PhD thesis with funding

Statistics and machine learning for predicting complex outputs with applications to nuclear safety

Scientific goals of the thesis:

The prediction of scalar-valued or vector-valued outputs is now well understood in statistics and machine learning. However, in many applications, one would be rather interested in predicting a quantity of interest taking values in more complex spaces, such as geometric manifolds, spaces of measures, and Euclidean subspaces with non-linear constraints. Importantly, the goal is to address output spaces with no linear structure, which makes most standard prediction methods not applicable. We refer to [1] for a review on this problem.

Several possibilities will be explored, focusing in particular on predicting measures. Some recent works [2, 3, 4] on measurement interpolation focus on the notion of splines to obtain a smooth reconstruction. The essence of the problem consists in building a coupling between the different measures to be interpolated. The notion of Wasserstein's barycenter [5] offers such a coupling and has been shown to be very efficient [6, 7]. Therefore, Wasserstein barycenters will be exploited in this thesis.

The implementation of a Bayesian model based on Gaussian processes will also be studied. Gaussian processes and their extensions are widely studied for vector or more complex inputs [8,9]. Nevertheless, there is little work on complex outputs, that is the focus of this thesis. A Bayesian model is particularly interesting for the implementation of iterative and probabilistic strategies for planning experiments [10].

The developed methodology will be applied to real problems in computational fluid dynamics, in collaboration with the Radioprotection and Nuclear Safety Institute (IRSN). Examples of outputs are then flow images and histograms of physical quantities of interest. The industrial motivation is then a better understanding of two-phase flow in steam generators of nuclear plants.

Supervising team and organizational details:

The thesis is funded for 3 years, in the frame of a collaboration project between Institut de Mathématiques de Toulouse (IMT), École Centrale de Marseille (ECM), IRSN and EDF R&D. The main location is IMT, and there will be regular visits to ECM and IRSN (Cadarache). Optionally, a funded master internship before the thesis is possible. The two directors are François Bachoc (IMT) and Jean Baccou (IRSN). Two other supervisors are Thibaut Le Gouic et Jacques Liandrat (ECM).

How to apply:

Applications will be considered starting in March 2023 and until the position is filled. The candidates should have master-level skills in mathematics / statistics / machine learning. Please send a CV (either in English or French) to

- François Bachoc : francois.bachoc@math.univ-toulouse.fr
- Jean Baccou : jean.baccou@irsn.fr
- Thibaut Le Gouic : thibaut.le-gouic@centrale-marseille.fr
- Jacques Liandrat : jacques.liandrat@centrale-marseille.fr

Bibliography:

[1] A. Menafoglio, P. Secchi, Statistical Analysis of Complex and Spatially Dependent Data: A Review of Object Oriented Spatial Statistics, European journal of operational research, 258(2), pp. 401-410, 2017.
[2] J.-D. Benamou, T. O. Galloüet, and F.-X. Vialard. “Second-Order Models for Optimal Transport and Cubic Splines on the Wasserstein Space”. In: Foundations of Computational Mathematics 19.5 (2019), pp. 1113-1143.
[3] Y. Chen, G. Conforti, and T. T. Georgiou. “Measure-Valued Spline Curves: An Optimal Transport Viewpoint”. In: SIAM Journal on Mathematical Analysis 50.6 (2018), pp. 5947-5968.
[4] S. Chewi et al. Fast and Smooth Interpolation on Wasserstein Space. Oct. 22, 2020. arXiv: 2010.12101 [math, stat]. url: http://arxiv.org/abs/2010.12101.
[5] M. Agueh and G. Carlier. “Barycenters in the Wasserstein Space”. In: SIAM Journal on Mathematical Analysis 43.2 (2011), pp. 904-924.
[6] J. Rabin et al. “Wasserstein Barycenter and Its Application to Texture Mixing”. In: International Conference on Scale Space and Variational Methods in Computer Vision. Springer, 2011, pp. 435-446.
[7] N. Bonneel, G. Peyré, M. Cuturi. “Wasserstein barycentric coordinates: histogram regression using optimal transport”. In: ACM Trans. Graph. 35.4 (2016), pp. 71-1.
[8] C. E. Rasmussen, K.E. Williams. “Gaussian processes for machine learning”. MIT Press, 2016.
[9] F. Bachoc, A. Suvorikova, D. Ginsbourger, J.M. Loubes, V. Spokoiny. “Gaussian processes for multidimensional distribution inputs via optimal transport and Hilbertian embedding”. In: Electronic journal of statistics. 14.2 (2020), pp. 2742-2772.
[10] C. Chevalier, J. Bect, D. Ginsbourger, E. Vazquez, V. Picheny, Y. Richet. “Fast parallel Kriging-based stepwise uncertainty reduction with application to the identification of an excursion set”. In: Technometrics. 56.4 (2014), pp. 455-465.