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.