# Combining statistical learning with deep learning for improved exoplanet detection and characterization

Olivier Flasseur*[1], Théo Bodrito[2], Julien Mairal[3], Jean Ponce[2,4], Maud Langlois[5], and Anne-Marie Lagrange[6,7]

[1]Centre de Recherche Astrophysique de Lyon – École Normale Supérieure - Lyon, Université Claude Bernard Lyon 1, Institut National des Sciences de l'Univers, Centre National de la Recherche Scientifique, Institut national des sciences de lÚnivers – France

[2]INRIA Paris – Inria, ENS, CNRS, PSL Research Univ. – France

[3]INRIA Grenoble – Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Inria Grenoble Rhone-Alpes, 655 av. de l'Europe, 38335 Montbonnot, France – France

[4]Courant Institute of Mathematical Sciences, Center for Data Science, New York University – United States

[5]Centre de Recherche Astrophysique de Lyon – École Normale Supérieure - Lyon, Université Claude Bernard Lyon 1, Institut National des Sciences de l'Univers, Centre National de la Recherche Scientifique, Institut national des sciences de lÚnivers – France

[6]Laboratoire d'études spatiales et d'instrumentation en astrophysique – Institut National des Sciences de l'Univers, Observatoire de Paris, Sorbonne Universite, Centre National de la Recherche Scientifique, Université Paris Cité – France

[7]Institut de Planétologie et dÁstrophysique de Grenoble – observatoire des sciences de l'univers de Grenoble – France

**Abstract**

In previous works, we have developed the PACO algorithm (1, 2, 3) dedicated to the post-processing of A(S)DI observations for exoplanet detection and characterization by direct imaging at high contrast. PACO captures locally the spatial correlations of the nuisance component (i.e., speckles plus other sources of noise) with a multi-variate scaled mixture of Gaussian models. It delivers reliable detection confidences with an improved sensitivity with respect to the classical processing methods of the field (e.g., cADI, PCA, TLOCI). However, it remains room for improvement, especially at short angular separations.

I will present the "deep PACO" algorithm (4, 5, 6) that combines the statistics-based model of PACO with a deep learning model in a three steps algorithm. First, the data are centered and whitened locally using the PACO framework to improve the stationarity and the contrast in a pre-processing step. Second, a convolutional neural network is trained from scratch, in a supervised fashion, to detect the signature of synthetic sources in the pre-processed science data. Finally, the trained network is applied to the pre-processed observations and delivers a detection map. Photometry of detected sources is estimated by a second deep neural network. Both models are trained from scratch with custom data augmentation strategy allowing to generate large training sets from a spatio(-spectro)-temporal dataset.

*Speaker

As a proof of concept, we applied our method on tens of datasets from VLT/SPHERE (both IRDIS and IFS instruments). We compared the proposed method against state-of-the-art algorithms of the field, including PACO. With ADI, the proposed method leads to a typical improvement by half a magnitude in terms of contrast with respect to the best comparative algorithm. The ultimate detection sensitivity driven by the fundamental photon noise limit can also be reached far from the star on some datasets. A joint processing of spatio-temporo-spectral observations obtained with ASDI allows to further improve the detection sensitivity.

The future thirty meters class telescopes will enable exploring much deeper the inner environment of nearby solar-type stars than existing facilities. This goal raises three challenges from a data science point of view: (i) approaching the ultimate performance of the instruments by an optimal extraction of the signals of the sought objects, (ii) capturing a highly spatially structured nuisance component subject to strong temporal fluctuations, and (iii) building a model of the nuisance component from several datasets to bypass the limits of ADI at very short angular separations. Concerning points (i) and (ii), data-driven approaches combining statistical modeling with deep learning would be highly valuable to model the complexity of such observations. We are considering to adapt and to apply our method on simulated ELT high-contrast data (e.g., from ELT/HARMONI) as a future work. Concerning point (iii), I will discuss some methodological developments we have started in that direction. The methodology we are targeting differs from RDI in the sense that a highly non-linear model will be learned from the observations and will exploit several prior domain knowledge.

(1) Flasseur+, A&A, 618, A138, 2018.

(2) Flasseur+, A&A, 634, A2, 2020.

(3) Flasseur+, A&A, 637, A9, 2020.

(4) Flasseur+, SPIE Adapt. Opt. Syst., 12185, 1154-1167, 2022.

(5) Flasseur+, to be submitted to MNRAS the 3rd of March 2023, ArXiv link available soon.

(6) Flasseur+, to be submitted to EUSIPCO conference the 6th of March 2023, ArXiv link available soon.

**Keywords:** statistical method, deep learning, exoplanet detection, exoplanet characterization