ISSN 1392 – 124X INFORMATION TECHNOLOGY AND CONTROL, 2013, Vol.42, No.1

An Approach to Identification of Dynamic Model for Optimization of Fed-Batch Fermentation Processes

Donatas Levišauskas, Tomas Tekorius

Process Control Department, Kaunas University of Technology
Studentų Str. 50, LT-51368 Kaunas, Lithuania
e-mail: donatas.levisauskas@ktu.lt, tomas.tekorius@ktu.lt

crossref http://dx.doi.org/10.5755/j01.itc.42.1.927

Abstract. A computation procedure is developed for identification of dynamic model intended for solving the biomass growth-related optimization problems at fed-batch fermentation processes. The model identification relies on exploiting a versatile structure model that covers several particular structure models commonly used in bioprocess engineering practice. The model parameters are found by using experimental data of batch or fed-batch fermentation processes and a consecutive identification procedure, which includes preliminary estimation of the parameters. Practical application of the proposed procedure of model identification is demonstrated for solving the feed-rate optimization problem in fed-batch culture *Escherichia coli*.

Keywords: Model identification; Optimization; Fed-batch fermentation process.

1. Introduction

The fed-batch culture technology is widely used in biotechnology for production of desired products: enzymes, antibiotics, recombinant proteins, etc. In fed-batch processes, the state of microorganisms' culture depends on concentrations of reacting substances that are controlled at desired levels by manipulating the feed-rate of substrate solution [1]. The feed-rate optimization is an urgent technological task at development of new and improvement of existent fermentation processes.

Modern approaches to dynamic optimization of fermentation processes refer to mathematical models, which are used for solving the practical optimization problems by mathematical methods [2-14]. Thereupon, identification of reasonable mathematical models of fermentation processes is the first stage of solving the model-based optimization problems.

The common type of models employed for optimization purposes are the first principle models, in which the rates of biochemical transformations in the mass balance equations are described by suitable kinetic relationships [15, 16], and the hybrid models, in which the rates are modeled by artificial neural networks [8-10].

Identification of the first principle models is not a trivial task and requires from the investigator specific experience and knowledge of the bioprocess modelling. The criterion of model quality usually is a coincidence between experimental data and model predictions; therefore, the model identification formalism is to find the most suitable kinetic relationships among the known ones that minimize the prediction error. It should be noted that a wide variety of kinetic relationships are developed for the modeling of particular bioprocesses [16]. The selection of suitable one requires knowledge of approximation abilities of functional relationships of various structures and detailed analysis of experimental data. In addition, identification of the nonlinear kinetic model parameters requires to predetermine or to guess initial values of the parameters in order to start an iterative identification procedure. Thereupon, identification of the first principle models by the traditional approach is a time-consuming task and is not attractive in daily bioengineering practice. It should be stressed that identification procedure of the hybrid models can be formalized [17], however, training of the neural networks requires sufficient statistics of experimental data. Thus, the hybrid models are not suitable at development of the new processes when only a limited amount of experimental data is available.

For solving the model-based feed-rate optimization problems several aproaches are commonly used. For simple mechanistic models of particular structures, analytical solutions and calculation algorithms can be obtained from the necessary conditions of optimality of the Maximum Principle [2-7]. For the models of

more complicated structure and the hybrid models, the parametric optimization approach is found to be effective for the feed-rate calculation [10-14]. In the parametric optimization procedure, the feed-rate time profiles are approximated by the time functions of universal structures, and the dynamic optimization problem is transformed into the nonlinear programming problem of optimization the parameters of the approximating function.

In this work we develop a universal optimization procedure for solving the model-based feed-rate time profile optimization problem of fed–batch fermentation process without external navigation directly from the process experimental data. In the optimization procedure, the dynamic model of universal structure is applied that covers the most common kinetic relationships and has a modest number of parameters to be identified. The presented approach allows overcoming the main problems of the model identification that complicate realization of automatic identification procedure: selection of relevant structure of dynamic model and determination of the first approach values of model parameters necessary to start the iterative identification procedure and to ensure fast convergence to the optimum values. The model-based optimal feed-rate calculation is performed by using the parametric optimization approach [14], in which parameters of the predetermined shape feed-rate time profile are calculated by the random search algorithm [18].

2. Dynamic model identification

The generalized model structure

The structure of dynamic model is chosen referring to analysis of kinetic relationships, which are used in mechanistic models for predicting the specific rates of biomass growth and substrate consumption [15, 16]. The selected structure model covers several typical mechanistic models that are most often used for modeling of batch and fed-batch fermentation processes:

$$\frac{dx}{dt} = \mu(s,x)x - u\frac{x}{V}, \quad (1)$$

$$\frac{ds}{dt} = -q(s,x)x + u\frac{s_f - s}{V}, \quad (2)$$

$$\frac{dV}{dt} = u, \quad (3)$$

$$\mu(s,x) = \mu_g(s,x) - \mu_d(x), \quad (4)$$

$$\mu_g(s,x) = \frac{\mu_{max}\, s}{\left(k_s + s + \frac{s^2}{k_i}\right)(1+\kappa_1 x)}, \quad (5)$$

$$\mu_d(x) = (k_x + k_{2x}), \quad (6)$$

$$q(s,x) = (Y_{xs} + \kappa_3 x)\mu_g(s,x) + \kappa_4, \quad (7)$$

where μ is specific biomass growth rate, q is specific substrate consumption rate, x is biomass concentration, s is substrate concentration; V is the volume of cultural liquid, u is feeding rate (control action), s_f is feeding substrate concentration, k_s, k_i, k_x, Y_{xs}, κ_1, κ_2, κ_3, κ_4, μ_{max} are model parameters subjected to identification.

In the presented model, equations (1)-(3) stand for dynamic mass balances on biomass, substrate and volume of cultural liquid, respectively. The specific growth rate of biomass (5) follows the Monod kinetics with extra regulatory terms that take into account the inhibitory effects of high substrate and biomass concentrations. The specific death rate (6) is assumed to be linear dependent on the biomass concentration. The specific substrate consumption rate (7) is proportional with the specific growth rate of biomass, and the extra terms are introduced to take into account the biomass concentration-related changes of yield coefficient and the biomass maintenance requirements.

Identification of parameters of the nonlinear model (1)-(7) by minimizing error between experimental data and the model predictions is not a trivial task because the first approach values of model parameters required in the iterative search algorithms are not known in advance. Setting of random initial values is not a reasonable approach as the random values may be far from those minimizing the fitting error, and convergence of numerical search procedure of the nine parameters may be poor. In addition, some parameters (Monod constant k_s, inhibitory constant k_i, yield coefficient Y_{xs}, maximum specific growth rate μ_{max}, etc) have physiological interpretation and random selection of the initial values of parameters does not guarantee their convergence to meaningful values.

The model identification difficulties also arise if the available number of experimental data points from investigated process is too small to ensure identification of model parameter values that provide the model with good prediction properties.

In this work, we develop a procedure for the model (1)-(7) identification, in which the experimental data are firstly approximated by a flexible logistic function that is well suited for approximation of s-type curves. The logistic function is further used to generate an extended number of data points for the dynamic model identification. The proposed identification procedure also includes estimation of the first approach values of model parameters that are further improved by an iterative search algorithm.

Procedure for identification of model parameters

The identification procedure consists of the following steps:

1. Approximation of experimental data from batch or fed-batch cultivation processes $x(t_i)$, $s(t_i)$, $V(t_i)$, $X(t_i) = x(t_i) \cdot V(t_i)$ and $S(t_i) = s(t_i) \cdot V(t_i)$, $i = 1, \dots, M$ (M is the number of experimental points) by the following logistic function:

$$y = \frac{b}{[1 + \exp(\sum_{l=0}^{m} a_l \cdot t^l)]}, \quad (8)$$

where t is the input variable (time), y is the output variable ($x, s, V,\ X\ or\ S$), b and a_l are identifiable parameters, m is the order of power series.

By processing the fed-batch culture experiments, the data from batch and feeding periods are approximated separately.

The parameters of the logistic function (8) are identified by transforming the function into a form that is linear with respect to parameters a_l:

$$ln\left(\frac{(b-y)}{y}\right) = \sum_{l=0}^{m} a_l \cdot t^l. \qquad (9)$$

With the experimental data points $x(t_i), s(t_i), V(t_i), X(t_i), S(t_i)$ and the predetermined value of parameter b, the parameters a_l can be directly calculated by the least squares method (LSM) [19]. The procedure of the parameters a_l calculation is repeated for the consistently increasing values of b until the best fitting of the model (8) predictions to the experimental data points (*minimum sum-of-squared error*) is obtained.

2. Generation of data points for the dynamic model identification.

An extended number of data points $x_j, s_j, V_j, X_j, S_j.$ $j = 1, \dots, N$ (N is the number of the generated data points) for identification of the dynamic model (1)-(7) is calculated from the identified logistic functions:

$$y_j = b \Big/ \left[1 + \exp\left(\sum_{l=0}^{m} a_l \cdot t_j^l\right)\right] \quad , j = 1, \dots, N \qquad (10)$$

The corresponding data points of the specific rates $\mu_j = \left(\frac{dX}{dt}\right) \cdot \frac{1}{X_j}$ and $q_j = \left[\left(\frac{dS}{dt}\right)_j - \left(\frac{dV}{dt}\right)_j \cdot s_f\right] \cdot \frac{1}{X_j}$ are calculated from the derivatives of the logistic function (8):

$$\left(\frac{dy}{dt}\right)_j = y_j \cdot \frac{y_j - b}{b} \sum_{l=0}^{m} \left(l \cdot a_l \cdot t_j^{l-1}\right). \qquad (11)$$

3. Estimation of the parameters k_s, k_i, k_x, μ_{max} of the model equations (4) - (6), assuming $\kappa_1 = 0$, $\kappa_2 = 0$.

With the above assumption, equation (4) is transformed into the inverse form:

$$\frac{s}{\mu + k_x} = \frac{k_s}{\mu_{max}} + \frac{1}{\mu_{max}} s + \frac{1}{k_i \mu_{max}} s^2. \qquad (12)$$

The advantage of structure of the equation (12) is that it is linear with respect to the model parameter relationships k_s/μ_{max}, $1/\mu_{max}$, $1/k_i\mu_{max}$. With the data points s_j, μ_j generated at Step 2 and the predetermined value of parameter k_x, the above parameter relationships can be directly calculated by using the LSM. The parameters calculation procedure is repeatedly performed for the consistently increasing values of $k_x$until the best fitting of model predictions to the data points (minimum sum-of-squared error) is obtained. The parameters k_s, k_i, μ_{max} are then calculated from the identified parameter relationships.

4. Estimation of the parameter Y_{xs} of the model equation (7), assuming $\kappa_3 = 0$, $\kappa_4 = 0$.

With the above assumption, the parameter Y_{xs} is estimated by using the LSM from the equation

$$q(s,x) = Y_{xs}\mu_g(s,x) \qquad (13)$$

by using the data s_j, q_j calculated at Step 2, and the corresponding data $(\mu_g)_j$ calculated from the equation (5) with the parameter k_s, k_i, μ_{max} values estimated at Step 4.

5. Improvement of the model parameter values by an iterative search algorithm.

For improvement of the parameter first approach values obtained at Steps 3, 4, the chemotaxis stochastic search algorithm [18] is applied. A criterion used in the iterative search procedure is minimization of the fitting function (sum-of-squared errors):

$$J = \min_{k_s, k_i, k_x, Y, \mu_{max}, \kappa_1 - \kappa_4} \sum_{j=1}^{N} \left(\frac{(x_j - x_{mj})}{x_{max}}\right) + \left(\frac{(s_j - s_{m,j})}{s_{max}}\right). \qquad (9)$$

where $x_{m,j}$, $s_{m,j}$ are model predictions of the data points x_j, s_j and x_{max}, s_{max} are maximum values from the data arrays.

3. Solving of the model-based optimization problem

The model-based feed-rate optimization task is solved by applying the parametric optimization approach [14] and the evolutionary programming algorithm [18]. Referring to investigation results of different shape feed-rate time-profiles used for model-based optimization of various fermentation processes [14], we applied in the parametric optimization procedure the exponential-type feed-rate time profile, which requires a modest number of parameters to be optimized and provides with good approximation of optimal feed-rate. The structure of the feed-rate control algorithm is as follows:

$$u(t) = \begin{cases} 0, & \text{if} \quad t < t_{feed\_start}, \\ \alpha\ \exp\left(\beta\left(t - t_{feed\_start}\right)\right), & \text{if} \quad t_{feed\_start} \le t \le t_{feed\_end}, \\ 0, & \text{if} \quad t > t_{feed\_end}, \end{cases} \qquad (15)$$

where t_{feed_start}, t_{feed_end}, α and β are parameters subjected to optimization.

The technological objective of optimization is to maximize the yield of biomass at the end of fed-batch cycle:

$$J = \max_{t_{feed\_start}, t_{feed\_end}, \alpha, \beta} V(t_e)x(t_e), \qquad (17)$$

where t_e is the process end time.

A typical technological restriction on the total substrate to be fed is equivalent to the restriction on the final volume of cultural liquid ($V(t_e) \le V_{max}$) and

can be easily introduced in the iterative optimization procedure.

Based on the presented methods and algorithms of the proposed model identification and the feed-rate optimization, a computation program can be developed, which provides direct calculation of an optimal feed-rate time profile for particular fermentation process from the relevant experimental data. The computation flow-chart is presented in Figure 1.

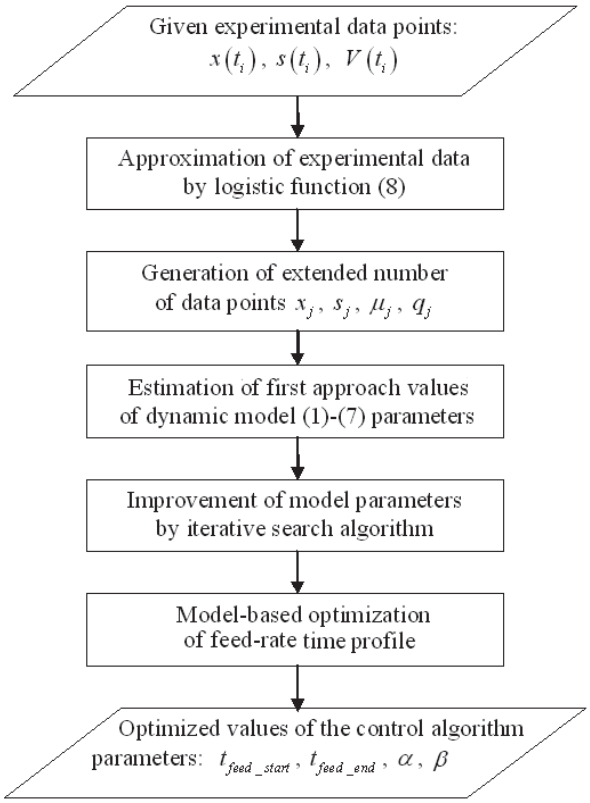

Figure 1. Flow-chart of the model-based optimal feed-rate computation procedure

4. Application example: optimization of fed-batch culture *E. coli*

Working capacities of the dynamic model identification and the model-based optimization algorithms were tested for optimization of recombinant *E. coli* fed-batch cultivation process. Experimental data points from fed-batch cultivation process are presented in Figure 2 (for the sake of confidentiality, experimental conditions and measurement units are not presented). Approximation of the experimental data with the logistic functions (8) (at power series order $m = 3$) is shown in Figure 2 by dotted lines. The parameters of the dynamic model (1)-(7) were identified from an extended number of data points ($N = 1000$) obtained from the logistic function. Initial values of process variables and restriction on final volume correspond to those in the model identification experiment: $x(t_s) = 0.104$, $s(t_s) = 8.87$, $V(t_s) = 7.0$, $V(t_e) \leq 8.8$, where t_s is process start time. The calculated values of the dynamic model parameters are given in Table 1. The process prediction by the dynamic model is shown in Figure 2 by solid lines.

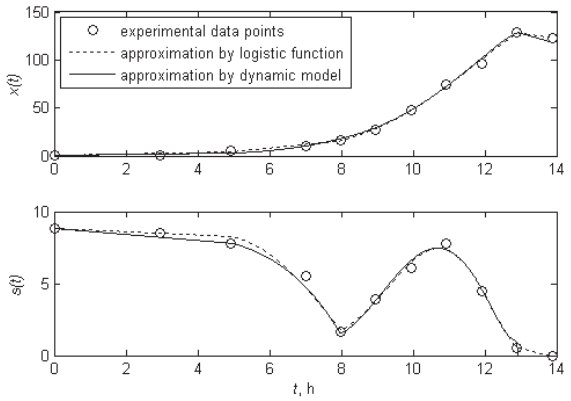

Figure 2. Identification results for the dynamic model (1)-(7)

The model identification results presented in Figure 2 demonstrate quite accurate fitting of the model to experimental data and validate practical efficiency of the model identification procedure.

The model-based process optimization was performed by applying the evolutionary programming algorithm with random settings of initial values of parameters of the control algorithm (15). The restriction was imposed on the total substrate to be fed, which corresponds to that in the model identification experiment. The calculated parameter values of the optimal control algorithm are the following: $t_{feed\_start} = 6.36h$, $t_{feed\_end} = 10.93h$, $\alpha = 0.233$, $\beta = 0.221$.

Figure 3 is representative of the model-based optimization results. The calculated optimal trajectories of the feed-rate, biomass concentration, substrate concentration and the total biomass accumulation are depicted by solid lines. For comparison, the time trajectories of the feed-rate and the total biomass at the process used for model identification are depicted by dashed lines.

Table 1. Values of dynamic model (1)-(7) parameters

Model parameters	k_s	k_i	k_x	Y_{xs}	κ_1	κ_2	κ_3	κ_4	μ_{max}
First approach values	2.95	55828	0.106	0.474	0	0	0	0	0.832
Final values	0.358	65565	0.0403	0.343	0.00483	0.000382	0.00196	0.0147	0.736

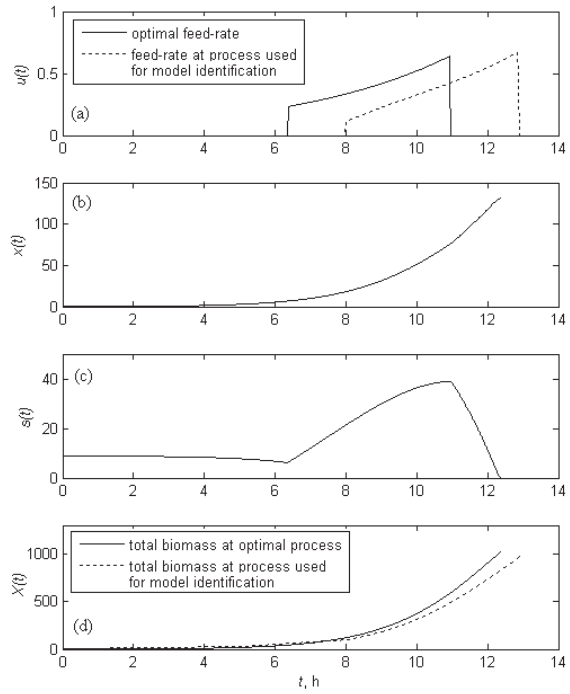

Figure 3. Time trajectories of the optimized process variables: (a)–feed-rate, (b)–biomass concentration, (c)-substrate concentration, (d)–total biomass (performance index)

As it follows from the optimization results presented in Figure 3, the optimal feed-rate control increases the yield of biomass by 3.5 % and decreases the process duration by 4.4 % compared with the results of fed-batch process used for identification of the dynamic model.

5. Conclusions

The identification procedure of fed-batch fermentation process model is developed that formalizes fitting of dynamic model of the versatile structure to particular experimental data. The steps of identification procedure include initial approximation of experimental data from batch or fed-batch fermentation processes by formal logistic relationships, estimation of the first approach values of dynamic model parameters, and iterative improvement of the parameter values. The main advantage of the proposed identification approach is formalization of the experimental data-based identification procedure and avoidance of testing kinetic models of various structures.

The developed identification and optimization procedures were tested for identification of mathematical model of particular fermentation process and model-based optimization of feed-rate time profile maximizing the yield of biomass at the end of cultivation cycle. The investigation results validate practical efficiency of the model identification and model-based optimization approach.

The presented model-based feed-rate optimization approach allows development a user-friendly computation program, which provides direct calculation of optimal feed-rate time profiles for particular fermentation processes from the relevant experimental data.

References

[1] **B. J. Minihane, D. E. Brown.** Fed-batch culture technology. In: *Biotechnology advances*, 1986, Vol. 4, No. 2, pp. 207-218.

[2] **H. C. Lim, K. S. Lee.** Control of bioreactor systems. In: *Biotechnology*, *Measurement, modeling and control*. *Schugerl* (*Ed*.), *VCH*, *Weinheim*, 1996, *Vol*. 4, pp. 506-560

[3] **J. M. Modak, H. C. Lim, Y. I. Tayeb.** General characteristics of optimal feed-rate profiles for various fed-batch fermentation processes. In: *Biotechnology and Bioengineering*, 1986, Vol. 28, pp. 1396-1407.

[4] **H. S. Shin, H. C. Lim.** Cell-mass maximization in fed-batch cultures. *Bioprocess Biosyst. Eng.*, 2006, Vol. 29, pp. 335-347.

[5] **H. S. Shin, H. C. Lim.** Maximization of metabolite in fed-batch cultures sufficient conditions for singular arc and optimal feed rate profiles. In: *Biochemical Engineering Journal*, 2007, Vol. 37, No. 1, pp. 62-74.

[6] **H. S. Shin, H. C. Lim.** Optimal fed-batch operation of recombinant cells subject to plasmid instability and death. In: *Bioprocess and Biosystems Engineering*, 2008, Vol. 31, No. 6, pp. 655-665.

[7] **W. Jin, D. D. Fan, Y. Luo, L. Chi, J. Q. Guo, G. P. Zhang, C. H. Zhu**. Optimizing the feeding operation of recombinant *Escherichia coli* during fed-batch cultivation based on Pontryagin's minimum principle. In: *African Journal of Biotechnology*, 2011, Vol. 10, No. 64, pp. 14143-14154.

[8] **S. Geethalakshmi, S. Narendran, N. Pappa, S. Ramalingam**. Development of a hybrid neural network model to predict feeding method in fed-batch cultivation for enhanced recombinant streptokinase productivity in *Escherichia coli*. In: *Journal of Chemical Technology and Biotechnology,* 2012, Vol. 87, No. 2, pp. 280-285.

[9] **P. R. Patnaik.** Design considerations in hybrid neural optimization of fed-batch fermentation for PHB production by ralstonia eutropha. In: *Food and Bioprocess Technology*, 2010, Vol. 3, No. 2, pp. 213-225.

[10] **D. Levišauskas, V. Galvanauskas, G. Žunda, S. Grigiškis.** Model-based optimization of biosurfatctant production in fed-batch culture In: *Azotobacter vinelandii*. *Biotechnology Letters*, 2004, Vol. 26, No. 14, p. 1141-1146.

[11] **A. Arpornwichanop, N. Shomchoam.** Studies on optimal control approach in a fed-batch fermentation. In: *Korean J. Chem. Eng*., 2007, Vol. 24, pp. 11-15.

[12] **M. Ronen, Y. Shabtai, H. Guterman.** Optimization of feeding profile for a fed-batch bioreactor by an evolutionary algorithm. In: *Journal of Biotechnology*, 2002, Vol. 97, No. 3, pp. 253-263.

[13] **C. Y. Liu.** Optimal control for nonlinear dynamical system of microbial fed-batch culture. In: *Journal of Computational and Applied Mathematics*, 2009, Vol. 232, No. 2, pp. 252-261.

[14] **D. Levišauskas, T. Tekorius.** Model-based optimization of fed-batch fermentation processes using predetermined type feed-rate time profiles: a comparative study. In: *Information Technology and Control*, 2005, Vol. 34, No. 3, pp. 231-236.

[15] **J. Nielsen, J. Villadsen, G. Liden.** Bioreaction Engineering Principles. In: *Kluwer Academic / Plenum Publishers*, *New York*, 2003.

[16] **V. V. Biriukov, V. M. Kantere**. Optimization of batch processes of microbiological synthesis. *Nauka, Moscow*, 1985 (in Russian).

[17] R. Oliveira, R. Simutis, S. F. Azevedo, A. Lubbert. HYBNET, an advanced tool for process optimization and control. In: *Proceedings of 7th IFAC International Conference on Computer Applications in Biotechnology (CAB7 1998)*, Osaka, Japan, May 31-June 4, 1998, pp. 289-294.

[18] **V. W. Porto, D. B. Fogel, L. J. Fogel.** Alternative neural network training methods. In: *IEEE Expert*, 1995, Vol. 10, No. 3, pp. 16-21.

[19] **D. C. Montgomery.** Design and Analysis of Experiments. *John Wiley&Sons*, 2001.

Received July 2011.