Int. J. Nonlinear Anal. Appl. 12 (2021) No. 2, 485-493
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2021.5075

Kalman filter and ridge regression backpropagation algorithms

Irtefaa A. Neamah[a,*], Zainab Abdul Redha[a,b]

[a] *Faculty of Computer Science and Mathematics, University of Kufa, Iraq*
[b] *Ministry of Education, Najaf, Iraq.*

(Communicated by Madjid Eshaghi Gordji)

Abstract

The Kalman filter (KF) compare with the ridge regression backpropagation algorithm (RRBp) by conducting a numerical simulation study that relied on generating random data applicable to the KF and the RRBp in different sample sizes to determine the performance and behavior of the two methods. After implementing the simulation, the mean square error (MSE) value was calculated, which is considered a performance measure, to find out which two methods are better in making an estimation for random data.
After obtaining the results, we find that the Kalman filter has better performance, the higher the randomness and noise in generating the data, while the other algorithm is suitable for small sample sizes and where the noise ratios are lower.

Keywords: Kalman Filter, Ridge Regression, Backpropagation Algorithms, Estimation.

1. Introduction

Rudolf Emil Kalman was born on May 19 th, 1930 in Budapest, Hungary. It was 1958 that he discovered the Kalman filter for the first time. He's revolutionized the subject in 1960 and 1961 with his articles on the Kalman filter [10] is a collection of mathematical equations that, in a way that reduces the mean squared error, provides an effective (iterative) computational method for estimating the state of a process [23]. A series of statistical measurements observed over time were used that contained noise of various types, and generating unknown estimates appeared to be more reliable than those measurement-based variables [14]. The Kalman filter is used successfully in many practical

*Corresponding author
Email addresses: irtefaa.radhi@uokufa.edu.iq (Irtefaa A. Neamah), zainaba.allahibi@student.uokufa.edu.iq (Zainab Abdul Redha)

Received: January 2021 *Accepted:* March 2021

applications such as target tracking, navigation, positioning and control, and signal processing due to the optimization, simple implementation, and low computational complexity [7].
Hoerl and Kennard were the first to suggest the principle of ridge regression in 1970 [25]. is a biased estimated regression way, it is basically the modified least squares way and uses the ridge regression way in analyzing interlaced linear data. This method acquires more accurate and realistic regression coefficients by abandoning non-bias in the least squares estimation [15].is one of the most intriguing areas of research. The relevance of the problem it addresses, collinearity in the form of multiple linear regression and as well as the suitability of the technique to be easily applied in practice based on an analysis of the so called ridge trace contributed to the popularity of this subject [16]. Ridge regression (RR),One of the linear regression techniques, has been used in a variety of science and technology areas [3].
Rumelhart and McClelland in 1986 suggested the Back Propagation Neural Network (BPNN). It is belong to belongs to machine learning modeling approach which is a multi-layer feed-forward network. [26]. This network is trained using error backpropagation algorithm. It is the most utilized neural network [26]. The explain of this algorithm was mentioned by Rumelhart (1986), he displaied severla real world contexts and systematic method for training multiple (three or more) layered artificial neural networks [1]. There are several BPNN implementations on marketing, bioinformatics, medicine, engineering, among others [9].
Using of Kalman filter was a time varying process disruption in a space of building. The process disturbance refers a synthetic composite oflosses and heat gains generated by airflows and internal heat sources. Internal heat sources and airflows are challenging to calculate and estimate due to their time-lag effect and dynamic nature on the indoor environment. An estimation of Kalman filter method was used to resolve this problem. Based on virtual and real-world tests, Find out by the researcher [9]. The Kalman filter can be used to estimate time-varying process disturbances in a building space.
Recent algorithms for subspace clustering that make use of sparse or low-rank representations perform clustering by factoring in errors and noises into their objective functions. After that, the similarity matrix is solved by alternating direction method of multipliers. These algorithms are time intensive to implement due to the iterative nature of the operation. The researcher presents an a new subspace clustering algorithm based on ridge regression. Experiments on face datasets show that the suggested approach increases the precision and robustness of clustering performance. Additionally, the proposed approach reduces the computational complexity [25].
The capacity of the neural network (ANN) was studied in regular NASDAQ stock exchange rate predicting, that was investigated by both of the researchers, [24]. The back propagation algorithm was used to train some feed forward ANNs, which were then evaluated. The short-term historical stock prices , as well as the day of week, were used as inputs in this study's methodology.
In 2019,the estimation and prediction by kalman filter has been widely used in the lumped system parameters [12].
In 2020 the researcher proposed a new error variance estimator and defined its approximate properties via on ridge regression and stochastic matrix theory. The suggested estimator worked well in both uninterrupted and sparse cases, and is accurate for both low and high dimensional models [13].
The word "voice recognition" is one of the most terms in the biometric technologies. It is used to authenticate any device using voice features rather than photos. The goal of the researcher's study is to develop a method for voice recognition by comparing the speaker's voice signal to previously registered voice signals in a database using back propagation algorithm in neural networks [6].
Corona Virus Disease 2019 (COVID19) has arisen as a global medical emergency in the contemporary moment. In 2020 an article was written by researchers [20]. after numerous studies and research on

the latest data on COVID19 diffusion. After collecting data from a variety of sources, it is passed and merged into variety models of machine learning. Random Forest Ensemble Learning Technique provides a high assessment score for the tested samples. This approach allows for the identification and analysis of many Important factors that contribute to diffusion. Additionally, linear correlations between different features are plotted using the Pearson Correlation matrix's heat map. Finally, the Kalman Filter is used to estimation the Future diffusion of SARS-Cov-2 which shows good results on tested samples.
Double Feed Induction Generator (BDFIG) has broad potential for its high reliability and low maintenance costs. To achieve high tool and control modeling, BDFIG impedance, and inductors are essential. Notwithstanding this, current identification methods require knowledge of the professional structure or excitement and special criteria, or only a discretionary portion of the parameters. As a result, the researcher proposed a model for determining complete multi-layered parameters based on the background diffusion algorithm of BDFIG, which was constructed using electrical quantities as nodes and parameters as adjustable weights and used electrical quantities measured from normal processes as data. Depending on the fitting error obtained by comparing the model output with the easy-to-measure references, the BP algorithm is applied to update the weights until the error is small enough [28].
In this article, the comparison between two estimation methods will implement in simulation study to know the performance of randomness estimation.

2. Kalman Filter (KF)

Kalman Filter algorithm was first time published in Rudolf E. Kalman's paper "A New Approach to Linear Filtering and Prediction Problems" in 1960 [27]. For optimal linear filtering of static random processes, it can be said that the Kalman filter (KF) is the most popular algorithm that provides optimal estimates in the minimum mean squared error (MMSE) [4]. A discrete-time Kalman filter is built, for state estimation x_k of the discrete models ,which includes a prediction phase and an updating phase, also known as a priori estimation $\hat{x}_k^-$ and a posteriori estimation, .as for the priori state estimation, there is a measurement up to $k-1$ steps, and in the subsequent there is a measurement up to k steps. We work to estimate the initial state of $\hat{x}_0$ and the covariance P_0 as a first step as follows :

$$\hat{x}_0 = E[x_0]$$
$$P_0 = E\left[(x_0-\hat{x}_0)(x_0-\hat{x}_0)^T\right]$$

The algorithm of the Kalman filter is summarized as follows:

★ Linear Process and Measurement Models

- Process Equation: $x_k = A_{k-1}x_{k-1} + w_{k-1}$
- Measurement Equation: $y_k = H_k x_k + v_k,$

where,
A: is the transition matrix taking the state x_{k-1} from time $k-1$ to time k.
w_k: The process noise is assumed to be additive, white, and Gaussian, with zero mean and with covariance matrix Q, i.e., $w_k \sim N(0, Q)$.
y_k: is the observable at time k.
H: is the measurement matrix.
v_k: The measurement noise is assumed to be additive, white, and Gaussian, with zero mean and with covariance matrix R, i.e., $v_k \sim N(0, R)$.

★ Prediction phase:

- Predicted state estimate $\hat{x}_k^- = A_{k-1}\hat{x}_{k-1}^+$
- Predicted error covariance $P_k^- = A_{k-1}P_{k-1}^+A_{k-1}^T + Q_{k-1}$

★ Update phase:

- Kalman gain $K_k = P_k^- H_k^T \left[H_k P_k^- H_k^T + R_k\right]^{-1}$
- Updated state estimate $\hat{x}_k = \hat{x}_k^- + K_k \left[y_k - H_k \hat{x}_k^-\right]$
- Updated error covariance $P_k = [I - K_k H_k] P_k^-$ [24, 4, 5]

3. Ridge Regression (RR)

The way to overcome the presence multicollinearity is called Ridge Regression (RR). This method works on the tying the constant P with the value of estimated parameters, wher P is a finite positive. Using this principle, it is equivalent to adding a constant k to the diagonal of the $X'X$ matrix. Ridge constant is the name given to this constant k. Ridge regression produces biased estimates based on the constant k. The modification of the ordinary Least Squares approach (OLS). This modification enables the coefficients of regression for the biased estimators. Instead of focusing on smaller MSE of regression coefficients, it must be used unbiased estimator to obtain the closest values of estimated parameters. Consider the following standard multiple linear regression model:

$$Y = X\beta + \varepsilon \tag{3.1}$$

Where, Y is vector $(n\times 1)$ of the dependent variable values ,X is a matrix with size $(n\times p)$, P predictor variables values, and this matrix $(1 \times p)$, β is vactor $(p \times 1)$ of unknown regression coefficients, and ε is vactor $(n \times 1)$ of random variables.
Let

$$\hat{\beta}_{Rid} = (X'X + KI)^{-1}.X'Y \tag{3.2}$$

be the estimator of ridge, then for any $k \geq 0$
Where,I is identity, and K value that can be calculated using the following formula:

$$K = \frac{ps_e^2}{\hat{\beta'}_{ols}\hat{\beta}_{ols}} \tag{3.3}$$

Where $\hat{\beta}_{ols}$ is the vector of estimators by (OLS)

$$\hat{\beta}_{ols} = (X'X)^{-1}.X'Y \tag{3.4}$$

and

$$s_e^2 = \frac{\sum_i^n e_i^2}{n-p} \tag{3.5}$$

Where, e^2 is the MSE by (OLS) [8, 18].

4. Backpropagation Algorithms (BP)

is a fast and simple iterative process that typically performs well, even with complex data [11]. Backpropagation of error Backpropagation is the most widely used algorithm for training deep neural networks and is the most successful learning procedure for these networks [19]. In the strategy of back propagation algorithm, the neural network received the input data repeatedly, and for each iteration of the training process, each neural network output presentation is compared to the result required to calculate the error. This error is transmitted to the neural network and is used to adjust the weights so that the error decreases with each iteration. So, the neural model gradually approaches the desired output [2].
By means of Eq.(4.1), the input data is presented to the hidden layer h_{i_1}, where x_{i_0} represents the data for the input layer and $w_{i_0 i_1}$ is weights

$$h_{i1} = \sum_{i_0=1}^{n0} w_{i_0 i_1} x_{i0} \tag{4.1}$$

In Eq.(4.2) The sigmoid activation function is used to obtain the answer of the hidden layer v_{i_1}

$$v_{i_1} = \frac{1}{1+e^{-h_{i_1}}} \tag{4.2}$$

The output to the output layer is N_{i_2}, which is expressed in Eq.(4.3) as follows

$$N_{i_2} = \sum_{i_1=1}^{n1} w_{i_1 i_2} v_{i_1} \tag{4.3}$$

In Eq.(4.4) The sigmoid activation function is used to obtain the answer of the output layer O_{i_2}

$$O_{i_2} = \frac{1}{1+e^{-N_{i_2}}}. \tag{4.4}$$

These weights should be adjusted (updated) to get the least possible error. Using Eq.(4.5), the mean squared error is determined as follows

$$Error = \frac{1}{2}\sum_{i_2=1}^{n2} \left(O_{i_2}^{d} - O_{i2}\right)^2, \tag{4.5}$$

where the desired (target) output is $O_{i_2}^{d}$ and ANN actual output is O_{i2}.
To obtain the updated weights through the standard gradient descent approach, Eq.(4.6) is given as follows

$$w_{i_1 i_2}(new) = w_{i_1 i_2}(old) + \eta \delta_{O_{i_2}}\ v_{i_1} \tag{4.6}$$

where $w_{i_1 i_2}$ are the weights associated between the neurons of the output layer and the neurons of the hidden layer, η is the learning rate and

$$\delta_{O_{i_2}} = \left(O_{i_2}^{d} - O_{i2}\right)\ O_{i_2}\ \left(1 - O_{i2}\right)$$

It is called the back propagation error from the neurons layer of output to the hidden layer neurons within weights. The answer of hidden layer is expressed as v_{i_1}. Same process is practiced to update

the weights between input layer and hidden layer neurons.
Associated gradient descent approach is shown as Eq.(4.7)

$$w_{i_1 i_2}\,(new) = w_{i_1 i_2}\,(old) + \eta \delta_{h_{i_1}}\ x_{i_0} \tag{4.7}$$

where $w_{i_1 i_2}$ are the weights associated between the neurons of the hidden layer and the neurons of the input layer,
and $\delta_{hi_1} = v_{i_1}\,(1 - v_{i_1}) \sum_{i_2=1}^{n_2} \delta_{O_{i_2}} w_{i_1 i_2}$ it is known as the error back propagated from the hidden layer neurons to the input layer neurons through weights.

5. Ridge Regression Backpropagation Algorithm (RRBp)

This method is based on ridge regression using the back propagation algorithm. This algorithm is called the Ridge Regression Backpropagation Algorithm.
ridge regression and the back propagation algorithm are combined in this method. First and foremost, in this context, the goal is:

$$t = f(x) \tag{5.1}$$

So, to calculate the desired output $\left(\hat{t} = \hat{f}(x)\right)$, which is to say, we're going to estimate the objective. They used the back propagation algorithm technique and ridge regression approach to find output after to calculate $\hat{\beta}$ as seen in Eq.(5.3) below. The weights will be modified as follows:

$$\delta = \left(t - \hat{t}\right) + \hat{\beta} * \left(1 - \hat{t}\right) * \hat{t}, \tag{5.2}$$

where,

$$\hat{\beta} = \left(X'X + KI\right)^{-1}.X'Y \quad [18] \tag{5.3}$$

6. Implementation of Simulation

In recent times, simulation has played a fundamental role in scientific and applied studies due to its prominent effect in saving effort, time and cost. Therefore, we decided to apply this study in a digital simulation by generating random data from different sample sizes (30, 50, 100, 500). A random data control was made for the accuracy of the results. These data were used in the two methods, the Kalman filter and a modified of Backpropegation, that was proposed in a previously published research paper, it is called ridge regression backpropagation (RRBp). The number of epoch was (10000). To measure the accuracy of the performance of the two methods, mean square error (MSE) was used. The following table shows the details of the results obtained after the implementation of the simulation.

Table 1: The Mean Square Error of Kalman Filter and Ridge Regression Backpropagation Algorithm.

Sample size	MSE	
	KF	RRBp
30	0.1268	0.0013
50	0.0911	0.0027
100	0.0411	0.0064
500	0.0084	0.0368

Through Table 1. above note that the following points:

1. The mean square error of a Kalman filter decreases as the sample size increases. This means that this filter has the ability to reduce the effects of random noise.
2. In the algorithm, we notice that the mean square error begins small at the sample size (30), but it increases increasing as the sample size increases. Despite this, this error value remains acceptable for these types of study.

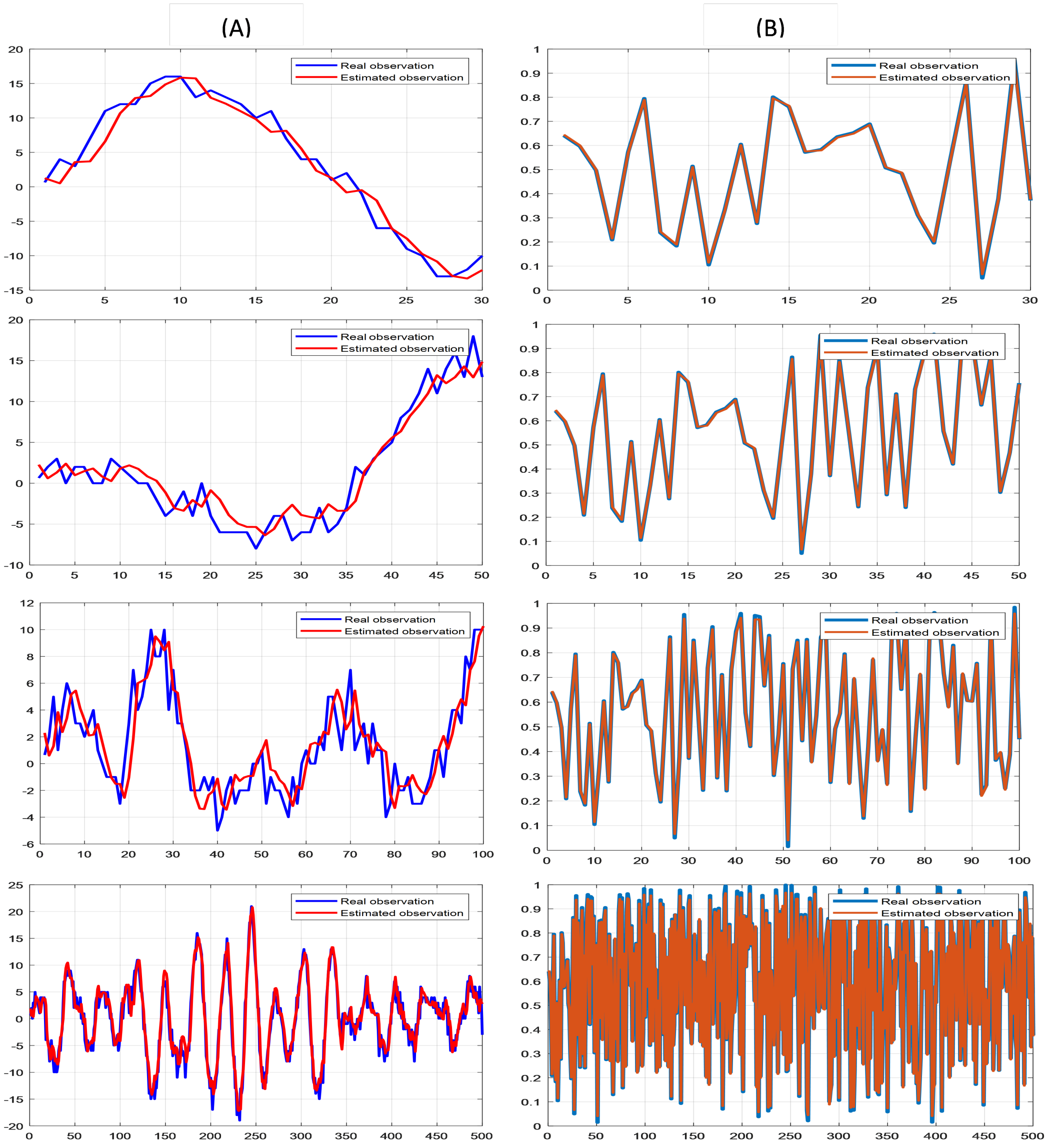

Figure 1: The Performance of Data in the Simulation with Sample Sizes (30, 50,100,500) Respectively. Column (A) Represent the KF and Coulumn (B) represent the RRBp

The above figure shows the behavior of random data estimating in the Kalman filter and the ridge regression backpropagation algorithm. The data represented with the real and estimated values for all sample sizes that were used in this study. The first column of Figure 1 has the four sample sizes plots for Kalman filter, while the second one has the plots for ridge regression backpropagation algorithm. As the first part shows the comparison between the algorithms in sample size (30), as well as for samples (50, 100 and 50), respectively. There is a clear difference between the real and the speculative data in the sample size (30) using the KF method while we do not notice this in the RRBp algorithm.

7. Conclusion

There are many estimation methods and they overlap with each other, and hybrid estimation methods have been found that combine the best of them in order to find ideal methods. In this study, the Kalman filter (KF) and the neural network method represented by the ridge regression backpropagation algorithm (RRBp) were compared through a simulation study and used of random data for different sample sizes to find out the behavior of these two methods. The conclusion of this study that the error value is the best possible in the Kalman filter when the sample size is large. While the performance of the RRBp algorithm is better at the smaller sample size.
This indicates the ability of the Kalman filter to control random noise in big data, while despite the quality of the RRBp algorithm, it loses its ability little by little with increasing randomness. This conclusion applies to the current study and similar studies.

References

[1] N. Bisoyi, H. Gupta, N. Padhy and G. Chakrapani, *Prediction of daily sediment discharge using a back propagation neural network training algorithm: A case study of the Narmada River, India*, Int. J. Sediment Res. 2018.

[2] L. Das, N. Kumar, R. Lather and P. Bhatia, *Emerging Trends in Mechanical Engineering: Select Proceedings of ICETMIE 2019*, Springer, 2021.

[3] Y. Deng, H. Zeng, Y. Jiang, G. Chen, J. Chen and L. Sun, *Ridge regression for predicting elastic moduli and hardness of calcium aluminosilicate glasses*, Materials Research Express, 2018.

[4] F. Govaers, *Introduction and Implementations of the Kalman Filter*, InTechOpen, 2018.

[5] A. Hameed, B. Karlik and M. S. Salman, *"Back-propagation Algorithm with Variable Adaptive Momentum*, Knowledge-Based Systems, 2016.

[6] S. Haykin, *Kalman Filtering and Neural Networks*, Wiley-Interscience, 2001.

[7] Y. Huang, Y. Zhang, Z. Wu, N. Li and J. Chambers, *A Novel Robust Student's t-Based Kalman Filter*, IEEE Transactions on Aerospace and Electronic Systems, 2017.

[8] N. Jeremia, S. Nurrohmah and I. Fithriani, *Robust Ridge regression to solve a multicollinearity and outlier*, J. Physics: Conference Series, 2020.

[9] D. Kim and C. Park, *Application of Kalman Filter for Estimating a Process Disturbance in a Building Space*, Sustainability, 2017.

[10] R. Kleinbauer, *Kalman Filtering Implementation with Matlab*, Universität Stuttgart, 2004.

[11] T. Lillicrap, A. Santoro, L. Marris, C. Akerman and G. Hinton, *Backpropagation and the brain*, Nature Reviews Neuroscience, 2020.

[12] X. Liu, S. Zheng and X. Feng, *Estimation of error variance via ridge regression*, Biometrika, 2020.

[13] A. Mansour, G. Salh and H. Zeen Alabdeen, *Voice recognition Using back propagation algorithm in neural networks*, Int. J. Comput. Trends Tech. 2015.

[14] H. Ma, L. Yan, Y. Xia and M. Fu, *Kalman Filtering and Information Fusion*, Springer, 2020.

[15] N. Ma, S. Zhao, Z. Sun, X. Wu and Y. Zhai, *An improved ridge regression algorithm and its application in predicting TV ratings*, Multimedia Tools Appl. 2017.

[16] G. McDonald, *Ridge Regression*, Wiley Interdisciplinary Reviews Computational Statistics (WIREs), 2009.

[17] A. Moghaddam, M. Moghaddam and M. Esfandyari, *Stock market index prediction using artificial neural network*, J. Econ. Fin. Adm. Sci. 2016.

[18] A. Muayad and A. Irtefaa, *Ridge Regression using Artificial Neural Network*, Indian J. Sci. Tech. 2016.

[19] N. Nawi, N. Zaidi, N. Hamid, M. Rehman, A. Ramli and S. Kasim, *Optimal parameter selection using three-term back propagation algorithm for data classification*, International Journal on Advanced Science, Engineering and Information Technology, 2017.

[20] K. Singh, S. Kumar, P. Dixit and M. Bajpai, *Kalman Filter Based Short Term Prediction Model for COVID-19 Spread*, Applied Intelligence, 2020.

[21] S. Setti and A. Wanto, *Analysis of backpropagation algorithm in predicting the most number of internet users in the world*, J. Online Info. (JOIN), 2018.

[22] J. Su, Y. Chen, D. Zhang and Y. Kang, *Full Parameter Identification Model based on Back Propagation Algorithm for Brushless Doubly Fed Induction Generator*, IEEE Transactions on Power Electronics, 2020.

[23] G. Welch and G. Bishop, *An Introduction to the Kalman Filter*, University of North Carolina at Chapel Hill , TR, 2006.

[24] J. Xie and S. Dubljevic, *Discrete-Time Kalman Filter Design for Linear Infinite-Dimensional Systems*, Processes, 2019.

[25] Y. Xu and X. Wu, *An affine subspace clustering algorithm based on ridge regression*, Pattern Analysis and Applications, 2016.

[26] X. Zhang, X. Chen and J. Li, *Improving Dam Seepage Prediction Using Back-Propagation Neural Network and Genetic Algorithm*, Mathematical Problems in Engineering, 2020.

[27] S. Zhao and B. Huang, *Trial-and-error or avoiding a guess? Initialization of the Kalman filter*, Automatica, 2020.

[28] K. R. *A new approach to linear filtering and prediction problems*, ASME. J. Basic Eng. 82(1) (1960) 35—45.