ICIC Express Letters ICIC International ©2020 ISSN 1881-803X
Volume **14**, Number **2**, February **2020** pp. **137–144**

THE PARSIMONIOUS AND ACCURATE CHARACTERIZATION OF AN ENERGY MANAGEMENT SYSTEM IN HYBRID VEHICLES BASED ON STATISTICAL ESTIMATION METHODS

Karin Kandananond

Faculty of Industrial Technology
Valaya Alongkorn Rajabhat University
1 Moo 20 Phaholyothin Road, Prathum Thani 13180, Thailand
karin@vru.ac.th

Received August 2019; accepted November 2019

Abstract. *The capability to efficiently and swiftly characterize the energy management system is one of the most important factors leading to the full achievement of the hybrid vehicle performance. The power train system substantially depends on the rechargeable batteries which are time dependent and have the characteristics of being non-linear. These characteristics are influenced by many factors, i.e., state of charge (SOC), vehicle speed, and supply voltage. The capability to accurately estimate the output parameter of the dynamic system is desirable. However, since the power train system itself is composed of many integrated parts, the projected relationship between inputs and outputs is possible but relies on a number of complex equations. Therefore, it might be a time-consuming and difficult job for operational engineers who have responsibilities including the diagnostics conduction when problems occur. To solve the problem, they are required to understand the complex structure of the energy management system. As a result, an alternative to the physical modeling has been introduced and these methods are statistical in nature, so they are understandable by most practitioners. The selected statistical methods in this study were linear regression, artificial neural network (ANN), and autoregressive with exogenous input (ARX), selected to determine the dynamic system of the energy management system. Three factors, SOC, supply voltage, and current were used as the inputs while the system output is the speed. The data regard the batteries, Lithium-ion (Li-ion) type, were collected and used for constructing models. The criteria for choosing the most appropriate method were based on two benchmarking criteria, accuracy and parsimony. The numerical performance of the accuracy depended on the estimation error in which a type of error measurement, mean squared error (MSE), was used. On the other hand, the selected model was parsimonious or the simplest model with the great predictive power, so the Akaike information criterion (AIC) was used to rank the model with the balance between parsimony and goodness of fit in order to avoid the overfitting. According to the results, the ARX method has outperformed both linear regression and neural network in term of MSE since its prediction error is the lowest. Afterwards, the order of autoregressive term in ARX was determined by considering the AIC, and the ARX model with $p = 1$ has been reported as the most parsimonious model compared to ARX with $p = 2$ and 3.*

Keywords: Accuracy, Artificial neural network (ANN), Autoregressive with exogenous input (ARX), Energy management system, Linear regression, Hybrid vehicles, Parsimony

1. **Introduction.** The ability to characterize the performance of a system is always desired by the practitioners in order to accurately predict the value of system output, so they can efficiently fine-tune or control the system. However, one of the dilemmas regarding the characterization is the structure of system itself, which is either black, grey, or white, so the method of system identification depends on the amount of information regarding the system. The energy management system of the power train system in hybrid vehicles

DOI: 10.24507/icicel.14.02.137

is an interesting example since the mechanical and electrical system is highly complex and needs a quick and accurate characterization for the engineers. In this case, although all information regarding the system is known, the system characterization is still difficult since there are many factors related to the system. As a result, the output estimation is not real-time and might consume a lot of time. However, on the practitioner's side, the quick estimation method with the acceptable degree of accuracy is always required, i.e., the parsimony seems to play an important role influencing the selection of the appropriate method. Therefore, the objective of this research is the benchmarking of different statistical estimation methods for predicting a system output with the decision criteria which are based on the parsimony and accuracy. A review regarding the estimation of a battery parameter, SOC, was conducted by [1], and the estimation methods used were model-based estimation methods, data-driven estimation methods, looking-up table-based methods, and Ampere-hour integral method. For the data driven estimation method, it was categorized into neural network, support vector machines, and uncertainty modeling. The results of the study led to a better operation of electric vehicles. The autoregressive with exogenous input (ARX) model was used by [2] to determine the dynamical battery model where there were following parameters as inputs: voltage output, current input, and other parameters, temperature, and state of charge (SOC). The estimated value was the discharge resistance, and the important finding was that the proposed model was robust to the uncertainty of the resistance values. A similar work was carried out by [3] who used the ARX and ARMAX method to identify the dynamic system of a servo motor. There are many research works focusing on the application of machine learning method in the estimation of system output. [4] reviewed the application of ANN to identify the system of the model predictive control (MPC) of the residential HVAC system. They also developed an algorithm called best network after multiple iterations (BNMI) which was utilized to determine the most appropriate ANN architecture. The results were applied for the better design of the MPC system. The study on the performance comparison of ARX and ANN method was performed in [5] since both ANN and ARX method were utilized to predict the time series data of natural gas consumption. According to the results, both methods have been capable of efficiently predicting the data with anomaly. Moreover, [6] used neural network and ARX model to predict the daily horizontal global solar radiation (DHGSR). The basis of this study depended on two parameters, sunshine hours and ambient temperature as the inputs. Another application of both ARX and neural network was shown in the study by [7], since these two methods were used to predict the temperature and humidity in the building. Different measures, i.e., goodness of fit, mean squared error (MSE), and mean absolute error (MAE) were utilized to assess the performance of each prediction method. [8] combined the application of ANN with ARX model in order to predict the noises generated by industrial equipment, i.e., vacuum pumps and compressors. The goal was to design the filter to reduce the noise in the system. Regarding the application of different methods, [9] estimated the discharge amount of Lithium-ion batteries by using four methods, i.e., ANN, ARX, ARMAX, and Kalman filter. Many research works focus on the characteristic study of SOC since it is an important factor affecting the battery management system of the electric vehicle. [10] used autoregressive moving average (ARMA) and neural network to determine the relationship between the battery open circuit voltage (OCV) and SOC. [11] utilized the ARX model to simulate the non-linear dynamic behavior of the battery in hybrid vehicles. The extended Kalman filter was later used to estimate the value of SOC. [12] applied the dual extended Kalman filter (DEKF) to update the ARX model in order to estimate the battery parameters, i.e., the battery open circuit voltage (OCV) while the input parameter was the SOC. Different machine learning algorithms were studied and assessed by [13]. The utilization was based on the estimation of Iron Phosphate Lithium-Ion battery in electric vehicles. These algorithms were decision trees, support vector-machines, regression models, and neural network. According to the literature, to fill a research gap, a complex scenario was picked, and different estimation methods were applied while two measures, prediction error and parsimony, were utilized to assess the performance of each method.

2. **Research Scenario.** This study has focused on the performance characterization of the dynamic system, i.e., the power train system of hybrid electric vehicle (HEV). The storage device used was the Li-ion batteries (220 V, 7 Ah). The schematic of the power train system is shown in Figure 1. As illustrated in Figure 1, the energy management system electronically controls the operation of batteries, the power control unit (AC/DC converter), generator, and engine. To conduct the experiment, a Simulink model created by [14] was used to simulate the actual mechanism of the power train, and the data was collected for the period of 1802 seconds. Three inputs were input voltage of the motor (V), battery SOC (percent) and current (A) supplied from the generator, and the output was corresponding speed (radian per second).

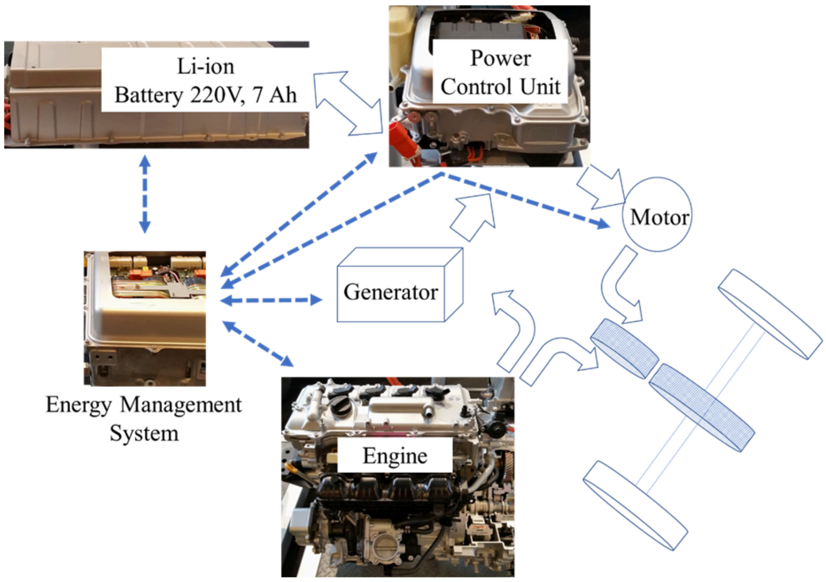

FIGURE 1. Schematic of HEV power train system

3. **Assessment of Estimation Methods.** Different methods were used to predict the motor speed of the car, and the selected method was expected to accurately determine the relationship between the designated inputs and the output. A machine learning method, artificial neural network (ANN), linear regression, and system identification method (Autoregressive with exogeneous input) were utilized to determine the relationship between inputs and output. A data set was collected and used, so the practitioners can compare the performance of different methods to predict the output of the system. Afterwards, the statistical package R was used to carry out the statistical analysis for each method as shown in Table 1.

TABLE 1. Prediction method and corresponding function in R

Prediction Method	**Function**
Autoregressive with exogenous input (ARX)	arx.ls
Multiple linear regression	lm
ANN	neuralnet

4. **Prediction Method.** Three prediction methods were utilized in this study as follows.

4.1. **ARX.** The system identification is based on the method to mathematically determine the relationship between the inputs and outputs of a dynamic system. The general equation of the system identification method is illustrated in (1).

$$y(t) = \left\{ \frac{B(q)}{A(q)} \right\} u(t) + \left\{ \frac{C(q)}{D(q)} \right\} e(t) \tag{1}$$

where $y(t)$ is the output at time t, $u(t)$ is the input at time t, $e(t)$ is the error at time t, $B(q)/A(q)$ is the relationship between output and input, $C(q)/D(q)$ is the relationship between output and noise. $A(q)$, $B(q)$, $C(q)$ and $D(q)$ are polynomial terms which are shown in the form of lagging operator: $A(q) = a_1 q^{-1} + \cdots + a_n q^{-n}$, $B(q) = b_1 q^{-1} + \cdots + b_n q^{-n}$, $C(q) = c_1 q^{-1} + \cdots + c_n q^{-n}$, and $D(q) = d_1 q^{-1} + \cdots + d_n q^{-n}$. This mathematical model is based on the time series model, and the values of the input in the lagging time, q^{-n}, is the shift operator. For example, the output signal, $q^{-1}u(t)$, is the output signal at time $t = t - 1$ or $u(t-1)$ so $q^{-n}u(t)$ is the output signal at time $t = t - n$ or $u(t-n)$. According to (1), the general equation of system identification technique is categorized into many subclasses, e.g., autoregressive (AR), output error (OE), Box-Jenkins (BJ) and autoregressive with exogenous input (ARX). The chosen model to predict the output in this research is the ARX which includes the lagging errors and inputs together. The equation of ARX is shown in (2).

$$A(q)y(t) = B(q)u(t) + e(t) \tag{2}$$

As shown in (2), it explicitly shows that ARX model is a special case of (1) with $C(q) = D(q)$. Therefore, the noise or error from the previous period has been accounted in the calculation. The diagram depicting the ARX model is illustrated in Figure 2.

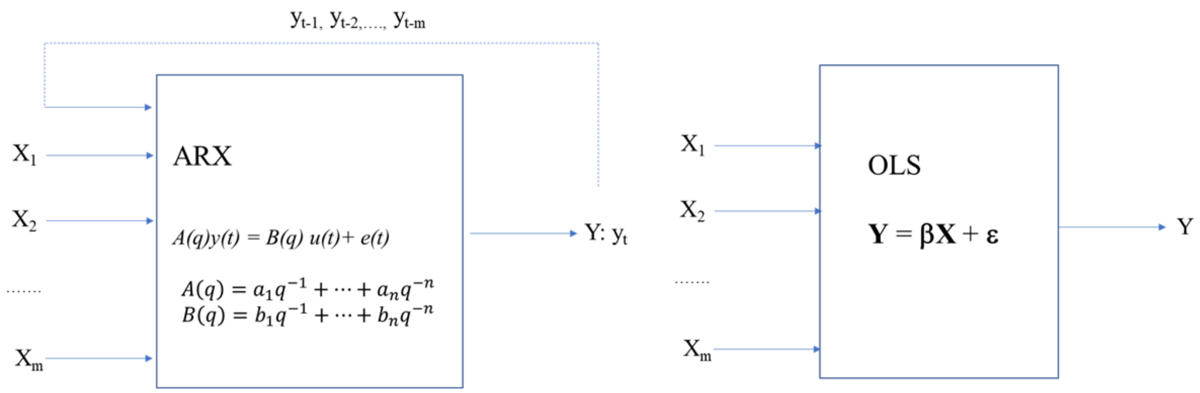

Figure 2. ARX and OLS model

4.2. **Regression analysis.** The regression analysis is the method which utilizes the ordinary least square (OLS) to estimate the unknown parameters of the system. On the other hand, the relationship between the independent variable (Y) and dependent variables (X) is analyzed by using this method. The diagram showing the mathematical system of OLS is also illustrated in Figure 2. The general equation of the model is $\mathbf{Y} = \beta\mathbf{X} + \varepsilon$, where $\beta = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$.

4.3. **Artificial neural network (ANN).** The artificial neural network is a network consisting of neurons which connect inputs to output(s). As shown in Figure 3, there are inputs (X_1, X_2, X_3) with the assigned weights ($w_1, w_2, \ldots, w_3$) which are either positive or negative values. After the values of inputs are multiplied by weights, these values are passed to the activation function in order to calculate the summation of multiplied

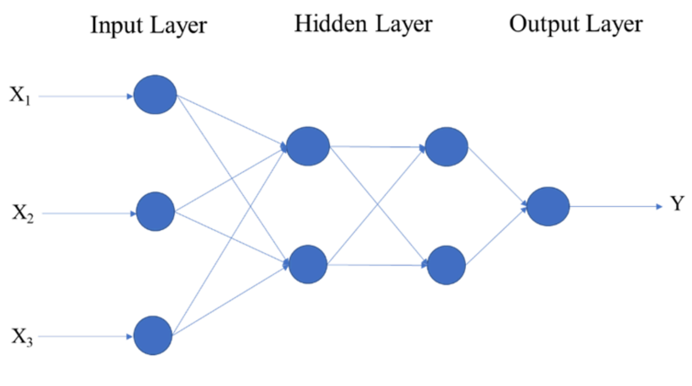

Figure 3. Example of ANN model

values: $f\{\sum(w_iX_i)\}$. The bias value may be added to the summation. The example of the artificial neural network is shown in Figure 3. In this research, the artificial neural network method is applied to the available data set for the purpose of the system identification.

5. **Results and Analysis.** After three different methods were selected and used to analyze the input and output data, the criteria for choosing the best method were decided. The performance assessment of each method was based on the prediction accuracy as well as the parsimony. There were three inputs, SOC, voltage, and current, and an output, speed. The four parameters related to the performance of hybrid vehicles were measured and shown in Figure 4.

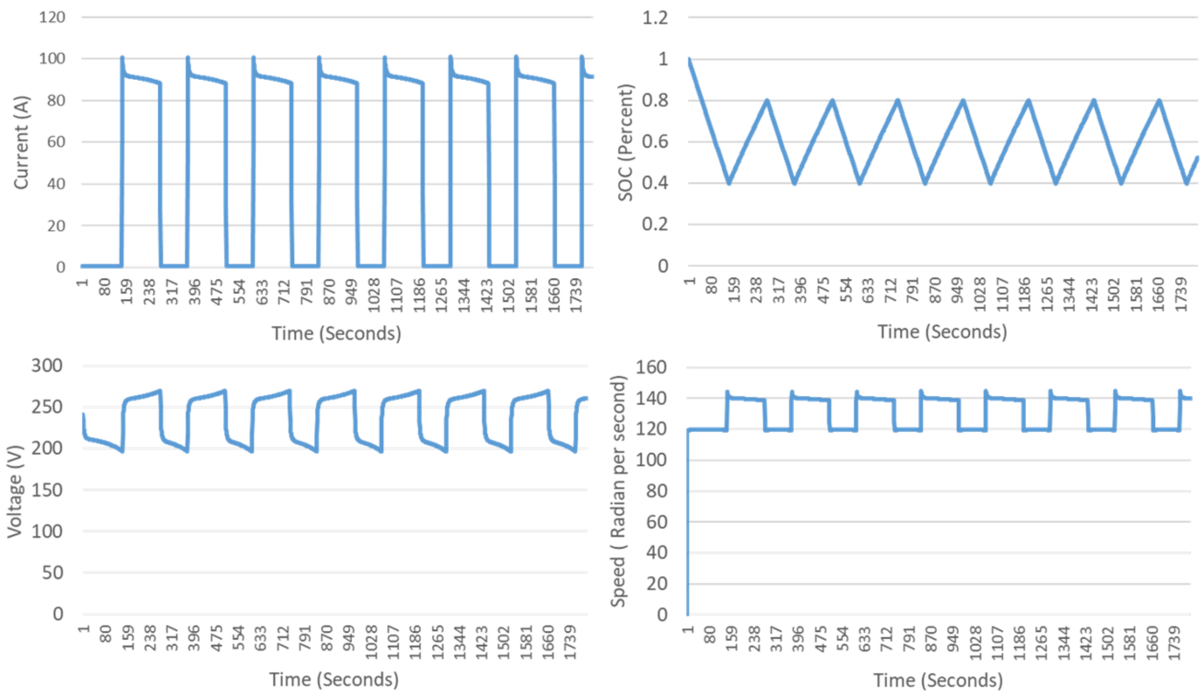

Figure 4. Inputs and output

Due to Figure 4, at the start, the battery supplied current to the motor. After the SOC of battery dropped below 40 percent, the generator was coupled with the transmission from the engine. Therefore, the battery was charged until its SOC reached 80 percent, and the torque was removed. It is interesting to note that the characteristics of the speed (radian per second) corresponds to the values of voltage (V), current (A) and SOC of batteries. Afterwards, the analysis of inputs and an output was carried out by three estimation methods as follows.

5.1. **ARX.** For the analysis, the arx.ls function in R was utilized. Three input variables and the lagged data of output were used as the source of data and the result is shown in the form of ARX model with $p = 1$ as follows: $y(t) - 0.01890211y(t-1) = 138.43059942 + 1.93805342x_{SOC}(t) - 0.08606644x_{Volt}(t) - 0.27913839x_{current}(t)$.

5.2. **Linear regression.** The linear regression model used all inputs and the output. After an R function, lm, was utilized, the results are shown in Figure 5.

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 156.75445    3.51152  44.640  < 2e-16 ***
xSOC          5.22432    1.00366   5.205 2.16e-07 ***
xvolt        -0.19636    0.01951 -10.062  < 2e-16 ***
xcurrent     -0.34516    0.01230 -28.064  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.749 on 1797 degrees of freedom
Multiple R-squared:  0.9316,    Adjusted R-squared:  0.9315
F-statistic:  8155 on 3 and 1797 DF,  p-value: < 2.2e-16
```

FIGURE 5. Linear regression results

According to the result in Figure 5, the linear regression equation is listed as follows:

$$y(t) = 156.75445 + 5.22432x_{SOC}(t) - 0.19636x_{Volt}(t) - 0.34516x_{current}(t) + \varepsilon(t).$$

5.3. **ANN.** After the neural network was applied to the data, the result is shown in Figure 6.

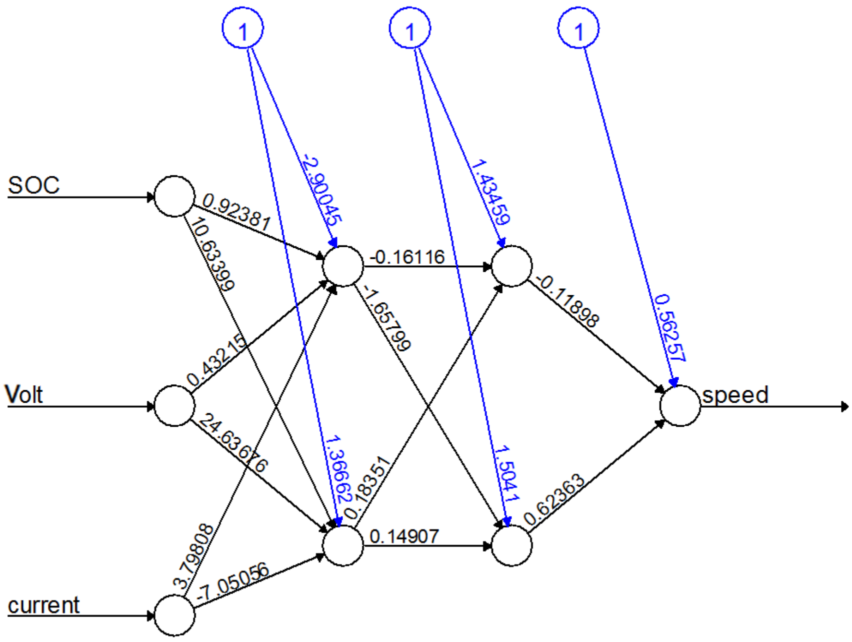

FIGURE 6. ANN results

Due to Figure 6, there are three inputs (SOC, Volt, and current) at the input layer. The connections between layers are shown in the lines assigned with the individual weight. Moreover, there are two hidden layers and each hidden layer had two nodes (two at the first layer and another two nodes at the second layer). The bias weight or the intercept is illustrated in the lines which are connected from node assigned with number 1. At the output node, there is an output which is the speed. The data was categorized into two groups (90% for training and 10% for testing), and the algorithm used was the backpropagation. The activation function used was the logistic function.

6. Benchmarking.

6.1. **MSE.** After three estimation methods were selected and utilized, the performance of each method was assessed. The prediction capability of each method was analyzed by considering the prediction errors (MSE: mean squared error) and shown in Table 2.

Table 2. Comparison of prediction errors

Criterion	**Method**		
	Linear Reg	**ANN**	**ARX**
MSE	0.3659	0.2583	$p=1$: 0.07588856, $p=2$: 0.0723863, $p=3$: 0.0674573

According to the MSE results, ARX leads to the lowest errors: 0.07588856 ($p = 1$), 0.0723863 ($p = 2$), 0.0674573 ($p = 3$) which are less than the ones from linear regression (0.3659) and ANN (0.2583).

6.2. **Parsimony.** As mentioned above, another criterion to decide the most appropriate model is the parsimony since this measure assists the practitioners to avoid the overfitting (including too many parameters in the model). Based on MSE, ARX has a better performance than the other two methods. However, the important key is the number of parameters to be included in the model, especially, ARX, which is dependent on the lagged data as the autoregressive input. If too many parameters are included in the model, the prediction model will be overfitting. As a result, the ARX models with different orders of autoregressive parameters were compared by considering the basis of parsimony, and Akaike information criterion (AIC) was the numerical parameter used to judge the best model as shown in Table 3.

Table 3. AIC comparison of different ARX models

Criterion	**The order of AR**			
	$p=0$	$p=1$	$p=2$	$p=3$
AIC	4.875094	0.2802087	0.237136	0.170794

Due to the results in Table 3, when the ARX with $p = 0$ (no autoregressive term), its AIC is equal to 4.875094. When the first order autoregressive term is added to the model ($p = 1$), the AIC significantly drops to 0.2802087. The percentage of data lost decreases by 94.25 percent after the lagged data (x_{t-1}) is included in the forecasting of the speed at time t. Since AIC represents the amount of information which is lost as well as the penalty when the parameter is added into the model, the lower AIC is, the better the model is. If the rule of thumb is considered, the approximate amount of AIC for 2 will be added as the penalized number when a parameter is included in the model. In this case, when the first order autoregressive is included in the model, the value of AIC is dropped by 4.5948853. Hence, the first order ARX substantially enhances the prediction quality. However, after the higher order of autoregressive term ($p = 2$ and 3) is included in the model, AIC slightly decreases to 0.237136 and 0.170794 consecutively. This result indicates that the ARX with $p = 1$ might be the most suitable model in term of parsimony.

7. **Conclusions and Discussions.** According to the results, the ARX model has outperformed other methods based on two criteria. For the accuracy, the prediction error of ARX has been the lowest. The distinct difference between ARX and the other models is that linear regression and ANN are mainly based on the data of three inputs. The discussion regarding this issue is that the inclusion of lagged data (autoregressive parameter) in

the ARX model might be useful for the prediction because it includes both the lagged data of the output and the input factors, so the inclusion of the previous information might help the engineers to predict the one-step ahead value. Therefore, the lagged speed data might play an important role in predicting the one-step ahead speed. However, it is interesting to note that ANN might have a better performance than ARX if a greater number of data is used to train the network. Regarding the parsimony, the different orders of autoregressive terms in ARX are put into the consideration, and their AICs indicate that ARX with $p = 1$ is not the lowest. Moreover it is not significantly different from the ones of ARX with $p = 2$ and 3. Therefore, the ARX with $p = 1$ is the most parsimonious (not overfitting with the 2nd and 3rd of autoregressive term) and also generates the accurate prediction. However, since the result of this study only applies to a system where there are three inputs, the prediction error from each method might be different if the future study is extended to a greater number of outputs (4 or more factors).

REFERENCES

[1] R. Xiong, J. Cao, Q. Yu, H. He and F. Sun, Critical review on the battery state of charge estimation methods for electric vehicles, *IEEE Access*, vol.6, pp.1832-1843, 2018.

[2] G. Giordano, V. Klass, M. Behm and G. Lindbergh, Model-based lithium-ion battery resistance estimation from electric vehicle operating data, *IEEE Trans. Vehicular Technology*, vol.67, no.5, 2018.

[3] F. Piltan, S. Tayebihaghighi and N. B. Sulaiman, Comparative study between ARX and ARMAX system identification, *International Journal of Intelligent Systems and Applications*, vol.2, pp.25-34, 2017.

[4] A. Afram, F. Janabi-Sharifi, A. S. Fung and K. Raahemifar, Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state-of-the-art review and case study of a residential HVAC system, *Energy and Buildings*, vol.141, no.15, pp.96-113, 2017.

[5] H. N. Akouemo and R. J. Povinelli, Data improving in time series using ARX and ANN models, *IEEE Trans. Power Systems*, vol.32, no.5, pp.3352-3359, 2017.

[6] S. Shamshirband, K. Mohammadi, J. Piri, D. Petković and A. Karim, Hybrid auto-regressive neural network model for estimating global solar radiation in Bandar Abbas, Iran, *Environmental Earth Sciences*, vol.75, no.2, 2016.

[7] G. Mustafaraj, G. Lowry and J. Chen, Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office, *Energy and Buildings*, vol.43, no.6, pp.1452-1460, 2011.

[8] R. S. Magalhães, C. H. O. Fontes, L. A. L. Almeida and M. Embiruçu, Identification of hybrid ARX – Neural network models for three-dimensional simulation of a vibroacoustic system, *Journal of Sound and Vibration*, vol.330, no.21, pp.5138-5150, 2011.

[9] E. F. Da Costa, D. A. De Souza, M. S. Araújo, V. P. Pinto, A. M. Peixoto and E. P. Da Costa Júnior, Estimate of discharge of lithium-ion batteries, in *New Knowledge in Information Systems and Technologies*, Á. Rocha, H. Adeli, L. Reis and S. Costanzo (eds.), Cham, Springer, 2019.

[10] X. Dang, L. Yan, H. Jiang, X. Wua and H. Sun, Open-circuit voltage-based state of charge estimation of lithium-ion power battery by combining controlled auto-regressive and moving average modeling with feedforward-feedback compensation method, *International Journal of Electrical Power & Energy Systems*, vol.90, pp.27-36, 2017.

[11] S. Yuan, H. Wu and C. Yin, State of charge estimation using the extended kalman filter for battery management systems based on the ARX battery model, *Energies*, vol.6, no.1, pp.444-470, 2013.

[12] N.-T. Tran, A. B. Khan and W. Choi, State of charge and state of health estimation of AGM VRLA batteries by employing a dual extended Kalman filter and an ARX model for online parameter estimation, *Energies*, vol.10, 2017.

[13] S. R. D. Santos, J. Aranha, T. C. D. Nascimento, D. Vieira, M. O. Eloy and F. Cerri, Study of machine learning algorithms to state of health estimation of iron phosphate lithium-ion battery used in fully electric vehicles, *SAE Technical Paper*, 2018.

[14] O. Tremblay, L.-A. Dessaint and A.-I. Dekkiche, A generic battery model for the dynamic simulation of hybrid electric vehicles, *IEEE Vehicle Power and Propulsion Conference*, Arlington, TX, USA, 2007.