

Seminari del Departament d'Estadística i Investigació Operativa

Universitat Politècnica de Catalunya

13th April 2026

Title: Rethinking Statistics in a Digitalized World: The Potential of Multivariate Statistical Models of Latent Variables in Industry and Health 4.0.

Author: Alberto José Ferrer Riquelme (Universitat Politècnica de València).

Web: <https://www.upv.es/ficha-personal/aferrer>.

Key Words: Digitalization; Data Science; Latent variable models; Principal Component Analysis; Partial least squares regression; Machine Learning

Abstract: The 4.0 era represents the Fourth Industrial Revolution, a new era of innovation marked by the integration of digital technologies such as Artificial Intelligence (AI), the Internet of Things (IoT) and Big Data, to create interconnected, automated and data-centric systems, which can improve efficiency, personalization and decision-making.

In this new era of digitalization, however, there is a prevailing belief that, due to the enormous amount and speed with which data is generated, these emerging digital technologies can solve important problems in society, industry, or healthcare solely through the analysis of empirical data, without the need for scientific models. theory, experience or knowledge of the sector. The idea is that causality no longer matters, just simple correlation is enough. Some even go so far as to claim that Statistics has died due to the rise of Data Science. They consider statistics obsolete and statisticians professionals specialized in techniques that are no longer useful, and concerned with issues that seem irrelevant in the face of the complexity of the challenges of the twenty-first century.

In this talk, through a SWOT (Strengths Weaknesses Opportunities and Threats) analysis of Statistics in the digitized world, I will try to reveal the reasons for these beliefs about the obsolescence of Statistics, and share my perspective on the key factors that could help Statistics regain its social recognition and its fundamental role as a key component to successfully address many of the problems facing our society.

Many of the ideas come from the teachings of Professor George E. P. Box, an eminent English statistician who died a decade ago, and Professor John MacGregor, a prominent Canadian chemical engineer and statistician, a disciple of Professor Box, who have been some of the most influential people in my professional career.

I will highlight the potential of Latent Variable Multivariate Statistical Models, and in particular Principal Component Analysis (PCA) and Partial Least Squares (PLS) Regression, to address some of the main challenges of industry and healthcare in highly digitized environments (Industry and Health 4.0). Among them, I will highlight the development of imaging biomarkers for the early diagnosis of cancer using non-orthogonal PCA, the design of Multivariate Statistical Process Control (MSPC) schemes and the multivariate analysis of images based on PCA, as well as the optimization of processes, the definition of the raw material design space and the development of multivariate capacity indices based on latent variables using historical data (not from any

experimental design) using PLS. These approaches leverage the PLS's ability to model causality in latent space, even from historical data, typical in highly digitized environments. I will also discuss the advantages/disadvantages of these techniques with others from Machine Learning such as Random Forests, Neural Networks, etc., and from Classical Statistics such as Multiple Linear Regression and Logistic Regression.

All topics addressed will be illustrated through real examples from industry (chemical, pharmaceutical and food) and health (cancer diagnosis by imaging) in which I have been involved in recent years, as a result of collaboration in research projects and consultancy work with several companies and hospitals.

References:

Borràs-Ferrís J, Palací-López D, Duchesne C, Ferrer A. Defining multivariate raw material specifications in industry 4.0. *Chemom Intel Lab Syst.* 2022;225, 104563.

Borràs-Ferrís J, Duchesne C, Ferrer A. Defining multivariate raw material specifications via SMB-PLS. *Chemom Intel Lab Syst.* 2023;240, 104912.

Borràs-Ferrís J, Duchesne C, Ferrer A. A latent space-based multivariate capability index: A new paradigm for raw material supplier selection in industry 4.0. *Chemom Intel Lab Syst.* 2025;258, 105339.

Box GEP. *Science and Statistics.* *J. Am. Stat. Assoc.* 1976;71(356):791-799.

Box GEP, Hunter JS, Hunter WG. *Statistics for Experimenters: Design, Discovery, and Innovation.* 2nd edn. John Wiley & Sons, Inc. 2005.

Ferrer, A. (2007). Multivariate statistical process control based on principal component analysis (MSPC-PCA): Some reflections and a case study in an autobody assembly process. *Qual Eng.* 2007;19:311–325.

Ferrer A. Latent structures-based multivariate statistical process control: a paradigm shift. *Qual Eng.* 2014;26(1):72-91.

Ferrer A. Discussion of “A review of data science in business and industry and a future view” by Grazia Vicario and Shirley Coleman. *Appl Stochastic Models Bus Ind.* 2020;36:23–29.

Ferrer A. Multivariate six sigma: A key improvement strategy in industry 4.0. *Qual Eng.* 2021;33:758–763.

Ferrer A, Borràs-Ferrís J, García-Carrión S. Data Analytics Strategies to Exploit Historical Databases for Process Optimization and Innovation in Digitalized Industry 4.0. In: di Bella, E., Gioia, V., Lagazio, C., Zaccarin, S. (eds) *Statistics for Innovation I.* SIS 2025. Italian Statistical Society Series on Advances in Statistics. Springer, Cham. 2025:33-37.

García-Carrión S, Pozueta L, Ferrer A. Enhancing Six Sigma with latent variable models: An industrial application of multivariate Six Sigma in the automotive sector. *Qual Eng.* 2026-02-17 | Journal article DOI: [10.1080/08982112.2026.2626827](https://doi.org/10.1080/08982112.2026.2626827)

Jaekle CM, MacGregor JF. Industrial applications of product design through the inversion of latent variable models. *Chemom Intel Lab Syst.* 2000;50:199-210.

MacGregor, JF. Empirical models for analyzing “big” data - what’s the difference. In: Spring AIChE Conf., Orlando, Florida, USA 2018.

Prats-Montalbán JM, Aguado-Sarrió E, Ferrer A. Multivariate Curve Resolution for Magnetic Resonance Image analysis: applications in prostate cancer biomarkers development, in: *Resolving Spectral Mixtures, with Application from Ultrafast Spectroscopy to Super-resolution Imaging, Data Handling in Science and Technology* vol. 30, Elsevier, Amsterdam, 2016, pp. 519–550.

Tomba E, Barolo M, García-Muñoz S. General framework for latent variable model inversion for the design and manufacturing of new products. *Ind Eng Chem Res.* 2012;51:12886-12900.

Bio:

Alberto Ferrer is a professor at the Department of Statistics and Applied Operations Research and Quality at the Universitat Politècnica de València - UPV (Spain). He is the researcher in charge of the Multivariate Statistical Engineering Research Group (GIEM), dedicated to the development and application of multivariate statistical methods based on latent variables for problem solving in industry, health and biotechnology. In line with his research activity, he has established a solid collaboration with national and international companies from different sectors (chemical, pharmaceutical, and food), as well as with researchers from different countries.

He has been an associate editor of *Technometrics* (2008-2009) and a member of the editorial board of *Quality Engineering* (2006-2013). Since 2019, he has been a member of the Editorial Board of *Chemometrics and Intelligent Laboratory Systems* and of the *Journal of Chemometrics*.

His main research interest is focused on the integration of machine learning and multivariate statistics in Data Science projects to address the challenges that digitalization is generating in industry, health and technology.

He is Scientific Director and Co-founder of Kenko Imalytics, S.L., a health technology spin-off company aimed at leveraging multivariate image analysis technology developed in the GIEM group to develop advanced imaging biomarkers for early cancer diagnosis from medical image analysis. He is also a scientific advisor and co-founder of Kensight Solutions, S.L., a spin-off consulting company specialized in improving the quality and productivity of industrial processes and the service sector through the intelligent application of strategies based on statistical models and machine learning.

He is a member of the International Association for Business and Industrial Statistics (ISBIS), the European Network for Business and Industrial Statistics (ENBIS) and the International Statistical Engineering Association (ISEA), as well as elected member of the International Statistical Institute (ISI).

In recognition of his outstanding contributions to the development and application of statistical methods in business and industry, he received the 2025 George Box Medal in 2025.