

French citizen

Birth date: October 15th 1970

Current position: Professor in Applied Statistics, Agrocampus, Rennes, France.

Head of the Master of Biological Data Science, Agrocampus, Rennes, France.

Head of the Department of Statistics and Computer Science, Agrocampus, Rennes, France.

Current President of the French region of the International Biometric Society.

Contact

Applied Mathematics Department
Agrocampus Ouest
CS 84215 - 65 rue de Saint-Brieuc
35042 Rennes cedex

Phone: +33.223.485.884

Email: david.causeur@agrocampus-ouest.fr

<http://math.agrocampus-ouest.fr/infoglueDeliverLive/membres/david.causeur>

Education

2005 **Research habilitation** in Mathematics, University of South Brittany (Vannes, France).

Title : Gauss-Markov type estimation in linear model with incomplete data & Spatial interaction between homologous factors

Committee

Michel Carbon, Michel Delecroix, Thierry Dhorne, Ion Grama

Reviewers: Jean-Marc Azaïs (University Paul Sabatier, Toulouse), Subir Ghosh, (University of California, Riverside) and Christian Robert (University Paris IX Dauphine)

1997 **Ph.D.** in Mathematics and Applications, University Rennes I (Rennes, France).

Title : Statistical inference in linear model with incomplete data.

Committee

Michel Bonneau, Jean-Baptiste Denis, Jean Deshayes

Supervisor: Thierry Dhorne

Reviewers: Jean-René Mathieu (University Paul Sabatier, Toulouse), Jean-Jacques Téchené (University of Pau)

1993 **Master of Science** in Pure Mathematics and Applications, University Rennes I

Positions

2007- **Professor** in Applied Statistics at Agrocampus, Rennes, France.

2005- **Research fellow** at IRMAR (National Center for Scientific Research), Rennes, France.

2002-2005 **Research fellow** at the Statistics and Modeling Department (Research Center for Economics and Statistics), Rennes, France.

1999-2001 **Assistant professor** in Statistics at University of South Brittany, Vannes, France.

1998-2001 **Research fellow** at the Applied Statistics Department of University of South Brittany, Vannes, France.

1998-2007 **Assistant professor** in Applied Mathematics at Agrocampus, Rennes, France.

1996-1997 **Teaching and Research Assistant** in Applied Mathematics at Agrocampus, Rennes, France.

Professional activities

2021-	Director of the Department of Statistics and Computer Science at Agrocampus, Rennes, France.
2021	Member of the program and organization committees of the <i>Journées de Biostatistique 2021</i> conference, Fully online.
2021	Member of the program and organization committees of the Channel Network Conference, IBS, Fully online.
2020-2022	President of the French region of the International Biometric Society.
2020	Jury of the Hans Van Houwelingen Award by the Dutch Biometric Society, Amsterdam, NL.
2019	Member of the program committee of the Channel Network Conference, IBS, Rothamsted Research Center, UK.
2017	Member of the program committee of the Channel Network Conference, IBS, Hasselt, Belgium.
2017	Member of the program committee of SMPGD (Statistical Methods for Post-Genomic Data), London, UK.
2016	Member of the program committee of JOBIM (Journées ouvertes en Biologie, Informatique et Mathématique), Lyon, France.
2015	Member of the program committee of the Channel Network Conference, IBS, Nijmegen, NL.
2012-2016	Director of the Department of Statistics and Computer Science at Agrocampus, Rennes, France.
2010-2012	Co-head of the Undergraduate program in Biology and Mathematics, University Paris VI, Paris, France.
2011	President of the organizing committee of the 8th Rennes Statistical Meeting, Rennes, France
2008	President of the organizing committee of the 6th workshop on Statistical Methods for Post-Genomic Data, Rennes, France.
2006-2008	Elected member of the scientific council of University Rennes II.
Since 2005	Elected member of the board of directors at Agrocampus, Rennes, France.
Since 2005	Head of the Master of Biological Data Science, Agrocampus, Rennes, France.
2006-2008	Elected member of the scientific council of Agrocampus, Rennes, France.
2005-2010	Co-organizer of the weekly statistics seminar at IRMAR (National Center for Scientific Research), Rennes, France.
2005	President of the organizing committee of the 2nd Rennes Statistical Meeting, Rennes, France

Refereeing:

Theoretical and Applied Statistics journals: Annals of Applied Statistics, Biometrical Journal, Biostatistics, Canadian Journal of Statistics, Communications in Statistics (Theory and Methods), Electronic Journal of Applied Statistical Analysis, ESAIM: Probability and Statistics, Journal of the American Statistical Association, Journal of the French Statistical Society, Journal of Statistical Planning and Inference, Metron, Scandinavian Journal of Statistics, Statistical Applications in Genetics and Molecular Biology, Statistics in Medicine, Wiley (books).

Bioinformatics and Computational Statistics: Bioinformatics, BMC Bioinformatics, Computational Statistics, Computational Statistics and data Analysis, IEE/ACM Transactions on Computational Biology and Bioinformatics, Journal of Computational and Graphical Statistics, Journal of Statistical Computation and Simulation.

Statistical Applications: Behavior Research Methods, International Journal of Statistics and Management Systems, Journal of Neuroscience Methods, Metabolites.

Teaching experience

1st year undergraduate program Biology & Mathematics

Differential equations and functional analysis.

1st year Master of Science

Data analysis, Elementary inferential statistics, Linear modeling, Generalized linear modeling.

2nd year Master of Science

Compartment modeling, Functional data analysis, High-dimensional data analysis, Nonlinear modeling, Spatial statistics, Statistical genetics, Statistical learning.

Research interests

Inference for high-dimensional data

Post-genomic data, ERP (EEG) data, Large-scale simultaneous inference, Factor analysis, Stability selection.

Incomplete observations

Estimation of variance components, Optimal sampling designs.

Spatial interaction between homologous factors

Generalized biadditive models, Bradley-Terry model, Random matrix theory, Dimensionality.

PhD director:

Defence	Topic
2022	Dynamic modeling of in vitro pepsinolysis (O. Suwareh).
2019	Block-level modeling of dependence in GWAS (F. Hébert).
2019	Statistical methodology for food product usage (M. Brard).
2017	Gauss-Markov estimation in survey sampling with non-response (B. Gelein).
2016	Assessment of statistical learning methods in genomic data analysis (R. Hornung).
2015	Stability of model selection in high-dimensional supervised classification (E. Perthame).
2012	Modeling gene regulatory networks using transcriptomic data - Application to eQTL detection. (Y. Blum).
2009	Linear gaussian state-space model with irregular sampling - Application to Sea Surface Temperature (P. Tandeo).
2009	Impact of dependence in large-scale multiple testing (C. Friguet).

Selected PhD committee memberships since 2012:

Defence	Topic
2021	Hierarchical clustering of Hi-C data under contiguity restrictions (N. Randriamihamison, PhD, Examiner).
2020	Latent variable Markov models with applications to pharmacology and epidemiology (M. Delattre, HdR, Examiner).
2020	Some statistical and computational contributions to genomic data analysis (G. Rigai, HdR, Reviewer).
2019	Regression models for heterogeneous functional data (M. Morvan, PhD, President).
2019	Regularized methods for high-dimensional data analysis: theory and applications (M. Perrot-Dockès, PhD, Reviewer).
2018	Dynamic of milk Ca content as an indicator of management system and diets (P. Gagnon, PhD, President).
2018	Classification and Regression Energy Tree for Functional Data (M. Brandi, PhD, Reviewer, University La Sapienza, Rome, Italy).
2018	Multifactorial analysis of crop yield (T. Galinier, PhD, Reviewer).
2017	Rare variants association studies in demographically stable populations (E. Persyn, PhD, Reviewer, President).
2017	Statistical methods and software for the analysis of transcriptomic data (A. Rau, Habilitation, Reviewer).
2016	Phenotypic effects of two mechanisms of activation of Wnt/ β -catenin signaling in hepatocellular carcinoma (R. Désert, PhD, Examiner).
2016	Contributions to categorical data analysis, data modeling and statistical inference in biostatistics (M. Emily, Habilitation, Examiner).
2016	Penalized learning for joint modeling of gene expression data and networks (T. Ha, PhD, President).
2015	Effects of partial asexuality on the dynamics of genotype frequencies in dominantly diploid populations (K. Reichel, PhD, President).
2015	Incorporating the linkage disequilibrium (LD) information in Genome-Wide association studies. (A. Dehman, PhD, Reviewer).
2015	Geometrical approach of genetical data (D. Laloe, Habilitation, Reviewer).
2015	Statistics for environmental, ecological, medical and archeological data (Lise Bellanger, Habilitation, President).
2014	Nonparametric modeling of electricity consumption (Leslie Hatton, PhD, President).
2014	Vietnamese food choice model: an approach from food products and consumers. (Nguyen Ba Thanh, PhD, Reviewer).
2014	Optimization of association genetics and genomic selection strategies for populations of different diversity levels. (R. Rincet, PhD, Examiner).
2014	Dealing with uncertainties in a complex marine ecosystem model. (L. Gasche, PhD, President).
2013	Accounting for the growing pigs' variability in models of nutritional needs. (B. Vautier, PhD, President).
2013	Statistics for genomic data analysis (M.-L. Martin-Magniette, Habilitation, Reviewer).
2012	Statistical methods for robust analysis of transcriptome data by integration of biological knowledge. (M. Jeanmougin, PhD, Reviewer).
2012	Phenotypic prediction and variable selection in high dimensional linear and linear mixed models. (F. Rohart, PhD, Reviewer).

Articles in refereed journals:

1. Hébert, F., **Causeur, D.** and Emily, M. (2021) Adaptive handling of dependence in high-dimensional regression modeling. *Journal of Computational and Graphical Statistics*. DOI: 10.1080/10618600.2022.2076687
2. Hébert, F., **Causeur, D.** and Emily, M. (2021) Global testing approach for gene-based gene-gene interaction. *Statistics in Medicine*. doi: 10.1002/sim.9389.
3. Prado, E., Eklouh-Molinier, C., Enez, F., Causeur, D., Blay, C., Dupont-Nivet, M., Labbé, L., Petit, V., Moreac, A., Taupier, G., Haffray, P., Bugeon, J., Corraze, G. and Nazabal, V. (2022) Prediction of fatty acids composition in the rainbow trout *Oncorhynchus mykiss* by using Raman micro-spectroscopy. ***Analytica Chimica Acta***. Volume 1191, <https://doi.org/10.1016/j.aca.2021.339212>.
4. Blay, C., Haffray P., D'Ambrosio, J., Prado, E., Dechamp, N., Nazabal, V., Bugeon, J., Enez, F., **Causeur, D.**, Eklouh-Molinier, C., Petit, V., Phocas, F., Corraze, G. and Dupont-Nivet, M. (2021) Genetic Architecture and Genomic Selection of Fatty Acid Composition Predicted by Raman Spectroscopy in Rainbow Trout. *BMC Genomics*. 22, 788. <https://doi.org/10.1186/s12864-021-08062-7>.
5. Comfort, W., Nunes de Andrade, B., Wingenbach, T.S.H, **Causeur, D.**, Boggio, P.S. (2021) Implicit Responses in the Judgement of Attractiveness in Faces with Differing Levels of Makeup. *Psychology of Aesthetics, Creativity and the Arts*. Advance online publication. <https://doi.org/10.1037/aca0000408>
6. Suwareh, O., **Causeur, D.**, Jardin, J., Briard-Bion, V., Le Feunteun, S., Pézenec, S., Nau, F. (2021) Statistical modeling of in vitro pepsin specificity. *Food Chemistry*. 130098, Volume 362. <https://doi.org/10.1016/j.foodchem.2021.130098>.
7. Hébert, F., **Causeur, D.** and Emily, M. (2021). An adaptive decorrelation procedure for signal detection. *Computational Statistics and Data Analysis*. 153. <https://doi.org/10.1016/j.csda.2020.107082>.
8. **Causeur, D.**, Sheu, C.-F., Perthame, E. and Rufini, F. (2019). A functional generalized F-test for signal detection with applications to event-related potentials significance analysis. *Biometrics*. 2019; 1–11. <https://doi.org/10.1111/biom.13118>
9. Aïchi, H., Fouad, Y., **Causeur, D.**, and Walter, C. (2020) Organic Carbon and Total Iron Effect on Soil Vis-SW NIR Spectra and Quantification of Their Contents Using PLS R Models. *Communications in Soil Science and Plant Analysis*, 51:9, 1253-1267, DOI: 10.1080/00103624.2020.1751187.
10. Désert, C., Baeza, E., Aite, M., Boutin, M., Le Cam, A., Montfort, J., Houée-Bigot, M., Blum, Y., Roux, P.-F., Hennequet-Antier, C., Berri, C., Metayer-Coustard, S., Collin, A., Allais, S., Lebihan, E., **Causeur, D.**, Gondret, F., Duclos, M., Lagarrigue, S. (2018). Multi-tissue transcriptomic study reveals the main role of liver in the chicken adaptive response to a switch in dietary energy source through the transcriptional regulation of lipogenesis. *BMC Genomics*. 19(1):187.
11. Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., **Causeur, D.**, Gilbert, H. and Louveau, I. (2017). A transcriptome multi-tissue analysis identifies biological pathways and genes associated with variations in feed efficiency of growing pigs. *BMC Genomics*. 18:244.
12. Hornung, R., **Causeur, D.**, Bernau, C. and Boulesteix, A.-L. (2017). Improving cross-study prediction through add-on batch effect adjustment and add-on normalization. *Bioinformatics*. doi: 10.1093/bioinformatics/btw650.
13. Blum, Y., Houée-Bigot, M., **Causeur, D.** (2016). Sparse factor model for co-expression networks with an application using prior biological knowledge. *Statistical Applications in Genetics and Molecular Biology*. 15(3), 253-272.
14. Hornung, R., Boulesteix, A.-L. and **Causeur, D.** (2016). Combining location-and-scale batch effect adjustment with data cleaning by latent factor adjustment. *BMC Bioinformatics*, 17(1), 1-19.
15. Gondret, F., Vincent, A., Houée-Bigot, M., Siegel, A., Lagarrigue, S., Louveau, I. and **Causeur, D.** (2016). Molecular alterations induced by a high-fat high-fiber diet in porcine adipose tissues: variations according to the anatomical fat location. *BMC Genomics*, 17(120).
16. Nyemb-Diop, K., **Causeur, D.**, Jardin, J., Briard-Bion, V., Guérin-Dubiard, C., Rutherford, S.-M., Dupont, D. and Nau, F. (2016) Investigating the impact of egg white gel structure on peptide kinetics profile during in vitro digestion. *Food Research International*. 88(B). 302-309.

17. Sheu, C.-F., Perthame, E., Lee, Y.-S., **Causeur, D.** (2016). Accounting for time dependence in large-scale multiple testing of event-related potential data. *Annals of Applied Statistics*. 10(1), 219-245.
18. Perthame, E., Friguet, C. and **Causeur, D.** (2015) Stability of feature selection in classification issues for high-dimensional correlated data. *Statistics and computing*. Doi 10.1007/s11222-015-9569-2. pp. 1-14.
19. Gelein, B., Haziza, D. and **Causeur, D.** (2014) Preserving relationships between variables with MIVQUE based imputation for missing survey data. *Journal of Multivariate Analysis*. **131**, 197-208
20. Nyemb, K., Jardin, J., **Causeur, D.**, Guérin-Dubiard, C., Dupont, D., Rutherford, S.M. and Nau, F. (2014) Investigating the impact of ovalbumin aggregate morphology on in vitro ovalbumin digestion using label-free quantitative peptidomics and multivariate data analysis. *Food Research International*. 63(B), 192–202.
21. **Causeur, D.**, Chu, M.-C., Hsieh, S. and Sheu, C.-F. (2012) A factor-adjusted multiple testing procedure for ERP data analysis. *Behavior Research Methods*. **44**, 635–643.
22. Mach, N., Blum, Y., Bannink, A., **Causeur, D.**, Houée-Bigot, M., Lagarrigue, S. and Smits, M.-A. (2012) Pleiotropic effects of polymorphism of the gene diacylglycerol-O-transferase 1 (DGAT1) in the mammary gland tissue of dairy cows. *Journal of Dairy Science*. **95** (9), 4989–5000.
23. Blum, Y., Le Mignon, G., **Causeur, D.**, Filangi, O. Désert, C. Demeure, O. Le Roy, P. and Lagarrigue, S. (2011) Complex trait subtypes identification using transcriptome profiling reveals an interaction between two QTL affecting adiposity in chicken. *BMC Genomics*. 12:567.
24. **Causeur, D.**, Friguet, C., Houée-Bigot, M. and Kloareg, M. (2011). Factor Analysis for Multiple Testing (FAMT): An R package for Large-Scale Significance Testing Under Dependence. *Journal of Statistical Software*. **40** (14), 1–19.
25. Friguet, C. and **Causeur, D.** (2011). Estimation of the proportion of true null hypotheses in high-dimensional data under dependence. *Computational Statistics and Data Analysis*, **55** (9), 2665–2676.
26. Caffier V., Didelot F., Pumo B., **Causeur D.** and Parisi L. (2010). Aggressiveness on a susceptible apple cultivar of eight *Venturia inaequalis* isolates virulent or avirulent towards the major resistance gene Vf. *Plant Pathology*. **59**(6). 1072–1080.
27. Blum, Y., Le Mignon, G., Lagarrigue, S. and **Causeur, D.** (2010). A factor model to analyse heterogeneity in gene expressions. *BMC Bioinformatics*. 11:368.
28. Friguet, C., Kloareg, M. and **Causeur, D.** (2009). A factor model approach to multiple testing under dependence. *Journal of the American Statistical Association*. **104** (488) 1406–1415.
29. **Causeur, D.**, Kloareg, M. and Friguet, C. (2009). Control of the FWER in multiple testing under dependence. *Communications in Statistics - Theory and Methods*. 38 (16 & 17) 2733–2747.
30. Ablain, W., Hallier Soulier, S., **Causeur, D.**, Gautier, M. & Baron, F. (2009) A simple and rapid method for the disruption of *Staphylococcus aureus*, optimized for quantitative reverse transcriptase applications: Application for the examination of Camembert cheese. *Dairy Science and Technology*. 89 (1) p 69.
31. Cutullic, E., Delaby, L., **Causeur, D.**, Michel, G. & Disenhaus, C. (2009) Hierarchy of factors affecting behavioural signs used for oestrus detection of Holstein and Normande dairy cows in a seasonal calving system. *Animal Reproduction Science*. 113 (1-4), 22–37
32. Kloareg, M. & **Causeur, D.** (2009). Double-sampling designs to reduce the non-discovery rate. Application to microarray data. *Journal of Data Science*. 7 (2), 219–234
33. **Causeur, D.** and Husson, F. (2008). Asymptotic power of double-sampling tests for general linear hypothesis. *Statistics*. 42, 2. 115-125.
34. Kloareg, M. and **Causeur, D.** (2007). Improving Type-II error rates of multiple testing procedures by use of auxiliary variables. application to microarray data. in Recent advances in Stochastic Modelling and data Analysis, Ed. Christos Skiadas, World Scientific Publishing, Co Pte Ltd.
35. **Causeur, D.** (2006) MIVQUE and maximum likelihood estimation for multivariate linear models with incomplete observations. *Sankhyā*. **68** (3), 408-434.
36. **Causeur, D.** (2005). Optimal sampling from concomitant variables for regression problems. *Journal of Statistical planning and Inference*. **128**, 289-301.
37. **Causeur, D.**, Dhorne, T. & Antoni, A. (2005). A two-way analysis of variance model with positive definite interaction for homologous factors. *Journal of Multivariate Analysis*. **95** (2), 431-448.
38. **Causeur, D.** & Husson, F. (2005). Causeur, D. & Husson, F. (2005). A 2-dimensional extension of the Bradley-Terry model for paired comparisons. *Journal of Statistical planning and Inference* **135**, 245-259.

39. Ledauphin, S., Vigneau, E. & **Causeur, D.** (2005). Functional approach for the analysis of Time Intensity curves using B-splines. *Journal of Sensory Studies*. **20** (4), 285-300.
40. **Causeur, D.** & Dhorne, T. (2003). Linear Regression Models under Conditional Independence Restrictions. *Scandinavian Journal of Statistics* **30** (3), 637-650.
41. Kamp, J.F., Marin-Mamellet, C., Forzy, J.F. and **Causeur, D.** (2001). HMI Aspects of the Usability of Internet Services with an In-car Terminal on a Driving Simulator. *IATSS Research*. **25**, 2, 29-40.
42. **Causeur, D.** (1999). Exact distribution of the regression estimator in double-sampling. *Statistics* **32**, 297-315.
43. **Causeur, D.** & Dhorne, T. (1998). Finite-sample properties of a multivariate extension of double-regression. *Biometrics* **54** (4), 1591-1601.
44. **Causeur, D.** (1998). Plan d'échantillonnage en plusieurs phases pour la réduction des coûts expérimentaux en régression linéaire. *Revue de Statistique Appliquée*. **XLVI** (4), 59-73.

Software development:

1. Hornung, R and Causeur, D. (2015). bapred: Batch Effect Removal (in Phenotype Prediction using Gene Data). R package version 0.2, <http://CRAN.R-project.org/package=bapred>
2. Causeur, D. and Sheu C.-F. (2014). ERP: Significance analysis of Event-Related Potentials data. R package version 1.0.1, <http://CRAN.R-project.org/package=erp>
3. Perthame, E., Friguet, C. and Causeur, D. (2014). FADA: Variable selection for supervised classification in high dimension. R package version 1.2., <http://CRAN.R-project.org/package=fada>
4. Causeur, D., Friguet, C., Houée-Bigot, M. and Kloareg, M. (2013). Factor Analysis for Multiple Testing (FAMT) : simultaneous tests under dependence in high-dimensional data. R package version 2.5., <http://CRAN.R-project.org/package=famt>

Selected talks since 2011:

1. Sparse inverse time correlation model for signal identification in functional Near Infrared Spectroscopy data. April 2021. 8th Channel Network Conference, 2021, Online conference.
2. Adaptive handling of dependence in high-dimensional regression modeling. November 2020. Invited talk at the Dutch Biometric Conference.
3. A functional generalized F-test for signal detection with applications to event-related potentials significance analysis. May 2020, Invited talk at Journées de Statistique, Nice.
4. Functional Analysis of Variance in the Comparison of Curves. November 2019. Invited talk at School of Occupational Therapy, National Taiwan University, Taipei
5. Ignoring dependence in large-dimensional testing issues: what consequences? July 2018. Channel Network Conference, Rothamsted Research Center, UK.
6. Handling dependence or not in statistical learning for high-dimensional data. December 2018, Invited Seminar, Department of Statistics, La Sapienza University, Roma, Italy.
7. Signal identification in ERP data by decorrelated Higher Criticism Thresholding, September 2015, CIBB 2015, Naples, Italy.
8. Large Scale Multiple Comparisons in Ranking Studies. Invited Seminar, November 2013. NCKU, Tainan, Taiwan.
9. Modèles à facteurs parcimonieux pour données à haut débit. Séminaire invité, Mars 2013, Université de Montpellier.
10. Sparse factor models for high-dimensional interaction networks. Invited talk, May 2013, 7th International Workshop on Simulation, Rimini, Italy.
11. Large scale significance testing under dependence. Invited Seminar, November 2012. NCCU, Chia-Yi, Taiwan.
12. Sparse factor model for high-dimensional data. Invited Seminar, November 2012. Academia Sinica, Taipei, Taiwan.
13. Sparse factor models for gene interaction networks. Invited Seminar, March 2012. University of Limerick, Ireland.
14. Sparse factor models for inference on gene regulatory networks. Invited Seminar, December 2011. IBE, University of Munich, Germany.
15. High-throughput data analysis in systems biology. 2nd international workshop in Biostatistics, Invited talk, Bio-SI 2011. Rennes, France.
16. Integrating biological knowledge in gene expression data analysis. Invited session, ISI 2011 (International Statistical Association), Dublin, Ireland.
17. Feature selection stability in high-dimensional heterogeneous data. Invited talk, Cladag 2011. Pavia, Italy.
18. Impact de la dépendance sur les procédures de tests multiples en grande dimension. Séminaire invité, Janvier 2011, Université de Besançon.