment phase, each point is assigned to the cluster with the closest centroid. Then, during the update phase, the mean of every cluster is recalculated. If the number of steps is not specified, the algorithm ends when there is no more change during an assignment phase. The resulting clusters represent the final partition. This way, the method aims to minimize within-cluster variances constructing clusters of similar returns (Macqueen, 1967).

The speed of the method and the final results depend on how the first clusters are initialized (Milligan, 1980). We use a classic k means algorithm from the scikit-learn package. Several ways exist to initialize the cluster centroids; we picked the Forgy method (Anderberg, 1973; Forgy, 1965). In this method, a k number of points are randomly chosen from the data set and are set as k cluster centroids. Then, the algorithm starts with the first assignment phase. Compared to the Random Partition method for which all points are first randomly assigned to a cluster and then an update phase is realized, Forgy initialization leads to better results for standard *K*-means method (Hamerly & Elkan, 2002; Peña et al., 1999). Controlling the randomization of the Forgy method to ensure that first points are far from each other gives even better results and defines the k-means++ initialization method (Arthur & Vassilvitskii, 2007) that we use here.

2.3. Optimization of Hyperparameters

2.3.1. Choice of *n* Variables

The first thing to consider before applying the *K*-means method to the MOLA data set is to choose how many variables are necessary to distinguish the three main type of laser returns: surface, noise, and atmospheric features. Note that atmospheric features are water and CO_2 ice clouds, ground fogs that would be closer to the surface than clouds, and dust structures. For this, we have several constraints.

The optimized number of clusters grows rapidly with the number of variables; so for easier interpretation, we try to work with only two variables. Moreover, because we want to work with the raw data with as little preprocessing as possible, the chosen variables must be directly derivable from the features available from MOLA data set. Not preprocessing the data was a choice to ease the reproducibility of the clustering in case of future studies with more complex clustering methods. Moreover, because of the variations of MOLA capacity during the mission duration, removing the surface signature cannot be done homogeneously on the whole data set. Another aspect is the duration of the mission. MOLA's transmitted energy decreased during the 2 years, so final variables should be normalized to take into account the decrease in performance and remain coherent for the whole data set.

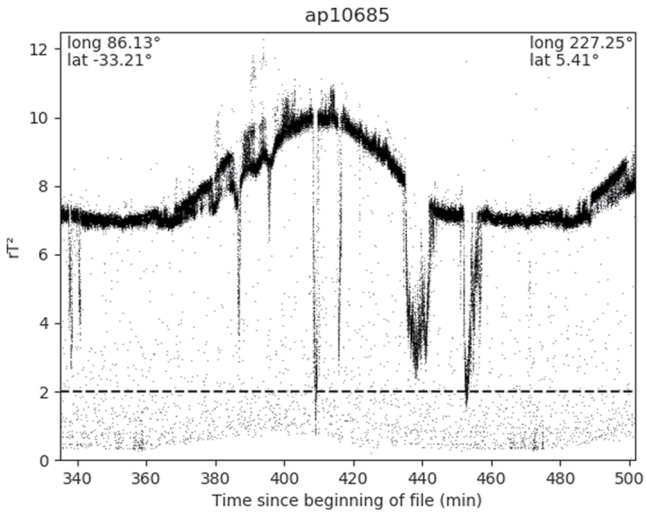

Figure 1. rT^2 against time for a ~15 m window from orbital phase ap10. Series of several low rT^2 returns are typical signatures of atmospheric structures while noise returns are sporadic decreases. Dashed line represents detection criteria used in Neumann et al. (2003). Some signatures appear as completely above this limit and may have been missed.

As reference studies pointed out (Neumann et al., 2003), the product of surface reflectivity with the two-way atmosphere transmissivity, hereafter referred as rT^2 product, can be seen as a "normalized returned energy" (Abshire et al., 2000) and is calculated by the following equation:

$$r_{\text{surf}} T_{atm}^2 = \frac{E_{rec}}{E_{trans}} z_{mars}^2 \frac{\pi}{t_r A_r} \quad (2)$$

where E_{rec} and E_{trans} are, respectively, the received and transmitted energy, z_{mars} is the distance between the spacecraft and the targeted point of Mars' surface, t_r is the optical transmission of MOLA receiver, and A_r is the area of the telescope. The transmitted energy was the main decreasing factor in MOLA capacity over the mission duration since aging of the detector was negligible enough for us to consider them constant. Another benefit of using rT^2 is that no sorting had to be done before applying the method. Indeed, some elements could affect reflectance such as the varying solar zenith angle or shadows at the surface. However, sunlight is already accounted as noise in MOLA detection and sunlight backscattered by the surface is much weaker than laser light, especially at this wavelength.

Figure 1 shows an example of the normalized return energy rT^2 against time. Surface returns are represented by a continuum, while atmospheric features, including clouds, cause a dip. Noise returns are seen as sporadic points. From this, we can deduce two variables that would distinguish the three types of laser returns: the rT^2 product and the average of the rT^2 product of the closest neighbor returns thereafter referred to as $\overline{rT^2}$. For surface returns, both of these variables should have a value close to the continuum value. For noise returns, rT^2 could be lower than continuum value but, because it is a single return, $\overline{rT^2}$ should not be affected that much and keep a value near the continuum one. Finally, for atmospheric features, both variables should clearly be lower than the continuum. The window around each return to compute $\overline{rT^2}$ will be limited to the two preceding and the two following returns since the probability that five consecutive returns are noise returns is lower than 10^{-5} (see Section 3.2.1 for noise returns rate). Keeping the number of returns in the window as low as possible gives more weight to weak rT^2 variations, so five was a good compromise between excluding noise while using a narrow window.

Overall, the energy transmitted by MOLA decreased by half over the mission duration (Neumann et al., 2003). This results in an increase in rT^2 product continuum level over time. Assumptions we made about rT^2 and $\overline{rT^2}$ behaviors for the different kind of returns are a comparison of their values with associated continuum of the same orbital phase. Yet, the continuum level of the first phases would be considered a low value for the latest phases. Thus, we have divided our data set into periods of approximately constant continuum level to apply our method. Since the data set is already divided into orbital phases representing approximately 10% of the whole data set, apart from the ones corresponding to the start and the end of the mapping or the ones containing a part of the solar conjunction (see Table 1 for more details), we successively apply our clustering method to each of these phases.

In conclusion, we decided to work with only two variables ($n = 2$) that are rT^2 and $\overline{rT^2}$ to apply K-means methods on MOLA data set.

2.3.2. Number of Clusters

The use of the K-means method implies an optimization of the number of clusters that is done by evaluating the clustering performance for a certain range of k. For this, we use three independent methods that compute different aspects of the partition for every value of k in a certain range:

- Elbow method (Thorndike, 1953) computes the total intra-cluster variation within cluster sum of squares that measures the variability of the two variables. The best number of clusters is the one corresponding to the elbow shape. After the elbow, increasing the number of cluster does not substantively improve the performance since the sum of squares is very slightly modified.
- Average silhouette method (Kaufman & Rousseeuw, 2005) computes the so-called silhouette score from the distance of each point to the centroid of its cluster and the second closest one, evaluating how distinct the clusters are. The best number of clusters is the first maximum larger than the smallest tested number of clusters.

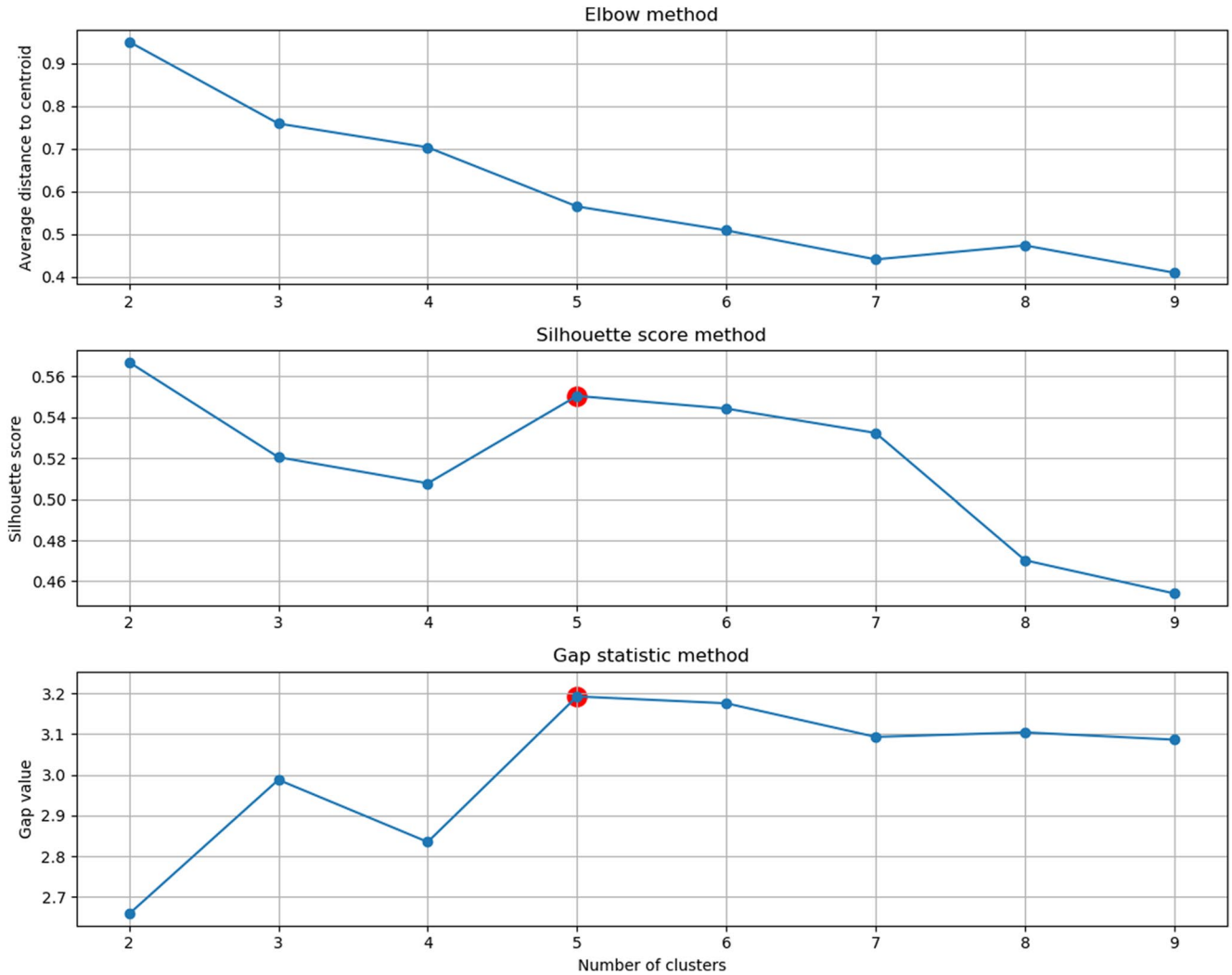

Figure 2. Optimization of the number of clusters with three independent clustering performance testing methods: elbow (top), silhouette score (middle), and gap statistic (bottom). K means method is applied with 2–9 clusters on mini batches from the data set. Red dots highlight optimized number of clusters. Elbow method does not furnish a stringent best number of clusters but its results are coherent with the two other methods, converging toward five clusters, with performance for $k = 6$ being relatively close.

- Gap statistic method (Tibshirani et al., 2001) computes how far the clustering structure is from a random, uniform distribution of points with the same limit values. Here again, the best number of clusters is the first maximum larger than the smallest tested number of cluster.

Figure 2 shows the result of all three methods. Despite testing totally independent features of the partition, two methods (silhouette score and gap statistic) converge toward the same number of five clusters. For the elbow method, the result is up to the choice of the user as the method only shows a suitable interval of number of clusters. We see that the gain between $k = 5$ and $k = 6$ is low compared to the one between $k = 4$ and $k = 5$. Beyond $k = 7$, there is no improvement when adding an additional cluster. Thus, any number of clusters between 5 and 7 can be picked because it corresponds to the "elbow" shape. All three methods have coherent results: this proves that the data set is indeed "clusterable," that is, different clusters have clearly different properties. However, we decided to visually inspect the clustering structure for four, five, and six clusters. While four clusters were not enough to clearly distinguish surface, noise, and atmospheric features, both five and six were potential candidates. Because the scores of $k = 6$ were nearly as good as $k = 5$ for at least two methods (silhouette score and gap statistic) and because it eases the physical interpretation of each cluster while minimizing the risk of false positives (discussed in details in Section 3.1.1), we chose six clusters in the continuation of our study. In particular,

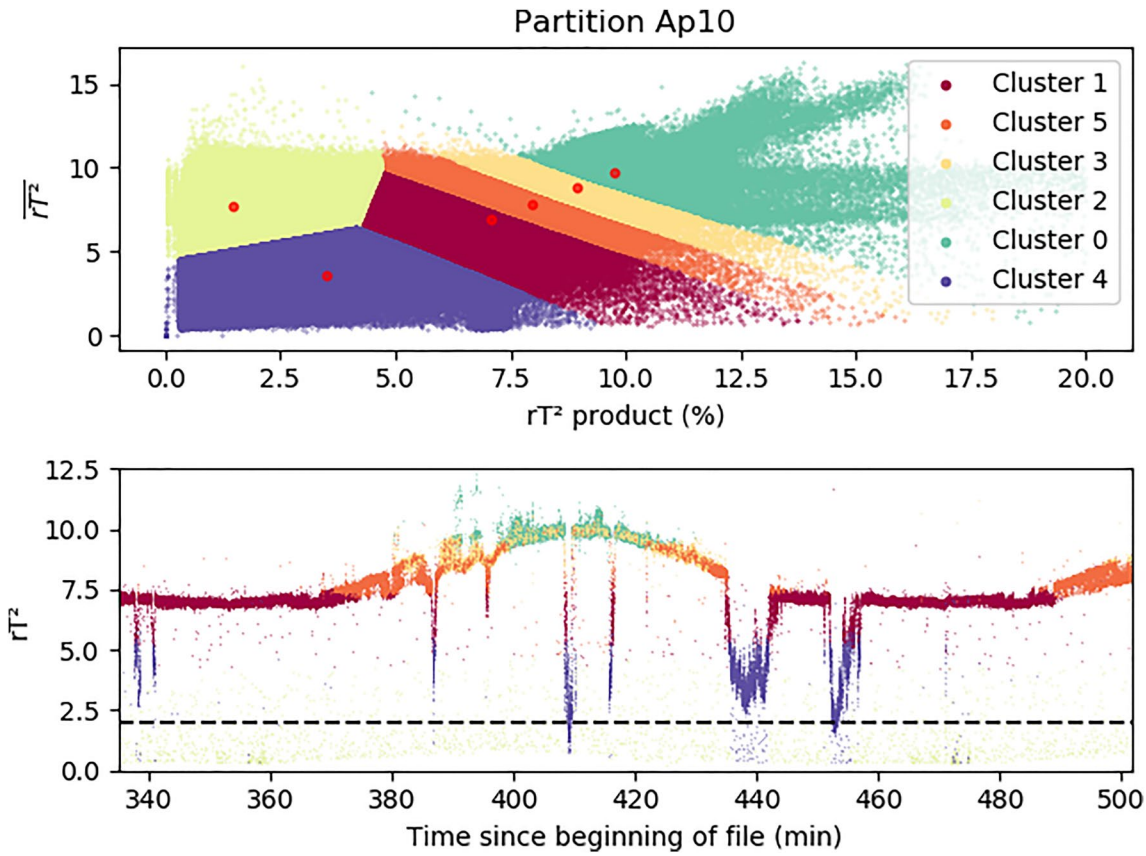

Figure 3. Example of clustering structure for ap10 orbital phase (a) and associated cluster distribution in the same rT^2 against time portion than Figure 1b. Cluster 4 gathers cloud returns (bottom of troughs), cluster 2 is made of noise returns (isolated points). Clusters 0, 3, and 5 are surface returns (different parts of the continuum). Cluster 1 is a limit between continuum and troughs and ensure that no surface returns are taken in our cloud cluster.

this will restrain atmospheric features to their most noticeable part, preventing doubtful cluster attributions for their borders.

3. Results

3.1. Clustering Results

3.1.1. Atmospheric Features Cluster

Typical final clustering structure is illustrated in Figure 3a. Note that only one orbital phase is exposed here. We clearly found the expected clusters. The blue cluster (cluster 4) corresponds to both a low rT^2 product and a low $\overline{rT^2}$. Figure 3b shows an example of rT^2 against time, colors corresponding to clusters: returns from cluster 4 are indeed at the bottom of the troughs and we can consider that the cluster regroups atmospheric features. Light yellow points (cluster 2) do not follow any specific distribution and are isolated points: this cluster regroups noise returns that are coherent with a high $\overline{rT^2}$ despite a low rT^2 product. All three green, dark yellow, and orange clusters (respectively, cluster 0, 3, and 5) are different parts of the continuum, so they correspond to surface returns. The surface returns are divided into three different clusters by the method because of the difference of altitude between both Mars' poles, south pole being approximately 6 km above the altitude of the north one. Therefore, the MOLA range, used to calculate rT^2 product, has a slight sinusoidal variation during an orbital phase from one pole to the other and gives this aspect to the continuum. Finally, red cluster (cluster 1) is the main reason that led us to choose six clusters instead of five. It includes two kinds of returns: the lowest part of the continuum but also the highest part of some troughs, what could be considered as the extreme borders of the atmospheric features. Still, to be sure that we limit as much as possible false positive detections, we do not keep cluster 1 returns in the rest of our study. The presence of such a cluster in the partition proves that cluster 4 returns are really inside the

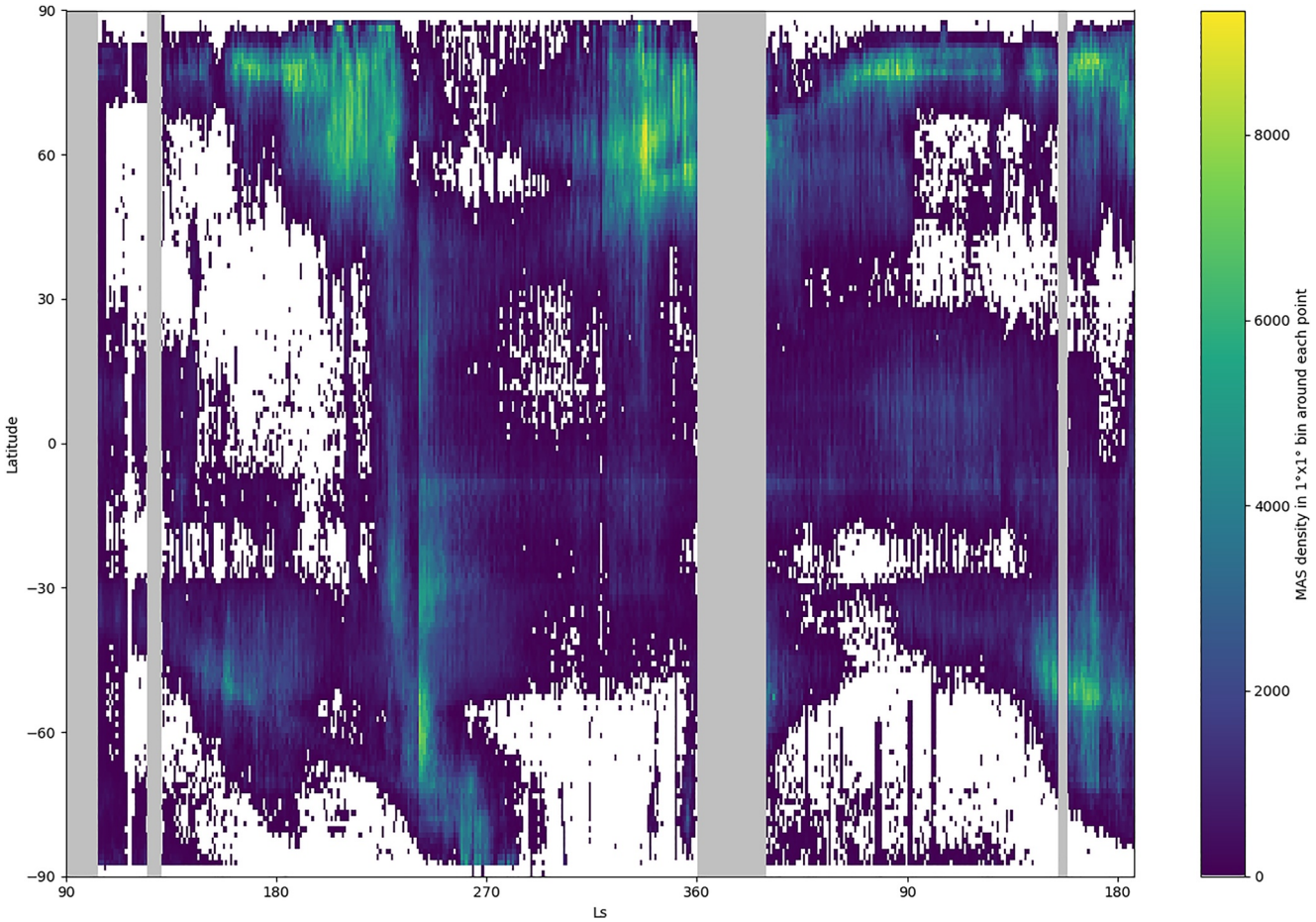

Figure 4. Seasonal and latitudinal distribution of atmospheric features returns from the clustering. Colors are based on number of returns in the 1° × 1° bin around each point. Gray brands correspond to no observations periods while white background represents a lack of Mars Orbiter Laser Altimeter (MOLA) Atmospheric Structure detection. Absence of returns at the start of MY25 is due to solar conjunction. The other two thin gray bands are related to spacecraft anomalies that forced MOLA to be turned off (see Section 2.1). L_s axis starts at 90° of MY24 to ease the visualization of the seasons.

troughs that would correspond to the core of atmospheric features. This study is a first approach on how basic clustering method can already bring significant results on such data sets, so being confident on our clustering returns appears as more valuable than trying to get every single one of them. However, this also means that the clustering has room for improvements with more complex methods that could make a clearer distinction between atmospheric structures borders and the lowest part of the continuum.

3.1.2. Comparison to a Previous Study

To verify the validity of our method, we compare our results with a previous study (Neumann et al., 2003) through seasonal and latitudinal distribution of atmospheric returns. Figure 4 shows that the distribution of returns from our atmospheric features cluster is coherent with the one obtained by overlapping both the reflective and absorptive clouds distributions from, respectively, Figures 5 and 6 of Neumann et al. (2003). Cloudiest areas such as the north pole during northern autumn, the band between 30 and 60°S at the end of southern winter, and global coverage at L_s 230° clearly appear in both distribution. The poles seem less dense initially. However, Neumann et al. (2003) divided the distribution in two distinct figures for reflective and absorptive clouds: the peak density of reflective clouds is nearly 10 times less than that of absorptive clouds. Thus, it is expected that in a single plot distribution that regroups both reflective and absorptive clouds such as Figure 4, poles would be less noticeable. Still, there are clearly clouds above poles in our distribution too and these results are discussed in detail later in the article (see Section 4.5). Even though we are talking about "atmospheric features" and not about clouds yet, we found around 33 times more returns than Neumann et al. (2003), thanks to use of clustering algorithms to analyze the data and also due to not having to impose strict constraints for detection.

3.2. Cloud and Dust Structure Catalog

3.2.1. Cloud Definition

At this point, we have a cluster of atmospheric returns, and we would like to group them into structures. Because we have no information about the composition of implied features, we define the alias MOLA Atmospheric Structure (MAS) that will be used hereafter and that regroups all possible atmospheric particle types: water clouds, CO_2 ice clouds, and dust structures. Choosing a good definition of the MAS in terms of returns will allow for the reduction of the remaining false-positive returns wrongly associated with the atmospheric features cluster. It will give information about the size of structures that can later be used as a comparison criteria. Here, size is used as a measure of the number of returns that constitute an atmospheric structure. It gives information about the spatial extent of the structure in one direction, following along-track position of the surface target. The along-track extent of the structure can then be calculated thanks to the space between two consecutive shots: $\text{size}_m = \text{size}_{\text{returns}} \times 300$. However, as we do not have information on the longitudinal extent of the cloud, we cannot, for example, estimate the actual cloud cover or fraction. This will also allow us to make a statistical analysis of the resulting MAS that we will gather in a unique catalog.

The first criterion to determine is the minimum number of consecutive atmospheric returns that is required to consider them forming a unique structure. For two returns to be considered as consecutive atmospheric returns, we must have two temporally successive returns that were both assigned to the atmospheric cluster. Temporally successive returns can also be considered as spatially successive returns in the tracking direction. Because there are approximately 300 m between two consecutive shots' targets, fixing a minimum of five consecutive returns defines atmospheric structures of at least 1.5 km, which corresponds to a typical size scale for Martian clouds (Clancy et al., 2017). The second criterion to determine is when we have reached the end of the considered MAS, that is to say that the next return in the atmospheric cluster may be part of another MAS. Noise returns, considered here as false triggers, have a uniform distribution over the mission duration with a constant probability of occurrence of about $P_{\text{FT}} = 1.25\%$ (see Section 2.3.1). The probability that 5 returns in a row are physically part of the considered MAS without being associated to the atmospheric cluster is equivalent to having 5 consecutive false positive, thus less than 1 in 10^5. Therefore, we fix the end of a MAS if 5 consecutive returns are all not assigned to atmospheric cluster. This also requires that we accept a "gap" made of a maximum of 4 consecutive returns that are not in the atmospheric cluster.

We apply this definition to all returns from our atmospheric cluster. Whenever a set of consecutive returns meets the first condition, they form a MAS, and we look for its end. Whenever the second condition is met, the end of the MAS has been reached, and we save it in our catalog. This way, a total of 261 862 MAS are formed.

3.2.2. Catalog Global Summary

The catalog is available alongside this article. It is presented as a text file in which the four first lines describe the parameters given for each cloud. The fifth line gives every parameters a column title to ease the read of the catalog by numerical libraries. All 261,862 MAS are then presented following the exact same scheme. An id number is attributed to every MAS for an easier identification of each of them and is widely used in our plotting program (given with the catalog).

We made the choice to reduce the number of parameters for each MAS compared to the raw data so that the catalog can be navigated easily. First three parameters allow finding the MAS in the PEDR data set: the file it comes from and the lines of the first and last returns of the MAS in that file. The size of the MAS is then given in terms of the number of returns. Note that size is different than the difference between the start and end lines because some returns in between may have not been associated to the atmospheric features cluster (and i.e., why we accepted some "gap" in our cloud definition). Longitude and latitude are then given through three parameters each: mean, start, and end. Solar longitude is calculated and only given as a mean for each cloud but ephemerid time is also given by mean, start, and end values. Finally, the triggered channel dominating in the cloud is also given. The version of the catalog with percent of occurrence of each channel for each cloud also exists in case of need. That makes a total of 16 parameters for each cloud. The distribution of clouds in the different orbital phases of the mission is given in Table 1.

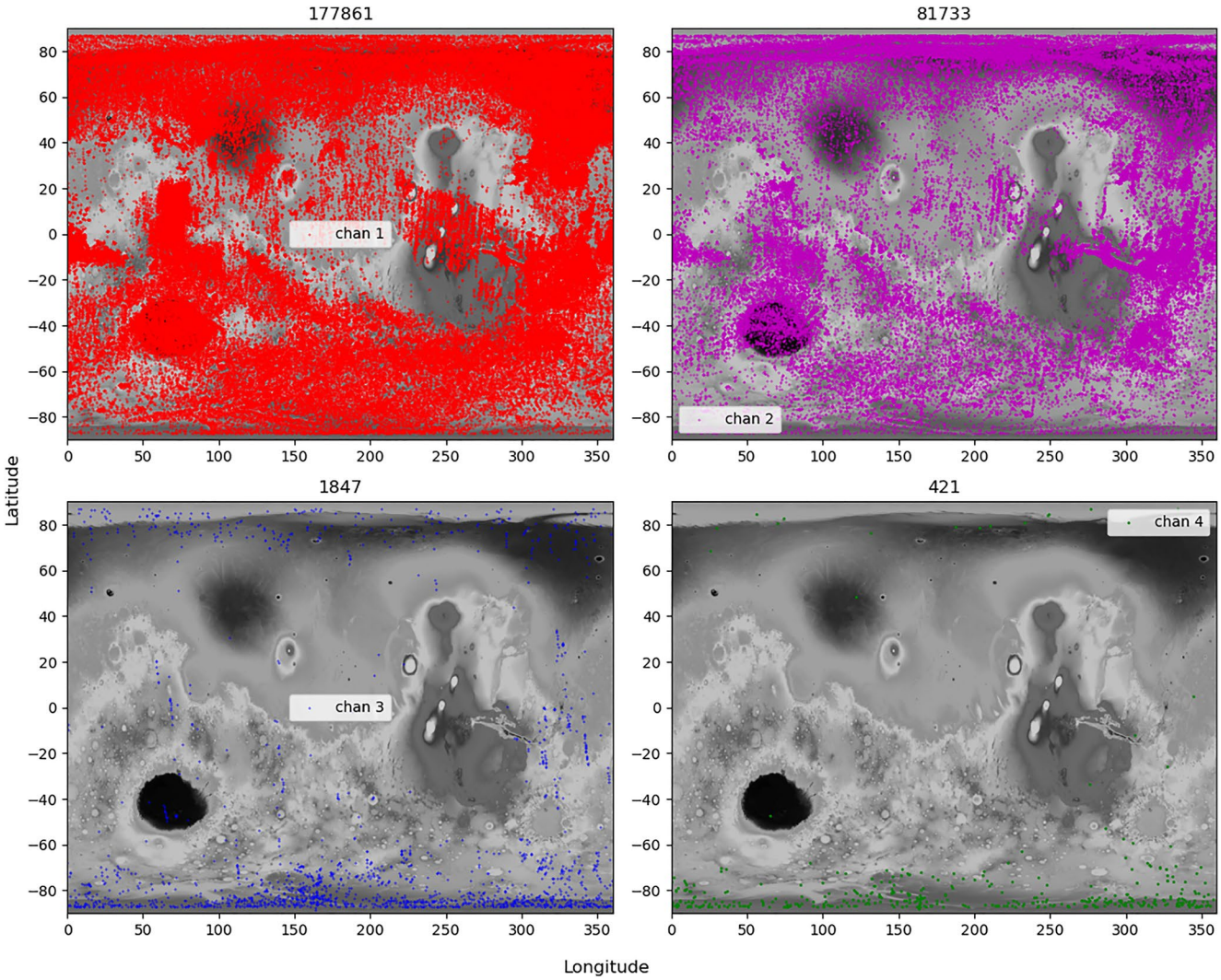

Figure 5. Atmospheric structures distribution by most triggered channel. Each structure is associated with the channel that is the most represented among the returns that constitute the structure. The size of markers for the channel 4 panel are bigger for visualization purpose only. Title of each panel corresponds to the number of structures.

4. Cloud Variability and Diversity

4.1. Size and Triggered Channel Distribution

In the following sections, we study the size of clouds in terms of number of returns, keeping in mind that two consecutive returns are spaced by approximately 300 m. We analyze the longitudinal and latitudinal distribution of MAS in function of their size. While structures smaller than approximately 500 returns, that is, 150 km, seem to follow the general distribution presented in Figure 4, 500+ returns MAS, hereafter called "bMAS" (big MAS) are not homogeneously distributed and exhibit some interesting areas. In particular, bMAS tend to form above certain structures such as basin or plains.

The MOLA channel that was triggered was a key point in the distinction of different kinds of clouds (Neumann et al., 2003). A first approach is to look at the dominant channel for each cloud (Figure 5). Channel 1 is by far the most represented channel, in opposition to Neumann et al. (2003). This tendency remains true even if we consider all atmospheric returns before classifying them as clouds. This means that our method notably allows detecting optically thinner structures. The number of clouds for a specific channel decreases with the increase of the target dispersion with a drastic fall between channel 2 and 3. Channel 1 and 2 clouds are distributed along the same lines as the general spatial distribution (Figure 4) while channel 3 clouds are mainly located at both poles, and channel

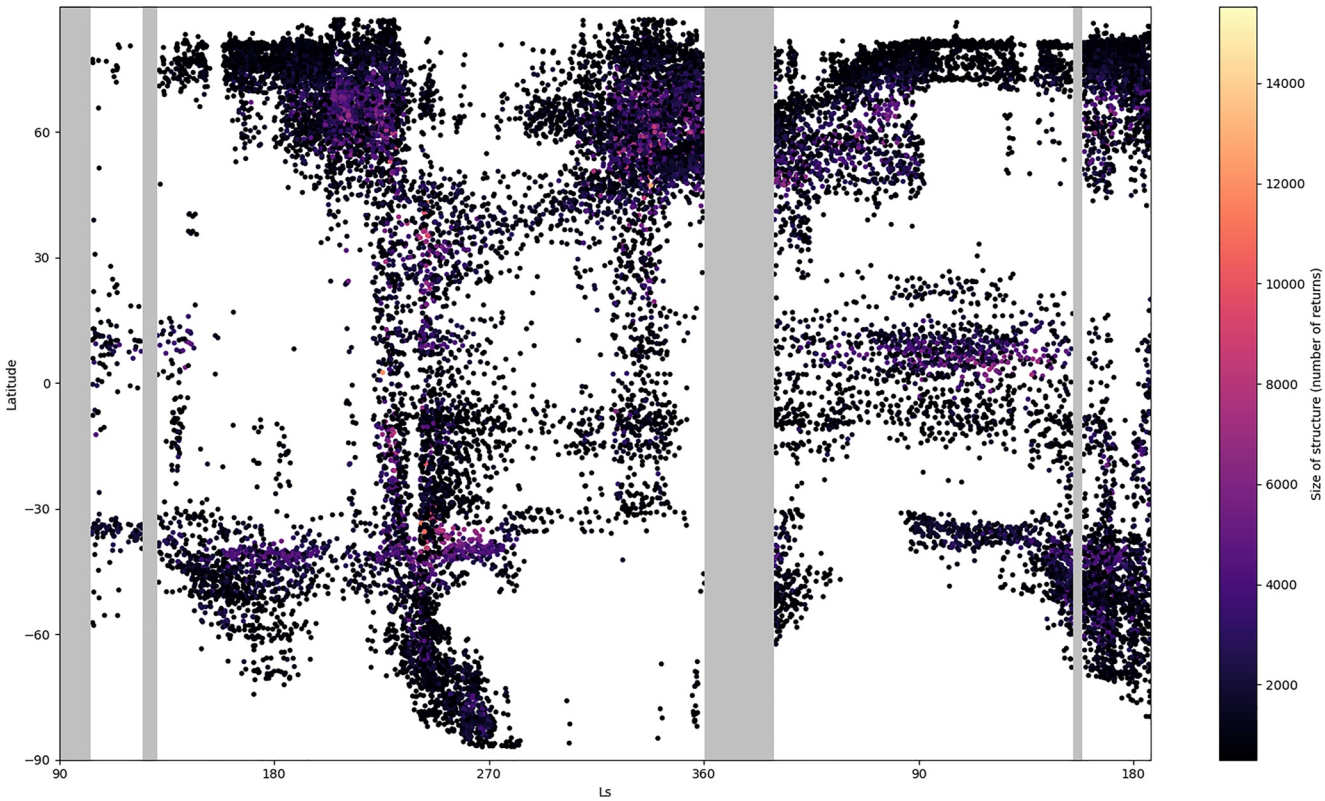

Figure 6. Distribution of big Mars Orbiter Laser Altimeter (MOLA) Atmospheric Structure (structures of at least 500 returns) over solar longitude and latitude. Gray bands represent periods without MOLA observations while white background means that there was no structure detection (see Figure 4 for more details about gray bands).

4 clouds are almost exclusively located at the south pole. Triggering channel 4 implies a dense vertical structure of at least 80 m. It means that something exclusive to the south pole allows the formation of globally optically thicker clouds, possibly the prevalence of CO_2 ice clouds over water clouds.

4.2. Temporal Variability

In the following section, we only consider bMAS (500+ returns) to highlight recurrent phenomenon that would imply the formation of big structures or reveal places where atmospheric conditions are more suitable for their formations. Distribution of bMAS over solar longitude and latitude is illustrated in Figure 6. Despite having only 1.5 Martian Years of data, we would like to compare the interannual bMAS variability. Note that the first year lacks the Southern hemisphere fall (start at $L_s = 103°$) while second year data stops after Southern hemisphere winter (end at $L_s = 187°$). Considering only the fraction of bMAS per day for each year, it appears as if the second year was less prolific in terms of bMAS. However, by comparing the only season we have in common for the 2 years, that is, Northern hemisphere summer/Southern hemisphere winter ($90° < L_s < 180°$), we see that, in fact, the second year presents more bMAS (comparison is made between five first panels of Figure 7 and five last panels of Figure 8). This discrepancy is explained by the absence of observations of the north pole winter for the second year while most bMAS seem to form above the north pole during this period (see Figures 7 and 8 and Section 4.5). bMAS could form more easily at the north pole thanks to the larger quantity of water vapor in the atmosphere, which would also mean that most of the bMAS are water clouds. Globally, the MAS distribution for each Martian Year looks very similar apart from Tharsis Montes (Section 4.3) that did not appear as a cloudy area for the first year while being one of the cloudier for the second year. The other difference is an apparent diminution of bMAS population at the south pole (latitude lower than 70°S in spite of winter).

During the observed part of MY24, northern summer is the one with less bMAS, mainly appearing over Hellas Basin, a little bit over Syrtis Major, Argyre Planitia, and at the north pole, more precisely over Acidalia Planitia.

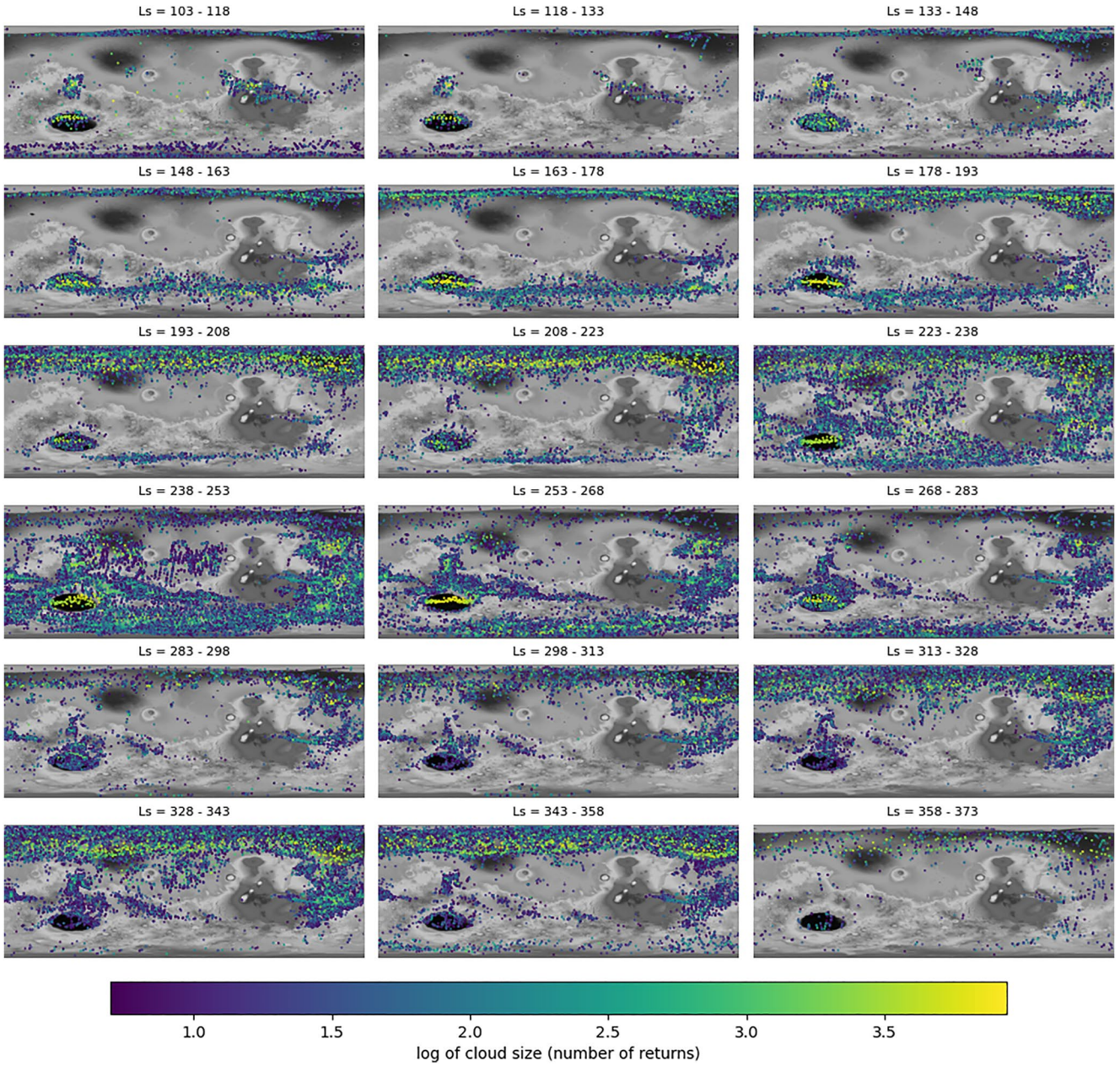

Figure 7. Fifteen degree window clouds distribution for MY24. Returns from our clustering are presented above Mars' topography map provided by NASA Goddard Space Flight Center. The end of the disparition of the aphelion cloud belt. Whole development of the south polar hood happens between $L_s = 160°$ and $L_s = 270°$. A dusty episode caused by a regional dust storm occurs around $L_s = 240°$.

Northern fall is more dense in terms of bMAS with the same areas of interest to which we can add both polar hood. North pole clouds also extend to a lower latitude over Utopia Planitia. Distributions at the poles seem to be correlated with the presence of ice at the surface. During northern summer, big clouds are almost all located around the north pole.

Concerning the second year, MY25, northern spring mainly exhibits big clouds in the northern hemisphere, above Syrtis Major and Acidalia Planitia. However, northern summer is notable because it appears to be very different from MY24 northern summer with way more bMAS. While bMAS are still found over the previously mentioned areas, there are also bMAS appearing over Tharsis Montes and in a band at 60°S latitude going from 0 to 150°E. This cloud cover is very similar to the early stages of the martian global dust storm that happened later this year, observed by MOC and TES, and described in Strausberg et al. (2005). These bMAS

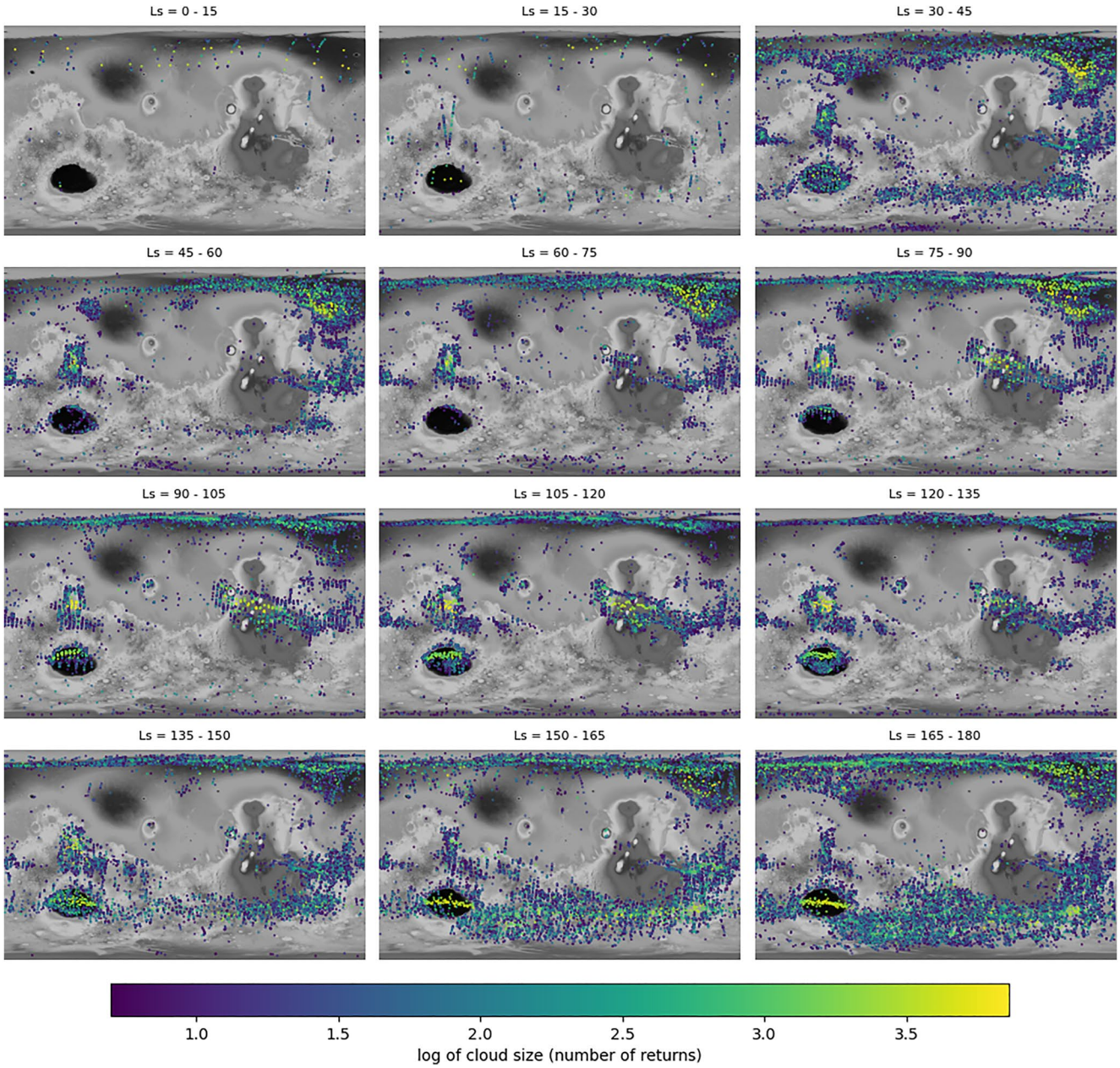

Figure 8. Fifteen degree window clouds distribution for MY25. Very few returns are observed on the first panels due to the solar conjunction. Aphelion cloud belt can be seen till $L_s = 160°$. Early stages of the upcoming global dust storm of MY25 are observed in the southern hemisphere from $L_s = 135°$.

could be dust structures but are too early in the year to be attributed to the global dust storm that started around $L_s = 177°$. However, their presence may suggest specific precursors for the necessary atmospheric conditions to support a global dust event, particularly given the absence of such features in MY24. Distribution of atmospheric structures at this solar longitudes are compared with the Mars Dust Activity Database results (J. M. Battalio & Wang, 2019) in Figure 9 bottom panel. Latitude and longitude extend of dust structures are in total agreement with our observations and with the band in the southern hemisphere, allowing us to hypothesize on their compositions. In addition, structures are observed earlier at Solis Planum south border in our observations than in the database, for which dust activity becomes important from L_s 180°. The two data sets may be complementary for a total understanding of how the dust developed in the low atmosphere step by step.

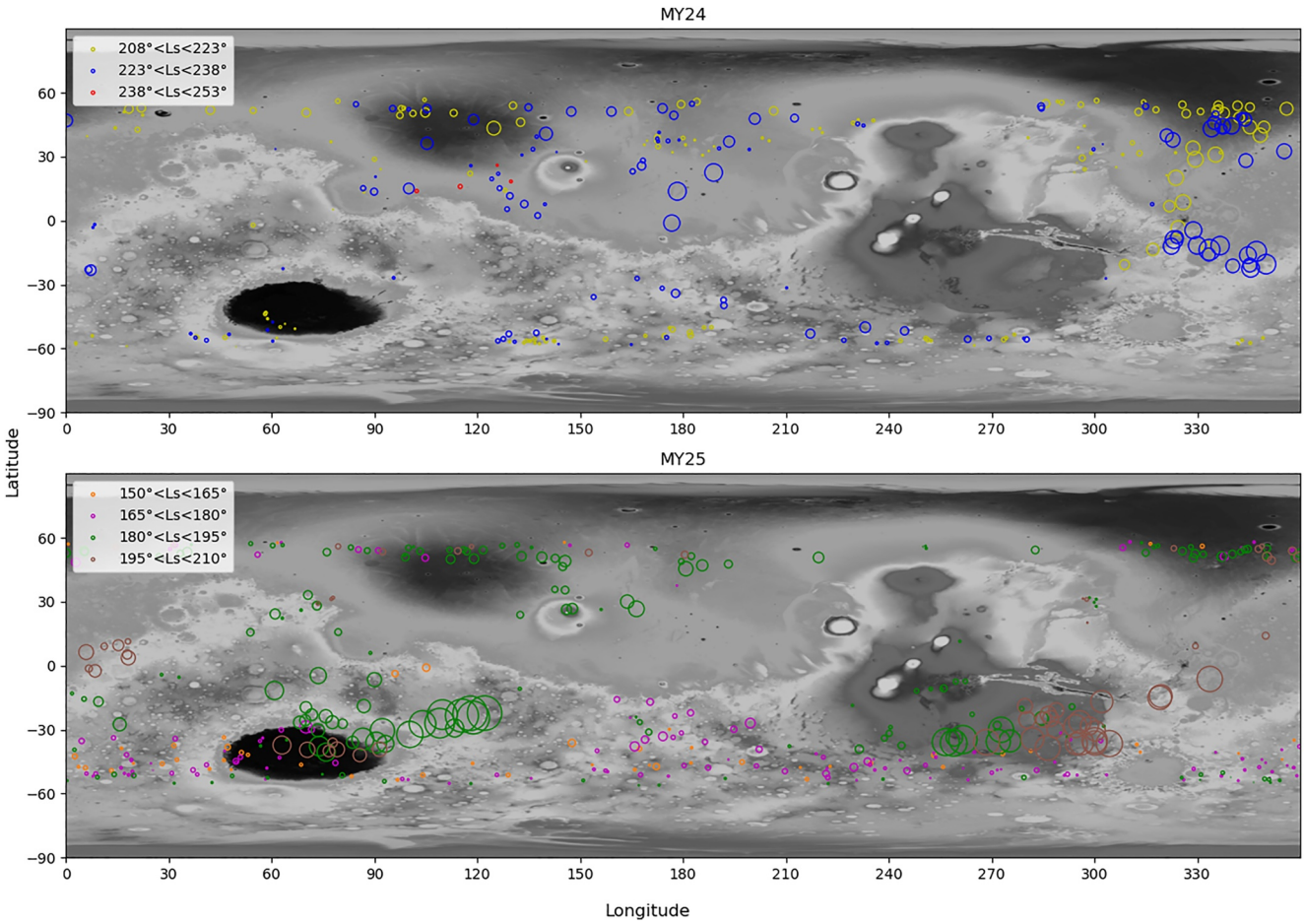

Figure 9. Results from the Mars Dust Activity Database (J. M. Battalio & Wang, 2019) for MY24 (top) and MY25 (bottom) for comparison with our observations on specific time periods. Markers sizes are proportional to the covered area given in the database for each event.

4.3. Tropical Regions

The first year observed by MOLA starts around L_s = 103°, when the tropical belt observed with MOC (Wang & Ingersoll, 2002) begins to disaggregate. However, MY25 shows its whole development. By comparison with other observations, we assume that MAS discussed in the following part are probably water clouds. As soon as L_s = 15°–30°, MAS start to form above Valles Marineris and Syrtis Major. MAS then appear east of Valles Marineris. The belt mainly consists of a band between 10°S and 25°N, and remains almost constant until L_s = 120°. Then it decreases slowly, a little bit later than what observed with MOC by Wang and Ingersoll (2002) but corresponds to the TES observation of the period (Hale et al., 2011). The belt is supposed to be longitudinally continuous around L_s = 60° but this never appears to be true in our observations, MAS being rare above Elysium and Amazonis Planitia. This probably means that these clouds are too high to be detected by our method since range gate of MOLA sets the maximum altitude around 10 km for most of the returns, while water clouds such as the ones in the tropical cloud belt can reach up to 20 km. MY24 allows the comparison of the cloud belt decay period for two distinct years. While it was still visible at L_s = 105°–120° for MY25, it is patchier than the first one. The aphelion cloud belt is a recurrent phenomenon in Mars' atmosphere, but its intensity seems to vary from 1 year to another.

The first part of the belt to develop is also the last to disappear in the end of northern summer. Indeed, clouds are permanent above the volcanoes of Syrtis Major and above Valles Marineris from L_s = 15° till L_s = 160° for MY25. In fact, MY24 shows that clouds can be associated with Valles Marineris for the majority of the Martian Year. Assuming that they do not decay during the solar conjunction for which we have no observation, they can be observed from L_s = 200° of MY24 till the end of the data set. Note that clouds from Valles Marineris are all

small in term of size, which agrees with observation of these clouds being in the canyons as observed by the High Resolution Stereo Camera and the imaging spectrometer Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité (OMEGA) (Inada et al., 2008) but also by MOC (Benson et al., 2003), suggesting that these structures could either be water clouds or dust. Syrtis Major tends to be slightly different with two distinct cloudy periods. The first one seems to be part of the aphelion cloud belt evolution, while the second one lasts from $L_s = 220°$ till $L_s = 280°$. This second period has not been observed in MOC images (Wang & Ingersoll, 2002), so these clouds may be different than the ones of the tropical cloud belt. Conversely, clouds can be found above Tharsis Montes only for the second year and mainly during the climax of the tropical cloud belt. The area is one of the first to disintegrate during the decay of the belt, and clouds are exclusively located above the volcanoes. Therefore, all three zones indicate that aphelion belt cloud formation and evolution are correlated with surface relief.

4.4. Southern Hemisphere

4.4.1. South Polar Hood

Another recurrent phenomenon can be seen twice in these observations. From $L_s = 140°$, a "moving" southern belt can be seen in a band between 45 and 55°S, almost longitudinally homogeneous from Hellas Basin to Argyre Planitia. It is first centered on latitude 50°S but appears to evolve toward the south pole during the end of southern winter where it finally decays at the end of southern spring. This belt has not been observed in MOC images (Wang & Ingersoll, 2002) nor in TES water ice cloud maps (Hale et al., 2011). However, a similar belt has been described as a second phase in the development of Mars' south polar hood (Benson et al., 2010) following the enrichment of the southern hemisphere atmosphere in water by the circulation of clouds in the region south of Tharsis. At the same time, the latitudinal evolution of this belt seems to correspond to the seasonal CO_2 polar cap edge (Giuranna et al., 2021). Thus, some clouds in the belt could also be CO_2 clouds, and the belt itself could be evidence of CO_2 circulation in Mars' southern atmosphere during local winter.

During MY24, an important number of returns can be observed from L_s 220° till L_s 260°, almost covering the entire surface of Mars. These MAS do not present any signatures neither in Hale et al. (2011) nor in Wang and Ingersoll (2002); consequently it is very unlikely that they are water clouds. Because of the period concerned and how it covers the majority of the planet surface, it is also very unlikely that all of them are CO_2 clouds. Thus, this is probably caused by a regional dust storm in the southern hemisphere. This result is consistent with TES observations (Smith, 2004) that show a sudden increase of dust optical depth starting at L_s 210° of MY24, and which is initially more intense in the southern hemisphere. This is also coherent with Montabone et al. (2015) observations of Mars Year 24 where a rise of the mean column dust optical depth was seen in this solar longitude interval. Observations in this period are compared with the Mars dust activity database results (J. M. Battalio & Wang, 2019) for MY24 in top panel of Figure 9. As early as L_s 200°, dust activity was observed around Acidalia Planitia longitude, which can also be seen in the corresponding panel of Figure 7 on the same time period. Dust then extends to Arcadia Planitia longitude. Differences occur between L_s 238° and L_s 253°, when dust activity is negligible in the dust database while still being very present in our observations. This is probably linked to some limitations on how the MDAD was built (M. Battalio & Wang, 2021). Indeed, dust edges were collected by eye, which becomes difficult whenever haze is too diffuse for the boundary between clear-air and dusty-air to be visible. Our clustering method collects all returns of higher opacity in the atmosphere, making it a good complement to MDAD for distinguishing dust haze and surface. Comparing these two data sets provides us additional observations of the phenomenon, allowing us to conclude on the dust composition of some of the atmospheric structures. Start and end of this rise in dust activity, as well as the longitude at which it takes place at, correspond to the definition of the "A" storm from Kass et al. (2016), which is a recurrent regional dust storm happening during Martian Years that are not affected by a global dust storm.

4.4.2. Hellas Basin

MAS can be seen above Hellas Basin for most of the observed period. During MY25, MAS start forming on the northern border of Hellas around $L_s = 60°$ and thicken until covering the entirety of the Basin at $L_s = 120°$ and then dissipating quickly. At $L_s = 160°$ the only MAS remaining form a line from the western to the eastern border of the Basin. The end phase of this same phenomenon can be seen in MY24 around the same L_s. During southern spring, MAS can be seen in Hellas Basin but cannot be distinguished from the southern polar hood. Observation of these MAS are coherent with observations of water clouds in Hellas Basin by the Mars Color Imager (Kahre

et al., 2019; Wolff et al., 2019). Hellas Basin is also the location for the largest MAS apart from the poles in our observations (around 12% of 500+ MAS across the mission duration) suggesting an important reservoir of water, in agreement with it playing a role in the atmospheric circulation of water from the north polar hood to the southern hemisphere.

4.5. Poles

4.5.1. North Pole

Combining the two observed Martian Years, we can see interannual variability of both poles apart from the window between solar longitudes ∼20° and 40° that corresponds to solar conjunction. The north polar hood develops through late northern summer. MAS are first observed above the pole itself till L_s = 160° when they can also be observed at lower latitudes, above Acidalia Planitia or the western side of Utopia Planitia. MAS then persist through northern fall and winter with almost global coverage of latitude higher than 50°N around L_s = 330°. Dust storm can either thicken or dissipate Mars' north polar hood (Martin, 1975). Indeed, large dust events induce a decrease in baroclinic wave activity because the atmosphere stabilizes by modifying the structure of the baroclinic zone, what reduces instabilities (M. Battalio & Wang, 2020, 2021; Hinson & Wilson, 2021). Moreover, the CO_2 ice cap that forms during the northern fall and winter is constrained by the alignment of CO_2 ice clouds with planetary waves (Kuroda et al., 2007, 2013). A planet-encircling dust storm such as the "A" storm observed earlier causes a massive decreased in traveling wave activity. Thus, this regional dust storm seems to have lowered the number of north pole MAS between L_s = 250° and 280°. This response from the northern hemisphere to a mid-latitude developing dust storm was also observed by Kass et al. (2016). Apart from this period, and assuming clouds did not disappear during solar conjunction, the north polar hood remains present from L_s = 150° to L_s = 30° as observed in Benson et al. (2011). After that, the only remaining MAS at north pole are located above Acidalia Planitia, until L_s = 90°, eventually suggesting specific local atmospheric conditions that allow clouds to remain longer than above pole itself.

4.5.2. South Pole

It is much harder to distinguish a clear evolution for the south pole. Clouds can be seen in the first observations of MY24 around L_s = 100° and could be the end of the first phase of Mars' south polar hood (Benson et al., 2010) but dissipate much faster than observed in MOC images (Wang & Ingersoll, 2002) since they can barely be seen as early as L_s 130°. Moreover, they are not observed at the same period in MY25. This would either mean that our method is not well adapted for south pole clouds or that they are too high in the atmosphere to be detected.

5. Conclusions

The MOLA instrument provides data for 1.5 Martian Years and has proven being able of detecting cloud signatures coming from the lower atmosphere, up to 20 km. Studying the data set with machine learning methods, and particularly clustering algorithms, bypasses the determination of a stringent limit between clouds and surface laser returns. We show that the MOLA data set is indeed clusterable but also that a certain number of clusters emerges using optimization methods. Clustering provides more detections than previous studies while maintaining a similar seasonal and latitudinal distribution, which confirms the viability of the process. We reduce the chances of having false positive in our atmospheric features cluster by picking the most appropriate number of clusters.

We gather consecutive atmospheric returns into structures to keep track of cloud or dust structure sizes. This way, some areas are revealed in the seasonal distribution as being more favorable for big atmospheric structures. Atmospheric structures are presented in a unique catalog regrouping results from the whole MOLA data set.

MOLA cloud observations are compared with two other instruments, MGS TES and MGS MOC, which have also observed clouds in the Martian atmosphere for the same time period. This comparison notably helps confirm assumptions on the composition of atmospheric structures and shows that both water and CO_2 ice clouds and important dust structures are found in the MOLA data set. Development of the aphelion cloud belt and the south polar hood are observed with a large temporal resolution. Because of the important proportion of channel 4 structures and by using both dust activity database and dust optical depth observations, we also showed that MOLA observed an important number of the dusty events that occurred during MY24 and MY25.

The results are also compared with more recent mission observations for the development of atmospheric structures within specific places like Valles Marineris or Hellas Basin. Hellas Basin seems to play a crucial role in the atmospheric circulation of water across a Martian Year, which is highlighted by the omnipresence of clouds above the basin. Seasonal variations above the north pole are important both in terms of cloud sizes and localization. However, even though it is apparently less cloudy, MOLA still seems to be less effective for the observation of clouds above the south pole.

The machine learning method optimization ensures we have as few false-positive atmospheric returns as possible. But that may have restrained the limits of the atmospheric cluster by a small margin. One way to improve our results, particularly on the determination of the borders of atmospheric features, could be using them as a training set for deep learning methods to recognize cloud signatures in rT^2 against time plots.

Data Availability Statement

The results of this paper, including the atmospheric structures catalog and the tool for cloud visualization are archived at the ESPRI MESOCENTRE and are publicly available (Caillé et al., 2022). Raw MOLA data files (PEDR) were downloaded from the MOLA page of PDS Geosciences Node (Ford et al., 1998). Mars' topography map used as background for clouds distribution was made by the MOLA Instrument and Science team and made available by NASA Goddard Space Flight Center Scientific Visualization Studio (https://svs.gsfc.nasa.gov/stories/MOLA/). The MOLA cloud returns from Neumann et al. (2003) are available at: https://pgda.gsfc.nasa.gov/products/62.

Acknowledgments
We thank the Agence National de la Recherche for funding (project MECCOM, ANR-18-CE31-0013) this research. Computational Resources for this project were provided through CICLAD server by the ESPRI mesocentre at the IPSL institute. We acknowledge the help of G. A. Neumann whose experience allowed us to understand the data set more easily. Finally, we deeply thank Michael Battalio and an anonymous reviewer for their very constructive reviews that helped improving this paper.

References

Abshire, J. B., Sun, X., & Afzal, R. S. (2000). Mars orbiter laser altimeter: Receiver model and performance analysis. *Applied Optics*, *39*(15), 2449–2460. https://doi.org/10.1364/AO.39.002449

Anderberg, M. R. (1973). Chapter 7 - Nonhierarchical clustering methods. In M. R. Anderberg (Ed.), *Cluster analysis for applications* (pp. 156–175). Academic Press. https://doi.org/10.1016/B978-0-12-057650-0.50013-2

Anderson, E., & Leovy, C. (1978). Mariner 9 television limb observations of dust and ice hazes on Mars. *Journal of the Atmospheric Sciences*, *35*(4), 723–734. https://doi.org/10.1175/1520-0469(1978)035<0723:MTLOOD>2.0.CO;2

Arthur, D., & Vassilvitskii, S. (2006). How slow is the *k*-means method? In *Proceedings of the twenty-second annual symposium on Computational geometry - SCG*, (Vol. *6*, pp. 144–153). https://doi.org/10.1145/1137856.1137880

Arthur, D., & Vassilvitskii, S. (2007). *k*-means++: The advantages of careful seeding. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027–1035). Society for Industrial and Applied Mathematics.

Azari, A. R., Biersteker, J., Dewey, R., Doran, G., Forsberg, E., Harris, C., et al. (2021). Integrating machine learning for planetary science: Perspectives for the next decade. *Bulletin of the AAS*, *53*(4). https://doi.org/10.3847/25c2cfeb.aa328727

Azari, A. R., Lockhart, J. W., Liemohn, M. W., & Jia, X. (2020). Incorporating physical knowledge into machine learning for planetary space physics. *Frontiers in Astronomy and Space Sciences*, *7*, 36. https://doi.org/10.3389/fspas.2020.00036

Battalio, J. M., & Wang, H. (2019). *The mars dust activity database (MDAD)*. Harvard Dataverse. https://doi.org/10.7910/DVN/F8R2JX

Battalio, M., & Wang, H. (2020). Eddy evolution during large dust storms. *Icarus*, *338*, 113507. https://doi.org/10.1016/j.icarus.2019.113507

Battalio, M., & Wang, H. (2021). The Mars dust activity database (MDAD): A comprehensive statistical study of dust storm sequences. *Icarus*, *354*, 114059. https://doi.org/10.1016/j.icarus.2020.114059

Benson, J. L., Bonev, B. P., James, P. B., Shan, K. J., Cantor, B. A., & Caplinger, M. A. (2003). The seasonal behavior of water ice clouds in the Tharsis and Valles Marineris regions of Mars: Mars Orbiter Camera observations. *Icarus*, *165*(1), 34–52. https://doi.org/10.1016/S0019-1035(03)00175-1

Benson, J. L., Kass, D. M., & Kleinböhl, A. (2011). Mars' north polar hood as observed by the Mars climate sounder. *Journal of Geophysical Research*, *116*(E3), E03008. https://doi.org/10.1029/2010JE003693

Benson, J. L., Kass, D. M., Kleinböhl, A., Mccleese, D. J., Schofield, J. T., & Taylor, F. W. (2010). Mars' south polar hood as observed by the Mars climate sounder. *Journal of Geophysical Research*, *115*(E12), E12015. https://doi.org/10.1029/2009je003554

Bufton, J., Blair, B., Cavanaugh, J., & Garvin, J. (1995). Shuttle Laser Altimeter (SLA): A pathfinder for space-based laser altimetry and LiDAR. In *The 1995 Shuttle Small Payloads Symposium* (Vol. *1*, pp. 83–90).

Caillé, V., Määttänen, A., Spiga, A., & Falletti, L. (2022). Mars Orbiter laser altimeter atmospheric (MOLA) features catalog and tool for cloud visualisation + MOLA brut dataset [Dataset]. ESPRI/IPSL. https://doi.org/10.14768/bb19aea0-264a-48da-961f-794f387a7349

Christensen, P. R., Bandfield, J. L., Hamilton, V. E., Ruff, S. W., Kieffer, H. H., Titus, T. N., et al. (2001). Mars global surveyor thermal emission spectrometer experiment: Investigation description and surface science results. *Journal of Geophysical Research*, *106*(E10), 23823–23871. https://doi.org/10.1029/2000JE001370

Clancy, R. T., Montmessin, F., Benson, J., Daerden, F., Colaprete, A., & Wolff, M. J. (2017). Mars clouds. In R. M. Haberle, R. T. Clancy, F. Forget, M. D. Smith, & R. W. Zurek (Eds.), *The atmosphere and climate of Mars* (pp. 76–105). Cambridge University Press. https://doi.org/10.1017/9781139060172.005

Curran, R., Conrath, B., Hanel, R., Kunde, V., & Pearl, J. (1973). Mars: Mariner 9 spectroscopic evidence for H_2O ice clouds. *Science*, *182*(4110), 381–383. https://doi.org/10.1126/science.182.4110.381

Farmer, C. B., Davies, D. W., Holland, A. L., Laporte, D. D., & Doms, P. E. (1977). Mars: Water vapor observations from the Viking orbiters. *Journal of Geophysical Research*, *82*(28), 4225–4248. https://doi.org/10.1029/JS082i028p04225

Ford, P., Neumann, G., & Smith, D. E. (1998). MOLA precision experiment data record [Dataset]. NASA Planetary Data System. https://doi.org/10.17189/1519520

Forgy, E. W. (1965). Cluster analysis of multivariate data: Efficiency versus interpretability of classifications.
French, R. G., Gierasch, P. J., Popp, B. D., & Yerdon, R. J. (1981). Global patterns in cloud forms on Mars. *Icarus*, *45*(2), 468–493. https://doi.org/10.1016/0019-1035(81)90047-6
Gardner, C. (1992). Ranging performance of satellite laser altimeters. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(5), 1061–1072. https://doi.org/10.1109/36.175341
Garvin, J., Bufton, J., Blair, J., Harding, D., Luthcke, S., Frawley, J., & Rowlands, D. (1998). Observations of the Earth's topography from the Shuttle Laser Altimeter (SLA): Laser-pulse echo-recovery measurements of terrestrial surfaces. *Physics and Chemistry of the Earth*, *23*(9), 1053–1068. https://doi.org/10.1016/S0079-1946(98)00145-1
Giuranna, M., Wolkenberg, P., Grassi, D., Aronica, A., Aoki, S., Scaccabarozzi, D., et al. (2021). The current weather and climate of Mars: 12 years of atmospheric monitoring by the Planetary Fourier Spectrometer on Mars Express. *Icarus*, *353*, 113406. https://doi.org/10.1016/j.icarus.2019.113406
Hale, A., Tamppari, L., Bass, D., & Smith, M. (2011). Martian water ice clouds: A view from Mars global surveyor thermal emission spectrometer. *Journal of Geophysical Research*, *116*(E4), E04004. https://doi.org/10.1029/2009JE003449
Hamerly, G., & Elkan, C. (2002). Alternatives to the *k*-means algorithm that find better clusterings. In *Proceedings of the eleventh international conference on Information and knowledge management - CIKM 02*, (pp. 600–607). https://doi.org/10.1145/584792.584890
Har-Peled, S., & Sadri, B. (2004). How fast is the *k*-means method? *Algorithmica*, *41*(3), 185–202. https://doi.org/10.1007/s00453-004-1127-9
Herr, K. C., & Pimentel, G. C. (1970). Evidence for solid carbon dioxide in the upper atmosphere of Mars. *Science*, *167*(3914), 47–49. https://doi.org/10.1126/science.167.3914.47
Hinson, D. P., & Wilson, R. J. (2021). Baroclinic waves in the northern hemisphere of Mars as observed by the MRO Mars Climate Sounder and the MGS thermal emission spectrometer. *Icarus*, *357*, 114152. https://doi.org/10.1016/j.icarus.2020.114152
Inada, A., Garcia-Comas, M., Altieri, F., Gwinner, K., Poulet, F., Bellucci, G., et al. (2008). Dust haze in Valles Marineris observed by HRSC and omega on board Mars Express. *Journal of Geophysical Research*, *113*(E2), E02004. https://doi.org/10.1029/2007JE002893
Ivanov, A. B., & Muhleman, D. O. (2001). Cloud reflection observations: Results from the Mars orbiter laser altimeter. *Icarus*, *154*(1), 190–206. https://doi.org/10.1006/icar.2001.6686
Kahre, M., Haberle, R., Hollingsworth, J., & Wolff, M. (2019). MARCI-observed clouds in the Hellas Basin during northern hemisphere summer on Mars: Interpretation with the NASA/Ames Legacy Mars Global Climate Model. *Icarus*, *338*, 113512. https://doi.org/10.1016/j.icarus.2019.113512
Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554. https://doi.org/10.1109/TKDE.2018.2861006
Kass, D. M., Kleinböhl, A., McCleese, D. J., Schofield, J. T., & Smith, M. D. (2016). Interannual similarity in the Martian atmosphere during the dust storm season. *Geophysical Research Letters*, *43*(12), 6111–6118. https://doi.org/10.1002/2016GL068978
Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data an introduction to cluster analysis*. Wiley-Interscience.
Kleinbohl, A., Schofield, J., Kass, D., Abdou, W., Backus, C., Sen, B., et al. (2009). Mars Climate Sounder limb profile retrieval of atmospheric temperature, pressure, and dust and water ice opacity. *Journal of Geophysical Research*, *114*(E10), E10006. https://doi.org/10.1029/2009JE003358
Kuroda, T., Medvedev, A. S., Hartogh, P., & Takahashi, M. (2007). Seasonal changes of the baroclinic wave activity in the northern hemisphere of Mars simulated with a GCM. *Geophysical Research Letters*, *34*(9), L09203. https://doi.org/10.1029/2006GL028816
Kuroda, T., Medvedev, A. S., Kasaba, Y., & Hartogh, P. (2013). Carbon dioxide ice clouds, snowfalls, and baroclinic waves in the northern winter polar atmosphere of Mars. *Geophysical Research Letters*, *40*(8), 1484–1488. https://doi.org/10.1002/grl.50326
Macqueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *5-th Berkeley Symposium on Mathematical Statistics and Probability* (pp. 281–297).
Malin, M., & Edgett, K. (2001). Mars global surveyor Mars orbiter camera: Interplanetary cruise through primary mission. *The Journal of Geophysical Research*, *106*(E10), 23429–23570. https://doi.org/10.1029/2000JE001455
Martin, L. J. (1975). North polar hood observations during Martian dust storms. *Icarus*, *26*(3), 341–352. https://doi.org/10.1016/0019-1035(75)90177-3
McCleese, D. J., Schofield, J. T., Taylor, F. W., Calcutt, S. B., Foote, M. C., Kass, D. M., et al. (2007). Mars Climate Sounder: An investigation of thermal and water vapor structure, dust and condensate distributions in the atmosphere, and energy balance of the polar regions. *Journal of Geophysical Research*, *112*(E5), E05S06. https://doi.org/10.1029/2006JE002790
Milligan, G. W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, *45*(3), 325–342. https://doi.org/10.1007/bf02293907
Montabone, L., Forget, F., Millour, E., Wilson, R., Lewis, S., Cantor, B., et al. (2015). Eight-year climatology of dust optical depth on Mars. *Icarus*, *251*, 65–95. https://doi.org/10.1016/j.icarus.2014.12.034
Neumann, G. A., Smith, D. E., & Zuber, M. T. (2003). Two Mars years of clouds detected by the Mars orbiter laser altimeter. *Journal of Geophysical Research*, *108*(E4), 5023. https://doi.org/10.1029/2002JE001849
Ni, H., & Benson, S. M. (2020). Using unsupervised machine learning to characterize capillary flow and residual trapping. *Water Resources Research*, *56*(8), e2020WR027473. https://doi.org/10.1029/2020WR027473
Pankine, A. A., Tamppari, L. K., Bandfield, J. L., McConnochie, T. H., & Smith, M. D. (2013). Retrievals of Martian atmospheric opacities from MGS TES nighttime data. *Icarus*, *226*(1), 708–722. https://doi.org/10.1016/j.icarus.2013.06.024
Peña, J., Lozano, J., & Larrañaga, P. (1999). An empirical comparison of four initialization methods for the *k*-means algorithm. *Pattern Recognition Letters*, *20*(10), 1027–1040. https://doi.org/10.1016/s0167-8655(99)00069-0
Selim, S. Z., & Ismail, M. A. (1984). *K*-means-type algorithms: A generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-*, *6*(1), 81–87. https://doi.org/10.1109/TPAMI.1984.4767478
Smith, D. E., Neumann, G. A., Ford, P. G., Arvidson, R. E., Guinness, E. A., & Susan, S. (1999). *Mars global surveyor laser altimeter precision experiment data record*. Tech. Rep. MGS-M-MOLA-3-PEDR-L1A-V1.0. NASA Planetary Data System.
Smith, D. E., Zuber, M. T., Frey, H. V., Garvin, J. B., Head, J. W., Muhleman, D. O., et al. (2001). Mars orbiter laser altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research*, *106*(E10), 23689–23722. https://doi.org/10.1029/2000JE001364
Smith, M. D. (2004). Interannual variability in TES atmospheric observations of Mars during 1999–2003. *Icarus*, *167*(1), 148–165. https://doi.org/10.1016/j.icarus.2003.09.010
Strausberg, M. J., Wang, H., Richardson, M. I., Ewald, S. P., & Toigo, A. D. (2005). Observations of the initiation and evolution of the 2001 Mars global dust storm. *Journal of Geophysical Research*, *110*(E2), E02006. https://doi.org/10.1029/2004JE002361
Thorndike, R. (1953). Who belongs in the family? *Psychometrika*, *18*(4), 267–276. https://doi.org/10.1007/BF02289263

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, *63*(2), 411–423. https://doi.org/10.1111/1467-9868.00293

Wagstaff, K., Lu, Y., Stanboli, A., Grimes, K., Gowda, T., & Padams, J. (2018). Deep Mars: CNN classification of Mars imagery for the Pds imaging atlas. In *Proceedings of the AAAI Conference on Artificial Intelligence*, (Vol. *32*, No. (1)). https://doi.org/10.1609/aaai.v32i1.11404

Wang, H., & Ingersoll, A. P. (2002). Martian clouds observed by Mars global surveyor Mars Orbiter Camera. *Journal of Geophysical Research*, *107*(E10), 8-1–8-16. https://doi.org/10.1029/2001JE001815

Wilson, A., & Chicarro, A. (2004). *Mars express - The scientific payloa*d (pp. 1–219). ESA Special Publication, SP-1240.

Wolff, M. J., Clancy, R. T., Kahre, M. A., Haberle, R. M., Forget, F., Cantor, B. A., & Malin, M. C. (2019). Mapping water ice clouds on Mars with MRO/MARCI. *Icarus*, *332*, 24–49. https://doi.org/10.1016/j.icarus.2019.05.041

Yu, L., Vanderburg, A., Huang, C., Shallue, C. J., Crossfield, I. J. M., Gaudi, B. S., et al. (2019). Identifying exoplanets with deep learning. III. automated triage and vetting of TESS candidates. *The Astronomical Journal*, *158*(1), 25. https://doi.org/10.3847/1538-3881/ab21d6

Zuber, M. T., Smith, D. E., Solomon, S. C., Muhleman, D. O., Head, J. W., Garvin, J. B., et al. (1992). The Mars observer laser altimeter investigation. *Journal of Geophysical Research*, *97*(E5), 7781–7797. https://doi.org/10.1029/92JE00341