Global analysis of gully composition using manual and automated exploration of CRISM imagery

Elyse Allender, Tomasz F. Stepinski*

Space Informatics Lab, University of Cincinnati, Cincinnati, OH 45221-0131, USA

Abstract

Gully formations on Mars have been the focus of many morphological and mineralogical studies aimed at inferring the mechanisms of their formation and evolution. In this paper we have analyzed 354 globally distributed gully-bearing Full Resolution Targeted (FRT) Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) images. The primary goal of the analysis was to identify all spectrally distinct deposits in these images (if any) and to classify them into hydrated and non-hydrated categories using only CRISM summary parameters (Viviano-Beck et al., 2014). Such approach makes possible to analyze a very large set of all distinct deposits in 354 images. We found that 68% of these images lack any distinct deposits, 8% of images contain non-hydrated deposits which coincide with the gullies and 24% of images contain hydrated deposits which coincide with the gullies. These results are compared with the recent analysis of 110 CRISM images by Nuñez et al. (2016) who also found that most gullies coincide with indistinct deposits, but, contrary to our findings, they found a predominance of non-hydrated minerals among distinct deposits. We attribute this discrepancy in part to their smaller and geographically biased sample of images, and in part to differing protocols of categorizing images. The discrepancy between the two surveys is further increased if we count all deposits in FRT gully-bearing images, not just deposits directly coinciding with the gullies, obtaining 44% indistinct, 15% non-hydrated, and 41% hydrated images. The secondary goal of this study was to perform the same image survey using a recently developed automated method in order to assess its accuracy and thus its feasibility for performing future surveys. We found the overall accuracy of the auto-mapper to be 76.2% but its accuracy for discovering distinct deposits, and in particular, distinct hydrated deposits was lower. We attributed the deficiencies of the auto-mapper primarily to its sensitivity to presence of noise in images and especially to presence of speckle noise. It is however worth noting that qualitatively both manual and automatic surveys arrived at the same overall conclusion.

Keywords: Mars, CRISM, Gullies formation, Mineral composition, Automatic analysis

1. Introduction

Gullies are among the youngest surface features on Mars; their topography is typically very sharply defined when not locally overlain by aeolian deposits or cut by faults (Treiman, 2003). In terms of structure, gullies consist of an alcove, channel, and depositional apron, while in terms of distribution, gullies are limited to latitudes greater than ± 30 degrees, and most frequently occur between 30 and 50 degrees in both hemispheres (Malin and Edgett, 2000).

The debate concerning the formation mechanism of gully features on Mars has been ongoing since the features were first observed from Mars Global Surveyor (MGS) Mars Orbiter Camera (MOC) in 1997 (Malin and Edgett, 2000). Their formation mechanism has been the focus of much speculation as the morphological characteristics of their terrestrial analogues imply that liquid water may have played a role in their formation. Fluidized mechanisms which may be responsible for gully emplacement include: subsurface aquifer release (Malin and Edgett, 2000; Gaidos, 2001), snowpack melting (Christensen, 2003; Williams et al., 2008) or near-surface ground ice (Mellon and Phillips, 2001; Costard et al., 2002) during periods of high-obliquity (Bridges and Lackner, 2006; Laskar et al., 2004). How-

*Corresponding author.

Email addresses: allendej@mail.uc.edu (Elyse Allender), stepintz@uc.edu (Tomasz F. Stepinski)

ever, as liquid water is not stable on Mars' surface given the current state of the atmosphere, several dry wasting mechanisms have also been proposed as the processes which emplaced the gullies. These are dry mass wasting processes (Treiman, 2003; Shinbrot et al., 2004; Pelletier et al., 2008), and CO_2 processes whether liquid (Musselwhite et al., 2001), or frost-based (Hoffman, 2002; Ishii and Sasaki, 2004; Mangold et al., 2008; Dundas et al., 2010, 2012; Raack et al., 2015; Pilorget and Forget, 2016). Kolb et al. (2010) suggests a dual mechanism, in which gullies were previously emplaced via a time-limited fluidized flow, and are currently modified through dry mass wasting, while Vincendon (2015) suggests several mechanisms (CO_2 and H_20) may be responsible for present-day gully activity.

Several different methodologies can be applied to investigate whether liquid water played a role in the formation of gullies, but for an investigation having a global scale and thus statistical meaning, perhaps the best approach is to utilize hyperspectral imagery taken by the Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) (Murchie et al., 2007) instrument onboard the Mars Reconnaissance Orbiter (MRO) to look for the presence or absence of hydrated minerals which coincide with gully locations. This was an approach taken in a recent study by Nuñez et al. (2016) who examined 110 globally distributed CRISM Map-projected Targeted Reduced Data Records (MTRDR) for associations between gully locations and their overlying mineralogy. They concluded that most gullies in their sample of images were spectrally indistinct from their surroundings, and, in cases when distinctions did exist, the mineralogy overlying the gullies varied without preference for hydrated minerals. On the basis of their findings they concluded that there was no clear evidence that long-term liquid water activity played a role in the formation and evolution of gullies.

It is worth noting that these Nuñez et al. (2016) conclusions were based on a relatively small sample of images. Globally, there are almost 5000 gullied landforms containing tens of thousands of individual gullies as documented in a database created by Harrison et al. (2015) on the basis of examining 54,023 CTX images. As gullies are features frequently targeted by CRISM, it is reasonable to expect that a larger sample of hyperspectral images containing gullies can be identified and examined to see whether the conclusions of Nuñez et al. (2016) will still hold.

This paper describes the result of such a larger survey. By overlaying the Harrison et al. (2015) gully database with image stamps of CRISM Full Resolution Targeted (FRT) images we identified 354 images containing gullies. For this set of images we performed an analysis aimed at identifying distinct mineral deposits co-located with gullies.

Conducting large-scale spectral surveys using a standard protocol based on matching spectra to specific mineral species is tedious work, which may be one of the reasons behind the relatively small size of the sample analyzed by Nuñez et al. (2016). We use a methodology based on summary parameters (Viviano-Beck et al., 2014) and browse products which may be less accurate but is more practical in application to large survey. Moreover, such methodology is underpinning recently developed (Allender and Stepinski, 2017) automated method for preliminary mapping of mineral deposits from CRISM images. Automating or at least semi-automating analysis of CRISM images makes possible even larger surveys. The secondary goal of this paper is to conduct the gully location/mineralogy analysis for a second time using our automated method. This makes it possible to assess the performance of the automated method on this relatively large sample of images for which, due to the primary goal of the paper, the ground truth has been established.

The rest of this paper is organized as follows. Section 2 explains the details of our dual methodology. Section 3 contains the results of manual survey. Section 4 is devoted to a comparison of the results of the manual and automatic surveys. Section 5 contains the discussion and conclusions. The supplement contains information about all images analyzed using both manual and automated methods.

2. Methodology

We use the Java Mission-Planning and Analysis for Remote Sensing (JMARS) software package (Christensen et al., 2009) to overlay stamps of CRISM FRT images with gully feature points from (Harrison et al., 2015) database. Note that Harrison et al. (2015) did not assign feature points to individual gullies but only to landmarks (predominantly craters) containing the gullies. A stamp is an image outline associated with an image ID, and a gully feature point is a pair of coordinates associated with a center of a landmark (crater). We only consider FRT images for inclusion in the sample to have a consistent image resolution. Using this procedure we will find all FRT images with gullies on rims of craters having radii equal to or smaller than the extent of an FRT image; we will not identify FRT images with gullies in larger craters as their stamps will not overlap with the centers of these craters. This procedure yielded 354 images, a sample over three times

the size of the Nuñez et al. (2016) sample. The 354 images were downloaded from the Planetary Data Systems (PDS) Geosciences Node and pre-processed using the standard algorithms provided in ENVI IDL 5.3 with the CRISM Analysis Tool (CAT) 7.3.1. The supplement to this paper consists of a table listing all of these images together with their parameters, catalogs of all mineral deposits (if any) found within their bounds using manual and automatic methods, and classification into one of five categories (see below). All further analysis (manual and automatic) is performed on these pre-processed images.

As our focus is on calculating the statistics of hydrated versus non-hydrated deposits underlying the gullies, we do not identify the specific mineralogy of each deposit but only classify all deposits as either indistinct, distinct hydrated, or distinct non-hydrated. A single FRT image may contain several gully features. We identify and classify all distinct deposits within an image but keep track of which deposits underlie gullies and which do not, yielding two additional classification categories. Our reason for keeping deposits that are not coincident with gullies is the small size of the area covered by an FRT image; the very existence of hydrated deposits in the vicinity of the gullies may be significant.

2.1. Manual methodology

Our manual analysis is aimed at the identification and classification of deposits and begins with the calculation of the suite of 26 second generation CRISM summary parameters (Viviano-Beck et al., 2014) for each of the 354 images. We then generate, for each image, the HYD, MAF, ICE, PHY and CAR browse products (each utilizing a triad of summary parameters (Viviano-Beck et al., 2014)) which focus on the detection of hydrated minerals, mafic minerals, ices, phyllosilicates, and carbonates respectively. Note that for each of the three summary parameters constituting a browse product a 0-99.9% image stretch was performed to ensure positive values of band depths and to omit outliers. These maximum stretch values are recorded in the Supplement to this paper so that the browse products may be reconstructed.

To analyze an image for mineral deposits all of its five browse products were visually examined. During the examination deposits, if present, were manually marked and labeled in accordance with the browse product keys in Viviano-Beck et al. (2014). For example, when examining the PHY browse product – whose constituent summary parameters are BD1900R2, BD2200, and BD2300 – a deposit with enhancements in the BD1900 and BD2300 summary parameters would appear in magenta color and be interpreted as an Fe/Mg phyllosilicate. All identified and labeled deposits are relabeled into two categories, hydrated or non-hydrated. The hydrated category's deposits are indicative of mineral alteration through contact with water. We consider deposits from the HYD, PHY, and CAR browse products hydrated. We also consider the H_2O ice deposit from the ICE product hydrated. Deposits from the MAF product are considered non-hydrated, as is the CO_2 ice deposit from the ICE product. An image which has no distinct deposits is labeled as 'uninteresting'.

It should be noted that in some images it was difficult to confirm the presence of spectrally distinct deposits as they appeared as linear features within shadowed gullies. If these linear features also appeared in shadowed regions across the rest of the image they were disregarded. If, however, the spectrally distinct deposits were present as linear features but in a spatial configuration that suggested underlying surface structure, then they were accepted as a viable detection.

Each map of labeled deposits was overlaid with an image reduced to a single band (1300 μm) which clearly shows surface features like gullies. By visually examining spatial relations between mineral deposits and gully features we classify images containing deposits into four categories, 'hydrated_gullies', 'non-hydrated_gullies', 'hydrated_non-gullies', 'non-hydrated_non-gullies'.

An image is classified as 'hydrated_gullies' if at least one of its hydrated deposits coincides with a gully feature. If none of the hydrated deposits coincide with gully features but at least one non-hydrated deposit does, the image is classified as 'non-hydrated_gullies'. If no deposits coincide with gullies but there is at least one hydrated deposit the image is classified as 'hydrated_non-gullies'. Finally, if no deposits coincide with gullies but there are non-hydrated deposits, the image is classified as 'non-hydrated_non-gullies'. Thus, at the end of our manual analysis each of the 354 images is classified into one of five categories including an 'uninteresting' category for images with no distinct deposits. Note that the label 'uninteresting' is used for consistency with the nomenclature of the automatic method (see next subsection), however, Nuñez et al. (2016) uses the label 'indistinct'.

2.2. Automated methodology

The most tedious part of manual analysis is the observation and marking of all possible deposits from the five CRISM browse products. Automating this task would make all surveys, particularly large surveys, of

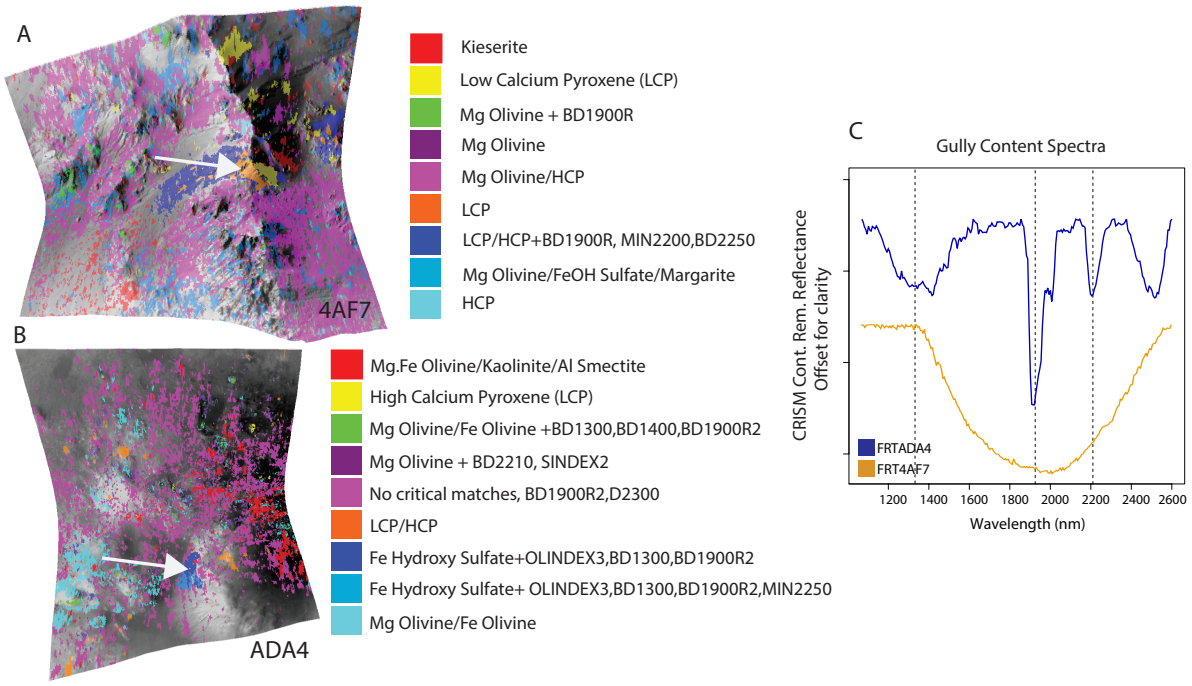

Figure 1: Examples of mineralogical maps generated by the auto-mapper from CRISM FRT images (A) 4AF7 (B) ADA4. Mineral signatures extracted from each of their respective main gully features are displayed in (C), with colors corresponding to their keys in (A) and (B). Arrows show the locations of extracted spectra in (A) and (B). Vertical dashed lines positioned at 1300 nm, 1900 nm, and 2200 nm illustrate hydration band depths for FRTADA4, while the FRT4AF7 LCP central band depth lies just above 1800 nm. The auto-mapper uses a LookUp Table (LUT) to assign mineral labels based on the presence of specific band depth features (process more fully described in Allender and Stepinski (2017). Specific mineral species may only be assigned based on the presence of certain band depths, as a basic example: in order for a deposit to be assigned as an Mg-smectite such as montmorillionite it must have absorption features at 1900, 2200, and 2300 nm (represented by CRISM summary products BD1900R_2, D2200, and D2300). If a deposit contains all of these features, but has additional enhancements in BD1400 and OLINDEX3 it would be labeled in the figure above as Mg-Smectite + BD1400 + OLINDEX3. If a deposit contained features from two mineral species who share common band depth features such as Kaolinite and Al-Smectite, the deposit would be labeled in the figure above as Kaolinite/Al-Smectite.

mineralogical deposits in CRISM images more attainable. Recently we introduced a method that, for a given CRISM FRT image, automatically marks and labels mineral units present in an image (Allender and Stepinski, 2017). The output of this method is a thematic map of mineral units. As auto-mapping is a new technology and has only been validated on a small sample of 20 images in the original Allender and Stepinski (2017) paper, the present study offers an opportunity to check its performance on a relatively large (354) sample of images.

Because of the way units are identified by the auto-mapper (by clustering vectors of values for 26 summary parameters) they are not necessarily mineralogically pure but may contain one, two, or more of 29 mineral species (see Table 1) that can be identified using the Viviano-Beck et al. (2014) parameters. Each unit's automatically generated label reflects species that are present in this unit. Thus, this method does not identify or map specific mineral species, instead it identifies and maps units having broader meanings. Fig. 1 shows two examples of the broad mineralogical maps automatically generated (with a legend) from FRT images by the auto-mapper. The legend refer to combinations of minerals and/or summary parameters showing enhancement from the background. This may be insufficient for identification of specific mineral species but sufficient for classification of deposits into 'hydrated' and 'non-hydrated'.

Fig. 1(C) shows spectral signatures extracted from main gully features in each image. The orange-colored spectrum is extracted from the region in the FRT4AF7 image auto-labeled as LCP. Its shape indeed indicates LCP, which is diagnosed by its broad absorption feature centered at around 1.8 microns. In this case both methods point to the same mineralogy. The dark blue-colored spectra is extracted from the region in FRTADA4 image auto-labeled as FeOH Hydroxy Sulfate + OLINDEX3, BD1300, BD1900R2, a description in-

dicating a mixture of minerals. This spectrum was interpreted by Carter et al. (2010) as Al-Smectite. In this case the auto-mapper did not identified a specific species but correctly indicated the presence of hydrated mineral which was the only information required.

We applied the auto-mapper to the batch of 354 images in our sample. This is a one-click operation resulting in the generation of 354 maps of mineral units. In the next step we classified all maps into the same five categories used in the manual analysis (see previous subsection). Table 1 divides all 29 mineral species, which are identifiable using Viviano-Beck et al. (2014) parameters, into hydrated (left column) and non-hydrated (right column) categories. If a unit has an auto-label containing species from the hydrated column in Table 1 we consider this unit 'hydrated', otherwise we consider it 'non-hydrated'. With individual units labeled, the images/maps are classified into 'hydrated_gullies', 'non-hydrated_gullies', 'hydrated_non-gullies', 'non-hydrated_non-gullies', and 'uninteresting' (images for which the auto-mapper did not find any distinct units) using the protocol described in the previous subsection. The decision of whether a deposit coincides with a gully is done manually.

Surprisingly, this procedure resulted in only 29 images being labeled as 'uninteresting' while 325 were given one of the four 'interesting' labels. A review of these 325 auto-maps corresponding to nominally 'interesting' images revealed that standard pre-processing (see the beginning of Section 2) did not eliminate all noise; 176 of the nominally 'interesting' images were really 'uninteresting' but, the residual noise was misinterpreted by the auto-mapper as a distinct unit. We manually reclassified these 176 images as 'uninteresting'. Note that this was a very fast fix as speckled images are very distinctive and can be reclassified without much effort. In our assessment of the performance of the auto-mapping (see Section 4) we didn't count these 176 cases as misclassifications because they are not related directly to the method itself. After this correction 205 images were assigned the 'uninteresting' label and 149 were assigned to one of the remaining four labels.

The outcome of both the manual and automated analysis is the classification of all images in our sample into five categories. This allows verification of the automated analysis in standard machine learning fashion. We consider the results of manual analysis as our "ground truth" and the results of automatic analysis as the predictions of a classifier. To assess the accuracy of its predictions we calculate a confusion matrix (Stehman, 1997). From this confusion matrix we calculate an overall classification accuracy and the values of recall and precision for each of the five categories.

Table 1: Mineral species allocated into hydrated/non-hydrated categories when using the automated method.

Hydrated	Non-Hydrated
H_2O Ice	CO_2 Ice
Kieserite	Low Calcium Pyroxene (LCP)
Alunite	High Calcium Pyroxene (HCP)
FeOHSulfate	Plagioclase
Jarosite	Fe Olivine
Polyhydrated Sulfate	Mg Olivine
Bassanite	
Kaolinite	
Margarite	
Gypsum	
Al Smectite	
Fe Smectite	
Mg Smectite	
Talc	
Serpentine	
Chlorite	
Prehnite	
Mg Carbonate	
Ca Carbonate	
Hydrated Silica	
Analcime	
Epidote	
Illite	

Take the category 'hydrated_gullies' as an example. The recall for this category is the ratio of the number of images correctly identified by the auto-mapper as containing gullies coincident with hydrated deposits to the number of all images containing such features in the manual analysis. Thus, recall gives us information about the auto-mapper's performance with respect to false negatives (how many hydrated deposits it missed). A low value of recall indicates that many images containing gullies overlaid by hydrated deposits are undetected by the auto-mapper. The precision for this category is the ratio of the number of images correctly identified by the auto-mapper as containing gullies overlaid by hydrated deposits to the number of all images identified (correctly or incorrectly) by the auto-mapper as such. Thus precision gives us information about the auto-mapper's performance with respect to false positives. A low value of precision indicates that among images identified by the auto-mapper as containing gullies overlaid by hydrated deposits, many do not truly contain these deposits.

3. Manual survey results

The manual survey of 354 images was conducted as described in Section 2.1. We labeled 155 of these images as 'uninteresting', meaning we could not find any distinct mineral deposits in them. In the remaining 199 images we identified distinct deposits. Table 2 shows how these 199 images breakdown into four categories according to our classification as described in Section 2.1. As can be seen from Table 2, 145 (73%) images with distinct deposits contain at least one hydrated deposit, and 84 (58%) of these have hydrated deposits co-located with gullies. Only 54 (27%) of all images contain non-hydrated deposits; these are split evenly between images in which deposits overlay gullies (28) and those in which they do not (26).

Table 2: Category breakdown of 'interesting' images as labeled by manual analysis

	Within gullies	Not within gullies	Total
Hydrated	84	61	145
Non-Hydrated	28	26	54
Total	112	87	199

Our result, summarized in Table 2, can be compared to the results of Nuñez et al. (2016). For this comparison we use 96 of the 110 images they evaluated, as the remaining 14 were not in FRT format. Nuñez et al. (2016) classified images into spectrally indistinct, phyllosilicates, mafics, and ices (CO_2/H_2O). We reclassified their labels into our categories as follows: spectrally indistinct $\longrightarrow$ uninteresting, mafics and CO_2 $\longrightarrow$ distinct non-hydrated within gullies, and phyllosilicates and H_2O $\longrightarrow$ distinct hydrated within gullies. With such a reclassification counting the Nuñez et al. (2016) labels (using the supplement to their paper) yields the following: uninteresting – 50 (52%), distinct non-hydrated – 33 (34%), and distinct hydrated – 13 (14%). This can be compared to our results, restricted to the 'within gullies' column, in Table 2: uninteresting (this includes 155 images which have no distinct deposits at all plus 87 images that have distinct deposits which do not overlay gullies) – 242 (68%), non-hydrated – 28 (8%), hydrated – 84 (24%).

Our results agree with the results of Nuñez et al. (2016) inasmuch as they confirm that in most images gullies do not coincide with distinct deposits, in fact we have found a larger percentage of images (68% versus 52%) in which gullies are spectrally indistinct from surroundings. However, we also have found that among the distinct deposits, hydrated deposits dominate non-hydrated deposits by a ratio of 84/28 = 3, whereas Nuñez et al. (2016) found that non-hydrated deposits dominate hydrated deposits by a ratio of 33/13 = 2.5. Thus, we have found that if deposits are distinct they are predominantly hydrated whereas Nuñez et al. (2016) found the opposite.

More insight follows from considering statistics for each hemisphere separately. Fig. 2A shows the global distribution of all of our 354 images, while Fig. 2B shows the global distribution of the 96 FRT images in Nuñez et al. (2016) for comparison). In this figure each image is indicated by one of five symbols according to its assigned category (see the legend). Note that some CRISM images in a sample are located very close to each other so their symbols on Fig. 2 may be obscured by a symbol indicating other close-by image. Table 3 enumerates images having different categories separately for northern and southern hemispheres. Our ratio of images in southern to northern hemispheres is 232/122 = 1.9 whereas the Nuñez et al. (2016) ratio is 76/20 = 3.8, thus our sample is much more balanced between the two hemispheres; it contains relatively more images in the north.

For the northern hemisphere the Nuñez et al. (2016) statistics are as follows: uninteresting – 15 (75%), non-hydrated – 1 (5%), hydrated – 4 (20%). This can be compared to our results restricted to hydrated_gullies column in Table 3, uninteresting – 78 (64%), non-hydrated – 3 (2%), hydrated – 41 (34%). For the northern hemisphere the difference between Nuñez et al. (2016) and our results is that we identified a larger percentage of distinct deposits (36% versus 25%) and a larger ratio of hydrated versus non-hydrated deposits (14 versus 4). For the southern hemisphere the Nuñez et al. (2016) statistics are as follows: uninteresting – 35 (46%), non-hydrated – 32 (42%), hydrated – 9 (12%). This can be compared to our results restricted to hydrated_gullies column in Table 3, uninteresting – 164 (71%), non-hydrated – 22 (10%), hydrated – 44 (19%). For the southern hemisphere the difference between Nuñez et al. (2016) and our results is that we identified a smaller percentage of distinct deposits (29% versus 54%) but a larger ratio of hydrated versus non-hydrated deposits (2 versus 0.37).

In addition to calculating statistics for images on the basis of deposits which coincide with gullies, as an alternative approach we also calculate statistics for all distinct deposits in gully-bearing images regardless of whether these deposits coincide with gullies. It is important to note that presence of hydrated deposits may not necessary be indicative of gully formation mechanism but rather reflect pre-existing hydrated minerals within the bedrock that have been remobilized by gully

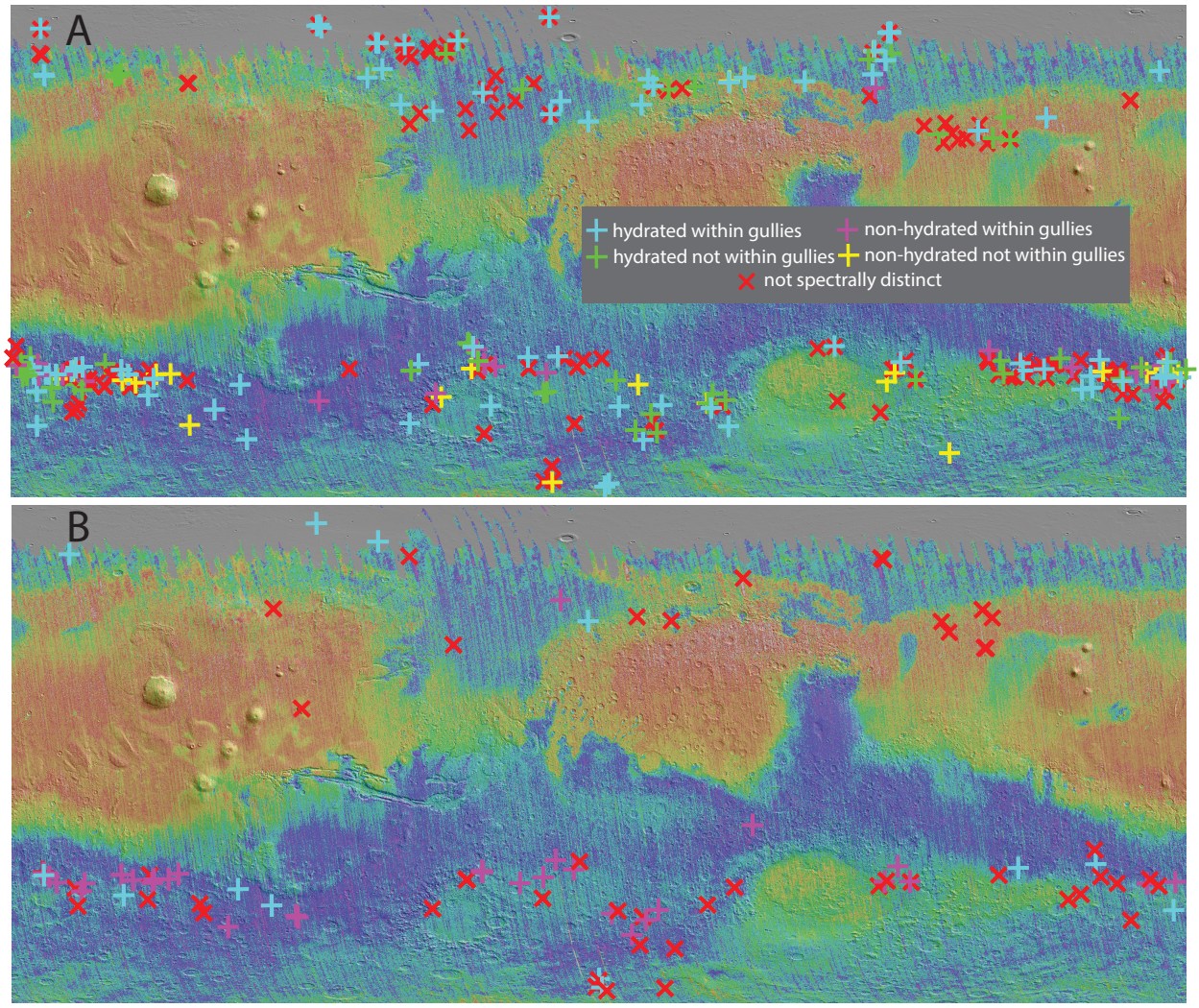

Figure 2: (A) Locations of the 354 FRT CRISM images in our sample. (B) Locations of 110 CRISM images in the Nuñez et al. (2016) sample. A map of the Dust Cover Index (DCI) (Ruff and Christensen, 2002), with the blue-to-red color gradient indicating the increasing presence of dust, is displayed in the background for reference. CRISM images are marked by symbols corresponding to our classification categories.

activity. Detection of hydrated deposits outside gullies would support such hypothesis. In addition, detection of nearby deposits is also important because modifications such as dust deposition can mask gully signatures. For this analysis both 'within gullies' and 'not within gullies' columns in Table 2 are counted, yielding the following: uninteresting – 155 (44%), non-hydrated – 54 (15%), and hydrated – 145 (41%). Such a counting protocol increases the percentage of images with distinct deposits (from 32% to 56%) and also increases the percentage of hydrated deposits from 24% to 41%).

For the northern hemisphere this protocol yields the following: uninteresting – 53 (43%), non-hydrated – 6 (5%), hydrated – 63 (52%) indicating the dominance of images with hydrated deposits. For the southern hemisphere such a protocol yields the following: uninteresting – 102 (44%), non-hydrated – 48 (21%), hydrated – 82 (35%). Again, as in the northern hemisphere this protocol not only increases the percentage of distinct deposits (as it is designed to do) but also increases the percentage of distinct hydrated deposits (which it is not specifically designed to do).

4. Validation of the automated survey

The automated survey of 354 images was conducted as described in Section 2.2. Table 4 is a confusion matrix between the results of the manual methodology

Table 3: North/South split of classified images

	non-hydrated_gullies	hydrated_gullies	non-hydrated_non-gullies	hydrated_non-gullies	uninteresting	Total
Northern Hemisphere	3	41	3	22	53	122
Southern Hemisphere	25	43	23	39	102	232
Total	28	84	26	61	155	354

Table 4: Confusion matrix between manual classification (ground truth) and automatic classification.

		Manual					
		non-hydrated_gullies	hydrated_gullies	non-hydrated_non-gullies	hydrated_non-gullies	uninteresting	N
Automatic	non-hydrated_gullies	17	5	0	1	0	23
	hydrated_gullies	1	47	0	2	0	50
	non-hydrated_notGullies	1	6	21	2	1	31
	hydrated_notGullies	1	8	0	33	3	45
	uninteresting	8	18	5	23	151	205
	N	28	84	26	61	155	354

Table 5: Values of recall and precision of the automated classifier for five different categories of images.

	non-hydrated_gullies	hydrated_gullies	non-hydrated_non-gullies	hydrated_non-gullies	uninteresting
Recall	61%	56%	81%	54%	98%
Precision	74%	94%	68%	73%	74%

(ground truth) and the automated methodology. The columns of Table 4 show how images assigned to a given category by the manual method were classified into various categories using the automated method. The last row in Table 4 (labeled N) lists the breakdown of images into five categories according to the manual method and corresponds to the values in Table 2. The rows of Table 4 show how images assigned to a given category using the automated method were classified into various categories using manual method. The last column in Table 4 (labeled N) lists the breakdown of images into five categories according to the automated method. Comparison of the last row with the last column of the table offers a quick comparison between the results of manual and automatic methods; essentially, classification via the automated method underestimates the number of images with hydrated deposits and overestimates the number of uninteresting (indistinct) images.

More specific information about the performance of the automated method can be obtained from Table 5 which lists the values of precision and recall for each image category. As mentioned in Section 2.2), recall provides information about the automated method performance with respect to false negatives. Low values of recall indicate that images classified manually as belonging to a given category were undetected as such with the automated method. According to the recall values in Table 5, the automated method fails to identify many images containing hydrated deposits (both within and outside of gullies). It has a better recall performance on images containing non-hydrated deposits, especially those outside the gullies, and has the best performance with respect to recall on images having no distinct deposits. Precision provides information about the performance of the automatic method with respect to false positives. Low values of precision indicate that among images identified using the automated method as belonging to a given category, many do not belong to this category according to the ground truth (manual classification). According to the precision values in Table 5 the automated method is very precise for images containing hydrated deposits within gullies and moderately precise for the remaining categories. It is least precise for images containing non-hydrated deposits outside of the gullies.

Overall, the auto-mapper tends to miss some distinct mineral deposits, especially distinct hydrated deposits. This results in too many images being labeled as indistinct. On the other hand, once the auto-mapper finds a hydrated deposit there is a high probability that it is a real deposit, this probability is somewhat lower for distinct non-hydrated minerals. The overall accuracy of the automated method (the number of correctly labeled images divided by the total number of images) is 75.9%. This value, if taken at its face value, is respectable, however, it hides the fact that different categories are assigned with different accuracies.

If we were to perform only the automated survey and rely on its results for drawing conclusions about the connection between locations of the gullies and their underlying mineralogy we would get the following statistics

Table 6: Major causes of mismatch between automatic and manual classification of images

Mismatch Type	non-hydrated_gullies	hydrated_gullies	non-hydrated_non-gullies	hydrated_non-gullies	uninteresting	Total
Sub-superpixel Scatter	4	0	1	2	13	20
Image Speckle	0	2	1	0	10	13
Not Rare Deposit	1	1	4	2	10	18
Noise More Interesting	1	0	4	8	21	34
Total	6	3	10	12	54	85

when counting only deposits underlying the gullies: uninteresting (indistinct) – 281 (79%), non-hydrated – 23 (7%), and hydrated – 50 (14%). If counting all deposits (using our alternative protocol) we would get: uninteresting – 205 (58%), non-hydrated – 54 (15%), and hydrated – 95 (27%). Although quantitatively these statistics differ from the manual survey, qualitatively they tell the same story – most gullies are indistinct from their surroundings, but, if distinct deposits are present they are predominantly hydrated.

To better understand the causes of the auto-mapper's misclassifications we examined all 85 (the sum of all non-diagonal entries in the confusion matrix in Table 4) images for which manually and automatically assigned labels differ. The results of this examination are summarized in Table 6. The rows in this table correspond to four causes of classification mismatch. *Sub-superpixel Scatter* is a mismatch caused when the distinct deposit is smaller than the size of a superpixel (see Allender and Stepinski (2017) for details) used by the auto-mapper; the auto-mapper can only recognize deposits larger than the user-defined size of a superpixel. Super-pixel scatter caused some images with distinct deposits to be auto-labeled as uninteresting. *Image Speckle* is a type of mismatch caused by mistaking a true deposit for speckle noise. As pointed out in Section 2.2 we reclassified 176 images with speckle noise to the uninteresting category, but as it turned out 13 of these images actually contained small distinct deposits in addition to noise. *Not Rare Deposit* is a type of mismatch caused when a gully is overlain by an interesting deposit but this deposit is also abundantly present in the rest of the image; thus, the superpixels underlying the gully were not outliers and were not considered by the auto-mapper to be distinct. For example, this happens when much of an image (including the gully features) contains H_2O ice; the ice will not stand out as a distinct deposit. *Noise More Interesting* is a type of mismatch caused by the auto-mapper limit on the number of outliers it considers as interesting deposits. If there are a lot of noisy superpixels in an image this limit is used up on them, leaving no room for outliers which are true deposits.

The columns in Table 6 correspond to categories of images misclassified by the auto-mapper. For example, the first row entry indicates that the auto-mapper misclassified 4 images as having non-hydrated deposits in gullies and determined 13 images were uninteresting (indistinct) due to sub-superpixel scatter. As can be seen from Table 6, regardless of the root cause of misclassification, in most cases the auto-mapper mistakenly classifies images as uninteresting, when in fact they do contain distinct deposits.

5. Discussion and conclusions

The primary purpose of this paper was to conduct a survey of 354 gully-bearing CRISM FRT images and identify all distinct mineral deposits, particularly those coincident with gully features. Statistics from such a survey can help to determine whether liquid water played a role in their formation and evolution and potentially refine the gully formation mechanism. The difference between our study and the recent study by Nuñez et al. (2016), which had a similar goal, is the size and content of image samples, and the methodology used to determine the character of deposits. The secondary purpose of this paper was to determine the degree to which we can rely on an automated method to find deposits in CRISM images (Allender and Stepinski, 2017) when conducting such a survey.

5.1. Gully content

The crucial issue addressed by this paper and by the Nuñez et al. (2016) paper is the same – what types of deposits, hydrated or non-hydrated, coincide with gullies? Both studies found that the majority of gullies are indistinct from their surrounding mineralogy. This is not a surprise as most gullies are expected to be mantled by dust just like their surrounding terrain, thus their underlying mineralogy is masked, and their hyperspectral images cannot be used to help determine their formation mechanism. However, both studies also identified gullies coincident with spectrally distinct deposits, which are potentially formation-indicating deposits. The main caveat then is that the formation mechanism governing gullies may have also simply exposed underlying, previously altered terrain, thus the presence of distinct mineralogy within a feature may simply be pre-existing

altered mineralogy (whether hydrated or non-hydrated) from within the bedrock which has been remobilized by gully activity. Indeed, we have found large number of hydrated deposits outside gullies which may support pre-existing hydrated mineralogy. The major and important difference between the two studies is that whereas Nuñez et al. (2016) found the majority of distinct deposits to be non-hydrated, we found the majority of distinct deposits to contain hydrated minerals. Two plausible explanations for this discrepancy are: (a) the different image samples and sizes, and (b) the different protocols on how to classify an image.

Our sample has over three times as many images as Nuñez et al. (2016) and they are more evenly distributed between the two hemispheres (see Fig. 2 and the description in Section 3). Visual comparison of Fig. 2A with Fig. 2B reveals that although in the southern hemisphere both sets of images sample approximately the same regions, this is not true of the northern hemisphere. In the northern hemisphere we sample more images from the low DCI regions of Acidalia Planitia and from higher latitudes. The fact that our sample contains a larger percentage of images in the northern hemisphere (34% versus 21% in the Nuñez et al. (2016) sample), and because many of our images in the northern hemisphere are in low dust areas and at higher latitudes not sampled by Nuñez et al. (2016) may explain why our overall statistics favor hydrated deposits. Additionally, small samples tend to be more biased than larger samples, and the Nuñez et al. (2016) sample is definitively biased inasmuch as the images in the northern hemisphere are under-represented, and, furthermore, the images in the low dust areas in the northern hemisphere are also under-represented.

We also acknowledge bias in our sample due to the selection process detailed in Section 2 – our sample only includes gully-bearing images in craters having a radius equal to or smaller than the size of an FRT image. It is possible that there could be a difference in the mineral signature of gullies depending on the size of craters they are located in; however, given that gullies are relatively young features mineral dependence on the size of the host crater is not expected. This is also supported by the fact that the 47 FRT images in the Nuñez et al. (2016) study which are also present in our sample (and thus are located in craters whose sizes were restricted by the size of FRT images) have a similar breakdown into hydrated, non-hydrated, and indistinct categories as the 49 FRT images that are not in our sample (and thus are presumably located in larger craters).

In the southern hemisphere both surveys sample approximately the same low DCI areas, yet we still found that among distinct deposits hydrated deposits are more numerous than non-hydrated deposits, whereas Nuñez et al. (2016) found the opposite. Because the sizes of the two samples in the south are quite different, 232 images in our sample versus the 76 in theirs, sample bias may again be a factor, but we also need to consider that differences in overall methodologies may contribute to differences in the results.

Our methodology of identifying distinct deposits and classifying them into categories (see Section 2.1) differs from what Nuñez et al. (2016) employed to obtain their results. In order to be able to process a large number of images (and even larger number of deposits within them) we observed and labeled deposits from browse products, whereas Nuñez et al. (2016) used browse products only as an intermediate step to highlight the presence/lack of deposits in an image. They then extracted spectra from these highlighted regions, averaged them, and assigned a label to the entire image by matching the averaged spectra to a library dataset. This invites a question about an accuracy of our method. Are summary parameters sufficient to classify deposits? In our previous paper (Allender and Stepinski, 2017) we use our method to conduct a min-survey of 20 FRT images with mineralogy of deposits previously determined via spectra matching. We found that our summary products-based method indicates mineralogies which are compatible with those indicated by spectra matching. Fig. 1C further illustrates feasibility of our method to identify deposits and classify them into hydrated and non-hydrated categories.

We note that images in our sample (which includes approximately half of the images analyzed by Nuñez et al. (2016)) contain multiple gullies and, on occasion, browse products indicate the presence of hydrated deposits in some of them while the others are flagged as underlain by non-hydrated deposits. Given this potential ambiguity we have developed a specific protocol (see Section 2.1) on how to label images. However, it is not clear from Nuñez et al. (2016) which gullies in the image contribute to the overall label determination or from which regions their contributing spectra have been extracted. Differences in these protocols may also be responsible for discrepancies between the results of the two surveys. In fact, out of the 47 images common to both surveys only 27 images were classified to the same category while 20 images were classified to differing categories. An example of one of these differing images (FRT9B59) is shown in Fig. 3. We classified this image as 'hydrated_gullies' and indeed spectrum shown in Fig. 3 indicates that H_2O/CO_2 ice mix is present. However, this image was marked as 'indis-

tinct' by Nuñez et al. (2016). This image may illustrate some of the frost-based current gully modification mechanisms at work on martian gullies as explored by Hoffman (2002); Ishii and Sasaki (2004); Mangold et al. (2008); Dundas et al. (2010, 2012); Raack et al. (2015); Pilorget and Forget (2016).

Overall, our survey of a large number of gully-bearing FRT images shows that the majority of gullies are indistinct from their surroundings. This is in agreement with the results of Nuñez et al. (2016). However, in images where we do observe distinct deposits we find that hydrated deposits are more numerous than non-hydrated deposits. This is in opposition to the results of Nuñez et al. (2016). We were not able to determine with certainty the root cause of this discrepancy due to the lack of detail of the labeling protocol of Nuñez et al. (2016), however, two factors are likely to be responsible: (a) the two surveys use different samples of images – with our sample being significantly larger and less spatially biased, and (b) the protocols used to label the images are different. Additionally, we also count all deposits in FRT images, regardless of whether they coincide with gullies. Using this second protocol the proportion of images labeled as containing hydrated deposits further increases. An example of a distinct, hydrated deposit not within a gully feature can be seen in Fig. 4, where a carbonate deposit has been identified just outside the rim of an impact crater. This may be an exposure of underlying altered terrain which was uplifted through impact and may offer insight into the composition of underlying units in this region where there was previous aqueous activity.

Though the conclusions of Nuñez et al. (2016) were reached using a small, geographically biased sample of images and their image labeling protocol – the key component in arriving at such a result – is not described in sufficient detail for us to make a direct comparison, overall our results agree with their conclusion that there is no clear indication for a role for liquid water in gully formation. From the example shown in Fig. 3 it can be seen that there is some evidence to support frost-based modification processes currently taking place. Vincendon (2015) reported that the majority of 38 active gully sites in the southern hemisphere contained evidence for the presence of CO_2 ice and frost. We found 24 alternative images in which ice detections were distinct within gullies, 50% of which were located in the southern hemisphere. Of these 12 images, only 3 contained evidence for CO_2 ice, suggesting that recent seasonal modification may have occurred at these sites.

The majority of images in our sample are characterized by the presence of distinct deposits both within and outside of gullies. This is most consistent with pre-existing altered mineralogy (whether hydrated or non-hydrated) within the bedrock which has been remobilized by gully activity. As the majority of gullies are found in Noachian and late Hesperian terrains (Tanaka et al., 2014) exposures of this underlying altered terrain are expected to contain aqueously altered materials.

We roughly estimate that there may be around 700 CRISM images containing gully features. We base this estimation on the fact that our sample (354 images) is restricted to gullies in craters having a radius equal or smaller than the size of an FRT image – approximately half of images in the Nuñez et al. (2016) sample (which has no restriction on the size of the host craters) also belong to our sample. Thus, the approximate number of FRT images containing gullies should be twice the number of images in our sample. It would be difficult to manually analyze 700 images, hence the importance of having a reliable algorithm for automated identification and labeling of deposits in CRISM images.

5.2. Reliability of the automated method

The secondary purpose of this study was to check the feasibility of using the recently proposed Allender and Stepinski (2017) semi-automated method for conducting a survey of CRISM images aimed at statistical description of mineral deposits co-located with gullies. The method, as it is described in its original paper, is in fact semi-automated because it generates labels which are general rather than specific, so a user can utilize the method to screen a large sample of images for those containing broadly defined deposits of interest and then perform detailed manual analysis (similar to that of Nuñez et al. (2016)) just on these deposits to extract their specific content. However, in the context of this paper, we were not interested in determining which specific minerals were present in a deposit, only in a broad indication of whether the deposit was spectrally distinct from its surroundings and whether it contained evidence for hydrated mineralogy. For such a purpose the method is fully automated.

The major rationale for performing an automated survey is the speed of analysis. Our automated survey of 354 images took ~80 hours (3.5 days) after which labeled maps of deposits were produced for all 354 images. This is in contrast to our manual analysis for which it took ~200 hours (2.5 weeks) to create and examine the five combinations of browse product for each image and note down if an interesting deposit was present (no interpretation of these results was performed in this stated time frame). These times are offered as a comparison of the time taken for the auto-mapper to

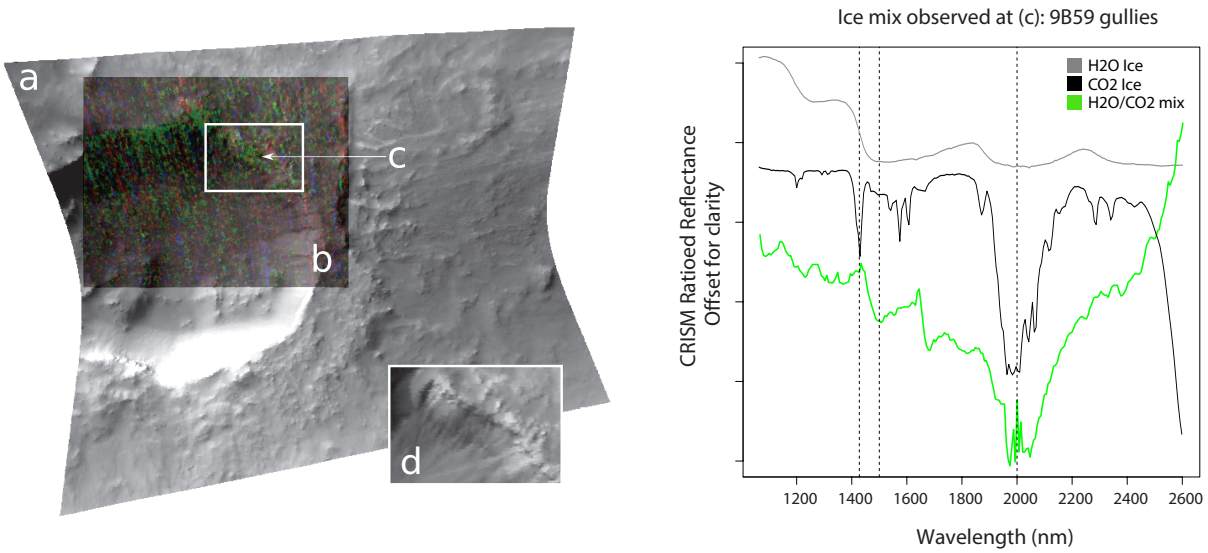

Figure 3: (Left) Single band greyscale image of FRT9B59 with locations of gullies overlain with the ICE browse product from Viviano-Beck et al. (2014): (a) Single band image approximating albedo (1300 nm) (b) Cropped portion of ICE browse product overlaying gullies (c) location of region of interest in a gully from which spectral signature is extracted. (d) Zoom image of greyscale underlying (c) illustrating presence of gullies contributing to ICE signature. (Right) Spectral signature extracted from region of interest (c) consistent with a mix of H_2O and CO_2 ices. Ice spectra from the MICA spectral library (Viviano-Beck, 2015) are plotted to show similarities in spectral shape and vertical dashed lines show the 1435 nm (CO_2), 1500 nm (H_2O) and 2000 nm (CO_2) features.

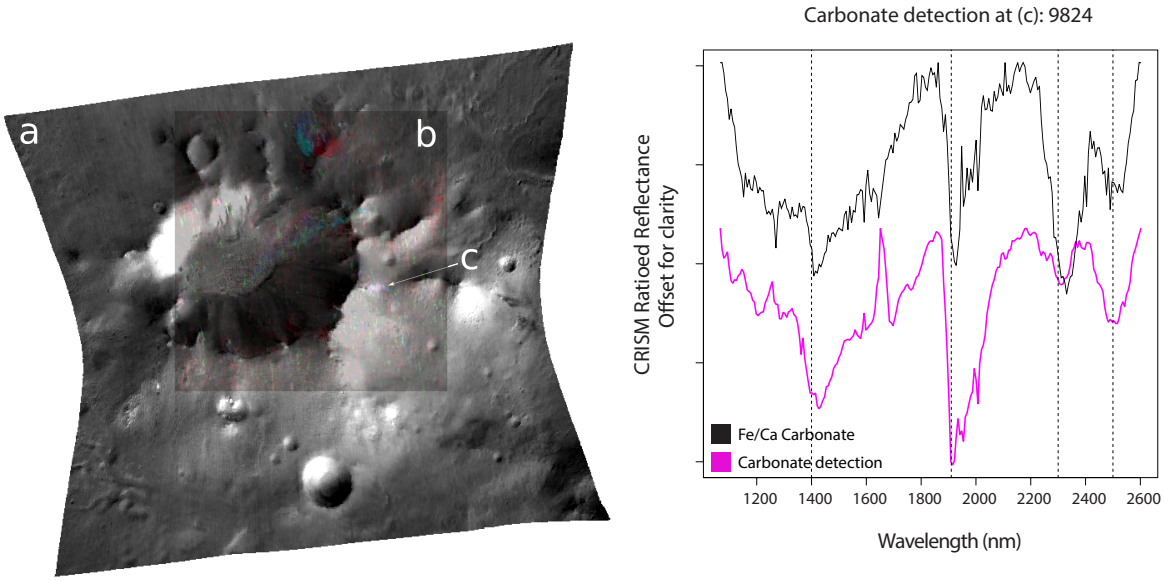

Figure 4: (Left) Single band greyscale image of FRT9824 with the region of interest overlain with the CAR browse product from Viviano-Beck et al. (2014): (a) Single band image approximating albedo (1300 nm) (b) Cropped portion of CAR browse product overlaying the location of region of interest (c) from which spectral signature is extracted. (Right) Spectral signature extracted from region of interest(c) consistent with a carbonate with absorption features at and around 1400, 1900, 2300, and 2500 nm indicated with dotted lines. The detected spectrum is plotted in magenta as this is a color the deposit appears in the CAR browse product. A MICA library spectrum (plotted in black) of Fe/Ca carbonate (Viviano-Beck, 2015) is included to show similarities in spectral shape.

run as opposed to the direct image processing effort required by an analyst. We show these numbers in order to demonstrate that automation may shorten the time required for analysis by an order of magnitude. Of course, the automated method is only valuable if it is accurate and the present study offered an opportunity to assess its accuracy.

As we documented in Section 4 the automated method is not yet as accurate as we would like it to be. Superficially, its overall accuracy of 75.9% on our sample of 354 images is on par with many commonly used classifiers (for example, terrestrial global land cover classifiers), however, the method is sensitive to the presence of noise, and to a lesser degree, to the values of selected parameters. The problem of noise is the primary reason for auto-mapper's misclassifications. As pointed out in Section 2.2 the presence of residual noise causes the auto-mapper to find spurious deposits. This type of misclassification is easily corrected, but the correction introduced its own error (see the "image speckle" row in Table 6). Noise is also responsible for an additional errors summarized in the "noise more interesting" row in Table 6. We project that the auto-mapper's performance, and its feasibility to perform surveys of large samples of CRISM images, will improve dramatically upon application to noise-reduced images. Modification of the auto-mapper to accommodate future surveys using the new MTRTR images, which have been subject to improved pre-processing and noise reduction, will likely yield superior results.

The secondary reason for auto-mapper's misclassifications is its (limited) sensitivity to free parameters. The misclassifications in the "sub-superpixel scatter" row in Table 6 are related to our selection of the size of superpixel. Finally, the design of the auto-mapper can occasionally cause a misclassification. The method assumes that a distinct deposit is a spectral outlier. However, there are situations where this is not the case. For example, if most of an image is covered by ice, and a small deposit of ice is located in an ice-free part of the image, it would not be identified as an outlier. Such cases are summarized in the "not rare deposit" row in Table 6.

Our analysis revealed that using an automated method for identifying and labeling deposits to survey CRISM images carries the risk of obtaining somewhat skewed results. If it is practical, a manual survey is still preferred as it is likely to yield more accurate results. The automated method would improve dramatically if we could remove more noise from images. Even with the current inaccuracies associated with the automated method, it should to be noted that the overall, qualitative conclusion from an automated survey of our sample of 354 images is the same as from our manual survey – most gullies are indistinct from their surroundings, yet if distinct deposits are found, most of them are hydrated. However, based on these results we cannot definitively conclude that a hydrated mechanism was involved in the emplacement and modification of gully features as gullied deposits contain pre-existing altered mineralogy (whether hydrated or non-hydrated) from within surrounding bedrock which has been remobilized by gully activity.

Acknowledgments. This work was supported by the University of Cincinnati Space Exploration Institute.

Allender, E., Stepinski, T., 2017. Automated, exploratory mineralogical mapping of CRISM imagery using summary product signatures. Icarus 281, 151–161.

Bridges, N. T., Lackner, C. N., 2006. Northern hemisphere Martian gullies and mantled terrain: Implications for near-surface water migration in Mars' recent past. Journal of Geophysical Research: Planets 111 (E9), e09014.

Carter, J., Poulet, F., Bibring, J., Murchie, S., 2010. Detection of hydrated silicates in crustal outcrops in the northern plains of Mars. Science 328, 1682–1686.

Christensen, P., 2003. Formation of recent martian gullies through melting of extensive water-rich snow deposits. Nature 422.

Christensen, P. R., Engle, E., Anwar, S., Dickenshied, S., Noss, D., Gorelick, N., Weiss-Malik, M., Dec. 2009. JMARS - A Planetary GIS. AGU Fall Meeting Abstracts.

Costard, F., Forget, F., Mangold, N., Peulvast, J. P., 2002. Formation of Recent Martian Debris Flows by Melting of Near-Surface Ground Ice at High Obliquity. Science 295 (5552), 110–113.

Dundas, C. M., Diniega, S., Hansen, C., Byrne, S., McEwen, A., 2012. Seasonal activity and morphological changes in martian gullies. Icarus 220 (1), 124–143.

Dundas, C. M., McEwen, A. S., Diniega, S., Byrne, S., Martinez-Alonso, S., 2010. New and recent gully activity on Mars as seen by HiRISE. Geophysical Research Letters 37 (7), l07202.

Gaidos, E. J., 2001. Cryovolcanism and the Recent Flow of Liquid Water on Mars. Icarus 153 (1), 218 – 223.

Harrison, T. N., Osinski, G. R., Tornabene, L. L., Jones, E., 2015. Global documentation of gullies with the Mars Reconnaissance Orbiter Context Camera and implications for their formation. Icarus 252, 236 – 254.

Hoffman, N., 2002. Active polar gullies on Mars and the role of carbon dioxide. Astrobiology 2 (3), 313–323.

Ishii, T., Sasaki, S., 2004. Formation of recent martian gullies by avalanches of CO2 frost. In: Lunar and Planetary Science Conference. No. 1556.

Kolb, K., McEwen, A., Pelletier, J., 2010. Investigating gully flow emplacement mechanisms using apex slopes. Icarus 208 (1), 132 – 142.

Laskar, J., Correia, A., Gastineau, M., Joutel, F., Levrard, B., Robutel, P., 2004. Long term evolution and chaotic diffusion of the insolation quantities of Mars. Icarus 170 (2), 343 – 364.

Malin, M. C., Edgett, K. S., 2000. Evidence for Recent Groundwater Seepage and Surface Runoff on Mars. Science 288 (5475), 2330–2335.

Mangold, N., Baratoux, D., Costard, F., Forget, F., 2008. Current gullies activity: Dry avalanches observed over seasonal forst as seen on HIRISE images. In: LPI workshop on martian gullies: theories and tests. No. 8005.

Mellon, M. T., Phillips, R. J., 2001. Recent gullies on Mars and the

source of liquid water. Journal of Geophysical Research: Planets 106 (E10), 23165–23179.

Murchie, S., Arvidson, R., Bedini, P., Beisser, K., Bibring, J.-P., Bishop, J., Boldt, J., Cavender, P., Choo, T., Clancy, R. T., Darlington, E. H., Des Marais, D., Espiritu, R., Fort, D., Green, R., Guinness, E., Hayes, J., Hash, C., Heffernan, K., Hemmler, J., Heyler, G., Humm, D., Hutcheson, J., Izenberg, N., Lee, R., Lees, J., Lohr, D., Malaret, E., Martin, T., McGovern, J. A., McGuire, P., Morris, R., Mustard, J., Pelkey, S., Rhodes, E., Robinson, M., Roush, T., Schaefer, E., Seagrave, G., Seelos, F., Silverglate, P., Slavney, S., Smith, M., Shyong, W.-J., Strohbehn, K., Taylor, H., Thompson, P., Tossman, B., Wirzburger, M., Wolff, M., 2007. Compact Reconnaissance Imaging Spectrometer for Mars (CRISM) on Mars Reconnaissance Orbiter (MRO). Journal of Geophysical Research: Planets 112 (E5), e05S03.

Musselwhite, D. S., Swindle, T. D., Lunine, J. I., 2001. Liquid CO2 breakout and the formation of recent small gullies on Mars. Geophysical Research Letters 28 (7), 1283–1285.

Nuñez, J. I., Barnouin, O. S., Murchie, S. L., Seelos, F. P., McGovern, J. A., Seelos, K. D., Buczkowski, D. L., 2016. New insights into gully formation on Mars: Constraints from composition as seen by MRO/CRISM. Geophysical Research Letters 43(17), 8893–8902.

Pelletier, J. D., Kolb, K. J., McEwen, A. S., Kirk, R. L., 2008. Recent bright gully deposits on Mars: Wet or dry flow? Geology 36 (3), 211–214.

Pilorget, C., Forget, F., 2016. Formation of gullies on Mars by debris flows triggered by CO2 sublimation. Nature Geoscience 9, 65–69.

Raack, J., Reiss, D., Appéré, T., Vincendon, M., Ruesch, O., Hiesinger, H., 2015. Present-day seasonal gully activity in a south polar pit (Sisyphi Cavi) on Mars. Icarus 251, 226–246.

Ruff, S. W., Christensen, P. R., 2002. Bright and dark regions on Mars: Particle size and mineralogical characteristics based on Thermal Emission Spectrometer data. Journal of Geophysical Research: Planets 107(E12).

Shinbrot, T., Duong, N.-H., Kwan, L., Alvarez, M. M., 2004. Dry granular flows can generate surface features resembling those seen in martian gullies. Proceedings of the National Academy of Sciences of the United States of America 101 (23), 8542–8546.

Stehman, S. V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62(1), 77–89.

Tanaka, K. L., Skinner, J. A., Dohm, J. M., Irwin III, R. P., Kolb, E. J., Fortezzo, C. M., Platz, T., Michael, G. G., Hare, T. M., 2014. Geologic map of Mars: USGS Scientific Investigations Map 3292. Tech. rep., U.S. Geological Survey, Flagstaff, AZ.

Treiman, A. H., 2003. Geologic settings of martian gullies: Implications for their origins. Journal of Geophysical Research: Planets 108 (E4), 8031.

Vincendon, M., 2015. Identification of Mars gully activity types associated with ice composition. Journal of Geophysical Research: Planets 120 (11), 1859–1879, 2015JE004909.

Viviano-Beck, C., 2015. MRO CRISM Type Spectra Library. NASA Planetary Data System.
URL `http://crismtypespectra.rsl.wustl.edu`

Viviano-Beck, C. E., Seelos, F. P., Murchie, S. L., Kahn, E. G., Seelos, K. D., Taylor, H. W., Taylor, K., Ehlmann, B. L., Wisemann, S. M., Mustard, J. F., Morgan, M. F., 2014. Revised CRISM spectral parameters and summary products based on the currently detected mineral diversity on Mars. Journal of Geophysical Research: Planets 119 (6), 1403–1431.

Williams, K., Toon, O., Heldmann, J., McKay, C., Mellon, M., 2008. Stability of mid-latitude snowpacks on Mars. Icarus 196 (2), 565 – 577.