CONDITION MONITORING OF ELECTRIC EQUIPMENT IN RAILWAY SUBSTATION BY ANALYZING MAINTENANCE DATA COLLECTED FROM MICS(MAINTENANCE INFORMATION COLLECTED SYSTEM)

Akira MATSUMOTO
East Japan Railway Company – Japan
a-matsumoto@jreast.co.jp

Shouji UETATSU
East Japan Railway Company – Japan
syouji-uematsu@jreast.co.jp

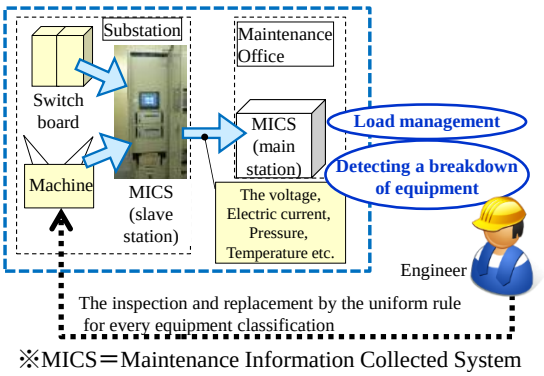
Hiroshi YAMAMOTO
East Japan Railway Company – Japan
yamamoto-hi@jreast.co.jp

Motohiko ONUKI
East Japan Railway Company – Japan
m-onuki@jreast.co.jp

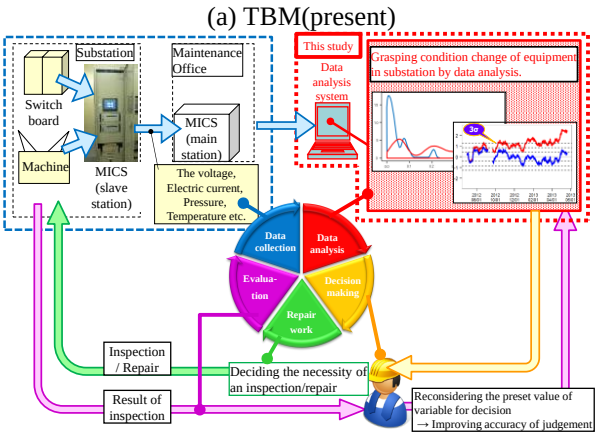
Shouta UKAI
Yokogawa Electric Corporation – Japan
Shouta.Ukai@jp.yokogawa.com

ABSTRACT

MICS(Maintenance Information Collected System) is the system that collects the information about the condition of electric equipment in substation such as voltage, current, pressure, temperature, etc. For determining optimization inspection timing of the equipment, we have examined and devised the method to construct the statistical model by a multiple regression analysis and a principal component analysis. In this method, the Residual are calculated. The Residual shows the degree of deviation between a normal condition and an abnormal condition. And we have showed that the time-series change of this Residual made it possible to monitor the equipment condition and grasp a deterioration trend of the equipment.



(a) TBM(present)



(b) CBM

figure1 The maintenance method in the railway substation

1 INTRODUCTION

In the current maintenance method of the railway substation, as shown in figure1(a), the inspection and replacement are performed with a period that is uniformly determined for each device type (TBM = Time Based Maintenance). However, the degree of deterioration is different for each equipment because of the environment where the equipment is used and individual differences. Therefore, despite the equipment has reached the inspection/replacement timing, it often hasn't been deteriorate too much. So, maintenance costs can be reduced if timing of inspection/ replacement can be determined based on the condition grasped for each equipment (CBM = Condition Based Maintenance). Figure1(b) shows the workflow of CBM on railway substation that we aim to. First, the data such as voltage, current, pressure and temperature are collected from the equipment in the substation. Second, the indicator to grasp deterioration of equipment is extracted by data analysis system. Third, the maintenance engineer determines the necessity of inspection/replacement by this indicator. Fourth, the maintenance engineer performs inspection/replacement of equipment. Finally, for improving the accuracy of determination, parameters used in data analysis system are updated by reviewing result of inspection / replacement.

East Japan Railway Company has automatically collected data such as temperature, pressure, bus voltage and load current from the equipment in the substations by MICS (Maintenance Information Collected System). The data collected from MICS (MICS data) are used for detecting a breakdown of equipment and load management.

We assumed that it was possible to extract the indicator to grasp the condition changing of the equipment by analysis of MICS data and to take advantage of this indicator to CBM. We examined and devised the method to grasp the condition changing of the equipment.

In our method, the statistical model is constructed by the statistical analysis of the past MICS data. This statistical model represents the normal condition of equipment. As the statistical analysis method to construct the statistical model, the multiple linear regression analysis (MLR) and the principal component analysis (PCA) are used. The Residual calculated by MLR and PCA indicates the difference between the current MICS data to the

Table1 The number of MICS Data used in this study

Substation	The number of data			The period
	AC current and voltage	DC current and voltage	Pressure, temperature , etc.	
A	16	9	20	2012/07/01～ 2013/12/31
B	7	10	6	
C	3	6	7	
D	20	9	12	2013/04/01～ 2013/07/31
E	15	6	17	
F	31	6	42	
G	16	10	8	
H	18	13	7	
I	14	21	14	
Total	140	90	133	

statistical model. The Residual is distributed almost near 0 in the normal condition, but increases / decreases in the abnormal condition. It is possible to grasp the condition changing of the equipment by monitoring the Residual.

2 CHARACTERISTICS OF THE MICS DATA

Characteristics of the MICS data are as follows.
(1) It is classified into measurement data (e.g. voltage of bus and current of feeder) and maintenance data (e.g. temperature and pressure of the equipment).
(2) Types of data can be collected are different for each substation. This difference is due to difference in a kind of equipment and time when the equipment has been manufactured (No sensors to collect maintenance data is installed in the old equipment).
(3) It is the maximum, minimum and average value of every hour automatically collected.
In this study, 9 substations were selected because of having relatively much type of maintenance data. The MICS data of these substations for each hour was used (Table1).

3 THE DESIGNED METHOD OF ANALYSIS

3.1 Constructing the statistical model by statistical analysis

In our method, the statistical model is constructed by statistical analysis of the MICS data that has been collected from the substation currently running normally. The model represents the correlation between each MICS data in the normal condition. MLR and PCA, which is the general statistical analysis method, are used for constructing the statistical model. The features and the reasons for selection of these analysis methods are shown as follows.

(1) MLR
MLR is the statistical analysis method that represents the relationship between objective variables and explanatory variables as the simple linear equation. The relationship represented by MLR is called MLR model. By MLR, it is possible to estimate the direct relationship between each physical factor such as current, pressure and temperature of a single machine. Therefore, we assume that MLR is suitable to grasp the condition changing of the equipment with a collapse of these relationships (e.g. gas leakage, abnormal heating). When the number of explanatory variables in MLR model is too much, the reproducibility of MLR model is reduced. This problem is called 'multicollinearity'. So, it is necessary to minimize the number of explanatory variable in MLR model. Table2 shows the factors selected as objective variables and

Table2 Objective variable and explanatory variable of MLR

Objective variable	Explanary Variable 1	Explanary Variable 2
Pressure	Current	Air temperature
Temperature	Current	Air temperature

Table3 Groups of factors for PCA

Group		Factor
Equipment Gr	Power receive	Pressure, temperature and current of power receiving GIS
	SRTr	Pressure ,temperature and current of rectifier transformer and sillicon rectifier
	DTr	Pressure ,temperature and current of distribution transformer
Data Type Gr	Current	Load current for each equipment
	Pressure	Pressure for each equipment
	Temperature	Temperature for each equipment and an air temperature

explanatory variables in MLR from the single machine.
(2) PCA
PCA is the statistical analysis method that decomposes the multivariate dataset into individual components by orthogonal rotation. The components are called the 'principal component'. The principal components have the characteristics of the dataset. PCA can represent the correlation between factors in the dataset. The correlation represented by PCA is called PCA model. Therefore, we assume that PCA is suitable to grasp the collapse of the correlation between factors having a similar trend in normal condition such as the temperature of the No.1 line transformer increasing more than one of other line transformers. By comprehensively investigating the correlation coefficient between factors shown in Table1, the groups in which the correlation coefficient between factors is high are selected. The groups are shown in Table3. The statistic model is constructed for each group. The Group is classified into Equipment Gr and Data Type Gr. Equipment Gr is the group including factors about the equipment of the same type (e.g. GIS, transformer). The Data Type Gr is the group including factors of the same type (e.g. temperature, pressure)

Since MLR and PCA respectively have the advantages and disadvantages as described below, we assume that it is possible to combine the advantages of both methods by the combination of both methods.

(1) To avoid multicollinearity, the MLR model can't only represent the correlation between the physical factors of a single machine. But it is easy to specific the machine where the condition changing has occurred because the model is constructed for each machine.
(2) The PCA model can represent the correlation to be able to grasp intuitively. But it is difficult to specific the equipment of which the change of the condition has occurred because many factors are included in model.

3.2 Grasping the change of the condition by the Residual

When the current MICS data don't conform to the statistical model described in 3.1, it confirms to be able to grasp the condition changing of the equipment. Therefore, in our method, the following Residual is used for the index to grasp the deviation of the current MICS data from the statistical model. Since the Residual is distributed almost near 0 in the normal condition, the

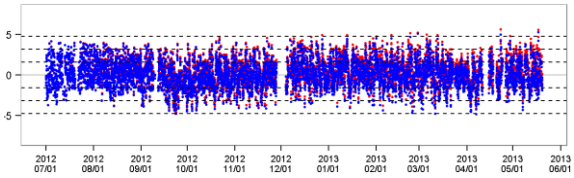
condition changing of the equipment can be grasped when the Residual gradually increases / decreases.

(1) The Residual of MLR is defined the difference between the predicted value, which is the value of the regression formula, of the MLR and the actual value.
(2) The residual component of PCA is defined the principal component excluded from one which the cumulative contribution rate is upper than 99%. The Residual of PCA is defined the square sum of the residual component.

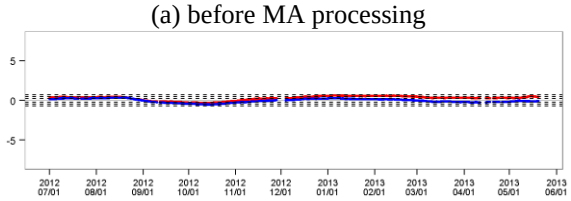
Figure2(a) shows the Residual calculated by MLR and PCA using the data of Table1 in normal condition. The Residual in normal condition isn't distributed completely around 0 and contains the noise. It is estimated that the cause of this noise is the correlation which can't be represented by MICS data. To remove this noise from the Residual, moving average (MA) processing is used. It is expected that the precision to grasp the change of the Residual in long-term is improved. As a result of attempts to some MA width, we could be removed almost noise by MA processing of 8 weeks (figure2 (b)).

4 ESTIMATION OF ANALYSIS METHODS

We evaluated that the condition changing of the equipment could be grasped by the Residual processed with MA. The method and result of the evaluation is shown as follow.



(a) before MA processing



(b) processed with MA

Figure2 The noise of the Residual (example of MLR)

4.1 Pseudo abnormal data that was used in the evaluation

While the equipment in the substation generally continues to deteriorate over more than 10 years, we currently have only MICS data of 1~2 years. Therefore, we hadn't had the MICS data collected when actually deterioration has occurred. To solve this problem, the pseudo abnormal data was used instead of the actually abnormal data. These pseudo abnormal data was created by adding a small change to the normal data. The pseudo abnormal data is shown in Table4. The policy of making abnormal data is as follow.

(1) Based the failure information on past, we assumed that the MICS data in Table4 changed slightly when the equipment deteriorated
(2) The size of changes in (1) was smaller than the size of changes that could be detected by the protective relay.
(3) The pseudo abnormal data was created by adding the changes in (1) to the normal data in 4~12 months.
(4) We assumed the data linearly changed in a period of (3). It wasn't necessarily that the data linearly changes with the actual deterioration. However, since the purpose of the evaluation was a confirmation that can grasp a small change in the data, the size of the change was more important than the slope of the change. Therefore the slope of the change was not considered.

4.2 Criteria for evaluation

When the Residual of pseudo abnormal data was gradually distributed away from 0, the condition changing the equipment is grasped. So, the quantitative criteria were made for evaluation of the change of the Residual. Figure3 is an illustration of the criteria for

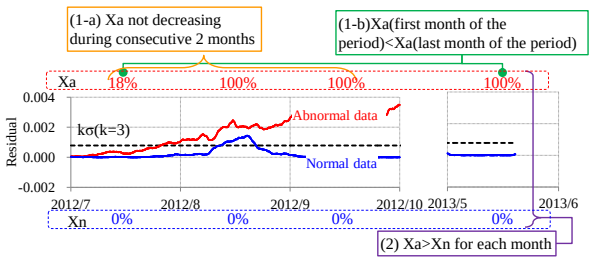


Figure3 The evaluation method

Table4 The Pseudo abnormal data

Substation	Factor			The range of normal data	The range of abnormal data	The supposed deterioration	How to create abnormal data
	No.	Name	Unit				
A	1	Temp. of sillicon rectifier	[℃]	0～44.88	0～44.95	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 12 months
	2	Press. of GIS	[kPa]	485～518	480～515	Pressure decrement because of SF_6 gas leakage	Decreasing from 0 to 10kPa during 12 months
B	1	Press. of rectifier transformer	[kPa]	50.0～760.0	49.6～759.6	Pressure decrement because of insulating oil leakage	Decreasing from 0 to 1% of pressure during 12 months
	2	Press. of GIS	[kPa]	49.0～73.0	48.3～72.95	Pressure decrement because of SF_6 gas leakage	Decreasing from 0 to 10kPa during 12 months
	3	Temp of distribution transformer	[℃]	4.9～44.9	6.2～45.5	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 12 months
C	1	Temp of distribution transformer	[℃]	14.7～46.9	14.7～47.1	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 4 months
G	1	Temp. of sillicon rectifier	[℃]	17.6～62.0	17.6～63.3	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 4 months
I	1	Press. of rectifier transformer	[kPa]	52.3～67.9	52.2～67.3	Pressure decrement because of insulating oil leakage	Decreasing from 0 to 1% of pressure during 12 months
	2	Temp. of sillicon rectifier	[℃]	28.4～59.7	28.4～68.8	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 4 months
H	1	Temp. of sillicon rectifier	[℃]	9.1～52.6	9.1～58.0	Temperature increment because of the fault in radiator	Increasing from 0 to 10% of primary current of transformer during 4 months

evaluation. The meanings of the variable in the criteria are as follow.

- σ : the standard deviation of the residuals of the normal data used for constructing the statistical model
- k : the variable used for the threshold of the Residual
- Xa(*month*)[%] : the rate that the residual for the abnormal data exceeds kσ during the *month*
- Xn(*month*) [%] : the rate that the residual for the normal data exceeds kσ during the *month*

The criteria for evaluation are as follow.

(1) The criteria that Residual increases.
(1-a) Xa not decreasing during two consecutive months
(1-b) Xa(*first month of the period*) <Xa (*last month of the period*)
(2) Xa> Xn for each month

The result of evaluation was judged to be 'Good' when the criteria (1) and (2) were satisfied and 'No good' when these criteria were not satisfied. The smaller the value of k is, the higher the detection rate and the false positive rate are. Conversely, the larger the k value is, the lower the detection rate and the false positive rate are. In this study, k was set to 3 because 3σ corresponded to the confidence interval 99.7% in normal distribution and it was considered to lower the false detection rate enough. To optimize the value of k is a problem to be solved. Moreover, we compared the results of the quantitative evaluation with the results of the evaluation by visual observation of human and confirmed that the result of the quantitative evaluation wasn't significantly different from the human intuition. The criteria of visual observation are as follows. The result of evaluation was judged to be 'Good' when the criteria of (a) and (b) were satisfied and 'No Good' when these criteria were not satisfied.

(a) The Residual of the abnormal data gradually increasing / decreasing
(b) The Residual of the abnormal data exceeding kσ more clearly than the residual of normal data

Table5 The result of evaluation

Substation	Factor			Result of evaluation		Month when the condition of equipment was judged to change			Difference between abnormal data and normal data when the condition of equipment was judged to change		
	No.	Name	Unit	MLR	PCA	MLR	PCA (equipment Gr)	PCA (data Type Gr)	MLR	PCA (equipment Gr)	PCA (data Type Gr)
A	1	Temp. of sillicon rectifier	[℃]	No Good	No Good	-	-	-	-	-	-
	2	Press. of GIS	[kPa]	-	Good	-	2012/08	2012/08	-	0.96~1.92	0.96~1.92
B	1	Press. of rectifier transformer	[kPa]	No Good	Good	-	2012/08	2012/09	-	0.51~1.39	0.88~1.68
	2	Press. of GIS	[kPa]	No Good	Good	-	2013/04	2012/09	-	8.48~9.41	1.92~2.85
	3	Temp of distribution transformer	[℃]	Good	Good	2013/09	2012/08	2012/08	0.37~1.01	0.00~1.84	0.00~1.84
C	1	Temp of distribution transformer	[℃]	No Good	Good	-	-	2013/07	-	-	0.07~0.58
G	1	Temp. of sillicon rectifier	[℃]	No Good	No Good	-	-	-	-	-	-
I	1	Press. of rectifier transformer	[kPa]	Good	Good	2013/05	2013/05	2013/05	0.14~0.32	0.14~0.32	0.14~0.32
	2	Temp. of sillicon rectifier	[℃]	Good	Good	2013/07	2013/05	2013/05		0.00~4.87	0.00~4.87
H	1	Temp. of sillicon rectifier	[℃]	Good	Good	2013/05	2013/05	2013/05	0.00~3.02	0.00~3.02	0.00~3.02

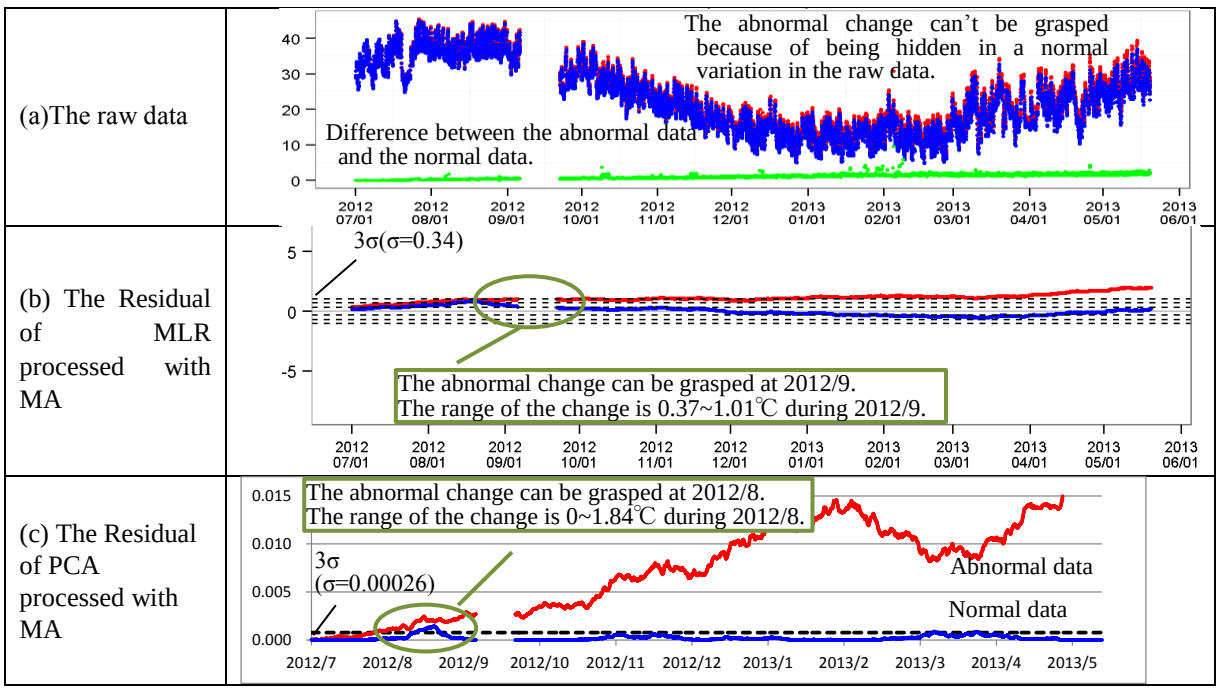


figure2 Grasping the deterioration by the Residual (temp. of distribution transformer in substation B)

4.3 The result of evaluation

The results of evaluation for all abnormal data are shown in Table5. The result was judged to be 'Good' in 8 of 10 cases. It was confirmed that criteria for the quantitative evaluation had been reasonable since the results of the quantitative evaluation matched to the result of the evaluation of visual observation. Also about 2 cases in which the results of evaluation was judged to be 'No Good', the Residual were increasing slightly. So, these 2 cases are expected to be 'Good' if k is set to the optimal value.
Figure4 shows the results of temperature of the distribution transformer in B substation. It is difficult to grasp the abnormality from the raw data because the abnormal change of the raw data is hidden in a normal variation in the raw data (shown in figure4 (a)). However, the Residual of MLR and PCA processed with MA gradually increases from 0 (shown in figure4 (b), (c)).
In balloons of figure4 (b), we note the first month when the result of evaluation was judge to be 'Good' and the rise of temperature at the month. It is clear that it was possible to grasp very small variations.

5 CONCLUSION

5.1 The result of this study

In order to take advantage of CBM, we have designed the method to grasp the condition changing of the equipment in the substation by the statistical analysis of MICS data. The method of analysis is as follows.

(1) The statistical model is constructed by the statistical analysis of the MICS data. The model indicates the condition of the equipment in the substation. As he statistical analysis method, MLR (Multiple Linear Regression analysis) and PCA (Principal Component Analysis) are used.
(2) The Residual are calculated. The Residual means the deviation of the current MICS data from the statistical model.
(3) The noise are removed from the Residual by MA (Moving Average) processing.
(4) When the Residual processed with MA becomes increasing, the condition changing of the equipment can be grasped.

By evaluation using pseudo abnormal data, we show that it is possible to grasp that the condition of the equipment in the substation slightly changes using this method.

5.2 The problem to be solved and the future prospects

We must solve the following problem in order to put this method into practical use. We think that these problems can be solved while we operate this method because much data are needed for evaluation of the solution of these problems.

(1) The deterioration mode that can't be grasped

In this method, it can't be grasped that the deterioration, such as following (a) and (b) because these deterioration has no relation to the MICS data. In order to solve this problem, it is necessary to capture the information in () to MICS.

(a) The failure of the operating mechanism of circuit breaker (the closing time / the magnitude of the vibration)
(b) The failure of elements in the control circuit (voltage / current of the control circuit)

(2) Optimization of parameters used for determining
In this study, the following values were used as the judgment parameters to be used to determine.

(a) MA width of the Residual (8 weeks)
(b) The threshold of the Residual that used for the quantitative evaluation criteria ($k\delta$, $k = 3$)
(c) The cumulative contribution rate used for calculating the Residual of PCA (99%)

Since optimum values of these parameters are different for each substation and equipment, it is necessary to propose the workflow in which these parameters can be updated while operating. To take advantage of this method to CBM, we will continue to evaluate reproducibility of the model by using the latest MICS data.

REFERENCES

[1] TH Wonnacott and RJ Wonnacott, 1998, *" Kaikibunseki to sono ouyou "*, Gendai-Sugakusha, Kyoto, Japan, pp.94-97, 568-575

[2] K.Vamuza and P.Filzmoser, 2009, " Introduction to Multivariate Statistical Analysis in Chemometrics ", *CRC Press*,.63-64.

[3] T Haga and S Hashimoto, 1980, *" Kaikibunseki to Shuseibunbunseki "*, JUSE Press, Ltd, Tokyo, Japan, pp.31-35, 162-164, 218-220