STOCHASTIC MODELLING OF THE GROWTH OF *C. ACETOBUTYLICUM* WITH MISSING DATA

NORLIANA BINTI MOHD. LIP

UNIVERSITI TEKNOLOGI MALAYSIA

STOCHASTIC MODELLING OF THE GROWTH OF

C. ACETOBUTYLICUM WITH MISSING DATA

NORLIANA BINTI MOHD. LIP

This thesis is submitted in fulfillment of the requirement for the award of the

Master of Science (Mathematics)

Faculty of Science

Universiti Teknologi Malaysia

NOV 2009

To my beloved parents, sisters, brothers and friends.

ACKNOWLEDGEMENT

First and foremost, all praise be to ALLAH, the Almighty, the Benevolent for HIS blessings and guidance for giving me the inspiration to embark on this thesis and instilling me the strength to see that this thesis becomes reality. I was able to complete my thesis successfully.

I would like to take this opportunity to express my deepest gratitude to my supervisor, Dr. Arifah binti Bahar who gave me her guidance and suggestions to me. Thank you for giving me her time and effort to complete this thesis successfully.

My sincere appreciation also extents to all my family members especially my parents who have been with me every steps of the way and give me support. I want to thank my fellow friends who have provided assistance at various occasions in order to complete this thesis as well as giving their great support and encouragements for me. Thanks for always being there to share all my happiness and sorrow.

Last but not least, I appreciate all the support and guidance provided by everyone who ever help me directly or indirectly. Thank you to all.

ABSTRACT

Stochastic influences play an important role in various areas especially in the area of biological process. Stochastic differential equation is the differential equation in which the terms of their characteristic involve stochastic process or 'white noise'. In this study, we used the stochastic differential equation to describe the population dynamics of the cell growth of *C. Acetobutylicum* in fermentation process. Stochasticity incorporated into the model via its growth coefficient$-\frac{\mu_{max}}{y_{max}}$. We used the model of stochastic logistic to model the growth of cell against time at different initial pH. The range of initial pH level is from 4.0 until 7.0. The missing data were estimated using expectation maximization (EM) and regression approach. The estimated parameters were obtained using simulated maximum likelihood. The estimated μ_{max} and ε values of stochastic differential equation at five different initial pH level (4.0, 4.5, 5.0, 6.0, and 7.0) are (0.1098, 0.09), (0.154, 0.04), (0.41, 0.01), (2.92, 0.113) and (0.341, 0.09) respectively. Five different trajectories for different initial pH were formed based on EM and regression approximation. It was found that all trajectories based on EM show a lower mean square error as compared to those approximated using regression. Thus, EM estimate is a better estimator for missing data and the model is adequate. It was also found that the means square error for stochastic are lower than deterministic model at five different initial pH. This implies that stochastic logistic model is better in describing the growth of cell *C.Acetobutylicum* in fermentation process compared to deterministic model.

ABSTRAK

Faktor-faktor rawak memainkan peranan yang penting dalam pelbagai bidang terutamanya yang melibatkan proses biologi. Persamaan pembezaan rawak merupakan persamaan terbitan dimana sebutan dalam persamaan tersebut melibatkan proses rawak. Kajian ini menggunakan persamaan pembezaan rawak untuk menggambarkan pergerakan populasi sel *C. Acetobutylicum* dalam proses fermentasi. Kerawakan dimasukkan ke dalam model terhadap pemboleh ubah pertumbuhan iaitu $-\frac{\mu_{max}}{y_{max}}$. Kami menggunakan model logistik rawak untuk menggambarkan pertumbuhan sel tersebut terhadap masa pada nilai pH awal yang berlainan. Julat nilai pH awal adalah di antara 4.0 hingga 7.0. Data yang tidak lengkap dianggarkan dengan menggunakan kaedah jangkaan maksima dan kaedah regresi. Nilai anggaran parameter diperolehi dengan menggunakan kaedah 'simulated maximum likelihood'. Nilai anggaran parameter μ_{max} dan ε bagi persamaan pembezaan rawak pada lima nilai pH (4.0, 4.5, 5.0, 6.0, dan 7.0) adalah (0.1098, 0.09), (0.154, 0.04), (0.41, 0.01), (2.92, 0.113), dan (0.341, 0.09). Sebanyak lima lintasan telah disimulasi berdasarkan kaedah nilai anggaran maksima dan nilai regresi. Hasil daripada simulasi tersebut, nilai kesalahan kuasa dua bagi kaedah nilai anggaran maksima lebih kecil berbanding menggunakan kaedah nilai regresi. Oleh itu, kaedah nilai anggaran maksima merupakan kaedah yang sesuai bagi menganggar nilai kehilangan data. Selain itu, nilai kesalahan kuasa dua pada 5 nilai pH yang berbeza bagi model rawak adalah lebih kecil berbanding model penentu. Ini menunjukkan bahawa model logistik rawak sesuai digunakan bagi menggambarkan pertumbuhan sel *C.Acetobutylicum* dalam proses fermentasi.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

This chapter discusses the background of the problem. It also includes statement of the problem, objectives of the study, scopes of the study, significance of the study and thesis organization.

1.2 Background of the Problem

A stochastic process which can be called as a random process is the counterpart to a deterministic process in probability theory. Instead of dealing with only one possibility of how the process might evolve under time for example the solutions of an ordinary differential equation, in a stochastic process there is some indeterminacy in its future evolution described by probability distributions. This means that even if the initial condition is known, there are many possibilities the process might go to, but some paths are more probable and others less.

Stochastic differential equation (SDE) is a differential equation where one or more of the terms involve in stochastic process. Typically, stochastic differential equation incorporates white noise which can be thought of as the derivative of Brownian motion. Non-stochastic differential equation is a model of dynamical systems where the state evolves continuously in time. If they are autonomous, then the state's future values depend only on the present state; if they are non-autonomous, it is allowed to depend on previous state as well.

With all these features, stochastic differential equation can be applied in such diverse areas such as neural networks, ecosystem dynamics, population genetics and macro-economic systems.

In this dissertation, stochastic differential equation had been used to describe the population dynamics of fermentation process. The proposed stochastic model is stochastic logistic model. The solution to stochastic logistic model was approximated using Euler - Maruyama method. The approach taken here was estimated the parameters for stochastic model using simulated maximum likelihood and also studied the missing value analysis to impute the missing values of the data before doing a simulation.

1.3 Statement of the Problem

Stochastic influences play an important role in biological processes. Thus, fermentation process which is one of the biological processes will be discussed in this study that may posses the stochasticity of the cell growth. The problem arises when the used data involve incompleteness data and expectation maximization (EM) and regression imputation will be carried out. Thus, we want to compare these two methods in dealing with missing value analysis. The approach taken here was also to estimate parameter estimations using simulated maximum likelihood and the data will be analyzed using Matlab. The cell growth of *C.Acetobutylicum* in fermentation process will be modelled using stochastic logistic model. Finally, we will make a comparison of this model with deterministic model in order to describe the cell growth of *C.Acetobutylicum* in fermentation process of sago starch.

1.4 Objective

The objectives of this study are:

i. To compare expectation maximization (EM) and regression as imputation methods in dealing with missing value analysis.
ii. To estimate the parameters of the model using simulated maximum likelihood.
iii. To model the growth of *C.Acetobutylicum* in fermentation process using the stochastic logistic model.

1.5 Scope of the Study

This study discussed the appropriateness of stochastic logistic equation as a model for the cell growth of *C.Acetobutylicum* in fermentation process. The approximate solution of the model was obtained using Euler- Maruyama approximation. This study focused on the model parameters which were estimated using simulated maximum likelihood. This study compared only EM and regression in estimating the stochastic differential equation based on Ito approach.

1.6 Significance of the Study

The contribution of this study is in findings an adequate model of *C.Acetobutylicum* cell growth in fermentation process in a more realistic way and also in modelling stochastic differential equation with missing values.

1.7 Thesis Organization

Chapter 1 discussed the background of problem stochastic differential equation (SDE) in general followed by statement of the problem, the objectives of the study, scope of the study as well as the significance of the study. Lastly, we included the thesis organization to review the overall of the study.

Chapter 2 introduced the fermentation process especially in microbiology. The review of researches development in fermentation process of sago starch will be discussed in this chapter. Besides, it was also included the discussion of stochastic differential equation in biological process. Lastly, the growth cycle of microbial as well as factors that influenced the bacterial growth also will be discussed in this chapter.

Chapter 3 discussed about the missing values analysis before numerically estimate the parameter based on the model of stochastic logistic. Besides, we discussed the numerically method of parameter estimations as well as the method of approximation to perform the simulation.

Chapter 4 discussed about imputation of missing data value in SPSS. The application of stochastic differential equation in fermentation process of sago starch to determine the values of parameter estimation will also be included here. Lastly, we discussed the results of the simulation.

Chapter 5 discussed the conclusion of the whole study and some recommendations for those who interested to pursue the study in various area of the stochastic differential equation.

REFERENCES

[1] M.A. Bazli, B. Arifah, M.S. Madihah, A.R. Haliza and R. Norhayati. (2009). *Stochastic Growth of the C.Acetobutylicum*, Universiti Teknologi Malaysia.

[2] R.D. Yates, and D.J. Goodman. (2005). *Probability and Stochastic Processes*, John Wiley, U.S.

[3] E.M.T. El-Mansi, C.F.A. Bryce, A.L.Demain, and A.R.Allman. (2007). *Fermentation Microbiology and Biotechnology*. Taylor & Francis Group, U.S.

[4] M.S. Madihah, A.B.Ariff, K.M. Sahaid, A.A. Suriani, M.I.A.Karim. (2001). *Direct fermentation of gelatinized sago starch to acetone, butanol, ethanol by C.Acetobutylicum.* World Journal of Microbiology and Biotechnology, Kluwer Academic, Netherland.

[5] D.G. Kim and M.S. Oh. (2008). *Fully Adaptive Simulated Maximum Likelihood Method*, Department of Applied Mathematics, Hanyang University and Department of Statistics, Ewha Womans University.

[6] D.Lies, O.Megumi, M.Shigenobu, Y.Masaaki, and O.Mitsuyasu, (2007). *Itaconic Acid Production using Sago Starch Hydrolysate by Aspergillus terreus TN484-M1*, Elsevier, Japan.

[7] M. Watanabe, and K.Yamaguchi. (2004). *The EM Algorithm and Related Statistical Models.* Marcel Dekker, US.

[8] J. Wolfgang. *Efficient Simulated Maximum Likelihood with an Application to Online Retailing.* Department of Decision and Information Technologies, University of Maryland.

[9] Muthuvelayudham, R and Viruthagiri, T. (2007). *Optimization and Modeling of Cellulase Protein from Trichoderma Reesei Rut C30 Using Mixed Substrate*. African Journal of Biotechnology. Vol.6(1), 041-046.

[10] Lawrence C. Evans. *An Introduction to Stochastic Differential Equation.* Department of Mathematics, Version 1.2, UC Berkeley.

[11] S. Naresh, K. L. Kalra, S.0. Harinder and B.Sunil. (2000).*Optimization of Fermentation Parameters for Production of ethanol from Kinnow Waste and Banana Peels by Simultaneous Saccharification and Fermentation.* Indian J. Microbiol.Vol.47, 310–316.

[12] M. Kinder and W. Wiechert. (1996). *Stochastic Simulation of Biotechnical Processes.* Department of Theoretical Biology, University of Bonn, Germany. Elsevier Science.

[13] L. Cobb. (1996). *Stochastic Differential Equation for Social Science.* (1981). Mathematical Frontiers of the Social and Policy Science. Westview Press.

[14] F. Molinari. (2007). *Missing Treatments.* Department of Theoretical Biology, Department of Economics, Connell University.

[15] H. Tim. (1999). *A Graphical Representation Little's test for MCAR.* MathSoft Inc. USA.

[16] Rebecca M. Weiss and David F. Olis. (1980). *Extracullar Microbial Polysaccharides. I. Substrate, Biomass and Product Kinetic Equations for Batch Xanthan Gum Fermentation.* Biotechnology and Bioengineering.Vol.XXII, pg: 859–873.

[17] N. Bouleau and D. Lepingle (1994). *Numerical Methods for Stochastic Processes.* Wiley Series, pg: 269–329.

[18] N.T. Longford. (2005). *Missing Data and Small Area Estimation: Modern Analytical Equipment for the Survey Statistician.* Springer Science + Business Media, Inc. US.

[19] R. Norhayati, B. Arifah, S.H. Yeak, A.R. Haliza, M.S. Madihah, and M.A. Bazli. (2009). *The Performance of Euler-Maruyama and 2-stage SRK in Approximating the Strong Solution of Stochastic Model*, Universiti Teknologi Malaysia.

[20] A.R. Haliza, B. Arifah, M.A. Bazli, R. Norhayati, M.S. Madihah and Gerhard-Wilhelm Weber. (2009). *Levenberg- Marquardt Parameter Estimation of Stochastic Logistic Model*, Universiti Teknologi Malaysia.

[21] Picchini. U. (2007). *SDE Toolbox: Simulation and Estimation Of Stochastic Differential Equations with MATLAB*, http://sdetoolbox.sourgeforge.net

[22] M.S. Madihah. (2002). *Direct Fermentation of Gelatinized Sago Starch toSolvent (Acetone Butanol-Ethanol) by Clostrium Acetobutylicum P262.* PhD Thesis, Universiti Putra Malaysia.