Associate Team proposal 2023-2025
Submission form

Title: Variance-reduced Optimization Methods and Bayesian Approximation Techniques for scalable inference.

Associate Team acronym: WOMBAT

Principal investigator (Inria): Florence Forbes, Statify Inria Grenoble, France

Principal investigator (Main team): Hien Duy Nguyen, University of Queensland, Brisbane, Australia

Other participants:

- Queensland University of Technology, Brisbane, Australia
- Swinburne University of Technology, Melbourne, Australia
- Griffith University, Brisbane, Australia
- Monash University, Melbourne, Australia
- University of Toulouse, France
- INRAE, Jouy-en-Josas, 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 Statify members

- **Julyan Arbel**: INRIA Senior researcher. His interests in the project are related to Bayesian neural networks, Bayesian parametric and nonparametric statistics, and more specifically approximate Bayesian computation (ABC). http://www.julyanarbel.com/
- **Florence Forbes**: INRIA Senior researcher, Head of the Statify team. Her interests in the project are related to Bayesian parametric and nonparametric statistics, simulation-based inference (SBI), ABC, 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
- **Pedro Rodrigues**: INRIA Junior researcher. His interests in the project are related to Bayesian inverse problems, SBI, invertible neural networks and medical imaging applications. https://plcrodrigues.github.io/
- **TrungTin Nguyen**: Post-doc fellow. Tin started his post-doc in the team in January 2022. His interests in the project are related to ABC, mixtures of expert models, neural networks and statistical model selection. https://trung-tinnguyen.github.io/
- **Kostas Pitas**: Post-doc fellow. Kostas started his post-doc in the team in February 2022. His interests in the project are related to Bayesian neural networks and theoretical tools to study their properties. https://www.konstantinos-pitas.com/
- **Pierre Wolinski**: Post-doc fellow. Pierre started his post-doc in the team in January 2021. His interests in the project are related to Bayesian and approximate inference, in particular in combination with neural networks. http://pierre-wolinski.fr/
- **Louise Alamichel**: PhD student, Statify and INRAE. Louise is supervised by J. Arbel and G. Kon Kam King since October 2021. Her works deals with Bayesian nonparametric models for genetic data. Thesis description
- **Geoffroy Oudoumanessah**: PhD student, Statify, Grenoble Institute of Neuroscience (GIN) and Creatis lab. Geoffroy has been an intern in the team since Spring 2022 and has started his PhD in October 2022. He is co-advised by F. Forbes, M. Dojat (GIN) and C. Lartizien (Creatis). His work deals with online algorithms for non Gaussian clustering and learning techniques for brain magnetic resonance (MR) imaging and anomaly detection. Linkedin.

Other French members

- **Gersende Fort**: CNRS senior researcher, IMT, University of Toulouse. Her interests in the project are related to stochastic approximation and optimization, online and federated learning, sampling and Monte-Carlo Markov Chain (MCMC) techniques. https://perso.math.univ-toulouse.fr/gfort/
- **Guillaume Kon Kam King**: INRAE senior researcher. His interests in the project are related to approximate Bayesian computation, Bayesian nonparametrics and mixture models. Guillaume is also the co-advisor of Louise Alamichel. https://sites.google.com/site/guillaumekonkamking/

- **Jean-Baptiste Durand**: CIRAD Senior researcher, AMAP laboratory, Montpellier. Jean-Baptiste has been in the Statify and the former Mistis team at Inria since their creation. He is now, starting from September 2022, a researcher at CIRAD but will continue to collaborate with the team in particular on Bayesian modelling and applications. His interests in the project are related to approximate Bayesian inference and statistical model selection. New web page

Australian partners

- **Hien Duy Nguyen**: Senior Lecturer, University of Queensland, Brisbane. Hien focuses on optimization and computational techniques for estimation and inference in complex data models. He is particularly devoted to the study of majorization–minimization algorithms and stochastic variants, finite mixture models and related approaches, and non-asymptotic methods for constructing hypothesis tests and confidence sets. Having completed his PhD in 2015, his research output has consisted of 61 peer-reviewed papers, 13 conference papers, and 3 book chapters, which have collectively resulted in over 900 citations, and a H-index of 20. Hien previously was awarded a prestigious ARC DECRA Fellowship in 2016. https://hiendn.github.io/

- **Geoff McLachlan**: Professor, University of Queensland, Brisbane. Geoff's research primarily consists of inferential methods of mixture models and their applications in computer science, engineering, genetics, and medical settings. He is particularly known for his popular books: The EM Algorithm and Extensions (with T. Krishnan), Finite Mixture Models (with D. Peel), Analyzing Micro Gene Expression Data (with K-A Do, and C. Ambroise), Discriminant Analysis and Statistical Pattern Recognition, and Mixture Models: Inference and Applications to Clustering (with K. Basford). His research output consists of over 200 peer-reviewed papers and over 50 conference papers that have collectively resulted in over 56000 citations and a H-index of 61. In 2015, Geoff was honoured with a Fellowship of the Australian Academy of Science for his academic achievements and service. https://people.smp.uq.edu.au/GeoffMcLachlan/

- **Sharon Lee**: Senior Lecturer, University of Queensland, Brisbane. Sharon's research consists largely of the development of novel distributions for modeling complex heterogeneous data, with particular emphasis on the application and estimation of such models. Her primary output consists of the characterization of skewness and heavy tail phenomena via parametric probability models. Her work has appeared in top journals such as Statistics and Computing and IEEE TNNLS, where it has attracted over 1000 citations. Sharon was previously awarded a prestigious ARC DECRA Fellowship in 2015.
http://researchers.uq.edu.au/researcher/11675

- **Chris van der Heide**: Post-doc Fellow, University of Queensland, Brisbane. Since completing his PhD in 2016 at the University of Queensland, on the topic of partial differential equations, Chris has conducted research as part of the nationwide ARC Centre of Excellence for Mathematical and Statistical Frontiers. His research focuses on simulation and optimization problems in large scale AI systems such as neural networks and normalizing flow models [29]. His outputs include papers at top AI and machine learning outlets such as the Conference on Uncertainty in Artificial Intelligence, AAAI and AISTATS.
http://researchers.uq.edu.au/researcher/15832

- **Daniel Ahfock**: Post-doc Fellow, University of Queensland, Brisbane. Daniel obtained his PhD in 2019 at the University of Cambridge. He has since conducted research at the

University of Queensland, under the guidance of Prof. Geoff McLachlan on the topic of semi-supervised learning via finite mixture models and ML methods. He also has strong research interests in Bayesian methods and matrix sketching approaches [28]. His research has appeared in top journals such as Biometrika, Statistics & Computing, and Bayesian Analysis. https://scholar.google.com/citations?user=mesoVwkAAAAJ

- **Chris Drovandi**: Professor in Statistics and Data Science, Queensland University of Technology, Brisbane. Chris currently holds an ARC Future Fellowship for his current research in scalable and robust Bayesian methods for implicit statistical models. He has particular interest in likelihood-free Bayesian approaches such as Approximate Bayesian Computation and Bayesian Synthetic Likelihood, as well as sampling approaches, such as sequential Monte Carlo methods. Chris was previously a recipient of an ARC DECRA Fellowship in 2016, and his research has cumulatively resulted in over 2700 citations and a H-index of 22. https://chrisdrovandi.weebly.com/

- **Leah South (also known as Leah Price)**: Lecturer at Queensland University of Technology (QUT). Leah's work is directed towards the Bayesian inference topics of Monte Carlo variance reduction techniques, applications of Stein's method, scalable Monte Carlo methods, Approximate Bayesian Computation, and Sequential Monte Carlo sampling. Leah has previously conducted research as a Senior Research Associate at Lancaster University and is the Secretary of the Computation Section of the International Society for Bayesian Analysis. Since completing her PhD in 2019, Leah has accrued over 290 citations from publications in major outlets, such as Biometrika, Bayesian Analysis, and the Journal of Computational & Graphical Statistics. https://sites.google.com/view/leahsouth

- **Darren Wraith**: Associate Professor, Queensland University of Technology, Brisbane. Darren received his PhD in 2008 and spent five years working internationally, as a Postdoc at Universite Paris-Dauphine and as a Research Engineer at Inria Grenoble Rhone-Alpes. His research is spread across topics in statistical methodology, with particular focus on mixture models, Bayesian inference, and high dimensional analysis, as well as applied statistics and biostatistics, including research in problem gambling, environmental sciences, and public health. Darren's current research is funded by a prestigious ARC Linkage Project grant, and he has previously been funded by the NHMRC: Australia's primary health research funding agency. Darren's research has resulted in over 1500 citations and a H-index of 21. http://staff.qut.edu.au/staff/wraith/

- **Mitchell O'Sullivan** PhD student, co-advised by C. Drovandi and L. South. Mitchell aims to develop new adaptive sequential Monte-Carlo ABC (SMC ABC) algorithms and create a Julia package for SMC ABC methods, potentially including new methods developed by the partners, to make them more accessible for practitioners.
https://research.qut.edu.au/qutcds/staff/mitchell-osullivan/

- **Ryan Kelly** PhD student, co-advised by C. Drovandi and L. South on scalable and robust Likelihood-Free Bayesian Inference. Ryan aims to consider scalable approaches for likelihood-free inference, such as Bayesian synthetic likelihood (BSL) and neural network based approaches, and advance them to make them robust to model misspecification. One possible added value to the project is a similar approach to make the other partners' methods (ABC) more robust. https://ryankellyaus.com/

- **Kai Qin**: Professor, Swinburne University of Technology, Melbourne. Kai currently serves as the Director of the Swinburne Intelligent Data Analytics Lab, the Deputy Director of

the Space Technology and Industry Institute, and the Program Lead for Data Analytics in the Data Science Research Institute. His primary research interests revolves around the field of Computational Intelligence, where he is largely interested in computational paradigms for solving complex real-world problems for which traditional methods are infeasible or ineffective. Notably, he is interested in the study of neural networks and deep learning systems, as well as remote sensing and computer vision applications. Kai is the chair of the IEEE Neural Networks Technical Committee and the Vice-Chair of the IEEE Emergent Technologies Task Force on "Multitask Learning and Multitask Optimisation". His research has collectively resulted in over 14000 citations and a H-index of 40. http://www.alexkaiqin.org/

- **David Frazier**: Associate Professor in Econometrics and Business Statistics at Monash University, and a current Australian Research Council (ARC) Discovery Early Career Research Award (DECRA) Fellow. David's research primarily focuses on the theoretical and computational aspects of simulation-based Bayesian inference, where he has made useful contributions to the underlying theory of simulation-based inference procedures such as approximate Bayesian computation and Bayesian synthetic likelihood. His research has featured in the most prestigious journals in statistics including Biometrika, Journal of the American Statistical Association, and the Journal of the Royal Statistical Society: Series B. https://dtfrazier.netlify.app/

- **Shu-Kay Angus Ng:** Professor of Biostatistics, Griffith University, Brisbane. Angus serves as a senior biostatistician in the Griffith University School of Medicine and Dentistry, where he applies his expertise in mixture modeling, random effects models, and machine learning and survival analysis to develop inferential approaches for bioinformatics, medical imaging, health economics, and clinical trials. Angus was appointed to the ARC College of Experts in 2022, where he provides advice and guidance on the targets of national research funding. Academically, Angus has an output of over 160 publications, which have jointly resulted in over 11000 citations and a H-index of 40. https://experts.griffith.edu.au/19026-shu-kay-angus-ng

1.2 Nature and history of the collaboration

This proposal follows the previous Associate Team: LANDER (2019–2022), whose full description is available on the team website. In the present proposal, WOMBAT, we plan to focus on simulation-based inference (SBI) approaches combined with stochastic approximation (SA) tools. The team composition includes many of the previous LANDER permanent researchers. New faculty members include: Chris Drovandi and Leah South (also known as Leah Price) from QUT, Brisbane, Shu-Kay Angus Ng from Griffith University, Brisbane, David Frazier from Monash University, Melbourne, Gersende Fort from IMT, University of Toulouse, Guillaume Kon Kam King from INRAE and Pedro Rodrigues from Statify. Their expertise in modern statistical techniques and algorithms will be invaluable. In particular, D. Frazier, C. Drovandi and L. South already have a number of common works and they provide, with P. Rodrigues, capacity in SBI techniques. A. Ng, G. Fort and G. Kon Kam King lend their abilities in statistical modelling, SA and stochastic optimization theory, respectively.

Despite the COVID-19 pandemic, the productivity of LANDER, in terms of collaborative research output, has been commendable with a total of 15 peer-reviewed journal publications and 7 conference papers; cf. the publication page, but travelling restrictions have limited the success of our students and early career researchers exchange plans. Aside from new research output, which we believe to be excellent, we plan to make the exchange component a key aim of

our present project. Between the partner institutions, the current situation is quite favourable to achieving this goal, as there are a good number of starting PhD students and experienced post-doc fellows within the proposed team composition.

The complementary nature of the makeup of the previous LANDER team is detailed in the 2019 LANDER proposal. We thus limit our description of the new faculty members, here. G. Fort is an expert in SA, stochastic processes, and online optimization, complementing the gap in advanced expertise in these topics from the other French members. A new collaboration with G. Fort on these topics has been initiated in early 2022, and has resulted on some preliminary work with H. Nguyen and F. Forbes [11]. As part of the LANDER project, a 20-day visit to UQ and QUT in Brisbane was organized for G. Fort and F. Forbes in April 2022. On the visit, G. Fort gave a UQ Statistics, Modelling and Operations Research Seminar on the topic of stochastic MM algorithms.

Towards our SBI targets, the participation of P. Rodrigues and G. Kon Kam King is natural due to their highly relevant research in the area. P. Rodrigues joined the Statify team in 2021 as a junior INRIA researcher. He is already an expert in SBI approaches, where he brings a perspective that is more colored by machine learning and neural networks than the Bayesian modeling focus of the rest of the team. His computational know-how and his knowledge of neuroscience also lend useful insight to the team, as we consider applications of our developed methodology. G. Kon Kam King has strong collaboration with the Statify team on ABC and Bayesian nonparametrics. He is currently involved in the PhD thesis of Louise Alamichel.

The collaboration of D. Frazier, C. Drovandi and L. South is logical as they are constant and regular contributors towards SBI and ABC research, often being present at ABC sessions at major conferences. Of particular note, recently, F. Forbes was invited to give a talk on SBI and ABC at the QUT Center for Data Science, at the Distinguished Visitor Seminar, and both F. Forbes and C. Drovandi were recently invited to the ABC session at the O-Bayes conference in September 2022. We also expect great outcomes from the software development know-how of L. South and C. Drovandi, as exemplified, in particular, by the highly popular BSL R package that would be a natural locus for our collaborative contributions. Two of their current PhD students, K. Ryan and M. O'Sullivan will also be of great help in making our potential developments more efficient and more visible.

To further assisting in our SA, stochastic optimization, and scalable optimization research, we have also saught the participation of A. Ng, from Griffith University. A. Ng is actively engaged with research in stochastic algorithms for feasible and scalable optimization, and has a strong history of research, together with G. McLachlan, on the topic of scalable EM algorithms [53, 54]. A. Ng will also contribute to the project via his experience of conducting biostatistical inference for biological imaging studies.

2 Scientific program

2.1 Context

Many inferential tools, such as machine learning algorithms and statistical models, require the estimation of model parameters, structures, quantities, and properties, from data. In practice, it is common that model characterizations are available through high-fidelity simulations of the data generating processes, but only through "black-boxes" that are poorly suited for optimization under uncertainty or conventional statistical inference procedures. The main statistical challenge is that model likelihoods are typically intractable or unavailable in closed form. Approaches suited for these scenarios are typically referred to as likelihood-free or simulation-based inference (SBI) methods, and have received a great deal of attention in recent years, with momentum coming from mixing of ideas from the interface between statistics and machine learning [63, 75]. However, most SBI methods scale poorly when the number of observations is too large, which makes them unsuitable for modern data, which are often acquired in real time, in an incremental nature, and are often available in large volume.

Computation of inferential quantities in an incremental manner may be forcibly imposed by the nature of data acquisition (e.g. streaming and sequential data) but may also be seen as a solution to handle larger data volumes in a more resource friendly way, with respect to memory, energy, and time consumption. To produce feasible and practical online algorithms for streaming data and complex models, we propose to study the family of stochastic approximation (SA) algorithms [61] that encompass and extend upon a wide variety of current algorithmic frameworks for optimization-based fitting of statistical and machine learning models. We are particularly interested in modern approaches that incorporate variance reduction techniques, as per [84]. Further, we plan to investigate the possibility to efficiently re-use simulations from one inferential problems to another, a quality which is referred to as amortization, and particularly useful when simulations are costly. We will then aim at combining both features, by incorporating active or sequential learning schemes in the inference process. Among SBI approaches, we will focus on approximate Bayesian computation (ABC) [85], Bayesian synthetic likelihood (BSL) [86, 37], and invertible neural networks [70, 82].

The overall goal of the project is to combine recent ideas from the SBI and SA literature, to propose efficient methods for handling complex inferential problems. We shall demonstrate our approaches via applications to problems in challenging domains, such as Magnetic Resonance Imaging (MRI) (G. Oudoumanessah's PhD with GIN and Creatis [52, 56]) or road network management (collaborations with Dr. Long Truong, La Trobe University [57, 49]) as initial targets. So doing, we hope to achieve both breakthroughs in applied methodology and the development of new SBI and SA techniques that wide-spread applicability.

2.2 Objectives (for the three years)

Our objectives can be summarized into three categorical themes: 1) The combination of existing methods to mitigate their respective limitations and leverage their strengths; 2) The establishment of theoretical properties to understand and guarantee the performance of existing and new approaches; and 3) The Design of new approaches. **Details are given in the next section.**

1. Determination of the ways in which SBI principles and constraints can be combined with stochastic optimization approaches in order to generate useful and practical methods for inference of complex models. More specifically, we plan to consider:

1.1 Online and Mini-Batch Majorization–Minimization (MM) algorithms.

1.2 Stochastic multi-task learning.

2. Establishment of theoretical properties and extension of existing stochastic algorithms in order to understand their, capabilities, performance, and limitations. More specifically, we aim at establishing:

2.1 Generic ABC consistency-like results.

2.2 Online model selection tools for streaming data.

3. Design new amortized SBI schemes. More specifically, we target the development of:

3.1 Amortized ABC and BSL schemes.

3.2 Invertible neural networks for approximate Bayesian inference.

2.3 Work-program (for the first year)

During the first year, we will mainly focus on Tasks 1.1, 2.1 and 3.1, with the following directions. We also provide a brief description of the three other tasks.

1.1. Online and Mini-Batch Majorization–Minimization (MM) algorithms
Participants: Seniors AN, DW, FF, GF, GM, HD, SL. Post-docs CH, DA, TN, PW. Student GO.

The problems associated with the incremental access of data can be handled with stochastic approximation (SA) and incremental sample averaging. A powerful class of algorithms for solving optimization problems is the class of majorization-minimization (MM) algorithms that subsumes many other common classes, such as the expectation-maximization (EM) algorithms, proximal point algorithms, convex–concave procedures, and block-coordinate descent, among many others [74]. For a certain class of models (*e.g.*, exponential family distributions and their mixtures), online EM solutions have been widely studied and applied, including by team members, *e.g.* [62, 51, 17, 10]. Of particular interest is the ability to adapt the more general MM algorithm framework via SA constructions [77, 83], which can generate a wide range of algorithms for incremental and real-time data inference beyond exponential family models. In the first year, building on positive preliminary work [11], we will seek to design stochastic MM algorithms, in particular for mixture of experts (MoE) models. MoEs are already used in our recent work on ABC [9] via the so-called GLLiM model [24]. One aim is to improve the GLLiM implementation, currently based on a batch EM algorithm, by considering SA and online MM approaches to improve computation time and memory usage, and to extend its applicability to streamed data. As a result, we also expect to contribute to the theoretical understanding of stochastic MM algorithms in more general use cases.

2.1. A generic ABC consistency-like result
Participants: Seniors CD, DF, FF, GF, HN, JA, GK. Post-docs CH, DA, TN, KP.

Approximate Bayesian inference can typically be described as the provision of approximations of intractable posterior distributions. Upon definition of an approximation approach, the subsequent pseudo or quasi-posterior distributions require study in terms of their theoretical properties. A natural locus of investigation is the establishment of consistency of distribution sequences, and concentration of the mass of such sequences upon the true parameters of the underlying data generating processes. More specifically, by focusing our attention on

discrepancy-based ABC methods, as a follow up of our results in [16], we plan to investigate the recent unifying framework proposed in [66], which analyses a number of discrepancy-based ABC schemes via the concept of integral probability metrics and Rademacher's complexity.

3.1. Amortized Approximate Bayesian computation and Bayesian Synthetic Likelihood schemes

Participants: Seniors CD, DF, FF, HN, LS, PR, GK, JBD. Post-docs CH, DA, TN. Students LA, RK, MO.

As an outcome of LANDER, we developed an ABC approach [9] that exhibits amortization properties via the use of the GLLiM inverse regression approach [24]. An interesting follow-up to our work is to investigate the use of GLLiM in the context of synthetic likelihood (SL) approaches. When used in a Bayesian framework, SL techniques can be viewed as alternatives to ABC, in which the intractable likelihoods are replaced by estimators of the likelihoods [37]. Since the seminal work of [86], several estimators have been proposed [55, 33, 48, 35], often derived from auxiliary models [50]. In our ABC work [9], GLLiM is used to provide approximate posteriors, but these posteriors are themselves obtained from approximate likelihoods that can lead to new SL procedures that can be fruitfully investigated.

Tasks for the following years

1.2. Stochastic multi-task learning

Participants: Seniors AN, FF, GF, HN, KQ. Post-docs CH, DA, PW. Students GO.

We plan to consider multi-tasking in an online MM, and more generally, an SA context. The multi-task setting can be used in various modes, following differing justification. Mimicking the setting described in [59], we will consider multi-task updates, in which each task obtains information from the other tasks via some transmission mechanism. We will first consider the case where all tasks target the same objective but may differ in their initialization, stepsize, their use of data, *etc.* The efficiency of their combination may depend on their number and the relative mechanisms of transfer of information between them. We propose to investigate different strategies that may result in more robust online algorithms that are less prone to arbitrary hyperparameter choices.

2.2. Online model selection for streaming data

Participants: Seniors AN, CH, CD, DW, FF, GF, GM, HN, SL, JBD. Post-docs TN. Students GO.

This is a task that presents numerous challenges but is worthwhile due to its wide impact, as model selection in practice is often equivalent to data-driven hyperparameter tuning, which is ubiquitous in modern learning machines and models. How we decide on the best model for a dataset without seeing all data points at once will be a central question of this research task. The challenge is that traditional information criteria such as BIC and AIC require estimation of the parameters that changes iteratively as more data are revealed to the practitioner. It is not yet clear to us which approach is the most promising, *e.g.* revisiting the Bayes factors [69] definition to investigate a recursive BIC computation, focus instead on non-asymptotic criteria like the slope heuristic [12], or attempt a Bayesian non-parametric [23] modelling not to commit to a specific model choice. Work on dynamic estimation such as [80, 73] also present good

starting points. Another promising avenue is the recently explored order confidence framework of [32].

3.2 Invertible neural networks for approximate Bayesian inference

Participants: Seniors CD, DF, FF, HN, JA, KQ, PR, JBD. Post-docs CH, DA, KP. PW. Students RK, MO.

Lastly, aside from the family of mixture of experts (MoE) models that we have used in our ABC work [9], mixture density networks [60] or normalizing flows [64, 70, 71, 29] are neural networks that have already been used in SBI to directly approximate likelihoods or posteriors. These approaches include sequential neural posterior estimation (SNPE), SNPE-A [81], SNPE-B [76], SNPE-C or AFT [65]. However, these methods do not scale well with the dimensionality. Overall, it is not clear whether the gain/compromise in flexibility/tractability would be positively or negatively comparable to Gaussian mixtures or MoEs, and a fair comparison of all these methods features would be of practical interest.

3 Data Management Plan

The project is mainly on developing new models and methods, that would be first validated on simulated or synthetic benchmark data, *e.g.* [75]. If successful, the next step would be to apply the models on real data and possibly to contribute to some more applied work. More specifically, in the context of G. Oudoumanessah's PhD, we are likely to use medical images from the M. J. Fox foundation, PPMI, and a local AGIR-Park study [52, 56]. This data will be used only locally by the French partners, and no transfer or exchange of data is planned in the project. The AGIR-Park data are maintained by the Grenoble Institute of Neuroscience and will not be hosted by the Inria team.

More generally, other sources of data are possible but have not been precisely identified yet. They will most probably not be hosted by the partners. As a consequence, no legal risks related to data ownership and transfer have been identified.

More details on the medical image databases:

PPMI. Data that may be used to illustrate our models can be obtained from the Parkinson's Progression Markers Initiative (PPMI) database . For up-to-date information on the study, visit http://www.ppmi-info.org. The Progression Parkinson's Marker Initiative [79] is a landmark study collaborating with partners around the world to create a robust open-access data set and biosample library to speed scientific breakthroughs and new treatments.

PPMI - a public-private partnership - is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Avid Ra- diopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, Celgene, Denali, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi Genzyme, Servier, Takeda, Teva, UCB, Verily, Voyager Therapeutics and Golub Capital.

AGIR-Park study. This dataset is hosted at Grenoble institute of neuroscience. It is a lot smaller compared to PPMI with only 8 controls subjects and 20 patients. All images have been acquired on a Phillips 3T scanner at the IRMaGe platform in Grenoble. This study is composed of 11 females and 9 males de novo PD patients that are between 53 and 92 years old, and 4 healthy females and males who are between 47 and 82 years old. All participants provided written informed consent to the study. The study was approved by the local ethical

committee (Comit. de Protection des Personnes Sud-Est V, ID RCB-2012-A00310-43 and ID RCB-2014- A01835-42) and registered with ClinicalTrial.gov (NCT02488395).

4 Budget

4.1 Budget (for the first year)

The estimation of the budget for 2023 is done using the following data:

- Cost for welcoming a researcher at Inria: 1000 euros / 10 days.
- Cost for welcoming a researcher in Brisbane: 2000 euros / 20 days.
- Return ticket France - Australia: 2000 euros

Total estimated costs: **17500 euros**

- 20-day visit of TT Nguyen, Inria post-doc fellow, to UQ and QUT, April 2023, Tasks 1.1 and 2.1.
  Cost = 2000 + 2000 = 4000 euros (**Inria**).
- 20-day visit of G. Oudoumanessah, Inria PhD student, to UQ and QUT, June 2023, Task 1.1.
  Cost = 2000 + 2000 = 4000 euros (**Inria**).
- 10-day visit of C. Drovandi or L. South, QUT researchers to Grenoble, Sept. 2023, Task 3.1.
  Cost = 2000 + 1000 = 3000 euros (**Inria**).
- 10-day visit of G. McLachlan or S. Lee, UQ researchers to Grenoble, July. 2023, Task 1.1.
  Cost = 2000 + 1000 = 3000 euros on **UQ funds**.
- 20-day visit of H. Nguyen, UQ researcher to Grenoble, July. 2023, Task 1.1 and 2.1.
  Cost = 2000 + 1500 = 3500 euros on **UQ funds**.

Funding requested from Inria: 17500 - 6500 = **11000 euros**

Total budget request for year 1 (Inria funding):. 11000 euros
Total amount of **secured** co-funding : 6500 euros.
Many of the Australian researchers are funded via various grant programs and thus a significant proportion of the collaborative travel plans will be borne by the individual researchers that are taking part in the activities.

4.2 Tentative Budget for second and third year

We expect a similar planning of 3 Inria-funded visits a year for the following years:
Year 2 estimated budget for Inria: 11000 euros
Year 3 estimated budget for Inria: 11000 euros

4.3 Strategy to get additional funding

No additional funding has been secured yet. We are still looking for opportunities. Some are listed below. We hope the existence of the associate team will help in the future to more easily obtain this kind of funding by formalizing and increasing the visibility of the collaboration.

- 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 mainly one program we could apply for, FASIC program 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 but the number of grants is quite limited too.
- 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 intend to organise a workshop on ABC, SBI and other computational Bayesian methods which will be organised and funded through the MATRIX research program. Our expression of interest is currently under preparation and will be submitted concurrently with this project.
- Together with Xin Guo at the University of Queensland, F. Forbes, G. Fort, and H. Nguyen have leveraged the success of the LANDER team to apply for an ARC Discovery Project, asking for 505,079 AUD, funded over 3 years, which will provide funding for a post-doc fellow and PhD student to work on SA and MM algorithms research. This grant, if successful when announced late in 2022, will provide significant funding towards the goals, plans, and activities of the WOMBAT team.
- Although it is not possible to mention a formal commitment at this stage, we hope to get funding from the Centre for Data Science (CDS) at QUT. It is likely that CDS could provide additional funding towards travel (particularly for students and early career researchers), and possibly short terms research assistants, and also a critical mass of researchers that might end up contributing to the research.
- Similarly, QUT has a program where co-funding of a joint PhD student can be guaranteed if an international institution is paying the other half. In Grenoble, such a possibility exists via some Idex call.

5 Added value

Both sides have expertise in statistical methods and complementary expertise in simulation-based inference and stochastic optimisation. The team consists of a good balance between theoretical and more applied competencies, on both French and Australian sides. The Australian partners have considerable backgrounds in EM and MM algorithms, specifically for solving clustering problems in non-Gaussian data settings, as well as backgrounds in statistical model selection, and Monte Carlo methods. The team also has acknowledged proficiency in statistical software development, and open source software production for the research community at large. The French partners provide complementary expertise in non-parametric statistics, Bayesian approaches, Graphical models, Markov chains, and stochastic approximation. On the application side, both the French and Australian partners have been engaged in neuroimaging

and MRI research. G. McLachlan and H. Nguyen have access to the resources and know-how 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. The Statify team has an active collaboration with the Grenoble Institute of Neuroscience for many years.

The added value of our association will be the ability to collectively combine our resources and data in order to develop tools that are more ubiquitous and universal than those that we could individually produce. The association will result in significant progress in the targeted applications and methodological developments.

Further, we hope to be able to exchange students on a regular basis. In particular, we wish to allow new Masters and PhD students to initiate topics covered by the team, as well as provide current PhDs and post-doc fellows the opportunity to extend their research experience and networks with short and longer term visits. Another objective is to organize a regular workshop on the associate team topics or special sessions at the ERCIM-CMSTATS or JSM conferences with the team partners. Here, we are currently actively engaged with the MATRIX Institute to organise such sessions.

6 Previous Associate Teams

F. Forbes and H.D. Nguyen were the PIs of the LANDER associate team (2019-2022).

7 Impact

Industrial prospects is not the primary goal of this collaboration but we will welcome any opportunities. We expect they may come from two main sources. Part of the proposed research may be useful to the Inria ROAD-AI Defi with Cerema, in which the Statify team is involved (J. Iollo's PhD). It should also be very relevant to the discussions between Inria and Nokia on a common federated learning project. Statify takes part on the discussions and the planned work on SA and MM in WOMBAT may directly benefit to our participation [25].

More generally, the research will generate mathematical and computational tools for conducting data analysis and inference that are directly applicable to a broad range of applied fields. We will work closely with experts from medical science, neuroscience (G. Oudoumanessah's PhD), transportation and infrastructure science, among others, to guarantee the usefulness of our methodological developments.

These algorithms can also be usefully applied to conducting real-time inference in critical domains, such as weather forecasting, water resources management, and logistics planning. Furthermore, the methodology permits the construction of feasible algorithms for the analysis of very large data sets that often arise in biological and spatial image analysis, bioinformatics, and social sciences data.

8 Intellectual Property Right Management

We plan to disseminate our results through the usual media including scientific journals, conferences, workshops, and seminars. Publications will be co-signed under the usual rules of our respective institutions. There will be public and pre-print versions of our manuscripts available on open-access repositories such as hal and arXiv. Our software development will follow the open-source access model and will be made available through dedicated servers, such as git-

lab.inria.fr or a similar service. We will also seek to publish our software on public repositories such as CRAN (the Comprehensive R Archive Network).

8.1 Background

Besides information already publicly available from the website, publications and software of each group, there is no specific background to be listed here. However, this is subject to be updated during the project by mutual agreement, if necessary.

Publications will follow traditional usage and cite previous works accordingly. During the project, each group expect to develop and use open source software.

8.2 Protective measures

The goal of the project is to exchange methodological tools and expertise with no confidential aspects a priori. The partners have already several years of research collaboration and know each other well. We do not at this stage identify real risks but the measures we can consider include to plan "convention de stage" or "convention d'accueil" for visitors.

In addition, if need be, depending on the project development, it will always be possible to sign a more formal collaborative research program agreement between the different parties, that could define the rights and obligations of the coordinators and the team members, in terms of the protection and use of results and associated intellectual property rights.

9 Ethical Issues

After consultation of the Coerle contact points in Grenoble, neither ethical nor legal risks have been identified for this project.

10 Others

For this associate team, specific efforts will be made to exchange students through visits of PhD students and post-doc fellows from the team but also through internships for new students. Previous attempts have been made in the LANDER team in 2019 with a common Master project proposed, but which was not conclusive probably for lack of proper advertisement. In the following years, the pandemics prevent from any similar plan. However, during the last visit of F. Forbes at QUT in April 2022, new interested Australian students have been identified and we plan to propose new internship possibilities for them.

11 References

11.1 Joint publications of the partners

Previous to the LANDER associate team

[1] 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.

[2] 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.

[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).

From the LANDER associate team

[6] Julyan Arbel, Stéphane Girard, Hien Nguyen, and Antoine Usseglio-Carleve. Multivariate expectile-based distribution: properties, Bayesian inference and applications, *Submitted*, 2022.

[7] Wraith, D., Maleki, M. and Forbes, F. Finite Mixtures of Multiple Scaled Generalized Hyperbolic Distributions using a Bayesian approach, *Communications in Statistics - Simulation and Computation,* under revision, 2022.

[8] J.-B. Durand, F. Forbes, C.D. Phan, L. Truong, H.D. Nguyen, F. Dama. Bayesian non parametric spatial prior for car crash risk mapping: a case study in Melbourne, Australia, *Australian and New Zealand Journal of Statistics*, 2022.

[9] Forbes, F., Nguyen, H. D., Nguyen, T., J. Arbel. Summary statistics and discrepancy measures for ABC via surrogate posteriors, *To appear in Statistics & Computing*, 2022.

[10] Nguyen, H. and Forbes, F. Global implicit function theorems and the online Expectation-Maximisation algorithm, *Australian and New Zealand Journal of Statistics*, 2022.

[11] Nguyen, H. D., Forbes, F., G. Fort. An online Minorization-Maximization algorithm, In *International Federation of Classification Societies (IFCS) conference, Porto, Portugal, July 19-23, 2022.*

[12] Nguyen, T., Nguyen, H. D., Chamroukhi, F., and Forbes, F. A non-asymptotic penalization criterion for model selection in mixture of experts models, *To appear in Electronic Journal of Statistics*, 2022.

[13] Nguyen, H. D., Nguyen, T., Chamroukhi, F., and McLachlan, G. J. Approximations of conditional probability density functions in Lebesgue spaces via mixture of experts models, *Journal of Statistical Distributions and Applications*, 8(1):13,2021

[14] Nguyen, T., Chamroukhi, F., Nguyen, H. D., and McLachlan, G. J. Approximation of probability density functions via location-scale finite mixtures in Lebesgue spaces. *Communications in Statistics - Theory & Methods*, 2021.

[15] M. Vladimirova, S. Girard, H. D. Nguyen, and J. Arbel. Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions, *Stat*, 2020.

[16] Nguyen, H. D., Arbel, J., Lu, H., and Forbes, F. Approximate Bayesian Computation Via the Energy Statistic. *IEEE Access*, 8:131683–131698, 2020.

[17] Nguyen, H. D., Forbes, F., and McLachlan, G. Mini-batch learning of exponential family finite mixture models. *Statistics & Computing*, 30:731–748, 2020.

[18] Nguyen, T., Nguyen, H. D., Chamroukhi, F., and McLachlan, G. J. Approximation by finite mixtures of continuous density functions that vanish at infinity, *Cogent Mathematics & Statistics*, 7(1):1750861, 2020.

[19] Redivo, E., Nguyen, H. D., and Gupta, M. Bayesian clustering of skewed and multimodal data using geometric skew normal distributions, *Computational Statistics & Data Analysis, 2020.*

[20] J. Arbel, O. Marchal, and H. D. Nguyen. On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables, *ESAIM: Probability & Statistics*, 2020.

[21] Nguyen, H. D., Chamroukhi, F., and Forbes, F. Approximation results regarding the multiple-output Gaussian gated mixture of linear experts model, *Neurocomputing*, 2019.

[22] Chamroukhi F, Nguyen H.D. Model-based clustering and classification of functional data, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 9.* 2019.

11.2 Main publications of the participants relevant to the project

All joint publications listed above are also relevant to the project.

Publications from Statify and other French collaborators:

[23] J. Arbel, P. De Blasi, and I. Prünster. Stochastic approximations to the Pitman–Yor process, *Bayesian Analysis*, 14(3):753–771, 2019.

[24] 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.

[25] A. Dieuleveut, G. Fort and H.-T. Wai. Federated Majorize-Minimization for large scale learning. *Preprint*, 2022.

[26] G Fort, E Moulines, and H-T Wai. A stochastic path-integrated differential estimator expectation maximization algorithm. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS), 2020.*

[27] Bystrova, D., Arbel, J., Kon Kam King, G., Deslandes, F. Approximating the clusters' prior distribution in Bayesian nonparametric models. *Third Symposium on Advances in Approximate Bayesian Inference*, 2021.

Publications from UQ:

[28] D.C. Ahfock, W.J. Astle, and S. Richardson, Statistical properties of sketching algorithms, Biometrika, 108(2), 283-297, 2021.

[29] L. Hodgkinson, C. van der Heide, F. Roosta, and M.W. Mahoney, Stochastic continuous normalizing flows: training SDEs as ODEs, *Uncertainty in Artificial Intelligence*, pp. 1130-1140, 2021.

[30] 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.

[31] S.X. Lee and G.J. McLachlan, 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.

[32] H.D. Nguyen, D. Fryer, and G.J. McLachlan, Order selection with confidence for finite mixture models. *Journal of the Korean Statistical Society*, to appear, 2022.

Publications from QUT and Monash University:

[33] An, Z., South, L. F., Nott, D. J., and Drovandi, C. C. (2019). Accelerating Bayesian Synthetic Likelihood With the Graphical Lasso. *Journal of Computational & Graphical Statistics*, 28(2):471–475. Publisher: Taylor & Francis.

[34] Drovandi, C. C. and Pettitt, A. N. (2011). Likelihood-free Bayesian estimation of multivariate quantile distributions. *Computational Statistics & Data Analysis*, 55:2541–2556.

[35] Frazier, D. T. and Drovandi, C. (2021). Robust Approximate Bayesian Inference With Synthetic Likelihood. *Journal of Computational & Graphical Statistics*, pages 1–19.

[36] Maleki, M., Wraith, D., Arellano-Valle, R. B. Robust finite mixture modeling of multivariate unrestricted skew-normal generalized hyperbolic distributions. *Statistics & Computing*, 2018.

[37] Price, L. F., Drovandi, C. C., Lee, A., and Nott, D. J. (2018). Bayesian Synthetic Likelihood. *Journal of Computational & Graphical Statistics*, 27(1):1–11.

Publications from Swinburne University of Technology:

[38] 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.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

Publications from Griffith University:

[43] X. Cheng, S. Wang, H. Wang, and S-K. Ng, Deep survival forests for extremely high censored data, *Applied Intelligence*, 1-15, 2022.

[44] Z. Zhang, Z. Shen, H. Wang, and S-K. Ng, A fast adaptive Lasso for the Cox regression via safe screening rules, *Journal of Statistical Computation and Simulation*, 91(14), 3005-3027, 2021.

[45] S-K. Ng, L. Xiang, and K-W. Yau, *Mixture Modelling for Medical and Health Sciences*, Chapman and Hall/CRC, 2019.

[46] S-K. Ng, A two-way clustering framework to identify disparities in multimorbidity patterns of mental and physical health conditions among Australians, *Statistics in Medicine*, 34(26), 3444-3460, 2015.

[47] S.K. Ng, L. Holden, and J. Sun, Identifying comorbidity patterns of health conditions via cluster analysis of pairwise concordance statistics, *Statistics in Medicine*, 31(27), 3393-3405, 2012.

11.3 Other references

From the partners:

[48] An, Z., Nott, D. J., and Drovandi, C. Robust Bayesian synthetic likelihood via a semi-parametric approach, *Statistics & Computing*, 30(3):543–557.

[49] C. De Gruyter, L. Davies, and L.T. Truong, Examining spatial variations in minimum residential parking requirements in Melbourne, *Journal of Transport Geography*, 94, 103096, 2021.

[50] Drovandi, C., Pettitt, T., and Lee, A. (2015). Bayesian indirect inference using a parametric auxiliary model. *Statistical Science*, 30(1):72–95.

[51] S Le Corff and G Fort. Online expectation maximization based algorithms for interence in hidden Markov models. *Electronic Journal of Statistics*, 7:763–792, 2013.

[52] V. Munoz-Ramirez, V. Kmetzsch, F. Forbes, S. Meoni, E. Moro, M. Dojat. Subtle anomaly detection in MRI brain scans: Application to biomarkers extraction in patients with de novo Parkinson's disease. *Artificial Intelligence in Medicine*, 2022.

[53] S.K. Ng, G.J. McLachlan (2004), Speeding up the EM algorithm for mixture model-based segmentation of magnetic resonance images, *Pattern recognition*, 37(8), 1573-1589, 2004.

[54] S.K. Ng, G.J. McLachlan, and A.H. Lee, An incremental EM-based learning approach for on-line prediction of hospital resource utilization, *Artificial Intelligence in Medicine*, 36(3), 257-267, 2006.

[55] Ong, V., Nott, D., Tran, M.-N., Sisson, S., and Drovandi, C. (2018). Likelihood-free inference in high dimensions with synthetic likelihood. *Computational Statistics and Data Analysis*, 128.

[56] G. Oudoumanessah. Unsupervised scalable anomaly detection: application to medical imaging. *Master thesis*, 2022.

[57] L.T. Truong and G. Currie, Macroscopic road safety impacts of public transport: A case study of Melbourne, Australia, *Accident Analysis and Prevention*, 132, 105270, 2019.

External to the partners:

[58] S Allassonniere and J Chevalier. A new class of stochastic EM algorithms. Escaping local maxima and handling inractable sampling, *Computational Statistics and Data Analysis*, 159: 107159, 2021.

[59] Lu Bai, Wu Lin, Abhishek Gupta, and Yew Ong. From multi-task gradient descent to gradient-free evolutionary multitasking: A proof of faster convergence, *IEEE trans. Cybernetics*, PP, 02 2021.

[60] Bishop, C. M. (1994). Mixture density networks. Technical report, Aston University, Birmingham.

[61] Borkar, V.S. Stochastic Approximation: A Dynamical Systems Viewpoint. *Cambridge University Press*, 2008

[62] O Cappé and E Moulines. On-line expectation-maximization algorithm for latent data models, *Journal of the Royal Statistical Society B*, 71:593–613, 2009.

[63] Cranmer, K. and Brehmer, J. and Louppe, G., The frontier of simulation-based inference, *PNAS*, 2020.

[64] Dinh, L., Krueger, D., and Bengio, Y. NICE: non-linear independent components estimation, In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*.

[65] Greenberg, D., Nonnenmacher, M., and Macke, J. Automatic posterior transformation for likelihood-free inference, In *International Conference on Machine Learning*, pages 2404–2414. PMLR.

[66] Legramanti, S. and Durante, D. and Alquier, P., Concentration and robustness of discrepancy-based ABC via Rademacher complexity, *arXiv*, 2022.

[67] B Karimi, B Miasojedow, E Moulines, and H-T Wai. Non-asymptotic analysis of biased stochastic approximation scheme. *Proceedings of Machine Learning Research, 99:0 1–31, 2019.*

[68] B Karimi, H-T Wai, R Moulines, and M Lavielle. On the global convergence of (fast) incremental expectation maximization methods. In *Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019.*

[69] Kass, R.E. and Raftery, A.E. Bayes factors, *Journal of the American Statistical Association*, 1995

[70] Kobyzev, I., Prince, S., and Brubaker, M. (2020). Normalizing Flows: An Introduction and Review of Current Methods, *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–1, 2020

[71] Kruse, J., Ardizzone, L., Rother, C., and Kothe, U. (2021). Benchmarking invertible architectures on inverse problems, *Workshop on Invertible Neural Networks and Normalizing Flows (ICML 2019), arXiv preprint arXiv:2101.10763.* 2021.

[72] E Kuhn, C Matias, and T Rebafka. Properties of the stochastic approximation EM algorithm with mini-batch sampling, *Statistics & Computing*, 30:1725–1739, 2020.

[73] M. Lambert and S. Bonnabel and F. R. Bach The recursive variational Gaussian approximation (R-VGA), *Statistics & Computing*, 2022.

[74] K. Lange. MM Optimization Algorithms *SIAM, Philadelphia, 2016.*

[75] Lueckmann, J.-M., Boelts, J., Greenberg, D. S., Gonçalves, P. J., and Macke, J. H. Benchmarking simulation-based inference, In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pages 343–351. PMLR, 2021

[76] Lueckmann, J.-M., Goncalves, P. J., Bassetto, G., Öcal, K., Nonnenmacher, M., and Macke, J. H. Flexible statistical inference for mechanistic models of neural dynamics, In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc. 2017

[77] J Mairal. Stochastic majorization-minimization algorithms for large-scale optimization, In *Advances in Neural Information Processing Systems*, 2013.

[78] F Maire, E Moulines, and S Lefebvre. Online EM for functional data, *Computational Statistics & Data Analysis*, 111:27–47, 2017.

[79] K. Marek, D. Jennings, S. Lasch, A. Siderowf, C. Tanner, T. Simuni, C. Coffey, K. Kieburtz, E. Flagg, S. Chowdhury, et al. The parkinson progression marker initiative (ppmi), *Progress in neurobiology, 95(4):629–635*, 2011

[80] Mark, C. and Metzner, C. and Lautscham, L. and Strissel, P. L. and Strick, R. and Fabry, B. Bayesian model selection for complex dynamic systems, *Nature communications*, 2018.

[81] Papamakarios, G. and Murray, I. Fast ε-Free Inference of Simulation Models with Bayesian Conditional Density Estimation. In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc.

[82] Papamakarios, G. and Nalisnick, E. and Rezende, D. J. and Mohamed, S. and Lakshminarayanan, B. Normalizing Flows for Probabilistic Modeling and Inference, *Journal of Machine Learning Research*, 2021

[83] M. Razaviyayn, M. Sanjabi, and Z.-Q. Luo. A stochastic successive minimization method for nonsmooth nonconvex optimization with applications to transceiver design in wireless communication networks, *Mathematical Programming*, 157:515–545, 2016.

[84] S.M. Ross, *Simulation*, Academic Press, 2022.

[85] Sisson, S. A., Fan, Y., and Beaumont, M. A., editors. *Handbook of Approximate Bayesian Computation*. CRC Press, Boca Raton, 2019.

[86] Wood, S. Statistical inference for noisy nonlinear ecological dynamic systems, *Nature*, 466(7310):1102–1104, 2010

12 Letter of Intent

THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

School of Mathematics and Physics

Head of School
Professor Joseph Grotowski

7 October 2022

Associate Team Program Officer,
Inria,

Dear Sir/Madam,

This letter confirms our intention to participate in the Inria Associate Team entitled WOMBAT: Variance-reduced Optimization Methods and Bayesian Approximation Techniques for scalable inference.

It is our understanding that the University of Queensland will be requested to sign an Associated Team agreement with Inria, for the Associate Team entitled WOMBAT to be retained.

Please see details of the corresponding partner institution provided below:

Associate Team acronym: WOMBAT

Principal investigator (Inria): Florence Forbes

Principal investigator (partner team):
Hien Nguyen, School of Mathematics and Physics, University of Queensland

Please contact me if you need any further information.

Yours sincerely

Professor Joseph Grotowski
Head of School
School of Mathematics and Physics
University of Queensland

School of Mathematics and Physics
The University of Queensland
Brisbane QLD 4072 Australia

T +61 7 3365 3405
F +61 7 3365 3328

E hos@smp.uq.edu.au
W smp.uq.edu.au

ABN: 63 942 912 684
CRICOS Provider 00025B