Post-doctoral research position:

Sampling Point Clouds with Point Processes and Determinantal Point Processes for Permutation-Invariant Learning, Optimization and Uncertainty Quantification

Host institution	INRIA Grenoble (Grenoble, France) or Ecole des Mines de Saint-Etienne (Saint-Etienne, France)
Duration	12 months
Academic partners	INRIA Grenoble (Olivier ZAHM) Ecole des Mines de Saint-Etienne (Victor TRAPPLER) CIROQUO2 scientific Consortium[1]
Industrial partners	Electricité de France (Julien PELAMATTI, Merlin KELLER) IFP Energies Nouvelles (Morgane MENZ)
Starting date	End of 2026, beginning of 2027, flexible

1 Context

Many scientific and engineering problems involve inputs that naturally take the form of **sets of vectors**, also referred to as **point clouds**. Examples include: wind farm layout design, sensor placement, experimental designs, neural architecture search, mixture-model based classification, optimal uncertainty quantification, etc...

Mathematically, these objects can be represented as

$$X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}, \quad \mathbf{x}_i \in \mathbb{R}^d, \quad n \in \mathbb{N} \tag{1}$$

and possess two characteristic properties:

- Permutation invariance: the order of the points $\mathbf{x}_i$ carries no meaning, and any permutation of their order results in an equivalent point cloud X.
- Variable cardinality: the number of points may vary from one configuration to another. In other words, for the same phenomenon of interest, the number n of points in a cloud X is not known beforehand.

[1] https://ciroquo.ec-lyon.fr/

As an illustrative example of point clouds, one might think about turbine placement in a wind-farm. In this case, the placement of each turbine can be represented as a 2-dimensional point $\mathbf{x}_i$, and the cloud of points X represents the overall layout of the wind-farm. Moreover, the number n of turbines comprising the wind-farm is not necessarily known beforehand, and may vary within given bounds. Finally, it should be clear that the order in which the set contains the various coordinate vectors has no influence on the overall wind-farm representation, thus resulting in a permutation invariant object.

The design and learning problems associated with point clouds cannot be addressed directly with standard methodologies developed for Euclidean spaces of fixed dimension. Recent research efforts have been initiated aimed at developing mathematical and computational tools for optimization, uncertainty quantification, and surrogate modeling/machine learning on spaces of point clouds. In this context, the issue of being able to properly sample point clouds is crucial, as it is a preliminary step for many of the aforementioned topics (e.g., generation of a machine learning training sample or a set of optimization starting solutions). The task of sampling point clouds raises several questions which are sparsely discussed in the literature, such as the definition of a probabilistic distribution over bounded point clouds, the efficient sampling of such distribution if it exists, and potentially the creation of *space-filling* samples.

2 Previous work

Recent preliminary work on this topic has been focused on the development of space-filling sampling methodologies for point-cloud spaces. The work investigated how classical design-of-experiments concepts such as maximin criteria and Latin Hypercube Sampling can be generalized to configuration spaces in which both the positions of points and their number must be optimized simultaneously. A major outcome of this work was the identification of suitable metrics for comparing point clouds of potentially different sizes. In particular, two complementary frameworks were explored:

- Distances derived from optimal transport, including Wasserstein and Sinkhorn divergences (Sow, 2024; Peyré and Cuturi, 2020)
- Kernel-based discrepancies such as the Maximum Mean Discrepancy (MMD)

These metrics, coupled with standard space filling methodologies, enabled the construction of the first sampling strategies in point configuration spaces and highlighted the importance of balancing global space coverage with a measure of topological diversity between sampled point clouds. Initial approaches based on repulsive mechanisms, quantization procedures and distance-based optimization produced promising results, while also revealing fundamental limitations of purely deterministic optimization-based designs. These observations naturally motivate the next phase of the project: the development of a rigorous probabilistic framework based on point process theory (Daley and Vere-Jones, 2003) for generating diverse and representative collections of point clouds. Among the available models, Determinantal Point Processes (DPPs) (Kulesza and Taskar, 2012; Tremblay et al., 2023) appear particularly promising because they provide an intrinsic notion of repulsion and diversity, properties that closely match the objectives of space-filling experimental design.

3 Research Objectives

The proposed postdoctoral project aims to establish the theoretical and algorithmic foundations of point-process-based sampling methods for spaces of point clouds, with a particular emphasis on DPPs and their applications to uncertainty quantification, surrogate modeling/machine learning and engineering optimization.

The project builds upon previous work tackling several aspects of the problem, namely:

- the formulation of point clouds as discrete probability measures defined on continuous spaces; the adaptation of Wasserstein/Sinkhorn and MMD distances to compare point clouds with varying cardinalities;
- the development of maximin-based sampling criteria on spaces of point clouds; the design of iterative optimization algorithms capable of moving, adding and removing points within configurations;
- the investigation of repulsion-based criteria to improve the diversity and coverage of collections of point clouds.

While these approaches demonstrated the feasibility of space-filling designs in configuration spaces, they also raised a number of open questions. In particular, they suggest that diversity should perhaps be enforced directly at the probabilistic level rather than through ad hoc optimization criteria. This observation provides a strong motivation for investigating modern point-process models and, in particular, determinantal point processes. DPPs offer a mathematically principled way to generate diversified samples and have already demonstrated remarkable success in machine learning, Bayesian optimization and experimental design. However, their extension to spaces whose elements are themselves point clouds remains largely unexplored. The central hypothesis of this project is that DPP-inspired sampling strategies can provide a new generation of space-filling designs for permutation-invariant objects, combining theoretical guarantees, computational tractability and practical efficiency for industrial applications.

The main objective of the project is to develop novel probabilistic and computational methodologies for generating representative and diverse samples of point clouds in high-dimensional configuration spaces. Such methodologies are expected to play a key role in: surrogate modeling/machine learning, design of experiments, uncertainty quantification, engineering design and decision-making (Mazoyer et al., 2020; Gautier et al., 2019; Neven and Goedemé, 2026; Bachoc et al., 2020)

The project will investigate how modern point process theory can be leveraged to construct space-filling sampling strategies for sets of vectors. A particular emphasis will be placed on Determinantal Point Processes (DPPs), whose intrinsic repulsive properties naturally promote diversity among samples and make them attractive candidates for experimental design and exploration tasks.

4 Scientific Challenges

The postdoctoral research will address several interconnected questions:

(i) Probabilistic Modeling of Point Clouds

Understand, study and define probability measures over spaces of finite point configurations. Investigate suitable representations of sets of vectors and variable-cardinality objects. Study geometric and statistical properties of such spaces.

(ii) Point Processes on Spaces of Point Clouds

Extend classical point process models to configuration spaces of point clouds. Characterize suitable notions of density and diversity. Study Poisson, Gibbs and determinantal point process formulations.

(iii) Determinantal Point Processes for Space-Filling Sampling

Construct DPPs based on distances or kernels defined between point clouds. Investigate kernels derived from: optimal transport embeddings, kernel mean embeddings (e.g., MMD), geometric representations of point sets, etc... Analyze theoretical guarantees regarding diversity, coverage and discrepancy.

(iv) Algorithms and Large-Scale Sampling

Develop efficient algorithms for sampling DPPs in high-dimensional spaces. Design scalable approximations adapted to high dimensional engineering applications and compare DPP-based approaches with state-of-the-art alternatives.

(v) Applications

The developed methods will be evaluated on both benchmark and industrial test cases, such as wind-farm layout simulation and optimization or sensor placement.

4.1 Expected Outcomes

The project is expected to deliver new theoretical results on point processes defined over spaces of point clouds, novel DPP formulations adapted to permutation-invariant objects, open-source software implementations in Python, numerical benchmarks on representative engineering problems, publications in leading journals and conferences in applied probability, statistics and machine learning.

5 Research Environment

The postdoctoral fellow will join a multidisciplinary collaboration at the interface of: applied probability, statistical learning, optimal transport, uncertainty quantification, design and analysis of computer experiments.

The project will benefit from close interactions between academic researchers and industrial partners through the CIROQUO research network. Regular exchanges with EDF R&D and IFPEN researchers will ensure strong connections between methodological developments and real-world applications.

6 Candidate profile

Applicants should hold a PhD in one of the following disciplines: Applied Mathematics, Probability and Statistics, Machine Learning, Computer Science, Operations Research, or a closely related field.

Particular expertise in several of the following areas is highly desirable:

- Point processes,
- Determinantal point processes,
- Numerical optimization,
- Probability theory,
- Optimal transport,
- Kernel methods,
- Statistical learning,
- Bayesian methods,
- Design of experiments.

Strong programming skills in Python are expected. Experience combining theoretical developments with computational implementations will be particularly appreciated.

7 Supervision

The successful candidate will be jointly supervised by researchers from: INRIA Grenoble (Olivier ZAHM, name.surname@inria.fr) and École des Mines Saint-Étienne (Victor TRAPPLER, name.surname@emse.fr), in collaboration with research engineers from EDF R&D (Julien PELAMATTI, Merlin KELLER, name.surname@edf.fr) and IFP Energies nouvelles (IFPEN) (Morgane MENZ, name.surname@ifpen.fr).

The position is part of a broader collaborative effort involving both academic and industrial partners working on uncertainty quantification, optimization and machine learning for complex engineering systems.

8 Application

To apply, send a Curriculum Vitae as well as a motivation letter to the supervision staff. The e-mail addresses can be found in the previous section.

Keywords

Point Processes • Determinantal Point Processes (DPPs) • Point Clouds • Set-Valued Inputs • Permutation Invariance • Space-Filling Designs • Experimental Design • Optimal Transport • Wasserstein Distance • Maximum Mean Discrepancy • Kernel methods, Uncertainty Quantification • Surrogate Modeling • Machine Learning • Engineering Design

References

Bachoc, François, Alexandra Suvorikova, David Ginsbourger, Jean-Michel Loubes, and Vladimir Spokoiny (Jan. 2020). "Gaussian Processes with Multidimensional Distribution Inputs via Optimal Transport and Hilbertian Embedding". In: *Electronic Journal of Statistics* 14.2, pp. 2742–2772. ISSN: 1935-7524, 1935-7524. DOI: 10.1214/20-EJS1725. (Visited on 07/24/2026).

Daley, DJ and D Vere-Jones (2003). *An Introduction to the Theory of Point Processes*. Probability and Its Applications. New York: Springer-Verlag. ISBN: 978-0-387-95541-4. DOI: 10.1007/b97277. (Visited on 07/24/2026).

Gautier, Guillaume, Rémi Bardenet, and Michal Valko (2019). "On Two Ways to Use Determinantal Point Processes for Monte Carlo Integration". In: *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc. (Visited on 07/30/2026).

Kulesza, Alex and Ben Taskar (Dec. 2012). "Determinantal Point Processes for Machine Learning". In: *Foundations and Trends® in Machine Learning* 5.2-3, pp. 123–286. ISSN: 1935-8237, 1935-8245. DOI: 10.1561/2200000044. arXiv: 1207.6083 [stat.ML]. (Visited on 07/24/2026).

Mazoyer, Adrien, Jean-François Coeurjolly, and Pierre-Olivier Amblard (Aug. 2020). "Projections of Determinantal Point Processes". In: *Spatial Statistics* 38, p. 100437. ISSN: 2211-6753. DOI: 10.1016/j.spasta.2020.100437. (Visited on 07/30/2026).

Neven, Robby and Toon Goedemé (July 2026). "Uncertainty-Aware DPP Sampling for Active Learning". In: *18th International Conference on Computer Vision Theory and Applications*, pp. 95–104. ISBN: 978-989-758-634-7. (Visited on 07/30/2026).

Peyré, Gabriel and Marco Cuturi (Mar. 2020). *Computational Optimal Transport*. DOI: 10.48550/arXiv.1803.00567. arXiv: 1803.00567 [stat.ML]. (Visited on 07/24/2026).

Sow, Babacar (Nov. 2024). "Optimizing and metamodeling functions defined over clouds of points". Theses. Ecole Nationale Supérieure des Mines de Saint-Etienne. DOI: 10.70675/9a1d3cf1z63e9z43d3z9e8cz URL: https://hal-emse.ccsd.cnrs.fr/tel-04894626.

Tremblay, Nicolas, Simon Barthelmé, Konstantin Usevich, and Pierre-Olivier Amblard (Feb. 2023). "Extended L-ensembles: A New Representation for Determinantal Point Processes". In: *The Annals of Applied Probability* 33.1, pp. 613–640. ISSN: 1050-5164, 2168-8737. DOI: 10.1214/22-AAP1824. (Visited on 07/24/2026).