Non-Homogeneous Hidden Markov Chain Models for Wavelet-Based Hyperspectral Image Processing

Marco F. Duarte and Mario Parente

***Abstract*— We consider the use of non-homogeneous Markov chain (NHMC) models for wavelet transformations of hyperspectral signatures to generate features for signal processing purposes. Inspired by the use of hidden Markov trees for natural images, the NHMC model enables the characterization of absorption bands and other structural features of mineral spectra that are used by experts in tasks like classification and unmixing, primarily in an *ad-hoc* fashion. We show that NHMC models can successfully identify and capture the information in a spectral signature dataset that can be exploited by standard classification algorithms to identify and differentiate spectral families. We also identify several metrics that can help determine whether each spectral band is informative to classification in a multiscale fashion.**

I. INTRODUCTION

Hyperspectral imaging systems (HSIs) encode electromagnetic radiation levels over hundreds of channels at each "pixel," spanning a wide range of wavelengths. Improvements in the quality, affordability, and deployability of HSIs have amplified their impact in a variety of scientific applications, including planetary and terrestrial geology, environmental monitoring, surveillance, and security [1]. Improvements in the spatial and spectral resolutions of HSIs enable scientists and engineers to consider complex information extraction applications that cannot be performed adequately with data acquired at low rates. Unfortunately, such improvements also strain the computational and visualization resources that are available to manage, understand, and decode the information that is present in the acquired hyperspectral data. Such bottlenecks are typical of data-rich settings, and its effects on science and technology have been characterized under the concept of *the data deluge* [2]. Practitioners must therefore navigate through massive quantities of very high-dimensional data to identify relevant features that encode the information desired for the application of interest; examples include spectral matching, image segmentation and denoising, and spectral unmixing.

Our focus in this paper is the facilitation of information extraction via the use of *probabilistic signal models for wavelet representations* of hyperspectral signals relevant to the particular problem of interest. While existing approaches that leverage wavelet representations are either *ad-hoc* in nature or limited to filtering techniques that manipulate the data but do not extract information [3, 4], we have recently introduced the use of *non-homogeneous hidden Markov chains* for wavelet-domain representations that provide a representation of the information contained in a hyperspectral dataset [5]. Wavelets are well suited for this purpose, as they compactly represent the fluctuations in reflectivity that practitioners rely on to perform signal processing with hyperspectral signatures. In this paper, we will show in more detail several examples of carefully obtained features from the wavelet model that can capture scientifically meaningful information on the spectral signatures for the minerals under study. In other words, we expand on the original aim of our model design in [5], which is to encode the physical information used by scientists to discriminate between the spectra of different minerals into quantitative features. While such "diagnostic" information can be extracted by experienced researchers and encoded by complicated *ad-hoc* rules, new rules have to be created for any additional spectral species not previously analyzed [6]. On the other hand, automatic techniques for spectral matching amount to calculating simple scalar scores, which encode some measure of similarity between spectral shapes, such as correlations, distances or vector angles [7]. The proposed model makes use of the full structural information of hyperspectral signals and interprets unknown information solely based on training data.

The authors are with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA. {mduarte,mparente}@ecs.umass.edu. This work was supported by the National Science Foundation under grant number IIS-1319585.

II. HYPERSPECTRAL IMAGE PROCESSING

A. Hyperspectral Imaging

Hyperspectral imagers (HSIs) or imaging spectrometers [8–16] measure electromagnetic energy scattered in their instantaneous field view in hundreds of narrow spectral channels. Hyperspectral datasets are organized into planes forming a data cube. Each plane corresponds to electromagnetic energy acquired over a spectral band for all pixels. Each spectral vector corresponds to the energy acquired at a given location (pixel) for all spectral bands. HSIs contribute significantly to earth and planetary observation and remote sensing [17–20]. Additional applications of HSIs include food safety [21, 22], pharmaceutical process monitoring and quality control [23, 24], and biology and medicine [25, 26], biometrics [27], and forensics [28].

Geological applications of HSIs take advantage of the unique capabilities of hyperspectral data to locate, map, and identify mineral assemblages present on the surface of Earth and other planetary bodies. Imaging spectroscopy gives outstanding results regarding the chemical composition and physical state of solid surfaces, providing clues about present and past activity and environmental conditions. The

identification of a mineral species from HSIs is routinely performed by planetary scientists and remote sensing experts using a wealth of techniques to match image pixel spectra to spectra of mineral samples acquired in a laboratory. This identification requires the extraction of *spectral reflectance*, a physical quantity measuring the percentage of light coming from a source (e.g., the Sun) which is reflected by the surface of a material.

B. Discrimination of Mineral Spectral Signatures

Spectral reflectance varies with wavelength for most materials because energy at certain wavelengths is scattered or absorbed to different degrees. These reflectance variations are evident if we compare spectral reflectance curves (plots of reflectance vs. wavelength) for different minerals, cf. Fig. 1. Pronounced downward deflections and dips of the spectral curves are known as *absorption bands* and mark the wavelength ranges for which the material selectively absorbs the incident energy. Their position, shape and strength can often be used to identify and discriminate different minerals.

The absorptions in reflectance spectra are due mainly to electronic and vibrational processes going on inside the mineral grain. Electronic absorptions occur due of the absorption of photons by atoms of transition elements (Ni, Cr, Co, Fe, etc.) in a crystal field [29]. These absorptions can occur for different energy levels depending on the atom and the crystal structure producing absorptions at different wavelengths, cf. Fig. 1. Crystal field processes are the source of the broad absorption bands found in the spectra of olivines (forsterite and fayalite) and pyroxenes in Figure 1(a). Vibrational excitations in minerals are produced by the partial rotational and translational movements of molecules within the crystal structure. These motions are called lattice modes and typically occur at very low energies (longer mid-infrared wavelengths), beyond about 20 μm. Still, their overtones and combinations, even if weaker in strength, are relevant at shorter wavelengths [30–32]. In reflectance spectroscopy, these weak absorptions can be measured easily and diagnostic information is routinely gained from second and third overtones and combinations. The sharp absorptions of carbonate minerals in Fig. 1(b) and of clay minerals in Fig. 1(c) are also due to vibrational overtones and combinations.

In order to discriminate between different mineral spectra, experts compare absorption band positions and shapes. In some cases the discrimination is simple, e.g., for the spectra of igneous minerals vs. phyllosilicate minerals. In other instances, the differences are quite subtle, as in the case of the clay minerals saponite and serpentine in Fig. 1, which present diagnostic absorptions at 2.314 and 2.321 μm, respectively, with the spectra being almost identical elsewhere.

III. Wavelet-Based Models

A. Wavelet Analysis

The wavelet transform is a widely used tool on signal and image processing. Specifically, the wavelet transform of a signal provides a multiscale space-frequency analysis of the signal's content, effectively encoding in a compact fashion the locations and scales in which the signal structure is present [33]. This energy compaction property is the main reason behind the popularity of wavelet transforms for signal processing and compression, including the state-of-the-art JPEG2000 image compression standard [34].

A 1D real-valued *undecimated wavelet transform* (UWT) of an N-sample signal $x \in \mathbb{R}^N$ is composed of *wavelet coefficients* $w_{s,n}$, each labeled by a scale $s \in \{1, \ldots, S\}$, $S \leq N$, and offset n, $1 \leq n \leq N$. A scaling coefficient w_0 captures the remaining energy of the signal. The coefficients are calculated using inner products $w_{s,n} = \langle x, \psi_{s,n} \rangle$, where $\psi_{s,n} \in \mathbb{R}^N$ denotes the mother wavelet function ψ dilated to scale s and translated to offset n:

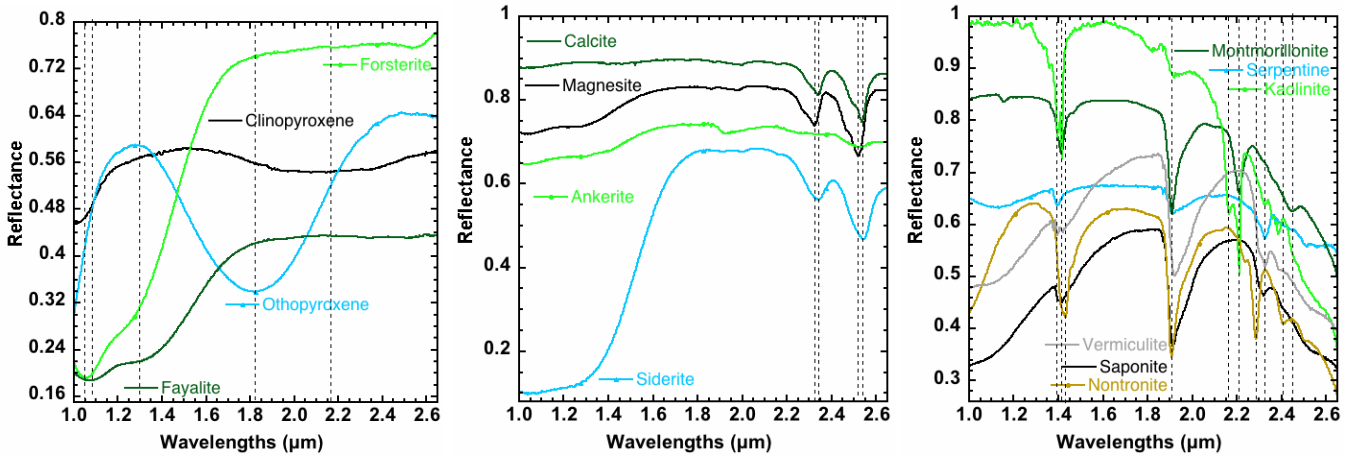
$$\psi_{s,n} = \frac{1}{\sqrt{s}} \phi \left(\frac{\lambda - n}{s} \right).$$

A coefficient $w_{s,n}$ at scale s describes a portion of the signal of size N/s. The coefficients can be organized into a 2-D array of size $S \times N$, where rows represent scales and columns represent samples. We say that each coefficient $w_{s,n}$, $s < S$, has a *child* coefficient $w_{s+1,n}$ at scale $s+1$; similarly, each coefficient $w_{s,n}$ at scale $s > 1$ has one *parent* $w_{s-1,n}$ at scale $s-1$.

Figure 2 shows an example hyperspectral signal and its undecimated wavelet coefficient array. A large wavelet coefficient (in magnitude) generally indicates the presence of a singularity inside its support; a small wavelet coefficient generally indicates a smooth region. This energy compaction property causes wavelet coefficients to have a peaky non-Gaussian distribution. Thanks to the nesting of child wavelets inside their parents, edges and discontinuities in general manifest themselves in the wavelet domain as chains of large coefficients propagating across scales from parents to children – a phenomenon known as *persistence*. Thus, wavelets both encode and exhibit structure from piecewise smooth signals, a property that has been exploited in a diverse array of areas, including image processing [33].

B. Non-Homogeneous Hidden Markov Chains

Inspired by the use of Hidden Markov Trees (HMTs) for statistical modeling of dyadic wavelet coefficients [35], we rely on a non-homogeneous hidden Markov chain (NHMC) to model the UWT coefficients. In contrast to HMTs, the choice of UWT yields a collection of NHMCs connecting each wavelet coefficient with its parent and child (if they exist). As in HMTs, our assumption is that large wavelet coefficients appear only sporadically at each wavelength and scale. Therefore, each wavelet coefficient $w_{s,n}$ is statistically modeled using a mixture of two Gaussians. The first component features a large variance $\sigma^2_{\mathsf{L},s,n}$ that models large nonzero coefficients and is anticipated to receive a small weight $p^{\mathsf{L}}_{s,n}$ during training, in order to encourage few such coefficients and preserve energy compaction. The second component features a small variance $\sigma^2_{\mathsf{S},s,n}$ that models small and zero-valued coefficients and is anticipated to receive a large weight $p^{\mathsf{S}}_{s,n} = 1 - p^{\mathsf{L}}_{s,n}$ during training due to the high likelihood of small and zero-valued wavelet coefficients.



(a) Igneous minerals (b) Carbonate minerals (c) Phyllosilicate (clay) minerals

Fig. 1: Examples of laboratory phyllosilicate (clay) mineral spectra and their absorption bands.

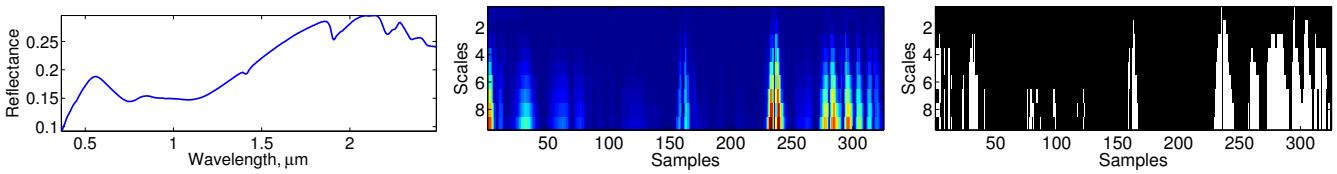


Fig. 2: *Left: Example spectral signature. Center: Corresponding wavelet coefficient array with $S = 9$ scales for $N = 325$ spectral samples using a Daubechies-4 undecimated wavelet transform. Small coefficient magnitudes are shown in blue, while the largest magnitudes are shown in red. Right: State labels obtained for the example spectral signature using a NHMC model. White corresponds to large labels; black to small.*

We distinguish these two components by associating to each wavelet coefficient $w_{s,n}$ an unobserved hidden state $\mathcal{S}_{s,n} \in \{\mathsf{S}, \mathsf{L}\}$, with probabilities $p(\mathcal{S}_{s,n} = \mathsf{S}) = p^{\mathsf{S}}_{s,n}$ and $p(\mathcal{S}_{s,n} = \mathsf{L}) = p^{\mathsf{L}}_{s,n}$. The value of $\mathcal{S}_{s,n}$ determines which of the two components of the mixture model is used to generate the probability distribution $f(w_{s,n})$ for $w_{s,n}$: $f(w_{s,n}|\mathcal{S}_{s,n} = \mathsf{S}) = \mathcal{N}(0, \sigma^2_{\mathsf{S},s,n})$ and $f(w_{s,n}|\mathcal{S}_{s,n} = \mathsf{L}) = \mathcal{N}(0, \sigma^2_{\mathsf{L},s,n})$, with $\sigma^2_{\mathsf{L},s,n} > \sigma^2_{\mathsf{S},s,n}$.

The persistence of large and small coefficients from parent to child is well-modeled by a Markov chain that links their coefficient states. This induces the NHMC graphical model on the coefficient array W, where the state $\mathcal{S}_{s,n}$ of a coefficient $w_{s,n}$ is affected only by the state $\mathcal{S}_{s-1,n}$ of its parent $w_{s-1,n}$. The NHMC is then completely determined by the set of likelihoods for the first state $\{p^{\mathsf{L}}_{1,n}\}$ and the set of state transition matrices for the different parent-child label pairs $(\mathcal{S}_{s,n}, \mathcal{S}_{s+1,n})$:

$$A_{s,n} = \begin{bmatrix} p^{\mathsf{S}\to\mathsf{S}}_{s,n} & p^{\mathsf{S}\to\mathsf{L}}_{s,n} \\ p^{\mathsf{L}\to\mathsf{S}}_{s,n} & p^{\mathsf{L}\to\mathsf{L}}_{s,n} \end{bmatrix}.$$

The persistence property implies that the values of $p^{\mathsf{L}\to\mathsf{L}}_{s,n}$ and $p^{\mathsf{S}\to\mathsf{S}}_{s,n}$ are significantly larger than their complements $p^{\mathsf{L}\to\mathsf{S}}_{s,n}$ and $p^{\mathsf{S}\to\mathsf{L}}_{s,n}$, respectively. Note that all the probabilities $p^{\mathsf{S}}_{s,n}$ and $p^{\mathsf{S}}_{s,n}$ can be computed from $\{A_{s,n}\}$, $p^{\mathsf{S}}_{1,n}$, and $p^{\mathsf{S}}_{1,n}$.

We separately train an NHMC on each of the N wavelengths or frequencies sampled by the hyperspectral acquisition device in order to capture the dynamics of observable spectral signatures for each wavelength individually. While the overlap between wavelet functions at a fixed scale and neighboring offsets introduces correlations between the corresponding wavelet coefficients, we consider each NHMC of parent-child wavelet coefficients independently for computational and interpretational reasons. The set of NHMC parameters Θ_n include the probabilities for the first hidden states $p^{\mathsf{S}}_{1,n}$ and $p^{\mathsf{L}}_{1,n}$, the state transition matrices $\{A_{s,n}\}^S_{s=1}$, and Gaussian variances $\{\sigma^2_{\mathsf{L},s,n}, \sigma^2_{\mathsf{S},s,n}\}^S_{s=1}$ — each of these for $1 \le n \le N$.

NHMC training (e.g., selecting the values of the model parameters Θ_n) is performed via an expectation maximization (EM) algorithm that maximizes the likelihood of a library of training data given the model parameters. The model training dataset is, ideally, as large and diverse as possible, ultimately aiming to obtain a universal model for all observable samples. Given the model, the state labels $\{\mathcal{S}_{s,n}\}$ for a given observation are obtained using a Viterbi algorithm [35, 36] that employs the Gaussian parameters and transition probabilities in $\{\Theta_n\}$. The algorithm also returns the likelihood $f(W|\Theta)$ of a wavelet coefficient array W under the model Θ as a byproduct. We propose the use of the parameters Θ, state labels $\mathcal{S}$, and likelihood $f(W|\Theta)$ as representations of the original hyperspectral signal x. As an example, one can use the binary state labeling vector as a feature for detection and classification of spectra in applications such as image segmentation and mineral identification [5].

IV. Benefits of NHMC Models for Hyperspectral Data

As previously illustrated, rule-based hyperspectral signal processing approaches focus on specific ranges of spectral bands and consider features at fixed scales that encode the structural information corresponding to the absorption bands in mineral spectra. However, different applications attribute different importance to particular regions of the spectrum during information extraction. For example, if the goal is spectral discrimination, the vibrational absorptions around 1.4 μm and 1.9 μm caused by water in the spectra of clays of Fig. 1 would not be useful due to their presence in all spectra. Similarly, spectral structure at certain scales is usually discarded, as this structure does not provide valuable information for discrimination; a prominent example is the removal of the spectrum continuum [4]. Based on these observations, the design of features to evaluate the discriminative power of specific regions of the spectral domain may allow practitioners to determine how to budget their sensing capabilities to match the regions and scales of the spectrum that are most informative to the task at hand.

Fortunately, the NHMC model provides several measures of the discriminative power of the spectrum at multiple scales by analyzing the available dataset and individually observing each spectral band at a multitude of scales. The modeling information for each spectral band can provide the practitioner with a quantification of the band's discriminative power. Recall that wavelet coefficients act as measures of the presence of discontinuities of different scales at different wavelengths of the hyperspectral signal. NHMC models leverage such a property to capture the frequency or dominance of discontinuities at each scale and wavelength, as captured by the likelihoods of and the transitions between small and large states.

Diversity in the presence of discontinuities at a given wavelength and scale is in general an indication that the corresponding hyperspectral signal information can be useful for hyperspectral signal processing problems, including classification. Features that are found to be redundant or irrelevant identify spectrum bands and scales that are non-informative for the problem at hand. A band with low discriminative power can be neglected. Similarly, the information for each scale can provide the practitioner with a quantification of the discriminative power achieved by a particular spectrum sampling density, which can be used to adjust the sampling rate of the hyperspectral sensor in the spectral domain appropriately.

Metrics of discriminative power can be obtained directly from the NHMC model parameters and spectral sample labels. For example, a large difference between the mixture Gaussian variances $\sigma^2_{\mathsf{S},s,n}$ and $\sigma^2_{\mathsf{L},s,n}$ is indicative of the presence of discontinuities at a given spectral band and scale. If the variances are close to each other, then a single Gaussian model suffices for the corresponding wavelet coefficient and the modeled does not discriminate between discontinuities and smooth regions. Similarly, if the model exhibits a state probability $p^{\mathsf{S}}_{s,n}$ or $p^{\mathsf{L}}_{s,n}$ that is very close to 0.5, then both states have equal likelihood and the model does not capture the anticipated sparse nature of the discontinuities among the dataset. Finally, one can consider the distribution of the labels L and S for each spectral band and scale among training samples to determine their usefulness in discriminating spectra, a process commonly known as *feature selection*. For example, a scale and wavelength pair whose state labels are consistently large (or consistently small) among all samples for a classification task are not relevant for labeling and should be excised during training and testing.

V. Experimental Evidence

We evaluate how the NHMC parameters and labels express discriminability information of the wavelet coefficients for two example hyperspectral signal processing problems. First, we evaluate the performance of classification using the NHMC-derived labels as features for the different spectra. Second, we analyze the information contained in the model parameters with regards to the discriminability for the aforementioned classification problems.

A. Classification with NHMC Labels

To test the performance of NHMC labels for classification, we consider a subset of the ENVI library corresponding to clays only, containing 57 samples and 12 classes; the list of minerals is provided in [4, Table 2]. The approach of [4] computes the 10-level undecimated wavelet decomposition of the mineral spectra and produces a signature by filtering the wavelet decomposition to ignore the 5 coarsest scales. Such signature is used as a feature for a minimum angle classifier. The reported accuracy is 89%. Figure 3 (left) shows the classification detail with misclassifications represented by points that deviate from the diagonal.

In comparison, we utilize the state labels $\mathcal{S}$ assigned to each material sample as its feature vector and employ a simple nearest-neighbor search classifier. Since the labels assume binary values, we use the Hamming distance measure to obtain the number of differing elements between two feature vectors. The results show that we significantly outperform the approach in [4], achieving a classification rate of 95% on the same dataset. The improved performance is likely due to the fact that in [4] the discrimination is based only on features extracted from single spectra, while our approach uses the richness of the whole spectral library to generate our model.

B. Discriminability From NHMC Parameters

To establish whether NHMC parameters are indicative of discriminability, we will consider two separate classification problems. The first problem considers a USGS ENVI library of endmember mineral spectral signatures totaling 481 samples and more than 200 classes. The second problem is as described in Section V-A.

For both examples given in this section, we consider three types of discriminability metrics. First, the ratio of variances $\sigma^2_{\mathsf{L}}/\sigma^2_{\mathsf{S}}$ provides a measure of the difference between the two

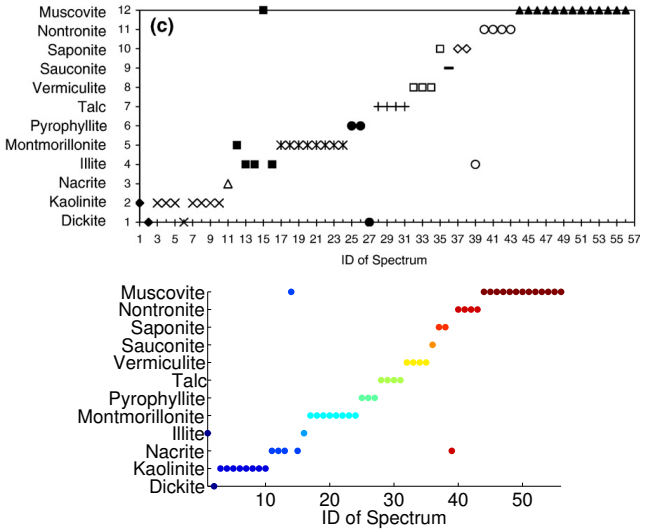


Fig. 3: *Classification results for USGS spectral library containing 57 clay samples from 12 classes. Left: Results from wavelet-filtered data with spectral angle mapper, achieving 89% classification rate (taken from [4]). Right: Results from NHMC state features with nearest neighbor search, achieving 95% classification rate.*

Gaussians present in the mixture as well as the samples matched to each one of the labels S and L. Second, the percentage of training samples that are labeled $\mathcal{S}_{s,n} = \mathsf{S}$ by the model provides a measure of the identifiability afforded by a given scale/wavelength across the training dataset; if this fraction is too low or too high, then the labels are likely not useful in signal processing tasks. Third, the probability of the small label according to the model $P(\mathcal{S}_{s,n} = \mathsf{S}|\Theta)$ provides a measure of the promotion of sparse structure provided by the model at each band and scale. When such probabilities are close to 0.5 (instead of the anticipated value near 1), one can consider the labeling provided by the model to be unreliable for the corresponding scale and wavelength.

The three types of metrics are shown for both problems in Figure 4. For the first problem, the first metric shows that all present spectral bands exhibit potential for discrimination through the disparity of the Gaussians present in each mixture. Furthermore, the second and third metrics identify the bands $1.34-1.48\mu$m, $1.86-1.98\mu$m, and $2.22-2.4\mu$m, that exhibit the highest variability in training labeling and model mixture probabilities, which is indicative of their particular importance in discriminating among sample classes. These metrics also coincide with the approach of [4] in assigning little discriminability to the coarsest scales of the wavelet transform. For the second problem, we see that all three types of features agree in identifying the spectral bands with indices $0.64-0.82\mu$m, $1.0-1.12\mu$m, and $1.42-1.66\mu$m as non-informative for the clay dataset. This agrees with inspection by human experts that determine lack of discriminating fluctuations in the spectra on the corresponding wavelengths.

C. Masking NHMC Labels

Finally, we evaluate the performance of this classification problem from NHMC labels that are masked according to their discriminating power. We repeat the classification experiment from Section V-A using NHMC labels, but in this case we select a subset of the labels by considering the discrimination metrics given in Section V-B.

We consider the three options as follows. First, we select labels for those wavelets coefficients for which the NHMC variances of the large and small states are different, i.e., for which $\sigma_{\mathsf{L}}/\sigma_{\mathsf{S}} > 1$. Second, we select labels for those wavelet coefficients for which the probability of the large and small states are unequal, i.e., $p^{\mathsf{S}}_{s,n}/p^{\mathsf{L}}_{s,n} \neq 1$. Third, we select labels for those wavelet coefficients for which the percentage of samples labeled large is neither 0% nor 100%.

Our results show that the outcome of the classification problem under these three maskings are each identical to that from using all of the labels computed, reaching the same 95% classification rate performance, verifying empirically that the choices of maskings derived from these metrics provide satisfactory identification of relevant information for the problem at hand.

VI. Conclusions and Future Work

We have showcased the improvements in standard signal processing tasks afforded by the proposed Non-Homogeneous Markov Chain model and the information conveyed by its parameters and the labeling of spectra provided by it. We have shown, for example, that the discriminability metrics proposed provide useful information for feature selection in spectrum classification. In future work, we expect to verify these properties and improvements for additional applications such as hyperspectral image segmentation. Additionally, we will elaborate on the robustness to variability of the acquired spectra of the proposed models, labels, and metrics. In particular, we will consider routinely observed variations in the continuum of the spectra and in the depth of the absorption bands that appear among samples of a given mineral.

References

[1] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Select. Top. Signal Proc.*, vol. 5, no. 2, pp. 354–379, Apr. 2012.

[2] R. G. Baraniuk, "More is less: Signal processing and the data deluge," *Science*, vol. 331, no. 6018, pp. 717–719, Feb. 2011.

[3] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2331–2338, Oct. 2002.

[4] B. Rivard, J. Feng, A. Gallie, and A. Sanchez-Azofeifa, "Continuous wavelets for the improved use of spectral libraries and hyperspectral data," *Remote Sensing of Environment*, vol. 112, pp. 2850–2862, 2008.

[5] M. Parente and M. F. Duarte, "A new semantic wavelet-based spectral representation," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, Gainesville, FL, June 2013.

[6] R. Clark, G. A. Swayze, K. Livo, S. Sutley, J. Dalton, R. McDougal, and C. Gent, "Imaging spectroscopy: Earth and planetary remote sensing with the USGS Tetracorder and expert systems," *J. Geophys. Res.*, vol. 108, no. E12, pp. 5131, doi:10.1029/2002JE001 847, 2003.

[7] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz, "The spectral image processing system (SIPS) – Interactive visualization and analysis of imaging spectrometer data," *Remote Sensing of Environment*, vol. 44, no. 2–3, pp. 145 – 163, May 1993.

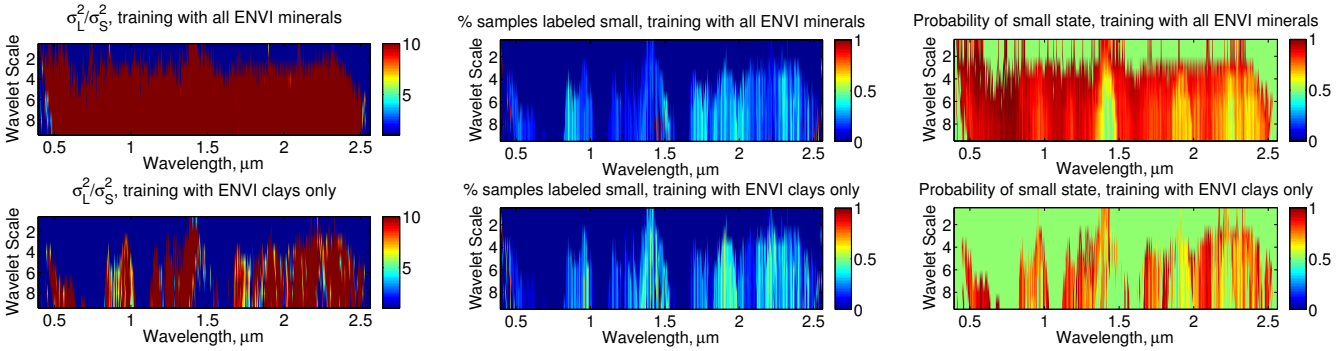


Fig. 4: *Metrics of important of wavelet coefficients learned by the NHMCs in two example hyperspectral signal processing examples. Top row: USGS ENVI endmember mineral dataset with 542 samples. Bottom row: 13-class clay mineral dataset extracted from ENVI with 56 samples used in [4]. Left column: Ratio of mixture Gaussian variances σ_L^2/σ_S^2. Values greater than 10 are displayed as red pixels. Small values (close to 1) signify poor discriminability caused by homogeneity in the training dataset. Center column: Percentage of training set samples that are labeled small (e.g. $\mathcal{S}_{s,n} = \mathsf{S}$). Percentages close to 0% or 100% are indicative of poor discriminability. Right column: Probability of small state (e.g., $P(\mathcal{S}_{s,n} = \mathsf{S}|\Theta)$) computed from NHMC model. Percentages close to 50% can be indicative of poor discriminability. In all cases, the presence of water absorption bands that are exclusive to clay minerals is evidenced by the lack of discriminability at the corresponding wavelengths when only clay minerals are used during training.*

[8] M. O. Smith, P. E. Johnson, and J. B. Adams, "Quantitative determination of mineral types and abundances from reflectance spectra using principal component analysis," in *Lunar and Planetary Sci. Conf.*, vol. 90, 1985, pp. 797–904.

[9] J. B. Adams, M. O. Smith, and P. E. Johnson, "Spectral mixture modeling: a new analysis of rock and soil types at the Viking Lander 1 site," *J. Geophys. Res.*, vol. 91, pp. 8098–8112, 1986.

[10] A. R. Gillespie, M. O. Smith, J. B. Adams, S. C. Willis, A. F. Fisher, and D. E. Sabol, "Interpretation of residual images: Spectral mixture analysis of AVIRIS images, Owens Valley, California," in *AVIRIS Workshop*, R. O. Green, Ed., vol. 90–54, 1990, pp. 243–270.

[11] G. Swayze, R. N. Clark, F. Kruse, S. Sutley, and A. Gallagher, "Ground-truthing AVIRIS mineral mapping at Cuprite, Nevada," in *JPL Airborne Earth Sci. Workshop*, 1992, pp. 47–49.

[12] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, *et al.*, "Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS)," *Remote Sens. Environment*, vol. 65, no. 3, pp. 227–248, 1998.

[13] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Proc. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

[14] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Proc. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[15] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging aplications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.

[16] D. Stein, S. Beaven, L. Hoff, E. Winter, A. Schaum, and A. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Proc. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[17] A. Plaza, J. A. Benediktsson, J. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, J. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environment*, vol. 113, pp. 110–122, 2009.

[18] M. E. Schaepman, S. L. Ustin, A. Plaza, T. H. Painter, J. Verrelst, and S. Liang, "Earth system science related imaging spectroscopy: an assessment," *Remote Sens. Environment*, vol. 3, no. 1, pp. 123–137, 2009.

[19] S. Murchie *et al.*, "Compact reconnaissance imaging spectrometer for Mars (CRISM) on Mars reconnaissance orbiter (MRO)," *J. Geophys. Res.*, vol. 112-5, 2007.

[20] ——, "Compact reconnaissance imaging spectrometer for mars investigation and data set from the mars reconnaissance orbiter's primary science phase," *J. Geophys. Res.*, vol. 114, no. E00D07, 2009.

[21] S. Mahest, A. Manichavsagan, D. Jayas, J. Paliwall, and N. White, "Feasibiliy of near-infrared hyperspeectral imaging to differentiate Canadian wheat classes," *Biosystems Eng.*, vol. 101, no. 1, pp. 50–57, 2008.

[22] R. Larsen, M. Arngren, P. Hansen, and A. Nielsen, "Kernel based subspace projection of near infrared hyperspectral images of maize kernels," *Image Analysis*, pp. 560–569, 2009.

[23] A. de Juan, M. Maeder, T. Hancewicz, L. Duponchel, and R. Tauler, "Chemometric tools for image analysis," *Infrared and Raman spectroscopic imaging*, pp. 65–109, 2009.

[24] M. B. Lopes, J.-C. Wolff, J. Bioucas-Dias, and M. Figueiredo, "NIR hyperspectral unmixing based on a minimum volume criterion for fast and accurate chemical characterization of counterfeit tablets," *Analytical Chemistry*, vol. 82, no. 4, pp. 1462–1469, 2010.

[25] G. Begelman, M. Zibulevsky, E. Rivlin, and T. Kolatt, "Blind decomposition of transmission light microscopic hyperspectral cube using sparse representation," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1317–1324, 2009.

[26] H. Akbari, Y. Kosugi, K. Kojima, and N. Tanaka, "Detection and analysis of the intestinal ischemia using visible and invisible hyperspectral imaging," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 8, pp. 2011–2017, 2010.

[27] A. Picon, O. Ghita, P. F. Whelan, and P. M. Iriondo, "Fuzzy spectral and spatial feature integration for classification of nonferrous materials in hyperspectral data," *IEEE Trans. Ind. Informat.*, vol. 5, no. 4, pp. 483–494, 2009.

[28] L. N. Brewer, J. A. Ohlhausen, P. G. Kotula, and J. R. Michael, "Forensic analysis of bioagents by X-ray and TOF-SIMS hyperspectral imaging," *Forensic Sci. Int.*, vol. 179, no. 2–3, pp. 98–106, 2008.

[29] R. Burns, *Mineralogical Applications of Crystal Field Theory,*, 2nd ed. Cambridge University Press, 1993.

[30] G. Hunt, "Spectral signatures of particulate minerals in the visible and near infrared," *Geophysics*, vol. 42, p. 501, 1977.

[31] ——, *Spectroscopic properties of rocks and minerals*, ser. Handbook of Physical properties of rocks. CRC Press, 1982.

[32] S. Gaffey, L. McFadden, D. Nash, and C. Pieters, *Ultraviolet, Visible, and Near-infrared Reflectance Spectroscopy: Laboratory spectra of Geologic Materials*, ser. Remote Geochemical Analysis: Elemental and Mineralogical Composition. Cambridge University Press, 1993.

[33] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA, USA: Academic Press, 1999.

[34] D. Taubman and M. Marcellin, *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer, 2001.

[35] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using Hidden Markov Models," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[36] L. R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, Feb. 1989.