International Journal of Biological, Life and Agricultural Sciences
ISSN: 2415-6612
Vol:8, No:9, 2014

State Estimation of a Biotechnological Process Using Extended Kalman Filter and Particle Filter

R. Simutis, V. Galvanauskas, D. Levisauskas, J. Repsyte, V. Grincas

***Abstract*—**This paper deals with advanced state estimation algorithms for estimation of biomass concentration and specific growth rate in a typical fed-batch biotechnological process. This biotechnological process was represented by a nonlinear mass-balance based process model. Extended Kalman Filter (EKF) and Particle Filter (PF) was used to estimate the unmeasured state variables from oxygen uptake rate (OUR) and base consumption (BC) measurements. To obtain more general results, a simplified process model was involved in EKF and PF estimation algorithms. This model doesn't require any special growth kinetic equations and could be applied for state estimation in various bioprocesses. The focus of this investigation was concentrated on the comparison of the estimation quality of the EKF and PF estimators by applying different measurement noises. The simulation results show that Particle Filter algorithm requires significantly more computation time for state estimation but gives lower estimation errors both for biomass concentration and specific growth rate. Also the tuning procedure for Particle Filter is simpler than for EKF. Consequently, Particle Filter should be preferred in real applications, especially for monitoring of industrial bioprocesses where the simplified implementation procedures are always desirable.

***Keywords*—**Biomass concentration, Extended Kalman Filter, Particle Filter, State estimation, Specific growth rate.

I. Introduction

ADVANCED control and optimization of biotechnological processes require on-line estimation of important process variables [1], [2]. Unfortunately, the key quantities of biotechnological processes state (e. g., biomass and product concentration, biomass specific growth rate, etc.) can't be measured online with sufficiently short response times. Hence the quantities must be determined indirectly using soft-sensors [3], [4]. The central idea behind a soft-sensor is to use (relatively) easily accessible on-line measurements for the estimation of other important process variables. One of the most often employed techniques for realization of soft-sensors in biotechnological processes is based on Kalman filter algorithms [5], [6]. The standard Kalman filter algorithm provides optimal estimates of measured and unmeasured state variables by combining information of a linear mathematical model and on-line measurements [7]. Since most practical systems are nonlinear, the Kalman filter was extended to Extended Kalman Filter (EKF). Here the nonlinear model of the system is simply linearized around the current estimate and then the standard Kalman filter algorithm is applied [7].The EKF algorithm works well only in the operating point of the system where the first-order Taylor series linearization adequately approximates the state. For complicated nonlinear systems, a more sophisticated Unscented Kalman filter (UKF) algorithm [8], [9] was proposed to improve the quality of estimates. The UKF generates a cloud of so-called sigma points based on the current mean and covariance of the process state. Then, using the unscented transformation, the sigma points are propagated through the nonlinear system to estimate the mean and covariance of the actual state of the nonlinear system. The comparison of EKF and UKF has been discussed in papers [9]-[11]. In addition, the applications of these filters for biotechnological processes were analyzed in papers [12], [13]. Despite the fact that EKF and UKF algorithms are widely discussed in theoretical works, practical application of these state estimation methods in industrial biotechnological processes is very rare. The main reason for that is that the tuning and implementation of EKF and UKF algorithms are relative complicated and consequently disliked in practical applications. With the increasing capacity of computers involved in industrial bioprocess monitoring systems, some Monte Carlo algorithms, e. g. Particle Filter method, become attractive for state estimation in biotechnological processes. The Particle Filter could be a more practical alternative for real-time applications and process monitoring tasks, conventionally approached by Kalman filter techniques. This filter is particularly useful for complicated nonlinear systems and non-Gaussian noise when computational power is rather cheap and the sampling rate is moderate [14]. The Particle Filter algorithm is a recursive implementation of the Monte Carlo method. The method approximates the Bayesian posterior probability density function of state vector with a set of randomly chosen, weighted samples. Each sample of the state vector is referred to as a particle. Posterior density function is estimated by directly implementing the Bayesian recursion equations to the system [14]. A complete review of Particle Filter algorithms can be found in [15]. Because the implementation and tuning of the Particle Filter algorithm is simpler than the EKF or UKF, this method could be attractive for practical applications in industrial bioprocesses. This paper investigates the quality of the EKF and PF algorithms for estimation of biomass concentration and specific growth rate in a typical fed-batch biotechnological process and provides practical recommendations for implementation of these algorithms. The paper is organized as follows. In Section II, the EKF and PF techniques are presented, in Section III, the biotechnological

R.Simutis,V. Galvanauskas, D. Levisauskas, J. Repsyte and V. Grincas are with the Automation Department, Kaunas University of Technology, Kaunas, Lithuania (phone: +370-37-300261; e-mail: Rimvydas.Simutis@ktu.lt, Vytautas.Galvanauskas@ktu.lt, Donatas.Levisauskas@ktu.lt, Jolanta.Repsyte@ktu.lt, Vykantas.Grincas@ktu.lt).

process models and measurement equations are introduced, in Section IV, the state estimation algorithms are applied for the fed-batch biotechnological process and the estimation quality of the EKF and PF are discussed. Finally, some conclusions are drawn in Section V.

II. State Estimation Algorithms

A. Extended Kalman Filter

To formulate the state estimation algorithms the authors assume that a biotechnological process could be represented by the following nonlinear stochastic model:

$$x_k = f(x_{k-1}, u_{k-1}) + w_{k-1} \quad (1)$$

$$y_k = h(x_k) + v_k \quad (2)$$

where x_k is state vector, u_k is control input, vector y_k is measurement vector and w_k, and v_k are process and measurement noise vectors, respectively. Those noises are uncorrelated Gaussian white noise - $w_k=N(0,Q_k)$, $v_k=N(0,R_k)$ and Q_k, R_k are assumed to be known. The EKF algorithm is realized using the following set of equations:

Predicted state:

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}) \quad (3)$$

Predicted covariance estimate:

$$P_{k|k-1} = A_{k-1}P_{k-1|k-1}A_{k-1}^T + Q_{k-1} \quad (4)$$

Measurement residual:

$$d_k = y_k - h(\hat{x}_{k|k-1}) \quad (5)$$

Residual covariance:

$$S_k = C_k P_{k|k-1} C_k^T + R_k \quad (6)$$

Kalman gain:

$$K_k = P_{k|k-1} C_k^T S_k^{-1} \quad (7)$$

Updated state estimate:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k d_k \quad (8)$$

Updated covariance estimate:

$$P_{k|k} = (I - K_k C_k) P_{k|k-1} \quad (9)$$

The matrices A_{k-1} and C_k are defined as the following Jacobians:

$$A_{k-1} = \frac{\partial f}{\partial x}\Big]_{\hat{x}_{k-1|k-1}}, \; C_k = \frac{\partial g}{\partial x}\Big]_{\hat{x}_{k|k-1}} \quad (10)$$

B. Particle Filter

Particle Filter approximates the Bayesian posterior probability density function (pdf) of state variables with a set of randomly chosen, weighted samples. The samples are generated using the process model equation. A sufficiently large number of the samples guarantee convergence to the true pdf of the state. The basic steps of the Particle Filter technique can be summarized as follows [13]:

1) *Initialization:* Generate initial population of the initial state vectors. Each sample of the state vector is referred to as a particle.

$$x_0^i = x_0 + N(0, Q_0), i = 1, \ldots, M \quad (11)$$

Assuming that all particles (state vectors) are equally likely at the start of the algorithm, set the weights of the particles to

$$W_0^i = \frac{1}{M}, i = 1, \ldots, M \quad (12)$$

2) *Measurement update:* Determine the contribution-weights of each particle in estimation procedure by the likelihood equation

$$W_k^i = W_{k-1}^i p(y_k \mid x_k^i) = W_{k-1}^i p_{e_k}(y_k - h(x_k^i)) \quad (13)$$

normalize to $W_k^i := W_k^i / \sum_i W_k^i$ and then compute the estimate

$$\hat{x}_k \approx \sum_{i=1}^M W_i^k x_k^i \quad (14)$$

where $p_{e_k}\big((y_k - h(x_k^i)) \mid 0, R_k\big)$ is normal distribution density functionof measurement errors.

3) *Re-sampling*: Re-sampling is used to eliminate particles with small weights and to stimulate particles with large weights. In this application re-sampling is implemented as follows:
   a) Take M samples with replacement from the existing particle set $\{x_k^i\}_{i=1}^M$ where the probability to take the sample i is W_k^i (Sampling Importance Re-sampling, *SIR*). After re-sampling set the value $W_k^i = 1/M$.
   b) Only re-sample as above when the effective number of samples is less than a threshold M_{th} ,

$$M_{eff} = \frac{1}{\sum_i (w_k^i)^2} < M_{th} \quad (15)$$

The M_{th} is chosen as $M_{th}=2M/3$ [14], [16].

4) *Prediction*: Take a model noise $w_k=N(0,Q_k)$ and simulate new position of particles

$$x_{k+1}^i = f(x_k^i, u_k) + N^i(0, Q_k), i = 1, \ldots, M. \quad (16)$$

5) *Cycle*: Set $k = k+1$ and iterate to Step 2.

III. Bioprocess Model

For testing the quality of the proposed state estimation algorithms a "virtual bioprocess" was prepared. This bioprocess was represented by a general mass-balance

equations model. The authors use this model to simulate a typical substrate limited bioprocess behavior in recombinant protein cultivation process. Cultivation time for this process was set t=12 h. In the first phase of the cultivation (t=0-7 h), relatively high specific biomass growth rate was maintained to achieve rapid accumulation of biomass in the bioreactor. During the second process phase, the cells were induced to produce a desirable product. In this phase, the specific biomass growth rate was set to be low. By formulation of the model structure, it was assumed that the changes in biomass concentration $X(t)$, specific growth rate $\mu(t)$ and culture weight $V(t)$ can be presented by the following equations:

$$\frac{dX}{dt} = \mu(t)X - u(t)X \quad (17)$$

$$\frac{dV}{dt} = F(t) \quad (18)$$

$$\mu(t) = \begin{cases} 0.8 - 0.06t, & \text{if } t \le 5 \text{ h} \\ 0.5, & \text{if } 5 < t \le 7\ h \\ 0.5 - 0.35(t-7), & \text{if } 7 < t \le 8\ h \\ 0.15, & \text{if } t > 8\ h \end{cases} \quad (19)$$

where $u(t) = F(t)/V(t)$ is controlled dilution term related to substrate supply to the bioreactor. For the simplicity reason, it was assumed that only the substrate feed flow $F(t)$ influences the reactor weight $V(t)$.This flow was adjusted according to the desired biomass growth rate $\mu_{set}(t)$ during the cultivation and was estimated using the equation:

$$F(t) = \frac{\mu_{set}(t)\ X(t)V(t)}{Y_{xs}S_F} \quad (20)$$

where Y_{xs} is model parameter, S_F – concentration of substrate in feed flow. The oxygen uptake rate (*OUR*) and base consumption (*BC*) are conventional measurements usually carried out during the bioprocess cultivation. These measurements are correlated with the biomass concentration and specific growth rate according the following equations:

$$OUR(t) = Y_{ox}\mu(t)\ X(t) + Y_{om}X \quad (21)$$

$$BC\ (t) = Y_{bx}X(t) \quad (22)$$

where Y_{ox}, Y_{om}, Y_{bx} are model parameters. The $OUR(t)$ and $BC(t)$ measurements usually were corrupted with measurement noises. The whole process model (17-22) was used to simulate typical process behavior in recombinant protein cultivation process using *E. coli* culture. The parameter values for the bioprocess model are given in Table I. Because the process parameters usually change significantly after induction, the parameter values are different for Phase I and Phase II. Fed-batch *E. coli* cultivation process was simulated within time range t=0-12 h., start biomass concentration was X_0=0.25g/kg, and bioreactor start-weight was V_0=1.0 kg. Substrate concentration in the feed was S_F= 500 g/kg. Simulation experiments were performed using Matlab programming environment.

TABLE I
PARAMETER VALUES FOR BIOPROCESS MODEL

Parameter	Value (Phase I)	Value (Phase II)
Y_{ox}	0.8 (g/g)	0.85 (g/g)
Y_{om}	0.12 (g/(g h))	0.15(g/(g h))
Y_{bx}	0.9 (g/g)	0.95 (g/g)
Y_{xs}	0.45 (g/g)	0.30 (g/g)

For implementation of EKF and PF algorithms, a simplified process model was used. This model doesn't require any special growth kinetic equations and consequently could be applied for state estimation in various bioprocesses. The process model is presented by the following difference equations:

$$X_k = X_{k-1} + \Delta t X_{k-1}(\mu_{k-1} - u_{k-1}) + w_{1,k-1} \quad (23)$$

$$\mu_k = \mu_{k-1} + w_{2,k-1} \quad (24)$$

where $\Delta t = 0.1\ h.$, is discretization step, u_k– known dilution term, and $w_{1,k-1}$ and $w_{2,k-1}$ are uncorrelated Gaussian white noises assumed to be proportional to the value of the state variables. The measurement model is described by equations:

$$OUR_k = Y_{ox}\mu_k X_k + Y_{om}X_k + v_{1,k} \quad (25)$$

$$BC_k = Y_{bx}X_k + v_{2,k} \quad (26)$$

where $v_{1,k}$ and $v_{2,k}$ are Gaussian white noises assumed to be proportional to the value of the measurements.

IV. SIMULATION RESULTS

Quality of the EKF and PF algorithms for estimation of biomass concentration and specific growth rate was investigated using data obtained from simulated bioprocess. The simulated measurement data (*OUR* and *BC*) were collected every time interval $\Delta t = 0.1\ h.$, and were corrupted by white Gaussian noises. The intensity of measurement noise was assumed to be known and it was varied during the accomplished simulation experiments. The obtained measurements were used by EKF and PF state estimation algorithms. It was assumed that process model noises in (23)-(24) are uncorrelated Gaussian white noises with zero mean and standard deviations are σ_x=0.03X_k, σ_μ=0.15μ_k. The number of particles M for the particle filter algorithm was chosen between 200 and 1000. The measurement data and state estimation results for a typical cultivation run (relative measurement noise intensity: 0.05OUR_k and 0.05BC_k, M=500) are shown in Figs. 1 and 2.

The absolute percentage estimation errors give a better measure for evaluation of the estimation errors during the cultivation process. These errors are presented in Fig. 3. It can be observed from the Figs. 2 and 3 that the errors both for biomass concentration estimation and for specific growth estimation are gradually lower using the Particle Filter method. The quality of the state estimation methods was also evaluated using mean absolute percentage error *(MAPE)* of

estimations defined by the equation:

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{k}\left|\frac{A_i - E_i}{A_i}\right| \quad (27)$$

where A_i is the true value of state variable and E_i is the estimated value. During the simulation experiments, the relative measurement noise intensity was varied both for oxygen uptake rate and base consumption (Range: 0.02-0.1 measurement value).These investigations are shown in Fig. 4.

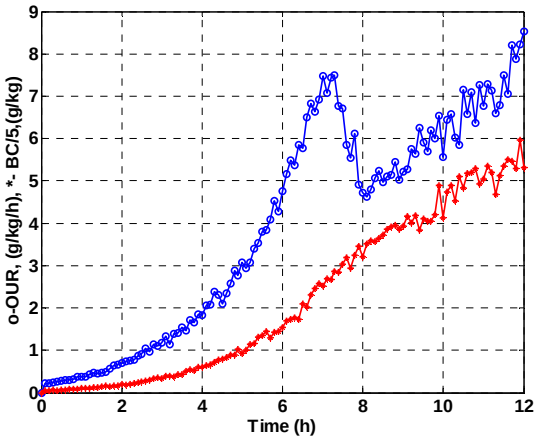

Fig. 1 Measurement data of a typical cultivation run

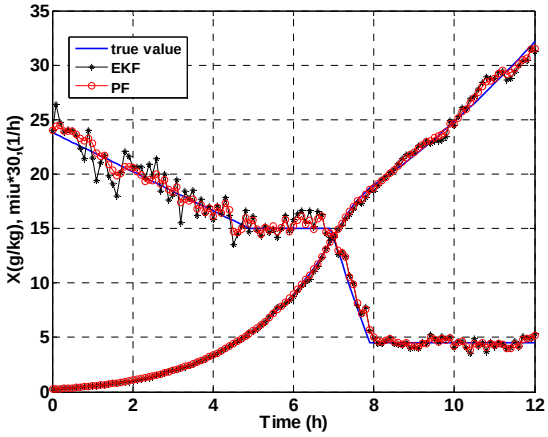

Fig. 2 State estimation results for a typical cultivation run

The average results from 10 cultivations presented in Fig.4 show that the estimation errors of the EKF and PF increase nearly proportionally while increasing of the measurement noises. Nevertheless, it is important to note that the Particle Filter gives slightly better estimation results in the whole investigated region of measurement noises. The quality of the investigated PF depends on the number of particles used in the filter. These investigations are presented in Fig. 5. It is important to note that for this application the Particle Filter becomes more accurate than the EKF when the number of particles is more than 30. The disadvantage of this filter is longer computation time needed for implementation of the filter. Table II provides relative computation time (RCT) needed for implementation of the EKF and Particle Filter algorithms together with the average estimation quality (10 cultivations) of both methods. The relative measurement noise intensity in these simulations was 0.05 OUR_k and 0.05 BC_k. As the table shows, the average estimation quality is better for the Particle Filter with more particles. Estimation quality of the Particle Filter decreases rapidly when the number of particles is lower than 200. On the other hand, the computation time for Particle Filter increases quite quickly with the increasing number of particles. Consequently, a suitable compromise between the number of particles and the estimation quality must be found when implementing the PF algorithm in practice.

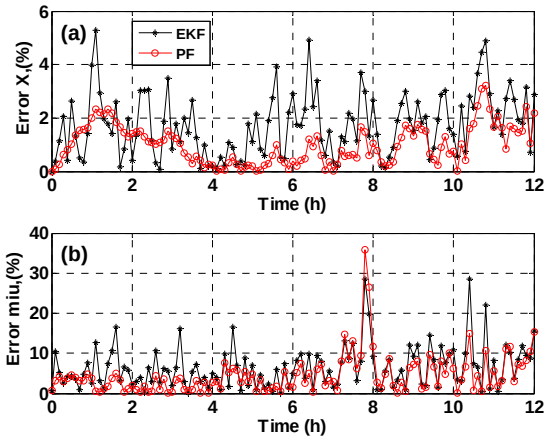

Fig. 3 (a) Percentage absolute X estimation errors (b) Specific growth rate estimation errors

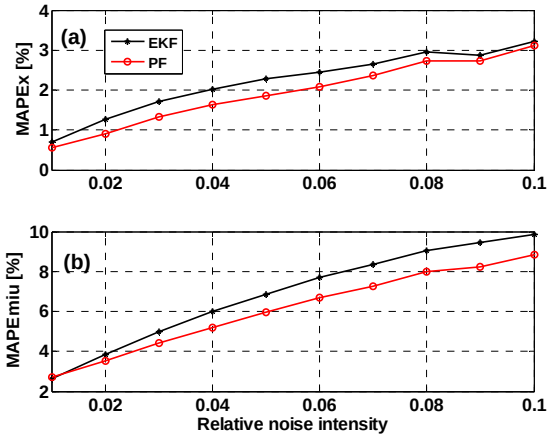

Fig. 4 (a) $MAPE_X$ (b) $MAPE_\mu$ as a function of relative noise intensity of the measurements

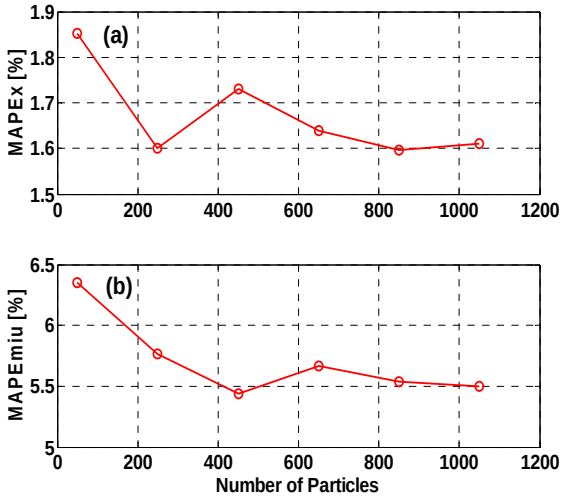

Fig. 5 (a) $MAPE_X$ (b) MAPEμ as a function of different number of particles in Particle Filter algorithm

TABLE II
COMMUTATION TIME FOR IMPLEMENTATION OF ESTIMATION ALGORITHMS

Technique	Rel.Comp.Time	$MAPE_X$	$MAPE_\mu$
EKF	1	2.4 %	6.7 %
PF(50)	5	1.9 %	6.3 %
PF(500)	14	1.7 %	5.6 %
PF(1000)	25	1.6 %	5.5 %

V. CONCLUSION

The analysis of investigated state estimation algorithms using the Extended Kalman Filter and Particle Filter has shown that the Particle Filter algorithm provides slightly better estimation results for biomass concentration and specific growth rate based on oxygen uptake rate and base consumption measurements. Also the tuning procedure of Particle Filter is much simpler in comparison with the Extended Kalman Filter. On the other hand, the computation time for Particle Filter increases quite quickly with the increasing number of particles. Consequently, the Particle Filter should be preferred for monitoring of industrial bioprocesses when sufficient computation capacity is devoted for the real-time process monitoring tasks.

ACKNOWLEDGMENT

This research was funded by a grant (No.MIP-056/2013) from the Research Council of Lithuania.

REFERENCES

[1] C.F. Mandenius, "Recent developments in the monitoring, modeling and control of biological production systems". *Bioproc. Biosyst*. Eng. 26, 2004, pp.347–351.
[2] B. Sonnleitner, "*Measurement, monitoring, modelling and control*, *in Basic Biotechnology*", 3rd ed., C. Ratledge, B. Kristiansen (eds). Cambridge University Press, 2006, pp. 251-270.
[3] P. Kadlec, B. Gabrys, S. Strandt, "Data-driven soft sensors in the process industry". *Comput. Chem. Eng*., vol. 33, 2009, pp. 795–814.
[4] R. Luttmann, D.G. Bracewell, G. Cornelissen, K. Gernaey, J. Glassey, V.C. Hass, C. Kaiser, C. Preusse, G. Striedner, C. Mandenius, "Soft sensors in bioprocessing: A status report and recommendations". *Biotechnology Journal*, vol. 7, no. 8, Sp. Iss. SI, 2012, pp. 1040-1048.
[5] G. Stephanopoulos, K.Y. San. "Studies on on-line bioreactor identification. Theory". *BiotechnolBioeng*, 26, 1984, pp. 1176-1188.
[6] "Automatic control of Bioprocesses", Ed. D.Dochain, John Willey & Sons, 2010.
[7] A.H. Jazwinski. "Stochastic Processes and Filtering Theory"; *Academic Press*: New York, NY, USA, 1970.
[8] S. Julier,J. Uhlmann, H. Durrant-Whyte. "A new approach for filtering nonlinear systems", *American Control Conference*, 1995, pp. 1628—1632.
[9] A. Romanenko, J. Castro, J. "The unscented filter as an alternative to the EKF for nonlinear state estimation: a simulation case study", *Computers and Chemical Engineering* 28, 2004, pp.347—355.
[10] E. Wan, R. van der Merwe. "The Unscented Kalman Filter". *Wiley Publishing*, 2001.
[11] F. Gustafsson, Hendeby F, Some G."Relations Between Extended and Unscented Kalman Filters," *Signal Processing, IEEE Transactions on* , vol.60, no.2,2012,pp.545-555.
[12] Q.Chai, B.Furenes,B.Lie, B. "Comparison of state estimation techniques, applied to a biological wastewater treatment process". *Proceedings of 10th International IFAC Symposium on Computer Applications in Biotechnology*, June 4--6, 2007, Cancun, Mexico, 2007, pp. 353--358.
[13] S. Schaepe, A. Kuprijanov, C.Sieblist, M. Jenzsch, R. Simutis, A. Lubbert. "Current advances in tools improving bioreactor performance". *Current Biotechnology*, 2014, v.2, p.1-12. DOI: 10.2174/2211550102666131217235246.
[14] F. Gustafsson, F. Gunnarsson, N. Bergman, J. Forssell, J.Jansson, R. Karlsson, P. Nordlund. "Particle filters for positioning, navigation and tracking", *IEEE Trans. Signal Process*., vol. 50, no. 2, 2002, pp. 425–437.
[15] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussianbayesian tracking", *IEEE Transactions on Signal Processing* 50 (2002), no. 2,p.p 174–188.
[16] A. Doucet, S.J. Godsill, and C. Andrieu. "On sequential simulation-based methods for Bayesian filtering". *Statistics and Computing*, 10(3), 2000, p.p. 197-208.