PEX 1. Multispectral expansion of residual speckles for planet detection

Nicholas Devaney, Éric Thiébaut

Monthly Notices
of the
ROYAL ASTRONOMICAL SOCIETY

MNRAS **472,** 3734–3748 (2017)
Advance Access publication 2017 August 30

doi:10.1093/mnras/stx2218

PEX 1. Multispectral expansion of residual speckles for planet detection

Nicholas Devaney[1]★ and Éric Thiébaut[2]

[1]*School of Physics, National University of Ireland, Galway, Ireland*
[2]*Univ Lyon, Univ Lyon1, Ens de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR5574, F-69230, Saint-Genis-Laval, France*

Accepted 2017 August 24. Received 2017 August 23; in original form 2016 November 15

ABSTRACT

The detection of exoplanets in coronographic images is severely limited by residual starlight speckles. Dedicated post-processing can drastically reduce this 'stellar leakage' and thereby increase the faintness of detectable exoplanets. Based on a multispectral series expansion of the diffraction pattern, we derive a multimode model of the residuals which can be exploited to estimate and thus remove the residual speckles in multispectral coronographic images. Compared to other multispectral processing methods, our model is physically grounded and is suitable for use in an (optimal) inverse approach. We demonstrate the ability of our model to correctly estimate the speckles in simulated data and demonstrate that very high contrasts can be achieved. We further apply our method to removing speckle from a real data cube obtained with the SPHERE integral field spectrograph instrument.

Key words: methods: data analysis – techniques: image processing – planets and satellites: detection.

1 INTRODUCTION

Direct detection of extra-solar planet images or spectra is extremely challenging because of the small angular separation and the enormous contrast ratio between the parent star and the planet. For ground-based detection of exoplanets, it is necessary to employ a very high-order adaptive optics (AO) system to reduce atmospheric wavefront errors to a very low level. In addition, a coronograph is employed to mask light from the parent star (Delacroix et al. 2012). Specialized instruments have recently been commissioned to search for exoplanets; SPHERE in the case of the 8 m VLT (Beuzit et al. 2008; Vigan et al. 2015), GPI in the case of the 8 m Gemini telescope (Macintosh et al. 2014; Kalas et al. 2015), SCExAO for the 8 m Subaru telescope (Guyon et al. 2011) and the Project 1640 system at the Palomar observatory (Crepp et al. 2011).

Some exoplanets have already been detected by direct imaging (Marois et al. 2008; Lagrange et al. 2009; Lafrenière, Jayawardhana & van Kerkwijk 2010; Ireland et al. 2011; Rameau et al. 2013; Bonavita et al. 2014), although the separations of the planets from the parent stars are relatively large. It has been found that the detection of exoplanets in AO-corrected images is severely limited by the presence of residual speckles (e.g. Janson et al. 2006). These speckles arise from uncorrected atmospheric errors, and uncorrected optical errors in the telescope and/or imaging instrument. The residual atmospheric errors are random and will average to a smooth halo if the exposure time is long compared with the atmospheric coherence time. However, the speckles due to uncorrected optical errors evolve slowly with time (Hinkley et al. 2007) and are difficult to distinguish from real point sources in the image. In order to reduce these static speckles, a number of differential imaging techniques have been proposed.

In angular differential imaging (ADI), the field can rotate during the observation (e.g. on an altitude–azimuth telescope), while the instrument and AO system are fixed in position; it is supposed that the system point spread function (PSF), including the quasi-static speckles, will remain approximately constant during the observation, while any exoplanets will rotate about the central star. In the LOCI approach which was first proposed by Lafrenière et al. (2007), the image is divided into annular regions defined by annuli and wedges, and the PSF is estimated and subtracted in each region. A temporal weighting of the frames making up the local PSF estimate can therefore vary over the image, reflecting the fact that the temporal evolution of speckles depends on their distance from the centre of the PSF. The direct application of LOCI to IFS data can lead to errors in the exoplanet spectrophotometry due to subtraction of residual starlight. Pueyo et al. (2012) proposed a modified algorithm referred to as '*Damped LOCI*' in which the cost function is modified to attempt to conserve companion flux. Marois et al. (2014) show that the performance can be improved by including a prior model of the planet spectrum – this is referred to as template-LOCI or simply TLOCI.

An alternative approach, sometimes referred to as *KLIP* (Karhunen–Loève Image Projection), to analysing ADI data is based on carrying out a principal component analysis (PCA) of the sequence of images (Soummer, Pueyo & Larkin 2012; Amara & Quanz 2012). The images are concatenated into a large array, and a singular value decomposition (SVD) carried out. This

★ E-mail: nicholas.devaney@nuigalway.ie

identifies orthogonal modes of variation. Each image is subsequently decomposed on to a number of principal components and the result is subtracted from the original images. The resulting residues are de-rotated and summed, following which any planet candidates are identified by thresholding. The number of modes to use is the result of a compromise between complete subtraction of the stellar signal and the removal of planet flux. The trade-off is usually carried out by injecting synthetic planets into the data and determining which number of modes will maximize the planet signal-to-noise ratio (S/N). In addition, the optimal number of modes will depend on the radial distance from the star at which it is required to search for a planet (Meshkat et al. 2014).

Another type of differential imaging relies on obtaining images at more than one wavelength. Spectral differential imaging (SDI, Smith 1987; Racine et al. 1999) uses images obtained (preferably simultaneously) at different wavelengths. If one of the images is taken inside a methane absorption line at 1.6 μm then it can be used as a good PSF estimate, since the signal from any exoplanet would be much weaker, at least if the planet is a gas giant. Marois et al. (2000) developed implementation details for the idea and this approach has been implemented in instruments used to survey stars for methane-rich companions (Marois et al. 2005; Biller et al. 2007). While the technique may be extended to include more wavelengths using double and higher order differencing, it is not straightforward and, in practice, is limited by noise (Marois et al. 2005).

Most of the instruments to be used for exoplanet detection will include an integral field spectrograph (IFS) in order to provide spectral information on any detected planet. It has been realized that the information present in IFS data may be used to reduce the effect of the stellar PSF, including any static speckles. Thatte et al. (2007) describe a technique in which the IFS data cube is radially rescaled to remove the wavelength-dependent scaling of the diffraction pattern. The PSF is then estimated by polynomial fitting along the wavelength axis for each pixel in the rescaled data cube. The polynomial order of the fit is chosen by the user – a higher order will tend to remove any planet signal and, in practice, a low order is used. Subtraction of the PSF estimate will allow detection of faint companions. This approach, which was originally proposed by Sparks & Ford (2002) is referred to as spectral deconvolution (SD). Crepp et al. (2011) describe a modification of the LOCI algorithm for use with IFS data cubes. A PSF reference is built up for each wavelength channel by combining images which are nearby in both wavelength and time (but separated enough in wavelength to ensure the speckles have moved significantly with respect to the diffraction limit). The frames at different wavelengths are scaled and combined by means of least squares. Using data from the Palomar observatory system, they find a small departure from linear scaling with wavelength, and attribute this to out-of-pupil aberrations. The PCA technique can also be applied to IFS data.

Here, we propose a new approach to the detection of exoplanets in multispectral data. Perrin et al. (2003) proposed that the PSF can be expanded in a power series of spatial modes which are functions of the residual phase in the pupil. The modes of odd order are antisymmetric, while the even order modes are symmetric. The dominant modes depend on the Strehl ratio of the corrected wavefront and the distance from the centre of the PSF. We extend this work to consider the spectral dependence of these modes, which turns out to be power laws. We examine the validity of this analysis using a SVD of simulated multispectral coronographic images, and show that using this approach can give excellent speckle suppression in simulated data. The approach has similarities with PCA, but there are some important differences. In particular, our method is based on imaging physics (the power series expansion of the PSF), which leads to insights not available with ad hoc processing. We will show that it is more flexible than approaches based on SVD, including PCA. Finally, it lends itself to an inverse approach for the joint estimation of the residual PSF plus detection of planets. A preliminary version of this work has been presented (Thiébaut et al. 2016). In this paper, we will provide more details about the derivation and the testing of our model and show how to use this model to perform planet detection in multispectral and multitemporal data. We present an example of planet detection using data from the SPHERE IFS, but the full exploitation of our technique in an inverse approach will be presented in a subsequent paper.

2 MODEL OF THE STELLAR LEAKAGE

In order to achieve the best detection sensitivity, it is necessary to reduce residual speckles as much as possible by appropriate post-processing of the images. We propose to tackle the removal of the stellar leakage as an *inverse problem* based on a proper modelling of the on-axis PSF. In order to constrain this model and thus achieve a very good level of '*soft coronography*', we exploit the chromatic behaviour of the speckle pattern.

2.1 Speckle alignment

As described in the Introduction, most speckle removal methods work on image differences where the observed images are obtained at different times and/or wavelengths and are subtracted after an interpolation and a multiplication by a scaling factor of one of the images. The interpolation implements a geometrical transformation of coordinates intended to align the speckles in the two images. Typically, this geometrical transform accounts for translation (to compensate for pointing errors and jitter), rotation (to compensate for pupil rotation during the night) and magnification (to compensate for chromatic geometrical effects due to diffraction). In principle, it is also possible to compensate for other geometrical effects such as distortion. The scaling factor accounts for any variation of the flux received from the star as it is seen in the two images (e.g. due to the variation of transparency or to the star spectrum if the two images are at different wavelengths). The resulting image differences are called ADI or SDI depending on whether the two images come from different exposures (and hence have different rotation angles) or from different spectral channels. The idea is that the residual speckles are cancelled or at least strongly reduced in the differential images compared to the original data. In principle, it is possible to combine images at different times *and* different wavelengths.

Since we want to first focus on speckle removal, we start by considering the case of observing an isolated star with no surrounding sources. In this case, the light distribution in the focal plane of a telescope is simply given by:

$$I(\boldsymbol{r}, \lambda, t) = F(\lambda)\, P(\boldsymbol{r}, \lambda, t) \tag{1}$$

where $\boldsymbol{r}$ is the position in the focal plane, λ is the wavelength, t is the time, $F(\lambda)$ is the spectral energy distribution (SED) of the star and $P(\boldsymbol{r}, \lambda, t)$ is the on-axis[1] PSF. Introducing 2D spatial coordinates

[1] What we call the *on-axis direction* is determined by the direction towards the star not that of the optical axis even though they usually coincide.

$\boldsymbol{s}$ in a reference coordinate system where the speckles are aligned, we can rewrite the on-axis PSF as:

$$P(\boldsymbol{r},\lambda,t) = P_{ref}(\boldsymbol{s},\lambda,t)\left|\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{r}}\right|, \tag{2}$$

where $P_{ref}(\boldsymbol{s},\lambda,t)$ is the distribution of speckles in the reference coordinate system and $|\partial \boldsymbol{s}/\partial \boldsymbol{r}|$ is the absolute value of the determinant of the Jacobian matrix of the coordinate transform $\boldsymbol{r} \mapsto \boldsymbol{s}$ and is needed to insure proper normalization. If there are no losses, the distributions of speckles are both normalized in their respective coordinate systems:

$$\iint P(\boldsymbol{r},\lambda,t)\,\mathrm{d}^2\boldsymbol{r} = \iint P_{ref}(\boldsymbol{s},\lambda,t)\,\mathrm{d}^2\boldsymbol{s} = 1\,. \tag{3}$$

This normalization condition insures that:

$$\iint I(\boldsymbol{r},\lambda,t)\,\mathrm{d}^2\boldsymbol{r} = F(\lambda)\,. \tag{4}$$

In these equations, the term $F(\lambda)$ is the SED of the star *as seen by the instrument*, i.e. it takes into account the transmission by the atmosphere and the instrument which may depend on wavelength. The star SED could be written as $F(\lambda, t)$, if the transmission also depends on time t.

The distribution $P_{ref}(\boldsymbol{s},\lambda,t)$ can also be seen as the PSF in the reference coordinate system for a source in the direction of the star. We will refer to it as the '*reference on-axis PSF*'. The mapping $\boldsymbol{r} \leftrightarrow \boldsymbol{s}$ is a general formalization of the geometrical coordinate transform implemented by existing methods such as ADI or SDI. The key idea of ADI- or SDI-based methods is that the reference on-axis PSF is approximately independent of the wavelength and/or the time and can thus be cancelled by means of image subtraction after interpolation.

2.2 Spatiospectral distribution of the speckles

Following the work of Perrin et al. (2003) who expanded the PSF as an infinite Taylor series with respect to the phase aberration, we propose to derive a model for the distribution of speckles which accounts for chromatic effects. Our intention is to use this model to achieve a better suppression of the stellar speckles.

Since the source is effectively at an infinite distance, the illumination in the image plane results from Fraunhofer diffraction and the on-axis PSF is given by:

$$P(\boldsymbol{r},\lambda,t) = \frac{1}{\rho(\lambda)}\left|\iint a(\boldsymbol{u},\lambda,t)\,\mathrm{e}^{\frac{\mathrm{i}\,2\pi}{f\,\lambda}\langle \boldsymbol{u},\boldsymbol{r}\rangle}\,\mathrm{d}^2\boldsymbol{u}\right|^2 \tag{5}$$

with $\rho(\lambda)$ a normalization factor, $a(\boldsymbol{u},\lambda,t)$ the complex amplitude transmission at the position $\boldsymbol{u}$ of the pupil plane and f the focal length. The expression $\langle \boldsymbol{u},\boldsymbol{r}\rangle$ denotes the usual scalar product of $\boldsymbol{u}$ by $\boldsymbol{r}$. In the conditions considered here, the distribution of the diffracted light depends on the phase aberrations in the pupil plane. Propagation of aberrations arising away from the pupil plane can give rise to a chromatic effect (Marois, Phillion & Macintosh 2006) which is not included in this analysis.

Also neglecting the chromaticism of the refractive index of air, the complex amplitude transmission is given by:

$$a(\boldsymbol{u},\lambda,t) = a_0(\boldsymbol{u})\,\exp\left(\frac{\mathrm{i}\,2\pi}{\lambda}\,\mathrm{d}(\boldsymbol{u},t)\right), \tag{6}$$

where $a_0(\boldsymbol{u})$ is the aberration-free telescope complex amplitude transmission (the so-called *pupil function*) and $\mathrm{d}(\boldsymbol{u},t)$ is an achromatic optical path difference due to the aberrations. The integral in equation (5) is directly related to the spatial (inverse) Fourier transform of the complex amplitude transmitted by the pupil:

$$\hat{a}(\boldsymbol{\omega},\lambda,t) = \iint a(\boldsymbol{u},\lambda,t)\,\mathrm{e}^{\mathrm{i}\,2\pi\,\langle \boldsymbol{u},\boldsymbol{\omega}\rangle}\mathrm{d}^2\boldsymbol{u}\,, \tag{7}$$

which is proportional to the complex amplitude in the focal plane. Introducing this quantity in equation (5) yields:

$$P(\boldsymbol{r},\lambda,t) = \frac{1}{\rho(\lambda)}\left|\hat{a}\left(\frac{\boldsymbol{r}}{f\,\lambda},\lambda,t\right)\right|^2. \tag{8}$$

The factor $\rho(\lambda)$ is such that the normalization in equation (3) holds and therefore:

$$\begin{aligned}\rho(\lambda) &= \iint\left|\hat{a}\left(\frac{\boldsymbol{r}}{f\,\lambda},\lambda,t\right)\right|^2\mathrm{d}^2\boldsymbol{r}\\ &= (f\,\lambda)^2\iint|\hat{a}(\boldsymbol{\omega},\lambda,t)|^2\,\mathrm{d}^2\boldsymbol{\omega}\\ &= (f\,\lambda)^2\iint|a_0(\boldsymbol{u})|^2\,\mathrm{d}^2\boldsymbol{u}\,,\end{aligned} \tag{9}$$

with $\boldsymbol{\omega} = \boldsymbol{r}/(f\,\lambda)$ and where the latter equation follows from Parseval's theorem and from equation (6).

Following Perrin et al. (2003), the exponential term of the complex amplitude transmitted by the pupil can be expanded in an absolutely convergent series in the optical path difference $\mathrm{d}(\boldsymbol{u},t)$:

$$a(\boldsymbol{u},\lambda,t) = a_0(\boldsymbol{u})\sum_{k\geq 0}\frac{1}{k!}\left(\frac{\mathrm{i}\,2\pi}{\lambda}\,\mathrm{d}(\boldsymbol{u},t)\right)^k.$$

At this point, it is convenient to introduce the phase aberration in the pupil at a given reference wavelength λ^{ref}:

$$\phi(\boldsymbol{u},t) = \frac{2\pi}{\lambda^{ref}}\,\mathrm{d}(\boldsymbol{u},t), \tag{10}$$

and to rewrite the expanded complex amplitude as:

$$a(\boldsymbol{u},\lambda,t) = a_0(\boldsymbol{u})\sum_{k\geq 0}\frac{\mathrm{i}^k\,\gamma(\lambda)^k\,\phi(\boldsymbol{u},t)^k}{k!}\,, \tag{11}$$

with:

$$\gamma(\lambda) = \frac{\lambda^{ref}}{\lambda}\,. \tag{12}$$

Then taking the inverse Fourier transform of equation (11) yields:

$$\hat{a}(\boldsymbol{\omega},\lambda,t) = \sum_{k\geq 0}\frac{\mathrm{i}^k\,\gamma(\lambda)^k\,\xi_k(\boldsymbol{\omega},t)}{k!}\,. \tag{13}$$

where:

$$\begin{aligned}\xi_{\mathrm{k}}(\boldsymbol{\omega},t) &= \iint a_0(\boldsymbol{u})\,\phi(\boldsymbol{u},t)^k\,\mathrm{e}^{\mathrm{i}\,2\pi\,\langle \boldsymbol{u},\boldsymbol{\omega}\rangle}\mathrm{d}^2\boldsymbol{u}\\ &= \hat{a}_0(\boldsymbol{\omega})\,\star^k\,\hat{\phi}(\boldsymbol{\omega},t)\end{aligned} \tag{14}$$

with $\hat{a}_0(\boldsymbol{\omega})$ and $\hat{\phi}(\boldsymbol{\omega},t)$, the inverse spatial Fourier transforms of the aberration-free pupil transmission and phase aberration at the reference wavelength. As in Perrin et al. (2003), we use $\star^k$ to denote multiple convolution products over the conjugate position $\boldsymbol{\omega}$:

$$\hat{a}_0\,\star^k\,\hat{\phi} \overset{\mathrm{def}}{=} \begin{cases}\hat{a}_0 & \text{if } \mathrm{k} = 0,\\ \hat{a}_0 \star \overbrace{\hat{\phi}\star\ldots\star\hat{\phi}}^{\mathrm{k\ terms}} & \text{if } \mathrm{k} > 0,\end{cases} \tag{15}$$

with $\star$ the ordinary convolution product. Taking the squared modulus of the complex amplitude in the focal plane and grouping the

terms of the same order with respect to the phase aberrations yields:

$$|\hat{a}(\boldsymbol{\omega},\lambda,t)|^2 = \sum_{k_1\geq 0}\sum_{k_2\geq 0}\frac{\mathrm{i}^{k_1-k_2}\,\gamma(\lambda)^{k_1+k_2}\,\xi_{k_1}(\boldsymbol{\omega},t)\,\xi^*_{k_2}(\boldsymbol{\omega},t)}{k_1!\,k_2!}$$
$$= \sum_{k\geq 0}\gamma(\lambda)^k\,(-\mathrm{i})^k\sum_{k'=0}^{k}\frac{(-1)^{k'}\,\xi_{k'}(\boldsymbol{\omega},t)\,\xi^*_{k-k'}(\boldsymbol{\omega},t)}{k'!\,(k-k')!}\,.$$

Taking the normalization in equation (9) into account, the on-axis PSF can finally be written:

$$\boxed{P(\boldsymbol{r},\lambda,t) = \textstyle\sum_{k\geq 0}\gamma(\lambda)^{k+2}\,p_k(\gamma(\lambda)\,\boldsymbol{r},t),} \tag{16}$$

with $\gamma(\lambda)=\lambda^{ref}/\lambda$ and where the *on-axis PSF modes* $p_\mathrm{k}(\boldsymbol{s},t)$, with $\boldsymbol{s}=\gamma(\lambda)\,\boldsymbol{r}$, are given by:

$$p_\mathrm{k}(\boldsymbol{s},t) = \frac{(-\mathrm{i})^k}{\rho(\lambda^{ref})}\sum_{k'=0}^{k}\frac{(-1)^{k'}\,\xi_{k'}(\boldsymbol{\omega},t)\,\xi^*_{k-k'}(\boldsymbol{\omega},t)}{k'!\,(k-k')!}\,, \tag{17}$$

where:

$$\boldsymbol{\omega} = \frac{\boldsymbol{s}}{f\,\lambda^{ref}} = \frac{\gamma(\lambda)}{f\,\lambda^{ref}}\,\boldsymbol{r}\,.$$

In words, the expansion in equation (16) shows that the change of the PSF with wavelength is a combination of chromatic magnification, by the factor $\gamma(\lambda)$, and amplifications, by powers of $\gamma(\lambda)$. The important point is that there are no other wavelength dependencies in the on-axis PSF. In particular, the PSF modes $p_\mathrm{k}(\boldsymbol{s},t)$ are achromatic; they only depend on the position $\boldsymbol{s}$ in the reference coordinate system and on the time t. The first term p_0 of the PSF expansion is the PSF without aberrations at the reference wavelength, while the other terms are due to the phase aberrations.

Our equations extend the work by Perrin et al. (2003), who considered the monochromatic case. In particular, they did not consider the chromatic magnification and amplification of the PSF modes. Some interesting properties of the on-axis PSF modes $p_\mathrm{k}(\boldsymbol{s},t)$ defined in equation (17) can be inferred from the paper of Perrin et al. (2003). First, all the terms $p_\mathrm{k}(\boldsymbol{s},t)$ of the PSF expansion are real-valued, the zeroth-order term $p_0(\boldsymbol{s},t)$ of this series is the unaberrated PSF and all the odd terms are spatially antisymmetric, while all the even terms are symmetric. Thus, for $k>0$, the terms $p_\mathrm{k}(\boldsymbol{s},t)$ have the same parity as k with respect to $\boldsymbol{s}$: $p_{2\mathrm{k}}(-\boldsymbol{s},t)=p_{2\mathrm{k}}(\boldsymbol{s},t)$, while $p_{2\mathrm{k}+1}(-\boldsymbol{s},t)=-p_{2\mathrm{k}+1}(\boldsymbol{s},t)$. Second, the model derived from the series expansion in the case of simple Fraunhofer diffraction remains approximately valid for an apodized Lyot coronograph. We therefore expect that the on-axis PSF model in equation (16) can serve as a good basis to remove the stellar leakage in the proposed inverse approach. Which terms of the expansion dominate depends on the Strehl ratio, distance from the axis and on the attenuation by the coronograph.

2.3 On-axis PSF in the reference coordinate system

If one simply defines the position in the reference coordinate system as:

$$\boldsymbol{s} = \gamma(\lambda)\,\boldsymbol{r}\,, \tag{18}$$

then noting that it yields:

$$|\partial\boldsymbol{s}/\partial\boldsymbol{r}| = \gamma(\lambda)^2\,, \tag{19}$$

and combining equations (2) and (16), the reference on-axis PSF is written:

$$\boxed{P_{ref}(\boldsymbol{s},\lambda,t) = \textstyle\sum_{k\geq 0}\gamma(\lambda)^k\,p_\mathrm{k}(\boldsymbol{s},t),} \tag{20}$$

with $\gamma(\lambda)=\lambda^{ref}/\lambda$. Since there is no chromatic spatial distortion in this expression, it is clear that the speckles are aligned in such a reference system.

The diffraction computations in the previous section were carried out assuming that the optical axis and the pupil orientation are the same at all wavelengths and times. This implies that a simple chromatic magnification between $\boldsymbol{s}$, the position in the reference coordinate system and $\boldsymbol{r}$, the position in the image coordinate system is sufficient to align the speckles. In practice, there may be several causes of misalignment that must be taken into account and a more complex mapping $\boldsymbol{s}\leftrightarrow\boldsymbol{r}$ may have to be considered. Nevertheless, the series in equation (20) and relation (2) remain valid in the general case and can be exploited to model the chromatic behaviour of the distribution of speckles.

3 VALIDATION OF THE MODEL

To validate our model, we propose to check how well it is able to fit realistic simulations of the stellar leakage. The model becomes separable if one considers resampled images in the reference coordinate system. Working with resampled images, truncated SVD (TSVD) provides an approximation which is different from our model as it makes no other assumptions besides separability. We use TSVD to exhibit the actual chromatic behaviour of the stellar leakage and compare it with the predictions of our model. The ability of our model to fit the stellar leakage is compared to TSVD which provides the best possible fit for a given number of modes.

3.1 Interpolated image in the reference coordinate system

The PSF expansion in equation (16) yields the following expression for the brightness distribution due to the star:

$$I(\boldsymbol{r},\lambda,t) = F(\lambda)\sum_{k\geq 0}\gamma(\lambda)^{2+k}\,p_\mathrm{k}(\boldsymbol{s}(\boldsymbol{r},\lambda,t),t)\,. \tag{21}$$

Considering this expression, it is useful to define:

$$A(\boldsymbol{s},\lambda,t) \equiv I(\boldsymbol{r}(\boldsymbol{s},\lambda,t),\lambda,t)$$
$$= \sum_{k\geq 0} g_\mathrm{k}(\lambda)\,p_k(\boldsymbol{s},t)\,, \tag{22}$$

where $g_\mathrm{k}(\lambda)$ are *spectral weighting functions* given by:

$$g_\mathrm{k}(\lambda) = F(\lambda)\,\gamma(\lambda)^{2+k}\,. \tag{23}$$

The quantity $A(\boldsymbol{s},\lambda,t)$ is the observed light distribution interpolated into the reference coordinate system. This distribution is clearly a separable expansion whose terms are the product of a chromatic weight, $g_\mathrm{k}(\lambda)$, by a spatiotemporal mode, $p_\mathrm{k}(\boldsymbol{s},t)$. We will see that this description is very useful as it allows us to introduce priors on the distribution of the speckles and to help suppress them.

The notations $\boldsymbol{s}(\boldsymbol{r},\lambda,t)$ in equation (21) and $\boldsymbol{r}(\boldsymbol{s},\lambda,t)$ in equation (22) make explicit the spatiospectrotemporal dependency between the image coordinates $\boldsymbol{r}$ and the position $\boldsymbol{s}$ in the reference coordinate system as discussed in Section 2.3. In order to simplify notation, this relationship will be implicitly assumed in what follows.

In practice, the data consist of a number of exposures acquired in different spectral channels. We denote the value measured by the *j*th pixel in the ℓth spectral channel during the *m*th exposure by:

$$I_{\mathrm{j},\ell,\mathrm{m}} \approx I\left(\boldsymbol{r}_{\mathrm{j},\ell,\mathrm{m}}, \lambda_\ell, t_\mathrm{m}\right)\ \Delta\Omega_{\mathrm{j},\ell,\mathrm{m}}\,. \tag{24}$$

Here $\boldsymbol{r}_{\mathrm{j},\ell,\mathrm{m}}$ is the position of the considered pixel, λ_ℓ is the effective wavelength of the spectral channel and t_m is the mean time of the exposure. The term $\Delta\Omega_{\mathrm{j},\,\ell,\,\mathrm{m}}$ accounts for the effective pixel area, spectral bandwidth and exposure duration. The $\approx$ sign in equation (24) accounts for any approximations (such as sampling the distribution instead of integrating it over $\Delta\Omega_{\mathrm{j},\,\ell,\,\mathrm{m}}$) and for the noise.

As described previously, in order to align the residual speckles, it is necessary to resample the data into the reference coordinate system. In practice, this amounts to applying a linear transform to the data. Formally:

$$A(\boldsymbol{s}_\mathrm{i}, \lambda_\ell, t'_\mathrm{i}) \approx A_{\mathrm{i},\ell} = \sum_{j,m} (\boldsymbol{R}_{\ell,\mathrm{m}})_{\mathrm{i,j}}\, I_{\mathrm{j},\ell,\mathrm{m}}\,, \tag{25}$$

where $\boldsymbol{R}_{\ell,\mathrm{m}}$ is an interpolation operator which maps a distribution sampled in the image frame of the ℓth spectral channel during *m*th exposure to the reference coordinate system. As explained in Section 2.3, the mapping to the reference coordinate system depends on the wavelength and on the time, hence the corresponding linear operator is indexed by ℓ and m. As implicitly assumed by equation (25), the resampling operator $\boldsymbol{R}_{\ell,\mathrm{m}}$ also scales the data to eliminate the $\Delta\Omega_{\mathrm{j},\,\ell,\,\mathrm{m}}$ factors. Combining equations (22) and (25), the stellar light in the resampled data is approximated by:

$$A_{\mathrm{i},\ell} \approx \sum_{k\geq 0} p_\mathrm{k}\left(\boldsymbol{s}_\mathrm{i}, t'_\mathrm{i}\right)\ g_\mathrm{k}(\lambda_\ell)\,. \tag{26}$$

In equations (25) and (26), $\boldsymbol{s}_\mathrm{i}$ and t'_i involve a sampling of the position in the reference coordinate system and of the time. Specific time sampling (i.e. not necessarily the same as the exposures) is required to correctly account for the temporal variations of the speckles and to ensure that the approximation assumed in equations (25) and (26) holds.

3.2 Truncated approximation

In practice, the model in equation (26) cannot be used with an unlimited number of modes. Perrin et al. (2003) have shown that the higher order modes are insignificant (the series is absolutely convergent) and that the first modes may also be negligible in case of images obtained using a coronograph. Limiting the series to K modes and assuming k_0 is the index of the first significant term in the series, our so-called PEX (Planet eXtractor) model of the brightness distribution in a reference coordinate system can be written:

$$A_{\mathrm{i},\ell} \approx \sum_{k=1}^{K} p_{\mathrm{k,i}}\, g_{\mathrm{k},\ell} \quad \text{with} \quad \begin{cases} p_{\mathrm{k,i}} = p_{\mathrm{k}+\mathrm{k}_0-1}(\boldsymbol{s}_\mathrm{i}, t'_\mathrm{i}) \\ g_{\mathrm{k},\ell} = g_{\mathrm{k}+\mathrm{k}_0-1}(\lambda_\ell) \end{cases} \tag{27}$$

According to the results of Perrin et al. (2003), we anticipate that $k_0 = 0$ without a coronograph and that $k_0 > 0$ with a coronograph. Using matrix notation, our approximation can be written in the more compact form:

$$\boldsymbol{A} \approx \sum_{k=1}^{K} \boldsymbol{p}_\mathrm{k} \cdot \boldsymbol{g}_\mathrm{k}^\mathrm{t} \tag{28}$$

where $\boldsymbol{p}_\mathrm{k} = (p_{\mathrm{k},1}\ p_{\mathrm{k},2}\ \ldots)^\mathrm{t}$ and $\boldsymbol{g}_\mathrm{k} = (g_{\mathrm{k},1}\ g_{\mathrm{k},2}\ \ldots)^\mathrm{t}$ are the (sampled) significant PSF modes and their spectral weights.

Introducing $F_\ell = F(\lambda_\ell)$ the SED of the star in the ℓth spectral channel, and $\gamma_\ell = \gamma(\lambda_\ell) = \lambda^{ref}/\lambda_\ell$, the specific form of the spectral weights assumed by our model is given by:

$$g_{\mathrm{k},\ell} = F_\ell\, \gamma_\ell^{\beta+k-1}\,, \tag{29}$$

where $\beta = k_0 + 2$ is the exponent for the first significant term of the expansion. With this notation, our approximation of the stellar leakage becomes:

$$A_{\mathrm{i},\ell} \approx F_\ell \sum_{k=1}^{K} \gamma_\ell^{\beta+k-1}\, p_{\mathrm{k,i}}\,. \tag{30}$$

Clearly, if the star SED, $F(\lambda)$ and the exponent β are both unknown, it is not possible to disentangle them from the resampled data alone without ambiguities. We therefore rewrite the model in equation (30) as:

$$A_{\mathrm{i},\ell} \approx q_\ell \sum_{k=1}^{K} \gamma_\ell^{k-1}\, p_{\mathrm{k,i}}\,, \tag{31}$$

where:

$$q_\ell = F_\ell\, \gamma_\ell^{\beta} \tag{32}$$

is the SED of the first significant term of the expansion.

3.3 Fitting the separable model

Assuming independent Gaussian noise for the images in a reference coordinate system, maximum likelihood estimation of the stellar speckles would be achieved by minimizing:

$$\chi^2 = \sum_{i,\ell} w_{\mathrm{i},\ell} \left(A_{\mathrm{i},\ell} - q_\ell \sum_{k=1}^{K} \gamma_\ell^{k-1}\, p_{\mathrm{k,i}}\right)^2, \tag{33}$$

where the statistical weights are given by:

$$w_{\mathrm{i},\ell} = \begin{cases} 0 & \text{if } A_{\mathrm{i},\ell} \text{ is unmeasured;} \\ 1/\,\mathrm{Var}\{A_{\mathrm{i},\ell}\} & \text{otherwise.} \end{cases} \tag{34}$$

Taking into account unmeasured data are an important feature as, after alignment and magnification, the images may have different supports in the considered reference coordinate system. To process the noiseless simulated images considered here, we set the weights to be equal to zero for unseen pixels and otherwise equal to one. Because of the resampling of the images, the values of $A_{\mathrm{i},\ell}$ are certainly correlated and this could be taken into account using non-diagonal statistical weights in the expression of the penalty χ^2. For the sake of simplicity, we consider independent statistics as assumed by equation (33) in the following.

For a given number of terms K in the expansion, the unknowns of the problem are the PSF modes, denoted by $\boldsymbol{p}$, and the SED of the first significant mode, denoted by $\boldsymbol{q}$. Fitting our model therefore amounts to solving the problem:

$$(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{q}}) = \arg\min_{\boldsymbol{p},\boldsymbol{q}} \chi^2\,, \tag{35}$$

where χ^2 is defined in equation (33). Solving this problem turns out to be a very difficult task because the model is bilinear in the parameters even though the penalty χ^2 is quadratic with respect to the model. Finding one of the components ($\boldsymbol{p}$ or $\boldsymbol{q}$) of the model given the other ($\boldsymbol{q}$ or $\boldsymbol{p}$) is comparatively trivial, as it requires to solve a weighted linear least-squares problem. In practice, it should not be too difficult to derive an estimation $\widehat{\boldsymbol{q}}$ of the first mode SED $\boldsymbol{q}$ and solving the difficult problem (35) can be avoided.

3.4 Approximation by a truncated singular value decomposition

Clearly, the model in equation (28) is a separable approximation of the interpolated distribution. The SVD invented by Eckart & Young (1939) and Mirsky (1960) is the perfect tool to extract a separable model from the resampled data. The SVD of $\boldsymbol{A}$ is written:

$$A_{\mathrm{i},\ell} = \sum_{k=1}^{\mathrm{rank}(\boldsymbol{A})} U_{\mathrm{i,k}}\, \sigma_{\mathrm{k}}\, V_{\ell,\mathrm{k}}\,, \tag{36}$$

or, using matrix notation:

$$\boldsymbol{A} = \boldsymbol{U} \cdot \boldsymbol{\Sigma} \cdot \boldsymbol{V}^{\mathrm{t}} = \sum_{k=1}^{\mathrm{rank}(\boldsymbol{A})} \sigma_{\mathrm{k}}\, \boldsymbol{u}_{\mathrm{k}} \cdot \boldsymbol{v}_{\mathrm{k}}^{\mathrm{t}}\,. \tag{37}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices whose kth columns are $\boldsymbol{u}_{\mathrm{k}}$ and $\boldsymbol{v}_{\mathrm{k}}$, the so-called left and right *singular vectors* of $\boldsymbol{A}$, and $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal elements are called the *singular values* of $\boldsymbol{A}$ denoted by σ_{k}. By convention, the singular values are all non-negative and sorted in descending order:

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{\mathrm{rank}(\boldsymbol{A})} > 0\,,$$

and all singular values for $k > \mathrm{rank}(\boldsymbol{A})$ are equal to zero.

According to the Eckart–Young–Mirsky theorem (Eckart & Young 1936; Mirsky 1960), the SVD truncated to the first K singular modes, provides the best approximation of this rank to the original matrix $\boldsymbol{A}$ in a least-squares sense. Thus, no other bilinear (separable) model with K modes can beat the one built from the truncated SVD. Approximating the resampled data by the TSVD is written:

$$\boldsymbol{A} \approx \sum_{k=1}^{K} \sigma_{\mathrm{k}}\, \boldsymbol{u}_{\mathrm{k}} \cdot \boldsymbol{v}_{\mathrm{k}}^{\mathrm{t}}\,, \tag{38}$$

with $K \leq \mathrm{rank}(\boldsymbol{A})$.

Using the SVD to determine a separable model is not new, it is for instance the method of choice to perform the PCA of data. The SVD has however some limitations in our context: (i) it yields the optimal separable decomposition in an ordinary least-squares sense but cannot cope with statistical weights or missing data;[2] (ii) it requires working with the interpolated data $\boldsymbol{A}$; (iii) it does not include any a priori behaviour that can be dictated by the physics and which could be introduced to improve the estimation. The interpolated data are necessarily correlated, while measurements in the raw data may be statistically independent. Using ordinary least squares is suboptimal compared to weighted least squares which can also cope with missing data. Finally, the series expansion based on physical considerations (diffraction) shows that the chromatic weights are fairly well constrained, while such constraints cannot be imposed in an SVD-based analysis.

SVD can however be used to investigate the chromatic behaviour of the distribution of speckles and to provide guidelines to design a more restrictive separable model as well as initial parameters for this model. This model can then be used to fit the data in an inverse approach in order to relieve all the drawbacks of SVD. Besides, since TSVD directly yields the best approximation of this rank, it can serve as a template to evaluate the precision achieved by any other approximation such as that in equation (28).

[2] Even though it is possible to discard some *bad pixels*, but this has to be done for all spectral channels at the same interpolated locations that is by removing some rows of the data matrix $\boldsymbol{A}$.

Due to the coronographic mask in the image plane, the model derived from the series expansion cannot apply everywhere and is certainly wrong in the central region – this is mostly critical for SVD. This is why, unless explicitly stated otherwise, we exclude the central region of the field of view in our subsequent SVD-based analysis.

3.5 Data simulation

In order to check the proposed separable approximation of the coronographic images, we simulated multispectral images with parameters typical of SPHERE (Beuzit et al. 2008): 8.2 m telescope, equipped with a Lyot coronograph with an apodized pupil as described by Carbillet et al. (2011) and a pixel size of 7.4 mas. Pixel integration was taken into account assuming a 100 % fill factor. We considered 21 spectral channels evenly distributed over the H band (1.46–1.84 μm). To account for imperfect wavefront correction, we introduced rather pessimistic phase aberrations of 70 nm rms with the same power spectrum as measured on SPHERE. Our simulations did not include filtering of turbulence-induced phase errors by the AO system.

Typical simulated images are shown in the left and centre panels of Fig. 1. The level of the brightest speckles in these images shows that a contrast of greater than 10^{-3} is needed to detect a planet using one of these coronographic images. Given these simulated images, we resampled the images in the different spectral channels to compensate for the chromatic magnification (see right panel of Fig. 1). Testing the model on the resampled data cube is described in the following sections.

3.6 Behaviour of the most significant mode

Comparing the TSVD factorization in equation (38) with our model in equation (28) yields the following correspondences:

$$\boldsymbol{p}_{\mathrm{k}} \approx \alpha_{\mathrm{k}}\, \boldsymbol{u}_{\mathrm{k}}\,, \tag{39a}$$

$$\boldsymbol{g}_{\mathrm{k}} \approx (\sigma_{\mathrm{k}}/\alpha_{\mathrm{k}})\, \boldsymbol{v}_{\mathrm{k}}\,, \tag{39b}$$

for some arbitrary factors $\alpha_{\mathrm{k}} \neq 0$ which must be introduced because the SVD modes $\boldsymbol{u}_{\mathrm{k}}$ are not normalized in the same way as our PSF modes $\boldsymbol{p}_{\mathrm{k}}$. Our model imposes more constraints than SVD does, and the above relations are therefore unlikely to be matched exactly. It is however interesting to investigate whether TSVD and our model yield similar results in the case of a single-mode approximation.[3] In this case, we expect that:

$$(\boldsymbol{v}_1)_\ell / F_\ell \approx (\alpha_1/\sigma_1)\,(\boldsymbol{g}_1)_\ell / F_\ell \propto \gamma_\ell^{\beta}\,,$$

with $F_\ell = F(\lambda_\ell)$, $\gamma_\ell = \lambda^{ref}/\lambda_\ell$ and where $\beta = k_0 + 2$ is the chromatic exponent associated with the first significant mode. Fig. 2 shows the values of $(\boldsymbol{v}_1)_\ell / F_\ell$ found by the SVD decomposition of the speckles of our simulation interpolated in a reference frame and for angular distance $\theta \geq 0.25$ arcsec. Clearly, a power law (the curve in Fig. 2):

$$(\boldsymbol{v}_1)_\ell / F_\ell \approx \eta\, \gamma_\ell^{\beta}\,, \tag{40}$$

with $\beta \simeq 3.7$ provides a perfect fit of the spectral weights estimated by SVD. This agreement between the spectral weights found by

[3] This would be a clear indication that the physical constraints correctly capture the relevant information.

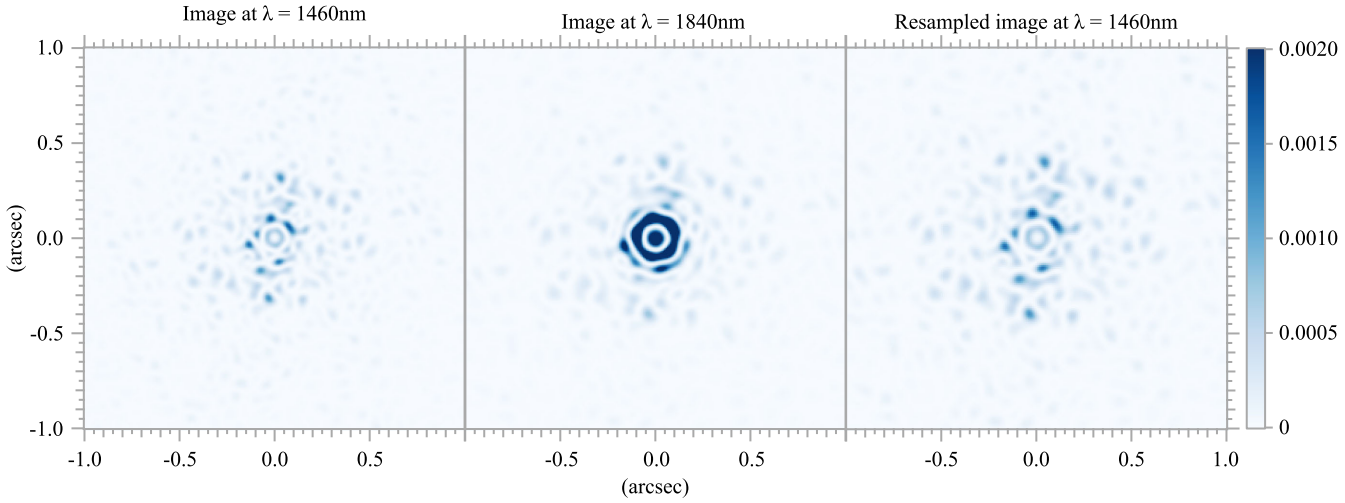

Figure 1. Simulated coronographic images. The leftmost (respectively central) image is for the spectral channel at the shortest (respectively longest) wavelength. The rightmost image is the image at the shortest wavelength magnified to match the diffraction pattern at the longest wavelength. The intensities have been normalized by the peak intensity without a coronograph. The simulations were carried out using parameters typical of SPHERE in the *H* band, see the text for details.

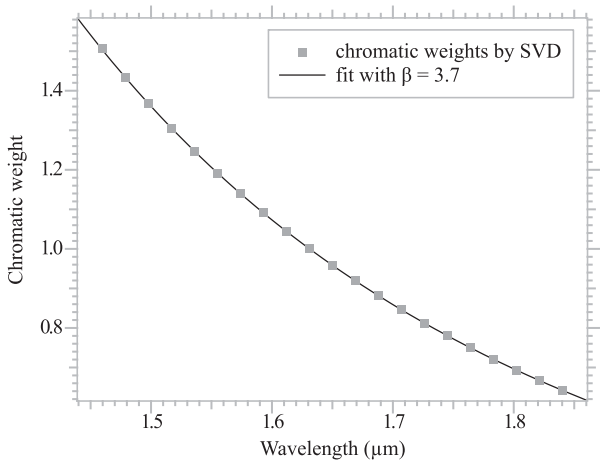

Figure 2. Chromatic weights computed by SVD. The symbols show the chromatic weights for the first mode of the SVD decomposition of the 21 spectral channels of the SPHERE-like simulation for an angular distance $\theta \geq 0.25$ arcsec. The curve is the result of a least-squares fit of a power law $\eta\,(\lambda^{ref}/\lambda)^{\beta}$ with $\beta \simeq 3.7$.

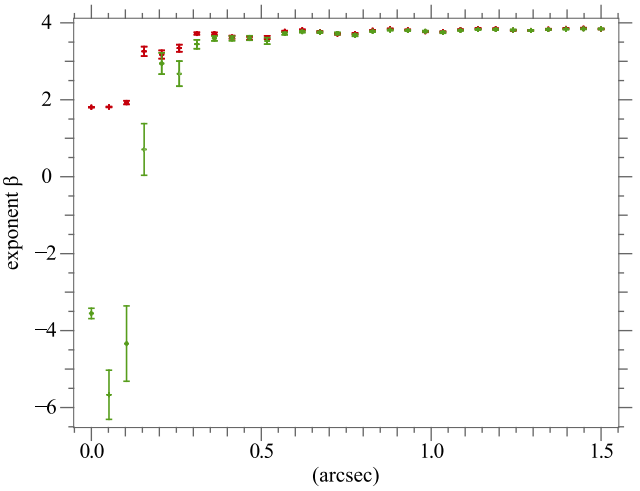

Figure 3. Chromatic exponent β. The spectral weights computed by SVD on annular regions (of thickness 0.05 arcsec) of the rescaled hyperspectral cube have been fitted by a power law $(\lambda^{ref}/\lambda)^{\beta}$. The points show the average values and their error bars computed from 50 SPHERE-like simulations in the *H* band with a 70 nm rms residual aberration. The red symbols are for a simulation with no coronograph and the green symbols are with an apodized Lyot coronograph.

SVD (which makes no specific assumptions about their chromatic behaviour) and the power law induced from the diffraction is a first validation of the proposed chromatic model.

In order to investigate whether the spectral behaviour depends on the angular distance, we extracted narrow annular regions (centred on the star position) from the interpolated cubes and computed the SVD of these data subsets. The chromatic behaviour of the first SVD right singular mode, $\boldsymbol{v}_1$, of these decompositions is then fitted with the power law in equation (40). The exponents β obtained for different angular distances θ from the centre are plotted in Fig. 3 (two cases are considered: with and without a coronograph). Without a coronograph, the exponent is $\beta \simeq 2$ for $\theta \leq 0.1$ arcsec which is exactly what is expected from diffraction in the aberration-free regime. For larger distances, the exponent grows rapidly to a flat level $\beta \simeq 3.7$ due to the aberrations.[4] With a coronograph, the exponent is very different near the centre where it can be as small as $\beta \simeq -6$ depending on the realization of the random aberrations; around the distance $\theta \simeq 0.15$ arcsec, the exponent grows rapidly to reach the same plateau at $\beta \simeq 3.7 \pm 0.1$ as in the case with no coronograph.

The curves in Figs 2 and 3 indicate that the best exponent for the $\theta \geq 0.2$ arcsec region is thus $\beta^{\star} \simeq 3.7 \pm 0.1$ and we observed the same behaviour with various aberration levels (in the range 60–100 nm). Remembering that for the first SVD mode, we should have

[4] We checked that without a coronograph or aberrations, the exponent is $\beta \simeq 2$ everywhere.

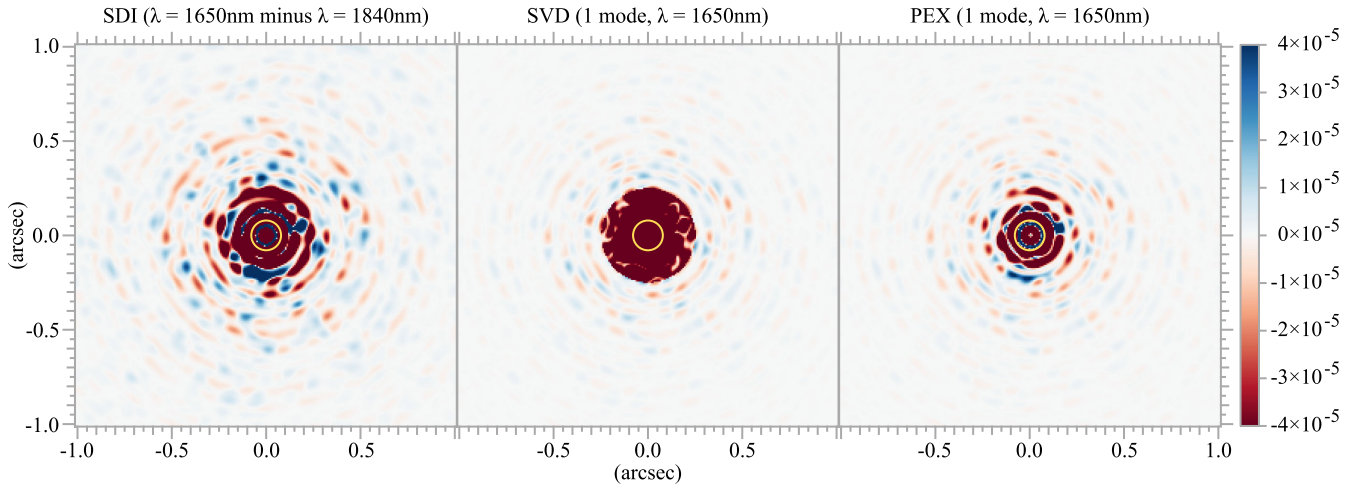

Figure 4. Residuals for a single-mode separable approximation of the multispectral data. The residuals are shown for the $\lambda = 1650$ nm channel and have been normalized by the peak intensity without a coronograph. Left: residuals by SDI; centre: residuals with truncated SVD; right: residuals with spectral weighting set to $(\lambda^{ref}/\lambda)^{3.7}$ and a linear least-squares fit of the spatial mode. The coordinates are in arcseconds (interpolated at the shortest wavelength). The values have been clipped to the range $[-4 \times 10^{-5}, +4 \times 10^{-5}]$ and colour scales are the same for all subfigures: blue for positive residuals, red for negative residuals and levels in relative contrast units. The coronograph mask is indicated by a yellow circle. See the text for details of the simulation.

$\beta^\star = k_0 + 2$ with k_0 the index of the most significant term in the chromatic expansion (22) and noting that $\beta^\star \simeq 3.7$ is close to 4 but is not integer, we deduce that $k_0 = 2$ is the most significant mode in the model given in equation (22) but that other modes are needed to correctly approximate the actual speckle pattern. The most significant mode has an even order, and indeed the most prominent speckles seem to be symmetrically distributed in Fig. 1. The index, $k_0 = 2$, of the first significant mode is an indication of the efficiency of the coronograph.

To support this deduction, we compared the results of the single-mode TSVD approximation which does not implement any specific chromatic behaviour with our model given in equation (30) with $K = 1$. With a single mode and assuming the star SED and the chromatic exponents are known, the solution to minimizing χ^2 defined in equation (33) with respect to $\boldsymbol{p}_1$ is given trivially by:

$$\widehat{p}_{1,\mathrm{i}} = \frac{\sum_\ell w_{\mathrm{i},\ell}\, F_\ell\, \gamma_\ell^\beta\, A_{\mathrm{i},\ell}}{\sum_\ell w_{\mathrm{i},\ell}\, F_\ell^2\, \gamma_\ell^{2\beta}}\,. \tag{41}$$

Figs 4 and 5 show the efficiency of the speckle suppression by different single-mode approximations: SDI, TSVD (with a mask to discard the central $\theta < 0.25$ arcsec region whose chromatic behaviour is very different from the other parts of the field of view) and our model computed according to equation (41). In Fig. 4 we took $\beta = 3.7$, while different values of β are considered in Fig. 5. In order to express the results in terms of the planet/star contrast achieved, the residuals have been normalized by the peak intensity in the image computed under the same conditions but with no coronograph. The profiles plotted in Fig. 5 are the root mean squared (rms) value of the normalized residuals for a given angular distance for all spectral channels (i.e. averaging is carried out for all azimuthal angles and wavelengths). At this point, only the distribution of speckles is considered, there is no noise in the simulations. The curves presented in Fig. 5, therefore really measure the ability of the various approximations to remove the stellar leakage. Fig. 5 shows that the best speckle suppression is achieved when $\beta \to \beta^\star \simeq 3.7$ as predicted from Fig. 3. Taking $\beta = 4$ is nearly as good as with $\beta = \beta^\star$ but any other integer values for β yield significantly worse results. Finally, the level of contrast achieved assuming a

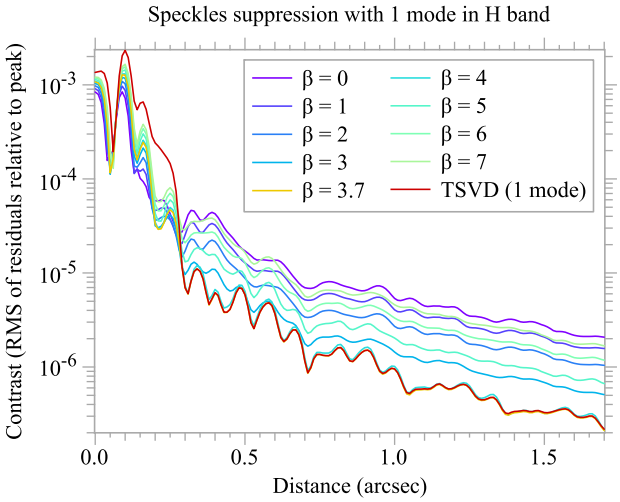

Figure 5. Residuals for a single-mode approximation of the multispectral distribution by truncated SVD (TSVD) or by the proposed model: $(\lambda^{ref}/\lambda)^\beta\, p(\boldsymbol{s})$ and various values of the exponent β. The computations were done for conditions typical of that of SPHERE in the *H* band, see the text for details.

specific chromatic behaviour (with a suitable exponent $\beta^\star$) is as good as what is obtained by TSVD which validates our approach. Fig. 4 clearly shows that the single-mode TSVD and PEX methods yield similar residuals in both distribution and magnitude for angular distances $\theta \geq 0.25$ arcsec where they are better than SDI. In our model, interpolated pixels are fitted independently, thus, contrary to TSVD, no masking of the central region is needed by PEX. Even though the residuals remain important in the central region, PEX is able to reduce the speckles for smaller angular distances than SVD.

As already noted, the speckles in the unprocessed images (Fig. 1) look mostly symmetrical in agreement with the even order of the most significant mode. On the contrary, the residual images in Fig. 4 appear to be nearly antisymmetrical. This is consistent with the diffraction-based expansion: as an even mode ($k_0 = 2$) has been removed, the next most significant mode should be an odd order

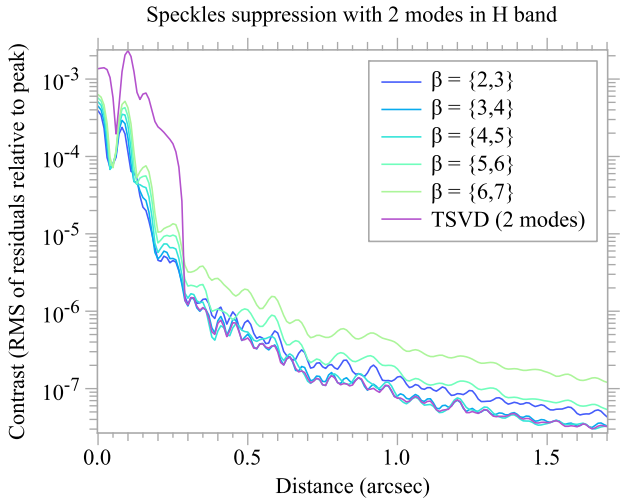

Figure 6. Residuals for a two-mode approximation of the multispectral distribution by truncated SVD (TSVD) or by the proposed model: $\sum_{k=1}^{2}(\lambda^{ref}/\lambda)^{\beta_k}\, p_k(\boldsymbol{s})$. The conditions are identical to those of Fig. 5.

mode at $k_0 \pm 1$ and thus antisymmetrical. We believe that this further supports the model in equation (31) and we examine the performance when more modes are removed.

3.7 Fitting multiple modes

In order to improve the speckle suppression, we now consider fitting more than one mode. As shown by the previous section, applying TSVD (with a single mode) to the resampled images yields a good estimate of the chromatic weights of the first PSF mode. Indeed equation (40) yields:

$$q_\ell = F_\ell\, \gamma_\ell^{\beta} \approx \eta\, (\boldsymbol{v}_1)_\ell\,, \tag{42}$$

where $\boldsymbol{v}_1$ is the first right singular vector of the SVD decomposition of the resampled data and η is an irrelevant normalization factor. Then, as noted before, the remaining unknowns of the problem, namely the PSF modes $\boldsymbol{p}$, are easily found as the result of a weighted linear least-squares fit:

$$\widehat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \sum_{i,\ell} w_{\mathrm{i},\ell} \left(A_{\mathrm{i},\ell} - q_\ell \sum_{k=1}^{K} \gamma_\ell^{k-1}\, p_{\mathrm{k,i}} \right)^2 . \tag{43}$$

Compared to the problem in equation (33), the new problem is trivial to solve. An additional simplification is that the problem is separable with respect to the spatiotemporal index i. In other words, for each spatiotemporal sample $(\boldsymbol{s}_\mathrm{i}, t'_\mathrm{i})$, one has to solve:

$$\{\widehat{p}_{\mathrm{k,i}}\}_{\mathrm{k}=1,\ldots,K} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^K} \sum_{\ell} w_{\mathrm{i},\ell} \left(A_{i,\ell} - q_\ell \sum_{k=1}^{K} \gamma_\ell^{k-1}\, x_\mathrm{k} \right)^2 , \tag{44}$$

which amounts to solving a linear system of only K unknowns (for each index i). This is similar to the *SD* method proposed by Sparks & Ford (2002) except that the chromatic exponents are not the same and that all terms are multiplied by a common SED $\boldsymbol{q}$.

To compare the multimode PeX model with TSVD, we assume again that the stellar SED is known and consider spectral exponents $\beta_\mathrm{k} = k_0 + k - 1$ for different values of the index k_0 of the most significant mode. The estimated PSF modes are then given by:

$$\{\widehat{p}_{\mathrm{k,i}}\}_{\mathrm{k}=1,\ldots,K} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^K} \sum_{\ell} w_{\mathrm{i},\ell} \left(A_{\mathrm{i},\ell} - F_\ell \sum_{k=1}^{K} \gamma_\ell^{\beta_\mathrm{k}}\, x_\mathrm{k} \right)^2 .$$

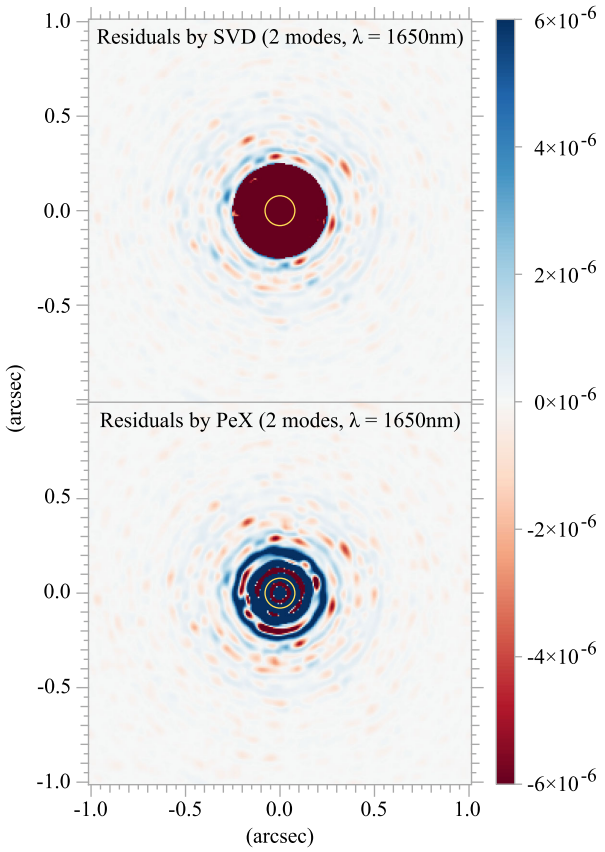

Figure 7. Residuals for a separable approximation of the multispectral data with two modes. This figure is similar to Fig. 4, except that two modes have been used with respective spectral weighting set to $(\lambda^{ref}/\lambda)^{\beta_k}$ with $\boldsymbol{\beta} = \{3, 4\}$ for the linear least-squares fit in the bottom panel.

Fig. 6 shows the rms level of the residuals after subtracting the two-mode models from the simulated images. Compared to TSVD, the PeX model achieves the same efficiency providing the correct spectral exponents are selected. In this case, $\boldsymbol{\beta} = \{3, 4\}$ or $\{4, 5\}$ are the best and correspond, respectively, to $k_0 = 3$ or 4. Compared to the unprocessed coronographic images, the gain is $\sim 10^{-7}$ in contrast; compared to the single-mode models, there is a $\sim$10 factor improvement. Fig. 7 displays the residuals in the $\lambda = 1650$ nm spectral channel. For an angular distance $\theta \geq 0.25$ arcsec, the residuals by TSVD and by the proposed method have almost exactly the same distribution. The central part has been masked for TSVD which therefore performs poorly compared to PeX.

With three modes, Fig. 8 shows that the efficiency of PeX no longer depends on the specific choice of the spectral exponents (even though we only checked for a limited range: $2 \leq k_0 \leq 6$). Except in the central part of the field of view, TSVD performs slightly better than PeX (both with three modes). This can be seen in the images of the residuals shown by Fig. 9. Compared to the two-mode models, the supplementary mode gains a $\sim$3–5 factor in the reduction of the level of the residuals depending on the distance from the centre.

Fig. 10 summarizes the performances of SDI, TSVD and the proposed spectral expansion, where the two latter methods are used with different numbers of modes (from 1 to 6). With more than three modes, we assumed that, as in the three-mode case, the efficiency does not so much depend on the list of exponents and we take $k_0 = 3$ as the index of the first significant mode which

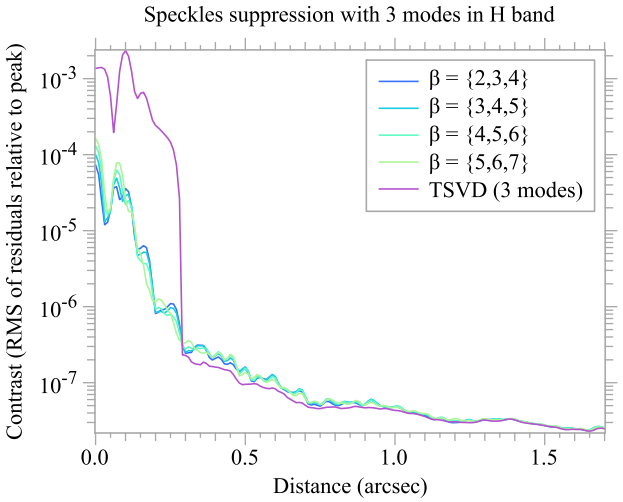

Figure 8. Residuals for a three-mode approximation of the multispectral distribution by truncated SVD (TSVD) or by the proposed model: $\sum_{k=1}^{3}(\lambda^{ref}/\lambda)^{\beta_k}\ p_k(\boldsymbol{s})$. The conditions are identical to those of Fig. 5.

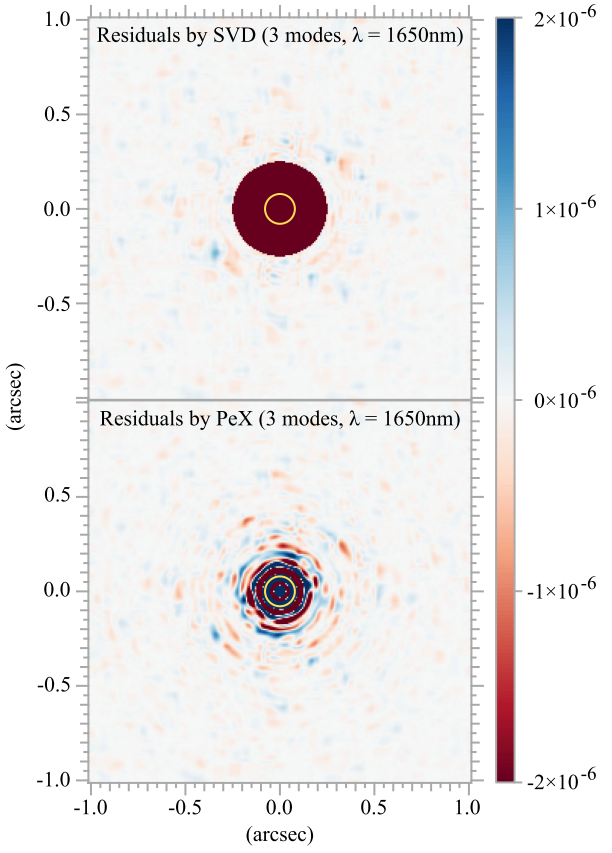

Figure 9. Residuals for a separable approximation of the multispectral data with three modes. This figure is similar to Fig. 4, except that three modes have been used with respective spectral weighting set to $(\lambda^{ref}/\lambda)^{\beta_k}$ with $\boldsymbol{\beta} = \{3, 4, 5\}$ for the linear least-squares fit in the bottom panel.

corresponds to chromatic exponents $\boldsymbol{\beta} = \{3, 4, 5, \ldots\}$. With a given number of modes (one for SDI), the different methods have roughly the same performance, increasing the number of modes improves significantly the achieved contrast: using two modes instead of a single mode improves the detection contrast by more than an order of magnitude. Increasing the number of modes also seems to flatten

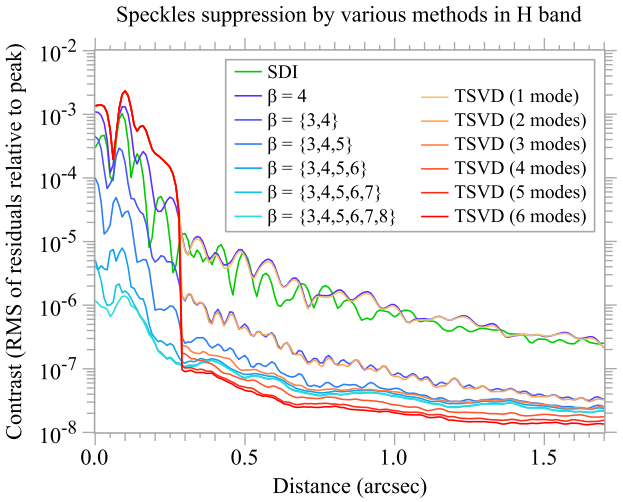

Figure 10. Residuals for different approximations of the multispectral distribution. SDI stands for simple resampled image subtraction; the curves labelled with $\beta = \ldots$ are for our model with 1–6 modes (the values are the chromatic exponents); TSVD is for truncated SVD. See Figs 5–8 and the text for more details.

the level of the residuals as a function of the distance. With six modes, a contrast of $\sim 2 \times 10^{-8}$ is reached at distances larger than 1 arcsec from the centre. Performances are worse near the centre but a contrast as low as 10^{-7} seems to be reachable near the edges of the mask with our method. This is very important for the detection of close companions.

4 APPLICATION TO EXOPLANET DETECTION

Assuming that the planet brightness is negligible compared to that of the stellar speckles and because the planet position does not vary with wavelength, our method for speckle removal should be rather insensitive to the presence of very faint planets. Even though a better approach that we will consider in a following paper would be to *jointly* perform speckle removal and planet detection, it is tempting to perform planet detection in the residuals obtained by subtracting the fitted model of the speckles from the observed images. This task is considered in this section. We first derive a detection test that can be applied to multivariate data (the considered images depend on the wavelength and on the exposure); we then apply this test to simulated data.

4.1 Criterion for detection in multivariate data

Planet detection amounts to deciding between two hypotheses: no planet is present ($\mathscr{H}_0$) or a planet is present ($\mathscr{H}_1$). If there is a planet at position $\boldsymbol{r}$ (hypothesis $\mathscr{H}_1$), the model of the image after removal of speckles is:

$$y_{j,\ell,m} = f_\ell\, h\left(\boldsymbol{r}_{j,\ell,m} - \boldsymbol{r}, \lambda_\ell\right) + \varepsilon_{j,\ell,m}\,, \tag{45}$$

where $y_{j,\ell,m}$ is the value of the residual image at the jth pixel of the ℓth spectral channel and the mth exposure, f_ℓ is the planet flux at wavelength λ_ℓ of the considered spectral channel, $h(\boldsymbol{r}, \lambda)$ is the off-axis PSF at sky position $\boldsymbol{r}$ and wavelength λ, and the term $\varepsilon_{j,\ell,m}$ accounts for the noise (and model errors). Of course, if there are no planets, then $f_\ell = 0$ $(\forall \ell)$ and the residuals are just due to the noise and the model under hypothesis $\mathscr{H}_0$ is just:

$$y_{j,\ell,m} = \varepsilon_{j,\ell,m}\,. \tag{46}$$

Deciding between the two hypotheses can be based on the level of the generalized likelihood ratio (GLR, see e.g. Kay 1998) which is:

$$\mathrm{GLR}(\boldsymbol{r}) = \frac{\max_f \Pr(\boldsymbol{y} \mid \boldsymbol{r}, \boldsymbol{f}, \mathscr{H}_1)}{\Pr(\boldsymbol{y} \mid \mathscr{H}_0)}\,, \tag{47}$$

where $\Pr(\boldsymbol{y} \mid \ldots)$ is the likelihood of the residuals $\boldsymbol{y}$ conditioned by the knowledge of some information or parameters represented by the ellipsis. The higher the GLR, the more likely is a detection and vice versa. For some chosen threshold τ, this is summarized by the notation:

$$\mathrm{GLR}(\boldsymbol{r}) \underset{\mathscr{H}_0}{\overset{\mathscr{H}_1}{\gtrless}} \tau\,,$$

which means that $\mathscr{H}_1$ is decided if $\mathrm{GLR}(\boldsymbol{r}) > \tau$, while $\mathscr{H}_0$ is decided if $\mathrm{GLR}(\boldsymbol{r}) < \tau$.

For Gaussian independent noise, the cologarithm of the GLR is:

$$\Lambda(\boldsymbol{r}) = -\log \mathrm{GLR}(\boldsymbol{r}) = \frac{1}{2}\left\{\sum_{j,\ell,m} w_{\mathrm{j},\ell,\mathrm{m}}\, y_{\mathrm{j},\ell,\mathrm{m}}^2 - \min_f \sum_{j,\ell,m} w_{\mathrm{j},\ell,\mathrm{m}}\,(y_{\mathrm{j},\ell,\mathrm{m}} - f_\ell\, h_{j,\ell,m}(\boldsymbol{r}))^2\right\},$$

where:

$$h_{\mathrm{j},\ell,\mathrm{m}}(\boldsymbol{r}) = h(\boldsymbol{r}_{\mathrm{j},\ell,\mathrm{m}} - \boldsymbol{r}, \lambda_\ell)\,,$$

and $w_{\mathrm{j},\ell,\mathrm{m}} \geq 0$ are statistical weights. Following the reasoning leading to equation (34), the weights are given by:

$$w_{\mathrm{j},\ell,\mathrm{m}} = \begin{cases} 0 & \text{for unmeasured values,} \\ 1/\operatorname{Var}\{y_{\mathrm{j},\ell,\mathrm{m}}\} & \text{otherwise.} \end{cases} \tag{48}$$

Expanding and simplifying $\Lambda(\boldsymbol{r})$ yields:

$$\Lambda(\boldsymbol{r}) = \sum_\ell \min_{f_\ell}\left\{b_\ell(\boldsymbol{r})\, f_\ell - \frac{1}{2}\, a_\ell(\boldsymbol{r})\, f_\ell^2\right\}, \tag{49}$$

with:

$$a_\ell(\boldsymbol{r}) = \sum_{j,m} w_{\mathrm{j},\ell,\mathrm{m}}\, h_{\mathrm{j},\ell,\mathrm{m}}^2(\boldsymbol{r})\,, \tag{50a}$$

$$b_\ell(\boldsymbol{r}) = \sum_{j,m} w_{\mathrm{j},\ell,\mathrm{m}}\, y_{\mathrm{j},\ell,\mathrm{m}}\, h_{\mathrm{j},\ell,\mathrm{m}}(\boldsymbol{r})\,. \tag{50b}$$

The above expression for $\Lambda(\boldsymbol{r})$ shows that obtaining the maximum likelihood estimator (MLE) of the planet SED (assuming the planet position) is simply a matter of solving separable simple quadratic problems for each spectral channel and yields:

$$\widehat{f}_\ell(\boldsymbol{r}) = \arg\min_{f_\ell}\left\{b_\ell(\boldsymbol{r})\, f_\ell - \frac{1}{2}\, a_\ell(\boldsymbol{r})\, f_\ell^2\right\} = \frac{b_\ell(\boldsymbol{r})}{a_\ell(\boldsymbol{r})}\,. \tag{51}$$

Substituting this result in equation (49) gives:

$$\Lambda(\boldsymbol{r}) = \frac{1}{2}\sum_\ell \frac{b_\ell^2(\boldsymbol{r})}{a_\ell(\boldsymbol{r})}\,. \tag{52}$$

The term $a_\ell(\boldsymbol{r})$ does not depend on the data, while the term $b_\ell(\boldsymbol{r})$ does depend on the data. The variance of this latter term can be computed as follows:

$$\begin{aligned}\operatorname{Var}\{b_\ell(\boldsymbol{r})\} &= \sum_{j,m} w_{\mathrm{j},\ell,\mathrm{m}}^2\, \operatorname{Var}\{y_{\mathrm{j},\ell,\mathrm{m}}\}\, h_{\mathrm{j},\ell,\mathrm{m}}^2(\boldsymbol{r}) \\ &= \sum_{j,m} w_{\mathrm{j},\ell,\mathrm{m}}\, h_{\mathrm{j},\ell,\mathrm{m}}^2(\boldsymbol{r}) \\ &= a_\ell(\boldsymbol{r})\,.\end{aligned}$$

Using this result, the MLE of the planet SED is given by:

$$\begin{aligned}\operatorname{Var}\{\widehat{f}_\ell(\boldsymbol{r})\} &= \frac{\operatorname{Var}\{b_\ell(\boldsymbol{r})\}}{a_\ell^2(\boldsymbol{r})} \\ &= \frac{1}{a_\ell(\boldsymbol{r})}\,.\end{aligned} \tag{53}$$

The cologarithm of the GLR can be finally put in the form:

$$\Lambda(\boldsymbol{r}) = \frac{1}{2}\sum_\ell \frac{\widehat{f}_\ell^2(\boldsymbol{r})}{\operatorname{Var}\{\widehat{f}_\ell(\boldsymbol{r})\}}\,, \tag{54}$$

which is a sum of the squared S/N of the MLE of the planet flux in each spectral channel. This is a generalization of a property demonstrated by Mugnier et al. (2009) to the case of multiframe data. We therefore introduce:

$$\mathrm{S/N}(\boldsymbol{r}) = \sqrt{2\,\Lambda(\boldsymbol{r})} = \sqrt{\sum_\ell \frac{b_\ell^2(\boldsymbol{r})}{a_\ell(\boldsymbol{r})}}\,. \tag{55}$$

which can be thought of as a *detection S/N* accounting for all the spectral channels.

Thiébaut & Mugnier (2006) have shown that accounting for additional constraints, notably the positivity and regularity of the SED, can greatly enhance the detection of faint sources. In our case, it is trivial to find the MLE of the planet SED subject to the constraint that it must be non-negative:

$$\begin{aligned}\widehat{f}_\ell^{+}(\boldsymbol{r}) &= \arg\min_{f_\ell \geq 0}\{b_\ell(\boldsymbol{r})\, f_\ell - \frac{1}{2}\, a_\ell(\boldsymbol{r})\, f_\ell^2\} \\ &= \frac{\max\{b_\ell(\boldsymbol{r}), 0\}}{a_\ell(\boldsymbol{r})}\,,\end{aligned} \tag{56}$$

which results from observing that $a_\ell(\boldsymbol{r}) > 0$ and is not more difficult to compute than the unconstrained estimator $\widehat{f}_\ell(\boldsymbol{r})$ in equation (51).

Combining multispectral and multitemporal data to perform planet detection has already been proposed by Thiébaut & Mugnier (2006) for the *Darwin* mission. However, as noted by Denis & Thiébaut (2015) and in our specific case, it turns out that computing $\Lambda(\boldsymbol{r})$ or $\mathrm{S/N}(\boldsymbol{r})$ for any assumed planet position $\boldsymbol{r}$ on an evenly spaced grid of positions $\{\boldsymbol{r}_\mathrm{i}\}_{\mathrm{i}=1,\ldots}$ can be done in a very economic way by means of fast Fourier transforms.

4.2 Application to simulated data

Fig. 11 shows a map of the detection S/N as defined in equation (55). To produce this map, we added planets along spiral tracks to the same simulated speckle patterns as used in the previous section. We then applied the proposed PEX method to estimate and remove the speckles (as if no planets were present) and compute the detection S/N map for all the 21 spectral channels. This map shows that, except under the coronographic mask and near its edges, our method is able to detect all the planets which all have the same contrast of 10^{-6} with respect to the host star. Effectively achieving such a contrast from a single exposure (i.e. without ADI) is very promising. We note that a contrast which is constant with wavelength implies that the planet's SED is the same as that of the host star (something which may occur for planets with a high albedo) and is the most unfavourable situation for the detection. With such a contrast, the planets cannot be seen without processing the original images which look exactly like the ones shown in Fig. 1.

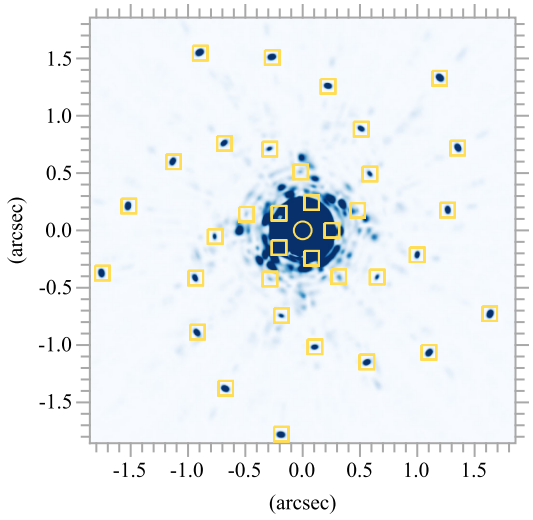

Figure 11. Detection map for planets with a 10^{-6} contrast, removal of speckles by PEX model with three modes and accounting for the positivity. The yellow boxes indicate the positions of planets in the simulation and the circle represents the coronograph mask.

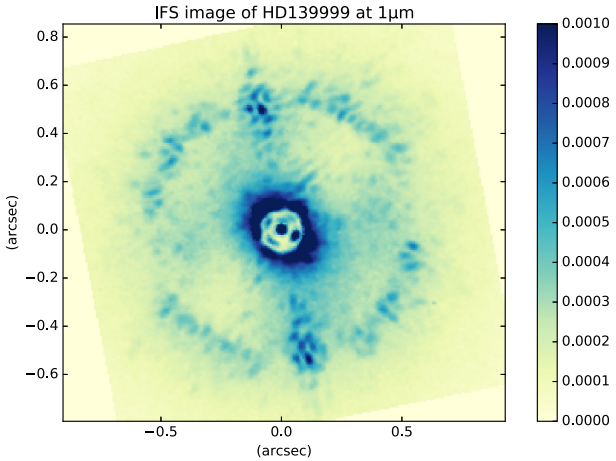

Figure 12. Real IFS image of HD139999 at 1 µm with additional fake planets (which cannot be seen in this image).

4.3 Application to real data

We also consider applying PEX to real SPHERE IFS data (Claudi et al. 2008) of the star HD139999A to which we added 25 fake planets with a contrast of 3×10^{-5}. The IFS image at 1 µm is shown by Fig. 12. As we wanted to demonstrate the ability of our approach to exploit the chromatism of the speckles, we process all available spectral channels[5] but only a *single exposure*. Note that the previously reported companion, HD139999Ab, which is at about 840 mas of HD139999A, is not in the field of the IFS in the considered exposure (Wagner et al. 2016).

[5] 39 spectral channels from $\lambda = 957.5$ to 1 635.8 nm.

Assuming a single planet at position $\boldsymbol{r}$ and using the same notation as in equations (31) and (45), the model of the measured data value in the jth pixel of the ℓth spectral image writes:

$$d_{\mathrm{j},\ell} = \underbrace{q_\ell \sum_i R_{\mathrm{i,j},\ell} \sum_{k=1}^{K} \gamma_\ell^{k-1}\, p_{\mathrm{k,i}}}_{u_{\mathrm{j},\ell}} + f_\ell\, h_{\mathrm{j},\ell}(\boldsymbol{r}) + \varepsilon_{\mathrm{j},\ell}\,, \tag{57}$$

where we drop the exposure index m to simplify the notation and introduce the linear operator $\boldsymbol{R}_\ell$ to interpolate[6] the diffraction-based model of the speckles (expressed in the reference coordinate system) at the positions of the pixels in the ℓth spectral image. Interpolating the model rather than the data avoids introducing more correlations in the data. The pixel size of SPHERE IFS images is 7.46 mas and we choose to sample the on-axis PSF modes $\boldsymbol{p}$ with an equivalent pixel size of 10 mas at the reference wavelength ($\lambda^{ref} = 1$ µm). This sampling size was found to be a good compromise between spatial smoothness of the speckle model and ability to fit the finest details.

The unknowns are $\boldsymbol{q}$, $\boldsymbol{p}$, $\boldsymbol{f}$ and $\boldsymbol{r}$. To follow the procedure described in Section 4.1, we first fit the speckle parameters ($\boldsymbol{q}$ and $\boldsymbol{p}$) on the IFS data (assuming $\boldsymbol{f} = 0$) and then run the detection tests on the residual multispectral images. As we already mention, fitting our model of the stellar speckles is difficult because the model is bilinear. Assuming a Gaussian distribution of the noise, we solve this problem by a hierarchical approach which consists in solving:

$$\widehat{\boldsymbol{p}} = \arg\min_{\boldsymbol{p}} \left\{ \min_{\boldsymbol{q}} \|\boldsymbol{d} - \boldsymbol{u}(\boldsymbol{p},\boldsymbol{q})\|_{\boldsymbol{W}}^2 \right\}, \tag{58}$$

where $\|\boldsymbol{d} - \boldsymbol{u}(\boldsymbol{p},\boldsymbol{q})\|_{\boldsymbol{W}}^2$ is the χ^2 of the data $\boldsymbol{d}$ given the model $\boldsymbol{u}(\boldsymbol{p},\boldsymbol{q})$ of the speckles introduced in equation (57). Here, $\|\boldsymbol{y}\|_{\boldsymbol{W}}^2 = \boldsymbol{y}^{\mathrm{t}} \cdot \boldsymbol{W} \cdot \boldsymbol{y}$ is a weighted quadratic norm and the weight $\boldsymbol{W}$ is the inverse of the noise covariance. Since the χ^2 is quadratic in $\boldsymbol{q}$, the innermost minimization in equation (58) is straightforward. To carry out the outermost minimization, we used a non-linear quasi-Newton method (Nocedal 1980) to optimize over the parameters $\boldsymbol{p}$. We found that, in practice, this hierarchical optimization strategy was very effective.

Compared to the simulations in the previous section, the data are corrupted by noise and a correct estimation of the statistical weights $\boldsymbol{W}$ is very important. As a first simplification, we assumed independent data and thus a diagonal weighting operator $\boldsymbol{W}$ whose components can be computed from the noise variance as in equation (48). Since no estimation of the noise variance is provided with SPHERE IFS data, we estimated this variance by assuming the following simple model (Foi et al. 2008):

$$\mathrm{Var}(\varepsilon_{\mathrm{j},\ell}) = \alpha_\ell\, \mathbb{E}\{d_{\mathrm{j},\ell}\} + \beta_\ell\,, \tag{59}$$

where $\alpha_\ell \geq 0$ and $\beta_\ell > 0$ are two unknown parameters which we assume to be the same for all the pixels of a given spectral channel ℓ. The term $\alpha_\ell\, \mathbb{E}\{d_{\mathrm{j},\ell}\}$ is the variance due to the photon noise, while β_ℓ is the variance of the detector noise. The parameters α_ℓ depend on the quantum efficiency and on the gain of the detector. We assumed that $\mathbb{E}\{\boldsymbol{d}\} \approx \boldsymbol{u}(\boldsymbol{p},\boldsymbol{q})$ and derive the noise model parameters from their maximum likelihood values:

$$(\widehat{\alpha}_\ell, \widehat{\beta}_\ell) = \arg\min_{\alpha,\beta} \sum_j \left\{ \frac{(d_{\mathrm{j},\ell} - u_{\mathrm{j},\ell})^2}{\alpha\, u_{\mathrm{j},\ell} + \beta} + \log(\alpha\, u_{\mathrm{j},\ell} + \beta) \right\}, \tag{60}$$

where the first term in the sum is the χ^2 of the data in the jth pixel of the ℓth spectral channel, while the logarithm term is due to the

[6] We used Catmull–Rom bicubic interpolation.

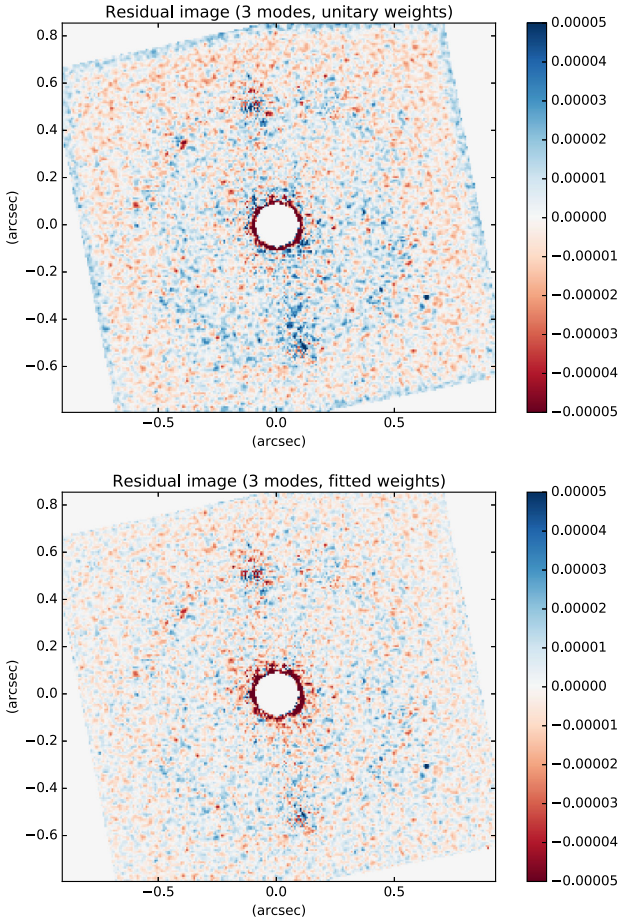

Figure 13. Residual images after removing the PEX model of the speckles. Here, only three modes have been used to model the speckles. The top image shows the residuals assuming uniform weights, while the weights computed from the simple model of the noise variance have been used for the bottom image.

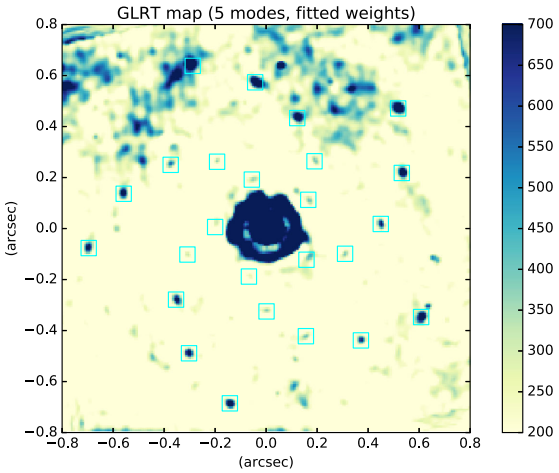

Figure 14. Map of the GLRT criterion. Five modes have been used to model the speckles. The light blue boxes indicate the locations of the fake planets, all have a contrast of 3×10^{-5}.

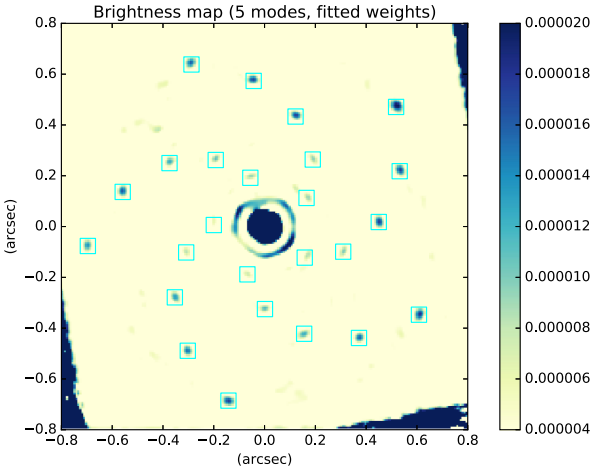

Figure 15. Map of the most likely brightness of point-like sources. Five modes have been used to model the speckles. The light blue boxes indicate the locations of the fake planets, all have a contrast of 3×10^{-5}.

normalization of the assumed Gaussian distribution. We use Powell's BOBYQA algorithm (Powell 2009) to solve the above problem. As the noise parameters depend on the model $\boldsymbol{u}(\boldsymbol{p}, \boldsymbol{q})$ of the speckles which themselves depend on the weights and hence on the noise model parameters, we apply the following alternating procedure: (i) assume uniform weights; (ii) fit the model of the speckles; (iii) fit the noise model parameters and (iv) update the weights and repeat starting at step (ii) until convergence. In practice, this procedure is stable and about 3–5 iterations are sufficient. The comparison of the residuals shown in Fig. 13 demonstrates that non-uniform weights fitted by the proposed alternating method yields smaller and less structured residuals.

Finally, we apply the detection tests described in Section 4.1 to the residuals $\boldsymbol{d} - \boldsymbol{u}(\widehat{\boldsymbol{p}}, \widehat{\boldsymbol{q}})$, Fig. 14 shows the GLR $\Lambda(\boldsymbol{r})$ for the data in the considered exposure. All the fake planets have a GLR which is a local maximum but not all can be detected without false alarms as there are several other positions where the criterion is higher (for instance, in the upper part of the field of view). Looking at equations (52), (50a) and (50b), it is evident that any errors in the magnitude of the statistical weights (which do not appear with the same power in the numerator and denominator of the GLR) could lead to a grossly wrong GLR. The assumed model of the noise variance, in equation (59) is too simple, at least because correlations in the data are ignored. Indeed, due to the way IFS multispectral images are produced, nearby pixels and spectral channels are strongly correlated. To mitigate this issue, we could have compared the GLR to its mean or median value along circular tracks at the same distance from the host star. We however note that the statistical weights appear with the same power in the numerator and denominator of the planet SED given by equation (51) and we therefore expect that errors in the magnitude of the weights somewhat compensate in the estimated SED even though the estimator is no longer optimal. We therefore assumed a constant SED for the sought planets (i.e. $f_\ell = f, \forall \ell$) and compute a map of the *best planet brightness* given its assumed position $\boldsymbol{r}$ which is simply given by:

$$\widehat{f}(\boldsymbol{r}) = \underset{f \geq 0}{\arg\min} \sum_\ell \left\{ b_\ell(\boldsymbol{r})\, f - \frac{1}{2}\, a_\ell(\boldsymbol{r})\, f^2 \right\} = \frac{\max(0, \sum_\ell b_\ell(\boldsymbol{r}))}{\sum_\ell a_\ell(\boldsymbol{r})}\,. \tag{61}$$

In this map, shown by Fig. 15, all the fake planets can be clearly seen with perhaps 1 or 2 false alarms. As a consequence of estimating the

stellar leakage and then the planetary signal (if any), the estimated planet brightnesses in Fig. 15 are always significantly lower than their true value: between $f \approx 2 \times 10^{-5}$ for the most remote planets and $f \approx 6 \times 10^{-6}$ for the ones close to the host star, while the truth is $f = 3 \times 10^{-5}$. A joint estimation of all these unknowns given the data should yield the best results and give an unbiased estimate of the planet SED.

These results on empirical data are very encouraging, notably because they were obtained with a single IFS exposure. To improve the detection limit, multiple exposures could be combined, but it is perhaps more important that the correlations in the data be taken into account. Performing a joint estimation of all the parameters would also be an improvement.

5 DISCUSSION

Our aim in this paper is to enhance the removal of residual speckle in multiwavelength images in order to improve exoplanet detection limits. In order to do this, we have extended the PSF expansion of Perrin et al. (2003) to explicitly take into account wavelength dependence. We show that the PSF may be written as a combination of spatial modes which spatially scale with the wavelength and which are multiplied by chromatic factors with a power law dependence on wavelength and mode number. The exact power law depends on which mode dominates the residuals, and has to be estimated from the data.

If the multiwavelength data are rescaled to a reference wavelength, then the model is separable and is given by an expansion of spatial modes multiplied by a wavelength-dependent factor. We refer to this as the PEX model of the data. The chromatic factor is the product of the SED of the star and a power law with an exponent β.

In order to verify our model, we simulated multiwavelength data (over the *H* band) from a coronographic system with characteristics similar to the SPHERE exoplanet IFS. The data cube is rescaled to 1.65 μm, and SVD used to fit a separable model. In the first instance, the SVD is truncated to fit a single mode and the fit is carried out in annuli centred on the axis. It is found that outside the coronographic mask a power law indeed provides an almost perfect fit to the corresponding chromatic factor. In the absence of aberrations, the fitted power-law exponent implies that the zeroth-order mode dominates, as expected. When aberrations are added to the simulation (70 nm rms), the fitted exponent depends on distance from the axis; but the second-order mode dominates in the wings. This is exactly what is expected from the Perrin et al. analysis. We investigated reducing the speckle residuals by subtracting single modes from the data, using either the SVD modes or fitting a mode obtained using the PEX model. When the PEX model uses the correct chromatic exponent, the performance is very similar to SVD, and in fact can reduce the speckles close to the edge of the coronographic mask.

The SVD fitting is subsequently carried out using multiple modes. It is found that the speckle suppression becomes insensitive to the exact choice of spectral exponents when more than three or four modes are fitted. In the simulated data, the level of suppression reaches 10^{-7} near the edge of the mask using just five modes.

Assuming Gaussian independent noise, which should be suitable for well-cleaned residuals, we derive the MLE for the planet flux at each wavelength, and its variance. The optimal detection criterion then amounts to finding the planet position which has the maximal detection S/N (cf. equation 55). By adding fake planets to the simulated data, we demonstrate detection down to a contrast ratio of 10^{-6} from a single exposure, although some speckle at this level can be seen near the edge of the coronographic mask. With real IFS data, we were able to achieve detection of fake planets from a *single exposure* with a contrast of 3×10^{-5} at 200 mas from the centre. This limit compares favorably to other methods, but is not as good as with simulated data. This is due to the noise in the real data (in our simulations there is no added noise) and to the assumption that pixels are independent (which is not the case with IFS multispectral images). By combining independent exposures and exploiting the apparent motion of the sources in the field of view, we however expect to improve the contrast limit by a factor roughly equal to the square root of the number of exposures.

Compared to other techniques for exoplanet detection in multispectral data, we believe that our approach offers some important advantages. It is based on a physical model of the residual PSF, which provides some insight compared to ad hoc approaches. It is well suited to simultaneous speckle suppression and planet detection, which we are developing for a future publication. In the current application, we fit modes to a data cube made up of the rescaled and interpolated narrow-band images. The fitting could be carried out on the original images by taking the rescaling into account explicitly, thereby removing the need for interpolation which can introduce artefacts. In fact, the inverse approach could be applied to the IFS raw data.

Most current approaches to processing ADI data are empirical and somewhat ad hoc. However, some efforts have been made to develop algorithms which are statistically optimal. For example, Smith, Ferrari & Carbillet (2009) describe a Maximum Likelihood approach to jointly estimate the stellar PSF and the planet position and intensity from the data, while Mugnier et al. (2009) describe a Maximum Likelihood approach to detecting planets in ADI images which have been pairwise subtracted. The analysis presented here can be used to process multitemporal data, taking into account any possible rotation or other transformation of the data as a function of time, as well as temporal correlation of the PSF modes. This work is under development and will be demonstrated in a subsequent paper.

ACKNOWLEDGEMENTS

The authors are very grateful to Maud Langlois for useful discussions to help understand the instrument and for carefully preprocessing the HD139999 data.

This work has made use of the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble), PYTHEAS /LAM/CeSAM (Marseille), OCA/Lagrange (Nice) and Observatoire de Paris/LESIA (Paris).

The research leading to these results has received support from the DETECTION project funded by the French CNRS (Mission pour l'Interdisciplinarité, DEFI IMAGIN) and from the *Programme Avenir Lyon Saint-Étienne Projet Emergent PALSE/2013/26*.

The simulations have been carried out using the YORICK language (Munro 1995), while the empirical data have been processed using the JULIA language (Bezanson et al. 2017).

REFERENCES

Amara A., Quanz S. P., 2012, MNRAS, 427, 948

Beuzit J.-L. et al., 2008, in McLean I. S., Casali M. M., eds, Proc. SPIE Conf. Ser. Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II. SPIE, Bellingham, p. 701418

Bezanson J., Edelman A., Karpinski S., Shah V. B., 2017, SIAM Rev., 59, 65

Biller B. A. et al., 2007, ApJS, 173, 143
Bonavita M., Daemgen S., Desidera S., Jayawardhana R., Janson M., Lafrenière D., 2014, ApJ, 791, L40
Carbillet M. et al., 2011, Exp. Astron., 30, 39
Claudi R. U. et al., 2008, in McLean I. S., Casali M. M., eds, Proc. SPIE Conf. Ser. Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II. SPIE, Bellingham, p. 70143E
Crepp J. R. et al., 2011, ApJ, 729, 132
Delacroix C. et al., 2012, Proc. SPIE, 8446, 84468K
Denis L., Thiébaut É., 2015, Détection robuste et rapide d'un motif connu dans une image, in XXVème Colloque GRETSI. Lyon. p. 374
Eckart C., Young G., 1936, Psychometrika, 1, 211
Eckart C., Young G., 1939, Bull. Am. Math. Soc., 45, 118
Foi A., Trimeche M., Katkovnik V., Egiazarian K., 2008, IEEE Trans. Image Process., 17, 1737
Guyon O., Martinache F., Clergeon C., Russell R., Groff T., Garrel V., 2011, Proc. SPIE, 8149, 814908
Hinkley S. et al., 2007, ApJ, 654, 633
Ireland M. J., Kraus A., Martinache F., Law N., Hillenbrand L. A., 2011, ApJ, 726, 113
Janson M., Brandner W., Henning T., Zinnecker H., 2006, A&A, 453, 609
Kalas P. G. et al., 2015, ApJ, 814, 32
Kay S. M., 1998, Fundamentals of Statistical Signal Processing: Detection Theory. Signal Processing Series Vol. 2, Prentice-Hall, Lebanon, IN
Lafrenière D., Marois C., Doyon R., Nadeau D., Artigau É., 2007, ApJ, 660, 770
Lafrenière D., Jayawardhana R., van Kerkwijk M. H., 2010, ApJ, 719, 497
Lagrange A.-M. et al., 2009, A&A, 493, L21
Macintosh B. et al., 2014, Proc. Natl. Acad. Sci., 111, 12661
Marois C., Doyon R., Racine R., Nadeau D., 2000, PASP, 112, 91
Marois C., Doyon R., Nadeau D., Racine R., Riopel M., Vallée P., Lafrenière D., 2005, PASP, 117, 745
Marois C., Phillion D. W., Macintosh B., 2006, in McLean I. S., Casali M. M., eds, Proc. SPIE Conf. Ser. Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy. SPIE, Bellingham, p. 3
Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafrenière D., Doyon R., 2008, Science, 322, 1348
Marois C., Correia C., Galicher R., Ingraham P., Macintosh B., Currie T., De Rosa R., 2014, Proc. SPIE9148, 91480U
Meshkat T., Kenworthy M. A., Quanz S. P., Amara A., 2014, ApJ, 780, 17
Mirsky L., 1960, Q. J. Math., 11, 50
Mugnier L. M., Cornia A., Sauvage J.-F., Rousset G., Fusco T., Védrenne N., 2009, J. Opt. Soc. Am. A, 26, 1326
Munro D. H., 1995, Comput. Phys., 9, 609
Nocedal J., 1980, Math. Comput., 35, 773
Perrin M. D., Sivaramakrishnan A., Makidon R. B., Oppenheimer B. R., Graham J. R., 2003, ApJ, 596, 702
Powell M., 2009, The BOBYQA algorithm for bound constrained optimization without derivatives, Technical Report NA2009/06, Department of Applied Mathematics and Theoretical Physics, Cambridge, England
Pueyo L. et al., 2012, ApJS, 199, 6
Racine R., Walker G. A. H., Nadeau D., Doyon R., Marois C., 1999, PASP, 111, 587
Rameau J. et al., 2013, ApJ, 772, L15
Smith W. H., 1987, PASP, 99, 1344
Smith I., Ferrari A., Carbillet M., 2009, IEEE Trans. Signal Process., 57, 904
Soummer R., Pueyo L., Larkin J., 2012, ApJ, 755, L28
Sparks W. B., Ford H. C., 2002, ApJ, 579, 543
Thatte N., Abuter R., Tecza M., Nielsen E. L., Clarke F. J., Close L. M., 2007, MNRAS, 378, 1229
Thiébaut É., Mugnier L., 2006, in Aime C., Vakili F., eds, IAU Colloq. 200: Direct Imaging of Exoplanets: Science and Techniques. Cambridge Univ. Press, Cambridge, p. 547
Thiébaut É., Devaney N., Langlois M., Hanley K., 2016, Proc. SPIE, 9909, 99091R
Vigan A., Gry C., Salter G., Mesa D., Homeier D., Moutou C., Allard F., 2015, MNRAS, 454, 129
Wagner K., Apai D., Kasper M., Kratter K., McClure M., Robberto M., Beuzit J.-L., 2016, Science, 353, 673

This paper has been typeset from a TEX/LATEX file prepared by the author.