

## Research Article

# Applying Artificial Neural Network to Predict Semiconductor Machine Outliers

Keng-Chieh Yang,<sup>1</sup> Conna Yang,<sup>2</sup> Pei-Yao Chao,<sup>3</sup> and Po-Hong Shih<sup>3</sup>

<sup>1</sup> Department of Information Management, Hwa Hsia Institute of Technology, No. 111, Gongzhuang Road, Zhonghe District, New Taipei City 235, Taiwan

<sup>2</sup> Institute of Business and Management, National Chiao Tung University, No. 118, Section 1, Jhongsiao W. Road, Jhongjheng District, Taipei City 100, Taiwan

<sup>3</sup> Institute of Information Management, National Chiao Tung University, No. 1001, University Road, Hsinchu City 300, Taiwan

Correspondence should be addressed to Keng-Chieh Yang; [andesyoung.tw@gmail.com](mailto:andesyoung.tw@gmail.com)

Received 28 June 2013; Revised 10 October 2013; Accepted 15 October 2013

Academic Editor: Jung-Fa Tsai

Copyright © 2013 Keng-Chieh Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Advanced semiconductor processes are produced by very sophisticated and complex machines. The demand of higher precision for the monitoring system is becoming more vital when the devices are shrunk into smaller sizes. The high quality and high solution checking mechanism must rely on the advanced information systems, such as fault detection and classification (FDC). FDC can timely detect the deviations of the machine parameters when the parameters deviate from the original value and exceed the range of the specification. This study adopts backpropagation neural network model and gray relational analysis as tools to analyze the data. This study uses FDC data to detect the semiconductor machine outliers. Data collected for network training are in three different intervals: 6-month period, 3-month period, and one-month period. The results demonstrate that 3-month period has the best result. However, 6-month period has the worst result. The findings indicate that machine deteriorates quickly after continuous use for 6 months. The equipment engineers and managers can take care of this phenomenon and make the production yield better.

## 1. Introduction

Advanced semiconductor manufacturing processes are all made of very sophisticated machines. Requirements on these processes need hundreds of control parameters of the machine [1]. If a slight deviation of the key values changed, it may cause the process deviation (excursion). And then, the production of wafer may be in reduction or even scrapped. For the normal operation and maintenance of equipment to ensure production, failures of the equipments must be diagnosed correctly and timely.

The semiconductor manufacturing processes are usually through FDC to collect a large number of status variable identification (SVID) as data in real-time processes. But these pieces of information are often used to adjust the machine parameters after the event, as shown in Figure 1. Unit-variable control charts are important tools to diagnose abnormalities of process. Using these charts, engineers or managers can understand the quality of the wafer. If the output of numerical

data is between the upper bound (UCL) and lower bound (LCL), it means the quality is qualified. Otherwise, the quality is faults. There are many mathematical methods in engineering and management [2, 3]. To efficiently analyze SVID data, artificial neural networks (ANNs) can provide good results for further controls. Artificial neural backpropagation network has some functions, including possess learning, fault tolerance, and the parallel computing [4, 5]. Applying these functions, artificial neural backpropagation network can develop a predictive method for outliers' machines and thus help the overall enhancement of yield in semiconductor manufacturing [6].

The objective of this study attempts to propose an effectively predictive model to detect abnormal values of FDC. We apply the neural network model combining with historical data for yielding learning variables to analyze the results. On the other hand, we also use the gray theory to further analyze the results of ANNs.

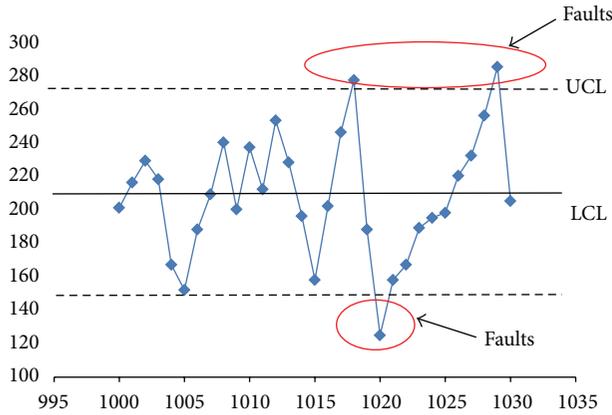


FIGURE 1: Univariable control chart.

Artificial neural networks are algorithms that can be used to perform nonlinear statistical modeling and provide a new alternative to logistic regression. Neural networks offer a number of advantages, including less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables, and the availability of multiple training algorithms. Disadvantages include its “black box” nature, greater computational burden, proneness to overfitting, and the empirical nature of model development [7].

Data analysis of this study can be divided into two stages: the data processing stage and the network training. In the data processing stage, in order to effectively control the constructing complexity of the neural network, this study applies the principal component analysis (PCA) and selects the stepwise variable to reduce the dimensions of the input variable. In the network training stage, we use the backpropagation neural network and gray relational analysis to detect the accuracy on prediction of model for the machine outlier. The research processes are shown in Figure 2.

## 2. Manufacturing Process Quality Control in Foundry

Manufacturing process quality control is a procedure or set of procedures intended to ensure that a product or service adheres to a defined set of quality criteria or meets the requirements of the client or customer. In this section, we introduce quality control methods in the semiconductor wafer manufacturing process. On the other hand, we also introduce the basic concepts of FDC in the semiconductor industry.

**2.1. Semiconductor Wafer Manufacturing Quality Control Methods.** In the semiconductor producing process, wafer manufacturing quality control methods can be classified into work-in process test (in-line) and the control wafer (dummy wafer) testing machine (off-line) [8, 9]. The former is done directly on the wafer testing. Wafer testing in the manufacturing process for its execution timing can be divided into

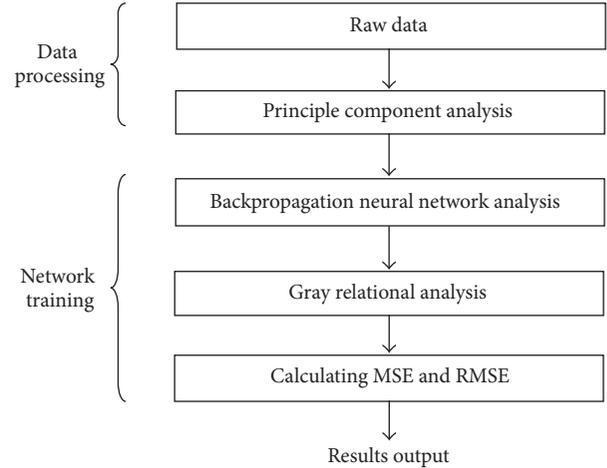


FIGURE 2: Research processes.

the front-end processing (front-end) for visual inspection, defect analysis (defect scan), wafer acceptance test (WAT), and posterior segment (back-end) test. The latter is based on the test piece for the machine to carry out its process capability. In this test, it usually obtains information entered to statistical process control (SPC) system. These quality control or testing methods are described below [10, 11].

- (1) Inspection: this is the appearance of deflections observed in the manufacturing site. Workers can view wafers in visual appearance or microscopic view. This is typically applying the sampling method and the information obtained can be qualitative or quantitative counts. The qualitative SPC data are not normally within the scope.
- (2) Offline measuring machines testing: this function is to simulate the dummy piece of result of the machine processes. Almost all of the semiconductor producing machines have this testing mode. When the wafer has been made completely, it will leave the producing machine and be immediately moved into the measuring machine. This is a SPC system with quantitative data entry. At this stage, this system is handled by the operator. Sometimes, computer-integrated manufacturing (CIM) systems will be placed in the counting function. If it exceeds its execution cycle, the CIM consignment will reject the manufacturing of this machine. However, in the advanced 300 mm factories, the above processes are run automatically and then detect the wafer to gain the data. This can reduce personnel operation and malfunction.
- (3) Defect analysis: this uses defect analysis instruments to scan the surface of wafer, typically applying sampling method. Information is obtained for the count of the number of defects.
- (4) Wafer acceptance test (WAT): when designing the electronic circuit, the test point for electrical testing has been placed. A wafer has five testing points and

each point represents one-fifth of the area that must be within die quality control.

- (5) Die test: this function are run in testing house. The testing machine detects each die in the maximum resolution, but the feedback is time-consuming.

The SPC system is one of the functions in MES system. It is also commonly used in semiconductor industry for quality control practices. In the producing process, the changes of product size are inevitable [12, 13]. Changes are divided into two types: normal and abnormal. The normal change is the inevitable factor. It has little effect on product quality. This situation is usually difficult to exclude. It is believed that the manufacturing process is affected by many factors beyond the control of variations. These variations are usually very small, and the impact on the quality is not great. In statistical quality control, these factors are called chance cause or common cause. Manufacturing process may be also influenced by some special factors (such as machine failures, operator error, or poor materials) and cause a large variation. Therefore, a great impact on the quality will lower down the quality level. These factors are called assignable causes or special causes. SPC uses control charts to detect events in the manufacturing process [13]. Another purpose of changes of the parameter in the manufacturing process is to eliminate or to avoid abnormal events, making the process in a normal state [9, 11].

Process control chart is usually recorded in the work-in product with the measurement data (inline monitor data). It also recorded the Dummy wafer with the measurement data (offline monitor data) [14]. In this chart, we calculate the sample statistics, such as mean and standard deviation. Besides, we randomly choose the sample and then input them into the control chart to determine whether the regulatory process is within the state. Finally, we get the results of capability of accuracy (Ca), capability of precision (Cp), process capability (Cpk), and so forth. We use Western Electric Rules to monitor stability of the process [15]. A typical control chart is composed of a center line [16] and two control limits: upper control line (UCL) and lower control line (LCL), as shown in Figure 3. SPC will determine when the data are out of the control line (UCL/LCL), and the engineers will take the warning for urgent treatment. This can improve the quality for better process control [4, 10].

**2.2. FDC in the Semiconductor Industry.** FDC contains two functions: fault detection and the fault classification. Engineers focused on results of fault detection testing to take some necessary actions. Different fault requires different corrective actions, while fault classification function is classified based on statistics eigenvalues. So engineers can quickly refer to the machine error code and restore the machine to normal state within the least time [4, 6, 8].

In the semiconductor wafer process, when the machine produces a certain number of wafer, some parameters will drift from original ones. So, at this moment, FDC can detect deviations within a short time. When the parameters deviate from the original value, and may be beyond the range of the set interval, the run-to-run are needed to be applied

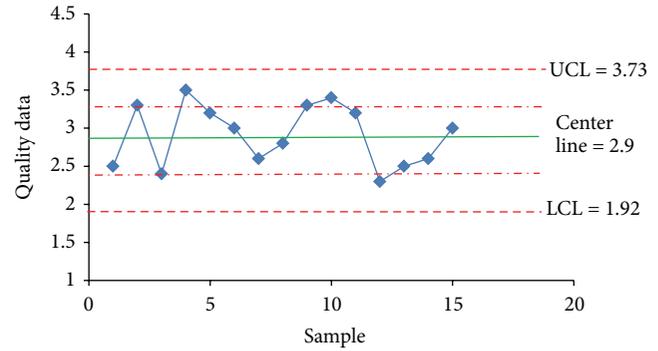


FIGURE 3: SPC control chart [15].

adjustments to modify the parameters directly and continuously collecting the running parameters of the machine and constant feedbacks. Based on the previous activities of quality control, engineers can adjust the machine parameters to ensure that the production is within normal operations [8].

Engineers use the FDC monitor to ensure the correctness of information of production status, including the manufacturing process, machine operating conditions, parameters, and use of the recipe. Engineers must check the machine status before the operation has issues; otherwise, when the production finishes, it will cause the business loss. The FDC monitor can avoid the waste of production capacity, reduce failures, and ensure the producing yield increase.

### 3. Methods

In this section, we introduce the researching method. First, we introduce artificial neural networks as the main researching method. Second, we adopt backpropagation neural network as the researching method to analyze the semiconductor manufacturing machines outliers. Lastly, after the ANN results, we use gray relational analysis to further justify the results.

**3.1. Artificial Neural Networks (ANNs).** Artificial neural networks are one kind of information processing systems that mimic biological neural networks. ANNs are defined as “computing systems that include software and hardware and use a lot of simple artificial neurons connected to mimic biological neural network artificial neurons [17].” These networks are simple simulation of biological neurons, which get information from the outside environment or other artificial neurons, making a very simple operation, and output the result to the outside environment or other artificial neurons [12, 13].

In other words, artificial neurons are computational models illuminated from the natural neurons. A natural neuron receives signals through synapses that are located on the dendrites or membrane of the neuron. When the neuron receives signals, it will be activated and emits a signal through the axon. This signal might be sent to another synapse and/or might activate other neurons [4].

The real neurons are complex when modeling artificial neurons. They are consisted of inputs, which are multiplied by weights, and then computed by a mathematical function which decides the activation of the neuron. Another function computes the output of the artificial neuron. ANNs combine artificial neurons in order to process information [11].

The strength of ganglion biological neural networks is the place to store information. Biological neural network learning is to adjust the intensity of ganglion [12]. Therefore, we can say that nerve cells in the input path tree pass through a lot of contact between the cell body to accept the ganglion cells around the body of the outgoing signals, and body axis of nerve cell is equivalent to the output path. We can transform information outside to the input signal into the input vector  $I_i$  and compute with the weighting value  $W_i$ . Artificial neural system can be divided into two parts. The front-end is a summation function of the input vector to be integrated and then the rear section by a simple transferring function for the message output. Finally, the output vector  $Y$  can be the form of other neurons as input. Transferring function can normally be sigmoid function [11, 12].

Artificial neurons that have the same function constitute a layer. In general, the structure of neural network includes input layer, hidden layer and output layer, wherein the input layer, and output layer are constituted with a single layer, but the hidden layer may have more layers, depending on the complexity of the problem.

**3.2. Backpropagation Neural Network.** Backpropagation neural network (BPN) model is a learning model in the neural network and the most representative one. Compared with perceptron network, backpropagation neural network has the following improvements [5, 18].

- (1) Increases the hidden layer: hence, this can perform the interaction between the processing units.
- (2) Uses a smooth differentiable transfer function: Therefore, the network can be applied to the steepest descent method and derives weighted correction formula for the network.

Backpropagation neural network includes the following [5].

- (1) Input layer: this layer is for the input of variables. The number of processing units is based on the complexity of the problem. This function uses a linear transferring function.
- (2) Hidden layer: this layer is for the input processing unit interacting between its processing units. We usually use the trial and error method to determine the number of processed units in this layer. This function usually uses a nonlinear transferring function.
- (3) Output layer: this layer is for the output of variables. The number of processing units is based on the complexity of problem. This function uses a nonlinear transferring function.

The most commonly used nonlinear conversion function of backpropagation neural network is sigmoid function, as

shown in formula (1). This function tends to a constant value when the dependent variable tends to positive/negative infinity [18]. The function value is often within the range  $[0, 1]$ :

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (1)$$

Backpropagation neural network applies Widrow-Hoff learning rule to generalize the multidifferentiable nonlinear transferring function [19, 20]. backpropagation neural network has partial weight (b) and the hidden layer is hyperbolic transferring function. The output layer is a linear transferring function. Using the known input vector and its corresponding target vector, together with a sufficient number of neurons in the hidden layer, this will enable the network approximate a finite number of discontinuities in any function [3, 17]. When appropriately trained backpropagation neural network is given new input vector, the network will calculate a reasonable output. Using generalized characteristic in the network, the new input vector can calculate output vector. In other words, when generalized characteristic of the network is achieved, we can use nontraining data in the network and this can produce a satisfactory output [18].

Multilayer networks for propagation algorithm is a generalized least mean squares (LMS) algorithm, and the backpropagation algorithm and LMS algorithms used mean square error (MSE) as performance indicators [18]. When each input vector is entered into the network, we can compare the gap between the network output and the target output to adjust the settings in the network variables. It generally uses minimum mean square error to measure the quality of learning [18]:

$$E = \frac{1}{N} \sum_j (t_j - a_j)^2, \quad (2)$$

where  $t$  is the target vector of the output layer and  $a$  is the output vector of the output layer.

Network learning is to minimize the error function. We usually use gradient decent to reach this goal. When entering a training example, the network slightly adjusts weights. This can make the size of the error function smaller and the sensitivity of the weight value proportional. In other words, the error function with weights is proportional to partial differential value [19]:

$$\Delta W_{ij} = \eta \frac{\partial E}{\partial W_{ij}}, \quad (3)$$

where  $W_{ij}$  is the weighted value between  $(n - 1)$ <sub>th</sub> layer within the  $i$ <sub>th</sub> processing unit and  $n$  layer of the  $j$ <sub>th</sub> processing unit.  $\eta$  is the learning rate. It is used to control the gradient decent method to minimize the error function. Gradient decent method is shown as follows [17, 19]:

$$\begin{aligned} \Delta W_{ij}^m &= \eta \frac{\partial E}{\partial W_{ij}^m}, \\ \Delta b_i^m &= \eta \frac{\partial E}{\partial b_i^m}. \end{aligned} \quad (4)$$

Using the chain rule we can obtain:

$$\begin{aligned}\frac{\partial E}{\partial b_{ij}^m} &= \frac{\partial E}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial W_{ij}^m}, \\ \frac{\partial E}{\partial b_i^m} &= \frac{\partial E}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial b_{ij}^m},\end{aligned}\quad (5)$$

where  $n$  is the activity function:

$$n_i^m = \sum_{j=1}^{s^{n-1}} W_{ij}^m a_j^{m-1} + b_i^m. \quad (6)$$

**3.3. Gray Relational Analysis.** Gray relational analysis is a system that includes quantitative description and comparative methods. A gray relation system means that a system where part of information is known, but part is unknown. In this situation, information quantity and quality form a continuum from a total lack of information to complete information—from black through grey to white [16, 21]. In this uncertain situation, one is always somewhere in the middle, somewhere between the extremes, and somewhere in the gray area.

In the gray relational analysis, if the gaps of the reference range are excessive, certain factors will be ignored. When the direction of each factor is inconsistent, the results may cause the deviation. Hence, we must do data preprocessing for the raw data [22]. We can do the initializing, averaging, or internalizing for the raw data. Each element of a sequence can satisfy two conditions: comparability and nonparallelism [21].

When we develop a gray relation system, if the trend of two factors is consistent, this means a higher degree of simultaneous change. It represents a higher degree of association between the two factors. Therefore, the gray relational analysis method is based on the similarity between factors or difference between them (i.e., gray relational degree) [21].

In the gray relation analysis, first we set the largest element in the matrix  $\Delta$  as  $\Delta_{\max}$  and the smallest element as  $\Delta_{\min}$ . And then, we define the gray relational coefficient  $\zeta$ , which is between 0 and 1 (this coefficient is decided by policymakers, usually set to 0.5). Finally, we calculate the gray relation coefficient  $\gamma_{0i}(k)$ , which is defined as follows [21]:

$$\gamma_{0i}(k) = \frac{\Delta_{\min} + \zeta \Delta_{\max}}{\Delta_{0i}(k) + \zeta \Delta_{\max}}. \quad (7)$$

Then, we take the average of gray relational coefficient gray relational degree  $\gamma(x_0, x_i)$ :

$$\gamma(x_0, x_i) = \frac{\sum_{k=1}^n \gamma_{0i}(k)}{n}, \quad (8)$$

where  $\gamma(x_0, x_i)$  represents the  $i_{\text{th}}$  comparison sequences (independent variables) on the reference sequence (the dependent variable) relation degree. Finally, we can sort the compared sequences and each reference sequence. This can explain the relation between variables and the system performance.

## 4. Experiment Results

This study adopts backpropagation neural network model and gray relational analysis for data analysis. We use a semiconductor machine as the experiment tool to analyze the results in the manufacturing process. The return value of this machine is as a network training inputting data for the analysis. This study uses Novellus Vector Machine and its Remote Process Controller (RPC) function to collect the data. The data collection period is between April 2008 and December 2011.

In this study, we use MATLAB 7.0, the neural network toolbox, to analyze the data [22]. Experiment processes include data preprocessing, network variables setting, the hidden layer neurons determination, the selection of the best combination of input variables, network output results determining principles, and sensitivity analysis [23, 24].

This study detects the gas transmission pressure of chamber. If the numerical data exceeds the upper bound or lower bound, this may cause the product failure. The gas may cause uneven distribution of the wafer surface, and the chemical change is likely to make the solid on the wafer surface not uniformly flat. This would seriously affect the chip yield. Therefore, this study adopts Novellus Vector Machines as the experiment target and observes the gas delivery pressure coefficient in the chamber for the research data source.

### 4.1. Backpropagation Neural Network Parameters Settings.

The setup of network parameters in an artificial neural network includes learning trials, learning rate, and the momentum correction coefficient. By setting up the learning rate at 0.1–0.3 for dynamic change, and the momentum correction coefficient at 0.01–0.02, the specific setup values differ according to the status of network convergence. When the MSE of the learning trail diminishes by 10–3/3000 epochs, the network learning process is completed. An average rate would be around 1,000–3,000 to achieve such standards [24].

Since the application of forward selection procedure will change the number of neurons in the hidden layer which leads to different groups of combinatorial optimization, this application should determine the number of neurons in the hidden layer prior to the selection process to prevent the process from being overly complicated.

To decide the number of neurons in the hidden layer, we should take the previously selected variable and substitute it into the network for training. Network training requires the exactly same setup which includes learning trials, learning rate, and the momentum correction coefficient for a final decision. In this study, the number of neurons in the hidden layer is tested by setting the range at 1–2 times of layers in the network. For example, if there were 10 input variables, then the number of layers would be 10. This research tests 1 to 20 neurons in the hidden layer by achieving the optimal result for the forward selection procedure that follows.

**4.2. Model Building.** Using technology of the process and monitor recipe as units, each of them has its own training dataset and artificial neural backpropagation network model.

The next step would be to analyze the characteristics of the data input. Specifically, the analysis should emphasize the column input and the column to see if it is global or partial. This means that it should be checked if the input space is mainly centered on specific areas and return its relationship with time. Also, one should seek the opinions of the process experts on whether the same input at a different period of time would lead to different results. In addition, to complete the process, it is necessary to discuss with process experts the prediction for each record and the acceptable period of time for building each training model. After the data analysis, if the number of data is insufficient, we need to go back to data preprocessing step to either gather more data or to decrease the input parameter dimensions.

The first artificial neural backpropagation network model that might be applicable would be the one that needs to take factors into consideration, such as the time building model and effectiveness. Also, the training data records, global/partial characteristics of the data, and number of parameters needed to include are all necessary elements that help one decide which type of artificial neural backpropagation network model could be used. When the parameters in the neuron model are too much and the dataset is not large enough, we must go back to the data preprocessing step as mentioned earlier to see if this problem can be solved by using an algorithm with lower efficiency or a model with less neurons.

When the model training is completed, it should be compared with the parameters offered by the machine and be checked to see if it falls within the error range. If it does not, then the training requires to be processed again.

Furthermore, the MSE of the model serves as an indicator too. When the model is lack of confidence and requires retraining, several markers, such as the confidence index, machine parameters, recipe, environment sensor parameter, and the measured figures, should be served as support whether or not to go through the training again. The decision of when to rebuild the model could be categorized into 2 situations: when the machine is undergoing maintenance or becoming low in accuracy.

- (1) Machine maintenance: when the machine is undergoing maintenance or preventive maintenance, this could change the machine situation. By using the actual parameters as a comparison, we could verify if the artificial neural backpropagation network model is still in an acceptable range.
- (2) Low accuracy: the actual parameters will be calculated in a fixed cycle. The calculated parameters could help determine if the model is still acceptable.

**4.3. Model Training.** There can be slight difference among machines after countless times of machine maintenance and wear and tear. Thus, network training should be conducted by the different selection of network input variables. For instance, data collected for the past six months, three months, and one month should be analyzed to observe the difference in each variable.

(1) *One-Month Network.* 790 records of data were collected from the transferring pressure of the Novellus Vector December 1, 2010 to December 31. The tested data retrieved from January 1, 2011 to January 5 consisted of 72 records. The learning times were set to 3,000, learning rate set to 0.1, and the momentum correction coefficient set to 0. Since the training results of the artificial neural backpropagation network do not always have the same result, this study conducts the same experiment 10 times to ensure the network stability.

The results of the tests show that each round of the training and tests is slightly different. However, all the results converge within the 1,100th cycle. As shown in Figures 5(a) and 5(b), the overall performance is fairly well with the MSE = 0.01641 and RMSE (Train-R) = 0.59956. Table 1 part (a) shows that by comparing the Train-R data and target data and depicting prediction output figure of the network training, the default mode and outlier can be predicted.

(2) *Three-Month Network.* 2,036 records of data were collected from the Transfer Pressure of the Novellus Vector from October 1, 2010 to December 31. The test data retrieved from January 1, 2011 to January 5 consisted of 72 records. The learning times were set to 3,000, learning rate set to 0.1, and the momentum correction coefficient set to 0. Since the training results of the artificial neural backpropagation network do not always have the same result, this study conducts the same experiment 10 times to ensure the network stability.

The results of the tests show that each round of the training and tests is slightly different. However, all the results converge within the 1,400th cycle. As shown in Figures 5(c) and 5(d), the overall performance is fairly well with the MSE = 0.01406 and RMSE (Train-R) = 0.66066. Table 1 part (b) shows that by comparing the Train-R data and target data and depicting prediction output figure of the network training, the default mode and outlier can be predicted.

(3) *Six-Month Network.* 4,214 records of data were collected from the Novellus Vector from July 1, 2010 to December 31 of the transfer pressure. The test data retrieved from January 1, 2011 to January 5 consisted of 72 records. The learning times were set to 3,000, learning rate set to 0.1, and the momentum correction coefficient set to 0. Since the training results of the artificial neural backpropagation network do not always have the same result, this study conducts the same experiment 10 times to ensure the network stability.

The results of the tests show that each round of the training and tests is slightly different. However, all the results converge within the 1,700 cycle. As shown in Figures 5(e) and 5(f), the overall performance is fairly well with the MSE = 0.01725 and RMSE (Train-R) = 0.55732. Table 1 part (c) shows that by comparing the Train-R data and target data and depicting prediction output figure of the network training, the default mode and outlier can be predicted.

We can see from Figure 4 that the parameters in the training process have a fairly well learning effect. However, the extreme values of learning ability are near to perfect with no influence from the network selection. Figure 4 shows

TABLE 1: Network training and performance tests.

Hidden neuron	1 month (a)		3 months (b)		6 months (c)	
	MSE	Train-R	MSE	Train-R	MSE	Train-R
1	0.0162	0.5873	0.0132	0.6773	0.0173	0.5643
2	0.0154	0.6032	0.0136	0.6832	0.0163	0.5932
3	0.0165	0.6244	0.0142	0.6834	0.0166	0.4934
4	0.0166	0.5873	0.0136	0.6583	0.0169	0.5274
5	0.0169	0.5972	0.0139	0.6572	0.0162	0.5843
6	0.0157	0.5878	0.0147	0.6978	0.0164	0.5032
7	0.0168	0.5983	0.0148	0.5783	0.0177	0.5836
8	0.0163	0.5973	0.0143	0.6373	0.0179	0.5787
9	0.0168	0.6036	0.0138	0.6439	0.0185	0.5897
10	0.0169	0.6092	0.0145	0.6899	0.0187	0.5554
Average	0.01641	0.59956	0.01406	0.66066	0.01725	0.55732

Denote: MSE is mean square error; Train-R is correlation coefficient.

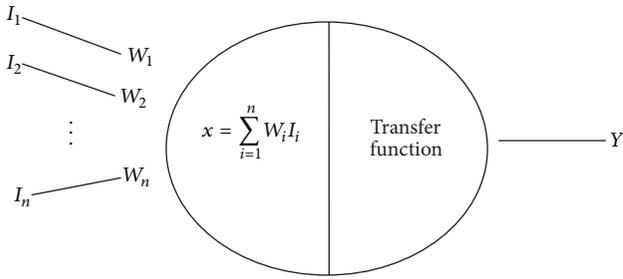


FIGURE 4: Artificial neural networks [11].

the change in the prediction results and actual concentration values. We can see that the artificial neural backpropagation network model performs better in extreme values, especially in very small values.

**4.4. Discussion.** This study uses the mean square error (MSE) of the training output data to analyze if training results are acceptable and converge. The MSE (9) can be seen below. The smaller for MSE, the better the network training results. In the network training process, the MSE might be unstable if the network is not converged. To improve such a situation, the network learning rate and momentum correction coefficient should be readjusted to reach convergence.

This study analyzes the network training and prediction results by using the correlation coefficient  $R$  and MSE (10). The  $n$  in the equation signifies the total number of data inputs,  $\mu$  denotes the arithmetic mean,  $\sigma$  denotes the standard deviation, the subscript  $i$  denotes the number of data, and subscripts  $t$  and  $a$  denote the actual value and the network output value. This study analyzes the network output by calculating the correlation coefficient of the network trail output and the actual value and chooses the higher one as the optimal result:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^n (t_i - a_i)^2, \quad (9)$$

$$R = \frac{(\sum t_i a_i) n \mu_t \mu_a}{(n-1) \sigma_t \sigma_a}. \quad (10)$$

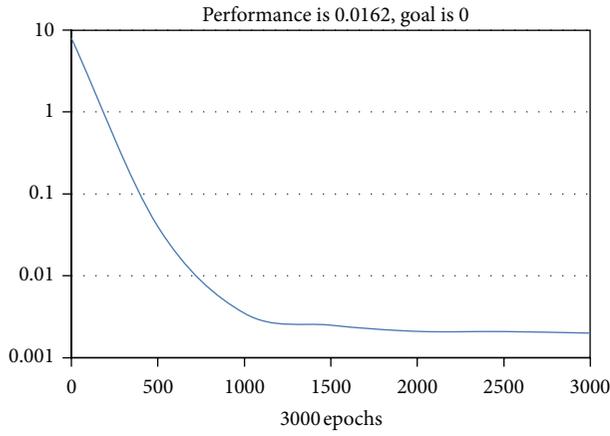
On the other hand, our research uses Novellus Vector Machine and its Remote Process Controller (RPC) function to collect the data. This study detects the gas transmission pressure of chamber. If the numerical data exceeds the upper bound or lower bound, this would cause the product failure. The gas may cause uneven distribution of the wafer surface, and the chemical change is likely to make the solid on the wafer surface not uniformly flat. This would seriously affect the chip yield.

After countless times of machine maintenance and wear and tear, the machines will need to be maintained to perform well again. After long functioning periods, the machines are likely to decrease in performance. Thus, the facility engineer should be aware of such a situation. The average time of maintenance for the machine would be around 3 months. It will take several times of adjustment to reach its original performance. Therefore, we suggest that the prediction and model training be around 3 months.

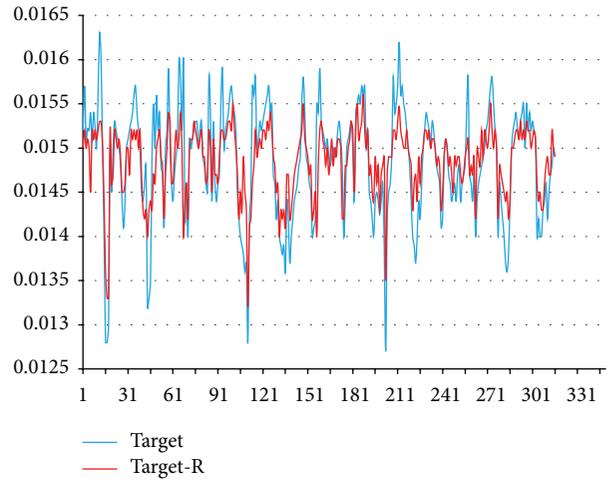
Our experimental results show that three-month period of network training data possesses the best results. Because the machine needs maintenance, the stability of machine in the first month is lower than that of three months. However, 3 months are the most stable situation in our study. Because the machine has been smoothly working, we get a better result in this experiment. But in the 6 months, the training results indicate that the MSE begins to deteriorate. This means that the machine needs maintenance again.

This study proposed neural networks as the research method to analyze the semiconductor machine outliers. Neural network analysis has been validated to show the capability of analyzing the plasma processing equipment [25], reactive ion etching [26], plasma etch process [27], chamber leak detector of plasma processing equipment [28], and so forth.

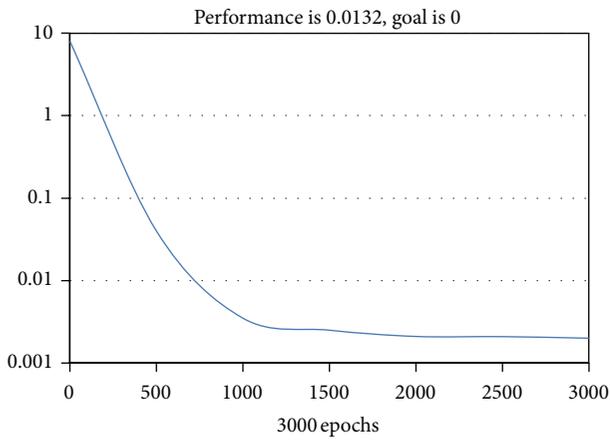
We believe that neural network method can provide an effective way to analyze the semiconductor machine outliers. In the previous studies, they seldom used neural network analysis with data mining technique for further analysis. This study clearly indicates the experimental results for practitioners and scholars as references.



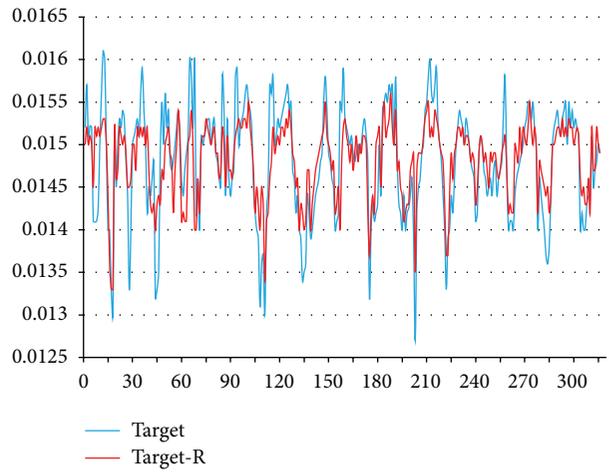
(a) Network convergence figure of one month



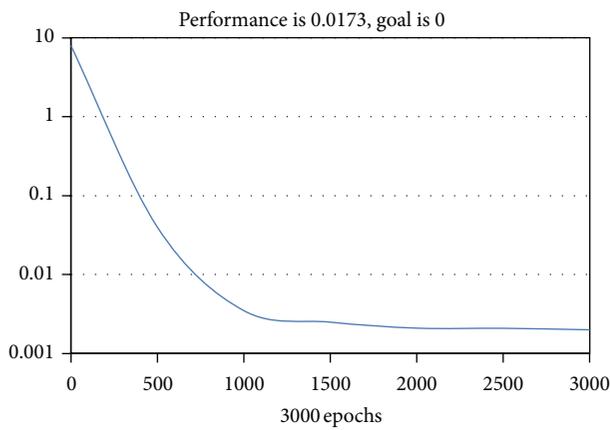
(b) Network training prediction output figure of one month



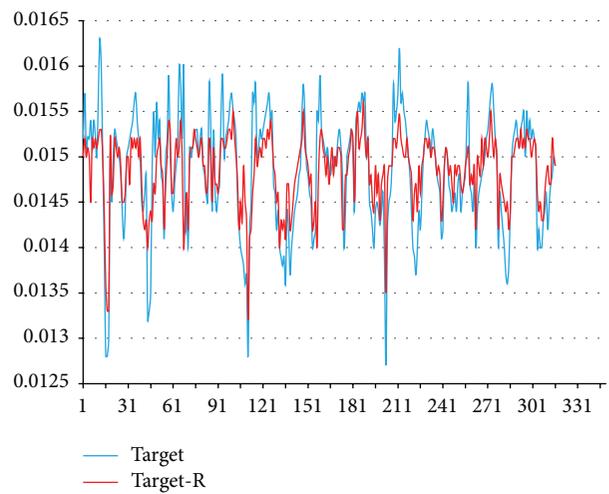
(c) Network convergence figure of 3 months



(d) Network training prediction output figure of 3 months



(e) Network convergence figure of 6 months



(f) Network training prediction output figure of 6 months

FIGURE 5: Network convergence and network training prediction output.

## 5. Conclusion

This study uses the artificial neural backpropagation network model to detect the outliers in semiconductor machines. Due to the complicity of the technology in the process and many types of machines in the semiconductor industry, we chose the most often used machine, Novellus Vector, which also had the hardest gas control and pressure control for the network training.

In the researching process, we have faced problems due to incompleteness or missing of the machine data. Due to the restriction of time and ability, the training of the artificial neural backpropagation network model is still not flawless. Therefore, we offer several suggestions for future studies.

- (1) This study uses Novellus Vector machines to conduct network training. But the semiconductor machines vary into many different types; we hope that in the future studies can involve more types of machines to test the results of this study.
- (2) By effectively controlling the abnormal situations in the machines, we can increase the yield rate that is one of the most important missions in the semiconductor industry. This study focuses on how to predict the outliers in the semiconductor machines. In the technology control system of the advanced process, immediate reactions are important. If the artificial neural backpropagation network model can detect the outliers using the automatic control system, and perform a real-time correction procedure, this will increase the yield rate greatly.

## Acknowledgments

The authors appreciate the anonymous referees for their constructive comments to improve this paper. They also appreciate the Guest Editor Professor Jung-Fa Tsai for his kind service.

## References

- [1] N. Kumar, K. Kennedy, K. Gildersleeve, R. Abelson, C. M. Mas-trangelo, and D. C. Montgomery, "A review of yield modelling techniques for semiconductor manufacturing," *International Journal of Production Research*, vol. 44, no. 23, pp. 5019–5036, 2006.
- [2] M.-H. Lin, J.-F. Tsai, and C.-S. Yu, "A review of deterministic optimization methods in engineering and management," *Mathematical Problems in Engineering*, vol. 2012, Article ID 756023, 15 pages, 2012.
- [3] H.-Y. Kao and C.-H. Huang, "Modeling supply chain diagnostics with fuzzy dynamic bayesian networks," *International Journal of Industrial Engineering*, vol. 15, no. 3, pp. 257–265, 2008.
- [4] L. Wang and K. Fu, *Artificial Neural Networks*, Wiley Online Library, 2008.
- [5] A. T. C. Goh, "Back-propagation neural networks for modeling complex systems," *Artificial Intelligence in Engineering*, vol. 9, no. 3, pp. 143–151, 1995.
- [6] E. D. Karnin, "Simple procedure for pruning back-propagation trained neural networks," *IEEE Transactions on Neural Networks*, vol. 1, no. 2, pp. 239–242, 1990.
- [7] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of Clinical Epidemiology*, vol. 49, no. 11, pp. 1225–1231, 1996.
- [8] K. Kerdprasop and N. Kerdprasop, "Computational intelligence techniques to fault detection in the semiconductor manufacturing process," *Applied Mechanics and Materials*, vol. 52–54, pp. 1171–1176, 2011.
- [9] G. Smith, *Statistical Process Control and Quality Improvement*, vol. 576, Prentice Hall, 1998.
- [10] C.-M. Fan, R.-S. Guo, S.-C. Chang, and C.-S. Wei, "SHEWMA: an end-of-line SPC scheme using wafer acceptance test data," *IEEE Transactions on Semiconductor Manufacturing*, vol. 13, no. 3, pp. 344–358, 2000.
- [11] Y. E. Shao and C. C. Chiu, "Developing identification techniques with the integrated use of SPC/EPC and neural networks," *Quality and Reliability Engineering International*, vol. 15, pp. 287–294, 1999.
- [12] C.-C. Chiu, Y. E. Shao, T.-S. Lee, and K.-M. Lee, "Identification of process disturbance using SPC/EPC and neural networks," *Journal of Intelligent Manufacturing*, vol. 14, no. 3-4, pp. 379–388, 2003.
- [13] Y. E. Shao, C.-J. Lu, and C.-C. Chiu, "A fault detection system for an autocorrelated process using SPC/EPC/ANN AND SPC/EPC/SVM schemes," *International Journal of Innovative Computing, Information and Control*, vol. 7, no. 9, pp. 5417–5428, 2011.
- [14] G. Wenski, T. Altmann, W. Winkler, G. Heier, and G. Hölker, "Doubleside polishing—a technology mandatory for 300 mm wafer manufacturing," *Materials Science in Semiconductor Processing*, vol. 5, no. 4-5, pp. 375–380, 2002.
- [15] J. S. Hunter, "A one-point plot equivalent to the Shewhart chart with Western Electric rules," *Quality Engineering*, vol. 2, pp. 13–19, 1989.
- [16] C. L. Lin, J. L. Lin, and T. C. Ko, "Optimisation of the EDM process based on the orthogonal array with fuzzy logic and grey relational analysis method," *International Journal of Advanced Manufacturing Technology*, vol. 19, no. 4, pp. 271–277, 2002.
- [17] K. Mehrotra, C. K. Mohan, and S. Ranka, *Artificial Neural Networks*, The MIT Press, 1997.
- [18] F.-C. Chen, "Back-propagation neural networks for nonlinear self-tuning adaptive control," *IEEE Control Systems Magazine*, vol. 10, no. 3, pp. 44–48, 1990.
- [19] M. T. Hagan, H. B. Demuth, and M. H. Beale, *Neural Network Design*, Pws Pub, Boston, Mass, USA, 1996.
- [20] B. Widrow and M. A. Lehr, "30 years of adaptive neural networks: perceptron, Madaline, and backpropagation," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1415–1442, 1990.
- [21] J. W. K. Chan and T. K. L. Tong, "Multi-criteria material selections and end-of-life product strategy: grey relational analysis approach," *Materials and Design*, vol. 28, no. 5, pp. 1539–1546, 2007.
- [22] H.-H. Lai, Y.-C. Lin, and C.-H. Yeh, "Form design of product image using grey relational analysis and neural network models," *Computers and Operations Research*, vol. 32, no. 10, pp. 2689–2711, 2005.

- [23] P. Vas, *Artificial-Intelligence-Based Electrical Machines and Drives: Application of Fuzzy, Neural, Fuzzy-Neural, and Genetic-Algorithm-Based Techniques*, vol. 45, Oxford University Press, 1999.
- [24] H. Demuth and M. Beale, *Neural Network Toolbox For Use With MATLAB*, The MathWork, 1993.
- [25] B. Kim and D. Kim, "Use of neural network to in situ conditioning of semiconductor plasma processing equipment," *Applied Soft Computing Journal*, vol. 12, no. 2, pp. 826–831, 2012.
- [26] S. J. Hong, G. S. May, and D.-C. Park, "Neural network modeling of reactive ion etching using optical emission spectroscopy data," *IEEE Transactions on Semiconductor Manufacturing*, vol. 16, no. 4, pp. 598–608, 2003.
- [27] J. P. Card, M. Naimo, and W. Ziminsky, "Run-to-run process control of a plasma etch process with neural network modelling," *Quality and Reliability Engineering International*, vol. 14, no. 4, pp. 247–260, 1998.
- [28] B. Kim and S. Kwon, "Wavelet-coupled backpropagation neural network as a chamber leak detector of plasma processing equipment," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6275–6280, 2011.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

