A Study On The Application Of AAKR Based Early Warning System For ICE

Hyungcheol Min[1], Heesoo Kim[2], Seokman Sohn[2] Yongchae Bae[2]

[1,2]*Korea Electronic Power Corporation Korea Electronic Power Research Institute, 105 Munji-ro, Yuseong-gu, Daejeon S.Korea*

jesuishc@kepco.co.kr
eldorado@kepco.co.kr
happysohn@kepco.co.kr
ycbaenw@kepco.co.kr

ABSTRACT

Internal Combustion Engine (ICE) is a major type of power generating plant in the islands area and Emergency Diesel Generator (EDG) for nuclear power plants. All the electricity workload in the islands area are supported by diesel engine but due to harsh environmental conditions and lack of manpower, diesel engines in islands are not managed properly. The components of diesel engines are mostly affected by electrical/mechanical problems. Therefore, engines installed in islands area are vulnerable to unexpected failure and replaced earlier than the expected product life.

In this paper, we suggest an early warning algorithm that detects machine breakdown before it happens to prevent unexpected machine failure. By applying this algorithm, we expect a life extension of the diesel engine through proactive and efficient maintenance. AAKR (Auto Associative Kernel Regression) is a multivariate state signal estimation model, which is a core algorithm of the early warning system. This algorithm compares saved normal state data with acquired present state data and calculates the weight to create estimation signals.

In this research, we analyzed 2 diesel engine operation data and confirmed that the algorithm works properly. In the future, based on this research result, we plan to build an IoT (Internet of Things) based diesel engine early warning system.

1. INTRODUCTION

ICE is a major type of power generating plant in the islands area and used as EDG for nuclear power plant. A monitoring system is needed for preventing electric and mechanical failure happens in same type engines. In islands area, due to harsh condition and lack of human resources diesel engine trip happens repeatedly. The monitoring system of these islands area is an elementary level applying algorithm which is monitoring diesel engine parameters by operating engineers or checking handwritten data recording. According to statistical result (Table 1) mechanical faults and electrical faults frequently occurs in islands areas. Most of breakdown types are inappropriate managing diesel engines and lack of human resources.

Table 1. The number of ICE Faults

	2014	2015
Mechanical faults	18	17
Electrical faults	41	35

According to the above Table 1, it seems that electrical faults frequently happen compared to mechanical faults. However, electrical faults cases, a majority are relatively mild faults that can be easily repaired. On the other hand, most of mechanical problems cases are critical to damage certain components of ICE. Therefore, JBC (Jeonwoo Business Company) which is managing power plants located in islands area realized necessity of building monitoring system. JBC pushes forward automation of recording diesel engine operation data by way of showing some islands area. Many engine diagnostics research has been developed in power generation field, most of them are focusing rotating machinery in power plants, i.e. steam turbine, gas turbine or reciprocating pump. It means that we cannot utilize ICE faults type or root-cause analysis. Therefore, we decided to build monitoring algorithm based on machine learning algorithm that requires only operation signal data.

Another challenge in the diesel engine diagnostics is related with lack of training set. In machine health identification, training sets do important role classifying abnormal condition and normal condition. The importance of building monitoring system of diesel engine system rises recent date so breakdown or trip operating record data is some insufficiency for classifying data. In this case, we decided to develop diagnostics algorithm based on normal operation signal. Using normal condition signal, the machine health can be inferred by comparing simulated normal condition signal and current signal. If the pattern of acquired signal and simulated signal has a big difference, machine trip can be inferred.

Therefore, the main purpose of this paper is developing early warning system for ICE and applying algorithm to test reliability of result by analyzing past ICE breakdown records.

2. STATE OF THE ART

There are various multivariate signal estimation methods, i.e. AAMSET (Auto Associative Multivariate State Estimation Technique) by Singer (1996), AANN (Auto Associative Neural Network) by Tim Hill (1994) and AAKR (Auto Associative Kernel Regression) Nardaraya -Watson (1964). AAMSET is protected by patent related with GE (General Electrics) SmartSignal and AANN is not proper for applying diagnostics algorithm due to its complicated and long calculation time. In this paper, we choose AAKR for signal estimation algorithm to develop diesel engine monitoring system.

A test set and training set is tested for effectiveness and reliability of algorithms.

Table 2. Correlation Coefficient Matrix

	x_1	x_2	x_3	x_4
x_1	1	-0.232	0.099	0.241
x_2	-0.232	1	0.922	-0.071
x_3	0.099	0.921	1	-0.124
x_4	0.241	-0.071	-0.124	1

The correlation coefficient matrix presented in Table 1. shows that the second and third inputs are highly correlated with each other, while inputs 1 and 4 have only slight correlations with the other inputs. Note that process models with small correlation, such as these, would probably not result in a usable model for calibration monitoring. The simplified example was developed with functions that have no true dependencies and therefore have marginal performance. If more cycles of sine wave were included, the performance would be worse yet. Higher correlated data would produce significantly better results.

For completeness of the architectures of the three empirical models are as follows: (1) AAKR—500 memory vectors, bandwidth of 0.2, Euclidean distance function, and Gaussian kernel; (2) AAMSET—16 memory vectors and bandwidth of 1.0; and (3) AANN—4 mapping/demapping neurons and two bottleneck neurons. The predictions for x3 the test data of each model are presented in Figure 1, Figure 2 and Figure 3.

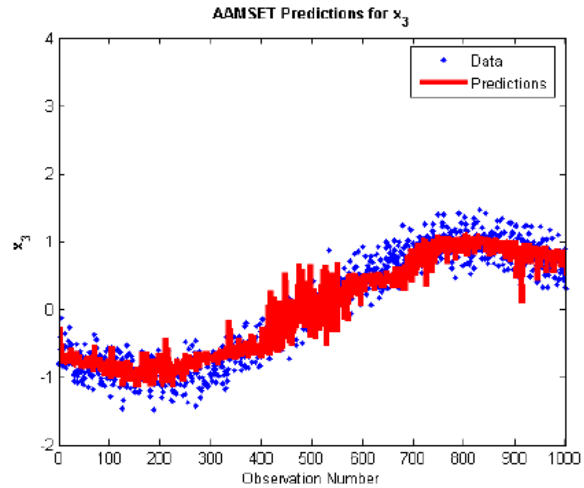

Figure 1. AAMSET Prediction

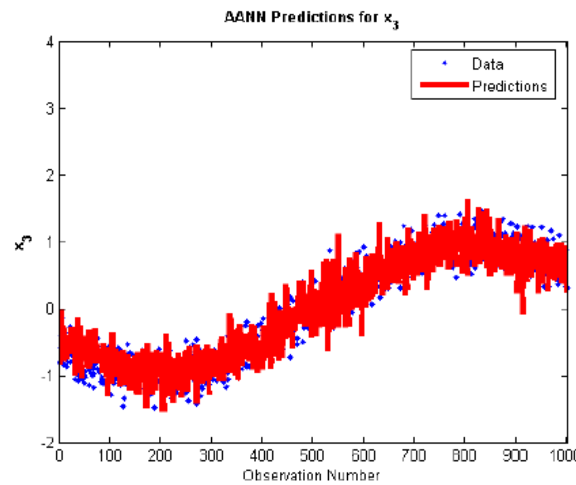

Figure 2. AANN Prediction

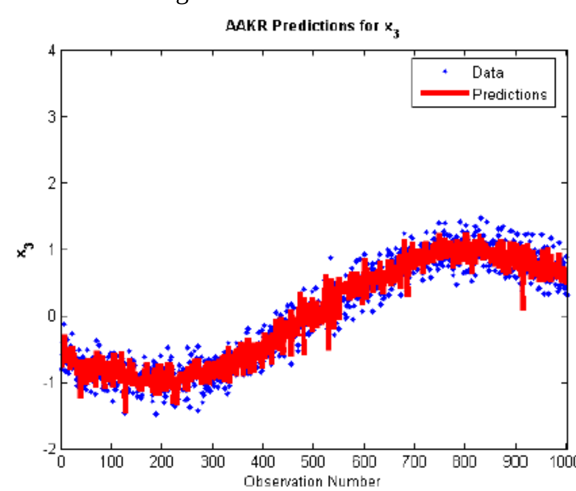

Figure 3. AAKR Prediction

Table 3. Correlation Coefficient Matrix

		x_1	x_2	x_3	x_4
AAKR	Accuracy	0.0114	0.0027	0.0153	0.0113
	Auto-Sensitivity	0.616	0.744	0.8398	0.6647
	Cross-Sensitivity	0.199	0.0729	0.2017	0.221
AAMSET	Accuracy	0.0717	0.0215	0.0409	0.1949
	Auto-Sensitivity	0.4767	0.3851	0.5366	0.4408
	Cross-Sensitivity	0.2176	0.0903	0.2871	0.2213
AANN	Accuracy	0.1311	0.0202	0.0585	0.0982
	Auto-Sensitivity	0.6138	0.4692	0.529	0.5223
	Cross-Sensitivity	0.2526	0.1277	0.4136	0.3957

It can be seen that the AAKR model has the smallest (best) accuracy metric, while the AAMSET and AANN models have the largest (worst) accuracy metric or worst accuracy performance.

Next, notice that the AAMSET model has the smallest auto-sensitivity metric, while the AAKR model has the largest (worst) auto-sensitivity metric. Again, the AANN performance is intermediate between AAKR and AAMSET. Therefore, for this example, the AAMSET model is the most robust, but sacrifices its accuracy performance. In other words, the predictions of the AAMSET model may be less accurate for good data, but less sensitive to changes in the inputs due to faults. It has a better ability to filter out faulty or corrupted signals.

Finally, notice that the cross-sensitivity performance of the models does not follow the pattern observed in the auto-sensitivity. Here, the AAKR model is best suited for estimating the true values of the other variables for a fault in one of the other variables. Also, notice that the AANN model has the worst cross-sensitivity performance, while the AAMSET model has an intermediate performance.

In this research, we chose AAKR algorithm to build early warning system due to its competitiveness of calculation speed.

3. THEORETICAL BACKGROUND OF AAKR

In statistics and empirical modeling, the process of estimating a parameter's value by calculating a weighted average of historical exemplar values is known as kernel regression Atkeson (1997). In general, kernel regression may be most compactly represented by the so-called Nardaraya -Watson (1964) estimator. Consider the simplest inferential model (i.e., one input X, one output y) whose Nardaraya-Watson estimator is given by the following equation:

$$\hat{y}(x) = \frac{\sum_{i=1}^{n}\left[K_h(X_i - x)Y\right]}{\sum_{i=1}^{n}K_h(X_i - x)} \tag{3.1}$$

Where X_i and Y_i are exemplar predictor and response values respectively; x is a query predictor vector, $K_h(X_i - x)$ is a weighting function or kernel function, which generates a weight similarity for a given difference of a query and exemplar vector, $\hat{y}(x)$ is an estimate of y, given x.

The nonparametric model operates by comparing a query x to past input examples X_i. The output is a weighted average of past examples Y_i. For example, if a query input is similar to stored inputs $X_{i=6,8,13}$, then the output is a weighted average of output examples $Y_{i=6,8,13}$. An alternative representation of Eq. (3.2) replaces the difference $X_i - x$ with a general distance function, $d_i(X_i, x)$ as follows:

$$\hat{y}(x) = \frac{\sum_{i=1}^{n}\left\{K_h\left[d_i(X_i, x)\right]Y_i\right\}}{\sum_{i=1}^{n}K_h\left[d_i(X_i, x)\right]} \tag{3.2}$$

A common distance function is the Euclidean distance, which is also known as the L^2 -norm and is given by the following equation:

$$d_i(X_i, x) = \sqrt{(X_i - x)^2} \tag{3.3}$$

More robust distance functions (i.e., erroneous data are less effective in degrading a parameter estimate) have also been investigated and include the L^1 -norm (Hines and Garvey 2005):

$$d_i(X_i, x) = |X_i - x| \tag{3.4}$$

In the one dimensional case these are equivalent but will not be when there are multiple model inputs.

Recall that the role of the kernel function is to generate a weight (similarity) for a given distance of a query vector from an exemplar vector. Therefore, it should have large values for small distances and small values for large distances. In other words, when a query vector is nearly identical to an exemplar, its distance should be small; therefore, that particular exemplar should receive a large weight. If a query vector did not resemble an exemplar, it should receive a small weight. One commonly used function that satisfies this criterion is the Gaussian kernel (Fan & Gijbels 1996):

$$K_h(d) = \frac{1}{\sqrt{2\pi h^2}} e^{\frac{-d^2}{2h^2}} \tag{3.5}$$

Here, h is commonly referred to as the kernel's bandwidth and is used to control what effective distances are deemed similar.

Next, a description of auto-associative kernel regression (AAKR) will be given. Because descriptions of AAKR do not yet appear in open literature, this derivation is based upon multivariate inferential kernel regression as derived by (Wand & Jones 1995).

In this report, the exemplar or memory vectors used to develop the empirical model are represented by the matrix X, where $X_{i,j}$ is the i th observation of the j th variable. For n_m memory vectors and p process variables, this matrix becomes:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,p} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_m,1} & X_{n_m,2} & \cdots & X_{n_m,p} \end{bmatrix} \tag{3.6}$$

Using this format, a query vector is represented by the $1 \times p$ matrix x:

$$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix} \tag{3.7}$$

The mathematical framework of this modeling technique is composed of three basic steps. First, the distance between a query vector and each of the memory vectors is computed. There are many different distance functions that may be used, but the most commonly used function is the Euclidean distance, for which the equation for the i th memory vector is as follows:

$$d_i(X_i, x) = \sqrt{\left(X_{i,1} - x_1\right)^2 + \left(X_{i,2} - x_2\right)^2 \cdots \left(X_{i,p} - x_p\right)^2} \tag{3.8}$$

For a single query vector, this calculation is repeated for each of the n_m memory vectors, resulting in an $n_m \times 1$ matrix of distances d:

$$d = \begin{bmatrix} d_1 \\ d_1 \\ \vdots \\ d_{n_m} \end{bmatrix} \tag{3.9}$$

These distances are used to determine weights by evaluating the Gaussian kernel, expressed by:

$$w = K_h(d) = \frac{1}{\sqrt{2\pi h^2}} e^{\frac{-d^2}{2h^2}} \tag{3.10}$$

where h is the kernel bandwidth; and

Finally, these weights are combined with the memory vectors to make predictions using the weighted average:

$$\hat{x} = \frac{\sum_{i=1}^{n_m} (w_1 X_i)}{\sum_{i=1}^{n_m} w_i} \tag{3.11}$$

If the scalar a is defined as the sum of the weights, that is,

$$a = \sum_{i=1}^{n_m} w_i \tag{3.12}$$

then Eqs. (3.12) can be represented by the following more compact matrix form:

$$\hat{x} = \frac{w^T X}{a} \tag{1}$$

4. ICE OPERATION DATA ANALYSIS

The experimental data represent operation data of ICE engines from power plant A, B and C.

Table 4. List of ICE

ICE	Manufacturer	Capacity
A	MAN B&W	1,900 kW
B	MAN B&W	1,900 kW

To verify early warning algorithm, ICE A mechanical faults data and power plant B had electrical faults data. All of data has been recorded by handwritten so it needs to be computerized by Microsoft Excel 2010. AAKR is programmed by MATLAB.

4.1. ICE A-Piston Damaged

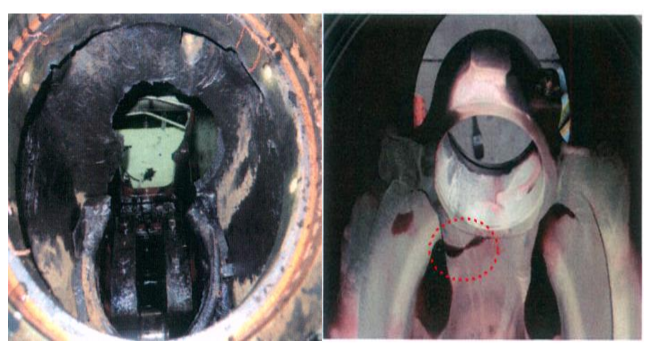

Figure 4. Piston damage of ICE A

ICE A had piston damaged in 2014.11.8. To verify early warning algorithm, 36 sensor output operation data from 2014.9.21 to 2014.10.15 is trained for AAKR learning data. From 2014.10.16 to 2014.11.8 operation data is tested for comparing estimated signal and real operation signal.

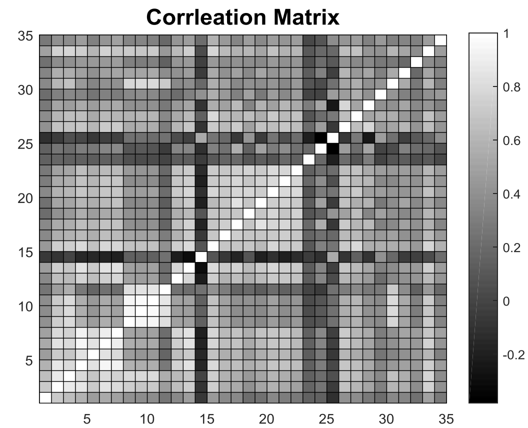

Figure 5. Correlation Plot of ICE A

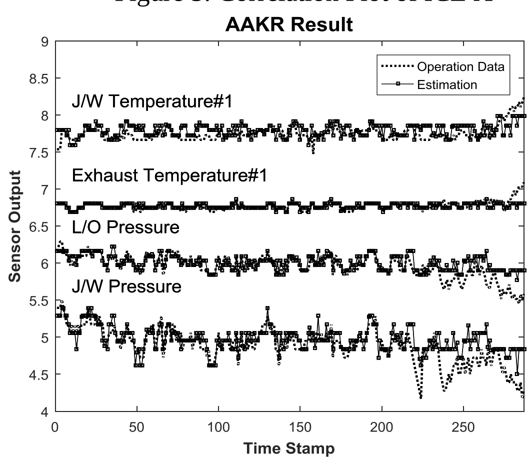

Figure 6. AAKR Result of ICE A

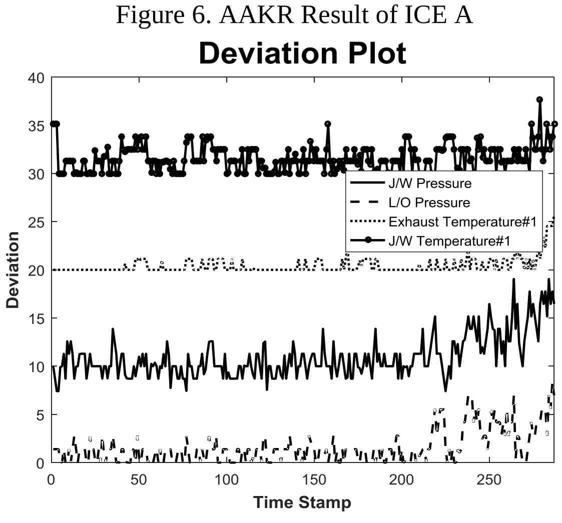

Figure7. Deviation Plot of ICE A

According to the correlation matrix, some of sensor output data can be filtered since correlation is relatively high. Based on correlation coefficient, 4 important sensor variables are selected which is jacket water pressure, lube oil pressure, exhaust temperature#1 and jacket water temperature#1. As represented in Figure 6., estimated signal by AAKR algorithm and operation data from 2014.9.21 to 2014.11.8 is plotted. At the last of x-axis which is the highest point of red line is the date when ICE A had piston damaged. As Figure 7. shows that the deviation between estimated signal and real operation data is getting bigger. If early warning system had been installed, the alarm would have noticed to operation engineers.

4.2. ICE B-AVR fault trip

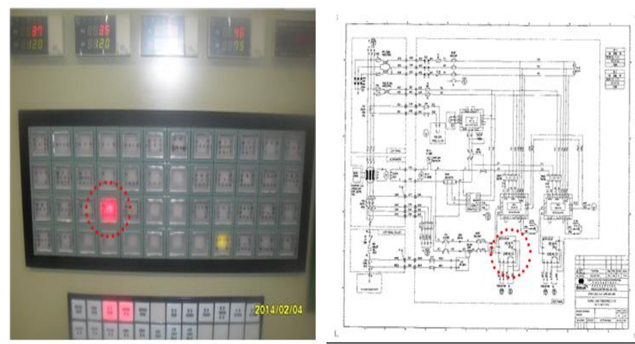

Figure 8. AVR Fault Trip of ICE B

ICE B had AVR fault trip in 2014.2.4. To verify early warning algorithm, 22 sensor output operation data from 2014.1.1 to 2014.1.16 is trained for AAKR learning data. From 2014.1.17 to 2014.2.4 operation data is tested for comparing estimated signal and real operation signal.

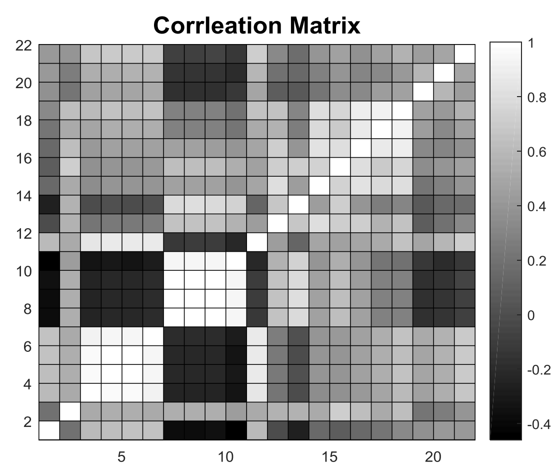

Figure 9. Correlation Plot of ICE B

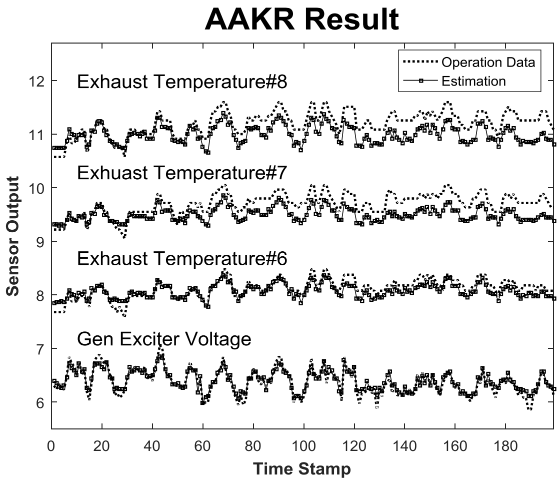

Figure 10. AAKR Result of ICE B

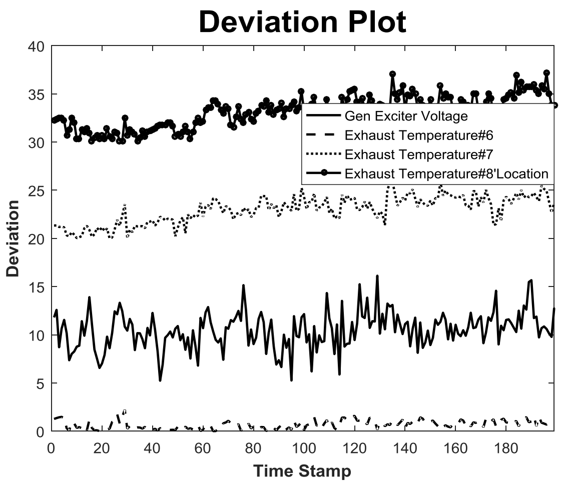

Figure 11. Deviation Plot of ICE B

According to the correlation matrix, some of sensor output data can be filtered since correlation is relatively high. Based on correlation coefficient, 4 important sensor variables are selected which is gen exciter voltage, exhaust temperature#6, exhaust temperature#7 and exhaust temperature#8. As represented in Figure 10., estimated signal by AAKR algorithm and operation data from 2014.1. 17 to 2014. 2.4 is plotted. At the last of x-axis which is the highest point of red line is the date when ICE B had AVR fault trip. As Figure 11., shows that the deviation between estimated signal and real operation data is getting bigger. Signal pattern of electrical fault is slightly different from that of mechanical problem. In further study, we will investigate more operation data from ICE and try to figure out that there is a typical signal pattern between mechanical faults and electrical faults. Again, if early warning system had been installed, the alarm would have noticed to operation engineers.

5. CONCLUSION

Table 5. The results of applying AAKR to ICE

ICE	Faults	Early Warning Alarm	Stop date
A	Piston damaged	2014.11.5	2014.11.8
B	AVR fault trip	2014.1.30	2014.2.4

We've tested 2 ICE operation data and found that early warning system worked at least 3 days to 5 days before ICE trip. In addition, the signal patterns are quite different between mechanical faults and electrical fault. In case of mechanical faults, the deviation between estimated signal and operation signal drastically increased from 3 days to 5 days before ICE fault. Electrical faults, the deviation between estimated signal and operation signal gradually increased from more than 5 days before ICE fault. The other ICE operation will be analyzed further study as operation data computerization finished.

The work is expected continued in the future. It is planned to extend the current amount of data, used to increase reliability of early warning system by analyzing various type of ICE fault problem.

Even though most of ICE operation data is recorded by handwritten, automation of operation data acquiring is planned to cover all ICE in islands area. With the installation of DAQ system in islands area, advanced ICE monitoring will be served together and also early warning system will be included in the system. The problems of the developing user-friendly program interface will be also considered.

REFERENCES

Jian-Da W., Cheng-Kai H., Yo-Wei C., Yao-Jung S. (2010). Fault diagnosis for internal combustion engines using intake manifold pressure and artificial neural network. *Expert Systems with Applications* vol. 37, pp949-958. doi: 10.1016/j.eswa.2009.05.082

Singer, R.M., Gross, K.C., Wegerich, S.Henke, D. (1996). MSET. Multivariate State Estimation Technique. USDOE Assistant Secretary for Nuclear Energy, Washington, DC (United States), ESTSC--001140MLTPL00

Tim H., Leoray M., Marcus O. C., William R.. (1994). Artificial neural network models for forecasting and decision making. *International Journal of Forecasting*. Volume 10, Issue 1, June 1994, Pages 5-15. doi: 10.1016/0169-2070(94)90045-0

Jian-Da W., Chiu-Hong L. (2009). An expert system for fault diagnosis in internal combustion engine using wavelet packet transform and neural network. *Expert*

Systems with Applications vol. 36, pp. 4278-4286. doi: 10.1016/j.eswa.2008.03.008

José M. L. , Vicente B. , Carlos G. , Ali A. (2010). A methodology for combustion detection in diesel engines through in-cylinder pressure derivative signal. *Mechanical Systems and Signal Processing* vol. 24, pp. 2261-2275. doi: 10.1016/ 10.1016/j.ymssp.2009.12.012

Jacobo P., Joaquín C., David P., José L. M.. (2011). Diesel engine condition monitoring using a multi-net neural network system with nonintrusive sensors. *Applied Thermal Engineering* vol. 31, pp. 4097-4105. doi: 10.1016/j.applthermaleng.2011.08.020

Jack L.B., Nandi A.K. (2002). Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms. *Mechanical Systems and Signal Processing* vol. 16, pp. 373-390. doi: 10.1006/mssp.2001.1454

E.A.Nadaraya. (1963). On Estimating Regression. *Theory of Probability and Its Applications* vol. 9, pp. 141-142. doi: 10.1137/1109020

BIOGRAPHIES

Hyungcheol Min received the Master degree from the Department of Mechanical Engineering, Ulsan National Institute of Science and technology, in 2014. Since 2014, he has been working with the Korea Electric Power Research Institute, Daejeon, Korea. His research interest includes signal processing, machine learning, diagnostics and prognostics of rotary machinery.

Heesoo Kim received the Ph.D. degree from the Department of Mechanical Engineering, Chonnam National University, Gwangju, Korea, in 2009. Since 1995, he has been working with the Korea Electric Power Research Institute, Daejeon, Korea. His research interests include vibration analysis/testing and mechanical integrity evaluation of generator stator winding.

Seokman Sohn received his M.S. in Dept. of Mechanical Engineering from Yonsei University in 1993 and KAIST in 1995 respectively. He completed Ph.D. in Southampton University in 2013. He has been a researcher at Korea Electric Power Research Institute from 1995. His main research interests include Monitoring, Diagnostics and Prognostics in Power Plant Machinery. He researched vibration reduction, pipe integrity evaluation, transformer noise reduction and vision measurement. Now his research interests includes prognostics and diagnostics in rotating machinery and 3D vision system for measuring cable vibration.

Yongchae Bae received the Ph.D. degree from the Department of Mechanical Engineering, Chonnam National University, Gwangju, Korea, in 1995. Since 1986, he has been working with the Korea Electric Power Research Institute, Daejeon, Korea. He is currently a Chief Researcher with the Power Generation Lab., KEPCO Research Institute, Daejeon. His research interests include vibration analysis, mechanical integrity of generator stator windings, and monitoring and diagnosis of main equipment in power plant.