

# Monitoring Nonlinear Profiles Using Support Vector Machines

Javier M. Moguerza<sup>1</sup>, Alberto Muñoz<sup>2</sup>, and Stelios Psarakis<sup>3</sup>

<sup>1</sup> University Rey Juan Carlos, Camino del Molino s/n, 28943 Fuenlabrada, Spain  
javier.moguerza@urjc.es

<sup>2</sup> University Carlos III, c/ Madrid 126, 28903 Getafe, Spain  
alberto.munoz@uc3m.es

<sup>3</sup> Athens Univ. of Econ. and Business, 76 Patission Str., 10434 Athens, Greece  
psarakis@aueb.gr

**Abstract.** In this work we focus on the use of SVMs for monitoring techniques applied to nonlinear profiles in the Statistical Process Control (SPC) framework. We develop a new methodology based on Functional Data Analysis for the construction of control limits for nonlinear profiles. In particular, we monitor the fitted curves themselves instead of monitoring the parameters of any model fitting the curves. The simplicity and effectiveness of the data analysis method has been tested against other statistical approaches using a standard data set in the process control literature.

**Keywords.** Kernel methods, Statistical Process Control, Support Vector Machines.

## 1 Introduction

The technique of Statistical Process Control (SPC) was introduced by Shewhart in 1924 [9]. The basic tools of SPC are the control charts. The central aim of control charts is the monitoring of processes, where the performance is determined by quality characteristics. In particular, control charts are useful to test the stability of a process. The process is often represented by a curve or function (called profile) that involves a response variable  $Y$  and a number of explanatory variables  $(X_1, X_2, \dots)$ . The simplest profiles are generated by linear models, but often more realistic nonlinear models are demanded.

In this paper, we are concerned with a different class of profile data that cannot be adequately represented by a linear structured model. This kind of data are generally known as nonlinear profiles. Very little work has been done for monitoring of nonlinear profiles (see [17,5,3]). For an overview of profile monitoring techniques see [18].

There are two phases involved in the SPC methodology. In Phase I, an historical amount of process data is available and has to be analyzed. The targets here are to test the stability of the process, to understand the variation in the process over time and to model the in-control process performance. In Phase II,

the goal is to monitor the process. Using on-line data, the system (adjusted in Phase I) has to perform a quick detection of shifts.

In Multivariate SPC, when the underlying distribution of the quality characteristics is not multivariate normal, the design of the control limits is not trivial, and regression Support Vector Machines (SVMs) may be used to design a control boundary limits which adapt to the shape of the data. In this work we propose a technique for Phase I analysis of nonlinear profiles using SVMs.

The rest of the paper is organized as follows. In Section 2 some background on profiles is given. In Section 3 SVMs for regression are briefly described. In Section 4 we show how to use SVMs for the monitoring of nonlinear profiles, and our method is compared to other state-of-the-art methods. Section 5 concludes.

## 2 Basics on Profiles

Profile data consist of a set of measurements with a response variable  $Y$  and one or more explanatory variables  $X^{(i)}$ ,  $i \in \{1, \dots, k\}$ . The explanatory variables are used to assess the quality of a manufactured item. In the case of linear profiles we have a linear regression model with a response variable  $Y$  and the vector of explanatory variables  $X = (X^{(1)}, \dots, X^{(k)})^T$ ,

$$Y = A_0 + A_1^T X + \varepsilon ,$$

where  $\varepsilon$  is the random error,  $A_0$  is a constant and  $A_1$  is a vector of parameters.

Applying the linear profile, we have to pool  $n$  set points  $x_1, \dots, x_n \in \mathbb{R}^k$ . For the sample  $i$  the process output is:

$$y_i = A_0 + A_1^T x_i + \varepsilon_i , \text{ with } i \in \{1, \dots, n\} ,$$

where  $\varepsilon_i$  are random variables, independent and normally distributed with mean zero and variance  $\sigma^2$ . The estimations of the vector  $A_1$  and the intercept  $A_0$  can be obtained using the least squares method (see [4]).

The extension of the linear model to the nonlinear case is

$$Y = f(X, \beta) + \varepsilon ,$$

where  $\varepsilon$  is the random error,  $\beta$  is a vector of parameters and  $f$  is a nonlinear function. Unlike linear profiles, which can be represented by a linear regression model whose regression model parameters can be used for monitoring and detection, nonlinear profiles are often sampled into high dimensional data vectors and analyzed by non-parametric methods. Examples of nonlinear profiles are given in [15,16].

## 3 Regression Support Vector Machines in a Nutshell

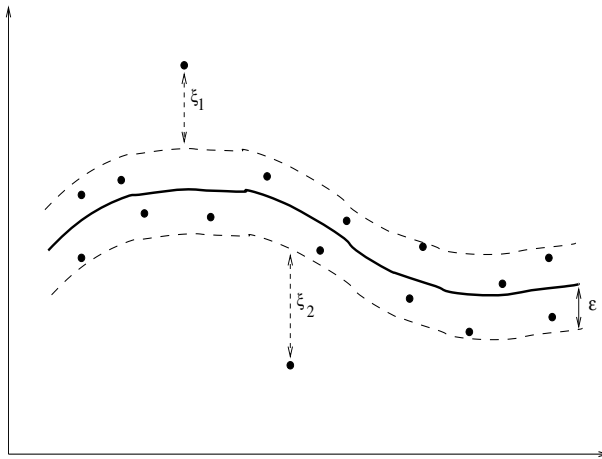
It is usual in SVMs tutorials to consider the regression problem from a geometric point of view. Consider a regression problem where the regression function is

nonlinear. The idea underlying Regression Support Vector Machines [13,2] is to map the data into a “Feature Space” so that the estimated regression function becomes linear. This mapping, denoted  $\Phi$  in the sequel, determines a dot product  $K$  in the Feature Space, called kernel, of the form  $K(x, y) = \Phi(x)^T \Phi(y)$ .

Regression SVMs work by solving the following optimization problem:

$$\begin{aligned}
 \min_{w, b, \xi, \xi'} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi'_i) \\
 \text{s.t.} & (w^T \Phi(x_i) + b) - y_i \leq \varepsilon + \xi_i \quad i = 1, \dots, n, \\
 & y_i - (w^T \Phi(x_i) + b) \leq \varepsilon + \xi'_i \quad i = 1, \dots, n, \\
 & \xi_i, \xi'_i \geq 0 \quad i = 1, \dots, n.
 \end{aligned}
 \tag{1}$$

Notice that  $\varepsilon$  appears only in the constraints, forcing the solution to be calculated taking into account a confidence band around the regression equation (see Figure 1).



**Fig. 1.** Support Vector Regression

The  $\xi_i$  and  $\xi'_i$  are slack variables that allow for some data points to stay outside the confidence band determined by  $\varepsilon$ . This is the standard support vector regression formulation. This optimization problem is convex and therefore, it has no local minima. The linear regression equation will take the form:

$$R^*(x) = w^{*T} \Phi(x) + b^* ,$$

where  $w^*$  and  $b^*$  are the values of  $w$  and  $b$  at the solution of the previous optimization problem. Notice that the inclusion of the mapping implies nonlinearity.

It can be shown that the problem described above can be stated as a regularization problem [12]. Regularization methods include, among others, smoothing splines [14] or kriging [1]. Regularization allows the construction of smooth

functions. In fact, it can be shown (see, for instance [7,8]) that the previous problem can be formulated as a regularization problem of the form:

$$\min_{f \in H_K} \frac{1}{n} \sum_{i=1}^n L(y_i, f(x_i)) + \mu \|f\|_K^2, \quad (2)$$

where  $\mu > 0$ ,  $H_K$  is the Reproducing Kernel Hilbert Space (RKHS) associated to the kernel  $K$ ,  $\|f\|_K$  denotes the norm of  $f$  in the RKHS,  $(x_i, y_i)$  are the sample data points, and  $L(y_i, f(x_i)) = (|f(x_i) - y_i| - \varepsilon)_+$ ,  $\varepsilon \geq 0$  is known as the ‘ $\varepsilon$ -insensitive loss’ function. By the Representer Theorem (see, for instance [7]), the solution to problem (2) has the form  $f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) + b$ , where  $x_i$  are the sample data points. Notice that the penalization term  $\mu \|f\|_K^2$  in eq. (2) favours the choice of smooth solutions to the optimization problem under consideration. It is immediate to show that  $\|f\|_K^2 = \|w\|^2$ , where  $w = \sum_i \alpha_i \Phi(x_i)$  and  $\Phi$  is the mapping defining the kernel function. In this way,  $R^*(\mu) = f(x)$ .

## 4 An Application of SVMs to Nonlinear Profiles

We will apply Regression SVMs to the example given in [15,16,18]. The data set is made up of 24 curves, each one corresponding to a vertical density profile (VDP). These data come from the manufacture of engineered woodboards. In the production of engineered wood boards the control of the density properties of the boards produced is crucial. Multiple measurements on a sample (usually a  $2 \times 2$  inch piece) to form the vertical density profile of the board have been taken. Figure 2 shows the 24 profiles. Each one consists of 314 measurements taken 0.002 inches apart. The VDPs are divided in three groups. Each group is made up by boards selected in the same 8-hour shift. Group 1 consists of 9 boards (1, . . . , 9), group 2 consists of 11 boards (10, . . . , 20), and group 3 consists of 4 boards (21, . . . , 24).

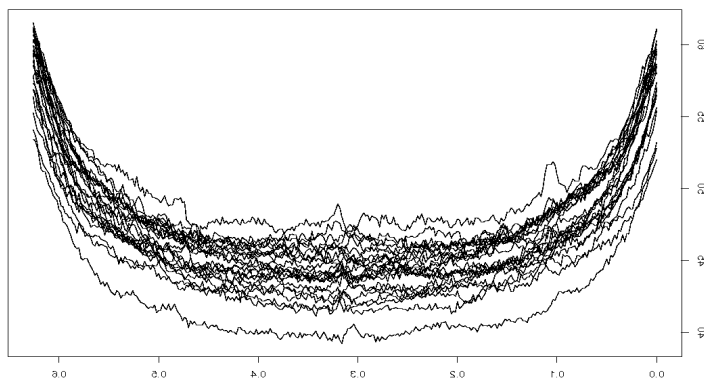
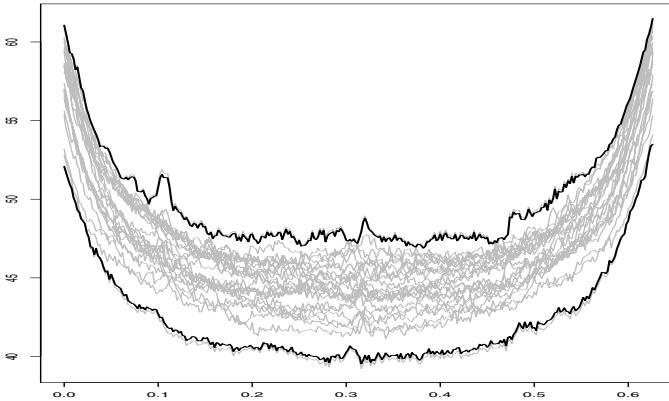


Fig. 2. 24 profiles. Each one consists of 314 measurements

The way in which we will apply Regression SVMs for profile monitoring falls into the area of Functional Data Analysis [11]. We will calculate the control limits by regularizing the percentile curves obtained from the profiles. Figure 3 shows, in bold, the 0.005 and 0.095 percentiles for the 24 curves. Between both percentiles, 99% percent of the curves should be contained. In the following application we will use the Gaussian kernel  $K(x, y) = e^{-\|x-y\|^2/\sigma}$  with  $\sigma = 0.025$ , and in the SVM formulation we will take  $C = 100$ . The parameters have been fixed by cross-validation (see [14] and [6] for details).

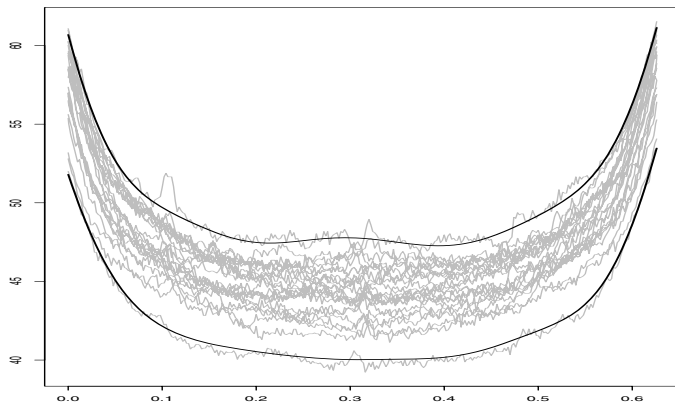


**Fig. 3.** 0.005 and 0.095 percentiles for the 24 curves

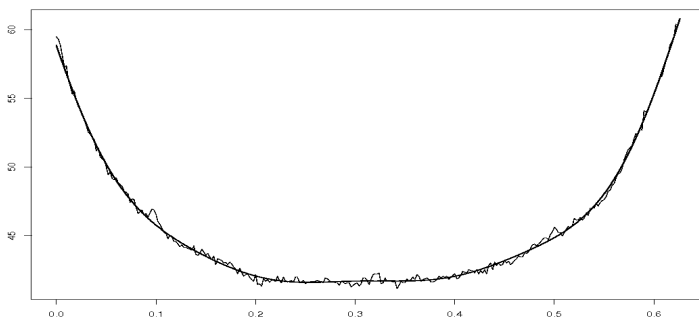
In order to obtain smooth control limits, we regularize these percentile curves using a Regression Support Vector Machine. Figure 4 shows the 24 curves and the regularized control limits in bold.

Using these control limits, outlying profiles can be calculated. In this application, we will consider a profile as out of bounds if its regularized version at some moment comes out of the band delimited by the upper or lower regularized control limits. The use of the regularized version of the profiles is justified by the need of avoiding the effect of the peaks of the nonregularized curves. Figure 5 shows the estimated curve for one of the profiles, in particular for the profile named 10. Figure 6 shows the regularized curves for all the profiles, and the control limits (the two bold curves).

Using our methodology, three profiles are obtained as out of bounds, in particular, profiles 3, 6 and 16. The regularized version of curve 3 is out of bounds 279 times, curve 6 is out of bounds 296 times and curve 16 is out of bounds 11 times. The remaining curves remain always between the control limits. Since each profile consists of 314 measurements, we may design a measure (taking values between 0 and 1) to decide how outlying is a profile (the proportion of time that its regularized version is out of control). In this way, the results for these three profiles are 0.89, 0.94 and 0.035, respectively. Figure 7 shows the outlying regularized profiles (black lines) and the control limits (dashed lines).



**Fig. 4.** The 24 curves and the regularized control limits in bold



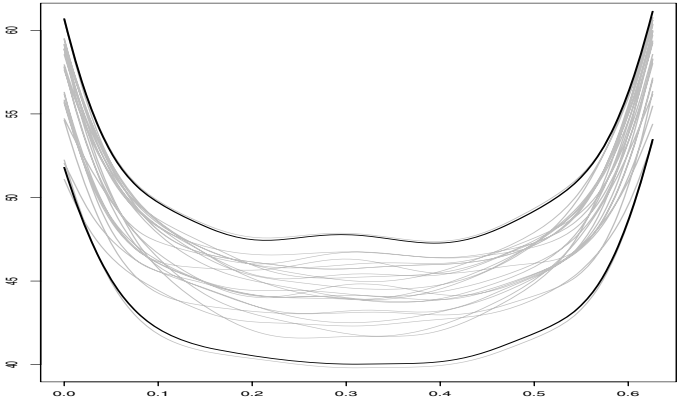
**Fig. 5.** Profile 10 and its corresponding regularized curve

Curves 3 and 6 are the ones clearly out of bounds, being their corresponding values for the outlying measure high, that is, 0.89 and 0.94 respectively. Curve 16 is out of bounds only in the tails of the lower control limit (being its measure 0.035, a low value). For the shake of completeness, Figure 8 shows the regularized control limits and the non-regularized version of the outlying curves.

Notice that using less restrictive criteria in order to decide whether a given curve is out of bounds (for instance, a percentage of times coming out of the limits), fewer curves would be considered as outlying ones.

#### 4.1 Comparison with Other Methods

The analysis of nonlinear profiles has been previously afforded using standard statistical methods. In particular, in [17] three methods are applied to this data set. These methods are based on the monitoring of the parameters obtained by estimating each VDP using a ‘bathtub’ function model. Given the data points



**Fig. 6.** The 24 regularized curves and the regularized control limits in bold

$(x_i^k, y_i^k)$  for the  $i^{th}$  profile (in our case  $i \in \{1, \dots, 24\}$ ,  $k \in \{1, \dots, 314\}$ ), the following model has to be estimated:

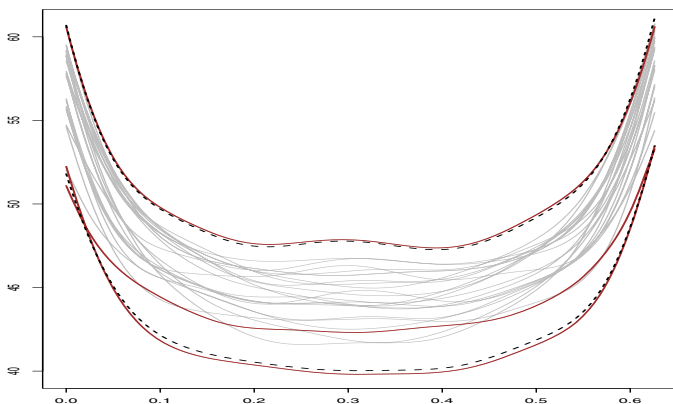
$$f_i(x_i^k, \beta_i) = \begin{cases} a_{i1}(x_i^k - d_i)^{b_{i1}} + c_i & , x_i^k > d_i , \\ a_{i2}(-x_i^k + d_i)^{b_{i2}} + c_i & , x_i^k \leq d_i , \end{cases} \quad (3)$$

where  $x_i^k$  is the  $k^{th}$  regressor variable value for the  $i^{th}$  profile, and  $\beta_i = (a_{i1}, a_{i2}, b_{i1}, b_{i2}, c_i, d_i)^T$  is the vector of parameters corresponding to the model estimating the  $i^{th}$  profile. Notice that there exists a ‘bathtub’ model for each of the 24 profiles. Once  $\beta_i$  has been estimated by  $\hat{\beta}_i$ ,  $i \in \{1, \dots, 24\}$ , three covariance matrices  $S_1$ ,  $S_2$  and  $S_3$  (for methods 1, 2 and 3, respectively) are calculated (see [17] for details). Then three  $T^2$ -statistics of the form  $T_i^2 = (\hat{\beta}_i - \tilde{\beta}_i)^T S_i^{-1} (\hat{\beta}_i - \tilde{\beta}_i)^T$ ,  $i \in 1, 2, 3$  are calculated in order to monitor the parameters simultaneously. These  $T_i$  follow standard statistical distributions and control limits can be calculated.

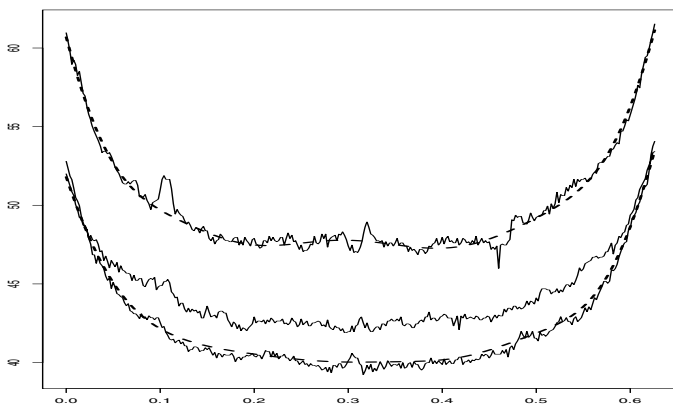
In [17] it is shown that, for the three statistics, board 15 is outlying. For two of the three statistics board 18 is borderline, while the third statistic indicates that boards 4, 9, 18 and 24 are outliers. It is apparent that none of these boards are outliers as they are included within the calculated control limits.

### 4.2 Principal Component Analysis

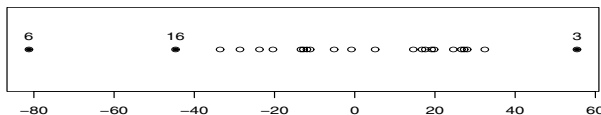
The results obtained with the regularization methodology are coherent with the ones of the following experiment. We have considered each profile as a single data point of dimension 314. Then, we have made a Principal Component Analysis (PCA) on the 24 profiles. The first principal component explains 98% of the variability. Figure 9 shows the 24 profiles projected on this first principal component, with profiles 3, 6 and 16 in bold. It is clear that the most outlying



**Fig. 7.** Outlying regularized curves. Control limits in dashed.



**Fig. 8.** Regularized control limits (dashed) and the non-regularized version of the outlying curves



**Fig. 9.** The 24 profiles projected on the first principal component, with profiles 3, 6 and 16 in bold

profiles are these three ones, being profiles 3 and 6 the most outlying ones. The PCA method can be considered a qualitative method rather than a quantitative one. Therefore, it does not provide a decision value to determine if a given profile can be considered as an outlier, while the method that we propose does.



## 5 Conclusions

In this work we have developed a new methodology based on Functional Data Analysis for the construction of control limits for nonlinear profiles. Our technique is not based in the monitoring of the parameters of any model fitting the curves. Instead, we monitor the fitted curves themselves. Therefore, we do not need to find a particular nonlinear regression model that must adapt to the particular shape of the profiles in each case. The use of the regularized curve corresponding to each profile has the effect of automatically considering the variability within the profile. Similarly, by building the percentile curves and the control limits we are taking into account the variability among different profiles.

The simplicity and effectiveness of the data analysis method has been tested against other statistical approaches using a standard data set in the process control literature. In particular we have shown that our method outperforms other state-of-the-art methods in term of outlying board detection.

Further research will include the study of density estimation methods to detect the outlying curves, by considering the profiles as data points in a functional data space. To this aim, we will extend the techniques presented in [10] to the functional data case.

**Acknowledgments.** This work was partially supported by spanish grants MEC MTM2004-02334, MEC MTM2006-14961-C05-05 and URJC-CM-2006-CET-0391.

## References

1. Cressie, N.: *Statistics for Spatial Data*. Wiley, Chichester (1993)
2. Cristianini, N., Shawe-Taylor, J.: *An Introduction to Support Vector Machines*. Cambridge University Press, Cambridge (2000)
3. Ding, Y., Zeng, L., Zhou, S.: Phase I Analysis for Monitoring Nonlinear Profiles in Manufacturing Processes. *Journal of Quality Technology* 38(3), 199–216 (2005)
4. Draper, N., Smith, H.: *Applied Regression Analysis*. Wiley, Chichester (1997)
5. Jensen, W.A., Birch, J.B., Woodall, W.H.: Profile Monitoring via Linear Mixed Models. *Journal of Quality Technology* (to appear, 2007)
6. Keerthi, S.S., Lin, C.J.: Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Computation* 15, 1667–1689 (2003)
7. Moguerza, J.M., Muñoz, A.: Support Vector Machines with Applications. *Statistical Science* 21(3), 322–336 (2006)
8. Moguerza, J.M., Muñoz, A.: Rejoinder to Support Vector Machines with Applications. *Statistical Science* 21(3), 358–362 (2006)
9. Montgomery, D.C.: *Introduction to Statistical Quality Control*, 2nd edn. Wiley, Chichester (1990)
10. Muñoz, A., Moguerza, J.M.: Estimation of high density regions using One-Class Neighbor Machines. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 28(3), 476–480 (2006)
11. Ramsay, J.O., Silverman, B.W.: *Functional Data Analysis*. Springer, Heidelberg (1997)

12. Tikhonov, A.N., Arsenin, V.Y.: Solutions of ill-posed problems. John Wiley & Sons, New York (1977)
13. Vapnik, V.: The Nature of Statistical Learning Theory. Springer, New York (1995)
14. Wahba, G.: Spline Bases, Regularization, and Generalized Cross Validation for Solving Approximation Problems with Large Quantities of Data. In: Cheney, W. (ed.) Approximation Theory III, pp. 905–912. Academic Press, London (1980)
15. Walker, E.: Determining the Smoothing Parameter in Penalized Regression Splines Using Model Selection Criteria. Talk to the IMPS, Osaka, Japan (2001)
16. Walker, E., Wright, W.P.: Comparing Curves with Additive Models. *Journal of Quality Technology* 34(1), 118–129 (2002)
17. Williams, J.D., Woodall, W.H., Birch, J.B.: Statistical Monitoring of Nonlinear Product and Process Quality Profiles. *Quality and Reliability Engineering International* (to appear, 2007)
18. Woodall, W.H., Spitzner, D.J., Montgomery, D.C., Gupta, S.: Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology* 36, 309–320 (2004)