Inria International programme
Associate Team proposal 2019-2021
Submission form

Title: Latent Analysis, Adversarial Networks, and DimEnsionality Reduction

Associate Team acronym: LANDER

Principal investigator (Inria): Florence Forbes, Mistis Inria Grenoble Rhône-Alpes, France

Principal investigator (Main team): Hien Nguyen, La Trobe University, Melbourne (Bundoora), Australia

Other participants:

- Queensland University of Technology, Brisbane, Australia
- University of Queensland, Brisbane, Australia: *The University of Queensland is the only Australian university to have obtained the top ranking of 5 ("well above world standard") in Statistics in all three Excellence in Research for Australia (ERA) evaluation exercises to date.*
- Swinburne University of Technology, Melbourne, Australia
- University of Caen, France

Key Words: A- Research themes on digital science: A6.2.4. Méthodes statistiques, A3.4.2. Apprentissage non supervisé, A3.4.5. Méthodes bayésiennes, A6.3.1. Problèmes inverses, A9.3. Analyse de signaux (vision, parole, etc.)

B- Other research themes and application areas: B9.4.5. Science des données, B2.6.1. Imagerie cérébrale, B3.3. Géosciences, B9.9.1. Risques environnementaux.

1 Partnership

1.1 Detailed list of participants

Inria Mistis members

- **Julyan Arbel**: Junior researcher, Mistis, Inria Grenoble Rhône-Alpes. His interests in the project are related to Bayesian parametric and nonparametric statistics. Julyan will also supervise the PhD work of F. Boux and V. Munoz with F. Forbes.
  http://www.julyanarbel.com/

- **Florence Forbes**: Senior researcher, Head of the Mistis team, Inria Grenoble Rhône-Alpes. Her interests in the project are related to clustering and regression techniques; in particular in non standard cases, including high dimensionality, heavy tail modeling and missing observations. She has been also working on medical imaging applications for many years.
  http://mistis.inrialpes.fr/people/forbes

- **Stéphane Girard**: Senior researcher, Mistis, Inria Grenoble Rhône-Alpes. His interests in the project are related to nonparametric statistics, regression and statistical learning in high dimension. Stéphane will also supervise the PhD work of B. Kuegler with F. Forbes
  http://mistis.inrialpes.fr/people/girard

- **Alexis Arnaud**: Post-doc, Mistis, Inria Grenoble Rhône-Alpes. Alexis' thesis was advised by F. Forbes, E. Barbier and B. Lemasson (GIN). His work dealt with non Gaussian clustering and learning techniques for brain multiparametric magnetic resonance (MR) analysis. Alexis will stay another year in Mistis as a post-doc to finalize ongoing work and software that will be used by the partners.
  http://mistis.inrialpes.fr/people/people/arnaud/

- **Benoit Kuegler**: PhD student, Mistis and Grenoble Planetology Institute (IPAG). Benoit started in September 2018 and is co-advised by Sylvain Douté (IPAG) and S. Girard and F. Forbes (Mistis). The objective is to implement a statistical learning technique capable of solving a complex inverse problem in planetary remote sensing. The main difficulties are the large number of observations to be reversed, their large size, the need to predict predictions for several correlated parameters and to provide quality and uncertainty measures.

- **Veronica Munoz**: PhD student, Mistis and Grenoble Institute of Neuroscience (GIN). Veronica started in October 2017. She is co-advised by M. Dojat (GIN), J. Arbel and F. Forbes (Mistis). The goal is to develop a pipeline based on robust non Gaussian clustering and model selection tools to exploit quantitative multiparametric MR data that integrate several microvascular MRI parameters and can be used to design new biomarkers for Parkinson's disease.

- **Fabien Boux**: PhD student, Mistis and Grenoble Institute of Neuroscience (GIN). Fabien started in September, 2017. He is co-advised by E. Barbier (GIN), J. Arbel and F. Forbes (Mistis). In his thesis we target the analysis of MR fingerprint data with regression and inversion techniques.

- **Steven Quinito:** Engineer, SED Inria Grenoble Rhône-Alpes, until December 2019. Steven is an expert in software development and will help with the implementation of

our methods in particular using GPU. A first specific target is the pipeline for anomaly detection developed during the PhD of Alexis Arnaud with a follow-up in the PhD of Veronica Munoz.

Other French members

- **Faicel Chamroukhi**: Professor at University of Caen and associate member of Inria team MODAL, his research interests include statistical learning, latent data models, multivariate and functional data analysis, pattern recognition, statistical signal and image processing, and their applications. F. Chamroukhi spent a year at INRIA Lille (MODAL) in 2015/16 and is since an associate member of the team. Some of his projects include INRIA partners (such as the upcoming ANR SMILES, nov-2018-2022).
  https://chamroukhi.users.lmno.cnrs.fr/

- **Trung Tin Nguyen**: PhD Student, University of Caen. Tin stars his PhD in October 2018 on a doctoral contract and is supervised by F. Chamroukhi. His thesis research is on unsupervised learning of feature hierarchies using deep mixtures of experts.

- F. Chamroukhi is also currently hiring a PhD student on an ANR contract (SMILES 2018-2022) and the expected starting date of the PhD is November 2018-January 2019. His/her thesis is on latent data models for large-scale clustering. This PhD student will also be involved in the team.

Australian partners

- **Luke Prendergast**: Associate Professor, Head of Department of Mathematics and Statistics, La Trobe University, Melbourne. Luke is an expert in dimensionality reduction techniques such as sliced inverse regression and single-index modelling. He has supervised 3 successful PhD candidates in dimensionality reduction.
  http://www.latrobe.edu.au/she/staff/profile?uname=lprendergast

- **Hien Nguyen**: Lecturer and Research Fellow, La Trobe University, Melbourne. Hien's research focuses on construction of computationally efficient and feasible algorithms for the estimation of mixture-type models, such as via minorization-maximization algorithms or maximum pseudolikelihood estimation. Hien completed is PhD in 2015 from University of Queensland and already has 210 citations (Google Scholar) for more than 35 publications including 25 journal papers. He received a very selective and prestigious Australian Research Council (ARC) DECRA fellowship in 2016. Hien supervises the PhD work of J. Bagnall.
  https://hiendn.github.io/

- **Geoff McLachlan**: Professor, University of Queensland, Brisbane. His interests in the project are related to classification, clustering and discriminant analyses, image analysis, machine learning, neural networks, and pattern recognition. The work of G. McLachlan has received over 42,000 citations on Google Scholar (Citations).
  https://people.smp.uq.edu.au/GeoffMcLachlan/

- **Sharon Lee**: Research Fellow, University of Queensland, Brisbane. Her interests in the project are related to flexible data modelling via skew mixture models. Sharon lends a significant amount of computational know-how to the project. Sharon Lee completed her PhD in 2014 and already has 566 citations (Google Scholar) for more than 19 publications

including 12 journal papers. She also received an ARC DECRA fellowship in 2015.
http://researchers.uq.edu.au/researcher/11675

- **Darren Wraith** : Senior Lecturer, Queensland University of Technology, Brisbane. His interests in the project are related to latent variable modelling and Bayesian statistics with application to medical imaging. He is one of the chief investigators in a project on prostate cancer radiotherapy that was funded in 2016 by the National Health and Medical Research Council (NHMRC), a highly competitive grant. ARC and NHMRC are the main funding organisations in Australia. D. Wraith is advising a post-doc Sun Yu in this context with Prof. A. Haworth from University of Sydney.
http://staff.qut.edu.au/staff/wraith/

- **Jessica Bagnall**: PhD student, La Trobe University. Jessica is supervised by H. Nguyen until 2020 and is currently working on a project revolving around the development of Boltzmann machine networks for statistical inference, and the use of archetypal analysis and similar techniques for data quantisation, especially in neuroimaging. Jessica also provides insight into the interpretation of neuroimaging results, from her previous research and studies in psychology.

- **Mason Terrett**: PhD student, La Trobe University. Mason is supervised by H. Nguyen until 2020. He is currently working on data mining and machine learning techniques for the analysis of satellite and radar imaging, for the purpose of invasive vegetation remote sensing. This project is in collaboration with the Commonwealth Scientific and Industrial Research Organisation of Australia (CSIRO).

- **Kai Qin**: Associate Professor and Director of the Intelligent Data Analytics Lab. at Swinburne University of Technology, Melbourne. Kai is currently interested in large-scale distributed learning and optimization on the basis of transfer learning and transfer optimization. In addition to his knowledge of deep neural networks, Kai will also bring to the project his expertise in optimization and computational techniques.
http://www.alexkaiqin.org/

1.2 Nature and history of the collaboration

The collaboration is based on three main points, in statistics, machine learning and applications: 1) clustering and classification (mixture models), 2) regression and dimensionality reduction (mixture of regression models and non parametric techniques) and 3) high impact applications (neuroimaging, satellite and radar imaging, and planetology). On the methodology point of view, the first topic is mainly led by the Australian partners while the second is led by the French partners although overlaps already exist.

G. McLachlan and S. Lee at the University of Queensland (UQ), D. Wraith at QUT, H. Nguyen at La Trobe, in Australia and F. Chamroukhi in Caen, have extensive expertise in clustering of non-Gaussian data, *e.g.* [14, 15, 21, 22, 27, 28]. This subject was more recently developed in Mistis and initiated with the post-doc (3 years) of D. Wraith in Mistis. D. Wraith graduated from QUT and is now a permanent researcher. He visited Mistis for 2 weeks back in 2015. This collaboration led to 2 journal papers [1, 2] and introduced a new subject in Mistis regarding the use of skew and heavy-tailed distributions in other fields, such as in regression with recent publications [11, 13]. The latter illustrates the complementary expertise brought by Mistis in non parametric statistics, Bayesian approaches and graphical Markov models to account for structures and dependencies between variables in datasets. Other common interests between the teams include MRI data analysis on the application side while Mistis is bringing

a new application field related to planetology. Lastly, K. Qin at Swinburne brings his complementary expertise in computational and optimization methods [31, 29, 30, 32].

F. Forbes visited Brisbane and Melbourne in April 2017, and initiated the idea of the proposed associated team with H. Nguyen (La Trobe), G. McLachlan (UQ), D. Wraith (QUT) and K. Qin (Swinburne). Soon after, in June 2017, H. Nguyen visited F. Chamroukhi in Caen and the Mistis team. Hien is now spending a month in Mistis (September 2018). Based on this planned association, we also obtained a FASIC funding from Hubert Curien Program that will be used by F. Forbes for a 2 week visit in Melbourne (October 2018).

In parallel, the partners applied to an AFRAN call (Australian-French Association for Research and Innovation) and received 2000 euros to organize a workshop entitled "Research School on Statistics and Data Science" to be held in June 2019 in Melbourne. This will be the third of an ongoing series of workshops that has been previously held in France. The first two editions were in Caen (attendees from more than 20 countries in 2018) and hosted by Prof. F. Chamroukhi. It is expected that this third edition will be the start of an alternating schedule between the two countries. This ongoing relationship will provide an annual avenue for researchers to strengthen bilateral academic networks.

We believe this associate team is a way to formalize our common interests in both methodological tools and applications into more publications and codes. Mistis has long term collaborations with experts in neuroimaging and MRI. The augmentation of the Mistis methodological expertise with that from the Australian partners would result in a more powerful group to address these challenging applications while also producing software tools for the statistics community (*e.g.* the numerous R packages produced by S. Lee, G. McLachlan and H. Nguyen).

2 Scientific program

2.1 Context

A wide class of problems from medical imaging [39, 43, 47, 50, 52] and astrophysics [11, 37, 41, 54] can be formulated as inverse problems. Solving an inverse problem means recovering an object Y (*e.g.* some parameters or a probability distribution) from indirect noisy observations X. It is often compounded by the presence of errors (noise) in data but also by other kinds of complexity such as the high dimensionality of the observations, the complex dependence structure of the parameters and the issue of missing data. Another challenge is to design numerical implementations that are computationally efficient, either leveraging high-performance computing such as GPU (Graphical Processing Unit) computing or building methods that make the best mathematical use of simulations or costly operations. We had initiated some work on GPU computing [3], while our main expertise is still essentially in methodology. Among probabilistic models, generative models have appealing properties to meet all the above constraints. They have been studied in various forms and rather independently, both in the statistical and machine learning literature, with different depths and insights, from the well established probabilistic graphical models to the more recent (deep) generative adversarial networks (GANs). The advantages of the latter are primarily computational and their disadvantages are the lack of theoretical statements [44, 51], in contrast to the former. However, recent results, such as that of [45] provide meaningful methods for interpreting the output of GANs. The overall goal of the collaboration is to build connections between statistical and machine learning tools used to construct and evaluate generative models with the resolution of real life inverse problems as a target. This induces in particular the need to scale models to high-dimensional data, while maintaining our ability to assess their correctness, typically the uncertainty associated with the provided solutions.

The more standard approaches are based on optimization techniques (eg. [4, 5]), possibly augmented with regularization tools (*e.g.* a priori knowledge) from statistical (Bayesian) perspectives (*e.g.* [50]). However, such methods cannot generally handle high-dimensional data and quantify uncertainty. A first direction of research is therefore to investigate beyond optimization techniques and consider non standard learning and regressions approaches, which are less commonly used to solve inverse problems. More specifically, we will focus on dimension reduction techniques [10, 11, 20] and latent variable modelling [12, 13, 24]. A second direction of research is to focus on the so-called implicit generative models [51] and to elaborate on the link between Approximate Bayesian Computation (ABC) in statistics and GANs in machine learning. The expectation is to design improved Bayesian inference tools for implicit generative models by combining the strengths of both approaches. At last, a third direction is to study the applicability of a recent variable selection procedure [38] to the estimation of the central subspace for dimension reduction in regression. The procedure in [38] relies on so-called knockoff variables and provides a very elegant tool. However, it is difficult to implement for real life inverse problems. This will also link with the two previous directions as we expect GANs to be useful to generate knockoffs.

2.2 Objectives (for the three years)

Task 1. Learning approaches for high-dimensional regression and scalable ABC. To handle regression tasks in high dimension, we adopt an inverse regression strategy, using two types of approaches, *i.e.* extensions of mixtures of experts (MoE) models [12, 13, 24, 14, 15] and sliced inverse regression (SIR) [40, 48]. Inverse regression techniques work quite well in terms of prediction performance but provide a point estimation of the parameters (usually a conditional mean) for which assessing uncertainty is not straightforward. Our Gaussian Locally Linear Mapping (GLLiM) model [12] and its extensions [13] predict a full distribution for the parameters but assume a parametric mixture model which may be far from the parameter posterior distribution as provided by non-parametric Monte Carlo Markov Chain (MCMC) or Approximate Bayesian Computation (ABC) methods. Starting from the work of [42] on ABC, our goal is therefore to design a new procedure that improves upon GLLiM by making it closer to Bayesian inversion (MCMC or ABC) while being more efficient in high dimension. Additional frameworks that will be investigated are the MoE models, which are similar to the GLLiM methodology and share many common features. For example, both can be easily estimated via MM-type (minorization-maximization; [19]) algorithms, which are numerically stable and convergent, and both are proved to generate universal approximating classes of mean functions [6] and conditional density models [9, 53, 6]. These models also generate design issues that we plan to address using model [12, 13] and variable [38] selection tools or Bayesian non-parametrics techniques [9, 34, 35]. At last, these derivations are initially based on Gaussian mixtures. We plan to design new regression models using more flexible distributions developed for clustering, such as generalized t and skew-t [1, 21, 22, 14] or non-elliptical distributions [2].

Task 2. Generative Adversarial Nets and Approximate Bayesian Computation. Two difficulties with ABC approaches are that the procedures may critically depend 1) on the choice of summary statistics and 2) on the choice of distances to compare them. In Task 1, we target the use of ABC to improve our regression models (GLLiM and MoE). This provides a first solution to the choice of summary statistics while improving in turn the scalability of ABC. In Task 1, the summaries are restricted to moments (*e.g.* means) and the distances to moment matching. In Task 2, the goal is to by-pass the choice of summaries by exploiting the fact that GLLiM and MoE models produce full parametric distributions. The latter can then be used as

summaries provided appropriate distances are used to compare them. In Task 2, we address the second difficulty by investigating various distances between probability distributions, some standard in statistics and others emerging from the Generative Adversarial Networks (GAN) literature. GANs [44] are rapidly growing generative models whose successes in deep learning so far have been less striking than that of discriminative networks but their connections with more standard statistical tools [51, 45, 46] are promising.

Task 3. Knockoff Sliced Inverse Regression. SIR includes a dimension reduction step which consists of estimating the central subspace. One difficulty is then to estimate the dimension of this subspace. To address this issue, we propose to develop an extension based on the so-called knockoff filters [38]. A first challenge is to adapt the framework [38] to SIR. This can be address by adopting a non-standard maximum likelihood formulation of SIR [10, 11, 40]. A second challenge is that knockoffs may be difficult to implement in practice in particular in very high-dimensional cases.

2.3 Work-program (for the first year)

Task 1. Learning approaches for high-dimensional regression and scalable ABC. F. Boux has been hired as a PhD student at the Grenoble Institute on Neuroscience (GIN) in September, 2017. He is co-advised by E. Barbier (GIN), F. Forbes and J. Arbel (Mistis). The goal is to exploit an emerging technique in MRI called MR fingerprinting [49]. The task can be formulated as an inverse problem. Previous approaches have been based on grid search methods [47] but they are likely to be unstable and become intractable for even a moderate number of parameters. As an alternative, we plan to develop new methods based on high-dimensional regression models with an additional ABC step to improve prediction accuracy. Finding the best models will require the joint expertise of the Mistis, Caen [12, 13, 41, 14, 15] and the Australian partners [18, 20, 23, 24]. The results of the investigation on the GIN data will then also benefit other neuroimaging researchers at the Centre for Advanced Imaging at UQ to which H. Nguyen and G. McLachlan are connected. This work will involved mainly F. Boux, A. Arnaud, J. Arbel and F. Forbes at Mistis; F. Chamroukhi, Trung Tin Nguyen in Caen, H. Nguyen and L. Prendergast at La Trobe; and G. McLachlan at UQ. Our partners at the GIN listed in section 5.2 will also be involved for the application to MRI.

Task 2. Bayesian inference in implicit models. We aim at generalizing ABC from standard moment summaries to full distribution summaries. This requires us to investigate theoretical properties as in [42] and to compare in practice different divergence measures between probability distributions [36]. More generally, there are some opportunities for exchange between GANs and ABC with regard to scalability, applications and theoretical understanding. We will then also investigate whether ideas from GANs could complement the work planned in Task 1 above. This task will mainly be initiated by J. Arbel and F. Forbes at Mistis; H. Nguyen at La Trobe; and D. Wraith at QUT.

Task 3. Knockoff sliced inverse regression. This SIR extension originally proposed in the PhD Thesis of A. Chiancone under the supervision of S. Girard in the Mistis team, makes use of knockoff filters [38]. The underlying idea of [38] is to construct copies (knockoffs) of the original covariates. From the comparison between the true and the fake variables, some information can then be used to decide weather a variable is active or not in the specific regression framework. One difficulty is to generate the knockoff variables in an efficient way, all the more so as the largest improvement is expected when the number of variables to duplicate is high.

We aim at studying whether GANs could help in generating the knockoffs. More generally, SIR will be considered as an alternative to Mixtures of Experts in the starting PhD of B. Kuegler in collaboration with the Grenoble Planetology Institute. This task will mainly be initiated by S. Girard and F. Forbes at Mistis; H. Nguyen and L. Prendergast at La Trobe.

Visits on associate team funding

- Visit of J. Bagnall, PhD student, to Mistis: 2 weeks, to work on neuroimaging projects.
- Visit of B. Kuegler, PhD student, to La Trobe: 2 weeks, to work on Task 3
- Visit of A. Arnaud, Post-doc to UQ and QUT: 2 weeks, to work on Task 1 at UQ and multiparametric MRI application (QUT).
- Visit of J. Arbel to UQ and QUT: 2 weeks, to work on Tasks 1 and 2.
- Visit of F. Forbes to La Trobe: 1 week to work on Task 2
- Visit of F. Chamroukhi to La Trobe: 1 week to work on Task 1.
- Visit of Trung Tin Nguyen to QUT: 2 weeks to work on Task 1.

Visits on other funding

- Visit of M. Terrett, PhD student, to Mistis: 2 weeks on CSIRO funds to work on Task 3.
- Visit of F. Forbes to La Trobe: 1 week on AFRAN funds (see Section 3.1)) for the workshop.
- Visit of F. Chamroukhi to La Trobe: 1 week on AFRAN funds for the workshop.
- Visit of D. Wraith on Lab. Jean Kuntzman funds (not secured yet): 3 weeks in Fall 2019 to work on Task 1.
- Visit of S. Yu to Mistis (using Australian NHRMC grant, see Section 3.1): 3 weeks to exchange on statistical methods for quantitative multiparametric MRI.

If additional co-funding available

- Visit of G. McLachlan to Mistis: 2 weeks to work on Task 1.
- Visit of F. Boux, PhD student, to La Trobe: 2 weeks, to work on neuroimaging projects.

3 Budget

3.1 Budget (for the first year)

The estimation of the budget for 2019 is done using the following data:

- Cost for welcoming a researcher (*e.g.* PhD) at Inria: 3000 euros / month.
- Return ticket France - Australia: 1500 euros

Estimated costs: **25000 euros**

- 2-week visit of J. Bagnall, PhD student , to Mistis.
  Cost = 1500 + 1500 = 3000 euros.
- 2-week visit of M. Terrett, PhD student , to Mistis.
  Cost = 1500 + 1500 = 3000 euros (**CSIRO funded**).
- 2-week visit of B. Kuegler, PhD student , to Melbourne.
  Cost = 1500 + 1500 = 3000 euros.
- 2-week visit of A. Arnaud, Post-doc , to Brisbane.
  Cost = 1500 + 1500 = 3000 euros.
- 2-week visit of F. Forbes (Mistis) to Melbourne.
  Cost = 1500 + 1500 = 3000 euros (1500 euros on **AFRAN funds**).
- 2-week visit of F. Chamroukhi (Caen) to Melbourne.
  Cost = 1500 + 1500 = 3000 euros (1500 euros on **AFRAN funds**).
- 3-week visit of D. Wraith to Mistis **on LJK funds**.
  Cost = 2000 + 1500 = 3500 euros.
- 3-week visit of S. Yu, PhD student to Mistis **on NHMRC funds**.
  Cost = 2000 + 1500 = 3500 euros.

Funding requested from Inria: 25000 - 13000 = **12000 euros**

Other secured funding

Darren Wraith is one of the chief investigators in the BiRT project (Biofocussed prostate cancer RadioTherapy): a project that was funded by the National Health and Medical Research Council (NHMRC). The grant is worth 753,000 Australian dollars over 5 years. The project is mainly on the medical physics side but the link with our project is on the use of the latest statistical approaches. The lead researcher is Prof. Annette Haworth from University of Sydney (http://sydney.edu.au/science/people/annette.haworth.php).
Expected cost for 2019: **3000 euros**.

Hien Nguyen is the coordinator of a project funded by AFRAN (Australian-French Association for Research and Innovation) whose main objective is to organize a workshop in June 2019. Funding: **2000 euros** . Internal funding of **3000 euros** has also been obtained, from La Trobe, for the purpose of hosting this workshop entitled "Research School on Statistics and Data Science".

Funding applied for

For the workshop expenses, a further **6000 euros** is currently sought from the AMSI (Australian Mathematical Science Institute) workshop funding scheme.

Travel funding and conference expenses are currently being sought for the work of M. Terrett at CSIRO.

In the context of developing collaborations between Data Institutes worldwide, the Grenoble Data Institute (Data@UGA project) is likely to fund part of the travelling expenses for Grenoble researchers to Melbourne. Indeed, the goal of F. Forbes visit to Melbourne in October 2018 is twofold. First the goal is to initiate and consolidate research collaboration with La Trobe related to the work of Mistis' PhD students, Benoit Kuegler, Fabien Boux, Veronica Munoz and Alexis Arnaud. Then, while in Melbourne, F. Forbes, as a member of Grenoble Data institute, will also visit K. Qin at Swinburne. Regarding the later, a specific target is to establish collaborations between Swinburne Data Science Research Institute and Grenoble Data Institute. Further connections will be sought between Grenoble Data Institute and the Centre for Data Analytics and Cognition at La Trobe. Among various opportunities to involve different data institutes, the idea would be to send Master students from Grenoble to Melbourne and to permit Australian students to study in Grenoble. Expected funds for 2019: **3000 euros**.

3.2 Strategy to get additional funding

No additional funding has been secured yet. We are still looking for opportunities. To our knowledge, the only regular funding program between France and Australia are listed on the French Embassy web site (link). It seems that there are two programs we could apply for but the respective grants cannot be cumulated. The scientific mobility program will only cover the expense of one economy return airfare. The other program is the FASIC that covers only a France to Australia mobility and is for PhDs after 2002 (maximum amount 7800 euros).

Other possibilities include the Inria international chair program where G. McLachlan could apply.

Besides, it seems possible to get some funding (travel money) locally for the Australian researchers for mobility and career development. Typically, in 2015 Darren Wraith got some funding from his university to spend 2 weeks in Grenoble.

We hope the existence of the associated team will help in the future to more easily obtain this kind of funding by formalizing and increasing the visibility of the collaboration.

4 Added value

The proposed associated team aims at strengthening the collaboration between Inria team Mistis, University of Caen, University of La Trobe and Swinburne University of Technology in Melbourne and University of Queensland and Queensland University of Technology in Brisbane. All teams have expertise in statistical methods and complementary expertise in cluster analysis and latent variable models on one hand and high-dimensional regression and structured data on the other hand. The Australian partners have considerable expertise in clustering of non Gaussian data and statistical model selection while the Mistis team provides complementary expertise in non parametric statistics, Bayesian approaches, graphical Markov models to account for structure and dependencies between variables in datasets. On the application side, both the Mistis and Australian partners have been heavily engaged in neuroimaging and MRI research. G. McLachlan and H. Nguyen have access to the resources and knowhow from the Centre for Advanced Imaging at UQ, and H. Nguyen has strong connections with the SIMEXP group at

the CRIUGM in Montreal and the Mouse Imaging Centre in Toronto. D. Wraith is supervising the post-doc of Sun Yu (NHMRC project) on the analysis of multiparametric MR data [25, 26] with similar data and issues encountered in the PhD theses of A. Arnaud, F. Boux and V. Munoz (Mistis and GIN). The Mistis team has for more than 10 years now an active collaboration with the Grenoble Institute of Neuroscience.

The added value of our association will be to collectively combine our resources and data in order to develop tools that are more ubiquitous and universal than we could have previously make, each on our own. The association will result in significant progress not only in medical imaging but potentially in a number of other application domains with high scientific, societal and economic impact, such as several other areas of signal, image and video analysis (*e.g.* remote sensing, audiovisual processing) and biology (*e.g.* plant modelling). The Mistis team will be able to provide its long term relationship with experts in the latter mentioned applications. Through his collaborations with CSIRO, H. Nguyen has also access to state of the art radar and satellite imaging data.

Also we hope to be able to exchange students on a regular basis, ongoing PhD students for short visits but also post-doc students that would have defended their PhD in one team and go to the other one for a post-doc. Another objective could be to organize a regular workshop on the associated team topics. A workshop is already planned in June 2019 in Melbourne (AFRAN funds) and could be followed or complemented yearly with a special session at the ERCIM-CMSTATS or JSM conferences with the team partners.

More generally, we will utilize the funds to seek and formalize further collaborations, and to organize workshops and lectures by additional participants who will help further the goals of the project. See section 5.2 below for our currently identified "extended" partners.

5 Other remarks

5.1 Industrial Links, impact and results

The Pixyl startup (http://pixylmedical.com/) was created in 2015 by Mistis and Grenoble Institute of Neuroscience (GIN) members. The company has still strong links with Mistis and GIN and would be able to use the development planned in the project especially for MRI applications. However, this is not the first goal of this proposal.

5.2 Others

5.2.1 Extension to other partners

Upon analysis of the current team proposal, and projecting into the future, we have identified partners who are not included in this first proposal that are likely to be included in the coming years.

Other data scientists. This project will fit in a larger framework involving other teams. This includes other INRIA teams: Antoine Deleforge recently joined an Inria team in Nancy and Mistis maintains strong collaboration with the Perception team in Grenoble; also Jerome Saracco from CQFD team in Bordeaux is an expert in inverse regression (SIR). We may also include Emeline Perthame from Institut Pasteur (https://research.pasteur.fr/fr/member/emeline-perthame/), a former Mistis post-doc and other Australian teams in Melbourne. We identified two researchers whose work on Bayesian statistics could be very relevant to the project, David Frazier and Gael Martin from Monash University in Melbourne.

Applied domain experts. We are involved in some common work with the Grenoble Institute of Neuroscience. Our main collaborators there are E. Barbier, B. Lemasson, J. Warnking and M. Dojat. We also have some collaboration [37, 41] with experts in astrophysics in Grenoble, including S. Douté (B. Kuegler PhD) and D. Fraix-Burnet, and F. Schmidt in Orsay (Multi-PlanNet network). S. Girard and J. Arbel (Mistis) are co-organizing with D. Fraix-Burnet the school of statistics for astrophysics every two years. We are then regularly exposed to some of their current issues (*e.g.* [54]) that can be seen as inverse problems and for which we will propose new solutions. Also we would interact with other Australian medical experts (Prof. A. Haworth team) from University of Sydney through the collaboration with D. Wraith.

6 References

6.1 Joint publications of the partners

Publications from Mistis and Queensland University of Technology:

[1] F. Forbes and D. Wraith. A new family of multivariate heavy-tailed distributions with variable marginal amounts of tailweight: application to robust clustering. *Statistics & Computing*, 24(6):971–984, 2014.

[2] D. Wraith and F. Forbes. Location and scale mixtures of Gaussians with flexible tail behaviour: Properties, inference and application to multivariate clustering. *Computational Statistics & Data Analysis*, 90:61–73, 2015.

Publications from Swinburne and Mistis:

[3] K. Qin, F. Raimondo, F. Forbes, Yew Soon Ong, An Improved CUDA-Based Implementation of Differential Evolution on GPU, Genetic and Evolutionary Computation Conference 2012 (Gecco 2012). Nominated for the best paper award (finalist) in the Digital Entertainment Technologies and Arts / Parallel Evolutionary Systems session.

[4] K. Qin, F. Forbes. Harmony Search with Differential Mutation Based Pitch Adjustment, Genetic and Evolutionary Computation Conference 2011 (Gecco 2011).

[5] K. Qin, F. Forbes. Dynamic Regional Harmony Search with Opposition and Local Learning, Genetic and Evolutionary Computation Conference 2011 (Gecco 2011).

Publication from La Trobe, University of Caen and Mistis:

[6] H. Nguyen, F. Chamroukhi and F. Forbes. Approximation results regarding the multiple-output mixture of linear experts model. *Submitted to Neural Computation*, *arXiv:1704.00946*, 2018.

Publications from La Trobe and University of Caen:

[7] F. Chamroukhi and H. D. Nguyen. Model-Based Clustering and Classification of Functional Data. *arXiv:1803.00276v2*, 2018.

[8] H. D. Nguyen and F. Chamroukhi, Practical and theoretical aspects of mixture-of-experts modeling: An overview, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2018.

6.2 Main publications of the participants relevant to the project

Publications from Mistis:

[9] J. Arbel, K. Mengersen, and J. Rousseau. Bayesian nonparametric dependent model for partially replicated data: the influence of fuel spills on species diversity. *Annals of Applied Statistics*, 10(3):1496–1516. (2016).

[10] C. Bernard-Michel, L. Gardes, and S. Girard. Gaussian regularized sliced inverse regression. *Statistics & Computing*, 19(1):85–98, Mar 2009.

[11] A. Chiancone, F. Forbes, and S. Girard. Student Sliced Inverse Regression. *Computational Statistics & Data Analysis*, August 2016.

[12] A. Deleforge, F. Forbes, and R. Horaud. High-Dimensional Regression with Gaussian Mixtures and Partially-Latent Response Variables. *Statistics & Computing*, 25(5):893–911, September 2015.

[13] E. Perthame, F. Forbes, and A.Deleforge. Inverse regression approach to robust non-linear high-to-low dimensional mapping. *J. of Multivariate Analysis*, nov. 2017.

Publications from University of Caen:

[14] F. Chamroukhi Skew t mixture of experts, *Neurocomputing*, vol.266, 390-408, 2017.

[15] F. Chamroukhi. Robust mixture of experts modeling using the t-distribution, *Neural Networks - Elsevier*, Vol. 79, 20-36, 2016.

[16] F. Chamroukhi Piecewise Regression Mixture for Simultaneous Functional Data Clustering and Optimal Segmentation, *Journal of Classification*, vol. 33, 374-411, 2016.

[17] F. Chamroukhi. Unsupervised learning of regression mixture models with unknown number of components, *Journal of Statistical Computation & Simulation*, vol. 86, 2308-2334, 2016.

Publications from La Trobe University:

[18] Garnham, A. L., and Prendergast, L. A. A note on least squares sensitivity in single-index model estimation and the benefits of response transformations. *Electronic Journal of Statistics*, 7: 1983–2004, 2013.

[19] Nguyen, H. D. An introduction to Majorization-Minimization algorithms for machine learning and statistical estimation. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(2), 2017.

[20] Prendergast, L. A., and Garnham, A. L. Response and predictor folding to counter symmetric dependency in dimension reduction. *Australian & New Zealand Journal of*

Statistics, 58(4):515–532, 2016.

Publications from University of Queensland:

[21] S. Lee and G.J. McLachlan. Finite mixtures of multivariate skew t-distributions: some recent and new results. *Statistics & Computing*, 24:181-202, 2014.

[22] Lee, S.X. and McLachlan, G.J. Finite mixtures of canonical fundamental skew t-distributions: the unification of the restricted and unrestricted skew t-mixture models. *Statistics and Computing* 26, 573-589, 2016.

[23] L.R. Lloyd-Jones, H.D. Nguyen, and G.J. McLachlan. A globally convergent algorithm for lasso-penalized mixture of linear regression models. *Computational Statistics & Data Analysis*, 2017.

[24] Nguyen, H. D., McLachlan, G. J., and Wood, I. A. Mixtures of spatial spline regressions for clustering and classification. *Computational Statistics & Data Analysis*, 93: 76–85, 2016.

Publications from Queensland University of Technology:

[25] S, Yu, Reynolds, H., Wraith, D., Williams, S., Finnegan, M. E., Mitchell, C., et al. Predicting prostate tumour location from multiparametric MRI using Gaussian kernel support vector machines: A preliminary study. *Australasian Physical & Engineering Sciences in Medicine*, 40(1): 39–49, 2017.

[26] Haworth, A., Reynolds, H., Mears, C., Betts, J., Finnegan, M., DiFranco, M., et al. Focal brachytherapy treatment planning using multi-parametric MRI and biological dose optimisation. *Brachytherapy*, 14(Supplement 1), S11-S12, 2015.

[27] Maleki, M. Wraith, D. Arellano-Valle, R.B. A flexible class of parametric distributions for Bayesian linear mixed models. *TEST. (In Press).*

[28] Maleki, M., Wraith, D., Arellano-Valle, R. B. Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Statistics and Computing.* Published online: 19 May, 2018.

Publications from Swinburne University of Technology:

[29] Feng, L., Zhou, L., Zhong, J., Gupta, A., Ong, Y.-S., Tan, K. C., Qin, A. K. Evolutionary Multitasking via Explicit Autoencoding. *IEEE Transactions on Cybernetics*, 2018.

[30] Liu, J., Gong, M., Qin, A. K., Zhang, P. A deep convolutional coupling network for change detection based on heterogeneous optical and radar images, *IEEE Transactions on Neural Networks and Learning Systems*, 29(3): 545-559, 2018.

[31] Li, G. Q., Zeng, F., Li, H. Q., Qin, A. K. Matrix function optimization problems under orthonormal constraint, *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* 2017.

[32] B. Y. Qu, B. F. Lang, J. J. Liang, A. K. Qin and O. D. Crisalle, Two-hidden-layer extreme learning machine for regression and classification, , 175: 826-834, 2016.

[33] L. Wan, K. Tang, M. Li, Y. Zhong and A. K. Qin, Collaborative active and semi-supervised learning for hyperspectral remote sensing image classification, *IEEE Transactions on Geoscience and Remote Sensing*, 53(5): 2384-2396, 2015.

6.3 Other references

[34] M. Albughdadi, L. Chaari, J-Y. Tourneret, F Forbes, and P. Ciuciu. A Bayesian non-parametric hidden Markov random model for hemodynamic brain parcellation. *Signal Processing*, 135, 01 2017.

[35] J. Arbel, G. Gayraud, and J. Rousseau. Bayesian optimal adaptive estimation using a sieve prior. *Scandinavian Journal of Statistics*, 40(3):549–570, 2013.

[36] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 214–223, 2017.

[37] C. Bernard-Michel, S. Douté, M. Fauvel, L. Gardes, and S. Girard. Retrieval of Mars surface physical properties from OMEGA hyperspectral images using Regularized Sliced Inverse Regression. *Journal of Geophysical Research: Planets*, 114(E6), 2009.

[38] E. J. Candes and R. Foygel Barber. Controlling the false discovery rate via knockoffs. *The Annals of Statistics*, 43(5):2055–2085, 2015.

[39] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu. Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach. *IEEE Trans. Med. Imaging*, 32(5):821–837, 2013.

[40] D. Cook. Fisher lecture: Dimension reduction in regression. *Statistical Science*, 22(1):1–26, 2007.

[41] A. Deleforge, F. Forbes, S. Ba, and R. Horaud. Hyper-Spectral Image Analysis with Partially-Latent Regression and Spatial Markov Dependencies. *IEEE Journal on Selected Topics in Signal Processing*, 9(6):1037–1048, September 2015.

[42] Fearnhead, P. and Prangle, D. Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474, 2012.

[43] A. Frau-Pascual, F. Forbes, and P. Ciuciu. Physiological models comparison for the analysis of ASL FMRI data. In *12th IEEE International Symposium on Biomedical Imaging, ISBI 2015*, pages 1348–1351, New York, United States, April 2015.

[44] I. J. Goodfellow, J.Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.

[45] Gutmann, M. U., Dutta, R., Kaski, S., Corander, J. Likelihood-free inference via classification. *Statistics and Computing*, 28(2), 411-425, 2018.

[46] Jiang, B. et al. Approximate Bayesian Computation with Kullback-Leibler Divergence as Data Discrepancy. *In International Conference on Artificial Intelligence and Statistics*, 711-1721, 2018.

[47] B. Lemasson, N. Pannetier, N. Coquery, Ligia S. B. Boisserand, Nora Collomb, N. Schuff, M. Moseley, G. Zaharchuk, E. L. Barbier, and T. Christen. MR Vascular Fingerprinting in Stroke and Brain Tumors Models. *Scientific Reports*, 6:37071, November 2016.

[48] Ker-Chau Li. Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, 86(414):316–327, 1991.

[49] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk, and M. A. Griswold. Magnetic Resonance Fingerprinting. *Nature*, 495(7440):187–192, March 2013.

[50] P. Mesejo, S. Saillet, O. David, C. Bénar, J. M. Warnking, and F. Forbes. A differential evolution-based approach for fitting a nonlinear biophysical model to fMRI BOLD data. *IEEE Journal of Selected Topics in Signal Processing*, 10(2):416–427, March 2016.

[51] S. Mohamed and B. Lakshminarayanan. Learning in implicit generative models. *CoRR*, abs/1610.03483, 2016.

[52] G. Naturaj, J-F. Nielsen, and J.A. Fessler. Dictionary-free MRI parameter estimation via kernel Ridge Regression. *ISBI 2017*, 2017.

[53] Norets, A., and Pelenis, J. Posterior consistency in conditional density estimation by covariate dependent mixtures. *Econometric Theory*, 30(3): 606–646, 2104.

[54] F. Schmidt and J. Fernando. Realistic uncertainties on Hapke model parameters from photometric measurements. *Icarus*, 260:73–93 (IF 2,84), 2015.