Supporting Information for "Machine-Learning-Driven New Geologic Discoveries at Mars Rover Landing Sites: Jezero and NE Syrtis"

Murat Dundar [1], Bethany L. Ehlmann [2,3], Ellen K. Leask [2]

[1]Computer & Information Science Dept., Indiana University-Purdue University, Indianapolis, IN, USA,
[2]Div. of Geological & Planetary Sciences, California Institute of Technology, Pasadena, CA, USA,
[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA

Contents

Two-layer Bayesian Gaussian Mixture Model and Derivation of the posterior predictive distribution

We use the following generative model to fit spectral data available in our training library.

$$\text{Data model: } \boldsymbol{x_{ijk}} \sim N(\boldsymbol{\mu_{jk}}, \Sigma_k) \tag{1}$$

$$\text{Local prior: } \boldsymbol{\mu}_{jk} \sim N(\boldsymbol{\mu}_k, \Sigma_k \kappa_1^{-1}) \tag{2}$$

$$\text{Global prior: } \boldsymbol{\mu_k} \sim N(\boldsymbol{\mu_0}, \Sigma_j \kappa_0^{-1}), \Sigma_k \sim W^{-1}(\Sigma_0, m) \tag{3}$$

where k, j, and i are indices used to indicate true patterns, their observed instances, and individual pixels, respectively. $W^{-1}(\Sigma_0, m)$ denotes the inverse Wishart distribution with scale matrix Σ_0 and degrees of freedom m. This model assumes that pixels $\boldsymbol{x_{ijk}}$ are distributed according to a Gaussian distribution with mean $\boldsymbol{\mu_{jk}}$ and covariance matrix Σ_k. Each true pattern is characterized by the parameters $\boldsymbol{\mu_k}$ and Σ_k. The parameter $\boldsymbol{\mu_0}$ is the mean of the Gaussian prior defined over the mean vectors of true patterns, κ_0 is a scaling constant that adjusts the dispersion of the centers of true patterns around $\boldsymbol{\mu_0}$.

Corresponding author: M. Dundar, mdundar@iupui.edu

A smaller value for κ_0 suggests that pattern means are expected to be farther apart from each other whereas a larger value suggests they are expected to be closer. On the other hand, Σ_0 and m dictate the expected shape of the pattern covariance matrices, as under the inverse Wishart distribution assumption the expected covariance is $E(\Sigma|\Sigma_0, m) = \frac{\Sigma_0}{m-d-1}$, where d denotes the number of channels used. The minimum feasible value of m is equal to d+2, and the larger the m is the less individual covariance matrices will deviate from the expected shape. The κ_1 is a scaling constant that adjusts the dispersion of the means of observed pattern instances around the centers of their corresponding true patterns. A larger κ_1 leads to smaller variations in instance means with respect to the means of their corresponding true pattern, suggesting small variations among observed instances of the pattern. On the other hand, a smaller κ_1 dictates larger variations among instances. In Bayesian statistics the likelihood of a pixel $\boldsymbol{x}$ originating from pattern k is obtained by evaluating the posterior predictive distribution (PPD) for pattern k. For our two-layer Gaussian mixture architecture PPDs are derived in the form of *student-t* distributions by integrating out unknown mean vectors and covariance matrices of the true pattern distributions and their observed instances. This directly links observed pattern data with the hyperparameters of the model (κ_0,κ_1,m,μ_0, Σ_0). Optimizing hyperparameters with pixel data from the training library encodes information about observed pattern variations into the model.

Let $\boldsymbol{x}$ be the spectral representation of a pixel in an image to be classified. To classify $\boldsymbol{x}$ we need to evaluate $P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k}, \ldots, \bar{\boldsymbol{x}}_{n_k k}, S_{1k}, \ldots, S_{n_k k})$ for each true pattern, where $\bar{\boldsymbol{x}}_{jk}$ and S_{jk} are the sample mean vector and sample covariance matrix of the observed instance j of pattern k. The derivation of $P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k}, \ldots, \bar{\boldsymbol{x}}_{n_k k}, S_{1k}, \ldots, S_{n_k k})$ can be carried out in four steps.

In step 1 we integrate out the observed pattern mean vector $\boldsymbol{\mu}_{jk}$ and connect sample mean with the unknown mean vector $\boldsymbol{\mu}_k$ of the true pattern.

$$P(\bar{\boldsymbol{x}}_{jk}|\boldsymbol{\mu}_k, \Sigma_k) \quad = \quad N(\boldsymbol{\mu}_k, \Sigma_k(\frac{1}{n_{jk}} + \frac{1}{\kappa_1})) \tag{4}$$

where n_{jk} is the number of pixels available for observed instance j of true pattern k in the training library.

In step 2 we derive the posterior distribution of the mean vector $\boldsymbol{\mu}_k$ by Bayes rule and show that the posterior mean is weighted average of the sample mean vectors of observed instances and the prior mean.

$$
\begin{aligned}
P(\boldsymbol{\mu}_k|\bar{\boldsymbol{x}}_{1k},\ldots,\bar{\boldsymbol{x}}_{n_k k},\boldsymbol{\mu}_0,\Sigma_k,\kappa_0) &= N(\bar{\boldsymbol{\mu}}_k,\bar{\Sigma}_k) \\
\bar{\boldsymbol{\mu}}_k &= \frac{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}\bar{\boldsymbol{x}}_{jk}+\kappa_0\boldsymbol{\mu}_0}{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0} \\
\bar{\Sigma}_k &= \bar{\kappa}_k^{-1}\Sigma_k \\
\bar{\kappa}_k &= (\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0)
\end{aligned}
$$

where n_k is the number of observed instances of pattern k, i.e., the number of training images in which pattern k is detected.

In step 3 we derive the posterior distribution for Σ_k by combining Wishart terms corresponding to all observed instances of pattern k.

$$
\begin{aligned}
P(\Sigma_k|S_{1k},\ldots,S_{n_k k}) &= IW(\bar{S}_s,\bar{m}_s) &(5)\\
\bar{S}_s &= \Sigma_0+\sum_{j=1}^{n_k}S_{jk} &(6)\\
\bar{m}_s &= m+\sum_{j=1}^{n_k}(n_{jk}-1)) &(7)
\end{aligned}
$$

Finally, in step 4 we derive the posterior predictive distribution for pattern k by integrating out parameters μ_k and Σ_k. Thanks to the conjugacy in our model this operation produces a closed form solution in the form of a *Student-t* distribution.

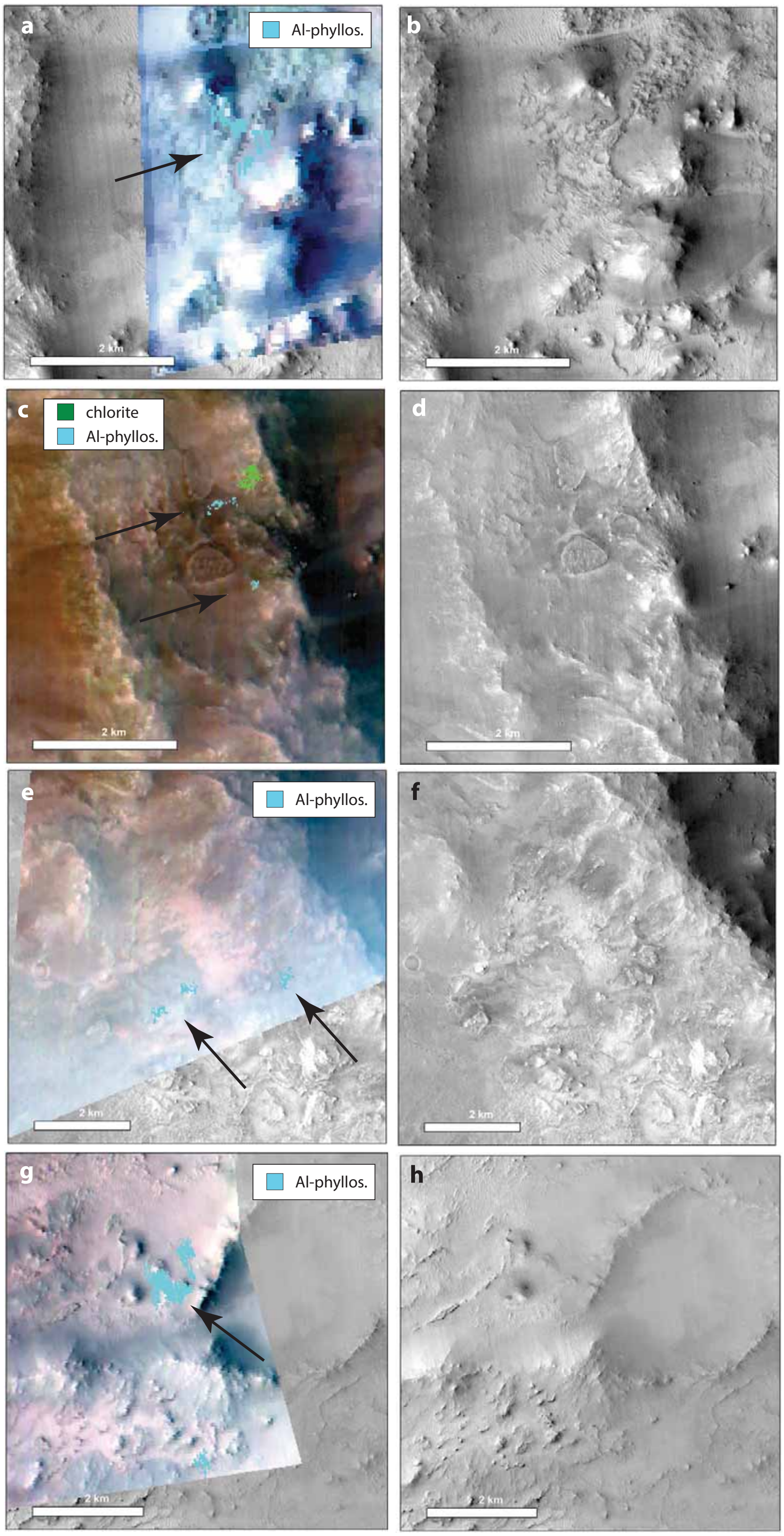
$$
\begin{aligned}
P(\boldsymbol{x}|\bar{\boldsymbol{x}}_{1k},\ldots,\bar{\boldsymbol{x}}_{n_k k},S_{1k},\ldots,S_{n_k k}) &= T(\boldsymbol{x}_{ji}|\bar{\boldsymbol{\mu}}_k,\bar{\Sigma}_s,\bar{v}_s) &(8)\\
\bar{\Sigma}_s &= \frac{\bar{S}_s}{\frac{\bar{\kappa}_s\bar{v}_s}{\bar{\kappa}_s+1}} \\
\bar{\kappa}_s &= \frac{(\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0)\kappa_1}{\sum_{j=1}^{n_k}\frac{n_{jk}\kappa_1}{(n_{jk}+\kappa_1)}+\kappa_0+\kappa_1} \\
\bar{v}_s &= m+\sum_{j=1}^{n_k}(n_{jk}-1)-d+1
\end{aligned}
$$

Context images of Jezero crater detections

Detections in Jezero crater and the corresponding CTX images are shown in Figures 1 and 2.

Dataset containing image coordinates of automated detections and ratioed spectra

Image coordinates of automated detections reported in this study are provided in the attached file.



64 **Figure 1.** (a) CRISM Al phyllosilicate detections in 040FF with (b) accompanying area in CTX. (c) CRISM Al phyllosilicate and chlorite detections in 05850 with (d) accompanying area in CTX. (e) CRISM Al phyllosilicate detections in 05850 with (f) accompanying area in CTX. (g) CRISM Al phyllosilicate detections in 1C558 with (h) accompanying area in CTX.

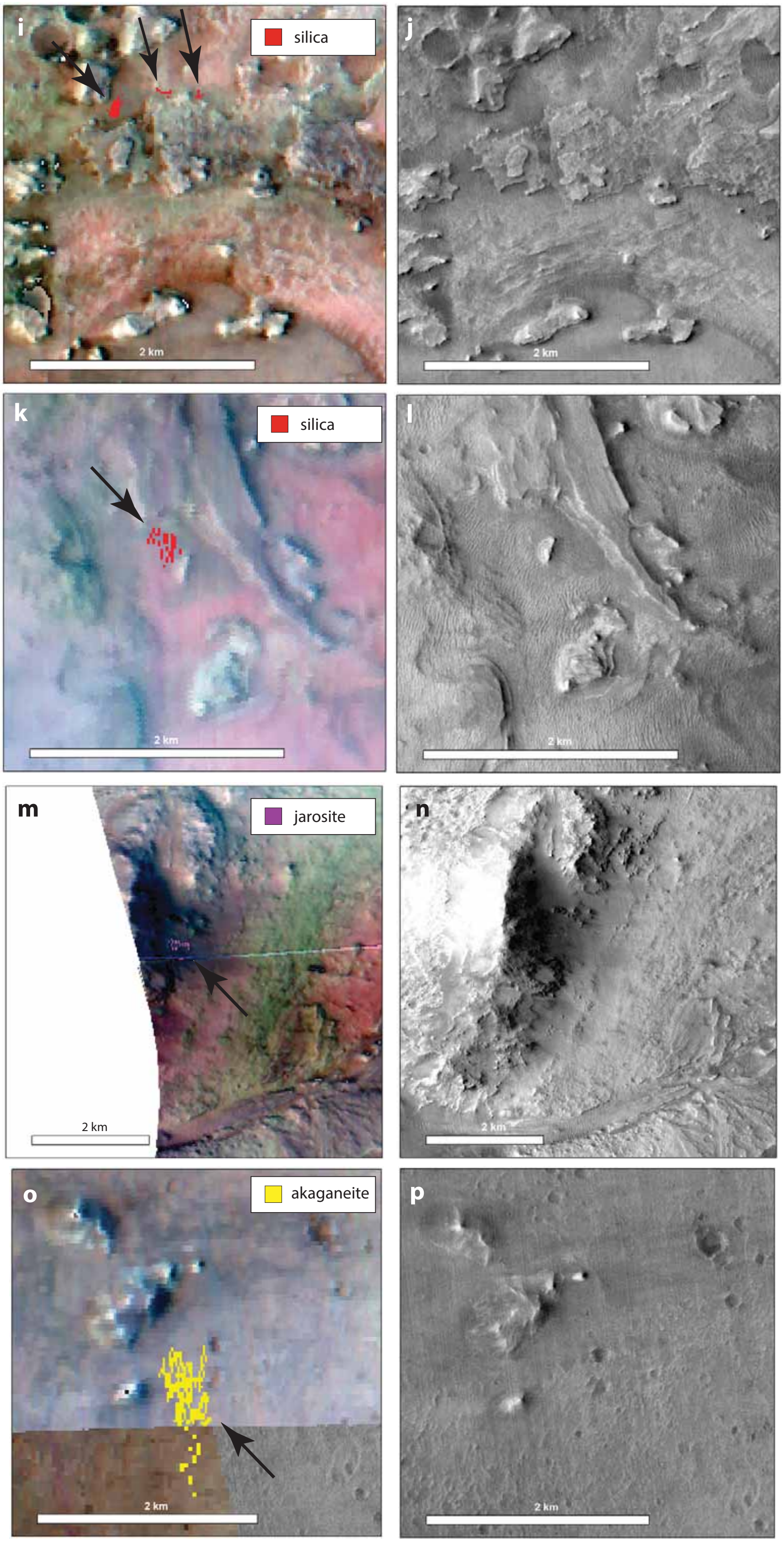


Figure 2. (i) CRISM silica detections in 047A3 with (j) accompanying area in CTX. (k) CRISM silica detection in 05C5E with (l) accompanying area in CTX. (m) CRISM jarosite detection in 05C5E with (n) accompanying area in CTX. (o) CRISM akaganeite detections in 05C5E and 040FF with (p) accompanying area in CTX.