

## Chapter 9<sup>1</sup>

# Multivariate Control Charts from a Data Mining Perspective

Giovanni C. Porzio

Department of Economics, University of Cassino,  
Via S. Angelo I-03043 Cassino (FR) - Italy, [porzio@eco.unicas.it](mailto:porzio@eco.unicas.it)

Giancarlo Ragozini

Department of Sociology, Federico II University of Naples  
Vico Monte di Pietà 1, I-80132 Naples - Italy, [gragoz@unina.it](mailto:gragoz@unina.it)

**Abstract:** This chapter aims at presenting our data mining vision on Statistical Process Control (SPC) analysis, specifically on the design of multivariate control charts for individual observations in the case of independent data and continuous variables. We first argue why the classic multivariate SPC tool, namely the Hotelling  $T^2$  chart, might not be appropriate for large data sets, and then we provide an up-to-date critical review of the methods suitable for dealing with data mining issues in control chart design. In order to address new SPC issues such as the presence of multiple outliers and incorrect model assumptions in the context of large data sets, we suggest exploitation of some multivariate nonparametric statistical methods. In a model-free environment, we present the way we handle large data sets: a multivariate control scheme based on the data depth approach. We first present the general framework, and then our specific idea on how to design a proper control chart. There follows an example, a simulation study, and some remarks on the choice of the depth function from a data mining perspective. A brief discussion of some open issues in data mining SPC closes the chapter.

**Key Words:** Data depth, Convex hull peeling, Hotelling's  $T^2$ , Nonparametric Control Charts, Outliers, Robustness, Skewness.

---

<sup>1</sup> Liao, T.W. and E. Triantaphyllou, (Eds.), **Recent Advances in Data Mining of Enterprise Data**, *World Scientific*, Singapore, pp. 413-462, 2007.