2015-07-01

Correlation of Fluorescence Spectroscopy and Biochemical Oxygen Demand (BOD5) Using Regression Analysis

Alexander Tetteh Narteh
Brigham Young University - Provo

Correlation of Fluorescence Spectroscopy and Biochemical Oxygen Demand (BOD_5) Using Regression Analysis

Alexander Tetteh Narteh

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

M. Brett Borup, Chair
William F. Christensen
Gustavious P. Williams
E. James Nelson

Department of Civil and Environmental Engineering

Brigham Young University

July 2015

ABSTRACT

Correlation of Fluorescence Spectroscopy and Biochemical Oxygen Demand (BOD_5) Using Regression Analysis

Alexander Tetteh Narteh
Department of Civil and Environmental Engineering, BYU
Master of Science

This research uses Regression analysis of fluorescence spectroscopy results to correlate with Biochemical Oxygen Demand (BOD_5. Fluorescence spectroscopy was applied to samples taken from seven sample sites in the Provo and Orem waste water treatment plants found in Utah County. A total of 161 samples were collected for this research. 23 samples each were taken from four sites in the Provo waste water treatment plant namely Provo head works, aeration basin, primary filter settlement basin and the Provo effluent basin. The Orem head works, the clarifier and the Orem effluent basin were the three sample sites in the Orem waste water treatment plant where 23 samples each were collected to carry out the analysis.

The fluorescent characteristics of the samples were determined using fluorescence spectrometry. These intensities were correlated with standard five day Biochemical Oxygen Demand (BOD_5 values which were used as a measure of the amount of biodegradable organic material present. Chemical oxygen demand (COD data were also taken from these treatment plants for correlation purposes.

Three different correlation analyses were made which were the correlation of fluorescence spectroscopy excitation-emission matrix (EEM against (1 individual sites BOD and COD values (2 Provo only and Orem only BOD and COD values (3 combined Provo and Orem BOD and COD values. The correlation of Individual site EEMs against BOD and COD values produced the best results. There was a higher correlation of EEM with BOD data than COD data. The R-squared for the combined Provo and Orem BOD data was 0.756 and that for COD was 0.729. Very high R-squared was obtained for Provo Influent data and Orem Influent data which were 0.955 and 0.946 respectively. This method can be used by wastewater stakeholders in deriving quick results in determining potential pollution events within a shorter time frame. This research demonstrates that there is a correlation between EEM and BOD/COD.

Keywords: regression analysis, biochemical oxygen demand, chemical oxygen demand, fluorescence spectroscopy, excitation-emission-matrices, tryptophan-like materials

ACKNOWLEDGEMENTS

I wish to thank Dr. Brett Borup for his guidance and enormous contribution to this research, and I also want to thank him for his tutelage and role in mentoring me on wastewater and drinking water processes. Special thanks also go to Dr. William F. Christensen for this expertise in Modern Regression methods and generating the programming code I used in my research.

I will also want to thank Michael Scheetz and Iva Riveros of the Provo Wastewater Treatment Plant in Provo, UT, for their immense help in aiding me with my samples and data collection in their facility. I also express immense thanks to L. G. Burton and Randy J. Sandoval of the Orem Wastewater Treatment Plant in Orem, UT, for the help they offered in my data and sample collection.

I also express gratitude to my beloved wife, Leanna Narteh, and my brother, Victor Narteh, for the encouragement and support they offered me in fulfilling this quest.

Special thanks also go to my friends Daniel Blake, Cari Eileen Ferrell, Zachary White, Zola Adjei and Brent Hargis for their encouragement and assistance as well.

TABLE OF CONTENTS

`

LIST OF TABLES

`

LIST OF FIGURES

1 INTRODUCTION

1.1 Problem Statement

The biochemical oxygen demand (BOD_5) of wastewater and surface water is the measurement of the amount of molecular oxygen required by micro-organisms in the biochemical oxidation of organic matter in water (i.e., degradation of the organic matter by micro-organisms into carbon dioxide and water with new cell growth). Although BOD_5 is a good indicator of the concentration of organic pollutants in the water, biochemical oxidation is a slow process, and the test, in its present form, takes 5 days until results are obtained.

The fluorescence spectroscopy has been used in the investigation of the composition, concentration, distribution and dynamics of organic matter from various sources in a range of aquatic environments. Past research has indicated there is a correlation between the fluorescent characteristics of a sample and the BOD of that sample (Karube, et al. 1977, Coble 1996, Reynold and Ahmad 1997, Parlanti, et al. 2000, N. Hudson, A. Baker, et al. 2008). Fluorescence spectroscopy also known as fluorometry or spectrofluorometry is a type of electromagnetic spectroscopy which analyzes fluorescence from a sample. It involves using a beam of light, usually ultraviolet light, that excites the electrons in molecules of certain compounds and causes them to emit light; typically, but not necessarily, visible light.

The Perkin-Elmer LS55 Fluorescence Spectrometers was the spectrometer used in this sample analysis. Fig 1-1 displays the excitation and signal viewing which occurs in a Perkin-Elmer LS55 Fluorescence Spectrometer.

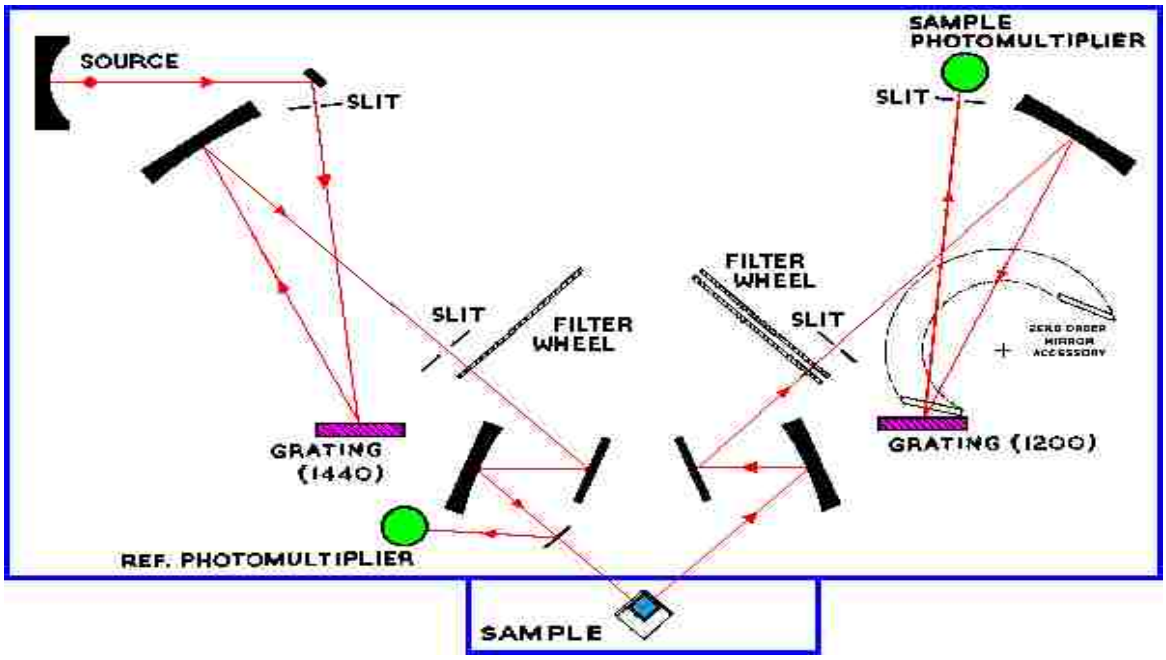


Figure 1-1 Perkin-Elmer LS55 Fluorescence Spectrometer Excitation & Signal Viewing (Hornak 1999)

Generally, the sample being examined has a ground electronic state (a low energy state) of interest, and an excited electronic state of higher energy. Within each of these electronic states are various vibrational states. The sample is first excited, by absorbing a photon, from its ground electronic state to one of the various vibrational states in the excited electronic state. Collisions with other molecules cause the excited molecule to lose vibrational energy until it reaches the lowest vibrational state of the excited electronic state (Mehta 2013).

The molecule then drops down to one of the various vibrational levels of the ground electronic state again, emitting a photon in the process. As molecules may drop down into any of several vibrational levels in the ground state, the emitted photons will have different energies, and thus frequencies. Therefore, by analyzing the different frequencies of light emitted in fluorescent spectroscopy, along with their relative intensities, the structure of the different vibrational levels can be determined.

In a typical fluorescence (emission) measurement, the excitation wavelength is fixed and the detection wavelength varies, while in fluorescence excitation measurement the detection wavelength is fixed and the excitation wavelength is varied across a region of interest.

An excitation emission matrix is measured by recording the emission spectra resulting from a range of excitation wavelengths and combining them all together. This is a three dimensional surface data set: fluorescence intensity as a function of excitation and emission wavelengths, and is typically depicted as a contour display as shown in Fig 1-2. The legends found at the in Fig 1-2 represent the various fluorescence intensities.

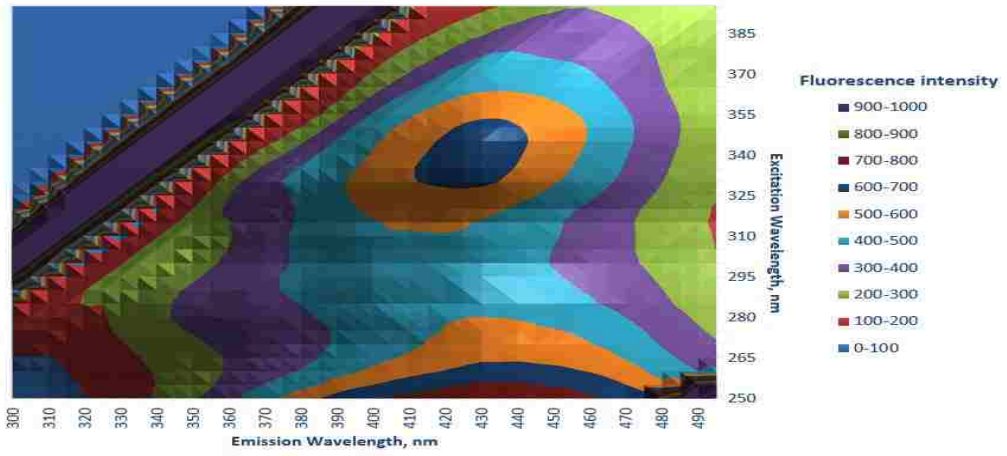


Figure 1-2 Three Dimensional Surface Data set Depicted as a Contour Display

The BOD_5 test is a crucial environmental index for monitoring organic pollutants in waste water but the conventional test is not suitable for process control and monitoring, where a rapid feedback is desirable. It is therefore of considerable interest to develop alternative methods that may replace this time- consuming test. It may be possible to determine the organic strength of a waste sample with the use of fluorescence spectroscopy in the range of 20-30 minutes. The purpose of this research is to determine if fluorescence spectroscopy method can be used to accurately estimate the BOD_5 and COD of wastewater samples. Emphasis will be put more on BOD_5 than COD in this research.

1.2 **Gaps in Research**

The current international standard for measuring organic strength of a water sample is the BOD_5 which takes 5-days e(Eaton, et al. 1995). Attempts have been made to simplify and speed up this test. The first BOD sensor was described by Karube, et al. (1977). Various rapid microbial biosensors have been devised to overcome the restrictive lag time inherent in the traditional BOD_5 test (Karube, et al. 1977, Princz and Olah 1990, Reiss, et al. 1998, R. Riedel, et al. 1988, Riedel, et al. 1998, Sangeetha, et al. 1996, Tanaka, et al. 1994) but these are limited either by the availability of oxygen, or by pure microbial cultures with a narrow substrate range, or they require calibration to a BOD_5 standard solution.

Many BOD sensors have been developed based on monitoring the dissolved oxygen (DO) consumption by immobilized microorganisms (Kulys and Kadziauskiene 1980, l. K. Riedel, et al. 1988, Marty, et al. 1997, G. Chee, et al. 1999). However, such BOD sensors using a membrane and an O_2 probe have many disadvantages such as membrane fouling, short-term stability, and calibration drift.

To date, rapid biochemical oxygen demand (BOD) techniques have combined microorganisms and oxygen electrodes to measure changes in dissolved oxygen concentrations at the sensor surface in order to quantify microbial activity (Slama, et al. 1996). However, the availability of oxygen has been shown to be a limiting factor for microbial catabolism and is a significant t factor contributing to the 5-day requirement of the traditional test (Reshetilov, et al. 1998)

Approaches that artificially increase the oxygen concentration have been used in an attempt to reduce the lag time (Reshetilov, et al. 1998). Another alternative approach has been to exclude oxygen altogether by substituting a mediator (Pasco, et al. 2000).

Fluorescence intensities reported in arbitrary fluorescence units (AFU) were correlated with standard five day BOD_5 values which were used as an indicator of the amount of biodegradable organic material present (Hudson et al. 2007). Research using EEM to correlate with BOD_5 have been done in the past (Hudson, Baker and Ward, et al. 2008) but not specifically to find a combination of specific cells which correlates with BOD_5 any time the model is run.

1.3 **Objectives**

This research uses Regression analysis of the EEM data to determine the BOD_5 of a wastewater sample. Regression analysis of fluorescence spectroscopy EEM data to determine COD of wastewater samples were also carried out but much emphasis will be placed on BOD_5 in this research. The fluorescence spectroscopy was applied to waste water samples which were collected from seven sites in the Provo and Orem Waste water treatment plant namely Provo head works, Provo primary filter settlement basin, Provo aeration basin, Provo effluent basin, Orem head works, Orem clarifier and the Orem effluent basin.

This research seeks to determine if one general model can predict BOD_5 using the EEM data and also to discover whether this model will work for every wastewater treatment plant, specific treatment plants or specific sample points in selected treatment plants.

2 LITERATURE REVIEW

2.1 Biochemical Oxygen Demand (BOD)

BOD is a measure of the dissolved oxygen consumed by microorganisms (aerobic bacteria) during the oxidation of substances in waters and wastes. Normal sources of BOD are readily biodegradable organic carbon (carbonaceous, CBOD) and ammonia (nitrogenous, NBOD) which are compounds consisting of metabolic byproducts of plant and animal wastes and human activities (domestic and industrial wastewaters). Severe dissolved oxygen (DO) depletion and fish kills in receiving water bodies are some of the effects of high level BOD discharge of wastes. Despite numerous shortcomings, standardized methods for the quantification of BOD in wastewaters have remained virtually unchanged for decades. Alternative techniques and estimation methods have been proposed. With the advancement in research in surface water quality the simulation of BOD discharges have been formulated into several mathematical models.

A variety of tests have been developed to measure the BOD and also estimate the rate of oxygen depletion in water or wastewater samples. Surface water quality models and wastewater treatment are mostly carried out using this oxidation rate. The diagram below (Fig 2-1) theoretically exhibits the BOD of a wastewater sample as a function of time.

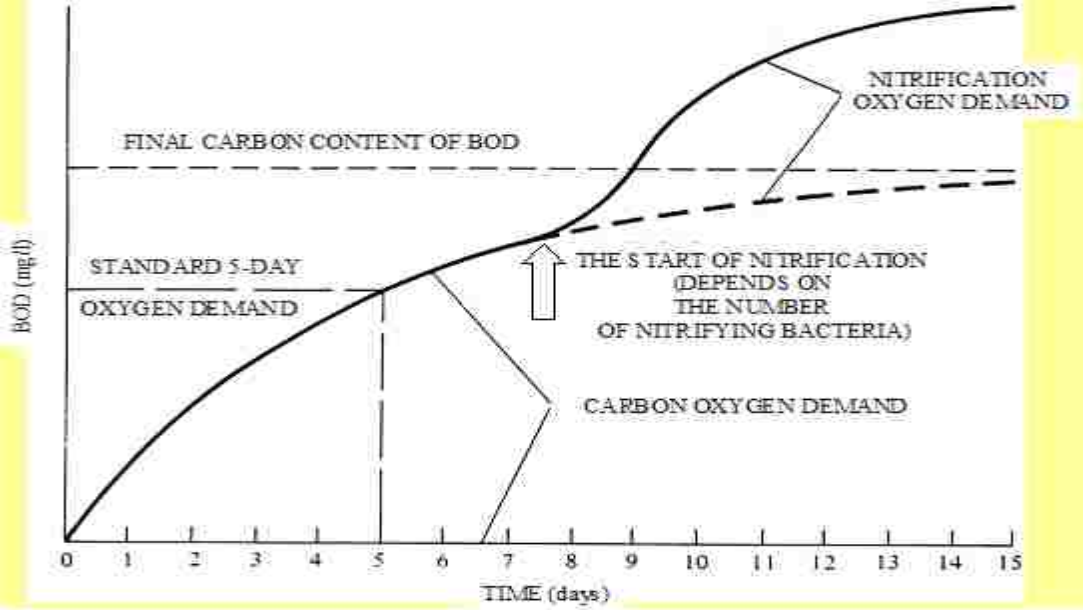


Figure 2-1 Biochemical Oxygen Demand and Time (Csaba and Csaba 2011)

2.1.1 Five-day BOD (BOD_5)

The BOD_5 test is a standardized test that provides information regarding the organic strength of wastewater. The amount of oxygen consumed in a sample within a five-day period is measured under carefully controlled and standardized conditions. The [BOD_5], expressed as part per million, ppm, is the difference between the initial DO reading and the corresponding (final) measurement made on the fifth day of incubation.

2.1.2 Problems Related to the BOD_5 Test

The BOD_5 test measures only the oxygen taken up during the biological oxidation of organic matter present. Its disadvantage is the long time required by the test which takes 5 days. It is run as a laboratory based biodegradation test and relies upon the presence of a thriving

microbial community that may be naturally present in the sample or artificially introduced by addition of a seed, commonly a known volume of sewage effluent of known BOD. A standard BOD test is run in the dark at a temperature of 20 degrees Celsius for 5 days. This test is temperature dependent and the reaction must occur in the dark because algae may be present and if the light is available the production of oxygen may occur.

There are several practical difficulties associated with the test including; the measurement depends on temperature, oxygen concentration, presence of toxins, plus the type, quantity and quality of seeding microorganisms. The test is labor intensive and the results highly vary.

Although BOD_5 is widely used and is very popular around the world, it has its down side to its application and its biggest inconvenience is the 5 day test period which delays analysis of potential pollution events. For this reason, surrogate techniques such as fluorescence spectroscopy which provides a quicker estimate of the polluting load based on fluoro-spectrometer measurements could provide several advantages. This method will provide quick results which will be of great use to environmental stake holders and regulators as well.

2.2 **Chemical Oxygen Demand (COD)**

Chemical oxygen demand (COD) is a measure of the capacity of water to consume oxygen during the decomposition of organic matter and the oxidation of inorganic chemicals such as ammonia and nitrite. COD measurements are commonly made on samples of waste waters or of natural waters contaminated by domestic or industrial wastes. Chemical oxygen demand is measured as a standardized laboratory assay in which a closed water sample is incubated with a strong chemical oxidant under specific conditions of temperature and for a

particular period of time. A commonly used oxidant in COD assays is potassium dichromate ($K_2Cr_2O_7$) which is used in combination with boiling sulfuric acid (H_2SO_4). Because this chemical oxidant is not specific to oxygen-consuming chemicals that are organic or inorganic, both of these sources of oxygen demand are measured in a COD assay.

COD is related to BOD; however, BOD only measures the amount of oxygen consumed by microbial oxidation and is most relevant to waters rich in organic matter. It is important to understand that COD and BOD do not necessarily measure the same types of oxygen consumption. For example, COD does not measure the oxygen-consuming potential associated with certain dissolved organic compounds such as acetate. However, acetate can be metabolized by microorganisms and would therefore be detected in an assay of BOD. In contrast, the oxygen-consuming potential of cellulose is not measured during a short-term BOD assay, but it is measured during a COD test.

2.3 **Fluorescence Spectroscopy**

Fluorescence is the phenomenon which occurs when an electron is excited to a higher energy level (electron orbit) by absorption of light energy, and then releases energy as light as it drops to a lower energy level. Chromophores and fluorophores are examples of fluorescent compounds. Chromophores absorb light (often pigments) and fluorophores absorb and remit light energy. An EEM can be created by simultaneously scanning excitation and emission wavelengths through set path length-ranges and measuring the fluorescent intensity.

Each resulting EEM consists of hundreds of measurement combinations of a single water sample, with excitation wavelength on one axis, emission wavelength is the second, and fluorescence intensity forms a third axis as displayed in Fig 2-2.

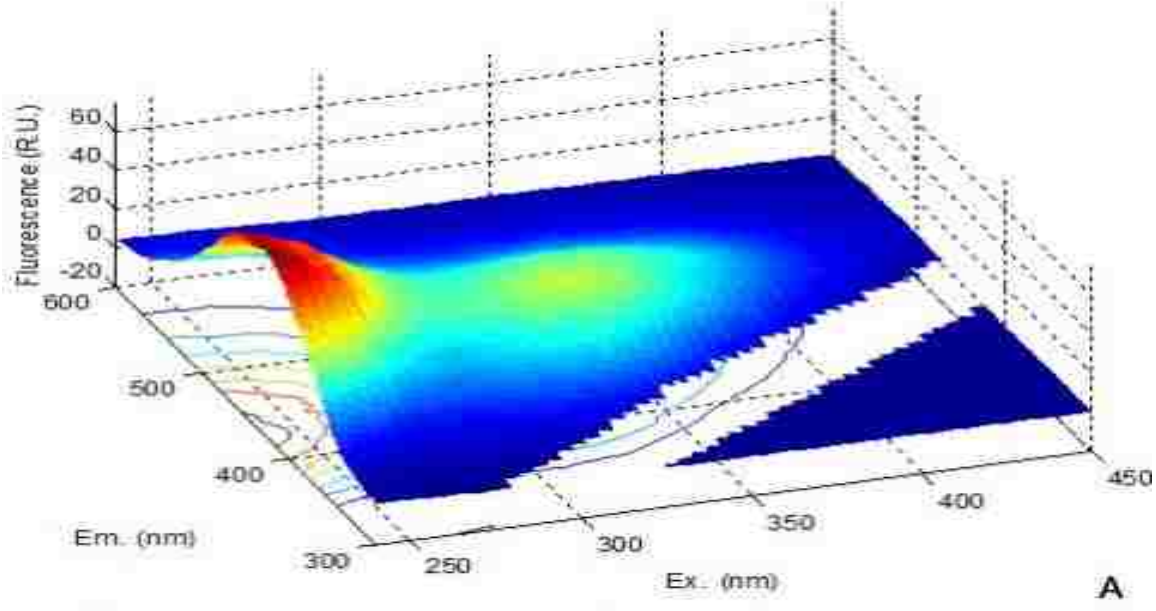


Figure 2-2 Three dimensional Contour Plot of an EEM (Traving, Riemann and Stedmon 2015) Instrumentation

The device that measures fluorescence is called a fluorometer or fluorimeter. The PerkinElmer LS55 luminescence spectrometer was the fluorimeter used for this research. The excitation source, sample cell and fluorescence detector are the main parts of the fluorimeter. Usually a deuterium or xenon lamp is used for the excitation of molecules in solution. The fluorescence is detected by a photomultiplier tube. The excitation spectrum and the fluorescence spectrum are found by scanning of the excitation monochromator and the fluorescence monochromator respectively.

Fluorescence is most often measured at a 90° angle relative to the excitation light. This geometry is used instead of placing the sensor at the line of the excitation light at a 180° angle in order to avoid interference of the transmitted excitation light. No monochromator is perfect and it will transmit some stray light, that is, light with other wavelengths than the targeted. An ideal monochromator would only transmit light in the specified range and have a high wavelength-independent transmission. When measuring at a 90° angle, only the light scattered by the sample causes stray light.

2.4 **Analysis of Data**

At low concentrations the fluorescence intensity will generally be proportional to the concentration of the fluorophore. Unlike in UV/visible spectroscopy, 'standard', device independent spectra are not easily attained. Several factors influence and distort the spectra, and corrections are necessary to attain 'true', i.e. machine-independent, spectra. The different types of distortions will here be classified as being either instrument- or sample-related. Firstly, the distortion arising from the instrument is discussed. As a start, the light source intensity and wavelength characteristics varies over time during each experiment and between each experiment. Furthermore, no lamp has a constant intensity at all wavelengths. To correct this, a beam splitter can be applied after the excitation monochromator or filter to direct a portion of the light to a reference detector.

Additionally, the transmission efficiency of monochromators and filters must be taken into account. These may also change over time. The transmission efficiency of the monochromator also varies depending on wavelength. This is the reason that an optional reference detector should be placed after the excitation monochromator or filter. The percentage

of the fluorescence picked up by the detector is also dependent upon the system. Furthermore, the detector quantum efficiency, that is, the percentage of photons detected, varies between different detectors, with wavelength and with time, as the detector inevitably deteriorates.

In the past decades, fluorescence spectroscopy has been used widely in different fields of research such as the biomedical applications of time-resolved fluorescence spectroscopy as a major research tool in bio-chemistry; biophysics and chemical physics (J. Lakowicz 1994), advances in laser technology and the development of long-wavelength probes (Berndt and Lakowicz 1992), the use of fluorophores with emission in the infrared upon short wavelength excitation such as serum and seawater (Thompson, et al. 2002) and the relationship between sample depth and the corresponding fluorescence excited signal.

Fluorescence spectroscopy has also been used widely in the water sciences. Biological systems contain many natural fluorophores; such as amino acids (tryptophan, tyrosine and phenylalanine), vitamins, coenzymes and aromatic organic matter in general that can be detected by fluorescence spectroscopy, regardless if they are intra- or extra-cellular (Galinha, et al. 2011). Therefore, this technique has great potential for real-time monitoring of biotechnological systems, as was previously demonstrated by Li and Humphrey (1991). Dissolved organic matter (DOM) is the most studied fraction; and fluoresces at various wavelengths (Coble, 1996, Hudson, Baker et al. 2007). Many studies demonstrate that DOM has an intrinsic fluorescence (Lochmueller and Saavedra 1986,Coble 1996, Baker 2001, Hudson, Baker et al. 2007).

These fluorescent properties are due to the presence of chromophores (particles that absorb light) and fluorophores (chromophores that absorb and then emit light at different wavelengths). The fluorescent fraction of DOM, or fluorometric dissolved organic matter (FDOM), is comprised of these compounds that emit light when excited (Mopper, Feng et al.

1996, Hudson, Baker et al. 2007). The intrinsic fluorescence of DOM has been investigated and studied extensively (Coble 1996, Patel-Sorrentino, Mounier et al. 2002, Kowalczuk et al. 2009, Murphy 2010, Blake 2014). The most commonly studied FDOM components include humic acids and amino acids in proteins and peptides. Humic acids are produced from the decomposition of natural plant material by biological and chemical processes in both terrestrial and aquatic environments (Baker 2001, Hudson, Baker et al. 2007, Ghervase, et al. 2010). FDOM exhibits characteristic peaks when subjected to fluorescence spectroscopy which correspond to amino acids such as tryptophan, fulvic acid, and tyrosine (Coble 1996, Baker 2001, Baker 2004, Hudson, Baker et al. 2007). Due to this characteristic, FDOM can be used as an indicator of the presence of organic matter from wastewater treatment processes.

Fluorescence spectroscopy has been used to characterize the quality of natural water bodies and to track anthropogenic pollution across a body of water (Stedmon, Markager and Bro 2003, Hall, Clow et al. 2005, Hudson, Baker et al. 2007, Guo, et al. 2010). In this way fluorescence spectroscopy operates as a fingerprint technique, allowing researchers to track the pollution through the ecosystem (Yan, Li and Myrick 2000, Baker 2001, Stedmon, Markager and Bro 2003, Hall, Clow et al. 2005, Ghervase et al. 2010, Blake 2014).

In natural waters, organic matter exists in dissolved, colloidal and particulate states with dissolved organic matter (DOM) being the most studied fraction, although some emphasis has been placed upon the colloidal fraction and its importance in water chemistry (Mopper, et al. 1996, Patel-Sorrentino, Mounier and Benaim 2002). The origin of most fluorophores in natural surface waters are from the breakdown of plant materials which are humic like in nature (peaks C and A, (Coble 1996)).

Waste waters including sewage effluents (Reynolds and Ahmad, 1997, Reynolds, 2002, Chen et al., 2003 and Ferrell 2009), farm wastes (Baker, 2002) and landfill leachates (Baker and Curry, 2004) have been found to be rich in microbial derived T and B fluorescence (Fig 2-3) and these peaks have been used as tracers of waste waters in natural waters (Baker and Inverarity, 2004 and Baker et al., 2004). Reynolds and Ahmad, (1997) determined that the sewage treatment process reduced peak T intensity to a much greater extent than the humic-like A and C peaks. This suggests that the T peak in untreated sewage, derived from anthropogenic activity, represents fresher, less degraded material with a high potential for oxidation, and that the fluorescence intensity of this peak may be used as a surrogate for the Biochemical Oxygen Demand (BOD) and COD test.

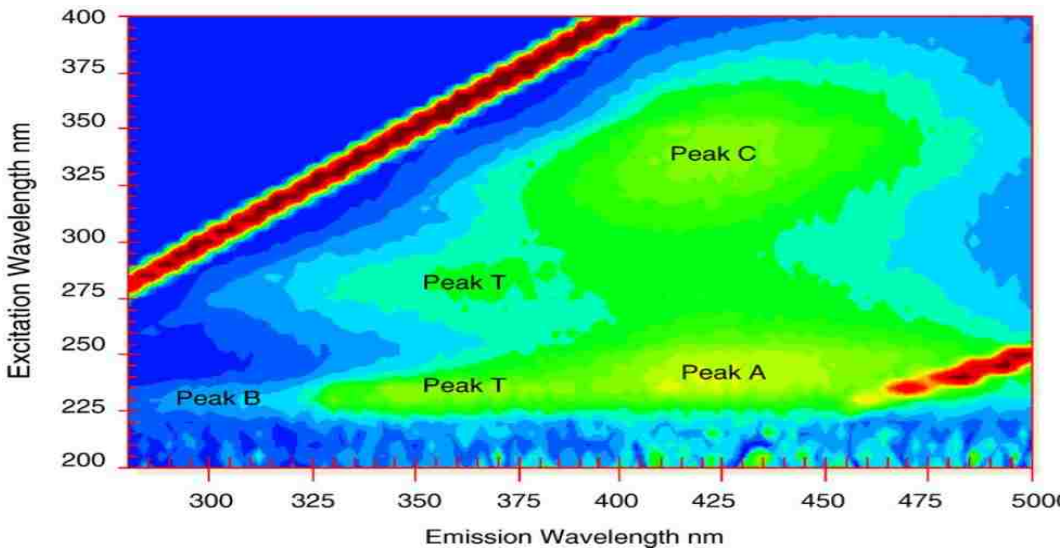


Figure 2-3 Example EEM illustrating positions of T_1, T_2, C and A Peaks. (Hudson, Baker and Ward, et al. 2008)

The fluorescence of a folded protein is a mixture of the fluorescence from individual aromatic amino acid residues (tryptophan, tyrosine, and phenylalanine). Protein fluorescence is generally excited at 280 nm or at longer wavelengths, usually at 295 nm. Most of the emissions are due to excitation of tryptophan residues, with a few emissions due to tyrosine and phenylalanine (Mocz n.d.).

Tryptophan has a much higher quantum yield and stronger fluorescence than the tyrosine and phenylalanine. The quantum yield, intensity and wavelength of maximum fluorescence emission o tryptophan is very dependent on solvent type. Tyrosine has a strong absorption bands at 280 nm and has characteristic emission profile when excited by light at this wavelength. Tyrosine is a weaker emitter than tryptophan, but it may still contribute greatly to protein fluorescence because it is usually present in large number.

Phenylalanine is made of only a benzene ring and a methylene group and the simple structure of this compound contributes to its weak fluorescence. The product of quantum yield and molar absorptivity maximum is especially low for this residue

The intensity derived from the EEM peak can be used as a measure of the concentration of the fluorophore to ppm or ppb levels, depending upon the fluorophore. Tryptophan-like fluorescence, which has been found to relate to the activity of the biological community showed the strongest correlation with BOD_5. Fluorescence analysis of the tryptophan-like peak (excitation/emission wavelength region 275/340 nm) is found to provide an accurate indication of the presence, and relative proportions of bioavailable organic material present (natural or anthropogenic). It therefore provides an insight relating to its oxygen depleting potential. Thus fluorescence spectroscopy is recommended as a portable or laboratory tool for the determination

of the presence of biodegradable organic matter with intrinsic oxidizing potential in natural waters (Hudson, Baker and Ward, et al. 2008).

This research study uses a regression analysis method to correlate excitation-emission matrices with the BOD_5 and COD. Regression analysis is a statistical tool for the investigation of relationships between variables. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables. Usually, the investigator seeks to ascertain the causal effect of one variable upon another.

More specifically, regression analysis enhances the understanding of how the typical value of the dependent variable (or 'criterion variable') changes when any one of the independent variables is varied, the other independent variables are held fixed.

The performance of regression analysis methods in practice depends on the form of the data generating process, and how it relates to the regression approach being used. Since the true form of the data-generating process is generally not known, regression analysis often depends to some extent on making assumptions about this process. These assumptions are sometimes testable if a sufficient quantity of data is available. Regression models for prediction are often useful even when the assumptions are moderately violated, although they may not perform optimally (Lindley 1987)

This study uses models based on Regression analysis of the EEM for it analysis to enhance its BOD prediction. The fluoro-spectrometer takes about 20 minutes to produce EEM data which can be analyzed in numerous ways.

Fluorescence excitation wavelength is mainly represented by (λ_{ex}) and the emission wavelength by (λ_{em}). Fluorophores exhibit fluorescence at excitation/emission wavelengths λ_{ex}

304–347 nm λ_{em} 405–461 nm (Peak C in Fig 2-3) and λ_{ex} 217–261 nm λ_{em} 395–449 nm (Peak A in Fig 2-3) (Hudson, Baker and Ward, et al. 2007). In addition to humic-like material, tryptophan-like and tyrosine-like material as "free" molecules or bound in amino acids and proteins (commonly referred to as peaks T and B respectively, (Coble 1996)) also exhibit fluorescence at distinctive wavelengths in natural waters. Tryptophan- like fluorescence (Peak T1 in Fig 2-3) occurs at λex/em 275–296/330–378 nm while tyrosine-like fluorescence (Peak B in Fig 2-3) was not commonly seen and is not addressed in this work. Peak T also has a shorter wavelength excitation/emission pair (named T2) with excitation at between λ_{ex} 216–247 nm and emission at between λ_{em} 329–378 nm.

Tryptophan-like fluorescence may be exhibited by natural waters where tryptophan is present as 'free' molecules or else bound in proteins, peptides or humic structures. Peaks T and B are related to microbial activity (Parlanti, et al. 2000)and may be transported into a system (allochthonous) or be created by microbial activity within a system (autochthonous).

Fluorescence spectroscopy has also been used in the study and identification of microbial communities both by fingerprinting and characterizing (Seaver, et al., 1998; Smith, et al., 2004; Elliott, et al., 2006), species identification (Gray, et al., 1998; Leblanc & Dufour, 2002) and in process monitoring (Saadi, et al., 2006; Farabegoli, et al., 2003). These works and that of (Cammack, et al. 2004) illustrate that fluorescence in specific spectral regions is associated with microbial activity. A summary of previous works relating fluorescence to BOD_5 is presented in (Table 2-1).

The variation in wavelengths is likely to be due to the physical characteristics of individual samples such as pH, metal ions, sample concentration (Vodacek and Philipot 1987).

Table 2-2 displays the different excitation and emission wavelengths used by different researchers and their component of study.

Table 2-1 Summary of Previous Work Relating Fluorescence to BOD_5 **(N. Hudson, A. Baker, et al. 2008)**

Analytical parameter	Optical parameter for correlation	Correlation coefficient	Number of samples	Comments	Water type	Author
Biochemical Oxygen Demand	Fluorescence intensity at excitation 250 and 350 nm, emission 430 nm	–	c.200 samples from different sites	No relationship found. Wrong wavelength pairs examined.	Rivers and sewage and industrial effluents, U.K.	(Comber et al., 1996)
Biochemical Oxygen Demand	Tryptophan (280/340 nm) (T_1)	0.94–0.97	129 samples, 3 sites	Relationship is site specific.	24 h composites of raw, settled and treated waste from 3 sewage works.	(Reynolds and Ahmad, 1997)
Biochemical Oxygen Demand	Tryptophan (248/340 nm) (T_2)	0.97	25 samples, 1 site	Up to 350 mgl^{-1} BOD. Relationship expected to be site specific.	Single sewage works.	(Ahmad and Reynolds, 1999)
Biochemical Oxygen Demand	Tryptophan (280/350 nm) (T_1)	0.96–0.99	101 samples, 2 sources		Single sewage works and a synthetic sewage.	(Reynolds, 2002)
Biochemical Oxygen Demand	Tryptophan (220–230 nm/340–370 nm) (T_2)	0.94–0.98	40 samples, 3 sites	Relationship site specific.	3 Landfill leachates, U.K.	(Baker and Curry, 2004)
Biochemical Oxygen Demand	Tryptophan (220/350 nm) (T_2)	0.85	434 samples, 62 sites	Samples not paired.	River Tyne catchment, N.E. England, U.K.	(Baker and Inverarity, 2004)
Biochemical Oxygen Demand	Tryptophan-like fluorescence (280/350 nm) (T_1)	0.906	294 samples, from 267 sites of which 141 effluents, 124 surface waters and 2 pollution incidents	Paired samples.	River, sewage effluent and industrial effluents S.W. England, U.K.	This work

Table 2-2 Wavelengths used in other Different Experiments

Excitation	Emission	Component	Study
220	350	Tryptophan	(Baker and Inverarity. 2004)
230	370	Tryptophan	(Baker and Curry. 2004)
260	380	Humic acid	(Coble 1996)
275	340	Tryptophan	(Coble 1996)
275	310	Tyrosine	(Coble 1996)
280	350	Tryptophan	(Reynolds 2002)
280	355	Tryptophan	(Hudson, Baker et al. 2008)
280	340	Tryptophan	(Reynolds and Ahmad, 1997)
280	340	Tryptophan	(Ahmad and Reynolds. 1999)
330	430		
330	440		
340	420		
340	430	Fulvic acid	(Hudson, Baker et al. 2008)
350	420	Humic acid	(Coble 1996)
350	430		
350	440		

3 MATERIALS AND METHODS

3.1 Sampling Locations and Sites

23 samples each were collected from a total of seven sites found in two locations. The locations where samples were collected were the Provo and Orem waste water treatment plants found in Utah County, Utah. Samples were taken from four sites in the Provo waste water treatment plant (Fig 3-1 and Fig 3-2) as displayed below namely "Provo Head Works (PI)", "Aeration Basin (PABI)" "Primary Filter Settlement Basin (PFSBE)", and the "Provo Effluent Basin (PE)".

Figure 3-1 Aerial View of the Provo Waste Water Treatment Plant

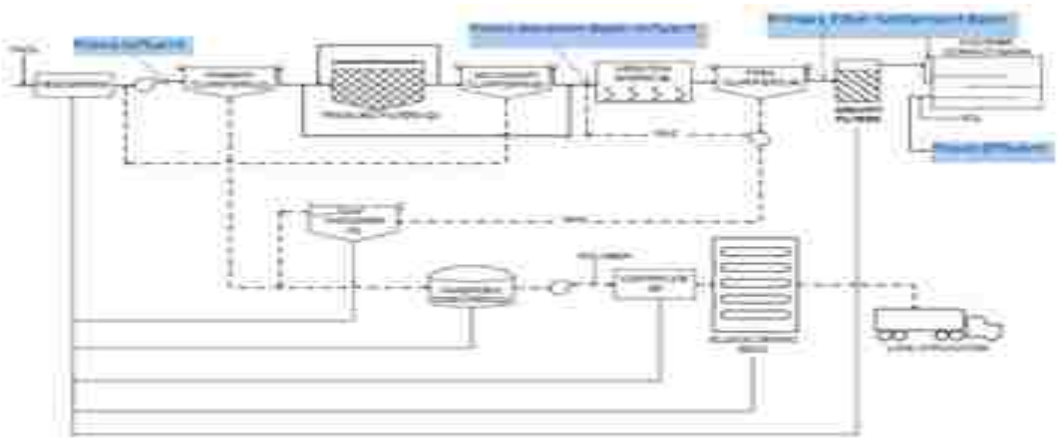


Figure 3-2 Flow Diagram of the Provo Waste Water Treatment Plant

The "Orem Head works (OI) ", "the Clarifier (OPE)" and the "Orem Effluent Basin (OE)" were the three sample sites in the Orem waste water treatment plant (Fig 3-3) where samples were collected from to carry out this research.

Figure 3-3 Aerial View of the Orem Waste water Treatment Plant

3.2 **Testing Procedures**

Samples collected for this research were collected over a period starting from January 2014 to April 2014. Samples collected were initially filtered using a Whatman brand medium flowrate cellulose filter paper, which has a pore size of 11µm. After the filtering process, samples were then diluted according to ratios displayed in Table 3-1 below so that the fluorescence intensity would be less than the maximum measurable limit of 1000.

Table 3-1 Dilution Ratios of Wastewater Samples

Sample Collection Sites	Dilution factor
Provo Influent	1/6
Provo ABI	1/3
Provo FSBE	1/3
Provo Effluent	1/3
Orem Influent	1/5
Orem PE	1/3
Orem Effluent	1/3

The PerkinElmer LS55 luminescence spectrometer (Fig 3-4) was then used in analyzing the various samples. The cuvettes and beakers used were rinsed thoroughly to avoid any form of contamination to samples. After the filtering and dilution process, the cuvette was then filled with sample and placed in the Luminescence spectrometer. A total of 30 scans were performed on each sample. The scans consisted of a range of excitation wavelengths from 250-400nm in 5nm increments, and fluorescence intensity was measured at corresponding emission wavelengths from 300-500nm with 5nm increments as well. Each scan generated one data file,

for a total of 30 files for each sample. The excitation and Emission slits were both 10nm and the Scan speed was 500 nm/min.

Figure 3-4 PerkinElmer LS55 Luminescence Spectrometer

BOD_5 values were determined by the treatment plant personnel using the 5210 B. (5-day BOD test) from the "Standard Methods manual for the Examination of Water and Wastewater". The COD values were also determined by using the Hach DR/4000 procedure (method 8000) (Hach, Klein and Gibbs 1997). Water samples were refluxed for 2 hours after which the photo-spectrometer was used to determine the COD values.

3.3 **Analysis Method**

Samples were run using a PerkinElmer LS55 luminescence spectrometer. An EEM was produced from the data obtained after 20 minutes of running water sample. Each resulting EEM consisted of 1200 of measurement combinations of a single water sample, with excitation

wavelength on one axis, emission wavelength on the second and fluorescence intensity forms a third axis. Fig 3-5 - Fig 3-11 are EEMs produced in this way on March 5, 2014.

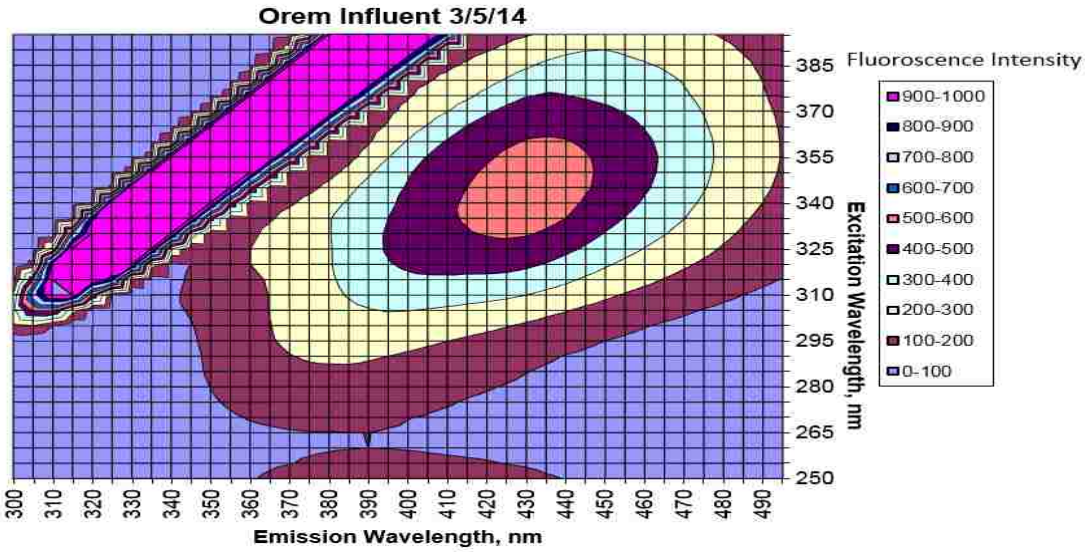


Figure 3-5 Orem Influent EEM Display 3/5/14

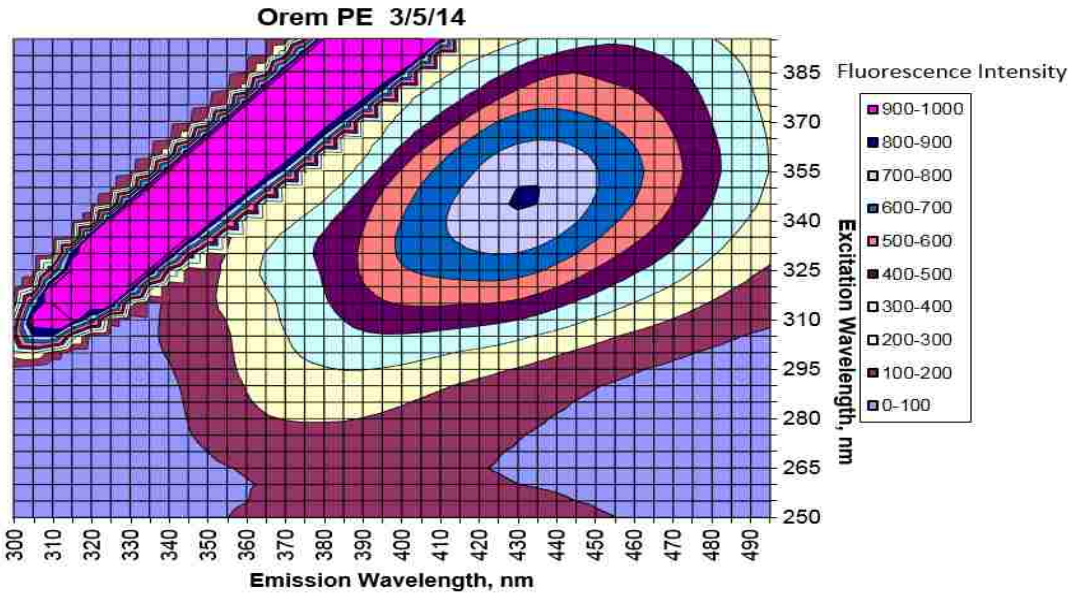


Figure 3-6 Orem PE EEM Display 3/5/14

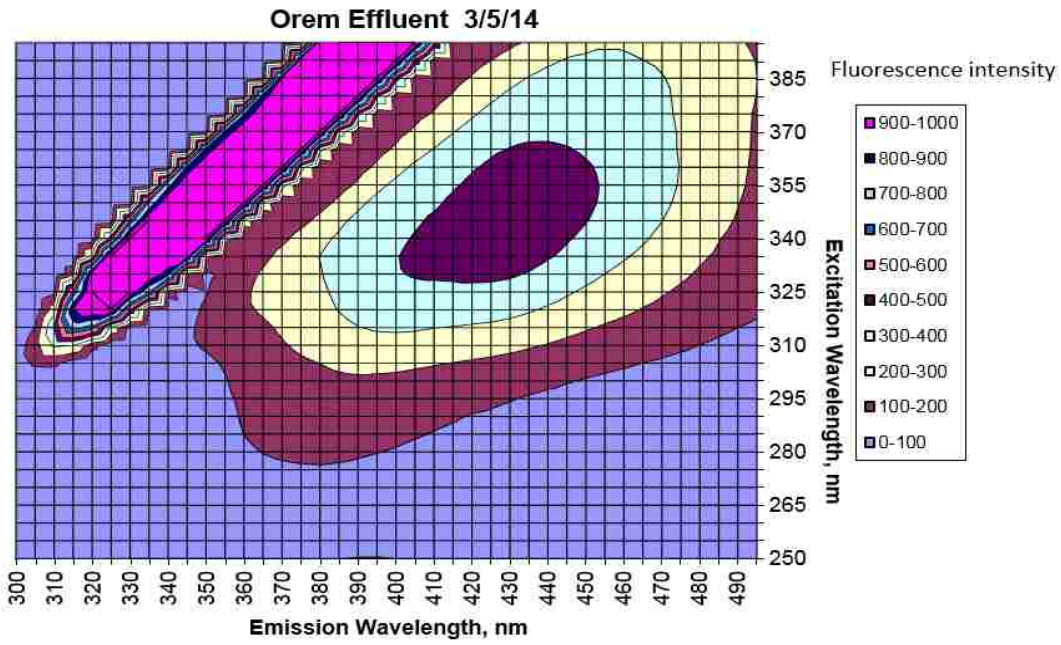


Figure 3-7 Orem Effluent EEM Display 3/5/14

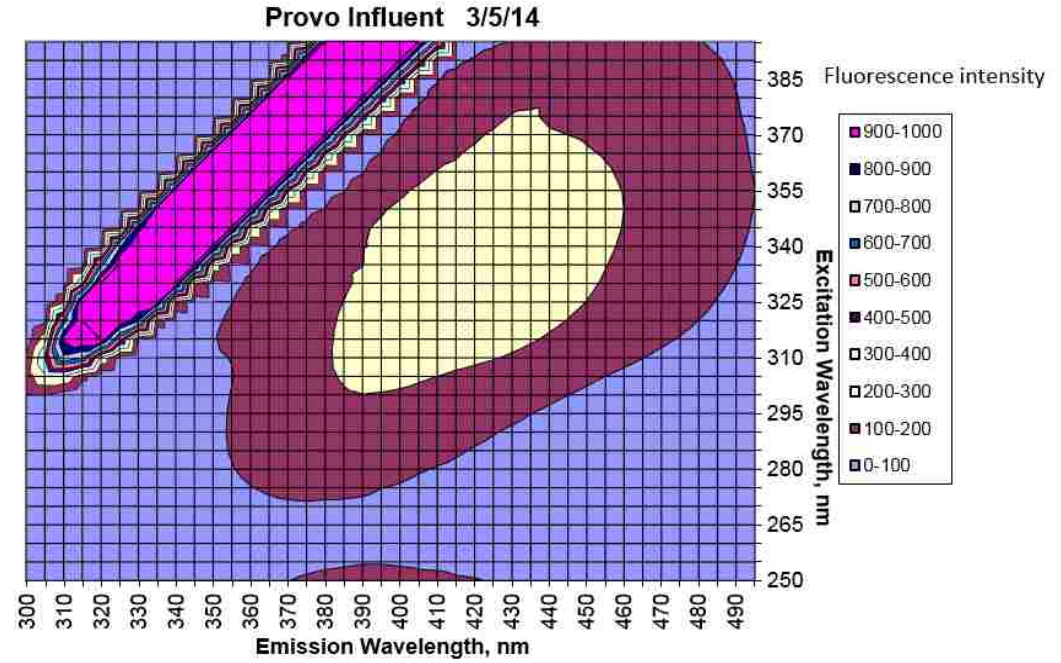


Figure 3-8 Provo Influent EEM Display 3/5/14

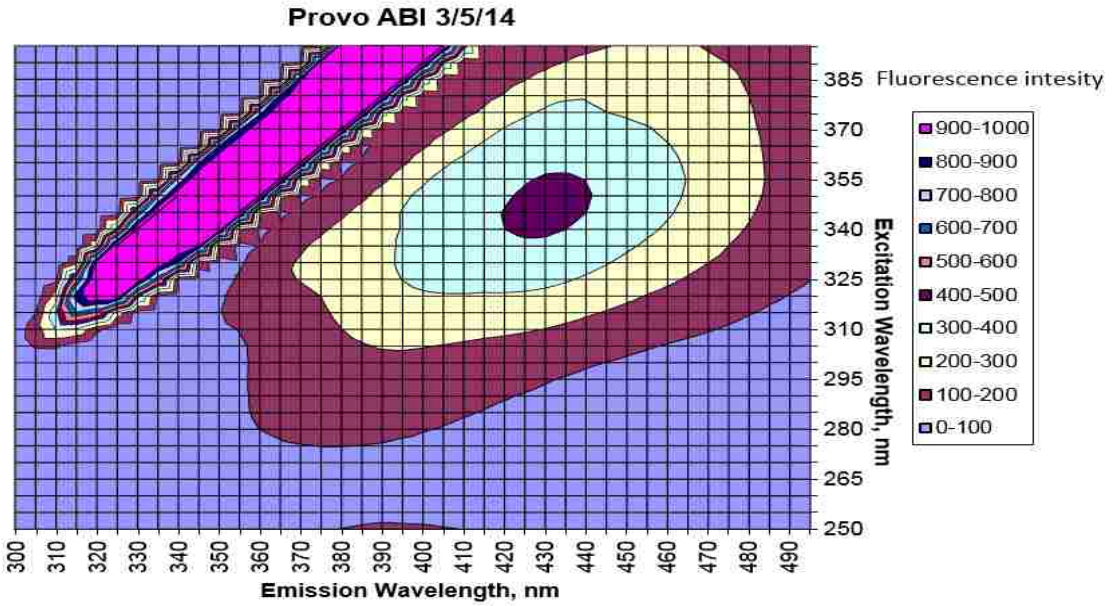


Figure 3-9 Provo ABI EEM Display 3/5/14

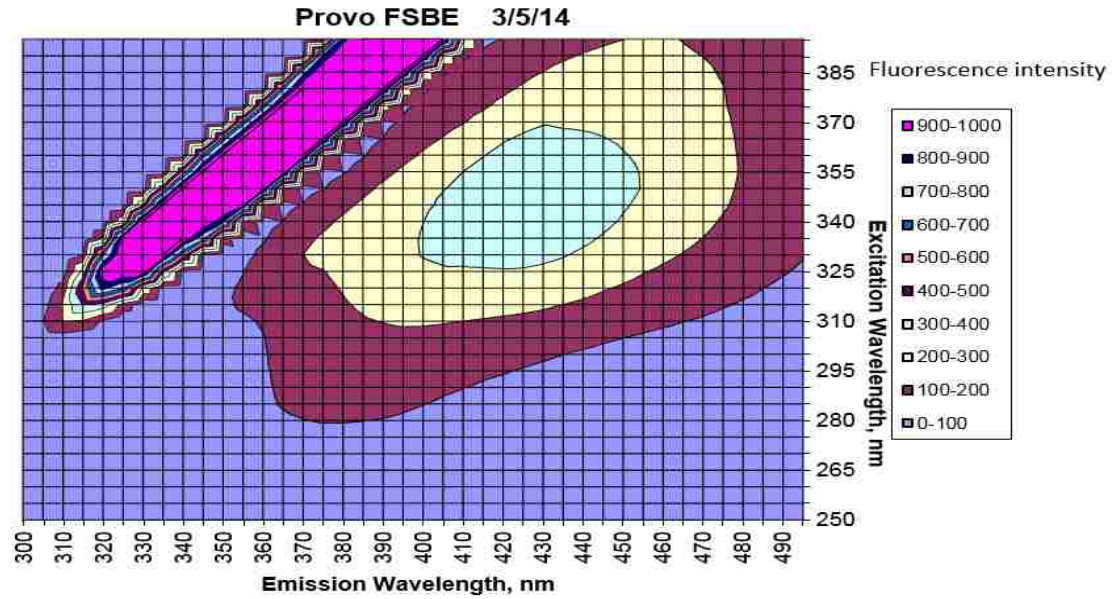


Figure 3-10 Provo FSBE EEM Display 3/5/14

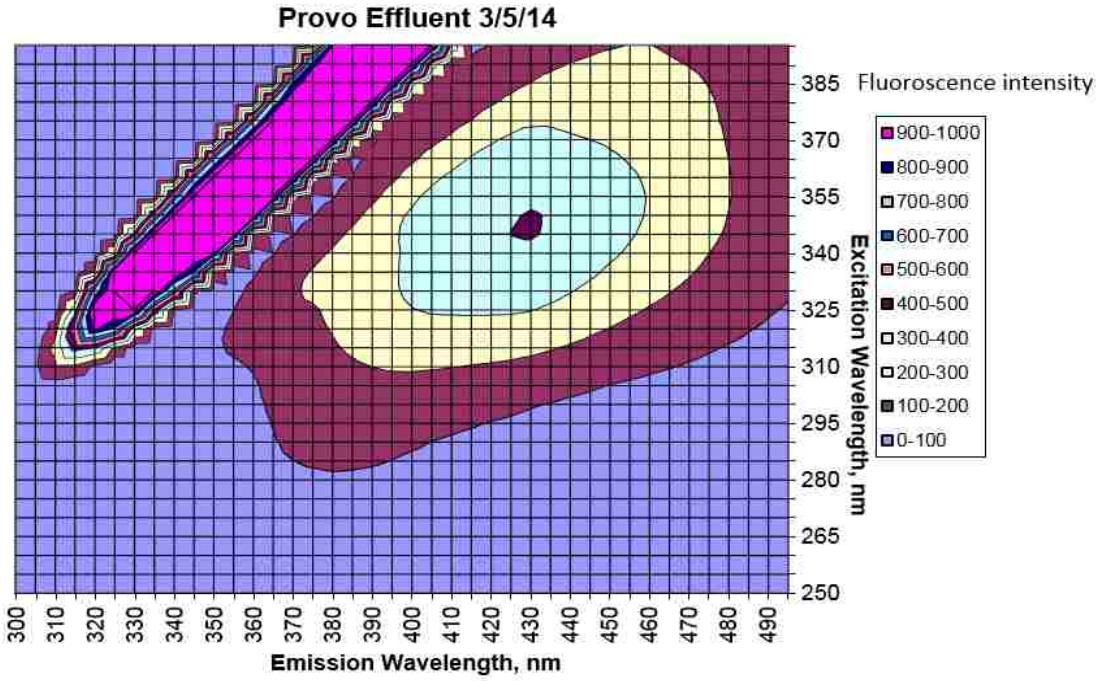


Figure 3-11 Provo Effluent EEM Display 3/5/14

For various phases within the water treatment process, multiple regression techniques were used to predict the values of the univariate BOD_5 response based on the values of the EEM matrix. The Brute Force approach was implemented using regressions to determine which EEM locations yielded the best predictors of BOD_5. 10,000 combinations of 10 possible cells were randomly selected in the EEM matrix. A multiple regression model with 10 covariates was then fit using these 10,000 combinations. Using R-squared, which describes how much variation of BOD_5 values can be explained by the 10 sites selected in each model, a site/cell selection was initiated to determine which cells were most predictive of R-squared. The R-squared were then arranged from the highest to the lowest and the predictive models with highest R-squared were saved. Each model is programmed to have 10 sites. In theory, the more times a site/cell showed up in these models, the better the site/cell was at predicting BOD values. This Brute force

process was ran 6 times on the selected data and the total amount was reduced from 1200 to 600 to 300 to 150 to 75 down to 18 cells. A subset selection of the 10 best possible sites was then selected from the 18 sites. This was done by testing every possible combination of 10 sites and the subset which produced the highest R-squared was selected.

This Brute force algorithm process was run according to the given pattern:

- Using both Provo and Orem EEM data together
- Using individual plants ie (Provo data only and Orem EEM data only)
- Using individual site EEM data for Provo(such as Provo influent, Provo ABI etc) and individual site data for Orem (such as Orem Influent etc)

The algorithm code can be found in the appendix section and the statistical model used in this analysis is listed below

BOD_5 Model equation =

$$B_1X_1+B_2X_2+B_3X_3+B_4X_4+B_5X_5+B_6X_6+B_7X_7+B_8X_8+B_9X_9+B_{10}X_{10}+Intercept$$

B = Coefficients

X = best cell/point value derived from the EEM

Intercept = needed intercept for prediction

4 RESULTS

4.1 Regression Results

A total of 161 samples from all the 7 sites were analyzed using the PerkinElmer LS55 luminescence spectrometer to produce Excitation Emission Matrices (EEMs). These samples were taken from 4 sites in the Provo wastewater treatment plant for 23 different days and 3 sites in the Orem wastewater treatment plant for 23 days as well. “Provo Head Works (PI)”, “Aeration Basin (PABI)” “Primary Filter Settlement Basin (PFSBE)”, “Provo Effluent basin (PE)”, “Orem Head works (OI) ”, “the Clarifier (OPE)” and the “Orem Effluent Basin (OE)” were the different sites of which samples were collected in the 2 treatment plants. BOD and COD tests were run by wastewater Lab personnels to obtain BOD and COD values for data analysis and correlation.

Results from the regression model show that there is a correlation between EEMs and BOD_5 as well as EEMs and COD.

4.2 Biochemical Oxygen Demand Data Results:

Table 4-1 shows the result of Provo Influent BOD and EEM data. An R-squared of 0.9547 was derived when 10 points were used in the correlation process.

Table 4-1 Results of Provo Influent BOD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	395,315	1.60E-03	-1.7235	0.1157
2	300,340	6.60E-03	-90.8879	0.264
3	305,370	1.63E-03	-40.3746	0.462
4	310,355	1.48E-07	70.076	0.669
5	330,390	3.20E-02	30.3062	0.7886
6	410,310	3.14E-03	1.8702	0.838
7	475,265	3.00E-03	9.8058	0.889
8	340,390	5.16E-02	-25.0818	0.933
9	485,360	1.61E-03	-10.9994	0.93712
10	300,345	5.67E-04	-24.5386	0.9547
Intercept		7.54E-06	128.2479	

Table 4-2 represents results of Orem influent BOD and EEM data. The model produced a very high R-squared of 0.988 when 10 points were used.

Table 4-2 Results of Orem Influent BOD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	360,380	1.38E-07	-13.559	0.0576
2	360,330	5.27E-05	19.669	0.0732
3	380,370	4.97E+06	-33.438	0.3597
4	365,335	4.45E-06	-26.529	0.6955
5	360,340	7.55E-04	-18.317	0.733
6	360,335	7.95E-06	35.419	0.8325
7	325,375	1.24E-06	-182.937	0.9323
8	305,370	6.45E-09	104.745	0.954
9	350,325	1.13E-03	-10.383	0.9683
10	355,375	4.16E+06	-0.652	0.9884
Intercept		5.10E-06	33259.406	

Table 4-3 represents results of Provo BOD and EEM data. The R-squared produced when 10 points were used was 0.7184.

Table 4-3 Results of Provo BOD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	315,315	1.45E-01	0.05277	0.351
2	455,335	7.16E-02	-2.857	0.481
3	425,300	1.70E-03	1.385	0.535
4	420,255	7.77E-05	-1.30068	0.685
5	495,315	4.72E-02	-7.458	0.695
6	475,315	9.71E-02	13.0015	0.702
7	480,315	3.32E-01	-7.8989	0.705
8	460,315	8.26E-02	-4.7081	0.712
9	430,330	2.46E-01	1.952	0.7151
10	440,330	2.66E-01	3.11482	0.7184
Intercept		8.47E-01	-4.46844	

Table 4-4 represents results of Orem BOD and EEM data .The R-squared produced when 10 points were used was 0.8839.

Table 4-4 Results of Orem BOD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	360,380	7.30E-06	-2.9235	0.195
2	360,330	1.10E-02	22.6873	0.658
3	380,370	3.97E-03	-0.6019	0.699
4	445,375	2.41E-02	2.906	0.771
5	360,335	6.84E-02	-31.214	0.817
6	360,340	5.10E-03	0.4055	0.848
7	335,325	9.89E-06	17.085	0.871
8	356,335	9.42E-02	29.16	0.874
9	325,380	1.42E-01	15.179	0.879
10	370,360	1.91E-07	-21.7012	0.8839
Intercept		6.04E-01	1861.9956	

Table 4-5 represents results of combined Provo and Orem BOD and EEM data. The R-squared produced when 10 points were used was 0.7566.

Table 4-5 Results of Combined Provo and Orem and EEM BOD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	350,305	1.97E-01	-2.6843	0.103
2	355,305	1.61E-03	0.8623	0.401
3	350,300	1.50E-01	-2.5632	0.576
4	440,320	6.08E-13	6.296	0.604
5	375,315	8.54E-09	-4.5533	0.735
6	405,250	3.34E-06	-1.555	0.736
7	495,260	3.69E-03	-4.9753	0.745
8	470,325	8.48E-03	-5.0096	0.753
9	485,255	3.06E-03	4.8247	0.755
10	190,335	3.20E-01	-1.3202	0.7566
Intercept		9.18E-01	2.3017	

4.3 Chemical Oxygen Demand Data Results

Table 4-6 shows the result of Provo Influent COD and EEM data. An R-squared of 0.9459 was derived when 10 points were used in the correlation process.

Table 4-6 Results of Provo Influent COD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	495,270	1.20E-03	45.013	0.04
2	310,370	1.10E-04	-216.455	0.043
3	315,360	6.32E-04	101.048	0.352
4	480,285	3.67E-03	35.464	0.69
5	430,260	5.46E-03	34.232	0.741
6	315,395	2.86E-04	95.854	0.78
7	425,265	4.29E-02	22.781	0.881
8	415,270	4.98E-02	14.59	0.889
9	470,260	2.32E-06	-92.331	0.9244
10	415,275	5.35E-05	-42.596	0.9459
Intercept		3.40E-07	477.852	

Table 4-7 shows the result of Orem Influent COD and EEM data. An R-squared of 0.9936 was derived when 10 points were used in the correlation process.

Table 4-7 Results of Orem Influent COD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	485,265	3.34E-06	-72.1558	0.002
2	475,280	3.59E-04	45.0283	0.546
3	450,270	1.79E-06	68.6163	0.647
4	300,255	3.12E-04	-171.69	0.6499
5	310,270	2.10E-06	116.3445	0.855
6	475,260	2.80E-06	-68.998	0.932
7	450,250	3.20E-03	-6.423	0.957
8	305,260	1.57E-04	-65.6263	0.974
9	485,275	3.48E-03	33.1808	0.9828
10	310,365	3.50E-07	-11.8	0.9936
Intercept		9.00E-04	152.164	

Table 4-8 represents results of Provo only COD and EEM data from the four sites in the Provo treatment plant location. The R-squared produced when 10 points were used was 0.7323.

Table 4-8 Results of Provo COD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	380,320	2.03E-02	-36.9696	0.331
2	430,315	4.10E-02	26.983	0.427
3	405,315	3.08E-04	16.867	0.446
4	385,320	2.75E-01	10.333	0.6
5	375,320	1.60E-03	27.806	0.634
6	385,315	9.68E-05	-15.585	0.661
7	435,325	2.76E-01	1.364	0.714
8	495,325	1.07E-01	-6.459	0.7195
9	435,315	1.60E-02	-30.757	0.727
10	320,315	2.36E-03	0.239	0.7323
Intercept		6.75E-02	-111.746	

Table 4-9 represents results of Orem COD and EEM data which is the combination of the data taken at the three sites in the Orem treatment plant location. The R-squared produced when 10 points were used was 0.8481.

Table 4-9 Results of Orem COD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	310,310	6.00E-02	2.9502	0.628
2	315,310	1.12E-01	-1.595	0.689
3	435,325	5.44E-02	12.204	0.709
4	465,315	1..72e-5	-13.393	0.767
5	440,325	5.68E-01	-5.7043	0.834
6	430,360	1.83E-03	-2.3101	0.836
7	310,315	1.32E-01	-1.1	0.845
8	445,330	7.55E-01	1.5108	0.847
9	470,290	1.58E-02	4.63	0.848
10	440,305	6.90E-01	-0.5793	0.8481
Intercept for formula		2.88E-02	162.4829	

Table 4-10 represents results of combined Provo and Orem COD and EEM data which is the combination of the data taken at the three sites in the Orem treatment plant location and the four sites in the Provo treatment plant. The R-squared produced when 10 points were used was 0.7286.

Table 4-10 Results of Combined Provo and Orem COD and EEM Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	405,315	7.30E-04	11.8993	0.0014
2	435,325	3.50E-04	11.4507	0.134
3	350,310	1.18E-07	29.1531	0.445
4	455,335	8.30E-04	-12.3344	0.5511
5	365,320	1.18E-02	-6.7484	0.651
6	455,330	3.79E-02	13.247	0.683
7	440,320	5.20E-04	-16.475	0.687
8	385,315	3.74E-03	-11.11	0.7073
9	345,305	7.00E-04	-16.858	0.7187
10	400,250	3.24E-08	-1.5301	0.7286
Intercept		6.21E-03	133.664	

Fig 4-1 shows the correlation of Provo predicted BOD_5 and the actual BOD_5 values from the treatment plant.

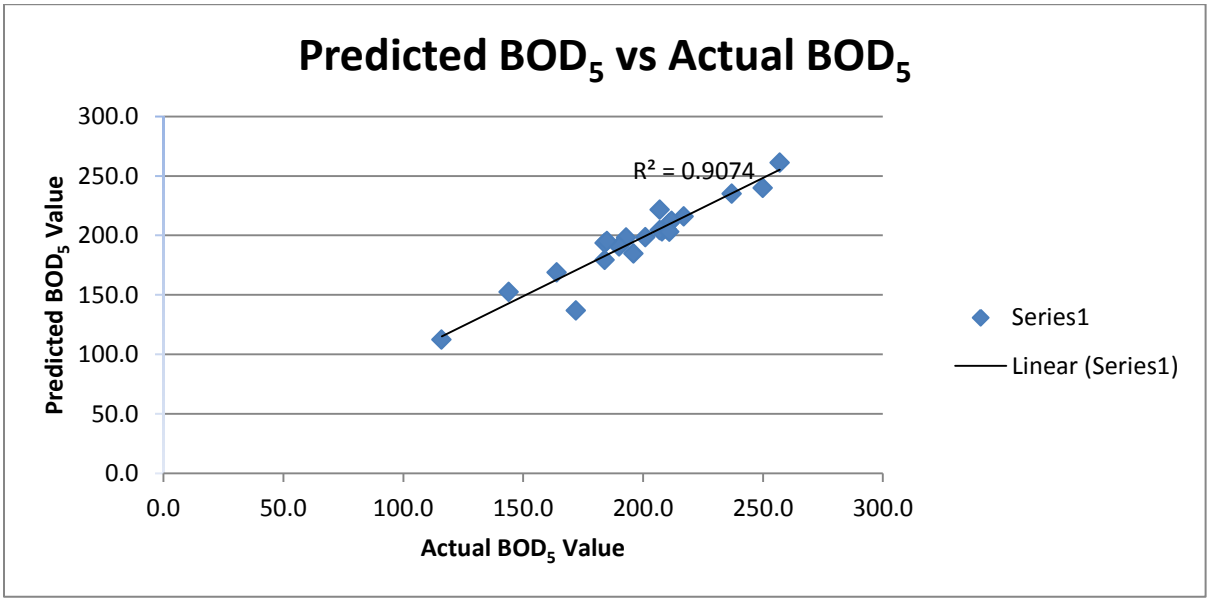


Figure 4-1 Provo Predicted BOD_5 vs Actual BOD_5

Table 4-11 shows the various days with their respective Provo predictive BOD_5 and actual BOD_5 values.

Table 4-11 Provo Predicted BOD_5 and Actual BOD_5

Date	Predicted BOD_5 Value	Actual BOD_5 Value
1/15/2014	210.1	211.0
1/16/2014	194.6	192.0
1/22/2014	215.8	217.0
1/23/2014	198.2	193.0
1/29/2014	203.3	208.0
1/30/2014	221.4	207.0
2/5/2014	112.2	116.0
2/6/2014	205.1	210.0
2/12/2014	168.8	164.0
2/13/2014	190.5	190.0
2/19/2014	179.2	184.0
2/20/2014	184.7	196.0
2/26/2014	235.0	237.0
3/5/2014	193.6	184.0
3/6/2014	136.9	172.0
3/12/2014	152.4	144.0
3/13/2014	198.3	201.0
3/26/2014	204.4	207.0
3/27/2014	261.0	257.0
4/3/2014	239.7	250.0
4/9/2014	212.2	212.0
4/16/2014	195.2	185.0
4/17/2014	203.0	211.0

Fig 4-2 shows the correlation of Orem Predicted BOD_5 and the Actual BOD_5 value using the Provo model.

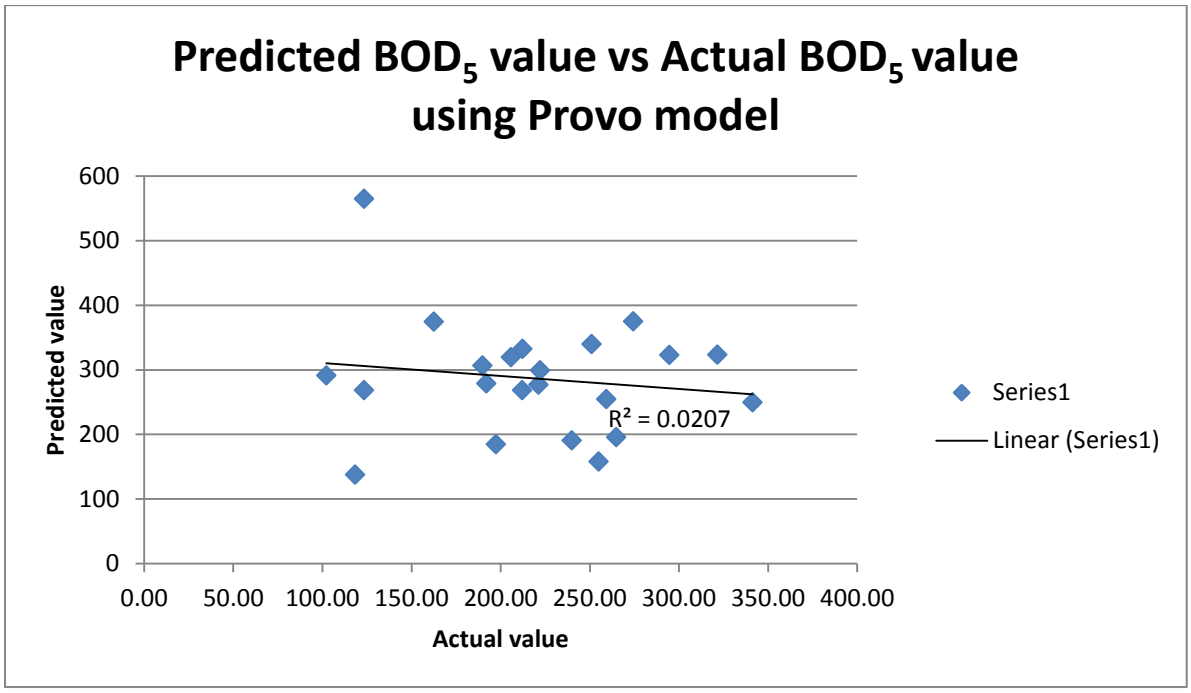


Figure 4-2 Orem Predicted BOD_5 value vs Actual BOD_5 value using Provo Model

Table 4-12 shows the various days with their respective Orem Predicted BOD_5 and the Actual BOD_5 value using the Provo model.

Table 4-12 Orem Predicted BOD_5 Values and Actual BOD_5 Values using Provo Model

Date	Predicted **BOD_5 Value**	Actual **BOD_5 Value**
1/15/2014	162.60	374.5
1/16/2014	191.92	278.75
1/22/2014	251.12	339.5
1/23/2014	274.35	375
1/29/2014	341.32	249.5
1/30/2014	221.26	276.75
2/5/2014	197.43	184.75
2/6/2014	240.01	190.5
2/12/2014	102.22	291.25
2/13/2014	118.40	137.75
2/19/2014	123.43	564.5
2/20/2014	123.43	268.5
2/26/2014	259.26	254.25
3/5/2014	321.58	323.25
3/6/2014	212.11	268.25
3/12/2014	222.23	299.25
3/13/2014	189.84	306.5
3/26/2014	205.91	319.25
3/27/2014	294.64	322.75
4/9/2014	264.93	195.5
4/16/2014	255.00	157.75
4/17/2014	212.21	332.25

5 DISCUSSION

Multiple regression technique predicts the values of the univariate BOD_5 response based on the values of the EEM. A total of 10 best fit cells were selected from 1200 cells using this method. These selected sites were reported in the Excitation Emission wavelength format (e.g. 280,330). For easy referencing, the EEM cell values were assigned as numbers ranging from 1 - 10 and simply called "Points" for easy graphing. Each result table comprised of specific rows such as Number of Points used, Best cells/sites, P-values, Coefficient and R- squared.

The P-value was calculated based on the respective least squares regression model. When running the algorithm code, the P-value was not used to determine the best model; however it helped in analyzing the model once a final model was selected. P-Values lower than .05 are excellent values for statistical significance (see Table 5-1). The lower the P-value the better. Using this concept, we concluded that points which had their P- values to be very small or very close to the zero were points which contributed significantly to a high R-squared.

The regression model provided a minimum correlation around the range of 0.7-0.72 in general and a maximum correlation from 0.98-0.99. Majority of the high correlations were derived when the model run individual data sites example Orem Influent. Depending on the sample data, some sites produced high R-squared when less than 10 points were used. Points used in the models were scattered and not specifically derived from a specific region. The EEM

displays of Fig 5.1 & Fig 5.2 are examples of scattered points selected from the Orem Influent and Provo Influent EEM data respectively. Even though all 10 points were needed to attain a high R-squared, some points were of more significance towards the R-squared than others due to their P value.

The coefficients were found using single least squares regression or multiple regression depending on the number of covariates been used in the model. Different graphs were plotted to show correlation between "Points" and "P value" as well as "Points", "R-squared" and "P value".

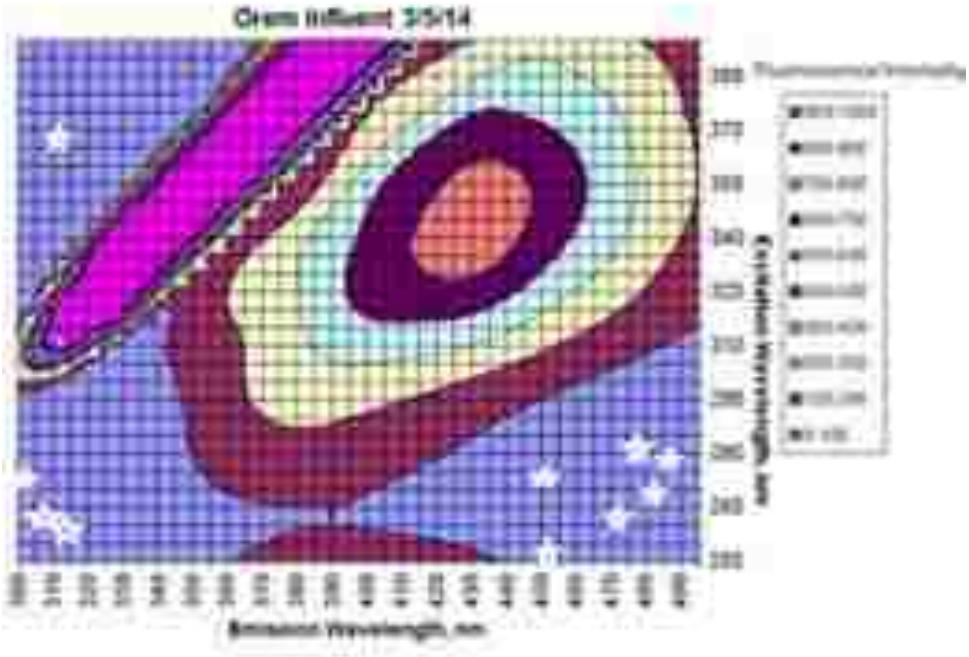


Figure 5-1 Scattered Points in Orem Influents EEM

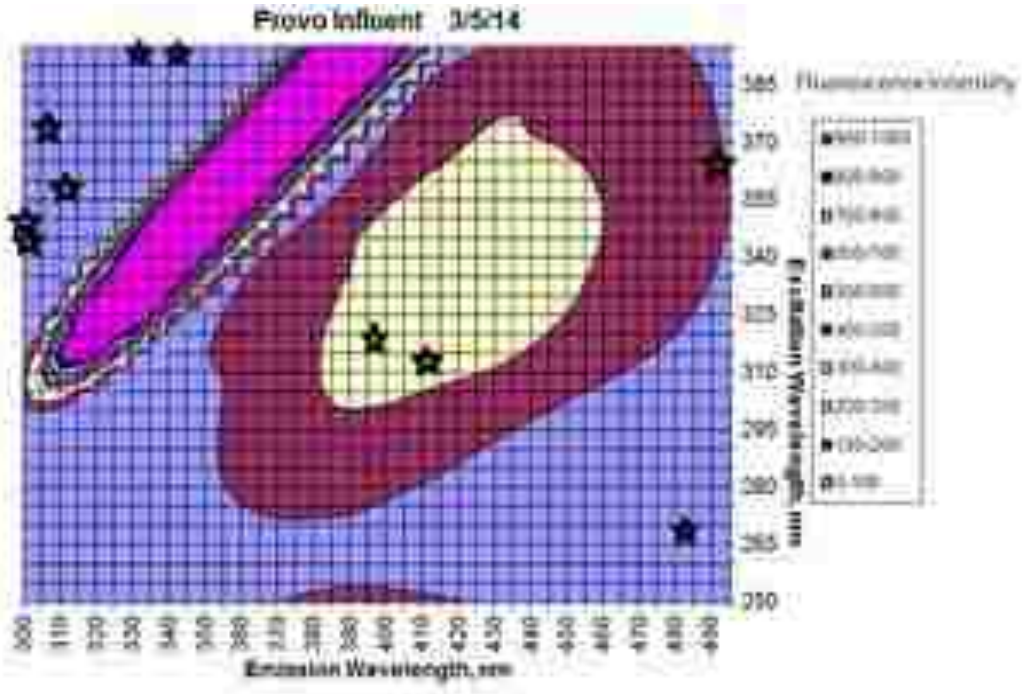


Figure 5-2 Scattered Points in Provo Influent EEM

Orem influent BOD and EEM data result produced a better R-squared of 0.98 which was bigger than the Provo BOD and EEM data result of 0.95. Comparing the correlation of EEM and BOD_5 as well as EEM and COD, it is hard to say which one correlated better. All the EEMs results used for our regression analysis were in their diluted state. The results of majority of the correlations were done using diluted EEM results and undiluted BOD_5 and COD values. Fig 4-1 gives a very good correlation of our Provo predicted BOD_5 and our actual BOD_5 values. The model was applied to the Provo EEM data which was in a diluted state.

Fig 4-2 is a display of a correlation of Orem Predicted BOD_5 values and Actual BOD_5 values using the Provo model. Before any correlation was made, the Orem EEM was initially converted from a diluted value to its original undiluted nature by multiplying the diluted values

by (5/6). The correlation derived from this model was poor as shown in Fig 4-2 and the R squared obtained was 0.02. This shows that each individual site is unique and requires its individual site model to produce the best correlation possible.

Table 5-1 P- Value and Interpretation

P- value	Interpretation
0.001 0.01 0.02 0.03	Highly Significant
0.04 0.049	Significant
0.05	Border line
0.051 0.06	On the edge of Significance
0.07 0.08 0.09 0.099 >or =0.1	Highly suggestive

A code was written in R language for the Regression analysis by Dr. William Christensen of the Department of Statistics at Brigham Young University. The code can be found in the appendix section.

5.1 **Consideration**

One factor to consider is the dilution of the wastewater samples. The samples were diluted according to the ratios found in Table 3-1 so that the measured intensity would not exceed the measurable limit of a 1000. Care was taken during the dilution process to avoid

contamination of samples. Nevertheless, this process of dilution using Table 3-1 ratio could result in EEMs not truly depicting the wastewater fluorescence.

Another factor of which much attention has to be given is the sampling frequency. A change in sample frequency could alter the entire result of an EEM. Many EEM data results for this research had to be rejected due to inconsistency and variation in scan speed input and emission wavelength of 300-500 nm/min as stated in page 30.

Accurate BOD and COD data was also another factor to consider. Some of the BOD and COD data received had inaccurate values recorded on certain days and or had no values recorded at all. These recordings were rejected and not included in the statistical process.

6 CONCLUSION

The purpose of this research was to determine if fluorescence spectroscopy EEM with BOD_5 and COD actual-values can be correlated using a regression analysis. This research validates the multiple regression technique as a viable and reliable method in which researchers and other wastewater stakeholders can use in determining BOD_5 and COD values within a shorter time frame of between 20-30 minutes.

Models generated for Provo and Orem individual sites can be used by Provo and Orem waste treatment plant personnels respectively in predicting their BOD_5 and COD results. Each individual site needs to have its individual model in other for a high correlation to occur.

The steps to follow in deriving BOD and COD are to dilute samples and run them in a fluoro-spectrometer. After EEM results are derived, each individual model is applied to its respective EEM to give a predictive BOD_5 or COD value. I will recommend that these models should be used side by side whiles the actual BOD_5 and COD test are still being done. This should be carried out for at least a month. If predicted values from the model match well with actual values from the test after the one month time frame then the model can be used solely to predict BOD_5 and COD without the need of carrying out the actual BOD and COD tests.

Regression analysis is a statistical house hold name and it is used in solving many correlational issues. This research finding can be improved if more data is taken all year round at

different seasons to see whether seasonal changes affect the model. Correlation of EEMs with BOD_5 as well as COD offers certain advantages over the time consuming 5 day BOD test as well as 2 hour COD test. Some of these advantages are that it is a less time consuming test and it is not cumbersome.

REFERENCES

Baker, A, and M. Curry. "Fluorescence of leachates from three contrasting landfills." *Water Resources*, 2004: 2605-2613.

Baker, A, and R. Inverarity. "Protein-like fluorescence intensity as a possible tool for determining river water quality." *Hydrology Process*, 2004: 2927-2945.

Baker, A. ""Fluorescence Excitation-Emission Matrix Characterization of Some Sewage-." *Environmental Science and Technology*, 2001: 948-953.

Baker, A. "Fluorescence properties of some farm wastes: implications for water quality monitoring." *Water Resources*, 2002: 189-195.

Baker, A. "Protein-like fluorescence intensity as a possible tool for determining river water quality." *Hydrological Processes*, 2004: 2927-2945.

Blake, DB. *Source Apportionment of Wastewater Using Bayesian Analysis Of Fluorescence Spectroscopy.* Provo, July 2014.

Cammack, WKL, J Kalf, YT Prairie, and EM. Smith. "Fluorescent dissolved organic matter in lakes: relationship with heterotrophic metabolism." *Limnol Oceanography* 49, no. 6 (2004): 2034-45.

Chee, GJ, Y Nomura, and I. Karube. "Biosensor for the estimation of low biochemical oxygen demand." *Analytical Chemistry*, 1999: 185-191.

Chen, W, P Westerhoff, J,A Leenheer, and K. Booksh. "Fluorescence excitation–emission matrix regional integration to quantify spectra for dissolved organic matter." *Environmental Sci Tech*, 2003: 5701-5710.

Coble, PG. "Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy." March 1996: 325-46.

Csaba, P, and J. Csaba. *Water Resources Management and Water Quality Protection.* Environmental Protection, 2011.

Eaton, AD, Clesceri LS, Greenberg, and AE. "Standard methods for the examination of water and wastewater,." *American Public Health Association*, 1995.

Elliott, S, JR Lead, and A. Baker. "Characterisation of the fluorescence from freshwater, planktonic bacteria." *Water Resources*, no. 10 (2006): 2075-83.

Farabegoli, G, C Hellinga, JJ Heijnen, and MCM. Van Loosdrecht. "Study on the use of NADH fluorescence measurements for monitoring wastewater treatment systems." *Water Resources*, no. 37 (2003): 2732-8.

Ferrell, CE. *Fluorescence Spectrophotometry and the mass balance: exitation-emission matrices.* Provo, April 2009.

Finazzi-Agro, A, G Rotilio, L Avigliano, P Guerrieri, V Boffi, and B. Mondovi. "Environment of copper in Pseudomonas fluorescens azurin: fluorometric approach." *Biochemistry*, 1970: 2009-2014.

Galinha, CF, G Carvalho, CAM Portugal, G Guglielmi, MAM Reis, and JG. Crespo. "Two-Dimensional Fluorescence as a fingerprinting tool for monitoring wastewater treatment systems." Feb 2011: 985-991.

Ghervase, L, E Carstea, G Pavelescu, and EB and A. Daskalova. ""Fluorescence evaluation of anthropogenic influence on rivers crossing Sofia."." *Romanian Reports in Physics*, 2010: 193-201.

Gray, PC, IR Shokair, SE Rosenthal, GC Tisone, JS Wagner, and LD. Rigdon. "Distinguishability of biologicalmaterial by use of ultraviolet multispectral fluorescence." *Applied Opt*, no. 25 (1998): 6037-41.

Guo, W, J Xu, J Wang, Y Wen, and J, Zhuo Y. Yan. "Characterization of dissolved organic matter in urban sewage using excitation emission matrix fluorescence spectroscopy and parallel factor analysis." *Journal of Environmental Sciences*, 2010: 1728-1734.

Hach, CC, RL Klein, and Charles, R. Gibbs. *Introduction to.* Hach company, 1997.

Hall, GJ, KE Clow, and JE. Kenny. "Estuarial fingerprinting through multidimensional fluorescence and multivariate analysis."." *Environmental Science & Technology*, 2005: 7560-7567.

Hornak, JK. *Molecular Fluorescence Spectroscopy.* 1999. http://jpkc.henu.edu.cn/sybpfx/scha/fluor.htm (accessed 6 7, 2015).

Huang, S, AA Heikal, and WW. Web. "Two-photon fluorescence spectroscopy and microscopy of NAD (P) H and flavoprotein." *Biophysical journal*, 2002: 2811-2825.

Hudson, N, A Baker, D Ward, DM, Brunsdon, C Reynolds, C Carliell-Marquet, and S. Browning. "Can fluorescence spectrometry be used as a surrogate for the Biochemical Oxygen Demand (BOD) test in water qualityassessment? An example from South West England." *Science for the Total Environment*, Dec 4, 2008: 150-157.

Hudson, N, A Baker, Reynolds, and D. "fluorescence analysis of dissolved organic matter in natural, waste and polluted waters." April 2007: 631-649.

JR, Lakowicz, and ed. "Topics in Fluorescence Spectroscopy." 1994.

Karube, I, T Matsunga, S Mitsuda, and S. Suzuki. "Microbial electrode BOD sensors." *Biotechnol. Bioeng*, 1977: 1535-1547.

Kowalczuk, P, M. J Durako, H Young, A Kahn, W. J Gonsior, and M. Cooper. ""Characterization of dissolved organic matter fluorescence in the South Atlantic Bight." *Marine Chemistry*, 2009: 182-196.

Kulys, J, and K. Kadziauskiene. "Yeast BOD sensor." *Biotechnology*, 1980: 221-226.

Leblanc, L, and E. Dufour. "Monitoring the identity of bacteria using their intrinsic fluorescence." *FEMS Microbiology*, no. 211 (2002): 147-53.

Li, JK, and AE. Humphrey. "Use of fluorometry formonitoring and control." *Biotechnol Bioengineering*, 1991: 37:1043-1049.

Lochmueller, CH, and SS. Saavedra. ""Conformational changes in a soil fulvic acid." *Analytical Chemistry*, 1986: 1978-1981.

Marty, JL, D Olive, and Y. Asano. "Measurement of BOD: correlation between 5-day BOD and commercial BOD biosensor values." *Environ. Technol*, 1997: 333-337.

Mehta, Akul. "Animation for the Principle of Fluorescence and UV-Visible Absorbance." *ANALYTICAL CHEMISTRY*, 2013.

Mocz, Gabor. *The Use of Fluorescence in Research into Protein Structures.* n.d. http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk/projects/gmocz/fluor.htm (accessed 4 11, 2015).

Mopper, K, RF Chen, Z Feng, and SB. Bentjen. "Effects of cross-flow filtration on the absorption and fluorescence propertieds of seawater." *Marine Chemistry*, 1996: 53-74.

Murphy, K. "Measurement of Dissolved Organic Matter Fluorescence in Aquatic Environments: An Interlaboratory Comparison." *Environmental Science and Technology*, 2010: 9405-9412.

Parlanti, E, K Wörz, L Geoffroy, and M. Lamotte. "Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs." *Organic Geochemistry*, no. 12 (2000): 1765-81.

Pasco, N, K Baronian, C Jeffries, and J. Hay. "Biochemical mediator demand–a novel rapid alternative for measuring biochemical oxygen demand." *Applied Microbiology and Biotechnology*, 2000: 613-618.

Patel-Sorrentino, N, S Mounier, and JY. Benaim. "Excitation-emission fluorescence matrix to study pH influence on organic matter fluorescence." *Water Research*, 2002: 2571-2581.

Princz, P, and J. Olah. *BOD measurement of non-toxic wastewaters with an improved microbial probe.* Proceedings of the 5th IAWPRC workshop held in Yokohama and Kyoto, Yokohama and Kyoto: Pergamon Press, 1990.

Reiss, M, A Heibges, J Metzger, and W. Hartmeier. "Determination of BOD-values of starch-containing waste water by a BOD-biosensor." *Biosens Bioelectron*, 1998: 1083-1090.

Reshetilov, AN, et al. "Effects of high oxygen concentrations on microbial biosensor signals. Hyperoxygenation by means of perfluorodecalin." *Biosens Bioelectron*, 1998: 795-799.

Reynold, DM, and SR. Ahmad. "Rapid and direct determination of wastewater BOD values using a fluorescence technique." *Water Resources*, 1997: 2012-2018.

Reynolds, DM. "The differentiation of biodegradable and non-biodegradable dissolved organic matter in wastewaters using fluorescence spectroscopy." *J Chem Technol Biotechnol*, 2002: 965-972.

Riedel, K, M Lehmann, K Tag, R Renneberg, and G. Kunze. "Arxula adeninivorans based sensor for the estimation of BOD." *Anal Lett*, 1998: 1-12.

Riedel, lK, R Renneberg, M Kühn, and F. Scheller. "Yeast BOD sensor ation of biochemical oxygen demand using microbial sensors." *Appl. Microbiol. Biotechnol*, 1988: 316-318.

Riedel, R, R Renneberg, M Kohn, and F. Scheller. "A fast estimation of biochemical oxygen demand using microbial sensors." *Appl Microbiol Biotechnol*, 1988: 316±318.

Saadi, I, M Borisover, R Armon, and Y. Laor. "Monitoring of effluent DOM biodegradation using fluorescence, UV and DOC measurements." *Chemosphere*, no. 63 (2006): 530–539.

Sangeetha, S, G Sugandhi, M Murugesan, VM Madhav, and Rajasekar R, Rajasekar S, Jeyakumar D, Rao GP. Berchmans S. "Torulopsis candida based sensor for the estimation of biochemical oxygen demand and its evaluation." *Electroanalysis*, 1996: 698-701.

Seaver, M, DC Roselle, JF Pinto, and JD. Eversole. "Absolute emission spectra from Bacillus subtilis and Escherichia coli vegetative cells in solution." *Appl Opt*, no. 22 (1998).

Slama, M, C Zaborosch, D Wienke, and F. Spener. "Simultaneous mixture analysis using a dynamic microbial sensor combined with chemometrics." *Anal Chem*, 1996: 3845±3850.

Smith, CB, JE Anderson, and SR. Webb. "Detection of Bacillus endospores using total luminescence spectroscopy." *Spectrochim Acta, Part A*, no. 11 (2004): 2517-21.

Stedmon, CA, S Markager, and R. Bro. "Source apportionment of pollutants and flows of combined sewer wastewater." *Water Research*, 2003: 1989-1998.

Tanaka, H, Nakamura E, Minamiyama Y, and Toyoda. T. "BOD biosensor for secondary effluent from wastewater treatment." *Water Sci Technol*, 1994: 215-227.

Thompson, RB, KF Joan, and RL. Joseph. "Phase fluorometry using a continuously modulated laser diode." *Analytical Chemistry*, 64: 2002.

Traving, SJ, L Riemann, and C. Stedmon. "Traits for bacterial carbon turnover in the marine environment: chemical complexity meets bacterial diversity." *Center for Ocean Life*, 2015.

Tutorvista. *TutorVista.* n.d. http://chemistry.tutorvista.com/analytical-chemistry/molecular-spectroscopy.html (accessed April 07, 2015).

Vodacek, A, and WD. Philipot. "Environmental effects on laser-induced fluorescence spectra of natural waters." *Remote Sens Environ*, 1987: 83-95.

Wieder, Irwin. US Patent 4058732. 11 15, 1977.

Yan, Y, H Li, and ML. Myrick. "Fluorescence Fingerprint of Waters: Excitation Emission Matrix Spectroscopy as a Tracking Tool." *Applied Spectroscopy*, 2000: 1539-1542.

.

APPENDIX A. DATA

A.1 Biochemical Oxygen Demand Data

Table A-1 Results of Provo ABI BOD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	320,365	6.40E-06	29.155	0.078
2	350,370	4.11E-03	0.84102	0.121
3	315,350	1.94E-03	-14.1827	0.613
4	300,350	9.20E-02	4.342	0.753
5	320,355	2.65E-02	-13.648	0.842
6	320,390	2.75E-05	-20.879	0.906
7	300,390	2.70E-04	-13.8075	0.934
8	300,250	3.23E-04	0.80165	0.948
9	310,390	2.39E-05	31.4634	0.9599
10	335,350	1.90E-01	-0.01363	0.9655
Intercept		1.43E-04	50.58662	

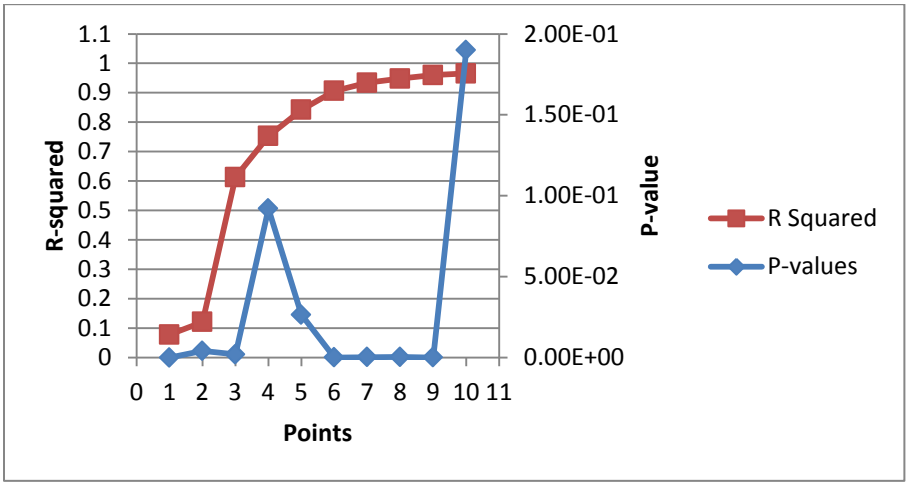


Figure A- 1 Representation of Provo ABI BOD P-values, R-squared and Points

Table A-2 Results of Provo FSBE BOD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	430,345	1.22E-04	0.28877	0.011
2	485,395	4.30E-05	-0.48843	0.565
3	455,395	6.96E-05	0.47306	0.567
4	425,345	1.42E-04	-0.29375	0.705
5	430,380	4.35E-04	-0.4057	0.834
6	435,330	2.63E-02	-0.2368	0.835
7	340,310	4.62E-04	-0.13085	0.849
8	455,325	9.97E-05	0.6148	0.919
9	435,330	1.22E-01	0.1469	0.946
10	380,360	1.36E-01	0.0241	0.9558
Intercept		2.28E-01	-2.4188	

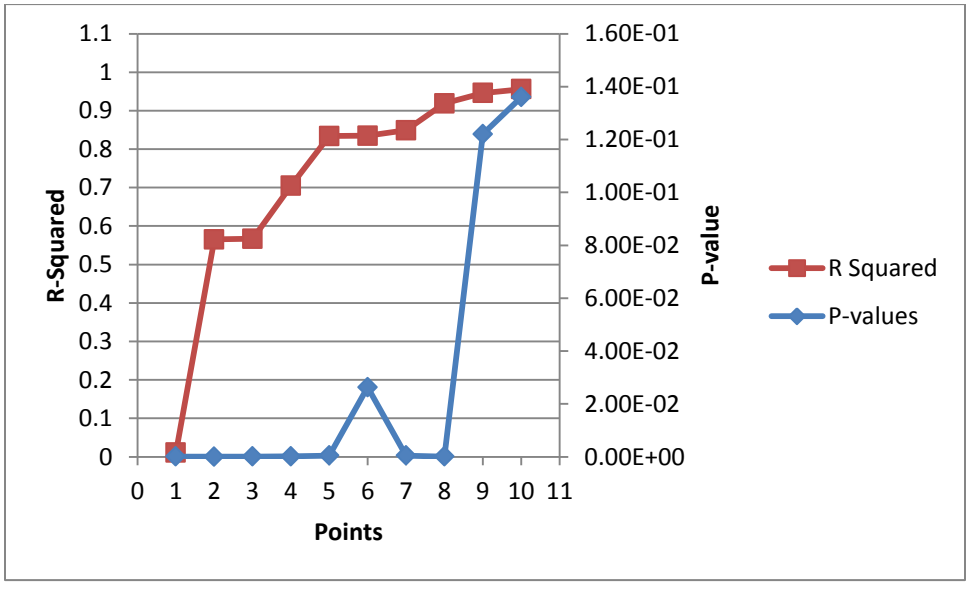


Figure A-2 Representation of Provo FSBE BOD P-values, R-squared and Points

Table A- 3Provo Effluent BOD data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	305,370	2.70E-03	-0.7575	0.131
2	410,380	4.29E-02	0.07407	0.275
3	415,375	6.34E-02	-0.0394	0.566
4	495,375	1.51E-03	0.2208	0.651
5	430,300	4.70E-04	-0.1288	0.737
6	405,375	1.18E-01	-0.02998	0.87999
7	335,395	4.90E-03	0.5846	0.908
8	430,305	1.60E-03	0.07098	0.948
9	410,280	6.03E-06	0.0577	0.953
10	475,270	2.74E-03	-0.18863	0.9618
Intercept		1.26E-03	1.1556	

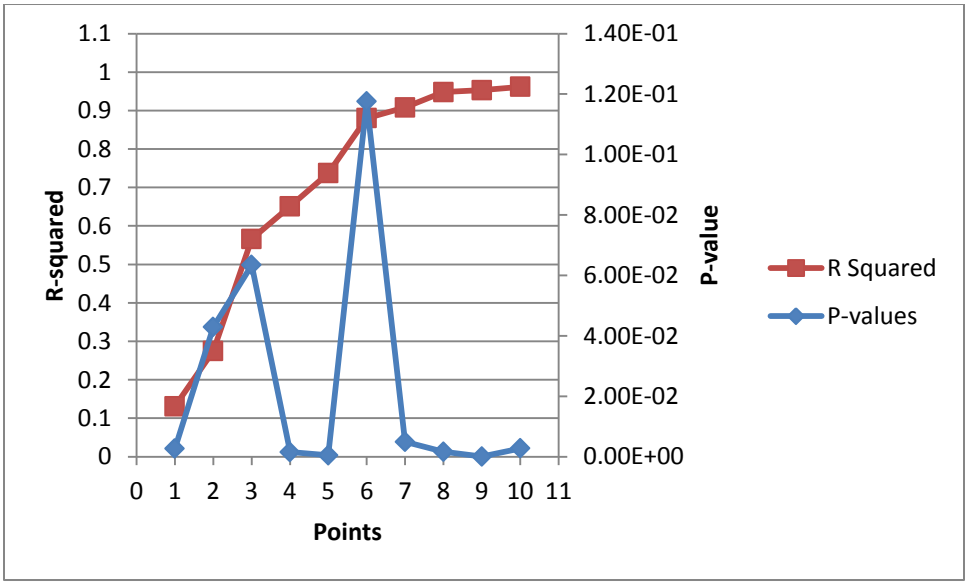


Figure A- 3 Provo Effluent BOD P-value, R-squared and Points

Table A- 4Results of Orem Effluent BOD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	310,375	5.29E-02	-8.046	0.0635
2	365,340	4.17E-03	0.29131	0.267
3	480,265	5.73E-03	2.489	0.3888
4	380,350	4.40E-03	-0.2179	0.676
5	355,395	1.55E-02	-6.614	0.873
6	315,395	4.61E-03	13.16084	0.932
7	330,395	2.06E-02	8.7353	0.94199
8	495,270	5.37E-06	-7.5851	0.9627
9	380,270	1.50E-03	0.4154	0.9708
10	485,255	2.07E-04	1.8569	0.9796
Intercept		1.02E-07	33.03	

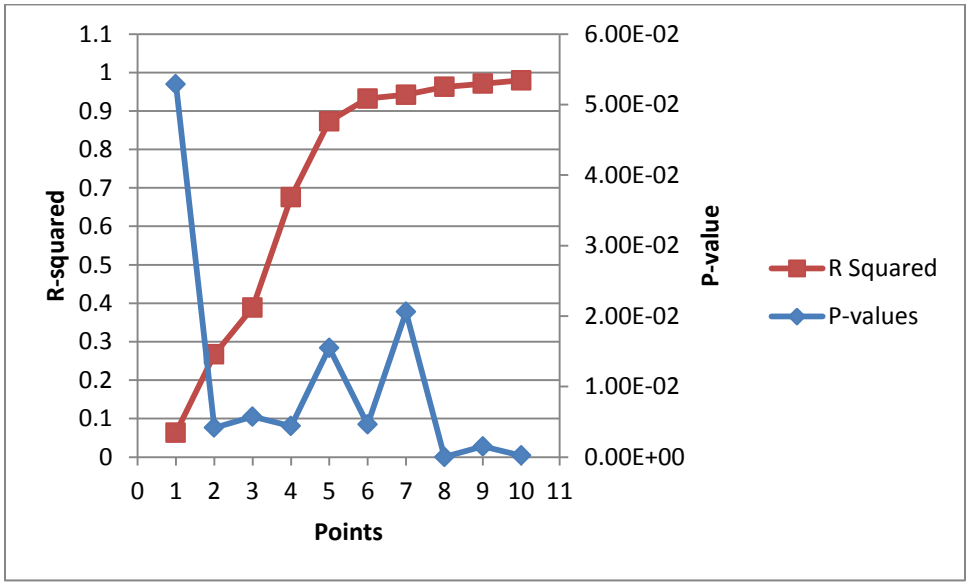


Figure A- 4 Representation of Orem Effluent BOD P-values, R-squared and Points

Table A- 5 Results of Orem PE BOD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	495,355	4.12E-04	-42.915	0.0025
2	490,355	4.22E-03	28.227	0.272
3	485,380	1.17E-03	12.701	0.482
4	490,290	5.40E-02	-5.831	0.7586
5	480,360	3.20E-04	-41.035	0.769996
6	490,365	2.40E-03	-33.424	0.838
7	455,370	1.90E-04	9.97	0.885
8	300,340	1.10E-04	10.026	0.922
9	485,365	5.68E-05	53.774	0.95766
10	455,340	1.03E-04	7.209	0.9703
Intercept		5.60E-03	165.509	

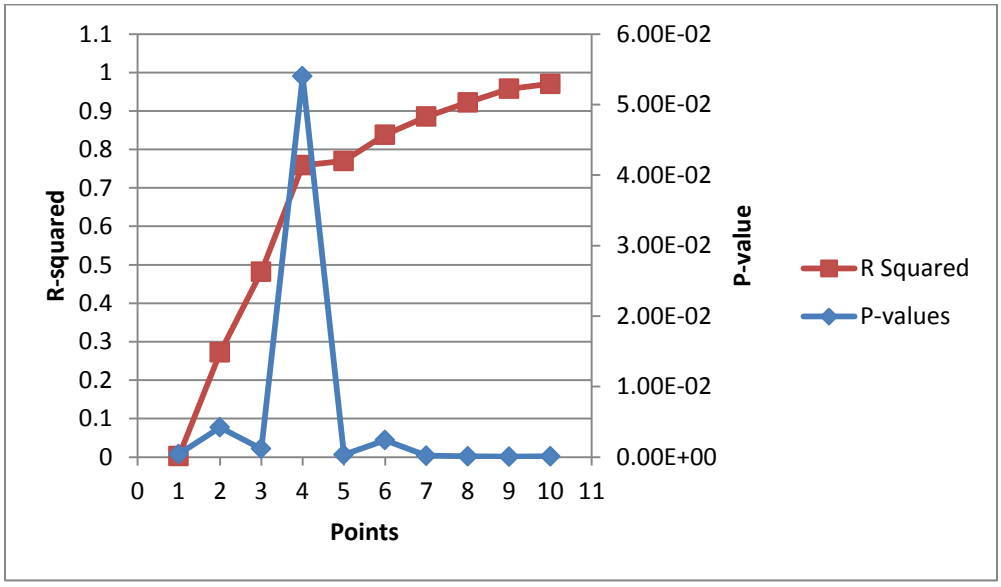


Figure A- 5 Representation of Orem Effluent BOD P-values, R-squared and Points

A.2 Chemical Oxygen Demand Data

Table A- 6 Results of Orem PE COD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R Squared
1	390,325	1.02E-05	0.9169	0.0119
2	460,275	1.10E-03	-23.575	0.179
3	480,265	1.30E-04	25.536	0.2245
4	310,275	8.97E-05	30.628	0.383
5	450,275	7.23E-03	-18.107	0.6532
6	440,260	4.95E-05	17.8049	0.7577
7	335,360	3.00E-03	7.3778	0.847
8	300,255	5.47E-05	-85.0906	0.906
9	300,370	6.86E-04	-26.1341	0.963
10	485,380	3.47E-05	-0.8982	0.984
Intercept		2.62E-06	245.329	

Table A- 7 Results of Orem Effluent COD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	485,265	3.34E-06	-72.1558	0.002
2	475,280	3.59E-04	45.0283	0.546
3	450,270	1.79E-06	68.6163	0.647
4	300,255	3.12E-04	-171.69	0.6499
5	310,270	2.10E-06	116.3445	0.855
6	475,260	2.80E-06	-68.998	0.932
7	450,250	3.20E-03	-6.423	0.957
8	305,260	1.57E-04	-65.6263	0.974
9	485,275	3.48E-03	33.1808	0.9828
10	310,365	3.50E-07	-11.8	0.9936
Intercept		9.00E-04	152.164	

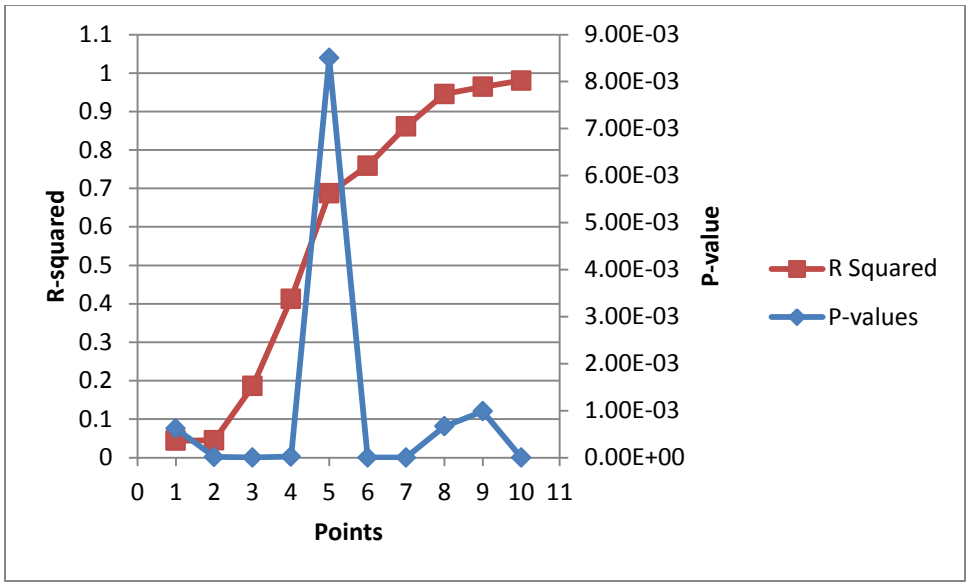


Figure A- 6 Representation of Orem Effluent COD P-values, R-squared and Points

Table A- 8 Results of Provo ABI COD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	355,395	4.76E-03	18.951	0.047
2	360,385	5.94E-03	-11.13	0.351
3	325,360	1.95E-03	-16.114	0.443
4	365,395	4.27E-01	-2.754	0.664
5	300,365	1.50E-01	5.199	0.686
6	335,320	1.54E-02	0.212	0.843
7	485,395	4.30E-02	-3.064	0.918
8	465,370	1.55E-01	0.447	0.93
9	340,365	3.12E-04	10.023	0.938
10	480,390	1.94E-01	2.124	0.941
Intercept		4.90E-06	-0.825	

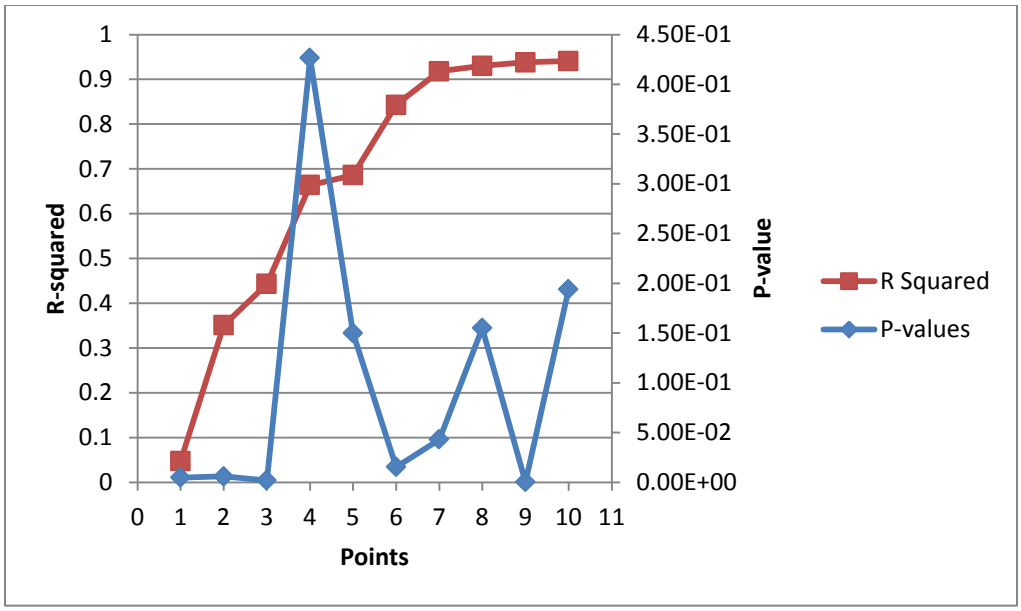


Figure A- 7 Representation of Provo ABI COD P-values, R-squared and Points

Table A- 9 Results of Provo FSBE COD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	485,315	2.76E-04	1.438	0.0597
2	395,320	2.17E-04	-0.4458	0.241
3	415,280	7.12E-05	2.222	0.258
4	385,285	2.90E-03	1.106	0.813
5	430,395	1.39E-02	0.894	0.85988
6	455,360	2.88E-02	-0.0887	0.8996
7	340,260	4.10E-05	-2.469	0.907
8	440,395	3.76E-02	-0.5784	0.925
9	410,290	8.70E-03	-1.661	0.941
10	400,290	1.95E-01	-0.7113	0.9489
Intercept		1.07E-04	12.0202	

Table A- 10 Results of Provo Effluent COD Data

Number Of Points Used	Best Cells/Sites	P-values	Coefficients	R-squared
1	480,270	1.40E-03	-0.168723	0.202
2	375,325	1.54E-04	-0.02537	0.442
3	335,395	2.08E-04	0.77045	0.532
4	415,375	5.00E-04	0.0346	0.791
5	435,340	2.67E-05	0.01323	0.819
6	430,300	4.68E-04	-0.027753	0.912
7	345,375	1.12E-04	-0.3661	0.915
8	485,275	2.10E-03	0.174	0.9232
9	420,375	1.60E-03	-0.02989	0.9447
10	425,295	1.78E-04	0.0404	0.9756
Intercept		1.50E-03	0.9028	

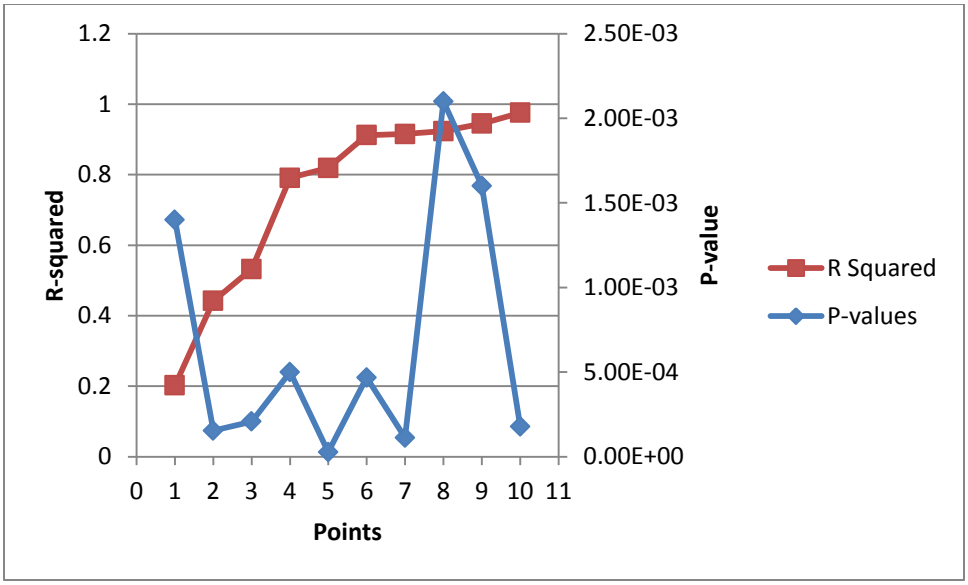
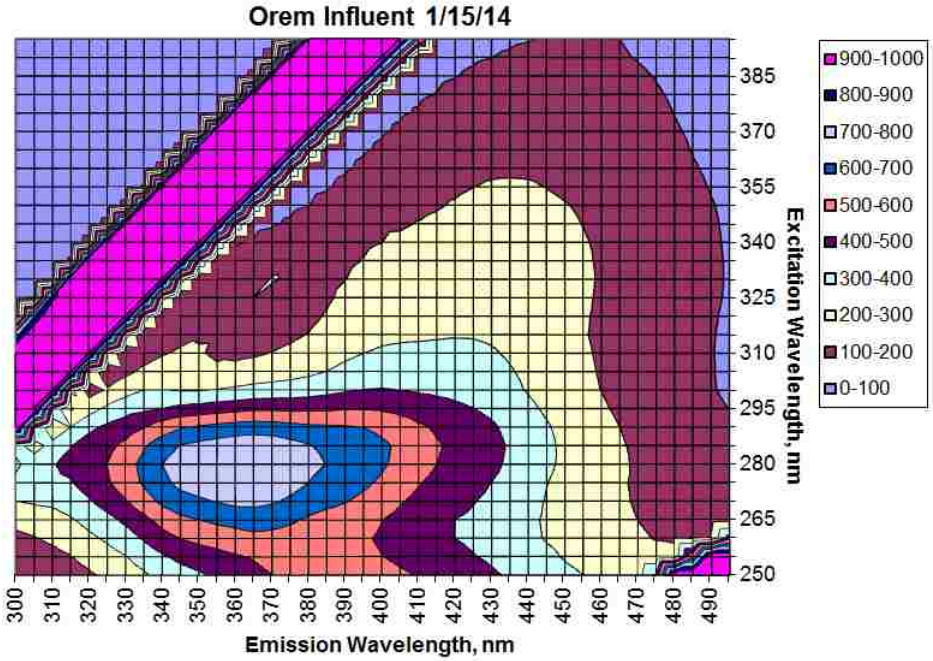
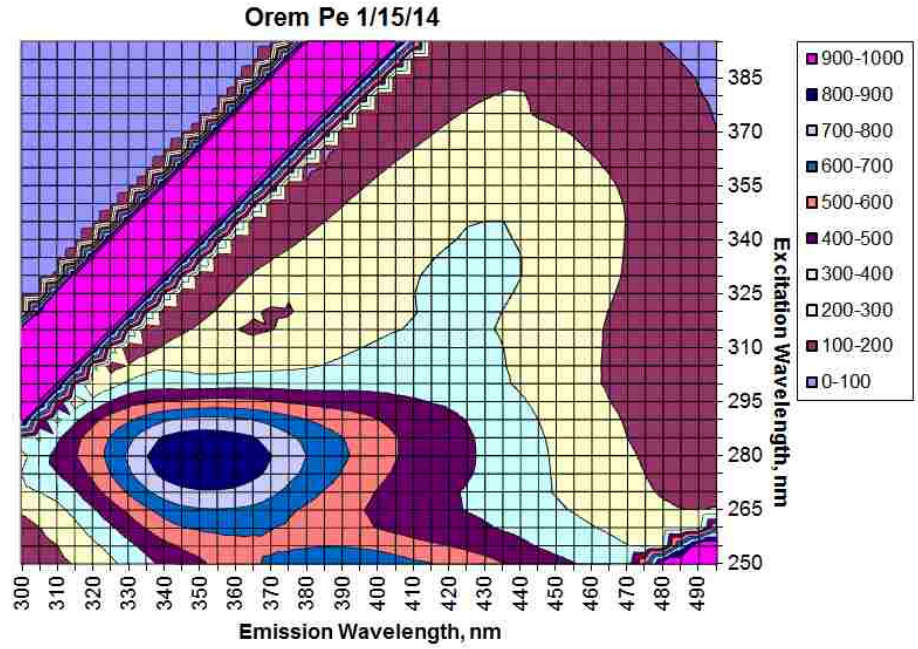
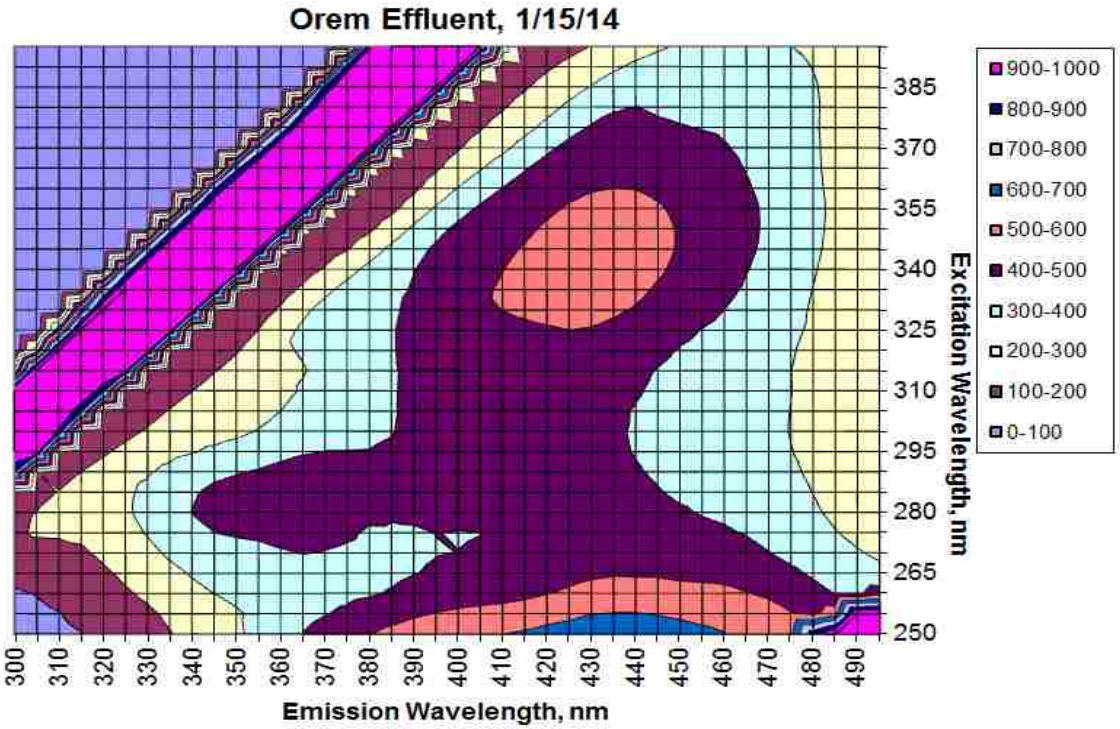


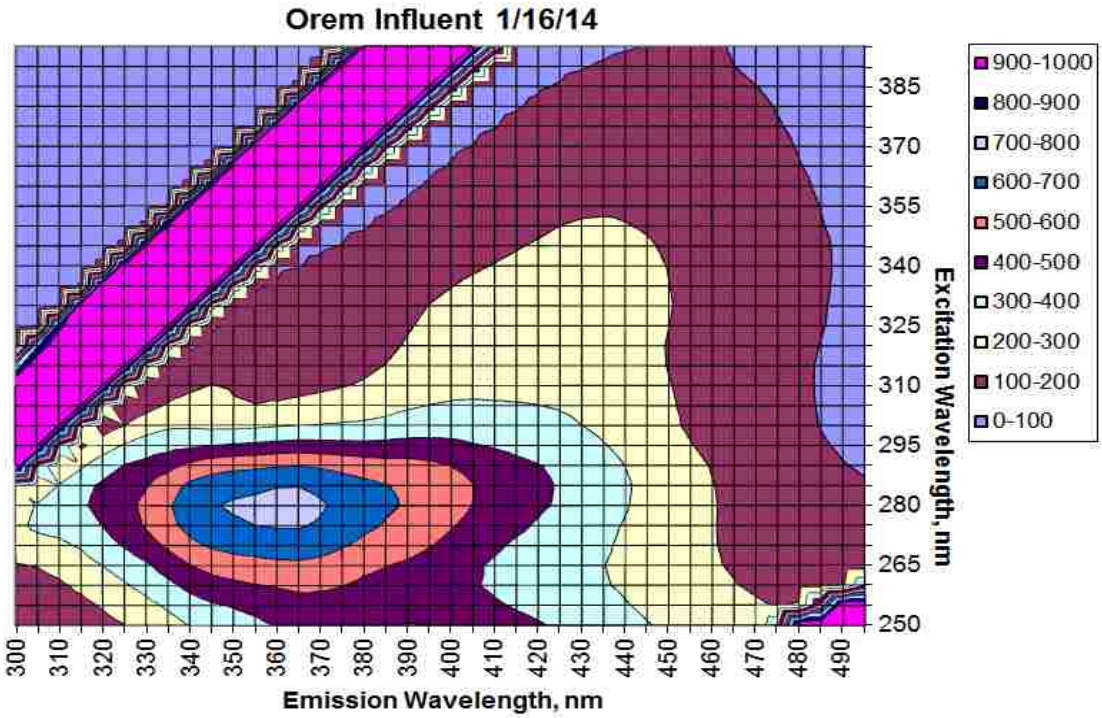
Figure A- 9 Representation of Provo Effluent COD P-values, R-squared and Points

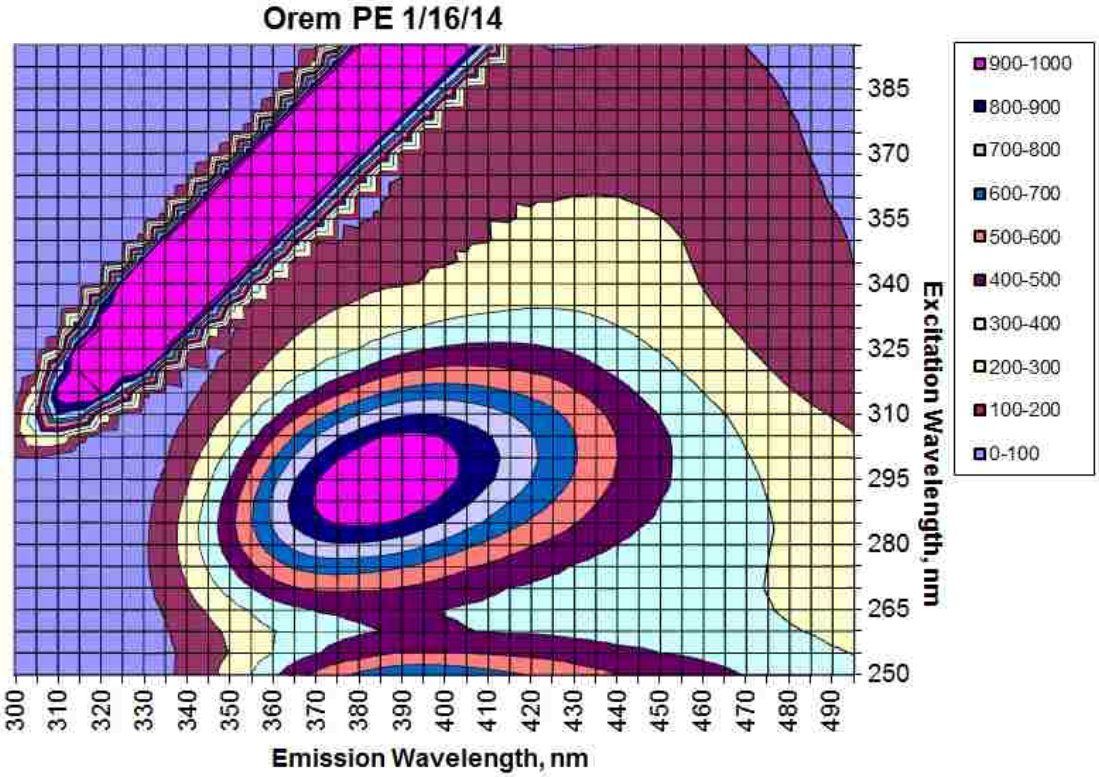
A.3 Orem Emission Excitation Matrix Data

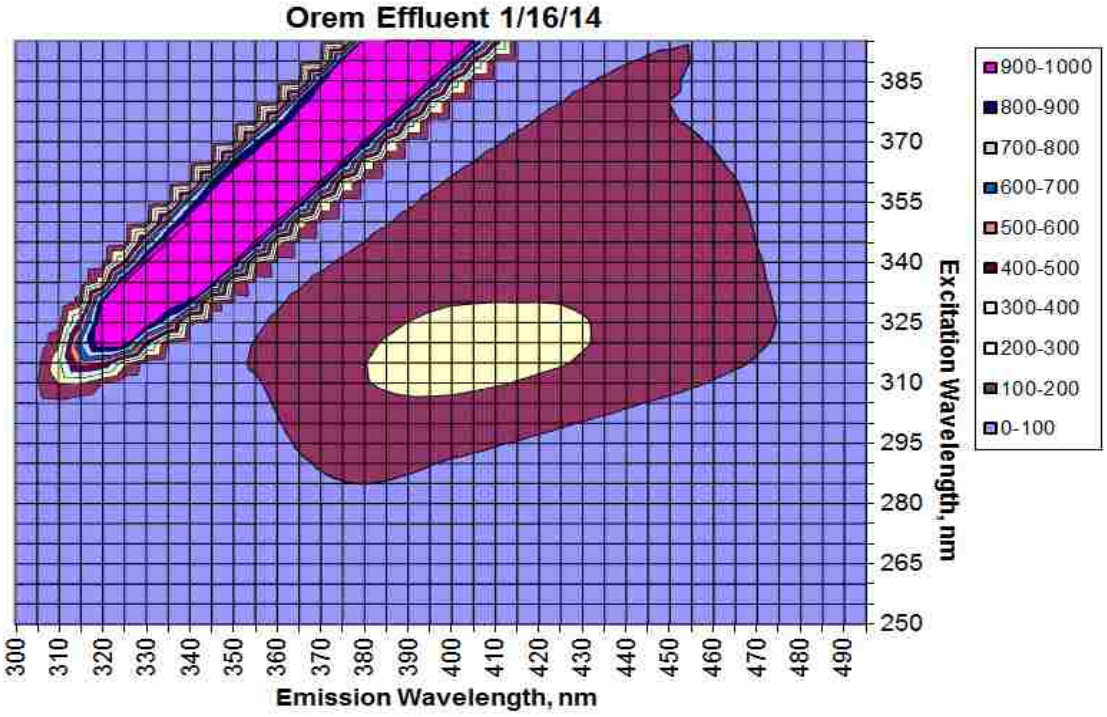


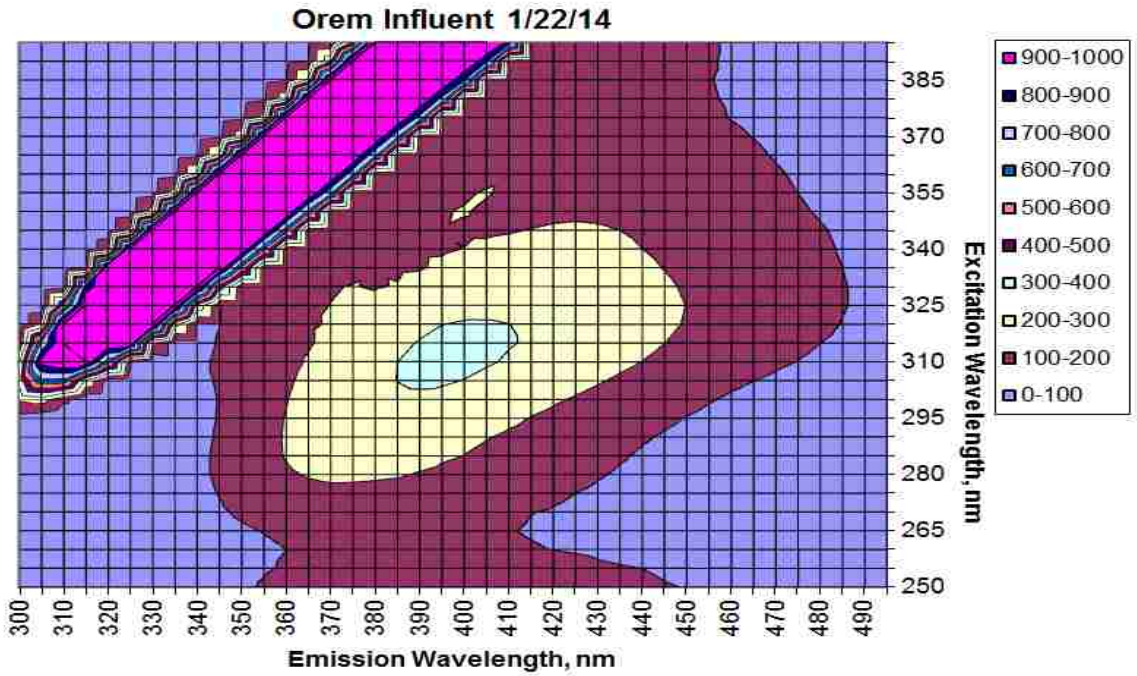


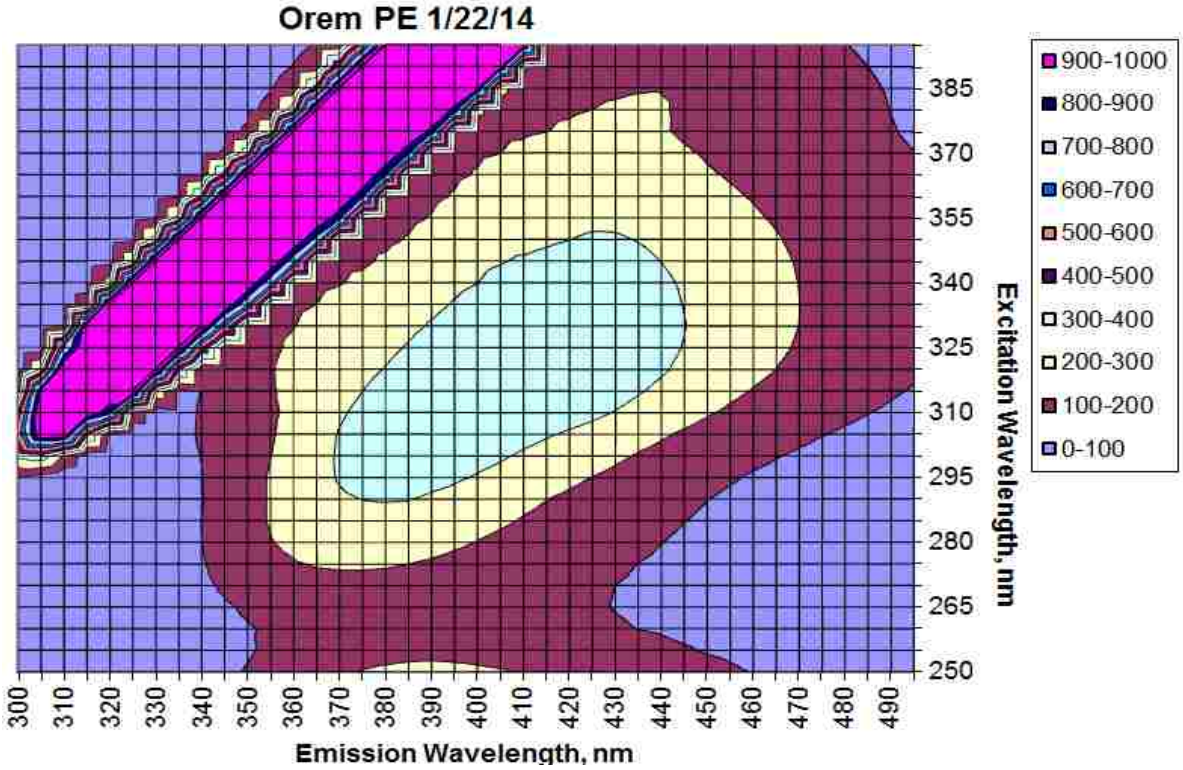


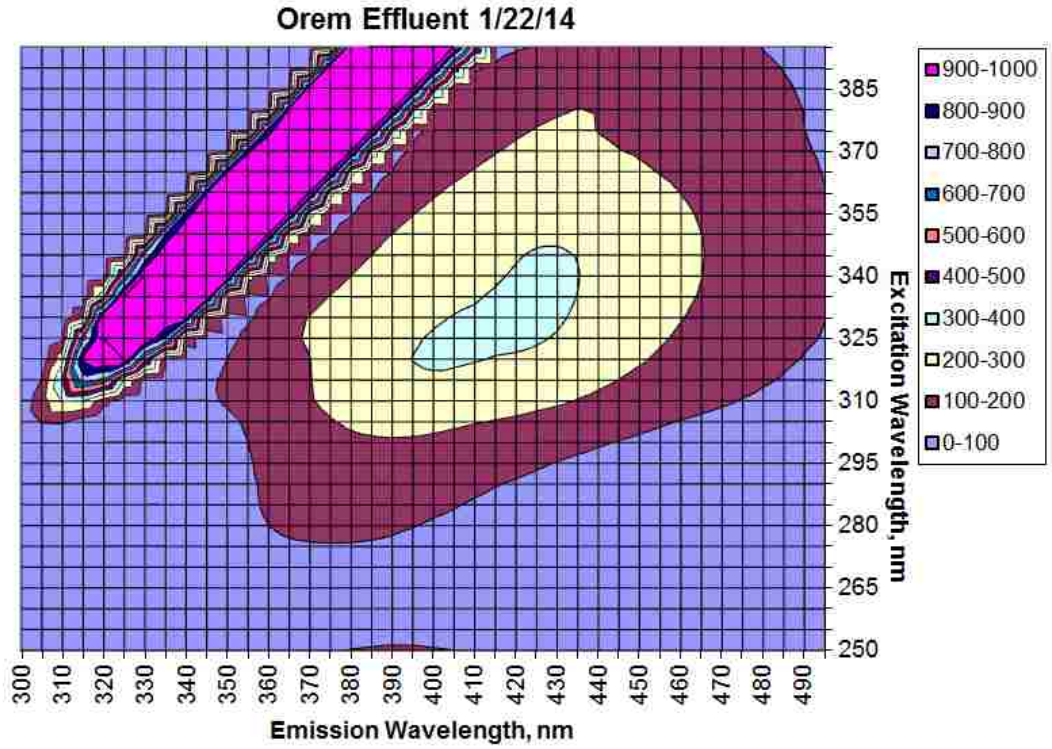


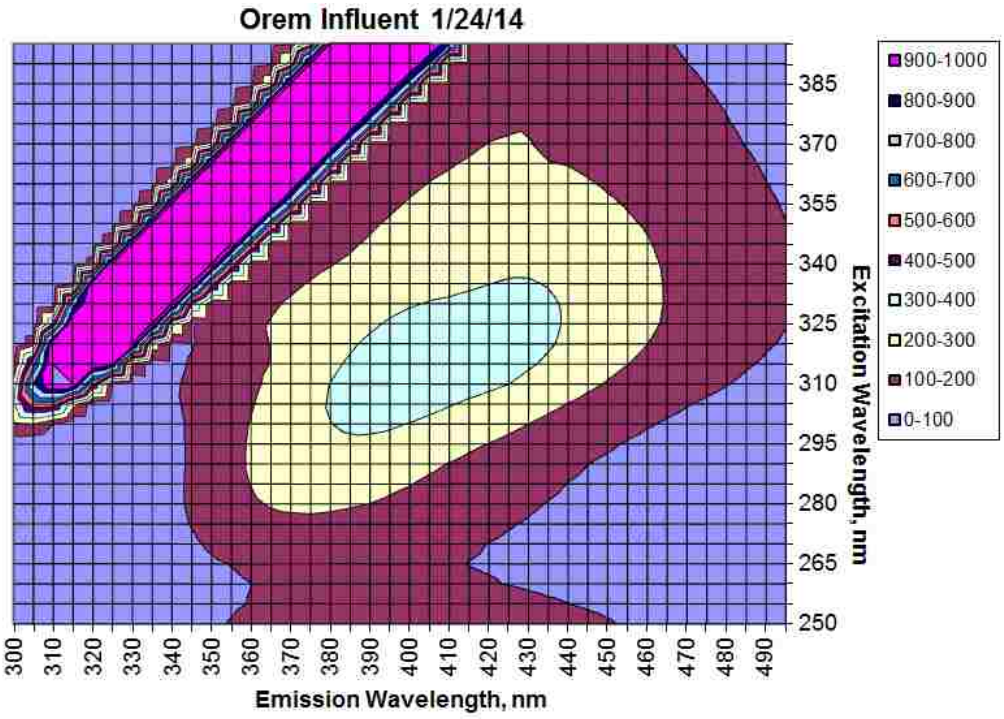


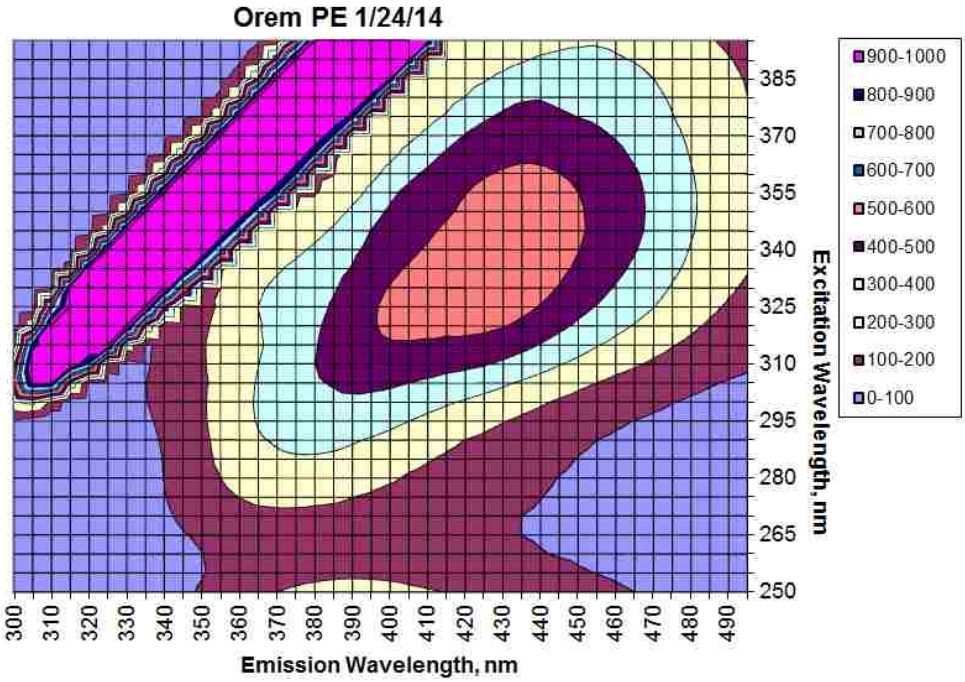


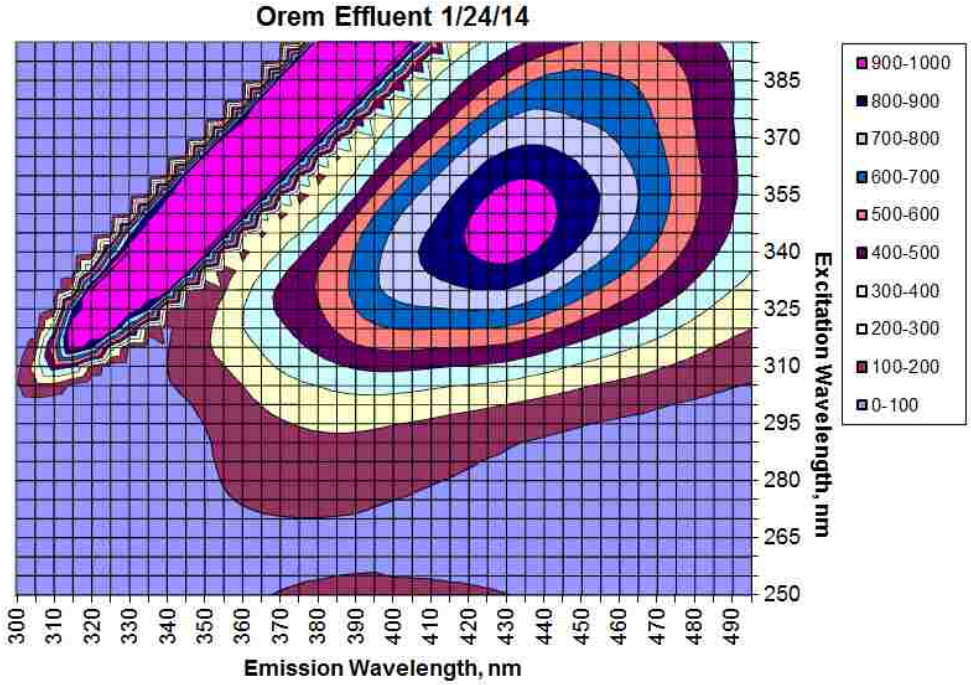


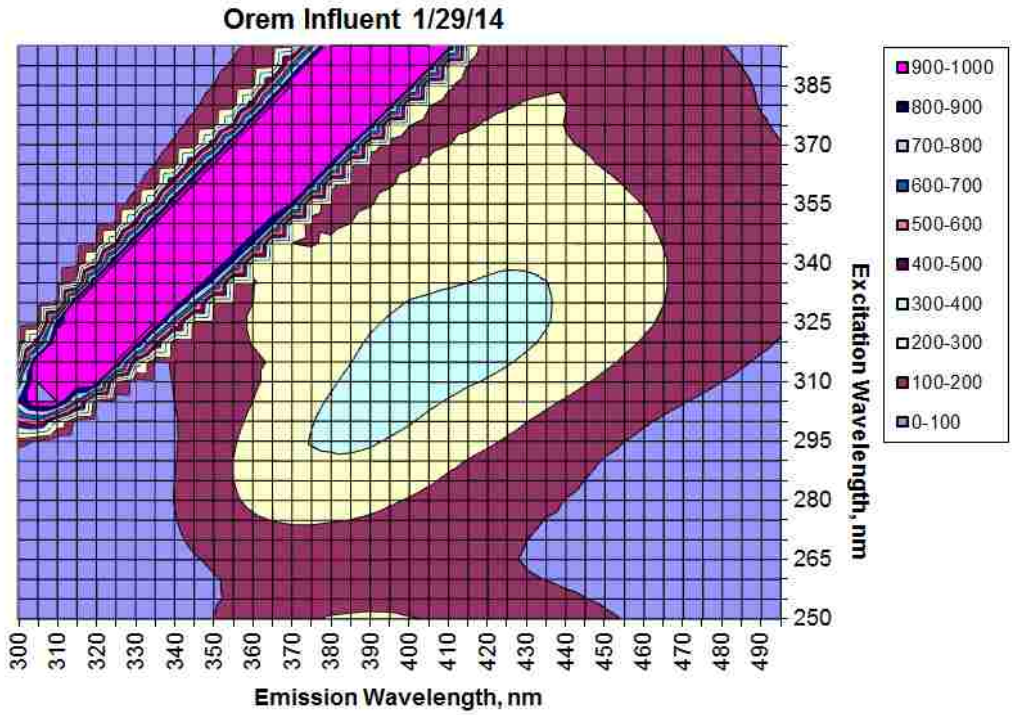


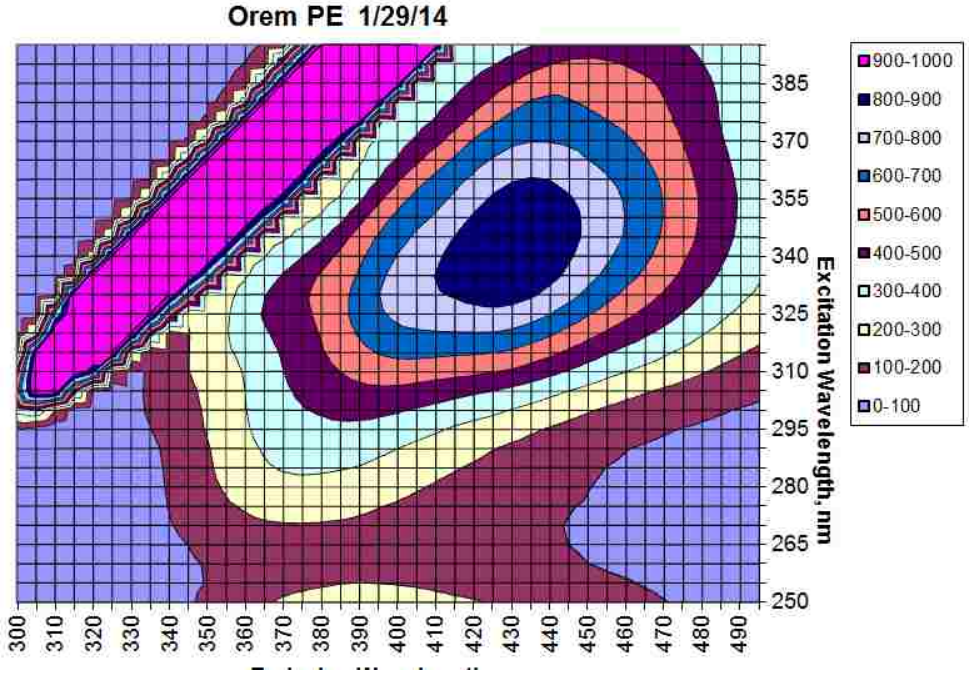


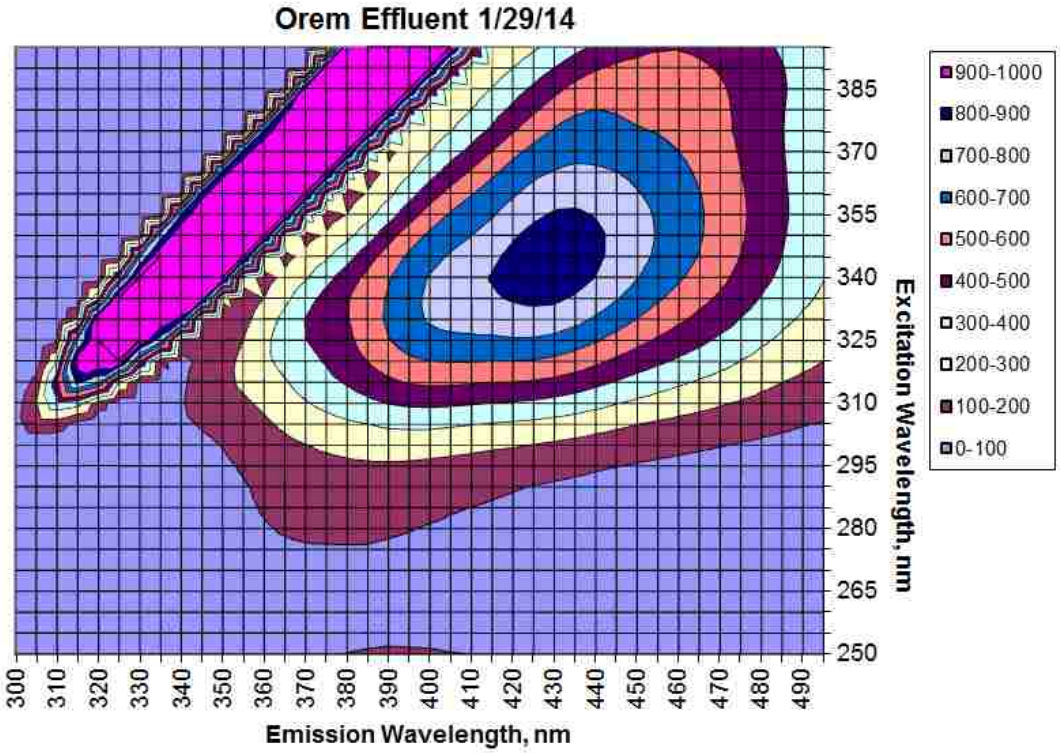


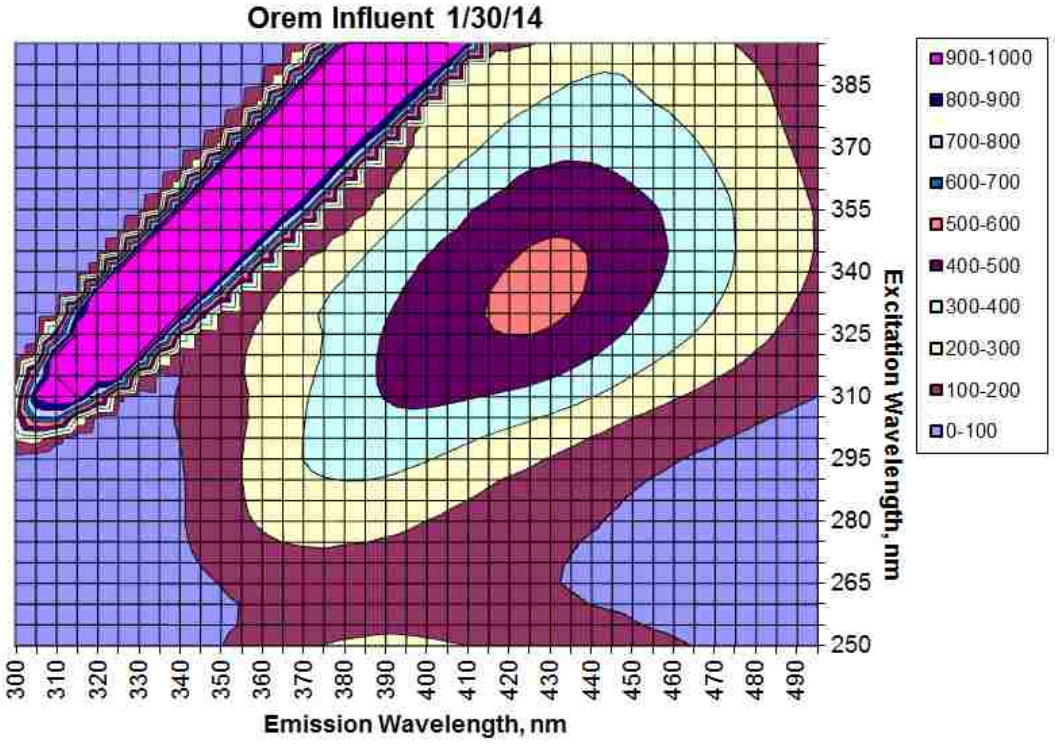


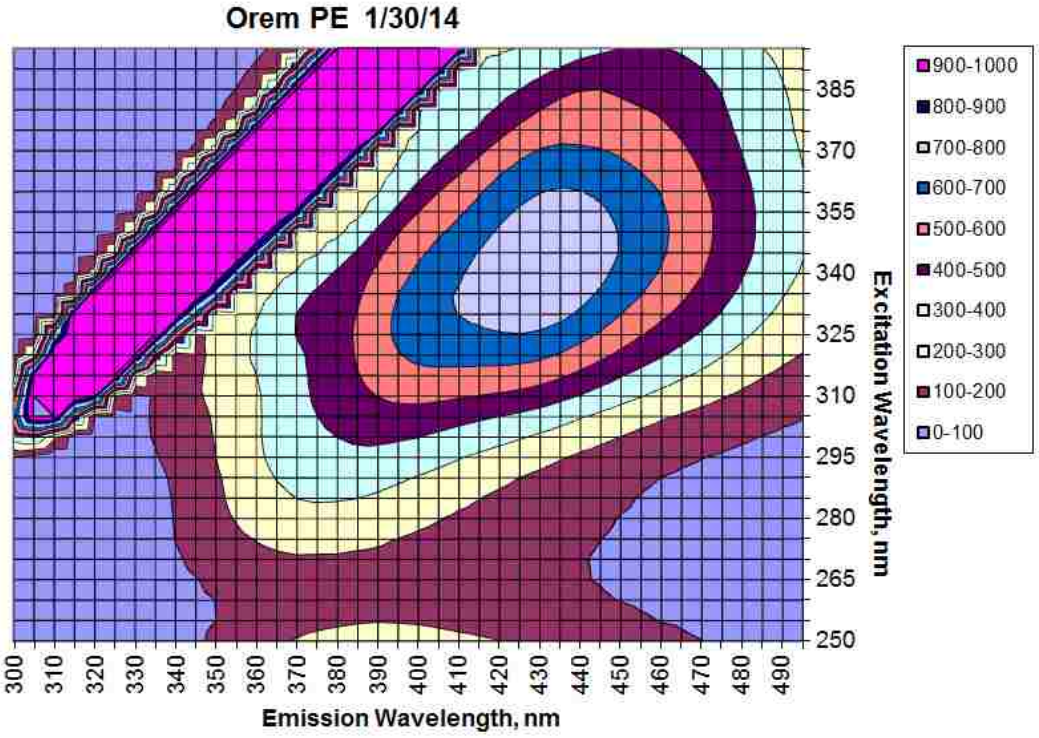


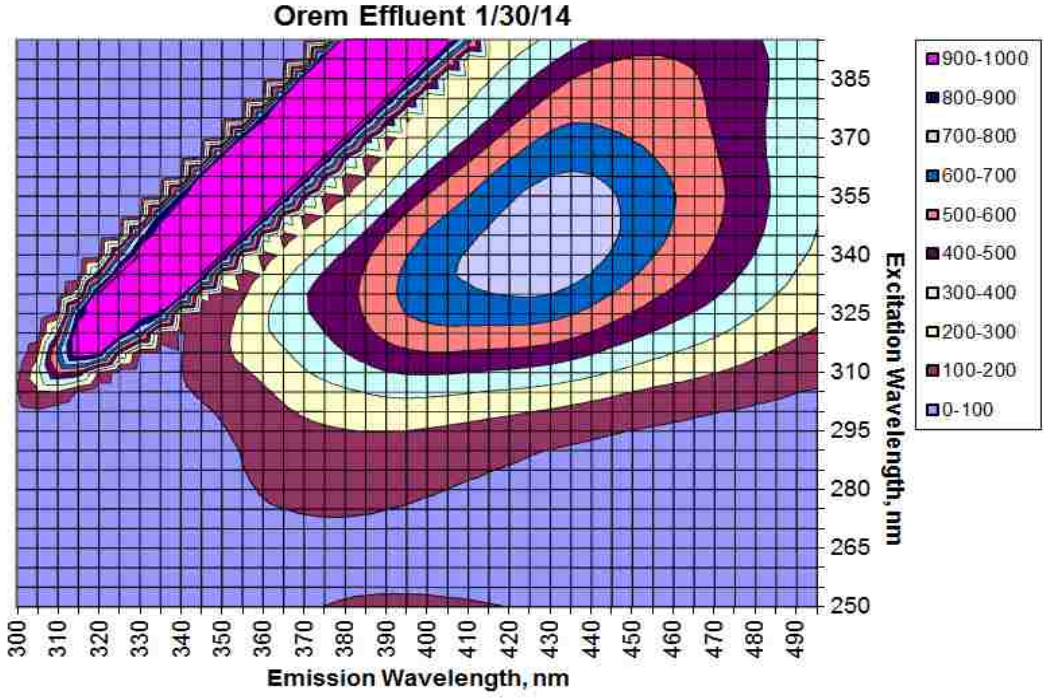


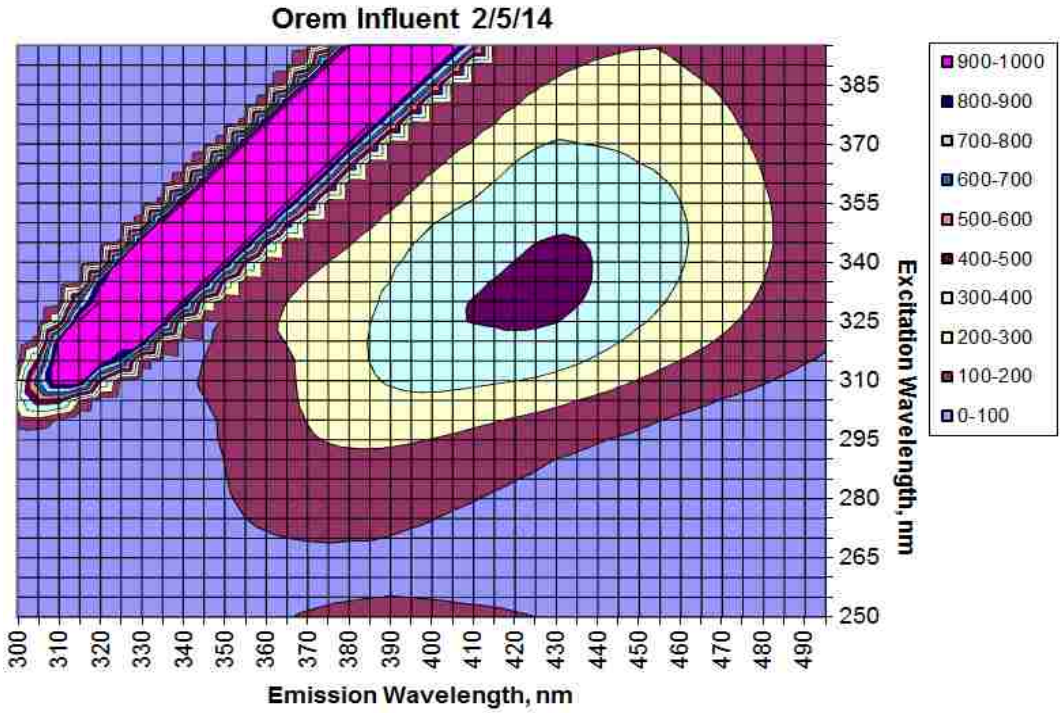


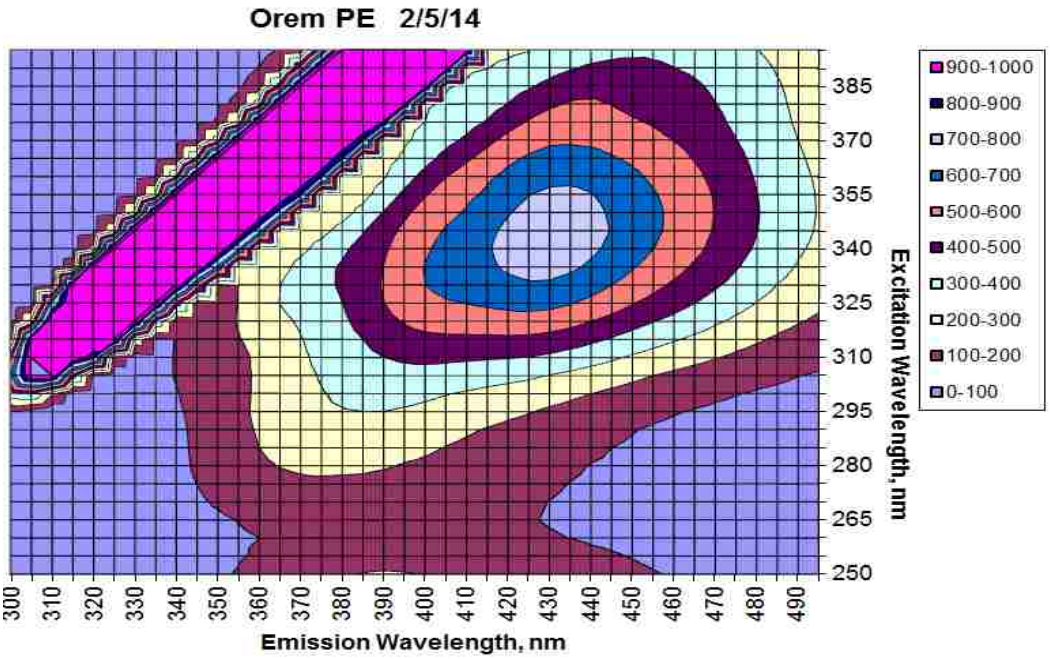


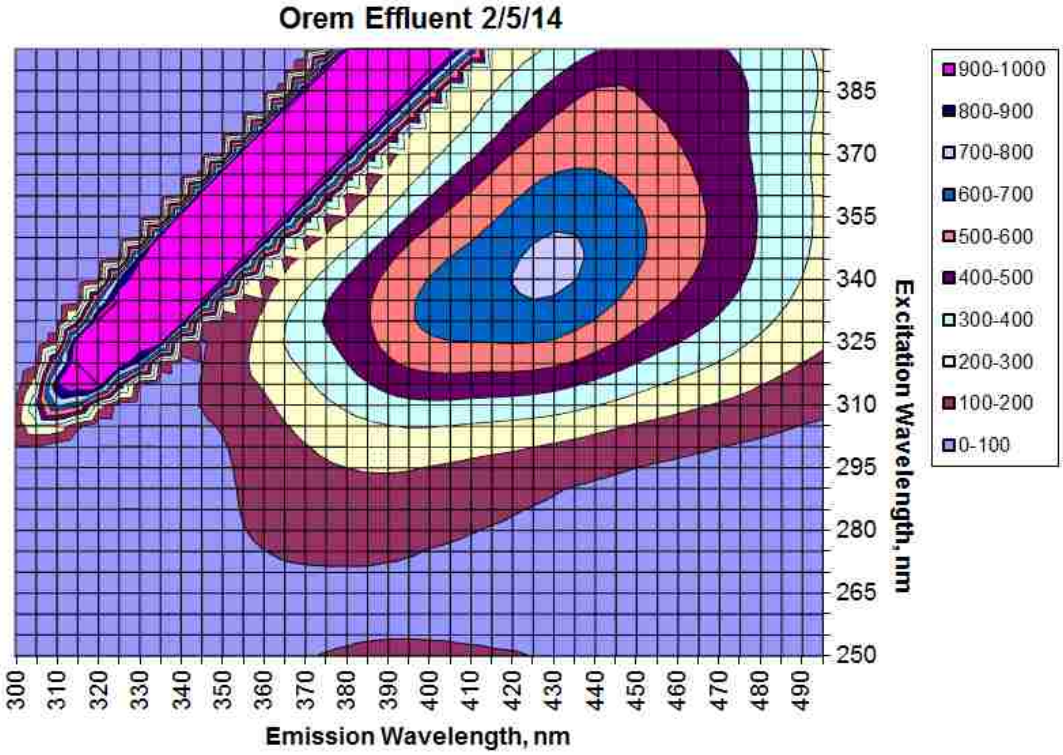


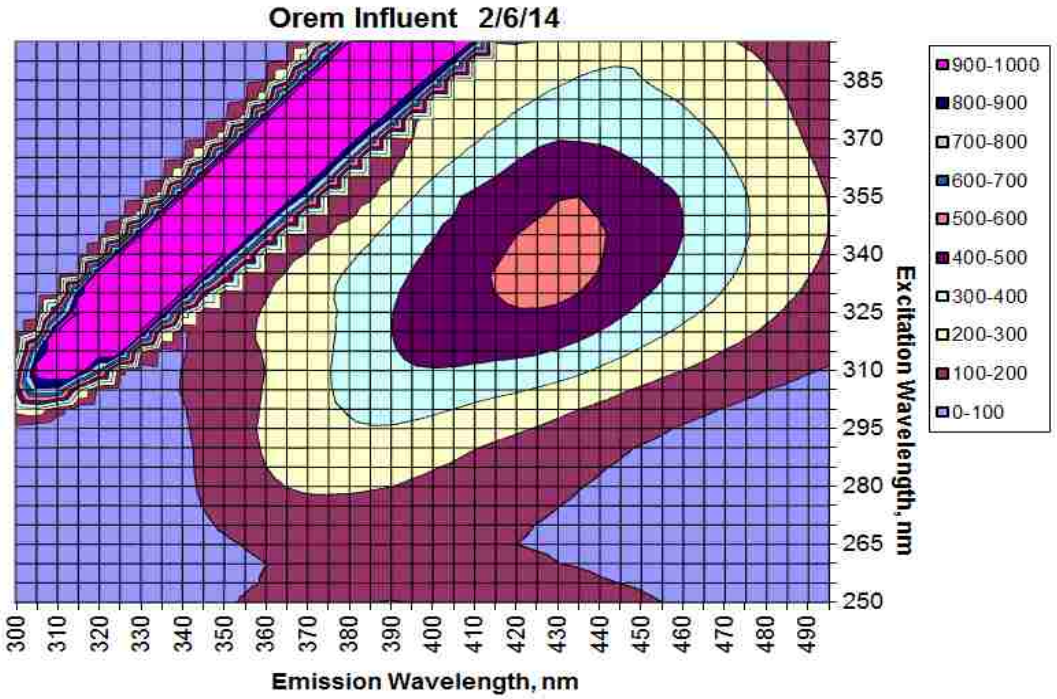


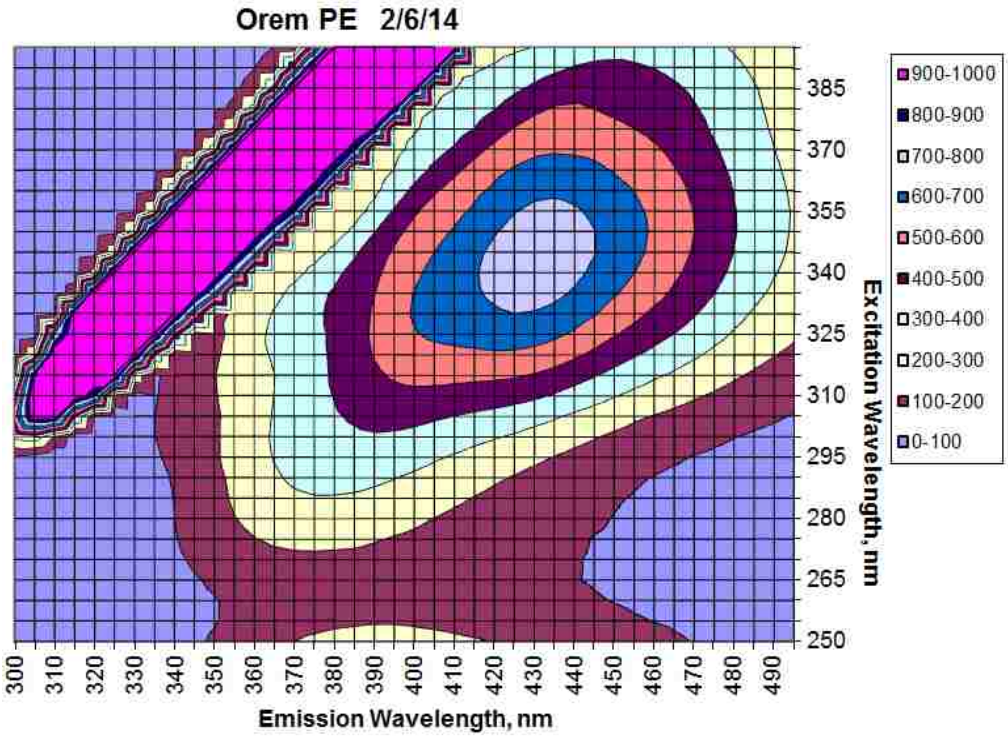


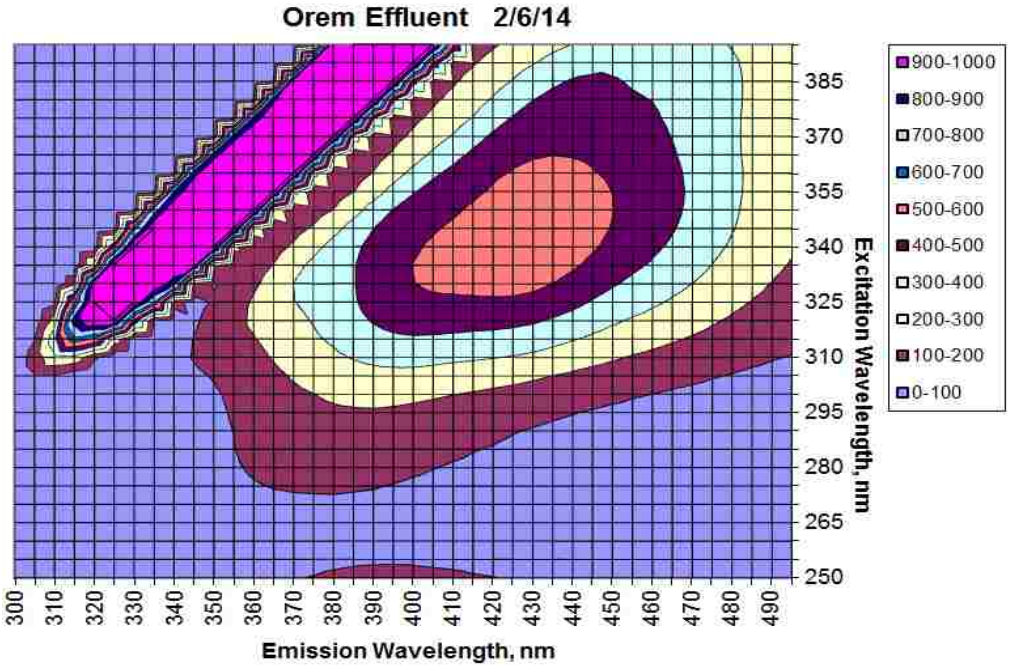


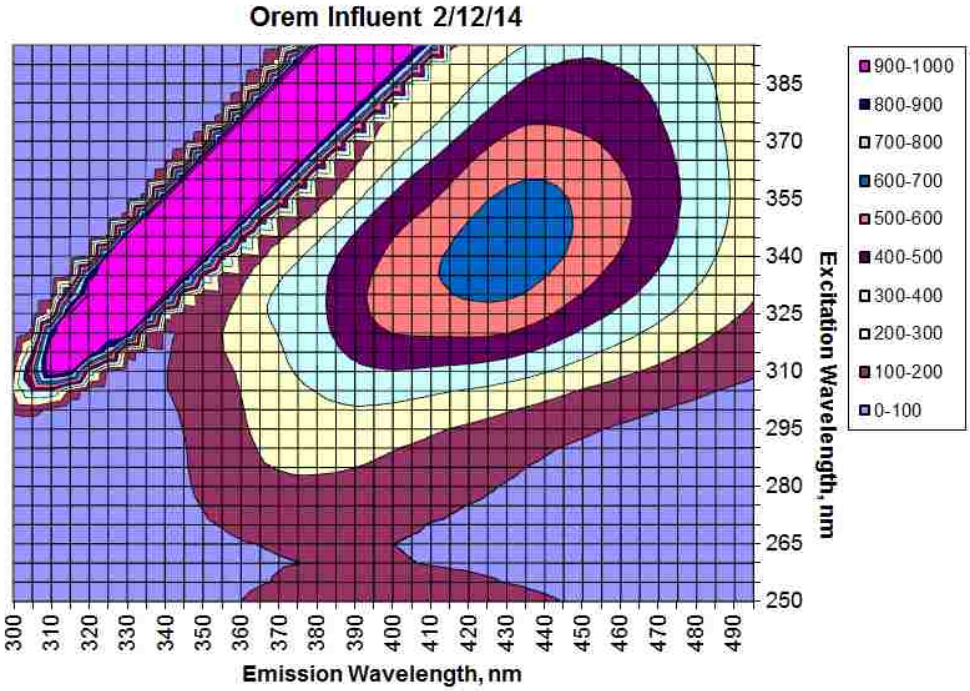


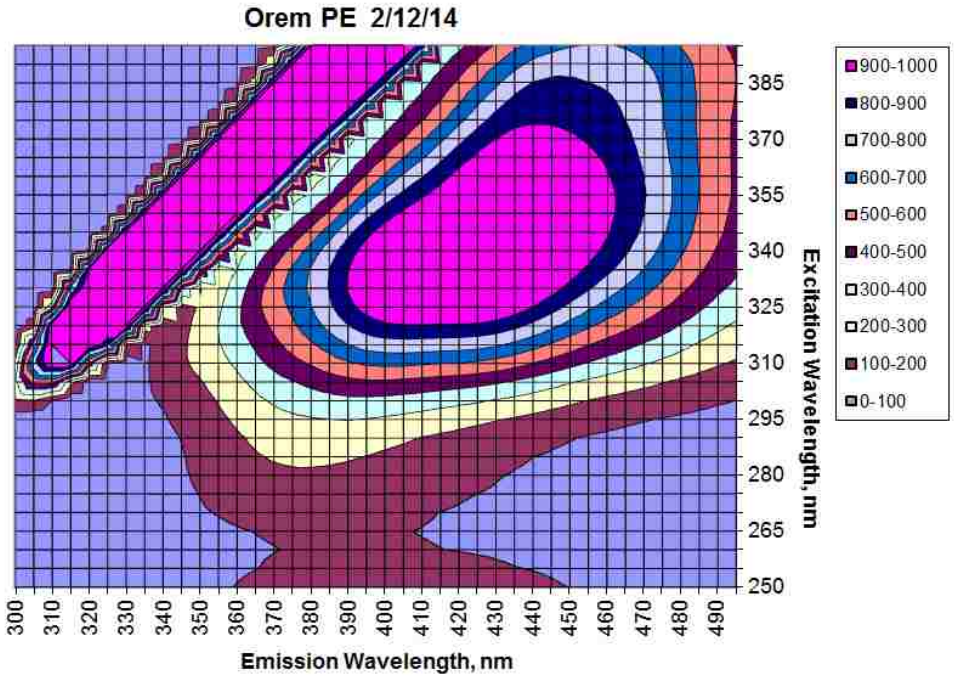


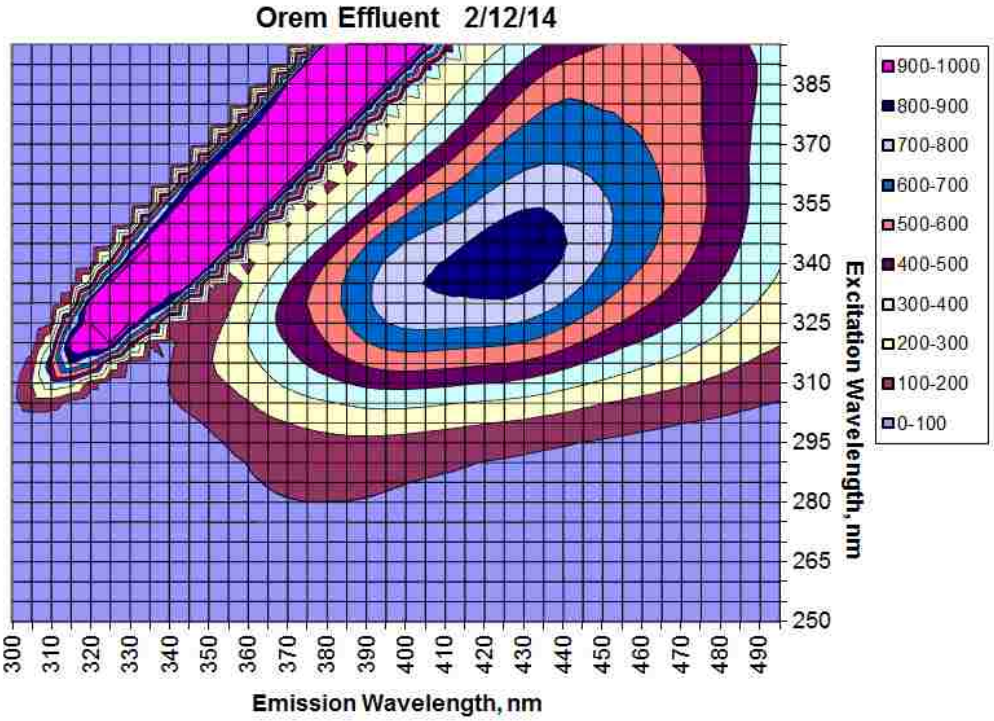


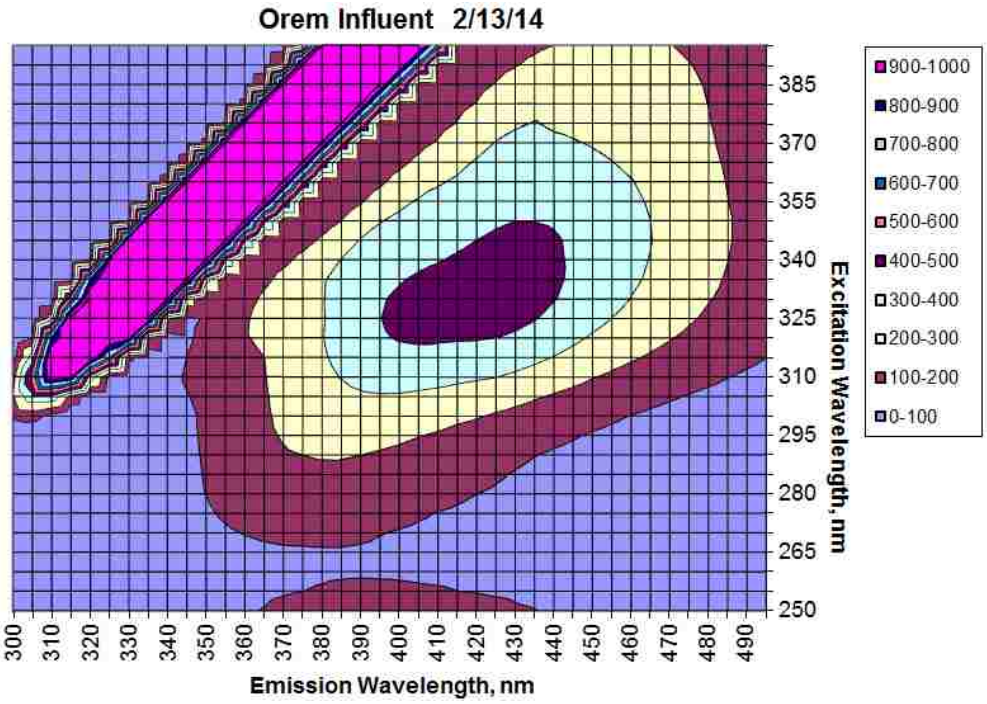


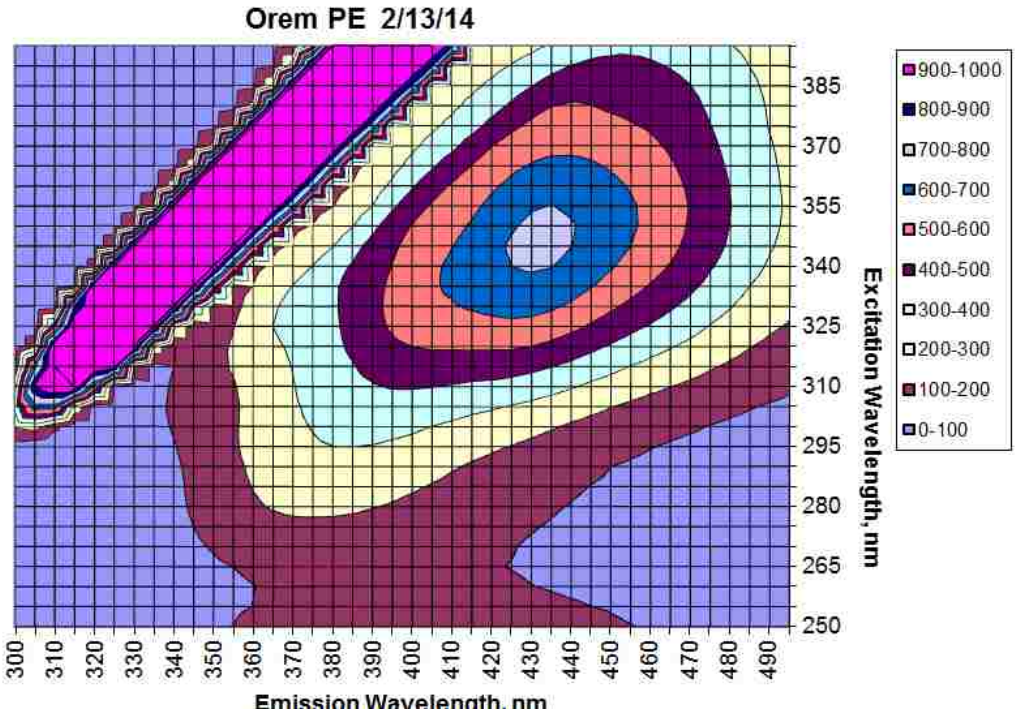


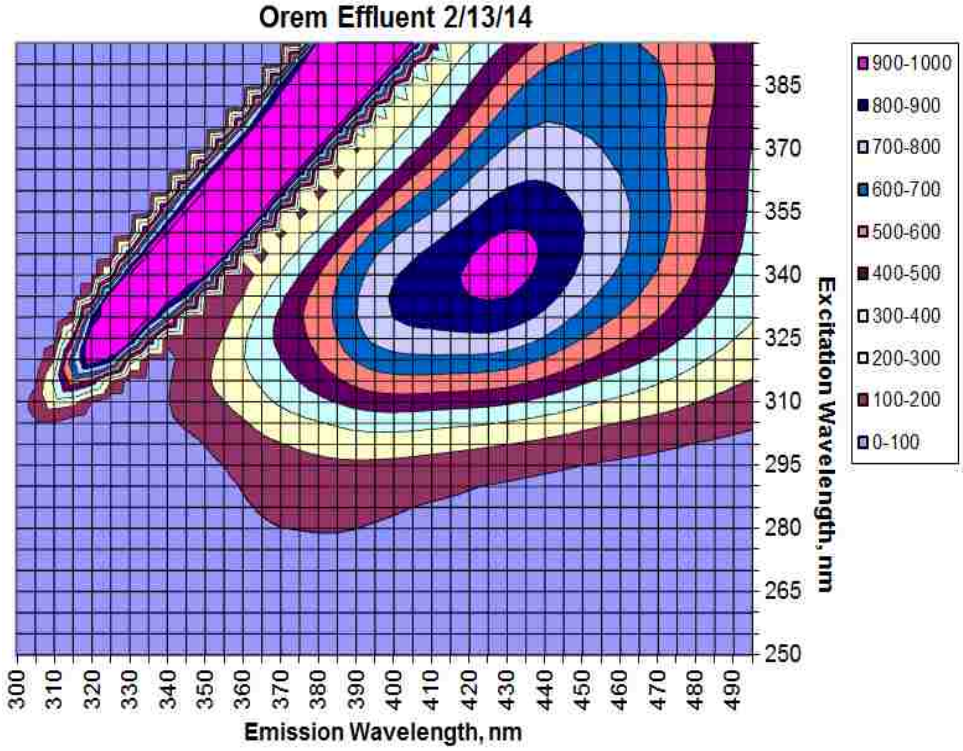


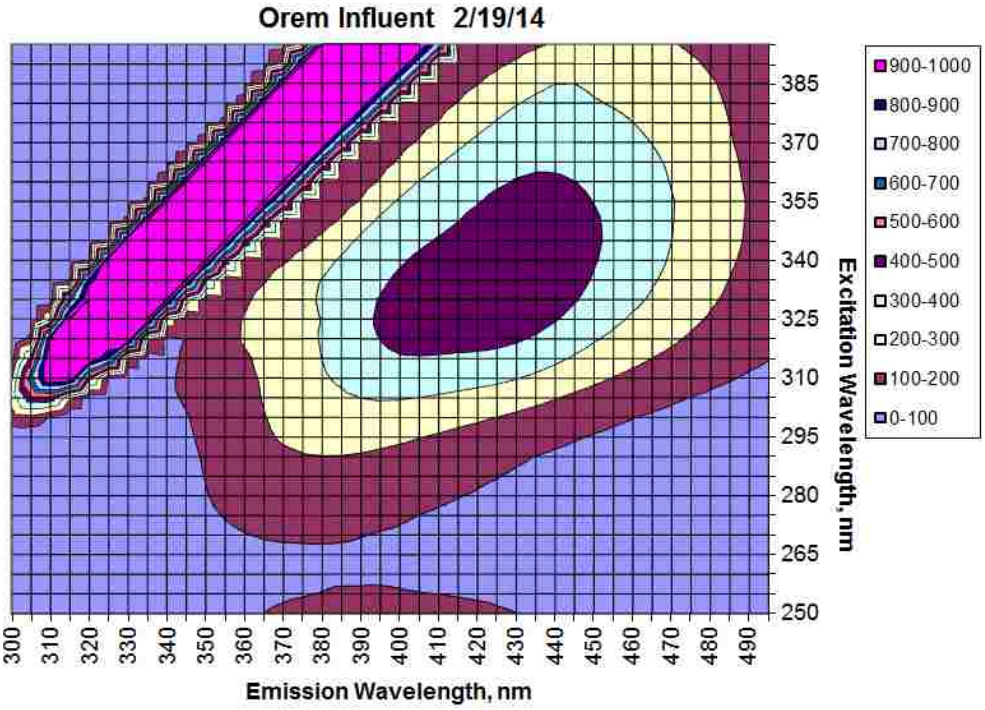


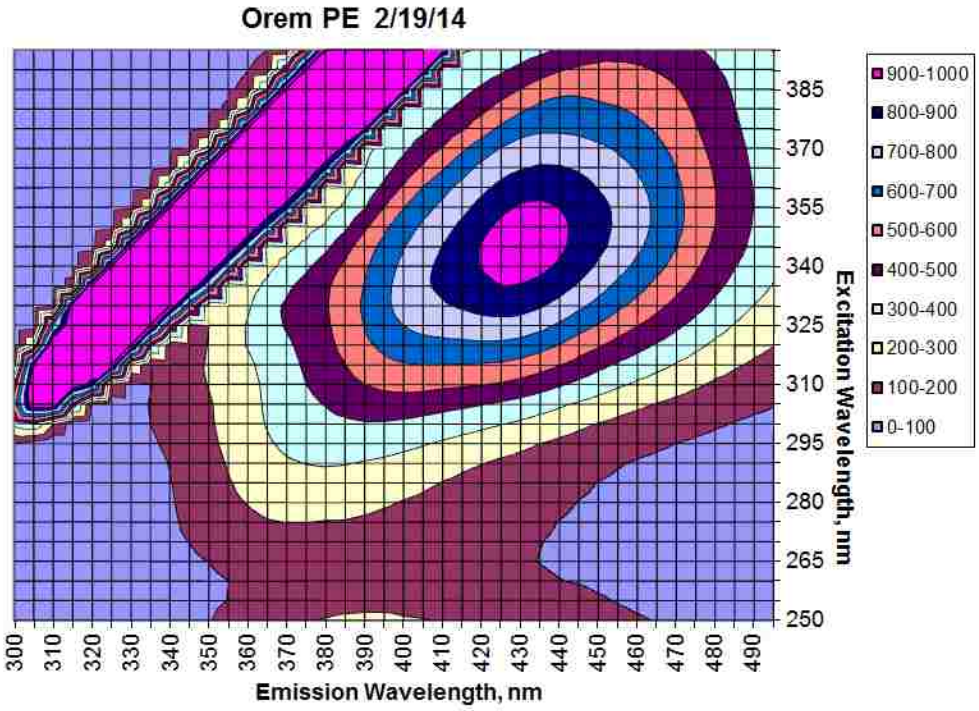


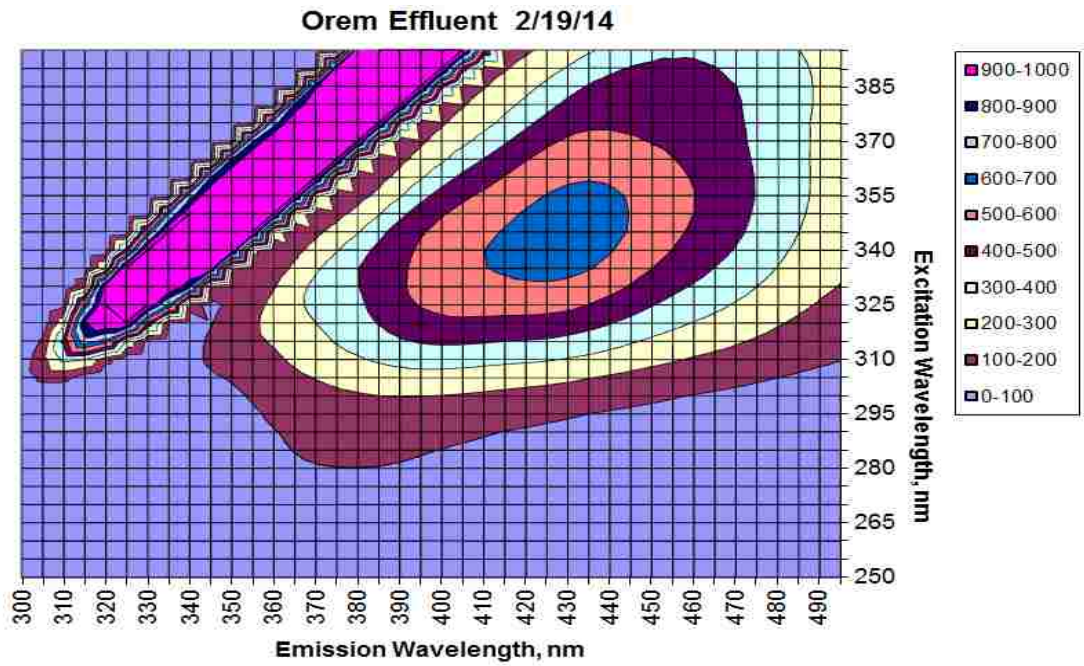


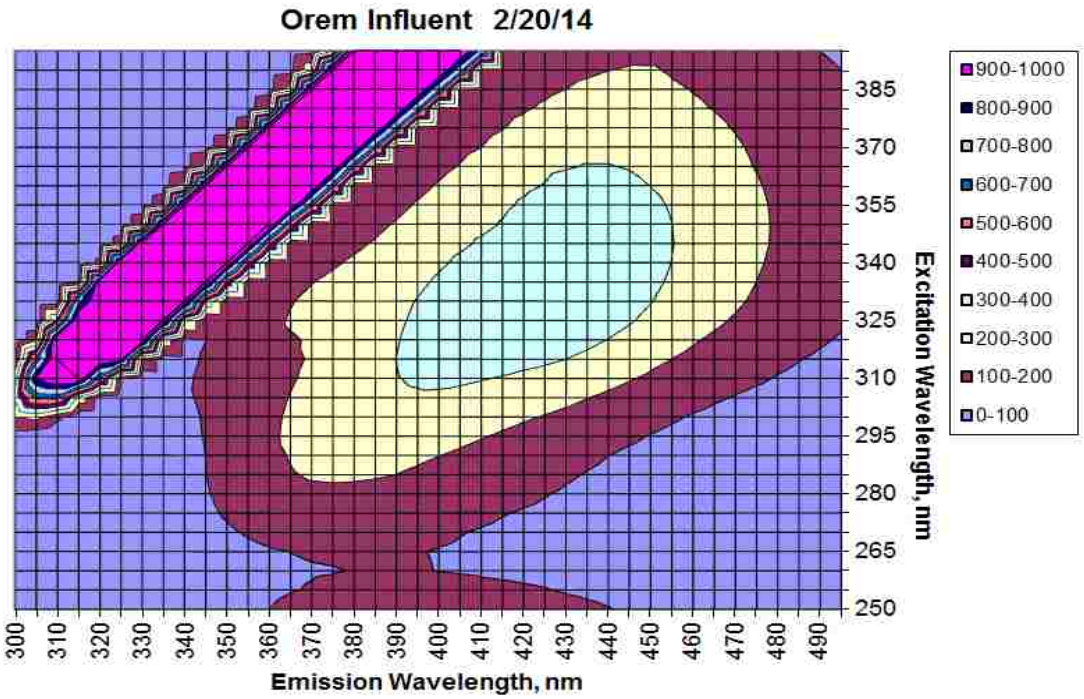


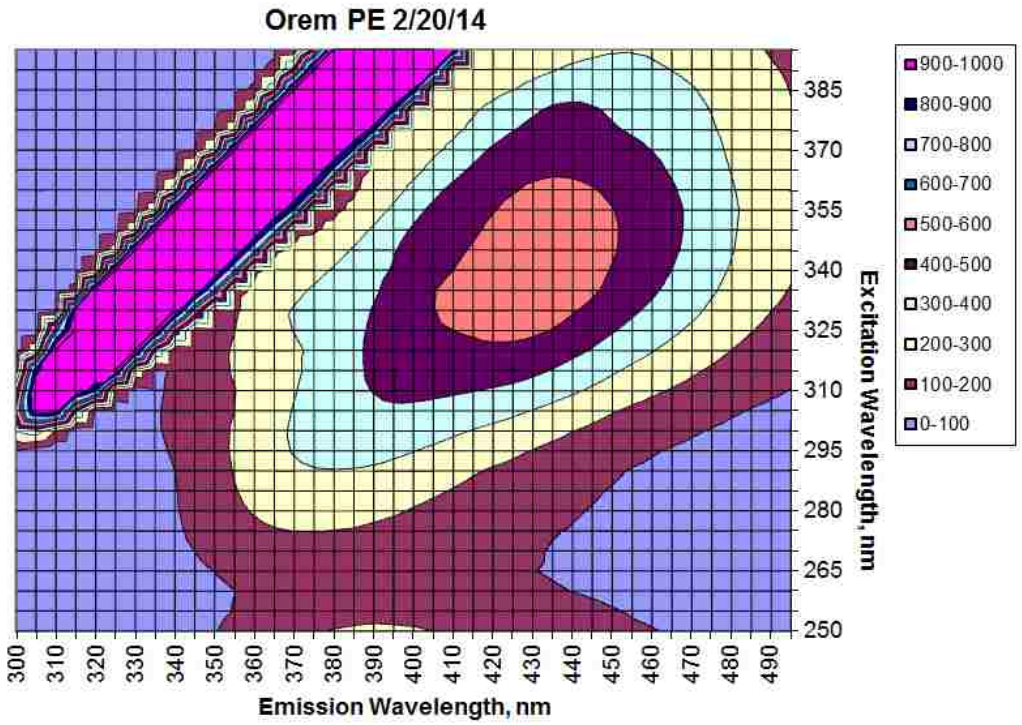


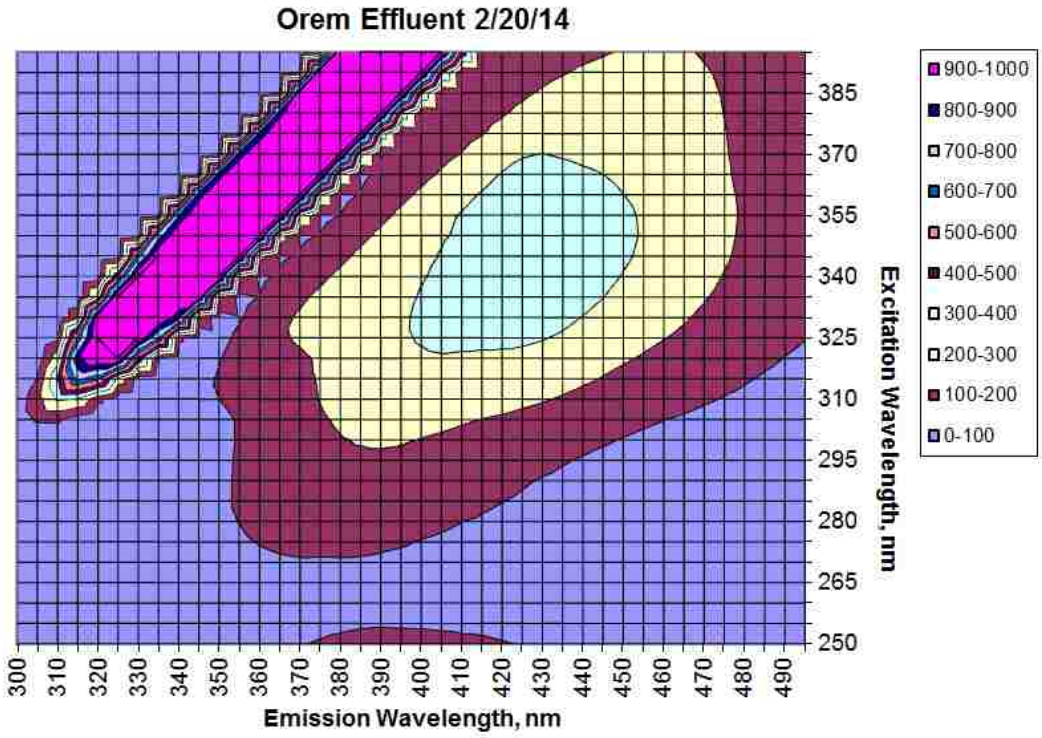


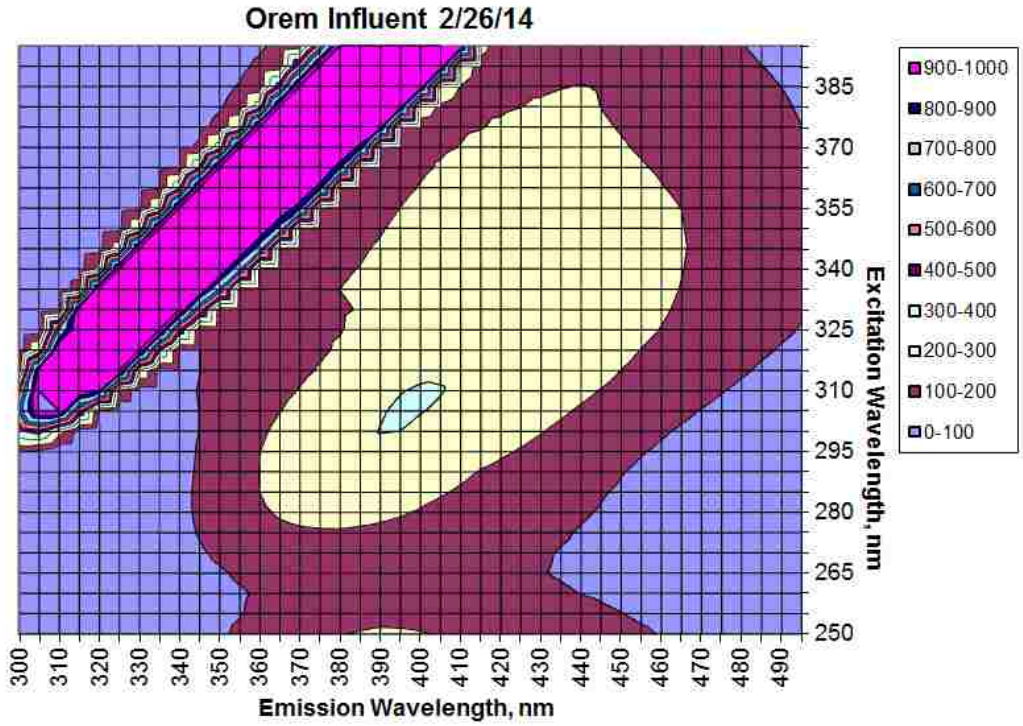


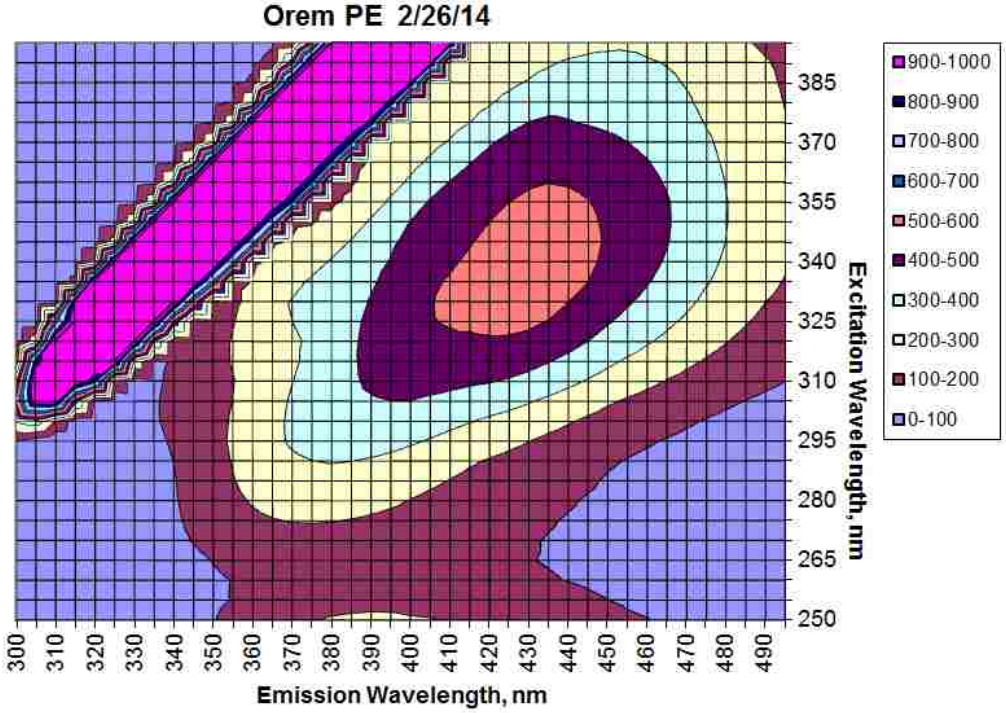


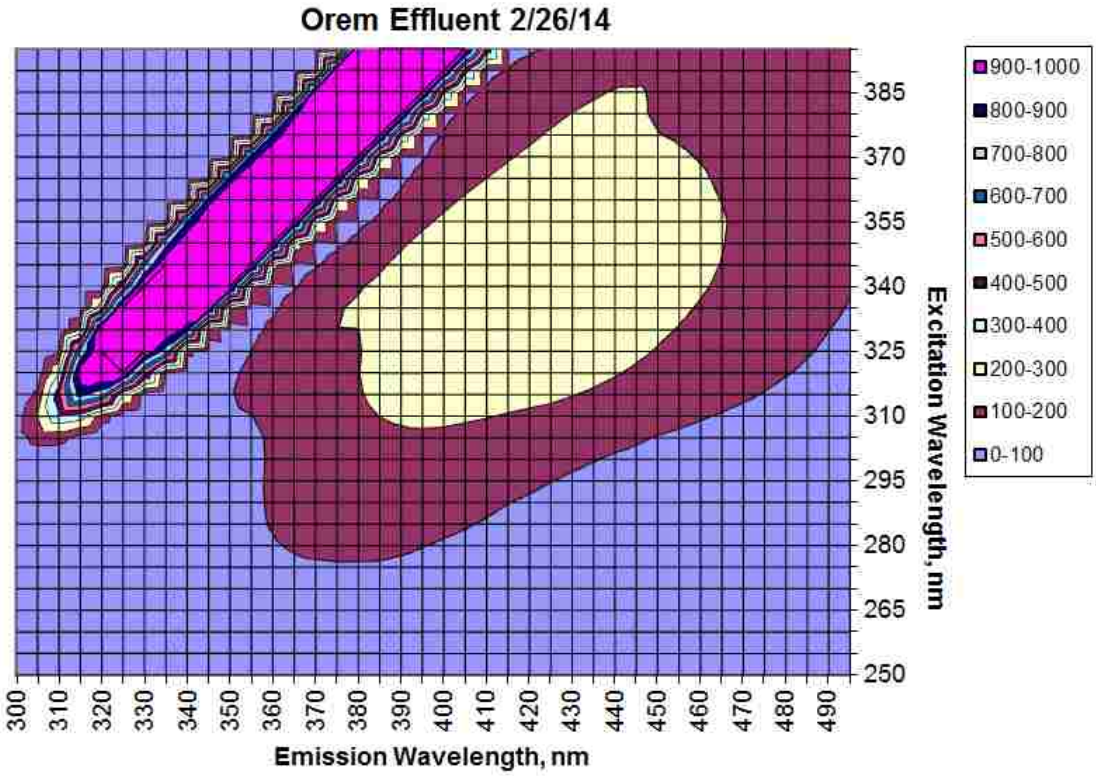


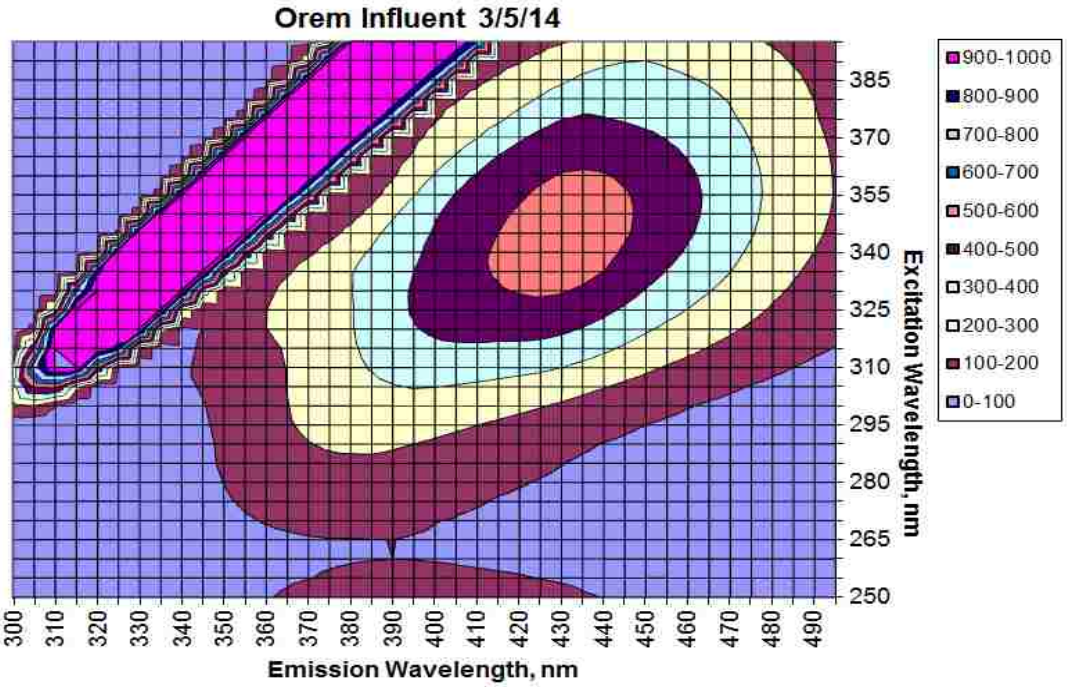


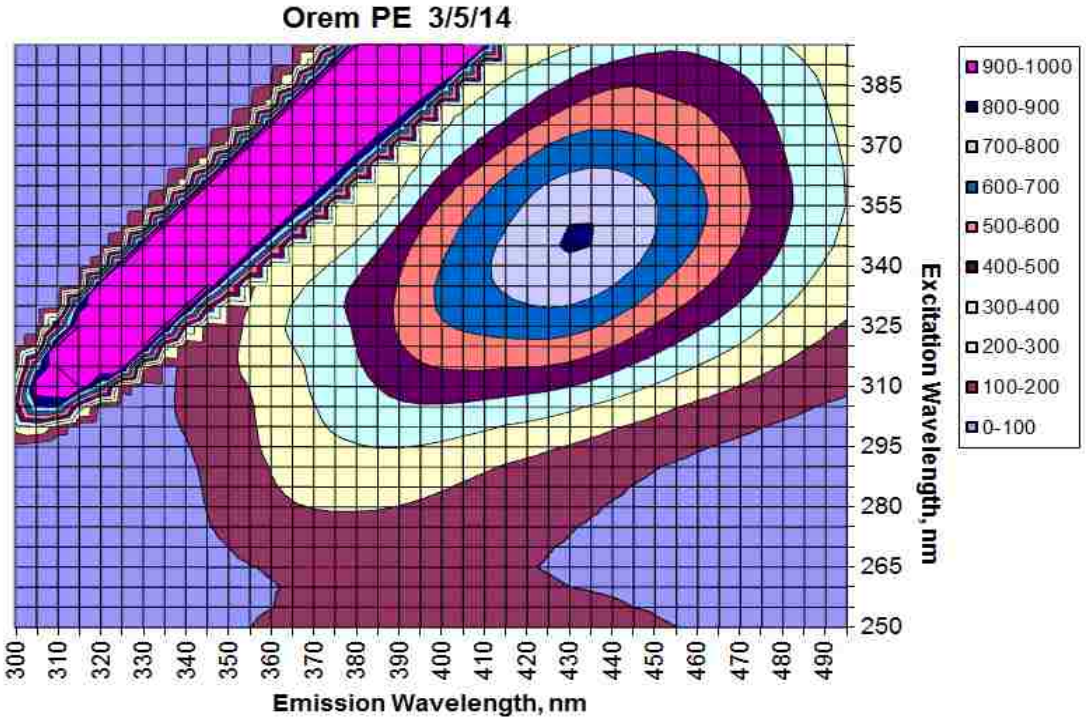


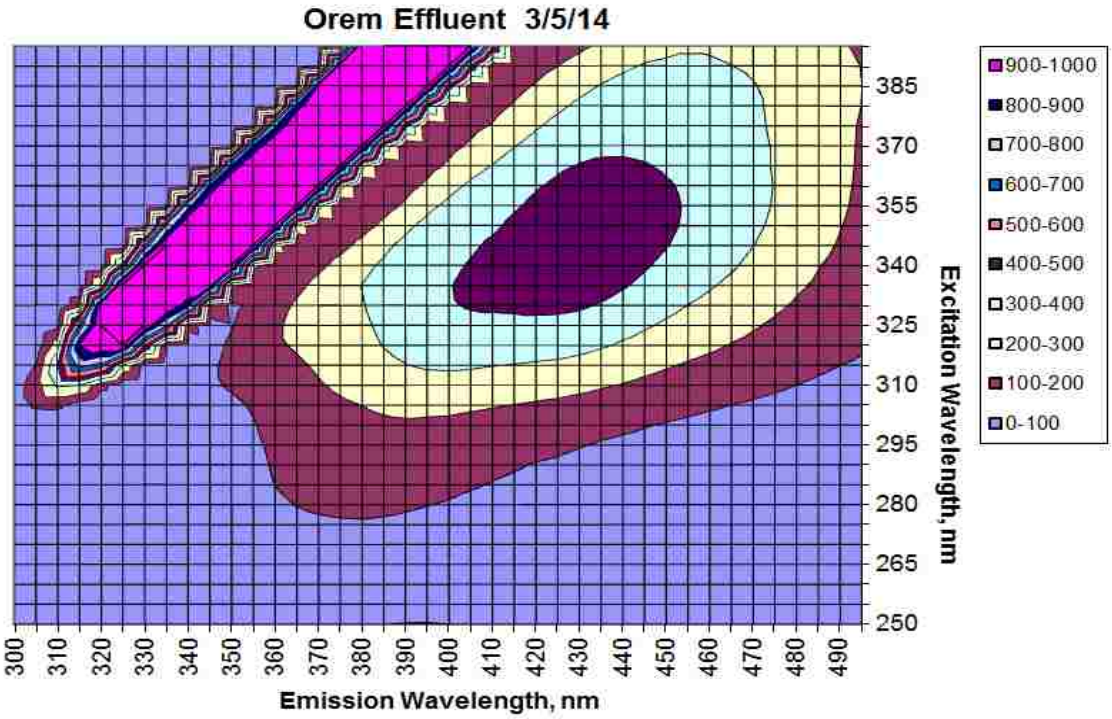


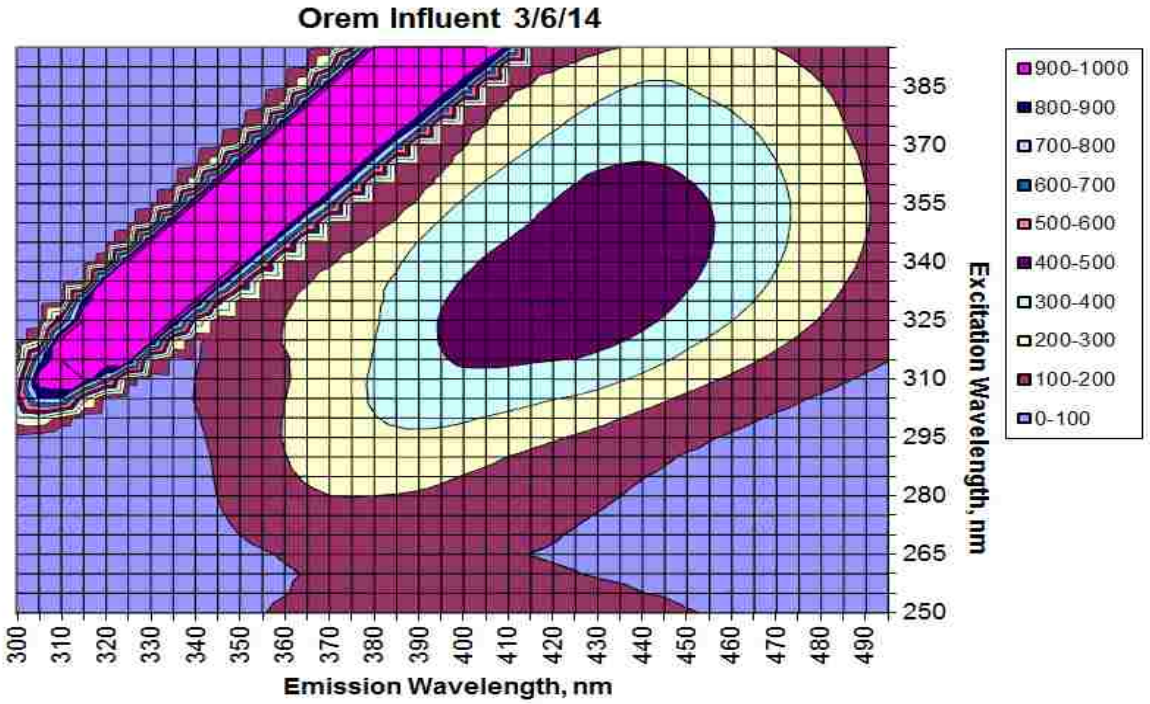


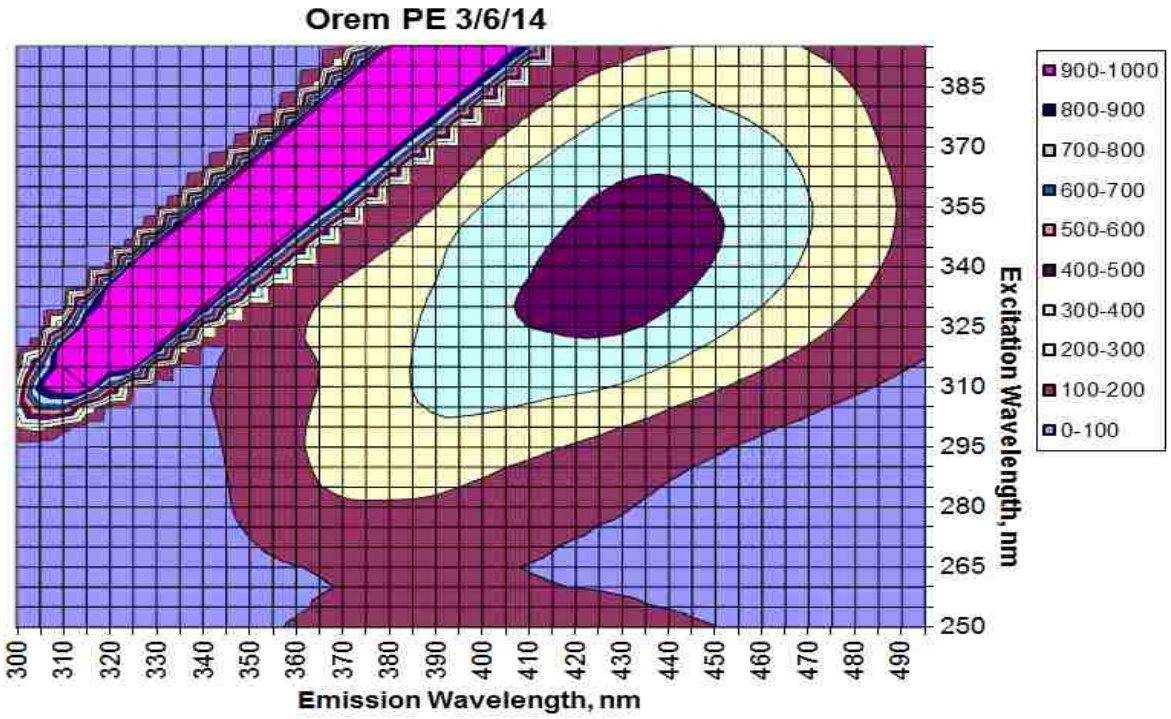


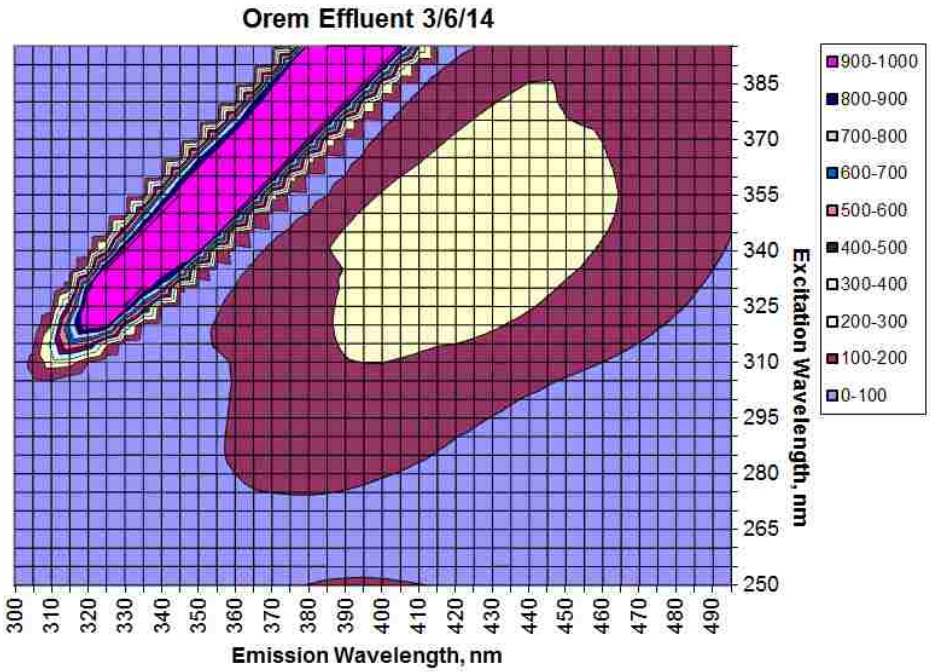


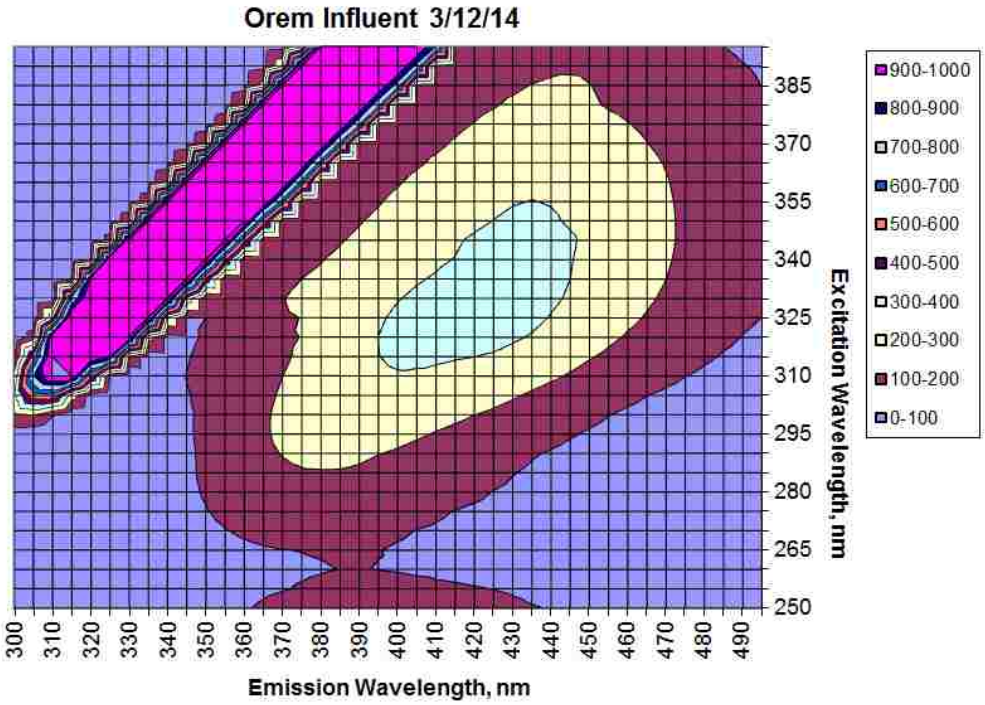


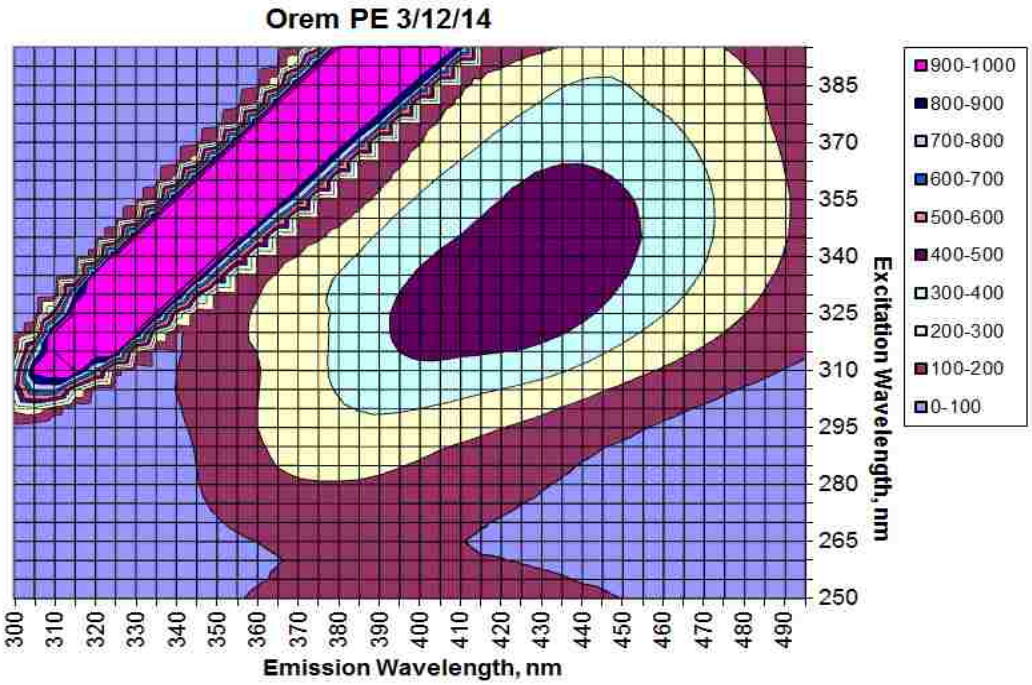


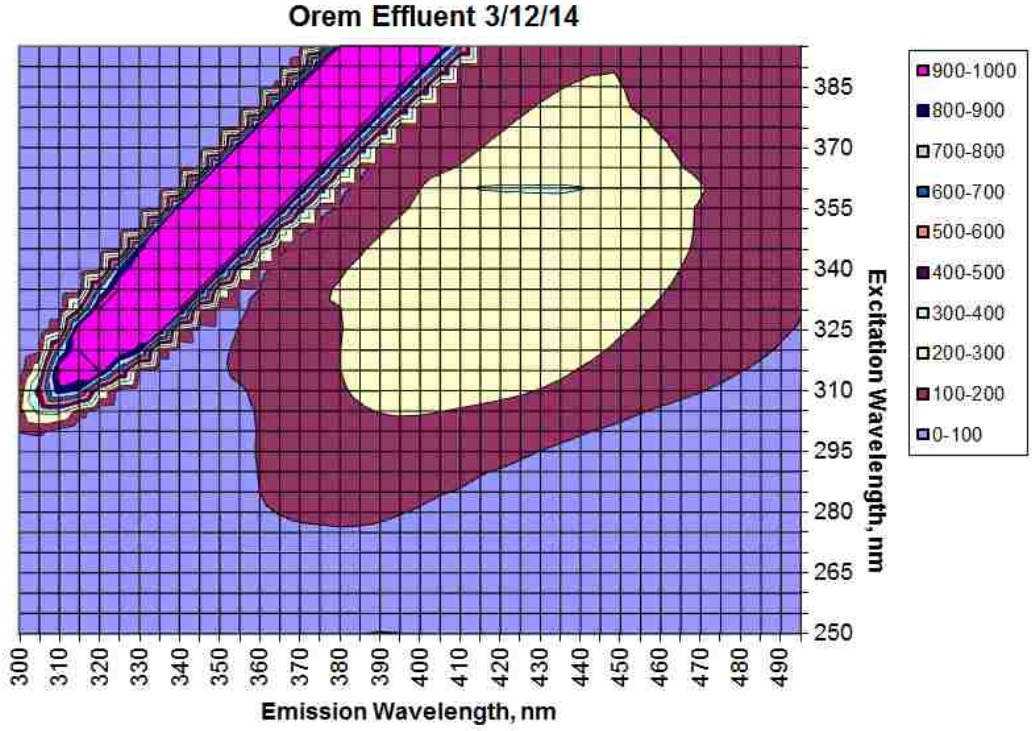


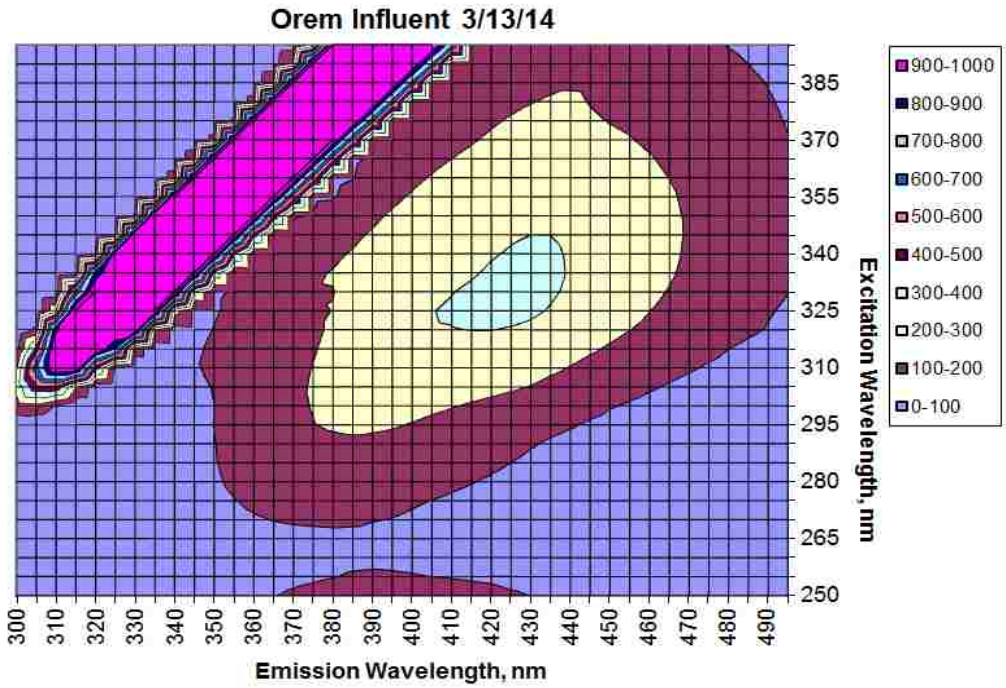


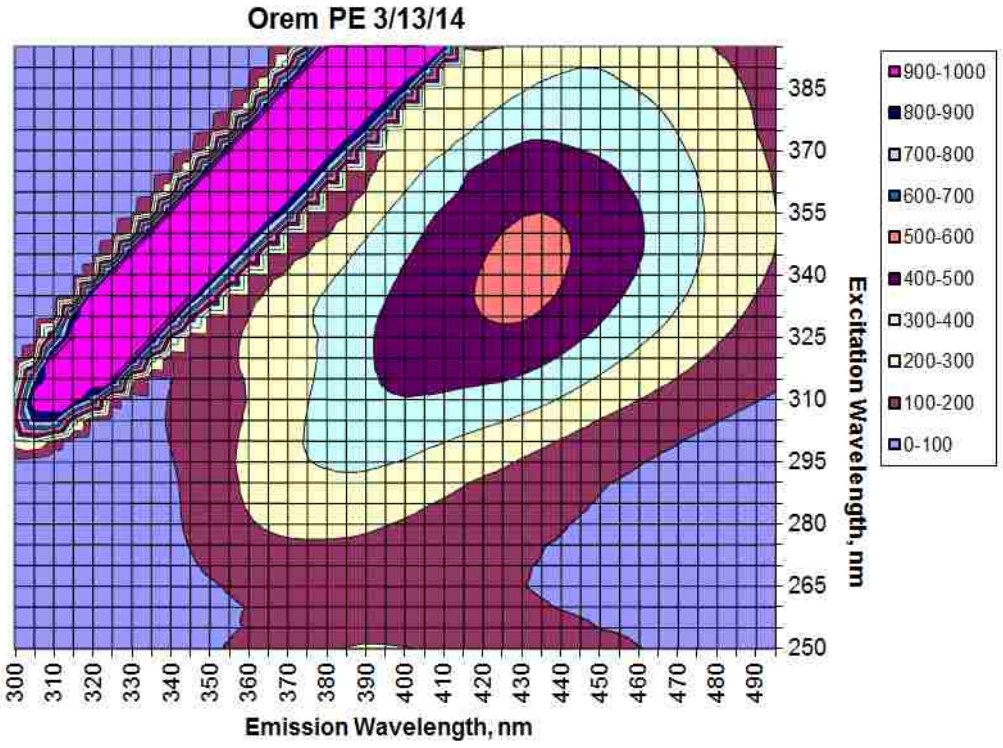


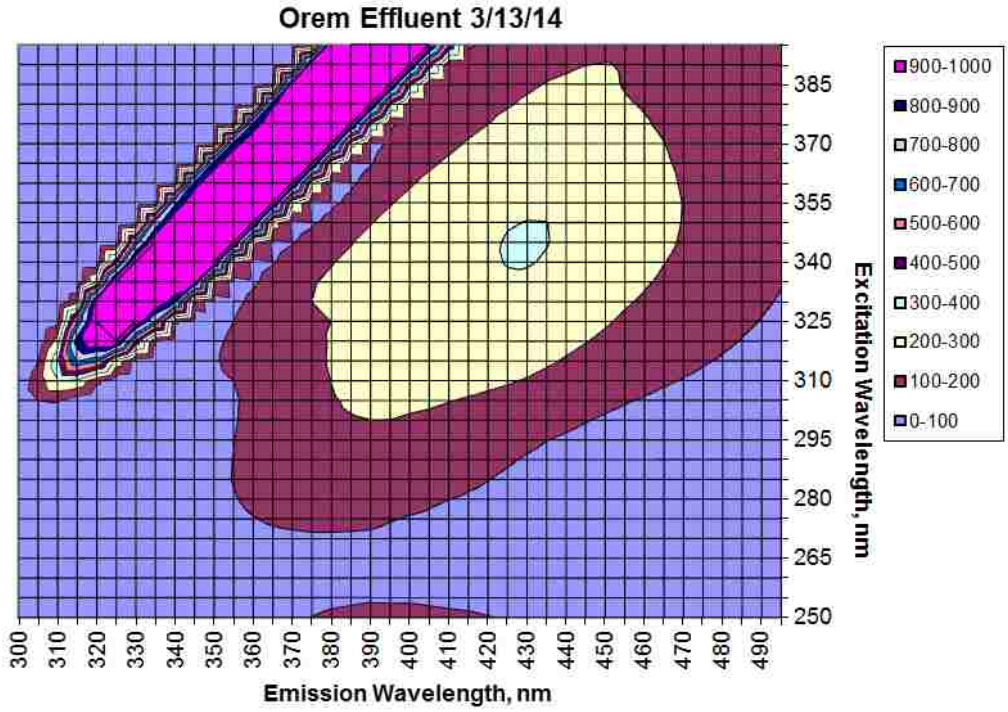


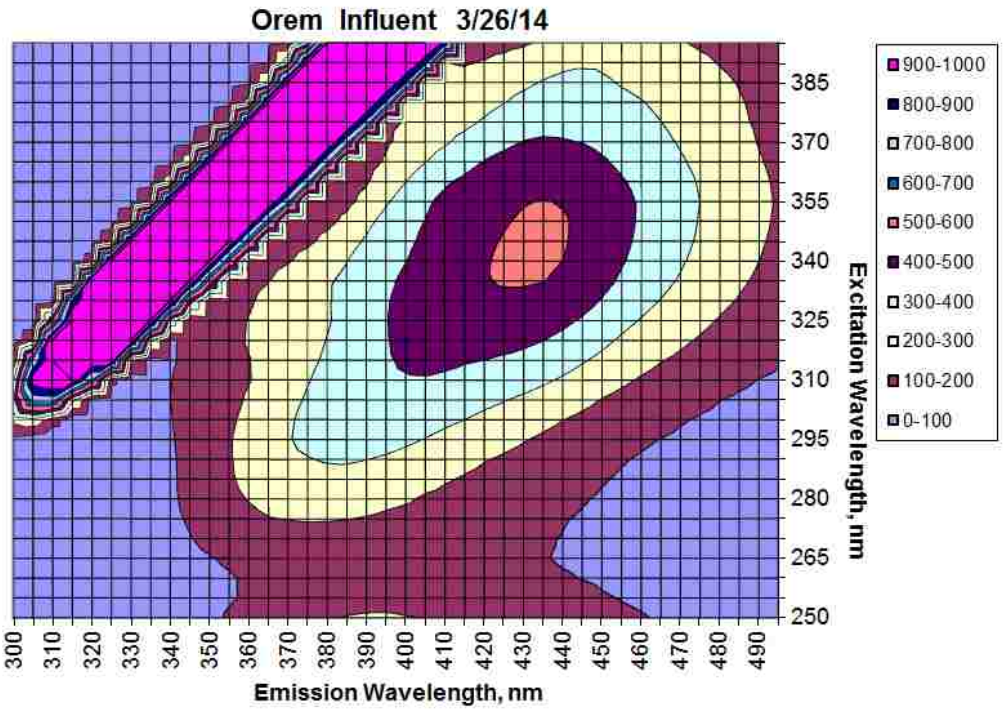


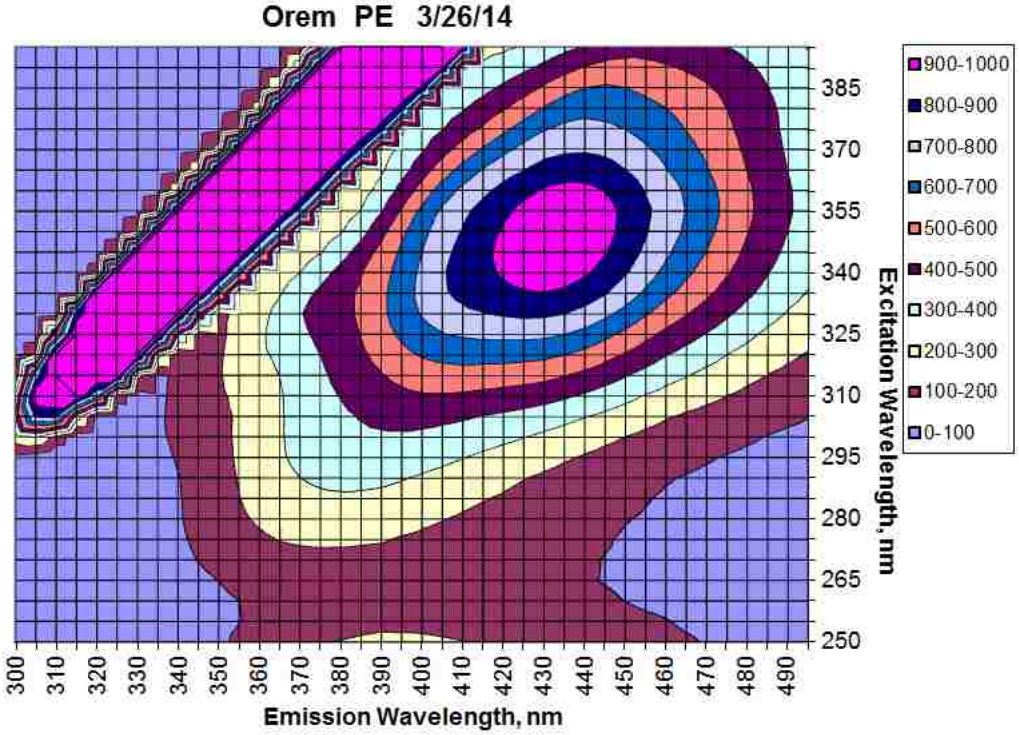


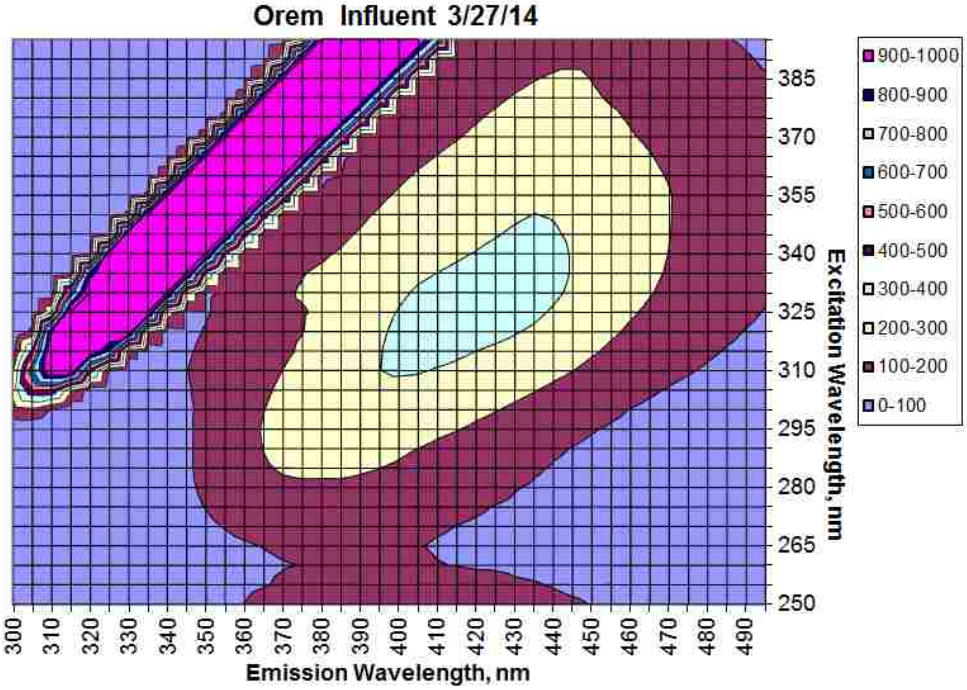


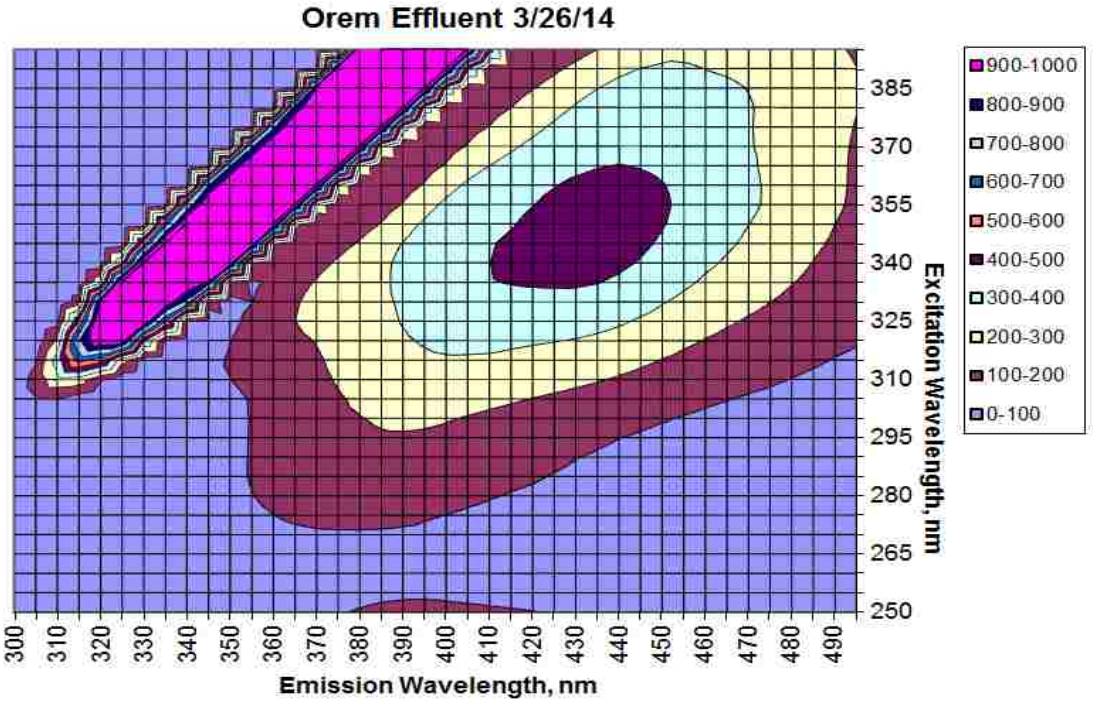


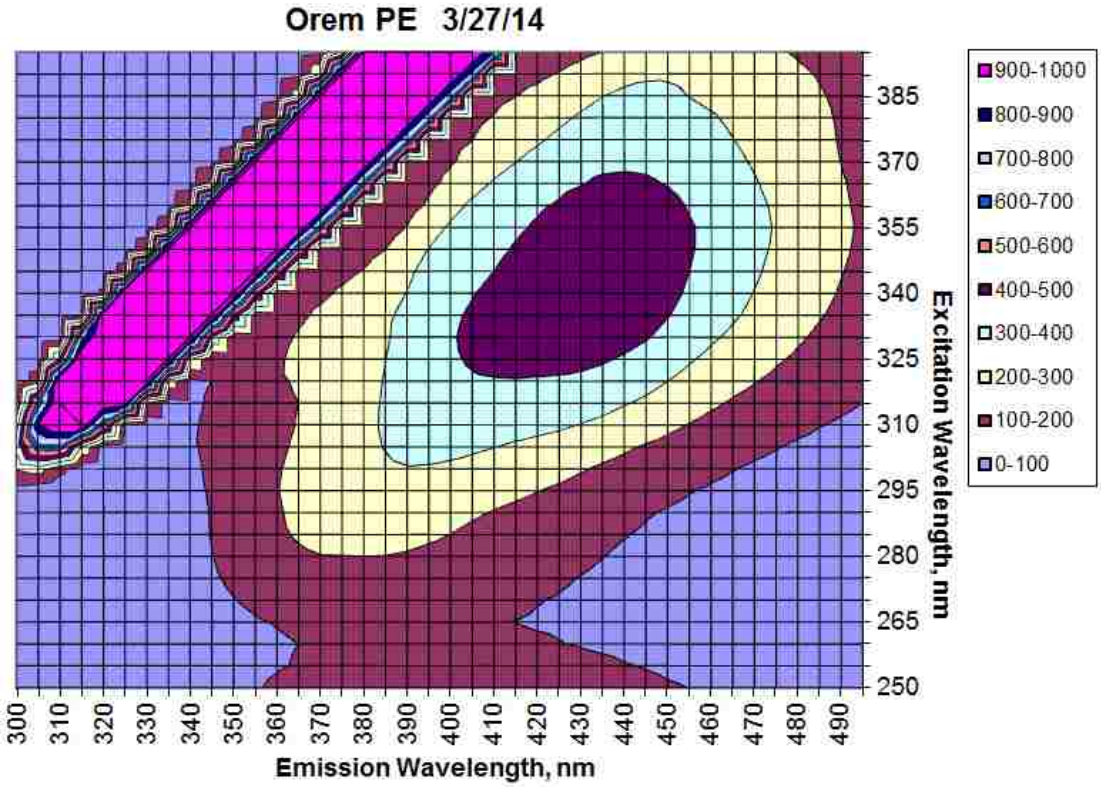


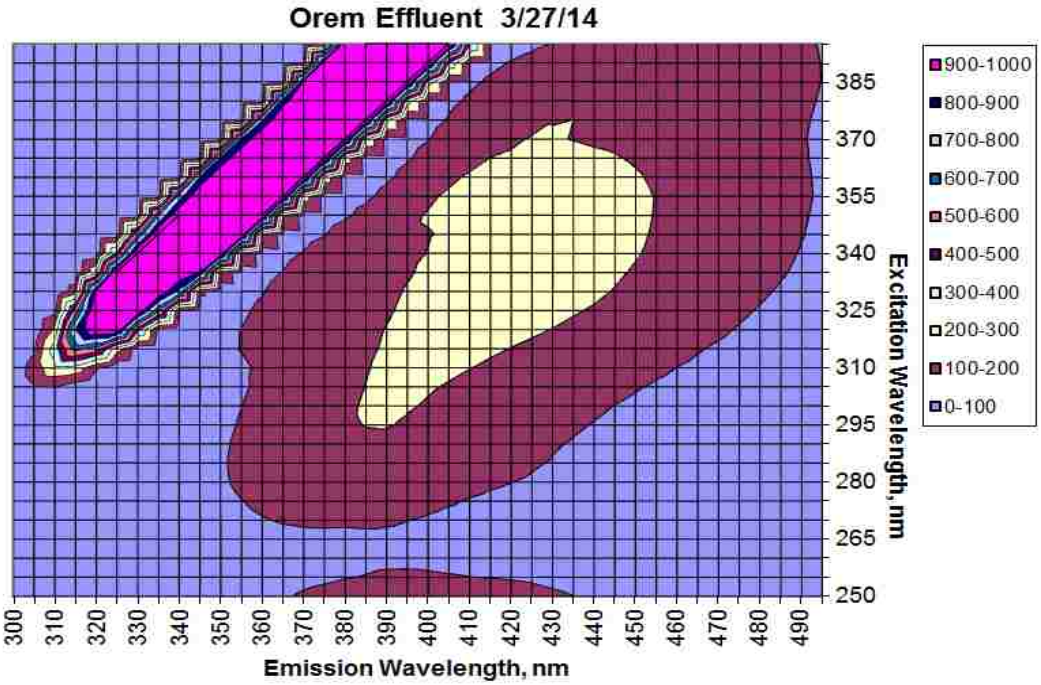


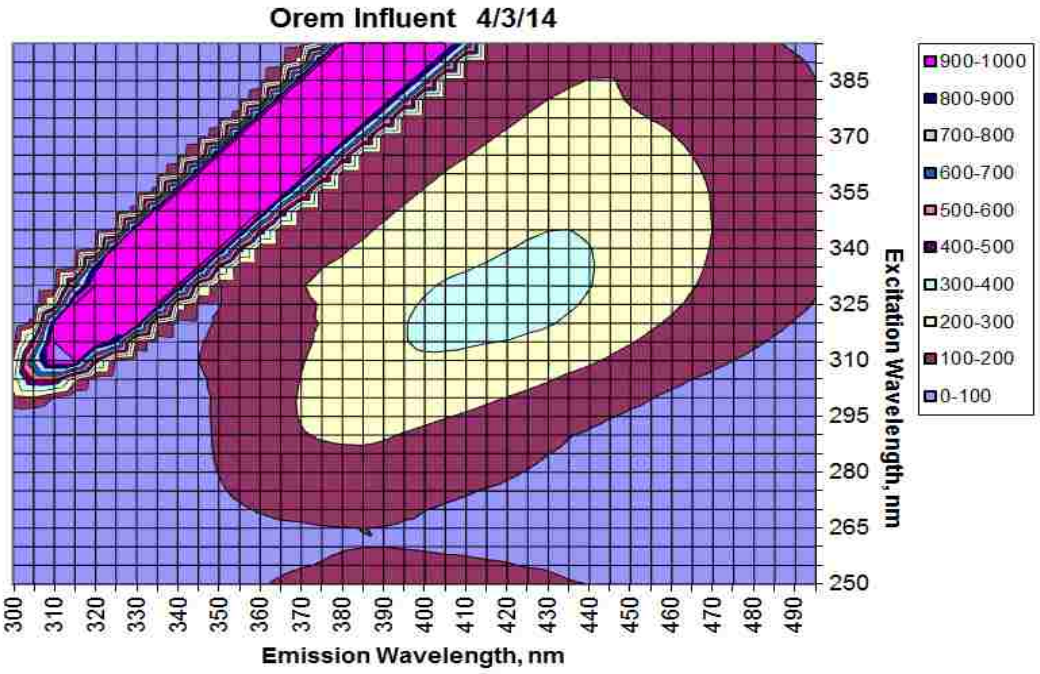


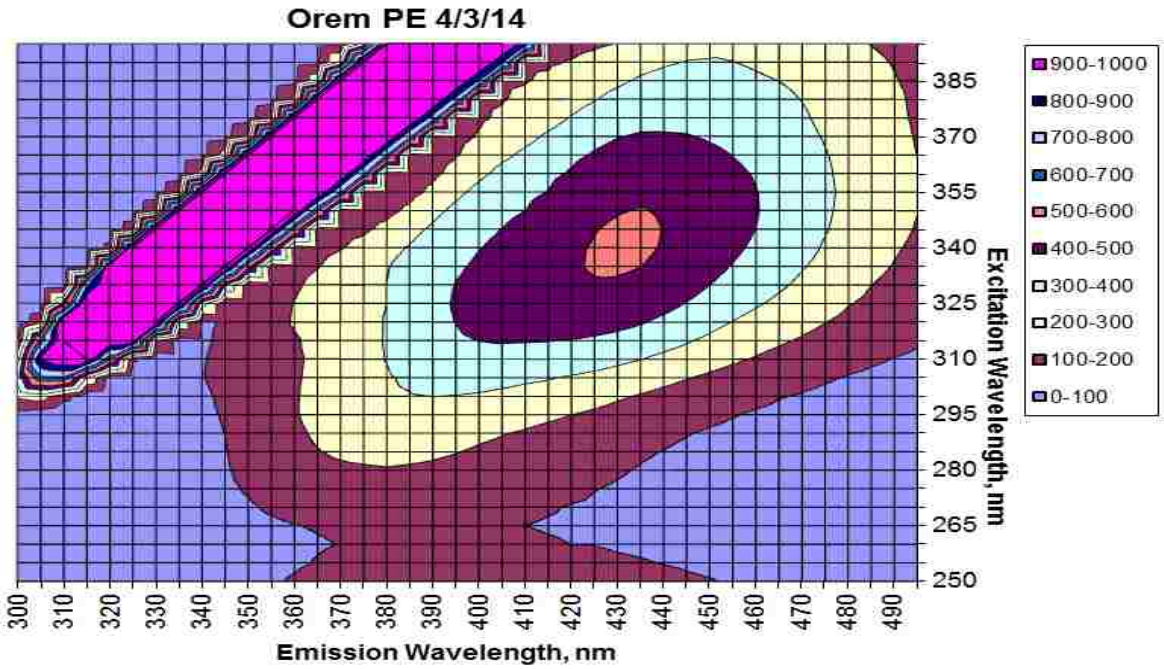


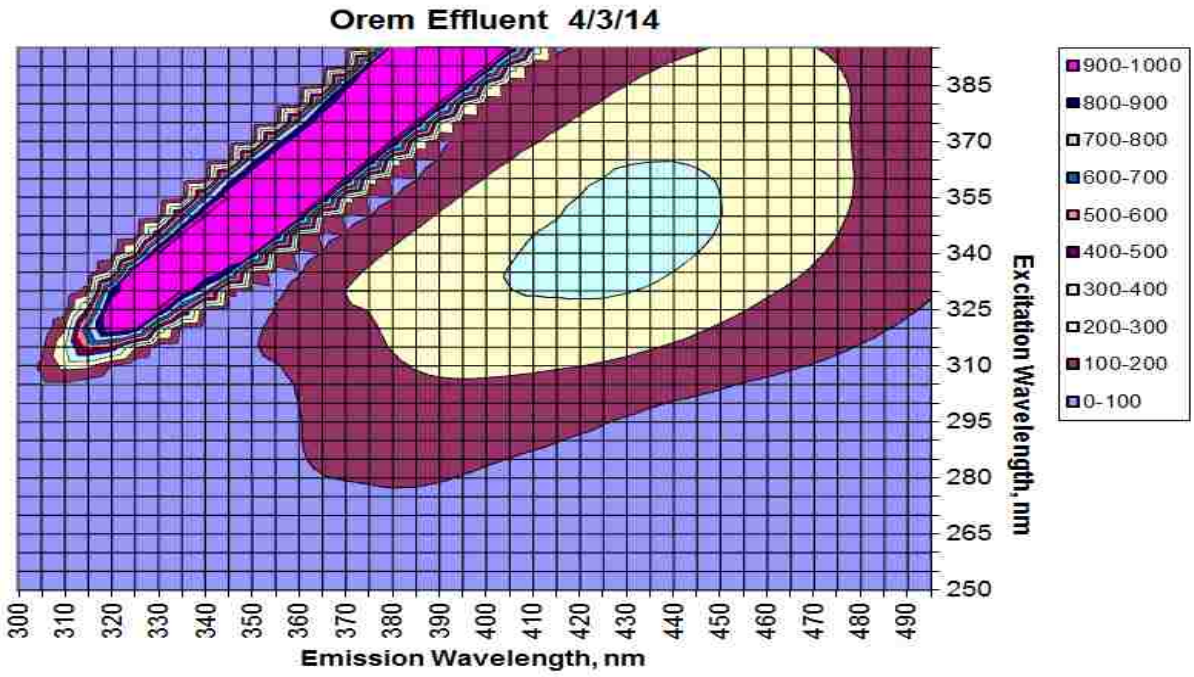


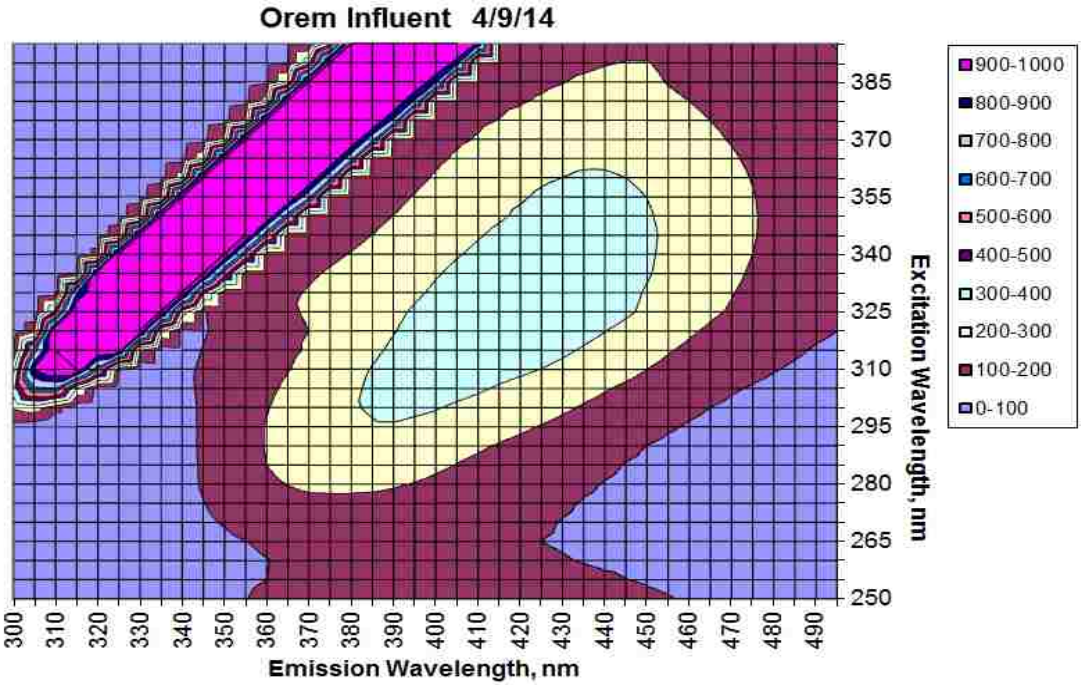


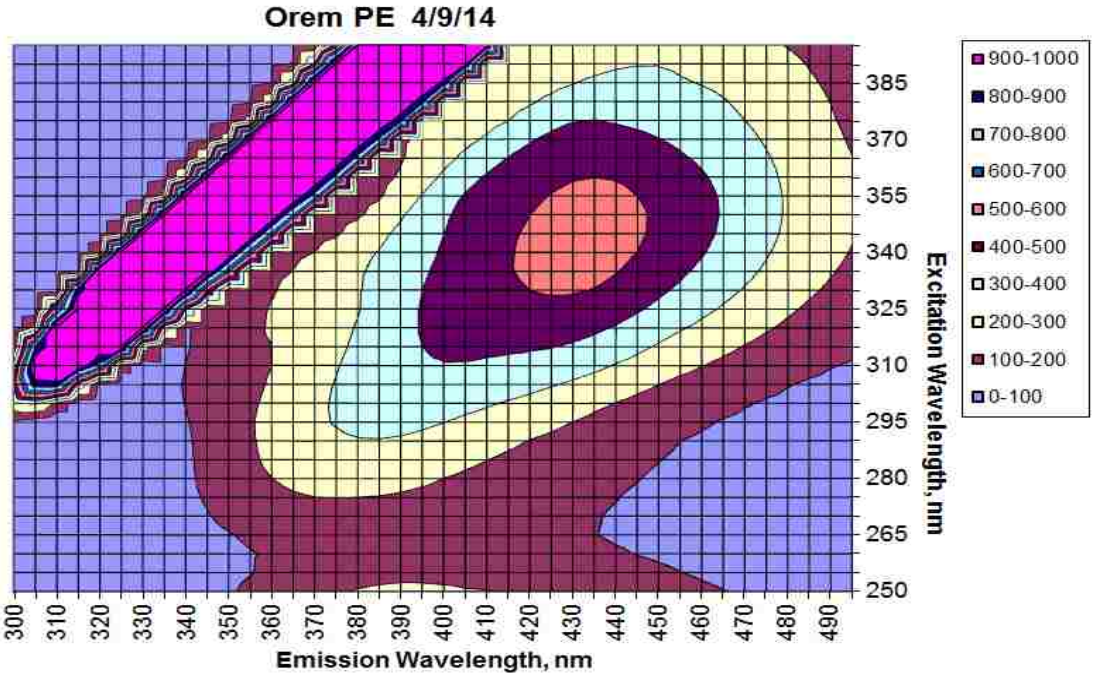


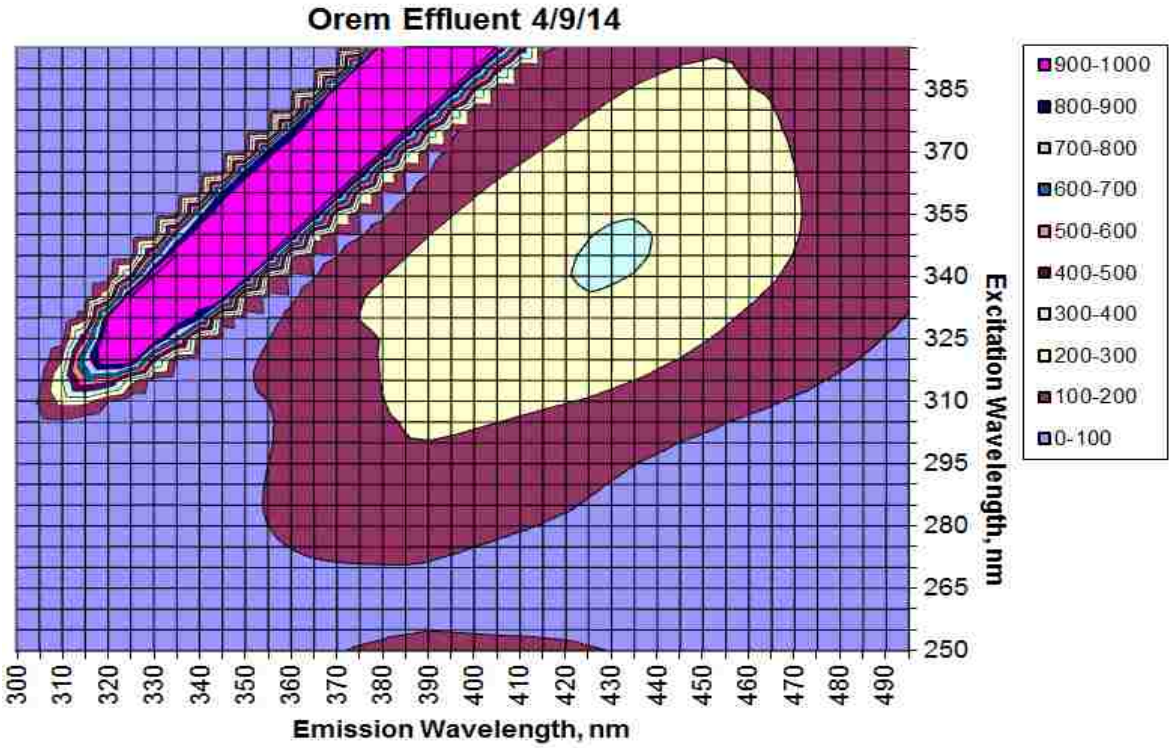


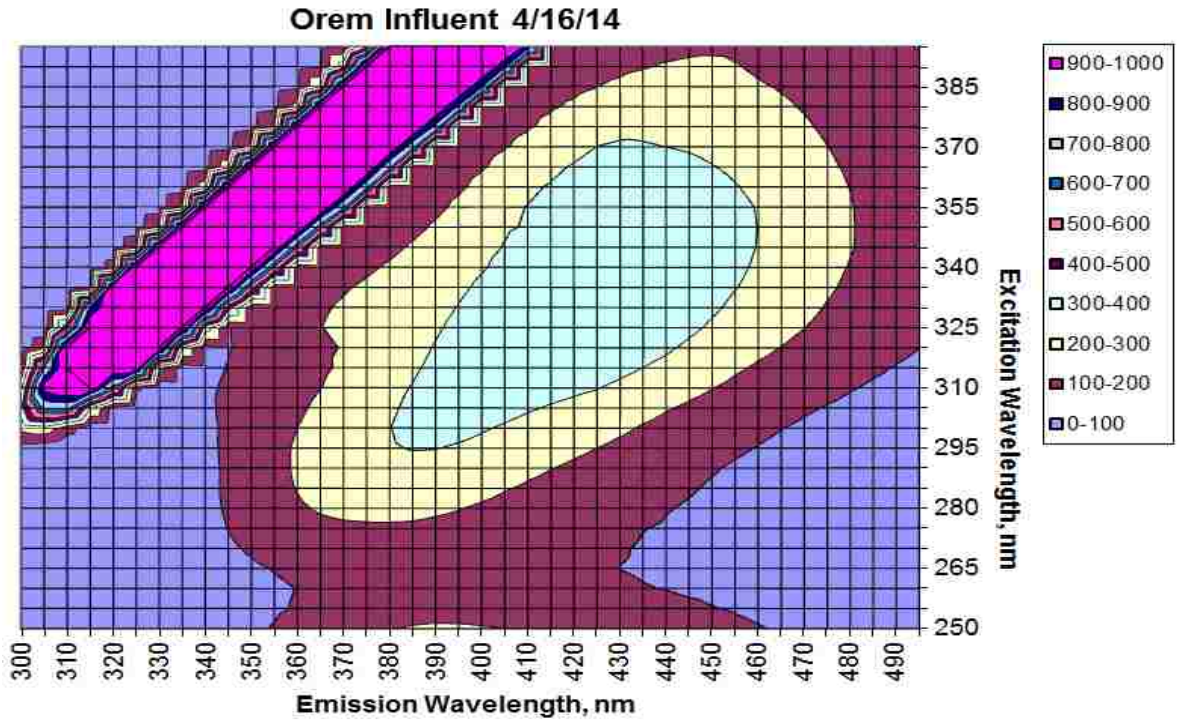


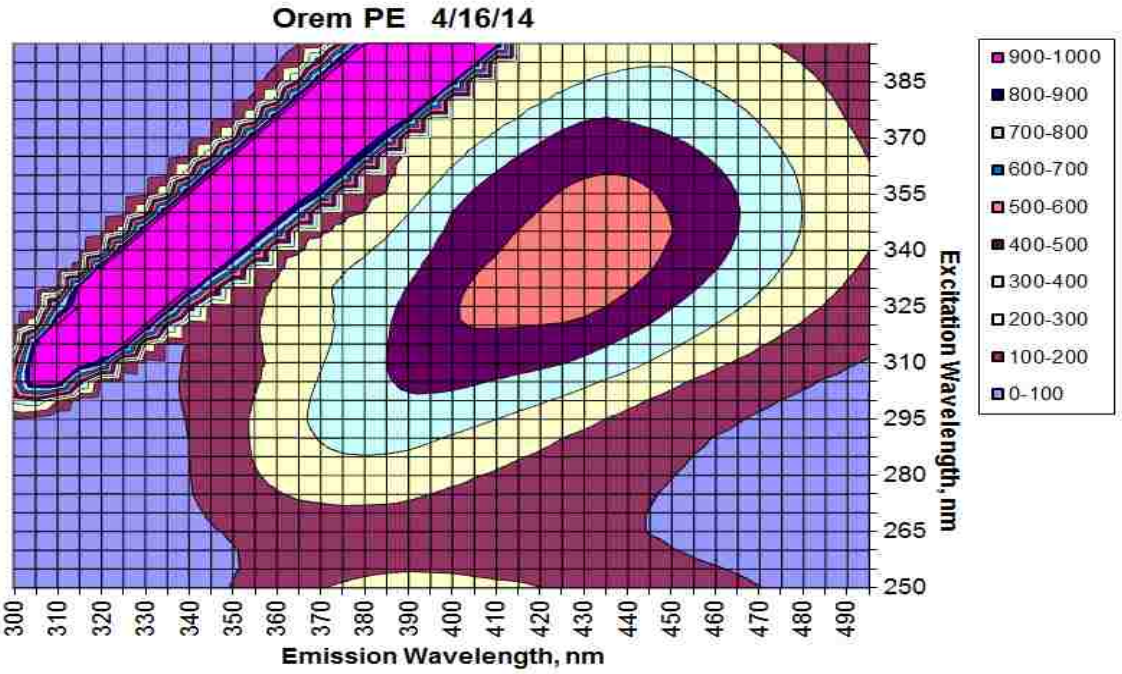


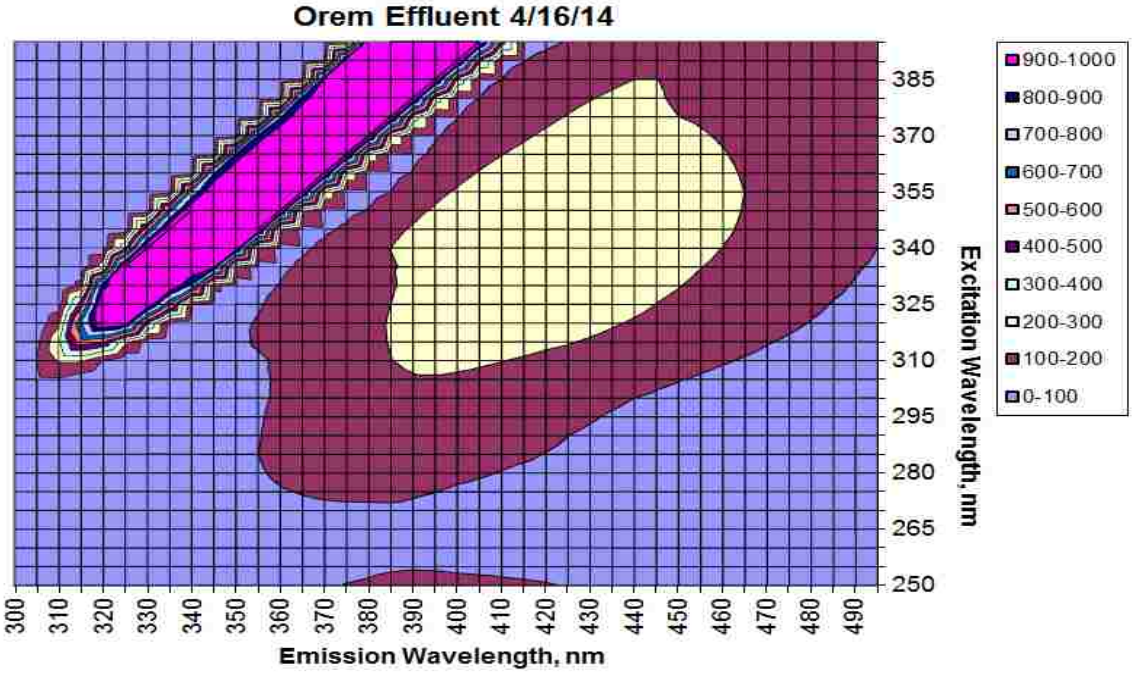


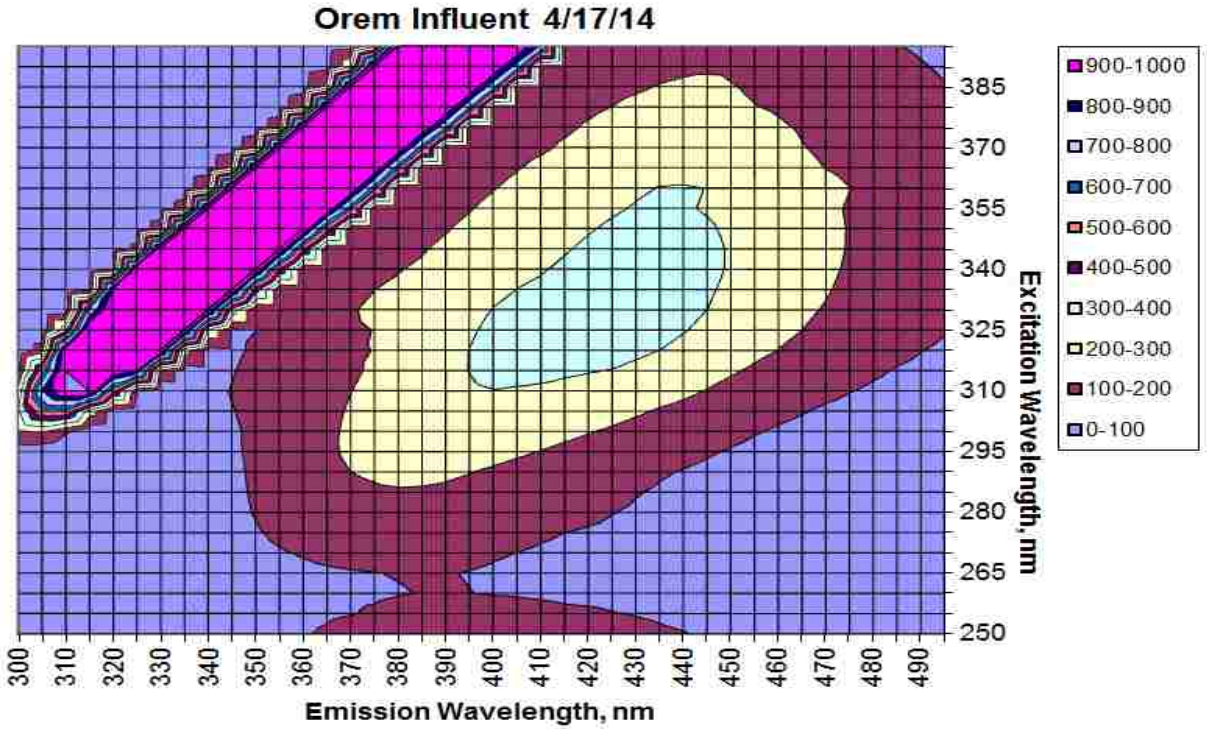


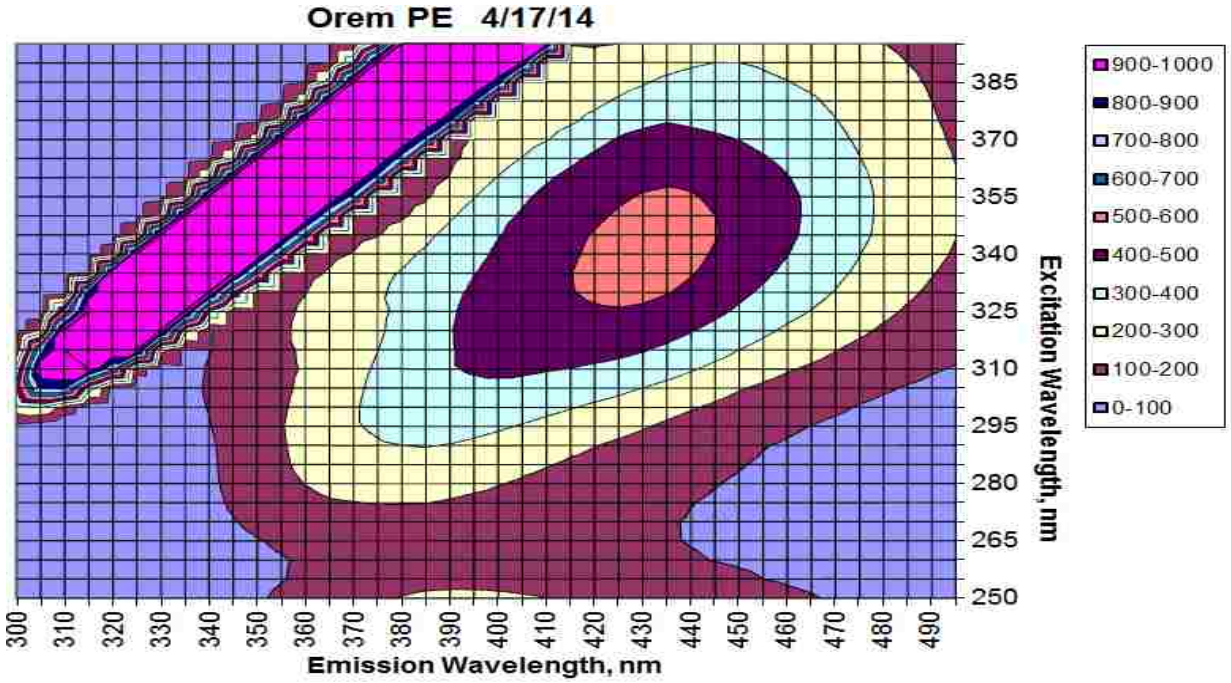


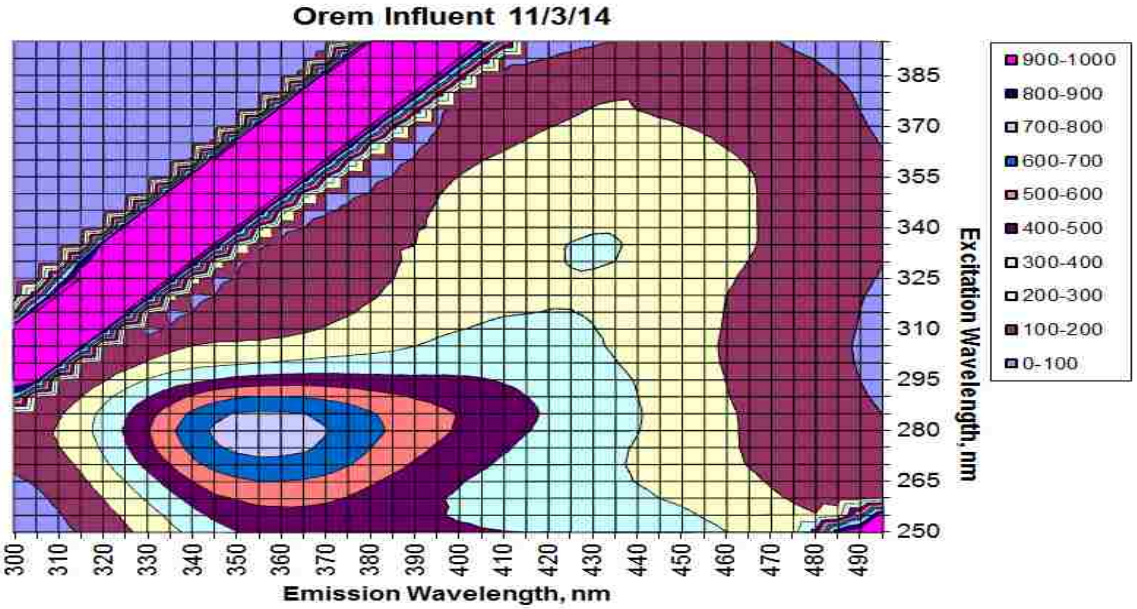


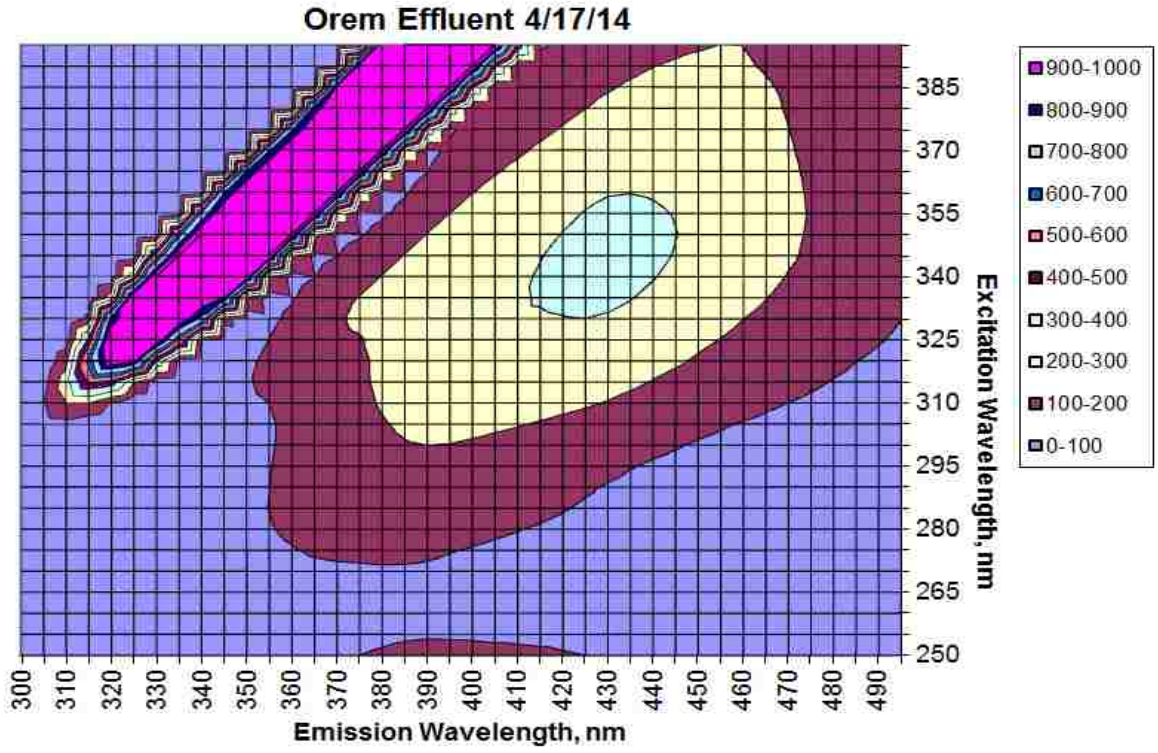




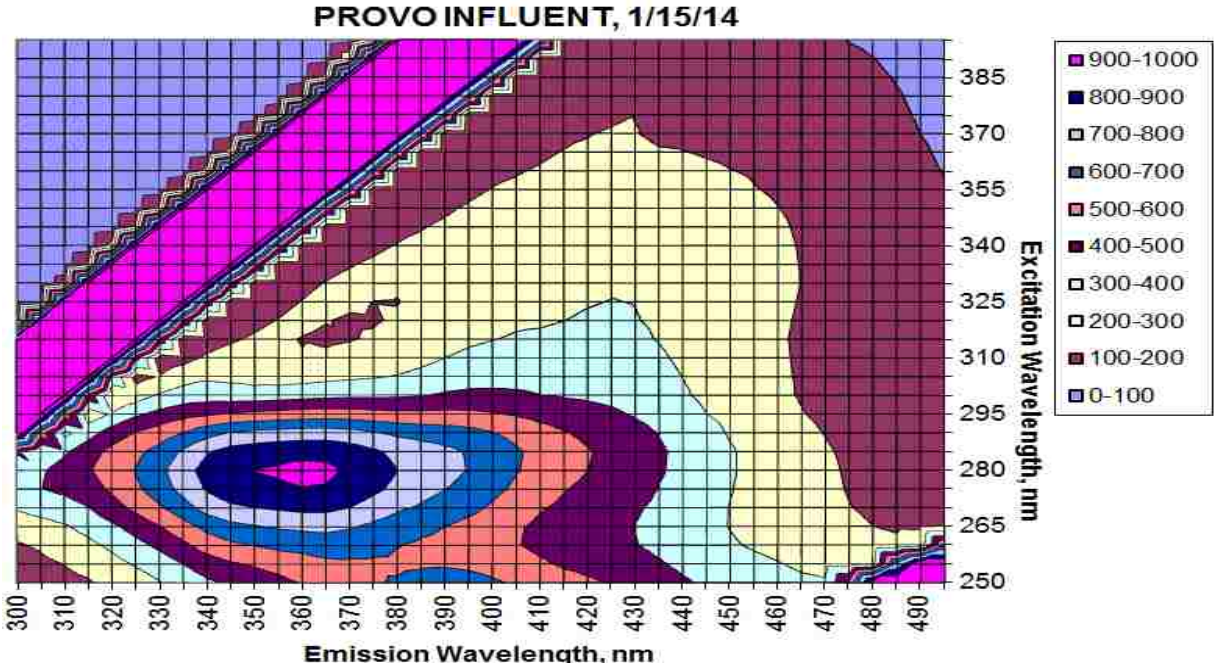


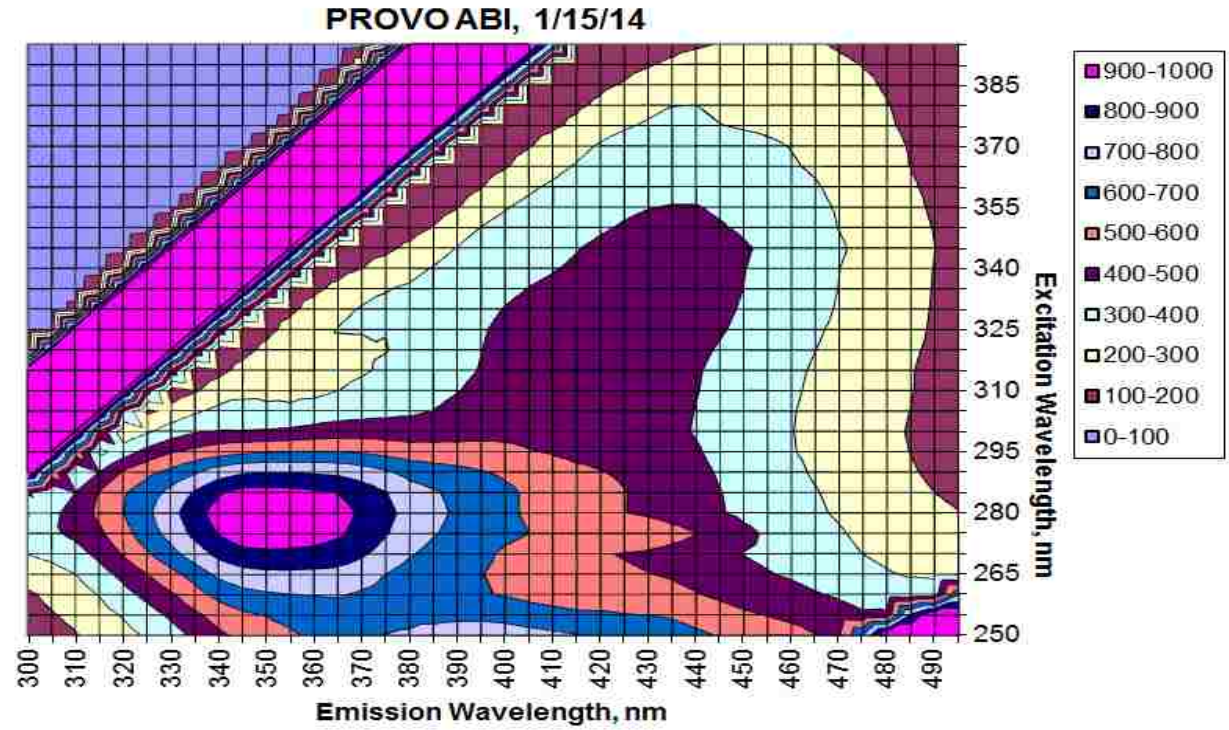


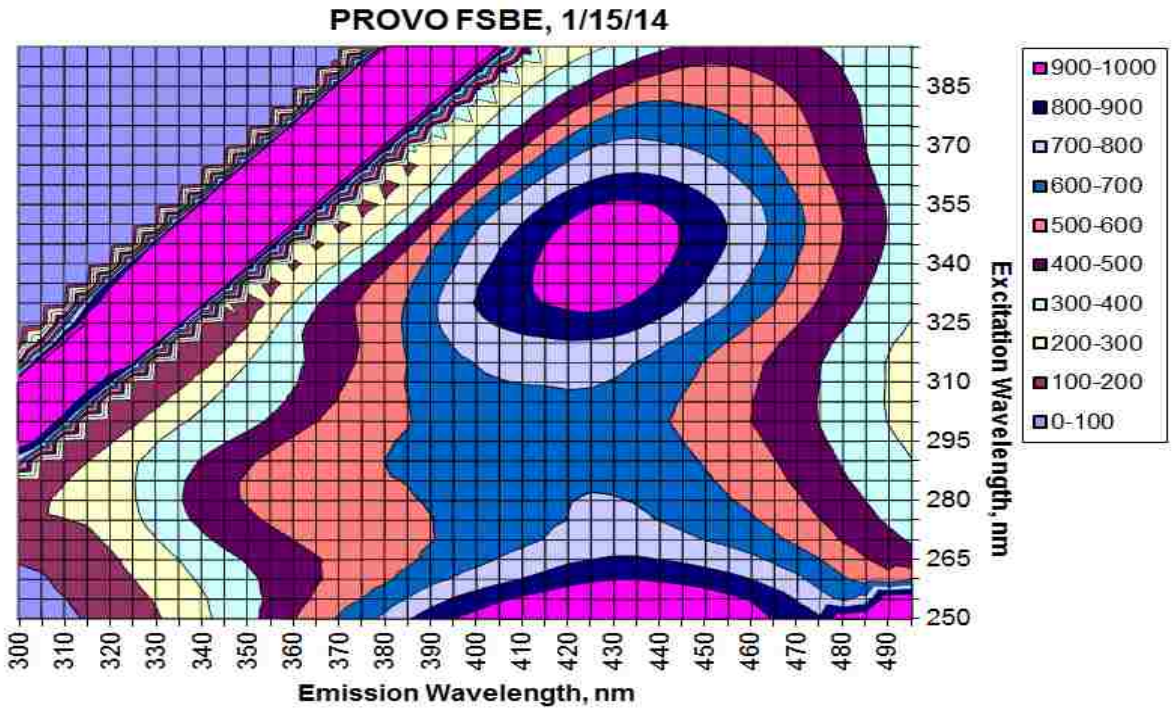


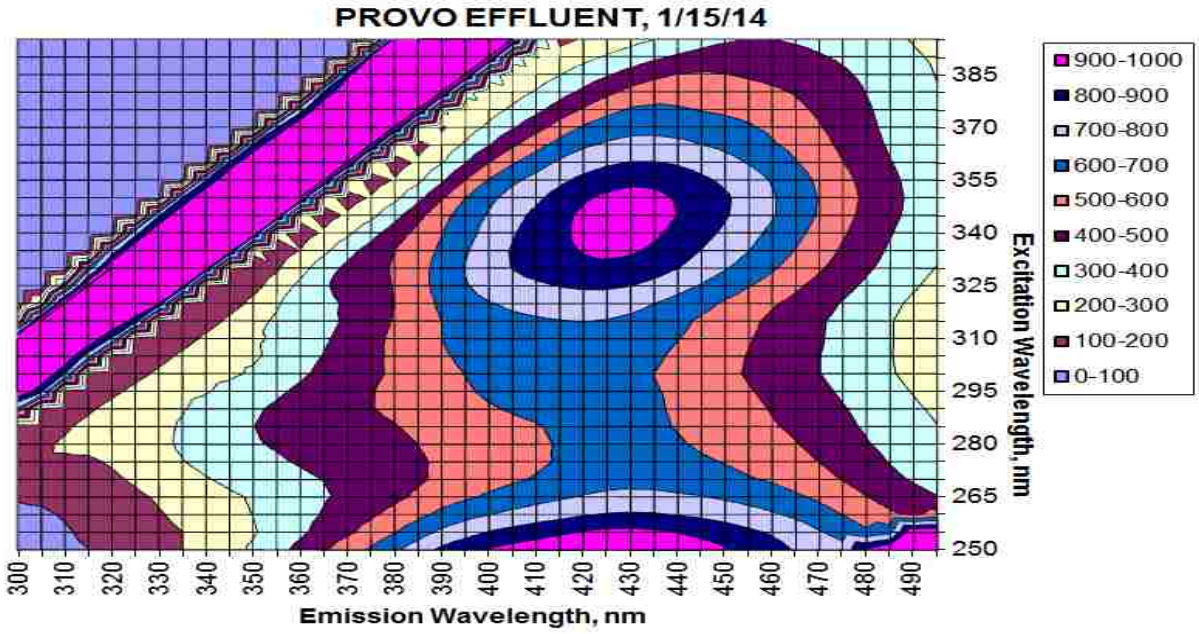


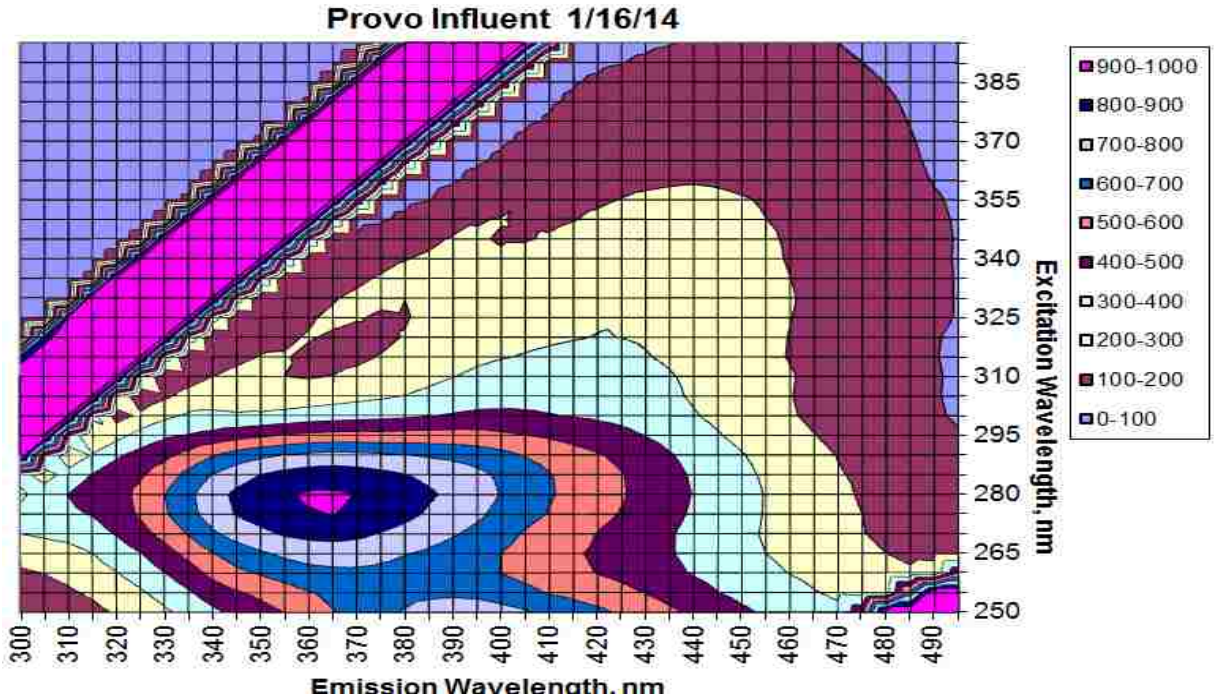
A.4 Provo Emission Excitation Matrix Data

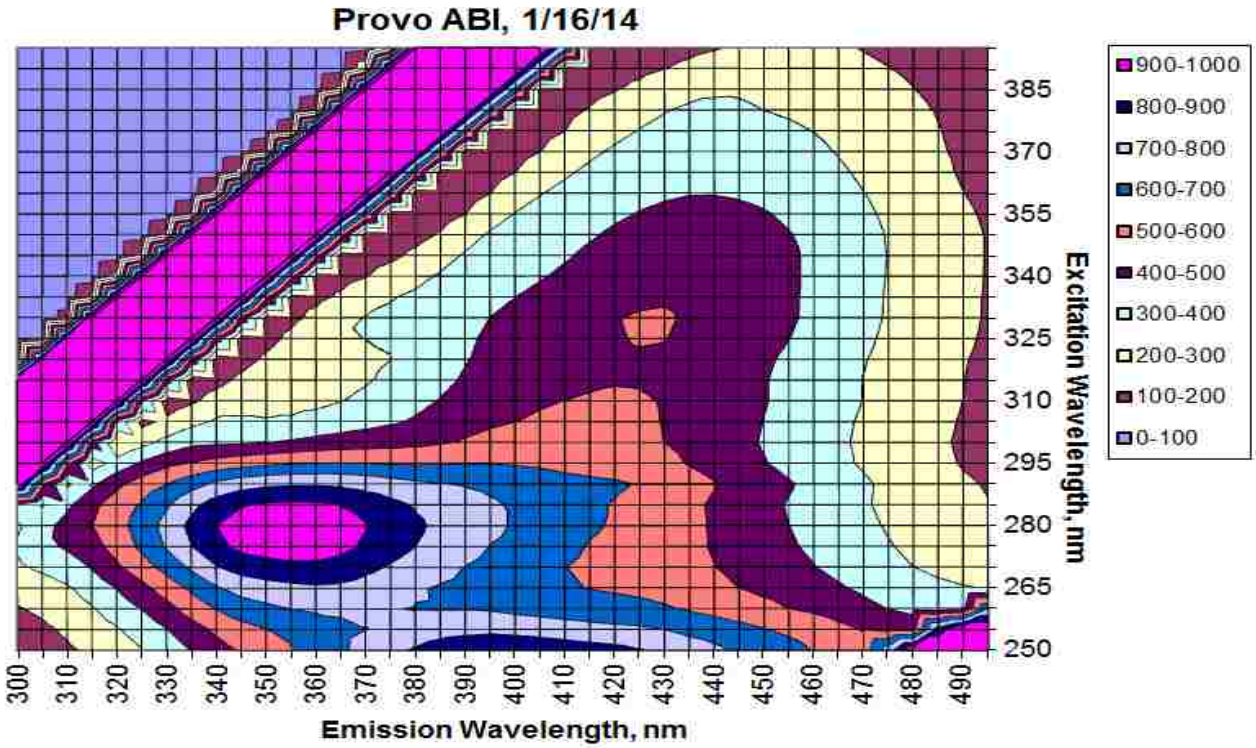


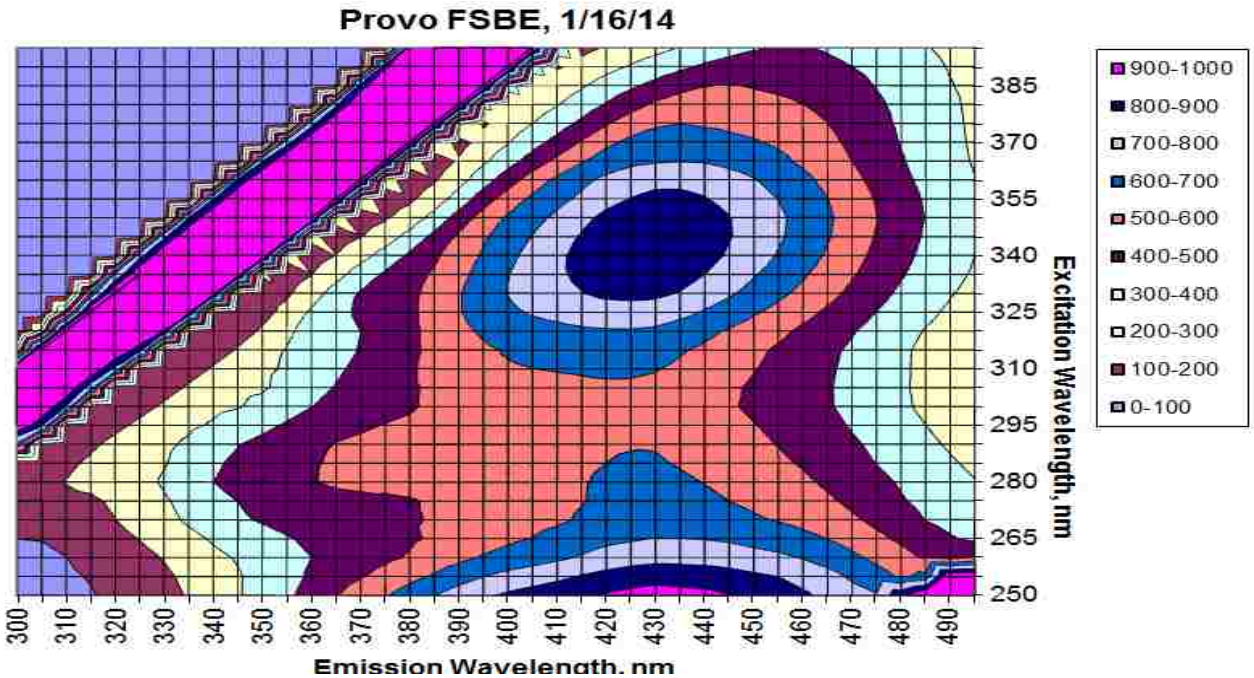


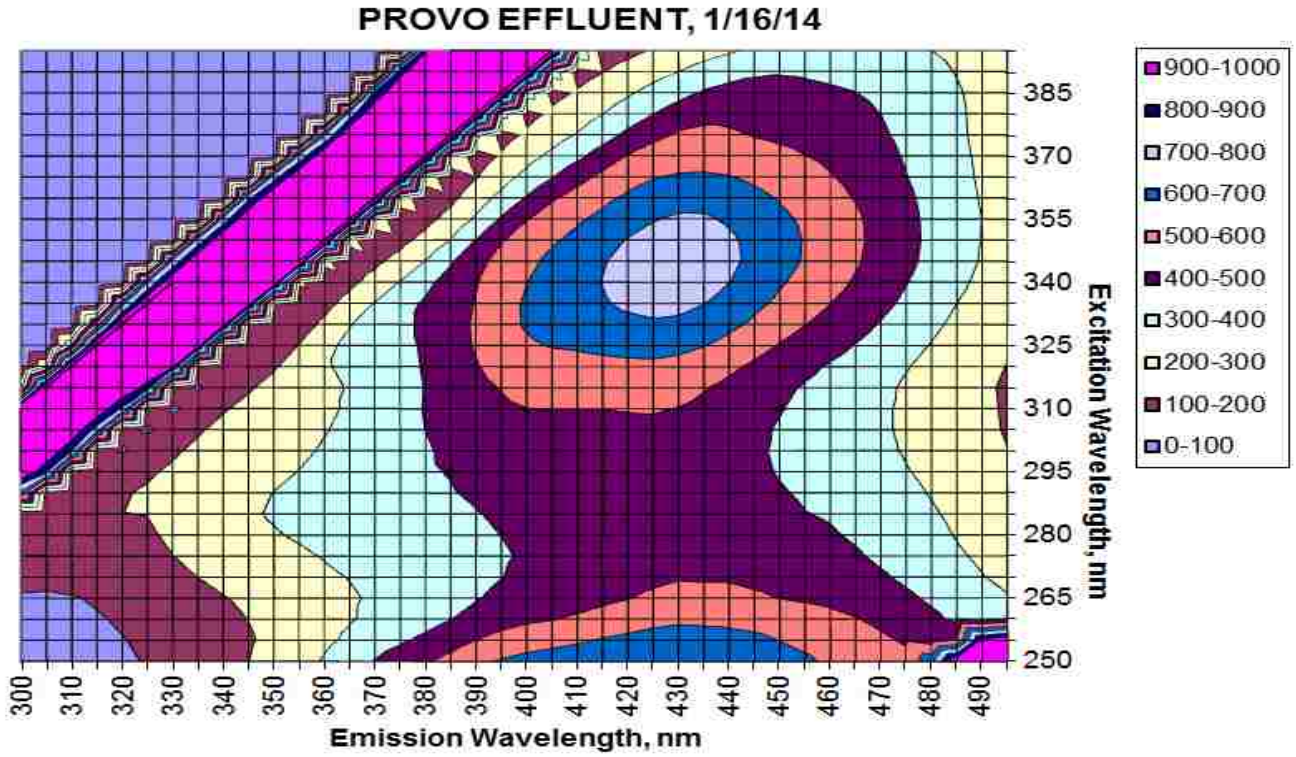


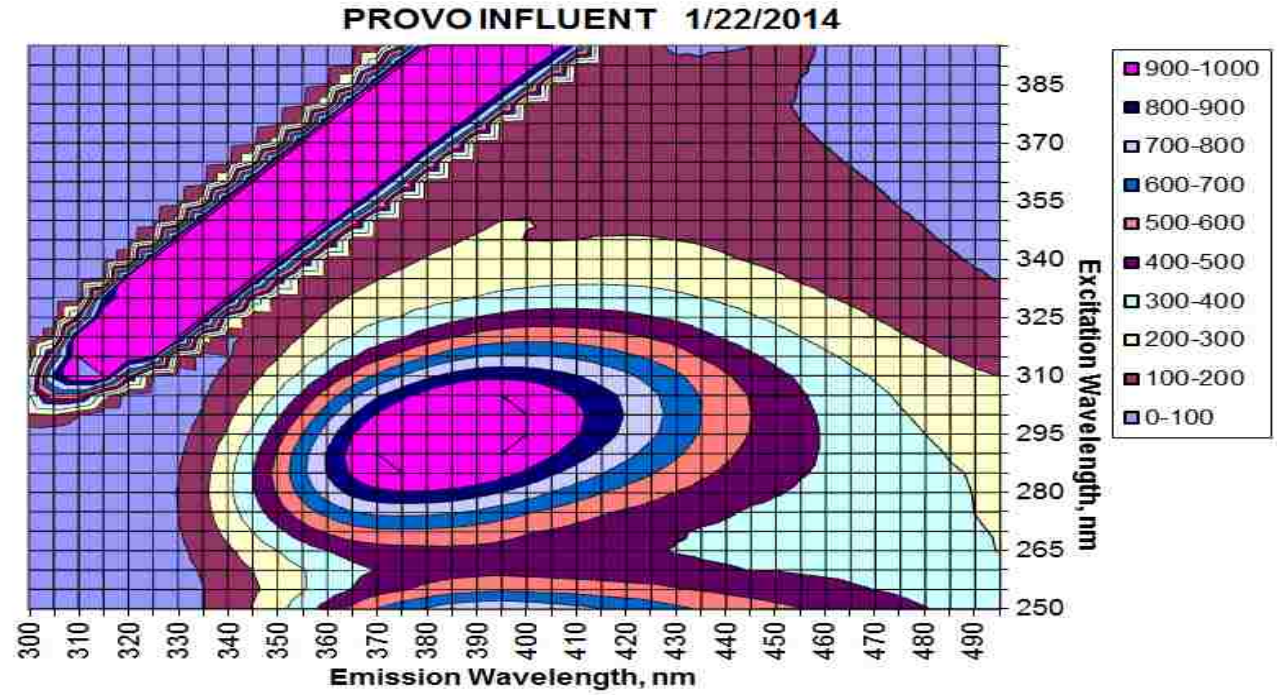


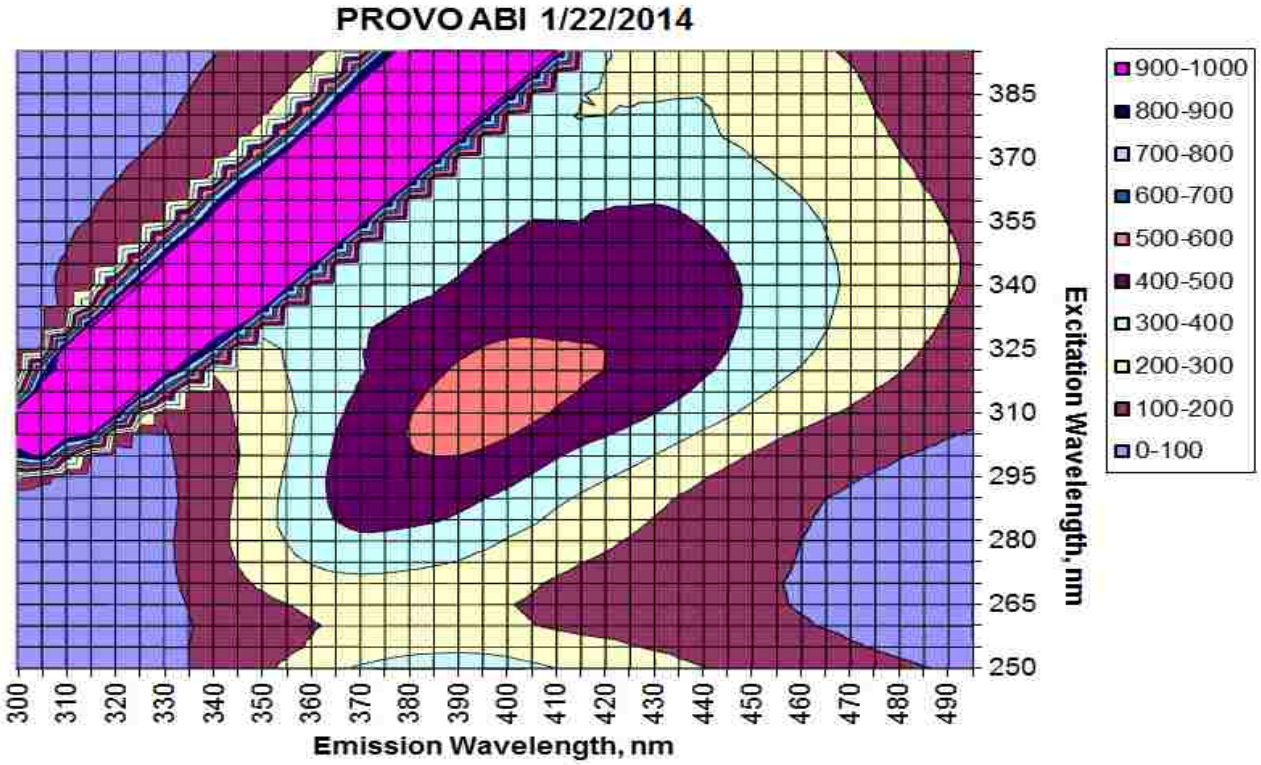


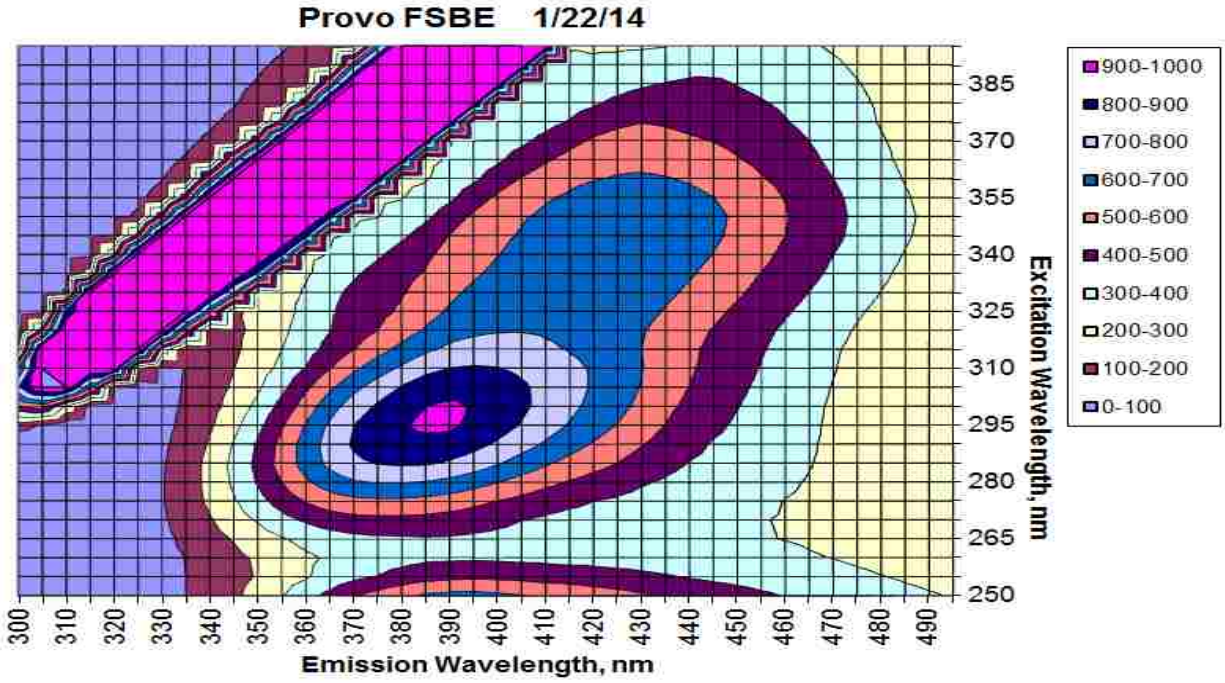


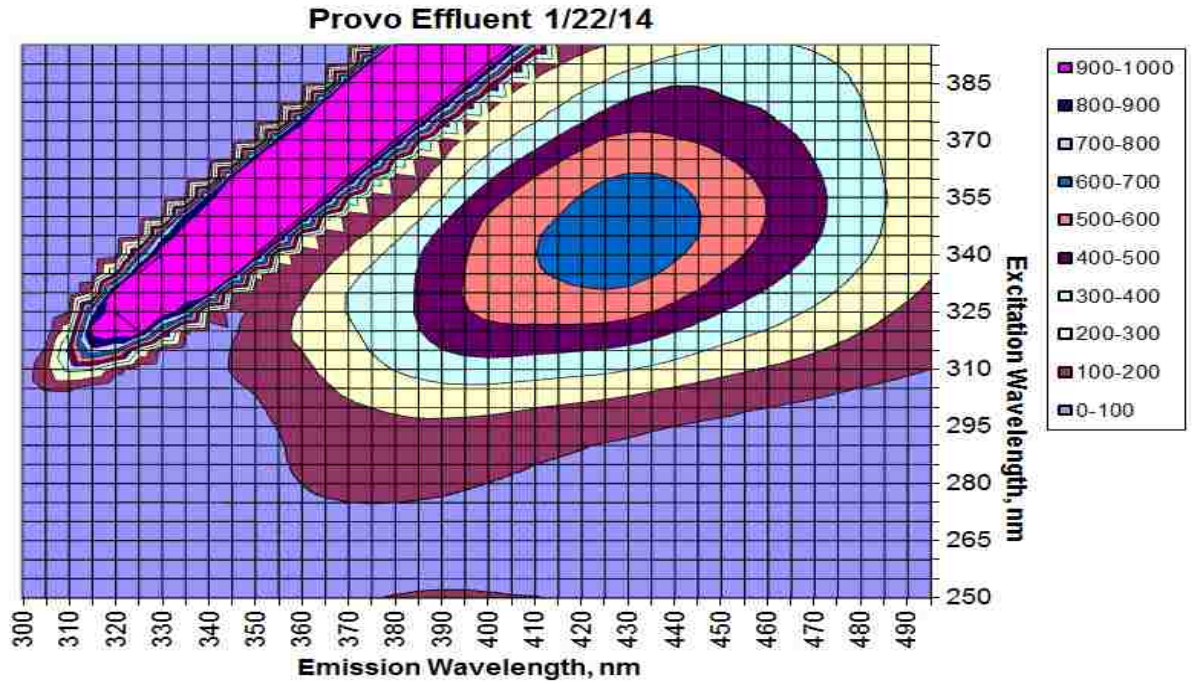


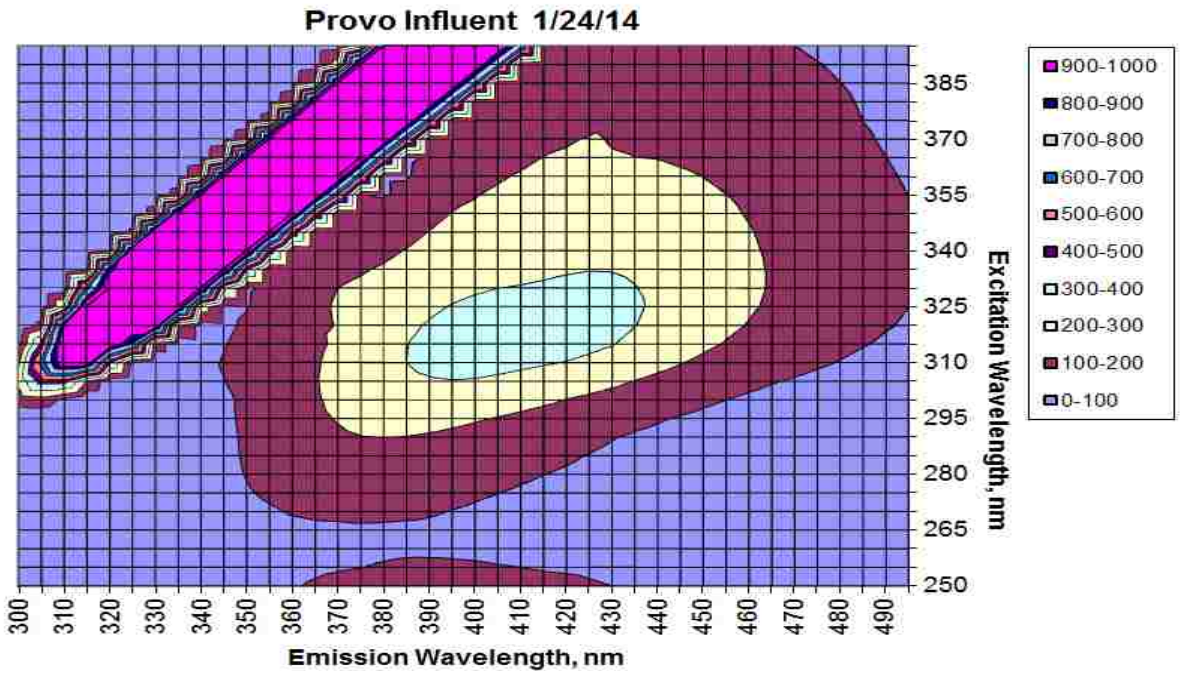


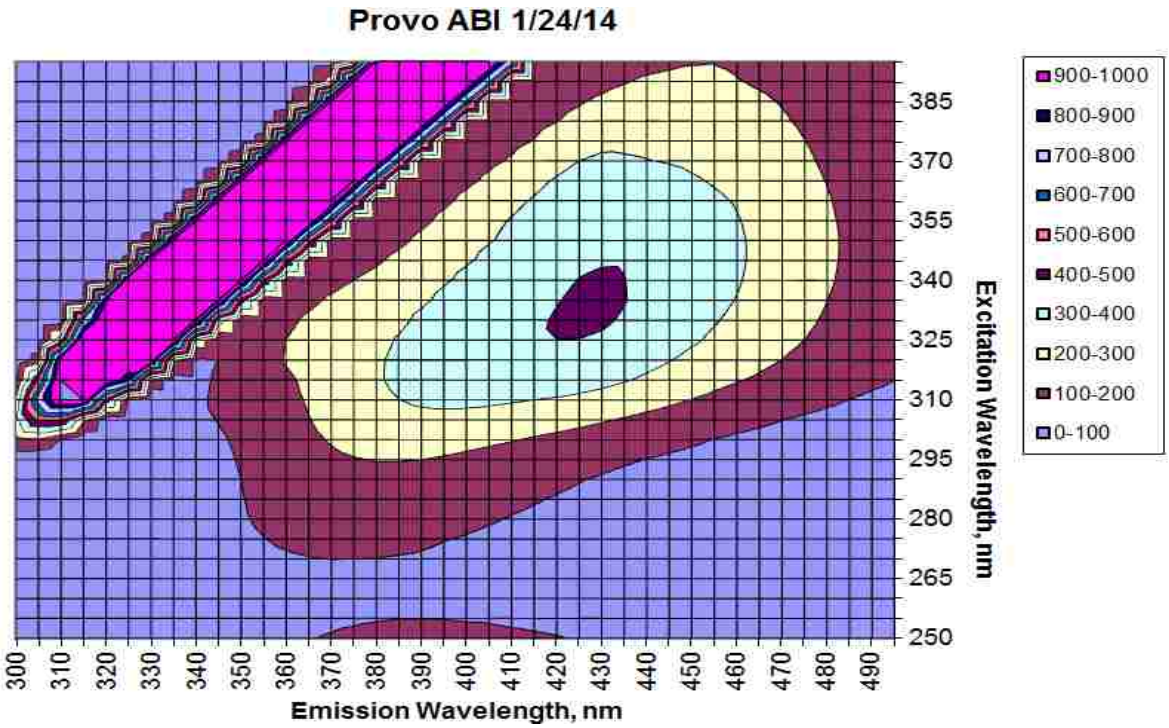


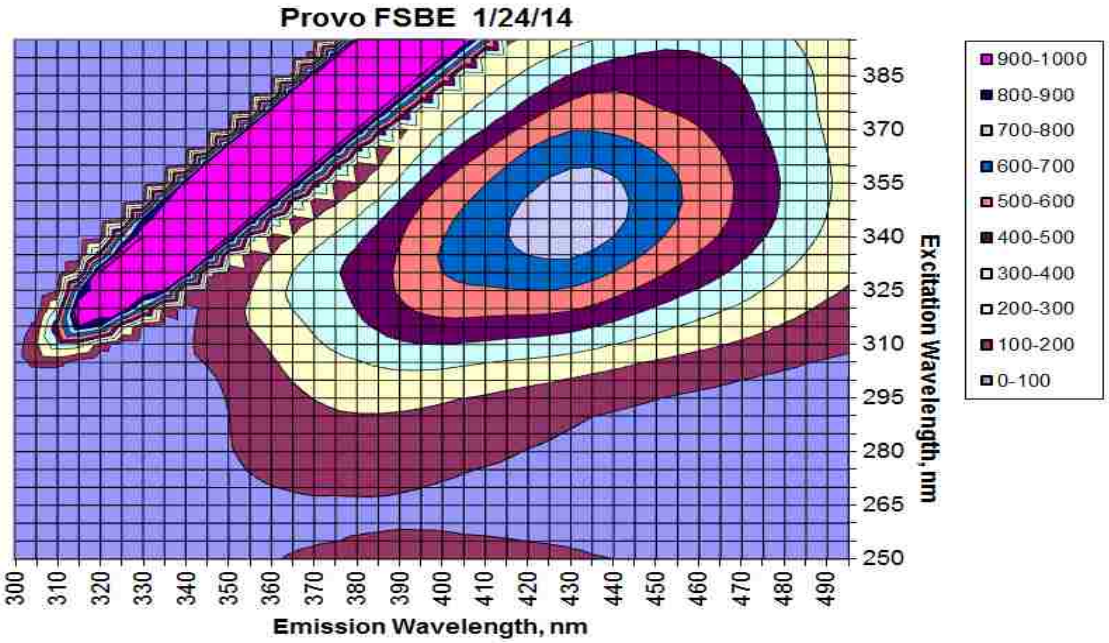


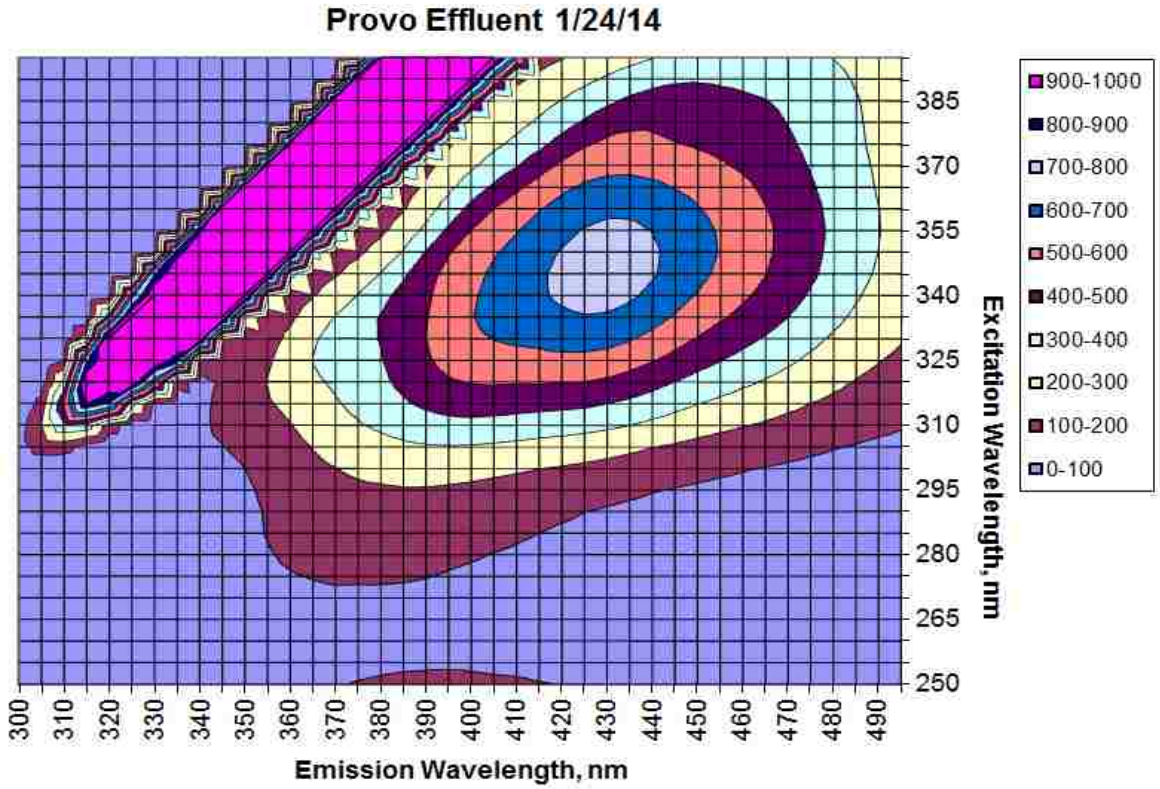


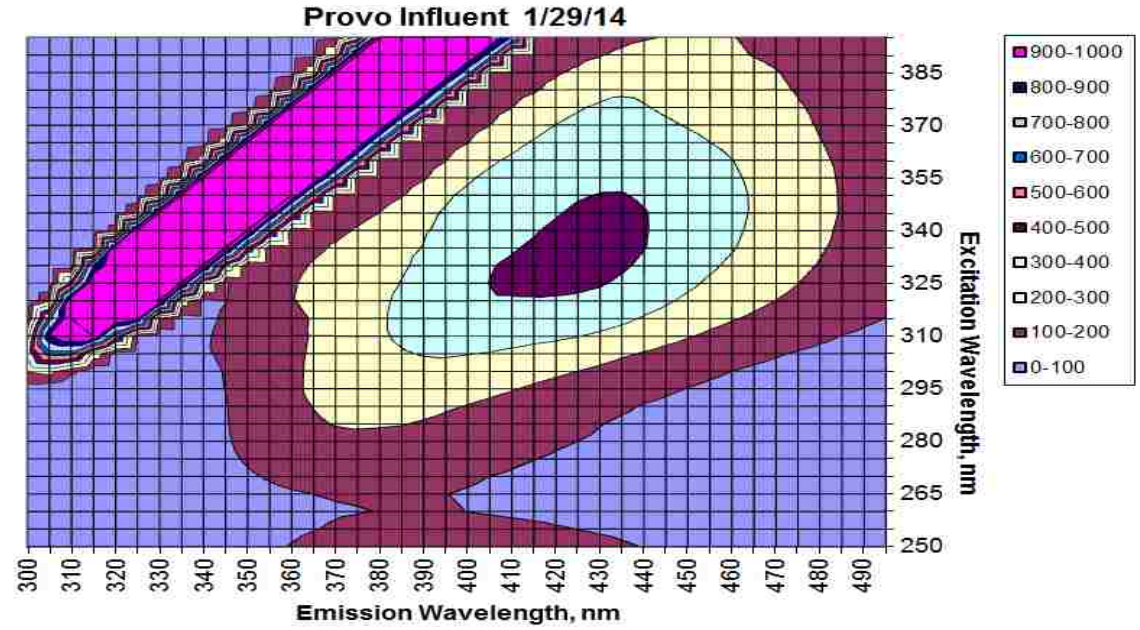


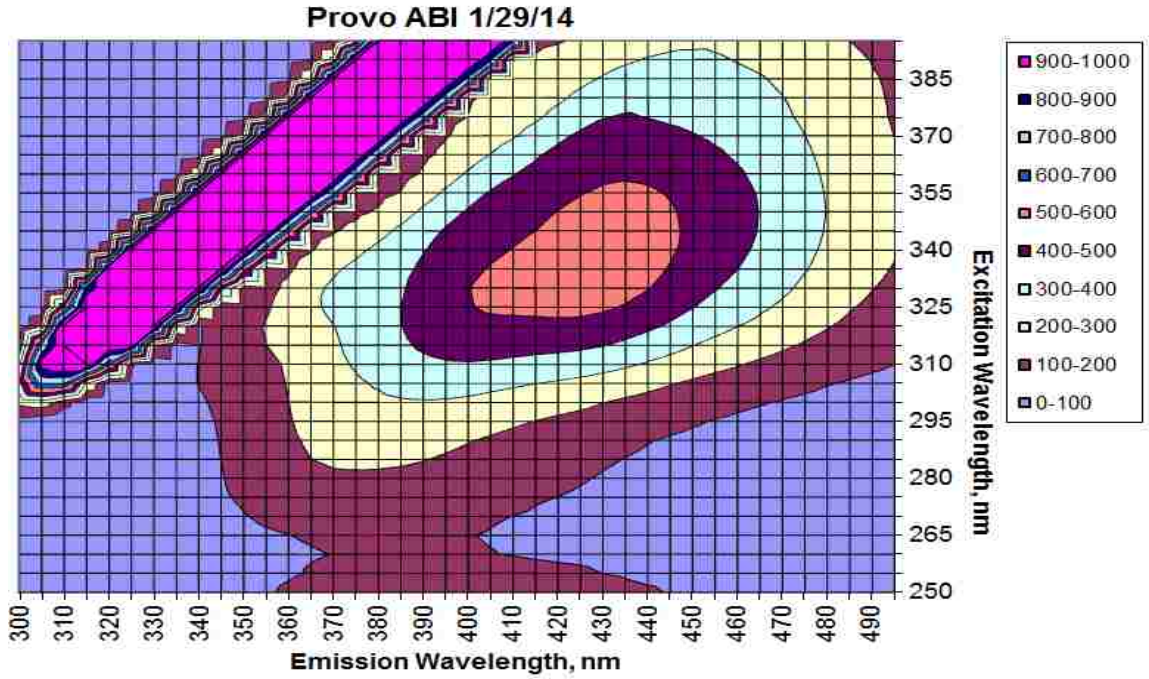


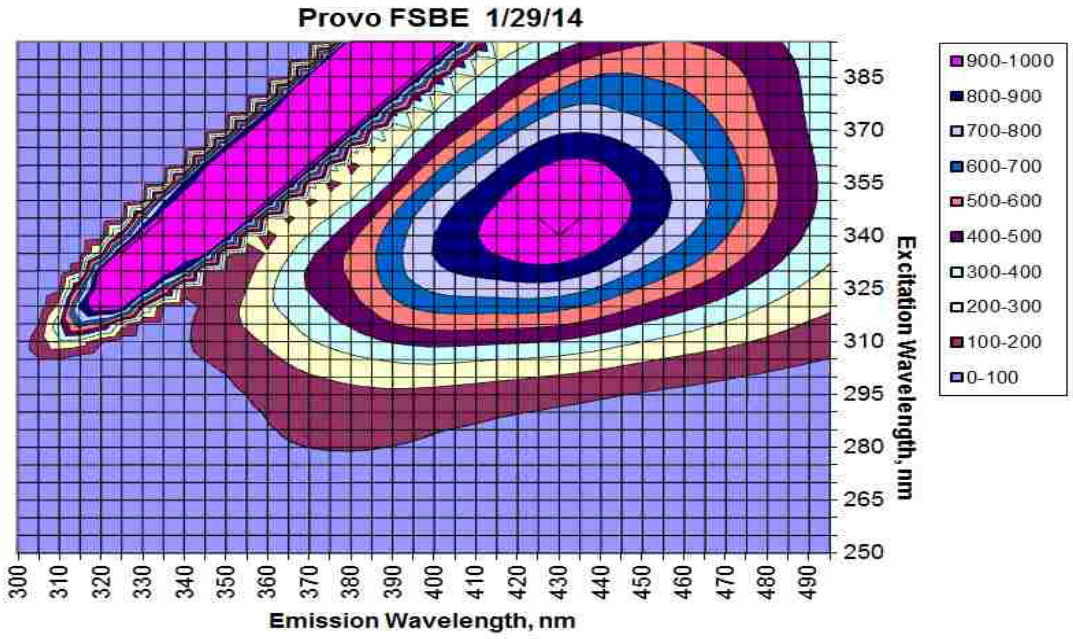


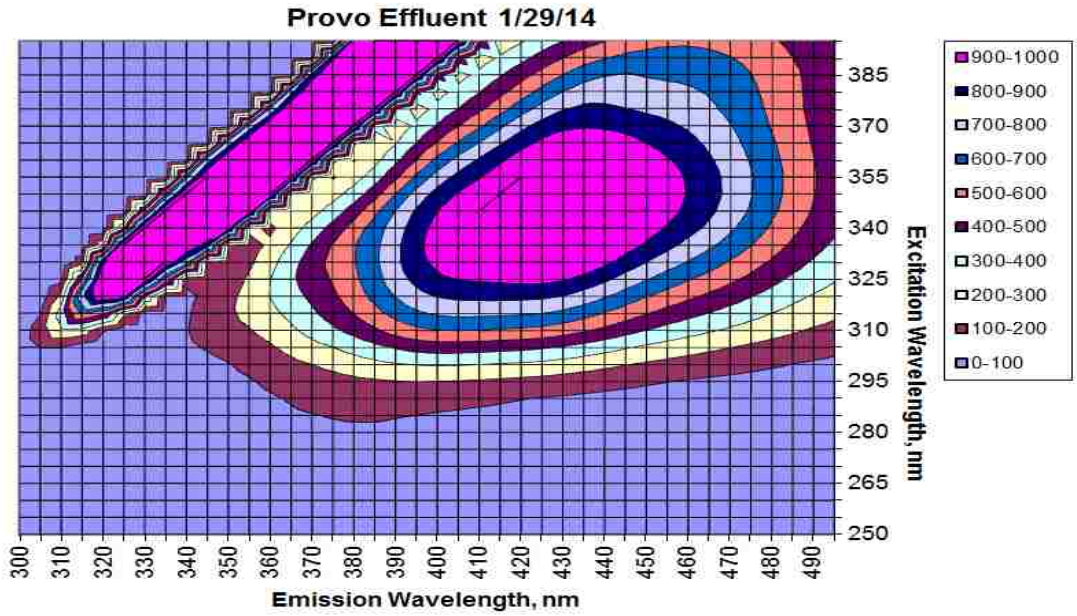


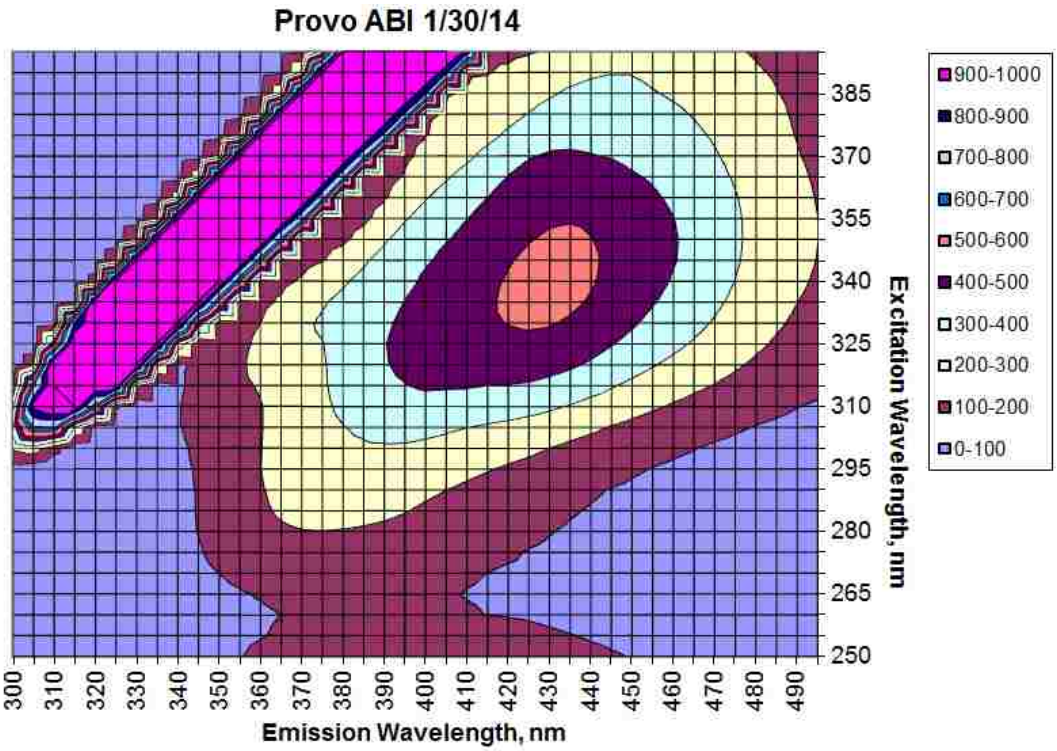


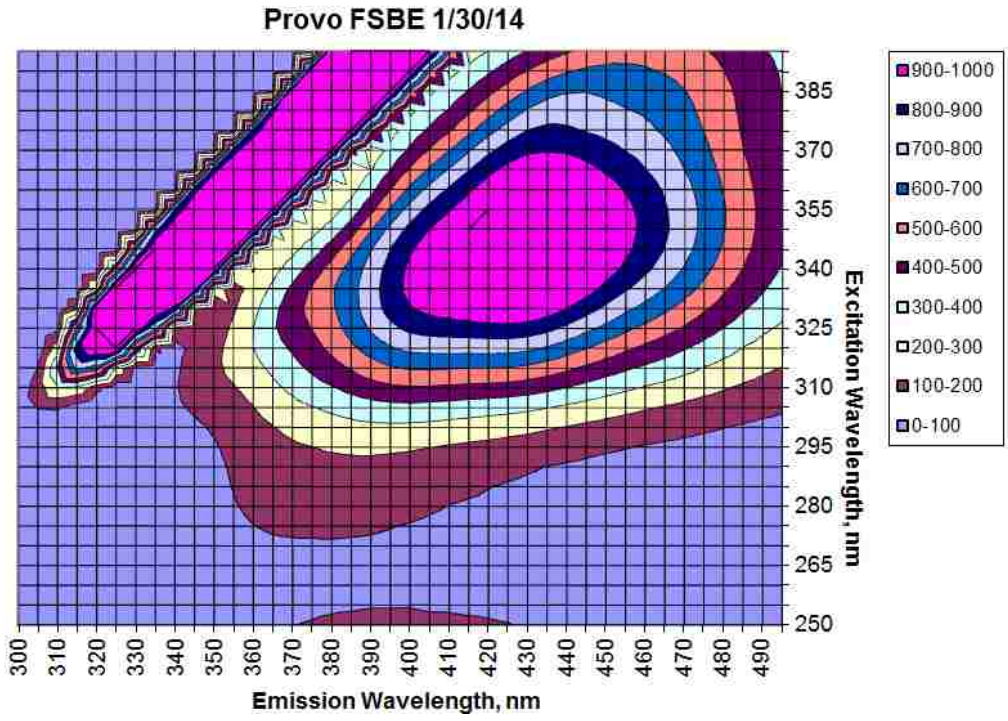


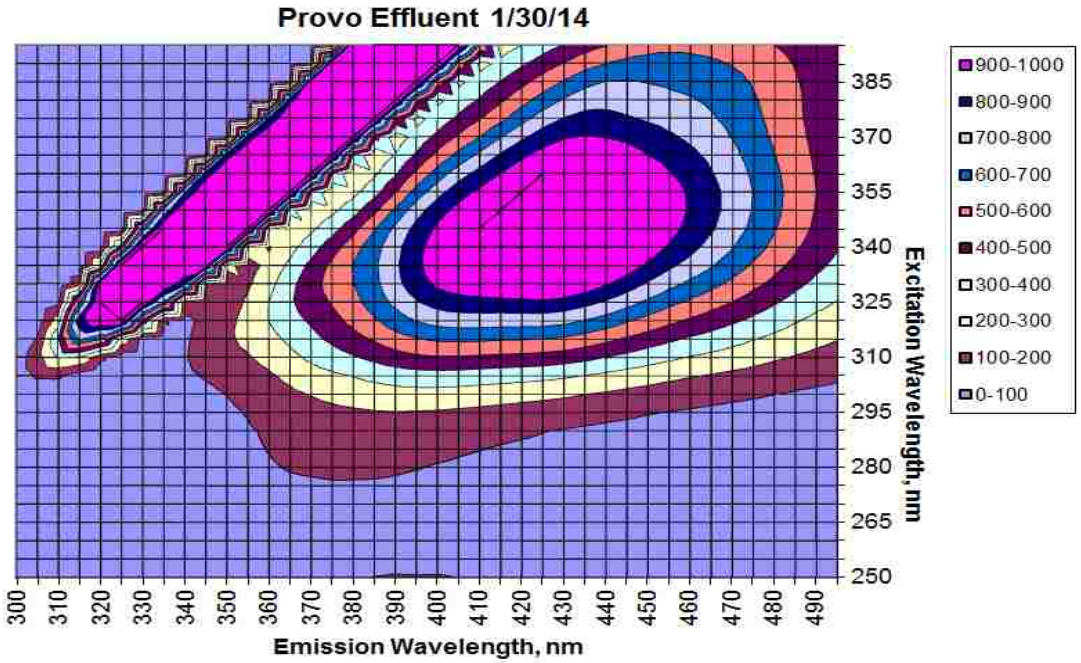


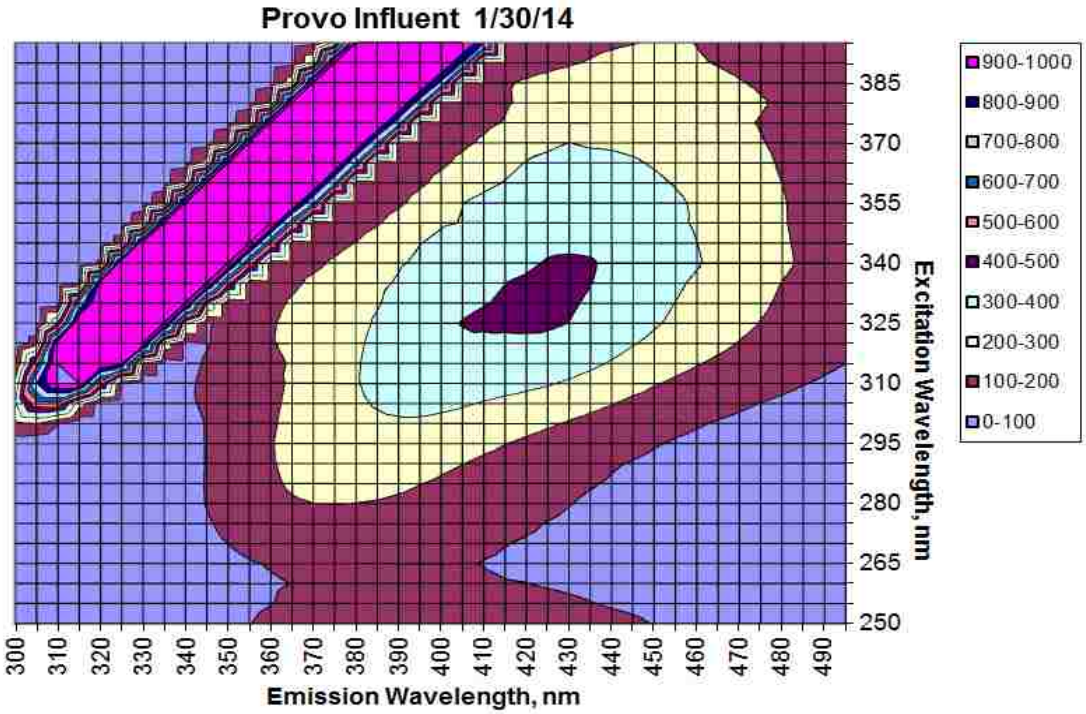


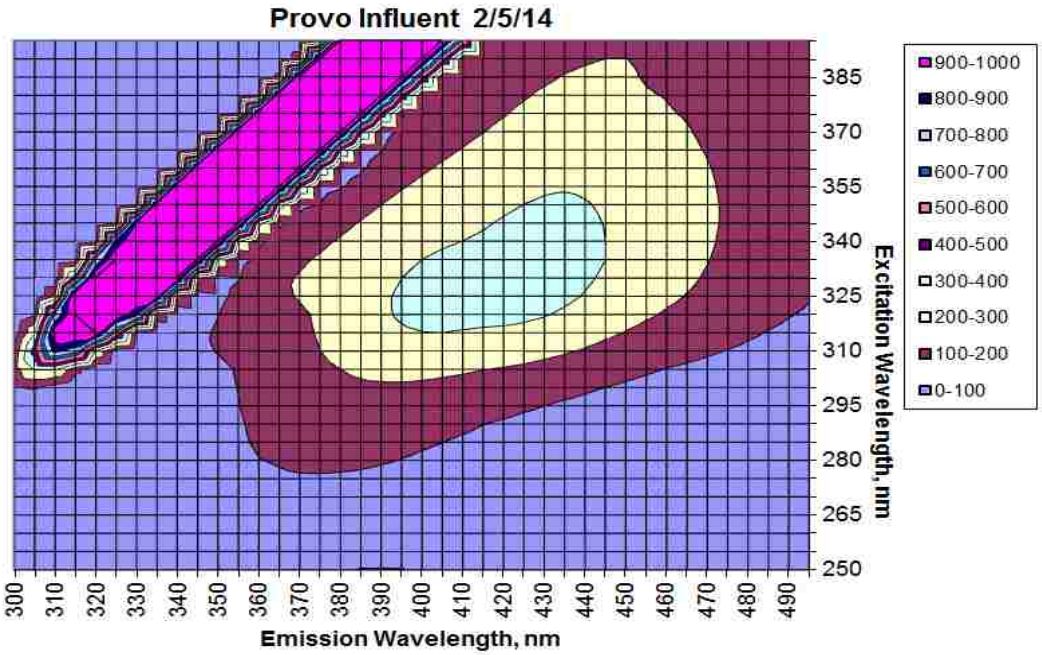


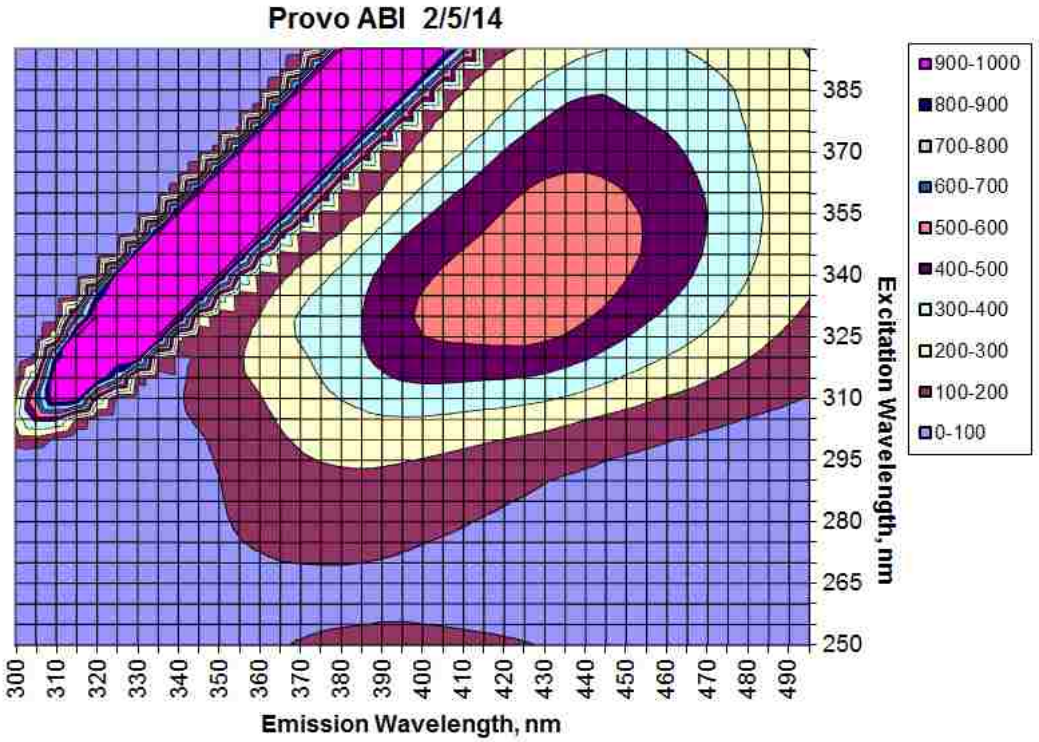


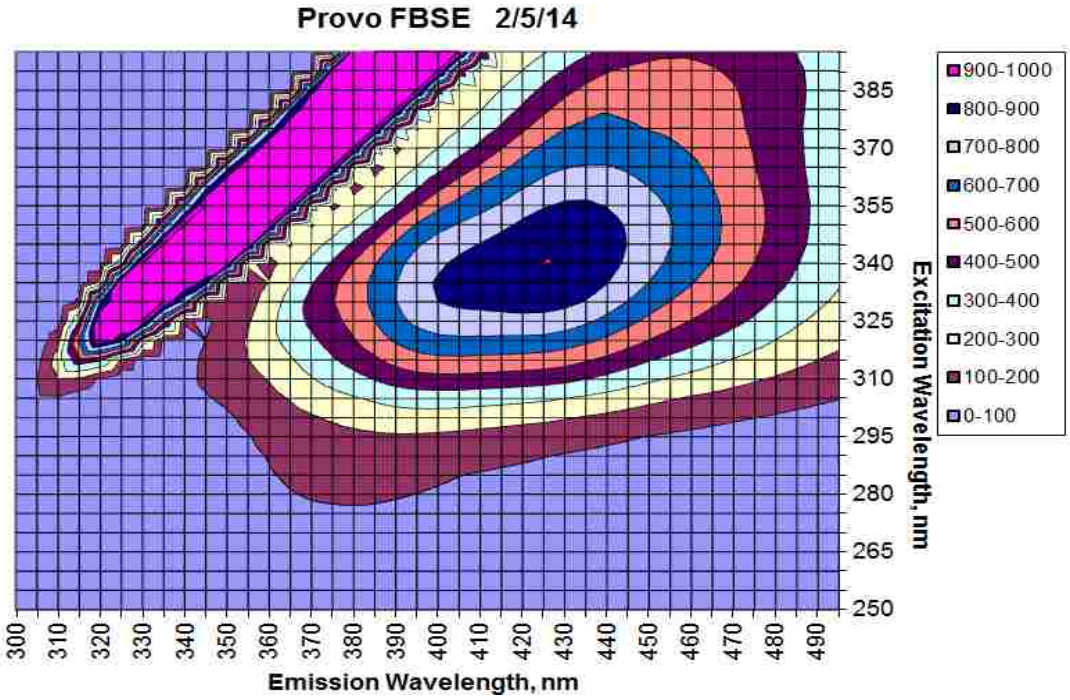


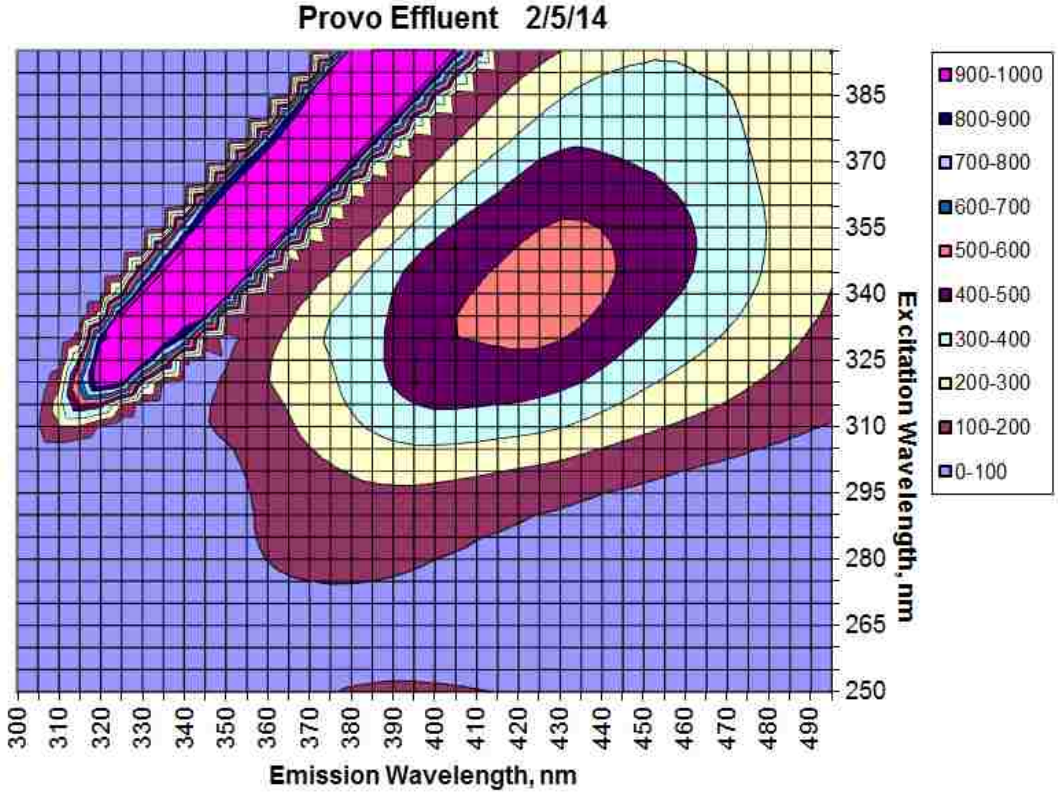


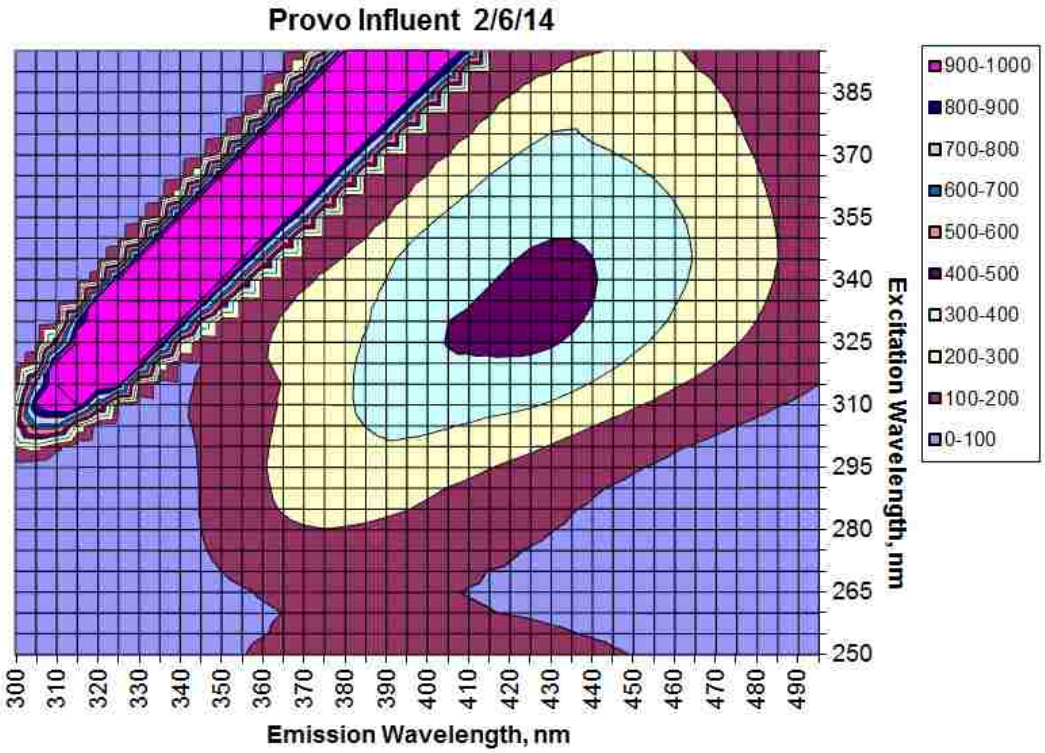


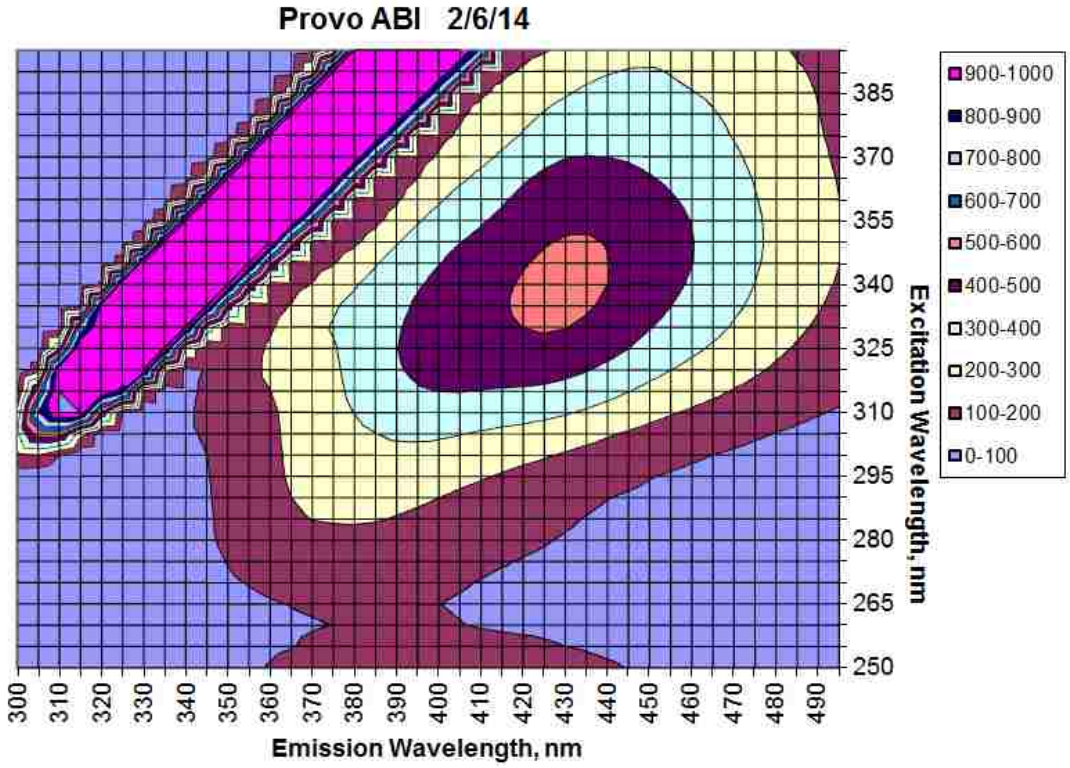


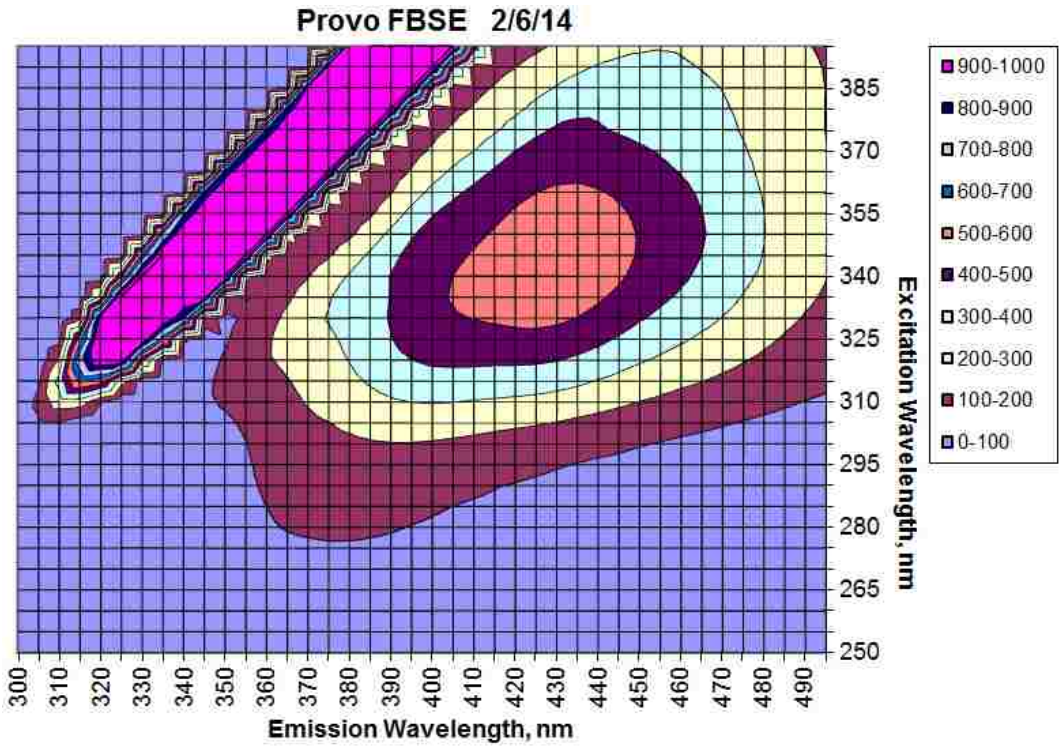


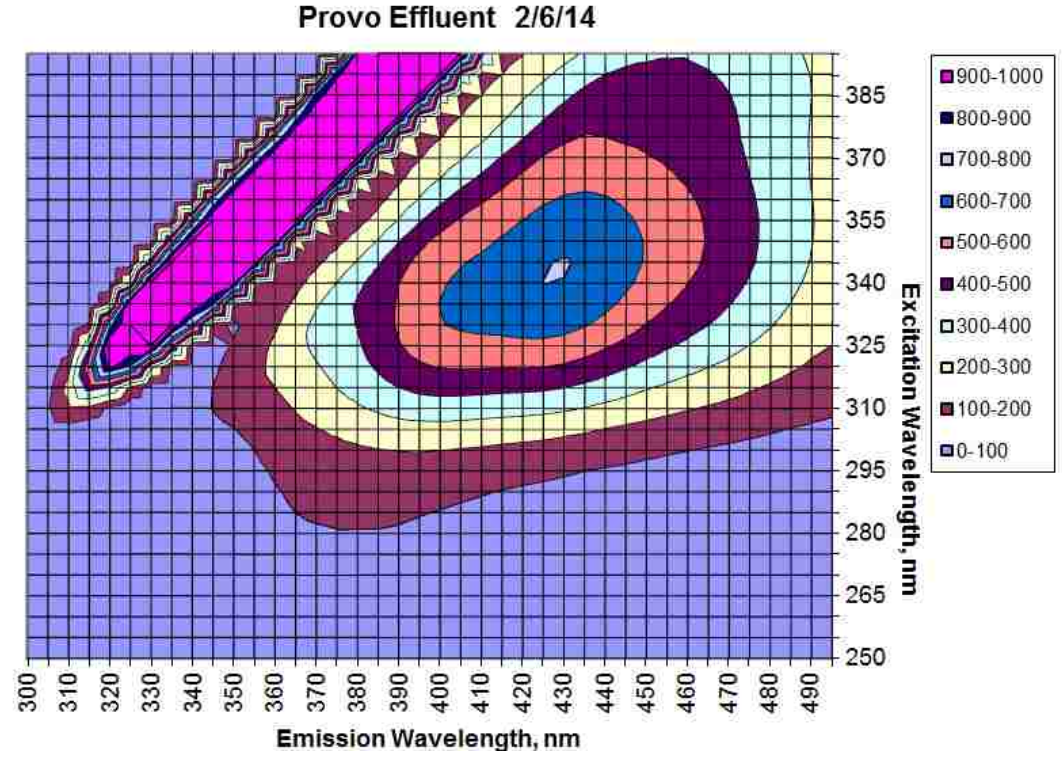


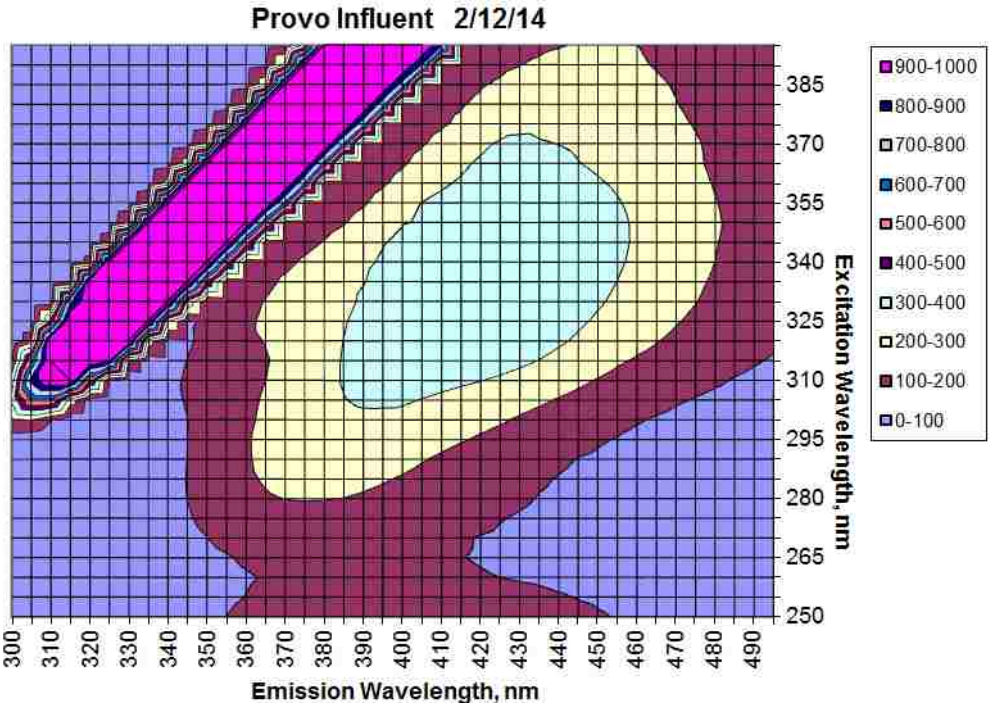


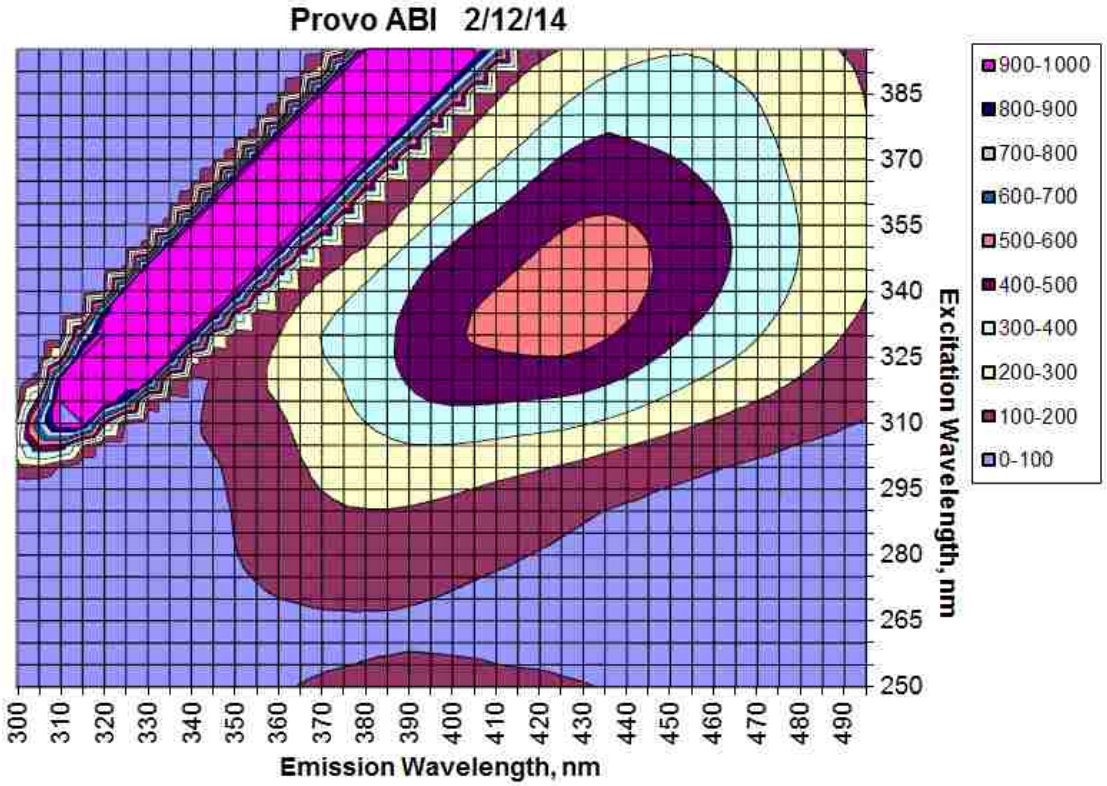


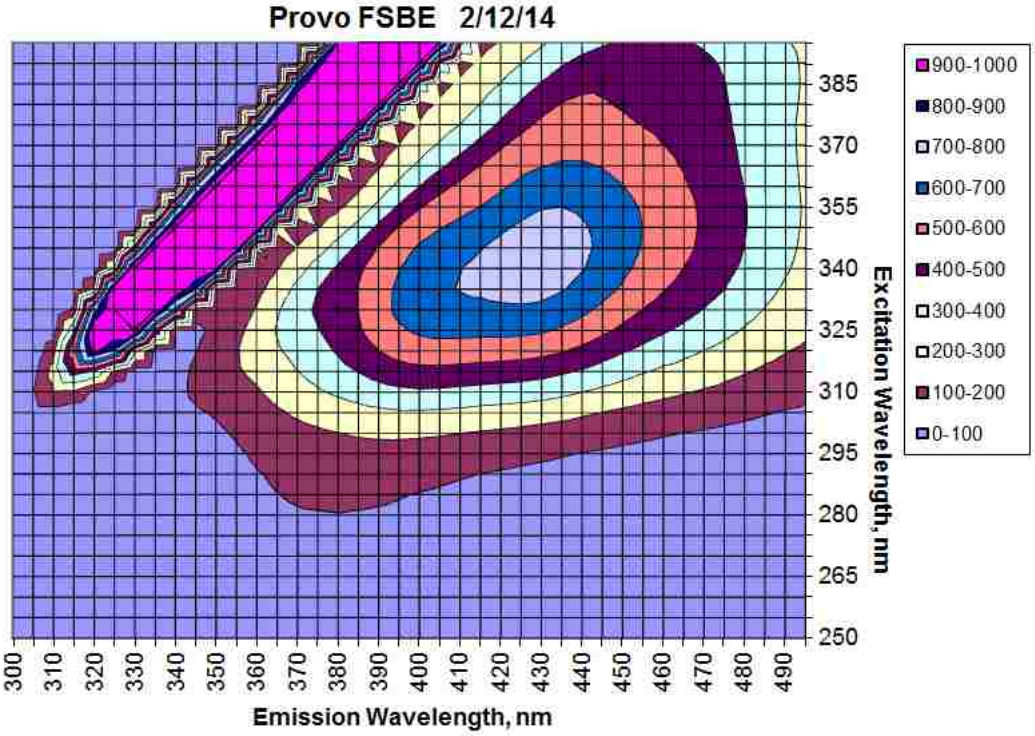


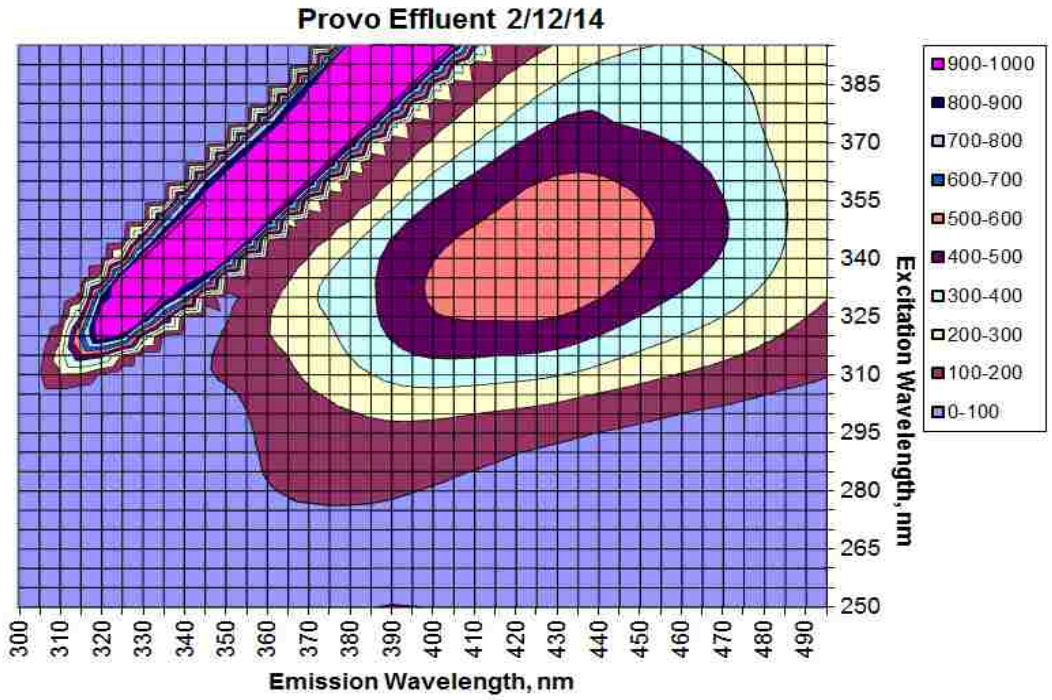


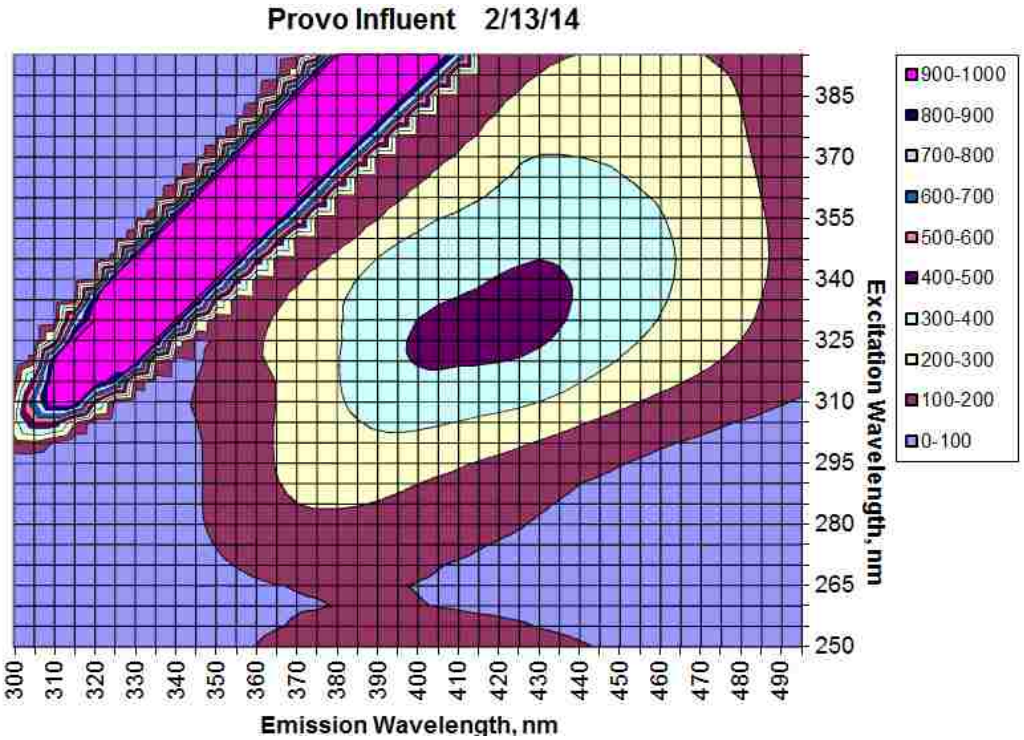


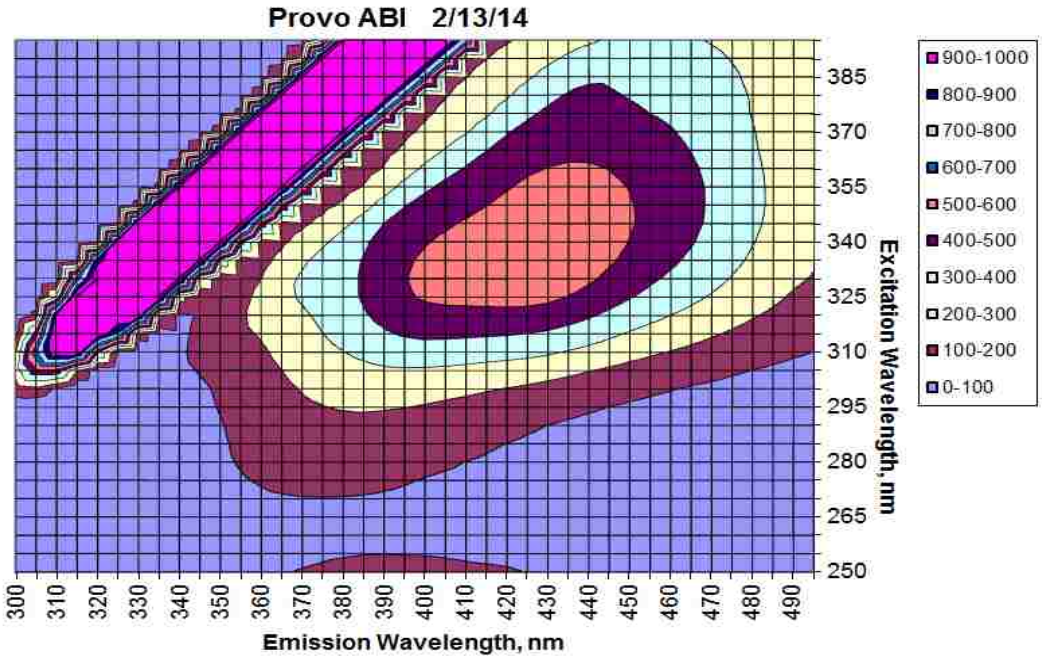


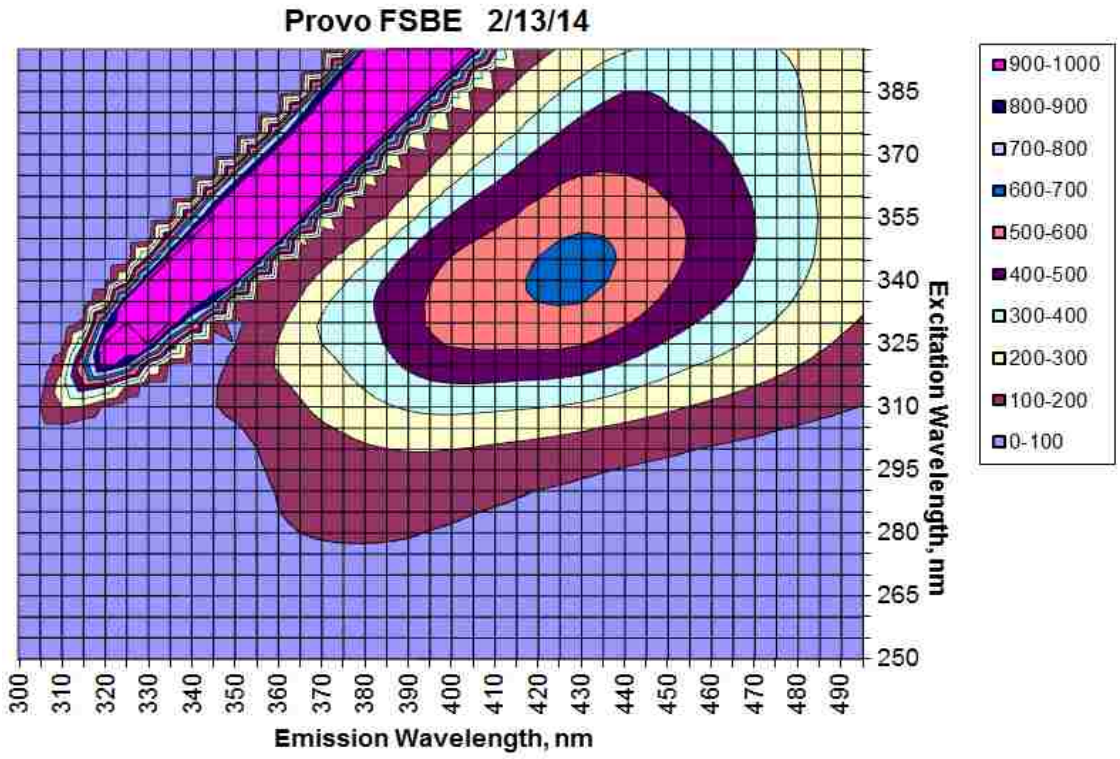


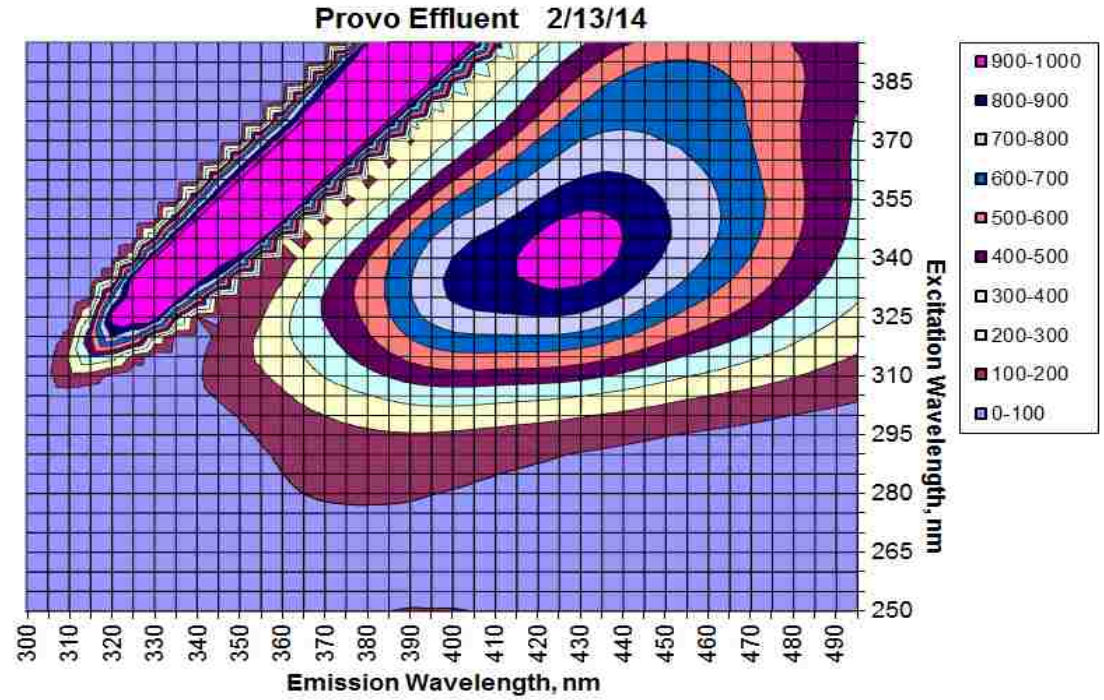


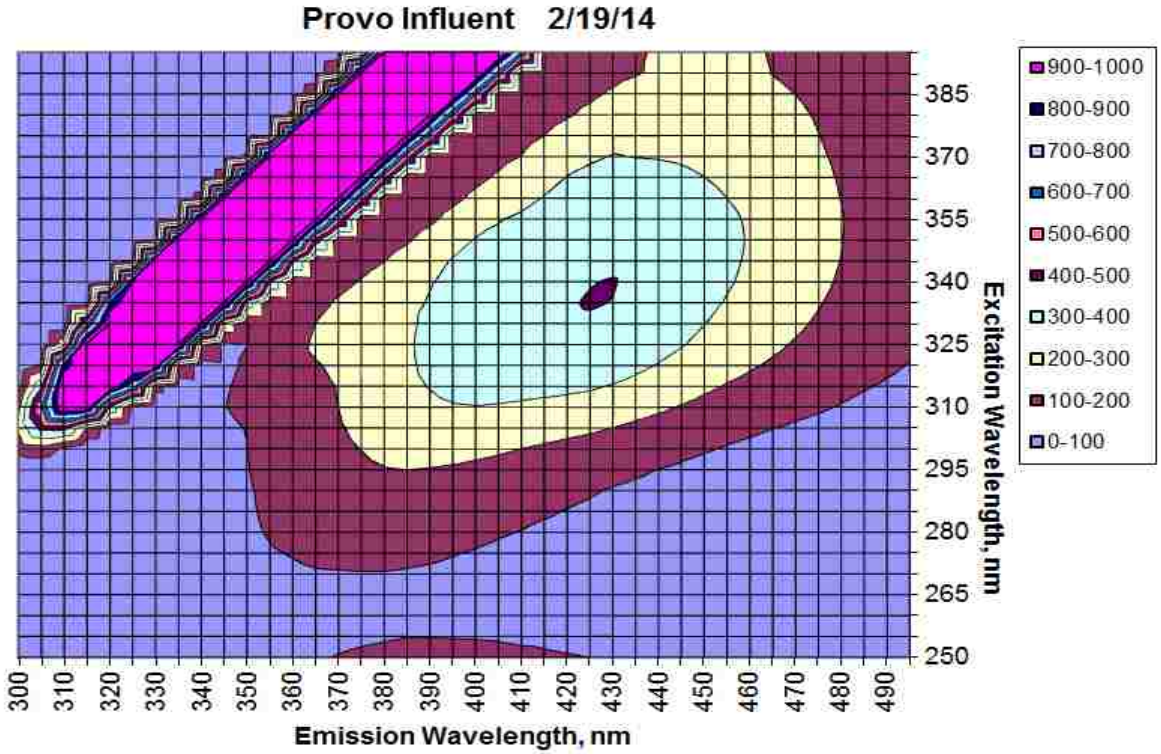


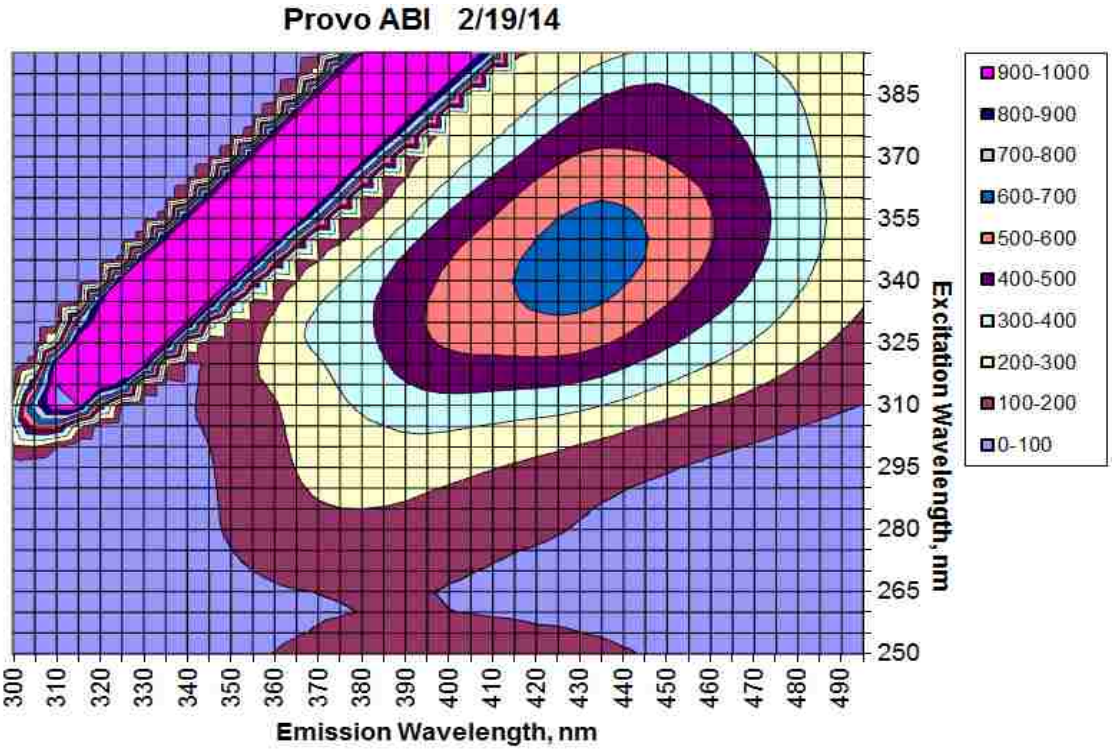


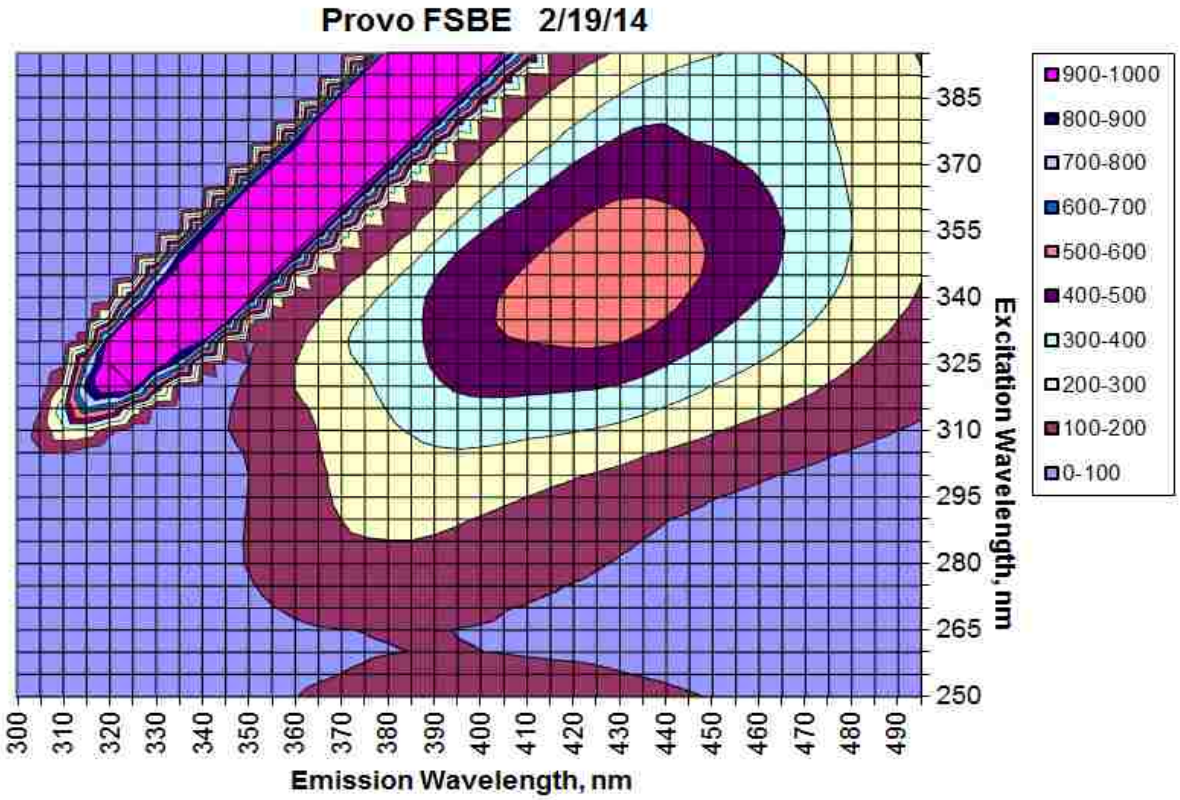


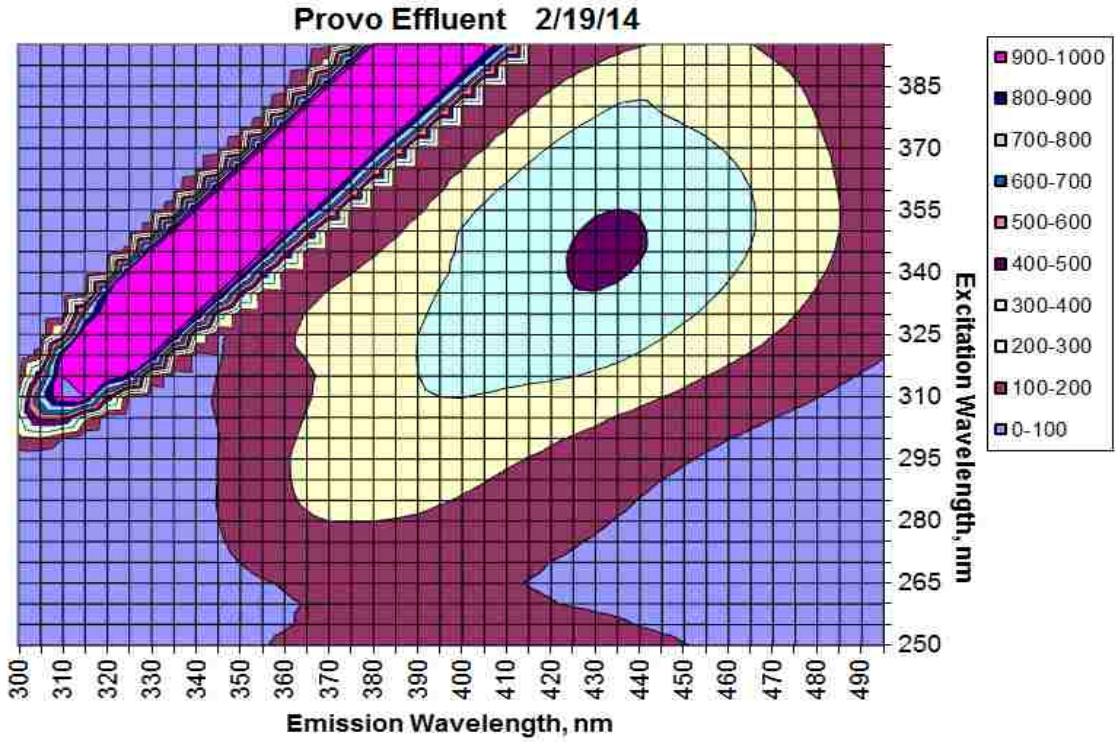


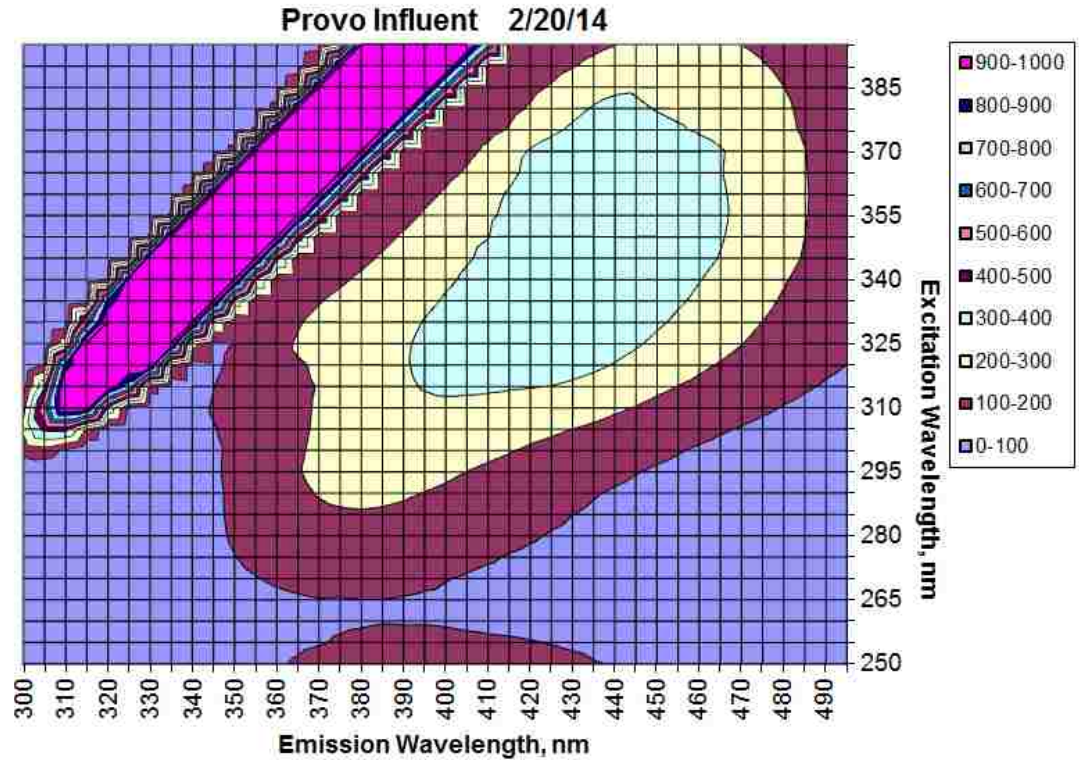


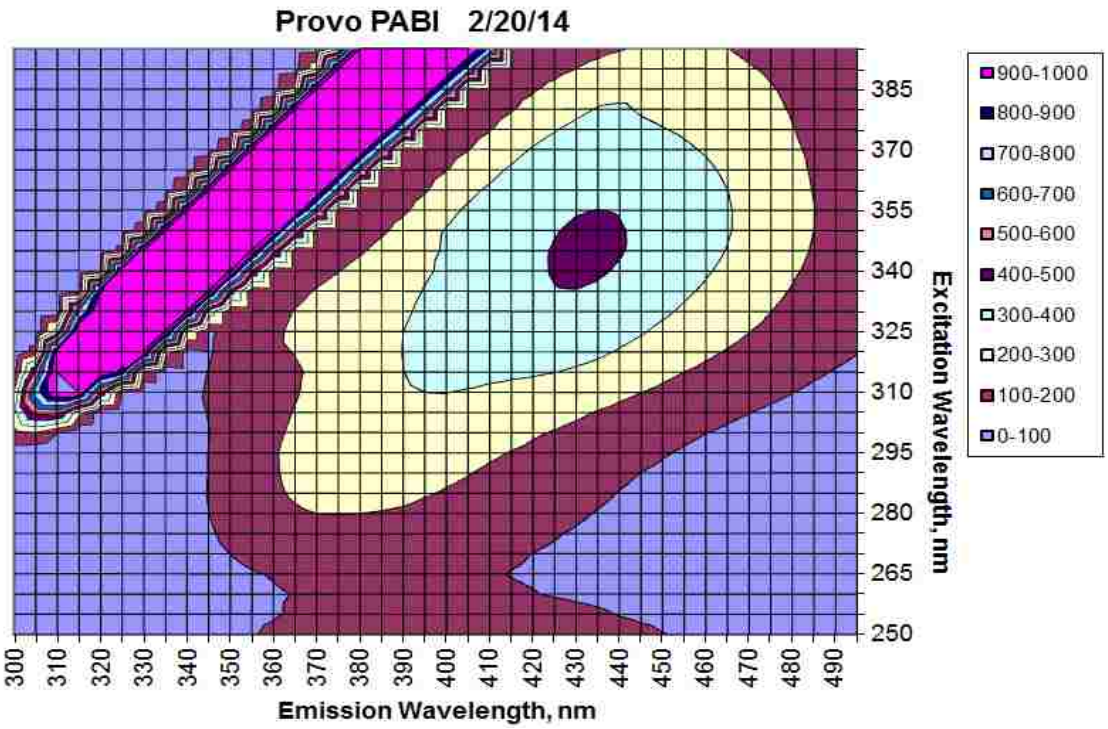


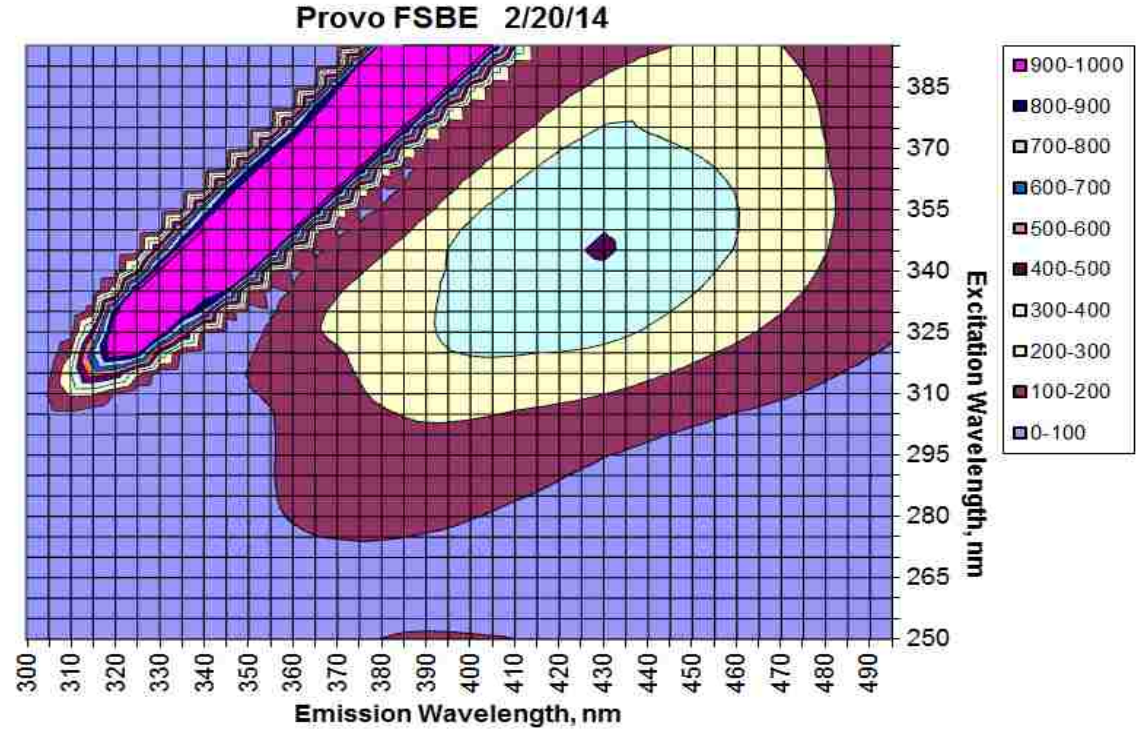


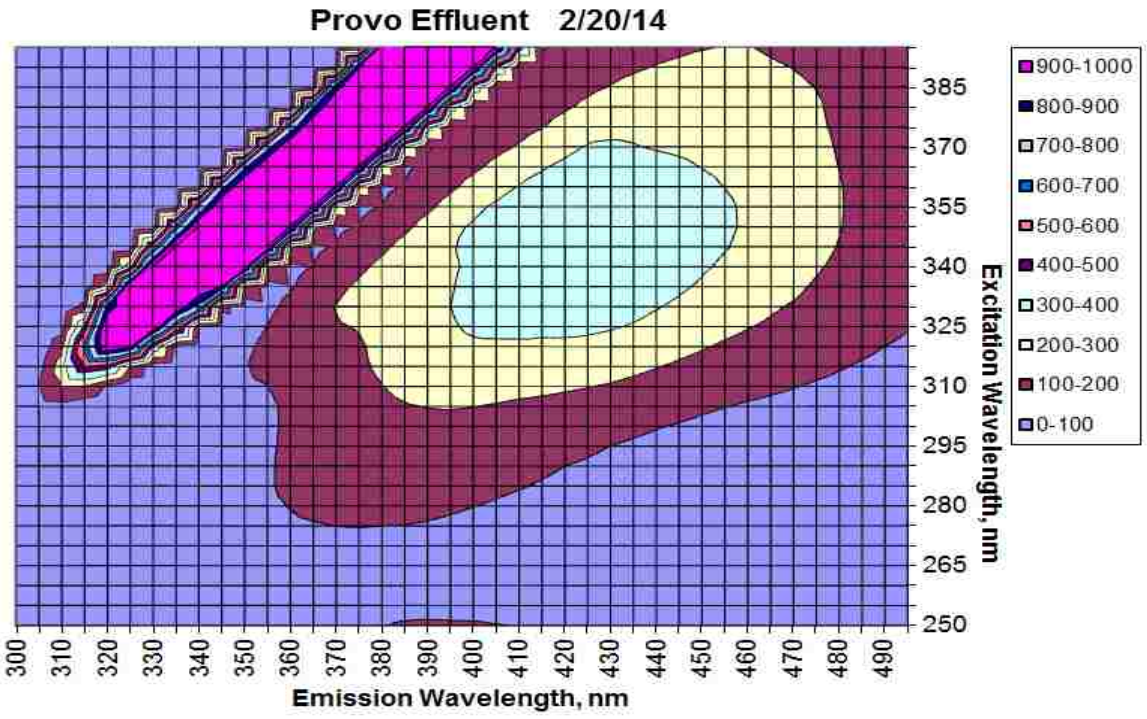


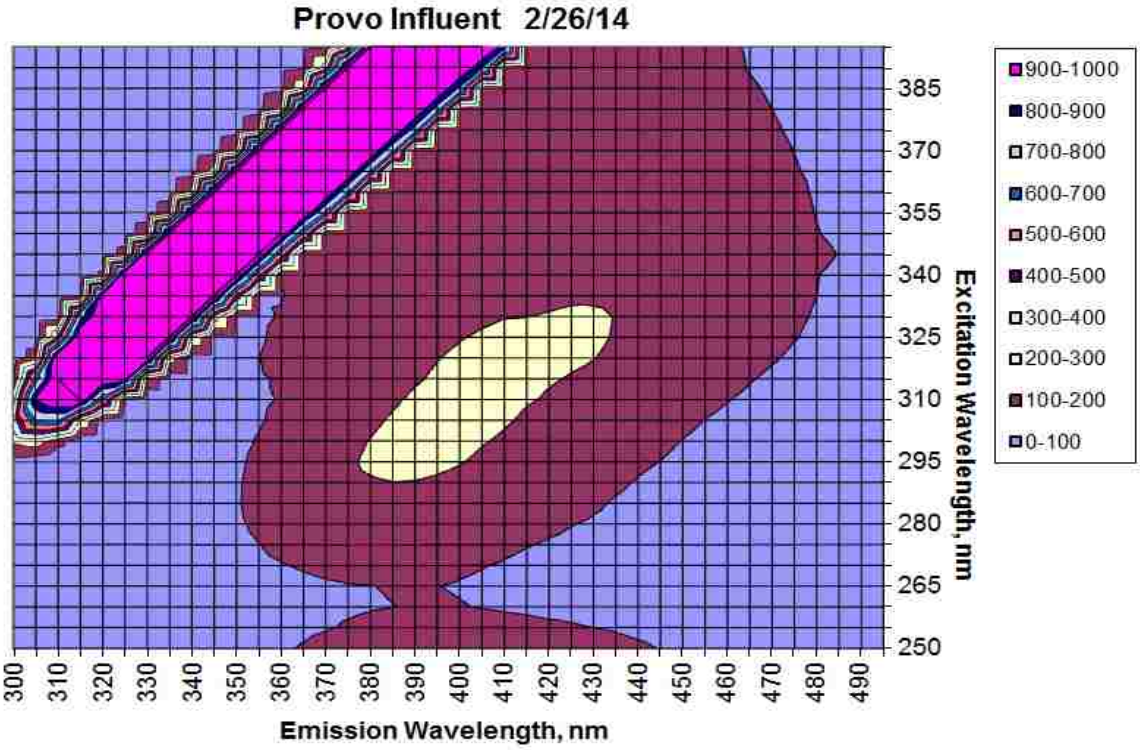


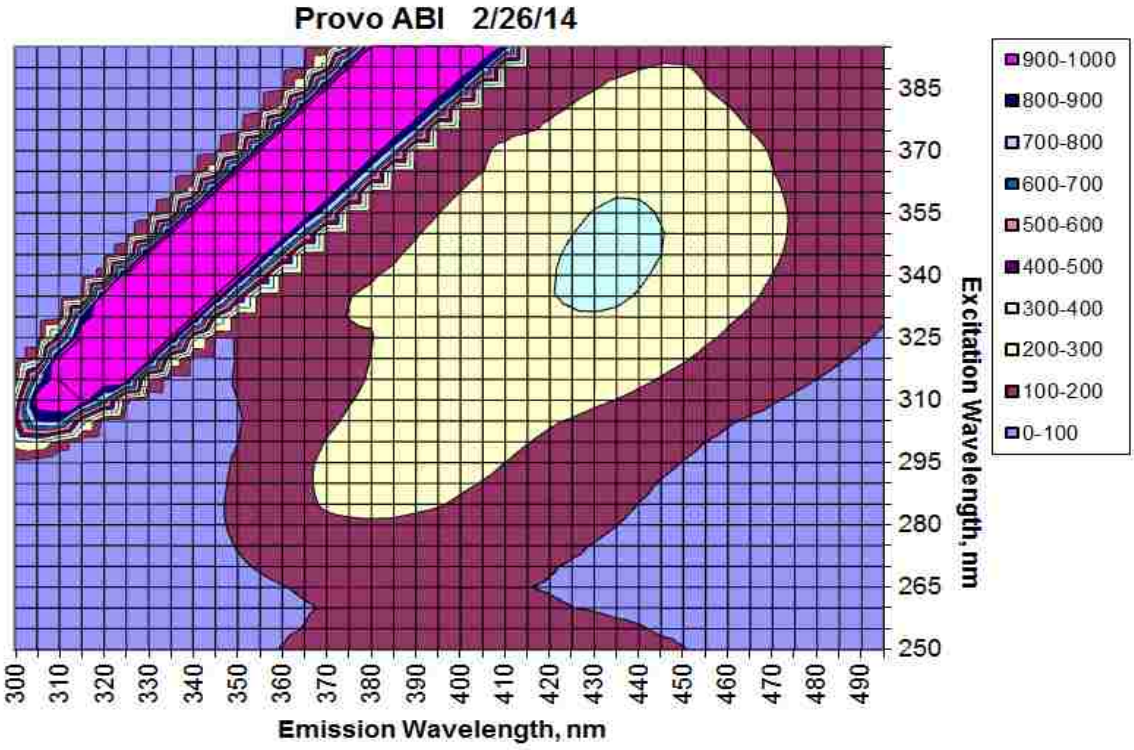


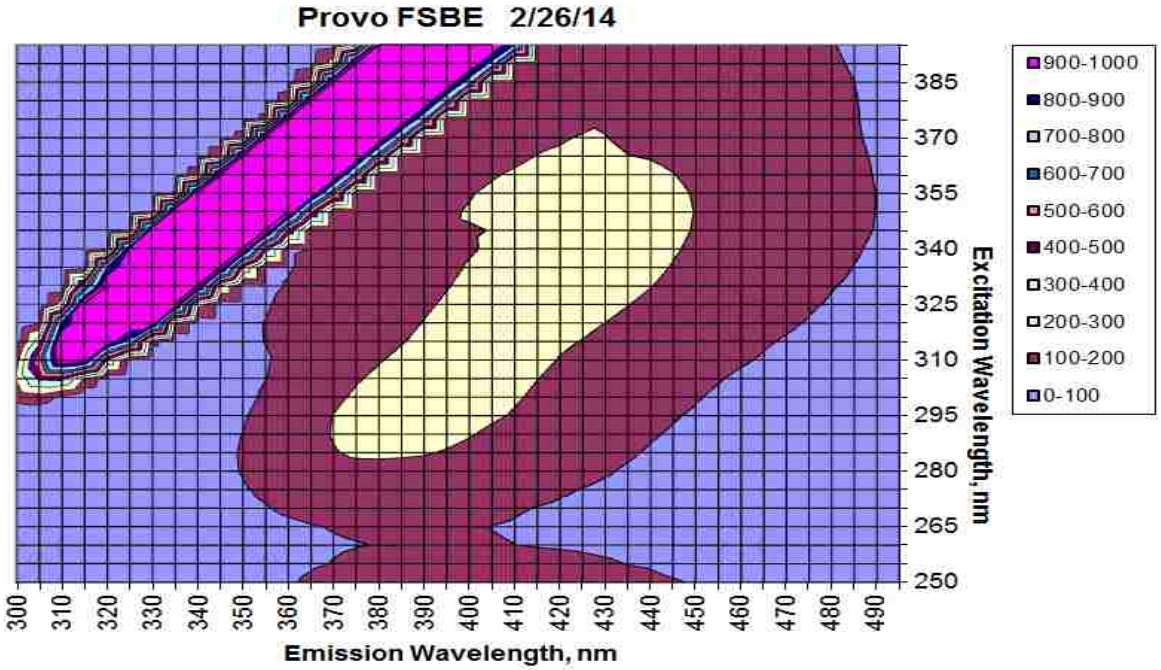


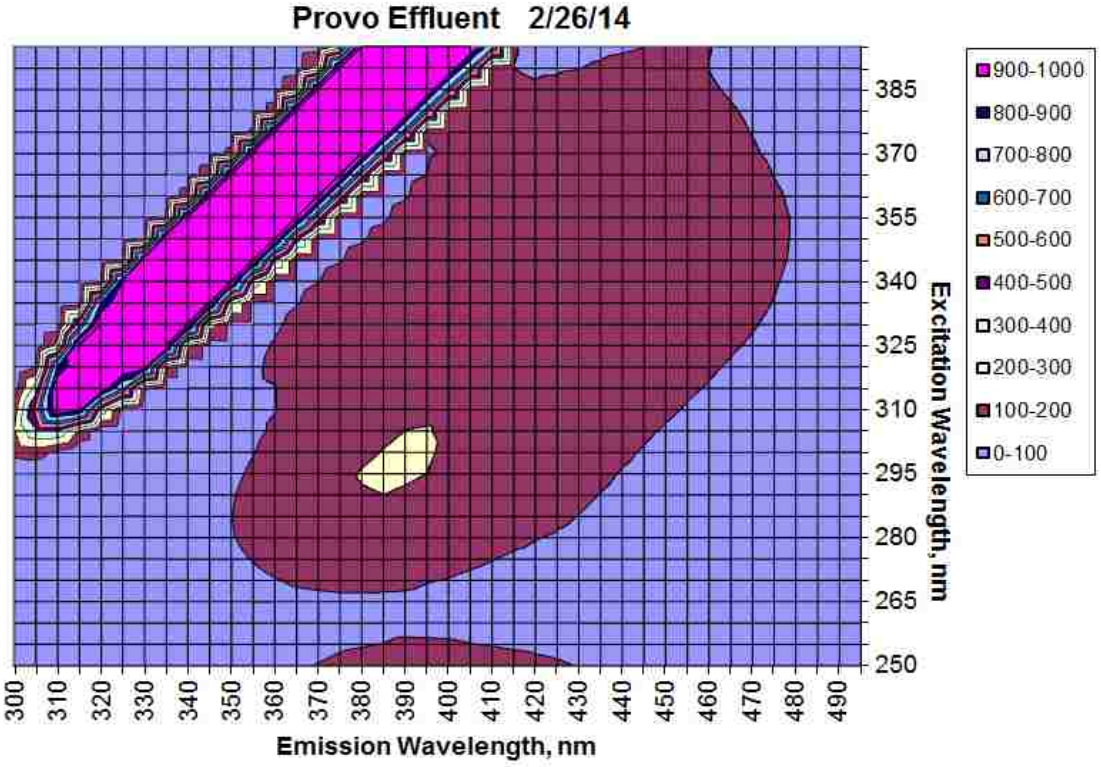


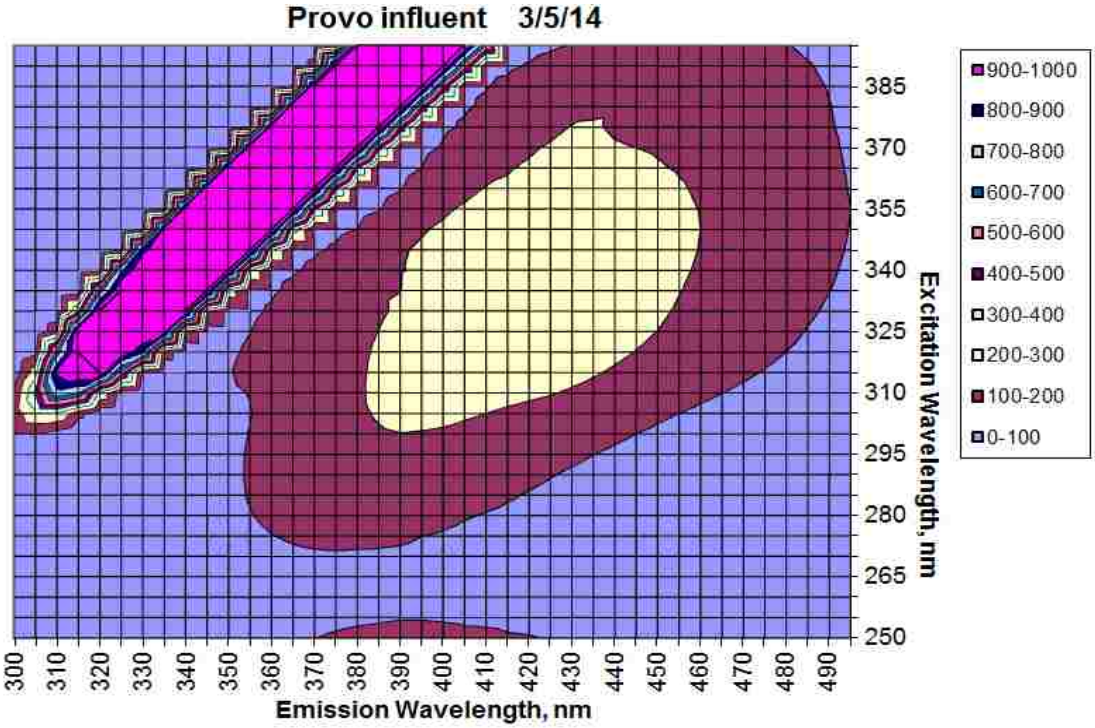


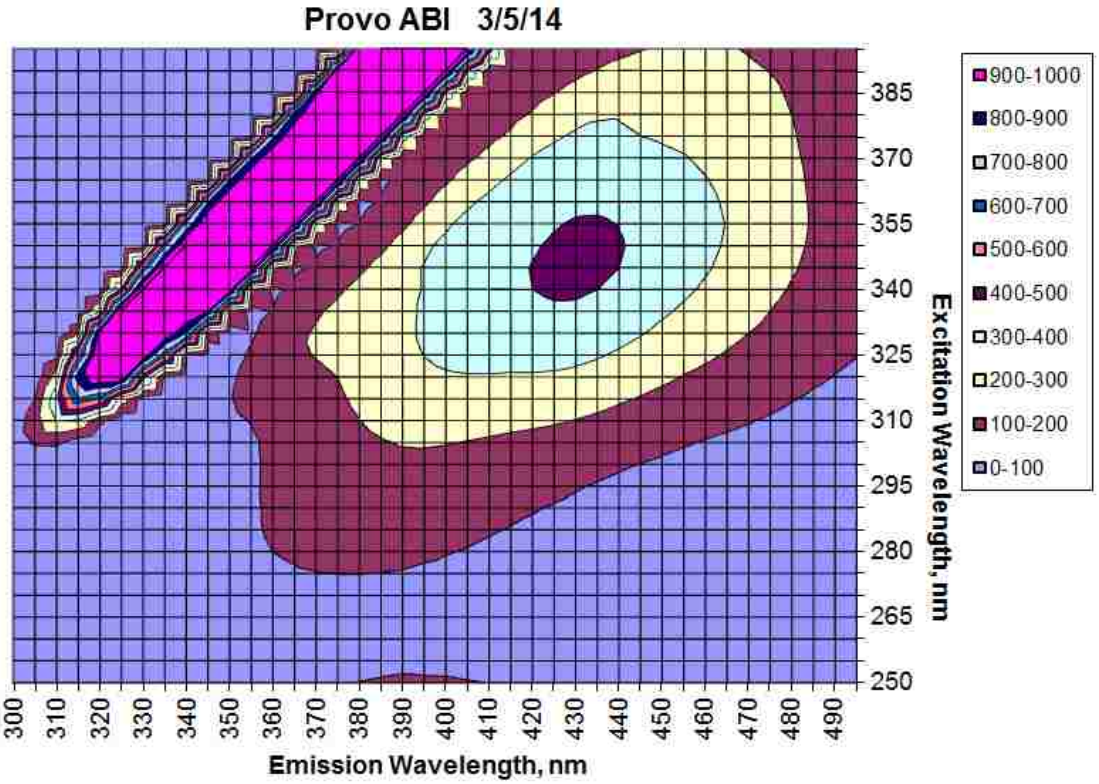


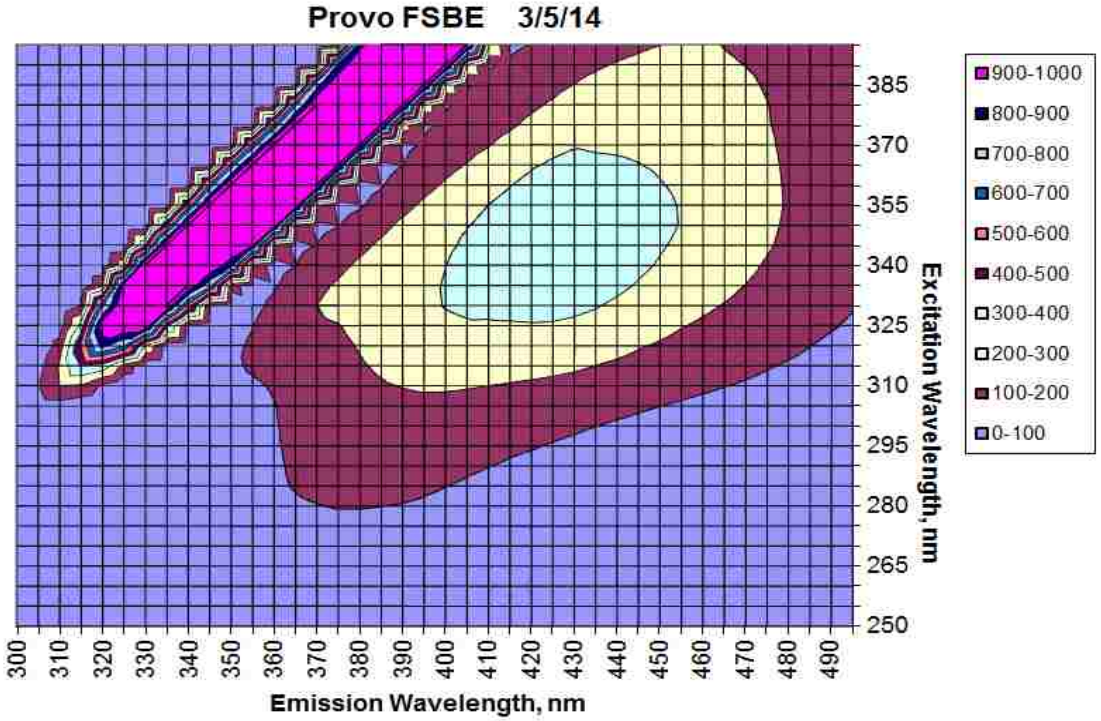


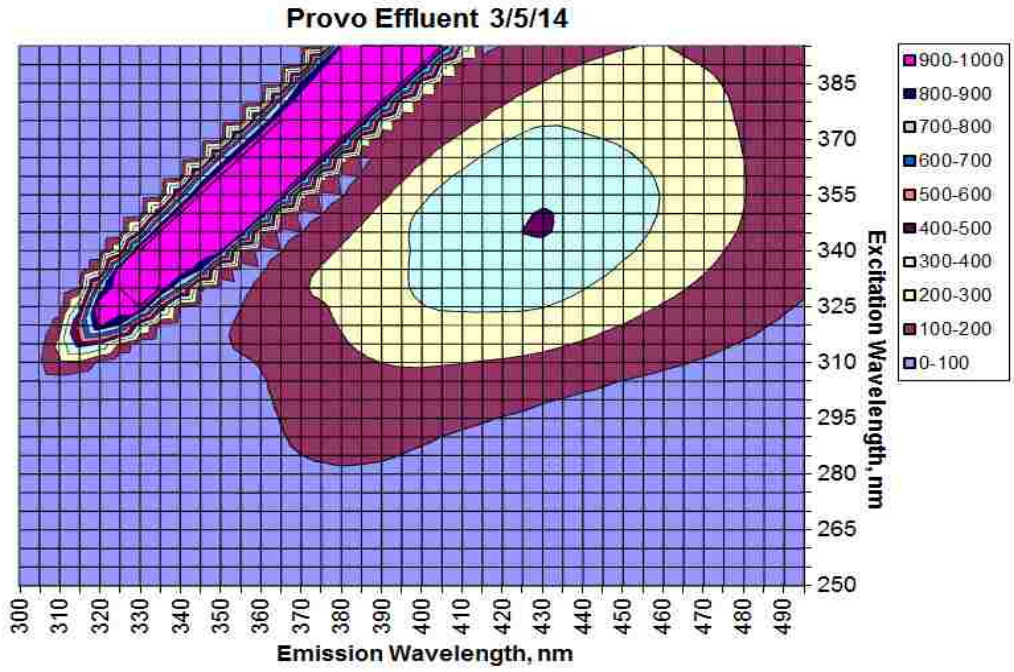


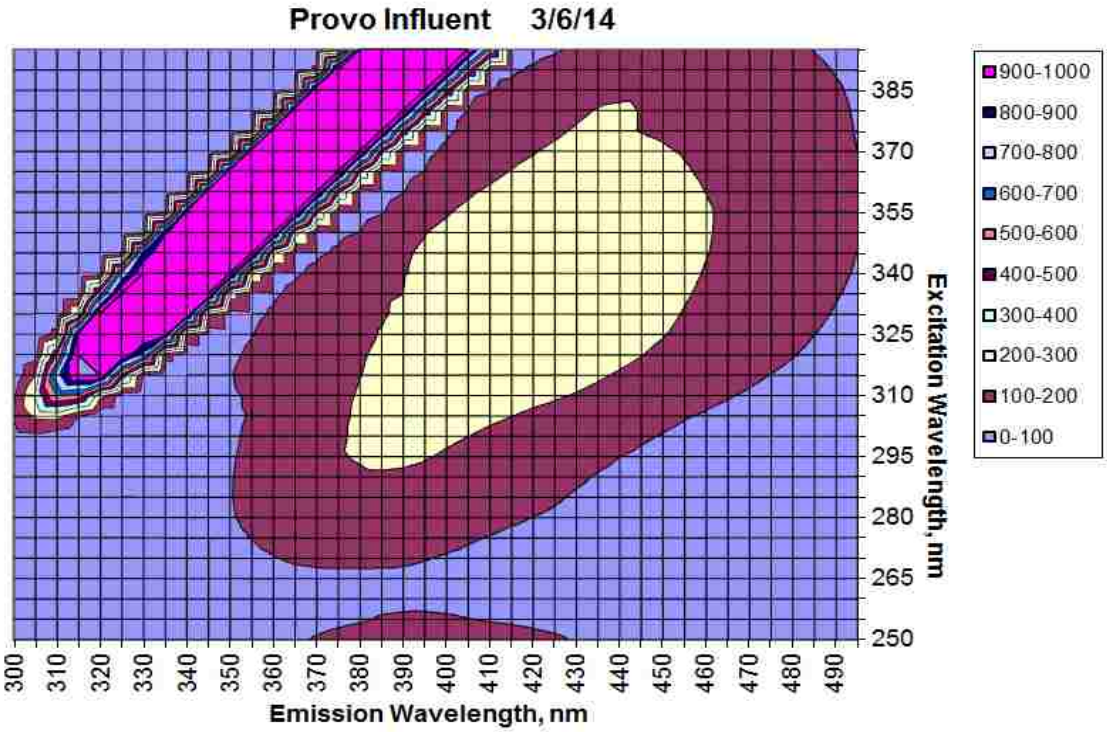


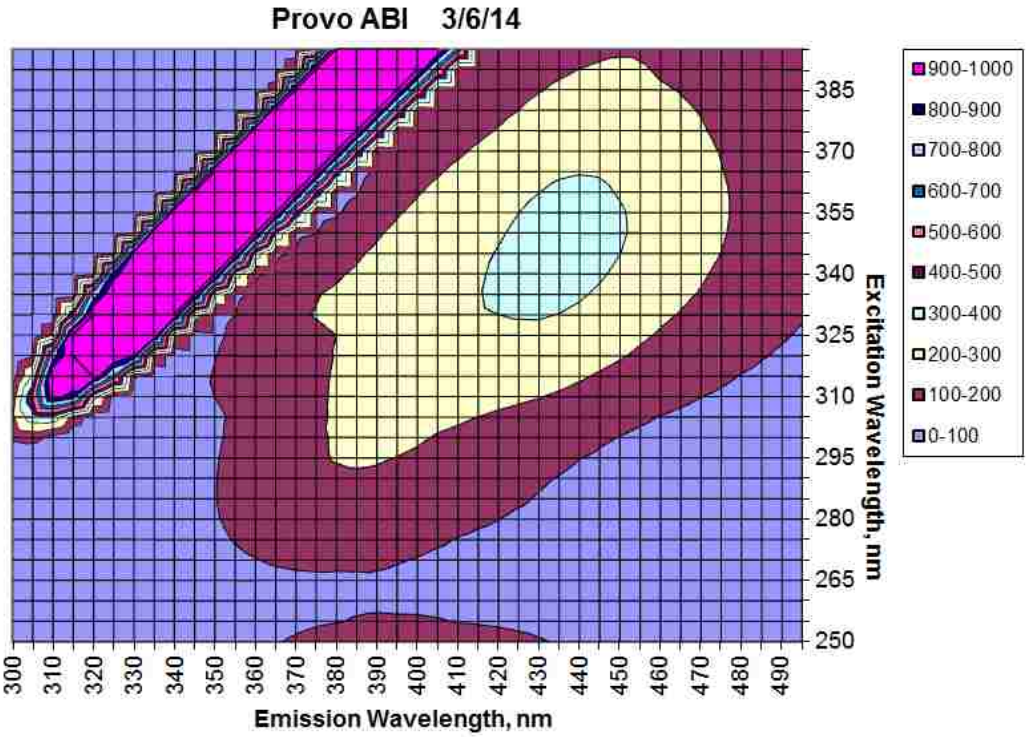


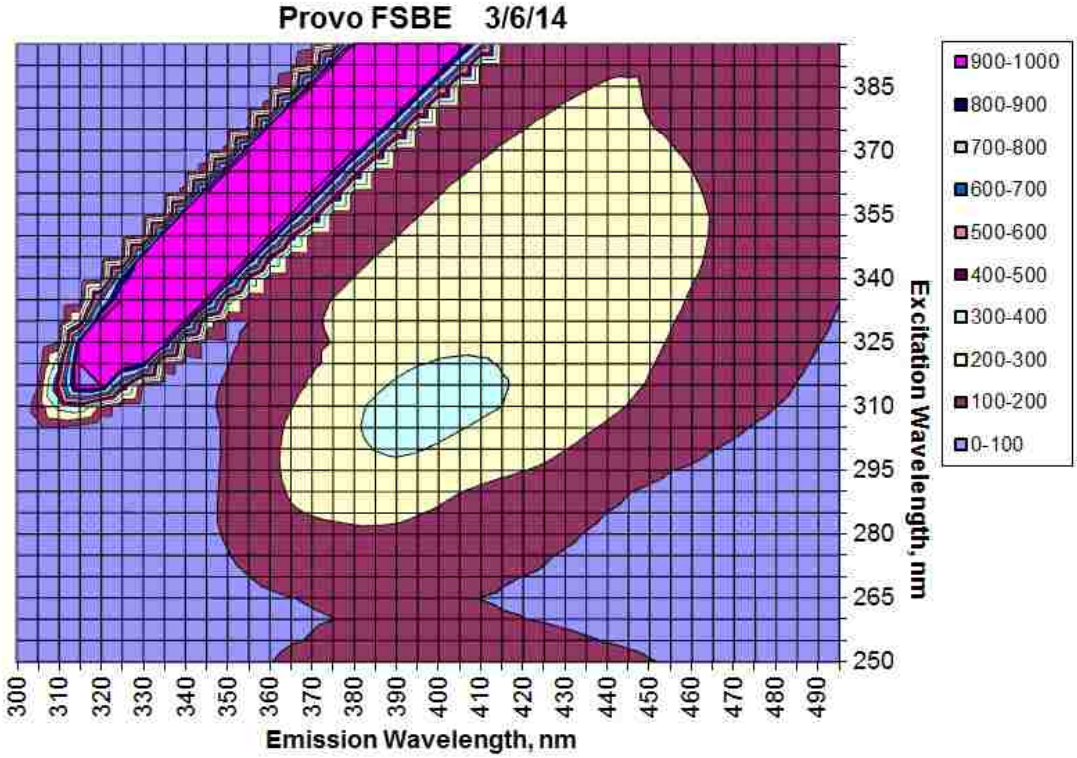


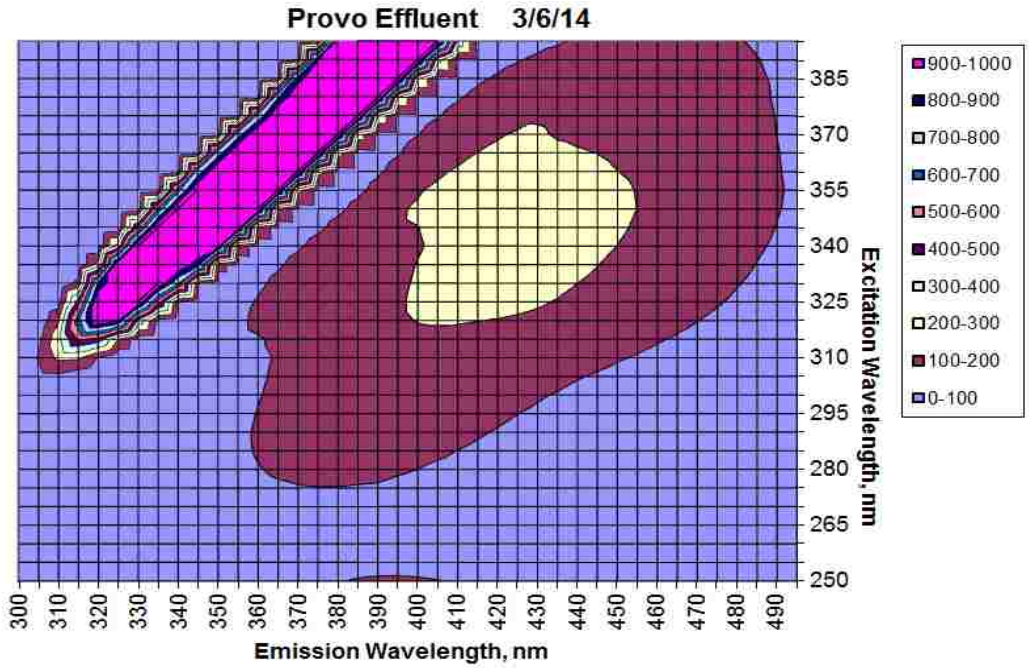


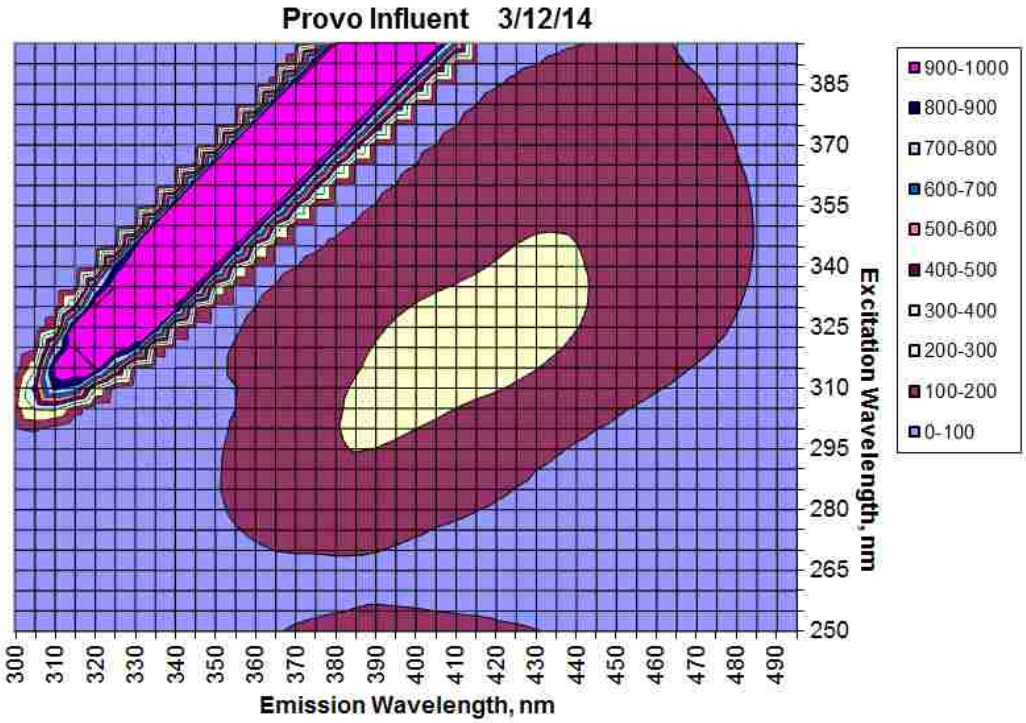


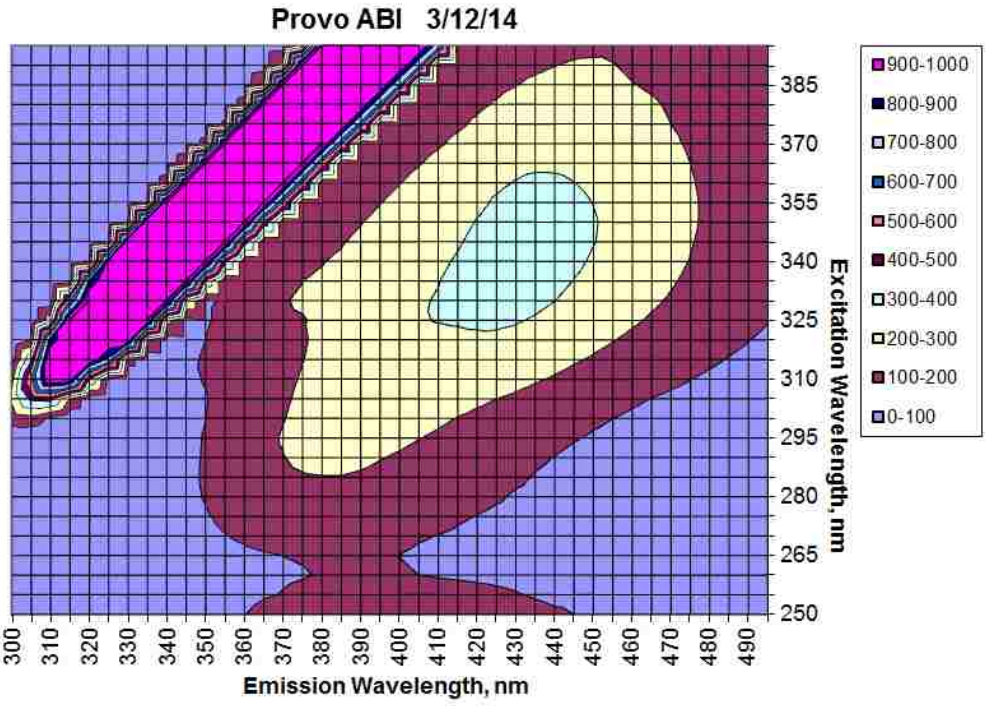


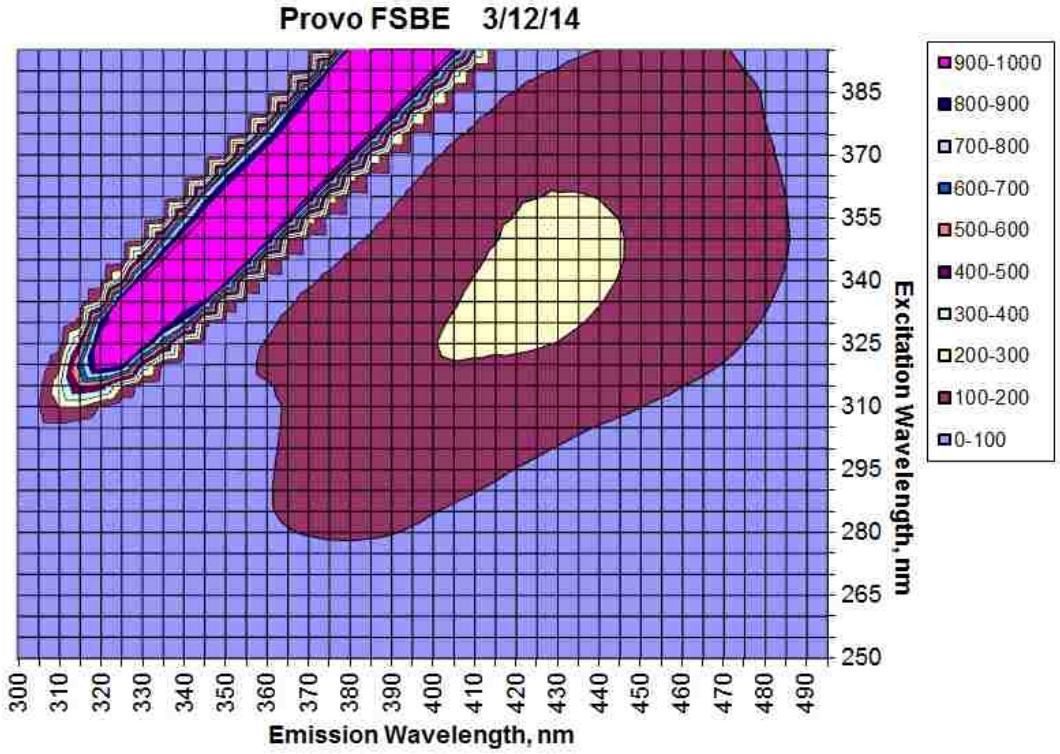


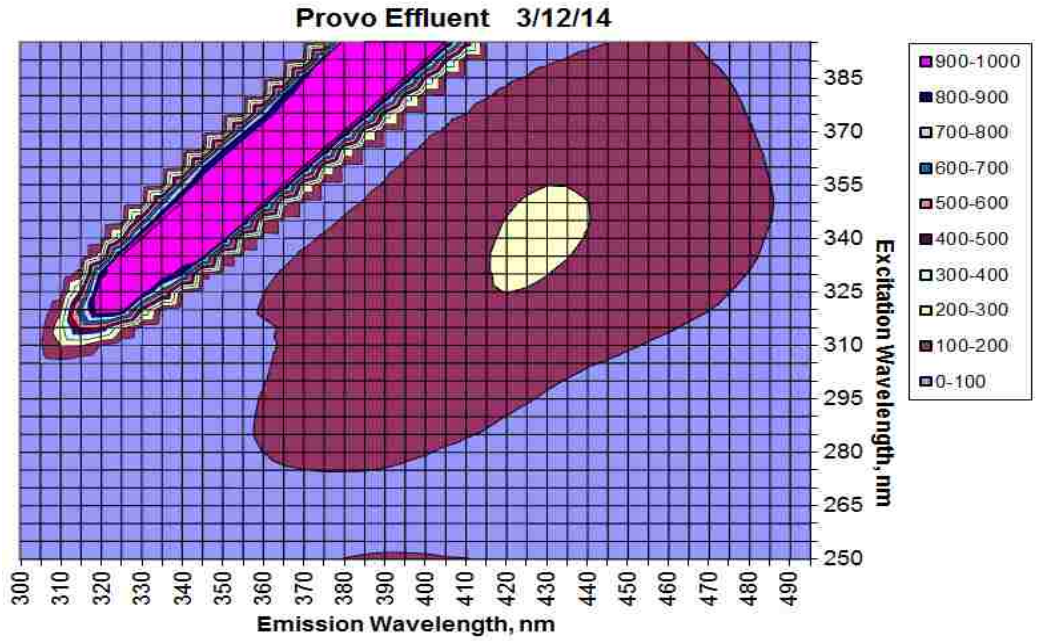


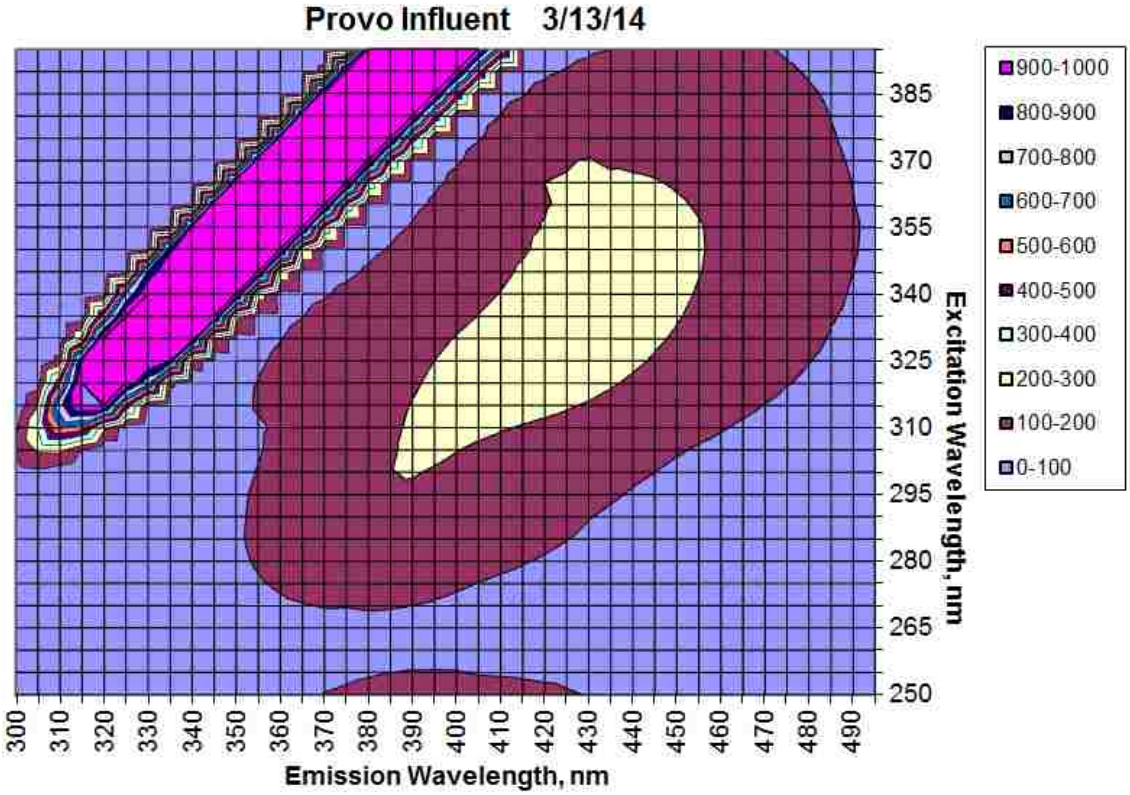


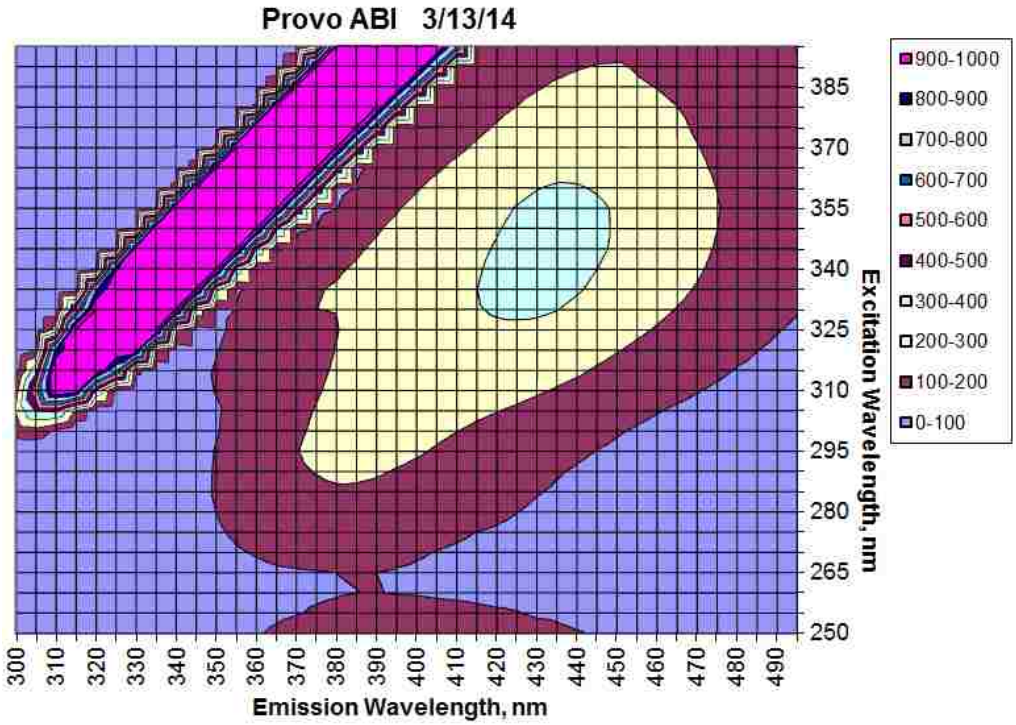


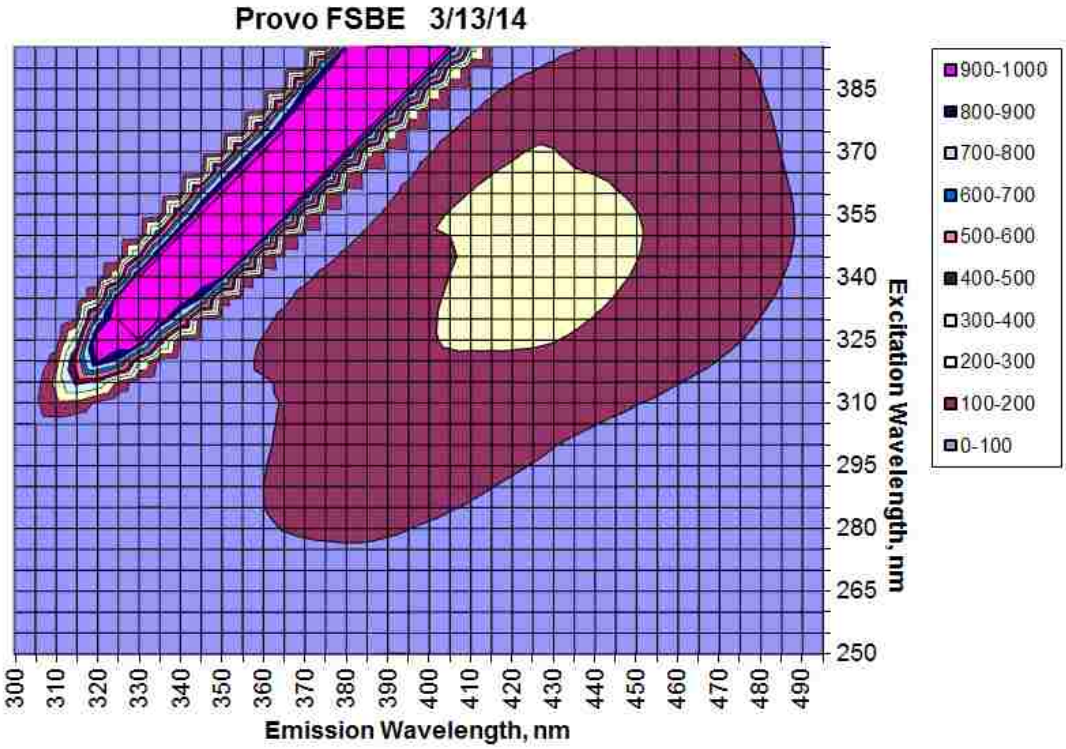


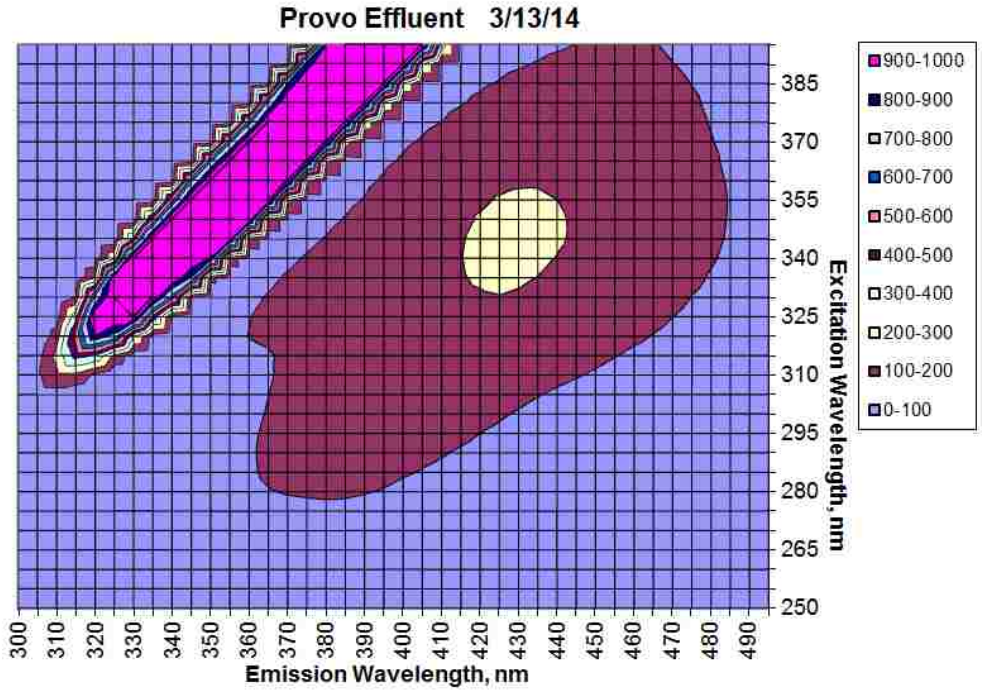


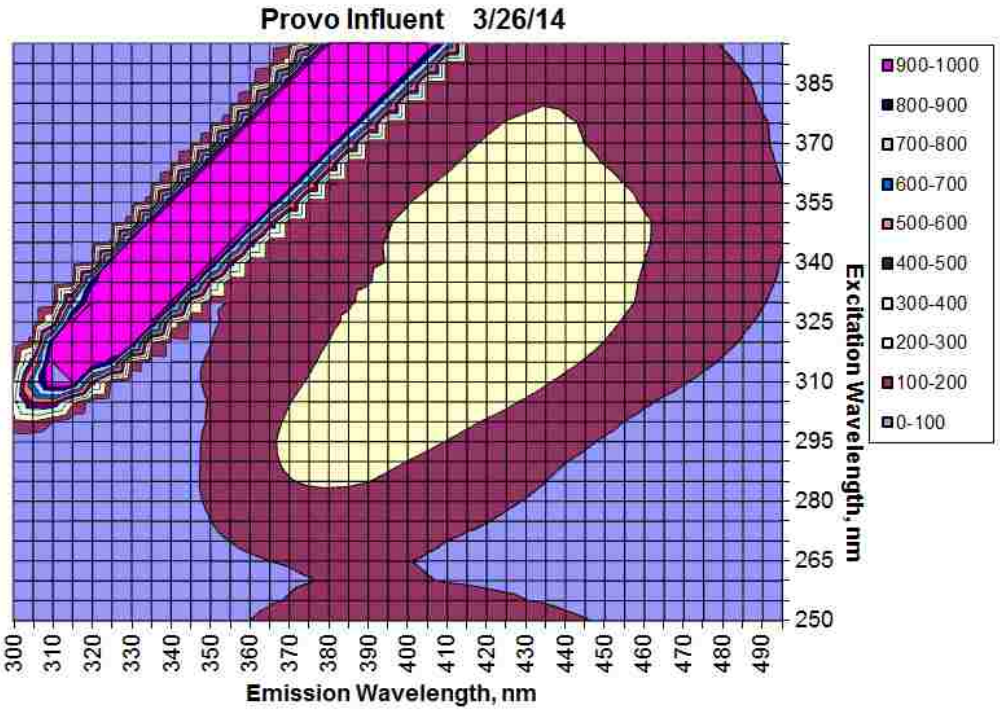


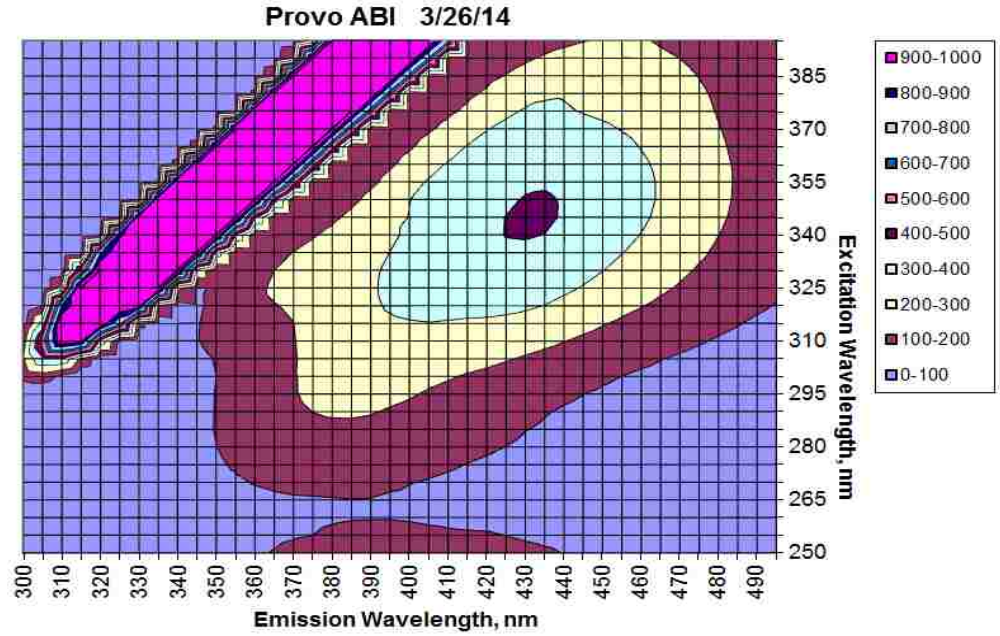


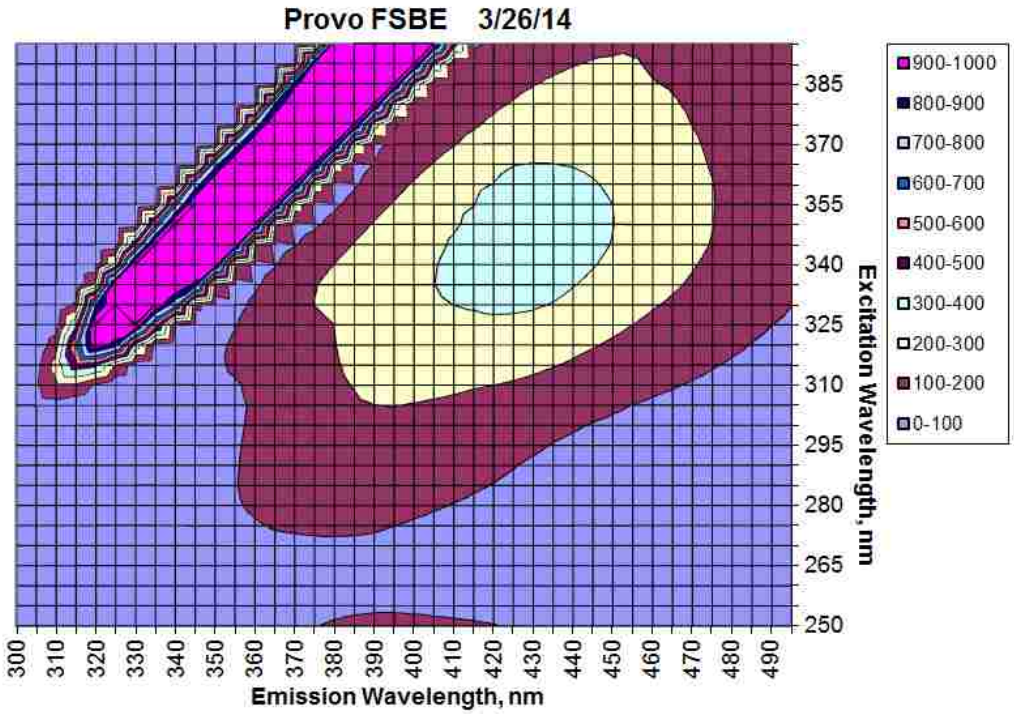


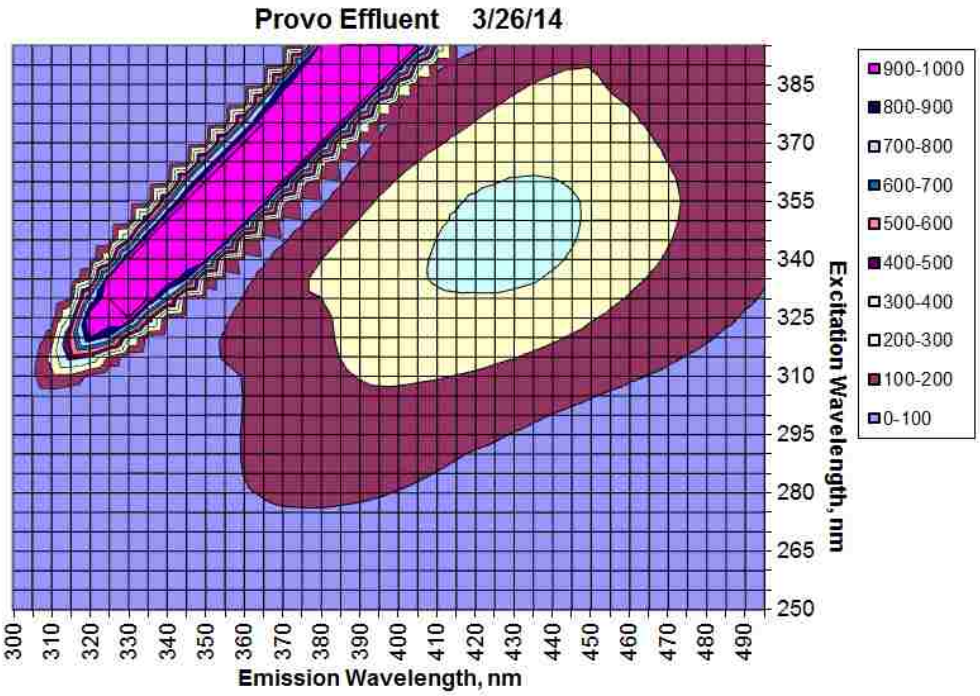


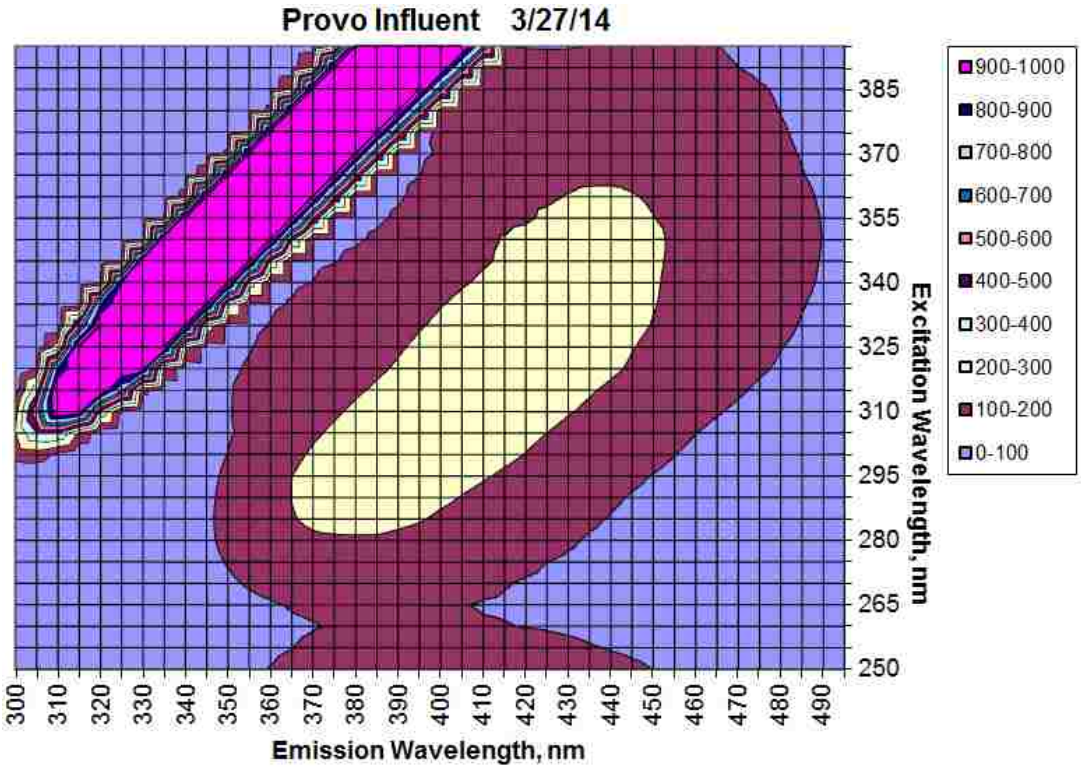


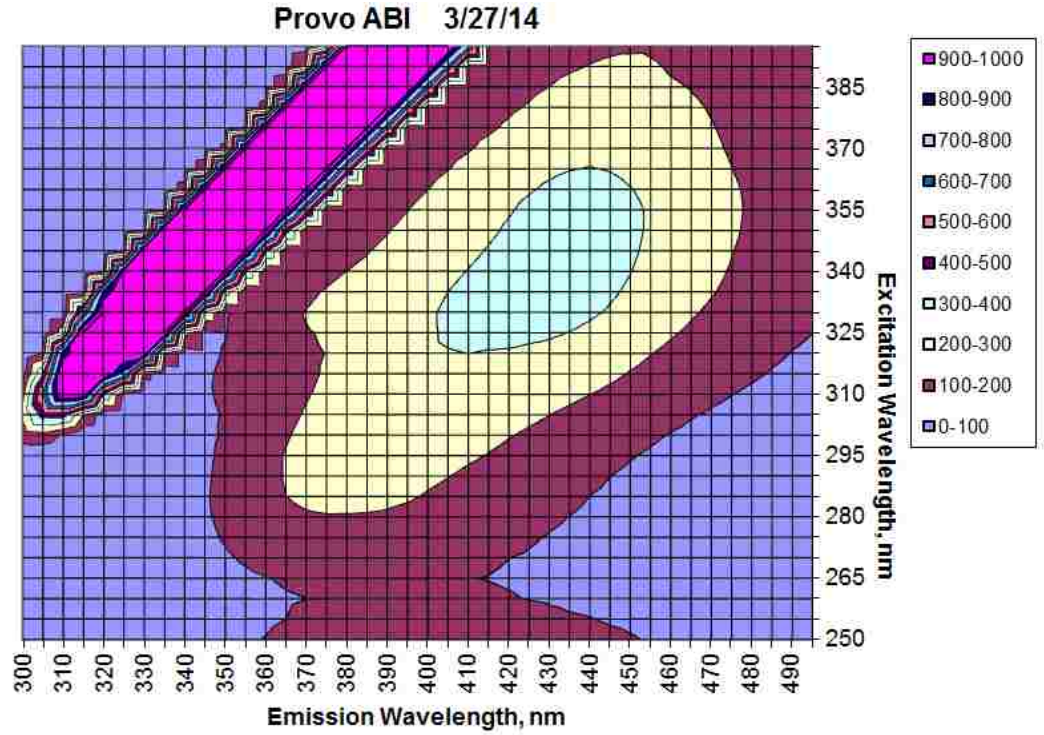


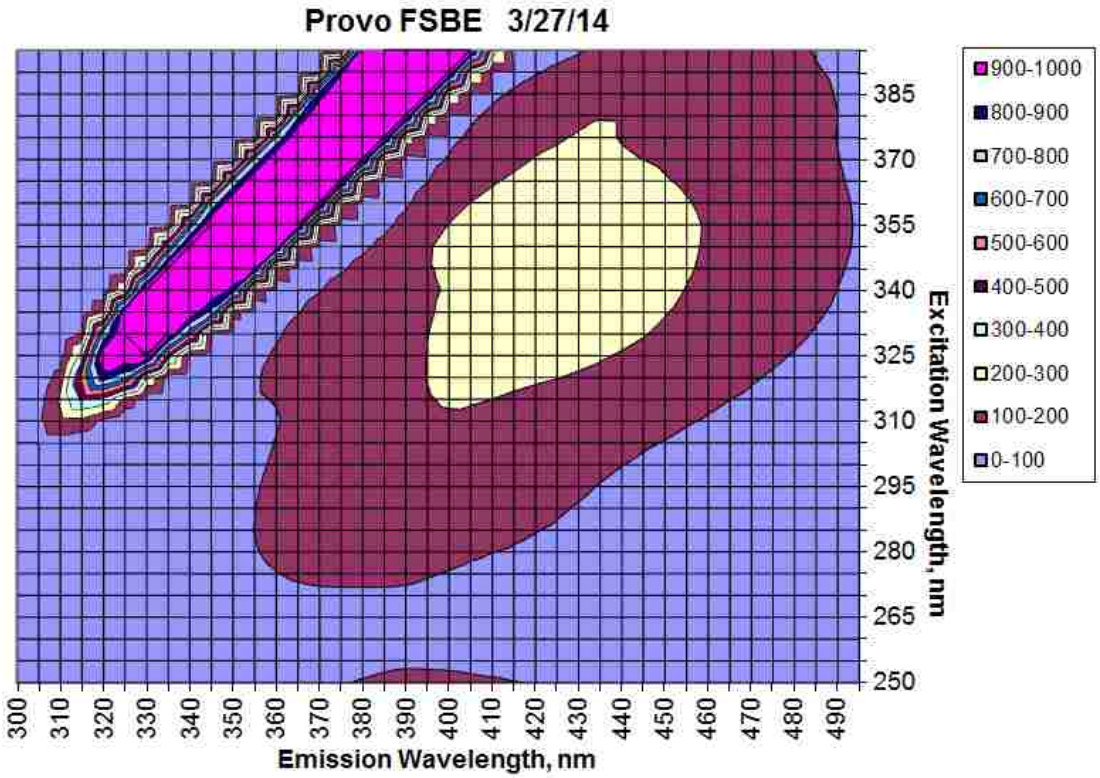


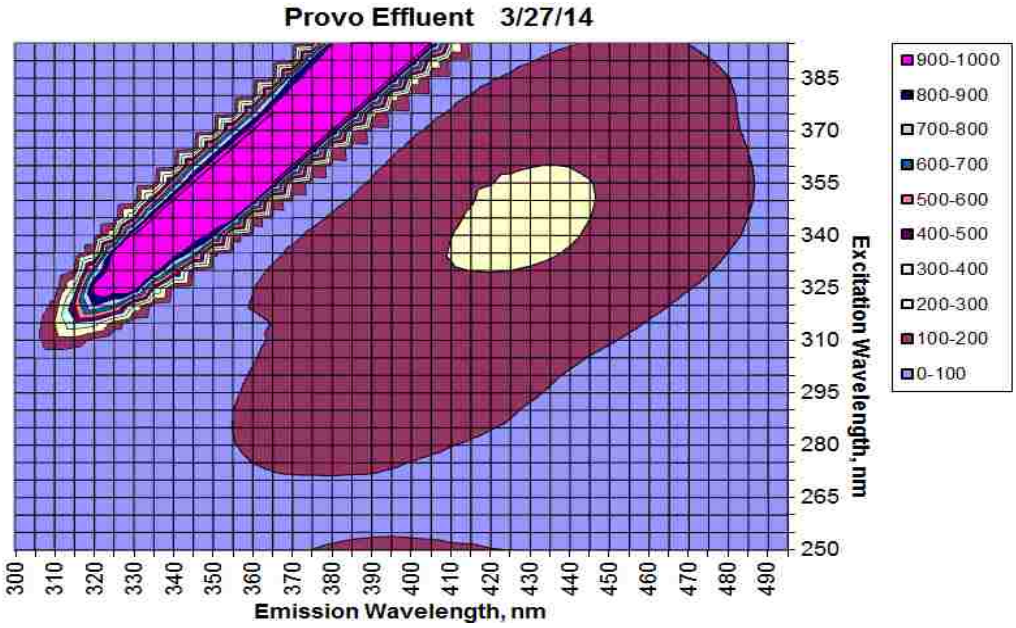


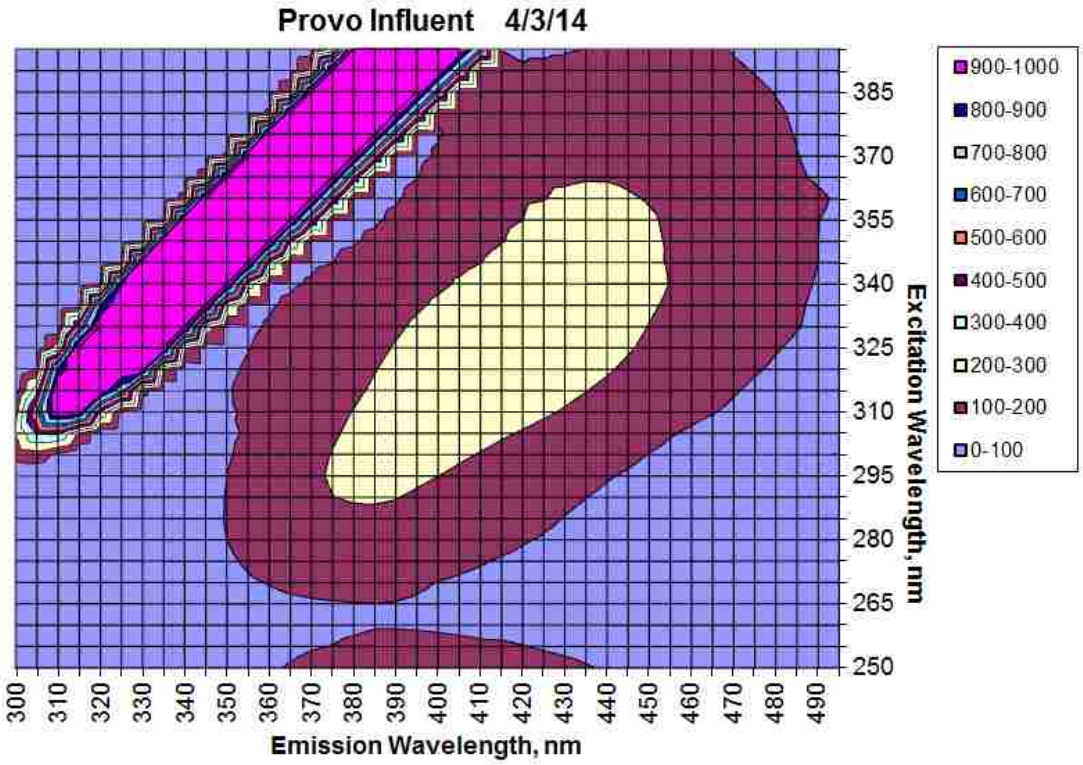


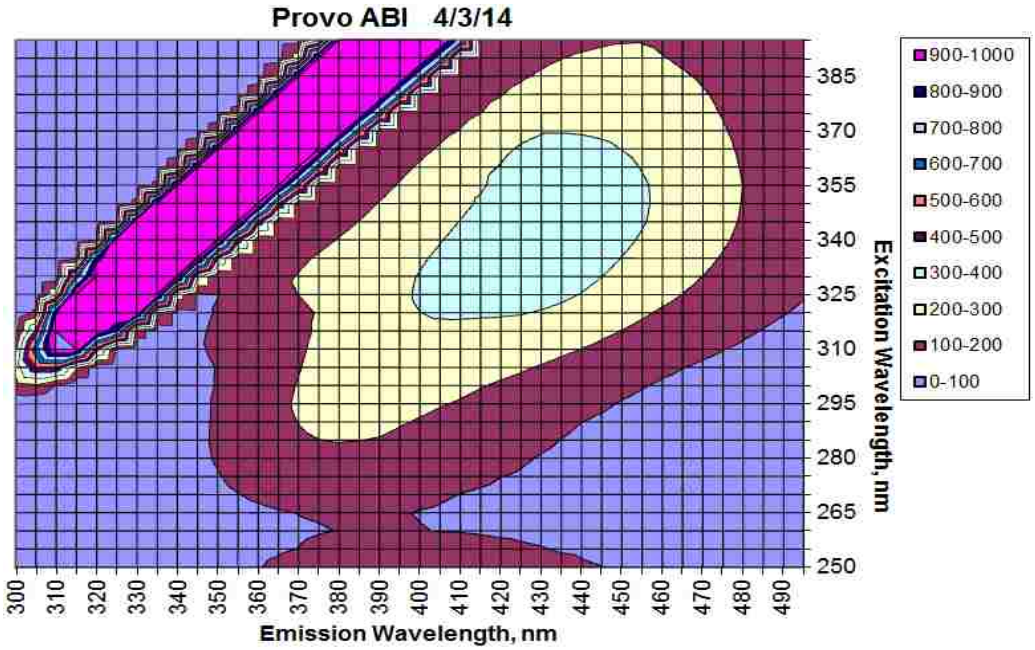


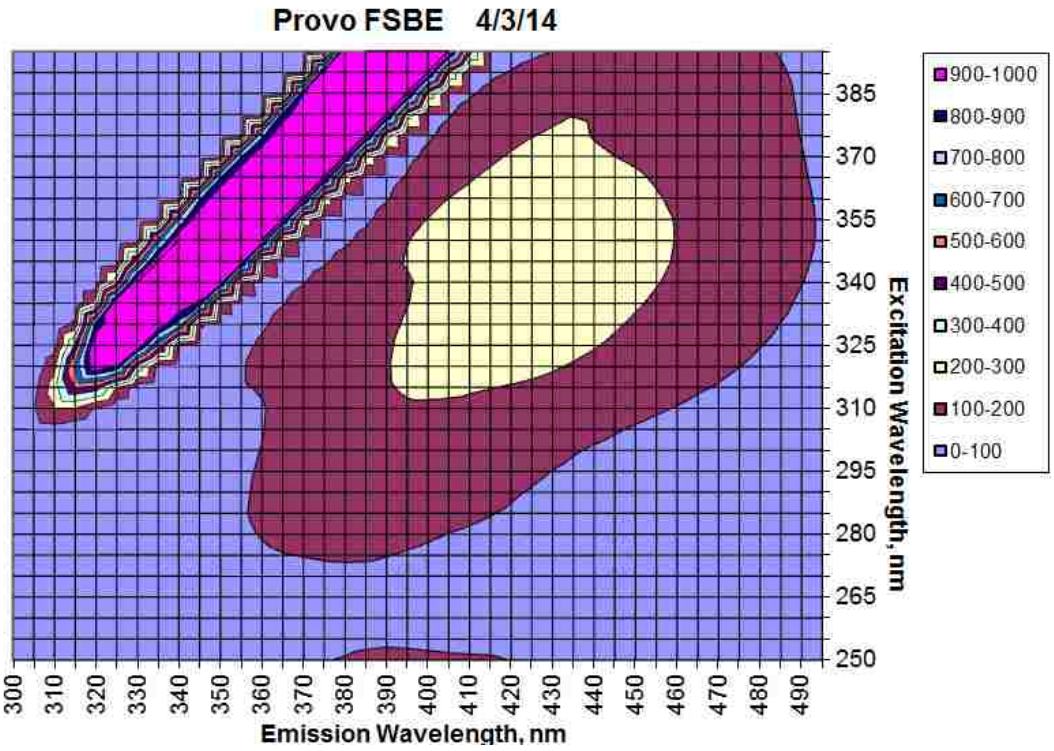


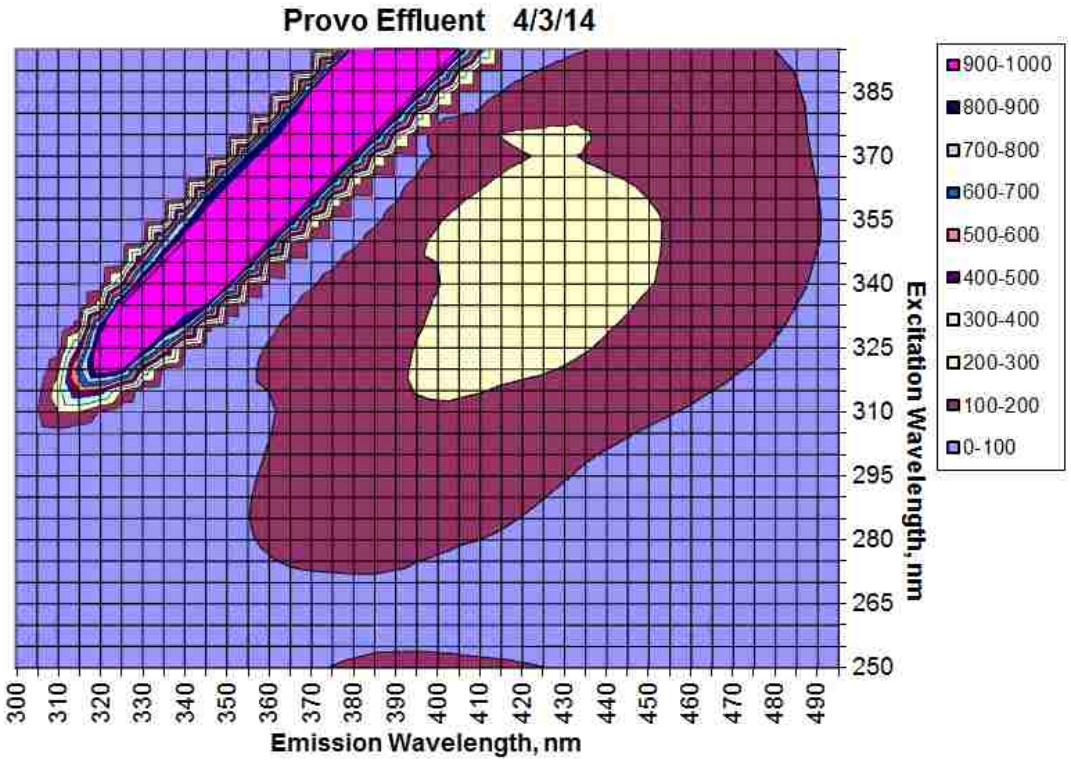


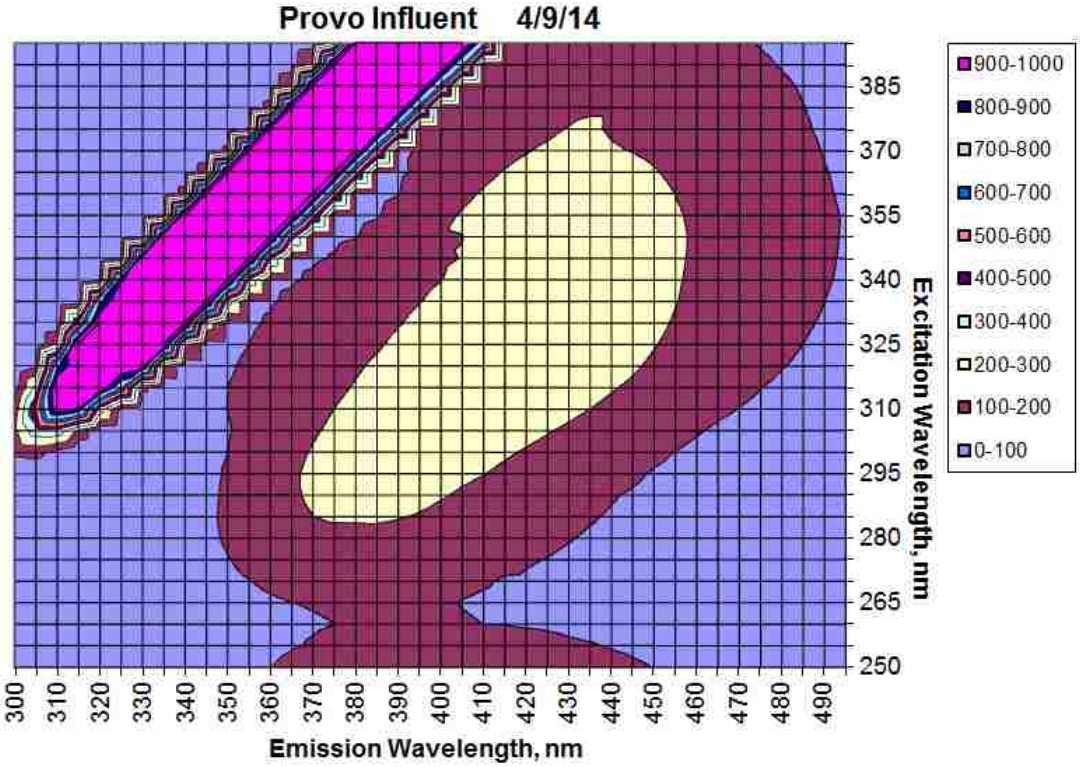


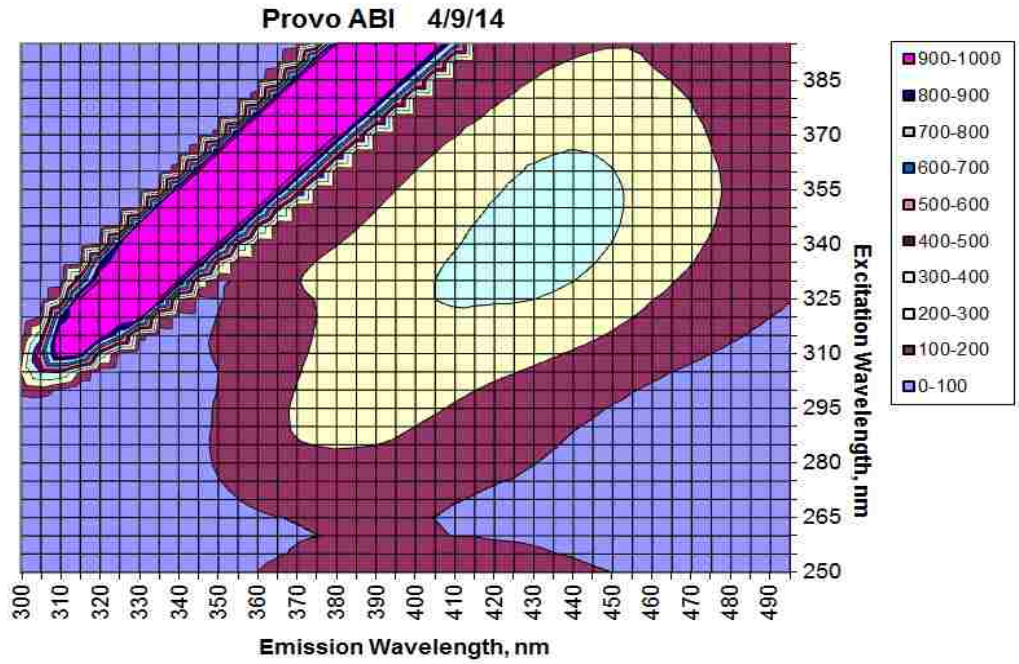


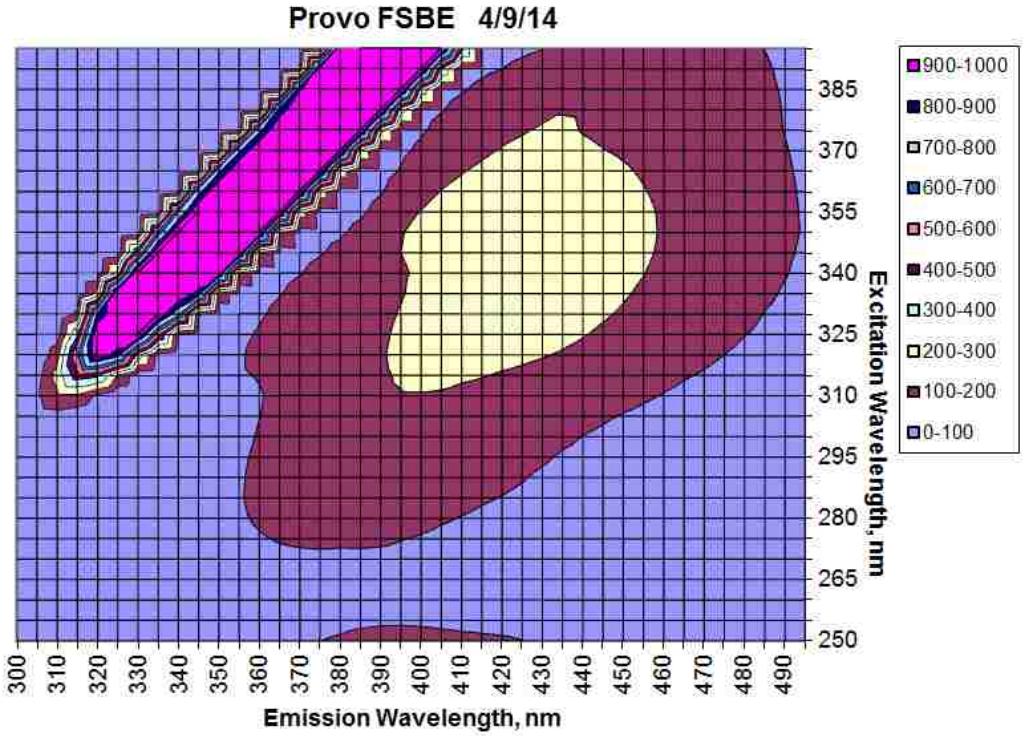


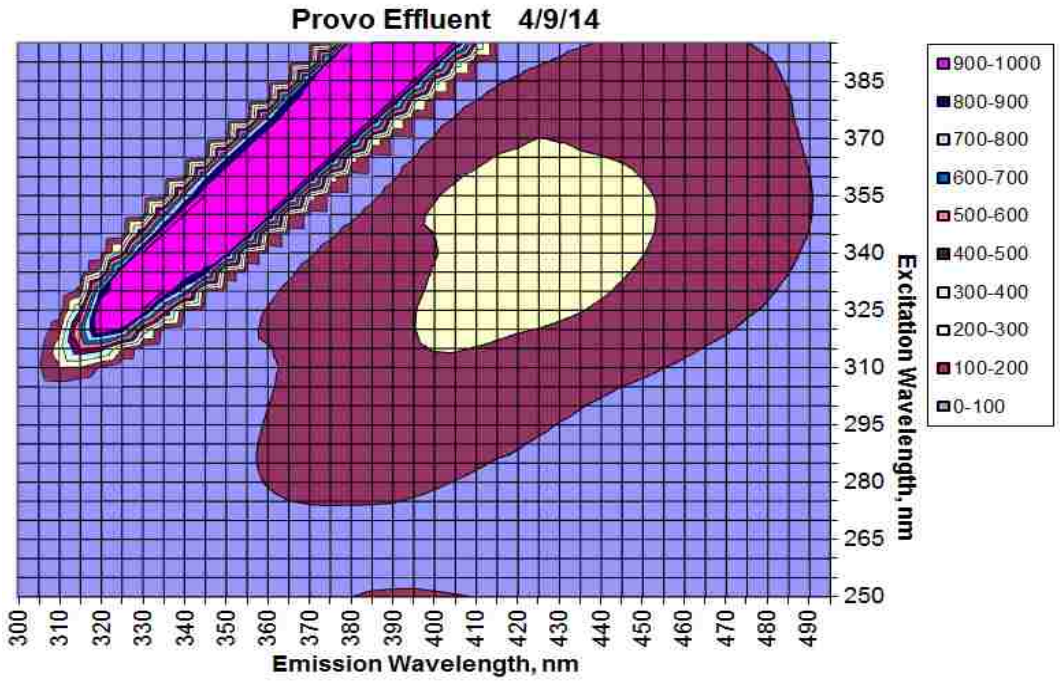


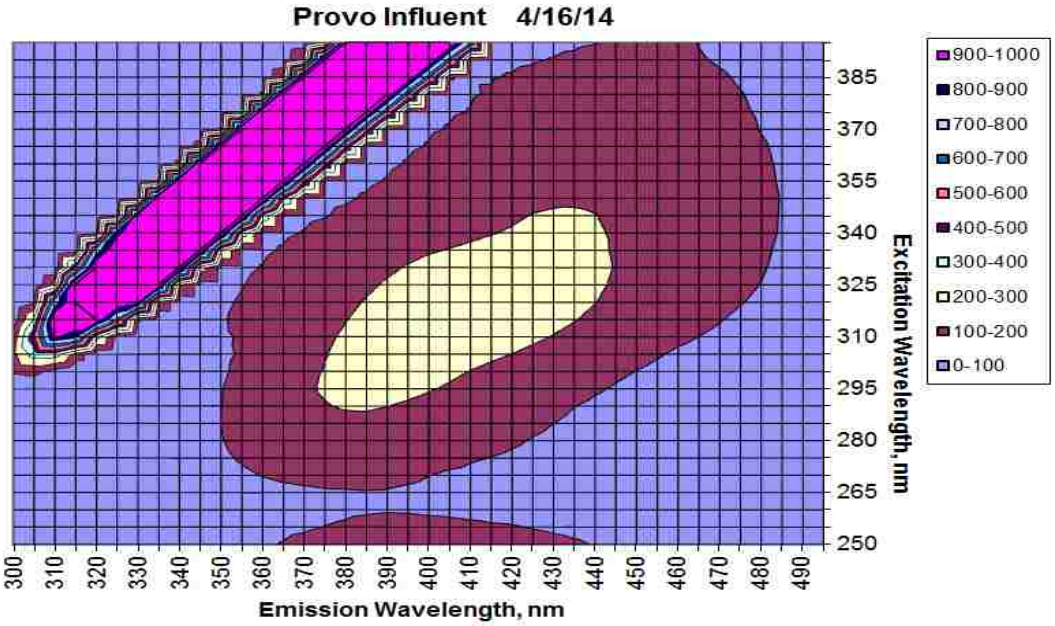


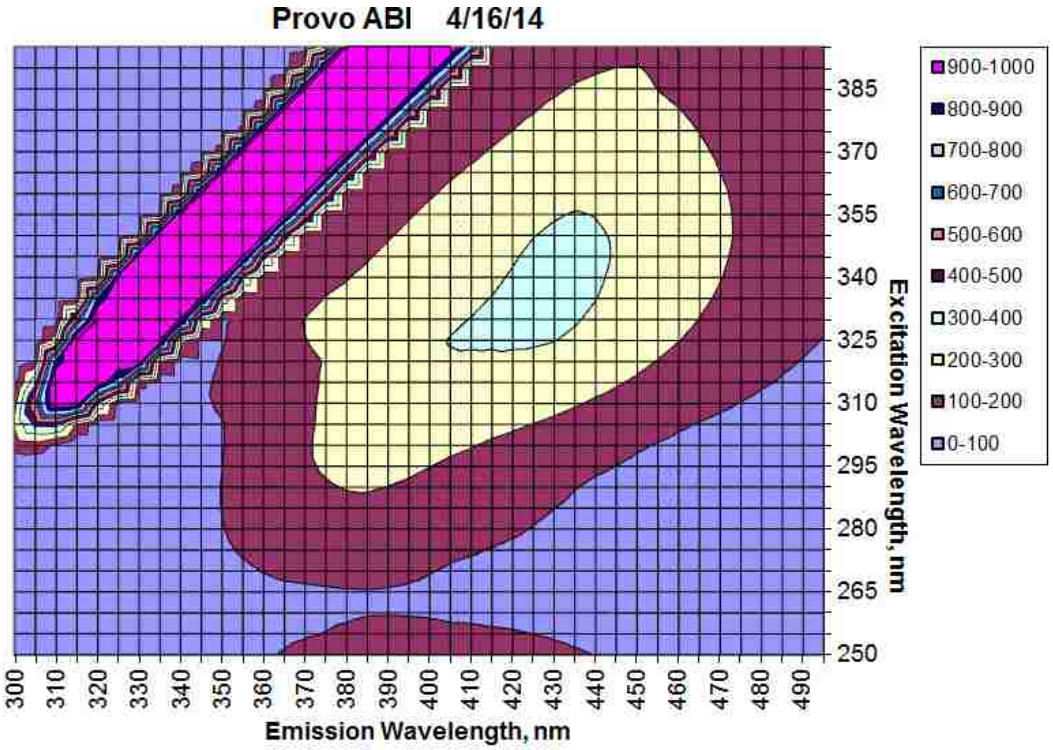


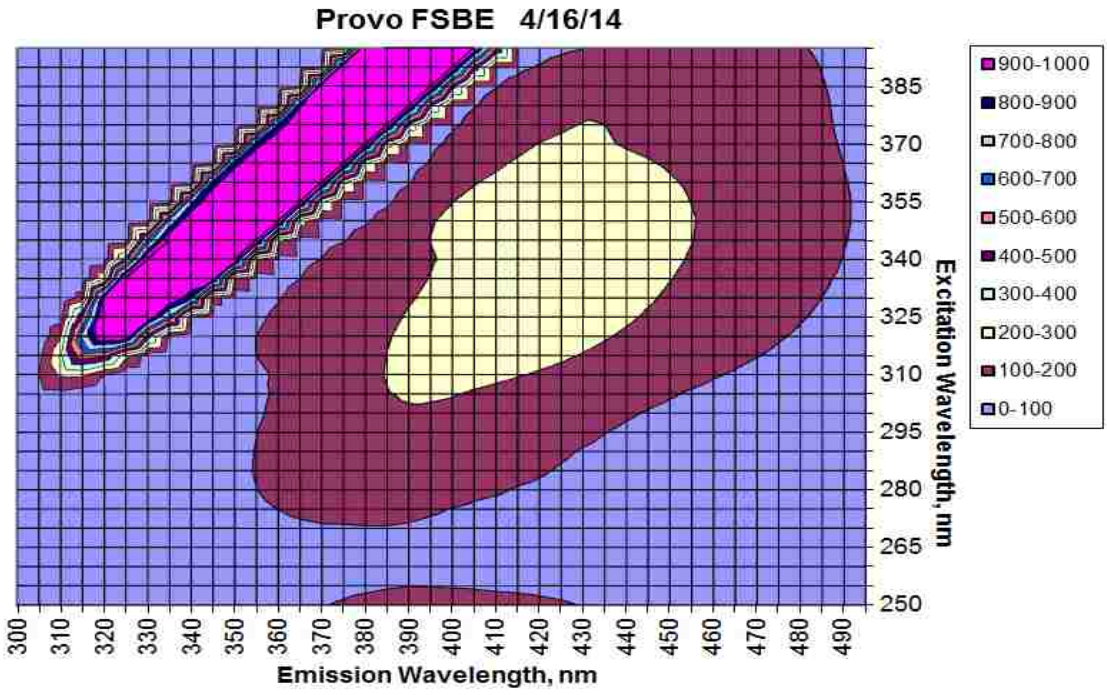


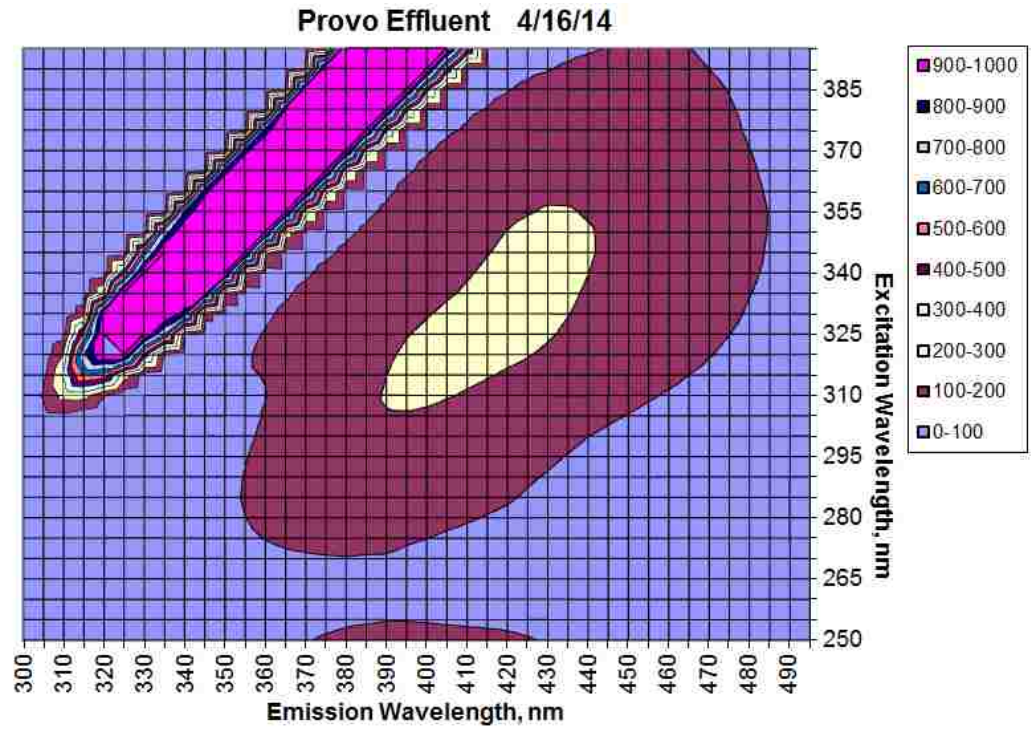


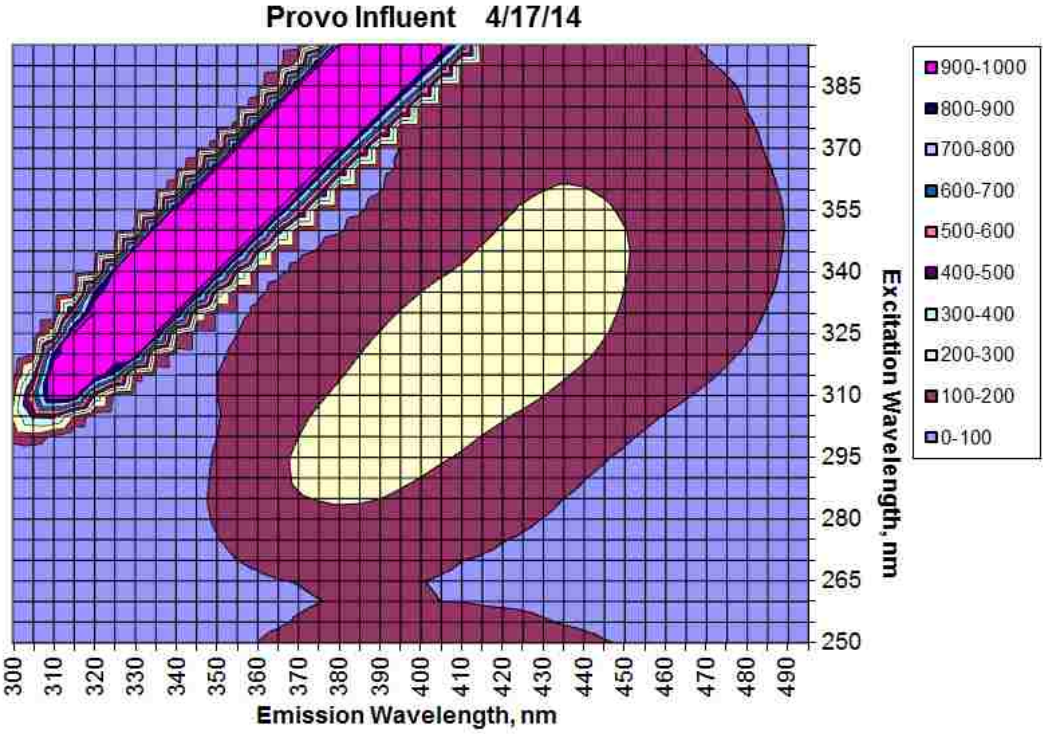


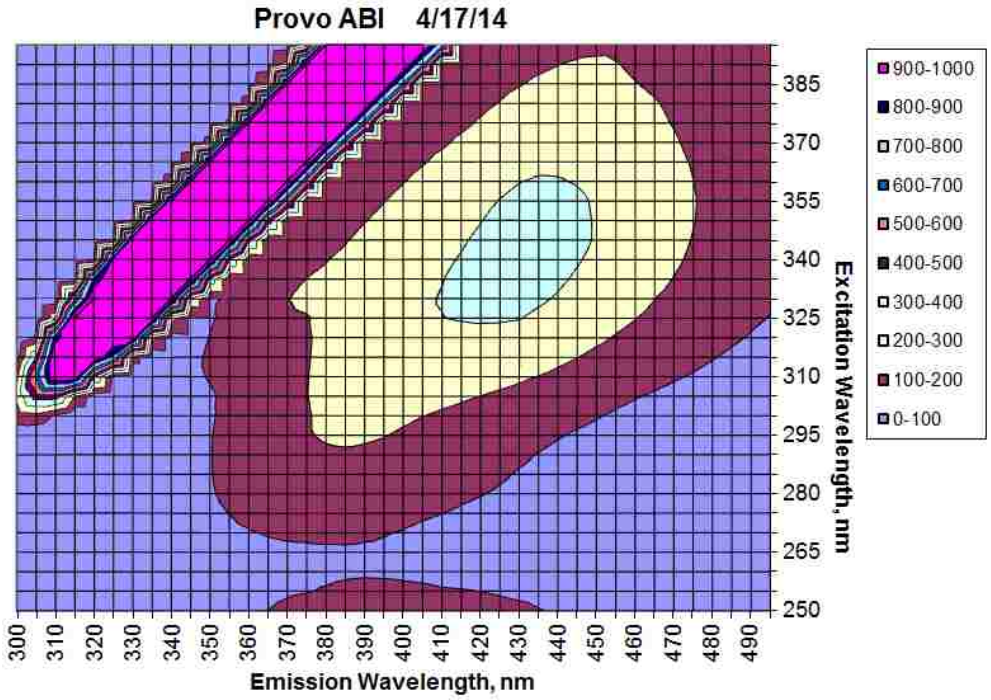


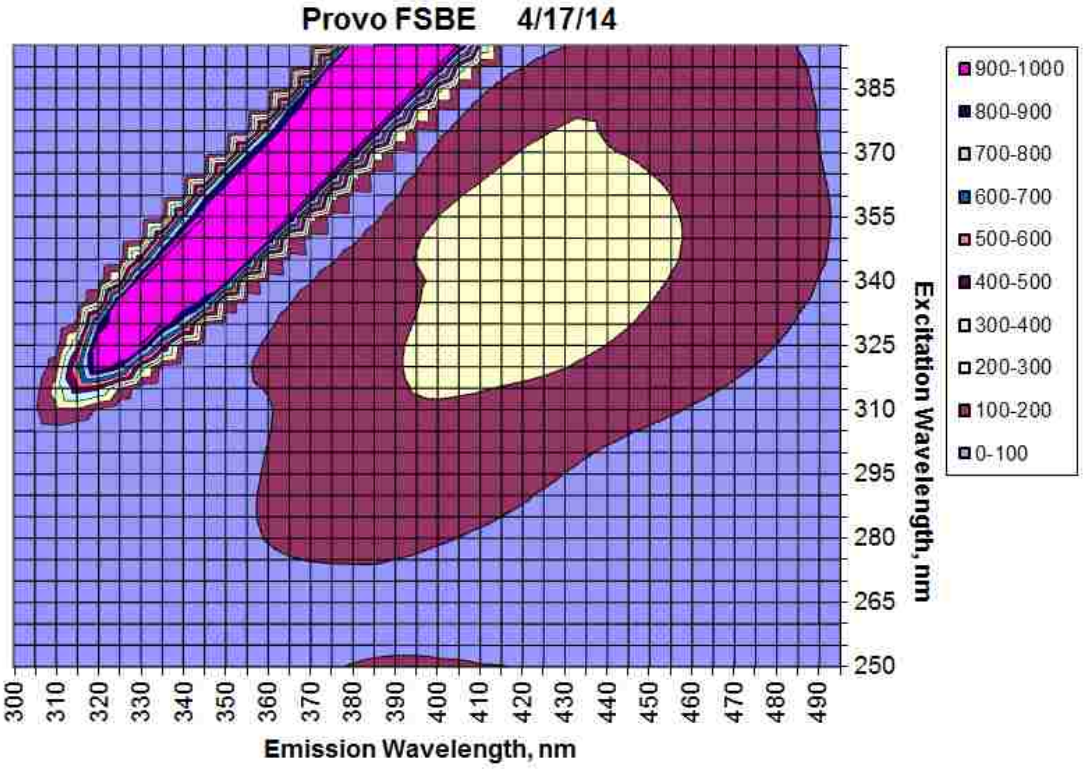


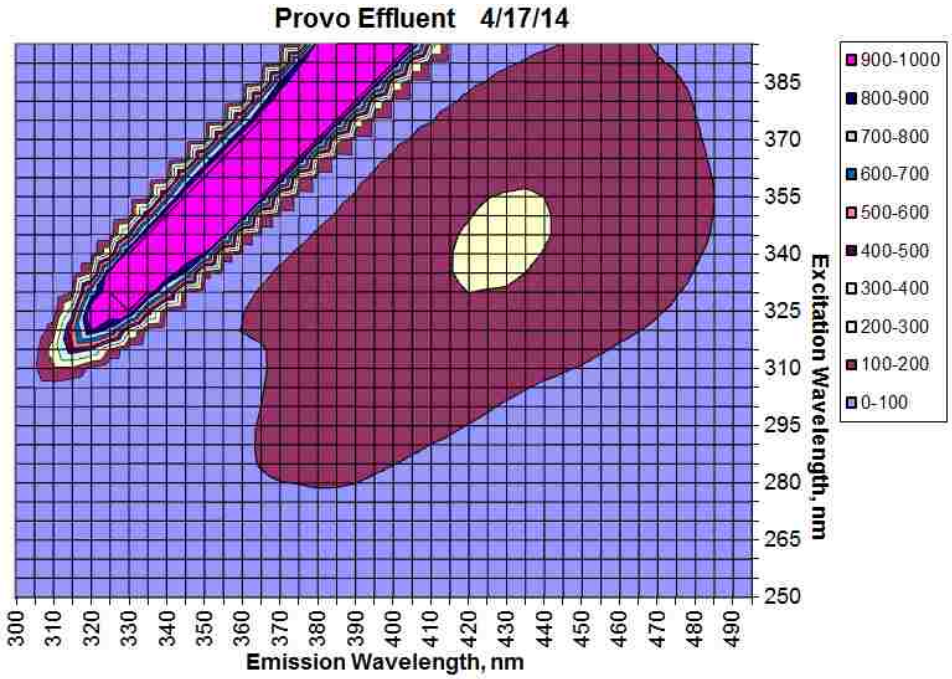












A.5 Algorithm Code

```
# Total
totalset <- rbind(oremset,provoset)
totalBOD <- c(oremBOD,provoBOD)


#Perform once
dropfrac<- 1/2
numX <- 10
cuts <- 6
totiter<- 100000

#Perform after each complettion
ourmax <- rep(NA,cuts)
ourmean <- rep(NA,cuts)
bestR-squared<- rep(NA,cuts)
R-squaredvals<- rep(NA,totiter)
randcolmat <- matrix(NA,totiter,numX)
currentdata <- as.matrix(totalset[,1:1200])



for(j in 1:cuts){
  for(i in 1:totiter){
    x <- dim(currentdata)
    randcolmat[i,]<- sample(1:x[2],10,replace=TRUE)
    x <- as.matrix(currentdata[,randcolmat[i,]])
    y <- summary(lm(totalBOD~x))
    R-squaredvals[i]<- y$r.squared
  }

  isitbest <- c(R-squaredvals > quantile(R-squaredvals,dropfrac))
  R-squaredvals1<- R-squaredvals[isitbest] #New vector with R-squared values that are best
  check <- rev(order(table(c(randcolmat[isitbest,]))))
  check1 <- check[1:(1200*(dropfrac)^j)]
  currentdata <- currentdata[,check1]
  ourmax[j] <- max(R-squaredvals)
  ourmean[j] <- mean(R-squaredvals)
}
#BEST Subset
```

```
require(leaps)
try4 <- leaps(currentdata,totalBOD, nbest = 3, method = "R-squared")
currentdata <- currentdata[,try4[[1]][28,]] #I need to work on this more.
#Fitting last model
final.lm <- summary(lm(totalBOD~currentdata))
finalpR-squared <- final.lm$r.squared
#Filling in the last values
tot.finalmeans[2,] <- ourmean
tot.finalmaxes[2,] <- ourmax
tot.finalset <- as.data.frame(currentdata)
tot.final.cols[2,] <- colnames(tot.finalset)
tot.final[2] <- finalpR-squared

int.total.model <- as.data.frame(totalset[,c(312,342,311,855,464,631,1173,1036,1112,1158)]) #
The columns we obtained.
int.total.model$result <- totalBOD
backward.var.selection <-
regsubsets(result~.,method="backward",nvmax=10,data=int.total.model)
stuff <- summary(backward.var.selection)
stuff$rsq

model10 <- summary(lm(result~.,data=int.total.model))



# Don't touch
tot.finalmeans <- matrix(NA,nrow=20, ncol = 6)
tot.finalmaxes <- matrix(NA,nrow=20, ncol = 6)
tot.final.cols <-  matrix(NA,nrow=20,ncol=10)
tot.final <- rep(NA, 20)
```