Investigating Systems of Varying Complexity Using Multidimensional Fluorescence and Parallel Factor Analysis

A dissertation submitted by

Joseph Chiarelli

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Chemistry

TUFTS UNIVERSITY

August 2017

Advisor: Professor Jonathan E. Kenny

Abstract

Steady-state fluorescence spectroscopy is used with Parallel Factor Analysis (PARAFAC) to investigate several systems of varying complexity. PARAFAC has been shown to be a useful tool to investigate multicomponent fluorescence systems. In Chapter 1, an introduction to molecular fluorescence and PARAFAC is discussed.

In Chapter 2, a novel method to estimate the fluorescence lifetimes in mixtures of five polycyclic aromatic hydrocarbons (PAHs): fluorene, pyrene, triphenylene, anthracene, and phenanthrene is presented. Score values obtained from oxygenated and deoxygenated samples in the analysis are used to determine Stern-Volmer constants. The determined constants are used to estimate fluorescence lifetimes of each PAH. An average quenching rate constant is determined using information from literature sources, to estimate the fluorescence lifetime. The method is applied to an experiment with gasoline to identify fluorescent components.

In Chapter 3, PARAFAC with soft independent modeling by class analogy (PARAFAC-SIMCA) was used to analyze fluorescence data from shrimp extracts to create geographic classification schemes for shrimp from three different locations on the Ecuadorian coast (Pedernales, Cojimies, Tonchigüe). The extraction process was improved from previous work with the use of microwave assisted extraction (MAE). This improved the processing rate from 1 sample per day to 12 samples per day, with minimal changes to other procedures. Each shrimp was extracted into aqueous and organic solvents. PARAFAC modeling of the fluorescence data produced a 3-component model for the aqueous samples and a 4-component model for the organic samples. Using PARAFAC scores from all the aqueous and some organic components, shrimp samples were classified by site at the 95% confidence level for 91% of samples.

In Chapter 4, cyclodextrin complexes with polycyclic aromatic hydrocarbons (PAHs) are investigated. Proposed is a method that allows for us to produce separate spectra and concentration data for free molecules and the complexes. This work specifically focuses on several complexes of β-cyclodextrin with naphthalene and anthracene. Specifically, 1:1 complexes and 2:2 complexes are investigated. Score values are used to calculate equilibrium constants for the various complexes and compared to literature data. The method is shown to be able to investigate mixtures of multiple PAHs with cyclodextrins simultaneously.

Acknowledgements

I have spent the better part of a decade reporting for duty to the Pearson basement, and I have had the good fortune of meeting many great people and making memories that I will look back on fondly for many more years to come.

First, I would like to thank my research advisor and fearless leader, Prof. Jonathan Kenny. It has been a pleasure to work with you the last eight years. Thank you for helping me discover my passion for teaching and preparing me for the next steps in my career. Next, I would like to thank my thesis committee. Prof. Arthur Utz, I had the pleasure of teaching with you several semesters ago, and I have always enjoyed the dialogue we have continued since then. Prof. Sam Thomas, I always enjoyed the classes I took with you and the interactions we shared from time to time outside of it. I would like to give a special thanks to my outside committee member Prof. Thandi Buthelezi of Wheaton College. Thank you for taking time out of your schedule to attend my defense. It has been a joy to work with you since we met during your sabbatical. You were a great mentor to me at Wheaton College this past year, and it is an opportunity I will always be grateful for.

The Kenny lab may be a small squad, but we are a proud one. I had the pleasure of working side-by side with many great individuals over the years. Mike, Jake, Chelsea, Stacey, Ben, Ashley, Alina, and Jae Woo, I learned a great deal working with you. Some of the work in this document would not be possible without your efforts. I would like to give special thanks to Tom Brownrigg. You were always a good friend and someone I could talk science with.

To my friends and family, you were all more supporting, and sometimes forgiving, than I could ask for. While getting through graduate school is very much an independent venture, I never once felt alone. To my fellow graduate students, you were all so supportive during my time here. Whether it was preparing for exams, practicing talks, or just going to support a local business in Davis Square, you were always there. I will especially miss my fellow basement dwellers and the community we share. To my parents, Jonathan, Nick, and Kaley, thank you for your support this whole time. I am sure I will see you this Sunday for dinner, no matter where or when you are reading this. Finally, to my lady friend Alyssa, thanks for helping me see the way to the finish line.

Table of Contents

List of Figures

List of Tables

List of Common Abbreviations

EEM ……. Excitation Emission Matrix

HOMO ……. Highest Occupied Molecular Orbital

IMP ……. Identity Match Plot

LUMO ….. Lowest Unoccupied Molecular Orbital

NaN …….. Not a number

PARAFAC ……. Parallel Factor Analysis

PIFE ……. Primary Inner Filter Effect

RIP……. Resample Influence Plot

SIFE ……. Secondary Inner Filter Effect

SSR ……. Sum of Squared Residuals

TCC ……. Tucker Congruency Coefficient

UV-Vis ……. Ultra Violet/Visible

Investigating Systems of Varying Complexity Using Multidimensional Fluorescence and Parallel Factor Analysis

Chapter 1: Introduction

1.1 Molecular Luminescence

The distribution of energy in a molecule can be described by the following expression:

$$E_{molecule} = E_{trans} + E_{rot} + E_{vib} + E_{elec} \quad \textbf{(1-1)}$$

The energy in a molecule is distributed among the available states. At room temperature conditions used in this work, the energy distribution for most molecules is well understood. The energy spacings between translation states is the smallest, much less than 1 cm^{-1}, and at these conditions many translational states are populated. Rotational energy spacings are further apart, about 1 cm^{-1}, but still populated enough at room temperature where several states are populated. Vibrational states are much further apart, on the order of 1000 cm^{-1}, and most of the molecules in a population occupy the ground vibrational state. Electronic states are extremely far apart, greater than 10,000 cm^{-1}, and all molecules are in the ground electronic state at room temperature. For a molecule to have sufficient energy to occupy upper electronic states, energy must be put into the system. This will concern the primary method for investigating chemical systems in this work which is molecular fluorescence.[1]

The relevant processes that are related to molecular fluorescence are summarized in Figure 1-1.

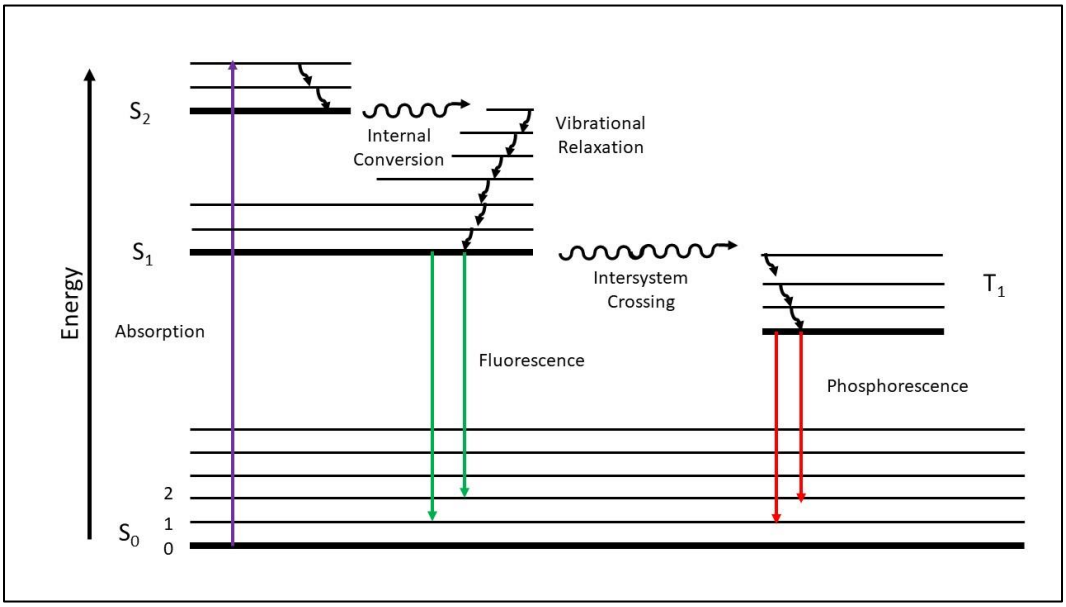


Figure 1-1: A Jablonski diagram describing the relevant photo-physical processes involved in this work. Radiative processes are indicates by straight arrows, while non-radiative processes are indicated by wavy arrows.[2]

This discussion will start with a molecule in the ground electronic state and ground vibrational state. We can describe the ground electronic state as a singlet state, S_0. A singlet state is an electronic state in which the two highest energy electrons occupy the highest unoccupied molecular orbital (HOMO) with spins paired in the ground state. From S_0, molecules can absorb light and have sufficient energy to occupy an upper electronic state as indicated by the purple arrow. This is a very fast process on the order of 10^{-15} s.[2] In the figure, the molecule is excited from the ground state, S_0, to the second excited singlet state, S_2. While the transition starts in the lowest vibrational state in the ground state S_0, it may be excited to other vibrational states in S_2 (v=0,1,2…) when it absorbs a

photon. Which vibrational state the molecule is excited to is based on Frank-Condon factors.[3] Once in the excited state, the molecule has different processes that allow it to relax back down to its ground state. These processes can be described as either radiative processes or non-radiative processes.

Regarding non-radiative processes, an excited molecule can transition to a lower energy state by vibrational relaxation, internal conversion, or intersystem crossing.[2] Vibrational relaxation, indicated by the short curved line in Figure 1-1, occurs due to the collisions between the excited molecule and other surrounding particles. This may be other molecules, solvent, or the walls of the container. Vibrational relaxation occurs at a very fast rate, about 10^{12} s^{-1}. The energy is released through inelastic collisions (via heat) until the molecule is in the ground vibrational state of that electronic state.

Internal conversion results in a molecule relaxing from an excited electronic state to a lower electronic state of the same spin multiplicity (i.e. singlet to singlet) in a non-radiative process. This is shown by the wavy black line in Figure 1-1 between the S_2 and S_1 states. This transition can happen between other states as well, but it is only shown once for simplicity.

In Figure 1-1, intersystem crossing involves the non-radiative transition of a molecule from the lowest vibrational level of S_1 to a degenerate vibrational level of the triplet state T_1. The triplet state is an electronic state where the electron spins are not paired. While this transition is considered forbidden, but it does happen for many molecules.[4] The lowest vibrational level of the triplet state is

always lower in energy than the lowest excited singlet state. Further vibrational relaxation results in some of the energy being lost in a non-radiative process.

The final way a molecule can relax that will be discussed here is through a radiative process, or emission of light. Two processes that are radiative are fluorescence and phosphorescence. These are indicated by the green and red arrows, respectively, in Figure 1-1. Fluorescence is almost always a result from a molecule in the first singlet excited state, according to Kasha's rule.[5] A famous exception to this rule is azulene, where the fluorescence originates from the S_2 excited state.[6] Phosphorescence is an analogous process from an excited triplet state. It is always lower in energy than the lowest excited singlet state of the same electron configuration, so longer wavelengths are observed for phosphorescence, than compared to the fluorescence of a molecule.

Concerning fluorescence, the focus in this work, it is important to emphasize a few points. At room temperature in solution, conditions dictate that fluorescence always happens from the same singlet excited state, independent of the excitation process. This results in the excitation spectrum of a molecule being independent of its emission spectrum. Also, due to the various nonradiative processes after absorption and prior to fluorescence, the photons resulting from fluorescence are always lower than or equal to the energy of the photons required for the molecule to absorb light. This results in emission spectrum always occurring at longer wavelengths, when compared to the excitation spectrum. These points will be important for the types of data analysis discussed later in this chapter.

1.2 Excitation-Emission Matrix (EEM)

The basic data unit in this work is the excitation-emission matrix (EEM). An EEM can be collected by exciting a sample at several different wavelengths, while measuring a range of emission wavelengths at each of these excitation wavelengths. This results in a three-dimensional spectrum that is ideal for samples containing more than one fluorophore. The three dimensions are excitation wavelength, emission wavelength and the intensity of fluorescence. For the EEM to be useful for the types of data analysis we will use, the EEMs we collect must meet certain criteria.

For a single fluorophore, the fluorescence intensity of a dilute solution can be described by equation (1-2).[7]

$$EEM_{ij} = 2.303\phi_f I_0(\lambda_i)\epsilon(\lambda_i)bcN(\lambda_j)\kappa(\lambda_j) \qquad (1\text{-}2)$$

In the above expression, $I_0(\lambda_i)$ is the intensity of the monochromatic excitation light at wavelength *i* directed at the sample. The factor $2.303\epsilon(\lambda_i)bc$ represents the optical density of the sample.[2] It is a product of the compounds molar extinction coefficient at a given wavelength $\epsilon(\lambda_i)$, the pathlength b, and the concentration of the fluorophore *c*. The quantum yield of fluorescence (ϕ_f) the ratio of photons emitted versus photons absorbed by the molecule. The fraction of fluorescence photons emitted at wavelength *j* is described by $N(\lambda_j)$. The instrumental influence on emission at wavelength *j* is described by $\kappa(\lambda_j)$.[7]

Assuming a dilute solution means the factor describing the optical density is much less than 1. In this case we can simplify expression (1-2) to get,

$$EEM_{ij} = \alpha_i \beta_j \gamma$$

(1-3)

where $\alpha_i = I_0(\lambda_i)\epsilon(\lambda_i)$, $\beta_j = N(\lambda_j)\kappa(\lambda_j)$ and $\gamma = 2.303\phi_f bc$. For a set of wavelengths, α_i can be represented as a vector that describes the observed relative excitation spectrum of the molecule, β_j can be represented as a vector that describes the observed relative fluorescence emission spectrum, and γ is a wavelength independent factor that contains all concentration information.[7]

Using expression (1-3), for an optically dilute solution with multiple fluorophores *r* we can state:

$$EEM_{ij} = \sum_{r=1}^{R} \alpha_{ir} \beta_{jr} \gamma_r$$

(1-4)

The EEM of a multicomponent solution of fluorophores, when optically dilute, results in the total fluorescence of the sample being a sum of individual contributions from each fluorophore.[7]

With optically dilute solutions, EEMs present several benefits over single two-dimensional fluorescence spectra. First, an EEM contains spectral qualities of all the fluorophores in the measured wavelength region. Some parts of the spectrum have signal from all the components in the sample, while some parts of the spectrum may only have signal from one component or no signal at all. An

example of an EEM is shown in Figure 1-2. The sample measured was made up of five different fluorophores: anthracene, fluorene, pyrene, triphenylene, and phenanthrene.

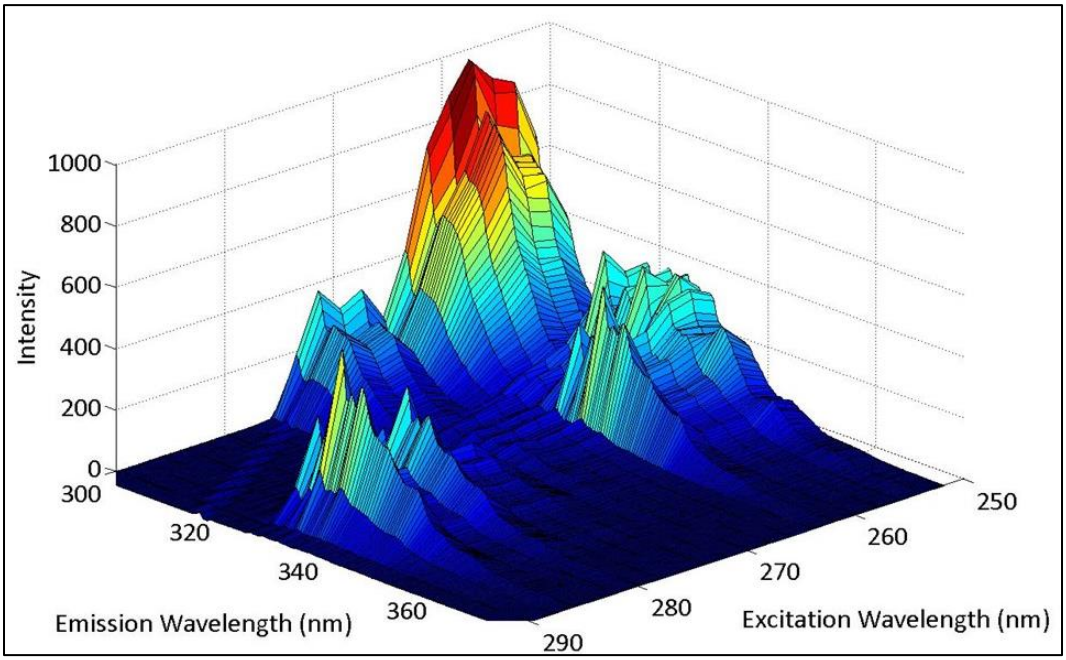


Figure 1-2: Excitation Emission Matrix (EEM) containing anthracene, fluorene, pyrene, triphenylene, and phenanthrene.

Second, the increased amount of signal from each sample can also have advantages for samples with multiple components. While many types of compounds share similarities in various regions of their excitation or emission spectra, rarely are the spectra near identical. EEMs allow for more wavelength regions to be measured to account for these differences. This has proven beneficial to areas of study that involve many different kinds of fluorophores such as: the study of dissolved organic matter (DOM) in natural waters,[8] monitoring the impact of sewage in rivers[9], or determination of pesticides.[10]

1.3 Multi-way analysis

The type of analysis we use to evaluate EEMs falls under the field of multi-way analysis. Simply put, multi-way analysis techniques describe a data set of several different dimension types. A single EEM already introduces two dimensions to our data set, emission and excitation wavelengths. The types of datasets differentiate multiway methods. One way of differing one dataset from another is by the number of dimensions the data set is comprised of, or ways. In general, we can refer to a dataset as having "N-ways" where N is the number of dimensions for the data set. Each way corresponds to a different type of variable.[11] In this work we will use 3-way datasets. Two ways are provided by the EEMs we collect, as previously stated. The third way is the number of EEMs we collect in a dataset. The third way is generally the sample number of the EEM, but it can correspond to other experimental variables such as the pH at which the sample EEM was collected.[12]

Multi-way analysis methods can be used in two different ways: exploratorily or confirmatory. Exploratory use of multi-way methods has the objective of developing a hypothesis to explain the problem at hand.[11] Confirmatory use of multi-way methods are used to validate or investigate an already constructed hypothesis.[11] While there are many different methods for multi-way analysis; the primary multi-way method used in this work is PARallel FACtor analysis (PARAFAC). The reason PARAFAC is used in EEM analysis is because it is constrained in a way that produces models that can be chemically interpretable.[13]

1.3.1 PARAFAC

The first papers to propose the use of PARAFAC model were published in 1970.[14,15] It wasn't until the 1990's that PARAFAC was used as a tool in multidimensional fluorescence studies.[16] The PARAFAC algorithm is used to explain three-way data that can be shown using (**1-5**).

$$\boldsymbol{X_{ijk}} = \sum_{\boldsymbol{r=1}}^{\boldsymbol{R}} \boldsymbol{a_{ir} b_{jr} c_{kr}} + \boldsymbol{e_{ijk}} \qquad (1\text{-}5)$$

Here the dataset of EEMs is represented by X, a three-way data set with dimensions i, j, and k. The dimensions for the data set correspond to excitation wavelength (*i*), emission wavelength (*j*), and sample number (*k*). For a PARAFAC model made up of *R* components, the dataset can be described by a sum of contributions for each individual component r. There are three factors that contribute to the model. The excitation and emission spectra, a_{ir} and b_{jr}, are calculated for each component r. A score value is calculated for each of the *k* samples for each component r. The term e_{ijk} represents the residual data not used in the PARAFAC model.[13]

There are three criteria a dataset must meet to be used with PARAFAC. First, the components of the dataset must not covary. This is based on the rule that no two fluorophores have the exact same spectrum. Second, the components of the dataset must be trilinear. This means the excitation and emission spectra must be independent of one another, as we showed for room temperature solutions, and the fluorescence intensity must increase due to a factor in the parameter γ in

equation (1-3). Third, the fluorescence intensity of the components must be additive, as we showed for an optically dilute solution in equation (1-4).

While the score value is generally considered to be proportional to concentration[17], it can more simply be described as the amount of fluorescence intensity attributed to a particular sample for a component of the model. The contributions to the fluorescence of a fluorophore were shown in equation (1-2). The contributions to the score value of the component will be restated with different variables for better use in later chapters. The properties of the fluorophore that contribute to the score value is shown in equation (1-6)[18].

$$\boldsymbol{c_{kr} = 2.303\phi_r \times [s]_{kr} \times \epsilon_{ir} \times em_{jr} \times f} \qquad (1\text{-}6)$$

Where ϕ_r the fluorescence quantum yield of component r, $[s]_{kr}$ is the concentration of component r in sample k, ε_{ir} is the molar absorptivity constant at excitation wavelength i for component r, em_{jr} is the relative fluorescence intensity at emission wavelength j for component r, and f is a coefficient comprising of instrumental factors.[13,19]

As can be seen in equation (**1-6**), score value and concentration are indeed proportional if all other factors are constant. In many experiments, the concentration of the component is changing for each sample, and it corresponds to a score value change. However, it is not always the reason for a score value change. Changes in ϕ_r or ϵ_{ir} over the range of samples in a dataset could result in score value changes as well. This is the case for some of the work in the following chapters.

Prior to the use of PARAFAC for analyzing multicomponent fluorescence spectra, methods such as rank annihilation required standard spectra in order to be useful in analysis of data.[20] A major benefit of using PARAFAC to analyze multicomponent datasets is that standard spectra are not necessary for the analysis.[13,17] The algorithm can fit spectra for each component in the model simultaneously, with no input needed other than the number of components in the model. This allows PARAFAC to be used in exploratory experiments where the actual fluorescent components are unknown, or the actual spectra are unobtainable otherwise.

1.4 Instrumentation

1.4.1 Steady-State Absorbance

A Varian Cary 300 is used for all absorbance measurements in this work. A tungsten lamp is used for wavelengths greater than 350 nm and a deuterium lamp is used for measurements below 350 nm. All samples are measured from 190 nm to 600 nm in any experiment. All samples are measured using a 1 cm quartz cuvette. Sample measurements are corrected by subtracting a solvent blank. Absorbance spectra of solvents used in samples are routinely measured prior to and during experiments to minimize introduction of any contaminants. All measurements are collected at room temperature, as the instrument does not have temperature control capability. The laboratory is climate controlled, so the temperature does not fluctuate by more than a few degrees on any given day. The lab temperature is normally around 20^{o} C.

1.4.2 Steady State Fluorescence

All fluorescence spectra were measured using a Varian Cary Eclipse fluorimeter. The instrument uses a xenon flash lamp, operating at 80 Hz with a pulse width at half max of about 2 μs, to excite the sample.[21] The excitation light is split, prior to being directed at the sample, and a fraction is directed to a reference detector to account for any fluctuations in the excitation source.[21] Emission is detected at a 90^{o} angle relative to the excitation source.[22,23] This is to limit the amount of scattered light detected by the instrument. Both the emission detector and the reference detector are R928 photomultiplier tubes. The photomultiplier tube has an adjustable bias voltage in the range of 400-1000 V. The useable wavelength range for excitation and emission is 190-600 nm. Samples are normally measured at 25 ^{o}C using a Peltier temperature controlled cuvette holder. Spectra are collected using uncorrected settings.

1.5 Preparation of EEMs for PARAFAC analysis.

Raw spectra of EEMs are not normally used for PARAFAC analysis. EEMs are prepared for modeling in two ways prior to modeling. The scattered light, Rayleigh or Raman, that appears in the fluorescence spectra is removed. EEMs are corrected to be optically dilute. Instrumental biases are corrected.

1.5.1 Removal of Scattered light

Scattered light is a normal interference in EEM collection and cannot be eliminated experimentally using the instrumentation we have available.[22,23] Therefore, it is removed prior to PARAFAC modeling, so it may not interfere with the model determination. Scattered light does not behave trilinearly, and

therefore cannot be modeled by the algorithm.[13] Nearly all of the light scattering in a spectrum is due to the solvent. The spectral signature of scattered light can vary with the excitation and emission wavelength. Rayleigh scattered light shows intensity at the excitation wavelength of the spectrum. Raman scattering shows intensity at a constant frequency difference compared to the excitation frequency.[24] The wavelength of Raman scattering is solvent dependent, but is usually tens of nanometers to the red of the excitation wavelength.[24] The intensity of scattered light also varies depending on the solvent. Raman scattering is less intense when compared to Rayleigh scattered light. Raman scattering is normally not corrected for, but if necessary, it can be removed by using an EEM of the solvent and subtracting it from the sample EEM.[13] The intensity of Rayleigh scattering is greater than Raman Scattering, and is generally on the same order of intensity as the emission intensity from the sample.

Rayleigh scattering is always corrected for in an EEM measurement by either limiting the wavelength range of the EEM or removing it from the spectrum. Scattered light can be in the same wavelength region as emission pertinent to the experiment. Preserving as much actual fluorescence signal is a primary concern when removing scattered light. To remove scattered light from the spectrum, it is changed to "Not-a-number" (NaN). This allows the wavelengths that scattered light appears to be ignored by MATLAB.[13] It should be noted, converting the wavelength values to NaN is not the same as converting them to zero intensity. EEMs are converted to NaN in MATLAB using a program

written by Gregory Hall.[25] An example of the result of scatter removal in an EEM is shown in Figure 1-3.

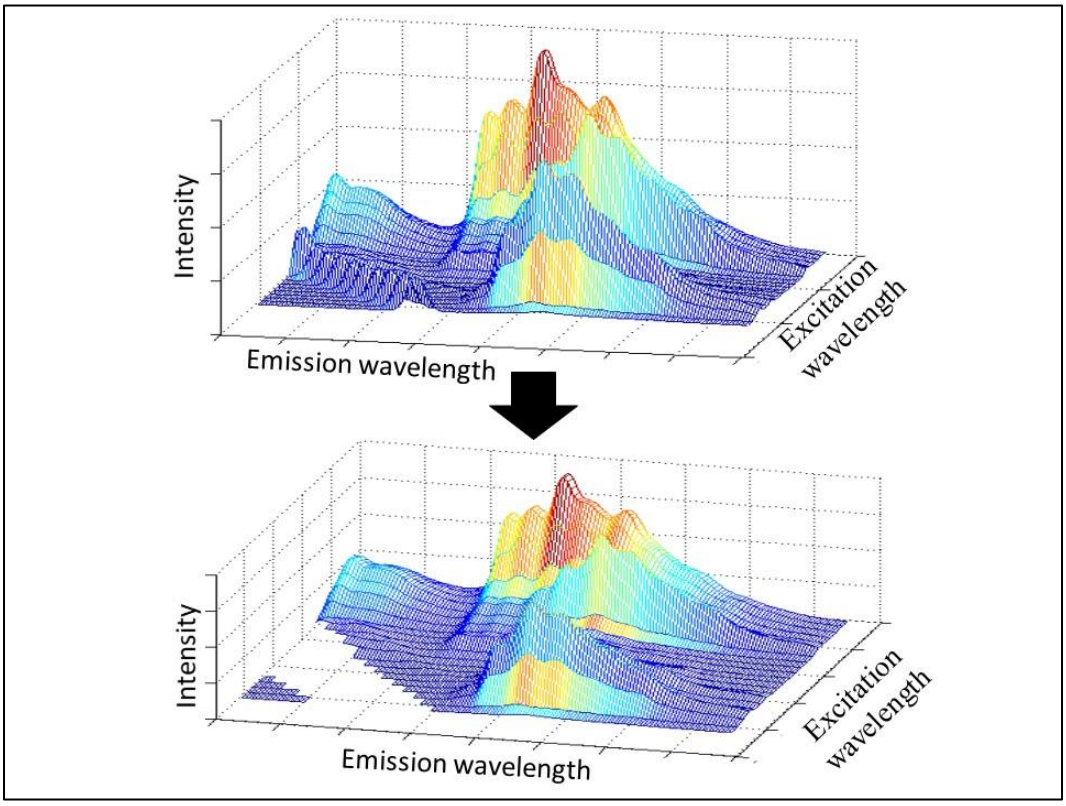


Figure 1-3: Depiction of the scatter removal process used for all EEMS prior to PARAFAC modeling. The section of the EEM containing scattered light is changed to not-a-number (NaN).

1.5.2 Inner Filter Corrections

In order for the fluorescence intensity of all components in a PARAFAC model to be calculated appropriately, the spectra must be optically dilute.[13] In other words, the sample's fluorescence intensity should be free of any influence from other interactions. One of the common complications with fluorescence spectra related to this issue are inner filter effects.[26] Inner Filter effects can be categorized as either primary (PIFE) or secondary (SIFE) inner filter effects. Primary inner filter effects involve the incident light from the excitation source

being absorbed by other molecules in solution before it can reach a fluorophore molecule that the optics of the instrument are focused on. The fluorophore molecule would as a result absorb less light and emit less fluorescence. Secondary inner filter effects concern the fluorescence emitted from a fluorophore molecule. When a photon of fluorescence is absorbed before it can reach the detector, less fluorescence intensity is measured in the spectrum. Both effects result in less fluorescence intensity than should be the case for a measured fluorescence spectrum. Fortunately, these inner filter effects can be corrected for.

Inner filter effects become relevant in measurements for samples that have absorbances higher than 0.05 in the wavelength range of interest.[2,23] Above this absorbance range, samples must be corrected to be optically dilute and the effects of this phenomena are minimized. The corrections applied to EEMs for this work are based on the method proposed by MacDonald et. al.[26] These corrections are implemented in a program initially written by Gregory Hall[25] and later modified by Hao Chen.[27] In general, absorbances for any sample are kept below 1.0 absorbance unit, as this is where the corrections are most effective. [26,28]

1.5.3 Instrumental Bias Corrections

The light source and detection components of the instrument do not perform equally over the wavelength range of the instrument.[13] The Xenon Flash lamp is more intense at certain wavelengths than others and the photomultiplier tube detects certain wavelengths better. There is also variance in the way certain wavelengths are diffracted by the gratings used in monochromators for excitation and emission light. Left uncorrected, the excitation and emission spectra would be

biased by these instrumental factors to reflect these inconsistencies. To account for these biases, spectra were corrected prior to use in PARAFAC analysis. The corrections used for the EEM spectra were rigorously described by Gregory Hall[25] and are based on corrections proposed by Melhuish.[29]

1.6 Modeling data with PARAFAC

1.6.1 Making the model

A PARAFAC model is constructed using an Alternating Least Squares (ALS) algorithm summarized in the following steps.[13] First, the user chooses the number of components to be used in the model. PARAFAC then constructs initial spectral loadings. It then calculates score values for each of the components in each sample. The amount of variance explained by the model is measured, and new estimates for the spectral loadings are then made to ensure a better fit. Score values are calculated for the new spectral loadings. This process repeats itself until the sum of squared residuals is minimized to the point where the change in the sum of squared residuals is below the convergence criteria.

There are several options implemented to ensure proper fitting models. One constraint put on the PARAFAC model is non-negativity constraints on all output. These constraints ensure that calculated spectra and score values cannot be negative. For spectra, there is a practical constraint as intensity cannot be negative if the instrument is calibrated properly. Since the score value is indicative of fluorescence intensity, it cannot be negative either. While the convergence criteria can be changed, it is normally set such that the relative change in fit is no less than 1×10^{-6}. Initialization of the loadings is completed by the calculation of 10

small models, starting with the best fitting loadings. While this setting increases initialization time as the dataset or number of components increases, it aids the constructing of more consistent models and avoids the problem of the algorithm converging on a local minimum.[30]

For any datasets in this work, PARAFAC modeling is used in as unbiased and objective a way as possible. For any PARAFAC model determination, models are constructed starting with one component, and successively constructed with increasing number of components until the user determines the number of components is obviously too high. While this decision is ultimately subjective, we have developed a routine for model determination that makes this a more objective decision.

1.6.2 Preliminary Diagnostics

The first step in evaluating which PARAFAC model is the best fit involves tabulating a few diagnostics common to each model. The table includes the sum of squared residuals, percentage of variance explained, number of iterations, and core consistency diagnostic. Each of these diagnostics is examined from model-to-model and any patterns are taken into consideration when deciding the model of best fit. While none of these diagnostics can independently show which PARAFAC model is a best fit, the table is usually very useful in determining the best model, or at least giving an indication as the models that should be examined more closely.

The aim of any PARAFAC model is to explain as much of the variance as possible within a given dataset, while still producing interpretable results.[13] This

is monitored in the percentage of variance explained and the Sum of squared residuals. Bigger models tend to explain the most variance within a dataset, and generally the best model will explain more than 99.9% of the variance within a dataset. This is reflected in the sum of squared residuals as well. The better models limit the amount of residual data not used in the model. When evaluating which model is best, we observe when the change in percentage of variance explained and the change in the sum of squared residuals is minimized. The best fit PARAFAC model tends to be the model where the change in these diagnostics is comparatively low, when compared to models with fewer components.

Another element of the diagnostic table that is useful is the number of iterations the algorithm used to reach the convergence criteria. The number of iterations essentially describes how difficult it is for the algorithm to converge. When a high spike in the number of iterations for the algorithm to converge is observed, this is usually a sign too many components are being used to fit the model.

The last diagnostic to be evaluated in the diagnostic table is the core consistency diagnostic. This is a diagnostic was proposed for choosing the best fit PARAFAC model.[31] This diagnostic is a good guideline to whether a PARAFAC model is a good fit. Simply put, a high core consistency value equal to or close to 100% indicates the calculated PARAFAC components are trilinear and not covarying, and the model is a good fit. The value for the diagnostic drops abruptly once the number of components chosen is a poor fit.[31]

Core consistency has been shown in our group's work to not always be appropriate in certain experiments involving energy transfer and analysis of shrimp samples.[32] In this example, spectra of components describing the donor-acceptor complex and the free molecules were too similar. Relatively recent papers, one co-authored by the original developer of the diagnostic Rasmus Bro, recommend not using core consistency when more complicated systems are being studied.[17,33]

1.6.3 Evaluation of PARAFAC Spectra

Once the diagnostic table has been evaluated, possible models are further evaluated by examination of the PARAFAC calculated spectra. Spectra for PARAFAC components should appear to have an appropriate shape and wavelength range. The excitation spectrum should have most of the intensity at a shorter wavelength when compared to the emission spectrum. There should only be one band for the emission spectrum, as it should only be a result of a transition from the lowest excited singlet state, as discussed earlier in the section on luminescence. Models that fit spectra not in accordance with these qualities tend to be disregarded.

1.6.4 Evaluation of Residual Data

Using PLS toolbox, evaluation of the residual data is possible. The leftover intensity of each EEM used in the PARAFAC model can be examined visually. At this stage in the evaluation process, most other criteria of the model have been evaluated (Diagnostic Table and calculated spectra) and the residuals are examined to ensure that we are not missing any appreciable intensity in our

model. Normally the resulting residual EEM will have leftover scattered light intensity or random instrumental noise. This is signal that is not trilinear. An example residual EEM is shown in Figure 1-4.

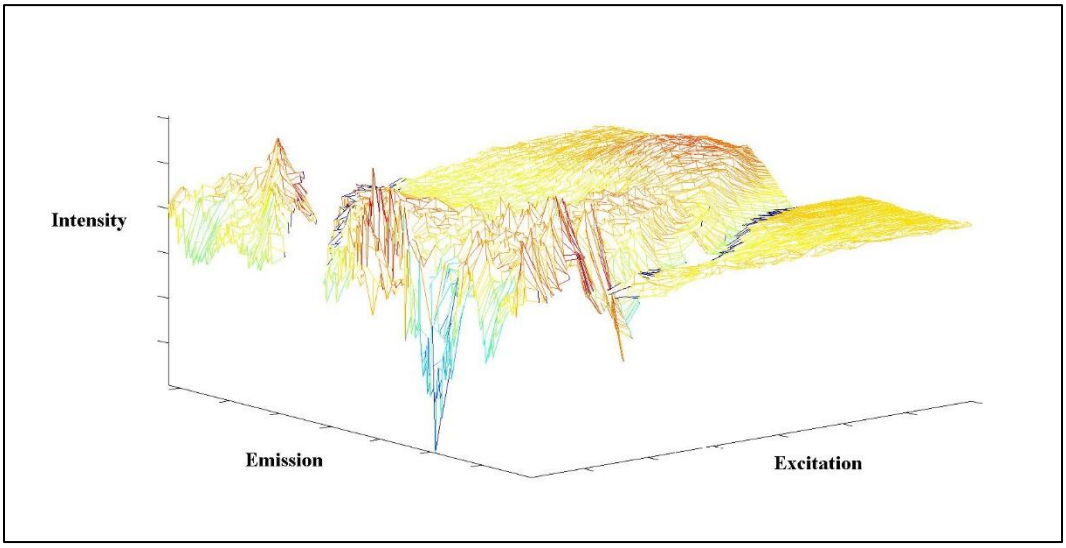


Figure 1-4: Example of an EEM made up of residual data not used in a PARAFAC model. These spectra should not show any intensity that could be discerned as another possible PARAFAC components.

1.6.5 Jack-Knife PARAFAC

The purpose of jack-knife PARAFAC is to examine the influence of a single sample on the chosen PARAFAC model. This method is used to determine outliers in a dataset used for a PARAFAC model.[34] To examine this, jack-knife PARAFAC calculates *k* PARAFAC models, where *k* is the number of samples in the dataset. For each separate jack-knife PARAFAC model, a sample is removed and the model is calculated using the remaining samples. A score value for each component in the model is then calculated for the sample omitted from the

dataset. Jack-knife models are then evaluated using the Resample Influence Plot (RIP) and Identity Match Plot (IMP).[34]

RIP is used for evaluating outliers with respect to the excitation and emission spectra calculated by PARAFAC. In RIP, the sum of squared residuals of the *k*th sample to the model calculated in the *k*th jack-knife model is shown on one axis versus the sum of squares of the difference between the excitation and emission spectra (second or third mode) obtained with each jack-knife segment and the overall PARAFAC model. [34] In a RIP, the data points for a sample will cluster together and any outliers will drift from the cluster. An example of a RIP with an outlier sample is shown in Figure 1-5.

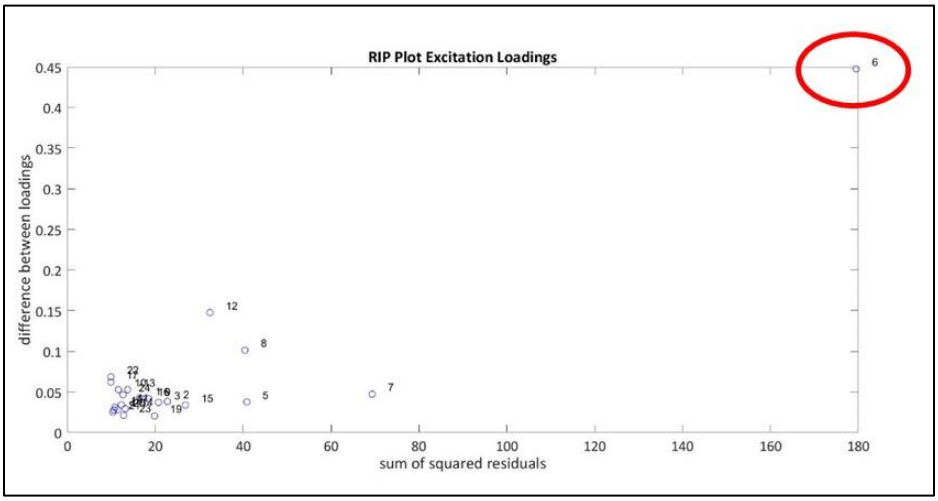


Figure 1-5: Example of a RIP with an outlier. Sample 6 (circled in red) is an outlier.

For evaluating outliers in the model based on the score values calculated by PARAFAC, IMPs are used. IMPs are made by placing the scores calculated by the overall PARAFAC model on one axis and plotting it against the scores predicted for the *k*th sample using the *k*th jack-knife model. If the scores are being predicted properly, the data points are arranged linearly. All data points should

follow a line, called the identity line. Outliers are indicated by deviating from the identity line significantly.

1.6.6 Split Half Analysis

In split-half analysis, a dataset used for PARAFAC modeling is divided into different halves and separate PARAFAC models are calculated for each portion.[35] The models for each half are compared to verify the number of components in the overall model. This is done quantitatively by calculating Tucker Congruency Coefficients (TCC) for each corresponding pair of components in each half's model.[36] If the coefficients are greater than 0.95, the components are considered the same. If both halves of the data set produce the same components separately, and match the components calculated by PARAFAC for the overall dataset, the model is considered validated.

1.7 References

(1) McQuarrie, D. A. *Quantum chemistry*; University Science Books: Mill Valley, Calif., 1983.

(2) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer US: Boston, MA, 2006.

(3) Hollas, J. M. *Modern spectroscopy*, 3rd ed.; J. Wiley: Chichester ; New York, 1996.

(4) Katsumata, T.; Sasajima, K.; Nabae, T.; Komuro, S.; Morikawa, T. *J. Am. Ceram. Soc.* **2005**, *81* (2), 413–416.

(5) Kasha, M. *Discuss. Faraday Soc.* **1950**, *9* (0), 14.

(6) Beer, M.; Longuet-Higgins, H. C. *J. Chem. Phys.* **1955**, *23* (8), 1390–1391.

(7) Warner, I. M.; Christian, G. D.; Davidson, E. R.; Callis, J. B. *Anal. Chem.* **1977**, *49* (4), 564–573.

(8) Coble, P. G. *Mar. Chem.* **1996**, *51* (4), 325–346.

(9) Baker, A. *Environ. Sci. Technol.* **2001**, *35* (5), 948–953.

(10) Zhu, S.-H.; Wu, H.-L.; Li, B.-R.; Xia, A.-L.; Han, Q.-J.; Zhang, Y.; Bian, Y.-C.; Yu, R.-Q. *Anal. Chim. Acta* **2008**, *619* (2), 165–172.

(11) Smilde, A. K.; Bro, R.; Geladi, P. *Multi-way analysis with applications in the chemical sciences*; J. Wiley: Chichester, West Sussex, England ; Hoboken, NJ, 2004.

(12) Chen, H.; Kenny, J. E. *Ann. Environ. Sci.* **2007**, *1* (0).

(13) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171.

(14) Carroll, J. D.; Chang, J.-J. *Psychometrika* **1970**, *35* (3), 283–319.

(15) R. Harshman. *UCLA Work. Pap. Phonetics* **1970**, No. 16, 1–84.

(16) Allen R. Muroski; Karl S. Booksh, † and; Myrick*, M. L. *Anal. Chem.* **1996**, *68* (20), 3534–3538.

(17) Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R.; Baker, A.; Stuetz, R.; Khan, S. J. *Anal. Methods* **2013**, *5* (23), 6557.

(18) Chen, H.; Kenny, J. E. *Analyst* **2010**, *135* (7), 1704–1710.

(19) Leurgans, S.; Ross, R. T. *Stat. Sci.* **1992**, *7* (3), 289–310.

(20) Ho, C. N.; Christian, G. D.; Davidson, E. R. *Anal. Chem.* **1980**, *52* (7), 1071–1079.

(21) *Cary Eclipse Hardware Manual and Specifications*.

(22) Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*, 6th ed.; Thomson Higher Education: Belmont, CA, 2007.

(23) Ingle Jr., J. D.; Crouch, S. R. *Spectrochemical Analysis*; Prentice Hall: Englewood Cliffs, NJ, 1988.

(24) Parker, C. A. *Analyst* **1959**, *84* (1000), 446.

(25) Hall, G. J. Chemometric Characterization and Classification of Estuarine Water Through Multidimensional Fluorescence, Tufts University, 2006.

(26) MacDonald, B. C.; Lvin, S. J.; Patterson, H. *Anal. Chim. Acta* **1997**, *338* (1–2), 155–162.

(27) Chen, H. Multi-way Fluorescence Studies: 1) Humic Substances; 2) Fluorescence Resonance Energy Transfer in Micelles, Tufts University, 2008.

(28) Gu, Q.; Kenny, J. E. *Anal. Chem.* **2009**, *81* (1), 420–426.

(29) Melhuish, W. H. *J. Opt. Soc. Am.* **1962**, *52* (11), 1256.

(30) Wise, B. M.; Gallagher, N. B.; Bro, R.; Shaver, J. M.; Windig, W.; Koch, R. S. *PLS_Toolbox 4.0 for use with MATLAB*; Eigenvector Research, Inc.: Wenatchee, WA, 2006.

(31) Bro, R.; Kiers, H. A. L. *J. Chemom.* **2003**, *17* (5), 274–286.

(32) Chen, H.; Kenny, J. E. *Analyst* **2012**, *137* (1), 153–162.

(33) Murphy, K. R.; Hambly, A.; Singh, S.; Henderson, R. K.; Baker, A.; Stuetz, R.; Khan, S. J. *Environ. Sci. Technol.* **2011**, *45* (7), 2909–2916.

(34) Riu, J.; Bro, R. *Chemom. Intell. Lab. Syst.* **2003**, *65* (1), 35–49.

(35) Stedmon, C. A.; Bro, R. *Limnol. Oceanogr.* **2008**, *6*, 572–579.

(36) Lorenzo-Seva, U.; Ten Berge, J. M. F. *Hogrefe Huber Publ. Methodol.* **2006**, *2* (2), 57–64.

Chapter 2: Estimating Fluorescence Lifetimes of Mixture Components using Steady-State Fluorescence Measurements

2.1 Introduction

2.1.1 Fluorescence Lifetime

There are situations where knowing approximate fluorescence lifetime can be very valuable, e.g., identifying unknowns.[1] Obtaining fluorescence lifetime data normally requires the use of a pulsed laser or flash lamp, or a high frequency phase modulation system, which can be expensive and time consuming to operate.[2] Furthermore, even using these techniques, it is difficult to determine the fluorescence lifetimes of components of an unknown mixture. Even if the components and their spectra are known, it is often impossible to choose a single excitation wavelength (or range) and a single emission wavelength (or range) at which it can be safely assumed that only one component contributes to the observed intensity. Further separation can be implemented prior to fluorescence measurement, but this chapter proposes a method that eliminates the need for this step.[3]

When a molecule is excited by a photon, it can relax to the ground electronic state by several different pathways. The rate at which an electronically excited molecule relaxes to the ground state by radiative means (fluorescence or phosphorescence, is related to the fluorescence lifetime of the molecule. The fluorescence lifetime in the absence of a quencher molecule τ_0 is related to the to the rate of the decay pathways of the molecule in equation (2-1)[4]:

$$\tau_0 = \frac{1}{k_r + k_{nr}} \qquad (2\text{-}1)$$

where k_r is the radiative rate constant and k_{nr} is the non-radiative rate constant. Since the τ_0 is dependent on the decay pathways of the molecule, and not the way it is measured, it is an intrinsic property of the molecule. This makes it a property of a molecule that can help identify it.

2.1.2 Stern-Volmer Relationship

The presence of dissolved oxygen in solution can complicate fluorescence analysis by dynamic or static quenching of excited fluorophores.[5] Due to the molecules chosen in this work, dynamic quenching is the dominant process.[5] In the case of a single fluorophore, fluorescence intensity, dissolved oxygen concentration, and fluorescence lifetime are related by the Stern-Volmer equation[4]:

$$\frac{F_0}{F} = 1 + \tau_0 k_{O_2}[O_2] \qquad (2\text{-}2)$$

where F_0 is the fluorescence intensity of the fluorophore in the absence of oxygen, F is the fluorescence intensity in the presence of dissolved oxygen at a concentration equal to $[O_2]$, k_{O_2} is the quenching rate constant for oxygen. The product of τ_0 and k_{O_2} is also referred to as a Stern-Volmer constant K_{SV}.

Another method for measuring τ_0 is varying the concentration of O_2 in solution. Determining τ_0 can be done from measurements of fluorescence intensity in the presence and absence of oxygen used in equation (2-2). This method suffers from the same difficulties as the other methods when involving measurement in mixtures. A potential solution to these problems is using PARAFAC to model the intensity attributed to the different components of the mixture and using the information from the model in equation (2-2).

2.1.3 Using PARAFAC with the Stern-Volmer equation.

The use of PARAFAC with steady state fluorescence intensity measurements can provide a convenient, economical way to obtain approximate lifetimes for individual components of complex mixtures. The PARAFAC algorithm is used to explain three-way data that can be shown using the following equation:

$$X_{ijy} = \sum_{r=1}^{R} a_{ir} b_{jr} c_{kr} + e_{ijy} \qquad (2\text{-}3)$$

Here the dataset of EEMs is represented by X, a three-way data set with dimensions *i*, *j*, and *y*. The dimensions for the data set correspond to excitation

wavelength (*i*), emission wavelength (*j*), and sample number (*k*). For a PARAFAC model made up of *R* components, the dataset can be described by a sum of contributions for each individual component r. There are three factors that contribute to the model. The excitation and emission spectra, a_{ir} and b_{jr}, are calculated for each component r. A score value is calculated for each of the *y* samples for each component r. The term e_{ijy} represents the residual data not used in the PARAFAC model.[6]

As discussed in the previous chapter, the score value describes the fluorescence intensity of component *r* in sample *y*. Assuming an optically dilute solution, the properties of the fluorophore that contribute to the score value are shown in equation (2-4)[7].

$$c_{kr} = 2.303\phi_r \times [s]_{yr} \times \epsilon_{ir} \times em_{jr} \times f \qquad \textbf{(2-4)}$$

Where ϕ_r the fluorescence quantum yield of component r, $[s]_{yr}$ is the concentration of component r in sample y, ε_{ir} is the molar absorptivity constant at excitation wavelength *i* for component *r*, em_{jr} is the relative fluorescence intensity at emission wavelength *j* for component *r*, and *f* is a coefficient comprised of instrumental factors.[6,8]

Score values for each determined component in the mixture can be used in place of fluorescence intensities in equation (2-2). For a single sample, the concentration, excitation and emission properties, and instrumental factors for each component is the same for an oxygenated or deoxygenated sample. The change in score value between the samples of varying oxygen concentration is

due to a difference in fluorescence quantum yield caused by the quencher oxygen.[8] When taking the ratio of score values in the absence and presence of oxygen, the resulting ratio is only dependent on the quantum yields of the fluorescent component in the presence, Φ, and absence of oxygen, Φ_0 and is the same as the ratio in equation (2-2) This relationship is shown below:

$$\frac{c_{kr_0}}{c_{kr}} = \frac{\Phi_{r_0}}{\Phi_r} = \frac{F_0}{F} \qquad (2\text{-}5)$$

where c_{yr_0} is the PARAFAC score value of the *r*th component of sample *y* in the absence of oxygen and c_{yr} is the corresponding score value of the same sample in the presence of dissolved oxygen. Given this relationship, the score value ratio can be substituted into Eq. 1 and rearranged to give:

$$\tau_0 k_{O_2} = K_{SV} = \frac{1 - \frac{c_{kr_0}}{c_{kr}}}{[O_2]} \qquad (2\text{-}6)$$

The concentration of oxygen in cyclohexane of 2.35 x 10^{-3} M, a well-known value[9], is used in equation . Since all of the variables on the right side of equation (2-6) are well known or calculated by PARAFAC, the Stern-Volmer constant, K_{SV}, for each component should be well determined.

However, for most PAHs, it is usually easier to find values for τ_0 in the literature. For example, a common resource available *Handbook of Fluorescence Spectra of Aromatic Molecules*, provides values for τ_0 for many different molecules. This quantity can also be measured directly for a single fluorophore

using more appropriate instrumentation. Therefore, it may be more useful for this method to determine τ_0, instead of K_{SV}. To determine τ_0, the quenching rate constant must be dealt with. The quencher rate constant, k_{O_2}, varies depending on the fluorophore. The quencher rate constant is described by the Smoluchowski equation[4]:

$$\boldsymbol{k = 4\pi RDN10^{-3}} \qquad (2\text{-}7)$$

where N is Avogadro's number, R is the collision radius, and D is the sum of the diffusion coefficients of the fluorophore and the quencher. The variation of k_{O_2} is primarily due to the diffusion coefficient of the fluorophore and its contribution to the collision radius. These properties of the fluorophore are governed primarily by the size and shape of the molecule.[4] Therefore, molecules of a similar size, shape, and make up should have relatively similar k_{O_2} value.[10]

In this work, an average value of the oxygen quenching rate constant is used to estimate τ_0 for each component. The average value of k_{O_2} was determined from literature values compiled for many similar molecules.[10–14] The PAHs used in this experiment all have similar size and shape and their respective k_{O_2} values do not vary considerably. Figure 2-1 shows a histogram of 72 k_{O_2}values in cyclohexane for 49 different molecules.[10–14] The arithmetic mean of this set of values is $\langle k_{O_2} \rangle = 2.64 \pm .34 \text{ x}10^{10}\ \text{M}^{-1}\text{s}^{-1}$.

This value is used to estimate τ_0. Equation (2-6) can be rearranged to get,

$$\tau_0 = \frac{1 - \frac{c_{kr_0}}{c_{kr}}}{\langle k_{O_2} \rangle [O_2]}$$

(2-8)

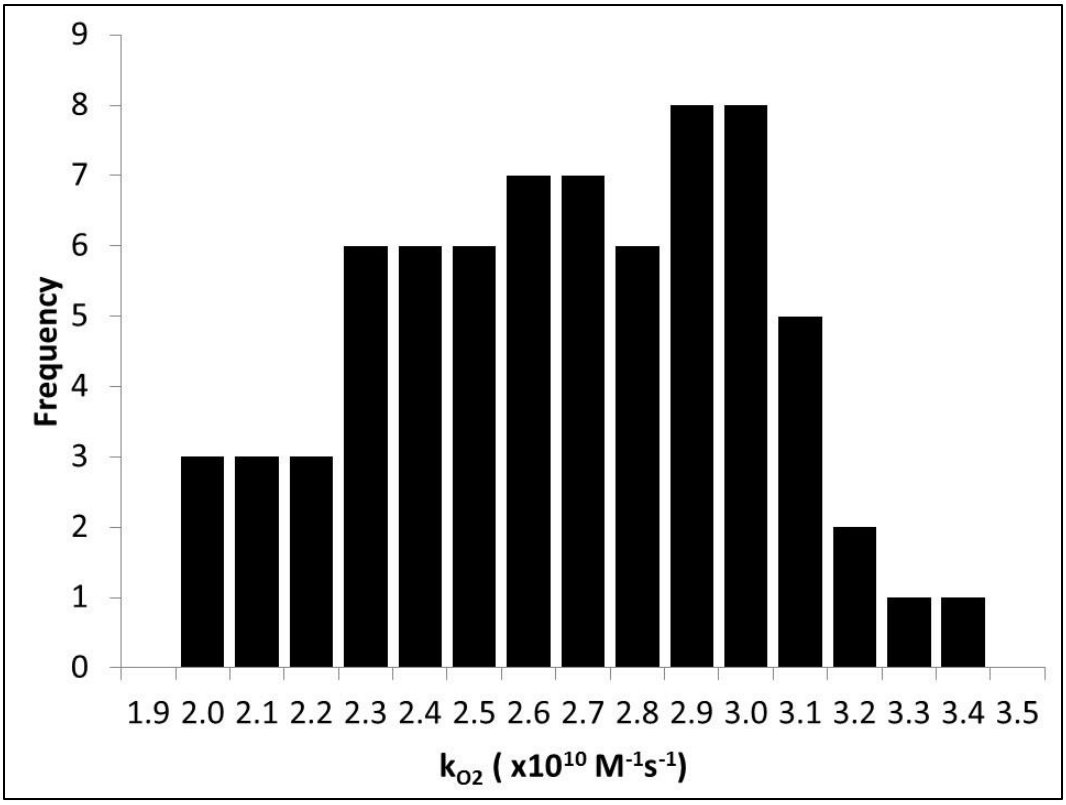


Figure 2-1: Distribution of k_{O_2} values for various PAHs. [10–14] **The arithmetic mean of this set of values is 2.64 ± 0.34 x 10^{10} $M^{-1}s^{-1}$**

One last method for calculating a fluorescence lifetime that is possible using this information from PARAFAC is by plotting the score value of each component against its concentration in the sample. For an optically dilute solution, this plot should be linear according to equation (2-4). An ideal example

of such a plot is shown in Figure 2-2. The slope of the trendline for this plot should be,

$$\text{slope} = 2.303\phi_r \times \epsilon_{ir} \times em_{jr} \times f \qquad \textbf{(2-9)}$$

If separate plots are made for the score values in the presence of quencher and in the absence of quencher, a ratio of the slopes gives a fluorescence intensity ratio as shown in equation (2-5). This fluorescence intensity can be used in equations (2-6) or (2-8) to calculate K_{SV} or τ_0, respectively. One drawback to this method, is the concentration must be known. The previous method for calculating the quantities of interest can be completed without knowing the concentration of the fluorophore.

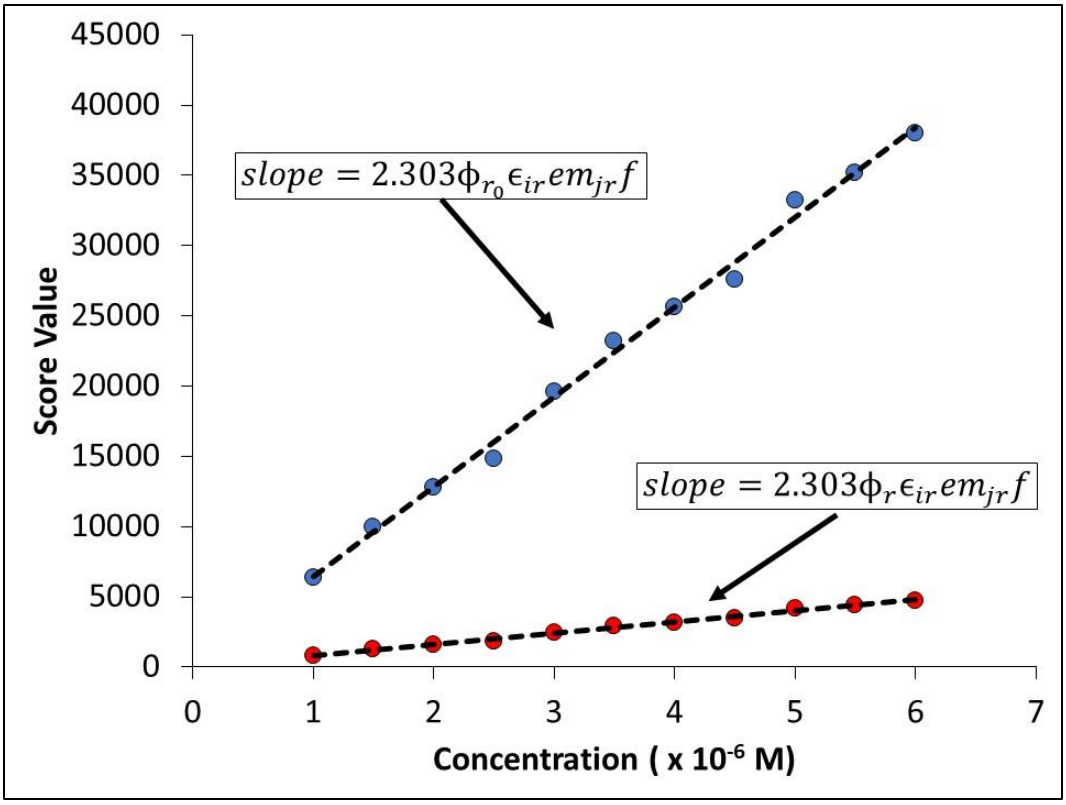


Figure 2-2: Example of plots of score value vs. concentration for a given component in the presence of oxygen (red), and in the absence of oxygen. (blue) Each plot is fit with a linear trend line and the ratio of the slopes can be used in (2-6) or (2-8) in calculating Ksv or τ_0,

In this work, we apply the PARAFAC analysis to a set of mixtures containing different combinations of 5 different PAHs, in what will be referred to as the PAH experiment. The PAHs used in this experiment are fluorene, pyrene, triphenylene, phenanthrene, and anthracene. These compounds were selected for their wide range of τ_0 values and availability within our research lab. From the data obtained in this experiment, we will determine K_{SV} and estimate τ_0 for each of the PAHs used. All methods for determining these quantities will be examined. Finally, we will apply this method to a dataset made up of gasoline samples. Gasoline is a substance that contains some of the PAHs used in the model

experiment,[15] and will provide some insight as to how applicable this method may be in identifying PAHs in mixtures where the contents are less controlled or unknown.

2.2 Experimental

2.2.1 PAH Experiment

Chemicals used were fluorene (98% Aldrich), pyrene (98% GC-grade Acros), triphenylene (98% Aldrich), anthracene (99% Acros) and phenanthrene (fluorescent grade, 98%). Each sample was prepared in cyclohexane (Acros HPLC-grade 99.99%).

Thirty samples were prepared of varying concentration regarding each compound. Samples contained 1 to 5 compounds. A table of the approximate concentrations for each molecule, in each sample, is shown below in Table 2-1. The values of concentrations are based on sample preparation. Samples were prepared using an HPLC syringe to add aliquots of a stock solution to each sample. The amounts of stock solution added to each sample were recorded, but an HPLC syringe is not as precise as a calibrated pipette. This was done for simplicity at the time. The concentration range for each compound was chosen so that the fluorescence intensity of each compound would be comparable to the other compounds under the same instrumental settings.

EEMs were collected using a Varian Cary Eclipse fluorescence spectrometer. Excitation wavelengths of 250-350 nm in 5nm increments were used. Emission was measured from 300-490 nm in 1 nm increments. Excitation

and emission band widths were 2.5 nm each. Photomultiplier tube voltage was set at 600 V. Absorbance spectra were measured using a Varian Cary 300. EEMs, and absorbance spectra were collected before and after purging with nitrogen for 15 min. This resulted in a total of 60 EEMs being used in the dataset.

EEMs were prepared for analysis by removing first-order Rayleigh scattering by setting it to "not-a-number" (NaN). The maximum absorbance of any sample in the wavelength range of interest was <0.4. EEM measurements were corrected to be optically dilute.[16] Instrumental effects were corrected for as well. [17]

	Concentration				
Sample	Fluorene ($\times 10^{-7}$ M)	Pyrene ($\times 10^{-7}$ M)	Triphenylene ($\times 10^{-7}$ M)	Anthracene ($\times 10^{-7}$ M)	Phenanthrene ($\times 10^{-8}$ M)
1	0	0	0	6	14
2	7.5	0	0	2	10
3	11	0	0	0	8
4	15	0	0	2	0
5	19	9	0	3	9
6	23	7	0	3	0
7	0	6	0	4	7
8	26	6	0	0	0
9	8	6	0	0	6
10	0	6	0	5	0
11	0	5	15	0	0
12	23	6	15	0	8
13	19	0	20	0	9
14	0	0	15	3	0
15	26	4	10	2	8
16	19	0	0	0	0
17	0	6	0	0	0
18	8	8	20	4	6
19	0	0	30	0	0
20	0	0	0	0	28
21	0	0	0	5	0
22	23	6	20	3	6
23	11	0	15	N/A	8
24	0	5	30	4	0
25	13	2	5	1	4
26	4	4	10	2	3
27	15	6	20	N/A	7
28	30	0	18	2	0
29	19	0	0	3	0
30	0	0	20	N/A	8

Table 2-1: Approximate concentration of each PAH added to each solution in the PAH experiment. Samples with a concentration of N/A indicate a contaminant.

2.2.2 Gasoline Experiment

Regular unleaded gasoline was obtained from a local gas station. A stock solution was prepared by diluting an aliquot of gasoline 1:1000 in cyclohexane (Acros HPLC-grade 99.99%). Three samples were prepared for EEM measurement by adding aliquots of the gasoline stock solution to cyclohexane in the following ratios: 1:100, 2:100 and 4:100 (stock:cyclohexane). EEMs and absorbance spectra were collected for each sample after several purge times (0 minutes, 1 minute, 5 minutes and 15 minutes). This resulted in 12 EEMs used in the dataset.

EEMs were collected for each sample using a Varian Cary Eclipse fluorescence spectrophotometer. The photomultiplier tube voltage (PMT) was set at 725 V. Excitation wavelengths of 240-350 nm in 5nm increments were used. Emission was measured from 250-500 nm in 1 nm increments. Excitation and emission band widths were 5 nm each. Absorbance spectra were measured using a Varian Cary 300. The maximum absorbance of any sample in the wavelength range of interest was <0.15. EEM measurements were corrected to be optically dilute.[16] Instrumental effects were corrected for as well. [17] EEMs were prepared for analysis by removing first-order Rayleigh scattering and first-order Raman scattering by setting it to "not-a-number" (NaN).[6]

2.3 Fitting PARAFAC models for PAH Experiment

2.3.1 Diagnostic Table

PARAFAC models were fit as described in Chapter 1. A table of the diagnostics is shown below in Table 2-2.

Number of Components	Number of Iterations	Percentage of Data Explained	Sum of Squared Residuals	Core Consistency Value
1	13	71.174	1.62E+08	100
2	16	89.974	5.52E+08	100
3	45	96.882	1.72E+08	99
4	55	99.530	2.59E+07	100
5	59	99.817	1.01E+07	100
6	145	99.864	7.50E+06	< 0
7	323	99.88	6.10E+06	< 0

Table 2-2: Diagnostics used for deciding most appropriate model in PAH experiment. The best fit model, the 5-component model, is shaded.

Examining the number of iterations, there is a small increase in the number of iterations as we proceed through models 1-5. For the 6-component model, the number of iterations needed for the model to converge triple. Therefore, this diagnostic suggests a 5-component model is the best fit.

The percentage of variance explained increases steadily as the number of components used in the model increases to 5 components. It does not increase very much for 6 and 7-component models, suggesting a 5-component model. The trend in the sum of squared residuals shares the same conclusion. The sum of squared residuals does not decrease considerably for models with more than 5 components. These diagnostics suggest a 5-compoent model is the best fit as well.

The core consistency diagnostic suggests a 5 component model is the best fit. For models with 1-5 components the core consistency value is high at around 100%. For 6 and 7-component models, the core consistency drops below 0. Since this is a model experiment, and not a very complex sample, core consistency is

expected to be useful in this experiment.[18] All the considerations in the diagnostic table result in a 5 component PARAFAC model being most appropriate.

2.3.2 Evaluation of PARAFAC Spectra

The excitation and emission spectra fit by the 5-component model are shown in Figure 2-3. All spectra match reference spectra for the five PAHs used very well.[10] Models with more than 5 components produce components with spectra that resemble fluorene. This is a sign of overfitting.[6] These observations, along with the other diagnostic data, supports the five component model to be most appropriate. Considering we added 5 PAHs to the samples in this experiment, this is a good result.

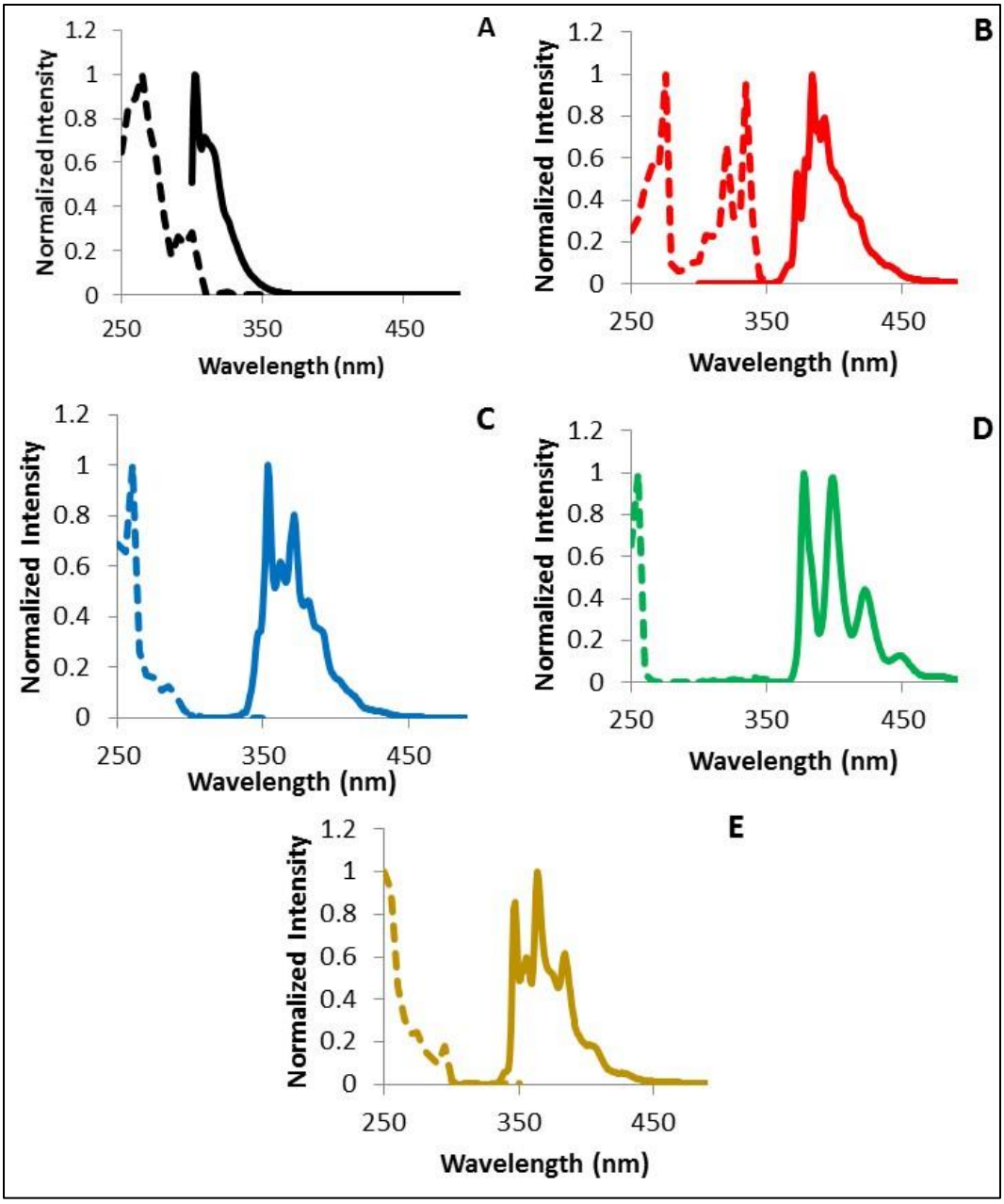


Figure 2-3: PARAFAC calculated excitation (dashed) and emission (solid) spectra for the PAH experiment (A) Component 1 Fluorene (B) Component 2 Pyrene (C) Component 3 Triphenylene (D) Component 4 Anthracene (E) Component 5 Phenanthrene

2.3.3 Variance Per Component

The percentage of variance explained by each component of the model is shown in Figure 2-4. Fluorene, the first component, explains the most variance in the model followed by pyrene, triphenylene, anthracene, and phenanthrene. The phenanthrene component (component 5) explains only 2 percent of the variance, and it may be a reason for the slight deviation in split -half spectra shown in Figure 2-7.

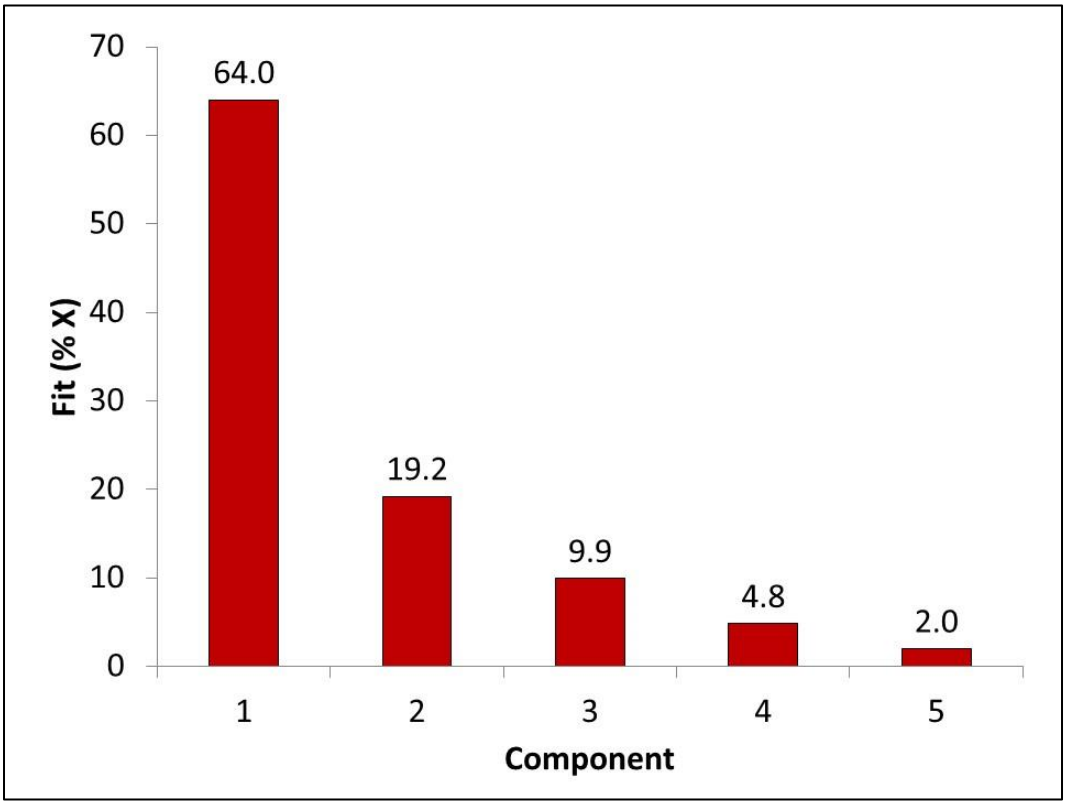


Figure 2-4: Variance explained by each component as a percentage of the model for the PAH experiment.

2.3.4 Jack-knife and Split Half Analysis

The five component model was evaluated for outliers using jack-knife PARFAC. [19] Jack-knife suggested no outliers for the dataset. The relevant plots can be found below in Figure 2-6 and Figure 2-5. Examination of RIP plots from

jack-knife suggested a low deviation in spectral loadings for each segment. The data points are well clustered and the outer data points do not deviate enough to suggest any are outliers.[19] Examination of IMP plots from jack-knife suggested no outliers regarding score values for each sample. All data points are on or very close to the identity line.

Split-half analysis was used to examine the consistency of the data set by modeling halves of the dataset separately.[20] Split half spectra are shown in Figure 2-7 and show consistent modeling for both halves. The halves were validated by Tucker congruency coefficients as well.[21] There are minimal differences between each half's spectra for a given component. The largest amount of difference between halves can be observed in component 5's spectra, used to describe phenanthrene. The differences between component 5's spectra were considered minimal.

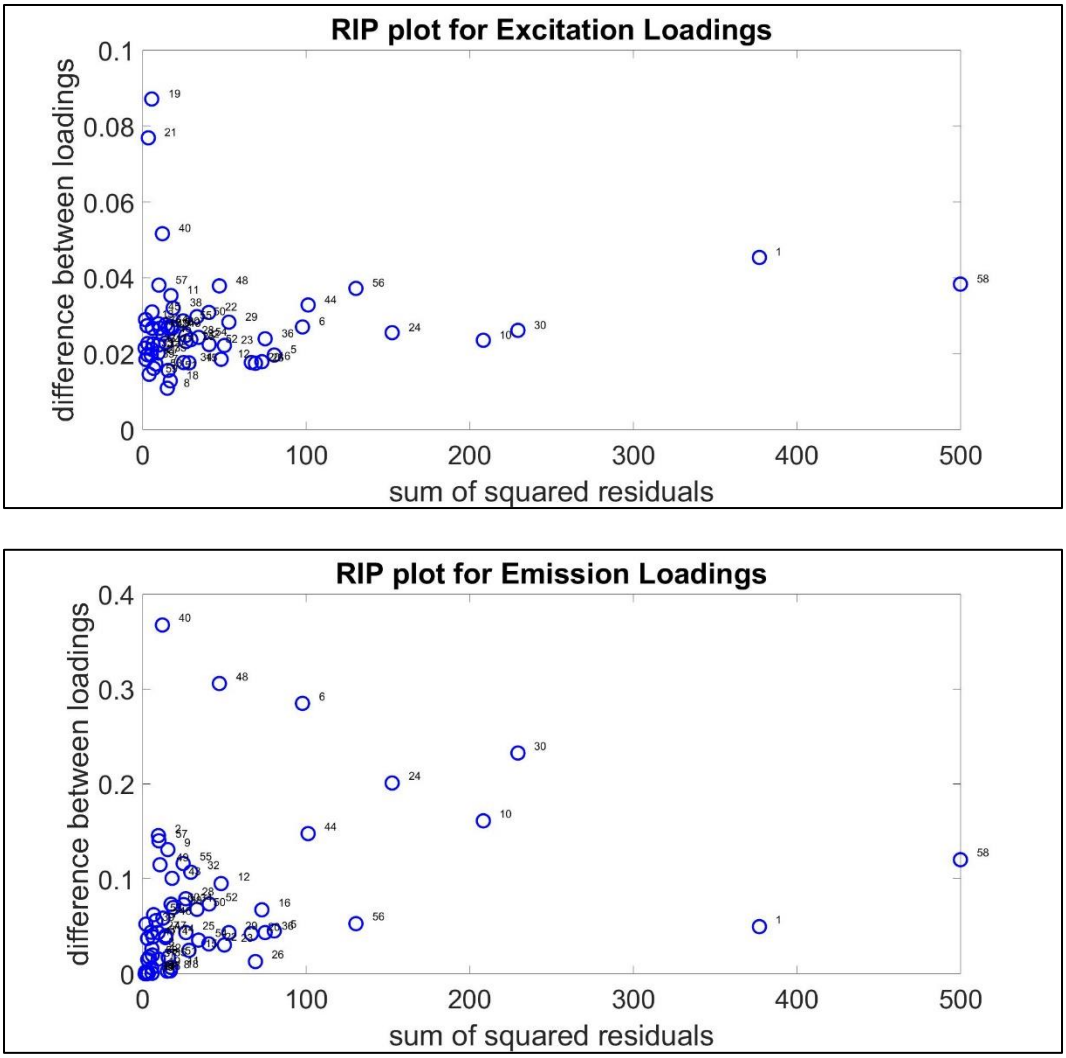


Figure 2-5: RIP plots for 5-component PARAFAC Model. No Data points deviate enough to be considered an outlier.

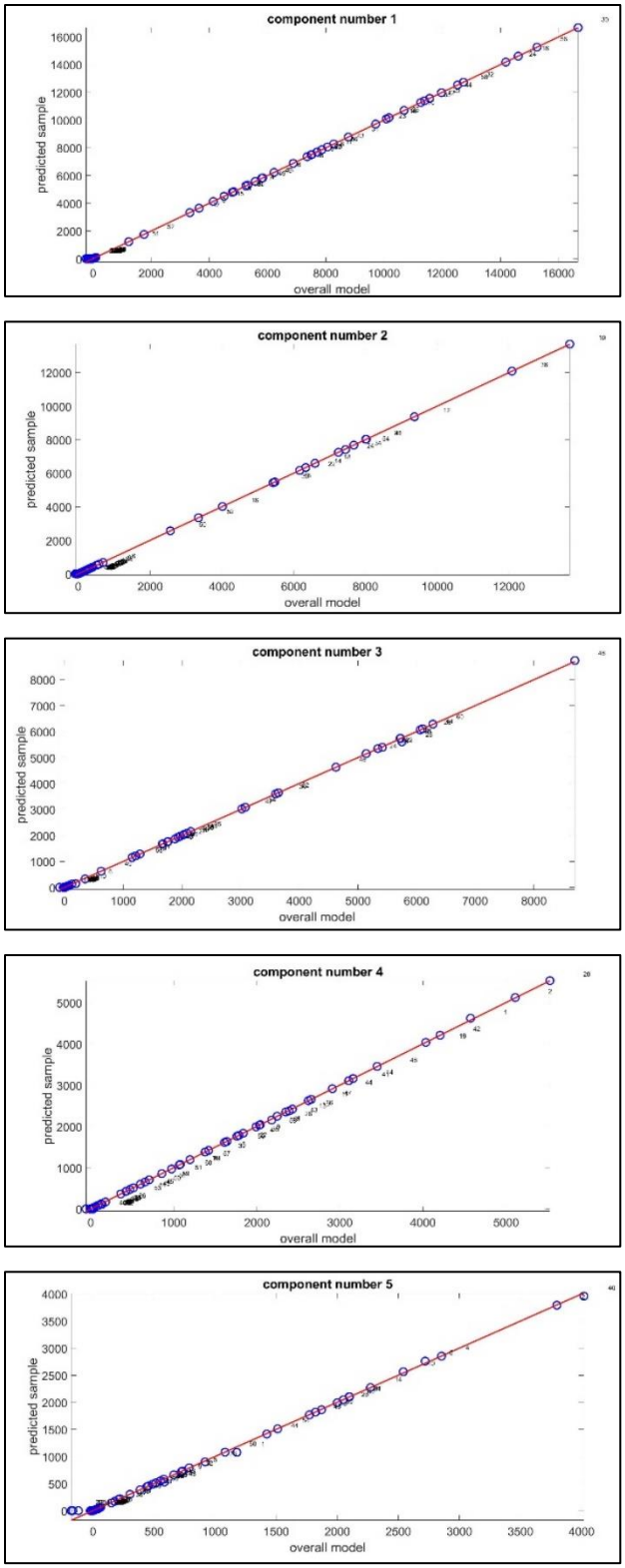


Figure 2-6: IMP Plots for each component of the 5 component model. No outliers were observed as each data point falls on the line.

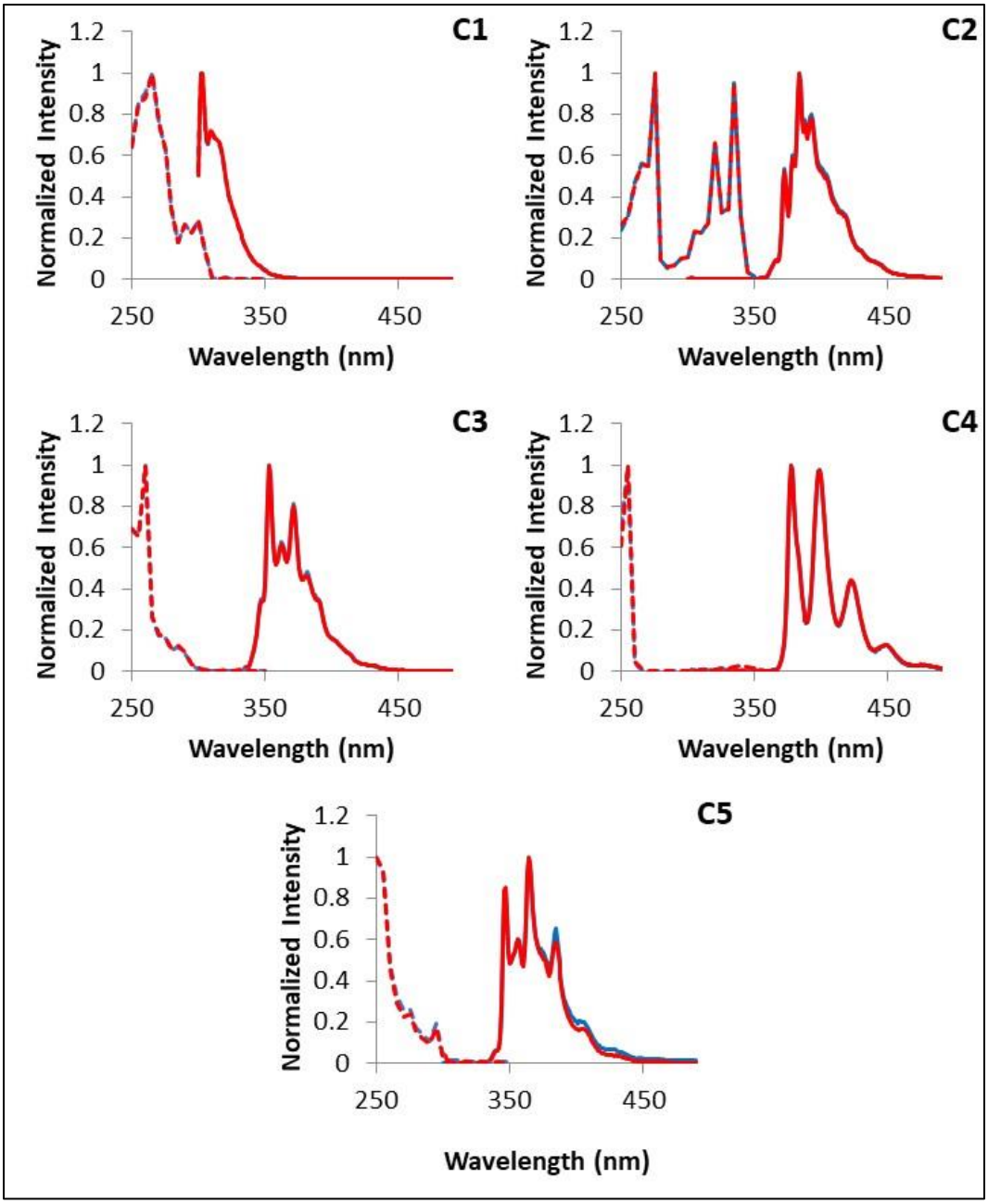


Figure 2-7: Split Half Analysis for 5 component model: Half 1 (blue) Half 2 (red) excitation (dashed) emission (solid) Each component is labeled C1-C5 accordingly. <u>Note:</u> Most spectra from each half overlay nearly perfect and it is difficult to discern spectra of halves.

2.3.5 Evaluation of Score Values

The score values for each component are shown in Figure 2-8 and Table 2-3. While for the most part, score values behave as expected, there was a problem. An issue that arose in this work was that some samples have non-zero score values for a component when the corresponding compound was not added. Unaltered EEMs were first examined to ensure proper composition. While some samples had an unintended anthracene addition, as indicated in Table 2-1, the rest of the samples cannot be explained this way. In the case of the anthracene contaminant, the raw spectra show visible traces of anthracene and the PARAFAC score values resulting from these samples can be used to calculate adequate K_{SV} values (shown later). Changing PARAFAC options, such as stopping criteria, to better fit the model and eliminate this error was not successful. The only explanation for this error is small sample to sample variations in spectra that are modeled improperly. Some of the PAHs spectra overlap considerably and one explanation is the fluorescence intensity is fit incorrectly. We have not been able to find any pattern in this error, and it appears to be random. Other contaminations are not believed to be the cause of this effect as a similar experiment been able to identify impurities as separate components (See Chapter 4). In almost all cases, the score values of these samples are much lower than the score value where the component is present. In the next section, a method for dealing with these non-zero score values is proposed.

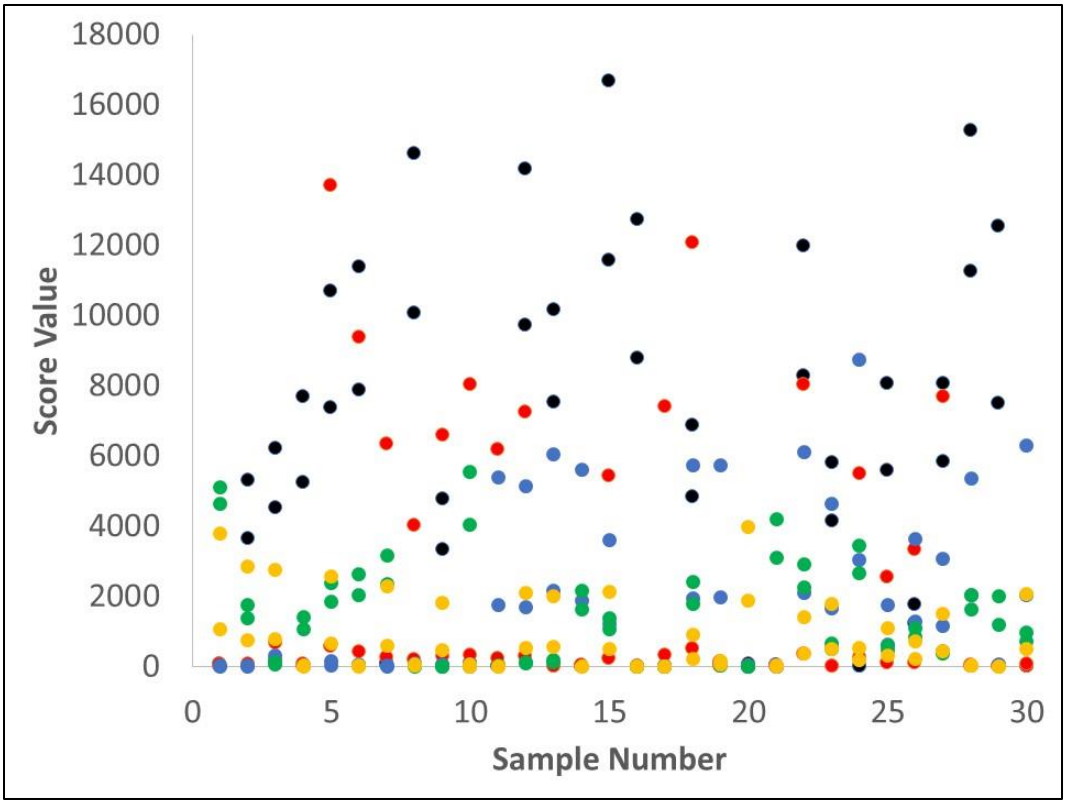


Figure 2-8: Plot of score values for each component in the 5-component model. Each sample has two score values for each component (oxygenated and deoxygenated). Component 1 (black) Component 2 (Red) Component 3 (Blue), Component 4 (Green), Component 5 (Yellow)

Sample		Component 1	2	3	4	5
1	U	55	0	0	4618	1077
	P	85	67	22	5117	3796
2	U	3645	15	8	1377	734
	P	5318	66	37	1763	2857
3	U	4507	190	120	63	791
	P	6226	679	319	112	2764
4	U	5251	14	17	1067	7
	P	7701	85	36	1419	24
5	U	7360	582	28	1835	663
	P	10697	13700	141	2373	2567
6	U	7865	406	21	2042	7
	P	11394	9364	56	2621	33
7	U	57	270	1	2347	581
	P	91	6336	21	3159	2274
8	U	10080	197	28	10	13
	P	14613	4012	76	37	59
9	U	3326	286	16	2	453
	P	4787	6592	44	0	1822
10	U	56	328	0	4035	12
	P	105	8022	61	5529	45
11	U	0	241	1760	13	2
	P	0	6168	5395	35	0
12	U	9715	282	1678	86	514
	P	14184	7249	5140	118	2107
13	U	7524	3	2149	121	555
	P	10172	0	6057	170	2000
14	U	0	0	1866	1634	0
	P	0	41	5606	2155	0
15	U	11570	221	1210	1074	513
	P	16670	5430	3592	1379	2112
16	U	8775	0	0	3	0
	P	12726	0	0	0	0
17	U	0	311	0	0	0
	P	0	7410	0	0	0
18	U	4835	515	1931	1784	209
	P	6881	12083	5732	2419	906
19	U	60	25	1966	26	57
	P	92	123	5734	83	147
20	U	43	0	20	0	1866
	P	62	0	0	0	3959
21	U	34	13	2	3106	11
	P	51	16	5	4206	21
22	U	8274	370	2084	2244	386
	P	11977	8023	6104	2912	1421
23	U	4134	5	1663	513	488
	P	5799	5	4623	652	1776
24	U	0	274	3021	2654	175
	P	0	5480	8733	3451	536
25	U	5574	99	622	470	303
	P	8065	2565	1752	600	1082
26	U	1230	113	1282	855	222
	P	1756	3346	3644	1078	724
27	U	5833	376	1150	361	443
	P	8059	7675	3080	427	1514
28	U	11265	22	2021	1611	18
	P	15259	35	5339	2028	32
29	U	7496	7	17	1194	0
	P	12535	0	48	1988	2
30	U	0	18	2030	703	488
	P	0	60	6280	967	2051

Table 2-3: Score values for PAH experiment. Row labeled U is for unpurged sample measurements, row labeled P is for Purged sample measurements.

2.4 Evaluation of Non-zero Score Values in Model Experiment

Score values were evaluated for further analysis using a two-step process. The first step of the process involved evaluating score value pairs, oxygenated and deoxygenated for each component of each sample, by comparing them to a minimum cut-off criterion. In this step, score pairs below the cut off are disregarded and not used for further analysis.

Where to set the cut off was examined by considering two types of errors: *Type 1* errors involve both score values of a sample being over the cut off criteria for a component corresponding to a compound that should not be in a sample. This would be considered a *false positive*. Type 2 errors involve a sample's score values not being selected, being below the cutoff, when the respective compound was added to the sample in Table 2-1. This would be a *false negative*.

A cut off value was set to 1.15% of the highest score value in the dataset. This cut-off was selected by changing the value to minimize the number of Type 1 or Type 2 errors. This criterion resulted in 98.7% correct classification of score values. There were two Type 2 errors (false negatives) for component 2, pyrene, due to the oxygenated score values being below the cutoff. These samples also had the lowest concentration of pyrene added, as indicated in Table 2-1. The samples with an unintended anthracene addition were included in the analysis, as they met the cut off criteria.

The score values that passed the cut off criteria were then used to determine Stern-Volmer constants, according to equation (2-6). Once K_{SV} values were determined for these samples, Grubbs' test[22] was used to identify

outlier constants. These outliers were disregarded from further consideration and not used in calculating the average K_{SV} value for each component.

2.5 Estimation of Fluorescence Lifetime in PAH Experiment

2.5.1 Stern-Volmer Constant Determination

Average Stern-Volmer constants determined from this work are summarized in Table 2-4. The values of K_{SV} obtained in this work are in good agreement, within one standard deviation, with calculated literature values. A K_{SV} value was calculated for each score pair approved for analysis by the criteria stated in the previous section.

Compound	Number of Score Pairs Included	Experimentally Determined K_{SV} (M^{-1})	Literature Value of k_{O_2} (x 10^{10} $M^{-1}s^{-1}$)	Literature Value of τ_0 (ns)	Calculated K_{SV} (M^{-1})
Fluorene	18	182 ± 16	2.04 [a]	10 [b]	204
Pyrene	14	9397 ± 782	2.19 [c]	405 [d]	8870
Triphenylene	15	808 ± 58	2.10 [e]	36.6 [e]	769
Anthracene	18	129 ± 20	2.60 [e]	4.9 [e]	127
Phenanthrene	17	1187 ± 132	2.32 [f]	57.5 [f]	1334

Table 2-4: In the third column, experimentally determined Stern-Volmer constants, K_{SV}, using an [O2] of 2.35 x 10^{-3} M from this work are listed. In the right-most column are calculated values for K_{SV} using various literature sources. a: Calculated from Ref 10 using $F_0/F = 1.48$ $[O_2] = 2.35 \times 10^{-3}$ M and $\tau_0 = 10$ ns , b: Ref 10 c: Calculated from Ref 11 using F0/F = 21.81 $[O_2] = 2.35 \times 10^{-3}$ M and $\tau_0 = 405$ ns d: Ref 11, e: Ref 12, f: Ref 13

2.5.2 Estimation of τ_0 from an Average Quenching Rate Constant

Estimated τ_0 values for each of the five components were calculated for each sample used in the K_{SV} determination for each component. For each compound, in Table 2-5, an average τ_0 and standard deviation is reported, alongside reference values from the literature. As can be seen, the estimated τ_0

values match well to literature values for each component. Each of the average τ_0 estimates fall within a few standard deviations of the corresponding literature ranges.

Compound	**Estimated Lifetime with St. Dev. (ns)**	**Lit. Lifetimes (ns)**
Fluorene	7 ± 1	10[a]
Pyrene	356 ± 30	370[b] 397[c] 400[d] 405[e], 450[f]
Triphenylene	31 ± 2	36.6[a]
Anthracene	5 ± 2	4.9[a,g]
Phenanthrene	45 ± 5	53[c] 56[h] 57.5[a]

Table 2-5: Estimated fluorescence lifetimes for each component with standard deviation and comparison to literature values. Estimated fluorescence lifetimes using average k_{O_2} value=2.64 x 10^{10} $M^{-1}s^{-1}$and $[O_2] = 2.35 \times 10^{-3}$M. a: Ref 10, b:Ref 23, c: Ref 5, d: Ref 24, e: Ref 11, f: Ref 25, g: Ref 14, h: Ref 26

2.5.3 Estimating τ_0 from Specific Quenching Rate Constants.

Using an average value for k_{O_2} is advantageous if the makeup of a mixture is completely unknown. Estimated lifetimes, as well as spectral information calculated by PARAFAC should be useful for further analyte identification. If any information about the mixture is determined, or other types of molecules are being measured, different k_{O_2} values can be used to improve estimation. In this work, the identity of each component is known. Specific k_{O_2} values can be substituted for each individual component, and improve the estimation of τ_0. These estimates are shown in Table 2-6. The use of specific quenching rate

constants greatly improves the estimation of fluorescence lifetime and each component matches literature values within one standard deviation.

Compound	**Determined τ_0 (ns)**	**Literature τ_0 (ns)**
Fluorene	9 ± 1	10[a]
Pyrene	429 ± 36	370[b] 397[c] 400[d] 405[e], 450[f]
Triphenylene	38 ± 3	36.6[a]
Anthracene	5 ± 1	4.9[a,g]
Phenanthrene	51 ± 6	53[c] 56[h] 57.5[a]

Table 2-6: Determined fluorescence lifetimes for each component with standard deviation and comparison to literature values. Estimated fluorescence lifetimes using k_{O_2} values referenced in Table 2 and $[O_2] = 2.35 \times 10^{-3}$M. a: Ref 10, b:Ref 23, c: Ref 5, d: Ref 24, e: Ref 11, f: Ref 25, g: Ref 14 ,h: Ref 26

2.5.4 Estimation of τ_0 using Score Value Plots

In this work, the approximate concentration of each PAH in each sample is known. As stated earlier, an alternate method, using the slopes of score vs. concentration plots, can be used to calculate K_{SV} and τ_0. Score values for each component were plotted as shown in Figure 2-2. The summary of the results using this method can be found in Table 2-7.

		Fluorene	Pyrene	Triphenylene	Anthracene	Phenanthrene
Oxygenated Score Plot	*Slope (M^{-1})*	4.00E+09	5.60E+08	9.64E+08	6.91E+09	7.00E+09
	Std. Dev (M^{-1})	4.40E+08	1.00E+08	2.10E+08	1.40E+09	1.30E+09
	R^2	0.93	0.68	0.48	0.81	0.83
Deoxygenated Score Plot	*Slope (M^{-1})*	5.69E+09	1.29E+10	2.80E+09	9.20E+09	2.63E+10
	Std. Dev (M^{-1})	6.80E+08	2.50E+09	6.50E+08	1.90E+09	4.10E+09
	R^2	0.91	0.65	0.45	0.81	0.86
Slope Ratio		1.43	23.0	2.92	1.33	3.75
Std. Dev.		0.23	6.1	0.93	0.38	0.91
K_{SV}		180	9400	820	140	1170
Std. Dev.		30	2500	260	41	290
τ_0 *(ns)*		6.8	350	31	5.3	44
Std. Dev. (ns)		1.4	110	11	1.7	12

Table 2-7: Results for Score v. Concentration plots of both oxygenated samples and deoxygenated samples. Lifetimes were calculated using $\langle k_{O_2} \rangle = 2.64 \times 10^{10}\ M^{-1}s^{-1}$ and $[O_2] = 2.35 \times 10^{-3}$ M.[9]

While the R^2 values do not imply a great linear relationship for some plots, most notably pyrene and triphenylene, the determined K_{SV} values compare nicely to those summarized in Table 2-4. Better R^2 values may have been obtained, had the concentration of each compound been known with a greater certainty. Considering the concentrations are approximate, the relationships follow an appropriate trend.

As stated before, this method requires concentration information for each sample, while the previous method is independent of concentration. Therefore, the first method is preferred unless reliable concentration information is known.

2.6 Estimation of τ_0 Gasoline Samples

2.6.1 Fitting of PARAFAC Model

The diagnostic table for this experiment is shown in Table 2-8. Evaluating the information in the table suggests either a 3 or 4-component PARAFAC model. The number of iterations steadily increase from 1 to 4 components with a large increase in the number of iterations when a fifth component is added to the fitting of the model. Both the sum of squared residuals and percentage of variance explained significantly change going from a 1-component model to a 4 component model. The core consistency diagnostic is only high for a 1 or 2-component model. A 3-component model still has a significant positive value for the core consistency. The core consistency drops below zero for all other models. Gasoline has many molecules[15] and can be considered a complex sample. Core consistency may not be useful in this experiment.[27] Many compounds in gasoline have very similar spectra and may be a reason for the lower core consistency value in the PARAFAC models.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	79	3.99E+06	97.964	100
2	362	5.85E+05	99.695	95
3	813	3.02E+05	99.842	45
4	227	1.78E+05	99.905	< 0
5	1327	1.23E+05	99.935	< 0
6	1225	9.50E+04	99.950	< 0

Table 2-8: Diagnostic table for PARAFAC models fitted in the gasoline experiment. The best fit model, the 3-component model, is shaded.

2.6.2 Evaluation of PARAFAC spectra

Looking at the PARAFAC spectra in Figure 2-9, all the spectra exhibit fluorescence consistent with PAHs expected in gasoline.[15] Components 1 and 2 correspond to a benzene-like and naphthalene-like molecule, respectively.[10] Component 3 was unidentifiable. It appears to be a mixture of more than one fluorophore, based on the complexity of the spectrum. In PARAFAC models fit to 4, 5 and 6 components, degenerate components are fitted and are not useful.[6] Possibilities for the identity of component 3 are an anthracene-like molecule or pyrene-like molecule. Both molecules exhibit fluorescence in this wavelength region and have a relatively high concentration in gasoline.[15] Models fit to 5 or 6 components, while not justifiable, have spectra that resemble these molecules.[10]

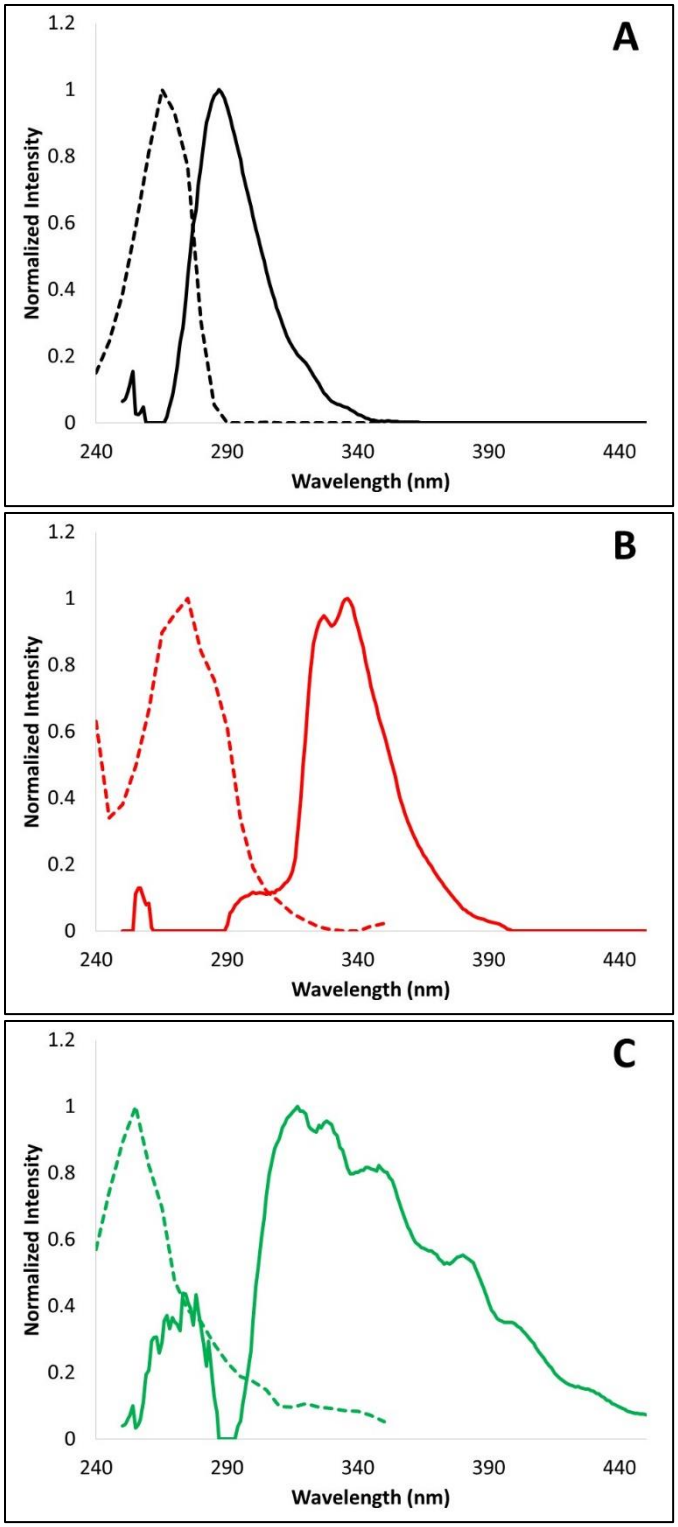


Figure 2-9: PARAFAC spectra for the 3-compoent model. A: Component 1 (Benzene-like) B: Component 2 (Naphthalene-like), C: Component 3 Excitation (dashed), Emission (solid)

2.6.3 Examination of Score Values

The score values for the 3-component model are shown in Figure 2-10. The score values trend as expected. All three components trend towards higher score values with each sample. This follows the concentration trend in the sample preparation. These score values will be used in estimating τ_0 for each component.

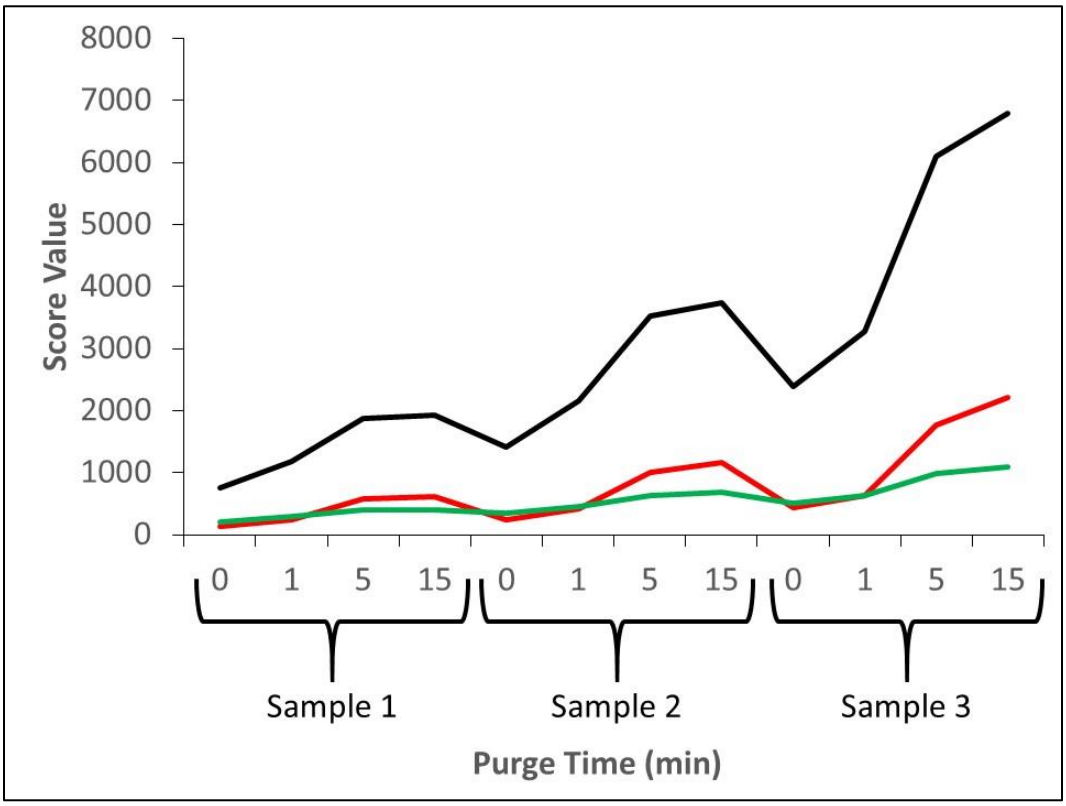


Figure 2-10: Plot of score values for the 3-component PARAFAC model. Component 1 (—); Component 2 (—); Component 3 (—); Component 4 (—)

2.6.4 Jack-knife and Split Half Analysis.

Jack knife analysis of the 3-component model did not reveal any outliers. While sample 12 (sample 3 after 15 minutes of purging with N_2) deviated the most of any sample with respect to the RIP plots shown in Figure 2-11, it is not considered an outlier. IMP plots, shown in Figure 2-12, do not suggest any

outliers as well. All data points lie on the identity line. The 3-component model was validated by split half analysis. All TCC values met the criteria for a valid model. Spectra for the analysis are shown in Figure 2-13. While the spectra for component 1 and 2 are very consistent in both halves, there are slight differences in the spectra of both halves for component 3. There are small differences in the structure of the spectra from 350 – 450 nm.

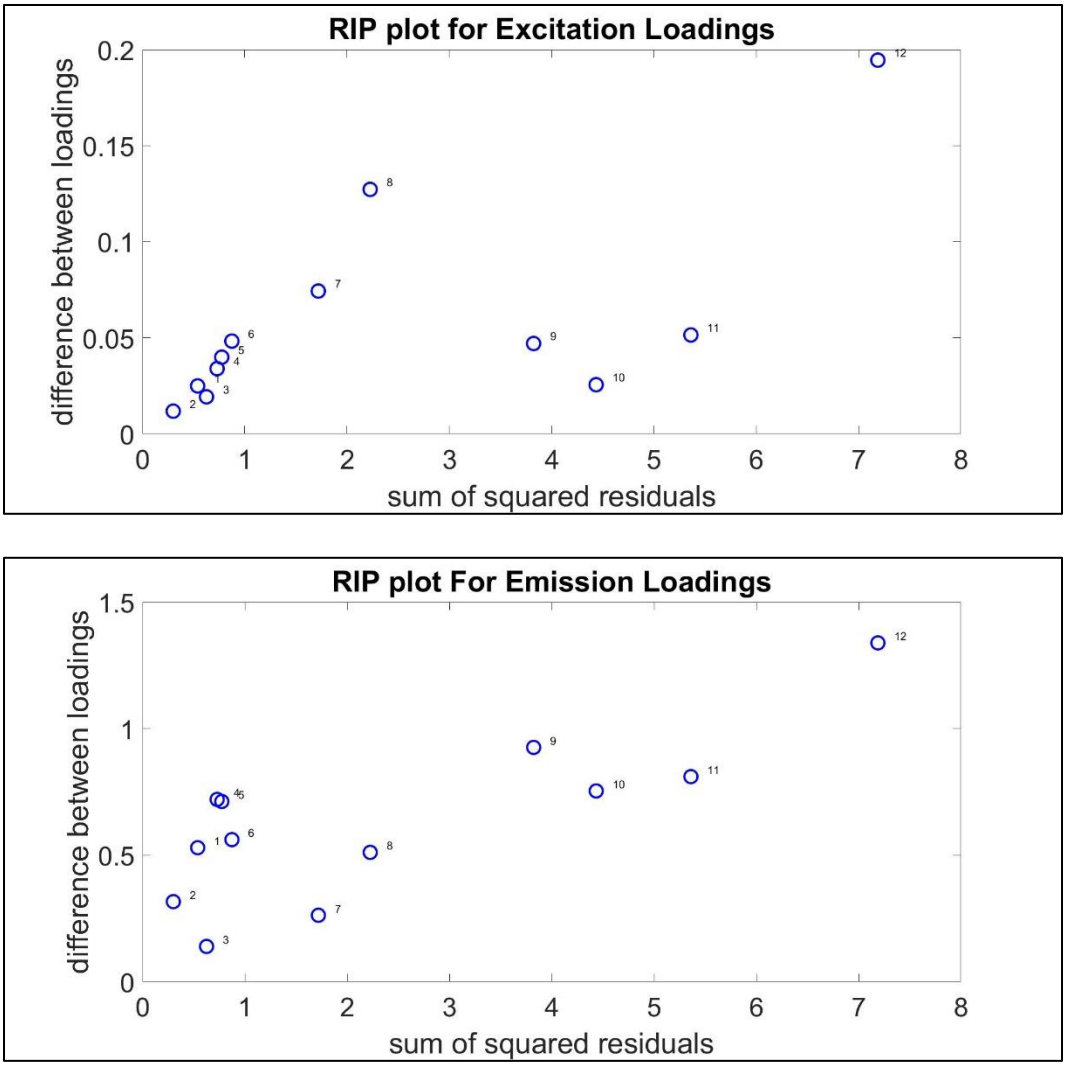


Figure 2-11: RIP plots for the 3-component PARAFAC model in the gasoline experiment.

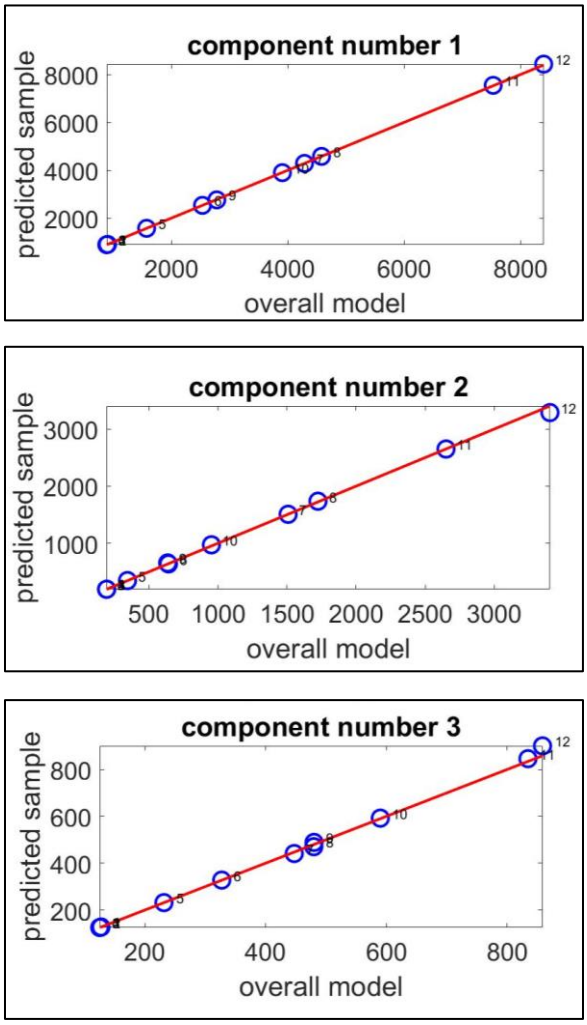


Figure 2-12: IMP plots for the 3-component PARAFAC model in the gasoline model.

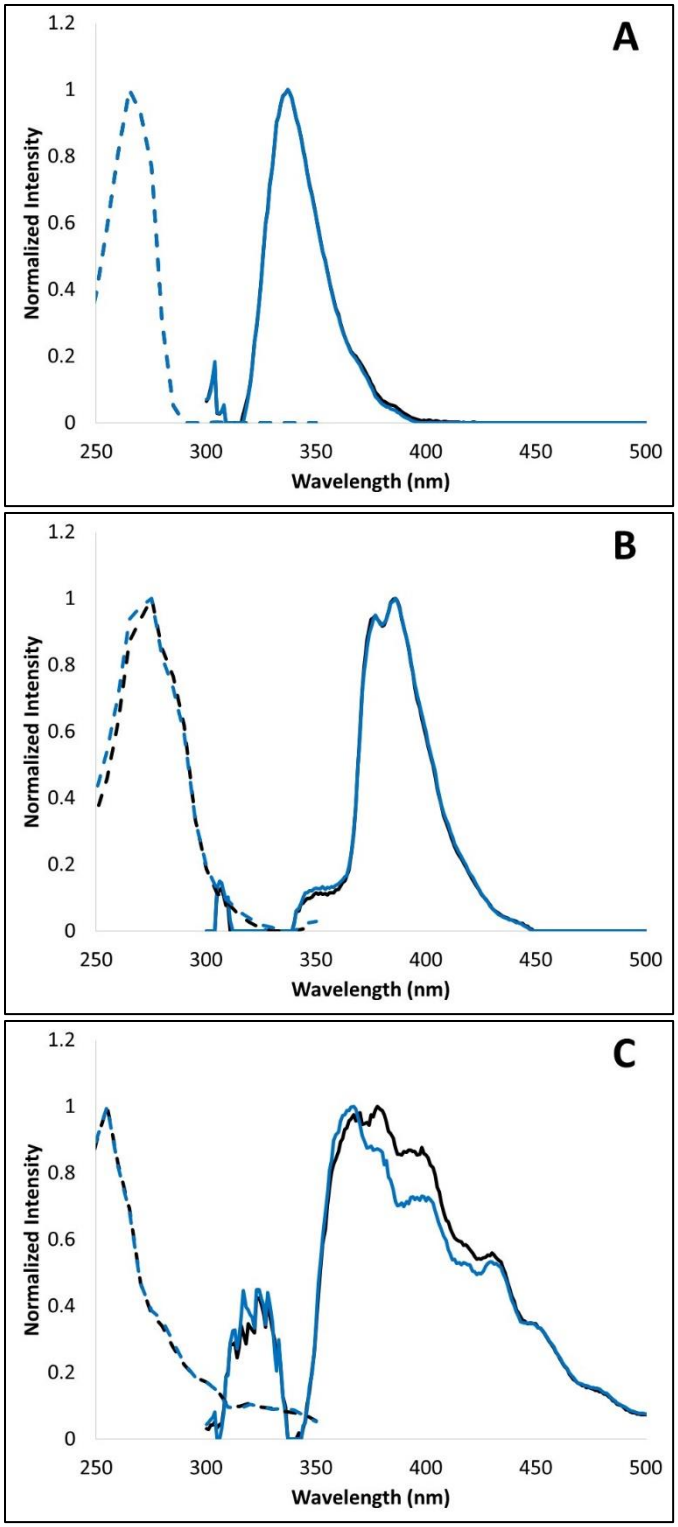


Figure 2-13: Split half spectra for the 3-component PARAFAC model in the gasoline. A: Benzene-like component; B: naphthalene-like component; C: Component 3 (Half A: black; Half B: blue; Excitation: dashed; Emission: solid)

2.6.5 Estimation of Fluorescence Lifetime

Score values were used in expression Figure 2-9 to estimate τ_0 for each component. The average quenching rate constant $\langle k_{O_2} \rangle = 2.64 \pm .34 \text{ x}10^{10}\ \text{M}^{-1}\text{s}^{-1}$ was used to estimate τ_0. The results of these estimates are shown in Table 2-9.

	Sample	Component		
		1	***2***	***3***
F_0/F	S1	2.56	4.53	1.94
	S2	2.65	4.91	2.01
	S3	2.84	5.15	2.13
τ_0 (ns)	S1	25.1	56.9	15.2
	S2	26.6	63.0	16.3
	S3	29.6	66.9	18.2
***Average τ_0* (ns)**		**27.1**	**62.3**	**16.6**
***Std. Dev.* (± ns)**		**2.3**	**5.0**	**1.5**

Table 2-9: Estimation of τ_0 for each component in the 4-component PARAFAC model.

These values estimated for τ_0 were compared to literature values. The τ_0 of benzene is 29 ns and many other substituted benzenes have a fluorescence lifetime in the absence of quencher of about 30 ns.[10] This corresponds very well to component 1. Component 2 will be described as a naphthalene-like molecule. Substituted naphthalenes have τ_0 values of about 60-70 ns.[10] This matches well with this component. Unsubstituted naphthalene has a τ_0 of 96 ns.[10] Component 3 is still not identified. The lifetime matches some methylated anthracene molecules, but the spectra are not a very good match.[10] Considering the complicated make up of gasoline,[15] this method work relatively well in identifying some of the fluorescent components.

2.7 Conclusions

The proposed methods for $\boldsymbol{K_{SV}}$ and τ_0 determination are straightforward and reliable. This method would be very accessible due to the instrumentation and methods used. It does not require the use of physical or chemical means to extract this information from a mixture. Estimated $\boldsymbol{\tau_0}$ values, combined with the spectral output PARAFAC analysis creates an interesting tool for identification of unknowns in a mixture or investigating mixture where little spectroscopic information is known.

2.8 References

(1) Vo-Dinh, T. *Talanta* **1998**, *47* (4), 943–969.

(2) Resch-Genger, U.; Ameloot, M. *Springer series on fluorescence,*. Springer: Berlin ; London 2008, p 1 online resource (xvi, 496).

(3) Desilets, D. J.; Kissinger, P. T.; Lytle, F. E. *Anal. Chem.* **1987**, *59* (14), 1830–1834.

(4) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, 3rd ed.; Springer US: Boston, MA, 2006.

(5) Brownrigg, J. T.; Kenny, J. E. *J. Phys. Chem. A* **2009**, *113* (6), 1049–1059.

(6) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171.

(7) Chen, H.; Kenny, J. E. *Analyst* **2010**, *135* (7), 1704–1710.

(8) Leurgans, S.; Ross, R. T. *Stat. Sci.* **1992**, *7* (3), 289–310.

(9) Pagano, T.; Carcamo, N.; Kenny, J. E. *J. Phys. Chem. A* **2014**, *118* (49), 11512–11520.

(10) Berlman, I. B. *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd ed.; Academic Press: New York, 1971.

(11) Pagano, T.; Biacchi, A. J.; Kenny, J. E. *Appl. Spectrosc.* **2008**, *62* (3), 333–336.

(12) Parmenter.CS; Rau, J. D. *J. Chem. Phys.* **1969**, *51* (5), 2242–2246.

(13) Saltiel, J.; Atwater, B. W. In *Advances in Photochemistry*; Volman, D. H., Hammond, G. S., Gollnick, K., Eds.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 1988; Vol. 14.

(14) Patterson, L. K.; Porter, G.; Topp, M. R. *Chem. Phys. Lett.* **1970**, *7* (6), 612–614.

(15) Marr, L. C.; Kirchstetter, T. W.; Harley, R. A.; Miguel, A. H.; Hering, S. V; Hammond, S. K. *Environ. Sci. Technol.* **1999**, *33* (18), 3091–3099.

(16) MacDonald, B. C.; Lvin, S. J.; Patterson, H. *Anal. Chim. Acta* **1997**, *338* (1–2), 155–162.

(17) Hall, G. J. Chemometric Characterization and Classification of Estuarine Water Through Multidimensional Fluorescence, Tufts University, 2006.

(18) Bro, R.; Kiers, H. A. L. *J. Chemom.* **2003**, *17* (5), 274–286.

(19) Riu, J.; Bro, R. *Chemom. Intell. Lab. Syst.* **2003**, *65* (1), 35–49.

(20) Stedmon, C. A.; Bro, R. *Limnol. Oceanogr.* **2008**, *6*, 572–579.

(21) Lorenzo-Seva, U.; ten Berge, J. M. F. *Methodology* **2006**, *2* (2), 57–64.

(22) Grubbs, F. E. *Technometrics* **1969**, *11* (1), 1.

(23) Hautala, R. R.; Schore, N. E.; Turro, N. J. *J. Am. Chem. Soc.* **1973**, *95* (17), 5508–5514.

(24) Kikuchi, K. *Chem. Phys. Lett.* **1991**, *183* (1–2), 103–106.

(25) Birks, J. B.; Dyson, D. J. *Proc. R. Soc. London Ser. a-Mathematical Phys. Sci.* **1963**, *275* (1360), 135–.

(26) Amata, C. D.; Burton, M.; Helman, W. P.; Ludwig, P. K.; Rodemeye.Sa. *J. Chem. Phys.* **1968**, *48* (5), 2374-.

(27) Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R.; Baker, A.; Stuetz, R.; Khan, S. J. *Anal. Methods* **2013**, *5* (23), 6557.

Chapter 3 : Improving Multidimensional Fluorescence Fingerprinting for Classification of Ecuadorian Shrimp Using Microwave-Assisted Extraction

3.1 Introduction

According to the National Fisheries Institute, shrimp has been the most consumed form of seafood in the United States.[1] This makes the quality of the shrimp available for consumption a primary concern. One method of investigating aquatic life is via molecular fluorescence. Fluorescence provides a fast measurement that can be useful in sample measurement, but sometimes it must be supplemented with more rigorous instrumentation such as gas or liquid chromatography.[2]

Coupling fluorescence with multiway analysis can provide much more information, as seen in the previous chapters of this work. Our research group has used PARAFAC with Soft Independent Modeling by Class Analogies (PARAFAC-SIMCA) for determination of geographic origin of samples of water and shrimp. [3,4] The motivations for the need of the classification has been described thoroughly in our previous work[4], and this current work builds on the established methodology.

3.1.1 SIMCA

Soft independent modeling of class analogy (SIMCA) is a classification method that uses principal component analysis (PCA) to describe sets of data representing defined classes. In a more traditional PCA analysis, a single PCA model would be used to differentiate the different classes of the dataset. This does not allow for any input of information specific to a class. SIMCA uses multiple PCA models to describe each class individually. A class is defined by a set of principal components specific to that class. A PCA model is built for each different class, using samples belonging to that class as calibration data. The total set of PCA models describing each class is considered one SIMCA model. Another set of samples, sometimes the same samples, is then projected onto each class's PCA model. The fit statistics for each PCA model, Hotelling's T^2 and the lack of fit statistic *Q*, are used to calculate a distance for each sample

$$d = \sqrt{(Q_r)^2 + (T_r^2)^2} \quad \textbf{(3-1)}$$

where within 1 distance unit means the sample is within the 95% confidence interval of the class. Samples with a calculated distance of greater than 1 unit do not belong to the class. The distinction of "soft" modeling indicates that samples can be assigned to any class, more than one class, or no class. "Hard" classification methods are stricter in the sense that they do not allow for various types of classification for a sample.[5]

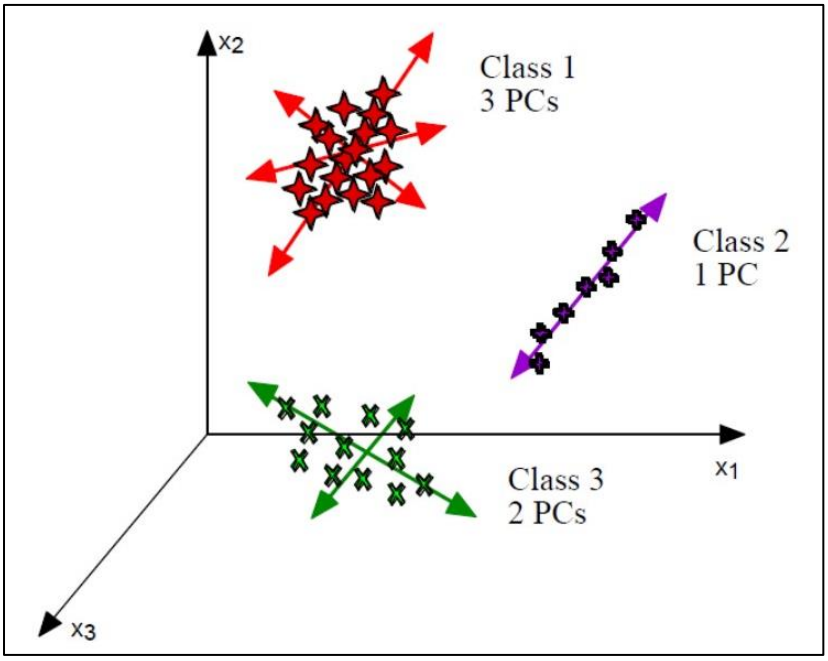


Figure 3-1: Graphical representation of SIMCA. For a given class, a PCA model is used to describe a calibration set of samples from that class. A SIMCA model is assembled from the principal components from each class's model. The test set of data is then projected onto the model and samples are classified.

Prior to the start of this project, two challenges were recognized. First, for the samples investigated, there was less geographic distance between sampling areas than the previous work.[4] . All the shrimp used in this work come from a single country, Ecuador. The locations of the three sites are indicated in Figure 3-2 and are within 100 kilometers of each other.[6] This differs greatly from our previous work, as the shrimp in that case came from various countries throughout the world and were thousands of kilometers apart. While the narrower geographical distance may cause our samples to have less diversity, and possibly more difficulty in classification, work by our group classifying water samples in a

narrow geographic region gave us reason to believe the shrimp could be classified.[7]

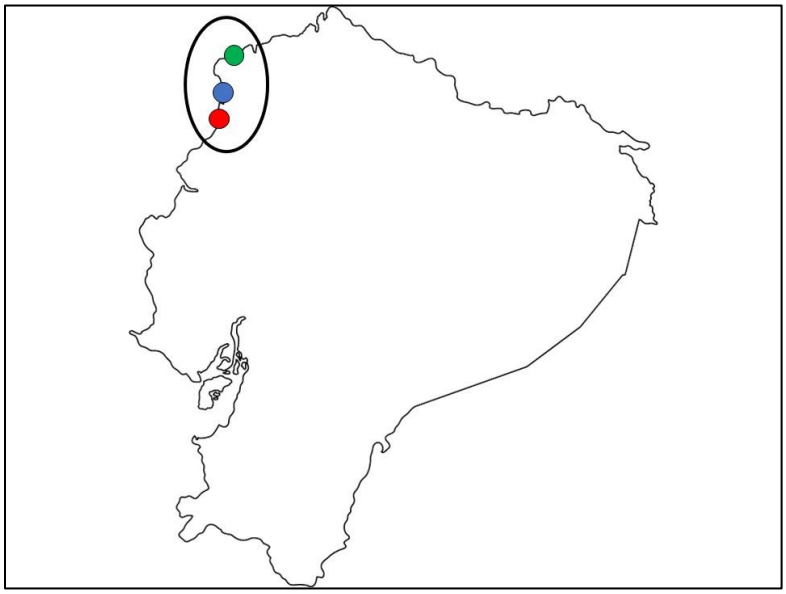


Figure 3-2: Sites where shrimp were collected for this work. The three sites were closest to the cities of Pedernales (red), Cojimies (blue) and Tonchigüe (green). The three sites are within 100 kilometers of each other. Adapted from Ref. [8]

Second, a larger number of shrimp were intended to be used in this experiment, when compared to previous work. There were originally over 100 samples intended in for this dataset. Later, in this project, difficulties with the samples required the dataset to be reduced significantly to its present size. Due to many samples needing to be processed in a limited window of time, the sample prep method from our previous work was not sufficient. To process the samples in the time required we have adapted the use of microwave extraction for this work. Microwaves were first used in the laboratory, for a purpose other than drying, by Abu-Samra et. al. to extract metals from an organic sample for further analysis.[9] Microwave extraction has been used to investigate PAHs in aquatic life in a method similar to the one used in this work.[10]

In a traditional reflux reaction, the vessel containing the reactant material is warmed using a rudimentary heat source such as a hot plate or Bunsen burner. This is an indirect method of heating the materials used in the reaction or extraction. Heat is transferred from the source, to a bath that is in contact with the reaction vessel, to the reaction vessel, to the solvent, and then finally to the reactants the experimenter is interested in. In the process of heat transfer from the source to the reactants, much of the energy is lost. A large amount of the energy input by the source is never used to heat the molecules of interest. This makes a more efficient heating method attractive.[11]

Microwaves have sufficient energy to activate rotation in polar molecules. When a microwave activates the rotation in a molecule, the energy absorbed by the molecule for this process is eventually released as heat. When the molecule is continuously bombarded by microwaves, more rotation occurs and more heat is released by the rotating molecule. In the case of reactions or extractions that involve microwaves, the solvent is normally the rotating molecule and the heat is released and absorbed by the reactants or reaction vessel. The more direct path between the heat source and the reactants results in a more efficient process. [11]

The purpose for using microwaves in the extraction process is to improve efficiency and scale for which we can process shrimp prior to EEM collection. The method for sample preparation in the previous work had a process rate of one shrimp per day. We find that the resulting extract is like the previous method and requires a fraction of the time to process each sample. The new method is able to process about 12 shrimp per day.

3.2 **Experimental**

3.2.1 Shrimp Collection

Shrimp used in this experiment were collected in three locations along the coast of Ecuador in Pedernales, Cojimies, and Tonchigüe. The shrimp were either collected by members of Dr. Acacia Alcívar-Warren's team or were purchased from local fishermen. Twelve shrimp were collected from each site. The shrimp were then frozen and stored in individual plastic bags. An approximately 5 g portion of the tail was removed for other work. The remainder of the shrimp, head and remainder of tail, was stored in a plastic bag. Shrimp were stored with dry ice in transport from Ecuador to our lab. The shrimp were stored in a -80 ^{o}C freezer until they were used for analysis.

3.2.2 Sample Preparation and Microwave Extraction

Prior to sample preparation, bagged shrimp were placed in warm water until they were thawed enough for further preparation. Once thawed, the shrimp was placed on a clean plastic cutting board and ~2 g of tissue was cut from the abdomen with a clean razor blade. The shell, and any legs, was removed prior to further preparation.

The original method used for sample preparation in our previous work was adapted from a method to extract PAHs from mussel tissue.[12] We modified our previous method to use microwaves with the guidance of an experimental procedure by McGowan and Leadbeater.[11] Similar methods involving microwave assisted extraction have been used in the past for shrimp and other aquatic

species. The 2 g portion was then placed in a glass microwave reaction vessel and 10 mL of 2 M KOH in 50:50 EtOH:H_2O solution was added. Samples were homogenized using a Polytron PT 10-35 homogenizer for approximately 2 minutes. After homogenization, a stir bar was added to the vessel prior to microwave extraction.

Once all the samples for a site (12 samples) had been processed as described, they were placed in a MARS-5 microwave oven (CEM) for extraction. The operating parameters for the MARS-5 microwave were: power 660W; ramp time of 6 minutes; temperature of 150 ^oC, low stirring and a run time of 15 minutes. In this step of the process, one shrimp sample from the Pedernales site was lost due to a loose cap on a reaction vessel.

After allowing the samples to cool, they were decanted into glass bottles. Each sample was filtered through a 0.45 μm Millipore nylon filter. The filtrate of each sample was then extracted with 10 mL of n-hexane. Aqueous and organic phases were collected. The organic phase extract was dried with Na_2SO_4. The organic and aqueous phase extracts were stored in glass bottles in the laboratory refrigerator until they were prepared for spectra measurement. At this point of the process, 35 shrimp samples were available for spectrum measurement, each with an organic phase portion and an aqueous phase portion for a total of 70 extracts. The organic phase extract was diluted by a factor of 13 in n-hexane and the aqueous phase extract was diluted by a factor of 250 with de-ionized water in preparation for EEM measurements. These dilutions were chosen after

preliminary experiments found an optimal dilution for obtaining samples with an absorbance of less than 1.

3.2.3 Spectra Collection

Spectra for each sample were collected on the same day as extraction. The absorbance spectrum of each organic and aqueous extract were collected using a Cary 300 UV-Vis spectrophotometer using a wavelength range of 230-600 nm in 1 nm intervals. EEMs for each extract were collected using a Cary-Eclipse spectrophotometer. The excitation wavelength range was 230-400 nm in 5 nm intervals and the emission range was 240-600 nm in 1 nm intervals. The photomultiplier tube voltage was set to 600 V. All measurements were taken in 1 cm quartz cuvettes.

3.2.4 Post-EEM Sample Evaluation

Prior to use in PARAFAC modeling, all 70 EEMs (aqueous and organic) were inspected manually. Several EEMs were deemed unsuitable for further analysis due to inadequate spectra. Eight shrimp samples were discarded from further analysis as either one the EEMs had intensity that had overloaded the signal. Two samples were from Pedernales; three samples were from Cojimies; and three samples were from Tonchigüe. Two samples from the Pedernales site were discarded due to low signal intensity. One Cojimies sample was discarded due to a contaminant. Based on examination of the spectrum, the contaminant was anthracene. This inspection resulted in 24 samples, for a total of 48 EEMs, being included in the PARAFAC models.

EEMs were prepared for analysis by removing first-order Rayleigh scattering by setting it to "not-a-number" (NaN). Instrumental effects were corrected for as well. Models were fit like work previously in this document. The results of the PARAFAC modeling were then used for SIMCA. Both PARAFAC and SIMCA analysis used MATLAB 2009a and PLS Toolbox (Eigenvector Research).

3.3 PARAFAC modeling

3.3.1 Organic Phase Samples

The following diagnostics are evaluated when choosing the best fitting PARAFAC model: core consistency, the percentage of the variance explained, the sum of squared residuals (SSR), and the number of iterations used by the algorithm. A summary of these diagnostics can be found in Table 3-1. These diagnostics are used to guide which models are considered for further evaluation and have been discussed in Chapter 1.[13]

Number of Components	Number of Iterations	Sum of Squared Residuals	Percentage of Data Explained	Core Consistency Value
1	115	4.97E+07	98.191	100
2	597	8.22E+06	99.700	99
3	739	2.88E+06	99.895	95
4	73	1.94E+06	99.929	76
5	947	1.25E+06	99.954	41
6	67	1.07E+06	99.961	< 0

Table 3-1: Diagnostic table for the organic phase models. A 4-component model was chosen for the organic phase samples and is shaded in grey.

Looking at the diagnostics, the number of iterations does not follow a pattern that lends consideration to any model. Therefore, the number of iterations for the algorithm to converge was not a primary consideration for choosing a

model. The percentage of variance explained increases as more components are fit to the data, which is expected. There is little change in the percentage of variance explained by a model once there are more than 3 components fit to it. The SSR decreases as more components are fit to the PARAFAC model. Less change is observed in the SSR as more than 4 components are used in the PARAFAC model. These diagnostics suggest a 3 or 4-component model is the best fit.

The core consistency values in Table 3-1 suggest a 3-component model is most appropriate. Other models with more components have a relatively low core consistency value.[14] This diagnostic was not considered in the previous work.[4] The core consistency diagnostic tends to not work well in experiments with higher complexity.[15] In samples that are not ideal, such as natural water samples, the best fit models exhibit negative core consistency values.[16] Therefore, the core consistency value will be noted, but not considered as a primary diagnostic.

Bearing in mind the information in Table 3-1, PARAFAC models with 3 and 4 components were considered for further evaluation. After evaluating spectra for both models, a 4-component model was determined to be the best model. The calculated spectra for the 4-component model of the organic phase samples are shown in Figure 3-3. The four components in this model are designated O1, O2, O3 and O4. The calculated spectra show appropriate excitation intensity prior to emission. There is an overall decrease in the smoothness of the spectra from component O1 to O4. Examination of the residuals did not show any evidence of other fluorescent components that were unaccounted for by the model.

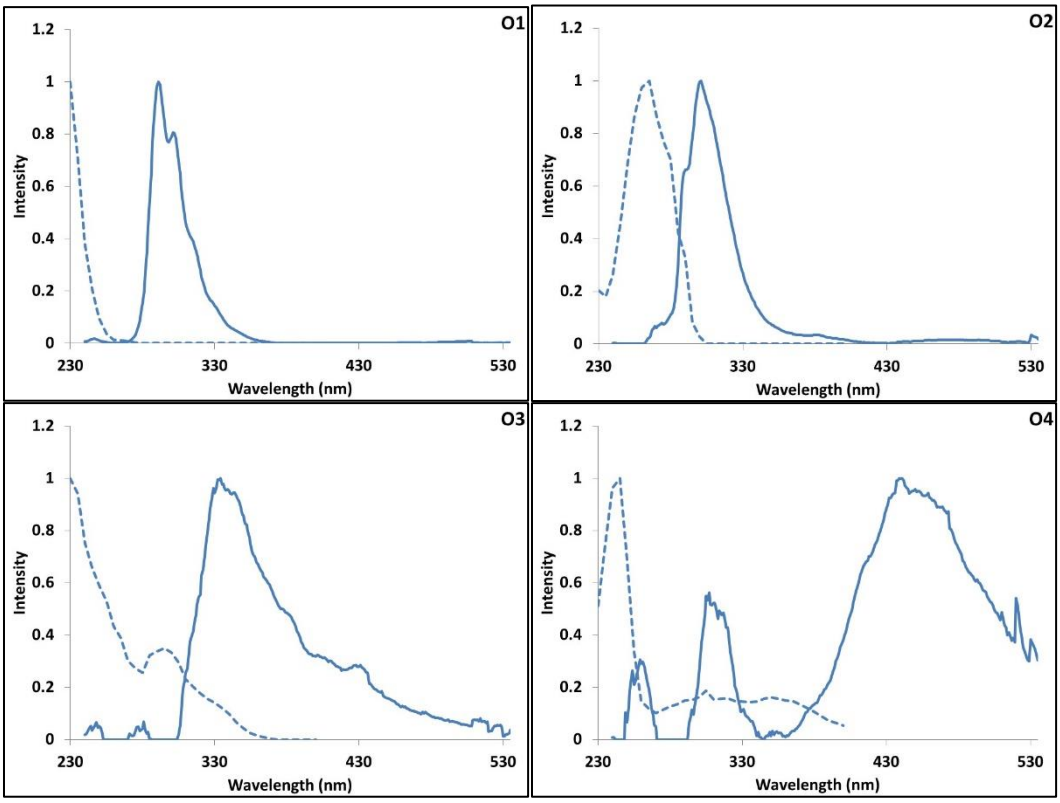


Figure 3-3: PARAFAC calculated Excitation (dashed) and Emission (solid) spectra for each of the components in a 4-component model (O1-O4) of the organic phase EEMs for 24 shrimp samples from Pedernales, Cojimies, and Tonchigüe.

Comparing the 4-component model to the previous work, there are some components that have a strong similarity. Components O1, O2, and O4 have a strong similarity to components 1, 3, and 5 of the organic phase in the previous work.[4] This helps give more support to a 4-component model. The similarities observed in spectra in both projects show some consistency between the two projects.

3.3.2 Aqueous Phase Samples

Next, the aqueous phase sample models were considered. Diagnostics for the aqueous phase models shown in Table 3-2 were evaluated similarly to the organic phase model assessment. The number of iterations did not show a pattern that suggests a certain number of components. The percentage of variance explained by increasing number of components shows little change after at least 3 components are used. Less change in the SSR is observed going from a 3-component model to a 4-component model. Both diagnostics would suggest either a 3 or 4-component model. The core consistency diagnostic suggests a 2 or 3 component model. The spectra for the 3 and 4 component models were chosen for further evaluation.

Number of PARAFAC Components	Number of Iterations	Sum of Squared Residuals	Percentage of Data Explained	Core Consistency Value
1	9	81830541	98.379	100
2	15	8807925	99.826	100
3	225	6545731	99.870	73
4	281	4097712	99.919	< 0
5	93	3665810	99.927	< 0
6	127	2520223	99.950	< 0

Table 3-2 Diagnostic table for the aqueous phase models. A 3-component model was chosen for the aqueous phase samples and is shaded in grey.

The decision to choose the 3-component model or the 4-component model was primarily based on a comparison of the first component in the 3-component model and the first two components in the 4-component model. A comparison of the calculated spectra for these components are shown in Figure 3-4. The first component of the 3-component model (3-1) shares a similar excitation spectra profile with the other two components of the 4-component model (4-1 and 4-2). Components 4-1 and 4-2 do not show unique excitation or emission spectra, when compared to the component 3-1. This is interpreted as an overfit of the model[13], and suggests a 3-component model is more appropriate.

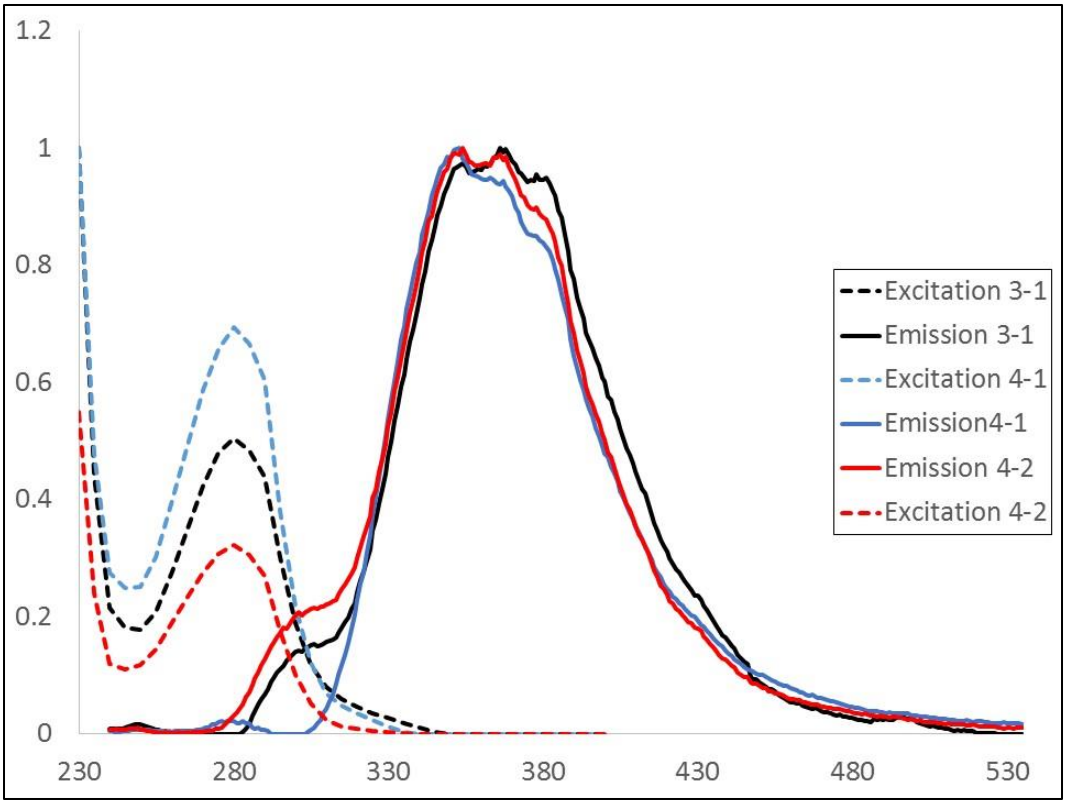


Figure 3-4: Comparison of the first component in a 3-component model (3-1) for the aqueous samples and the first 2 components of the 4-component model (4-1 and 4-2).

Looking at the variance explained per component can gives us further information for the spectra in the comparison of 3 and 4 component models. Table 3-3 shows the variance explained per component for 2, 3, 4, and 5-component models. From a 3-component model to a 4-component model, there is a noticeable difference in the variance attributed to the first two components in the 3-component model and the first two components in the 4-component model. This change helps affirm the assumption made when comparing the spectra calculated in both models.

Number of components in Model	Component's Fit Percentage of Model				
	1	***2***	***3***	***4***	***5***
2	95.5	4.5			
3	78.9	15	6.2		
4	47.8	44.8	5.9	1.5	
5	55.7	33.5	5.6	3.7	1.5

Table 3-3: Percentage of variance attributed to each component in the aqueous phase models. The percentage is calculated based on the amount of variance attributed to the individual component compared to the total amount of variance used in the PARAFAC model.

The spectra for the 3-component model are shown in Figure 3-5. The components of the PARAFAC model are labeled A1, A2 and A3. These spectra reflect realistic fluorophores and do not show any odd characteristics. Examination of the residuals showed no signs of other fluorescent components that should have been accounted for in the model. The components determined in this work have some similarities to the previous project[4], but they are not as comparable when compared to the organic phase model. Component A3 shares the same emission spectra as component one in the old work, but its excitation spectrum is more like component 2 in the old work. Component A2 has some spectral similarity with component 2 in the old work as well. Considering the two extraction methods were different, the amount of continuity between the two datasets is promising for future experiments and can help guide modeling.

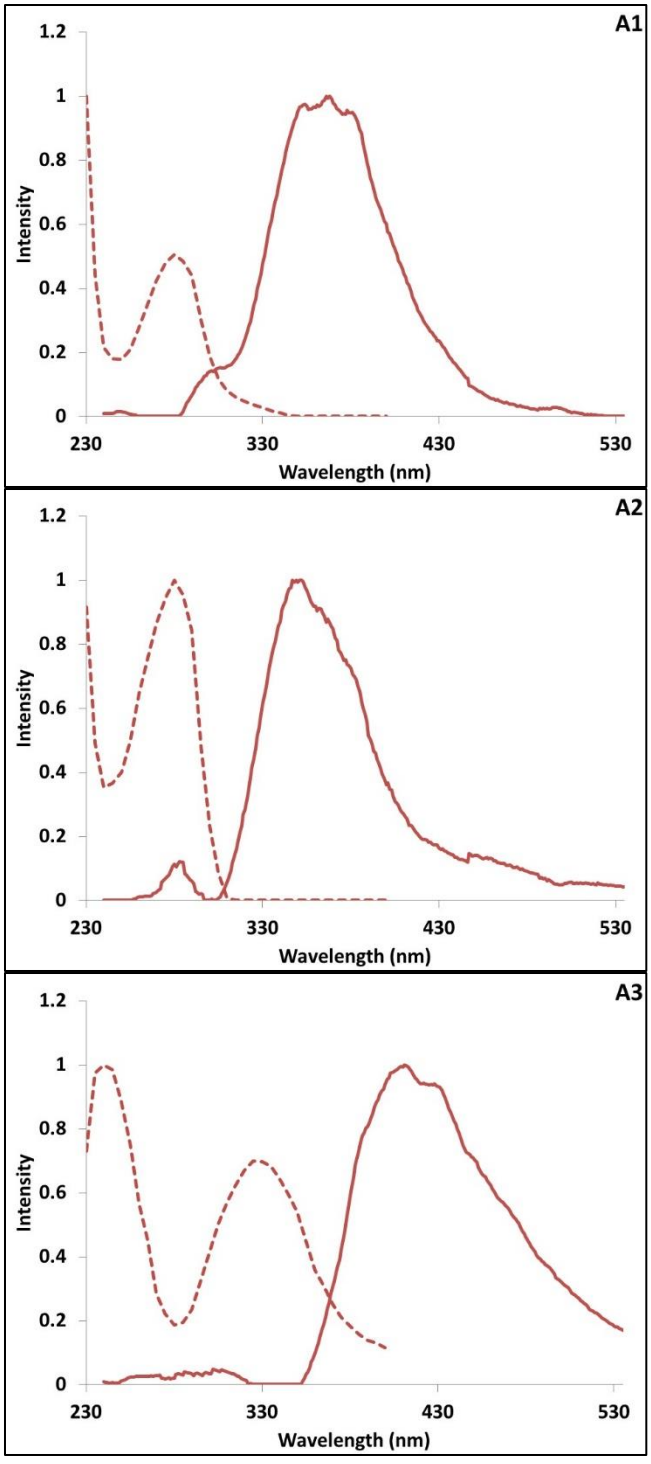


Figure 3-5: PARAFAC calculated spectra for a 3-component model (A1-A3) of the aqueous phase EEMs for 24 shrimp samples from Pedernales, Cojimies, and Tonchigüe.

3.4 Outlier and Split Half Analysis of PARAFAC Models

3.4.1 Outlier Determination for Aqueous and Organic Phase Models

Both models were evaluated for outliers using jack-knife PARAFAC. Jack-knife PARAFAC investigates the amount of influence each sample has on the overall PARAFAC model. IMP and RIP plots were produced from the jack-knife analysis and are used in the evaluation as described in Chapter 1. These plots are shown in Figure 3-6, Figure 3-7, Figure 3-8, Figure 3-9. Looking at these plots, it is obvious that samples 6 and 7 in the dataset are outliers. These samples originate from the Pedernales site. In the RIP plots for the excitation and emission loadings of the organic phase model, in Figure 3-7, both samples are separated from the rest of the dataset. Sample 6 does not cluster with the dataset in the RIP plots for the aqueous phase model in Figure 3-9. In both sets of IMP plots, samples 6 and 7 deviate from the line for component 4 in the organic phase model and in component 3 in the aqueous phase model. These samples were eliminated from the data set prior to SIMCA analysis.

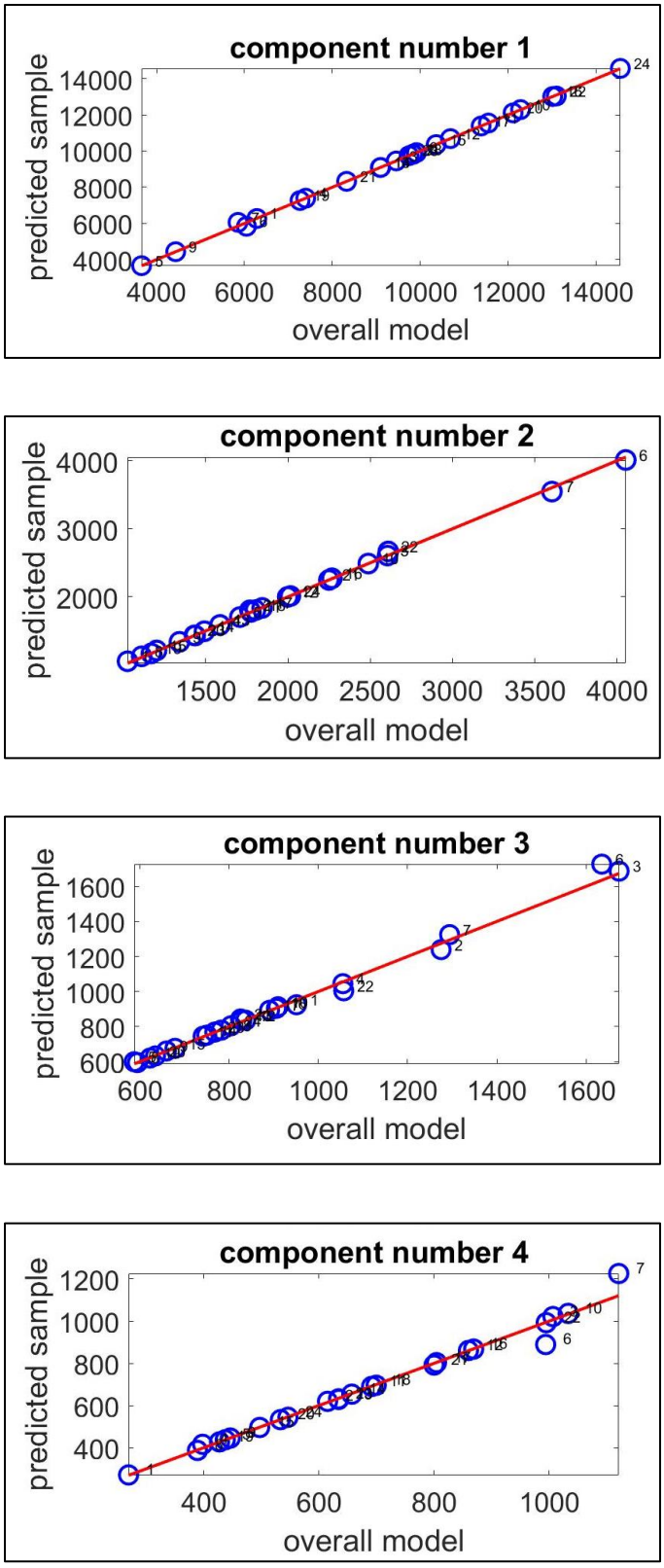
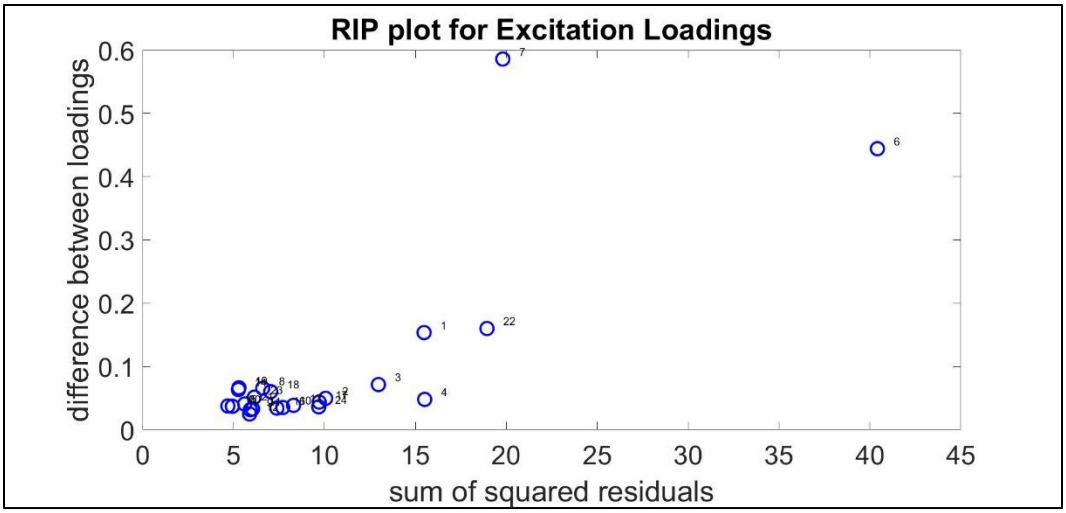


Figure 3-6: IMP plots for organic phase 4-component model.



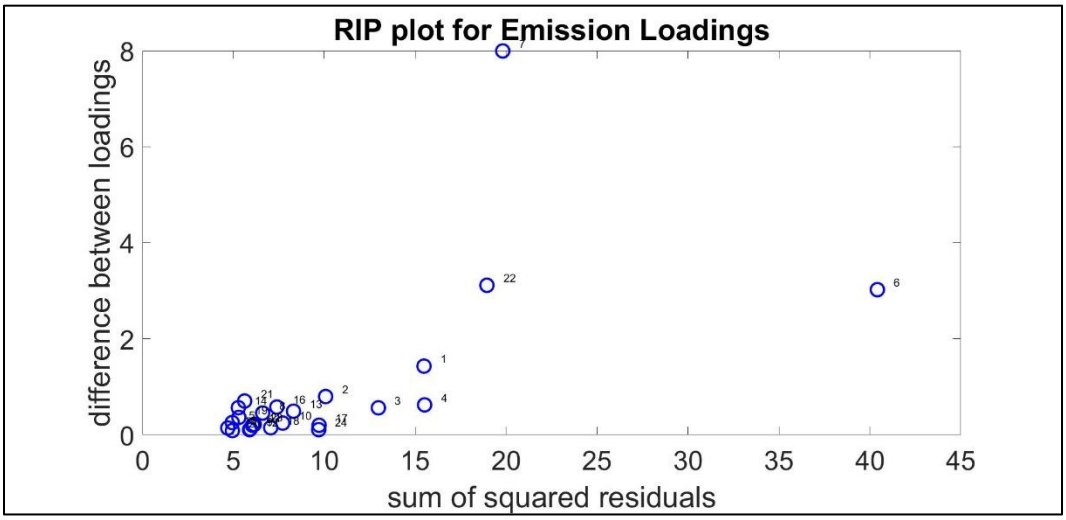


Figure 3-7: RIP plots for the 4 component organic phase model.

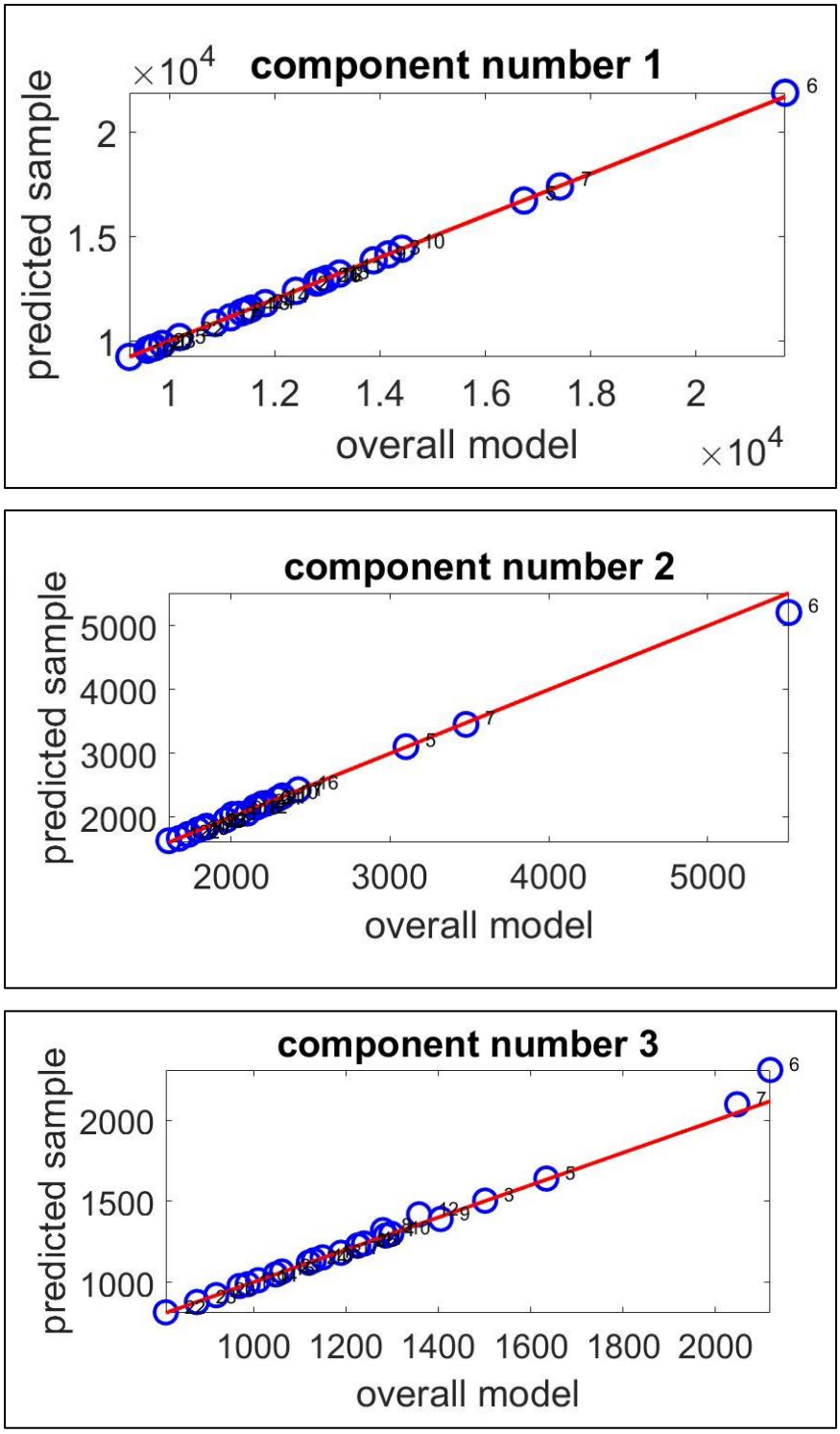


Figure 3-8: IMP plots for the aqueous phase model resulting from the jack-knife PARAFAC analysis.

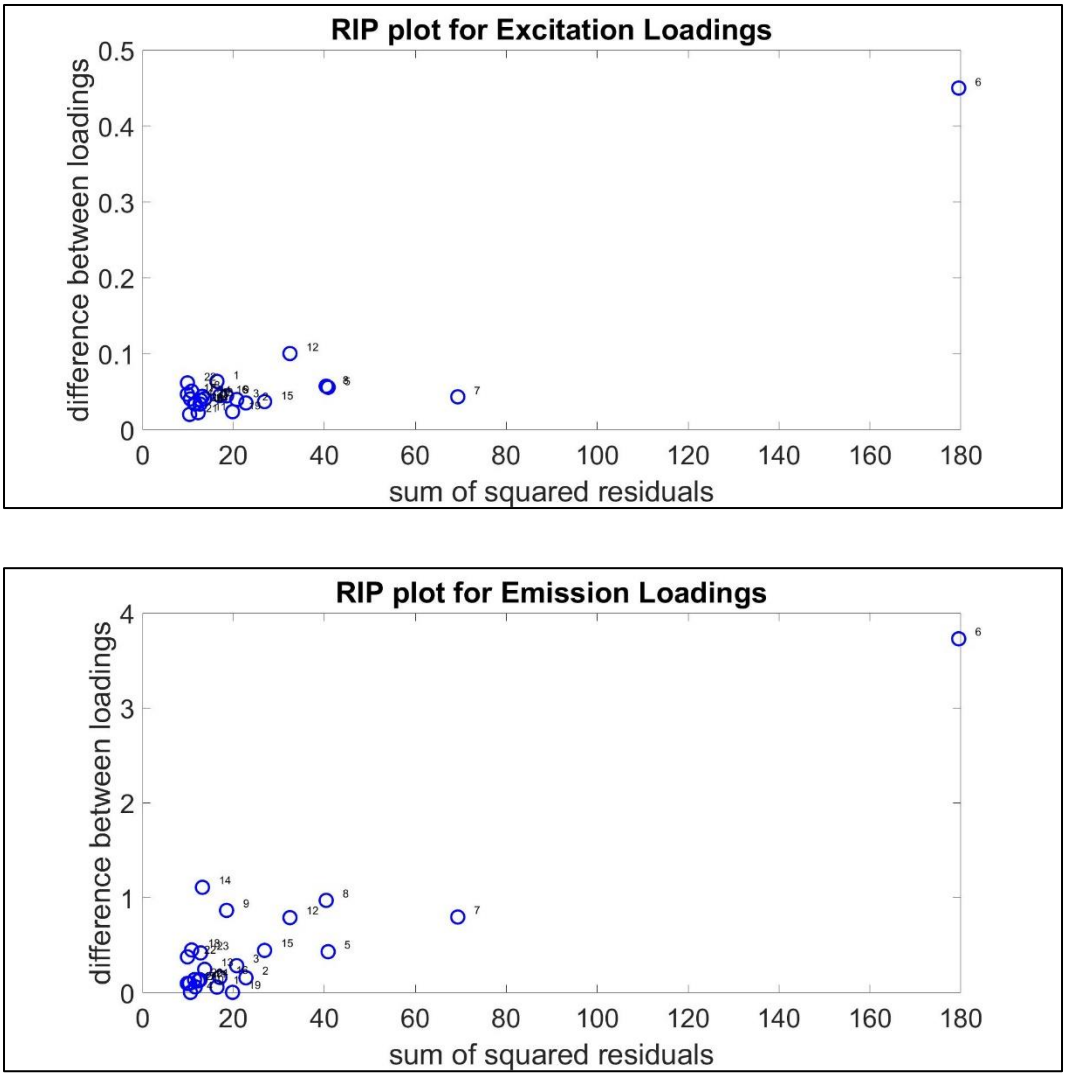


Figure 3-9: RIP plots from jack-knife analysis of the 3-component aqueous model

3.4.2 Split Half Analysis of Aqueous and Organic Phase Models

Split half validation creates separate PARAFAC models for two halves of the dataset and compares the models. Looking at the split half analysis spectra in Figure 3-10 there is a lot of similarity between all the spectra. Spectra from each half have either relatively perfect overlap, or they share much of the same features upon visual inspection. The program used for split half analysis in MATLAB was obtained through the DOMFluor toolbox.[17] It also validates the analysis by calculating Tucker congruency coefficients for the components of each half.[18] For this split half analysis, each component is validated with a calculated coefficient of 1 for each comparison. This validates our 4-component organic phase model and 3-component aqueous phase model. The score values from these models will be used in SIMCA modeling to classify the shrimp by location.

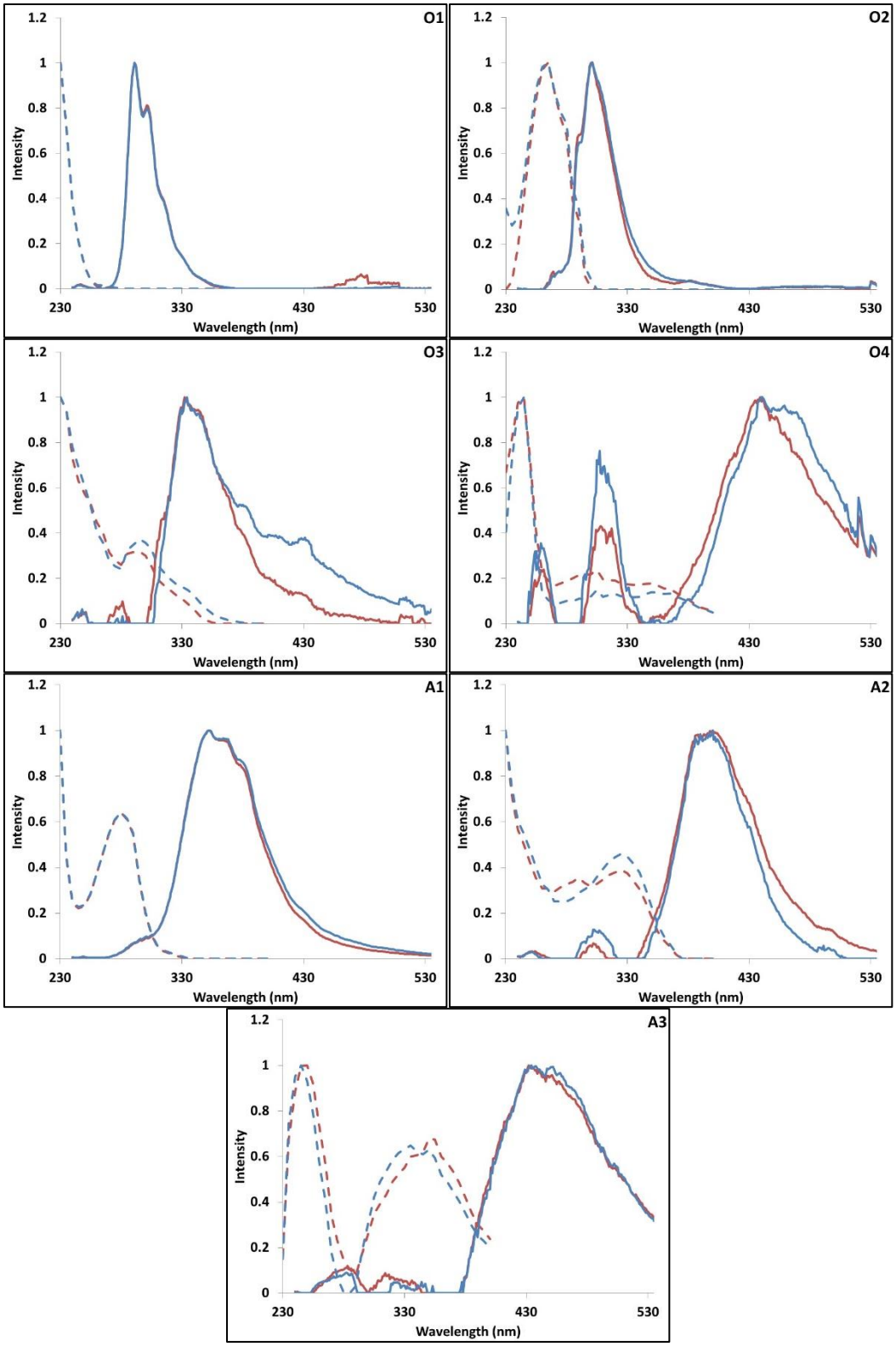


Figure 3-10: Split Half Analysis results for the organic phase (O1-O4) and aqueous phase (A1-A3) models. Excitation (dashed) and emission spectra (solid) are shown for half 1 (red) and half 2 (blue).

3.5 SIMCA Classification by Geographic Location

Twenty-two samples were used for SIMCA classification. The same samples were used for both the calibration and test sets. Prior to use in SIMCA, score values from PARFAC were normalized by dividing score values by the sum of all score values for that sample's PARAFAC component. This is to ensure no component is weighted more so than another component. This is particularly useful in this case as we are using score values from separate PARAFAC models (Organic and Aqueous). This SIMCA analysis used the score values from all 3 of the 4 organic phase components and all 3 aqueous phase components. The fourth component of the organic phase was omitted from the SIMCA analysis as it explains very little variance, as can be seen in Figure 3-11. This component was also omitted from SIMCA analysis in the previous work.[4]

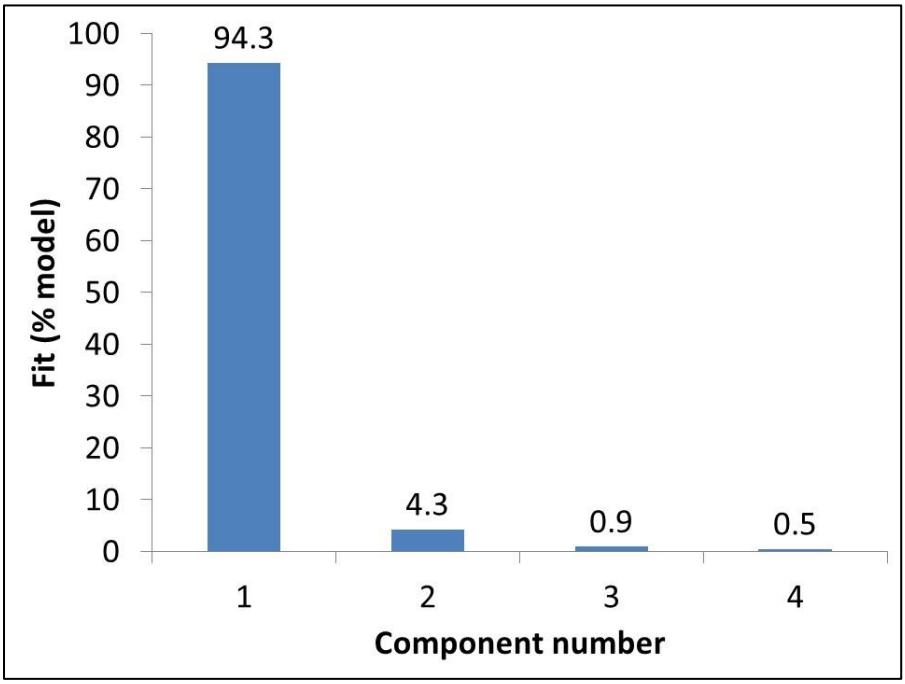


Figure 3-11: Variance explained by each component as a percentage of the model. Component 4 was omitted from SIMCA analysis.

All of the aqueous phase components were used in the SIMCA analysis as they each explained greater than 5 % of the variance, indicated in Table 3-3.

Each sample was labeled for classification by its site of origin (Pedernales, Cojimies, or Tonchigüe), making for a total of three classes. Each class was modeled using PCA components to explain a at least 98% of the variance. This resulted in the Pedernales class using 5 principal components, and the Cojimies class and Tonchigüe class using 4 principal components and 5 principal components, respectively. The SIMCA distances from this model are shown in Figure 3-12. In this case all samples classify properly with 91% accuracy. Samples were considered misclassified if: the sample did not classify their class or classified for another class it did not belong in. For this dataset, all samples classified properly, except for two Tonchigüe samples. While they did classify within the Tonchigüe class, they also classified for the Cojimies class. These samples are the 5^{th} and 9^{th} samples for the Tonchigüe class. These samples are indicated by the red bars in the SIMCA distance plot for Cojimes in Figure 3-12.

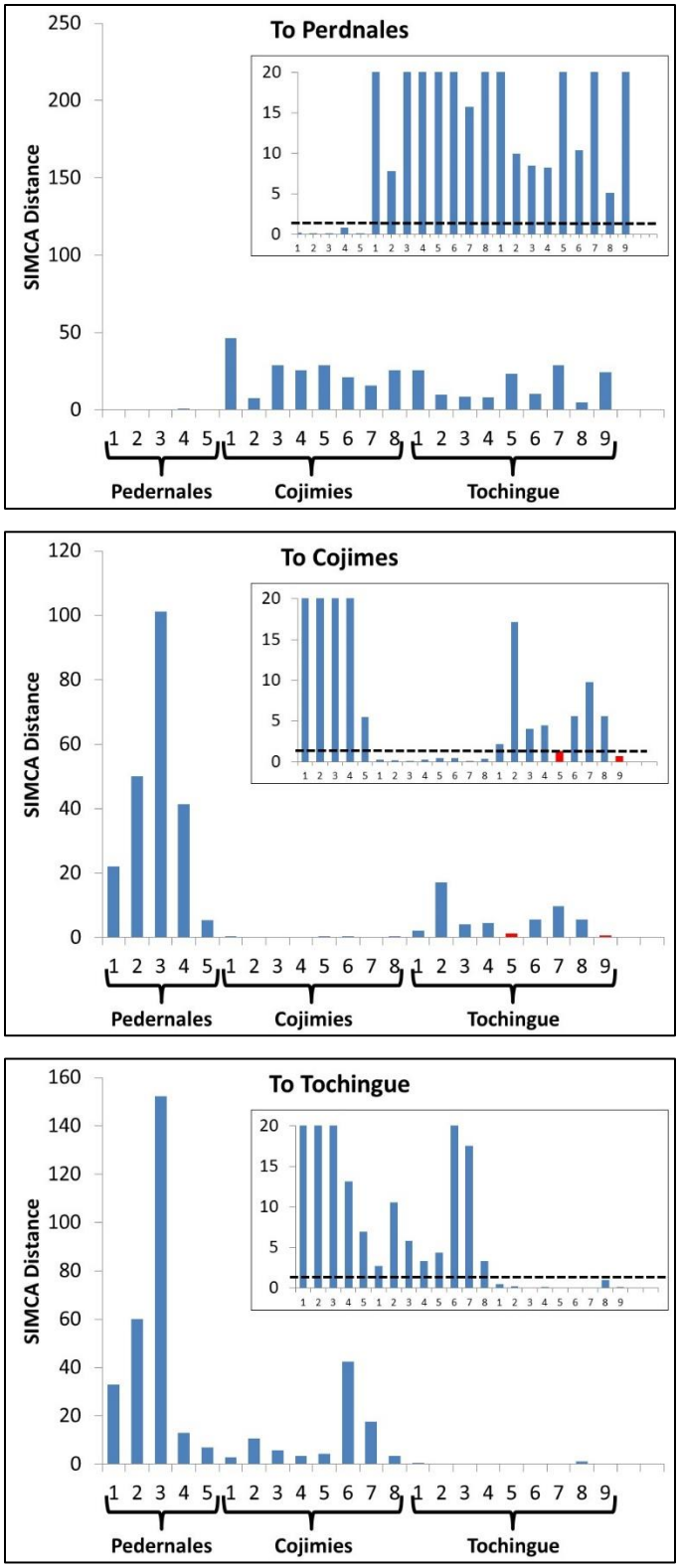


Figure 3-12: SIMCA Distances using all components of both PARAFAC models. Samples that did not classify properly are indicated by red bars.

3.6 Conclusions and Future Work

Each site originally started with 12 shrimp per site. The method in the previous work used more conventional reflux extraction and could allow one shrimp to be completely extracted and measured using our spectrometers in one day. Microwave extraction was implemented to improve sample processing. Using the microwave for extraction did improve the rate at which shrimp samples were processed and allowed all the shrimp in each site to be evaluated on the same day.

The microwave used in this work could hold 12 reaction vessels which allowed for all samples within a site to undergo the same exact extraction process. While we did not investigate the variation of samples within a given site with any other means other than what has already been presented, the success of the work here implies that there was no variation from sample to sample in a single site that impacted the results. While one of the samples was lost during the microwave extraction portion, this was due to a mistake when sealing the vessel, and was not a common problem in this experiment.

The use of microwave extraction is also more energy and cost efficient than the traditional reflux method. While there was not a rigorous cost benefit comparison between the two methods, there were several aspects of the procedure that are obvious improvements in this regard. First, the amount of heating time per sample was drastically reduced. In the previous work, a sample was refluxed for several hours.[4] In this work, the sample was heated for minutes. This is a fraction of the time and reduced the amount of energy needed per sample.

Combine this savings with the other benefit of being able to heat twelve samples at the same time, and the energy savings become even greater. As a group that teaches the "12 Principles of Green Chemistry"[19] in any class we teach, I think it is worth acknowledging efforts that achieve the goals of these guiding ethics.

One of the original goals that was not realized in this work was the use of separate samples in calibration and test sets. With the original number of shrimp per site, it was anticipated this could be another improvement. Initial results from splitting the datasets have been unsatisfying, but work is currently being done by another member of our lab to examine how successful this approach can be with the current dataset.

3.7 Acknowledgements

This work was accomplished with support from the Tufts University Summer Scholars program and from the Fundación para la Conservación de la Biodiversidad Acuática y Terrestre de Ecuador (FUCOBI) in Guayaquil, Ecuador. Additionally, this work would not have been possible without use of the laboratory microwave in the Chemistry Department at Merrimack College.

3.8 References

(1) National Fisheries Institute. U.S. Per-Capita Consumption By Species in Pounds https://www.aboutseafood.com/wp-content/uploads/2015/11/Top-Ten-Seafood-2015.pdf (accessed Jun 14, 2017).

(2) Morel, G.; Samhan, O.; Literathy, P.; Al-Hashash, H.; Moulin, L.; Saeed, T.; Al-Matrouk, K.; Martin-Bouyer, M.; Saber, A.; Paturel, L.; Jarosz, J.; Vial, M.; Combet, E.; Fachinger, C.; Suptil, J. *Fresenius. J. Anal. Chem.* **1991**, *339* (10), 699–715.

(3) Hall, G. J.; Kenny, J. E. *Anal. Chim. Acta* **2007**, *581* (1), 118–124.

(4) Eaton, J. K.; Alcivar-Warren, A.; Kenny, J. E. *Environ. Sci. Technol.* **2012**,

46 (4), 2276–2282.

(5) WOLD, S.; SJÖSTRÖM, M. In *Chemometrics: Theory and Application*; American Chemical Society, 1977; pp 243–282.

(6) Google Maps https://www.google.com/maps (accessed Jun 15, 2017).

(7) Hall, G. J.; Clow, K. E.; Kenny, J. E. *Environ. Sci. Technol.* **2005**, *39* (19), 7560–7567.

(8) Ecuador Outline Map http://www.worldatlas.com/webimage/countrys/samerica/outline/ecout.htm (accessed Jun 12, 2017).

(9) Abu-Samra, A.; Morris, J. S.; Koirtyohann, S. R. *Anal. Chem.* **1975**, *47* (8), 1475–1477.

(10) Zhang, H.; Xue, M.; Dai, Z. *J. Food Compos. Anal.* **2010**, *23* (5), 469–474.

(11) Leadbeater, N.; McGowan, C. *Clean, fast organic chemistry: Microwave-assisted laboratory experiments*; CEM Pub: Matthews, NC, 2006.

(12) Baumard, P.; Budzinski, H.; Garrigues, P. *Mar. Pollut. Bull.* **1998**, *36* (8), 577–586.

(13) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171.

(14) Bro, R.; Kiers, H. A. L. *J. Chemom.* **2003**, *17* (5), 274–286.

(15) Murphy, K. R.; Stedmon, C. A.; Graeber, D.; Bro, R.; Baker, A.; Stuetz, R.; Khan, S. J. *Anal. Methods* **2013**, *5* (23), 6557.

(16) Murphy, K. R.; Hambly, A.; Singh, S.; Henderson, R. K.; Baker, A.; Stuetz, R.; Khan, S. J. *Environ. Sci. Technol.* **2011**, *45* (7), 2909–2916.

(17) Stedmon, C. A.; Bro, R. *Limnol. Oceanogr.* **2008**, *6*, 572–579.

(18) Lorenzo-Seva, U.; ten Berge, J. M. F. *Methodology* **2006**, *2* (2), 57–64.

(19) Anastas, P. T.; Warner, J. C. *Green chemistry : theory and practice*; Oxford University Press, 1998.

Chapter 4: Investigation of β-Cyclodextrin Complexes with Polycyclic Aromatic Hydrocarbons using PARAFAC

4.1 Introduction

Cyclodextrins have been used in many different research fields for some time since their discovery in 1891.[1] Applications include chemical separation;[2,3] extraction of pollutants[4] and increasing drug solubility.[5] While these are only a few applications, it is obvious in a brief literature search, that these types of molecules still generate a lot of interest.

Generic cyclodextrins are made up of D(+)glucopyranose units connected by glyosidic bonds to form a cone shape. The most common commercially available cyclodextrins are made up of 6, 7 or 8 units and are named α, β, or γ-cyclodextrin, respectively. These cyclodextrins have a n inner cavity diameter that ranges from about 5 to 8 Å, with the α form being the smallest and the γ form being the largest form.[1]

This work concerns the host-guest interactions in water between β-cyclodextrin and polycyclic aromatic hydrocarbons, specifically naphthalene and

anthracene. The cyclodextrin has a relatively nonpolar cavity when compared to water[6] and PAH molecules of an appropriate size can form stable complexes within the cyclodextrin cavity.[1]

The equilibrium reaction for the 1:1 Naphthalene/β-cyclodextrin complex in water is:

$$\boldsymbol{N + C \rightleftharpoons NC} \qquad \textbf{(4-1)}$$

where naphthalene is labeled as $\boldsymbol{N}$, β-cyclodextrin is labeled as $\boldsymbol{C}$, and the complex is given by $\boldsymbol{NC}$. This complex was shown to be a 1:1 complex by Hamai.[7] Hamai measured the equilibrium constant using the steady state fluorescence using a Benesi-Hildebrand plot.[8] The 1:1 complex is an easy complex to measure this way, due to the large increase in fluorescence intensity of naphthalene in the cyclodextrin cavity.[7] This change in the fluorescence intensity of naphthalene with increasing β-cyclodextrin concentration can be seen below in Figure 4-1. The 1:1 complex has also been investigated spectroscopically by other researchers.[9–12] Similar cyclodextrin complexes have also been investigated using chromatographic methods.[13]

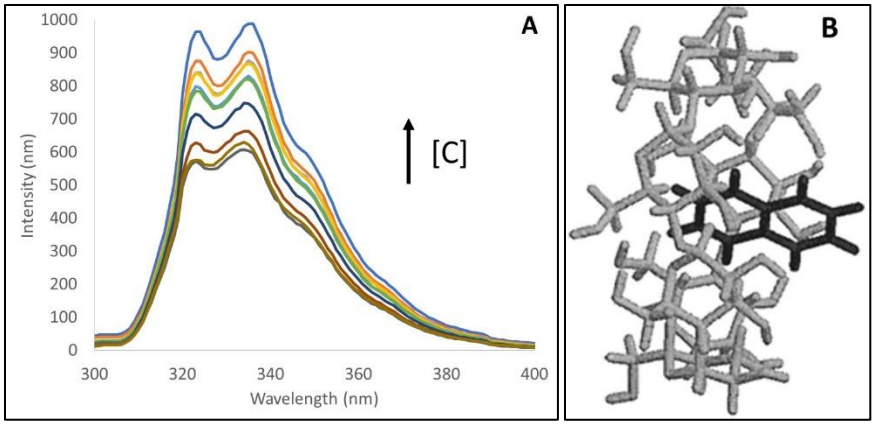


Figure 4-1: (A) The fluorescence spectra (excitation at 275 nm) for a sample of constant naphthalene concentration with an increasing β-cyclodextrin concentration in water. (B) Minimum energy structure of a 1:1 complex between naphthalene an β-cyclodextrin.[14]

At naphthalene concentrations greater than [N] = 5 x 10^{-6} M , a 2:2 complex can form,

$$NC + NC \leftrightharpoons (NC)_2$$

(4-2)

Where *NC* is the 1:1 complex between naphthalene and β-cyclodextrin and $(NC)_2$ is a 2:2 complex, or dimer between two 1:1 complexes.[7] For clarity, the 2:2 complex will be labeled as *D,* in expressions throughout the rest of this work. There is a local minimum energy structure for the 2:2 complex shown in Figure 4-2A.[14] The corresponding fluorescence of a sample containing a 2:2 complex is shown as well in Figure 4-2B. The structure shown is believed to be responsible for the emission at wavelengths greater than 375 nm. It is uncertain whether the 2:2 complex emission is separate from the naphthalene monomer.[14] There are other lower energy orientations of the naphthalene molecules within the 2:2 complex cavity that would not result in dimer fluorescence. It is proposed they

may fluoresce like any other 1:1 complexed naphthalene. Since there is no way of physically isolating the complexes in this system, it presents an interesting opportunity for PARAFAC analysis. Our group has investigated a similar system with micelles as the host molecule.[15]

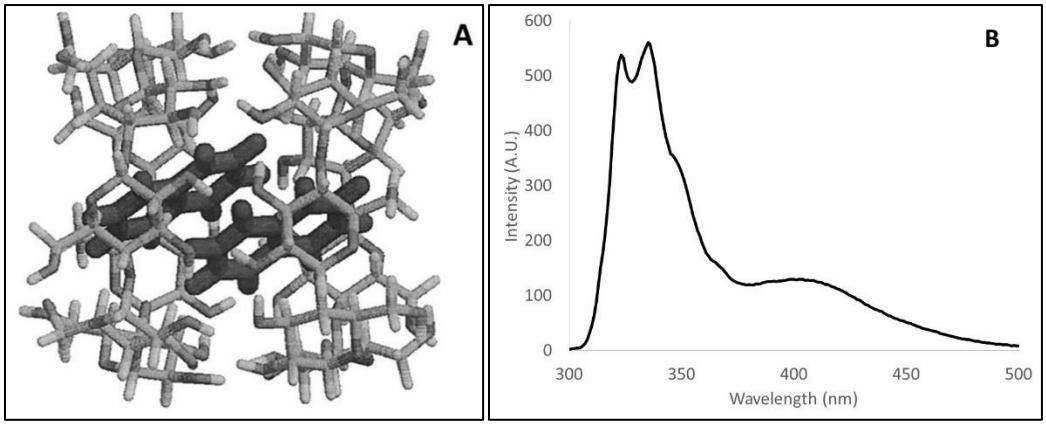


Figure 4-2: (A) Local minimum structure of a 2:2 complex for naphthalene and β-cyclodextrin most likely responsible for dimer fluorescence.[14] (B) Fluorescence spectra of naphthalene 2:2 excimer emission.

Naphthalene is not the only PAH that forms complexes with β-cyclodextrin, in fact many different PAHs form various complexes. Anthracene also forms a 1:1 complex with β-cyclodextrin. Few formation constants are reported in the literature for anthracene and β-cyclodextrin (2300 and 3870).[10,16] The spectrum for anthracene in a solution with β-cyclodextrin does not very much change. Neither source in the literature for this complex used a direct spectroscopic method for determining formation constants in water.

In this chapter, we will attempt to model the fluorophores in these systems with β-cyclodextrin using PARAFAC. We will validate these models by calculating formation constants for host-guest interactions between β-cyclodextrin

and the PAHs naphthalene and anthracene. For naphthalene, we will investigate the 1:1 complex and the 2:2 complex.[7] We will show a calculated PARAFAC spectrum for the 2:2 complex which has been hypothesized in previous research.[14] There will be two datasets for the 2:2 complex at different temperatures. We will also model a 1:1 complex between anthracene and β-cyclodextrin. One potential advantage to using PARAFAC is the ability to investigate multiple PAH/β-cyclodextrin complexes simultaneously and a dataset will be presented to showcase this application.

4.2 Experimental

4.2.1 Chemical Preparation

β-Cyclodextrin was purchased from Sigma Aldrich and recrystallized with Nanopure water (18 MΩ ionic purity) prior to use. Naphthalene (Aldrich 99+%) was recrystallized prior to use using methanol. Anthracene (Aldrich 99%) was recrystallized in methanol prior to use. All samples were prepared using Nanopure water (18 MΩ ionic purity).

4.2.2 Experimental Timing

One complication with using β-cyclodextrin is that it has poor solubility in water, and is known to aggregate in solution.[17,18] It is used in these experiments due to its availability and its common use in other research. To minimize the effects of this issue, experiments were done in as timely a manner as possible. At the least samples were prepared and measured within 24 hours of the β-cyclodextrin being recrystallized prior to use.

4.2.3 1:1 Naphthalene β-cyclodextrin Experiment

For the naphthalene 1:1 complex experiment, a stock solution of naphthalene was prepared by stirring a saturated solution of naphthalene in conductive grade water for approximately 2 days and then filtering with a 0.45 μm filter. A stock solution of β-cyclodextrin was prepared to 0.010 M in conductive grade water and stirred until dissolved under low heat. The resulting solution was then filtered using a 0.45 μm filter. This solution was cooled to room temperature prior to use.

Samples were prepared using the stock solutions to have a concentration of naphthalene of about 5×10^{-6} M and varying concentrations of β-cyclodextrin from 0 to 10 mM. A total of 10 samples were used in the dataset presented. Naphthalene concentrations were kept constant and verified to be about 5×10^{-6} M to ensure no formation of other complexes.[7]

EEMs were collected for each sample using a Varian Cary Eclipse fluorescence spectrophotometer. The photomultiplier tube voltage (PMT) was set at 725 V. Excitation wavelengths of 250-300 nm in 5nm increments were used. Emission was measured from 300-400 nm in 1 nm increments. Excitation and emission band widths were 5 nm each. Absorbance spectra were measured using a Varian Cary 300. The maximum absorbance of any sample in the wavelength range of interest was <0.4. Therefore, fluorescent measurements were corrected to be optically dilute.[19] EEMs were prepared for analysis by removing first-order Rayleigh scattering and first-order Raman scattering by setting it to "not-a-

number" (NaN).[20] Instrumental effects were corrected for as well.[21] Models were fit according to the details outlined in Chapter 1.[20]

4.2.4 2:2 Complex for naphthalene and β-cyclodextrin

Stock solutions were prepared similarly to the 1:1 experiment. A total of 11 samples were prepared for the dataset. The total concentration of naphthalene in each sample was 5 x 10^{-5} M. This concentration was chosen such that it would be high enough to form a 2:2 complex, but leave the concentration of β-cyclodextrin constant. Samples were measured on the same instrumentation used in the 1:1 naphthalene complex experiment.

The wavelength range used for EEMs was 250-300 nm, in 5 nm steps, for excitation and 300-500 nm, in 1 nm steps, for emission. The PMT voltage for the fluorometer was 550 V. The slit widths of the monochromators for excitation and emission were 5 nm. Measurements for each sample were made at 25 ^{o}C and 10 ^{o}C. Samples were cooled in the temperature controlled cuvette holder for 10 min prior to measurements. EEMs were corrected the same way prior to PARAFAC modeling as in the 1:1 naphthalene complex experiment.

4.2.5 1:1 Complex for Anthracene and β-Cyclodextrin

A solution of anthracene in conductive grade water was made by adding several drops of a saturated solution of anthracene in methanol. The solution was heated until slightly warm, then allowed to cool to room temperature. This solution was then filtered using a 0.45-micron filter. This solution was then added to various amounts of β-cyclodextrin to make up the samples used in this dataset.

Concentrations for β-cyclodextrin in each sample ranged from 0 mM to 10 mM. The total solution volume of each sample was 4.00 mL. A total of 5 samples were prepared for the dataset. The total concentration of anthracene in each sample was assumed to be 2 x 10^{-7} M, the solubility limit of anthracene at 25 ^{o}C.[22] Samples were measured on the same instrumentation used in the 1:1 naphthalene complex experiment. The wavelength range used for was 250-350 nm, in 5 nm steps, for excitation and 300-500 nm, in 1 nm steps, for emission. The PMT voltage for the fluorometer was 700 V. The slit widths of the monochromators for excitation and emission were 5 nm. EEMs were corrected the same way prior to PARAFAC modeling as in the 1:1 naphthalene complex experiment.

4.2.6 Simultaneous Anthracene and Naphthalene Experiment

The samples in this experiment contained both PAHs anthracene and naphthalene. The total concentration of each PAH was constant for all samples. Samples from the 1:1 anthracene complex experiment were used for this experiment as well. 0.100 mL of Naphthalene stock was added to each sample resulting in a total solution of 4.10 mL. The approximate total concentration of naphthalene and anthracene in each sample was 4 x 10^{-6} M and < 2 x 10^{-7} M, respectively. Samples were measured on the same instrumentation used in the 1:1 naphthalene complex experiment. The wavelength range used for was 250-350 nm, in 5 nm steps, for excitation and 300-500 nm, in 1 nm steps, for emission. The PMT voltage for the fluorometer was 700 V. The slit widths of the monochromators for excitation and emission were 5 nm. EEMs were corrected the

same way prior to PARAFAC modeling as in the 1:1 naphthalene complex experiment.

In this experiment, due to the small number of samples used, several PARAFAC settings were changed to better fit the model to the dataset. The maximum number of iterations was increased from 10,000 to 20,000. The stop criteria for convergence of the algorithm was decreased to a change in fit of 1×10^{-10}. These changes were made when initial PARAFAC models were determined to be unsatisfactory.

4.3 PARAFAC Modeling

4.3.1 1:1 Complex for Naphthalene and β-Cyclodextrin

As with any other experiment, the first step in choosing the best fit model is examining the PARAFAC diagnostic table shown below in Table 4-1. Examining the number of iterations, there is a large increase in the number of iterations from a 1-component model to a 2-component model. There is a comparable increase in the number of iterations when adding a 3rd component and a 4th component as well. While it is obvious the algorithm is working hard to converge on a model, there is a large spike in the number of iterations from a 1-component model to a 2-component model. This diagnostic suggests either a 1 or 2-component model.

The rest of the diagnostics suggest a 2-component model as well. The sum of squared residuals (SSR) has a sizeable decrease going from a 1-component to a 2-component model. This decrease in the sum of squared residuals is not

significant in models with more than 2 components. The percentage of variance explained is increased when a second component is added to the model, but there is no significant improvement when adding a third or fourth component to the PARAFAC model. The core consistency value also drops dramatically when a third component is added.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	39	2.50E+06	99.923	100
2	1131	1.50E+05	99.995	100
3	2277	1.40E+05	99.996	<0
4	3783	6.52E+04	99.998	<0

Table 4-1: Diagnostic table for PARAFAC models used in 1:1 Complex experiment for Naphthalene and β-cyclodextrin. The best fit model, the 2-component model, is shaded.

Examination of the PARAFAC spectra for a 2-component model helps confirm its validity. Using the spectra, we can suggest that component 1 describes the naphthalene complexed with β-cyclodextrin. Since the first component of a PARAFAC model explains the most variance[20], this should make sense as the fluorescence intensity of naphthalene when complexed with β-cyclodextrin is much higher.[7,14] The second component describes the naphthalene free in the solution. Component 2 matches very well to the spectrum of naphthalene in water.[23] Since the cavity of a cyclodextrin is relatively nonpolar, when compared to water, one might assume that the spectrum should match the spectrum of naphthalene in a nonpolar solvent. This is true for component 1, as it matches the spectrum of naphthalene in cyclohexane very well.[23] There is also noticeable

redshift in the excitation spectrum for the complexed naphthalene, which has been observed in other work.[7,14]

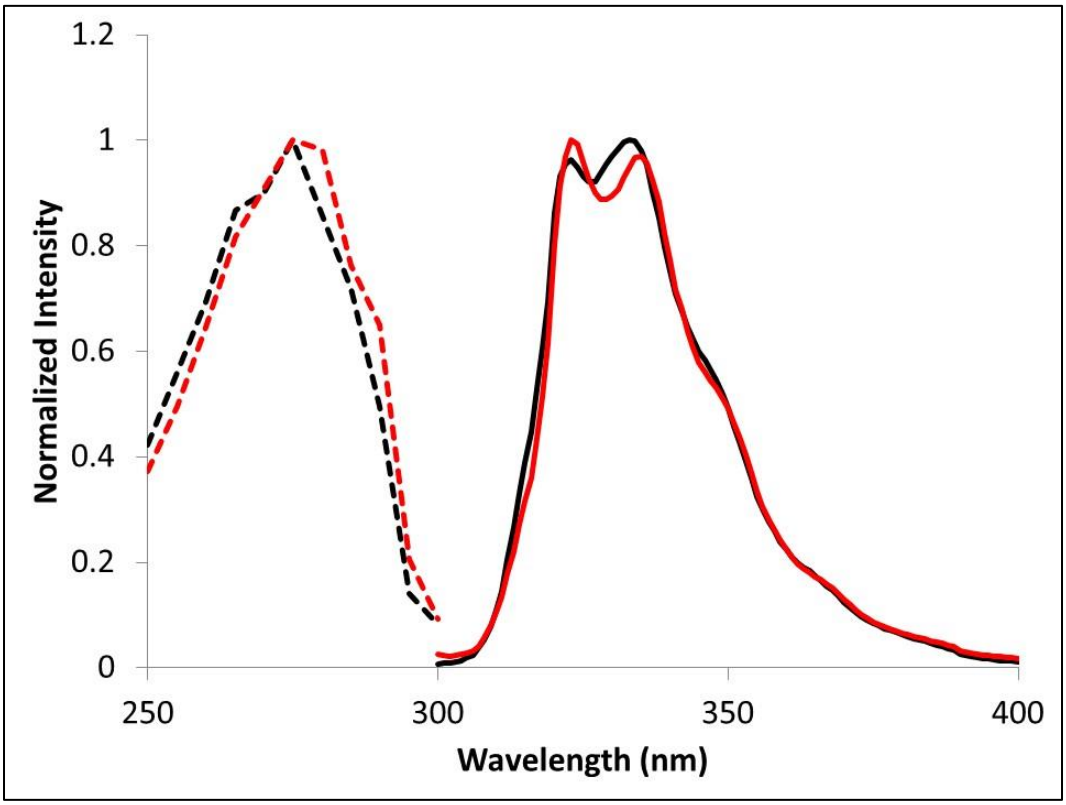


Figure 4-3: PARAFAC calculated spectra for the free naphthalene component (Excitation --- Emission —) and for the complexed naphthalene component (Excitation --- Emission —)

A plot of the score values for each component is shown in Figure 4-4. The scores for each component show the trends we hypothesized. The score value for the free naphthalene should have the highest value for the sample with 0 mM β-cyclodextrin. The sample should also have a score value of zero for the 1:1 complexed naphthalene component, as there is no β-cyclodextrin present. As the concentration of β-cyclodextrin increases, the score value for the free naphthalene gradually decreases and the score value for the 1:1 complexed naphthalene component gradually increases. If the score values had units of pure

concentration, they would add up to a constant value for each sample, since the amount of naphthalene is constant. The reason the score values of the 1:1 complexed naphthalene trend to a higher value than the maximum score value of free naphthalene is due to the higher molar absorptivity of 1:1 complexed naphthalene.[7]

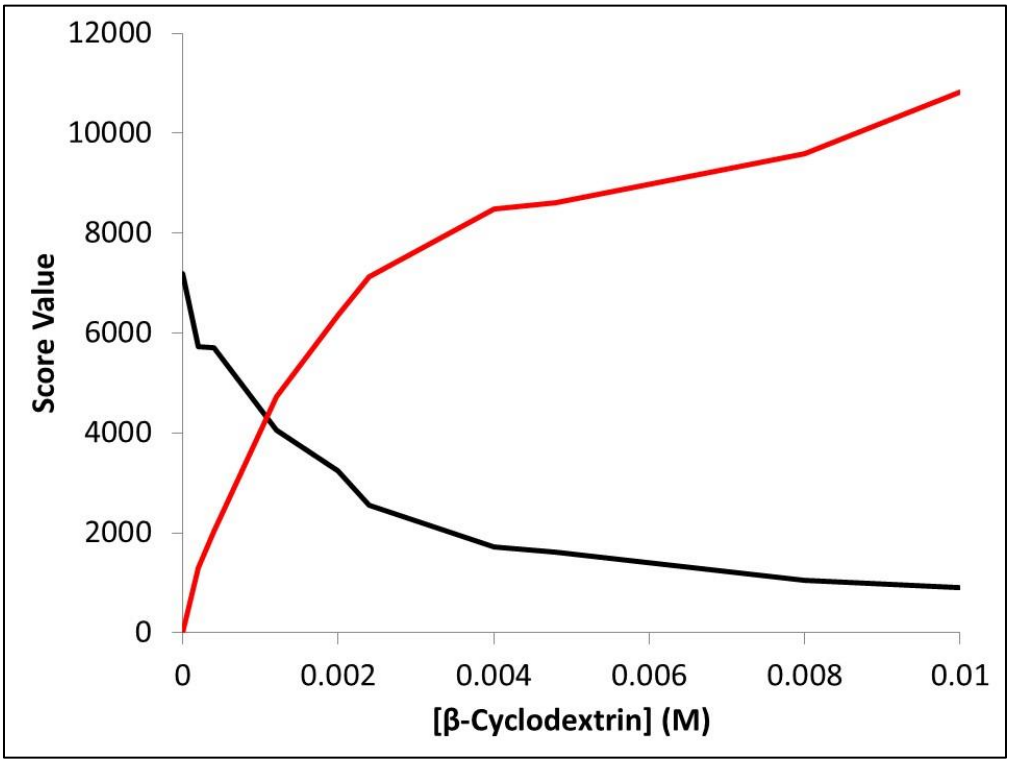


Figure 4-4: Score value plot for the 2-component model in the 1:1 naphthalene complex experiment. Score values are plotted for the PARAFAC components that describe free naphthalene (—) and 1:1 complexed naphthalene (—)

[B-Cyclodextrin] (M)	Score Value	
	Free	***Complexed***
0.01	10824	908
0.008	9582	1061
0.0048	8613	1622
0.004	8495	1714
0.0024	7131	2565
0.002	6365	3236
0.0012	4722	4064
0.0004	2042	5716
0.0002	1306	5730
0	0	7189

Table 4-2: Score Values for 1:1 complex experiment for naphthalene and β-cyclodextrin

The 2-component model was evaluated for outliers using the jack-knife PARAFAC method.[24] The RIP plots and the IMP plots can be found in Figure 4-5 and Figure 4-6. Examination of the RIP plots result in no outliers being identified. Sample 1 appears slightly deviated from the rest of the dataset, but it does not appear to be an obvious deviation as has been seen in published examples.[24] There are no outliers identified via the IMP plots either. While some of the points deviate slightly from the line, none appear to be an outlier.

Split half analysis was not as useful for validating this dataset. The analysis resulted in the data being deemed "invalid". Due to the similarity between the free and complexed spectra, Tucker congruency coefficients (TCC) were equal to 1 for both compoenents in both halves.[25] A model is validated by this analysis, if it calculates a TCC equal to 1 for two corresponding components in separate halves, and calculates a TCC equal to 0 for other components within the same half.[26] Split half spectra are shown for each component in Figure 4-7. While the method may not have quantitavely validated the model, examination of

the split half spectra in Figure 4-7 show that both halves produce a nearly identical model.

In this case, the 2-component model is the most appropriate. The diagnostic table suggests a 2 component model. The spectra of the the 2 component model show features that we expect from previous literature[7,14] Therefore this model will be used for determination of the formation constant, K_1, for the 1:1 complex of naphthalene and β-cyclodextrin.

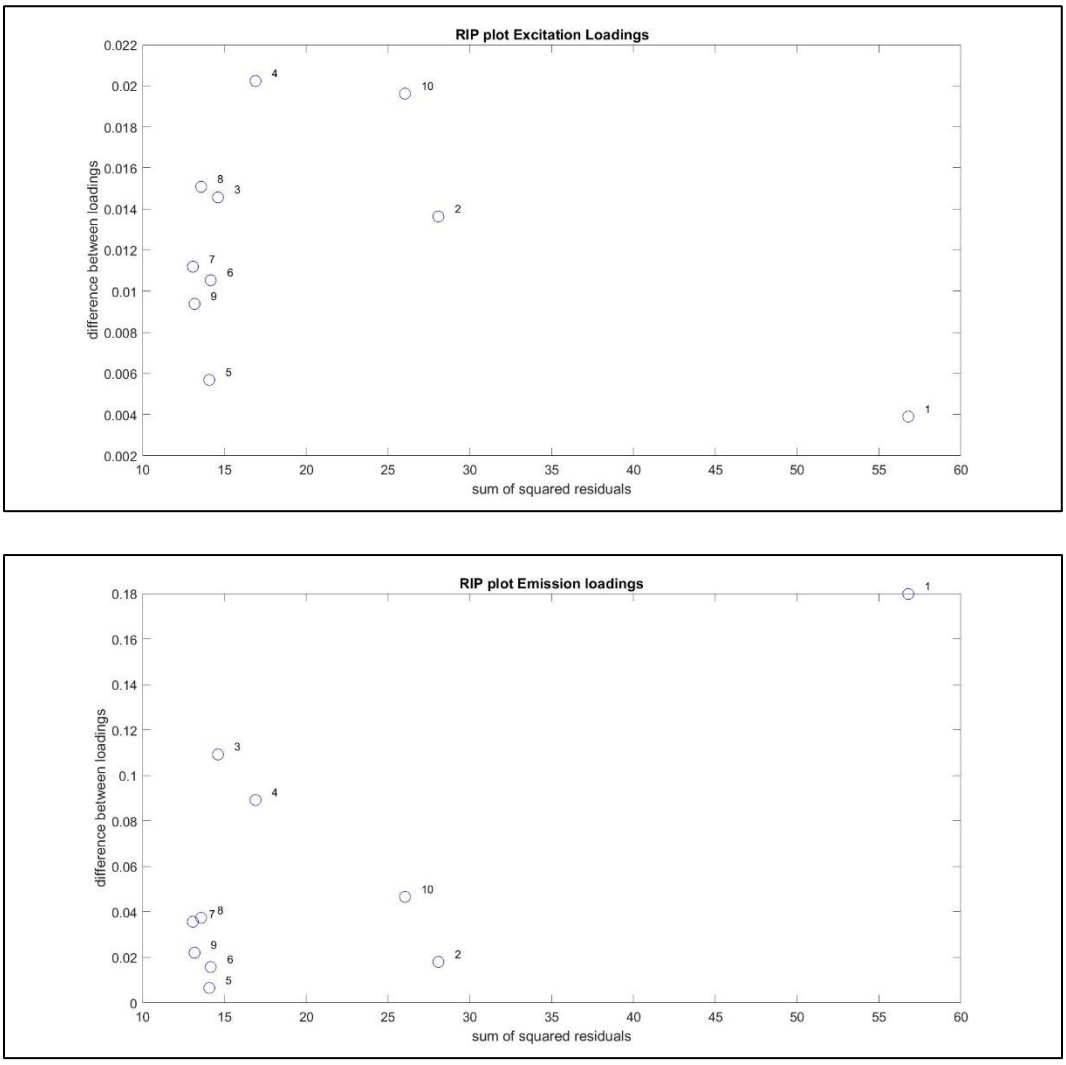


Figure 4-5: RIP plots for the excitation and emission spectra for the 2-component PARAFAC model of the 1:1 complex between naphthalene and β-cyclodextrin.

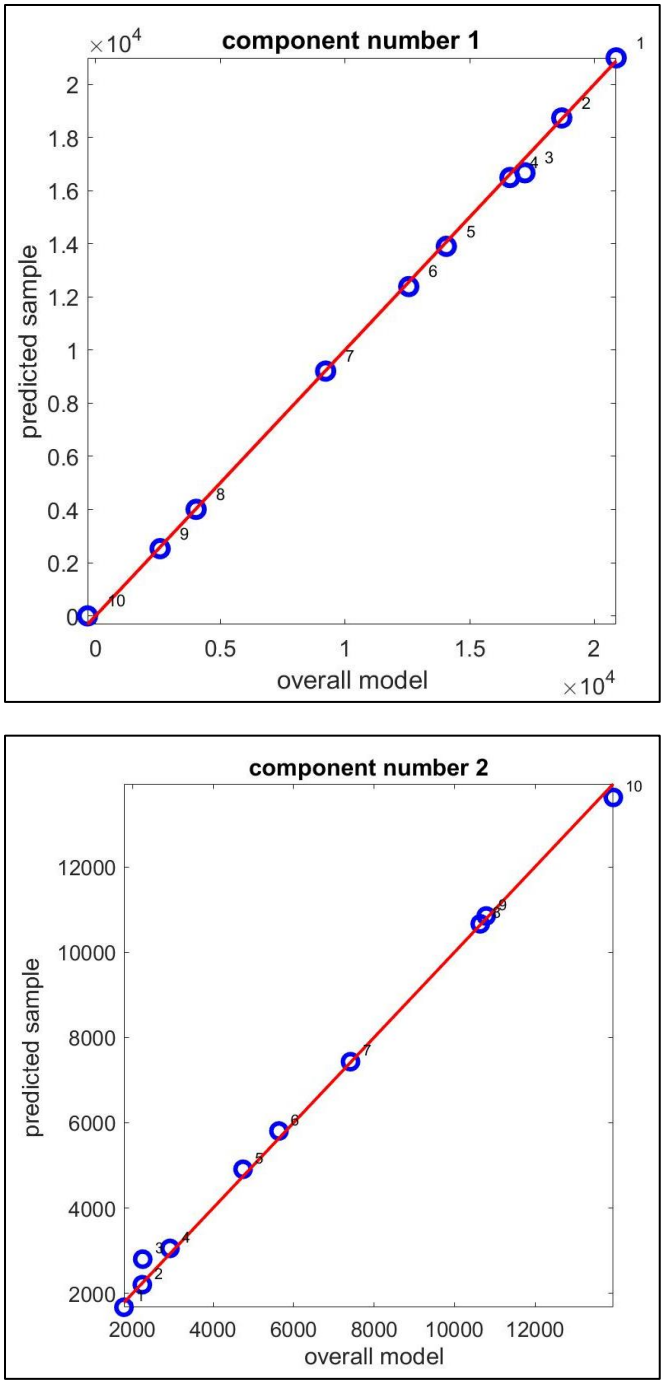


Figure 4-6: IMP plots for the excitation and emission spectra for the 2-component PARAFAC model of the 1:1 complex between naphthalene and β-cyclodextrin.

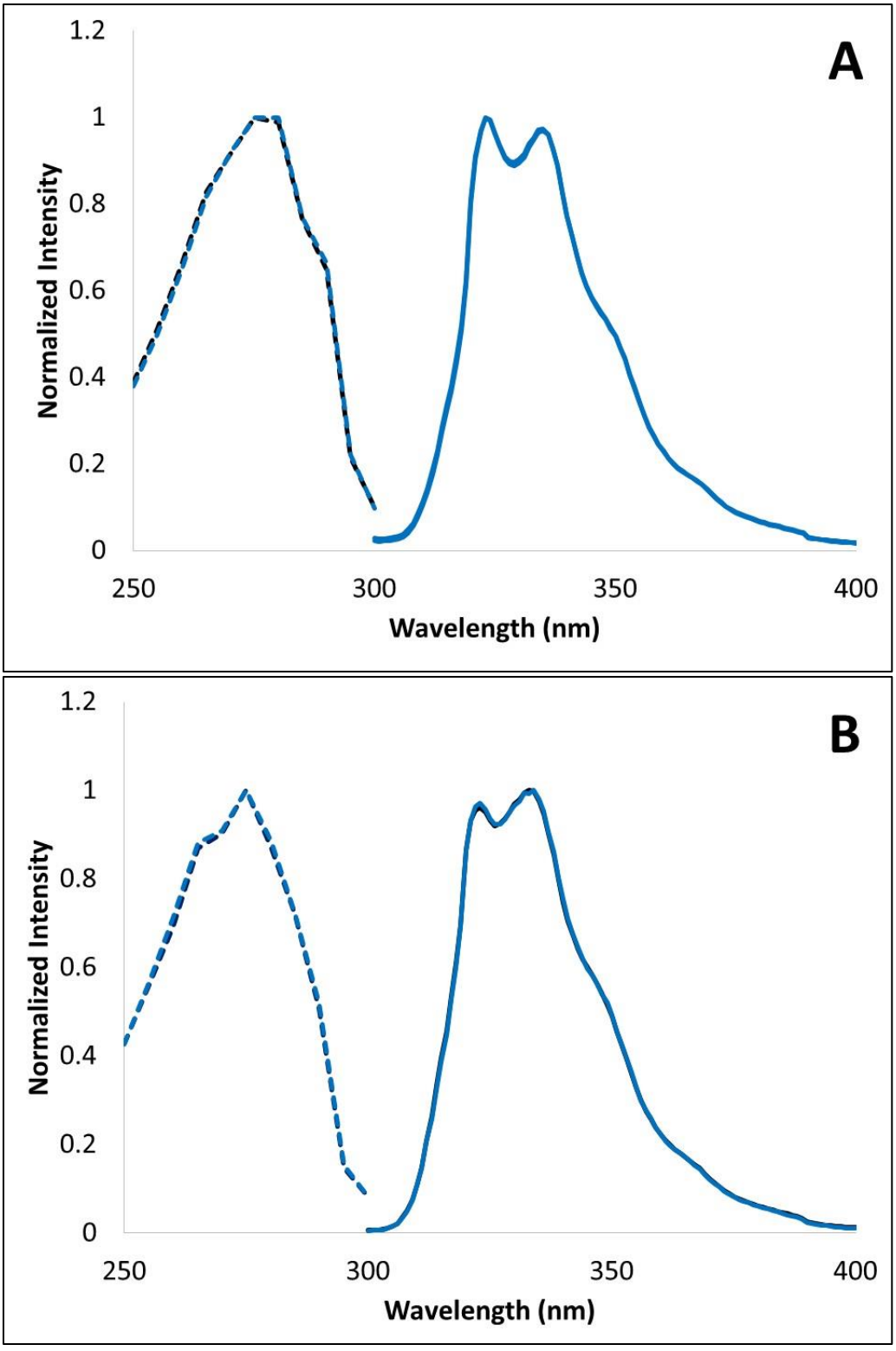


Figure 4-7: Split half spectra for the 2-component PARAFAC model for the 1:1 complexed naphthalene experiment A: 1:1 complexed naphthalene component; B: free naphthalene component (Half A: black; Half B: blue; Excitation: dashed; Emission: solid)

4.3.2 2:2 Complex Naphthalene Experiment

The diagnostic table for the PARAFAC models is shown in Table 4-3. The number of iterations does not show a pattern that suggests a model. The number of iterations increases drastically going from a 2-component model to a 3-component model. Going from a 3-component model to a 4-component model, there is a drastic reduction in the number of iterations. The lack of a pattern suggests this diagnostic is not useful in determining an appropriate model in this case.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	4	8.42E+06	99.803	100
2	41	1.03E+06	99.976	99
3	659	9.90E+04	99.998	65
4	83	9.87E+04	99.998	< 0
5	235	9.69E+04	99.998	< 0

Table 4-3: Diagnostic table for PARAFAC models used in 2:2 complex experiment for Naphthalene and β-cyclodextrin. The best fit model, the 3-component model, is shaded.

Looking at the sum of squared residuals, the percentage of variance explained by the model and the core consistency value, these diagnostics suggest a 3-component model. There is a significant drop in the SSR with increasing component number until the 4-component model. The reduction in the SSR is negligible in models with 4 or 5 components. The percentage of variance explained by the model stops increasing with models of 4 or 5 components. The core consistency value is relatively high for models with 1-3 components, and then drops below zero for models with 4 and 5 components. Therefore, after

evaluating the diagnostic table, a 3-component model was chosen for further evaluation.

The spectra from the 3-component PARAFAC model is shown in Figure 4-8. Spectra for the free naphthalene component and the 1:1 complexed naphthalene match the spectra calculated in the 1:1 complex experiment. Interestingly, the third component of this experiment attributed to the fluorescence due to the 2:2 complex appears to be mostly separate from the rest of the naphthalene spectra. While there is some noise level signal in the region naphthalene normally fluoresces at, most of the fluorescence is in the region with a maximum around 400 nm. Examining the spectra in the 4 component model, it appears the fourth component modeled is because of scattered light.

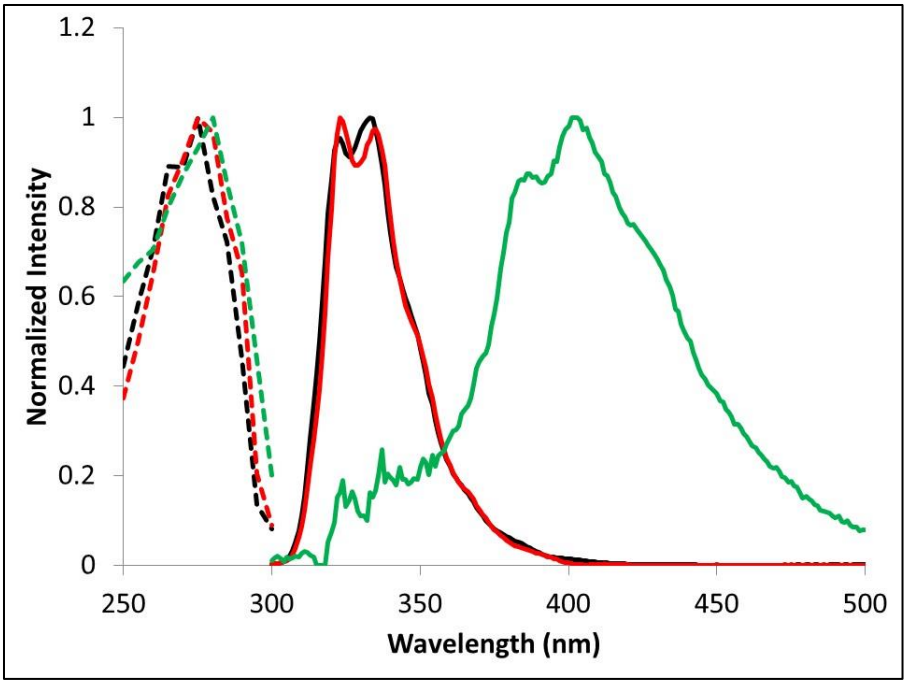


Figure 4-8: PARAFAC spectra for the 3-component model of the 2:2 complex experiment at 25 ºC. free naphthalene component (Excitation --- Emission —); 1:1 complexed naphthalene component (Excitation --- Emission —); and the 2:2 complex (Excitation --- Emission —).

There was one issue with the 3-component model. The original PARAFAC analysis calculated score values that were non-zero for components that should not be present in one sample. In the sample with no β-cyclodextrin present, there is no possibility for the 1:1 complex or 2:2 complex to form. These components had non-zero score values for this sample, however. Including a fourth component improved the score values, but did not completely bring the score values to zero. With the assistance of Alex Liu, we could make some small modifications to the PARAFAC program written by Bro.[27] Alex was able to edit the program to zero the score value for certain components in a sample, while still allowing the intensity no longer attributed to this part of the model to be used by other components. The models produced by this method were of comparable

quality when compared to using the original program. The score values calculated using Alex's program are the ones presented in Figure 4-9.The modified program will be used to calculate score values at 10 ^{o}C. Score values, using both the default program and Alex's modified program, for the experiment at 25 ^{o}C can be compared in Table 4-4 and Table 4-5.

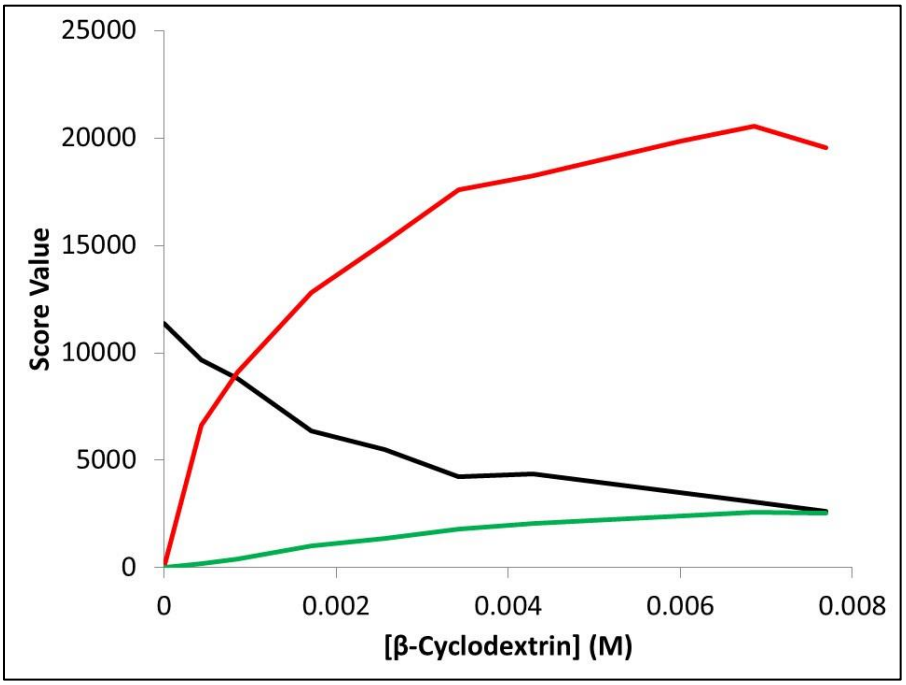


Figure 4-9: Score value plot for the 3-component model for the 2:2 complex system at 25 ^{o}C. Score values are calculated for the components that describe free naphthalene (—), 1:1 complexed naphthalene (—), and 2:2 complexed naphthalene (—). Calculated using Alex Liu's program

[B-Cyclodextrin] (M)	Score Value		
	Free	*Complexed*	*Dimer*
0.0077	3300	18866	2526
0.0069	3820	19788	2571
0.006	4240	19096	2400
0.0043	5106	17497	2072
0.0034	4984	16846	1816
0.0026	6253	14428	1349
0.0017	7105	12100	1007
0.0009	9564	8330	429
0.0004	10413	5878	186
0	12103	1534	0

Table 4-4: Score Values for 2:2 complex experiment at 25 ºC calculated using default PARAFAC program.

[B-Cyclodextrin] (M)	Score Value		
	Free	*Complexed*	*Dimer*
0.0077	2609	19552	2528
0.0069	3075	20528	2574
0.006	3490	19841	2403
0.0043	4350	18249	2074
0.0034	4252	17574	1818
0.0026	5515	15164	1351
0.0017	6383	12823	1009
0.0009	8805	9092	430
0.0004	9672	6623	186
0	11384	0	0

Table 4-5: Score Values for 2:2 complex experiment at 25 ºC calculated using program written by Alex Liu.

Finally, the score plot for the 3-component model, shown in Figure 4-9, was examined. As the concentration of β-cyclodextrin increases, the score value for each of the complexes increases. The score value of the 2:2 complex is rather low when compared to the other two components, but this is due to the low amount of fluorescence visible in the EEMs. The score values of the free naphthalene and the 1:1 complex naphthalene exhibit similar behavior to the 1:1

complex experiment. These score values will be used to determine the K_2 for the formation of the 2:2 complex.

The 3-component model for this system was analyzed further using jack-knife and split half analysis. The RIP plots for the excitation and emission loadings are shown in Figure 4-10. No obvious outliers appear in these plots. Examination of the IMP plots suggest no outliers as well. Split half spectra are shown for each component in Figure 4-11. For the free naphthalene component and the 1:1 complex component, the spectra match nearly identically. For the third component, which describes the 2:2 complex, there is a little deviation between the spectra of the two halves in the wavelength region from 325-350 nm. This is the region where naphthalene normally fluoresces in water.[23] Split half analysis was not able to quantitatively validate the model, for similar reasons as the 1:1 naphthalene complex experiment. The TCC calculated for each component resulted in the components being too alike to be considered valid.

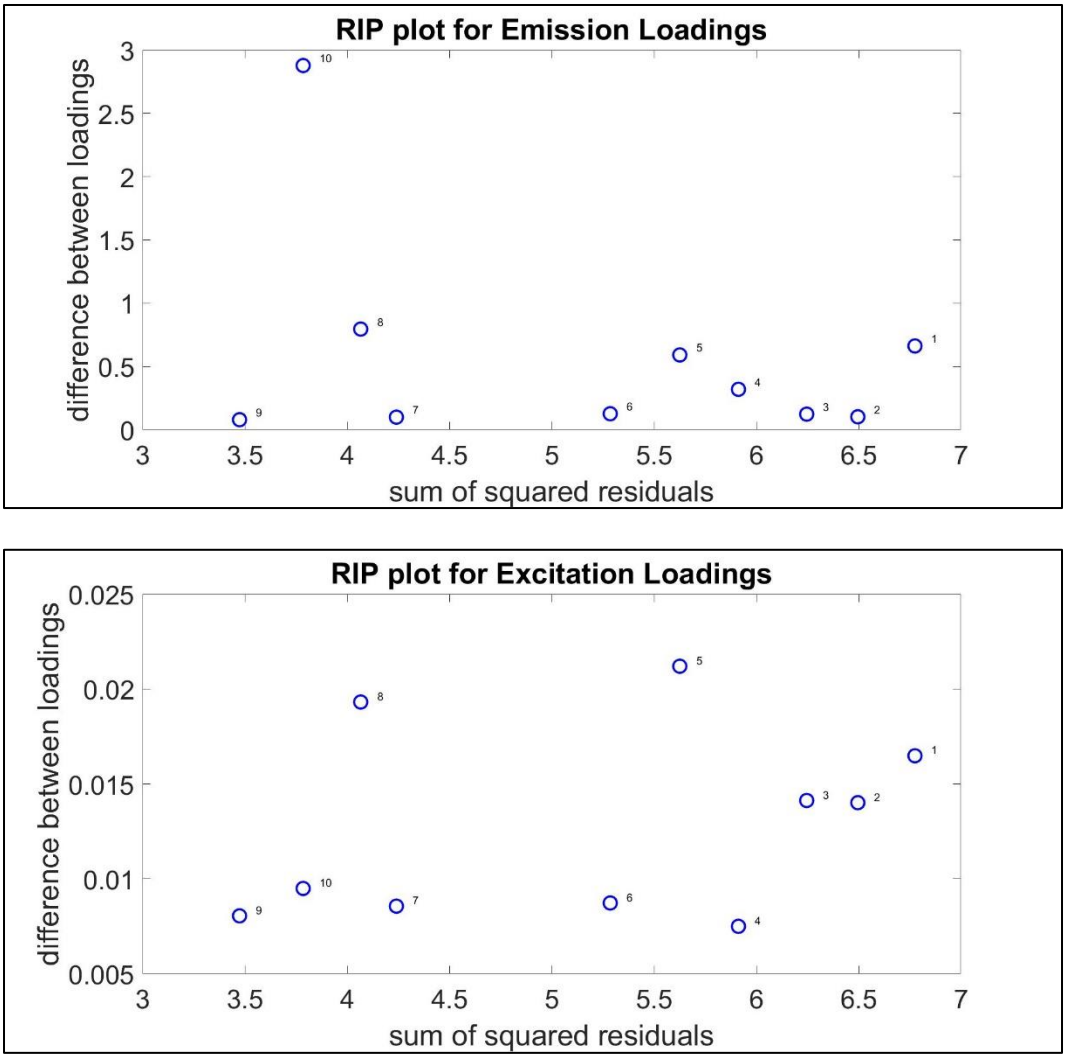


Figure 4-10: RIP plots for the excitation and emission loadings of the 3-component model used in the experiment investigating the 2:2 complex between naphthalene and β-cyclodextrin.

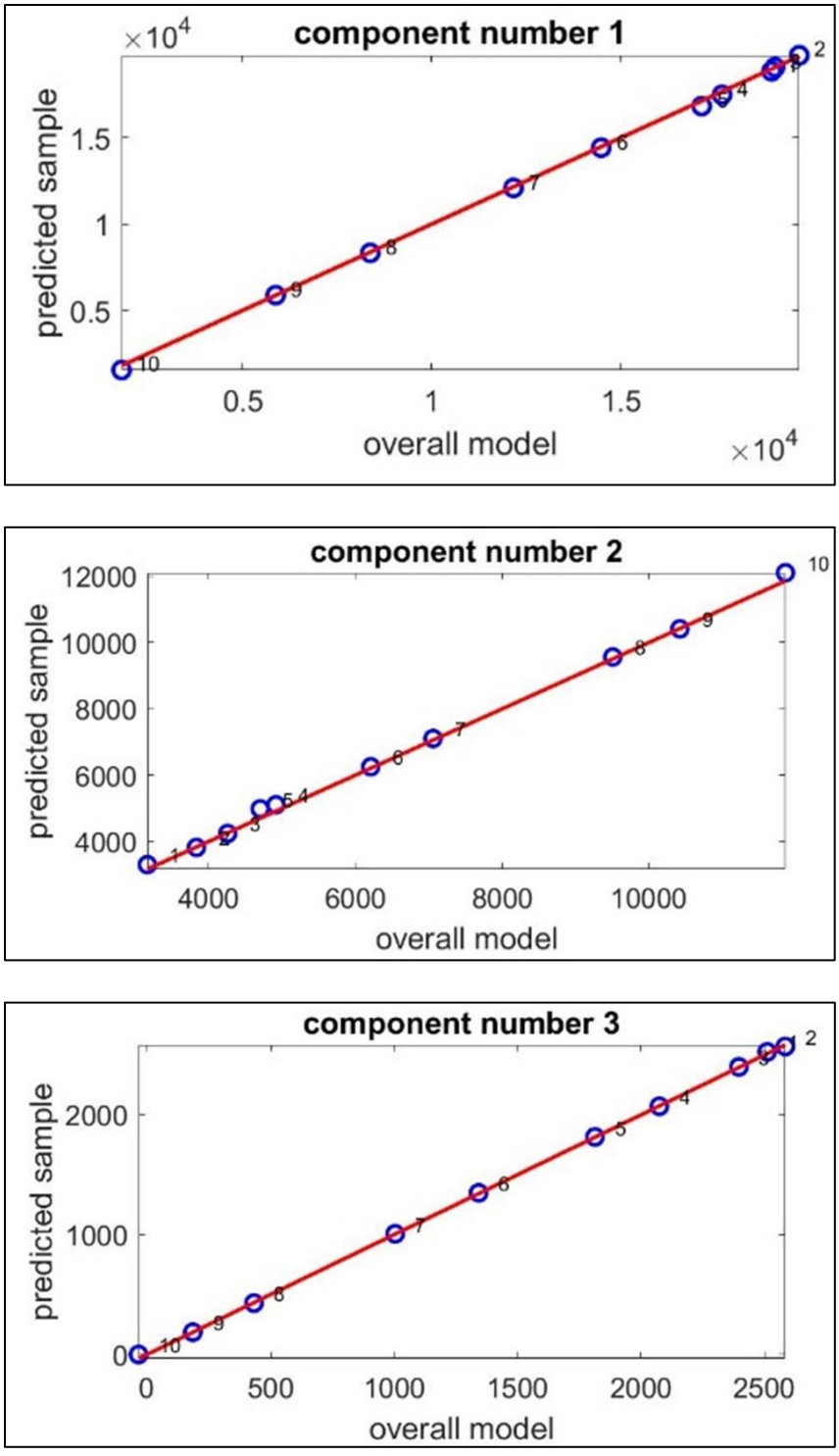


Figure 4-11: IMP plots for the three components in the PARAFAC model that describes the experiment for the 2:2 complex between naphthalene and β-cyclodextrin.

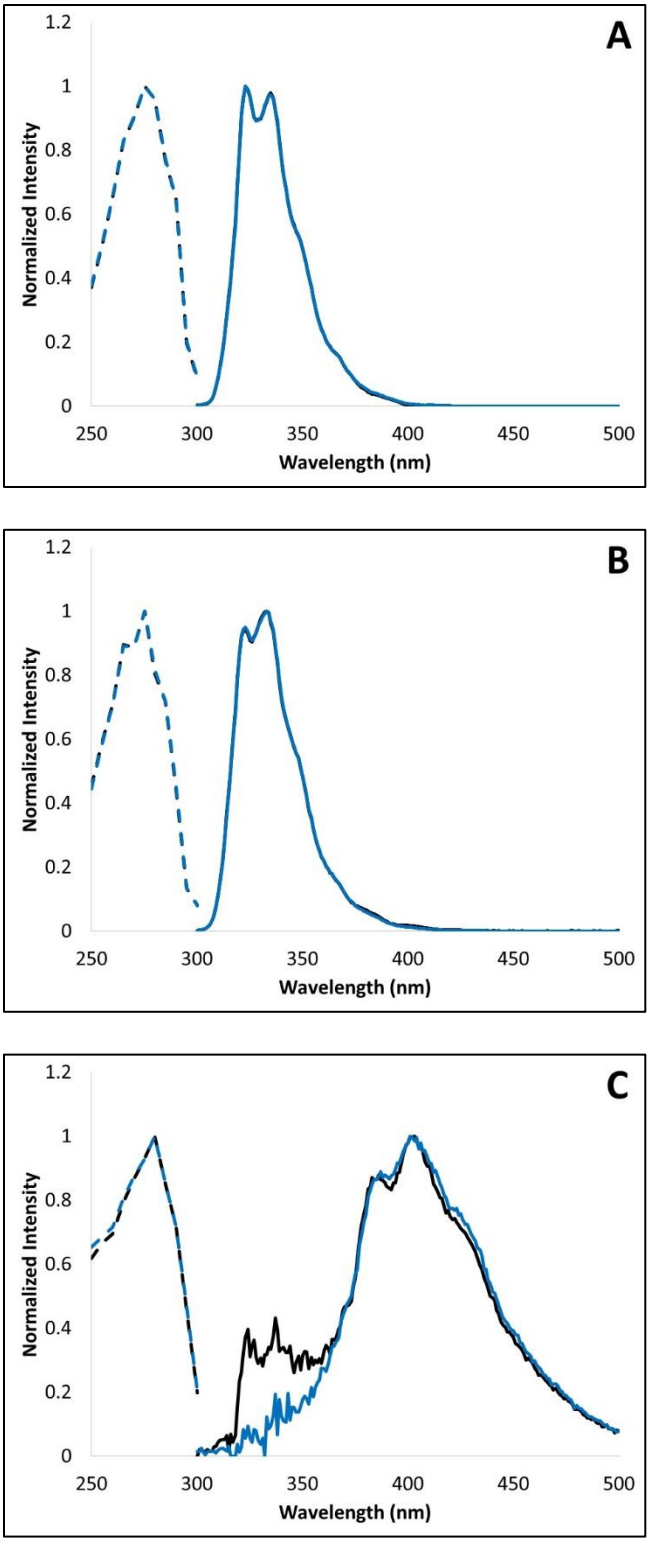


Figure 4-12: Split half spectra for the 3-component PARAFAC model in the 2:2 naphthalene complex experiment at 25 ºC. A: 1:1 complexed

naphthalene; B: free naphthalene; C: 2:2 complexed naphthalene (Half A: black; Half B: blue; Excitation: dashed; Emission: solid)

As stated previously, samples for this experiment were also measured at 10 ^{o}C. The fluorescence from the 2:2 complex has an increased intensity with temperature.[14] It should be noted that all models constructed for this experiment were used the PARAFAC algorithm in PLS Toolbox. This experiment encountered the same non-zero score value error in the experiment at 25 ^{o}C. Therefore, the modified program by Alex Liu was used to calculate score values for the final model.

The diagnostic table for the PARAFAC models constructed for this experiment can be found in Table 4-6. Examining the number of iterations used to converge on each model, there are large spikes in the number of iterations once 3 and 4 components are used in the PARAFAC model. The sum of squared residuals and the percentage of variance explained each see significant change until 4-components are used in the model. The core consistency value is high for a 1 or 2 component model, but drops once 3 components are used. The core consistency drops below zero for a 4 and 5 component model. The diagnostic table suggests either a 2 or 3 component model.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	4	4.22E+07	98.905	100
2	37	2.48E+06	99.936	100
3	535	5.99E+05	99.984	31
4	1689	5.85E+05	99.985	< 0
5	111	5.85E+05	99.985	< 0

Table 4-6: Diagnostic table for PARAFAC models used in 2:2 Complex experiment for Naphthalene and β-cyclodextrin. The best fit model, the 3-component model, is shaded.

The spectra for the 3-component PARAFAC model are shown in Figure 4-13. The first two components of the PARAFAC model, describing free naphthalene and 1:1 complexed naphthalene, are both very similar to the spectra determined in the 1:1 complex experiment and the 2:2 complex experiment at 25 ^{o}C. The spectra for the third component, which describes the 2:2 complex looks slightly different. The region with a maximum at 400 nm remains very much the same, but the region between 300 to 350 nm has more relative intensity when compared to the experiment at 25 ^{o}C.

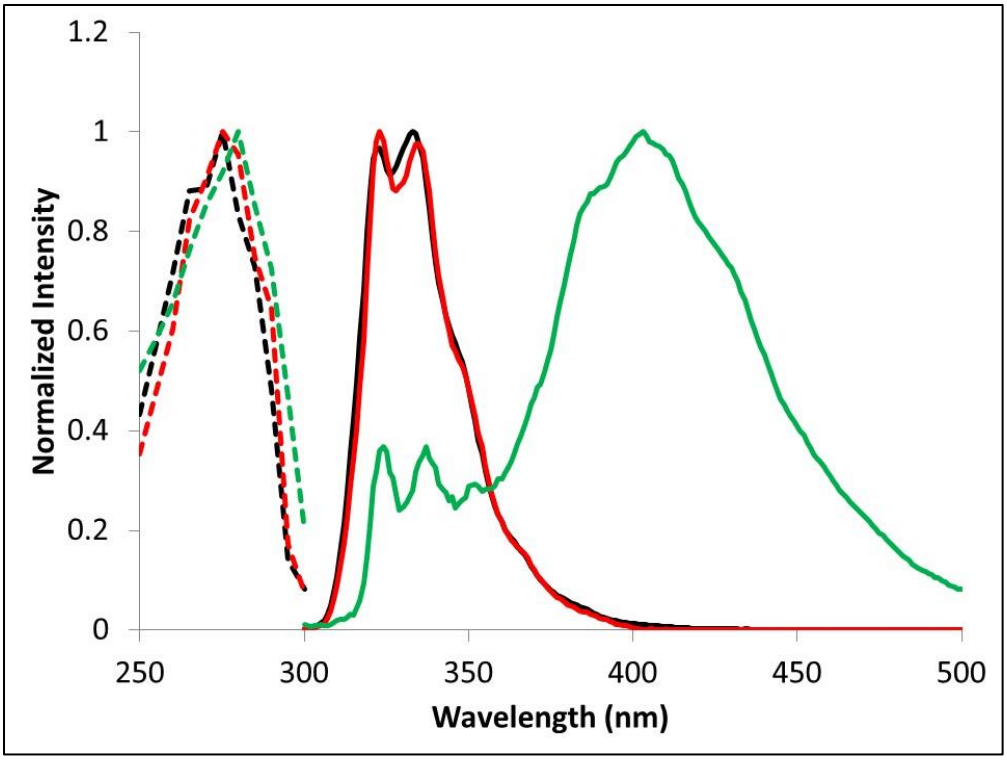


Figure 4-13: PARAFAC spectra for the 3-component model of the 2:2 complex experiment at 10 °C. free naphthalene component (Excitation --- Emission —); 1:1 complexed naphthalene component (Excitation --- Emission —); and the 2:2 complex (Excitation --- Emission —).

The score value plots, shown in Figure 4-14, are not as well behaved as in the experiment at 25 °C. This may be due to inconsistent cooling from sample to sample as the cuvettes were only cooled by a cuvette holder. Despite this error, the overall trend of the score values reflects expected behavior. The score values for the free naphthalene and 1:1 complexed naphthalene still exhibit the previously observed trend from other experiments. It is also interesting that the score value trend for these two components mirror each other almost exactly. As for the third component, which describes the 2:2 complex, the increased intensity

observed in the excimer fluorescence at 10 ^{o}C corresponds to an increase in the score values for the third component.

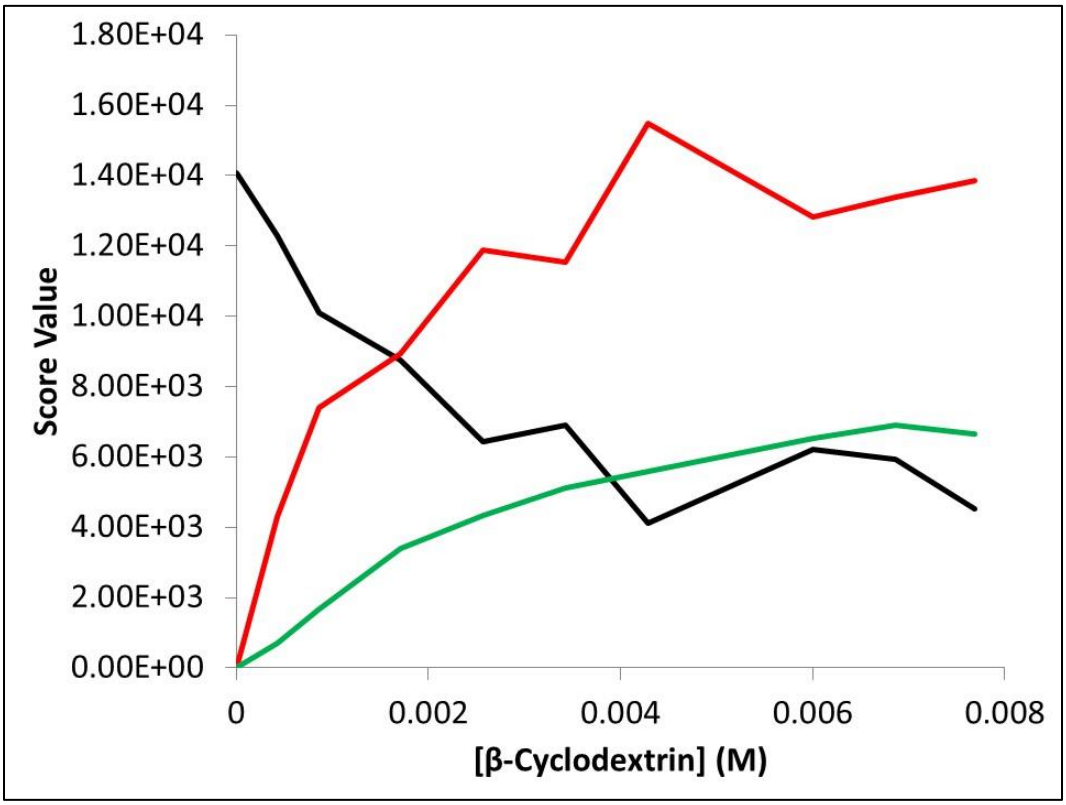


Figure 4-14: Score value plot for the 3-component model for the 2:2 complex system at 10 ^{o}C. Score values are calculated for the components that describe free naphthalene (—), 1:1 complexed naphthalene (—), and 2:2 complexed naphthalene (—).

[B-Cyclodextrin] (M)	Score Value		
	Free	***Complexed***	***Dimer***
0.0077	2802	16491	6413
0.0069	4177	16091	6655
0.006	4490	15446	6276
0.0034	5459	13705	4919
0.0026	5093	13819	4164
0.0017	7572	10579	3270
0.0009	9205	8550	1601
0.0004	11641	5052	662
0	13942	133	0

Table 4-7: Score Values for 2:2 complex experiment at 10 ºC calculated using default PARAFAC program.

[B-Cyclodextrin] (M)	Score Value		
	Free	***Complexed***	***Dimer***
0.0077	1961	17321	6416
0.0069	3260	16997	6658
0.006	3578	16348	6280
0.0034	4551	14606	4922
0.0026	4203	14702	4167
0.0017	6647	11501	3273
0.0009	8248	9506	1603
0.0004	10663	6034	663
0	13007	0	0

Table 4-8: Score Values for 2:2 complex experiment at 10 ºC calculated using program written by Alex Liu.

The 3 component model was evaluated for outliers using jackknife and subjected to split half analysis as well. Samples 4 and 10 appear to be outliers according to the RIP plots shown in Figure 4-15. They are kept in the data set for better comparison to the experiment at 25 ºC. Split half analysis could not be used to validate the model quantitatively, but the spectra are shown in Figure 4-17.

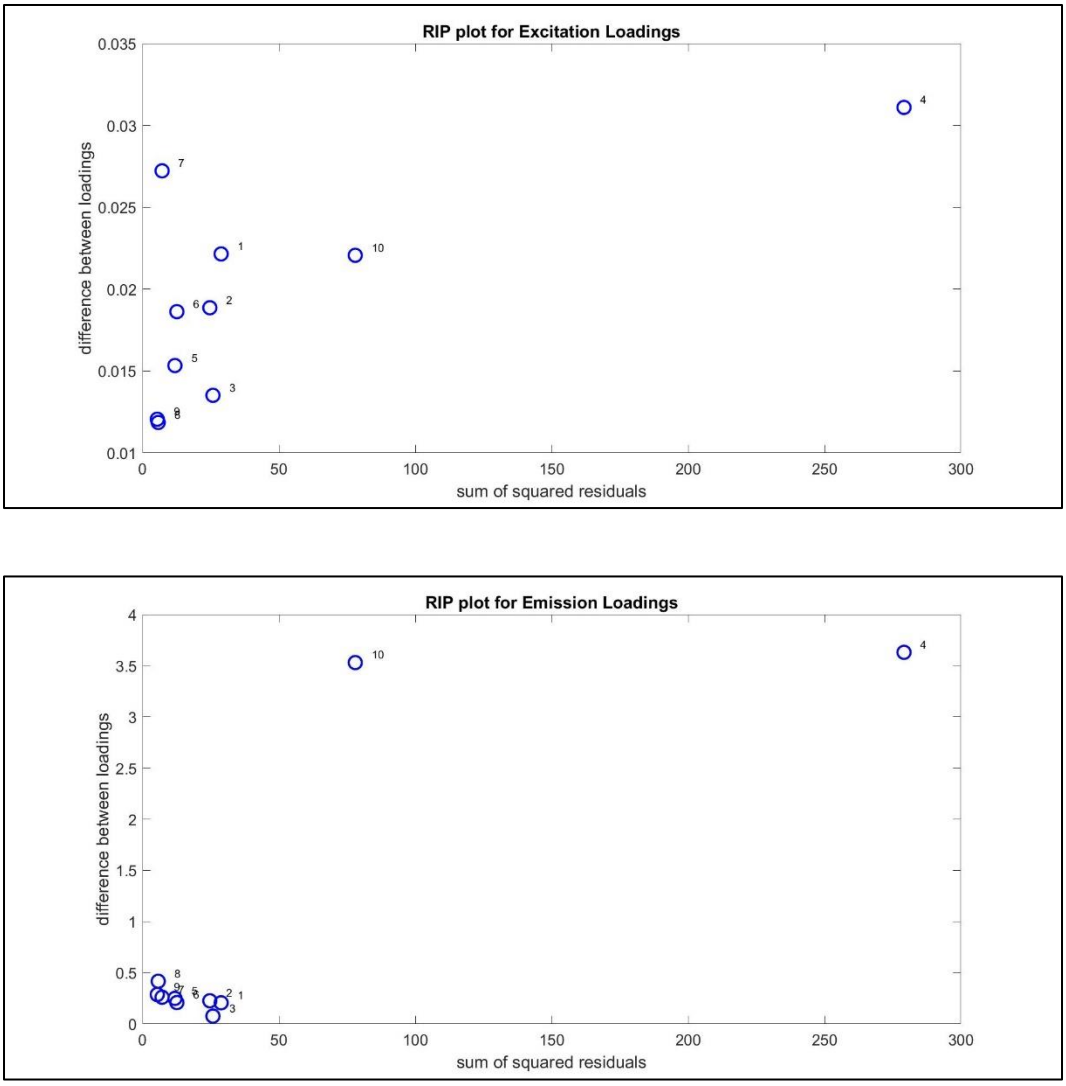


Figure 4-15: RIP plots for the excitation and emission loadings of the 3-component model used in the experiment investigating the 2:2 complex between naphthalene and β-cyclodextrin at 10 ºC.

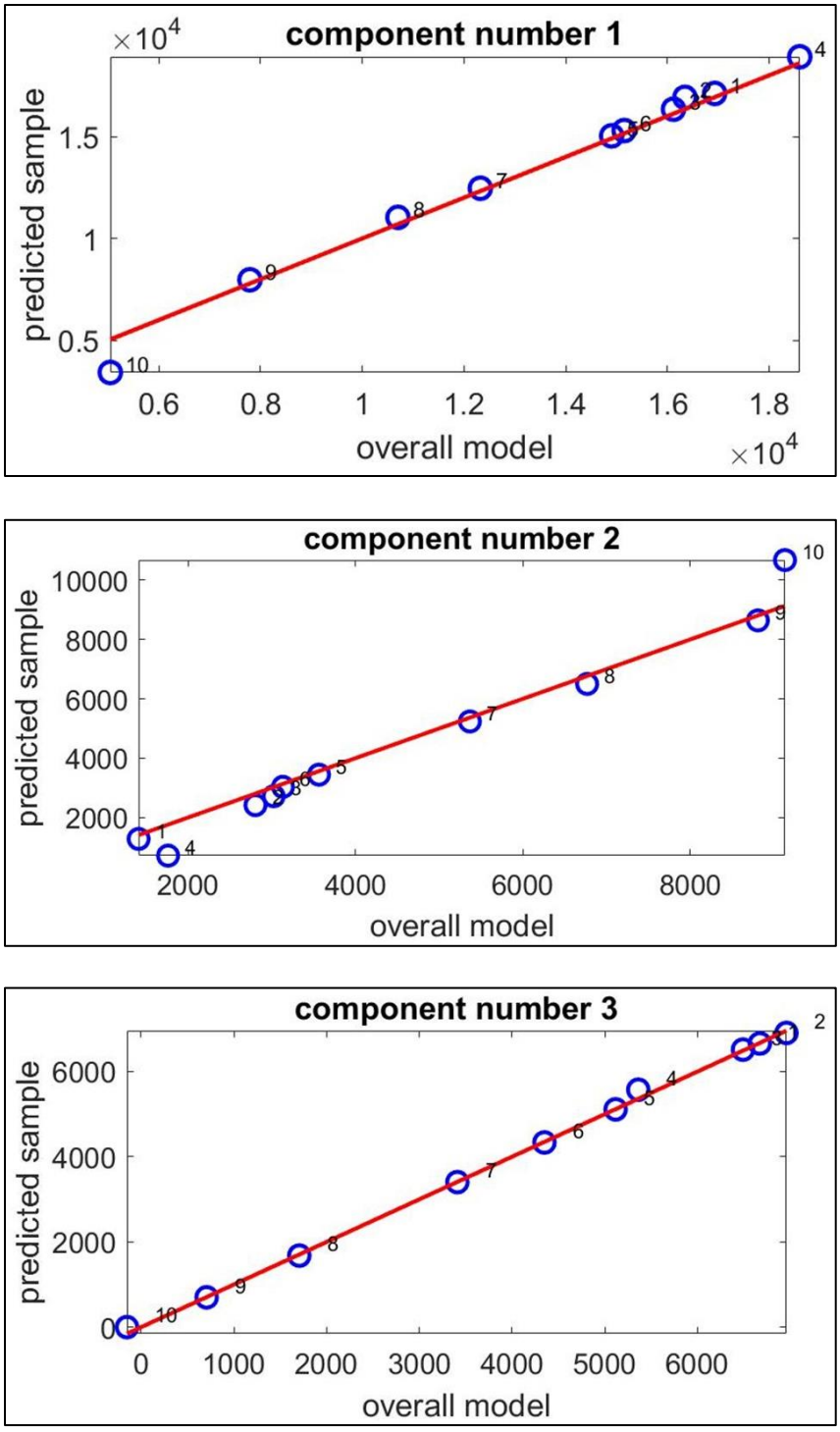


Figure 4-16: IMP plots for the three components in the PARAFAC model that describes the experiment for the 2:2 complex between naphthalene and β-cyclodextrin at 10 ºC.

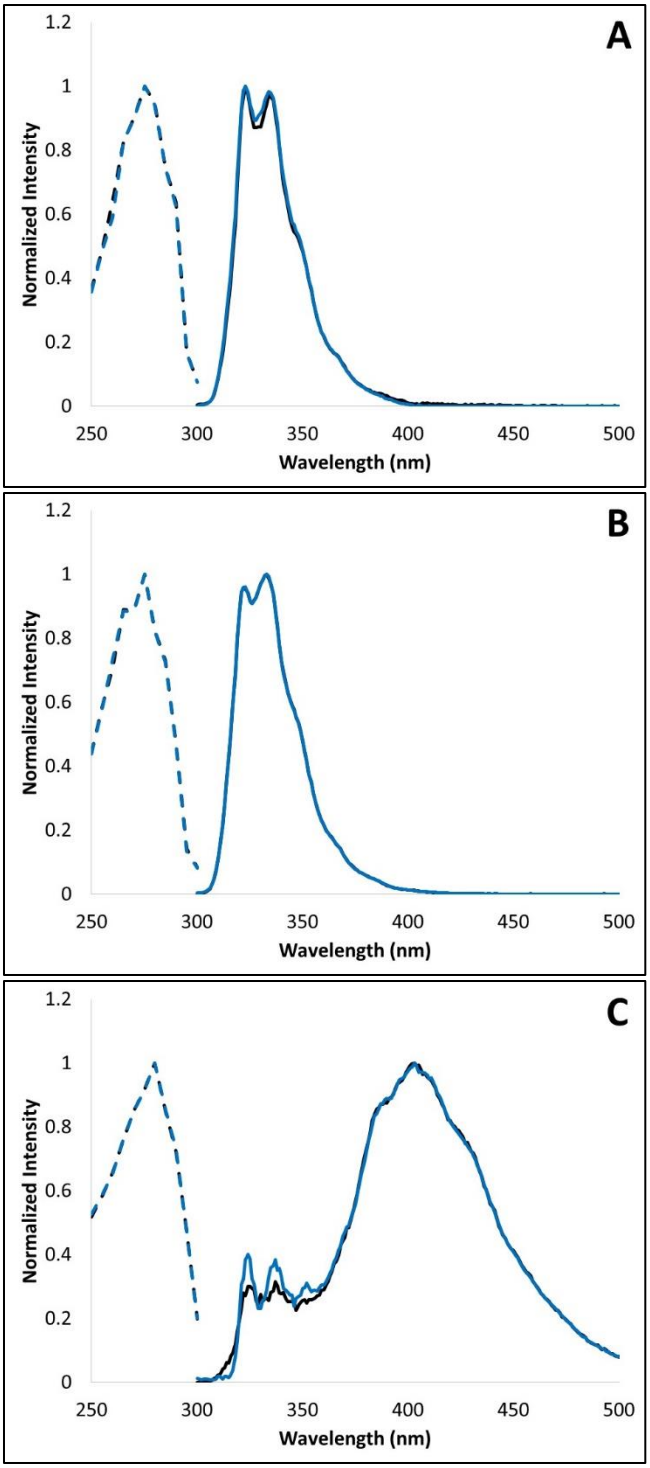


Figure 4-17 Split half spectra for the 3-component PARAFAC model in the 2:2 naphthalene complex experiment at 25 ºC. A: 1:1 complexed naphthalene; B: free naphthalene; C: 2:2 complexed naphthalene (Half A: black; Half B: blue; Excitation: dashed; Emission: solid)

The PARAFAC model chosen to describe the 2:2 experiment is the 3-component model. This model matches the number of fluorophores we hypothesize to be in the system, assuming no contaminants or unexpected complexes. While there was a slight difference between the spectra for the 2:2 complexed naphthalene component, they provide further insight into the complex. The score values from this model will be used to calculate the formation constant of the 2:2 complex.

4.3.3 1:1 Anthracene Complex Experiment

The diagnostic table shown in Table 4-9 is for the experiment involving the 1:1 complex between anthracene and β-cyclodextrin. The diagnostics suggest a 3-component model. The number of iterations show a drastic increase once a fourth component is added to the model. The sum of squared residuals shows insignificant decrease once a fourth component is added. The percentage of variance explained by the model does not increase significantly after three components have been added to the model. The core consistency for the 3-component model is very high with a drop once a fourth component is added. This table results in a 3-component model being very likely.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	26	2.12E+06	99.299	100
2	385	2.44E+05	99.919	100
3	673	2.94E+04	99.990	93
4	5357	2.40E+04	99.992	25
5	10000	2.29E+04	99.992	< 0

Table 4-9: Diagnostic table for PARAFAC models used in 1:1 Complex experiment for anthracene and β-cyclodextrin. The best fit model, the 2-component model, is shaded.

The evidence in the diagnostic table does not match the original hypothesis for this experiment. The 1:1 complex system between anthracene and β-cyclodextrin should result in a 2-component model. The components in the PARAFAC model should describe the fluorescence intensity due to the free anthracene in solution and the fluorescence intensity due to the 1:1 complexed anthracene. A third component was very much unexpected.

Examination of the score values and spectrum for the third component lead us to believe that it is possibly a contaminant. The score values for the third component are very low, and present in each sample. This would coincide with the working hypothesis that it is a contaminant in the experiment. One benefit of PARAFAC analysis is that it will determine the fluorophores in the dataset without the need for standards. This allows for the discovery of contaminants, or other unexpected fluorophores. It is unlikely that the third component is an unexpected complex between anthracene and β-cyclodextrin. The spectrum is vastly different from either of the other spectra of anthracene and is severely blue shifted. This does not match the expected shifts of any complex we have looked at. The spectra of the free anthracene component match the expected spectra of anthracene.[23] The spectrum of the 1:1 complexed anthracene appears to be redshifted, which is consistent with the other spectra we have observed so far in this experiment.

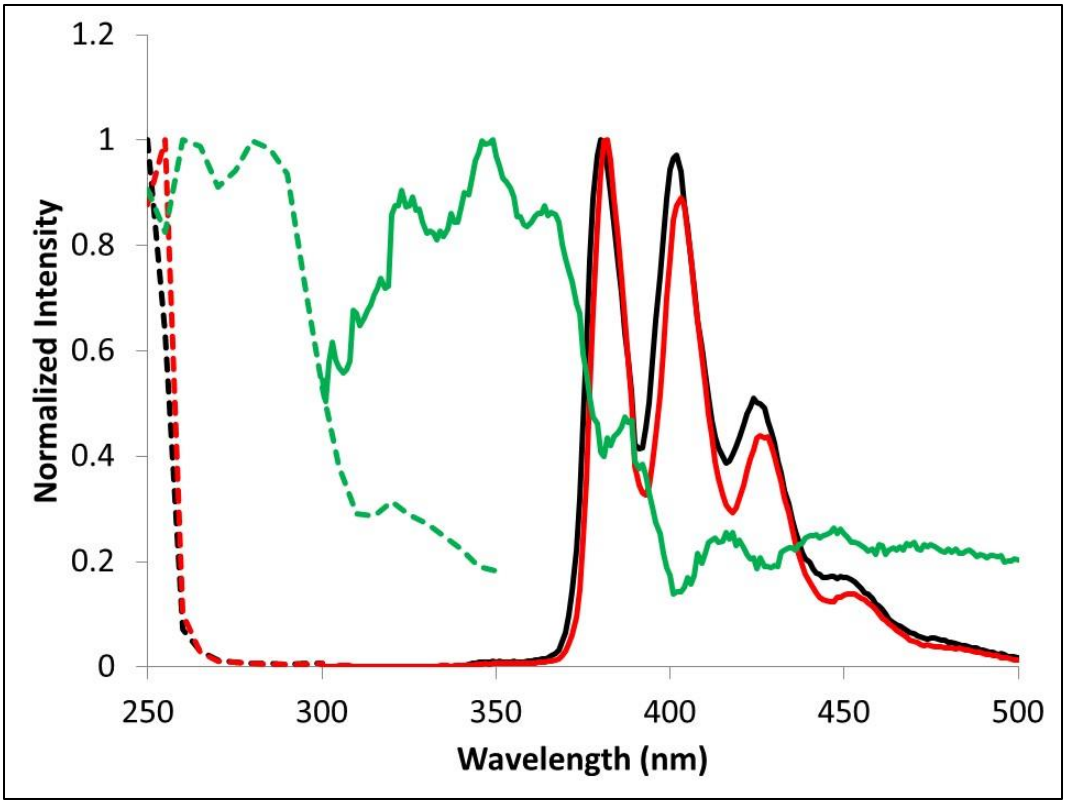


Figure 4-18: PARAFAC spectra for the 3-component model of the 1:1 complex experiment for anthracene and β-cyclodextrin. free anthracene component (Excitation --- Emission —); 1:1 complexed anthracene component (Excitation --- Emission —); and unknown component (Excitation --- Emission —).

Despite the presence of the unknown component, the score values for the free anthracene and 1:1 complexed anthracene still behave as expected. The score values for all three components are shown in Figure 4-19 and the score values calculated using Alex Liu's modified program. The score values for the unknown third component are very low when compared to the other two components. There is a slight increase in the score values for the third component as the concentration of β-cyclodextrin increases. This may be indicative of the fluorophore interacting with the β-cyclodextrin molecules. If this is the case, it may influence the formation constant calculated for the 1:1 complex in this system.

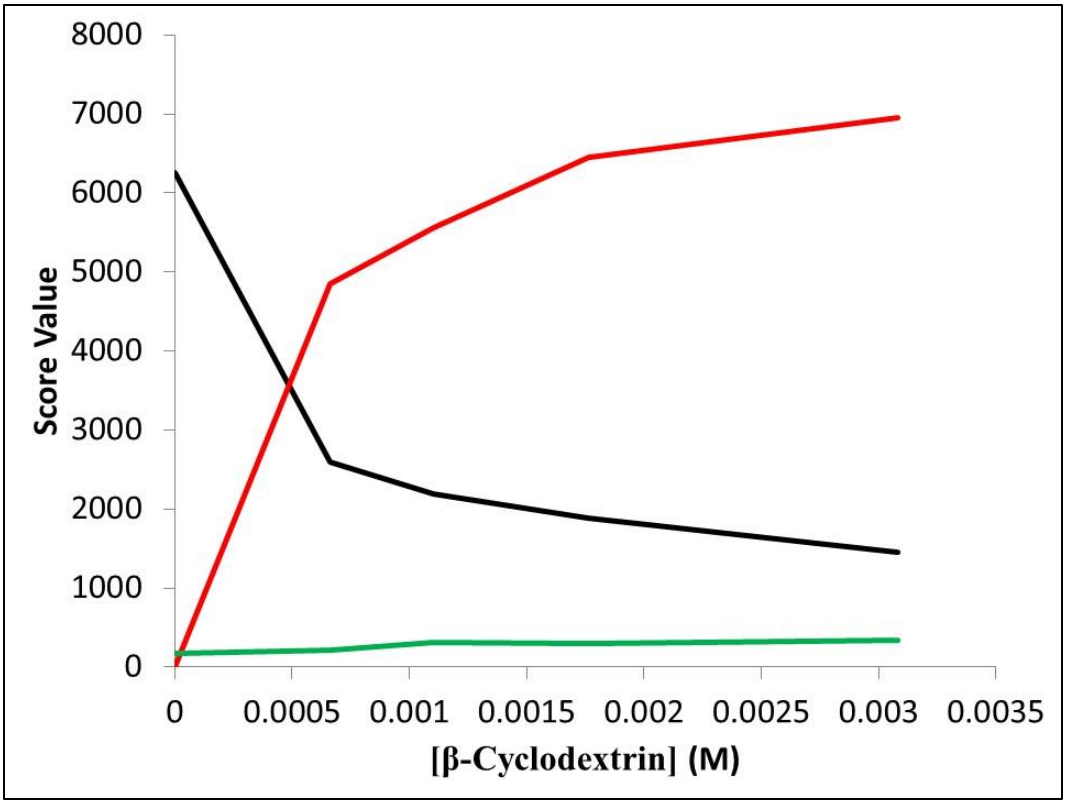


Figure 4-19: Score value plot for the 3-component model for the 1:1 anthracene complex system. Score values are calculated for the components that describe free anthracene (—), 1:1 complexed anthracene (—) and unknown component (—). Calculated using Alex Liu's program

[B-Cyclodextrin] (M)	Score Value		
	Comp Anth.	Free Anth.	Contaminant
0.0030	6950	1454	335
0.0017	6449	1890	297
0.0011	5554	2191	305
0.0006	4848	2601	209
0	712	6258	179

Table 4-10: Score values for 1:1 anthracene experiment using default program

[B-Cyclodextrin] (M)	Score Value		
	Free. Anth.	Comp. Anth.	Contaminant
0.0030	1473	6981	335
0.0017	1907	6477	297
0.0011	2209	5582	305
0.0006	2613	4867	209
0	6275	0	179

Table 4-11: Score values for 1:1 anthracene experiment using Alex Liu's modified program.

Due to the small number of samples in this data set, analysis using jackknife or split half are not useful. This dataset was originally prepared with 11 samples, but there were issues with the solubility of β-cyclodextrin in some of the more concentrated samples. This was indicated by the formation constant plots when calculating K_1 for this complex. To minimize this effect, samples were used that had a concentration of β-cyclodextrin of about 3 mM. This is a concentration that has been found to have minimal aggregation or solubility problems.[17]

4.3.4 Multiple Complex Experiment

There were 5 samples used in this dataset as well, the samples were made from samples originally used in the anthracene 1:1 complex experiment. The diagnostic table for this dataset is shown in Table 4-12. As stated previously, some of the usual PARAFAC settings needed to be changed to fit a model that best explained the dataset. Some of these changes result in some of the diagnostics not being as reliable in suggesting the best fit model for the dataset. The number of iterations increases drastically from a 3-component model to a 4-component model. The sum of squared residuals and the percentage of variance explained both have significant changes as each component is added to the model,

up until when four components are used. The core consistency value is only high for a model with 1 or 2 components.

Number of Components	Number of Iterations	Sum of Squared Residuals	Percent of Variance Explained	Core Consistency Value
1	9	2.19E+08	74.940	100
2	23	2.59E+06	99.711	100
3	921	2.25E+06	99.749	< 0
4	10993	8.12E+04	99.991	< 0
5	8030	2.56E+05	99.971	< 0
6	20000	5.45E+04	99.994	< 0

Table 4-12: Diagnostic table for PARAFAC models used in the multicomplex experiment. The best fit model, the 4-component model, is shaded.

The number of iterations and the core consistency diagnostic do not suggest anything beyond a 3-component model. Evaluation of the spectra or these models, however, do not suggest these models are the best fit either. Focusing on just the SSR and the percentage of variance explained, both diagnostics strongly suggest a 4-component model. This is the hypothesized model in this system, as there should be four different types of fluorophores: free naphthalene, free anthracene, 1:1 complexed naphthalene, and 1:1 complexed anthracene.

The spectra for the 4-component model are shown in Figure 4-20. All the spectra calculated by the PARAFAC model match the spectra from previous experiments very well. All the expected components are accounted for and there are no unexpected components. Using a 5-component model would result in some scattered light to be modeled. Interestingly, the unknown component from the 1:1 anthracene experiment would show up in a 6-component model. The samples in this data set were made from the samples used in that experiment, so if the

unknown component is truly a chemical contaminant, it should be present in the PARAFAC model.

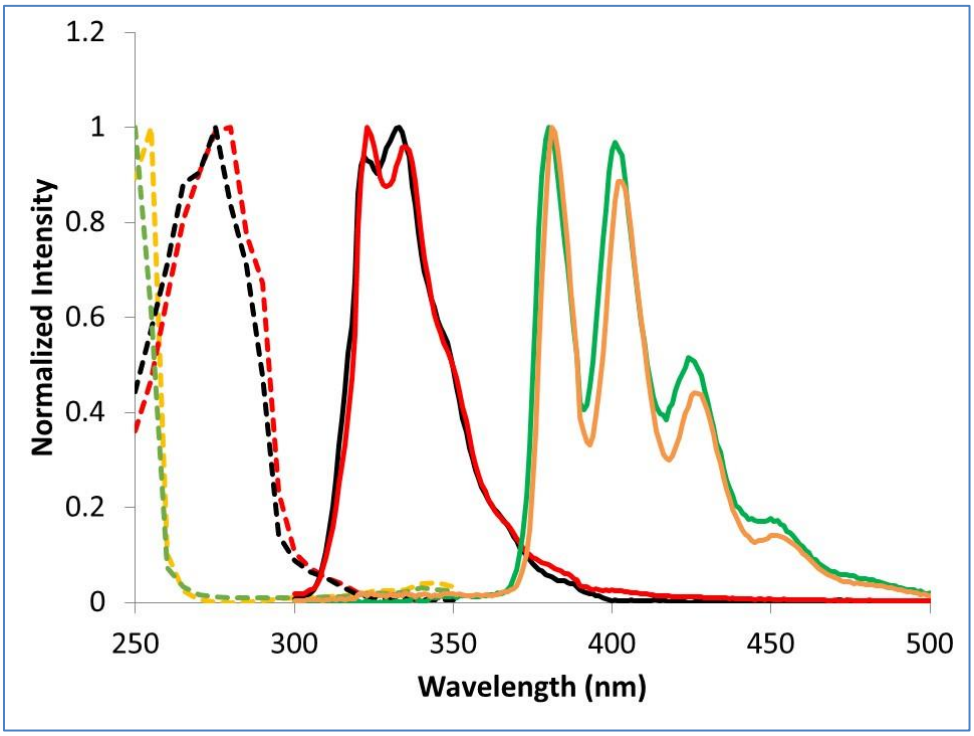


Figure 4-20: PARAFAC spectra for the 4-component model of the simultaneous 1:1 complex experiment. free naphthalene component (Excitation --- Emission —); 1:1 complexed naphthalene component (Excitation --- Emission —); free anthracene component (Excitation --- Emission —); and 1:1 complexed anthracene (Excitation --- Emission —).

The score values for the components in the 4-component model are shown in Figure 4-21. The score values for each component exhibit the same trends that appeared in previous experiments. The initial PARAFAC modeling calculated non-zero score values for the complexed components when no β-cyclodextrin was present. This resulted in the modified program written by Alex Liu to be used again. The score values calculated using this program are the score values presented in Figure 4-21.

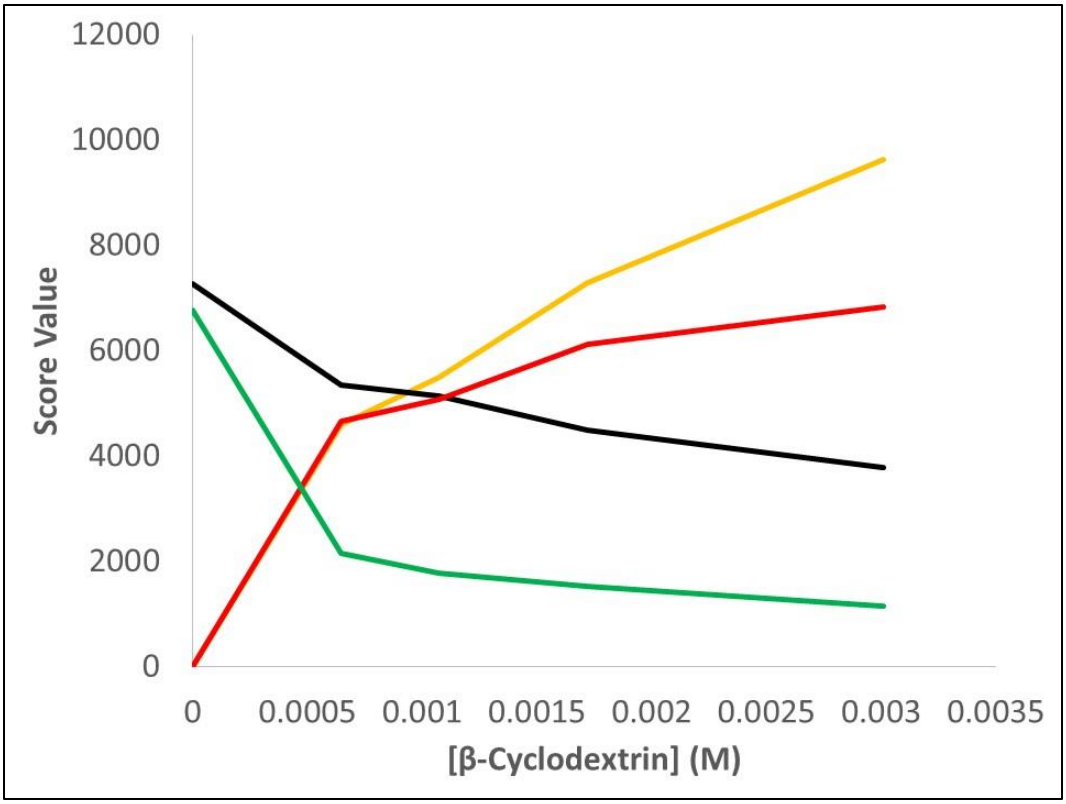


Figure 4-21: Score value plot for the 4-component model for the simultaneous complex system. Score values are calculated for the components that describe free naphthalene (—), 1:1 complexed naphthalene (—), free anthracene (—) and 1:1 complexed anthracene (—). Calculated using Alex Liu's program

[B-Cyclodextrin] (M)	Score Value			
	Comp. Anth.	Comp. Naph.	Free Naph.	Free Anth.
0.0030	10028	6834	3393	1158
0.0017	7643	6122	4146	1528
0.0011	5826	5085	4827	1779
0.0006	4912	4657	5044	2155
0	1858	1129	6991	6773

Table 4-13: Score values for multi-determination experiment using default program.

[B-Cyclodextrin] (M)	Score Value			
	Comp. Anth.	Free Naph	Comp. Naph.	Free Anth.
0.0030	9632	3790	6834	1159
0.0017	7290	4500	6122	1529
0.0011	5502	5151	5085	1780
0.0006	4606	5350	4657	2155
0	0	7272	0	6773

Table 4-14: Score values for multi-determination experiment using Alex Liu's modified program.

The small number of samples make the use of further analysis, such as jack-knife and split half analysis, not useful. In this case the model is being chosen as an appropriate fit based on the other models previously chosen in other experiments. While some of the diagnostics agree with this choice, this is a rare case when the more objective information we usually use to decide the best fit model is not useful. The information from the 4-component model in this experiment will be used to calculate formation constants for each of the 1:1 complexes in this system.

4.4 Formation Constant Determination

To determine the formation constants for each complex, we will have to derive several expressions to use the PARAFAC score value in our calculation of these constants. All the results for the formation constant determination are summarized in

4.4.1 1:1 Complex of Naphthalene and β-cyclodextrin

First, the expression for the equilibrium constant for the 1:1 complex between naphthalene and β-cyclodextrin is:

$$K_1 = \frac{[NC]}{[N][C]}$$

(4-3)

where K_1 is the equilibrium constant for reaction (4-1), $[NC]$ is the concentration of the 1:1 complex at equilibrium, $[N]$ is the concentration of free naphthalene at equilibrium, and $[C]$ concentration of free β-cyclodextrin at equilibrium. For this experiment, it is assumed there are only 3 possible species in each sample (naphthalene, β-cyclodextrin, and the 1:1 complex). Only two of these species fluoresce, free naphthalene and complexed naphthalene.

To use expression (4-3) to determine K_1, an expression to relate the score values of the PARAFAC components to their concentration must be determined. As stated previously in Chapter 1, the fluorescence intensity of a molecule can be attributed to the following factors,

$$\boldsymbol{c_{kr} = 2.303\phi_r \times [s]_{kr} \times \varepsilon_{ir} \times em_{jr} \times f}$$

(4-4)

where ϕ_r the fluorescence quantum yield of component r, $[s]_{kr}$ is the concentration of component r (corresponding to some species *s*) in sample k, ε_{ir} is the molar absorptivity constant at excitation wavelength *i* for component *r*, em_{jr} is the relative fluorescence intensity at emission wavelength *j* for component *r*, and *f* is a coefficient comprising of instrumental factors.[20,28]

Since the only concern is concentration at this moment, the above expression can be simplified by summarizing the other factors in a single constant.

$$\alpha_r c_{kr} = [s]_{kr}$$

(4-5)

We will define α_r as the *sensitivity factor* for component *r*. It must be noted that α_r is not an intrinsic characteristic of component *r*. Since it contains an instrumental factor *f*, as seen in expression (4-4), it is only valid for fluorophores measured at the same instrumental settings. We can use expression (4-5) to relate score values for PARAFAC components to actual concentrations.

Using expression (4-5) to incorporate the score values for free and complexed naphthalene components into expression (4-3) the following expression is obtained:

$$K_1 = \frac{\alpha_{NC} \mathrm{c}_{NCk}}{\alpha_N \mathrm{c}_{Nk}[C]}$$

(4-6)

Where α_N is the sensitivity factor for free naphthalene, α_{NC} is the sensitivity factor for complexed naphthalene, c_{NCk} is the PARAFAC score value for complexed naphthalene in sample *k*, and c_{Nk} is the PARAFAC score value for free naphthalene in sample *k*. In the experiments involved in this work, the concentration of β-cyclodextrin will be much larger than the other species in the system. Therefore, we will consider it to remain constant, a similar assumption made by Hamai.[7] For expression (4-6) to be useful in determining K_1, the sensitivity factors for each component need to be determined.

To solve for the sensitivity factors, a different expression is required. The following expression is true for a 1:1 complex between naphthalene and β-cyclodextrin at equilibrium:

$$[N]_{tot} = [N]_{free} + [N]_{complexed} \quad \textbf{(4-7)}$$

Where $[N]_{tot}$ is the total concentration of naphthalene in the sample, $[N]_{free}$ is the concentration of free naphthalene in the sample and $[N]_{complexed}$ is the concentration of complexed naphthalene. Substituting in score values in for the concentrations of each species and applying the expression to a data set of *k* samples, we get the following:

$$[N]_{totk} = \alpha_N \mathrm{c}_{Nk} + \alpha_{NC} \mathrm{c}_{NCk} \quad \textbf{(4-8)}$$

where $[N]_{totk}$ is the total concentration of naphthalene in sample *k*, α_N is the sensitivity factor for free naphthalene, c_{Nk} is the score value of the PARAFAC component that describes the free naphthalene in sample *k*, α_{NC} is the sensitivity factor for complexed naphthalene and c_{NCk} is the score value of the PARAFAC component that describes complexed naphthalene in sample *k*. If we assume a constant total concentration of naphthalene for each sample in the data set, the expression can be rearranged to be more useful in the form.

$$\mathrm{c}_{Nk} = -\left(\frac{\alpha_{NC}}{\alpha_N}\right)\mathrm{c}_{NCk} + \left(\frac{[N]_{totk}}{\alpha_N}\right) \quad \textbf{(4-9)}$$

The above expression can be used in a plot of the score value of free naphthalene against the score value of complexed naphthalene. It will provide both a slope and y-intercept that can be used to calculate information regarding the sensitivity factors for each component. The slope of this plot will calculate a sensitivity factor ratio $\left(\frac{\alpha_{NC}}{\alpha_N}\right)$. The plot resulting from the 1:1 experiment using this expression is shown in Figure 4-22.

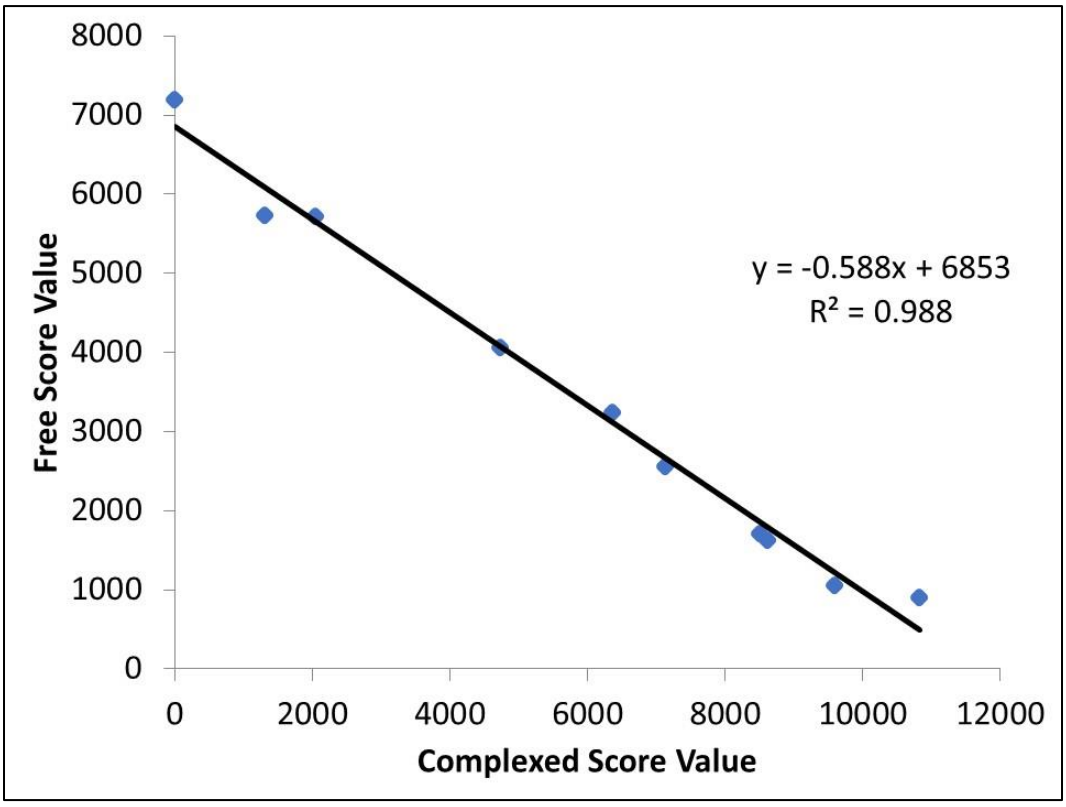


Figure 4-22: Plotting the free naphthalene score value against the complexed naphthalene score value can be used to calculate the sensitivity factor ratio.

The plot calculates a sensitivity factor ratio of 0.588 from the slope of a linear regression of the data. This ratio can now be used to help calculate K_1 in expression (4-10).

$$\left(\frac{\alpha_{NC}}{\alpha_N}\right)\left(\frac{\mathrm{c}_{NCk}}{\mathrm{c}_{Nk}}\right) = K_1[C]$$

(4-10)

Plotting a ratio of the score values, multiplied by a sensitivity factor ratio, against the concentration of β-cyclodextrin should allow for a linear regression that calculates a slope equal to K_1. The left side of expression (4-10) will be referred to as the weighted score ratio. Plotting the data from the 2-component model from the 1:1 complex experiment calculates a formation constant of 690 ±10. This value matches other values in the literature very well.[7,9,10,14] In other trials, K_1 would range from about 200 to 700. This is most likely due to varying β-cyclodextrin quality and solubility. This range of values is also similar in the literature,[9,10] so it is acceptable.

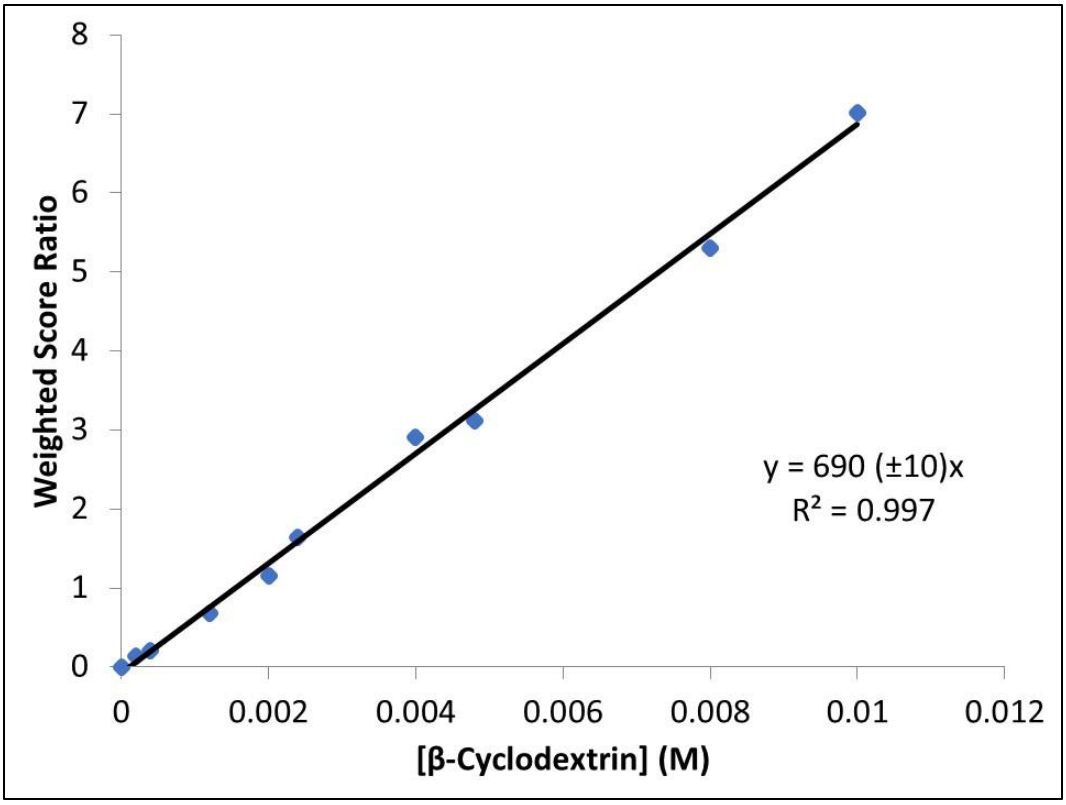


Figure 4-23: Plot to determine the formation constant K_1 for the 1:1 complex between naphthalene and β-cyclodextrin. The slope indicates a formation constant of 694 for this complex.

4.4.2 1:1 Complex for Anthracene and β-Cyclodextrin

The determination of the formation constant for the 1:1 anthracene complex was determined using the same method as the 1:1 complex between naphthalene and β-cyclodextrin. Similar expressions and plots were used in determining K_1 for the 1:1 complex between anthracene and β-cyclodextrin. Using the score values from the 2-component PARAFAC model, a formation constant of 1020 ± 140 was calculated for the 1:1 Anthracene-β-cyclodextrin complex. This is at least similar to the formation constant found in the literature of about 2000.[10] A sensitivity factor ratio of 0.693 was calculated during the determination.

4.4.3 2:2 Complex for Naphthalene and β-Cyclodextrin

While similar approach to the other complexes will be used in determining the formation constant for the 2:2 complex, it is slightly more complicated. There are more species to deal with in this system. The equilibrium constant expression for the system in reaction (4-2) is given by:

$$K_2 = \frac{[(NC)_2]}{[NC]^2}$$

(4-11)

In this system, three fluorescent species are expected to be present in each sample: free naphthalene, 1:1 complexed naphthalene, and 2:2 complexed naphthalene. Each species was described by a separate PARAFAC component. Using the PARAFAC score values for each component, K_2 can be determined using the appropriate substitutions in either expression (4-11) .

$$K_2 = \frac{\alpha_D c_{Dk}}{(\alpha_{NC} c_{NCk})^2}$$

(4-12)

This expression is used for determining K_2, sensitivity factors must be determined for each component. Similarly, to the approach in the 1:1 complex, each species can be related to the total concentration of naphthalene:

$$[N]_{tot} = [N]_N + [N]_{NC} + 2[N]_D$$

(4-13)

Where $[N]_N$, $[N]_{NC}$ and $[N]_D$ are the concentrations of free naphthalene, 1:1 complexed naphthalene and 2:2 complexed naphthalene (dimer), respectively. Substituting in score values for each component appropriately:

$$[N]_{tot} = \alpha_N c_{Nk} + \alpha_{NC} c_{NCk} + 2\alpha_D c_{Dk}$$

(4-14)

Where $\alpha_{N,}$ α_{NC} and α_D are the sensitivity factors for free naphthalene, 1:1 complexed naphthalene and 2:2 complexed naphthalene, respectively; and $c_{Nk,}$ c_{NCk} and c_{Dk} are the score values used to describe free naphthalene, 1:1 complexed naphthalene and 2:2 complexed naphthalene in sample *k*, respectively.

At this point, there are several ways to get the information regarding the three sensitivity factors. If a sample in the dataset used for PARAFAC does not contain β-cyclodextrin, and the concentration of naphthalene in the sample is known, the sensitivity factor for free naphthalene can easily be determined for this sample. However, using the rest of the information in the dataset is more desirable. Therefore, a system of equations can be used to solve for the sensitivity factors in Excel. It turns out this method could be used for the 1:1 complex as well and will be compared to the method previously described. The sensitivity factors calculated from the 2:2 experiment are shown in Table 4-15. These sensitivity factors are then used to calculate K_2 for each sample assuming a total concentration of naphthalene of 5 x 10^{-5} M. Plotting the data using the following expression:

$$\alpha_D c_{Dk} = K_2((\alpha_{NC} c_{NCk})^2)$$

(4-15)

A K_2 is determined to be about 12000 ±400 at 25 ^oC and about 4100 ± 200 at 10 ^oC. The value at 25 ^oC is comparable to the value determined by Hamai of 4000.[7] A value at 10 ^oC was not available..

The values for K_1 in these trials were calculated as well. The same method was used, as in the previous 1:1 experiment. The results are in Table 4-15. The trend in K_1 for these trials also follows the trend expected of an expected increase with a decrease in temperature. The fluorescence intensity of naphthalene increases measurably with a decrease in temperature, corresponding to more complex formation.[7,14]

The K_1 values were lower than the previous 1:1 experiment. This is probably due to imperfect starting material β-cyclodextrin. This type of result was observed many times in other 1:1 complex trials. It was either not recrystallized enough or it was contaminated at some other point of the process. The less than perfect R^2 values also indicate it may be a solubility issue as well. Either way the calculated equilibrium constants are still adequate and follow expected trends.

4.4.4 Multiple Complex Experiment

Using the score values from the 4-component model, formation constants were determined by the same method used in the separate 1:1 experiments. These results are summarized in Table 4-15. These formation constants compare favorably to the constants in the literature and the previous experiments.

Experiment	**α ratio**	**K_1 (± std. error)**	**R^2**	**K_2 (± std. error)**	**R^2**	**α_D ($x10^{-9}$)**	**α_N ($x10^{-9}$)**	**α_{NC} ($x10^{-9}$)**
A 1:1	0.588	1020 (140)	0.949					
N 1:1	0.693	690 (10)	0.997					
N 2:2 (25 °C)	0.478	450 (20)	0.991	12000 (400)	0.972	2.63	4.34	1.19
N 2:2 (10 °C)	0.632	570 (70)	0.896	4100 (200)	0.982	0.624	3.72	1.85

A 1:1 (Multi)	0.852	1560 (170)	0.964	
N 1:1 (Multi)	0.362	288 (20)	0.982	

Table 4-15: Summary of formation constant determinations. Constants for anthracene (A) and Naphthalene (N) are calculated from the individual experiments and the simultaneous experiment.

Since this experiment calculated equilibrium constants at two different temperatures, Van't Hoff plots can be used to calculate other thermodynamic quantities for the 1:1 and 2:2 complex. Assuming the enthalpy change is constant in this temperature range we can us the Van't Hoff equation below:[29]

$$lnK = -\frac{\Delta H}{RT} + \frac{\Delta S}{R}$$

(4-16)

Where K is the equilibrium constant, ΔH is the change in enthalpy, ΔS is the change in entropy, T is temperature in Kelvin, and R is the gas constant 8.314 J/mol·K. Using the data from our experiments, we obtain the plots are shown for each complex in

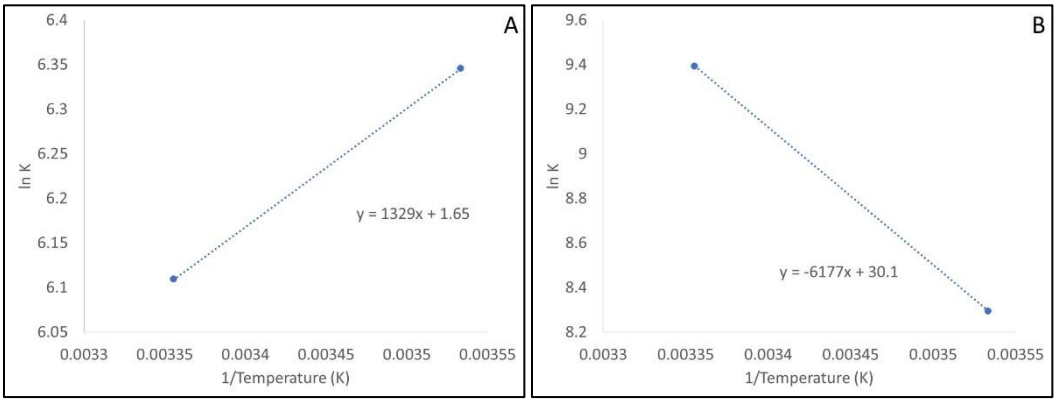


Figure 4-24: Van't Hoff Plots for 1:1 Complex (A) and 2:2 complex (B)

Using the information from these plots, the Gibbs free energy change can be calculated as well using:

$$\Delta G = \Delta H - T\Delta S \qquad \textbf{(4-17)}$$

The thermodynamic information for each complex using the data from this work is summarized in Table 4-16.

Complex	ΔH (J/mol)	ΔS (J/mol·K)	ΔG (J/mol)
1:1	-11049	13.7	6963
2:2	51353	250.4	-122221

Table 4-16: Thermodynamic data for Naphthalene complex experiments

4.5 Conclusions and Future Work

This work has shown that multidimensional fluorescence with PARAFAC is capable of investigating host-guest systems that are otherwise not investigated extensively using simple spectroscopic techniques. Interesting directions for this work would be investigating other complexes that have not been thoroughly investigated. Also, investigating the maximum number of complexes PARAFAC can model simultaneously would be worthwhile. This would require relatively larger datasets than the ones presented here, although, the amount of data extracted from relatively few samples in this case was very worthwhile. This work could adequately model datasets with a limited number of components and still calculate relatively good formation constants.

It is also recommended that other cyclodextrins be used in future experiments. Regular β-cyclodextrin is not very easy to work with, due to poor

solubilty and aggregation problems. A cyclodextrin such as (2-Hydroxypropyl)-β-cyclodextrin is very comparable and much more soluble. This would likely result in more consistent experiments as well.

4.6 References

(1) Bortolus, P.; Monti, S. In *Advances in Photochemistry*; John Wiley & Sons, Inc., 1996; pp 1–133.

(2) Fujimura, K.; Ueda, T.; Ando, T. *Anal. Chem.* **1983**, *55* (3), 446–450.

(3) Sand, D. M.; Schlenk, H. *Anal. Chem.* **1961**, *33* (11), 1624–1625.

(4) Serio, N.; Chanthalyma, C.; Prignano, L.; Levine, M. *ACS Appl. Mater. Interfaces* **2013**, *5* (22), 11951–11957.

(5) Brewster, M. E.; Loftsson, T. *Adv. Drug Deliv. Rev.* **2007**, *59* (7), 645–666.

(6) Hamilton, J. A.; Chen, L. *J. Am. Chem. Soc.* **1988**, *110* (17), 5833–5841.

(7) Hamai, S. *Bull. Chem. Soc. Jpn.* **1982**, *55*, 2721–2729.

(8) Benesi, H. A.; Hildebrand, J. H. *J. Am. Chem. Soc.* **1949**, *71* (8), 2703–2707.

(9) Evans, C. H.; Partyka, M.; Van Stam, J. *J. Incl. Phenom. Macrocycl. Chem.* **2000**, *38* (1/4), 381–396.

(10) Sanemasa, I.; Takuma, T.; Deguchi, T. *Bull. Chem. Soc. Jpn.* **1989**, *62* (10), 3098–3102.

(11) García-Zubiri, Í. X.; González-Gaitano, G.; Isasi, J. R. *J. Incl. Phenom. Macrocycl. Chem.* **2007**, *57* (1–4), 265–270.

(12) Hashimoto, S.; Thomas, J. K. *J. Am. Chem. Soc.* **1985**, *107* (16), 4655–4662.

(13) Tang, J.-J.; Cline Love, L. J. *Anal. Chim. Acta* **1997**, *344*, 137–143.

(14) Grabner, G.; Rechthaler, K.; Mayer, B.; Köhler, G.; Rotkiewicz, K. *J. Phys. Chem. A* **2000**, *104* (7), 1365–1376.

(15) Chen, H.; Kenny, J. E. *Analyst* **2010**, *135* (7), 1704–1710.

(16) Poh, B.-L.; Chow, Y. M. *J. Incl. Phenom. Mol. Recognit. Chem.* **1992**, *14* (2), 85–90.

(17) Bonini, M.; Rossi, S.; Karlsson, G.; Almgren, M.; Nostro, P. Lo; Baglioni,

P. *Langmuir* **2006**, *22* (4), 1478–1484.

(18) Simona Rossi; Massimo Bonini; Nostro, P. Lo; Baglioni, P. *Langmuir* **2007**, *23* (22), 10959–10967.

(19) MacDonald, B. C.; Lvin, S. J.; Patterson, H. *Anal. Chim. Acta* **1997**, *338* (1–2), 155–162.

(20) Bro, R. *Chemom. Intell. Lab. Syst.* **1997**, *38* (2), 149–171.

(21) Hall, G. J. Chemometric Characterization and Classification of Estuarine Water Through Multidimensional Fluorescence, Tufts University, 2006.

(22) *Handbook of Chemistry and Physics*, 98th ed.; Rumble, J. R., Ed.; CRC Press: Boca Raton, FL, 2017.

(23) Berlman, I. B. *Handbook of Fluorescence Spectra of Aromatic Molecules*, 2nd ed.; Academic Press: New York, 1971.

(24) Riu, J.; Bro, R. *Chemom. Intell. Lab. Syst.* **2003**, *65* (1), 35–49.

(25) Lorenzo-Seva, U.; ten Berge, J. M. F. *Methodology* **2006**, *2* (2), 57–64.

(26) Stedmon, C. A.; Bro, R. *Limnol. Oceanogr.* **2008**, *6*, 572–579.

(27) Andersson, C. A.; Bro, R. *Chemom. Intell. Lab. Syst.* **2000**, *52* (1), 1–4.

(28) Leurgans, S.; Ross, R. T. *Stat. Sci.* **1992**, *7* (3), 289–310.

(29) McQuarrie, D. A. *Quantum chemistry*; University Science Books: Mill Valley, Calif., 1983.