

## Master 2 internship – Direct detection and characterization of exoplanets: statistical learning, multi-epoch and multi-spectral data fusion

**Keywords:** statistical modeling, data-driven approaches, data fusion, inverse problems, nuisance modeling, multivariate data from VLT/SPHERE, high-angular resolution & high-contrast imaging, exoplanet detection & characterization.

**Scientific Context:** The direct observation of the close environment of stars can reveal the presence of exoplanets and circumstellar disks, providing crucial insights into the formation, evolution, and diversity of planetary systems [1]. Given the very small angular separation with respect to the host star and the huge contrast between the (very bright) star and the (very faint) exoplanets and disks, imaging the immediate vicinity of a star is extremely challenging. To overcome these difficulties, advanced observational techniques are used. They include (i) extreme adaptive optics, which compensates in real time for wavefront distortions caused by atmospheric turbulence; (ii) coronagraphy, which partially blocks the star light; and (iii) observing strategies which introduce diversity among the different signals to be unmixed [2]. Dedicated processing methods that combine the recorded spatio-temporo-spectral image series form the last corner-stone of direct imaging and they aim to efficiently suppress the nuisance component (i.e., speckles and noise) corrupting the signals of interest [3]. In this context, data science developments are decisive to improve the detection sensitivity of exoplanets and the accuracy of their physical characterization.

Beyond optimal post-processing of individual observations, fusing multiple observations of the same star taken over different epochs can significantly improve the detection sensitivity. The key challenge in this approach lies in accounting for both the nuisance statistics and the orbital motion of the exoplanet across epochs. To address this, the PACOME algorithm (for PACO Multi-Epoch; [4]) has been recently introduced. PACOME leverages statistical modeling of the nuisance component and its correlations at the local scale within a small pixel patch. This approach is inherited from the PACO algorithm, specifically designed for exoplanet detection from individual (mono-epoch) dataset of observations. The by-products of PACO from each epoch provide sufficient statistics that can be optimally combined using PACOME, while efficiently exploring the Keplerian motion of exoplanets. This multi-epoch strategy yields a combined detection score that is directly interpretable as a measure of detection confidence. In addition to improving sensitivity, PACOME enables the estimation of orbital parameters, along with their joint and marginal distributions. Although PACOME achieves state-of-the-art performance, there remains room for improvement, especially near the star. Here, the assumption of a local-scale statistical description of the nuisance component overlooks larger-scale spatial correlations, thus limiting the method's detection sensitivity.

In this context, data science developments are decisive to improve the detection sensitivity of exoplanets and the accuracy of the estimation of their orbit.

**Research directions:** This project will build on recent advancements in modeling the nuisance component that corrupts high-contrast total intensity observations. The focus will be on improving exoplanet detection and characterization. Possible research directions include:

1/ *Modeling large-scale nuisance correlations:* To address the limitations discussed, the goal is to integrate a more refined modeling of the nuisance component within multi-epoch detection algorithms. This can be achieved using the ASAP approach [5], which approximates the precision matrix (i.e., inverse of the covariance matrix) with a structured, sparse model. This flexible statistical model of the full-field covariance, with reduced complexity and learned directly from the science data, is expected to better capture large-scale correlations than PACO.

**2/ Joint spatio-spectral modeling of large-scale correlations:** Building on point 1/, the objective is to develop a joint spatio-spectral model of the nuisance that accounts for large-scale correlations across both spatial and spectral dimensions jointly.

**Data:** The project will focus on developing / improving new processing algorithms using spectroscopic total intensity observations (i.e., spatio-temporal-spectral data recorded with an Integral Field Spectrograph) from the SPHERE instrument, currently operating on the Very Large Telescope (VLT). Several multi-epochs observations are available to both ground the performance of the proposed algorithm and to search for new exoplanets!

Once a proof of concept is established, simulations for HARMONI, one of the first-light instruments of the upcoming Extremely Large Telescope (ELT), may be considered. In this case, the algorithm will be adapted to account for HARMONI's specific features, particularly its higher spectral resolution. Achieving the required contrast with this instrument will require extended total exposure times on a single star, making a multi-epoch strategy indispensable.

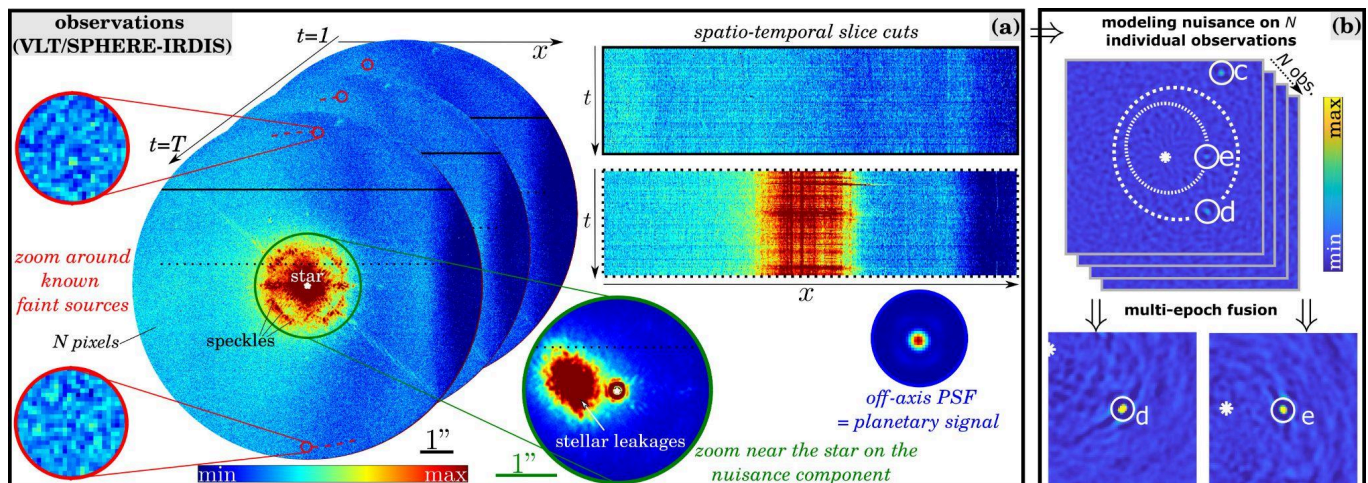
**Desired Skills:** The MSc candidate should have a strong background in signal and image processing, applied mathematics, numerical methods, machine learning or related fields. A strong interest in physics, pluri-disciplinary research and scientific applications is a plus.

**Team and Collaborations:** The MSc candidate will integrate a collaborative project within the [AIRI team](#) at the [Astrophysics Research Center of Lyon](#) (CRAL). The AIRI team has extensive expertise in high-angular-resolution and high-contrast imaging, both in instrumentation and data science. The workplace is CRAL, 9 avenue C. André, 69230 Saint-Genis-Laval.

**Starting date:** As soon as possible.

**Nature of the financial support:** Team or project funding; about 600-700 euros/month.

**Contacts:** [eric.thiebaut@univ-lyon1.fr](mailto:eric.thiebaut@univ-lyon1.fr), [olivier.flasseur@univ-lyon1.fr](mailto:olivier.flasseur@univ-lyon1.fr)



**Panel (a):** A typical observation series from the IRDIS imager of the VLT/SPHERE instrument, showing spatio-temporal diversity for a given wavelength. The exoplanet signals (red circles), appearing as off-axis instrumental PSFs, are corrupted by a strong, multi-correlated nuisance component in the form of speckles and stellar leakages. Spatio-temporal slice cuts along the solid and dashed black lines are shown on the right to illustrate these correlations. **Panel (b):** The top part illustrates detection maps obtained from different observation series of the same star at different epochs. The bottom part shows the fusion of sufficient statistics from individual detection maps, improving the detection significance of exoplanets.

**References (co-signed by members of the team in bold):**

- [1] [Currie+, "Direct imaging and spectroscopy of extrasolar planets", Protostars and Planets VII, 534, 799, 2023](#)
- [2] [Follette, "An introduction to high contrast differential imaging of exoplanets and disks", Publications of the Astronomical Society of the Pacific, 135\(1051\), 2023](#)
- [3] [Pueyo+, "Direct imaging as a detection technique for exoplanets", chapter in Handbook of Exoplanets, 2018](#)
- [4] [Dallant+, "\*\*PACOME: Optimal multi-epoch combination of direct imaging observations for joint exoplanet detection and orbit estimation.\*\*" \*Astronomy & Astrophysics\*, 679, A38, 2024](#)
- [5] [Thiébaud+, "\*\*Beyond FRiM, ASAP: a family of sparse approximation for covariance matrices and preconditioners.\*\*" \*Adaptive Optics Systems VIII\*. Vol. 12185. SPIE, 2022](#)