JJMIE

Volume 5, Number 4, Aug. 2011
ISSN 1995-6665
Pages 353 - 357

Jordan Journal of Mechanical and Industrial Engineering

Modeling of the MEMS Reactive Ion Etching Process Using Neural Networks

M. Ashhab*[,a], N. Talat[b]

[a]*Department of Mechanical Engineering, The Hashemite University, Zarqa 13115, Jordan*

[b]*Mechanical Engineering Department, University of Jordan, Amman 11942, Jordan*

Abstract

Reactive ion etch (RIE) is commonly used in microelectromechanical systems (MEMS) fabrication as plasma etching method, where ions react with wafer surface substrate in plasma environment. Due to the importance of RIE in the MEMS field, two prediction models are established to predict the wafer status in reactive ion etching process: back-propagation neural network (BPNN) and principle component analysis BPNN (PCABPNN). These models have the potential to reduce the overall cost of ownership of MEMS equipment by increasing the wafer yield, and not depend upon monitoring wafers or expensive metrology rather it will enable inexpensive real-time wafer-to-wafer control applications in RIE. The artificial neural net (ANN) is trained with historical available input-output process data. Once trained, the ANN forecasts the process output rapidly if given the input values.

Keywords: MEMS; Reactive ion etching; Modelling; Neural networks

1. Introduction

Micro-Electro-Mechanical Systems (MEMS) is an emerging technology, which uses the tools and techniques that were developed to build microscopic machines, sensors, actuators, and electronics. MEMS enable expanding the space of possible designs and applications of nearly every product category on a common silicon substrate through microfabrication technology. MEMS devices are manufactured using batch fabrication techniques similar to those used for integrated circuits. The fabrication of modern semiconductor products requires thousands of processing steps. A key element in achieving high yields during semiconductor fabrication is to minimize the amount of defected wafers. Therefore, detecting the defected wafers is a very important issue.

Neural networks have seen an explosion of interest over the last few decades. They have been utilized to study the possibility of designing control techniques to significantly improve the performance of the RIE process [1]. Optical emission spectroscopy data were used to construct neural network models of plasma etch process in [2]. A virtual metrology system for MEMS is proposed in [3] that fulfills real-time quality measurement of each wafer and detects the performance degradation of the corresponding machines from the information of manufacturing processes. Neural networks are being successfully applied in microfabrication inspection systems [4-8]. The major reason for adopting neural networks is because neural networks have potential capability in modelling and control of non-linear systems. Moreover neural networks have the ability of learning arbitrary nonlinear mappings between noisy sets of input and output data. Back-propagation neural network (BPNN) is currently the most popular learning rule used in supervised learning, which is also known as feed forward neural network and multilayer perceptron (MLP).

Back propagation is a very powerful tool with application to solve the problems of prediction, optimization, control, and diagnosis in the MEMS manufacturing processes [9-12]. Most of the literature adopt BPNN because it has the advantages of an easier-comprehended theory, faster recalling speed and higher learning accuracy. However, the determination of the structure architecture and the parameters under this network is difficult.

Many researchers have studied pattern classification by using BPNN for the automatic inspection system in the MEMS industry [13-15]. Zoroofi et al. [13] used curve recognition to detect the contamination on a wafer surface during semiconductor production. Three conventional classification models: a back-propagation technique, a minimum distance algorithm and a maximum likelihood classifier, were used and the performance of these three models was compared. The results showed that the back-propagation classifier has a better classification performance. Su et al. [14] proposed a neural-network approach for semiconductor wafer post-sawing inspection. BPNN, radial basis function network (RBFN), and learning vector quantization (LVQ) were employed in the inspection models. The inspection results showed that both BPNN and LVQ have excellent prediction results with 100% accuracy. Chen et al. [15] used BPNN in the etch semiconductor process to identify and classify endpoint curves. By real-time monitoring of changes in the endpoint

* Corresponding author. e-mail: sami@hu.edu.jo

curve, the abnormalities of products can be detected immediately. The system can reduce the uncertainty in the process curve classification and provide machine shut-down suggestion immediately when necessary. In this respect, back propagation neural network is utilized to identify the wafer status during reactive ion etching, RIE, which is an important MEMS fabrication process.

2. Reactive Ion Etching

Reactive ion etching (RIE) is a very difficult process to control, since the physical mechanism of this process is not well understood. Consequently wafer defect occurs in RIE when there is a sudden change in the etching behavior. This change can happen due to operator errors or machine errors, such as gas leak, power fault, and pressure fault. The main defect in oxide RIE is un-open etch. This defect costs 10%-20% yield loss in a factory. Un-open etch signify the inadequate etching space in the wafer surface as shown in Figure 1.

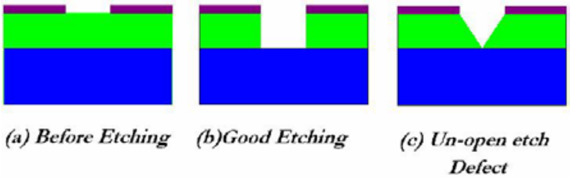

Figure 1: Wafer surface shape before and after etching process.

A silicon dioxide film etching without any errors is a complex task. One of the significant detractive defect in this process is un-opened etch. Predicting wafer status in RIE is important to enhance yield, quality, and efficiency. Towards this end, thorough analysis has been done to get improved prediction models with the purpose of providing valuable benefits. One of the attractive prediction models is artificial neural networks. Data preparation for neural networks has been discussed and applied in different ways.

3. Case Study: Oxide RIE

The advanced reactive ion etching equipment used in the factory is 2300 Exelan Flex made up of two chambers as shown in Fig. 2. The main etching chamber is configured inside the vacuum RIE chamber for optimal efficiency, where these two chambers are separated by quartz confinement rings. The RIE chamber consists of two parallel plates, RF power supply and pumping system.

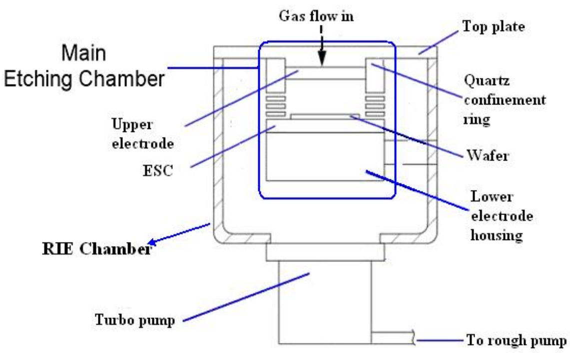

Figure 2: Sketch of the main etching chamber inside the RIE chamber.

Forty different signals were collected from 2300 Exelan Flex equipment for the reactive ion etching process. Twenty two signals are carefully chosen from all the signals. The chosen signals directly impact the wafer status. Table 1 provides more elaborations about RIE factors.

Table 1: RIE factors and their clarification.

#	Factor	Factor explanation
1	Bias Voltage	Direct-current voltage at electrode. The value of the bias voltage depends on the size of the electrode and on the gas pressure.
2	ESC Clamp Voltage	Electrostatic chuck.(a mechanism for holding wafers using electrostatic attraction.)
3	ESC Current1	The Current pass in the first selected point on ESC
4	ESC Current2	The Current pass in the second selected point on ESC
5	ESC Temperature	Electrostatic chuck. Temperature (a mechanism for holding wafers using electrostatic attraction.)
6	Fluorine Pressure	The mixed Gases pressure inside the pipe
7	Forward Power 27 MHZ	Applied 2 MHZ for dissociation
8	Forward Power 2MHz	Applied 2 MHZ for bombard
9	Gas 1	C_4F_8 mass flow rate
10	Gas 10	Xe mass flow rate
11	Gas 11	C_4F_6 mass flow rate
12	Gas 4	Ar mass flow rate
13	Gas 7	O_2 mass flow rate
14	He Flow Inner	Mass flow rate for Helium gas inside the Inner tube of cooling system
15	He Flow Outer	Mass flow rate for Helium gas inside the outer tube of cooling system
16	He Pressure Inner	Helium gas Pressure inside the Inner tube of cooling system
17	He Pressure Outer	Helium gas Pressure inside the outer tube of cooling system
18	Pressure	The pressure value inside the RIE chamber
19	Process Time	The time from the initial to the end process
20	Reflect Power 27 MHz	The reflected 27 MHZ which has not been used at the process
21	Reflect Power 2 MHz	The reflected 2 MHZ which has not been used at the
22	Top Plate Temperature	The temperature on the upper plate

4. Data Preparation

This section explains the details of the data preparation performed in this study. Fig. 3 presents the percentages of the training and test wafers where the total number of wafers is one hundred twenty. The wafers were collected from the 2300 Exelan Flex machine. The ratio of the number of training wafers to the number of test wafers is three. Fourteen wafers (12%) from the ninety training wafers (75%) stand for unopened etch defected wafers, and five wafers (4%) from the thirty test wafers (25%) stand for unopened etch defected wafers.

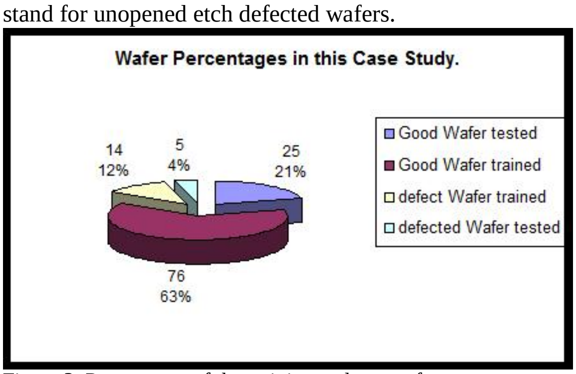

Figure 3: Percentages of the training and test wafers.

Data preparation techniques are used to obtain good prediction results. Three different data preparation techniques are suggested: raw data, sampling data and statistical summary data.

- Raw data preparation: one hundred eighty four data points is the minimum number of data points from the collected data. Thus the first one hundred eighty data points are suggested as raw data inputs for offline prediction models and twenty data points are suggested as raw data inputs for online prediction models.

- Sampling: non-symmetric sampling is the second suggested preparation technique, which has ability to cover all etching steps, at the same time focusing on the main three etching steps (step 4, step 5 and step 10). Table 2 shows the number of captured samples in each step, where two samples are captured from stabilization steps (steps 1, 2, 3, 7, 8, and 9) and two from plasma ramp down steps (steps 6 and 11). More than half of the samples are captured from the main etching steps (steps 4, 5, and 10). As a result thirty-four captured samples cover all the etching steps. These thirty-four captured samples are used as inputs for offline prediction models.

Table 2: Number of suggested sampling for each step in the sampling technique.

Step Number	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	Step 7	Step 8	Step 9	Step 10	Step 11
Number of sample	2	2	2	6	6	2	2	2	2	6	2

Figure 4 illustrates the position of captured samples, where the first data point of each step is captured, and the suggested sampling rate for the main etching steps is five whereas it is equal to three for other steps. These samples include the first data point of each stabilization step.

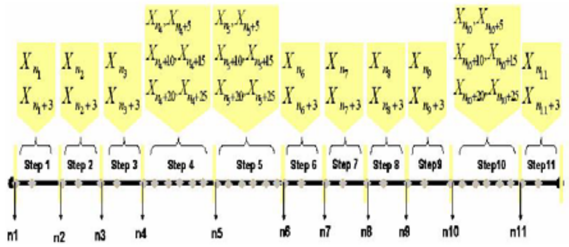

Figure 4: The position of captured samples.

- Statistical summary preparation: the last suggested preparation technique depends on mean and standard deviation values. Since many samples have no data for the sixth step it became out of interest. This means ten steps will be statistically summarized and applied in prediction models.

5. Principal Component Analysis

Principal component analysis (PCA) is an important analysis technique in multivariate statistics. It was first suggested in 1901 by Pearson [16], and formally developed by Hotelling [17]. The main idea of principal component analysis (PCA) is to represent a number of correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component (PC) accounts for the variability in the data as much as possible, the second PC is the linear combination with the second largest variance and orthogonal to the first PC, and so on. There are as many PCs as the number of the original variables. For many datasets, the first several PCs explain most of the variance, so that the rest can be disregarded with minimal loss of information. The objectives of using PCA are to reduce the dimensionality of a data set and to identify new underlying variables that are now orthogonal.

To enhance performance of the prediction model in this study, PCA is suggested to represent the RIE factors, since simple neural networks with few nodes and connections tend to have better generalization capability. In this section, PCA technique automatically extracts three principle components (PCs) from all RIE factors (twenty two factors). It is important to treat each etching step separately in PCA, because each step has different inherent physical/chemical characteristics. Considering the overall process characteristics and the objective of model simplicity, it was decided that utilizing one PCA for each of the eleven steps would yield a better solution than utilizing a single PCA for the entire process. In this paper, principal component analysis was utilized for 90 training wafers. The principle components are found by computing the sample covariance matrix and selecting its eigenvectors (loading vectors) for the largest eigenvalues.

6. Architecture of Prediction Models

As stated before, this study combines back propagation neural network (BPNN) and principle component analysis (PCA) to construct the two prediction models. Prediction models are concerned with all etching process steps to predict the wafer status at the end of the etching process.

The BPNN in this study consists of three layers of neurons: the input layer, hidden layer, and output layer. The input layer receives external information such as RIE processing factors or principle components. From the output layer, predictions are produced with binary values to represent the wafer status. Since the network output is between zero to one, the zone that is smaller than a minimum threshold value is set to zero and the zone that is greater than a maximum threshold value is set to one. If the network output value is equal to one that means the wafer status is good, and otherwise it is defected. When the network output value is between the minimum and maximum values, then the network fails to predict the wafer status.

The BPNN also incorporates hidden layers of neurons, which do not interact with the outside world, but assist in performing nonlinear feature extraction on the data provided by the input and output layers. The number of hidden layers was set to one in this application. Training matters have to be settled with respect to the description of the BPNN network structure.

6.1. Training:

During training, the network is trained to associate outputs with input patterns. This principle is referred to as supervised learning. The training is continued until the training reached the maximum number of epochs or training neural network has MSE (mean square error) less than 10-6. The maximum number of epochs used during training the networks is set to 10000.

After training, the prediction performance of the models is evaluated with two test sets. In the first test set computation of the prediction error for new data points is performed. Data of twenty-five good wafers and five-defected wafers stand for the test data set. Two types of errors are obtained in the first test: type I prediction error occurred when good wafers are predicted as defected wafers, and type Π prediction error occurred when defected wafers are predicted as good wafers.

The second test set depends on the recognition / rejection rate. The recognition rate is the percentage of test

 © 2011 Jordan Journal of Mechanical and Industrial Engineering. All rights reserved - Volume 5, Number 4 (ISSN 1995-6665)

samples recognized correctly and the output value is located outside the range between the minimum and maximum threshold values. The rejection rate is the percentage of input samples that could not be assigned to any particular class; because the output value is located somewhere between the minimum and maximum threshold values. The minimum and maximum values are determined for every prediction model after testing the training wafers, where the minimum value is the highest output value for a defected wafer, and the maximum value indicates the lowest value for a good wafer.

7. Evaluation of Prediction Models

The major aspect in this section is to evaluate the performance of the prediction models, and decide the best prediction model with respect to a constraint. The same training and test samples are used for all prediction models. The experimental data examined were collected from reactive ion etching of silicon dioxide thin film.

7.1. BPNN prediction model:

Figure 5 illustrates the offline BPNN prediction model, where the five significant factors (bias voltage, He inner flow, He outer flow, pressure, and reflect power 2MHZ) are the model inputs. The number of input neurons of the BPNN is different for each of the three data preparation techniques described above. When the raw data preparation technique is applied in the offline BPNN the number of input neurons is nine hundred, whereas it is one hundred seventy when captured sampling is applied, and one hundred for the statistical summary preparation technique.

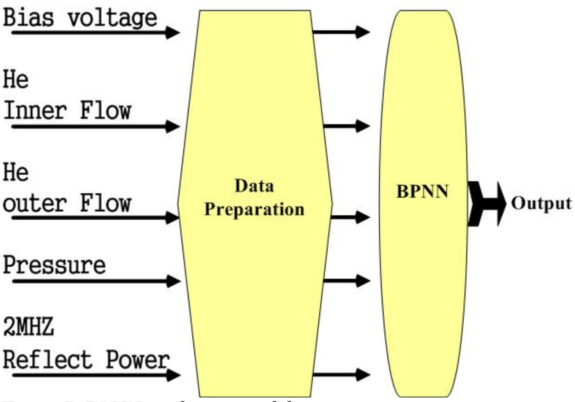

Figure 5: BPNN prediction model.

The raw data preparation technique does not prepare data as well as other data preparation techniques, even if it has good performance for predicting the wafer status by using two factors (He outer flow and bias voltage). In addition, the raw data preparation technique has no ability to cover all etching steps. Table 3 illustrates the offline prediction model performance by using sampling and statistical summary preparation techniques. Both data preparation techniques assist the offline BPNN prediction model to achieve zero error and 100% recognition rate. Recognition rate represents the percentage of test samples recognized correctly and the corresponding output values are located outside the range between the minimum and maximum threshold values, namely, 0.1 and 0.9, respectively.

Table 3: The performance of offline BPNN prediction model.

	BPNN Characteristic	
	Sampling	**Statistical summary**
Recognition rate	100%	100%
Training MSE	1.94311E-07	1.47594E-06
Testing MSE	5.12016E-09	7.46886E-07
# of input neurons	170	100
Error prediction	0%	0%

7.2. PCABPNN prediction model:

To construct offline PCABPNN prediction model the principle component analysis (PCA) and back propagation neural network (BPNN) are combined together. First, the PCA is adopted to extract valuable information from the twenty two RIE factors for each step. Then the extracted principle components for all the steps are combined together and prepared by the three different data preparation techniques described earlier in this paper. The neural network is trained afterwards by the prepared PC data of 90 training wafers and tested for prediction accuracy by the prepared PC data of 30 test wafers.

Figure 6 illustrates the offline PCABPNN prediction model. Principle component analysis is applied in this model to extract the input parameters for the neural network. Moreover the required time to find the principle components is much less than the required time to find the significant parameters.

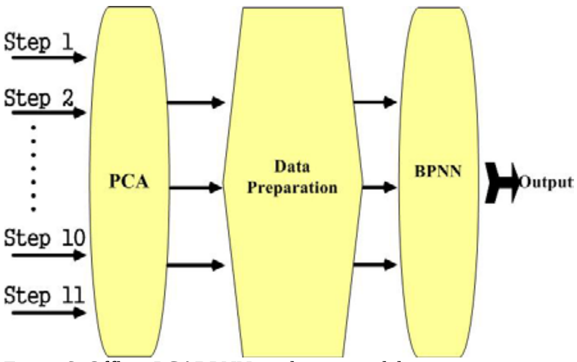

Figure 6: Offline PCABPNN prediction model.

One hundred eighty data points for each principle component parameter are prepared by the raw data preparation technique. Totally, there are five hundred and forty neuron inputs for the three PCs. Since the raw data preparation technique covers the first one hundred eighty data points of the PCs, the rest of data points are ignored. The ignored points of PCs may contain important information. Three defected wafers are incorrectly predicted as good wafers by using the raw data preparation technique.

In order to reduce the input data size one hundred and two captured samples from the three principle components are trained and tested using the BPNN (see Table 4). Four samples out of thirty tested samples were clamped between 0.1 and 0.9 (the minimum and maximum threshold values) by applying the captured sampling in the offline PCABPNN prediction model. Applying the statistical summary data preparation technique yields no samples in the failure zone, due to high accuracy (100% recognition rate) of the prediction model. In general, the statistical summary data preparation technique has the best ability to enhance the accuracy of the offline PCABPNN prediction model.

Table 4: The performance of the offline PCABPNN prediction model.

	PCABPNN Characteristics		
	Raw PC	Sampling	Statistical summary
Recognition rate	96.67%	86.67%	100%
Training MSE	1.29021E-06	2.4213E-05	4.52214E-06
Testing MSE	0.101489096	0.041548198	4.05693E-05
# of input neurons	540	102	60
Error prediction	10.00%	3.33%	0%

8. Conclusion

Two prediction models for etching of silicon dioxide thin film in the MEMS fabrication process reactive ion etching were developed to predict the wafer status correctly by using statistical summary preparation techniques. Back propagation neural network (BPNN) is the first prediction model and is the backbone for the second prediction model, namely, the principle component analysis BPNN (PCABPNN). The two models achieve the objective of achieving fast and robust predictions. This paper describes the potential of these prediction models to reduce the overall cost of MEMS equipment, to achieve high yields and throughput during MEMS fabrication. The two prediction models do not depend upon monitoring wafers or expensive metrology rather they will enable inexpensive real-time wafer-to-wafer inspection application. The results from the evaluation of the prediction models indicate that robust, accurate and stable predictors have been constructed. Furthermore, a greater accurate performance of prediction models has been achieved by online BPNN prediction model.

References

[1] N. Tudoroiua , R.V. Patelb, K. Khorasania, “Neural Network Based Control Strategies for Improving Plasma Characteristics in Reactive Ion Etching”. Neurocomputing, Vol. 69, 2006, 786-802.

[2] B. Kim, M. Kwona, “Prediction of Plasma Etch Process by Using Actinometry-based Optical Emission Spectroscopy Data and Neural Network”. Journal of Materials Processing Technology, Vol. 209, 2009, 2620-2626.

[3] P. H. Choua, M. J. Wua, K. K. Chenb, “Integrating Support Vector Machine and Genetic Algorithm to Implement Dynamic Wafer Quality Prediction System”. Expert Systems with Applications, Vol. 37, 2010, 4413-4424.

[4] S. R. Bhatikar, R. L. Mahajan, “Artificial Neural-Network-Based Diagnosis of CVD Barrel Reactor”. IEEE Transactions on Semiconductor Manufacturing, Vol. 15, 2002, 71-78.

[5] F. L. Chen, S. F. Liu, “A Neural-Network Approach to Recognize Defect Spatial Pattern in Semiconductor Fabrication”. IEEE Transactions on Semiconductor Manufacturing, Vol. 13, 2000, 366-373.

[6] J. H. Lee, S. J. You, S. C. Park, “A New Intelligent SOFM-based Sampling Plan for Advanced Process Control”. Expert Systems with Applications, Vol. 20, 2001, 133-151.

[7] F. L. Chen, S. F. Liu, K. Y. Y. Doong, K. L. Young, “LOGIC Product Yield Analysis by Wafer Bin Map Pattern Recognition Supervised Neural Network”. IEEE International Symposium on Semiconductor Manufacturing, Vol. 1, 2003, 501-504.

[8] F. Di Palma, G. D. Nicolao, G. Miraglia, E. Pasquinetti, F. Piccinini, “Unsupervised Spatial Pattern Classification of Electrical-Wafer-Sorting Maps in Semiconductor Manufacturing”. Pattern Recognition Letters, Vol. 26, 2005, 1857-1865.

[9] S.S. Han, M. Ceiler, S. A. Bidstrup, P. Kohl, G. May, “Modeling the Properties of PECVD Silicon Dioxide Films Using Optimized Back Propagation Neural Networks”. IEEE Transactions on Components, Packing, and Manufacturing Technology, Vol. 17, 1994, 174-182.

[10] S. S. Han, G. S. May, “Using Neural Network Process Models to Perform PECVD Silicon Dioxide Recipe via Genetic Algorithms”. IEEE Transactions on semiconductor manufacturing, Vol. 10, 1997, 279-287.

[11] B. Kim, D. W. Kim, G. T. Park, “Prediction of Plasma Etching Using a Polynomial Neural Network”. IEEE Transactions on Plasma Science, Vol. 31, 2003, 1330-1336.

[12] B. Kim, W. S. Hong, “Use of Neural Network to Characterize a Low Pressure Temperature Effect on Refractive Property of Silicon Nitride Film Deposited by PECVD”. IEEE Transactions on Plasma Science, Vol. 32, 2004, 84-89.

[13] R. A. Zoroofi, H. Taketani, S. Tamura, Y. Sato, K. Sekiya, “Automated Inspection of IC Wafer Contamination”. Pattern Recognition, Vol. 34, 2001, 1307-1317.

[14] C. T. Su, T. Yang, C. M. Ke, “A Neural-Network Approach for Semiconductor Wafer Post-Sawing Inspection”. IEEE Transactions on Semiconductor Manufacturing, Vol. 15, 2002, 260-266.

[15] W. C. Chen, C. T. Chen, T. H. Ho, J. H Chen, L. J. Sheu, “Use of Neural Network in Pattern Recognition of Semiconductor Etching Process”. The Proceedings of the 11th International Conference on Industrial Engineering and Engineering Management, Vol. 1, 2005, 719-725.

[16] K. Pearson, “On Lines and Planes of Closest Fit to Systems of Points in Space”. Philosophical Magazine, Vol. 2, 1901, 559–572.

[17] H. Hotelling, “Analysis of a Complex of Statistical Variables into Principal Components”. Journal of Educational Psychology, Vol. 24, 1933, 417–441.