Spring 5-15-2022

Monitoring Bacteria Cultures Using Near Infrared (nir) Binary Spectronephelometry (bsn), Raman Spectra and Principal Component Analysis (pca)

Steven Ortiz
Syracuse University

Abstract

Current noninvasive methods cannot continuously and simultaneously monitor the concentrations of cells and media components that define the state of native bacterial cultures, because of changing turbidity. A new technique, binary spectronephelometry (BSN) has the same or better sensitivity and precision for population monitoring as optical density at 600nm (OD_{600}), while simultaneously measuring metabolic processes. The BSN algorithm uses laser induced emission to probe mildly turbid media i.e., propagation of light occurs in the single scattering regime. A BSN "training set" associates a grid of elastic emission measurements, comprising Rayleigh and Mie scattering, and inelastic emission measurements, comprising fluorescence and Raman emission, with a grid of varying known concentrations of bacteria and media. For the normal lifecycle culture and for some but not all growth media, the results confirmed the linear dependence of the two measured signals on the concentration of bacteria and medium dilutions, and agreement with OD_{600}. Having thus confirmed basic assumptions and a calibrated algorithm, various bacteria cultures were monitored over time, producing quantitative "growth curves" for the bacteria and "depletion curves" for media. Simultaneously obtained Raman spectra demonstrate glucose uptake and conversion into bacterial membrane phospholipids. Beyond the normal life cycle, a few examples of using BSN to monitor cultures subjected to external perturbations are shown. Specific characteristics of Principal Component Analysis (PCA) of Raman spectra, and their variation in bacterial cultures, as well as assessing the role of bacterial transport e.g., diffusion in performing and analyzing BSN, are presented.

Monitoring Bacteria Cultures Using Near Infrared (NIR) Binary Spectronephelometry (BSN), Raman Spectra and Principal Component Analysis (PCA)

by

Steven Ortiz

B.S., Rhode Island College, 2012
M.S., Syracuse University, 2015
M.Phil., Syracuse University, 2016

Dissertation
Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Chemistry.

Syracuse University
May 2022

Acknowledgements

The assistance from Dr. Hewen Zheng of the Luk Research Group, Dr. Jarrod Grossman of the Kahan Research Group and Dr. Elizabeth Raymond-Castellano of the Korendovych Research Group. Collaboration from Professor Goodisman, Dr. Richard McDonough, Dr. Paul Dent, Dr. Seth Fillioe and Minh Dinh were invaluable and appreciated. And the funding provided by LighTouch, Inc. and instrumental support from Lambda Solutions Inc. for this and other research studies is also very much appreciated

Table of Contents

List of Illustrative Material

1. Introduction

Whether for research involving bacteria in small-scale cuvette-based experiments or monitoring large-scale bioreactor performance i.e., the physical and chemical states of the culture and the medium, there is always an advantage in being able to do so continuously, noninvasively and without physical sampling of the culture. Spectroscopic methods have evolved with the enabling technologies and today near infrared (NIR) Raman spectroscopy shows considerable promise for such applications. Despite producing well-known autofluorescence and photobleaching behavior, the absorption of NIR is small enough to allow laser probing of biological materials in general without cell damage.[1] The turbidity of biological materials presents a challenge because of the requirement to account for the light propagation effects in addition to absorption i.e., simple Beer's Law analysis. In this thesis we explore the use of a new algorithm that increases the spectroscopic and physical optical information that can be collected from the remitted light produced by such laser probing.

The traditional *in situ* physical optics-based method employed by biochemists for determining the concentration of bacteria in media solution is the attenuation of light passing through a sample at 600nm i.e. "turbidometry" and is also known as the optical density or OD_{600}.[2] Developed with the appearance of commercial spectrometers as an improvement on "dry weight" as a measure of culture density, OD_{600} gives users limited information, at best only the concentration of cells in solution and no chemical information such as media composition. It is also susceptible to artifacts due to absorption or scattering losses not related to the bacteria concentration. Since bacteria are known to create biofilms that can occur in concert with various environmental and metabolic changes, systematic scattering losses have probably always been an

artifact of the use of OD_{600} to provide any kind of continuous measurement of a culture's "vital signs".[3]

1.1. Binary Spectronephelometry

Optical probing of bacterial cultures is one example of analyzing a mildly turbid sample i.e. an optically thin or dilute sample, using a general approach based on radiation transfer theory devised by Goodisman and Chaiken to simultaneously obtain turbidity and composition information in other biological systems.[4, 5] For various reasons when this algorithm is applied to noninvasive *in vivo* blood and tissue analysis in humans it is called either PVOH or PV[O]H, but when it is applied to *in vitro* systems, it is called "Binary Spectronephelometry" (BSN). BSN was developed to determine bacteria concentration simultaneously with analyzing the media chemically e.g., glucose or other feedstock concentration or products of metabolism. Since there are better i.e., more sensitive, accurate and precise measures of various culture properties e.g., flow cytometry, high performance liquid chromatography, the basic advantage of BSN is that it noninvasive, nondestructive and continuously provides multiple simultaneous measures of culture status.

For practical purposes it would be helpful to have analytics of the kinds of responses a culture produces in response to various external stimuli. The essentially continuous responses that can be observed using BSN present the opportunity to fully automate and document the maintenance of large-scale bacterial cultures. For example, added bleach and/or exposure to ultraviolet light would be expected to affect the bacteria and/or the medium. With BSN it is possible to discern such changes in real time as we show in this thesis.

Because the spectra obtained from cultures are very complex, mostly due to the chemical complexity of legacy media like lysogeny broth (LB) media itself, we explore the use of principal component analysis (PCA) for reducing the spectral information to a more finite number of variables. Lysate derived culture media e.g., LB media have disadvantages from the point of view of large-scale biotech application and so there has been a drive toward using "chemically defined" media in which minimal media (MM) is used. MM contain a single well-defined carbon source and only other chemicals deemed essential i.e., a single source of nitrogen, phosphorus, and other elements etc. In this case the possibility of using Raman features for the individual components is quite enticing because of the degree of control that might impart. In this thesis we detail the study of BSN applied to three types of chemically defined media i.e., lactate and glucose as only carbon source, and metal containing medium. We have set the stage for a broad application of big data analytics to commercial bacterial chemical production using BSN to provide the relevant real time input data.

With accurate knowledge of all the atoms available to the organisms, it should be possible to continuously monitor and confirm mass and energy balance during a culture "lifetime". Although we still have further to go to achieve this goal, we believe we have observed the Raman signature of production of a "mucus like" substance by *Pseudomonas aeruginosa* cultures although we cannot unequivocally associate exclusively with specific Raman features at this time. But if we could identify the chemical structure of this substance, we could balance it against the consumption of carbon containing media molecules. This is complicated at best in LB media, but seemingly possible in minimal media and with various sources of carbon e.g., lactic

acid, glucose.

In special cases such as bacteria that can process e.g., metals and develop static electrical potentials across the culture, the possibility of balancing mass and redox potential exists. We explore this in the case of *Shewanella oneidensis* and although we do not achieve that goal in this thesis, we believe that significant progress was made. We show that complexation of copper and possibly other metals with media components produces coordination compounds that absorb and fluoresce in the NIR, and because of hydrolysis, the BSN calibration procedure cannot presently be followed.

BSN has limitations like standard nephelometry instrumentation or perhaps flow cytometry or other cell counting approaches. For the purposes of demonstrating the applicability of BSN to monitor bacterial cultures, the present apparatus was adequate. A highly optimized and engineered BSN system will minimize 1) the interaction of stray incoming laser light with the sample container e.g., a cuvette or viewing port into a large bioreactor, 2) manage any reflection of incoming light off the top liquid surface if applicable and 3) cope with gross spatial inhomogeneity of the culture due to macroscopic deposits and floaters due to fouling. We have documented both effects 1 and 2 before and we do so again in this thesis.[4] We also anticipate that with enough experience on monitoring such systems, the raw data analysis itself will reveal macroscopic fouling when problematic. Two main advantages of BSN will be the minimization or simplification of sample preparation and the ability to monitor various processes *in situ*.

Indeed, OD_{600} was intended for use only in small, lab scale experimentation but we believe BSN

could be optimized for use in large scale commercial operations involving bacterial cultures. To this end we describe the use of Principal Component Analysis in a manner that may find particular application towards that end. We hope this approach will lead to possible versions of a BSN based devices that do not use multichannel detection, can be miniaturized and require less power. Portable battery-operated systems for monitoring large scale biotech operations would be a long-range goal. Possibly a BSN based device could find application in on-site monitoring of water based environmental chemistry.

BSN is a form of multispectral imaging or analysis.[6] Arguably, the simplest type of multispectral optical system that can probe any medium involves only one light source and two light detectors. This type of system would seem to invite small embodiments that utilize only optical filters for spectral dispersion single channel detectors. In fact, we recently demonstrated progress in miniaturization in the context of the latest PVOH device.[7, 8]

Finally, and possibly to that end, one of the bacteria we have chosen to study is capable of metabolizing metals to produce electrochemical gradients. Some have suggested that such bacteria could someday be used to produce electricity at wastewater treatment plants.[9,10,11] Such small, portable BSN devices/systems could be useful both in exploring the possibilities in the lab and implementing actual practical systems in the field. Indeed, OD600 was originally implemented commercially using simple one or at most two cuvette devices at a wavelength selected by dial. Although crude by today's standards, the practical impact of replacing dry weight with OD600 as a means of establishing bacterial density in many kinds of research and

development involving bacteria nearly 50 years ago cannot be over-estimated. We suspect that a similar evolution could occur around BSN.

To establish the theoretical validity of the BSN method the relationship between the scattered light signal and the concentration of bacteria in situ within the single scattering range is ascertained by a training set. We have learned that the application of the BSN to bacterial growth in laboratory conditions, and probably other situations, benefits from being able to characterize propagation of cells through the medium. Both issues are dealt with together in the first part of the results section below. The biochemical implication of the Raman spectrum and its analysis by Principal Component Analysis follows.

Appendix A contains a complete, peer reviewed, published account of some of the research presented in this thesis that is presently in the primary scientific literature. This thesis overlaps with that manuscript, but contains much more results, analysis and discussion involving other bacteria, other media, and other aspects of BSN e.g., transport of bacteria in relation to optical probing, applied to various aspects of bacterial cultures in general. The next chapter provides a general summary of BSN as applied to *in vitro* bacterial cultures. The theory of the BSN algorithm and demonstrate its application to real-time monitoring of bacterial cultures. Raman spectra of cultures obtained simultaneously are analyzed by Principal Component Analysis (PCA) and can provide real-time metabolic information concerning the bacteria and its effect on the medium. We spend considerable attention to PCA in this thesis because we find confusion in the application of such chemometric techniques in the literature and because a dedicated chapter could prove useful to later group members in their research.

Beyond that we find that whereas a complex legacy medium like Lysogeny Broth (LB) presents no serious problems in the application of BSN, use of a metal supplemented minimal medium has resisted the empirical calibration procedure we have already accomplished for both LB and metal-free minimal medium. This includes minimal with either lactate or glucose as a sole carbon source.

We provide minimal background on the particular bacteria selected for experimentation because our immediate goal is to explore the generality of BSN as a solution to the challenge of obtaining spectroscopic information from turbid systems. “Obtaining information” implies that some probing light must enter the turbid system and exit as light of the same wavelength due to elastic interactions within or as light of a different wavelength due to inelastic interactions. BSN requires both kinds of light be remitted.

As long as the “turbid system” comprises a volume containing entities that can elastically scatter light and/or produce inelastically scattered light, which can include fluorescence, there is the potential that an optical system can be designed that could probe the volume collecting only single scattered photons from the probed volume. In those cases, BSN can provide information as advertised. For the first time in our experience found a system, certain metal supplemented media, for which this does not work, and we show why.

1.2. BSN Algorithm and Training Sets

The empirical observations in Figure 1 frames our approach to presenting the BSN algorithm. Using a modified Raman spectrometer, one incident laser produces an emission spectrum containing two types of light, scattering and fluorescence. The first is the Elastic Emission (EE) that is displayed in the Rayleigh line region of the spectrum i.e., Raman shift = 0 cm^{-1}. As the light from the laser passes through the sample some of the photons experience physical i.e., Rayleigh or Mie scattering, producing no shift in wavelength and appearing in the spectrum as a narrow peak. The second type of signal is Inelastic Emission (IE) and arises from inelastic interactions between the incident photons and molecules. The molecules absorb/exchange the energy from/to the laser photons and fluoresce, or change their vibrational mode, in either case causing detectable scattered light with a shift in wavelength. This can be seen in the Raman and fluorescence region of the spectrum as shown in Figure 1.

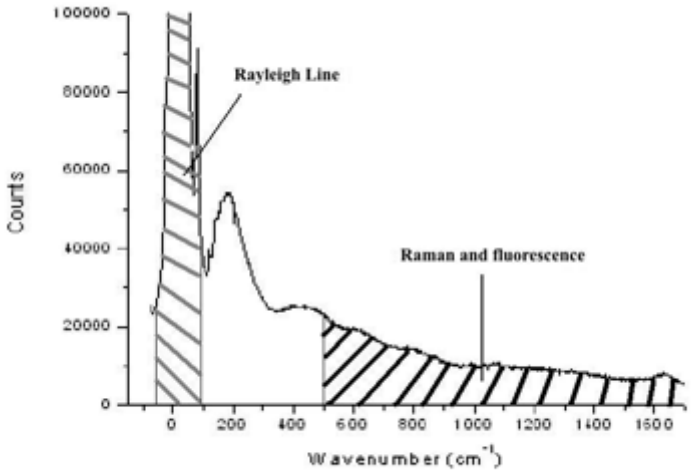

Figure 1: Spectrum showing the wavenumber vs. photon count, the Rayleigh line (EE) and the Raman and fluorescence (IE) regions. Note that the Rayleigh line is cut off roughly an order of magnitude below its actual maximum to preserve dynamic rage in the representation of the remitted light caused by one incident near infrared (NIR) laser.

The EE and IE an algorithm was developed by Chaiken and Goodisman.[12] The algorithm relates the number of photons scattered elastically and inelastically to both the volume fraction of bacteria (ϕ_{bac}) and the volume fraction of media (ϕ_{med}). It was necessary to develop the algorithm because the bacteria cells not only contribute to the EE scattering but since the cells themselves are composed of molecules such as lipids, saccharides and DNA, they affect the IE signal as well. The liquid medium phase also contributes both EE and IE and we explore the effect of these multiple sources on the efficacy of the algorithm.

The sum of the volume fractions of the bacteria, media and the scattered light caused by the bottom of the cuvette accounts for the entire volume of the solution and can be set to equal one. Although light from the cuvette can usually be avoided with good optical design, as we show below, typical medium and growing conditions produce elastic scattering from debris in the medium justifying the inclusion of a generalized source of EE i.e., ϕ_{cuv}.

$$1 = \phi_{bac} + \phi_{med} + \phi_{cuv} \quad [1]$$

If we take the probed volume to be a constant, then sum of any changes of these volume fractions will be zero. The volume fractions can change in *in vivo* applications due to e.g., blood movement or *in vitro* evaporation of water from liquid media. In reality, this assumption is effective so long as the probing regime is not near the transition turbidity between the single scattering and multiple scattering regime.

$$0 = d\phi_{bac} + d\phi_{med} + d\phi_{cuv} \quad [2]$$

From radiation transfer theory,[1] and given latitude in designing the optical geometry for probing, in principle it is always possible to find a linear range for propagation losses in turbid solutions between zero and some higher concentration of physical scatterer e.g., quartz spheres, bacteria or red blood cell concentration.

Once experimental data has confirmed empirically that the relationship between the volume fractions and the scattered light is linear for a specific range of concentrations, then equations 3 and 4 apply giving the relationship between the EE and IE signals to each of the volume fractions.

$$EE = \vartheta_1 + \vartheta_2\phi_{bac} + \vartheta_3\phi_{med} \quad [3]$$

$$IE = \vartheta_4 + \vartheta_5\phi_{bac} + \vartheta_6\phi_{med} \quad [4]$$

As stated before, the bacteria cells contribute to both the EE and IE and the media could potentially do the same so both independent variables are needed in calculating both the EE and IE. As long as the causes of EE and IE are physically independent processes, Equations 3 and 4 are linearly independent equations in two variables that can be measured. Hence, they can be algebraically inverted so that the measured quantities (EE and IE) are the independent variables, and the volume fractions of bacteria and media can be determined as shown in equations 5 and 6.

$$\phi_{bac} = a + b(EE) + c(IE) \quad [5]$$

$$\phi_{med} = d + e(EE) + f(IE) \qquad [6]$$

The coefficients a through f can be expressed in terms of scattering and absorption coefficients, geometry, laser and other properties of the experiment but in practice they can always be found experimentally by measuring EE and IE on a "training set" of systematically varying concentrations of bacteria and media, using e.g., phosphate buffered saline (PBS) to adjust volume and media fractions as needed. With actual measurements, geometric and other practical issues are incorporated empirically into the training set. A bilinear regression of e.g., ϕ_{bac} on EE and IE produces parameters a-f.

The choice of reference solutions to calibrate the algorithm can accentuate the ability to observe small changes in specific characteristics. For example, we may anticipate that a real culture may produce particulate debris that scatters differently from the bacteria themselves. By choosing to include or exclude such particulates e.g., by filtration from the media during calibration we could "build-in" the EE produced by the media particulates. We might minimize interference in measuring the number density of bacteria although that could be completely ameliorated if the bacteria and particulates each produced distinctive Raman features.

A further trade-off involves the choice of linear range over which to calibrate the algorithm. The narrower the range, and the lower the average turbidity over the range, the better the linearity. But if utility requires a broader range, or that it extend to higher turbidity e.g., for bacteria concentration, some systematic error may be acceptable, because it can be compensated. For *in*

vivo systems, maintaining homeostasis is a driving force and the ability to monitor a narrow range of parameter space surrounding the optimal homeostasis conditions may be the only requirement to be clinically or technologically useful.[12]

1.3. Bacterium of Examination

1.3.1. Pseudomonas aeruginosa PA01

Pseudomonas aeruginosa PA01 (PA) is a gram-negative bacterium capable of organizing individual cells into an aggregate that form a biofilm protecting the collective from harsh environmental conditions.[13] PA is also a pathogen capable of causing infections in patients with immune compromised respiratory diseases like cystic fibrosis (CF) and chronic obstructive pulmonary disease (COPD) as well as people with physical trauma such as burns and abrasions.[14] Among patients with immunosuppressed diseases the mortality rate from PA infections is 40% to 60%, including patients in the burn and palliative care units in hospitals.[15] In the United States over 200,000 cases per year of severe infections arising from hospital visits alone have been reported with PA being the leading causative agent.[16] Since we know PA produces such films we expect that OD_{600} might be susceptible to measurements artifacts if used to monitor a single culture in a cuvette over time. Providing researchers with a more efficient tool for studying *PA* cells and their environment can reveal more information on their behavior as they transition from planktonic to sessile cells.

There have been studies that involve spectroscopic analyses of PA; some of these studies use Surface Enhanced Raman Scattering (SERS) or Raman microscopy amid claims that it can be utilized to identify bacteria at the single cell level.[15] During growth PA produces a mucus-like

substance believed to be composed of polysaccharides that float to the surface, during experimentation this substance interferes with the scattering of the laser which makes BSN monitoring of overnight cultures impossible.

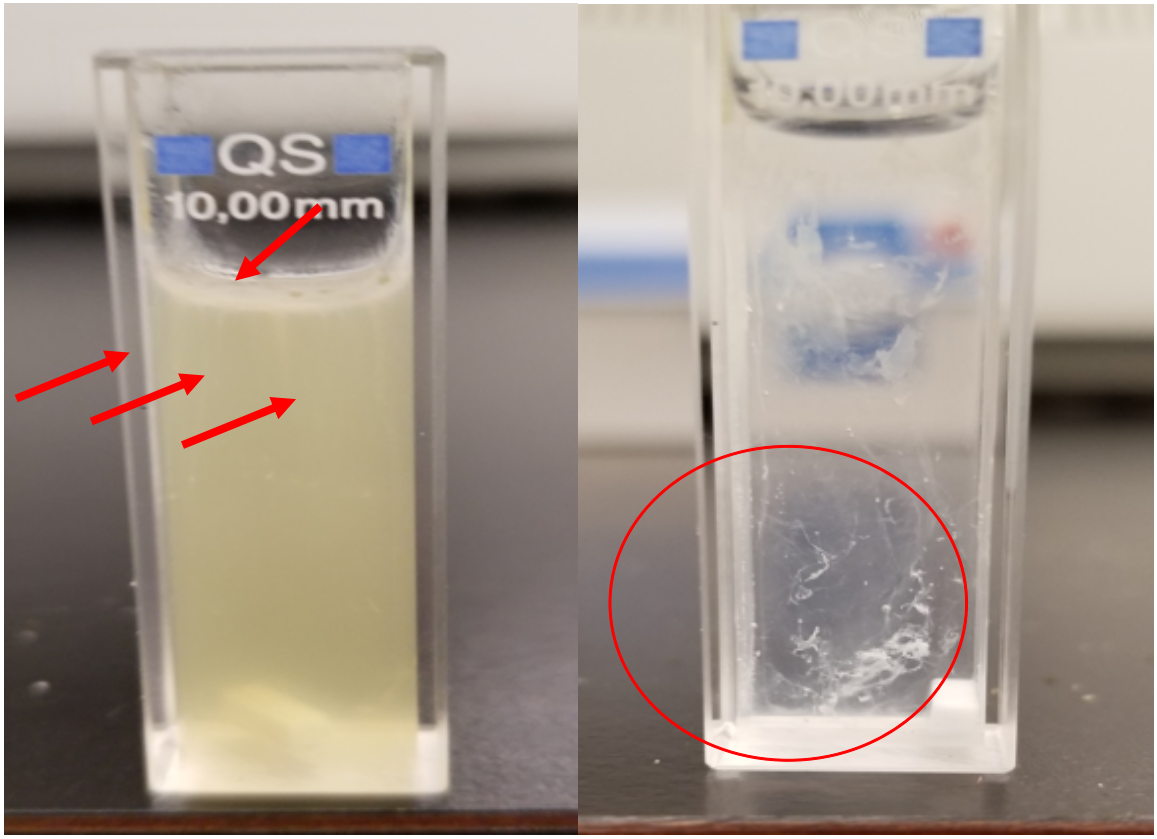

Figure 2: A photo of a PA overnight culture which shows the mucus-like substance at the surface and extending downward through the media (left) indicated by the red arrows and circled region. Then the culture was allowed to stand for an hour to settle, the media was then pipetted out and the cuvette filled with PBS (right).

The current technique for determining the concentration of bacterial cells is by measuring OD_{600} but this method provides no information on the medium. Crystal violet (CV) dye assay is used to determine the thickness of biofilm formed by PA. The dye is diffused through the cell membranes, then washed with a saline solution to remove excess CV and analyzed through UV-Vis spectroscopy.[17] However this process is limited because the diffusion of CV into the biofilm

is random and the number of wash cycles varies from group to group, all of which can affect the data collected and make this process time consuming. The purpose of this research is to determine whether light scattering and the BSN algorithm can be used to monitor quantitatively the physical and chemical environment of viable bacterial cells in solution.

1.3.2. Shewanella oneidensis MR-1

Shewanella oneidensis MR-1 (SO) is a wild type of Gram-negative rod-shaped bacterium capable of respiring aerobically and anaerobically. During anaerobic respiration, it has been suggested that SO is proficient at electron transfer across its outer membrane through a porin-cytochrome to extracellular deca- or undeca-heme cytochrome acceptors that then act as reductase of multivalent metals.[18,19] This makes SO a possible choice for microbial fuel cells (MFC) which are being studied as a potential for alternative energy. *S. oneidensis* is also being researched as a means of cleaning up hazardous metals, even in mineral form. One goal for many environmental scientists is to develop fuel cells that generate electricity during water treatment.[20] Whether it is for environmental or engineering purposes, having the ability to continuously monitor the concentration of bacteria and media constituents should prove valuable for all applications of *SO* and might also be useful in the practical applications themselves. This is especially true for MFC where the electromotive force i.e., voltage of the cell is dependent on the bacterial cell population.[21]

That's where we hope this research can lead, an experiment to continuously observe a bacterial culture in an operating MFC using BSN while monitoring the cell voltage. The effects on the MFC voltage/current characteristic of adding extra fuel e.g., lactate or different fuel e.g., glucose

at a specific bacteria concentration can be compared to the scaling of voltage with increasing bacteria concentration at constant fuel input. At this point, in order to validate an overall MFC model, we make this comparison to check for internal consistency between our measured parameters e.g., rate of glucose or lactate consumption and published MFC voltage-current characteristics.

1.4. Bacterial Diffusion

We observe that under our growth conditions PA growth curves, for example, appear roughly concave up whereas for *Escherichia coli*, figure 3, the shape is consistently concave down. We believe these contrasting shapes reflect differences in transport of bacteria e.g., by diffusion as the total concentration of the bacteria increases and the difference in growth rates for the bacteria. We also observe that the progress of the E. Coli growth is always relatively smooth whereas the PA growth is always uneven, although the total factor increase over six hours for the PA is nearly as much as for the E. Coli.

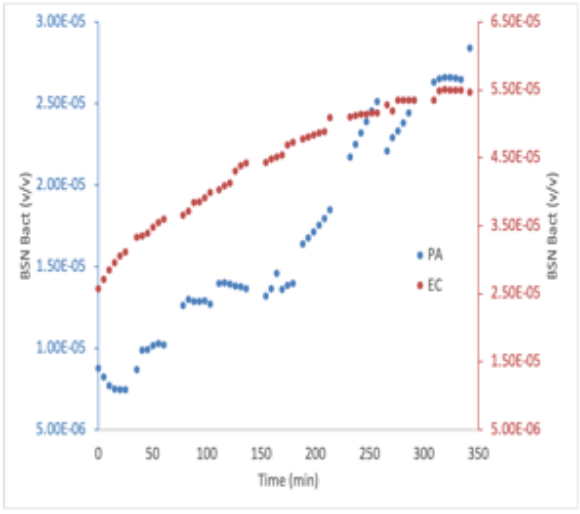

Figure 3: The comparison of the growth curves for E. coli (red) and P. aeruginosa (blue).

Conventionally, as bacteria concentration is tracked by e.g., physically sampling and measured by OD_{600}, if an aliquot of "stationary phase" culture is added to sterile fluid culture medium e.g., LB, the concentrations is observed to increase over time and the resultant growth curve is described in four phases. The first phase of growth is called the lag phase where the cells are just beginning to multiply, and bacteria concentration appears to increase linearly. Next is the "exponential" or "log" phase when every cell split into two new cells. How often this occurs is characterized by the "doubling time". As resources start to decrease the growth will enter the "stationary phase", cells stop splitting and the bacteria concentration becomes constant.

These names for the different phases of culture phases follow the observation that a graph of OD_{600} bacteria concentration versus time during exponential phase on a logarithmic scale is linear. We have done this for our systems, and this is not the case. It should be pointed out that normally a 1cm cuvette is probed at least halfway down the cuvette in making an OD_{600} measurement whereas the BSN results in this paper are made more near the surface of the liquid, near the medium-air interface. The position of the probed volume is near the surface in BSN monitoring 1) in order to reduce the amount of elastic and inelastic scattering as these photons traverse to the eventual detector and 2) because we calibrated the BSN algorithm for the physical arrangement of the probed volume described below. This paper is meant to explore the source of this observation.

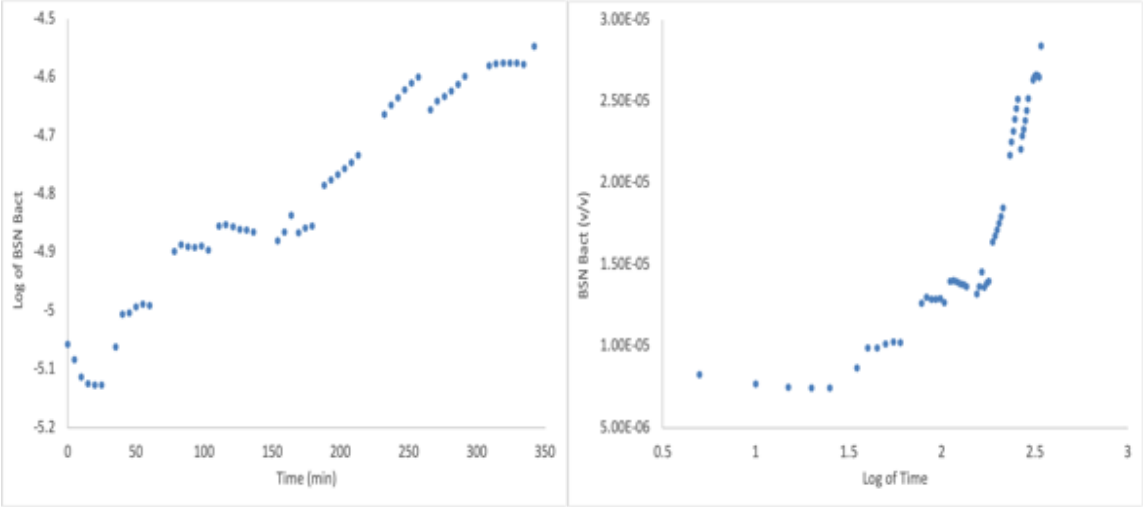

Figure 4: The six-hour growth curve for PA with the BSN bacteria volume fraction on the logarithmic scale (left) and the time on the logarithmic scale.

The observed increase in bacteria concentration during a culture lifetime i.e., a "growth curve" reflects the change in bacteria number with time as well as the transport mechanisms in play that distribute the bacteria into and out of the BSN probed volume. We also note that the medium "volume fraction" i.e., the materials in the medium that contribute to the IE simultaneously decreases in the same time frame. This reflects the fact that, other than evaporative loss of water, which is replaced manually at specified time intervals, producing an unmistakable (but sometimes useful) artifact in both the bacterial and medium volume fractions because of our apparatus optical design, we observe the matter in the medium being transformed into bacterial cells and mucous over time.

The use of OD_{600} was intended to reduce/eliminate the need for using dry weight to characterize bacterial concentrations in various media. An optical growth curve using OD_{600} the measurements can be made with or without stirring, with an additional contribution by any biofilms that may form. Since our main interest in this thesis is to demonstrate the basic

phenomenology for using the BSN algorithm to characterize bacterial systems, we note that the spatial inhomogeneity of an unstirred or unshaken culture would seem to be greater than for an agitated culture. Larger bioreactor applications could also involve stirring and so we chose to include stirring in these experiments because it may help in avoiding fouling of observation ports in larger systems.

We posit that with stirring, nonviable bacteria would be distributed in the available medium volume in accordance with their density relative to water i.e., a gravity enforced bacterial density gradient. In addition, there could be aggregation effects involving not just the bacteria but also the particulates that are inherent in the media. These effects are likely to be complicated and exist to varying extent from specific culture to specific culture, but with stirring, we hoped to create a region of relative quasi-equilibrium i.e., spatiotemporal homogeneity allowing smooth variation of bacterial number density. The smooth and very consistent variation of all E. Coli growth curves reinforced our optimism, but the PA results are not smooth and monotonic and visual observation of the PA samples in the cuvettes over time did not reveal substantial aggregation, involving either PA and/or particulates or bubbles. As the bacteria reproduce, their total number increase, and at any location inside the cuvette we expect bacteria to flow into and out of a probed volume, even as the total amount increases. The expected movement of the bacteria can be passive and/or active.[22,23]

BSN utilizes a laser focused to create a “probed volume” that is much larger than bacteria i.e., $\sim 10^2$ microns in all dimensions, but orders of magnitude smaller than the total volume of the cuvette, that is located ~3mm below the surface of the liquid sample, and ~2cm above the axis of

rotation of the micro stir bar that's located near the bottom of the cuvette. We chose this arrangement for the reasons stated above and so even nonviable bacteria cells would have to travel through the medium, producing fluctuations in the gravitationally enforced bacterial density gradient, to interact with the light to be detected by BSN. This motion could be actively affected by live bacteria themselves and/or e.g., by passive diffusion of particles through a medium as described by the Einstein-Smoluchowski equation.

The bacterial number density in this region of spatiotemporal homogeneity is indicated by the BSN bacterial volume fraction, increasing as the total bacterial number density increases. Note that in the death phase the apparent volume fraction doesn't decrease to zero quickly and the medium volume fraction doesn't immediately rebound. The spectroscopic and physical optical characteristics of viable and recently deceased bacterial cells are nearly unchanged at the level of precision afforded by this choice of IE.

1.5. Principal Component Analysis

The spectroscopy of biological systems in vivo i.e., the nature of the light emitted from a section of biological matter e.g., skin, blood or a bacterial culture is complicated, but amenable to analysis by the same principles as all other systems. For example, pulse oximetry relies on 1) the absorption spectroscopy of hemoglobin and 2) an algorithm that correctly treats the propagation of light in the appropriate limit of turbidity.[24,25] Biological systems in vivo have many and varied outside forms, spanning a range of spatial scales. Internally, biological systems are highly organized structurally and functionally, manifesting extensively and finely correlated chemistry and physics to fulfill the basic requirements for being alive. Because light must traverse from the

outside to the inside and back again, in the analysis of spectra of living systems, it is essential to solve both the problems of spectroscopy and light propagation simultaneously. This chapter deals with the use of statistical analysis aimed at understanding spectroscopic data, but it could be used to deal with issues related to physical form as well.[26]

Any approach to the spectroscopy of such systems is governed by ethical considerations since in vivo biological systems cannot be manipulated with impunity. Living biological systems can be characterized by 1) a state of "homeostasis", as defined by the relative constancy of ethically acceptable vital signs, and 2) how those vital signs transition between states of homeostasis, given some form of external stimulus or change in environment. The time development of the spectra of a biological system whether at homeostasis or in transition produces a series of spectra, reflecting the correlated chemical and physical processes in effect maintaining homeostasis or finding new set points. A dose response study based on spectra of a set of systems at specific time point, being treated with different amounts of e.g., bleach or light could also constitute a homologous series amenable to analysis by PCA. Although not pursued in detail in this thesis, examples of the raw data that could support such an effort are shown later.

General statistical techniques exist to analyze the *internal* correlations within and between analogous series of data. In the context of chemistry i.e., "chemometrics" or more specifically principal component analysis (PCA) presents a relatively unbiased approach to unraveling the spectroscopy of biological systems, assuming only that complex series of spectra can be decomposed in a linear fashion. In this thesis we observe cultures over time from seeding to death phase, always at homeostasis but transforming over time according to a well-defined life

cycle. Additionally, we utilize a few different approaches to perturb bacterial cultures, producing opportunities to observe series of spectra or vital signs derived from the spectra e.g., BSN.

In the context of analysis of human blood in fingertip capillary beds, “tissue modulation” involves the use of thermal and/or mechanical stimulus to produce particular spatiotemporal distributions of mobile tissues, i.e., capillary blood, among nonmobile tissues, i.e., epidermis.[27] In the context of bacterial cultures the BSN monitoring of bacterial density versus time and/or the density of culture medium versus time defines two vital signs. But BSN does not contain the same information as the associated series of complete Raman spectra. The intricate chemical and physical relationships and internal correlations constituting the metabolism of the culture over the life cycle of the bacteria are reflected in those spectra.

One type of internal correlation involves the fact that many biological systems are mostly closed systems, exchanging only certain materials with their surroundings over time e.g., respiration, perspiration and insensible perspiration.[28] A bacterial culture is closed if the cuvette holding the medium is sealed to prevent exchange/loss of gases and water. The same atoms are present for all time, even if the bacteria die, only the chemical bonds changing over time. Thus, for every Raman feature lost due to breakage of the associated chemical bonds, formation of other bonds creates new Raman active vibrational modes. Death corresponds to equilibrium i.e., the Raman features are unchanging with time, except under the effects of passive chemistry.

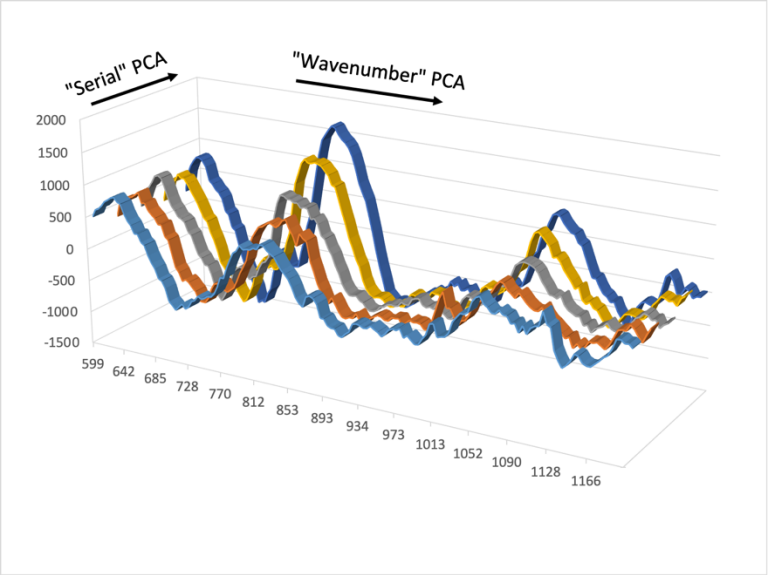

Figure 5: A visual depiction of the treatment of data by the two methods of PCA. “Serial” PCA analyzes the data across spectra and “Wavenumber” PCA across the Raman shift of each spectrum.

Given the nature of the raw Raman Spectra data as in figure 5, chemometric analysis can consider either 1) the variance of the observed Raman intensity across the wavenumber range spanned by each spectrum averaged over the e.g., time series or 2) the variance of the Raman intensity at a specific Raman shifted wavenumber across the homologous series, e.g., a time series or dose response series. There may be other possibilities, but the first approach is useful for 1) assigning features of the Raman spectra to particular molecules and 2) identifying when two or more features change in a manner consistent with a stoichiometric relationship. Similarly, the latter approach lends itself to discerning the optimal Raman shift(s) to monitor correlated variations over the homologous series of time or treatment. See section 4.3.3. below for examples involving data in the training set procedure.

The literature often presents PCA in terms unfamiliar to scientists because PCA is applicable to many types of data and very general terms are used. A primary goal of this chapter is to clarify

the identification of spectral terms and other terms with the terms used to describe PCA in the statistical and mathematical literature.

2. Experimental and Results

Preceding a comprehensive discussion of all results in Section 5, the different types of experiments are presented in Sections 2 and 3 along with the raw results. A detailed presentation of PCA and results of applying it to our data is presented in Section 4.

2.1. Binary Spectronephelometry

Since the BSN algorithm is only rigorously valid across the linear propagation range, the first task was to find this range of concentrations for PA in LB media. The Luk research group, Department of Chemistry, Syracuse University, provided all *P. aeruginosa* PA01cultures used in this research. LB media was prepared with 10g/L Tryptone, 10g/L Yeast extract and 5g/L NaCl in DI water. Then the media were autoclaved at 120°C for 20 minutes. Phosphate Buffered Saline (PBS) was made using crushed tablets (Sigma-Aldrich) and dissolved in 100mL of DI water for 1xPBS solution. Bacteria samples were grown in LB media overnight at 37°C then diluted with media/PBS solution. The training set could be conveniently formed using serial half dilutions of a culture with Lysogeny Broth (LB) media and requiring a constant total volume of 3.5mL in the cuvette to produce a calibration curve. It is essential that each sample i.e., solution have the same volume since the reflection off the top of the solution produces a contribution to the collected EE. The notch filter was positioned in order to provide a numerical balance i.e., comparable signal to noise of EE and IE. This did not necessarily correspond to the optimal position of the filter to achieve the best Raman spectra but once calibrated, the system produces

highly reproducible data. The system used for this pilot study is by no means optimized and a properly engineered system would possess much improved sensitivity, reproducibility and precision.

The OD_{600} of each of the diluted cultures were obtained, immediately followed by a Raman scan with the following settings. The scan was set to 'Accumulation' mode, the number of accumulations was 1500, exposure time was 0.2 seconds, and the cycle time was 0.2 seconds, for a total scan time of five minutes. After subtracting the dark current, the sum of the signals from about $-7cm^{-1}$ to $50cm^{-1}$ was taken to calculate the signals contributed from the EE. The IE signals were calculated by summing up the signals from about $900cm^{-1}$ to the end of the spectrum i.e., about $1900cm^{-1}$, all calculations were performed using Microsoft Excel. The EE and IE signals were then transferred to an Origin Labs Origin worksheet; this program was used to produce all graphs.

To obtain transient data as in Growth Curves the Raman spectrometer was setup with the following settings: Acquisition mode was Kinetic Series, Exposure Time was 0.2sec, Accumulate Cycle Time was 0.2sec, Number of Accumulations was 1500, Number in Kinetic Series was 6, the program should automatically set the Kinetic Cycle Time to 300sec. These settings will give a series of six scans each five minutes long for a total scan time of thirty minutes. A custom-made cuvette holder was used, it was designed to allow heated water to circulate around the cuvette, a hot water thermostat was used to keep the solution at 37°C. At the side of the cuvette holder a modified magnetic stirrer was placed with the magnet nearly parallel

with the closest vertical side of the cuvette. This arrangement insured that the incident laser light did not directly impinge on the micro stir-bar.

To begin obtaining a growth curve, a culture of PA was diluted using LB media to an OD_{600} of between 0.02 to 0.03; 3.5mL of this diluted solution was placed into cuvette along with a tiny magnetic stir bar. Once the cuvette is placed into the cuvette holder the stirrer was powered on and the holder was raised to the highest position so that the light is focused consistently on the same position below the surface of the solution. After each 30-minute kinetic series 25µL of LB media was added to the cuvette to reduce the effect of evaporation of the solvent and after every two series or every hour the cuvette was removed and the OD_{600} taken. The data obtained from the growth curves were analyzed the same way as with the calibration curve, a dark current scan was taken before and after the experiment then the average of the two was subtracted from the spectra.

The volume fraction of bacteria is the volume that the bacteria cells occupy with respect to the volume of the entire solution. To calculate the volume fraction of bacteria an OD_{600} of 1 was set as 6.0×10^8cells/mL[29,30] and the average cell size was taken as $1.5\mu m$ in length (l) and $0.6\mu m$ wide (w).[31,32] Using this information the average volume that the bacteria cells occupy was calculated using the equation below.

$$\phi_{bac} = \frac{Vol\ of\ Cells(mL)}{Vol\ of\ Solution(mL)} = \frac{(OD_{600})(6.0\times10^{-4})(2\pi w^2 \times l)}{V_{sol}} \qquad [7]$$

Once the volume fraction of bacteria has been found for each solution in the calibration curve the

Origin worksheet was used to create a graph of the EE and IE signals versus the volume fractions of the bacterial cultures. Figure 6 below shows the calibration curve that was used to determine the maximum bacterial concentration in the "linear" range i.e., figure 6 (right) which was found to correspond to an OD_{600} of about 0.4. Then a second calibration curve was produced with this maximum concentration to confirm that the range is in fact linear. We know that there will be some systemic error despite the high linear correlation but in appropriate cases it is a worthwhile trade off and can be ameliorated to some extent if a redundant signal is available e.g., Raman features that can be assigned to specific materials. Note that the EE does not decrease between zero and the first solution containing bacteria because the laser impinges on the cuvette bottom in both cases.

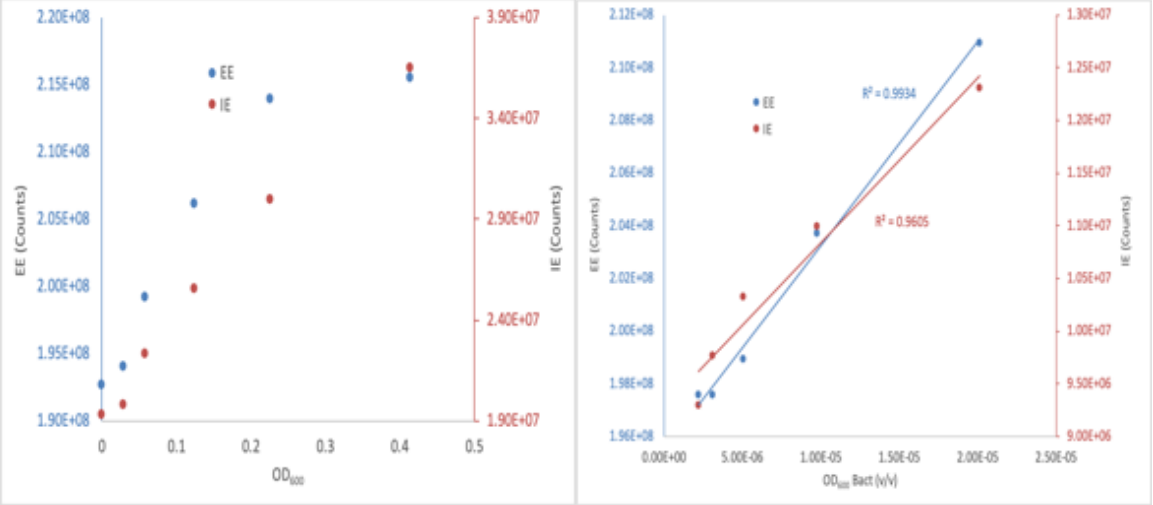

Figure 6: Shows the EE and IE signals as the concentration of bacteria (OD_{600}) increases into the nonlinear range (left). And the volume fractions in the linear range with respect to both EE and IE signals (right).

2.2. Instrumentation and Experimental Setup

All samples were analyzed using both a Perkin-Elmer 950 Lambda UV-Vis spectrometer to obtain OD_{600} and a custom-made 785nm excitation Raman spectrometer. Figure 7 schematically shows the setup of the Raman instrument, the incident light path passes through a clean-up filter

and then a collimator lens. The light is then reflected downward by a 45° angled mirror through a 3 mm aperture drilled through a second downward facing 45° angled metal mirror and focused several millimeters below the surface of the sample solution. With non-turbid samples the laser light impinges on the bottom of the cuvette. Some penetrates both the front and back surfaces of the cuvette bottom and enters a Wood's Horn. The rest is reflected upwards. With no turbidity the laser spot is roughly 3 mm in diameter on the bottom of the cuvette, but the base of the cuvette is slightly tilted so that the majority of the specular reflection is spatially filtered out of the collection optics train. The back-scattered EE and IE light is reflected by the mirror, notch-filtered to remove residual incident laser light, and dispersed by a Kaiser® f~1.4 Holospec spectrograph onto an Andor® Charge-Coupled Device (CCD) camera that measures intensities versus the Raman shift.

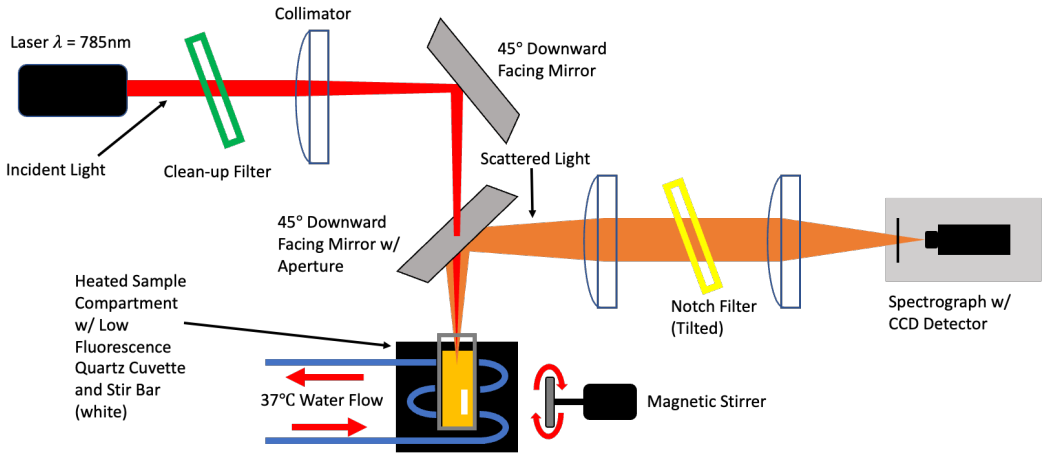

Figure 7: The instrument setup for the custom-made Raman spectrometer with an excitation wavelength of 785nm.

The CCD camera is thermoelectrically cooled to -55°C. The Andor® MCD operating system allows the user to select various scanning options such as mode, scan time, exposure time, etc. which controls the camera. For this research experiment two scanning modes were used during different phases, the first was 'Accumulation' mode that is a single scan for a set time (5

minutes) that produces one spectrum used for the training sets. And the second was 'Kinetic Series' mode, as the name suggests it produces a series of spectra each at a specific length of time (6 scans each for 5 minutes for a total scan time of 30 minutes) used during the growth curves. The dead time between scans is <1 second and between kinetic series is 15 minutes. For every experiment one or more 'Dark Current' spectra were recorded so that it could be subtracted from the sample spectra, this was done by completing an 'Accumulation' scan with the laser powered down. The UV-Vis spectrometer was used to measure the bacteria concentration for each sample at 600nm (OD_{600}) for comparison. For all scans performed using the Raman spectrometer the solvents (LB media and PBS) were filtered with Anatop® 0.2µm 25mm inorganic membrane filters into a clean beaker that is immediately covered with Parafilm®. This is done to ensure that the bacteria cells are the main objects in solution that contribute to the EE; the bacteria culture was not filtered and is known to contain some particulates. All experiments were completed using a quartz cuvette (Amersil®101-QS; 1cm pathlength) with a low fluorescence background.

2.3. Training Set and Bi-Linear Fit

Next in the procedure was the development of a "training set" i.e., a series of solutions with systematically varying concentrations of bacteria culture, LB media and Phosphate Buffered Saline (PBS) chosen to span the linear range. Starting with a recently grown bacteria culture we dilute it with LB media so that it has an OD_{600} of 0.4 or below. Then we prepare six "solvents" that will be used to dilute the culture in each series of the training set; the solvents are serial half dilutions of filtered LB media with PBS. Then each series involved half dilutions of the culture with the corresponding solvent to create a six-by-six training set, each with a total volume of

3mL. Each solution in the training set was analyzed by taking OD_{600} and BSN measurements i.e., EE and IE using the same settings as the calibration curve.

	Sample 1	Sample 2	⋯	Sample 6
Series 1	3.5mL Culture 0mL Media 0mL PBS	1.75mL Culture 1.75mL Media 0mL PBS	⋯	0mL Culture 3.5mL Media 0mL PBS
Series 2	3.5mL Culture 0mL Media 0mL PBS	1.75mL Culture 0.875mL Media 0.875mL Media	⋯	0mL Culture 1.75mL Media 1.75mL Media
⋮	⋮	⋮	⋱	⋮
Series 6	3.5mL Culture 0mL Media 0mL Media	1.75mL Culture 0mL Media 1.75mL PBS	⋯	0mL Culture 0mL Media 3.5mL PBS

Table 1: The volumes of the components (bacteria culture, media and PBS) in each of the samples for the training set. The overnight bacteria culture was diluted to the linear range of the EE signal (OD_{600} = ~0.4) before being used in the training set.

The training set data was collected, analyzed and graphed in the same manner as the calibration curve data, the following show the EE and IE signals as the concentration of bacteria and LB media are independently changed (figure 8). The training set is necessary so that the coefficients a through f in equations 5 and 6 i.e., the BSN algorithm can be determined using a bi-linear regression in the Origin program. Starting with the OD_{600} taken at the start of each series of the training set we calculated the concentration of each dilution. The solvent concentration was calculated by percent volume of half dilutions relative to standard concentrations of PBS.

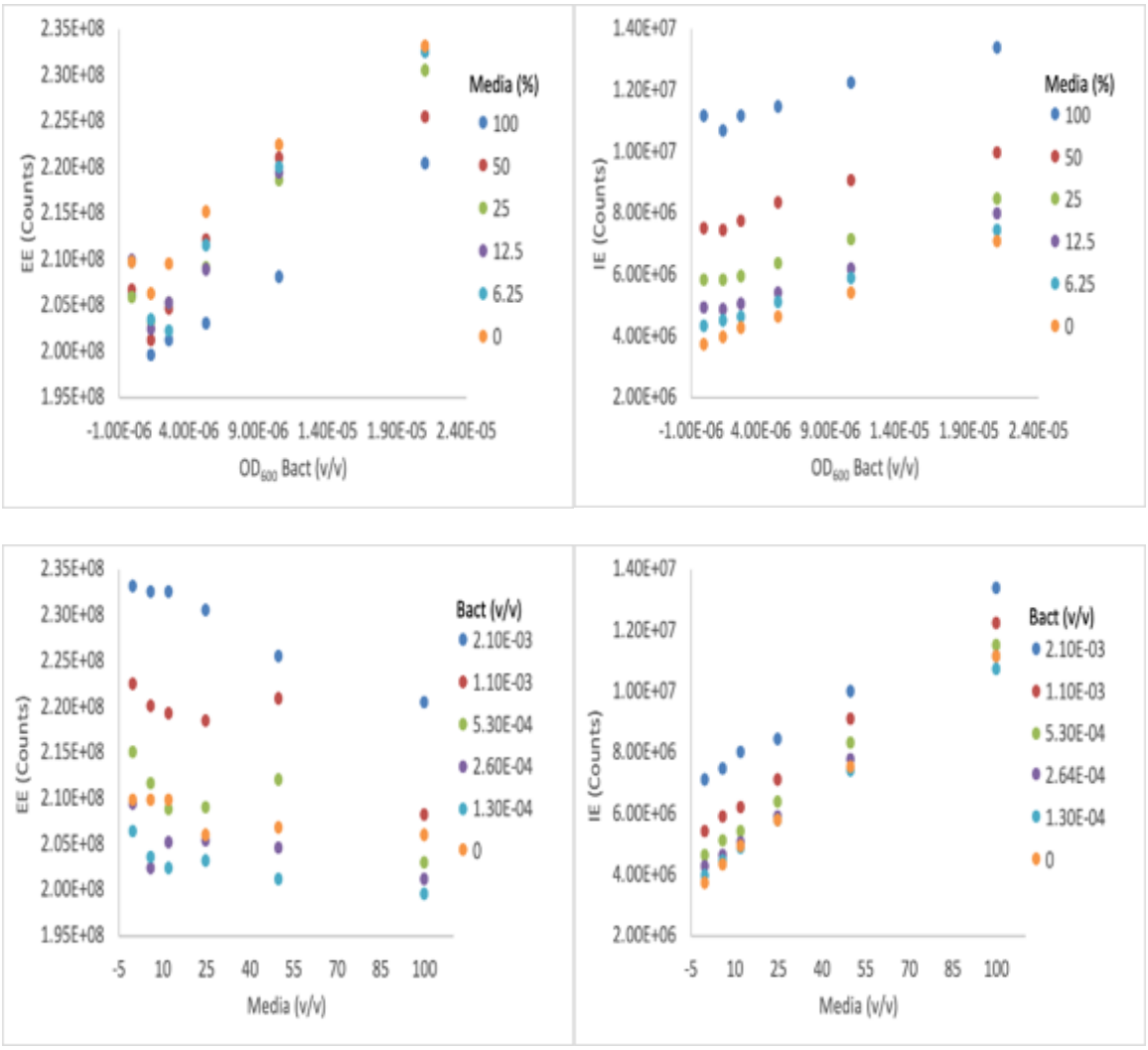

Figure 8: The EE and IE signals for *P. aeruginosa* PA01 as the concentrations of bacteria (top left and right) and media (bottom left and right) change. Note that the EE increases when there are no bacteria in the cuvette because the incoming probing laser light penetrates to the bottom of the cuvette.

A bilinear regression fit was applied to the training set by first converting the OD_{600} for each of the solutions to volume fraction of bacteria this was called the OD_{600} volume fraction of bacteria (OD_{600} Bact) using equation [7]. Origin was used to plot the actual volume fraction of bacteria as the dependent variable and both the corresponding EE and IE as independent variables in

accordance with equation [5]. From the fitting option in the analysis menu of the Origin program, multiple regressions fit was selected, and the program returned a bilinear fit and r^2 value. This process was repeated for the actual volume fraction of media diluted with PBS and the same EE and IE signals used for the bacteria bilinear fit. Below is a table showing the results of the bilinear fit for PA find the coefficients *a* through *f*.

	Intercept (*a* and *d*)	Uncertainty Intercept	EE (*b* and *e*)	Uncertainty EE	IE (*c* and *f*)	Uncertainty IE
ϕ_{bac} for PA	-1.476E-4	8.966E-6	7.029E-13	4.248E-14	8.271E-13	1.608E-13
ϕ_{med} for PA	2.328	0.135	-1.374E-8	6.404E-10	1.285E-7	2.245E-9

Table 2: The coefficients (*a* through *f*) for the BSN algorithm equation of PA.

These values were called the “BSN Bact” and “BSN Med”. The BSN Bact was plotted against the actual values from the OD_{600} to produce the graphs in Figure 9 (left). However, the OD_{600} does not give information on the media concentration so the BSN Med is compared to the dilution factor used in the training set to produce Figure 9 (right).

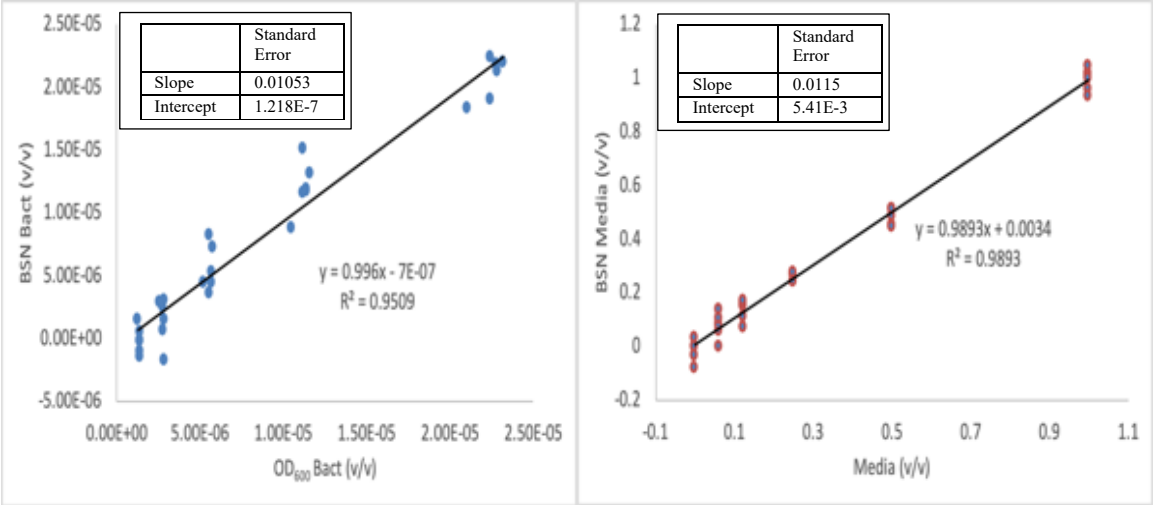

Figure 9: A plot of the BSN volume fraction of bacteria versus the OD_{600} volume fraction of bacteria for PA (left). The BSN volume fraction of media versus the media composition by dilution factor for PA (right).

The Kahan research group, Department of Chemistry, Syracuse University, supplied a frozen stock of *S. oneidensis* MR-1. LB media was prepared with 10g/L Tryptone, 10g/L Yeast extract and 5g/L NaCl in DI water. Lactate media was made using 2.5g/L of $NaH_2PO_4 \bullet H_2O$, 4.5g/L of Na_2HPO_4, 500mg/L of NH_4Cl, 190mg/L of KCl, 70mg/L of $CaCl_2$, 10mL/L of Wolfe's trace mineral and 100µL/L of lactic acid. Then the media were autoclaved at 120°C for 20 minutes. Phosphate Buffered Saline (PBS) was made using crushed tablets and dissolved in DI water, 1g/100mL for 1xPBS solution. Samples of bacteria were grown in LB media overnight at 35°C then diluted with media/PBS solutions. The average cell size of *SO* was taken as $1.8\mu m$ in length (l) and $0.45\mu m$ wide (w) which allows us to calculate the volume per organism.[33,34] The training set of SO in LB and lactate media was carried out in the same manner as the PA training set to produce the same graphs in figure 10.

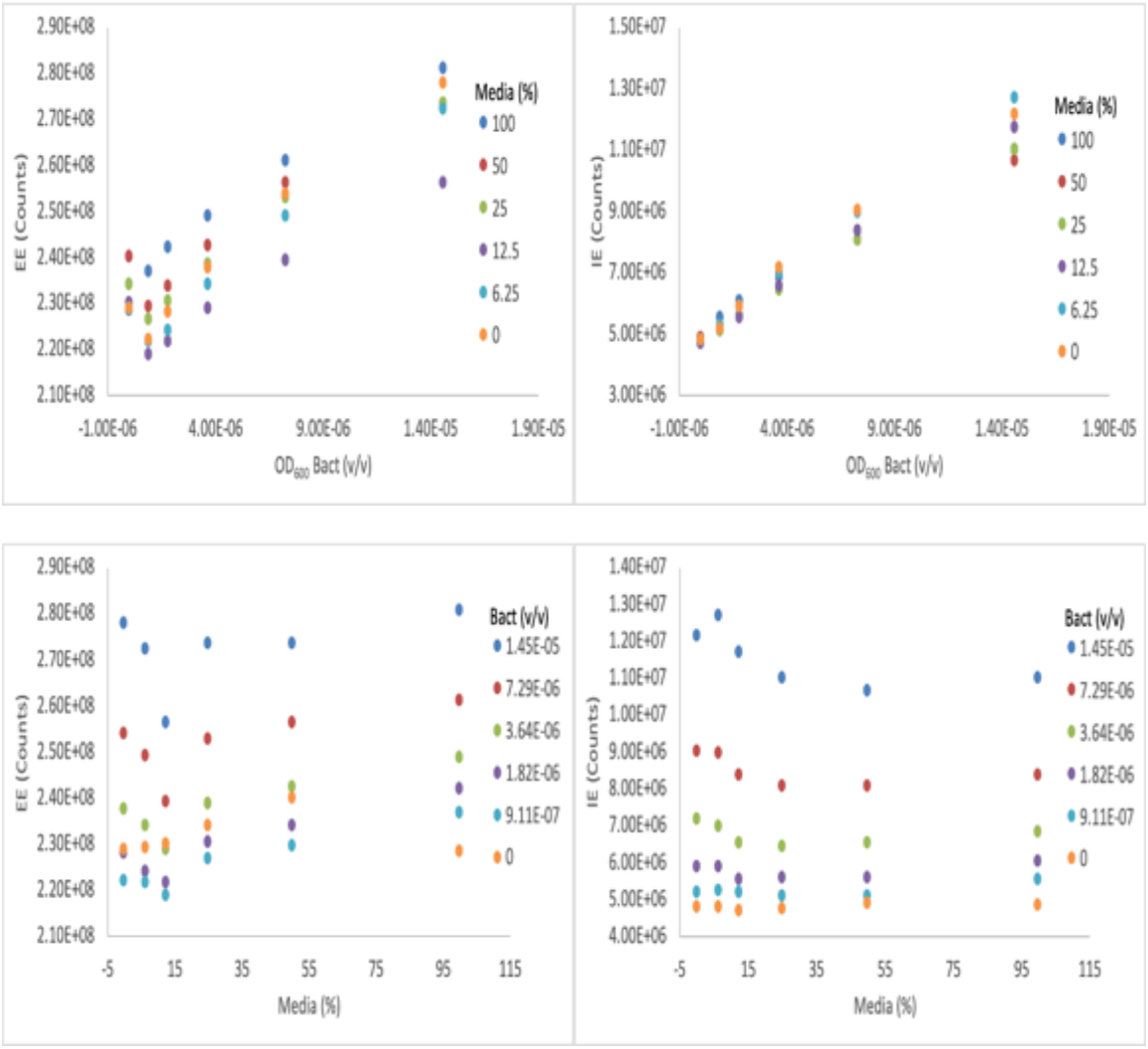

Figure 10: The relationship of EE and IE for *S. oneidensis* MR1 as the concentrations of bacteria (top left and right) and media (bottom left and right) increase.

As with PA, a bilinear fit was used on the SO training set to find the coefficients of the BSN algorithm for this bacterium. Table 3 shows the results of that bilinear fit. And just like with PA these values were used to create the equation to determine the volume fraction of bacteria ("BSN Bact") and media ("BSN Med"). These values were compared to the actual volume fractions in Figure 11.

	Intercept (a and d)	Uncertainty Intercept	EE (b and e)	Uncertainty EE	IE (c and f)	Uncertainty IE
ϕ_{bac} for SO	-1.449E-5	3.483E-6	2.318E-14	1.776E-14	1.915E-12	1.303E-13
ϕ_{med} for SO	-1.175	0.246	2.471E-9	1.257E-9	1.277E-7	9.225E-9

Table 3: The coefficients (a through f) for the BSN algorithm equation of SO.

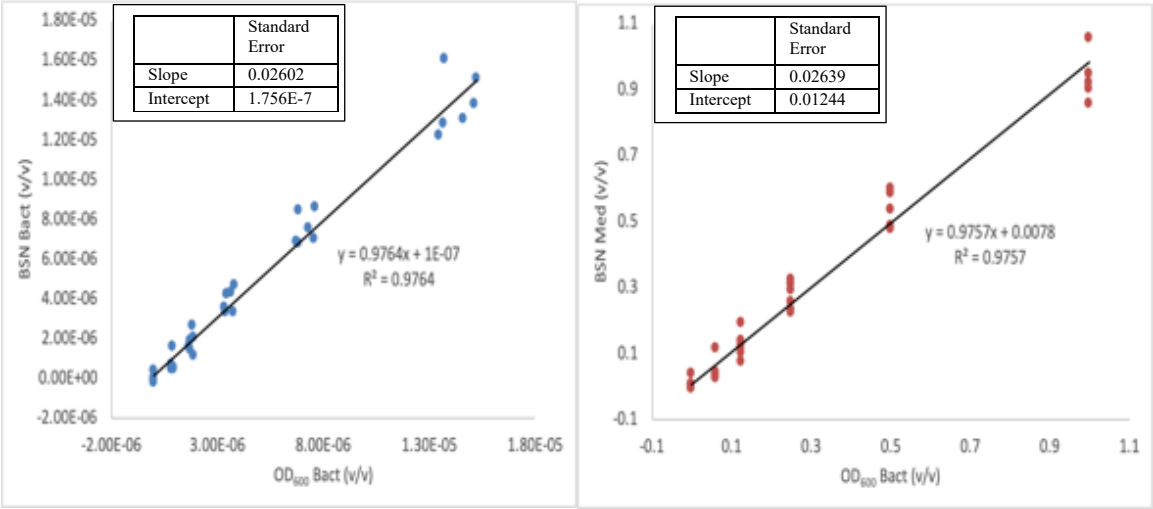

Figure 11: The BSN volume fraction of bacteria versus the OD_{600} volume fraction of bacteria for SO (left). The BSN volume fraction of media versus the media composition by dilution factor for SO (right).

2.4. Bacterial Growth and Raman Spectra

Next, we applied the calibrated BSN algorithm obtained from the training set to the PA growth to find ϕ_{bac} and ϕ_{med} respectively from the EE and IE signals obtained from the growth curves. Then the results were plotted versus time and a growth curve was created using the UV-Vis data so that they can be compared. Figures 12 and 13 display the growth/decay curves and the comparison of calculated versus actual bacteria volume fractions for twelve hours of PA growth.

The following procedure and conditions were used to obtain the BSN parameters a-f in equations 5 and 6, allowing the BSN algorithm to be applied to a realistic situation that occurs in laboratories and commercial bioreactors i.e., growth curves. Two PA growth curves were produced, one for six hours and the other for twelve hours. Overnight cultures of PA were grown in LB media under the same conditions for the training sets and then 20µL aliquot was diluted into 3.48mL of LB media in the quartz cuvette. The cuvette was thermostat controlled to 37°C. A blank solution (LB media without PA) was exposed to growth conditions and used as a control experiment. The wavenumber limits defining the EE and IE from the raw CCD data for the calibration curve and training set were the same as were used to analyze the data from the growth curves and the control.

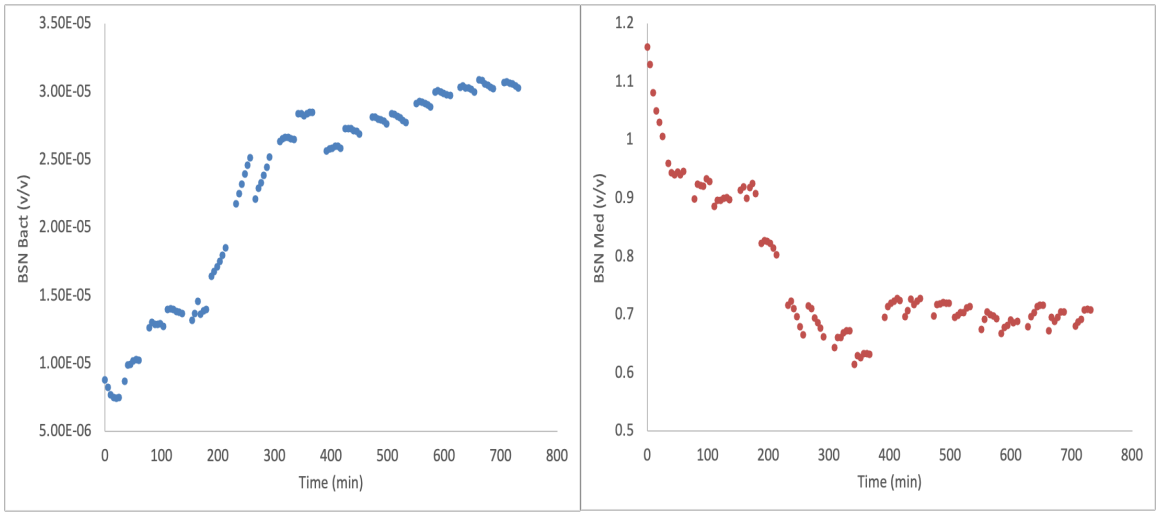

Figure 12: Volume fraction of bacteria from the BSN algorithm versus time for twelve-hour growth (left) and volume fraction of media from the BSN algorithm versus time for six-hour growth curve (right).

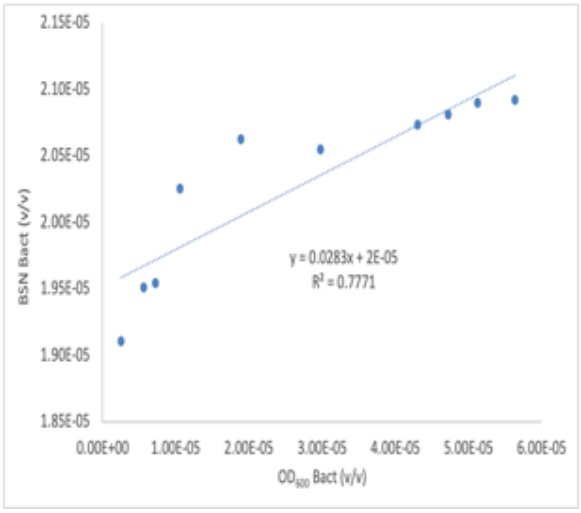

Figure 13: A graph displaying the BSN ϕ_{bac} versus OD_{600} ϕ_{bac} for the twelve-hour curve. Note that the limit of linearity of the single scattering range is 2.00 x 10^{-5} v/v.

The following figures show the BSN volume fractions of bacteria and media for PA growth with and without exposure to UV light (figure 14) and the LB media without PA (blank solution) under bacteria growth conditions with and without exposure to UV light (figure 15).

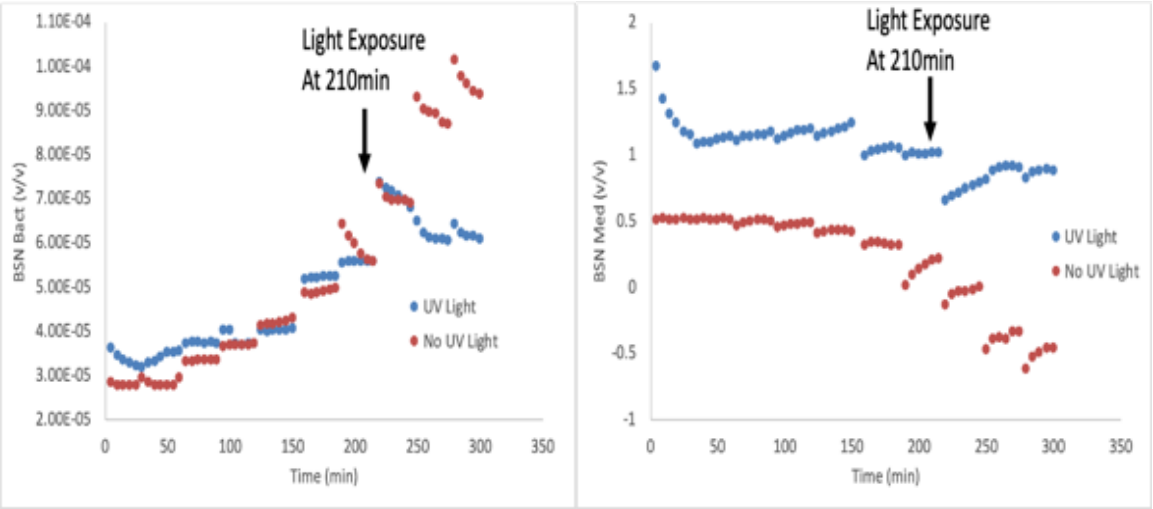

Figure 14: Comparison of BSN volume fractions of bacteria and media for PA growth without UV light exposure (red) and UV light exposure at 210min (blue).

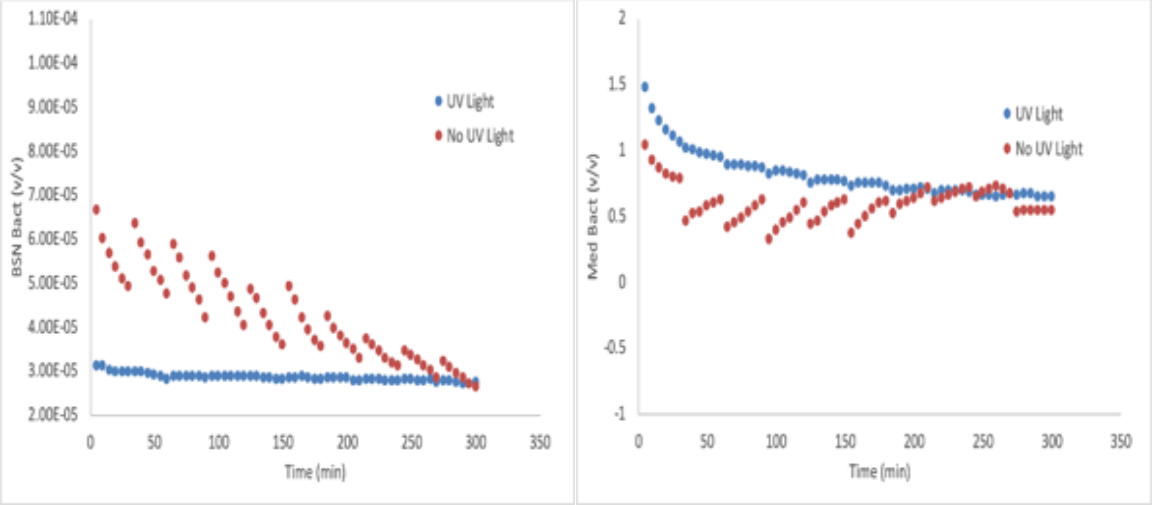

Figure 15: Comparison of BSN volume fractions of bacteria and media for LB media without bacteria under growth conditions without UV light exposure (red) and with UV light exposure at 150min (blue). Note that BSN Bact for zero bacteria (left) is dominated by reflection of incoming light off the bottom of the cuvette and top surface of medium.

The growth curve experiments were repeated, and the solution was exposed for 5 min to UV light from a lamp at 210 min of the experiment. Another experiment was performed with 25µL of bleach added to the solution instead of LB media at 120min, 150min, 180min and 210min. Each of these experiments was repeated with a blank solution as a control experiment.

The following figures show the comparison of the volume fractions of bacteria and media for PA growth with and without bleach added (figure 16) and the blank solution under growth conditions with and without bleach added (figure 17).

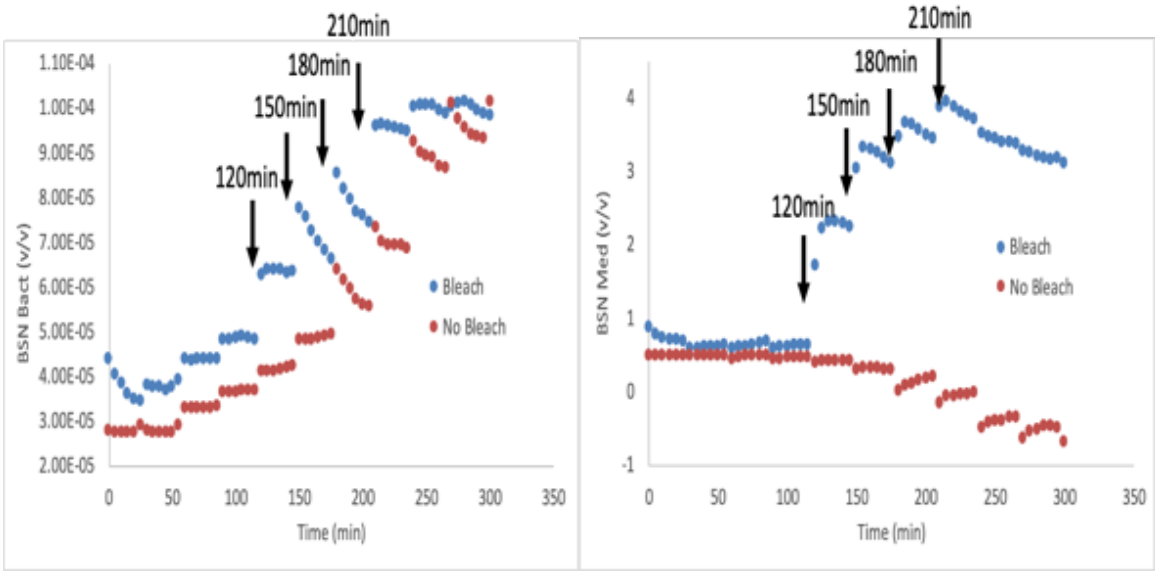

Figure 16: Comparison of BSN volume fractions of bacteria and media for PA growth without bleach (red) and with bleach added at 120min, 150min, 180min and 210min (blue). Note that bacteria level is slightly different at the outset for the two different runs.

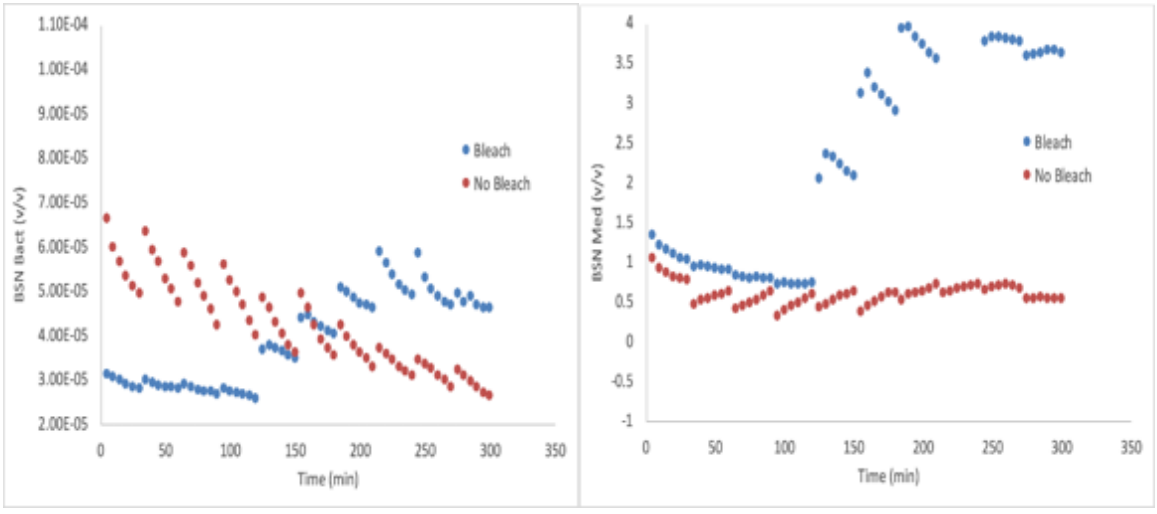

Figure 17: Comparison of BSN volume fractions of bacteria and media for LB media without bacteria under growth conditions without bleach (red) and with bleach added at 120min, 150min, 180min and 210min (blue).

Raman spectra can be used to keep account of the chemical species in the bacterial culture by analyzing the feature that change as the bacteria grow and the feedstock is consumed. However,

to do so Raman features must be identified with their chemical component *in situ*. To do this the literature Raman spectra of some components that are known to be in the bacteria culture were studied as their concentration changed to determine which if any Raman features correlate. One component that is certain to be present in the culture are the bacteria and the phospholipid bilayer of the outer membrane will be the part of the cell that contributes to the spectra. Figure 18 shows the spectra of DL-dipalmitoylphosphatidylcholine at various temperatures as an analogue for the biomass contributed by bacteria cells.

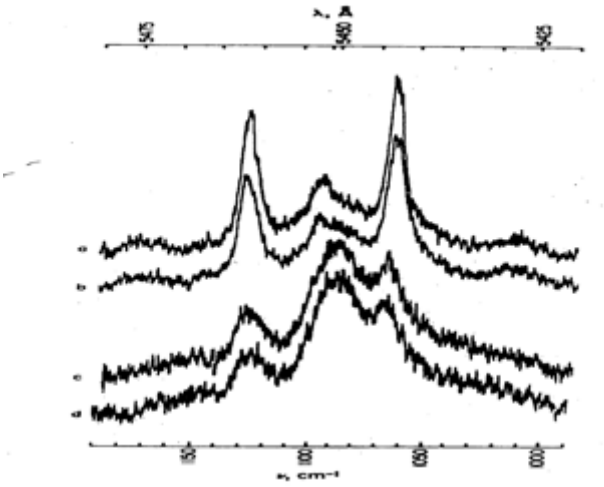

Figure 18: Raman spectrum of 20% DL-dipalmitoylphosphatidylcholinesonicates in water showing the peak at about 1065cm^{-1}. This scan was taken at various temperatures (a) 20°C (b) 30°C (c) 40°C (d) 50°C.[35]

The spectrum of DL-dipalmitoylphosphatidylcholine shows a noticeable peak at about 1065cm^{-1} especially at 40°C which is near the temperature that the bacteria were incubated during the growth curves. Another chemical contributor to Raman spectra that is known to be present is glucose. As such a calibration curve of glucose was created to determine what Raman feature corresponds with the varying concentration. Figure 21 shows the Raman spectra of D-glucose and the calibration curve that was constructed by comparing the integration of the 1120cm^{-1} peak and the actual concentration.

All biological Raman spectra with NIR excitation contain a broad autofluorescence with Raman emission superimposed and presently it is not known how to separate the autofluorescence from the Raman emission deterministically. Therefore, a baseline correction called a “101-7” was performed on the cropped raw IE so that Raman spectral features can be more easily observed for comparison. This was done by first cropping the spectrum of every 20-minute series of scans for each of the wavenumbers from 600cm^{-1} to 1900cm^{-1}, this corresponds to the Raman/fluorescence and ensures a reasonable separation from the Rayleigh/Mei region of the spectrum. Then the data with their corresponding wavenumbers were transferred to an Origin worksheet.

The 101-7 baseline subtraction involves first smoothing the experimental spectrum using simple 101-point adjacent averaging. The smoothed spectrum resembles the underlying fluorescence spectrum and is then subtracted from the experimental spectrum to leave the narrower Raman features. The difference i.e., the Raman spectrum is then smoothed using a 7-point adjacent averaging window to reduce high frequency noise thereby producing a “101-7 baseline corrected spectrum”. Figure 19 (right) shows the baseline corrected Raman spectra displaying all possible chemical peaks more clearly than the raw spectra shown in figure 19 (left). Operational details on the 101-7 baseline correction can be found in Appendix A although we note here that the procedure produces baseline corrected Raman spectra that has zero mean and standard deviation of 1 i.e., standard normal variant (SNV) transformation without assumption or bias. SNV transformation is the first step in any chemometric analysis.

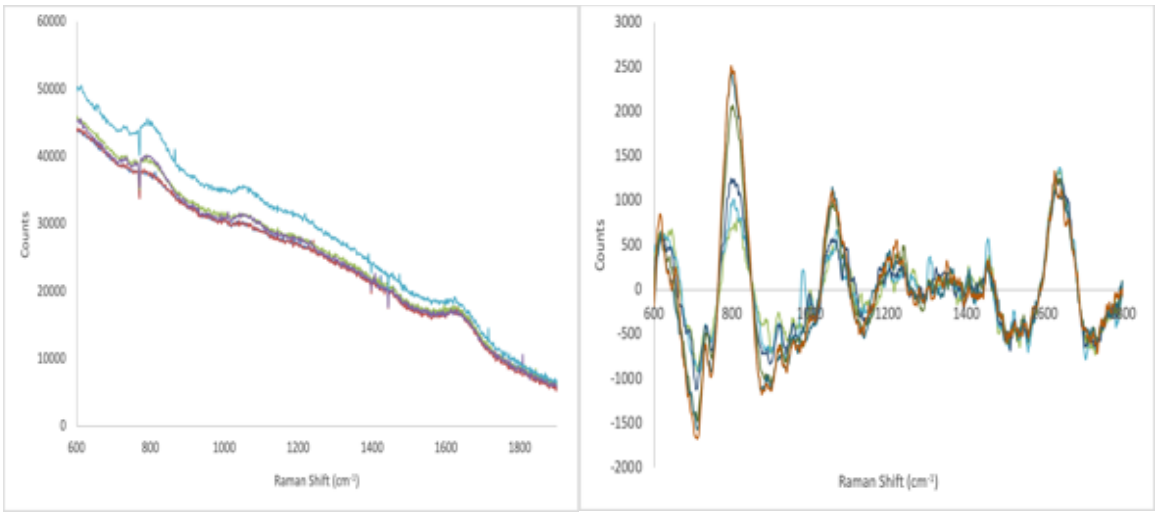

Figure 19: An example of raw Raman spectra (left) and the 101-7 baseline corrected spectra of a six-hour PA growth curve (right).

We performed some crude integration of peaks that were obviously varying in a systematic manner. The 1065cm^{-1} peak was integrated using the baseline corrected spectra at six-time intervals (60, 120, 180, 240, 300 and 330min). Then the graph was cropped to show only the region from 950cm^{-1} to 1150cm^{-1} and a line was drawn from 1014cm^{-1} to 1075cm^{-1} (figure 20 black line) this range was chosen to include as much of the peak without including the possible peak at 1120cm^{-1}. The counts corresponding to the horizontal line were subtracted from the signals along that line and the results were summed up to find the area of the peak. This procedure was repeated for the peak at 1120cm^{-1} for the region from 1097cm^{-1} to 1148cm^{-1} (figure 20 red line). Figure 21 left is the spectra of D-Glucose at varying concentrations in DI water. Figure 21 right displays the change in areas for the 1120cm^{-1} peak as the concentration of glucose in DI water increases. This procedure has been shown[36] to give good correlation with calibration curves for solutions containing known amounts of analytes e.g., glucose in PBS.

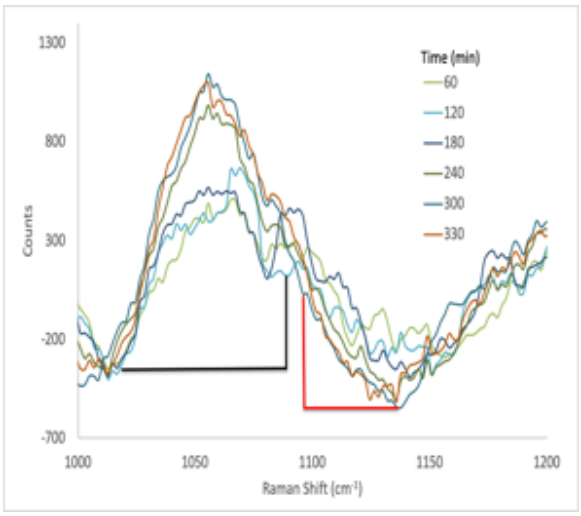

Figure 20: The baseline corrected spectra showing the range of the peak integration at 1065cm^{-1} (black) and 1120cm^{-1} (red).

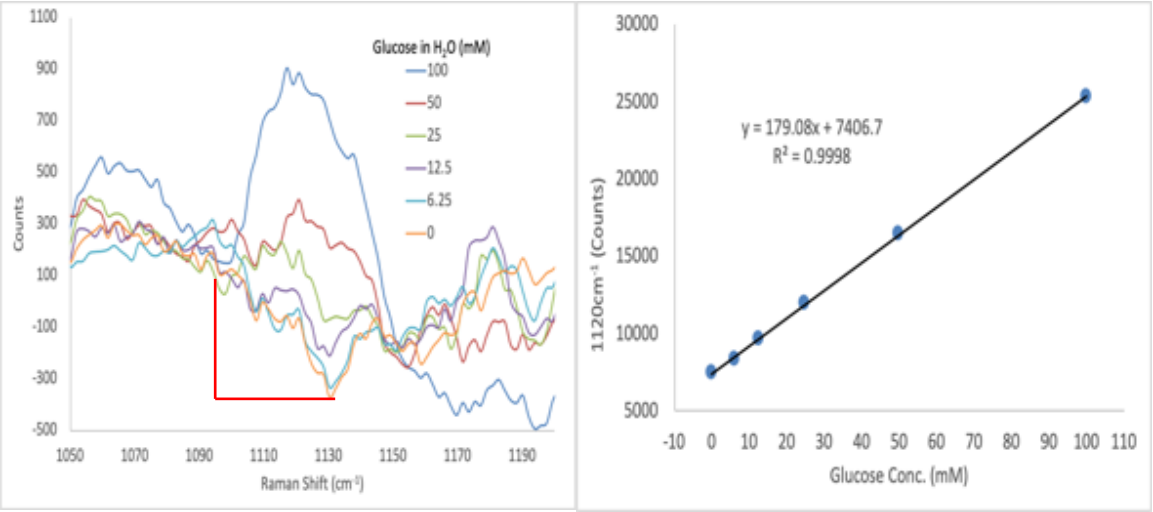

Figure 21: The spectrum of glucose in DI water showing the peak at about 1120cm^{-1} and the region of integration in the red boundary (left). The calibration curve of the integrated 1120cm^{-1} peak for glucose in DI water (right).

Integration of the 800cm^{-1}, 1065cm^{-1} and 1120cm^{-1} peaks were calculated from the six-hour growth curve of PA to produce figure 22. And the integrated 1120cm^{-1} peak from both the six- and twelve-hour growth curves of PA to calculate the concentration of glucose using the calibration curve of D-glucose (figure 21 right) to create figure 23.

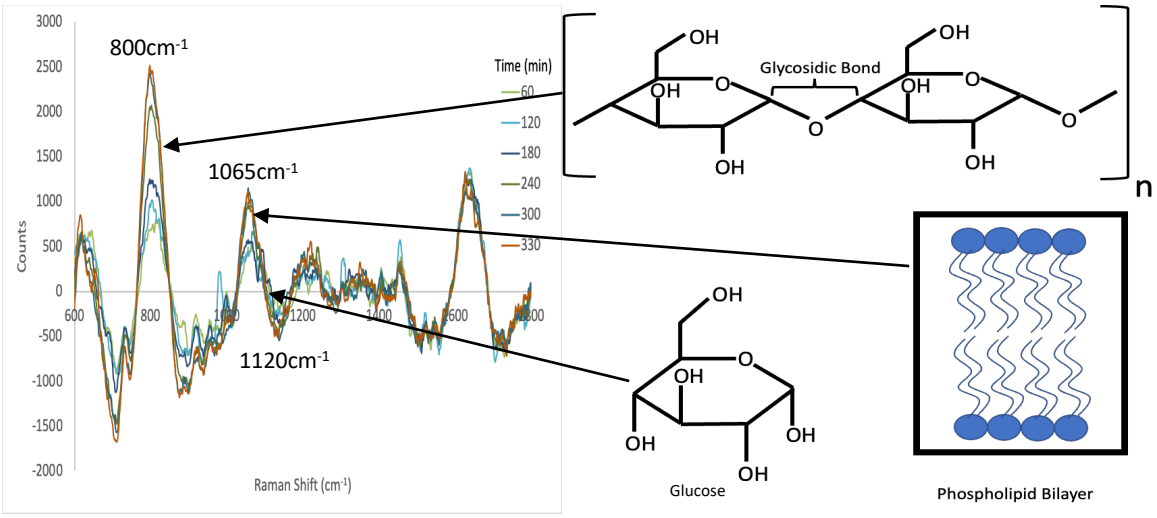

Figure 22: The 800cm^{-1}, 1065cm^{-1} and 1120cm^{-1} peaks of the PA growth Raman spectra and their chemical contributors.

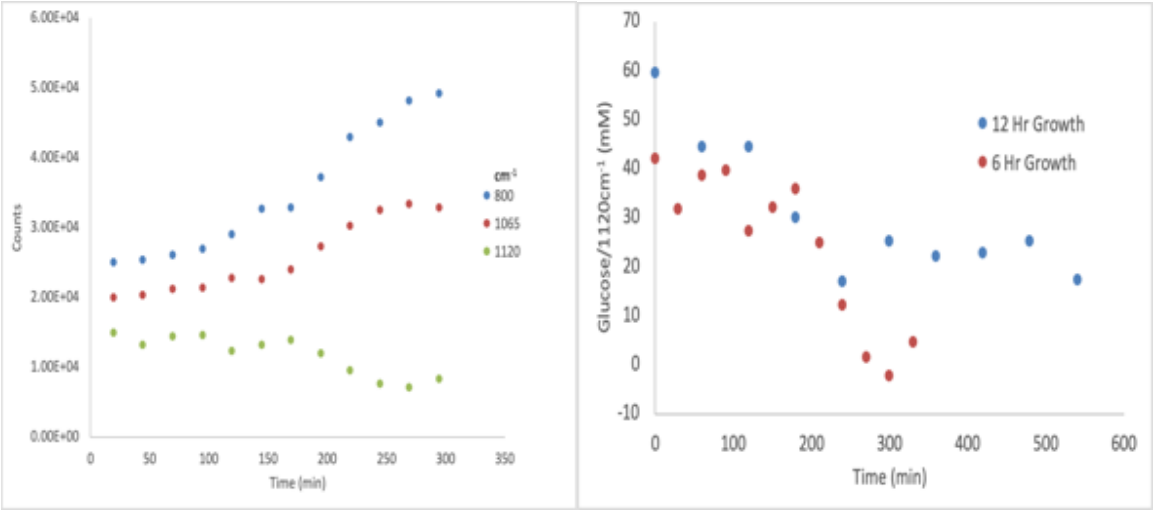

Figure 23: The change of Raman peaks at 800cm^{-1}, 1065cm^{-1} and 1120cm^{-1} for the PA twelve-hour growth. Concentration of glucose derived from the calibration curve of 1120cm^{-1} peak of glucose in DI water for two independent PA growths for six and twelve hours.

2.5. Limitations with Minimal Media

The following graphs show the results of the BSN analysis when performed on SO cultures in lactate media. As figure 24 shows that unlike with SO in LB media, the lactate media training did not yield a linear trend for the EE as the concentration increased. Not only is there no linear trend in the same concentration range as the LB media training set but also the EE signal seems

to decrease as the concentration increases. Figure 25 shows the baseline corrected Raman spectra from SO in lactate media. And figure 26 shows the spectra of lactic acid and lactate media minus the Wolfe's trace mineral.

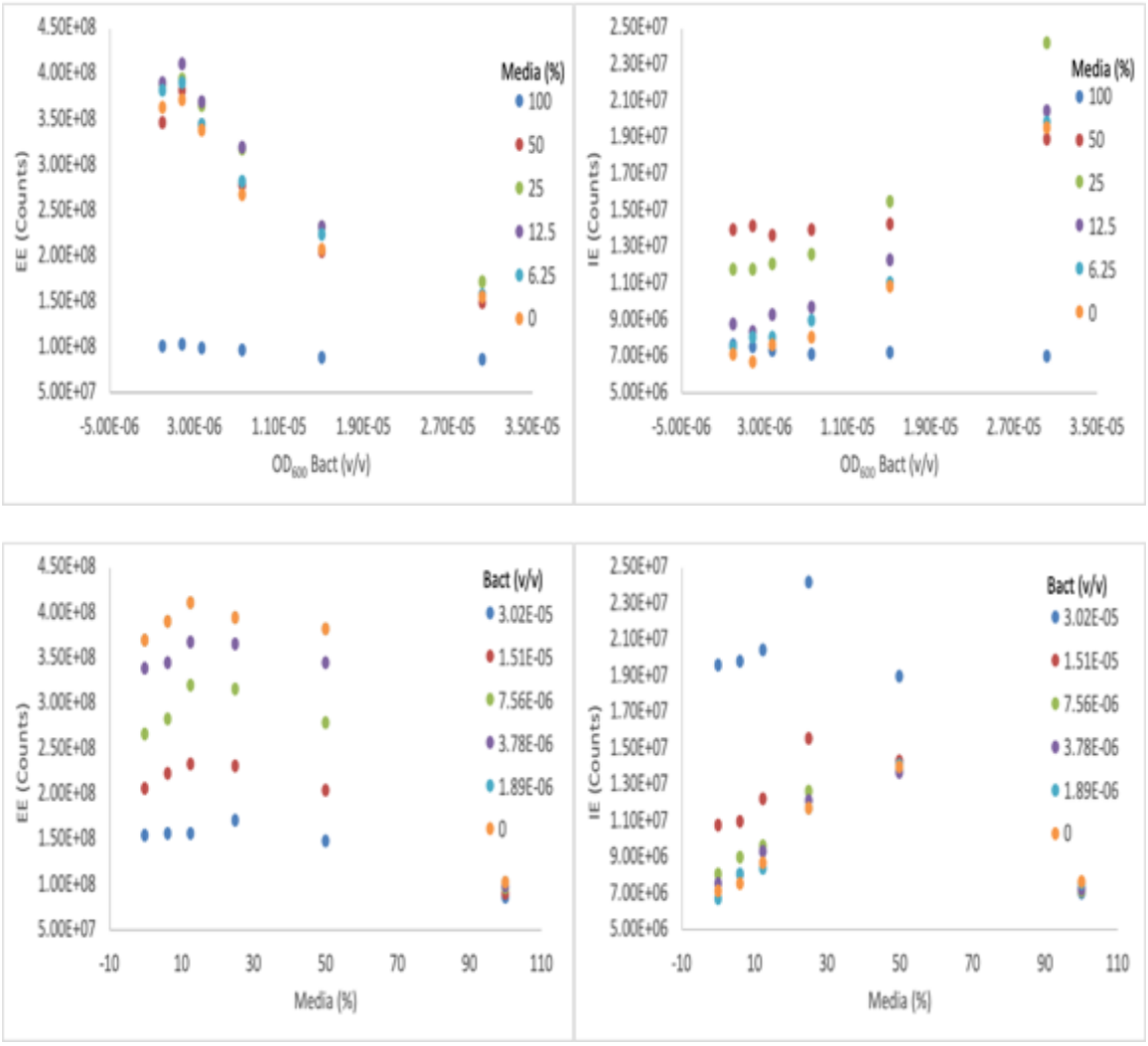

Figure 24: The training set of SO in lactate media, the EE (left) and IE (right) as bacteria (top) and media (bottom) concentrations increase. There is very little linear correlation and a loss of EE and IE signals at the highest concentration of media.

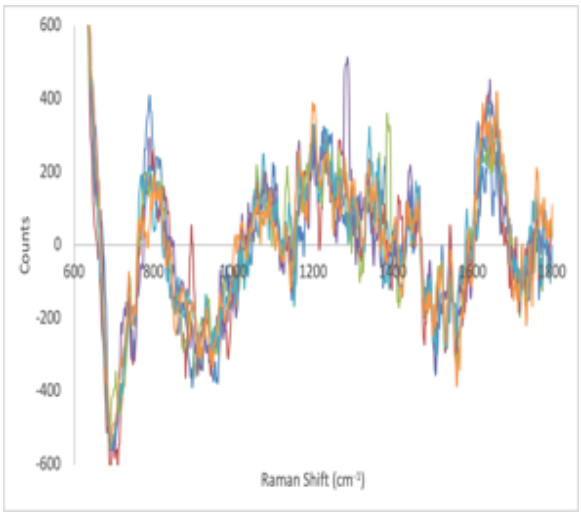

Figure 25: The 101-7 baseline corrected Raman spectra of SO in lactate media as the concentration of bacteria decreases.

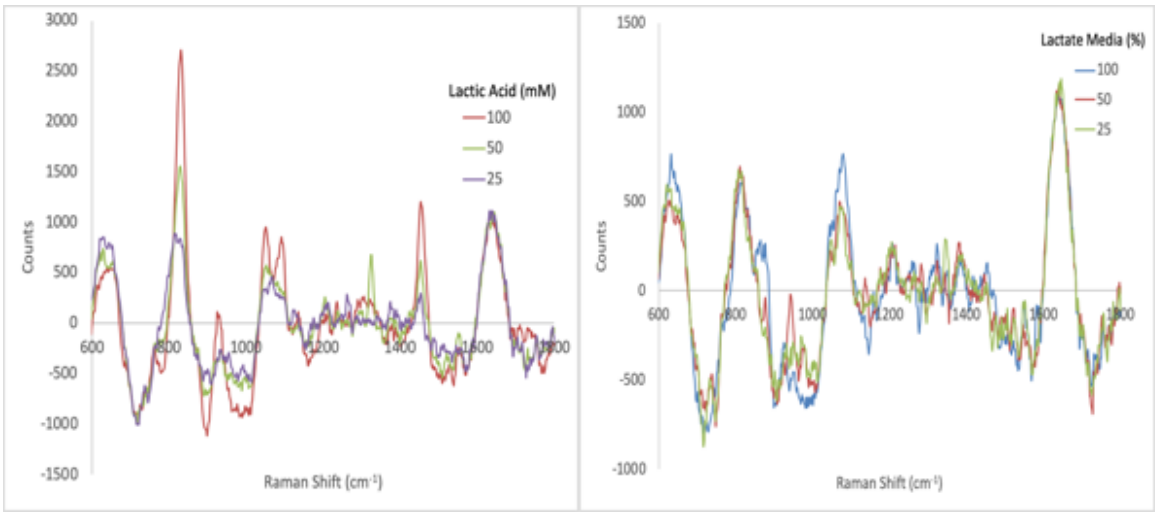

Figure 26: The 101-7 baseline corrected Raman Spectra of lactic acid in DI water (left) and lactate media minus Wolfe's Trace Mineral (right) at various concentrations.

Since the Raman spectra for both the lactic acid and the lactate media without the Wolfe's trace mineral showed less noise when compared to the spectra from the training set. It was believed that there is some substance in the Wolfe's trace mineral that is interfering with the Raman signal production and/or detection. The UV-Vis and Raman spectra of 100mM solutions of metal ions

found in the trace mineral were used to create figure 27.

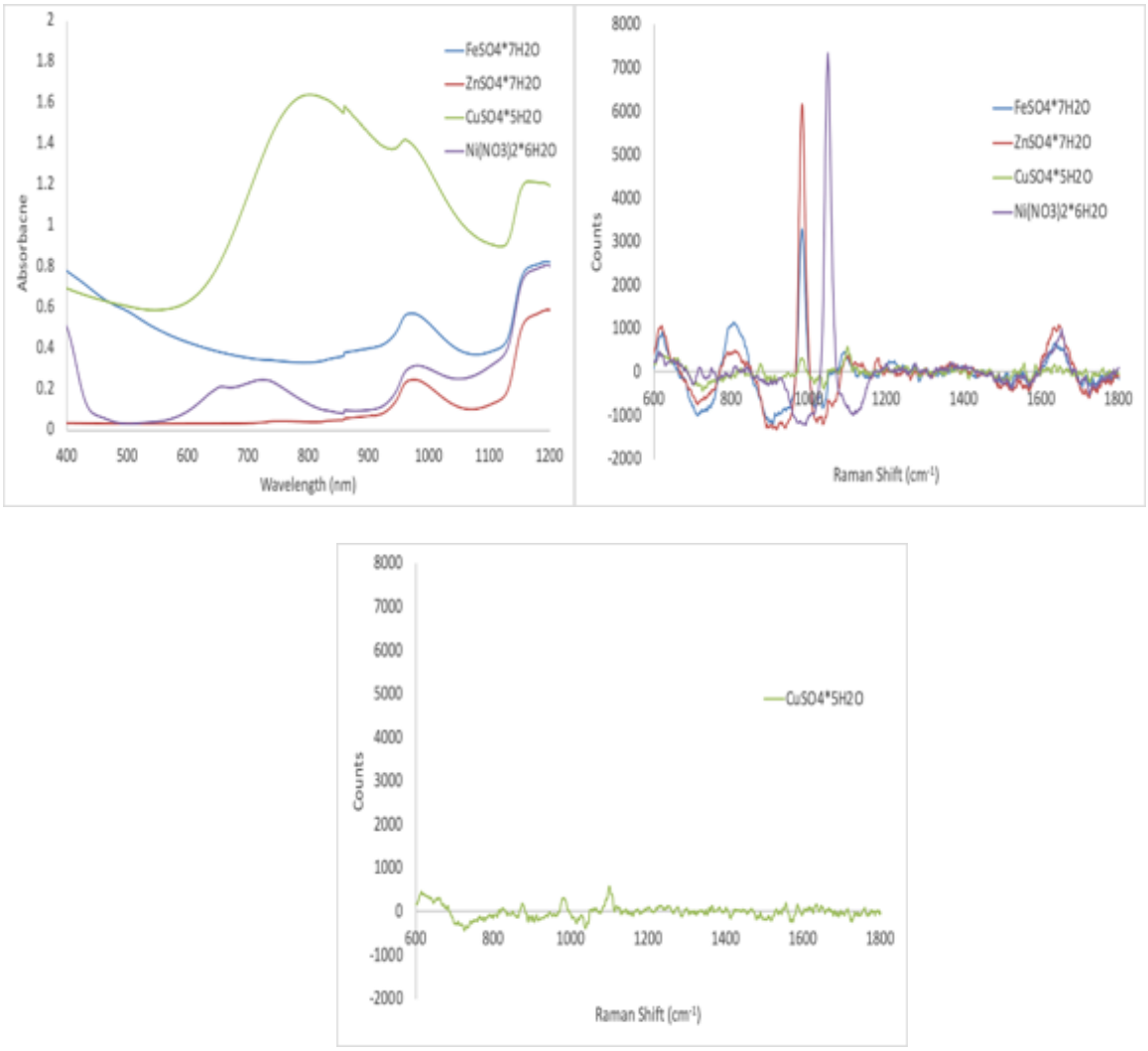

Figure 27: The UV-Vis spectra (top left) and 101-7 baseline corrected Raman spectra (top right) of 100mM concentration of $FeSO_4 \cdot 7H_2O$, $ZnSO_4 \cdot 7H_2O$, $CuSO_4 \cdot 5H_2O$ and $Ni(NO_3)_2 \cdot 6H_2O$. The bottom graph shows the baseline corrected Raman spectrum of 100mM $CuSO_4 \cdot 5H_2O$.

3. Diffusion of Bacteria

3.1. Einstein-Smoluchowski Equation

Mass diffusion is a form of transport while the first attempt to model diffusion was formulated

by Adolf E. Fick a German physician.[37] Fick's equation follows a similar design as Fourier's for heat conduction, the equation for momentum transport by Newton and Ohm's law for electric current. For all of these, the flux is proportional to a driving force and in the case of diffusion the "force" is a concentration gradient with the proportionality constant being the diffusion coefficient (D). Brownian motion has been used as a model for diffusion in a uniform media and the Einstein-Smoluchowski equation (eq 12) is used to simulate the data, which relates the mean square displacement in a three-dimensional space from an origin location at t = 0, to the diffusion coefficient and time (τ).

$$x^2 = 4Dt \qquad [12]$$

Where x is the mean square displacement from its initial position travels e.g., on a grid through the medium for time t. A factor of 4 as shown in [12] corresponds to diffusion i.e., a random walk on a 2-dimensional grid whereas for three dimensions we have a random walk in a 3-dimensional volumetric grid and a factor of 6. This assumes that the volume the particles traverse is isotropic but since the cuvette has one length much longer than the other two, and the average density induced concentration gradient is in one dimension. On average there are no gradients in the other two spatial dimensions, the bacteria are otherwise unrestricted. We know that the dimensions of the bacteria are very small with respect to the dimensions of the cuvette, but the culture clearly extends to the cuvette walls. To the extent the bacteria behave socially, or even independently, their motion may or may not be random We hypothesize that a correct value for the dimensionality will be $1 < D < 3$ and can be comparable to other relevant measurements.

3.2. Propagation of Bacteria Through the Media

The case of bacteria that is analogous to diffusion of colloidal particles in that the particle i.e.,

the bacteria/colloidal particle is much larger than the molecules of the solvent this a condition for the observation of Brownian motion in a Newtonian fluid. Early experiments on Diffusion of colloidal particles (which are much larger than solvent molecules) using dynamic light scattering were performed on polystyrene spheres.[38] Research has also been conducted using *Salmonella typhimurium* that were placed between two glass slides spaced by parafilm and observed by differential dynamic microscopy as opposed to dynamic light scattering.[39] Another experiment was performed on E. coli that were fluorescently tagged and then their image taken by an inverted fluorescent microscope.[40] However, the bacteria in these experiments were conducted in room temperature and no stirring/shaking; their motility through the medium was by flagella in addition to possible diffusion. This experiment involves a larger volume, and the driving force of the diffusion is the increase in bacteria concentration. For our BSN experiments the diffusion of bacteria is detected by the increase of EE signal as the cells traverse the plane that intersects with the laser focal point and accumulate in the region above that plane. This is indicated by the increase in color intensity of the plane in figure 28.

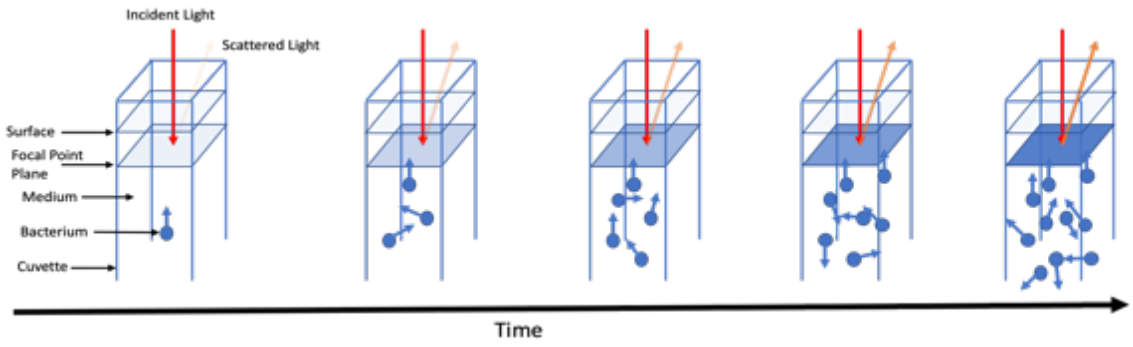

Figure 28: A diagram that shows the relationship between the intensity of scattered light, i.e., EE and IE signals, and the diffusion of bacteria cells.

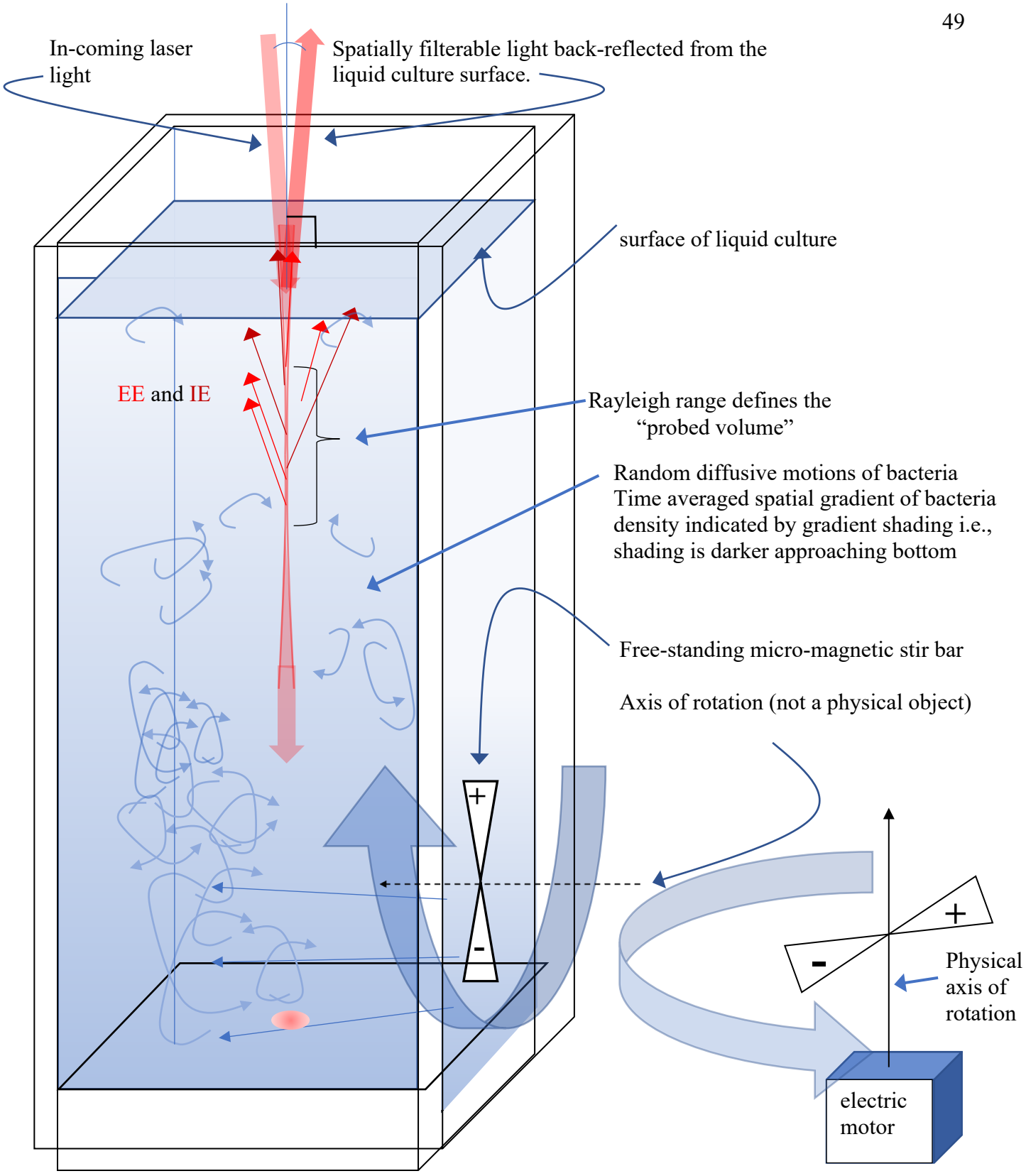

Figure 29: A diagram depicting the bacteria culture in the cuvette, the laser light path, scattered light, magnetic micro stir bar and stirrer.

3.3. Calculation of Diffusion Coefficient

The following figures (30-33) show the growth of PA by time in minutes (left) and the square root of time (right) as well as their linear trendline. The arrows show the point where the bacteria volume fraction suddenly increases as the point where the cells first enter the plane that intersects the laser focal point.

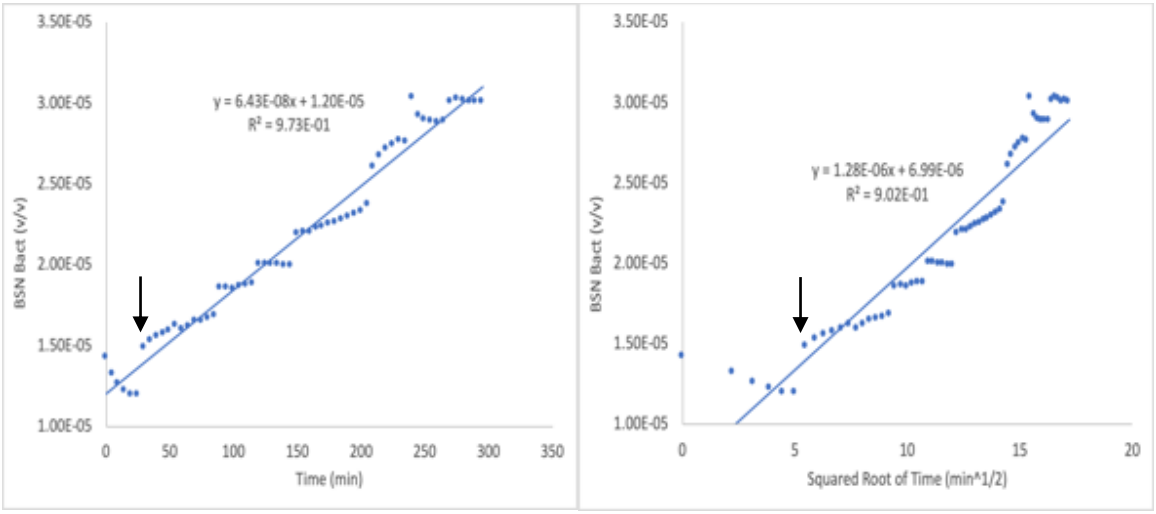

Figure 30: The six-hour growth curve (left). The same growth curve plotted with the square root of time which shows the relationship to the diffusion of bacteria through the media.

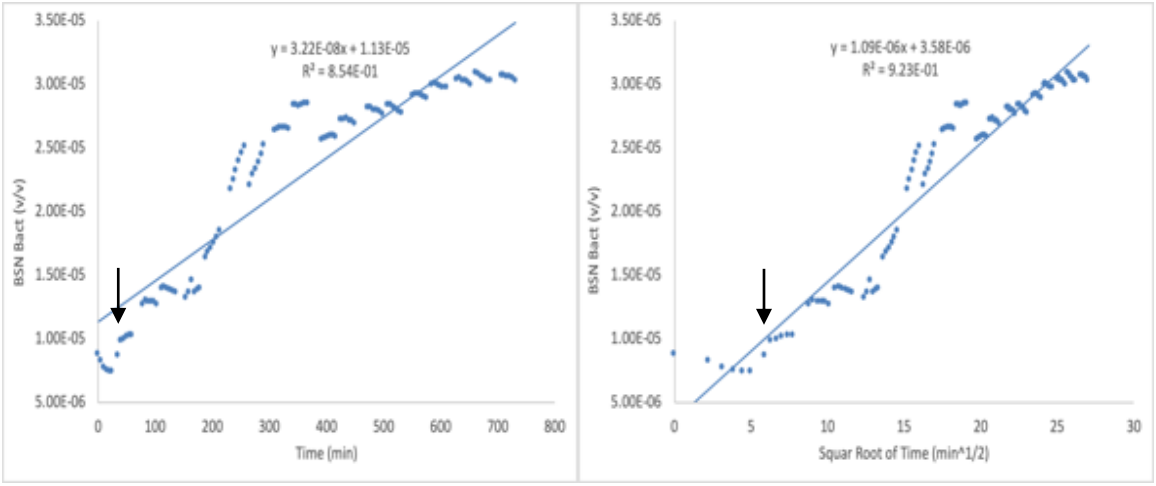

Figure 31: A twelve-hour growth curve (left) and the same curve plotted against the square root of the time (right).

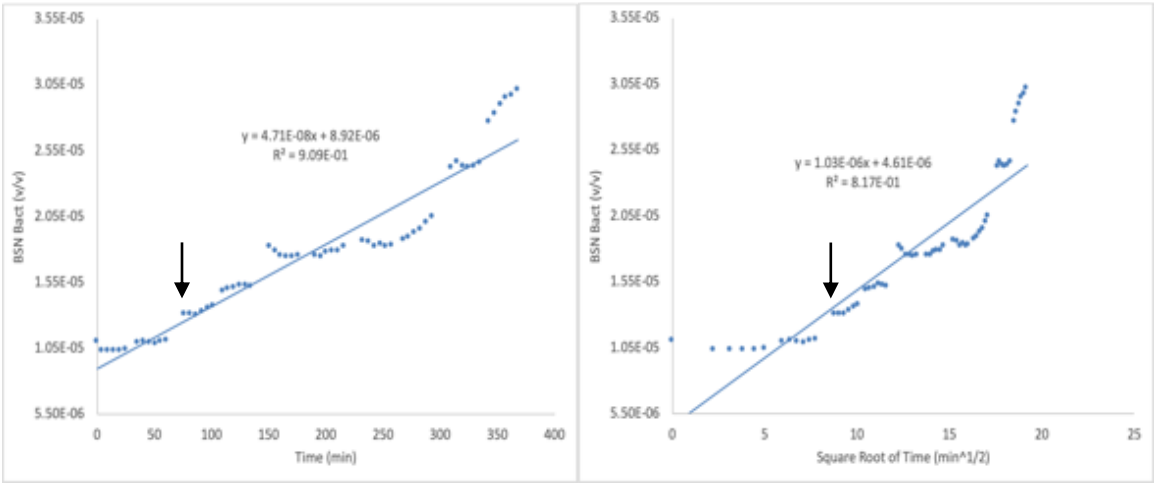

Figure 32: A six-hour growth curve (left). The same growth curve plotted against the square root of time (right).

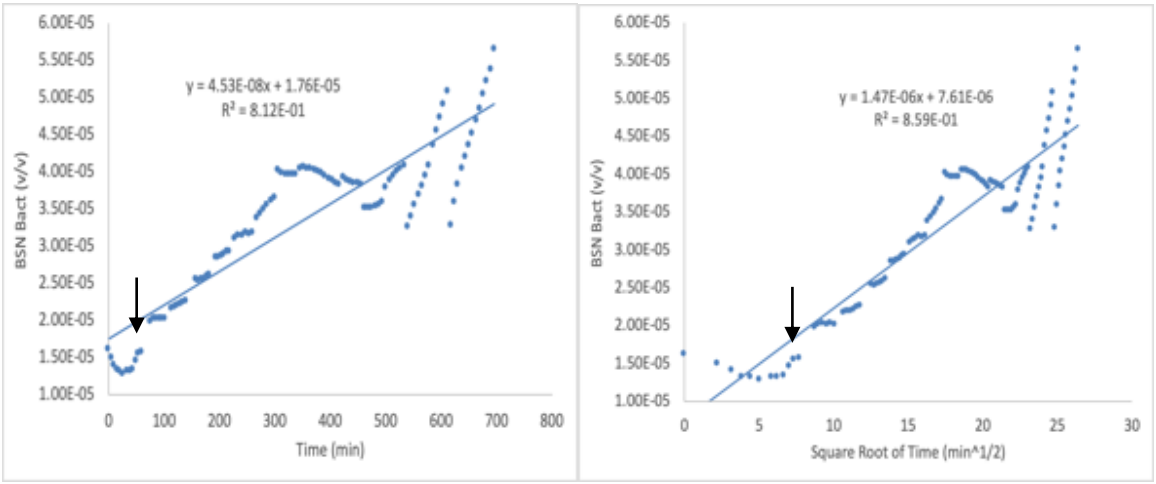

Figure 33: A twelve-hour growth curve (left). The same growth curve plotted against the square root of the time (right).

Table 4 shows the slope and calculated diffusion coefficient for each of the growth curves in figures 30-33. To calculate the diffusion coefficient for the growth curves the transport distance of the bacteria was taken as 1cm.

Time of bacteria detection (min)	Square root of bacteria detection time ($\sqrt{\text{min}}$)	D_2 for 1 dimension ($\frac{m^2}{s}$) $\times 10^{-8}$	D_4 for 2 dimensions ($\frac{m^2}{s}$) $\times 10^{-9}$	D_6 for 3 dimensions ($\frac{m^2}{s}$) $\times 10^{-9}$
30	5.48	1.39	3.47	1.54
35	5.91	1.19	2.98	1.32
39	6.24	1.07	2.67	1.19
77	8.77	0.541	1.35	0.601
Average Diffusion Coefficient		1.05 ± 0.362	2.62 ± 0.906	1.16 ± 0.403

Table 4: The growth curves of PA over various time lengths and square root of time, their slopes and diffusion coefficients.

4. Principal Component Analysis of BSN Raman Spectra

4.1. Theoretical Background of PCA

"Chemometrics" is a term[41] used to describe the general process of analyzing a complex stream of quantitative data and obtaining analytical information. To retrieve spectroscopic information for use in algorithm development we employ chemometric techniques to the raw resolved Raman spectra. We seek the Raman spectral features that contain the most information i.e., variance with respect to changing culture characteristics with respect to time, composition, bacterial number density and biological activity.

To assist with the identification of Raman features with the maximum variance as the different components change, principal component analysis (PCA) was performed on the Raman spectra.

PCA takes a dataset as a matrix $\boldsymbol{X}$ with rows m (objects) and columns n (variables), $\boldsymbol{X} = m \times n$, then converts M and N into new scaled matrices, $\boldsymbol{M}$ and $\boldsymbol{N}$ respectively. The new matrices transformed to a P-dimensional coordinate system with each axis a "principal component" of the dataset.

First, matrix $\boldsymbol{X}$ is converted to a Hermitian covariance matrix, then singular value decomposition (SVD) expresses it as a factor of three matrices ($\boldsymbol{LS^2L^T}$). The "loading" matrix ($\boldsymbol{L}$) consists of the eigenvectors of the covariance matrix and a diagonalized "latent" matrix ($\boldsymbol{S}$) of their corresponding eigenvalue.[42,43] These matrices are used to extrapolate information about the dataset, such as the "score" matrix ($\boldsymbol{T}$) which is the variance magnitude of the data. And the "variance explained" matrix ($\boldsymbol{VE}$) that shows the weight of each principal component contributes to the variance.[44] The following sections explain both PCA and SVD in detail.

4.2. Mathematical Explanation of PCA and SVD

Given a matrix of data points $\boldsymbol{X}$ with M rows and N columns, the covariance matrix is calculated by subtracting average of data points along each column ($\bar{x}_n$) from the corresponding column of $\boldsymbol{X}$ to form a new matrix $\boldsymbol{K}$. Then multiply the transposed new matrix ($\boldsymbol{K^T}$) with $\boldsymbol{K}$ to form the square matrix $\boldsymbol{A}$ with the variances of $\boldsymbol{X}$ in the diagonal and the covariances as the nondiagonal elements. Finally, $\boldsymbol{A}$ is divided by the number of rows minus one which forms the covariance matrix $\boldsymbol{C}$, making the covariance matrix Hermitian, that is, a square matrix equal to its transpose.[45,46]

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{bmatrix}$$

$$\boldsymbol{X} - \overline{\boldsymbol{X}} = \begin{bmatrix} x_{11} - x_1 & x_{12} - x_2 & x_{13} - x_3 \\ x_{21} - x_1 & x_{22} - x_2 & x_{23} - x_3 \\ x_{31} - x_1 & x_{32} - x_2 & x_{33} - x_3 \\ x_{41} - x_1 & x_{42} - x_2 & x_{43} - x_3 \\ x_{51} - x_1 & x_{52} - x_2 & x_{53} - x_3 \end{bmatrix} = \boldsymbol{K}$$

$$\boldsymbol{K}^T\boldsymbol{K} = \begin{bmatrix} k_{11} & k_{21} & k_{31} & k_{41} & k_{51} \\ k_{12} & k_{22} & k_{32} & k_{42} & k_{52} \\ k_{13} & k_{23} & k_{33} & k_{43} & k_{53} \end{bmatrix} \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \\ k_{41} & k_{42} & k_{43} \\ k_{51} & k_{52} & k_{53} \end{bmatrix} = \boldsymbol{A}$$

$$\boldsymbol{A} = \begin{bmatrix} \sum_{m=1}^{M} k_{m1}^2 & \sum_{m=1}^{M} k_{m1}k_{m2} & \sum_{m=1}^{M} k_{m1}k_{m3} \\ \sum_{m=1}^{M} k_{m2}k_{m1} & \sum_{m=1}^{M} k_{m2}^2 & \sum_{m=1}^{M} k_{m2}k_{m3} \\ \sum_{m=1}^{M} k_{m3}k_{m1} & \sum_{m=1}^{M} k_{m3}k_{m2} & \sum_{m=1}^{M} k_{m3}^2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

$$\boldsymbol{C} = \frac{\boldsymbol{A}}{M-1} = \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix}$$

4.2.1. Singular Value Decomposition

Singular value decomposition (SVD) is performed to find the loadings and latent matrices. SVD factorizes any matrix into three new matrices: $\boldsymbol{U}$, $\boldsymbol{V}$, and $\boldsymbol{S}$ with their relationship shown below using a general non-square matrix $\boldsymbol{X}$.

$$\boldsymbol{X}(m \times n) = \boldsymbol{U}(m \times m)\boldsymbol{S}(m \times n)\boldsymbol{V}^T(n \times n)$$

The matrices $\boldsymbol{U}$ and $\boldsymbol{V}$ are both orthogonal, i.e., $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. The following shows SVD performed on the covariance matrix $\boldsymbol{X}$ by first finding the product of its

transpose and itself to form matrix $\boldsymbol{W}$.

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{21} & x_{31} \dots x_{M1} \\ x_{12} & x_{22} & x_{32} \dots x_{M2} \\ x_{13} & x_{23} & x_{33} \dots x_{M3} \\ \vdots & \vdots & \vdots \dots \vdots \\ x_{1N} & x_{2N} & x_{3N} \dots x_{MN} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} \dots x_{1M} \\ x_{21} & x_{22} & x_{23} \dots x_{2M} \\ x_{31} & x_{32} & x_{33} \dots x_{3M} \\ \vdots & \vdots & \vdots \dots \vdots \\ x_{N1} & x_{N2} & x_{N3} \dots x_{NM} \end{bmatrix} = \boldsymbol{W}$$

$$\boldsymbol{W} = \begin{bmatrix} \sum_{m=1}^{M} x_{m1}^2 & \sum_{m=1}^{M} x_{m1}x_{m2} & \sum_{m=1}^{M} x_{m1}x_{m3} & \dots & \sum_{m=1}^{M} x_{m1}x_{mN} \\ \sum_{m=1}^{M} x_{m2}x_{m1} & \sum_{m=1}^{M} x_{m2}^2 & \sum_{m=1}^{M} x_{m2}x_{m3} & \dots & \sum_{m=1}^{M} x_{m2}x_{mN} \\ \sum_{m=1}^{M} x_{m3}x_{m1} & \sum_{m=1}^{M} x_{m3}x_{m2} & \sum_{m=1}^{M} x_{m3}^2 & \dots & \sum_{m=1}^{M} x_{m3}x_{mN} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \sum_{m=1}^{M} x_{mM}x_{m1} & \sum_{m=1}^{M} x_{mM}x_{m2} & \sum_{m=1}^{M} x_{mM}x_{m3} & \dots & \sum_{m=1}^{M} x_{MN}^2 \end{bmatrix}$$

The relationship between the matrix ($\boldsymbol{W}$), its eigenvalues (λ) and eigenvectors (h) and the identity matrix are expressed by the following equation $\boldsymbol{W}h = \lambda \boldsymbol{I}h$ algebraically rearranged as $(\boldsymbol{W} - \lambda \boldsymbol{I})h = 0$.

$$\boldsymbol{W} - \lambda \boldsymbol{I} = \begin{bmatrix} w_{11} - \lambda & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} - \lambda & \dots & w_{2N} \\ \vdots & \vdots & \dots & \vdots \\ w_{M1} & w_{M2} & \dots & w_{MN} - \lambda \end{bmatrix}$$

To determine the eigenvalues, use the determinant for the above equation and simplify into polynomial form then solve for the roots which are λ.

$$|\boldsymbol{W} - \lambda \boldsymbol{I}| = [(w_{11} - \lambda)(w_{22} - \lambda)(w_{33} - \lambda) \dots] + [w_{12}w_{23}w_{31} \dots] + [w_{13}w_{21}w_{32} \dots]$$
$$- [w_{13}(w_{22} - \lambda)w_{31} \dots] - [w_{12}w_{21}(w_{33} - \lambda) \dots] - [(w_{11} - \lambda)w_{23}w_{32} \dots] \dots$$
$$|\boldsymbol{W} - \lambda \boldsymbol{I}| = q_1\lambda^N + q_2\lambda^{N-1} + q_3\lambda^{N-2} + \dots + q_N\lambda^0$$

Where all q are constants. There are as many eigenvalues as there are columns of $\boldsymbol{W}$ and their square root is listed in descending order along the diagonal of the singular values matrix also

called the latent matrix $\boldsymbol{S}$ with zeroes as the other elements.

$$\boldsymbol{S} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & \dots 0 \\ 0 & \sqrt{\lambda} & 0 & \dots 0 \\ 0 & 0 & \sqrt{\lambda_3} & \dots 0 \\ \vdots & \vdots & \vdots & \dots 0 \\ 0 & 0 & 0 & \dots \sqrt{\lambda_N} \end{bmatrix}$$

Where $\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_N \geq 0$. Each eigenvalue is then substituted into the $\boldsymbol{W} - \lambda_n \boldsymbol{I}$ and row operations are performed on the resultant matrix to convert it into reduced row echelon form to find the eigenvectors (x) that correspond with that eigenvalue.

$$\boldsymbol{W} - \lambda \boldsymbol{I} = \begin{bmatrix} w_{11} - \lambda & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} - \lambda & \dots & w_{2N} \\ \vdots & \vdots & \dots & \vdots \\ w_{M1} & w_{M2} & \dots & w_{MN} - \lambda \end{bmatrix} \xrightarrow{row\ reduction}$$

$$\begin{bmatrix} 1 & 0 & 0 & \dots & d_1 \\ 0 & 1 & 0 & \dots & d_2 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & d_M \end{bmatrix} = \begin{matrix} h_1 + 0h_2 + 0h_3 + \cdots + d_1 h_N = 0 \\ 0h_1 + h_2 + 0h_3 + \cdots + d_2 h_N = 0 \\ \vdots \quad \vdots \quad \vdots \quad \dots \quad \vdots \\ 0h_1 + 0h_2 + 0h_3 + \cdots + d_M h_N = 0 \end{matrix} \rightarrow h_n = \begin{bmatrix} v_{1n} \\ v_{2n} \\ \vdots \\ v_{Mn} \end{bmatrix}$$

The eigenvectors for each of the eigenvalues corresponds to a column in matrix $\boldsymbol{V}$.

$$\boldsymbol{V} = \begin{bmatrix} v_{11} & v_{12} & v_{13} & \dots & v_{1N} \\ v_{21} & v_{22} & v_{23} & \dots & v_{2N} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ v_{M1} & v_{M2} & v_{M3} & \dots & v_{MN} \end{bmatrix}$$

Once $\boldsymbol{S}$ and $\boldsymbol{V}$ are found then $\boldsymbol{U}$ can be determined using $\boldsymbol{U} = \boldsymbol{C}\boldsymbol{S}^{-1}\boldsymbol{V}$, where $\boldsymbol{S}^{-1}$ is the pseudoinverse of $\boldsymbol{S}$, or by repeating the above procedure using $\boldsymbol{X}\boldsymbol{X}^T$ instead of $\boldsymbol{X}^T\boldsymbol{X}$.

4.2.2. Application of SVD to The Covariance Matrix

However, since SVD is being conducted on the covariance matrix $\boldsymbol{C}$ which is a Hermitian matrix then $\boldsymbol{U} = \boldsymbol{V}$, as shown below.

$$\boldsymbol{C}^T\boldsymbol{C} = (\boldsymbol{USV}^T)^T(\boldsymbol{USV}^T)$$

$$\boldsymbol{C}^T\boldsymbol{C} = (\boldsymbol{VSU}^T)(\boldsymbol{USV}^T)$$

$$\boldsymbol{C}^T\boldsymbol{C} = \boldsymbol{VS}^2\boldsymbol{V}^T = \boldsymbol{LS}^2\boldsymbol{L}^T$$

Where $\boldsymbol{L}$ is the loadings matrix and $\boldsymbol{S}$ is the latent matrix; each column of both matrices corresponds to the principal components.

$$\begin{matrix} PC1 & PC2 & PC3 \end{matrix}$$

$$\boldsymbol{L} = \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}$$

$$\begin{matrix} PC1 & PC2 & PC3 \end{matrix}$$

$$\boldsymbol{S} = \begin{bmatrix} s_{11} & 0 & 0 \\ 0 & s_{22} & 0 \\ 0 & 0 & s_{33} \end{bmatrix}$$

The variance explained is calculated using the latent matrix and show the percent contributed by each principal component to the variance of the data.

$$\boldsymbol{VE} = \frac{s_{mm}}{\sum_{m=1}^{M} s_{mm}} \times 100$$

Finally, the scores matrix ($\boldsymbol{T}$) is determined by taking the product of the original data matrix and the loadings. When the scores for the first and second principal components are graphed against each other the points furthest from the origin correspond with data with the greatest variance.

$$\begin{matrix} PC1 & PC2 & PC3 \end{matrix}$$

$$\boldsymbol{T} = \boldsymbol{XL} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & t_{43} \\ t_{51} & t_{52} & t_{53} \end{bmatrix}$$

4.3. Application of PCA to Raman Spectra

For the experimental data collected from the serial dilution of glucose, PA and SO training sets and PA growth were exported to Microsoft Excel spreadsheets. Then the wavenumbers and their corresponding counts are imported into MATLAB as two matrices and labeled accordingly. PCA was performed on the counts matrix as described in Appendix 2, all the resultant output matrices are exported to Excel spreadsheets. Two methods of calculating PCA were performed, one called the "Serial" method and the other called the "Wavenumber" method. The "Serial" method was performed with the wavenumbers of the spectrum as the rows and the series, e.g., changing bacteria concentration, as the columns of the data matrix before PCA is conducted.

$$\begin{array}{l} \text{Series } 1 \quad 2 \quad 3 \\ \text{Raman Shift} \downarrow \end{array} \boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \\ x_{41} & x_{42} & x_{43} \\ x_{51} & x_{52} & x_{53} \end{bmatrix} \overset{PCA}{\longrightarrow} \boldsymbol{T} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & t_{43} \\ t_{51} & t_{52} & t_{53} \end{bmatrix}$$

The variance of the data is extrapolated from the scores by graphing the first two principal components, for example figure 34 right. For the "Wavenumber" method the data matrix is transposed so that the series are the rows, and the wavenumbers are the columns then PCA is performed.

$$\boldsymbol{X}^T = \begin{array}{c} \text{Raman Shift} \longrightarrow \\ \begin{bmatrix} x_{11} & x_{21} & x_{31} & x_{41} & x_{51} \\ x_{12} & x_{22} & x_{32} & x_{42} & x_{52} \\ x_{13} & x_{23} & x_{33} & x_{43} & x_{53} \end{bmatrix} \end{array} \begin{array}{l} \text{Series} \\ 1 \\ 2 \\ 3 \end{array} \overset{PCA}{\longrightarrow} \boldsymbol{L} = \begin{bmatrix} l_{11} & l_{12} & l_{13} & l_{14} & l_{15} \\ l_{21} & l_{22} & l_{23} & l_{24} & l_{25} \\ l_{31} & l_{32} & l_{33} & l_{34} & l_{35} \\ l_{41} & l_{42} & l_{43} & l_{44} & l_{45} \\ l_{51} & l_{52} & l_{53} & l_{54} & l_{55} \end{bmatrix}$$

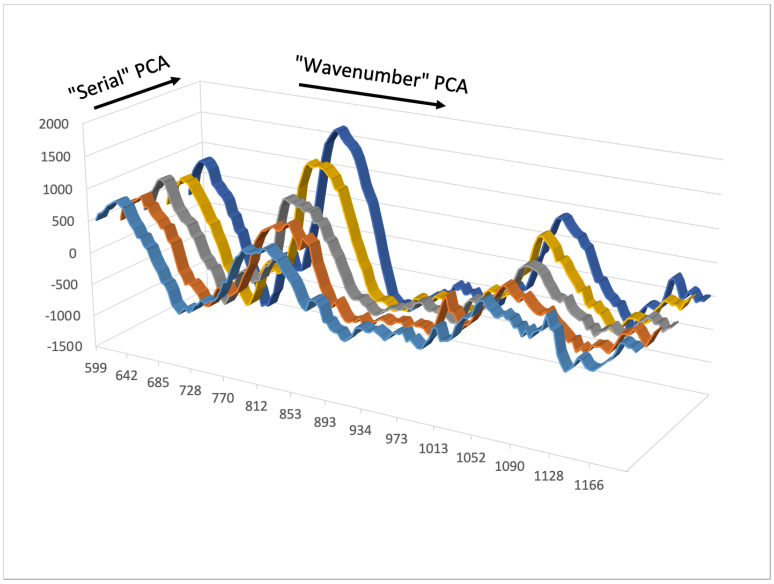

Figure 34: A visual depiction of the treatment of data by the two methods of PCA. “Serial” PCA analyzes the data across spectra and “Wavenumber” PCA across the Raman shift of each spectrum.

This method produces as many rows in the loadings matrix as wavenumbers in the spectrum and the first two principal components are graphed against the Raman shift to produce a graph that resembles the spectrum, for example figure 41 right.

To help interpret the results of PCA of real Raman spectra of bacterial cultures i.e., how the cultures and biological processes relate to their corresponding spectra, three series of model series were constructed with plausible trends in changing “Raman” features from spectrum to spectrum chosen to mimic time related data and demonstrate a use of the Serial method approach to PCA. Differences in how the synthesized Raman features were constructed provide a basis for judging how sensitive the Wavenumber approach is to complex line shapes created by overlapping features and in some cases by the process of baseline subtraction.

These model series were constructed using “peaks” that have a rectangular base and attaching a pointed cap that is the same for all constructed peaks. The rectangular base is made longer/shorter, but not wider, to cause a peak to increase/decrease. Wavenumber method PCA of changing sets of features created and scaled in this way produces the rectangular base and a derivative-like line shape as the first two components, that when scaled and combined, using the PCA coefficients, reproduce the >99% of the raw input data variance.

When a peak is decreasing, the sense of the “derivative” is opposite from when the peak is increasing across the time series. This line shape effect is associated with the exact approach for constructing these spectra and not for realistic data, but we do think the general behavior has implications for recognizing correlated changes across a series of Raman spectra.

4.3.1. Model Series

Figures 35 through 40 show the results of the PCA on the model series using the serial method. The scores for the first two principal components of each model series are displayed in figures 35 right, 37 right and 39 right along with their respective spectra to the left. While figures 36, 38 and 40 show the percent variance explained by each of the principal components.

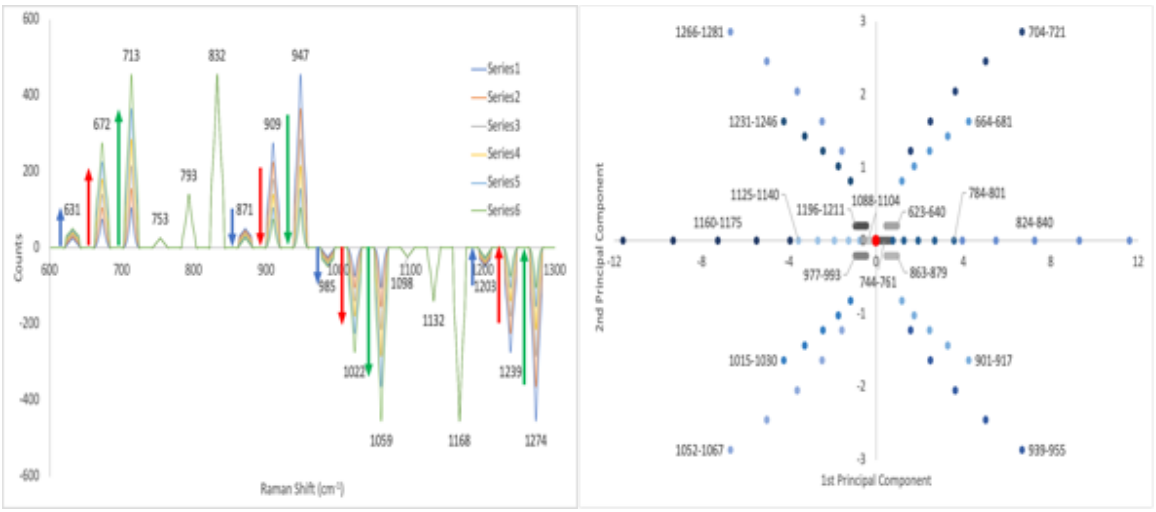

Figure 35: Model series one with peaks that are symmetrical about the x-axis (left). PCA scores for the 1st and 2nd Principal Components and their corresponding spectra regions (right). Using the serial method. Arrows point in the direction of varying peak counts from series 1 to 6, same color indicate peaks that similar magnitude variation.

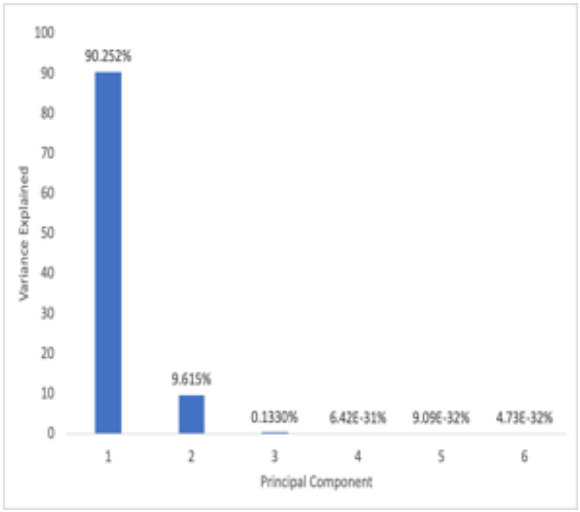

Figure 36: The percent variance explained by each of the principal components for model series one. Serial method.

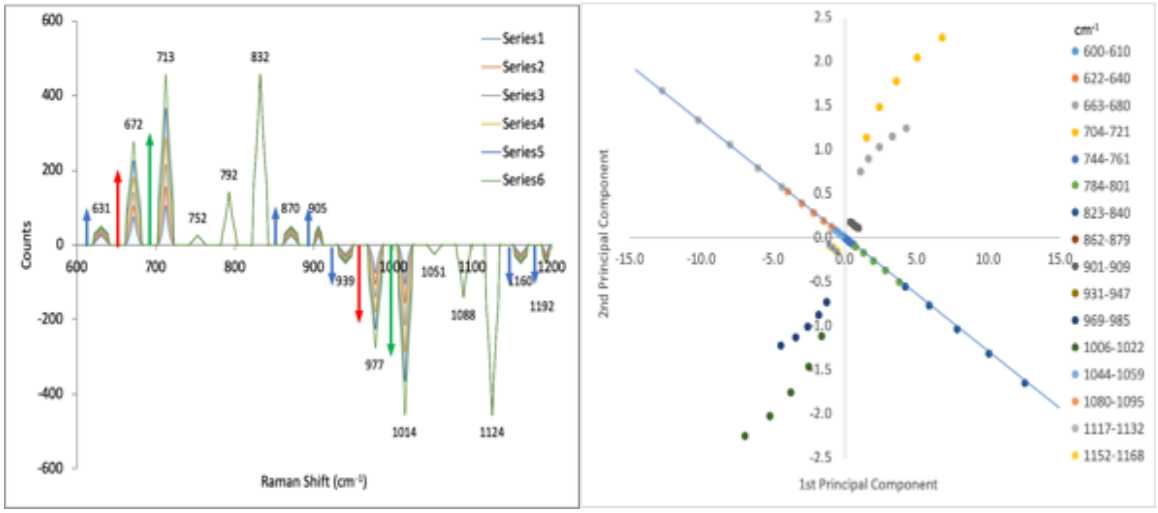

Figure 37: Model series two with peaks of various size and symmetrical about the x-axis (left). PCA scores for the 1st and 2nd Principal Components for the model series (right). Using the serial method. Arrows point in the direction of changing peak counts from series 1 to 6. Arrows with the same color indicate peaks that vary in the same magnitude.

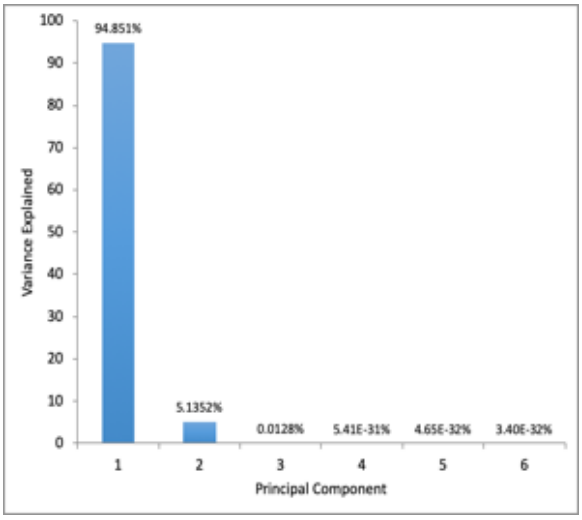

Figure 38: The percent variance explained by each of the principal components for model series two. Serial method.

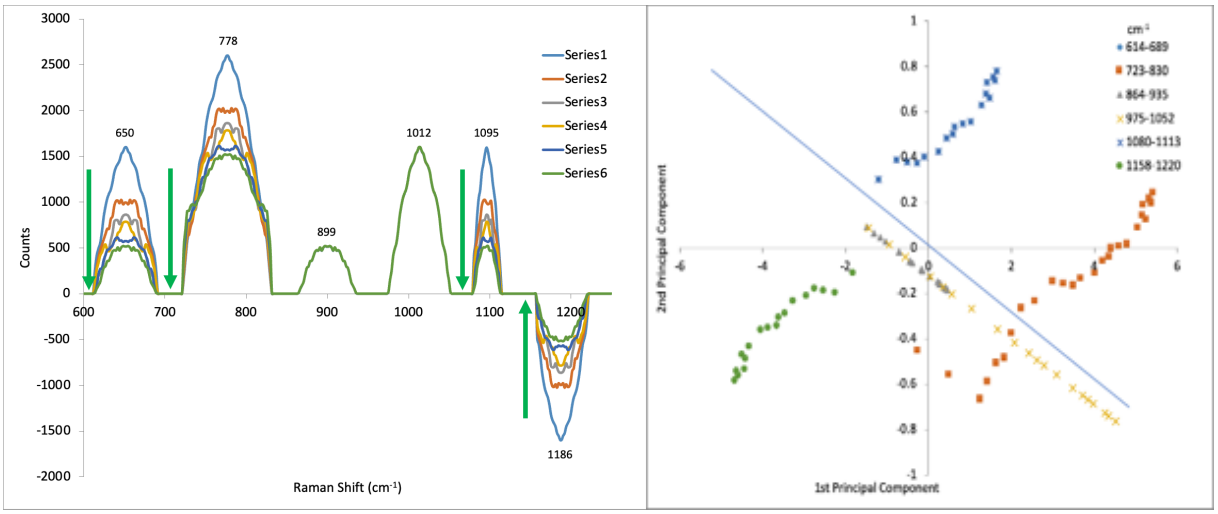

Figure 39: Model series three with peaks of various sizes (left). PCA scores for the 1st and 2nd Principal Components for the model series (right). Serial method. Arrows point in the direction of changing peak counts from series 1 to 6. Arrows with the same color indicate peaks that vary in the same magnitude.

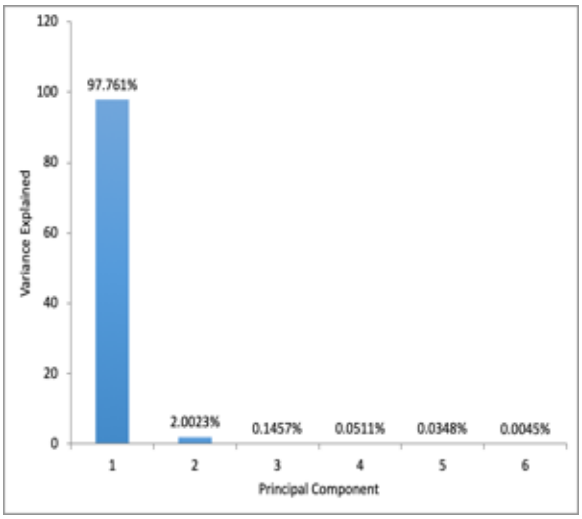

Figure 40: The percent variance explained by each of the principal components for model series three. Serial method.

The following figures show the PCA results for the same model series using the wavenumber method so that the data can be represented on the same Raman shift scale as the spectra.

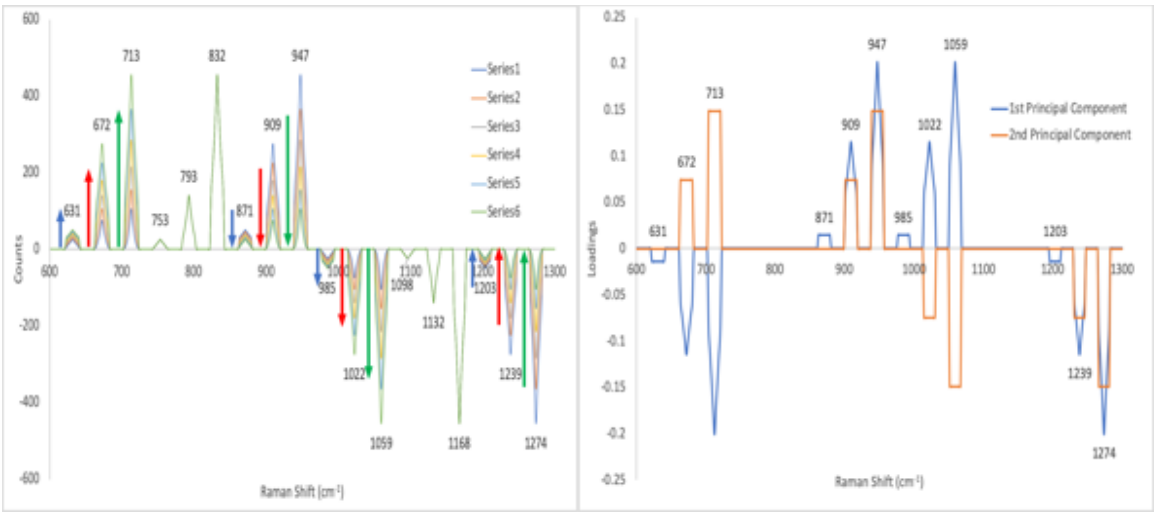

Figure 41: Model series one (left) and the loadings of the first two principal components graphed versus the wavenumbers for model series one (right). Wavenumber method. Arrows point in the direction of changing peak counts from series 1 to 6. Arrows with the same color indicate peaks that vary in the same magnitude.

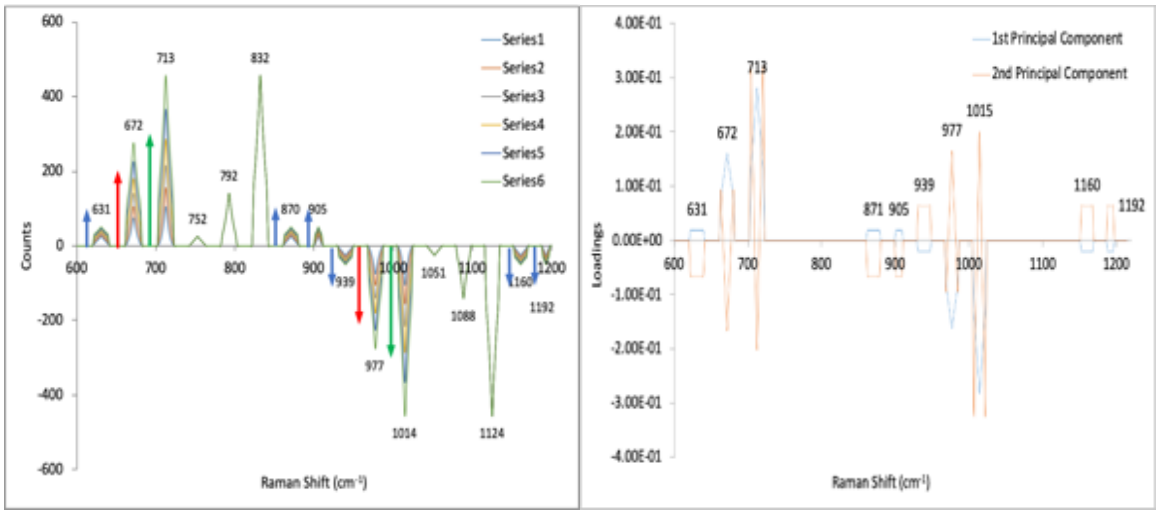

Figure 42: Model series two (left) and the loadings of the first two principal components graphed versus the wavenumbers for model series two (right). Wavenumber method. Arrows point in the direction of changing peak counts from series 1 to 6. Arrows with the same color indicate peaks that vary in the same magnitude.

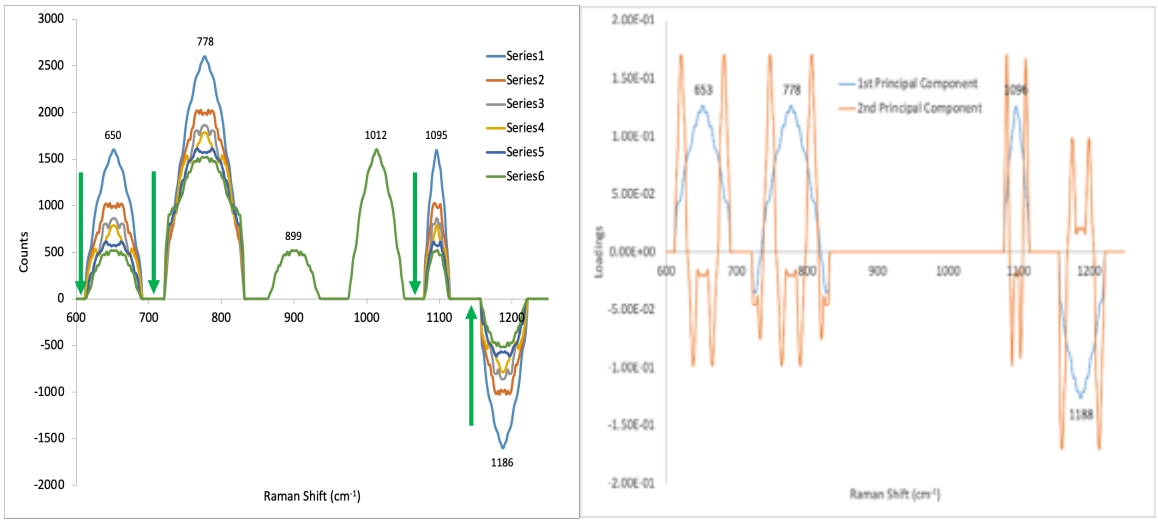

Figure 43: Model series three (left) and the loadings of the first two principal components graphed versus the wavenumbers for model series three (right). Wavenumber method. Arrows point in the direction of changing peak counts from series 1 to 6. Arrows with the same color indicate peaks that vary in the same magnitude.

4.3.2. Glucose Calibration Curve

PCA was performed on the spectra of the glucose calibration curve to further aid in interpreting the data. The following graphs show the results of the analysis for the glucose spectra. Glucose was selected because of the correlation between the 1120cm^{-1} peak in its Raman spectrum and the concentration shown in figure 21.

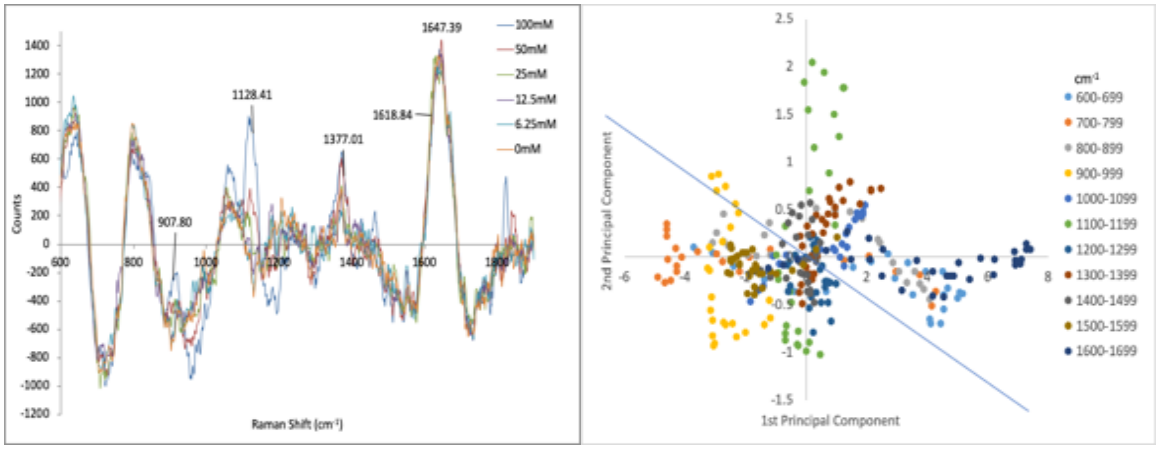

Figure 44: The spectra of glucose at various concentrations (left) and the scores of the first two principal components of the serial method (right).

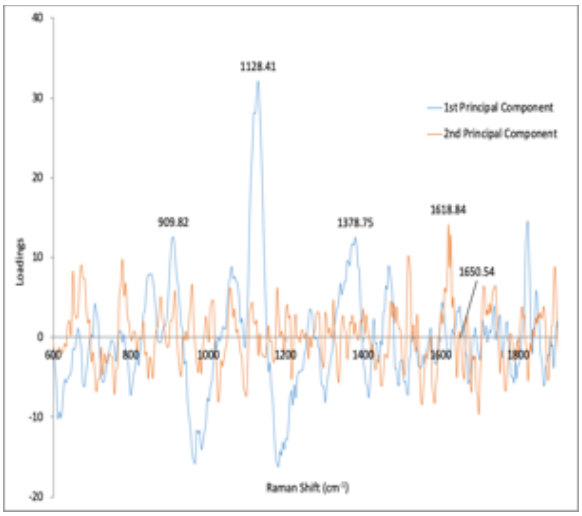

Figure 45: The wavenumber method of PCA for the glucose spectra.

4.3.3. PA and SO Training Sets

The following graphs show the results of PCA on the training set Raman spectra of *P. aeruginosa* and *S. oneidensis* for both the serial and wavenumber methods.

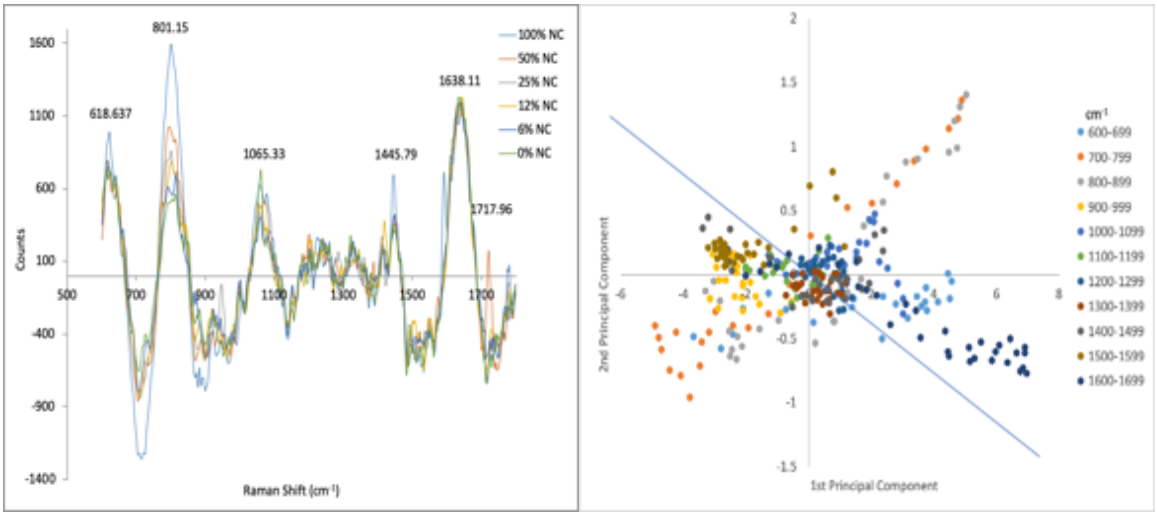

Figure 46: The Raman spectra for the training set of PA (left) and the serial method of PCA showing the scores for the different regions of the spectra (right).

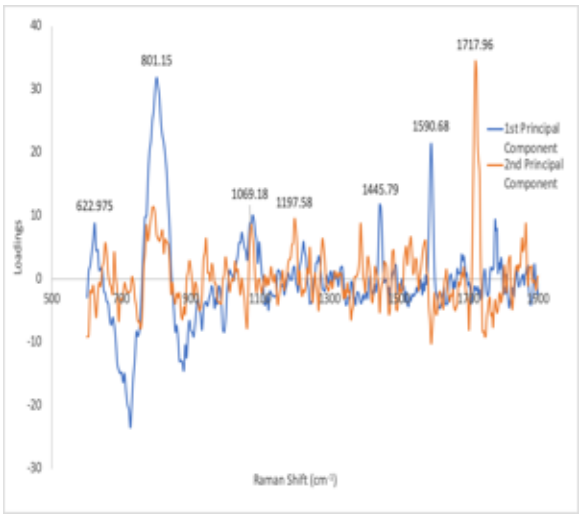

Figure 47: The wavenumber method of PCA for the spectra of the training set for PA.

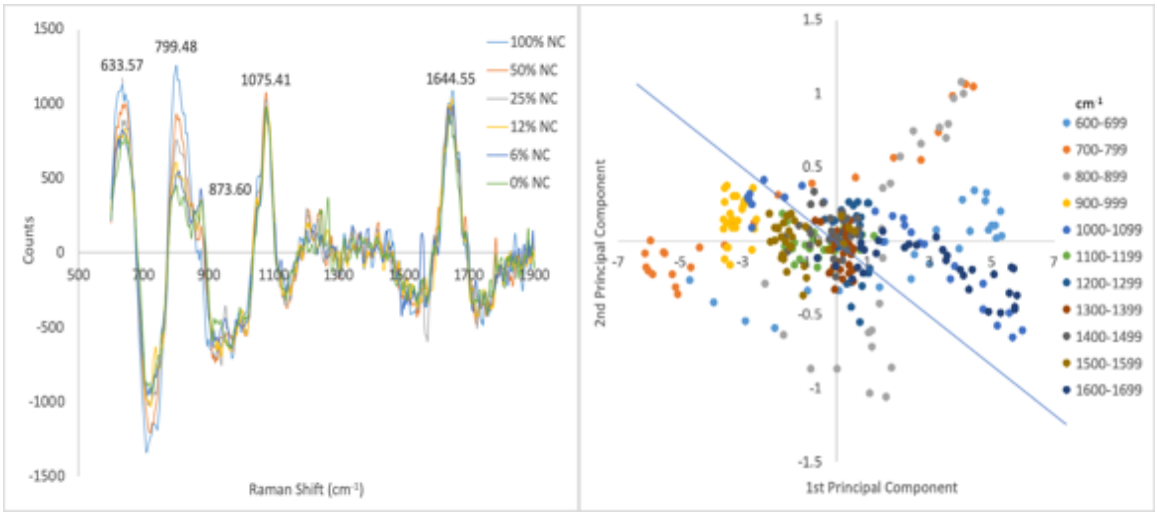

Figure 48: The Raman Spectra of the training set for SO (left) and the serial method of PCA showing the scores for the different regions of the spectra (right).

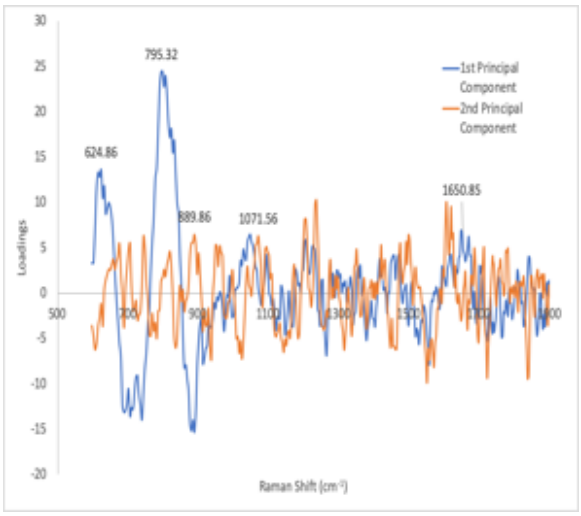

Figure 49: The wavenumber method of PCA for the spectra of the training set for SO.

4.3.4. Bacterial Growth Curve

Figures 50 and 51 show the results of both the serial and wavenumber methods of PCA performed on the Raman spectra of the twelve-hour PA growth.

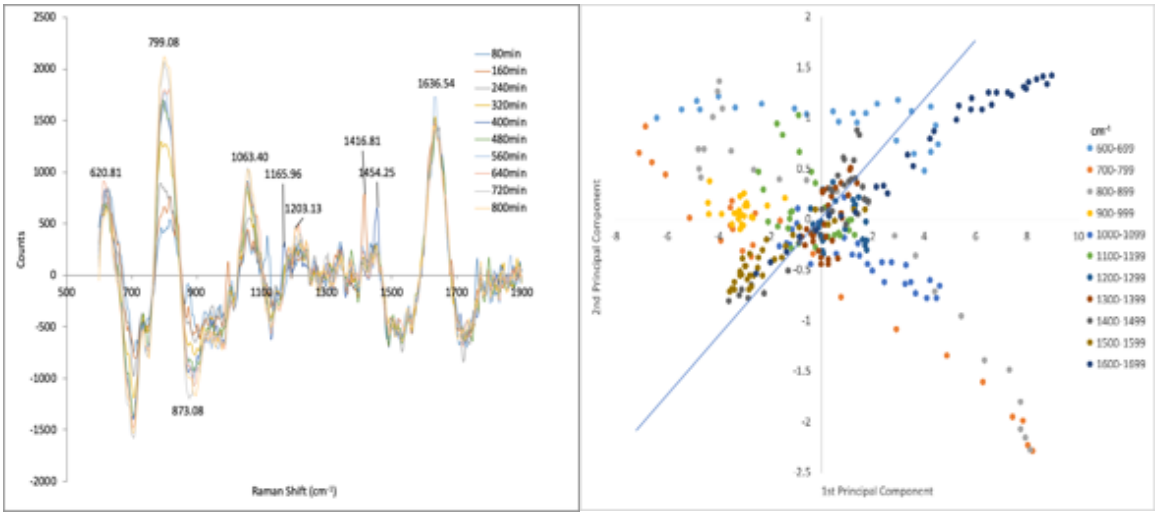

Figure 50: The Raman spectra of the twelve-hour growth for PA (left) and the serial method of PCA showing the scores for the different regions of the spectra (right).

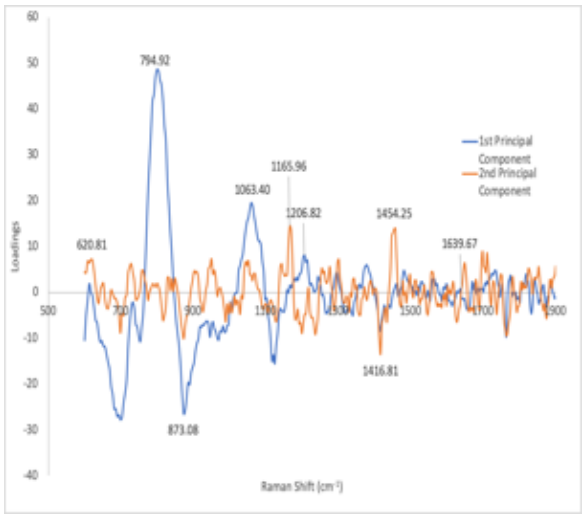

Figure 51: The wavenumber method of PCA for the spectra of the twelve-hour growth for PA.

Much like with the biochemical analysis of the training sets and the calibration curve of glucose some of the Raman peaks were integrated. To help determine what Raman peaks in the spectra appear as prominent features in the serial method of the PCA the labeled peaks in the spectra were integrated to show their change over time for PA growth (figures 52-54).

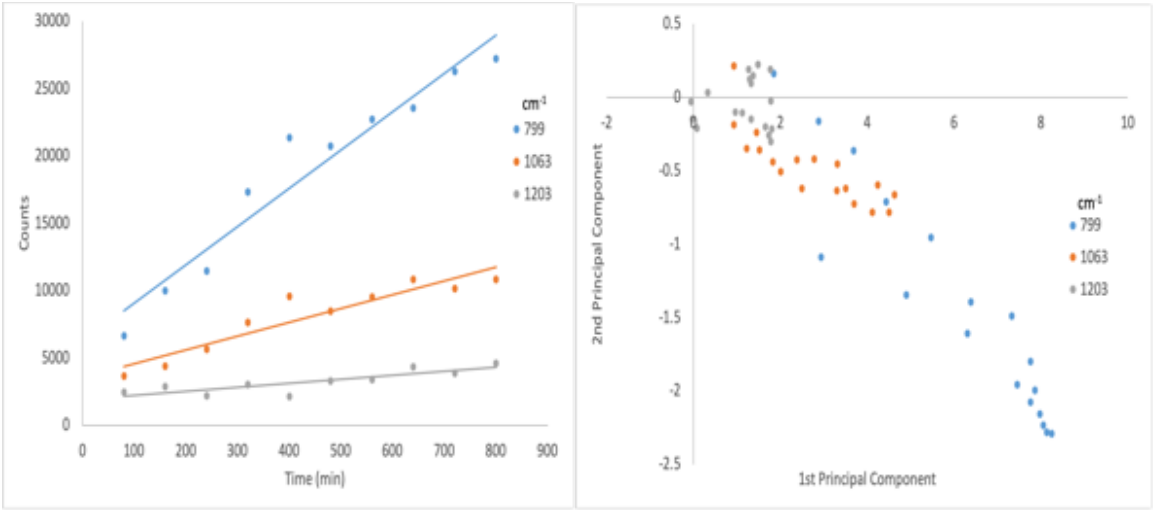

Figure 52: The integration of Raman peaks at about 799cm^{-1}, 1063cm^{-1} and 1203cm^{-1} (left) and their associated PCA scores for the serial method (right).

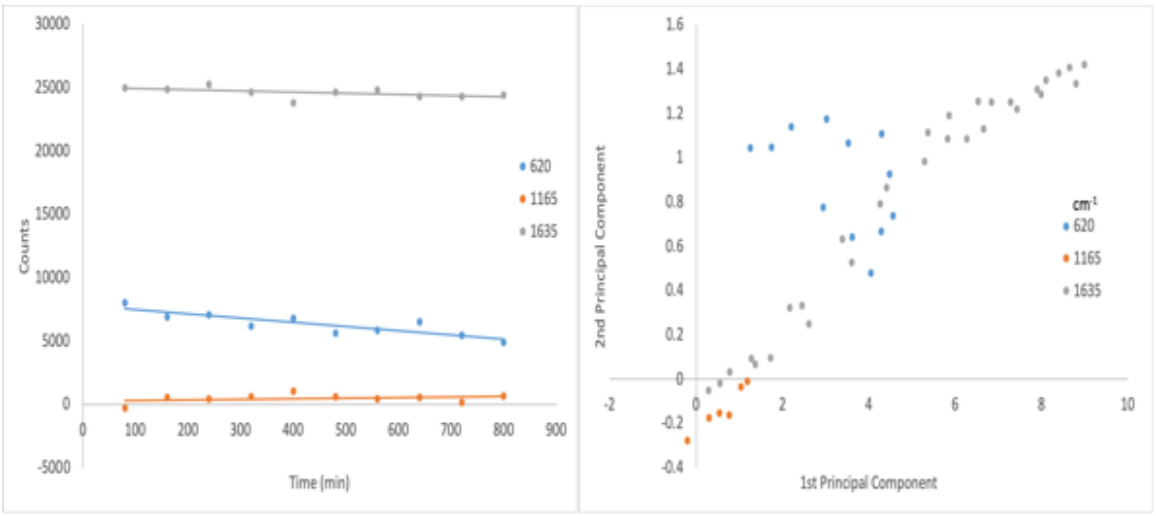

Figure 53: The integration of Raman peaks at about 620cm^{-1}, 1165cm^{-1} and 1635cm^{-1} (left) and their associated PCA scores for the serial method (right).

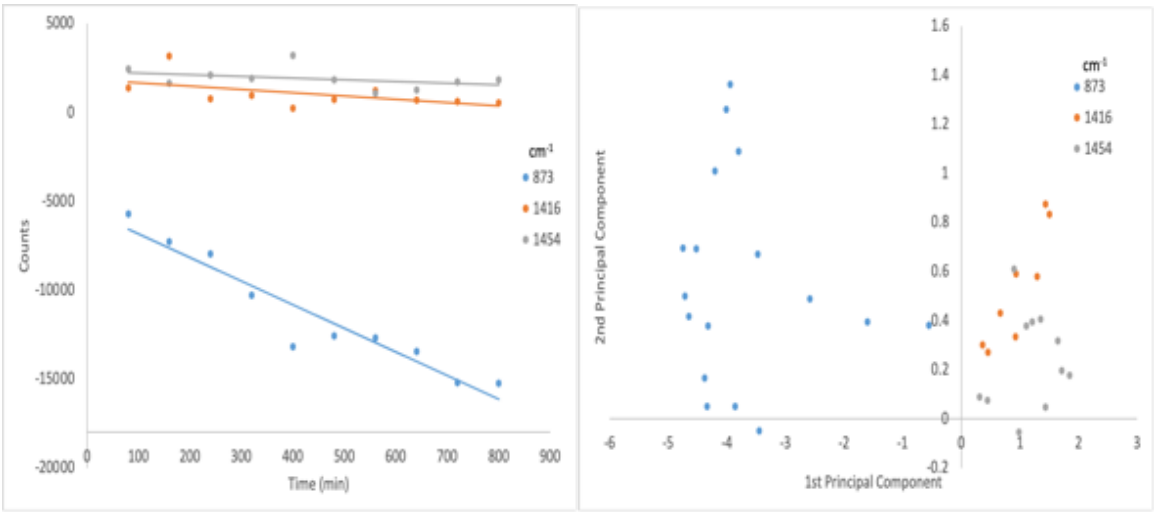

Figure 54: The integration of Raman peaks at about 873cm^{-1}, 1416cm^{-1} and 1454cm^{-1} (left) and their associated PCA scores for the serial method (right).

5. Discussion

Figures 9 and 11 (pg. 32 and 34) demonstrate the validity of the BSN algorithm, with the OD_{600} ϕ_{bac} and ϕ_{med} plotted against their respective calculated values from equations [5] and [6]; the slopes would be expected to be 1 and the intercept at 0. With uncertainties of fitted parameters denoted as ±σ i.e., one standard deviation, at 95% confidence i.e., at two standard deviations this

is observed in figure 9 for PA ϕ_{bac} with a slope of 0.9045 ± 0.04976 and intercept of 6.74×10^{-7} ± 5.27×10^{-7}, for the ϕ_{med} that has a slope of 0.989 ± 0.01762 and intercept at 0.0034 ± 0.0083. This is evidence of the accuracy of the algorithm in the specified range. Another aspect of the graphs that show the validity of the BSN algorithm is the grouping of the data points. The precision of the algorithm is displayed by the close grouping of the data points that distinctly show each of the series of the training set. This process was repeated for the twelve-hour growth curve of *P. aeruginosa* to create figure 13 (pg. 36).

Figures 12 - 13 that display the BSN versus time and OD_{600} ϕ_{bac} for the twelve-hour growth curve shows that the trend is very close to being linear and that the linearity in the trend stops about halfway. This corresponds to an OD_{600} of about 0.4 according to figure 6 (pg. 26), when creating the calibration curve, it was determined to be the end of the linear range of bacteria concentration as related to the EE and IE signals. Only including the first five data points of figure 12 gives a linear graph with r^2 value of 0.904. In an earlier study using quartz spheres and aqueous porphyrin solutions,[4] showed that terms higher order than linear in the EE or IE can be added to the simple linear equations [3] through [6] given above to extend the range of utility at least to some extent above the single scattering limit.

To demonstrate the ability of the BSN method to detect disruptions to PA; growth curves were conducted then were either exposed to UV light or bleach. For comparison control experiments shown in figure 15 (pg. 37), were conducted with and without disruption, these controls were blank LB media solutions (without bacteria) exposed to the same conditions as the growth curves with bacteria present. Figure 14 (pg 36) shows the comparison of the volume fractions of

bacteria (left) and media (right) for the growth curve with and without UV light exposure. At 210min into the growth there is an immediate decrease in the rate of growth of the volume fraction of bacteria, corresponding to turning on the UV light, that levels off, at lower, preexposure level as the UV stays on. Apparent consumption of medium i.e., decreases in the media concentration also decreases compared to the non-UV exposed culture. While the growth without UV light exposure continues to increase. As for the volume fraction of media, there is a sudden decrease before the system recuperates and the media shows a similar constant trend as the growth without the exposure. Some of the apparent sudden decrease of the volume fraction of media can be due to the effect that UV light has on fluorescent molecules that may be in the media. Photooxidation to produce carbonyls and other emissive molecules is well known. While the blank LB media under growth conditions with and without bleach (figure 15, pg 37) show no effect to the volume fractions of bacteria and media.

The UV light and bleach are different in that the UV light immediately penetrates the culture deeply, completely or at least until it is all absorbed. The bleach must be placed at the surface of the culture and must diffuse throughout before having an effect on the probed volume. Both the bacteria and media produced apparent abrupt increases on addition of bleach to the blank solutions. Note that the figure 15 BSN bacteria measurements is very low and shows reduced evaporation effects with UV light but effect on the media measurement is much less. We suggest that in the blanks, bleach itself may interact with the laser light producing emission and the UV exposure causes photocoagulation of particles in the media. These larger particles accumulate at the surface of the cuvette producing the observed effect.

For the bleach inoculated growth curve (figure 16, pg 38) the volume fraction of bacteria seems to be unaffected by the addition of bleach at 120min, 150min 180min or 210min. However, the volume fraction of media increases starting at the first inoculation of bleach. The bleach was added in 25µL aliquots, which may not be a sufficient amount to affect bacteria growth but does affect the volume fraction of media. This may be cause by the reaction between sodium hypochlorite (NaOCl) and the media to increase the fluorescence; this is supported by the data attained from the blank media. Figure 17 (pg 38) shows that both volume fractions increase when the samples are inoculated with bleach even in the absence of bacteria. One can expect that almost any oxidizing agent added to a growth medium will produce fluorescent carbonyl compounds. The presence of oxidizable "sacrificial" medium could protect the bacteria from oxidation.

Just with the training set for PA, figure 11 (pg 34) shows the validity of the algorithm for SO ϕ_{bac} with a slope of 0.9764 ± 0.02602 and intercept of $1.09\text{x}10^{-7} \pm 1.76\text{x}10^{-7}$, SO ϕ_{med} has a slope of 0.9757 ± 0.02639 and intercept at 0.00783 ± 0.01244. And the graphs show the validity of the BSN algorithm via the grouping of the data points. The precision of the algorithm and reproducibility of the experimental procedure is displayed by the close grouping of the data points that distinctly show each of the series of the training set.

The training set process was repeated for SO using a lactate-based media, which is used for microbial fuel cells (MFC). However unlike with the SO training set in LB media there was no linear correlation between bacteria or media concentration and the signals, i.e., EE and IE. Figure 24 (pg. 44) shows the EE and IE signals for the varying concentrations of bacteria and media.

Note that for the top two graphs of the figure the series with 100% media shows the lowest signal counts and are nearly constant despite variations to the bacteria concentration. It seemed suspicious that at 100% lactate media the EE and IE signals would be at their lowest range, and it was concluded that the media might be interfering with the back-scattered light. Figure 25 (pg. 45) shows the baseline corrected Raman spectra that appear with an abundance of noise when compared to the spectra for LB media (figure 19 right, pg. 41). Figure 26 (pg. 45) shows the baseline-corrected Raman spectra for the lactic acid used to prepare the media (left) and the lactate media minus the Wolfe's trace mineral (right).

After reviewing these data, it was determined that a chemical species in the Wolfe's trace mineral may be interfering with the Raman spectra. UV-Vis analysis confirmed that the Wolfe's trace mineral absorbs light in a range from about 600nm to 800nm, around the wavelength of the incident light for our Raman spectrometer (785nm). Wolfe's trace mineral is a mixture of EDTA and various inorganic salts; some of the metal cations found in the solution are Fe, Zn, Cu and Ni.

Because of their electron configuration it was believed that one or more of these cations were responsible for the absorbance of the incident light. 100mM each of $FeSO_4 \cdot 7H_2O$, $ZnSO_4 \cdot 7H_2O$, $CuSO_4 \cdot 5H_2O$ and $Ni(NO_3)_2 \cdot 6H_2O$ were prepared and analyzed by UV-Vis and Raman spectrometer. Figure 27 (pg. 46) shows the UV-Vis spectra (top left) and baseline-corrected Raman spectra (top right) of the cation solutions. For the UV-Vis spectra, Cu^{2+} is the cation that shows the most absorbance in the range of the incident light for the Raman spectrometer though there is some absorbance from Fe^{2+} as well. The Raman spectra show a pronounced narrow peak

at about 990cm^{-1} for the compounds containing sulfate and another at about 1070cm^{-1} peak for nitrate, despite having an absorbance the sulfate with Fe^{2+} can be seen. However, for the Cu^{2+} solution there are no discernible peaks in either of these regions and no peak at about 1600cm^{-1} for water.

As is shown in figure 22 (pg. 43) both the 800cm^{-1} and 1065cm^{-1} peaks increase as the growth progresses and the 1120cm^{-1} decreases. Associating these Raman features with the chemical species and components of the bacterial metabolic cycles we suspect is supported by the trends displayed in figure 22. Bacteria cells multiply, producing more phospholipid membranes and increasing the 1065cm^{-1} peak of the Raman spectra. Conversely, increased bacteria lead to increased consumption of glucose and a decrease of the 1120cm^{-1} peak. For both the six- and twelve-hour growths of PA the 1120cm^{-1} feature decreases as glucose is consumed. But for the twelve-hour growth curve the change in the 1120cm^{-1} peak seems to become constant which would be expected if the bacteria growth slows along with the consumption of glucose.

The association of specific chemical species in the culture with Raman features in the spectra could be a lengthy process if attempting to do so by simply integrating peaks. Often the biochemical literature contains a reference to functional group frequencies as a strong basis for and “assignment”. We have focused our attention on PCA precisely because it is a rigorous method of finding multiple changing Raman peaks that change in a correlated and deterministic manner as the concentration(s) of stoichiometrically associated analyte(s) change, i.e., bacteria and feedstock change in a directly related manner. But in these results, we also found the first medium system which could not be easily calibrated empirically.

In making comparisons between media, we also realized we needed to discover and explain other possible causes and effects that may be operating in growth curve data. In observing the effects of bleach addition and UV light exposure to cultures in real-time we anticipate experiment would be expected to be different to that of a 400,000-liter capacity, continuous, bioreactor operating as a tightly controlled open system being subjected to analogous stimulus and probes. Probing such a system through a viewing port in the side of a big tank may limit the regions of such a large culture that can be observed.

Specifically, as suggested by the microscopic environment implied by figures 28 and 29 (pg. 48 and 49), diffusion is a plausible explanation for the difference in the shape for *E. coli* (EC) and *P. aeruginosa* growth curves given that EC has a faster doubling time than PA. The lag in growth rate between PA and EC would account for the differences in growth curves. A higher growth rate means that there are more collisions between bacteria cells and a faster buildup of diffusion driving force. This means that EC cells travel to the plane of intersection with the laser focal point much faster than PA, creating the apparent lag phase in the first ~25 minutes of the PA growth curve and the absence of one in the EC growth curve. Previous research has reported the diffusion coefficient of *P. aeruginosa* as 2.1 x 10^{-9} m^2s^{-1} and another study performed by this lab has measured the diffusion coefficient of *E. coli* as 1.56 x 10^{-9} m^2s^{-1}.[6,47] And the laboratory doubling time for PA and EC being reported as 30min and 20min respectively.[48] Although the diffusion coefficient of EC is smaller than that of the experimental PA diffusion coefficient of 2.1 x 10^{-9} m^2s^{-1} the doubling time is greater than that of PA. This suggests that there is no correlation between the bacteria growth and their diffusion. And the bacteria may be providing the driving force for the observed diffusion kinetics, both PA and EC are motile bacteria having

flagella. When motile bacteria (PA) were compared to immotile bacteria (*Klebsiella pneumoniae*) the diffusion coefficient of the motile bacteria was found to be 2.3 time greater.[47]

Our goal was not and could not settle these and related issues here. It was only to show that BSN provides a new unexplored path towards bringing quantitative metrics to characterization of viable microbial systems in various useful contexts. Regardless of the values for D we ultimately calculate, we can state that the temporal variation in bacteria concentration is neither constant in time nor location within the cuvette. This effect must be considered when applying BSN to the various situations we envision i.e., from large scale bioreactors to microfluidic systems and simple cuvettes. BSN could provide metrics for quantifying viability and motility to characterize the state of a culture, and possibly provide feedback to control local environment e.g., temperature, presence of nutrients and waste, flow rates in large scale and continuous applications.

BSN monitoring provides a natural approach for simultaneously determining viability and number density. We suggest that viability can be accessed by measuring consumption of nutrients, production of metabolites and by apparent mobility as determined by these kinds of growth curves. The apparent reduction in the concentration of "medium" i.e., materials that produced IE in the system is at least as much by the bacteria displacing the medium from the probed volume as it is by the actual chemical/physical rearrangement of the nutrients in the medium to being inside the bacteria as catabolites, anabolites and generic metabolites.

To be useful as an analytical tool, the information that PCA provides must be interpretable.

Some model series were constructed and analyzed to display the output provided by PCA under such conditions as a closed cuvette with most chemical processes involving metabolism. For model series one (figure 35 left, pg. 61) the serial method yielded a scores graph (figure 35 right, pg. 61) that is easily interpretable, all regions of the spectra that remain constant from series to series have their corresponding scores located on the x-axis. Since the spectra are symmetrical around the x-axis, the mean counts are at zero and all regions that are constant at zero counts have their scores located at the origin. All other regions of the spectra have their scores located at different quadrants of the graph depending on whether they are above the x-axis and how they are changing between series. Simple growth curves contain both catabolism and anabolism i.e., the decrease of Raman features because of bacterial consumption and the correlated increase of other features because of anabolism.

Using conventional quadrant number classification where quadrant I correspond to points with coordinates (+, +), II with (-, +), III with (-, -) and IV with (+, -). All scores found in quadrant I are associated with regions in the spectra that are above the x-axis and have increasing counts between series i.e., anabolism. The scores for regions that are above the x-axis that have decreasing counts between series are in quadrant IV i.e., catabolism. And all regions that are below the x-axis have their scores located in quadrants II and III with the ones that have increasing counts in II and decreasing counts in III. The greater the changes between the series for a region of the spectra the further from the origin its score will be with respect to the 2nd principal component axis, increasing changes have more positive and decreasing having greater negative values.

As for the wavenumber method of PCA (figure 41 right, pg. 64) seems to show similar results in a format that makes it easier to determine regions of the spectra with the greatest changes. The first three regions of the spectra (labeled 631, 672 and 713cm^{-1}) were created to resemble peaks that increase in total intensity and changes between series. Their corresponding regions in the spectral graph shows them to have increasingly negative values for the 1st principal component as the overall size of the spectra regions increase. However, the 2nd principal component for these regions shows them to be increasingly positive and the 631cm^{-1} peak has no corresponding region in the spectral graph.

This might be due to the 631cm^{-1} peak changing between the series at equal increments while the other two (672 and 713cm^{-1}) are increasing at greater intervals. The second group of three spectra regions (753, 793 and 832cm^{-1}) are not changing between the series and so have no corresponding regions in the spectral graph for either principal component. The third group of three spectra peaks (871, 909 and 947cm^{-1}) are the same size as the first three but are changing between series in a manner that is proportionately inverse. The 871cm^{-1} peak only shows a corresponding region for the 1st principal component but nothing for the 2nd. But the peaks at 909cm^{-1} and 947cm^{-1} show regions in the spectral PCA graph that are positive for both principal components.

Just like with model series one, model two (figure 37 left, pg. 62) was designed to be symmetrical around the x-axis so that the mean counts will be zero. The difference in model series two is that each positive peak increase between the series and has a corresponding negative peak that decreases proportionally. Figure 37 right (pg. 62) shows the scores graph of

the serial method of PCA and like model series one all regions of the spectra with zero (mean) counts have their scores located at the origin. However, unlike model series one, all regions with non-zero counts that are invariant between the series have scores that are not located directly on the x-axis (1st principal component). All regions that have non-zero counts that are invariant between series have their scores located along a diagonal line that passes through the origin and extends into quadrants II and IV, the blue line in figure 37 right (pg. 63). Regions with counts greater than the mean are located further from the origin and in quadrant IV; regions with counts lower than the mean are in quadrant II.

Other than this difference, the scores graph for model series two resembles that of model series one with the least variant regions of the original spectra having scores that are parallel to the invariant scores. And regions of the original spectra that have increasing variations between the series have scores that become closer to being perpendicular to the diagonal line formed by the invariant scores. For the wavenumber method of PCA (figure 42 right, pg. 64) the 1st principal component resembles the original spectra with regions that have greater variation having larger corresponding PCA peaks. However, the 2nd principal component shows different peaks for some portions of their corresponding regions of the original spectra. The beginning and ending of each peak have proportionally correlating peaks but have inverted peaks in the 2nd principal component at their center.

Model series three (figure 39 left, pg. 63) was created to resemble more closely the shape of peaks found in an actual spectrum but to aid in interpreting the result only a few varying regions are included. Although the peaks are different sizes they are varying from series to series by the

same increments. And this model series has only one negative peak to determine what the affect would be with a non-zero mean. The scores graph from the serial method of PCA (figure 39 right, pg. 63) shows a similar configuration to the one displayed for model series two. All invariant regions of the spectra that are at non-zero counts form a line that is diagonal to the x-axis and extend through quadrants II and IV. And all the regions that are varying between series have scores that branch out perpendicularly from the invariant scores line; the region with the greater counts having scores further down the line formed by the invariant scores. Unlike the statistical method results for the other model series, this one shows no scores at the origin of the graph and the invariant scores are mostly located in quadrant IV. Figure 43 right (pg. 65) shows the results of the wavenumbers method and how it resembles the loadings method for model series two (figure 42 right, pg. 64) with respect to both principal components.

Since a calibration curve for glucose had already been developed, both methods of PCA were performed on those spectra to produce figures 44 right (serial) and 45 (wavenumber). The Raman spectra of glucose (figure 44 left, pg. 65) has a known peak at about 1120cm^{-1} that proportionally correlates with decreasing concentration. In the scores graph of the serial method (figure 44 right, pg. 65) there are two prominent groups, one corresponding to about the 1120cm^{-1} and the other to the 1650cm^{-1} region of the spectrum. The scores that correspond to the 1650cm^{-1} region are located parallel to the 1st principal component axis near zero relative to the 2nd principal component. Conversely the 1120cm^{-1} region scores are located parallel to the 2nd principal component near zero relative to the 1st principal component. This resembles the pattern seen for model series one where the scores corresponding to the invariant regions of the spectra are located along the 1st principal component. Its resemblance to model series one could be due

to it being a simple two component solution, glucose and water. The greater the variation of the region the closer their scores come to being parallel to the 2nd principal component axis.

As for the loadings graph of the wavenumber method (figure 45, pg. 66) the 1128cm^{-1} peak that is associated with glucose can be prominently seen for the 1st principal component but does not appear for the 2nd principal component. However, the 1650cm^{-1} (water) does not appear for either principal component just as the case with the invariant regions of the model series. Another possible Raman feature for glucose is the peak at 1377cm^{-1} that appears to increase with the concentration in both the scores graph (brown points in figure 44 right, pg. 65) and the loadings graph (figure 45, pg. 66).

The results of PCA performed on the training set of PA and SO are displayed in figures 46 to 49 (pg. 66 - 68). Figure 46 left (pg. 66) shows the Raman spectra of PA for the training set and is decreasing in concentration of bacteria between series and there are some prominently displayed peaks, most notably the regions about 620, 800, 1065 and 1640cm^{-1}. For the scores graph of the PCA (figure 46 right, pg. 66) the scores that correspond to peaks 620 and 1640cm^{-1} are shown to follow a trend that is similar to the invariant regions of model series two. While the scores that correspond to the 800 and 1065cm^{-1} peaks are located diagonally from the origin and extend into quadrant I much like the variant scores of the model series. The wavenumber method is displayed in figure 47 (pg. 67) and shows that of the four peaks of interest in the Raman spectra, 1640cm^{-1} is the only one to not appear in the graph. The peak at 620cm^{-1} is the only one to appear for the 1st principal component but does not appear for the 2nd principal component. However, the peaks at about 800 and 1065cm^{-1} appear for both principal components.

As for the training set for SO, figure 48 left (pg. 67) shows the Raman spectra that displays nearly the same prominent peaks shown for the spectra of PA, peaks about 630, 800, 1075 and 1645cm^{-1}. The serial method displayed in figure 47 right (pg. 67) resembles the results of those for PA with two noticeable exceptions. The scores corresponding to the 630cm^{-1} peak are located mostly in quadrant I along with the scores for the 800cm^{-1} peak and the scores for the 1075cm^{-1} are in quadrant IV with the scores for 1645cm^{-1}. Figure 49 (pg. 68) displays the results for the wavenumber method and shows that the peaks about 630, 1075 and 1645cm^{-1} all only appear for the 1st principal component and not the 2nd. The peak about 800cm^{-1} is the only one to appear for both principal components.

The twelve-hour growth of PA was analyzed by PCA, and the results are displayed in figures 50 and 51 (pg. 68 and 69). The Raman spectra for the growth (figure 50 left, pg. 68) show the exact same prominently shown peaks as the training set for PA; 620, 800, 1065 and 1640cm^{-1}. But in figure 50 right (pg. 68) which displays the scores of the peak finding method resembles an inverted version of the scores graph for the training set. This is because as the concentration of bacteria was diluted from series to series for the training set, the growth has the concentration increasing as the bacteria reproduced. So that scores corresponding to invariant peaks of the spectra are now located in quadrant I & III and the variant scores are in quadrant II & IV. The scores that correspond to the peak at 1120cm^{-1} i.e., glucose, appear in quadrant II. This is because it is decreasing during the growth as bacteria consume glucose. These scores appear closer to the origin for the PCA analysis of the training sets because the glucose concentration is constant. In the spectral method (figure 51, pg. 69), once again the invariant peaks 620 and

1640cm^{-1} do not appear for either principal component, the 800cm^{-1} peak only appears in the 1st principal component and the 1065cm^{-1} appears in both.

Figures 52 to 54 (pg. 69 – 70) shows the variation of peak intensities for the PA growth and reinforces the interpretation of the peaking finding method of PCA. Raman features with little to no variation have their corresponding PCA scores found in quadrant I and small peak intensity scores located near the origin; increasing Raman features have their scores in quadrant IV and decreasing peak scores located in quadrant II. Although, the identity of the chemical species that contribute to these Raman features has not been completely established there are some that we have provided evidence of their contribution, e.g., glucose. And the complex composition of the LB media, i.e., peptides and saccharides, makes the process of accounting for every chemical species' contribution to the Raman spectrum impossible in detail, general categories like lipid, protein and carbohydrate are likely within reach. However, PCA can be used to determine the change of one chemical species relative to another. In figure 50 right (pg. 68), scores for the Raman feature about 799cm^{-1} and 1065cm^{-1} are in quadrant IV and the regions on either side of these peaks have their scores located in quadrant II. This means that while 799cm^{-1} and 1065cm^{-1} are increasing as the regions of the spectrum at about 700cm^{-1} and 900cm^{-1} are decreasing, in terms of a chemical equation this can be written as below.

$$A + B \rightarrow AB \qquad [8]$$

Where AB is the chemical species that contributes to the 799cm^{-1} and 1065cm^{-1} peaks and A and B are the species that can contribute to the 700cm^{-1} and 900cm^{-1} regions, the identity of substances A and B are unknown. Another possibility is that this is representative of a decomposition reaction with A and B both contributing 799cm^{-1} and 1065cm^{-1} peaks. But since

the 1065cm^{-1} peak is believed to be associated with the bacterial membrane and is increasing monotonically in mass as the culture grows then equation 8 will fit this model.

Another reaction that is detectable in the peak-finding method of PCA is the conversion of glucose to carbon dioxide and water represented by equation 9. A previous chapter of this dissertation has established that glucose contributes to the 1120cm^{-1} peak and other research has attributed aqueous carbon dioxide to the Raman features at 1274cm^{-1} and 1383cm^{-1}.[49] Figure 55 shows the PCA scores for water (1636cm^{-1}), glucose (1120cm^{-1}) and carbon dioxide (1274cm^{-1} and 1383cm^{-1}).

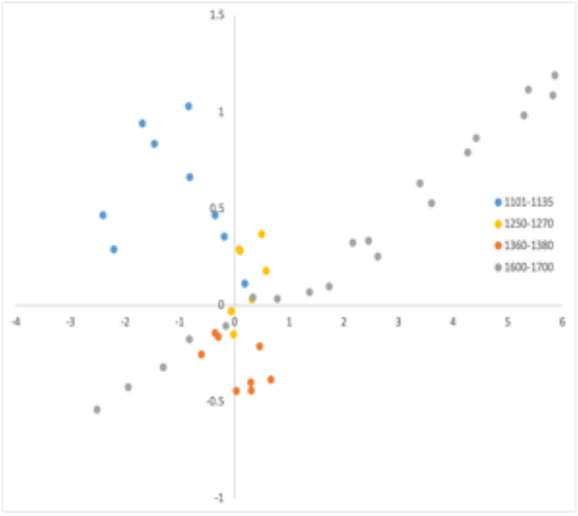

Figure 55: The PCA scores for the peak-finding method displaying the peaks at 1120cm^{-1}, 1274cm^{-1}, 1383cm^{-1} and 1636cm^{-1}.

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6H_2O + 6CO_2 \quad [9]$$

Following the same pattern of interpreting the PCA scores as the model series, we can see that most of the scores for the 1120cm^{-1} peak are in quadrant II meaning that it is decreasing as compared to the 1636cm^{-1} peak which remains constant. And most of the scores for the 1384cm^{-1} peak are in quadrant IV suggesting that it is increasing, however the 1274cm^{-1} remains constant

with its scores in the same quadrant as the water peak at 1636cm^{-1}.

6. Conclusion and Future Work

Experimentation has determined that Binary Spectronephelometry is a very useful tool for the analysis of bacteria cultures in the same way that in vivo hematocrit monitoring has been useful for detecting plasma/red blood cell volume. By detecting for EE and IE scattered light data for the population of bacteria cell can be collected while simultaneously keep track of the information for the chemical environment. This research also shows the sensitivity of the collected data by displaying the effect of the EE and/or IE signals when either the bacteria growth is disrupted by UV light, or the sample is contaminated by bleach. BSN is conducted against a zero background, unlike analysis by turbidity which uses the ratio of transmitted and incident light, e.g., UV-Vis spectroscopy. There are however limitations to BSN analysis, when used on bacteria cultures in media containing copper salts the scattered signals are absorbed limiting the amount of data collected. This is a previously unknown challenge for the development of BSN for the analysis of bioreactors that grow bacteria like *S. oneidensis* that reduce metals in wastewater for metabolization and are a subject of study for their use in microbial fuel cells. Future research will have to be conducted to find a potential work around or solution for this problem.

Chemometric analysis shows to be a useful tool for extrapolating data for the Raman spectrum of samples that is collected along with the volume fraction of bacteria. Simple integration of Raman features helps the user determine concentration of chemical species in the sample which can relay information on the rate of consumption of nutrients by the bacteria. Glucose was used in

this research as an example of a chemical species that can be tracked in real-time during bacteria growth but in order for BSN to be an effective tool for this purpose all possible Raman features associated with glucose will have to be determined.

Principal component analysis is a useful mathematical tool for associating chemical species with their Raman features by emphasizing peaks of the spectrum with the greatest variation while changing the concentration of a particular chemical. PCA was used in this study to help identify a 1377cm^{-1} peak in the Raman spectrum as increasing along with the concentration of glucose. This research also used PCA to emphasize peaks that are varying as the concentration of bacteria changes, helping to confirm that the 800 and 1065cm^{-1} peaks are associated with the bacteria cells. Future research on legacy media will have to continue to identify the other components of e.g., LB media with Raman features to help determine which nutrients are consumed and at what rate of consumption during bacteria growth.

Binary spectronephelometry has been used in this research study to track the diffusion of bacteria cells through the media. There is a difference in the appearance of growth curve shapes between PA and EC that maybe explained by their diffusion coefficients and growth rates. Despite EC having a greater growth rate, doubling time is 20min, than PA which has a doubling time of 30min; PA has a slightly greater diffusion coefficient than EC. This is a clear indicator that the growth of the bacteria is not the driving force for their diffusion, but that the motile cells are diffusing on their own. In the future experiments could be performed on immotile bacteria to determine whether BSN can be used to find their diffusion coefficients and compare them to those of motile bacteria.

7. Appendices

7.1. Appendix A: Procedure for 101-7 Baseline Correction of Raman Spectra

To produce the 101-7 baseline corrected spectra first sum up the signals of every 20-minute series of scans for each of the wavenumbers from 901cm^{-1} to 1901cm^{-1} and transferring that data with their corresponding wavenumbers to an Origin worksheet. Then three new columns were created in the worksheet. Highlight the first column of the signal data by clicking the top of the column, from the analysis menu go to the signal processing option then smooth and select open dialog. In the signal-processing window select adjacent averaging from the method drop down menu. Type in 101 in the points of window field and select the first newly created column as the output by clicking on the adjacent arrow and selecting the column. This will automatically place the results of the 101-point averaging in the first new column. Highlight the second new column and press control Q on the computer keyboard; this brings up the set values window. Then subtract the 101-point averaging from the original data, example Col(B)-Col(Y). With the second new column highlighted perform another smoothing with a 7-point averaging by entering 7 in the points of window field and selecting the third new column as the output. This will be the final result of the 101-7 corrections and was repeated for each set of data then were plotted against the wavenumbers to create Figure 6 right.

7.2. Appendix B: MatLab Procedure for Principal Component Analysis

Launch MatLab R2017a, note that previous versions of MatLab use the princomp command instead of the pca command given in this instruction. From a Microsoft Excel spreadsheet, the spectra were cropped from 600cm^{-1} to 1900cm^{-1} and every other row is deleted to reduce the number of data points. Import the wavenumbers and signals as two different matrices from the

Excel spreadsheet by copy and paste the data into the command window. To do this, in the MatLab command window enter the following:

$$>> wavenumbers = [n_1$$
$$n_2$$
$$\vdots$$
$$n_i];$$

This creates an *i* numbered row and one column matrix called wavenumbers, the same was done with the signals by entering:

$$>> counts = [x_{11}\ x_{12} \cdots\ x_{1j}$$
$$x_{21}\ x_{22}\ \cdots\ x_{2j}$$
$$\vdots$$
$$x_{i1}\ x_{i2}\ \cdots\ x_{ij}];$$

This is the counts matrix with *i* rows and *j* columns that contains the signals. Create an array with the labels for the bacteria dilutions by entering:

$$>> s = \{'100\%\ bacteria'\text{,}'50\%\ bacteria'\text{,}'25\%\ bacteria'\text{,}'12\%\ bacteria' \ldots$$
$$'6\%\ bacteria'\text{,}'0\%\ bacteria'\};$$

Then convert the array to a character array called series by entering:

$$>> series = char(s);$$

To ensure that the matrices and array are structured correctly a graph of the Raman spectra was produced using them (Figure 54). The commands to create the graph are:

$$>> figure()$$

$$>> plot(wavenumbers, counts)$$

>> xlabel('Raman Shift (1/cm)')

>> ylabel('Counts')

>> legend(series)

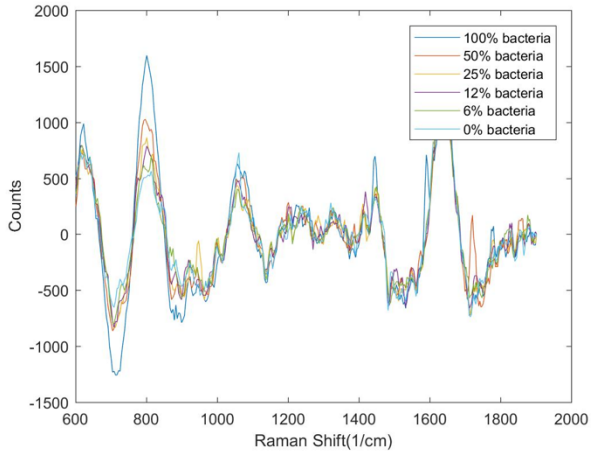

Figure 56: Raman spectra of the bacteria dilution from the training set created from MatLab.

Before performing a PCA on the data, the weighted variance of the counts was calculated by taking the inverse of the variance and defining it as a new matrix called w.

>> w = 1./var(counts);

Then we perform the PCA on the counts matrix with the following command.

>> [wcoeff, score, latent, tsquared, explained] = pca(counts, 'VariableWeights', w);

This returns the five outputs as a set of matrices. The wcoeff contains the coefficients for the principal components. Scores are contained with the score matrix and corresponds to a wavenumber with the same index, that is how they are identified to create the score graph e.g. Figure 55. Latent matrix contains the eigenvalues for the principal components and the tsquared is calculated using Hotelling's T^2 statistics and shows the data with the greatest variance. And the explained matrix shows the percent that each principal component explains the variance of the data. Next we transform the coefficients for the principal components to be orthogonal.

*>> coefforth = inv(diag(std(counts)))*wcoeff;*

And then we confirm that the coefficients are in fact orthogonal by multiplying the compliment of coefforth with the coefforth itself and creating the identity matrix (I).

*>> I = coefforth'*coefforth;*

When the first three rows and columns of the identity matrix is displayed the following result should be returned.

>> I(1:3, 1:3)

ans =

1.0000 0.0000 -0.0000

0.0000 1.0000 -0.0000

-0.0000 -0.0000 1.0000

The following commands were used to create the scores graph (Figure 55) with respect to the first two principal components and the gname command allows the user to select points to display that score's numbered index in the matrix.

>> figure()

>> plot(score(: ,1), score(: ,2), '>')

>> xlabel('1st Principal Component')

>> ylabel('2nd Principal Component')

>> gname

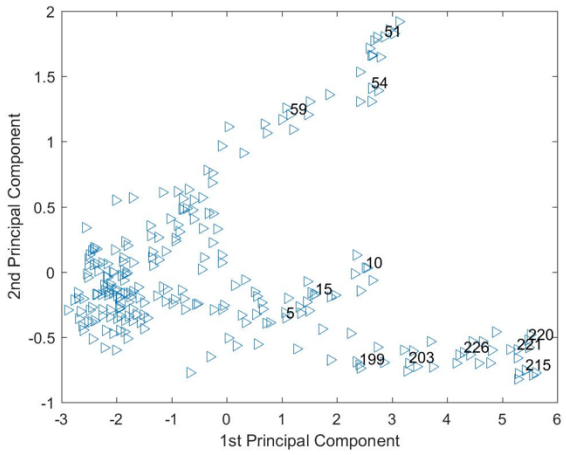

Figure 57: An example of the scores for the training set with respect to the 1st and 2nd principal component with their labeled indices.

The wavenumbers that correspond to the indices of the selected data points are given with the following commands.

```
>> selected = [5 10 15 51 54 59 199 203 215 220 221 226];
>> wavenumbers(selected, : )
ans =
        1.0e+03 *
          0.6012
          0.6403
          0.7720
          0.8094
          1.0499
          1.0769
          1.4492
          1.6129
```

1.6444

1.6753

The bar graph (Figure 56) that shows the percent explained for the principal components with the following.

>> figure()

>> bar(explained)

>> xlabel('Principal Component')

>> ylabel('Variance Explained (%)')

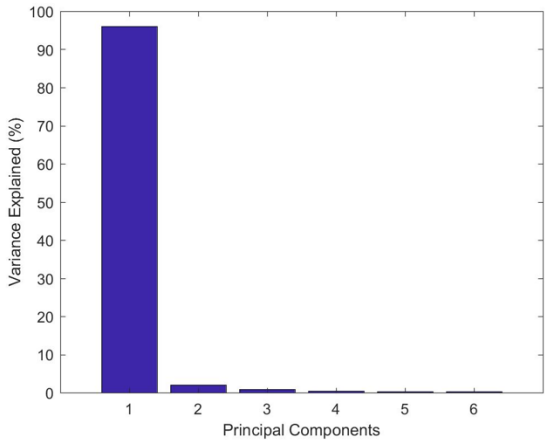

Figure 58: The percent variance explained for the principal components.

And finally, to create the graph that shows the vectors for the variables as they are related to the first two principal components and the scores (Figure 57).

>> biplot(coefforth(: , 1:2), 'scores', score(: , 1:2), 'varlabels', series);

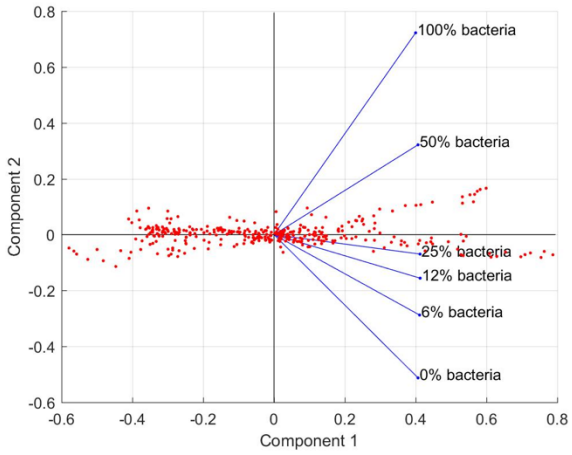

Figure 59: An example of the vectors of the variables plotted in relation to the scores for the first two principal components.

7.3. Appendix C: Published Paper

Coupled Turbidity and Spectroscopy Problems: A Simple Algorithm for Volumetric Analysis of Optically Thin or Dilute, In Vitro Bacterial Cultures in Various Media

Steven Ortiz, Richard T. McDonough, Paul Dent, Jerry Goodisman, and Joseph Chaiken

Abstract

An approach binary spectronephelometry (BSN) to perform real-time simultaneous noninvasive in situ physical and chemical analysis of bacterial cultures in fluid media is described. We choose to characterize cultures of *Escherichia coli* (NC), *Pseudomonas aeruginosa* (PA), and *Shewanella oneidensis* (SO) in the specific case of complex media whose Raman spectrum cannot be unambiguously assigned. Nevertheless, organism number density and a measure of the chemical makeup of the fluid medium can be monitored noninvasively, simultaneously, and

continuously, despite changing turbidity and medium chemistry. The method involves irradiating a culture in fluid medium in an appropriate vessel (in this case a standard 1cm cuvette) using a near infrared laser and collecting all the backscattered light from the cuvette, i.e., the Rayleigh–Mie line and the inelastically emitted light which includes unresolved Raman scattered light and fluorescence. Complex ‘‘legacy’’ media contain materials of biological origin whose chemical composition cannot be fully delineated. We independently calibrate this approach to a commonly used reference, optical density at 600 nm (OD600) for characterizing the number density of organisms. We suggest that the total inelastically emitted light could be a measure of the chemical state of a biologically based medium, e.g., lysogeny broth (LB). This approach may be useful in a broad range of basic and applied studies and enterprises that utilize bacterial cultures in any medium or container that permits optical probing in the single scattering limit.

Introduction

From basic research in fields as disparate as evolutionary biology or the mechanisms of photosynthesis, to applied research on subjects from biofuel production to nuclear waste remediation, there is general interest in improving our basic understanding of bacteria and bacterial cultures.[1–10] In a variety of important situations, bacterial cultures in fluid media are optically thin or optically dilute systems such that light can be propagated into a culture and subsequently detected outside, all in the single scattering regime. This paper is about obtaining quantitative information from such mildly turbid systems, noninvasively and continuously.
In this limit nearly all collected light shown in Fig. 1 has endured essentially no more than two elastic scattering events involving either (1) the incoming probing light, (2) Rayleigh–Mie scattered light, or (3) inelastically scattered light as it propagates from where it was created to

where it is detected. For any given probing/collection geometry this implies an upper limit to the number density of bacteria and/or the physical thickness of the probed system, above which the single scattering requirement is not met.

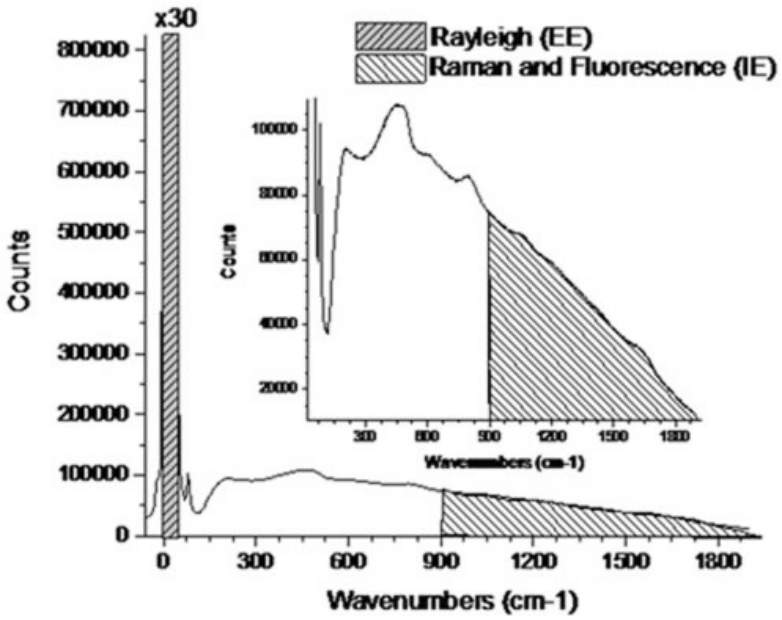

Figure 1. Using a 785 nm laser, an example of a spectrum taken with 200 ms frame of Andor CCD for a NC culture suspended in LB. The range used to calculate IE was ~ 900 to 1902 cm^{-1} and the range used to calculate EE was -5 to +46 cm^{-1} which are highlighted.

We report here an approach that allows spectroscopic and physical scattering probing of such mildly turbid systems with a single light source, thereby noninvasively, simultaneously and quantitatively monitoring the number density of bacteria and the chemical state of the medium, without physical sampling. As shown in Fig. 1, a single light source produces two fundamentally different types of remitted/detectable light. We detect some light that is physically scattered, as in nephelometry, as well as light that contains spectroscopic information, and because in the simplest limit we can consider the sample to contain only two phases, a particulate/scattering phase and a non- turbid fluid phase, we call our approach and algorithm binary

spectronephelometry (BSN).

There are better methods for number density measurements, e.g., flow cytometry and medium composition and measuring the effects of metabolism like HPLC,[11–15] but these require physical sampling and cannot employed simultaneously to the same sample nondestructively and continuously. They do not allow real-time, simultaneous recording of properties as a culture develops. Real-time continuous monitoring of organisms in their natural habitat could allow high-throughput measurements of the effects of many kinds of stimuli, e.g., added chemicals and gases, temperature, light, or pressure, relative to baseline behavior, as we will show in another paper.

Problems in Characterizing and Monitoring Bacterial Cultures

Conventional bacteriology practice[16–18] often relies on measuring the apparent absorbance or optical density of a culture at 600nm (OD600) as a quick and convenient method to establish number density of organisms at the outset of an experiment and possibly later as well. This practice began with Koch[19] and others[20,21] and was put on a quantitative basis by establishing an empirical correlation between the dry weight of bacteria recovered from a culture and the OD600 that could be measured with newly available ultraviolet–visible (UV–Vis) absorption spectrometers. Substituting OD600 measurements for dry weight measurements improved the throughput and quality of experimentation and thus accelerated the rate of discovery. The technologies of spectroscopic probing have improved steadily over the last forty years to the point where, for bacterial cultures in fluid media, it is possible to improve on OD600 as a basis for estimating number density, culture viability, and phenotypic expression while improving the reproducibility of experimental conditions.

To appreciate the improvement possible note that the use of OD600 by itself cannot differentiate between scattering losses due to bacteria, scattering/absorption losses due to debris/particulates in the culture and on the cuvette walls, and absorption by chemical species or films. As chemical species are consumed or produced during the lifespan of the culture, absorption changes cannot be separated from variation in scattering losses, i.e., turbidity changes as the number density of organisms and/or their cellular size varies. Thus, by itself OD600 cannot provide chemical information about the medium. Furthermore, OD600 is susceptible to a small but systematic error if a film is deposited on the surfaces of the container, e.g., the cuvette, as we have shown elsewhere.[22] Note that a film/ debris that causes a change in apparent absorbance of 0.01 AU corresponds to a 3% discrepancy in organism apparent number density per surface, and there must be at least two surfaces in probing the absorbance of a cuvette and its contents.

In choosing a course to improve upon OD600, there is considerable latitude depending on the specific goals of an experiment or enterprise. Our general goal is to provide a reproducible, convenient, noninvasive method for (1) establishing the initial number density of organisms and concentrations of nutrients and (2) allowing continuous, noninvasive monitoring of number density increase and measures of metabolic activity, with no physical sampling. Then, more experiments can be performed on the same culture, as the life cycle of the culture plays out. By establishing a quantitative connection with OD600, we establish full backward compatibility since we also establish a connection with dry weight. We note that literally all samples probed with a laser produce a remitted light spectrum containing broadband fluorescence and weaker Raman emission like that shown in Fig. 1. The emission near 0cm^{-1} Raman shift is comprised of both Rayleigh and Mie scattered light and is termed elastic emission (EE). In our present experimental configuration, it is measured by integrating from -5 to +46 cm^{-1}. We refer to the

inelastic emission (IE) that in our present experimental configuration is measured by integrating Stokes emission from ~900 to 1802 cm^{-1}. The integration range chosen for the IE maximizes the IE collected while minimizing the amount of EE contamination included in the IE. We note that either kind of emission presents specific possibilities for revealing information about the bacterial system using BSN. Variations of Fig. 1 have been shown in earlier publications where the algorithm was described in different contexts.[23,24]

While attempting to improve our capacity to monitor bacterial cultures, the spectroscopy of traditional or ‘‘legacy’’ media like lysogeny broth (LB) medium must be considered, if only to provide some backward compatibility in comparing future results with the existing extensive literature. Indeed, the ambiguity of the composition of LB leads to ambiguity in the spectroscopy and the interpret- ability of the behavior of the various contributors to the IE. This situation prevails in spite of a strong shift in research and applied bacteriology[25–30] toward use of so-called ‘‘chemically defined’’ media (CD), of which minimal medium (MM) is an example.

Media such as MM are isotonic fluids that contain single inorganic sources of the essential elements needed to sup- port culture health as well as a single well-defined source of molecular carbon, e.g., glucose or lactate. CD media pro- vide improved control and reproducibility of the cultures.[25–30] LB media contain a broad range of materials, such as proteins, carbohydrates, and lipids as relatively unrefined biological sources, e.g., yeast extract or trypsinized milk. They are thus complex mixtures that defy the detailed analysis that is possible with CD media. In this work we show how to analyze for LB medium taken as a composite material. We also show that it is possible to achieve the single-scattering regime in bacterial cultures with a complicated fluid phase, e.g., filtered or unfiltered LB medium and LB can produce more particles over time or settling can occur. We have previously shown[23] that this can be discerned by BSN monitoring.

What chemical information can be obtained depends on the spectroscopic assignment of the sources of the various kinds of emission, and the propagation of light in the sample. Minimal media, despite comprising only a single molecular source of carbon, e.g., glucose or lactate, along with inorganic nitrogen, phosphorus, and sulfur can still produce complicated spectroscopic behavior. Absorbance can be as important as elastic scattering because it affects both EE and IE. Early experiments in our lab have shown that inclusion of transition metals in MM, e.g., Wolfe's trace metal mixture with copper,[31,32] that absorbs near infrared emission, can modulate the detected IE in complicated ways. Transition metals can complex to the nitrogen, sulfur and phosphorous donors, shifting absorption/emission profiles and further complicating the IE.

The IE includes fluorescence and/or phosphorescence as well as Raman scattering. In this paper we concentrate on using the IE as a composite emission. i.e., unresolved fluorescence–phosphorescence and Raman, to show that it should be possible to implement BSN monitoring using only a single laser, a single filter and two single channel detectors. This will facilitate installing BSN systems in working bioreactors, reducing the power and spatial footprints and overall complexity of continuously monitoring bacterial and similar materials. A cuvette is used in this study only because it is easily reproduced by anyone interested in exploring this approach for their application. We will present a detailed analysis of the Raman spectra in a separate publication as well as examples of monitoring the time dependent behavior of these bacterial cultures.

Either fluorescence or phosphorescence is usually more intense than any spontaneous Raman scattered emission component. We have previously[23] surveyed various kinds of biological materials that are potential sources of Raman scattering and fluorescence/or phosphorescence

under NIR excitation[33] and that display photobleaching as is common for materials in biological media.[34] In vitro spectra of authentic samples of materials and constructs, e.g., tissue phantoms, can provide inferences or support for a hypothetical assignment. While specific assignments are rarely unequivocal it might be useful to discern if there are materials in e.g., LB medium that are either consumed or transformed from an emissive material into a non- emissive material or vice versa. We make this suggestion because of the observation of NIR induced autobleaching[33,34] and also of that inflammation in injured spinal cord tissue in vivo or ex vivo is accompanied by increased autofluorescence.[35–40]

BSN Algorithm

The system is closed except for possible evaporation/condensation loss/gain of water, or loss of carbon as gaseous CO or CO_2. With those exceptions, the atoms that were present initially in a culture sample, or were added during the culture lifetime, remain in the culture. Since there may be a metabolic redistribution of the chemical bonds present, we might expect the Raman and/or fluorescence spectra to change, perhaps in complicated ways that cannot be fully accounted for using the BSN approach. For a complex medium, the approach we describe provides a sensitive and potentially quantitative probe of changes in the medium optical or spectroscopic properties due to culture growth and metabolism.

In collecting the EE, we sometimes inadvertently include some light from reflections from the liquid sample surface or the bottom or other surfaces of the cuvette. This artifact must be managed as discussed below in order to produce reliable inferences concerning the behavior of the culture. Similar management is required for the IE because the IE can be contaminated by fluorescence or Raman from the cuvette produced by stray incident light. We closely monitored

the presence of foam or bubbles on the surface of the culture and it was not challenging for these experiments. Nevertheless, their presence must be considered in applying BSN to, e.g., bacteria that generate gasses metabolically.

For a three-phase system like those considered previously,[23] there are two independent intensive parameters, taken as the volume fractions of phases 1 and 2, i.e., $\phi_1$and ϕ_2. (Obviously, $\phi_1 + \phi_2 + \phi_3 = 1$). In the present application, phases 1 and 2 will be NC and LB. Two spectroscopic parameters, IE and EE, are measured (see above). If the sample volume is fixed, and assuming we are in the single- scattering regime, these two parameters are linearly dependent on ϕ_{LB} and ϕ_{NC}. Consequently, ϕ_{LC} and ϕ_{NC} are linear functions of IE and EE. This linear dependence is defined by six constants, a through f.

$$\phi_{NC} = a + bIE + cEE \quad (1)$$

$$\phi_{LB} = d + eIE + fEE \quad (2)$$

Once the values of these six parameters are established, a measurement of IE and EE becomes a measurement of the three volume fractions.

In general, we can determine the values of two independent intensive variables from measured EE and IE. In the case of a bacterial culture, which is a two-phase system, one of these parameters is the volume fraction of bacteria ϕ_1. The volume fraction of medium is just 1 – ϕ_1. The second intensive parameter, ϕ_2, may be the concentration of some biologically produced material like tryptone or peptone, or of specific chemicals/molecules added as sources of carbon, sulfur, phosphorus or nitrogen, like glucose.

In this paper we present and analyze results of probing an open-top culture with a single laser

and show the advantages over measuring OD600. In general, an absorption measurement has smaller signal-to-noise[41,42] than an emission measurement, sometimes by as much as three orders of magnitude. The improved enabling technologies of emission spectroscopy make it relatively easy to use a single color laser to induce both elastic and inelastic emission essentially simultaneously. Also, it is likely possible to achieve higher performance in smaller, less expensive and more highly integrated packages, so that the approach could enjoy easy adoption in production facilities. Separating emission into elastic and inelastic contributions is easily accomplished using commercial off the shelf filters and optical components. The total remitted inelastic light (IE) is comprised of two components, narrow band, i.e., Raman scattered light and broadband, or fluorescence and/or phosphorescence. Since there is no background emission, it is possible to measure the IE to better than ± 0.01% precision, so small changes in IE can be easily discerned. Since Raman features can be localized to spectral intervals as narrow as ≈10–20 cm^{-1}, very specific and small changes can be observed even with the background produced by complicated media. Since spontaneous Raman scattering is usually a parts per thousand analytical technique,[42] this can be a limitation on the proposed approach. If the chemical information is contained in the fluorescence, the sensitivity can be better in which less than parts per million is common place. Thus, complex media can be probed for changes in chemical state. We show that for the complex medium LB, we can con- struct an empirical calibration system that enables quantitative monitoring of chemical and physical changes of the bacteria and the medium during the life of a culture. However, it may prove difficult to assign the source(s) of the changes. For CD media like MM on the other hand[25–30] spontaneous Raman spectroscopy provides a simple, direct approach to establish the chemical state of the medium without physical sampling. In this paper we show sufficient Raman spectra to demonstrate that such features can

serve as IE for practical purposes. Results for NC are shown in the main body of this paper but corresponding results and experimental details for PA and OE are available in the Supplemental Materials. Spectra relevant to MM calibration are also available. After presenting experimental details, we show results from application of the BSN algorithm with typical calibration results and tests of internal and external consistency. We suggest that if such measurements can be done with equal accuracy and precision it will be possible to characterize other processes in bacterial cultures.

Materials and Methods

BL21 Gold (DE3) Escherichia coli (NC) with a pET27b(+) sapB vector insert was obtained from Dr. R.P. Doyle (Syracuse University) grown in LB with 5 μL of kanamycin (35 mg/mL) (IBI Scientific). Phosphate-buffered saline (PBS) (1 x solution) (ICN Biomedical Inc.), Difco M9 Minimal Salts and NaCl (Fisher Scientific), Bacto Tryptone (Becton, Dickson, and Company), and yeast extract (aMReSCO) were used without further purification.

Films grown on quartz utilized Hellma Analytics High Precision 1cm cuvettes made of Quartz Suprasil (#101- 10-40). The surface quality, i.e., roughness of the quartz cuvettes or substrates, is assumed to be λ/10 across the entire UV–NIR range based on the specifications of comparable quartz discs produced by Heraeus, the entity that we believe produces the quartz for both Hellma cuvettes and Starnacells. High-quality fused silica is the material of choice in order to minimize fluorescence from the cuvette produced by stray incident light although we have used inexpensive polystyrene successfully too.

Nearly all OD600 spectra were obtained using a PerkinElmer Lambda 950 UV–Vis spectrophotometer but some results were obtained using an Agilent Cary 60 UV– Vis. Spectral

scans were set to a 2 nm slit width, 1.0 s integration time, and 1.0 nm data interval. Before scanning the sample, the spectrometer was normalized (‘‘auto-zeroed’’) with a cuvette containing 3 mL of LB over the same wave- length range.

Using a flexible temperature probe attachment connected to a Vernier LabQuest 2 Logger Pro, temperature data were collected. We employ a 785 nm, continuous wave external cavity laser (Process Instruments) that can generate a maximum of 450 mW at the location of the sample in a bandwidth of $1.5cm^{-1}$ within a multimode spatial distribution. We employ ~180 mW of laser power in a roughly square shaped spatial distribution that is focused to a spot about 0.125 mm wide inside the sample.

The sample is housed in a custom made temperature controlled base as shown in Figs. 2a and 2b. The base height is 3.50 cm and the cuvette height is 4.50 cm. These dimensions were used to ensure the sample is covered by the walls of the base to allow homogeneity in heating and or cooling of the sample, but still allow easy removal of the cuvette from the base. To set up a system that produces valid measurements across a range of variable turbidity, under conditions of relatively low turbidity, which restricts the light which can be collected to a smaller spatial region, the focus of the laser was preset to begin ~4mm below the surface of the liquid. The depth of focus was such that all focused incident light was contained at a location ~1 ± 0.5 cm below the top surface of the culture, or ~2 ± 1 cm from the bottom of the culture. A magnetic stirrer was added to the setup in order to maintain constant stirring of samples during the experiments. To insure that the stir bar did not directly scatter light itself, the stirrer was place off center of the cuvette so that the stir bar was actually spinning parallel to one of its vertical sides, i.e., around an axis normal to a cuvette side.

The following times/CCD temperature were chosen to allow data collection to occur in the shot

noise limit. The data for the training set were collected on the Andor CCD Camera cooled to ~55°C, after dispersion using a Kaiser Optical f/2.0 Holospec spectrograph, had an acquisition time of 5 min. with 1500 accumulations of 200 ms frames for each spectrum. The experimental setup is shown in Figs. 2a and 2b.[23] Unlike standard Raman apparatus, this system is not intended to completely suppress the Rayleigh–Mie scattered light, i.e., the EE. The idea is to measure the Rayleigh line without saturation and obtain a Raman spectrum at the same time. Since the available filter set is commercial off the shelf (Semrock) and not being used as designed/intended, detuning the edge filter induces some oscillations in the spectrum at Raman shifts lower than 800 cm^{-1}. Aliquot experiment data had an acquisition time of 30s with 150 accumulations of 200 ms frames for each spectrum, and the counts were then scaled up to match that of training set data. The culture was grown and analyzed over a 24 h period while stirred at 37°C. Addition of 25 mL of medium every 30 min was required to keep the culture at a constant volume, and the focal point of the laser fully submerged.

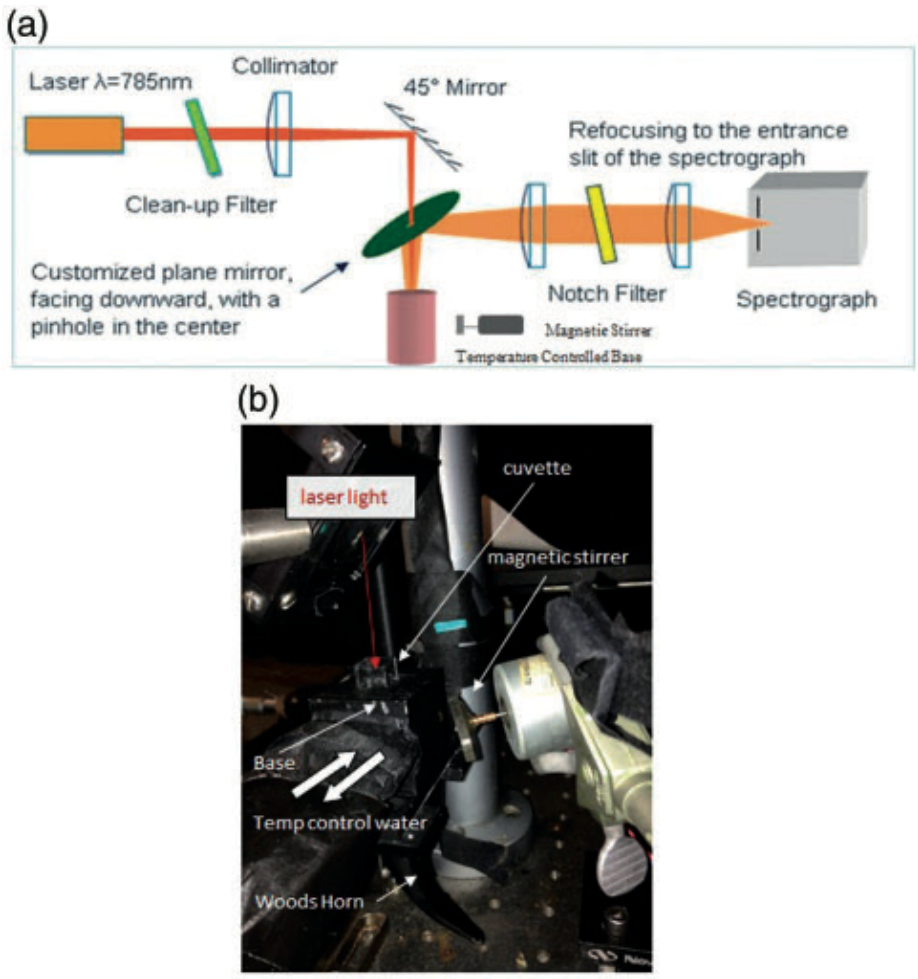

Figure 2. (a) Experimental setup of the 785 nm NIR laser with an added stirrer and temperature controlled base that was used in our apparatus. (b) Close-up photo showing incoming laser light, cuvette, Woods Horn, temperature controlling base with temperature controlled water inlet/outlet. Micro magnetic stir bar is inside cuvette/base and magnetic stirrer/motor (outside as shown) has strong magnet on axle.

The sample preparation of the LB medium contained 10 g/L tryptone, 5 g/L yeast extract, and 10 g/L NaCl at 37°C. The pH of the medium was maintained at 7.0 with NaOH. The solution was stirred to homogeneity, and then diluted to a final volume of 500mL with DI water. The medium was then autoclaved at 121°C for 15 min in a water bath.

The sample preparation for the aliquot experiment consisted of 3 mL of LB with 3μL of kanamycin that was placed into a quartz cuvette with or without a biofilm present.[22] At ~45s into

each 3min kinetic series with a spec time of 30s, 5μL of NC (OD600 = ~1.000 in the same LB the bacteria were grown in), or pure unfiltered LB was added to the cuvette over a 2h period. At the temperature employed, i.e., 37°C the 5μL aliqouts were chosen because they were close to the evaporation loss thus the total volume of the cuvette remained constant.

A 101 and seven-point (101-7) arbitrary baseline correction process that allows narrow peaks on a sloping or broadly ‘‘hump-shaped’’ fluorescence background to be studied, was applied to all raw spectra.[23,43] The 101-7 adjacent averaging was chosen partly by empirical observation after noting how others performed similar corrections.[44–46] From our own studies,[23,43] a 101-7 was shown to produce very reliable quantitative calibration curves as well as spectra that had the most realistic appearance of pure Raman spectra based on subjective judgment/visual inspection. Note that the instrument function for our spectroscopic system is not applied to any of the results shown because it does not facilitate the analysis or the background subtraction process.

The 101-7 procedure is statistically unbiased, i.e., it produces a baseline-corrected spectrum with zero mean without making assumptions about what part of the total IE spectrum has little or no Raman activity. The effect on a spectrum can be expressed analytically aiding further data analysis and algorithm development.[23,43] Note that before statistical/chemometric analysis, raw spectra are usually processed as a set and the standard normal variant trans- form is used for normalization, i.e., zero mean and unit standard deviation. The 101-7 baseline correction achieves zero mean in a single step while improving the dynamic range of the apparent Raman emission and providing a quantitative estimate of the amount of underlying fluorescence.

Results

Empirical Calibration of the BSN Algorithm: Training Sets

Sample preparation for training sets included four vials containing 5mL of LB with 5μL of kanamycin were inoculated with NC and incubated over a 24 h period at 37°C with shaking. The contents of these four vials were then combined into one larger sterile vial. Six stock solutions of PBS and LB were prepared and placed into sterile plastic vials for all trials. Vial one contained 15.00mL of filtered LB. Vial two contained 13.50mL of filtered LB and 1.50mL of filtered PBS. Vial three contained 12.00mL of filtered LB and 3.00mL of filtered PBS. Vial four contained 10.50mL of filtered LB and 4.50mL of filtered PBS. Vial five contained 9.00mL of filtered LB and 6.00mL of filtered PBS. Vial six contained 7.50mL of filtered LB and 7.50mL of filtered PBS. 1.50mL of NC and 1.50mL of respective stock. Serial half dilutions were made of each sample producing a training set containing 36 samples.

Since each sample was a mixture of NC, LB, and PBS, the volume fractions of NC and LB, ϕ_{NC} and ϕ_{LB}, were taken as the two independent intensive parameters, with $\phi_{PBS} = 1 - \phi_{LB} + \phi_{NC}$. Since each sample was made from equal amounts of an LB-PBS mixture with volume fraction of LB v_{LB} and an NC-PBS mixture with volume fraction of NC v_{NC}, the actual volume fractions of LB and NC in the sample are $\phi_{LB} = 0.5\ v_{LB}$ and $\phi_{NC} = 0.5\ v_{NC}$. The measured EE and IE (counts) are plotted in Figs. 3a to 3d versus ϕ_{LB} or ϕ_{NC}.

We collected data in the shot-noise limit, which means the large number of counts for both the raw EE and IE makes the experimental error bars for each datum smaller than the symbols used in the graphs. The deviations from the lines are therefore due to errors in volume measurement and/or perhaps originate in inhomogeneity of the sample with changing concentration. We chose the concentration ranges to avoid nonlinear dependence of counts on concentration (see below). Since one needs to state a tolerance to make such a definition, there is some subjectivity involved. We compared changing NC concentration with constant LB

concentration and changing LB concentration with constant NC concentration (Figs. 3a to 3d). The linear range for NC is determined by the behavior of the EE data and for a 1 cm cuvette and our geometry was shown to extend to at least EE = 2.45×10^8 counts ≈ 0.323 AU OD600 (Fig. 3c).

The training set results were used to find the values of the six parameters in Eq. 1. Then, as will be shown in a subsequent paper, the BSN algorithm, with these parameters, can be used to quantitatively analyze to quantify the time development of NC culture and the depletion of nutrients.

To estimate the importance of the quadratic and combination terms EE^2, IE^2, and EE*IE, which may arise from the turbidity, they were added to Eq. 1. Results are shown in Figs. 4a and 4b. The reduction in residual sum of squares of deviations (RSS) was compared to the increase in number of parameters using the F-test. Inclusion of a quadratic term in EE was not statistically significant at 95% confidence and inclusion of other terms by themselves or in combination was also not statistically significant. Including only the linear terms produces a model that predicts, within experimental error, both the NC and LB concentrations. Plotting calculated versus actual LB concentration produces a line with a slope of 1, an intercept of 0 and with the slope known to better than ±2% precision at 2σ. This implies that, in the range studied, the variation of the IE with changing LB concentration is linear with a slope independent of the NC concentration.

For the training set results in Fig. 3b, we observed that while the intercepts of the linear IE dependence on LB concentration depends strongly on the NC concentration, the slopes do not. Thus, IE is a bilinear function of NC and LB with no cross term. Also, NC is independent of IE. This has important ramifications for the use of Raman spectra to monitor specific components in the fluid phase. In particular, a simple Raman calibration curve using non- turbid solutions can

be obtained for some component such as glucose. Then, within a constant additive factor, it will be valid for measuring glucose concentration for all bacterial volume fractions spanning the linear range of turbidity. The constant additive factor can be obtained from the training set data so long as the training set and the Raman measurements utilize the same optical probing geometry. A complete and general example of this approach is beyond the scope of the present paper and will be presented separately later, but we note that obtaining the additive factor involves using the fractional increase in IE with fractional increase in scattering phase, i.e., Figs. 3b and 3c. We denote the volume fractions of the three liquids in the samples in the training set by ϕ_{LB}, ϕ_{NX} and $\phi_{PBS} = 1 + \phi_{LB} + \phi_{NC}$. The parameters in the BSN algorithm (Eq. 1), determined from the training set, are given in Table I.

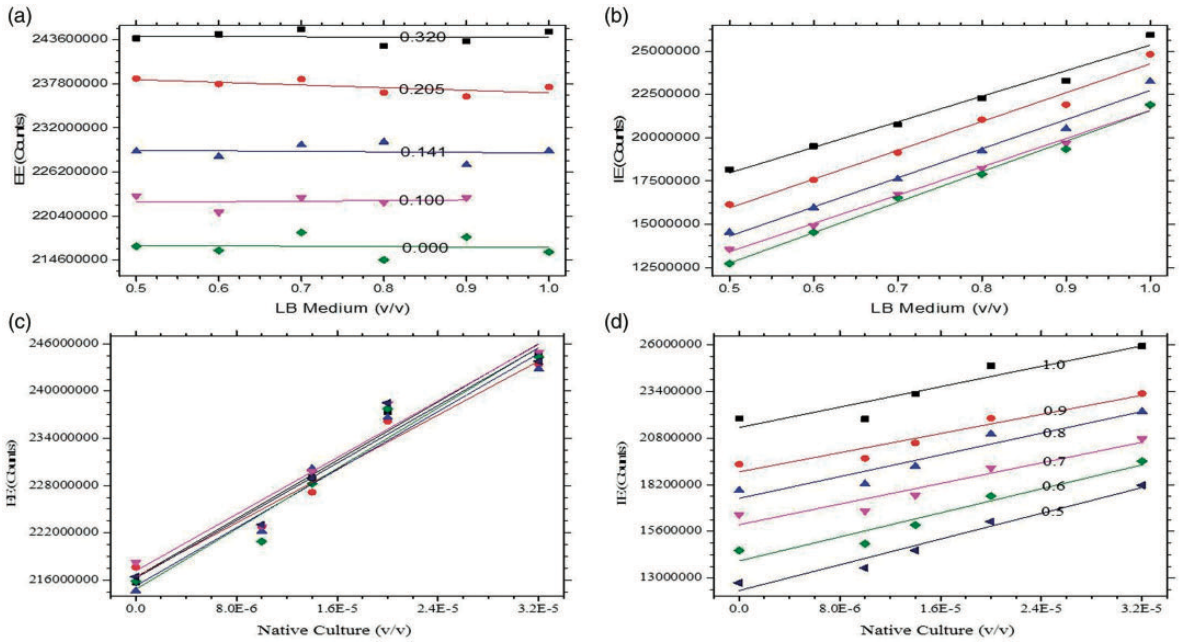

Figure 3. Training set data for both EE and IE for NC and LB. The EE (-5 to +46cm^{-1}) and the IE (900–1902cm^{-1}) are summed up after an average dark current subtraction of each spectrum has been performed. The (a) EE and (b) IE of changing LB at constant NC (v/v): 3.20 E-5 (#), 2.05 E-5 ( ), 1.41 E-5 ( ), 1.00 E-5 ( ), 0.00 ( ). The EE (c) and IE (d) of changing NC (v/v) at

constant LB (v/v): 1.00 ( ), 0.900 ( ), 0.800 ( ), 0.700 ( ), 0.600 ( ), and 0.500 ( ). One outlier at 1.00 E-5 NC (v/v) in 1.00 v/v LB was removed from the training set with a certainty greater than 99% due to the Q test.

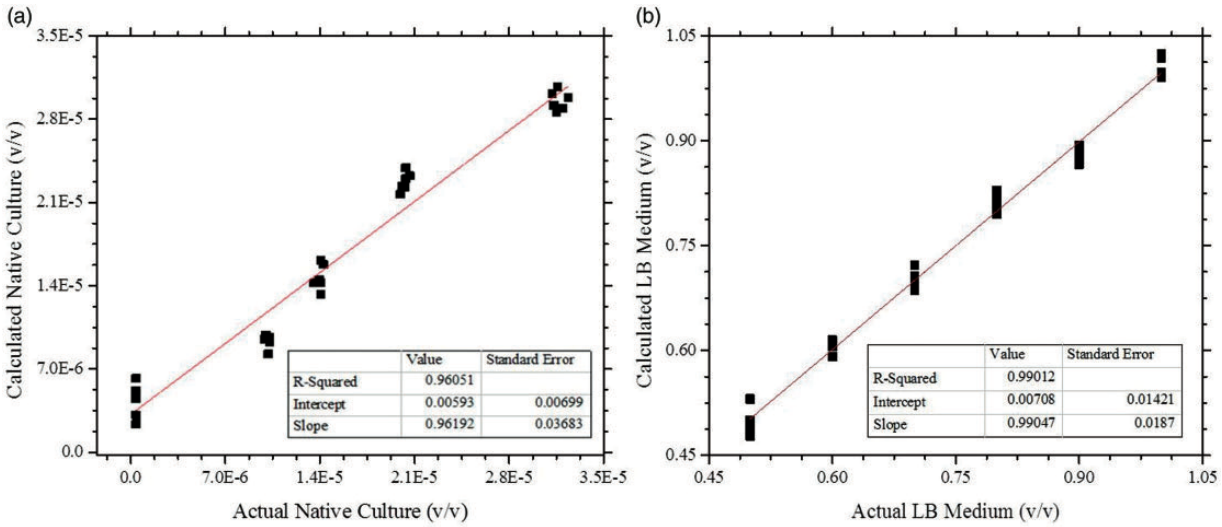

Figure 4. Calculated data for NC and LB generated by the algorithm when applied to the training set data. (a) The volume fraction of NC calculated using BSN without the quadratic terms plotted versus the actual volume fraction of NC calculated from OD 600 measurements. (b) The volume fraction of LB calculated using BSN without the quadratic terms plotted versus the actual volume fraction of LB based on the known dilution factor from the standard recipe.

Application of the BSN Algorithm

All the samples in the training set are mixtures of LB, PBS, and NC. PBS makes only a small contribution to IE and EE, and LB makes only a small contribution to EE. The algorithm allows us to calculate the volume fractions of the three components (ϕ_{LB}, ϕ_{NC} and ϕ_{PBS}) in an unknown mixture from measured EE and IE.

Table I. Parameter values obtained from applying the BSN algorithm to the training set data.

Parameter	**Value**	**Standard error**
a	-2.28×10^{-4}	0.105×10^{-4}
b	1.07×10^{-12}	0.0510×10^{-12}
c	-3.01×10^{-14}	1.55×10^{-13}

d	1.87	0.130
e	-9.70×10^{-9}	0.629×10^{-9}
f	5.86×10^{-8}	0.192×10^{-8}

We note the organic fraction of the LB mixture contributing to the Raman scattering comprises tryptone and yeast extract, i.e., enzymatic protein digests. Because the IE includes a large frequency range, we must group the contributions of tryptone and yeast extract together. Since both substances are mostly composed of amino acids, with average molecular mass of 137.20 g/mol (average over the 20 amino acids, essential and non-essential), we can convert the 15 g/L to an amino acid concentration of (15/137.20) M. Thus, the concentration of amino acids is $[\phi_{LB}][0.109\ \text{M}]$.

Results: Known Artifacts and Internal Consistency

Although we explore the possibility of using the BSN approach in various types of research and production/reactor systems, e.g., closed, open, batch, flow and hybrids, our current apparatus was open requiring maintenance. After dozens of preliminary experiments to provide a basis, we devised the aliquot experiment that we executed and completely repeated to produce the results in Figs. 5a and 5b. We show one full data set because it allows facile anticipation/demonstration of the appropriate BSN response to common manipulations of a culture. The volume of the aliquots was chosen to replace the net loss of culture volume due to apparent evaporation as closely as possible, so that the total volume in the cuvette remained constant (3mL). Aliquots of either NC or LB were added at equal time intervals to fresh, non-bacteria-inoculated, kanamycin-protected LB. If NC contains only bacteria and LB, aliquots of NC should result in a chemistry

concentration (v/v) NC change per aliquot of (2.6 x 10^6 μm^3)/(3 x 10^{12} μm^3) = 0.87 x 10^{-6}. If LB is a homogeneous phase, the LB v/v calculated by BSN should be unchanged from 1.0 with added LB aliquots, if the aliquots replace the exact volumes of that homogeneous phase that leaves by evaporation. These two expectations are indicated by green lines in Figs. 5a and 5b. However, LB medium is demonstrably not homogeneous. First, the concentrations of fluorescence/Raman producing chemicals would be expected to increase over time if only water evaporates. Second, the presence of particulates in fresh or old LB is easily demonstrable. Third, particulates must also be transferred when either NC or LB aliquots are added. And finally, autofluorescence and photobleaching are well-known contributors to IE and its variation over time for NIR irradiation of biological systems in general.[47,48] Furthermore, biofilms contain bacteria, particulates and chemicals from whatever medium is present internally and on the surface.

Biofilms[22] form on surfaces in contact with bacterial cultures, whether in closed, open or flowing systems. Thus we included experiments in which a cuvette had its internal surfaces prepared, i.e., biofilm coated by a culture for 24h before the old culture was poured out and fresh, noninoculated LB was placed in the cuvette to start an aliquot experiment. All these effects are visible in the qualitative response of EE and IE with the added aliquots in the form of the calculated NC and LB volume percentages in Figs. 5a and 5b.

At certain times the calculated volume fractions of NC and/or LB are either negative or greater than one relates (1) to the design of our optical system, (2) to the presence of particulates in LB, and (3) the manner in which the BSN training set model was constructed and calibrated. Furthermore, and not surprising for a manually executed experiment, there are occasional discontinuities, e.g., at about 30min and again near 60min, in the calculated LB and NC when

LB were added with a biofilm present. The presence of an unwanted bubble, which can produce unwanted reflections affecting both the EE and IE, cannot always be anticipated or seen before making a measurement. Each data set contains 150, 5mL aliquots and one aliquot involving a partially plugged/cracked pipettes tip will produce a response unlike the others.

In the construction of the training set, PBS was used to dilute full strength LB, i.e., LB v/v = 1 (by definition) to produce lower values of LB v/v. PBS does not contain particulates whereas LB necessarily contains particulates and in this approach to BSN calibration both, the chemicals in LB and the particulates change together, in constant proportion. But the inherent particulate fraction of LB itself must change over time if only because water can evaporate but particles cannot. This should affect the relative variation in EE and the production/collection of IE due to light propagation effects.

With regard to the design of our optical system, if the top level of the liquid in the cuvette varies by about 10^2 microns or more, the baseline raw EE and IE signal values change. The movement defeats the spatial filtering meant to block unwanted reflections from the top surface mostly affecting the EE. The movement also affects the collection efficiency of emission, i.e., both EE and IE from within the cuvette. If the cuvette liquid surface in the aliquot experiment is shifted from the position used in constructing the training set, the position of 0 NC v/v or 0 LB v/v will be shifted. We also expect a proportional and opposite response of the calculated NC and LB responses reflecting the basic assumption of the model, $\phi_{NC} + \phi_{LB} + \phi_{PBS} = 1$.

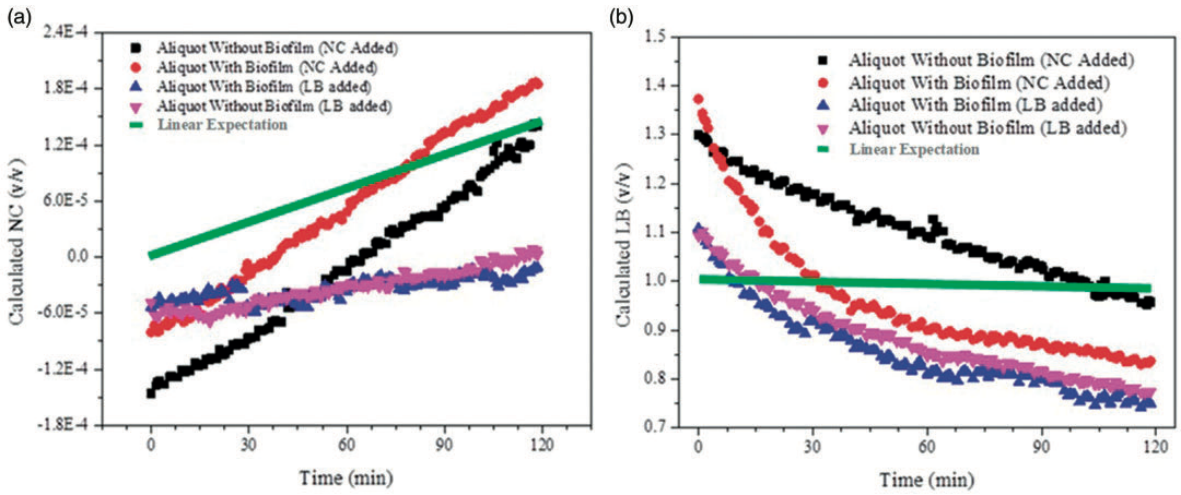

Figure 5. At ~45 s into each 3 min kinetic series with a specified time of 30 s, 5 mL of NC (OD600 = ~1.000 in the same LB the bacteria were grown in) or pure unfiltered LB was added to the cuvette over a 2 h period. At the temperature employed, i.e., 37°C the 5 μL aliqouts were chosen because they were close to the evaporation loss thus the total volume of the cuvette remained constant. (a) Observed, i.e., calculated from IE and EE, NC concentration with timed addition of aliquots of NC (5 μL of NC, OD600 ~ 1.0) from a stock culture. (b) Observed LB concentration with addition of successive aliquots of LB from a stock LB.

As depicted in Figs. 5a and 5b, the BSN-calculated NC response to added E. coli (NC) is linear regardless of whether the cuvette contained a biofilm. This is expected for the biofilm-free case since the aliquots were a small volume compared to the total cuvette volume, and they were delivered at equal time intervals. If the bacteria were sticking to the cuvette surfaces we might also expect a nonlinear effect. In contrast, as shown in a separate paper, a growth curve reflecting the increase in NC concentration obtained by normal reproduction after starting with, e.g., an initially full cuvette containing NC is not linear over time. If the LB associated particulates dominate the production of EE after the bacteria, then the difference between the slope of the calculated value (green line) and observed value is mostly due to the scattering by the particulates.

In all experiments, once begun, the aliquots were added without moving the cuvette. The

apparent constant shift between the average NC concentrations for the two experiments, i.e., with and without biofilm present is due to the shift in cuvette position since the two experiments were executed on separate days. With no biofilm present and NC added, the calculated LB response was also linear, decreasing in a manner consistent with the basic assumption of the model. Quantitatively, the values of LB measured using BSN should be 1.0 if the combined fluorescence and Raman neither changes over time nor does the EE. We observe the calculated LB versus time decreases monotonically and has an average value of somewhat less than ≈ 1.0 over 2 h. There is a significant difference between whether there is a previous biofilm or not. The presence and inevitable production of particulates in LB independent of bacterial activities should cause a decrease in the apparent LB, as observed because the particles occupy volume. Note that there is no constraint that the observed variation in IE and EE be consistent with the overall model adding credibility to the inherent assumptions.

Whether or not a biofilm was present before NC was added, the calculated NC response was linear. However, when a biofilm was present initially the corresponding calculated LB response was not linear and decreased more rapidly than when no biofilm was present for the first 30 min. After the first 30 min, the BSN calculated LB decreased in a manner consistent with the variation with LB and NC aliquots. After the first 30min the LB slopes were all roughly equal. To help associate Raman features with either NC or medium, the variation of the Raman signals with LB and NC proportions is shown in Figs. 6a and 6b. For use with MM cultures and other situations, such an association would be required to implement a BSN training set based on using Raman features as the IE. Relevant results are available in the Supplemental Material for this paper, but we will discuss the effects of turbidity induced spectral dis- tortion[49–51] in constructing Raman based BSN training sets in detail in a separate publication. We used 101-7 baseline- corrected

raw data for 900–1902cm^{-1} to minimize the fluorescence contribution. The accessible Raman shift range for these baseline-corrected spectra reflects the filter arrangement needed to isolate the IE scattering with minimal contribution from EE scattering. Although we shall make some assertions as to possible assignments of Raman features to actual chemical entities, we harbor no delusions that these assignments are proven. We note that, while LB is a very complicated mixture, Raman features at ~1412 cm^{-1} and 1455 cm^{-1} (Fig. 6d) correlate strongly with (C(6)–H_2·COH) and (CH_3) stretches and that usually such moieties do not hydrolyze in aqueous media.[52] The most likely source of such vibrational motions, alkyl chains, are present extensively in both LB and the organisms themselves, i.e., NC. The effect on the Raman spectrum of replacing one source of these Raman photons (NC) with another source (LB) is determined by the different CH_2 and CH_3 content of these sources. Thus the strength of these two features generally increases with increasing LB above 0.6 v/v and constant NC. However, over the whole range in Figs. 6a and 6b and in detail, their change was complicated and different for the two Raman features.

On spectroscopic grounds the feature at ~1060cm^{-1} could come from a ν(C–O–C) asymmetric stretch or the (C–C) stretch, and be associated with the amount of ‘‘biomass’’ as shown in Figs. 6a and 6c and more specifically in Fig. 7 and Figure S1.[52,53] These types of moieties can certainly hydrolyze non-enzymatically, particularly if they are associated with ester linkages which would be depleted in LB. However, in viable bacteria the biomass materials are bacterial membranes, e.g., phospholipids or triglycerides and these are not labile so long as the bacteria are viable.

Graphing the area under the curve (AUC) data of the peaks from Figs. 6a and 6b in Fig. 7, we observe that the strength of the 1060 cm^{-1} feature increases roughly linearly with the proportion

of the culture that is NC, until the turbidity-limited value is reached. Thus this feature largely comprises NC but may have LB contribution, or not, since the noise between spectra may be comparable with the changes induced by varying LB and NC proportions. The situation is better for the 1410 cm^{-1} and the 1455 cm^{-1} features as they too increase linearly in proportion to the amount of LB present in the sample apparently with less noise. The only exception is the 1410 cm^{-1} feature where the lowest concentration is higher due to interference from signal coming from the cuvette.[54] The 101-7 background-corrected Raman spectra, displaying some internal consistency, suggest that changes in the Raman spectra with perturbations caused by the cultures can be rationalized.

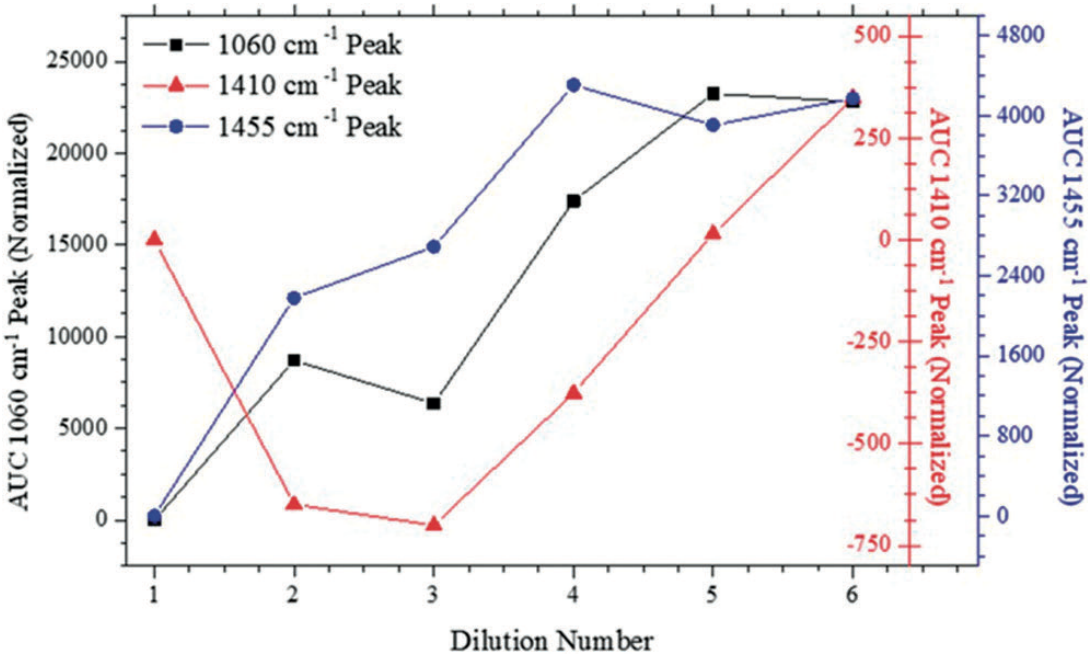

Figure 7. The 1060 cm^{-1} peak increases monotonically with the proportion of NC while the 1410 cm^{-1} and 1455 cm^{-1} peaks increase monotonically with proportion of LB. The dilution number refers to the six serial dilutions used in Figure 6 in which 1–6 represent NC 0.000-0.503 and LB 0.50–1.00, respectively. All data in this figure were normalized for simpler comparison.

Discussion

The purpose of this study was to explore the applicability of modern spectroscopic technology to

the noninvasive, in vivo real-time study of bacterial cultures and similar mildly turbid systems. The BSN algorithm is an improvement on OD600 in several ways. It offers the potential to simultaneously, continuously and noninvasively monitor two culture parameters, e.g., NC concentration and nutrient concentration. By itself, OD600 can give information only on the concentration of organisms and that information is rendered ambiguous by the potential presence of debris and other sources of turbidity in addition to the bacteria. Thus, BSN provides a more complete picture of a bacterial culture than OD600.

The information given by BSN concerning legacy media like LB maybe helpful in discerning culture-induced modification(s) of the fluid environment even though it is not currently possible to assign Raman or other spectroscopic features associated with LB. Another advantage is that bacterial cultures can be studied more quickly and with greater flexibility, improving throughput and possibly increasing the rate of discovery, production and innovation. We here dis- cuss mostly methodological issues and defer most discussion of the results concerning bacteria and cultures to future papers.

In the present work, we parametrized the BSN algorithm for mixtures of PBS, LB, and NC, and applied the parametrized algorithm to growth curves. We showed how the calculated volume fractions of NC correlate with OD600 and vary in a predictable linear manner with additions of aliquots of bacteria We anticipate that observing time dependent behavior, e.g., growth curves would seem to be reproducible independent of the starting OD and BSN would allow experiments in which several cultures could be run in parallel to test, e.g., some hypothesis that effects the bacterial growth rate.

To apply the BSN algorithm, EE and IE must be measured for a range of bacterial and media dilutions to assure that they are linearly related to concentrations. In this study, we found it easy

to meet these requirements without extensive design and fabrication of specialized apparatus. This con- firms that the EE and IE are produced and collected in the single scattering limit. Note that the calibration parameters a–f obtained from the bilinear fits are dependent on the excitation/collection geometry. The training set calibration lines comparing BSN-calculated NC and LB concentrations to the actual values have slopes of 1 and intercepts of 0 within experimental error. Thus, our present optical arrangement was adequate to demonstrate the applicability of BSN to bacterial cultures and that development of commercial turnkey systems is possible.

Indeed, improvements in design and engineering for specific goals are possible. For example, monitoring the back reflection from the top surface of the liquid to drive a servo to keep the reflection constant would (1) remedy the drift due to evaporation of the liquid over time and (2) assess the presence of foam or debris on the surface. With attention to the artifacts revealed in this work, specialized apparatus and procedures for BSN can be developed for a wide range of possible scenarios. Clearly, there will be differences in engineering how BSN is used in cuvettes, 400 000 L bioreactors, 96-well plates and micro fluidic environments but many of the basic ideas and remedies for artifacts are anticipated here. It may also be possible to use the BSN approach in other situations, e.g., flow cytometry.

There are various practical issues that could contribute to ‘‘heterogeneity’’ that can be mitigated to some extent. For example, there may be fouling of windows if taking measurements continuously through observation ports or sides of culture tubes. However, we have found that examining the data stream in real time can reveal such artifacts. Aggregates or window fouling produces bigger intensity spikes than smaller fragments. In contrast, gradual fouling/fogging of an observation port window or fiber bundle, possibly with/without a catastrophic endpoint,

would present a different time course for the distribution of intensity in each of the two channels of remitted light shown in Fig. 1.

Engineering the capacity to move the probing light to different parts of an, e.g., observation port or culture tube wall can also mitigate that problem. In approaching the problem of macroscopic aggregates, which would need to be discriminated from foam and other bubbles and debris, remitted light data streams contain temporal characteristics of movement in a fluid, both bulk physical movement and diffusion/osmosis types of motion.

Conclusion

The BSN algorithm can be applied to a variety of bacterial and other turbid systems, i.e., cell cultures. Using the BSN algorithm can accurately determine the cell count in a culture without taking OD600 readings throughout the experiment. We have demonstrated that continuous, non-invasive monitoring in real time of biomass, i.e., membrane production and nutrient depletion for NC culture is possible in both LB and MM, using the 1060cm^{-1} and 1120 cm^{-1} Raman features respectively. The BSN algorithm provides a more comprehensive picture of the chemical and physical state of bacterial cultures than OD600 alone.

Acknowledgments

This work would not have been possible without the expert assistance of the R.P. Doyle Group, Hewen Zheng of the Luk Group, and the Hougland Group in the Chemistry Department of Syracuse University.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

Funding for this study was provided by Syracuse University.

ORCID iD Joseph Chaiken https://orcid.org/0000-0002-7715-3852

Supplemental Material

All supplemental material mentioned in the text, including training set data for SO and PA and experimental detail for culturing, is available in the online version of the journal.

References

1. C.E. Bagwell, S. Bhat, G.M. Hawkins, et al. ‘‘Survival in Nuclear Waste, Extreme Resistance, and Potential Applications Gleaned from the Genome Sequence of Kineococcus radiotolerans’’. PLoS One. 2008. 3(12): e3878.

2. M.H. Choi, S.-W. Jeong, H.E. Shim, et al. ‘‘Efficient Bioremediation of Radioactive Iodine Using Biogenic Gold Nanomaterial-Containing Radiation-Resistant Bacterium, Deinococcus radiodurans R1’’. Chem. Commun. 2017. 53(28): 3937–3940.

3. R.J. Di Mario, M.C. Machingura, G.L. Waldrop, et al. ‘‘The Many Types of Carbonic Anhydrases in Photosynthetic Organisms’’. Plant Sci. 2018. 268: 11–17.

4. Y. Doi. ‘‘L-Lactate Production from Biodiesel-Derived Crude Glycerol by Metabolically Engineered Enterococcus Faecalis: Cytotoxic Evaluation of Biodiesel Waste and Development of a Glycerol- Inducible Gene Expression System’’. Appl. Environ. Microbiol. 2015. 81(6):

2082–2089.

5. E.A. Ferreira, C. Da Conceicao De Matos, E.A. Barbosa, et al. ‘‘Cassava Physiological Responses to the Application of Herbicides’’. Semina: Cienc. Agrar. 2015. 36(2): 645–656.

6. H. Friman, A. Schechter, Y. Nitzan, et al. ‘‘Phenol Degradation in Bio- Electrochemical Cells". Int. Biodeterior. Biodegrad. 2013. 84: 155–160. 7. S. Mandal, A.-M. Carey, J. Locsin, et al. ‘‘Mechanism of Triplet Energy Transfer in Photosynthetic Bacterial Reaction Centers’’. J. Phys. Chem. B. 2017. 121(27): 6499–6510.

8. N.H.A. Nguyen, R. Spanek, V. Kasalicky, et al. ‘‘Different Effects of Nano-Scale and Micro-Scale Zero-Valent Iron Particles on Planktonic Microorganisms from Natural Reservoir Water’’. Environ. Sci.: Nano. 2018. 5(5): 1117–1129.

9. M.I. Sarro, A.M. Garcia, D.A. Moreno. ‘‘Biofilm Formation in Spent Nuclear Fuel Pools and Bioremediation of Radioactive Water’’. Int. Microbiol. 2005. 8(3): 223–230.

10. J. Wang. ‘‘Engineered Radiation-Resistant Bacteria and Their Application in Bioremediation of Radioactive Waste-Contaminated Environment’’. J. Radiat. Res. Radiat. Process. [Fushe Yanjiu Yu Fushe Gongyi Xuebao]. 2004. 22(5): 257–260.

11. A. Adan, G. Alizada, Y. Kiraz, et al. ‘‘Flow Cytometry: Basic Principles and Applications’’. Crit. Rev. Biotechnol. 2017. 37(2):163–176.

12. G. Freer, L. Rindi. ‘‘Intracellular Cytokine Detection by Fluorescence-Activated Flow Cytometry: Basic Principles and Recent Advances’’. Methods. 2013. 61(1): 30–38.

13. D.J. Weiss, M.J. Wilkerson. ‘‘Flow Cytometry’’. In: D. Weiss, D.J. Weiss, editors. Schalm’s Veterinary Hematology. Ames, IA: Wiley-Blackwell, 2010Chap. 12, Pp.

14. J. Fyrestam, C. Oestman. ‘‘Determination of Heme in Microorganisms Using HPLC-MS/MS and Cobalt(III) Protoporphyrin IX Inhibition of Heme Acquisition in Escherichia Coli’’. Anal.

Bioanal. Chem. 2017. 409(30): 6999–7010.

15. J. Jiang, B. Wang, J. Li, et al. ‘‘Total Coumarins of Hedyotis diffusa Induces Apoptosis of Myelodysplastic Syndrome SKM-1 Cells by Activation of Caspases and Inhibition of PI3K/Akt Pathway Proteins’’. J. Ethnopharmacol. 2017. 196: 253–260.

16. R.S. Ur, P.G. Casey, C. Hill, et al. ‘‘Development of a Click Beetle Luciferase Reporter System for Enhanced Bioluminescence Imaging of Listeria Monocytogenes: Analysis in Cell Culture and Murine Infection Models’’. Front Microbiol. 2017. 8: 1797.

17. R.B. Santos, B. Chandrasekar, M.K. Mandal, et al. ‘‘Low Protease Content in Medicago truncatula Cell Cultures Facilitates Recombinant Protein Production’’. Biotechnol. J. 2018. 13(7)e1800050.

18. X. Jia, Y. Liu, C. Li, et al. ‘‘Optimum Growth Conditions of Klebsiella pneumoniae for Degradation of Petroleum Contaminants and Tolerability Research Methods’’. 2016. Tianjin University, China.

19. A.L. Koch. ‘‘Turbidity Measurements of Bacterial Cultures in Some Available Commercial Instruments’’. Anal. Biochem. 1970. 38(1): 252–259.

20. G. Bratbak, I. Dundas. ‘‘Bacterial Dry Matter Content and Biomass Estimations’’. Appl. Environ. Microbiol. 1984. 48(4): 755–757.

21. M. Loferer-Krossbacher, J. Klima, R. Psenner. ‘‘Determination of Bacterial Cell Dry Mass by Transmission Electron Microscopy and Densitometric Image Analysis’’. Appl. Environ. Microbiol. 1998. 64(2): 688–694.

22. R.T. McDonough, H. Zheng, M.A. Alila, et al. ‘‘Optical Interference Probe of Biofilm Hydrology: Label-Free Characterization of the Dynamic Hydration Behavior of Native Biofilms’’. J. Biomed. Opt. 2017. 22(3): 035003/1–035003/13.

23. P. Dent, B. Deng, J. Goodisman, et al. ‘‘Coupled Turbidity and Spectroscopy Problems: A Simple Algorithm for the Volumetric Analysis of Optically Thin or Dilute Two-Phase Systems’’. Appl. Spectrosc. 2015. 69(3): 377–388.

24. J. Chaiken, J. Goodisman, B. Deng, et al. ‘‘Simultaneous, Noninvasive Observation of Elastic Scattering, Fluorescence and Inelastic Scattering as a Monitor of Blood Flow and Hematocrit in Human Fingertip Capillary Beds’’. J. Biomed. Opt. 2009. 14(5): 050505/ 1–050505/3.

25. I.K. Leth, K.A. McDonald. ‘‘Media Development for Large Scale Agrobacterium Tumefaciens Culture’’. Biotechnol. Prog. 2017. 33(5): 1218–1225.

26. R. Chelikani, F.H. Bacon, H. Yuan. Overcoming the Challenges of Moving from Hydrolysate-Containing Media to Chemically Defined Media for a High Cell Density Perfusion Cell Culture Process. Washington, DC: American Chemical Society, 2017.

27. L.B. Hazeltine, K.M. Knueven, Y. Zhang, et al. ‘‘Chemically Defined Media Modifications to Lower Tryptophan Oxidation of Biopharmaceuticals’’. Biotechnol. Prog. 2016. 32(1): 178–188.

28. M. Krajcovic, Y. Ilin, S. Hutchins, et al. ‘‘What Cells Want Versus What Cells Need: Simplification of Chemically Defined Media for Optimal Performance’’. 2018. American Chemical Society.

29. V. Sublett, B. Cooper, J.O. Boles. ‘‘Use of Enhanced Chemically Defined Media to Improve Uptake of L-Telluromethionine by Escherichia coli DL41(DE3)’’. 2015. Technical Session presented at: 249th American Chemical Society National Meeting. Colorado Convention Center, Denver, Colorado; 22–26 March 2015.

30. J. Xu, M.S. Rehmann, X. Xu, et al. ‘‘Improving Titer While Maintaining Quality of Final

Formulated Drug Substance Via Optimization of CHO Cell Culture Conditions in Low-Iron Chemically Defined Media''. mAbs. 2018. 10(3): 488–499.

31. J.A. Silver. ''Frequency-Modulation Spectroscopy for Trace Species Detection: Theory and Comparison Among Experimental Methods''. Appl. Opt. 1992. 31(24): 4927.

32. J.A. Silver. ''Frequency-Modulation Spectroscopy for Trace Species Detection: Theory and Comparison Among Experimental Methods: Errata''. Appl. Opt. 1992. 31(6): 707–717. doi:10.1364/AO.31.004927.

33. B. Deng, A. Simental, P. Lutz, et al. ''Singlet Oxygen Induced Advanced Glycation End-Product Photobleaching of in Vivo Human Fingertip Autofluorescence''. Proc. SPIE. Biomedical Vibrational Spectroscopy V: Advances in Research and Industry. 2012. 8219: 82190D/ 1–82190D/11.

34. B. Deng, C. Wright, E. Lewis-Clark, et al. ''Direct Noninvasive Observation of Near Infrared Photobleaching of Autofluorescence in Human Volar Side Fingertips in Vivo''. Proc. SPIE. Biomedical Vibrational Spectroscopy IV: Advances in Research and Industry. 2010. 7560: 75600P/1–75600P/11.

35. K. Bae, W. Zheng, Z. Huang. ''Quantitative Assessment of Spinal Cord Injury Using Circularly Polarized Coherent Anti-Stokes Raman Scattering Microscopy''. Appl. Phys. Lett. 2017. 111(6): 063704/ 1–063704/4.

36. T. Saxena, B. Deng, E. Lewis-Clark, et al. ''Near Infrared Raman Spectroscopic Study of Reactive Gliosis and the Glial Scar in Injured Rat Spinal Cords''. Proc. SPIE. Biomedical Vibrational Spectroscopy IV: Advances in Research and Industry. 2010. 7560: 75600I/1–75600I/12.

37. T. Saxena, B. Deng, D. Stelzner, et al. ''Raman Spectroscopic Investigation of Spinal Cord

Injury in a Rat Model''. J. Biomed. Opt. 2011. 16(2): 027003/1–027003/14.

38. J. Shen, L. Fan, J. Yang, et al. ''A Longitudinal Raman Microspectroscopic Study of Osteoporosis Induced by Spinal Cord Injury''. Osteoporosis Int. 2010. 21(1): 81–87.

39. S. Tamosaityte, R. Galli, E. Koch, et al. ''Inflammation-Related Alterations of Lipids After Spinal Cord Injury Revealed by Raman Spectroscopy''. J. Biomed. Opt. 2016. 21(6): 61008.

40. S. Wang, Z. Liang, Y. Gong, et al. ''Confocal ''Raman Microspectral Imaging of Ex Vivo Human Spinal Cord Tissue''. J. Photochem. Photobiol., B. 2016. 163: 177–184.

41. T.L. Chester, J.J. Fitzgerald, J.D. Winefordner. ''Theoretical Comparison of the Signal-to-Noise Ratios of Fourier Transform Spectrometry with Single Slit Linear and Slewed Scan Spectrometric Methods for the Photon Noise Limited Situation''. Anal. Chem. 1976. 48(4): 779–783.

42. J.D. Ingle Jr., S.R. Crouch. ''Molecular Scattering Methods''. In: Spectrochemical Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1988.

43. J. Chaiken, W.F. Finney, X. Yang, et al. ''Progress in the Noninvasive In- Vivo Tissue-Modulated Raman Spectroscopy of Human Blood''. Proc. SPIE Biomedical Diagnostic, Guidance, and Surgical-Assist Systems III. 2001. BiOS 2001 The International Symposium on Biomedical Optics. San Jose, CA; 4 Jun 2001. 4254: 216–227.

44. B.D. Beier, A.J. Berger. ''Method for Automated Background Subtraction from Raman Spectra Containing Known Contaminants''. Analyst. 2009. 134(6): 1198–1202.

45. P.J. Cadusch, M.M. Hlaing, S.A. Wade, et al. ''Improved Methods for Fluorescence Background Subtraction from Raman Spectra''. J. Raman Spectrosc. 2013. 44(11): 1587–1595.

46. K. Lin, Z. Wang, S. Ju, et al. ''Based on Removing Spectral Raman Spectroscopy Fluorescence Background Subtraction Method''. 2018. Xidian University, China. 11 pp.

47. A.C. Croce, G. Bottiroli. ‘‘Autofluorescence Spectroscopy and Imaging: A Tool for Biomedical Research and Diagnosis’’. Eur. J. Histochem. 2014. 58(4): 2461.

48. G.H. Patterson, D.W. Piston. ‘‘Photobleaching in Two-Photon Excitation Microscopy’’. Biophys. J. 2000. 78(4): 2159–2162.

49. M.G. Mu¨ller, I. Georgakoudi, Q. Zhang, et al. ‘‘Intrinsic Fluorescence Spectroscopy in Turbid Media: Disentangling Effects of Scattering and Absorption’’. Appl. Opt. 2001. 40(25): 4633–4646.

50. I. Barman, G.P. Singh, R.R. Dasari, et al. ‘‘Turbidity Corrected Raman Spectroscopy for Blood Analyte Detection’’. Anal. Chem. 2009. 81(11): 4233–4240.

51. J.V. Sinfield, C.K. Monwuba. ‘‘Assessment and Correction of Turbidity Effects on Raman Observations of Chemicals in Aqueous Solutions’’. Appl. Spectrosc. 2014. 68(12): 1381–92.

52. A.T. Tu. Raman Spectroscopy in Biology: Principles and Applications. New York, NY: John Wiley & Sons, 1982.

53. R. Mendelsohn, S. Sunder, H.J. Bernstein. ‘‘Deuterated Fatty Acids as Raman Spectroscopic Probes of Membrane Structure’’. Biochim. Biophys. Acta. 1976. 443(3): 613–617.

54. J. Chaiken, W.F. Finney, C.M. Peterson, et al. ‘‘Noninvasive In-Vivo Tissue-Modulated Near-Infrared Vibrational Spectroscopic Study of Mobile and Static Tissues: Blood Chemistry’’. Proc. SPIE Biomedical Spectroscopy: Vibrational Spectroscopy and Other Novel Techniques. 2000. 3918: 135–143.

8. References

1 V. Tuchin, *Tissue Optics: Light Scattering Methods and Instruments for Medical Diagnosis*, SPIE Press, Bellingham, WA, 2000.

2 A.L. Koch, "Turbidity Measurements of Bacterial Cultures in Some Available Commercial Instruments", *Anal. Biochem.*, 38, 252-259 (1970).

3 R.T. McDonough, et al, "Optical Interference Probe of Biofilm Hydrology: Label-Free Characterization of the Dynamic Hydration Behavior of Native Biofilms", J. Biomed. Opt., 22(3), 1-13 (2017).

4 P. Dent, et al, "Coupled Turbidity and Spectroscopy Problems: A Simple Algorithm for Volumetric Analysis of Optically Thin or Dilute Two Phase System", *Applied Spectroscopy*, 69(3), 377-388 (2015).

5 J. Chaiken, et al, "Simultaneous, Noninvasive Observation of Elastic Scattering, Fluorescence and Inelastic Scattering as a Monitor of Blood Flow and Hematocrit in Human Fingertip Capillary Beds", *J. Biomed. Opt*., 14(5), 1-3 (2009).

6 S. Ortiz, et al, "Label-free, Noninvasive Optical Measurement of Bacterial Mobility in Native Culture Using Binary Spectronephelometry", *Manuscript in Progress*, (2019).

7 D. Rice, et al, "PV[O]H: Noninvasive Enabling Technology, New Physiological Monitoring, and Big Data", Military Medicine, 158, 458-464 (2021).

8 D. Rice, et al, "The PVOH device: Our First Stop on the Path to Small and Very Small Physical Embodiments of the PV[O]H Algorithm", *Photonic Diagnosis, Monitoring, Prevention, and Treatment of Infections and Inflammatory Diseases 2020*, Proc. SPIE 11223, 1-9 (2020).

9 L. Xiang, L. C. Chan and J. W. C. Wong, “Removal of Heavy Metals From Anaerobically Digested Sewage Sludge by Isolated Indigenous Iron-Oxidizing Bacteria”, *Chemosphere*, 41, 283-287 (2000).

10 J. Kodha, et al, “Measurement of the Glycerol Concentration During the Microbial Treatment of the Wastewater From the Biodiesel Fuel Production Plant Using Near Infrared Spectroscopy”, *J. Near Infrared Spec.*, 16, 199-204 (2008).

11 N. Tabbsum, N. Islam and S. Ahmed, “Progress in Microbial Fuel Cells For Sustainable Management of Industrial Effluents”, *Process Biochem.*, 106, 20-41 (2021).

12 J. Chaiken, et al, “Analyzing Near-Infrared Scattering from Human Skin to Monitor Changes in Hematocrit”, *J. Biomed. Opt.*, 16(9), 1-19 (2011).

13 J.W. Costerton, P.S. Stewart and E.P. Greenberg, “Bacterial Biofilms: A Common Cause of Persistent Infections”, *Science*, 284(5418), 1318-1322 (1999).

14 J.B. Lyczak, C.L. Cannon and G.B. Pier, “Establishment of *Pseudomonas aeruginosa* Infection: Lessons From a Versatile Opportunist”, *Microbes and Infection*, 2, 1051-1060 (2000).

15 X. Wu, et al, “Culture-Free Diagnostics of *Pseudomonas aeruginosa* Infection by Silver Nanorod Array Based SERS from Clinical Sputum Samples”, *Nanomedicine: NBM*, 1-8 (2014).

16 H. Chai, et al, “Spectroscopic Investigations of the Binding Mechanisms Between Antimicrobial Peptides and Membrane Models of *Pseudomonas aeruginosa* and *Klebsiella pneumonia*”, *Bioorganic & Medicinal Chemistry*, 4210–4222 (2014).

17 M. Arabski, S. Wąsik, and Z. Drulis-Kawa, “Laser Interferometry Analysis of Ciprofloxacin Diffusion through *Pseudomonas aeruginosa* Biofilm”, *Clin. Microbiol*. 2, 105-110 (2013).

18 M. J. Edwards, et al, “The X-ray Crystal Structure of Shewanella oneidensis OmcA Reveals New Insight at the Microbe–Mineral Interface”, *FEBS Letters*, 588, 1886-1890 (2014).

19 F. Golitsch, C. Bücking and J. Gescher, “Proof of Principle for an Engineered Microbial Biosensor Based on Shewanella oneidensis Outer Membrane Protein Complexes”, *Biosensors and Bioelectronics*, 47, 285-291 (2013).

20 K. Artyushkova, et al, “Relationship between surface chemistry, biofilm structure, and electron transfer in Shewanella anodes”, *Biointerphases*, 10, 1-12 (2015).

21 S. Babanova, et al, “Innovative Statistical Interpretation of *Shewanella oneidensis* Microbial Fuel Cell Data”, *Phys. Chem. Chem. Phys.*, 16, 8956-8969 (2014).

22 N. Singh, et al., “Chemical Signals of Synthetic Disaccharide derivatives Dominate Rhamnolipids at Controlling Multiple Bacterial Activities”, *ChemBioChem*, 17, 107-111 (2016).

23 G. S. Sheyte, et al., “Specific Maltose Derivatives Modulate the Swarming Motility of Nonswarming Mutant and Inhibit Bacterial Adhesion and Biofilm Formation by *Pseudomonas aeruginosa*”, *ChemBioChem*, 15, 1514-1523 (2014).

24 V. Twersky, “Multiple Scattering of Waves and Optical Phenomena”, *J. Opt. Soc. Am.*, 52 (2), 145 – 171 (1962).

25 V. Twersky, “Absorption and Multiple Scattering of Biological Suspensions”, *J. Opt. Soc. Am.*, 60 (8), 1084 – 1093 (1969).

26 J. Chaiken, et al., “Instrument for Near Infrared Emission Spectroscopic Probing of Fingertip *In Vivo*”, *Rev. Sci. Instrum.*, 81, 1-10 (2010).

27 J. Chaiken, et al., “Noninvasive, in vivo, Modulation Near Infrared Vibrational Spectroscopic Study of Mobile and Static Tissues: Blood Chemistry”, *Proc. SPIE*, 3918, 134 – 143 (2000).

28 P. Mitchell, “General Theory of Membrane Transport from Studies of Bacteria”, *Nature* 180, 134 – 136 (1957).

29 D. K. Shen, et al, “PsrA is a Positive Transcriptional Regulator of the Type III Secretion System in *Pseudomonas aeruginosa*”, *Infection and Immunity* 74(2), 1121-1129 (2006).

30 D. K. Shen, et al, “High-Cell-Density Regulation of the Pseudomonas aeruginosa Type III Secretion System: Implications for Tryptophan Catabolites”, *Microbiology* 154, 2195-2208 (2008).

31 M. Baniasadi, et al, “Nano Indentation of *Pseudomonas aeruginosa* Bacterial Biofilm Using Atomic Force Microscopy”, *Materials Research Express 1*, 1-15 (2014).

32 O. Stukalov, et al, “Use of Atomic Force Microscopy and Transmission Electron Microscopy for Correlative Studies of Bacterial Capsules”, *Appl. Environ. Microbiol.* 74(17), 5457-5465 (2008).

33 S. Bose, et al, “Bioreduction of hematite nanoparticles by the dissimilatory iron reducing bacterium Shewanella oneidensis MR-1”, *Geochimica et Cosmochimica Acta*, 73, 962-976 (2009).

34 T. Ogi, et al, “Direct room-temperature synthesis of a highly dispersed Pd nanoparticle catalyst and its electrical properties in a fuel cell”, *Powder Technology*, 205, 143-148 (2011).

35 A. T. Tu, *Raman Spectroscopy in Biology: Principles and Applications*, “Lipids and Biological Membranes”, John Wiley & Sons, New York, NY, 1982.

36 R. T. McDonough, et al, “Coupled turbidity and spectroscopy problems: a simple algorithm for volumetric analysis of optically thin or dilute, in vitro bacterial cultures in various media”, Manuscript in Preparation, *Applied Spectroscopy*.

37 L. F. M. Franco, M. Castier and I. G. Economou, “Diffusion in Homogeneous and Inhomogeneous Media: A New Unified Approach”, *Journal of Chemical Theory and Computation*, 12, 5247-5255 (2016).

38 G. L. Paul and P. N. Pusey, “Observation of a Long Time Tail in Brownian Motion”, *Journal of Physics A: Mathematical and General*, 14, 3301-3327 (1981).

39 D. Germain, M. Leocmach and T. Gibaud, “Differential Dynamic Microscopy to Characterize Brownian Motion and Bacteria Motility”, *Condensed Matter*, 1-10 (2015).

40 D. J. Rowland, H. H. Tuson and J. S. Biteen, “Resolving Fast, Confined Diffusion in Bacteria with Image Correlation Spectroscopy”, *Biophysical Journal*, 110, 2241-2251 (2016).

41 M. L. Adams, *Chemometrics in Analytical Spectroscopy*, Royal Society of Chemistry Thomas Graham House Science Park Milton Road Cambridge CB4 0WF UK (2004).

42 S. Wold, K. Esbensen and P. Geladi, “Principal Component Analysis”, *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52 (1987).

43 I. T. Jolliffe and J. Cadima, “Principal Component Analysis: A Review and Recent Developments”, *Phil. Trans. R. Soc. A*, 374, 1-16 (2016).

44 D. Ballabio, “A MATLAB Toolbox for Principal Component Analysis and Unsupervised Exploration of Data Structures”, *Chemometrics and Intelligent Laboratory Systems*, 149, 1-9, (2015).

45 G. W. Stewart, “On the Early History of the Singular Value Decomposition”, *SIAM Review*, 35(4), 551-566 (1993).

46 M. Zare, M. Eftekhari and G. Aghamollaei, “Supervised Feature Selection Via Matrix Factorization Based on Singular Value Decomposition”, *Chemometrics and Intelligent Laboratory Systems*, 185, 105-113 (2019).

47 Y. -C. Kim, “Diffusivity of Bacteria”, *Korean Journal of Chemical Engineering*, 13, 282-287 (1996).

48 B. Gibson, et al, “The Distribution of Bacteria Doubling Times in the Wild”, *Proc. R. Soc. B*, 285, 1-9 (2019).

49 G. R. Anderson, “The Raman Spectra of Carbon Dioxide in Liquid H2O and D2O”, *The J. of Phys Chem*, 81(3), 273-276 (1977).

---

9. Curriculum Vitae

Steven Ortiz email: stortiz@syr.edu

Education

Ph.D., Chemistry, Syracuse University Expected Dec. 2021

Continued thesis research with Dr. Chaiken entitled "Monitoring Bacteria Cultures using Near Infrared (NIR) Binary Spectronephelometry (BSN), Raman Spectra and Principal Component Analysis (PCA)".

M.S., Chemistry, Syracuse University May 2015

Completed research and defended a thesis entitled "Label Free Analysis of Planktonic *P. aeruginosa* Cells and Growth Media by Near Infrared (NIR) Binary Spectronephelometry (BSN)".

B.S., Chemistry, Rhode Island College May 2012

Industry Experience

- Quality Control Analysist, Byrne Ultra Dairy Jan 2021 to Dec 2021

  Extract A1 and A2 beta casein proteins in milk products and analyze by UPLC.

- Chemical Lab Technician, Interplex Engineered Products June 2012 to July 2013

  Perform analytical tests on production hull cells at a metal plating factory for electrical components.

Publications/Presentations

- S. Ortiz, R.T. McDonough, P. Dent, J. Goodisman and J. Chaiken, "Coupled Turbidity and Spectroscopic Problems: A Simple Algorithm for Volumetric Analysis of Optically Thin or Dilute, In Vitro Bacterial Cultures in Various Media", *Applied Spectroscopy*, 74(3), 261-274 (2020) — Jul. 2019
- S. Fillioe, K. K. Bishop, J. V. S. Jannini, J. J. I. Kim, R. McDonough, S. Ortiz, J. Goodisman, J. Hasenwinkel, C. M. Peterson and J. Chaiken, "In Vivo, Noncontact, Real-Time, PV[O]H Imaging of the Immediate Physiological Response to Spinal Cord Injury in a Rat Model", *Journal of Biomedical Optics*, 25(3), 1-12 (2019) — Oct. 2019
- P. Dent, R. McDonough, K. Rawlins, S. Ortiz, J. Goodisman and J. Chaiken, "Spectroscopic Analysis of Weakly Turbid Systems: A Generally Useful Simple Algorithm", Abstracts of Papers, 249th ACS National Meeting and Exposition (2015) — March 2015
- S. Ortiz and K. H. Almeida, "Molecular Dimerization State of NAMPT", Poster Presentation at the RI-INBRE Symposium. — Aug. 2011

Research Experience

- Graduate Research Assistant: Research involves the use of Raman, fluorescence and UV-Vis instrumentation to analyze — Nov. 2013 to Present

bacteria cultures and the development of an algorithm to extrapolate data.

- Undergraduate Research Assistant: Involved using biochemical techniques such as restriction enzyme digest, gel electrophoresis and plasmid purification on bacteria and yeast cells. April 2011 to Dec. 2011

Teaching Experience

- Graduate Teaching Assistant, Syracuse University, CHE 347: Physical Chemistry Lab and CHE 107 & 117: General Chemistry Lab I & II. Sept. 2013 to April 2018
- Physical Sciences Tutor, Rhode Island College, General, Organic and Biochemistry and Physics. Sept. 2010 to April 2012

Professional Development

- Graduate courses included organic spectroscopy, perspectives in biochemistry, crystallography, vibrational spectroscopy, advanced physical chemistry and metals in medicine.
- Relevant undergraduate courses are biochemistry, analogue electronics, analytical chemistry and introduction to instrumentation.

References

- Joseph Chaiken Ph.D., Professor of Chemistry, Syracuse University, CST 2-010 Syracuse, NY 13244 (315) 443-4285 jchaiken@syr.edu
- Karen Almeida Ph.D., Professor of Chemistry, Rhode Island College Clarke Science Bldg. 108, 600 Mt. Pleasant Ave. Providence, RI 02908 (401) 456-9665 kalmeida@ric.edu