SELECTION OF MULTISPECTRAL BAND COMBINATION FOR IDENTIFICATION OF EVAPORITE MINERALS ON MARS.

G. R. L. Kodikara and L. J. McHenry, Dept. of Geosciences, University of Wisconsin-Milwaukee, Milwaukee, WI 53211, gayantha@uwm.edu, lmchenry@uwm.edu.

Introduction: Evaporitic minerals on Mars have been identified by orbital instruments such as CRISM [1] and OMEGA [2], from Mars meteorites such as ALH84001 [3] and nakhlites [4], and by the in-situ rover missions such as Opportunity [5] and Curiosity [6]. The identification and characterization of evaporative environments are important to reconstruct the climatic history and to characterize the ancient aqueous fluids once present at the Martian surface and subsurface [5]. The CRISM summary products (simple band ratios and index maps) are commonly used to identify a diverse range of aqueous minerals, including sulfates, carbonates, phyllosilicates, and hydrated silica [7]. However, these methods can be time consuming and require attention to detail. Therefore, in this study, we demonstrate the capability of statistical learning algorithms (also called Machine Learning (ML) algorithms) to better identify the different evaporitic spectra from common surface mineralogies on Mars using a minimum number of bands/band combinations.

Methods: We use 150 reflectance spectra combining the ten most likely evaporite mineral species with five abundant surface minerals on Mars to help select the most suitable bands and/or band combinations and ML algorithm. The evaporite mineral group includes anhydrite, aragonite, calcite, dolomite, epsomite, gypsum, halite, magnesite, thenardite, and trona. The non-evaporite mineral group (common surface minerals on Mars) includes clinopyroxene, hematite, olivine, orthopyroxene, and plagioclase. Each mineral type was represented by ten spectra covering different particle sizes. Spectra were taken from RELAB, USGS, JPL and PSF spectral databases. Non-evaporite mineral group spectra were assigned as one common type called "surface" (50 spectra), while evaporite minerals were treated separately (100 spectra, with 10 spectra for each type). The band depths of all spectra were calculated using the "Convex hull" continuum removal methods and then the maximum band depth and their variations were calculated (Figure. 1). Based on the band depths and their variations, the wavelength range between 1000 nm - 2500 nm was selected for further analysis.

All spectra were resampled using Gaussian spectral response functions defined by the fwhm (full-width-half maximum) values of the CRISM Multispectral Survey (MSP) mode IR (infrared) data [2]. We created a Spectral Resampling Bandpass Filter, removing some close bands in MSP mode (Figure 1). Eighteen band indices were initially calculated. These include the existing CRISM spectral summary product parameters [7] and new band indices based on the general shape of the spectra and characteristics of the absorption features.

Table 1: Calculated spectral parameters (band indices).

Index	Formulation
BI01	$(R_{1079} – R_{2390})/(2390.58 – 1079.96)$
BI02	$R_{1875} – R_{1928}$
BI03	$R_{1690} – R_{1750}$
BI04	$R_{1079} – R_{2250}$
BI05	$R_{1079} – R_{2231}$
BI06	$R_{1079} – R_{2291}$
BI07	$R_{1079} – R_{2317}$
BI08	$R_{1079} – R_{2331}$
BI09	$R_{1079} – R_{2350}$
BI10	$R_{1079} – R_{2390}$

The Learning Vector Quantization (LVQ) model is used to estimate the feature importance and the best feature combination to identify different evaporite minerals from common "surface" materials [8]. This algorithm learns to find features automatically by viewing all the spectra and deciding which parts of the spectra and which indices are most useful for differentiating between those minerals. We used stratified random sampling with proportional allocations to split the entire dataset into two sets as training (to train the models) and validation (to evaluate their performance). Training sets include 80% of the observations (120 observations) and the rest are assigned as validation (30 observations) and kept aside to measure the accuracy of the winning ML algorithm. We adopted three ML algorithms, including Artificial Neural Network (nnet), Support Vector Machine (svm), and Random Forest (rf). The k-fold cross validation method was used to estimate the test error associated with each ML algorithm to evaluate their performance on the training dataset [9]. For that, the training dataset was split into ten parts, nine to train and one to test, and program runs were conducted for all combinations of train-test splits. The same process was repeated three times for each algorithm with different feature combinations to achieve the most reliable results with minimum band combinations. Finally, model performance was measured using the validation data set with two statistical measures, Overall accuracy and Kappa [10]. The entire study was done using the R statistical software package.

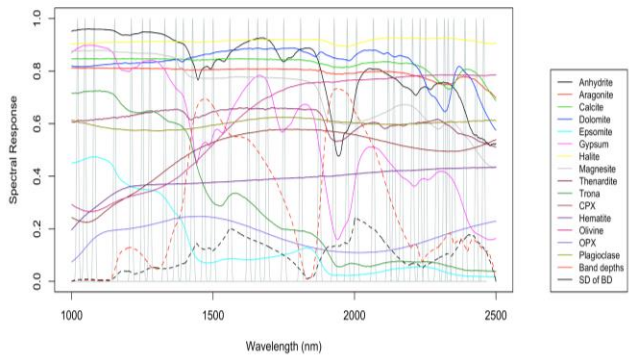

Figure 1. Representative spectra of each mineral type, Maximum band depth (red dashed line) and the variations of the band depths (black dashed line). CRISM MSP IR band responses are shown in gray.

Results: We achieved the highest accuracy with a minimum of ten features (indices) calculated over 12 bands (Table 1). Artificial Neural Network algorithm was the best ML algorithm for identifying the evaporites using the ten selected spectral parameters (Figure 2). Accuracy tells us the percentage of observations that the model classified correctly, while the kappa statistics tell us how well two evaluators can classify an observation correctly. It was able to differentiate the "surface" spectra from the evaporite spectra with 100% accuracy (Table 2). Based on the results, within the evaporite minerals, there is a 50% chance to misclassify calcite as dolomite, gypsum as anhydrite, halite as thenardite, and thenardite as "surface" materials, etc. Most of the above misclassifications are acceptable based on their spectral characteristics. The spectral response of both halite and thenardite are the same and only represent the O-H and H-O-H absorption features at 1.4 µm and 1.9 µm [11]. Calcite and dolomite also show similar spectral characteristics and the increase of Fe^{2+} content in dolomite causes the carbonate band in their spectra to shift to longer wavelengths [12]. Therefore, the band position differences in calcite and dolomite could be too small to detect using current band passes. Even though anhydrite should be anhydrous and thus featureless near 1.4 and 1.9 µm, most of our anhydrite spectra shows weak features indicating that they are partially hydrated and possibly in transition to gypsum or bassanite [13].

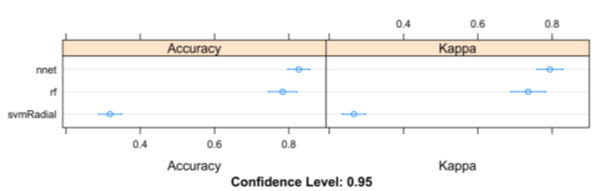

Figure 2: Accuracy and interrater reliability of adopted Machine learning methods. The Neural Network method yielded an accuracy around 80%.

Table 2. Confusion matrix of the test results.

	Reference										
Prediction	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*
A	1	0	0	0	0	1	0	0	0	0	0
B	0	2	0	0	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	0	0	0	0
D	0	0	1	2	0	0	0	0	0	0	0
E	0	0	0	0	2	0	0	0	0	0	0
F	0	0	0	0	0	1	0	0	0	0	0
G	1	0	0	0	0	0	1	0	1	0	0
H	0	0	0	0	0	0	0	2	0	0	0
I	0	0	0	0	0	0	1	0	1	0	0
J	0	0	0	0	0	0	0	0	0	2	0
K	0	0	0	0	0	0	0	0	0	0	10

(**A**: anhydrite, **B**: aragonite, **C**: calcite, **D**: dolomite, **E**: epsomite, **F**: gypsum, **G**: halite, **H**: magnesite, **I**: thenardite, **J**: Trona, **K**: surface)

Future Work: Since we used the laboratory spectra to train the model, the trained model cannot be used to identify the evaporites directly from the CRISM MSP data because the image spectra are different from the laboratory spectra mainly due to the additive black noise, multiplicative noise, dust cover on the surface and atmospheric effects [14]. Therefore, we are continuing our study to find out/ model the image spectra for the above mineral types to identify the evaporites on Mars using the proposed method. We will also incorporate the effect of the dust cover to the spectral behaviors.

References: [1] Murchie, et al. (2009) Journal of Geophysical Research, 114 (E00D06). [2] Carter, et al. (2013) Journal of Geophysical Research: Planets, 118, 831-858. [3] Warren (1998) Journal of Geophysical Research 103 (E7): 16, 759–16, 773. [4] Bridges and Grady (2000) Earth and Planetary Science Letters 176: 267–79. [5] Tosca, et al. (2005) Earth and Planetary Science Letters 240: 122–48. [6] McLennan, et al. (2014) Science 343 (1244734). [7] Viviano-Beck, et al. (2014) Journal of Geophysical Research: Planets, 119, 1403-1431. [8] Kohonen (2001) Self-Organizing Maps, 501 pp. [9] Kuhn and Johnson (2016) Applied Predictive Modeling, 600 pp. [10] James, et al. (2017) An Introduction to Statistical Learning with Applications in R, 426 pp. [11] Crowley (1991) Journal of Geophysical Research, 96 (B10), 16,231-16,240. [12] Gaffey (1986) American Mineralogist, 71, 151-162. [13] Bishop, et al. (2014) American Mineralogist, 99, 2105-2115. [14] Bultel, et al. (2015) IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. doi: 10.1109/JSTARS.2015.2405095.