

## PhD thesis – Statistical modeling and physics-informed deep learning for exoplanet detection and characterization at high-contrast from multi-dimensional data

**Keywords:** statistical modeling, physics-informed deep learning, data-driven approaches, inverse problems, hybrid approaches, instrumental modeling, nuisance modeling, multivariate data, 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 leveraging the telescope's pupil-tracking mode, which introduces 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 (i.e., spectrum and orbit estimation).

**Research Objectives:** The project is structured around three complementary research directions.

*1/ Fine modeling of the nuisance component:* Modeling the spatio-temporo-spectral correlations of the nuisance component is crucial for its efficient suppression. Typically, spatial correlations are local due to speckles (resulting from diffraction in the presence of residual aberrations) but can also extend to larger spatial scales due to stellar light leaking from the coronagraph. State-of-the-art approaches model the nuisance correlations either locally using a statistical model built on small patches [4], or at larger scales through machine learning techniques [5,6,7]. While machine learning approaches are effective at modeling larger-scale correlations, they often lack interpretability. In this context, two main strategies for estimating these large-scale correlations can be explored: (i) flexible statistical models of the full-field covariance [8], with reduced complexity, learned from the science data, (ii) information extraction from massive data archives using deep learning, which should provide interpretable detection scores and reliable estimates.

*2/ Integrating physics-based prior knowledge:* Such information will be leveraged to condition the nuisance model by knowledge on both the image formation process and on the physics of the objects we are observing. Regarding the image formation process, the project will consider integrating instrumental (i.e., optical) models [9] in the learning process to improve the modeling of the nuisance component by, for instance, linking typical aberrations of phase maps to typical structures (e.g., symmetries) of the speckles field. Regarding the objects of interest, the project will consider generating prior information on typical exoplanet spectra. Such prior information can be leveraged from physics-based simulations that account for atmospheric chemistry and cloud presence [10]. Key research questions center on designing effective architectures and learning strategies to create and integrate these priors — whether by integrating them as regularization terms in inverse-problem approaches or employing them as generative models —. For both types of approaches, such hybrid methods aim to bridge the gap between fully data-driven approaches

(which may lack interpretability) and fully parametric approaches (which may lack flexibility and fidelity to the complexity of the observations).

**3/ Data fusion of multiple & heterogeneous data series:** Beyond the optimal processing of individual observations, combining multiple observations of the same star taken at different epochs (i.e., dates) effectively increases the total time of observations of the star's close environment, thereby significantly improving detection sensitivity. The key challenge in this approach lies in accounting for both nuisance statistics and the Keplerian orbital motion of the exoplanet across epochs. To address this, a multi-epoch data fusion algorithm was recently developed [11]. While demonstrating powerful capabilities, this approach is inherently limited by its assumption of a local-scale statistical description of the nuisance, which overlooks larger-scale spatial correlations. This limitation particularly affects detection sensitivity near the star, where strong and large-scale stellar leakages corrupt the observations. In this context, the objective of the thesis project will be to develop nuisance models (see points 1/ and 2/) that provide sufficient statistics for each individual observation. Such by-products will be optimally fused —without loss of information— through a multi-epoch strategy. Methodological developments on the multi-epoch strategy itself are also anticipated, e.g. on the optimized detection score that should robustly account for the uneven quality of data due to evolving observing conditions across epochs.

**Data, Instruments, and Versatility of the Developments:** This project will focus on developing new processing algorithms for total intensity observations (imaging and spectroscopy) from [SPHERE](#), the state-of-the-art high-contrast imaging instrument on the [VLT](#). Available multi-epoch observations from archival data will be leveraged for improved detection sensitivity. Once the proof of concept is established, the methods will be extended to other instruments and imaging modalities to demonstrate their versatility. In particular, data fusion will be explored for heterogeneous datasets from complementary high-contrast instruments, such as [VLT/SPHERE](#) and [JWST](#), which differ in spatial and spectral resolution. The developed approach will generate sufficient statistics for each individual dataset, intrinsically accounting for uncertainties, and opening the door to the robust fusion of heterogeneous high-contrast data. Beyond these aspects, simulations from [HARMONI](#), a first-light instrument of the upcoming ELT, may be considered. Achieving the required detection sensitivity with this instrument will require both modeling the large-scale correlations in the data—due to the ELT's segmented mirror, which results in a highly structured PSF—and extending the total exposure time by observing a target star over multiple epochs, as the exoplanets of interest are extremely faint. These peculiarities make this thesis project both highly timely and valuable for the scientific exploitation of next-generation instruments.

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

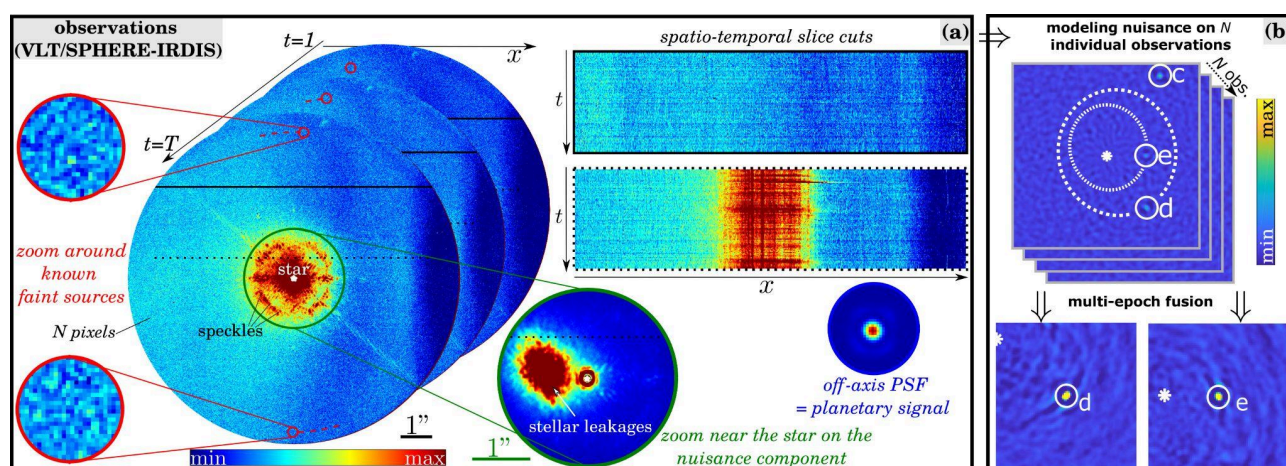
**Team and Collaborations:** The PhD candidate will integrate a collaborative project between the [AIRI team](#) at the [Astrophysics Research Center of Lyon](#) and the [THOTH/WILLOW](#) teams at [Inria](#) Grenoble/Paris. The AIRI team has extensive expertise in high-angular-resolution and high-contrast imaging, both in instrumentation and data science. The THOTH and WILLOW teams have a longstanding expertise in machine learning. Additional collaborations would involve experts in observational astrophysics, including Maud Langlois ([CRAL](#), Lyon), Anne-Marie Lagrange and Anthony Boccaletti ([LIRA](#), Paris), as well as experts in signal processing (e.g., at the [Hubert Curien Laboratory](#)). The candidate will work in a dynamic environment, involving other students and researchers on related projects, including [PEPR Origins](#), which e.g. involves many experts in deep learning.

**Starting Date:** As soon as possible. Possibility to begin as part of an M2 internship or with a short-term pre-doctoral engineering contract.

**Originality and Impact:** This thesis project is inherently interdisciplinary, being at the interface between data science and astronomy. The physics of both the observed objects and the instruments used is central to the

envisioned methodological developments. A key objective is to incorporate this physical knowledge as prior information into learning-based approaches that leverage large data archives. Additionally, the thesis aims to advance statistical modeling of observational noise, with a particular focus on: (i) accurately capturing noise correlations in high-dimensional settings, and (ii) enabling the robust fusion of heterogeneous data while accounting for noise statistics.

These developments are essential for high-contrast imaging, yet no existing approach effectively integrates all these crucial components—both for optimizing the exploitation of current observations and for preparing advanced processing techniques for next-generation instruments. Finally, the application domain (exoplanet imaging) is a hot topic in modern astrophysics, with significant implications for state-of-the-art advancements in data science. The developed methods will be applied to real data from world-class high-contrast imaging instruments, including VLT/SPHERE and JWST, which are the most advanced facilities for exoplanet imaging.



**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.

**Contacts:** [olivier.flasseur@univ-lyon1.fr](mailto:olivier.flasseur@univ-lyon1.fr), [jean.ponce@inria.fr](mailto:jean.ponce@inria.fr), [julien.mairal@inria.fr](mailto:julien.mairal@inria.fr), [theo.bodrito@inria.fr](mailto:theo.bodrito@inria.fr)

#### 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] Flasseur+, "PACO ASDI: an algorithm for exoplanet detection and characterization in direct imaging with integral field spectrographs", 637(A9), *Astronomy & Astrophysics*, 2020
- [5] Flasseur+, "deep PACO: Combining statistical models with deep learning for exoplanet detection and characterization in direct imaging at high contrast", *Monthly Notices of the Royal Astronomical Society*, 527(A1), 2024
- [6] Bodrito+, "MODEL&CO: Exoplanet detection in angular differential imaging by learning across multiple observations", *Monthly Notices of the Royal Astronomical Society*, 534(A2), 2024
- [7] Bodrito+, "A new statistical model of star speckles for learning to detect and characterize exoplanets in direct imaging observations", in *CVPR*, 2025
- [8] Thiébaud+, "Beyond FRIM. ASAP: a family of sparse approximation for covariance matrices and preconditioners", *Adaptive Optics Systems VIII*, SPIE, 2022
- [9] Feng+, "Exoplanet detection via differentiable rendering", *IEEE Transactions on Computational Imaging*, 2025
- [10] Charnay+, "A self-consistent cloud model for brown dwarfs and young giant exoplanets: comparison with photometric and spectroscopic observations", *The Astrophysical Journal*, 854(2), 2018
- [11] Dallant+, "PACOME: Optimal multi-epoch combination of direct imaging observations for joint exoplanet detection and orbit estimation", *Astronomy & Astrophysics*, 679((A38)), 2023