Anomaly Detection for Noisy Data with the Mahalanobis–Taguchi System

Masato Ohkubo, Yasushi Nagata

ABSTRACT

Purpose: Condition-based maintenance requires an accurate detection of unknown yet-to-have-occurred anomalies and the establishment of anomaly detection procedure for sensor data is urgently needed. Sensor data are noisy, and a conventional analysis cannot always be conducted appropriately. An anomaly detection procedure for noisy data was therefore developed.

Methodology/Approach: In a conventional Mahalanobis–Taguchi method, appropriate anomaly detection is difficult with noisy data. Herein, the following is applied: 1) estimation of a statistical model considering noise, 2) its application to anomaly detection, and 3) development of a corresponding analysis framework.

Findings: Engineers can conduct anomaly detection through the measurement and accumulation, analysis, and feedback of data. Especially, the two-step estimation of the statistical model in the analysis stage helps because it bridges technical knowledge and advanced anomaly detection.

Research Limitation/implication: A novel data-utilisation design regarding the acquired quality is provided. Sensor-collected big data are generally noisy. By contrast, data targeted through conventional statistical quality control are small but the noise is controlled. Thus various findings for quality acquisition can be obtained. A framework for data analysis using big and small data is provided.

Originality/Value of paper: The proposed statistical anomaly detection procedure for noisy data will improve of the feasibility of new services such as condition-based maintenance of equipment using sensor data.

Category: Research paper

Keywords: engineered system; Gaussian graphical model; liner Gaussian model; statistical modelling; Taguchi method

1   INTRODUCTION

In recent years, the condition-based maintenance of equipment using sensor data has been put into practical use in the Japanese manufacturing industry based on the progress of technologies related to the Internet of Things (IoT). Condition-based maintenance requires an accurate detection of unknown anomalies that have yet to occur, and thus the establishment of an anomaly detection procedure for sensor data is needed. However, because sensor data are noisy with features observed by adding noise to the true measurement value, a conventional analysis procedure cannot always be used to conduct an appropriate analysis. The purpose of this study is to develop a novel anomaly detection procedure for noisy data.

Figure 1 shows a schematic diagram illustrating the process of generating noisy data. The symbols x, y, and e in the figure are P-dimensional random variables indicating the true measurement value, observed value, and noise, respectively. For example, when assuming the formation process of a porous film, the true value of measurement x takes a value measured using precision-measuring equipment as the physical properties of the porous film (e.g., film thickness, film width, and pore diameter). By contrast, the observed value y takes a value measured using sensors for such physical properties. Here, noise is caused by measurement errors derived from the sensors or by various other factors. Thus, such noise is additively superimposed over the true value of the measurement, and $y = x + e$ is established. As described in Section 3, this model is also related to the engineered system used in the Taguchi method.

When dealing with such noisy data, problems can arise in which it becomes difficult for engineers to consider adopting anomaly detection algorithms from the viewpoint of intrinsic technology. Taking the example of the formation process of a porous film, it can be stated that it is easy for skilled engineers to consider the relation between the physical properties of the porous film from the true measurement value. By contrast, in the case of noisy data, correlations that should be found among the physical properties of a porous film cannot be inferred from the data. Considering the influence of noise, anomaly detection based on a model in which the essential correlation structure in the true measurement value is ignored can be useful for improving the anomaly detection. Therefore, there is a risk that the adopted anomaly detection algorithms will lack validity from the viewpoint of intrinsic technology.
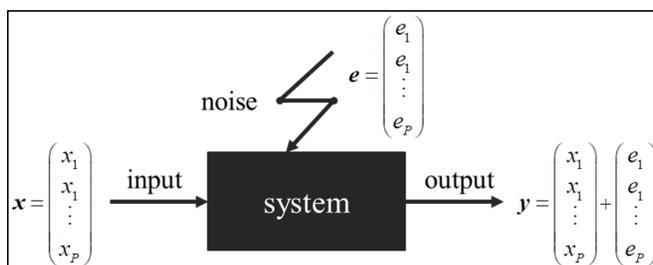
Figure 1 – Process of Generating Noisy Data
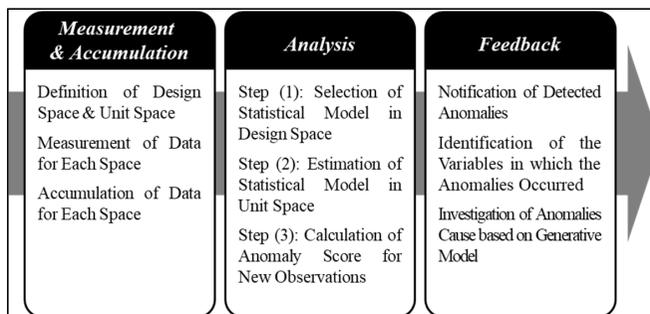
Figure 2 – Overview of Proposed Procedure

To solve this problem, a novel anomaly detection procedure is proposed that improves the Mahalanobis–Taguchi (MT) method developed by Taguchi and Jugulum (2002). Figure 2 shows an overview of the proposed procedure. From Figure 2, the proposed procedure consists of three stages, namely the measurement and accumulation, analysis, and feedback of data. Through these stages, an anomaly detection procedure that allows engineers to consider algorithms from the viewpoint of intrinsic technology while maintaining a predictive performance is expected to be realised.

In the first stage, the measurement and accumulation of data are conducted for two spaces, the design space and the unit space. The design space is a novel concept not found in a conventional MT method. The concept of a design space is proposed with the intention of incorporating the technical knowledge of the engineer into the analysis procedure. The details of this are described in Section 3.2.1.

In the following analysis stage, anomaly detection is conducted using three steps: (1) the selection of a statistical model in the design space, (2) an estimation of a statistical model in the unit space, and (3) the calculation of an anomaly score for new observations. Step (1) is a novel stage. It has been recognised that statistical models that are consistent with the engineer's technical knowledge often differ from models showing a high predictive performance. The proposed procedure is expected to fill in the gap between these two models. The details of this are provided in Section 3.2.2.

In the final feedback stage, anomalies detected in the analysis stage are notified to the engineers. This helps investigate the causes of anomalies by identifying the variable that is suspected of incurring an anomaly. It is also expected to make it possible to investigate the causes of anomalies more accurately by analysing the statistical model estimated during the analysis stage according to the proposal by Ohkubo and Nagata (2017). The details of this are described in Section 3.2.3.

The structure of this paper is as follows. In Section 2, the MT method as used in previous studies is applied and its anomaly detection procedure is described. In Section 3, a novel anomaly detection procedure for noisy data is proposed. In Section 4, the usefulness of the proposed procedure is confirmed through a

Monte Carlo simulation. Finally, in Section 5, the implications of the proposal are discussed.

2   MAHALANOBIS–TAGUCHI SYSTEM

The MT system is a generic term for multivariate analysis methods proposed using the Taguchi method. There are many applications in various fields centring on the manufacturing industry (e.g., Jin and Chow, 2013; Peng et al., 2017). Among the different MT systems, in recent years, the MT method, which is a multivariate analysis procedure for anomaly detection, has attracted the attention of practitioners. In fact, the MT method has been used as an algorithm forming the core of an anomaly prediction system of equipment using sensor data (Takahama and Mikami, 2012). In this section, an overview of the MT method is described in terms of the measurement and accumulation, analysis, and feedback to facilitate a comparison with the proposed procedure.

2.1   Measurement and Accumulation Stage

In the measurement and accumulation stage, after defining a dataset called a unit space, the measurement and accumulation of data are conducted. A unit space is a group forming a homogeneous population. In general, in the field of statistical anomaly detection, after sampling normal data, namely data with a normal label, the data are used as training data to estimate a statistical model. The unit space applied in the MT method can be considered training data consisting only of data with normal labels.

Here, when conducting anomaly detection, it is necessary to prepare test data to evaluate whether the performance of the algorithm is sufficient for practical use. Test data consist of data with normal labels and data with anomaly labels. In the proposed procedure, although test data are also required, the description is omitted because it becomes redundant. In addition, it is necessary to prepare a dataset called "signal data" separately from the unit space in the MT method. However, it has been pointed out that the definition and use of signal data remain controversial (e.g., Woodall et al., 2003; Inoh et al., 2012). In this paper, signal data are interpreted as test data, and signal data or the analysis procedure related to it are not discussed.

2.2   Analysis Stage

In the analysis stage, anomaly detection is conducted using two steps: (1) an estimation of the Mahalanobis distance in the unit space and (2) the calculation of an anomaly score for new observations. In general, in the field of anomaly detection, the deviation from the normal state is quantified by a certain scale, and an anomaly is considered if the scale exceeds a predetermined threshold, and a normal state is determined if it does not. The MT method can be said to be an

anomaly detection procedure that uses the Mahalanobis distance (MD) as its scale.

First, the MD is estimated from an observed value in the unit space. Let $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ be the population mean vector and population covariance matrix of the P-dimensional variable $\boldsymbol{y}$ observed for a population. Then, the MD on the population is defined as follows:

$$\Delta(\boldsymbol{y}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) = (\boldsymbol{y}-\boldsymbol{\mu})^{T}\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}), \tag{1}$$

where T is the transpose of a vector or matrix. Note that the vectors in this paper are vertical vectors, and their transposes are horizontal vectors. Then, with the MT method, the MD on the population is estimated using the sample mean vector and sample covariance matrix obtained from N individuals belonging to the unit space as the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

In (2), the anomaly score for the new observations is calculated. Here, a function for quantifying the degree of deviation from the normal state is called an anomaly score function. In addition, the value of this function taken in practice is called an anomaly score. The anomaly score function in the MT method is the following function in which $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in equation (1) are replaced by their respective estimator $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$:

$$\Delta(\boldsymbol{y}\,|\,\hat{\boldsymbol{\mu}},\hat{\boldsymbol{\Sigma}}) = (\boldsymbol{y}-\hat{\boldsymbol{\mu}})^{T}\,\hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{y}-\hat{\boldsymbol{\mu}}). \tag{2}$$

At this time, after substituting the new observed value $\boldsymbol{y}_{\text{new}}$ into equation (2) and calculating the anomaly score, an anomaly is determined if the value exceeds the predetermined threshold; otherwise, it is judged as normal.

2.3 Feedback Stage

In the feedback stage, anomalies detected during the analysis stage are notified to the engineers. In addition, the variable suspected of having the anomaly is identified through a causal diagnosis using an orthogonal array. Here, a causal diagnosis using an orthogonal array is conducted as follows: (1) assign each variable on a two-level orthogonal array as a factor, taking 1 when using the j-th variable of observed variable $\boldsymbol{y}$, and 0 when not using it, (2) calculate the anomaly score through the combination specified in each row, (3) calculate the factorial effects of each variable, and (4) create a graph of the factorial effects. See Taguchi and Jugulum (2002) for more details on this.
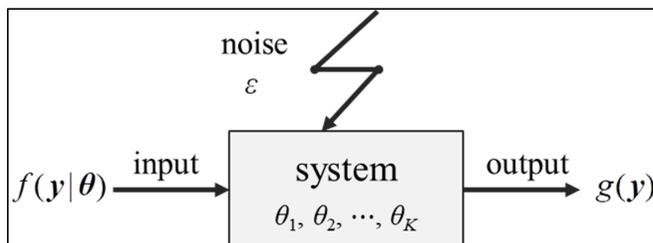
Figure 3 – Modelling Philosophy of the Proposed Procedure

3   PROPOSED PROCEDURE

In this section, the proposed anomaly detection procedure is described, which is an improvement over the MT method under the premise of its application to noisy data.

Figure 3 shows a schematic diagram illustrating the modelling philosophy, following the parameter design of the Taguchi method. The symbols $f(y|\theta)$, $g(y)$, and ε in the figure indicate parametric statistical models, the true distribution of y, and noise, respectively. The symbols θ_1, θ_2, ⋯, θ_K in the figure are the elements of θ in $f(y|\theta)$, respectively. Figure 3, which shows a system chart of the Taguchi method, illustrates the estimation of $f(y|\theta)$ that minimizes the effect of noise ε by appropriately applying the set of parameters θ_1, θ_2, ⋯, θ_K. In addition, the entire system shown in the system chart is called an engineered system.

The novel anomaly detection procedure is formulated according to this modelling philosophy. The aim is to prove that the proposed procedure will allow engineers to consider the algorithm from the viewpoint of intrinsic technology while maintaining a predictive performance by introducing some concepts from the Taguchi method. The details of the proposed procedure are described for each stage, namely the measurement and accumulation, analysis, and feedback of data, in the following sections. Note that Taguchi, Chowdhury and Wu (2005) is used to indicate the Taguchi method.

3.1   Measurement and Accumulation Stage

In the measurement and accumulation stage, after defining two datasets, called the "design space" and "unit space", a measurement and accumulation of the data are conducted for each space. In this section, the definition of each space and the established relation between the true measurement value x, observed value y, and noise e in the design and unit spaces, respectively, are described.

First, the design space is defined as a set of observed values in an ideal state. Such data are actively measured and accumulated by engineers mainly using

precision instruments with the purpose of performing statistical modelling from the viewpoint of intrinsic technology. At this time, it is assumed that $y = x$ is established between the true value of measurement x and the observed value y.

Next, the unit space is defined as a set of observed values in a usual state. Such data are measured and accumulated passively from sensors with the purpose of applying statistical modelling from the viewpoint of a predictive performance. At this time, it is assumed that the observed value y is generated along with the noise e according to the generation process shown in Figure 1. That is, $y = x + e$ is established.

Here, the concept of a design space is additionally described. Although both data types in the design and unit spaces are P-dimensional observation variables with a normal label, the data collection method and generative model differ, as described above. Therefore, merging these data and treating them as single learning data is not recommended. The intention of using the design space is considered the same as using a test piece with the Taguchi method. In the Taguchi method, it is considered useful to create a model, called a test piece, to capture the essence of a technology. The design space is also measured and accumulated to grasp the essential structure of the data.

Note that the notations of the probability density function in each space are as follows. First, the true distribution of y, i.e. $g(y)$, is described as $g_0(y)$ when it emphasises the distribution of the design space. Similarly, when it emphasises the distribution of the unit space, it is described as $g_1(y)$. Next, the true distribution of x is described as $s(x)$ in both cases, assuming it does not change in the design or unit space. Finally, the parametric statistical model $f(y|\theta)$ is described as $f_0(y|\theta_0)$ and $f_1(y|\theta_1)$ in the design and unit spaces, respectively, as in the case of $g(y)$.

3.2 Analysis Stage

In the analysis stage, anomaly detection is applied using three steps: (1) selection of the statistical model in the design space, (2) estimation of the statistical model in the unit space, and (3) the calculation of an anomaly score for new observations.

In (1), statistical modelling is conducted on the set of observed values in the design space under the assumption of a Gaussian graphical model. At this time, the best model is selected from a family of models while actively using the technical knowledge. In (2), statistical modelling is conducted on the set of observed values in the unit space under the assumption of a linear Gaussian model. That is, an estimation of a statistical model considering the existence of noise is conducted using the information of the model selected in (1). In (3), anomaly detection is conducted based on the statistical model estimated in (2) according to the general theory of anomaly detection. The details of this procedure are described below in order of (1) to (3).

3.2.1 Selection of Statistical Model in Design Space

Now, when a random variable z follows a P-dimensional normal distribution with the P-dimensional vector μ and the $P \times P$ symmetric positive-definite matrix Σ as parameters, its probability density function is described as $\mathcal{N}(z|\mu,\Sigma)$. At this time, assume $g_0(y) = \mathcal{N}(y|\mu,\Sigma)$ and $s(x) = \mathcal{N}(x|\mu,\Sigma)$. In this step, Gaussian graphical modelling (GGM) described by Lauritzen (1996) is conducted on the set of observations valued in the design space under this assumption.

GGM is one of the basic methods used in a multivariate analysis, which is conducted by applying the following steps: (1) estimating the parameters in a multivariate normal distribution and (2) evaluating and selecting statistical models, assuming that most of the non-diagonal elements of Ω take a value of zero where the inverse matrix of the covariance matrix Σ is Ω. There are many applications of GGM; applied procedures have also been proposed in the fields of anomaly detection (e.g., Ide et al., 2009). Ohkubo and Nagata (2017) proposed introducing the MT method, and indicated that the method is useful for both improving the anomaly detection performance and pursuing the cause of the anomaly.

In this study, GGM is conducted to understand the generative model, as applied by Ohkubo and Nagata (2017), for the usefulness of the feedback step described later. Therefore, when using the information criteria, it is recommended to use criteria that have the same concept as the Bayesian information criteria (BIC) developed by Schwarz (1978). However, a statistical model should be selected from the viewpoint of intrinsic technology while referring to such model of the evaluation criteria. As mentioned later, the model selection in this step does not significantly affect the anomaly detection performance, and it can therefore be useful to obtain a model to understand the essential correlational structure of the data.

3.2.2 Estimation of Statistical Model in Unit Space

In this step, estimating a parametric statistical model $f_1(y|\theta_1)$ against a set of observed values in the unit space is considered. Here, a linear Gaussian model is assumed for the true distribution $g_1(y)$ in the unit space. Now, assume that the probability density function of noise e is $\mathcal{N}(e|\mathbf{0}_P,\Lambda^{-1})$, where $\mathbf{0}_P$ is a P-dimensional zero vector and Λ is a $P \times P$ symmetric positive-definite matrix. The observed value y then follows a multivariate normal distribution from the assumption in the signal space, $y = x + e$. In addition, from the reproducibility of the normal distribution, the marginal distribution of y is described as follows:

$$g_1(y) = \mathcal{N}(y|\mu,\Omega^{-1}+\Lambda^{-1}), \tag{3}$$

where $s(\boldsymbol{x}) = \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}^{-1})$ from the assumption of the design space, $\boldsymbol{y} = \boldsymbol{x}$. At this time, an estimation of the parameters can be conducted under the model $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1)$, which is assumed to be a linear Gaussian model where $\boldsymbol{\theta}_1 = \{\boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Lambda}\}$. Note that the expectation–maximisation (EM) algorithm can be easily implemented for this model (e.g., Bishop, 2006).

3.2.3   Calculation of Anomaly Score for New Observations

In this step, an anomaly score function is defined based on the general theory of anomaly detection (e.g., Yamanishi and Takeuchi, 2002). Specifically, $f_1(\boldsymbol{y} \mid \hat{\boldsymbol{\theta}}_1)$ is defined, which is a statistical model in which the parameters of $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1)$ have been replaced by the estimators $\hat{\boldsymbol{\theta}}_1$, as the anomaly score function:

$$Score\left(\boldsymbol{y} \mid \hat{f}_1\right) = -\ln f_1(\boldsymbol{y} \mid \hat{\boldsymbol{\theta}}_1). \tag{4}$$

Here, the plug-in estimator is applied such that the estimator of the parameter is plugged into the original probability density function as an estimator of the parametric statistical model, and $\hat{f}_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1) = f_1(\boldsymbol{y} \mid \hat{\boldsymbol{\theta}}_1)$ is established. At this time, after substituting the new observed value $\boldsymbol{y}_{\text{new}}$, which is the target of judgement, into equation (4) and calculating the anomaly score as in the case of the MT method, an anomaly is determined if the value exceeds the predetermined threshold; otherwise, the value is judged as normal.

3.2.4   Relationship with Modelling Philosophy

This sub-subsection additionally describes the relation between the analysis stage in the proposed procedure and the modelling philosophy.

According to the modelling philosophy shown in Figure 3, the purpose of the proposed procedure is to estimate a $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1)$ that minimizes the effect of noise ε by setting $\boldsymbol{\theta}_1$ appropriately. At this time, it is assumed that the following relation holds among $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1)$, $g_1(\boldsymbol{y})$,   and ε:

$$Score\left(\boldsymbol{y} \mid g_1\right) = Score\left(\boldsymbol{y} \mid f_1\right) + \varepsilon, \tag{5}$$

where the *Score* function has the same definition as in equation (4). Equation (5) can be interpreted as the state in which the *Score* functions of $g_1(\boldsymbol{y})$ and $f_1(\boldsymbol{y} \mid \boldsymbol{\theta}_1)$ match, which can be interpreted as the ideal state. In addition, it can be stated that minimising noise ε while aiming at an ideal state is a means to realise the above-mentioned modelling philosophy. Note that, in the case of $\varepsilon = 0$, equation (5) is called an ideal function in the Taguchi method.

When aiming at such an ideal function, it is possible to consider the expected noise value ε against $g_1(\boldsymbol{y})$ as in the following evaluation index:

$$E[\varepsilon] = \int \left\{ Score\left(\boldsymbol{y}\,|\,g_1\right) - Score\left(\boldsymbol{y}\,|\,f_1\right) \right\} g_1(\boldsymbol{y}) dy$$

$$= -\int \left\{ \ln \frac{g_1(\boldsymbol{y})}{f_1(\boldsymbol{y}\,|\,\boldsymbol{\theta}_1)} \right\} g_1(\boldsymbol{y}) dy \tag{6}$$

Here, the right-hand side of the second equal sign is the negative Kullback–Leibler (KL) divergence from the equation form. Because a KL divergence is non-negative (e.g., Bishop, 2006), the minimisation problem in equation (6) can be said to be the KL divergence minimisation problem. Note that this evaluation index can be stated to be the same concept as the signal-to-noise (SN) ratio in the Taguchi method.

From the above discussion, the rational modelling procedure in (2) is a modelling procedure that aims to estimate $f_1(\boldsymbol{y}\,|\,\boldsymbol{\theta}_1)$, which accurately approximates $g_1(\boldsymbol{y})$ in equation (3). That is, the estimated covariance structure is an estimated value of $\boldsymbol{\Omega}^{-1} + \boldsymbol{\Lambda}^{-1}$, where the effect of noise is added to the essential correlation structure. Therefore, it is difficult to grasp the essential correlation structure only through (2). In the same way, the same problem occurs in the conventional MT method. By contrast, with the proposed procedure, it is possible to realise the modelling philosophy by executing (2) after (1) while grasping the essential correlation structure.

3.3   Feedback Stage

In the feedback stage, anomalies detected during the analysis stage are notified to the engineers. In addition, as with a conventional MT method, identifying the variable that is suspected of having an anomaly through a causal diagnosis using an orthogonal array helps in investigating the cause of the anomaly. At the same time, information on the covariance structure among the parameters of the statistical model previously estimated during the analysis stage is simultaneously provided.

Ohkubo and Nagata (2017) proposed applying the GGM framework to investigate the cause of an anomaly when it occurs. It is possible to consider the cause of an anomaly through the conditional independence between variables by observing the precision matrix, which is the analysis result of the GGM. However, in the case of noisy data, because the covariance structure is affected by noise in the unit space, it is difficult to grasp the essential structure of the correlation even if the precision matrix is observed. By contrast, with the proposed procedure, even in the presence of noise, it is possible to obtain an estimate of the precision matrix $\boldsymbol{\Omega}$ with high accuracy, and thus it is possible to capture an essential correlation structure. Therefore, the proposed procedure is expected to be useful for investigating the cause of an anomaly more accurately even in the case of noisy data.

4   MONTE CARLO SIMULATION

In this Section, through Monte Carlo simulation, the usefulness of the proposed procedure is verified from two perspectives: predictive performance and consistency with the intrinsic technology.

4.1   Dataset Overview

In this experiment, 100 sets of design space, unit space, and test data are prepared. Here, the dimension of each data is 10 and the sample size is 50, 200, 50,000 in order.

The generative model is $\mathcal{N}(y\,|\,\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0^{-1})$ for design space and $\mathcal{N}(y\,|\,\boldsymbol{\mu}_0, \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Lambda}_0^{-1})$ for unit space, using $\boldsymbol{\mu}_0$, $\boldsymbol{\Omega}_0$, and $\boldsymbol{\Lambda}_0$ described later. That is, according to the generation process shown in Figure 1, it is assumed that the observed value y is generated with noise. The test data is composed of data with normal labels and anomaly labels, and its generative model of normal state is the same as unit space. By contrast, the generative model of anomaly state is $\mathcal{N}(y\,|\,\boldsymbol{\mu}_0, c^2\boldsymbol{\Omega}_0^{-1} + c^2\boldsymbol{\Lambda}_0^{-1})$ where $c^2 = 5$.

Here, $\boldsymbol{\mu}_0$, $\boldsymbol{\Omega}_0$, and $\boldsymbol{\Lambda}_0$ are set as follows. First, let $\boldsymbol{\mu}_0$ be a 10-dimensional zero vector. Next, $\boldsymbol{\Omega}_0$ is set so that most of its off-diagonal elements are zero while satisfying positive definite. Specifically, we use the following procedure applying Cholesky decomposition: (1) prepare a lower triangular matrix $\mathbf{B}_{\boldsymbol{\Omega}}$ whose elements are all 1 including diagonal elements, (2) replace 20% of off-diagonal elements in $\mathbf{B}_{\boldsymbol{\Omega}}$ with a value of zero, (3) standardize so that the diagonal elements of $\mathbf{B}_{\boldsymbol{\Omega}}\mathbf{B}_{\boldsymbol{\Omega}}^T$ are 1, and (4) let $\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}^T$ be the result of procedure in (3), and then let $\boldsymbol{\Omega}_0$ be $(1-\alpha)^{-1}\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}^T$ weighted with α ($0 < \alpha < 1$). Let $\boldsymbol{\Lambda}_0$ be $\alpha^{-1}\tilde{\mathbf{B}}_{\boldsymbol{\Lambda}}\tilde{\mathbf{B}}_{\boldsymbol{\Lambda}}^T$ according to the same procedure. Note that we call "uncorrelated noise" when the ratio of non-zero elements to the all non-diagonal elements of $\tilde{\mathbf{B}}_{\boldsymbol{\Lambda}}$ is 0%, and call "correlated noise" when is 100%. From above, if we define $\boldsymbol{\Sigma}_0$ as $\boldsymbol{\Sigma}_0 = \boldsymbol{\Omega}_0^{-1} + \boldsymbol{\Lambda}_0^{-1}$, the following relationship holds:

$$\boldsymbol{\Sigma}_0 = (1-\alpha)\left(\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}\tilde{\mathbf{B}}_{\boldsymbol{\Omega}}^T\right)^{-1} + \alpha\left(\tilde{\mathbf{B}}_{\boldsymbol{\Lambda}}\tilde{\mathbf{B}}_{\boldsymbol{\Lambda}}^T\right)^{-1}. \tag{7}$$

From equation (7), it can be seen that α ($0 < \alpha < 1$) is a weight for covariance matrix of noise. As α approaches 0, the effect of noise decreases, and as it approaches 1, the effect of noise increases.

4.2   Evaluation Objects

In this experiment, Proposed MT, Predictive MT, and Interpretive MT are compared. First, Proposed MT means our proposed procedure. Next, Predictive MT means MT method with emphasis on predictive performance. In this case, it is MT method using sample mean vector and sample covariance matrix

calculated from the observation values of unit space. Finally, Interpretive MT means MT method which emphasizes that it can be interpreted from the viewpoint of intrinsic technology. In this case, it is MT method using the sample mean vector and sample covariance matrix estimated from the observed values of unit space, under the known graph structure of Ω_0 (i.e., the position of the nonzero element of Ω_0). We note that the MT method using true parameters is called Ideal MT. Then, it is judged that the closer to the performance of Ideal MT, the better the anomaly detection procedure.

4.3   Evaluation criteria

In this experiment, the performance of the evaluation objects shown in Section 4.2 is evaluated from the viewpoint of prediction performance and consistency with the intrinsic technology. Here, the consistency with the intrinsic technology in this experiment is evaluated based on whether the generative model of x in the design space can be accurately estimated. The following describes specific evaluation criteria.

First, the evaluation criterion for measuring the prediction performance is the positive discrimination rate of test anomaly samples (hereinafter referred to simply as the positive discrimination rate) of each procedure. The positive discrimination rate takes a value from 0 to 100 and is better if close to 100. The threshold value is set at 1% of the negative discrimination rate of test normal samples (hereinafter referred to simply as the negative discrimination rate).

Next, the evaluation criterion to measure the consistency with the intrinsic technology is the KL divergence between the distribution $s(x)$ of the true measured value x and its predicted distribution $\hat{s}(x) = \mathcal{N}(x \mid \hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Omega}}_0^{-1})$. Here, Predictive MT and Interpretive MT in the previous section use $\hat{\boldsymbol{\Sigma}}_0^{-1}$ as $\hat{\boldsymbol{\Omega}}_0$.

4.4   Experimental Result

The results of this experiment are shown in Figure 4 and Figure 5. Note that the proposed procedure is conducted by a maximum likelihood estimation supposed that the structure of Ω_0 is given when we estimate parameters from the design space in order to align conditions with Interpretive MT.

Figure 4 shows graphs comparing the evaluation objects from the viewpoint of prediction performance. The left side corresponds to the experimental results for uncorrelated noise and the right side corresponds to the experimental results for correlated noise. The vertical axis of each figure is the positive discrimination rate, and the horizontal axis is the weight α for the covariance matrix of noise.
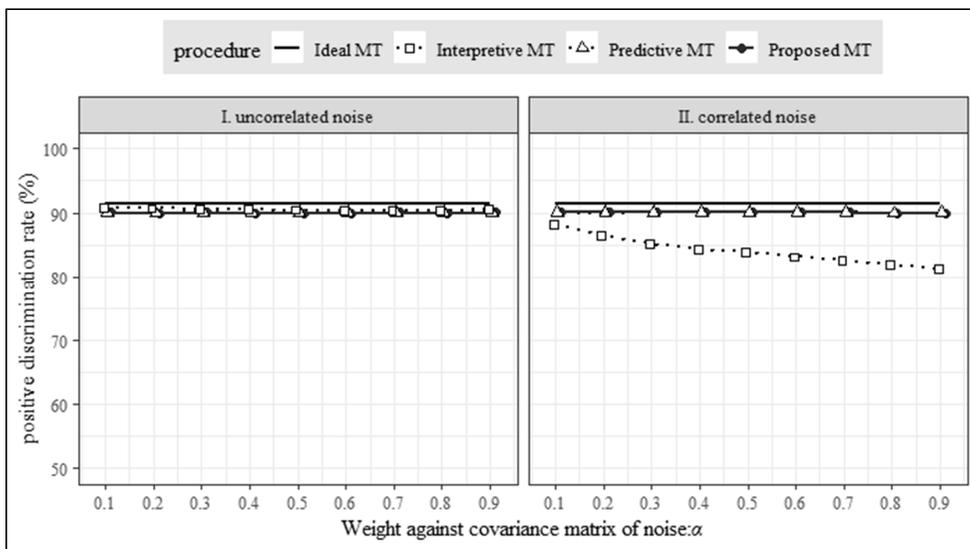
Figure 4 – Evaluation of Predictive Performance In Each Procedure

From Figure 4, in the case of correlated noise, it can be observed that the positive discrimination rate of Interpretive MT decreases as α increases. In the case of correlated noise, this is caused by the difference between the graph structure of x and y, occurring due to the correlation between the variables of noise. Because the prior knowledge used by Interpretive MT is a graph structure of x, the influence of incorrect model setting becomes stronger as the influence of noise becomes stronger. By contrast, although Proposed MT also uses the graph structure of x, appropriate analysis is possible because the model is corrected from data in the unit space.

Figure 5 shows graphs comparing the evaluation objects from the viewpoint of consistency with the intrinsic technology. The left side corresponds to the experimental results for uncorrelated noise and the right side corresponds to the experimental results for correlated noise. The vertical axis of each figure is the KL divergence, and the horizontal axis is the weight α for the covariance matrix of noise.

It can be observed from Figure 4 that in the case of correlated noise, Proposed MT has obtained a predicted distribution with smaller KL divergence than Predictive MT and Interpretive MT. The cause of this phenomenon is that in Predictive MT and Interpretive MT, estimation is conducted on the population distribution of y, not x. In the case of correlated noise, the correlation structure of y becomes different from the correlation structure of x as the influence of the noise becomes stronger. Therefore, in the case of correlated noise, it can be said that it becomes difficult to estimate the generative model of x. By contrast, Proposed MT can learn the generative model of unit space while maintaining the information of the generative model of x.
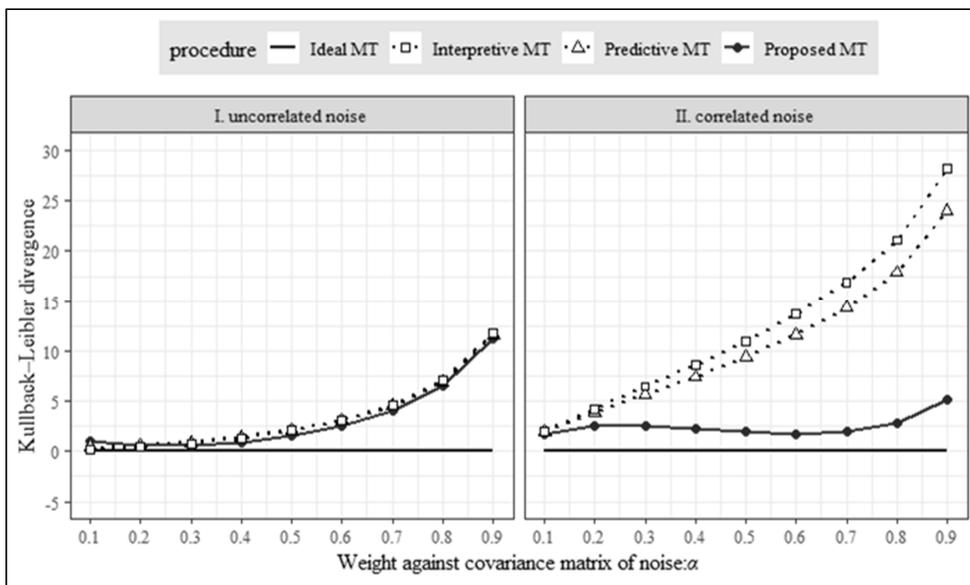
Figure 5 – Evaluation of Consistency with Intrinsic Technology in Each Procedure

5 DISCUSSION

In this research, a novel anomaly detection procedure was proposed for noisy data. Specifically, the anomaly detection procedure used in the MT method, which is a representative methodology based on the Taguchi method, has been improved such that noisy data can be properly analysed. Through each stage, namely the measurement and accumulation, analysis, and feedback of data, quality-management engineers can conduct anomaly detection while reflecting on their technical knowledge in an analysis. Although the proposal covers all three stages of data utilisation, a two-step estimation of the statistical model in the analysis stage is useful in the sense that it fills in the gap between the engineer's technical knowledge and advanced anomaly detection. The usefulness of the proposed procedure was thus shown through a theoretical examination and numerical experiment using a Monte Carlo simulation.

In recent years, the utilisation of IoT-related technologies in the manufacturing industry has become increasingly popular. Because IoT-related technology has also been reviewed as an important factor in establishing a competitive advantage in the manufacturing sector (Porter and Heppelmann, 2014), active discussions should be made on how to use IoT-related technologies in all aspects of corporate activities, including quality management. In fact, the impact of IoT on quality management has been considered from various perspectives, not only in industry but also in the academic field (e.g., Foidl and Felderer, 2015; Park et al., 2017; Shin et al., 2018). Thus, how to use data for quality management has been recognised as a common key factor.

This proposal should lead to a data utilisation design when considering the aspect of quality acquisition. In recent years, although many arguments have been made regarding the use of so-called big data collected from sensors and smart devices, it is difficult to obtain technical knowledge from such data because much of the data are noisy. By contrast, data targeted using a conventional statistical quality control and the Taguchi method are small but their noise is controlled. Thus, from such small data, various findings for a quality acquisition can be determined. This proposal will lead to the provisioning of a novel framework for conducting a data analysis while taking advantage of both big and small data.

The proposed procedure enables the learning of an anomaly detection model to achieve a high performance from the unit space corresponding to big data while obtaining technical knowledge from the design space corresponding to small data. Even in the field of anomaly detection, high-performance machine learning methodologies have been proposed, e.g., the local outlier factor developed by Breunig et al. (2000) and the one-class support vector machine developed by Schölkopf et al. (2001). However, many of these methodologies are black-boxed algorithms. By contrast, although conventional procedures of MT systems have clear algorithms, it is difficult to achieve as high a performance as a machine learning methodology. The proposed procedure is an example of a methodology for achieving a high performance while ensuring validity from the viewpoint of intrinsic technology, whereas the assumptions for statistical models are severe.

Finally, the contributions and limitations of this research are summarised as follows. The main contributions of this study are to establish a novel anomaly detection procedure for noisy data and improve the feasibility of new services, such as condition-based maintenance of the equipment using sensor data. A limitation of this study is its inability to quantitatively express the degree of abnormality for each occurrence factor. In the field of multivariate control charting, originating from the T^2 chart proposed by Hotelling (1947), when an anomaly occurs, the anomaly cause, which is a breakdown of the essential correlation structure or fluctuation from noise, is expressed as individual statistics (e.g., Jackson and Mudholkar, 1979). Similar proposals have also been made in the field of MT systems (Ohkubo and Nagata, 2018). From this perspective, a future study will be applied to improve the proposed procedure. It is also necessary to apply the proposed procedure to real cases such as condition-based equipment maintenance.

ACKNOWLEDGEMENTS

REFERENCES

Bishop, C.M., 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Breunig, M.M., Kriegel, H.P., Ng, R.T. and Sander, J., 2000. LOF: Identifying density-based local outliers. *ACM SIGMOND Record*, [e-journal] 29(2), pp.93-104. DOI: 10.1145/342009.335388.

Foidl, H. and Felderer, M. 2015. Research challenges of industry 4.0 for quality management. In: M. Felderer, F. Piazolo, W. Ortner, L. Brehm, HJ. Hof, eds., *Proceedings of the International Conference on Enterprise Resource Planning Systems*. Munich, Germany. 16-17 November. pp.121-137. DOI: 10.1007/978-3-319-32799-0_10.

Hotelling, H., 1947. Multivariate quality control - illustrated by the air testing of sample bombsights. In: C. Eisenhart, M.W. Hastay and W.A. Wallis, eds. 1947. *Techniques of Statistical Analysis*. New York: McGraw-Hill. pp.113-184.

Ide, T., Lozano, A.C., Abe, N. and Liu, Y., 2009. Proximity-based anomaly detection using sparse structure learning. In: Ch. Apte, H. Park, K. Wang and M.J. Zaki, *Proceedings of the 2009 Society for Industrial and Applied Mathematics (SIAM) International Conference on Data Mining*. Sparks, Nevada. 30 April – 02 May. pp.97-108. DOI: 10.1137/1.9781611972795.9.

Inoh, J., Nagata, Y., Horita, K. and Mori, A., 2012. Prediction accuracies of improved Taguchi's T methods compared to those of multiple regression analysis. *Journal of The Japanese Society for Quality Control*, [e-journal] 42(2), pp.265-277. DOI: 10.20684/quality.42.2_265.

Jackson, J.E. and Mudholkar, G.S., 1979. Control procedures for residuals associated with principal component analysis. *Technometrics*, [e-journal] 21(3), pp.341-349. DOI: 10.1080/00401706.1979.10489779.

Jin, X. and Chow, T.W., 2013. Anomaly detection of cooling fan and fault classification of induction motor using Mahalanobis–Taguchi system. *Expert Systems with Applications*, [e-journal] 40(15), pp.5787-5795. DOI: 10.1016/j.eswa.2013.04.024.

Lauritzen, S.L., 1996. *Graphical Models*. Oxford: Oxford University Press.

Ohkubo, M. and Nagata, Y., 2017. The Mahalanobis–Taguchi method based on graphical modeling. *Japanese Journal of Applied Statistics*, [e-journal] 46(1), pp.13-26. DOI: 10.5023/jappstat.46.13.

Ohkubo, M. and Nagata, Y., 2018. Anomaly detection in high-dimensional data with the Mahalanobis–Taguchi system. *Total Quality Management & Business Excellence*, [e-journal] 29(9-10), pp.1213-1227. DOI: 10.1080/14783363.2018.1487615.

Park, S.H., Shin, W.S., Park, Y.H. and Lee, Y., 2017. Building a new culture for quality management in the era of the Fourth Industrial Revolution. *Total Quality Management & Business Excellence*, [e-journal] 28(9-10), pp.934-945. DOI: 10.1080/14783363.2017.1310703.

Peng, C.F., Ho, L.H., Tsai, S.B., Hsiao, Y.C., Zhai, Y., Chen, Q., Chang, L.C. and Shang, Z., 2017. Applying the Mahalanobis–Taguchi system to improve tablet PC production processes. *Sustainability*, [e-journal] 9(9), 1557. DOI: 10.3390/su9091557.

Porter, M.E. and Heppelmann, J.E., 2014. How smart, connected products are transforming competition. *Harvard Business Review*, 92(11), pp.64-88.

Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J. and Williamson, R.C., 2001. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7), pp.1443-1471.

Schwarz, G., 1978. Estimating the dimension of a model. *The Annals of Statistics*, [e-journal] 6(2), pp.461-464. DOI: 10.1162/089976601750264965.

Shin, W.S., Dahlgaard, J.J., Dahlgaard-Park, S.M. and Kim, M.G., 2018. A quality scorecard for the era of industry 4.0. *Total Quality Management & Business Excellence*, [e-journal] 29(9-10), pp.959-976. DOI: 10.1080/14783363.2018.1486536.

Taguchi, G., Chowdhury, S., and Wu, Y., 2005. *Taguchi's Quality Engineering Handbook*. Hoboken, NJ: Wiley.

Taguchi, G. and Jugulum, R., 2002. *The Mahalanobis–Taguchi Strategy: A Pattern Technology System*. New York: John Wiley & Sons.

Takahama, M. and Mikami, N., 2012. Detection of Abnormal Signs for Gas Turbine Power Plant. *Journal of Quality Engineering Society*, [e-journal] 20(4), pp.437-443. DOI: 10.18890/qes.20.4_45.

Woodall, W.H., Koudelik, R., Tsui, K.L., Kim, S.B., Stoumbos, Z.G. and Carvounis, C.P., 2003. A review and analysis of the Mahalanobis—Taguchi system. *Technometrics*, [e-journal] 45(1), pp.1-15. DOI: 10.1198/004017002188618626.

Yamanishi, K. and Takeuchi, J.I., 2002. A unifying framework for detecting outliers and change points from non-stationary time series data. In: O.R. Zaïane and R. Goebel, *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Edmonton Alberta, Canada, July, 2002. New York, NY: Association for Computing Machinery. pp.676-681. DOI: 10.1145/775047.775148.

ABOUT AUTHORS

Masato Ohkubo – (M.O.) Toyo University, Tokyo, Japan, Department of Business Administration, Faculty of Business Administration, e-mail: m.ohkubo@toyo.jp, Author's ORCID: 0000-0002-1885-6945.

Yasushi Nagata – (Y.N.) Waseda University, Tokyo, Japan, Department of Industrial and Management Systems Engineering, School of Creative Science and Engineering, e-mail: ynagata@waseda.jp, Author's ORCID: 0000-0002-4947-1204.

AUTHOR CONTRIBUTIONS

M.O. – conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, visualization; Y.N. – writing—review and editing, supervision; M.O. and Y.N. – project administration, funding acquisition.

CONFLICTS OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.