Towards Optimal Classifier of Spectroscopy Data

David D. Pokrajac, Poopalasingam Sivakumar, Yuri Markushin, Mukti Rana, Gary Holness, Jinjie Liu, Noureddine Melikechi, **Delaware State University**

Daniela Milovic, **University of Nis**

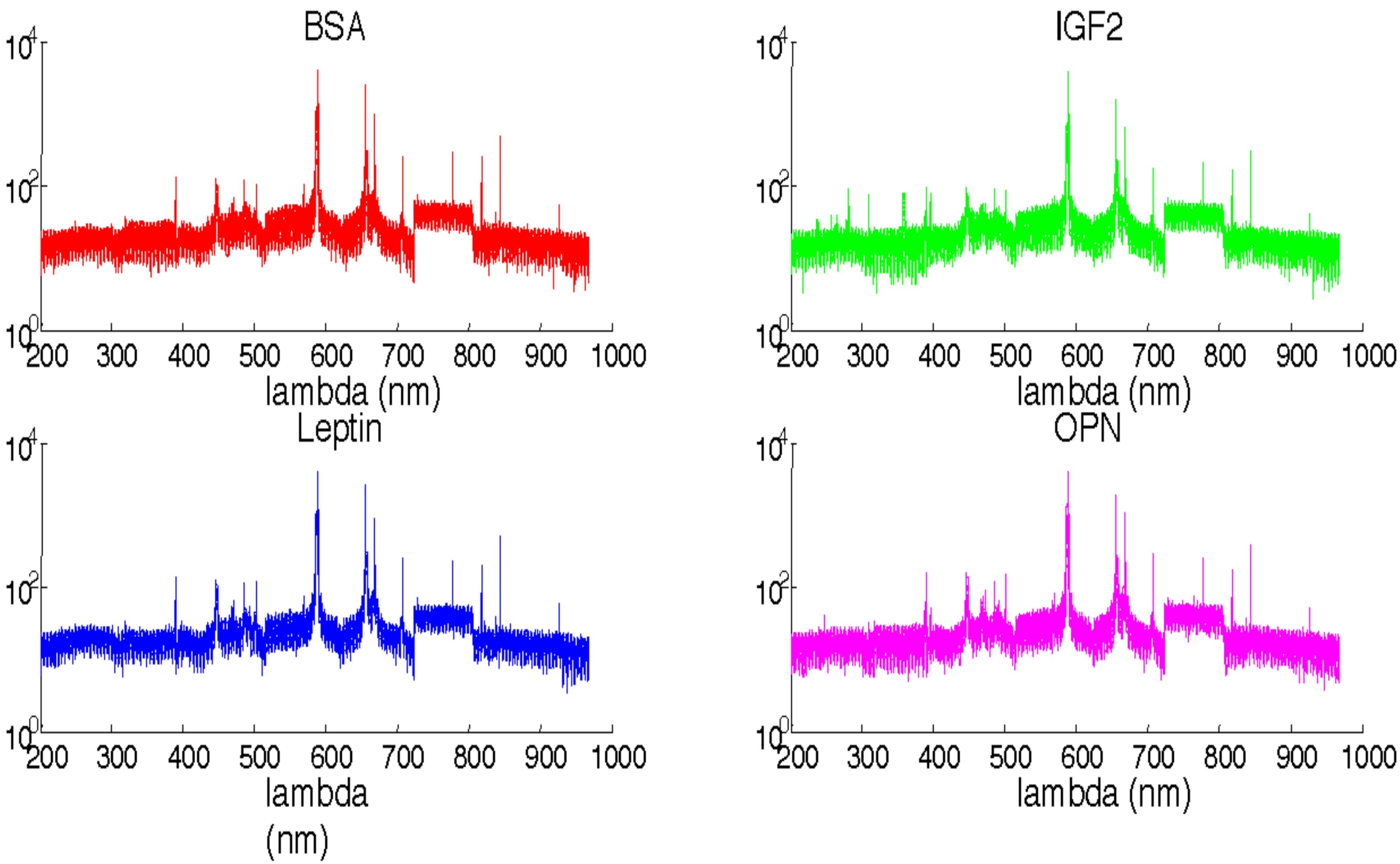
icETRAN,

Vrnjacka Banja, 6/5/2014

Motivation

- Laser spectroscopy produced vast amounts of data
- Need for automatization of classification and discrimination of spectra
- Classification techniques are ad-hoc and do not have theoretical justification
- No assurance of optimality from statistical theory of detection point of view

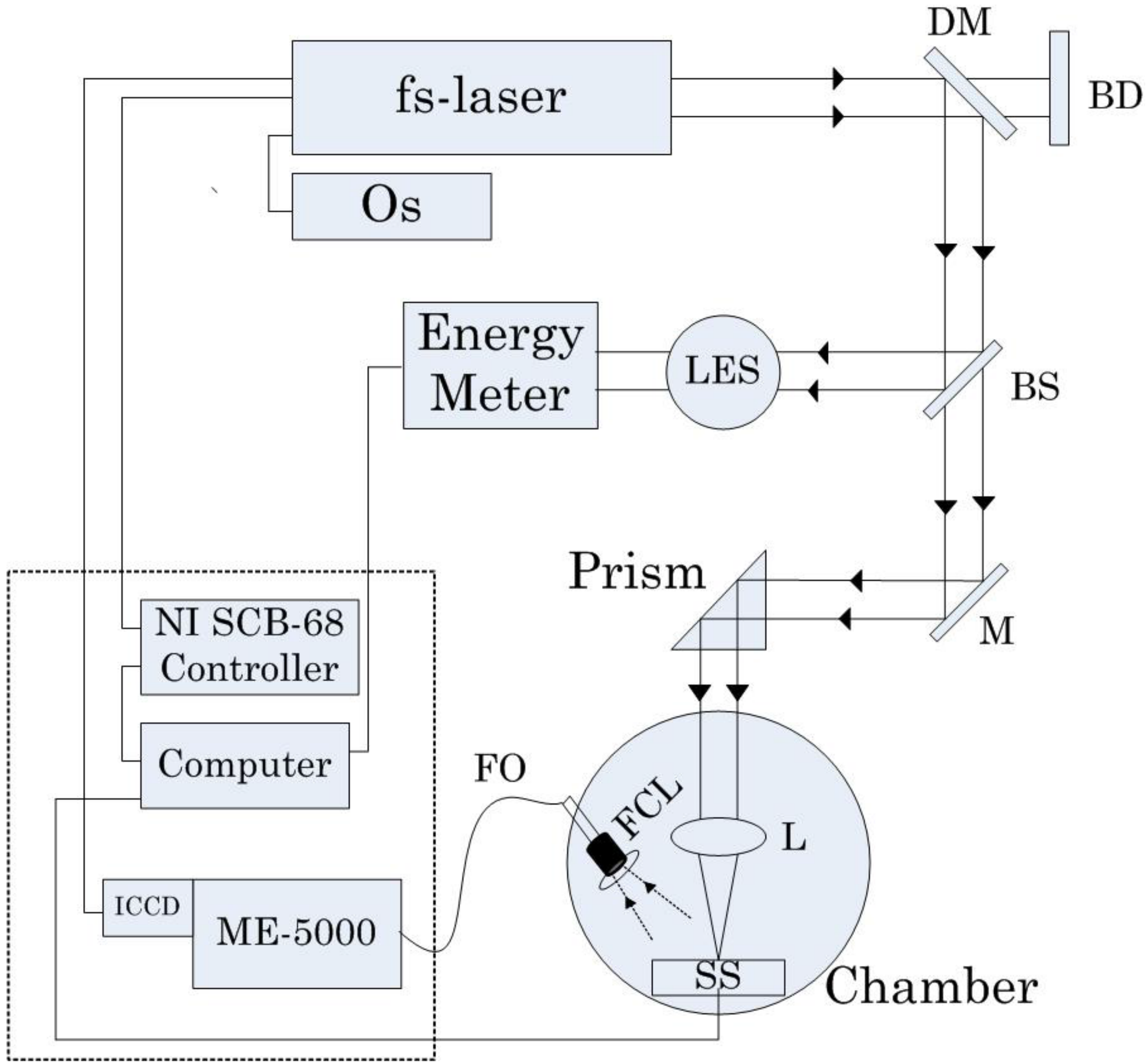
Examples of Spectroscopy Data



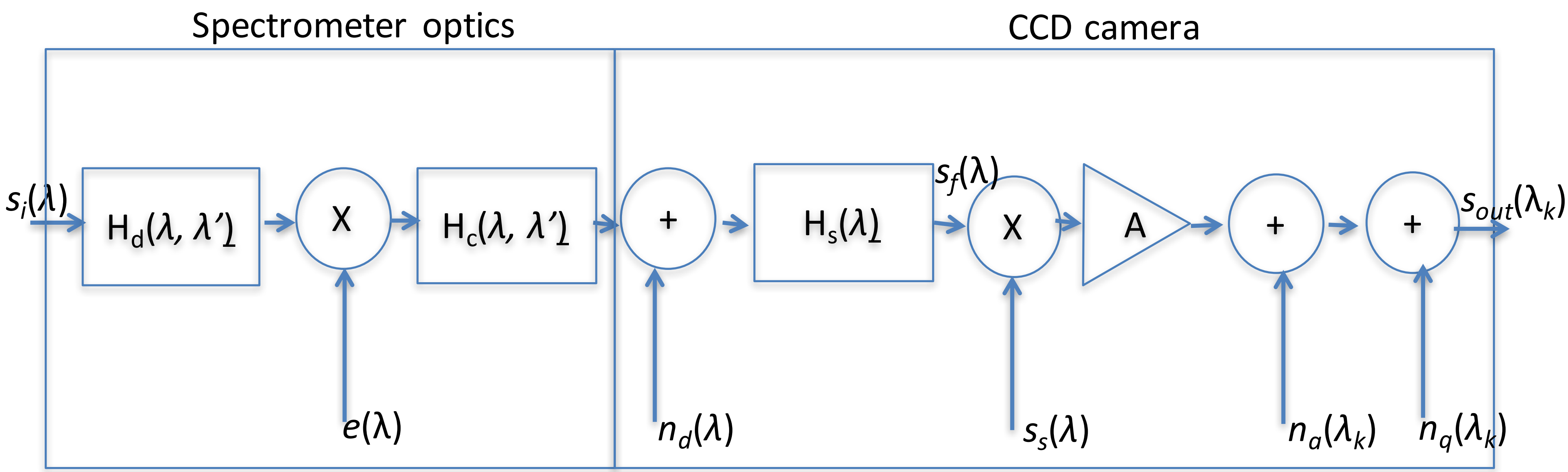
Goal

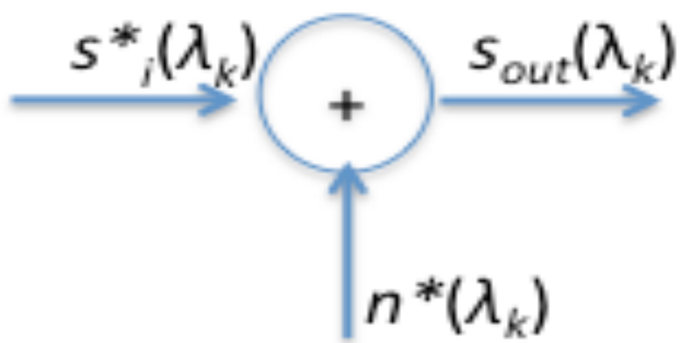
- Develop optimal classifier for spectroscopy data
- Consider echelle spectrograph with an Intensified Charge Coupled Device (ICCD) sensors
- Verify model assumptions using experimental data

Laser Induced Breakdown Spectroscopy (LIBS) System

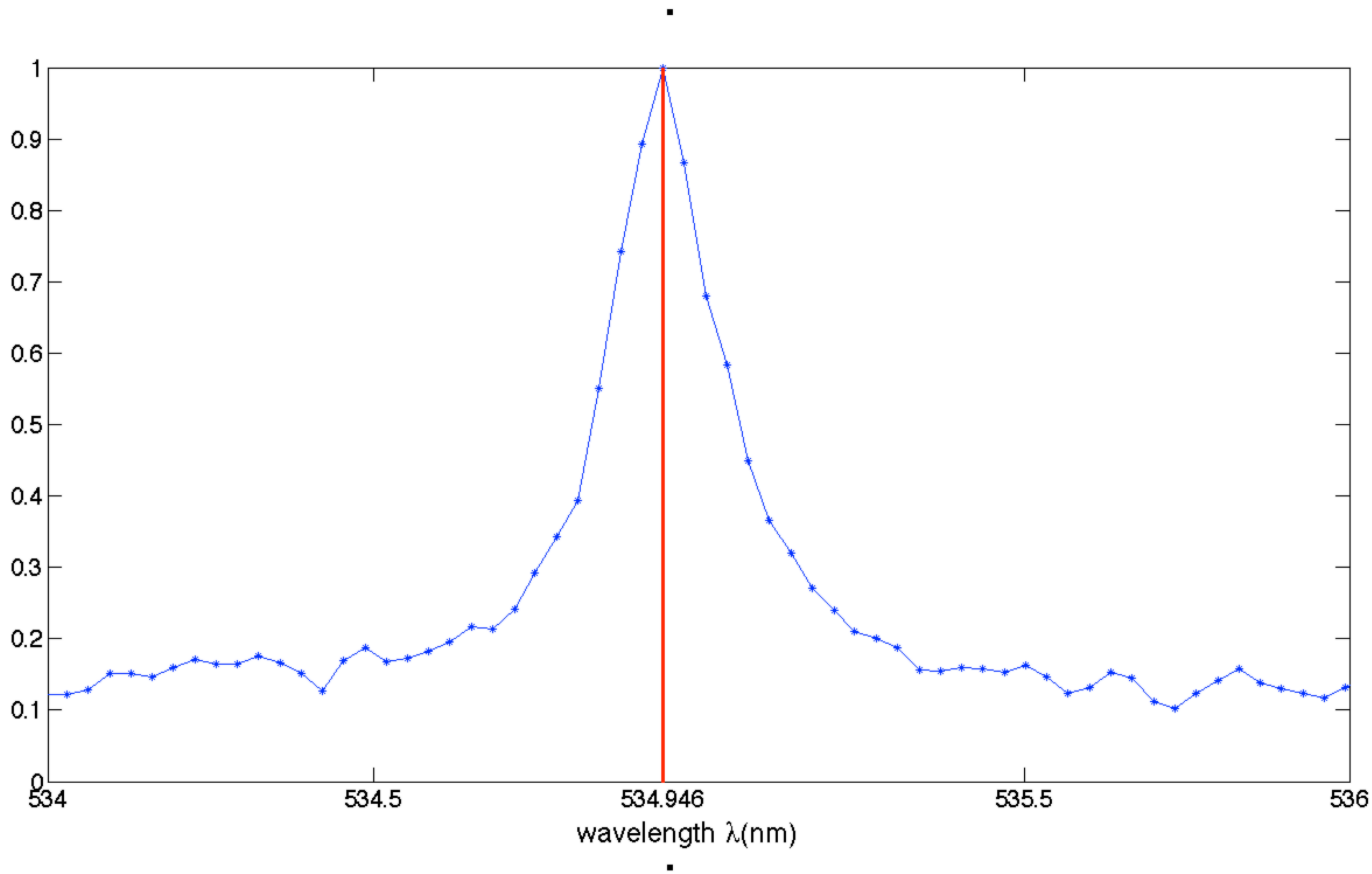


Block Diagram of System





Effects of Spectral Line Widening



Optimal Classifier of Spectroscopy Data

$$H_1: s_{out}(\lambda_k) = s^*_{i,1}(\lambda_k) + n^*(\lambda_k), k = 1, \ldots K$$
$$H_2: s_{out}(\lambda_k) = s^*_{i,2}(\lambda_k) + n^*(\lambda_k), k = 1, \ldots, K$$

- Detection of Gaussian signal in Gaussian noise!

$$H_1: \mathbf{s}_{out} = \mathbf{r}_1,$$
$$H_2: \mathbf{s}_{out} = \mathbf{r}_2,$$

$\mathbf{r}_1, \mathbf{r}_2$ are K-variate Gaussian vectors

$$p(\mathbf{s}_{out}|H_i) = \frac{1}{(2\pi)^{K/2}|\mathbf{\Sigma}_i|^{1/2}} e^{-\frac{1}{2}(\mathbf{s}_{out}-\mathbf{m}_i)^T \mathbf{\Sigma}_i^{-1}(\mathbf{s}_{out}-\mathbf{m}_i)}$$

- Likelihood ratio test

$$\Lambda(\mathbf{s}_{out}) \triangleq \frac{p(\mathbf{s}_{out}|H_1)}{p(\mathbf{s}_{out}|H_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta.$$

- **Log-likelihood test—quadratic decision boundary**

$$l(\mathbf{s}_{out}) = \mathbf{s}_{out}^T \mathbf{A}\, \mathbf{s}_{out} + \mathbf{b}^T \mathbf{s}_{out} \underset{H_1}{\overset{H_2}{\gtrless}} \gamma$$

$$\mathbf{A} \triangleq \frac{1}{2}(\mathbf{\Sigma}_1^{-1} - \mathbf{\Sigma}_2^{-1})$$
$$\mathbf{b} \triangleq \mathbf{\Sigma}_2^{-1}\mathbf{m}_2 - \mathbf{\Sigma}_1^{-1}\mathbf{m}_1$$
$$\gamma \triangleq ln\, \eta + \frac{1}{2}(\ln|\mathbf{\Sigma}_2| - \ln|\mathbf{\Sigma}_1| + \mathbf{m}_2^T\mathbf{\Sigma}_2^{-1}\mathbf{m}_2 - \mathbf{m}_1^T\mathbf{\Sigma}_1^{-1}\mathbf{m}_1).$$

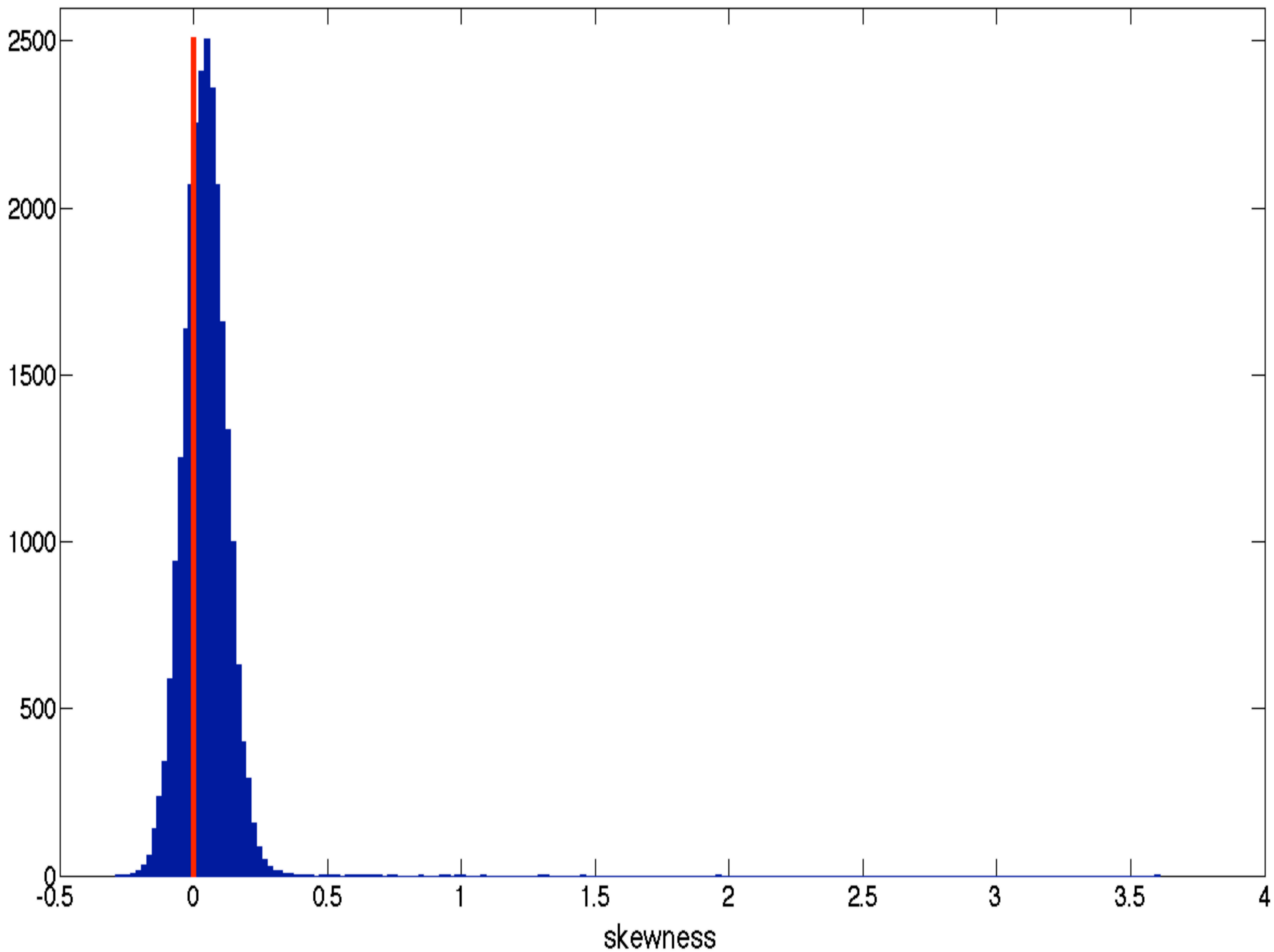
Experimental Results

- Andor Mechelle ME5000 spectrograph with an ICCD camera (iStar, Andor Technology, DH734-18F 03)
- The total number of channels: 26,040.
- Wavelength range: 199.04—974.83nm.
- The spectrometer used orders *m*=21-100.
- The grating with 52.13 line/mm; grating constant *d* 5-30μm, blazed at 32.35 degrees.
- Plasma excited with a broadband CPA-Series Ti-Sapphire ultra-short laser (Clark-MXR, Inc, Model: 2210) generating 150 fs lon pulses operating at 775nm
- Experiments performed with:
  - “Dark signal”
  - NIST standardized glass

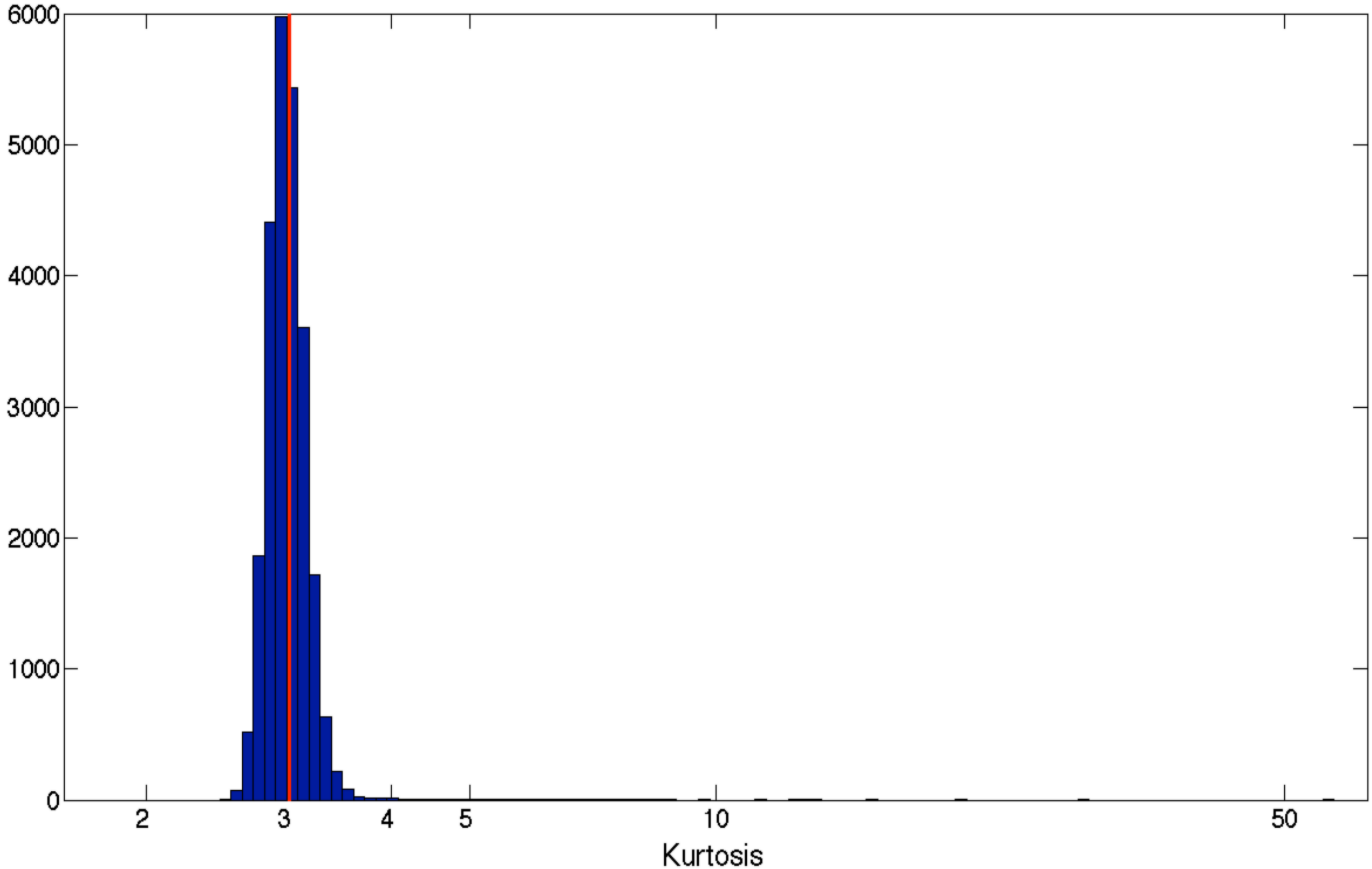
Hypotheses Tested

- H_{01}: $s_{\mathrm{out}}(\lambda_k)$ follows Gaussian distribution, $\lambda_k \in$[200.33nm, 909.45nm] (for dark signal and NIST glass)
  - Tested using Kolmogorov-Smirnov, Lilliefors tests and by inspection of skewness and kurtosis
- H_{02}:$s_{out}(\lambda_i)$, $s_{out}(\lambda_j)$ are uncorrelated when $\lambda_i \neq \lambda_j$ (for dark signal)
  - Tested by inspection of estimated normalized autocorrelation

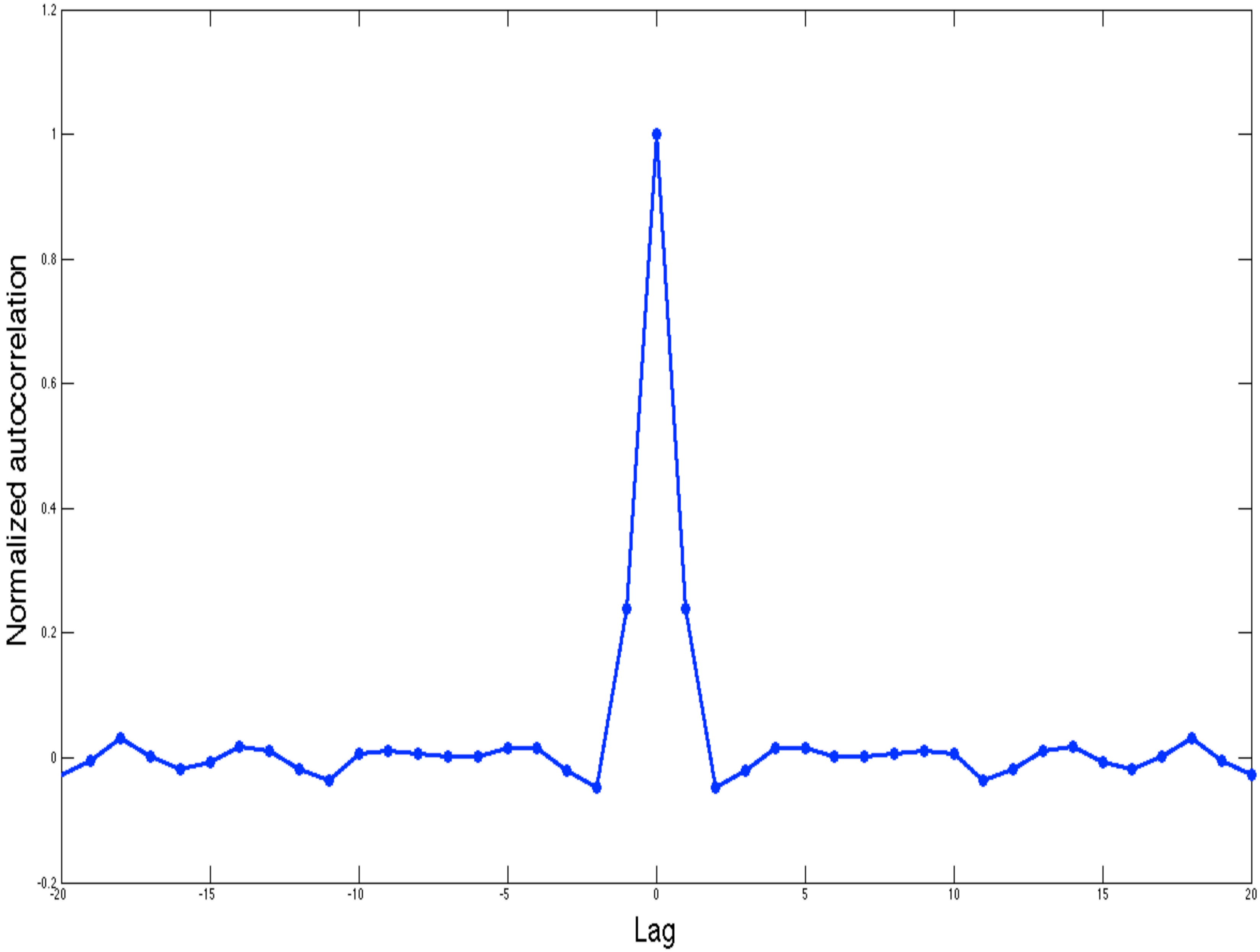
Histogram of Skewness of “Dark Signal”



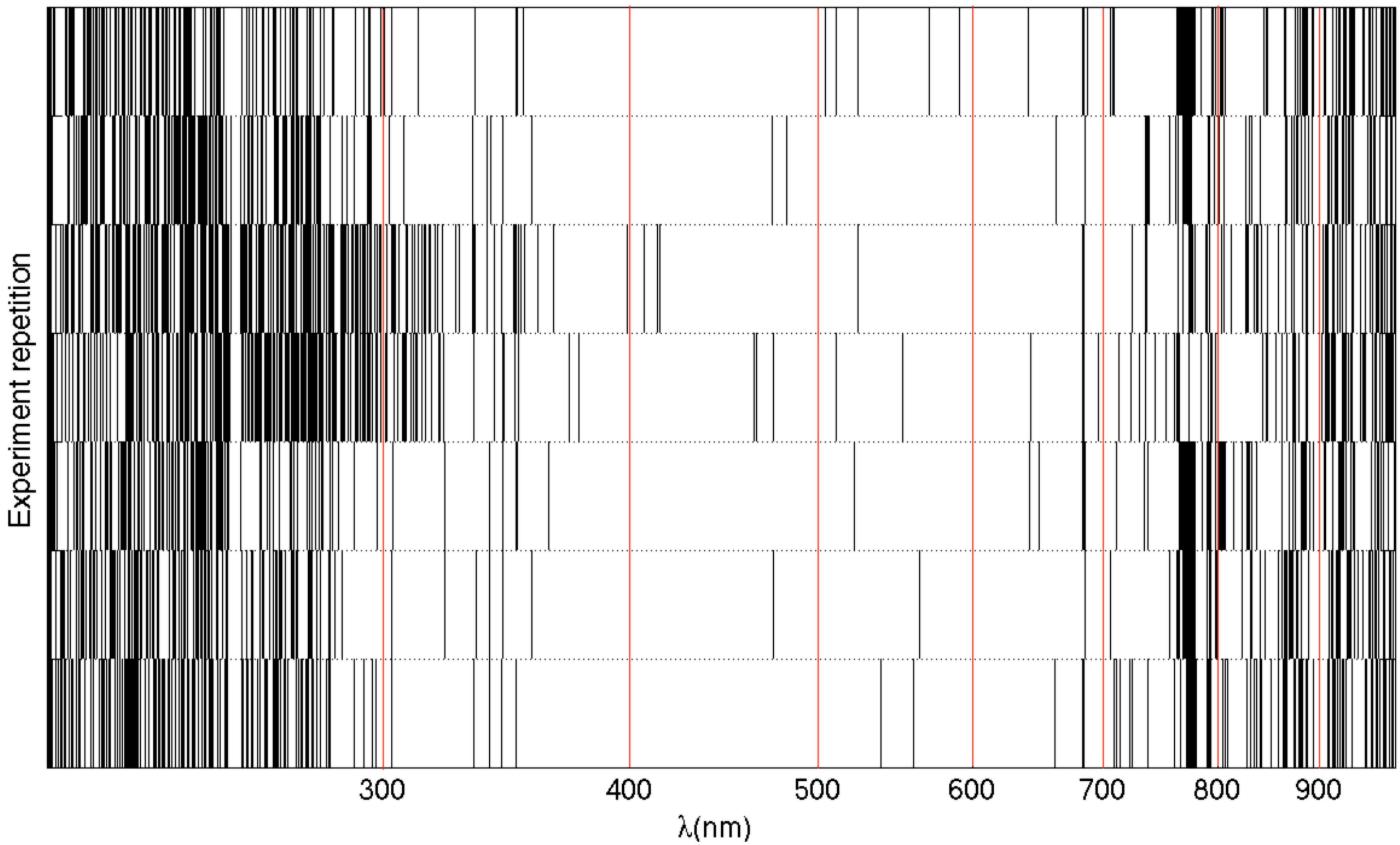
Histogram of Kurtosis of “Dark Signal”



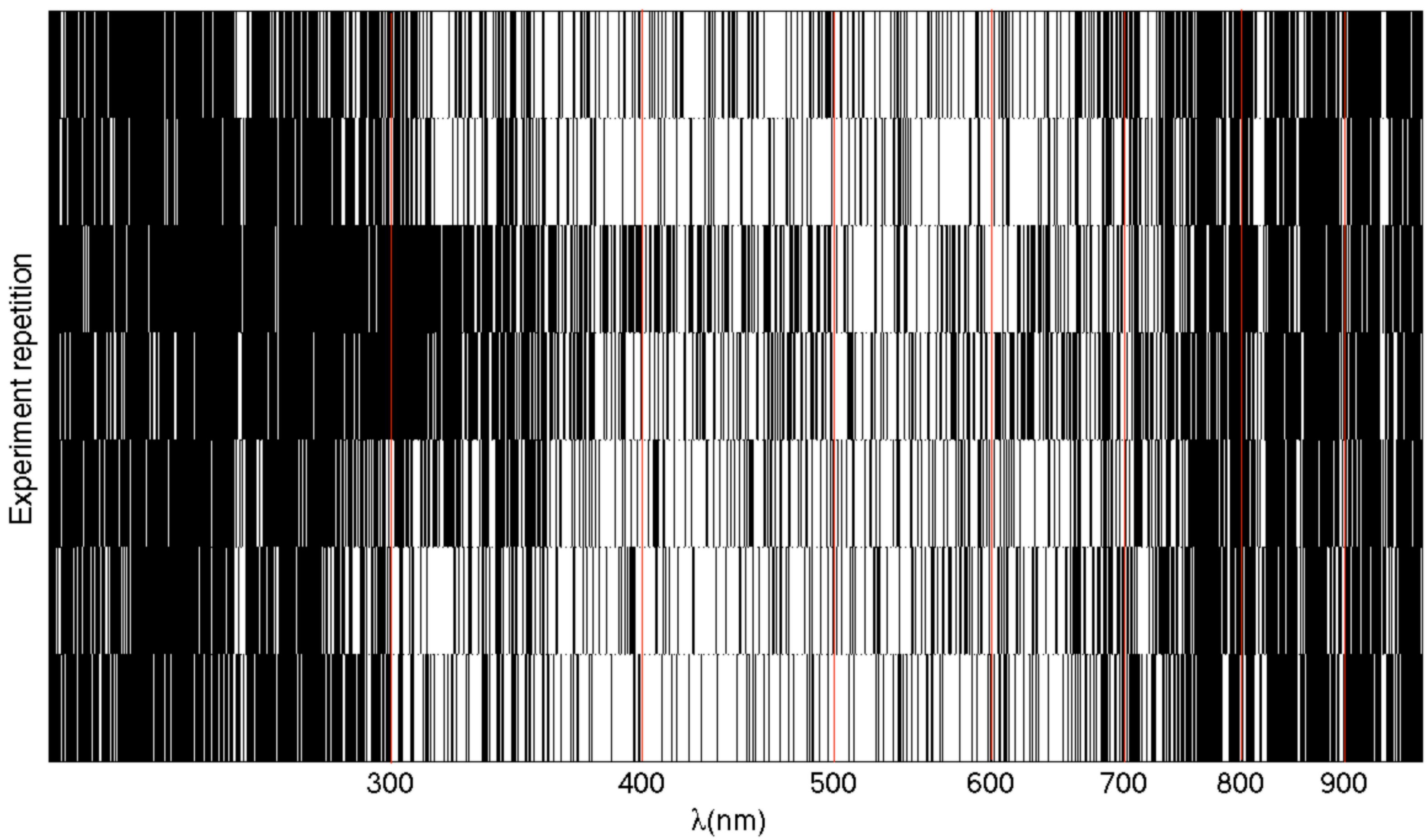
Autocorrelation of "Dark Signal"



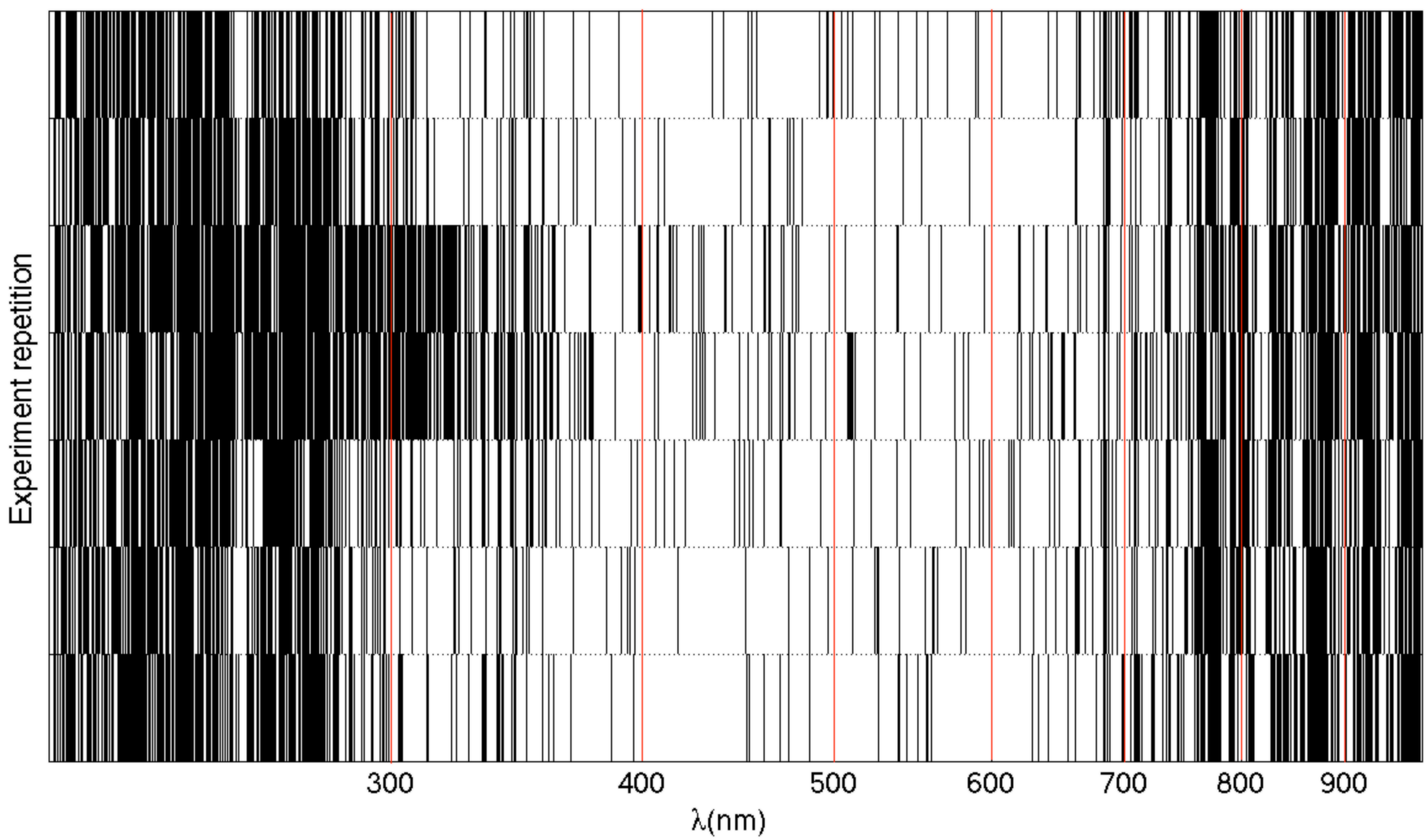
Kolmogorov-Smirnov Test for Gaussianity of NIST Glass Spectrum (α=0.05)



Lilliefors Test for Gaussianity of NIST Glass Spectrum (α=0.05)



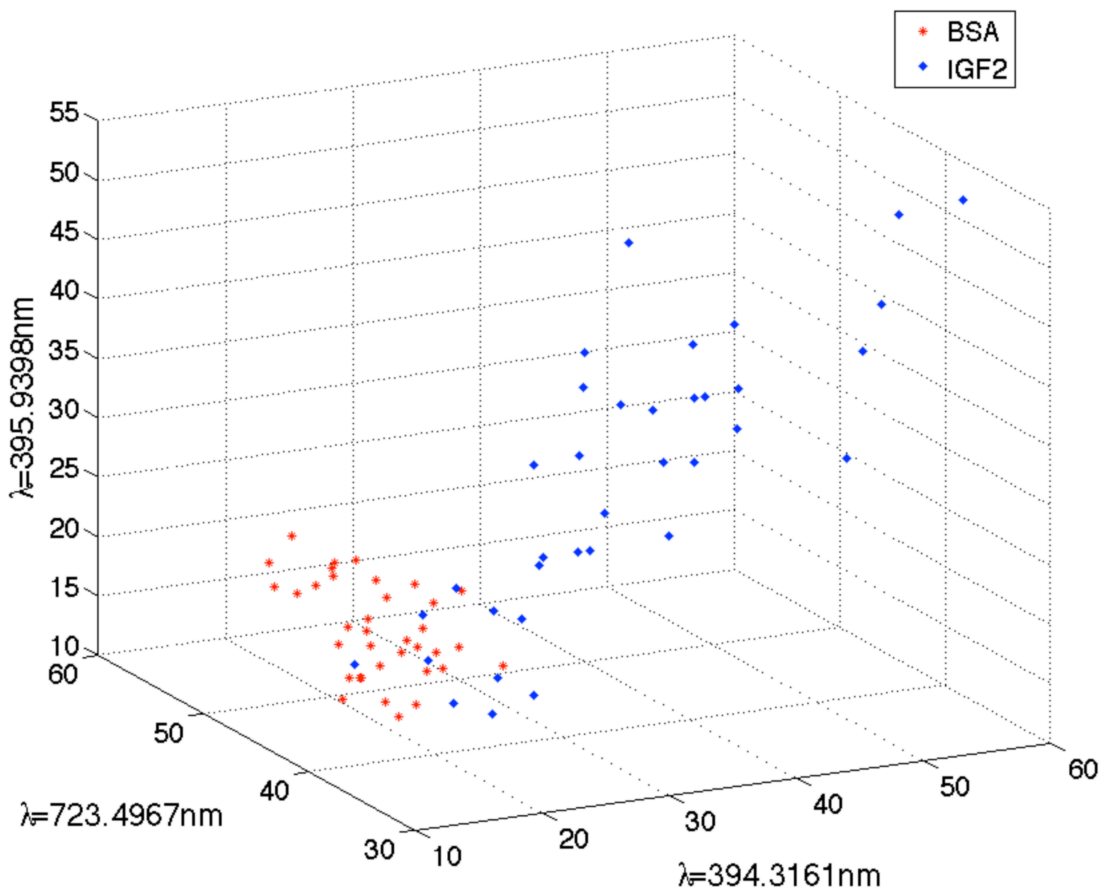
Lilliefors Test for Gaussianity of NIST Glass Spectrum (α=0.005)



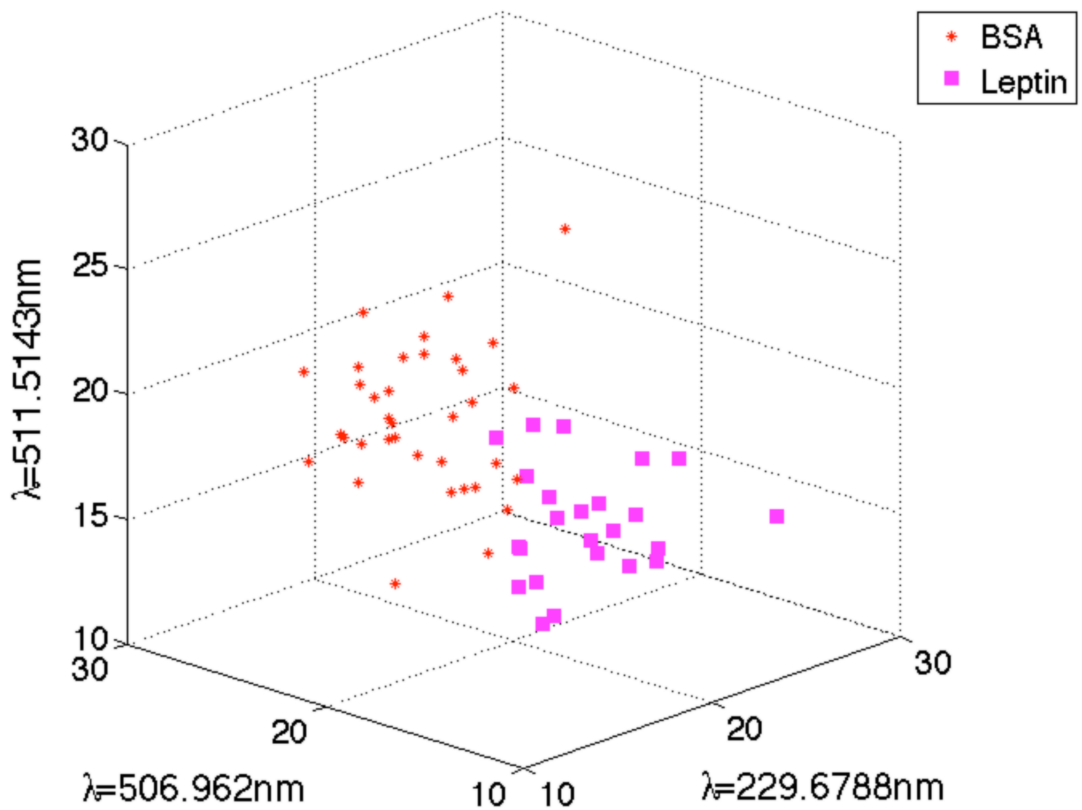
Discussion and Conclusions

- “Dark signal”:
  - Gaussian for almost all wavelengths
  - Observably correlated only with the samples at adjacent wavelengths
- NIST glass signals $s_{out}(\lambda_k)$ approximately Gaussian for a large range of $\lambda_k \in [400\text{nm}, 700\text{nm}]$
- Optimal classifier for spectroscopy data has quadratic decision boundary
- **Optimal classifier is applicable if:**
  - **The number of samples is sufficiently large**
  - **Feature selection to determine discriminatory wavelengths is applied**

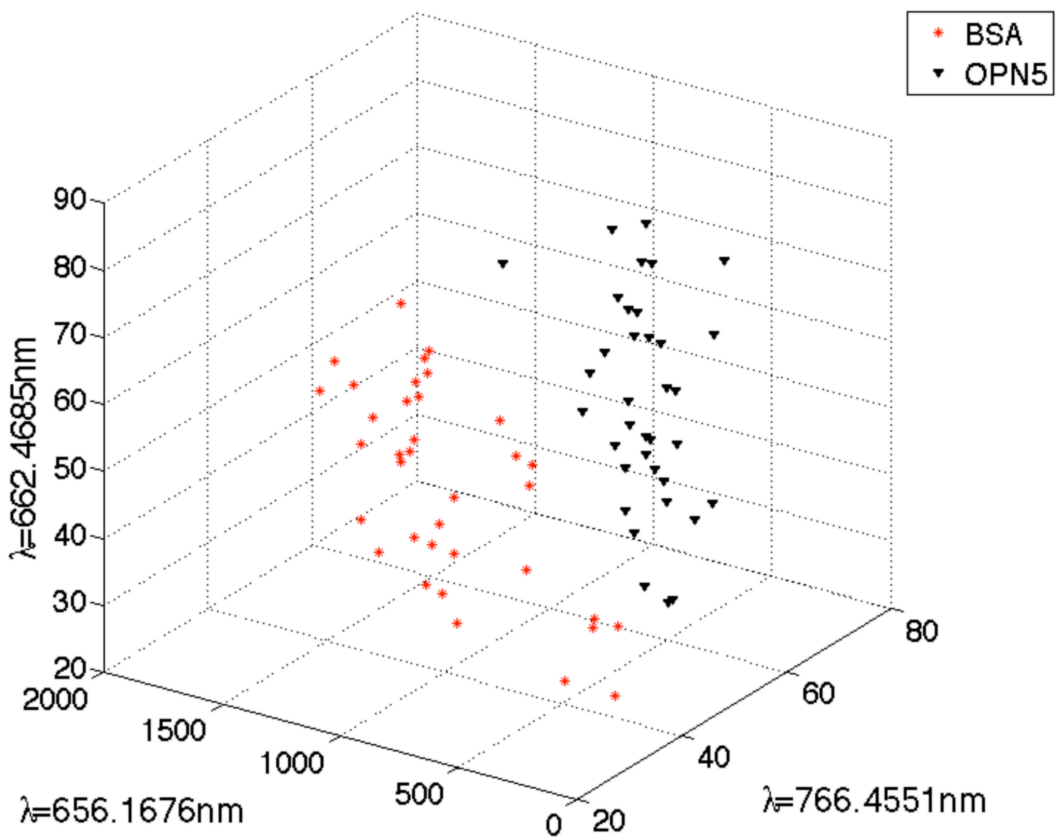
Examples: Discrimination of Protein Classes using Selected Features



Examples: Discrimination of Protein Classes using Selected Features



Examples: Discrimination of Protein Classes using Selected Features



Acknowledgements

- US Department of Defense Breast Cancer Research Program (HBCU Partnership Training Award #BC083639)
- US National Science Foundation (CREST grant #HRD-1242067), CREST-8763
- NASA URC 7658
- US Department of Defense/Department of Army (45395-MA-ISP, #54412-CI-ISP, W911NF-11-2-0046)