https://communications.science.ankara.edu.tr

Commun.Fac.Sci.Univ.Ank.Ser. A1 Math. Stat.
Volume 69, Number 2, Pages 1205–1214 (2020)
DOI: 10.31801/cfsuasmas.749624
ISSN 1303–5991 E-ISSN 2618-6470

Received by the editors: June 08, 2020; Accepted: July 21, 2020

AN ADAPTIVE EXTENDED KALMAN FILTERING APPROACH TO NONLINEAR DYNAMIC GENE REGULATORY NETWORKS VIA SHORT GENE EXPRESSION TIME SERIES

Levent ÖZBEK

Department of Statistics, Faculty of Science, Ankara University, Ankara, TURKEY

Abstract. Sleep spindles, which are believed to have important role of reinforcing the sleep duration, are the characteristic wave shapes that are seen in non-REM sleep stage. Detecting and analyzing the wave forms of spindles as well as determining the areas and durations of sleep spindles are quite important to understand the sleeping process thoroughly. However, the fact that spindles have temporary regime features and lower amplitudes compared to the background EEG signals makes resolving and distinguishing between them difficult. Although there have been extensive research on the decomposition of EEG signals and about the general characteristics of the spindles, the existing studies do not decompose the components in a dynamic fashion. This study takes this argument as its starting point and comes up with a methodology to detect the spindles in the sleep EEG. In particular, this study separates EEG signals into trend and cycle components via frequency analysis, where the methodology allows for system parameters and the components to be estimated simultaneously. Since the methodology allows for the parameters to vary over time, observing the time patterns of the estimated parameters have the potential to reveal further information about the sleep process.

1. Introduction

Gene regulation is one of the most amazing processes taking place in living cells. From the sequences of hundreds of thousands of genes, cells must decide which genes to express at a particular time. As the development of the cell evolves, different conditions and functions require an efficient mechanism to turn on the required genes leaving the others behind. Cells may also activate new genes to respond to the environmental changes effectively and play specific roles. The knowledge of which

2020 *Mathematics Subject Classification.* Primary 05C38, 15A15; Secondary 05A15, 15A18.
Keywords and phrases. KF, EKF, AEKF, Gene regulatory networks.

ozbek@science.ankara.edu.tr
0000-0003-1018-3114.

gene triggers a particular genetic condition may help preventing the potentially harmful effects by turning that gene off. For instance, cancer may be controlled by deactivating the gene that causes it.

Gene expression is the production of functional gene materials, e.g., mRNA. The level of gene functionality may be measured using microarrays or gene chips to produce data on gene expression. Using this data reasonably may help us to have an understanding of how the genes are interacting in a living organism.

Different genes may cooperate to produce a particular reaction while a gene may repress other genes as well. The potential benefits of gene regulation may be obtained if only a complete and accurate picture of gene interactions is available. A network specifying how different genes are interconnected may go a long way in helping us to understand the gene regulation mechanism. The control and interaction of genes may be described through a gene regulatory network.

DNA microarray technology has provided an efficient way of measuring the expression levels of thousands of genes in a single experiment on a single â€œchip.â€İ It enables the monitoring of expression levels of thousands of genes simultaneously. Measuring gene expression levels in different conditions may prove useful techniques in medical diagnosis, treatment, and drug design. In order to infer useful biological information and determine the relationships between individual genes, many research efforts have currently focused on clustering.

Recently, there has been an increasing interest of research to reconstruct models for gene regulatory networks from time series data. Obviously, choosing a good model that fits gene regulatory networks is essential to make a meaningful analysis on the expression data.

Many gene expression experiments produce short time series data with only a few time points due to its high measurement costs. The time series usually represents the dynamic response of an organism to a change in conditions, e.g., application of some drug or other treatment. Therefore, it is highly desired to extract the functional information from the data on the time series of gene expressions, and the modeling of gene expression time series has become an increasingly interesting field of research.

Since it is well known that the gene expression is an inherently stochastic phenomenon, the network should be of a â€œstochasticâ€İ nature. Recently, dynamic modeling of gene regulatory networks from time series data has received more and more research interest.

The state-space model assumes that the gene expression value depends not only on the current internal state variables but also on the external inputs, which reflects the nature of a dynamic network. Unfortunately, most results reported on state-space models have been focused on linear systems, and therefore, the non-linear phenomenon of the gene networks may not be taken into account. Most of the literature available concerning the modeling of the time series of gene expressions have not explicitly dealt with these two features, and therefore, there is a need

to seek alternative approaches to identify the parameters of a nonlinear stochastic gene regulatory network through real-time gene expression time series. In search of such an approach, EKF approach appears to be an appropriate candidate.

The traditional KF addresses the general problem of estimating the state of a discrete-time system governed by a linear stochastic difference equation. EKF linearizes about the current mean and covariance, and therefore may handle nonlinearities that may be associated either with the process model or with the observation model, or with the both. On the other hand, EKF is known as an effective recursive estimator of process variables, which may be suitable for identifying large number of parameters using a short time series.

In paper [1], the gene regulatory network is considered as a nonlinear dynamic stochastic model that consists of the gene measurement equation and the gene regulation equation. In order to reflect the reality, it is considered that the gene measurement from microarray was noisy; and it is assumed that the gene regulation equation was a nonlinear dynamic process which is autoregressively stochastic where the nonlinearity stems from the inherently non-linear regulatory relationship and the degree among genes. After specifying the model structure, they applied the EKF algorithm for identifying both the parameters of the model and the actual value of the levels of gene expression. Note that the EKF algorithm is an online estimation algorithm that may identify a large number of parameters (including parameters of nonlinear functions) through iterative procedure by using a small number of observations. Four sets of data regarding the real-world gene expression were processed to demonstrate the effectiveness of the EKF algorithm, and the obtained models are evaluated from the aspect of bioinformatics.

The EKF is extensively used in nonlinear state estimation problems. As long as the system characteristics are correctly known, EKF gives the best performance. However, when the system information is partially known or incorrect, EKF may diverge or give biased estimates. An extensive number of works has been published to improve the performance of EKF.

Many researchers have proposed the introduction of a forgetting factor, both into the KF and EKF, to improve the performance. However, there are two fundamental problems with this approach: the incorporation of the optimal forgetting factor into EKF and the selection of the optimal forgetting factor.

In paper [2–5], they proposed a new AEKF with a forgetting factor, and two methods are analyzed for the selection of the optimal forgetting factor. The stability properties of the proposed filter are also investigated. Results of the stability analysis show that the proposed filter is an exponential observer for nonlinear deterministic systems.

In this study, application of the developed model on the gene regulatory networks has been examined. With the aim of corroborating estimation method, it has been decided that the AEKF was proper for being used and malaria gene expression has been applied for the set of data on the time series. A results have been compared

with the results of the former research [1], and it has been understood that the estimation results obtained through the developed model were more preferable.

2. GENE MODEL AND PROBLEM FORMULATION

The measured gene expression levels may be modeled as

$$y_i(k) = x_i(k) + v_i(k) \quad i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, m. \tag{1}$$

where $y(k) = [y_1(k), y_2(k), \ldots, y_n(k)]^T$ is the measurement data from microarray experiments at time k with $y_i(k)$ describing the ith gene expression levels at time k, $x_i(k)$ are the actual levels of ith gene expression which stand for mRNA concentrations and/or protein concentrations at time k, $v_i(k)$ is the measurement noise, n is the number of the genes, and m is the number of the measurement time points. Here, $v(k) = [v_1(k), v_2(k), \ldots, v_n(k)]^T$ s assumed to be a zero-mean Gaussian white noise sequence with constant covariance $R > 0$, i.e., $v(k)$ $N(0, R)$. The gene regulatory network containing n genes is described by the following discrete-time nonlinear stochastic dynamical system **??**:

$$x_i(k+1) = \sum_{j=1}^{n} a_{ij} x_j(k) + \sum_{j=1}^{n} b_{ij} f_{ij}\left(x_j(k), \mu_j\right) + I_{ai} + \xi_i(k) \tag{2}$$
$$i = 1, 2, \ldots, n \quad k = 1, 2, \ldots, m-1.$$

Where $A = (a_{ij})_{nn}$ is the linear regulatory relationship and the degree among genes, $B = (b_{ij})_{n \times n}$ represents the nonlinear regulatory relationship and degree among genes; $I_0 = [I_{01}, I_{02}, \ldots, I_{0n}]^T$ is the constant vector with I_{0i} standing for the external bias on the ith gene; $\xi(k) = [\xi_1(k), \xi_2(k), \ldots, \xi_n(k)]^T \sim N(0, Q_0)$; and the nonlinear function $f_j\left(x_j, \mu_j\right)$ is given by

$$f_j\left(x_j, \mu_j\right) = \frac{1}{1 + e^{-\mu_j x_j}} \tag{3}$$

with μ_j being a parameter to be identified. Setting

$$\mu(k) = [\mu_1, \mu_2, \ldots, \mu_n]^T \tag{4}$$

and

$$f(x(k), \mu) = [f_1\left(x_1(k), \mu_1\right), f_2\left(x_2(k), \mu_2\right) \ldots, f_n\left(x_n(k), \mu_n\right)]^T \tag{5}$$

we can rewrite 1 and 2 in the following vector form:

$$x(k+1) = Ax(k) + Bf(x(k), \mu) + I_0 + \xi(k) \tag{6}$$

$$y(k) = x(k) + v(k) \tag{7}$$

Letting

$$A_e = [a_{11}, a_{21}, \ldots, a_{n1}, a_{12}, a_{22}, \ldots, a_{n2}, a_{1n}, a_{2n}, \ldots, a_{nn}]^T \tag{8}$$

$$B_e = [b_{11}, b_{21}, \ldots, b_{n1}, b_{12}, b_{22}, \ldots, b_{n2}, b_{1n}, b_{2n}, \ldots, b_m]^T \tag{9}$$

$$\mu(k) = [\mu_1, \mu_2, \ldots, \mu_n]^T \tag{10}$$

$$\theta = \left[A_e^T B_e^T \mu^T I_0^T\right]^T \tag{11}$$

all the parameters to be estimated are denoted by $\theta = \left[A_e^T B_e^T \mu^T I_0^T\right]^T$ In order to establish the gene expression model 2, it is necessary to identify the parameter vector θ. In this paper, we aim at estimating the parameters of the model2 via the AEKF method from the measurement data.

3. THE ADAPTIVE EKF APPROACH TO PARAMETER ESTIMATION

The data set is from the time series of malaria gene expression [7]. It consists of 530 genes expressed in 48 equally spaced time points. We choose the time series of expressions of the first six genes given by $z = [z_1, z_2, z_3, z_4, z_5, z_6]$

In this study, regulatory network models have been examined; and in the framework of the model 6, the AEKF proposed in [2,3] and used in the real data (Figure 7, Table 1) application studies have been conducted. Estimation results were given in Figure 2- 7 .State estimation results were given in Figure 2- 6.

In order to compare the estimated observation and the squares of actual observation values, error criterion is used and given in Figure-1. As it may be seen in Figure 1 adaptive EKF has a value of estimation more accurate than the normal EKF.

4. Conclusion

In this paper research, application of the developed model on the gene regulatory networks has been examined. With the aim of corroborating the Kalman Filter estimation method, it has been decided that the adaptive extended Kalman filter was proper for being used and malaria gene expression has been applied for the set of data on the time series. The results have been compared with the results of the former research [1] and it has been understood that the estimation results obtained through the developed model were more preferable. AEKF has a value of estimation more accurate than the normal EKF.

5. Extended Kalman Filter

The optimum linear filtering and prediction methods introduced by Kalman (1960) have been considered as one of the greatest achievements among the theories of estimation. The Kalman Filter solves the problem of estimating the instantaneous states of a linear dynamic system distorted by Gaussian white noise, using measurements that are linear functions of the system state and corrupted by additive white noise. Therefore, it is the appropriate estimation procedure for the state space systems. However, since the the simultaneous estimation of the parameters and the state problem has a nonlinear nature, the standard linear KF needs to be modified to solve such a problem. The EKF is one of the most popular estimation techniques largely investigated for state estimation of nonlinear systems. It consists

TABLE 1. Data Set.

Z=[4,314	2,271	2,789	3,788	4,162	2,208
3,2789	1,8179	2,3653	2,5943	2,9244	2,0724
1,6684	0,7923	1,4219	1,2601	0,9809	0,9977
1,7445	1,2726	1,3902	1,8115	2,1758	1,3763
1,0716	0,7282	1,068	0,9243	0,9998	0,7307
0,9868	0,5669	0,8739	0,8472	0,8891	0,4528
0,99	0,528	0,649	0,831	0,745	0,489
0,778	0,4488	0,7413	0,624	0,5897	0,5092
0,8355	0,5778	0,5219	0,9553	0,9722	0,4854
0,5796	0,3129	0,5056	0,4316	0,3823	0,3545
0,491	0,254	0,368	0,423	0,37	0,258
0,3782	0,2401	0,3691	0,2943	0,3343	0,2504
0,3446	0,2036	0,3232	0,2634	0,3019	0,2264
0,146	0,126	0,173	0,136	0,128	0,117
0,1465	0,1608	0,1002	0,128	0,1482	0,1313
0,2114	0,1577	0,1133	0,1028	0,1168	0,1554
0,2061	0,171	0,1239	0,0811	0,1129	0,1666
0,172	0,211	0,101	0,097	0,118	0,214
0,1678	0,2138	0,0642	0,0518	0,0839	0,2089
0,17	0,262	0,063	0,049	0,081	0,279
0,2155	0,3233	0,0632	0,0427	0,0948	0,2675
0,2226	0,2806	0,0655	0,0524	0,0917	0,3096
0,2101	0,3582	0,0467	0,0496	0,0995	0,3894
0,1976	0,4357	0,028	0,0469	0,1074	0,4691
0,2375	0,3711	0,0608	0,0544	0,1016	0,4062
0,2131	0,4639	0,041	0,0475	0,109	0,5582
0,253	0,641	0,044	0,075	0,128	0,592
0,1947	0,6707	0,0391	0,0707	0,1381	0,7738
0,2148	0,8082	0,085	0,1066	0,1739	0,8656
0,2349	0,9458	0,1309	0,1425	0,2098	0,9574
0,265	1,144	0,205	0,21	0,303	1,251
0,6056	1,3391	0,3874	0,5808	0,5905	1,2578
1,013	1,9144	0,9661	1,0017	0,8967	1,9266
1,4945	2,0826	1,3078	1,7174	1,6631	2,0004
1,991	2,319	1,8535	1,9343	1,7467	2,4258
2,5285	2,5555	2,493	2,2905	2,3982	2,4844
1,7578	2,9656	1,7872	2,0121	1,8186	2,8291
1,8211	2,3457	2,0033	1,9548	1,5144	2,3201
2,5851	3,3361	3,4185	4,0059	3,6226	7,6102
3,884	3,2779	4,6765	4,5845	2,8834	2,9527
3,8805	3,1208	4,7711	5,1805	3,6588	2,7262
6,0726	4,1553	6,6787	6,1378	6,9146	4,197
5,4836	2,2738	4,1907	4,4675	5,1801	2,3114
4,6334	2,0388	4,6189	4,125	4,6347	2,3628
3,2207	1,8348	2,5593	3,2643	3,9337	2,0484
1,0636	1,5575	2,3816	1,9541	2,8011	1,6607
1,561	1,9512	2,9104	2,6247	3,4341	2,0003
1,1717	1,4513	2,3003	1,9389	2,1344	1,3854]

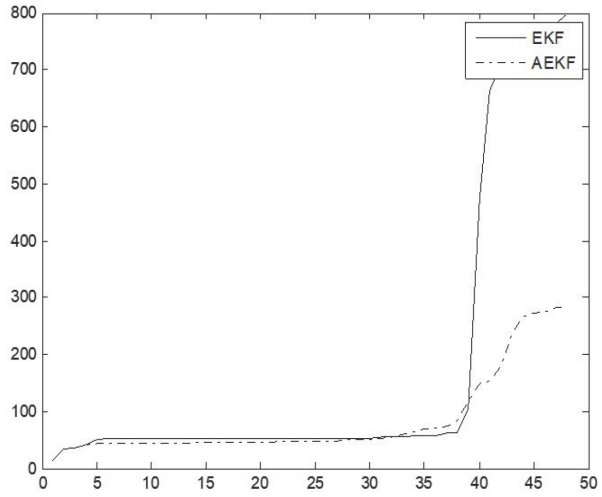

FIGURE 1. Squares Error

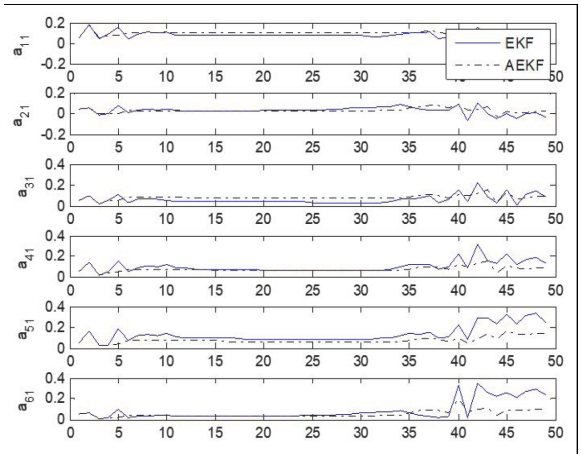

FIGURE 2. Estimation of parameters

of using the standard Kalman filter equations to the first-order approximation of the nonlinear model about the last estimate. It should also be noted that the EKF is very sensitive to its initialization and filter divergence is inevitable if the arbitrary matrices have not been chosen appropriately. [2, 8]

A non-linear state space model can be written as

$$x_t = f\left(x_{t-1}, t-1\right) + G_{t-1} w_{t-1} \tag{12}$$

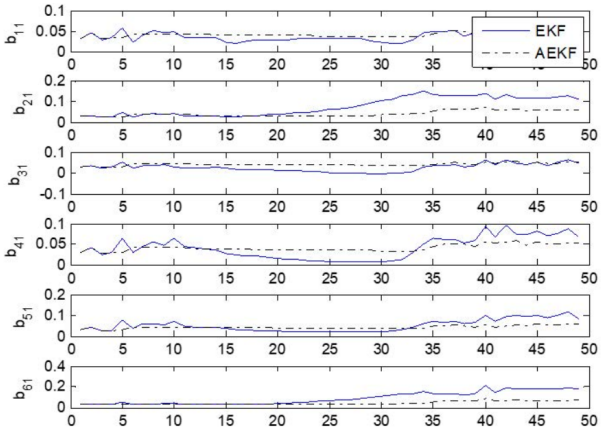

FIGURE 3. Estimation of parameters

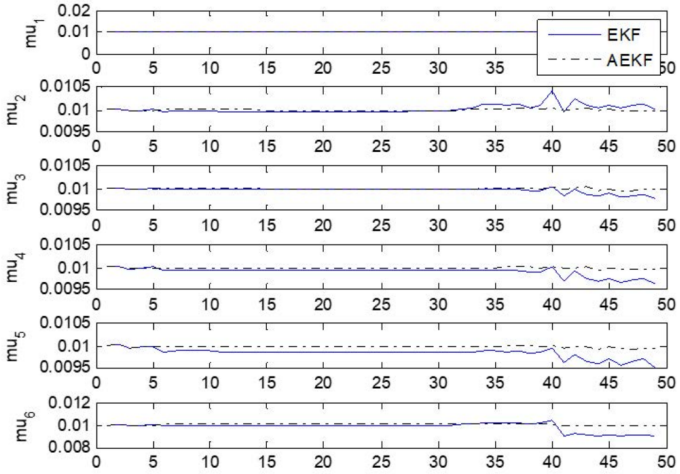

FIGURE 4. Estimation of parameters

$$y_t = h(x_t, t) + v_t \tag{13}$$

where f_t and h_t are vector-valued functions, W_t and v_t are uncorrelated zero mean white noise sequences with covariance matrix Q_t and R_t respectively. The EKF algorithm is

$$P_0 = \mathrm{Cov}(x_0) \tag{14}$$

$$\bar{x}_0 = E(x_0) \tag{15}$$

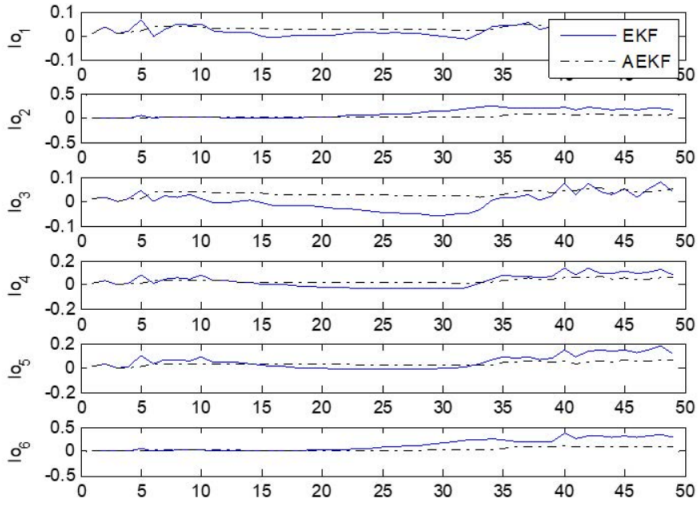

FIGURE 5. Estimation of parameters

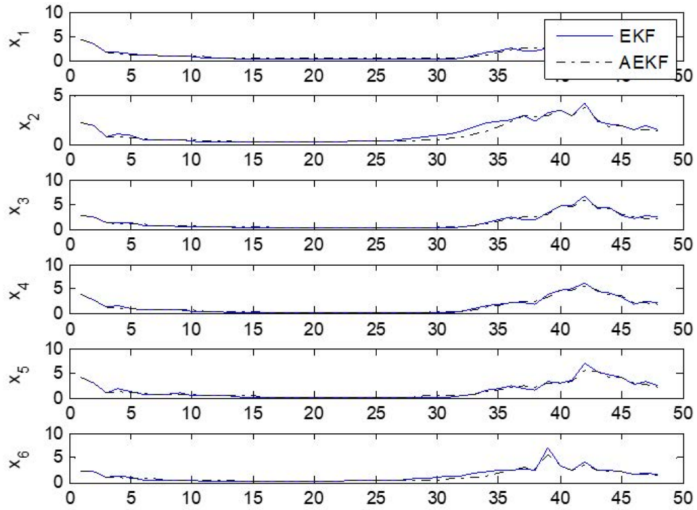

FIGURE 6. Estimation of parameters

As it is shown in [2] and [8], the updating equations are:

$$P_{t|t-1} = \alpha_t \left[\frac{\partial f_{t-1}}{\partial x_{t-1}}\left(\hat{x}_{t-1}\right)\right] P_{t-1} \left[\frac{\partial f_{t-1}}{\partial x_{t-1}}\left(\hat{x}_{t-1}\right)\right] + \alpha_t G_{t-1} Q_{t-1} G_{t-1} \tag{16}$$

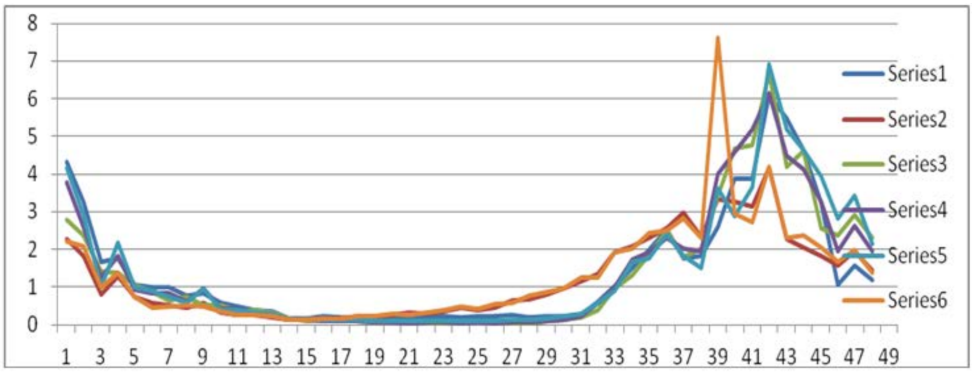

FIGURE 7. Real data

$$\hat{x}_{t|t-1} = f_{t-1}\left(\hat{x}_{t-1}\right) \tag{17}$$

$$K_k = P_{t|t-1}\left[\frac{\partial h_t}{\partial x_t}\left(\hat{x}_{t|t-1}\right)\right]\left[\left[\frac{\partial h_t}{\partial x_t}\left(\hat{x}_{t|t-1}\right)\right]P_{t|t-1}\left[\frac{\partial h_t}{\partial x_t}\left(\hat{x}_{t|t-1}\right)\right]' + R_t\right]^{-1} \tag{18}$$

$$P_t = \left[I - t_t\left[\frac{\partial h_t}{\partial x_t}\left(\hat{x}_{t|t-1}\right)\right]\right]P_{t|t-1} \tag{19}$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t\left[y_t - h_t\left(\hat{x}_{t|t-1}\right)\right] \tag{20}$$

$t = 1, 2, \ldots$

REFERENCES

[1] Wang, Z., Liu, X., Liu, Y., Liang, J., Vinciotti, V., An Extended Kalman Filtering Approach to Modeling Nonlinear Dynamic Gene Regulatory Networks via Short Gene Expression Time Series, Zidong Wang, Xiaohui Liu, Yurong Liu, Jinling Liang, and Veronica Vinciotti, *IEEE/ACM Transactions on Computational Biology and Bioinformatic*, 6 (3) (2009), 410-419.

[2] Özbek, L, Efe, M., An Adaptive Extended Kalman Filter with Application To Compartment Models, *Communications In Statistics-Simulation and Computation*, 33 (1) (2004), 145-158.

[3] Özbek, L., Aliev, F.A., Comments on "Adaptive Fading Kalman Filter With An Application", *Automatica*, 34 (12) (1998), 1663-1664.

[4] Jwo, D., Weng, T., An Adaptive Sensor Fusion Method with Applications in Integrated Navigation, *The Journal of Navigation*, 61 (2008), 705-721.

[5] Biçer, C., Babacan., E.K., Özbek, L., Stability of the adaptive fading extended Kalman filter with the matrix forgetting factor, *Turk. J. Elec. Eng. and Comp. Sci.*, 20 (5) (2012), 819-833.

[6] Chen, L., Aihara, K., Chaos and Asymptotical Stability in Discrete-Time Neural Networks, *Physica D: Nonlinear Phenomena*, 104 (1997), 286-325.

[7] Bozdech, Z., Llinas, M., Pulliam, B.L., Wong, E.D., Zhu, J., The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium Falciparum, *PLoS Biology*, 1 (1) (2003), 85-100.

[8] Özbek, L., Kalman Filtresi, Akademisyen Yayınevi, 2018.