รายงานวิจัยฉบับสมบูรณ์

โครงการ การสร้างและการเปรียบเทียบแบบจำลองสำหรับกระบวนการผลิตไซลิทอล

Development and Comparison of Models for Xylitol Production Process

โดย ผศ.ดร. รวิพิมพ์ ฉวีสุข และคณะ

30 ธันวาคม พ.ศ. 2553

สัญญาเลขที่ MRG4780025

รายงานวิจัยฉบับสมบูรณ์

โครงการ การสร้างและการเปรียบเทียบแบบจำลองสำหรับกระบวนการผลิตไซลิทอล

Development and Comparison of Models for Xylitol Production Process

คณะผู้วิจัย	สังกัด
1. ผศ.ดร. รวิพิมพ์ ฉวีสุข	มหาวิทยาลัยเกษตรศาสตร์
2. รศ.ดร. สาโรจน์ ศิริศันสนียกุล	มหาวิทยาลัยเกษตรศาสตร์

สนับสนุนโดยสำนักงานคณะกรรมการการอุดมศึกษา

และสำนักงานกองทุนสนับสนุนการวิจัย

(ความเห็นในรายงานนี้เป็นของผู้วิจัย สกอ. และ สกว. ไม่จำเป็นต้องเห็นด้วยเสมอไป)

บทคัดย่อ

รหัสโครงการ MRG4780025

ชื่อโครงการ การสร้างและการเปรียบเทียบแบบจำลองสำหรับกระบวนการผลิตไซลิทอล

ชื่อนักวิจัย รวิพิมพ์ ฉวีสุข มหาวิทยาลัยเกษตรศาสตร์
สาโรจน์ ศิริศันสนียกุล มหาวิทยาลัยเกษตรศาสตร์

E-Mail Address : ravipim.c@ku.ac.th

ระยะเวลาโครงการ : 2 ปี

การสร้างแบบจำลองนั้นมีประโยชน์ต่อการศึกษาพฤติกรรม การติดตาม และปรับปรุงระบบของกระบวนการชีวภาพซึ่งมีความซับซ้อนอยู่โดยธรรมชาติ งานวิจัยนี้ศึกษาความเป็นไปได้ในการใช้แบบจำลองการถดถอย แบบจำลอง Dual kriging และแบบจำลองเครือข่ายประสาทเทียม 2 ชนิด คือ Backpropagation (BPN) และ Cascade Correlation (CCLN) ในการจำลองกระบวนการผลิตไซลิทอล ข้อมูลที่ใช้ในการสร้างแบบจำลองรวบรวมจากกระบวนการผลิตไซลิทอลแบบต่อเนื่องจาก *Candida mogii* ในระบบหมุนเวียนเซล (Cell recycling) ซึ่งได้ตรวจสอบอิทธิพลของอัตราการหมุนเวียนเซล (Recycle ratio) และอัตราการให้อากาศ (Aeration rate) ต่อปริมาณเซลและผลผลิตไซลิทอล แบบจำลองที่ศึกษาจะแสดงความสัมพันธ์ระหว่างสภาวะในการผลิตไซลิทอลได้แก่ อัตราการหมุนเวียนเซล อัตราการให้อากาศ และเวลาการหมัก กับผลลัพท์ของกระบวนการได้แก่ปริมาณเซลและผลผลิตไซลิทอล โดยสร้างแบบจำลองแยกกันระหว่างผลลัพท์ที่สนใจแต่ละตัว แบ่งข้อมูลทั้งหมดเป็น 3 ส่วนเพื่อสร้างแบบจำลอง เลือกพารามิเตอร์ที่เหมาะสม และทวนสอบแบบจำลอง ในการเลือกพารามิเตอร์นี้ได้ทดลองใช้รูปแบบของแบบจำลองการถดถอยแบบขั้นบันไดหลายระดับ และทดลองใช้โครงสร้างและพารามิเตอร์ในการเรียนรู้ของแบบจำลอง BPN และ CCLN หลายประเภท วัดความสามารถในการใช้งานทั่วไปของแบบจำลองที่สร้างขึ้นโดยการประเมินด้วยข้อมูลชุดทวนสอบ เปรียบเทียบประสิทธิภาพของแบบจำลองด้วยค่าความถูกต้องในการทำนายค่าผลลัพท์ของกระบวนการและค่าความลำเอียงของแบบจำลองจากข้อมูลทุกชุด สำหรับแบบจำลองที่ใช้ในการทำนายปริมาณเซลได้ถูกต้องแม่นยำ มีความสามารถในการใช้งานทั่วไปสูงที่สุด และไม่ลำเอียงได้แก่แบบจำลอง CCLN ในขณะที่แบบจำลองที่ใช้ในการทำนายปริมาณผลผลิตไซลิทอลได้ถูกต้องแม่นยำ และมีความสามารถในการใช้งานทั่วไปสูงที่สุดได้แก่แบบจำลอง BPN อย่างไรก็ตามแบบจำลอง BPN นี้มีความลำเอียงเชิงลบเล็กน้อย การนำไปใช้จึงต้องตระหนักถึงประเด็นนี้เสมอ โดยรวมแล้วแบบจำลองทางสถิติคือ แบบจำลองการถดถอย และ Dual kriging นั้นมีประสิทธิภาพต่ำกว่าแบบจำลองชนดกล่องดำหรือแบบจำลองเครือข่ายประสาทเทียมในการจำลองความสัมพันธ์ของกระบวนการผลิตไซลิทอลในงานวิจัยนี้

คำสำคัญ: การผลิตไซลิทอล; แบบจำลองการถดถอย; แบบจำลองเครือข่ายประสาทเทียม; แบบจำลอง Backpropagation; แบบจำลอง Cascade Correlation; แบบจำลอง Dual kriging

ABSTRACT

Project Code : MRG4780025

Project Title : การสร้างและการเปรียบเทียบแบบจำลองสำหรับกระบวนการผลิตไซลิทอล

Investigator : รวิพิมพ์ ฉวีสุข มหาวิทยาลัยเกษตรศาสตร์

E-Mail Address : ravipim.c@ku.ac.th

Project Period : 2 ปี

Modeling approach can be useful in understanding the behavior, monitoring and improving of the bioprocess system which is generally very complicated in nature. The potential use of polynomial regression, dual kriging, backpropagation neural network (BPN) and cascade correlation neural network (CCLN) in empirically modeling the xylitol production process were studied. The data used were collected from a continuous xylitol production by *Candida mogii* cell recycling where the effects of recycle ratio and aeration rate on the cell biomass and xylitol concentration were investigated. This research attempted to approximate the relationships between fermentation conditions (recycle ratio, aeration rate and fermentation time) and outputs of the system (cell biomass and xylitol cocentration). A separated model was developed for each output of interest. The entire data were divided into three data sets for building and selecting proper models, and validating them. Various functional forms of stepwise polynomial regression, various architectures, training parameters of BPN and CCLN were explored. Generalization capability of the models was evaluated using the validation data set. The performance of the model was assessed by the prediction accuracy across all data sets and by the model bias. For the cell biomass predictive model, the CCLN model was superior to BPN, dual kriging and polynomial regression models in terms of prediction accuracy and generalization capability with no bias. For the xylitol concentration predictive model, the BPN model outperforms the CCLN, dual kriging and polynomial regression models with respect to prediction accuracy. However, this BPN model is very slightly underestimates the data. Consequencely, care must be taken when using it. On the whole, the statistical based models as polynomial regression and dual kriing were quite inferior to the alternative block box modeling techniques such as neural networks in approximate the relationship of the xylitol production process in this research.

Keywords: Xylitol production; Modeling techniques; Polynomial regression; Backpropagation neural network; Cascade correlation learning neural networks

Contents

Development and Comparison of Models for Xylitol Production Process

Introduction

Xylitol is a pentahydroxy sugar alcohol that is as approximately sweet as sucrose. It readily dissolves in water, has pleasant cool and fresh sensation and can reduce the formation of dental caries (Greenby, 1992). Due to these properties, xylitol has found its wide applications as a sweetener in various food products such as chewing gum, candies, ice cream, beverages and some pharmaceutical products. These food products are fast growing products in the current market. New products are launched to the market every few month. Bakery products, spices, jams, jellies and dessert represent potential applications of xylitol in food products (Emodi, 1978) making it of high value to food industry. Study on every aspect of the xylitol production would therefore be very beneficial. Xylitol can be produced by several processes such as extraction from fruits and vegetables, chemical reduction of xylose, and bioprocess. Extraction is an uneconomical method due to high cost and relatively low xylitol content in fruits and vegetables (Washuttl et al., 1978). Chemical reduction is commercially used (Counsell, 1978) but is very expensive and high level of contaminants from the production process makes it difficult to be purified. This might limit its use in the industries. As a result, bioprocess receives more attention from researchers during the past decade. Xylitol can be fermented from bacteria, fungi or yeast. Yeasts are the best xylitol producer, particularly *Candida spp* (Winkelhausen and Kuzmanova, 1998).

Bioprocess is a complicated dynamic system. Understanding process behavior or system identification such that the process can be well controlled, predicted or redesigned is always a difficult task. Traditionally, mathematical models are used to represent complex effects of processing inputs on the productivity or quality of the outputs. Building mathematical models require a high level of a *priori* insight about the system that is rarely available due to the complex and nonlinear behavior of the bioprocess. Consequently, building these types of models is time

consuming and may not well represent the system. If a certain level of insight about the system exists, an empirical statistical model such as regression is an alternative. When little knowledge of the system is known, an empirical black-box model such as artificial neural network (ANN) is a sound choice.

This research will focus on modeling xylitol production using regressions, dual kriging (a geostatisical model) and artificial neural networks and compare their performances. It is an extension of the work of Sirisansaneeyakul and Tochampa on xylitol production using *Canida mogii* (Sirisansaneeyakul et al., 2000[1,2]). Regression and response surface methodology (RSM) are the most popular nonlinear modeling method. Although the regression and RSM have a good theoretical background and is straightforward to implement, they requires restrictive assumptions on the error terms and their performance depends on the appropriateness of the polynomial functional forms. ANN has been recognized as an alternative for modeling nonlinear system in the past decade. The model requires little or no priori assumption of functional forms and rather it attempts to learn from the training input-output examples or the so-called “learning by example”. It is also robust to deviations from traditional statistical assumptions such as independently normal random errors, common error variance and multicollinearity. An ANN with nonlinear transfer functions can theoretically model any relationship to an arbitrary accuracy and is thus termed a universal approximator (Funahashi, 1989; Hornik et al., 1989). A wide range of applications has utilized these features of ANN. These include pattern classification, speech production and recognition, function approximation, signal processing, image compression, associative memory, clustering, combinatorial optimization, nonlinear modeling, and control. ANN applications in bioprocess are more widespread in alcoholic fermentation and recombinant fermentation. It is observed that all of these applications utilize the most popular backpropagation neural network (BPN).

This research attempts to explore a potential use of two types of function approximation ANNs: BPN and cascade-correlation learning network (CCLN). A BPN is a feedforward multilayer neural network trained by gradient descent (Rumehart and McClelland, 1986) that minimizes the

total squared error of the output computed by the network. The training algorithm involves three stages: the feed-forward of input training set, the calculation and backpropagation of the error and the adjustment of the weights. Drawbacks to BPN are large computational time due to back propagating the errors and adjusting all the weights simultaneously as well as the difficulty in selecting the proper architecture, i.e., the number of hidden neurons and hidden layers. The CCLN was developed by Fahlman and Lebiere (Touretzky, 1990) and incorporates two key ideas: cascade architecture and the maximization of the correlation between a new unit's output and the residual error during learning. The cascade architecture starts with only input and output neurons and connection weights are adjusted to minimize the total squared error. Candidate hidden neurons are then added, one at a time, to reduce the error. Due to its constructive algorithm, the CCLN will automatically find the proper architecture of the network, however since the final number of hidden neurons is unbounded an overparameterized network may result (Tang and Wah, 1996).

While regression models have restricted assumptions of the uncorrelated error components, correlation might exist among the sample data observed from physical phenomena as in bioprocess. Data close together, in time or in space, are likely to be correlated and should be modeled as such (Cressie, 1991). Dual kriging is a modeling technique that allows the incorporation of spatial correlation into the interpolation or estimation process. Accordingly, it might be an alternative modeling technique to better represent the input-output relationship from xylitol production which appears to be a set of spatial data. Dual kriging has been adopted in several applications including mining, environment, physical and chemical compositions and behaviors and financial analysis. This research will be a pioneer to employ dual kriging in bioprocess modeling. Though the work centers on xylitol production, the knowledge earned here will definitely be fruitful to other bioprocesses.

Objectives

1. Examine the potential use of several-order polynomial regression, backpropagation neural network (BPN), cascade correlation learning network (CCLN) and dual kriging in modeling xylitol production process.

2. Compare the performance of the best identified regression, dual kriging, BPN and CCLN models and make a recommendation on xylitol production.

Literature Review

The Occurrence and Properties of Xylitol

Xylitol ($C_5H_{12}O_5$) is a naturally occurring pentahydroxy sugar alcohol in many fruits and vegetables. Yellow plum, strawberry, cauliflower, raspberry, lettuce, spinach, onion, carrot, grape, and banana are examples of xylitol's natural sources. Among these, yellow plum is the richest source, containing almost 1% on a dry basis (Aminoff, et. al, 1978). Yeast, lichens, seaweed and mushroom are other natural sources of xylitol. Xylitol is also a metabolic intermediate in mammalian carbohydrate metabolism. In human adults, 5 to15 grams of xylitol can be produce per day. Since xylitol is metabolized independently of insulin, it will not fluctuate the insulin and glucose blood levels and thus can be used as diabetic sweetener (Touster, 1974; Bassler, 1978; Emodi, 1978; Bar, 1991; Makinen,1992). In addition, this property is useful for post-operative or post-traumatic states of patients as well as for correction of catabolic disorders (peripheral lipolysis, stimulation of glucogenesis, and degradation of muscle protein) (Forster, 1974; Ritzel and Brubacher, 1976). As xylitol does not react with amino acid, its utilization for parenteral nutrition is then possible. Moreover, its metabolism does not involve glucose-6-phosphate dehydrogenase and is therefore an ideal sweetener for glucose-6-phospahte dehydrogenase-deficient population. Xylitol has also an anti-ketonic effect and is very well received in post-surgery infusions in patients with difficulty in metabolizing sugar (Sanronan et al., 1991).

Xylitol possesses many advantageous characteristics and has thereby received much research attention as food ingredient in the last three decades (Aminoff, et. al, 1978; Emodi, 1978; Ylikahri, 1979; Pepper and Olinger, 1988; Pepper, 1989). It does not undergo Maillard reaction which leads to food browning and reduction in nutritional of protein value. The addition of xylitol in food products can improve the color and taste without undesirable changes during their storage. Xylitol is as sweet as sucrose, nearly twice as sweet as sorbitol and approximately three times as sweet as mannitol. Its caloric content is equal to that of sucrose, 17kJ / kg. Xylitol, alone or in combination with other sugars, is shown to be a beneficial sweetener in yoghurt, jams and frozen desserts as it provides better texture, color and taste and stability compare to sucrose (Abril et al., 1982). Xylitol produces a cool and fresh sensation on oral and nasal cavities as a result of its negative heat of dissolution. It can then be used as part of coating of confectionary or

pharmaceutical products such as vitamins or expectorants (Pepper and Olinger, 1988) and in the formulation of dietary complements (Petrovich, 1988).

The most significant characteristic of xylitol in commercial implications is anticariogenic property. Xylitol is not utilized by the acid producing, cariogenic bacteria in human oral cavity and therefore inhibits their growth, formation of plaque and deminerization of tooth enamel and the formation of new dental caries (Bar, 1988). Bar (1991) and Makinen (1992) consider it as the best alternative sweetener for caries prevention. As a consequence, much is consumed in chewing gum, confectionary, mouthwash and toothpaste. In toothpaste, xylitol also shows ability to retain moisture (Mori and Saraya, 1988).

Xylitol Production

Three major procedures are available for xylitol production: solid-liquid extraction, chemical synthesis and bioprocess.

1. Solid-liquid extraction

Natural xylitol found in fruits, vegetables and other natural sources can be recovered from by solid-liquid extraction. However, due to its low concentration in these sources, the extraction becomes difficult and uneconomical (Hyvonen et al., 1982; Pepper and Olinger, 1988).

2. Chemical synthesis

At present, xylitol is commercially produced by chemical synthesis. The general procedures composes of 4 main steps: (1) acid-catalyzed hydrolysis of plant materials; (2) purification of the hydrolysate to xylose solution or a pure crystalline xylose; (3) hydrogenation of the xylose to xylitol; and (4) crystallization of the xylitol (Aminoff et al., 1978). The major raw materials for manufacturing xylitol are xylans, which are present in hardwoods (birch and beech trees and some plant structural tissues such as corn-stalks, wheat, flax and rice straw, cotton seeds, sunflower or coconut hulls, sugarcane bagasse and wood pulp). These materials can be hydrolyzed to D-xylose and other sugars such as L-arabinose, D-mannose, and D-galactose (Krull and Inglett, 1980) with D-xylose as a major component (80-85%). These contaminating sugar can complicate crystallization and purification of xylose. As such, the critical step in the process is purification of xylose from the hydrolysate. This is achieved by employing ion-exchange chromatography. Activated carbon is also used to remove color. After that, catalytic hydrogenation of the purified xylose is carried out. The resulting solution requires chromatographic fractionation

and concentration before crystallization into purified xylitol. The intensive purification and separation steps are very expensive and thus making the production cost about ten times higher than that of other sugar alcohols undergone similar process. This limits the commercial use of xylitol despite its wide range of applications. Other concerns for xylitol production via chemical method are high pollution levels and waste-treatment.

3. Bioprocess

Biotechnological approach for xylitol production is based on the utilization of microorganisms and/or enzymes. It is an alterative process that might offer some cost and environmental friendly advantages over the chemical process. Bacteria, fungi and yeasts are capable of assimilating and fermenting xylose to xylitol, ethanol, and other compounds. Among these organisms, the yeasts are considered to be the best xylitol producers and thereby receiving most attention from researchers. Winkelhausen and Kuzmanova (1998) collected and compared the performance of xylitol production among various yeast strains from many publications and concluded that the genus *Candida* is the best xylitol producers. This summary coincides with the work of Ojamo (1994) who compared more than 30 yeast strains. The yeast conversion of D-xylose to xylitol begins with a transport of xylose across the yeast's cell membrane. The D-xylose uptake in *Candida moogii* ATCC 18364 were found to follow Michaelis-Menten kinetics which suggested a carrier-mediated facilitated diffusion transport system (Sirisansaneeyakul et. al., 1995). The xylose metabolism in yeasts was extensively studied and described (Barbosa, et al., 1988; Rizzi et al., 1988; Prior et al., 1989; Hahn-Hägerdal, 1994). In general, xylose will undergo an oxido-reductive route via two sequential reactions. Firstly, xylose reductase transforms xylose into xylitol in the presence of NADH and/or NADPH. Subsequently, xylitol is either secreted from the cell or oxidized to xylulose by xylitol dehydrogenase in the presence of either NAD+ or NADP+. Xylulose is then transformed via phosphorylation by xylulokinase to xylulose-5-phosphate which enters the pentose phosphate pathway (PPP). The PPP consists of an oxidative phase that leads to NADPH regeneration and a non-oxidative phase that produces glyceraldehydes-3-phosphate and fructose-6-phosphate. Both non-oxidative products can be converted to pyruvate in the Embden-Meyerhof-Parnas pathway. Pyruvate can either be decarboxylated and reduced to ethanol or can enter the tricarboxylic acid cycle. Figure 1 shows a simplified scheme of xylose metabolism in yeasts.

Some of the metabolic products and the cofactor regeneration are required for cell growth. One cannot just stop when xylose is converted to xylitol. In order to obtain good yields of xylitol, a

balance between the amount of xylose being converted to xylitol and the amount of xylitol being available for further metabolism must be obtained (Winkelhausen and Kuzmanova, 1998).

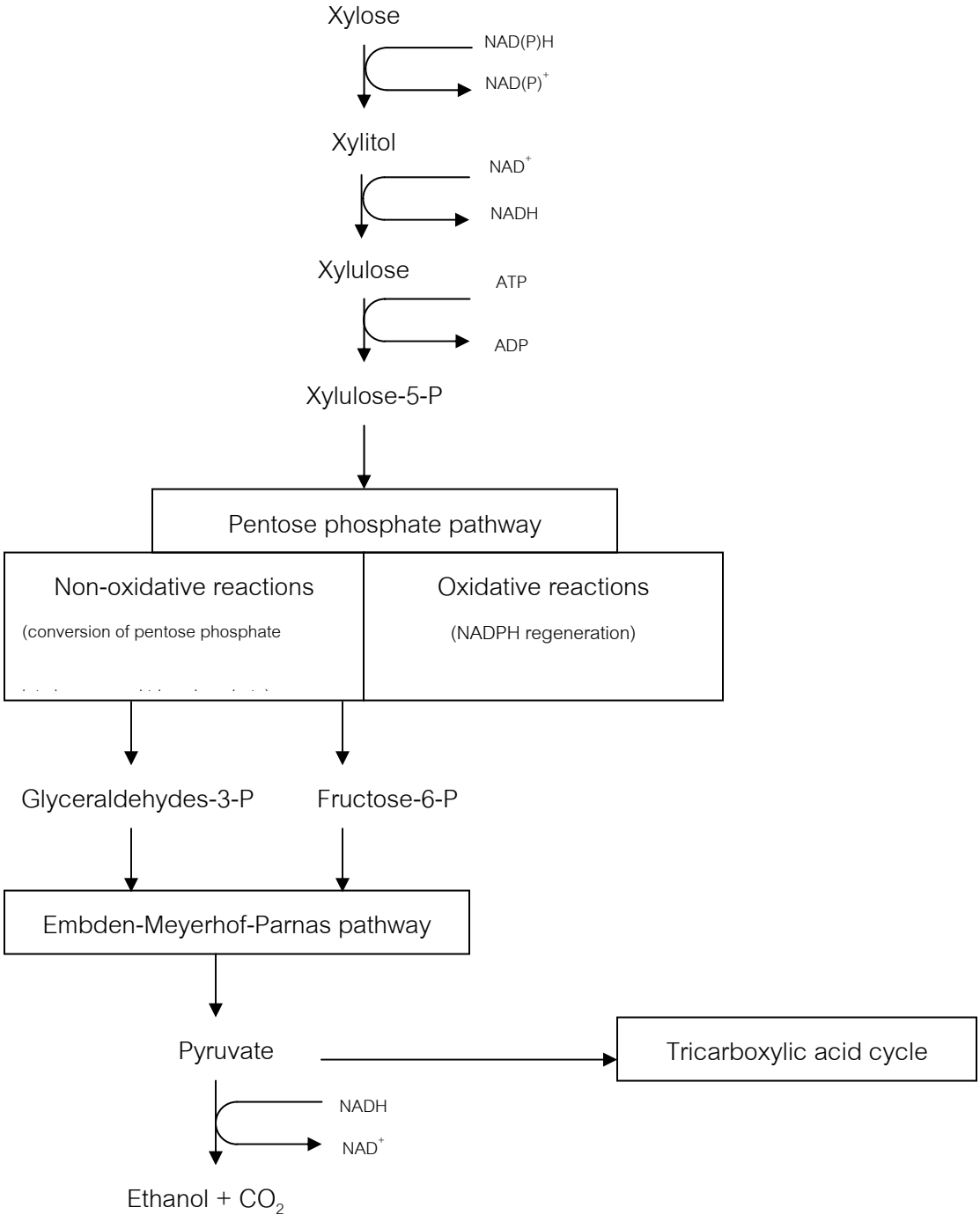

Figure 1 A simplified scheme of xylose metabolism in yeasts

Process Variables Influencing Xylitol Production

A number of experimental conditions are considered to influence the xylitol production from the yeasts with respect to yields and productivities (Nigam and Singh, 1995; Parajo et al., 1998a, 1998b; Winkelhausen and Kuzmanova, 1998). These conditions include nutrients, initial cell concentration, the culture age, temperature, pH, substrate composition and concentration, product concentration and aeration.

In general, the suitable temperature for xylitol production was observed to be 30^{O}C. When the yeast was cultured in a range of 30^{O}C and 37^{O}C, the xylitol cocnetration was mostly temperature independent. From the industrial perspective, lower temperature implies lower costs and more managerial conditions (Sanchez et al., 2004). The proper initial pH value would be between 4 and 7 depending on the yeast species and fermentation culture (batch or fed-batch or continuous). D-xylose concentration significantly affects the growth and the fermentation. High xylose concentration induces xylitol fermentation but inhibits ethanol production. For most yeasts, the initial xylose concentrations between 100 and 200 g/l would produce the highest yields (Winkelhausen and Kuzmanova, 1998).

When glucose was used as a co-substrate in xylitol production, many researchers (Nolleau *et al.,* 1995; Yahashi *et al.,* 1996a, 1996b; Sreenath and Jeffries, 1996) found that it would improve overall process. In the presence of glucose, xylose could be converted to xylitol more efficiently leading to faster cell growth than when xylose was the only substrate. On the other hands, other researchers reported that using glucose as a co-substrate led to faster cell growth but lower xylitol cocnetration (Silva et al., 1996; Vandeska, 1996). This findings was attributable to a partial inhibition of xylose reductase in the presence of glucose. The yeasts will consume glucose first and then use xylose once the glucose is completely utilized.

Xylitol is not produced under fully aerobic conditions while the yeasts fail to grow on xylose under anaerobic conditions (Parajo *et al.,* 1998b; Faria *et al*., 2002). Under a limited oxygen supply, NADH cannot be oxidized to NAD+, resulting in an inhibition of NAD+-linked xylitol dehydrogenase and thereby a decrease in the oxidation of xylitol to xylulose with an increase in xylitol accumulation. On the contrary, a sufficient oxygen supply will enable the oxidation of xylitol to xylulose and thus an increase in cell growth. Consequently, dissolved oxygen (DO) plays a very important role.

Modeling in Bioprocess

Bioprocess is a complicated dynamic system. Understanding process behavior or system identification such that the process can be well monitored and controlled, predicted, optimized or redesigned is always a difficult task. By construction and analysis of certain models, better knowledge of real-world bioprocess could be achieved. An accurate models is thereby a building block in improving the performance of a process through control, optimization and redesign.

Modeling techniques are divided into three major types:

1. Mathematic, mechanistic or white box modeling techniques

This type of model is constructed based on the underlying process principles (first principles) such as mass, energy and momentum balances. Although the model structure comes from the first principles, the model parameters are obtained from fitting the model structure to empirical data. However, developing an accurate model requires a considerable knowledge of the bioprocess physics, chemistry and microbiology that is rarely available. Consequently, the model constructed may not well represent the system and is too costly in practice since much time and effort must be consumed during the construction and validation.

2. Empirical modeling techniques

The construction of this type of model is based on empirical data of the process's behavior. The structure of the models is generic and cannot be interpreted in terms of mechanistic laws. However, little process knowledge is required. As a result, the cost of building this model is bearable. Either experimental data or actual plant data could be used to fit the model. Meanwhile, the application region of both data types will be different. The process model obtained from experimental data cannot be directly applicable to a real plant without some modifications. Several empirical modeling techniques are available. These include polynomial regression models, Taguchi models, generalized linear models, splines, radial basis functions, kernel smoothing, spatial correlation models (kriging), frequency-domain approximation, and artificial neural networks.

Barton (1992) suggests the following criteria to be considered in choosing an empirical modeling technique:

(1) The ability to gain insight from the form of the model.

(2) The ability to capture the shape of arbitrary smooth functions.

(3) The ability to characterize the accuracy of the fit.

(4) The robustness of the prediction away from observed (x, y) pairs.

(5) The ease of computation of the approximate the function of interest.

(6) The numerical stability of the computations and consequent robustness of prediction to small changes in the parameters defining function.

(7) The existence of software for computing the model, characterizing its fit, and using it for prediction.

3. Hybrid modeling techniques

A hybrid modeling technique can start by deriving a model based on the process principles and then includes black box elements as parts of the white box (Braake et al., 1998). In other words, the basic structure of the models is from the first principles while important relationships are modeled by mixed empirical / mechanistic relations. For instance, a structure of a process to be modeled is known priori and the neural network is trained to estimate the time varying unknown process variable (te Braake et al., 1998) or Psichogios and Ungar (1991) used a multi-layer feed forward neural network as the non-parametric estimator for the unknown process parameters in the first principle model.

This type of model can also obtained by incorporating prior knowledge of a process into a black box model during process modeling (van Deventer *et al*., 2004). For example, Lindskog and Ljung (1994) searched for the combinations or transformations of the input signals corresponding to physical variables and used the resulting signals in an empirical model.

In this research, only empirical modeling techniques are studied. Three of them are of interest and will be discussed in details. They are polynomial regression models, artificial neural networks and spatial correlation models.

Polynomial Regression Models

1. Polynomial regression models

Regression analysis is one of the most widely used of all statistical tools for modeling the input-output relationship. It serves three major purposes:

(1) to make inferences about the regression parameters

(2) to estimate the mean response for a given set of input variables

(3) to predict a new response for a given set of input variables

The polynomial regression model is the most frequently used curvilinear response model in practice. There are two types of variables in any regression model: the independent or predictor or

input variables (X) and the dependent or response variables (Y). Polynomial regression models can contain one, two, or more than two independent variables while each independent variable can be present in various powers. A polynomial regression model for n observations of pairs (x_i, y_i) can be expressed as:

$$Y_i = \sum_{j=1}^{l} \sum_{k=1}^{p} \beta_k Z_k(x_{ij}) + \varepsilon_i \qquad \textit{For } i = 1, 2, \ldots, n \tag{1}$$

where there are *p* power functions $Z_k(x_{ij})$; for example the power function might be 1, x_{i1}, x_{i2}, $x_{i1}x_{i2}$, x_{i1}^2, x_{i1}^3, or $x_{i3}^2 x_{i4}^3$. β_k are the regression coefficients which are to be estimated from the observed pairs of data points via least squares or maximum likelihood estimation. ε_i is a normal random error term with mean $E\{\varepsilon_i\} = 0$ and common variance $\sigma^2\{\varepsilon_i\} = \sigma^2$ so that the errors are not correlated with each other, i.e., the covariance $\sigma\{\varepsilon_i, \varepsilon_j\} = 0$ for all i, j; $i \neq j$.

When using a polynomial model as an approximation to the true regression function, a second-order or third-order model is often fitted and the possible adequacy of a lower-order model is then explored (Neter *et al*., 1990).

Multicollinearity or intercorrelation, i.e. where the independent variables are correlated among themselves, is unavoidable in polynomial models especially for high-order polynomials. A high degree of multicollinearity does not inhibit a good fit nor does it tend to affect the inference about mean responses or prediction of new responses, provided that these inferences are made within the region of observations. However, standard interpretations based on the regression coefficients, such as a large coefficient for a linear term indicating a significant effect of the independent variable, a large coefficient for a quadratic term indicating a non linear response, and a large coefficient for a cross product term ($x_i x_j$) indicating a change in the effect of one independent variable as a function of the value of the other (Barton, 1992), are often unwarranted. This is due to the large sampling variability of the estimated regression coefficients when the multicollinearity exists. In order to avoid this situation, all polynomial regression models should be formulated in terms of deviations, i.e. the independent variable is expressed as a deviation about its mean ($\overline{X}$) or $x_i = X_i - \overline{X}$ (Neter *et al*., 1990).

2. Strengths and weaknesses of polynomial regression models

Barton (1992) points out that polynomial regression models perform well with criteria (1) and (3)-(7) in Section ”modeling in bioprocess”. While straightforward to implement, the regression models require restrictive assumptions on the error terms. Their performance also depends on the

appropriateness of the polynomial functional forms. Polynomial regressions of all types may provide good fits but the accuracy of the predicted response will degrade with increasing distance from the experimental observation. The higher the order of the polynomials, the more rapidly the accuracy degrades.

3. Its applications in bioprocess

Its applications in bioprocess range from production of citric acid (Chen, 1996), streptomycin (Saval et al., 1993) and tannin acyl hydolase (Lekha et al., 1994) to cellulose (Shi and Weimer, 1992). The method has been used in xylitol production from *Candida tropicalis* (Horitsu et al., 1992) and *Candida duilliermondii* (Roberto et al., 1995). However, these works explored only up to second-order models and none was mentioned on the validity of the underlying assumptions that is very important for the model reliability.

Classical Kriging and Dual Kriging

Kriging is an estimation technique proposed in 1951 by D.G. Krige, a mining engineer, for gold deposit evaluations. Similar to polynomial regression, the kriging technique is also associated with the acronym BLUE, “Best Linear Unbiased Estimator” of a random function and is ‘best’ in terms of aiming at minimizing the variance of estimation error among all linear estimators (Poirer and Taniwa, 1991). Geologists and environmental engineers have been using kriging technique to estimate the measurements or characteristics of hydraulic properties or contaminant concentrations in air, water, or soil in regions that were inaccessible or unobserved. Later its use was extended to simulation community. The theory of classical kriging is well-covered by Journel and Huijbregts (1978). Classical kriging is usually implemented as a local estimation method. That is, its procedure requires the solution of a new system of equations for each interpolated value. According to Trochu (1993), a global estimation kriging technique called “dual kriging” was developed in 1985. Under dual kriging, the kriging system is evaluated only once for the whole domain by simultaneously using the information provided by all data points. The development of classical kriging equations and derivation of dual kriging is discussed based on Journel and Huijbregts (1978), Porier and Taniwa, 1991), and Trochu (1993) as follows.

1. Theory of classical kriging

Basically, the purpose of kriging is to estimate the value of a random function $U(X)$ at a specified point or location X, given a set of measurements or computed samples $U(X_i)$ taken at location X_i for $i = 1, 2, \ldots, N$. The original theory of kriging was formulated for dealing with one, two, or three dimensional problems, i.e., when X represents the position vector $X = x$ or $X = (x, y)$ or $X = (x, y, z)$. However, it can be generalized to an L dimension problem, i.e., $X = (x^1, x^2, \ldots, x^L)$.

The estimation of $U(X)$ can be obtained as a linear combination of the observed data point X_i where $i = 1, 2, \ldots, N$:

$$u^*(X) = \sum_{i=1}^{N} \lambda_i U(X_i) \tag{2}$$

As a BLUE, a set of weights λ_i must be determined in such as way that (1) the expected values of $U(X)$ and $u^*(X)$ are identical, i.e., $E[U(X)] = E[u^*(X)]$ and (2) the variance of the estimation error $Var[U(X) - u^*(X)]$ is minimized.

In kriging, the random function $U(X)$ is comprised of the sum of two terms:

$$U(X) = a(X) + b(X) \tag{3}$$

where $a(X)$ is a drift function representing the average behavior of $U(X)$ or $a(X) = E[U(X)]$, and $b(X)$ is a stationary fluctuation with $E[b(X)] = 0$.

The kriging system can be derived so as to minimize the variance of the estimation error under the constraints of unbiased conditions as follows:

From the unbiased condition, $E[U(X)] = E[u^*(X)]$, equation (2) can be expressed as

$$E[U(X)] = \sum_{i=1}^{N} \lambda_i E[U(X_i)] \tag{4}$$

Since the drift function represents the expected value of $U(X)$, equation (4) can be represented by

$$a(X) = \sum_{i=1}^{N} \lambda_i a(X_i) \tag{5}$$

In general, the drift function is built up from M basis functions, $p_l(X)$, l=1, 2, …, M and thus the conditions of unbiased become

$$p_l(X) = \sum_{i=1}^{N} \lambda_i p_l(X_i), \qquad l=1, 2, \ldots, M \tag{6}$$

The variance of the estimation error is calculated as follows

$$Var[U(X) - u^*(X)] = \sigma_R^2 = Var[U(X)] - 2Cov[U(X), u^*(X)] + Var[u^*(X)]$$

with $Var[U(X)] = \sigma_{U(X)}^2$

$$Cov[U(X), u*(X)] = Cov\left\{\left[\sum_{i=1}^{N}\lambda_i U(X_i)\right], U(X)\right\} = \sum_{i=1}^{N}\lambda_i Cov[U(X), U(X_i)]$$

$$Var[u*(X)] = Var\left[\sum_{i=1}^{N}\lambda_i U(X_i)\right] = \sum_{i=`}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j Cov[U(X_i), U(X_j)]$$

Combining these three terms again, the variance of estimation error can be expressed as

$$\sigma_R^2 = \sigma_{U(X)}^2 - 2\sum_{i=1}^{N}\lambda_i Cov[U(X), U(X_i)] + \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j Cov[U(X_i), U(X_j)] \quad (7)$$

This error variance is minimized subject to *M* unbiased conditions in (6). A Lagrangian technique is used to convert a constrained minimization problem into an unconstrained one by introducing *M* Lagrange multipliers, μ_l, *l*=1, 2,…, *M*, associated with the constraints. The solution is then characterized by a linear system of *N*+*M* equations in *N*+*M* unknowns $\lambda_1, \ldots, \lambda_N$ and $\mu_1, \ldots, \mu_M$:

$$\sum_{j=1}^{N}\lambda_j Cov[U(X_i), U(X_j)] + \sum_{l=1}^{M}\mu_l p_l(X_i) = Cov[U(X), U(X_j)], \quad i=1, 2, \ldots, N$$

(8)

$$\sum_{j=1}^{N}\lambda_j p_l(X_j) = p_l(X), \quad l=1, 2, \ldots, M.$$

This system is called the “kriging system” and can be written in matrix form

$$\begin{bmatrix} C_{ii} & \cdots & C_{1N} & p_1(X_1) & \cdots & p_M(X_1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C_{N1} & \cdots & C_{NN} & p_1(X_N) & \cdots & p_M(X_N) \\ p_1(X_1) & \cdots & p_1(X_N) & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p_M(X_1) & \cdots & p_M(X_N) & 0 & \cdots & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ \mu_1 \\ \vdots \\ \mu_M \end{bmatrix} = \begin{bmatrix} C_1 \\ \vdots \\ C_N \\ p_1(X) \\ \vdots \\ p_M(X) \end{bmatrix} \quad (9)$$

where C_{ij} denotes the covariance between sample points X_i and X_j, or $Cov[U(X_i), U(X_j)]$, and C_i is the covariance between sample points X_i and a point X, or $Cov[U(X), U(X_i)]$, in which the value of $U(X)$ is to be estimated. Solving this system yields the optimal values of the λ_i, *i*=1, 2, .., *N*, at the point *X*.

2. Dual formulation of kriging

The kriging system of equations (9) depends on the covariance between the sample point X_i and the point X. That is, the solution of system λ_i will depend on the point X. Therefore, a new kriging system would be needed for each estimated value. This can be computationally expensive for large problems. The dual formulation of kriging was developed to provide independent λ_i and thus eliminate this limitation of the classical kriging procedure. Dual kriging can be formulated from equation (9) as follows. When the matrix in system (9) is inverted, the following expression is obtained:

$$\begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ - \\ \mu_1 \\ \vdots \\ \mu_M \end{bmatrix} = \left[\begin{array}{ccc|c|ccc} & & & & & & \\ & Q & & & R & & \\ & & & & & & \\ \hline - & - & - & + & - & - & - \\ \hline & & & & & & \\ & R^T & & & S & & \\ & & & & & & \end{array}\right] \cdot \begin{bmatrix} C_1 \\ \vdots \\ C_N \\ - \\ p_1(X) \\ \vdots \\ p_M(X) \end{bmatrix} \tag{10}$$

By substituting this solution into equation (2), the estimated value $u^*(X)$ can be expressed as:

$$u^*(X) = \begin{bmatrix} U(X_1) & \cdots & U(X_N) \end{bmatrix} \cdot Q \cdot \begin{bmatrix} C_1 \\ \vdots \\ C_N \end{bmatrix} + \begin{bmatrix} U(X_1) & \cdots & U(X_N) \end{bmatrix} \cdot R \cdot \begin{bmatrix} p_1(X) \\ \vdots \\ p_M(X) \end{bmatrix} \tag{11}$$

By the symmetry of the kriging matrix, a new set of coefficients is defined as:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} = Q \cdot \begin{bmatrix} U(X_1) \\ \vdots \\ U(X_N) \end{bmatrix}, \qquad \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix} = R^T \cdot \begin{bmatrix} U(X_1) \\ \vdots \\ U(X_N) \end{bmatrix} \tag{12}$$

and thus equation (11) becomes

$$u^*(X) = \sum_{j=1}^{N} b_j C_j + \sum_{l=1}^{M} a_l p_l(X) \tag{13}$$

Equation (13) is called dual kriging. The coefficients a_l, l=1, …, M and b_j, j=1, …, N can be written in matrix form as:

$$\begin{bmatrix} b_1 \\ \vdots \\ b_N \\ - \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} & Q & & | & & A & \\ - & - & - & + & - & - & - \\ & R_T & & | & & B & \end{bmatrix} \cdot \begin{bmatrix} U(X_1) \\ \vdots \\ U(X_N) \\ - \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{14}$$

where the matrices A and B are arbitrary. By choosing $A = R$ and $B = S$, the matrix of equation (10), which is the inverse kriging matrix, appears again in equation (12). Hence, the coefficients a_l's and b_j's are solutions of

$$\begin{bmatrix} & C_{ij} & & | & & p_l(X_i) & \\ - & - & - & + & - & - & - \\ & p_l(X_j) & & | & & 0 & \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_N \\ - \\ a_1 \\ \vdots \\ a_M \end{bmatrix} = \begin{bmatrix} U(X_1) \\ \vdots \\ U(X_N) \\ - \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{15}$$

This system of linear equations together with the dual kriging model in equation (13) constitutes the dual formulation of kriging.

3. The drift function and the covariance functions

According to Journel and Huijbregts (1978), a polynomial drift function is normally used in geostatistics applications. This can be a complete polynomial of order k (k dimensions) composed of all possible subsets of variables of size 1 to k. For example, an order three basis is expressed as:

$$a(x) = a_0 + \sum_{i=1}^{L} a_i x_i + \sum_{i=1}^{L} \sum_{j=i}^{L} a_{ij} x_i x_j + \sum_{i=1}^{L} \sum_{j=i}^{L} \sum_{k=j}^{L} a_{ijk} x_i x_j x_k \tag{16}$$

It is apparent that dual kriging also requires the knowledge of the covariance between two points or locations. The covariance between two points or locations is assumed to depend only on the Euclidean distance h between X_i and X_j, and not on the particular positions X_i or X_j and is represented by $C(h)$. In general, the covariance function decreases from its maximum value at $C(0)$ since the degree of correlation between two locations decreases as the distance h between them increases. Ratle (1998) summarizes two approaches to be used for obtaining the covariance.

The first approach is to use an arbitrary theoretical covariance function. These functions are called shape functions rather than covariance functions since they have no relationship to the actual covariance. Kriging under these conditions is considered to be an exact interpolator. The other approach is to use the estimation of an experimental covariance function from the observed data. Under this condition, kriging is employed as an estimator. However, it is difficult to estimate a covariance function from the experimental data because it requires the knowledge of the unknown mean. Consequently, only theoretical covariance will be considered in this research.

Ratle (1998) has described three common theoretical covariance functions. The first covariance function is the pure nugget effect model which is the limiting case where the fluctuations around the samples are assumed to be insignificant. This model is appropriate for noisy data as well as for problems where only a rough estimate of the solution is required. The pure nugget effect covariance is written as:

$$C(h) = \begin{cases} 1 & if \quad h = 0; \\ 0 & otherwise. \end{cases} \tag{17}$$

Under the pure nugget effect, there is no correlation between two points regardless of distance *h*. The kriging model does not pass anymore through all the data points and it reduces to a simple polynomial regression on the drift function basis.

The other two models are based on the notion of distance of influence as introduced by Trochu [11]. The models assume that the correlation or actual covariance between two very distance points is negligible or zero. The general covariance *C*(*h*) may be designed in such a way that *C*(*h*) = 0 if *h* > *d*, where *d* is a predefined threshold. The first model is the linear model. It assumes that the covariance decreases linearly from a maximal value at *h* = 0 to zero at *h* = *d*. The linear covariance is expressed as:

$$C(h) = \begin{cases} 1 - h/d & if \quad h < d; \\ 0 & otherwise. \end{cases} \tag{18}$$

The other model is the cubic covariance. This model ensures continuity by imposing the nullity of the first derivative of *C*(*h*) at the points *h* = 0 and *h* = *d*. Two other conditions are *C*(*0*) = 1 and *C*(*d*) = 0. The covariance function is defined as:

$$C(h) = \begin{cases} 1 - 3(h/d)^2 + 2(h/d)^3 & if \quad h < d; \\ 0 & otherwise. \end{cases} \tag{19}$$

4. Strengths and weaknesses of dual kriging

Correlation might exist among the sample data observed from physical or social phenomena. Data close together, in time or in space, are likely to be correlated and should not be modeled as statistically independent (Cressie, 1991). The covariance function in dual kriging allows the incorporation of spatial correlation into the interpolation or estimation process. In addition, kriging has been suggested to be a useful technique in small sample estimation of environmental and mining data.[69,73] It appears that an dual kriging modeling technique can satisfy Barton's criteria (1), (3) and (4) for choosing modeling techniques discussed in Section "modeling in bioprocess". However, major applications of kriging and dual kriging are in geostatistics where spatial data are collected from locations defined on two or three dimensions. As a result, the available software has been also limited to three dimensions.

5. Its applications in bioprocess

Dual kriging has been adopted in several applications including stress analysis (Poirer and Tinawi, 1991), a contouring program (Trochu, 1993), shrinkage analysis (Mamat et. al., 1995), modeling of failure behavior for composite materials (Echaabi et al., 1995; 1996), a surrogate for fitness landscape in evolutionary optimization (Rattle, 1998), and a metamodeling technique for capital investment evaluation (Chaveesuk, 2000; Cahveesuk and Smith, 2005). However, there is no implementation in bioprocess modeling. Thus, this research will be the first research group to employ dual kriging in bioprocess modeling.

Artificial Neural Networks

An artificial neural network (ANN) is a parallel computational model consisting of a large number of simple, and highly interconnected adaptive processing elements (artificial neurons) in an architecture inspired by the structure of the biological nervous systems. The majority of the ANNs are closely related to traditional mathematical and/or statistical models such as non-parametric pattern classifiers, clustering algorithms, nonlinear filters, and statistical regression models. An important feature of the ANN lies in its adaptive nature where learning by example becomes a key factor in solving problems. This feature is very appealing in applications where little is known about the problem while observed data is readily available. A wide range of applications have

utilized these features of ANN. These include pattern classification, speech production and recognition, function approximation, signal processing, image compression, associative memory, clustering, combinatorial optimization, nonlinear modeling, and control.

ANN is characterized by (1) its architecture (a pattern of connections between the artificial neurons), (2) its training or learning algorithm (a method of determining the weights of the connections), and (3) its activation function (a function of the input that each neuron has received) (Fausett, 1994). Typically, the artificial neurons are organized into a sequence of layers with full or random connections between layers. Associated with each connection is a weight that represents the information being used to solve the problem. An example of a two-layer feedforward neural architecture is illustrated in figure 2 where the net comprises of a hidden layer and an output layer.

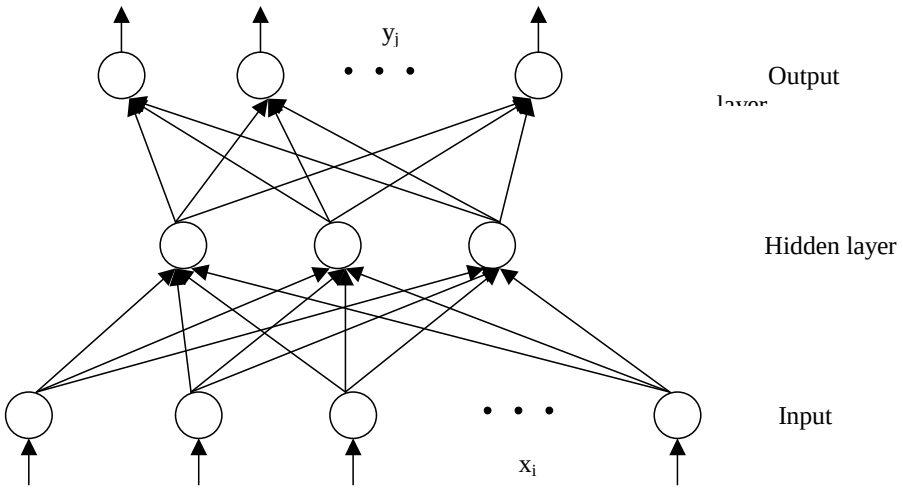

Figure 2 Example of a Typical ANN

Each neuron *j* in the hidden or the output layer sums its input signals x_i weighted by the connection weights w_{ij}, and applies an activation function to determine its output signal y_j. There exist many activation functions *f(z)*, ranging from a simple threshold function to complex non linear functions such as sigmoid, hyperbolic tangent and logistic functions. Figure 3 shows a schematic representation of a computing neuron.

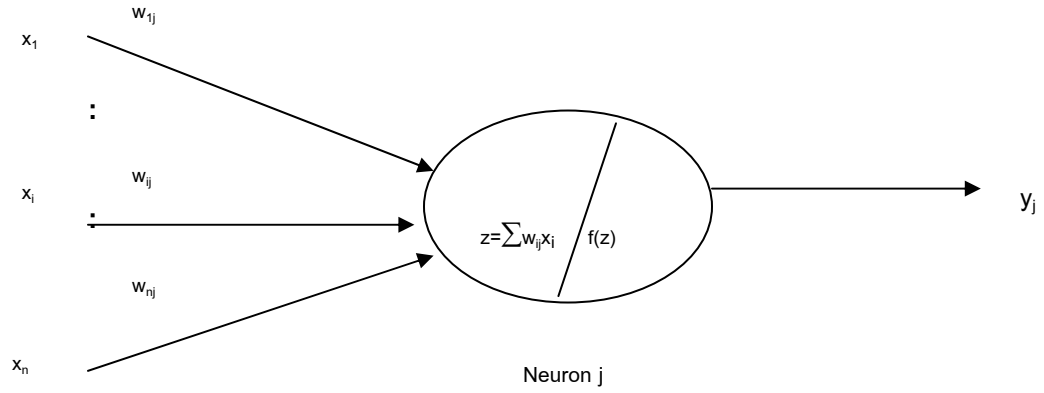

Figure 3 Schematic Representation of a Computing Neuron

An ANN is trained to learn the relationship between the input and output of a system via two main types of learning (training) algorithms: supervised and unsupervised. A supervised learning algorithm adjusts the weights of inter-neuron connections to minimize the difference between the desired and actual network outputs corresponding to a given input. An unsupervised learning algorithm does not require the desired or known output. During the training, only input patterns are presented to the neural network which will adapt the weights of its connections to cluster the input patterns into groups with similar features.

At the beginning of the training algorithm, a neural network starts with a randomized state of small initial weights. The weights are then iteratively adjusted until the ANN reaches a fixed and stable state where the appropriate weights are obtained to solve the problem. Typically, the algorithm is iterative until a balance between the ability to respond correctly to the input patterns that are used for training (memorization) and the ability to give good responses to input that is similar to that used in training (generalization) is achieved. It is common to use two sets of disjoint data from the same population during training: a training set and a testing set. The training patterns are used to adjust the weights during training whereas the testing patterns are used to estimate the generalization ability of the network at intervals during training. As training continues, the errors on the training set will continue to decrease while the errors on the testing set will decrease and then increase again when so-called overtraining occurs. Overtraining implies that the network loses its ability to generalize. One popular training stoppage criterion is, consequently, at the point where the minimum error on the testing set is reached.

There are various types of ANN models. This research will focus on supervised training neural networks. The most popular one used in function approximations of the nonlinear relationships is the backpropagation network (BPN). Alternative to this backpropagation is the cascade-correlation learning network (CCLN). Both networks will be examined in this research and thus deserve more discussion in detail.

1. Backpropagation network

A BPN is a feedforward multilayer neural network trained by the backpropagation training algorithm. The backpropagation, or the generalized delta rule, training algorithm was first proposed by Werbos (1974) as part of his Ph.D. dissertation. However, the elucidation of this training algorithm by Rumelhart *et al*. (1986) became the key step in reemergence of neural networks as a tool in solving a wide variety of problems. Basically, the backpropagation training algorithm is a gradient descent method which attempts to minimize the total squared error of the output computed by the network. The traditional and most applicable activation function employed in backpropagation algorithm is the sigmoid function and the hyperbolic tangent (TanH) as shown in equation (20) and (21), respectively.

$$f_i(z) = \frac{1}{1+e^{-z}} \tag{20}$$

$$f_i(z) = \frac{1-e^{-2z}}{1+e^{-2z}} \tag{21}$$

where z is the weighted sum of the input signals to the ith hidden- or output-layer neuron.

This supervised learning algorithm encompasses three stages: the feeding forward of the input training pattern, the backpropagation of the associated error, and the adjustment of the weights to reduce the squared error (Fausett, 1994). During the feeding forward, each input neuron receives an input signal and broadcasts it to the hidden neurons, each of which computes its activation and sends its output signal to each output neuron. Each output neuron then computes its activation to form the response of the net for the given input pattern. In the second stage, the output from each output neuron is compared to the target value to calculate the associated error. The error correction weight adjustment for each output neuron is then computed

and propagated back to all neurons in the hidden layer. Similarly, the error correction weight adjustment for each hidden neuron is computed for updating the weights between the hidden layer and the input. Finally, the weights for all layers are adjusted simultaneously based on the computed weight adjustment and the activation of each neuron. The BPN is considered to be a universal approximator since it can perform any continuous function approximation to an arbitrary accuracy (Funayashi, 1989; Hornik et al., 1989).

Major limitations associated with applying BPN are difficulty in selecting the proper architecture and learning parameters as well as convergence to local minima and instability (Leondes, 1998).

2. Cascade-correlation learning network

The cascade-correlation learning algorithm network (CCLN) was developed by Fahlman and Lebiere (1990) in an attempt to improve a slow pace of learning, which is a major drawback in BPN during those days.

Fahlman and Lebiere's CCLN incorporates two key ideas: the cascade architecture and the maximization of correlation between a new unit's output and the residual error during the learning algorithm. The cascade architecture starts off with input buffers and output units but with no hidden units. All input buffers and output units are directly connected with an adjustable weight. These connections are trained to minimize the total squared error. The output units may just produce a linear sum of their weighted inputs, or they may employ some non-linear activation function, particularly the hyperbolic tangent as previously described.

When there is no significant reduction in error, candidate hidden units are then added one at a time to incrementally reduce some residual error. The creation and training of these candidate units consists of two phases. In the first phase, the input of each candidate hidden unit is connected to every input buffer and also the output of every previously installed hidden unit whereas its output is not yet connected to the active network. Each new unit therefore adds a new one-unit layer to the network, leading to cascade connections. These cascade connections enable the development of the higher-order feature detecting capability. The weights on the input side of the new candidate unit are then trained using a gradient ascent to maximize S, the sum over all output units *o* of the magnitude of the correlation, or more precisely the covariance, between *V*, the candidate unit's value or activation, and E_o, the residual output error observed at unit *o*. Fahlman and Lebiere define S as:

$$S = \sum_o \left| \sum_p \left(V_p - \overline{V}\right)\left(E_{p,o} - \overline{E_o}\right) \right| \tag{22}$$

where *o* is the network output at which the error is measured,

p is the training pattern and

$\overline{V}$ and $\overline{E_o}$ are the values of *V* and E_o averaged over all patterns.

When there is no significant improvement in *S*, the training is terminated and the candidate unit is said to be tenured. This training can be thought of as a maximal alignment with the residual error (Phatak and Koren, 1994). If a candidate hidden unit can be trained to correlate positively with the error at a given output unit, it will cancel some of the residual error by developing a negative connection to the corresponding output-layer weight (to be trained later), and vice versa for the negative correlations (Teng and Wah, 1996). The input side connection weights of the tenured unit are frozen hereafter. In the second phase, the candidate unit's output is then connected to all the network's output units. These weight connections feeding the output units as well as those from the existing hidden units are trained to minimize the error without backpropagating the error through the hidden units. Phatak and Koren (1994) consider this phase as canceling the error as much as possible by exploiting the alignment accomplished in the first phase. This process is continued, and new units are added until the desirable error is obtained.

Using a CCLN offers two benefits. First, it will automatically find the size and the topology of the resulting ANN. The problem of overspecifying the number of hidden units, and thus overtraining, could therefore be alleviated. Second, learning in CCLN is fast since it only updates the weights for the new candidate hidden units added, and the weights of previously added hidden units are fixed after they are installed. One problem with the CCLN is that the final number of hidden units is unbounded due to its dynamically constructive algorithm. Not bounding the number of hidden units in training may lead to overtraining as well (Teng and Wah, 1996). Another drawback stems from the cascade architecture. Each new hidden unit will in effect add a new layer, leading to a very deep structure; the number of connections for the n*th* hidden unit increases as O(*n*). This will give rise to the problem that the generalization performance of the network may be degraded when *n* is large as some of these parameters may be irrelevant to the prediction of the output (Phatak and Koren,1994; Kwok and Yeung, 1997).

3. Strengths and weaknesses of artificial neural networks

The major strength of ANNs in the development of empirical models lies in the theoretic ability to universally model any relationship to any degree of accuracy. This helps eliminate potential error in selecting an appropriate functional form. ANN models are also robust to deviations from traditional statistical assumptions such as normal random errors, common error variance, no multicollinearity of independent variables, and no autocorrelation. ANNs can accommodate a combination of continuous variables and discrete numeric variables. Moreover, most of ANN paradigms, such as BPN, are global models so that a single neural network could be developed to model an entire response surface. Lastly, the parallel architecture offers robustness to data that is incomplete or contains errors. It appears that an ANN modeling technique can satisfy Barton's criteria (1) – (5) and (7) for choosing modeling techniques discussed in Section "modeling in bioprocess".

In the presence of many favorable features, ANNs also exhibit some drawbacks. These include the limitation in their approximation capability by finite and imperfect data sets. The accuracy and precision of an ANN model will depend on the quantity and appropriateness of training data. Too few training data can result in imprecision and inaccuracy in the ANN models, particularly when the model is overtrained or used for extrapolation. Finally, the opponents of ANNs criticizes them as being "Black Boxes" due to the difficulty in explaining exactly why an ANN produces a certain output.

4. Its applications in bioprocess

ANN applications in bioprocess are more widespread in alcoholic fermentation (Cleran et al., 1991; Insa et. al., 1995; Oishi et al., 1992; Vlassides et al., 2001; Honda et al., 1998) and recombinant fermentation (Glassey et al., 1994; Yang, 1992). Only one article has been found to use genetic algorithms coupling with backpropagation neural networks in xylitol production for medium optimization (Fang, et al., 2002). It is observed that all of these works utilize the most popular backpropagation neural network (BPN).

Research Methodology

This research comprises three parts: (1) data collection and preparation; (2) model constructions and validations and (3) comparison of model performance

1. Data collection and preparation

1.1 The data from the continuous production of xylitol by *Candia mogii* ATCC 18364 recycling system, using xylose as a substrate was collected from Tochampa (1998). In his work, two major factors affecting the xylitol production were studied :

(1) Recycle ratio

(2) Aeration rate

Each factor was varied at two levels, leading to 4 experiments as shown in table 1.

Table 1 Experimental Design

Experiment	Recycle Ratio (R)	Aeration rate (vvm)
1	0.50	0.3
2	0.50	1.0
3	0.75	0.3
4	0.75	1.0

In each experiment, the data on biomass (g/l) and xylitol concentration (g/l) were collected from the start of the fermentation for every 2 or 4 hours up to 60 hours, making up 69 data points. The data were arranged in an input-output pattern with recycle ratio, aeration rate and fermentation time as inputs and the cell biomass and xylitol cocnetration as the output. The following table shows all data collected.

Table 2 Data from the Continuous Production of Xylitol by *Candia mogii* ATCC 18364

Time (hr)	Recycle	Aeration	Biomass	Xylitol
0	0.5	1.0	6.93	0.07
4	0.5	1.0	8.42	0.37
8	0.5	1.0	8.77	1.97
12	0.5	1.0	9.52	3.42
16	0.5	1.0	11.84	3.54
20	0.5	1.0	14.8	1.77
24	0.5	1.0	18.61	0.24
28	0.5	1.0	20.20	0.20
32	0.5	1.0	21.37	0.15
36	0.5	1.0	22.92	0.15
40	0.5	1.0	23.85	0.13
44	0.5	1.0	25.09	0.20
48	0.5	1.0	24.86	0.20
52	0.5	1.0	25.13	0.13
56	0.5	1.0	24.54	0.06
0	0.5	0.3	5.01	0.15
4	0.5	0.3	5.12	0.16
8	0.5	0.3	5.10	0.36
12	0.5	0.3	4.86	0.59
16	0.5	0.3	4.74	0.96
20	0.5	0.3	4.58	1.32
24	0.5	0.3	4.43	1.56
28	0.5	0.3	4.44	1.76
32	0.5	0.3	4.29	1.97
36	0.5	0.3	4.21	2.06
40	0.5	0.3	4.00	1.79
44	0.5	0.3	3.93	1.57
48	0.5	0.3	4.06	1.38
52	0.5	0.3	4.06	1.37
56	0.5	0.3	3.53	1.48

Table 2 Data from the Continuous Production of Xylitol by *Candia mogii* ATCC 18364 (continued)

Time (hr)	Recycle	Aeration	Biomass	Xylitol
0	0.75	1.0	6.59	0.08
2	0.75	1.0	10.51	0.68
4	0.75	1.0	7.86	1.29
6	0.75	1.0	9.54	1.57
8	0.75	1.0	11.22	1.60
10	0.75	1.0	14.21	1.36
12	0.75	1.0	17.27	0.95
18	0.75	1.0	22.31	0.39
20	0.75	1.0	22.78	0.27
22	0.75	1.0	23.76	0.21
24	0.75	1.0	24.54	0.20
26	0.75	1.0	25.25	0.18
28	0.75	1.0	26.18	0.18
30	0.75	1.0	26.07	0.18
32	0.75	1.0	27.05	0.20
34	0.75	1.0	27.55	0.19
36	0.75	1.0	28.07	0.23
42	0.75	1.0	28.02	0.52
44	0.75	1.0	28.63	0.43
46	0.75	1.0	29.47	0.25
48	0.75	1.0	29.32	0.25
50	0.75	1.0	29.88	0.21
52	0.75	1.0	30.05	0.18
0	0.75	0.3	6.99	0.05
4	0.75	0.3	0.99	0.27
8	0.75	0.3	8.21	0.58
12	0.75	0.3	7.92	1.16
16	0.75	0.3	8.64	1.85
20	0.75	0.3	7.68	2.36
24	0.75	0.3	7.55	3.21
28	0.75	0.3	7.48	3.70
32	0.75	0.3	7.63	4.06
36	0.75	0.3	7.80	4.54
40	0.75	0.3	7.35	3.78
44	0.75	0.3	7.24	3.86
48	0.75	0.3	7.60	3.96
52	0.75	0.3	7.91	3.59
56	0.75	0.3	8.10	3.18
60	0.75	0.3	8.49	2.50

1.2 All 69 data points were randomly divided into three disjointed data sets:

(1) Training or fitting data set for model construction (41 data points)

(2) Testing data set for selection of proper model parameters (14 data points)

(3) Validating data set for estimation of model generalization capability (14 data points)

2. Model Development

In this research, empirical models were constructed to approximate the relationship between three input factors, i.e., recycle ratio (R), aeration rate (vvm) and fermentation time, and two output variables or responses, i.e., cell biomass (g/l) and xylitol cocnetration (g/l). One model was built for one response as shown in figure 4 and 5. Four types of models were examined, i.e., polynomial regression, dual kriging, backpropagation neural network (BPN) and cascade correlation neural network (CCLN).

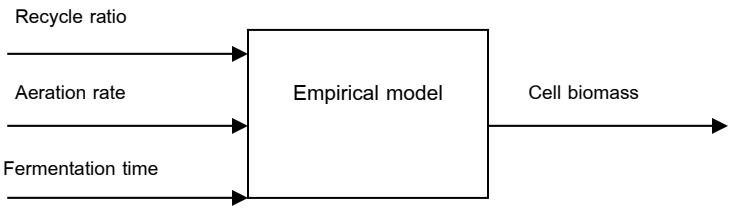

Figure 4 Empirical model for cell biomass

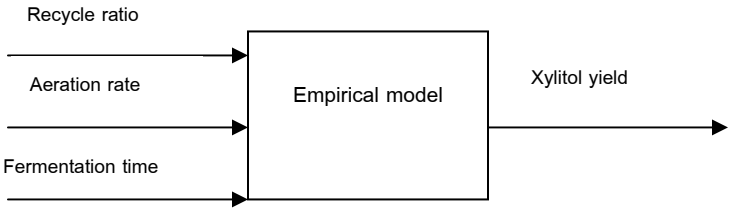

Figure 5 Empirical model for xylitol concentration

In terms of polynomial regression models, a second-order polynomial regression was selected to capture the nonlinearity that exists in most bioprocess. SPSS software version 12 was used to build the second-order stepwise polynomial regression models from the fitting data set. In order to minimize the multicollinearity effect, each independent or input variable was expressed as a deviation around its mean. Both forward and backward stepwise regressions were used with the

probability to enter and remove of 0.05. The aptness of the polynomial regression model was investigated using residual plots and a normal probability plot. The variance inflation factor (VIF) was calculated to examine the presence of muticollinearity. Proper model parameters were selected based on the test of models on the testing data set.

Dual kriging models were built from the fitting datat set using the code developed by Rattle (1998). All variables were normalized between –1 and 1. The order of drift basis function, the covariance function and the Euclidean distance of influence are the key parameters for the dual kriging metamodel. First and second-order drift functions, three types of covariance models: pure nugget effect, linear, and cubic covariance functions and various distances of influence (d) varying between 0.1, 0.2, …, 1.0 were explored to select the proper parameter values via the testing data set.

Both BPN and CCLN models were constructed from the fitting data set using NeuralWare Explorer software. Input neurons were used to represent time, recycle ratio and aeration rate while an output neuron was used to represent the NPV of each alternative. All variables were normalized between –1 and 1 to avoid pathological problems during training of the network. Building a useful ANN model requires proper selection of its architecture and the learning parameters. Various architectures and the learning parameters for both BPN and CCLN were investigated via 8 x 2 x 3 x 2 x 3 factorial design or making up of 288 experiments as follows:

(1) Number of hidden neurons : 1, 2, 3, 4, 5, 6, 7, 8
(2) Learning rule : delta rule and extended rule
(3) Initial learning rate : 0.1, 0.25 and 0.5
(4) Transfer function : sigmoid and TanH transfer function
(5) Random initial weights : change 3 random seeds

Both models utilized the momentum of 0.4 as recommended in default function by Neuralware (1994). The use of momentum stems from the fact that when a very unusual pair of training pattern is learned, it is desirable to use small learning rate to avoid a major disruption in the direction of learning. When the training data are relative similar, it is preferable to train at the fairly rapid rate. Momentum allows net to make reasonable large weight adjustments as long as the adjustments are in the same general direction for several patterns while using a smaller learning rate to prevent a large response to the error from any one training pattern (Fausett, 1994). In order to avoid overtraining which may lead to the problem of degredation in generalization capability or model overfitting, Save Best option in Neuralware software was employed. Each

neural network was trained using the training set for 1,000 iterations and stop to evaluate for their accuracy with the tesing set. Trainng or learning was stopped when the error measure of the testing data set continued to increase. The proper architecture and learning parameters were selected based on the error of this testing data set.

Once the models were built from the fitting data set, their accuracy must be assessed to select the model with the most appropriate parameter values. This is accomplished by using the models to predict the response of a testing data set. The predicted response, together with the actual response, were used to compute the accuracy measures: the root mean square error (RMSE), defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{n}}$$

where y_i denotes the actual response value of data point *i*

$\hat{y}_i$ denotes the predicted response value of data point *i*

n denotes the total number of data points

After the proper model parameters were selected, a validation data set was used to assess the model generalization capability. The extent of model deterioration and overfit (overtrain) were examined by comparing the error measurements from this independent validation data set with the ones computed from the fitting data set. A large increase in the magnitude of error measures indicates overtraining, i.e., memorization of fitting data set with poor generalization capability.

3. Performance Comparison of Various Modeling Techniques

Good estimation model must possess high prediction accuracy as well as be not bias. Polynomial regression, dual krigning, BPN and CCLN models were then compared based on the following criteria :

3.1 Prediction accuracy

This is achieved by comparing the error measurements from the testing data set and independent validation data set with the ones computed from the fitting data set. The lower the

error measure across all data set, the higher the prediction accuracy of the model. In addition, the accuracy ccan be evaluated by a plot of the predicted values against the actual value. A 45° straight line through the origin indicates that the model is highly accuarte.

3.2 Model bias

A good estimator must be unbiased or exhibits as less bias as possible. Bias is a systematic distribution of residuals (predicted output – actual output). A quantitative method to point out the bias of the microbial growth models is to compute a bias factor (B_f) as follows (Jeyamkondan et al., 2001) :

$$B_f = 10^{\sum_{i=1}^{n} \frac{\log(\hat{y}_i / y_i)}{n}}$$

If a bias factor is close or equal to 1, the model is unbiased. A bias factor much greater than 1 indicates that the model overestimates the data while a value much less than 1 indicates that it underestimates the data.

4. Comparison of Biomass Predictive Models and Xylitol Concentration Models

To compare modeling performance of varius responses, the mean absolute percentage error (MAPE) is generally used.

$$MAPE = \frac{\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100}{n}$$

Results and Discussions

1. Modeling the cell biomass

1.1 Model parameters

All empirical models were constructed with various structures and / or model parameters to approximate the relationship between input variables (fermentation time, recycle ratio, and aeration rate) and a response i.e. cell biomass. Table 3 summarizes the final selection of model parameters from continuous xylitol production data.

Table 3 The model parameters for cell biomass

Model Type	Final Structure or Parameters	RMSE (g/l)	
		Fitting Set	Testing Set
Polynomial regression	Second-order stepwise regression	1.50	1.87
Dual kriging	First-order drift function with linear covariance function and distances of influence of 0.9	0.00	1.07
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.78	0.73
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	1.04	0.99

The final polynomial regression models obtained from a stepwise procedure is as follows:

$$y_1 = 0.07331 + 0.48210x_2x_3 + 11.31331x_1x_2 - 0.00467x_3^2 + 0.26346x_1x_3$$

where y_1 = cell biomass (g/l)

x_1 = recycle ratio (R)

x_2 = aeration rate (vvm)

x_3 = fermentation time (hours)

It is observed that the relationship is nonlinear with only fermentation time included as a quadratic term in the model. The interaction between recycle ratio and aeration rate, between recycle ratio and fermentation time, and between recycle ratio and fermentation time, and between aeration rate and fermentation time exist. The coefficient of determination (R^2) for the fitted model is 0.9743. R-square is a measure of a proportion of total variation in response y_1 that is explained by a set of input variables x's or interactions among them. R-square close to 1 indicates that most of variability in response y_1 is explained by the regression model. It appears that this model fits the fitting data set quite well.

Since regression models are always constructed based on rigid statistical assumptions, the reliability of these models will definitely depend on the validity of these assumptions as well. The results from the plot of residuals against predicted values, the normal probability plot of residuals and the calculation of Variance Inflation Factor (VIF) points out that the model errors are normally and independently distributed with constant variance and that multicollinearity does not significantly present. As a result, the model aptness is verified and one can rely on its subsequent use.

As theoretical covariance functions are employed for dual kriging model in this research, the dual krigning models then become exact interpolators as observed in the error measure of 0 in the fitting data set. The best dual kriging model identified uses first-order drift function and exhibit higher prediction accuracy than second-order regression model. Dual kriging requires no rigid assumption except the unbiased estimator which would then be tested later.

There is no underlying statistical assumption for BPN and CCLN models. As a consequent, they are ready for validation. It is observed for both BPN and CCLN that hyperbolic tangent transfer function and a small initial learning rate of 0.1 are proper parameters. Under these sets of parameters, the models yield the least RMSE in the testing data set. As recommended by Neuralware (1994), the hyperbolic tangent transfer function works well for a real world data as is this case. The learning rate is generally set between 0 and 1. Too small learning rate yields slow learning whereas too large leaning rate may cause a large error reduction and a major disruption of the direction of learning and thus the net might get stuck in local minimum rather than achieving global minimum of error. The best BPN model requires 7 hidden neurons while the best CCLN has 5 neurons. However, the prediction accuracy for these two construction data sets of BPN is better than that of CCLN.

1.2 Model validation

The best models selected from section 1.1 were validated using the validation data set. The results are summarized in table 4.

Table 4 Validations results of various empirical models for cell biomass

Model Type	Model Parameters	RMSE (g/l)	
		Fitting Set	Validation Set
Regression	2^{nd} order polynomial	1.50	2.03
Dual kriging	1^{st} order drift function with linear covariance function and distances of influence of 0.9	0.00	1.93
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.78	1.59
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	1.04	1.13

It is seen that the CCLN model outperforms the regression, dual kriging and BPN models in terms of prediction accuracy in the validation data set, i.e. the set that has not been used for model constructions. In other words, the CCLN model exhibits better generalization capability than the others. Meanwhile, the second order regression model shows the worst generalization capability. Apart from the dual kriging model which is an exact interpolator, the prediction accuracy of the BPN model is highest in the fitting data set but this performance deteriorates in the validation data set. The BPN is quite sensitive to the overparametization and overtraining, leading to a loss in generalization property. Overparametization refers to too many free parameters or connecting weights which arise from too many hidden neurons and might subsequently leads to model overfitting, i.e. performing well in the fitting data but not the validation data. Unlike the BPN, the CCLN starts off with no hidden neurons and it will automatically find the size and the topology of the resulting ANN. The problem of overspecifying the number of hidden units and thus overfitting could therefore be alleviated. The remedies, however, for overparametization and overfitting is by reducing the network size such as pruning the trained connections with small weights.

1.3 Comparison of Model Performances

Four types of model were compared based on prediction accuracy and bias as described below.

1.3.1 Prediction accuracy

Prediction accuracy is one of the criteria for choosing an empirical modeling technique. Prediction accuracy in terms of RMSE for models developed in section 1.2 is depicted in table 5.

Table 5 Prediction accuracy of various models for cell biomass

Model Type	Model Parameters	RMSE (g/l)		
		Fitting Set	Testing Set	Validation Set
Regression	2^{nd} order polynomial	1.50	1.87	2.03
Dual kriging	1^{st} order drift function with linear covariance function and distances of influence of 0.9	0.00	1.07	1.93
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.78	0.73	1.59
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	1.04	0.99	1.13

It is observed that the ANN models exhibit higher prediction accuracy than the second-order regression and dual kriging models across nearly all data sets. The BPN outperforms the CCLN only in the fitting data set and testing data set but not the validation data set which was used to assess the generalization capability. However, the choice of an empirical model depends on the decision problem. Kleijnen and Sargent (2000) assert that a high accuracy model is critical for prediction while a crude model may suffice for understanding the behavior of the system of interest. Since this is a prediction problem, the CCLN model might be the best choice in terms of its accuracy.

Figure 6-9 shows scatter plots between actual value and predicted value of cell biomass for the fitting data set from regression, dual kriging, BPN and CCLN models with the R^2 of 0.97, 1.00, 0.99, and 0.985, repectively. It is quite obvious that dual kriging and BPN have a better fit to the

fitting data set or more accurate than the CCLN and regression models. However, Figure 10-13 illustarte scatter plots between actual value and predicted value of cell biomass for the validating data set from regression, dual kriging, BPN and CCLN models with the R^2 of 0.947, 0.95, 0.97, and 0.98, repectively. These results confirm that all model's accuracy drop from the fitting data set to the validating one. Overall, the CCLN's accuracy deteriorates the least compared to the others.

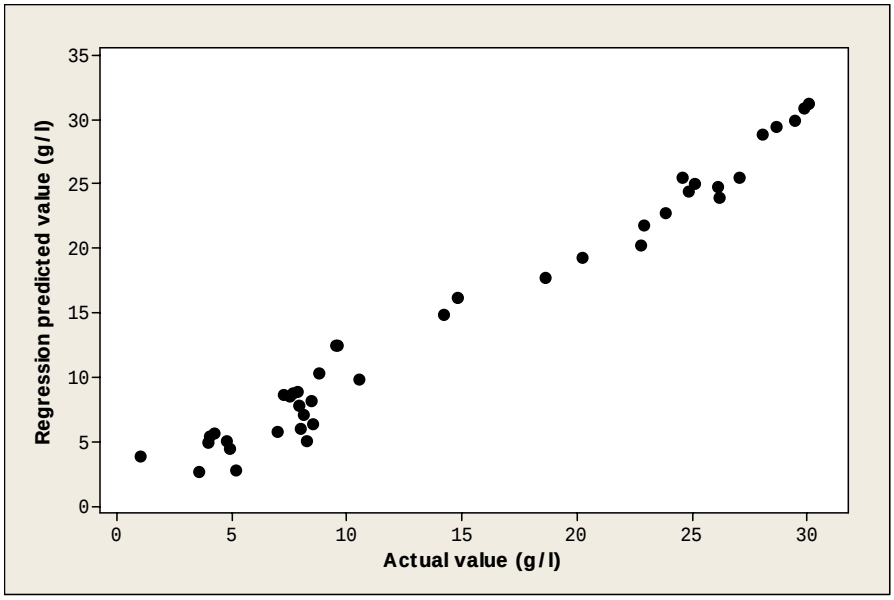

Figure 6 A scatter plot between actual value and predicted value of cell biomass from regression model on the fitting data set (R^2 = 0.97)

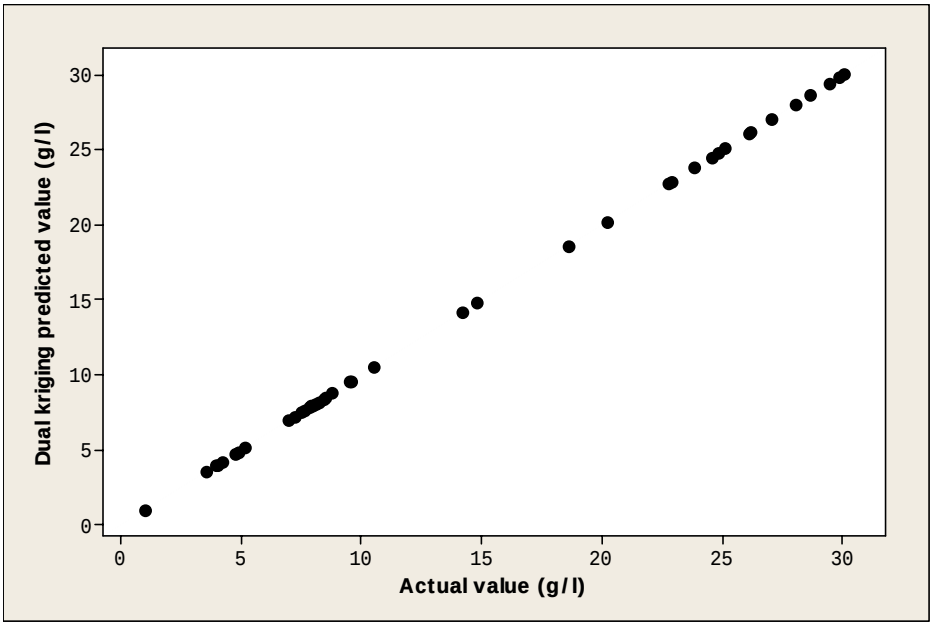

Figure 7 A scatter plot between actual value and predicted value of cell biomass from dual kriging model on the fitting data set (R^2 = 1.00)

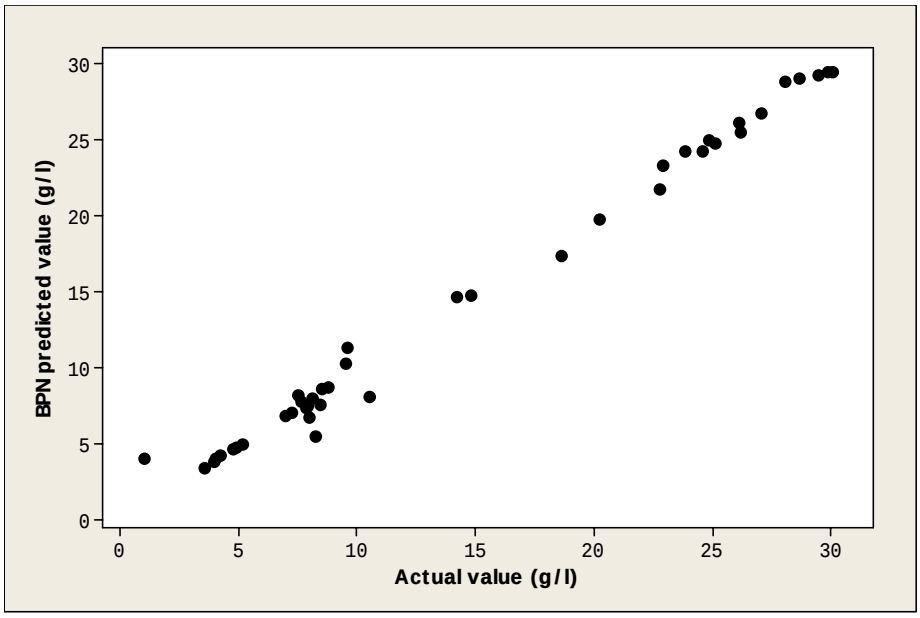

Figure 8 A scatter plot between actual value and predicted value of cell biomass from BPN model on the fitting data set (R^2 = 0.99)

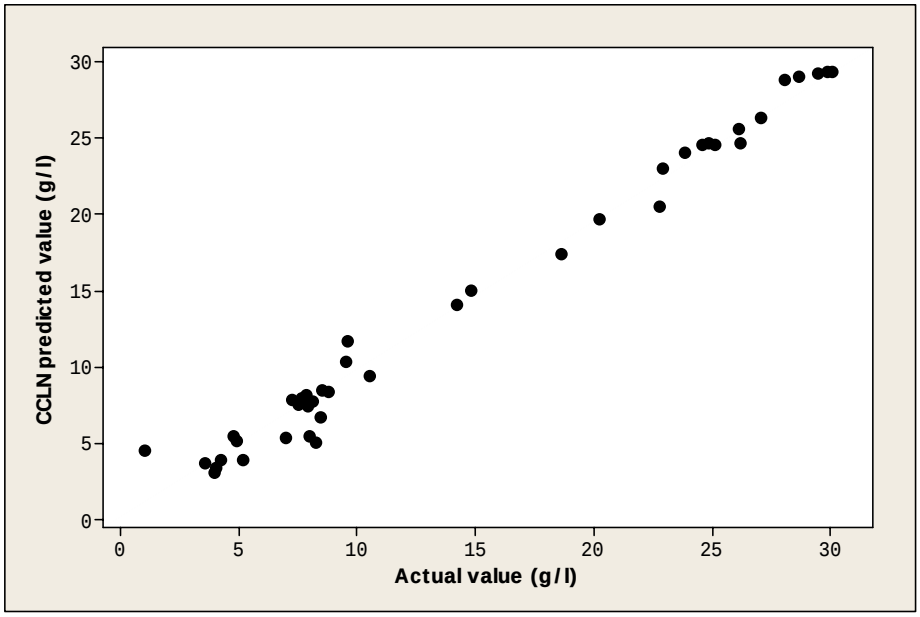

Figure 9 A scatter plot between actual value and predicted value of cell biomass from CCLN model on the fitting data set (R^2 = 0.985)

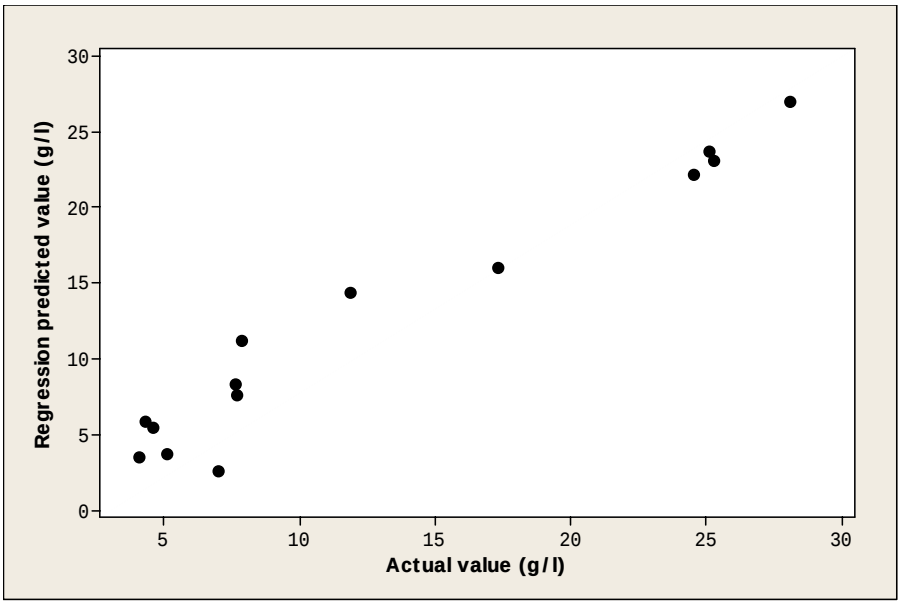

Figure 10 A scatter plot between actual value and predicted value of cell biomass from regression model on the validating data set (R^2 = 0.947)

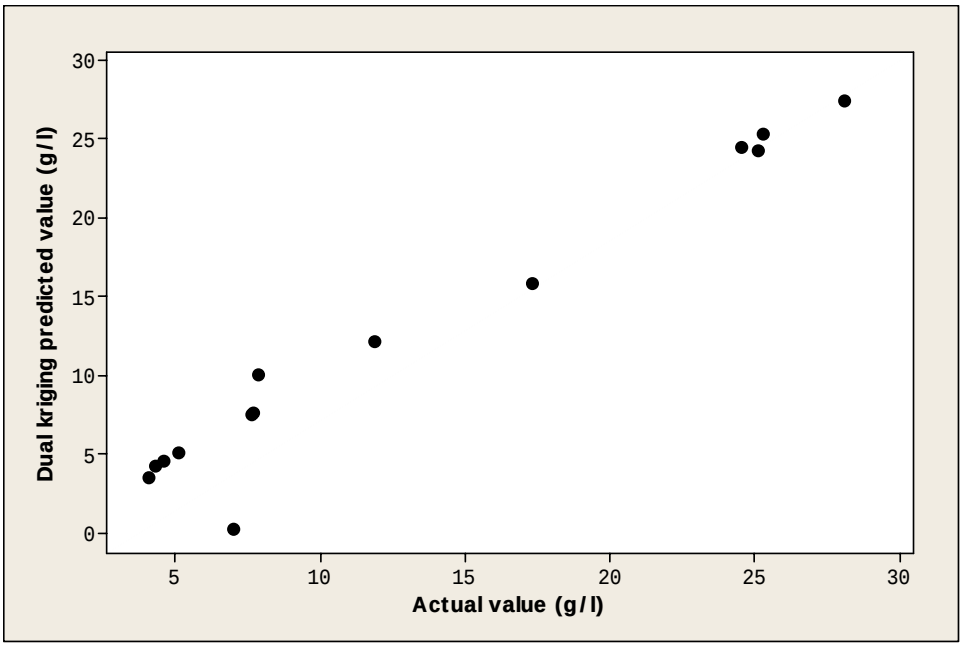

Figure 11 A scatter plot between actual value and predicted value of cell biomass from dual krigig model on the validating data set (R^2 = 0.95)

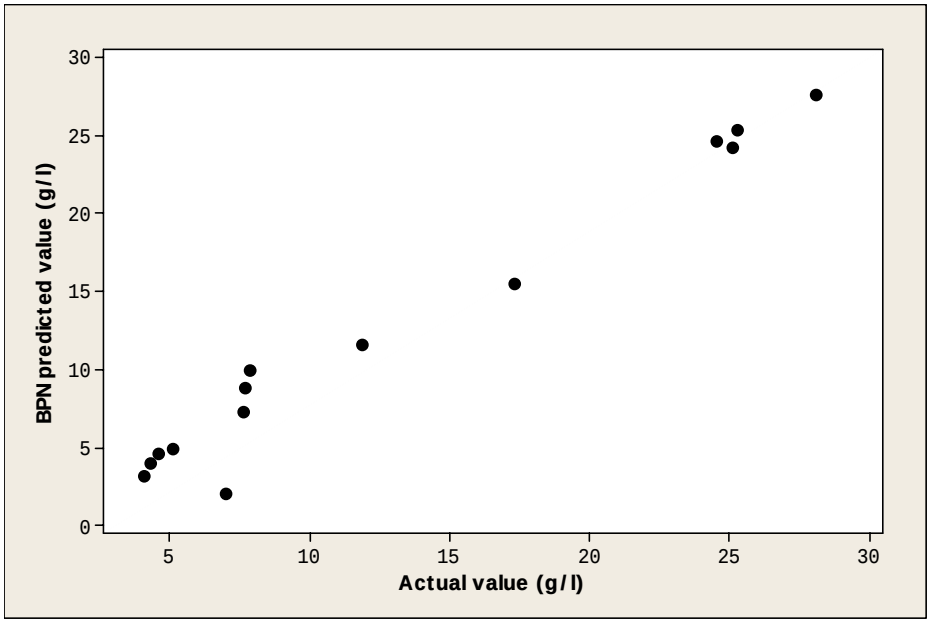

Figure 12 A scatter plot between actual value and predicted value of cell biomass from BPN model on the validating data set (R^2 = 0.97)

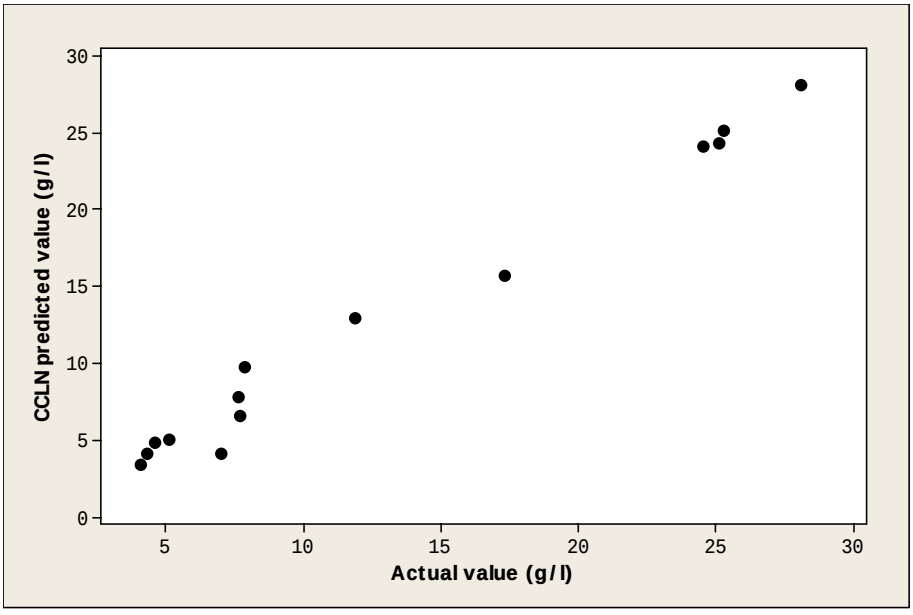

Figure 13 A scatter plot between actual value and predicted value of cell biomass from CCLN model on the validating data set (R^2 = 0.98)

1.3.2 Model bias

Table 6 shows bias factor values for both fitting and validating data sets in prediction of cell biomass. It is observed that all models are not biased in the fitting data set. In fact both regression and dual kriging are constructed to be unbiased in nature. Dual kriging is perfectly unbiased with a bias factor value of 1.00 since it is an exact interpolator. For the validating data set, it is revealed that all four models slightly underestimated the data. A plot between actual and predicted observations could also be used to examine the bias type of any models. Figure 14-17 display this plot of regression, dual kriging, BPN and CCLN models for fitting data set while Figure 18-21 exhibit the same plots for validating data set. It is apparent that the predicted values of all models are pretty much on the actual values indicating that they are unbiased. These results in the fitting data are congruence with the bias factors. On the other hand, it is revealed in the validating data set that all models are little biased downwards, i.e., they slightly underestimate the data (the lower predicted value compared to the actual value) especially the ones with low cell biomass values which often occurred at the start of the xylitol production process. In general, neural network models are known to be biased in nature. Their bias is often take the form of undershoot, i.e. where the network model does not reach the upper and lower extreme of the actual or target data, especially for the network with sigmoid transfer function (Twomey and Smith, 1996). This can be remedied by training the network on the expanded normalization range.

Table 6 Bias Factor of various models for cell biomass

Model Type	Model Parameters	Bias Factor (B_f)	
		Fitting Set	Validation Set
Regression	2^{nd} order polynomial	1.02	0.95
Dual kriging	1^{st} order drift function with linear covariance function and distances of influence of 0.9	1.00	0.80
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	1.02	0.91
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	0.99	0.95

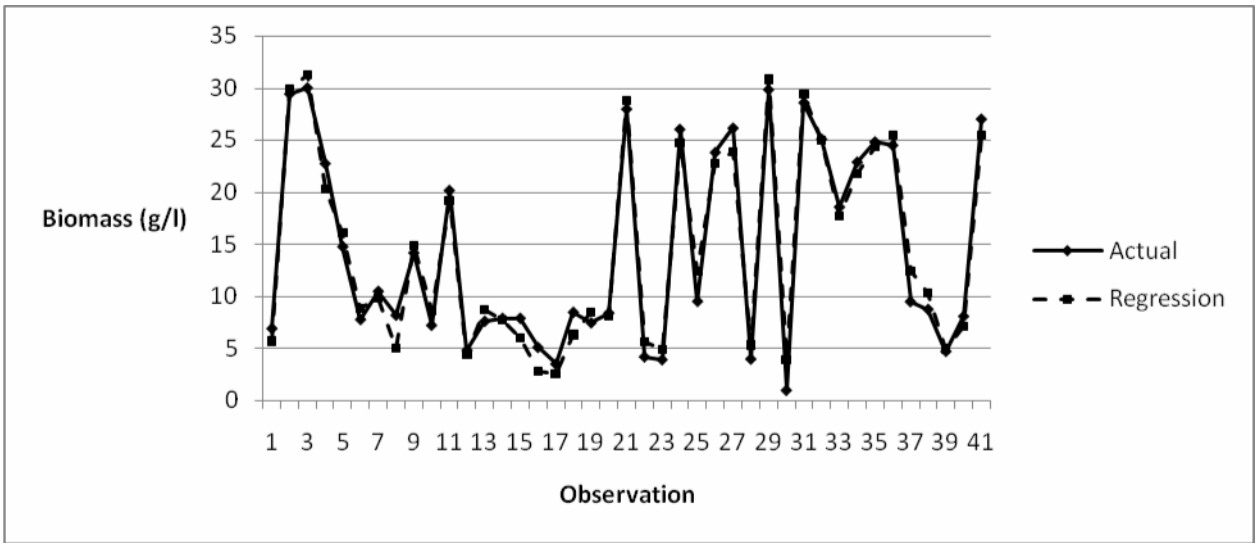

Figure 14 A plot between actual and predicted values of cell biomass from regression model on the fitting data set

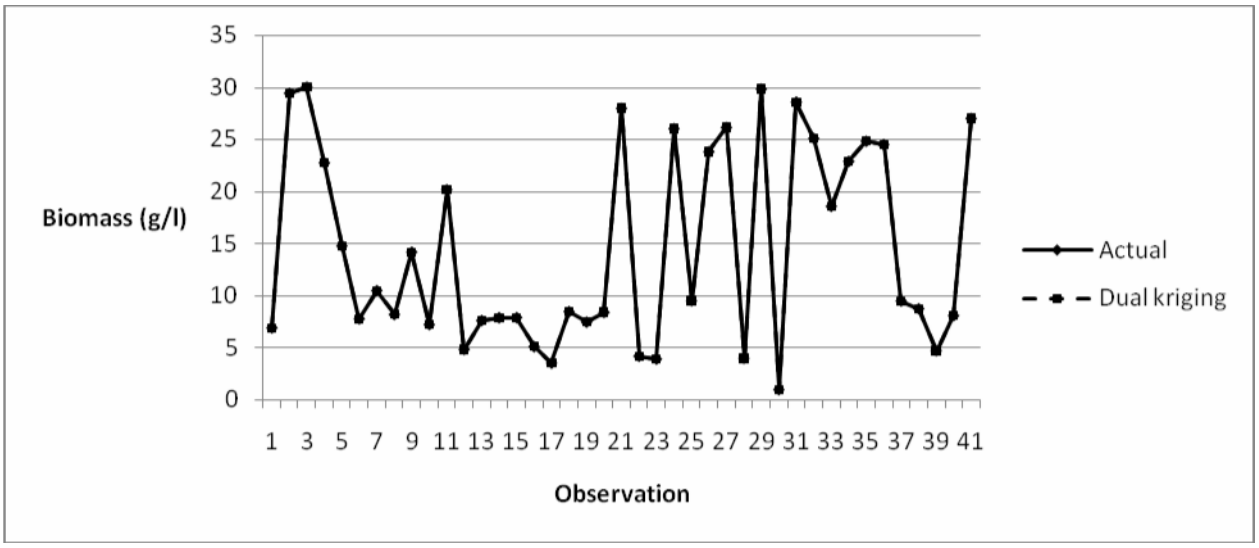

Figure 15 A plot between actual and predicted values of cell biomass from dual kriging model on the fitting data set

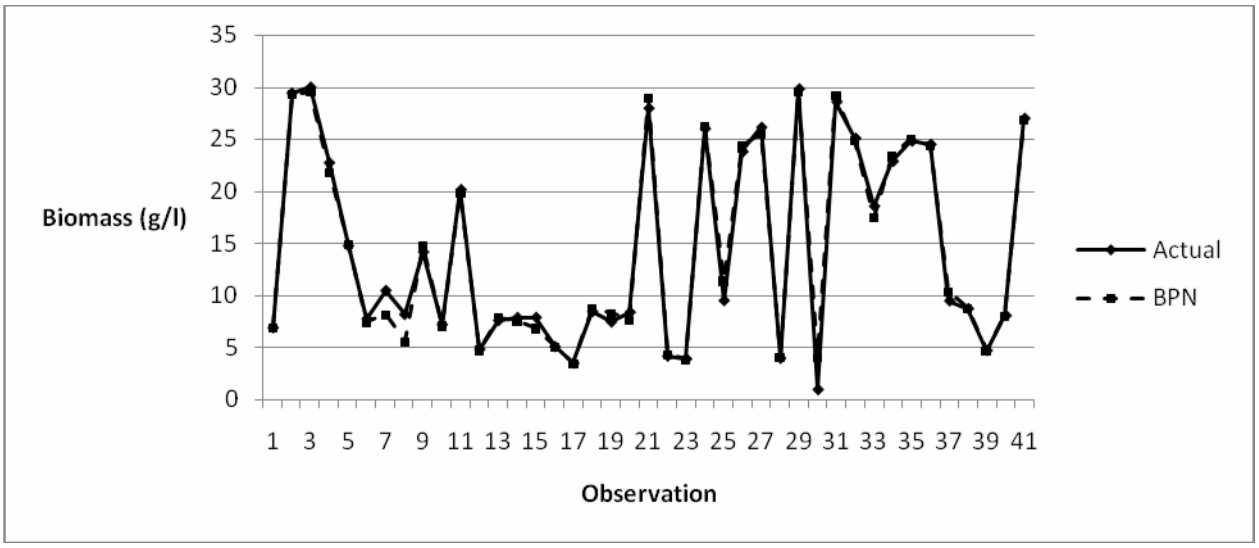

Figure 16 A plot between actual and predicted values of cell biomass from BPN model on the fitting data set

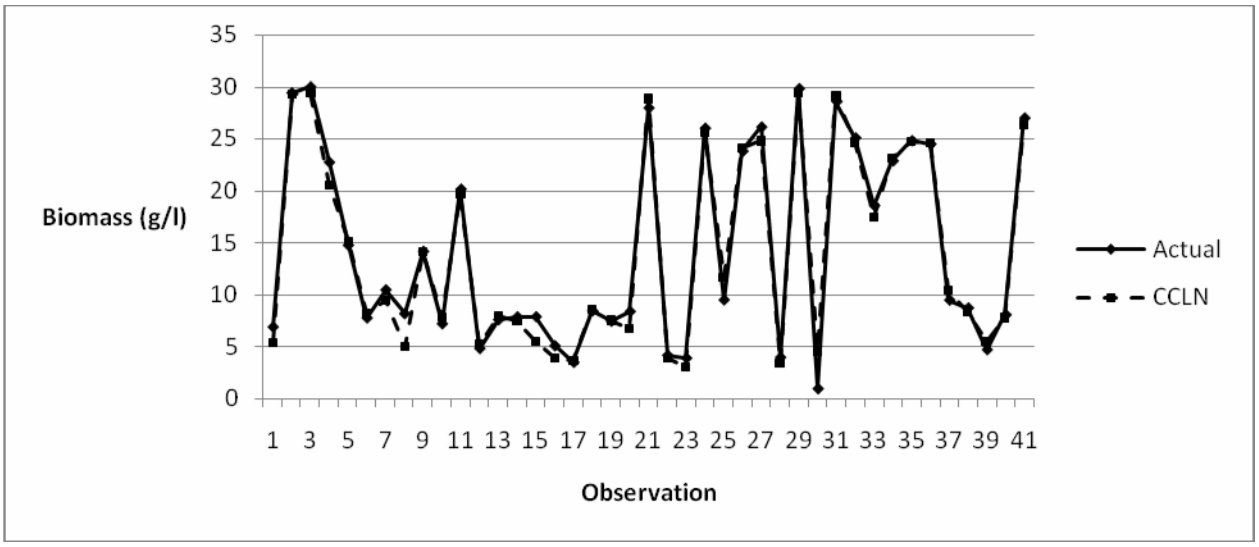

Figure 17 A plot between actual and predicted values of cell biomass from CCLN model on the fitting data set

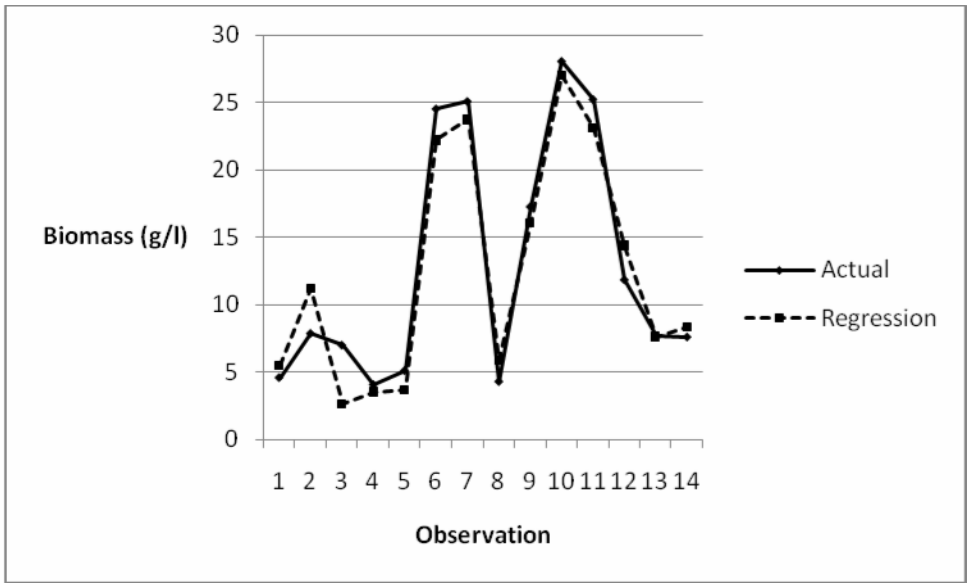

Figure 18 A plot between actual and predicted values of cell biomass from regression model on the validating data set

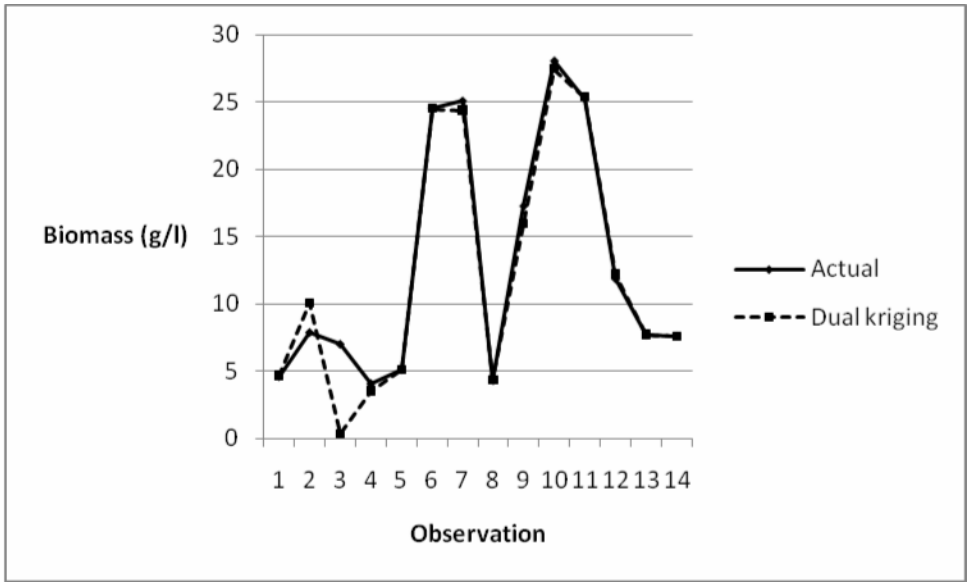

Figure 19 A plot between actual and predicted values of cell biomass from dual kriging model on the validating data set

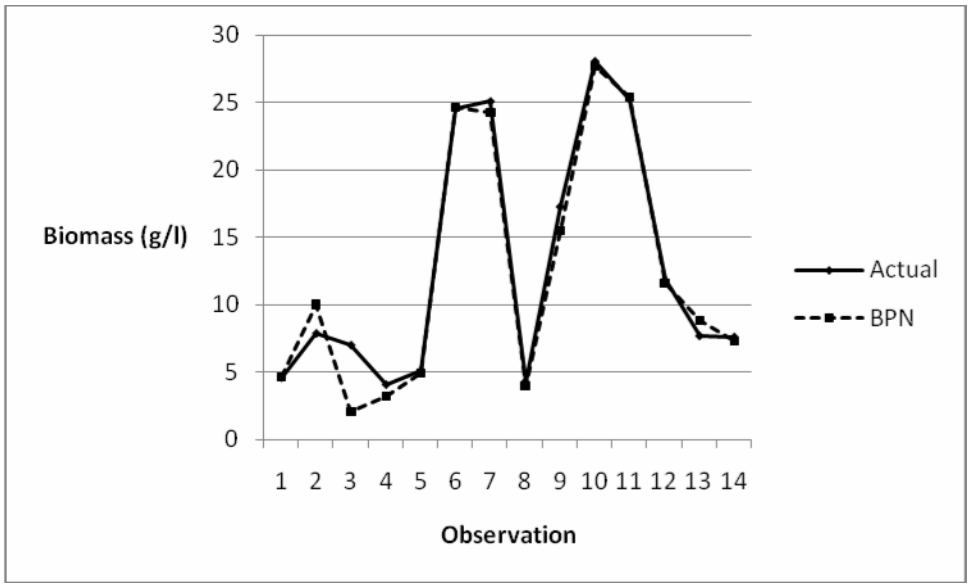

Figure 20 A plot between actual and predicted values of cell biomass from BPN model on the validating data set

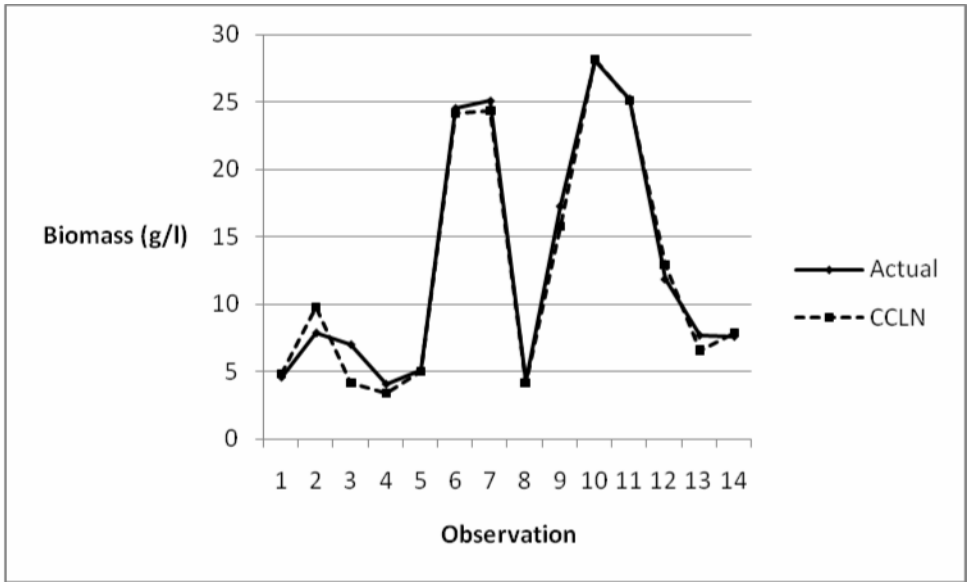

Figure 21 A plot between actual and predicted values of cell biomass from CCLN model on the validating data set

2. Modeling the xylitol concentration

2.1 Model parameters

All empirical models were again constructed with various model parameters to approximate the relationship between input variables (fermentation time, recycle ratio, and aeration rate) and xylitol concentration as a response. Table 7 summarizes the final selection of model parameters from continuous xylitol production data.

Table 7 The model parameters for xylitol cocnetration

Model Type	Final Structure or Parameters	RMSE (g/l)	
		Fitting Set	Testing Set
Polynomial regression	First-order stepwise regression	0.92	0.77
Dual kriging	Second-order drift function with linear covariance function and distances of influence of 0.9	0.00	0.23
BPN	1-layer with 8 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.12	0.11
CCLN	6 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.60	0.33

The final polynomial regression models obtained from a stepwise procedure is as follows:

$$y_2 = 0.90080 + 0.09276x_1x_3 - 0.06831x_2x_3$$

where y_2 = xylitol concentration (g/l)

x_1 = recycle ratio (R)

x_2 = aeration rate (vvm)

x_3 = fermentation time (hours)

It is apparent that the relationship is still nonlinear though there is no quadratic term in the model. All input variables in the model are in the forms of interaction terms. An interaction between recycle ratio and fermentation time indicates a change in the effect of recycle ratio as a

function of the value of fermentation time. Similarly, an interaction between aeration rate and fermentation time indicates a change in the effect of aeration rate as a function of the value of fermentation time. The Coefficient of Determination (R^2) for the fitted regression model is 0.5331. That is the variation in response y_2 is not quite well explained by a set of interactions among input variables included in the model. Adding more input variables in various powers or interaction terms may help increase the R-square. However, high R-Square does not imply that the model will be useful, i.e., with respect to general prediction accuracy (Neter *et al*., 1990). The results from the plot of residuals against predicted values, the normal probability plot of residuals and the calculation of Variance Inflation Factor (VIF) show that all underlying assumptions are valid. That is, the model errors are normally and independently distributed with constant variance and the multicollinearity does not exist. Consequently, the model is quite reliable for subsequent use.

Similar to the results for cell biomass prediction, the dual kriging model selected for xylitol prediction is also an exact interpolator with the error measure of 0 in the fitting data set. However, this dual kriging model employs second-order drift function and still exhibit higher prediction accuracy than the regression model across 2 data sets.

For ANN models, both BPN and CCLN models, constructed with hyperbolic tangent transfer function, extended delta rule and a small initial learning rate of 0.1, yield the smallest RMSE in the testing data set. The best BPN model requires more hidden neurons that the best CCLN model. Nevertheless, the prediction accuracy for these two data sets of BPN is much better than that of CCLN. It is quite remarkable that the RMSEs of the testing data set for regression, BPN and CCLN models are smaller than those of the fitting data set. It appears that the BPN model outperforms the others in terms of prediction accuracy.

2.2 Model validation

The best models selected from section 2.1 were validated using the validation data set. The results are summarized in table 8.

Table 8 Validations results of various empirical models for xylitol concentration

Model Type	Model Parameters	RMSE (g/l)	
		Fitting Set	Validation Set
Regression	First-order stepwise regression	0.92	0.78
Dual kriging	Second-order drift function with linear covariance function and distances of influence of 0.9	0.00	0.23
BPN	1-layer with 8 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.12	0.16
CCLN	6 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.60	0.52

Both polynomial regression and CCLN models exhibit pretty good generalization capability. Their prediction accuracy does not deteriorate at all from the fitting data set to validation data set. However, though the accuracy of the dual kriging and BPN models decline slightly, both appears to be more accurate than the regression and CCLN models. On a whole, the BPN model outperforms the other models with repect to its prediction accuracy.

2.3 Comparison of Model Performances

Four types of model were compared based on prediction accuracy and bias as described below.

2.3.1 Prediction accuracy

Table 9 compares the prediction accuracy in terms of RMSE for models developed in section 2.2. It is apparent that the BPN model possesses considerably higher prediction accuracy than the dual kriging, CCLN and polynomial regression models.

Table 9 Prediction accuracy of various models for xylitol concentration

Model Type	Model Parameters	RMSE (g/l)		
		Fitting Set	Testing Set	Validation Set
Regression	2^{nd} order polynomial	0.92	0.77	0.78
Dual kriging	1^{st} order drift function with linear covariance function and distances of influence of 0.9	0.00	0.23	0.23
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.12	0.11	0.16
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	0.60	0.33	0.52

Figure 22-25 shows scatter plots between actual value and predicted value of xylitol concentration for the fitting data set from regression, dual kriging, BPN and CCLN models with the R^2 of 0.52, 1.00, 0.99, and 0.71, repectively. Similar to biomass prediction, the dual kriging and BPN models have a better fit to the fitting data set or are more accurate than the CCLN and regression models. Figure 26-29 display scatter plots between actual value and predicted value of xylitol concentration for the validating data set from regression, dual kriging, BPN and CCLN models with the R^2 of 0.58, 0.97, 0.99, and 0.82, repectively. These results reveal that dual kriging's and BPN's accuracy drop somewhat from the fitting data set to the validating one whereas those from the regression's and CCLN's stay put. However, the overall dual krging's and BPN's prediction accuracy are higher.

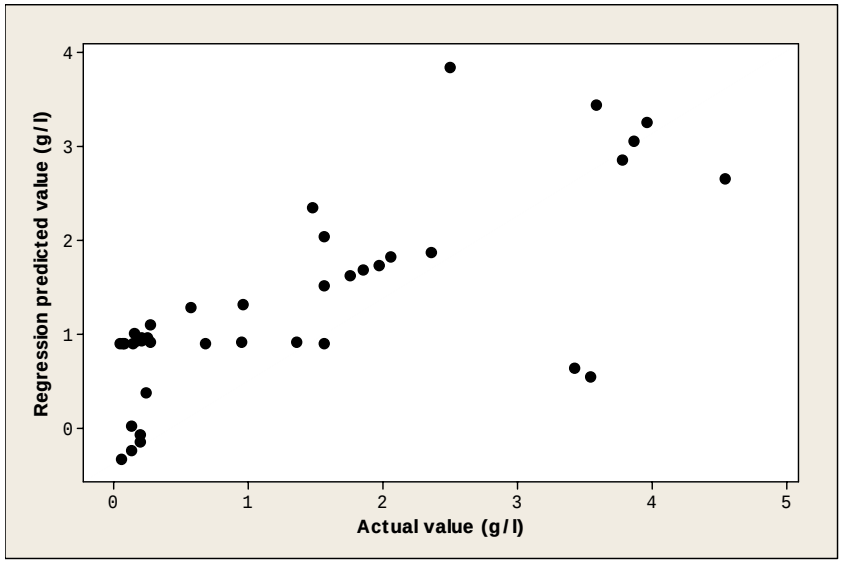

Figure 22 A scatter plot between actual value and predicted value of xylitol concentration from regression model on the fitting data set (R^2 = 0.52)

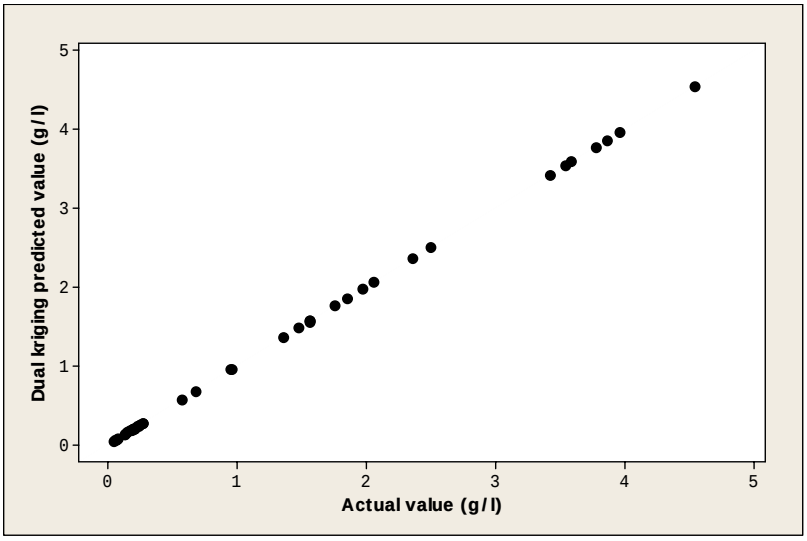

Figure 23 A scatter plot between actual value and predicted value of xylitol concentration from dual kriging model on the fitting data set (R^2 = 1.00)

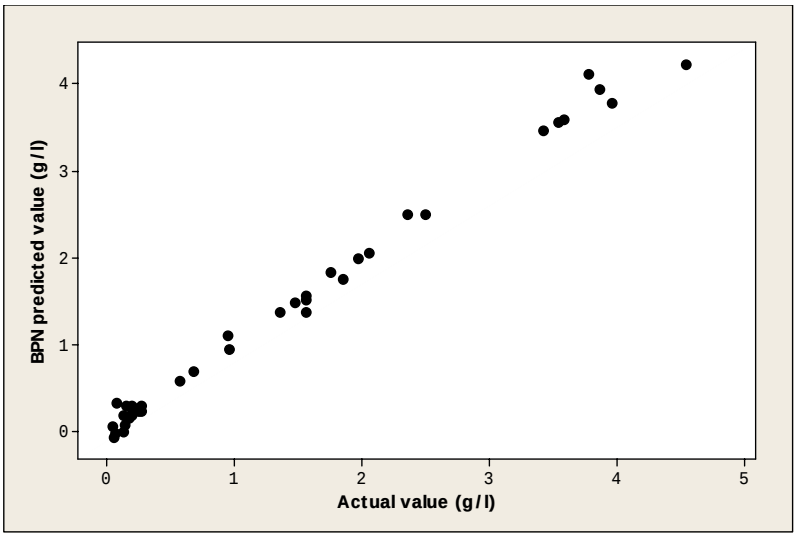

Figure 24 A scatter plot between actual value and predicted value of xylitol concentration from BPN model on the fitting data set (R^2 = 0.99)

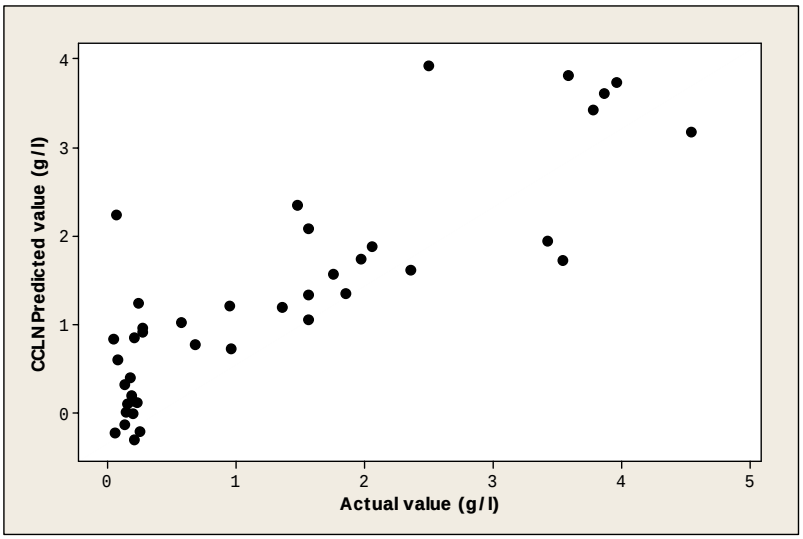

Figure 25 A scatter plot between actual value and predicted value of xylitol concentration from CCLN model on the fitting data set (R^2 = 0.71)

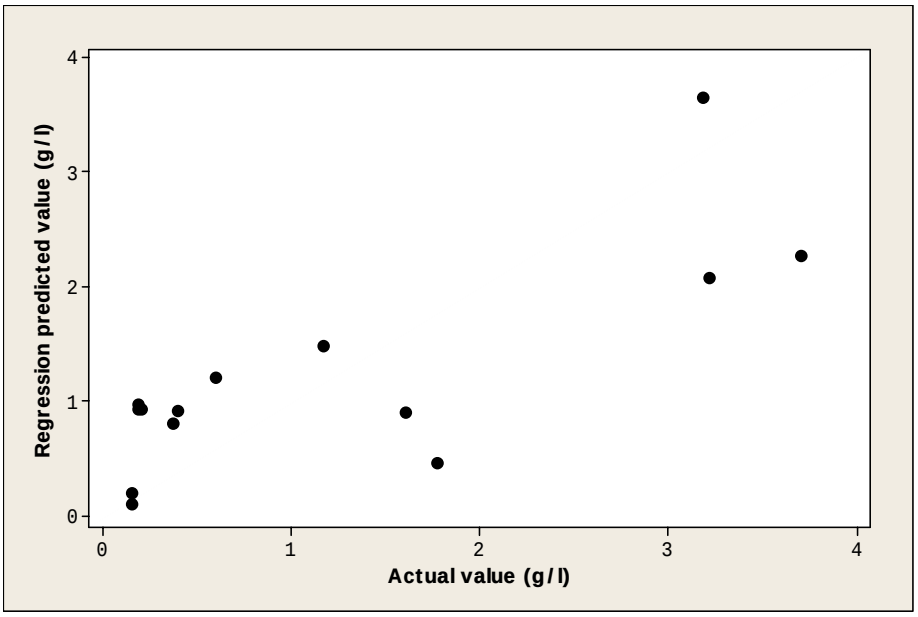

Figure 26 A scatter plot between actual value and predicted value of xylitol concentration from regression model on the validating data set (R^2=0.58)

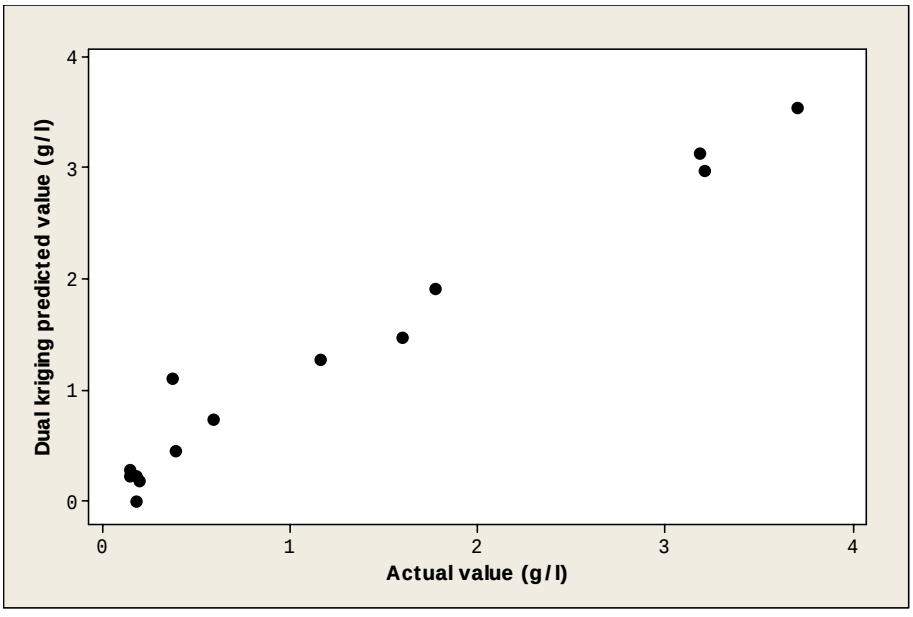

Figure 27 A scatter plot between actual value and predicted value of xylitol concentration from dual kriging model on the validating data set (R^2=0.97)

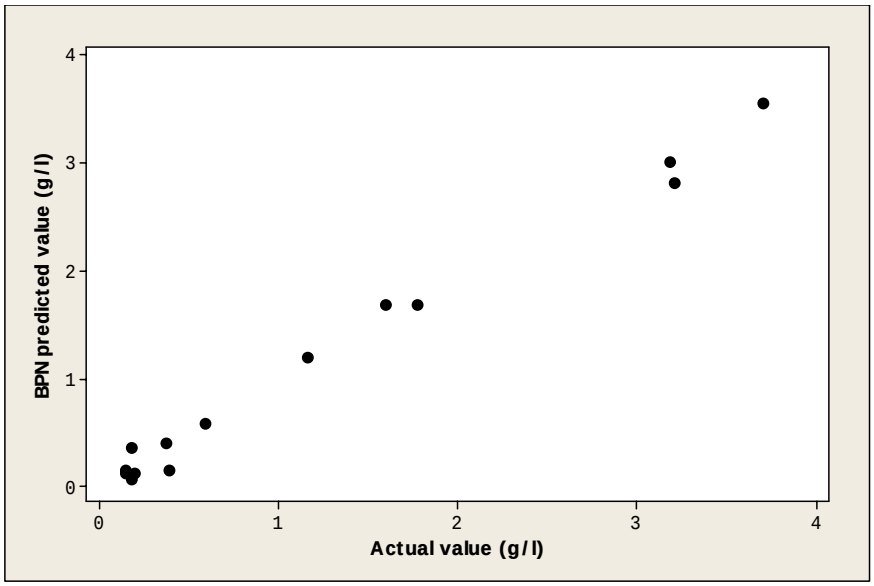

Figure 28 A scatter plot between actual value and predicted value of xylitol concentration from BPN model on the validating data set (R^2 =0.99)

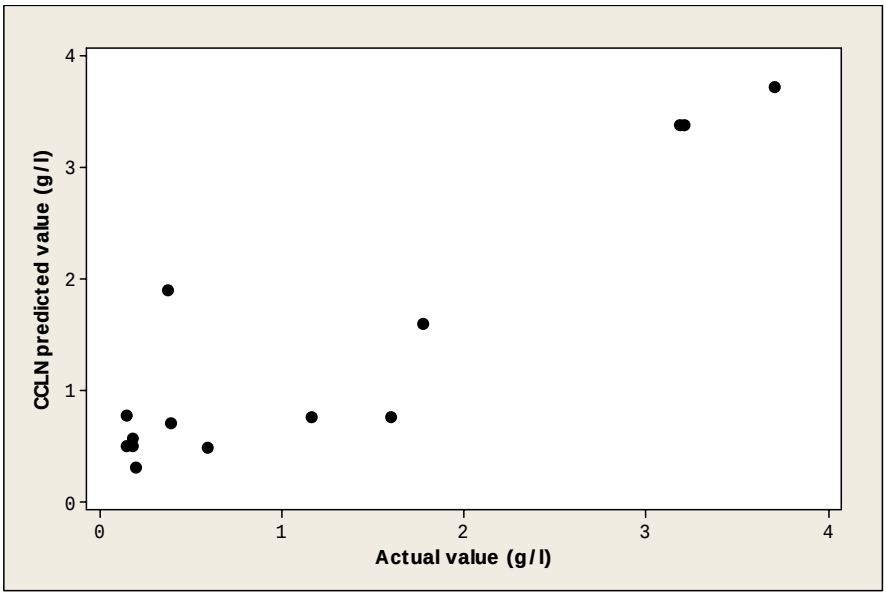

Figure 29 A scatter plot between actual value and predicted value of xylitol concentration from CCIN model on the validating data set (R^2 =0.82)

2.3.2 Model bias

Table 10 illustrates bias factor values for both fitting and validating data sets in prediction of xylitol concentration. It is observed that regression, dual kriging and BPN models are not biased in the fitting data set whereas the CCLN models are biased upwards. For the validating data set, dual kriging is slightly biased upwards (overestimates the data) and regression and CCLN models relatively overestimated the data while the BPN model slightly underestimates the data. A plot between actual and predicted observations could also be used to examine the bias type of any models. Figure 30-33 display plots between actual and predicted observations of regression, dual kriging, BPN and CCLN models, respectively for fitting data set while Figure 34-37 exhibit the same plots for validating data set. These results are consistent with the bias factor for both data sets. It is furher discerened that the bias upwards in the validating data set of the regression and CCLN models often occurs with the lower levels of xylitol concentration or at the start of the process as seen in the predciton of cell biomass.

Table 10 Bias Factor of various models for xylitol concentration

Model Type	Model Parameters	Bias Factor (B_f)	
		Fitting Set	Validation Set
Regression	2^{nd} order polynomial	1.05	1.41
Dual kriging	1^{st} order drift function with linear covariance function and distances of influence of 0.9	1.00	1.19
BPN	1-layer with 7 hidden neurons, TanH transfer function, extended delta rule, initial learning rate of 0.1, momentum of 0.4	0.98	0.87
CCLN	5 hidden neurons, TanH transfer function, delta rule, initial learning rate of 0.1, momentum of 0.4	1.29	1.58

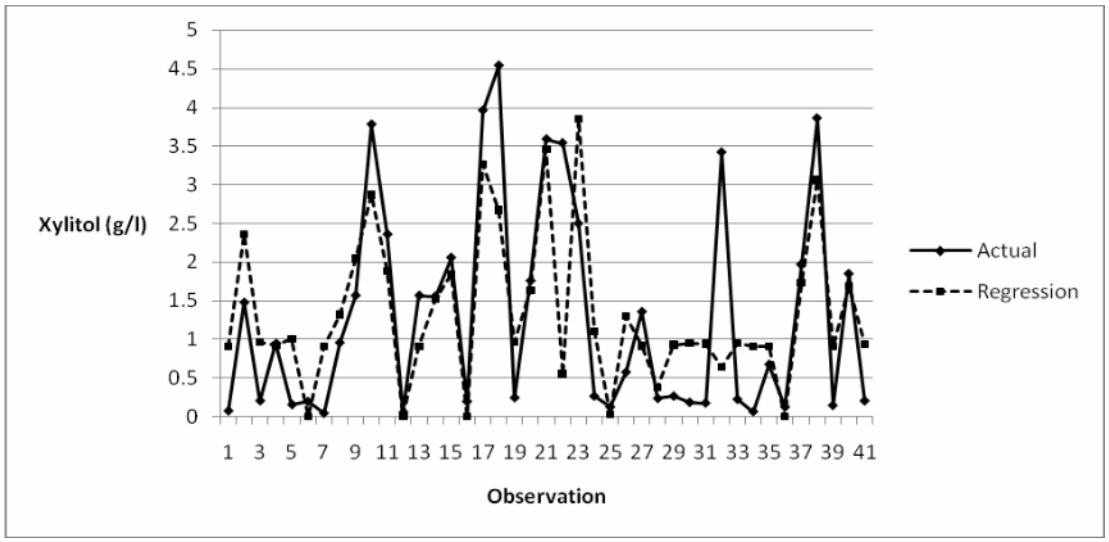

Figure 30 A plot between actual and predicted values of xylitol concentration from regression model on the fitting data set

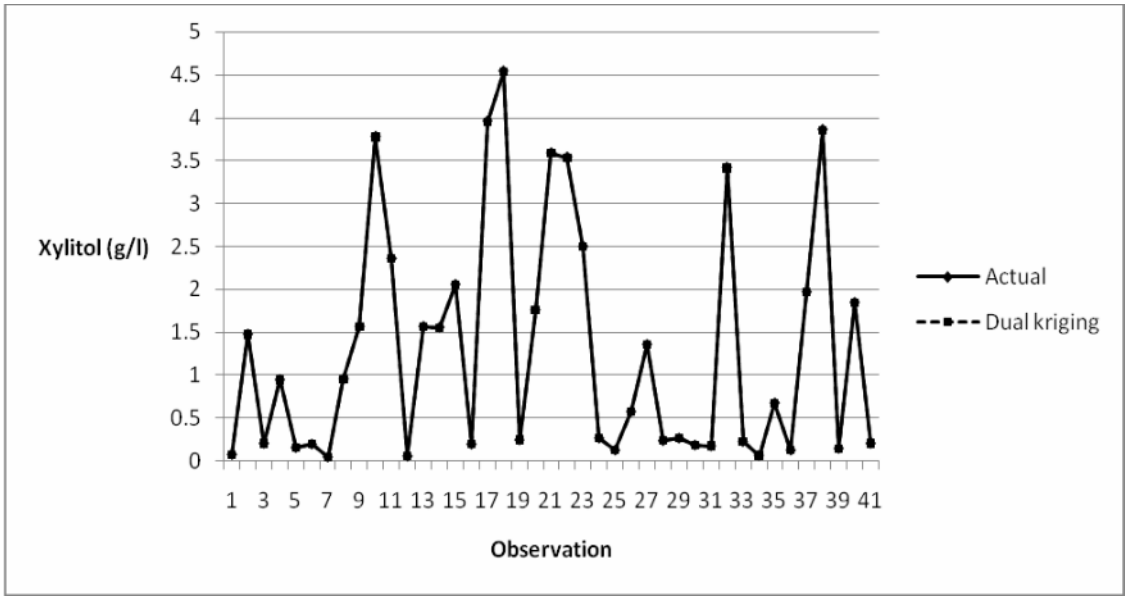

Figure 31 A plot between actual and predicted values of xylitol concentration from dual kriging model on the fitting data set

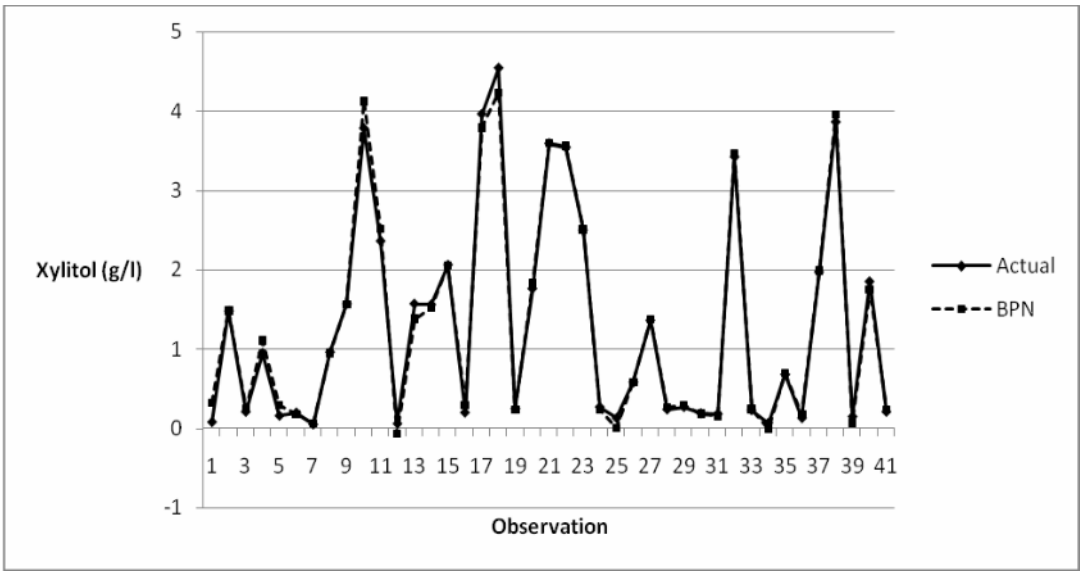

Figure 32 A plot between actual and predicted values of xylitol concentration from BPN model on the fitting data set

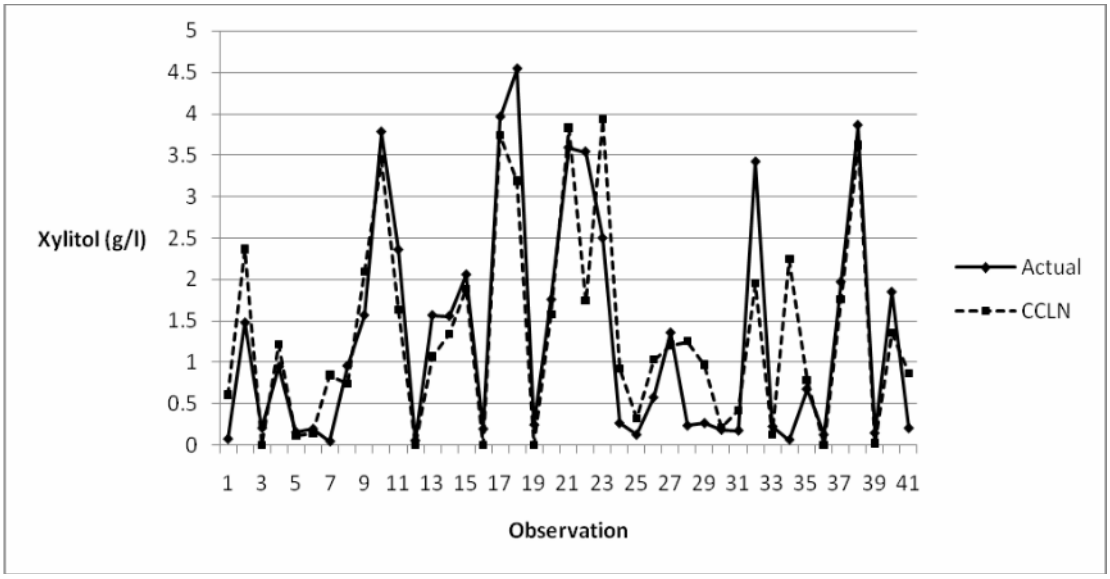

Figure 33 A plot between actual and predicted values of xylitol concentration from CCLN model on the fitting data set

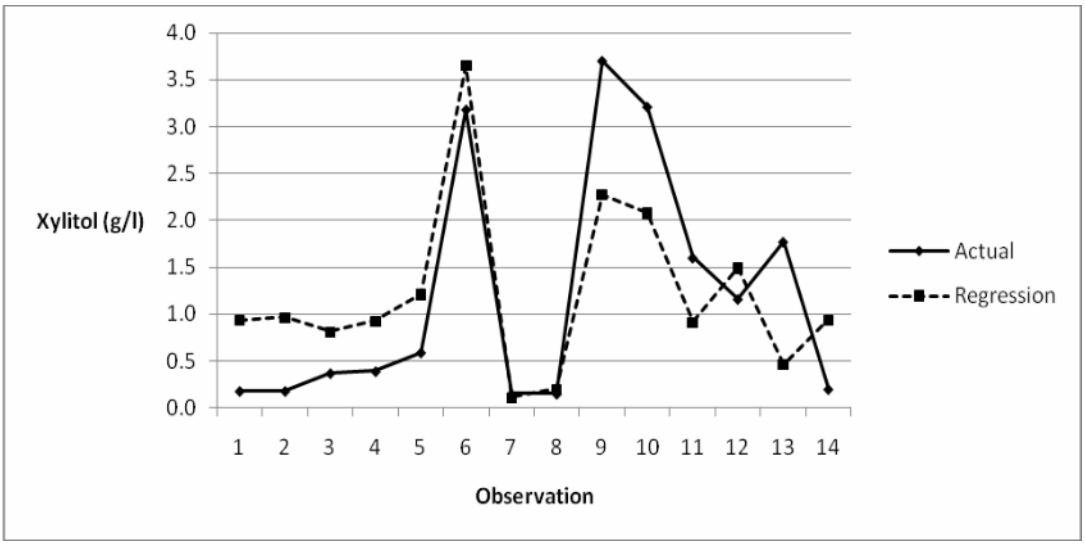

Figure 34 A plot between actual and predicted values of xylitol concentration from regression model on the validating data set

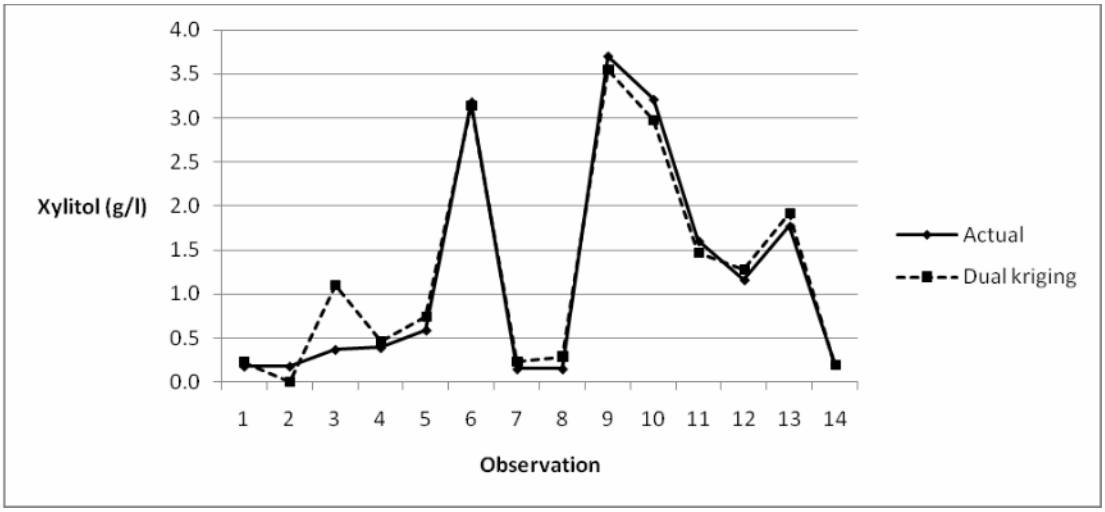

Figure 35 A plot between actual and predicted values of xylitol concentration from dual kriging model on the validating data set

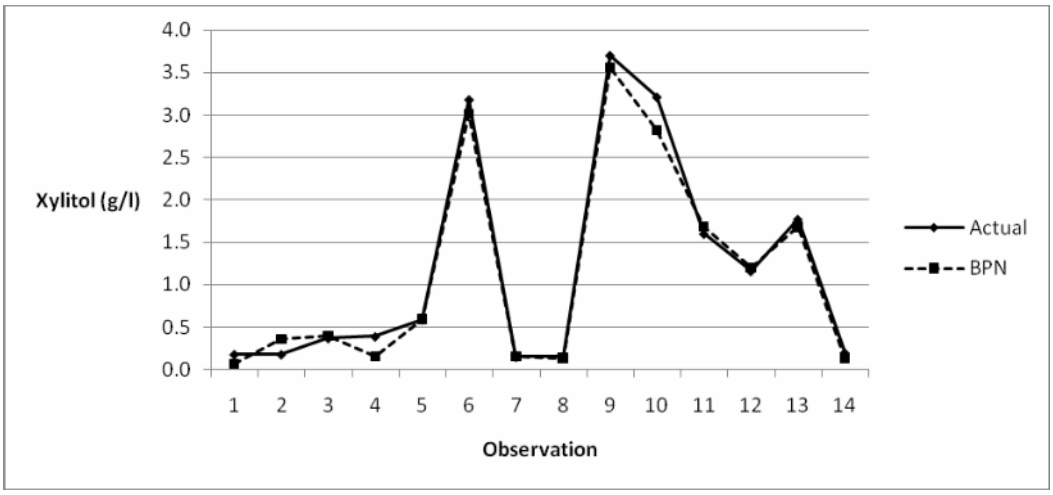

Figure 36 A plot between actual and predicted values of xylitol concentration from BPN model on the validating data set

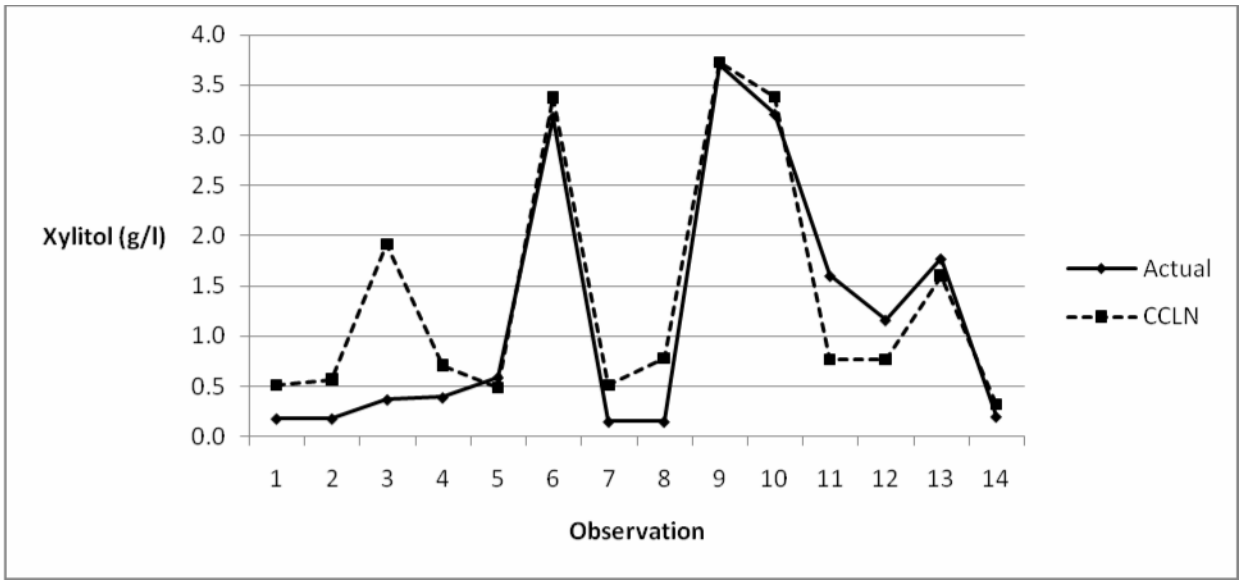

Figure 37 A plot between actual and predicted values of xylitol concentration from CCLN model on the validating data set

3. Comparison between biomass predictive models and xylitol predictive models

Table 10 compares the performance of biomass predictive models and xylitol predictive models in terms of prediction accuracy using mean absolute percentage error (MAPE) and model bias. It is recoginized based on prediction accuracy (MAPE) that cell biomass predictive models are higher. In other words, it is easier to predict the cell biomass than the xylitol concentration. A reason accounts for this result is that the xylitol concentration value (0-5 g/l) is quite lower than that (1-30 g/l) of the cell biomass. With such low values, uncertainty in its measurement and rounding off in calculation may lead to inaccuracy. This might also contribute to the higher in the xylitol concentration's model bias than the cell biomass's.

Table 10 Comparison of the performance of biomass and xylitol predictive models

Prediction Type	Model	MAPE (%)		Model Bias	
		Fitting Set	Validating Set	Fitting Set	Validating Set
Cell biomass	2^{nd} order polynomial	20	19	Unbiased	Unbiased
	1^{st} order drift function	0	11	Unbiased	Slightly underestimated
	3-7-1* BPN	12	12	Unbiased	Unbiased
	3-5-1* CCLN	16	9	Unbiased	Unbiased
Xylitol concentration	1^{st} order polynomial	216	134	Unbiased	Fairly overestimated
	2^{nd} order drift function	0	39	Unbiased	Slightly overestimated
	3-8-1* BPN	28	22	Unbiased	Very slightly underestimated
	3-6-1* CCLN	198	124	Slightly overestimated	Fairly overestimated

* Number of input neurons - number of hidden neurons - number of output neurons of the neural network.

Conclusions and Recommendations

This research investigates the potential use of 4 types of empirical models i.e. regression, dual kriging, back propagation neural network (BPN) and cascade correlation neural network (CCLN) in modeling the relationship between 3 input factors, i.e., recycle ratio, aeration rate, and fermentation time of the xylitol production from *Candida mogii* on 2 process outputs, i.e., cell biomass and xylitol concentration. Each ouput is modeled separately. For cell biomass prediction model, the 3-5-1 CCLN model outperforms the other 3 models with respect to its generalized prediction accuracy with RMSE of 1.13 g/l and MAPE of 9% in the validating data set and it demonstrates an unbiased type. For xylitol concentration prediction model, 3-7-1 BPN model exhibits highest generalized prediction accuracy with RMSE of 0.16 g/l and MAPE of 22% in the validating data set. However, this BPN model very slightly underestimates the data. As a consequence, care must be taken when using this model. All in all, artificial neural network (ANN) models (BPN and CCLN) are identified to be more accurate and more relaiable than the statistical models like regression and dual kriging in predicting cell biomass and xylitol concentration in xylitol production process. In terms of ease of building the models, several commercial softwares are available for regression as well as ANN models. Nevertheless, ANN model bulding and validation requires longer time than statistical models. One then needs to tradeoff between its performance and development time and costs.

References

Abril, J. R., Stull, J. W., Taylor, R. R., Angus, R. C. and Daniel, T. C., Journal of Food Science, Vol. 47, (1982), pp. 472-475.

Aminoff, C., Vanninen, E., and Doty, T. E., "The Occurrence, Manufacture and Properties of Xylitol," In Counsell J. N. ed., Xylitol, (London: Applied Science Publisher, 1978), pp. 1-9.

Barbosa, M. F. S. de Medeiros, M. B. de Mancilha, I. M., Scheider, H., and Lee, H., "Screening of Yeasts for Production of Xylitol from D-Xylose and Some Factors which Affect Xylitol Yield in *Candida guilliermondii*," J. Ind. Microbiol., Vol. 3, (1988), pp. 241-251.

Bär, A., "Caries Prevention with Xylitol: A Review of the Scientific Evidence," Wld. Rev. Ntr. Diet., Vol. 55, (1988), pp. 183-209.

Bär, A., "Xylitol," In Nabors, L. O. and Gelardi, R.C. eds., Alternative Sweeteners, 2nd edition, (NY: Marcel Dekker,1991), pp. 349-379.

Barton, R.R, "Metamodels for Simulation Input-Output Relations," Proceedings of the 1992 Winter Simulation Conference, (1992), pp. 289-299.

Bassler, K. H. "Biochemistry of Xylitol,", In Counsell, J. N. eds., Xylitol, (London: Applied Science Publishers, 1978), pp. 35-41.

Chaveesuk, R., "A Metamodeling Approach to Sensitivity Analysis of Capital Investments," (Ph.D. dissertation, Department of Industrial Engineering, University of Pittsburgh, USA), 2000.

Chaveesuk, R. and Smith, A.E. 2005. Dual Kriging : An Exploratory Use in Economic Modeling, The Engineering Economist, Vol. 50 (3), (2005), pp. : 247-271.

Chen, H-C, "Optimizing the Concentration of Carbon, Nitrogen and Phosphorus in a Citric acid Fermentation with response Surface Method," Food Biotechnology, Vol. 10, (1996), pp. 13-27.

Cleran, Y., Thibault, J., Cheruy, A., and Corrieu, G., "Comparison of Prediction Performances between Models Obtained by the Group Method of data handling and Neural Networks for the Alcoholic Fermentation Rate in Ecology," J. Ferment Bioeng.,Vol. 71, (1991), pp.356-362.

Cressie, N. A. C., Statistics for Spatial Data, (NY: John Wiley & Son, 1991), pp. 3.

Counsell, J.N., ed., Xylitol, "The Occurrence, Manufacture and Properties of Xylitol," by Aminnoff, C., Vanninen, E., and Doty, T.E., (Applied Science Publishers, London, 1978), pp. 1-9.

Echaabi, J., Trochu, F., and Gauvin, R., “A general strength theory for composite materials based on dual kriging interpolation,” Journal of Reinforced Plastics and Composites, Vol. 14, (1995), pp. 211-232.

Echaabi, J., Trochu, F., and Gauvin, R., “Review of failure criteria for fibrous composite materials,” Polimer Composites, Vol. 17(6), (1996), pp.786-798.

Emodi, A., “Xylitol: Its Properties and Food Applications,” Food Technology, January, (1978), pp. 28-32.

Fang, B., Chen, H., Xie, X., Wan, N., and Hu, Z., “Using Genetic algorithms Coupling Neural Networks in a Study of Xylitol Production : Medium Optimization,” Process Biochemistry, (2002), pp. 1-6.

Faria, L. F. F., Gimenes, M. A. P., Nobrega, R., and Pereira, Jr. N., “Influence of Oxygen Availability on Cell Growth and Xylitol Production by *Candida guilliermondii*,” Appl. Biochem Biotechnol., Vol. 98, (2002), pp. 449-458.

Fausett, L., Fundamental of Neural Networks: Architectures, Algorithms, and Applications, (Englewood Cliffs, NJ: Prentice-Hall, 1994), p. 3.

Förster, H., In Sippple, H. L. and McNutt, K. W. eds, Sugars in Nutrition, (NY: Academic Press, 1974), p. 259.

Funahashi, K., “On the Approximate Realization of Continuous Mappings by Neural Networks,” Neural Networks, Vol. 2 (1989), pp. 183-192.

Glassey, J., Montague, G.A., Ward, A.C., and Kara, B.V., “Enhanced Supervision of Recombinant E. coli Fermentations via Artificial Neural Networks, Proc. Biochem, Vol. 29, (1994), pp. 387-398.

Greenby, T.H. ed., *Progress in Sweetener,* 2nd edn., “Latest Dental Studies on Xylitol in Caries limiation,” by Makinen, K.K., (Elsevier Applied Science, London, 1992), pp. 331-362.

Hahn-Hägerdal, B., Jeppsson, H. Skoog, K., and Prior B. A., “Biochemistry and Physiology of Xylose Fermentation by Yeasts,” Enzyme. Microbiol. Technol., Vol. 16, (1994), pp. 933-943.

Honda, H. Hanai, T., Katayama, A., Tihoyama, H., and Kobayashi, T, “Temperature Control of Ginjo Saki Mashing Process by automatic Fuzzy Modeling Using Fuzzy Neural Networks,” J. Fermentation Bioeng., Vol. 85, (1998), pp. 107-112.

Horitsu, H., Yahashi, Y., Takamizawa, K., Kawai, K., Suzuki, T., and Watanabe, N., “Production of Xylitol from D-Xylose by *Candia tropicalis*: Optimization of Production Rate,” Biotechnol. Bioeng., Vol.40, (1992), pp.1085-1091.

Hornik, K., Stinchcombe, M., and White, H., “Multilayer Feedforward Networks are Universal Approximators,” Neural Networks, Vol. 2 (1989), pp. 359-366.

Hyvönen, L. Koivisitoinen, P., and Voirol, F., “ Food Technological Evaluation of Xylitol,” In Chichester, C. O., Mrak, E. M., and Stewart, G. eds, Advances in Food Research, Vol. 28, (NY: Academic Press, 1982), pp. 373-403.

Insa, G. Sablayrolles, J-M., and Douzal, V., “Alcoholic Fermentation under Oenological Conditions: Use of a Combination of Data Analysis and Neural Networks to Predict Sluggish and Stuck Fermentations,” Bioproc. Eng., Vol. 13, (1995), pp. 171-176.

Jeyamkondan, S., Jayas, D.S., and Holley, R.A., “Microbial Growth Modeling with Artificial Neural Networks,” International Journal of Food Microbiology, Vol. 64, (2001), pp. 343-354.

Journel, A.G. and Huijbregts, Ch. J., Mining Geostatistics, (NY: Academic Press, 1978).

Kleijnen, J.P.C. and Sargent, R.G., “A Methodology for Fitting and Validating Metamodels in Simulation,” European Journal of Operational Research, Vol. 120 (2000), pp. 14-29.

Krull, L. N. and Inglett, G. E., “Analysis of Neutral Carbohydrates in Agricultural Residues by GLC,” J. Agric. Food Chem., Vol. 28, (1980), pp. 17-19.

Kwok, T. and Yeung, D., “Constructive Algorithm for Structure Learning in Feedforward Neural Networks for Regression Problems,” IEEE Transactions on Neural Networks, Vol. 8, No. 3 (May 1997), pp. 630-645.

Lekha, P.K., Chand, N., and Losane, B.K., “Computerized Study of Interactions among factors and Their Optimization through Response surface Methodology for the Production of Tannin Acyl Hydolase by *Aspergillus niger* PKL104 under Solid State Fermentation,” Bioproc. Eng. Vol. 7, (1994), pp.7-15.

Leondes, C.T., ed., Industrial and Manufacturing Systems. Vol. 4 of Neural Network Systems Techniques and Applications, “Dynamic Nonlinear Systems, by Y.C. Shin”, (San Diago, CA:Academic Press, 1998), pp. 348.

Lindskog, P. and Ljung, L. “Tools for Semi-Physical Modeling,” Proc. Of the IFAC SYSID 1994 Conference, Copenhagen, Denmark, (1994), p. 237-242.

Mäkinen,K. K., ”Latest Dental Studies on Xylitol and Mechanism of Action of Xylitol in Caries Limitation,” In Greenby, T. H. eds., Progress in Sweeteners, 2nd edition, (NY: Elsevier Applied Science, 1992)

Mamat, A., Trochu, F. and Sanschagrin, B., “Analysis of shrinkage by dual kriging for filled and unfilled polypropylene moldes parts,” Polymer Engineering and Science, Vol. 35(19), (1995), pp.1511-1520.

Mori S. and Saraya, J., Journal of Japanese Kokai Tokkyo Koho, Vol. 63, (1988), pp. 135-322.

Neter, J., Wasserman, W., and Hutner, M.H., Applied Linear Statistical Models, (3rd edition; Burr Ridge, IL: Richard D. Irwin, 1990).

Nigam, P. and Singh, D., “Process for Fermentative Production of Xylitol: A Sugar Substitute,” Process Biochem., Vol. 35, (1995), pp. 117-124.

Nolleau, V., Preziosi-Belloy, L., and Navarro, J.M., “The Reduction of Xylose to Xylitol by *Candida* guilliermondii and *Candida* parpsilosis: Incidence of Oxygen and pH,” Biotechnol. Lett., Vol. 17, (1995), pp. 417-422).

Oishi, K., Tominaga, M., Kawato, A., and Imayasu, S., “Analysis of the state Characteristics of Sake Brewing with a Neural Network,” J. Ferm. Bioeng., Vol. 73, (1992), pp. 153-158.

Ojamo, H., Yeast Xylose Metabolism and Xylitol Production, Technical Research Centre of Finland, (Espoo, Finlandia: VTT Publications, 1994).

Parajo, J. C., Dominguez, H., and Dominguez, J. M., “Biotechnological Production of Xylitol Part1: Interest of Xylitol and Fundamental of Its Biosynthsis,” Bioresource Echnology, Vol. 65, (1998a), pp. 191-201.

Parajo, J. C., Dominguez, H., and Dominguez, J. M., “Biotechnological Production of Xylitol Part2: Operation in Culture Media Made with Commercial Sugars,” Bioresource Echnology, Vol. 65, (1998b), pp. 203-212.

Pepper, T and Olinger, P. M., “Xylitol in Sugar-Free Confections,” Food Technology, Vol. 10, (1988), pp. 98-106.

Pepper, T., “The Use of Xylitol in Confectionary Production,” Confectionary Prod., Vol, 3, (1989), pp. 253-256.

Phatak, D.S. and Koren, I., “Connectivity and Performance Tradeoffs in the Cascade Correlation Learning Architecture,” IEEE Transactions on Neural Networks, Vol. 5, No. 6 (November 1994), pp. 930-935.

Porier, C. and Tinawi, R. “Finite element stress tensor fields interpolation and manipulation using 3D dual kriging,” *Computers & Structures*, (1991), pp. 211-222.

Pooirier, C. and Tinawi, R., “Finite Element Stress Tensor Fields Interpolation and Manipulation Using 3D Dual Kriging,” *Computers & Structures*, Vol. 40(2) pp. 211-222, 1991.

Prior, B. A., Killian, S.G., and du Preez, J. C., "Fermentation of D-Xylose by the Yeast *Candida Shehatae* and *Pichia* sipitis," Proc. Biochem., Vol. 24, (1989), pp. 21-32.

Psichogios, D. C. and Ungar, L. H., "A Hybrid Neural-Netwok-First Principles Approach to Process Modeling," A. I. Ch. E. Journal, Vol. 38, (1991), p. 1499.

Rattle, A., "Accelerating the Convergence of Evolutionary Algorithms by Fitness Landscape Approximation," in Proceedings of the 5th International Conference on Parallel Problem Solving from Nature-PPSN V*, Amsterdam, The Netherlands*, September, (1998), pp. 87-96.

Rizzi, M., Erlemann, P., Bui-Thanh, N. A., Dellweg, H., "Xylose Fermentation by Yeasts: Purification and Kinetic Studies of Xylose Reductase from *Pichia stipitis.,*" Appl. Microbiol. Biotechnol., Vol. 29, (1988), pp. 148-154.

Roberto, I.C. Sato, S.,de Mancilha, I.M., and Taqueda, M.E.S., "Influence of Media Composition on Xylitol fermentation by Candida guiliiermondii Using Response Surface Methodology," Biotechnol. Lett., Vol.10, (1995), pp. 1223-1228.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J., "Learning Internal Representations by Error Propagation," in Rumehart, D. E. and McClelland, J. L. eds., Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations, (Cambridge, MA: MIT Press, 1986), pp. 318-362.

Rumehart, D.E. and McClelland, J.L. eds., "Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations, Learning Internal Representations by Error Propagation," by D.E. Rumelhart, G.E. Hinton, and R.J. Williams", (Cambridge, MA: MIT Press), (1986), pp. 318-362.

Sanronan M. A., Dominguez, M. J., Nunez, J. M. Lema, J. M., Afinidad XLVIII 434, (1991), p. 215.

Saval, S., Pablos, L. and Sanchez, S., "Optimization of a Culture Medium for Streptomycin Production Using Response Surface Methodology," Bioresource Technol., vol. 43, (1993), pp.19-25.

Shi, Y. and Weimer, P.J., "Response Surface Analysis of the effects of pH and dlution rate on Ruminococcus flavefaciens FD-1 in cellulose-fed continuous culture," Appl. Environ. Microbiol., Vol.58, (1992), pp. 2583-2591.

Sirisansaneeyakul, S., Staniszewski, M., and Rizzi, M., "Screening of Yeasts for Production of Xylitol from D-Xylose," J. Feremnt. Bioeng., Vol. 6, (1995), pp. 564-570.

Sirisansaneeyakul, S., Tochampa, W., Bashir, I., Rizzi, M., and Bhuwapathanapun, S., "Kinetic Modeling of pH Affecting Xylitol Production by *Candida mogii*," Thai J. Agric. Sci., Vol. 33(3-4), (2000), pp.159-166.

Sirisansaneeyakul, S., Tochampa, W., Bashir, I., Rizzi, M., and Bhuwapathanapun, S., "Continuous Production of Xylitol by Cell Recycling System," Thai J. Agric. Sci. Vol. 33(3-4), (2000), pp. 99-106.

Silva, S. S. Roberto, I. C., Felipe, M. G. A., and Mancilha, I. M., "Batch Fermentation of Xylose for Xylitol Production in Stirred Tank Bioreactor," Process Biochem., Vol. 31, (1996), pp. 549-553.

Sreenath H. K. and Jeffries, T. W., Applied Biochemistry and Biotechnology, Vol. 57(8), (1996), pp. 551-561.

Tochampa, W., "Xylitol by Cell Recycling with Hollow Fibers," (M.S. thesis, Department of Biotechnology, Faculty of Agro-Industry, Kasetsart University, Thailand,1998).

Touster, O., In Sipper, H. L. and McNutt, K. W. eds, Sugars in Nutrition, (NY: Academic Press, 1974) , p. 229.

Winkelhausen, E. and Kuzmanova, S., "Review: Microbial Conversion of D-Xylose to Xylitol," Journal of Fermentation and Bioengineering, Vol. 86(1), (1998), pp. 1-14.

te Braake, H. A. B., van Can, H. J. L., Vebruggen, H. B., "Semi-mechnaistic Modeling of Chemical Process with Neural Networks," Engineering Applications of Artificial Intelligence, Vol. 11, (1998), p. 507-515.

Teng, C. and Wah, B.W., "Automated Learning for Reducing the Configuration of a Feedforward Neural Network," IEEE Transactions on Neural Networks, Vol. 7, No. 5 (September 1996), pp. 1072-1085.

Touretzky, D.S. ed., *Advances in Neural Information Processing Systems 2*, "The Cascade-Correlation Learning Architecture, by S.E. Fahlman and C. Lebiere," (San Mateo, CA: Morgan Kaufmann Publishers, 1990), pp. 524-532.

Trochu, F., "A Contouring Program Based on Dual Kriging Interpolation," Engineering with Computers, (1993), pp. 160-177.

Twomey, J. M. and Smith, A. E., “Validation and Verification,” In Kartam, N. Flood, I., and Garett, J. eds., Artificial Neural Networks for Civil Engineer: Fundamental and Applications, (ASCE Press, 1996), pp. 44-64.

van Deventer, J. S. J., Kam, K. M., van der Walt, T. J., “Dynamic Modeling of a Carbon-in-Leach Process with the Regression Network,” Chemical Engineering Science, Vol. 59, (2004), pp. 4575-4589.

Vandeska, E., Amartey, S., Kuzmanova, S., and Jeffries, T. W., Fed-Batch Culture for Xylitol Production by *Candida boidinii*, Process Biochem., Vol. 31, (1996), pp. 265-270.

Vlassides, S., Ferrier, J.G., and Block, D.E., “Using Historical Data for Bioprocess Optimization: Modeling Wine Characteristics Using Artificial Neural Networks and Archived process Information,” Biotechnol. Bioeng., Vol. 73, (2001), pp.55-68.

Washuttl, J., Reiderer, P., and Bancher, E., “A Qualitative and Quantitative Study of Sugar-Alcohols in Several Foods,” J. Food Sci., Vol. 38(1978), pp. 1262-1263.

Winkelhausen, E. and Kuzmanova, S., “Microbial Conversion of D-Xylose to Xylitol,” Journal of Fermentation and Bioengineering, Vol. 86(1), pp. 1-14, 1998.

Worbos, P.J., “Beyond Regression: New Tools for Prediction and Analysis in the Behaviorial Sciences.” (unpublished Ph.D. dissertation, Department of Statistics, Harvard University, 1974).

Yahashi, Y., Horitsu, H., Kawai, K., Suzuki, T., and Takamizawa, K., Journal of Fermentaion Bioengineering, Vol. 81, (1996a), pp. 148-152.

Yahashi, Y., Hatsu, M., Horitsu, H., Kawai, K., Suzuki, T., and Takamizawa, K., Biotechnology Letters, Vol. 18, (1996b), pp. 1395-1400.

Yang, X-M, “Optimization of a cultivation process for recombinant Protein Production by Escherichia coli, J. Biotechnol., Vol. 23, (1992), pp. 271-289.

Ylikahri, R. In Chichester, C. O., Mrak, E. M., and Stewart, G. eds, Advances in Food Research, Vol. 25, (NY: Academic Press, 1979), pp. 159-180.

OUTPUT ที่ได้จากโครงการ

1. ผลงานวิจัยนี้นำไปใช้ในการเรียนการสอนวิชา Artificial neural networks and Its Application in Agro-Industry

2. ผลงานวิจัยนี้นำไปเป็นต้นแบบในการทำงานวิจัยเรื่องการประยุกต์แบบจำลองเชิงประจักษ์ในการทำนายการเจริญเติบโตของเชื้อจุลินทรีย์ *Vibrio parahaemolyticus* ในกุ้งขาวแช่เยือกแข็ง และผลิตนิสิตระดับปริญญาโทสาขาการจัดการเทคโนโลยีอุตสาหกรรมเกษตร ภาควิชาเทคโนโลยีอุตสาหกรรมเกษตร คณะอุตสาหกรรมเกษตรจำนวน 1 คน คือนายผดุงเดช พูลสุข