Benoit Liquet

Professor at University of Pau et Pays de L'Adour

✉ benoit.liquet@univ-pau.fr
http://benoit-liquet.github.io

Education

2009	**HDR (PhD Supervisor diploma)**, *Bordeaux Segalen Univ.*, Bordeaux.
2002	**PhD Biostatistics**, *Bordeaux Segalen Univ.*, Bordeaux.
1999	**MD Public Health**, *Bordeaux Segalen Univ.*, Bordeaux.
1998	**MD Stochastic modelling** & **Operational research**, *University Bordeaux I*, Bordeaux.

Position

2015 -	**Professor**, *University of Pau et Pays de l'Adour*, (France), Member of LMAP.
2015 -	**Senior Research Fellow**, *ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS)*, Queensland University Of Technology, From August 2015.
2015 -	**Honorary Associate Professor**, *School of Mathematics & Physics*, The University of Queensland,(UQ).
2013 - 2015	**Senior Lecturer**, *School of Mathematics and Physics (SMP)*, The University of Queensland (UQ).
2012 - 2013	**Senior Investigator Statistician**, *Medical Research Council Biostatistics*, Cambridge.
2012 - Sept	**Sabbatical visitor researcher**, *Medical Research Council Biostatistics*, Cambridge.
2011 - 2012	**Sabbatical visitor researcher**, *Queensland Facility for Advanced Bioinformatics*, Brisbane.
2011 - Oct	**Sabbatical visitor researcher**, *Department of Biostatistics*, Copenhagen.
2007 - 2011	**Associate professor**, *Bordeaux Segalen Univ.*, Bordeaux, ISPED.
2004 - 2007	**Associate professor**, *University Pierre Mendes France*, Grenoble.
2003 - 2004	**Lecturer**, *Mathematics Department*, University Montpellier II.
2002 - 2003	**Lecturer**, *Mathematics Department*, University Bordeaux I.
2000 - 2002	**Assistant/Lecturer**, *Bordeaux Segalen Univ.*, Bordeaux, ISPED.

Administrative responsibilities

2018 -	**Head of the team "Probability and Statistics" of the LMAP** (Laboratory of Mathematics and Their Applications) from University of Pau et Pays de l'Adour.
2016 -	**Board Member of the committee of the LMAP**
2016 -	**Board Member of the committee of the CT3** (Commission Thématique Applications Scientifiques) from University of Pau et Pays de l'Adour.
2016 -	**In charge of the dissemination of the new pedagogic tools** (Interaction learning, e-learning, ..,) at Anglet campus from University of Pau et Pays de l'Adour.
2015 - 2018	**Board Member of the academic consultative committee** section 25/26 University of Pau et Pays de l'Adour.
2015 -	**Board Member of MIRA (Milieux et Ressources Aquatiques)** University of Pau et Pays de l'Adour.
2014 - 2015	**Coordinator** of the statistics centre at UQ.

2014 - 2015	**Representative** of SMP for collaboration and student exchange with Ecole Centrale (consortium of Engineering Universities in France).
2013 - 2015	**Treasurer** of the SFB (French Biometrics Society).
2012 -	**Board Member of the R Committee** (French organization for coordinating R conferences).
2010 - 2015	**Board Member of the SFB.**
2010 - 2013	**Board Member of the academic consultative committee** section 26 of Bordeaux University.
2009 - 2012	**Board Member of the Sport Department** - University of Bordeaux 2.
2008 - 2011	**Board Member of the coordination and development committee on distance learning** -ISPED.

Research Management

2019-January	**Designer and co-coordinator** with Kerrie Mengersen of the "2-day Statistical Methods for the Analysis of High-Dimensional and Massive Data set" at Queensland University of Technology. This workshop has attracted more than 200 attendees during the two days.
2018 - June	**Co-designer and co-coordinator** of the first summer school (July 2017) at Anglet campus in collaboration with Imperial College London. This summer school is the follow-up of the one-week workshop delivered in December at Imperial college on "Statistical approaches to characterize the exposome".
2017 - June	**President of the Organization Committee** - "Les sixièmes Rencontres R Anglet 2017" (`http://angletr2017.com/`).
2015 - 2016	**Member of the Scientific Committee**, 31th International Workshop on Statistical Modelling (Rennes).
2014 - 2015	**Co-organisator** of Statistics seminar at SMP.
2012 - July	**President of the Scientific Committee** - "Rencontres R BoRdeaux 2012" (Bordeaux).
2011 - April	**Organization Committee Member**, International Biometrics Society Channel Network (Bordeaux).
2009 - May	**Organization Committee Secretary**, XXXXIth Conference of SFdS (Bordeaux).
2004 - May	**Organization Committee Member**, XXXVIth Conference of SFdS (Montpellier).
2005 -	**Member** of SFDS, of the SFB, and of the IBS (International Biometrics Society).
2004 -	**Reviewer**, for the following journals: Biometrics, Statistics, Statistica Sinica, Journal of Multivariate Analysis, Journal of Statistical Theory and Practice, Biometrical Journal, Revue d'Epidémiologie et de Santé Publique, Journal de la Société Française de Statistique, Bioinformatics, Australian & New Zealand Journal of Statistics.
2012 - 2013	**Expertise for IReSP** (French institute in public health) regarding grant project about Mathematics and Statistics research applied in oncology.

Industrial Collaborations and Grants

2018 - 2018	Contract with Capionis: regarding a research project on : "Machine learning methods to analyse cross-omics data".
2018 - 2018	Contract with HUPI (`http://www.hupi.fr`) regarding the project **KEOLIS**: "Anticipate the maintenance operations to be performed on each bus from Bordeaux".
2016 - 2017	Industrial Grant from a collaboration with "CAPIONIS" Statistical Consulting Company in Bordeaux. Project on : "Health-Economic modeling using Markov model and application in R". (5,000 euros and the salary for one master student training from March 2017 to August 2017').
2016 - 2021	Investigator in "MICROPOLIT" project funded by Aquitaine Region.
2016 - 2019	"ASMAH" (Advanced Statistical Methods for the analysis of High-dimensional data) project granted by joint call from UPPA and UPV/EHU for funding a Phd.
2015 - 2017	Partner investigator in "Mechanomics" project granted by the Cancer Research UK calls
2015 - 2016	"EOLE" project, Funded by UPPA as a BQR ("Bonus Qualité Recherche")

2014 - 2014	Two teaching & Learning grants from School of Mathematics and Physics at UQ.
2014 - 2015	Advanced statistical methods for analysis of multidimensional databases of human brain Imaging from IdEx Bordeaux Centre InterLabex CPU-TRAIL.
2013 - 2013	Grant Visiting Lecturer FMIPA UNS from Ministry Education Indonesia.
2010 - 2013	Public Health ANR project: Dynamic Biostatistics Models for Epidemiology.
2010 - 2013	Grant from Danone for a Phd Studentship.
2011 -	Expert for Danone Research on statistical methodology.
2009 - 2010	Contract with Voies Navigables de France: "modelling of waiting time at locks on Canal du Midi".
2009 - 2010	Courses on multiple comparison procedure for statisticians from Danone Research.
2006 - 2007	Statistical Consultant to BioArtificial Gel Technologies (Skin and pharmaceutical company in Montréal).
2003 - 2004	Statistician in the governmental project (Ministry of Food, Agriculture, Fishing and rural affairs).

Student Supervision

2019 -	Supervision of a Phd candidate (Charnel Tabet) at University of Pau et Pays de L'Adour
2019 -	Supervision of a Phd candidate (Alexandre Lefranc) at University of Pau et Pays de L'Adour
2018 -	Supervision of a Phd candidate (Aurelien Callens) at University of Pau et Pays de L'Adour
2018 -	Supervision of a Phd candidate (Sebastien Coube) at University of Pau et Pays de L'Adour
2018 -	Co-Supervision of a Phd candidate (Floren Hugon) at University of Pau et Pays de L'Adour
2017 -	Co-Supervision of a Phd candidate (Marc Antoine Poncet) at University of Pau et Pays de L'Adour
2017 -	Co-Supervision of a Phd candidate (Sophie Defontaine) at University of Pau et Pays de L'Adour
2016 -	Co-Supervision of a Phd candidate (Camilo Broc) at University of Pau et Pays de L'Adour
2016 -	Co-Supervision of a Phd candidate (Matthew Sutton) at Queensland University Technology.
2014 -	Supervision of a Phd candidate (Yifan Jiang) at the University of Queensland.
2013 - 2017	Co-supervision of a Phd candidate (Azam Asanjarani) at the University of Queensland.
2018 -	Supervision of a Master Student in Statistics (Anglet campus from UPPA).
2017 -	Supervision of a Master Student in Biostatistics (Anglet campus from UPPA).
2017 -	Supervision of a Master Student regarding the MICROPOLIT project (Anglet campus from UPPA).
2016 -	Co-supervision of a Master Student regarding the EOLE project (Anglet campus from UPPA).
2013 -	Supervision Honours and Master students at the University of Queensland.
2010 - 2013	Supervision of a PhD candidate (Jéremie Riou) on the topic of "Multiple testing in clinical studies".
2006 - 2009	Supervision of a PhD candidate (Moliere N'Guile Makao) on the research topic of "Prediction of nosocomial pneumonia using multi- state models".
2007 - 2012	Supervision of first and second year Master of Public Health trainees.
2004 - 2007	Supervision of undergraduate students in Statistics and Computer Science (IUT-STID).

Educational responsibilities

2016 -	Created an **Inter-university cooperation (ACI)** between Sebelas Maret University (Solo in Indonesia)" and (UPPA). Related to this ACI, I visited Sebelas Maret University (December 2016 and February 2017) to give some lectures and to attract students for studying in Pau.
2016 -	Coordinator of "Advanced Data Mining" for the MSID Master at UPPA.
2016 -	Coordinator of all the courses in Mathematics and Statistics at the Ecology Department from Anglet Campus (UPPA).
2014 - 2014	Coordinator of the course "Analysis of Scientific Data" for 900 undergraduate students (UQ).
2013 - 2015	Coordinator of the course "Experimental Design" (UQ).
2007 - 2011	Coordinator of several statistical courses from the Master of Public Health (ISPED).
2009 - 2010	Coordinator of all the statistical courses of the first year of the the Master of Public Health.

2008 - 2010	Coordinator of statistical courses for distance learning (ISPED).
2010 - 2011	Coordinator of the course "R software" during the summer school (ISPED).

Research development

MBSGS	**Maintainer of the R package** on Multivariate Bayesian Sparse Group Selection with Spike and Slab. (with Matthew Sutton)
sgPLS	**Maintainer of the R package** on Sparse Group Partial Least Square regression models. (with Pierre Lafaye de Micheaux)
CEoptim	**Maintainer of the R package** on cross-entropy method for optimization problem (with Tim Benham, Qibin Duan and Dirk P. Kroesse).
mixOmics	**Contributor of the R package** on statistical integrative techniques to analyse highly dimensional data sets (with Sebastien Dejean, Ignacio Gonzalez and Kim-Anh Le Cao).
ClustOfVar	**Developer of the R package** on cluster analysis of a set of variables. Variables can be quantitative, qualitative or a mixture of both (with Marie Chavent, Vanessa Kuentz and Jérome Saracco).
PCAmixdata	**Contributor of the R package** on multivariate analysis of mixed data (with Marie Chavent, Vanessa Kuentz, Amaury Labenne and Jérome Saracco).
lcmm	**Developper of the R package** on estimation of various latent class mixed models and joint latent class mixed models, available on CRAN (with CÃ©cile Proust-Lima).
edrGraphicalTools	**Maintainer of the R package** which provides tools for dimension reduction methods (with JÃ©rome Saracco and Raphael Coudret).
ConvergenceConcepts	**Developer of the R package** to check and visualize convergence of a sequence of random variables in law, in probability, in r-th mean and almost surely (with Pierre Lafaye de Micheaux).
PIGE	**Maintainer of the R package** on Pathway Interaction Gene Environment which is an extension on the ARTP method (with Thérèse Truong).
R2GUESS	**Maintainer of the R package** on Bayesian variable selection regression of multiple correlated phenotypes (with Marc Chadeau-Hyam, Gianlucca Campanella and Sylvia Richardson).
CPMCGLM	**Contributor of the R package** on the correction of the pvalue after multiple coding in generalized linear model (with Jéremie Riou).
rPowerSampleSize	**Contributor of the R package** on the sample size determination with multiple correlated endpoints (with Jéremie Riou and Pierre Lafaye de Micheaux).
CPMTL	**Developer of the Fortran program** to correct the p-value after multiple coding of an explanatory variable in a logistic regression.
Language	**Programming in many languages** (C, Fortran 77, 90, Delphi) and advanced use of statistical software: R (`http://www.biostatisticien.eu/springeR`) SPlus, SPSS, SAS, Minitab.

Research Output

Topics	My research topics since 1999 have focused on model selection, multi-state models, survival analysis, multiple testing problems, and dimension reduction methods applied to a combination of clinical, health and biological data mainly from epidemiological studies. All of these have a large degree of relevance to my recent interest (from 2011) in the analysis of "Omics" data and to research on advanced methods for analysis of high-dimensional data.
Award (2009)	**Scientific Excellence Award** from the University of Bordeaux II.
Award (2016)	**PEDR** which is equivalent to Scientific Excellence Award from the University of Pau et Pays de L'Adour.
Award (2018)	**Lindley Prize Award**. The Lindley Prize is awarded for innovative research in Bayesian Statistics that is presented at an ISBA World Meeting and accepted for publication in the journal Bayesian Analysis. It is related to our publication [A45].
Publications	**53 articles in international peer-reviewed** journals. In addition, 3 in review, 3 book sections and 5 articles in preparation.

Books	**Five books on R software** which is the Swiss knife of data science. We wrote a best-selling book (1st on amazon.fr, Nov. 2013, in the category mathematics applied to computer science), entitled *The R software. Fundamentals of Programming and Statistical Analysis*. Several editions of this book have been published, in French, English, Chinese and Indonesian. A Spanish version is in progress. Each one of these books has around 600 pages. The French edition was nominated for the Roberval Prize (category "university education"). **Two Books on Dynamical Biostatistical models** (one in French and one in English).
Publication Metrics	My publication and citation rate are clearly increasing since 2008. My h-index is 21 (Google scholar, December, 2018).
Conferences - Seminar	I presented my research as invited speaker in several conferences (Australia, France, Canada, Indonesia, England). Since 2009, I have been invited to present my research in different universities (UNSW, Macquarie University New South Wales, University of Aukcland, Queensland University of Technology, MRC Cambridge, University of Copenhagen, The George Institute for Global Health (Sydney)).
Collaborations	I have been invited to work with international research teams by Sylvia Richardson (MRC Cambridge), Thomas Gerds (Copenhagen University), Pierre Lafaye de Micheaux (Université Montrél), Desgagné Alain (Université du Quebec, Montréal), Christian Léger (Université de Montréal), Kim Ahn Le Cao (QFAB, The University of Queensland) and Su Sutanto (University of Sebalas, Indonesia).
Grants for invited Professor	I have secured several grant funding to invite international Professor/Researcher: Thomas Gerds (Copenhagen University) in June (2017), Nathoo Farouk (University of Victoria, Canada) in June 2017, Ricardo Lemos (Climate Coorporation, USA) in September 2017.
Some current projects (since 2015)	Since my new position as Professor at "Université de Pau et Pays de l'Adour" (UPPA), I am an active member of the LMAP ("Laboratory of Mathematics and Their Applications-UMR CNR 5142)" and a member of MIRA ("Milieux et resources aquatiques"). The research federation (MIRA) aims to promote collaborative links between research teams from different institutions (UPPA, CNRS, INRA, IFREMER) sharing common interests on aquatic environments and their resources. In the following, I list some collaborative projects related to the LMAP and/or MIRA:
2015-2016. "EOLE" project	Funded by UPPA as a BQR ("Bonus Qualité Recherche") EOLE aims to introduce particularly innovative action that involves the synthesis of knowledge and analysis of metadata on the environment and fishery resources of the "Golf de Gascogne". Co-supervision with Noelle Bru (LMAP), Gilles Morandeau and Nathalie Caill-Milly (Ifremer) of a research internship student (6 months).
2015-2017 "Mecha- nomics" project	"Investigating molecular mechanisms involved in smoking-induced lung cancer". This project granted by the Cancer Research UK calls will extend novel mathematical modelling approaches to integrate exposure history, genetic variants, and identify the pathological stage(s) at which they exert their effects, and if/when they contribute to the dynamics of disease progression. Our international project involves partners from Europe (UK, The Netherlands, France) and from the USA. In this project I am in charge, in collaboration with Marc Chadeau (Leader of the project from Imperial College,) of the statistical developments of two aims of the project including the implementation of multi-state models and in the development of computationally optimized methods to handle high-dimensional data.
2016-2021. "MICROP- OLIT" project	This projet supported by all teams from MIRA aims to "revisit" the concept of quality of the coastal environment in a multi-disciplinary approach and spatio-temporal to construct indicators, allowing managers a more global view, more ecosystem and thereby foster a more meaningful management. It aims to assess the past and current quality of the environment, but also to predict future changes and to propose possible remedies measures. I am involved in several actions of this project and the coordinator of one action (" Modelisation and Simulation")
2016-2019 "ASMAH" . project	ASMAH "Advanced Statistical Methods for the analysis of High-dimensional data" aims to boost and enhance the current toolkit for integrative analysis of massive datasets, especially for longitudinal design. Using high resolution/ high quality real data this project will generate innovative methodologies that overcome the big data challenge in a real-life context and thus help the full exploration of existing and yet under exploited data.

2018-2020 "CCGIP" project	Principal Investigator of project entitled "Cross Cancer Genomic Investigation of Pleiotropy Effect at Pathway and GxE interactions levels: application to breath and thyroid cancers" which has just been granted by " La Ligue contre le Cancer". This project involves 3 teams: Inserm U1018 (Epidemiology of cancers: genes and environment), LMAP and Imperial College London (Department of Mathematics). The MRC Biostatistics Unit in Cambridge is also a partner of this project. I am the coordinator of the consortium.
2018-2021 "BBCEE" project	Principal Investigator of project entitled "Big model and Big data in Computational Ecology and Environmental Sciences" granted by the "E2S-UPPA 2018 New challenges". This project puts forward innovative and novel methodology to develop tools that will overcome the big data challenge in Ecology and Environmental Sciences and thus reveal nature's secrets, currently hidden in big data.

Publications: refereed journal articles

[A53] B. Liquet and J. Riou. CPMCGLM: An R package for p-value adjustment when looking for an optimal transformation of a single explanatory variable in generalized linear models. *BMC Medical Research Methodology* In press (2019).

[A52] C. Broc, B. Calvo and B. Liquet . Penalized Partial Least Square applied to structured data. *Arabian Journal of Mathematics* pp 1â16 (2019).

[A51] M. Sutton, K. Mengersen, and B . Liquet. Sparse Subspace Constrained Partial Least Squares. *Statistical Computation and Simulation* 1-15, (2018).

[A50 D. Morichon, I. de Santiago, M. Delpey, T. Somdecoste, A. Callens, B. Liquet, P. Liria. Assessment of flooding hazards at an engineered beach during extreme events : Biarritz, SW France. *Journal of Coastal Research* 85, 801- 805 (2018) .

[A49] C. Broc, M. Evangelou, T. Truing, and B. Liquet. Investigating Gene- and Pathway-environment Interaction analysis approaches. *Journal de la Société Française de Statistique* Vol. 159, 2 (2018).

[A48] M. Sutton, R. Thiebaut, and B. Liquet. Sparse partial least squares with group and subgroup structure. *Statistics in Medicine* 37(23), 3338 - 3356 (2018).

[A47] P. Jain, B. Liquet, J. Vlaanderen, B. Bodinier, K. Van Veldhoven, M. Kogevinas, C. Villanueva, R.Vermeulen, P. Vineis, M. Chadeau-Hyam. A multivariate approach to investigate molecular effects of multiple exposures. *Journal of Epidemiology & Community Health* (2018) `doi:10.1136/jech-2017-210061`.

[A46] R. Vermeulen, F. Saberi Hosnijeh, B. Bodinier, L. Portengen, B. Liquet, et al. Prediagnostic blood immune markers, incidence and progression of B-cell lymphoma and multiple myeloma: univariate and functionally-informed multivariate analyses. *International Journal of Cancer* (2018) `doi:10.1002/ijc.31536`.

[A45] Liquet B., K. Mengersen, A. N. Pettitt, and M. Sutton. Bayesian Variable Selection Regression Of Multivariate Responses For Group Data. *Bayesian Analysis* 12, 4 (2017).

[A44] H. P. Chong, Froen F., Richardson S., Liquet B., S. Charnock-Jones and Smith G. C, Age at menarche and the risk of operative delivery. *The Journal of Maternal-Fetal & Neonatal Medicine* 28:1-8 (2017).

[A43] Picat M. Q., Pellegrin I., Bitard J., Wittkop L., Proust-Lima C., Liquet B., Moreau J. F., Bonnet F., Blanco P., Thiebaut R., Integrative analysis of immunological data to explore chronic immune activation in successfully treated HIV patients. Plos One 12(1) (2017). `https://doi.org/10.1371/journal.pone.0169164`.

[A42] Benham T., Duan Q., Kroese D.P., Liquet B., CEoptim: Cross-Entropy R package for optimization. *Journal of Statistical Software* 76 (8) (2017).

[A41] Proust-Lima C., Philipps V., Liquet B., Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: the R package lcmm. *Journal of Statistical Software* 78(2) (2017).

[A40] Delorme P., Lafaye de Micheaux P., Liquet B. and Riou J., Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination. *Statistics in Medicine* 35,1 (2016).

[A39] Liquet B., and Saracco J., BIG-SIR a Sliced Inverse Regression Approach for Massive Data. *Statistics and Its Interface* 9, 4 (2016).

[A38] Liquet B., Lafaye de Micheaux P, Hejblum B., Thiebaut R., Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. *Bioinformatics* 32:1, 35-42 (2016).

[A37] Chong H., Cordeaux Y., Froen F., Richardson S., Liquet B., Charnock-Jones S. and Smith G. C., Age-related changes in murine myometrial transcript profile are mediated by exposure to the female sex hormones. *Aging Cell* 15, 1 (2016).

[A36] Liquet B., Bottolo L., Campanella G., Richardson S. and Chadeau-Hyam M., R2GUESS: GPU-based R package for Bayesian variable selection regression of multivariate responses. *Journal of Statistical Software* 69, 2 (2016).

[A35] Liquet B., and Nazarathy Y., A dynamic view to moment matching of truncated distributions. *Statistics and Probability Letters* 104 87-93 (2015).

[A34] Commenges D., Proust-Lima C., Samieri C., and Liquet B., A universal approximate cross-validation criterion and its asymptotic distribution. *The International Journal of Biostatistics* 11:1, 51-67 (2015).

[A33] Truong T., Liquet B., Menegaux F., Plancoulaine S., Laurent-Puig P., Mulot C., Cordina-Duverger E., Sanchez M., Arveux P., Kerbrat P., Richardson S., Guénel P., Breast cancer risk, nightwork, and circadian clock gene polymorphisms. *Endocr Relat Cancer* 21:4, 629-38 (2014).

[A32] Chavent M., Girard S., Kuentz V., Liquet B., Nguyen T. M. N., Saracco J., A sliced inverse regression approach for data stream. *Computational Statistics* 29:5, 1129-1152 (2014)

[A31] Coudret R., Liquet B., Saracco J., Comparison of sliced inverse regression approaches for underdetermined cases. *Journal de la Société Française de Statistique* Vol 155:2 (2014).

[A30] Lafaye de Micheaux P., Liquet B., Marques S. and Riou J., Power and sample size determination in clinical trials with multiple primary continuous correlated end points. *Journal of Biopharmaceutical Statistics* 24:2, 378-97, (2014).

[A29] Bottolo L., Chadeau-Hyam M., Hastie D. I., Zeller T., Liquet B., Castagne R., Wild P. S., Schillert A., Munzel T., Tregouet D., Cambien F., Petretto E., Blankenberg S., Tiret L. and Richardson S., GUESS-ing polygenic associations with multiple phenotypes using a GPU-based Evolutionary Stochastic Search algorithm. *Plos Genetics* 9:8, (2013).

[A28] Chadeau-Hyam M., Jombart T., Campanella G., Bottolo L., Vineis P., Liquet B., Vermeulen R., Deciphering the complex: Methodological overview of statistical models to derive OMICS-based biomarkers. *Environmental and Molecular Mutagenesis* 54:7, 542-557, (2013).

[A27] Liquet B., and Riou J., Correction of the significance level after multiple coding of an explanatory variable in generalized linear model. *BMC Medical Research Methodology* 13:75, (2013).

[A26] Liquet B., Le Cao K., Hocini H., and Thiebaut R., A novel approach for biomarker selection and the integration of repeated measure experiments from two platforms. *BMC, Bioinformatics* 13:325, (2012).

[A25] Liquet B., Rondeau V., and Timsit J. F., Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. *BMC Medical Research Methodology* 12:79, (2012).

[A24] Coeurjolly J.F., Makao M., Timsit J.F. and Liquet B., Attributable risk estimation for adjusted disability multistate models: application to nosocomial infections. *Biometrical Journal* 54:5, 600-616, (2012).

[A23] Chavent M., Kuentz V., Liquet B., and Saracco J., ClustOfVar: An R Package for the Clustering of Variables. *Journal of Statistical Software* 50:13, 1-16, (2012).

[A22] Josse J., Chavent M., Liquet B., and Husson F., Handling missing values with Regularized Iterative Multiple Correspondence Analysis. *Journal of Classification* 29:1, 91-116, (2012).

[A21] Commenges D., Liquet B. and Proust-Lima C., Choice of prognostic estimators by estimating difference of expected conditional Kullback-Leibler risks. *Biometrics* 68:2, 380-387, (2012).

[A20] Liquet B. and Saracco, J., A graphical tool for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. *Computational Statistics* 27:1, 103-125, (2012).

[A19] Chavent M., Kuentz V., Liquet B. and Saracco J., A sliced inverse regression approach for a stratified population. *Communications in Statistics - Theory and Methods* 40:21, (2011).

[A18] Laboute E., Liquet B., Savalli L., Puig P., Trouve P., Influence of the type of knee brace on clinical postoperative evolution after anterior cruciate ligament reconstructions in competitive sportspeople. *Journal de Traumatologie du Sport*, 28: 3, 103-125, (2011)

[A17] Liquet B. and Commenges D., Choice of estimators based on different observations : Modified AIC and LCV criteria. *Scandinavian Journal of Statistics* 38, 268-287, (2011).

[A16] Liquet B., Choix d'estimateurs basé sur le risque de Kullback-Leibler. *Journal de la Société Française de Statistique* 151:1, (2010).

[A15] Chavent M., Liquet B. and Saracco J., A semiparametric approach for a multivariate sample selection model. *Statistica Sinica* 20:2, 513-536, (2010).

[A14] Kuentz V., Liquet B. and Saracco J., "Bagging" version of Sliced Inverse Regression. *Communications in Statistics- Theory and Methods* 39, 1985-1996, (2010).

[A13] Lafaye de Micheaux P. and Liquet B., ConvergenceConcepts: an R package to investigate various modes of convergence. *The R Journal* 1, 18-25, (2009).

[A12] Lafaye de Micheaux P. and Liquet B., Understanding Convergence Concepts: A Visual-Minded and Graphical Simulation-Based Approach. *The American Statistician* 63:2, 173-178, (2009).

[A11] Commenges, D. and Liquet B., Asymptotic distribution of score statistics for spatial cluster detection with censored data. *Biometrics* 64, 1-8, (2008).

[A10] Cenier T., Amat C., Litaudon P., Garcia S.StatLearn 2013, Lafaye de Micheaux P., Liquet B., Roux S. and Buonviso N., Odor vapor pressure and quality modulate local field potential oscillatory patterns in the olfactory bulb of the anesthetized rat. *European Journal of Neuroscience* 27:6, 1432-1440, (2008).

[A9] Liquet B. and Saracco, J. Application of the bootstrap approach to the choice of dimension and the α parameter in the SIR_α method. *Communication in Statistics - Simulation and computation* 37:6, 1198-1218, (2008).

[A8] Rondeau V., Michiels S., Liquet B. and Pignon J. P., Investigating trial and treatment heterogeneity in an individual patient data meta-analysis of survival data by means of the penalized maximum likelihood approach. *Statistics in Medicine* 27, 1894-1910, (2008).

[A7] Commenges D., Joly P., Gégout-Petit A. and Liquet B., Choice between semi-parametric estimators of Markov and non-Markov multi-state models from generally coarsened observations. *Scandinavian Journal of Statistics* 34, 33-52, (2007).

[A6] Liquet B. and Saracco, J., Pooled marginal slicing approach via SIR_α with discrete covariables. *Computational Statistics* 22:4, 599-617, (2007).

[A5] Liquet B., Saracco, J. and Commenges D., Selection between proportional and stratified hazards models based on Expected Log-likelihood. *Computational Statistics* 22:4, 619-634, (2007).

[A4] Liquet B. and Commenges D., Computation of the p-value of the maximum of score tests in the generalized linear model; application to multiple coding. *Statistics & Probability Letters* 71, 33-38, (2005).

[A3] Liquet B. and Commenges D., Estimating the expectation of the log-likelihood with censored data for estimator selection. *Lifetime data analysis* 10, 351-367, (2004).

[A2] Liquet B., Sakarovitch C. and Commenges D., Bootstrap choice of estimators in non-parametric families: an extension of EIC. *Biometrics* 59, 172-178, (2003).

[A1] Liquet B. and Commenges D., Correction of the p-value after multiple coding of an explanatory variable in logistic regression. *Statistics in Medicine* 20, 2815-2826, (2001).

Publications: refereed conference papers

[RC1] Thiebaut R., Liquet B., Hocini H., Hue S. Richert L. Raimbault M. Le Cao K. and Levy Y. A new method for integrated analysis applied to gene expression and cytokines secretion in response to LIPO-5 vaccine in HIV-negative volunteers. *Retrovirology* 9:2, 121, (Poster presentation at AIDS Vaccine 2012).

Publications: scholarly books

[B7] Lafaye de Micheaux P., Drouilhet R. and **Liquet B.** (2016), "Perangkat Lukan R - Dasar-dasar Pemrograman dan Analisis Statistika", *Statistika dan Pemrograman.* (In Indonesian.)

[B6] Lafaye de Micheaux P., Drouilhet R., **Liquet B.**, Pan D., Tang N. and Li Q. (2015), "R:软件教程——从入门到精通统计分析 (R: the textbook - Master the language & Perform statistical analyses)", *Higher Education Press* (largest textbook publisher in China). (In Chinese.)

[B5] Commenges D.; Jacqmin-Gadda H.(eds); Alioum A.; Joly P.; Leffondre K.; Liquet B.; Proust-Lima C.; Rondeau V. & Thiébaut R. (2015) Dynamical Biostatistical models, *Series: Chapman & Hall/CRC Biostatistics Series*

[B4] Commenges D.; Jacqmin-Gadda H.(eds); Alioum A.; Joly P.; Leffondre K.; Liquet B.; Proust-Lima C.; Rondeau V. & Thiébaut R. (2015) Modèles Biostatistiques pour l'Epidémiologie. *De Boeck Superieur*

[B3] Lafaye de Micheaux P., Drouilhet R. and **Liquet B.** (2014), "Le logiciel R, Maitriser le langage - Effectuer des analyses (bio)statistiques", *Springer, Collection: Statistique et probabilités appliquées, Vol. 1*, 2nd Edition., 674 p., Broché, ISBN: 978-2-8178-0534-4. (In French.)

[B2] Lafaye de Micheaux P., Drouilhet R. and **Liquet B.** (2014), "The R software. Fundamentals of Programming and Statistical Analysis", *Springer, Collection: Statistics and Computing, vol. 40*, 655 p., ISBN: 978-1-4614-9019-7.

[B1] Lafaye de Micheaux P., Drouilhet R. and **Liquet B.** (2011), "Le logiciel R, Maitriser le langage - Effectuer des analyses statistiques", *Springer - Collection: Statistique et probabilités appliquées, Vol. 1*, 1st Edition., XVI, 527 p., Broché, ISBN: 978-2-8178-0114-8. (In French.) **Nominated for the Roberval Prize.**

Publications: book sections

[C3] Saracco J., Gannoun A., Guinot C. and Liquet B., A semiparametric approach to estimate reference curves for biophysical properties of the skin. In *Statistical Methods for Biostatistics and Related Fields.* Berlin: Springer, 181-205, (2007).

[C2] Gannoun A., Liquet B., Saracco J. and Wolfgang U., A kernel method in analysis of replicated micro-array experiments. In *Statistical Methods for Biostatistics and Related Fields.* Berlin: Springer, 45-61, (2007).

[C1] Liquet B. and Commenges D., Selecting a semi-parametric estimator by the expected log-likelihood. In *Probability, Statistics and Modelling in Public Health*, Springer, 332-349, (2005).

Invited conferences/lectures

[IV28] 2019 (June), Liquet B, Invited speaker at International workshop on "Perspectives On High-dimensional Data Analysis" (https://indico.uu.se/event/526/) Uppsala University in Sweden

[IV27] 2019 (May), Liquet B, Invited Professor, seminar and lectures for "Advanced Statistical Topics in Health Research", organized at University of Copenhagen (Danemark).

[IV26] 2019 (March), Invited Professor (one week) Seminar and running a workshop on "Statistical methods for high dimensional and Massive Data", Sebelas Maret University at Solo (Indonesia)

[IV25] 2018 (October), Liquet B, Invited speaker at "1er événement DATA & IA à Pau les 28, 29 et 30 septembre 2018", organized at Pau https://www.ia-pau.com/#speakers.

[IV24] 2018 (September), Liquet B, Invited speaker at "CIBB2018: Computational Intelligence methods for Bioinformatics and Biostatistics", organized at Caparica (Portugal) https://eventos.fct.unl.pt/cibb2018/pages/program-invited.

[IV23] 2018 (July), Liquet B, Invited Professor, seminar and lectures on "Statistical methods for high dimensional data", organized at the University of Victoria (Canada).

[IV22] 2018 (June), Liquet B, Invited speaker at "Workshop in honor of Daniel Commenges' 70th birthday", organized at Bordeaux University (Bordeaux).

[IV21] 2018 (May), Liquet B, Invited Professor, seminar and lectures on "Statistical methods for analyzing Big-DATA", organized at the University of Copenhagen (Danemark).

[IV20] 2018 (April), Liquet B, Invited speaker at International workshop on "Perspectives On High-dimensional Data Analysis" (http://hdda-viii.uca.ma)

[IV19] 2018 (Avril), Liquet B, Invited lectures the Short course on "Stat-XP: statistics to analyse OMICs data and characterise the exposome", organized Imperial College London, England.

[IV18] 2017 (December), Liquet B. Organiser of a Contributed Session at the 2017 IASC-ARS/NZSA Conference. "Statistical Methods for the Analysis of High-Dimensional and Massive Data", New Zealand.

[IV17] 2017 (October), Liquet B. Organiser of a Contributed Session at the 2017 WANRM workshop "Mathematical Methods for Modelling Natural Resources", The University of Queensland (Australia).

[IV16] 2017 (October), Invited Professor (one week) Seminar and running a workshop on "Reproducible workflow analysis using R", Sebelas Maret University at Solo (Indonesia)

[IV15] 2016 (December), Lafaye de Micheaux P., Liquet B. and Riou J., "Type-II generalized family-wise error rate formulas with application to sample size determination", 9th International Conference of the ERCIM (European Research Consortium for Informatics and Mathematics) Spain.

[IV14] 2016 (December), "Bayesian Variable Selection Regression Of Multivariate Responses For Group Data", Invited to the Big Bayes Session at the Australian Statistical Conference.

[IV13] 2016 (December), "Statistical Methods For Analysing High-Dimensional Data and Massive Data", International Conference on Mathematics : Education, Theory, and Application (ICMETA). Universitas Sebelas Maret (Indonesia).

[IV12] 2016 (December), Two invited lectures: "A tutorial for penalized regression models" and "A tutorial for PLS and Bayesian Variable Selection". Short course on "Stat-XP: statistics to analyse OMICs data and characterise the exposome", organized Imperial College London, England.

[IV11] 2016 (December) Sutton M., Liquet B., and Thiébaut R. "Sparse Group Subgroup PLS for Genomics". International Symposium for Big Data Visualisation Analytics (BDVA) Sydney.

[IV10] 2016 (October) Lafaye de Micheaux P., Liquet B. and Sutton M., "A Unified Regularized Group PLS Algorithm Scalable to Big Data", journée STAtistique de Rennes (jSTAR), 13rd edition on Big Data, Rennes, France.

[IV9] 2016 (July) Dimension Reduction approaches for BIG-DATA. Workshop on Big-Data organized by ACEMS at Queensland University of Technology.

[IV8] 2016 (February) Statistical Methods for Analysising High-dimensional Data and Massive Data. Lecture at the University of Melbourne.

[IV7] 2015 (December), Multivariate approaches: Dimension reduction for big data sets. Short course on "Statistical approaches to characterize the exposome: Overview and Perspective", organized Imperial College London, England.

[IV6] 2014 (December), Multivariate approaches: Dimension reduction for big data sets. Short course on "Statistical approaches to characterize the exposome: Overview and Perspective", organized Imperial College London, England.

[IV5] 2014 (November), R2GUESS: A Graphic Processing Unit-Based R Package for Bayesian Variable Selection Regression of Multivariate Responses, Bayes on the Beach conference organized in Gold Coast, Australia.

[IV4] 2013 (October), Multi-state models and nosocomial infections. Workshop on "dynamic predictions for repeated markers and repeated events: models and validation in cancer", organized in Bordeaux, France.

[IV3] 2013 (April), Strategies to analyse 'omics' data: From standard T-tests in genomics to a more general Bayesian approach. Workshop Statlearn'13. Bordeaux University (France).

[IV2] 2009 (November), Choice of estimators based on different observations: Modified AIC and LCV criteria. Colloque CRM-ISM-GERAD de Statistique. McGill University (Canada).

[IV1] 2009 (October), Choix d'estimateurs basés sur des observations différentes. GDR "Statistique et Santé", Université Paris-Descartes.

Invited seminar (2011-2017)

[IS28] 2017 (May) Bayesian Network approach for predicting coastal risk. SIAME seminar, Université de Pau et Pays de l'Adour.

[IS27] 2017 (May) Bayesian Variable Selection Regression Of Multivariate Responses For Group Data. LMAP seminar, Université de Pau et Pays de l'Adour.

[IS26] 2016 (November) A Unified Regularized Group PLS Algorithm Scalable to Big Data. ISG (Intelligent Systems Group) seminar at UPV/EHU University (Spain).

[IS25] 2016 (November) Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination. LMAP seminar, Université de Pau et Pays de l'Adour.

[IS24] 2016 (October) Bayesian Variable Selection Regression Of Multivariate Responses For Group Data. Seminar at the OMMAS team from the IMB (Institute Matheémetics de Bordeaux) University of Bordeaux.

[IS23] 2016 (August) BIG-SIR a Sliced Inverse Regression Approach for Massive Data. Seminar at the Centre for Statistics from The University of Queensland.

[IS22] 2016 (May) BIG-SIR a Sliced Inverse Regression Approach for Massive Data. LMAP, Université de Pau et Pays de l'Adour.

[IS21] 2016 (February) Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination. Victorian Centre For Biostatistics Seminar, Melbourne.

[IS20] 2016 (February) Semi-parametric regression approach for massive data. Maths seminar series at Queensland University of Technology.

[IS19] 2016 (February) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. SSAI Queensland Branch Event.

[IS18] 2015 (November) Statistical Methods for the Analysis of High-dimensional data Applied in Biostatistics Context. MIRA, Milieux et ressources aquatiques. Université de Pau et Pays de l'Adour.

[IS17] 2015 (October) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. ISPED, University of Bordeaux.

[IS16] 2015 (October) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. LMAP, Université de Pau et Pays de l'Adour.

[IS15] 2015 (August) BIG-SIR a Sliced Inverse Regression Approach for Massive Data. ACEMS seminar, Queensland University of Technology.

[IS14] 2015 (August) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. Queensland Brain Institute, Australia.

[IS13] 2015 (June) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. Imperial College, London.

[IS12] 2015 (June) Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context. ACEMS seminar, University of Queensland.

[IS11] 2015 (January) R2GUESS: A Graphic Processing Unit-Based R Package for Bayesian Variable Selection Regression of Multivariate Responses. Queensland Brain Institute, Australia.

[IS10] 2014 (December) Strategies to analyse 'omics' data. Grenoble University, France.

[IS9] 2013 (November) Type-II Generalized Family-Wise Error Rate Formulas with Application to sample Size Determination. Universitas Sebelas Maret, Indonesia.

[IS8] 2012 (April) Multi-state approaches to model nosocomial infection in intensive care units. Universitas Sebelas Maret, Indonesia.

[IS7] 2012 (March) On Akaike and likelihood cross-validation criteria for model selection. Queensland University of Technology (QUT), Brisbane.

[IS6] 2012 (March) A novel approach for the integration of repeated measures experiments and biomarker selection . University Queensland Seminar Series - Computational Biology.

[IS5] 2012 (March) Choice of Estimators by Estimating Differences of Expected Kullback-Leibler Risks, Faculty of Science Macquarie University New South Wales.

[IS4] 2012 (March) Multistate approaches to model nosocomial pneumonia disease in intensive care units. The George Institute for Global Health, Sydney.

[IS3] 2012 (March) On Akaike and likelihood cross-validation criteria for model selection. School of Mathematics and Statistics. The University of New South Wales, Sydney.

[IS2] 2012 (February) On Akaike and likelihood cross-validation criteria for model selection. Department of Statistics. The University of Auckland.

[IS1] 2011 (September) On Akaike and likelihood cross-validation criteria for model selection. Department of Biostatistics. University of Copenhagen.

International and national communications, proceeding (2007-2018)

[CO33] Liquet B. " Bayesian Variable Selection Regression Of Multivariate Responses For Group Data". Bayesian Statistics in the Big Data Era, CIRM, Marseille Luminy, France, (November 2018).

[CO32] Liquet B. " Statistical Methods for the Analysis of High-Dimensional and Massive Data using R". Australasian Applied Statistics Conference New Zealand, (December 2018).

[CO31] Liquet B., and Saracco J. "BIG-SIR a Sliced Inverse Regression Approach for Massive Data", speaker in Contributed Session at the 2017 IASC-ARS/NZSA, New Zealand, (December 2017).

[CO30] Liquet B., Bayesian Variable selection regression of multivariate responses for group data, ISBA 2016 World Meeting International Society for Bayesian Analysis, Sardinia (Italy) June 2016.

[CO29] Liquet B., Statistical Methods for the Analysis of High-dimensional Data and Massive Data, Workshop on Computational and Mathematical Foundations for Big Data Analytics, QUT, Australia (July 2016).

[CO28] Liquet B., Lafaye de Micheaux P, Hejblum B., Thiebaut R., Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context., The R User Conference 2016, Stanford, California (US), June 2016.

[CO27] Liquet B., Lafaye de Micheaux P, Hejblum B., Thiebaut R., Group and Sparse Group Partial Least Square Approaches Applied in Genomics Context., Les cinquièmes Rencontres R, Toulouse, France, June 2016.

[CO26] Liquet B., and Saracco J., BIG-SIR a Sliced Inverse Regression Approach for Massive Data, Quatrièmes rencontres R, Grenoble (France), June 2015.

[CO25] Lafaye de Micheaux P., Liquet B., Perminder S., Anbupalam T. and Wei W., New Statistical tools to study heritabiliy of the brain, Australian Statistical Conference in conjunction with the Institute of Mathematical Statistics Annual Meeting, Sydney (Australia), July 2014.

[CO24] Delorme, P., Lafaye de Micheaux P., Liquet, B. and Riou, J., Type-II Generalized Family-Wise Error Rate Formulas with Application to Sample Size Determination, 8th International Conference on Multiple Comparison Procedures, Southampton (England), July 2013.

[CO23] Delorme, P., Lafaye de Micheaux P., Liquet, B. and Riou, J., Package SSDDA : Sample Size Determination and Data Analysis in the context of continuous co-primary endpoints in clinical trials, Deuxièmes rencontres R, Lyon (France), June 2013.

[CO22] Delorme, P., Lafaye de Micheaux P., Liquet, B. and Riou, J., Calcul de taille d'échantillon dans le cadre de critères de jugements multiples avec un contrôle de la r-power et du gFWER, 44 Journées de Statistique, Toulouse (France) Mai 2013.

[CO21] Liquet B., Chadeau-Hyam M., Bottolo L., and Richardson S. R2GUESS: a GPU-based R package for a Bayesian variable selection model accommodating multivariate responses, Statistical Methods for (Post)-Genomics Data 2013, Amsterdam, January 2013.

[CO20] Chavent M., Genuer R., Liquet B., and Saracco J., ClustOfVar: an R package for dimension reduction via clustering of variable. Application to supervised classification and variables selection for omics data (Poster presentation)„Statistical Methods for (Post)-Genomics Data 2013, Amsterdam, January 2013.

[CO19] Liquet B., Chadeau-Hyam M., Bottolo L., and Richardson S., GUESS: GPU-based C^{++} software for Bayesian variable selection regression of multiple responses (Poster presentation), Australasian Applied Statistics Conference 2012, Queenstown (New Zeland) December 2012.

[CO18] Liquet B., Le Cao K., Hocini H., and Thiebaut R., A novel approach for biomarker selection and the integration of repeated measure experiments from two platforms, NZSA 2012 Conference, Dunedin (New Zealand), December 2012.

[CO17] Delorme P., Lafaye de Micheaux P., Liquet B. and Riou J., Power and sample size computation for a control of the "r-power" (Poster presentation), NZSA 2012 Conference, Dunedin (New Zealand), December 2012.

[CO16] Lafaye de Micheaux P., Liquet. B., Marque S. and Riou J., Power and sample size determination in clinical trials with multiple primary continuous endpoints, XXVIth International Biometric Conference, Kobe (Japan), August 2012.

[CO15] Liquet B., Le Cao K., Hocini H., and Thiebaut R., A novel approach for biomarker selection and the integration of repeated measure experiments from two platforms, XXVIth International Biometric Conference, Kobe (Japan), August 2012.

[CO14] Liquet B., Commenges D. and Proust-lima C., Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback-Leibler risks, Biometrics by the Blowholes, Kiama, New South Wales, December 2011, Australia.

[CO13] Riou J., Liquet B., Correction of the significance level after multiple coding of an explanatory variable in generalized linear model., 7th International Conference on Multiple Comparison Procedures, Washington D.C., USA, August 2011.

[CO12] Liquet B., Proust-Lima C., lcmm: an R package for estimation of latent class mixed models and joint latent class models. The R User Conference 2011. Coventry, UK.

[CO11] Chavent M., Girard S., Kuentz V., Liquet B., Nguyen T. M. N., Saracco J., An adaptive SIR method for block-wise evolving data streams. ASMDA -XIV International Conference, Rome Italy, 2011.

[CO10] Riou J., Liquet B., Méthodes de correction du degré de signification pour une recherche de codage optimal dans un modèle linéaire généralisé 43èmes Journées de Statistique (SFdS), Tunis, Mai 2011.

[CO9] Chavent M., Kuentz V., Liquet B., Saracco J., Classification de variables: le package ClustOfVar. 43èmes Journées de Statistique (SFdS), Tunis, Mai 2011.

[CO8] Liquet B., Choice of estimators based on different observations: Modified AIC and LCV criteria. XXVèmes International Biometric Conference, Florianopolis (Brazil), december 2010.

[CO7] Chavent, M., Kuentz, V., Liquet, B., Saracco, J., Regression inverse par tranches pour une population stratifiee. 42èmes Journées de Statistique (SFdS), Marseille, Mai 2010.

[CO6] Lafaye de Micheaux, P. and Liquet, B., Le Package R ConvergenceConcepts : un nouvel outil graphique pour l'étude de quelques modes de convergence de variables aléatoires, Marseille, Mai 2010.

[CO5] Chavent M., Kuentz V. and Liquet B., Données manquantes en ACM : l'algorithme NIPALS. XVIèmes Rencontres de la Société Francophone de Classification, Grenoble 2009.

[CO4] Liquet B. and Saracco J., A criterion for selecting the number of slices and the dimension of the model in SIR and SAVE approaches. ASMDA -XIII International Conference, Vilnius Lithuania, 2009.

[CO3] Nguile Makao M., Timsit J.F., Coeurjolly J.F., and Liquet B., Prédiction de la pneumonie nosocomiale Ã l'aide d'un modèle multi-états., XXXXIe Journées de Statistique, Bordeaux, 2009.

[CO2] Chavent M., Liquet B. and Saracco J., Two steps estimation in a multivariate semiparametric sample selection model XXXXe Journées de Statistique, Montréal, 2008.

[CO1] Desjardins S., Desgagné A., Lafaye de Micheaux P., Liquet B., Generalization of the Paired T-Test for the Missing Values Case, Joint Statistical Meeting, Salt Lake City, Poster, 2007.

Research activities (summary)

Research field:

Statistics
Applied Statistics
Biostatistics
Mathematical Statistics

My research topics since 1999 have focused on model selection, multi-state models, survival analysis, multiple testing problems, and dimension reduction methods applied to a combination of clinical, health and biological data mainly from epidemiological studies. All of these have a large degree of relevance to my recent interest (from 2011) in the analysis of "Omics" data (transcriptomic, proteomic, metabolomic,...). As outlined below, I have made significant contributions to all of these areas.

Model selection

After my PhD thesis ([A1],[A2],[A3]), I worked on developing an information criterion for choosing between a proportional hazard model and a stratified model [A5]. I also focused on the choice of semi-parametric estimators in multi-state models [A7]. I proposed a criterion for choosing between Markovian and non-Markovian models. I have been working on the development of an Akaike Information Criterion (AIC) to select models based on different observations [A17]. I was specifically interested in the estimation of the variability of the developed criteria for model selection. Moreover, following the work on the choice of estimators based on different observations, in collaboration with Daniel Commenges (Bordeaux University), we developed a selection criterion to choose between several prognostic models. Particularly, in the case of prostate cancer, a joint modeling of the marker and the risk of PSA relapse of prostate cancer predicts relapse of cancer, taking into account the progression of the illness. The aim of this work was to propose an information criterion and a practical tool to choose among several statistical models, such as joint models including data modeling of repeated PSA markers [A20]. Finally, we proposed a universal criterion based on cross-validation for choosing between different estimators [A34].

Dimension reduction and semi-parametric model

My main contribution concerns SIR (Sliced Inverse Regression) methods introduced by Li (1991), for estimating the parametric part of a very general semi-parametric regression model: $y = f(x'b_1, ..., x'b_K, \epsilon)$ where the distribution free error ϵ is independent of the covariate x, the link function f is unknown and arbitrary, and K is strictly inferior to the dimension p of x. In collaboration with Jerome Saracco, we developed an extension of these methods when the dependent variable y is multidimensional and in presence of a qualitative covariate in the model ([A6], [A18]). We also proposed a graphical approach based on the bootstrap method to the simultaneous choice of size K and α parameter for the SIR_α method [A9]. An approach based on the "bagging" method has also been developed for SIR to estimate more efficiently space dimension reduction when samples are small [A14]. In collaboration with colleagues from Bordeaux university, we focused on data arriving sequentially by blocks in a stream. In this context, a semi-parametric regression model involving a common EDR (Effective Dimension Reduction) direction is assumed in each block. Then, our goal was to estimate this direction at each arrival of a new block [A31]. This approach is particularly relevant for avoiding the storage of big data sets. In [A30] we proposed an extension of the SIR method based on the QZ algorithm and on the Moore-Penrose pseudo-inverse, when the dimension of the covariate is greater than the number of observations. This represents a major break through for SIR based dimension reduction because our developed method enables to select the most relevant predictors (e.g., genes) associated to the response variable without linearity assumptions typically employed by other methods in the context of genome wise association studies. By exploiting the "divide and conquer" principle we have recently proposed in [A40] a new SIR estimator (BIG-SIR) for analyzing massive data sets.

Multi-state model and survival model

Besides the various works in connection with the model selection theme, several of my contributions are also focusing specifically on the following models:

Joint model:

In collaboration with Virginie Rondeau (INSERM, Bordeaux), we proposed to estimate both the heterogeneity in survival time between trials and the heterogeneity in terms of treatment in a meta-analysis on individual data. A model with multivariate frailties with correlated random effects acting multiplicatively on the baseline hazard function or in interaction with the treatment is estimated using the penalized likelihood method with a Laplace approximation [A8]. This original approach enables to give smoothed estimates of the hazard function that represent incidence in epidemiology taking into account heterogeneity between trials which is an important issue the context of meta-analysis.

Multi-state model:

With my PhD student Molière Nguilé (defended in 2009), we were interested in modeling nosocomial pneumonia in intensive care units (ICU) with a multi-state model. Several approaches were compared, including semi-Markov models and non-homogeneous Markov models. We also focused on the estimation of attributable mortality to nosocomial pneumonia [A23]. We defined a new estimation procedure to evaluate adjusted attributable risk using a multi-state model including covariates. Attributable risk has become an important concept in clinical epidemiology and the novel approach is useful for clinicians to make a decision regarding a patient's follow-up treatment. This work was a part of the ANR Public Health (2010-2013) project Dynamic Models for Epidemiology Biostatistics (MOBIDYQ). Several model extensions have been proposed. I put forward a method to integrate frailties in the multi-state model to take into account the heterogeneity on the data [A24].

Multiple testing

I proposed a method for correcting the significance level when multiple coding of an explanatory variable is tested in a generalized linear model [A4]. In collaboration with Daniel Commenges, we have adapted this method for the detection of geographic clusters in the presence of censored data [A11]. The proposed approach has the important advantage to not require an independence assumption between clusters. With my PhD student Jeremie Riou, we have extended, based on resampling approaches, the correction of the significance level for quadratic statistics [A26]. We also recently developed new innovative techniques to determine sample size in the context of multiple co-primary endpoints [A29]. We provided formulas for computation of sample size, and for data analysis. This novel approach has been used by Danone research to design clinical trials. This work has been greatly extended recently in [A39], where we defined a new type-II r-generalized family-wise error rate. We then obtained very general formulas to compute the probability of rejecting at least r false null hypotheses. Our R package (`rPowerSampleSize`) enables one to compute all these quantities. This paper puts forward a fundamental new concept. It is foreseen that this paper, together with its associated package, will modify the current practice of several clinical research studies in the near future.

Omics data

I present in the following some works related to the analysis of Omics data.

Integrative analysis in context of HIV vaccination study:

In the context of HIV vaccine evaluation, we have proposed a dimension reduction method to analyze the relationship between transcriptomic profiles and cytokines production while taking into account prior knowledge on biological pathways. We defined in [A25] a two-step analysis combining a multilevel approach and a multivariate approach to reveal separately the effects of conditions within subjects from the biological variation between subjects. The approach is extended to two-factor designs and to the integration of two matched data sets. Importantly, this enables to reveal clusters of predictors (e.g., genes), that are associated with the phenotypes (e.g., cytokines secretion). These clusters can be seen as gene signatures to predict future cytokine response.

Sparse Bayesian variable selection model:

In collaboration with Sylvia Richardson and colleagues from Imperial College in London, we are working on a sparse Bayesian variable selection regression of multiple correlated phenotypes. The method is based on Evolutionary Stochastic Search algorithms to explore the huge dimensional space model in context of genome wide association studies. We have recently showed the advantage of the proposed method in comparison to others Bayesian and non-Bayesian approaches [A28]. I have also developed with colleagues from Imperial College in London `R2GUESS` ([A38]) an R package which is a wrapper of the GUESS (Graphical processing Unit Evolutionary Stochastic Search) program. GUESS is a computationally optimised C++ implementation of a fully Bayesian variable selection approach that can analyse, in a genome-wide context, single and multiple responses in an integrated way. The program uses packages from the GNU Scientific Library (GSL) and offers the possibility to re-route computationally intensive linear algebra operations towards the Graphical Processing Unit (GPU) through the use of a proprietary CULA-dense library.

Gene- and pathway-environment interaction analysis:

I have recently developed in collaboration with the Research Center for Epidemiology (INSERM U1018; France) a method for computing gene and pathway p-values using the Adaptive Rank Truncated test (ARTP). This method can analyze pathways/genes based on a genetic association study, with a binary case-control outcome or a survival outcome ([A32]). I provided an R package `PIGE` based on parallel computing and on an efficient permutation algorithm to analyze gene- and pathway-environment interaction analysis. I was participating to a grant project on hormonal, environmental and genetic risk factors of breast, lung, and head and neck cancers among women. We used recent development methods on profile regression to define profiles based on genetic and environmental characteristics of study subjects associated with a particular level of cancer risk (Article in preparation).

Clustering of variables:

In collaboration with Marie Chavent (Bordeaux University) we have developed an R package `clustOfvar` for the clustering of variables [A23]. This package, especially devoted to mixtures of quantitative and qualitative variables, could be an useful tool for dimension reduction and variable selection. This method and the associated algorithm will be the key concept to study association between qualitative clinical predictors and genetic predictors.

Teaching activities (summary)

I have participated actively in university-level teaching in my career. Since completing my PhD studies, I have had many opportunities to give lectures at different levels for students who come from different disciplines: Mathematics, Statistics, Medicine and Biology. Theses experiences have allowed me to develop my teaching abilities and also a passion for this important discipline.

2015 - 2017 **Professor at Pau et Pays de l'Adour (192 h/year) (Anglet and Pau Campus).**

Coordinator/Instructor of the course *Biostatistics 1* - L2 (undergraduate second year students).

Coordinator/Instructor of the course *Data Analysis* - L2.

Coordinator/Instructor of the course *Biostatistics 2* - L3 (undergraduate third year students).

Coordinator/Instructor of the course *Probability and Statistics* - L1 (undergraduate first year students).

Coordinator/Instructor **Advanced Data Mining/Machine Learning** for MSID Master Students.

Instructor of the course **Time Series Analysis** for DynEA (Aquatic Ecosystem Dynamics) Master Students.

Instructor of the course **Generalized Linear Model** for DynEA Master Students.

2013 - 2015 **Senior Lecturer at The university of Queensland (Brisbane).**

Co-designer with Dirk Kroese of the new course STAT1301 *Advanced Analysis of Scientific Data.*

Coordinator of the course STAT1201 *Analysis of Scientific Data* (900 undergraduate students).

Coordinator of the course STAT3003 *Experimental Design.*

Instructor of the course STAT4401 *Advanced Statistics.*

Instructor of the course STAT1201 *Analysis of Scientific Data.*

Supervision of honours and Masters students.

2013 - 2014 **Volunteer for teaching at Sebelas Maret University at Solo (Indonesia)**

Expanding teaching activities through collaboration between SMP and the faculty of Science and Mathematics from Universitas Sebelas Maret.

Designer of a sequence of lectures on R software

Coordinator and instructor of the course STAT1301 *Advanced Analysis of Scientific Data*

Develop a web teaching activity for supporting these courses (forum, recording).

2007 - 2011 **Associate Professor Bordeaux 2 University (192 h)**

Coordinator of the course "Initiation to R software" during the summer school proposed by ISPED.

Coordinator of the course "Modelling", Master of Public Health.

Coordinator of the course "Introduction to Statistics" - Master of Public Health.

Coordinator of the course "Inferential Statistics" - Master of Public Health.

Introduction to "modelling" - Master of Public Health (face-to-face and web-based).

Coordinator of the course "Statistical methods in Health": web-based teaching - Master degree.

Instructor of the course "R software" - Master of Public Health, Biostatistics.

Instructor in the course "Algorithms and numerical methods" - Master of Public Health, Biostatistics.

Instructor in the course "Multidimensional Data Analysis" - Master of Public Health, Biostatistics.

Teaching: "Data Analysis" Master of Public Health year 2, speciality Biostatistics.

Trainee supervision Master Public Health year 1 and 2.

2004 - 2007 **Associate Professor Grenoble 2 University: IUT STID (192 h)**

Teaching at undergraduate level (IUT) : Inferential Statistics, analysis, probability, algebra, anova, linear regression; lectures, tutorials and practical courses (SAS, SPSS, MINITAB, R).

Trainee supervision students in IUT STID.

Teaching of a Master in Statistics University Pierre Mendàs France: "statistic, combinatory and computing" (lecture and practical course).

Supervision (with J. F. Timsit - INSERM Grenoble) of a Master Master trainee in statistics - Grenoble (from Feb to Sept 2006).

2003 - 2004 Lecturer Montpellier 2 University (96h)

Teaching Inferential Statistics for Mathematic students (degree): tutorials and practical courses. (Splus and R software).

Teaching Statistics in Master in Applied Mathematics and Social Sciences (MASS) : tutorials and practical courses. (Splus software).

Lecture and tutorials in Discriminant analysis and Scoring (licence MASS).

Practical courses on Statistics, MASS (SAS software).

Supervision a Master trainee in Biostatistics.

2002 - 2003 Lecturer Bordeaux 1 University (96h)

Teaching Analysis and Algebra for Mathematics Students: tutorials.

Teaching Probability and Statistics in Mathematics Master.

Practical courses in Master MSRO (Stochastic Modelling and Operational Research).

Supervision of Master MSRO trainees.

2000 - 2002 Lecturer at Bordeaux 2 University (2x64h)

Statistics for first year medical students.

Lectures and practical courses in Informatics, Master of of Public Health year 1 : database and Access software.

Computer Science for medical student: Excel software, Windows, network, word.

1999 - 2000 Assistant Lecturer assistant at Bordeaux 2 University (96h)

Tutorials on Probability, algebra, statistic and informatics for biology Students.

Practical courses in Statistic for Biology students: SPSS software.

Practical courses in Informatics for Sport students: Word and Excel software.