

CURRICULUM VITAE Gilbert SAPORTA

Born : April, 8, 1946.

Emeritus Professor of Applied Statistics
Conservatoire National des Arts et Métiers (CNAM)
292 rue Saint Martin 75141 PARIS Cedex 03
Phone : +33 1 40 27 22 68
e-mail: gilbert.saporta@cnam.fr
webpage : <http://cedric.cnam.fr/~saporta>

Guest professor 2017-2022
Beihang University
Xue Yuan Road No.37, HaiDian District,
Beijing 100191
China.

EDUCATION

Ingénieur Ecole Centrale Paris (1968).
Master in mathematical statistics, Université Paris VI (1969).
Graduated from Institut de Statistique de l'Université de Paris (1970).
Docteur (Ph.D.) Statistique, Université Paris VI (1975).
Docteur ès Sciences Mathématiques, Université Paris VI (1981).

TEACHING EXPERIENCE

at Conservatoire National des Arts et Métiers (1984-2014):

- Unit STA 108 « Survey Sampling »
- Unit STA 201 « Advanced Multivariate Analysis »

others:

1969-1984 Assistant professor **Paris Institute of Technology**
1970-1972 lecturer Bachelor in Psychology **Université Paris X**
1971-1978 lecturer **Université de Paris VI**.
1970-1983 Associate professor **Ecole Nationale Supérieure des Mines de Paris**.
1977-1983 Associate professor **Ecole Nationale Supérieure du Pétrole et des Moteurs (IFP school)**
1983-1991 Associate professor **Ecole Centrale Paris**
1983-2007 lecturer **Ecole Nationale de la Statistique et de l'Administration Economique**
1986-2000 lecturer **Institut de Statistique de l'Université Paris VI (ISUP)**

E-learning

Co direction of [StatNet](http://www.agro-montpellier.fr/cnam-lr/statnet/) : a web-based course in Introductory Statistics <http://www.agro-montpellier.fr/cnam-lr/statnet/> .

RESEARCH

Topics:

Data science

- Categorical and mixed data analysis by means of optimal scaling techniques and correspondence analysis
- Functional Data Analysis
- Supervised classification with applications in risk analysis (credit scoring)
- Confidence region and validity in data analysis
- Data mining, Big Data and Statistical Learning Theory

Head of the research team MSDMA « Statistical Methods for Data Mining and Learning » at CEDRIC-CNAM (15 researchers) until 2014

Ph.D. supervision:

- 27 Ph.D ; 2 in progress

Associations:

- Honorary President of [Société Française de Statistique](#)
- President of the non-profit Foundation "La Science Statistique"
- Past President of IASC ([International Association for Statistical Computing](#)) 2005-2007
- Former Vice-president of ISI (International Statistical Institute). 2005-2007
- Former President of Société de Statistique de France & Société de Statistique de Paris.
- Former president of ASU (Association pour la Statistique et ses Utilisations) 1986-1988;
- Member of [Société Francophone de Classification](#), [Psychometric Society](#), [International Association for Statistical Education](#), [Societa Italiana di Statistica](#),
- Honorary member of *Societatea de Probabilitati si Statistica din Romania*

Editorial boards

- [Revstat](#), [Advances in Data Analysis and Applications](#) , [Big Data's Applications](#), [CSBIGS](#), [Journal of Classification](#), [Journal de la Société Française de Statistique](#)

Working group:

Member of the Big Data group of CROS- Eurostat <http://www.cros-portal.eu/content/big-data>

CONSULTING

L'Oréal Paris (Cosmetics), IPSOS (Market Research), Tricast (Insurance Business Intelligence)

RECENT PUBLICATIONS (after 2005)

Books :

- « Probabilités, analyse des données et statistique », 3rd edition, 656 pages, Editions Technip, Paris, 2010
- « L'analyse des données » Que Sais-Je n°1854 , 50 000 copies (9th edition)
- Co-editor of « Modèles statistiques pour données qualitatives », Technip, 2005
- Co-editor of « Analyse statistique des données spatiales », Technip, 2006
- Co-editor of « Analyse statistique des données longitudinales », Technip, 2010
- Co-editor of « Approches non paramétriques en régression », Technip, 2011
- Co-editor of « Modèles à variables latentes et modèles de mélange », Technip, 2013
- Co-editor of « Approches statistiques du risque », Technip, 2014
- Co-editor of « Méthodes robustes en statistique », Technip, 2015
- Co-editor of “The Multiple Facets of Partial Least Squares Methods”, Springer Proceedings in Mathematics and Statistics, pp. 268, 2016

Pair-reviewed journals :

- C. Preda et **G. Saporta**. Clusterwise PLS regression on a stochastic process . *Computational Statistics & Data Analysis*, 49(1): 99-108, 2005.
- C. Preda et **G. Saporta**. PLS regression on a stochastic process . *Computational Statistics & Data Analysis*, 48(1): 149-158, 2005
- A. Tenenhaus, A. Giron, E. Viennet, M. Béra, **G. Saporta** et B. Fertil. Kernel logistic PLS: A tool for supervised nonlinear dimensionality reduction and binary classification . *Computational Statistics & Data Analysis*, 51(9): 4083-4100, 2007.
- C. Preda, **G. Saporta** et C. Lévêder. PLS classification of functional data . *Computational Statistics*, 22(2): 223-235, 2007
- M. Plasse, N. Niang, **G. Saporta**, A. Villemainot et L. Leblond. Combined use of association rules mining and clustering methods to find relevant links between binary rare attributes in a large data set. *Computational Statistics & Data Analysis*, 52(1): 596-613, 2007
- S. Benammou, G. Saporta et B. Souissi. Une procédure de réduction du nombre de paires en analyse conjointe *Journal de la Société Française de Statistique*, 148(4): 57-76, 2007.
- M. Vichi et **G. Saporta**. Clustering and Disjoint Principal Component Analysis . *Computational Statistics & Data Analysis*, 52(0), 2008
- S. Zaugg, **G. Saporta**, E. van Loon, H. Schmaljohann et F. Liechti. Automatic identification of bird targets with radar via patterns produced by wing flapping . *J. R. Soc. Interface*, vol. 5(26), pp. 1041-1053, 2008
- M. Charrad, Y. Lechevallier , **G. Saporta**, M. Ben Ahmed - Web Content Data Mining : la classification croisée pour l'analyse textuelle d'un site Web, *Revue des Nouvelles Technologies de l'Information* (RNTI-E-11), vol. 1, pp. 43-54, 2008
- H. Wang , M. Chen , **G. Saporta**. Gram-Schmidt regression and application in cutting tool abrasion prediction, *Journal of Beijing University of Aeronautics and Astronautics*, vol. 34(6), pp. 729-733, 2008 (in chinese)
- H. Wang , M. Ye , **G. Saporta**. Classification for Multiple Linear Regression Methods, *Journal of System Simulation*, vol. 21(22), pp. 7048-7056, 2009 (in chinese)
- G. Youness, **G. Saporta**: Comparing partitions of two sets of units based on the same variables, *Advances in Data Analysis and Classification*, vol. 4(1), pp. 53-64, 2010
- A. Aguilera , M. Escabias , C. Preda , **G. Saporta** - Using basis expansions for estimating functional PLS regression: Applications with chemometric data, *Chemometrics and Intelligent Laboratory Systems*, vol. 104(2), pp. 289-305, 2010
- C. Preda , **G. Saporta** , M. Hadj Mbarek - The NIPALS Algorithm for Missing Functional Data, *Revue Roumaine de Mathématiques Pures et Appliquées*, vol. 55(4), pp. 315-326, 2010

- H. Wang , M. Chen , **G. Saporta**. Variable selection in discriminant analysis based on Gram-Schmidt process, *Journal of Beijing University of Aeronautics and Astronautics*, vol. 37(8), pp. 958-961, 2011 (in chinese)
- J. Séguéla, **G. Saporta**. A hybrid recommender system to predict online job offer performance, *Revue des Nouvelles Technologies de l'Informatique*, vol. RNTI-E-25, pp. 177-197, 2013
- L. Berger , J. Macé , J. Ricco , **G. Saporta**. Methodology for the evaluation of vascular surgery manpower in France, *Public Health*, vol. 127, pp. 65-71, 2013
- N. Niang-Keita, F. Fogliatto, **G. Saporta**. Non parametric on-line control of batch processes based on STATIS and clustering, *Journal de la Société Française de Statistique*, vol. 154(3), pp. 124-142, 2013
- C. Gomes , H. Nocairi , M. Thomas , J. Collin , **G. Saporta**. A simple and robust scoring technique for binary classification, *Artificial Intelligence Research*, vol. 3(1), pp. 52-58, 2014,
- H. Hassani , **G. Saporta**, E. Silva. Data Mining and Official Statistics: The Past, the Present and the Future, *Big Data*, vol. 2(1), pp. 1-9, 2014
- G. Saporta - A Conversation with Jean-Louis Bodin, *International Statistical Review*, vol. 83(1), pp. 2-16, 2015,
- H. Nocairi , C. Gomes , M. Thomas , **G. Saporta**. "Improving Stacking Methodology for Combining Classifiers; Applications to Cosmetic Industry", *Electronic Journal of Applied Statistical Analysis*, vol. 9(2), pp. 340-361, 2016
- F. Husson, J. Josse , **G. Saporta**. "Jan de Leeuw and the French School of Data Analysis", *Journal of statistical software*, vol. 73(6), pp. xx, 2016
- B. Zaffora, M. Magistris, **G. Saporta**, F. La Torre. "Statistical sampling applied to the radiological characterization of historical waste", *EPJ Nuclear Sci. Technol.* , vol. 2, pp. 11, 2016
- M. Hocine, K. Ait-Bouziad , P. Legeron, W. Dab, **G. Saporta**. "How to Identify and Prioritize Psychosocial Factors Impacting Stress Level", *PLoS ONE*, vol. 11(6), pp. 1-11, 2016
- B. Zaffora, M. Magistris, J. Chevalier, C. Luccioni, **G. Saporta**, L. Ulrici. "A new approach to characterize very-low-level radioactive waste produced at hadron accelerators", *Applied Radiations and Isotopes*, vol. 122, pp. 141-147, 2017, (doi: <http://dx.doi.org/10.1016/j.apradiso.2017.01.019>)
- J. Geronimi, **G. Saporta**. "Variable selection for multiply-imputed data with penalized generalized estimating equations", *Computational Statistics & Data Analysis*, vol. 110, pp. 103-114, 2017, (doi:<http://dx.doi.org/10.1016/j.csda.2017.01.001>)
- C. Liberati , F. Camillo , **G. Saporta**. Advances in credit scoring: combining performance and interpretation in kernel discriminant analysis, *Advances in Data Analysis and Classification*, Vol. 11(1), pp 121-138, 2017
- B. Zaffora, M. Magistris, **G. Saporta**, J. Chevalier. "Uncertainty quantification applied to the radiological characterization of radioactive waste", *Applied Radiations and Isotopes*, vol. 127, pp. 142-149, 2017
- G. Saporta. "Quelle statistique pour le Big Data?", *Statistique et Société*, vol. 5(1), pp. 31-36, 2017

Conference proceedings :

- D. Costanzo, C. Preda et **G. Saporta** - Anticipated prediction in discriminant analysis on functional data for binary response . In COMPSTAT2006, pp. 821-828, Physica-Verlag, 2006.
- G. Saporta - Model Selection and Predictive Inference, in Trends and Challenges in Applied Mathematics, Bucarest, pp.92-96, 2007
- G. Saporta**. Models for Understanding versus Models for Prediction . In COMPSTAT 2008, Porto, 24-29 aout, pp. 315-322, Physica Verlag, 2008.
- M. Charrad, Y. Lechevallier , M. Ben Ahmed , **G. Saporta** - Block clustering for web pages categorization, Intelligent Data Engineering and Automated Learning - IDEAL 2009, Vol. 5788, pp.260-267, Series Lecture Notes in Computer Science, 2009
- C. Preda , **G. Saporta** , B. Hadj Mbarek - PLS Regression with Functional Predictor and Missing Data, PLS'09, 6th Int. Conf. on Partial Least Squares and Related Methods, Pékin, pp.17-22, 2009
- S. Wang , J. Wang , H. Wang , **G. Saporta**- An Approach for PLS Regression Modeling of Functional Data, PLS'09, 6th Int. Conf. on Partial Least Squares and Related Methods, Pékin, pp.28-33, 2009
- N. Niang Keita, F. Fogliatto, **G. Saporta** - Batch Process Monitoring by Three-way Data Analysis Approach, ASDMA'09, Vilnius, pp.294-298, 2009

- M. Charrad, M. Ben Ahmed , Y. Lechevallier , **G. Saporta** - Determining the number of clusters in CROK12 algorithm, MSDM'09 Meeting on Statistics and Data Mining, pp.143-148, Hammamet, 2009
- C. Preda , **G. Saporta** , M. Hadj Mbarek - Anticipated and Adaptive Prediction in Functional Discriminant Analysis, COMPSTAT'2010, 19th International Conference on Computational Statistics, Paris, August 2010, pp.189-198,
- F. de A. T. de Carvalho , **G. Saporta**, D. N. Queiroz - A Clusterwise Center and Range Regression Model for Interval-Valued Data, COMPSTAT'2010, 19th International Conference on Computational Statistics, Paris, August 2010, pp.461-468, Paris, France,
- S. Demeyer, N. Fischer, **G. Saporta** - Contributions to Bayesian Structural Equation Modeling, COMPSTAT'2010, 19th International Conference on Computational Statistics, Paris, August 2010, pp.469-476, Paris, France,
- S. Benammou , B Souissi , **G. Saporta** - On the use of Weighted Regression in Conjoint Analysis, COMPSTAT'2010, 19th International Conference on Computational Statistics, Paris, August 2010, pp.777-784,
- M. Charrad, Y. Lechevallier, **G. Saporta**, M. Ben Ahmed - Détermination du nombre de classes dans les méthodes de bipartitionnement, 17ème Rencontres de la Société Francophone de Classification, June 2010 , June 2010, pp.119-122, Saint-Denis de la Réunion, La Réunion,
- M. Charrad, Y. Lechevallier, M. Ben Ahmed , **G. Saporta**- On the Number of Clusters in Block Clustering Algorithms, 23rd International FLAIRS Conference, May 2010, pp.xx, Series AAAI, Florida, USA,
- M. Charrad, Y. Lechevallier, M. Ben Ahmed , **G. Saporta** - WCUM pour l'analyse d'un site Web, 10ème conférence internationale sur l'Extraction et la Gestion des Connaissances EGC'2010 , January 2010, pp.43-50, Hammamet, Tunisia,
- J. Séguéla, **G. Saporta**- A Semi-Supervised Hybrid System to Enhance the Recommendation of Channels in terms of Campaign ROI, CIKM'11, 20th ACM Conference on Information and Knowledge Management, October 2011, pp.2265-2268, Glasgow, UK,
- A. Bernard, C. Guinot , **G. Saporta** - Sparse principal component analysis for multiblock data and its extension to sparse multiple correspondence analysis, Compstat 2012, August 2012, pp.99-106, Limassol, Chypre,
- C. Gomes , H. Nocairi , M. Thomas , F. Ibanez , J. Collin , **G. Saporta** - Stacking prediction for a binary outcome, Compstat 2012, August 2012, pp.271-282, Limassol, Chypre,
- C. Liberati , F. Camillo , **G. Saporta** - Kernel discrimination and explicative features: an operative approach, Compstat 2012, August 2012, pp.507-518, Limassol, Chypre,
- N. Niang Keita, **G. Saporta**, F. Fogliatto - Contrôle non paramétrique de procédés par lots basé sur STATIS et la classification , Agrostat2012, 12èmes journées Européennes Agro-Industrie et Méthodes Statistiques 28, 29 Février & 1-2 , March 2012, pp.163-169, Paris
- G. Saporta** , H. Hassani - Data Mining, Machine Learning and Official Statistics, Conference of European Statistics Stakeholders, November 24-25, 2014, November 2014, Rome , Italie,
- D. Beaton , A. Bernard, H. Abdi, **G. Saporta**- An integration of partial least squares and sparsified multiple correspondence analysis applied to genetic data, PLS2014, May 2014, Paris, France,
- M. Hocine , G. Russolillo, **G. Saporta**- A PLS Path Model for Predicting Impact of Job Characteristics on Work-Related Stress, PLS2014, May 2014, Paris, France,
- C. Preda , **G. Saporta** - PLS regression for multivariate functional data, PLS2014, May 2014, Paris, France,
- G. Giordano, **G. Saporta**, M. Vitale. "Exploratory data analysis and contiguity relations: An outlook", CFE-CMStatistics 2015, December 2015, pp.96, Londres, Grande Bretagne,
- N. Niang Keita, S. Bougeard, **G. Saporta**. "Clusterwise multiblock PLS regression", CFE-CMStatistics 2015, December 2015, pp.195, Londres, Grande Bretagne,
- N. Niang Keita, S. Bougeard, **G. Saporta**, H. Abdi. "Clusterwise multiblock PLS", CARME 2015, September 2015, pp.58, Naples, Italie,
- H. Abdi, D. Beaton, **G. Saporta**. "Generalizing partial least squares and correspondence analysis methods to predict categorical and heterogeneous data", CARME 2015, September 2015, pp.12, Naples, Italie,
- H. Nocairi , **G. Saporta**. "Stacking Predictions for Multiclass Outcomes", 60th ISI World Statistics Congress, July 2015, pp.xx, Rio de Janeiro, Brésil,
- M. Hocine, K. Ait-Bouziad , **G. Saporta**. "A methodological tool for work-related stress management: Importance-Performance Analysis used complementary to PLS path modelling", ASMDA 2015, July 2015, pp.xx, Le Pirée, Grèce,

J. Geronimi, **G. Saporta**. "The effect of missing visits on GEE, a simulation study", ASMDA 2015, July 2015, pp.269-276, Piraeus, Greece,

B. Zaffora, M. Magistris, **G. Saporta**. "A New Approach to Estimate Uncertainty in Waste Characterization", Advances in Nuclear Nonproliferation Technology and Policy Conference, September 2016, pp.pp, Santa Fe, USA,

J. Geronimi, **G. Saporta**. "Variable selection for longitudinal biomarkers constrained by a detection limit", Compstat 2016, August 2016, pp.23-24, Oviedo, Spain,

S. Bougeard, N. Niang Keita, **G. Saporta**. "Regularized clusterwise multiblock regression", Compstat 2016, August 2016, pp.14, Oviedo, Spain,

J. Geronimi, **G. Saporta**. "Variable selection for multiply-imputed data with penalized generalized estimation equations", XXVIIIth International Biometric Conference, July 2016, pp.xx, Victoria, Canada,

S. Bougeard, N. Niang Keita, **G. Saporta**. "Prediction for regularized clusterwise multiblock regression", ASMDA 2017, June 2017, pp.42-43, Londres, UK,

M. Hocine, N. Feropontova, N. Niang Keita, K. Ait-Bouziad, **G. Saporta**. "Importance of factors contributing to work-related stress: comparison of four metrics", ASMDA 2017, June 2017, pp.97-98, Londres, UK,

Book chapters :

G.Saporta : Les classes latentes in J.J.Droesbeke,M.Lejeune,G.Saporta ,éditeurs. « Modèles statistiques pour données qualitatives », Technip, 2005. pp. 71-82

G.Saporta, N.Niang : Correspondence analysis and classification in J.Blasius & M.Greenacre (editors) "Multiple Correspondence Analysis and Related Methods", Chapman & Hall. pp. 371-392, 2006

G.Saporta: Epidémiologie et Data Mining. In « L'épidémiologie humaine », Rapport sur la Science et la Technologie n°23, sous la direction d'Alain-Jacques Valleron, Académie des Sciences, 2006 pp. 129-136

C. Preda, **G.Saporta**: PCR and PLS for Clusterwise Regression on Functional Data, in "Selected Contributions in Data Analysis and Classification", P.Brito & al. editors. Studies in Classification, Data Analysis, pp. 589-598, Springer, 2007.

G. Saporta, N. Niang Keita- Principal Component Analysis: application to Statistical Process Control, in: "Data Analysis", ISTE-Wiley, pp. 1-23, 2009

V. Stan, **G. Saporta** - Conjoint Use of Variables Clustering and PLS Structural Equations Modeling, in: "Handbook of Partial Least Squares", Springer, pp. 235-246, 2010

S. Demeyer, J. Foulley, N. Fischer, **G. Saporta** - Bayesian analysis of structural equation models using parameter expansion, in: "Statistical learning and data science", Chapman & Hall/CRC, pp. 135-145, 2012

L. Lebart, **G. Saporta** - Historical Elements of Correspondence Analysis and Multiple Correspondence Analysis, Titre du livre: "Visualization and Verbalization of Data", Chapman and Hall/CRC, pp. 31-44, 2014

G. Saporta. Actualité de l'analyse des données. in "Alfredo Rizzi, l'uomo e l'opera", RCEMultimedia, Naples, pp. 43-52, 2015

G. Saporta. "Des méthodes classiques d'analyse des données au Big Data", in "Le big data pour les compagnies d'assurance", September 2017, ISTE Editions, pp. 41-55, (isbn: 978-1-78405-298-0)

Invited and keynote lectures:

G. Saporta. Some Statistical Aspects of Credit Scoring . Plenary invited lecture, *3rd world conference on Computational Statistics & Data Analysis*, Limassol, Chypre, 28-31 octobre, 2005

G. Saporta. Credit scoring, statistique et apprentissage, conférence plénière. *EGC06*, Lille, 18-20 janvier, 2006.

G. Saporta. Model Selection and Predictive Inference . In *Trends and Challenges in Applied Mathematics*, Bucarest, 20-23 juin, 2007

G. Giordano et **G. Saporta**. The Response Surface as Explorative Tool in Multiple Correspondence Analysis . In *CARME 07, Correspondence Analysis and Related Methods*, Rotterdam, 25-27 juin, 2007

G. Saporta. Discriminant analysis on functional data . *XV Congresso Annual da Sociedade Portuguesa de Estadística*, Lisboa, 19-21 août, 2007

G. Saporta, C. Preda - Adaptive Forecasting on Functional Data, *SIS'08*, Univ. Calabria, January 2008

- G. Saporta - Models for Understanding versus Models for Prediction, *COMPSTAT'08*, Porto, Portugal, January 2008, pp.315-322
- G. Saporta - Supervised and Unsupervised Linear Methods for Functional Data Analysis, *ASMDA'09*, Vilnius, July 2009, Vilnius, Lituanie,
- G. Saporta - Modèles pour comprendre, ou modèles pour prévoir?, *MSDM'09*, Hammamet, Tunisie, March 2009, Hammamet, Tunisie
- G. Saporta - Classification supervisée et données fonctionnelles, *SFC'2010*, 17èmes rencontres de la Société Francophone de Classification, Saint Denis de la Réunion, Juin 2010
- G. Saporta - A survey of some sparse methods for high-dimensional data, *SADA'13*, Cotonou, Benin Republic, March 2013
- G. Saporta - From Sparse Regression to Sparse Multiple Correspondence Analysis, *European Conference on Data Analysis*, July 2013, Luxembourg
- A. Bernard, **G. Saporta** - De l'ACP sparse à l'ACM sparse, *Agrostat 2014*, Rabat, Maroc, 2014
- G. Saporta - Quelle statistique pour les Big Data? , *Symposium "Risques et modèles en épidémiologie humaine, animale et végétale"* 19 et 20 mai 2015, Université des Antilles, May 2015, Pointe à Pitre
- G. Saporta - Conjoint Analysis: past, present and issues, *Choice and preference analysis for quality improvement and seminar on experimentation*, July 2015, Bari, Italie,
- G. Saporta**, S. Bougeard , N. Niang Keita - Les méthodes « clusterwise » : principes et applications, *XXII èmes rencontres de la Société Francophone de Classification*, September 2015, Nantes, France,
- G. Saporta. "Some Sparse Methods for High Dimensional Data", *H2DM'16*, June 2016, pp.xx, Naples, Italie,
- G. Saporta. "Predictive versus Generative Modelling: a Challenge for (Social) Sciences", *DSSR*, February 2016, pp.xx, Naples, Italie,
- G. Saporta. "50 years of data analysis: from EDA to predictive modelling and machine learning ", *ASMDA 2017*, June 2017, Londres,