

Military Technical College
Kobry El-Kobbah,
Cairo, Egypt.



15th International Conference
on Applied Mechanics and
Mechanical Engineering.

SCORING MODELS FOR PROCESS MONITORING IN INDUSTRY A MECHATRONIC APPLICATION

R. M. Shahin*, W. M. Hussein*, A. F. El-Sherif* and H. M. Mahgoub*

ABSTRACT

To compete on a global level, manufacturers must strive to find new solutions to ensure high product quality while maximizing productivity. One way to address many of the challenges this creates is to turn increasingly to mechatronic design approach and automation. To successfully realize automation, a mechatronic system for on-line process monitoring is required to take the place of an expert's judgment. This paper outlines the use of a statistical multivariate technique called Projection to Latent Structure (PLS), and applies it to the monitoring of a machining process as an application. This approach is used to integrate machine tool sensory data from a milling machine. Experiments were conducted on a milling machine under three conditions, sharp, worn and chatter tools. The score models were tested under different conditions with the results showing that the proposed technique can be used for tool wear monitoring and can successfully differentiate between new process conditions.

KEY WORDS

Process monitoring, multivariate analysis, Sensors, surface roughness, Projection to Latent Structure, Milling, Mechatronic system.

* Egyptian Armed Forces.

INTRODUCTION

Many manufacturing companies are equipped with unmanned or partially manned production lines made up of multiple machines working together in cells. Thus industrial machines need to be assessed to get more insight about their performance with minimal human supervision. To emulate a human's monitoring action it is highly desirable to implement a mechatronic system for on-line process monitoring which will allow the supervision of the performance of industrial processes. Also it is valuable to provide immediate contribution information identifying common failure mechanisms that can be assigned to the machine tool responsible for the process shift.

Most monitoring techniques presently available focus on one or a limited number of failure mechanisms like tool wear, breakage or chatter using a heuristic approach which gives limited information about product quality [1]. In other words, the used signals depend on a special problem as well as on the type of machine. A review of different technologies in process fault detection and monitoring applications can be found in [2-10].

They identify three characteristics that are commonly considered for future industrial researches and applications of tool monitoring systems. They are the use of multiple sensors [11], the integration of control signals from the machine controller into the monitoring system [12] and finally the use of wireless and telecommunication based systems to facilitate dissemination of information [13-14]. Among these techniques, artificial neural network ANNs and its combination with other methods have been the most predominant to date [15-18]. ANNs techniques are characterized as black box approaches which model the relation between different variables to a desired response without giving any information about what happens inside the process. Also some of these approaches are susceptible to missing data due to sensor reliability issues, and some applications use many charts to monitor individual process. In addition, issues associated with collinearity and dimensionality needs to still be specifically addressed in these techniques. Looking for a generic approach to solve many of these limitations can be addressed through the use of multivariate statistical analysis.

Multivariate statistical process control (MSPC) [20], takes a different approach as compared to the other methods mentioned previously. The most fundamental difference with this approach is that the model is based on non-causal empirical correlations extracted from normal plant operating data when only common cause variation exists [21]. The simplicity of this approach is that there is no need for a fundamental model of the system and only data from normal production needs to be used, which is generally available in some form at most factories. Two of the approaches used in multivariate analysis are: a projection method called principal component analysis (PCA) [22-23] and projection to latent structure (PLS) [24]. Many applications of these two techniques have been successfully applied in other fields of process monitoring [20, 27, 29] ranging from batch to continuous processes.

This paper outlines the use of the multivariate score models to develop a mechatronic system to monitor an industrial process. By integrating different sensory data such as force, spindle and table vibration signals from a milling machine under

different conditions, models based on PLS and its extension PLS-DA [19] can be developed into a single user interface. From this interface surface roughness and tool wear, for example, can be monitored using the score space of these models.

MULTIVARIATE PLS BASIC THEORY

PLS is defined as a generalized multiple regression method based on relating two blocks of data made up of input and output variables. This method is particularly effective at modelling and analyzing variables that are correlated; which is the case for many process data.

In our case, we test the collinearity within variables by checking the loading plot, a scatter plot between the model weights in the first 2 principal components as shown in Fig.1. It shows that a positive correlation exists between feed/tooth, rpm, vibration, current and the forces in (y, z)-directions in the direction of the first principal components (x-axis). Also, it shows a negative correlation between the mentioned factors with the acoustic emission and the feed force. An overview about the method and its applications can be found in [31]. This commonly occurring collinearity between variables means that the data will have some dominant type of variability that carries most of the information. To visualize these hidden dominant directions, variables are projected onto a new subspace. This implies the reduction of a large data set from many variables to a few factor scores (t,u) and loadings (w,p,q). The point behind using 2 loadings (w,p) to describe x-variables, is that in PLS algorithm, the y's is allowed to intervene directly during the input variables decomposition which decreases the variance explanation of the input variables towards the maximum predictability of the response variable. In this case we use two loadings to describe the input variables, the w's to calculate principal components and the p's when calculating the residuals for each component because it better describes input variables. Scores can be described as new generalized variables while loadings build the connection between the original space and the new subspace. The approach works by selecting factors of input variables in a sequence which successively maximizes the explained covariance between the input and output variables. Given a matrix of input data X, and output data, Y, a factor of the input data, t₁, and output data, u₁, is evaluated such that:

$$X = \sum_{k=1} t_k \bullet p_k^T + E \quad \text{and} \quad Y = \sum_{k=1} u_k \bullet q_k^T + F$$

These equations are referred to as the outer relationships where the "t" vectors are mutually orthogonal. These vectors together with the u_k terms are selected so as to maximize the covariance between each pair, (t_k, u_k). Linear regression is performed between the t_k and the u_k terms to produce the inner relationship, such that $u_k = b_k \bullet t_k + \varepsilon_k$, where b_k is a regression coefficient, and ε_k refers to the prediction error. The PLS method provides the potential for a regularized model through selecting an appropriate number of latent variables, u_k in the model. The major benefit of this approach is that it builds a model between the inputs as well as the outputs and finds the maximum correlation between them in the direction of maximum prediction of the output response. The benefit of this approach is twofold. First, it has the advantage of examining any change that could happen in the new

input variables, that is the sensors, before using the model for prediction. It does this by monitoring the input variables using a multivariate chart called the DModX (distance to the X-model). Second, it helps to isolate any shifts that could happen by providing variables that are responsible for these shifts through details available in contribution plots. Furthermore, it is often found that a relatively small number of low-index latent variables can explain the greater part of the variation in both the input and output variables. The algorithm used to calculate the PLS is superficially "similar" to the NIPALS (Nonlinear Iterative Partial Least Squares) algorithm [22]. Sometimes factors are introduced that describe the data from different groups (machines). So, we must assign some distinct levels to these factors to help discriminate between different machines. We call these factors qualitative or dummy variables. PLS-DA uses dummy variables [32] in the output variable list to separate between several groups of data. Mathematically this can be illustrated by adding new variables "G" to the model with some regression coefficients say "c", and set G = 1 for data from one machine and G = 0 for data from another machine in the following way: $\hat{Y} = \hat{f} + c \cdot G$. Suppose that \hat{f} represents the fitted model, so when G = 0, the data from the first machine is highlighted: $\hat{Y} = \hat{f}$ and by putting G = 1, the data from the second machine is highlighted: $\hat{Y} = \hat{f} + c$ where c estimates the difference in levels between the responses of two groups of machine data. So, data from different machines will be clustered depending on their regression coefficient values.

EXPERIMENTAL SET-UP

The experimental validation was performed on a CNC milling machine (FADAL vmc15). It involved the side end-milling process for X axis feed of 6065-Aluminum parts (65×23×3 mm) using a 6 mm, 2-flute high speed steel (HSS) tool.

Two tools were used, one was as supplied (sharp) and the other was worn, prepared manually with flank wear of VB=0.2 mm as measured using a toolmaker's microscope. The system used to collect the data has a total of three sensory signals.

Table 1. Sample data.

#	F/t	DOC	rpm	Ra (µm)
Sharp tool	0.05	0.45	5000	0.58
Sharp tool	0.075	0.65	7000	0.57
Sharp tool	0.05	0.65	7000	0.43
Worn tool (one tooth)	0.05	0.45	5000	1.84
Worn tool (one tooth)	0.075	0.65	6500	1.06
Worn tool (one tooth)	0.05	0.65	5000	1.44
Worn tool (two teeth)	0.05	0.45	4500	2.42
Worn tool (two teeth)	0.075	0.65	7500	2.84
Worn tool (two teeth)	0.05	0.65	6500	2.05

One accelerometer (B&K) is used to monitor the vibration of the spindle it is connected to an amplifier (B&K). Other accelerometer (B&K) is used to monitor the vibration of the table it is connected to an amplifier (B&K), The force signals are monitored using a 3-component dynamometer (Kistler 2825A). These signals are sampled at 20 kS/s on a PC with a National Instruments data acquisition card (PCI-6033E) running LabView software. The experimental set-up is shown in Fig. 2.

The test plan involved dry machining with different combinations of depth of cut (0.45-0.65mm) and feed/tooth (0.05-0.075mm) for a range of spindle speeds based on the capability of the machine. Test runs started from 3,000 RPM and ending at 7,500 RPM for the machine (40 sharp and 60 worn). The number of experiments is arbitrary chosen to cover the combination of feed and depth of cut for different rpm and to test the capability of the approach to deal with the machine with all observations. Surface roughness Ra values were measured as the relevant quality indicator of the machined parts for this study. It was measured in an area of (0.7×8 mm) along the feed direction using a Mitutoyo sj-201 stylus. A sample of Ra data is shown in Table 1.

MODEL BUILDING AND CONFIGURATION

The multivariate PLS model is a linear model given in matrix form by: $Y = X B + E$. However, there exists non-linear PLS versions even by augmenting the original matrix with the non-linear factors or by building a non-linear relation between score factors [33]. To build the model using process and machine variables, let X include variables containing the information in the sensory data (forces, base vibration and spindle vibration), and the cutting data (feed/tooth, cutting speed and depth of cut). These sensors were believed to have a direct correlation with the machined surface roughness and in the same time cover many sources of variations that could happen during the machining process. Each signal from used sensors was averaged every 60 msec by taking the mean of the readings to form one observation for each set of cutting conditions. Then let Y consist of the machining response which is the surface roughness Ra values of the machined parts in this case. Finally, the PLS regression is made between X and Y based on an algorithm which is superficially "similar" to NIPALS. All the models were built using SIMCA-P code developed by Umetrics. To examine the relationship between the inputs and the output in the latent space figure 3 shows a scatter plot between the t-u scores in the direction of the first principal component. It illustrates a dispersed linear trend between them which imply the possibility to use the approach for further analysis.

There are different approaches to deal with the available data depending on their type. First, we can use only normal data based on common cause variation and build the model, then use the model to detect any shift or fault that will violate the correlation structure established between the variables during the process. Then by examining the contribution plots we can isolate the variables known as X-predictors that are most responsible for this shift. This approach is applied using PCA when one type of data only (X-variables) is available, or in our case, using PLS when two blocks of data, that is both predictors and responses, are available. The second approach is based on using data from both good and bad parts. In this case PCA can be applied for data clustering and prediction on new incoming data, but in our

case, we use PLS-DA, which is generally known to provide better data separation than PCA, by assigning different groups of data to the response dummy variables. Then the model is used to discriminate between them and to predict the nature of new incoming data.

The first model was built using PLS-DA on three groups of existing data (sharp, worn and chatter conditions) from the machine with three principal components and the following percentage of explanation: X-variables model explanation R2X (cum): 0.64, Y-variable model explanation R2Y (cum): 0.80, Y-variable model predictability Q2 (cum): 0.65. PLS-DA uses three “y” variables in this case as dummy variables to separate between sharp (y1), worn (y2) and chatter (y3) data at the training stage. Table 2 illustrates the configuration of the X-matrix in the 2 models used in this paper.

Table 2. X-Y matrix.

X-matrix			PLS-DA							Y-matrix		
fx	fy	fz	BVib	SVib	F/t	DOC	rpm		Y1	Y2	Y3	
									sharp	worn	chatter	
									1	0	0	
									0	1	0	
									0	0	1	
* Sharp, wear and chatter data were used in this model and then learned with different conditions. Sharp data were normal data that gave accepted surface roughness.												
PLS												
fx	fy	fz	BVib	SVib	F/t	DOC	rpm		Y=Ra			
* Only sharp data were used that gave accepted Ra, then latent control charts (Score and DModX) plots were used to detect abnormalities during the process like wear.												

RESULTS AND ANALYSIS

Before starting the discussion we need to know some aspects of the approach. First, all data were mean centered and scaled to unit variance. Second, control limits in the latent space were established using F-distribution based on reference distribution provided by the dataset [34]. Also, t_1 and t_2 are the first two principal components that capture most of the variance-covariance between the X-Y relationships. Figure 4 illustrates how the score plot is built for a simple case 3-variables and 2-scores. After determining the direction of maximum variation by iterative steps and get the second orthogonal direction by the same way after subtracting the first component, we rotate the new plane determined by the new score variables t_1 and t_2 and then monitor the

movement of the process variables in the reduced dimensional space during machining.

Figure 5 shows the PLS-DA latent space score plot of the data from the milling machine (sharp, worn and chatter conditions). It shows the capability of the approach to simultaneously monitor industrial process subject to different faulty mechanisms tool wear and chatter. It also shows the directions of the wear and chatter occurrence machine where the different clusters can be separated by straight lines. The normal data group came from conditions that produced an acceptable surface roughness. Figure 6 shows the results of testing the model under different conditions (normal, excess feed, chatter and wear with excessive feed). It is clear that the model successfully separates between different conditions based on their violation of the existing correlation pattern between the variables.

As a result, the model can be used to monitor different failure mechanisms. More investigation can be carried out using contribution plots to quickly assess which factors affect these different shifts and assigning them to the machine responsible for this quality issue. Fig.7 shows a contribution plot (They are the loading weights multiplied by the change in X-variables between two observations) for an observation with chatter; we can notice that vibration provided the largest contributing factor in this case. Also, we can see that this fault occurred in machine one (LX1) from the positive value of dummy variable X1.

A second model was built based on PLS to relate surface roughness Ra with sensor outputs and the main cutting parameters (feed/tooth, rpm, D.O.C.) using two principal components with the following percentage explanation: R2X (cum): 0.55, R2Y (cum): 0.59, Q2 (cum): 0.488. The model was built using only normal data from the machine that gave acceptable quality as indicated by the measured Ra value. The model was tested with wear and chatter conditions and monitored in the latent space by Hotelling's T2 and DModX Shewart charts as shown in Fig.8. Hotelling's T2 combines all significant scores together in one chart and detects major changes that could happen inside the model. DModX considers model residuals which can identify variable correlation changes or the occurrence of a new event that is not modeled. We can clearly notice that T2 detected the wear and chatter conditions while DModX detected also the tool wear and chatter events.

Furthermore, in addition to monitoring the latent space, we can use the original space to monitor the real values of the surface roughness Ra in a time series sequence. Data from the milling machine using sharp tool (common cause variation) under different cutting conditions are shown in Fig.9. The model gives a good approximation of actual Ra values with model prediction ability Q2 = 59% and with accuracy 83 %.

CONCLUSION AND FUTURE WORK

A new mechatronic system to monitor industrial machine tool based on multivariate scoring models was developed. The approach uses experimental data for model building and testing. The superiority of this method over other existing approaches is the modeling of the X-space which facilitates the detection of process change due to

wear and other conditions where in the same time monitoring Ra with an acceptable accuracy. Results can be summarized as follows:

- A new scoring charts based on PLS-DA is introduced to deal with multiple faults on machine tools. The model was tested under different conditions including: normal, worn tool, as well as chatter and was able to successfully differentiate between them.
- A new approach based on PLS is illustrated for surface roughness prediction using data from milling machine tools.
- System dimensions were reduced from 8 to 3 which can be handled by two charts (score and contribution plots). This greatly simplifies the monitoring of machines and can help for process planning and machine rescheduling.

This approach doesn't build a causality direction between variables rather than a correlation direction inside a bounded region except if the model was based on designed experiments. Further analysis is needed to include tool chipping and some machine faults (spindle bearing, parts wear). Also, there is a need for more investigation to include more features like factor variances, frequency components and nonlinear terms into the X-matrix for better fault separation and to improve model predictability. Finally, including different machining processes, testing new material and tools is highly recommended.

REFERENCES

- [1] Ahmed A.D. Sarhan; and R. M. El-Zahry; "Monitoring of tool wear and surface roughness in endmilling for intelligent machining", 2011, International Journal of the Physical Sciences Vol. 6(10), pp. 2380-2392.
- [2] Prickett, P.W.; Johns, C.; "An overview of approaches to end milling tool monitoring"; 1999, International journal of machine tools & manufacture 39:105-122.
- [3] Tseng, P.C.; Chou, A.; "The intelligent on-line monitoring of end milling"; 2002, International journal of machine tools & manufacture 42: 89-97.
- [4] Prickett, P.W.; and Johns, C.; "The development of an on-line milling cutter tooth breakage detection system"; 2001, Proc. Instn Mech. Engrs, part B, Journal of Engineering Manufacture, Vol 215.
- [5] LandersSadettin Orhan, Ali Osman Er, Necip Camus-cu, Ersan Aslan; "Tool wear evaluation by vibration analysis during end milling of AISI D3 cold work tool steel with 35 HRC hardness"; 2007, NDT&E International 40 (2007) 121–126.
- [6] BISU Claudiu1,a, GERARD Alain, ZAPCIU Miron and CAHUC Olivier; " The milling process monitoring using 3D envelope method"; 2011, ICASAAM Department of Machines and Systems of Production.
- [7] Tunde Isaac Ogedengbe, Robert Heinemann and Srichand Hinduja; "Feasibility of Tool Condition Monitoring on Micro-Milling Using Current Signals"; 2011, AU J.T. 14(3)
- [8] M.Y.M. Yunus, J. Zhang; " A Multidimensional Scaling based Process Monitoring Technique"; 2010, 20th European Symposium on Computer Aided Process Engineering – ESCAPE20.
- [9] X. Li, H. Zeng, J. H. Zhou, S. Huang, T. B. Thoe, K. C. Shaw, and B. S. Lim; "Multi-modal sensing and correlation modelling for condition-based monitoring in milling machine"; 2011SIMTech technical reports Volume 8 Number 1.

- [10] Eyup Bagci; "Monitoring and analysis of MRR-based federate optimization approach and effects of cutting conditions using acoustic sound pressure level in free-form surface milling"; 2011, Scientific Research and Essays Vol. 6(2), pp. 256-277.
- [11] Liu, C.; and Wang, B.; "A case study on multisensor data fusion for imbalance diagnosis of rotating machinery"; 2001, Artificial intelligence for engineering design, Analysis and manufacturing, 15: 203-210, U.S.A.
- [12] Bianca Maria Colosimo, Giovanni Moroni, and Marco Grasso; "Real-Time Tool Condition Monitoring in Milling by Means of Control Charts for Auto-Correlated Data"; 2010, Journal of Machine Engineering, Vol. 10, No. 2.
- [13] Lee, J.; and Wang, B.; "Computer-Aided maintenance: Methodologies and Practices"; 1999, Kluwe Academic, The Netherlands.
- [14] Paul Wright, David Dornfeld, and Nathan Ota; "Condition Monitoring in End-Milling Using Wireless Sensor Networks (WSNs)"; 2008, Transactions of NAMRI/SME Volume 36.
- [15] Sick, B.; "On-line and indirect tool wear monitoring in turning with artificial neural networks: A review of more than a decade of research"; 2002, Mechanical system and signal processing 16(4):487-546.
- [16] Mota-Valtierra G.C., Herrera-Ruiz G., Franco-Gasca L.A., and Macias-Bobadilla G.; "ANN Based Tool Condition Monitoring System for CNC Milling Machines"; 2011, Ingeniería Investigación y Tecnología. Vol. XII, Núm. 4.
- [17] Li, A.U.; Elbestawi, M.A.; "Fuzzy clustering for automated tool condition monitoring in machining"; J. Mech. Systems and Signal Processing 10 (5), (1996), pp. 533-550.
- [18] Dražen Bajić, Sonja Jozić, Luka Celent; "Modeling The Surface Roughness During Milling in Off- Line Monitoring"; 2011, 15th International Research/Expert Conference.
- [19] Eriksson, L.; Johansson, E.; Wold, N.K.; and Wold, S.; "Multi- and Megavariate data analysis, principles and applications"; 2001, Umetrics AB, Sweden.
- [20] Kourti, T.; and MacGregor, J. F.; "Multivariate SPC methods for process and product monitoring"; 1996, Journal of quality technology, Vol.28, 4, October.
- [21] Yoon, S.; and MacGregor, J.F.; "Fault diagnosis with multivariate statistical models part 1: using steady state fault signatures"; 2001, Journal of process control 11 : 387- 400.
- [22] Wold, S.; Esbensen, K.; and Geladi, P.; "Principal component analysis"; 1987, Chemometrics and intelligent laboratory systems, 2: 37-52.
- [23] Jackson, J.E.; "A User's Guide to Principal Components"; 1991, Wiley , New York.
- [24] Wold, S.; Sjostrom, M.; and Eriksson, L.; "PLS-regression: a basic tool of chemometrics"; 2001, Chemometric and intelligent laboratory systems, 58:109-130.
- [25] Lennox, B.; Hiden, H.G.; Montague, G.A.; Kornfeld, G.; and Goulding, P.R.; "Application of multivariate statistical process control to batch operation"; 2000, Computers and chemical engineering, v24, n2, July, PP. 291-296.
- [26] Goulding, P.R.; Lennox, B.; Sandoz, D.J.; Smith, K.J.; and Marjanovic; "Fault detection in continuous processes using multivariate statistical methods"; 2000, International journal of systems science, 31(11), pp. 1459-1471.

- [27] Yon, S.; Landry, J.; Wold, N.K.; Pepe; and Wold, S.; “Multivariate process monitoring and early fault detection (MSPC) using PCA and PLS”; 2003, Plant automation and decision support conference, September 21-24, Hyatt Regency, San Antonio, Texas.
- [28] Chen, J.; and Liu, K.C.; “On-line batch process monitoring using dynamic PCA and dynamic PLS models”; 2002, Chemical engineering science 57:63-75.
- [29] Kourti, T.; “Abnormal situation detection, three-way data and projection methods robust data archiving and modelling for industrial applications”; 2003, Annual reviews in control 27 : 131-139.
- [30] Tsung, F.; “Improving Automatic-Controlled process quality using adaptive principal component monitoring”; 1999, Quality Reliability Engineering Int. 15: 135-142.
- [31] MacGregor, J.F.; Yu,H.; Munoz, S.; Cerrillo, J.; ” Data-based latent variable methods for process analysis, monitoring and control”; 2005, Computers and Chemical Engineering 29, 1217-1223.
- [32] Draper, N. R.; and Smith, H.; “Applied Regression Analysis”; 1998, Wiley Series in Probability and Statistics.
- [33] Azmi, A. I.†, Lin, R.J.T. and Bhattacharyya, D.; “ Modelling of Tool Wear in End Milling of GFRP Composites Using Multiple Regression Analysis and Radial Basis Function Neural”; 2010,The 11th Asia Pacific Industrial Engineering and Management Systems Conference.
- [34] Wold, S.; Kettaneh-Wold,N.; Skagerberg, B.; ” Nonlinear PLS modelling ”; 1989, Chemometrics and Intelligent Laboratory Systems, 7, 53.
- [35] MacGregor, J.F.;" Using on-line process data to improve quality"; 1995, ASQC Statistics Division Newsletter, Vol. 16, No.2

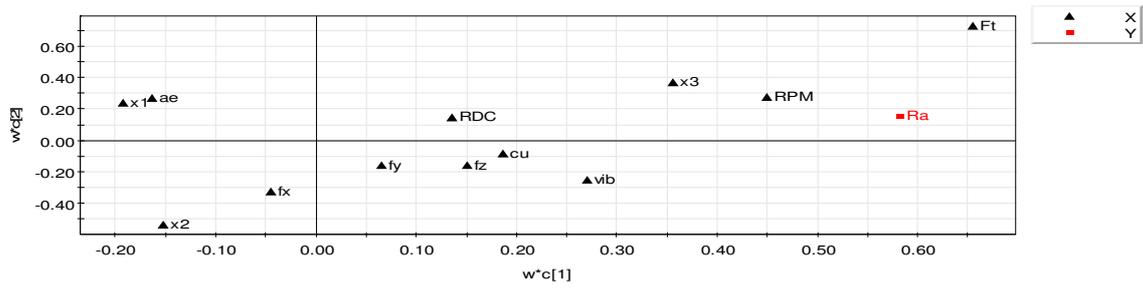


Fig.1. PLS-model loading plot.

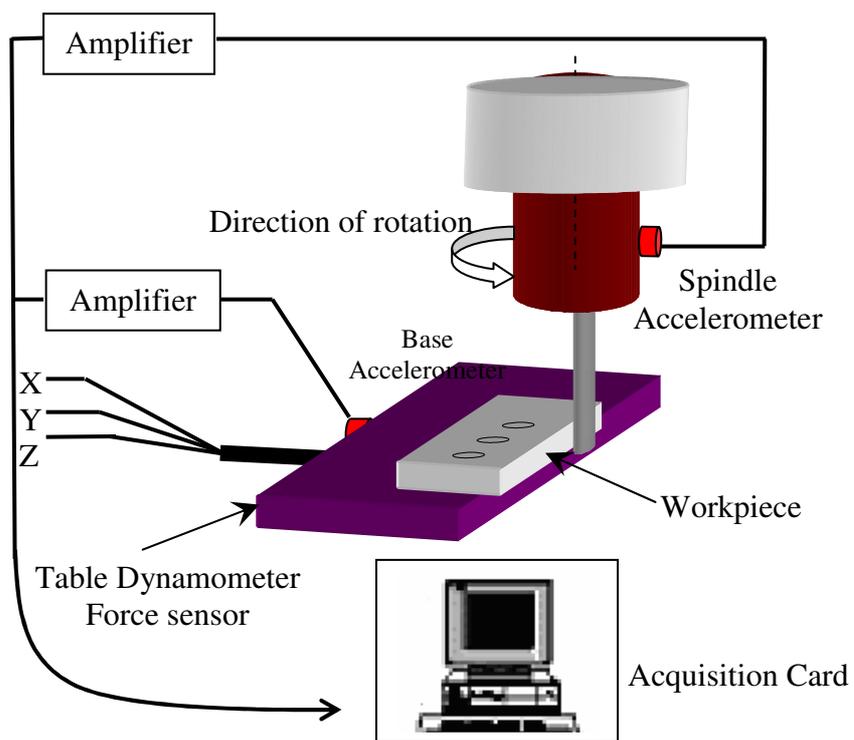


Fig. 2. Experimental setup.

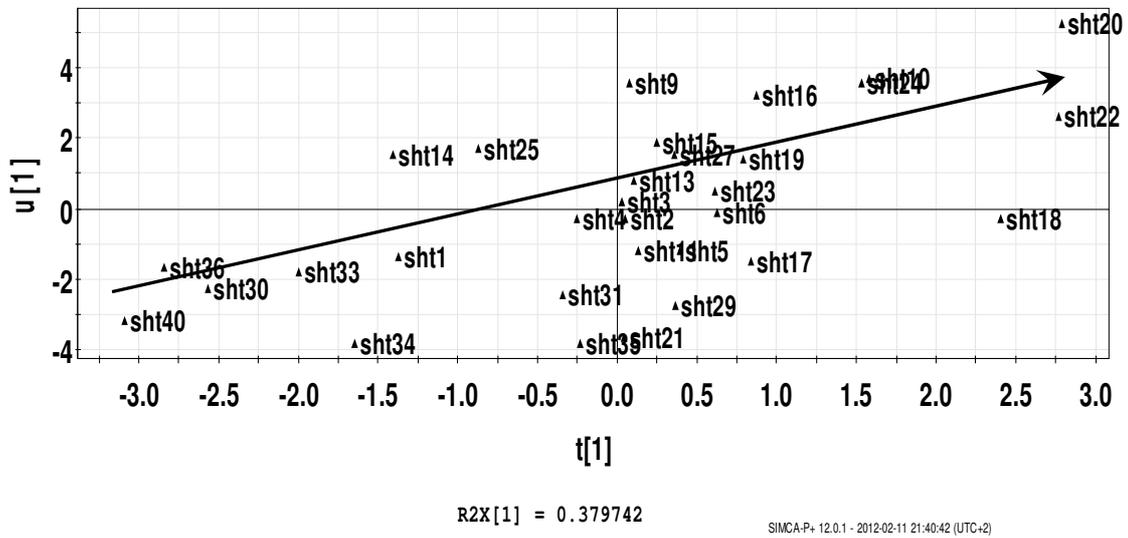


Figure 3. PLS t-u Score plot.

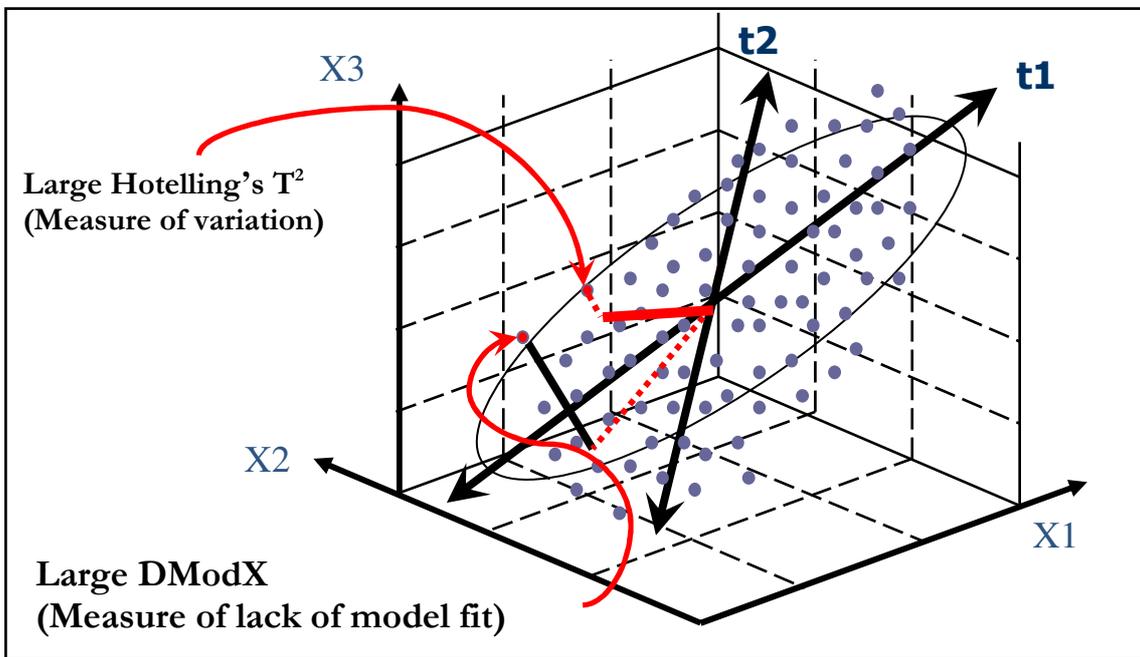


Figure 4. Establishing score plot for three variables.

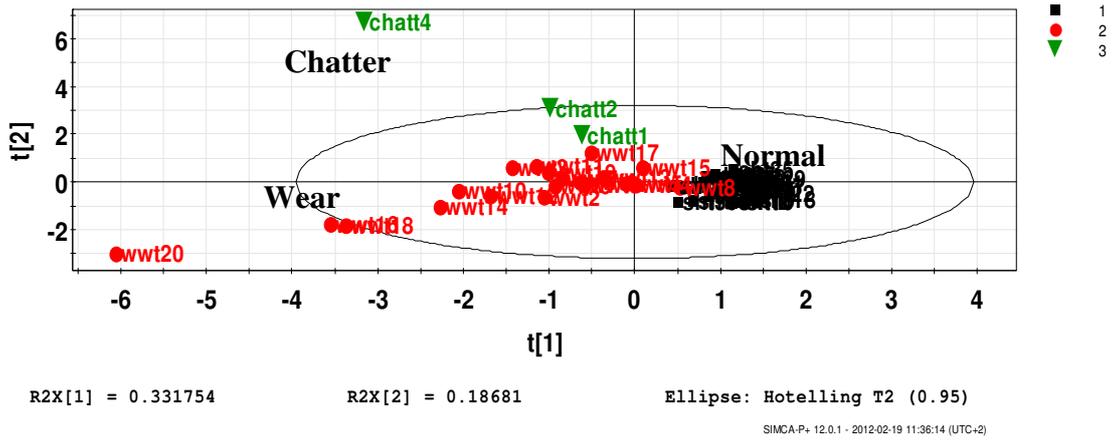


Figure 5. PLS-DA score plot.

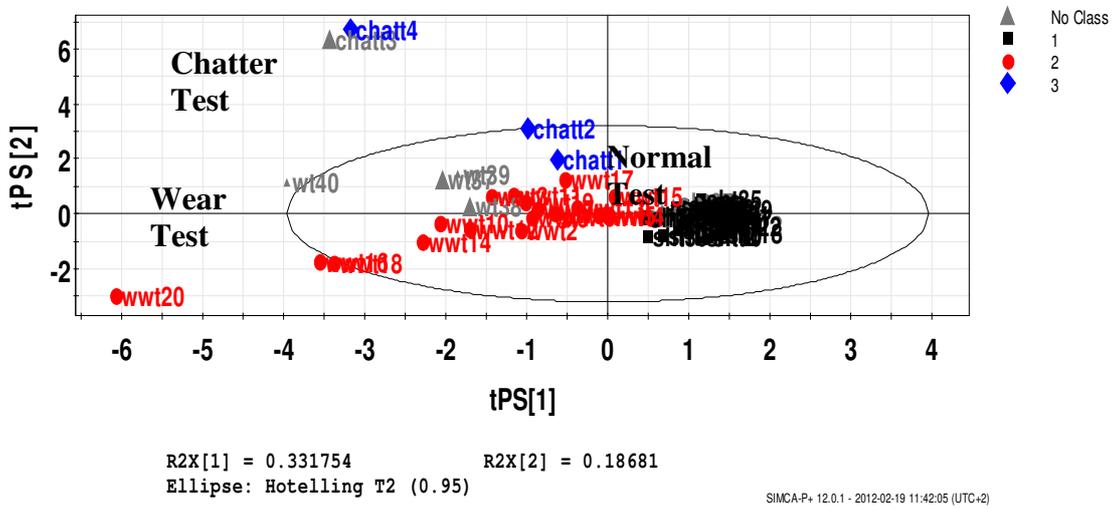


Figure 6. PLS-DA including training and testing-sets score plot.

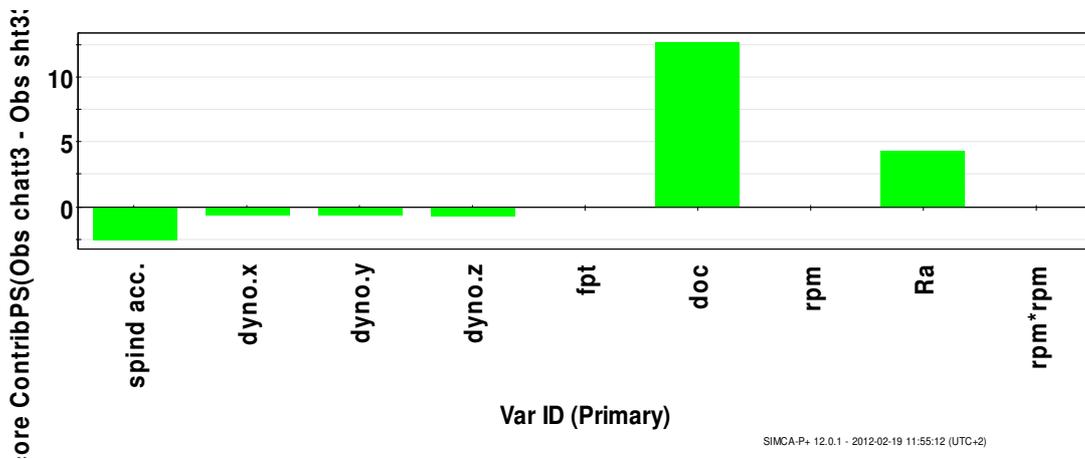


Figure 7. PLS-DA contribution plot (Observation with Chatter).

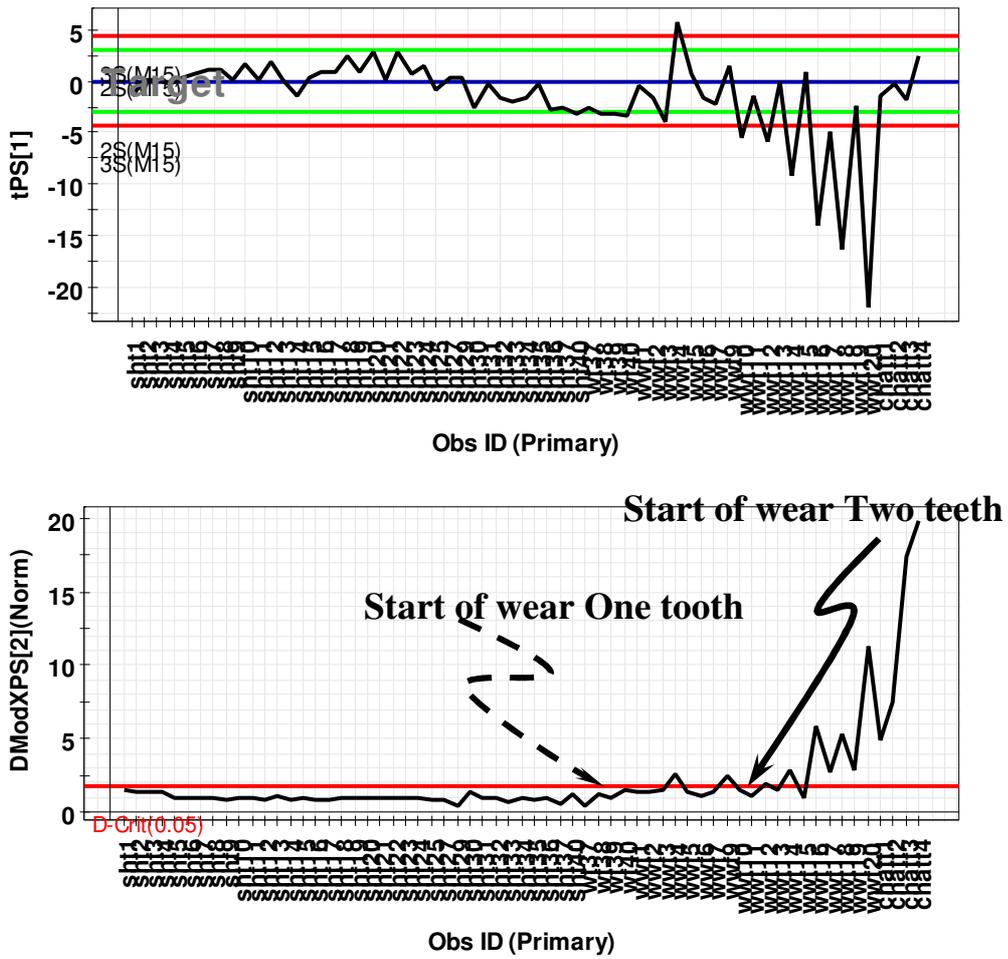


Figure 8. PLS Shewart control chart.

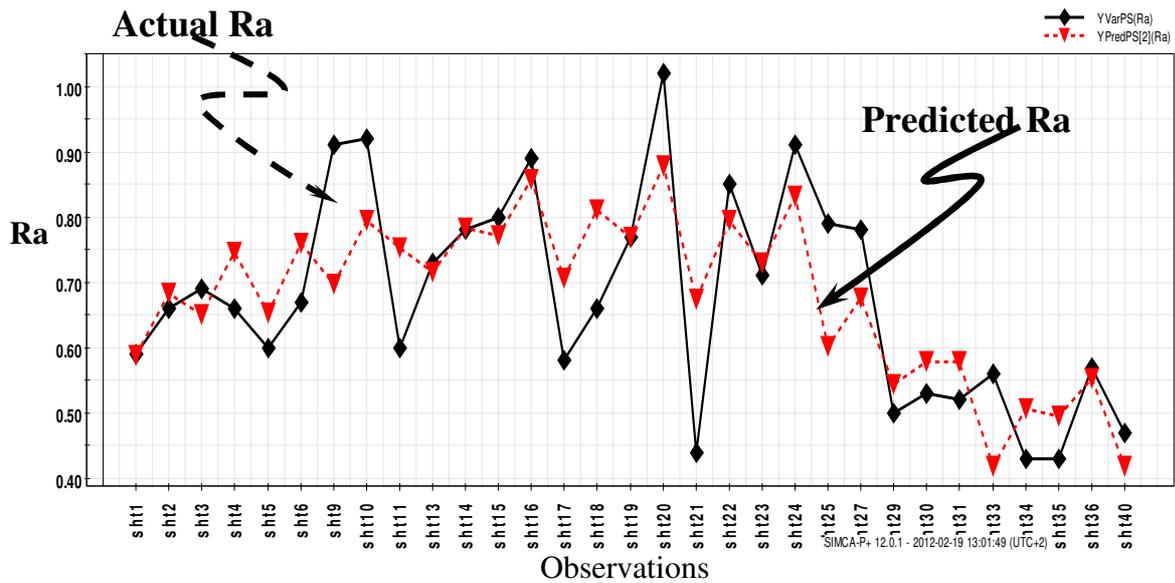


Figure 9. Time series of actual and predicted Ra for three machines.