

# TAILORING NON-HOMOGENEOUS MARKOV CHAIN WAVELET MODELS FOR HYPERSENSPECTRAL SIGNATURE CLASSIFICATION

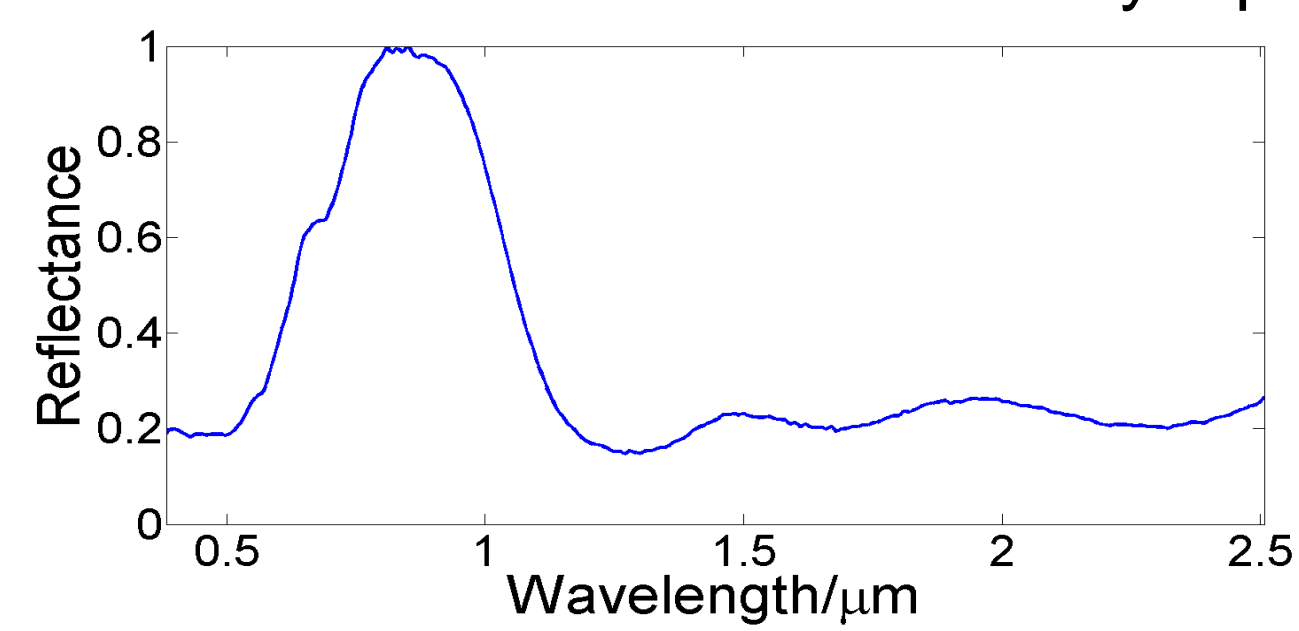
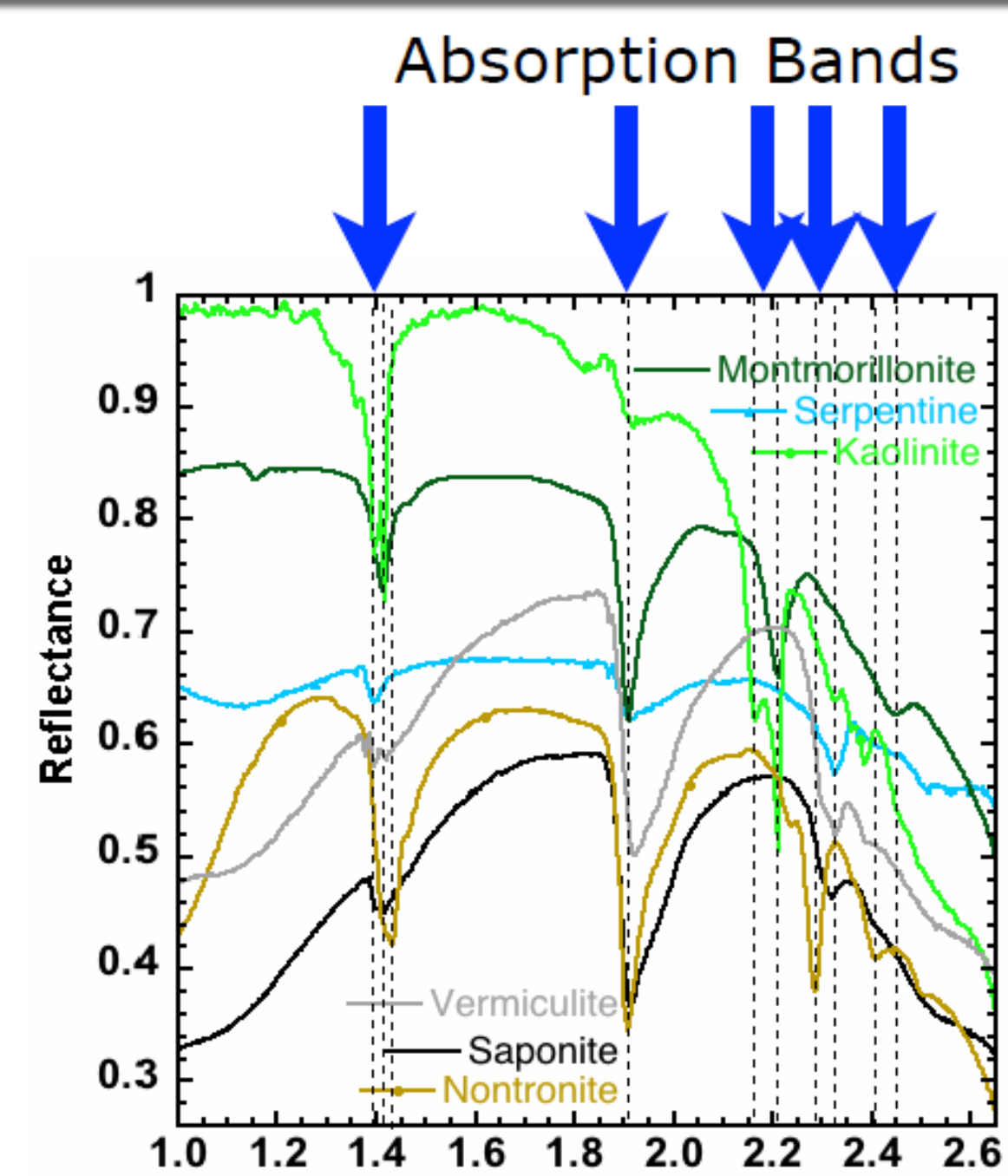
Siwei Feng, Yuki Itoh, Mario Parente, Marco F. Duarte

## Introduction

- Improvements in hyperspectral imaging systems (HSIs) – spatial and spectral resolution – enable researchers to consider complex feature extraction from hyperspectral data.
- Practitioners have to face massive amounts of high-dimensional data.
- Experienced researchers characterize minerals by shape and positions of fluctuations in their reflectance spectral curves. Current automated information extraction approaches rely on complicated ad-hoc rules (e.g., Tetracorder).
- We focus on the facilitation of information extraction process by using mathematical models for hyperspectral signals. We select a non-homogeneous hidden Markov chain (NHMC) model [Duarte and Parente '13] and present modifications to the original model to enable it to characterize spectral data as fully as possible

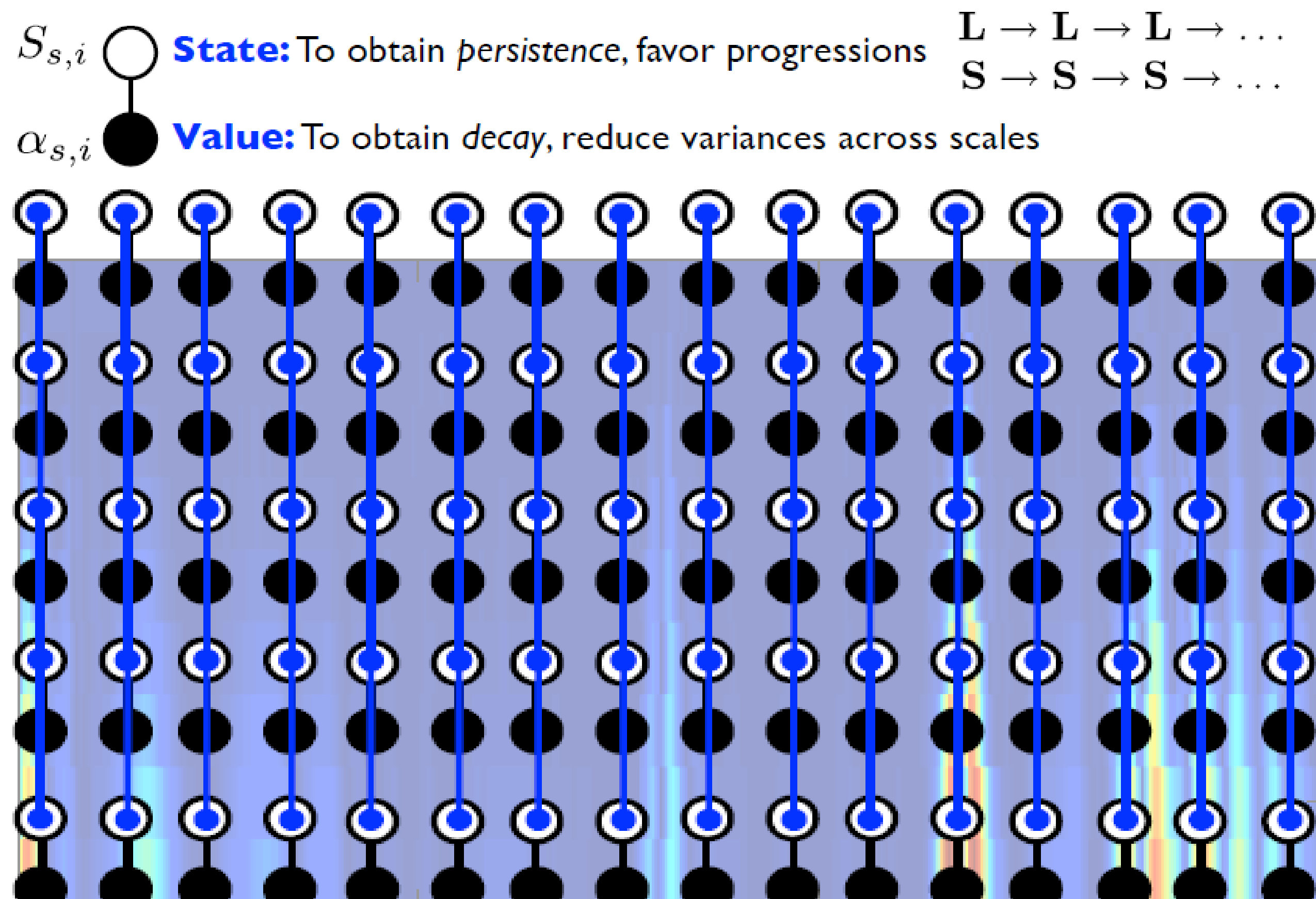
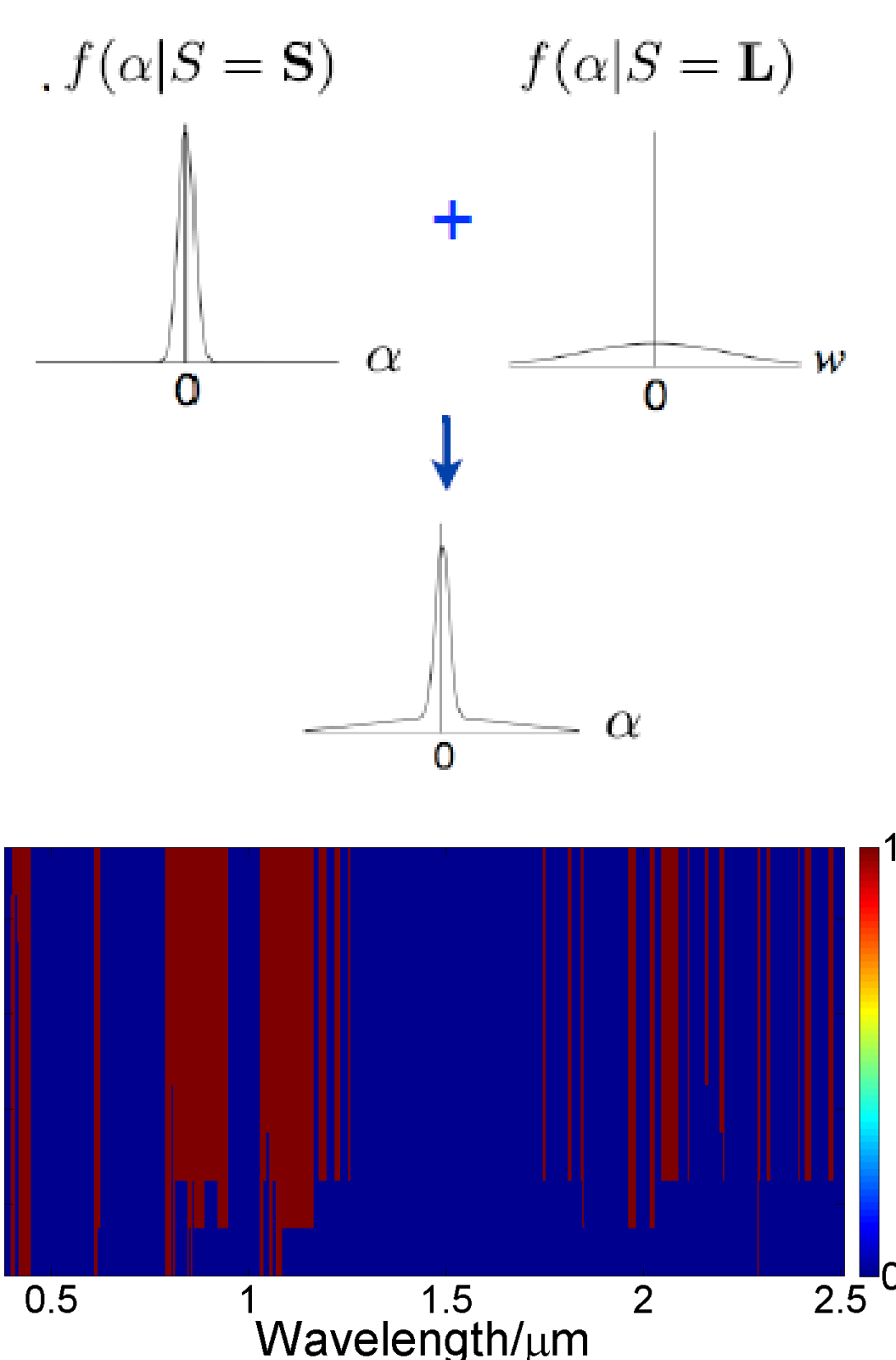
## Wavelet Representations of Hyperspectral Signatures

- Shape and positions of certain fluctuations (e.g. absorption bands) in hyperspectral signature curves provide identification of mineral species.
- Wavelet coefficients act as “detectors” of fluctuations in hyperspectral signatures.
- Wavelet type: undecimated wavelet transform (UWT) – provides maximum flexibility and resolution on the set of scales and offsets.
- Mother wavelet: Haar – encodes both magnitudes and fluctuations of slopes.
- Final representation: 2-D wavelet array. Rows: scales; columns: offsets/wavelengths.
- Structure of wavelet coefficient array: sparsity and persistence.

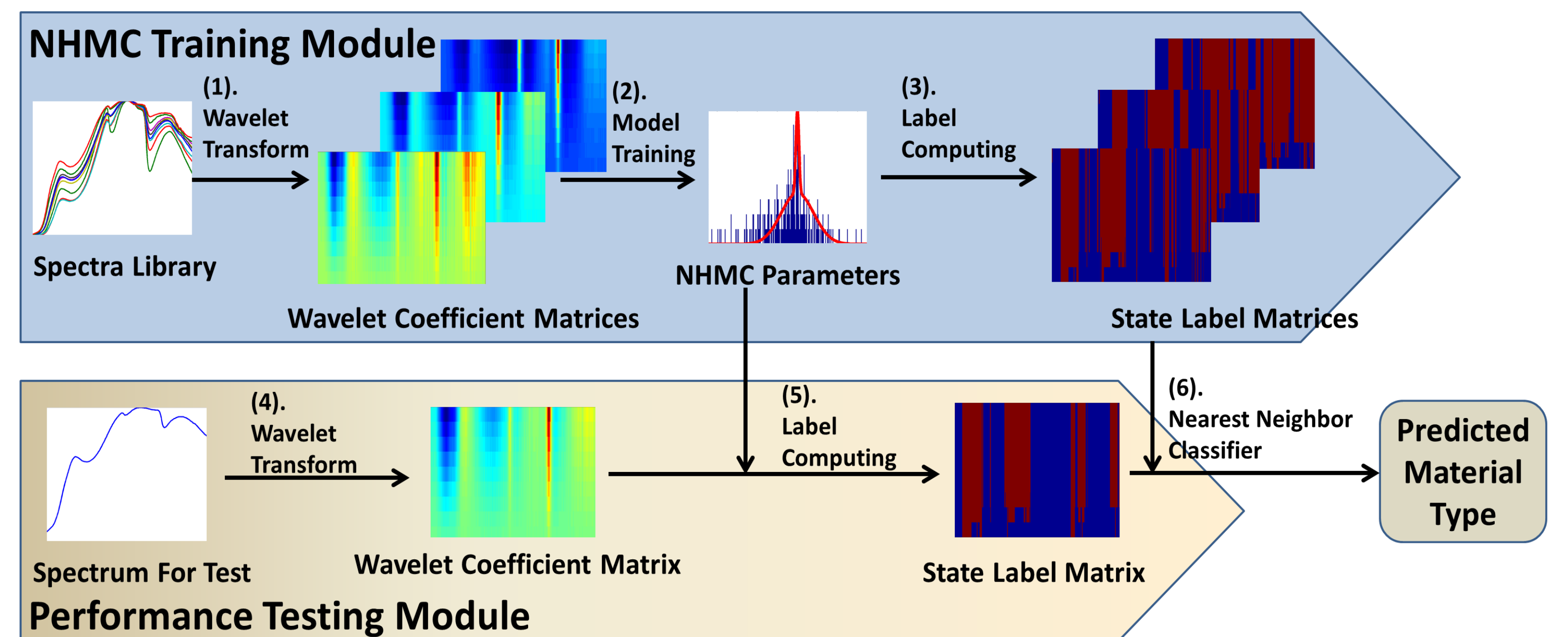


## Non-homogeneous Hidden Markov Chain Model

- Motivation: to extract scientifically meaningful information on the spectral signatures.
- Implementation: non-homogeneous hidden Markov chain (NHMC): a statistical model that encodes the structure of wavelet coefficients.
- Wavelet coefficient characterization: zero-mean Gaussian mixture model (GMM).
- GMM parameter training algorithm: Expectation-maximization (EM) algorithm.
- Classification feature: 2-D state label array of corresponding 2-D wavelet coefficient array.
- State labels are generated using Viterbi algorithm.

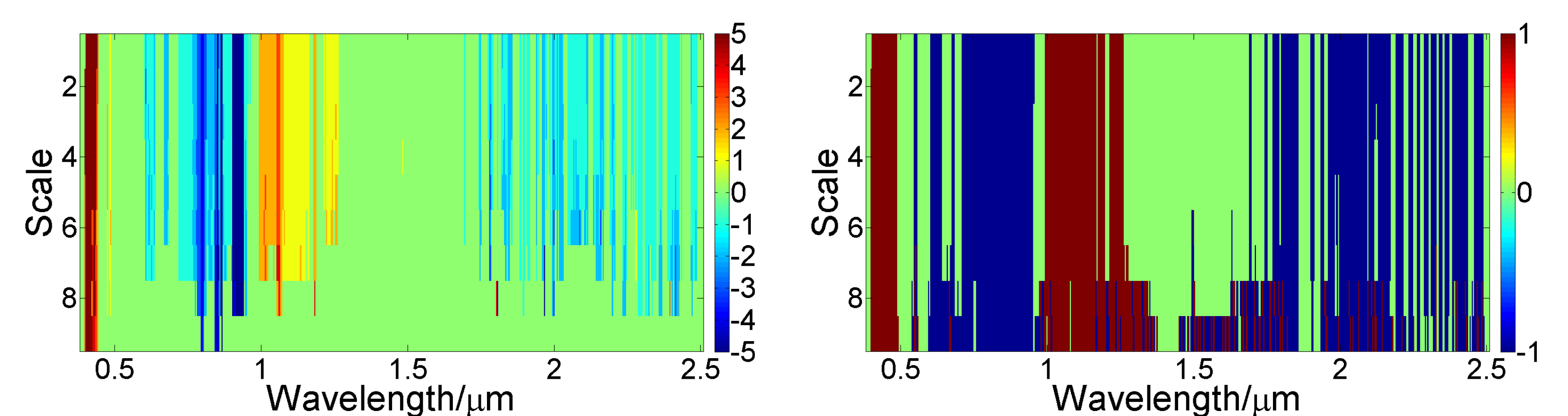
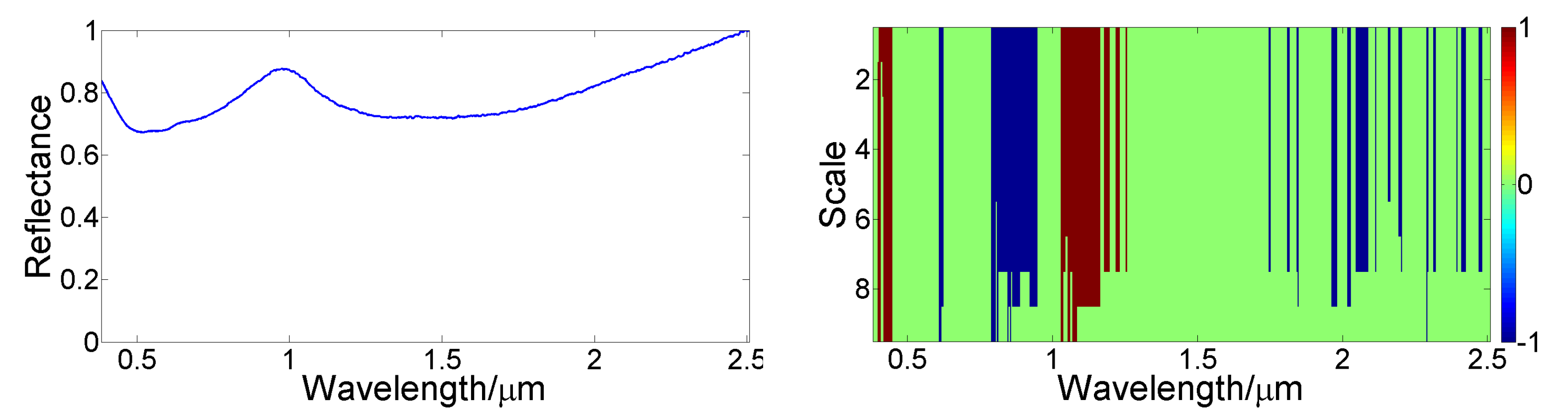


## NHMC-Based Hyperspectral Classification



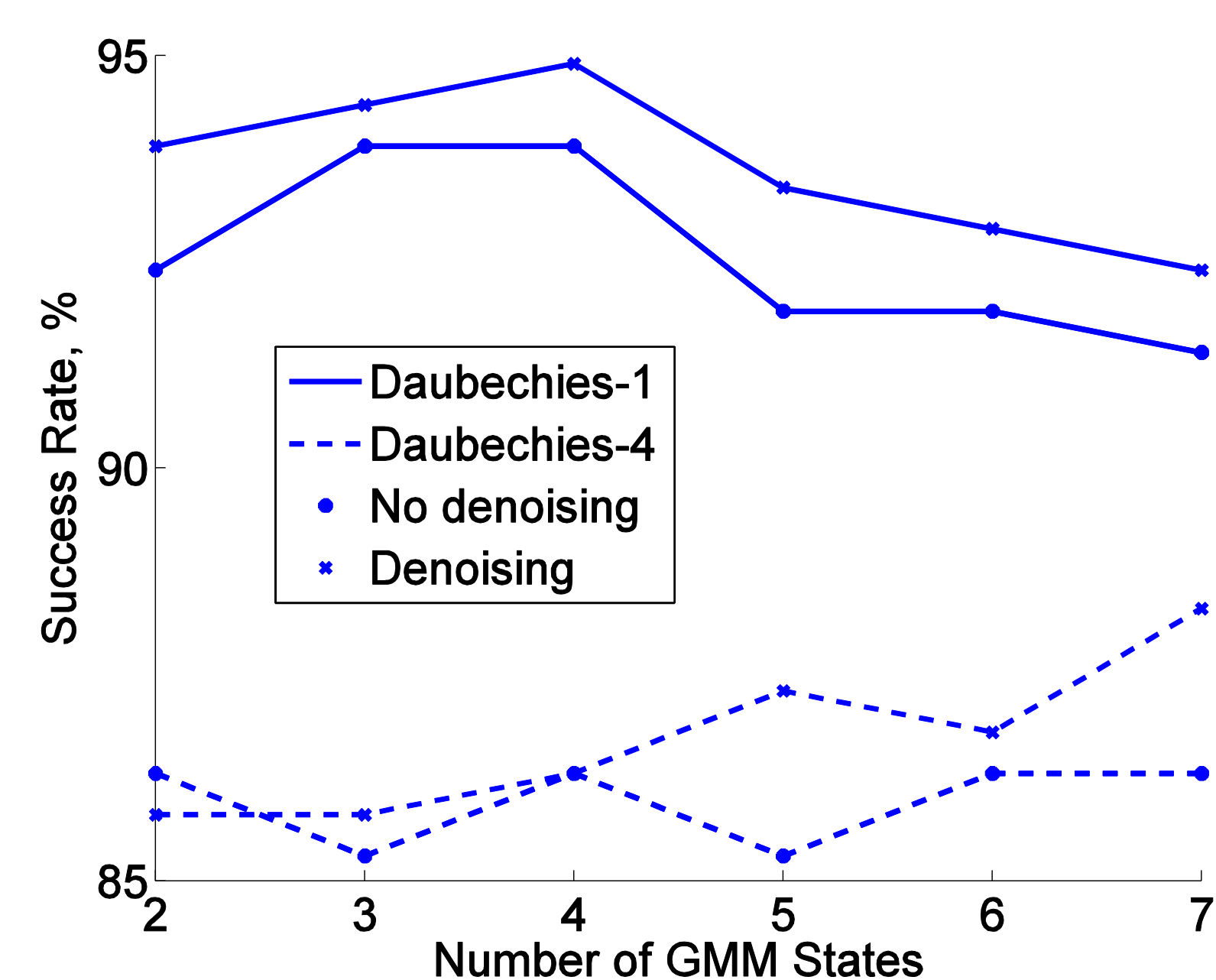
## NHMC Improvements

- Signs:** Add the signs of Haar wavelet coefficients to corresponding state labels to enable the state label array to reflect orientations of slopes in spectral curve.
- States:** Two states not sufficient to discriminate smooth regions from fluctuations: Increase number of GMM Gaussian components to capture moderate-level fluctuations.
- Mixture of Gaussians (MoG):** Large number of states might lead to multitude of label arrays for a given class: Combine the advantages of a binary-state GMM and  $k$ -state GMM through keeping the original “smallest” state as the “small” state and combining the other  $k-1$  states as the “large” state, i.e., large state is a MoG.
- Denosing:** Noise appears to be present in spectral libraries; perform signature denosing in wavelet domain via soft thresholding. [Donoho '93]



## Experimental Results

- USGS AVIRIS remote sensing database: minerals, vegetation, etc.
- 244 samples, 26 classes, 5+ samples per class.
- Each reflectance spectrum is normalized by its maximum value.
- Model selection: Use 5-fold cross validation (CV) to pick the model with highest CV accuracy. Best performance obtained with Daubechies-1 (Haar) wavelet and 4 GMM states on denoised signatures.



NHMC, Haar, 4-GMM, Denoised	91.5%
[Rivard et al. '08]	85.1%
Spectral Information Divergence (Nearest Neighbor)	83.0%
Spectral Correlation Measure (Nearest Neighbor)	80.9%
Spectral Angle Measure (Nearest Neighbor)	80.9%
Euclidean Distance (Nearest Neighbor)	80.9%