2014-08-24

The Detection of Analytes Using Spectroscopy: A Dempster-Shafer Approach

Nicholas J. Napoli
University of Miami, nick1nap@gmail.com

UNIVERSITY OF MIAMI

THE DETECTION OF ANALYTES USING SPECTROSCOPY: A DEMPSTER-SHAFER APPROACH

By

Nicholas J. Napoli

A THESIS

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Master of Science

Coral Gables, Florida

August 2014

UNIVERSITY OF MIAMI

A thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science

THE DETECTION OF ANALYTES USING SPECTROSCOPY: A DEMPSTER-SHAFER APPROACH

Nicholas J. Napoli

Approved:

_____
Mohamed Abdel-Mottaleb, Ph.D
Professor of Electrical and
Computer Engineering School

_____
Manohar N. Murthi, Ph.D.
Associate Professor of
Electrical and Computer
Engineering

_____
Kamal Premaratne, Ph.D.
Professor of Electrical and
Computer Engineering

_____
M. Brian Blake, Ph.D.
Dean of the Graduate School

_____
James N. Wilson, Ph.D.
Assistant Professor of Chemistry

NAPOLI, NICHOLAS J. (M.S., Electrical and Computer Engineering)
The Detection of Analytes Using Spectroscopy: (August 2014)
A Dempster-Shafer Approach
Abstract of a thesis at the University of Miami.

Thesis supervised by Professor Kamal Premaratne.
No. of pages in text. (109)

The Environmental Protection Agency (EPA) has set forth guidelines for drinking water to ensure the safety of the public from harmful contaminants and pollutants. Current standardized methods for the detection of water borne toxins and pollutants are expensive, and vague in their analysis: qualitative and quantitatively. We introduce a data fusion model using Dempster-Shafer Theory to qualitatively detect multiple combinations of analytes suspended in Toluene. The benefits of data fusion model are its ability to be extended for additional sources of evidence such as pH, turbidity, and electrical conductivity and its ability to handle epistemic uncertainty. In addition, we develop a method of modeling spectroscopy data and an ability to synthetically add spectroscopy noise and perturbations to the signal. This novel chemometric detection method that is introduced has reported 99% detection under the most extreme noise condition of η=2.0, using cepstral coefficients as an evidence source when fused over all the simulated spectra data. This was an increase in the averaged recognition using correlation coefficients by 46.3 percent.

This is dedicated to people who have entered into my life that have shaped me into who I am today! My mother who has always been there for me and has provided me the tools to accomplish my goals and dreams. My grandfather, Joseph Fiero, who has made into the man I am today. My brother who has guided me and has watched over me. I love you. Additionally, Alex Vallejo, Alex Castro, and Avtar Singh three influential people when I was an undergraduate that taught me how to believe in myself.

"If you hear a voice within you say,
'You cannot paint,'
then by all means paint,
and that voice will be silenced"
-Vincent Van Gogh

Acknowledgements

I would like to especially thank my immediate adviser, Dr. Kamal Premaratne for this time and assistance with Dempster-Shafer Theory and in the preparation of this thesis.

To Dr. Manohar Murthi, for his guidance in spectral modeling techniques to be used as evidence sources.

To Dr. James Wilson, for his extensive time and enthusiasm working with me on the spectroscopy modeling. Without your open door, none of this would have been possible.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Purpose

The technology of UV/Vis spectroscopy allows the modern user to accurately measure light absorption and emission within the ultraviolet-visible region of the electromagnetic spectrum. UV/Vis spectroscopy is routinely employed to determine the presence, and/or concentration of a wide array of analytes in a sample. This method is preferred over other methods because it allows for both qualitative and quantitative analysis. UV/Vis spectroscopy has high levels of accuracy, sensitivity, reproducibility, and is cost effective. These advantages of using UV/Vis spectroscopy makes it particularly well-suited for water safety testing.

In the United States, the Environmental Protection Agency (EPA) is responsible for ensuring water safety by establishing guidelines that mandate methods and schedules for testing. The EPA's National Primary Drinking Water Regulations identifies microorganisms, disinfectants and their byproducts, inorganic and organic chemicals, and radionuclides as regulated contaminants. Other recommended, but non-enforceable guidelines, include acceptable pH levels, turbidity, and odor.

Innovative methods are being developed for an efficient and integrated system for the detection of water contaminants destined for consumption. These detection

methods extract key variables (components or factors) and then apply clustering schemes for classification. Parallel Factor Analysis (PARAFAC), Principle Component Analysis (PCA), and various other clustering schemes are currently being explored. Although these methods have been shown to be effective, there are some disadvantages.

For example, PARAFAC may be ineffective when confronted with missing values [3] [4]. PARAFAC's effectiveness is dependent on linear proportionality factors [5], and outliers can reduce PARAFAC's effectiveness [4]. In the case of PCA, the rotational scheme prevents direct measurement [6], and in non-ideal cases, noise can be potentially modeled [6] [7].

In addition to these disadvantages, the aforementioned methods do not account for *epistemic uncertainty*. Epistemic uncertainty is the lack of knowledge about the system, which can be reduced by increasing the amount of relevant data or evidence [8]. The implementation of uncertainty to represent a process when using mathematical models is an essential part of the numerical representation. The aforementioned methods only account for *aleatory uncertainty*, which is caused by random variability and is not reducible [8]. This can be modeled from historic data sets via probability distribution models [9] [8]. Dempster-Shafer (DS) Theory is a method that accounts for epistemic uncertainty, which combines evidence to reduce uncertainty while achieving a more accurate decision.

Our initial intention was to develop a detection method for water contaminants destined for consumption. However, the scope of the project was narrowed as a direct result of complications that arose during the acquisition of water fluorescent data. Therefore, the study proceeded with the purpose of achieving two aims. First, we aimed to design a functional synthetic spectroscopy database to resemble a realistic spectroscopy experiment. The second aim was to investigate the effectiveness of DS Theory as a novel qualitative method for detection of analytes using spectroscopy. Despite the specific focus being amended, applications for this work remain wide-ranging and have significant merit in the field of chemometrics.

1.2 Background

The physical mechanisms governing spectroscopy are a product of the interactions between electromagnetic energy and matter. This section discusses the basics of the physics and chemistry that are related to spectroscopy. This begins the foundation for our discussion on how we develop the spectroscopy database in Chapter 2.

1.2.1 Electromagnetic Radiation

Electromagnetic Radiation is a type of energy that is emitted or absorbed by particles. Light is a form of electrical magnetic radiation [10]. We classify the type of electromagnetic radiation (radiowaves, microwaves, infrared radiation, visible light, ultraviolet radiation, etc.) by either wavelength or frequency, which are specific ranges in the electromagnetic spectrum. UV/Vis spectroscopy's spectrum ranges from approximately $185 - 700\,nm$, shown in Figure 1.1. The wavelength is related to the frequency and speed of light by

$$\nu = \frac{c}{\lambda}, \tag{1.1}$$

where ν is the frequency, c is the speed of light $2.998 \cdot 10^8\,m/s$, and λ is the wavelength.

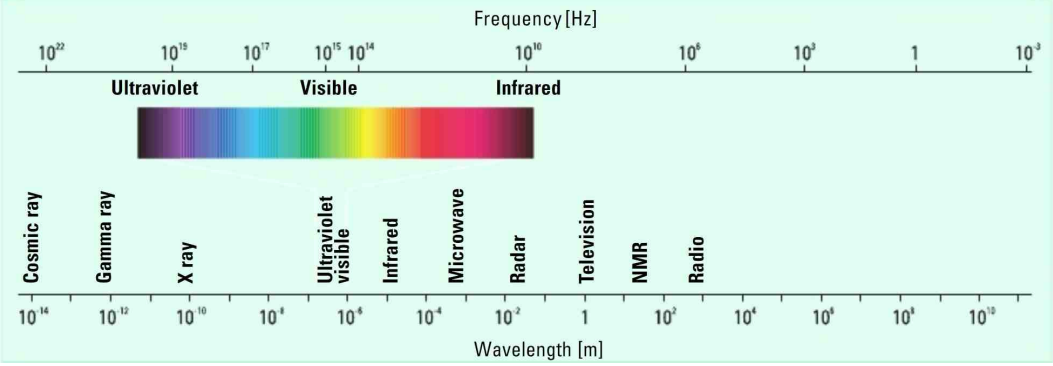


Figure 1.1: Electromagnetic Spectrum.

The Planck-Einstein equation 1.2 defines the relationship between the energy of a photon, E, and the corresponding frequency of radiation, where h is Planck's constant

$6.626 \cdot 10^{-34}\, m^2\, kg/s$:

$$E = h \cdot \nu = \frac{ch}{\lambda} \tag{1.2}$$

Equations 1.1 and 1.2 describe the proportional relationship between photon energy and frequency, ν, and the inverse proportionality between photon energy, E, and wavelength, λ.

Energy and Matter

Specific wavelengths of light interact with matter differently. This interaction alters the wavelength and intensity properties of the incident beam of light. The product of this interaction is then used to qualitatively and quantitatively analyze matter by the amount of energy the specimen absorbs or emits. This absorption (blue) or emission (red) of energy is represented as an intensity that changes as a function of wavelength, shown in Figure 1.2. In Figure 1.2, the absorption (blue) and emission (red) intensity

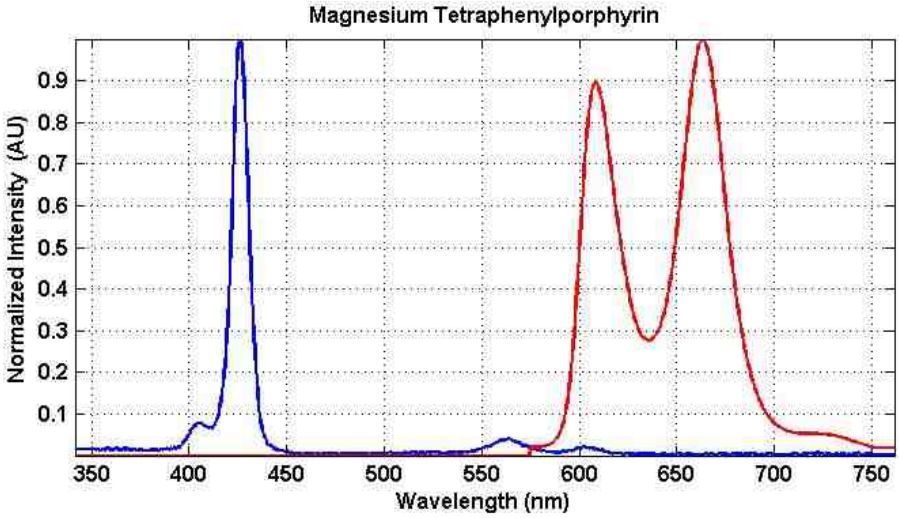


Figure 1.2: Absorption and Emission Spectra.

is normalized to graphically represent both absorption and emission simultaneously.

1.2.2 Absorption Spectroscopy

The designs of absorption spectroscopy instruments vary by light source, collimators, and photomultiplier tubes (PMTs). The general conceptual design of absorption spectroscopy is demonstrated in Figure 1.3. An incident beam of light, I_0, is projected through the sample to detect the transmitted light intensity, I, by a PMT. Beer-

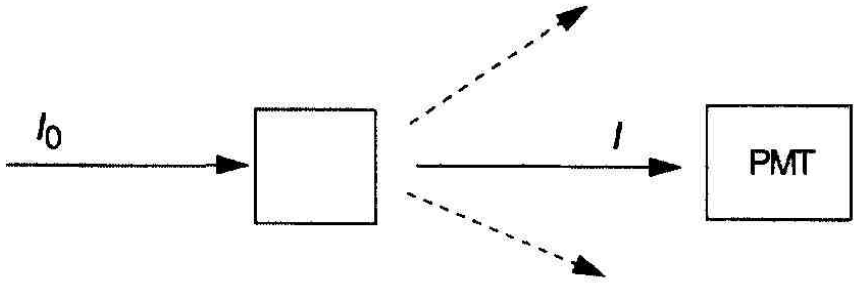


Figure 1.3: Overview of Spectral Absorbance.

Lambert's Law explains how the incident beam changes as it travels through the medium and its relationship to absorption. There are two basic rules, which combined constitute Beer-Lambert's Law [10]. First, *"Lambert's Law states that the fraction of light absorbed by a transparent medium is independent of the incident light intensity, and each successive layer of the medium absorbs an equal fraction of the light passing through it"*. These successive layers of medium that absorb the light lead to an exponential decay of the light intensity given by

$$\log_{10}\left(\frac{I_o}{I}\right) = k\ell, \tag{1.3}$$

where ℓ is the *pathlength* (the distance light travels through the medium), and k is the *medium constant* (a unique constant that is dependent on the chemical composition of the medium). Second, *"Beer's Law claims that the amount of light absorbed is proportional to the number of molecules of the chromophore through which the light passes [10]"*. A *chromophore* is defined as a molecule that absorbs light at certain wavelengths. Therefore, the medium constant, k, is a function of the concentration

of chromophores. This is defined by

$$k = \varepsilon C, \tag{1.4}$$

where C is the concentration of the chromophore, ε is the *molar extinction coefficient*. Therefore, equation 1.3 can be rewritten as

$$A = \log_{10}\left(\frac{I_o}{I}\right) = \varepsilon C \ell, \tag{1.5}$$

where A is the *absorbance*. Although the molar extinction coefficient changes as a function of the wavelength, it is typically denoted in the general form as ε. An alternative form explicitly denotes the molar extinction coefficient and absorbance with superscripts (A^λ and ε^λ) to indicate their dependence on the wavelength.

1.2.3 Emission Spectroscopy

Emission spectroscopy is widely used in analytical measurements and scientific investigations. The emission spectrum provides an abundance of information about the chemical composition by the way the molecule absorbs and emits energy. In order for a molecule to emit electromagnetic radiation, it first must absorb the electromagnetic radiation energy. This process of absorption can only occur when the difference in energy between the *ground state*, E_2, and *excited states*, E_1, of an atom is equivalent to the energy of the electromagnetic radiation applied to the molecule. Therefore, the difference in energy between the two states corresponds to the *absorbed photon energy*, shown by

$$h\nu = E_2 - E_1. \tag{1.6}$$

When the sample absorbs electromagnetic radiation, the electrons in its orbitals are excited and then relaxed causing an emission of light. This excitation causes

the electrons to jump from the lowest unoccupied molecular orbital (LUMO) to its highest occupied molecular orbital (HOMO). This emitted energy is detected by the spectrometer and recorded in the form of an emission spectrum. A general diagram of this process is demonstrated in Figure 1.4, where $h\nu_1$ is the incident beam of a specific excitation wavelength, and $h\nu_2$ is the emitted energy at a new wavelength. The wavelengths, at which these absorptions occur are unique to the types of atoms or molecules present within the sample. This provides a qualitative and quantitative analysis of the sample. The emission spectrum that is generated is a spectral signature unique to the sample.

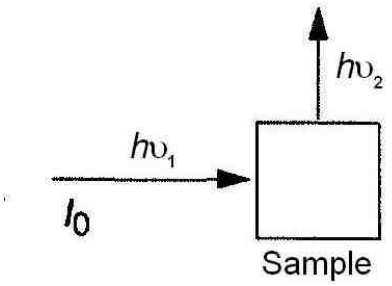


Figure 1.4: Overview of Excitation/Emission.

Chapter 2

Generating a Synthetic Spectroscopy Database

The unavailability of open source spectroscopy data and the difficulty in direct acquisition of spectroscopy data via experiments are significant hurdles in the development of effective detection algorithms associated with such data in chemometrics. To address this issue, we sought to develop a method to generate a synthetic spectroscopy data bank which attempts to faithfully capture the operationally relevant features of fluorescent spectra. We wanted this method to possess the ability to account for the variations arising from different excitation wavelengths, multiple combinations of chemical mixtures, different apparatus and effects of user error.

2.1 Data from PhotoChemCAD

To generate such a synthetic database, we started with a sparse amount of reputable data from PhotoChemCAD [11, 12]. To achieve our objective with this limited set of experimental data, we made several assumptions:

- There is no chemical quenching.
- The spectral resolution is accurately depicted through interpolation and decimation.
- All path lengths are $1\,cm$.

- Accurate depiction of the molar absorptivity is projected by scaling the absorption spectrum by cited molar extinction coefficients.
- Beer-Lambert's law is obeyed.
- Concentrations are below $1\mu M$ to minimized inner filter effects to achieve additive absorption and emission spectra.

From PhotoChemCAD, we chose multiple chemicals within the same solvent, Toluene. The analytes selected were from two chemical classes: oligopyrrole and polycyclic aromatic hydrocarbons. These two criteria were selected to obtain data that are similar in chemical composition. The resulting strong spectral similarities create a "non-trivial" data set for classification and quantification purposes. The chemicals that were used in the database are listed in Table 2.1 and 2.2.

The data from PhotoChemCAD provide neither the concentration of the analytes nor the emission characteristics at different excitation wavelengths. In our data set, we acquire these two quantities directly from published results for spectral modeling see Table 2.1 and 2.2. Epsilon, ε, enables us to account for different concentrations of the chemical; and the quantum yield, Φ_F, enables us to calculate different emissions spectra at various excitation wavelengths. With these quantities accounted for, we are able to generate a more realistic data set.

2.1.1 Preprocessing of Spectral Data

The spectral data from PhotoChemCAD is an agglomeration of various sources, where acquisition parameters are distinctive from each other. The spectral data from the selected chemicals are different in their wavelength range and optical resolution. To ensure proper manipulation among different chemical spectral vectors, the length and indexing of the vectors are required to be equivalent. The optical sampling of the spectral vectors obtained were either $0.25\,nm$, $0.5\,nm$, or $1.0\,nm$; hence, the vectors were interpolated or decimated to get values corresponding to $0.5\,nm$. Each vector was padded with elements of value $1.0 \cdot 10^{-20}$ to provide a uniform wavelength range

Chemical	**Solvent**	**Epsilon (ε) in cm^{-1}/M at λ_{Ex}**	**Quantum Yield (Φ_F)**	**Cited**
5,10-Diaryl Chlorin	Toluene	$89,100$ at $414\,nm$	0.260	[13]
5,10-Diaryl Mg-oxoChlorin	Toluene	$191,000$ at $408\,nm$	0.100	[13]
5,10-Diaryl oxoChlorin	Toluene	$174,000$ at $414\,nm$	0.130	[13]
5,10-Diaryl Zn-Chlorin	Toluene	$186,000$ at $412\,nm$	0.083	[14, 13]
5,10-Diaryl Zn-oxoChlorin	Toluene	$209,000$ at $408\,nm$	0.040	[13]
Bis(5-mesityldiprinato)zinc	Toluene	$115,000$ at $487\,nm$	0.360	[15]
Bis(5-phenyldiprinato)zinc	Toluene	$115,000$ at $485\,nm$	0.006	[15]
Magnesium Octaethylporphyrin	Toluene	$408,300$ at $410\,nm$	0.150	[16, 17]
Magnesium Tetramesityporphyrin	Toluene	$446,700$ at $426.5\,nm$	0.170	[18, 17]
Magnesium Tetraphenylporphyrin	Toluene	$22,000$ at $564\,nm$	0.150	[19, 20]

Table 2.1: Chemicals in the Database: Oligopyrrole.

throughout the entire database of selected chemicals. The value $1.0 \cdot 10^{-20}$ is used to avoid absolute zero errors and to circumvent subsequent complications with vector and matrix manipulations.

2.1.2 Generating Spectra Corresponding to Various Concentrations

A spectrometer takes measurements of light absorption, producing a unique spectral absorption signature. When taking measurements, the concentration and pathlength are held constant. The epsilon value, which is a function of the excitation wavelength, is an intrinsic property of the measured chemical that defines the spectral waveform

Chemical	**Solvent**	**Epsilon (ε) in cm^{-1}/M at λ_{Ex}**	**Quantum Yield (Φ_F)**	**Cited**
Perylene-diimide	Toluene	$44,000$ at $490\,nm$	0.97	[21]
Perylene-Monoimide	Toluene	$32,000$ at $511\,nm$	0.86	[22]
Perylene-Monoimide(OR)3	Toluene	$32,000$ at $479\,nm$	0.91	[22]
Perylene-Monoimide (OR)	Toluene	$40,000$ at $507\,nm$	0.82	[22]

Table 2.2: Database Chemicals: Polycyclic Aromatic Hydrocarbons.

characteristics. The selected data only provides the absorbance, giving no insight into the concentration or pathlength. We are therefore unable to distinguish the epsilon values due to the unknown collection parameters from the various sources provided by PhotoChemCAD. The measured absorbance of the sample is proportional to the number of absorbing molecules from the incident light of the spectrometer and it is essential that the absorbance value is corrected for a meaningful comparison [23]. This correction for absorption is referred to as *molar absorptivity* or *molar extinction coefficients*, which serves to compare spectra and evaluate the relative strength of the absorbance. In order perform a proper comparison between spectra, we scale the spectral vector with respect to epsilon at its appropriate listed excitation wavelength from Table 2.1 and 2.2.

Consider a measured absorption data vector from Table 2.1:

$$\mathbf{D}_{Ab_j} = \begin{bmatrix} d_{Ab_{j1}} & d_{Ab_{j2}} & \cdots & d_{Ab_{jN}} \end{bmatrix}^T, \tag{2.1}$$

where $d_{Ab_{jk}} \in [0, A], \forall k \in \overline{1, N}$, A is an arbitrary positive number, and j is a specific chemical. These optical absorption measurements were scaled to coincide with cited molar extinction coefficients (i.e., epsilon) at the corresponding wavelength from

Table 2.1 and 2.2 via the following equation:

$$\mathbf{S}_{Ab_j} = \mathbf{D}_{Ab_j} \frac{\varepsilon_{\lambda_{Ex}j}}{D_{Ab_j\lambda_{Ex}}} C, \tag{2.2}$$

where $\varepsilon_{\lambda_{Ex}j}$ is the molar extinction coefficient of chemical j from Table 2.1 and 2.2, C is the concentration, and $D_{Ab_j\lambda_{Ex}}$ is the element in $\mathbf{D}_{Ab_j}$ that is associated with epsilon at the specific excitation wavelength of λ_{Ex}. With the spectral vector properly scaled, we assume that each spectral element depicts its appropriate molar extinction coefficient for all wavelengths. We can now apply Beer-Lambert's law to expand the database by altering the concentration while we hold constant the pathlength at 1 cm.

2.1.3 Expanding the Data Set Via Quantum Yield

When a molecule is excited to a higher quantum state of a particle and it transitions to a lower state, the molecule emits a photon. The more the molecule absorbs energy the higher potential for it to elicit more photons. The amount of fluorescence emission is a function of the amount of light absorbed by a molecule. This function is known as the *quantum yield of the fluorescence,* Φ_F. It is defined as the number of photons emitted over the total number of photons absorbed [24]. Based on the intrinsic nature of the molecule and its absorption properties, specific wavelengths are more prone to be absorbed than others. It is apparent that the excitation wavelength affects the total intensity of the absorption and concurrently affects the total emission intensity. By dynamically changing the excitation wavelength, we generate various emission spectra accounting for the effect of quantum yield.

Consider a measured fluorescence emission spectrum vector from Table 2.1 and 2.2:

$$\mathbf{D}_{Em_j} = \begin{bmatrix} d_{em_{j1}} & d_{em_{j2}} & \cdots & d_{em_{jN}} \end{bmatrix}^T, \tag{2.3}$$

where $D_{em_{jk}} \in [0, A], \forall k \in \overline{1, N}$, A is an arbitrary positive number, and j is a specific chemical. In order to generate further data and for a realistic simulation of different

excitation wavelengths, we consider $I_{o\lambda}$ as the intensity of incident light to excite the sample. We are able to quantify the summed intensity of the emission fluorescence, S_{Em}, as

$$\mathbf{S}_{Em} \propto \mathbf{N}_{Em_j} S_{Ab_j\lambda_\varepsilon} \Phi_{F_j} I_{o\lambda}, \tag{2.4}$$

where $S_{Ab_j\lambda_\varepsilon}$ is the element in $\mathbf{S}_{Ab_j}$ that is associated with epsilon at λ_{Ex} and $\mathbf{N}_{Em_j}$ is the normalized vector of $\mathbf{S}_{Em_j}$, i.e.,

$$\mathbf{N}_{Em_j} = \frac{\mathbf{D}_{Em_j}}{\sum_{k=1}^{N} \mathbf{D}_{Em_j}}. \tag{2.5}$$

Note that the summed fluorescence emission is dependent on the incident light intensity, the absorbance magnitude at a particular λ_{Ex}, and Φ_F [24]. As shown in Figure 2.1, different λ_{Ex}s obtained from the absorption spectrum elicit different energy contributions to the spectral topology of the emission signal.

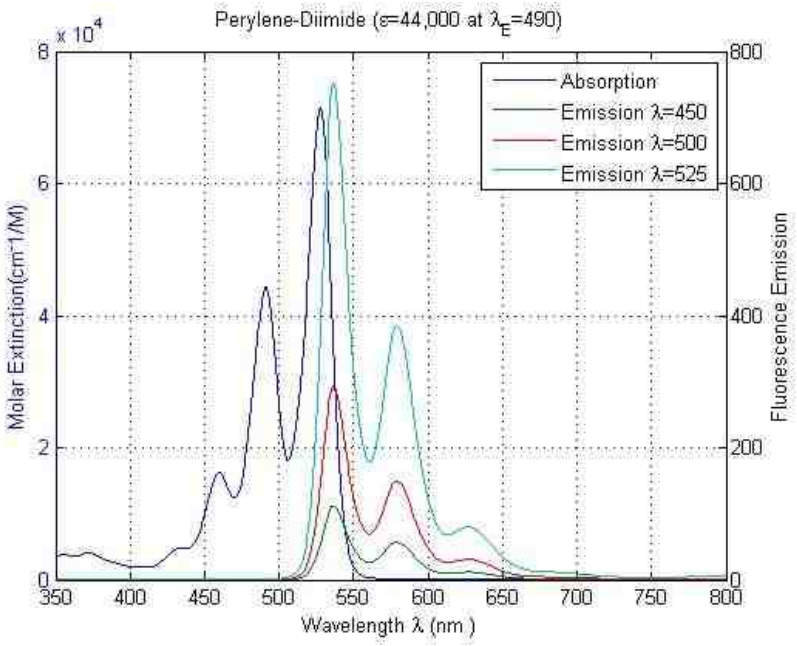


Figure 2.1: Emissions as a Function of λ and Φ_F.

2.1.4 Generating Spectra Corresponding to Combinations of Chemicals

Contingent on our assumption that the absorption and emission spectra are additive, we expand our database to include linear combinations of chemicals. This assumption is valid only when Beer-Lambert's law is obeyed and the inner filter effect is minimized, thus allowing us to define the linear combination process at the wavelength of interest as

$$A_{x+y}^{\lambda_1} = A_x^{\lambda_1} + A_y^{\lambda_1} = \epsilon_x^{\lambda_1} b C_x + \epsilon_y^{\lambda_1} b C_y. \tag{2.6}$$

One way to introduce an amount of different chemicals is by sampling without replacement and without ordering. This can be accomplished via 2.7 and expressed as a binomial coefficient, where k is the number of analytes that are chosen from a set of n total number of analytes [25]:

$$C_k^n = \frac{n(n-1)\dots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!} = \binom{n}{k} \tag{2.7}$$

However, our intention is to evaluate these analytes over different combinations of k. A visual example of four different chemicals is shown in Table 2.3.

$k=1$	$k=2$	$k=3$	$k=4$
Chem(1)	Chem(1,4)	Chem(1,2,3)	Chem(1,2,3,4)
Chem(2)	Chem(4,2)	Chem(1,2,4)	
Chem(3)	Chem(4,3)	Chem(1,3,4)	
Chem(4)	Chem(3,1)	Chem(2,3,4)	
	Chem(3,2)		
	Chem(2,1)		
Total = 4	Total = 6	Total = 4	Total = 1

Table 2.3: Evaluating Chemical Permutations.

Hence, we need to define equation 2.7 over a sum of all k:

$$\sum_{k=1}^{k=n} \frac{n!}{k!(n-k)!} = \sum_{k=1}^{k=n} \binom{n}{k}. \tag{2.8}$$

It is apparent that, as we evaluate various n number of sets using equation 2.8, we can get $2^n - 1$ different chemical combinations (the -1 is due to the fact that we are not considering the scenario where there is no chemical present within the database). For software implementation purpose, we use a binary representation for chemical presence/absence within the sample. This is illustrated by amending Table 2.3 with the appropriate coded binary representation, where each bit represents chemical presence/absence (1=Chemical Present and 0=Chemical Not Present).

$k = 1$	$k = 2$	$k = 3$	$k = 4$
Chem(1)=[0001]	Chem(1,4)=[1001]	Chem(1,2,3)=[0111]	Chem(1,2,3,4)=[1111]
Chem(2)=[0010]	Chem(4,2)=[1010]	Chem(1,2,4)=[1011]	
Chem(3)=[0100]	Chem(4,3)=[1100]	Chem(1,3,4)=[1101]	
Chem(4)=[1000]	Chem(3,1)=[0101]	Chem(2,3,4)=[1110]	
	Chem(3,2)=[0110]		
	Chem(2,1)=[0011]		

Table 2.4: Binary Coding of Chemicals.

2.2 Errors in Spectroscopy Measurements

Spectroscopy signals are also affected by electronic noise, stray light, light scattering, wavelength accuracy, resolution, stability, baseline flatness, effects of sampling geometry, and user error [26, 1]. While it is not realistic to accommodate all these types of errors, we now discuss how several additional sources of error are introduced into our synthetic spectroscopy data set.

2.2.1 User Error

Use error is quite common in spectroscopy data. The most common user error involves a lack of concentration, usually associated with pipetting chemical dilutions at low concentrations. This error can be further exacerbated if the molar absorptivity is high. Applying Beer-Lambert's law to this error, one would expect changes in peak height and overall spectral area. Another user error involves fluorescent contamination of the measured sample, or when the detected light is contaminated by Rayleigh or Raman scatter. This is also contingent on the particle size of the analyte, which is a function of the variance within the measured spectrum [27]. Figure 2.2 shows how the emission remains the same with different excitation wavelengths in a pure sample, and how a contaminate alters the emission spectrum topology at $420\,nm$ [1].

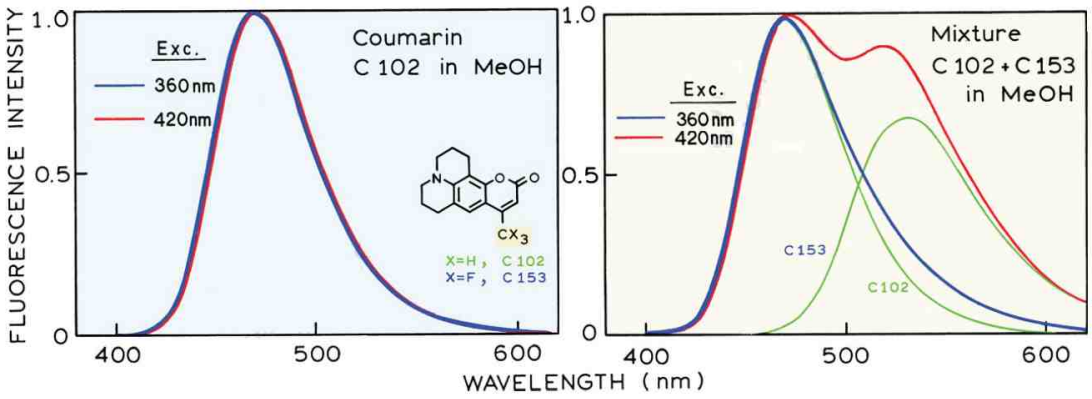


Figure 2.2: Emission Spectra of $C102$ and a Mixture of $C102$ with the Fluorescent Impurity $C153$ [1].

2.2.2 Stray Light

Stray light is the measured light of any wavelength reaching the detector that is not associated with the bandwidth of the selected wavelength [2]. Stray light manifests itself as an apparent deviation in Beer-Lambert law. The effects of stray light is a decrease in absorbance and a reduction of the perceived projected linearity of the absorbance. This can be described by equation 2.9, where I is the transmitted light,

I_s is the stray light, and I_o is the incident light:

$$Absorbance = -\log\left(\frac{I + I_s}{I_o + I_s}\right). \tag{2.9}$$

Figure 2.3 shows the effect of stray light on the absorbance [2].

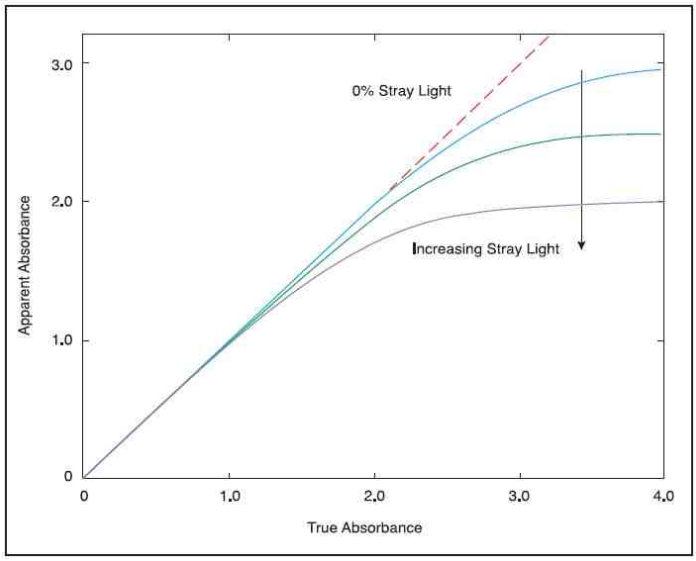


Figure 2.3: Apparent Absorbance vs True Absorbance with Increasing Stray Light [2].

2.2.3 Wavelength Accuracy

Wavelength accuracy is the inability to preserve the wavelength scaling at the detector or emitter. This scaling error introduces a shift in the measured wavelength. This causes our perception of the true λ_{max} to be inaccurate [1, 2]. See Figure 2.4.

2.2.4 Self-Absorption

Self-absorption depends upon the geometric arrangement observing the fluorescence and high optical densities, which can cause intensity distortion within specific wavelength ranges. As can be seen in Figure 2.4, the error causes a shifting of the spectrum. Figure 2.5 is an example of a right-angle observation, where short

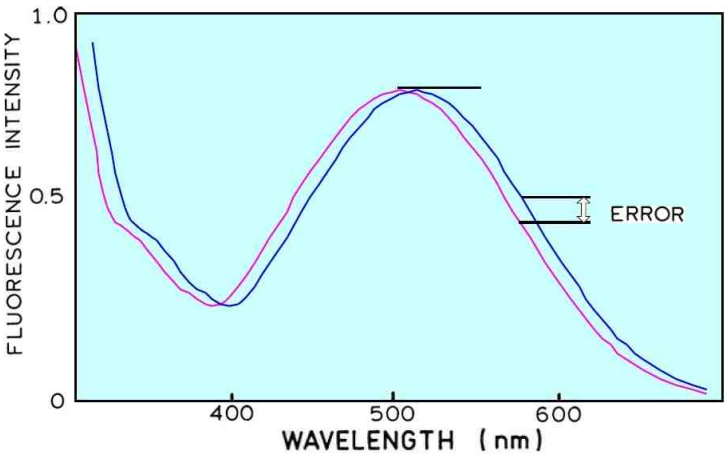


Figure 2.4: Wavelength Accuracy [1].

wavelength emissions are attenuated by the analyte Anthracence's absorbance at the shorter wavelengths [1].

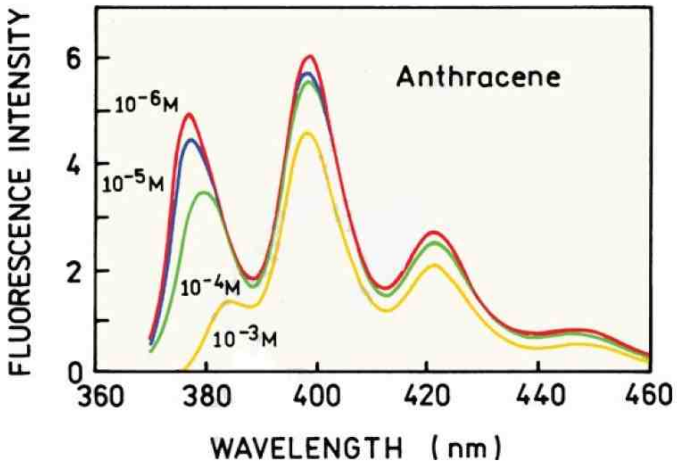


Figure 2.5: Effects of Self-Absorption.

2.2.5 Overview of Spectroscopy Errors

From the discussion above, one notices that a large proportion of the errors associated with spectroscopy measurements cause the nominal spectrum to be perturbed in a "smoother" manner. Other types of errors that generate higher frequency

perturbations in the spectroscopy measurements are typically modeled as additive white gaussian noise (AWGN). For the current purpose, our intention is to model the errors that have a higher potential to elicit a misclassification within the database. This is what we undertake in the following sections.

2.3 Insertion of Spectroscopy Measurement Perturbations

The introduction of perturbations to the database allows for the creation of more realistic test samples in our prototype data set. This enables us to explore how signal degradation can affect classification performance of our algorithm.

2.3.1 Windowing

We employ different windowing functions as the basic strategy to alter the spectral emission and absorption vectors. The *windowing foundation* is only the basis of defining, where the types of window functions will be implemented in the spectra. These locations for each type of windowing function shares a relationship with the spectral peaks of emission and absorption vectors. In later sections, the windowing functions are modified to provided either compression or dilation to the spectra within a specified design range.

Consider the following spectral data vector (corresponding to emission or absorption):

$$\mathbf{S} = \begin{bmatrix} S_1 & S_2 & \cdots & S_N \end{bmatrix}^T, \tag{2.10}$$

where $S_k \in [0, A]$, $\forall k \in \overline{1, N}$, and A is an arbitrary real positive constant. The peaks of this vector were examined to adaptively create perturbations in the vicinity of these peaks. This was done to create a unique correlated noise (Dilation or Compression) to individual spectral vector. This correlated noise is dependent on $\mathbf{S}$'s spectral peak

"shape", the windowing function type (t), and the given window size L_w, where $L_w \ll N$. The peaks were sought by implementing the Matlab function `findpeaks`, where we set a minimum window distance L_w between peaks. The `findpeaks` function yields a vector $\mathbf{P}$, where the maximum peak location for the processed spectra are represent by the elements of $\mathbf{P}$:

$$\mathbf{P} = [k_1, \quad k_2, \ \cdots k_K]^T, \tag{2.11}$$

where $k_p \in [1, N], \forall p \in \overline{1, K}$, and $K \leq N/L_w$. Given L_w and N, we determine the number of windows that will be designed by taking the integer quotient $\lfloor N/L_w \rfloor = C$, where C is the number of windows to be designed. This yields $C+1$, window segments. We determine the type of windowing function that will be employed to each window segment, i, based on our *peak indictor* vector, $\mathbf{I}_p$. The peak indictor vector informs us at what window segment a peaks occur by taking the integer quotient plus a unit ones vector $\mathbf{U}_K$, where K is the length.

$$\mathbf{I}_p = \lfloor \mathbf{P}/L_w \rfloor + \mathbf{U}_K \tag{2.12}$$

We use vector $\mathbf{I}_p$ to determine, t, the type of windowing function to implement for $\mathbf{W}_{i,t}$. Each $\mathbf{W}_{i,t}$ is characterized with one of five possible windowing type functions, t, on the i^{th} window segment, where $\forall i \in \overline{1, C+1}$. The i^{th} window segment is associated to $\mathbf{S}$'s spectral data by $[S_{1+(i-1)\cdot L_w}, S_{i\cdot L_w(i-1)+L_w}]$.

$$\mathbf{W}_{i,t} = [W_1, \quad W_2, \ \cdots \ W_{L_w}]^T, \tag{2.13}$$

where, $W_j \in [0, 1], \forall j \in \overline{1, L_w}, \forall t \in \overline{1, 5}, \forall i \in \overline{1, C}$. Equation 2.14, handles the residual data of S, since we only designed C windows.

$$\mathbf{W}_{(C+1),t} = [W_1, \quad W_2, \ \cdots \ W_{N-(L_w \cdot C)}]^T, \tag{2.14}$$

where, $W_j \in [0,1], \forall j \in \overline{1, L_w}, \forall t \in \overline{1,5}, \forall i \in \overline{1, C}$. The type of windowing function that is implemented for each i window segment is determined by recursively examining each case in numerical order until the specific conditions of a case is in accordance of the criteria. The windowing functions and conditioned criteria are defined by the following five cases, where for all cases $\alpha \in [0,1]$, $\gamma = (L_w - 1)$, and $\beta = (1 - \frac{\alpha}{2})$.

Windowing Case 1: Hanning Window

$$\mathbf{W}_{i,1}(j) = .5(1 - \cos(2\pi(\frac{j}{L_w}))) \tag{2.15}$$

if and only if $\exists i \in \mathbf{I}_p$

Windowing Case 2: Tukey Window

$$\mathbf{W}_{i,2}(j) = \begin{cases} \frac{1+\cos\left(\pi(\frac{2(j-1)}{\alpha(\gamma)})-1\right)}{2}, & \text{for } 1 \leqslant j \leqslant \frac{\alpha(\gamma)}{2} \\ 1, & \text{for } \frac{\alpha(\gamma)}{2} \leqslant j \leqslant (\gamma)(\beta) \\ \frac{1+\cos\left(\pi(\frac{2(j-1)}{\alpha(\gamma)})-\frac{2}{\alpha}+1\right)}{2}, & \text{for } (\gamma)(\beta) \leqslant j \leqslant (\gamma) \end{cases} \tag{2.16}$$

if and only if $\exists i \notin \mathbf{I}_p \wedge \exists(i-1) \in \mathbf{I}_p \wedge \exists(i+1) \in \mathbf{I}_p$

Windowing Case 3: Modified Left Tukey Window

$$\mathbf{W}_{i,3}(j) = \begin{cases} \frac{1+\cos\left(\pi(\frac{2(j-1)}{\alpha(\gamma)})-1\right)}{2}, & \text{for } 1 \leqslant j \leqslant \frac{\alpha(\gamma)}{2} \\ 1, & \text{for } \frac{\alpha(\gamma)}{2} \leqslant j \leqslant (\gamma) \end{cases} \tag{2.17}$$

if and only if $\exists i \notin \mathbf{I}_p \wedge \exists(i-1) \in \mathbf{I}_p \wedge \exists(i+1) \notin \mathbf{I}_p$

Windowing Case 4: Modified Right Tukey Window

$$\mathbf{W}_{i,4}(j) = \begin{cases} 1, & \text{for } 1 \leqslant j \leqslant (\gamma)(\beta) \\ \frac{1+cos(\pi(\frac{2(j-1)}{\alpha(\gamma)})-\frac{2}{\alpha}+1)}{2}, & \text{for } (\gamma)(\beta) \leqslant j \leqslant (\gamma) \end{cases} \tag{2.18}$$

if and only if $\exists i \notin \mathbf{I}_p \wedge \nexists(i-1) \in \mathbf{I}_p \wedge \exists(i+1) \in \mathbf{I}_p$

Windowing Case 5: Null Variance Window

$$\mathbf{W}_{i,5}(j) = 1, \qquad \text{for } \forall j \tag{2.19}$$

if and only if $\exists i \notin \mathbf{I}_p \wedge \exists(i-1) \notin \mathbf{I}_p \wedge \exists(i+1) \notin \mathbf{I}_p$

Each individual $\mathbf{W}_{i,t}$ window vector that is designed will be cascaded in numerical order to construct, $\mathbf{J}_F$, the foundation for creating our dilation compression vector to modify vector $\mathbf{S}$. Hence, $\mathbf{J}_F$ is defined as:

$$\mathbf{J}_F = [\mathbf{W}_{1,t},\ \mathbf{W}_{2,t},\ \ \cdots\ \mathbf{W}_{C,t},\ \mathbf{W}_{(C+1),5}]^T, \tag{2.20}$$

where $\mathbf{W}_{i,t} \in [0, A], \forall i \in \overline{1, C}, \forall t \in \overline{1, 5}$, and $\mathbf{J} \in [0, A], \forall i \in \overline{1, (C+1)}$. However, note we only design C windows, the (C+1) window will always default to case 5 to handle the residual data of $\mathbf{S}$ thats smaller than the specified window size, L_w.

2.3.2 Basic Design for Dilation Noise

Once our foundation vector is set, we modify, $W_{i,t}$. We demonstrate a few different methods to modify $W_{i,t}$ to create the most ideal synthetic noise representation starting from the most basic. As well, we review the down falls in order to improve upon each method. We first exemplify synthetic noise to the data by simply dilating the window locations where the peaks occur. This dilation is scaled by a constant η. Therefore,

we revamp $W_{1,t}$, to be scaled by η and force other Tukey windows $W_{2-4,t}$ to one:

$$\hat{\mathbf{W}}_{\hat{D}_{1,t}} = (\eta) \cdot \mathbf{W}_{1,t} + U_{L_w}, \tag{2.21}$$

$$\hat{\mathbf{W}}_{\hat{D}_{2-5,t}} = (0) \cdot \mathbf{W}_{2-5,t} + U_{L_w}, \tag{2.22}$$

$$\mathbf{J}_{\hat{D}} = [\hat{\mathbf{W}}_{\hat{D}_{1,t}},\ \hat{\mathbf{W}}_{\hat{D}_{2,t}},\quad \cdots\ \hat{\mathbf{W}}_{\hat{D}_{C,t}},\ \mathbf{W}_{(C+1),5}]^T, \tag{2.23}$$

$$\mathbf{S}_{S+N} = \mathbf{J}_{\hat{D}}\ diag(\mathbf{S}) \tag{2.24}$$

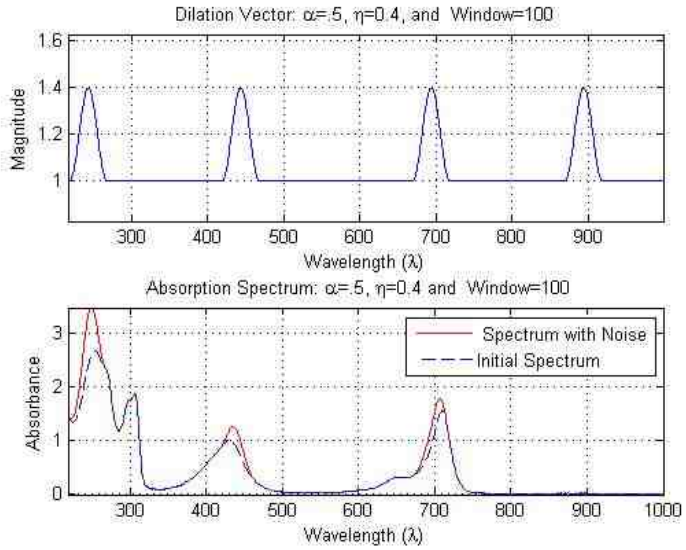


Figure 2.6: Absorbance Spectrum with Dilation Noise.

We can see in figure 2.6, that by implementing this method we are only given the option of dilating the spectrum at a constant η. In addition, its worth noting due to the fixed windowing segments, the alignment of the peaks are not centered directly over the windowed segment causing shifting of the peak wavelength. In some scenarios, this maybe considered ideal for further perturbations of the signal, where the shifting of the peak wavelength is a function of η.

2.3.3 Basic Design for Compression Noise

Given that we can dilate the signal, in order for us to compress the signal, a compression vector is designed as such by equation 2.25, 2.26, and 2.27, which is pictorially represented in Figure 2.8:

$$\hat{\mathbf{W}}_{\hat{C}_{i,t}} = \eta \cdot \mathbf{W}_{i,t} + (U_{L_w} - \eta U_{L_w}) \tag{2.25}$$

$$\mathbf{J}_{\hat{C}} = [\hat{\mathbf{W}}_{\hat{C}_{1,t}},\ \hat{\mathbf{W}}_{\hat{C}_{2,t}},\ \cdots\ \hat{\mathbf{W}}_{\hat{C}_{C,t}},\ \mathbf{W}_{(C+1),5}]^T, \tag{2.26}$$

$$\mathbf{S}_{S+N} = \mathbf{J}_{\hat{C}}\ diag(\mathbf{S}) \tag{2.27}$$

where $D_k \in [0, A], \forall k \in \overline{1, N}$

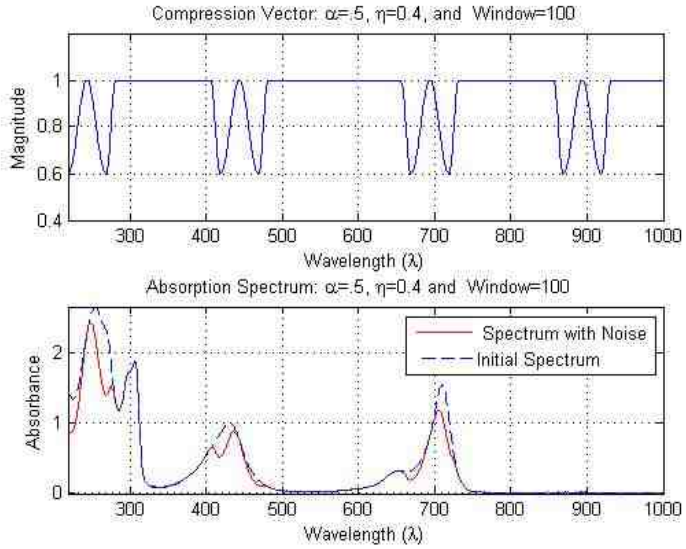


Figure 2.7: Absorbance Spectrum with Compression Noise.

2.3.4 An Axiomatic Approach for Designing Appropriate Noise

In the previous sections, we discussed the foundation of the windowing and the functions that were integral for dilation and compression of the signal by a factor of η. It's only sensible to anticipate the above methods as a combination of random

dilation and compression for a more convincing simulated noise. Considering these two functions as a whole, comprised of different adjacent windowing techniques, it is careless to assume that η's effect on the signal is equivalent. A pragmatic approach of declaring guidelines for the noise algorithm must be established to meet the appropriate standards to have unbiased detection simulations when noise is introduced. In order to fortify this concept of why guidelines are required and to exam what guidelines that need to be put in place, we introduce an algorithm that functionally fails as a dilating compressing algorithm.

Failure for Proper Dilation and Compression

$$\hat{\mathbf{W}}_{i,1} = \eta X \mathbf{W}_{i,1} + (1 - \frac{\eta}{2}) U_{L_w}, \tag{2.28}$$

$$\hat{\mathbf{W}}_{i,2-4} = (\eta) \mathbf{W}_{i,2-4} + (1 - \frac{\eta}{2}) U_{L_w}), \tag{2.29}$$

$$\hat{\mathbf{W}}_{i,5} = \mathbf{W}_{i,5},, \tag{2.30}$$

$$\mathbf{J}_{\hat{CD}} = [\hat{\mathbf{W}}_{\hat{D}_{1,t}}, \ \hat{\mathbf{W}}_{\hat{D}_{2,t}}, \quad \cdots \ \hat{\mathbf{W}}_{\hat{D}_{C,t}}, \ \mathbf{W}_{(C+1),5}]^T, \tag{2.31}$$

$$\mathbf{S}_{S+N} = \mathbf{J}_{\hat{CD}} \ diag(\mathbf{S}) \tag{2.32}$$

where X is a random variable uniformly distributed $X \in [0, 1]$ Visually examining the noise vector, we can note that the hanning windows exceeds the peak magnitude of one. We may assume the data vector will undergo dilation and compression, but this is deceiving. When the filter is applied to the data vector the data is not properly dilated as expected. It begins to become apparent, graphically in Figure 2.8 that, the magnitude of dilation on the sample may not be comparable to the compression. However, the improper dilation may be attributed to that particular sample, therefore simulations are done in the preceding section to examine how the intensity changes are distributed across numerous samples.

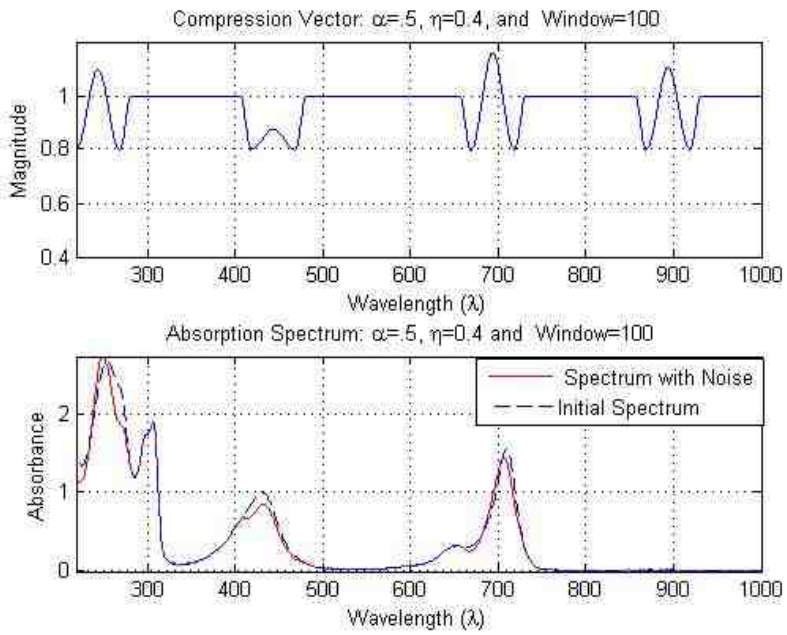


Figure 2.8: Absorbance Spectrum with Noise Failing to Dilate.

Noise Design Guidelines

The previous example shows the possible cause of how the noise is distributed over the vector. When examined over numerous trials, it is possible that it may not be ideal for testing our detection algorithm in the later chapters. Hence, we set guidelines of what our ideal noise distribution should be in order to achieve an appropriate testing scenario for our detection algorithm. The guidelines are verified by visual inspection from simulations that produce the noise distribution.

Synthetic Noise Distribution Ideal Guidelines

1. η will be random for each designed window.
2. Within a single noise vector, Dilation and Compression can occur at various locations.
3. As η increases, the variance of the distribution will increase.

4. The noise distribution when evaluating the magnitude between corresponding elements should be symmetric around zero.

5. The distribution of intensity loss and gain for a entire vector should be symmetric around zero.

These guidelines will assist in the development of a pragmatic noise that will better challenge the subsequent detection process. The simulation that is imposed to validate if the noise distribution meets the guideline criteria is created with a database size of eleven Chemicals yielding $2^{Chem} - 1 = 2047$ chemical combinations , to process 6000 randomly picked chemicals. The correlated noise algorithm applied the following noise equations 2.28 ,2.28, and 2.28. A normalization for element intensity in equation 2.33 was imposed, since the absorbance can vary so greatly, N_{EI}.

$$\mathbf{N}_{EI} = \frac{\mathbf{S}}{max(\mathbf{S})} \tag{2.33}$$

In order to create a robust synthetic noise model, we need to account for dilation and compression equally. We are able to note the displacement of the intensity to the corresponding elements with a histogram representation of $\Delta\mathbf{I}_E$, defined as 2.34. The parameter η is varied over four simulation to evaluate the effects of the variance on the distribution, as shown in Figure 2.9, identifying the asymmetry in the noise distribution. This current noise model does not meet the guidelines stated and is therefore non-ideal. The model provided would provide questionable results in the detection algorithm, where it is ambiguous if the algorithm failed or if it was a slight change in η causes a large shift in the attenuation of the original signal.

$$\Delta\mathbf{I}_E = \mathbf{J}_{\hat{CD}}\ diag(\mathbf{N}_{EI}) - \mathbf{N}_{EI} \tag{2.34}$$

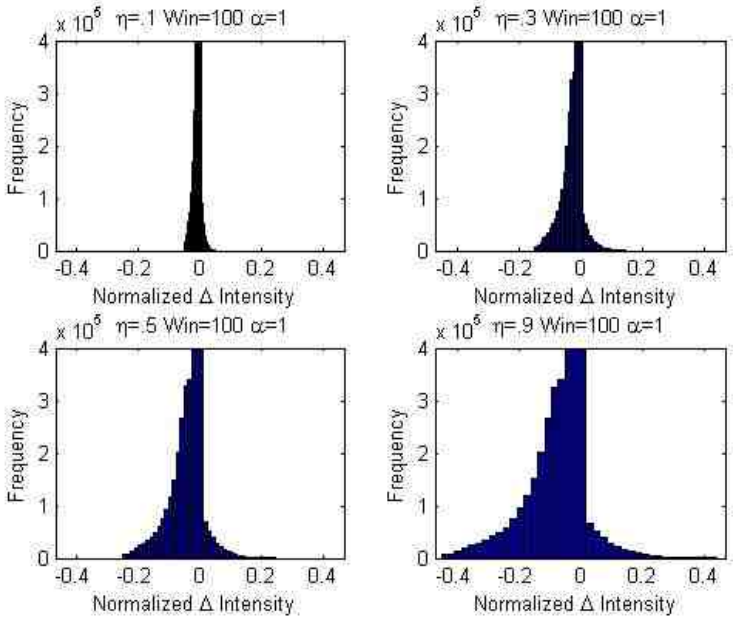


Figure 2.9: Histogram of the Change in Intensity of Corresponding Elements.

Figure 2.10 examines the change in intensity over the entire vector, which solidifies the need for the stated guidelines of (4) and (5). We can can see that from the distribution that a shifting is occurring as η increases, preventing any type of dilation to occur within the spectrum. The change of intensity over the entire vector, $\mathbf{N}_{VI}$, of the normalized spectrum, $\mathbf{N}_{VI}$, was calculated by equations 2.35 and 2.36, to achieve the histogram Figure 2.10.

$$\mathbf{N}_{VI} = \frac{\mathbf{S}}{\sum_{k=1}^{N}(\mathbf{S})} \tag{2.35}$$

$$\Delta\mathbf{I}_V = \sum_{k=1}^{N}(\mathbf{J}_{\hat{CD}}\ diag(\mathbf{N}_{VI}) - \mathbf{N}_{VI}) \tag{2.36}$$

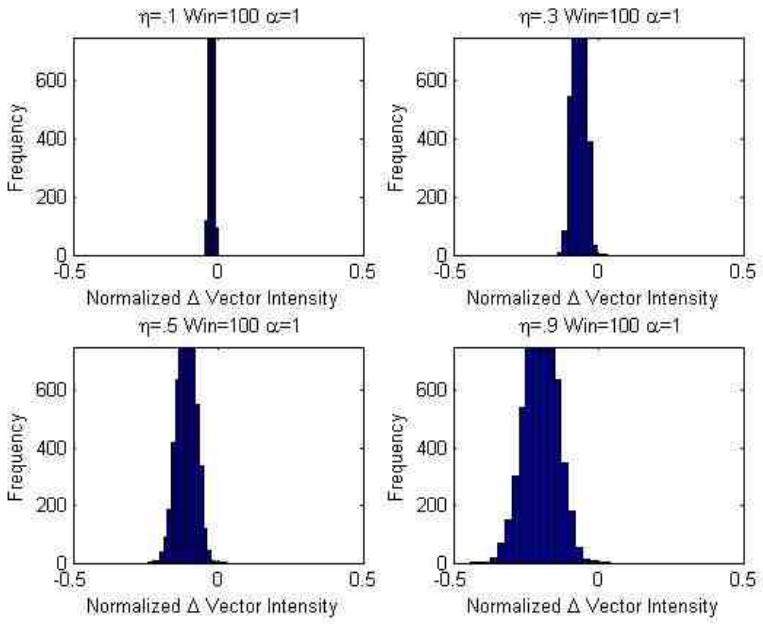


Figure 2.10: Histogram of the Change in Intensity of Entire Vector.

The cause can be further explained by examining the expectation of the designed noise vector $W_{i,1}$ elements' j and the adjacent windowing vectors.

$$E[\hat{\mathbf{W}}_{i,1_j}] = E[\eta \cdot X \cdot W_{i,1_j} + U_j - \frac{\eta}{2}U_j]$$
$$E[\hat{\mathbf{W}}_{i,1_j}] = \eta \cdot W_{i,1_j}E[X] + E[U_j] - E[\frac{\eta}{2}U_j]$$
$$E[\hat{\mathbf{W}}_{i,1_j}] = \eta \cdot W_{i,1_j}E[X] + E[1] - E[\frac{\eta}{2}]$$
$$E[\hat{\mathbf{W}}_{i,1_j}] = \eta \cdot W_{i,1_j}\frac{1}{2} + 1 - \frac{\eta}{2}$$

The adjacent windowing elements are constants and do not need to be evaluated ($W_{i,3}$ $W_{i,4}$), but should be examined pictorially with $W_{i,1}$ in Figure 2.11.

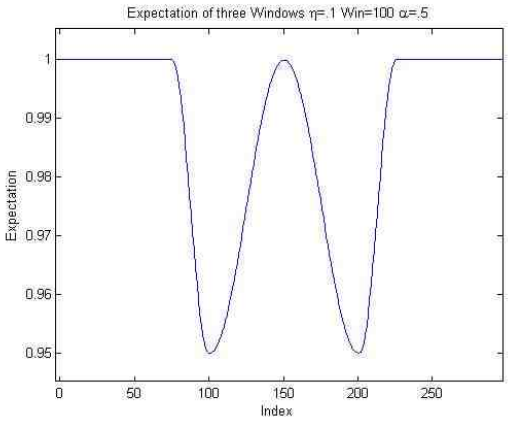


Figure 2.11: Windowing Segment Examining Expectation.

We can now note, pictorially from Figure 2.11, that none of the elements expectations surpass one, ultimately causing a summed loss of intensity or compression of the signal. We can also note that anytime these adjacent windows occur ($W_{i,2}$ $W_{i,3}$ $W_{i,4}$), we will always encounter a further induced intensity loss. This intensity lose is not equally compensated by dilation. Furthermore, even when dilation occurs within the function 2.28 it still compresses the width of the peak, even though it's an effective method for dilation. This deficiency can be attributed to the innate way we window, since the peak of the signal can range anywhere with the window length L_w. However, we should not see this as a shortcoming since it causes further realist noise by shifting the spectrum by a function of η.

2.3.5 Approximating an Advantageous Synthetic Noise

We propose the following combined dilation and compression noise functions to fulfill the following guidelines, based on the windowing foundation function 2.20. In order to satisfy these requirements, constraints are imposed to the functions. However, we maintain the functions robustness to dynamically be altered for compression and dilation.

Defining a Robust function

This is done by using a binomial distribution to create an indictor function **I** to determine what $\mathbf{W}_{i,1}$ function to implement, compression $\hat{\mathbf{W}}_{C_{i,1}}$ or dilation $\hat{\mathbf{W}}_{D_{i,1}}$, where $p = .5$, in equation 2.38. This allows further manipulation of the function for us to engage the guidelines criteria more stringently by enabling two functions to competent against each other to achieve an approximate zero mean distribution.

$$\mathrm{I} = \begin{cases} 1, & \text{with probablity p} \\ 0, & \text{with probablity (1-p)} \end{cases} \tag{2.37}$$

The adjacent window's elements are dependent on the design of $\hat{\mathbf{W}}_{i,1}$. We designed $\hat{\mathbf{W}}_{i,1}$ from the fundamentals of 2.21 and 2.25 for the compression and dilation vectors. Thus, the following equations 2.39, 2.40 were developed for the $\mathbf{W}_{i,1}$ windowing case.

$$\hat{\mathbf{W}}_{B_{i,1}} = \hat{\mathbf{W}}_{C_{i,1}}\mathrm{I} + \hat{\mathbf{W}}_{D_{i,1}}(\mathrm{I}-1) \tag{2.38}$$

$$\hat{\mathbf{W}}_{C_{i,1}} = (A_{c1}X_2\eta)\mathbf{W}_{i,1} + \mathbf{U}_{L_w}(1 - A_{c2}\eta + A_{c3}\eta X_3) \tag{2.39}$$

$$\hat{\mathbf{W}}_{D_{i,1}} = (A_{d1}X_1\eta)\mathbf{W}_{i,1} + \mathbf{U}_{L_w} \tag{2.40}$$

where $\eta \in [0,1], X_i$ is a uniform random variable $\in [0,1], \forall i \in \overline{1,3}$, and $A_x i$ are coefficients to control the functions. In order to avoid discontinuities between the adjacent windows, due to biasing from each window from the random variable X_3, the following adjustments were made to the following cases:

$$\hat{\mathbf{W}}_{i,2} = \mathbf{W}_{i,2}(1 - \hat{\mathbf{W}}_{i+1,1}(1)) + \hat{\mathbf{W}}_{i+1,1}(1), \qquad \text{for } (\gamma)(\beta) \leqslant j \leqslant (\gamma) \tag{2.41}$$

$$\hat{\mathbf{W}}_{i,2} = \mathbf{W}_{i,2}(1 - \hat{\mathbf{W}}_{i-1,1}(L_W)) + \hat{\mathbf{W}}_{i-1,1}(L_W), \qquad \text{for } 1 \leqslant j \leqslant \tfrac{\alpha(\gamma)}{2} \tag{2.42}$$

$$\hat{\mathbf{W}}_{i,3} = \mathbf{W}_{i,3}(1 - \hat{\mathbf{W}}_{i+1,1}(1)) + \hat{\mathbf{W}}_{i+1,1}(1), \tag{2.43}$$

$$\hat{\mathbf{W}}_{i,4} = \mathbf{W}_{i,4}(1 - \hat{\mathbf{W}}_{i-1,1}(L_W)) + \hat{\mathbf{W}}_{i-1,1}(L_W), \tag{2.44}$$

Constraint A:

Constraint A is to bound the expected dilation and compression functions to be equivalent. This will assist one functions maximum from overpowering the other and maintain a equivalent magnitude changes.

$$
\begin{aligned}
1 - E[\min_{j \in \overline{1, L_W}} (\hat{W}_{C_{i,1}j})] &= E[\max_{j \in \overline{1, L_W}} (\hat{W}_{D_{i,1}})] - 1 \\
1 - E[1 - A_{c2}\eta + A_{c3}\eta X_3] &= E[A_{d1} X_1 \eta + 1] - 1 \\
1 - E[1] + E[A_{c2}\eta] - E[A_{c3}\eta X_3] &= E[A_{d1} X_1 \eta] + E[1] - 1 \\
E[A_{c2}\eta] - E[A_{c3}\eta X_3] &= E[A_{d1} X_1 \eta] \\
\eta A_{c2} - A_{c3}\eta E[X_3] &= A_{d1}\eta E[X_1] \\
A_{c2} - A_{c3} E[X_3] &= A_{d1} E[X_1] \\
A_{c2} - \frac{1}{2} A_{c3} &= \frac{1}{2} A_{d1} \\
2A_{c2} - A_{c3} &= A_{d1}
\end{aligned}
$$

Constraint B:

Constraint B is to bound the variance of the dilation and compression functions to be equivalent. This will prevent the dilation and compression functions from causing an asymmetry within the noise distribution.

$$Var[\min_{j\in\overline{1,L_W}}(\hat{W}_{C_{i,1}j})] = Var[\max_{j\in\overline{1,L_W}}(\hat{W}_{D_{i,1}})]$$
$$Var[1 - A_{c2}\eta + A_{c3}\eta X_3] = Var[A_{d1}X_1\eta + 1]$$
$$Var[A_{c3}\eta X_3] = Var[A_{d1}X_1]$$
$$A_{c3}^2\eta^2 Var[X_3] = A_{d1}^2\eta^2 Var[X_1]$$
$$A_{c3}^2 = A_{d1}^2$$

Constraint C:

Constraint C is to bound the compression function's maximum element in that vector to never surpass the maximum value of the dilation function's maximum element.

$$\min_{j\in\overline{1,L_W}}(\hat{W}_{C_{i,1}j}) \leq \max_{j\in\overline{1,L_W}}(\hat{W}_{D_{i,1}})$$
$$A_{c1}\eta + 1 - A_{c2}\eta + A_{c3}\eta X_3 \leq A_{d1}X_1\eta + 1$$
$$A_{c1}\eta - A_{c2}\eta + A_{c3}\eta X_3 \leq A_{d1}X_1\eta$$
$$A_{c1} - A_{c2} + A_{c3} \leq A_{d1}$$

Constraint D:

Recalling Figure 2.11, based on the topology of design there will always be a greater intensity loss with the compression function is implemented. To achieve an approximate zero mean intensity loss for the vector, stated by the fifth guideline, we minimize the expected intensity loss of the vector when compared in a random process of dilation and compression by exploiting the binomial equation. Using the binomial equation to our advantage, we can design the $\hat{\mathbf{W}}_{C_{i,1}}$ and $\hat{\mathbf{W}}_{D_{i,1}}$ functions accordingly having them achieve an approximate averaged expected intensity value over all elements, 2.45. In order to account for this average expectation, we account for the effects of the adjacent windows as well for a proper approximation. In

the dilations case we used $\mathbf{W}_D = [\mathbf{W}_{i,5}, \hat{\mathbf{W}}_{i,1}, \mathbf{W}_{i,5}]$ and for the compression case we used $\mathbf{W}_C = [\hat{\mathbf{W}}_{i,3}, \hat{\mathbf{W}}_{i,1}, \hat{\mathbf{W}}_{i,4}]$. Therefore, base on the design of $\hat{\mathbf{W}}_{i,1}$ for the compression and dilation we can manipulate our indictors function's p value to further optimize the expected intensity to have an approximate equivalent intensity deviation of compression and dilation.

$$E_{win} = p\frac{1}{3 * L_w}(\sum_{j=1}^{3 \cdot Lw} E[W_C]) + (1-p)\frac{1}{3 * L_w}(\sum_{j=1}^{3 \cdot Lw} E[W_D]) \approx 1 \tag{2.45}$$

The following approximation was done with $\alpha = 1, \eta = .1, A_{c1} = A_{c2} = A_{c3} = A_{d1} = .5$ and $L_w = 100$, where $E_{win} = 0.9918$ with a $p = .33$, implementing this optimize noise at these parameters we manifest others noise vectors at various different η's assuming the changes are minute. If desired you can optimize the p for different etas. In Figure 2.12, we used the aforementioned parameters to designed noise at $\eta = 1$, where simulations of the distributions of element and vector intensity is shown in 2.13.

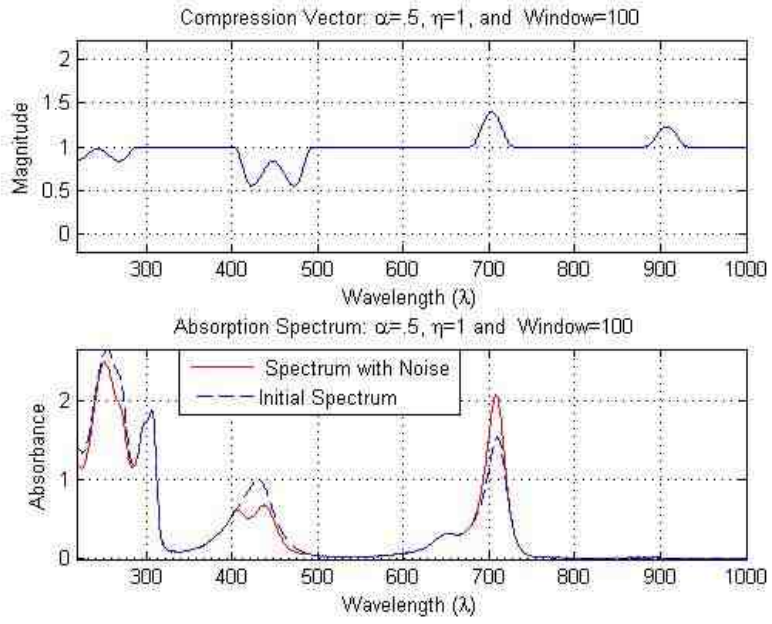
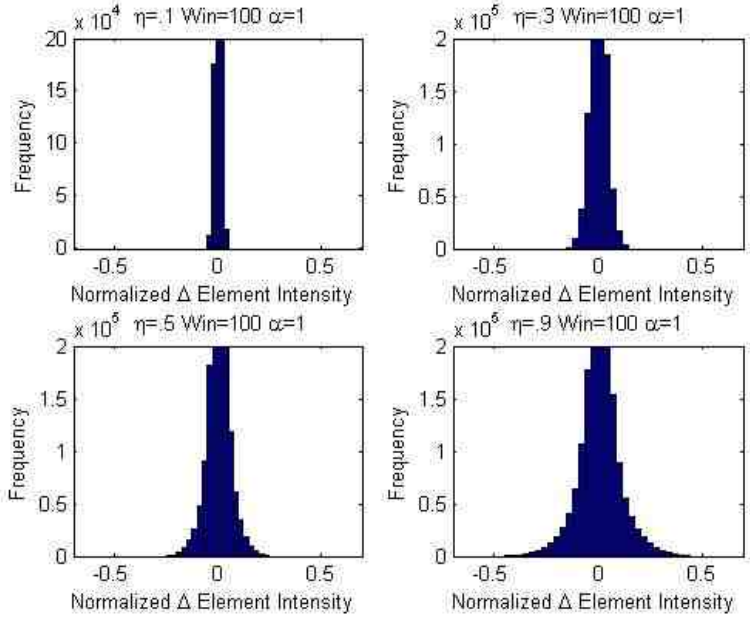
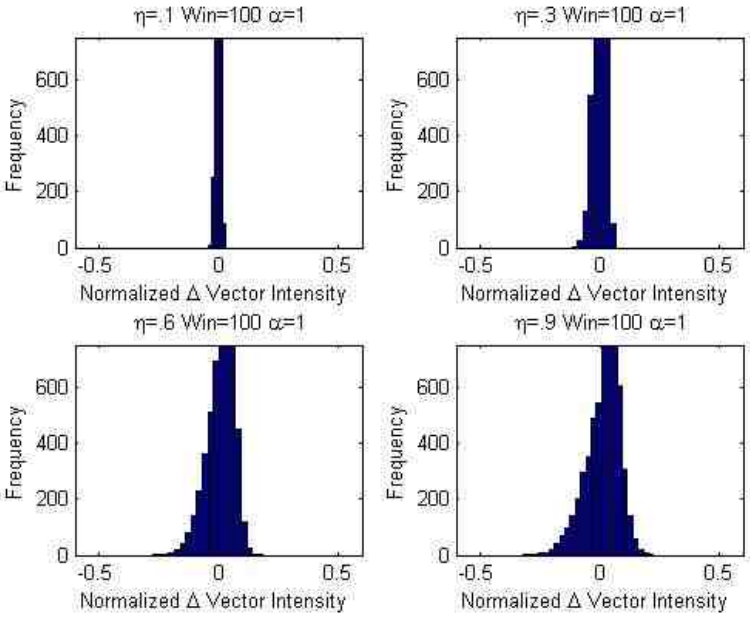


Figure 2.12: Simulated Optimized Noise.



(a) Element Intensity



(b) Vector Intensity

Figure 2.13: Histograms of Intensity.

Chapter 3

Conventional Methods of Chemometrics

In the field of spectroscopy and chemometrics, Beer-Lambert's law can be considered a major cornerstone for detection-related problems. As was discussed in Section 1.2.2, the molar extinction coefficients for an analyte are unique and they are exploited to identify compounds and their compositions. The majority of chemometric methods use the linear relationship of the absorptivity coefficients betoken by the Beer-Lambert's law for quantitative and qualitative analysis.

3.1 Quantitative Analysis Methods

The most common practice of UV/Vis spectroscopy is quantitative measurement of an analyte's concentration using the absorption spectrum [28]. Calibration methods or standards are implemented to determine the concentration by graphically extrapolating the data or by using regression methods [28]. There are various types of methods of regression analysis, each possessing its own advantages and disadvantages. *Monocomponent analysis* is the most basic and easiest method, but it is based on many assumptions and lacks the realistic occurrences in an experimental design. *Multicomponent analysis* methods contain fewer assumptions, but are typically more complex thus preventing many chemists from utilizing the methods.

3.1.1 Simple Linear Regression

Linear regression is the simplest of the quantitative methods; it falls into the monocomponent analysis category. Linear regression is typically implemented when we have an isolated compound and the molar absorptivity is unknown [29]. This method uses part of the Beer-Lambert's law which states that the absorption and concentration are proportional. A reliable measurement of concentration is calculated by creating a calibration curve, which visually plots absorbance versus concentration. The way we graph absorbance points on the calibration curve is dependent on the spectral resolution of the instrument. We may examine the peak intensity, A_{peak}, if the resolution is high. If the resolution is low, we evaluate the area of the peak, A_{area}, as in

$$C_x = \beta_1 A_{Area} + \beta_2, \tag{3.1}$$

where C is the concentration of the compound and β is the calibration coefficients. Figure 3.1 shows an ideal example (with zero error) of a calibration curve, which would be simple to solve for the calibration coefficients.

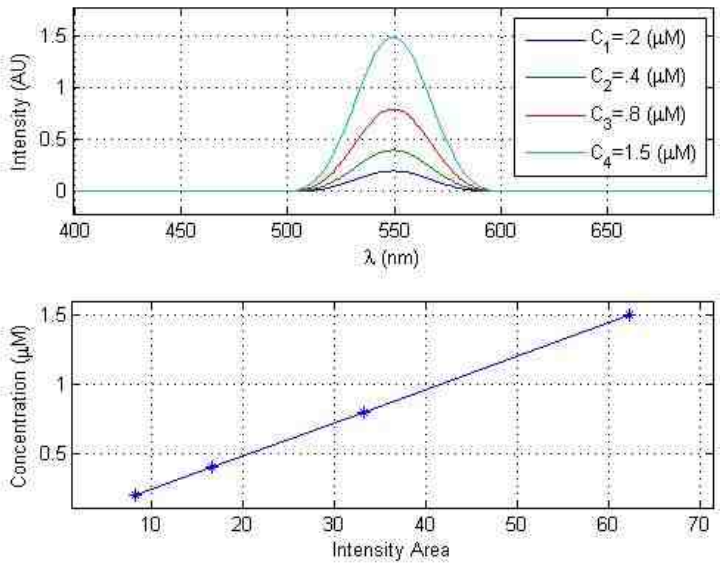


Figure 3.1: An Ideal Calibration Curve.

When instrumentation or user error is introduced there is no perfect solution for the calibration coefficients. We are required to find the coefficients by linear regression. The coefficient values that are chosen produce the least amount of error in approximating the linear relationship [30]. Linear Regression coefficients are defined as:

$$\beta_1 = \frac{\sum_i [(A_i - \bar{A})(C_i - \bar{C})]}{\sum_i (A_i - \bar{A})^2}; \quad \beta_2 = \bar{C} - \beta_1 \bar{A}, \tag{3.2}$$

where $(\bar{A}, \bar{C})$ are the means of A and C values, respectively. This position $(\bar{A}, \bar{C})$ is referred to as the *centroid* [30]. In Figure 3.2, noise has been introduced into the system depicted in Figure 3.1 and the calibration line was determined by linear regression.

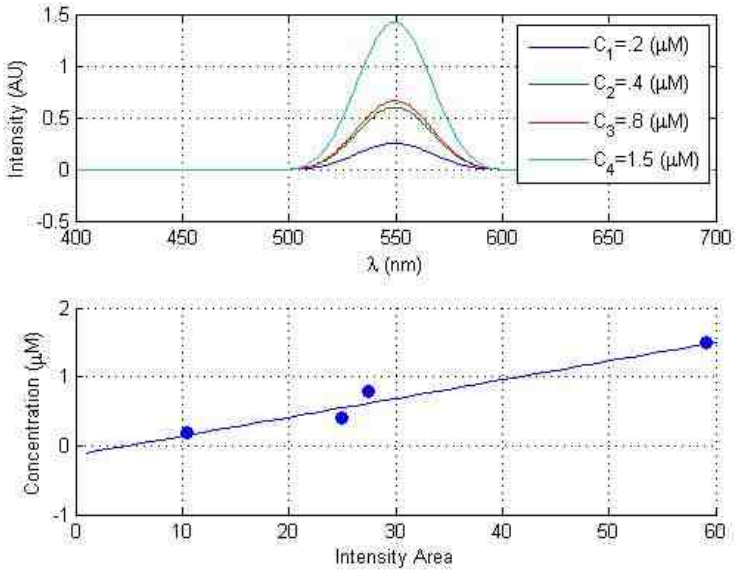


Figure 3.2: Calibration Curve with Linear Regression

This monocomponent analysis method however is incapable of examining more then one compound. So, this particular analysis method is only effective for an analyte that has been physically isolated. [31]. If this method was to be implemented in the evaluation of a secondary compound, in order to achieve accurate results the spectral bands could not have any interaction, and would require an additional equation.

Another disadvantage of using this method is the assumption that only one absorption value is estimated, making the calibration regression line more susceptible to noise. Overall, this method's simplicity limits its reliability and general functionality.

3.1.2 Classical Least Squares Method

The classical least square method of implementing the Beer-Lambert's law allows for the examination of multiple spectral wavelengths, compounds, and concentrations. This method does however make the assumption that the pathlength is kept constant for all data comparisons. Defining Beer-Lambert's law in equation: 3.3,

$$A_{\lambda p} = \beta_{m,\lambda p} C_m, \tag{3.3}$$

where the molar absorptivity and pathlength are substituted under β, p is the specified wavelength, and m is the chemical. We could easily solve for a single constant of β, but this would be no different than implementing the linear regression model that was previously discussed in Section 3.1.1. The purpose of classical least mean squares is to examine multiple βs, where a β coefficient describes the linear relation of absorption by multiple chemicals at a specific wavelength. Equation 3.4, describes a simple example of this linear relationship evaluating two different chemicals at two different wavelengths:

$$\begin{pmatrix} A_{\lambda 1} \\ A_{\lambda 1} \end{pmatrix} = \begin{pmatrix} \beta_{a,\lambda 1} & \beta_{b,\lambda 1} \\ \beta_{a,\lambda 2} & \beta_{b,\lambda 2} \end{pmatrix} \begin{pmatrix} C_a \\ C_b \end{pmatrix} + \begin{pmatrix} E_{\lambda 1} \\ E_{\lambda 1} \end{pmatrix}. \tag{3.4}$$

We extend this model over multiple spectra samplings, chemicals and wavelengths by equation 3.5:

$$\mathbf{A} = \boldsymbol{\beta}\mathbf{C}, \tag{3.5}$$

where

$$\mathbf{A} = \begin{pmatrix} A_{1,1} & \cdots & A_{N,1} \\ \vdots & \ddots & \vdots \\ A_{1,P} & \cdots & A_{N,P} \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \beta_{1,1} & \cdots & \beta_{M,1} \\ \vdots & \ddots & \vdots \\ \beta_{1,P} & \cdots & \beta_{M,P} \end{pmatrix}; \mathbf{y} = \begin{pmatrix} C_{1,1} & \cdots & C_{1,N} \\ \vdots & \ddots & \vdots \\ C_{M,1} & \cdots & C_{M,N} \end{pmatrix}. \tag{3.6}$$

Here, N is the number of samples of experiments run on the spectrometer, P is the number of wavelengths chosen, and M is the number of chemicals. In order to approximate β by $\hat{\beta}$, we are required to invert C by using the psuedoinverse method since the matrix is not square. This yields

$$\hat{\boldsymbol{\beta}} = \mathbf{A}\mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}. \tag{3.7}$$

3.2 Qualitative Analysis Methods

Qualitative analysis methods of UV/Vis spectroscopy for the identification of analytes are often accomplished by correlating the sample to a known database [32, 33]. In Section 1.2.3, we discussed the correlation of the spectrum to the molecular orbital location. We are therefore able to gain insight into the molecular orbital features identifying their ground and excited states. This ability to determine electron placement is incorporated to processes in titrations, for determining titration end points and equilibrium constants, as well [31].

3.2.1 Principle Component Analysis

Principle component analysis (PCA) generates a new coordinate system by an orthogonal linear transformation of the previous coordinate system. The new variables are known as *principle components*. For example, equation **??** and **??**, demonstrates the linear transformation of a multivariate case of N variables defining

the first two principle component Z_1 and Z_2 .

$$Z_1 = a_{11}X_1 + a_{12}X_2 + a_{13}X_3 + \cdot a_{1n}X_N \tag{3.8}$$

$$Z_2 = a_{21}X_1 + a_{22}X_2 + a_{23}X_3 + \cdot a_{2n}X_N \tag{3.9}$$

We are able to reduce the dimensionality of a multivariate problem using PCA by describing the data set by only the initial principle components.These principle components enhance the variables of the data to aid in their discrimination between each other. The initial principle components offer the most variability of the data set, therefore are the most discriminative. The higher the principle component is, the less variability and hence less discriminative. Figure 3.3 is an example in two-dimensional space that illustrates how PCA rotates the data.

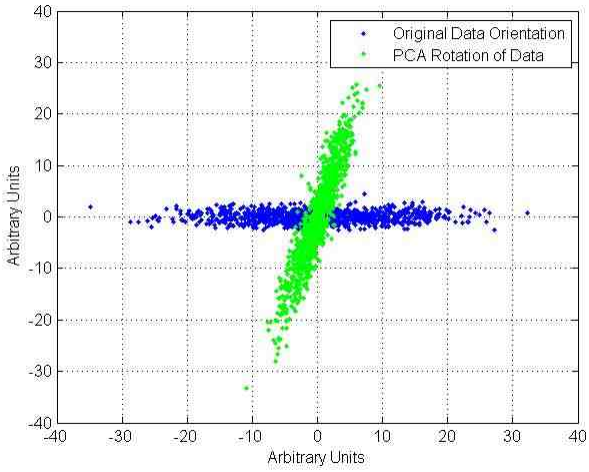


Figure 3.3: Illustration of PCA Rotating Data to Another Orthogonal Plane.

These dimensionally methods are instrumental in chemometrics because of the size of the multivariate analysis that is undertaken. PCA is a beneficial way of aiding in the discrimination of data sets by rotation, however it is still necessary to implement additional numerical methods to ultimately discriminate different spectra, such as linear regression, Gaussian mixture models, or a clustering scheme.

3.2.2 Parallel Factor Analysis

Parallel factor analysis (PARAFAC) is a method used to determine the optimal way for orientation of the axes to discriminate spectra. PARAFAC uses the intrinsic properties of the data under study to determine these unique latent factors (i.e., hidden factors) to find the optimal orientation. Cattell provided the conceptual framework for PARAFAC. He examined multiple principles in determining the optimal discriminating axes for the data set proposing parallel proportional profiles as the most fundamental principle for determining best fitting axis orientation [34]. The benefits of parallel proportional profiles is not a rotational indeterminate formulation, unlike PCA, where rotation is implemented to determine the best fitting axes orientation. This rotational scheme prevents the reconstruction ability. The concept of parallel proportional profiles exploits the fact that the variation of two data set matrices should be some factor of the other. This is done by extracting the best-fitting unique factor axes (i.e., latent factors) by decomposing the three-way data cube of factors into individual rank-1 matrices, as shown in Figure 3.4. PARAFAC uses an iterative procedure performed by an alternating least squares algorithm to converge on the set of best-fitting axes that can best explain the patterns of expansion and contraction across each data set slice from the three-way data cube PARAFAC.

When implemented within the context of chemometrics, PARAFAC employs the following decomposition of the data:

$$x_{ijk} = \sum_{f=1}^{F} a_{if} b_{if} c_{kf} + e_{ijk}, \tag{3.10}$$

where i is the sample, j is the emission and k is the excitation, a_{if} is the concentration, b_{if} is emission data, and c_{kf} is the excitation spectrum, f being the vector [4]. See Figure 3.4.

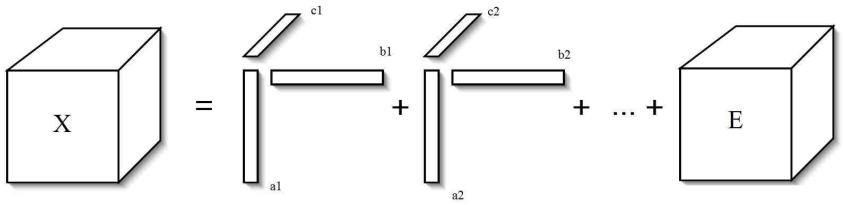


Figure 3.4: A PARAFAC Model for the X Data Cube.

As previously mentioned in Section 3.2.1, this method is solely for the manipulation of the axis to aid in the discrimination of features. Additional pattern recognitions algorithms are necessary to for the detection of a chemical.

3.2.3 Cluster Analysis Methods

Clustering is a method that could be applied after PCA or PARAFAC. Clustering will enable further classification in a single or in multiple planes, where PCA and PARAFAC may have difficulty. Clustering has many methods for determining the cluster size and distance between two clusters. The main concept of cluster analysis consists of grouping objects that are close together in distance into different classes [30].

Typical methods of examining distance on a plane are Euclidean distance and Mahalanobis distance. The simplest of the two is to implement the Euclidean distance which is just the traditional distance measure defined in

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}. \tag{3.11}$$

Mahalanobis distance differs fundamentally by using the statistics of the data (in particular, the covariance). The distance between points is then used to form the clusters using methods such as single linkage method, complete linkage clustering and average linkage clustering. These are schemes that utilize distance measurements to determine the final clustering of data.

Chapter 4

A Dempster-Shafer Theoretic Method for the Spectroscopy Detection Problem

4.0.1 General Discussion

In this chapter, we briefly discuss the basics of *Dempster-Shafer (DS) theory* and highlight its differences with the Bayesian approach. We then propose how we develop a DS theoretic method to address our spectroscopy detection problem. Our approach consists of comparing the feature vectors of an *unknown* sample (i.e., a sample containing unknown analytes) with that of a *prototype* sample (i.e., a sample from the database containing known chemical analytes) to estimate the concentrations associated with the unknown sample. For this comparison, we use the correlation coefficient between the feature vectors of the unknown and prototype samples. When an unknown sample is compared to all the prototype samples, we get a correlation vector corresponding to the given unknown sample. We then propose a method whereby each correlation vector is transformed into a DS theoretic model. With the different types of evidence sources, we obtain multiple DS theoretic models which are then fused (using a DS theoretic fusion strategy, such as the *Dempster's combination rule (DCR),* to arrive at a decision regarding the type of analytes in the unknown sample.

4.0.2 DS Theory and the Bayesian Approach

DS theory is known for its ability to handle imperfect data in an effective and more intuitive manner [35]. DS theory, otherwise known as *belief theory,* can be thought of as a generalization of probability theory [**?** 36]. One main difference between the two methods is, while the Bayesian approach assigns probabilities according to axioms of probability [37, 38], the axiom of additivity is relaxed in DS theory. For example, consider a case with three possible outcomes $\Theta = \{X, Y, Z\}$. In the Bayesian approach, $Pr(X) + Pr(Y) + Pr(Z) = 1$, so that $Pr(X^c) = Pr(Y) + Pr(Z)$, where X^c is the complement of the proposition X. This is referred to as the axiom of additivity. DS Theory does not impose this requirement in the way it assigns "supports" to propositions. In addition, DS theory allows "supports" to be assigned to the complete power set of possibilities. Contrast this with probability theory where the probabilities assigned to the "singletons" (in the example, $\{X\}$, $\{Y\}$ and $\{Z\}$) *completely* determines the probabilities associated with other propositions (e.g., $Pr(\{X \cup Y\}) = Pr(\{X\}) + Pr(\{Y\})$). However, DS theory provides an easy transition to-and-from probability, a feature that sets it apart from other frameworks for uncertainty handling (e.g., fuzzy sets).

4.0.3 Uncertainty in Evidence

When a decision is to be rendered, there are two main types of uncertainties one may encounter: *aleatory uncertainty* and *epistemic uncertainty.* Aleatory uncertainty refers to a system behaving in a "random" way. This is accounted for by using historic data otherwise known as the *frequentist approach* [9]. Epistemic uncertainty is caused by a lack of knowledge which is reduced through increased understanding [9, 8]. Traditionally, epistemic uncertainty has been accounted for by Bayesian methods, but this requires historic information to form a probability of the event [9]. When historic information is not readily available, it is modeled through Laplace's inference

by implementing a uniform distribution, referred to as the *principle of insufficient reason* [9].

In the above example, if a probability was assigned to $\{X\}$ by some past information, we would be forced to equally allocate the rest of the probability to $\{Y\}$ and $\{Z\}$. This is due to the axioms of probability and the lack of historical information, thus constraining us to rely on Laplace's inference. When the axiom of additivity is not used as a constraint, we may encounter the mass's inability to sum to one [9]. This is associated with the sources reporting the same or conflicting information. DS Theory is a major tool which allows us to decipher uncertainty and permits data fusion techniques to reduce uncertainty from imperfect data (such as, source information that is conflicting or sources reporting similar information) [39, 40, 9].

4.0.4 Basic Notions of DS Theory

The set of mutually exclusive and exhaustive propositions of interest is referred to as the *frame of discernment (FOD).* We take the FOD Ω to be a finite set (i.e., $\Omega = \{\theta_1, \ldots, \theta_n\}$). The *basic probability assignment (BPA),* otherwise referred to as a *basic belief assignment* or *mass function* is a function $m : 2^\Omega \to [0, 1]$, where 2^Ω is the power set of Ω, such that

$$m(\emptyset) = 0; \quad \sum_{A_i \subseteq 2^\Omega} m(A_i) = 1. \tag{4.1}$$

In order to develop a mass function, evidence is required. Evidence is typically defined subjectively by experts. This evidence is then fused to dynamically update the mass function to aid in the reduction of uncertainty and to redistribute mass to the propositions. Perhaps the most common method employed for evidence fusion in

DS theory is the *Dempster's combination rule (DCR)* [41]:

$$m(A_i) = \frac{\sum_{A_p \cap A_q = A_i} m_1(A_p)\, m_2(A_q)}{\sum_{A_p \cap A_q = \emptyset} m_1(A_p)\, m_2(A_q)}, \tag{4.2}$$

where the evidence provided by the mass functions m_1 and m_2 are combined to get the fused mass function m. This is usually denoted as

$$m = m_1 \oplus m_2. \tag{4.3}$$

The DCR is commutative and associative, thus making it convenient to fuse multiple sources of evidence as

$$m = m_1 \oplus m_2 \oplus \ldots \oplus m_M. \tag{4.4}$$

For our purposes, we view the outputs of various feature extraction techniques as sources of evidence or "experts". These evidence sources are then combined for evidence fusion used in decision-making.

4.1 Evidence Sources

In Chapter 2, we developed fluorescent and absorption spectra by the generation of the spectroscopy database. Using these spectra, we extract four different attribute or feature vectors associated with each fluorescent and absorption spectra. These feature vectors are then used as evidence sources to be later fused (or combined) using DS theoretic techniques. See Figure 4.1. We now describe the four types of feature vectors generated by the evidence sources: discrete cepstral coefficients, energy content, spectral differential, and a matched filter output.

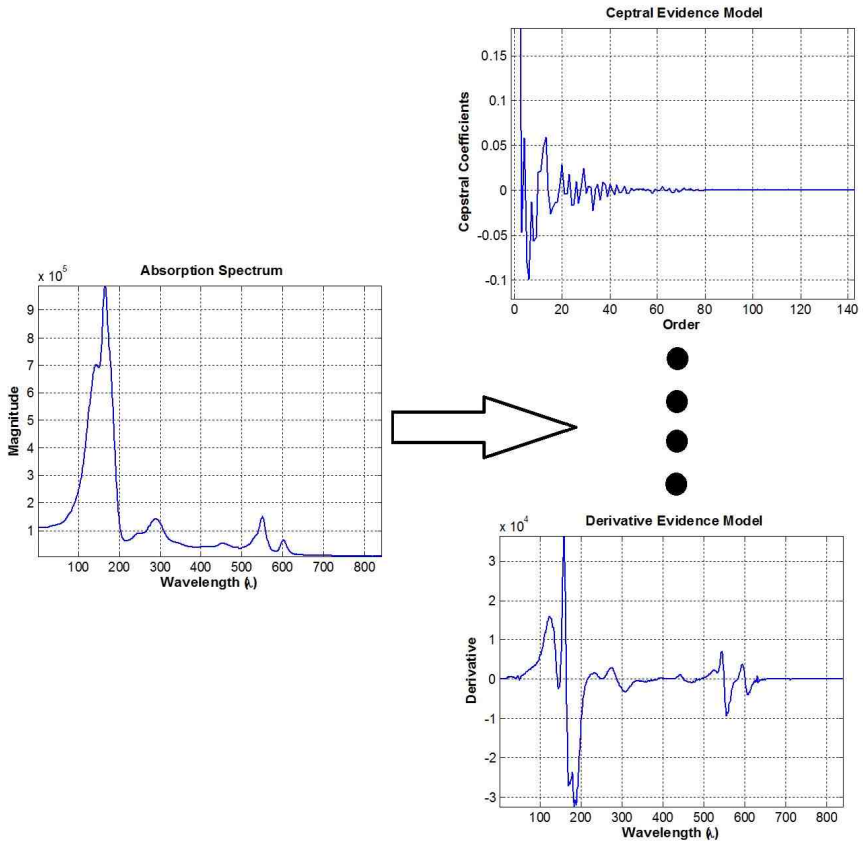


Figure 4.1: Various Attribute or Feature Vectors Extracted from an Absorption Spectrum.

4.1.1 Discrete Cepstral Coefficients

Due mainly to the noise robustness, linear predictive coding (LPC) coefficients associated with all-pole models of the spectrum are typically used in communications to capture the spectral envelope information.

One disadvantage of the LPC model is that the predicted envelope spectrum tends to overshoot and undershoot spectral regions when there are sudden changes [42]. Discrete cepstrum based methods were designed to combat this issue and these methods have been refined for enhanced computational speed [42]. We capture the envelope of the spectrum magnitude by the cepstrum based estimation scheme described in [43]. The work in [43] utilizes a regularization technique to prevent the

"ill conditioning" problem associated with this method. We follow the same strategy in our work.

The relationship between the real cepstrum coefficients and the spectral magnitude is given by

$$\log|S(f;c)| = c_0 + 2\sum_{i=1}^{p} c_i \cos(2\pi f i). \tag{4.5}$$

With regularization, the spectrum estimation error is given by

$$\epsilon_r = \sum_{k=1}^{L} \| \log a_k \log|S(f;c)| \|^2 + \lambda\, R[S(f;c)], \tag{4.6}$$

where

$$R[S(f;c)] = \int_{f=1/2}^{1/2} \left[\frac{d}{df}\log|S(f;c)|\right]^2 df. \tag{4.7}$$

The error is minimized in the least-squares sense by

$$\mathbf{c} = (\mathbf{M}^T\mathbf{M} + \lambda\mathbf{R})^{-1}\mathbf{M}^T\mathbf{a}, \tag{4.8}$$

where

$$\begin{aligned}
\mathbf{c} &= \begin{bmatrix} c_0 & c_1 & \cdots & c_p \end{bmatrix}^T; \\
\mathbf{a} &= \begin{bmatrix} \log(a_1) & \log(a_2) & \cdots & \log(a_L) \end{bmatrix}^T; \\
\mathbf{R} &= 8\pi^2\, diag\begin{bmatrix} 0 & 1^2 & 2^2 & \cdots & p^2 \end{bmatrix}; \\
\mathbf{M} &= \begin{bmatrix} 1 & 2\cos(2\pi f_1) & 2\cos(2\pi f_1 2) & \cdots & 2\cos(2\pi f_1 p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 2\cos(2\pi f_L) & 2\cos(2\pi f_L 2) & \cdots & 2\cos(2\pi f_L p) \end{bmatrix}.
\end{aligned} \tag{4.9}$$

4.1.2 Energy Content

Energy content at various regions within the spectrum has the ability to provide information about the concentrations of chemicals. A filter bank design was employed

to generate a feature vector that captures the energy content at localized regions from the absorption and emission spectra. We used a very simple triangle filter bank design, which is typically used for speech processing. We drew inspiration from this model to implement the linear portion of the filter bank under $1\,kHz$. We used the method in [44] to design the linear triangle filters. Figure 4.2 shows a design of the linear filter bank with 29 cascaded filters for our spectrum analysis of energy content. The feature vector generated captures the energy content, which contains coefficients associated from the filter bank. Each coefficient generated corresponds to a single triangle filter that is multiplied with the spectral signal. Therefore, the example provided below of the 29 cascaded triangle filters will yield 29 coefficients that make up the feature vector.

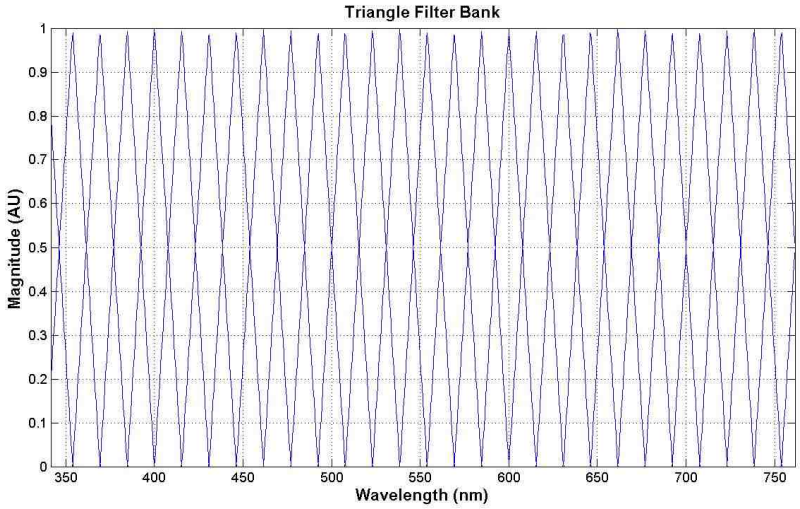


Figure 4.2: Triangle filter bank.

4.1.3 Spectral Differential

Differential of the spectral intensity of the signal has been used for many years to provide more valuable information than the intensity spectrum itself [45, 10]. This data is typically used to extract information when there are multiple peaks near each other to highlight the trough between the adjacent peaks [10]. We use the forward

Euler approximation for the derivative:

$$\frac{dy}{d\lambda} = \frac{y_{i+1} - y_i}{\Delta\lambda}, \tag{4.10}$$

where y is the intensity of the spectrum and λ is the wavelength.

4.1.4 Matched Filter Output

Matched filters are commonly used in telecommunications and radar systems to detect an unknown deterministic signal by correlating it to a set of known signals. We apply the same concept of detecting an unknown deterministic spectrum by correlating it to our known set of prototype spectra. We use matched filtering as a method of evidence for chemical spectra by designing filters to develop maximized signal-to-noise ratio (SNR) values for each spectrum as a mode of evidence.

Let us consider the output of a linear filter with an impulse response h corresponding to the input x:

$$y[n] = \sum_{k=-\infty}^{\infty} h[n-k]\, x[k]. \tag{4.11}$$

For our purposes, we assume that the unknown spectrum signal is a member of the set of prototype spectra and it has been distorted by Gaussian white noise. With this noise model in place, we designed matched filters for each prototype spectrum signal within the set. Each matched filter is optimum in the sense that it maximizes the signal-to-noise ratio (SNR) with respect to additive noise. The impulse response of the matched filter is given by

$$h = \frac{1}{\sqrt{s^H R_v^{-1} s}} R_v^{-1} s, \tag{4.12}$$

where s is the prototype spectrum signal and R_v is the covariance matrix associated with the noise v.

We apply the unknown noise distorted spectrum signal to the database of prototype filters in obtaining optimized filter outputs. These outputs are calculated by convolving the unknown signal with the prototype set of spectra. The optimized filter output y can also be thought of as the inner product of the filter, h, and the signal with additive noise, $s + v$:

$$y = \sum_{k=-\infty}^{\infty} h^*[k]\, x[k] = h^H x = h^H s + h^H v = y_s + y_v. \tag{4.13}$$

The associated SNR can be expressed as

$$\text{SNR} = \frac{|y_s|^2}{E\{|y_v|^2\}} = \frac{|h^H s|^2}{E\{|h^H v|^2\}} = \frac{|h^H s|^2}{h^H R_v h}. \tag{4.14}$$

Using the SNR as the objective function, we can evaluate the strongest prototype matches by calculating the SNR produced by each designed matched filter and evaluating it with its prototype spectrum. This will produce an array of N SNR values associated to N spectra. The higher the SNR value that is achieved by using filter h_i, the more likely the unknown spectrum is it is the unknown spectrum s_i. We use this array of normalized SNR values as our matched filter output feature vector.

4.2 Proposed DS Theoretic Model

The DS theoretic notion of mass captures the "support" allocated to each proposition (which may consist of singleton and composite propositions). Contextual considerations (e.g., accuracy, source reliability, source conflicts, etc.) all play a role in determining the mass to be allocated [9].

4.2.1 Evidence Models to Correlation Coefficients

The prototype samples were processed using the aforementioned methods to generate *prototype feature vectors.* These feature vectors corresponding to the complete

prototype fluorescence spectrum data can be arranged into a matrix form to generate a *prototype feature template:*

$$\mathbf{T}_{ij} = \begin{pmatrix} X_{11} & X_{21} & \cdots & X_{T1} \\ X_{12} & X_{22} & \cdots & X_{T2} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1N} & X_{2N} & \cdots & X_{TN} \end{pmatrix}, \tag{4.15}$$

where the subscript i in $\mathbf{T}_{ij}$ identifies one of the two types of spectrum data (i.e., absorption or emission at a specific excitation wavelength) and the subscript j in $\mathbf{T}_{ij}$ identifies the one of four types of features (i.e., discrete cepstral coefficients, energy content, spectral differential, and matched filter output). Here, N denotes the number (or size) of the prototype spectrum data set and T denotes the length of feature vector.

This feature template is associated with the "ideal" fluorescence spectroscopy data without any additional spectrum perturbations having been introduced. To determine the prototype spectrum which "best matches" an unknown chemical, we employ the correlation coefficient which provides a simple yet effective means of statistically describing the relationship between two sets of data. For our purpose, given an unknown spectrum, we generate its four feature vectors and then compute the associated correlation coefficients between these feature vectors of the unknown spectrum and feature vectors of the prototypes in $\mathbf{T}_{ij}$. A normalized correlation vector generated thus takes the form

$$\mathbf{V} = \begin{bmatrix} V_1 & V_2 & \cdots & V_N \end{bmatrix}^T, \tag{4.16}$$

where V_i, $i = 1, \ldots, N$, denotes the positive normalized correlation coefficient between a feature vector corresponding to the unknown specimen and the i^{th} prototype spectrum in the data set.

4.2.2 Correlation Coefficients to DS Mass

In this section, we explore how a DS model can be fitted to represent the normalized correlation coefficient vector so that it may be treated as a single evidence source. Regarding this task, not only must we obtain a valid DS model as described in 4.1, but more importantly, we must ensure that the model captures potential conflicts reasonably well among prototype spectra that are "competing" for a match with the unknown specimen. For example, a correlation vector with multiple values of 1 would indicate that multiple prototype compounds are perfect matches for the unknown specimen.

Weighting Matrix

To proceed, let us generate the following matrix associated with the normalized correlation vector in 4.16:

$$\begin{aligned}\Delta\mathbf{V} &= \mathbf{J}_{NN}\mathbf{D}_N - \mathbf{D}_N\mathbf{J}_{NN} \\ &= \begin{pmatrix} (V_1 - V_1) & (V_2 - V_1) & \cdots & (V_N - V_1) \\ (V_1 - V_2) & (V_2 - V_2) & \cdots & (V_N - V_2) \\ \vdots & \vdots & \ddots & \vdots \\ (V_1 - V_N) & (V_2 - V_N) & \dots & (V_N - V_N) \end{pmatrix},\end{aligned} \tag{4.17}$$

where $\mathbf{J}_{NM}$ denotes the $N \times M$ matrix with each entry being 1 and $\mathbf{D}_N = diag\,[V_1, V_2 \cdots, V_N]$ denotes the diagonal matrix with the diagonal entries being $\{V_1, V_2, \cdots, V_N\}$.

This $\Delta\mathbf{V}$ matrix examines how each correlation coefficient's magnitude deviates relative to all other elements in the correlation coefficient vector in 4.16. The difference taken between the correlations coefficients can be conceptually thought of as a "distance" and the sum of the columns of the $\Delta\mathbf{V}$ is the "sum of distances" relative to each element, V_i. However, the elements in the $\Delta\mathbf{V}$ matrix still lack information about the overall magnitude of the elements in $\mathbf{V}$.

For example, consider the case where $V_1 = 0.4$, $V_2 = 0.3$, $V_3 = 0.7$, and $V_4 = 0.6$, yields $(V_1 - V_2) = (V_3 - V_4) = 0.1$, irrespective of the fact that the values V_3 and V_4 are significantly higher (thus more correlated with the unknown specimen) than the values V_1 and V_2 (thus less correlated with the unknown specimen).

To capture the overall magnitude of each element in $\mathbf{V}$, we utilize a weighting strategy as

$$\begin{aligned} \Delta\mathbf{W} &= \Delta\mathbf{V} \circ \mathbf{J}_{NN}\mathbf{D}_N \\ &= \begin{pmatrix} V_1\Delta V_{11} & V_2\Delta V_{21} & \cdots & V_N\Delta V_{N1} \\ V_1\Delta V_{12} & V_2\Delta V_{22} & \cdots & V_N\Delta V_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ V_1\Delta V_{1N} & V_2\Delta V_{2N} & \ldots & V_N\Delta V_{NN} \end{pmatrix}, \end{aligned} \tag{4.18}$$

where $\circ$ denotes the matrix Hadamard product, $\Delta V_{ij} = V_i - V_j \in [-1, +1]$, $\forall i, j \in \overline{1, N}$. The columns of $\Delta\mathbf{W}$ compare the distance of an element V_i with all the elements in $\mathbf{V}$ and weights V_{ij} with V_i . This weighting of V_i could be thought of as an indication of the "strength" of the corresponding prototype being a match.

Let us take the same example as before: $V_1 = 0.4$, $V_2 = 0.3$, $V_3 = 0.7$, and $V_4 = 0.6$, would yield $V_1\Delta V_{12} = (0.4)(0.4 - 0.3) = 0.04$ and $V_3\Delta V_{34} = (0.7)(0.7 - 0.6) = 0.07$, which clearly emphasize the correlation coefficients possessing a higher magnitude.

This example demonstrates the need of having both the "strength" and "distance" components. However, it is important to observe how these components interact with each other to generate $\Delta\mathbf{W}$. Both "strength" and "distance" are required to be high, in order to yield a high value in $\Delta\mathbf{W}$. When one or both components are low, low values in $\Delta\mathbf{W}$ are generated. So, $\Delta\mathbf{W}$ captures not only how the elements in $\mathbf{V}$ are distributed, but also their relative strengths.

For example, consider the case where, $V_5 = 1.0$, $V_6 = 0.98$, $V_7 = 1.0$, and $V_8 = 0.1$. Then $V_5(\Delta V_{56}) = 1(1 - 0.98) = 0.02$ and $V_7(\Delta V_{78}) = 1(1 - 0.1) = 0.90$. This shows that when a high correlation value, V_5, is compared to another high correlation value,

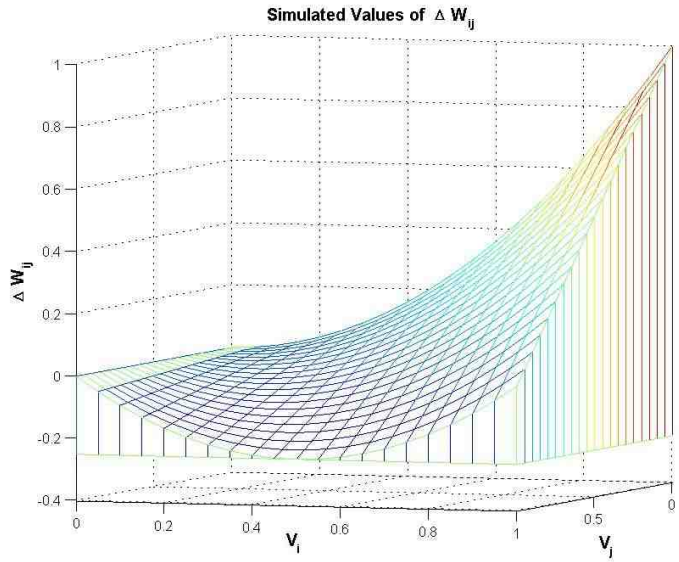


Figure 4.3: Evaluation of the Range of $\Delta W_{ij} = V_i \Delta V_{ij}$

V_6, a low value is generated for the associated element in $\Delta\mathbf{W}$. However, when a high correlation coefficient value, V_7, is compared to a low correlation value, V_8, a high value is generated for the associated element in $\Delta\mathbf{W}$. A simulation is provided in Figure 4.3 to demonstrate the ranges and attributes of the $\Delta W_{ij} = V_i \Delta V_{ij}$ function. The ranges should be noted, where the maximum achievable value is 1 and the minimum achievable value is .25. The attributes that are worth noting are the exponential reward of obtaining high "strength" and "distance", as well as the exponential penalty of obtaining a low "strength" and "distance".

Column Weights Vector

Each column of $\Delta\mathbf{W}$ evaluates a specific V_i against the entire set of correlation coefficients within $\mathbf{V}$. This captures the strength of each correlation coefficient V_i relative to all the other correlation coefficients in the correlation coefficient vector, $\mathbf{V}$. Thus, the summation of the column vectors informs us how each element in $\mathbf{V}$ is different from the other elements and how strongly it matches the unknown specimen. Let us refer to the sum of all the elements in a column vector as its *column weight*.

The *column weights*, C_i, are calculated as

$$\mathbf{C} = \begin{bmatrix} C_1 & C_2 & \cdots & C_N \end{bmatrix}^T = (\mathbf{J}_{1N}\Delta\mathbf{W})^T, \tag{4.19}$$

where $C_i \in [-(N-1)/4, (N-1)], \forall\ i \in \overline{1, N}$. These column weights allow us to identify the rival correlation coefficients that are legitimately competing to be a match for the unknown sample. To demonstrate this, take the $N = 4$ case.

Case 1: Let $\boldsymbol{V} = [1,\ 0,\ 0,\ 0]^T$. This indicates that V_1 constitutes a perfect match with no competing matches. Note that

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \implies \mathbf{C} = [3,\ 0,\ 0,\ 0]^T.$$

Notice how $\mathbf{C}$ puts the maximum weight on V_1.

Case 2: $\boldsymbol{V} = [1,\ 1,\ 0,\ 0]^T$. This indicates that both V_1 and V_2 are competing for being a match. Note that

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} \implies \mathbf{C} = [2,\ 2,\ 0,\ 0]^T.$$

Notice how $\mathbf{C}$ puts equal weights for V_1 and V_2. This weight is less than what V_1 is given in Case 1 because of the uncertainty generated by having two competing prototypes.

Case 3: $\boldsymbol{V} = [1,\ 1,\ 0.9,\ 0]^T$. This indicates 3 prototypes competing for being a match, but with V_3 being the slightly weaker candidates. Note that

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & -0.09 & 0 \\ 0 & 0 & -0.09 & 0 \\ 0.1 & 0.1 & 0 & 0 \\ 1 & 1 & 0.81 & 0 \end{pmatrix} \implies \mathbf{C} = [1.1,\ 1.1,\ 0.63,\ 0]^T.$$

Notice how $\mathbf{C}$ distributes its weights among V_1, V_2 and V_3.

Case 4: $\boldsymbol{V} = [1,\ 1,\ 0.9,\ 0.1]^T$. In this case,

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & -0.09 & -0.09 \\ 0 & 0 & -0.09 & -0.09 \\ 0.1 & 0.1 & 0 & -0.08 \\ 0.9 & 0.9 & 0.72 & 0 \end{pmatrix} \implies \mathbf{C} = [1.0,\ 1.0,\ 0.54,\ -0.26]^T.$$

Notice how $\mathbf{C}$ de-emphasizes V_4 which is in no position to compete with V_1, V_2, and V_3.

Identification of Potential Focal Elements

The vector $\mathbf{C}$ plays a critical role in our development because we can utilize the column weights C_i to identify the potential focal elements of the DS theoretic model we propose. To explain, take the "extreme case" where no conflict is present,

$$\mathbf{V} = [1,\ 0,\ 0,\ \cdots,\ 0]^T, \tag{4.20}$$

which yields

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \end{pmatrix} \implies \mathbf{C} = [(N-1),\, 0,\, 0,\, \cdots,\, 0]^T. \tag{4.21}$$

When observing this extreme case, we may say that, as the column weight approaches $(N-1)$, the uncertainty regarding the corresponding prototype being a match decreases. Conversely, the higher the difference between $(N-1)$ and a column weight, the less likely the corresponding prototype is the correct match. When the column weight is zero or negative, we become more certain that the corresponding prototype is not a match.

The inability of the prototype associated to V_i competes against the other prototypes causing the corresponding C_i to be negative. This is exactly what occurs in Case 4 in Section 4.2.2, where C_4's magnitude was not high enough to compete against the others. So, in our model, we neglect the prototypes corresponding to non-positive values of C_i, thus preventing them from becoming focal elements in our DS theoretic model. This strategy restricts the domain of candidate prototypes that are potential focal elements. This avoids having to assign DS masses to "unnecessary" or "weak" candidates.

We propose two methods of reducing the domain of candidates to be focal elements:

- *Column weights associated method:* This is described above where the focal elements are restricted by the criterion associated with the column weights vector, $\mathbf{C}$. The criteria to determine the number of focal elements, P, is

$$P = \sum_{i=1}^{N} T_i, \tag{4.22}$$

where

$$T_i = \begin{cases} 1, & \textit{if } \boldsymbol{C}_i > 0; \\ 0, & \textit{if } \boldsymbol{C}_i \leq 0. \end{cases} \tag{4.23}$$

The vector $\mathbf{T} = [T_1, T_2 \cdots, T_N]$ is an indicator vector which identifies the candidates that will later constitute the focal elements of our DS theoretic model. Let us define a *mass measure vector* $\mathbf{H} = [H_1, H_2 \cdots, H_N]$ as

$$H_i = \frac{C_i + |C_i|}{2}. \tag{4.24}$$

Clearly, $H_i \in [0, (N-1)]$, $\forall\ i \in \overline{1, N}$. Note that, $\mathbf{H}$ is identical to $\mathbf{C}$, except that it substitutes 0 for all the non-positive elements of $\mathbf{C}$.

- *Column+row weights associated method:* This method restricts the candidate selection even further by placing constraints using both the column weights vector and the *row weights vector,* R_i, which is defined as

$$\mathbf{R} = \begin{bmatrix} R_1 & R_2 & \cdots & R_N \end{bmatrix}^T = \Delta \mathbf{W}\, \mathbf{J}_{N1}. \tag{4.25}$$

The criterion used in this column+row weights associated method uses the indicator $\mathbf{T} = [T_1, T_2, \cdots, T_N]^T$, where

$$\mathbf{T}_i = \begin{cases} 1 & \textit{if } R_i \leq 0 \textit{ and } C_i > 0; \\ 0, & \textit{otherwise.} \end{cases} \tag{4.26}$$

Correspondingly, the number of focal elements is given by

$$P = \sum_{i=1}^{N} T_i, \tag{4.27}$$

which generates the following *mass measure vector:*

$$\mathbf{H} = \mathbf{C}^T diag(\mathbf{T}). \tag{4.28}$$

DS Mass Allocation

The proposed mass measure vector contains information regarding the strength of the propositions that warrant being focal elements. Crucial variables for the allocation of the DS theoretic masses are the mass measure vector elements (H_i), the amount of focal elements (P), and the length of the mass measure vector (N). These variables (P, N, H_i) adjust in accordance to each unique scenario.

The method that is proposed for determining the DS Mass allocation is designed around the previous discussed *extreme case scenario* (ECS). The amount of scenarios were limited since the incorporation of every scenario would be a monumental task and computationally complex. However, the design around the ECS method allows for the generalization to all other cases. Although the design is constrained to fit this specific paradigm of the ECS, we show through numerical examples in the later section that the paradigm has validity beyond this special case.

In order to incorporate each possible scenario to properly depict appropriate DS theoretic masses for various values of N and P, we define three types of functions *minimum uncertainty ratio* $Mu(P, N)$, *maximum mass measure*, $Mm(P, N)$, and *total mass measure,* $Tm(P, N)$. The designed quantities mentioned above are formulated to logically fit the DS theoretic framework around the ECS of having only ones and zeros as correlation coefficients.

Maximum Mass Measure: The maximum mass measure is achieved only when the correlation coefficients V_i, are restricted to the values 1 s and 0 s only (the ECS). Maximum mass measure informs us of the maximum possible mass that is achievable when all the elements in the mass measure vector are summed. The achievable maximum mass measure, for given values of N and P, will always be the same,

regardless of how $\mathbf{V}$ is distributed. In Section A.1, we show that

$$Mm(P, N) = -P^2 + NP, \tag{4.29}$$

where N is the correlation coefficient vector length, and P is the number of focal elements that will be assigned.

Minimum Uncertainty Ratio: When the maximum mass measure is achieved for a specific combination of N and P values, the ECS occurs. We are then provided with P conflicting perfect matches (obtaining a perfect match constitutes optimal "strength" and "distance"). Thus due to the conflict an uncertainty must be accounted and as more conflict is presented, the uncertainty should increase.

The uncertainty assignment is consequently formulated to contain meaning regarding the relationship between the total amount of candidates, N, and the reduced set of focal elements P. Every element that is assigned a one will be a focal element, and every zero that is assigned will be a proposition of zero mass measure. This yields P ones, and $(N-1-P)$ zeros. Examples of coherent uncertainty assignments regarding the relationship between N and P are demonstrated in Table 4.1.

Table 4.1: Assignment of Uncertainty with Regards to P and N.

P	N	$M(\Theta)$
1	128	0
32	128	.25
64	128	.50
96	128	.75
128	128	1.0

In Table 4.1, we can note when the $\Delta\mathbf{W}$ matrix reduces all the candidates down to one focal element, no conflicting information is presented, consequently $M(\Theta) = 0$. However, if we are only able to reduce half of the candidates shown in the example when $N = 128$ and $P = 64$ then, $M(\Theta) = 0.50$. Therefore, based on this unique ones and zeros, case our uncertainty should be approximately set to $M(\Theta) = \frac{P}{N}$. This ratio, $M(\Theta) = \frac{P}{N}$, reflects the calculated values set in Table 4.1, except for when

$P = 1$. In order to account for all ranges of P for the ones and zeros case, we can approximate the assignment of the uncertainty, $M(\Theta)$, by the following ratio,

$$\frac{P-1}{N-1}. \tag{4.30}$$

Total Mass Measure: Applying the maximum mass measure and the minimum uncertainty we obtain the total mass measure function based around this unique ones and zeros case. Let us calculate $Tm(P,N)$ by using $Mm(P,N)$, and the minimum uncertainty ratio defined in equation 4.30 by

$$\begin{aligned}
\frac{Tm(P,N) - Mm(P,N)}{Tm(P,N)} &= \frac{P-1}{N-1} \\
\frac{1}{Tm(P,N)} &= (-1 + \frac{P-1}{N-1})\frac{1}{-Mm(P,N)} \\
\frac{1}{Tm(P,N)} &= (\frac{P-N}{(N-1)(P^2 - NP)} \\
Tm(P,N) &= (N-1)P.
\end{aligned} \tag{4.31}$$

Equation 4.31 can be conceptualized as the maximum possible attainable mass given the number of focal elements, P. Since the diagonal elements in $\Delta\mathbf{W}$ will always be zero, we do not want to account for these elements in the total mass measure. Note that in equation 4.31, N-1 is formulated to remove these diagonal elements for the calculation of the total mass measure.

Recall in Subsection 4.2.2, we presented various cases that demonstrate the effects of competing values on the column weights C_i. The subsection demonstrated that when increased competition between correlation coefficient values are presented, the column weights are reduced. This reduction of column weight is dependent on the adjacent competing correlation coefficient values.

The total mass measure treats adjacent competing values independently, and consequently are not affected by adjacent competing column vectors. Therefore, each independent column produces a weight of $(N-1)$, defined in Equation 4.25. Thus,

P independent columns (corresponding to each focal element), yields a total mass measure of $(N-1)P$.

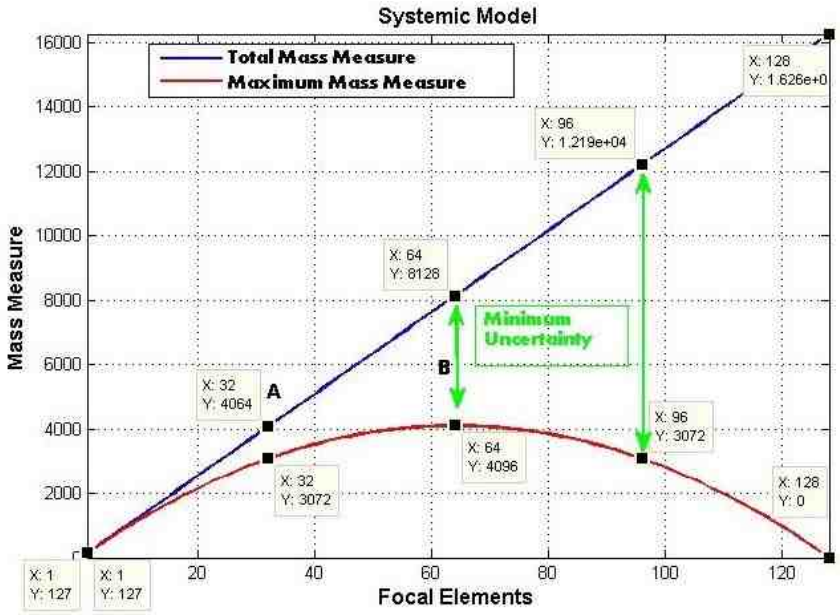


Figure 4.4: Mass Measure in Relation to Number of Focal Element

Systematic Example Model: In Figure 4.4, the *Maximum Mass Measure* function (red), and total mass measure (blue), are modeling a mass measure vector with 128 elements, over the entire range of focal elements, P. The *Maximum Mass Measure* function is a parabola that is concave down and symmetric around the vertex at $P = N/2$ (Location B). The parabola's shape is associated to how ΔW elicits competition. Recall, Subsection 4.2.2 for cases 1 and 2: two competing $V_i = 1$'s can reduce an individual's column vector's weight. However, these cases demonstrate how column weights have the potential to be redistributed to adjacent columns. In Figure 4.4, the redistribution of weights to adjacent columns causes an increase in $Mm(P, N)$ between the interval of $1 < P > N/2$ (Locations A and B). When $P > N/2$ occurs, there are too many competitors and $\mathbf{\Delta W}$ is unable to reduce the candidates causing $Mm(P, N)$ to decrease. Therefore, Figure 4.4 demonstrates that the *Maximum Proposal Weight*

dynamically changes as we introduce additional V_i's, which meet the focal element criteria.

In Figure 4.4, there are five sets of labels that are shown. These labels contain the coordinates X and Y, where X is P, and Y is the mass measure. These labels are presented to demonstrate the assigned mass measure regarding its relationship to the uncertainty for the unique ones and zeros case. Table 4.2 demonstrates the uncertainty assignment regarding the following labels in Figure 4.4 using the maximum mass measure and total mass measure.

Table 4.2: Assignment of Uncertainty with Regards to Systemic Weights.

P	N	$Mm(P,N)$	$Tm(P,N)$	$M(\Theta)$
1	128	127	127	0.0
32	128	3072	4064	0.244
64	128	4096	8128	0.496
96	128	3072	12192	0.748
128	128	0	16256	1.0

The *Maximum Mass Measure* is only obtained when the correlation coefficients are solely assigned ones and zeros. When the maximum mass measure is not obtained, a residual mass measure, $Residual = Mm(P) - \sum_{i=1}^{N} \mathbf{H}$, is formed and relocated to the uncertainty. The residual mass measure occurs when the sum of H is less then $Mm(P)$.

DS Masses Defined: Thus, we calculate the uncertainty and adjust the mass measure vector's elements to DS masses as

$$M(A) = \begin{cases} 1 - \left(\dfrac{\sum_{i=1}^{N} \mathbf{H}_i}{\frac{i=1}{Tm(P,N)}} \right), & \text{for } A = \Theta; \\ H_i \left(\dfrac{\frac{1-m(\Theta)}{N}}{\sum_{i=1} H_i} \right) & \text{for } A = H_i, \end{cases} \tag{4.32}$$

where $\Theta = \{V_1, V_2, \cdots, V_N\}$ is the the FoD consisting of the singletons H_i, $i \in \overline{1, N}$.

General Examples for Arbitrary Correlation Coefficients

General cases of arbitrary correlation coefficient values (Y and Z) are provided to demonstrate the assignment of DS masses and uncertainty. The three examples below are demonstrated with the $N = 4$ case.

Example i: $\boldsymbol{V} = [Y, Y, Y, Y]^T$. This example demonstrates the special case when all the correlation coefficients are equivalent. In this case, the weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [0, 0, 0, 0]^T.$$

Implementing the *column weights associated method*, we obtain zero focal elements ($P = 0$), and a *mass measure vector*,

$$\mathbf{H} = [0, 0, 0, 0]^T.$$

The assigned DS masses and uncertainty are

$$\mathbf{M} = \begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

Notice, that all the prototypes are equally competitive (neither prototype is a better match for the unknown specimen). This produces a scenario of maximum conflict generating an uncertainty, $M(\Theta = 1)$.

Example ii: $\boldsymbol{V} = [Y, Y, Y, Z]^T$. This example presents the scenario when all the correlation coefficients are equivalent, except for one, where $Z > V$. In this case, the weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & 0 & Z(Z-Y) \\ 0 & 0 & 0 & Z(Z-Y) \\ 0 & 0 & 0 & Z(Z-Y) \\ Y(Y-Z) & Y(Y-Z) & Y(Y-Z) & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [Y(Y-Z),\, Y(Y-Z),\, Y(Y-Z),\, 3Z(Z-Y)]^T.$$

Implementing the *column weights associated method*, we obtain a single focal element ($P = 1$), and a *mass measure vector*,

$$\mathbf{H} = [0,\, 0,\, 0,\, 3Z(Z-Y)]^T.$$

The assigned DS masses and uncertainty are

$$\mathbf{M} = \begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{3Z(Z-Y)}{1(3)} \\ 1 - \frac{3Z(Z-Y)}{1(3)} \end{pmatrix},$$

Notice, the mass assignment for $M(V_4)$ and $M(\Theta)$ are dependent on the "distance" between Z and Y. As the "distance" between Z and Y increases, the mass of $M(\Theta)$

decreases and the mass of $M(V_4)$ increases. The maximum "distance" achievable is when $Z = 1$ and $Y = 0$, yielding $\mathbf{M} = [0,\, 0,\, 0,\, 1,\, 0]^T$, thus representing no conflict. However, if the "distance" between Z and Y decreases, the mass of $M(\Theta)$ increases, thus causing the mass of $M(V_4)$ to decrease. The minimum achievable "distance" occurs when Z approaches Y, where $Z \approx Y$. As we approach this minimum distance, we converge to the scenario of maximum conflict (as shown in *Example i*), where $\mathbf{M} = [0,\, 0,\, 0,\, 0,\, 1]^T$.

Example iii: $\boldsymbol{V} = [Y,\, Y,\, Z,\, Z]^T$. This presents half the correlation coefficients in vector as one value and the other half as another competing value, where $Z > Y$. In this case, the weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0 & Z(Z-Y) & Z(Z-Y) \\ 0 & 0 & Z(Z-Y) & Z(Z-Y) \\ Y(Y-Z) & Y(Y-Z) & 0 & 0 \\ Y(Y-Z) & Y(Y-Z) & 0 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [2Y(Y-Z),\, 2Y(Y-Z),\, 2Z(Z-Y),\, 2Z(Z-Y)]^T.$$

Implementing the *column weights associated method*, we obtain a two focal elements ($P = 2$), and a *mass measure vector*,

$$\mathbf{H} = [0,\, 0,\, 2Z(Z-Y),\, 2Z(Z-Y)]^T.$$

Compare this mass assignment to *Example ii*, although the mass assignment is still dependent on the "distance" between Z and Y, the magnitude of the mass assignment was reduced from $3Z$ to $2Z$. The assigned DS masses and uncertainty are

$$\mathbf{M} = \begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \frac{2Z(Z-Y)}{(3)(2)} \\ \frac{2Z(Z-Y)}{(3)(2)} \\ 1 - (\frac{4Z(Z-Y)}{(3)(2)}) \end{pmatrix}.$$

This reduction of magnitude is attributed to having additional competing prototypes. Furthermore, notice how the range of $M(\Theta)$ is constrained from the previous example's competing prototypes. $M(\Theta)$ will never achieve zero, even when $M(V_3)$, and $M(V_4)$ reaches a maximum "distance" between Z and Y. The range of $M(\Theta)$ is bounded between .33 and 1 because of the amount of focal elements present in the vector. This boundary is defined by the minimum uncertainty which was describe by the ratio of P and N, in Equation 4.30. For this example, we want an uncertainty range from .5 to 1, which fits the ideal ratio of $\frac{P}{N}$. As N increases we approach a better fit to this ideal ratio. In typical applications, the N value is not small, allowing a better approximated value of $\frac{P}{N}$. As N increases, we obtain a closer approximation to this ideal uncertainty range, seen in previous Table 4.2.

Numerical Examples

Example i: $\boldsymbol{V} = [0.01,\ 0.03,\ 0.12,\ 0.98]^T$. This example presents V_1, as the only strong candidate in the set. This generates a *weight matrix* of

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [-0.0110,\ -0.0306,\ -0.0792,\ 2.7244]^T.$$

Implementing the *column weights associated method*, we obtain one focal element ($P = 1$), and a *mass measure vector*,

$$\mathbf{H} = [0,\ 0,\ 0,\ 2.7244]^T,$$

where

$$\sum_{i=1}^{N}(H_i) = 2.7244.$$

The assigned DS masses and uncertainty are

$$\begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0.9081 \\ 0.0919 \end{pmatrix}.$$

Example ii: $\boldsymbol{V} = [0.98,\ 0.83,\ 0.40,\ 0.30]^T$. This example presents V_1 and V_2, as strong candidates in the set. The weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & -0.125 & -0.232 & -0.204 \\ 0.147 & 0 & -0.172 & -0.159 \\ 0.568 & 0.357 & 0 & -0.030 \\ 0.666 & 0.440 & 0.040 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [1.3818,\ 0.6723,\ -0.3640,\ -0.3930]^T.$$

Implementing the *column weights associated method*, we obtain a two focal elements ($P = 2$), and a *mass measure vector*,

$$\mathbf{H} = [1.3818,\ 0.6723,\ 0,\ 0]^T,$$

where

$$\sum_{i=1}^{N}(H_i) = 2.0541.$$

The assigned DS masses and uncertainty are

$$\begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0.2303 \\ 0.1120 \\ 0 \\ 0 \\ 0.6577 \end{pmatrix}.$$

Example iii: $\boldsymbol{V} = [0.31,\ 0.32,\ 0.43,\ 0.44]^T$. This example presents multiple weak candidates in the set. In this case, the weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0.0032 & 0.0516 & 0.0572 \\ -0.00310 & 0.0473 & 0.0528 & \\ -0.0372 & -0.0352 & 0 & 0.0044 \\ -0.0403 & -0.0384 & -0.0043 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [-0.0806,\ -0.0704,\ 0.0946,\ 0.1144]^T.$$

Implementing the *column weights associated method*, we obtain two focal elements ($P = 2$), and a *mass measure vector*,

$$\mathbf{H} = [0,\ 0,\ 0.0946,\ 0.1144]^T,$$

where

$$\sum_{i=1}^{N}(H_i) = 0.2090.$$

The assigned DS masses and uncertainty are

$$\begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0.0158 \\ 0.0191 \\ 0.9652 \end{pmatrix}.$$

In this example, we illustrate that with a weak set of correlation coefficients, masses are still assigned to the stronger candidates, but the uncertainty increases.

Example iv: $\boldsymbol{V} = [0.31,\ 0.32,\ 0.73,\ 0.74]^T$. We use this example to compare *Example iii* to demonstrate how uncertainty decreases if V_3 and V_4 become stronger candidates. In this case, the weight matrix is

$$\Delta\mathbf{W} = \begin{pmatrix} 0 & 0.0032 & 0.3066 & 0.3182 \\ -0.0031 & 0 & 0.2993 & 0.3108 \\ -0.1302 & -0.1312 & 0 & 0.0074 \\ -0.1333 & -0.1344 & -0.0073 & 0 \end{pmatrix},$$

which yields the following *column weights*,

$$\mathbf{C} = [-0.2666,\ -0.2624,\ 0.5986,\ 0.6364]^T.$$

Implementing the *column weights associated method*, we obtain a two focal elements ($P = 2$), and a *mass measure vector*,

$$\mathbf{H} = [0,\ 0,\ 0.5986,\ 0.6364]^T,$$

where

$$\sum_{i=1}^{N}(H_i) = 1.2350.$$

The assigned DS masses and uncertainty are

$$\begin{pmatrix} M(V_1) \\ M(V_2) \\ M(V_3) \\ M(V_4) \\ M(\Theta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0.0998 \\ 0.1061 \\ 0.7942 \end{pmatrix}.$$

Chapter 5

Simulations

The simulations in this chapter examine the functionality of the proposed method and the feasibility of using DS theory for qualitative chemometric analysis. The data for these simulations are obtained from the PhotoChemCAD and consequently processed with the method discussed in Section 2.1.

We present these simulations under idealistic and adverse conditions, by adding synthetic noise that was previously described in Section 2.3. Our method uses the evidence models discussed in Section 4.1, where these models process the absorption and emission spectrums under various noise conditions. DS theory's Dempster's combinations rule is then applied to determine the analyte combination present within a simulated unknown specimen.

5.1 Overview of the Proposed Method

In our work, we restricted our attention to seven chemicals, each of which is set to have a uniform concentration of $0.5\mu M$. The absorption spectra of these seven chemicals are shown in Figures 5.1 and 5.2. These seven chemicals correspond to the first seven chemicals in Tables 2.1 and 2.2 located in Section 2.1. The reason for restricting our attention to these seven chemicals was mainly related to computational complexity associated with running numerous permutations within the data set.

This is done to examine the effects of different noise levels, generating different evidence sources, and combination of evidence from these sources. The computation complexity associated with the application of the proposed algorithm for classification is not overly significant.

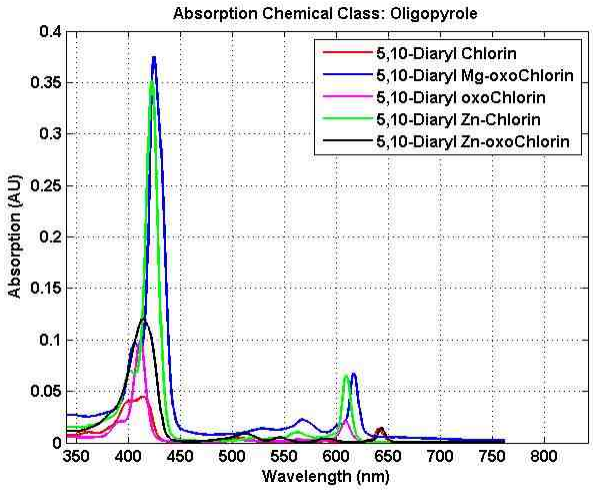


Figure 5.1: Oligopyrrole Absorption Data at $.5\mu M$

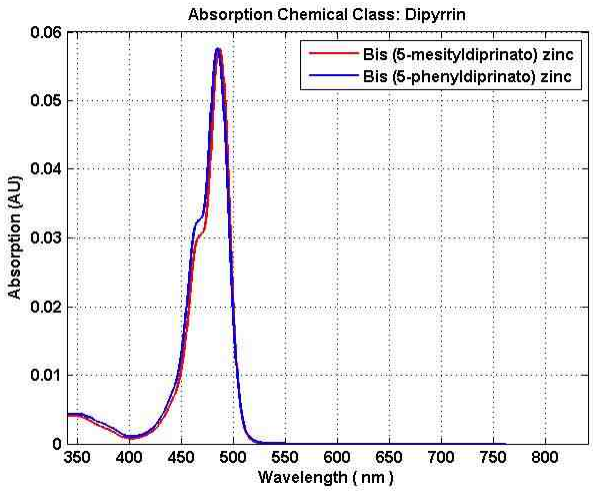


Figure 5.2: Dipyrrin Absorption Data at $.5\mu M$

Beer-Lamberts law of Additivity enabled us to produce $2^7 - 1 = 127$ different chemical combinations (a sample contains at least one chemical) from the original

seven chemicals. Figure 5.3 shows the close spectral similarities corresponding to these different combinations.

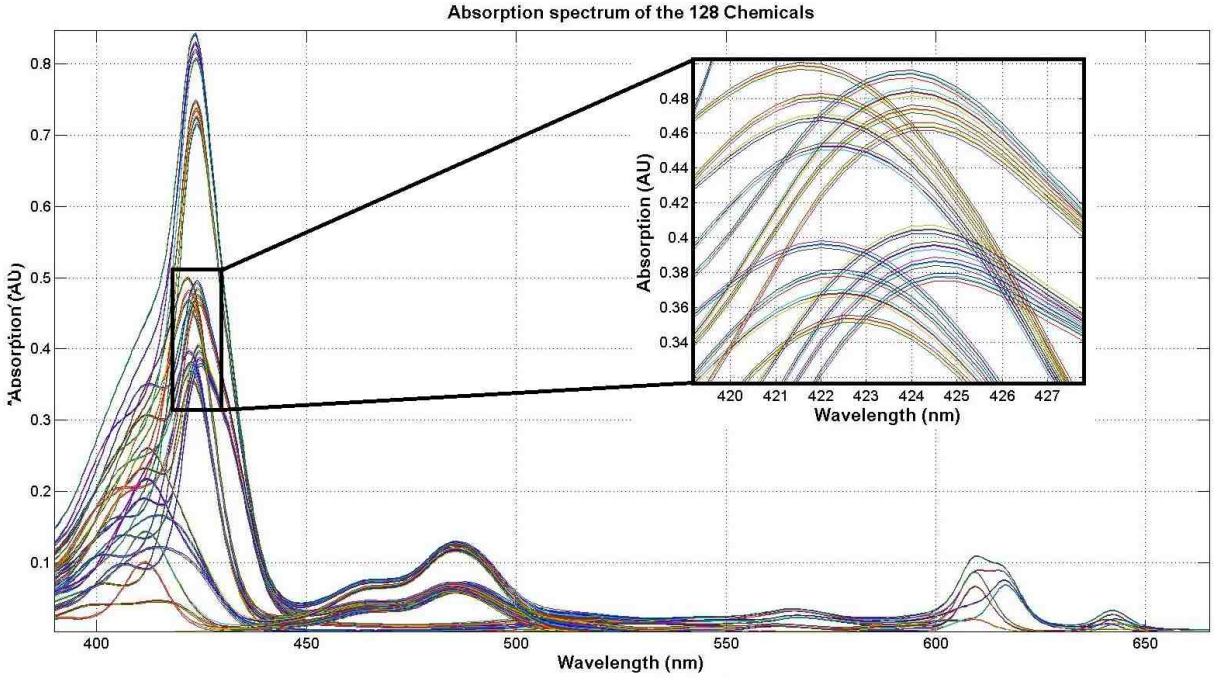


Figure 5.3: Absorption Spectra of the 127 Combinations at $.5\mu M$

We simulated the detection of these 127 combinations by running 1500 random detection experiments to guarantee the examination of all 127 combinations. The simulation encompasses different noise parameters of η, ranging from zero to two, by incrementing η in 0.5 intervals.

As we discussed in Section 2.1.3, the quantum yield affects the intensity of the emission spectrum. This value quantifies the dispersion of energy as emitted light by the molecule. We incorporated this into our fluorescent model's emission data by implementing equation 2.4 and simulating excitation wavelengths from $400nm$ to $650nm$ by incrementing every $25nm$. These simulations produce the eleven emission spectra in Figures 5.4 and 5.5.

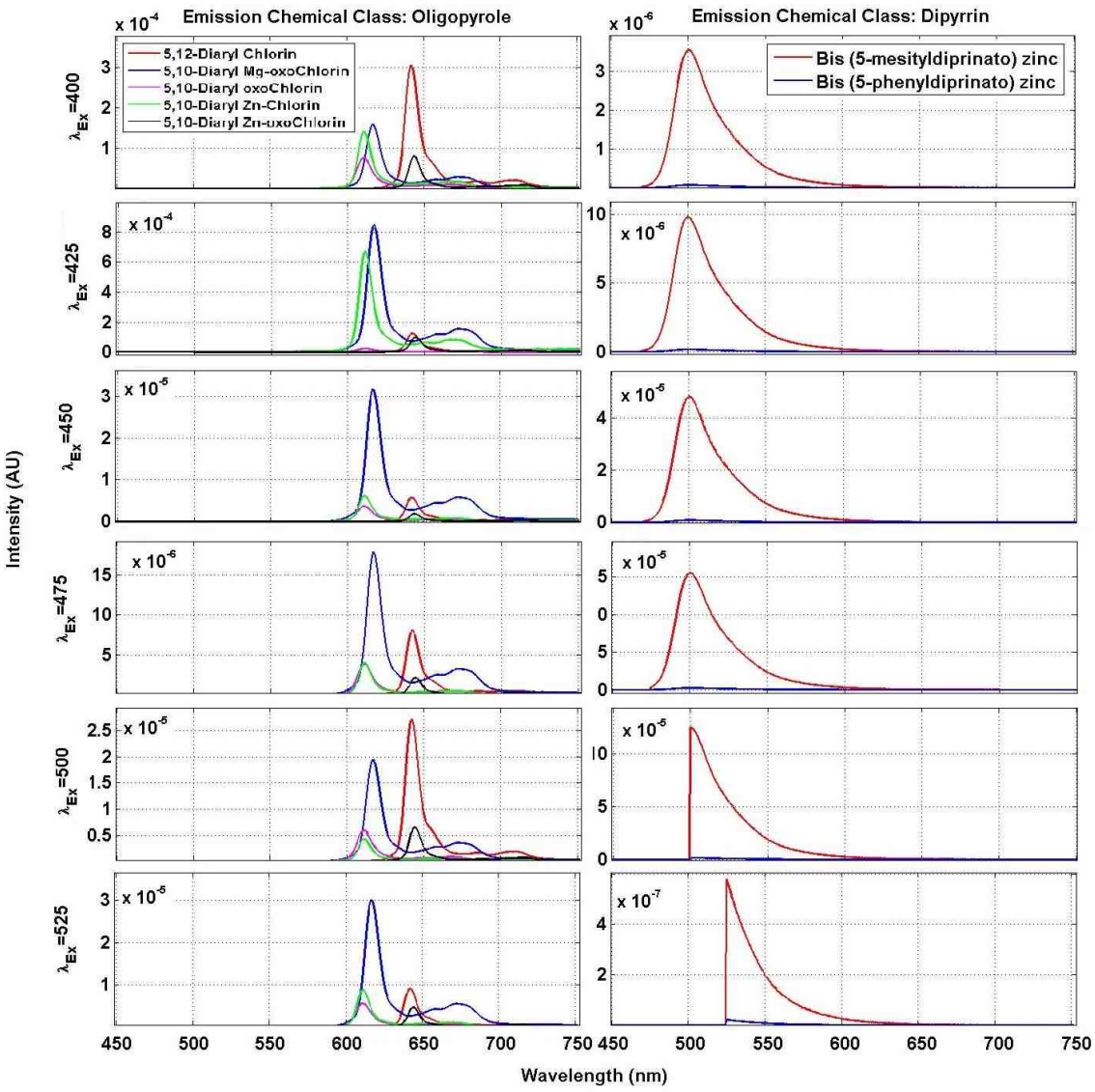


Figure 5.4: Fluorescent Emission from λ_{Ex} $400nm$ to $525nm$

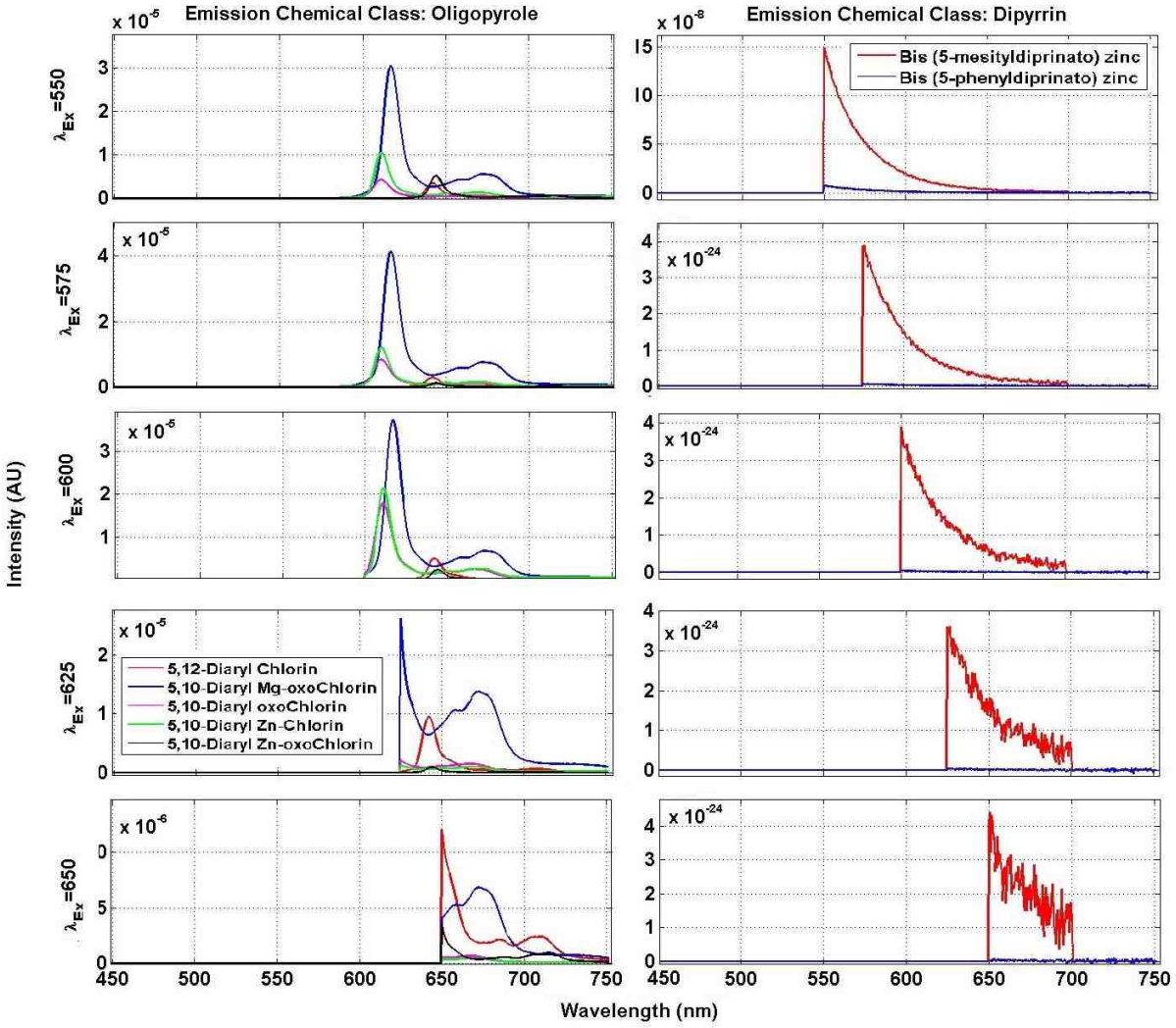


Figure 5.5: Fluorescent Emission from λ_{Ex} $550nm$ to $650nm$

5.1.1 Overview of Detection Process and Experimental Simulations

The detection process and experimental simulations are summarized in the flow charts in Figures 5.6 and 5.7. The process begins in Figure 5.6, where a random chemical combination, X_{Exp}, is picked from the 127 possibilities. This random chemical combination,X_{Exp}, will produce one unique absorption spectrum and eleven unique emission spectra from the eleven excitation wavelengths. Each individual spectra is

then processed to extract its associated four features: central coefficients, spectral peaks, spectral intensity differential, and matched filter. In the flow chart, this is referred to as "feature extraction". The extracted feature is then compared against the same feature corresponding to each possible chemical combination in the database, The comparison is implemented via correlation coefficients. We then convert our correlation coefficients to DS masses described in Section 4.2 and apply Dempster's combination rule (DCR) to obtain the fused DS masses. In the next step, we sort the DS masses to evaluate each evidence model technique over the multiple fusions of spectra shown in Figure 5.7.

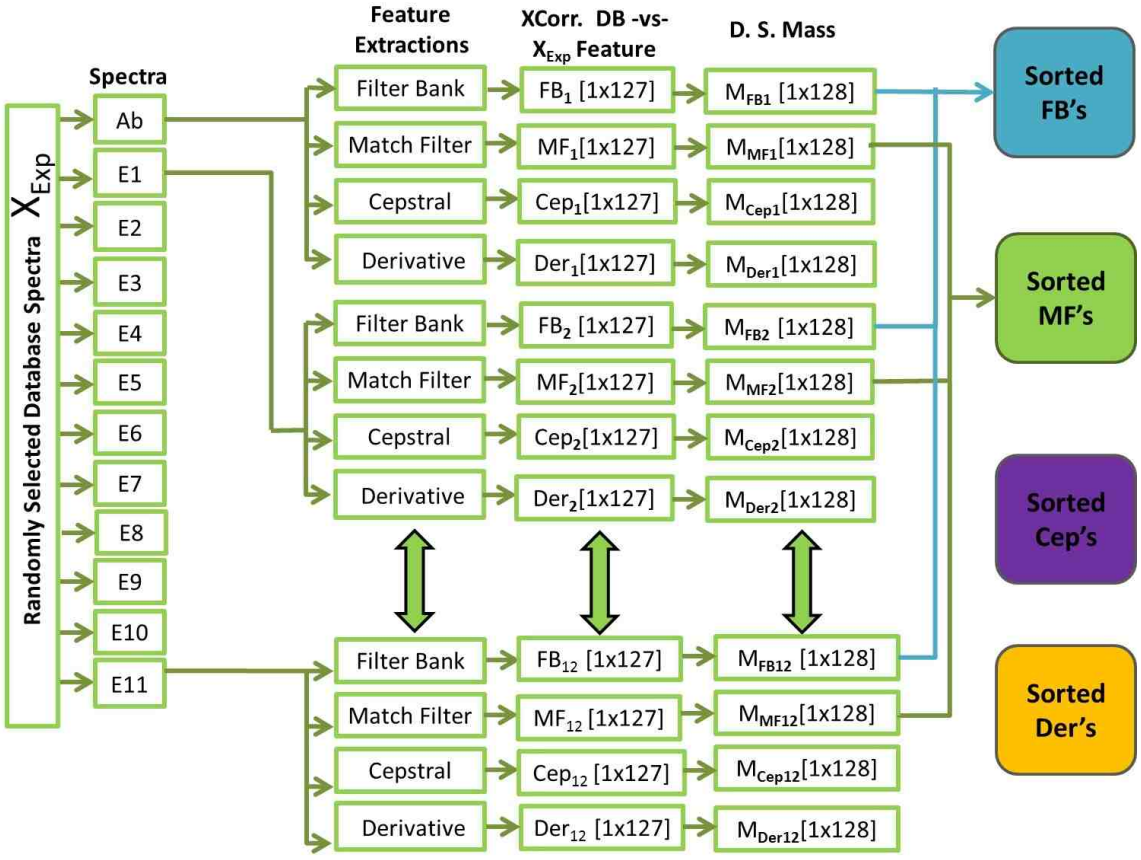


Figure 5.6: Part 1: An Overview of the Simulation Process

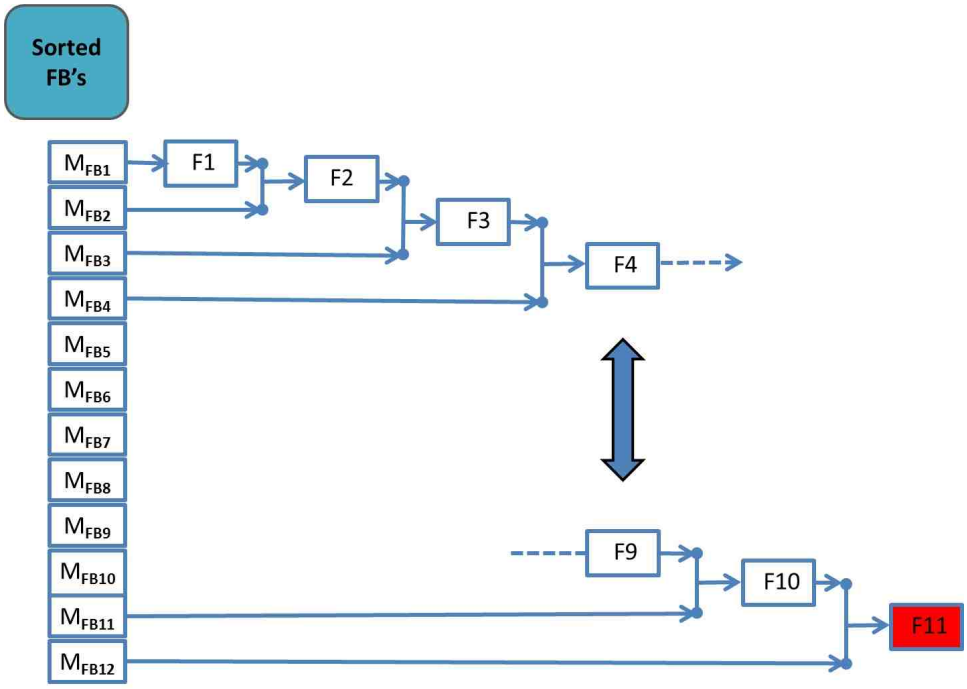


Figure 5.7: An Overview of DCR implementing Filter Bank Evidence

5.2 Colored Compression and Dilation Noise

The simulations that are provided examine the detection of the exact chemical combination when noise is introduced into the spectrum. The following sections presents noise levels ranging from η beginning at zero to two. When η is zero there is obviously no noise presented within the system, and the pure signal is simulated for detection. When η is two, it presents the most extreme noise situations for the spectral signal. The spectral signal at localized areas is capable of being compressed to zero or dilated to double in magnitude. The following simulations below are presented with the noise parameters of a window length of $50nm$ and an $\alpha = 1$. In Figure 5.8, this magnitude affect of colored compression and dilation noise on the fluorescent spectrum, as a function of η, is evident. This spectral sample is from the absorption spectrum and lies at index 56, which is 0111000 in binary code. This binary code represents the three additive chemicals that are presented in the figure; Bis (5-mesityldiprinato) zinc, 5,10-Diaryl Zn-Chlorin and 5,10-Diaryl Zn-oxoChlorin.

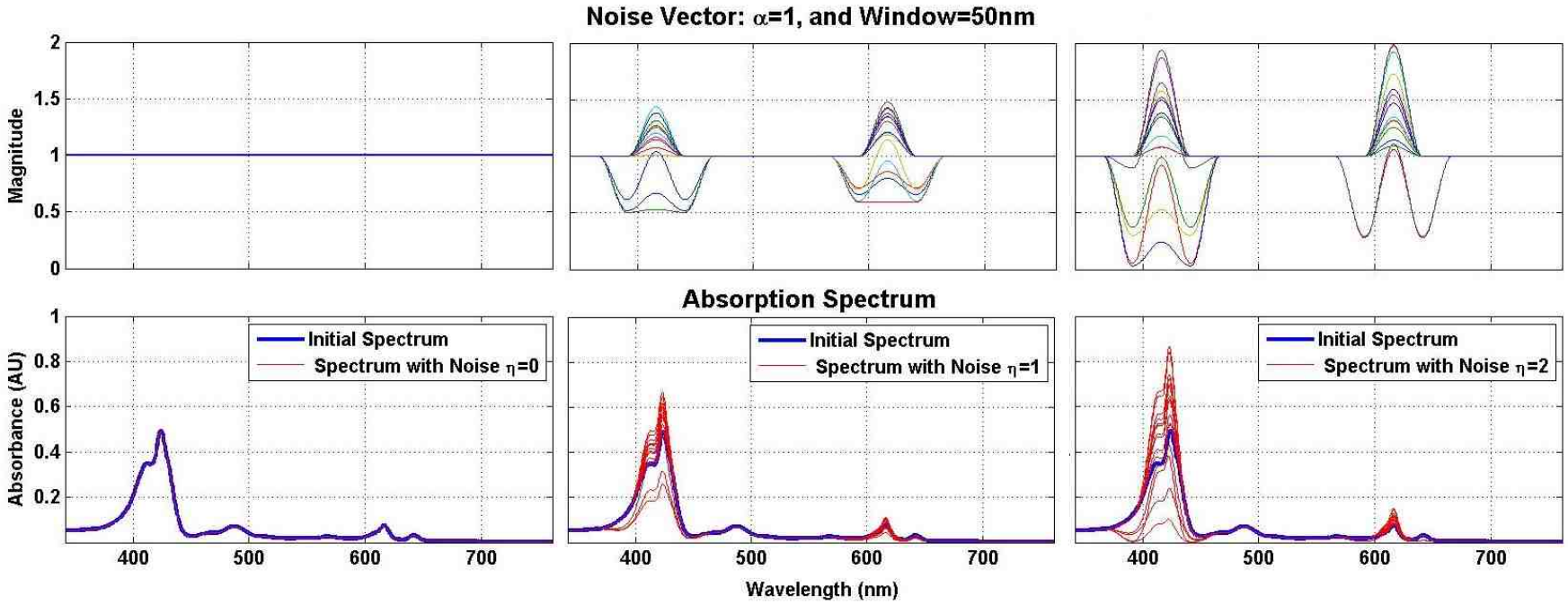


Figure 5.8: Different Noise Parameters of η

5.3 The Assessment of Feature Extractions Methods For Evidence

Through the simulation process, we examined various stages of the algorithm at different noise conditions. One of the stages is the evidence model's correlation coefficients. We evaluate this stage by examining if the random chemical in the experiment is detected as one of the top 5 highest correlation coefficients in the set. The top five correlation coefficients are represented by a Rank, where Rank 1 is the first highest correlation coefficient and Rank 5 is the fifth highest. We initially simulated the detection process with no noise for the data set to develop a baseline of our detection capability. We then proceeded with increased conditions of noise. This allows for the comparison of the Dempster Shafer Model's detection results to be evaluated downstream against the results of the correlation coefficients. The following results are depicted in Figures 5.9 and 5.10.

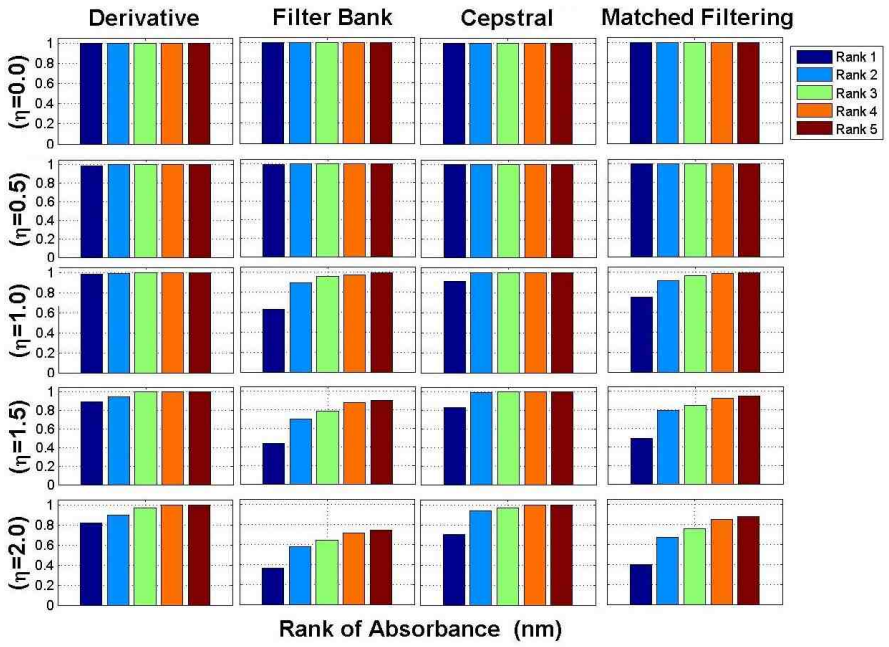


Figure 5.9: Detection of the Absorption Spectrums Under Different Noise Parameters of η Using Feature Extraction's Correlation Coefficients

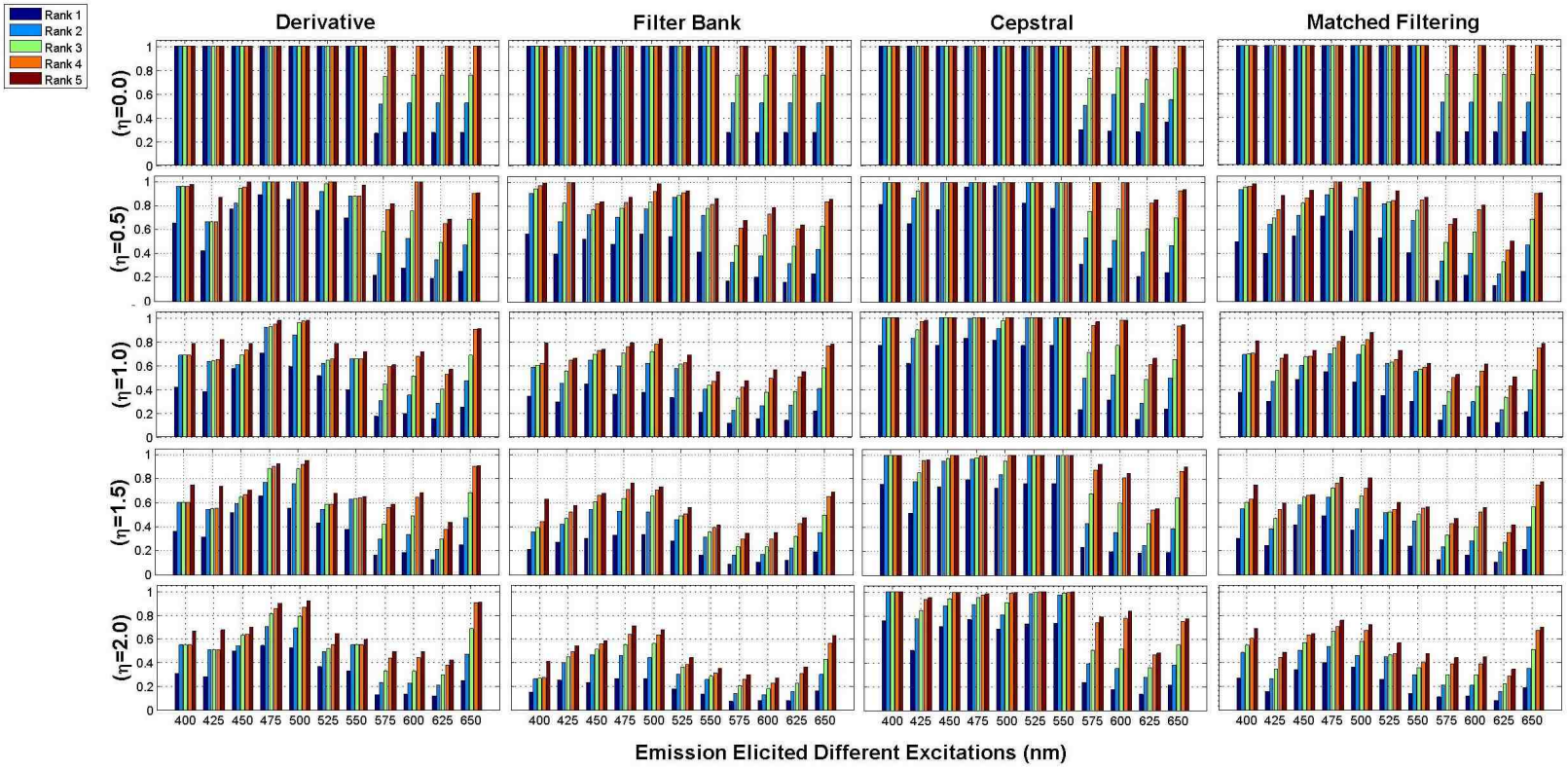


Figure 5.10: Detection of the Emission Spectrums Under Different Noise Parameters of η Using Feature Extraction's Correlation Coefficients

	Matched Filter				
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**
$\boldsymbol{\eta = 0.0}$	75.9%	84.1%	91.9%	100%	100%
$\boldsymbol{\eta = 0.5}$	45.5%	66.7%	75.3%	83.6%	87.5%
$\boldsymbol{\eta = 1.0}$	35.0%	53.4%	60.8%	67.5%	72.4%
$\boldsymbol{\eta = 1.5}$	28.7%	46.4%	54.4%	61.5%	66.3%
$\boldsymbol{\eta = 2.0}$	23.8%	38.3%	46.8%	54.3%	59.6%
	Filter Bank				
$\boldsymbol{\eta = 0.0}$	75.9%	84.1%	91.9%	100%	100%
$\boldsymbol{\eta = 0.5}$	44.1%	65.6%	74.7%	84.1%	87.2%
$\boldsymbol{\eta = 1.0}$	30.0%	49.3%	57.8%	64.6%	69.8%
$\boldsymbol{\eta = 1.5}$	23.6%	39.6%	47.3%	54.0%	59.2%
$\boldsymbol{\eta = 2.0}$	18.9%	32.6%	39.0%	44.8%	50.1%
	Derivative				
$\boldsymbol{\eta = 0.0}$	75.8%	84.1%	91.8%	100%	100%
$\boldsymbol{\eta = 0.5}$	58.1%	74.9%	83.0%	89.9%	93.6%
$\boldsymbol{\eta = 1.0}$	44.5%	61.4%	68.6%	74.8%	80.3%
$\boldsymbol{\eta = 1.5}$	40.2%	55.9%	64.0%	69.7%	75.0%
$\boldsymbol{\eta = 2.0}$	36.0%	50.7%	58.1%	64.0%	70.1%
	Cepstral				
$\boldsymbol{\eta = 0.0}$	77.0%	84.7%	92.3%	100%	100%
$\boldsymbol{\eta = 0.5}$	65.3%	81.8%	89.8%	98.0%	98.3%
$\boldsymbol{\eta = 1.0}$	59.5%	79.3%	87.3%	95.1%	96.0%
$\boldsymbol{\eta = 1.5}$	55.8%	74.6%	84.3%	92.1%	93.2%
$\boldsymbol{\eta = 2.0}$	53.1%	72.3%	79.5%	88.5%	90.0%

Table 5.1: Average Detection Across Spectrum

5.4 Accuracy Assessment of Correlation Coefficient Evidence Vs DCR Fusion

This section presents the accuracy of detecting unknown chemicals using correlation coefficients and the fusion of the correlation coefficient information that was manipulated to fit the DS theory framework. The dempster combination rule was applied for this data to be fused. The following Figures: 5.6.3, 5.12, 5.11, and 5.6.3 show the accuracy of detection as a function of η, the noise. These figures only analyze Rank 1, which is when the highest correlation coefficient or DS Mass is chosen. In these Figures two curves are bolded DCR Fusion and Average. DCR Fusion is the fusion of all 12 spectra, over a single evidence technique. DCR Fusion is bolded to highlight the importance and the final results of the decision. Average is the average accuracy of the correlation coefficients from the 12 spectra that are examined. This is bolded to simply obtain a single reference for comparison against the DCR Fusion result in order to gauge the algorithms improvement.

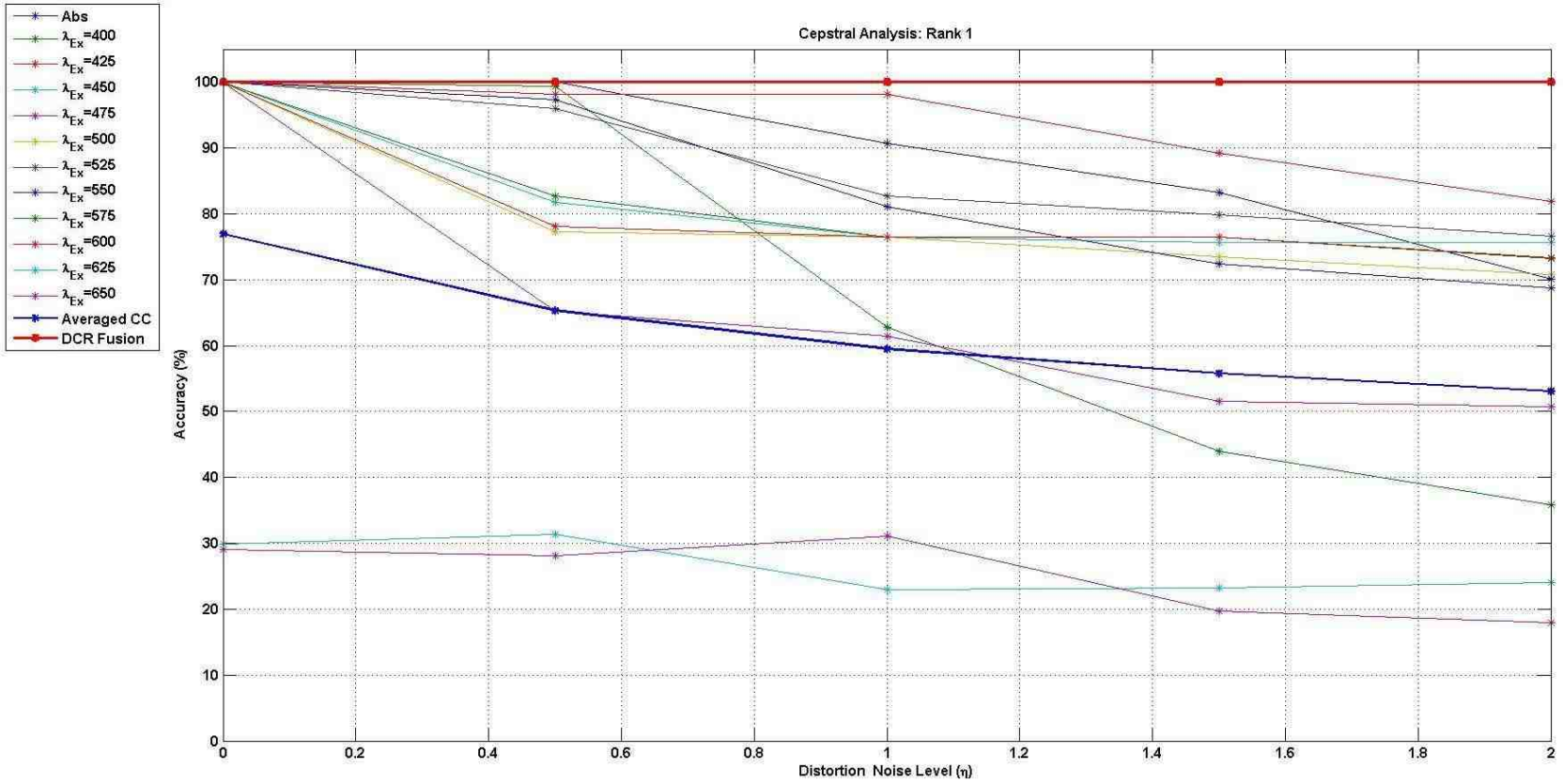


Figure 5.11: Cepstral Analysis: Detection Comparison of Dempster Combination Rule and Correlation Coefficients

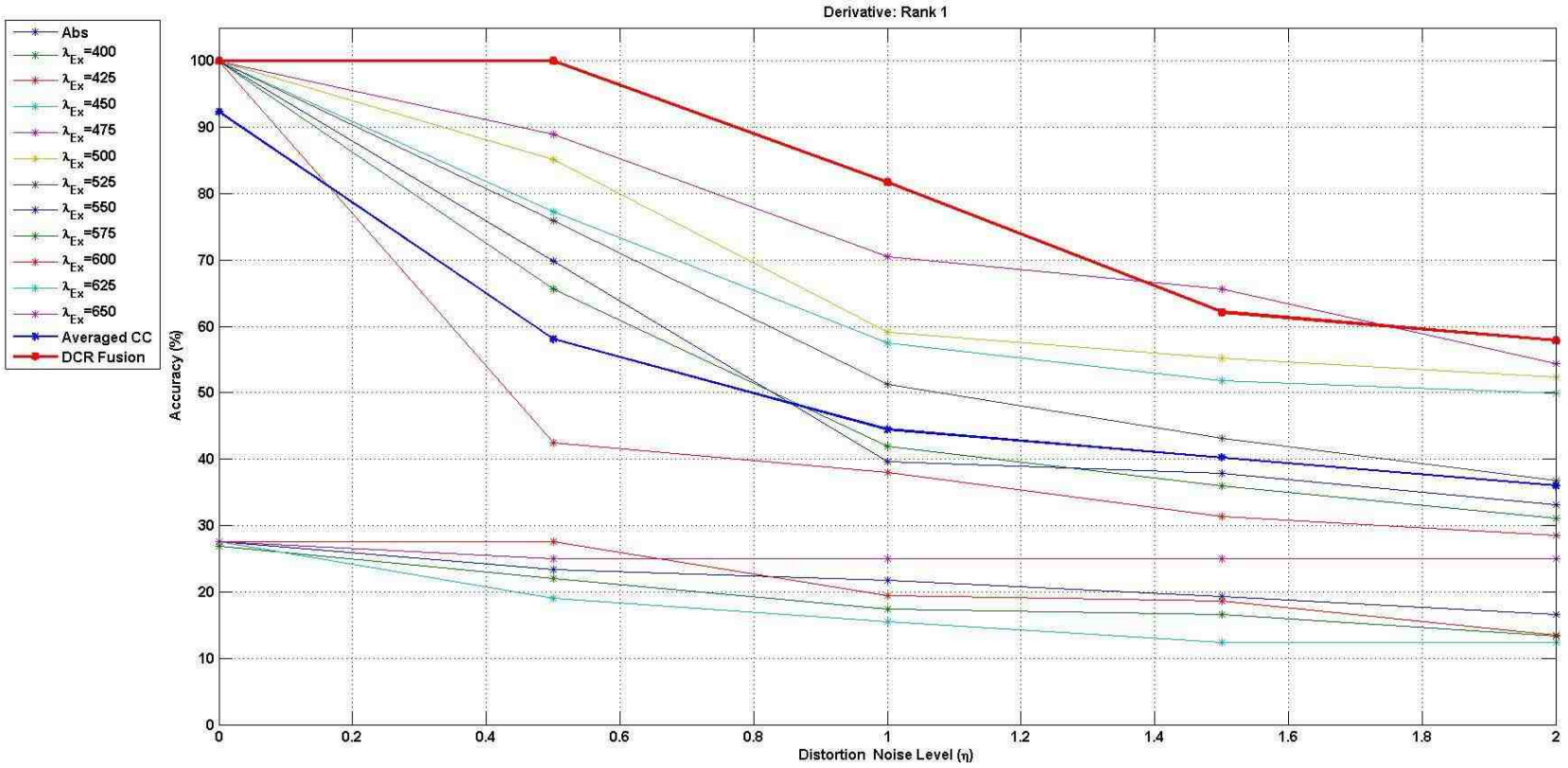


Figure 5.12: Derivative Analysis: Detection Comparison of Dempster Combination Rule and Correlation Coefficients

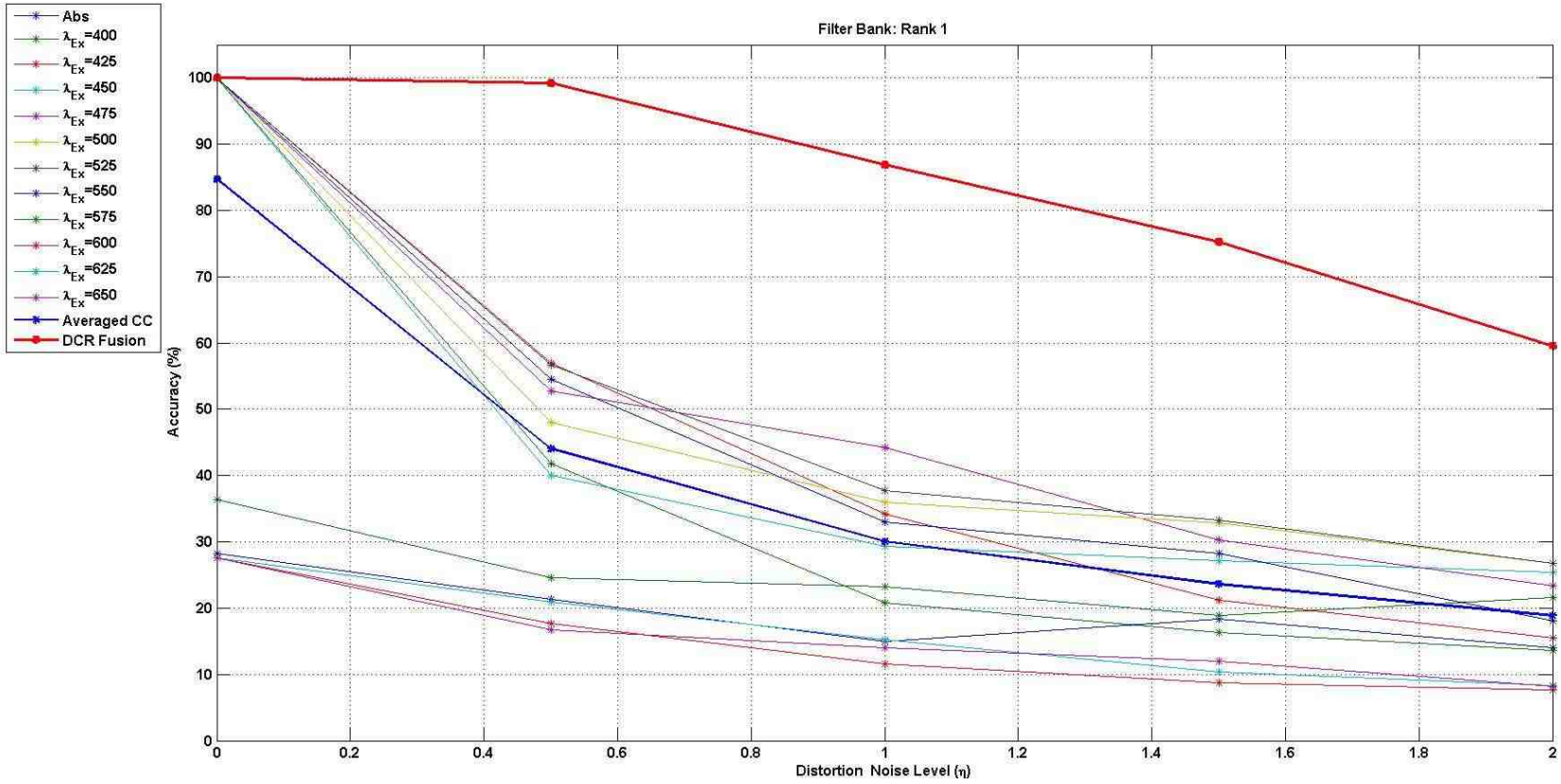


Figure 5.13: Filter Bank Analysis: Detection Comparison of Dempster Combination Rule and Correlation Coefficients

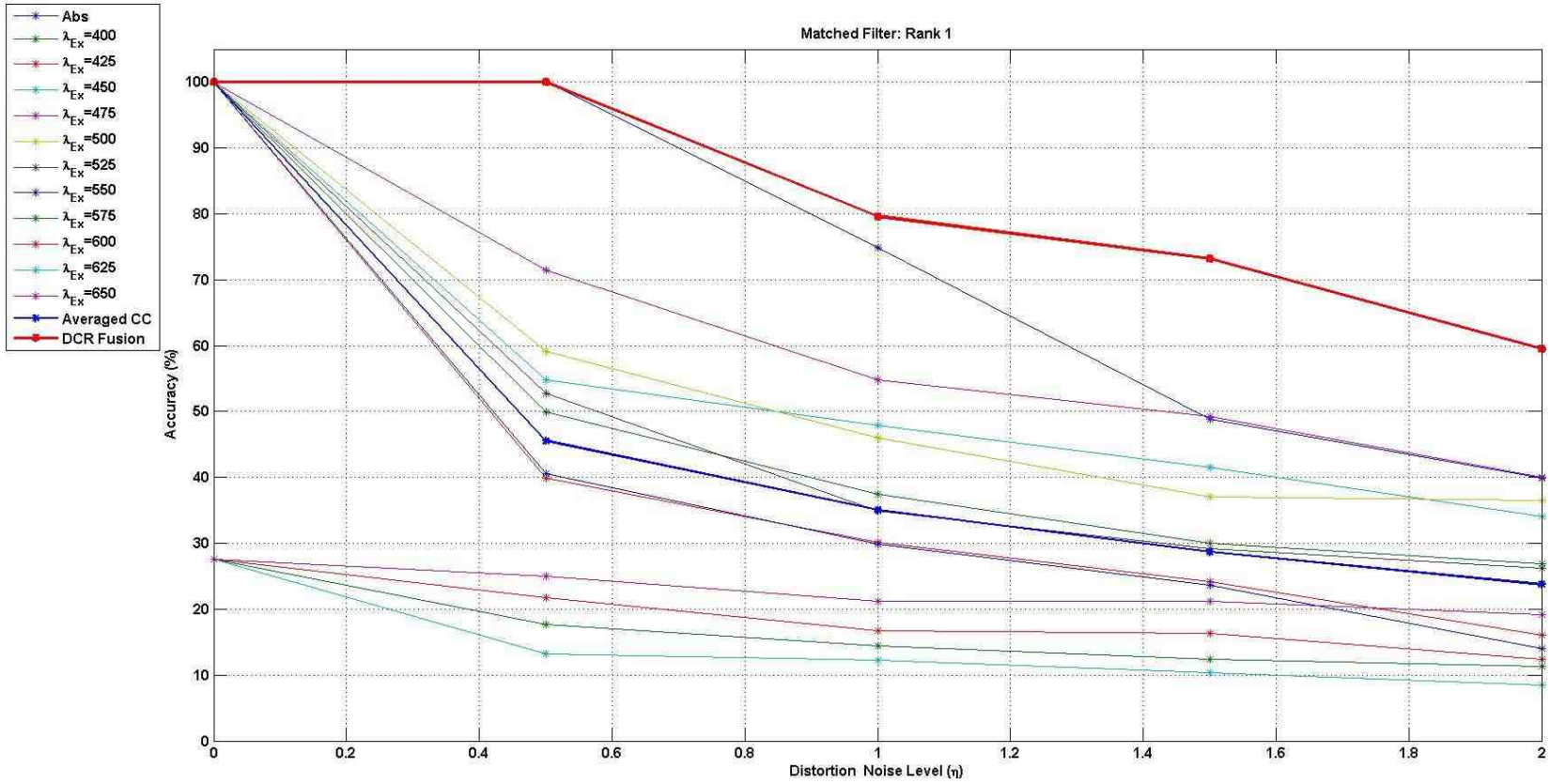


Figure 5.14: Matched Filter Analysis: Detection Comparison of Dempster Combination Rule and Correlation Coefficients

5.4.1 Coefficient Distributions and Dempster Uncertainty

As we discussed in Section 4.2, the correlation coefficients of the evidence model are used to determine the DS mass and uncertainty. This section examines the uncertainty from the simulation data and then compares it to correlation coefficient's distribution. The following Table 5.2 is the average uncertainty assignment for each associated spectrum and evidence model, over the 1500 simulations, for the noise parameter when $\eta = 1.5$.

Table 5.2: Averaged Uncertainty Specific to Each Evidence

Absorption: Averaged Uncertainty										
Cepstral			*Filter Bank*			*Derivative*			*Matched Filter*	
0.95			0.932			0.711			0.913	
Emission: Averaged Uncertainty										
Cepstral										
λ_{400}	λ_{425}	λ_{450}	λ_{475}	λ_{500}	λ_{525}	λ_{550}	λ_{575}	λ_{600}	λ_{625}	λ_{650}
0.987	0.986	0.986	0.987	0.999	0.999	0.999	0.998	0.999	.999	0.999
Filter Bank										
λ_{400}	λ_{425}	λ_{450}	λ_{475}	λ_{500}	λ_{525}	λ_{550}	λ_{575}	λ_{600}	λ_{625}	λ_{650}
0.842	0.807	0.676	0.632	0.623	0.865	0.873	0.896	.909	0.942	0.970
Derivative										
λ_{400}	λ_{425}	λ_{450}	λ_{475}	λ_{500}	λ_{525}	λ_{550}	λ_{575}	λ_{600}	λ_{625}	λ_{650}
0.718	0.534	0.573	0.554	0.583	0.642	0.605	0.554	0.632	.790	0.962
Matched Filter										
λ_{400}	λ_{425}	λ_{450}	λ_{475}	λ_{500}	λ_{525}	λ_{550}	λ_{575}	λ_{600}	λ_{625}	λ_{650}
0.825	0.753	0.681	0.633	0.625	0.815	0.815	0.808	0.853	.886	0.945

The histograms presented in Figure 5.15, are the 127 correlation coefficients that were assigned to each simulation for the 1500 simulations that were generated. These correlation coefficient histograms are from the emission spectra at the λ_{425}

excitation wavelength with respect to a specific evidence model. This specific excitation wavelength and their respective evidence models were chosen since their respective assigned uncertainty was wide ranging relative to each other. This best demonstrates the algorithms uncertainty assignment and its association with the correlation coefficient's distribution over the set.

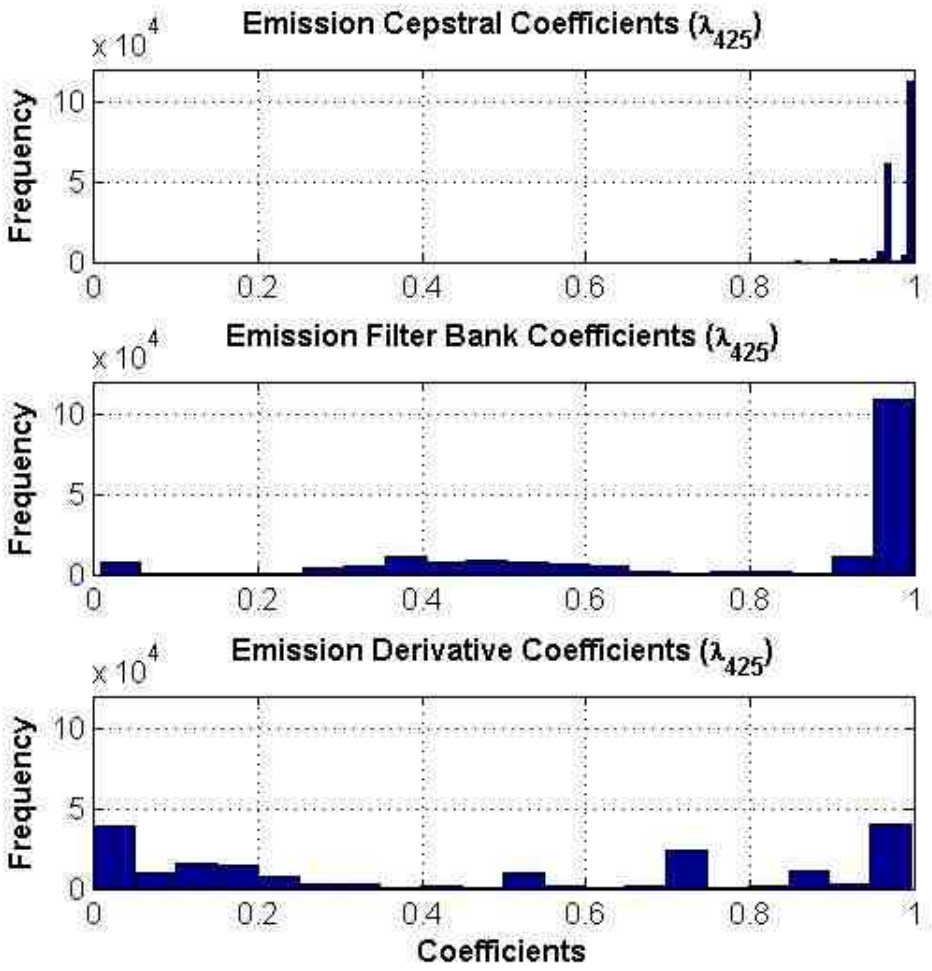


Figure 5.15: Histogram of Emission Feature's Coefficients (λ_{425})

5.5 Fusion Over Spectra

In process flow chart 5.7, we demonstrated the fusion of the 12 spectra with respect to an individual evidence model. The final fusion result will be altered, based on the order of the fusion of spectra. The order of how the spectra is fused is irrelevant, since the dempster combination rule is associative and distributive. Figure 5.16 and 5.17 shows the accuracy of the fusion as each spectrum is fused with the previous while keeping a single evidence model constant over the fusion process. This fusion is shown for two noise levels $\eta = 0.5$ and $\eta = 2.0$.

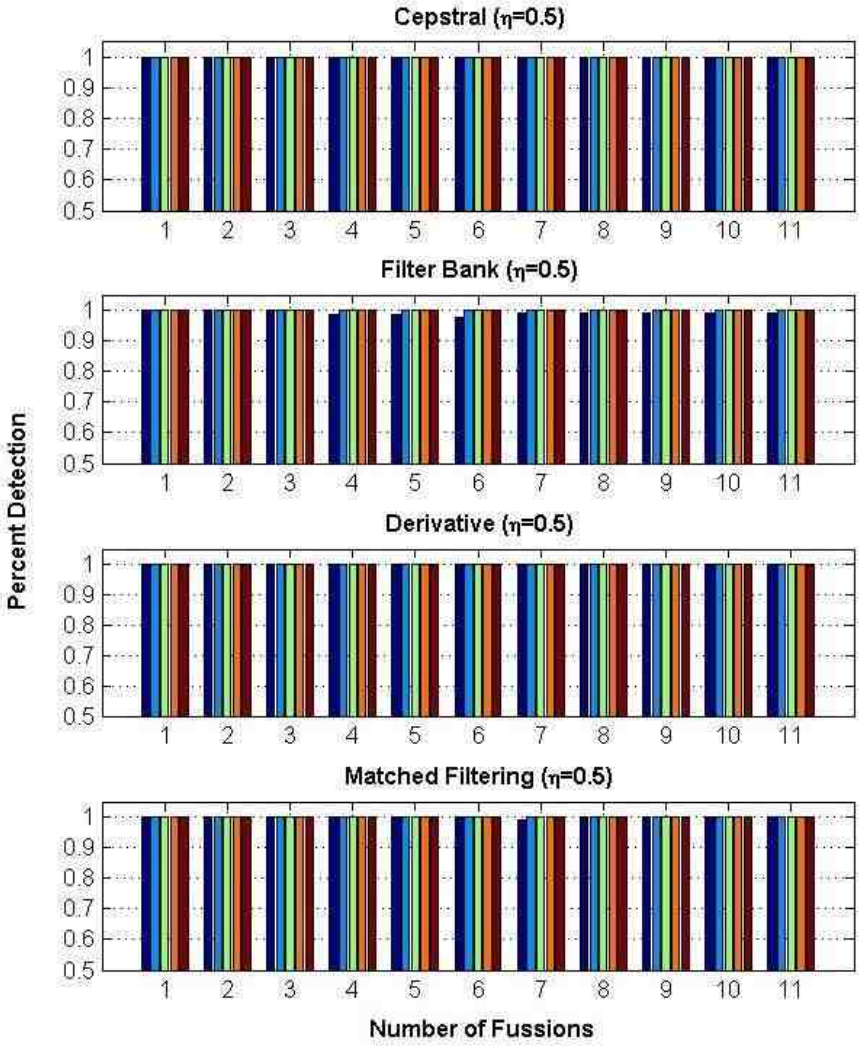


Figure 5.16: Fusion Across Numerous Spectra at $\eta = 0.5$

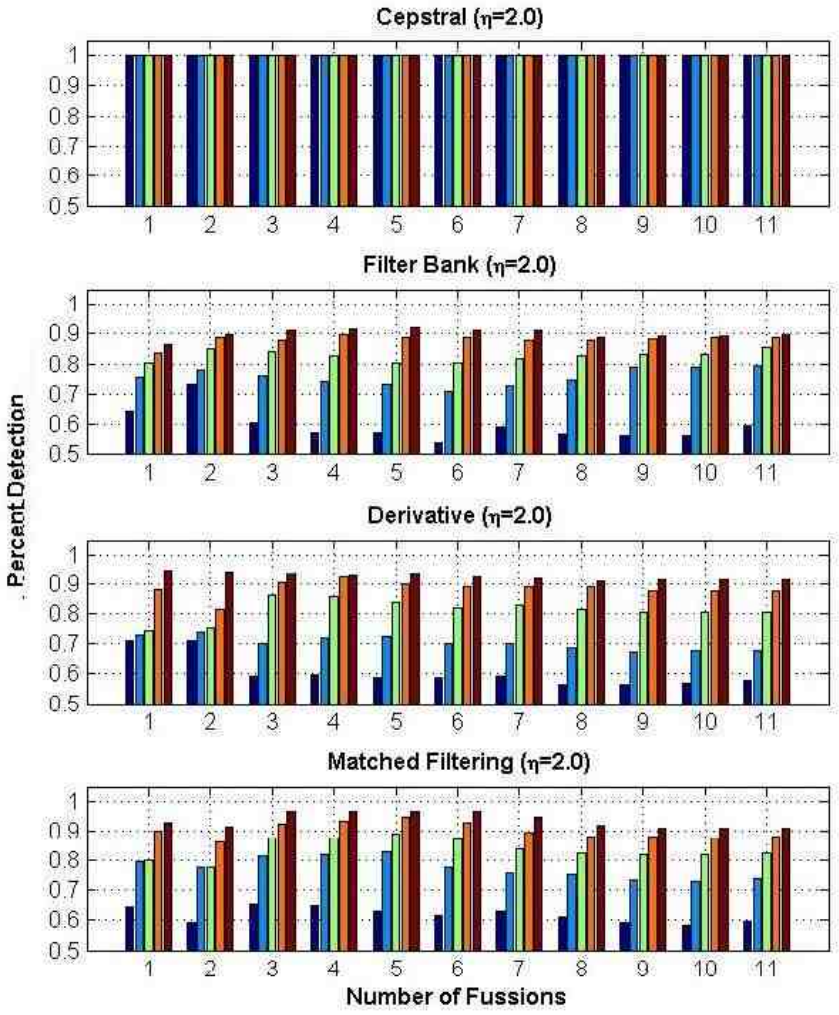


Figure 5.17: Fusion Across Numerous Spectra at $\eta = 2.0$

5.6 Combinations of Different Variations of Evidence

This section reports the fusion accuracy over the first five ranks and their associated assigned uncertainty. We then examine the fusion accuracy of various combinations of evidence methods and their associated uncertainty.

5.6.1 Fusion of One Form of Evidence

Table 5.3: Fusion of One Form of Evidence

	Matched Filter					
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	Θ
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.232
$\boldsymbol{\eta = 1.0}$	79.6%	91.8%	97.5%	99.1%	100%	.231
$\boldsymbol{\eta = 1.5}$	73.2%	81.5%	89.6%	92.9%	97.0%	.230
$\boldsymbol{\eta = 2.0}$	60.0%	73.9%	82.3%	88.3%	90.9%	.229
	Filter Bank					
$\boldsymbol{\eta = 0.5}$	99.2%	100%	100%	100%	100%	.265
$\boldsymbol{\eta = 1.0}$	86.9%	93.9%	97.4%	99.3%	100%	.266
$\boldsymbol{\eta = 1.5}$	75.3%	85.4%	88.5%	91.8%	98.3%	.266
$\boldsymbol{\eta = 2.0}$	59.6%	79.3%	85.5%	88.6%	88.9%	.268
	Derivative					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.119
$\boldsymbol{\eta = 1.0}$	81.7%	89.6%	96.3%	96.3%	98.4%	.120
$\boldsymbol{\eta = 1.5}$	62.1%	72.5%	89.3%	91.3%	93.7%	.123
$\boldsymbol{\eta = 2.0}$	57.9%	68.0%	80.7%	87.8%	91.4%	.127
	Cepstral					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.900
$\boldsymbol{\eta = 1.0}$	100%	100%	100%	100%	100%	.900
$\boldsymbol{\eta = 1.5}$	100%	100%	100%	100%	100%	.900
$\boldsymbol{\eta = 2.0}$	99.4%	100%	100%	100%	100%	.900

5.6.2 Fusion of Two Forms of Evidence:

Table 5.4: Fusion of two Methods of Evidence Part A

	Matched Filter and Filter Bank					
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	Θ
$\boldsymbol{\eta = 0.5}$	99.2%	100%	100%	100%	100%	.140
$\boldsymbol{\eta = 1.0}$	82.8%	94.1%	99.5%	100%	100%	.140
$\boldsymbol{\eta = 1.5}$	76.4%	85.4%	91.3%	96.4%	98.0%	.139
$\boldsymbol{\eta = 2.0}$	64.5%	77.1%	86.3%	91.6%	95.4%	.139
	Derivative and Filter Bank					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.087
$\boldsymbol{\eta = 1.0}$	83.2%	90.7%	96.3%	96.3%	100%	.088
$\boldsymbol{\eta = 1.5}$	68.1%	83.3%	91.7%	93.7%	95.1%	.090
$\boldsymbol{\eta = 2.0}$	58.6%	69.4%	82.2%	89.3%	91.9%	.092
	Derivative and Cepstral					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.117
$\boldsymbol{\eta = 1.0}$	81.7%	90.0%	96.3%	96.8%	98.9%	.119
$\boldsymbol{\eta = 1.5}$	62.1%	76.7%	89.3%	92.6%	95.1%	.122
$\boldsymbol{\eta = 2.0}$	58.5%	70.2%	82.1%	88.7%	92.3%	.125
	Matched Filter and Cepstral					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.228
$\boldsymbol{\eta = 1.0}$	80.4%	91.8%	99.1%	99.1%	100%	.227
$\boldsymbol{\eta = 1.5}$	74.8%	81.5%	89.6%	93.9%	97.0%	.225
$\boldsymbol{\eta = 2.0}$	61.7%	74.6%	82.9%	89.7%	92.3%	.224

Table 5.5: Fusion of two Forms Evidence Part B

	Matched Filter and Derivative					
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	Θ
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.083
$\boldsymbol{\eta = 1.0}$	83.2%	90.0%	96.3%	96.3%	98.9%	.083
$\boldsymbol{\eta = 1.5}$	67.5%	80.8%	89.0%	92.8%	94.7%	.085
$\boldsymbol{\eta = 2.0}$	57.9%	69.4%	82.0%	88.5%	90.6%	.087
	Cepstral and Filter Bank					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.260
$\boldsymbol{\eta = 1.0}$	87.7%	94.8%	97.4%	99.3%	100%	.260
$\boldsymbol{\eta = 1.5}$	76.1%	85.4%	89.0%	93.9%	99.3%	.260
$\boldsymbol{\eta = 2.0}$	61.5%	80.5%	86.2%	89.0%	90.9%	.261

5.6.3 Fusion of Multiple Methods of Evidence

Table 5.6: Fusion of Multiple Forms of Evidence

	Derivative, Cepstral, and Filter Bank					
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	Θ
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.086
$\boldsymbol{\eta = 1.0}$	83.2%	90.7%	96.3%	96.8%	100%	.087
$\boldsymbol{\eta = 1.5}$	68.1%	83.7%	91.7%	94.1%	95.7%	.089
$\boldsymbol{\eta = 2.0}$	59.2%	71.3%	82.2%	90.1%	92.9%	.091
	Derivative, Filter Bank, and Matched Filter					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.065
$\boldsymbol{\eta = 1.0}$	83.2%	90.7%	96.3%	96.8%	100%	.066
$\boldsymbol{\eta = 1.5}$	69.8%	82.5%	91.7%	94.1%	95.7%	.067
$\boldsymbol{\eta = 2.0}$	58.6%	69.4%	82.2%	89.3%	91.5%	.068
	Derivative, Cepstral, and Match					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.082
$\boldsymbol{\eta = 1.0}$	83.2%	90.7%	96.3%	96.8%	100%	.083
$\boldsymbol{\eta = 1.5}$	67.5%	82.0%	90.2%	94.1%	94.7%	.084
$\boldsymbol{\eta = 2.0}$	58.5%	71.3%	82.2%	90.1%	92.4%	.086
	Cepstral, Filter Bank and Matched Filtering					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100%	.138
$\boldsymbol{\eta = 1.0}$	83.6%	94.1%	99.5%	100%	100%	.138
$\boldsymbol{\eta = 1.5}$	78.0%	85.4%	92.1%	96.9%	98.0%	.138
$\boldsymbol{\eta = 2.0}$	68.1%	78.5%	87.1%	91.6%	95.4%	.137

Table 5.7: Fusion of Multiple Forms of Evidence

	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	Θ
	Cepstral, Filter Bank, Matched Filtering, and Derivative					
$\boldsymbol{\eta = 0.5}$	100%	100%	100%	100%	100	.065
$\boldsymbol{\eta = 1.0}$	83.2%	91.1%	96.3%	96.8%	100%	.066
$\boldsymbol{\eta = 1.5}$	69.8%	82.5%	91.7%	94.1%	95.7%	.067
$\boldsymbol{\eta = 2.0}$	59.2%	71.3%	82.0%	91.1%	92.8%	.068

Chapter 6

Conclusion and Future Results

We have proposed a novel qualitative algorithm using Dempster-Shafer Theory to be introduced into the chemometric field for the detection of specific analyte combinations, which are close in chemical composition. The design of this novel method led in the development of an algorithm for converting correlation coefficients to fit the Dempster-Shafer Theory's framework for their respective masses and uncertainty. In the process of the development of this algorithm, we have developed unique modeling techniques for situations when spectroscopy data is scarce or unavailable.

The developed spectroscopy modeling techniques, were designed to mimic the typical behavior of absorption and emission spectrum in order to obtain a challenging and legitimate spectroscopy detection problem. This detection challenge could be seen in Figure 5.5 in the emission spectra for excitation wavelengths $575nm$, $600nm$, $625nm$, and $650nm$ by examining the magnitude difference of the two chemical classes, where Oligopyrole is 10^{-5} and Dipyrrin is 10^{-24}. The evidence methods implemented were not capable of resolving such differences in resolution at those excitation wavelengths. This can be seen in Figure 5.10, under ideal conditions, $\eta = 0$, where this difficulty is demonstrated in our detection at the respective emission spectrums. The spectroscopy model only produced one spectrum for a chemical's absorption and respective excitation. We overcame this obstacle with our unique

development of a synthetic correlated noise algorithm to produce additional testing data under various conditions.

The performance of the evidence methods (using the highest correlation coefficient as a match) seen in Figures 5.9 and 5.10, show that the best methods, respectively: Cepstral, Derivative, Matched Filtering and Filter Banks. This performance of the evidence methods is more apparent in Table 5.1, where the detection accuracy over the 12 spectra is averaged with respect to individual evidence methods. In Table 5.2, the uncertainty with respect to each individual evidence source across spectra has a small variance. The design for the assignment of the uncertainty was based on the correlation coefficient's magnitude and distance relative to the entire set of coefficients. The assignment of the uncertainty could be thought as being associated to how the correlations coefficients are distributed within the set. This association of uncertainty and the correlation coefficient's distribution is shown in Figure 5.15, where three evidence methods (Cepstral, Filter Bank, and Derivative) are presented from the emission spectrum at λ_{425}. This respective uncertainty for these methods were .986, .807, and .534. In Figure 5.15, the Cepstral evidence distribution is favored in reporting only high correlation coefficient values. Hence, the uncertainty is the high due to the conflict. Likewise, as the distributions of the correlation coefficients tend to spread, yielding less conflict, the uncertainty decreases in the cases of the filter bank and derivative evidence. Since, there is a small variance between the uncertainty across spectra with respect to an individual evidence method. We can deduce that the evidence method has strong association with distinguishing conflict within the set and creating distinctive features to classify individual spectrums from the set is the integral component to reduce the uncertainty. This also highlights that over the set of spectra presented there not an ideal spectrum that best differentiates a chemical better then another spectrum.

In figures 5.11, 5.12, and we can note, that regardless of incorrect information from different spectra using different evidence methods when the dempster combination rule is implemented to the data for fusion we can obtain a more accurate detection

rate. However, in Figures 5.16 and 5.17 we can note that regardless of the *eta* presented after first initial couple of fusions over the spectra their detection improvements in the algorithm are limited. We can infer from this that all 11 DCR fusion are unnecessary for the improvement in the detection. The proposed DS Theory method, under the most extreme noise conditions,$\eta = 2.0$, has shown a detection accuracy of 99 percent at Rank 1, using solely the cepstral based evidence source. This is an increase of 46.3 percent detection when compared to choosing the Rank 1 averaged correlation coefficient over the 12 spectra and 1500 simulations. Cepstral was the largest detection gain when $\eta = 2.0$, when compared to the other evidence methods. The sole implementation of the Cepstral based evidence source using Dempster-Shafer Theory has proven the best indictor of detecting analytes in this study. This is possibly due to its robustness against noise and by taking the log of the spectra to highlight subtle changes from underlying chemicals. Despite the high level of detection, the uncertainty remains high at .90. However, there are other fusion combinations of evidence that provide low uncertainty and suitable levels of detection above .90 at Rank 5.

In future studies, different scaling schemes and implementations of logarithmic functions should be applied to handle such resolutions issues. For the purposes of water borne toxins and water safety, this method has merit for the groundwork as an initial alert system in the detection of chemicals in a water sources. By examining the concurrent ranks, we can achieve above 90 percent detection of the top five most probable chemicals, with a low uncertainty. In addition, further work could be added for the decision process of the DS Theory Model. Such information as pH, turbidity, and electrical conductivity are all feasible and insightful information sources that could provide additional fusion techniques for the future as chemical indicators.

Bibliography

[1] J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*. Springer Science and Business Media, 2006. ix, 15, 16, 17, 18

[2] M. Allen, "Stray light-measurement and effect on performance in uv-visible spectrophotometry," Thermo Fisher Scientific, Madison, WI, USA, TechnicalNote 51170, July 2012. ix, 16, 17

[3] G. Tomasi, "Practical and computational aspects in chemometric data analysis," Ph.D. dissertation, The Royal Veterinary and Agricultural University, May 2006. 2

[4] C. Anderson and R. Bro, "Practical aspects of parafac modeling of fluorescence excitation-emission data," *Journal of Chemometrics*, vol. 17, pp. 200–515, 2003. 2, 42

[5] R. Harshman and M. E. Lundy, "Parafac: Parallel factor analysis," vol. 18, pp. 39–72, 1994. 2

[6] S. Matero, "Chemometric methods in pharmaceutical tablet development and manufacturing unit operation," Ph.D. dissertation, University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, FINLAND, June 2010. 2

[7] B. R. Mathon, "Assessing uncertainty assocaiated with groundwater and watershed problems using fuzzy mathematics and generalized regression neural networks," Ph.D. dissertation, The University of Vermont, 2011. 2

[8] L. P. Swiler, T. L. Paex, and R. L. Mayes, "Epistemic uncertainity quantification tutorial," *Proceedings of the IMAC-XXVII*, 2009. 2, 45

[9] K. Sentz and S. Ferson, "Combination of evidence in dempster-shafer theory," SAND, Binghamton University, P.O. Box 6000 Binghamton, NY 13902-6000, April 2002. 2, 45, 46, 52

[10] R. K. Poole and U. Kalnenieks, *Spectrophotometry and Spectrofluorimety*. Oxford, 2000, ch. Fluroscence principle and measurement, p. 31. 3, 5, 50

[11] H. Du, R.-C. A. Fuh, J. Li, A. Corkan, and J. S. Lindsey, "Photochemcad. a computer-aided design and research tool in photochemistry and photobiology," *Photochem.Photobiology*, vol. 68, pp. 141–142, 1998. 8

[12] J. Dixon, M. Taniguchi, and J. Lindsey, "Photochemcad two. a refined program with accompanying spectral databases for photochemical calculations," *Photochem.Photobi*, vol. 81, pp. 212–213, 2005. 8

[13] M. Taniguchi, H.-J. Kim, J. K. S. D. Ra, C. Kirmaier, E. Hindin, J. R. Diers, S. Prathapan, D. F. Bocian, D. Holten, and J. S. Lindsey, "Synthesis and electronic properties of regioisomerically pure oxochlorins," vol. 67, pp. 7329–7342, 2002. 10

[14] J. Strachan, D. F. OShea, T. Balasubramanian, and J. S. Lindsey, "Rational synthesis of meso-substituted chlorin building blocks," vol. 65, pp. 3160–3172, 2000. 10

[15] I. V. Sazanovich, C. Kirmaier, E. Hindin, L. Yu, D. Bocian, J. S. Lindsey, and D. Holten, "Structural control of the excited-state dynamics of bis(dipyrrinato)zinc complexes: self-assembling chromophores for light-harvesting architecture," vol. 126, pp. 2664–2665, 2004. 10

[16] E. Zass, H. P. Isenring, R. Etter, and A. Eschenmoser, "Der einbau van magnesium in liganden der chlorophyll-reihe mit (2,6-di-t-butyl-4-methylphenoxy)magnesiumjodid," vol. 63, pp. 1048–1067, 1990. 10

[17] S. I. Yang, J. Seth, J.-P. Strachan, S. Gentemann, D. Kim, D. Holten, J. S. Lindsey, and D. F. Bocian, "Ground and excited state electronic properties of halogenated tetraarylporphyrins: Tuning the building blocks for porphyrin-based nanostructures," vol. 3, pp. 117–147, 1999. 10

[18] J. S. Lindsey and J. N. Woodford, "A simple method for preparing magnesium porphyrins," vol. 34, pp. 1063–1069, 1995. 10

[19] J. R. Miller and G. D. Dorough, "Pyridinate complexes of some metallo-derivatives of tetraphenylporphine and tetraphenylchlorin," vol. 74, pp. 3977–3981, 1952. 10

[20] J. P. Strachan, S. Gentemann, J. Seth, W. A. Kalsbeck, J. S. Lindsey, D. Holten, and D. F. Bocian, "Effects of orbital ordering on electronic communication in multiporphyrin arrays," vol. 119, pp. 11 191–11 201, 1997. 10

[21] S. Prathapan, I. Yang, J. Seth, M. A. Miller, D. F. Bocian, D. Holten, and J. S. Lindsey, "Synthesis and excited-state photodynamics of perylene-porphyrin dyads. 1. parallel energy and charge transfer via a diphenylethyne linker," *Journal of Physical Chemistry*, vol. B 105, pp. 8237–8248, 2001. 11

[22] K. Tomizaki, R. S. Loewe, C. Kirmaier, J. K. Schwartz, J. L. Retsek, D. F. Bocian, D. Holten, and J. S. Lindsey, "Synthesis and photophysical properties of light-harvesting arrays comprised of a porphyrin bearing multiple perylene-monoimide accessory pigments," *Journal of Organic Chemistry*, vol. 67, pp. 6519–6534, 2002. 11

[23] W. Reusch. Michigan state university: Visible and ultraviolet spectroscopy @ONLINE. [Online]. Available: http://www2.chemistry.msu.edu/faculty/reusch/VirtTxtJml/intro1.htm 11

[24] A. G. Szabo, *Spectrophotometry and Spectrofluorimety.* Oxford, 2000, ch. Fluroscence principle and measurement, p. 40. 12, 13

[25] A. Leon-Garcia, *Probability,Statistics and Random Processes for Electrical Engineering.* Pearson Prentice Hall, 2008. 14

[26] H. Lam, "Performance of uv-vis spectrophotometers," 2012. 15

[27] R. Barnes, M. Dhanoa, and S. Lister, "Standard normal variate transformation and de-trending of near-infared diffuse reflectance spectra," vol. 43, pp. 772–777, 1989. 16

[28] J. A. Howell, *Handbook of Instrumental Techinques for Analytical Chemistry.* Prentice Hall PTR, 1997, ch. Ultraviolet and Visible Molecular Absorption Spectrometry. 36

[29] H. Mark and J. Workman, *Statistics in Spectroscopy.* Academic Press, 1991. 37

[30] J. N. Miller and J. C. Miller, *Statistics and Chemometrics for Analytical Chemistry.* Pearson Education Limited, 2005. 38, 43

[31] J. H. Miyawa and S. G. Schulman, *Handbook of Pharmaceutical Analysis.* Marcel Dekker, Inc., 2002, ch. Chap5: Ultraviolet-Visible Spectroscopy. 38, 40

[32] L. Lang, *Absorption Spectra in the Ultraviolet and Visible Regions.* New York: Academic Press, 1961. 40

[33] S. R. Laboratories, *Ultraviolet Reference Spectra.* Philadelphia:Sadtler Research Laboratories, Updated at regular intervals. 40

[34] R. Harshman, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, pp. 1–84, 1970. 42

[35] E. Lefevre, P. Vannorenberghe, and O. Colot, "Using information criteria in dempster-shafer basic belief assignment," *IEEE International Fuzzy Systems Conference Proceedings*, no. I, pp. 173–178, August 1999. 45

[36] D. Dewasurendra, P. Bauer, and K. Premaratne, "Distributed evidence filtering in networked embedded systems," *In Networked Embedded Sensing and Control Lecture Notes in Control and Information*, pp. 173–178, 2010. 45

[37] A. Gelman, "The boxer, the wreslter and the coin flip: A paradox of robust bayesian inference and belief functions," *The American Statistician*, vol. 60, no. 2, pp. 146–150, May 2006. 45

[38] G. Fioretti, "A mathematical theory of evidence for g.l.s. shackle," *Mind and Society*, vol. 2, pp. 77–97, 2001. 45

[39] M. Wafa and E. Zagrouba, "Estimation of mass function in evidence theory for fusion of gray level based images," *The International Conference on Signal and Electronic Systems*, September 2010. 46

[40] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Information Fusion*, pp. 149–162, 2002. 46

[41] P. K. Black, "Is shafer general bayes," *Int. J. Approx. Reasoning*, 1988. 47

[42] H. Gu and S. Tasi, "A discrete-cepsturm based spectrum-envelope estimation scheme and its example of application of voice transformation," *Computional Linguistics and Chinese Language Processing*, vol. 14, no. 4, pp. 363–381, December 2009. 48

[43] O. Cappe and E. Moulines, "Regularization techniques for discrete cepstrum estimation," *IEEE Signal Processing Letters*, vol. 3, no. 4, pp. 100–102, April 1996. 48

[44] S. Singh and E. Rajan, "Vector quantization approach for speaker recognition using mfcc and inverted mfcc," *International Journal of Computer Applications*, vol. 17, no. 1, March 2011. 50

[45] M. J. Adams, *Chemometrics in Analytical Spectroscopy.* The Royal Society of Chemistry, 1995. 50

Appendix A

Chapter 4 Proofs

A.1 Proof: Maximum Mass Measure

Consider the correlation coefficient vector

$$\mathbf{V} = \begin{bmatrix} V_1 & V_2 & \cdots & V_N \end{bmatrix}^T, \text{ where } V_i \in [0,1], \ \forall i \in \overline{1,N}. \tag{A.1}$$

Generate the following matrix associated with $\mathbf{V}$:

$$\Delta\mathbf{W} = \begin{bmatrix} V_1\Delta V_{11} & V_2\Delta V_{21} & \dots & V_N\Delta V_{N1} \\ V_1\Delta V_{12} & V_2\Delta V_{22} & \dots & V_N\Delta V_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ V_1\Delta V_{1N} & V_2\Delta V_{2N} & \dots & V_N\Delta V_{NN} \end{bmatrix}, \tag{A.2}$$

where $\Delta V_{ij} = (V_i - V_j)$. Next, generate a column weights vector, $\mathbf{C}$, by summing the elements of each column of $\Delta\mathbf{W}$:

$$\mathbf{C} = \begin{bmatrix} N\,V_1^2 - V_1\sum_{i=1}^{N} V_i & N\,V_2^2 - V_2\sum_{i=1}^{N} V_i & \dots & N\,V_j^2 - V_j\sum_{i=1}^{N} V_i & \dots & N\,V_N^2 - V_N\sum_{i=1}^{N} V_i \end{bmatrix}. \tag{A.3}$$

Let the summation of the elements in vector $\mathbf{C}$ be Mm. Then,

$$Mm = \sum_{j=1}^{N} C_j = \sum_{j=1}^{N} \left(N\, V_j^2 - V_j \sum_{i=1}^{N} V_i \right) = N \sum_{j=1}^{N} V_j^2 - \left(\sum_{j=1}^{N} V_j \right)^2 . \tag{A.4}$$

We now consider the following problem:

$$\text{find } Mm_{\max} = \max \left\{ Mm : \sum_{i=1}^{N} V_i \text{ is kept constant with } V_i \in [0,1],\ \forall i \in \overline{1,N} \right\}. \tag{A.5}$$

To address this problem, let us 'separate' out the two arbitrary terms V_k and V_ℓ from Mm:

$$Mm = N \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i^2 + V_k^2 + V_\ell^2 \right) - \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i + V_k + V_\ell \right)^2 . \tag{A.6}$$

Suppose, for some $\rho \geq 0$, we change V_k to $\widehat{V}_k = V_k - \rho$ and change V_ℓ by the same amount to $\widehat{V}_\ell = V_\ell + \rho$, while ensuring that the updated values of $\widehat{V}$ and $\widehat{V}_\ell$ remain bounded in $[0,1]$. We will refer to this 'value migration' from V_k to V_ℓ as *skewing from V_k to V_ℓ*. Then the resulting expression corresponding Mm yields

$$\begin{aligned} \widehat{Mm} &= N \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i^2 + \widehat{V}_k^2 + \widehat{V}_\ell^2 \right) - \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i + \widehat{V}_k + \widehat{V}_\ell \right)^2 \\ &= N \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i^2 + (V_k - \rho)^2 + (V_\ell + \rho)^2 \right) - \left(\sum_{\substack{i=1 \\ i \neq k,\ell}}^{N} V_i + (V_k - \rho) + (V_\ell + \rho) \right)^2 . \end{aligned} \tag{A.7}$$

So, the change in value of Mm due to skewing from V_k to V_ℓ os

$$\widehat{Mm} - Mm = 2N\rho\,(V_\ell - V_k) \geq 0 \iff V_\ell \geq V_k. \tag{A.8}$$

Note that, in order to ensure that $\widehat{V}_k \in [0,1]$ and $\widehat{V}_\ell \in [0,1]$, we must have

$$\widehat{V}_k = V_k - \rho \geq 0 \text{ and } \widehat{V}_\ell = V_\ell + \rho \leq 1 \iff \rho \leq V_k \text{ and } \rho \leq 1 - V_k \iff \rho \leq \min\{V_k, 1 - V_\ell\}. \tag{A.9}$$

So, we must upper bound ρ by

$$\rho_{\max} = \min\{V_k, 1 - V_\ell\}. \tag{A.10}$$

From (A.8), it is clear that $\widehat{Mm}$ achieves a maximum when $\rho = \rho_{\max}$. Note the following:

$$\begin{aligned} \text{If } \rho_{\max} = V_k: \quad & \widehat{V}_k = V_k - V_k = 0; \\ & \widehat{V}_\ell = V_\ell + V_k; \\ \text{If } \rho_{\max} = 1 - V_\ell: \quad & \widehat{V}_k = V_k - 1 + V_\ell = (V_k + V_\ell) - 1; \\ & \widehat{V}_\ell = V_\ell + 1 - V_\ell = 1. \end{aligned} \tag{A.11}$$

So, skewing from V_k to V_ℓ to maximize $\widehat{Mm}$ renders either V_k to reach 0 (when $\rho_{\max} = V_k$) or V_ℓ to reach 1 (when $\rho_{\max} = 1 - V_\ell$).

To determine $Mm_{\max}$ in (A.5), we may continue the above process as follows:

I. Pick a pair $\{V_i, V_j\}$, where $V_i \neq 0$ and $V_j \neq 1$.

II. Skew from $\min\{V_i, V_j\}$ to $\max\{V_i, V_j\}$ if $V_i < V_j$, or skew from V_i to V_j if $V_i = V_j$, until either V_i reaches 0 or V_j reaches 1.

III. Repeat Step I until no pair $\{V_i, V_j\}$, where $V_i \neq 0$ and $V_j \neq 1$, can be found. The resulting vector $\mathbf{V}_{\max}$ generates the maximum value $Mm_{\max}$. This vector $\mathbf{V}_{\max}$ takes one of two forms:

III.a. All entries of $\mathbf{V}_{\max}$ are either 0 or 1.

III.b. All entries, except one entry V_r where $V_r \in (0,1)$, of $\mathbf{V}_{\max}$ are either 0 or 1.

Suppose the number of 1s in $\mathbf{V}_{\max}$ is K_1. Then,

$$Mm_{\max} = N\left(\sum_{j=1}^{K_1} 1^2 + V_r^2\right) - \left(\sum_{j=1}^{K_1} 1 + V_r\right)^2 = N(K_1 + V_r^2) - (K_1 + V_r)^2. \tag{A.12}$$

Note that

$$K_1 = \text{Integer part of } \sum_{i=1}^{N} V_i; \quad V_r = \text{Fractional part of } \sum_{i=1}^{N} V_i. \tag{A.13}$$

Moreover, the elements of $\mathbf{V}_{\max}$ need not be ordered in a particular manner.