

Mathieu Carrière

Topological Data Analysis
Statistical Machine Learning

☎ +33(0) 4 92 38 77 57
✉ mathieu.carriere@inria.fr
Address: Antibes, France
Skype: mathieu.carriere
French/American

Short Bio

I did my PhD at Inria Saclay in the DataShape team, under the supervision of Steve Oudot, and a postdoc of two years in the Rabadán Lab, at the Department of Systems Biology of Columbia University, under the supervision of Raúl Rabadán. My research focuses on topological data analysis and statistical machine learning, with applications in, e.g., bioinformatics. I contributed to the analysis of topological descriptors and their use in machine learning tools such as kernel methods and deep learning. All details can be found on my website: <https://www-sop.inria.fr/members/Mathieu.Carriere/>.

Education

- 2020-Present **Permanent Research Scientist (CRCN)**, *DataShape team, Centre Inria d'Université Côte d'Azur, Sophia Antipolis, France.*
- 2018-2020 **Postdoc. Research Fellow**, *Rabadán Lab, Columbia University, New York, USA.*
- 2014-2018 **Ph.D. in Computer Science**, *EDSTIC (Saclay), Inria Saclay, Palaiseau, France.*
Title: On metric and statistical properties of topological descriptors for geometric data.
- 2011-2014 **Engineering Degree**, *Ecole Centrale Paris, Châtenay-Malabry, France.*
- 2013-2014 **M.Sc. in Mathematics, Vision and Learning (MVA)**, *ENS Cachan, Cachan, France.*

Selected Publications

I am mostly working at the intersection of topological data analysis and statistical machine learning, as well as their applications. Below, I provide a curated list of some of my most representative publications, in computational geometry and topology, as well as artificial intelligence proceedings and journals. My full list of publications is available at: <https://www-sop.inria.fr/members/Mathieu.Carriere/publis/publis.html>

Structure and stability of the one-dimensional Mapper.

Mathieu Carrière and Steve Oudot.

Foundations of Computational Mathematics (FoCM), 18(6):1333–1396, 2017.

Sliced Wasserstein kernel for persistence diagrams.

Mathieu Carrière, Marco Cuturi, and Steve Oudot.

In 34th International Conference on Machine Learning (ICML 2017), volume 70, pages 664–673, PMLR, 2017.

Statistical analysis and parameter selection for Mapper.

Mathieu Carrière, Bertrand Michel, and Steve Oudot.

Journal of Machine Learning Research (JMLR), 19(12):1–39, 2018.

PersLay: a neural network layer for persistence diagrams and new graph topological signatures.

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda.

In 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020), pages 2786–2796, 2020.

Multiparameter persistence image for topological machine learning.

Mathieu Carrière and Andrew Blumberg.

In Advances in Neural Information Processing Systems 33 (NeurIPS 2020), pages 22432–22444, 2020.

Optimizing persistent homology based functions.

Mathieu Carrière, Frédéric Chazal, Marc Glisse, Yuichi Ike, Hariprasad Kannan, and Yuhei Umeda.

In 38th International Conference on Machine Learning (ICML 2021), volume 139, pages 1294–1303, 2021.

Topological Uncertainty: monitoring trained neural networks through persistence of activation graphs.

Théo Lacombe, Yuichi Ike, Mathieu Carrière, Frédéric Chazal, Marc Glisse, and Yuhei Umeda.

In 30th International Joint Conference on Artificial Intelligence (IJCAI 2021), pages 2666–2672, 2021.

Persistent homology based characterization of the breast cancer immune microenvironment: a

feasibility study.

Andrew Aukerman, Mathieu Carrière, Chao Chen, Kevin Gardner, Raúl Rabadán, and Rami Vanguri.
Journal of Computational Geometry (JoCG), 12(2), 2022.

A gradient sampling algorithm for stratified maps with applications to topological data analysis.

Jacob Leygonie, Mathieu Carrière, Théo Lacombe, and Steve Oudot.
Mathematical Programming, 2023.

Selected Invited Talks and Courses

I have participated to many different seminars and conferences in computational topology and machine learning. Below, I provide a curated list of recent scientific events for which I was invited and funded by the organizers for giving a talk/keynote/introductory course. My full list of talks is available at: <https://www-sop.inria.fr/members/Mathieu.Carriere/talk/talk.html>

Talk: "Probabilistic and Statistical Analysis of the Mapper algorithm in Topological Data Analysis".

Symposium on Probabilistic Methods in Biological Sciences, 6th March 2020,
Columbia University, New-York, USA.

Keynote: "An introduction to Topological Data Analysis".

Le Monde des Mathématiques Industrielles, 8th–9th March 2021,
Centre Inria d'Université Côte d'Azur (online), Sophia Antipolis, France.

Introductory Course to the Gudhi library.

2nd Workshop on Topological Methods in Data Analysis, 4th–6th October 2021,
Heidelberg University (online), Heidelberg, Germany.

Talk: "Topology identifies emerging adaptive mutations in SARS-CoV-2".

When AI meets Biology: a workshop, 1st October 2021,
Université de Lyon (online), Lyon, France.

Talk: "An introduction to Topological Machine Learning".

Interpretable and higher-order statistics for late-time cosmology, 27th June–1st July 2022,
Institute for Fundamental Physics of the Universe, Trieste, Italy.

Course: "Basic concepts in topological data analysis".

Introductory School to IHP's thematic quarter on Geometry and Statistics in Data Sciences, 5th–9th Sept. 2022,
IESC Cargèse, Corsica, France.

Course: "Topological Data Analysis".

École d'hiver Statski, 1st–3rd February 2023,
Université de Franche Comté, Metabief, France.

Code Contributions

I am part of the editorial board of the Gudhi library for topological data analysis, to which I am heavily involved in the development and maintenance. I have created several Python submodules (with backend in C++) of the Gudhi library, with Scikit-Learn- and TensorFlow-based APIs for easy integration with these libraries and better accessibility to machine learning practitioners. More precisely, I have developed the following modules:

Representations module: Scikit-Learn-like classes for persistence diagram representations in ML pipelines.

<https://gudhi.inria.fr/python/latest/representations.html>.

Cover complex module: Scikit-Learn-like classes for cover complexes computation in data visualization.

https://gudhi.inria.fr/python/latest/cover_complex_sklearn_user.html.

PersLay module: TensorFlow layer for optimizing representations of persistence diagrams.

https://gudhi.inria.fr/python/latest/representations_tflow_itf_ref.html.

Differentiation module: TensorFlow models for computing and optimizing persistence diagrams.

https://gudhi.inria.fr/python/latest/cubical_complex_tflow_itf_ref.html.

https://gudhi.inria.fr/python/latest/rips_complex_tflow_itf_ref.html

https://gudhi.inria.fr/python/latest/ls_simplex_tree_tflow_itf_ref.html

I have also been involved in the development of tutorials for the Gudhi library, which concretely describe how to

implement machine learning pipelines on real-world data sets with topological tools. See <https://github.com/GUDHI/TDA-tutorial> and the tutorials therein corresponding to the modules described above.

Other coding projects outside of Gudhi can be accessed there: <https://www-sop.inria.fr/members/Mathieu.Carriere/code/code.html>.

Grants and Prizes

I received the following prizes for my work:

- 2017 Best Scientific Contribution 2017 (2nd Prize – 600 euros) from the STIC doctoral school.
- 2018 Thiessé de Rosemont / Schneider Prize (10,000 euros) from Chancellerie des Universités de Paris.

I have been awarded the following grants:

- 2016 Mobility Grant (1000 euros) from the DAAD exchange program.
- 2023 ANR JCJC TopModel (213,993 euros) on multiparameter topological data analysis methods for cosmology and single-cell data.

Teaching Activities

I am supervising the following Ph.D. students:

- 2021–on David Loiseaux, *Multivariate topological data analysis for statistical machine learning*, funded by 3IA Côte d’Azur Institute.
<https://www-sop.inria.fr/members/David.Loiseaux/>
- 2023–on Ziyad Oulhaj, *Statistical contributions to the Mapper algorithm*, funded by ANR GeoDSIC.

I am currently teaching M2 classes on topological data analysis and its interactions with machine learning and applications. More precisely, I am in charge (with colleagues) of the following courses:

- 2021–2023 *Foundations of Geometric Methods in Data Analysis*, MSc. Data Science and Artificial Intelligence (MScDSAI), Université Côte d’Azur, Nice, France.
<http://www-sop.inria.fr/abs/teaching/uca-master-data-science-GTML/uca-master-data-science-GTML--cazals-boissonnat-carriere.html>
- 2020–2023 *Foundations of Geometric Methods in Data Analysis*, 3rd Year of Engineering Program, CentraleSupélec, Gif-sur-Yvette, France.
<http://www-sop.inria.fr/abs/teaching/centrale-FGMDA/centrale-FGMDA--cazals-carriere.html>

As a Ph.D. student, I was a teaching assistant for the following courses:

- 2015–2017 *Topological Data Analysis*, Ecole Polytechnique, Palaiseau, France.
<https://www.enseignement.polytechnique.fr/informatique/INF556/>
- 2016–2017 *Basics of Algorithmic and Programming*, Ecole Polytechnique, Palaiseau, France.
<https://www.enseignement.polytechnique.fr/informatique/INF411/>

Outreach Activities

I have been a reviewer for many different journals and conferences in computational topology and machine learning. For instance, I have regularly participated to the reviewing process of:

- ML/AI ICML, ICLR, NeurIPS, JMLR, TMLR.
- Comp. Topo. SoCG, DCG, JACT, JoCG, Proc. AMS.

I am also a jury member for the Gilles Kahn Ph.D. prize (<https://www.societe-informatique-de-france.fr/recherche/prix-de-these-gilles-kahn/>).

I was part of the organizing committee of the following scientific events:

- May 2022 *Computational Geometry Days 2022*, Online.
<https://project.inria.fr/jga2022/>
- 2019–2020 *New-York Applied Topology Meeting Group*, Columbia University, New-York, USA.
<https://psoc.c2b2.columbia.edu/index.php/new-york-applied-topology-meeting-group/>
- Nov. 2019 *Symposium on Random Matrix Theory*, Columbia University, New-York, USA.
<http://mc4660-projects.s3-website-us-east-1.amazonaws.com/rmtsymp2019/index.html>

Moreover, I wrote a few Medium blog posts about topological data analysis and its usefulness in data science, available at: <https://medium.com/@mathieu.carriere3>

References

Steve Oudot
GeomeriX team
Inria Saclay
91120 Palaiseau, France
steve.oudot@inria.fr
+33(0) 1 74 85 42 16

Marco Cuturi
CREST - ENSAE
Université Paris-Saclay
91120 Palaiseau, France
marco.cuturi@ensae.fr
+33(0) 1 70 26 68 57

Raúl Rabadán
Systems Biology Department
Columbia University
New-York, USA
rr2579@columbia.edu