

RESEARCH CENTRES

**Inria Saclay Center  
at Université Paris-Saclay**

**Inria Center  
at Université Côte d'Azur**

IN PARTNERSHIP WITH:

**Université Paris-Saclay, CNRS**

2022

ACTIVITY REPORT

Project-Team  
DATASHAPE

## Understanding the shape of data

IN COLLABORATION WITH: Laboratoire de mathématiques d'Orsay de  
l'Université de Paris-Sud (LMO)

### DOMAIN

**Algorithmics, Programming, Software  
and Architecture**

### THEME

**Algorithmics, Computer Algebra and  
Cryptology**

*Inria*

# Contents

<b>Project-Team DATASHAPE</b>	<b>1</b>
<b>1 Team members, visitors, external collaborators</b>	<b>2</b>
<b>2 Overall objectives</b>	<b>3</b>
<b>3 Research program</b>	<b>4</b>
3.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis	4
3.2 Statistical aspects of topological and geometric data analysis	4
3.3 Topological and geometric approaches for machine learning	5
3.4 Experimental research and software development	5
<b>4 Application domains</b>	<b>5</b>
<b>5 Social and environmental responsibility</b>	<b>5</b>
5.1 Footprint of research activities	5
<b>6 Highlights of the year</b>	<b>6</b>
6.1 PhD defense	6
6.2 Events	6
<b>7 New software and platforms</b>	<b>6</b>
7.1 New software	6
7.1.1 GUDHI	6
<b>8 New results</b>	<b>7</b>
8.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis	7
8.1.1 Swap, Shift and Trim to Edge Collapse a Filtration	7
8.1.2 Nearly Tight Convergence Bounds for Semi-discrete Entropic Optimal Transport	7
8.1.3 Quantitative Stability of Barycenters in the Wasserstein Space	7
8.1.4 Local Criteria for Triangulating General Manifolds	8
8.2 Statistical aspects of topological and geometric data analysis	8
8.2.1 Efficient Approximation of Multiparameter Persistence Modules	8
8.2.2 Statistical analysis of Mapper for stochastic and multivariate filters	8
8.2.3 Nonparametric estimation of a multivariate density under Kullback-Leibler loss with ISDE	9
8.2.4 Deconvolution of spherical data corrupted with unknown noise	9
8.2.5 Euler and Betti curves are stable under Wasserstein deformations of distributions of stochastic processes	10
8.2.6 On $C^0$ -persistent homology and trees	10
8.3 Topological and geometric approaches for machine learning	10
8.3.1 RipsNet: a general architecture for fast and robust estimation of the persistent homology of point clouds	10
8.3.2 Topological phase estimation method for reparameterized periodic functions.	10
8.4 Algorithmic and Combinatorial Aspects of Low Dimensional Topology	11
8.4.1 Localized Geometric Moves to Compute Hyperbolic Structures on Triangulated 3-Manifolds	11
8.5 Miscellaneous	11
8.5.1 Covering families of triangles	11
8.5.2 Topological Data Analysis and its usefulness for precision medicine studies	12
8.5.3 Persistent homology based characterization of the breast cancer immune microenvironment: a feasibility study	12
8.5.4 Asymptotic-Möbius maps	12

8.5.5	Alpha Wrapping with an Offset	12
<b>9</b>	<b>Bilateral contracts and grants with industry</b>	<b>13</b>
9.1	Bilateral contracts with industry	13
<b>10</b>	<b>Partnerships and cooperations</b>	<b>14</b>
10.1	International research visitors	14
10.1.1	Visits of international scientists	14
10.1.2	Visits to international teams	14
10.2	National initiatives	14
10.2.1	ANR	14
10.2.2	Collaboration with other national research institutes	16
10.3	Regional initiatives	16
<b>11</b>	<b>Dissemination</b>	<b>17</b>
11.1	Promoting scientific activities	17
11.1.1	Scientific events: organisation	17
11.1.2	Scientific events: selection	17
11.1.3	Journal	17
11.1.4	Invited talks	17
11.1.5	Leadership within the scientific community	18
11.1.6	Research administration	18
11.2	Teaching - Supervision - Juries	18
11.2.1	Teaching	18
11.2.2	Supervision	18
11.2.3	Juries	20
11.3	Popularization	20
11.3.1	Articles and contents	20
<b>12</b>	<b>Scientific production</b>	<b>20</b>
12.1	Major publications	20
12.2	Publications of the year	21

## **Project-Team DATASHAPE**

*Creation of the Project-Team: 2017 July 01*

### **Keywords**

#### **Computer sciences and digital sciences**

- A3. – Data and knowledge
- A3.4. – Machine learning and statistics
- A7.1. – Algorithms
- A8. – Mathematics of computing
- A8.1. – Discrete mathematics, combinatorics
- A8.3. – Geometry, Topology
- A9. – Artificial intelligence

#### **Other research topics and application domains**

- B1. – Life sciences
- B2. – Health
- B5. – Industry of the future
- B9. – Society and Knowledge
- B9.5. – Sciences

# 1 Team members, visitors, external collaborators

## Research Scientists

- David Cohen-Steiner [Team leader, INRIA, Researcher]
- Frederic Chazal [Team leader, INRIA, Senior Researcher, HDR]
- Jean-Daniel Boissonnat [INRIA, Emeritus, HDR]
- Mathieu Carrière [INRIA, Researcher]
- Marc Glisse [INRIA, Researcher]
- Jisu Kim [INRIA, Starting Research Position]
- Clément Maria [INRIA, Researcher]

## Faculty Members

- Gilles Blanchard [UNIV PARIS SACLAY, Professor, HDR]
- Blanche Buet [UNIV PARIS SACLAY, Associate Professor]
- Pierre Pansu [UNIV PARIS SACLAY, Professor, HDR]

## Post-Doctoral Fellows

- Charles Arnal [INRIA]
- Felix Hensel [INRIA]
- Kristof Huszar [INRIA, until Sep 2022]

## PhD Students

- Charly Boricaud [UNIV PARIS SACLAY]
- Jeremie Capitaio-Miniconi [UNIV PARIS SACLAY]
- Antoine Commaret [ENS PARIS]
- Bastien Dussap [UNIV PARIS SACLAY]
- Laure Ferraris [INRIA, from Apr 2022]
- Georg, Alexander Gruetzner [STUDIENSTIFTUNG]
- Alexandre, Guerin [Sysnav]
- Olympio Hacquard [UNIV PARIS SACLAY]
- Hugo Henneuse [UNIV PARIS SACLAY, from Oct 2022]
- Etienne Lasalle [UNIV PARIS SACLAY]
- Vadim Lebovici [ENS PARIS]
- David Loiseaux [INRIA]
- Daniel Perez [Paris Sciences et Lettres, until Sep 2022]
- Wojciech Reise [INRIA]
- Owen Rouille [INRIA, until Aug 2022]
- Christophe Vuong [TELECOM PARIS]

### Technical Staff

- Hind Montassif [INRIA, Engineer, until Sep 2022]
- Vincent Rouvreau [INRIA, Engineer]
- Hannah Schreiber [INRIA, Engineer]

### Interns and Apprentices

- Simon Delalande [INRIA, from Sep 2022]
- Ottavio Khalifa [INRIA, from May 2022 until Sep 2022]

### Administrative Assistants

- Aissatou-Sadio Diallo [INRIA, from May 2022]
- Sophie Honnorat [INRIA]

### Visiting Scientists

- John Harvey [UNIV CARDIFF, from Aug 2022]
- Wolfgang Polonik [University of California Davis, from Sep 2022]

### External Collaborators

- Bertrand Michel [CENTRALE NANTES, HDR]
- Martin Royer [SYSTEMX, from Jun 2022]

## 2 Overall objectives

DataShape is a research project in Topological Data Analysis (TDA), a recent field whose aim is to uncover, understand and exploit the topological and geometric structure underlying complex and possibly high dimensional data. The overall objective of the DataShape project is to settle the mathematical, statistical and algorithmic foundations of TDA and to disseminate and promote our results in the data science community.

The approach of DataShape relies on the conviction that it is necessary to combine statistical, topological/geometric and computational approaches in a common framework, in order to face the challenges of TDA. Another conviction of DataShape is that TDA needs to be combined with other data science approaches and tools to lead to successful real applications. It is necessary for TDA challenges to be simultaneously addressed from the fundamental and applied sides.

The team members have actively contributed to the emergence of TDA during the last few years. The variety of expertise, going from fundamental mathematics to software development, and the strong interactions within our team as well as numerous well established international collaborations make our group one of the best to achieve these goals.

The expected output of DataShape is two-fold. First, we intend to set up and develop the mathematical, statistical and algorithmic foundations of Topological and Geometric Data Analysis. Second, we intend to pursue the development of the GUDHI platform, initiated by the team members and which is becoming a standard tool in TDA, in order to provide an efficient state-of-the-art toolbox for the understanding of the topology and geometry of data. The ultimate goal of DataShape is to develop and promote TDA as a new family of well-founded methods to uncover and exploit the geometry of data. This also includes the clarification of the position and complementarity of TDA with respect to other approaches and tools in data science. Our objective is also to provide practically efficient and flexible tools that could be used independently, complementarily or in combination with other classical data analysis and machine learning approaches.

## 3 Research program

### 3.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis

TDA requires to construct and manipulate appropriate representations of complex and high dimensional shapes. A major difficulty comes from the fact that the complexity of data structures and algorithms used to approximate shapes rapidly grows as the dimensionality increases, which makes them intractable in high dimensions. We focus our research on simplicial complexes which offer a convenient representation of general shapes and generalize graphs and triangulations. Our work includes the study of simplicial complexes with good approximation properties and the design of compact data structures to represent them.

In low dimensions, effective shape reconstruction techniques exist that can provide precise geometric approximations very efficiently and under reasonable sampling conditions. Extending those techniques to higher dimensions as is required in the context of TDA is problematic since almost all methods in low dimensions rely on the computation of a subdivision of the ambient space. A direct extension of those methods would immediately lead to algorithms whose complexities depend exponentially on the ambient dimension, which is prohibitive in most applications. A first direction to by-pass the curse of dimensionality is to develop algorithms whose complexities depend on the intrinsic dimension of the data (which most of the time is small although unknown) rather than on the dimension of the ambient space. Another direction is to resort to cruder approximations that only captures the homotopy type or the homology of the sampled shape. The recent theory of persistent homology provides a powerful and robust tool to study the homology of sampled spaces in a stable way.

### 3.2 Statistical aspects of topological and geometric data analysis

The wide variety of larger and larger available data - often corrupted by noise and outliers - requires to consider the statistical properties of their topological and geometric features and to propose new relevant statistical models for their study.

There exist various statistical and machine learning methods intending to uncover the geometric structure of data. Beyond manifold learning and dimensionality reduction approaches that generally do not allow to assert the relevance of the inferred topological and geometric features and are not well-suited for the analysis of complex topological structures, set estimation methods intend to estimate, from random samples, a set around which the data is concentrated. In these methods, that include support and manifold estimation, principal curves/manifolds and their various generalizations to name a few, the estimation problems are usually considered under losses, such as Hausdorff distance or symmetric difference, that are not sensitive to the topology of the estimated sets, preventing these tools to directly infer topological or geometric information.

Regarding purely topological features, the statistical estimation of homology or homotopy type of compact subsets of Euclidean spaces, has only been considered recently, most of the time under the quite restrictive assumption that the data are randomly sampled from smooth manifolds.

In a more general setting, with the emergence of new geometric inference tools based on the study of distance functions and algebraic topology tools such as persistent homology, computational topology has recently seen an important development offering a new set of methods to infer relevant topological and geometric features of data sampled in general metric spaces. The use of these tools remains widely heuristic and until recently there were only a few preliminary results establishing connections between geometric inference, persistent homology and statistics. However, this direction has attracted a lot of attention over the last three years. In particular, stability properties and new representations of persistent homology information have led to very promising results to which the DATASHAPE members have significantly contributed. These preliminary results open many perspectives and research directions that need to be explored.

Our goal is to build on our first statistical results in TDA to develop the mathematical foundations of Statistical Topological and Geometric Data Analysis. Combined with the other objectives, our ultimate goal is to provide a well-founded and effective statistical toolbox for the understanding of topology and geometry of data.

### 3.3 Topological and geometric approaches for machine learning

This objective is driven by the problems raised by the use of topological and geometric approaches in machine learning. The goal is both to use our techniques to better understand the role of topological and geometric structures in machine learning problems and to apply our TDA tools to develop specialized topological approaches to be used in combination with other machine learning methods.

### 3.4 Experimental research and software development

We develop a high quality open source software platform called GUDHI which is becoming a reference in geometric and topological data analysis in high dimensions. The goal is not to provide code tailored to the numerous potential applications but rather to provide the central data structures and algorithms that underlie applications in geometric and topological data analysis.

The development of the GUDHI platform also serves to benchmark and optimize new algorithmic solutions resulting from our theoretical work. Such development necessitates a whole line of research on software architecture and interface design, heuristics and fine-tuning optimization, robustness and arithmetic issues, and visualization. We aim at providing a full programming environment following the same recipes that made up the success story of the CGAL library, the reference library in computational geometry.

Some of the algorithms implemented on the platform will also be interfaced to other software platforms, such as the R software for statistical computing, and languages such as Python in order to make them usable in combination with other data analysis and machine learning tools. A first attempt in this direction has been done with the creation of an R package called TDA in collaboration with the group of Larry Wasserman at Carnegie Mellon University (Inria Associated team CATS) that already includes some functionalities of the GUDHI library and implements some joint results between our team and the CMU team. A similar interface with the Python language is also considered a priority. To go even further towards helping users, we will provide utilities that perform the most common tasks without requiring any programming at all.

## 4 Application domains

Our work is mostly of a fundamental mathematical and algorithmic nature but finds a variety of applications in data analysis, e.g., in material science, biology, sensor networks, 3D shape analysis and processing, to name a few.

More specifically, DATASHAPE is working on the analysis of trajectories obtained from inertial sensors (PhD theses of Wojtek Riese Alexandre Guérin with Sysnav, participation to the DGA/ANR challenge MALIN with Sysnav) and, more generally on the development of new TDA methods for Machine Learning and Artificial Intelligence for (multivariate) time-dependent data from various kinds of sensors in collaboration with Fujitsu, or high dimensional point cloud data with Metafora.

DATASHAPE is also working in collaboration with the University of Columbia in New-York, especially with the Rabadan lab, in order to improve bioinformatics methods and analyses for single cell genomic data. For instance, there is a lot of work whose aim is to use TDA tools such as persistent homology and the Mapper algorithm to characterize, quantify and study statistical significance of biological phenomena that occur in large scale single cell data sets. Such biological phenomena include, among others: the cell cycle, functional differentiation of stem cells, and immune system responses (such as the spatial response on the tissue location, and the genomic response with protein expression) to breast cancer.

## 5 Social and environmental responsibility

### 5.1 Footprint of research activities

The weekly research seminar of DATASHAPE is now taking place in hybrid mode. The travels for the team members have decreased a lot these years, mainly because of the COVID-19 pandemic but also to take care of the environmental footprint of the team.



## 6 Highlights of the year

### 6.1 PhD defense

- Daniel Perez. Homologie persistante des processus stochastiques et leurs fonctions zeta. Juillet 2022.
- Owen Rouillé. Calculs à grande échelle d'invariants de 3-variétés. Septembre 2022.
- Louis Pujol. Modélisation des données de cytométrie et classification non supervisée en dimension modérée sous l'hypothèse de structure d'indépendance. Décembre 2022.
- Étienne Lasalle. Quelques contributions à l'analyse statistique de données à structure de graphe. Decembre 2022.
- Alex Delalande. Quantitative Stability in Quadratic Optimal Transport. December 2022.

### 6.2 Events

- Co-organization of the thematic trimester at Institut Henri Poincaré, [Geometry and Statistics in Data Sciences](#). Sept-Dec. 2022.
- We organized a one week team workshop in May 2022, giving the opportunity to all the PhD students, post-doc and researchers of the team to present their work and discuss scientific questions all together. Some researchers, Elisabeth Gassiat (Univ. Paris-Saclay - deconvolution with unknown noise), Indira Chatterji (Univ. Nice - fundamental group) and Pierre Pansu (Near homology) were also invited to give mini-courses.

## 7 New software and platforms

### 7.1 New software

#### 7.1.1 GUDHI

**Name:** Geometric Understanding in Higher Dimensions

**Keywords:** Computational geometry, Topology, Clustering

**Scientific Description:** The Gudhi library is an open source library for Computational Topology and Topological Data Analysis (TDA). It offers state-of-the-art algorithms to construct various types of simplicial complexes, data structures to represent them, and algorithms to compute geometric approximations of shapes and persistent homology.

The GUDHI library offers the following interoperable modules:

. Complexes: + Cubical + Simplicial: Rips, Witness, Alpha and Čech complexes + Cover: Nerve and Graph induced complexes . Data structures and basic operations: + Simplex tree, Skeleton blockers and Toplex map + Construction, update, filtration and simplification . Topological descriptors computation . Manifold reconstruction . Topological descriptors tools: + Bottleneck and Wasserstein distance + Statistical tools + Persistence diagram and barcode

**Functional Description:** The GUDHI open source library will provide the central data structures and algorithms that underly applications in geometry understanding in higher dimensions. It is intended to both help the development of new algorithmic solutions inside and outside the project, and to facilitate the transfer of results in applied fields.

**News of the Year:** - TensorFlow interface - Improved Čech complex, edge collapses - Packaging for recent Mac

**URL:** <https://gudhi.inria.fr/>

**Publication:** [hal-01108461](#)

**Contact:** Marc Glisse

**Participants:** Clément Maria, François Godi, David Salinas, Jean-Daniel Boissonnat, Marc Glisse, Mariette Yvinec, Pawel Dlotko, Siargey Kachanovich, Vincent Rouvreau, Mathieu Carrière, Clément Jamin, Siddharth Pritam, Frederic Chazal, Steve Oudot, Wojciech Reise, Hind Montassif

**Partners:** Université Côte d'Azur (UCA), Fujitsu

## 8 New results

### 8.1 Algorithmic aspects and new mathematical directions for topological and geometric data analysis

#### 8.1.1 Swap, Shift and Trim to Edge Collapse a Filtration

**Participant:** Marc Glisse.

*In collaboration with Siddharth Pritam (Shiv Nadar University, India)*

Boissonnat and Pritam introduced an algorithm to reduce a filtration of flag (or clique) complexes, which can in particular speed up the computation of its persistent homology. They used so-called edge collapse to reduce the input flag filtration and their reduction method required only the 1-skeleton of the filtration. In this paper [26] we revisit the use of edge collapse for efficient computation of persistent homology. We first give a simple and intuitive explanation of the principles underlying that algorithm. This in turn allows us to propose various extensions including a zigzag filtration simplification algorithm. We finally show some experiments to better understand how it behaves.

#### 8.1.2 Nearly Tight Convergence Bounds for Semi-discrete Entropic Optimal Transport

**Participant:** Alex Delalande.

We derive nearly tight and non-asymptotic convergence bounds for solutions of entropic semi-discrete optimal transport. These bounds quantify the stability of the dual solutions of the regularized problem (sometimes called Sinkhorn potentials) w.r.t. the regularization parameter, for which we ensure a better than Lipschitz dependence. Such facts may be a first step towards a mathematical justification of  $\epsilon$ -scaling heuristics for the numerical resolution of regularized semi-discrete optimal transport. Our results also entail a non-asymptotic and tight expansion of the difference between the entropic and the unregularized costs [25].

#### 8.1.3 Quantitative Stability of Barycenters in the Wasserstein Space

**Participant:** Alex Delalande.

*In collaboration with Guillaume Carlier (CEREMADE) and Quentin Mérigot (Laboratoire de Mathématiques d'Orsay)*

Wasserstein barycenters define averages of probability measures in a geometrically meaningful way. Their use is increasingly popular in applied fields, such as image, geometry or language processing. In these fields however, the probability measures of interest are often not accessible in their entirety and the

practitioner may have to deal with statistical or computational approximations instead. In this article, we quantify the effect of such approximations on the corresponding barycenters. We show that Wasserstein barycenters depend in a Hölder-continuous way on their marginals under relatively mild assumptions. Our proof relies on recent estimates that quantify the strong convexity of the dual quadratic optimal transport problem and a new result that allows to control the modulus of continuity of the push-forward operation under a (not necessarily smooth) optimal transport map [36].

#### 8.1.4 Local Criteria for Triangulating General Manifolds

**Participant:** Jean-Daniel Boissonnat.

*In collaboration with Ramsay Dyer, Arijit Ghosh (Indian Statistical Institute) and Mathijs Wintraecken (IST Austria).*

We present [13] criteria for establishing a triangulation of a manifold. Given a manifold  $M$ , a simplicial complex  $A$ , and a map  $H$  from the underlying space of  $A$  to  $M$ , our criteria are presented in local coordinate charts for  $M$ , and ensure that  $H$  is a homeomorphism. These criteria do not require a differentiable structure, or even an explicit metric on  $M$ . No Delaunay property of  $A$  is assumed. The result provides a triangulation guarantee for algorithms that construct a simplicial complex by working in local coordinate patches. Because the criteria are easily verified in such a setting, they are expected to be of general use.

## 8.2 Statistical aspects of topological and geometric data analysis

### 8.2.1 Efficient Approximation of Multiparameter Persistence Modules

**Participant:** David Loiseaux, Mathieu Carrière.

*In collaboration with Andrew Blumberg (Columbia University)*

Topological Data Analysis is a growing area of data science, which aims at computing and characterizing the geometry and topology of data sets, in order to produce useful descriptors for subsequent statistical and machine learning tasks. Its main computational tool is persistent homology, which amounts to track the topological changes in growing families of subsets of the data set itself, called filtrations, and encode them in an algebraic object, called persistence module. Even though algorithms and theoretical properties of modules are now well-known in the single-parameter case, that is, when there is only one filtration to study, much less is known in the multi-parameter case, where several filtrations are given at once. Though more complicated, the resulting persistence modules are usually richer and encode more information, making them better descriptors for data science. In this article [39], we present the first approximation scheme, which is based on fibered barcodes and exact matchings, two constructions that stem from the theory of single-parameter persistence, for computing and decomposing general multi-parameter persistence modules. Our algorithm has controlled complexity and running time, and works in arbitrary dimension, i.e., with an arbitrary number of filtrations. Moreover, when restricting to specific classes of multi-parameter persistence modules, namely the ones that can be decomposed into intervals, we establish theoretical results about the approximation error between our estimate and the true module in terms of interleaving distance. Finally, we present empirical evidence validating output quality and speed-up on several data sets.

### 8.2.2 Statistical analysis of Mapper for stochastic and multivariate filters

**Participant:** Mathieu Carrière.

*In collaboration with Bertrand Michel (Ecole Centrale de Nantes)*

Reeb spaces, as well as their discretized versions called Mappers, are common descriptors used in Topological Data Analysis, with plenty of applications in various fields of science, such as computational biology and data visualization, among others. The stability and quantification of the rate of convergence of the Mapper to the Reeb space has been studied a lot in recent works [BBMW19, CO17, CMO18, MW16], focusing on the case where a scalar-valued filter is used for the computation of Mapper. On the other hand, much less is known in the multivariate case, when the codomain of the filter is  $\mathbb{R}^p$ , and in the general case, when it is a general metric space  $(Z, d_Z)$ , instead of  $\mathbb{R}$ . The few results that are available in this setting [DMW17, MW16] can only handle continuous topological spaces and cannot be used as is for finite metric spaces representing data, such as point clouds and distance matrices. In this article [16], we introduce a slight modification of the usual Mapper construction and we give risk bounds for estimating the Reeb space using this estimator. Our approach applies in particular to the setting where the filter function used to compute Mapper is also estimated from data, such as the eigenfunctions of PCA. Our results are given with respect to the Gromov-Hausdorff distance, computed with specific filter-based pseudometrics for Mappers and Reeb spaces defined in [DMW17]. We finally provide applications of this setting in statistics and machine learning for different kinds of target filters, as well as numerical experiments that demonstrate the relevance of our approach.

### 8.2.3 Nonparametric estimation of a multivariate density under Kullback-Leibler loss with ISDE

**Participant:** Louis Pujol.

In this paper [44], we propose a theoretical analysis of the algorithm ISDE, introduced in previous work. From a dataset, ISDE learns a density written as a product of marginal density estimators over a partition of the features. We show that under some hypotheses, the Kullback-Leibler loss between the proper density and the output of ISDE is a bias term plus the sum of two terms which goes to zero as the number of samples goes to infinity. The rate of convergence indicates that ISDE tackles the curse of dimensionality by reducing the dimension from the one of the ambient space to the one of the biggest blocks in the partition. The constants reflect a combinatorial complexity reduction linked to the design of ISDE.

### 8.2.4 Deconvolution of spherical data corrupted with unknown noise

**Participant:** Jérémie Capitao Miniconi.

*In collaboration with Elisabeth Gassiat (Laboratoire de Mathématiques d'Orsay)*

We consider the deconvolution problem for densities supported on a  $(d-1)$ -dimensional sphere with unknown center and unknown radius, in the situation where the distribution of the noise is unknown and without any other observations. We propose estimators of the radius, of the center, and of the density of the signal on the sphere that are proved consistent without further information. The estimator of the radius is proved to have almost parametric convergence rate for any dimension  $d$ . When  $d=2$ , the estimator of the density is proved to achieve the same rate of convergence over Sobolev regularity classes of densities as when the noise distribution is known [35].

### 8.2.5 Euler and Betti curves are stable under Wasserstein deformations of distributions of stochastic processes

**Participant:** Daniel Perez.

Euler and Betti curves of stochastic processes defined on a  $d$ -dimensional compact Riemannian manifold which are almost surely in a Sobolev space  $W^{n,s}(X, \mathbb{R})$  (with  $d < n$ ) are stable under perturbations of the distributions of said processes in a Wasserstein metric. Moreover, Wasserstein stability is shown to hold for all  $p > d/n$  for persistence diagrams stemming from functions in  $W^{n,s}(X, \mathbb{R})$ . [41]

### 8.2.6 On $C^0$ -persistent homology and trees

**Participant:** Daniel Perez.

In this paper [42] we give a metric construction of a tree which correctly identifies connected components of superlevel sets of continuous functions  $f : X \rightarrow \mathbb{R}$  and show that it is possible to retrieve the  $H_0$ -persistence diagram from this tree. We revisit the notion of homological dimension previously introduced by Schweinhart and give some bounds for the latter in terms of the upper-box dimension of  $X$ , thereby partially answering a question of the same author. We prove a quantitative version of the Wasserstein stability theorem valid for regular enough  $X$  and  $\alpha$ -Hölder functions and discuss some applications of this theory to random fields and the topology of their superlevel sets.

## 8.3 Topological and geometric approaches for machine learning

### 8.3.1 RipsNet: a general architecture for fast and robust estimation of the persistent homology of point clouds

**Participant:** Felix Hensel, Marc Glisse, Frederic Chazal, Thibault de Surrél, Mathieu Carrière.

*In collaboration with Théo Lacombe (Université Gustave Eiffel), Hiroaki Kurihara (Fujitsu Ltd) and Yuichi Ike (University of Tokyo).*

The use of topological descriptors in modern machine learning applications, such as persistence diagrams (PDs) arising from Topological Data Analysis (TDA), has shown great potential in various domains. However, their practical use in applications is often hindered by two major limitations: the computational complexity required to compute such descriptors exactly, and their sensitivity to even low-level proportions of outliers. In this work [47], we propose to bypass these two burdens in a data-driven setting by entrusting the estimation of (vectorization of) PDs built on top of point clouds to a neural network architecture that we call RipsNet. Once trained on a given data set, RipsNet can estimate topological descriptors on test data very efficiently with generalization capacity. Furthermore, we prove that RipsNet is robust to input perturbations in terms of the 1-Wasserstein distance, a major improvement over the standard computation of PDs that only enjoys Hausdorff stability, yielding RipsNet to substantially outperform exactly-computed PDs in noisy settings. We showcase the use of RipsNet on both synthetic and real-world data. Our implementation will be made freely and publicly available as part of the open-source library Gudhi

### 8.3.2 Topological phase estimation method for reparameterized periodic functions.

**Participant:** Frédéric Chazal, Wojciech Riese.

*In collaboration with Thomas Bonis (Univ. Gustave Eiffel) and Bertrand Michel (Ecole Centrale Nantes).*

We consider a signal composed of several periods of a periodic function, of which we observe a noisy reparametrisation. The phase estimation problem consists of finding that reparametrisation, and, in particular, the number of observed periods. Existing methods are well-suited to the setting where the periodic function is known, or at least, simple. We consider the case when it is unknown and we propose an estimation method based on the shape of the signal. We use the persistent homology of sublevel sets of the signal to capture the temporal structure of its local extrema. We infer the number of periods in the signal by counting points in the persistence diagram and their multiplicities. Using the estimated number of periods, we construct an estimator of the reparametrisation. It is based on counting the number of sufficiently prominent local minima in the signal. This work is motivated by a vehicle positioning problem, on which we evaluated the proposed method [34].

## 8.4 Algorithmic and Combinatorial Aspects of Low Dimensional Topology

### 8.4.1 Localized Geometric Moves to Compute Hyperbolic Structures on Triangulated 3-Manifolds

**Participant:** Clément Maria, Owen Rouillé.

A fundamental way to study 3-manifolds is through the geometric lens, one of the most prominent geometries being the hyperbolic one. We focus on the computation of a complete hyperbolic structure on a connected orientable hyperbolic 3-manifold with torus boundaries. This family of 3-manifolds includes the knot complements. This computation of a hyperbolic structure requires the resolution of gluing equations on a triangulation of the space, but not all triangulations admit a solution to the equations. In this paper [27], we propose a new method to find a triangulation that admits a solution to the gluing equations, using convex optimization and localized combinatorial modifications. It is based on Casson and Rivin's reformulation of the equations. We provide a novel approach to modify a triangulation and update its geometry, along with experimental results to support the new method.

## 8.5 Miscellaneous

### 8.5.1 Covering families of triangles

**Participant:** Marc Glisse.

*In collaboration with Olivier Devillers and Ji-Won Park (Inria team Gamble) and Otfried Cheong (KAIST, Korea).*

A cover for a family  $F$  of sets in the plane is a set into which every set in  $F$  can be isometrically moved. We are interested in the convex cover of smallest area for a given family of triangles. Park and Cheong conjectured that any family of triangles of bounded diameter has a smallest convex cover that is itself a triangle. The conjecture is equivalent to the claim that for every convex set  $X$  there is a triangle  $Z$  whose area is not larger than the area of  $X$ , such that  $Z$  covers the family of triangles contained in  $X$ . We prove this claim [17] for the case where a diameter of  $X$  lies on its boundary. We also give a complete characterization of the smallest convex cover for the family of triangles contained in a half-disk, and for the family of triangles contained in a square. In both cases, this cover is a triangle.

### 8.5.2 Topological Data Analysis and its usefulness for precision medicine studies

**Participant:** Mathieu Carrière, Frédéric Chazal.

*In collaboration with Raquel Iniesta (King's College London), Ewan Carr (King's College London), Naya Yerole mou (University of Oxford, Alan Turing Institute), Bertrand Michel (Ecole Centrale Nantes)*

Precision medicine allows the extraction of information from complex datasets to facilitate clinical decision-making at the individual level. Topological Data Analysis (TDA) offers promising tools that complement current analytical methods in precision medicine studies. We introduce the fundamental concepts of the TDA corpus (the simplicial complex, the Mapper graph, the persistence diagram and persistence landscape). We show how these can be used to enhance the prediction of clinical outcomes and to identify novel subpopulations of interest, particularly applied to understand remission of depression in data from the GENDEP clinical trial [21].

### 8.5.3 Persistent homology based characterization of the breast cancer immune microenvironment: a feasibility study

**Participant:** Mathieu Carrière.

*In collaboration with Andrew Aukerman (Columbia University), Chao Chen (Stony Brook University), Kevin Gardner (Columbia University), Raúl Rabadán (Columbia University), Rami Vanguri (Columbia University)*

Persistent homology is a powerful tool in topological data analysis. The main output, persistence diagrams, encode the geometry and topology of given datasets. We present a novel application of persistent homology to characterize the biological environment surrounding breast cancers, known as the tumor microenvironment. Specifically, we will characterize the spatial arrangement of immune and malignant epithelial (tumor) cells within the breast cancer immune microenvironment. Quantitative and robust characterizations are built by computing persistence diagrams from quantitative multiplex immunofluorescence, which is a technology which allows us to obtain spatial coordinates and protein intensities on individual cells. The resulting persistence diagrams are evaluated as characteristic biomarkers predictive of cancer subtype and prognostic of overall survival. For a cohort of approximately 700 breast cancer patients with median 8.5-year clinical follow-up, we show that these persistence diagrams outperform and complement the usual descriptors which capture spatial relationships with nearest neighbor analysis. Our results [11] thus suggest new methods which can be used to build topology-based biomarkers which are characteristic and predictive of cancer subtype and response to therapy as well as prognostic of overall survival.

### 8.5.4 Asymptotic-Möbius maps

**Participant:** Georg Gruetzner.

Roughly speaking, a map between metric spaces is asymptotically Möbius if it induces quasi-Möbius maps on asymptotic cones. We show [38] that under such maps, some large-scale notions of dimension increase: asymptotic dimension for finitely generated nilpotent groups, telescopic dimension for CAT(0) spaces.

### 8.5.5 Alpha Wrapping with an Offset

**Participant:** David Cohen-Steiner.

*In collaboration with Pierre Alliez and Cédric Portaneri (Inria team Titane), Mael Rouxel-Labbé (Geometry Factory) and Michael Hemmer (Independent researcher).*

Given an input 3D geometry such as a triangle soup or a point set, we address the problem of generating a watertight and orientable surface triangle mesh that strictly encloses the input [23]. The output mesh is obtained by greedily refining and carving a 3D Delaunay triangulation on an offset surface of the input, while carving with empty balls of radius alpha. The proposed algorithm is controlled via two user-defined parameters: alpha and offset. Alpha controls the size of cavities or holes that cannot be traversed during carving, while offset controls the distance between the vertices of the output mesh and the input. Our algorithm is guaranteed to terminate and to yield a valid and strictly enclosing mesh, even for defect-laden inputs. Genericity is achieved using an abstract interface probing the input, enabling any geometry to be used, provided a few basic geometric queries can be answered. We benchmark the algorithm on large public datasets such as Thingi10k, and compare it to state-of-the-art approaches in terms of robustness, approximation, output complexity, speed, and peak memory consumption. Our implementation is available through the CGAL library.

## 9 Bilateral contracts and grants with industry

### 9.1 Bilateral contracts with industry

- **Participants:** Alexandre Guerin, Frédéric Chazal.

Collaboration with **Sysnav**, a French SME with world leading expertise in navigation and geopositioning in extreme environments, on TDA, geometric approaches and machine learning for the analysis of movements of pedestrians and patients equipped with inertial sensors (CIFRE PhD of Alexandre Guérin).

- **Participants:** Felix Hensel, Theo Lacombe, Marc Glisse, Mathieu Carrière, Frédéric Chazal.

Research collaboration with **Fujitsu** on the development of new TDA methods and tools for Machine learning and Artificial Intelligence (started in Dec 2017).

- **Participants:** Louis Pujol, Bastien Dussap, Marc Glisse, Gilles Blanchard.

Research collaboration with **MetaFora** on the development of new TDA-based and statistical methods for the analysis of cytometric data (started in Nov. 2019).

- **Participants:** David Cohen-Steiner.

Collaboration with **Dassault Systèmes** and Inria team **Geomerix** (Saclay) on the applications of methods from geometric measure theory to the modelling and processing of complex 3D shapes (PhD of Lucas Brifault, started in May 2022).



## 10 Partnerships and cooperations

### 10.1 International research visitors

#### 10.1.1 Visits of international scientists

##### Other international visits to the team

###### Wolfgang Polonik

**Status** researcher

**Institution of origin:** UC Davis

**Country:** USA

**Dates:** September to December 2022

**Context of the visit:** research stay

**Mobility program/type of mobility:** research stay and lecture.

###### John Harvey

**Status** researcher

**Institution of origin:** Cardiff University, School of Mathematics

**Country:** U.K.

**Dates:** September to December 2022

**Context of the visit:** research stay

**Mobility program/type of mobility:** research stay

#### 10.1.2 Visits to international teams

##### Sabbatical programme

###### Clément Maria

**Visited institution:** Escola de Matemática Aplicada de la Fondation Getúlio Vargas (Brésil)

**Dates of the stay:** From Sat Oct 01 2022 to Sat Sep 30 2023

**Summary of the stay:** This is a one year sabbatical stay whose main goal is to set up and reinforce new long term collaborations in computational and applied topology and topological data analysis.

### 10.2 National initiatives

#### 10.2.1 ANR

##### ANR ASPAG

**Participants:** Marc Glisse.

- Acronym : ASPAG.
- Type : ANR blanc.
- Title : Analysis and Probabilistic Simulations of Geometric Algorithms.
- Coordinator : Olivier Devillers (équipe Inria Gamble).
- Duration : 4 years from January 2018 to December 2021, extended to June 2022.
- Others Partners: Inria Gamble, LPSM, LABRI, Université de Rouen, IECL, Université du Littoral Côte d'Opale, Telecom ParisTech, Université Paris X (Modal'X), LAMA, Université de Poitiers, Université de Bourgogne.

- Abstract:

The analysis and processing of geometric data has become routine in a variety of human activities ranging from computer-aided design in manufacturing to the tracking of animal trajectories in ecology or geographic information systems in GPS navigation devices. Geometric algorithms and probabilistic geometric models are crucial to the treatment of all this geometric data, yet the current available knowledge is in various ways much too limited: many models are far from matching real data, and the analyses are not always relevant in practical contexts. One of the reasons for this state of affairs is that the breadth of expertise required is spread among different scientific communities (computational geometry, analysis of algorithms and stochastic geometry) that historically had very little interaction. The Aspaga project brings together experts of these communities to address the problem of geometric data. We will more specifically work on the following three interdependent directions.

(1) Dependent point sets: One of the main issues of most models is the core assumption that the data points are independent and follow the same underlying distribution. Although this may be relevant in some contexts, the independence assumption is too strong for many applications.

(2) Simulation of geometric structures: The phenomena studied in (1) involve intricate random geometric structures subject to new models or constraints. A natural first step would be to build up our understanding and identify plausible conjectures through simulation. Perhaps surprisingly, the tools for an effective simulation of such complex geometric systems still need to be developed.

(3) Understanding geometric algorithms: the analysis of algorithm is an essential step in assessing the strengths and weaknesses of algorithmic principles, and is crucial to guide the choices made when designing a complex data processing pipeline. Any analysis must strike a balance between realism and tractability; the current analyses of many geometric algorithms are notoriously unrealistic. Aside from the purely scientific objectives, one of the main goals of Aspaga is to bring the communities closer in the long term. As a consequence, the funding of the project is crucial to ensure that the members of the consortium will be able to interact on a very regular basis, a necessary condition for significant progress on the above challenges.

- See also: <https://members.loria.fr/Olivier.Devillers/aspaga/>

## ANR Chair in AI

**Participants:** Frédéric Chazal, Marc Glisse, Louis Pujol, Wojciech Riese.

- Acronym : TopAI
- Type : ANR Chair in AI.
- Title : Topological Data Analysis for Machine Learning and AI
- Coordinator : Frédéric Chazal
- Duration : 4 years from September 2020 to August 2024.
- Others Partners: Two industrial partners, the French SME Sysnav and the French start-up MetaFora.
- Abstract:

The TopAI project aims at developing a world-leading research activity on topological and geometric approaches in Machine Learning (ML) and AI with a double academic and industrial/societal objective. First, building on the strong expertise of the candidate and his team in TDA, TopAI aims at designing new mathematically well-founded topological and geometric methods and tools for Data Analysis and ML and to make them available to the data science and AI community through state-of-the-art software tools. Second, thanks to already established close collaborations and the strong involvement of French industrial

partners, TopAI aims at exploiting its expertise and tools to address a set of challenging problems with high societal and economic impact in personalized medicine and AI-assisted medical diagnosis.

#### ANR ALGOKNOT

**Participants:** Clément Maria.

- Acronym : ALGOKNOT.
- Type : ANR Jeune Chercheuse Jeune Chercheur.
- Title : Algorithmic and Combinatorial Aspects of Knot Theory.
- Coordinator : Clément Maria.
- Duration : 2020 – 2023 (3 years).
- Abstract: The project AlgoKnot aims at strengthening our understanding of the computational and combinatorial complexity of the diverse facets of knot theory, as well as designing efficient algorithms and software to study their interconnections.
- See also: <https://www-sop.inria.fr/members/Clement.Maria/>

#### ANR GeMfaceT

**Participants:** Blanche Buet.

- Acronym: GeMfaceT.
- Type: ANR JCJC -CES 40 – Mathématiques
- Title: A bridge between Geometric Measure and Discrete Surface Theories
- Coordinator: Blanche Buet.
- Duration: 48 months, starting October 2021.
- Abstract: This project positions at the interface between geometric measure and discrete surface theories. There has recently been a growing interest in non-smooth structures, both from theoretical point of view, where singularities occur in famous optimization problems such as Plateau problem or geometric flows such as mean curvature flow, and applied point of view where complex high dimensional data are no longer assumed to lie on a smooth manifold but are more singular and allow crossings, tree-structures and dimension variations. We propose in this project to strengthen and expand the use of geometric measure concepts in discrete surface study and complex data modelling and also, to use those possible singular discrete surfaces to compute numerical solutions to the aforementioned problems.

#### 10.2.2 Collaboration with other national research institutes

##### IFPEN

**Participants:** Frédéric Chazal, Marc Glisse, Jisu Kim.

Research collaboration between DataShape and IFPEN on TDA applied to various problems issued from energy transition and sustainable mobility.

#### 10.3 Regional initiatives

##### Metafora

**Participants:** Gilles Blanchard, Bastien Dussap, Marc Glisse, Louis Pujol.

- Type : Paris Region PhD<sup>2</sup> - PhD 2021.
- Title : Analyse de données cytométriques.

The Île-de-France region funds two PhD theses in collaboration with Metafora biosystems, a company specialized in the analysis of cells through their metabolism. The first (Louis Pujol) is supervised by Pascal Massart (Inria team Celeste) and Marc Glisse, and its goal is to improve clustering for this particular type of data. The second one (Bastien Dussap) is supervised by Gilles Blanchard and Marc Glisse and aims to compare samples instead of analyzing just one sample.

## 11 Dissemination

### 11.1 Promoting scientific activities

#### 11.1.1 Scientific events: organisation

##### Member of the organizing committees

- Frédéric Chazal co-organized the thematic trimester at Institut Henri Poincaré, [Geometry and Statistics in Data Sciences](#) from September to December 2022.
- Gilles Blanchard co-organized the [7e Journée Statistique & Informatique pour la Science des Données à Paris-Saclay](#) held at the Institut des Hautes Etudes Scientifiques, January 26, 2022.

#### 11.1.2 Scientific events: selection

##### Member of the conference program committees

- David Cohen-Steiner was a member of the program committee of the Symposium on Geometry Processing 2022.

#### 11.1.3 Journal

##### Member of the editorial boards

- Gilles Blanchard was member of the following journal editorial boards: Annals of Statistics, Electronic Journal of Statistics, Bernoulli.
- Frédéric Chazal is a member of the following journal editorial boards: Discrete and Computational Geometry (Springer), Graphical Models (Elsevier).
- Frédéric Chazal is the Editor-in-Chief of the Journal of Applied and Computational Topology (Springer).

#### 11.1.4 Invited talks

- Pierre Pansu gave a talk on distances between chain complexes in [Montpellier](#) and at the colloquium of Laboratoire de Mathématiques Blaise Pascal.
- Gilles Blanchard gave invited talks at the workshop [Mathematics of machine learning](#), BCIM, Bilbao, Spain; at the workshop [Non-Linear and High Dimensional Inference](#) at IHP (part of the thematic trimester "Geometry and Statistics in Data Science"); at the workshop [3-Day Meeting of statisticians in Paris](#) at IHP; at the [Symposium on Inverse Problems: From experimental data to models and back](#) at the University of Potsdam, Germany.
- Blanche Buet gave talks at the workshop [Conference Shape Optimization, related topics and applications](#); at the workshop Measure-theoretic approaches and Optimal Transportation in Statistics at IHP (part of the thematic trimester "Geometry and Statistics in Data Science").

### 11.1.5 Leadership within the scientific community

- Frédéric Chazal is the Director of the [DATAIA Institute](#) at Université Paris-Saclay.
- Frédéric Chazal is a member of the board of directors of the DIM project AI4IDF of the Région Ile-de-France.
- Clément Maria is co-head (with Théo Lacombe) of the GT GeoAlgo within the [GdR IM\(\)](#).

### 11.1.6 Research administration

- Pierre Pansu is deputy director of the [FMJH](#).
- Marc Glisse is president of the CDT at Inria Saclay.
- Blanche Buet is member of the CCUPS (Commission Consultative de l'Université Paris-Saclay), Laboratory council and "comité parité" of LMO.

## 11.2 Teaching - Supervision - Juries

### 11.2.1 Teaching

- GESDA Introductory School (IHP thematic quarter): Mathieu Carrière and Marc Glisse, 11h15 eq-TD, IESC Cargèse, France.
- Master: Frédéric Chazal, Analyse Topologique des Données, 30h eq-TD, Université Paris-Sud, France.
- Master: Marc Glisse and Clément Maria, Computational Geometry Learning, 36h eq-TD, M2, MPRI, France.
- Master: Frédéric Cazals and Mathieu Carrière, Foundations of Geometric Methods in Data Analysis, 24h eq-TD, M2, École Centrale Paris, France.
- Master: Frédéric Cazals and Jean-Daniel Boissonnat and Mathieu Carrière, Geometric and Topological Methods in Machine Learning, 30h eq-TD, M2, Université Côte d'Azur, France.
- Master: Frédéric Chazal and Julien Tierny, Topological Data Analysis, 38h eq-TD, M2, Mathématiques, Vision, Apprentissage (MVA), ENS Paris-Saclay, France.
- Master: Gilles Blanchard, Mathematics for Artificial Intelligence 1, 70h eq-TD, IMO, Université Paris-Saclay, France.
- Master: Blanche Buet, TD-Distributions et analyse de Fourier, 60h eq-TD, M1, Université Paris-Saclay, France.
- Master: Marc Glisse, Conception et analyse d'algorithmes, 40h eq-TD, M1, École Polytechnique, France.
- Undergrad: Marc Glisse, Mécanismes de la programmation orientée-objet, 40h eq-TD, L3, École Polytechnique, France.

### 11.2.2 Supervision

- PhD: Louis Pujol, Partitionnement de données cytométriques, Pascal Massart and Marc Glisse, defended in December 2022.
- PhD: Etienne Lasalle, TDA and statistics on graphs. Frédéric Chazal and Pascal Massart. Defended in December 2022.
- PhD: Alex Delalande, optimal transport. Frédéric Chazal and Quentin Mérigot. Defended in December 2022.

- PhD: El Mehdi Saad, Efficient online methods for variable and model selection. Gilles Blanchard and Sylvain Arlot. Defended in December 2022.
- PhD: Owen Rouille, Large scale computations of 3-manifold invariants. Jean-Daniel Boissonnat et Clément Maria. Defended in September 2022.
- PhD: Daniel Perez, Homologie persistante des processus stochastiques et leurs fonctions zeta. Claude Viterbo and Pierre Pansu. Defended in July 2022.
- PhD in progress: Vadim Lebovici, Laplace transform for constructible functions. Started September 2020. Steve Oudot and François Petit.
- PhD in progress: Christophe Vuong, Random hypergraphs. Started November 2020. Laurent Decreusefond and Marc Glisse.
- PhD in progress: Bastien Dussap, Comparaison de données cytométriques, started October 1st, 2021, Gilles Blanchard and Marc Glisse.
- PhD in progress: Olympio Hacquard, Apprentissage statistique par méthodes topologiques et géométriques, 2020, Gilles Blanchard and Clément Levrard.
- PhD in progress: Hannah Marienwald, Transfer learning in high dimension. Started September 2019. Gilles Blanchard and Klaus-Robert Müller.
- PhD in progress: Jean-Baptiste Fermanian, Estimation de Kernel Mean Embedding et tests multiples en grande dimension. Started September 2021. Gilles Blanchard and Magalie Fromont-Renoir.
- PhD in progress: Antoine Commaret, Persistent Geometry. Started September 2021. David Cohen-Steiner and Indira Chatterji.
- PhD in progress: Lucas Brifault, Théorie de la mesure géométrique appliquée pour la modélisation de formes complexes. Started May 2022. David Cohen-Steiner and Mathieu Desbrun.
- PhD in progress: David Loiseaux, Multivariate topological data analysis for statistical machine learning. Started November 2021. Mathieu Carrière and Frédéric Cazals.
- PhD in progress: Wojciech Rieser, TDA for curve data. Started October 2020. Frédéric Chazal and Bertrand Michel.
- PhD in progress: Alexandre Guérin, Movement analysis from inertial sensors. Started on October 2021. Frédéric Chazal and Bertrand Michel.
- PhD in progress: Jérémie Capitaou-Miniconi, deconvolution for singular measures with geometric support. Started on October 2020. Frédéric Chazal and Elisabeth Gassiat.
- PhD in progress: Charly Boricaud, Geometric inference for Data analysis: a Geometric Measure Theory perspective. Started on October 2021. Blanche Buet, Gian Paolo Leonardi et Simon Masnou.
- PhD in progress: Hugo Henneuse. Statistical Foundations of Topological Data Analysis for multidimensional random fields. Started on October 2022. Frédéric Chazal and Pascal Massart.
- PhD in progress: Laure Ferraris. Measure-dependent metric learning and applications in Topological Data Analysis. Started on October 2022. Frédéric Chazal.
- PhD in progress: Georg Grützner. Espaces de Möbius et géométrie à grande échelle. Started in May 2020. Pierre Pansu.

### 11.2.3 Juries

- Frédéric Chazal was the president of the PhD committee of Louis Pujol.
- Frédéric Chazal was the president of the HDR committee of Nicolas Chenavier (Université du littoral).
- Pierre Pansu was a member of the PhD committee of Oussama Bensaid (Université Paris-Cité).
- Pierre Pansu was a member of the comité de sélection PR in Didactique des Sciences at Université Paris-Saclay.
- Mathieu Carrière was a member of the jury for the Gilles Kahn PhD prize awarded by SIF
- Gilles Blanchard was the president of the PhD committee of Perrine Lacroix (University Paris-Saclay)
- Gilles Blanchard was a reviewer for the thesis and president of the PhD committee of Thibault Randrianarisoa (Sorbonne Université)
- Gilles Blanchard was a reviewer for the thesis and member of the PhD committee of Ulysse Marteau-Ferey (Inria/SIERRA)
- Gilles Blanchard was a reviewer for the HDR and member of the HDR committee of Christophe Denis (University Paris-Est)
- Blanche Buet was a member of the PhD committee of Romain Petit (INRIA - Paris Dauphine).
- Blanche Buet was a member of the comité de sélection MC 25/26 no. 103/832 « Analyse et géométrie in Nice.
- Blanche Buet was a member of the comité de sélection MC 25/26 "Analyse numérique, modélisation, optimisation, apprentissage, EDP" in Orsay.

## 11.3 Popularization

### 11.3.1 Articles and contents

- Blanche Buet wrote a dissemination paper *Varifolds : des films de savon aux surfaces discrètes*, published in *Maths en pleines formes Express* (2022) for the "23ème salon culture et jeux mathématiques".

## 12 Scientific production

### 12.1 Major publications

- [1] D. Attali, U. Bauer, O. Devillers, M. Glisse and A. Lieutier. 'Homological Reconstruction and Simplification in R<sup>3</sup>'. In: *Computational Geometry* (2014). DOI: [10.1016/j.comgeo.2014.08.010](https://doi.org/10.1016/j.comgeo.2014.08.010). URL: <https://hal.archives-ouvertes.fr/hal-01132440>.
- [2] J.-D. Boissonnat, R. Dyer and A. Ghosh. 'Delaunay Triangulation of Manifolds'. In: *Foundations of Computational Mathematics* 45 (2017), p. 38. DOI: [10.1007/s10208-017-9344-1](https://doi.org/10.1007/s10208-017-9344-1). URL: <https://hal.inria.fr/hal-01509888>.
- [3] J.-D. Boissonnat, R. Dyer, A. Ghosh and S. Y. Oudot. 'Only distances are required to reconstruct submanifolds'. In: *Computational Geometry* 66 (2017), pp. 32–67. DOI: [10.1016/j.comgeo.2017.08.001](https://doi.org/10.1016/j.comgeo.2017.08.001). URL: <https://hal.inria.fr/hal-01583086>.
- [4] J.-D. Boissonnat, K. C. Srikanta and S. Tavenas. 'Building Efficient and Compact Data Structures for Simplicial Complexes'. In: *Algorithmica* (Sept. 2016). DOI: [10.1007/s00453-016-0207-y](https://doi.org/10.1007/s00453-016-0207-y). URL: <https://hal.inria.fr/hal-01364648>.

- [5] B. Buet, G. P. Leonardi and S. Masnou. ‘A Varifold Approach to Surface Approximation’. In: *Archive for Rational Mechanics and Analysis* 226.2 (Nov. 2017), pp. 639–694. DOI: [10.1007/s00205-017-1141-0](https://doi.org/10.1007/s00205-017-1141-0). URL: <https://hal.archives-ouvertes.fr/hal-02141325>.
- [6] F. Chazal, D. Cohen-Steiner and A. Lieutier. ‘A Sampling Theory for Compact Sets in Euclidean Space’. In: *Discrete Comput. Geom.* 41.3 (2009), pp. 461–479. URL: <http://dx.doi.org/10.1007/s00454-009-9144-8>.
- [7] F. Chazal, D. Cohen-Steiner and Q. Mérigot. ‘Geometric Inference for Measures based on Distance Functions’. Anglais. In: *Foundations of Computational Mathematics* 11.6 (2011). RR-6930, pp. 733–751. DOI: [10.1007/s10208-011-9098-0](https://doi.org/10.1007/s10208-011-9098-0). URL: <http://hal.inria.fr/inria-00383685>.
- [8] F. Chazal, S. Y. Oudot, M. Glisse and V. De Silva. *The Structure and Stability of Persistence Modules*. SpringerBriefs in Mathematics. Springer Verlag, 2016, pp. VII, 116. URL: <https://hal.inria.fr/hal-01330678>.
- [9] L. J. Guibas, S. Y. Oudot, P. Skraba and F. Chazal. ‘Persistence-Based Clustering in Riemannian Manifolds’. Anglais. In: *Journal of the ACM* 60.6 (Nov. 2013), p. 38. URL: <http://hal.archives-ouvertes.fr/hal-00923563>.
- [10] M. Mandad, D. Cohen-Steiner, L. Kobbelt, P. Alliez and M. Desbrun. ‘Variance-Minimizing Transport Plans for Inter-surface Mapping’. In: *ACM Transactions on Graphics* 36 (2017), p. 14. DOI: [10.1145/3072959.3073671](https://doi.org/10.1145/3072959.3073671). URL: <https://hal.inria.fr/hal-01519006>.

## 12.2 Publications of the year

### International journals

- [11] A. Aukerman, M. Carriere, C. Chen, K. Gardner, R. Rabadán and R. Vanguri. ‘Persistent homology based characterization of the breast cancer immune microenvironment: a feasibility study’. In: *Journal of Computational Geometry*. Special Issue of Selected Papers from SoCG 2020 12.2 (2022), pp. 183–206. URL: <https://hal.inria.fr/hal-03912324>.
- [12] N. Berkouk and G. Ginot. ‘A derived isometry theorem for sheaves’. In: *Advances in Mathematics* 394 (Jan. 2022), p. 108033. DOI: [10.1016/j.aim.2021.108033](https://doi.org/10.1016/j.aim.2021.108033). URL: <https://hal.archives-ouvertes.fr/hal-03864712>.
- [13] J.-D. Boissonnat, R. Dyer, A. Ghosh and M. Wintraecken. ‘Local Criteria for Triangulating General Manifolds’. In: *Discrete and Computational Geometry* (30th Sept. 2022). DOI: [10.1007/s00454-022-00431-7](https://doi.org/10.1007/s00454-022-00431-7). URL: <https://hal.inria.fr/hal-03794702>.
- [14] B. Buet, G. P. Leonardi and S. Masnou. ‘Weak and approximate curvatures of a measure: a varifold perspective’. In: *Nonlinear Analysis* (2022). DOI: [10.1016/j.na.2022.112983](https://doi.org/10.1016/j.na.2022.112983). URL: <https://hal.science/hal-02141391>.
- [15] B. Buet and M. Rumpf. ‘Mean curvature motion of point cloud varifolds’. In: *ESAIM: Mathematical Modelling and Numerical Analysis* 56.5 (2022), pp. 1773–1808. DOI: [10.1051/m2an/2022047](https://doi.org/10.1051/m2an/2022047). URL: <https://hal.archives-ouvertes.fr/hal-03019123>.
- [16] M. Carriere and B. Michel. ‘Statistical analysis of Mapper for stochastic and multivariate filters’. In: *Journal of Applied and Computational Topology* 6.3 (Sept. 2022), pp. 331–369. DOI: [10.1007/s41468-022-00090-w](https://doi.org/10.1007/s41468-022-00090-w). URL: <https://hal.inria.fr/hal-03912325>.
- [17] O. Cheong, O. Devillers, J.-W. Park and M. Glisse. ‘Covering families of triangles’. In: *Periodica Mathematica Hungarica* (2022). URL: <https://hal.inria.fr/hal-03662311>.
- [18] D. Cohen-Steiner, A. Lieutier and J. Vuillamy. ‘Lexicographic Optimal Homologous Chains and Applications to Point Cloud Triangulations’. In: *Discrete and Computational Geometry* (13th Sept. 2022). DOI: [10.1007/s00454-022-00432-6](https://doi.org/10.1007/s00454-022-00432-6). URL: <https://hal.archives-ouvertes.fr/hal-03870128>.
- [19] O. Hacquard, K. Balasubramanian, G. Blanchard, C. Levrard and W. Polonik. ‘Topologically penalized regression on manifolds’. In: *Journal of Machine Learning Research* 23.161 (Apr. 2022), pp. 1–39. URL: <https://hal.archives-ouvertes.fr/hal-03402076>.



- [20] K. Huszár. ‘On the pathwidth of hyperbolic 3-manifolds’. In: *Computing in Geometry and Topology* 1.1 (2nd Feb. 2022), 1:1–1:19. DOI: [10.57717/cgt.v1i1.4](https://doi.org/10.57717/cgt.v1i1.4). URL: <https://hal.archives-ouvertes.fr/hal-03373577>.
- [21] R. Iniesta, E. Carr, M. Carriere, N. Yerolemou, B. Michel and F. Chazal. ‘Topological Data Analysis and its usefulness for precision medicine studies’. In: *Sort: Statistics and Operations Research Transactions* (2022). DOI: [10.2436/20.8080.02.120](https://doi.org/10.2436/20.8080.02.120). URL: <https://hal.inria.fr/hal-03912322>.
- [22] V. Lebovici. ‘Hybrid Transforms of Constructible Functions’. In: *Foundations of Computational Mathematics* (2022). DOI: [10.1007/s10208-022-09596-2](https://doi.org/10.1007/s10208-022-09596-2). URL: <https://hal.inria.fr/hal-03908728>.
- [23] C. Portaneri, M. Rouxel-Labbé, M. Hemmer, D. Cohen-Steiner and P. Alliez. ‘Alpha Wrapping with an Offset’. In: *ACM Transactions on Graphics*. Proceedings of the ACM SIGGRAPH conference 41.4 (1st June 2022), pp. 1–22. DOI: [10.1145/3528223.3530152](https://doi.org/10.1145/3528223.3530152). URL: <https://hal.inria.fr/hal-03688637>.

### National journals

- [24] C. Icubam, L. Bonnasse-Gahot, M. Dénès, G. Dulac-Arnold, S. Girgin, F. Husson, V. Iovene, J. Josse, A. Kimmoun, F. Landes, J.-P. Nadal, R. Primet, F. Quintao, P. G. Raverdy, V. Rouvreau, O. Teboul and R. Yurchak. ‘ICU Bed Availability Monitoring and analysis in the Grand Est region of France during the COVID-19 epidemic’. In: *Statistique et Société* (17th Mar. 2022). DOI: [10.1101/2020.05.18.20091264](https://doi.org/10.1101/2020.05.18.20091264). URL: <https://hal.science/hal-02620018>.

### International peer-reviewed conferences

- [25] A. Delalande. ‘Nearly Tight Convergence Bounds for Semi-discrete Entropic Optimal Transport’. In: AISTATS 2022 - 25th International Conference on Artificial Intelligence and Statistics. Vol. 151. PMLR. Valencia / Virtual, Spain, 28th Mar. 2022, pp. 1619–1642. URL: <https://hal.science/hal-03396206>.
- [26] M. Glisse and S. Pritam. ‘Swap, Shift and Trim to Edge Collapse a Filtration’. In: SoCG 2022 - 38th International Symposium on Computational Geometry. Berlin, Germany, 7th June 2022. DOI: [10.4230/LIPIcs.SoCG.2022.44](https://doi.org/10.4230/LIPIcs.SoCG.2022.44). URL: <https://hal.inria.fr/hal-03791448>.
- [27] C. Maria and O. Rouillé. ‘Localized Geometric Moves to Compute Hyperbolic Structures on Triangulated 3-Manifolds’. In: 30th Annual European Symposium on Algorithms (ESA 2022). Potsdam, Germany, 5th Sept. 2022. URL: <https://hal.science/hal-03929416>.

### Scientific book chapters

- [28] J.-D. Boissonnat, F. Chazal and B. Michel. ‘Topological Data Analysis’. In: *Novel Mathematics Inspired by Industrial Challenges*. Vol. 38. Mathematics in Industry. Springer International Publishing, 31st Mar. 2022, pp. 247–269. DOI: [10.1007/978-3-030-96173-2\\_9](https://doi.org/10.1007/978-3-030-96173-2_9). URL: <https://hal.inria.fr/hal-03813929>.

### Doctoral dissertations and habilitation theses

- [29] A. Delalande. ‘Quantitative Stability in Quadratic Optimal Transport’. Université Paris-Saclay, 14th Dec. 2022. URL: <https://theses.hal.science/tel-03935445>.
- [30] E. Lasalle. ‘Contributions to statistical analysis of graph-structured data’. Université Paris-Saclay, 5th Dec. 2022. URL: <https://theses.hal.science/tel-03941869>.
- [31] D. Perez. ‘Persistent homology of stochastic processes and their zeta functions’. ENS Paris - Ecole Normale Supérieure de Paris; Université Paris - Saclay, 11th July 2022. URL: <https://hal.archives-ouvertes.fr/tel-03905418>.
- [32] O. Rouillé. ‘Large scale computations of 3-manifolds invariants’. Université Côte d’Azur, 5th Sept. 2022. URL: <https://theses.hal.science/tel-03902902>.

**Reports & preprints**

- [33] J.-D. Boissonnat and K. Dutta. *Dimensionality Reduction for Persistent Homology with Gaussian Kernels*. Jan. 2023. URL: <https://hal.inria.fr/hal-03935155>.
- [34] T. Bonis, F. Chazal, B. Michel and W. Reise. *Topological phase estimation method for reparameterized periodic functions*. 3rd June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03687686>.
- [35] J. Capitao-Miniconi and É. Gassiat. *Deconvolution of spherical data corrupted with unknown noise*. 7th Mar. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03584579>.
- [36] G. Carlier, A. Delalande and Q. Merigot. *Quantitative Stability of Barycenters in the Wasserstein Space*. 20th Sept. 2022. URL: <https://hal.science/hal-03781835>.
- [37] M. Glisse and S. Pritam. *Swap, Shift and Trim to Edge Collapse a Filtration*. 29th Sept. 2022. URL: <https://hal.inria.fr/hal-03791455>.
- [38] G. A. Gruetzner. *Asymptotic-Möbius maps*. 16th Dec. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03746707>.
- [39] D. Loiseaux, M. Carriere and A. Blumberg. *Efficient Approximation of Multiparameter Persistence Modules*. 7th June 2022. URL: <https://hal.inria.fr/hal-03689199>.
- [40] D. Perez. *An attempt of reproduction of Sovacool et al.'s "Differences in carbon emissions reduction between countries pursuing renewable electricity versus nuclear power"*. 26th June 2022. URL: <https://hal.archives-ouvertes.fr/hal-03170325>.
- [41] D. Perez. *Euler and Betti curves are stable under Wasserstein deformations of distributions of stochastic processes*. 19th Dec. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03906966>.
- [42] D. Perez. *On  $C^0$ -persistent homology and trees*. 23rd May 2022. DOI: 10.48550/arXiv.2012.02634. URL: <https://hal.archives-ouvertes.fr/hal-03040819>.
- [43] L. Pujol. *ISDE : Independence Structure Density Estimation*. 5th May 2022. URL: <https://hal.archives-ouvertes.fr/hal-03401530>.
- [44] L. Pujol. *Nonparametric estimation of a multivariate density under Kullback-Leibler loss with ISDE*. 5th May 2022. URL: <https://hal.archives-ouvertes.fr/hal-03660157>.
- [45] E. M. Saad and G. Blanchard. *Constant regret for sequence prediction with limited advice*. 4th Oct. 2022. URL: <https://hal.archives-ouvertes.fr/hal-03797597>.
- [46] T. de Surrel, F. Hensel, M. Carriere, T. Lacombe, Y. Ike, H. Kurihara, M. Glisse and F. Chazal. *RipsNet: a general architecture for fast and robust estimation of the persistent homology of point clouds*. 7th Feb. 2022. URL: <https://hal.inria.fr/hal-03560450>.

**Other scientific publications**

- [47] F. Hensel, M. Glisse, F. Chazal, T. de Surrel, M. Carriere, T. Lacombe, H. Kurihara and Y. Ike. 'RipsNet: a general architecture for fast and robust estimation of the persistent homology of point clouds'. In: *Proceedings of Machine Learning Research*. Proceedings of Topological, Algebraic, and Geometric Learning Workshops 2022 196 (2022), pp. 96–106. URL: <https://hal.inria.fr/hal-03867083>.