~~Charles B. Sither~~

~~School of Health Sciences~~

~~Western Carolina University~~

~~Cullowhee, NC 28723~~

~~Email: cbsither@gmail.com~~

~~Phone: (828) 734-0682~~

**Mid-Infrared Spectroscopy Detects Gonotrophic Status in *Aedes triseriatus* (Say)**

Charles B. Sither1\*, Bradley F. Guilliams2, Mark A. Rothermund2, Brian D. Byrd1, and Scott W. Huffman2

1. School of Health Sciences, Western Carolina University, USA
2. Department of Chemistry and Physics, Western Carolina University, USA

\*Corresponding Author

**Abstract**

**Introduction**

d

**Methods**

*Specimen Rearing and Storage*

We reared *Aedes triseriatus* mosquitoes (MSU strain) at standard laboratory rearing conditions (27oC, 75% RH, 16:8 day/night photoperiod). Emerged adults (>500) were placed in a single rearing cage and held for 7 days allowing for mating and maturation. Approximately 250 mosquitoes were removed from the original cage and fed to repletion on human blood. We provided ovipositioning substrate 72 hours post blood feeding. After ovipositioning, mosquitoes were concurrently removed from both the original cage (non-bloodfed; nulliparous) and the secondary cage (bloodfed; suspected parous). Parous and nulliparous mosquitoes were killed by freezing and stored at -20oC. A total of 170 mosquitoes were then selected for MIRS analysis: 99 previously blood-fed, and 71 not blood-fed.

*Mid-Infrared Spectroscopy*

We measured the mid-infrared region (4000–650 cm−1) using a NicoletTM Centaurμs™ housing a Nicolet™ IS™ 10 FTIR spectrophotometer equipped with a potassium bromide (KBr) beamsplitter at 4 cm-1 resolution.

A single IR spectrum of each mosquito tibia was acquired using a ## SPEC MODEL HERE Brads work: Nicolet model Centaurus infrared microscope ## FT-IR spectrophotometer equipped with ## MEASUREMENT APPARATUS AND SPEC MEASUREMENT TYPE (CAVITY RING DOWN?) ## and transformed with the ## SOFTWARE HERE ## software. To ensure consistency amongst measurement locations between mosquito specimens, the tibia from the midleg of each mosquito was removed and slide mounted prior to spectral acquisition. Background and MIR spectra were determined by averaging over >60 scans at a resolution of 4cm-1 with a spectral range 4000 to 650cm-1

*Gonotrophic Status Prediction*

We implemented a Partial Least Squares regression (PLS) algorithm with Python3 (v3.12.0) in the scikit-learn package (v1.5.1) to distinguish between nulliparous and parous *Aedes triseriatus* mosquitoes. The analysis workflow is visually summarized in figure 1. To ensure data quality, we subjected all spectra to three preprocessing algorithms derived from Siria *et al*. (2022) which filter spectra that are low intensity, contain high amounts of atmospheric interference, or distorted spectra. Additionally, we transformed all data first by performing an L2 normalization (CITE) then a Savitzky-Golay filter (CITE). After preprocessing and prior to model training and assessment, we randomly assigned 80% of the data to a training dataset and 20% to a test dataset stratified by prediction status (nulliparous/parous) and sampling date. The PLS algorithm requires a single hyperparameter, number of components, that must be decided in advance of model training. To determine the optimal hyperparameter value, we performed a grid search over *n* components from *n* = 2 to *n* = 12 and selected the optimal components based on which component minimized mean squared error (MAE) over 10 shuffled K-folds. We attempted a second optimization step involving Variable Importance in Project (VIP) (CITE HERE) excluding variables with a VIP score < 1. Using these selected features, we performed an additional grid search on VIP scores ≥ 1 then trained a second model on said VIP regions with VIP scores ≥ 1. From here on out, Base Model refers to the non-VIP optimized model, while VIP Model refers to the model that underwent VIP feature selection. We compared both models by evaluating accuracy, recall, precision, and F1 scores with confidence intervals calculated by bootstrapping test data predictions.

**Results**

We obtained MIR spectra in in the 4000cm-1 to 650cm-1 region for all 170 female *Ae. triseriatus* mosquitoes. During data preprocessing, 4 spectra were discarded due to distorted backgrounds yielding a final sample size of 166 spectra. All L2 normalized spectra were transformed with a Savitzky-Golay filter containing the following parameters: window size = 5, polynomial order = 2, and derivative order = 2 (Figure 2). Training and test data stratified by sampling date and parity status yielded 56 nulliparous and 76 parous *Ae. triseriatus* mosquitoes in the training data along with 21 parous and 13 nulliparous mosquitoes in the test dataset.

**Discussion**

**References**

1. Detinova TS. Age-grouping methods in Diptera of medical importance, with special reference to some vectors of malaria. World Health Organization Monograph Series. 1962;13–191.

2. Joy TK, Gutierrez EHJ, Ernst K, Walker KR, Carriere Y, Torabi M, et al. Aging Field Collected Aedes aegypti to Determine Their Capacity for Dengue Transmission in the Southwestern United States. PLOS ONE. 2012;7:e46946.

3. Colchero F, Clark JS. Bayesian inference on age-specific survival for censored and truncated data. Journal of Animal Ecology. 2012;81:139–49.

4. Wagner I, Grigoraki L, Enevoldson P, Clarkson M, Jones S, Hurst JL, et al. Rapid identification of mosquito species and age by mass spectrometric analysis. BMC Biol. 2023;21:10.

5. Siria DJ, Sanou R, Mitton J, Mwanga EP, Niang A, Sare I, et al. Rapid age-grading and species identification of natural mosquitoes for malaria surveillance. Nat Commun. 2022;13:1501.

6. Mwanga EP, Siria DJ, Mshani IH, Mwinyi SH, Abbasi S, Jimenez MG, et al. Rapid classification of epidemiologically relevant age categories of the malaria vector, Anopheles funestus. Parasites Vectors. 2024;17:143.

7. Jiménez MG, Babayan SA, Khazaeli P, Doyle M, Walton F, Reedy E, et al. Prediction of mosquito species and population age structure using mid-infrared spectroscopy and supervised machine learning. Wellcome Open Research [Internet]. 2019 [cited 2024 Sep 18];4. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6753605/

**Tables**

Table 1. Base model and VIP model metrics on test data.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Metric | Value | CI (2.5%) | | CI (97.5%) |
|  | Recall/Sensitivity | 0.905 | 0.765 | 0.99 | |
| Base Model | Precision/Specificity | 0.864 | 0.7 | 0.99 | |
|  | Accuracy | 0.853 | 0.735 | 0.971 | |
|  | F1 Score | 0.884 | 0.765 | 0.971 | |
|  | Recall/Sensitivity | 0.905 | 0.766 | 0.99 | |
| VIP Model | Precision/Specificity | 0.905 | 0.762 | 0.99 | |
|  | Accuracy | 0.882 | 0.765 | 0.971 | |
|  | F1 Score | 0.905 | 0.791 | 0.98 | |

**Figures**

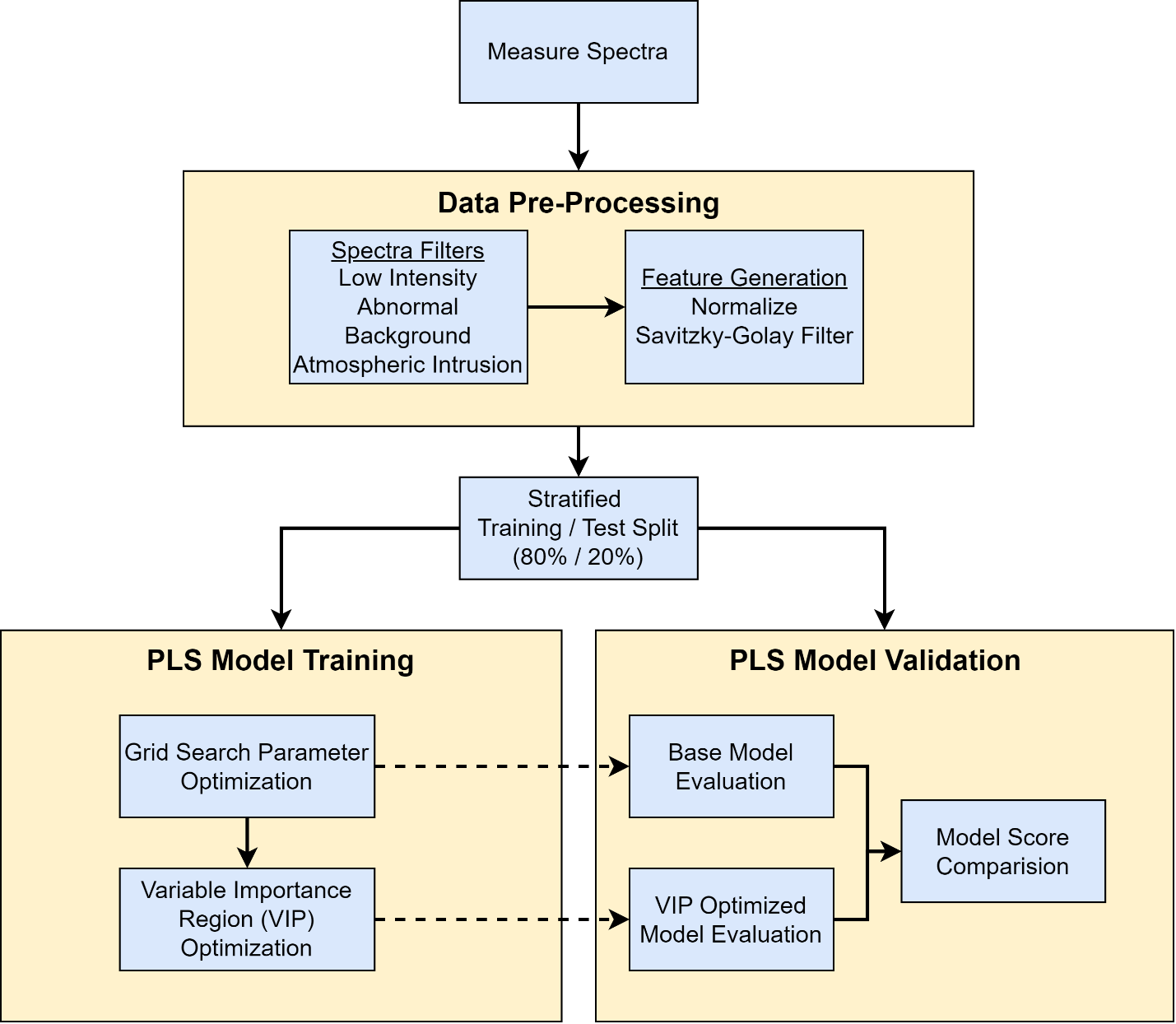


Figure 1. Workflow diagram for PLS model training and evaluation.

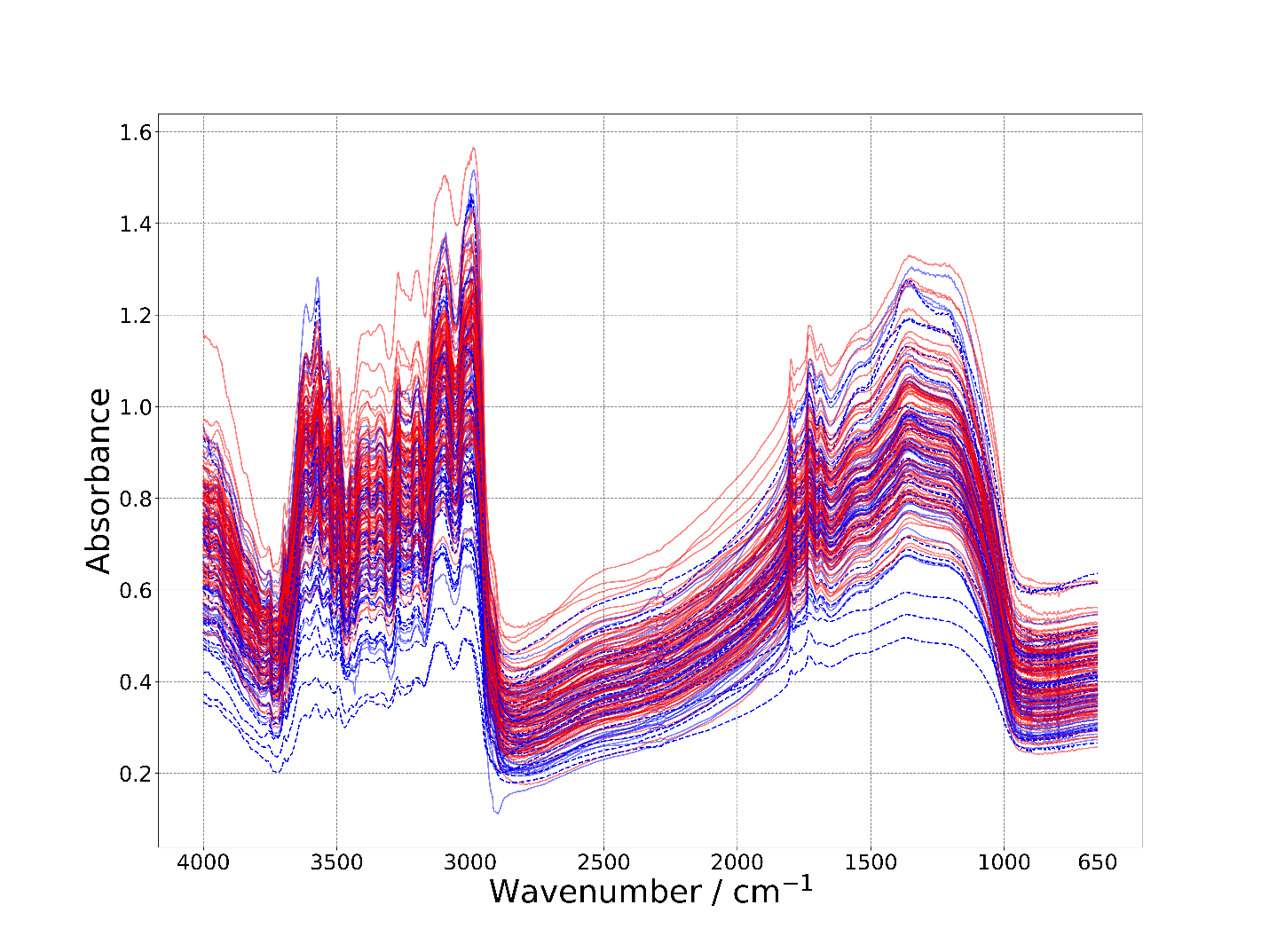


Figure 2. Untransformed MIR spectra from *Aedes triseriatus*. Parous and nulliparous readings are labeled red and blue, respectively.

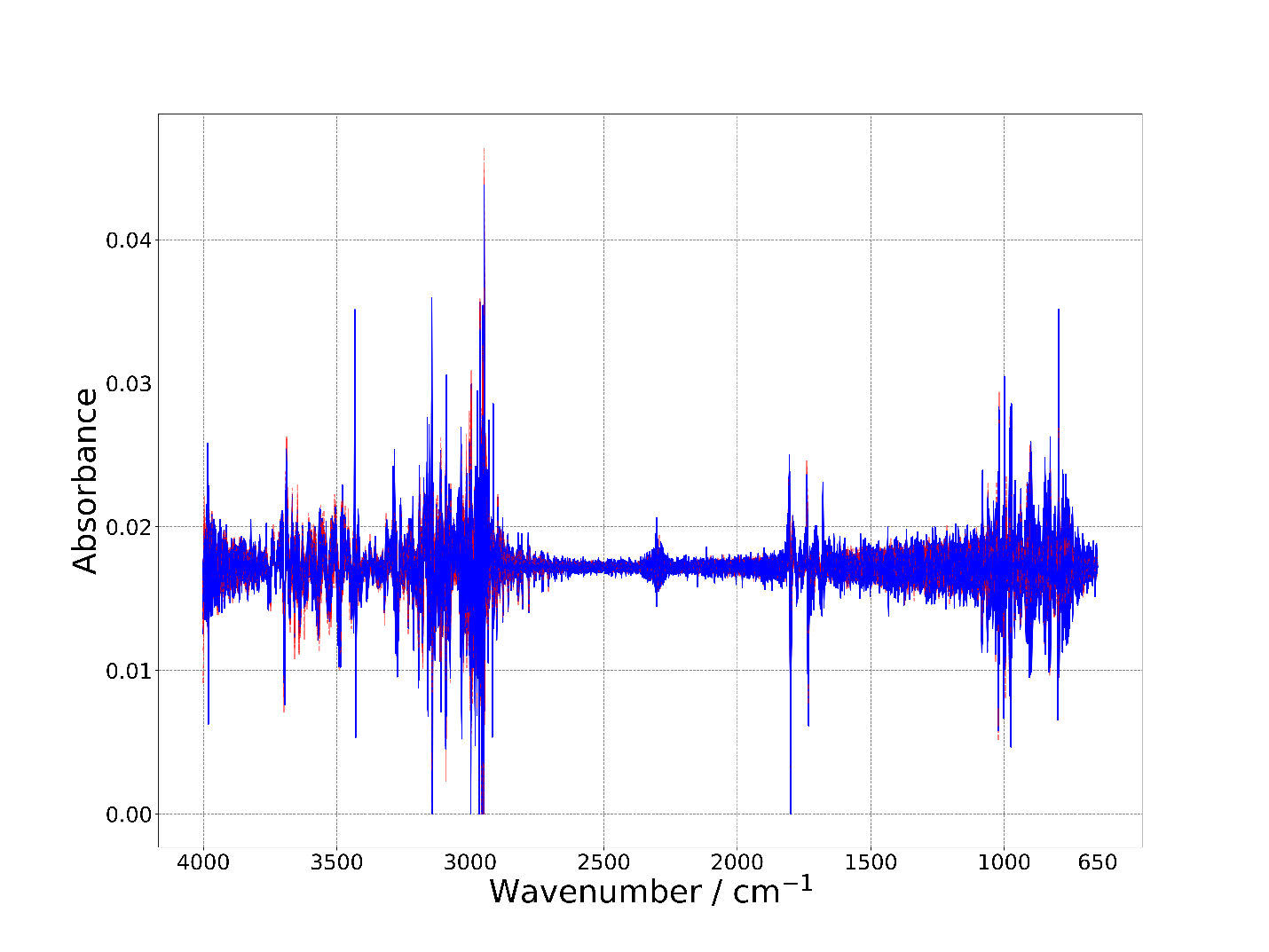


Figure 3. L2 normalized and Savitzky-Golay filtered MIR spectra.

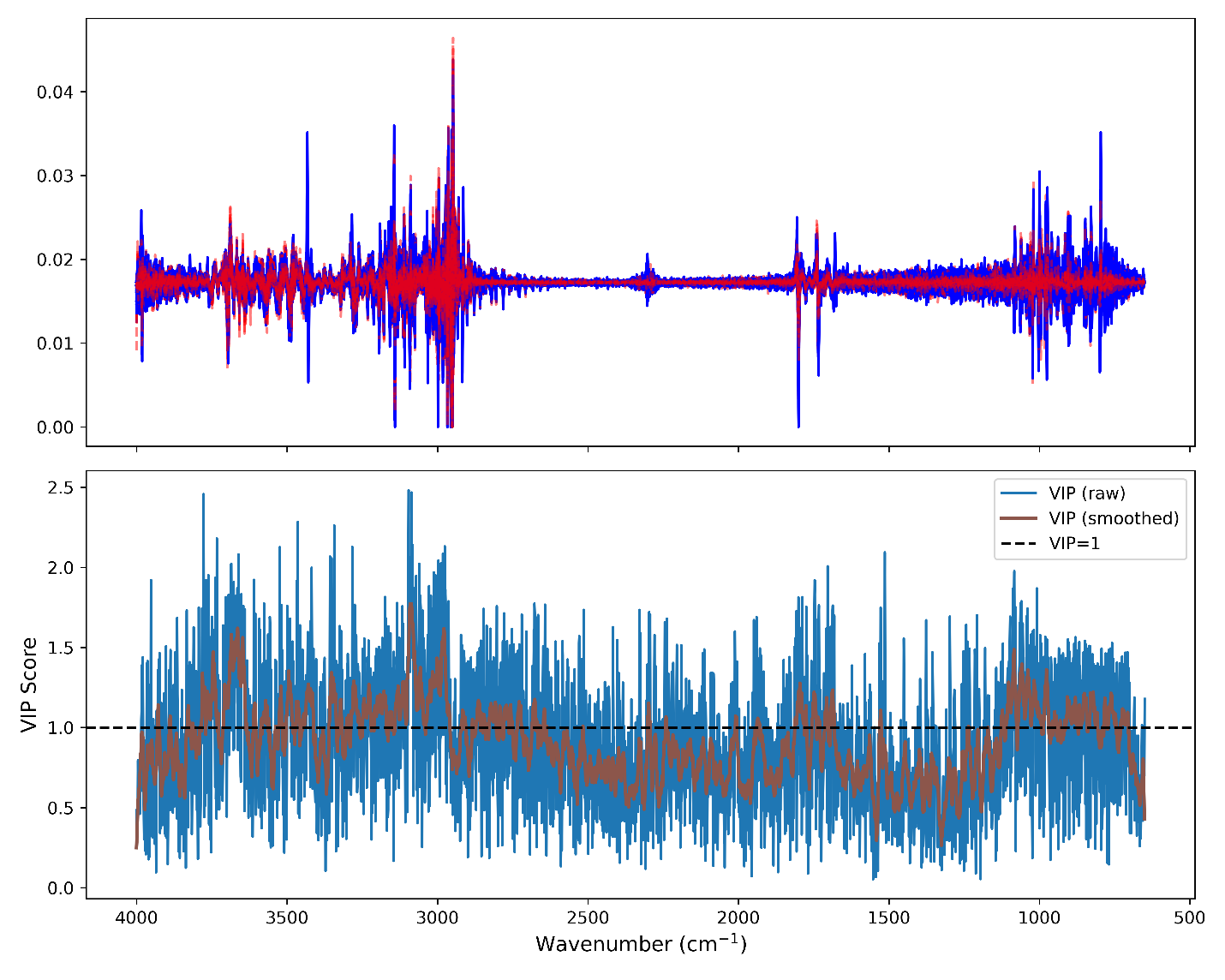


Figure 4. VIP regions (bottom) and transformed spectra (top). VIP scores ≥ 1 are considered important features.