Advanced modelling, monitoring, and process control of bioconversion systems

Elliott C. Schmitt

A dissertation

Submitted in partial fulfillment of the

Requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Richard Roy Gustafson, Chair

Renata Bura

Fernando Resende

Program Authorized to Offer Degree:

School of Environmental and Forest Resources

University of Washington

Abstract

Advanced modelling, monitoring, and process control of bioconversion systems

Elliott C. Schmitt

Chair of the Supervisory Committee:

Dr. Richard Roy Gustafson

School of Environmental and Forest Resources

Production of fuels and chemicals from lignocellulosic biomass is an increasingly important area of research and industrialization throughout the world. In order to be competitive with fossil-based fuels and chemicals, maintaining cost-effectiveness is critical. Advanced process control (APC) and optimization methods could significantly reduce operating costs in the biorefining industry. Two reasons APC has previously proven challenging to implement for bioprocesses include: lack of suitable online sensor technology of key system components, and strongly nonlinear first principal models required to predict bioconversion behavior.

To overcome these challenges batch fermentations with the acetogen *Moorella thermoacetica* were monitored with Raman spectroscopy for the conversion of real lignocellulosic hydrolysates and a kinetic model for the conversion of synthetic sugars was developed. Raman spectroscopy was shown to be effective in monitoring the fermentation of

sugarcane bagasse and sugarcane straw hydrolysate, where univariate models predicted acetate concentrations with a root mean square error of prediction (RMSEP) of 1.9 and 1.0 g L^{-1} for bagasse and straw, respectively. Multivariate partial least squares (PLS) models were employed to predict acetate, xylose, glucose, and total sugar concentrations for both hydrolysate fermentations. The PLS models were more robust than univariate models, and yielded a percent error of approximately 5% for both sugarcane bagasse and sugarcane straw. In addition, a screening technique was discussed for improving Raman spectra of hydrolysate samples prior to collecting fermentation data. Furthermore, a mechanistic model was developed to predict batch fermentation of synthetic glucose, xylose, and a mixture of the two sugars to acetate. The models accurately described the bioconversion process with an RMSEP of approximately 1 g L^{-1} for each model and provided insights into how kinetic parameters changed during dual substrate fermentation with diauxic growth.

Model predictive control (MPC), an advanced process control strategy, is capable of utilizing nonlinear models and sensor feedback to provide optimal input while ensuring critical process constraints are met. Using the microorganism *Saccharomyces cerevisiae,* a commonly used microorganism for biofuel production, and work performed with *M. thermoacetica,* a nonlinear MPC was implemented on a continuous membrane cell-recycle bioreactor (MCRB) for the conversion of glucose to ethanol. The dilution rate was used to control the ethanol productivity of the system will maintaining total substrate conversion above the constraint of 98%. PLS multivariate models for glucose (RMSEP 1.5 g L^{-1}) and ethanol (RMSEP 0.4 g L^{-1}) were robust in predicting concentrations and a mechanistic kinetic model built accurately predicted continuous fermentation behavior. A setpoint trajectory, ranging from 2 – 4.5 g L^{-1} h^{-1} for productivity was closely tracked by the fermentation system using Raman measurements and

an extended Kalman filter to estimate biomass concentrations. Overall, this work was able to demonstrate an effective approach for real-time monitoring and control of a complex fermentation system.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

First and foremost I would like to thank my two main advisors, Dr. Rick Gustafson and Dr. Renata Bura. I have known both of them since I started my Masters at UW in Mechanical Engineering in 2007. Throughout this entire process they have given me unconditional support, while allowing me the freedom to explore and research the topics that I found most interesting. In addition, they always provided insight, motivation, and challenged me to become a better researcher for which I am truly grateful.

I would like to thank my committee members Dr. Fernando Resende, Dr. Dan Smith, and Dr. Brad Holt for many useful conversations and for allowing me the opportunity to present my research to them.

I want to give a special thanks to my fellow Biofuels and Bioproducts Lab colleagues for their support. First, Dr. Shannon Ewanick and Dr. Azra Vajzovic, who I met when I first started doing my research and they have always provided valuable mentorship as well as friendship. Thank you to my travel companions to many a conferences, Erik Budsberg, Jordan Crawford, Rodrigo Morales, and my South American travel buddy, Chang Dou; my 108 lab colleagues operating on the thermal side of bioconversion, Oliver Jan and Guanqun Luo. Most importantly I would like to thank Mandana Ehsanipour who was a major contributor to my research, and I could not have done half my work without her. Thank you all for everything, I'm grateful to have had the chance to work with you and know you.

Last but certainly not least I would like to thank my family and friends for their continuous support throughout my life. To my parents, Raymond Schmitt and Claudia Vandervord, thank you for being a source of inspiration that made me the man I am today.

Chapter 1: Introduction

Portions of this chapter were published in Pure and Applied Chemistry. Volume 86, Issue 5, pages 867 – 879. (2014).

1.1 CURRENT STATE OF ADVANCED BIOFUELS PRODUCTION

In 2007 the United States passed the second version of the Renewable Fuels Standard (RFS2), designed to promote the sustainable production and consumption of advanced biofuels. Today, majority of biofuel production is conducted at first generation biorefineries that utilize corn starch as the major feedstock for bioethanol fuel production, with co-production of dried grains (DDG) used for animal feed and residue material that is used for energy production. Second generation facilities, which seek to utilize lignocellulosic biomass for the production of biofuels, are expected to reach total production capacity that would eventually surpass that of first generation corn ethanol by 2022 [1]. In 2013, the RFS2 mandate called for a billion gallons of cellulosic biofuel production, but less than a million gallons of total production was estimated [2]. There are a number of factors that have hindered the biofuels industry, including: limited investment from private sector, the ethanol blend wall, lack of infrastructure to promote E85 or higher fuel grades, the high production costs associated with processing lignocellulosic materials for the production of biofuels and feedstock cost, which is estimated to make up at least a third of the total fuel cost [3].

Much like petroleum, a multitude of products can be produced from lignocellulosic biomass including: transportation fuels, commodity chemicals and chemical precursors to higher value products. However, current second generation biofuel production cost estimates are two to three times that of petroleum fuels on an energy equivalent [4]. Biofuel and bio-based chemical production processes must become more efficient to be competitive with comparable fossil fuel based production. Fuels and chemicals manufacturing processes from petroleum have a wealth of

online analytical sensors that permit them to operate at or near capacity with optimal yields. The ability to provide online measurements leads to automation and advanced process control, where production costs can be further reduced by increasing efficiencies. This hyper efficiency is a necessary condition for profitability in manufacturing high volume, narrow-profit margin products such as fuels and commodity chemicals.

Advanced process control has a significant impact in refining and petrochemicals industry due to the low-profit margins, driving the importance of even incremental performance increases [5]. Majority of the current research focuses on reducing operating costs of biofuels and biochemical by finding cheaper feedstocks, developing efficient and robust microorganisms, process integration, and co-product utilization [4]. There has been relatively little emphasis on process control and optimization methods, which have significantly reduced operating costs in similar industries. The need for process efficiency – and hence the need for online sensors – is especially acute in biomass fed biorefineries due to the complexity and expense of the feedstock. Process improvements (feedstock pretreatment, microorganisms, enzymes, etc.) are expected to reduce costs in the near future, but in the long-term, improving the efficiency of existing operations will likely have a significant effect on overall process economics.

1.2 ADVANCED PROCESS CONTROL

The goal of process control is to keep conditions of a set of controlled variables at a desired or optimal value in the presence of process variations. Advanced process control (APC) extends this by handling multivariable and nonlinear systems, with constraints if necessary, while achieving real-time optimization. APC systems are common practice throughout various industries for production of petrochemicals, chemicals, polymers and energy [6]. Benefits include increasing profit margins through increases in throughput, process stability improvement, energy

consumption reduction and increased yield of valuable products [7]. However, the literature concerning control systems for lignocellulosic bioprocesses is relatively scarce. A review of controls for bioreactors from Rani and Rao [8], discussed various control techniques including dynamic programming, online adaptive control, online optimization, nonlinear control, and optimal control. The two major reasons for the minimal application of APC for bioreactors include: highly nonlinear behavior of the bioprocess itself and lack of reliable technology for online measurements of major chemical and biochemical concentrations [9].

The type of control strategy implemented depends on the bioreactor operating mode (batch, fed-batch, continuous), availability of online measurements, and the accuracy of the bioreactor model [9]. Development of control systems that use feedback measurements from bioreactors is critical for lignocellulosic biorefineries. Over the last decade model predictive control (MPC) or receding horizon control (RHC) has become commonplace in many chemical and manufacturing industries [6]. In 2003, it was reported that there were over 5,000 applications of MPC, with majority of the applications being in refining, petrochemicals, chemicals, and pulp and paper industries [10]. More recently, an international survey of 66 APC experts, 38 APC users, and 28 APC suppliers showed that over two thirds of the respondents used MPC [7]. In the literature, the majority of the applications of MPC to fermentation systems have been based on simulation. These include both fed-batch [11], [12] and continuous [13]–[15] systems, all of which used some form of nonlinear MPC or neural network. Though academic research continues to develop the theory and algorithms for MPC, there exists a wealth of experience that can be applied to biorefining from related industries. For the last few decades, MPC has become the method of choice for difficult, multivariable control problems with significant interactions between inputs and outputs in refineries and petrochemical plants [6].

1.2.1 Model predictive control

Model predictive control (MPC) or receding horizon control (RHC), is an industry established optimal control strategy dating back to 1970 [6]. In general, the goal of optimal control theory is to minimize a cost function by determining a control law that achieves a desired optimal criterion. If the system is linear and the cost function quadratic without any constraints, the optimal control law can be found analytically over an infinite control horizon using a linear quadratic controller (LQC) design. However, if the system is nonlinear or system constraints are necessary, than another control strategy, such as MPC, is better suited. MPC is an online model-based control strategy for real-time optimization of a process. A model of the plant is used to make open-loop predictions of the future behavior of the system over a finite horizon, and the solution to a cost function (usually quadratic) determines the optimal input for the closed-loop system. At each sample instance the MPC algorithm is repeated, and the first control action in a sequence of predicted future control actions is implemented. The calculation of the optimal input at each sample instance is the main difference to conventional control techniques. MPC can handle linear or nonlinear equality or inequality constraints, multiple inputs and outputs (MIMO), and linear or nonlinear models. Furthermore, MPC naturally utilizes both elements of feedforward and feedback control. For further discussion of linear MPC see [16], [17], and for nonlinear MPC [18], [19].

In MPC a model is required to predict future system states of the plant. A model can be represented in continuous or discrete form, though in practice most models are implemented in discrete form. A nominal nonlinear discrete model of a plant can be given as,

$$x_{k+1} = f(x_k, u_k) \tag{1.1}$$

$$y_k = h(x_k, u_k)$$

where $x_k \in \mathbb{R}^n, u_k \in \mathbb{R}^m$ $y_k \in \mathbb{R}^p$ are the system state, control input, and output, respectively, at the current sample instance k. The state model can be a nonlinear function $f(x_k, u_k)$ with a nonlinear output function $h(x_k, u_k)$. Then the optimization problem can be formulated as follows,

$$min\ x'_{k+p} P x_{k+p} + \sum_{i=0}^{N_p-1} x'_{k+i} Q x_{k+i} + \sum_{i=0}^{N_m-1} u'_{k+i} R u_{k+i} \tag{1.2}$$

Subject to process constraints

$$E x_k + F u_k \leq G \tag{1.3}$$

where, N_p denotes the length of the prediction horizon, and N_m denotes the length of the control horizon ($N_m \leq N_p$). The weighting matrices *Q, R,* and *P* are positive definite, and can be used as tuning parameters, along with the horizon constants N_p and N_m. The first term of the objective function is called a terminal cost as it penalizes any solutions not found in a feasible region. This is not always implemented, if the system is well-behaved; however it is often used to make the algorithm stable. The second term drives the current state or controlled variable towards the desired setpoint. The third term is used to penalize the manipulated variable from any major deviations. The algorithm for implementing MPC is summarized in Algorithm 1.1, and visually shown in Figure 1.1.

Algorithm 1.1 (Basic MPC controller)

1. Measure x_k
2. Obtain u_k by solving a finite horizon optimal control problem
3. Apply the first element $u_k = u_{k|k}$

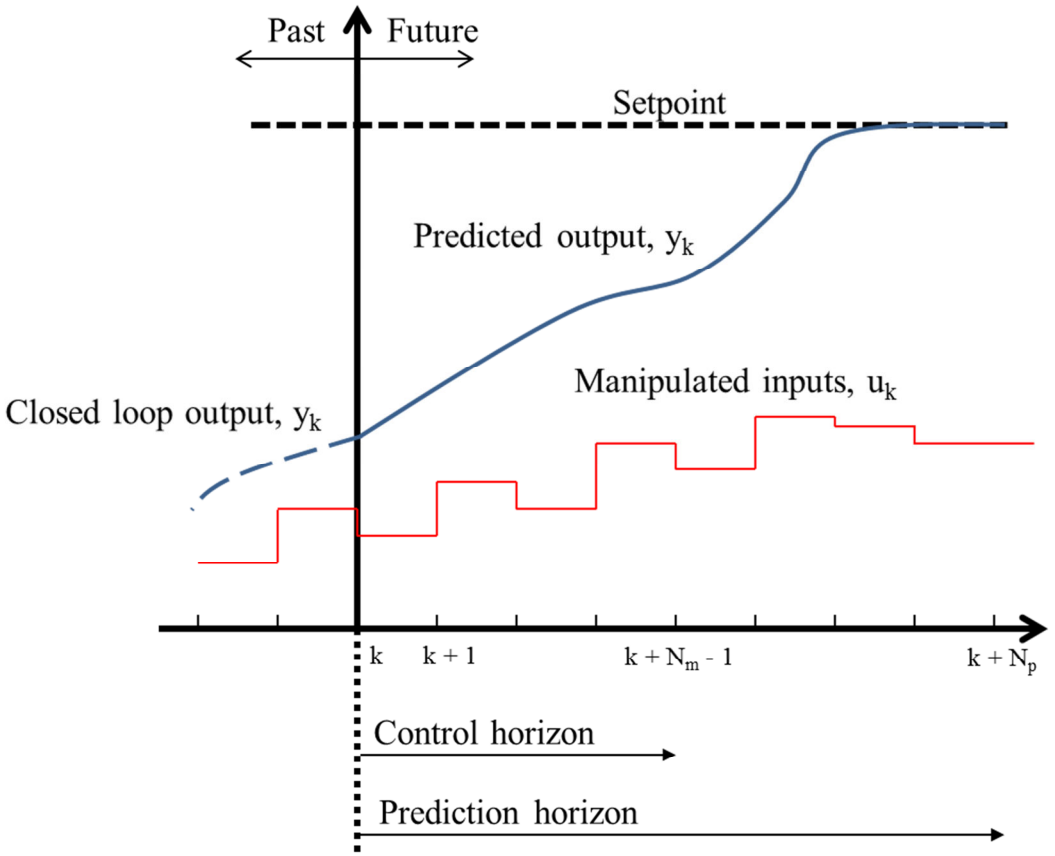

Figure 1.1. Basic concept of model predictive control (MPC), demonstrating how at time step k, the predicted output (y_k) is calculated over the prediction horizon (N_p), to calculate the optimal input sequence (u_k). This process is repeated at every time step.

1.2.2 State Estimation

In a deterministic system, free of noise and model mismatch, a robust model could provide perfect prediction of the future plant state. However, a more realistic assumption is that the plant and model are not the same, and in general the model parameters have some uncertainty built into them. In addition, disturbances can enter the controller in the form of input or output noise. To account for the plant-model mismatch and disturbances, an augmented model is written as,

$$\begin{aligned} x_{k+1} &= f(x_k, u_k, d_k) \\ d_{k+1} &= d(k) \\ y_k &= h(x_k, u_k, d_k) \end{aligned} \tag{1.4}$$

where $d_k \in \mathbb{R}^{n_d}$ is the disturbance model added to the original model in equation (1.1). In this augmented form, disturbances enter both in the input and output. The disturbance is often assumed to be additive and expressed as zero mean white noise. The goal now is to estimate what the disturbance using a state observer or state estimator, preferably an optimal one, to achieve an offset free controller.

A state estimator is an important component of a modern model-based control system [6]. Most modern control techniques (e.g. MPC) require full state information at each sample time, which is often not available in chemical or bioprocesses. The state estimator or state observer can be used to convert available measurement information into an estimate of the whole state based on known dynamic relationships. The Kalman filter (KF) is the most popular state estimation technique, as it provides optimal exponential observations by minimizing the mean square estimation error. However, the Kalman filter is designed for linear systems. For nonlinear systems the extended Kalman filter (EKF) was developed, though it is suboptimal to the KF. The EKF uses a first-order linearization along the state estimate trajectory to compute the evolution of the covariance matrix. Another well-known state observer for bioprocesses is the asymptotic observer. The asymptotic observer does not rely on perfect knowledge of the model and can be utilized to make estimates with unknown reaction kinetic. The use of such observers in the context of bioprocesses has been reviewed [20].

1.3 APPLICATION OF RAMAN TO CONTINUOUS, ONLINE MEASUREMENTS

One of the main setbacks of utilizing advanced process control (APC) in bioprocesses has been the lack of suitable in vitro measurement technology. With the development of chemometric methods and Raman technology, we now are at a point where this paradigm could change. Using

online Raman measurements of key process parameters will allow for the development of more robust system models, which in turn can be used for prediction and optimization.

The Raman effect was discovered in 1928 and is observed upon irradiating solid, liquid or gaseous samples with monochromatic light and measuring the inelastic scattering that is exhibited by 10^{-6} of the incident photons [21]. Unlike similar spectroscopic methods such as near infrared (NIR), water is virtually invisible to Raman, which allows its use in aqueous solutions [22]. In addition, Raman spectra have narrower peak widths and cover a wide spectral range, improving the potential for analyzing complex matrices of compounds [23]. Benefits of using a spectroscopic method are its speed, lack of sample preparation, and ability to scale from micro scale up to production scale. The diversity of compounds present in lignocellulosic processes calls for an analytical method that can simultaneously measure multiple compounds with high sensitivity at both high and low concentrations. Other methods such as HPLC, UV-Vis spectroscopy, capillary electrophoresis, gas chromatography/mass spectrometry and electrochemical methods are unable to quantify multiple compounds simultaneously and/or require destructive and time-consuming sample preparation and analysis [23].

The primary advantages of using Raman, or other spectroscopic methods, over traditional analysis methods are increased speed of analysis and a reduction in user interaction. With online Raman spectroscopy, a full spectrum can be collected in less than a minute to provide immediate data regarding the concentration of any compounds of interest, once appropriate models have been developed. This real-time, continuous output of information enables application of advanced process controls. While the use of Raman spectroscopy as an analytical technique is widely used in other disciplines, its application for continuous monitoring of fermentation of lignocellulosics is still being developed. Small scale ethanol fermentations have been measured

by Raman spectroscopy through periodic removal and measurement of samples [24], and continuous monitoring of nano-scale reactions with in situ measurement [25] has been conducted.

The method used for sample collection is an important consideration both for building a robust process sensor and for improving the quality of data. Following excitation with a laser beam, the Raman scattered light is filtered and collected on a detector. A Raman microscope or immersion probe can be used to measure individual samples, or to continuously monitor a reaction by pumping the reaction mixture through a sampling loop. A probe can also be immersed in the reaction solution to enable continuous remote sampling. Such probes are typically composed of a fiber optic portion connected to a stainless steel probe fitted with a window at the tip. Compared to a standard probe with a flat window, the use of a sapphire-tipped hemispherical ball probe tip reduces the focal length and increases the number of Raman events that can be measured [26]. In reactions subject to high levels of fluorescent interference, the reduced focal length results in less noise and higher quality data.

1.3.1 Challenges of using Raman to measure lignocellulosic bioprocesses

While Raman spectroscopy has the distinct advantage of being practical in aqueous solutions, it is subject to interference by other means. Such interference falls into two categories – biological and chemical. Both cause an elevated spectral background signal that can mask the desired Raman features of the spectra representing the compounds of interest. Biological interference is present in any system containing living cells, and results from the heterogeneous nature of biological systems. The second means of interference is the influence of compounds present in the media, even in the absence of any cells. In lignocellulosic fermentations, these are often represented by lignin-derived compounds. Lignin is made up of highly conjugated phenolic

groups that fluoresce when excited by the laser [27]. In addition to process interference, measurement noise caused by instrumentation, background light sources, and cosmic rays must also be filtered before deducing the true spectra. Figure 1.2 shows how typical Raman spectra and its constituent components are broken down.

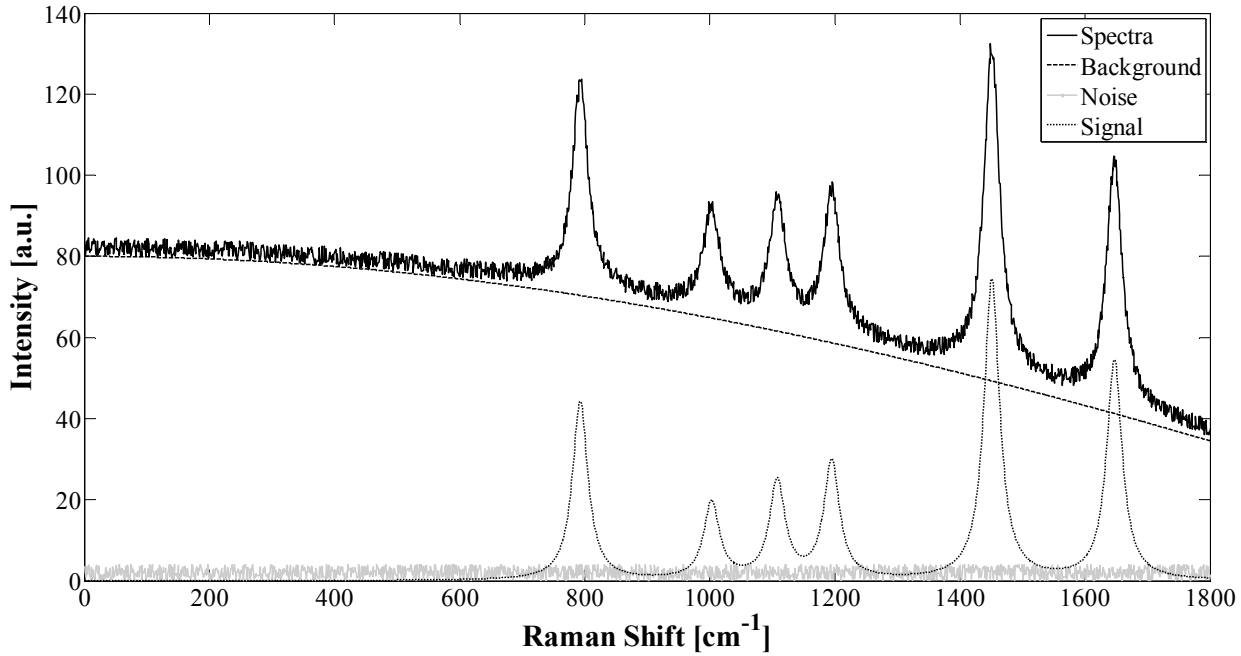

Figure 1.2. Components of simulated Raman spectra. The Spectra = Signal + Background + Noise. The goal is to reduce the spectra to its true signal.

The impact of this background interference can be minimized in a number of ways. Cells can be physically removed from the media prior to measurement by means of a filter [24] and chemical interference can be minimized by using low fluorescence medium [25]. Alternatively, many different chemometric methods have been developed to mathematically remove the fluorescent background and noise from the collected spectra and reveal the spectral components of the compounds of interest [28], [29]. In addition, the choice of excitation wavelength is important. In solutions without the presence of fluorophores, lower wavelength visible light such as 514 nm provides high sensitivity but greatly increased fluorescence [25], [30]. Higher wavelength, near-IR excitation, may be used to minimize fluorescence, but at the expense of

peak intensity. For lignocellulosics, 785 nm wavelength is most often utilized [23], [31], [32], but 993 nm and 1064 nm have also been used to further reduce the fluorescent background [33], [34].

1.3.2 Quantification of Raman data

Robust models must be constructed to use the large amounts of spectral data generated in a given process and to convert Raman signal intensities into units of concentration. Unlike chromatographic methods, any compound measured by Raman can present multiple peaks, many of them overlapping. In addition, some compounds are strong scatterers, presenting high intensity peaks, while others scatter weakly. Ethanol, for example, presents a strong, well-isolated C-C bond stretching peak at 883 cm^{-1}. This single peak allows for univariate analysis of ethanol by measuring the peak area of a number of samples and comparing the areas to the concentration of the same samples measured by a validation method such as HPLC [34]. This calibration model is then validated by predicting the concentration of a second set of spectra, and comparing the predicted values to the actual values. For weaker scatterers or complex mixtures of samples with overlapping peaks, multivariate methods must be employed. Similarly, in the presence of a high background, even strong scattering compounds at low concentrations may require multivariate analysis [23]. Typically these models are constructed using partial least squares (PLS) methods.

The goal of PLS is to predict a set of dependent variables, Y, from a set of independent variables, X. In the case of developing a model for Raman spectra for lignocellulosics, the independent variables are the Raman spectra and the dependent variables are concentration of substrates and products usually measured on an HPLC. A PLS algorithm is then implemented to simultaneously decompose these variables into scores, T and U, loadings, P and Q, and a matrix

of error, *E*, with the goal of explaining as much as possible of the covariance between *X* and *Y*. A linear relationship is established according to:

$$X = TP^T + E_X = \sum_{i=1}^{I} t_i p_i^T + E_X \tag{1.5}$$

$$Y = UQ^T + E_Y = \sum_{i=1}^{I} u_i q_i^T + E_Y \tag{1.6}$$

$$U = bT \tag{1.7}$$

where $b = U^T T / T^T T$ is the regression coefficient vector of the linear model.

1.3.3 Use of Raman spectroscopy to measure fermentation of sugars to ethanol

The conversion of glucose to ethanol is one of the oldest forms of fermentation and has been a widely studied application of Raman spectroscopy for continuously measuring fermentation. During fermentation with *Saccharomyces cerevisiae,* measurable components include ethanol, glycerol and the cell concentration, all of which can be quantified to measure productivity, stress, and process efficiency. Shaw et al [24] used a 780 nm laser to measure the progress of a fermentation of pure glucose by continuously pumping the fermentation solution through a quartz cell in a Raman microscope. Samples were filtered through a 0.2 μm filter prior to analysis every 0.5 to 1 hour, removing much of the background fluorescence from the cells. A multivariate PLS model developed from this fermentation was then used to predict glucose and ethanol concentrations in a different fermentation with a root mean square error of prediction (RMSEP) of 8.3% and 1.8% for glucose and ethanol, respectively [24]. Ávila et al [35] developed a multivariate PLS model based on measurements of glucose, ethanol, glycerol and cell concentrations by 785 nm Raman taken at 15 minute intervals through a flow cell pumped continuously from the fermentation reactor without any filtration. Their model found RMSEP

values of 0.5%, 0.3%, 0.02% and 1 g L^{-1} for glucose, ethanol, glycerol and cell concentration respectively. The limit of detection (LOD) can be calculated by doubling the RMSEP.

The use of Raman spectroscopy to follow the progress of reactions using pretreated lignocellulosic material is just beginning. Initial work by Shih and Smith [32] measured the progress of ethanol production from the enzymatic hydrolysate of both dilute acid and ammonia pretreated corn stover using *S. cerevisiae* by analyzing offline aliquots of the fermentation broth. In the pretreated corn stover hydrolysate, LOD's for glucose and ethanol were determined to be 8 and 6 g L^{-1} respectively based on quantification of peaks at 1128 cm^{-1} and 883 cm^{-1} for glucose and ethanol respectively.

Ewanick et al. at the University of Washington continuously measured the fermentation of both pure glucose in water and a switchgrass-derived lignocellulosic hydrolysate [31]. Measurement of both reactions was accomplished by means of an immersion ball probe inserted in a fast loop circulating media continuously from the fermentation reactor, with measurements taken every minute. Both reactions required mathematical removal of the fluorescent background through use of a polynomial fitting algorithm. In the pure glucose fermentation most of the background could be attributed to the cell biomass. The background was four times greater in the lignocellulosic fermentation due to lignin-derived components of the switchgrass hydrolysate. With proper signal processing, however, this additional background in the lignocellulosic fermentation only increased the glucose limit of detection (LOD) from 1 to 2 g L^{-1} and the ethanol LOD from 0.8 to 1.2 g L^{-1}. Both ethanol and glucose were modeled based on principle component analysis (PCA) of the spectrum and PLS comparison of the main spectral components to HPLC data. These results showed for the first time that Raman could successfully follow the progress of fermentation of a highly fluorescent lignocellulosic liquid stream [31]. By

collecting data continuously and comparing it to HPLC, data, a model was developed with a low LOD compared to previous research.

Raman spectroscopy will continue to be investigated as a means to measure the progress of both enzymatic hydrolysis and fermentation. In particular, monitoring of simultaneous saccharification and fermentation (SSF) will prove highly valuable to improve the efficiency of a widely used industrial process [36]. Raman monitoring of SSF of corn mash (starch-based) has already been successfully proven and shown to be able to measure not only monomeric glucose, but oligomeric and polymeric glucose as well as ethanol [34]. Application of these techniques to lignocellulosic hydrolysis and SSF will prove more challenging, but holds the potential to reduce reaction time and cost by increasing process efficiency. Future work at the University of Washington related to Raman spectroscopic measurement of lignocellulosic biorefinery processes will focus on simultaneous quantification of multiple sugars, particularly xylose, and on alternative microbial products such as acetic acid. The use of pretreatment hydrolysates from diverse feedstocks and different pretreatment processes will introduce new combinations of fluorophores that may affect the Raman measurements and require new models and techniques to reduce the fluorescent background.

1.3.4 Characteristic Raman bands for fermentation of lignocellulosic components

A review of the literature regarding Raman spectroscopic methods for fermentation of bio-based substances shows common Raman bands (cm^{-1}) can be found. For an exhaustive list of Raman bands for biological components the reader is referred to [37]. Most researchers find a strong visible peak for ethanol at approximately 880 cm^{-1} representing C-C stretching. Often this peak is used to monitor the progress of fermentation, as it is a strong peak that can be easily deciphered from the data. While other peaks for ethanol, such as those at the region near 2900

cm^{-1} are often disregarded since most organic compounds present similar bands [38]. Similarly, glucose has a strong distinguishable peak around 1120 cm^{-1}, which in turn can be followed. However, when dealing with multiple sugar systems there exists overlapping Raman bands from various sugars due to spectral interference between peaks making isolation difficult. For fermentation of sugars the bands between 300 – 1500 cm^{-1} are rich with structural information. Raman bands in the 920 – 960 cm^{-1} region are strong from α-1,4 glycosidic linkages in glucose oligomers and polysaccharides [34], [38]. As the fermentation progresses, the substrate is consumed and metabolites are produced showing major peaks disappearing and being formed, but generally chemometric methods are still required to deconvolute these systems into corresponding components of interest.

1.4 KINETIC MODELING OF BIOPROCESSES

In addition to sensor development, deriving appropriate bioreactor system models is critical for model predictive control. Linear systems have been the focus of control system design for decades, and there exists well defined methods for determining appropriate controller design in such cases. However, when dealing with real processes, most often than not, nonlinear behavior is apparent and a linear model approximation can be inadequate to describe key process dynamics. To capture the inherent nonlinearity of a process sufficiently, it is often necessary to use a first principle modeling approach, which typically leads to a set of nonlinear differential or nonlinear differential algebraic equations. In addition, chemical reactions are continuous in nature, whereas implementation of a control system through the use of a computer requires a discrete model. The discretization of a nonlinear model requires an approximation to the nonlinear system of equations, which often is based on some form of Taylor series expansion to approximate nonlinear terms. This can increase computational complexity as suitable numerical

algorithms are required to integrate the model solution. Thus, while accuracy of the model is important, one must often make trade-offs between model accuracy and computational simplicity for its practical implementation in a control system.

Mechanistic models for fermentation systems are derived from theoretical mathematical formulation of key internal biochemical system dynamics. The goal of a mechanistic model is to be able to predict key components of bioprocess states, which usually consist of cell mass, or biomass (not to be confused with the term biomass that describes lignocellulosic material), substrate(s), and major product(s), under specified conditions. These models can be classified as segregated or unsegregated, and structured or unstructured. An unsegregated or continuum model, assumes a homogeneous reactor environment where the average cell behavior can be used to predict the kinetics of the system, thus the cell population is treated as a continuum or lumped biophase [39]. A segregated or corpuscular model, considers the behavior of individual cells in a heterogeneous environment, where each cell in a population will have different characteristics at any point in time during the process (e.g. age, cell mass, viability). Furthermore, this system can be broken down into structured, where biomass is described by several mechanistic variables, or unstructured, where a single variable describes the biomass.

A general dynamic model for bioreactors, which can be used to describe multiple microorganism, substrate, and product concentration can be written as [40],

$$\dot{\xi} = K\phi(\xi, t) - D\xi - Q(\xi) + F \tag{1.8}$$

where ξ represents a vector of bioreactor states; ϕ are reaction rates, which are usually dependent on the current state and dynamics; K are reaction rate constants, such as product yield; D is the dilution rate, describing the influent flow and outfluent flow of the system based on a constant volume; Q represents gaseous outflow of any products, such as, carbon dioxide in yeast

fermentation; biochemical kinetics of the process are described within the first term, $K\phi(\xi, t)$; and the remaining terms describe the transport dynamics of the components through the bioreactor. Functions ϕ are the rate equations which may depend on several state and environmental variables.

From the general dynamic model, one can develop a series of relationships describing various physical behaviors of biprocesses. The kinetics of cellular growth accounts for majority of the system nonlinearity. A number of growth kinetic terms have been presented in the literature, but by far the most utilized of these is based on the model developed by Monod in 1956 [41]. The manner in which a microorganism grows on substrate can further complicate the model. Diauxic growth describes the consumption of two or more substrates in sequence, where additional substrates are only utilized after complete exhaustion of the preferred substrate [42].

1.5 OBJECTIVES AND HYPOTHESIS

The overall objective of this body of research is the implementation of advanced process control, modeling, and monitoring for the production of valuable biochemicals from carbohydrates. This is achieved by implementing a model predictive control strategy that can handle nonlinear systems with Raman spectroscopy providing online measurements for output feedback. Specific research objectives include:

(1) Development of a first principal kinetic model that describes the fermentation of glucose, xylose and a mixture of both sugars to acetate during batch fermentation with acetogenic bacterium *Moorella thermoacetica*.

(2) Development of a strategy to utilize online Raman measurements to monitor the conversion of lignocellulosic hydrolysate to acetate with the microorganism *Moorella thermoacetica*.

(3) Development of a nonlinear model predictive control (NMPC) strategy to stabilize the bioreactor under continuous fermentation for setpoint tracking of productivity of valuable end-products, while maximizing total substrate conversion.

Chapter 2. Kinetic Modeling of *Moorella thermoacetica* Growth on Single and Dual Substrate Systems

To be submitted for publication in an international journal.

Abstract

Acetic acid is an important chemical raw material that can be produced directly from sugars in lignocellulosic hydrolysates. Development of kinetic models that capture the bioconversion dynamics of multiple sugar systems will be critical to optimization and process control in future lignocellulosic biorefinery processes. In this work a kinetic model was developed for the single and dual substrate conversion of xylose and glucose to acetic acid using the acetogen *Moorella thermoacetica*. Batch fermentations were performed experimentally at 20 g L^{-1} total sugar concentration using synthetic glucose, xylose, and a mixture of glucose and xylose at a 1:1 ratio. The product yield, calculated as total product formed divided by total sugars consumed, was 79.2%, 69.9%, and 69.7% for conversion of glucose, xylose, and a mixture of glucose and xylose (1:1 ratio), respectively. During dual substrate fermentation *M. thermoacetica* demonstrated diauxic growth where xylose (the preferred substrate) was almost entirely consumed before consumption of glucose began. Kinetic parameters were similar for the single substrate fermentations and a strong linear correlation was determined between the maximum specific growth rate $\boldsymbol{\mu_{max}}$ and substrate inhibition constant, $\boldsymbol{K_s}$. Parameters estimated for the dual substrate system demonstrated changes in the specific growth rate in particular changed the most. Thus, kinetic growth is affected when multiple substrates are present in a fermentation system and models should be developed to reflect these features.

2.1 INTRODUCTION

The sustainable production of chemicals and fuels from renewable feedstocks is of global importance. Lignocellulosic biomass is the basis for second generation fuels and chemicals is expected to tremendously impact the renewable fuels industry. Organic acid production is one such area that could benefit from microbial fermentation of lignocellulosic hydrolysates [43]. Acetic acid is an important chemical raw material used in the production of chemicals such as, vinyl acetate monomers, acetic anhydride, and ethyl-esters. Approximately eleven million tons per year of acetic acid is produced, with about 190,000 tons coming from microbial processes. Commercial facilities are currently being demonstrated for the production of acetic acid from hybrid poplar in the Pacific Northwest [44]. Advances in the processing of lignocellulosic

feedstocks could provide a significant increase in the sustainable production of acetic acid and other organic acids.

Lignocellulosic biomass, such as switch grass or hybrid poplar, has three major components in cellulose, hemicellulose, and lignin. The goal of pretreatment for lignocellulosic material is to breakdown these components to access carbohydrates, which can subsequently be converted into higher value products. The most commonly used microorganism in the biofuels industry is *Saccharomyces cerevisiae,* which can convert hexoses to ethanol, but other organisms have been used to produce xylitol, acetic acid, butyric acid and many other products [45]. Acetogenic bacteria, or acetogens, are anaerobes that use the acetyl-CoA pathway (or more commonly called the Wood-Ljungdahl pathway) for the reduction of CO_2 to acetic acid or acetate. *Moorella thermoacetica* (formerly known as *Clostridium thermoaceticum*) was one of the first acetogens to be studied [46]. *Moorella thermoacetica* is a rod-shaped, gram positive, anaerobic, thermophilic, homoacetogenic bacteria that is both autotrophic and heterotrophic, allowing it to consume a wide range of substrates including: carbohydrates, gases, alcohols, organic acids, and methoxylated aromatic compounds. It was first noticed for its ability to convert glucose to acetate at nearly 100% theoretical mass yield. From an industrial perspective this organism is attractive for its high carbon conversion efficiency, in addition to being both anaerobic and thermophilic, which could potentially lower fermentation cost compared to common aerobic fermentation. In order to utilize all the sugars, microorganisms with the ability to convert both hexose and pentose carbohydrates are critical for the efficient and economical utilization of lignocellulosic hydrolysate.

Organisms that consume multiple sugars often exhibit diauxic behavior, preferential consumption of a particular substrate prior to another [42]. This catabolic repression can change

the characteristics of fermentation compared to single substrate systems, such as increases or decreases in cellular growth rate, which ultimately leads to shorter or longer residence times. The dynamic behavior of such a process can be further complicated by the bioreactor operation mode (e.g. batch, fed-batch, or continuous), where substrate feed concentrations could potentially be varied by upstream processing and feedstock irregularities. Kinetic models are employed to understand how upstream bioconversion processes affect fermentation, since experimental studies are time consuming and costly. In addition, these models can be utilized to develop operation and process control strategies for bioconversion.

Kinetic models for fermentation systems can be experimental, mechanistic, or a combination of the two. Mechanistic models, based on fundamental mass/energy balances, can be classified as segregated or unsegregated, and structured or unstructured. Unsegregated-unstructured models assume a homogeneous reactor environment where the average cell behavior can be used to predict the kinetics of the system with minimal kinetic parameters, and are most commonly found in literature [39]. However, for fermentation systems of multiple substrates, unsegregated-structured models and in some cases segregated-structured have been employed. Segregated models consider the behavior of individual cells in a heterogeneous environment, where each cell in a population will have different characteristics at any point in time during the process (e.g. age, cell mass, viability), and structured models describes cellular growth by several mechanistic variables. These models are often computationally much more complex, which makes their implementation often cumbersome and in some cases impossible to solve in a manageable time frame for optimization studies. There has been relatively minimal research in developing kinetic models for multiple substrate systems using simple unsegregated models.

To be useful for modeling lignocellulosic hydrolyzates, kinetic models of microorganism growth will need to include multiple substrates consumption, as well as additional inhibition terms to accurately reflect the system. The aim of this research is to develop a mechanistic model for fermentation of glucose and xylose, main components of lignocellulosic hydrolyzate, to acetic acid using the acetogen *M. thermoacetica*. The parameters for the model were estimated using experimental data obtained from batch fermentation of single substrate consumption of glucose and xylose, and a mixture of glucose and xylose at a 1:1 ratio.

2.2 KINETIC MODEL DEVELOPMENT

2.2.1 Batch kinetic model

An unstructured-unsegregated model was developed for batch fermentation of glucose, xylose, and a mixture of glucose and xylose to acetic acid. Several assumptions were used in the model development including: the medium is well-mixed and homogenous, due to the small volume used in experimentation; temperature and pH are held constant during fermentation; a maintenance term is neglected due to the growth rate being much greater; product formation is assumed to be only growth associated, shown experimentally.

Given the aforementioned assumptions, a model for batch growth of *M. thermoacetica* on carbohydrates can be written as,

$$\frac{dX}{dt} = r_x X \tag{2.1}$$

$$\frac{dS_1}{dt} = -r_{s_1} X \tag{2.2}$$

$$\frac{dS_2}{dt} = -r_{s_2} X \tag{2.3}$$

$$\frac{dP}{dt} = r_p X \tag{2.4}$$

where X, S_i, and P, are the biomass, substrate ($i = 1$ for xylose, and $i = 2$ for glucose) and product concentration, respectively, with kinetic rates of microorganisms growth (r_x), substrate consumption (r_{s_i}), and product formation (r_p).

For this work, growth kinetics include both substrate and product inhibition characteristics, and catabolic repression is based off the work of [47],

$$r_x = [\mu_1 + \mu_2] - K_d \tag{2.5}$$

$$\mu_1 = \frac{\mu_{max,1} S_1}{K_{s_1} + S_1}\left(1 - \frac{P}{P_{max}}\right) \tag{2.6}$$

$$\mu_2 = \frac{\mu_{max,2} S_2}{K_{s_2} + S_2}\left(\frac{1}{1 + K_1 S_1}\right)\left(1 - \frac{P}{P_{max}}\right) \tag{2.7}$$

$$\mathrm{r}_{s_1} = \frac{\mu_1}{Y_{xs,1}} \tag{2.8}$$

$$\mathrm{r}_{s_2} = \frac{\mu_2}{Y_{xs,2}} \tag{2.9}$$

$$\mathrm{r}_p = \mu_1 Y_{px,1} + \mu_2 Y_{px,2} \tag{2.10}$$

where μ_i is the specific growth rate of the biomass of component i ($i = 1$ for xylose, and $i = 2$ for glucose); K_d is the rate of cell death (considered constant); $\mu_{max,i}$ is the maximum growth rate of cells; $K_{s,i}$ is the substrate saturation constant; P_{max} is the maximum acetic acid concentration before 90% inhibition; $Y_{xs,i}$ and $Y_{px,i}$ are the cell yield and product yield constants, respectively; and K_i is an inhibitory constant of the substrate causing catabolic repression (if no catabolic repression exists, then $K_i = 0$). The growth rate expressions described in equations (2.5) – (2.7) include four main terms: a Monod growth component with saturation kinetics, where growth is inhibited by low substrate concentration; a catabolic repression term where the glucose is inhibited by xylose; an end-product inhibition term where growth is reduced as product concentration increases; and a cell death term.

2.2.2 Parameter estimation

Parameter estimation is found by minimizing the chi-squared error between the model predictions and experimental data from the batch system. This is determined through minimization of the following objective function,

$$\chi^2(\theta) = min \sum_{j=1}^{n} \left[\frac{\left(X_j - X_{m,j}\right)^2}{\sigma^2} + \frac{\left(S_{1,j} - S_{1,m,j}\right)^2}{\sigma^2} + \frac{\left(S_{2,j} - S_{2,m,j}\right)^2}{\sigma^2} + \frac{\left(P - P_{m,j}\right)^2}{\sigma^2} \right] \quad (2.15)$$

where θ is the vector of kinetic parameters; X_j, $S_{i,j}$, and P_j are experimental data of biomass, substrate ($i = 1$ for xylose, and $i = 2$ for glucose), and product concentrations at sampling time j; $X_{m,j}$, $S_{m,j}$, and $P_{m,j}$ are the concentrations computed by the model in each sampling time; and σ^2 is the variance between experimental replicates for total of samples n. The parameters are subjected to upper and lower bounds to ensure realistic values,

$$l_q \leq \theta_q \leq u_q$$

where θ_q is parameter q, $l_q$is the lower bound for parameters, and u_q is the set of upper bounds on the parameters.

2.3 MATERIALS AND METHODS

2.3.1 Microorganism cultivation

Moorella thermoacetica (ATCC 39073) was acquired from the American Type Culture Collection (Manassas, VA) grown in medium containing in gL^{-1} of yeast extract, 10; glucose/xylose/1:1 ratio 20; L-Cysteine Hydrochloride Monohydrate 1.0; KH_2PO_4 7.0; $NaHCO_3$ 16; K_2HPO_4 5.5; $(NH_4)_2SO_4$ 1.0; $MgCl_2$-$6H_2O$, 0.33; $CaCl_2$ 0.05; $NaMoO_4$-$2H_2O$ 0.0025; $Fe(NH_4)_2SO_4$-$6H_2O$ 0.04; $Co(NO_3)_2$-$6H_2O$ 0.03; Na_2SeO_3 0.0002; $NiCl_2$-$6H_2O$ 0.0002. The pH of the medium was adjusted to pH 7.4 using 5M NaOH. Inoculants were grown at 58°C in

stoppered and crimp-sealed 150 mL Wheaton serum bottles sparged with filter-sterilized 100% CO_2, inoculated with 10% by volume of a 36 hr culture (Total volume of inoculant 100 mL). Cells were grown in the media through three transfers before they were used for analyses of growth yields and kinetic model [48].

2.3.2 Experimental conditions

Batch fermentations were performed in quadruplicate using a 1.3 L Bioflo 115 (New Brunswick Scientific Co., Inc., Edison, NJ) bioreactor with 500 mL of working volume under anaerobic conditions. Sterilization of the fermenter vessel containing carbohydrates was performed in an autoclave at 121^{o}C for 30 min. Deoxygenated CO_2 was filtered and sparged at a flow rate of 0.5 L min^{-1} into the bottom of the vessel continuously. Temperature of the bioreactor was held at 58^{o}C and pH 6.8. A 5M buffer of sterilized NaOH was added to maintain pH throughout fermentation. Inoculation was performed with 10% v/v of three-stage seeded organism grown on glucose, xylose, or glucose and xylose in 1:1 ratio. Four batch fermentations were performed for approximately 20 g L^{-1} initial concentrations of glucose, xylose, and a mixture of glucose/xylose at a 1:1 ratio (10 g L^{-1}, 10 g L^{-1}) each.

2.3.3 Growth, acetic acid and carbohydrates analysis

A 1 mL sample was taken from the fermenter at specified time points and used for analysis. The sample was centrifuged for 5 min at 10,000 rpm, removing the supernatant for High-Performance Liquid Chromatograph (HPLC) and then the pellet was re-suspended in 1mL of de-ionized water. The supernatant was then put through a 0.2 micrometer filter to remove any remaining cells. The cell free supernatant was used for off-line determination of sugars (glucose and xylose), and acetic acid in a Shimatzu Prominence LC HPLC with Phenomenex RHM

monosaccharide H+ (ion exclusion) column. Bacteria concentration was determined by optical density (OD) at 600nm. The OD was converted to cell dry mass following NREL protocol [49].

2.4 RESULTS AND DISCUSSION

2.4.1Fermentation of single sugars

Fermentation of single sugars (glucose and xylose) was conducted under anaerobic conditions in quadruplicates and monitored for 48 hours. Sample points for the replicates were taken at the same time points. *M. thermoacetica* consumed all glucose (initial concentration of 18.4 g L^{-1}), and xylose (initial concentration of 17.4 g L^{-1}) in under 40 hours of total fermentation time. The product yield, calculated as total product formed divided by total sugars consumed, was 79% and 70% for glucose and xylose, respectively. Acetic acid was the only major metabolite found and its production was mainly associated with cell growth for both substrates. The maximum volumetric production rate was 0.4 and 0.3 g L^{-1} h^{-1} for glucose and xylose, respectively. Maximum cell mass was approximately 16.5 g L^{-1} for both xylose and glucose. The characteristic growth curve of the organism was similar from batch-to-batch and for both sugars. This indicates that *M. thermoacetica* is a stable microorganism that can achieve predictable behavior under similar growth conditions for both glucose and xylose. Experimental data was broken into calibration and validation data sets, where 75% of the collected data was used for calibration and 25% was used for validation. Results from single substrate fermentation experiments and associated prediction model are shown in Figure 2.1. The model shows the model predictions compared to the validation data set.

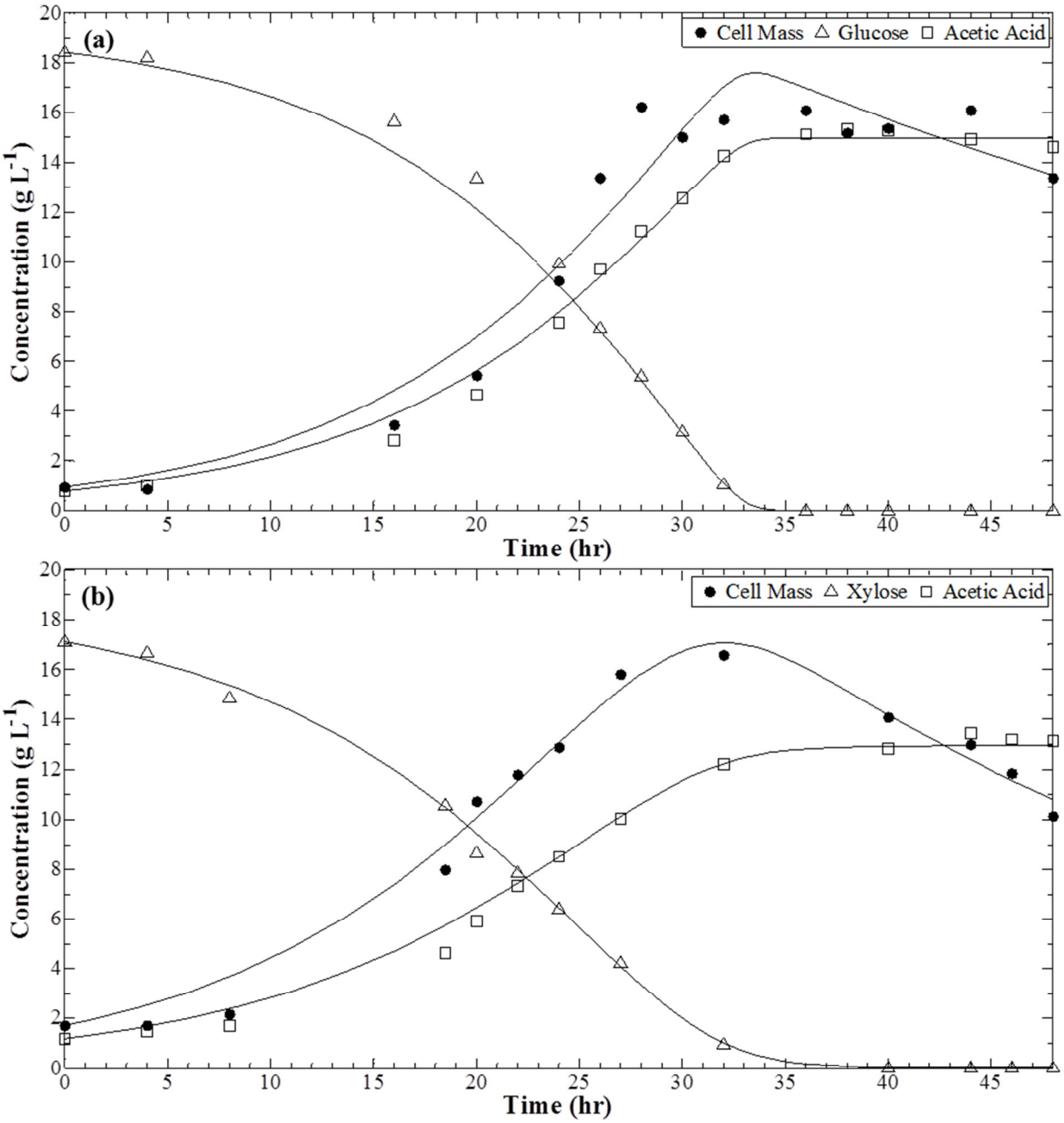

Figure 2.1. Comparison of model prediction (solid lines) and experimental data for batch culture of M. thermoacetica (a) on glucose, and (b) on xylose.

2.4.2 Fermentation of multiple sugars

The dynamic behavior of dual substrate system differed from the single sugar fermentations. The growth characteristics are representative of catabolic repression, with xylose repressing the

utilization of glucose. Figure 2.2 shows the growth, acetic acid production and substrate utilization by *M. thermoacetica* grown on a mixture of glucose (~9.9 g L^{-1}) and xylose (~9.0 g L^{-1}). The initial growth rate was slower while both sugars were present compared to a single substrate fermentation. Consumption of glucose does not begin until the preferred substrate, xylose, is over 70% (~24h) utilized. However, once the xylose was entirely consumed the growth on glucose was much faster increasing the maximum cell concentration to 15.4 g L^{-1}. This is in agreement with previous studies showing fermentation with mixed sugars [48], [50]. The average yield for the dual substrate fermentation was similar to single substrate fermentation at 70%.

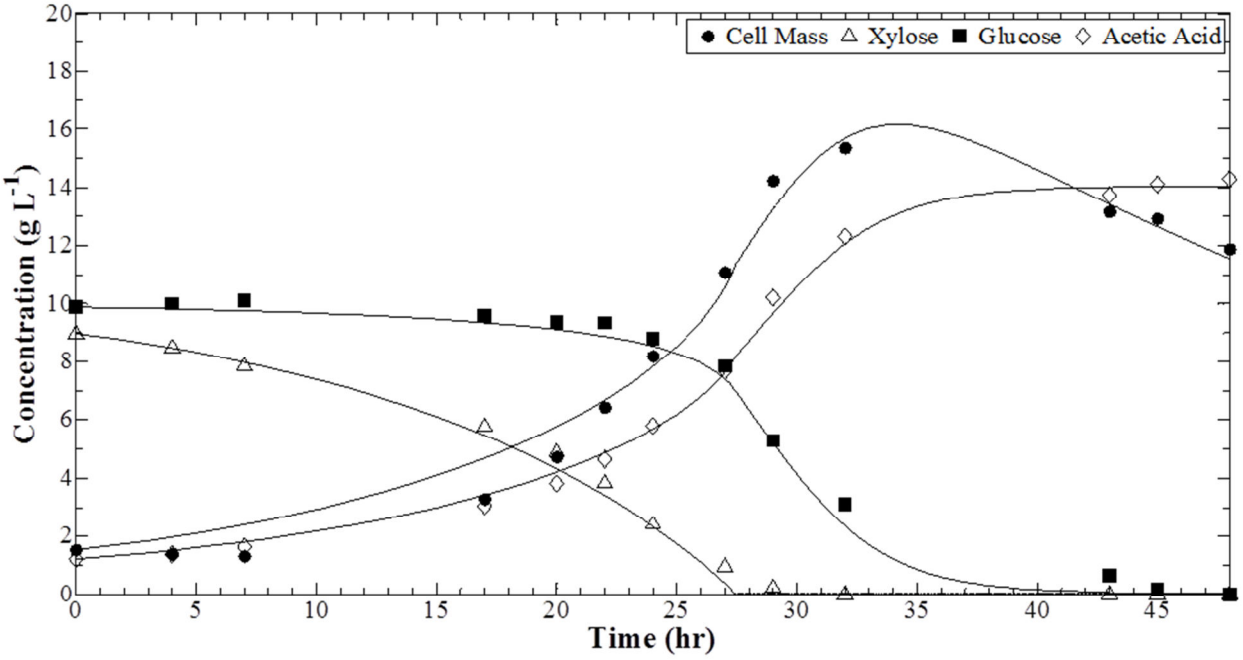

Figure 2.2. Comparison of model prediction (solid lines) and experimental data (markers) for batch culture of M. thermoacetica a mixture of glucose and xylose (1:1 ratio).

2.4.3 Kinetic model estimation and identification

For the single substrate models five parameters were estimated from experimental data and these values are shown in Table 2.1 along with their 95% confidence interval. The maximum acetate concentration has been well documented in literature to be in the range of 45 – 60 g L^{-1}, for our

model P_{max} is held constant at 48 g L^{-1} [51]. The parameters were ranked from least identifiable to most identifiable based an orthogonalization algorithm [52]. The results show that the value of K_s is the least identifiable parameter. The correlation matrix indicates the linear dependence of parameters upon each other; these values are shown in Table 2.1. A strong correlation (value close to 1.0) reveals that the value of one parameter has a strong dependency on the value of another. All parameters show high correlation with at least one other parameter. The root mean squared error (RMSE) was calculated as,

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n}\left(\hat{y}_j - y_j\right)^2}{n}} \tag{19}$$

which demonstrates the spread between predicted and experimental data. Both single substrate models show good agreement between model and data.

Table 2.1. Kinetic parameters for growth and production from batch culture of *M. thermoacetica* on single-substrates with 95% confidence intervals. Sensitivity ranking of most sensitive to least sensitive parameters based on work from [52]. Linear correlation matrix values are given for the estimated parameters.

Parameters	**Glucose**	**Xylose**	**Rank**	**Correlation Matrix**				
	$\theta \pm 95\%\ CI$	$\theta \pm 95\%\ CI$		μ_{max}	K_s	K_d	Y_{xs}	Y_{px}
RMSEC	0.57	0.74						
RMSEP	0.92	0.93						
$\mu_{max}\ (h^{-1})$	0.13±0.02	0.15±0.02	3	1.00				
$K_s\ (g\ l^{-1})$	0.61±1.26	2.30±1.34	1	0.93	1.00			
$K_d\ (h^{-1})$	0.02±0.01	0.02±0.01	5	0.57	0.24	1.00		
$Y_{xs}\ (g\ g^{-1})$	1.15±0.16	1.53±0.15	4	0.61	0.30	0.95	1.00	
$Y_{px}\ (g\ g^{-1})$	0.67±0.09	0.45±0.04	2	-0.54	-0.25	-0.86	-0.91	1.00
$P_{max}\ (g\ l^{-1})$	48.0	48.0						

The maximum growth rates estimated by the model were 0.13 and 0.15 h^{-1} for glucose and xylose, respectively. This agrees with previous literature estimates that estimated maximum growth rates between 0.1 – 0.22 h^{-1} [51], [53]. Product formation can be seen to be largely growth associated, which confirms previous assumptions. The yield constants, Y_{px} and Y_{xs}

estimated from the nonlinear parameter estimation were also in good agreement with experimentally derived product yield constant as calculated by the effective product yield, ($Y_{ps} = Y_{px} \times Y_{xs}$), as 77% (79.2% from data) and 69% (69.9% from data) for glucose and xylose, respectively compared to the experimental yields.

The kinetic parameters were re-estimated for the dual substrate growth model with experimental data from batch culture containing a 1:1 ratio of glucose and xylose. The kinetic parameters estimated for the dual growth model are shown in Table 2.2. The major difference compared to the single substrate model were the maximum growth rate, μ_{max}, and substrate saturation constant, K_s, which were decreased for xylose consumption and increased for glucose consumption. This demonstrates the catabolic repression that occurs with the mixed sugars as the organism is slower to consume the preferred substrate xylose, and then rapidly consumes glucose once majority of the xylose is consumed.

Table 2.2. Kinetic parameters for growth and production from batch culture of *M. thermoacetica* on dual substrate.

Parameters*	**Glucose/Xylose Mixture**	**± 95% CI**
RMSEC	0.81	
RMSEP	1.02	
$\mu_{max,1}$ (h^{-1})	0.08	0.05
$K_{s,1}$ $(g\ l^{-1})$	1.23	0.07
$Y_{xs,1}$ $(g-cells/g-substrate)$	1.12	0.09
$Y_{px,1}$ $(g-product/g-cells)$	0.45	0.04
$\mu_{max,2}$ (h^{-1})	0.44	0.05
$K_{s,2}$ $(g\ l^{-1})$	10.8	0.40
$Y_{xs,2}$ $(g-cells/g-substrate)$	1.71	0.14
$Y_{px,2}$ $(g-product/g-cells)$	0.49	0.06
K_d (h^{-1})	0.03	0.005
K_1 $(l^{-1}\ g)$	1.37	0.10
P_{max} $(g\ l^{-1})$	48.0	

2.5 CONCLUSION

In this work growth models for the acetogen *M. thermoacetica* were developed. An unstructured-unsegregated mechanistic model was used to predict fermentation of both single sugars (glucose and xylose) and a mixture of sugars with diauxic growth. The model for both single and dual substrate fit the data well; even with high correlation between parameters the simple model of a mixed sugar fermentation system provides adequate representation of diauxic growth. Through nonlinear parameter estimation, it was shown that growth parameters can vary greatly between single and dual substrate systems. This is important fact to consider when developing models for future lignocellulosic biomass bioconversion processes. As these feedstocks will contain various concentrations of monosaccharides achieved by different feedstock types, as well as processing technologies.

2.6 NOMENCLATURE

S	Substrate concentration	$g\ L^{-1}$
X	Cell mass concentration	$g\ L^{-1}$
P	Product concentration	$g\ L^{-1}$
r_x	Rate of cell growth	h^{-1}
r_s	Rate of substrate consumption	h^{-1}
r_p	Rate of product formation	h^{-1}
μ	Specific growth rate	h^{-1}
μ_{max}	Maximum specific growth rate	h^{-1}
K_s	Substrate saturation constant	$g\ L^{-1}$
K_1	Substrate inhibition constant	
P_{max}	Maximum product concentration before inhibition	$g\ L^{-1}$
K_d	Cell death rate constant	h^{-1}
Y_{xs}	Cell growth yield from substrate	$g\ cells\ g^{-1}\ substrate$
Y_{px}	Product yield from cells	$g\ product\ g^{-1}\ cells$
Y_{ps}	Product yield from substrate, or effective product yield	$g\ product\ g^{-1}\ substrate$

Acknowledgements

This project is supported by Agriculture and Food Research Initiative Competitive Grant no. 2011-68005-30407 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of

the U.S. Department of Agriculture. The University of Washington Denman professorship fund provided financial support.

Chapter 3: Real-time Monitoring of Lignocellulosic Hydrolysate to Acetate using Raman Spectroscopy

To be submitted for publication in an international journal.

Abstract

In this work the conversion of lignocellulosic hydrolysate to acetate using *Moorella thermoacetica* was monitored using Raman spectroscopy. The signal to noise ratio (SNR) and limit of detection (LOD) for biomass from switchgrass, wheat straw, sugarcane bagasse, and sugarcanes straw were approximated to determine effects of fluorescence background. The LOD for hydrolysates were as high 25 g L^{-1} and as low as 5 g L^{-1}. Using overliming detoxification increased the SNR and correspondingly reduced the LOD. Univariate and multivariate models were developed for predicting acetate concentrations during fermentation of sugarcane straw and sugarcane bagasse hydrolysates. Three calibration models based on aqueous solutions, synthetic sugar fermentation, and hydrolysate fermentation were compared for the univariate models for acetate. Results show that calibration models built from hydrolysate fermentation were more accurate. Multivariate models were used to predict acetate, xylose, glucose, and total sugar concentrations. Much lower prediction errors were found with the multivariate models compared to the univariate models.

3.1 INTRODUCTION

Real-time continuous monitoring technology is critical to the future success of biorefining fuels and chemicals from lignocellulosic biomass [45]. In recent years a number of studies have been published that utilize Raman spectroscopy for qualitative and quantitative monitoring of analyte concentrations in microbial fermentation. The majority of these studies have focused on bioethanol production from sugars using *S. cerevisiae* [24], [25], [35], [54], [55], and only a few recent studies have performed in situ bioreactor monitoring in the presence of real lignocellulosic hydrolysates [31], [32], [34], [56]. Other biofuels related Raman spectroscopy studies have investigated lactic acid fermentation [57] and biochemical production using *E. coli* [58] from synthetic sugars. However, there has been little attention to monitoring other important chemical compounds with Raman spectroscopy that can be produced directly from microbial fermentation of cellulosic biomass.

The volume of organic acids produced from microbial processes is rapidly increasing with advances in metabolic engineering, systems biology, synthetic biology, and bioprocess engineering. Monomeric sugars derived from lignocellulosic biomass are well-suited for the production of various organic acids, ranging from C_2 to C_6's [59]. Acetic acid, an important C_2 chemical precursor used in the production of various chemicals, is one such organic acid that can be produced directly from fermentation. Global production of acetic acid is estimated at 10 million MT (Metric tons) per year and has a market value nearly twice that of fuel grade bioethanol [60]. The majority of industrial acetic acid production is from fossil fuels and chemical methods, such as methanol carbonylation, acetaldehyde oxidation, and butane liquid-phase oxidation processes [59]. In the food industry, acetic acid or vinegar, is most commonly produced by a two-stage process of first fermenting sugars to ethanol using *S. cerevisiae* followed by fermentation of ethanol to acetic acid using bacteria from *Acetobacter* or *Gluconobacter* genera [61]. A commercial facility in the United States for the microbial production of acetic acid from hybrid poplar is currently under development [44]. Utilization of lignocellulosic materials for the production of biochemicals, such as the current organic acids, could significantly increase the sustainable production of not just fuel but important commodity chemicals.

Acetogens are anaerobes that use the acetyl-CoA pathway (or more commonly called the Wood-Ljungdahl pathway) for the reduction of CO_2 to acetic acid or acetate. *Moorella thermoacetica* (formerly known as *Clostridium thermoaceticum*), is one such acetogen that can efficiently convert monosaccharides directly to acetic acid [46]. This organism is well-suited for conversion of lignocellulosic biomass as it has the ability to convert both hexose and pentose carbohydrates. Acetogens in general have not been used to commercially produce acetic acid,

due to their low tolerance to the produced acetic acid. This yields a much lower titer than acetate produced from ethanol oxidation by acetic acid bacteria (AAB). However, the production rate is much higher for acetogens as they can metabolize a wide range of carbon sources in a simple one-step process [62]. Real-time monitoring in combination with continuous fermentation processes could help overcome the low tolerance of *M. thermoacetica* by providing on-line measurements and continuous removal of acetic acid concentrations [45].

Raman spectroscopy is based on Raman scattering or inelastic scattering of light from vibrational states of molecules. Raman scattering is in general weaker compared to Rayleigh or elastic scattering, which other spectroscopic techniques measure, such as, NIR or MIR. However, an advantage of Raman spectroscopy is that water has a much weaker Raman scattering compared to organic molecules, such that aqueous solutions can be easily measured. Nonetheless, the signal can be corrupted by a number of other effects, such as the appearance of fluorescence background, background noise from the charged-cooling device (CCD), gaussian noise, cosmic rays, and scattering caused by particle interference. The major challenge in utilizing Raman spectroscopy in the presence of organic compounds, such as lignin, is fluorescence interference background signal that can reduce, if not completely mask Raman scattering. The exact cause of laser-induced fluorescence (LIF) from lignin is not completely understood. A recent study investigating fluorescence induced from model lignin compounds determined that flexible structures emit much greater fluorescence than rigid structures [63].

In order to overcome the effects of LIF, both chemometric data analysis and physical treatments have been employed. Previous studies have shown that the pretreatment methods prior to bioconversion affect the background of Raman spectra of the hydrolysate, and subsequently the quantitation of monosaccharides and other metabolites [23]. However,

thermochemical pretreatment of lignocellulosic biomass can lead to inhibitory compounds such as furans, phenolic compounds and soluble lignin components for fermentation [64]. Prior to fermentation, detoxification steps can be used to remove these inhibitive compounds created by the degradation of lignocellulosic compounds in pretreatment. One common method for detoxification is overliming, which involves raising the pH of the hydrolysate from a low to a high pH. The exact mechanism on how inhibitory compounds are removed by this method is still under investigation [65]. In combination with physical treatments, preprocessing methods are mathematical data treatments that are applied post data collection to enhance important features of the Raman spectra. These include background removal methods, filtering or smoothing methods, scaling and scatter correction techniques, which all can be used independently or in combination. A review of several preprocessing techniques for biological Raman spectra showed that not only is selection of the appropriate preprocessing technique critical, but also the order in which these methods are applied [28].

In order to utilize Raman spectroscopy to quantitatively monitor concentrations in situ, calibration models must be developed based on known concentrations measured on other instruments, usually performed off-line with high-performance liquid chromatography (HPLC). Calibration models can be developed using either univariate or multivariate analysis. Univariate analysis generally utilizes the intensity of a single Raman band or the area under the peak, while multivariate analysis can utilize a large number of variables or bands to develop a regression model. In this study, the monitoring of acetate in several pretreated lignocellulosic biomass samples were investigated using Raman spectroscopy. The objectives were to (1) determine how the background signal caused by LIF can affect the acetate signal, specifically the signal to noise ratio (SNR) and the limit of detection (LOD), and (2) develop predictive models for the

quantitative measurement of acetate and sugar concentrations during the fermentation of lignocellulosic hydrolysates to acetate using *M. thermoacetica*.

3.2 MATERIALS AND METHODS

3.2.1 Pretreatment of Lignocellulosic Biomass

In this study, fermentation was performed on lignocellulosic hydrolysate from sugarcane bagasse and sugarcane straw. The biomass was soaked in water overnight, dewatered and impregnated for 12 hrs with anhydrous 3% (w/w) SO_2. Samples of 200 g oven-dried weight (ODW) were loaded in a 1.5L batch steam gun (HM^3 Energy Inc, Gresham OR) and held for 5 minutes at a temperature of 195°C followed by explosive decompression and recovery of the pretreated slurry. The water-soluble fraction from steam explosion of biomass was separated from the slurry by vacuum filtration and kept at 4°C until use. Further details of the steam explosion can be found in [66].

Prior to fermentation, each hydrolysate was conditioned by an overliming (OL) process in which the pH was increased from 1.9 to 10.0 with $Ca(OH)_2$ at 50^oC for 30 min using a heater stir plate. The pH was then adjusted to 5.0 with sulfuric acid. The general protocol used for overliming process was based on a previously developed method [65]. After overliming, sugarcane bagasse and sugarcane straw hydrolysates used for fermentation were supplemented with xylose to bring the total concentration to 20 g L^{-1}.

3.2.2 Microorganism Cultivation

Moorella thermoacetica (ATCC 39073) was acquired from the American Type Culture Collection (Manassas, VA) grown in synthetic medium containing (in g L^{-1}) yeast extract, 10; glucose, 10; xylose, 10; L-cysteine hydrochloride monohydrate 1.0; KH_2PO_4 7.0; $NaHCO_3$ 16; K_2HPO_4 5.5; $(NH_4)_2SO_4$ 1.0; $MgCL_2$-6H2O, 0.33; $CaCl_2$ 0.05; $NaMoO_4$-$2H_2O$ 0.0025;

$Fe(NH_4)_2SO_4$-$6H_2O$ 0.04; $Co(NO_3)_2$-$6H_2O$ 0.03; Na_2SeO_3 0.0002; $NiCl_2$-$6H_2O$ 0.0002. The pH of the medium was adjusted to pH 7.3 using 5M NaOH. Inoculants were grown at 58°C in stoppered and crimp-sealed 250 mL Wheaton serum bottles containing 180 mL growth medium. The bottles were first sparged with filter-sterilized 100% CO_2, inoculated with 10% by volume for a 36 hr culture (Total volume of inoculant 200 mL). After 36 hours of growth, cell cultures were harvested and centrifuged at 4000 rpm (Beckman, GS-15R, Germany) for 5 minutes. They were then decanted to yield a cell concentration of 7 g $DCWL^{-1}$ (dry cell weight per liter). Cell concentrations were determined by measurement of optical density at 600 nm on a spectrophotometer (Shimadzu UV-1700, Columbia, MD) using standard curves relating absorbance to oven dried cell weight [49].

3.2.3 Fermentation Conditions for Synthetics and Hydrolysates

Batch fermentations were performed under anaerobic condition in a 1.3 L Bioflo 115 (New Brunswick Scientific Co., Inc., Edison, NJ) bioreactor with 500 mL working volume. Sterilization of the fermenter was performed in an autoclave at 121^{o}C for 30 min. The solution to be fermented was pumped aseptically into the vessel, and deoxygenated CO_2 was sparged at a flow rate of 1 standard liter per minute (SLPM) into the bottom of the vessel continuously for an hour prior to inoculation. After inoculation CO_2 was continuously pumped into the headspace to prevent oxygen from entering the system. The bioreactor was maintained at a temperature of 58^{o}C, pH 6.8, and continuous agitation with a Rushton impeller at 175 rpm. 5M NaOH was automatically added to maintain the pH throughout fermentation. One milliliter samples were taken at the time of inoculation and at specific intervals thereafter. Samples were immediately centrifuged at 10,000 rpm for 5 min to separate microbial cell and supernatant. The supernatant was filtered by using a 0.2 μm syringe filter and then stored at -20°C until analysis. Optical

density (OD) of the cell mass was measured at 600 nm. The OD was converted to cell dry mass (g L^{-1}) as described in the previous section. Control fermentations were performed for each hydrolysate using synthetic sugars.

3.2.4 HPLC analysis of Hydrolysate Fermentation

The concentration of monomeric sugars (arabinose, galactose, glucose, xylose, and mannose) was measured on a Dionex (Sunnyvale, CA) high-performance liquid chromatography (HPLC, ICS-3000) system equipped with an AS (auto sampler), ED (electrochemical detector), dual pumps and anion exchange column (Dionex, CarboPac PA1). Standards were prepared containing sufficient arabinose, galactose, glucose, xylose and mannose to encompass the same range of concentrations as the samples. Fucose was added to all samples as an internal standard.

Acetic acid, hydroxymethylfurfural (HMF) and furfural were measured using refractive index detection on a Shimadzu Prominence LC. Acetate concentration is reported as acetic acid. Separation of these compounds were achieved by an anion exchange column (REZEX RHM-Mono saccharide H+(8%); Phenomenex, Inc., Torrance, CA, USA) with an isocratic mobile phase that consisted of 5 mM H_2SO_4 at flow rate of 0.6 mL min^{-1}. Standards were prepared and used to quantify the unknown samples.

3.2.5 Raman data collection and analysis

Real-time analysis was performed using a RamanRXN2 analyzer (Kaiser Optical Systems, Ann Arbor, MI) across 150 – 3425 cm^{-1} wavenumber. Sample excitation came from an Invictus 785 nm laser with power at the sample of 225 mW. Spectra were collected at 20 second exposures with 10 accumulations every 20 minutes. Spectral acquisitions utilizing cosmic ray removal, where done within HoloPro environment. In situ data was collected using fiber optics connected to a blue sapphire tip ballprobe immersion optic (MarqMatrix Solutions, WA) inserted into the

bioreactor vessel. The bioreactor vessel was covered to protect the probe from any outside light sources.

Calibration models were created and analyzed using Matlab (MathWorks, MA) and the PLS Toolbox (Eigenvector Research Inc., WA) from data collected during batch fermentation within bioreactors. Univariate modeling was performed using a single peak area calculated by the trapezoidal rule. For multivariate modeling, preprocessing of Raman spectra consisted of a 1st-derivative using a 2nd order polynomial and a 15pt window, then normalized using standard normal variate (SNV) method, and mean centering. Model performance was assessed based on several measures: the root mean square error of calibration (RMSEC), the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP), correlation of determination (R^2) and the percent error of prediction. The percent error of prediction is calculated by dividing the RMSEP by the total validation range. Cross validation is performed using the leave-one-out (LOO) method.

3.2.6 Analysis of Signal to Noise and Limit of Detection for Hydrolysates Spiked with Acetate

In order to determine the effect of LIF on the acetate Raman signal, the signal to noise ratio (SNR) was calculated for lignocellulosic hydrolysate from wheat straw, switchgrass, sugarcane straw, and sugarcane bagasse. The biomass samples were pretreated using steam explosion as described above. After pretreatment, each sample was spiked to 20 g L^{-1} of acetate and the pH was adjusted to 6.8 using 5M NaOH. The acetate C-C stretching band at 928 cm^{-1} was used by selecting the spectral range in 900 – 950 cm^{-1}. The SNR was calculated as follows,

$$SNR = \frac{S}{\lambda\sigma} \tag{3.1}$$

where S is the peak intensity of a Raman spectrum of interest, σ is the root mean square error between the a high-SNR sample and a low-SNR reference, and λ is a scaling coefficient obtained by dividing the sum of all the intensity values in the reference Raman spectrum by that in the corresponding low-SNR Raman spectrum [67]. This method for the SNR calculation was used because it directly compares fluorescence free pure spectra with the signal affected by LIF over a particular region of interest.

The limit of detection (LOD) can be defined as the minimum analyte concentration that is detectable with a certain degree of confidence [68]. Typically the LOD is calculated after taking several samples at various concentrations as well as instrument blank samples and developing calibration curves. In this study, an approximation of the LOD is used based on SNR calculations without developing calibration curves. The LOD was calculated from the SNR as follows,

$$LOD = 3\frac{X_{20}}{SNR} \tag{3.2}$$

where X_{20} is the concentration of the samples spiked with acetate.

After the initial analysis of the raw and spiked hydrolysates, overliming was performed on sugarcane straw hydrolysate to determine its effects on the Raman signal. Three target pH conditions were assessed: a low level condition (target pH 9), a medium level (target pH 10), and a high level (target pH 11). For each target condition the reaction was stirred at 50^{o}C for 30 minutes as described in the methods above. The detoxified sugarcane straw hydrolysate was subsequently spiked with acetate to 20 g L^{-1}, and the SNR and LOD were calculated for each corresponding detoxification condition. This allowed the SNR and LOD to be used as a screening method for detoxification prior to fermentation.

3.3 RESULTS AND DISCUSSION

3.3.1 Characteristics of lignocellulosic hydrolysates

Raman spectroscopy is especially useful for monitoring carbohydrates in aqueous solutions, since the Raman bands of water do not mask the vibrational bands of sugars [69]. The Raman spectra of 10 g L^{-1} aqueous solutions of glucose, xylose, arabinose, mannose, galactose, and acetate (pH 6.8) samples in the spectral range of 400 – 1500 cm^{-1} are shown in Figure 3.1. Several distinguishable Raman bands for acetate were found at 652, 928, 1349, and1416 cm^{-1}, which are assigned to COO deformation, C-C stretching, CH_3 deformation, and C-O stretching, respectively [70]. Acetate is the conjugate base of acetic acid, and in solutions above pH 5.5 acetic acid converts to the acetate anion. The C-C stretching band of acetate has the strongest intensity at 928 cm^{-1}, however in acetic acid form this band scatters at 891 cm^{-1} [71]. This suggests potential for using Raman spectroscopy to not only monitor in situ concentrations but also changes in pH. For the five monosaccharides, major bands tend to be close together in the fingerprint region. The strongest peaks for glucose occur at 519 cm^{-1} (C2-C1-O1 bending) and 1120 cm^{-1} (COH bending) [72], xylose at 536, 900, and 1122 cm^{-1} (COH bending), arabinose has a strong peak at 575 cm^{-1} and several medium peaks between 800 – 1100 cm^{-1}, galactose has its strongest peak at 525 cm^{-1} and similarly mannose has a strong peak at 523 cm^{-1}. All samples contained bands in the C-H stretching region between 2800 - 3000 cm^{-1}, but these peaks were not used for analysis as they tend to be weaker and overlapping in a much narrower spectral region.

Unlike the simple sugar solutions measured above, the lignocellulosic hydrolysate fermentation broth contains various quantities of monosaccharides, degradation products (e.g. HMF, furfural, and phenolics), and a complex matrix of compounds from the fermentation

media. Therefore, convoluted peaks are expected, especially in the 400 –1500 cm^{-1} spectral region.

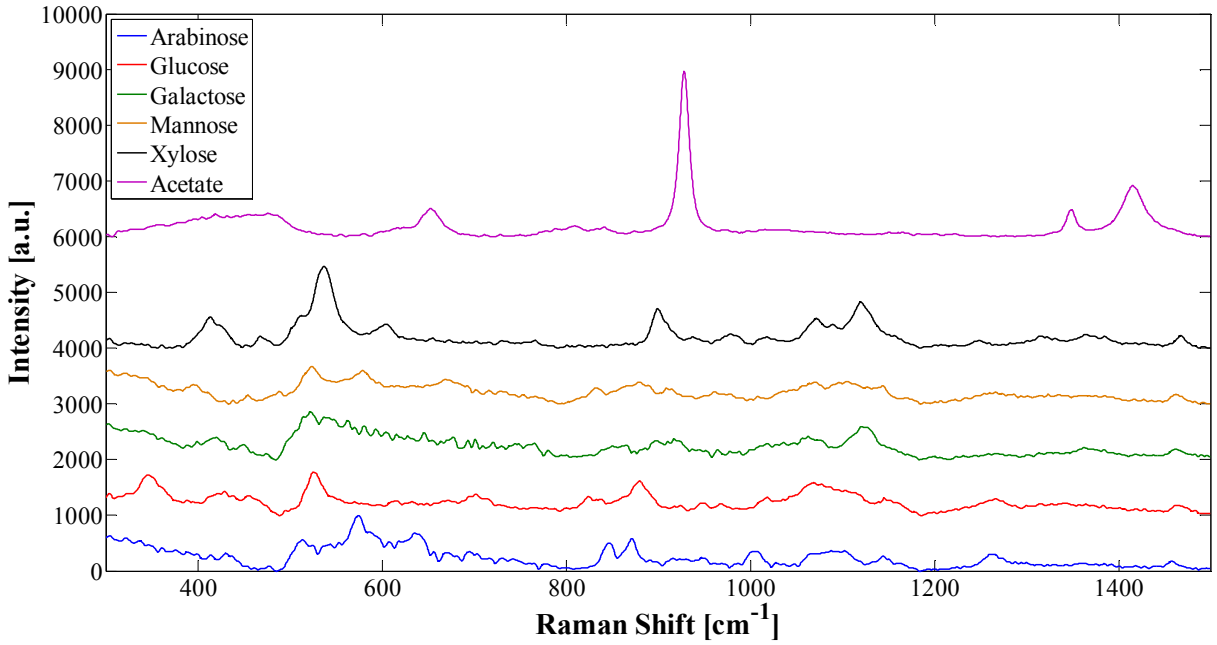

Figure 3.1. Aqueous solutions of 10 g L^{-1} samples of monosaccharides and acetate. Spectra are offset for clarity.

Calculation of the signal to noise ratio (SNR) was used to quantitatively determine the influence of fluorescence background from hydrolysates on spectral bands of interest. Liquid hydrolysates were prepared by pretreating switchgrass, wheat straw, sugarcane bagasse, and sugarcane straw with steam explosion. The raw Raman spectra of the four different hydrolysate samples spiked to 20 g L^{-1} acetate are shown in Figure 3.2. In the hydrolysate spectra, the dominant band of acetate at 928 cm^{-1} is visually difficult to distinguish. Using equations (3.1) and (3.2), the SNR and limit of detection (LOD) values for the spiked hydrolysates are calculated and shown in Table 3.1. As defined in equation (3.2), a low SNR will produce a high LOD. Thus, large amounts of noise from background fluorescence greatly reduce the ability to detect

low concentrations of analytes with strong Raman bands. The lowest LOD was obtained for sugarcane straw with a LOD of 5 g L^{-1}, while sugarcane bagasse had the highest.

In this study, the LOD was approximated as a function of SNR to quickly ascertain how background fluorescence in hydrolysates affects the Raman signal prior to calibration model development. In comparison, previous studies using corn stover hydrolysate the detection limit was calculated for ethanol and glucose using univariate calibration models, with an LOD of 6 g L^{-1} for ethanol and glucose ranged from 4 – 20 g L^{-1}, depending on the pretreatment of the hydrolysate ([23], [32]). A much lower LOD of 1 g L^{-1} was found in switchgrass hydrolysate for both ethanol and glucose, calculated by taking two times the RMSECV of the multivariate calibration model [31]. There is no internationally agreed upon standard method for calculating the LOD from multivariate models, however recently an approach consistent with the International Union of Pure and Applied Chemistry (IUPAC) has been proposed ([73]).

Table 3.1. Measured acetate concentrations in hydrolysate and calculated SNR/LOD for untreated hydrolysates.

Sample	Acetate (g L^{-1})	SNR	LOD (SNR = 3)
Switchgrass	24.4	3.84	19.8
Wheat Straw	24.4	5.02	14.6
Sugarcane Bagasse	23.6	2.64	24.9
Sugarcane Straw	22.0	14.0	5.21

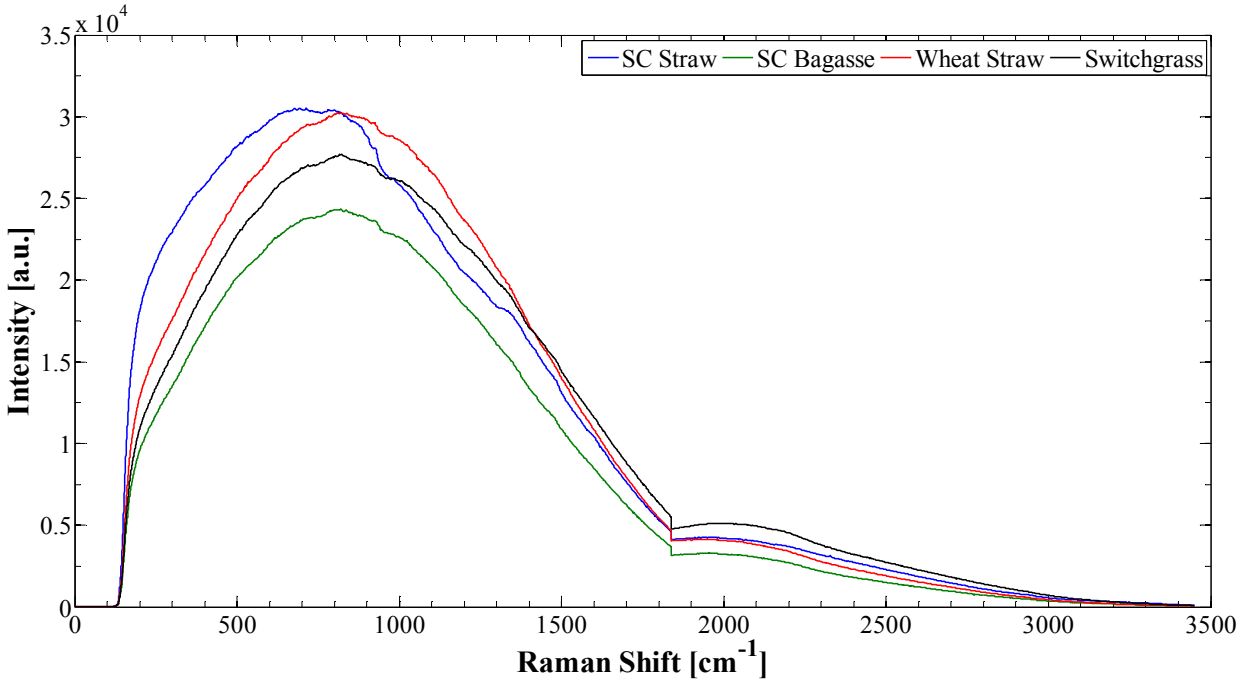

Figure 3.2. Raman spectra of unprocessed lignocellulosic hydrolysates spiked with 20 g L^{-1} of acetate**.**

Lignocellulosic hydrolysates can contain a wide range of sugar concentrations depending on the feedstock and pretreatment used. Steam pretreated sugarcane bagasse has been shown to yield monomeric xylose 0.3 – 15.5 g per 100 g of dry biomass concentrations, depending on the harshness of the pretreatment [74]. Theoretically an organism such as *M. thermoacetica* could convert 100% of xylose to acetate. The detection limit of untreated hydrolysate (24.9 g L^{-1}) is therefore too high, and this material requires further pretreatment or chemometric processing. In order to enhance the SNR and reduce the LOD, detoxification by overliming was used to reduce the amount of soluble lignin in the hydrolysate. Overliming has been shown to improve fermentation yields; however, using too harsh of a condition can dramatically reduce the amount of sugars in the media [65]. The reaction was stirred at 50°C for 30 minutes at three target pH conditions: a low level condition (target pH 9), a medium level (target pH 10), and a high level (target pH 11). The low level condition increased the SNR to 19.7 (LOD 3.5 g L^{-1}), the medium increased to 21.4 (LOD 3.2 g L^{-1}), and the high level decreased the SNR to 12.4 (LOD 5.9 g L^{-1}).

Thus, physical pretreatments have the potential to greatly enhance the effectiveness of Raman spectroscopy for monitoring hydrolysate fermentation.

3.3.2 Fermentation of Lignocellulosic Hydrolysates

Raman spectroscopy was used to monitor the fermentation of lignocellulosic derived sugars to acetate using the liquid hydrolysate fraction of pretreated sugarcane straw and sugarcane bagasse in duplicate. In addition, fermentation of synthetic media as controls for both hydrolysates was monitored using Raman. The change in the spectral range of 400 – 1500 cm^{-1} over the 76 hour fermentation, with baseline removed for visual purposes only, for both hydrolysate fermentations is represented in Figure 3.3. The spectra of sugarcane bagasse are much noisier than sugarcane straw, though distinguishable features can be found in both. Acetate clearly has the most dominant peak in the Raman spectra at 928 cm^{-1}, and minor ones can be found at 1344 and 1414 cm^{-1}. The peak at 980 cm^{-1} is associated with sulfates present in the media and was assumed to change in concentration due to sodium hydroxide added to the system during fermentation. Peaks for sugar consumption can be seen in the spectral regions of 500 – 600 cm^{-1} and 1000 – 1200 cm^{-1}, though a significant amount of overlap exists due to the presence of the complex matrix of the media and multiple sugar concentrations. The most dominant sugar peaks are associated with xylose, as it is present in much larger concentrations in all fermentations. Xylose has two distinguishable peaks at 536 cm^{-1} and 900 cm^{-1} and glucose at 519 cm^{-1}. Strong Raman bands are also present during in situ fermentation from the sapphire crystal tip of the probe at 414, 440, 573, and 748 cm^{-1} [75].

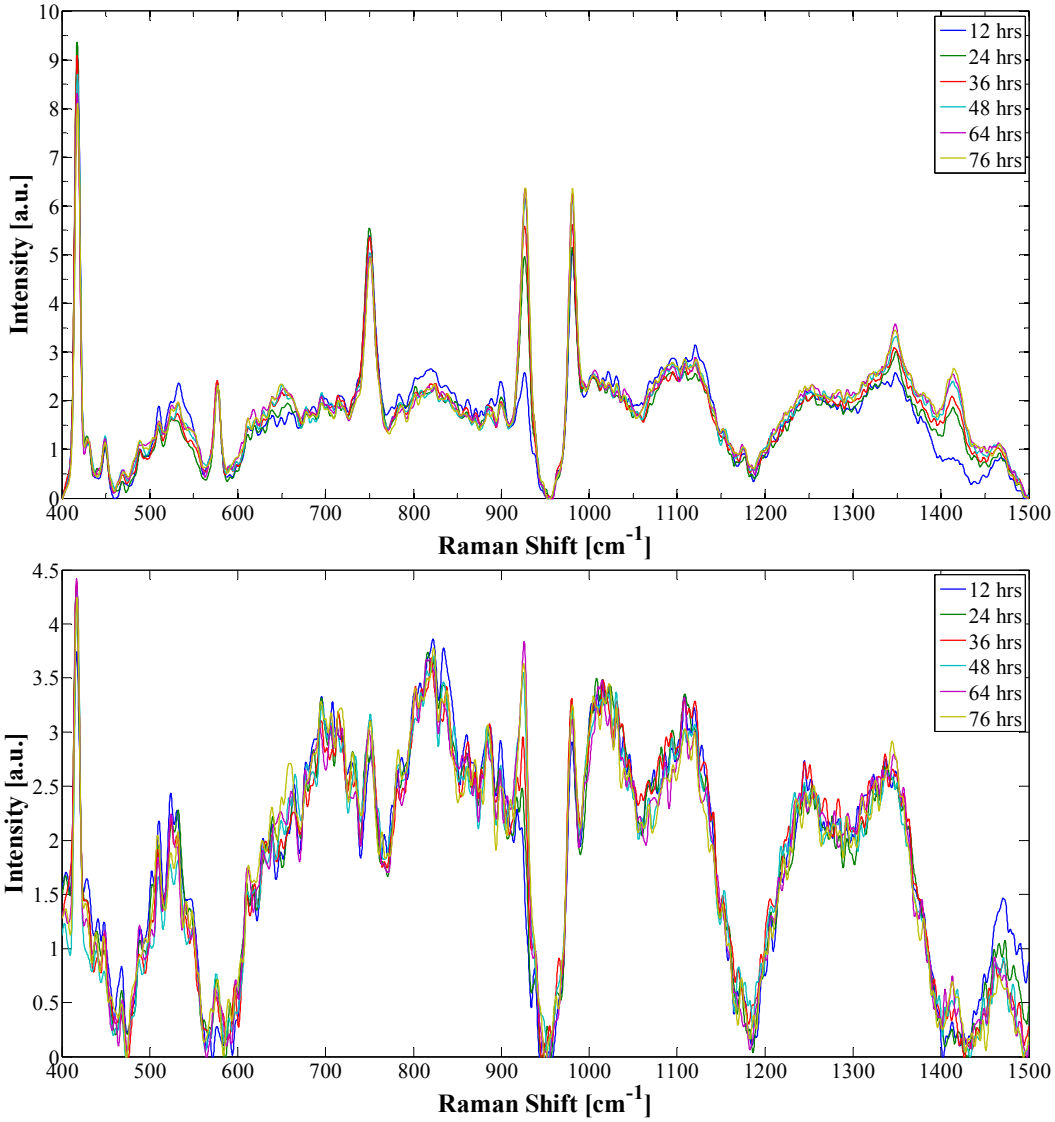

Figure 3.3. Baseline removed Raman spectra samples from fermentation of (top) sugarcane straw, and (bottom) sugarcane bagasse.

3.3.3 Univariate Model Calibration

A univariate calibration model was developed based on the peak area of acetate at 928 cm^{-1}. The area was chosen as opposed to the peak intensity as the peak location shifted during fermentation. The area was calculated using the trapezoidal rule for the spectral range of 900 – 950 cm^{-1}. Raman measurements were calibrated with HPLC data for three calibration data sets: a single set of aqueous solutions of acetate at concentrations ranging from 0 – 20 g L^{-1} (6 reference

HPLC data points), using data from two batch fermentations in synthetic media representing controls for hydrolysate fermentation (13 reference HPLC data points per batch), and data from two batch fermentations using sugarcane straw or sugarcane bagasse hydrolysate (26 reference HPLC data points for each hydrolysate). The calibration models were validated on both hydrolysates to determine the effects of using various calibration data for model development. Table 3.2 shows the results the model diagnostics from the univariate model. The errors for calibration, cross validation, and prediction are presented along with correlation coefficients (R^2) for calibration and prediction.

Table 3.2. Univariate model prediction results of acetate for sugarcane bagasse and sugarcane straw fermentations, using calibration data from aqueous acetate solution, synthetic media fermentation, and hydrolysate fermentation. The area under the peak intensity at 928 cm^{-1} in the spectral range of 900 – 950 cm^{-1} was used for calibration with HPLC results.

Data Set	R_c^2	RMSEC	RMSECV	R_p^2	RMSEP	% Error
Sugarcane Bagasse						
Aqueous Sol.	0.99	0.18	0.41	0.91	2.23	11.2
Synth Media	0.97	0.47	0.69	0.96	3.85	25.3
Hydrolysate	0.97	0.86	1.31	0.91	1.89	8.76
Sugarcane Straw						
Aqueous Sol.	0.99	0.18	0.41	0.93	1.63	7.85
Synth Media	0.97	0.47	0.69	0.93	1.422	6.86
Hydrolysate	0.99	0.25	0.6	0.96	1.045	5.04

Prediction of sugarcane straw was much more robust than that of sugarcane bagasse. The RMSEC and RMSECV values using aqueous acetate were much lower than the fermentation models. This is expected as no fluorescence background was present in the samples or scattering effects from particles in the fermentation. However, calibration models built from hydrolysates were more accurate than those built from standard concentrations of aqueous acetate solutions and from those of synthetic fermentation. This implies that calibration models should be built from samples containing a matrix of compounds as close as possible to those being predicted.

Furthermore, all calibration models show higher RMSEP values than RMSECV, which emphasizes the importance of using validation datasets.

Previous studies using univariate modeling for the conversion of corn mash to ethanol demonstrated good prediction capability with a RMSEP of 0.6 and a percent prediction error of 6.1% for ethanol [34]. In a more recent study using sugarcane bagasse hydrolysate for ethanol fermentation, univariate models for ethanol yielded a RMSEP of 0.98 g L^{-1} for ethanol in situ with scatter correction, and a relative percent error of about 12% [56]. The low prediction error for acetate suggests that univariate analysis is useful for monitoring major hydrolysate fermentation products, such as ethanol and acetate, given an adequate SNR and appropriate calibration data.

3.3.4 Multivariate Calibration

Multivariate modeling has been commonly used in real-time monitoring of bioconversion systems using spectroscopic analysis. Typically regression models are developed based on principal component regression (PCR) or partial least squares (PLS) algorithms. PLS models were developed from batch fermentation of hydrolysate (13 reference HPLC data points) and separate batch fermentations were used for validation (13 reference HPLC data points). The models were used to predict acetate, xylose, glucose, and total sugar concentration of the validation set. Total sugar is calculated as the sum total concentration of xylose, glucose, arabinose, galactose, and mannose at each sample point. The idea behind using total sugars is that it may not always be critical to know the concentrations of individual sugars, if the fermenting microorganism can consume them all. Separate models were developed for each analyte of interest based on the lowest number of latent variables that reduced the prediction error. Variables in the spectral range of 500 – 950 cm^{-1} were used for model development. Other

variables in the fingerprint region were assessed in the model development, but this range yielded the best performance and contained majority of the major bands for acetate, xylose, and glucose. Table 3.3 shows the calibration and prediction data from the PLS multivariate modeling of acetate, xylose, glucose, and total sugar concentration for both fermentation using sugarcane bagasse and sugarcane straw.

Table 3.3. Multivariate model prediction results of acetate, xylose, glucose, and total sugars for sugarcane bagasse and sugarcane fermentations. The variables in the spectral range of 500 – 950 cm^{-1} were used. Preprocessing consisted of 1st – order derivative with 15-pt window, SNV, and mean centering.

Data Set	LV	R_c^2	RMSEC	RMSECV	R_p^2	RMSEP	% Error
Sugarcane Bagasse							
Acetate	4	0.99	0.15	0.99	0.99	1.06	4.91
Xylose	3	0.99	0.31	1.24	0.95	2.59	13.6
Glucose	3	0.99	0.22	0.72	0.87	1.28	25.5
Total Sugars	3	0.99	0.14	1.29	0.98	1.69	6.40
Sugarcane Straw							
Acetate	3	0.99	0.26	0.66	0.98	0.97	4.68
Xylose	3	0.99	0.22	0.96	0.99	1.35	6.63
Glucose	4	0.99	0.04	0.12	0.97	0.30	17.3
Total Sugars	3	0.99	0.23	1.00	0.97	1.50	6.19

The validation batch predictions are plotted against fermentation time in Figure 3.4. Overall the results from multivariate calibration demonstrated accurate prediction results for both sugarcane bagasse and straw. Qualitatively the predictions for sugarcane bagasse demonstrate the data trend; however, quantitatively sugar estimates are fall off towards the end of fermentation. The number of latent variables required for each set was either 3 or 4. Sugarcane straw models were still much more robust than sugarcane bagasse, as with the univariate model. However, a significant improvement from the univariate model was that the percent prediction error of acetate was less than 5% for both hydrolysates. The RMSEC and RMSEP for xylose, glucose and total sugars are relatively low with high coefficient of determinations (with the exception of glucose in bagasse), showing model stability. The percent prediction error of glucose is much

higher than for xylose, due to the low concentration of glucose in the hydrolysates (less than 5 g L^{-1}). Total sugar estimates are closely related to xylose predictions, suggesting that individual sugar concentrations above their individual LOD should be used in calculating total sugar estimates. As with the univariate modeling, the calibration and cross validation error generally underestimate the actual predictive error.

Several recent studies have utilized multivariate models to predict concentration of fermentation analytes with Raman spectroscopy. In a study monitoring the conversion of sugars from grape juice for vinegar production, multivariate models predicting acetic acid concentration yielded a RMSEP of 0.25 g L^{-1} [71]. Other multivariate models for biofuel fermentation yielded predictive error of about 2 – 6.5% for ethanol ([31], [34], [35]), and a much larger predictive error for lactic acid at about 43% [57]. Previous percent errors for Raman-based predictions of glucose from hydrolysate using multivariate models range from about 12 - 21% for glucose estimates ([31], [34]). The percent error for predicting glucose in this study was high for both sugarcane bagasse and sugarcane straw, but comparable to many published studies.

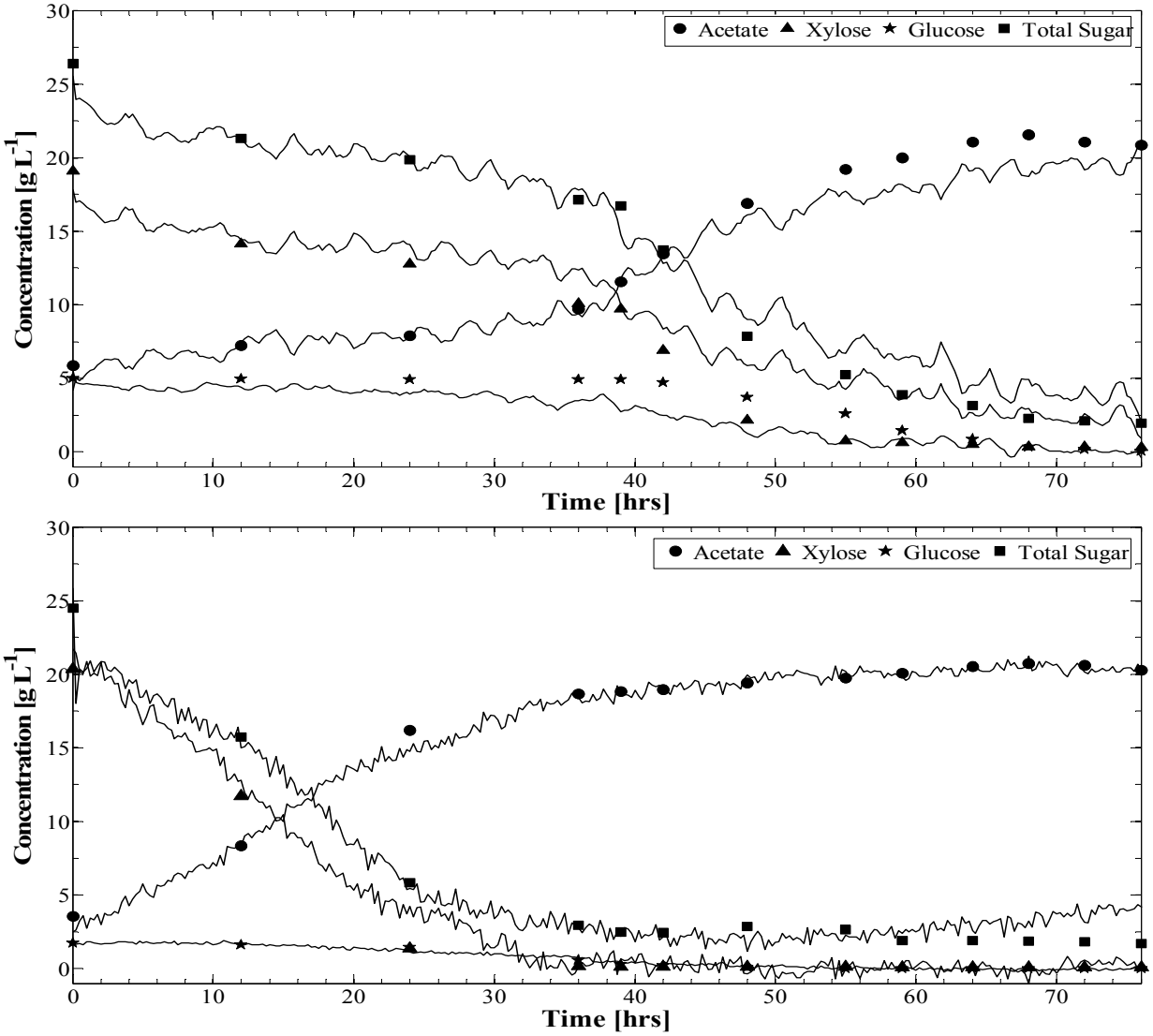

Figure 3.4. Predicted time elapsed concentrations of acetate, xylose, glucose, and total sugar from multivariate calibration model during batch fermentation of (top) sugarcane bagasse, and (bottom) sugarcane straw.

3.4 CONCLUSION

Fluorescence interference displays a stronger signal than the inelastic Raman scattering which leads to large backgrounds in Raman spectra as well as degradation of the SNR of the inelastic scattering data of interest. Four different untreated lignocellulosic hydrolysates all had a very low SNR, which yields a high LOD making them unsuitable for real-time monitoring without

additional physical or chemometric treatments. Overliming proved to be effective in increasing the SNR of sugarcane straw hydrolysate, though further investigation should be conducted to determine optimal overliming conditions for each type of hydrolysate.

Raman spectroscopy based univariate models were developed for predicting acetate, as well as multivariate models for predicting acetate, xylose, glucose, and total sugar concentrations during the fermentation of two lignocellulosic hydrolysates using *M. thermoacetica*. It was shown that utilizing an appropriate calibration data set is essential for good predictive results. The high accuracy multivariate predictions of acetate (less than 5% prediction error for both hydrolysates) in this work are significant given the highly complex nature of the fermentation medium when compared to the simpler media of previously published work. The multivariate models proved to be more accurate than the univariate models. The multivariate model was also able to predict xylose, glucose, and total sugar consumption at relatively low RMSEP for sugarcane straw. Sugarcane bagasse yielded less accurate carbohydrate models in terms of predictive abilities, but still demonstrated adequate error in terms of acetate prediction and sugars.

Acknowledgements

This project is supported by Agriculture and Food Research Initiative Competitive Grant no. 2011-68005-30407 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. The University of Washington Denman professorship fund provided financial support.

Chapter 4: NMPC of an MCRB using Raman spectroscopy and EKF for full-state feedback

Abstract

In this work Raman spectroscopy is used to monitor a continuous membrane cell-recycle bioreactor (MCRB) in real-time for the bioconversion of glucose to ethanol with *S. cerevisiae*. A first principle kinetic model and multivariate calibration models for Raman were developed based on batch fermentation data. The multivariate model had a percent error of prediction for ethanol of 2% and 3% for glucose. These low prediction errors based on a validation data set were attributed to utilizing an external microfiltration membrane to separate the glucose and ethanol stream from growing yeast cells prior to Raman measurements. The batch models were then used to monitor the continuous MCRB system in open-loop. Both the kinetic model and the multivariate Raman sensor model were effective in monitoring the MCRB fermentation. Finally the ethanol productivity of the MCRB was controlled utilizing a nonlinear model predictive controller (NMPC) with total substrate conversion as a constraint. This work demonstrates the effectiveness of real-time monitoring and control of a continuous MCRB bioethanol fermentation.

4.1 INTRODUCTION

Production of fuels and chemicals from lignocellulosic biomass is an increasingly important area of research and industrialization throughout the world. In order to be competitive with fossil-based fuels and chemicals, maintaining cost-effectiveness is critical. Majority of the current research focuses on reducing operating costs of biofuels and biochemicals by finding cheaper feedstocks, developing efficient and robust microorganisms, process integration, and co-product utilization. There has been relatively little emphasis on process control and optimization methods, which have significantly reduced operating costs in similar industries [76].

Shifting from traditional batch fermentation to continuous fermentation has been proposed as one such method to reduce manufacturing costs [77]. Continuous fermentation has several advantages over traditional batch fermentation, including: reduced operating costs, minimal processing footprint, higher productivity, continuous product removal, which reduces inhibitor concentrations in the bioreactor, and less down-time for clean-up and sterilization [77]. In general, the productivity in a continuous fermentation system is much higher than in a batch

reactor [78]. Productivity of a continuous fermentation process can be further increased by continuous separation of microorganism cells from the product stream, this process is known as membrane cell recycle bioreactors (MCRB). In this setup the effluent of the bioreactor is passed through a membrane to form two streams: one stream containing the cells recycled back into the bioreactor (the retentate) and another stream containing product (the permeate). A recent review of these technologies and their potential role in biorefineries are presented in [79].

Automation of continuous processes could further reduce overall production costs [80]. Benefits of bioprocess automation include: process variability reduction, productivity improvements, increased on-line monitoring and troubleshooting capability [80]. A major ingredient to process automation is real-time measurements of key bioprocess properties and states. Commonly used on-line sensors for bioprocesses monitor temperature, gas flow rates, agitation rate, oxidation reduction (redox), dissolved oxygen, and pH. Other probes that have been implemented on-line or at-line, include dielectric probes for monitoring cell growth [81], and spectroscopic probes (e.g. NIR, MIR, Raman) for monitoring analyte concentrations [82]. Raman spectroscopy has proven to be a useful measurement tool for real-time analysis of fermentation systems [45]. However, quantitative measurements of bioconversion processes today mainly rely on off-line technologies, such as, high-performance liquid chromatography (HPLC) for monitoring concentrations. Performing analysis on an off-line instrument delays critical system evaluation, making it difficult to perform process control. In addition to hard sensors, soft sensors, otherwise known as state observers, can be implemented to estimate unknown state measurements ([20], [40]). These are mathematical techniques, rather than actual sensors, that utilize knowledge of the relationships between known system measurements and unknown system states to calculate the unknown states.

Model predictive control (MPC) is becoming an industry standard in chemical, petrochemical, and bioprocess related industries. Over the last decade model predictive control (MPC) or receding horizon control (RHC) has become commonplace in many chemical and manufacturing industries. In 2003, it was reported that there were over 5,000 applications of MPC, with majority of the applications being in refining, petrochemicals, chemicals, and pulp and paper industries [10]. More recently, an international survey of 66 APC experts, 38 APC users, and 28 APC suppliers showed that over two thirds of the respondents used MPC [7]. The benefit of linear model predictive control without constraints is that an optimal solution can be found using efficient quadratic programming with guaranteed stability. In the literature, majority of the applications of MPC for fermentation systems have been based on simulation. These include both fed-batch [11], [83] and continuous [13]–[15] systems, all of which used some form of nonlinear MPC or neural network. One of the major reasons for the minimal application of APC for bioreactors is lack of reliable technology for online measurements of major chemical and biochemical concentrations [9]. A recent study utilized Raman spectroscopy sensors with nonlinear MPC (NMPC) strategy to control a fed-batch fermentation of mammalian cell growth [84]. The researchers were successful in implementing an optimal feeding strategy. An advantage of using NMPC is that nonlinear models can be used directly for prediction. However, NMPC can lead to nonconvex optimization that may result in suboptimal local minima solutions.

The objective of this investigation was to develop a MCRB fermentation system for continuous ethanol production using the microorganism *S. cerevisiae.* Raman spectroscopy sensors were used for online measurements of the substrate (glucose) and product (ethanol) concentrations, providing output feedback to a nonlinear model predictive control (NMPC). First, the system was operated in batch mode to develop kinetic models for the NMPC, and

multivariate calibration models for the Raman sensors. Second, the system was operated in open-loop without the controller to determine the nominal performance of the MCRB system and compare off-line HPLC measurements, with the Raman measurements, and model predictions for a continuous system. Finally, the NMPC is applied to the MCRB using the Raman sensors and an extended Kalman filter (EKF) to provide estimates of biomass concentration for full state feedback. The goal of the NMPC is to follow a setpoint trajectory for the productivity, while holding total substrate conversion as a constraint.

4.2 KINETIC MODEL DEVELOPMENT

4.2.1 MCRB Kinetic Model

A mass balance of a membrane cell-recycle bioreactor (MCRB) system can be defined by a system of ordinary differential equations (ODE) of the major states. The model in this investigation consists of three nonlinear ODEs, with state variables for biomass, substrate, and product concentration defined by the following system:

$$\frac{dX}{dt} = -BDX + r_x X \tag{4.1}$$

$$\frac{dG}{dt} = D(G_f - G) - r_s X \tag{4.2}$$

$$\frac{dE}{dt} = -DE + r_p X \tag{4.3}$$

where X, G, and E, are the biomass, glucose, and ethanol concentration, respectively, with kinetic rates of microorganisms growth (r_x), glucose consumption (r_s), and ethanol production (r_p). Environmental variables include the dilution rate, D (where $D = F/V$, total flowrate divided by the volume); B the cell bleed ratio (where $B = F_b/F$); G_f is the feed concentration of glucose into the bioreactor; F is the total volumetric flow rate into the reactor for glucose feed;

F_b is the flow rate of the bleed, and V is the reactor volume (assumed constant). It is assumed that no cells exit through the membrane or enter with the feed. Similarly, no product enters with the feed stream.

An unstructured-unsegregated model was used to describe the growth kinetics of *S. cerevisiae* based on the work of [85]. Several common assumptions were used in the model development including: the medium is well-mixed and homogenous; temperature and pH are held constant during fermentation; and product formation is assumed to be both growth associated and non-growth associated. The growth term is based on Monod type kinetics with a cell growth inhibition term. Given the aforementioned assumptions the kinetic model is written as,

$$r_x = \mu_{max} \frac{G}{K_S + G} (1 - \frac{X}{X_{max}}) \tag{4.4}$$

$$\mathrm{r}_s = \frac{r_x}{Y_{xs}} + m_s \tag{4.5}$$

$$\mathrm{r}_p = Y_{px} r_x + m_p \tag{4.6}$$

where μ_{max} is the maximum growth rate of cells; K_s is the substrate saturation constant for cell growth; X_{max} is the biomass concentration when cell growth ceases; Y_{px} is Luedeking-Piret growth-associated constant and m_p is the Luedeking-Piret non-growth associated constant; and Y_{xs} is the cell yield from substrate and m_s is the maintenance coefficient for cells based on sugar consumed.

4.2.2 Parameter estimation

Kinetic parameters were determined by setting the dilution rate, $D = 0$ (batch mode) in equations (4.1) – (4.3) and minimizing the chi-squared error between the model predictions and

experimental data for a batch system. This is determined through minimization of the following objective function,

$$\chi^2(\theta) = min \sum_{j=1}^{n} \left[\frac{\left(X_j - X_{mod,j}\right)^2}{\sigma^2} + \frac{\left(G_j - G_{mod,j}\right)^2}{\sigma^2} + \frac{\left(E - E_{mod,j}\right)^2}{\sigma^2} \right] \quad (4.7)$$

where θ is the vector of kinetic parameters; X_j, G_j, and E_j are experimental data of biomass, substrate, and product concentrations at sampling time j; $X_{mod,j}$, $G_{mod,j}$, and $E_{mod,j}$ are the concentrations computed by the model in each sampling time; and σ^2 is the variance between experimental replicates for total of samples n. The parameters are subjected to upper and lower bounds to ensure realistic values,

$$l_q \leq \theta_q \leq u_q$$

where θ_q is parameter q, $l_q$is the lower bound for parameters, and u_q is the set of upper bounds on the parameters. Initial parameter values were obtained from literature.

4.3 MATERIALS AND METHODS

4.3.1 Microorganism Cultivation

S. cerevisiae (ATCC96851, American Type Culture) was streaked onto YPD agar plate cultures, and allowed to grow for 48 hours. Prior to fermentation, pre-culture cells were grown adding one colony from the plate to liquid media containing 10 g L^{-1} each of glucose, peptone, and yeast extract. After 24 hours of growth at 30ºC and 175 rpm orbital shaking, the cells were centrifuged and the spent supernatant removed and replaced with fresh media. The cells were then grown for another 24 hours under the same conditions; the cells were again spun down, washed twice in deionized water, and then re-suspended in deionized water. The cell concentration was

determined by measuring the optical density of the suspension at 600 nm and comparing to a calibration curve prepared using oven dried cells at varying optical densities.

Synthetic fermentation broth solutions were adjusted to pH 6 using 2M HCL and 5M NaOH. Nutrients in the form of 10 g L^{-1} peptone, 10 g L^{-1} yeast extract, 2 g L^{-1} ammonium phosphate, 2g L^{-1} sodium nitrate, and 0.2 g L^{-1} sodium sulfate. Glucose was added at 45 g L^{-1} for a set of batch fermentations, 60 g L^{-1} for a continuous MCRB fermentation without controller, and 45 g L^{-1} for fermentation with controller. All media was filtered through a 0.2 μm sterile filter before fermentation. The initial cell concentration used for fermentation was 5 g L^{-1}. The total solution volume was 1000 mL and the mixture was stirred continuously with a Rushton impellor at 175 rpm.

4.3.2 Fermentation Conditions

Batch and continuous fermentations were performed in a 2L BioFlo 115 bioreactor (New Brunswick Scientific Co., Inc., NJ) equipped with an external cross-flow microfiltration membrane with a surface area of 0.1 m^2 and nominal pore size 0.22 μm (Pellicon 2 , EMD Millipore, MA). A schematic diagram of the continuous MCRB system is shown in Figure 4.1. During batch fermentation permeate and retentate streams were both recycled back into the fermenter, while during continuous fermentation permeate was continuously removed from the system. Before each run the membrane system was cleaned and sterilized by circulating Tergazyme ® at 0.2% concentration at temperature 45^{o}C for 30 – 60 minutes, followed by circulating a liter of 5% dilute NaOCl solution for 15 – 30 minutes. The system was then flushed using 5 L of sterile deionized H_2O. Prior to continuous fermentation the system was operated in batch mode for 6 hours until majority of the glucose had been consumed. Once continuous

fermentation was initiated, fresh media containing glucose was continuously pumped into the bioreactor and filtered broth was pumped out of the system in the permeate stream.

The bioreactor unit was equipped with a pH probe, temperature probe, level probe, exhaust condenser, and acid/base ports. The bioreactor environmental conditions were held constant at 30°C, pH 6.0, and 175 rpm continuous agitation using the PID controller of the BioFlo 115 control tower (New Brunswick Scientific Co., Inc., Edison, NJ). Three external peristaltic pumps were used for control of the glucose feed (Masterflex L/S, EW-07551-20, Cole-Parmer, with an easy-load II pump head), the recycle loop (Masterflex L/S, EW-07551-20, Cole-Parmer, with high-performance pump head), the permeate (Masterflex L/S, EW-07551-20, Cole-Parmer, with an easy-load II pump head), and the bleed flow pump was set as 10% of the dilution rate on a persistaltic pump connected to the BioFlo 115 control tower. The transmembrane pressure (TMP) was monitored using pressure transducers (PX309-100A-5V, Omega Eng.), and manually controlled using a back pressure regulator on the retentate side (EB1HF1, Equilibar). The pressure was maintained at 1.5 bars and the recycle flow rate was held at 0.5 L h^{-1}. Data acquisition (DAQ) was performed using a National instrument DAQ (NI-cDAQ-9174) and analysis was performed in Matlab 8.1 (Mathworks, MA).

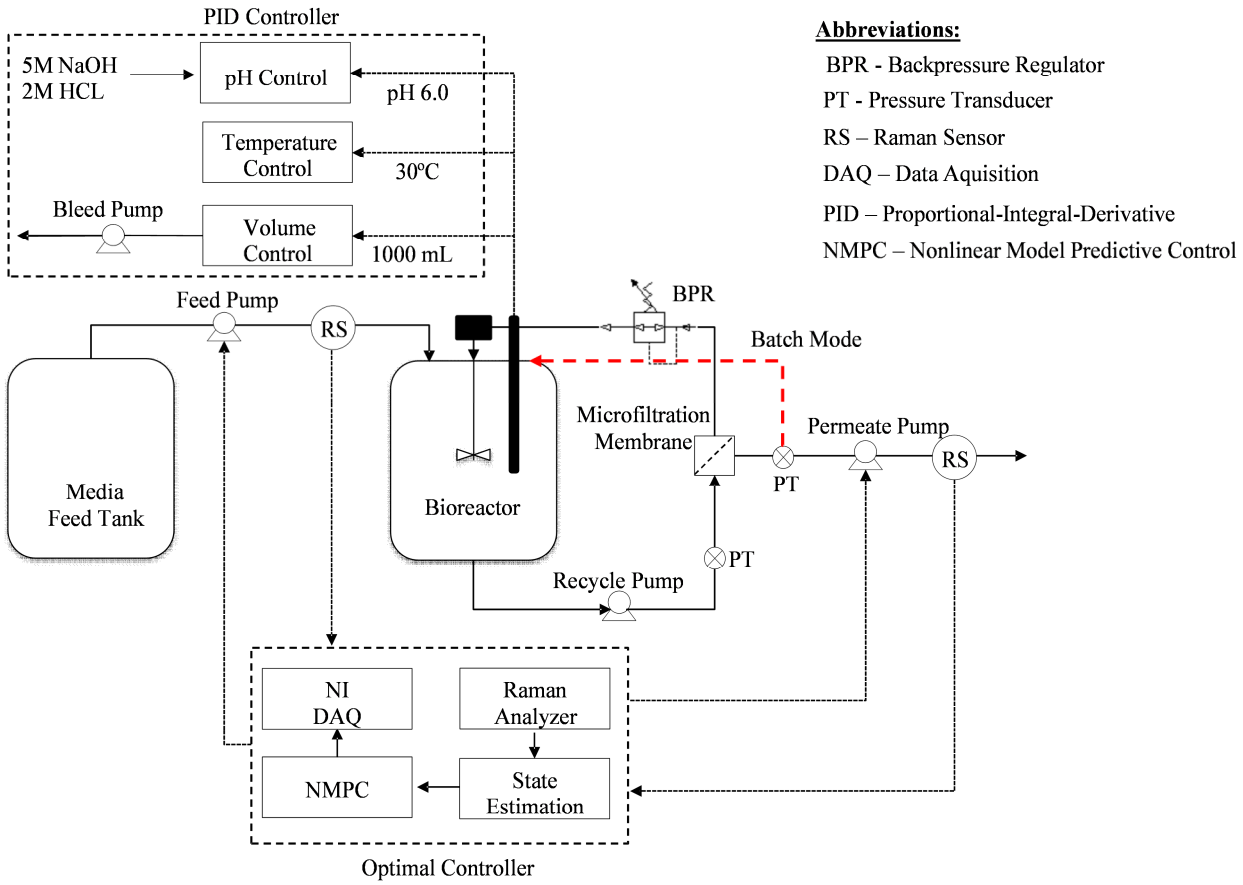

Figure 4.1. Schematic diagram of MCRB system used in experiment.

4.3.3 HPLC analysis

One milliliter samples were removed from the fermenter and the permeate stream at specified time points and used for analysis. The samples were centrifuged for 5 min at 10,000 rpm, removing the supernatant for High-Performance Liquid Chromatograph (HPLC) and then the pellet was re-suspended in 1 mL of deionized water. The supernatant was then put through a 0.2 µm filter to remove any remaining cells. The cell free supernatant was used for off-line determination of glucose, and ethanol in a Shimadzu Prominence LC. Separation of these compounds were achieved by an anion exchange column (REZEX RHM-Monosaccharide H+(8%); Phenomenex, Inc., Torrance, CA) with an isocratic mobile phase that consisted of 5 mM H_2SO_4 at flow rate of 0.6 mL min^{-1}. Standards were prepared and used to quantify the

unknown samples. Yeast cell concentration was determined by optical density (OD) at 600nm. The OD was converted to cell dry mass following established protocol [49].

4.3.4 Raman data collection and analysis

Real-time analysis was performed using a RamanRXN2 analyzer (Kaiser Optical Systems, MI) across 150 – 3425 cm^{-1} wavenumber. Sample excitation came from an Invictus 785 nm laser with power at the sample of 225 mW. Spectra were collected at 10 second exposures with 5 accumulations every 15 minutes. Spectral acquisitions utilizing cosmic ray removal, where done within HoloPro environment. In situ data was collected using fiber optics connected to a blue sapphire tip ballprobe immersion optic (MarqMatrix Solutions, WA) inserted directly into the permeate stream using a swagelock t-connection compression fitting. This point of measurement was selected, since the membrane removes cells from the glucose and ethanol permeate stream, reducing scattering caused by the biomass particles. A second Raman probe was inserted into the media feed stream to monitor incoming glucose concentration.

Multivariate partial least square (PLS) calibration models were created and analyzed using Matlab 8.1 (MathWorks, MA) and the PLS Toolbox (Eigenvector Research Inc., WA) from data collected during batch and continuous fermentation within bioreactors. Multivariate modeling was performed with preprocessing consisting of a 2^{nd} order polynomial baseline removal method [29], then normalized using standard normal variate (SNV) method, and mean centering in the spectral range between 800 to 1200 cm^{-1}. Model performance was assessed based on several measures: the root mean square error of calibration (RMSEC), the root mean square error of cross-validation (RMSECV), the root mean square error of prediction (RMSEP), correlation of determination (R^2) and the percent error of prediction. The percent error of

prediction was calculated by dividing the RMSEP by the total validation range. Cross validation was performed using the leave-one-out (LOO) method.

The limit of detection (LOD) was calculated for PLS multivariate models based on published work [73]. The method is consistent with international (IUPAC) recommended calculation of LOD for univariate models. The pseudounivariate calculation of LOD was used in this investigation, based on the following equation,

$$LOD = \frac{3.3}{S_{pu}} \sqrt{\left(1 + h_{0,min} + \frac{1}{I}\right) var_{pu}} \tag{4.8}$$

where S_{pu} is the slope of the pseudounivariate line, I is the number of calibration samples, var_{pu} is the variance of the regression residuals, and $h_{0,min}$ is the minimum sample leverage when the analyte concentration is zero. It was shown that the analyte sample leverage can be simply calculated by,

$$h_{0,min} = \frac{\bar{y}_{cal}^2}{\sum_{i=1}^{I} y_i^2} \tag{4.9}$$

where $\bar{y}_{cal}$ is the mean of the calibration samples, and y_i is the centered concentration for the *i*th sample.

4.3.5 Nonlinear model predictive control with output feedback for an MCRB

In most continuous fermentation systems the goal is to maximize productivity. The productivity is defined as the amount of product formed over time,

$$\mathrm{Pr} = \bar{E}D \tag{4.10}$$

where $\bar{E}$ is the steady state ethanol concentration of the continuous system, and D is the dilution rate. In addition to productivity, it is also important to maintain a high level of substrate

conversion to ensure all the feed substrate is utilized. Total substrate conversion can be defined as,

$$Con = \frac{G_f - \bar{G}}{G_f} \tag{4.11}$$

where the steady state conversion of the continuous system is a function of the incoming substrate feed concentration G_f, and the steady state glucose concentration $\bar{G}$.

A nonlinear model predictive control (NMPC) strategy was utilized to track productivity with the dilution rate as the manipulated variable, while holding the substrate conversion as a constraint. The nonlinear model described in equations (4.1) – (4.6) can be written in state space form as:

$$\begin{aligned} x_{k+1} &= f(x_k, u_k) \\ y_k &= C x_k \end{aligned} \tag{4.12}$$

where x_k is the state vector at the current time sample k and x_{k+1} represents the state predicted at the next sample instant, u is the vector of manipulated inputs, y is a linear vector function of outputs, C is a diagonal matrix for full state output, and $f(x, u)$ is the nonlinear bioreactor model. The NMPC requires full state output feedback of the major states including: biomass, glucose, and ethanol concentration. However, the only state measurements known at each sample instant are for glucose and ethanol. Therefore, an observer is used for state estimation of the biomass concentration. In this study, the extended Kalman filter (EKF) was used. The EKF uses a first-order linearization along the state estimate trajectory to compute the evolution of the covariance matrix. The EKF in (equation) is formulated as:

$$x_{k+1}^{-} = f(\hat{x}_k, u_k) \tag{4.13}$$

$$\hat{x}_k = x_k^{-} + K_k(y_k - C x_k^{-}) \tag{4.14}$$

where x_k^- and $\hat{x}_k$ are called the *a priori* and *a posteriori* estimate, respectively. K_k is the Kalman gain calculated by:

$$P_{k+1}^- = A_k P_k^+ A_k^T + Q \tag{4.15}$$

$$K_k = P_k^- C^T (C P_k^- C^T + R)^{-1} \tag{4.16}$$

$$P_k^+ = (I - K_k C) P_k^- \tag{4.17}$$

where $A_k = (\partial f / \partial x)|(\hat{x}_k, u_k)$ is the linearization of the model; Q (diag[1e-5, 1, 1]) and R (diag[0.4, 1.5]) are symmetric positive definite matrices used as tuning parameters for the EKF; P is the covariance matrix, which was initial set by taking half upper and lower bounds of the initial state estimation as described in [86]; equation (4.15) represents a partial update of the covariance matrix used in the calculation at the current sample for the Kalman gain in equation (4.16), and the algorithm ends by updating the covariance matrix for the next sample in equation (4.17).

Finally the optimal control move over a finite prediction horizon is calculated by optimizing a cost function. The cost function over the finite horizon is defined by:

$$\min \sum_{i=0}^{N_p-1} \left[W_q \left(Pr_i - Pr_{sp} \right)^2 + W_r (u_{i+1} - u_i)^2 \right] + F(Pr_{sp}) \tag{4.18}$$

$$s.t. \quad x_{k+1} = f(x_k, u_k)$$

$$t \in [k, k + N_p]$$

$$D \geq 0$$

$$Con \geq 0.98$$

where the summation is minimized over the prediction horizon with the first term controls the predicted system state and the second term controls the input sequence. The weighting parameters W_q and W_r used to penalize setpoint tracking and manipulated input movement,

respectively, are both set to one. A terminal cost set to the productivity setpoint value is used to ensure that the optimal solution lands in a feasible region. The inequality constraint on the dilution rate is used to ensure realistic values and the constraint on substrate conversion is set to 98%. In addition by selecting a relative long prediction horizon ($N_p = 20$) helps the system achieve a stable solution for the system [87]. The objective function is evaluated and the selection of a control move is repeated until the optimum is obtained. The optimization problem was solved using an interior point method, and the integration of the nonlinear model was solved using a 4th order runga-kutta method, which were solved sequentially in Matlab 8.1.

4.4 RESULTS AND DISCUSSION

4.4.1 Kinetic modeling and Raman-based monitoring of batch ethanol fermentation

Two batch fermentations where performed with the microorganism *S. cerevisiae* and an initial glucose concentration of 45 g L^{-1}. During fermentation online measurements were made with Raman sensors. Prior to Raman measurements cell mass was continuously separated from the fermentation broth with an external microfiltration membrane (as described in the methods above) and subsequently recycled back into the fermenter. Raman measurements were taken every 15 minutes, and HPLC samples were taken every 30 minutes over 8 hours. Data was preprocessed as previously described in the spectral range of 800 – 1200 cm^{-1}. One batch fermentation data set (17 reference HPLC data points) was used to build partial least squares (PLS) calibration models between Raman spectra and HPLC measurements for glucose and ethanol concentrations. A second independent data set (17 reference HPLC data points) was used for validation of the multivariate model. Separate PLS models were developed for both ethanol and glucose.

All sugars were consumed within 7 hours with an average total sugar to ethanol yield of 80% (g g^{-1}) of the theoretical yield. The corresponding average ethanol productivity for batch fermentation was 2.9 g L^{-1} h^{-1}. Prediction results from the multivariate PLS models are shown in Figure 4.2. In the preprocessed Raman spectra, major Raman bands for ethanol (883 cm^{-1}, C-C stretch) and glucose (1123 cm^{-1}, COH bending) were readily distinguishable. The boxes in the top two plots of Figure 4.2 show the statistical data from the calibration models, while the circles show the prediction results for the validation set. Based on minimal RMSEP of the validation data set, six latent variables (LVs) were used for the ethanol model yielding a RMSEP of 0.4 g L^{-1} and three for glucose yielding a RMSEP of 1.5 g L^{-1}. Model predictions also produced a high correlation of determination of 0.99 for both ethanol and glucose. From equations (4.8) the limit of detection (LOD) was calculated as 0.7 g L^{-1} and 2.0 g L^{-1} for ethanol and glucose, respectively. The low error, low LOD, and high R^2 values for model predictions demonstrate the robustness of the models over the validation range.

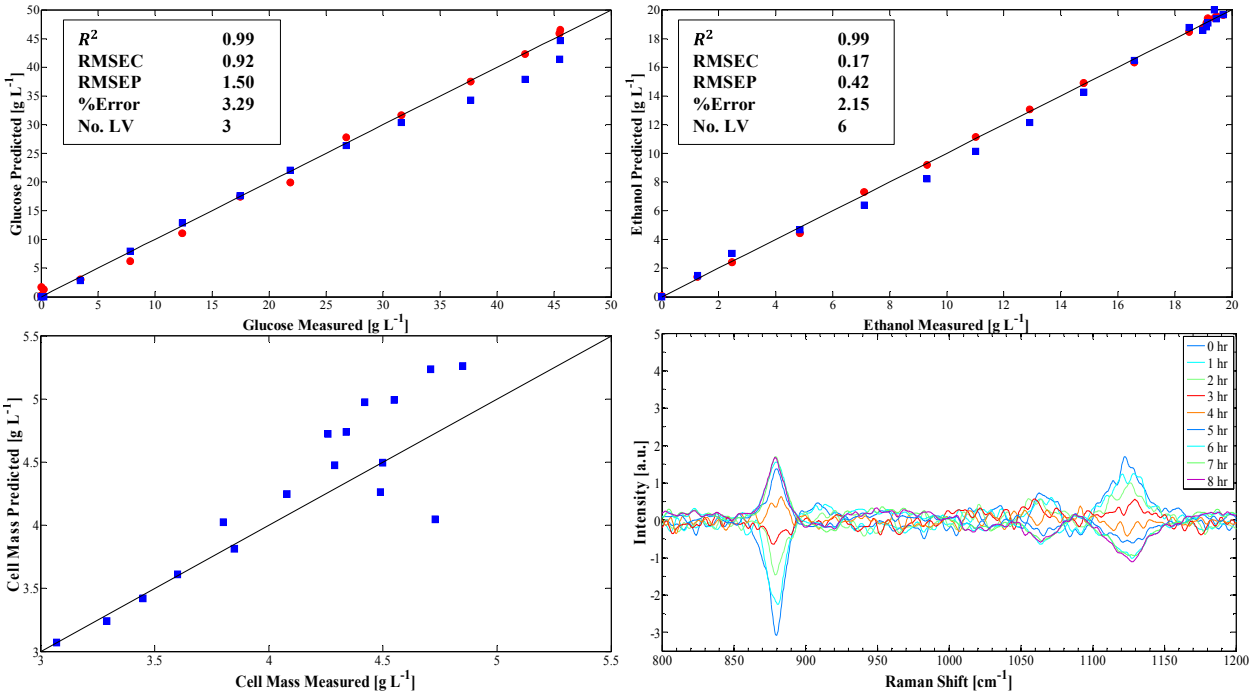

Figure 4.2. Multivariate PLS model results (red circles) and kinetic model results (blue squares) compared to measured HPLC data of batch ethanol fermentation. Line represents unitary model prediction. Prediction curves for ethanol (top left), glucose (top right), cell mass (bottom left), and preprocessed Raman spectra (bottom right). Statistical data for PLS model shown in top plots for ethanol and glucose. Preprocessing consisted of a 2nd-order polynomial baseline removal method, SNV, and mean centering.

Similar to the development of the multivariate PLS model, a single batch set of HPLC measurements were used for the nonlinear regression of kinetic parameters identified for equations (4.1) – (4.6) (in batch mode, $D = 0$). The nominal values of the seven estimated parameters are shown in Table 4.1 along with their 95% confidence interval. The nominal values represent those found through nonlinear regression using 50 passes of a multi-start global search method in Matlab 8.1 (MathWorks, MA). The parameters were ranked from least identifiable to most identifiable based on an orthogonalization algorithm [52]. The least identifiable parameter or the parameter with the most uncertainty was the maximum cell concentration X_{max}, an inhibition parameter added to account for reduction in the rate of cell growth by the accumulation of cell concentration in the bioreactor. This parameter also has a 95% confidence interval that includes zero in its bounds and thus, is not statistical significant. The nominal parameter values yielded a RMSE of 1.8 g L^{-1} between the model and the data. The results from the model prediction are also shown in Figure 4.2. Total cell mass was only estimated with the kinetic model.

Table 4.1. Kinetic parameters for growth and production from batch culture of *S. cerevisiae* on glucose.

Parameters	**Nominal** $\theta \pm 95\%\ CI$	**Rank**
RMSE	1.75	
$\mu_{max}\ (h^{-1})$	0.13±0.03	7
$m_p\ (g\ g^{-1}\ h^{-1})$	0.63±0.14	2
$m_s\ (g\ g^{-1}\ h^{-1})$	1.59±0.33	3
$K_S\ (g\ l^{-1})$	0.008±0.04	6
$Y_{px}\ (g\ g^{-1})$	2.11±1.20	5
$Y_{xs}\ (g\ g^{-1})$	0.20±0.11	4
$X_{max}\ (g\ l^{-1})$	93.9±140	1

Noise in the Raman spectra during fermentation due to scattering caused by increasing cell concentration can greatly reduce the prediction performance of calibration models. This was demonstrated in a recent study that used Raman spectroscopy to monitor *in situ* batch ethanol fermentation [55]. In that study, two models were compared, one that utilized an approximation of the scattering caused by cells to correct for the noise in the Raman signal, and the other without. Univariate models for ethanol and glucose with scatter correction yielded a RMSEP of 1.6 and 2.4 g L^{-1}, for ethanol and glucose, respectively; while without scatter correction the prediction error was much higher at 7.1 and 8.9 g L^{-1} for ethanol and glucose, respectively. This yielded a percent error ranging approximately from 5 – 22% for ethanol and 8 – 36% for glucose, depending on the effect of cell scattering. In this study, cells were separated prior to Raman measurement through a microfiltration membrane, which eliminated the need to correct for cell scattering.

However, the difference in prediction error could also be attributed to the use of a multivariate model versus a univariate model. The percent error is a useful comparison of prediction error of different models in relation to the concentration range of the predicted component. Similar PLS multivariate models for batch ethanol fermentation have been previously reported, yielding predictive error ranging from about 2 – 7% for ethanol [31], [34],

[35]. Previous percent errors for Raman-based predictions of glucose from hydrolysate using multivariate models ranged from about 12 - 21% for glucose estimates [34], [31]. This study demonstrates a similar percent error for ethanol at 2%, and a much lower prediction error for glucose at 3%.

4.4.2 Open-loop performance of MCRB system

The open-loop performance of the continuous MCRB was assessed at different dilution rates. A total volume of 1 L synthetic media with a constant substrate feed concentration of 60 g L^{-1} glucose in a 2 L bioreactor was used during the experiment. The system ran in continuous mode for a total of 64 hours and monitored online with Raman spectroscopy. Four different dilution rates of 0.05, 0.1, 0.15, 0.2 and 0.3 h^{-1} were tested at a constant bleed ratio of 0.1. The calibration model and kinetic model parameters determined during batch fermentation were utilized to make both Raman predictions and a model simulation of the continuous fermentation. The time elapsed results of the continuous fermentation with Raman predicted measurements of glucose and ethanol, and corresponding HPLC measurements are shown in Figure 4.3. Substrate conversion remained relatively constant until a dilution rate of 0.2 h^{-1} was reached. At a dilution rate of 0.3 h^{-1}, the substrate concentration steadily increased while the ethanol concentration decreased until the steady state washout of the system was reached and fermentation ended. Approximate stable steady states were achievable with the continuous open-loop MCRB, with productivity ranging from 1 - 5 g L^{-1} h^{-1} and total substrate conversion greater than 95%. Productivity reached nearly 7 g L^{-1} h^{-1} for the continuous fermentation at a dilution rate of 0.3 h^{-1}, over twice the productivity during batch fermentation. A comparison of the average productivity and substrate conversion obtained from measurements of offline HPLC, Raman spectra, and kinetic model simulation of the continuous MCBR system are shown in Table 4.2.

The model was set with constant bleed ratio of 0.1 and given the initial conditions measured from the HPLC at the start of the continuous fermentation. The kinetic model, Raman measurements, and HPLC results demonstrate strong agreement. The percent difference between HPLC and Raman measurements ranged from 1 – 4% and 0 – 4% for productivity and conversion, respectively. The percent difference was slightly higher between HPLC and model predictions ranging from 0 – 11% and 0 – 5% for productivity and conversion, respectively. This suggests that the kinetic parameters estimated from the batch fermentation as well as the calibration model developed during batch are both suitable for continuous fermentation predictions.

Table 4.2. Average productivity and conversion at different dilution rates calculated from HPLC data, Raman measurements, and continuous MCRB model.

Dilution	Productivity			Conversion		
Rate	HPLC	Raman	Model	HPLC	Raman	Model
0.05	1.27	1.30	1.18	0.99	0.98	0.99
0.1	2.59	2.63	2.52	1.00	0.99	0.99
0.15	3.87	3.73	3.87	1.00	0.97	0.99
0.2	4.66	4.50	5.19	0.95	0.98	0.99
0.3	6.71	6.66	6.79	0.87	0.84	0.86

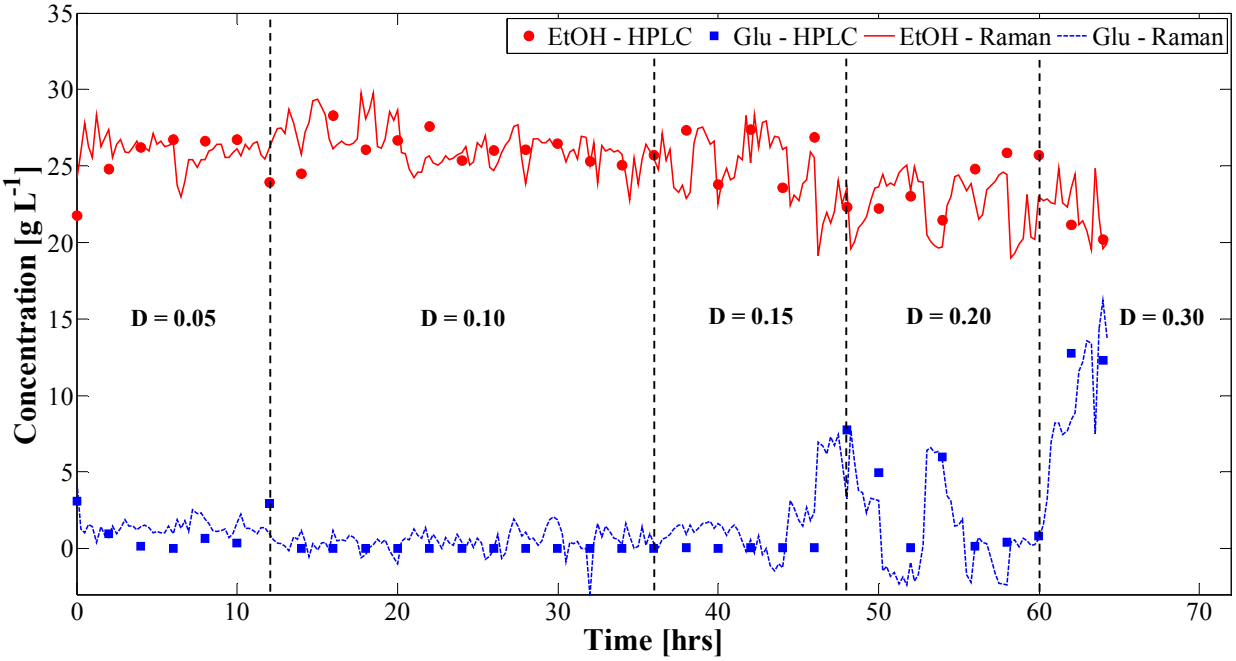

Figure 4.3. Comparison of Raman measurements (lines) of glucose and ethanol and HPLC data (markers) measured during open-loop continuous fermentation in MCRB.

4.4.3 Closed-loop performance of MPC of MCRB

Simulations of the closed-loop performance of the MCRB system were performed with the NMPC described above using the nominal kinetic parameters for the plant model with state estimates from the EKF and dilution rate as input from the NMPC. Substrate feed was simulated as having a nominal value of 45 g L^{-1} $\pm$ 10% white noise and the bleed ratio was held constant at 0.1. The first simulation tested plant-model mismatch, by adjusting the model kinetic parameters within the NMPC to be $\theta \pm 50\%$ of their nominal values, in addition, the EKF simulated Raman measurements from the plant added an additional $\pm$ 20% white noise. The control objective was to follow the productivity setpoint trajectory, while optimizing the control moves of the manipulated variable (D). As can be seen in Figure 4.4, the controller performs well in the presence of the plant-model mismatch. At times the plant productivity deviates from the setpoint,

but the offset is minimized. When the productivity jumps from 2.5 to 4 g L^{-1} h^{-1}, the plants total substrate conversion drops down below the constraint, but regains 100% conversion within 2 hrs.

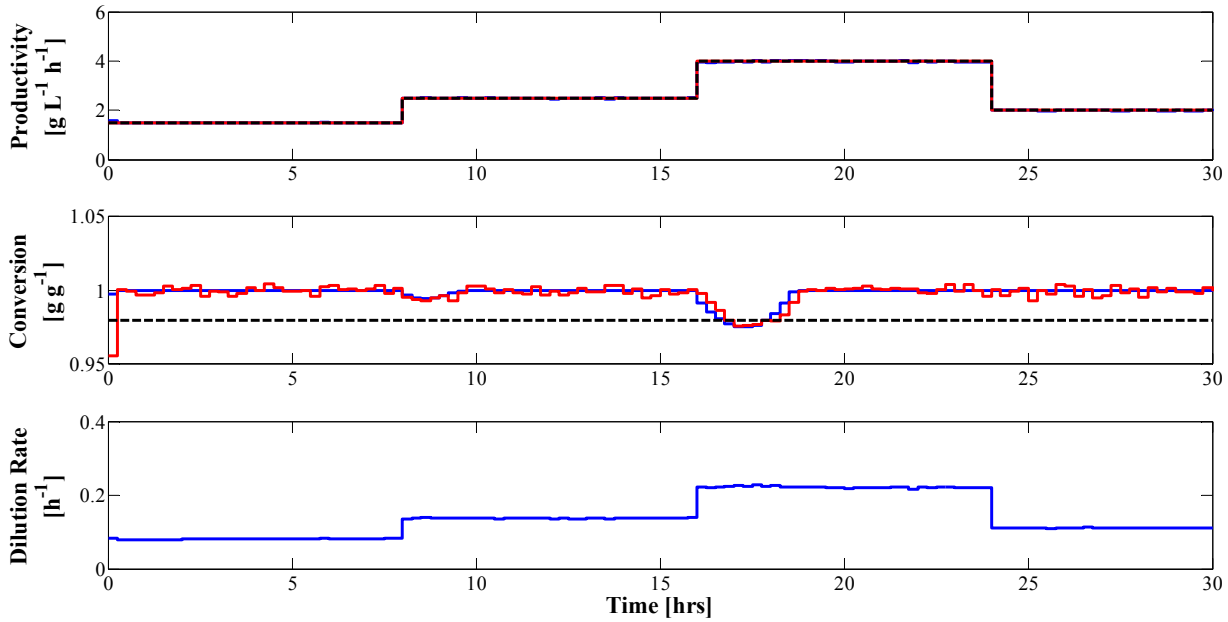

Figure 4.4. Simulation of closed-loop NMPC MCRB with plant-model mismatch and white noise added to measurements of glucose and ethanol. (Top) productivity of plant (red – line), model (blue – line), and setpoint (black – dashed line); (Middle) total substrate conversion of plant (red – line), model (blue – line) and constraint (black – dashed line); (Bottom) dilution rate as input to plant from NMPC.

The second simulation, tested the system when there is an abrupt disturbance in the substrate feed concentration. This is a realistic disturbance expected in biorefineries as the feedstock entering can have variable properties in the amount of sugars available for fermentation [88]. Between operational hours 16 and 24 the substrate feed was reduced to 30 g L^{-1}, from the nominal 45 g L^{-1} setting. It can be seen in Figure 4.5 that the NMPC reacts to this change by stepping up the dilution rate to counteract decreasing ethanol concentration. The controller is unable to reject this disturbance and the plant productivity shows offset from the setpoint. This could be overcome in future designs by including disturbance rejection for offset free tracking into the MPC model [89].

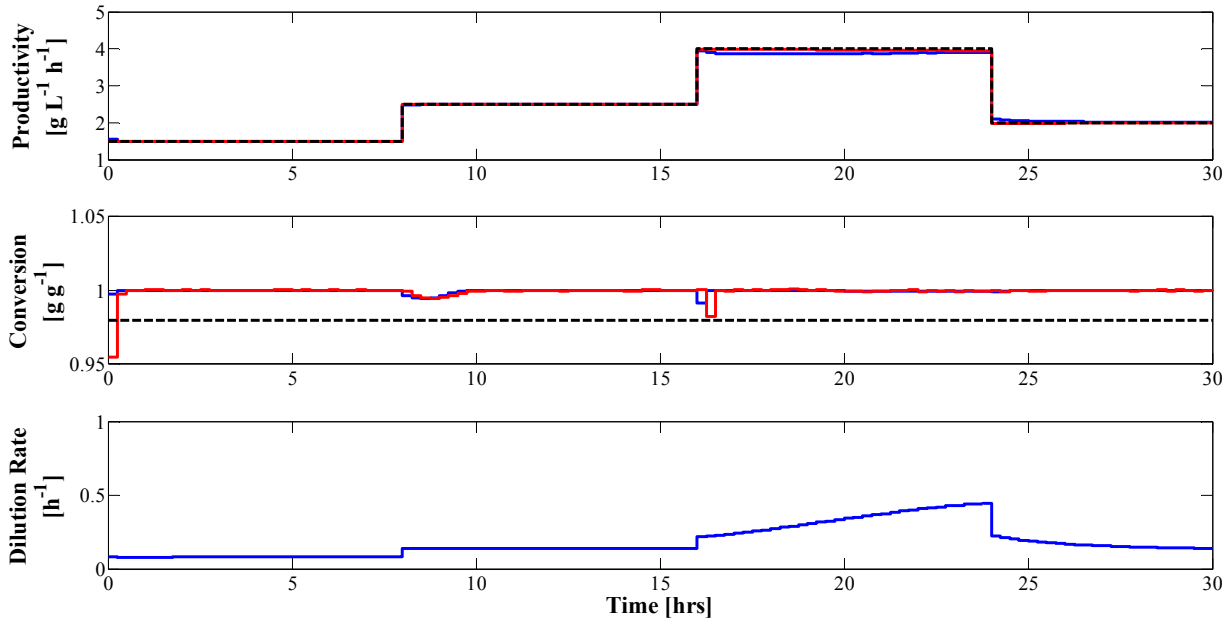

Figure 4.5. Simulation of closed-loop NMPC MCRB with disturbance in substrate feed concentration from 16 – 24 hrs going from 45 to 30 g L^{-1}. (Top) productivity of plant (red – line), model (blue – line), and setpoint (black – dashed line); (Middle) total substrate conversion of plant (red – line), model (blue – line) and constraint (black – dashed line); (Bottom) dilution rate as input to plant from NMPC.

The NMPC design was tested in the experimental MCRB with real-time monitoring using Raman sensors and full output feedback from the EKF. Measurements of glucose and ethanol concentrations were made in real-time at 15 minute intervals using Raman spectroscopy. At each measurement the EKF and NMPC algorithms were updated. The plot in Figure 4.6 shows the time elapsed time measurements of glucose (blue line – Raman, blue squares - HPLC), ethanol (red line – Raman, red circles – HPLC), cell concentration (black line – EKF, black triangles – OD), and substrate feed (green line – Raman). Raman and EKF measurements were started during the batch mode, which lasted six hours. Compared to the HPLC measurements, the RMSEP for real-time measured glucose, ethanol and cell mass (EKF estimates) were 1.5, 1.1, and 2.3 g L^{-1}, respectively. The Raman measurements closely followed the off-line HPLC measurements taken, while the EKF converged to the off-line cell mass measurements once the continuous mode was initiated, but seemed to diverge towards the end. The performance of the

EKF could be improved by periodically reinitializing the EKF with more off-line measurements. The nominal HPLC concentration for the substrate feed was 45 g L^{-1} of glucose.

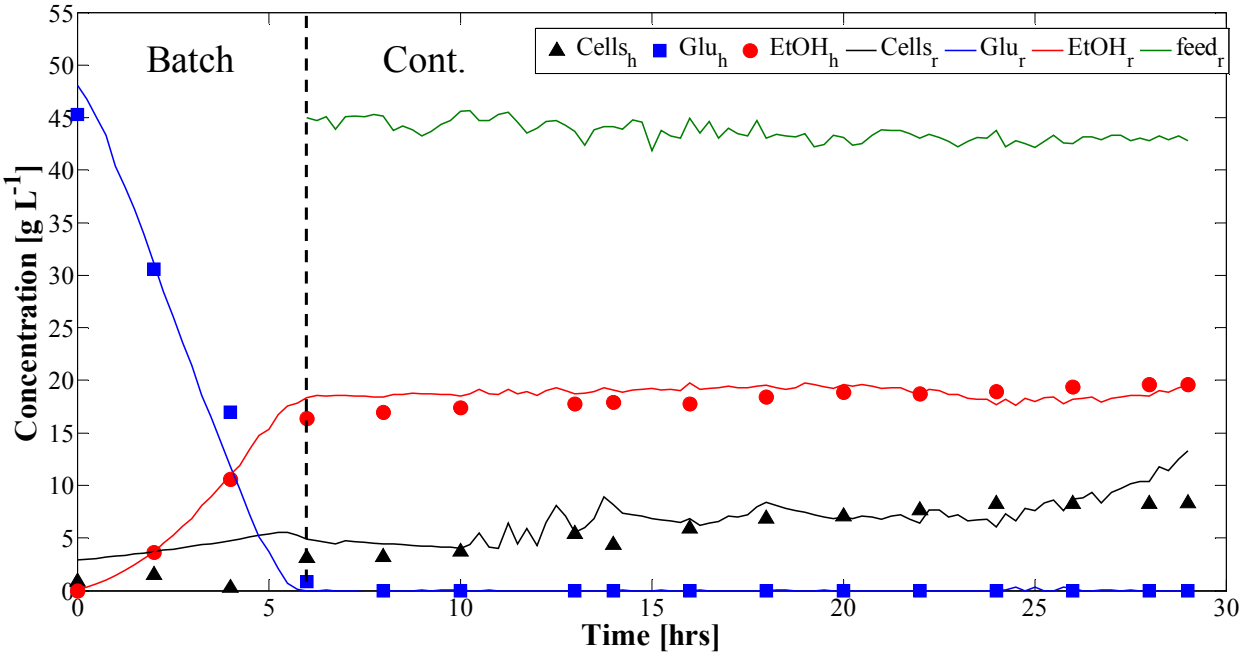

Figure 4.6. Time elapsed Raman measurements of glucose (blue – line), ethanol (red – line), and substrate feed (green – line); EKF estimates of cell mass concentration (black – line); off-line measurements of glucose (blue – squares), ethanol (red – circles), and cell mass (black – triangles) concentrations for experimental continuous fermentation in MCRB.

In Figure 4.7, the productivity and conversion measured by Raman sensors (blue lines) and off-line HPLC measurements (red circles), as well as the setpoint/constraints (black dashed line), and dilution rate from the NMPC are shown. The productivity was set to four different values of 2, 2.5, 3.5, and 4.5 g L^{-1} h^{-1}. The NMPC was able to handle changes in the setpoint productivity trajectory with some undershoot over the entire fermentation based on the Raman measurements, and closely followed the calculations of productivity from the off-line HPLC measurements of ethanol. The average productivity based on HPLC for the four setpoint regimes was 1.8, 2.3, 3.2, and 4.5 g L^{-1} h^{-1}. As with the simulation, incorporating disturbance rejection into the controller could reduce the amount of offset from the setpoint. The total substrate conversion remained greater than the constraint of 98% for the entire duration of the

fermentation. The dilution rate operated between 0.1 – 0.25 h^{-1} with the upper end occurring when the productivity was set to 4.5 g L^{-1} h^{-1}.

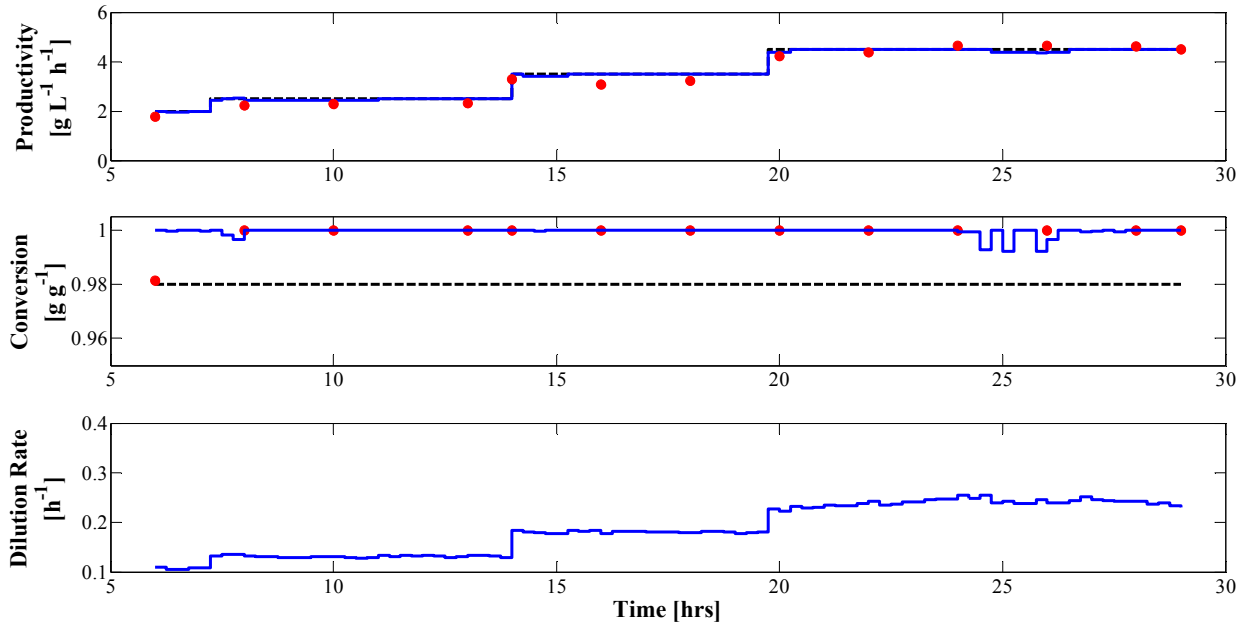

Figure 4.7. NMPC performance during experimental continuous fermentation in MCRB. (Top) productivity of based on Raman (blue – line), off-line HPLC (red – circles), and setpoint (black – dashed line); (Middle) total substrate conversion based on Raman (blue – line), off-line HPLC (red – circles), and constraint (black – dashed line); (Bottom) dilution rate as input to pumps from NMPC.

4.5 CONCLUSION

Control of bioprocesses has previously been difficult to achieve due to lack of suitable online measurement techniques and inherent nonlinearity in first principle bioprocess models. In this study Raman spectroscopy was shown to be especially useful for monitoring ethanol and glucose measurements in real-time for both batch and continuous fermentation systems. The multivariate PLS models yielded a percent prediction error of 2% and 3% for ethanol and glucose, respectively. The low percent error was achieved by appropriate preprocessing conditions and physically separating yeast cells from the analyte stream, which reduced cell scattering noise in the spectra. A first principle kinetic model for *S. cerevisiae* was developed based on batch fermentation data, and proved to be effective for predicting continuous fermentation behavior.

Simulations of the MCRB system using the kinetic model yielded less than 11% and 5% difference between offline HPLC measurements of productivity and total substrate conversion from experimental data.

A nonlinear model predictive control (NMPC) was applied to an experimental MCRB system. The NMPC utilized Raman measurements feedback measurements of glucose and ethanol, while an extended Kalman filter (EKF) was used to estimate the unknown biomass concentration for full-state feedback. The EKF demonstrated good agreement with offline measurements of the biomass concentration. The NMPC calculated an optimal dilution rate as input into the experimental fermentation system. This allowed the fermentation to fellow a productivity setpoint trajectory while achieving high substrate conversion. However, the error from the Raman measurements as well as process-model mismatch caused offset from the setpoint during the fermentation. This work shows that advanced process control can effectively be used for nonlinear bioprocesses.

Acknowledgements

This project is supported by Agriculture and Food Research Initiative Competitive Grant no. 2011-68005-30407 from the USDA National Institute of Food and Agriculture. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the view of the U.S. Department of Agriculture. The University of Washington Denman professorship fund provided financial support.

Chapter 5: Conclusion and Future Work

5.1 CONCLUSION

Advanced process control (APC) is critical to the future success of the biorefining industry. Spectroscopic technologies, such as Raman spectroscopy, have made it possible to monitor in real-time bioconversion reactions that can provide feedback measurements to APC systems. First principle models based on simple unstructured-unsegregated kinetics have proven to be useful for developing predictive models of complex fermentation systems. Model predictive control (MPC), an advanced process control strategy, is capable of utilizing nonlinear models and sensor feedback to provide optimal input while ensuring critical process constraints are met. The combination of modeling, monitoring and advanced process control of bioconversion systems could lead to improvement in process efficiency and reduce overall production costs, which are necessary to make lignocellulosic biofuels and biochemicals competitive with fossil-based fuels and chemicals.

In this work growth models for the acetogen *M. thermoacetica* were developed. This was the first attempt at developing a full kinetic model for this important microorganism. An unstructured-unsegregated mechanistic model was used to predict fermentation of both single sugars and a mixture of sugars (glucose and xylose) with diauxic growth. The model for both single and dual substrate fit the data well and captured the major features of the fermentation including catabolic repression. The model provided insights into how the maximum specific growth rate (μ_{max}) for *M. thermoacetica* can change in the presence of multiple sugars; showing a decrease in μ_{max} associated with the preferred substrate xylose and an increase in μ_{max} associated with glucose. This behavior is an important factor to consider when developing

models for future lignocellulosic bioconversion processes that will include multiple sugars in the fermentation media.

Monitoring of a complex anaerobic fermentation of lignocellulosic hydrolysate using *M. thermoacetica* was performed with Raman spectroscopy. Fermentation was performed using real hydrolysate from steam exploded pretreated sugarcane bagasse and sugarcane straw. Univariate models for predicting acetate concentrations performed best when the calibration data was from the same media complex as the fermentation data being predicted yielding a root mean square error of prediction (RMSEP) of 1.9 and 1.0 g L^{-1} for sugarcane bagasse and sugarcane straw, respectively. Multivariate partial least squares (PLS) models with appropriate preprocessing where much more robust in predicting acetate concentrations yielding a percent error of 5% for both sugarcane bagasse and sugarcane straw. The PLS models were also effective in predicting xylose, glucose, and total sugar concentrations. Furthermore, preliminary work has shown that calculations of the signal-to-noise ratio (SNR) and an approximation of the limit of detection (LOD) prior to conducting fermentation can help improve the overall calibration model development. Using this screening strategy, it was shown that detoxification of the hydrolysate using overliming was able to decrease the approximated LOD. This work was the first time Raman spectroscopy has been shown as an effective tool for monitoring anaerobic fermentation with the microorganism *M. thermoacetica.*

The application of a nonlinear model predictive control (NMPC) with closed-loop feedback from Raman sensors was applied to a continuous membrane cell-recycle bioethanol fermentation system. The dilution rate was used to control the ethanol productivity of the system will maintaining total substrate conversion above the constraint of 98%. The fermentation was performed using *S. cerevisiae* in both batch and continuous fermentation mode. PLS multivariate

models for glucose and ethanol were robust in predicting concentrations in both batch mode and continuous mode of operation. In addition, a kinetic model built upon data from batch fermentation was sufficient for predicting continuous fermentation behavior. A setpoint trajectory for productivity was closely tracked by the fermentation system using Raman measurements of glucose and ethanol concentrations and an extended Kalman filter to estimate biomass concentrations. Even with the inherently strong nonlinear behavior of the bioconversion system, stable steady states were achieved within appropriate operating conditions. This work demonstrates the effectiveness of this approach for real-time monitoring and control of a complex fermentation system. This was an important research achievement, as to our knowledge no experimental evidence has been published as to how the combination of these systems would perform in context of biofuels production.

5.2 FUTURE WORK

The following items are recommended for future work:

- An important area of research that still demands more attention for the improvement of processing lignocellulosic material is detoxification prior to bioconversion. These physical treatments have been demonstrated to improve not only fermentation results, but also reduce the impact of lignin induced fluorescence (LIF) on Raman spectra. A screening method was discussed in regards to improving the spectra of lignocellulosic hydrolysate prior to performing fermentation. Much work still needs to be done in regards to understanding how detoxification effects lignocellulosic hydrolysate and samples from Raman sensors. Monitoring the progress of these reactions in real-time could provide useful information regarding the optimization of process conditions, as well, as potentially provide insights into the unknown elements of the reaction.

- A model was developed for anaerobic fermentation of glucose and xylose using the acetogen *M. thermoacetica*. In addition, Raman spectroscopy was used to monitor the fermentation of lignocellulosic hydrolysates to acetate using this important bacterium. Thus an obvious extension to this work would be to continue the development of model predictive control of a membrane cell-recycle bioreactor with *M. thermoacetica* and lignocellulosic hydrolysate.

Chapter 6: References

[1] T. R. Brown, "A techno-economic review of thermochemical cellulosic biofuel pathways.," *Bioresour. Technol.*, vol. 178, pp. 166–176, Sep. 2014.

[2] R. Schnepf and B. Yacobucci, "Renewable fuel standard (RFS): overview and issues," *CRS Rep. Congr.*, vol. R40155, 2013.

[3] M. a. Carriquiry, X. Du, and G. R. Timilsina, "Second generation biofuels: Economics and policies," *Energy Policy*, vol. 39, no. 7, pp. 4222–4234, Jul. 2011.

[4] V. Balan, "Current Challenges in Commercially Producing Biofuels from Lignocellulosic Biomass," *ISRN Biotechnol.*, vol. 2014, pp. 1–31, 2014.

[5] T. Samad and A. M. Annaswamy, "The impact of control technology," 2011.

[6] J. H. Lee and J. M. Lee, "Progress and challenges in control of chemical processes.," *Annu. Rev. Chem. Biomol. Eng.*, vol. 5, pp. 383–404, Jan. 2014.

[7] M. Bauer and I. Craig, "Economic assessment of advanced process control–a survey and framework," *J. Process Control*, vol. 18, no. 1, pp. 2–18, 2008.

[8] K. Yamuna Rani and V. Ramachandra Rao, "Control of fermenters-a review," *Bioprocess Biosyst. Eng.*, vol. 21, no. 1, pp. 77–88, 1999.

[9] M. Henson, "Exploiting cellular biology to manufacture high-value products," *IEEE Control Syst. Mag*, vol. 26, no. 4, pp. 54–62, 2006.

[10] S. J. Qin and T. a. Badgwell, "A survey of industrial model predictive control technology," *Control Eng. Pract.*, vol. 11, no. 7, pp. 733–764, Jul. 2003.

[11] A. Ashoori, B. Moshiri, A. Khaki-Sedigh, and M. R. Bakhtiari, "Optimal control of a nonlinear fed-batch fermentation process using model predictive approach," *J. Process Control*, vol. 19, no. 7, pp. 1162–1173, Jul. 2009.

[12] a U. M. Kiran and A. K. Jana, "Control of continuous fed-batch fermentation process using neural network based model predictive controller.," *Bioprocess Biosyst. Eng.*, vol. 32, no. 6, pp. 801–8, Oct. 2009.

[13] S. Ramaswamy, T. J. Cutright, and H. K. Qammar, "Control of a continuous bioreactor using model predictive control," *Process Biochem.*, vol. 40, no. 8, pp. 2763–2770, Jul. 2005.

[14] Z. K. Nagy, "Model based control of a yeast fermentation bioreactor using optimally designed artificial neural networks," *Chem. Eng. J.*, vol. 127, no. 1–3, pp. 95–109, Mar. 2007.

[15] P. Mhaskar and S. Aumi, “Transition from Batch to Continuous Operation in Bio-Reactors : A Model Predictive,” *Can. J. Chem. Eng.*, vol. 85, pp. 416–423, 2007.

[16] M. Morari and J. H. Lee, “Model predictive control : past , present and future,” vol. 23, pp. 667–682, 1999.

[17] D. Q. Mayne, J. B. Rawlings, C. V Rao, and P. O. M. Scokaert, “Constrained model predictive control : Stability and optimality,” *Automatica*, vol. 36, pp. 789–814, 2000.

[18] R. Findeisen, L. Imsland, F. Allgower, and B. Foss, “State and output feedback nonlinear model predictive control: An overview,” *Eur. J. Control*, vol. 9, pp. 179–195, 2003.

[19] M. Cannon, “Efficient nonlinear model predictive control algorithms,” *Annu. Rev. Control*, vol. 28, no. 2, pp. 229–237, Jan. 2004.

[20] D. Dochain, “State and parameter estimation in chemical and biochemical processes: a tutorial,” *J. Process Control*, vol. 13, no. 8, pp. 801–818, Dec. 2003.

[21] I. R. Lewis and H. G. M. Edwards, *Handbook of Raman spectroscopy: from the research laboratory to the process line*, vol. 28. New York, NY: CRC Press, 2001.

[22] M. H. Brooker, G. Hancock, B. C. Rice, and J. Shapter, “Raman frequency and intensity studies of liquid H2O, H218O and D2O,” *J. Raman Spectrosc.*, vol. 20, pp. 683–694, 1989.

[23] C.-J. Shih, J. S. Lupoi, and E. A. Smith, “Raman spectroscopy measurements of glucose and xylose in hydrolysate: role of corn stover pretreatment and enzyme composition,” *Bioresour. Technol.*, vol. 102, no. 8, pp. 5169–5176, Apr. 2011.

[24] A. Shaw and N. Kaderbhai, “Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive Raman spectroscopy and chemometrics,” *Appl. Spectrosc.*, vol. 53, no. 11, pp. 1419–1428, 1999.

[25] A. Picard, I. Daniel, G. Montagnac, and P. Oger, “In situ monitoring by quantitative Raman spectroscopy of alcoholic fermentation by *Saccharomyces cerevisiae* under high pressure,” *Extremophiles*, vol. 11, no. 3, pp. 445–452, May 2007.

[26] T. M. Battaglia, E. E. Dunn, M. D. Lilley, J. Holloway, B. K. Dable, B. J. Marquardt, and K. S. Booksh, “Development of an in situ fiber optic Raman system to monitor hydrothermal vents.,” *Analyst*, vol. 129, no. 7, pp. 602–6, Jul. 2004.

[27] U. P. Agarwal, “An overview of Raman spectroscopy as applied to lignocellulosic materials,” in *Advances in lignocellulosics characterization*, D. S. Argyropoulos, Ed. Atlanta GA: TAPPI Press, 1999, pp. 201–225.

[28] N. Afseth, V. Segtnan, and J. Wold, “Raman spectra of biological samples: A study of preprocessing methods,” *Appl. Spectrosc.*, vol. 60, no. 12, pp. 1358–1367, 2006.

[29] C. Lieber and A. Mahadevan-Jansen, “Automated method for subtraction of fluorescence from biological Raman spectra,” *Appl. Spectrosc.*, vol. 57, no. 11, pp. 1363–1367, 2003.

[30] T. B. Shope, T. J. Vickers, and C. K. Mann, “The direct analysis of fermentation products by Raman spectroscopy,” *Appl. Spectrosc.*, vol. 41, no. 5, pp. 908–912, 1987.

[31] S. M. Ewanick, W. Thompson, B. Marquardt, and R. Bura, “Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy,” *Biotechnol. Biofuels*, vol. 6, no. 28, 2013.

[32] C.-J. Shih and E. a Smith, “Determination of glucose and ethanol after enzymatic hydrolysis and fermentation of biomass using Raman spectroscopy.,” *Anal. Chim. Acta*, vol. 653, no. 2, pp. 200–6, Oct. 2009.

[33] L. S. Mendes, F. C. C. Oliveira, P. A. Z. Suarez, and J. C. Rubim, “Determination of ethanol in fuel ethanol and beverages by Fourier transform (FT)-near infrared and FT-Raman spectrometries,” *Anal. Chim. Acta*, vol. 493, no. 2, pp. 219–231, 2003.

[34] S. R. Gray, S. W. Peretti, and H. H. Lamb, “Real-time monitoring of high-gravity corn mash fermentation using in situ raman spectroscopy.,” *Biotechnol. Bioeng.*, vol. 110, no. 6, pp. 1654–62, Jun. 2013.

[35] T. C. Avila, R. J. Poppi, I. Lunardi, P. a G. Tizei, and G. a G. Pereira, “Raman spectroscopy and chemometrics for on-line control of glucose fermentation by Saccharomyces cerevisiae.,” *Biotechnol. Prog.*, vol. 28, no. 6, pp. 1598–604, 2012.

[36] K. Olofsson, M. Bertilsson, and G. Lidén, “A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks.,” *Biotechnol. Biofuels*, vol. 1, no. 1, p. 7, Jan. 2008.

[37] J. De Gelder, K. De Gussem, P. Vandenabeele, and L. Moens, “Reference database of Raman spectra of biological molecules,” no. April, pp. 1133–1147, 2007.

[38] M. R. Almeida, R. S. Alves, L. B. L. R. Nascimbem, R. Stephani, R. J. Poppi, and L. F. C. de Oliveira, “Determination of amylose content in starch using Raman spectroscopy and multivariate calibration analysis.,” *Anal. Bioanal. Chem.*, vol. 397, no. 7, pp. 2693–701, Aug. 2010.

[39] K. V Gernaey, A. E. Lantz, P. Tufvesson, J. M. Woodley, and G. Sin, “Application of mechanistic models to fermentation and biocatalysis for next-generation processes.,” *Trends Biotechnol.*, vol. 28, no. 7, pp. 346–54, Jul. 2010.

[40] L. Dewasme and P. Bogaerts, “Monitoring of Bioprocesses: Mechanistic and Data-Driven Approaches,” *Tech. Bioprocess*, pp. 57–97, 2009.

[41] Y. Liu, “Overview of some theoretical approaches for derivation of the Monod equation.,” *Appl. Microbiol. Biotechnol.*, vol. 73, no. 6, pp. 1241–50, Jan. 2007.

[42] W. Harder, L. Dijkhuizen, and J. R. Postgate, “Strategies of Mixed Substrate Utilization in Microorganisms,” *Phil. Trans. R. Soc. Lond. B*, vol. 297, pp. 459 – 480, 1982.

[43] M. Sauer, D. Porro, D. Mattanovich, and P. Branduardi, “Microbial production of organic acids: expanding the markets,” *Trends Biotechnol.*, no. January, 2008.

[44] D. Bacovsky, N. Ludwiczek, M. Ognissanto, and M. Wörgetter, “Status of Advanced Biofuels Demonstration Facilities in 2012: A Report to IEA Bioenergy Task 39,” no. March, 2013.

[45] S. Ewanick, E. Schmitt, R. Gustafson, and R. Bura, “Use of Raman spectroscopy for continuous monitoring and control of lignocellulosic biorefinery processes,” *Pure Appl. Chem.*, vol. 86, no. 5, pp. 867–879, 2014.

[46] H. L. Drake and S. L. Daniel, “Physiology of the thermophilic acetogen Moorella thermoacetica.,” *Res. Microbiol.*, vol. 155, no. 6, pp. 422–36, 2004.

[47] Y. Kwon and C. Engler, “Kinetic models for growth and product formation on multiple substrates,” *Biotechnol. Bioprocess Eng.*, vol. 10, no. 1, pp. 587 – 592, 2005.

[48] J. R. Andreesen, a Schaupp, C. Neurauter, a Brown, and L. G. Ljungdahl, “Fermentation of glucose, fructose, and xylose by Clostridium thermoaceticum: effect of metals on growth yield, enzymes, and the synthesis of acetate from CO 2 .,” *J. Bacteriol.*, vol. 114, no. 2, pp. 743–51, May 1973.

[49] N. Dowe and J. M. Nrel, “SSF Experimental Protocols — Lignocellulosic Biomass Hydrolysis and Fermentation Laboratory Analytical Procedure (LAP) Issue Date : 10 / 30 / 2001 SSF Experimental Protocols — Lignocellulosic Biomass Hydrolysis and Fermentation Laboratory Analytical Pro,” no. January, 2008.

[50] N. Balasubramanian, J. S. Kim, and Y. Y. Lee, “Fermentation of xylose into acetic acid by Clostridium thermoaceticum.,” *Appl. Biochem. Biotechnol.*, vol. 91–93, pp. 367–76, Jan. 2001.

[51] G. Wang and D. Wang, “Elucidation of growth inhibition and acetic acid production by Clostridium thermoaceticum,” *Appl. Environ. Microbiol.*, vol. 47, no. 2, pp. 294–298, 1984.

[52] K. Z. Yao, B. M. Shaw, B. Kou, K. B. McAuley, and D. W. Bacon, "Modeling Ethylene/Butene Copolymerization with Multi-site Catalysts: Parameter Estimability and Experimental Design," *Polym. React. Eng.*, vol. 11, no. 3, pp. 563–588, Jan. 2003.

[53] K. Sugaya, D. Tuse, and J. Jones, "Production of acetic acid by Clostridium thermoaceticum in batch and continuous fermentations," *Biotechnol. Bioeng.*, vol. 28, no. 5, pp. 673 – 683, 1986.

[54] S. Sivakesava, J. Irudayaraj, and a Demirci, "Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy.," *J. Ind. Microbiol. Biotechnol.*, vol. 26, no. 4, pp. 185–90, Apr. 2001.

[55] J. Iversen, R. Berg, and B. Ahring, "Quantitative monitoring of yeast fermentation using Raman spectroscopy," *Anal. Bioanal. Chem.*, pp. 4911–4919, 2014.

[56] J. a Iversen and B. K. Ahring, "Monitoring lignocellulosic bioethanol production processes using Raman spectroscopy.," *Bioresour. Technol.*, vol. 172, pp. 112–20, Nov. 2014.

[57] S. Sivakesava, J. Irudayaraj, and D. Ali, "Simultaneous determination of multiple components in lactic acid fermentation using FT-MIR, NIR, and FT-Raman spectroscopic techniques," *Process Biochem.*, vol. 37, no. 4, pp. 371–378, Dec. 2001.

[58] H. Lee, P. Boccazzi, and N. Gorret, "In situ bioprocess monitoring of Escherichia coli bioreactions using Raman spectroscopy," *Vib. Spectrosc.*, vol. 35, pp. 131–137, 2004.

[59] Y.-S. Jang, B. Kim, J. H. Shin, Y. J. Choi, S. Choi, C. W. Song, J. Lee, H. G. Park, and S. Y. Lee, "Bio-based production of C2-C6 platform chemicals.," *Biotechnol. Bioeng.*, vol. 109, no. 10, pp. 2437–59, Oct. 2012.

[60] B. Schiel-Bengelsdorf and P. Dürre, "Pathway engineering and synthetic biology using acetogens.," *FEBS Lett.*, vol. 586, no. 15, pp. 2191–8, Jul. 2012.

[61] I. Y. Sengun and S. Karabiyikli, "Importance of acetic acid bacteria in food industry," *Food Control*, vol. 22, no. 5, pp. 647–656, May 2011.

[62] R. M. Busche, "Extractive fermentation of acetic acid: Economic trade-off between yield of CLostridium and concentration of acetobacter.," *Appl. Biochem. Biotechnol.*, vol. 28–29, pp. 605–621, 1991.

[63] A. Lähdetie, P. Nousiainen, J. Sipilä, T. Tamminen, and A.-S. Jääskeläinen, "Laser-induced fluorescence (LIF) of lignin and lignin model compounds in Raman spectroscopy," *Holzforschung*, vol. 67, no. 5, pp. 531–538, Jan. 2013.

[64] A. Vajzovic, R. Bura, K. Kohlmeier, and S. Doty, “Novel endophytic yeast *Rhodotorula mucilaginosa* strain PTD3 II: production of xylitol and ethanol in the presence of inhibitors,” *J. Ind. Microbiol. Biotechnol.*, vol. 39, no. 10, pp. 1453–63, Oct. 2012.

[65] A. Mohagheghi, M. Ruth, and D. J. Schell, “Conditioning hemicellulose hydrolysates for fermentation: Effects of overliming pH on sugar and ethanol yields,” *Process Biochem.*, vol. 41, no. 8, pp. 1806–1811, Aug. 2006.

[66] S. Ewanick and R. Bura, “The effect of biomass moisture content on bioethanol yields from steam pretreated switchgrass and sugarcane bagasse.,” *Bioresour. Technol.*, vol. 102, no. 3, pp. 2651–8, Feb. 2011.

[67] S. Chen, X. Lin, C. Yuen, S. Padmanabhan, R. W. Beuerman, and Q. Liu, “Recovery of Raman spectra with low signal-to-noise ratio using Wiener estimation.,” *Opt. Express*, vol. 22, no. 10, pp. 12102–14, May 2014.

[68] A. Olivieri, “Analytical Figures of Merit: From Univariate to Multiway Calibration,” *Chem. Rev.*, vol. 114, no. 10, pp. 5358–78, May 2014.

[69] M. Kacuráková and M. Mathlouthi, “FTIR and laser-Raman spectra of oligosaccharides in water: characterization of the glycosidic bond.,” *Carbohydr. Res.*, vol. 284, no. 2, pp. 145–57, May 1996.

[70] K. Ito and H. Bernstein, “The vibrational spectra of the formate, acetate, and oxalate ions,” *Can. J. Chem.*, vol. 34, no. 2, pp. 170–178, 1956.

[71] R. S. Uysal, E. A. Soykut, I. H. Boyaci, and A. Topcu, “Monitoring multiple components in vinegar fermentation using Raman spectroscopy.,” *Food Chem.*, vol. 141, no. 4, pp. 4333–43, Dec. 2013.

[72] M. Mathlouthi and D. Luu, “Laser-Raman spectra of d-glucose and sucrose in aqueous solution,” *Carbohydr. Res.*, vol. 81, pp. 203–212, 1980.

[73] F. Allegrini and A. Olivieri, “IUPAC-Consistent Approach to the Limit of Detection in Partial Least-Squares Calibration,” *Anal. Chem.*, vol. 86, no. 15, pp. 7858–66, Aug. 2014.

[74] C. Carrasco, H. M. Baudel, J. Sendelius, T. Modig, C. Roslander, M. Galbe, B. Hahn-Hägerdal, G. Zacchi, and G. Lidén, “SO2-catalyzed steam pretreatment and fermentation of enzymatically hydrolyzed sugarcane bagasse,” *Enzyme Microb. Technol.*, vol. 46, no. 2, pp. 64–73, Feb. 2010.

[75] M. Kadleõ, “Raman spectra of synthetic sapphire,” vol. 32, pp. 955–958, 2001.

[76] T. F. Edgar, “Control and operations: when does controllability equal profitability?,” *Comput. Chem. Eng.*, vol. 29, no. 1, pp. 41–49, Dec. 2004.

[77] S. Brethauer and C. E. Wyman, “Bioresource Technology Review : Continuous hydrolysis and fermentation for cellulosic ethanol production,” *Bioresour. Technol.*, vol. 101, no. 13, pp. 4862–4874, 2010.

[78] H. N. Chang, I. K. Yoo, and B. S. Kim, “High density cell culture by membrane-based cell recycle,” *Biotechnol. Adv.*, vol. 12, no. 3, pp. 467–487, 1994.

[79] P. Wei, L. Cheng, L. Zhang, and X. Xu, “A review of membrane technology for bioethanol production,” *... Sustain. Energy ...*, vol. 30, pp. 388–400, 2014.

[80] J. S. Alford, “Bioprocess control: Advances and challenges,” *Comput. Chem. Eng.*, vol. 30, no. 10–12, pp. 1464–1475, Sep. 2006.

[81] A. P. Ferreira, L. M. Vieira, J. P. Cardoso, and J. C. Menezes, “Evaluation of a new annular capacitance probe for biomass monitoring in industrial pilot-scale fermentations.,” *J. Biotechnol.*, vol. 116, no. 4, pp. 403–9, Apr. 2005.

[82] N. D. Lourenço, J. a Lopes, C. F. Almeida, M. C. Sarraguça, and H. M. Pinheiro, “Bioreactor monitoring with spectroscopy and chemometrics: a review.,” *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 1211–37, Sep. 2012.

[83] A. U. M. Kiran and A. K. Jana, “Control of continuous fed-batch fermentation process using neural network based model predictive controller.,” *Bioprocess Biosyst. Eng.*, vol. 32, no. 6, pp. 801–8, Oct. 2009.

[84] S. Craven, J. Whelan, and B. Glennon, “Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller,” *J. Process Control*, vol. 24, no. 4, pp. 344–357, Apr. 2014.

[85] E. Rivera, A. Costa, and R. Andrade, “Development of adaptive modeling techniques to describe the temperature-dependent kinetics of biotechnological processes,” *Biochem. ...*, vol. 36, pp. 157–166, 2007.

[86] R. Schneider and C. Georgakis, “How To NOT Make the Extended Kalman Filter Fail,” *Ind. Eng. Chem. Res.*, vol. 52, no. 9, pp. 3354–3362, Mar. 2013.

[87] L. Grüne and Jürgen Pannek, *Nonlinear model predictive control*. London: Springer, 2011.

[88] W. C. Wong, H.-S. Song, J. H. Lee, and D. Ramkrishna, “Hybrid cybernetic model-based simulation of continuous production of lignocellulosic ethanol: Rejecting abruptly changing feed conditions,” *Control Eng. Pract.*, vol. 18, no. 2, pp. 177–189, Feb. 2010.

[89] M. Morari and U. Maeder, “Nonlinear offset-free model predictive control,” *Automatica*, vol. 48, no. 9, pp. 2059–2067, Sep. 2012.