

RESEARCH CENTRE

**Inria Center
at Université Grenoble Alpes**

IN PARTNERSHIP WITH:

CNRS, Université de Grenoble Alpes

2022

ACTIVITY REPORT

Project-Team

STATIFY

**Bayesian and extreme value statistical
models for structured and high
dimensional data**

IN COLLABORATION WITH: Laboratoire Jean Kuntzmann (LJK)

DOMAIN

**Applied Mathematics, Computation and
Simulation**

THEME

**Optimization, machine learning and
statistical methods**

Inria

Contents

Project-Team STATIFY	1
1 Team members, visitors, external collaborators	2
2 Overall objectives	3
3 Research program	4
3.1 Mixture models	4
3.2 Graphical and Markov models	5
3.3 Functional Inference, semi- and non-parametric methods	5
3.3.1 Modelling extremal events	6
3.3.2 Level sets estimation	6
3.3.3 Dimension reduction	7
4 Application domains	7
4.1 Image Analysis	7
4.2 Biology, Environment and Medicine	7
5 Social and environmental responsibility	8
5.1 Footprint of research activities	8
5.2 Impact of research results	8
6 Highlights of the year	8
6.1 New projects	8
6.2 Outstanding papers	8
6.3 Awards	9
7 New software and platforms	9
7.1 New software	9
7.1.1 Planet-GLLiM	9
7.1.2 Kernelo	9
8 New results	10
8.1 Mixture models	10
8.1.1 An online Minorization-Maximization algorithm	10
8.1.2 Global implicit function theorems and the online expectation-maximisation algorithm	10
8.1.3 Concentration results for approximate Bayesian computation without identifiability	10
8.1.4 Online EM algorithm for robust clustering	11
8.1.5 A non-asymptotic approach for model selection via penalization in high-dimensional mixture of experts models	11
8.1.6 Bayesian mixture models (in)consistency for the number of clusters	11
8.1.7 Joint supervised classification and reconstruction of irregularly sampled satellite image times series	12
8.2 Semi and non-parametric methods	12
8.2.1 Extreme events and neural networks	12
8.2.2 Estimation of extreme risk measures	13
8.2.3 Conditional extremal events	13
8.2.4 Estimation of multivariate risk measures	14
8.2.5 Dimension reduction for extremes	14
8.2.6 Bayesian inference for extreme values	14
8.2.7 Diagnosing convergence of Markov chain Monte Carlo	15
8.2.8 Dimension reduction with Sliced Inverse Regression	15
8.2.9 Latent factor models: a tool for dimension reduction in joint species distribution models	15
8.3 Graphical and Markov models	16

8.3.1	A Pre-Screening Approach for Faster Bayesian Network Structure Learning	16
8.3.2	Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis.	16
8.3.3	Beyond Voxel Prediction Uncertainty: Identifying brain lesions you can trust.	17
8.3.4	Multi-Scale Evaluation of Uncertainty Quantification Techniques for Deep Learning based MRI Segmentation.	17
8.3.5	Improving Uncertainty-based Out-of-Distribution detection for medical image segmentation	17
8.3.6	Brain subtle anomaly detection based on Auto-Encoders latent space analysis: Application to <i>de novo</i> Parkinson patients	18
8.3.7	Bayesian nonparametric models for hidden Markov random fields on count variables and application to traffic accidents	18
8.3.8	Hidden Markov models for the analysis of eye movements	19
8.3.9	Estimation of leaf area densities in tropical forests	19
8.3.10	Bayesian neural networks	20
8.3.11	Graph comparisons	20
8.3.12	Spatio-temporal data	21
8.4	Inverse problems	21
8.4.1	Sequential Bayesian experimental design for inverse problems.	21
8.4.2	Bayesian nonparametric mixture of experts for high-dimensional inverse problems.	22
8.4.3	Hierarchical Bayesian models for simulation-based inference	22
8.4.4	Bayesian inference on a large-scale brain simulator	22
9	Bilateral contracts and grants with industry	23
9.1	Bilateral contracts with industry	23
10	Partnerships and cooperations	23
10.1	International initiatives	23
10.1.1	Inria associate team not involved in an IIL or an international program	23
10.2	International research visitors	24
10.2.1	Visits of international scientists	24
10.2.2	Visits to international teams	24
10.3	National initiatives	25
10.3.1	Networks	26
11	Dissemination	26
11.1	Promoting scientific activities	26
11.1.1	Scientific events: organisation	26
11.1.2	Scientific events: selection	27
11.1.3	Journal	27
11.1.4	Invited talks	28
11.1.5	Research administration	28
11.2	Teaching - Supervision - Juries	28
11.2.1	Teaching	28
11.2.2	Supervision	29
11.2.3	Juries	30
11.3	Popularization	30
11.3.1	Articles and contents	30
12	Scientific production	30
12.1	Major publications	30
12.2	Publications of the year	31
12.3	Other	36
12.4	Cited publications	36

Project-Team STATIFY

Creation of the Project-Team: 2020 April 01

Keywords

Computer sciences and digital sciences

- A3.1.1. – Modeling, representation
- A3.1.4. – Uncertain data
- A3.3.2. – Data mining
- A3.3.3. – Big data analysis
- A3.4.1. – Supervised learning
- A3.4.2. – Unsupervised learning
- A3.4.4. – Optimization and learning
- A3.4.5. – Bayesian methods
- A3.4.7. – Kernel methods
- A5.3.3. – Pattern recognition
- A5.9.2. – Estimation, modeling
- A6.2. – Scientific computing, Numerical Analysis & Optimization
- A6.2.3. – Probabilistic methods
- A6.2.4. – Statistical methods
- A6.3. – Computation-data interaction
- A6.3.1. – Inverse problems
- A6.3.3. – Data processing
- A6.3.5. – Uncertainty Quantification
- A9.2. – Machine learning
- A9.3. – Signal analysis

Other research topics and application domains

- B1.2.1. – Understanding and simulation of the brain and the nervous system
- B2.6.1. – Brain imaging
- B3.3. – Geosciences
- B3.4.1. – Natural risks
- B3.4.2. – Industrial risks and waste
- B3.5. – Agronomy
- B5.1. – Factory of the future
- B9.5.6. – Data science
- B9.11.1. – Environmental risks

1 Team members, visitors, external collaborators

Research Scientists

- Florence Forbes [Team leader, INRIA, Senior Researcher, HDR]
- Sophie Achard [CNRS, Senior Researcher, HDR]
- Julyan Arbel [INRIA, Researcher, HDR]
- Pedro Luiz Coelho Rodrigues [INRIA, Researcher]
- Stephane Girard [INRIA, Senior Researcher, HDR]
- Pierre Wolinski [UNIV OXFORD, Starting Research Position, from Oct 2022]

Faculty Members

- Jean-Baptiste Durand [GRENOBLE INP, Associate Professor]
- Olivier Francois [GRENOBLE INP, Professor]

Post-Doctoral Fellows

- Jhouben Janyk Cuesta Ramirez [Inria, from Oct 2022]
- Trung Tin Nguyen [Inria]
- Konstantinos Pitas [INRIA, from Mar 2022]

PhD Students

- Louise Alamichel [UGA]
- Yuchen Bai [UGA]
- Daria Bystrova [UGA]
- Lucrezia Carboni [UGA]
- Jacopo Iollo [INRIA]
- Benjamin Lambert [PIXYL, from Feb 2022]
- Hana Lbath [UGA]
- Minh Le [Invensense, CIFRE]
- Theo Moins [INRIA]
- Geoffroy Oudoumanessah [INSERM, from Oct 2022]

Technical Staff

- Pascal Dkengne Sielenou [INRIA, Engineer]

Interns and Apprentices

- Geoffroy Oudoumanessah [INSERM, from Apr 2022 until Sep 2022]

Administrative Assistant

- Marion Ponsot [INRIA]

Visiting Scientists

- Filippo Ascolani [Bocconi University, from Nov 2022 until Nov 2022]
- Alberto Gonzales Sanz [University of Toulouse, from Nov 2022]
- Hien Nguyen [UNIV QUEENSLAND, from Jul 2022]
- Kai Qin [UNIV SWINBURNE, from Jul 2022]

External Collaborator

- Jonathan El Methni [UNIV PARIS]

2 Overall objectives

The STATIFY team focuses on statistics. Statistics can be defined as a science of variation where the main question is how to acquire knowledge in the face of variation. In the past, statistics were seen as an opportunity to play in various backyards. Today, the statistician sees his own backyard invaded by data scientists, machine learners and other computer scientists of all kinds. Everyone wants to do data analysis and some (but not all) do it very well. Generally, data analysis algorithms and associated network architectures are empirically validated using domain-specific datasets and data challenges. While winning such challenges is certainly rewarding, statistical validation lies on more fundamentally grounded bases and raises interesting theoretical, algorithmic and practical insights. Statistical questions can be converted to probability questions by the use of probability models. Once certain assumptions about the mechanisms generating the data are made, statistical questions can be answered using probability theory. However, the proper formulation and checking of these probability models is just as important, or even more important, than the subsequent analysis of the problem using these models. The first question is then how to formulate and evaluate probabilistic models for the problem at hand. The second question is how to obtain answers after a certain model has been assumed. This latter task can be more a matter of applied probability theory, and in practice, contains optimization and numerical analysis.

The STATIFY team aims at bringing strengths, at a time when the number of solicitations received by statisticians increases considerably because of the successive waves of *big data*, *data science* and *deep learning*. The difficulty is to back up our approaches with reliable mathematics while what we have is often only empirical observations that we are not able to explain. Guiding data analysis with statistical justification is a challenge in itself. STATIFY has the ambition to play a role in this task and to provide answers to questions about the appropriate usage of statistics.

Often statistical assumptions do not hold. Under what conditions then can we use statistical methods to obtain reliable knowledge? These conditions are rarely the natural state of complex systems. The central motivation of STATIFY is to establish the conditions under which statistical assumptions and associated inference procedures approximately hold and become reliable.

However, as George Box said "Statisticians and artists both suffer from being too easily in love with their models". To moderate this risk, we choose to develop, in the team, expertise from different statistical domains to offer different solutions to attack a variety of problems. This is possible because these domains share the same mathematical food chain, from probability and measure theory to statistical modeling, inference and data analysis.

Our goal is to exploit methodological resources from statistics and machine learning to develop models that handle variability and that scale to high dimensional data while maintaining our ability to assess their correctness, typically the uncertainty associated with the provided solutions. To reach this goal, the team offers a unique range of expertise in statistics, combining probabilistic graphical models and mixture models to analyze structured data, Bayesian analysis to model knowledge and regularize ill-posed problems, non-parametric statistics, risk modeling and extreme value theory to face the lack, or

impossibility, of precise modeling information and data. In the team, this expertise is organized to target five key challenges:

1. *Models for high dimensional, multimodal, heterogeneous data;*
2. *Spatial (structured) data science;*
3. *Scalable Bayesian models and procedures;*
4. *Understanding mathematical properties of statistical and machine learning methods;*
5. *The big problem of small data.*

The first two challenges address sources of complexity coming from data, namely, the fact that observations can be: 1) high dimensional, collected from multiple sensors in varying conditions *i.e.* multimodal and heterogeneous and 2) inter-dependent with a known structure between variables or with unknown interactions to be discovered. The other three challenges focus on providing reliable and interpretable models: 3) making the Bayesian approach scalable to handle large and complex data; 4) quantifying the information processing properties of machine learning methods and 5) allowing to draw reliable conclusions from datasets that are too small or not large enough to be used for training machine/deep learning methods.

These challenges rely on our four research axes:

1. *Models for graphs and networks;*
2. *Dimension reduction and latent variable modeling;*
3. *Bayesian modeling;*
4. *Modeling and quantifying extreme risk.*

In terms of applied work, we will target high-impact applications in neuroimaging, environmental and earth sciences.

3 Research program

3.1 Mixture models

Participants: Jean-Baptiste Durand, Florence Forbes, Stephane Girard, Julyan Arbel, Olivier Francois, Daria Bystrova, Giovanni Poggiato, Geoffrey Oudoumanessah, Louise Alamichel.

Keywords: Key-words: mixture of distributions, EM algorithm, missing data, conditional independence, statistical pattern recognition, clustering, unsupervised and partially supervised learning..

In a first approach, we consider statistical parametric models, θ being the parameter, possibly multi-dimensional, usually unknown and to be estimated. We consider cases where the data naturally divides into observed data $y = \{y_1, \dots, y_n\}$ and unobserved or missing data $z = \{z_1, \dots, z_n\}$. The missing data z_i represents for instance the memberships of one of a set of K alternative categories. The distribution of an observed y_i can be written as a finite mixture of distributions,

$$f(y_i; \theta) = \sum_{k=1}^K P(z_i = k; \theta) f(y_i | z_i; \theta). \quad (1)$$

These models are interesting in that they may point out hidden variables responsible for most of the observed variability and so that the observed variables are *conditionally* independent. Their estimation is often difficult due to the missing data. The Expectation-Maximization (EM) algorithm is a general and now standard approach to maximization of the likelihood in missing data problems. It provides parameter estimation but also values for missing data.

Mixture models correspond to independent z_i 's. They have been increasingly used in statistical pattern recognition. They enable a formal (model-based) approach to (unsupervised) clustering.

3.2 Graphical and Markov models

Participants: Jean-Baptiste Durand, Florence Forbes, Julyan Arbel, Sophie Achard, Olivier Francois, Mariia Vladimirova, Lucrezia Carboni, Hana Lbath, Minh-tri Le, Yuchen Bai.

Keywords: Key-words: graphical models, Markov properties, hidden Markov models, clustering, missing data, mixture of distributions, EM algorithm, image analysis, Bayesian inference..

Graphical modelling provides a diagrammatic representation of the dependency structure of a joint probability distribution, in the form of a network or graph depicting the local relations among variables. The graph can have directed or undirected links or edges between the nodes, which represent the individual variables. Associated with the graph are various Markov properties that specify how the graph encodes conditional independence assumptions.

It is the conditional independence assumptions that give graphical models their fundamental modular structure, enabling computation of globally interesting quantities from local specifications. In this way graphical models form an essential basis for our methodologies based on structures.

The graphs can be either directed, e.g. Bayesian Networks, or undirected, e.g. Markov Random Fields. The specificity of Markovian models is that the dependencies between the nodes are limited to the nearest neighbor nodes. The neighborhood definition can vary and be adapted to the problem of interest. When parts of the variables (nodes) are not observed or missing, we refer to these models as Hidden Markov Models (HMM). Hidden Markov chains or hidden Markov fields correspond to cases where the z_i 's in (1) are distributed according to a Markov chain or a Markov field. They are a natural extension of mixture models. They are widely used in signal processing (speech recognition, genome sequence analysis) and in image processing (remote sensing, MRI, etc.). Such models are very flexible in practice and can naturally account for the phenomena to be studied.

Hidden Markov models are very useful in modelling spatial dependencies but these dependencies and the possible existence of hidden variables are also responsible for a typically large amount of computation. It follows that the statistical analysis may not be straightforward. Typical issues are related to the neighborhood structure to be chosen when not dictated by the context and the possible high dimensionality of the observations. This also requires a good understanding of the role of each parameter and methods to tune them depending on the goal in mind. Regarding estimation algorithms, they correspond to an energy minimization problem which is NP-hard and usually performed through approximation. We focus on a certain type of methods based on variational approximations and propose effective algorithms which show good performance in practice and for which we also study theoretical properties. We also propose some tools for model selection. Eventually we investigate ways to extend the standard Hidden Markov Field model to increase its modelling power.

3.3 Functional Inference, semi- and non-parametric methods

Participants: Julyan Arbel, Daria Bystrova, Giovanni Poggiato, Stephane Girard, Florence Forbes, Pedro Coelho Rodrigues, Pascal Dkengne Sielenou, Meryem Bousebata, Theo Moins, Pierre Wolinski, Sophie Achard

Keywords: Key-words: dimension reduction, extreme value analysis, functional estimation..

We also consider methods which do not assume a parametric model. The approaches are non-parametric in the sense that they do not require the assumption of a prior model on the unknown quantities. This property is important since, for image applications for instance, it is very difficult to introduce sufficiently general parametric models because of the wide variety of image contents. Projection methods are then a way to decompose the unknown quantity on a set of functions (e.g. wavelets). Kernel methods which rely on smoothing the data using a set of kernels (usually probability distributions) are other examples. Relationships exist between these methods and learning techniques using Support

Vector Machine (SVM) as this appears in the context of *level-sets estimation* (see section 3.3.2). Such non-parametric methods have become the cornerstone when dealing with functional data [88]. This is the case, for instance, when observations are curves. They enable us to model the data without a discretization step. More generally, these techniques are of great use for *dimension reduction* purposes (section 3.3.3). They enable reduction of the dimension of the functional or multivariate data without assumptions on the observations distribution. Semi-parametric methods refer to methods that include both parametric and non-parametric aspects. Examples include the Sliced Inverse Regression (SIR) method [90] which combines non-parametric regression techniques with parametric dimension reduction aspects. This is also the case in *extreme value analysis* [87], which is based on the modelling of distribution tails (see section 3.3.1). It differs from traditional statistics which focuses on the central part of distributions, *i.e.* on the most probable events. Extreme value theory shows that distribution tails can be modelled by both a functional part and a real parameter, the extreme value index.

3.3.1 Modelling extremal events

Extreme value theory is a branch of statistics dealing with the extreme deviations from the bulk of probability distributions. More specifically, it focuses on the limiting distributions for the minimum or the maximum of a large collection of random observations from the same arbitrary distribution. Let $X_{1,n} \leq \dots \leq X_{n,n}$ denote n ordered observations from a random variable X representing some quantity of interest. A p_n -quantile of X is the value x_{p_n} such that the probability that X is greater than x_{p_n} is p_n , *i.e.* $P(X > x_{p_n}) = p_n$. When $p_n < 1/n$, such a quantile is said to be extreme since it is usually greater than the maximum observation $X_{n,n}$.

To estimate such quantiles therefore requires dedicated methods to extrapolate information beyond the observed values of X . Those methods are based on Extreme value theory. This kind of issue appeared in hydrology. One objective was to assess risk for highly unusual events, such as 100-year floods, starting from flows measured over 50 years. To this end, semi-parametric models of the tail are considered:

$$P(X > x) = x^{-1/\theta} \ell(x), \quad x > x_0 > 0, \quad (2)$$

where both the extreme-value index $\theta > 0$ and the function $\ell(x)$ are unknown. The function ℓ is a slowly varying function *i.e.* such that

$$\frac{\ell(tx)}{\ell(x)} \rightarrow 1 \text{ as } x \rightarrow \infty \quad (3)$$

for all $t > 0$. The function $\ell(x)$ acts as a nuisance parameter which yields a bias in the classical extreme-value estimators developed so far. Such models are often referred to as heavy-tail models since the probability of extreme events decreases at a polynomial rate to zero. It may be necessary to refine the model (2,3) by specifying a precise rate of convergence in (3). To this end, a second order condition is introduced involving an additional parameter $\rho \leq 0$. The larger ρ is, the slower the convergence in (3) and the more difficult the estimation of extreme quantiles.

More generally, the problems that we address are part of the risk assessment theory. For instance, in reliability, the distributions of interest are included in a semi-parametric family whose tails are decreasing exponentially fast. These so-called Weibull-tail distributions [8] are defined by their survival distribution function:

$$P(X > x) = \exp\{-x^\theta \ell(x)\}, \quad x > x_0 > 0. \quad (4)$$

Gaussian, gamma, exponential and Weibull distributions, among others, are included in this family. An important part of our work consists in establishing links between models (2) and (4) in order to propose new estimation methods. We also consider the case where the observations were recorded with a covariate information [9]. In this case, the extreme-value index and the p_n -quantile are functions of the covariate. We propose estimators of these functions by using moving window approaches, nearest neighbor methods, or kernel estimators.

3.3.2 Level sets estimation

Level sets estimation is a recurrent problem in statistics which is linked to outlier detection. In biology, one is interested in estimating reference curves, that is to say curves which bound 90% (for example) of the

population. Points outside this bound are considered as outliers compared to the reference population. Level sets estimation can be looked at as a conditional quantile estimation problem which benefits from a non-parametric statistical framework. In particular, boundary estimation, arising in image segmentation as well as in supervised learning, is interpreted as an extreme level set estimation problem. Level sets estimation can also be formulated as a linear programming problem. In this context, estimates are sparse since they involve only a small fraction of the dataset, called the set of support vectors.

3.3.3 Dimension reduction

Our work on high dimensional data requires that we face the curse of dimensionality phenomenon. Indeed, the modelling of high dimensional data requires complex models and thus the estimation of high number of parameters compared to the sample size. In this framework, dimension reduction methods aim at replacing the original variables by a small number of linear combinations with as small as possible loss of information. Principal Component Analysis (PCA) is the most widely used method to reduce dimension in data. However, standard linear PCA can be quite inefficient on image data where even simple image distortions can lead to highly non-linear data. Two directions are investigated. First, non-linear PCAs can be proposed, leading to semi-parametric dimension reduction methods [89]. Another field of investigation is to take into account the application goal in the dimension reduction step. One of our approaches is therefore to develop new Gaussian models of high dimensional data for parametric inference [1]. Such models can then be used in a Mixtures or Markov framework for classification purposes. Another approach consists in combining dimension reduction, regularization techniques, and regression techniques to improve the Sliced Inverse Regression method [90].

4 Application domains

4.1 Image Analysis

Participants: Florence Forbes, Jean-Baptiste Durand, Stephane Girard, Pedro Coelho Rodrigues, Geoffroy Oudoumanessah, Sophie Achard.

As regards applications, several areas of image analysis can be covered using the tools developed in the team. More specifically, in collaboration with team PERCEPTION, we address various issues in computer vision involving Bayesian modelling and probabilistic clustering techniques. Other applications in medical imaging are natural. We work more specifically on MRI and functional MRI data, in collaboration with the Grenoble Institute of Neuroscience (GIN). We also consider other statistical 2D fields coming from other domains such as remote sensing, in collaboration with the Institut de Planétologie et d'Astrophysique de Grenoble (IPAG) and the Centre National d'Etudes Spatiales (CNES). In this context, we worked on hyperspectral and/or multitemporal images. In the context of the "pole de compétitivité" project I-VP, we worked on images of PC Boards.

4.2 Biology, Environment and Medicine

Participants: Florence Forbes, Stephane Girard, Jean-Baptiste Durand, Julyan Arbel, Sophie Achard, Pedro Coelho Rodrigues, Olivier Francois, Yuchen Bai, Theo Moins, Daria Bystrova, Meryem Bousebata, Lucrezia Carboni, Hana Lbath.

A third domain of applications concerns biology and medicine. We considered the use of mixture models to identify biomarkers. We also investigated statistical tools for the analysis of fluorescence signals in molecular biology. Applications in neurosciences are also considered. In the environmental domain, we considered the modelling of high-impact weather events and the use of hyperspectral data as a new tool for quantitative ecology.

5 Social and environmental responsibility

5.1 Footprint of research activities

The footprint of our research activities has not been assessed yet. Most of the team members have validated the "charte d'éco-responsabilité" written by a working group from Laboratoire Jean Kuntzmann, which should have practical implications in the near future.

5.2 Impact of research results

A lot of our developments are motivated by and target applications in medicine and environmental sciences. As such they have a social impact with a better handling and treatment of patients, in particular with brain diseases or disorders. On the environmental side, our work has an impact on geoscience-related decision making with e.g. extreme events risk analysis, planetary science studies and tools to assess biodiversity markers. However, how to truly measure and report this impact in practice is another question we have not really addressed yet.

6 Highlights of the year

6.1 New projects

- Australian Research Council Discovery project (2023-25): F. Forbes is coPI of a recently (Nov. 2022) funded project for 3 years.
- ANR project PEG2 (2022-26) on Predictive Ecological Genomics: STATISFY is involved in this 4-year project recently accepted in July 2022. The PI is prof. Olivier Francois who spent 2 years (2021-22) in the team on a *Delegation* position.
- Plan de Relance project with GE Healthcare (2022-24). The topic of the collaboration is related to early anomaly detection of failures in medical transducer manufacturing.
- MIAI: Pedro Rodrigues was awarded 13K euros for his project on *Machine Learning for Experimental Data* via the IRGA 2022 call for projects.
- MIAI: F. Forbes, J-B. Durand and Y. Bai were awarded 4.5K euros for their project on *Developing IA for estimating leaf area density area in tropical forests from LiDAR data*, Nov. 2022.
- A new collaboration has recently started with the PhD co-supervision of Julien Zhou, involving J. Arbel, P. Gaillard from Inria THOTH team, and T. Rahier from Criteo Grenoble. This PhD will address bandit problems from a Bayesian viewpoint.
- ANR project NODAL (2022-2026) « Network-based biomarker discovery of neurodegenerative diseases using multimodal connectivity ». Sophie Achard is involved in this projet where the PI is in Rennes Julie Coloigner
- ANR DynaSTI : « modélisation de la dynamique spatio-temporelle de la connectivité fonctionnelle en IRMF de repos. » Sophie Achard is involved in this projet where the PI is in Strasbourg Céline Meillier

6.2 Outstanding papers

A recent work has been accepted in Journal of Machine Learning Research, which is one of the top journals in the domain [13].

6.3 Awards

M. Vladimirova received the best poster award for her work at the *BAYesian Young Statisticians Meeting* in Université de Montréal, Canada, in June 22-23, 2022.

H. Lbath received the student award for her work on Clusted-based inter-regional correlation estimation presented at *2022 IMS International Conference on Statistics and Data Science (ICSDS)* December 13-16, 2022, Florence, Italy, [56].

7 New software and platforms

7.1 New software

7.1.1 Planet-GLLiM

Name: Planet-GLLiM

Keyword: Inverse problem

Functional Description: The application implements the GLLiM statistical learning technique in its different variants for the inversion of a physical model of reflectance on spectro-(gonio)-photometric data. The latter are of two types: 1. laboratory measurements of reflectance spectra acquired according to different illumination and viewing geometries, 2. and 4D spectro-photometric remote sensing products from multi-angular CRISM or Pléiades acquisitions.

URL: <https://gitlab.inria.fr/kernelo-mistis/planet-gllim-front-end/-/wikis/Home>

Publications: [insu-03705153](#), [hal-02908364](#)

Contact: Sylvain Douté

Participants: Florence Forbes, Benoit Kugler, Sami Djouadi, Samuel Heidmann, Stanislaw Borkowski

Partner: Institut de Planétologie et d'Astrophysique de Grenoble

7.1.2 Kernelo

Name: Kernelo-GLLiM

Keywords: Inverse problem, Clustering, Regression, Gaussian mixture, Python, C++

Scientific Description: Building a regression model for the purpose of prediction is widely used in all disciplines. A large number of applications consists of learning the association between responses and predictors and focusing on predicting responses for the newly observed samples. In this work, we go beyond simple linear models and focus on predicting low-dimensional responses using high-dimensional covariates when the associations between responses and covariates are non-linear.

Functional Description: Kernelo-GLLiM is a Gaussian Locally-Linear Mapping (GLLiM) solver. Kernelo-GLLiM provides a C++ library and a python module for non linear mapping (non linear regression) using a mixture of regression model and an inverse regression strategy. The methods include the GLLiM model (Deleforge et al (2015)) based on Gaussian mixtures.

URL: <https://gitlab.inria.fr/kernelo-mistis/kernelo-gllim-is/-/wikis/home>

Publications: [hal-00863468](#), [hal-02908364](#)

Contact: Florence Forbes

Participants: Florence Forbes, Benoit Kugler, Sami Djouadi, Samuel Heidmann, Stanislaw Borkowski

Partner: Institut de Planétologie et d'Astrophysique de Grenoble

8 New results

8.1 Mixture models

8.1.1 An online Minorization-Maximization algorithm

Participants: Florence Forbes.

Joint work with: Hien Nguyen, University of Queensland, Brisbane Australia, Gersende Fort, IMT Toulouse and Olivier Cappé ENS Paris.

Modern statistical and machine learning settings often involve high data volume and data streaming, which require the development of online estimation algorithms. The online Expectation-Maximization (EM) algorithm extends the popular EM algorithm to this setting, via a stochastic approximation approach. We show that an online version of the Minorization-Maximization (MM) algorithm, which includes the online EM algorithm as a special case, can also be constructed in a similar manner. We demonstrate our approach via an application to the logistic regression problem and compare it to existing methods [36].

8.1.2 Global implicit function theorems and the online expectation-maximisation algorithm

Participants: Florence Forbes.

Joint work with: Hien Nguyen, University of Queensland, Brisbane Australia.

Due to the changing nature of data, online and mini-batch variants of EM and EM-like algorithms have become increasingly popular. The consistency of the estimator sequences that are produced by these EM variants often rely on an assumption regarding the continuous differentiability of a parameter update function. In many cases, the parameter update function is often not in closed form and may only be defined implicitly, which makes the verification of the continuous differentiability property difficult. We demonstrate how a global implicit function theorem can be used to verify such properties in the cases of finite mixtures of distributions in the exponential family and more generally when the component specific distribution admits a data augmentation scheme in the exponential family. We demonstrate the use of such a theorem in the case of mixtures of beta distributions, gamma distributions, fully-visible Boltzmann machines and Student distributions. Via numerical simulations, we provide empirical evidence towards the consistency of the online EM algorithm parameter estimates in such cases [27].

8.1.3 Concentration results for approximate Bayesian computation without identifiability

Participants: Florence Forbes, Julyan Arbel, Trung Tin Nguyen.

Joint work with: Hien Nguyen, University of Queensland, Brisbane Australia.

We study the large sample behaviors of approximate Bayesian computation (ABC) posterior measures in situations when the data generating process is dependent on non-identifiable parameters. In particular, we establish the concentration of posterior measures on sets of arbitrarily small size that contain the equivalence set of the data generative parameter, when the sample size tends to infinity. Our theory also makes weak assumptions regarding the measurement of discrepancy between the data set and simulations, and in particular, does not require the use of summary statistics and is applicable to a broad class of kernelized ABC algorithms. We provide useful illustrations and demonstrations of our theory in

practice, and offer a comprehensive assessment of the nature in which our findings complement other results in the literature.

8.1.4 Online EM algorithm for robust clustering

Participants: Florence Forbes, Geoffroy Oudoumanessah.

Joint work with: Hien Nguyen, University of Queensland, Australia and Michel Dojat Grenoble Institute of Neurosciences.

A popular way to approach clustering tasks is via a parametric mixture model. The vast majority of the work on such mixtures has been based on Gaussian mixture models. However, in some applications the tails of Gaussian distributions are shorter than appropriate or parameter estimations are affected by atypical observations (outliers). To address this issue, mixtures of so-called multiple scale Student distributions have been proposed and used for clustering. In contrast to the Gaussian case, no closed-form solution exists for such mixtures but tractability is maintained via the use of the expectation-maximisation (EM) algorithm. However, such mixtures require more parameters than the standard Student or Gaussian mixtures and the EM algorithm used to estimate the mixture parameters involves more complex numerical optimizations. Consequently, when the number of samples to be clustered becomes large, applying EM on the whole data set (Batch EM) may become costly both in terms of time and memory requirements. A natural approach to bypass this issue is to consider an online version of the algorithm, that can incorporate the samples incrementally or in mini batches. In this work, we proposed to design a tractable online EM for mixtures of multiple scale Student distributions in order to use it then to detect subtle brain anomalies from MR brain scans for patients suffering from Parkinson disease. The application to Parkinson disease will be carried out jointly with Grenoble Institute of Neuroscience.

8.1.5 A non-asymptotic approach for model selection via penalization in high-dimensional mixture of experts models

Participants: Florence Forbes, Trung Tin Nguyen.

Joint work with: Faicel Chamroukhi, University of Caen and Hien Nguyen, University of Queensland, Brisbane, Australia.

Mixture of experts (MoE) are a popular class of statistical and machine learning models that have gained attention over the years due to their flexibility and efficiency. In this work, we consider Gaussian-gated localized MoE (GLoME) and block-diagonal covariance localized MoE (BLoME) regression models to present nonlinear relationships in heterogeneous data with potential hidden graph-structured interactions between high-dimensional predictors. These models pose difficult statistical estimation and model selection questions, both from a computational and theoretical perspective. This paper [84] is devoted to the study of the problem of model selection among a collection of GLoME or BLoME models characterized by the number of mixture components, the complexity of Gaussian mean experts, and the hidden block-diagonal structures of the covariance matrices, in a penalized maximum likelihood estimation framework. In particular, we establish non-asymptotic risk bounds that take the form of weak oracle inequalities, provided that lower bounds for the penalties hold. The good empirical behavior of our models is then demonstrated on synthetic and real datasets.

8.1.6 Bayesian mixture models (in)consistency for the number of clusters

Participants: Julyan Arbel, Louise Alamichel, Daria Bystrova.

Joint work with: Guillaume Kon Kam King (INRAE).

Bayesian nonparametric mixture models are common for modeling complex data. While these models are well-suited for density estimation, their application for clustering has some limitations. Recent results proved posterior inconsistency of the number of clusters when the true number of clusters is finite for the Dirichlet process and Pitman–Yor process mixture models. In [68], we extend these results to additional Bayesian nonparametric priors such as Gibbs-type processes and finite-dimensional representations thereof. The latter include the Dirichlet multinomial process, the recently proposed Pitman–Yor, and normalized generalized gamma multinomial processes. We show that mixture models based on these processes are also inconsistent in the number of clusters and discuss possible solutions. Notably, we show that a post-processing algorithm introduced for the Dirichlet process can be extended to more general models and provides a consistent method to estimate the number of components.

8.1.7 Joint supervised classification and reconstruction of irregularly sampled satellite image times series

Participants: Alexandre Constantin, Stephane Girard.

Joint work with: Mathieu Fauvel, INRAE

Recent satellite missions have led to a huge amount of earth observation data, most of them being freely available. In such a context, satellite image time series have been used to study land use and land cover information. However, optical time series, like Sentinel-2 or Landsat ones, are provided with an irregular time sampling for different spatial locations, and images may contain clouds and shadows. Thus, pre-processing techniques are usually required to properly classify such data. The proposed approach is able to deal with irregular temporal sampling and missing data directly in the classification process. It is based on Gaussian processes and allows to perform jointly the classification of the pixel labels as well as the reconstruction of the pixel time series. The method complexity scales linearly with the number of pixels, making it amenable in large scale scenarios. Experimental classification and reconstruction results show that the method does not compete yet with state of the art classifiers but yields reconstructions that are robust with respect to the presence of undetected clouds or shadows and does not require any temporal preprocessing [16].

8.2 Semi and non-parametric methods

8.2.1 Extreme events and neural networks

Participants: Stephane Girard.

Joint work with: M. Allouche and E. Gobet (CMAP, Ecole Polytechnique)

Feedforward neural networks based on Rectified linear units (ReLU) cannot efficiently approximate quantile functions which are not bounded, especially in the case of heavy-tailed distributions. We thus propose a new parametrization for the generator of a Generative adversarial network (GAN) adapted to this framework, basing on extreme-value theory. We provide an analysis of the uniform error between the extreme quantile and its GAN approximation. It appears that the rate of convergence of the error is mainly driven by the second-order parameter of the data distribution. The above results are illustrated on simulated data and real financial data [13].

A similar investigation has been conducted to simulate fractional Brownian motion with ReLU neural networks [12].

In [69], we propose new parametrizations for neural networks in order to estimate extreme quantiles in both non-conditional and conditional heavy-tailed settings. All proposed neural network estimators feature a bias correction based on an extension of the usual second-order condition to an arbitrary order. The convergence rate of the uniform error between extreme log-quantiles and their neural network

approximation is established. The finite sample performances of the non-conditional neural network estimator are compared to other bias-reduced extreme-value competitors on simulated data. It is shown that our method outperforms them in difficult heavy-tailed situations where other estimators almost all fail. Finally, the conditional neural network estimators are implemented to investigate the behaviour of extreme rainfalls as functions of their geographical location in the southern part of France. The results are submitted for publication.

8.2.2 Estimation of extreme risk measures

Participants: Stephane Girard.

Joint work with: J. El-Methni (Univ. Paris Cité), G. Stupfler (Univ. Angers) and A. Usseglio-Carleve (Univ. Avignon).

One of the most popular risk measures is the Value-at-Risk (VaR) introduced in the 1990's. In statistical terms, the VaR at level $\alpha \in (0, 1)$ corresponds to the upper α -quantile of the loss distribution. Weissman extrapolation device for estimating extreme quantiles (when $\alpha \rightarrow 0$) from heavy-tailed distributions is based on two estimators: an order statistic to estimate an intermediate quantile and an estimator of the tail-index. The common practice is to select the same intermediate sequence for both estimators. In [11], we show how an adapted choice of two different intermediate sequences leads to a reduction of the asymptotic bias associated with the resulting refined Weissman estimator. This new bias reduction method is fully automatic and does not involve the selection of extra parameters. Our approach is compared to other bias reduced estimators of extreme quantiles both on simulated and real data.

The Value-at-Risk however suffers from several weaknesses. First, it provides us only with a pointwise information: $\text{VaR}(\alpha)$ does not take into consideration what the loss will be beyond this quantile. Second, random loss variables with light-tailed distributions or heavy-tailed distributions may have the same Value-at-Risk. Finally, Value-at-Risk is not a coherent risk measure since it is not subadditive in general. A first coherent alternative risk measure is the Conditional Tail Expectation (CTE), also known as Tail-Value-at-Risk, Tail Conditional Expectation or Expected Shortfall in case of a continuous loss distribution. The CTE is defined as the expected loss given that the loss lies above the upper α -quantile of the loss distribution. This risk measure thus takes into account the whole information contained in the upper tail of the distribution.

Risk measures of a financial position are, from an empirical point of view, mainly based on quantiles. Replacing quantiles with their least squares analogues, called expectiles, has recently received increasing attention [3]. The novel expectile-based risk measures satisfy all coherence requirements. Currently available estimators of extreme expectiles are typically biased and hence may show poor finite-sample performance even in fairly large samples. In [22], we focus on the construction of bias-reduced extreme expectile estimators for heavy-tailed distributions. The rationale for our construction hinges on a careful investigation of the asymptotic proportionality relationship between extreme expectiles and their quantile counterparts, as well as of the extrapolation formula motivated by the heavy-tailed context. We accurately quantify and estimate the bias incurred by the use of these relationships when constructing extreme expectile estimators. This motivates the introduction of a class of bias-reduced estimators whose asymptotic properties are rigorously shown, and whose finite-sample properties are assessed on a simulation study and three samples of real data from economics, insurance and finance.

8.2.3 Conditional extremal events

Participants: Stephane Girard.

Joint work with: G. Stupfler (Univ. Angers) and A. Usseglio-Carleve (Univ. Avignon).

As explained in Paragraph 8.2.2, expectiles have recently started to be considered as serious candidates to become standard tools in actuarial and financial risk management. However, expectiles and their sample versions do not benefit from a simple explicit form, making their analysis significantly harder than that of quantiles and order statistics. This difficulty is compounded when one wishes to integrate auxiliary information about the phenomenon of interest through a finite-dimensional covariate, in which case the problem becomes the estimation of conditional expectiles.

We exploit the fact that the expectiles of a distribution F are in fact the quantiles of another distribution E explicitly linked to F , in order to construct nonparametric kernel estimators of extreme conditional expectiles. We analyze the asymptotic properties of our estimators in the context of conditional heavy-tailed distributions. Applications to simulated data and real insurance data are provided [21]. The extension to functional covariates is investigated in [20].

8.2.4 Estimation of multivariate risk measures

Participants: Julyan Arbel, Stephane Girard.

Joint work with: H. Nguyen, University of Queensland, Brisbane, Australia and A. Usseglio-Carleve (Univ. Avignon).

Expectiles form a family of risk measures that have recently gained interest over the more common value-at-risk or return levels, primarily due to their capability to be determined by the probabilities of tail values and magnitudes of realisations at once. However, a prevalent and ongoing challenge of expectile inference is the problem of uncertainty quantification, which is especially critical in sensitive applications, such as in medical, environmental or engineering tasks. In [14], we address this issue by developing a novel distribution, termed the multivariate expectilebased distribution (MED), that possesses an expectile as a closed-form parameter. Desirable properties of the distribution, such as log-concavity, make it an excellent fitting distribution in multivariate applications. Maximum likelihood estimation and Bayesian inference algorithms are described. Simulated examples and applications to expectile and mode estimation illustrate the usefulness of the MED for uncertainty quantification.

8.2.5 Dimension reduction for extremes

Participants: Meryem Bousebata, Stephane Girard.

Joint work with: G. Enjolras (CERAG).

In the context of the PhD thesis of Meryem Bousebata, we propose a new approach, called Extreme-PLS, for dimension reduction in regression and adapted to distribution tails. The objective is to find linear combinations of predictors that best explain the extreme values of the response variable in a non-linear inverse regression model. The asymptotic normality of the Extreme-PLS estimator is established in the single-index framework and under mild assumptions. The performance of the method is assessed on simulated data. A statistical analysis of French farm income data, considering extreme cereal yields, is provided as an illustration [15].

8.2.6 Bayesian inference for extreme values

Participants: Julyan Arbel, Theo Moins, Stephane Girard.

Joint work with: A. Dutfoy (EDF R&D).

Combining extreme value theory with Bayesian methods offers several advantages, such as a quantification of uncertainty on parameter estimation or the ability to study irregular models that cannot be handled by frequentist statistics. However, it comes with many options that are left to the user concerning model building, computational algorithms, and even inference itself. Among them, the parameterization of the model induces a geometry that can alter the efficiency of computational algorithms, in addition to making calculations involved. In [75], we focus on the Poisson process characterization of extremes and outline two key benefits of an orthogonal parameterization addressing both issues. First, several diagnostics show that Markov chain Monte Carlo convergence is improved compared with the original parameterization. Second, orthogonalization also helps deriving Jeffreys and penalized complexity priors, and establishing posterior propriety. The analysis is supported by simulations, and our framework is then applied to extreme level estimation on river flow data. The results are submitted for publication.

8.2.7 Diagnosing convergence of Markov chain Monte Carlo

Participants: Julyan Arbel, Theo Moins, Stephane Girard.

Joint work with: A. Dutfoy (EDF R&D).

Diagnosing convergence of Markov chain Monte Carlo (MCMC) is crucial in Bayesian analysis. Among the most popular methods, the potential scale reduction factor (commonly named \hat{R}) is an indicator that monitors the convergence of output chains to a stationary distribution, based on a comparison of the between- and within-variance of the chains. Several improvements have been suggested since its introduction in the 90'ss. We analyse some properties of the theoretical value R associated to \hat{R} in the case of a localized version that focuses on quantiles of the distribution. This leads to proposing a new indicator, which is shown to allow both for localizing the MCMC convergence in different quantiles of the distribution, and at the same time for handling some convergence issues not detected by other \hat{R} versions. This work is submitted for publication [74].

8.2.8 Dimension reduction with Sliced Inverse Regression

Participants: Stephane Girard.

Joint work with: H. Lorenzo and J. Saracco (Inria Bordeaux Sud-Ouest).

Since its introduction in the early 90's, the Sliced Inverse Regression (SIR) methodology has evolved adapting to increasingly complex data sets in contexts combining linear dimension reduction with non linear regression. The assumption of dependence of the response variable with respect to only a few linear combinations of the covariates makes it appealing for many computational and real data application aspects. In [19], we propose an overview of the most active research directions in SIR modeling from multivariate regression models to regularization and variable selection.

8.2.9 Latent factor models: a tool for dimension reduction in joint species distribution models

Participant: Julyan Arbel, Daria Bystrova, Giovanni Poggiato.

Joint work with: Wilfried Thuiller, LECA - Laboratoire d'Ecologie Alpine.

We investigate modelling species distributions over space and time which is one of the major research topics in both ecology and conservation biology. Joint Species Distribution models (JSDMs) have recently been introduced as a tool to better model community data, by inferring a residual covariance matrix

between species, after accounting for species' response to the environment. However, these models are computationally demanding, even when latent factors, a common tool for dimension reduction, are used. To address this issue, previous research proposed to use a Dirichlet process, a Bayesian nonparametric prior, to further reduce model dimension by clustering species in the residual covariance matrix. Here, we built on this approach to include a prior knowledge on the potential number of clusters, and instead used a Pitman-Yor process to address some critical limitations of the Dirichlet process. We therefore propose a framework that includes prior knowledge in the residual covariance matrix, providing a tool to analyze clusters of species that share the same residual associations with respect to other species. We applied our methodology to a case study of plant communities in a protected area of the French Alps (the Bauges Regional Park), and demonstrated that our extensions improve dimension reduction and reveal additional information from the residual covariance matrix, notably showing how the estimated clusters are compatible with plant traits, endorsing their importance in shaping communities. A book chapter describing latent factor models as a tool for dimension reduction in joint species distribution models is also available.

8.3 Graphical and Markov models

8.3.1 A Pre-Screening Approach for Faster Bayesian Network Structure Learning

Participants: Florence Forbes.

Joint work with: Thibaud Rahier from Criteo and Sylvain Marié from Schneider.

Learning the structure of Bayesian networks from data is a NP-Hard problem that involves optimization over a super-exponential sized space. Still, in many real-life datasets a number of the arcs contained in the final structure correspond to strongly related pairs of variables and can be identified efficiently with information-theoretic metrics. In this work, we propose a meta-algorithm to accelerate any existing Bayesian network structure learning method. It contains an additional arc pre-screening step allowing to narrow the structure learning task down to a subset of the original variables, thus reducing the overall problem size. We conduct extensive experiments on both public benchmarks and private industrial datasets, showing that this approach enables a significant decrease in computational time and graph complexity for little to no decrease in performance score [37].

8.3.2 Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis.

Participants: Florence Forbes, Benjamin Lambert.

Joint work with: Senan Doyle, Alan Tucholka from Pixyl and Michel Dojat from Grenoble Institute of Neurosciences.

The full acceptance of Deep Learning (DL) models in the clinical field is rather low with respect to the quantity of high-performing solutions reported in the literature. Particularly, end users are reluctant to rely on the rough predictions of DL models. Uncertainty quantification methods have been proposed in the literature as a potential response to reduce the rough decision provided by the DL black box and thus increase the interpretability and the acceptability of the result by the final user. In this review, we propose an overview of the existing methods to quantify uncertainty associated to DL predictions. We focus on applications to medical image analysis, which present specific challenges due to the high dimensionality of images and their quality variability, as well as constraints associated to real-life clinical routine. We then discuss the evaluation protocols to validate the relevance of uncertainty estimates. Finally, we highlight the open challenges of uncertainty quantification in the medical field [72].

8.3.3 Beyond Voxel Prediction Uncertainty: Identifying brain lesions you can trust.

Participants: Florence Forbes, Benjamin Lambert.

Joint work with: Senan Doyle, Alan Tucholka from Pixyl and Michel Dojat from Grenoble Institute of Neurosciences.

Deep neural networks have become the gold-standard approach for the automated segmentation of 3D medical images. Their full acceptance by clinicians remains however hampered by the lack of intelligible uncertainty assessment of the provided results. Most approaches to quantify their uncertainty, such as the popular Monte Carlo dropout, restrict to some measure of uncertainty in prediction at the voxel level. In addition not to be clearly related to genuine medical uncertainty, this is not clinically satisfying as most objects of interest (e.g. brain lesions) are made of groups of voxels whose overall relevance may not simply reduce to the sum or mean of their individual uncertainties. In this work, we propose to go beyond voxel-wise assessment using an innovative Graph Neural Network approach, trained from the outputs of a Monte Carlo dropout model. This network allows the fusion of three estimators of voxel uncertainty: entropy, variance, and model's confidence; and can be applied to any lesion, regardless of its shape or size. We demonstrate the superiority of our approach for uncertainty estimate on a task of Multiple Sclerosis lesions segmentation [35].

8.3.4 Multi-Scale Evaluation of Uncertainty Quantification Techniques for Deep Learning based MRI Segmentation.

Participants: Florence Forbes, Benjamin Lambert.

Joint work with: Senan Doyle, Alan Tucholka from Pixyl and Michel Dojat from Grenoble Institute of Neurosciences.

Deep Learning (DL) techniques have become the gold standard for biomedical image segmentation. Although extensively used, they tend to be considered as black-boxes, preventing their full acceptance in clinical routine. This opacity is partly due to the inability of neural networks to express the uncertainty in their predictions. In recent years, tremendous work has been carried out to alleviate this limitation and develop models that know when they don't know 1. In this work, we propose an in-depth evaluation of 3 state-of-the-art approaches to quantify uncertainty attached to DL predictions : Monte- Carlo Dropout, Deep Ensemble, and Heteroscedastic network. We performed a multi-scale analysis of these techniques by evaluating uncertainty estimates at the voxel, lesion and image levels. We illustrate this comparison on an automatic segmentation task to detect White-Matter Hyperintensities (WMH) from T2-weighted FLAIR MRI sequences of multiple sclerosis patients [54].

8.3.5 Improving Uncertainty-based Out-of-Distribution detection for medical image segmentation

Participants: Florence Forbes, Benjamin Lambert.

Joint work with: Senan Doyle, Alan Tucholka from Pixyl and Michel Dojat from Grenoble Institute of Neurosciences.

Deep Learning models are easily disturbed by variations in the input images that were not seen during training, resulting in unpredictable behaviours. Such Out-of-Distribution (OOD) images represent a

significant challenge in the context of medical image analysis, where the range of possible abnormalities is extremely wide, including artifacts, unseen pathologies, or different imaging protocols. In this work, we evaluate various uncertainty frameworks to detect OOD inputs in the context of Multiple Sclerosis lesions segmentation. By implementing a comprehensive evaluation scheme including 14 sources of OOD of various nature and strength, we show that methods relying on the predictive uncertainty of binary segmentation models often fails in detecting outlying inputs. On the contrary, learning to segment anatomical labels alongside lesions highly improves the ability to detect OOD inputs.

8.3.6 Brain subtle anomaly detection based on Auto-Encoders latent space analysis: Application to *de novo* Parkinson patients

Participants: Florence Forbes, Geoffroy Oudoumanessah.

Joint work with: Michel Dojat from Grenoble Institute of Neurosciences, Carole Lartizien, Nicolas Pinon, Robin Trombetta from Creatis.

Neural network-based anomaly detection remains challenging in clinical applications with little or no supervised information and subtle anomalies such as hardly visible brain lesions. Among unsupervised methods, patch-based auto-encoders with their efficient representation power provided by their latent space, have shown good results for visible lesion detection. However, the commonly used reconstruction error criterion may limit their performance when facing less obvious lesions. In this work, we design two alternative detection criteria. They are derived from multivariate analysis and can more directly capture information from latent space representations. Their performance compares favorably with two additional supervised learning methods, on a difficult *de novo* Parkinson Disease (PD) classification task.

8.3.7 Bayesian nonparametric models for hidden Markov random fields on count variables and application to traffic accidents

Participants: Julyan Arbel, Jean-Baptiste Durand, Florence Forbes.

Joint work with: Hien Nguyen from University of Queensland Brisbane Australia and Fatoumata Dama, Laboratoire des Sciences du Numérique de Nantes, France

Hidden Markov random fields (HMRFs) have been widely used in image segmentation and more generally, for clustering of data indexed by graphs. Dependent hidden variables (states) represent the cluster identities and determine their interpretations. Dependencies between state variables are induced by the notion of neighborhood in the graph. A difficult and crucial problem in HMRFs is the identification of the number of possible states K . Recently, selection methods based on Bayesian non parametric priors (Dirichlet processes) have been developed. They do not assume that K is bounded a priori, thus allowing its adaptive selection with respect to the quantity of available data and avoiding costly systematic estimation and comparison of models with different fixed values for K . Our previous work [85] has focused on Bayesian nonparametric priors for HMRFs and continuous, Gaussian observations. In this work, we consider extensions to non-Gaussian observed data. A first case is discrete data, typically issued from counts. A second is exponential-distributed data. We defined and implemented Bayesian nonparametric models for HMRFs with Poisson- and exponential-distributed observations. Inference is achieved by Variational Bayesian Expectation Maximization (VBEM).

We proposed an application of the discrete-data model to a new risk mapping model for traffic accidents in the region of Victoria, Australia [17]. The partition into regions using labels yielded by HMRFs was interpreted using covariates, which showed a good discrimination with regard to labels.

As a perspective, Bayesian nonparametric models for hidden Markov random fields could be extended to non-Poissonian models (particularly to account for zero-inflated and over-/under-dispersed cases of

application) and to regression models. Further perspectives of this work include the improvement of the convergence in the VBEM algorithm, since however the KL divergence between the posterior distribution and its approximation converges, the sequence of optimizing parameters is shown to diverge in our current approach.

8.3.8 Hidden Markov models for the analysis of eye movements

Participants: Jean-Baptiste Durand, Sophie Achar.

Joint work with: Anne Guérin-Dugué (GIPSA-lab) and Benoit Lemaire (Laboratoire de Psychologie et Neurocognition)

This research theme is supported by a LabEx PERSYVAL-Lab project-team grant.

In the last years, GIPSA-lab has developed computational models of information search in web-like materials, using data from both eye-tracking and electroencephalograms (EEGs). These data were obtained from experiments, in which subjects had to decide whether a text was related or not to a target topic presented to them beforehand. In such tasks, reading process and decision making are closely related. Statistical analysis of such data aims at deciphering underlying dependency structures in these processes. Hidden Markov models (HMMs) have been used on eye-movement series to infer phases in the reading process that can be interpreted as strategies or steps in the cognitive processes leading to decision. In HMMs, each phase is associated with a state of the Markov chain. The states are observed indirectly through eye-movements. Our approach was inspired by Simola *et al.* (2008) [91], but we used hidden semi-Markov models for better characterization of phase length distributions (Olivier *et al.*, 2021) [29]. The estimated HMM highlighted contrasted reading strategies, with both individual and document-related variability. New results were obtained in the standalone analysis of the eye-movements: 1) a statistical comparison between the effects of three types of texts was performed, considering texts either closely related, moderately related or unrelated to the target topic; 2) a characterization of the effects of the distance to trigger words on transition probabilities and 3) highlighting a predominant intra-individual variability in scanpaths.

As a perspective of this work, the segmentation induced by our eye-movement model could be used to obtain a statistical characterization of functional brain connectivity through simultaneous EEG recordings. This should lead to some integrated models coupling EEG and eye movements within one single HMM for better identification of strategies.

The results of this study have been partially included in a dissemination work at Inria Interstice journal. [86].

8.3.9 Estimation of leaf area densities in tropical forests

Participants: Jean-Baptiste Durand, Florence Forbes, Yuchen Bai.

Joint work with: Grégoire Vincent, IRD, AMAP, Montpellier, France.

Covering just 7% of the Earth's land surface, tropical forests play a disproportionate role in the biosphere: they store about 25% of the terrestrial carbon and contribute to over a third of the global terrestrial productivity. They also recycle about a third of the precipitations through evapotranspiration and thus contribute to generate and maintain a humid climate regionally, with positive effects also extending well beyond the tropics. However, the seasonal variability in fluxes between tropical rainforests and atmosphere is still poorly understood. Better understanding the processes underlying flux seasonality in tropical forests is thus critical to improve our predictive ability on global biogeochemical cycles. Leaf area, one key variable controlling water efflux and carbon influx, is poorly characterized. To monitor

evolutions of biomass, leaf area density (LAD) or gas exchange, aerial and terrestrial laser scanner (LiDAR) measurements have been frequently used.

The principle is, for different LiDAR shoots assumed as independent, to measure the portions of beam lengths between successive hits. Possible censoring comes from beams not being intercepted within a given voxel. Current approaches aim at connecting LAD to the distribution of beam lengths through some statistical model. Such a simplified model does not currently take into account several effects that may impact either LAD or beam lengths: heterogeneity and dependencies in the vegetation properties in different voxels on the one hand and the nature of hit material on the other hand: wood vs. leaves, leading to biases or deteriorated uncertainty in estimation.

This collaboration, supported by Y. Bai's PhD work, aims at developing machine learning methods to address these issues. Current work is now focusing on 1) Applying 3D point convolution neural network to discriminate wood from leaves 2) Modeling the effect of underdetection of vegetal elements due to gradual loss of laser power through different censoring assumptions and validating these assumptions on simulated and real data sets.

8.3.10 Bayesian neural networks

Participants: Julyan Arbel, Pierre Wolinski, Konstantinos Pitas.

The study of feature propagation at initialization in neural networks lies at the root of numerous initialization designs. An assumption very commonly made in the field states that the pre-activations are Gaussian. Although this convenient *Gaussian hypothesis* can be justified when the number of neurons per layer tends to infinity, it is challenged by both theoretical and experimental works for finite-width neural networks. In [79], our major contribution of this work is to construct a family of pairs of activation functions and initialization distributions that ensure that the pre-activations remain Gaussian throughout the network's depth, even in narrow neural networks. In the process, we discover a set of constraints that a neural network should fulfill to ensure Gaussian pre-activations. Additionally, we provide a critical review of the claims of the Edge of Chaos line of works and build an exact Edge of Chaos analysis. We also propose a unified view on pre-activations propagation, encompassing the framework of several well-known initialization procedures. Finally, our work provides a principled framework for answering the much-debated question: is it desirable to initialize the training of a neural network whose pre-activations are ensured to be Gaussian?

We also investigate the cold posterior effect through the lens of PAC-Bayes generalization bounds. We argue that in the non-asymptotic setting, when the number of training samples is (relatively) small, discussions of the cold posterior effect should take into account that approximate Bayesian inference does not readily provide guarantees of performance on out-of-sample data. Instead, out-of-sample error is better described through a generalization bound. In this context, we explore the connections of the ELBO objective from variational inference and the PAC-Bayes objectives. We note that, while the ELBO and PAC-Bayes objectives are similar, the latter objectives naturally contain a temperature parameter λ which is not restricted to be $\lambda = 1$. For classification tasks, in the case of Laplace approximations to the posterior, we show how this PAC-Bayesian interpretation of the temperature parameter captures important aspects of the cold posterior effect.

8.3.11 Graph comparisons

Participants: Sophie Achard, Lucrezia Carboni.

Joint work with: Michel Dojat from GIN, Univ. Grenoble Alpes

In recent accepted publication, we worked on the notion of graph comparisons. Node role explainability in complex networks is very difficult, yet is crucial in different application domains such as social

science, neurosciences or computer science. Many efforts have been made on the quantification of hubs revealing particular nodes in a network using a given structural property. Yet, in several applications, when multiple instances of networks are available and several structural properties appear to be relevant, the identification of node roles remains largely unexplored. Inspired by the node automorphically equivalence relation, we define an equivalence relation on graph's nodes associated with any collection of nodal statistics (i.e. any functions on the node set). This allows us to define new graph global measures, the power coefficient and the orthogonality score to evaluate the collection parsimony and heterogeneity of a given nodal statistics collection. In addition, we introduce a new method based on structural patterns to compare graphs that have the same vertices set. This methods assigns a value to a node to determine its role distinctiveness in a graph family. Extensive numerical results of our method are conducted on both generative graph models and real data concerning human brain functional connectivity. The differences in nodal statistics are shown to be dependent on the underlying graph structure. Comparisons between generative models and real networks combining two different nodal statistics reveal the complexity of human brain functional connectivity with differences at both global and nodal levels. Using a group of 200 healthy controls connectivity networks, our method is able to compute high correspondence scores among the whole population, to detect homotopy, and finally to quantify differences between comatose patients and healthy controls.

8.3.12 Spatio-temporal data

Participants: Sophie Achard, Hana Lbath.

Joint work with: Alex Petersen, Brigham Young University, US and Wendy Meiring, University Santa Barbara California, US

A novel non-parametric estimator of the correlation between regions, or groups of arbitrarily dependent variables, is proposed in the presence of noise. The challenge resides in the fact that both noise and low intra-regional correlation lead to inconsistent inter-regional correlation estimation using classical approaches. While some existing methods handle one of these issues or the other, none tackle both at the same time. To address this problem, we propose a trade-off between two approaches: correlating regional averages, which is not robust to low average intra-regional correlation, and averaging pairwise inter-regional correlations, which is not robust to high noise. To that end, we project the data onto a space where the Euclidean distance can be used as a proxy for the sample correlation. We then leverage hierarchical clustering to gather together highly correlated variables within each region prior to averaging. We prove our estimator is consistent for an appropriate cut-off height of the dendrogram. We also empirically show our approach surpasses popular estimators in terms of quality and provide illustrations on real-world datasets that further demonstrate its usefulness. [56]

8.4 Inverse problems

8.4.1 Sequential Bayesian experimental design for inverse problems.

Participants: Florence Forbes, Jacopo Iollo.

Bayesian Optimal Experimental Design is a technique that allows practitioners to efficiently make use of limited experimental resources. In the sequential case however, we often have to run from start computationally intensive inference techniques such as Markov Chain Monte Carlo or Variational Inference to obtain the current estimate of the posterior. In this study, we show how we can leverage Sequential Monte Carlo Samplers in order to simultaneously estimate the best possible design and the current posterior. This iterative procedure make use of past information in order to avoid to have to run an inference algorithm from start at each iteration. We demonstrate its applicability in a source location inverse problem.

8.4.2 Bayesian nonparametric mixture of experts for high-dimensional inverse problems.

Participants: Julyan Arbel, Florence Forbes, Trung Tin Nguyen.

A wide class of problems can be formulated as inverse problems where the goal is to find parameter values that best explain some observed measures. Typical constraints in practice are that relationships between parameters and observations are highly nonlinear, with high-dimensional observations and multi-dimensional correlated parameters. To handle these constraints, we consider probabilistic mixtures of locally linear models, which can be seen as particular instances of mixtures of experts (MoE). We have shown in previous studies that such models had a good approximation ability provided the number of experts was large enough. This contribution is to propose a general scheme to design a tractable Bayesian nonparametric (BNP) MoE model to avoid any commitment to an arbitrary number of experts. A tractable estimation algorithm is designed using a variational approximation and theoretical properties are derived on the predictive distribution and the number of components. Illustrations on simulated and real data show good results in terms of selection and computing time compared to more traditional model selection procedures.

8.4.3 Hierarchical Bayesian models for simulation-based inference

Participants: Pedro Rodrigues, Julia Linhart.

Joint work with: Thomas Moreau and Alexandre Gramfort from Inria Saclay and Gilles Louppe from Université de Liège

Inferring the parameters of a stochastic model based on experimental observations is central to the scientific method. A particularly challenging setting is when the model is strongly indeterminate, i.e. when distinct sets of parameters yield identical observations. This arises in many practical situations, such as when inferring the distance and power of a radio source (is the source close and weak or far and strong?) or when estimating the amplifier gain and underlying brain activity of an electrophysiological experiment. In a recent work, we have proposed the hierarchical neural posterior estimation (HNPE), a novel method for cracking such indeterminacy by exploiting additional information conveyed by an auxiliary set of observations sharing global parameters. This method extends recent developments in simulation-based inference (SBI) based on normalizing flows to Bayesian hierarchical models. We validated HNPE quantitatively on a motivating example amenable to analytical solutions and then applied it to invert a well known non-linear model from computational neuroscience, using both simulated and real EEG data.

8.4.4 Bayesian inference on a large-scale brain simulator

Participants: Pedro Rodrigues.

Joint work with: Nicholas Tolley and Stephanie Jones from Brown University, Alexandre Gramfort from Inria Saclay

The Human Neocortical Neurosolver (HNN) is a framework whose foundation is a cortical column model with cell and circuit level detail designed to connect macroscale signals to meso/microcircuit level phenomena. We apply this model to study the cellular and circuit mechanisms of beta generation using local field potential (LFP) recordings from the non-human primate (NHP) motor cortex. To characterize beta producing mechanisms, we employ simulation based inference (SBI) in the HNN modeling tool. This framework leverages machine learning techniques and neural density estimators to characterize the relationship between a large space of model parameters and simulation output. In this setting, Bayesian inference can be applied to models with intractable likelihood functions (Gonçalves 2020, Papamakarios

2021). The main goal of this project is to provide a set of guidelines for scientists that wish to apply simulation-based inference to their neuroscience studies with a large-scale simulator such as HNN. This involves developing new methods for extracting summary features, checking the quality of the posterior approximation, etc. This work is mostly carried out by the Ph.D. student Nicholas Tolley from Brown University.

9 Bilateral contracts and grants with industry

9.1 Bilateral contracts with industry

Participants: Florence Forbes, Pedro Luiz Coelho Rodrigues, Stephane Girard, Julyan Arbel.

Plan de Relance project with GE Healthcare (2022-24). The topic of the collaboration is related to early anomaly detection of failures in medical transducer manufacturing. The financial support for STATIFY is of 155K euros.

Contract with EDF (2020-2023). Julyan Arbel and Stéphane Girard are the advisors of the PhD thesis of Théo Moins founded by EDF. The goal is to investigate sensitivity analysis and extrapolation limits in extreme-value theory Bayesian methods. The financial support for STATIFY is of 150 keuros.

Contract with TDK-Invensense (2020-2023). Julyan Arbel is the advisor of the PhD thesis of Minh Tri Lê founded by TDK-Invensense. The goal is to apply deep learning methods on small size systems, thus investigating compression methods in deep learning. The financial support for STATIFY is of 150 keuros.

Contract with Phimeca (2021-2022). Stéphane Girard is supervising the work of Valentin Pibernus (engineer at Phimeca) funded by a PEPS-AMIES-2 project and a contract between STATIFY and Phimeca. The goal is to implement auto-associative models for assessing the dispersion of pollutants in the atmosphere. The total financial support for STATIFY is of 50 keuros.

Contract with Valeo (2021-2022). Stéphane Girard is supervising the work of Pascal Dkengne Sielenou (engineer at STATIFY) funded by ca ontract between STATIFY and Valeo. The goal is to design statistical method for autonomous vehicle data analysis [70]. The total financial support for STATIFY is of 50 keuros.

10 Partnerships and cooperations

10.1 International initiatives

10.1.1 Inria associate team not involved in an IIL or an international program

Participants: Florence Forbes, Julyan Arbel, Jean-Baptiste Durand, Stephane Girard, Trung Tin Nguyen.

LANDER stands for "Latent Analysis, Adversarial Networks, and DimEnsionality Reduction" and is an associate team with researchers in Australia that has started in 2019. Details can be found at [Lander website](#).

During the year, a number of journal publications have been finalized on subjects directly linked to the project, mixture of expert models and variants related to high dimensional regression, distributions and properties for non-Gaussian data, clustering and classification, Approximate Bayesian Computation (ABC), Bayesian non parametrics and Markov random fields, Expectiles, etc. The corresponding publications and working papers are listed on the website.

All this work served as a basis for a new submission to the Australian Research Council (ARC), which has been recently granted.

10.2 International research visitors

10.2.1 Visits of international scientists

Other international visits to the team

Kai Qin

Status (Professor)

Institution of origin: Swinburn University

Country: Australia

Dates: July 2022

Context of the visit: Lander associate team

Hien Duy Nguyen

Status (Professor)

Institution of origin: University of Queensland

Country: Australia

Dates: July 2022

Context of the visit: Lander associate team

Filippo Ascolani

Status (PhD Student)

Institution of origin: Bocconi University, Milano

Country: Italy

Dates: November 2022

Context of the visit: collaboration

Alberto Gonzales Sanz

Status (PhD Student)

Institution of origin: University of Toulouse

Country: France

Dates: November 2022

Context of the visit: collaboration

10.2.2 Visits to international teams

Research stays abroad

Konstantinos Pitas

Visited institution: Riken

Country: Japan

Dates: October-November 2022

Context of the visit: Bayes-Duality ANR project

Hana Lbath

Visited institution: Brigham Young University

Country: USA

Dates: May-June 2022

Context of the visit: QFunc ANR project

Sophie Achard

Visited institution: Brigham Young University

Country: USA

Dates: May 2022

Context of the visit: QFunc ANR project

Florence Forbes

Visited institution: University of Queensland and Queensland University of Technology

Country: Australia

Dates: April 2022

Context of the visit: Lander associate team

10.3 National initiatives

Participants: Jean-Baptiste Durand, Florence Forbes, Julyan Arbel, Sophie Achard, Stephane Girard, Giovanni Poggiato, Pedro Luiz Coelho Rodrigues.

ANR

- STATIFY is involved in the ANR project **GAMBAS** (2019-2023) hosted by CIRAD, Montpellier. The project Generating Advances in Modeling Biodiversity And ecosystem Services (GAMBAS) develops statistical improvements and ecological relevance of joint species distribution models. The project supports the PhD thesis of Giovanni Poggiato.
- an ANR project RADIO-AIDE (2022-26) for *Radiation induced neurotoxicity assessed by Spatio-temporal modeling and AI after brain radiotherapy* coordinated by S.Ancelet from IRSN has been granted for 4 years starting from April 2022. It involves STATIFY, Grenoble Insitute of Neurosciences, Pixyl, ICANS, APHP, ICM and ENS P.Saclay. The available funding for STATIFY is 94K euros.

- ANR project PEG2 (2022-26) on Predictive Ecological Genomics: STATIFY is involved in this 4-year project recently accepted in July 2022. The PI is prof. Olivier Francois who spent 2 years (2021-22) in the team on a *Delegation* position.
- Julyan Arbel is coPI of the Bayes-Duality project launched with a funding of \$2.76 millions by Japan JST - French ANR for a total of 5 years starting in October 2021. The goal is to develop a new learning paradigm for Artificial Intelligence that learns like humans in an adaptive, robust, and continuous fashion. On the Japan side the project is led by Mohammad Emtiyaz Khan as the research director, and Kenichi Bannai and Rio Yokota as Co-PIs.
- STATIFY is involved in the 4-year ANR project ExtremReg (2019-2023) hosted by Toulouse University. This research project aims to provide new adapted tools for nonparametric and semiparametric modeling from the perspective of extreme values. Our research program concentrates around three central themes. First, we contribute to the expanding literature on non-regular boundary regression where smoothness and shape constraints are imposed on the regression function and the regression errors are not assumed to be centred, but one-sided. Our second aim is to further investigate the study of the modern extreme value theory built on the use of asymmetric least squares instead of traditional quantiles and order statistics. Finally, we explore the less-discussed problem of estimating high-dimensional, conditional and joint extremes The financial support for STATIFY is about 15.000 euros.

Grenoble Idex projects

- **MIAI**, Multidisciplinary Institute in Artificial Intelligence: In the context of the MIAI institute, S. Achard is co-PI of a chair (2020-23) on *Toward Robust and Understandable Neuromorphic Systems* funding one PhD student and several post-doc fellows.
- MIAI: Pedro Rodrigues was awarded 13K euros for his project on *Machine Learning for Experimental Data* via the IRGA 2022 call for projects.
- MIAI: F. Forbes, J-B. Durand and Y. Bai were awarded 4.5K euros for their project on *Developing IA for estimating leaf area density area in tropical forests from LiDAR data*, Nov. 2022.

10.3.1 Networks

MSTGA and AIGM INRAE (French National Institute for Agricultural Research) networks: F. Forbes and J.B Durand are members of the INRAE network called AIGM (ex MSTGA) network since 2006, [website](#), on Algorithmic issues for Inference in Graphical Models. It is funded by INRAE MIA and RNSC/ISC Paris. This network gathers researchers from different disciplines. STATIFY co-organized and hosted 2 of the network meetings in 2008 and 2015 in Grenoble.

11 Dissemination

Participants: Jean-Baptiste Durand, Florence Forbes, Julyan Arbel, Sophie Achard, Stephane Girard, Pedro Luiz Coelho Rodrigues.

11.1 Promoting scientific activities

11.1.1 Scientific events: organisation

Member of the organizing committees

- F. Forbes is a member of the organizing committee of the MCM2023 14th international conference on Monte-Carlo methods and applications.

- F. Forbes is a member of the organizing committee of the 1st French workshop on AI for biomedical imaging IABM2023 and of the scientific committee for the Bayesian autumn school at CIRM in October 2023.

11.1.2 Scientific events: selection

Member of the conference program committees

- Stéphane Girard, Member of the Scientific Program Committee of the 15th International Conference of the ERCIM WG on computational and methodological statistics, London, 2022. He organized an invited session entitled "*Machine learning for extremes*" and he was chair of a contributed session entitled "*Extreme values*".
- Julyan Arbel, Member of the scientific committee of the Statistical Methods for Post Genomic Data analysis (SMPGD) meeting, in January 2022.
- Julyan Arbel, Member of the scientific committee of the One World Approximate Bayesian Computation (ABC) Seminar, from 2020 to 2022.

Reviewer

- Julyan Arbel, Reviewer, *Advances in Neural Information Processing Systems 2022*.
- Julyan Arbel, Reviewer, *Symposium on Advances in Approximate Bayesian Inference (AABI) 2022*.

11.1.3 Journal

Member of the editorial boards

- Stéphane Girard, Associate Editor, *Revstat - Statistical Journal* since 2019.
- Stéphane Girard, Member of the Advisory Board, *Dependence Modeling* since 2015.
- Stéphane Girard, Associate Editor, *Journal of Multivariate Analysis* from 2016 to 2022.
- Julyan Arbel, Associate Editor, *Bayesian Analysis* since 2019.
- Julyan Arbel, Associate Editor, *Australian and New Zealand Journal of Statistics* since 2019.
- Julyan Arbel, Associate Editor, *Statistics & Probability Letters* since 2019.
- Julyan Arbel, Associate Editor, *Computational Statistics & Data Analysis* since 2020.
- Florence Forbes, Associate Editor, *Australian and New Zealand Journal of Statistics* since 2019.

Reviewer - reviewing activities

- Stéphane Girard, Reviewer, *Advances in Water Resources*.
- Stéphane Girard, Reviewer, *Extremes*.
- Stéphane Girard, Reviewer, *Computational Statistics and Data Analysis*.
- Stéphane Girard, Reviewer, *Advances in Data Analysis and Classification*.
- Julyan Arbel, Reviewer, *Computo*.
- Julyan Arbel, Reviewer, *Journal of Machine Learning Research*.

11.1.4 Invited talks

- F. Forbes, Automatic learning of functional summary statistics for approximate Bayesian computation, O'Bayes 2022 - Objective Bayes Methodology Conference, Santa Cruz, United States, September 2022, [46].
- F. Forbes, Learning approaches for Bayesian inverse problems, UGA-McMaster joint workshop 2022, Hamilton, Canada, June 2022, [47].
- F. Forbes, Online Majorization Minimization algorithms, CIRM Workshop on Computational Methods for Unifying Multiple Statistical Analyses (Fusion), Luminy, France, October 2022, [48].
- F. Forbes, Simulation based inference for high dimensional inverse problems: application to magnetic resonance fingerprinting, Colloque Intelligence Artificielle et santé : approches interdisciplinaires 2022, Nantes, France, June 2022, [49].
- A. Usseglio-Carleve, S. Girard, G. Stupfler, Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models, Insurance Data Science Conference 2022, Milano, Italy, June 2022.
- S. Achard, Statistical comparisons of spatio-temporal networks, CIRM Workshop on GraphLearn: Machine Learning and Signal Processing on Graphs, Luminy, France, November 2022,

11.1.5 Research administration

- S. Achard, since Nov. 2020, has been elected as the head of Pole MSTIC (with J.P. Jamont, K. Altisen and C. Lescop) at University of Grenoble. Before she was co-director of pole MSTIC since 2017.
- Julyan Arbel, Member of the scientific committee of the Data Science axis of Persyval Labex.
- Julyan Arbel, Board of Directors member of ISBA, the International Society for Bayesian Analysis, 2022-2025.
- F. Forbes is a member of the advisory committee of the Helmholtz AI Cooperation Unit since 2019.
- F. Forbes is Scientific Advisor since March 2015 for the Pixyl company.
- F. Forbes and S. Achard are members of the EURASIP Technical Area Committee BISA (Biomedical Image & Signal Analytics) since January 2021 for a 3 years duration.
- J. Arbel has been a member of the Comité des Emplois Scientifiques at Inria Grenoble since 2019.
- P. Rodrigues is a member of the Comité D'éveloppement Technologique, a committee for evaluating software development projects, at Inria Grenoble since 2021.
- F. Forbes is a member of the Comité d'organisation stratégique of Inria Grenoble since 2017.
- F. Forbes was deputy head of science at Inria Grenoble, from Sep. 2021 to Sep. 2022.

11.2 Teaching - Supervision - Juries

11.2.1 Teaching

- Master: Stéphane Girard, *Statistique Inférentielle Avancée*, 18 ETD, M1 level, Ensimag. Grenoble-INP, France.
- Master: Stéphane Girard, *Introduction to Extreme-Value Analysis*, 18 ETD, M2 level, Univ-Grenoble Alpes (UGA), France.
- Master: Julyan Arbel, *Bayesian nonparametrics and Bayesian deep learning*, Master Mathématiques Apprentissage et Sciences Humaines (M*A*S*H), Université PSL (Paris Sciences & Lettres), 25 ETD.

- Master: Julyan Arbel, *Bayesian deep learning*, Master Intelligence Artificielle, Systèmes, Données (IASD), Université PSL (Paris Sciences & Lettres), 12 ETD.
- Master, Julyan Arbel, *Bayesian machine learning*, Master Mathématiques Vision et Apprentissage (MVA), École normale supérieure Paris-Saclay, 36 ETD.
- Master, Julyan Arbel, *Bayesian statistics*, Ensimag Grenoble INP, 27 ETD.
- Master: Jean-Baptiste Durand, *Statistics and probability*, 192H, M1 and M2 levels, Ensimag Grenoble INP, France. Head of the MSIAM M2 program, in charge of the data science track.
- Jean-Baptiste Durand is a faculty member at Ensimag, Grenoble INP.
- Sophie Achard M1 course Théorie des graphes et réseaux sociaux, M1 level, MIASHS, Université Grenoble Alpes (UGA), 14 ETD.

11.2.2 Supervision

- Stéphane Girard was co-supervisor of the PhD thesis of Michaël Allouche (with Emmanuel Gobet, Ecole Polytechnique) "*Contributions to generative modeling and dictionary learning: Theory and application*", Institut Polytechnique de Paris, defended on December, 2022.
- Stéphane Girard was co-supervisor of the PhD thesis of Meryem Bousebata (with Geoffroy Enjolras, UGA), "*Bayesian estimation of extreme risk measures: Implication for the insurance of natural disasters*", UGA, defended on March 2022.
- Julyan Arbel was supervisor of the PhD thesis of Mariia Vladimirova, "*Bayesian Neural Networks' Distributional Properties*", Inria, defended on March 2022.
- PhD in progress: Daria Bystrova, "Joint Species Distribution Modeling: Dimension reduction using Bayesian nonparametric priors", started on October 2019, Julyan Arbel and Wilfried Thuiller, Université Grenoble Alpes.
- PhD in progress: Giovanni Poggiatto, "Scalable Approaches for Joint Species Distribution Modeling", started on November 2019, Julyan Arbel and Wilfried Thuiller, Université Grenoble Alpes.
- PhD in progress: Théo Moins "Quantification bayésienne des limites d'extrapolation en statistique des valeurs extrêmes", started on October 2020, Stéphane Girard and Julyan Arbel, Université Grenoble Alpes.
- PhD in progress: Minh Tri Lê, "Constrained signal processing using deep neural networks for MEMs sensors based applications.", started on September 2020, Julyan Arbel and Etienne de Foras, Université Grenoble Alpes, CIFRE Invensense.
- PhD in progress: Hana Lbath, "Advanced Spatiotemporal Statistical Models for Quantification and Estimation of Functional Connectivity", started in October 2020, supervised by Sophie Achard and Alex Petersen (Brigham Young University, Utah, USA).
- PhD in progress: Lucrezia Carboni, "Graph embedding for brain connectivity", started in October 2020, supervised by Sophie Achard and Michel Dojat (GIN).
- PhD in progress: Louise Alamichel. "Bayesian Nonparametric methods for complex genomic data" Inria, started in October 2021, advised by Julyan Arbel and Guillaume Kon Kam King (INRAE).
- PhD in progress: Julien Zhou. "Learning combinatorial bandit models under privacy constraints" Inria-Criteo, started in November 2022, advised by Julyan Arbel, Pierre Gaillard and Thibaud Rahier.
- PhD in progress: Mohamed-Bahi Yahiaoui. "Computation time reduction and efficient uncertainty propagation for fission gas simulation" CEA Cadarache-Inria, started in October 2021, advised by Julyan Arbel, Loïc Giraldi, Geoffrey Daniel.

- PhD in progress: Yuchen Bai, "Hierarchical Bayesian Modelling of leaf area density from UAV-lidar", started in October 2021, supervised by Jean-Baptiste Durand, Florence Forbes and Gregoire Vincent (IRD, Montpellier).
- PhD in progress: Julia Linhart, "Simulation based inference with neural networks: applications to computational neuroscience", started in November 2021, supervised by Pedro Rodrigues and Alexandre Gramfort (DR Inria Saclay).
- PhD in progress: Jacopo Iollo, started in January 2022, supervised by F. Forbes, P. Alliez (DR Inria Sophia) and C. Heinkele (Cerema, Strasbourg).
- PhD in progress: Geoffroy Oudoumanessah, started in October 2022, supervised by F. Forbes, C. Lartizien (Creatis, Lyon) and M. Dojat (GIN).
- PhD in progress: Theo Sylvestre, started in October 2022, supervised by F. Forbes and S. Ancelet (IRSN).
- PhD in progress: Benjamin Lambert, started in January 2022, supervised by F. Forbes and M. Dojat (GIN).

11.2.3 Juries

- Stéphane Girard, Member of the PhD thesis committee of Gloria Buritica, "*Assessing the time dependence of multivariate extremes for heavy rainfall modeling*", Sorbonne Université, May 2022.
- Julyan Arbel, Rapporteur of the PhD thesis of Gabriel Ducrocq, Institut Polytechnique de Paris, France.
- Julyan Arbel, Rapporteur of the PhD thesis of Ioanni Mitro, University College Dublin, Ireland.
- Julyan Arbel, Rapporteur of the PhD thesis of Yuqing Hu, IMT Atlantique & Orange, France.
- Julyan Arbel, Member of the PhD thesis of Meryem Bousebata, Inria Grenoble, France.
- F. Forbes, Member of the PhD thesis of Grégoire Aufort, "Statistique Computationnelle Bayésienne pour l'étude des Distributions Spectrales d'énergie des galaxies", October 2022, University Aix-Marseille.
- F. Forbes, Reviewer for the Master Thesis of John Davey from Adelaide University, Australia, July 2022.
- F. Forbes, Reviewer for the CSI, mid-term examination, for the PhD thesis of Thomas Coudert, GIN, University of Grenoble, October 2022.

11.3 Popularization

11.3.1 Articles and contents

S. Achard and J-B. Durand published an illustration of the process of statistical modeling, in the Inria journal for popularization of science (in French). The topic was illustrated through the analysis of eye movements to infer cognitive processes.

12 Scientific production

12.1 Major publications

- [1] C. Bouveyron, S. Girard and C. Schmid. 'High dimensional data clustering'. In: *Computational Statistics and Data Analysis* 52 (2007), pp. 502–519.

- [2] F. Boux, F. Forbes, J. Arbel, B. Lemasson and E. L. Barbier. ‘Bayesian inverse regression for vascular magnetic resonance fingerprinting’. In: *IEEE Transactions on Medical Imaging* 40.7 (July 2021), pp. 1827–1837. DOI: [10.1109/TMI.2021.3066781](https://doi.org/10.1109/TMI.2021.3066781). URL: <https://hal.archives-ouvertes.fr/hal-02314026>.
- [3] A. Daouia, S. Girard and G. Stupfler. ‘Estimation of Tail Risk based on Extreme Expectiles’. In: *Journal of the Royal Statistical Society series B* 80 (2018), pp. 263–292.
- [4] A. Deleforge, F. Forbes and R. Horaud. ‘High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables’. In: *Statistics and Computing* (Feb. 2014). DOI: [10.1007/s11222-014-9461-5](https://doi.org/10.1007/s11222-014-9461-5). URL: <https://hal.inria.fr/hal-00863468>.
- [5] F. Forbes and G. Fort. ‘Combining Monte Carlo and Mean field like methods for inference in hidden Markov Random Fields’. In: *IEEE trans. Image Processing* 16.3 (2007), pp. 824–837.
- [6] F. Forbes, H. D. Nguyen, T. T. Nguyen and J. Arbel. ‘Summary statistics and discrepancy measures for ABC via surrogate posteriors’. In: *Statistics and Computing* 32.85 (2022). DOI: [10.1007/s11222-022-10155-6](https://doi.org/10.1007/s11222-022-10155-6). URL: <https://hal.archives-ouvertes.fr/hal-03139256>.
- [7] F. Forbes and D. Wraith. ‘A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweights: Application to robust clustering’. In: *Statistics and Computing* 24.6 (Nov. 2014), pp. 971–984. DOI: [10.1007/s11222-013-9414-4](https://doi.org/10.1007/s11222-013-9414-4). URL: <https://hal.inria.fr/hal-00823451>.
- [8] S. Girard. ‘A Hill type estimate of the Weibull tail-coefficient’. In: *Communication in Statistics - Theory and Methods* 33.2 (2004), pp. 205–234.
- [9] S. Girard, G. C. Stupfler and A. Usseglio-Carleve. ‘Extreme Conditional Expectile Estimation in Heavy-Tailed Heteroscedastic Regression Models’. In: *Annals of Statistics* 49.6 (Dec. 2021), pp. 3358–3382. DOI: [10.1214/21-AOS2087](https://doi.org/10.1214/21-AOS2087). URL: <https://hal.archives-ouvertes.fr/hal-03306230>.
- [10] H. Lu, J. Arbel and F. Forbes. ‘Bayesian nonparametric priors for hidden Markov random fields’. In: *Statistics and Computing* 30 (2020), pp. 1015–1035. DOI: [10.1007/s11222-020-09935-9](https://doi.org/10.1007/s11222-020-09935-9). URL: <https://hal.archives-ouvertes.fr/hal-02163046>.

12.2 Publications of the year

International journals

- [11] M. Allouche, J. El Methni and S. Girard. ‘A refined Weissman estimator for extreme quantiles’. In: *Extremes* (2022). URL: <https://hal.inria.fr/hal-03266676>.
- [12] M. Allouche, S. Girard and E. Gobet. ‘A generative model for fBm with deep ReLU neural networks’. In: *Journal of Complexity* 73 (Dec. 2022), p. 101667. DOI: [10.1016/j.jco.2022.101667](https://doi.org/10.1016/j.jco.2022.101667). URL: <https://hal.archives-ouvertes.fr/hal-03237854>.
- [13] M. Allouche, S. Girard and E. Gobet. ‘EV-GAN: Simulation of extreme events with ReLU neural networks’. In: *Journal of Machine Learning Research* 23.150 (2022), pp. 1–39. URL: <https://hal.archives-ouvertes.fr/hal-03250663>.
- [14] J. Arbel, S. Girard, H. D. Nguyen and A. Usseglio-Carleve. ‘Multivariate expectile-based distribution: properties, Bayesian inference, and applications’. In: *Journal of Statistical Planning and Inference* 225 (2023), pp. 146–170. URL: <https://hal.inria.fr/hal-03428827>.
- [15] M. Bousebata, G. Enjolras and S. Girard. ‘Extreme Partial Least-Squares’. In: *Journal of Multivariate Analysis* 194 (2023), p. 105101. DOI: [10.1016/j.jmva.2022.105101](https://doi.org/10.1016/j.jmva.2022.105101). URL: <https://hal.inria.fr/hal-03165399>.
- [16] A. Constantin, M. Fauvel and S. Girard. ‘Mixture of multivariate Gaussian processes for classification of irregularly sampled satellite image time-series’. In: *Statistics and Computing* 32 (2022), p. 79. DOI: [10.1007/s11222-022-10145-8](https://doi.org/10.1007/s11222-022-10145-8). URL: <https://hal.science/hal-03280484>.
- [17] J.-B. Durand, F. Forbes, C. D. Phan, L. Truong, H. D. Nguyen and F. Dama. ‘Bayesian nonparametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia’. In: *Australian and New Zealand Journal of Statistics*. Special Issue: Geoff McLachlan Festschrift 64.2 (June 2022), pp. 171–204. DOI: [10.1111/anzs.12369](https://doi.org/10.1111/anzs.12369). URL: <https://hal.inria.fr/hal-03138803>.

- [18] F. Forbes, H. D. Nguyen, T. T. Nguyen and J. Arbel. ‘Summary statistics and discrepancy measures for ABC via surrogate posteriors’. In: *Statistics and Computing* 32.85 (2022). DOI: [10.1007/s11222-022-10155-6](https://doi.org/10.1007/s11222-022-10155-6). URL: <https://hal.archives-ouvertes.fr/hal-03139256>.
- [19] S. Girard, H. Lorenzo and J. Saracco. ‘Advanced topics in Sliced Inverse Regression’. In: *Journal of Multivariate Analysis* 188 (2022), p. 104852. DOI: [10.1016/j.jmva.2021.104852](https://doi.org/10.1016/j.jmva.2021.104852). URL: <https://hal.inria.fr/hal-03367798>.
- [20] S. Girard, G. Stupfler and A. Usseglio-Carleve. ‘Functional estimation of extreme conditional expectiles’. In: *Econometrics and Statistics* 21 (Jan. 2022), pp. 131–158. DOI: [10.1016/j.ecosta.2021.05.006](https://doi.org/10.1016/j.ecosta.2021.05.006). URL: <https://hal.inria.fr/hal-03117547>.
- [21] S. Girard, G. Stupfler and A. Usseglio-Carleve. ‘Nonparametric extreme conditional expectile estimation’. In: *Scandinavian Journal of Statistics* 49.1 (Mar. 2022), pp. 78–115. DOI: [10.1111/sjos.12502](https://doi.org/10.1111/sjos.12502). URL: <https://hal.science/hal-02114255>.
- [22] S. Girard, G. Stupfler and A. Usseglio-Carleve. ‘On automatic bias reduction for extreme expectile estimation’. In: *Statistics and Computing* 32 (9th Aug. 2022), p. 64. DOI: [10.1007/s11222-022-10118-x](https://doi.org/10.1007/s11222-022-10118-x). URL: <https://hal.science/hal-03086048>.
- [23] M. Jallais, P. L. Coelho Rodrigues, A. Gramfort and D. Wassermann. ‘Inverting brain grey matter models with likelihood-free inference: a tool for trustable cytoarchitecture measurements’. In: *Journal of Machine Learning for Biomedical Imaging* (8th May 2022), pp. 1–27. DOI: [10.48550/arXiv.2111.08693](https://doi.org/10.48550/arXiv.2111.08693). URL: <https://hal.inria.fr/hal-03417406>.
- [24] B. Kugler, F. Forbes and S. Douté. ‘Fast Bayesian Inversion for high dimensional inverse problems’. In: *Statistics and Computing* 32.31 (22nd Mar. 2022). DOI: [10.1007/s11222-021-10019-5](https://doi.org/10.1007/s11222-021-10019-5). URL: <https://hal.archives-ouvertes.fr/hal-02908364>.
- [25] T. Mistral, P. Roca, C. Maggia, A. Tucholka, F. Forbes, S. Doyle, A. Krainik, D. Galanaud, E. Schmitt, S. Kremer, A. Kastler, I. Troprès, E. Barbier, J.-F. Payen and M. Dojat. ‘Automated Quantification of Brain Lesion Volume From Post-trauma MR Diffusion-Weighted Images’. In: *Frontiers in Neurology* 12 (23rd Feb. 2022), p. 740603. DOI: [10.3389/fneur.2021.740603](https://doi.org/10.3389/fneur.2021.740603). URL: <https://hal.archives-ouvertes.fr/hal-03593111>.
- [26] V. Muñoz-Ramírez, V. Kmetzsch, F. Forbes, S. Meoni, E. Moro and M. Dojat. ‘Subtle anomaly detection: Application to brain MRI analysis of *de novo* Parkinsonian patients’. In: *Artificial Intelligence in Medicine* 125 (Mar. 2022), p. 102251. DOI: [10.1016/j.artmed.2022.102251](https://doi.org/10.1016/j.artmed.2022.102251). URL: <https://hal.science/hal-03578350>.
- [27] H. D. Nguyen and F. Forbes. ‘Global implicit function theorems and the online Expectation-Maximisation algorithm’. In: *Australian and New Zealand Journal of Statistics* 64.2 (June 2022), pp. 255–281. DOI: [10.1111/anzs.12356](https://doi.org/10.1111/anzs.12356). URL: <https://hal.archives-ouvertes.fr/hal-03110213>.
- [28] H. D. Nguyen, S. Lee and F. Forbes. ‘A Festschrift for Geoff McLachlan’. In: *Australian and New Zealand Journal of Statistics* 64.2 (June 2022), pp. 111–116. DOI: [10.1111/anzs.12372](https://doi.org/10.1111/anzs.12372). URL: <https://hal.archives-ouvertes.fr/hal-03883742>.
- [29] B. Olivier, A. Guérin-Dugué and J.-B. Durand. ‘Hidden Semi-Markov Models to Segment Reading Phases from Eye Movements’. In: *Journal of Eye Movement Research* 15.4 (30th Sept. 2022), p. 5. DOI: [10.16910/jemr.15.4.5](https://doi.org/10.16910/jemr.15.4.5). URL: <https://hal.inria.fr/hal-03155843>.
- [30] S. M. Potin, S. Douté, B. Kugler and F. Forbes. ‘The impact of asteroid shapes and topographies on their reflectance spectroscopy’. In: *Icarus* 376 (Apr. 2022), 114806:1–24. DOI: [10.1016/j.icarus.2021.114806](https://doi.org/10.1016/j.icarus.2021.114806). URL: <https://hal.archives-ouvertes.fr/hal-03500909>.
- [31] E. Roger, L. Rodrigues de Almeida, H. Loevenbruck, M. Perrone-Bertolotti, E. Cousin, J.-L. Schwartz, P. Perrier, M. Dohen, A. Vilain, P. Baraduc, S. Achard and M. Baciú. ‘Unraveling the functional attributes of the language connectome: crucial subnetworks, flexibility and variability’. In: *NeuroImage* 263 (Nov. 2022), p. 119672. DOI: [10.1016/j.neuroimage.2022.119672](https://doi.org/10.1016/j.neuroimage.2022.119672). URL: <https://hal.univ-grenoble-alpes.fr/hal-03811050>.

- [32] S. Yassine, U. Gschwandtner, M. Auffret, S. Achard, M. Verin, P. Fuhr and M. Hassan. ‘Functional Brain Dysconnectivity in Parkinson’s Disease: A 5-Year Longitudinal Study’. In: *Movement Disorders* 37.7 (July 2022), pp. 1444–1453. DOI: [10.1002/mds.29026](https://doi.org/10.1002/mds.29026). URL: <https://hal.archives-ouvertes.fr/hal-03656851>.

International peer-reviewed conferences

- [33] S. Achard, I. Gannaz and K. Polissano. ‘Génération de modèles graphiques’. In: GRETSI 2022 - XXVIIIème Colloque francophone de traitement du signal et des images. Nancy, France, 2022, pp. 1–3. URL: <https://hal.archives-ouvertes.fr/hal-03811642>.
- [34] G. J.-P. C. Becq, A. Bhanot, C. Meillier, S. Achard and E. L. Barbier. ‘Effect of filtering on the evaluation of dynamic functional connectivity and application to data from functional MRI of rats’. In: *28e Colloque sur le traitement du signal et des images*. Grets2022 - 28° Colloque sur le traitement du signal et des images. Nancy, France, 9th Sept. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03868431>.
- [35] B. Lambert, F. Forbes, S. Doyle, A. Tucholka and M. Dojat. ‘Beyond Voxel Prediction Uncertainty: Identifying brain lesions you can trust’. In: *Lecture Notes in Computer Science*. iMIMIC 2022 - Workshop on Interpretability of Machine Intelligence in Medical Image Computing. Vol. 13611. Singapore, Singapore, 7th Oct. 2022, pp. 61–70. DOI: [10.1007/978-3-031-17976-1_6](https://doi.org/10.1007/978-3-031-17976-1_6). URL: <https://hal.archives-ouvertes.fr/hal-03883749>.
- [36] H. D. Nguyen, F. Forbes, G. Fort and O. Cappé. ‘An online Minorization-Maximization algorithm’. In: 17th Conference of the International Federation of Classification Societies. Proceedings of the 17th Conference of the International Federation of Classification Societies. Porto, Portugal, July 2022. URL: <https://hal.archives-ouvertes.fr/hal-03542180>.
- [37] T. Rahier, S. Marié and F. Forbes. ‘A Pre-Screening Approach for Faster Bayesian Network Structure Learning’. In: ECML-PKDD 2022 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. ECML-PKDD. Grenoble, France, 19th Sept. 2022, pp. 1–16. URL: <https://hal.archives-ouvertes.fr/hal-03873684>.

Conferences without proceedings

- [38] L. Alamichel, D. Bystrova, J. Arbel and G. Kon Kam King. ‘On the consistency of Bayesian nonparametric mixtures for the number of clusters’. In: ISBA 2022 - World Meeting International Society for Bayesian Analysis. Montreal, Canada, 26th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03866460>.
- [39] M. Allouche, J. El Methni and S. Girard. ‘A refined Weissman estimator for extreme quantiles’. In: Compstat 2022 - 24th International Conference on Computational Statistics. Bologna, Italy, 23rd Aug. 2022. URL: <https://hal.science/hal-03764409>.
- [40] M. Allouche, J. El Methni and S. Girard. ‘A refined Weissman estimator for extreme quantiles’. In: EcoSta 2022 - 5th International Conference on Econometrics and Statistics. Kyoto, Japan, 4th June 2022. URL: <https://hal.science/hal-03688584>.
- [41] M. Allouche, S. Girard and E. Gobet. ‘Estimation of extreme quantiles from heavy-tailed distributions with neural networks’. In: CMStatistics 2022 - 15th International Conference of the ERCIM WG on Computational and Methodological Statistics. London, United Kingdom, 17th Dec. 2022. URL: <https://hal.science/hal-03910645>.
- [42] M. Allouche, S. Girard and E. Gobet. ‘EV-GAN: Simulation of extreme events with ReLU neural networks’. In: EcoSta 2022 - 5th International Conference on Econometrics and Statistics. Kyoto, Japan, 4th June 2022. URL: <https://hal.science/hal-03689703>.
- [43] S. Borkowski, S. Douté, F. Forbes and S. Heidmann. ‘Planet-GLLiM: software for scalable Bayesian analysis of multidimensional data in astrophysics’. In: Journées de l’Action Spécifique Numérique Astrophysique (ASNUM 2022). Lyon, France, 12th Dec. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03883020>.

- [44] M. Bousebata, S. Girard and G. Enjolras. ‘Extreme partial least-squares’. In: EcoSta 2022 - 5th International Conference on Econometrics and Statistics. Kyoto, Japan, 4th June 2022. URL: <https://hal.science/hal-03692286>.
- [45] J. El Methni and S. Girard. ‘A refined extreme quantiles estimator of Weibull tail-distributions’. In: CMStatistics 2022 - 15th International Conference of the ERCIM WG on Computational and Methodological Statistics. London, United Kingdom, Dec. 2022. URL: <https://hal.science/hal-03910631>.
- [46] F. Forbes. ‘Automatic learning of functional summary statistics for approximate Bayesian computation’. In: O’Bayes 2022 - Objective Bayes Methodology Conference. Santa Cruz, United States, 6th Sept. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03874003>.
- [47] F. Forbes. ‘Learning approaches for Bayesian inverse problems’. In: UGA-McMaster joint workshop 2022. Hamilton, Canada, 20th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03874054>.
- [48] F. Forbes. ‘Online Majorization Minimization algorithms’. In: CIRM Workshop on Computational Methods for Unifying Multiple Statistical Analyses (Fusion). Luminy, France, 24th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03874028>.
- [49] F. Forbes. ‘Simulation based inference for high dimensional inverse problems: application to magnetic resonance fingerprinting’. In: Colloque Intelligence Artificielle et santé : approches interdisciplinaires 2022. Nantes, France, 29th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03876812>.
- [50] F. Forbes. ‘Simulation-based approaches to Bayesian inverse problems’. In: ICMS workshop on Interfacing Bayesian statistics, machine learning, applied analysis, and blind and semi-blind imaging inverse problems. Edimburgh, United Kingdom, 24th Jan. 2023. URL: <https://hal.archives-ouvertes.fr/hal-03874057>.
- [51] F. Forbes. ‘Summary statistics and discrepancy measures for approximate Bayesian computation via surrogate posteriors’. In: BayesComp 2023. Levi, Finland, 15th Mar. 2023. URL: <https://hal.archives-ouvertes.fr/hal-03874011>.
- [52] F. Forbes, H. D. Nguyen, T. Nguyen and J. Arbel. ‘Mixture of expert posterior surrogates for approximate Bayesian computation’. In: SFdS 2022 - 53èmes Journées de Statistique de la Société Française de Statistique. Lyon, France, 13th June 2022, pp. 1–6. URL: <https://hal.archives-ouvertes.fr/hal-03679688>.
- [53] B. Lambert, F. Forbes, S. Doyle, A. Tucholka and M. Dojat. ‘Fast Uncertainty Quantification for Deep Learning-based MR Brain Segmentation’. In: EGC 2022 - Conference francophone pour l’Extraction et la Gestion des Connaissances. Blois, France, 24th Jan. 2022, pp. 1–12. URL: <https://hal.archives-ouvertes.fr/hal-03498120>.
- [54] B. Lambert, F. Forbes, A. Tucholka, S. Doyle and M. Dojat. ‘Multi-Scale Evaluation of Uncertainty Quantification Techniques for Deep Learning based MRI Segmentation’. In: ISMRM-ESMRMB & ISMRT 2022 - 31st Joint Annual Meeting International Society for Magnetic Resonance in Medecine. London, United Kingdom, 7th May 2022, pp. 1–3. URL: <https://hal.archives-ouvertes.fr/hal-03578023>.
- [55] H. LBATH, A. Petersen and S. Achard. ‘Large-Scale Correlation Screening Under Dependence for Brain Functional Connectivity Inference’. In: JSM 2022 - Joint Statistical Meetings. Washington, United States, 6th Aug. 2022. URL: <https://hal.inria.fr/hal-03867423>.
- [56] H. LBATH, A. Petersen, W. Meiring and S. Achard. ‘Clustering-Based Inter-group Correlation Estimation’. In: ICSDS 2022- IMS International Conference on Statistics and Data Science. Florence, Italy, 13th Dec. 2022. URL: <https://hal.inria.fr/hal-03867463>.
- [57] J. Linhart, P. L. Coelho Rodrigues, T. Moreau, G. Louppe and A. Gramfort. ‘Neural Posterior Estimation of hierarchical models in neuroscience’. In: GRETSI 2022 - XXVIIIème Colloque Francophone de Traitement du Signal et des Images. Nancy, France, 6th Sept. 2022, pp. 1–3. URL: <https://hal.archives-ouvertes.fr/hal-03858828>.

- [58] J. Linhart, A. Gramfort and P. L. Coelho Rodrigues. ‘Validation Diagnostics for SBI algorithms based on Normalizing Flows’. In: NeurIPS 2022 - the 36th conference on Neural Information Processing Systems - Machine Learning and the Physical Sciences workshop. New Orleans, United States, 28th Nov. 2022, pp. 1–7. URL: <https://hal.archives-ouvertes.fr/hal-03856444>.
- [59] T. Moins, J. Arbel, A. Dutfoy and S. Girard. ‘A local version of R-hat for MCMC convergence diagnostic’. In: SFdS 2022 - 53èmes Journées de Statistique de la Société Française de Statistique. Lyon, France, 13th June 2022, pp. 1–6. URL: <https://hal.inria.fr/hal-03683927>.
- [60] T. Moins, J. Arbel, A. Dutfoy and S. Girard. ‘Improving MCMC convergence diagnostic with a local version of R-hat’. In: CMStatistics 2022 - 15th International Conference of the ERCIM WG on Computational and Methodological Statistics. London, United Kingdom, 17th Dec. 2022. URL: <https://hal.science/hal-03910658>.
- [61] T. Moins, J. Arbel, A. Dutfoy and S. Girard. ‘On the use of a local R-hat to improve MCMC convergence diagnostic’. In: Energy Forecasting Innovation Conference 2022. Londres, United Kingdom, 24th May 2022. URL: <https://hal.inria.fr/hal-03683896>.
- [62] T. Nguyen, F. Chamroukhi, H. D. Nguyen and F. Forbes. ‘Model selection by penalization in mixture of experts models with a non-asymptotic approach’. In: JDS 2022 - 53èmes Journées de Statistique de la Société Française de Statistique (SFdS). Lyon, France, 13th June 2022, pp. 1–6. URL: <https://hal.archives-ouvertes.fr/hal-03663348>.
- [63] J. Saracco, H. Lorenzo and S. Girard. ‘Advanced topics in Sliced Inverse Regression V2’. In: JMVA 2022 - 50th Anniversary Jubilee. Virtual Meeting, United States, 20th Apr. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03842478>.
- [64] A. Usseglio-Carleve, S. Girard and G. Stupfler. ‘Extreme conditional expectile estimation in heavy-tailed heteroscedastic regression models’. In: Insurance Data Science Conference 2022. Milan, Italy, 15th June 2022, pp. 1–65. URL: <https://hal.science/hal-03683646>.
- [65] P. Wolinski and J. Arbel. ‘Imposing Gaussian Pre-Activations in a Neural Network’. In: JDS 2022 - 53es Journées de Statistique de la Société Française de Statistiques (SFdS). Lyon, France, 13th June 2022. URL: <https://hal.science/hal-03853790>.

Edition (books, proceedings, special issue of a journal)

- [66] L. Alamichel, J. Arbel, D. Bystrova and G. Kon Kam King, eds. *Bayesian nonparametric mixtures inconsistency for the number of clusters*. 17th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03866522>.

Reports & preprints

- [67] S. Achard and I. Gannaz. *Whittle estimation with (quasi-)analytic wavelets*. 18th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03272326>.
- [68] L. Alamichel, D. Bystrova, J. Arbel and G. Kon Kam King. *Bayesian mixture models (in)consistency for the number of clusters*. 22nd Nov. 2022. DOI: 10.48550/arXiv.2210.14201. URL: <https://hal.archives-ouvertes.fr/hal-03866434>.
- [69] M. Allouche, S. Girard and E. Gobet. *Estimation of extreme quantiles from heavy-tailed distributions with neural networks*. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03751980>.
- [70] P. A. Dkengne Sielenou and S. Girard. *Estimation of upper bounds for return levels: Methodology and illustrations on univariate random sequences*. Inria - Research Centre Grenoble – Rhône-Alpes, 9th Feb. 2022. URL: <https://hal.inria.fr/hal-03562999>.
- [71] M. N. Hounkonnou and P. D. Sielenou. *Conservation laws for under determined systems of differential equations*. 4th Feb. 2022. URL: <https://hal.inria.fr/hal-03557495>.
- [72] B. Lambert, F. Forbes, A. Tucholka, S. Doyle, H. Dehaene and M. Dojat. *Trustworthy clinical AI solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis*. 7th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03806630>.

- [73] N. Mahouton and P. D. Sielenou. *Extremum conditions for functionals involving higher derivatives of several variable vector valued functions*. 4th Feb. 2022. URL: <https://hal.inria.fr/hal-03557517>.
- [74] T. Moins, J. Arbel, A. Dutfoy and S. Girard. *On the use of a local \hat{R} to improve MCMC convergence diagnostic*. 7th Mar. 2022. URL: <https://hal.inria.fr/hal-03600407>.
- [75] T. Moins, J. Arbel, S. Girard and A. Dutfoy. *Reparameterization of extreme value framework for improved Bayesian workflow*. 7th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03806159>.
- [76] M. Ohlmann, G. Poggiato, S. Dray, W. Thuiller, C. Matias and V. Miele. *Quantifying the overall effect of biotic interactions on species distributions along environmental gradients*. 9th Jan. 2023. URL: <https://hal.science/hal-03172480>.
- [77] G. Oudoumanessah, M. Dojat and F. Forbes. *Unsupervised scalable anomaly detection: application to medical imaging*. Inria Grenoble - Rhône-Alpes; Grenoble Institut des neurosciences, 21st Oct. 2022, pp. 1–43. URL: <https://hal.archives-ouvertes.fr/hal-03824951>.
- [78] K. Pitas and J. Arbel. *Cold Posteriors through PAC-Bayes*. 29th Sept. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03791457>.
- [79] P. Wolinski and J. Arbel. *Gaussian Pre-Activations in Neural Networks: Myth or Reality?* 25th Dec. 2022. DOI: [10.48550/arXiv.2205.12379](https://doi.org/10.48550/arXiv.2205.12379). URL: <https://hal.science/hal-03933169>.

Other scientific publications

- [80] L. Alamichel, D. Bystrova, J. Arbel and G. Kon Kam King. ‘Bayesian mixture models (in)consistency for the number of clusters’. In: 13th Bayesian nonparametrics (BNP) conference. Puerto Varas, Chile, 24th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03866441>.
- [81] L. Carboni, M. Dojat and S. Achard. ‘Nodal statistics-based structural pattern detection for graph collections characterization’. In: Conference of Complex System. Palma de Mallorca, Spain, 17th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03866319>.
- [82] T. Moins, J. Arbel, S. Girard and A. Dutfoy. ‘A Local Version of R to Improve MCMC Convergence Diagnostic’. In: BAYSM 2022 - Bayesian Young Statisticians Meeting. Montréal, Canada, 22nd June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03765339>.
- [83] T. Moins, J. Arbel, S. Girard and A. Dutfoy. ‘A Local Version of R-hat to Improve MCMC Convergence Diagnostic’. In: ISBA 2022 - World Meeting of the International Society for Bayesian Analysis. Montréal, Canada, 27th June 2022, pp. 1–1. URL: <https://hal.archives-ouvertes.fr/hal-03765331>.

12.3 Other

Scientific popularization

- [84] T. Nguyen, H. D. Nguyen, F. Chamroukhi and F. Forbes. ‘A non-asymptotic approach for model selection via penalization in high-dimensional mixture of experts models’. In: *Electronic Journal of Statistics* 16.2 (27th Sept. 2022), pp. 4742–4822. DOI: [10.1214/22-EJS2057](https://doi.org/10.1214/22-EJS2057). URL: <https://hal.archives-ouvertes.fr/hal-03214606>.

12.4 Cited publications

- [85] H. Lu, J. Arbel and F. Forbes. ‘Bayesian nonparametric priors for hidden Markov random fields’. In: *Statistics and Computing* 30 (2020), pp. 1015–1035. DOI: [10.1007/s11222-020-09935-9](https://doi.org/10.1007/s11222-020-09935-9). URL: <https://hal.archives-ouvertes.fr/hal-02163046>.
- [86] J.-B. Durand and S. Achard. *Comprendre un processus cognitif grâce à l’analyse statistique du mouvement des yeux*. Tech. rep. Interstices, INRIA, hal-03505113, 2021.

-
- [87] P. Embrechts, C. Klüppelberg and T. Mikosh. *Modelling Extremal Events*. Vol. 33. Applications of Mathematics. Springer-Verlag, 1997.
 - [88] F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, Springer, 2006.
 - [89] S. Girard. 'Construction et apprentissage statistique de modèles auto-associatifs non-linéaires. Application à l'identification d'objets déformables en radiographie. Modélisation et classification'. PhD thesis. Université de Cergy-Pontoise, Oct. 1996.
 - [90] K. Li. 'Sliced inverse regression for dimension reduction'. In: *Journal of the American Statistical Association* 86 (1991), pp. 316–327.
 - [91] J. Simola, J. Salojärvi and I. Kojo. 'Using hidden Markov model to uncover processing states from eye movements in information search tasks'. In: *Cognitive Systems Research* 9.4 (Oct. 2008), pp. 237–251.