Dynamic PCA structure induced autocorrelation

Ana S. Matos, Teresa S. Neves, and João T. Mexia

New University of Lisbon, Portugal

Abstract

In the last years, multivariate statistical process control (MSPC) methods, and namely principal component analysis (PCA), have shown to provide a powerful approach to detection and isolation of abnormal conditions, in process industries with highly correlated variables. The use of PCA in MSPC assumes implicitly that the observations at one time are statistically independent of past observations and that the latent variables follow a Gaussian distribution. However, in chemical or biological processes, variables are driven by random noise and uncontrollable disturbances, which may cause process variables to be auto-correlated. Several statistical methods have been developed and applied to overcome the existence of autocorrelation in data. In this article, a comparative study of the performance between the well known Hotelling T^2 control chart, using residuals or one-step-ahead predictions, and the dynamic PCA (DPCA) method based on time lagged measurement vector, proposed by Ku et al. (1995), is presented. The approach developed to compare those charts is described in detail, using the average run length (ARL) as a performance indicator. Monte Carlo experiments are used to simulate three first order autoregressive models (AR(1)), with different autoregressive parameters and different variances, without cross correlation between them. ARL results for a step perturbation introduced in one, two, or all variables simultaneously show that the DPCA chart presents a steady good performance for all shift magnitudes, when compared with Hotelling T^2 chart, with significant reduction in SDRL (standard deviation run length), evidencing a better reliability in the fault detection process. Although this study, with a small dimension of variables, has shown that DPCA method achieves better results on detecting and monitoring disturbances in dynamic processes (recommended by other authors like Lee et al. (2004) and Chen and Liu (2002) as an example), the dynamic PCA cannot eliminate data autocorrelation, regardless of the time lag considered. As a consequence of this evidence, the influence of the dynamic PCA structure on data is investigated using different time series (AR, IMA, ARMA) and also random noise series. Time lag is also considered as a variable in this extension of the study. Through Monte

Carlo simulation is proven that dynamic PCA structures induce dynamics into the score variables, even if all the variables correspond to random noise series. The main advantages and disadvantages of each chart are pointed out, in a practical perspective of those who intent to use MSPC to monitoring dynamic continuous processes with a small number of variables to be controlled. Simulated and real data are used for illustration.

Keywords

Multivariate statistical process control (MSPC), Hotelling T^2 control chart, Dynamic principal component analysis (DPCA), Average Run Length (ARL).

References

Chen, J. and Liu, K.-C. (2002). On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chemical Engineering Science 57*, 63-75.

Lee, J.-M., Yoo, C., and Lee, I.-B. (2004). Statistical monitoring of dynamic processes based on dynamic independent component analysis. *Chemical Engineering Science 59*, 2995–3006.

Ku, W., Storer, R.H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems 30*, 179-196.