STIM map: detection map for exoplanets imaging beyond asymptotic Gaussian residual speckle noise

Benoît Pairet [1]★ Faustine Cantalloube,[2] Carlos A. Gomez Gonzalez,[3] Olivier Absil[4] and Laurent Jacques [1]

[1]*ISPGroup, ELEN/ICTEAM, UCLouvain, B-1348 Louvain-la-Neuve, Belgium*
[2]*Max Planck Institute for Astronomy, D-69117 Heidelberg, Germany*
[3]*Université Grenoble Alpes, IPAG, F-38000 Grenoble, France*
[4]*Space Sciences, Technologies, and Astrophysics Research (STAR) Institute, University of Liège, B-4000 Liège Belgium*

ABSTRACT

Direct imaging of exoplanets is a challenging task as it requires to reach a high contrast at very close separation to the star. Today, the main limitation in the high-contrast images is the quasi-static speckles that are created by residual instrumental aberrations. They have the same angular size as planetary companions and are often brighter, hence hindering our capability to detect exoplanets. Dedicated observation strategies and signal processing techniques are necessary to disentangle these speckles from planetary signals. The output of these methods is a detection map in which the value of each pixel is related to a probability of presence of a planetary signal. The detection map found in the literature relies on the assumption that the residual noise is Gaussian. However, this is known to lead to higher false positive rates, especially close to the star. In this paper, we re-visit the notion of detection map by analysing the speckle noise distribution, namely the Modified Rician distribution. We use non-asymptotic analysis of the sum of random variables to show that the tail of the distribution of the residual noise decays as an exponential distribution, hence explaining the high false detection rate obtained with the Gaussian assumption. From this analysis, we introduce a novel time domain detection map and we demonstrate its capabilities and the relevance of our approach through experiments on real data. We also provide an empirical rule to determine detection threshold providing a good trade-off between true positive and false positive rates for exoplanet detection.

Key words: methods: statistical – techniques: high angular resolution – techniques: image processing – planets and satellites: detection.

1 INTRODUCTION

In the field of exoplanet study, high contrast imaging (HCI) provides valuable information to study planetary systems properties since it gives access to the spectral features of the exoplanet's atmosphere (Konopacky et al. 2013; Crossfield 2015), its mass determination via orbital follow-up (Bonnefoy et al. 2014; Pueyo et al. 2015), and the study of its interactions with its environment such as other planets or circumstellar discs (Espaillat et al. 2014; Hughes, Duchene & Matthews 2018). This information can constrain planetary system formation models and improve our understanding of the nature of exoplanets (see Bowler & Nielsen 2018, for a review).

Only a few tens of exoplanets have been directly detected around the hundreds of stars observed within surveys led during the last decade (Chauvin et al. 2015). This low number of detections tells us that either the types of planets accessible through direct imaging are indeed rare or that our sample is strongly biased by our technical limitations. The main challenge of exoplanet imaging is that exoplanets are faint objects located in close vicinity to their host star that is much brighter. Emitted light from young Jupiter-like planets are typically 10^{-6}–10^{-4} fainter than their host star in the near infrared, where the contrast is favourable, and the typical projected separation between the planet and its host star is of 0.1 arcsec.

To reach such an angular resolution, 10 m class ground-based telescopes are used in combination with adaptive optics (AO) systems which correct for the resolution loss induced by the atmospheric turbulence (Guyon 2005). Coronagraph devices

★ E-mail: benoit.pairet@uclouvain.be

are then used to increase the dynamic range by removing the coherent part of the starlight which is hiding the faint circumstellar signals (Guyon et al. 2006). With dedicated instruments such as VLT/SPHERE (Beuzit et al. 2008), Gemini/GPI (Macintosh et al. 2008), KecK/NIRC2 (McLean & Chaffee 2000), Subaru/SCeXAO (Jovanovic et al. 2015), or LBTI/LMIRCam (Hinz et al. 2016), a typical contrast of 10^{-4} is obtained at a separation of 0.5 arcsec. However, residual aberrations induce the presence of speckles in the images, which are of the same angular size as a point source and are often brighter than the exoplanets signal. These speckles are quasi-statics and hence cannot be calibrated or averaged through longer exposures. Post-processing techniques are then applied to disentangle the planetary signals from these starlight residuals and thus reach a contrast of down to 10^{-6} at 0.5 arcsec (Mawet et al. 2012).

To disentangle signals, post-processing methods require diversity within the data. The diversity is obtained through specific observations strategies (such as pupil tracking, dual band imaging or dual polarization imaging). The most widely used methods today rely on angular differential imaging (ADI; Marois et al. 2006) which makes use of pupil tracking mode observations. This mode keeps the speckle field almost constant during the observation while any circumstellar signal rotates with a deterministic velocity given by the parallactic angles. Most methods today consist in empirically estimating the speckle field, then subtracting it from each frame of the image cube, and then combining the frames to form the so-called *processed frame*. In this processed frame the residual speckles add up incoherently and thus average to small values whereas planetary signals are aligned and average to the actual values of the intensity of the planet.

Once this processed frame is computed, one has to perform the detection, which is usually visually performed by the user. A robust detection requires the use of statistical tests from which a detection map is built. In this context, a good knowledge of the underlying distribution of the residual noise is crucial. As discussed further below, current estimation techniques rely on asymptotic analysis of random variables that predicts a Gaussian distribution. As it is known that a Gaussian based detection leads to high false positive rate (FPR), it has been proposed to modify the detection procedure to take into account the deviation from Gaussianity. However, to the best of the knowledge of the authors, no study has yet proposed a method to measure how far off the residuals are from a Gaussian distribution.

In this work, we argue that the tail of the distribution is the important feature as it will decide the outcome of the statistical test. We empirically and theoretically discuss the decay of the tail of the distribution of the residual noise after post-processing. This analysis is non-asymptotic and the resulting estimate of the residual noise level depends on the number of frames. Then we introduce and study a new detection map that is built in light of our results. Although our detection map is designed for an ADI sequence of images, we stress that the non-asymptotic analysis of the speckle noise distribution does not depend on the observation strategy and can be applied to any images where speckle noise is present.

1.1 Contributions

This work brings several contributions, summarized below.

First, we study the tail decay of the residual noise distribution on the processed frame for several datasets. We show numerically that in each case the (empirical) quantiles of the noise distribution are closer to the quantiles of a Laplacian distribution than to those of a Gaussian distribution. This indicates that the residual noise exhibits the exponential tail decay of a Laplacian noise.

Secondly, we prove theoretically that the MR distribution is actually sub-exponential (Vershynin 2010), i.e. it belongs to a class of distributions – including the Laplacian distribution – whose tails decay exponentially. This allows us to use non-asymptotic statistical tools, based on measure concentration, to determine a meaningful estimate of the residual noise level on the processed frame in function of the number of frames. Moreover, we study the sensitivity of our analysis with respect to the fraction of frames that can be considered as statistically independent. In particular, we quantify how the temporal dependence between frames slows down the tail decay of the residual noise. We note that all these observations could not be reached with classical asymptotic analyses, e.g. relying on the central limit theorem (CLT).

Thirdly, we leverage this non-asymptotic analysis to introduce a novel detection map, the standardized trajectory intensity mean map (or STIM map). Our theoretical analysis explains the distribution of pixel intensities in this new map, for which exoplanets are associated with clear outliers in a hypothesis testing context. Comparatively to the SNR_t map (Mawet et al. 2014), we observe that the STIM map distribution is also more concentrated where there is no exoplanet signal; exoplanet detection is thus made more stable with a single thresholding procedure.

Finally, by establishing an automatic estimation of a reliable detection threshold, we demonstrate the capabilities of the STIM map through numerous experiments involving real datasets.

1.2 Paper structure

The rest of the paper is organized as follows. We present the state-of-the-art of the reduction techniques and of the detection procedure, our notations, and our conventions in Section 2. In Section 3, after a brief presentation of the speckle noise distribution, we study the statistical properties of the residual speckle noise both empirically and theoretically, through a rigorous non-asymptotic statistical analysis. Then, we present how this analysis applies to the tail decay of the residual noise on the processed frame. We finally discuss the impact of non-independence of the residual speckle noise. In Section 4, we review the current detection map procedure and we use our previous results to introduce and justify our new detection map, the STIM map. In Section 5, we demonstrate the efficacy of our approach by computing the STIM map for several on-sky data from different instruments. We finally conclude and give perspectives for further applications of this work in Section 6. The information about the datasets used throughout the paper and the mathematical developments can be found in the appendices.

2 FRAMEWORK AND PRELIMINARIES

2.1 Current ADI-based post-processing techniques

An ADI dataset is a volume with T images (or frames) of size $n \times n$ that is reshaped into $\mathbb{R}^{T \times N}$ matrices with $N = n^2$. In the same fashion, $n \times n$ images are depicted as $\mathbb{R}^N$ vectors. We denote matrices by capital bold symbols, e.g. $\boldsymbol{Y}$, and vectors by lowercase bold symbols, e.g. $\boldsymbol{f}$.

ADI-based post-processing methods can be summarized in three steps described in Fig. 1: *(i)* a point-spread function (PSF) model[1]
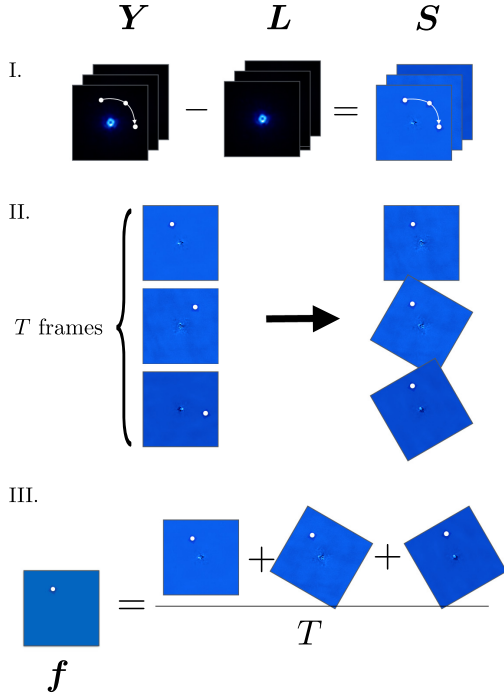
Figure 1. Schematic representation of the ADI method. Y is the data cube, L is the model PSF, S the subtracted data cube, and T the number of frames. f is the processed frame.

L, containing only the host star signal with neither planetary companions nor circumstellar discs signals, is estimated empirically from the data cube Y, i.e. $L = \Theta(Y)$ for some function Θ discussed below, *(ii)* this PSF model is subtracted from the cube to form the volume $S = Y - L$, and *(iii)* the individual frames of S are aligned to a common direction for the potential companions and collapsed into a processed frame f. The temporal correlation of S is known to be lower than that of the initial volume, resulting in residual speckles being considerably less correlated. Thus in step *(iii)*, the residual speckles average to a mean close to zero while the planetary signals add-up. In practice, part of the planetary signal is present in the model PSF L and the intensity of the planetary signals extracted from the resulting processed frame f is an underestimate of their true intensity.

The processed frame f computed in step *(iii)* of the ADI-based post-processing is often computed as the pixelwise median of the aligned frames of S. In order to ease the theoretical analysis, we use in this study the mean instead of the median. Moreover, our methods are conveniently described according to the following mathematical conventions. When referring to a certain pixel in an image, we use a single index g instead of a tuple (i, j) such that an element of the processed frame is depicted as f_g. Instead of defining an

[1]We note that, in HCI, the PSF has a slightly different meaning than in other fields such as signal processing. In the context of HCI for exoplanet detection, the PSF refers to the response of the whole observation system, from the star to the detector, going through the atmosphere, the telescope pupil, the AO, the coronagraph and any optical device in the light path. Hence here, the PSF model refers to the approximation of the whole starlight signal, accounting for the speckle field. Planetary signals have a different response because the coronagraph is designed to act mainly (if not only) on on-axis signals.

aligned volume as in step *(iii)* above, we first collect in a vector $s^{[g]} = (s_1^{[g]}, \cdots, s_T^{[g]})^\top \in \mathbb{R}^T$ the values of the pixels of S that are part of the trajectory induced by the parallactic angles starting from the first line of the matrix S at index g. In this context, the gth element of the processed frame is the computed mean of the trajectory g in the volume S:

$$f_g = \hat{\mu}_g \equiv \hat{\mu}(s^{[g]}) = \frac{1}{T} \sum_{i=1}^{T} s_i^{[g]}. \tag{1}$$

This description of the post-processing in terms of trajectories is also illustrated in Fig. 5 in the context of the detection maps.

A wide variety of methods exists to construct the model PSF L (that is to say different ways to define the function Θ), such as c-ADI (classical ADI; Marois et al. 2006, using the median of the data cube), LOCI (Locally Optimized Combination of Images; Lafrenière et al. 2007, using a linear combination of patches of the images), or PCA (Principal Component Analysis; Soummer, Pueyo & Larkin 2012; Amara & Quanz 2012, using the first principal components of the data cube). Other methods based on ADI perform different steps to obtain the processed frame f, such as LLSG (Gonzalez et al. 2016, that separates the volume in a low-rank part for the star PSF plus a sparse part for the planetary signal). We refer to this class of method as the speckle subtraction methods.

There exists another class of post-processing techniques based on the inverse problem approach that perform a maximum likelihood estimation (equivalent to matched filtering under the Gaussian hypothesis) of the companion flux and position, producing directly a detection map. The techniques, pioneered by ANDROMEDA (Cantalloube et al. 2015), have been extended in different fashions, such as multispectral data (see for instance the FMMF technique, Ruffio et al. 2017).

These two classes of post-processing methods are complementary in exoplanets detection. They are often used together, as it is illustrated, for instance, in Delorme et al. (2017a).

The aim of the detection map that we introduce in Section 4.2 is to provide a more robust detection procedure for speckle subtraction methods. For this reason, we do not consider the second class of methods in this paper. However, it is worth mentioning that our theoretical findings have implications in the maximum likelihood estimation on which they rely. This is briefly discussed in the conclusion.

2.2 Detection procedure

We want to test the absence or the presence of a planet for each pixel g on the processed frame f. In this context, we define on the trajectory supported by g the null-hypothesis H_0 as the absence of a planet and the research hypothesis H_1 as the presence of a planet. Mathematically,

$$H_0: \quad f_g = I_{\text{noise}}, \tag{2}$$

$$H_1: \quad f_g = I_{\text{planet}} + I_{\text{noise}}, \tag{3}$$

where I_{planet} is the intensity of a hypothetical planet on location g and I_{noise} is a random variable describing the residual noise at this location.

Nothing is known a priori about the value of I_{planet}, hence we accept H_1 by rejecting H_0. Given an observed intensity I, how likely is it that the null-hypothesis accounts for this observation? In other words, what is the probability that the random variable I_{noise}

takes a value equal or greater than I. Or mathematically, how large is $P(f_g \geq I | H_0) = P(I_{\text{noise}} \geq I)$. If it is unlikely that the residual noise explains the observed values, then it is likely that something else, such as an off-axis signal, explains it.

In this context, it is thus important to have a realistic estimate of the distribution of the residual noise. Then one can select a detection threshold in order to have a fixed confidence level, i.e. a fixed probability, for a planetary signal to be detected.

Because the usual data reducing techniques have a whitening effect on the residuals, justifying the independence of the random variables summed together into the processed frame and the number of frames being typically large, the CLT is generally invoked to state that the residual noise in the processed frame follows a Gaussian distribution (Marois et al. 2008; Mawet et al. 2014). Hence, under the null hypothesis, pixels values are assumed to be drawn from a Gaussian distribution with zero mean and standard deviation σ. The probability of observing a value of 5σ is below 3×10^{-7}. Hence, rejecting H_0 when $I_g > 5\sigma$ yields a confidence level of $1 - 3 \times 10^{-7}$ under a Gaussian assumption. However, it is known that the Gaussian assumption leads to high FPRs (Marois et al. 2008; Mawet et al. 2014). We briefly summarize how this non-Gaussianity is currently accounted for in Section 4.1, before introducing the proposed detection procedure based on our theoretical analysis.

In the context of direct imaging, the outcome of the hypothesis testing is a detection map that assigns a value to each pixel in the processed frame. The larger this value, the less likely the null-hypothesis is verified, hence the more likely a planetary signal is present at that location.

2.3 Planet-free datasets

In what follows, we will study the distribution of the residual speckle noise in the processed frame. The presence of a planet disturbs the tail distribution as planets are precisely detected as outliers in the processed frame. Hence the study of the tail decay requires planet-free datasets. An option is to remove all known planetary signals with, for instance, the negative companion injection method (NEGFC; Marois, Macintosh & Véran 2010). However, this method does not guarantee that the totality of the planetary signal is removed as it cannot be used for extended sources or faint planetary signals that are not previously detected.

For these reasons, we decided to use another method to significantly reduce the influence of potential exoplanets or other on-sky signals in the dataset. We consider the trajectory groups $\bar{g} \in \bar{\mathcal{G}}$ that we obtain using the opposite values of the parallactic angles (Marois et al. 2008). This way, we obtain a similar temporal dependence of the residual speckles but circumstellar signals will be all averaged to negligible values and have larger standard deviation. Processed frames obtained with this method are referred to as *opposite angles processed frames* in the text. We discuss the applicability of this method in Section 5.

In the following, we will use three datasets described in Appendix A: β-Pic using the VLT-NACO instrument, and HD 206893 and 51 Eri taken with the VLT/SPHERE-IRDIS instrument. Each target hosts a planetary signal.

3 RESIDUAL SPECKLES STATISTICS

In this section, we first present the statistics followed by the speckle noise. Then we empirically show that the distribution of the residual noise on the processed frame exhibits a slower decay than expected with a Gaussian distribution. After that, we introduce the concept of sub-exponentiality and show that the MR distribution is sub-exponential. We use this newly demonstrated property of the speckle noise to characterize the residual noise on the processed frame using non-asymptotic statistics. We end this section by an analysis of the impact of the non-independence of the residuals on S.

3.1 Speckle noise statistics

The mean intensity I for an AO-corrected long exposure can be modelled as the sum of the static coherent point spread function (only due to the diffraction by the telescope aperture) I_c and a random speckle noise intensity I_s. It has been shown that the total intensity I follows a Modified Rician (MR) distribution:

$$p_{\text{MR}}(I, I_c, I_s) = \frac{1}{I_s} \exp\left(-\frac{I + I_c}{I_s}\right) \mathcal{I}_0\left(\frac{2\sqrt{I I_c}}{I_s}\right), \tag{4}$$

where $\mathcal{I}_0$ is the modified Bessel function of the first kind. This equation, first derived for laser (Goodman 1975), was adapted to HCI for exoplanet detection (Fitzgerald & Graham 2006; Soummer et al. 2007; Marois et al. 2008, and references therein).

The expectation and variance of I are given by (Soummer et al. 2007):

$$\mathbb{E}I = I_s + I_c, \tag{5}$$

$$\sigma_I^2 = I_s^2 + 2 I_s I_c. \tag{6}$$

In the following, we will consider that in high flux regime the other sources of noise (photon and detector noise) are negligible, especially close to the star, and hence $\sigma_{\text{total}} = \sigma_I$.

The parameters I_c and I_s are not constant throughout the field of view, however they can be consider as constant at a given radius, i.e. at constant separation from the star. Hence the mean and standard deviation of the speckle is a function of the radial distance from the star.

When I_c equals zero, the intensity distribution turns into a pure speckle exponential statistics. Note that when using a coronagraph, the unaberrated term I_c tends towards 0 since the coronagraph is designed to remove the static diffraction pattern. Moreover, thanks to the post-processing, the PSF model L which is subtracted to the data cube includes the static features and hence $I_s \gg I_c$.

Another important aspect of the speckle noise is the different time-scales appearing in the speckle formation (Hinkley et al. 2007). A thorough analysis of the impact these different time-scales is beyond the scope of this paper. However, we argue (Soummer et al. 2007) that this effect can reasonably be modelled by changing I_s to $I_{s_1} + I_{s_2}$, where I_{s_1} and I_{s_2} are the random speckle noise intensity of time-scales τ_1 and τ_2, respectively. This substitution has no impact in the scope of this work, hence, for simplicity and without loss of generality, we consider the distribution displayed in equation (4) for the speckles.

3.2 Tail-decay of the residual noise on the processed frame

We use a statistical visualization tool, namely a quantile–quantile plot (Q–Q plot), to empirically show that, under the null-hypothesis, the tail decay of pixel intensities on the processed frame is better explained by a Laplacian distribution than by a Gaussian distribution. This empirical observation is, later in this paper, supported by a theoretical analysis of the MR distribution.

A Q–Q plot is a statistical tool to compare two distributions by plotting their respective quantiles against each other (see e.g.
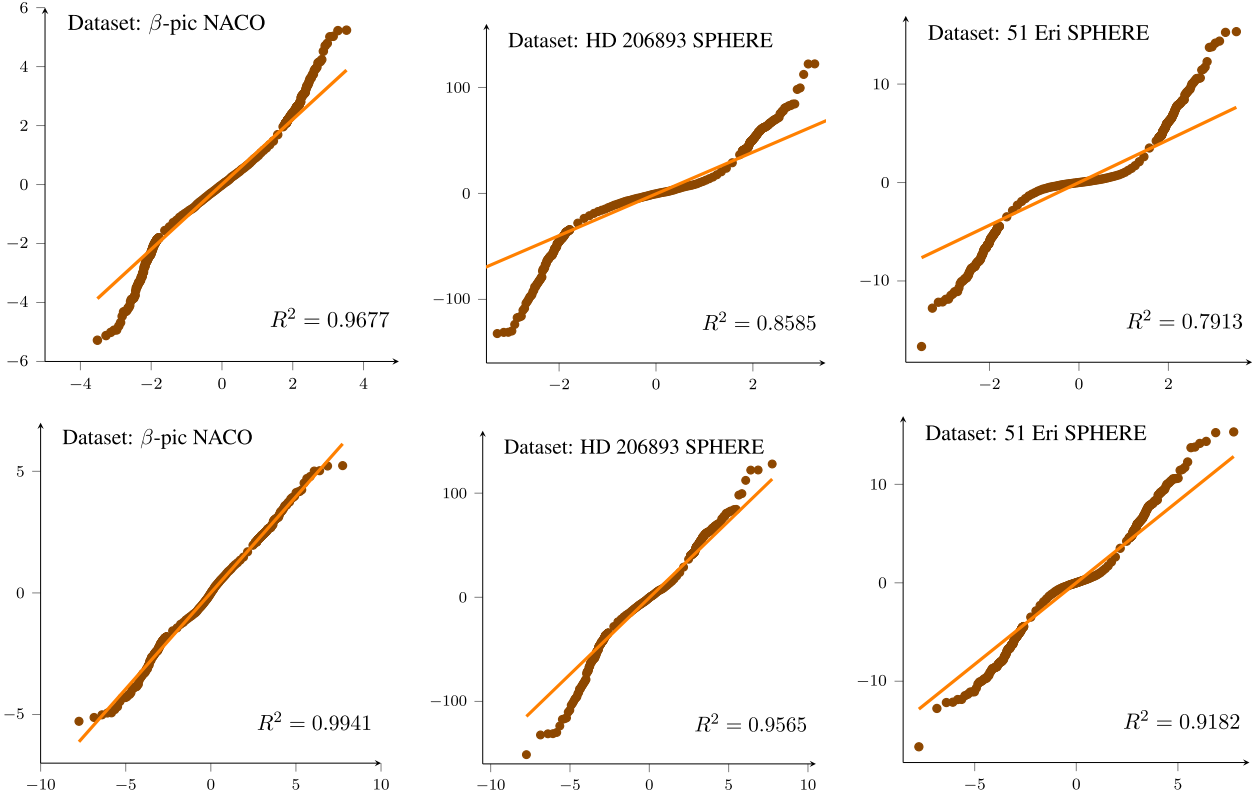
Figure 2. Q–Q plots for the three datasets. The sample is drawn from the centre circle of radius of $8\lambda/D$ of the opposite angles processed frame, using respectively 10, 17, and 5 principal components. Top, we compare the distribution of the pixel values to the Gaussian distribution. We can see that the fit is accurate for the centre of the distribution but becomes increasingly bad for the end of the tails. Bottom, we make the comparison with the Laplace distribution. The corresponding Q–Q plot being closer to a straight line than under the Gaussian assumption. We display R^2, the coefficient of determination, i.e. the Pearson correlation between the paired of quantiles. The closer R^2 is to 1, the better the fit is, i.e. if the two distributions are linearly related, then $R^2 = 1$. Its values confirms here that the Laplacian fit is better than the Gaussian fit.

Heiberger & Holland 2004). It thus allows one to compare the empirical distribution of some data to the distribution they are assumed to follow. If the assumed distribution is correct, the resulting Q–Q plot will approximately be a straight line. When the distribution parameters (e.g. mean and variance) are identical, the resulting straight line is the bisector. Deviations from the straight line indicate differences between the two distributions, such as different skewness or kurtosis. In the context of exoplanet detection, we are particularly interested in the tail of the distribution, as this indicates when the H_0 hypothesis can be rejected. As we are only interested in the positive values, the information about the tail is extracted from the right side of the plot. If the points of the Q–Q plot are above the straight line in the right side of the plot, it indicates that the distribution on the y-axis (data) has a heavier tail than the distribution on the x-axis (test distribution).

We use the opposite angles processed frames for all three datasets in order to significantly reduce the influence of any circumstellar signal which could bias the residual speckles distribution (as described at the end of Section 2). To build the Q–Q plots, we arbitrarily select the pixels within the central annulus of radius $8\lambda/D$ in the opposite angles processed frame as our data sample. Note that we observe that the results are similar when performed with a different radius or annulus-wise.

We compare the pixel intensities distribution of our sample to the Gaussian distribution in the Q–Q plot displayed in Fig. 2 (top). We see that the Gaussian fit is good for the first quantiles but becomes increasingly discrepant for higher quantiles. As a consequence, the

likelihood that residual noise has a large value is greater than what we can expect from a Gaussian random variable. In the framework of detection, it means that a given probability of presence under the Gaussian assumption will in reality yield a larger number of false alarms.

Interestingly, we found that a Laplace distribution yields a better fit for the tail decay, as shown in Fig. 2 (bottom). The Laplace distribution follows a probability density function $f(x|\mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$, where μ is the mean and b refers to as the diversity, which is linked to the variance as $\sigma^2 = 2b^2$. For $x \geq \mu$, the cumulative distribution function (CDF) is given by

$$F_x(x) = P(X \leqslant x) = 1 - \frac{1}{2} \exp\left(-\frac{x-\mu}{b}\right).$$

In other words, this means that the probability that a Laplacian random variable is larger than the mean, $P(X - \mu \geqslant x) = \frac{1}{2} \exp\left(-\frac{x}{b}\right)$ decreases exponentially. In comparison, for a Gaussian r.v.'s, we have $P(X - \mu \geqslant x) \leqslant \exp\left(-\frac{x^2}{2\sigma^2}\right)$ when $x \geq 0$ increases.

To ease the comparison between the Gaussian and the Laplacian plots, we also displayed in Fig. 2 the coefficient of determination R^2 that is the Pearson correlation between the paired sample quantiles. For two compared samples X and Y, it is given by:

$$R^2 = \left(\frac{\mathrm{cov}(X, Y)}{\hat{\sigma}(X)\hat{\sigma}(Y)}\right)^2,$$

where $\text{cov}(X, Y)$ is the sample covariance between samples X and Y. The closer R^2 is to 1, the closer the distribution X is to the distribution Y. We can see that R^2 is always closer to one in the Laplacian probability plots than in the Gaussian probability plots, thus supporting the observed trends.

We emphasize here that the only information provided by Fig. 2 is that the tail decay is better explained by a Laplacian (i.e. it has an exponential tail decay) but not that the actual distribution of the residual noise on the processed frame is Laplacian. A bound for the actual tail decay of the residual noise on the processed frame is derived in Section 3.4, based on the theoretical results presented in Section 3.3.

3.3 Sub-exponentiality of the MR distribution

We prove now that an MR distribution has an exponential tail day, i.e. it belongs to the class of *sub-exponential* random variables (Vershynin 2010). Consequently, an MR random variable Z is such that $P(Z > z)$ decays like $\mathcal{O}(e^{-z})$ when the level z increases, and not like the Gaussian tail decay $\mathcal{O}(e^{-z^2})$. As described below (see Theorem 1), this has a clear impact on the sum of m independently and identically distributed (i.i.d.) MR random variables; this sum, seen as a random variable, has a tail that also decays exponentially when the level increases beyond a value only depending on the MR characteristics.

Mathematically, a sub-exponential random variable is such that (Vershynin 2010),

$$P\left(|X - E(X)| \geqslant t\right) \leqslant c_1 \exp(-c_2 t),$$

for all $t \geq 0$, where c_1, $c_2 > 0$ are two universal constants. Equivalently, sub-exponential random variables can be defined as follows (Vershynin 2010).

DEFINITION 1 **(Sub-exponential random variables)** A random variable X is called sub-exponential if its sub-exponential norm $\|X\|_{\psi_1}$, defined as

$$\|X\|_{\psi_1} := \sup_{p \geqslant 1} p^{-1} \left(E|X|^p\right)^{1/p}, \tag{7}$$

is bounded, i.e. $\|X\|_{\psi_1} < +\infty$.

Interestingly, as proved in Appendix B, MR random variables are sub-exponential.

PROPOSITION 1. *Let* $X \sim MR(\alpha, \beta)$ *be an MR random variable with*

$$MR(\alpha, \beta) \sim \frac{1}{\beta} \exp\left(-\frac{t + \alpha}{\beta}\right) \mathcal{I}_0\left(\frac{2\sqrt{t\alpha}}{\beta}\right),$$

i.e. with mean $\alpha + \beta$ *and variance* $\beta^2 + 2\alpha\beta$*. Then* X *is sub-exponential with* $\|X\|_{\psi_1} \leqslant 6\beta$*.*

Therefore, sums of MR random variables enjoy the following concentration phenomenon shared by all sub-exponential random variables.

THEOREM 1 *(Vershynin (2010)) Let* $X_1, \ldots, X_m$ *be independent* m *centred sub-exponential random variables, and let* $K = \max_i \|X_i\|_{\psi_1}$*. Then, for every* $\epsilon \geq 0$*, we have*

$$P\left(\sum_{i=1}^m X_i \geqslant \epsilon m\right) \leqslant \exp\left[-c \min\left(\frac{\epsilon^2}{K^2}, \frac{\epsilon}{K}\right) m\right], \tag{8}$$

where $c > 0$*.*

We can thus bound the complementary CDF (i.e. $1 - \text{CDF}$) of $\sum_i X_i$ and state that for small values of the level ϵ, sub-exponential

random variables behave as Gaussian random variables while for larger values of ϵ, they exhibit an exponential decay when ϵ increases. As detailed in Section 3.4, this change of behaviour is critical to detect outliers to the speckle distribution, e.g. planetary signals.

3.4 Non-asymptotic analysis of the residual noise

The takeaway message of Proposition 1 is that we can apply Theorem 1 to characterize the tail decay of pixel intensities in the processed frame under the null-hypothesis (in the absence of a planetary signals). We first show how this can be done for a processed frame computed with the c-ADI algorithm. Then we argue that, even if in the case of PCA this analysis is more complicated and beyond the scope of this paper, we can still expect our analysis to hold.

For the c-ADI algorithm, L consists of T copies of an image whose pixel intensities are the temporal median (or mean) of the data sequence of the corresponding trajectories. Hence the mean of the distribution is shifted towards zero but otherwise remains MR and under the i.i.d. hypothesis, the sum along trajectories (step 3 in Fig. 1) satisfies equation (8). The processed frame is $f_g = \sum_i s_i^{[g]}/T$ and for a trajectory g, under the null-hypothesis, we have

$$P\left(\frac{1}{T}\sum_i s_i^{[g]} \geqslant \epsilon\right) \leqslant \exp\left(-c \min(\epsilon^2/K^2, \epsilon/K)T\right). \tag{9}$$

From equation (4) and Proposition 1, we note that if $I \sim MR(I_c, I_s)$, then $K \leq 6I_s$ and the variance is equal to $\sigma_I^2 = I_s^2 + 2I_c I_s \geqslant K^2/36$. Therefore, $K \leq \sigma_I/6$ and the bound (9) implies

$$P\left(\frac{1}{T}\sum_i s_i^{[g]} \geqslant \epsilon\right) \leqslant \exp\left(-\tilde{c} \min\left(\bar{c}\epsilon^2/\sigma_I^2, \epsilon/\sigma_I\right) T\right), \tag{10}$$

for some $\tilde{c} > 0$ and $\bar{c} > 0$. For small values of ϵ/σ_I, we do observe a Gaussian bound in equation (10). On the contrary, for large values of ϵ/σ_I the bound displays an exponential decay, i.e. the tail decreases exponentially as ϵ increases. As the confidence level of the detection depends on the probability of having an outlier, it is important to take into account this phenomenon to avoid high FPRs. Therefore we can reject the null-hypothesis by a careful selection of a threshold driven by the bound (10). This is the theoretical motivation of our detection map presented in Section 4.2.

Moreover, the bound in equation (10) provides the information that the tail of distribution of the residual noise on the processed frame decays exponentially with the number of frames. Indeed, we see that the more i.i.d. random variables are added together, the larger the probability that the sum does not deviate from the mean. This effect is called the concentration of measure. The non-asymptotic nature of our analysis lies in this explicit dependence on the number of frames.

To the best of the authors' knowledge, these effects had not been fully theoretically assessed. This is the key theoretical contribution of this paper as it explains the lower confidence level observed in the literature and it is the basis of the detection map proposed in Section 4.2.

In the case of PCA, the theoretical analysis is more difficult and is left for future work. Nevertheless, we observed from the empirical Q–Q plots in Fig. 2 that PCA-generated processed frames exhibit residual noise with exponential tail decay. Furthermore, we argue that since L is built with few principal components is the low-rank structure of the data sequence, i.e. that it captures the slowly varying parts of the data volumes, L remains close to the temporal mean. Hence the residual noise of a PCA-processed frame is expected to
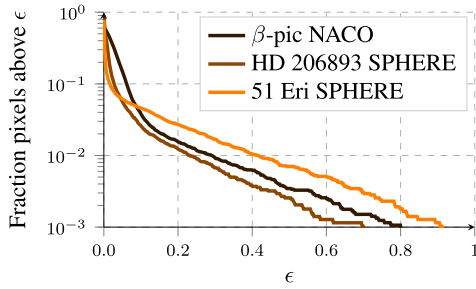
Figure 3. The fraction of pixels in the opposite angles processed frame above a certain threshold ϵ as ϵ increases for three different datasets. We can see the fast decay for small values of ϵ followed by a smaller decay. Each opposite angles processed frame is obtained using one principal component.

be similar to that of ADI-processed frame. We also note that PCA subtraction removes highly temporally correlated speckles, leaving lesser correlated speckles (Mawet et al. 2014) and thus the i.i.d. hypothesis of Theorem 1 is more likely to be verified for PCA. Indeed, we show in the Section 3.5 that the non-independence of the residual noise along trajectories results in a slowdown of the concentration towards the mean and that the temporal correlation drops quickly as the number of principal components increases.

To furthermore illustrate that we can use equation (10) to bound the probability that residual corruption on a PCA-processed frame reaches a value ϵ, we exhibit the sub-exponential behaviour expected from equation (10) in three different datasets using PCA in the reduction. To do so, we compute the empirical complementary CDF of the opposite angles processed frame of the three datasets. We count n_ϵ the number of pixels on the processed frame that are larger than ϵ, for an ϵ ranging from zero to the maximal values taken on the processed frame. We display in Fig. 3 the evolution of n_ϵ/n as ϵ increases for the three considered datasets. We can see that for all datasets the decay of the complementary CDF exhibits two distinct behaviours as expected from the bound of equation (10). First a fast decay that is compatible with a quadratic (thus Gaussian) decay. Then a slower decay that displays a linear (thus exponential) trend.

For these reasons, we used a PCA to process the data throughout this paper and we characterized the resulting processed frame using our analysis. As demonstrated in Section 5, we obtained convincing results regarding the capability of our method to detect exoplanets.

3.5 Time-dependent residual speckles

As it is common in the literature, we assumed so far that the whitening effect of the reduction procedure is sufficiently strong to consider that the residual noise along trajectories is made of independent temporal components. That is the underlying assumption when one uses the CLT to state that the noise on the processed frame asymptotically follows a Gaussian distribution if the number of frames is large. In our analysis, we also sum random variables along trajectories and the bound given by equation (10) only holds for i.i.d. random variables $s_i^{[g]}$, i.e. $s_i^{[g]}$ and $s_j^{[g]}$ are independent for $i \neq j$.

It is thus important to estimate the length of the temporal dependence of the residual speckle and its impact on the confidence level for the detection procedure. Due to its asymptotic nature, it is unclear how one can estimate the impact that the non-independence of the random variables has on the CLT, whereas this estimation is possible for our analysis. We here show how, under mild assumptions, the tail bound in equation (10) can be modified to
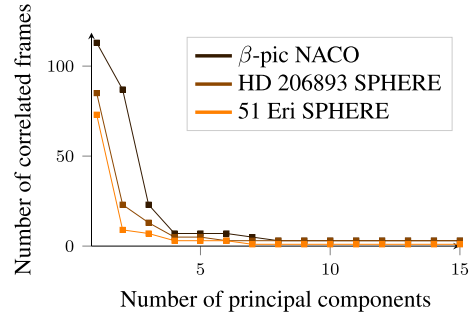

Figure 4. Correlation length (in number of frames) with respect of the number of principal components used in the processing for different dataset.

account for the non-independence of the random variables $s_i^{[g]}$. We did not consider spatial dependence because it does not interfere with the assumptions of Theorem 1.

We are here interested in the typical temporal dependence along trajectories after subtraction of the model PSF *in terms of number of frames* as it determines the number of i.i.d. random variables appearing in the sum of equation (10). We assume a length of dependence τ, such that τ successive elements of a trajectory are dependent. In other words, two elements of a trajectory x_i and x_j are independent if $|i - j| > \tau$ (τ being an integer). The result concerning the sub-exponential decay of the tail is summarized in the following corollary, the proof is postponed to Appendix C.

COROLLARY 1. *With a length of dependence τ, equation (10) becomes*

$$P\left(\frac{1}{T}\sum_i s_i^{[g]} \geqslant \epsilon\right) \leqslant \tau \, exp\left(-\tilde{c}\min\left(\frac{\tilde{c}\epsilon^2}{\sigma_I^2}, \frac{\epsilon}{\sigma_I}\right)\frac{T}{\tau}\right). \qquad (11)$$

Corollary 1 tells us that the temporal dependence slows down the concentration of measure around the mean.

It is in practice impossible to know the temporal dependence. Hence, we use the temporal autocorrelation (for each trajectory) after the subtraction of the model PSF in order to estimate the dependence of the residual noise. For a signal with non-correlated samples, the autocorrelation displays a large central peak surrounded by white noise. We averaged the normalized autocorrelation of all trajectories into a total autocorrelation. We define the correlation length as the number of frames for which the averaged autocorrelation is above a certain threshold.

We display in Fig. 4 the evolution of the correlation length, using a threshold of 0.2, for our three different datasets processed using PCA when the number of principal components increases. Although the typical correlation length decreases quickly as the number of principal components increases, its impact cannot be disregarded and the bound of equation (11) is more accurate than the bound of equation (10). However, the estimation of the factor τ depends on the threshold one uses to compute the correlation length. Hence, the presence of τ (along with the other constants) in the bound of equation (10) hinders our capability to exactly estimate the inverse CDF of the distribution of the residual corruption. We show in Section 5 that the inverse CDF can be estimated from the opposite angles processed frame.

4 DETECTION MAP

In this fourth section, we first briefly summarize a popular approach to compute a detection map from the processed frame, i.e. the SNR_t map (Mawet et al. 2014), and list the shortcomings of this method.
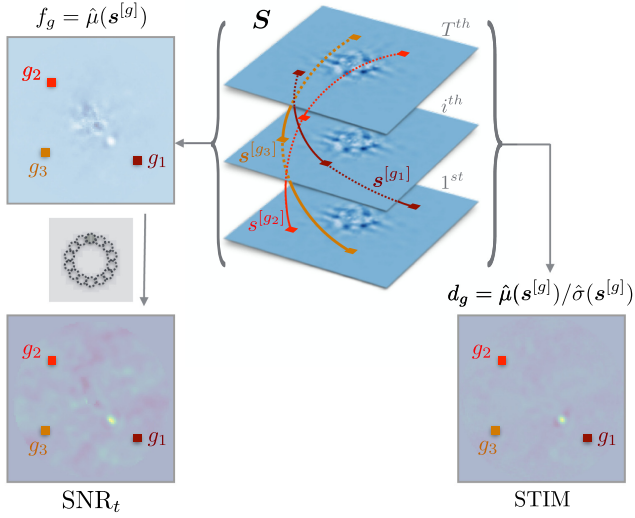
Figure 5. Illustrative views of the computations of both the SNR_t and the STIM maps. We depicted three arbitrary trajectories $s^{[g_i]}$ and the corresponding elements g_i of each frame to illustrate the notations used in the paper. The SNR_t is built from the processed frame, comparing the statics of resolution elements located at the same radial distance from the centre of the frame. The STIM map is constructed directly from the residual cube S.

We then propose our novel detection map, the STIM map, and provide statistical guarantees for its subsequent thresholding based on our theoretical analysis.

4.1 Current detection procedure for exoplanets direct imaging

As the statistical properties of the processed frame are known to be radius dependent, the null-hypothesis is tested for each radius separately. And because of the diffractive effect, resolution elements of diameter λ/D are considered. As there are fewer such resolution elements closer to the star, the confidence level of the detection suffers from small sample statistics.

To deal with the small sample statistics and the high FPR obtained when relying directly on Gaussian confidence level, Mawet et al. (2014) proposed another detection map based on the t-student statistics. The t-test is argued to be more robust with respect to small sample statistics and with respect to the deviation from a Gaussian distribution for the null-hypothesis. The test is defined by computing the following map from the pixels intensity on the processed frame

$$\mathrm{SNR}_t(x) = \frac{\bar{x}_1 - \bar{x}_2}{s_2 \sqrt{1 + \frac{1}{n_2}}}, \qquad (12)$$

where $\bar{x}_1$ is the mean flux of intensity inside the circle surrounding the regarded pixel centred on x, n_2 is the number of the other λ/D resolution elements located at the same radial distance from the centre of the frame, and $\bar{x}_2$ and s_2 are respectively the mean fluxes and sample variances of all such resolution elements. The computation of the SNR_t map is illustrated in Fig. 5. The resulting SNR_t map and the processed frame from which it is built can be seen in Fig. 6 for the three tested datasets. These images were built using the VORTEX IMAGE PROCESSING (VIP) PYTHON package (Gonzalez et al. 2017b).[2]

Despite its capacity to detect faint objects (see e.g. Currie et al. 2015; Quanz et al. 2015), there are two main drawbacks for this SNR_t.

First, although the t-test is more robust with respect to deviation from a Gaussian distribution, this deviation is not quantified. We can hope that this distribution will be close to Gaussian if we add enough frames. But we have no knowledge of how robust the validity of the Gaussian hypothesis is with respect of the number of frames.

Secondly, the SNR_t compares the statistical properties of a pixel against pixels at same angular separation from the star. If there exist multiple planets at the same radius, or extended structures such as circumstellar discs, the map will overestimate the noise at that radius, hence increasing the likelihood of a planet to be considered as a residual speckle. Furthermore, any test performed directly on the processed frame suffers from the small sample statistics when considering pixels near the centre of the processed frame, i.e. when attempting to detect planets close to their host star.

Our objective is to define a robust detection procedure, still efficient at very close separation to the star in order to be sensitive to faint planetary signals all over the field of view, even at locations where the starlight residuals are very intense and varying fast (typically below a few λ/D from the star).

The proposed detection map is computed in the temporal domain and independently for each trajectory. Hence it has a sample statistics that does not depend on the radial distance from the centre and the presence of multiple planets does not hinder the likelihood of a detection. Furthermore, the analysis of the distribution of the residual noise is non-asymptotic with respect to the number of co-added frames, consequently it is possible to estimate the confidence of a detection depending on the number of frames available.

4.2 STIM: a time domain sub-exponential detection map

We now present our detection map and justify its use in the light of the results of the previous section. The presence of the standard deviation in the exponential in equation (10) induces a slower decay rate in areas of S with a larger standard deviation. Thus trajectories with larger temporal standard deviation are more likely to yield a significantly large value on the processed frame under the null hypothesis. For this reason, we propose to compute the map d whose components are given by

$$d_g = \frac{\hat{\mu}_g}{\hat{\sigma}_g}, \qquad (13)$$

where $\hat{\mu}_g$ is the computed mean of trajectory g in equation (1) and $\hat{\sigma}_g$ its standard deviation; e.g. the square root of its computed variance

$$\hat{\sigma}_g^2 \equiv \hat{\sigma}\left(s^{[g]}\right)^2 = \frac{1}{T-1} \sum_{i=1}^{T} \left(s_i^{[g]} - \hat{\mu}_g\right)^2. \qquad (14)$$

We use d as a detection map and we name it the standardized trajectory intensity mean map or STIM map for short. Its computation is illustrated in Fig. 5.

Another way to introduce the STIM map would be the following. The quantity $\hat{\sigma}_g$ is proportional to the standard deviation of the computed mean $\hat{\mu}_g$. Indeed, denoting the standard deviation of $\hat{\mu}_g$ by $\sigma(\hat{\mu}_g)$, it is given by

$$\sigma\left(\hat{\mu}_g\right) = \frac{\sigma\left(s^{[g]}\right)}{\sqrt{T}}.$$

[2]The code is publicly available at https://github.com/vortex-exoplanet/VIP.
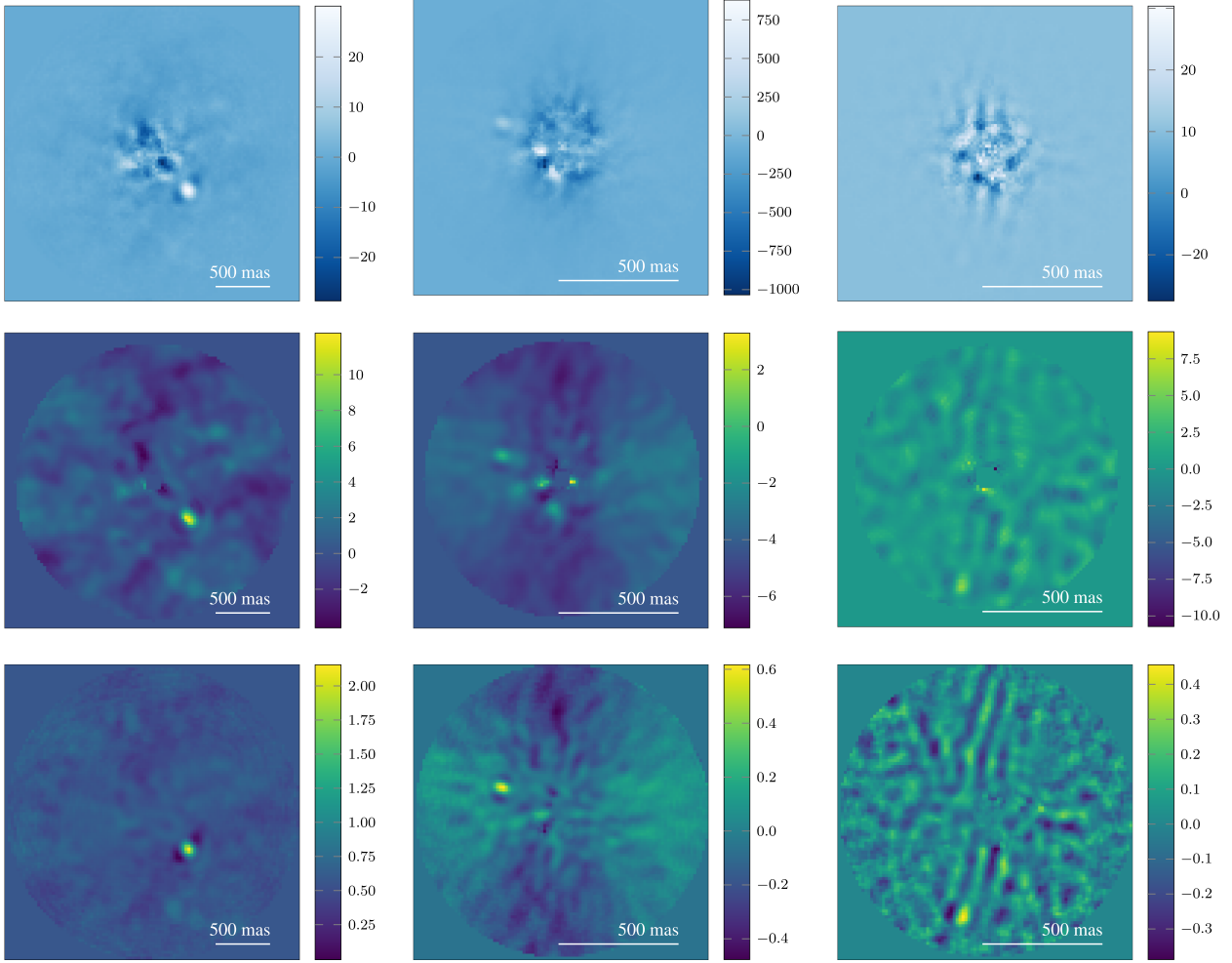
Figure 6. Illustrative results for the three datasets: β-pic NACO (left), HD 206893 SPHERE (centre), and Eri 51 SPHERE (right). Top: the processed frame, obtained with PCA, using respectively 3, 10, and 5 principal components. Middle: SNR_t map. Bottom: STIM map.

Replacing the standard deviations by the estimator (14), we get that the STIM map is given by

$$d_g = \sqrt{T} \frac{\hat{\mu}_g}{\hat{\sigma}(\hat{\mu}_g)}.$$

We display the STIM maps obtained for the three considered datasets in Fig. 6. In Fig. 7, we also display a one-dimensional plot of both the SNR_t and the STIM map for the HD 206893 SPHERE dataset in order to better compare the behaviour of the two maps.

Intuitively, because residual speckles on S are spread on multiple trajectories, for a trajectory g that is free of an exoplanet signal, the computed mean is expected to be small after the PCA subtraction. However, since most temporally correlated speckles are absorbed in L (see Section 2), the computed variance of the intensities along the trajectory g is expected to be large. Thus, we can expect to have $\hat{\mu}_g/\hat{\sigma}_g \ll 1$. On the other hand, since most the planets flux is still present on S, for a trajectory p containing a planet, we expect the computed mean to be large. Hence, we expect $\hat{\mu}_p/\hat{\sigma}_p$ to be larger than $\hat{\mu}_g/\hat{\sigma}_g$.

We now show how we can use sub-exponentiality of the MR distribution to characterize the proposed detection map. Using the argument from Section 3.2, we can apply Theorem 1 with $\epsilon \leftarrow \sigma_I \epsilon$

and equation (10) becomes

$$P\left(\frac{1}{T}\sum_i s_i^{[g]} \geqslant \sigma_I \epsilon\right) \leqslant \exp\left(-\tilde{c}\min\left(\bar{c}\epsilon^2, \epsilon\right) T\right). \quad (15)$$

As we noted, σ_I depends on the separation from the star. We assume that there are enough frames so that the computed standard deviation along a trajectory is a good estimate of σ_I for that trajectory and we can set $\epsilon \leftarrow \epsilon T \hat{\sigma}_g \approx \epsilon T \sigma_I$ and rearrange

$$P\left(\frac{\sum_i s_i^{[g]}}{T\hat{\sigma}_g} \geqslant \epsilon\right) \leqslant \exp\left(-\tilde{c}\min\left(\bar{c}\epsilon^2, \epsilon\right) T\right). \quad (16)$$

By definition $\sum_i s_i^{[g]}/T = \hat{\mu}_g$, hence

$$P\left(\frac{\hat{\mu}_g}{\hat{\sigma}_g} \geqslant \epsilon\right) \leqslant \exp\left(-\tilde{c}\min\left(\bar{c}\epsilon^2, \epsilon\right) T\right). \quad (17)$$

In other words, the probability that the residual speckle noise reaches a large value on the STIM map decays exponentially with respect to T, the number of frames.

We argue that the proposed detection map does not suffer from the two drawbacks of SNR_t described in Section 4.1.

First, by accessing the temporal domain, it is not impacted by the small sample statistics when considering planets close to their host (compared to the SNR_t that only analyses the processed frame).
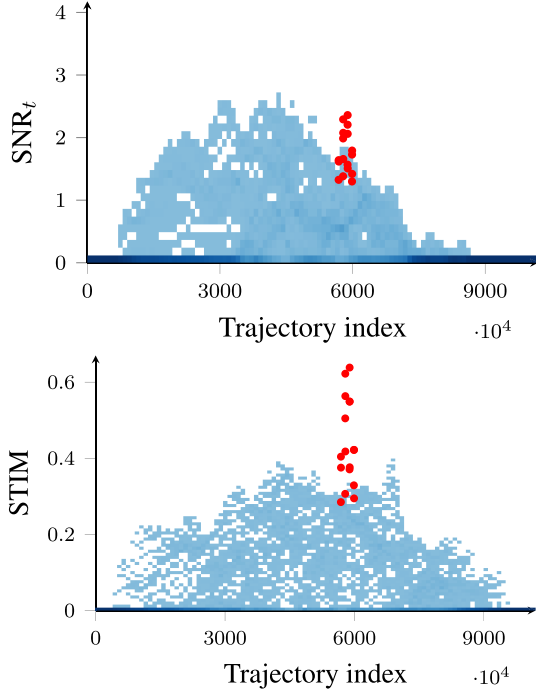
Figure 8. Plot of the fraction of pixels in the STIM map above τ in three different settings: (i) dark, the planet is present, (ii) medium, the planet is removed with the negative companion injection method, and (iii) light, the planet is present but the opposite parallactic angles are used in the processing. As the decay is indistinguishable for the cases (ii) and (iii), we added dots (light) and crosses (medium) to ease the comparison. On the other hand, the decay is significantly different when a planet is present. These results are obtained on the β-pic dataset processed using a PCA with 10 principal components.

Figure 7. Comparison of the SNR_t map (top) and the STIM map (bottom) in a one-dimensional plot for the HD 206893 dataset. The planetary trajectories are represented with dots. Other trajectories are displayed as rectangles in a transparency fashion, the darker the more trajectories take this value. We can see that for the SNR_t map, it is not possible to select a detection threshold such that planetary trajectories are the only to take values larger than the threshold. For the STIM map, such a threshold exists. In Section 5, we describe how to set automatically this threshold from the complementary CDF.

Indeed, the sample statistics of the STIM map is the number of frames T and is the same for all radii.

We mention that if T is small (or if the temporal correlation is strong, see Section 3.5), $\hat{\sigma}$ is not necessarily a good estimator of the standard deviation. Nevertheless, the asymptotic analysis behind the state-of-the-art detection procedures suffers from the same small sample statistics. Indeed, if T is small or strongly correlated, the hypothesis that the number of *i.i.d.* random variables summed-up can be considered as infinite is severely hindered. Furthermore, the non-asymptotic nature of our analysis allows one to evaluate the quality of the estimator $\hat{\sigma}$ using standard statistical tools.

Secondly, by construction, the intensity of a given pixel on d does not depend on the intensity of other pixels. There is thus no influence of other planets that could be located on the same radius. We also note that the STIM map is easy to implement and fast to compute. For the illustrative examples of Fig. 6, the STIM maps only required a third of a second to be completed using a single CPU. A python implementation of the STIM map is available in the VIP toolbox.

The detection map suffers the same concentration of measure slowdown as described in Section 3.5. With a length of dependence τ, equation (17) becomes

$$P\left(\frac{\hat{\mu}_g}{\hat{\sigma}_g} \geqslant \epsilon\right) \leqslant \tau \exp\left(-\tilde{c}\min\left(\bar{c}\epsilon^2, \epsilon\right) T/\tau\right). \tag{18}$$

In Section 5.1, we use the opposite angles detection map to estimate the typical values observed on the detection map under the null hypothesis. As with the opposite angles procedure the plan-
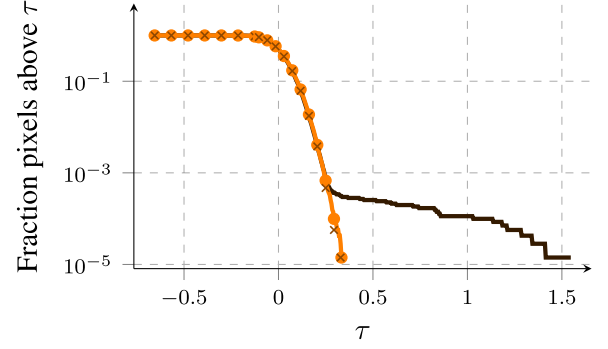
etary signal is significantly reduced while preserving the speckle temporal dependence, the concentration of measure slowdown of equation (18) is then accounted for.

Before moving on, we note that the bound similar to the one in equation (17) would still hold if the mean is replaced by the median (Wainwright 2019). However, when performing experiments with median, we did not find significant improvement over the mean (See Appendix D).

5 EXPERIMENTS AND DISCUSSION

In this section, we provide some numerical experiments to demonstrate the capabilities of our approach.

5.1 Detection threshold estimation

We know from sub-exponentiality of the MR distribution that the ratio $\hat{\mu}_g/\hat{\sigma}_g$ is close to zero with high probability under the null-hypothesis. Unfortunately, the unknown constants $\tilde{c}$ and $\bar{c}$ in equation (17) and the potential temporal dependence in the residual noise prevents us to leverage this bound to determine a detection threshold rejecting H_0. We here propose to select a detection threshold by estimating the complementary CDF of the residual noise on the processed frame, i.e. estimating $P(\hat{\mu}_g/\hat{\sigma}_g > \epsilon|H_0)$. If this function is known, a threshold can be set so that it is unlikely that the residual noise reaches its value. As we do not have access to the true complementary CDF of the residual noise, we propose to estimate it from the empirical CDF computed from the opposite angles detection map. As outlined previously, the temporal dependence of the speckle noise is the same as in the initial dataset. Therefore, the concentration of the $\hat{\mu}/\hat{\sigma}$ around 0 is also preserved with the same potential concentration of measure slowdown (see Section 3.5). Mathematically, $P(\hat{\mu}_g/\hat{\sigma}_g > \epsilon|H_0) \approx P(\hat{\mu}_{\bar{g}}/\hat{\sigma}_{\bar{g}} > \epsilon)$ holds because H_1 does not arise in the opposite angles processed frame. We estimate the complementary CDF as the number of pixels taking a value larger than a threshold for increasing values of this threshold. We thus compute $n_\epsilon = \sum_g \chi_+(\hat{\mu}_g/\hat{\sigma}_g - \epsilon)$, where $\chi_+(x) = 1$ if $x > 0$ and 0 otherwise. If n is sufficiently large, we have $n_\epsilon/n \approx P(\hat{\mu}_g/\hat{\sigma}_g > \epsilon)$ and n_ϵ/n is then a good approximation of the complementary CDF.
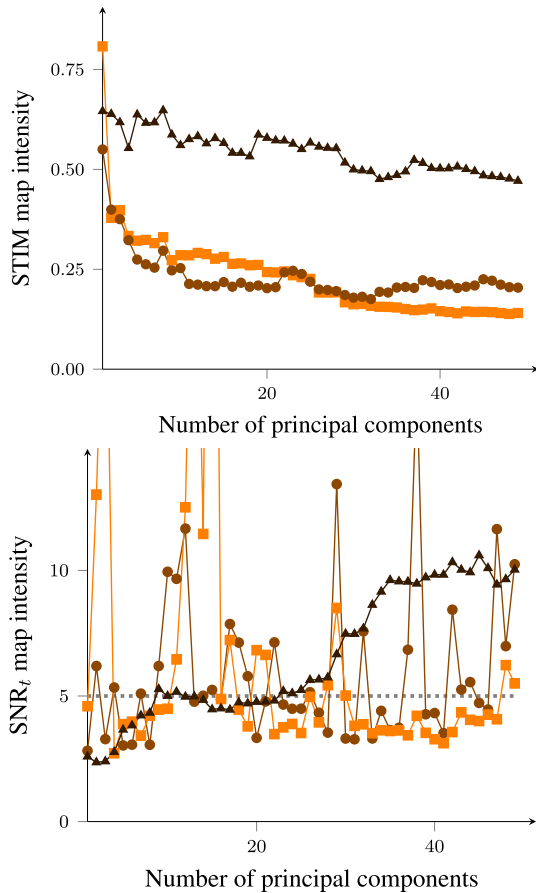
Figure 9. Detectability level for STIM map (top) and for SNR_t map (bottom) obtained with the HD 206893 dataset. In dark (triangle), intensity at planet location, and in light (rectangle), detectability level estimated by inverse trajectories. In medium (dots), detectability level estimated by looking at the second largest spot on the detection map. On the SNR_t map we overplotted in grey dashed the 5σ threshold. The STIM map detects the planet for all rank larger than 1 while SNR_t map requires large number of principal components to yield sufficiently large values of the SNR.

To show that we can indeed use the opposite parallactic angles technique to estimate the residual noise level, we compare the empirical complementary CDF of the detection map in three different cases: (i) the planet is present in the dataset, (ii) the planet signal is strongly attenuated using the negative companion injection method, and (iii) the planet is present but we used the opposite parallactic angles in the processing. We have observed that the curves in cases (ii) and (iii) are practically indistinguishable, hence the complementary CDF of the residual noise can be estimated from the detection map obtained with the opposite parallactic angles. We display the comparison for the β-pic dataset in Fig. 8 where we see that $\tau = 0.5$ is an adequate detection threshold. In Fig. 11, we see that this threshold allows to detect the three injected companions without any false positive.

We also show the sensitivity of the threshold with respect of the number of principal components used in the PCA. We compare the intensity of the planetary signal on the detection maps with *(i)* the largest intensity observed on the opposite angles detection map and *(ii)* the largest intensity on the detection map under the null-hypothesis. Fig. 9 displays the evolution of these three quantities as the number of principal components increase from 1 to 50, for the STIM and SNR_t maps obtained with the HD 206893 dataset.
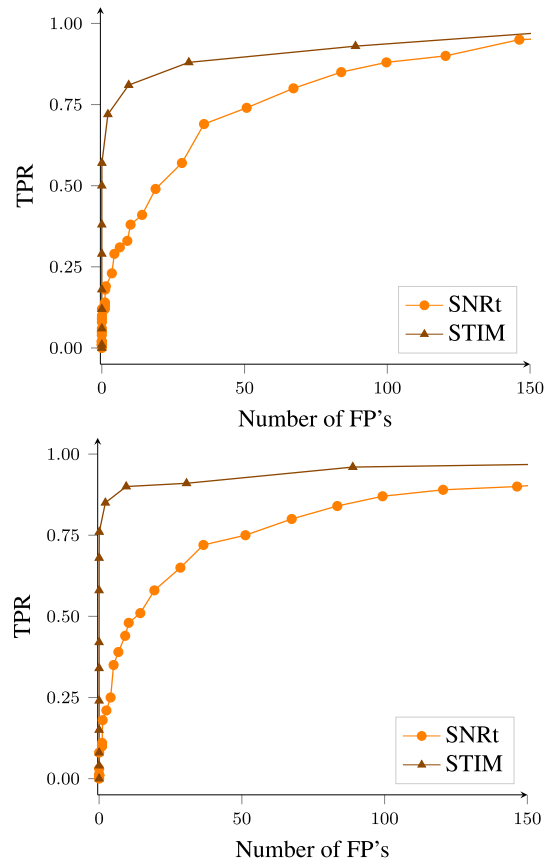
Figure 10. ROC curves obtained for 100 fake companion injections, all processed using four principal components. Top: at radial separation 250 mas with contrast between 2.75×10^{-5} and 4.12×10^{-5}. Bottom: at radial separation 375 mas with contrast between 1.37×10^{-5} and 2.75×10^{-5}.

For the STIM map, the opposite angles detection map yields a good approximation of the maximal noise intensity on the actual detection map. Furthermore, this approximation can be used to estimate a detection threshold such that the planetary signal is detected without false positives. In contrast, for the SNR_t, the maximal values on the opposite angles map is not a good estimate for the maximal value of the detection map under the null-hypothesis. In addition, for a wide range of numbers of principal components, the SNR_t map is unable to detect the planet without false positive. We observed a similar trend for the other datasets.

5.2 Receiver operating characteristic curves

Receiver operating characteristic (ROC) curve is a widely used tool to compare classifiers. It consists of plotting the true positive rate (TPR) as a function of the FPR. A good classifier yields a good trade-off between large TPR for small FPR.

We propose to build a ROC curve that is localized following the procedure described in Gonzalez, Absil & Van Droogenbroeck (2017a). The resulting ROC curve is close to an alternative free-response operating characteristic (AFROC) curve where one plots the fraction of objects detected versus the fraction of images with one or more false positives (Metz 2006). Given a detection threshold, we say there is a TP if there is, within a λ/D diameter area around the planet's position, one pixel whose intensity is above the threshold. We count the number of FP's as the number of λ/D circular areas that contain at least one pixel above the threshold, the
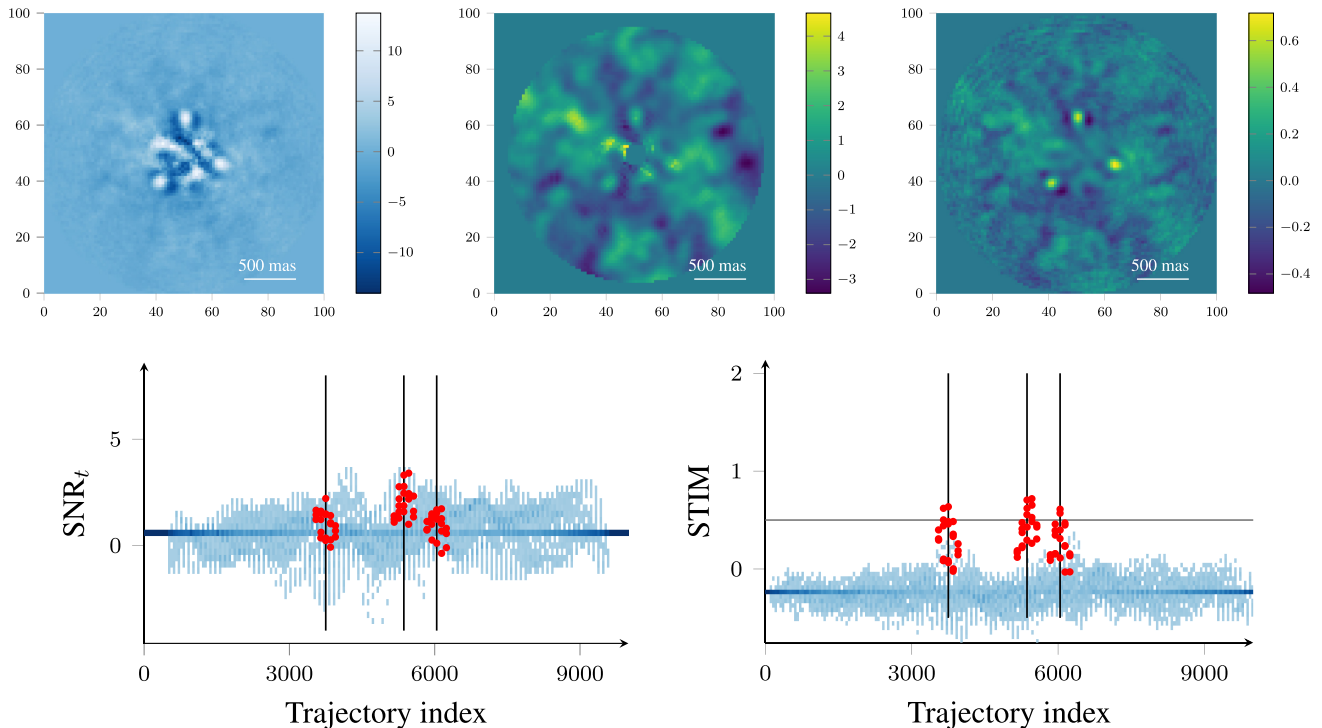
Figure 11. Results with three synthetic exoplanets injected at the same separation in the β-pic NACO data. Top row: processed frame using PCA with five principal components (left), SNR_t map (centre), and STIM map (right). Bottom row: intensities of the SNR_t map (left) and of the STIM map (right) for each trajectory. For the SNR_t map, we see that no planet stands above the noise level, while for the STIM map, all three planets are detected without false positive when using the 0.5 threshold derived in Section 5.1.

experiment is repeated 100 times. We computed the TPR and the FPR for a given radius and for a given intensity range.

We display in Fig. 10 the ROC curves obtained for the β-pic dataset with synthetic planets injected at small inner working angles (top) and high contrast (bottom). In both cases, we observe a gain in terms of TPRs versus FPR ratio. For the small inner working angles case, the TPR for no false positive is 9 per cent for SNR_t map, whereas the STIM map reaches 57 per cent TPR for no false positive. For high contrast, these percentages are respectively 10 per cent and 76 per cent. We conclude that in both cases, a significant improvement is achieved.

5.3 Example with multiple planets at same radius

In Section 4, we stated that when multiple planets are located at the same radial distance from the host star, or when extended structures are present, the SNR_t tends to overestimate the residual noise and this hinders its capability to detect planets. We illustrate this effect in this first experiment. We inject three planets at the same radial separation in a planet-free dataset and show how the SNR_t and STIM maps behave.

We perform this experiment on the β-pic dataset using the opposite angles to significantly reduce the effect of the known companion. Then we injected three synthetic planetary signals, at a radial separation of 352 mas and position angles of $133°$, $270°$, and $15°$ with contrast of 4.41×10^{-5}, 4.81×10^{-5}, and 3.73×10^{-5}, respectively (which is below the 5σ detection limit presented in Absil et al. 2013). We processed the data with a PCA, using five principal components. Fig. 11 shows the processed frame, the SNR_t map, and the STIM map. We see that no planet is found above the

noise level in the SNR_t map whereas all three planets are above the noise level in the STIM map.

6 CONCLUSION AND FUTURE WORK

In this paper, we first showed empirically that the tail decay of the residual noise on the processed frame is better explained by a Laplacian distribution than by a Gaussian distribution. From this qualitative observation, we then theoretically proved that the MR distribution is sub-exponential which means that its tail decays as an exponential. We used non-asymptotic statistical analysis to show that the tail distribution of the residual noise on the final post-processed frame indeed decays as an exponential. Compared to the commonly used CLT approach, this non-asymptotic analysis directly estimates the sensitivity of the detection procedure with respect to the number of frames we have in a given dataset and with respect to the temporal correlation of the residuals.

Based on our statistical analysis, we introduced a novel detection map, called STIM map, and studied its theoretical properties. We then used thorough numerical experiments to demonstrate its capabilities on real data. We also provided a method to automatically estimate the detection threshold from any dataset. This detection map thus enables direct and automatic detection in PCA-processed frames with a higher true positive to false positive ratio than state-of-the-art SNR_t map used for exoplanet imaging today. In particular, our performance analysis showed that the proposed STIM map reaches a significant gain in terms of detection, especially at small inner working angles where the poor field rotation and the high speckle noise variance makes it very difficult to extract signals and reveal the presence of fainter signals. This usual limitation is bypassed thanks to the time-domain approach of the STIM map

that, to the best knowledge of the authors, has never been used in the framework of HCI post-processing.

On top of being directly used after speckle subtraction algorithms, our analysis has implications on other algorithms. For instance inverse problems approaches, such as ANDROMEDA or FMMF, rely on a maximum a posteriori (MAP) estimate thus the knowledge of the actual distribution of the residual noise allows to take it into account when solving the problem. Indeed, these authors assume that the residual distribution is Gaussian, which implies the use of an ℓ_2-norm estimation. We have shown in this paper that the decay of the tail is exponential and better accounted for by a Laplace distribution. It turns out that the optimal norm to use in the MAP estimation for the Laplace distribution is an ℓ_1-norm. This solution is being implemented in such methods and published in a forthcoming paper.

Future work would involve studying the specific case of extended features, such as faint debris discs with sharp edges (see e.g. Lee & Chiang 2016, for a gallery) or bright protoplanetary discs showing blunt structures such as spiral arms (see e.g. Benisty et al. 2015, 2017). Another step is to apply this method on images taken using different diversity such as the spectral diversity provided by integral field spectrograph, where the planet extraction and characterization is difficult due to the degeneracy of the obtained signal-to-noise ratio with planetary spectrum.

REFERENCES

Absil O. et al., 2013, A&A, 559, L12
Amara A., Quanz S. P., 2012, MNRAS, 427, 948
Benisty M. et al., 2015, A&A, 578, L6
Benisty M. et al., 2017, A&A, 597, A42
Beuzit J.-L. et al., 2008, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 7735, Ground-based and Airborne Instrumentation for Astronomy II. SPIE, Bellingham, p. 701418
Boccaletti A. et al., 2008, in Hubin N., Max C. E., Wizinowich P. L., eds, Proc. SPIE Conf. Ser. Vol. 7015, Adaptive Optics Systems. SPIE, Bellingham, p. 10
Bonnefoy M. et al., 2014, A&A, 567, L9
Bowler B. P., Nielsen E. L., 2018, Occurence Rates fro Direct Imaging Surveys, Handbook of Exoplanets, 1
Cantalloube F. et al., 2015, A&A, 582, A89
Chauvin G. et al., 2015, A&A, 573, A127
Crossfield I. J. M., 2015, PASP, 127, 941
Currie T., Cloutier R., Brittain S., Grady C., Burrows A., Muto T., Kenyon S. J., Kuchner M. J., 2015, ApJ, 814, L27
Delorme P. et al., 2017a, A&A, 608, A79
Delorme P. et al., 2017b, A&A, 608, A79
Espaillat C. et al., 2014, in Beuther H., Klessen R. S., Dullemond C. P., Henning T., eds, Protostars and Planets VI. University of Arizona Press, Tucson, p. 914
Fitzgerald M. P., Graham J. R., 2006, ApJ, 637, 541
Gonzalez C. G., Absil O., Absil P., Van Droogenbroeck M., Mawet D., Surdej J., 2016, A&A, 598, A54
Gonzalez C. A. G., Absil O., Van Droogenbroeck M., 2018, A&A, 613, A71
Gonzalez C. A. G. et al., 2017b, AJ, 154, 7
Goodman J. W., 1975, in Laser Speckle and Related Phenomena. Springer-Verlag, Berlin, p. 9
Guyon O., 2005, ApJ, 629, 592
Guyon O., Pluzhnik E. A., Kuchner M. J., Collins B., Ridgway S. T., 2006, ApJS, 167, 81
Heiberger R. M., Holland B., 2004, *Statistical Analysis and Data Displays* Springer, Berlin
NIST Digital Library of Mathematical Functions, 2018, Available at: https://dlmf.nist.gov
Hinkley S. et al., 2007, ApJ, 654, 633
Hinz P. M. et al., 2016, in Malbet F., Creech-Eakman M. J., Tuthill P. G., eds, Proc. SPIE Conf. Ser. Vol. 9907, Optical and Infrared Interferometry and Imaging V. SPIE, Bellingham, p. 14
Hughes A. M., Duchene G., Matthews B., 2018, ARA&A, 56, 541
Jovanovic N. et al., 2015, PASP, 127, 890
Konopacky Q. M., Barman T. S., Macintosh B. A., Marois C., 2013, Science, 339, 1398
Lafrenière D., Marois C., Doyon R., Nadeau D., Artigau É., 2007, ApJ, 660, 770
Lagrange A.-M. et al., 2010, Science, 329, 57
Lee E. J., Chiang E., 2016, ApJ, 827, 125
Macintosh B. et al., 2015, Science, 350, 64
Macintosh B. A. et al., 2008, Adaptive Optics Systems, International Society for Optics and Photonics Vol. 7015,, p. 701518
Marois C., Lafreniere D., Doyon R., Macintosh B., Nadeau D., 2006, ApJ, 641, 556
Marois C., Lafreniere D., Macintosh B., Doyon R., 2008, ApJ, 673, 647
Marois C., Macintosh B., Véran J.-P., 2010, Adaptive Optics Systems II, International Society for Optics and Photonics Vol. 7736. p. 77361J
Mawet D., Riaud P., Absil O., Surdej J., 2005, ApJ, 633, 1191
Mawet D. et al., 2012, Space Telescopes and Instrumentation 2012 : Optical, Infrared, and Milimeter Wave, Vol. 8442, p. 844204
Mawet D. et al., 2013, in Shaklan S., ed., Proc. SPIE Conf. Ser. Vol. 8864, Techniques and Instrumentation for Detection of Exoplanets VI. SPIE, Bellingham, p. 9
Mawet D. et al., 2014, ApJ, 792, 97
McLean I. S., Chaffee F. H., 2000, in Iye M., Moorwood A. F., eds, Proc. SPIE Conf. Ser. Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. SPIE, Bellingham, p. 2
Metz C. E., 2006, J. Am. Coll. Radiol., 3, 413
Milli J. et al., 2017, A&A, 597, L2
Papoulis A., Pillai S. U., 2002, Probability, Random Variables, and Stochastic Processes. Tata McGraw-Hill Education
Pueyo L. et al., 2015, ApJ, 803, 31
Quanz S. P., Amara A., Meyer M. R., Girard J. H., Kenworthy M. A., Kasper M., 2015, ApJ, 807, 64
Ruffio J.-B. et al., 2017, ApJ, 842, 14
Samland M. et al., 2017, A&A, 603, A57
Soummer R., 2005, ApJ, 618, L161
Soummer R., Ferrari A., Aime C., Jolissaint L., 2007, ApJ, 669, 642
Soummer R., Pueyo L., Larkin J., 2012, ApJ, 755, L28
Vershynin R., 2010, preprint (arXiv:1011.3027)
Wainwright M. J., 2019 , High-dimensional Statistics: A Non-asymptotic Viewpoint, Vol. 48. Cambridge Univ. Press, Cambridge

Table A1. Description of the three datasets used in this paper. The total number of images constituting the data cube is noted N_{images}. The integration time used for each exposure of this data cube is noted t_{int}. The average turbulence coherence time during the observation is noted t_{coh}.

Name	Date	Instrument (VLT)	Coronagraph	Filter	N_{images}	Total field rotation (deg)	t_{int} (s)	t_{coh} (ms)
β-pic	Feb 2013	NACO	AGPM	Lp (3.5–4.1 μm)	612	83.0	8	2.1
HD 206893	Oct 2015	SPHERE-IRDIS	APLC	H (1.48–1.77 μm)	570	49.3	4	2.8
51 Eri	Sep 2015	SPHERE-IRDIS	APLC	K1 (2.0–2.2 μm)	256	41.6	16	6.2

APPENDIX A: PRESENTATION OF THE DATASET USED IN THIS PAPER

Throughout this paper, we used three different representative dataset to test our approach: (1) Beta Pictoris (HIP 27321) observations using the VLT/NaCo instrument, published in Absil et al. (2013) (ESO program ID 60.A-9800), (2) HD 206893 (HIP 107412) observations using the VLT/SPHERE-IRDIS instrument, published in Milli et al. (2017) (ESO program ID 96.C-0388), and (3) 51 Eridani (HIP 21547) observations using the VLT/SPHERE-IRDIS instrument, published in Samland et al. (2017) (ESO program ID 095.C-0298). Each target has one companion discovered by imaging: (1) β-pic b (Lagrange et al. 2010) is a 8–15 Jupiter mass exoplanet orbiting its host star at 5–10 au, (2) HD 206893 b (Milli et al. 2017) is a 12–50 Jupiter mass companion orbiting its host star at 10–15 au (Delorme et al. 2017b), and (3) 51 Eri b (Macintosh et al. 2015) is a 2–10 Jupiter mass companion orbiting its host star at 10–14 au (Samland et al. 2017).

Table A1 gathers the main information about each dataset taken with instrument located at the ESO Paranal observatory. The three dataset make use of a coronagraph: the β-pic data are taken with an Annular Groove Phase Mask (AGPM) vector vortex coronagraph (Mawet et al. 2005) optimized in *L*' band (Mawet et al. 2013) and the two SPHERE datasets are taken using the Apodized Lyot Coronagraph (APLC, Soummer 2005) optimized for the YJH bands (focal mask diameter of 185 mas) and including a Lyot stop (Boccaletti et al. 2008).

APPENDIX B: PROOF OF SUB-EXPONENTIALITY OF MR DISTRIBUTION

We prove Proposition 1, i.e. let $X \sim \text{MR}(\alpha, \beta)$ with

$$\text{MR}(\alpha, \beta) \sim \frac{1}{\beta} \exp\left(-\frac{t+\alpha}{\beta}\right) \mathcal{I}_0\left(\frac{2\sqrt{t\alpha}}{\beta}\right),$$

then X is sub-exponential with $\|X\|_{\psi_1} \lesssim 6\beta$.

Proof. The moments of X are given by

$$\mathbb{E}X^p = \int_0^{+\infty} t^p \frac{1}{\beta} \exp\left(-\frac{t+\alpha}{\beta}\right) \mathcal{I}_0\left(\frac{2\sqrt{t\alpha}}{\beta}\right) \, \mathrm{d}t,$$

by change of variable $t = s^2$, $\mathrm{d}t = 2s\,\mathrm{d}s$, we get

$$\mathbb{E}X^p = \int_0^{+\infty} s^{2p} \left(\frac{2s}{\beta} \exp\left(-\frac{s^2+\alpha}{\beta}\right) \mathcal{I}_0\left(\frac{2s\sqrt{\alpha}}{\beta}\right)\right) \, \mathrm{d}s$$

$$= \frac{1}{\sqrt{2}} \mathbb{E}Y^{2p},$$

where Y follows a Rice distribution (Papoulis & Pillai 2002):

$$Y \sim \text{Rice}(\nu, \sigma) \sim \frac{s}{\sigma^2} \exp\left(-\frac{s^2+\nu^2}{2\sigma^2}\right) \mathcal{I}_0\left(\frac{s\nu}{\sigma^2}\right)$$

with $\nu = \sqrt{\alpha}$ and $\sigma = \sqrt{\frac{\beta}{2}}$. We thus have to prove that $\frac{1}{p}(\mathbb{E}Y^{2p})^{1/p}$ is bounded. The raw moment of Y are given by[3]

$$\mathbb{E}Y^{2p} = \sigma^{2p} 2^p \Gamma(1+p) L_p\left(-\frac{\nu^2}{2\sigma^2}\right),$$

where $\Gamma(\cdot)$ is the gamma function and $L_\gamma(\cdot)$ is the Laguerre polynomial of degree γ.

Since for $x > 0$, $\Gamma(x) < \sqrt{2\pi} e^{-(11/12)x} x^{x-1/2}$ [section 5.6] (NIST), we find, for $p \geq 1$,

$$\Gamma(1+p) < \sqrt{2\pi}(1+p)^{\frac{1}{2}+p},$$

where we use the fact that $e^{-(11/12)x} < 1$. Moreover, for $n \in \mathbb{N}$, $L_n(\cdot)$ [equation 18.14.8] (NIST):

$$|L_n(x)| \leqslant \frac{1}{n!} \exp(x/2),$$

Gathering all these observations yields:

$$\mathbb{E}(Y^{2p}) < \left(\sigma^{2p} \frac{2^p}{p!} \sqrt{2\pi}(1+p)^{\frac{1}{2}+p} \exp\left(-\frac{\nu^2}{4\sigma^2}\right)\right).$$

Thus for $p \geq 1$,

$$p^{-1}(\mathbb{E}Y^{2p})^{\frac{1}{p}} < p^{-1}\left(\sigma^{2p} \frac{2^p}{p!} \sqrt{2\pi}(1+p)^{\frac{1}{2}+p}\right)^{\frac{1}{p}}.$$

Recalling that $\mathbb{E}Y^{2p} = \sqrt{2}EX^p$, we get

$$p^{-1}(\mathbb{E}X^p)^{\frac{1}{p}} < p^{-1}\left(\sigma^{2p} \frac{2^p}{p!} \sqrt{\pi}(1+p)^{\frac{1}{2}+p}\right)^{\frac{1}{p}}$$

$$< p^{-1}\left(\sigma^{2p} \frac{2^p}{p!} \sqrt{\pi}(1+p)^{\frac{1}{2}}(1+p)^p\right)^{\frac{1}{p}}$$

$$< \sigma^2 2\sqrt{\pi}^{\frac{1}{p}} \left(\frac{\sqrt{1+p}}{p!}\right)^{\frac{1}{p}} (1/p + 1).$$

To find a bound for $\left(\frac{\sqrt{1+p}}{p!}\right)^{\frac{1}{p}}$, we consider two cases. First, the case $\frac{\sqrt{1+p}}{p!} \leqslant 1$, then $\left(\frac{\sqrt{1+p}}{p!}\right)^{\frac{1}{p}} \leqslant 1$. Otherwise, $\left(\frac{\sqrt{1+p}}{p!}\right)^{\frac{1}{p}} \leqslant \frac{\sqrt{1+p}}{p!} \leqslant \frac{\sqrt{1+p}}{p} = \sqrt{1/p^2 + 1/p}$. Then we observe that $\sqrt{1/p^2 + 1/p}$ is a monotonically decreasing function reaching its maximum for $p = 1$ with value $\sqrt{2}$. Hence in both cases, we conclude $\left(\frac{\sqrt{1+p}}{p!}\right)^{\frac{1}{p}} \leqslant \sqrt{2}$. For $(1/p + 1)$, we also use the fact that it is a monotonically decreasing function reaching its maximum for $p = 1$ with value 2. Collecting these results and observing that $\sqrt{\pi}^{1/p} \leqslant \sqrt{\pi}$, we get

$$\|X\|_{\psi_1} \leqslant \sigma^2 4\sqrt{\pi}\sqrt{2} = \beta 2\sqrt{2\pi} < 5.02\beta < 6\beta.$$

Hence we have shown that $X \sim \text{MR}(\alpha, \beta)$ is sub-exponential with $\|X\|_{\psi_1} < 6\beta$. $\square$

[3]https://reference.wolfram.com/language/ref/RiceDistribution.html

APPENDIX C: PROOF OF THE CONCENTRATION OF MEASURE SLOWDOWN INDUCED BY NON-INDEPENDENCE

We prove Corollary 1, we first present a proposition.

PROPOSITION 2.   *Assume $B = \sum_{i=1}^{\Delta} A_i$ and $\epsilon > 0$. Then*

$$P(|B| \geqslant \epsilon) \leqslant \Delta \max_i P(|A_i| \geqslant \epsilon/\Delta). \tag{C1}$$

Proof. *We have*

$$P\left(|\sum_{i=1}^{\Delta} A_i| \geqslant \epsilon\right) \leqslant P\left(\sum_{i=1}^{\Delta} |A_i| \geqslant \epsilon\right) \tag{C2}$$

$$\leqslant P\left(\sum_{i=1}^{\Delta} \max_i |A_i| \geqslant \epsilon\right) \tag{C3}$$

$$= P(\Delta \max_i |A_i| \geqslant \epsilon). \tag{C4}$$

Using a union bound, one gets

$$P\left(|\sum_{i=1}^{\Delta} A_i| \geqslant \epsilon\right) \leqslant \Delta \max_i P\left(|A_i| \geqslant \epsilon/\Delta\right).$$

Corollary 1 follows from Proposition 2. We assume we have a dependence length of τ. We put the elements of the trajectory into subgroups of $\nu = T/\tau$ independent elements.

We write $B = \sum_{i=1}^{\tau} A_i$, where $A_i = \sum_{j=1}^{\nu} s_{(i+j\cdot\tau)}^{[g]}$. Then we apply Proposition 2 with $\Delta = \tau$ and $\epsilon \leftarrow \epsilon T$. We get

$$P\left(|\hat{\mu}_g| \geqslant \epsilon\right) = P\left(|\sum_{i=1}^{\tau} \sum_{j=1}^{\nu} s_{(i+j\cdot\tau)}^{[g]}| \geqslant \epsilon T\right) \tag{C5}$$

$$\leqslant \tau \max_i P\left(|\sum_{j=1}^{\nu} s_{(i+j\cdot\tau)}^{[g]}| \geqslant \epsilon T/\tau\right). \tag{C6}$$

For a given i, the $s_{(i+j\cdot\tau)}^{[g]}$ are i.i.d. sub-exponential random variables and hence we can apply Theorem 1 and get

$$P\left(|\hat{\mu}_g| \geqslant \epsilon\right) \leqslant 2\exp\left(-c\min(K^2/\sigma^2, \epsilon/K)T/\tau\right). \tag{C7}$$

APPENDIX D: COMPARISON BETWEEN THE MEDIAN AND THE MEAN

In order to compare the capability of the STIM when using the mean (STIMean) or the median (STIMedian), we reproduce two figures from the text.
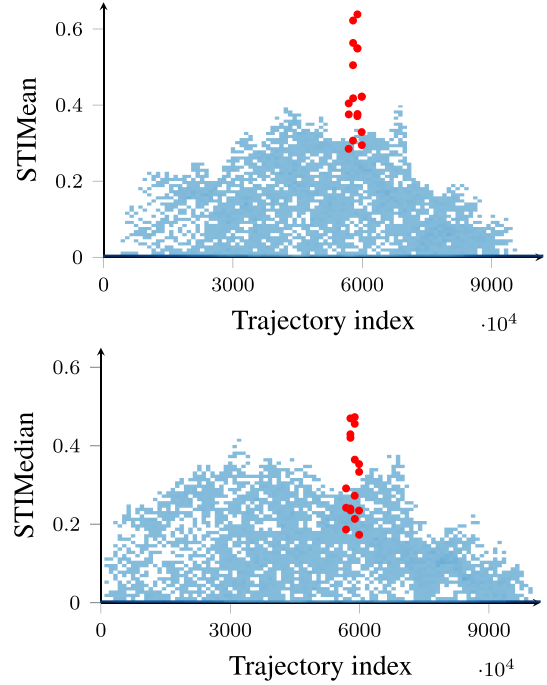

Figure D1. Reproduction of Fig. 7 for the mean (top) and for the median (bottom). The planetary signal is found to emerge from the noise more clearly for the mean than for the median.

In Fig. D1, we compare the capability of the STIMean and the STIMedian on HD 20689. We can see that the median performs poorly compared to the mean.

We also reproduce the ROC curves of section with the STIMedian added in Fig. D2. We do not observe any significant difference between the detection capabilities of the two maps.
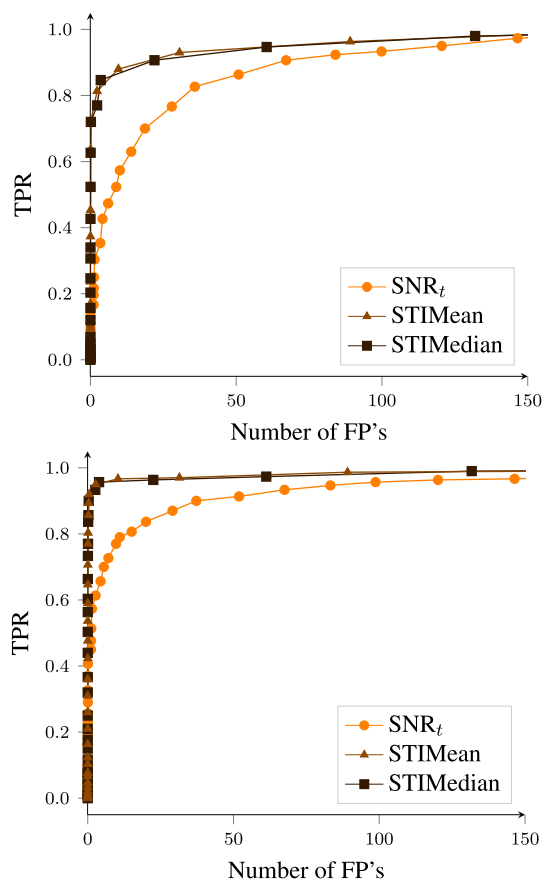
Figure D2. Reproduction of Fig. 10, with the STIMedian map added. Top: radial separation of 250 mas with contrast between 2.75×10^{-5} and 4.12×10^{-5}. Bottom: radial separation of 375 mas with contrast between 1.37×10^{-5} and 2.75×10^{-5}. We do not observe a significant difference between the use of the median or the mean in the computation of the STIM map.

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