

Sophie Mauran

Ph.D. in Applied Mathematics and Computer Science

APO Team
Toulouse INP

✉ sophie.mauran@laposte.net



Academic Degrees

February 2025 **PhD in Applied Mathematics and Computer Science**, *Toulouse INP (Toulouse National Polytechnic Institute)*, Toulouse, France.

- **Title:** Introduction of kernel methods for ensemble data assimilation : algorithmic and applications to geophysical fluids
- **Keywords:** Data assimilation – Kernel methods – Machine Learning – Ensemble Kalman filters
- **Advisors:** Ehouarn Simon (IRIT), Laurent Bertino (NERSC) and Sandrine Mouysset (IRIT)

September 2021 **Engineer Diploma – French Grande École**, *Toulouse INP–ENSEEIH*, Toulouse, France.
Major in Applied Mathematics and Computer Science

June 2015 **Baccalauréat**, *Lycée Pierre Paul Riquet*, Saint-Orens de Gameville, France.
Major in Maths, Physic, Chemistry

Baccalauréat is a French national academic qualification that students can obtain at the completion of their secondary education. Wikipedia page: [https://w.wiki/9w\\$M](https://w.wiki/9w$M)

Research experience

September 2024 – **Research and Teaching Assistant (ATER)**, *Toulouse INP*, Toulouse, France.

- Continuation of doctoral work

February 2025 ◦ **96 hours** of teaching in the department of Computer Science and Telecommunication, bachelor and master students (Lebesgue integral, Fonctionnal Programming, Concurrent Systems, Numerical Optimisation, Data Classification & Analysis).

October 2021 – August 2024 **Doctoral grant (Research Grant by the French Minister of Higher-Education and Research)**, *Toulouse INP*, Toulouse, France.

- **Summary:** In data assimilation, the Kalman filter is the best linear unbiased estimator of the filtering distribution. However, its estimation becomes sub-optimal as soon as the system presents non-linearities and distributions deviate from the Gaussian distribution. This thesis proposes to use the properties of kernel methods, widely used in machine learning, to transform variable distributions and improve filter estimation. A non-linear data assimilation algorithm (KETKF) has been developed based on a generalization of the linear Ensemble Transform Kalman Filter method, with the kernel as the main parameter. An application to the Lorenz 63 model shows the interest of the algorithm for non-linear kernels. A localization of the KETKF algorithm has also been proposed (LKETKF), based on a division of the domain into clusters defined according to the physical state of the system. This slicing algorithm was also developed during this thesis, and when applied to the LETKF for the QG model, it shows an advantage in terms of ensemble error and standard deviation in areas of high system variability. As for the LKETKF applied to the QG model, the results are encouraging and suggest further development.
- **Co-supervision:**
 - Ehouarn Simon and Sandrine Mouysset, IRIT (Toulouse, France), APO (Parallel Algorithms and Optimisation) team (<https://www.irit.fr/departement/calcul-intensif-simulation-optimisation/equipe-apo/>)
 - Laurent Bertino, NERSC (Bergen, Norway), Data Assimilation group (<https://nersc.no/en/gruppe/data-assimilation-group/>).
2 stays: 1 month in June 2022 and 3 months between April and June 2023

- October 2021 **Teaching Assistant**, *Toulouse INP*, Toulouse, France.
- August 2024 **64 hours** of teaching per year in the department of Computer Science and Telecommunication, bachelor and master students (Hilbertian Analysis and Applications, Optimisation, Lebesgue integral, Fonctionnal Programming, Concurrent Systems, Numerical Optimisation, Data Classification & Analysis, Language translation and compiling).
- March 2021 – **Research internship**, *Toulouse INP*, Toulouse, France.
- August 2021
- Preparation of thesis work :
 - Theory on the generalization of ETKF with kernel methods
 - First version of the KETKF and first applications to the Lorenz 63 model
 - **Supervisors** : Ehouarn Simon and Serge Gratton
- September 2020 – **Semester Project**, *EPFL (Ecole Polytechnique Fédérale de Lausanne)*, Lausanne, Switzerland.
- January 2021
- Project carried out one day a week during a semester abroad at EPFL as part of my curriculum at Toulouse INP
 - **Topic**: Theory for model-parallel training of deep neural networks
 - Modeling the Delayed Descent Gradient method
 - Convergence proof in the case of 2 machines with weak assumptions and development of leads to relax the strong assumptions in the general case
 - **Supervisor**: Martin Jaggi, Machine Learning and Optimization (MLO) laboratory (<https://www.epfl.ch/labs/mlo/>)
- June 2020 – **Research internship**, *CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique)*, Toulouse, France.
- August 2020
- **Topic**: Solving linear systems $Ax = b$ with the dimension of A of the order of a million with the iterative Block-Cimmino method.
 - Implementing the triangular augmentation method in C++
 - Method parallelization in MPI
 - **Supervisor**: Philippe Leleux, Parallel Algorithms & sCientific sOftware Operational Performances (Algo-Coop) team (<https://cerfacs.fr/parallels-algorithms/>)
- June 2019 – **Research internship**, *UNIGE (University of Geneva)*, Geneva, Switzerland.
- August 2019
- **Topic**: Participation in the European Transkribus project for the digital preservation of ancient manuscripts
 - State of the art on the various training methods considered
 - Application of the algorithms developed for the project to manuscripts by Ferdinand de Saussure
 - **Supervisors**: Gilles Falquet and Luka Nerima, Language Processing and Analysis laboratory (<https://www.unige.ch/lettres/linguistique/research/lat1>)

Publications

International conferences with selection committee

- 2023 Sophie Mauran, Sandrine Mouysset, Ehouarn Simon, and Laurent Bertino. A kernel extension of the Ensemble Transform Kalman Filter. In *23rd International Conference on Computational Science (ICCS 2023)*, volume 10476 of *Lecture Notes in Computer Science book series (LNCS)*, pages 438–452, Prague, Czech Republic, July 2023. Springer Nature Switzerland.

Short papers, posters and international workshops

- 2024 **ISDA (International Conference on Intelligent Systems Design and Applications)**, *Kobe, Japan*.
- A clustering domain-based localisation strategy for ensemble Kalman filters, Sophie Mauran, Sandrine Mouysset, Ehouarn Simon, Laurent Bertino (poster)
 - Towards the localization of a kernel extension of the ETKF: Sophie Mauran, Sandrine Mouysset, Ehouarn Simon, Laurent Bertino (poster)
- 2023 **EnKF Workshop**, *Norheimsund, Norway*.
- A kernel extension of the Ensemble Transform Kalman Filter: Sophie Mauran, Sandrine Mouysset, Ehouarn Simon, Laurent Bertino (talk, extended abstract)
- 2022 **Sparse Days**, *Saint-Girons, France*.
- Introduction of kernel methods in data assimilation: Sophie Mauran, Sandrine Mouysset, Ehouarn Simon, Laurent Bertino (poster)

Ongoing work

Mauran Sophie, Ehouarn Simon, Sandrine Mouysset and Laurent Bertino, *A clustering domain-based localisation strategy for ensemble Kalman filters*.

Mauran Sophie, Ehouarn Simon, Sandrine Mouysset and Laurent Bertino, *A kernel extension of the Ensemble Transform Kalman Filter : Part II*.

Teaching

Pedagogical approach:

The courses are developed by the teaching team with a view to promoting cross-disciplinarity and autonomy, and preparing students for their professional lives. The concepts presented enable students to easily adapt to all kinds of technologies, without being restricted to a particular application framework. The courses are designed to be progressive and interactive, so as to encourage students' progressive autonomy and overall understanding of the content: the first part generally consists of a presentation of the theoretical aspects and methodology, which are then applied to practical exercises carried out and corrected in class and with the active participation of the students. Each subject usually concludes with a small-group self-directed project.

Main Subjects taught

Hilbertian Analysis and Applications, 2nd year students (engineering curriculum – master), 13 hours/year.

Definitions and formal properties of Hilbert spaces, Riesz representation theorem, introduction of the kernel trick and generalization of linear optimization problems using RKHS (Ridge regression, Support Vector Machine, Principal Components Analysis).

Numerical Optimisation, 2nd year students (engineering curriculum – master), 21 hours/year.

Implementation of several optimisation algorithms with and without constraints (Newton method; Trust Regions method with optimisation of the quadratic model on the trust region with truncated conjugate gradient method and Cauchy step method; augmented Lagrangian method.)

Additional input: Contributing to the development of the half-year project

Concurrent Systems, 2rd year students (engineering curriculum – master), 28 hours/year.

Introduction to parallel computing, handling parallel computing tools (semaphores, monitors, communication channels), design patterns and solving concurrent algorithmic problems.

Additional input: Course coordinator for the HPC-Big Data curriculum (2023-2025) : syllabus design and implementation, exam elaboration.

Data analysis and classification, 3rd year students (engineering curriculum – master), 9 hours/year.

Application of various classification methods (unsupervised and supervised methods) to a data set of heartbeats in order to recognize different beat anomalies.

Scientific skills

Data Assimilation	Ensemble Data Assimilation Methods and Algorithms, Localisation Methods, Non-linear Data Assimilation, Resampling Techniques
Machine Learning	Reproducing Kernel Hilbert Spaces, Kernel Methods, Clustering Methods, Data Preprocessing
High Performance Computing	Parallel Computing, Sparse Linear Algebra, Matrix Approximation
Programming	Functionnal Programming (OCaml), Imperative Programming (C, Java, Python, Julia)

Referees

Dr. Ehouarn Simon

*Associate Professor at
Toulouse INP, IRIT*

✉ ehouarn.simon@toulouse-inp.fr

Dr. Sandrine Mouysset

*Associate Professor at
Université Paul Sabatier, IRIT*

✉ sandrine.mouysset@toulouse-inp.fr

Dr. Laurent Bertino

*Research Leader, Senior Researcher at
NERSC*

✉ Laurent.Bertino@nersc.no

Dr. Philippe Leleux

*Associate Professor
INSA Toulouse*

✉ philippe.leleux@insa-toulouse.fr