

PACO ASDI: an algorithm for exoplanet detection and characterization in direct imaging with integral field spectrographs

Olivier Flasseur¹, Loïc Denis¹, Éric Thiébaud², and Maud Langlois²

¹ Université de Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, 42023 Saint-Etienne, France

e-mail: surname.name@univ-st-etienne.fr

² Université de Lyon, Université Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR 5574, 69230 Saint-Genis-Laval, France

e-mail: surname.name@univ-lyon1.fr

Received 3 December 2019 / Accepted 6 March 2020

ABSTRACT

Context. Exoplanet detection and characterization by direct imaging both rely on sophisticated instruments (adaptive optics and coronagraph) and adequate data processing methods. Angular and spectral differential imaging (ASDI) combines observations at different times and a range of wavelengths in order to separate the residual signal from the host star and the signal of interest corresponding to off-axis sources.

Aims. Very high contrast detection is only possible with an accurate modeling of those two components, in particular of the background due to stellar leakages of the host star masked out by the coronagraph. Beyond the detection of point-like sources in the field of view, it is also essential to characterize the detection in terms of statistical significance and astrometry and to estimate the source spectrum.

Methods. We extend our recent method PACO, based on local learning of patch covariances, in order to capture the spectral and temporal fluctuations of background structures. From this statistical modeling, we build a detection algorithm and a spectrum estimation method: PACO ASDI. The modeling of spectral correlations proves useful both in reducing detection artifacts and obtaining accurate statistical guarantees (detection thresholds and photometry confidence intervals).

Results. An analysis of several ASDI datasets from the VLT/SPHERE-IFS instrument shows that PACO ASDI produces very clean detection maps, for which setting a detection threshold is statistically reliable. Compared to other algorithms used routinely to exploit the scientific results of SPHERE-IFS, sensitivity is improved and many false detections can be avoided. Spectrally smoothed spectra are also produced by PACO ASDI. The analysis of datasets with injected fake planets validates the recovered spectra and the computed confidence intervals.

Conclusions. PACO ASDI is a high-contrast processing algorithm accounting for the spatio-spectral correlations of the data to produce statistically-grounded detection maps and reliable spectral estimations. Point source detections, photometric and astrometric characterizations are fully automatized.

Key words. techniques: image processing – techniques: high angular resolution – methods: statistical – methods: data analysis

1. Introduction

While most exoplanet detections (Schneider et al. 2011) have been obtained using indirect methods (Santos 2008), such as transit photometry or radial velocity techniques (Lovis & Fischer 2010), direct imaging (Traub & Oppenheimer 2010) appears to be a method of choice for the detection and characterization of young and massive exoplanets. To date, only a few exoplanets (over the 4300 known ones) have been successfully detected (Marois et al. 2008; Lagrange et al. 2009; Nielsen et al. 2012; Bailey et al. 2013; Macintosh et al. 2015; Chauvin et al. 2017; Keppler et al. 2018, see also Bowler 2016; Pueyo 2018; Nielsen et al. 2019 for recent reviews on direct imaging) since the emergence of this technique in the early 2000s. This can be explained by the difficulty in detecting the very faint signature of the exoplanets (typical levels of contrast with the host star are between 10^{-5} and 10^{-6} in the near infrared). The use of a coronagraph to mask out the host star jointly with an extreme adaptive optics system is mandatory for reaching these challenging contrasts. Exoplanet hunting facilities dedicated to direct observations such as VLT/SPHERE (Beuzit et al. 2008,

2019), GEMINI/GPI (Macintosh et al. 2014), Subaru/SCEXAO (Jovanovic et al. 2015), Keck/VORTEX (Howard et al. 2010), and Magellan/MagAO (Morzinski et al. 2014) are equipped with these cutting-edge optical systems.

Several observation strategies can be used for direct imaging. The most popular observation mode is angular differential imaging (ADI, Marois et al. 2006). It consists of tracking the target star and maintaining the telescope pupil stable over time while the field of view rotates. Such observational sequence produces 3D datasets (2D + time) in which the speckles (structured background resulting from instrumental aberrations) are quasi-static, while the signature of the companions describes an apparent circular motion along around its host star. Spectral differential imaging (SDI, Racine et al. 1999) consists of recording simultaneously images in several spectral channels using an integral field spectrograph (IFS). Reduced 3D datacubes (2D + spectral) are obtained by mapping raw observations of the IFS cameras into a multi-spectral cube (Pavlov et al. 2008). In the reduced datacubes, the stellar speckles that are due to diffraction are very similar from one wavelength to the other, apart from

their chromatic scaling (Perrin et al. 2003). After compensating for this scaling, speckles are aligned and can be combined to cancel out thus revealing the presence of off-axis sources whose positions are achromatic. A natural extension of ADI and SDI is to use them simultaneously.

This hybrid observation mode called angular and spectral differential imaging (ASDI) produces 4D datasets (2D + time + spectral), combining the properties of both ADI and SDI. Using ASDI datasets such as the ones obtained with the VLT/SPHERE-IFS instrument brings a spectral diversity (Macintosh et al. 2014; Gerard et al. 2019) compared to simple ADI. The discrimination between the signal from off-axis sources and the background signal due to stellar speckles is thus improved. In addition, ASDI datasets allow both the detection and the characterization of the detected exoplanets (Beuzit et al. 2019). Such a characterization is performed by fitting orbit models and exoplanet atmospheric models to the estimated astrometry and photometry (Vigan et al. 2010). Physical information such as age, effective temperature, composition or surface gravity can be derived for the detected exoplanets (Müller et al. 2018; Cheetham et al. 2019; Mawet et al. 2019; Claudi et al. 2019; Mesa et al. 2019a).

Whatever the observation mode, the recorded images are combined by a post-processing method in order to cancel out as much as possible the stellar speckles which largely dominate the exoplanet signal. Current state-of-the-art detection algorithms applied in direct imaging can be split into two families: algorithms specifically designed to work in SDI mode and algorithms initially designed to work in ADI mode which were later adapted to also work in ASDI mode.

There are few methods specific to SDI. They are mainly based on physical modeling of the stellar point spread function (PSF). The PeX algorithm (Thiébaud et al. 2016; Devaney & Thiébaud 2017) derives a model of the chromatic dependence of the speckles based on diffraction theory. The MEDUSAE method (Ygouf 2012; Cantalloube 2016; Cantalloube et al. 2018) uses an analytic model of the coronagraphic PSF and performs speckle modeling by an inverse problem approach that estimates phase aberrations from the measurements. It also includes an object restoration step via a deconvolution procedure combined with suitable regularization penalties.

There is a much larger variety of ADI processing methods. Several of them are derived from the LOCI algorithm (Lafrenière et al. 2007) in which a stellar PSF image is created by combining images selected in a library of data acquired under experimental conditions similar to those of the observation of interest. The combined images are selected and weighted in order to minimize the residual noise inside annular regions of the images. This combined image is then subtracted from the recorded images to attenuate the speckles and enhance the exoplanet signal. Several adaptations of this general principle have been proposed in the literature, such as the ALOCI (Currie et al. 2012a,b), TLOCI (Marois et al. 2013, 2014), or MLOCI (Wahhaj et al. 2015) algorithms. Among these, the TLOCI algorithm has become one of the gold-standard methods to process ADI datasets. It differs from the standard LOCI algorithm on the construction of the reference stellar PSF. Instead of only minimizing the residual noise, TLOCI also maximizes the exoplanet throughput. Another group of methods models the fluctuations of the stellar speckles (i.e., the on-axis PSF) by a low-dimensional subspace. The exoplanet signal is thus captured on the subspace orthogonal to the subspace that captures fluctuations of the stellar speckles. The data are projected on an orthogonal basis created by principal components analysis (PCA). This is the principle behind the widely used KLIP

algorithm (Soummer et al. 2012; Absil et al. 2013) which builds a basis of the subspace capturing the stellar PSF by performing a Karhunen-Loève transform that is, a PCA¹ of the images from the reference library. To obtain a model of the stellar PSF to subtract in order to attenuate the speckles, the science data is projected onto a predetermined number of modes. Interestingly, as demonstrated by Savransky (2015), the (A,M,T)LOCI- and KLIP-type methods can be seen as two closely related instances of the general class of algorithms called Blind Source Separation (BSS). The LLSG method (Gonzalez et al. 2016) is also based on subspace approaches since it decomposes the datasets into low-rank, sparse and Gaussian components. Other ADI processing methods are based on a statistical framework. For example, the MOODS method (Smith et al. 2009) performs a joint estimation of the exoplanet amplitude and of the stellar speckles. The ANDROMEDA algorithm (Mugnier et al. 2009; Cantalloube et al. 2015) forms differences in temporal images to suppress stellar speckles and performs the detection of differential off-axis PSFs. A generalized likelihood ratio test is then evaluated to perform the detection. A matched filter approach can also be used on the PCA residuals to perform the detection (Ruffio et al. 2017). All these algorithms based on a statistical framework encompass the estimation of the exoplanet throughput and detection confidences through a maximum likelihood approach under a white and Gaussian hypothesis which is a rather crude assumption for this kind of data. The PACO algorithm (Flasseur et al. 2018a) is also based on a maximum likelihood approach but with a more consistent statistical model self-calibrated on the data and accounting for the spatial covariances of the speckles at the scale of small patches. Recently, some works have also applied deep learning techniques to direct imaging (Gonzalez et al. 2018; Yip et al. 2019).

Most of these algorithms are subject to different limitations. Generally, they are not fully unsupervised so that the tuning of several parameters is often mandatory to reach the best performance of the method. Such tuning is time-consuming and should ideally be repeated for each dataset since it depends on the dataset properties (considered spectral bands, number of temporal and spectral frames, quality of the observations, amount of parallactic rotation, etc.). For both SDI and ASDI processing, the recorded images at wavelength λ are scaled by a factor $\lambda_{\text{ref}}/\lambda$, where λ_{ref} is a reference wavelength, so that the on-axis PSF and the speckle field are approximately aligned throughout the ASDI stack (reduced chromatic variations). Due to the difference between the scaling factor applied respectively to the shortest and the longest wavelengths, only a central area of the field of view is covered by all rescaled images. Some source detection techniques process only that area common to all wavelengths. This leads to a drastic reduction of the field of view (typically 25% to 50%), which limits the ability to detect sources. In addition, all ADI and ASDI processing methods are subject to the so-called self-subtraction phenomenon. By combining information (either by image subtraction as in the TLOCI type methods, or by modes subtraction as in the PCA type methods) at different times or wavelengths to attenuate the speckle background, the signal of the exoplanets is also attenuated. Consequently, the photometry is not intrinsically preserved so that an additional calibration step via Monte-Carlo injections is mandatory to compensate for the exoplanet self-subtraction. Finally, the main limitation of existing approaches is the lack of control of the probability of false alarms on the detection maps and contrast curves (see Sect. 5

¹ For discrete data, the Karhunen-Loève transform and the PCA are identical procedures.

for a discussion). It is common for state-of-the-art methods to produce detection maps with many more false alarms than theoretically expected.

Based on an analysis of the limitations of existing algorithms for ASDI data processing and of the needs driven by the new planet finder instruments, the following desirable specifications for exoplanet detection algorithms may be listed as:

- automatic source detection with statistical guarantees (i.e., control over the probability of false alarms),
- characterization of the sources detected: subpixel astrometry and unbiased estimation of the relative spectrum², with reliable confidence intervals,
- ability to process the whole field of view covered by the instrument without artifacts,
- computation of a map reporting the contrast achieved for a source to be detected at a given detection threshold.

In this paper, we attempt to address these different points by deriving an algorithm from a data-driven statistical modeling of ASDI observations. The proposed algorithm, named PACO ASDI, is an extension of our ADI exoplanet detection method PACO introduced in Flasseur et al. (2018a). PACO can be used to process independently each spectral channel of ASDI datasets (Flasseur et al. 2018b). The main methodological adaptations of PACO ASDI for joint ASDI processing are the following: (i) modeling of local covariances based both on temporal and spectral information (see Sect. 2), (ii) adaptation to the time-specific and wavelength-specific magnitude of the background fluctuations (see Sect. 2), (iii) a general approach combining detection maps at different wavelengths that accounts for spectral correlations (see Sect. 3.2) that is also beneficial to combine detection maps from other existing algorithms (see Appendix F), and (iv) estimation of the spectrum of sources, including a parameter-free spectral smoothing (see Sect. 4.2).

Figure 1 gives the general scheme of PACO ASDI to which we will refer throughout the paper. It illustrates the four main steps of the algorithm:

- learning of a model of the background that accounts for the patch spatial covariance (step ①, detailed in Sect. 2),
- single-wavelength detection, by application of detection theory to our statistical model of the background (step ②, detailed in Sect. 3.1),
- multi-wavelength detection, by combining the single-wavelength detection maps; this is achieved by learning the spectral correlations between single-wavelength detection maps (steps ③–⑥) and introducing a (coarse) prior information on the spectrum of the source (detailed in Sect. 3.2),
- once a source has been detected, its astrometry and photometry is estimated during a characterization step (step ⑦, described in Sect. 4), by iteratively refining the source parameters (angular location, spectrum and total flux) and the statistical model of the background (spatial and spectral correlations).

Statistical modeling of the spatial, temporal, and spectral fluctuations of ASDI datasets is a guiding thread throughout the paper for grounding the detection and estimation method and to obtain reliable indications on the probability of false alarms, astrometric, and photometric confidence intervals. Those statistical guarantees are essential to the astronomers to automatize the detection and analysis of ASDI datasets, for the scientific exploitation of the results, and also for characterizing the performance of the instrument (detection limits and photometric

accuracy depending on the observation strategy, the observation conditions, and the performance of the adaptive optics + coronagraph).

This paper is organized as follows. We describe in Sect. 2 our statistical modeling of the background fluctuations for ASDI datasets. In Sect. 3, we explain how to obtain single wavelength and combined detection maps at a controlled probability of false alarms. Section 4 details our parameter-free and regularized spectrum estimation procedure applied to the detected sources. In Sect. 5, we illustrate on VLT/SPHERE-IFS datasets the performance of the proposed PACO ASDI algorithm in terms of detection maps, achievable contrast, and spectrum estimation. Finally, Sect. 6 presents the paper’s conclusions.

2. Statistical modeling of background fluctuations in ASDI

After speckle alignment by spectral image magnification, background structures (i.e., stellar speckles) are approximately constant (up to a multiplication by a chromatic factor accounting for the star spectrum) through time and the wavelengths. A closer observation reveals some temporal and spectral fluctuations. These fluctuations are spatially structured. It is essential to model these fluctuations in order to discriminate between an insignificant change of the background and a point source. We describe in this section a statistical model of the background fluctuations. The detection and source characterization algorithm PACO ASDI is grounded on this model.

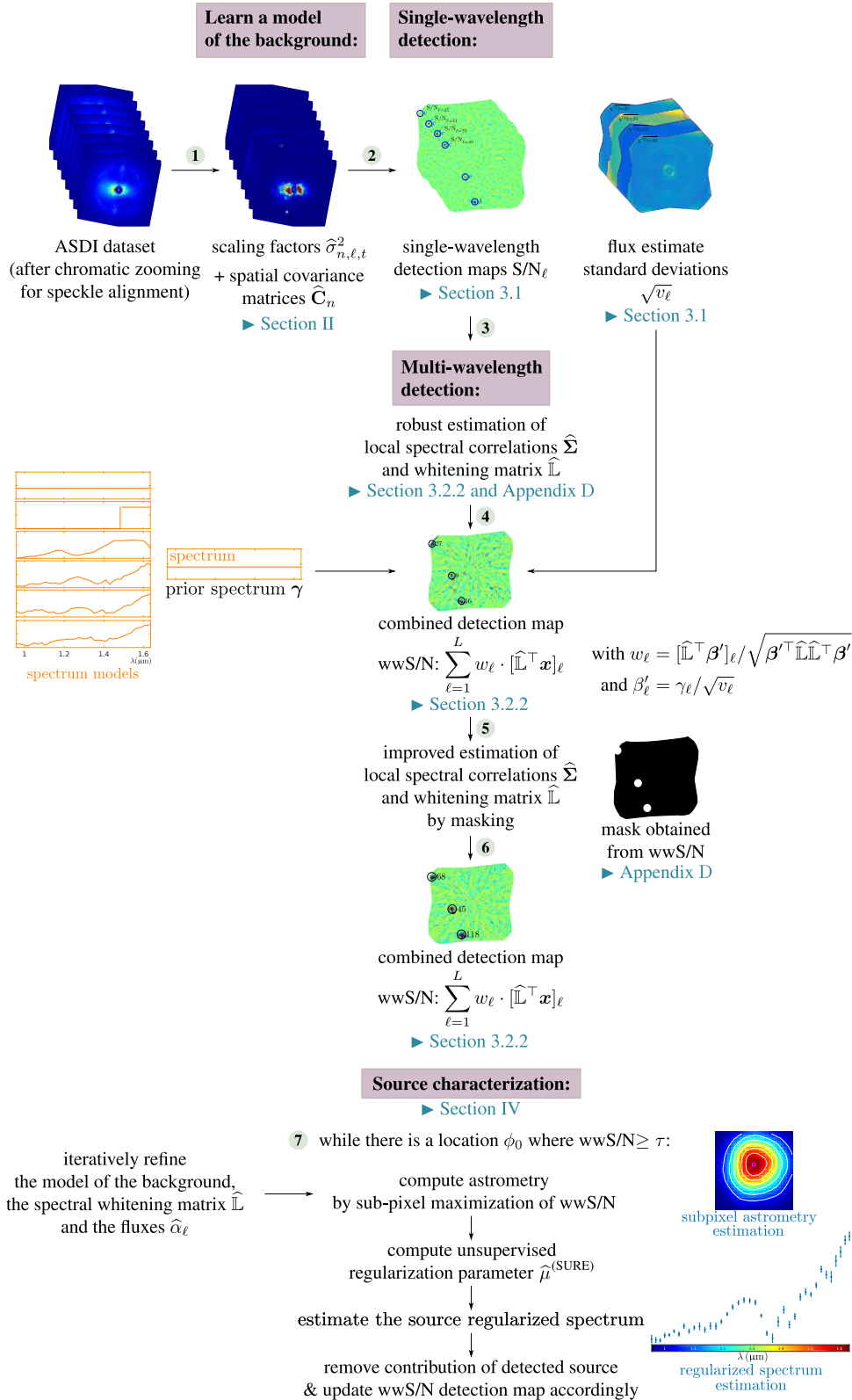
2.1. Local multivariate Gaussian model

Our modeling of the spatial covariances in ADI datasets with PACO algorithm led to two conclusions: (i) to account for the nonstationarity of the background, local modeling is necessary; and (ii) given the limited number of samples available at any given location, a trade-off must be found between the size of the covariance matrices and the estimation variance.

To extend the modeling from ADI datasets to the 4D spatio-temporo-spectral datacubes of ASDI, we keep a local Gaussian modeling: parameters of the Gaussians are estimated by analyzing patches extracted at a given location. For the n th pixel³ of the field of view (identified by its 2D angular direction θ_n on the sky with respect to the stellar center), we extract $T \cdot L$ patches (where T is the number of temporal frames and L is the number of spectral channels), each made of K pixels. The patch size is constant for a given instrument and is fixed with the same empirical rule than the one derived in the PACO algorithm: it should be chosen so that twice the full width at half maximum (FWHM) of the off-axis PSF is encompassed by the patches. Those patches $\mathbf{r}_{n,\ell,t}$ are all centered on the same sky location θ_n but correspond to different frames and spectral channels leading to the local collection $\{\mathbf{r}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T}$, where ℓ indicates the spectral channel and t the frame index. If this collection contains no off-axis point source, we model each patch $\mathbf{r}_{n,\ell,t}$ as a random realization of the K -dimensional Gaussian $\mathcal{N}(\mathbf{m}_{n,\ell}, \sigma_{n,\ell,t}^2, \mathbf{C}_n)$. The mean patch $\mathbf{m}_{n,\ell}$ is the same for all t but is chromatic. The $K \times K$ covariance matrix is modeled as a product of two factors: a time and wavelength-dependent scaling $\sigma_{n,\ell,t}^2$ and a spatial covariance matrix \mathbf{C}_n that are constant for a given patch collection extracted around pixel n . This modeling follows the two guidelines: (i) local

² In this paper, the term relative spectrum (or more simply spectrum) refers to the estimated spectrum of the detected sources before correction for the stellar spectrum.

³ We note that a reminder of the main notations used throughout the paper is given in Table 1.


Fig. 1. Scheme of PACO ASDI algorithm.

adaptivity to account for background nonstationarities, in particular, the model is specific to a given spatial location and captures different fluctuation magnitudes for different wavelengths or different temporal frames; and (ii) a limitation of the number of parameters that have to be estimated from the collection of patches by neglecting temporal and spectral correlations. Several variants of this modeling have been evaluated in experiments,

not reported here, that led to worse detection performances⁴. The effectiveness of introducing a temporal scaling factor in ADI

⁴ In particular, the choice of a temporal versus spectral mean has been studied. A temporal mean was found better suited to model the background of SPHERE-IFS data. It is straightforward to replace the temporal mean by a spectral mean if needed to model the fluctuations of other instruments.

Table 1. Reminder of the main notations.

Not.	Definitions
▷ Constants and related indexes	
N	Number of pixels in a frame
n	Pixel index
L	Number of spectral frames
ℓ	Spectral index
T	Number of temporal frames
t	Temporal index
K	Number of pixels in a patch
▷ Data quantities	
r	Observed intensity
f	Background
h	Off-axis PSF
▷ Estimated quantities	
<i>Background statistics</i>	
$\widehat{\sigma}^2$	Temporo-spectral weight
$\widehat{\mathbf{m}}$	Temporal mean
$\widehat{\mathbf{S}}$	Spatial (empirical) covariance
$\widehat{\mathbf{C}}$	Spatial (shrunk) covariance
$\widehat{\mathbf{N}}$	Cholesky factorization of $\widehat{\mathbf{C}}$: $\widehat{\mathbf{C}}^{-1} = \widehat{\mathbf{N}}\widehat{\mathbf{N}}^\top$
$\widehat{\mathbf{\Sigma}}$	Spectral (shrunk) covariance
$\widehat{\mathbf{L}}$	Cholesky factorization of $\widehat{\mathbf{\Sigma}}$: $\widehat{\mathbf{\Sigma}}^{-1} = \widehat{\mathbf{L}}\widehat{\mathbf{L}}^\top$
<i>Source quantities</i>	
$\widehat{\boldsymbol{\alpha}}$	Vector of source fluxes (spectrum)
$\widehat{\mathbf{v}}$	Standard deviations of the source fluxes
$\widehat{\boldsymbol{\beta}}$	Expected spectral S/N: $\widehat{\beta}_\ell = \alpha_\ell / \sqrt{v_\ell}$
$\widehat{\boldsymbol{\gamma}}$	Prior on the source spectrum
$\widehat{\boldsymbol{\beta}}'$	Expected spectral S/N if $\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\gamma}}: \widehat{\beta}'_\ell = \gamma_\ell / \sqrt{v_\ell}$
\mathbf{w}	Spectral weights
▷ Detection criterions	
S/N_ℓ	Spectral S/N at wavelength ℓ
\mathbf{x}	Vector of spectral S/N: $x_\ell = S/N_\ell$
$\widehat{\mathbf{S}}$	Vector of whitened spectral S/N := $\widehat{\mathbf{L}}^\top \mathbf{x}$
wwS/N	Weighted whitened S/N := $\sum_\ell w_\ell [\widehat{\mathbf{L}}^\top \mathbf{x}]_\ell$
GLRT $_\ell$	Spectral GLR: = S/N_ℓ^2
GLRT $^+$	Combined GLR: = $\sum_\ell [\widehat{\alpha}_\ell]_+^2 / v_\ell$
wGLRT	Whitened GLR: = $\ \widehat{\mathbf{L}}^\top \mathbf{x}\ _2^2$
▷ Positions	
ϕ_0	2D angular loc. of a source in a ref. frame
$\phi_{\ell,t}$	2D angular loc. at wavelength ℓ and time t
$\lfloor \phi_{\ell,t} \rfloor$	Closest on-grid location to $\phi_{\ell,t}$

has been recently demonstrated (Flasseur et al. 2020). Neglecting the spectral correlations of the background may seem a crude approximation. There are indeed some strong correlations but these correlations are difficult to capture at the scale of patches given our limited number of samples. In Sect. 3.2, we describe how to account for spectral correlations at a later stage of the algorithm, with satisfying results.

2.2. Local learning of the parameters

Since a different multivariate Gaussian model is defined for each angular location θ_n , the estimation of the parameters $\mathbf{m}_{n,\ell}$, $\sigma_{n,\ell}^2$ and \mathbf{C}_n can be performed independently on each collection

Algorithm 1: Local background statistics estimation.

Input: $\{\mathbf{r}_{n,1,1}, \dots, \mathbf{r}_{n,L,T}\}$ (stack of LT patches)
 (each patch has K pixels)

Output: $\{\widehat{\mathbf{m}}_{n,\ell}\}_{\ell \in 1:L}$ (mean patches)

Output: $\widehat{\mathbf{C}}_n$ ($K \times K$ spatial covariance)

Output: $\{\widehat{\sigma}_{n,1,1}, \dots, \widehat{\sigma}_{n,L,T}\}$ (scaling factors)

$\widehat{\mathbf{C}}_n \leftarrow \mathbf{I}$ (initialize at identity matrix)

$\forall \ell, \forall t, \widehat{\sigma}_{n,\ell,t}^{(\text{old})} \leftarrow 1$

for $\ell \leftarrow 1$ **to** L **do**

$\widehat{\mathbf{m}}_{n,\ell} \leftarrow \frac{1}{T} \sum_{t=1}^T \mathbf{r}_{n,\ell,t}$ (sample mean)

do

$\widehat{\mathbf{C}}_n \leftarrow \frac{K}{\text{tr}(\widehat{\mathbf{C}}_n)} \widehat{\mathbf{C}}_n$ (normalization)

▷ **Step 1:** Estimate scaling parameters

for $\ell \leftarrow 1$ **to** L **do**

for $t \leftarrow 1$ **to** T **do**

$\widehat{\sigma}_{n,\ell,t}^2 \leftarrow \frac{1}{K} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})^\top \widehat{\mathbf{C}}_n^{-1} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})$

▷ **Step 2:** Update the mean patches

for $\ell \leftarrow 1$ **to** L **do**

$s \leftarrow \sum_{t=1}^T 1/\widehat{\sigma}_{n,\ell,t}^2$ (normalization factor)

$\widehat{\mathbf{m}}_{n,\ell} \leftarrow \frac{1}{s} \sum_{t=1}^T \frac{1}{\widehat{\sigma}_{n,\ell,t}^2} \mathbf{r}_{n,\ell,t}$ (weighted mean)

▷ **Step 3:** Update the spatial covariance

$\widehat{\mathbf{S}}_n \leftarrow \sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1/(TL)}{\widehat{\sigma}_{n,\ell,t}^2} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})(\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})^\top$

$\widetilde{P} \leftarrow (\sum_{\ell,t} 1/\widehat{\sigma}_{n,\ell,t}^2) / (\sum_{\ell,t} 1/\widehat{\sigma}_{n,\ell,t}^4)$ (equivalent number of patches)

$\widehat{\mathbf{C}}_n \leftarrow \text{Alg2}(\widehat{\mathbf{S}}_n, \widetilde{P})$ (shrinkage estimator)

while $\max_{\ell,t} |\widehat{\sigma}_{n,\ell,t} - \widehat{\sigma}_{n,\ell,t}^{(\text{old})}| \geq \epsilon;$

$\{\mathbf{r}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T}$ of 2D patches centered on a given location θ_n . Under our assumptions of negligible temporal and spectral correlations, the co-log-likelihood \mathcal{L}_n of the collection can be written:

$$\mathcal{L}_n = -\log p(\{\mathbf{r}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T} \mid \{\mathbf{m}_{n,\ell}\}_{\ell \in 1:L}, \{\sigma_{n,\ell,t}^2\}_{\ell \in 1:L, t \in 1:T}, \mathbf{C}_n) \quad (1)$$

$$\begin{aligned} \Rightarrow \mathcal{L}_n &= \frac{LTK}{2} \log 2\pi + \sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1}{2} \log \det(\sigma_{n,\ell,t}^2 \mathbf{C}_n) \\ &+ \sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1}{2} (\mathbf{r}_{n,\ell,t} - \mathbf{m}_{n,\ell})^\top (\sigma_{n,\ell,t}^2 \mathbf{C}_n)^{-1} (\mathbf{r}_{n,\ell,t} - \mathbf{m}_{n,\ell}). \end{aligned} \quad (2)$$

We show in Appendix A that the maximum likelihood estimates for the Gaussian parameters are the solution to the following system of non-linear equations:

$$\begin{cases} \widehat{\mathbf{m}}_{n,\ell} = \frac{1}{\sum_{t \in 1:T} 1/\widehat{\sigma}_{n,\ell,t}^2} \cdot \sum_{t \in 1:T} \frac{1}{\widehat{\sigma}_{n,\ell,t}^2} \mathbf{r}_{n,\ell,t} \\ \widehat{\sigma}_{n,\ell,t}^2 = \frac{1}{K} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})^\top \widehat{\mathbf{S}}_n^{-1} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell}) \\ \widehat{\mathbf{S}}_n = \frac{1}{TL} \sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1}{\widehat{\sigma}_{n,\ell,t}^2} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})(\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell})^\top, \end{cases} \quad (3)$$

where the maximum likelihood estimate of the covariance $\widehat{\mathbf{S}}_n$ is not directly used as an estimate of the covariance $\widehat{\mathbf{C}}_n$, but

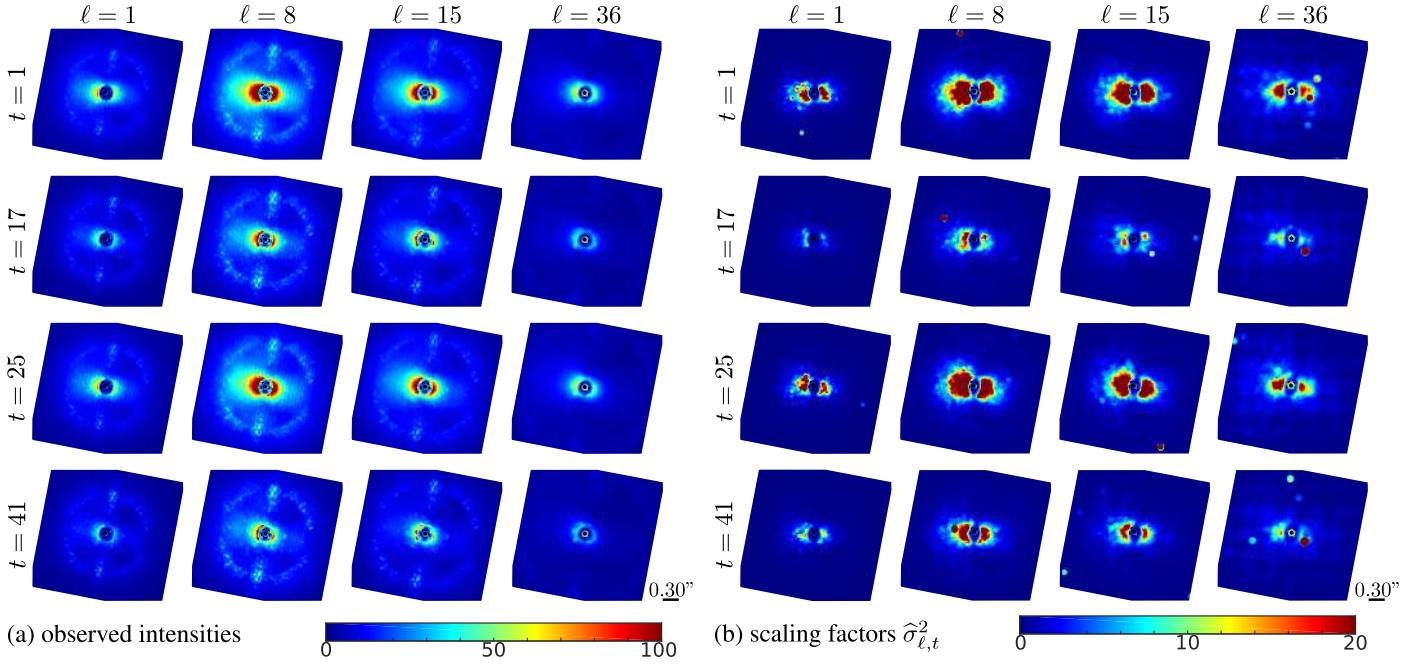


Fig. 2. Accounting for temporal and spectral fluctuations with time and wavelength-specific scaling factors: *a*: observed intensities, for some selected frames (4 wavelengths \times 4 exposures); *b*: corresponding spatial distribution of scaling factors.

Algorithm 2: Shrinkage covariance estimator.

Input: $\widehat{\mathbf{S}}_n$ ($K \times K$ sample covariance matrix)

Input: \widetilde{P} (equivalent number of patches)

Output: $\widehat{\mathbf{C}}_n$ (covariance matrix estimate)

$$\widehat{\rho} \leftarrow \frac{\text{tr}(\widehat{\mathbf{S}}_n) + \text{tr}^2(\widehat{\mathbf{S}}_n) - 2 \sum_{i=1}^K [\widehat{\mathbf{S}}_n]_{ii}^2}{(\widetilde{P}+1) \left(\text{tr}(\widehat{\mathbf{S}}_n) - \sum_{i=1}^K [\widehat{\mathbf{S}}_n]_{ii}^2 \right)} \quad (\text{shrinkage factor})$$

for $i \leftarrow 1$ **to** K **do**

$[\widehat{\mathbf{C}}_n]_{ii} \leftarrow [\widehat{\mathbf{S}}_n]_{ii}$ (copy diagonal)

for $j \leftarrow i + 1$ **to** K **do**

$[\widehat{\mathbf{C}}_n]_{i,j} \leftarrow (1 - \widehat{\rho}) [\widehat{\mathbf{S}}_n]_{i,j}$ (shrink off-diagonal)

$[\widehat{\mathbf{C}}_n]_{j,i} \leftarrow (1 - \widehat{\rho}) [\widehat{\mathbf{S}}_n]_{j,i}$ (shrink off-diagonal)

is replaced by a more reliable estimator with a smaller risk, as described in the following paragraphs.

We solve the System (3) by the method of fixed-point iteration, that is, by alternatively updating each unknown until convergence. This leads to Algorithm 1, where we chose the arbitrary normalization $\text{tr}(\widehat{\mathbf{C}}_n) = K$ for matrix $\widehat{\mathbf{C}}_n$ (some form of normalization is necessary to remove the scaling degeneracy in the product $\widehat{\sigma}_{n,\ell,t}^2 \widehat{\mathbf{C}}_n$).

For locations θ_n outside of the central region of the field of view, some patches $\mathbf{r}_{n,\ell,t}$ fall outside of the measured area for the largest wavelengths. In that case, the sum for the computation of \mathbf{S}_n in the System (3) and in Algorithm 1 is restricted to the wavelengths ℓ for which the patch is measured and the normalization factor $1/TL$ is corrected to match the actual number of terms in the sum. Given the severe reduction in the number of patches actually used close to the borders of the field of view, it is important to regularize the sample covariance to reduce the estimation variance and to prevent obtaining singular or ill-conditioned matrices. As in Flasseur et al. (2018a), we use a shrinkage estimator, implemented according to Algorithm 2.

Because of the weighting by factors $1/\widehat{\sigma}_{n,\ell,t}^2$, some patches have more importance than others and an equivalent number \widetilde{P} of patches is used in the shrinkage formula, step 3 of Algorithm 1, see also Flasseur et al. (2020). The closed form expression of \widetilde{P} is derived in Appendix B. The rationale behind this equivalent number of patches comes from the variance reduction when performing the weighted mean.

Figure 2a depicts the observed intensities in some frames of an ASDI dataset. Fluctuations can be noted both through time and through the wavelengths. In Fig. 2b, maps of the time and wavelength-specific scaling factors $\widehat{\sigma}_{n,\ell,t}^2$ are displayed for 16 pairs (ℓ, t) . At a given location n , large values of this scaling factor compared to other frames t or other wavelengths ℓ indicate that the corresponding patches have a moderate or negligible weight when estimating the mean background and the spatial covariance matrix. In the source detection and characterization steps described in the following sections, patches with comparatively larger scaling factors $\widehat{\sigma}_{n,\ell,t}^2$ also play a minor role. The method is thus robust to the presence of outliers⁵ in the data, see Flasseur et al. (2020). A close inspection of the maps in Fig. 2b reveals the presence of outliers: when an outlier affects a patch, the whole patch is discarded, outliers are thus visible as a disk-shaped area of large $\widehat{\sigma}_{n,\ell,t}^2$ values (corresponding to all spatial locations n that contain the outlier, that is, the disk shape of our patches).

The convergence of Algorithm 1 is illustrated in Fig. 3. Three different locations in the field of view, depicted by a red dot in the insert, are selected: a small angular separation in Fig. 3a, an intermediate separation in 3b and a large separation in 3c. In each case, 1000 different random draws were used as an initialization. The graphs report the normalized distance to the solution found with a constant initialization after a large

⁵ Outliers are artifacts taking the form of unexpected fluctuations. These artifacts can be spatially localized (e.g., defective pixels) or can impact a larger part of the field of view when, e.g., a sudden degradation of the adaptive optics correction occurs.

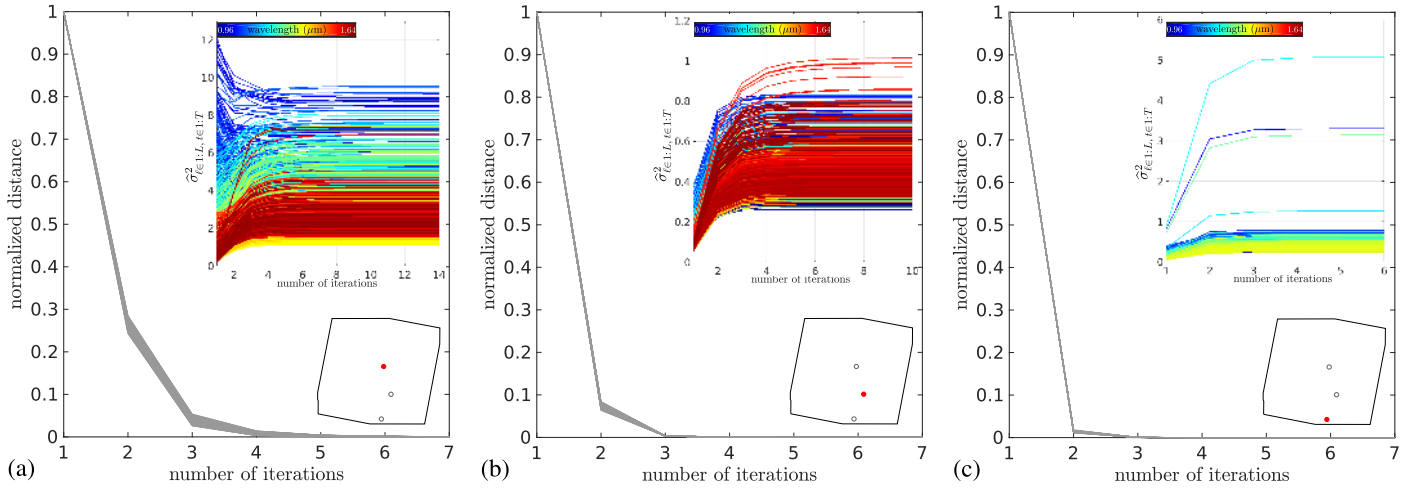


Fig. 3. Convergence of the scaling factors, starting from many random initializations. In the inserts, the location in the field of view is indicated as well as the evolution of the weights until the convergence criterion is reached.

number of iterations. Convergence to the same solution is observed experimentally in all cases. An insert also gives the evolution of each scaling factor $\hat{\sigma}_{n,\ell,t}^2$ with the iterations, until the convergence criterion is reached. A satisfactory convergence is reached in about 10 iterations. At large angular separations, as in 3c, only the shortest wavelengths are available after the speckles are aligned by spectral zooming. The convergence is even faster in this case.

Figure 4 evaluates the statistical accuracy of our modeling of the background on HR 8799 ASDI dataset. The left column gives the values and empirical distributions of the collection of mean-subtracted patches $\{\mathbf{r}_{n,\ell,t} - \mathbf{m}_{n,\ell}\}_{\ell \in 1:L, t \in 1:T}$, at a location n near the coronagraph (rows a and b), at a location farther from the coronagraph (rows c and d), and for all patches from the field of view (row e). Since only the modeling of the background is considered here, all patches around and at the location of the 3 known point-like sources were excluded. Simply removing an average background per wavelength is not satisfactory: values are not distributed according to a Gaussian distribution, there are numerous large deviations. The central column of Fig. 4 gives the intensity values and the empirical distributions when only a spatial whitening is applied, using the same spatial covariance matrix for all frames and all wavelengths. Each mean-subtracted patch intensity of the collection $\{\bar{\mathbf{r}}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T} = \{\mathbf{r}_{n,\ell,t} - \bar{\mathbf{m}}_{n,\ell}\}_{\ell \in 1:L, t \in 1:T}$ is multiplied by the $K \times K$ Cholesky factor $\hat{\mathbf{N}}_n^T$ of the spatial covariance $\hat{\mathbf{C}}_n$ such as $\hat{\mathbf{N}}_n \hat{\mathbf{N}}_n^T = \hat{\mathbf{C}}_n^{-1}$. If the statistical modeling captures accurately the fluctuations of the background, the vectors $\{\hat{\mathbf{N}}_n^T \bar{\mathbf{r}}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T}$ should be distributed according to $\mathcal{N}(\mathbf{0}, \mathbf{I})$ under \mathcal{H}_0 : the linear filter $\hat{\mathbf{N}}_n^T$ whitens the vectors $\{\bar{\mathbf{r}}_{n,\ell,t}\}_{\ell \in 1:L, t \in 1:T}$. The introduction of the whitening step leads to residuals more closely following a standard Gaussian distribution. The right column considers the case of spatial whitening by a covariance matrix scaled by the time and wavelength-specific factors $\hat{\sigma}_{n,\ell,t}$. The empirical distributions follow more closely a standard Gaussian, yet the match is not perfect close to the coronagraph. Accounting for the spectral or temporal correlations would probably further improve the statistical modeling of the background. Such modeling, however, seems difficult to carry out given the limited number of samples and is left to further studies. It is shown in the following sections that the proposed modeling already provides consistent results.

3. Detection maps

The statistical model of the background in ASDI datasets introduced in the previous section is essential to derive the detection and characterization method. Backgrounds at all wavelengths are combined to estimate the parameters of this model. We first describe how this multi-wavelength background model can be applied to produce a detection map at a single wavelength. We then discuss the combination of detection maps at several wavelengths.

3.1. Detection at a single wavelength

Let ϕ_0 be the hypothetical location of a point source in some reference frame. If a point source is present at that location, with a flux α_ℓ in the ℓ th band of the spectrum, then the signal of that source corresponds, at time t and in the n th patch, to $\alpha_\ell \mathbf{h}_{n,\ell}(\phi_{\ell,t})$, with $\mathbf{h}_{n,\ell}(\phi_{\ell,t})$ the zoomed-in off-axis PSF centered at the sub-pixel location $\phi_{\ell,t}$ of the source at the ℓ th wavelength and n th frame. Given the scarcity of sources in the field of view, it is safe to suppose that, within a small patch of a few tens of pixels, only a single source may be present. Detecting a point source at location ϕ_0 then amounts to deciding for one of two hypotheses:

$$\begin{cases} \mathcal{H}_0 : \{\mathbf{r}_{[\phi_{\ell,i}],\ell,t}\}_{i \in 1:T} & = \{\mathbf{f}_{[\phi_{\ell,i}],\ell,t}\}_{i \in 1:T} \\ \text{(background only)} & \\ \mathcal{H}_1 : \{\mathbf{r}_{[\phi_{\ell,i}],\ell,t}\}_{i \in 1:T} & = \alpha_\ell \{\mathbf{h}_{[\phi_{\ell,i}],\ell}(\phi_{\ell,t})\}_{i \in 1:T} \\ \text{(background+source)} & \quad + \{\mathbf{f}_{[\phi_{\ell,i}],\ell,t}\}_{i \in 1:T}, \end{cases} \quad (4)$$

where \mathbf{f} is the notation for patches that contain pure background. The collection of patches considered in this hypothesis test corresponds to all patches that would contain the source if it was present: patches centered at pixel locations $[\phi_{\ell,t}]$ that match the location of the source at time t and wavelength ℓ due to the rotation of the field of view and the zoom applied to align the speckles at all wavelengths, see Fig. 5. Under hypothesis \mathcal{H}_0 , the collection of patches corresponds to pure background: no source is present at location ϕ_0 . Under hypothesis \mathcal{H}_1 , the patches result from the superimposition of an off-axis PSF and of the background.

Under our statistical model of the background given in Eq. (2), the likelihood of each hypothesis can be compared for a

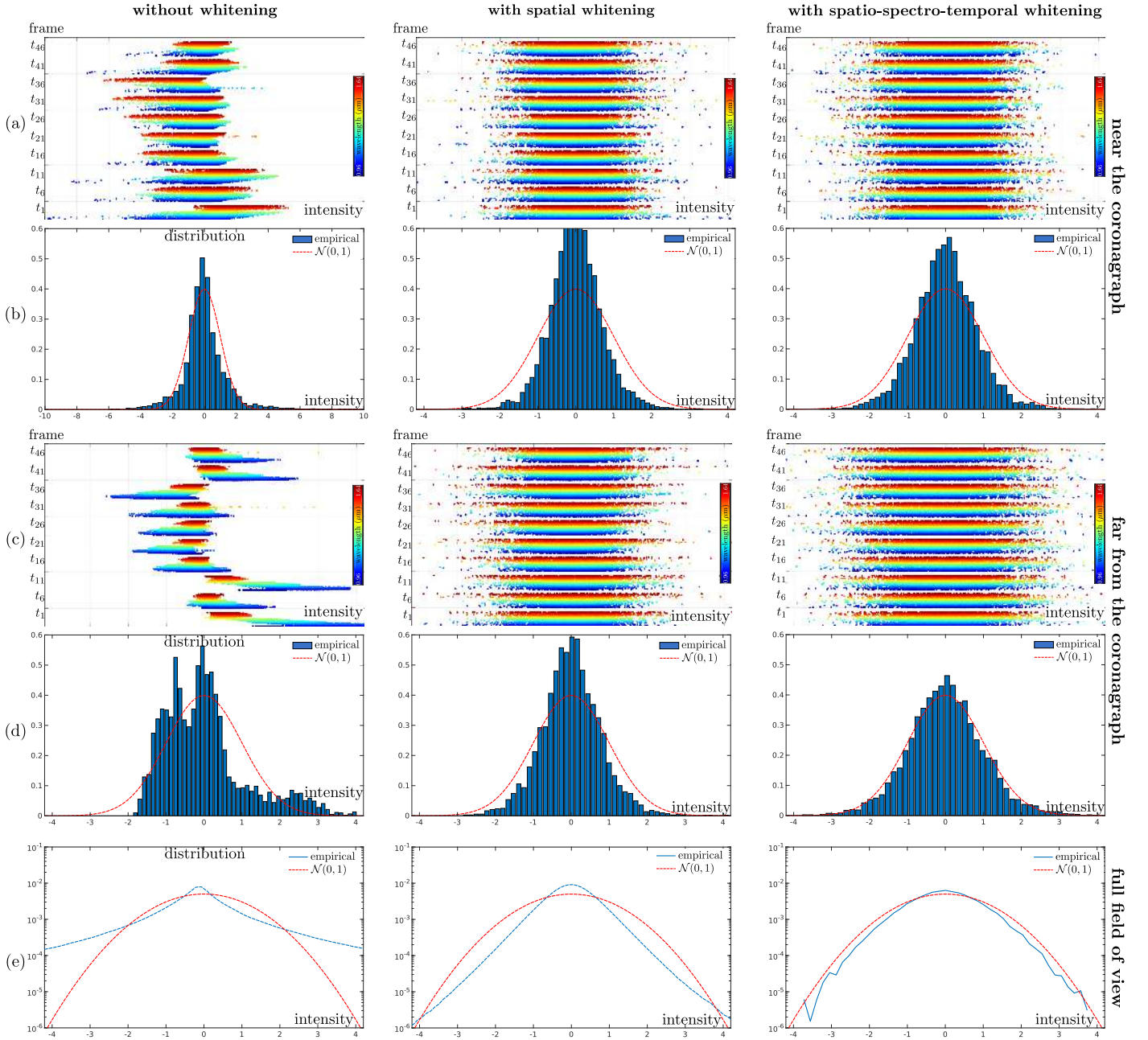


Fig. 4. Distribution of the centered patches: *left*: without whitening; *center*: after spatial whitening; *right*: after spatial whitening and correction by the wavelength and time-specific scaling factors. Rows *a* and *b*: location selected at a small angular separation; rows *c* and *d*: location at a larger angular separation; row *e*: empirical distribution computed over the whole field of view. Patches represented in this figure contain no point-source.

given flux α_ℓ :

$$2 \log \frac{p(\{\mathbf{r}_{[\phi_{\ell,i}],\ell,t}\}|\mathcal{H}_1, \alpha_\ell)}{p(\{\mathbf{r}_{[\phi_{\ell,i}],\ell,t}\}|\mathcal{H}_0)} = \alpha_\ell \sum_{t=1}^T \mathbf{u}_{[\phi_{\ell,i}],\ell,t}^\top \widehat{\mathbf{C}}_{[\phi_{\ell,i}]}^{-1} \mathbf{h}_{[\phi_{\ell,i}],\ell}(\phi_{\ell,t}) \quad (5)$$

$$\text{with } \mathbf{u}_{n,\ell,t} = \frac{1}{\hat{\sigma}_{n,\ell,t}^2} (\mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell} - \alpha_\ell \mathbf{h}_{n,\ell}(\phi_{\ell,t})).$$

Since the flux α_ℓ is generally not known beforehand, it has to be estimated from the data. The maximum likelihood estimator,

under our model of the background, is:

$$\widehat{\alpha}_\ell = \frac{\sum_{t=1}^T b_{\ell,t} / \sqrt{\hat{\sigma}_{[\phi_{\ell,i}],\ell,t}^2}}{\sum_{t=1}^T a_{\ell,t} / \sqrt{\hat{\sigma}_{[\phi_{\ell,i}],\ell,t}^2}}, \quad (6)$$

with

$$a_{\ell,t} = \mathbf{h}_{[\phi_{\ell,i}],\ell}(\phi_{\ell,t})^\top \widehat{\mathbf{C}}_{[\phi_{\ell,i}]}^{-1} \mathbf{h}_{[\phi_{\ell,i}],\ell}(\phi_{\ell,t}) \quad (7)$$

and

$$b_{\ell,t} = \mathbf{h}_{[\phi_{\ell,i}],\ell}(\phi_{\ell,t})^\top \widehat{\mathbf{C}}_{[\phi_{\ell,i}]}^{-1} (\mathbf{r}_{[\phi_{\ell,i}],\ell,t} - \widehat{\mathbf{m}}_{[\phi_{\ell,i}],\ell}). \quad (8)$$

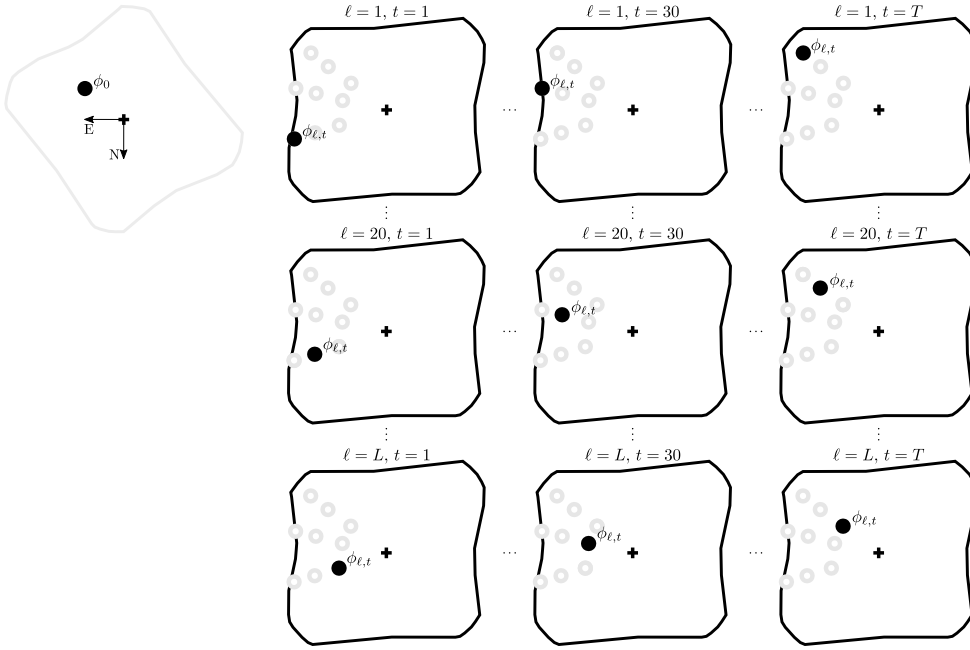


Fig. 5. Evolution of the 2D location $\phi_{\ell,t}$ of a source in a speckle-aligned ASDI dataset: ϕ_0 defines the 2D angular location of a point source in a reference frame; the apparent location of the point source in the t th observation and the ℓ th spectral band is indicated by a black disk; the apparent locations of a point source at other observation times and spectral bands are indicated by gray circles. The location $\phi_{\ell,t}$ describes a radial motion with the wavelength and a rotation about the optical axis over time.

Substituting α_ℓ with its estimate $\widehat{\alpha}_\ell$ in Eq. (5) leads to the generalized likelihood ratio test:

$$\text{GLRT}_\ell : \frac{\left(\sum_{t=1}^T b_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2 \right)^2}{\sum_{t=1}^T a_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta. \quad (9)$$

Only positive flux estimates are physically meaningful for point sources. The test can then be improved by discarding locations leading to negative flux estimates $\widehat{\alpha}_\ell$ (see also Flasseur et al. 2018a):

$$\text{S/N}_\ell : \frac{\sum_{t=1}^T b_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2}{\sqrt{\sum_{t=1}^T a_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \tau, \quad (10)$$

which matches GLRT_ℓ when $\tau = \sqrt{\eta}$ and $\widehat{\alpha}_\ell \geq 0$. As noted by Mugnier et al. (2009), the ratio in (10) corresponds to a signal-to-noise ratio. It is obtained by linearly transforming the data and accounts for the local, time and wavelength-specific covariance of the background. The variance of the estimator $\widehat{\alpha}_\ell$, hereafter noted v_ℓ , is:

$$\text{Var}[\widehat{\alpha}_\ell] = \left(\sum_{t=1}^T a_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2 \right)^{-1}. \quad (11)$$

The signal-to-noise ratio of the flux estimate (S/N_ℓ) therefore corresponds to the ratio $\widehat{\alpha}_\ell / \sqrt{v_\ell}$ and is distributed as a standard normal variate under \mathcal{H}_0 (Mugnier et al. 2009).

Figure 6 compares the detection maps S/N_ℓ computed with Eq. (10) and detection maps obtained by PACO on ADI subsets (i.e., by processing the data one wavelength at a time). The major difference between the two approaches is that PACO ASDI combines information from all wavelengths to learn the background model (more specifically, the covariance matrices). Therefore, a more accurate model is obtained and point-sources are better

discriminated against the background: the signal-to-noise ratio of the sources is improved at all wavelengths while the fluctuations in the absence of sources are comparable.

Beyond the improvement of the detection map at a given wavelength, PACO ASDI also benefits from combining detection maps at different wavelengths to better detect sources, as described in the next section.

3.2. Combining multiple detection maps

3.2.1. Combination assuming spectral independence

The detection of point sources can be largely improved by combining information from different wavelengths. The most straightforward approach consists of the extension of the hypothesis test in Eq. (4) in order to include the patches at all times and all wavelengths (all locations depicted in Fig. 5 rather than a single row). Under the assumption that spectral channels are independent, the likelihood can be factored as a product over all channels and fluxes α_ℓ can be separately estimated. Deciding for the presence of a source at location ϕ_0 based on the generalized likelihood ratio amounts to:

$$\text{GLRT} : \sum_{\ell=1}^L \frac{\left(\sum_{t=1}^T b_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2 \right)^2}{\sum_{t=1}^T a_{\ell,t} / \widehat{\sigma}_{[\phi_{\ell,t}, \ell, t]}^2} = \sum_{\ell=1}^L \frac{\widehat{\alpha}_\ell^2}{v_\ell} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta. \quad (12)$$

To account for the imposed non-negativity of source fluxes, the test can be slightly modified, see Flasseur et al. (2018b):

$$\text{GLRT}^+ : \sum_{\ell=1}^L \frac{[\widehat{\alpha}_\ell]_+^2}{v_\ell} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} \eta, \quad (13)$$

where $[x]_+ = \max(x, 0)$ is the positive part of x .

Under \mathcal{H}_0 and the assumption that S/N_ℓ values are independent from one another (a hypothesis that will be rejected in the following paragraphs), the distribution of GLRT^+ is given by (see the derivation in Appendix C):

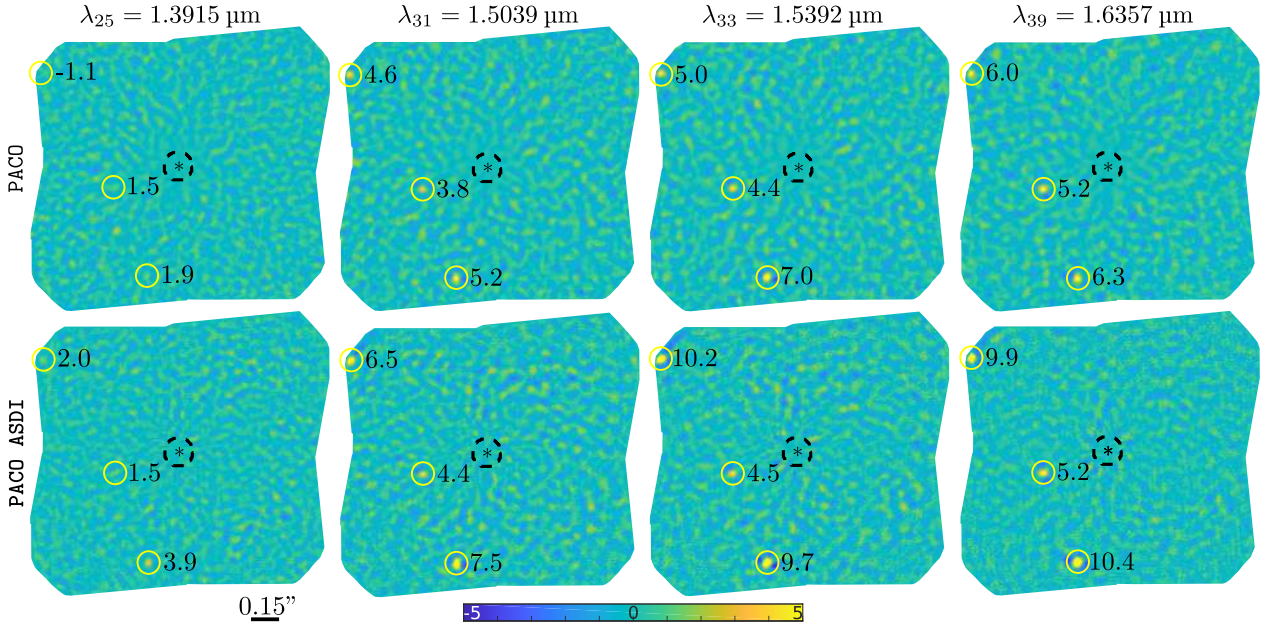


Fig. 6. PACO versus PACO ASDI: impact of learning background structures at a single wavelength (PACO, *first row*) or jointly from all the wavelengths (PACO ASDI, *second row*). Single-wavelength detection maps S/N_ℓ are shown for PACO ASDI. The differences between the 2 lines are related to the estimation of the covariance matrices \mathbf{C} and the scaling parameters σ . The combination of those maps leads to a single detection map, not shown here, with improved sensitivity (see text and Fig. 8).

$$p(\text{GLRT}^+ | \mathcal{H}_0) = \frac{1}{2^L} \delta_0(\text{GLRT}^+) + \sum_{\ell=0}^{L-1} \frac{L!}{2^L \ell! (L-\ell)!} \chi_{L-\ell}^2(\text{GLRT}^+), \quad (14)$$

where δ_0 is a Dirac mass centered in 0 and $\chi_{L-\ell}^2$ is a Chi-square distribution with $L - \ell$ degrees of freedom.

Figure 7a displays the GLRT^+ obtained with Eq. (13) on an ASDI dataset of HR 8799 obtained with SPHERE-IFS. Three point-sources can be detected in this dataset, at the locations marked c, d, and e. To perform an automated detection, it is necessary to set a threshold corresponding to a fixed probability of false alarm. The empirical distribution of GLRT^+ values, excluding the three regions that contain the point-sources, is shown at the right of the detection map of Fig. 7a. This empirical distribution is compared to the theoretical distribution of GLRT^+ under \mathcal{H}_0 , drawn in dashed line. A strong mismatch is observed. Due to this discrepancy, it is not possible to derive a detection threshold from the model of Eq. (14). The empirical distribution is shifted to the left as if the number of wavelengths was smaller than L . An effective number of wavelengths could be derived by fitting the parameter L in Eq. (14) to the empirical distribution. This effective number of wavelengths would account for the correlations of S/N values between adjacent wavelengths. To limit the number of false alarms, the detection threshold is typically set in order to reach probabilities as low as 10^{-7} . A mismatching of the right tail of the distribution may have a large impact on the value of the threshold. Rather than estimating an effective number of wavelengths to adjust the model (14), we model the distribution of the S/N_ℓ values by accounting for the wavelength correlations.

3.2.2. Accounting for spectral correlations

Signal-to-noise ratio values S/N_ℓ defined in Eq. (10) are Gaussian distributed. However, they are not mutually independent

because the background patches, in a given frame, are very similar for adjacent wavelengths. Before combining detection maps, it is necessary to learn the spectral correlations between the maps S/N_ℓ . This learning is performed locally, on a small region of the maps S/N_ℓ . Since there might be a point source within the region, it is necessary to use a robust estimator $\widehat{\Sigma}$ of the spectral covariance (otherwise, spectral whitening would suppress the source). There are several robust estimators for the covariance, see for example the review in Hubert et al. (2008). The minimum covariance determinant (MCD) method identifies a subset of observations of a fixed size whose covariance matrix has the lowest determinant. To identify this subset quickly, we use the algorithm FAST-MCD introduced in Rousseeuw & Driessen (1999). The region over which an estimate $\widehat{\Sigma}$ is computed must be large enough to guarantee that the area of large S/N_ℓ values corresponding to a point source be considered as an outlier. Appendix D discusses how to set the size of this region and how $\widehat{\Sigma}$ can be improved, in a second step, by masking.

The vector \mathbf{x} of S/N_ℓ values is a sufficient statistic for the fluxes $\widehat{\alpha}_\ell$ of a point source. The detection of a point source can thus be defined directly on the vector \mathbf{x} :

$$\begin{cases} \mathcal{H}_0 : \mathbf{x} = \boldsymbol{\epsilon} \\ \text{(no source)} \\ \mathcal{H}_1 : \mathbf{x} = \boldsymbol{\epsilon} + \boldsymbol{\beta}, \\ \text{(a point source is present)} \end{cases} \quad (15)$$

where $\boldsymbol{\beta} \in \mathbb{R}^L$ is the vector of expected S/N values at each of the L wavelengths: $\beta_\ell = \alpha_\ell / \sqrt{v_\ell}$, and $\boldsymbol{\epsilon}$ is a random vector accounting for the fluctuations of S/N_ℓ values. According to our model of spectral correlations, $\boldsymbol{\epsilon}$ follows the Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$. Replacing the unknown vector $\boldsymbol{\beta}$ by its maximum likelihood estimate $\widehat{\boldsymbol{\beta}} = \mathbf{x}$ gives an approximation of the likelihood of \mathcal{H}_1 and leads to the following GLR test:

$$\text{wGLRT} : \quad \|\widehat{\mathbb{L}}^\top \mathbf{x}\|_2^2 = \mathbf{x}^\top \widehat{\boldsymbol{\Sigma}}^{-1} \mathbf{x} \underset{\mathcal{H}_0}{\underset{\mathcal{H}_1}{\geq}} \eta, \quad (16)$$

where $\widehat{\mathbb{L}}$ is the $L \times L$ whitening matrix obtained by Cholesky factorization, i.e., such that $\widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top = \widehat{\Sigma}^{-1}$. Similarly to the spatial whitening introduced in Sect. 2.2, the matrix $\widehat{\mathbb{L}}$ is such that $\widehat{\mathbb{L}}^\top \mathbf{x}$ be distributed according to $\mathcal{N}(\mathbf{0}, \mathbf{I})$ under \mathcal{H}_0 : it whitens vectors of S/N_ℓ values, hence the name of the corresponding detection criterion.

In the absence of spectral correlations (i.e., $\Sigma = \mathbf{I}$), $\mathbb{L} = \mathbf{I}$ and the wGLRT equals the GLRT defined in Eq. (12). Under \mathcal{H}_0 , the wGLRT follows a χ^2 distribution with L degrees of freedom.

Figure 7b displays the wGLRT detection map and, by masking out the three sources, the distribution of wGLRT under \mathcal{H}_0 . The comparison with GLRT⁺ shows that the spectral whitening reduces artifacts in the absence of sources (periodic structures observed in Fig. 7a are no longer visible in Fig. 7b). The empirical distribution of wGLRT is much closer to the expected distribution, however, the match is not perfect, which motivates considering another approach to the combination of detection maps.

3.2.3. Improving the detection based on a prior spectrum model

If a coarse model of the spectrum of the point source under study is available prior to the detection, this model can be used to improve the detection performance by giving more weight to spectral bands where larger values are expected. Let $\boldsymbol{\gamma}$ be the spectrum of a point source. Given that spectrum, the hypothesis test (15) takes the simplified form:

$$\left\{ \begin{array}{ll} \mathcal{H}_0 : & \mathbf{x} = \boldsymbol{\epsilon} \\ \text{(no source)} & \\ \mathcal{H}_1 : & \mathbf{x} = \boldsymbol{\epsilon} + \alpha^{\text{int}} \cdot \underbrace{\begin{pmatrix} | \\ \gamma_\ell / \sqrt{v_\ell} \\ | \end{pmatrix}}_{\boldsymbol{\beta}'}, \\ \text{(a point source is present)} & \end{array} \right. \quad (17)$$

where α^{int} is the spectrally integrated flux, that is, the flux such that $\alpha_\ell = \alpha^{\text{int}} \gamma_\ell$, for all ℓ . In contrast to the hypothesis test (15), this new test requires estimating a single scalar parameter: α^{int} . The maximum likelihood estimator for the integrated flux α^{int} is:

$$\widehat{\alpha}^{\text{int}} = \frac{\mathbf{x}^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}{\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}, \quad (18)$$

where $\boldsymbol{\beta}'$ is the vector of \mathbb{R}^L whose ℓ th element is equal to $\gamma_\ell / \sqrt{v_\ell}$ (the expected S/N_ℓ value if α_ℓ was equal to γ_ℓ , i.e. $\alpha^{\text{int}} = 1$). The variance of the estimator $\widehat{\alpha}^{\text{int}}$ is $(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{-1}$. By substituting α^{int} with its estimate $\widehat{\alpha}^{\text{int}}$, another GLRT is obtained:

$$\text{wwGLRT} : \quad 2 \log \frac{p(\mathbf{x}|\mathcal{H}_1, \widehat{\alpha}^{\text{int}})}{p(\mathbf{x}|\mathcal{H}_0)} = \frac{(\mathbf{x}^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^2}{\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}. \quad (19)$$

Since only positive integrated fluxes α^{int} make sense, vectors \mathbf{x} such that $\mathbf{x}^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}' < 0$ can be discarded. The square root of the GLRT then leads to the test:

$$\text{wwS/N} : \quad \frac{\mathbf{x}^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}{\sqrt{\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}} = \sum_{\ell=1}^L w_\ell \cdot \left[\widehat{\mathbb{L}}^\top \mathbf{x} \right]_\ell \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \quad (20)$$

which takes the form of a linear combination of the *whitened* vector of S/N_ℓ values, with weights w_ℓ defined by $w_\ell = \left[\widehat{\mathbb{L}}^\top \boldsymbol{\beta}' \right]_\ell / \sqrt{\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'}$, hence the name wwS/N. Like in our previous derivation of S/N_ℓ , wwS/N can be interpreted as a signal to noise ratio: $\text{wwS/N} = \widehat{\alpha}^{\text{int}} / \sqrt{\text{Var}[\widehat{\alpha}^{\text{int}}]}$.

Interestingly, wwS/N corresponds to the optimal linear combination of the S/N_ℓ values, in the sense that the probability of detection is maximized for all false alarms rates, i.e. a matched filter, see Appendix E.

Comparison between wGLRT and wwS/N. Both wGLRT and wwS/N encompass a whitening of the spectral correlations, but wwS/N also includes an additional spectral weighting strategy based on a prior spectrum model. In order to select between these two different detection criteria, several aspects must be considered: (i) does the criterion follow the expected distribution under \mathcal{H}_0 , that is, can detection thresholds be set for prescribed false alarms rates? (ii) How large is the gain obtained when the spectrum of the source is available? (iii) How does wwS/N degrade if the assumed spectrum differs from the true spectrum?

Under \mathcal{H}_0 , wGLRT is expected to follow a χ^2 distribution with L degrees of freedom, while wwS/N is expected to follow a standard Gaussian distribution. The analysis of Fig. 7b led to the conclusion that the fit with a χ^2 distribution with L degrees of freedom was not accurate. Figure 8 illustrates the empirical distribution of wwS/N for several prior spectra. For all the spectra considered, the empirical distribution is in very good fit with a centered standard Gaussian. A possible explanation for this better fit of wwS/N with the theoretical distribution under \mathcal{H}_0 is that the weights w_ℓ give more importance to spectral channels of good quality (i.e., with a low variance v_ℓ) and that these channels follow more closely our Gaussian model. The good fit of wwS/N under \mathcal{H}_0 with the expected distribution makes it possible to reliably set detection thresholds for a prescribed false alarms rate.

In order to assess the gain in detection performance brought by the prior knowledge of the spectrum of the source, we compare the contrasts achievable at a 5σ false alarms rate. We derive the theoretical contrast values based on our statistical modeling even if, in practice, a deviation is observed between the theoretical distribution and the empirical distribution of wGLRT. Under \mathcal{H}_0 , wGLRT is expected to follow a χ^2 distribution with L degrees of freedom. The probability of false alarms is thus: $\text{PFA} = P(\chi_L^2 > \eta)$. The probability of detection of a source of flux $\alpha^{\text{int}}\boldsymbol{\gamma}$, $\text{PD} = P(\text{wGLRT} > \eta|\mathcal{H}_1)$, corresponds to the probability that a noncentral χ^2 distribution with L degrees of freedom and noncentrality parameter $(\alpha^{\text{int}})^2 \boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}'$ exceeds the detection threshold η . This probability corresponds to $Q_{L/2}(\alpha^{\text{int}}(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{1/2}, \sqrt{\eta})$, where $Q_M(a, b)$ is Marcum Q-function (Simon 2007). Hence, the theoretical 5σ contrast reached by wGLRT can be computed by first solving the equation $\frac{1}{\Gamma(L/2)} \gamma(L/2, \eta/2) = \Phi(5)$ for η (where γ is here the lower incomplete gamma function and Φ is the cumulative distribution function of the standard normal distribution), and then solving $Q_{L/2}(\alpha^{\text{int}}(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{1/2}, \sqrt{\eta}) = 1/2$ for α^{int} . For example, when $L = 39$ (as is the case of SPHERE-IFS), we find $\eta \approx 100$ and $\alpha^{\text{int}} \approx 7.87(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{-1/2}$.

Since the expectation $\mathbb{E}[\text{wwS/N}|\mathcal{H}_1]$ is equal to $\alpha^{\text{int}}(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{1/2}$, the 5σ contrast reached by wwS/N is readily obtained: $\alpha^{\text{int}} = 5(\boldsymbol{\beta}'^\top \widehat{\mathbb{L}}\widehat{\mathbb{L}}^\top \boldsymbol{\beta}')^{-1/2}$. Including the prior knowledge of the source spectrum, therefore, improves the contrast by a factor 1.57 (wwS/N reaches a theoretical contrast that is 1.57 times better than wGLRT).

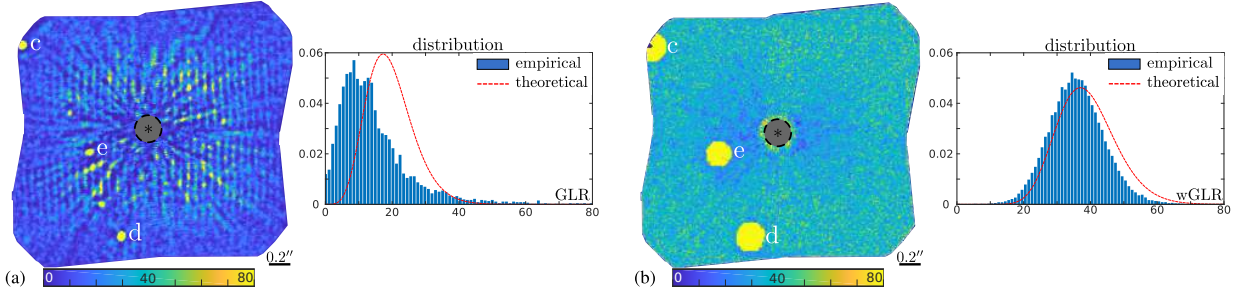


Fig. 7. Combined detection maps computed on HR 8799 ASDI dataset: *a*: GLRT⁺ criterion and its distribution in the absence of sources; *b*: wGLRT criterion, including a spectral whitening operation, and its distribution. The three sources are excluded for the computation of the empirical distributions.

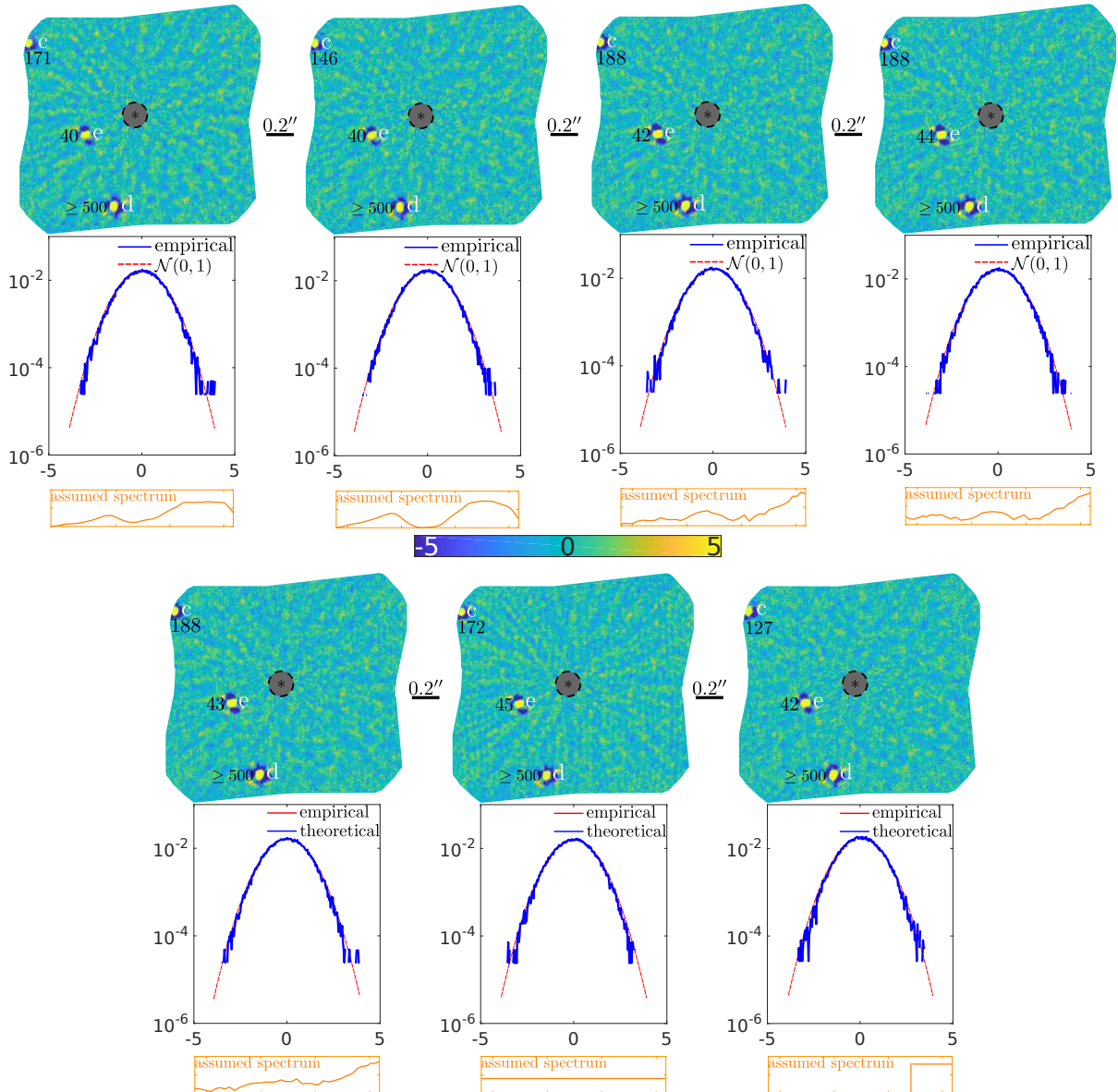


Fig. 8. Combined detection map with spectrum priors: in the absence of sources the empirical distribution matches very closely a Gaussian distribution (red parabola in the log-scale representations).

Rather than expressing the contrast in terms of the value of the integrated flux α^{int} required in order to achieve the detection, it can also be expressed as the flux of the source, at a given wavelength, so that the detection using jointly all wavelengths

is possible. This wavelength-specific contrast corresponds to the values $\alpha_\ell = \alpha^{\text{int}} \gamma_\ell$. For example, the achievable contrast using only a single detection map S/N_ℓ is $5\sqrt{v_\ell}$. When the multi-wavelength criterion wwS/N is applied, if $\hat{\mathbf{L}} = \mathbf{I}$ and if the prior

actually matches the true spectrum of the source, then the achievable contrast corresponds to a flux $5\gamma_{\ell}/\sqrt{\sum_{\ell'}\gamma_{\ell'}^2/v_{\ell}}$ in the ℓ th channel. Therefore, with respect to the single-wavelength map, the contrast is improved by a factor:

$$\sqrt{1 + \sum_{\ell' \neq \ell} \frac{(\gamma_{\ell'}/\gamma_{\ell})^2}{v_{\ell'}/v_{\ell}}}, \quad (21)$$

which depends on the spectrum γ_{ℓ} of the source and on the variance v_{ℓ} of the source flux. This factor is strictly greater than one (i.e., the contrast is strictly improved) provided that there is at least one wavelength ℓ' , different from ℓ , such that the spectrum is non zero ($\gamma_{\ell'} \neq 0$) and the variance is finite ($v_{\ell'} < \infty$). Obviously, if these two conditions are not met, either the source emits no light at the additional wavelengths or no meaningful measurement is available so that the performance cannot be improved compared to when using a single wavelength detection map. In all other cases, there is a gain, that is, the combined detection map leads to a better sensitivity than even the detection map S/N_{ℓ} at the wavelength providing the best contrast. In particular, if the spectrum is flat ($\forall \ell, \gamma_{\ell} = 1/L$) and the variances v_{ℓ} are all equal, the contrast is improved by a factor \sqrt{L} , which is to be expected when combining L measurements of identical statistical weight. This factor should, however, be considered as an upper bound that cannot be reached in practice for several reasons: (i) the whitening matrix $\widehat{\mathbb{L}}$ differs from the identity because of the correlations between channels, the effective number of (independent) channels is, in fact, smaller than L , (ii) the estimation of matrix $\widehat{\mathbb{L}}$ is performed in two steps and relies, in the second step, on a thresholding strategy that requires a detection to prevent the attenuation of point-like sources, (iii) neither the spectrum nor the variance are constant with respect to the wavelength, (iv) the true spectrum of the source may differ from the prior spectrum used in wwS/N.

Impact of a mismatch between the true and assumed spectrum in wwS/N. Finally, the impact of a mismatch between the assumed spectrum in wwS/N and the true spectrum of the source needs to be assessed. This impact can be evaluated by comparing the contrast that is reached when the actual spectrum is used with respect to the contrast when an incorrect prior spectrum is used in wwS/N. Let γ_{\star} be the true spectrum and β'_{\star} the vector such that $\beta'_{\star \ell} = \gamma_{\star \ell} / \sqrt{v_{\ell}}$ for all ℓ . The achievable contrast under the true spectrum prior is:

$$5(\beta'_{\star} \widehat{\mathbb{L}} \beta'_{\star})^{-1/2}. \quad (22)$$

Under the incorrect prior γ , wwS/N is distributed, under \mathcal{H}_1 , according to the Gaussian $\mathcal{N}(\alpha^{\text{int}}(\beta'_{\star} \widehat{\mathbb{L}} \beta')(\beta'_{\star} \widehat{\mathbb{L}} \beta')^{-1/2}, 1)$. The contrast that is achieved is thus equal to $5(\beta'_{\star} \widehat{\mathbb{L}} \beta')^{1/2}/(\beta'_{\star} \widehat{\mathbb{L}} \beta')$. With respect to the ideal case where the true spectrum γ_{\star} is used as a prior, the achievable contrast is degraded by a factor:

$$\frac{\sqrt{\beta'_{\star} \widehat{\mathbb{L}} \beta'} \sqrt{\beta'_{\star} \widehat{\mathbb{L}} \beta'_{\star}}}{\beta'_{\star} \widehat{\mathbb{L}} \beta'}, \quad (23)$$

which corresponds to the inverse of the normalized correlation between the whitened true spectrum and the whitened assumed spectrum.

Table 2 reports the factors by which the achievable contrast is degraded when the prior spectrum differs from the true spectrum. Due to the symmetry in Eq. (23), the role of the prior spectrum and of the true spectrum can be interchanged. These factors

Table 2. Degradation of the achievable contrast when the prior spectrum differs from the true spectrum.

true SED:				
prior SED:				
	1.07 ± 0.02	1.11 ± 0.03	1.005 ± 0.001	1.48 ± 0.14
		1.015 ± 0.004	1.04 ± 0.01	1.22 ± 0.06
			1.08 ± 0.02	1.12 ± 0.03
				1.4 ± 0.1

Notes. Contrast ratio with respect to a detection with the true spectrum, closer to 1 is better.

have been computed for 1681 whitening filters computed on a SPHERE-IFS dataset around HR 8799. The mean factor and its standard deviation are reported in the table. When the true spectrum and the prior spectrum are very close (third true spectrum and first prior, in the table) the contrast degradation is negligible (the factor is not significantly greater than 1). Even when the spectrum differs significantly (last true spectrum and first or last prior), the contrast degradation remains modest (at most a factor 1.48 in this case) and smaller than that observed when replacing wwS/N by wGLRT (a factor 1.57 was predicted in the previous paragraph).

Selected detection criterion. These comparisons between wGLRT and wwS/N lead to a clear conclusion: wwS/N is to be preferred since (i) its distribution under \mathcal{H}_0 is more accurately modeled (so that detection thresholds can be automatically set to reach prescribed false alarms rates), (ii) the detection performance is higher when the spectrum of the source is known in advance, (iii) even if the prior spectrum used in wwS/N differs significantly from the true spectrum of the source, the detection performance of wwS/N is higher than that of wGLRT.

Other algorithms for exoplanet detection in ASDI datasets can produce a detection map (signal-to-noise ratio) per wavelength. We show in Appendix F that our strategy for combining multiple detection maps is also beneficial to those algorithms.

4. Source characterization

So far, we introduced two modelings of the data: (i) the background model introduced in Sect. 2, accounting for spatial covariances as well as wavelength-specific and time-specific scaling factors, and (ii) the model of the spectral covariances of vectors \mathbf{x} of S/N_{ℓ} values. The second model, based on the intermediate detection maps S/N_{ℓ} , includes both the patch covariances (through the computation of S/N_{ℓ} values) and the spectral covariances. Rather than performing the astrometric and photometric characterizations of a detected point source based on the co-log-likelihood \mathcal{L}_n introduced in Eq. (2), which does not account for spectral correlations, we define the co-log-likelihood \mathcal{C} on the vectors of S/N_{ℓ} values:

$$\begin{aligned} \mathcal{C}(\phi_0, \alpha) &= -\log p(\mathbf{x} | \phi_0, \alpha, \{v_{\ell}\}_{\ell \in 1:L}, \mathbb{L}) \\ &= \frac{1}{2} \left\| \mathbb{L}^T \left(\mathbf{x} - \frac{\alpha_{\ell}}{\sqrt{v_{\ell}}} \right) \right\|_2^2 + \text{const.}, \end{aligned} \quad (24)$$

where the vector \mathbf{x} of S/N_{ℓ} values is extracted at the integer location $[\phi_0]$ of the field of view, and variance values v_{ℓ} depend on the level of background fluctuations in the patches extracted from

spectral channel ℓ . The constant term depends only on L and on the determinant of the whitening matrix \mathbb{L} .

Similarly to PACO algorithm (Flasseur et al. 2018a), when characterizing a point source found in the detection map, the background statistics are re-estimated jointly with the determination of the subpixel location and flux of the source. This prevents any bias that may occur due to self-subtraction (computation of the mean patches $\widehat{\mathbf{m}}_{n,\ell}$ without accounting for the presence of the source). An alternating estimation strategy is carried out by iteratively applying the following steps, see also Fig. 1: (i) Algorithm 1 is applied to the residual patches $\{\mathbf{r}_{[\phi_{\ell,t}],\ell,t} - \alpha_{\ell} \mathbf{h}_{[\phi_{\ell,t}],\ell}(\phi_{\ell,t})\}_{\ell \in 1:L, t \in 1:T}$, with α initially set to $\mathbf{0}$, to learn the local background statistics, (ii) S/N $_{\ell}$ values x_{ℓ} and variances v_{ℓ} are computed for each wavelength, (iii) the spectral covariance Σ under \mathcal{H}_0 is estimated based on the vectors of values $x_{\ell} - \alpha_{\ell} / \sqrt{v_{\ell}}$ in a local area centered at $[\phi_0]$; the whitening matrix \mathbb{L} is then derived by Cholesky factorization of the inverse of Σ , (iv) the subpixel location ϕ_0 and the flux values α_{ℓ} are estimated, then all the steps are repeated to improve the background modeling and progressively separate source and background.

The last step, corresponding to the astrometric and photometric estimations, is detailed in the following two paragraphs.

4.1. Astrometric estimation

The estimation of the location ϕ_0 , with subpixel accuracy, can be performed by maximizing one of the combined-wavelengths detection criteria over a grid. In practice, we maximize wwS/N over a refined subpixel grid of locations ϕ_0 with the current flux estimates as a prior spectrum: $\gamma = \alpha / (\sum_{\ell} \alpha_{\ell})$. This corresponds to jointly maximizing $\mathcal{C}(\phi_0, \mathbf{0}) - \mathcal{C}(\phi_0, \alpha^{\text{int}} \gamma)$ with respect to the location ϕ_0 and the integrated source flux α^{int} .

Beyond the unbiased estimation of the astrometry and the photometry, it is critical to characterize the variance of these two quantities. In this context, the variances and covariances are predicted at each location of the field of view through the so-called Cramér–Rao lower bounds (CRLBs) of the vector \mathbf{p} of parameters (the 2D angular location ϕ_0 and the spectrum α) that characterizes a point source. The CRLB is a good estimate of the covariance of the maximum likelihood estimator when the number of samples is large enough (Kendall et al. 1948). We follow the approach of Flasseur et al. (2018a, 2020) for ADI processing with the PACO algorithm: the Fisher information matrix \mathbf{I}^{F} on the vector \mathbf{p} is given by:

$$\begin{aligned} [\mathbf{I}_{[\phi_0]}^{\text{F}}(\mathbf{p})]_{i,j} &= \sum_{\ell=1}^L \sum_{t=1}^T \frac{1}{\hat{\sigma}_{[\phi_0],\ell,t}^2} \frac{\partial [\alpha_{\ell} \mathbf{h}_{[\phi_0],\ell}(\phi_0)]^{\text{T}}}{\partial p_i} \\ &\cdot \widehat{\mathbf{C}}_{[\phi_0]}^{-1} \frac{\partial [\alpha_{\ell} \mathbf{h}_{[\phi_0],\ell}(\phi_0)]}{\partial p_j}, \end{aligned} \quad (25)$$

where the first two components of \mathbf{p} represent the two components of the 2D location ϕ_0 . The product $\alpha_{\ell} \mathbf{h}_{[\phi_0],\ell}$ models the signal of a point source of flux α_{ℓ} in the ℓ th spectral channel. As in Flasseur et al. (2018a, 2020), we use a continuous model of the off-axis PSF (for instance, an isotropic Gaussian) to simplify the computation of the spatial derivatives. The standard deviations δ_i for each of the parameters are obtained from the diagonal of the inverse of Fisher information matrix:

$$[\delta(\mathbf{p})]_i = \sqrt{[\mathbf{I}^{\text{F}}(\mathbf{p})^{-1}]_{i,i}}. \quad (26)$$

4.2. Estimation of the source spectrum

At a given source location ϕ_0 , estimating the vector of source fluxes α by minimizing \mathcal{C} leads to the following maximum likelihood estimates:

$$\widehat{\alpha} = \mathbf{V}^{-1} \mathbf{x}, \quad (27)$$

where \mathbf{V} is a diagonal matrix with diagonal entry $[\mathbf{V}]_{\ell,\ell} = 1 / \sqrt{v_{\ell}}$. It means that for each wavelength ℓ , the estimated flux $\widehat{\alpha}_{\ell}$ is:

$$\widehat{\alpha}_{\ell} = \sqrt{v_{\ell}} x_{\ell}, \quad (28)$$

which corresponds to the same flux estimates as obtained when computing the S/N $_{\ell}$ values channel by channel with Eq. (6) (i.e., accounting for the spectral correlations does not lead to a different estimator because \mathbb{L} is non-singular). The estimator covariance, on the other hand, reflects that flux variations are correlated:

$$[\text{Cov}[\widehat{\alpha}]]_{i,j} = \sqrt{v_i v_j} [\Sigma]_{i,j}. \quad (29)$$

When using instruments with many contiguous spectral bands, a spectral smoothness can also be enforced by favoring fluxes with small variations from one spectral band to the other, as captured by the following regularization term:

$$\mathcal{R}(\alpha) = \frac{1}{2} \|\mathbf{D}\alpha\|_2^2, \quad (30)$$

with \mathbf{D} the matrix of the finite differences.

If a large library of spectra is available, an a priori covariance Γ of the spectrum can be learned, providing a richer modeling than the simple smoothness prior. In the definition of \mathcal{R} , the matrix \mathbf{D} is then replaced by $\mathbf{D} = \Gamma^{-1/2}$, that is, $\mathcal{R}(\alpha) = -\log p(\alpha) + \text{const} = \frac{1}{2}(\alpha - \bar{\alpha})^{\text{T}} \Gamma^{-1} (\alpha - \bar{\alpha}) = \frac{1}{2} \|\Gamma^{-1/2}(\alpha - \bar{\alpha})\|_2^2$, where the prior distribution $p(\alpha)$ is a centered multivariate Gaussian with mean vector $\bar{\alpha}$ and covariance matrix Γ , see for example (Tarantola 2005).

When a regularization term is considered, the estimation of the fluxes α corresponds to a maximum a posteriori (MAP):

$$\widehat{\alpha}^{(\text{MAP})} = \arg \min_{\alpha} \mathcal{C}(\phi_0, \alpha) + \mu \mathcal{R}(\alpha), \quad (31)$$

where μ is a hyperparameter that controls the amount of smoothing introduced by the regularization term. Since both \mathcal{C} and \mathcal{R} are quadratic in α , the MAP estimate can be obtained in closed form and corresponds to the following linear transform of the vector \mathbf{x} of S/N $_{\ell}$ values:

$$\widehat{\alpha}^{(\text{MAP})} = \underbrace{(\widehat{\mathbf{v}} \Sigma^{-1} \mathbf{V} + \mu \mathbf{D}^{\text{T}} \mathbf{D})^{-1}}_{\mathbf{M}(\mu)} \widehat{\mathbf{v}} \widehat{\mathbf{L}} \widehat{\mathbf{S}}, \quad (32)$$

where $\widehat{\mathbf{S}} = \widehat{\mathbf{L}}^{\text{T}} \mathbf{x}$ is the whitened⁶ vector of spectral S/N $_{\ell}$ values, and $\mathbf{M}(\mu)$ is the matrix defining the linear estimator $\widehat{\alpha}^{(\text{MAP})}$, parameterized by the smoothing parameter μ . It can be noted that removing the regularization term (i.e., taking $\mu = 0$) in Eq. (32) yields the same estimator as the one in Eq. (27).

Setting the value of the hyperparameter μ requires some adaptation to both the spectrum smoothness and the integrated flux. In order to obtain a detection and characterization method that is fully automatic, we investigated several strategies to select automatically the value of μ : (i) the generalized maximum likelihood (GML), also known as the evidence method

⁶ We remind that $\widehat{\mathbf{L}} \widehat{\mathbf{L}}^{\text{T}} = \widehat{\Sigma}^{-1}$.

(Wahba 1985; MacKay 1992), which first marginalizes the joint distribution $p(\mathbf{x}, \boldsymbol{\alpha} | \mu, \phi_0)$ with respect to the unknown spectrum $\boldsymbol{\alpha}$, then maximizes the so-called generalized likelihood $p(\mathbf{x} | \mu, \phi_0)$ with respect to μ ; (ii) the generalized cross-validation (Craven & Wahba 1978; Golub et al. 1979), which approximates the error obtained by a leave-one-out validation strategy and is agnostic of the noise variance; (iii) Stein’s unbiased risk estimator (SURE; Stein 1981). Estimating μ with the SURE led to the best overall performance in our simulations, with a slight improvement over GML and a clear gain with respect to GCV, see Appendix G.

After spectral whitening, the vector of whitened S/N_ℓ values is distributed according to a centered Gaussian distribution. The unbiased risk estimate provided by SURE is then (Thompson et al. 1991):

$$\widehat{\text{risk}} = \frac{1}{L} \|(\mathbf{I} - \mathbf{A}(\mu))\widehat{\mathbb{S}}\|_2^2 + \frac{2}{L} \text{tr}(\mathbf{A}(\mu)) - 1, \quad (33)$$

where $\mathbf{A}(\mu) = \mathbb{L}^T \mathbf{V} \mathbf{M}(\mu)$. Therefore, the parameter $\widehat{\mu}^{(\text{SURE})}$ that minimizes the risk⁷ estimate is given by:

$$\widehat{\mu}^{(\text{SURE})} = \arg \min_{\mu} \|(\mathbf{I} - \mathbf{A}(\mu))\widehat{\mathbb{S}}\|_2^2 + 2 \text{tr}(\mathbf{A}(\mu)). \quad (34)$$

Estimating $\widehat{\mu}^{(\text{SURE})}$ is a mono-dimensional minimization that can be performed, for instance, by a golden section search (Brent 1973). Once $\widehat{\mu}^{(\text{SURE})}$ is obtained, the vector of fluxes is computed: $\widehat{\boldsymbol{\alpha}}^{(\text{MAP})} = \mathbf{A}(\widehat{\mu}^{(\text{SURE})})\widehat{\mathbb{S}}$.

5. Results

5.1. Datasets description

In this section, we assess the performance of the proposed PACO ASDI algorithm both in terms of source detection and spectrum estimation. The results are compared to two standard algorithms: TLOCI and KLIP. These two algorithms are currently used for the exploitation of the SPHERE science data (Lagrange et al. 2019a; Mesa et al. 2019b; Gratton et al. 2019; Gibbs et al. 2019; Maire et al. 2019). In the following, they are applied both in ADI and ASDI mode. We used the SpeCal (Galicher et al. 2018) implementation of TLOCI-A(S)DI and KLIP-A(S)DI algorithms which is the current reference and well documented post-processing standard of the SPHERE consortium. The TLOCI implementation is based on the algorithm as described in Galicher et al. (2011), Marois et al. (2014). In particular, it implements a frame selection strategy for the estimation of the stellar PSF to mitigate the self-subtraction of the putative sources. A flat spectral template was considered in our experiments. The KLIP implementation is based on the algorithm as described in Soummer et al. (2012). No frame selection strategy was implemented to minimize the self-subtraction of the putative sources when conducting the PCA. There are several variants of these implementations, especially for the KLIP-ASDI method. In this paper, we used only the publicly available variant of the SPHERE consortium. For all algorithms from the SpeCal software, the S/N maps are computed after annular normalization of the residual images (obtained after subtraction of the estimated on-axis PSF) by the empirical standard deviation of the noise. Besides, with SpeCal, the self-subtraction phenomenon is calibrated before the reduction using massive injections of fake exoplanets to derive the radial throughput of

the algorithms as a function of the angular separation. The estimated exoplanet spectra are then compensated for this calibrated correction, see Delorme et al. (2017) for detailed procedures. In ADI mode, each spectral channel is processed independently. A combined detection map is obtained by a simple summation of the detection maps from the different spectral channels. The spectrum estimation is obtained by a photometry estimation per wavelength using the ADI PSF model. The reason we considered the ADI mode is that it is a common practice to process independently each spectral channel of an ASDI series because the exoplanet signature are usually present in the redder channels.

For the comparisons, we selected four datasets from the SPHERE-IFS instrument obtained in various conditions of observation leading to different degrees of difficulty for the detection and estimation tasks. The four datasets are obtained from the SPHERE-IFS raw data using the pre-reduction and handling pipeline of the SPHERE consortium (Pavlov et al. 2008). Background, flat-field, bad pixels, registration, true-North, wavelength and astrometric calibrations are also performed during this step, see Pavlov et al. (2008), Zurlo et al. (2014), Maire et al. (2016) for the detailed procedures. These processings are followed by additional steps implemented at the SPHERE Data Center (Delorme et al. 2017) to refine the wavelength calibration, reduce the cross-talk and somewhat cope with bad pixels. We did not apply other post-processings such as high-pass filtering. Three of these calibrated datasets are dedicated to the evaluation of the detection performance on fields reported in the literature to contain point sources. The fourth one is used for the evaluation of the spectrum recovery performance. To do so, we numerically injected fake point sources at low fluxes to perform this evaluation. The four datasets considered were recorded around the following reference targets:

HR 8799 (HIP 114189) which is a A5V type star located in the Pegasus constellation. It hosts four confirmed exoplanets, only three of them (HR 8799 c, d, and e) are within the SPHERE-IFS field of view (the last one, HR 8799 b, is visible with the larger field of SPHERE-IRDIS). All of them were discovered (Marois et al. 2008), confirmed (Marois et al. 2010), and widely studied (Bowler et al. 2010; Currie et al. 2011; Marley et al. 2012) by direct imaging. In the following, this dataset is used as a baseline to compare the overall performance of the detection algorithms on sources at a standard level of contrast (between 10^{-6} and 10^{-5}) in the considered spectral band.

β Pictoris (HIP 27321) which is a A6V type star located in the Pictor constellation. It hosts two known exoplanets (β Pictoris b and c) as well as a protoplanetary disk made of gas and dust. β Pictoris b was discovered (Lagrange et al. 2009) and confirmed (Lagrange et al. 2010) by direct imaging. β Pictoris c was discovered more recently by the radial velocities method (Lagrange et al. 2019b). In the following, we do not consider the presence of β Pictoris c⁸. New datasets around this star allowed to constrain better the orbit of the β Pictoris b and to refine its photometry (Lagrange et al. 2019a). In the following, we use a dataset at a challenging epoch (before the conjunction reported

⁸ β Pictoris c has a very small angular separation with its host star (0.10-0.15 arcsec at its maximal elongation). For comparison, the inner working angle of the coronagraphs of SPHERE is 0.125 arcsec. Besides, β Pictoris c has a very low contrast given its separation (about 1×10^{-4} in the H Johnson’s band). For these two reasons, it could be detected by direct imaging only when it is at its maximal elongation, given the current instrumental and processing capabilities, see Lagrange et al. (2019b).

⁷ The risk is defined by: $\text{risk} = \mathbb{E} [\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}\|^2]$.

Table 3. Log information for the considered SPHERE-IFS datasets.

Target	ESO ID	Observation date	T	Δ_{par}	Seeing	Strehl ratio
HR 8799	095.C-0298	2015-07-04	46	16.4°	1.43	78%
β Pictoris	097.C-0865	2016-09-16	116	38.5°	0.45	94%
HD 131399	095.C-0389	2015-06-12	49	37.2°	1.30	78%
HD 172555	095.C-0192	2015-07-11	31 ^(*)	12.9°	1.20	–

Notes. Columns are: target name, ESO survey ID, observation date, number T of available temporal frames, total amount of parallactic rotation Δ_{par} of the field of view, seeing value at the beginning of the observations, and Strehl ratio at 1.6 μm . ^(*)The recorded dataset around HD 172555 was made of 62 temporal frames but we selected $T = 31$ frames by discarding one of them over two, preserving the amount of the parallactic rotation Δ_{par} to slightly increase the difficulty of the spectrum estimation task.

in Lagrange et al. 2019a) since the angular separation of the exoplanet is smaller than 0.15 arcsec, which corresponds to nine pixels away from the coronagraph mask.

HD 131399 (HIP 72940) which is a A1V type star located in the Centaurus constellation. It forms a triple system with two other stars (HD 131399 B and C) located about 349 au from the brightest star HD 131399 A (De Zeeuw et al. 1999; Dommanget & Nys 2002; Pecauc & Mamajek 2013). This system also hosts a faint point source (HD 131399 Ab) discovered by direct imaging (Wagner et al. 2016), at first supposed to be a bounded exoplanet. However, further joint analysis of GEMINI/GPI and VLT/SPHERE datasets to refine the astrometry and the spectrum estimation of the candidate companion led to the conclusion that HD 131399 Ab is more likely to be a background brown dwarf (Nielsen et al. 2017). In the following, this dataset is used to compare the behavior of detection algorithms in the case of a faint point source falling close to the limit of the instrument’s field of view.

HD 172555 (HIP 92024) which is a A5V type star located in the Pavo constellation (Schütz et al. 2005; Lisse et al. 2009). The analysis of directly imaged (A)(S)DI series was conducted in several surveys but, to the best of our knowledge, no point source was ever detected (Nielsen et al. 2008; Nielsen & Close 2010). We use this dataset to conduct the spectrum estimation of numerically injected fake point sources.

Table 3 summarizes the main observation parameters of these datasets. It can noted that the conditions of observations were not particularly good for three of these datasets (seeing values between 1.20 and 1.43).

Besides, a MATLABTM implementation of the PACO ASDI routines for computing spectrally combined wwS/N maps is available online⁹.

5.2. Detection performance

5.2.1. Detection results

In the following, the detection criteria is based on wwS/N map with PACO ASDI since it offers interesting properties in terms of detection sensitivity and controlled false alarms rate when thresholding at 5σ (see Sect. 3.2.3). In our comparison to the state-of-the-art algorithms, we use the final signal-to-noise map (denoted “combined S/N”) provided by the different pipelines. The “combined S/N” map is generally obtained by a weighted mean of the signal-to-noise ratio computed in each channel. This combination is generally followed by a post-processing step via

⁹ <http://doi.org/10.5281/zenodo.3679426>

so-called “unsharp filtering” (high pass filtering) to improve the visual quality of the combined S/N map by attenuating some spurious background artifacts. In PACO ASDI, the spatial whitening and spectral whitening operations can be seen as data-driven and locally-adaptive filters. No additional filtering is required to enhance the detection maps produced by PACO or PACO ASDI.

Figures 9 and 10 show the combined detection maps around HR 8799, β Pictoris, and HD 131399 obtained with TLOCI-A(S)DI, KLIP-A(S)DI, and PACO ASDI for two sets of color bars. The detection is performed by thresholding the maps at $\tau = 5$ (corresponding to a PFA = 2.87×10^{-7}), as classically done in direct imaging.

State-of-the-art algorithms lead to strong radial background artifacts likely caused by a miss-modeling of the spatial and spectral correlations since they take the form of the typical structures observed in the GLRT⁺ maps obtained with PACO ASDI (see Fig. 7a) when the spectral correlations are neglected. In addition, the detection seems very difficult in the regions close to the borders of the instrument’s field of view due to strong artifacts, in particular with TLOCI-ADI and KLIP-ADI, possibly due to a miss-modeling of the aberrant data occurring on the borders. For example, distinguishing the signature of the exoplanet HR 8799 c (top-left corner of the field of view) from the artifacts seems almost impossible with TLOCI-ADI and KLIP-ADI. In ASDI mode, the quality of the detection maps is generally not significantly better in these areas due to artifacts (in particular with TLOCI-ASDI) or regions with zeros or saturated values obtained with KLIP-ASDI (indicated by arrows in the figures), possibly caused by the absence of explicit modelization of areas with missing data for the longest wavelengths. All these sources of artifacts cause a severe limitation of the workable field of view in which an exoplanet can be actually detected. The portion of the field of view impacted by missing or aberrant data increases with the parallactic rotation and with the ratio $\lambda_{\text{max}}/\lambda_{\text{min}}$. It reaches more than 20% for TLOCI-ASDI on the β Pictoris and HR 8799 datasets. PACO ASDI provides stationary detection maps on the whole field of view (including at the vicinity of the host star and close to the borders of the field of view) so that a unique detection threshold can be set. The stationarity is explained both by the local statistical modeling of the spatio-spectral correlations of the background and the explicit consideration of the missing or aberrant data which are flagged as outliers.

Regarding the detection accuracy of the algorithms, only PACO ASDI ensures a statistically-grounded control of the PFA in the sense that no false alarm is generally observed at the 5σ threshold. Unlike PACO ASDI, all other algorithms considered lead to several false detections in the field of view (many more than expected at 5σ). In practice, astronomers are familiar with the “false alarms issue” and generally differentiate candidate companions from the false alarms by visual inspection of the multi-spectral data and of the reduction results.

Regarding the detection sensitivity of the tested algorithms, only PACO ASDI leads to detection peaks significantly higher than the conventional detection threshold ($\tau = 5$) for all the known point sources, even when they are close to the host star (as β Pictoris b) or close to the borders of the field of view (as HD 131399 Ab). For example, for HR 8799, only one source can be unambiguously detected at 5σ with state-of-the-art algorithms on the dataset we have reduced. The other sources can be distinguished by a visual inspection of the detection maps but require to lower the detection confidence at about 3.5 for the exoplanet HR 8799 e and at 4.3 for the exoplanet HR 8799 d. The best results with existing methods are obtained with KLIP-ADI, leading to the detection of HR 8799 c (top-left corner of

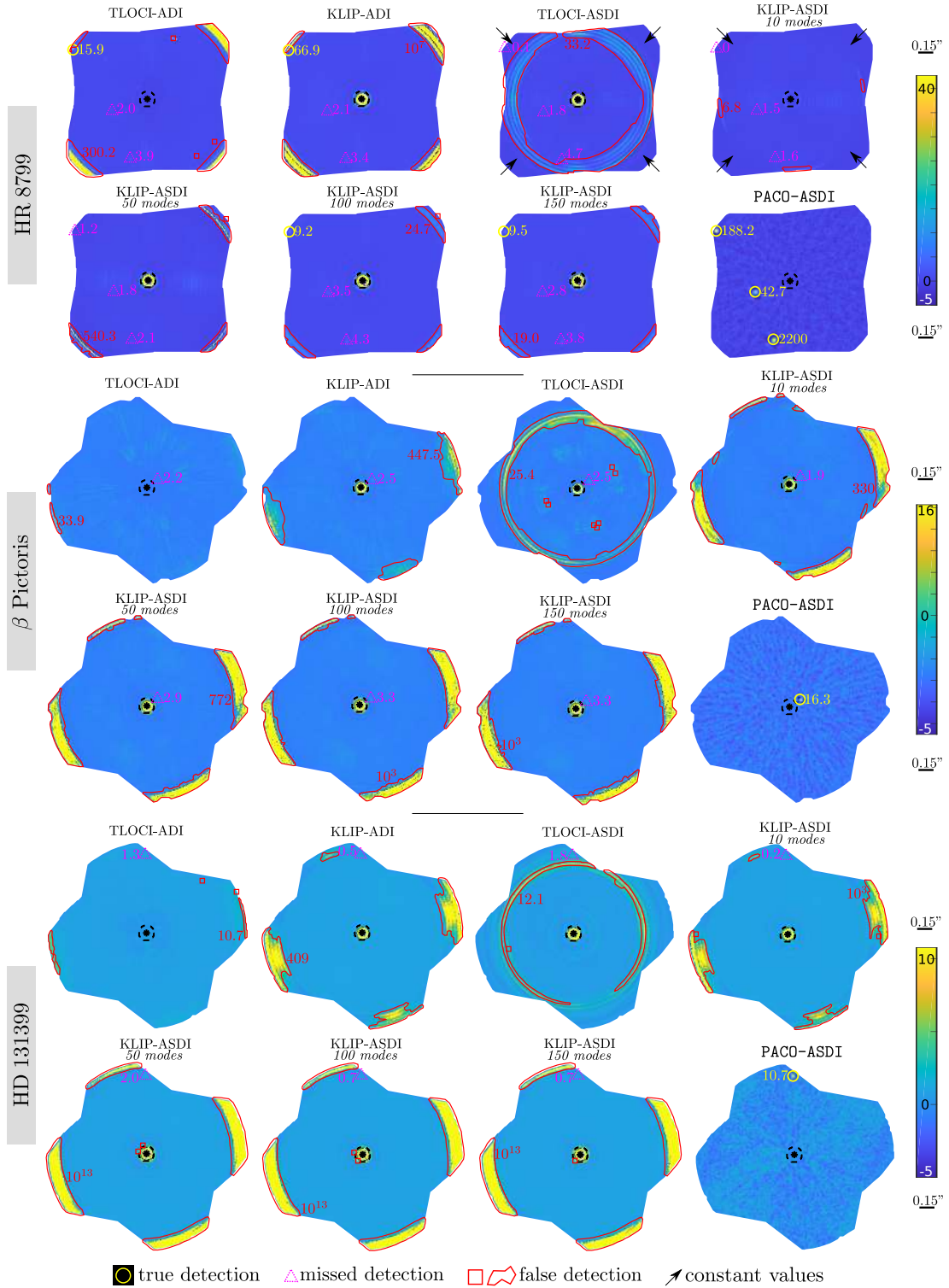


Fig. 9. Detection maps (wWS/N for PACO ASDI and combined S/N for the other algorithms) around HR 8799, β Pictoris and HD 131399 obtained with TLOCI-ADI, KLIP-ADI (50 modes), TLOCI-ASDI, KLIP-ASDI (10, 50, 100 and 150 modes) and the proposed PACO ASDI algorithm. The color bars are common for all methods and are set between -5 and the highest true detection peak provided by PACO ASDI (excepted for HR 8799 for which the color bar is set between -5 and 42.7 corresponding to the wWS/N value of HR 8799 e with PACO ASDI). The detection threshold is set at $\tau = 5$ and the values above this threshold are classified as true detections (yellow circles) and false detections (red squares and polygons). Missed detections are indicated by pink triangles. The value of the largest false alarm is also indicated in red on each map. Black arrows point to areas with constant values.

the field of view) at a combined S/N equal to 66.9. On the same dataset, the three known exoplanets are detected with PACO ASDI and the wWS/N reaches 188.2 for HR 8799 c. For β Pictoris and HD 131399, only PACO ASDI can detect the known

sources without any false alarm in the field of view. Lowering the detection threshold with TLOCI-A(S)DI and KLIP-A(S)DI leads to several false alarms in the field of view thus preventing their automatic detection. Moreover, our experiments

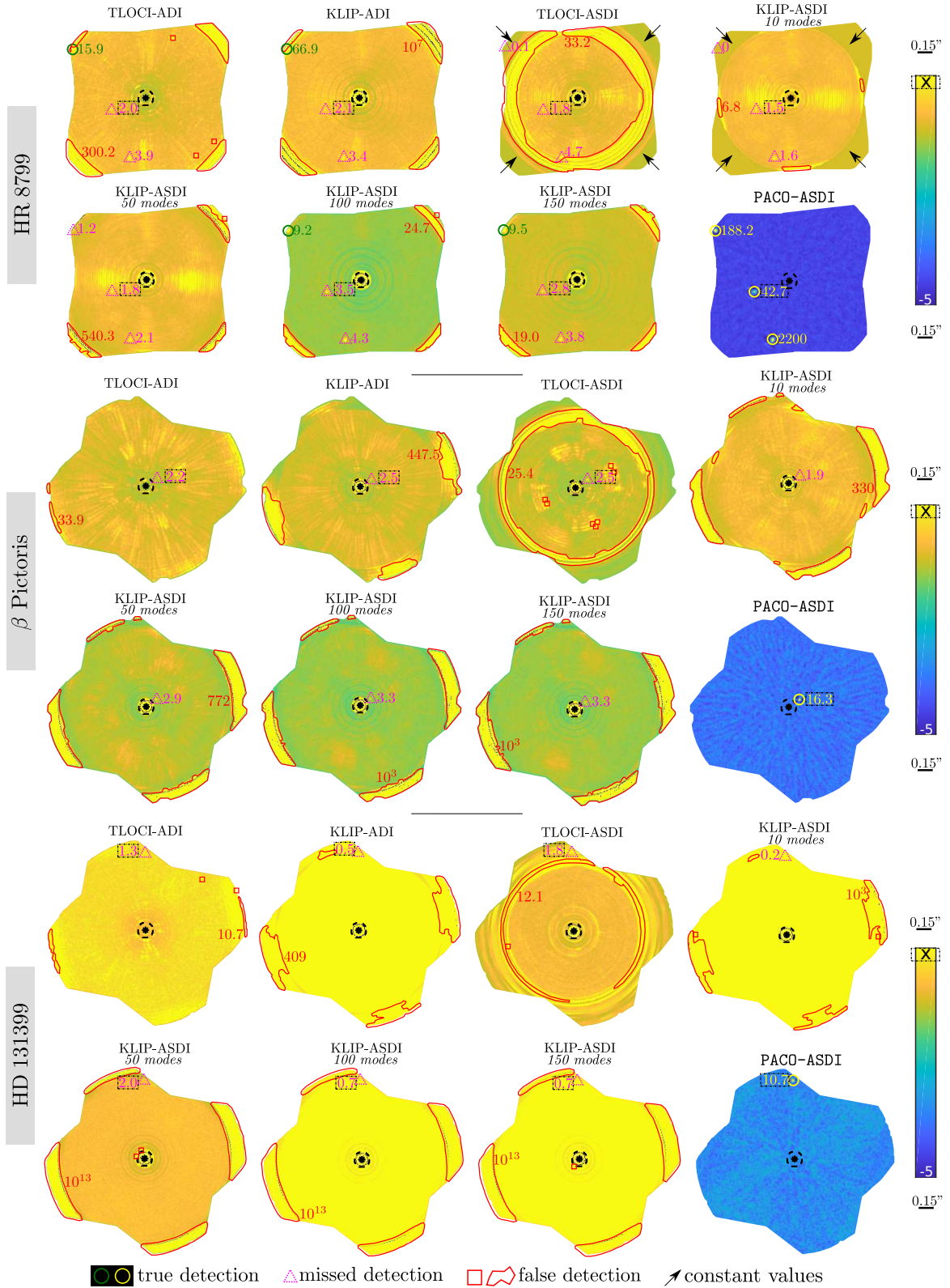


Fig. 10. Same caption than Fig. 9. The color bars are adapted to each method and are set between -5 and the detection peak associated with one of the real sources to be detected (respectively HR 8799 e which is the closest to the host star, β Pictoris b, and HD 131399 Ab).

tend to show that in certain cases it is not so easy to distinguish true detections from false alarms via visual inspection: on the considered β Pictoris dataset processed with TLOCI-A(S)DI and KLIP-A(S)DI algorithms, it seems difficult to visually discriminate β Pictoris b (combined S/N $\in [1.9; 3.3]$) and

HD 131399 Ab (combined S/N $\in [0.2; 2.0]$) from false alarms since the shape of the detection peaks (blobs spatially correlated on a few pixels) are sometimes quite similar. The case of HD 131399 is interesting since the dataset we consider in this paper was already processed by other authors. For example,

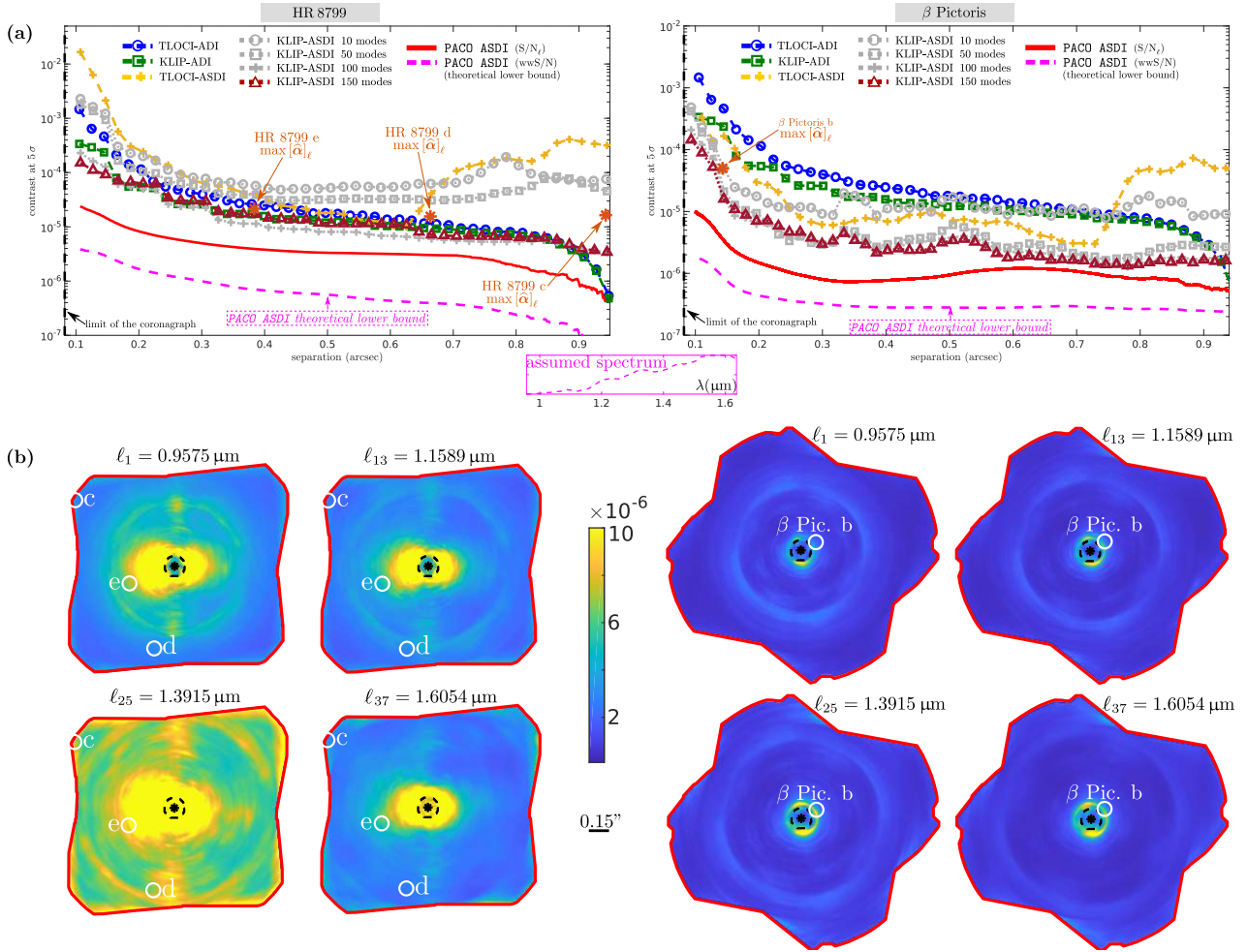


Fig. 11. Achievable 5σ contrast on HR 8799 and β Pictoris. *a*: contrast curves obtained with TLOCI-A(S)DI, KLIP-A(S)DI, and PACO ASDI. All curves correspond to the mean contrast along spectral channels. For PACO ASDI, the solid red line is for the spectral mean S/N_ℓ contrast while the dashed pink line is for the spectral mean wwS/N contrast. The wwS/N contrast is the theoretical lower bound given by Eq. (22) when several spectral channels are combined. Contrast curves as provided by KLIP and TLOCI do not correspond to a 5σ false alarms rate contrarily to the contrast curves of PACO ASDI. The achievable contrasts are thus significantly over-optimistic for KLIP and TLOCI, see discussion in the text (Sect. 5.2.2). *b*: examples of 2D S/N_ℓ contrast maps obtained with PACO ASDI for four spectral channels: $\ell_1 = 0.9575 \mu\text{m}$, $\ell_{13} = 1.1589 \mu\text{m}$, $\ell_{25} = 1.3915 \mu\text{m}$ and $\ell_{37} = 1.6054 \mu\text{m}$. The superimposed white circles represent the locations of the known exoplanets.

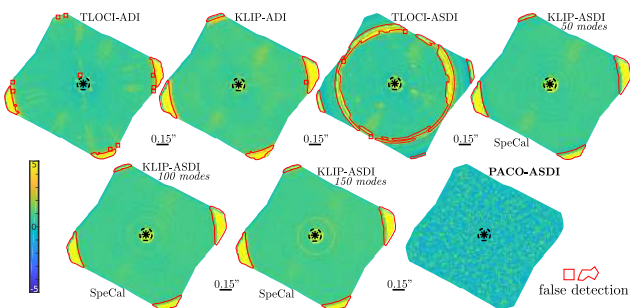


Fig. 12. Detection maps (wwS/N for PACO ASDI and combined S/N for the other algorithms) around HD 172555 in the absence of fake point sources.

Nielsen et al. (2017) report the detection of HD 131399 Ab at a combined S/N between 4.0 and 6.3 with the cADI algorithm (Marois et al. 2006; Lagrange et al. 2010). The detection values reported with the cADI method are higher than the values obtained in our experiments with the TLOCI-A(S)DI and KLIP-A(S)DI algorithms but they remain significantly lower than the

value obtained with PACO ASDI ($wwS/N = 10.3$). The difference between TLOCI & KLIP and cADI could be explained by the low “aggressivity” of the cADI method (i.e., low self-subtraction and reduced amount of artifacts due to the ADI strategy) which seems to be particularly adapted to this target located very near the borders of the field of view. However, while the cADI detection map presented in Fig. 2 top-left corner of Nielsen et al. (2017) allows to identify HD 131399 Ab, several point-like features are also falsely detected with a comparative or higher level of S/N , especially near the coronagraph.

5.2.2. Achievable contrast

In this section, we compare the minimal contrast required to achieve a detection with PACO ASDI to the contrasts of TLOCI and KLIP. As is classically done in the literature, we derive the so-called “ 5σ contrast curves” representing the minimum contrast of a source to still be detected with a probability of detection $PD = 0.5$ when the detection threshold is set to obtain a probability of false alarms $PFA = 2.87 \times 10^{-7}$. This achievable contrast can be computed for the single-wavelength detection maps.

Table 4. Angular separation, minimum, maximum, and mean contrast of the 12 fake point sources injected in the fourth dataset.

Id.	Sep.	min[α_ℓ]	max[α_ℓ]	mean[α_ℓ]
#1	0.220''	2.3×10^{-5}	8.0×10^{-5}	5.0×10^{-5}
#2	0.306''	1.2×10^{-5}	5.0×10^{-5}	2.8×10^{-5}
#3	0.264''	1.3×10^{-5}	2.0×10^{-5}	1.8×10^{-5}
#4	0.399''	6.2×10^{-8}	3.0×10^{-5}	1.1×10^{-5}
#5	0.454''	3.3×10^{-6}	1.0×10^{-5}	6.6×10^{-6}
#6	0.682''	1.9×10^{-7}	9.0×10^{-5}	3.3×10^{-5}
#7	0.166''	1.8×10^{-6}	6.6×10^{-6}	3.8×10^{-6}
#8	0.187''	3.6×10^{-6}	1.5×10^{-5}	8.5×10^{-6}
#9	0.219''	4.1×10^{-9}	2.0×10^{-6}	7.5×10^{-7}
#10	0.318''	2.6×10^{-6}	4.0×10^{-6}	3.6×10^{-6}
#11	0.504''	3.3×10^{-7}	1.0×10^{-6}	6.6×10^{-7}
#12	0.618''	2.1×10^{-6}	7.0×10^{-6}	4.4×10^{-6}

Notes. The three sources (#5, #10, and #12) on which we perform additional Monte–Carlo estimations are emphasized in bold font.

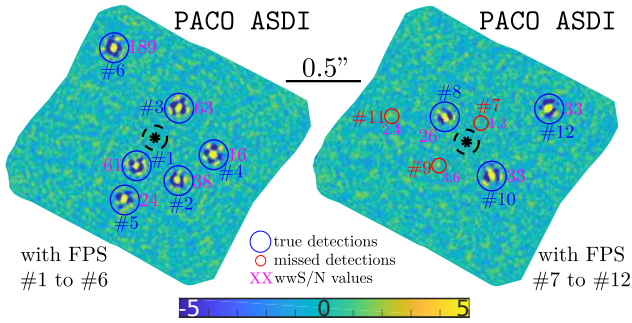


Fig. 13. wwS/N maps obtained with PACO ASDI around HD 172555 with injected fake point sources #1 to #6 (left) and #7 to #12 (right).

As detailed in Sect. 3.2.3, the 5σ contrast in channel ℓ is $5\sqrt{v_\ell}$ (the minimum contrast α_ℓ of the source in the spectral channel ℓ so that $P(S/N_\ell > \tau) = 0.5$). As detailed in Sect. 3.2.3, the combination of the S/N_ℓ maps improves the achievable contrast. When the combined detection map wwS/N is used, the achievable contrast is given by Eq. (22) if the prior spectrum perfectly matches the source spectrum. In practice, this theoretical lower bound is not reached, for the reasons discussed in Sect. 3.2.3 and, like with PACO, because at the detection stage the S/N_ℓ values are underestimated, the background statistics being estimated in the presence of the source. From our experience, values of the contrast achieved for single-wavelength detections are typically reached in practice with S/N_ℓ and can thus be used as a safe value of the achievable contrast.

Figure 11a gives the achievable 5σ contrast curves obtained with TLOCI-A(S)DI, KLIP-A(S)DI, and PACO ASDI on HR 8799 and β Pictoris. For PACO ASDI, both the S/N_ℓ contrast and the combined wwS/N contrast are represented. Considering the S/N_ℓ contrast curves of PACO ASDI, a clear gain is observed at small angular separations (≤ 0.7 arcsec) comparatively to the state-of-the-art algorithms. At larger separations, this gain is maintained except for KLIP-ASDI which can reach better contrasts. However, as already observed and discussed in our previous paper (Flasseur et al. 2018a) on ADI series, contrast curves produced by state-of-the-art algorithms are often optimistic both in terms of PD and PFA. These previous observations also verify here in ASDI mode: all detection maps of

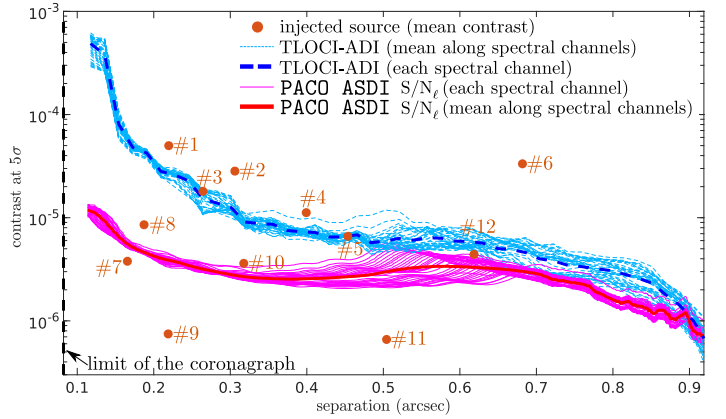


Fig. 14. Contrast curves at 5σ obtained on HD 172555 with PACO ASDI comparatively to TLOCI-ADI. The mean contrast of the fake faint point sources #1 to #12 is marked by orange points.

state-of-the-arts algorithms present many more false alarms than what would be expected at 5σ . According to Fig. 11, far from the β Pictoris star, the best achievable contrast is reached when using KLIP-ASDI with 150 modes and PACO ASDI that converge towards the same detection limit. However, Figs. 9 and 10 show that KLIP-ASDI produces several false alarms in the field of view and a lower signal-to-noise ratio for the exoplanet compared to PACO ASDI. With KLIP-ASDI, the largest value of a false alarm is significantly higher than 5.0, while it should be very unlikely to have a background value larger than 4.0. This illustrates that the 5σ contrast of state-of-the-art algorithms does not correspond to the expected level of PFA and can thus only be used to perform relative comparisons. We also give the contrast reached by PACO ASDI when single-wavelength detection maps are combined. A theoretical gain (see Eq. (21)) slightly less than one order of magnitude is expected due to the combination, according to Fig. 11. When comparing the S/N_ℓ values of the point sources in the single-detection maps of HR 8799 shown in Fig. 6b (values in the range 1.5–5.5) to the values in the combined detection maps of Figs. 9 and 10 ($wwS/N = 42.7$), the improvement is in relatively good agreement with the contrast curves of Fig. 11a.

As for PACO (Flasseur et al. 2018a), the achievable contrast of PACO ASDI can be computed at every point of the field of view. Figure 11b gives examples of 2D contrast maps obtained with PACO ASDI for some selected spectral channels. They show that the detection is more favorable on certain spectral channels than others. For example, the achievable contrast on spectral channel $\ell_{25} = 1.3915 \mu\text{m}$ is about twice worse than the one obtained on spectral channel $\ell_{13} = 1.1589 \mu\text{m}$. This can be explained by the presence of large spectral variations of the intensity fluctuations or additional noise probably caused by the low atmospheric transmission. Interestingly, these maps indicate that the achievable contrast varies significantly along an annulus of fixed angular separation. It is particularly the case near the host star since the residual central halo is not isotropic. Thus, the 2D contrast information can be useful to derive an accurate estimation of the achievable contrast given the angular location of a detected source.

5.3. Spectrum estimation performance

In this section, we evaluate the performance for the spectrum extraction of point sources. As mentioned in Sect. 1, the

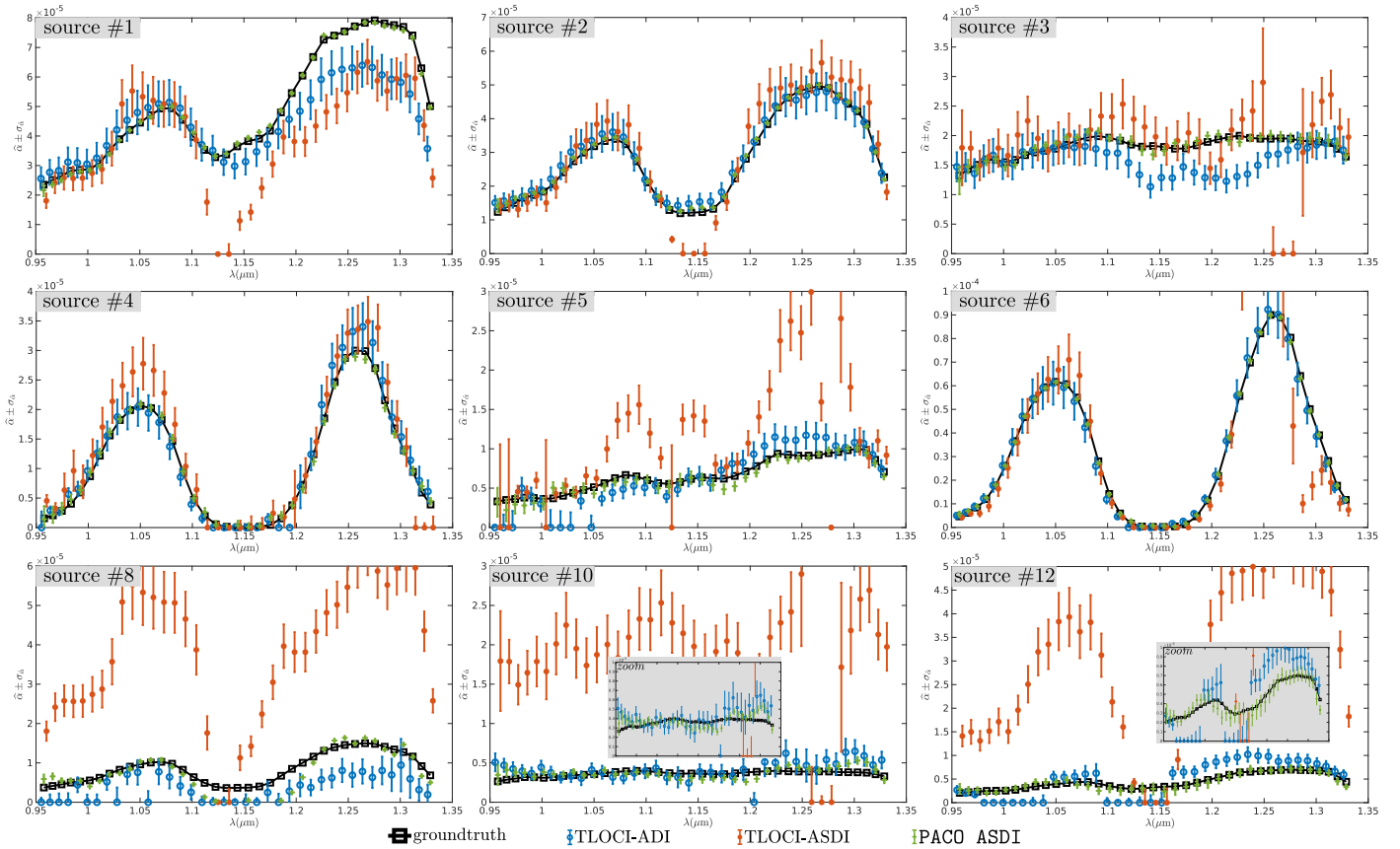


Fig. 15. Estimated spectrum of the detectable fake faint point sources (#1 to #6 plus #8, #10, and #12) obtained with TLOCI-ADI (blue), TLOCI-ASDI (orange), and PACO ASDI (green). The spectrum ground truths of the different fake faint point sources are marked by black lines. The given 1σ confidence intervals are those predicted by the considered algorithms. A zoom around the ground truth is added for sources #10 and #12.

recovered spectra are not corrected for the stellar spectrum in this paper.

We first use numerical injections of fake point sources for the quantitative characterization of PACO ASDI. For this purpose, we use the dataset around HD 172555 (see Sect. 5.1) with no detectable source. Figure 12 gives the wS/N map obtained with PACO ASDI and the combined signal-to-noise ratio maps obtained with other algorithms on this dataset, before injecting the fake point sources. While state-of-the-art algorithms reveal several areas above the detection threshold at $\tau = 5$, experts did not identify consistent detection peaks by closer inspection. In addition, PACO ASDI does not identify any significant detection peak at 5σ . We inject 12 fake point sources in the field of view with a variety of mean contrast and true spectra. Table 4 gives astrometric and photometric information about the fake point sources. Sources #1 to #6 have a mean flux lower or equal to 5×10^{-5} while sources #7 to #12 have a mean flux lower or equal to 8.5×10^{-6} . Figure 13 presents the wS/N maps obtained with PACO ASDI around HD 172555 with fake point sources #1 to #6 (left) and #7 to #12 (right) simultaneously injected. Figure 14 gives contrast curves at 5σ obtained on the considered dataset with PACO ASDI comparatively to TLOCI-ADI. Contrast curves indicate that sources #1 to #6, #8, #10, and #12 can be detected at 5σ by PACO ASDI while the sources #7, #9, and #11 are too faint to be detected from the S/N_ℓ maps at the considered angular separations, see Fig. 13.

Figure 15 shows the spectrum estimation for the nine detectable fake sources obtained by PACO ASDI, TLOCI-ADI, and TLOCI-ASDI. These results show that the spectrum

estimations provided by PACO ASDI are in good agreement with the ground truth since no systematic photometric bias can be observed. We note only one significant discrepancy between the estimated spectrum by PACO ASDI and the ground truth occurring for source #8 between $1.07 \mu\text{m}$ and $1.22 \mu\text{m}$. This discrepancy can be explained both by the very faint source contrast in this spectral band (lower than 5×10^{-6}), the proximity with the host star (angular separation equals to 0.187 arcsec) and the presence of a “dark” speckle near the injection leading to a negative estimated source flux (thresholded at 0) at the corresponding wavelengths. The predicted spectrum confidence intervals also seem coherent with the empirical standard deviation of estimation. The spectrum estimates obtained by PACO ASDI are qualitatively much better than those obtained with TLOCI.

To complete this statistical study, we consider three sources from the 12 previous ones (sources #5, #10, and #12) for which we perform 30 Monte–Carlo injections/spectrum estimations over a circular annulus (i.e., at constant angular separations). Figure 16 gives the 30 estimated spectra obtained with TLOCI-A(S)DI, KLIP-ADI, and PACO ASDI and the mean estimations. The confidence intervals provided by the different algorithms can be compared to the empirical 1σ confidence intervals. Table 5 complements this figure with statistical results in terms of photometric bias, agreement between the predicted and the empirical confidence intervals, and mean square error (MSE). These results show that PACO ASDI is photometrically unbiased in the sense that the photometric bias is negligible (about $\pm 1\%$ of the mean contrast of the sources) without resorting to Monte–Carlo methods to estimate and compensate the

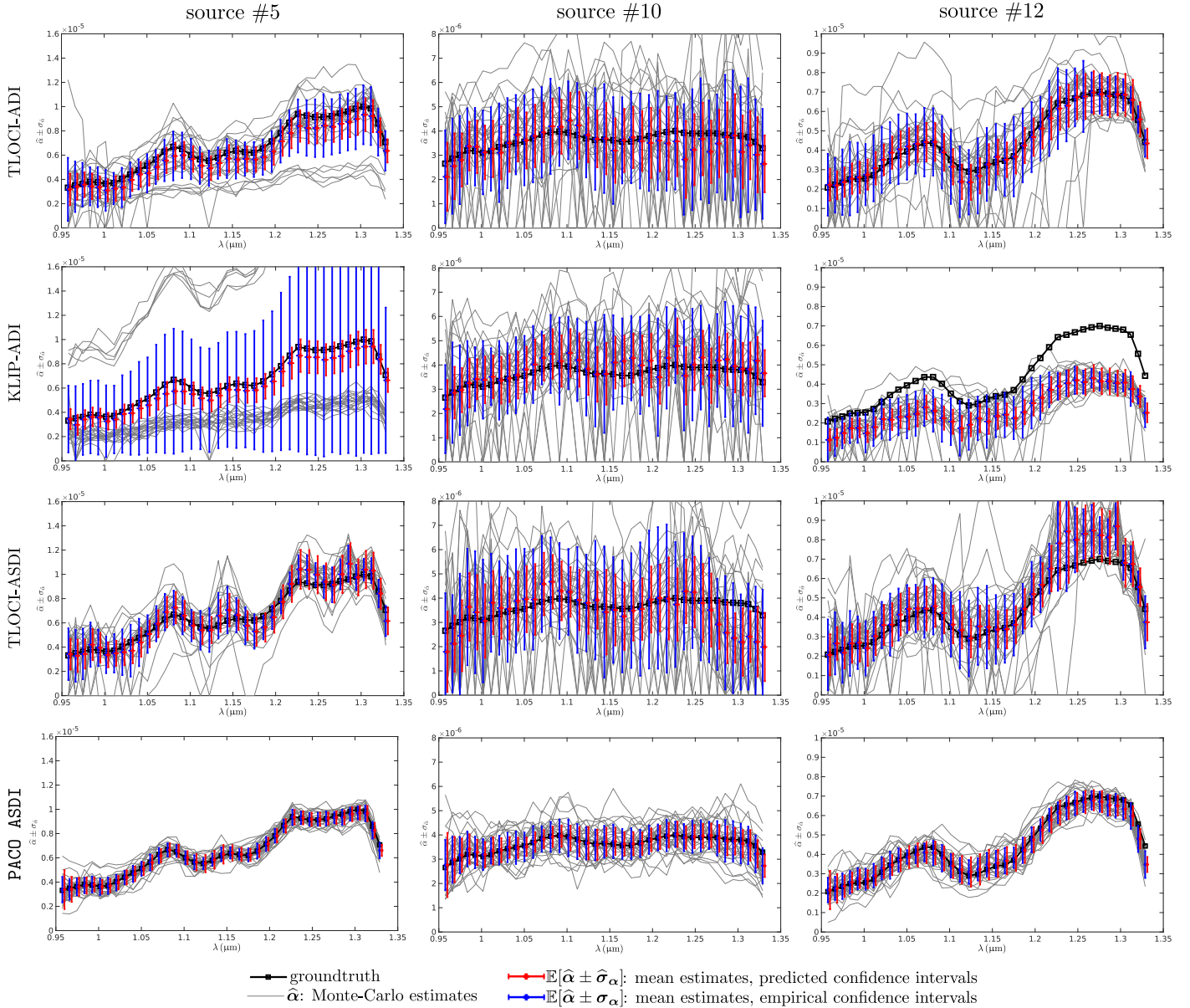


Fig. 16. Monte–Carlo estimated spectrum for fake point sources #5, #10 and #12 obtained with TLOCI-ADI, KLIP-ADI (5 modes), TLOCI-ASDI and PACO-ASDI. For each of the three considered sources, the 30 Monte–Carlo spectrum estimations are given in gray line. Red and blue lines compare the 1σ predicted confidence intervals to the empirical ones centered on the mean estimated spectra over the 30 Monte–Carlo estimations. The spectrum ground truth of the considered sources is in black.

potential source self-subtraction phenomenon, as is common practice with most of the state-of-the-art algorithms. In comparison, this relative photometric bias reaches in most of the cases 4% of the mean contrast of the sources with other methods (excepted for source #12 with TLOCI-ADI). The results of state-of-the-art methods are generally not better in ASDI than in ADI mode. This could be explained by the stronger source self-subtraction that occurs when several spectral channels are processed jointly. This observation could illustrate why experts tend to apply also ADI detection and/or characterization algorithms on ASDI datasets by processing each spectral channel independently (Nielsen et al. 2017; Perrot et al. 2019; Gibbs et al. 2019; Maire et al. 2020, see also Maire et al. 2014; Rameau et al. 2015 for the challenge of ASDI processing). The empirical confidence intervals are also smaller with PACO-ASDI. The RMSE of the estimated spectra is reduced by a factor at least two by PACO

ASDI with respect to other algorithms (on average, by a factor 3.6 for the three sources analyzed in Table 5). Moreover, the confidence intervals provided by the method are in good match with the observed standard deviations of the Monte–Carlo simulation: the ratio between the predicted standard deviation and the empirical standard deviation is between 0.94 and 1.21 for PACO-ASDI, which is closer to one than for other methods (i.e., PACO-ASDI confidence intervals are more reliable).

We also illustrate astrometric and spectrum estimations on the real point sources in the first three datasets. Table 6 presents the estimated astrometry and Fig. 17 gives the estimated spectra with PACO-ASDI of the exoplanets HR 8799 c-d-e, β Pictoris b, and the background source HD 131399 Ab. The datasets are processed in a totally automatic fashion as described in Flasseur et al. (2018c) for ADI datasets. A joint refinement of the astrometry and photometry estimations of the source with

Table 5. Monte–Carlo validation of spectrum estimation methods.

	TLOCI-ADI	KLIP-ADI	TLOCI-ASDI	PACO ASDI
<i>Source #5: $\langle\alpha_\ell\rangle_\ell = 6.6 \times 10^{-6}$</i>				
bias: $\langle\frac{\widehat{\alpha}_\ell - \alpha_\ell}{\langle\alpha_\ell\rangle_\ell}\rangle_\ell$	-9.2%	-7.4%	+1.3%	-0.7%
predicted std: $\langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	1.1×10^{-6}	1.0×10^{-6}	1.3×10^{-6}	6.2×10^{-7}
empirical std: $\langle\sigma_{\alpha_\ell}\rangle_\ell$	1.8×10^{-6}	5.4×10^{-6}	1.5×10^{-6}	6.0×10^{-7}
ratio $\langle\sigma_{\alpha_\ell}\rangle_\ell / \langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	1.65	5.24	1.13	0.97
RMSE: $\sqrt{\langle\ \widehat{\alpha} - \alpha\ ^2\rangle}$	1.9×10^{-6}	5.7×10^{-6}	1.6×10^{-6}	6.1×10^{-7}
<i>Source #10: $\langle\alpha_\ell\rangle_\ell = 3.6 \times 10^{-6}$</i>				
bias: $\langle\frac{\widehat{\alpha}_\ell - \alpha_\ell}{\langle\alpha_\ell\rangle_\ell}\rangle_\ell$	-4.2%	+5.8%	-4.7%	-0.1%
predicted std: $\langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	1.3×10^{-6}	1.0×10^{-6}	1.3×10^{-6}	5.5×10^{-7}
empirical std: $\langle\sigma_{\alpha_\ell}\rangle_\ell$	2.0×10^{-6}	1.9×10^{-6}	2.3×10^{-6}	6.7×10^{-7}
ratio $\langle\sigma_{\alpha_\ell}\rangle_\ell / \langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	1.54	1.90	1.78	1.21
RMSE: $\sqrt{\langle\ \widehat{\alpha} - \alpha\ ^2\rangle}$	2.0×10^{-6}	1.9×10^{-6}	2.4×10^{-6}	6.9×10^{-7}
<i>Source #12: $\langle\alpha_\ell\rangle_\ell = 4.4 \times 10^{-6}$</i>				
bias: $\langle\frac{\widehat{\alpha}_\ell - \alpha_\ell}{\langle\alpha_\ell\rangle_\ell}\rangle_\ell$	-0.3%	-38.5%	+10.0%	+1.1%
predicted std: $\langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	9.4×10^{-7}	6.0×10^{-7}	1.1×10^{-7}	6.0×10^{-7}
empirical std: $\langle\sigma_{\alpha_\ell}\rangle_\ell$	1.5×10^{-6}	9.2×10^{-7}	1.5×10^{-6}	5.7×10^{-7}
ratio $\langle\sigma_{\alpha_\ell}\rangle_\ell / \langle\widehat{\sigma}_{\alpha_\ell}\rangle_\ell$	1.66	1.53	1.29	0.94
RMSE: $\sqrt{\langle\ \widehat{\alpha} - \alpha\ ^2\rangle}$	1.6×10^{-6}	2.0×10^{-6}	1.7×10^{-6}	6.2×10^{-7}

Notes. Best results are highlighted in blue while the worst results are in red.

Table 6. Estimated astrometry (separation and true-north aligned angle) of the real faint point sources known in the used datasets with PACO ASDI.

Source	Est. sep. (arcsec)	Est. angle (°)
HR 8799 c	0.9425 ± 0.0031	328.31 ± 0.19
HR 8799 d	0.6636 ± 0.0065	218.94 ± 0.51
HR 8799 e	0.3834 ± 0.0068	274.80 ± 1.09
β Pictoris b	0.1423 ± 0.0063	216.25 ± 1.70
HD 131399 Ab	0.8412 ± 0.0065	194.74 ± 0.59

Notes. The confidence intervals are given at 1σ .

the highest wwS/N is performed. Its estimated flux contribution is then subtracted to the data and the wwS/N map is updated using a conventional cleaning approach (see Fig. 17 for cleaned wwS/N examples). This procedure is repeated until no source at a significant level of signal-to-noise ratio is detected in the wwS/N map. The estimated spectra are quite smooth and coherent between one spectral channel to the other with realable confidence intervals. The spectrum extraction is especially challenging for HD 131399 Ab since it is located near the borders of the SPHERE-IFS field of view (see Figs. 9 and 10), and the observing conditions of the considered dataset were not particularly good (seeing about 1.30, see Table 3). Nielsen et al. (2017) use photometric and astrometric estimations from VLT/SPHERE and GEMINI/GPI to show that HD 131399 Ab is more likely a background brown dwarf. This result was a revision of the previous status of HD 131399 Ab firstly considered as an exoplanet just after its discovery and confirmation on the basis of (noisy) extracted astrometry and photometry (Wagner et al. 2016). The difficulty to ascertain that HD 131399 Ab was an exoplanet based on the first observations illustrates the importance of algorithms that can provide reliable astrometric and spectrum estimations. We expect PACO ASDI to help refining the estimated orbit and the spectral characterization of future candidate companions.

6. Conclusion

ASDI observations provide very rich data for the detection and characterization of point sources such as exoplanets. Data processing algorithms form an important component in high contrast imaging. Despite extensive efforts in the design of coronagraphic instruments, the separation of the signal of interest (off-axis sources) from the background signal of the on-axis star has to be performed under adverse conditions: large temporal and spectral fluctuations of the intensity of the background, strong correlations and contamination by outliers. Data processing algorithms must be designed with the aim to be robust to all these characteristics of the noise. We have shown in this paper that data-driven statistical modeling paved the way to reliable source detection and characterization methods.

PACO ASDI, the data processing algorithm introduced in this paper, produces detection maps with improved sensibility compared to existing methods. In particular, the detection maps are obtained through a whitening and weighting strategy accounting for the spectral correlations. We show that this general principle is also beneficial to combine detection maps from other existing detection methods. An important additional feature of PACO ASDI is the control of the probability of false alarms: detection maps can be reliably thresholded. Using the conventional 5σ threshold generally produces no false alarm in an IFS field. This contrasts with detection maps obtained with other methods for which many false alarms are observed, in particular at very small angular separations and close to the borders of the field of view. Full exploitation of the field of view seems to be a feature of PACO ASDI that is shared by few other methods.

The elaborate statistical model of PACO ASDI, accounting for spatial, spectral, and temporal fluctuations, is also used to characterize the astrometry and photometry of the detected point sources. By refining the model of the background jointly with the estimation of the source spectrum, the bias due to source self-subtraction is prevented. The spectrum is estimated using a parameter-free spectral regularization. Our numerical

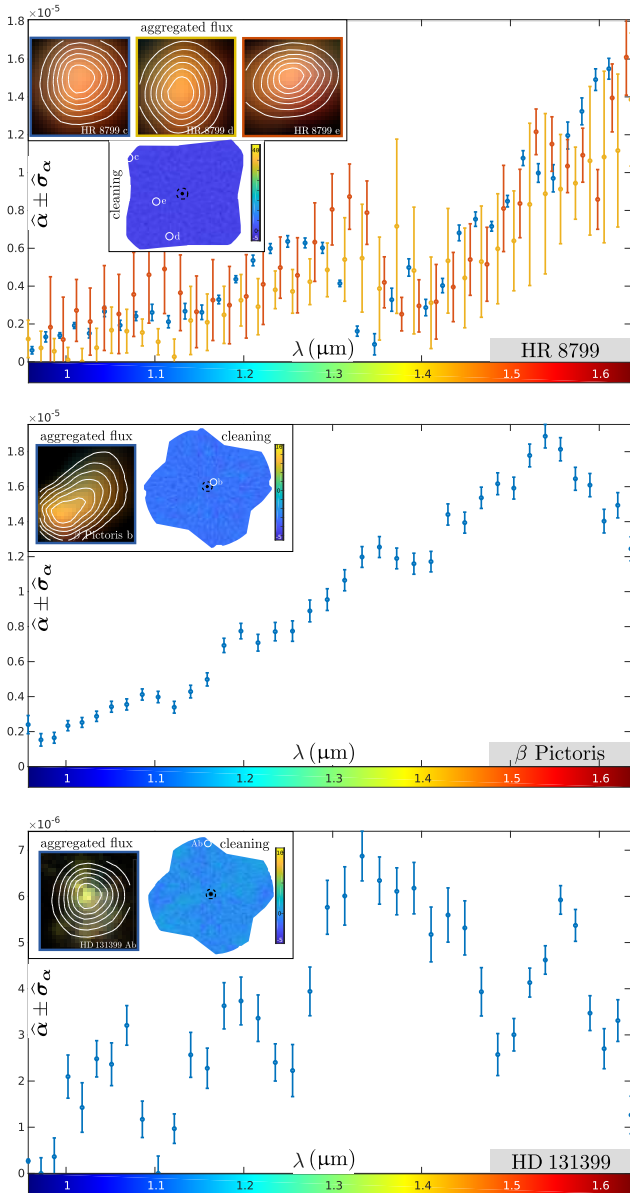


Fig. 17. Estimated spectra using PACO ASDI of the known real faint point sources of the considered datasets (*top*: HR 8799 c–d–e, *middle*: β Pictoris b, *bottom*: HD 131399 Ab). The inserts show a residual $wW/S/N$ map after “cleaning” the contribution of the detected sources. Synthetic subpixel views (4 nodes per pixels) show with false colors the aggregated flux of the detected sources along the different spectral channels (blue for $\ell_1 = 0.9575 \mu\text{m}$ and red for $\ell_{L=39} = 1.6357 \mu\text{m}$).

experiments show reduced estimation errors compared to standard methods.

Beyond the direct analysis of ASDI datasets, PACO ASDI also provides information on the achievable contrast, photometric and astrometric accuracies that are reached for given instrumental and observational conditions. The impact of different observation scenarios (spectral coverage, parallactic rotation) can then be assessed using a data-driven model whose prediction accuracy has been validated on real data.

Acknowledgements. We thank A. Boccaletti (LESIA, Paris, France) who provided the transmission of the SPHERE coronagraphs. We also thank the anonymous Referee for his careful reading of the manuscript as well as his insightful comments and suggestions. This work has made use of the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble, France), PYTH-

EAS/LAM/CESAM (Marseille, France), OCA/Lagrange (Nice, France), Observatoire de Paris/LESIA (Paris, France), and Observatoire de Lyon/CRAL (Lyon, France).

References

- Absil, O., Milli, J., Mawet, D., et al. 2013, *A&A*, 559, L12
 Bailey, V., Meshkat, T., Reiter, M., et al. 2013, *ApJ*, 780, L4
 Beuzit, J.-L., Feldt, M., Dohlen, K., et al. 2008, *Proc. SPIE*, 7014, 701418
 Beuzit, J.-L., Vigan, A., Mouillet, D., et al. 2019, *A&A*, 631, A155
 Bowler, B. P. 2016, *PASP*, 128, 102001
 Bowler, B. P., Liu, M. C., Dupuy, T. J., & Cushing, M. C. 2010, *ApJ*, 723, 850
 Brent, R. P. 1973, *Algorithms for Minimization without Derivatives* (Englewood Cliffs, NJ: Prentice-Hall)
 Cantalloube, F. 2016, Ph.D. Thesis, Grenoble Alpes
 Cantalloube, F., Mouillet, D., Mugnier, L., et al. 2015, *A&A*, 582, A89
 Cantalloube, F., Ygouf, M., Mugnier, L., et al. 2018, ArXiv e-prints [arXiv:1812.04312]
 Chauvin, G., Desidera, S., Lagrange, A.-M., et al. 2017, *A&A*, 605, L9
 Cheetham, A., Samland, M., Brems, S., et al. 2019, *A&A*, 622, A80
 Claudi, R., Maire, A.-L., Mesa, D., et al. 2019, *A&A*, 622, A96
 Craven, P., & Wahba, G. 1978, *Numer. Math.*, 31, 377
 Currie, T., Burrows, A., Itoh, Y., et al. 2011, *ApJ*, 729, 128
 Currie, T., Debes, J., Rodigas, T. J., et al. 2012a, *ApJ*, 760, L32
 Currie, T., Fukagawa, M., Thalmann, C., Matsumura, S., & Plavchan, P. 2012b, *ApJ*, 755, L34
 De Zeeuw, P., Hoogerwerf, R. V., de Bruijne, J. H., Brown, A., & Blaauw, A. 1999, *AJ*, 117, 354
 Delorme, P., Meunier, N., Albert, D., et al. 2017, *SF2A-2017: Proceedings*, 347
 Devaney, N., & Thiébaud, É. 2017, *MNRAS*, 472, 3734
 Domnanget, J., & Nys, O. 2002, *VizieR Online Data Catalog*: I/274
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2018a, *A&A*, 618, A138
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2018b, in *International Society for Optics and Photonics*, Adapt. Opt. Syst. VI, 10703, 10703R
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2018c, in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2735
 Flasseur, O., Denis, L., Thiébaud, É., & Langlois, M. 2020, *A&A*, 634, A2
 Galicher, R., Marois, C., Macintosh, B., Barman, T., & Konopacky, Q. 2011, *ApJ*, 739, L41
 Galicher, R., Boccaletti, A., Mesa, D., et al. 2018, *A&A*, 615, A92
 Gerard, B. L., Marois, C., Currie, T., et al. 2019, *AJ*, 158, 36
 Gibbs, A., Wagner, K., Apai, D., et al. 2019, *AJ*, 157, 39
 Golub, G. H., Heath, M., & Wahba, G. 1979, *Technometrics*, 21, 215
 Gonzalez, C. G., Absil, O., Absil, P.-A., et al. 2016, *A&A*, 589, A54
 Gonzalez, C., Absil, O., & Van Droogenbroeck, M. 2018, *A&A*, 613, A71
 Gratton, R., Ligi, R., Sissa, E., et al. 2019, *A&A*, 623, A140
 Howard, A. W., Johnson, J. A., Marcy, G. W., et al. 2010, *ApJ*, 721, 1467
 Hubert, M., Rousseuw, P. J., & Van Aelst, S. 2008, *Stat. Sci.*, 92
 Jovanovic, N., Martinache, F., Guyon, O., et al. 2015, *PASP*, 127, 890
 Kendall, M. G., Stuart, A., & Ord, J. K. 1948, *The Advanced Theory of Statistics* (JSTOR), 1
 Keppler, M., Benisty, M., Müller, A., et al. 2018, *A&A*, 617, A44
 Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, E. 2007, *ApJ*, 660, 770
 Lagrange, A.-M., Gratadour, D., Chauvin, G., et al. 2009, *A&A*, 493, L21
 Lagrange, A.-M., Bonnefoy, M., Chauvin, G., et al. 2010, *Science*, 329, 57
 Lagrange, A.-M., Boccaletti, A., Langlois, M., et al. 2019a, *A&A*, 621, L8
 Lagrange, A.-M., Meunier, N., Rubini, P., et al. 2019b, *Nat. Astron.*, 1
 Lisse, C. M., Chen, C., Wyatt, M., et al. 2009, *ApJ*, 701, 2019
 Lovis, C., & Fischer, D. 2010, *Exoplanets*, 27
 Macintosh, B., Graham, J., Barman, T., et al. 2015, *Science*, 350, 64
 Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, *Proc. Nat. Acad. Sci.*, 111, 12661
 MacKay, D. J. 1992, *Neural Comput.*, 4, 415
 Maire, A.-L., Boccaletti, A., Rameau, J., et al. 2014, *A&A*, 566, A126
 Maire, A.-L., Langlois, M., Dohlen, K., et al. 2016, in *International Society for Optics and Photonics*, Ground-based Airborne Instrum. Astron. VI, 9908, 990834
 Maire, A.-L., Rodet, L., Cantalloube, F., et al. 2019, *A&A*, 624, A118
 Maire, A.-L., Baudino, J.-L., Desidera, S., et al. 2020, *A&A*, 633, L2
 Marley, M. S., Saumon, D., Cushing, M., et al. 2012, *ApJ*, 754, 135
 Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, *ApJ*, 641, 556
 Marois, C., Macintosh, B., Barman, T., et al. 2008, *Science*, 322, 1348

- Marois, C., Zuckerman, B., Konopacky, Q. M., Macintosh, B., & Barman, T. 2010, *Nature*, **468**, 1080
- Marois, C., Correia, C., Veran, J.-P., & Currie, T. 2013, *Proc. Int. Astron. Union*, **8**, 48
- Marois, C., Correia, C., Galicher, R., et al. 2014, in *International Society for Optics and Photonics*, Adapt. Opt. Syst. IV, 9148, 91480U
- Mawet, D., Hirsch, L., Lee, E. J., et al. 2019, *AJ*, **157**, 33
- Mesa, D., Keppler, M., Cantalloube, F., et al. 2019a, *A&A*, **632**, A25
- Mesa, D., Bonnefoy, M., Gratton, R., et al. 2019b, *A&A*, **624**, A4
- Morzinski, K. M., Close, L. M., Males, J. R., et al. 2014, in *International Society for Optics and Photonics*, Adapt. Opt. Syst. IV, 9148, 914804
- Mugnier, L. M., Cornia, A., Sauvage, J.-F., et al. 2009, *J. Opt. Soc. Am. A*, **26**, 1326
- Mller, A., Keppler, M., Henning, T., et al. 2018, *A&A*, **617**, L2
- Nielsen, E. L., & Close, L. M. 2010, *ApJ*, **717**, 878
- Nielsen, E. L., Close, L. M., Biller, B. A., Masciadri, E., & Lenzen, R. 2008, *ApJ*, **674**, 466
- Nielsen, E. L., Liu, M. C., Wahhaj, Z., et al. 2012, *ApJ*, **750**, 53
- Nielsen, E. L., De Rosa, R. J., Rameau, J., et al. 2017, *AJ*, **154**, 218
- Nielsen, E. L., De Rosa, R. J., Macintosh, B., et al. 2019, *AJ*, **158**, 13
- Pavlov, A., Mller-Nilsson, O., Feldt, M., et al. 2008, in *International Society for Optics and Photonics*, Adv. Softw. Control Astron. II, 7019, 701939
- Pecaut, M. J., & Mamajek, E. E. 2013, *ApJS*, **208**, 9
- Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, *ApJ*, **596**, 702
- Perrot, C., Thebault, P., Lagrange, A.-M., et al. 2019, *A&A*, **626**, A95
- Pueyo, L. 2018, *Handbook of Exoplanets*, 705
- Racine, R., Walker, G. A., Nadeau, D., Doyon, R., & Marois, C. 1999, *PASP*, **111**, 587
- Rameau, J., Chauvin, G., Lagrange, A.-M., et al. 2015, *A&A*, **581**, A80
- Rousseeuw, P. J., & Driessen, K. V. 1999, *Technometrics*, **41**, 212
- Ruffio, J.-B., Macintosh, B., Wang, J. J., et al. 2017, *ApJ*, **842**, 14
- Santos, N. C. 2008, *New Astron. Rev.*, **52**, 154
- Savransky, D. 2015, in *International Society for Optics and Photonics*, Tech. Instrum. Detection Exoplanets VII, 9605, 96050R
- Schneider, J., Dedieu, C., Le Sidaner, P., Savalle, R., & Zolotukhin, I. 2011, *A&A*, **532**, A79
- Schtz, O., Meeus, G., & Sterzik, M. 2005, *A&A*, **431**, 175
- Simon, M. K. 2007, *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists* (Springer Science & Business Media)
- Smith, I., Ferrari, A., & Carbillet, M. 2009, *IEEE Trans. Signal Process.*, **57**, 904
- Soummer, R., Pueyo, L., & Larkin, J. 2012, *ApJ*, **755**, L28
- Stein, C. M. 1981, *Ann. Stat.*, 1135
- Tarantola, A. 2005, *Inverse Problem Theory and Methods for Model Parameter Estimation* (SIAM), 89
- Thibaut, ., Devaney, N., Langlois, M., & Hanley, K. 2016, in *International Society for Optics and Photonics*, SPIE Astron. Telescopes+ Instrum., 99091R
- Thompson, A. M., Brown, J. C., Kay, J. W., & Titterton, D. M. 1991, *IEEE Trans. Pattern Anal. Mach. Intell.*, 326
- Traub, W. A., & Oppenheimer, B. R. 2010, *Exoplanets*, 111
- Vigan, A., Moutou, C., Langlois, M., et al. 2010, *MNRAS*, **407**, 71
- Wagner, K., Apai, D., Kasper, M., et al. 2016, *Science*, **353**, 673
- Wahba, G., et al. 1985, *Ann. Stat.*, **13**, 1378
- Wahhaj, Z., Cieza, L. A., Mawet, D., et al. 2015, *A&A*, **581**, A24
- Ygouf, M. 2012, Ph.D. Thesis, Grenoble
- Yip, K. H., Nikolaou, N., Coronica, P., et al. 2019, ArXiv e-prints [arXiv:1904.06155]
- Zurlo, A., Vigan, A., Mesa, D., et al. 2014, *A&A*, **572**, A85

Appendix A: Estimation of the local statistics of the background in ASDI

In this appendix, we derive the maximum likelihood estimator of the Gaussian parameters. The expression of this estimator is obtained by minimizing the co-log-likelihood \mathcal{L}_n defined in Eq. (2). The first-order optimality condition $\nabla \mathcal{L}_n = \mathbf{0}$ leads to equations defining each parameter.

The condition $\left. \frac{\partial \mathcal{L}_n}{\partial \mathbf{m}_{n,\ell}} \right|_{\mathbf{m}_{n,\ell} = \widehat{\mathbf{m}}_{n,\ell}} = 0$ gives:

$$\mathbf{C}_n^{-1} \sum_{t \in 1:T} \frac{1}{\sigma_{n,\ell,t}^2} (\widehat{\mathbf{m}}_{n,\ell} - \mathbf{r}_{n,\ell,t}) = \mathbf{0}.$$

Since \mathbf{C}_n^{-1} is necessarily non-singular, we obtain the expression of the wavelength-specific average patch:

$$\widehat{\mathbf{m}}_{n,\ell} = \frac{1}{\sum_{t \in 1:T} 1/\sigma_{n,\ell,t}^2} \cdot \sum_{t \in 1:T} \frac{1}{\sigma_{n,\ell,t}^2} \mathbf{r}_{n,\ell,t}, \quad (\text{A.1})$$

which corresponds to a weighted average of the spatial patches, computed over time indices 1 to T , with weights $1/\sigma_{n,\ell,t}^2$ that reduce the impact of frames and spectral channels displaying a large variance $\sigma_{n,\ell,t}^2$.

The condition $\left. \frac{\partial \mathcal{L}_n}{\partial \sigma_{n,\ell,t}^2} \right|_{\sigma_{n,\ell,t}^2 = \widehat{\sigma}_{n,\ell,t}^2} = 0$ leads to:

$$\frac{K}{2\widehat{\sigma}_{n,\ell,t}^2} - \frac{1}{2\widehat{\sigma}_{n,\ell,t}^4} \bar{\mathbf{r}}_{n,\ell,t}^\top \mathbf{C}_n^{-1} \bar{\mathbf{r}}_{n,\ell,t} = 0,$$

with $\bar{\mathbf{r}}_{n,\ell,t} = \mathbf{r}_{n,\ell,t} - \widehat{\mathbf{m}}_{n,\ell}$ the residual patches. This gives:

$$\widehat{\sigma}_{n,\ell,t}^2 = \frac{1}{K} \bar{\mathbf{r}}_{n,\ell,t}^\top \mathbf{C}_n^{-1} \bar{\mathbf{r}}_{n,\ell,t}. \quad (\text{A.2})$$

The time and wavelength-specific scaling factors $\sigma_{n,\ell,t}$ are thus obtained by computing the variance of each spatially whitened patch.

Finally, the condition $\nabla_{\mathbf{C}_n} \mathcal{L}_n \big|_{\mathbf{C}_n = \widehat{\mathbf{S}}_n} = \mathbf{0}$ gives:

$$\frac{TL}{2} \widehat{\mathbf{S}}_k^{-1} - \widehat{\mathbf{S}}_k^{-1} \left(\sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1}{2\sigma_{n,\ell,t}^2} \bar{\mathbf{r}}_{n,\ell,t} \bar{\mathbf{r}}_{n,\ell,t}^\top \right) \widehat{\mathbf{S}}_k^{-1} = \mathbf{0},$$

leading to:

$$\widehat{\mathbf{S}}_k = \frac{1}{TL} \sum_{\substack{\ell \in 1:L \\ t \in 1:T}} \frac{1}{\sigma_{n,\ell,t}^2} \bar{\mathbf{r}}_{n,\ell,t} \bar{\mathbf{r}}_{n,\ell,t}^\top, \quad (\text{A.3})$$

which is the sample covariance matrix of the spatial patches, each rescaled by the corresponding time and wavelength-specific factor.

Appendix B: Derivation of the equivalent number of patches

The equivalent number of patches \widetilde{P} corresponds to the number of samples if all weights are equal and is smaller when some weights differ.

Let us assume that $\{r_t\}_{t \in 1:T}$ be a collection of T independent and identically distributed random variables. The weighted mean $\widehat{m} = \sum_{t=1}^T w'_t r_t$, where $w'_t \geq 0$ are normalized weights ($w'_t = w_t / \sum_{t=1}^T w_t$), is an unbiased estimator of $\mathbb{E}[r]$ with a

variance $\text{Var}[\widehat{m}] = \sum_{t=1}^T \text{Var}[w'_t r_t]$ (by independence of the r_t), which leads to $\text{Var}[\widehat{m}] = \text{Var}[r]/\widetilde{P}$, with

$$\widetilde{P} = 1 / \sum_{t=1}^T w_t'^2 = \left(\sum_{t=1}^T w_t \right)^2 / \left(\sum_{t=1}^T w_t^2 \right), \quad (\text{B.1})$$

the effective number of samples.

If all weights are equal, $\widetilde{P} = T$: the effective number of samples is equal to the total number of samples. If all weights but one are zero, $\widetilde{P} = 1$. In our case, the weights w_t correspond to $1/\widehat{\sigma}_{n,\ell,t}^2$, which leads to the formula to compute \widetilde{P} in Algorithm 1, step 3. In practice, the samples $\{r_t\}_{t \in 1:T}$ are not identically distributed (their variances differ), but \widetilde{P} still indicates if the mean is reliable.

Appendix C: Derivation of the distribution of GLRT⁺

The GLRT⁺ is defined as the sum $\sum_{\ell=1}^L \frac{[\widehat{\alpha}_\ell]_+^2}{v_\ell} = \sum_{\ell=1}^L s_\ell$ with $s_\ell = [\widehat{\alpha}_\ell]_+^2 / v_\ell$. In the absence of source and under the simplifying assumption of an absence of spectral correlation of the backgrounds (i.e., under \mathcal{H}_0 and within our Gaussian model with statistically independent channels) the terms s_ℓ are independent and identically distributed. Due to the thresholding of negative values, the distribution of each s_ℓ corresponds to a mixture of a χ^2 random variable and a Dirac mass at 0:

$$p(s_\ell | \mathcal{H}_0) = \frac{1}{2} \delta_0(s_\ell) + \frac{1}{2} \chi_1^2(s_\ell), \quad (\text{C.1})$$

where the Dirac mass δ_0 centered in 0 accounts for the probability 1/2 that S/N_ℓ be negative and the Chi-square distribution with one degree of freedom χ_1^2 corresponds to the distribution of the square of a standard Gaussian variable. By independence of the s_ℓ , the distribution of their sum, that is to say GLRT⁺, is given by the convolution product:

$$\begin{aligned} p(\text{GLRT}^+ | \mathcal{H}_0) &= \underbrace{(p(s_1 | \mathcal{H}_0) \cdots p(s_L | \mathcal{H}_0))}_{L \text{ times}} (\text{GLRT}^+) \\ &= \frac{1}{2^L} \delta_0(\text{GLRT}^+) \\ &\quad + \sum_{\ell=0}^{L-1} \frac{L!}{2^{\ell+1} \ell! (L-\ell)!} \chi_{L-\ell}^2(\text{GLRT}^+), \end{aligned} \quad (\text{C.2})$$

by application of the binomial expansion and the property $\chi_a^2 * \chi_b^2 = \chi_{a+b}^2$ (the sum of two independent χ^2 random variables with respective degrees of freedom a and b is a χ^2 -distributed random variable with $a+b$ degrees of freedom).

Appendix D: Robust estimation of the spectral correlations

We described in Sect. 3.2.2 how modeling the covariance between the detection maps S/N_ℓ at each wavelength could be used in order to combine the spectral information. In this appendix, we discuss the estimation process of the spectral covariances Σ .

When estimating covariance matrices Σ , we face two difficulties: (i) the estimation must be local in order to capture the nonstationarities of those correlations and (ii) the estimation cannot be performed in the presence of sources (since the spectral correlations would then also depend on the spectrum of the source and on the values of v_ℓ).

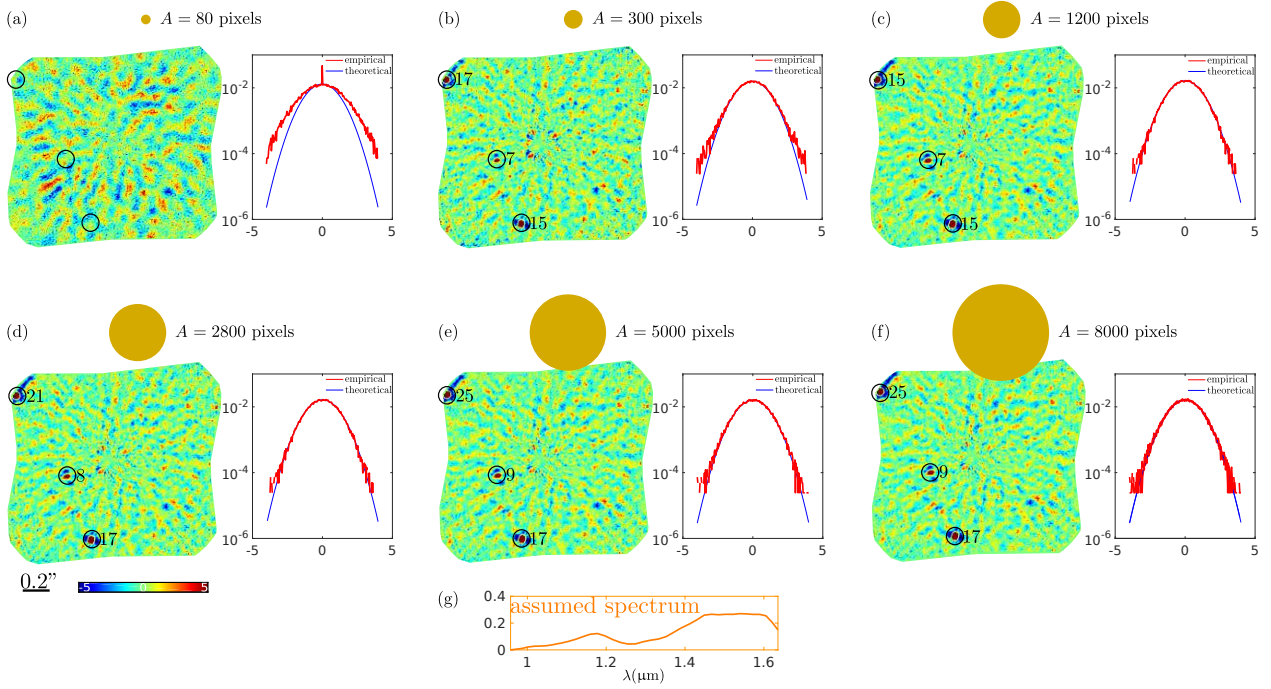


Fig. D.1. Influence of the size A (given in pixels and displayed as a disk at the scale of the field of view) of the region over which the spectral covariances $\widehat{\Sigma}$ are estimated. The S/N_f maps are combined assuming the spectrum shown in (g), for whitening matrices $\widehat{\mathbf{L}}$ obtained from each estimate of the spectral covariance. From *a* to *f*: combined maps $w\widehat{S}/N$ and the empirical distribution of $w\widehat{S}/N$ under \mathcal{H}_0 are shown side by side.

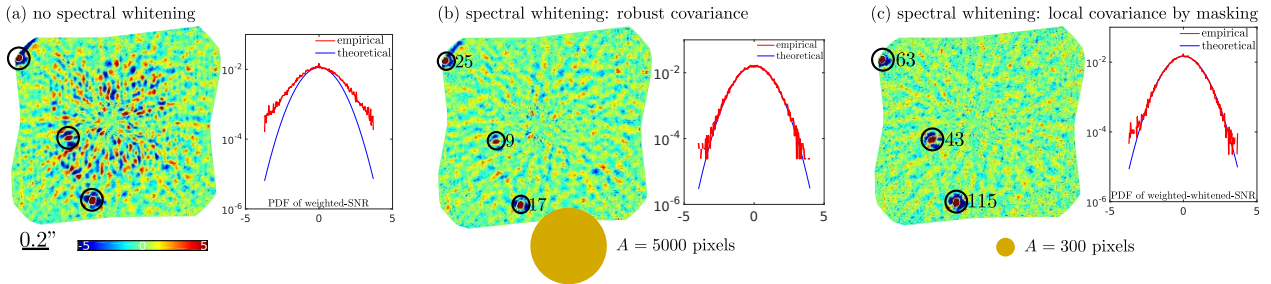


Fig. D.2. Comparison of three spectral whitening strategies: *a*: no whitening; *b*: spectral whitening based on a robust estimate of the covariance computed over a large area ($A = 5000$ pixels, shown at the bottom); *c*: spectral whitening based on a local estimate of the covariance computed over a smaller area ($A = 300$ pixels, shown at the bottom) by masking out regions of high $w\widehat{S}/N$ given by method (*b*).

As described in Sect. 3.2.2, the presence of sources can be circumvented by applying a robust estimator like the MCD to a region large enough so that the portion corresponding to a point source is always marginal. Figures D.1a–f displays the combined detection maps $w\widehat{S}/N$ obtained on HR 8799 when assuming the spectrum plotted in (g). The size of the region over which the robust estimation of Σ is performed is given both in terms of pixels and by a disk at the same scale as the detection map. If the spectral covariances are learned on a region that is too small, as in Fig. D.1a, the detection map is flattened even at the location of the sources. By increasing the size of the region, the robust estimator of the spectral covariance correctly captures the correlations in the absence of sources. However, when the region gets too large, the whitening operation slightly lacks locality.

It can be observed in Fig. D.1 that the empirical distribution in the absence of source correctly matches the expected standard Gaussian model. From a detection map like Fig. D.1e, it is

then possible to detect point sources by thresholding at $\tau = 5$, and a binary mask can be obtained in order to mask the point sources. In a second step, the spectral covariance matrices Σ can be re-estimated on much smaller windows by excluding all pixels that fall in the binary mask. Figure D.2 compares the detection map obtained D.2a without spectral whitening, D.2b with spectral whitening performed after estimating the spectral covariances over a large area with a robust estimator, and D.2c with spectral whitening performed on small areas by computing the sample covariance after exclusion of the pixels around the point sources. This last strategy can be applied to small areas ($A = 300$ pixels), and thus, better eliminate spurious structures in the background. However, it requires a two-step processing: first the computation of the whitened detection map D.2b with the robust covariance estimator, then the formation of the exclusion map, the re-estimation of the covariance matrices and the re-computation of a new detection map.

Appendix E: Optimality of the detection criterion wwS/N

In this part, we show that the optimal linear combination of S/N_ℓ values corresponds to the wwS/N test that we derived from the GLRT. The general form of a test based on a linear combination of (whitened) S/N_ℓ values takes the form:

$$\text{wwS/N} : \sum_{\ell=1}^L w_\ell \cdot [\widehat{\mathbf{L}}^\top \mathbf{x}]_\ell \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau, \quad (\text{E.1})$$

where w_ℓ are weights whose value is to be determined.

Under hypothesis \mathcal{H}_1 , the value of wwS/N is to be maximized, while the variance of wwS/N under \mathcal{H}_0 remains equal to one, so that under \mathcal{H}_0 wwS/N is a standard Gaussian variate and a detection threshold can be straightforwardly set.

Since the vector $\widehat{\mathbf{L}}^\top \mathbf{x}$ follows $\mathcal{N}(\mathbf{0}, \mathbf{I})$ under \mathcal{H}_0 , the variance of wwS/N under \mathcal{H}_0 equals $\sum_{\ell=1}^L w_\ell^2$. The constraint that wwS/N has unit variance leads to the condition $\sum_{\ell=1}^L w_\ell^2 = \|\mathbf{w}\|_2^2 = 1$.

Under \mathcal{H}_1 , the expected value of wwS/N is

$$\mathbb{E}_{\mathcal{H}_1} [\text{wwS/N}] = \sum_{\ell=1}^L w_\ell \cdot [\widehat{\mathbf{L}}^\top \mathbb{E}_{\mathcal{H}_1} [\mathbf{x}]]_\ell, \quad (\text{E.2})$$

where $[\mathbb{E}_{\mathcal{H}_1} [\mathbf{x}]]_\ell = \alpha^{\text{int}} \beta'_\ell$. Equation (E.2) is a scalar product between the vector of weights w_ℓ and the whitened expected S/N_ℓ values. Given the normalization constraint $\|\mathbf{w}\| = 1$, Eq. (E.2) is maximized for a vector of weights that has unit Euclidean norm and is collinear to $\widehat{\mathbf{L}}^\top \mathbb{E}_{\mathcal{H}_1} [\mathbf{x}]$. This leads to the following definition of optimal weights:

$$\mathbf{w} = \frac{\widehat{\mathbf{L}}^\top \boldsymbol{\beta}'}{\|\widehat{\mathbf{L}}^\top \boldsymbol{\beta}'\|}, \quad (\text{E.3})$$

which corresponds to the values of the weights obtained in Eq. (20).

Appendix F: Combination of S/N maps with spectral whitening

Our approach to combine detection maps computed at different wavelengths is general and can also apply to the output of other algorithms, as illustrated in Fig. F.1. We processed the signal-to-noise ratio maps produced by TLOCI and KLIP algorithms. In F.1a, we show the combined signal-to-noise ratio obtained by simple averaging, in F.1b and F.1c we apply a spectral whitening and the prior spectrum of Fig. D.1g according to the definition of wwS/N (we have set $v_\ell = 1$ for both TLOCI and KLIP). In D.1b, the whitening matrix $\widehat{\mathbf{L}}$ is computed from the robust covariance estimator applied on a large area, in D.1c the two-step approach with masking of the point sources in the second step is applied. The detection maps are clearly improved by our spectral whitening scheme. Compared to PACO ASDI, the combined detection maps display a lower value for the 3 sources of the field of view as

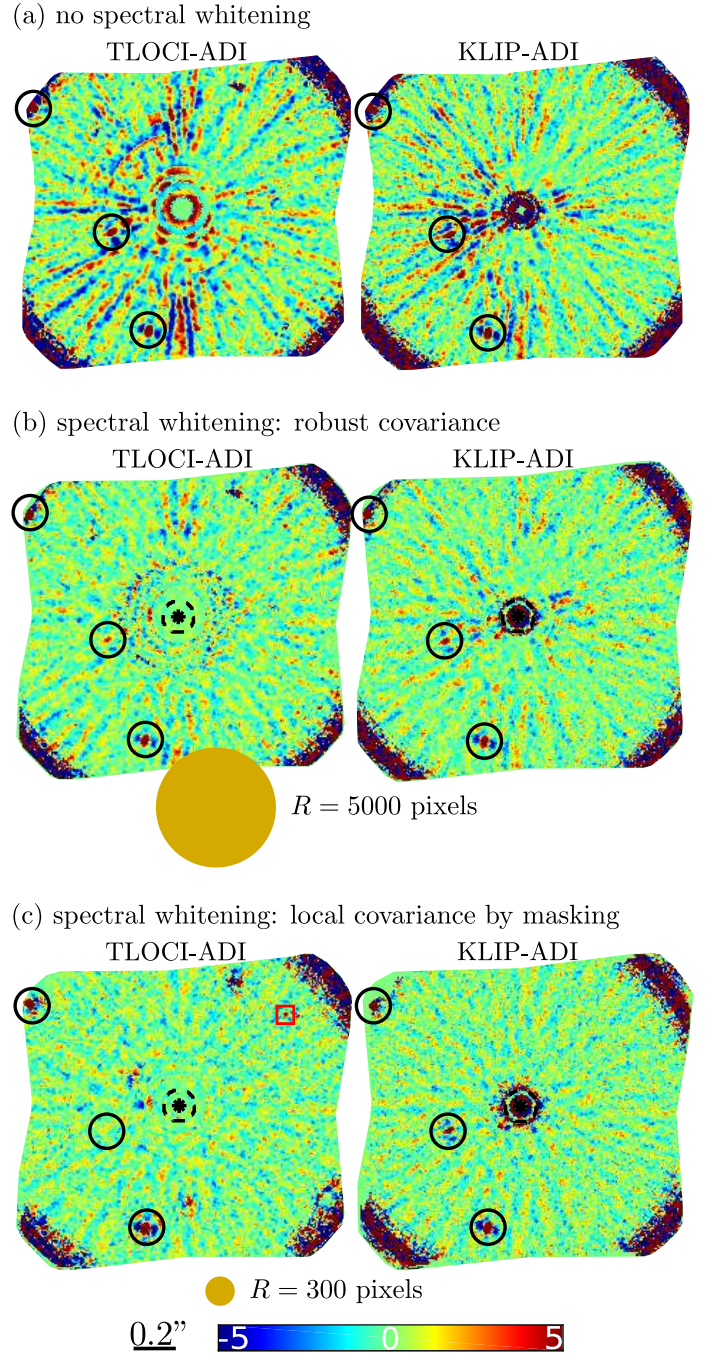


Fig. F.1. Combination of S/N maps with our spectral whitening strategy: *a*: simple spectral averaging of TLOCI and KLIP S/N maps, *b–c*: combination with spectral whitening and the spectrum shown in Fig. D.1g.

well as some border artifacts, which indicates that modeling the nonstationary spatial covariance also plays an important role in PACO ASDI.

Appendix G: Automatic setting of the smoothing parameter μ for spectrum estimation

As detailed in Sect. 4.2, it is useful to enforce a spectral smoothness during the spectrum estimation of a detected point source. The expected gain is a reduction of the MSE of the estimation, in particular when the source contrast is very weak. In this appendix, we numerically compare three well-known regularization strategies of the estimated spectra: GCV (Craven & Wahba 1978; Golub et al. 1979), GML (Wahba 1985; MacKay 1992) and SURE (Stein 1981) (see Sect. 4.2 for a short description). For this purpose, we perform 30 Monte–Carlo injections and spectrum estimations on 4 sources (#2, #5, #10, #12, see Sect. 5.3). Figure G.1 compares the GCV, GML, and SURE regularization strategies in terms of MSE and agreement between the estimates with the empirical 1σ confidence intervals. A comparison is also given with the results obtained when the regularization hyperparameter μ is set in an “oracle” mode that is, by selecting μ that minimizes the MSE between the spectrum estimate and the spectrum ground truth. These experiments illustrate that the GML and SURE approaches lead to very similar results with a slight improvement brought by SURE with respect to GML. In comparison, the GCV leads to significantly worse results. This can be explained by the fact the GCV is generally used when the noise variance is unknown. Regularizing the spectrum estimation with GML or SURE is beneficial since it reduces the MSE. As expected, the gain brought by the regularization is larger when the contrast of the source is weak and for sources located near the host star, that is, when the estimated spectrum is very noisy. For example, the MSE is reduced by about 37% for the source #2 (brightest one of the four considered) while it is reduced up to 67% for the source #10 which is the faintest one. In addition, the results obtained with the automatic setting of the hyperparameter μ by GML or SURE are not very far from the optimal results achieved by the oracle. Finally, as shown by Fig. G.1 (bottom), the estimated confidence intervals are in good match with the empirical ones (a factor between 0.94 and 1.21 is observed in our experiments) when the regularization is performed with the SURE approach.

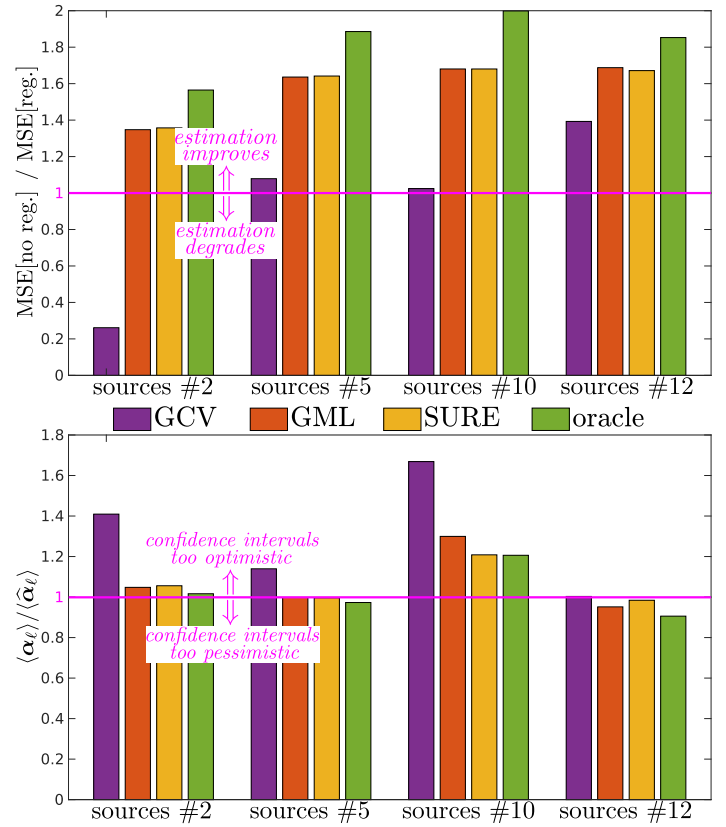


Fig. G.1. Comparison of the GCV, GML, and SURE regularization strategies on 30 Monte–Carlo injections / spectrum estimations for sources #2, #5, #10, and #12. *Top:* gain in terms of MSE reduction comparing with the absence of spectral regularization (values higher than one indicates a decrease of the MSE). *Bottom:* comparison between the empirical 1σ confidence intervals predicted by PACO ASDI and the empirical ones (values higher than one indicate that predicted confidence intervals are smaller than the empirical ones so that the algorithm estimation is too optimistic).