UCLA
UCLA Previously Published Works

Title
Improving and Assessing Planet Sensitivity of the GPI Exoplanet Survey with a Forward Model Matched Filter

Authors
Ruffio, JB
Macintosh, B
Wang, JJ
et al.

Improving and Assessing Planet Sensitivity of the GPI Exoplanet Survey with a Forward Model Matched Filter

Jean-Baptiste Ruffio[1], Bruce Macintosh[1], Jason J. Wang[2], Laurent Pueyo[3], Eric L. Nielsen[1,4], Robert J. De Rosa[2], Ian Czekala[1], Mark S. Marley[5], Pauline Arriaga[6], Vanessa P. Bailey[1], Travis Barman[7], Joanna Bulger[8], Jeffrey Chilcote[9], Tara Cotten[10], Rene Doyon[11], Gaspard Duchêne[2,12], Michael P. Fitzgerald[6], Katherine B. Follette[1], Benjamin L. Gerard[13,14], Stephen J. Goodsell[15], James R. Graham[2], Alexandra Z. Greenbaum[16], Pascale Hibon[17], Li-Wei Hung[6], Patrick Ingraham[18], Paul Kalas[2,4], Quinn Konopacky[19], James E. Larkin[6], Jérôme Maire[19], Franck Marchis[4], Christian Marois[13,14], Stanimir Metchev[20,21], Maxwell A. Millar-Blanchaer[22,30], Katie M. Morzinski[23], Rebecca Oppenheimer[24], David Palmer[25], Jennifer Patience[26], Marshall Perrin[3], Lisa Poyneer[25], Abhijith Rajan[26], Julien Rameau[11], Fredrik T. Rantakyrö[17], Dmitry Savransky[27], Adam C. Schneider[26], Anand Sivaramakrishnan[3], Inseok Song[10], Remi Soummer[3], Sandrine Thomas[18], J. Kent Wallace[22], Kimberly Ward-Duong[26], Sloane Wiktorowicz[28], and Schuyler Wolff[29]

[1] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA, 94305, USA; jruffio@stanford.edu
[2] Astronomy Department, University of California, Berkeley; Berkeley CA, 94720, USA
[3] Space Telescope Science Institute, Baltimore, MD, 21218, USA
[4] SETI Institute, Carl Sagan Center, 189 Bernardo Avenue, Mountain View, CA, 94043, USA
[5] NASA Ames Research Center, Mountain View, CA, 94035, USA
[6] Department of Physics & Astronomy, University of California, Los Angeles, CA, 90095, USA
[7] Lunar and Planetary Laboratory, University of Arizona, Tucson AZ, 85721, USA
[8] Subaru Telescope, NAOJ, 650 North A'ohoku Place, Hilo, HI 96720, USA
[9] Dunlap Institute for Astronomy & Astrophysics, University of Toronto, Toronto, ON, M5S 3H4, Canada
[10] Department of Physics and Astronomy, University of Georgia, Athens, GA, 30602, USA
[11] Institut de Recherche sur les Exoplanètes, Départment de Physique, Université de Montréal, Montréal QC, H3C 3J7, Canada
[12] Univ. Grenoble Alpes/CNRS, IPAG, F-38000 Grenoble, France
[13] University of Victoria, 3800 Finnerty Road, Victoria, BC, V8P 5C2, Canada
[14] National Research Council of Canada Herzberg, 5071 West Saanich Road, Victoria, BC, V9E 2E7, Canada
[15] Gemini Observatory, 670 N. A'ohoku Place, Hilo, HI, 96720, USA
[16] Department of Astronomy, University of Michigan, Ann Arbor MI, 48109, USA
[17] Gemini Observatory, Casilla 603, La Serena, Chile
[18] Large Synoptic Survey Telescope, 950N Cherry Avenue, Tucson, AZ, 85719, USA
[19] Center for Astrophysics and Space Science, University of California San Diego, La Jolla, CA, 92093, USA
[20] Department of Physics and Astronomy, Centre for Planetary Science and Exploration, The University of Western Ontario, London, ON N6A 3K7, Canada
[21] Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY 11794-3800, USA
[22] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91125, USA
[23] Steward Observatory, University of Arizona, Tucson AZ, 85721, USA
[24] Department of Astrophysics, American Museum of Natural History, New York, NY, 10024, USA
[25] Lawrence Livermore National Laboratory, Livermore, CA, 94551, USA
[26] School of Earth and Space Exploration, Arizona State University, P.O. Box 871404, Tempe, AZ, 85287, USA
[27] Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY, 14853, USA
[28] Department of Astronomy, UC Santa Cruz, 1156 High Street, Santa Cruz, CA, 95064, USA
[29] Department of Physics and Astronomy, Johns Hopkins University, Baltimore, MD, 21218, USA

Abstract

We present a new matched-filter algorithm for direct detection of point sources in the immediate vicinity of bright stars. The stellar point-spread function (PSF) is first subtracted using a Karhunen-Loéve image processing (KLIP) algorithm with angular and spectral differential imaging (ADI and SDI). The KLIP-induced distortion of the astrophysical signal is included in the matched-filter template by computing a forward model of the PSF at every position in the image. To optimize the performance of the algorithm, we conduct extensive planet injection and recovery tests and tune the exoplanet spectra template and KLIP reduction aggressiveness to maximize the signal-to-noise ratio (S/N) of the recovered planets. We show that only two spectral templates are necessary to recover any young Jovian exoplanets with minimal S/N loss. We also developed a complete pipeline for the automated detection of point-source candidates, the calculation of receiver operating characteristics (ROC), contrast curves based on false positives, and completeness contours. We process in a uniform manner more than 330 data sets from the Gemini Planet Imager Exoplanet Survey and assess GPI typical sensitivity as a function of the star and the hypothetical companion spectral type. This work allows for the first time a comparison of different detection algorithms at a survey scale accounting for both planet completeness and false-positive rate. We show that the new forward model matched filter allows the detection of 50% fainter objects than a conventional cross-correlation technique with a Gaussian PSF template for the same false-positive rate.

Key words: instrumentation: adaptive optics – methods: statistical – planetary systems – surveys – techniques: high angular resolution – techniques: image processing

---

[30] NASA Hubble Fellow.

1. Introduction

Direct-imaging techniques spatially resolve exoplanets from their host star by using high-contrast imaging instruments, usually combined with the power of large telescopes, adaptive optics, coronagraphs, and sophisticated data processing. This technique currently allows the detection of young (<300 Myr), massive ($>2\,M_{\mathrm{Jup}}$), self-luminous exoplanets at host-star separations that are not yet covered by indirect methods ($a > 5$ au), and it therefore helps to constrain planet population statistics. The Gemini Planet Imager (GPI) (Macintosh et al. 2014), operating on the Gemini South telescope, is one of the latest generation of high-contrast instruments with extreme adaptive optics. The GPI Exoplanet Survey (GPIES) is targeting 600 young stars and has to date observed more than half of them. As part of the survey, it imaged several known systems and discovered the exoplanet 51 Eridani b (Macintosh et al. 2015).

High-contrast images suffer from spatially correlated noise, called speckles, which originates from optical aberrations in the instrument, and also from a diffuse light component resulting from the time-averaged uncorrected atmospheric turbulence. The correlation length of the speckles in a raw image is equal to the size of the unocculted point-spread function (PSF). Speckles are often described as quasi-static, since they are correlated across the observed spectral band and across the exposures of a typical observation (Bloemhof et al. 2001; Sivaramakrishnan et al. 2002; Perrin et al. 2003). It has been shown that the probability density function (PDF) of the speckle noise is not Gaussian, but rather better described by a modified Rician distribution (Bloemhof et al. 2004; Soummer & Aime 2004; Fitzgerald & Graham 2006; Hinkley et al. 2007; Soummer et al. 2007; Marois et al. 2008; Mawet et al. 2014). The Rician PDF has a larger positive tail than a Gaussian distribution, yielding a comparatively higher number of false positives at constant signal-to-noise ratio (S/N).

In a single image, disentangling the signal of a faint planet buried under speckle noise is challenging because the spatial scale of speckle noise and the signal from a planet are similar, corresponding to the size of the PSF. However, speckles and astrophysical signals behave differently with time and wavelength. This diversity can be used to build a model of the speckle pattern and subtract it from all images. Two observing strategies are commonly used for speckle subtraction with instruments such as the GPI, angular differential imaging (ADI) (Marois et al. 2006), and spectral differential imaging (SDI) (Marois et al. 2000; Sparks & Ford 2002). The observing setup for ADI is different from traditional imaging with altitude/azimuth telescopes, as the instrument field derotator is switched off or adjusted to keep the telescope pupil fixed with respect to sky rotation. As a consequence, the astrophysical signal rotates on the detector, following the sky rotation with respect to the telescope (i.e., parallactic angle evolution), while the speckle pattern remains relatively stable. In a similar tactic, SDI exploits the radial linear dependence of the speckle pattern with wavelength to separate it from the planet signal, which remains at the same position at all wavelengths. By definition, an observation with an integral field spectrograph (IFS), such as the GPI, provides both temporal and spectral diversity of the speckle pattern for ADI or SDI to be used separately or in tandem.

The most common spectral subtraction algorithms are a locally optimized combination of images (LOCI), Lafrenière

et al. (2007a), which uses a least-squares approach to optimally subtract the speckle noise, and Karhunen-Loéve image processing (KLIP), Soummer et al. (2012), which regularizes the least-squares problem by filtering out the high-order singular modes. Following speckle subtraction, point sources can be searched for by using a matched filter or a more general Bayesian model comparison framework, as shown in Kasdin & Braems (2006). However, the distortion of the planet PSF caused by the speckle subtraction algorithm, referred to as self-subtraction, can make it difficult to define an accurate matched-filter template. The self-subtraction can be accurately modeled in simplified cases, as shown in the matched-filter approach by Cantalloube et al. (2015) using ADI-subtracted pair images. In the context of planet characterization, the self-subtraction also biases the photometry and the astrometry of the object. The inverse problem is usually solved by injecting negative planets in the raw images and iteratively minimizing the image residuals (Marois et al. 2010; Morzinski et al. 2015). Recently, Pueyo (2016) derived a closed-form approximation of the self-subtraction in KLIP, but without applying it in the context of a matched filter. Wang et al. (2016) used this new forward model in a Bayesian framework to estimate the astrometry of β Pictoris b.

The main challenge with uniformly and systematically characterizing the detections of a high-contrast imaging exoplanet survey is the high number of false positives even at relatively high S/N. The detection threshold is hard to define because the PDF of the residual noise is generally unknown and depends on the instrument, the choice of data processing, and the data set itself. The lack of well-known false-positive rates makes it very difficult to evaluate the performance of algorithms relative to one another. Currently, candidates are discarded as false positives by visual inspection, which does not permit a rigorous calculation of planet completeness. In order to accurately characterize planet-detection statistics and ultimately constrain the underlying planet population in a uniform and unbiased manner, it is important to improve and systematize exoplanet detection methods.

The goal of this paper is to define a systematic and rigorous approach for exoplanet detection in large direct-imaging surveys such that the long-period exoplanet population can be inferred in a meaningful manner. Using the latest KLIP framework, we develop an automated matched-filter-based detection algorithm that includes a forward model of the planet self-subtraction (e.g., Pueyo 2016) and accounts for the noise variations in the spatial, temporal, and spectral dimensions of a data set. As of the end of 2016, GPIES has already observed 330 stars, which allows us to precisely estimate the false-positive rate and define meaningful detection thresholds for the entire survey. We conduct a rigorous set of tests to characterize state-of-the-art detection algorithms and demonstrate that a forward model matched filter (FMMF) most effectively recovers a planet signal while reducing the number of false-positive detections. The paper is structured as follows.

1. GPIES observations and data reduction are presented in Section 2,
2. the matched filter is described in Section 3,
3. the optimization of the reduction parameters for GPIES is presented in Section 4,
4. the residual noise is characterized for the different algorithms in Section 5, including the calculation of receiver operating characteristics (ROC),

5. the detection sensitivity as a function of separation, referred to as the contrast curve, is calculated in Section 6, where the contrast is defined as the companion-to the host-star brightness ratio in a spectral band,

6. the follow-up strategy and vetting of point-source candidates is discussed in Section 7,

7. the contours of the planet completeness, which is the fraction of planets that could have been detected, are then derived in Section 8, and finally,

8. we conclude in Section 9.

We refer any reader who is not familiar with the data processing of high-contrast images to Appendix A, which includes a detailed description of KLIP, the matched filter, and the planet PSF forward model. The mathematical notations are summarized in Appendix B.

2. Observations and Data Reduction

2.1. Observations

In this paper, we use 330 observations from the GPI Exoplanet Survey (GPIES) (Gemini programs GS-2014B-Q-500, GS-2015A-Q-500, and GS-2015B-Q-500; PI: B. Macintosh) to construct and test our matched filter. A typical GPIES epoch consists of 38 one-minute exposures in H band (1.5–1.8 μm). The number of usable raw spectral cubes can be lower because of degrading weather condition or isolated star tracking failures. We arbitrarily consider any data set with more than 20 usable exposures as complete. In some cases, a few exposures were added to the observing sequence in order to compensate for poor conditions, which resulted in a number of data sets with more than 38 exposures. For each star, we only consider the first complete epoch and ignore any follow-up observations that may have been made. We have not considered data sets with visible debris disks in order to avoid biasing the contrast curves. In H band, a GPI spectral cube has 37 wavelength channels and 281×281 pixels in the spatial dimensions, half of which are not filled with data, however, because of the tilted IFS field of view. Therefore, a typical GPIES data set includes approximately 1400 images at different position angles and wavelengths.

2.2. Raw Data Reduction

Spectral data cubes are built from raw IFS detector images using standard recipes from the GPI Data Reduction Pipeline[31] version 1.3 and 1.4 (Perrin et al. 2016). The process includes correction for dark current, bad pixels, correlated read noise, and microphonics induced by cryocooler vibration (Ingraham et al. 2014). Flexure in the instrument slightly shifts the position of the lenslet micro-spectra on the detector. The offset is calibrated using argon arc lamp images at the target elevation, which are then compared with wavelength solution references taken at zenith (Wolff et al. 2014). Finally, each spectral cube is also corrected for optical distortion according to Konopacky et al. (2014). GPI images also contain four fainter copies of the unblocked PSF called satellite spots (Marois et al. 2006; Sivaramakrishnan & Oppenheimer 2006). The spots are used to estimate the location of the star behind the focal plane mask, and their flux allows the photometric calibration of the images. The GPI satellite spot-to-star flux

ratio used here is 2.035×10^{-4} for H band (Maire et al. 2014). In the following, the satellite spots are also median combined to estimate an empirical planet PSF that is wavelength dependent.

2.3. Use of Simulated Planets

Simulated planets are necessary to optimize and characterize a detection algorithm because real point sources in high-contrast images are scarce. We decided to neglect for now the PSF smearing that is due to the sky rotation in a single exposure. The planetary spectra are selected from atmospheric models described in M. S. Marley et al. (2017, in preparation) and Saumon et al. (2012). For a given cloud coverage, the only model parameter with a significant effect on the shape of the spectrum in a single band is the temperature. However, this work is not about atmospheric characterization, so physically unrealistic model temperatures for a given object are not an issue as long as the shape of the spectrum matches. In the following, the references to T-type and L-type planets correspond to the analogous spectra of brown dwarfs. T-type spectra have strong methane absorption features, while L-type spectra are cloudy objects whose spectra are dominated by H_2O and CO. Of the known extrasolar planets, 51 Eridani b is an example of the T-type (Macintosh et al. 2015) and β Pictoris b of the L-type (Morzinski et al. 2015). The transmission spectrum of the Earth's atmosphere combined with that of the instrument is estimated by dividing the satellite spot spectrum with a stellar spectrum, which is then used to translate the spectrum from physical units into raw pixel values from the detector. The spectrum of the star is interpolated from the Pickles atlas, Pickles (1998), based on the star's spectral type.

2.4. Speckle Subtraction with KLIP

Each individual image in the data set is speckle subtracted using a Python implementation of KLIP called PyKLIP[32] (Wang et al. 2015). The KLIP algorithm consists of building and subtracting a model of the speckle pattern in an image, called science image, from a set of reference images that can be selected from the same data set or from completely different observations. In this paper, we use a combination of ADI and SDI strategies. The KLIP mathematical framework is summarized in Appendix A.1.1. A side effect of the speckle subtraction is the distortion of the planet PSF, referred to as self-subtraction. The self-subtraction is fully characterized in Appendix A.1.2.

All individual images are first high-pass filtered by subtracting a Gaussian-convolved image with a full-width at half-maximum (FWHM) of 12 pixels. Then, they are aligned and the reference images are scaled to the same wavelength as the science image. In order to account for spatial variation of the speckle behavior, KLIP is independently applied on small subsections of the field of view. Each image is therefore divided into small 100 pixel arcs to which a 10 pixel wide padding is added, as illustrated in Figure 1. For each sector, the reference library is built according to an exclusion criterion based on the displacement and the flux overlap (Marois et al. 2014) of the planet PSF between the science image and its reference images. The exclusion criterion and the reference library selection is described in Appendix A.1.3. The assumed spectrum for the companion used in the exclusion parameter

---
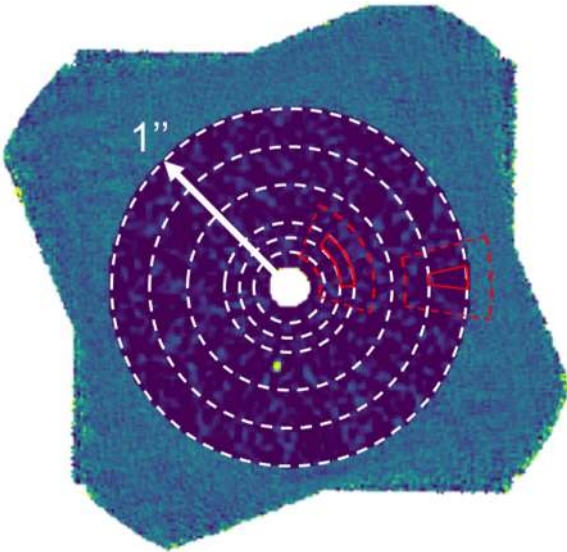
Figure 1. Definition of the padded sectors dividing the image. The annuli boundaries used for the unpadded sectors are shown as white dashed circles. The first three annuli are thinner to account for the rapidly varying noise standard deviation close to the focal plane mask. Examples of sectors are drawn in red with the solid line containing 100 pixels and the dashed line delimiting the padded area. The outer working angle has been set to 1″.

is referred to as the reduction spectrum. In order to speed up the reduction, we only include the $N_R = 150$ most correlated images. This means that most of the images satisfying the exclusion criterion are in practice not used. In Section 4.1 it is shown that the exclusion criterion has a soft maximum around 0.7 for T-type planets, which is the value used in the following. The number of Karhunen-Loève modes of the reference library kept for the speckle subtraction is set to $K = 30$. This value has been chosen as a reasonable guess based on our experience, but it should be rigorously optimized in the future. In order to limit computation time, we have arbitrarily defined the outer working angle of the algorithm to 1″ ($\approx$71 pixels).

3. Matched Filter

3.1. Concept

In the field of signal processing, a matched filter is the linear filter maximizing the S/N of a known signal in the presence of additive noise (Kasdin & Braems 2006; Röver 2011; Cantalloube et al. 2015). A detailed description of the matched filter is presented in appendix Appendix A.2 and we only summarize the key results here. If the noise samples are independent and identically distributed, the matched filter corresponds to the cross-correlation of a template with the noisy data. In the context of high-contrast imaging, the pixels are neither independent nor identically distributed (i.e., heteroskedastic), which introduces a local noise normalization in the expression of the matched filter.

In a data set, each image is indexed by its exposure number τ and its wavelength λ. We define the vector $\boldsymbol{p}_l$, with $l = (\tau, \lambda)$, as a specific speckle-subtracted image. Similarly, we define the matched-filter template $\boldsymbol{m}_l$ as the model of the planet signal in the corresponding processed image normalized such that it has the same broadband flux as the star. The whitening effect of the speckle subtraction allows one to assume uncorrelated residual

noise, which significantly simplifies the matched filter. However, this assumption is not perfectly verified, and its consequence is discussed in Section 3.6. The maximum likelihood estimate of the planet contrast at separation ρ and position angle θ is then given by

$$\tilde{\epsilon}(\rho, \theta) = \frac{\sum_l \boldsymbol{p}_l^\top \boldsymbol{m}_l / \sigma_l^2}{\sum_l \boldsymbol{m}_l^\top \boldsymbol{m}_l / \sigma_l^2}, \tag{1}$$

where σ_l is the local standard deviation at the position (ρ, θ), assuming that it is constant in the neighborhood of the planet. Note that the planet model $\boldsymbol{m}_l$ approaches zero rapidly when moving away from its center (ρ, θ), allowing one to only consider postage-stamp sized images containing the putative planet instead of the full images in Equation (1). Then, the theoretical S/N of the planet can be written

$$\mathcal{S}(\rho, \theta) = \frac{\sum_l \boldsymbol{p}_l^\top \boldsymbol{m}_l / \sigma_l^2}{\sqrt{\sum_l \boldsymbol{m}_l^\top \boldsymbol{m}_l / \sigma_l^2}}. \tag{2}$$

A detection can be claimed when the S/N is such that the observation cannot be explained by the null-hypothesis.

3.2. Matched-filter Template and Forward Model

The calculation of the matched-filter template $\boldsymbol{m}_l$ is complicated by the distortion of the planet PSF after speckle subtraction. The characteristic effect of the self-subtraction manifests as negative wings on each side of the planet PSF and a narrowing of its central peak. It is common practice to define the matched-filter template as a 2D Gaussian, but this model fails to include most of the information about the distortion. In this paper, we present a novel matched-filter implementation using KLIP and the forward model of the planet PSF from Pueyo (2016) illustrated in Figure 2. The forward model is a linearized closed-form estimate of the distorted PSF, which is detailed in Appendix A.3. It is built from the same satellite spot-based PSF as the simulated planet injection in Section 2.3 and from the reduction spectrum. The forward model is a function of the PSF location on the image and therefore needs to be calculated for each pixel.

In this paper, we define the throughput as the ratio of the estimated contrast of a planet after speckle subtraction over its original contrast. This definition might differ from the conventional idea of throughput in the data processing community for direct imaging. It is a measure of the ability of an algorithm to recover an unbiased contrast estimate of the point source. This can then be used to validate our implementation of the forward model. A low throughput indicates that the planet signal has been distorted during the speckle subtraction. The flux of each point source is estimated with Equation (1) and then compared to the known true contrast of the simulated planet. Figure 3 compares the throughput as a function the exclusion criterion, when using either the forward model or only the original PSF as a template. The estimation algorithm is otherwise identical in both cases. Figure 3 demonstrates that the forward model accurately models the self-subtraction because the throughput is close to one even for very aggressive reduction, while it drops very quickly to zero with the original PSF. However, as the exclusion parameter becomes too small, the linear approximation of the forward model breaks down and the throughput
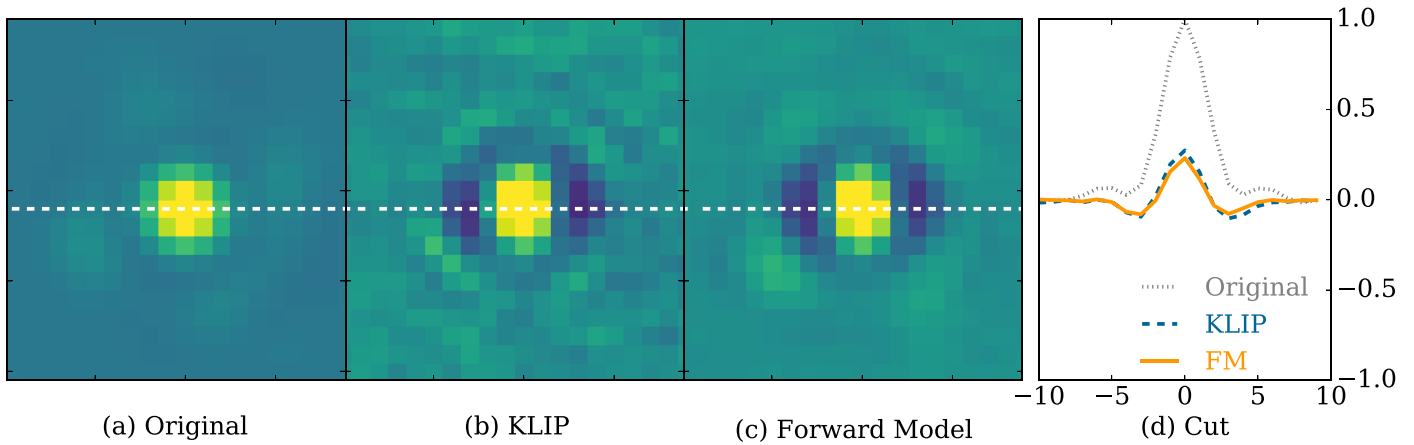
Figure 2. KLIP self-subtraction and forward-modeled PSF. The three panels from left to right represent (a) the original broadband PSF calculated from the GPI satellite spots, (b) the speckle-subtracted image of a simulated planet using KLIP, and (c) the KLIP forward model of the PSF calculated at the position of the simulated planet. All three images are collapsed in time and wavelength and have been scaled to their peak value. The last panel (d) includes horizontal cuts of the different PSF in (a)–(c). The negative ring and negative lobes around the central peak are characteristic of the self-subtraction. The shape and the final amplitude of the PSF are both successfully recovered by the forward model. The simulated planet was injected in 38 cubes of the 51 Eridani b GPIES discovery epoch on 2014 December 18, which is characterized by remarkably stable observing conditions. The star is located approximately 30 pixels (0″4) above the planet and is not visible here.

drops. As expected, the scatter in Figure 3 decreases with lower values of the exclusion criterion. The test was performed with two representative data sets at two different separations: (a) HR 7012 at 0″3 and 0″6, and (b) HD 131435 at 0″4 and 0″8. The data sets were chosen to represent different regimes of data quality as measured by their exoplanet contrast sensitivity in the fully reduced image; HR 7012 is an example of a good data set, while HD 131435 is of medium quality. A total of 35 simulated exoplanets were injected at each separation in seven copies of the same data set. For each copy of the data set, the five simulated planets, which are 72° apart, were rotated by 10° in position angle to better sample the image.

The new algorithm is compared to two simpler matched filters, all using the same KLIP implementation, but involving a simple Gaussian PSF and no modeling of the planet self-subtraction. The three algorithms are referred to in this paper as Gaussian cross-correlation (GCC), Gaussian matched filter (GMF), and forward model matched filter (FMMF). The three methods are detailed in this section and illustrated in Figure 4. The different methods also differ in the way the data set is collapsed before the matched filter is performed. It is beyond the scope of this paper to evaluate the effect of each particular difference. All the algorithms presented in this paper are available in PyKLIP (Wang et al. 2015). A significant fraction of the code is shared with the Bayesian KLIP-FM Astrometry (BKA) method developed in Wang et al. (2016).

3.3. Gaussian Cross-correlation

The Gaussian cross-correlation (GCC) is the baseline algorithm because it is commonly used in high-contrast imaging. The overlapping sectors are mean combined after speckle subtraction in order to limit edge effects. Then, the processed data set is first derotated, coadded, and then collapsed using the reduction spectrum. The resulting image is cross-correlated with a 2D Gaussian kernel with a FWHM of 2.4 pixels. The size of the Gaussian has been chosen to be significantly smaller than the FWHM of the original PSF ($\approx$3.5 pixels) to account for the self-subtraction. The cross-correlation is nothing more than a matched filter where the noise is spatially identically distributed. This assumption is not verified for GPI images, as discussed in

Section 5.1. In practice, the noise properties only need to be azimuthally uniform; the theoretical S/N from the matched filter is always rescaled as a function of separation by estimating the standard deviation in concentric annuli, as discussed in Section 3.6. To summarize the Gaussian cross-correlation for GPIES, the template is a 20×20 pixel stamp projected on a 281×281 pixel image.

3.4. Gaussian Matched Filter

The Gaussian matched filter (GMF) projects the PSF template on the coadded cubes directly, as illustrated in Figure 4. Therefore, the template is a stamp cube with both spatial and wavelength dimensions. It also uses a Gaussian PSF with a 2.4 pixel FWHM, and the reduction spectrum is used to scale the template as a function of wavelength. The matched filter is calculated according to Equation (2) considering a single combined cube. As a consequence, this implementation does not assume identically distributed noise. The local noise standard deviation is estimated at each position in a 20×20 pixel stamp from which a central disk with a 2.5 pixel radius is removed. To summarize the GMF for GPIES, the template is a $20 \times 20 \times 37$ pixel cube stamp, which includes the spectral dimension, projected on a $281 \times 281 \times 37$ pixel datacube.

3.5. Forward Model Matched Filter

The forward model matched filter (FMMF) is performed on an uncollapsed data set according to Equation (2), as illustrated in Figure 4. It uses the forward model as a template. The sector padding provides the necessary margin to perform the projection of the template anywhere in the image. The local standard deviation in each image is estimated at the position (ρ, θ) in a local arc defined by $[\rho - \Delta\rho, \rho + \Delta\rho] \times [\theta - \Delta\rho/\rho, \theta + \Delta\rho/\rho]$ and $\Delta\rho = 10$ px. In this case, the center of the arc is not masked, which results in an overestimation of the standard deviation in the presence of a planet signal. This does not significantly change the detectability of real objects, because the overestimated local standard deviation plays against both real objects and false positives.
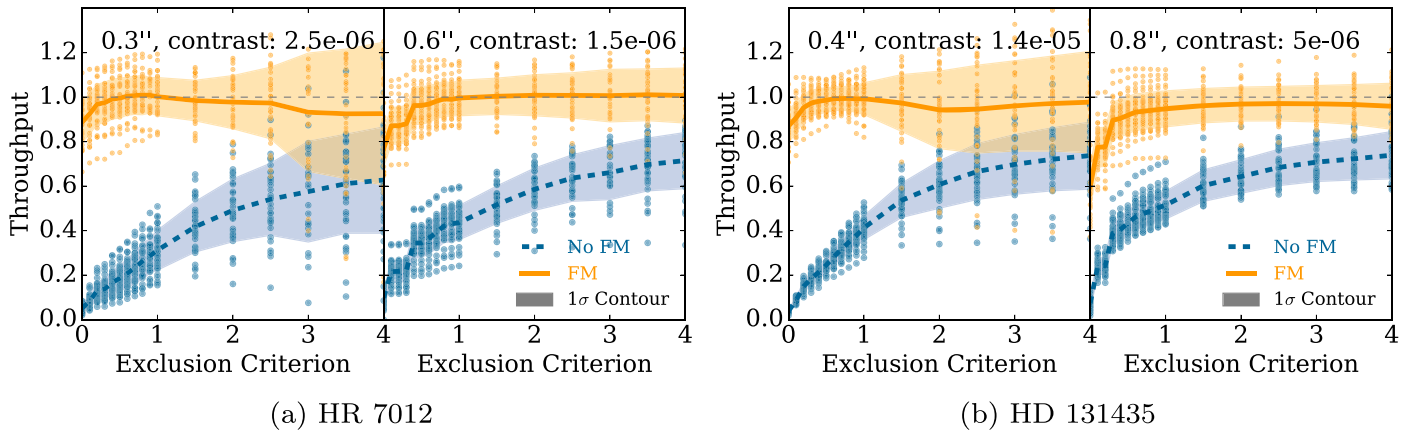
Figure 3. Algorithm throughput as a function of the exclusion criterion and the planet model. The throughput is estimated for both the forward model (FM, solid orange) and the original PSF (No FM, dashed blue). The conversion factor has no physical units, although it matches pixels when the reduction spectrum is constant. The test was performed with two representative data sets at two different separations: (a) HR 7012 at $0.''3$ and $0.''6$, and (b) HD 131435 at $0.''4$ and $0.''8$. The data sets were chosen to represent different regimes of data quality; HR 7012 is an example of a good data set, while HD 131435 is of medium quality. (Dots) Throughput for the 35 simulated exoplanets injected at different position angles in seven different copies of the data set. (Shaded) 1σ spread of the throughput. The original contrast of the simulated planets is indicated at the top of the plot. A spectrum with a methane signature is assumed for the planets and for the reference library selection.

In practice, only the values of the inner products $p_l^\top m_l$, $m_l^\top m_l$ and $\sigma_l(\rho, \theta)$ are saved while the sectors are speckle subtracted in order to limit computer memory usage. The final sum of Equation (2) is performed at the very end. One advantage of not combining the data is that it is not necessary to derotate the speckle-subtracted images, as long as one accounts for the movement of the model in the data as a function of time and wavelength. This removes the image interpolation associated with derotation and therefore limits interpolation errors. The matched-filter calculation is run on a discretized grid, in this case, centered on each pixel in the final image. The matched filter is therefore slightly less sensitive to planets that are not also centered on a pixel. Assuming spatially randomly distributed planets, we estimate that the discretization results in an average loss of S/N of a few percentage points, compared to the case where every planet is centered on a pixel. To summarize the FMMF for GPIES, the template is a $20 \times 20 \times 37 \times 38$ pixel multidimensional stamp, therefore it includes both spectral and time dimensions, projected on an uncombined $281 \times 281 \times 37 \times 38$ pixel data set.

The FMMF reduction of a typical GPI 38 exposure data set requires around 30 wall clock hours on a computer equipped with 32 2.3 GHz cores and using $\sim$20 GB of random-access memory (RAM). As a consequence, a supercomputer is necessary to process an entire survey. Each data set is independent and can be run on separate nodes without sharing memory. In this paper, we have used the SLAC bullet cluster to process the entire campaign using on the order of 10^6 CPU hours including the data processing necessary for the work presented in the remainder of the paper.

3.6. S/N Calculation

In practice, the theoretical S/N defined in Equation (2) needs to be empirically calibrated by estimating the standard deviation from the matched-filter map itself (Cantalloube et al. 2015). Indeed, the S/N is overestimated due to overly optimistic assumptions on the noise distribution. The residual noise is mostly white and Gaussian in areas that are not dominated by speckle noise, but both assumptions break down close to the mask because of speckle noise, causing the matched filter to lose some validity, although it remains

relatively effective nonetheless. It is also hard to estimate the local noise accurately, because the noise properties vary from pixel to pixel, adding a layer of uncertainty on the theoretical S/N. This is why an empirical standard deviation is estimated in the matched-filter map as a function of separation using a 4-pixel wide annulus, as illustrated in Figure 5. In order to prevent a planet from biasing its own S/N, the standard deviation is calculated at each pixel while masking a disk with a 5-pixel radius centered on that pixel from the annulus. In addition, all the known astrophysical objects are also masked. In the particular case of injected simulated planets, the standard deviation is estimated from the planet free reduction. In the following, unless specified otherwise, any reference to an S/N relates to the calibrated S/N.
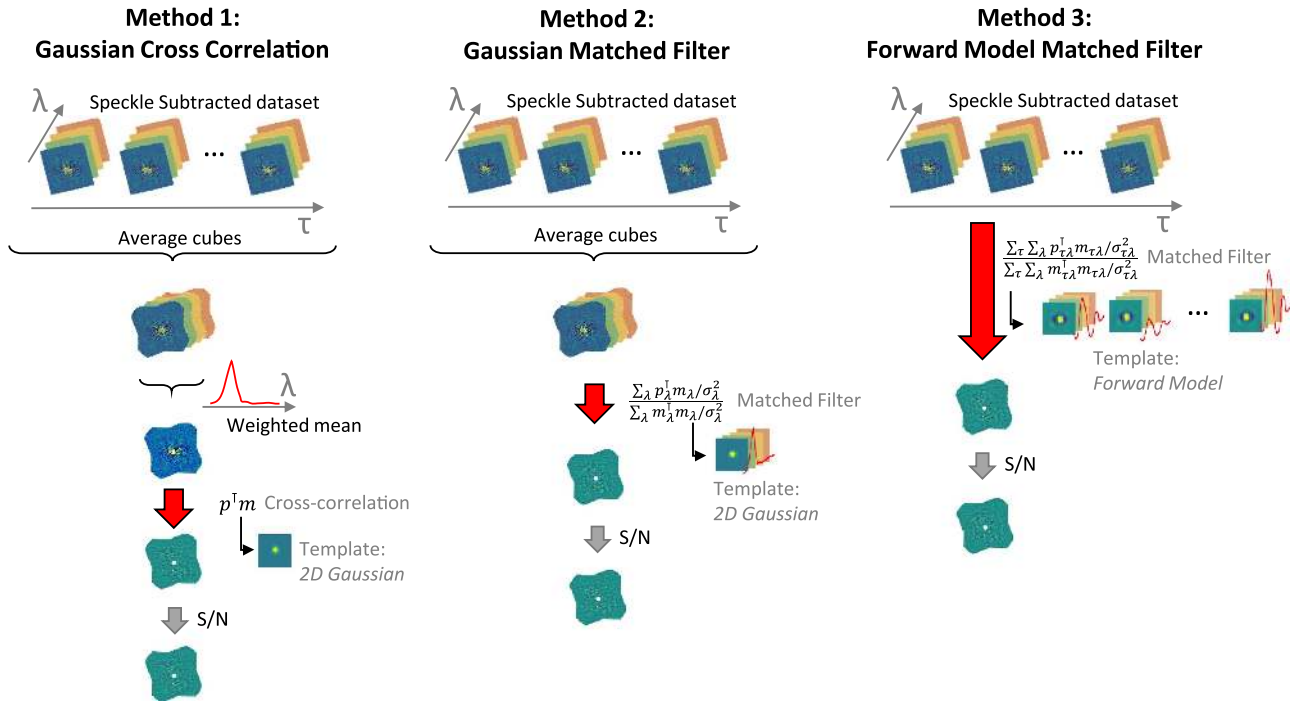
For a centered Gaussian noise, the S/N is related to the false-positive rate (FPR) following

$$\text{FPR} = \frac{1}{2}\left(1 - \text{erf}\left(\frac{\text{S/N}}{\sqrt{2}}\right)\right). \quad (3)$$

It is common practice to choose a 5σ detection threshold or higher, as discussed in Marois et al. (2008). For a Gaussian distribution, a 5σ threshold corresponds to an FPR equal to $\text{FPR} = 2.9 \times 10^{-7}$, which represents a false detection every 3.4×10^6 independent samples, or equivalently, on the order of 1000 GPI epochs. The deviation from Gaussianity of the residual noise in the final image will dramatically increase this false-positive fraction, as shown in Section 5.2. Small sample statistics (Mawet et al. 2014) also increases the false-positive fraction, but it is not expected to be a dominant term in this work and has therefore been neglected. Indeed, the inner working angle in the code is larger than three resolution elements at 1.6 μm.

3.7. Example Reduction

Figure 6 gives an example of a reduction using the three different algorithms on the HR 7012 GPIES data set observed on 2015 April 08. This data set does not contain any visible astrophysical signal. Simulated planets were injected according to Section 2.3. With the exception of the anomaly at $0.''4$, the FMMF consistently yields a higher S/N than the two other methods, and the final image shows fewer residual features. In

Method 1:
Gaussian Cross Correlation

Method 2:
Gaussian Matched Filter

Method 3:
Forward Model Matched Filter

Figure 4. Illustration of three different matched filters. They differ by their template, Gaussian, or forward model, and by the way the data set is combined before the matched filter.
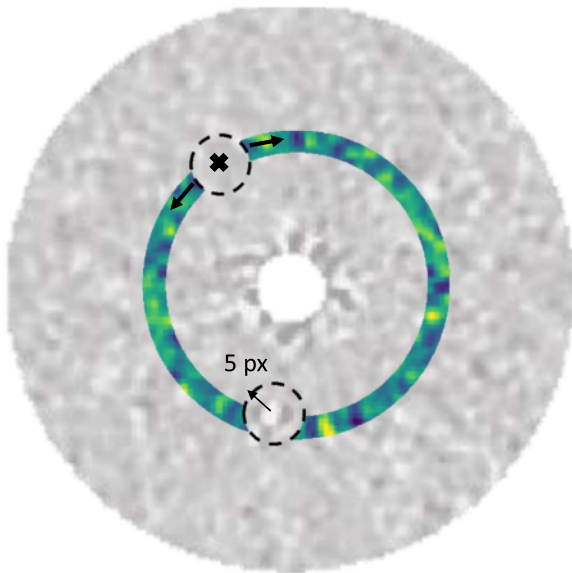
Figure 5. Illustration of the estimation of the standard deviation. For each pixel (cross), the standard deviation is empirically calculated in a 4-pixel wide annulus from which the surroundings of the current pixel as well as any known astrophysical signal have been masked.

the following, the simulated data sets are only reduced at the position of the simulated planets in order to speed up the algorithm. Figure 7 shows the processed data for several follow-up epochs of 51 Eridani also using the three algorithms. We show that FMMF would have marginally detected 51 Eridani b in all epochs, while this is not true for the GMF and the GCC.

4. Optimization

The free parameters for KLIP and the matched filter are the exclusion criterion, the reduction spectrum, the number of Karhunen-Loéve modes, the number of reference images, and the shape of the sectors. In this paper, we are only optimizing with respect to the exclusion criterion and the spectrum of the template. The exclusion criterion is only optimized for T-type objects, but we do not expect a significant difference for L-type objects. Preliminary tests show that the performance of the algorithm is less sensitive to the choice of the other parameters.

4.1. Exclusion Criterion

The exclusion criterion is defined in Appendix A.1.3 and its optimal value is a trade-off between achieving a better speckle subtraction and maintaining a stronger planet signal. The forward model helps to keep the throughput close to unity for more aggressive reductions, therefore improving the overall S/N.

The optimal value of the exclusion criterion is found by calculating the S/N of simulated planets for different values of the parameter. We used similar simulated planets as for Figure 3 with a 1000 K cloud-free model spectrum and a 600 K cloud-free spectrum for the reduction. The choice of spectra is discussed in Section 4.2. Figure 8 shows that the exclusion parameter has a soft optimum around 0.7. The optimal exclusion criterion does not seem to significantly depend on separation or data set quality. This apparent stability of the FMMF optimum suggests that a single value of the exclusion criterion can be used for the entire survey. Consequently, all following reductions in *H* band will be performed with an exclusion criterion of 0.7. The optimal exclusion criterion is expected to change depending on the filter used for the observation, and future work should involve separate optimizations in the different spectral bands. The FMMF almost always yields a better S/N than the other two algorithms, but this does not necessarily mean that it has a better detection efficiency. Interestingly, neither GCC nor GMF have a consistent or even a well-defined optimum. It is very common in the field to
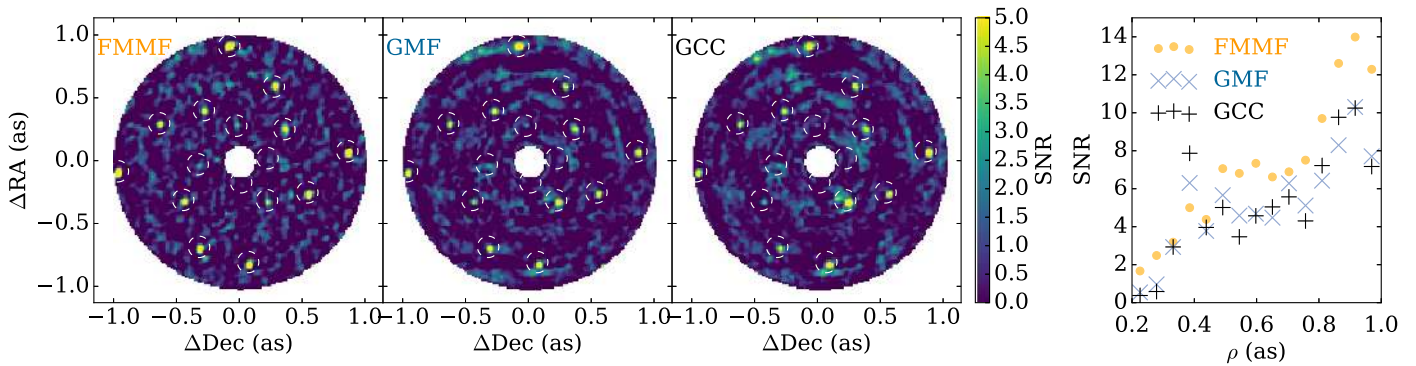
Figure 6. Reduction of the HR 7012 epoch including 16 simulated T-type planets with three different algorithms, from left to right: FMMF, GMF, and GCC. Each image corresponds to an S/N map where the simulated planets have been circled. The rightmost figure shows the S/N of the simulated planets as a function of separation for the three algorithms. A T-type spectrum similar to 51 Eridani b was used for the simulated planets, and a spectrum with a sharper methane signature was used for the reduction and the matched filter.
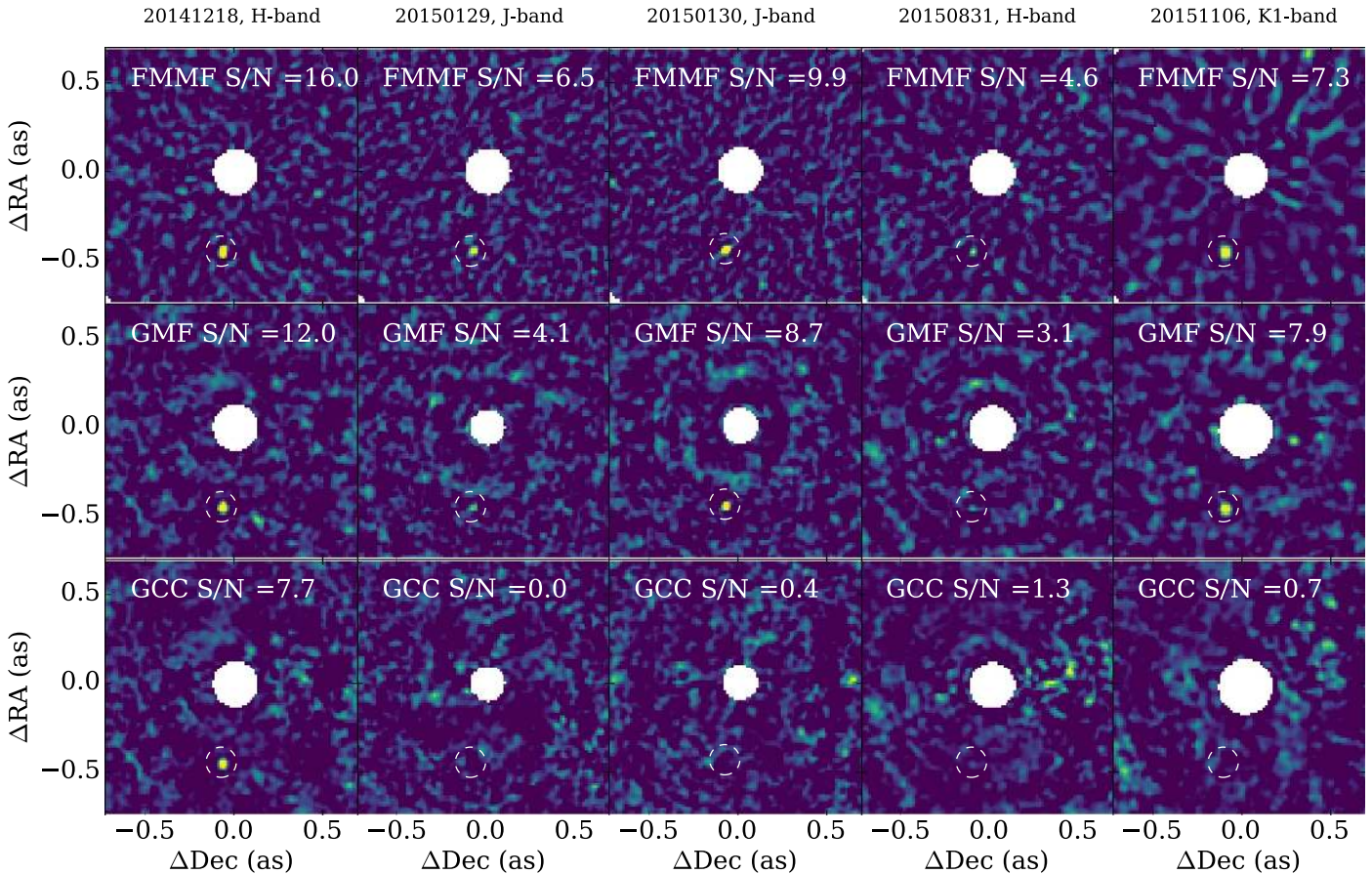
Figure 7. Reduction of several 51 Eridani epochs with three different algorithms, from top to bottom: FMMF, GMF, and GCC. Each column corresponds to a specific data set. The images correspond to S/N maps and the local S/N maximum at the location of the planet (dashed circle) is printed in each stamp.

reduce a data set with different sets of parameters and select the best one a posteriori, but we believe that the FMMF limits the need to fine-tune the parameters for each data set.

4.2. Spectral Mismatch

In this section we discuss the optimization of the reduction spectrum used in the forward model and the reference library selection. The goal is to estimate the number of reduction spectra that should be used to recover the widest variety of planets. Currently detectable exoplanets and brown dwarfs are

expected to feature spectra ranging from the T to the L spectral types. T-type objects are characterized by methane absorption bands visible in H band and an energy peak around 1.6 μm, while L-type objects feature a cloudy atmosphere with a flatter spectrum that peaks in the second half of the band.

We created ten copies of the HD 131435 GPIES data set, each with 16 simulated planets injected according to the spiral pattern from Figure 6. In each copy of the data set, the simulated planets were injected with one of ten spectra. The spectra were selected from a list of cloud-free and cloudy atmosphere models such that the most likely spectra from both

(a) HR 7012   (b) HD 131435
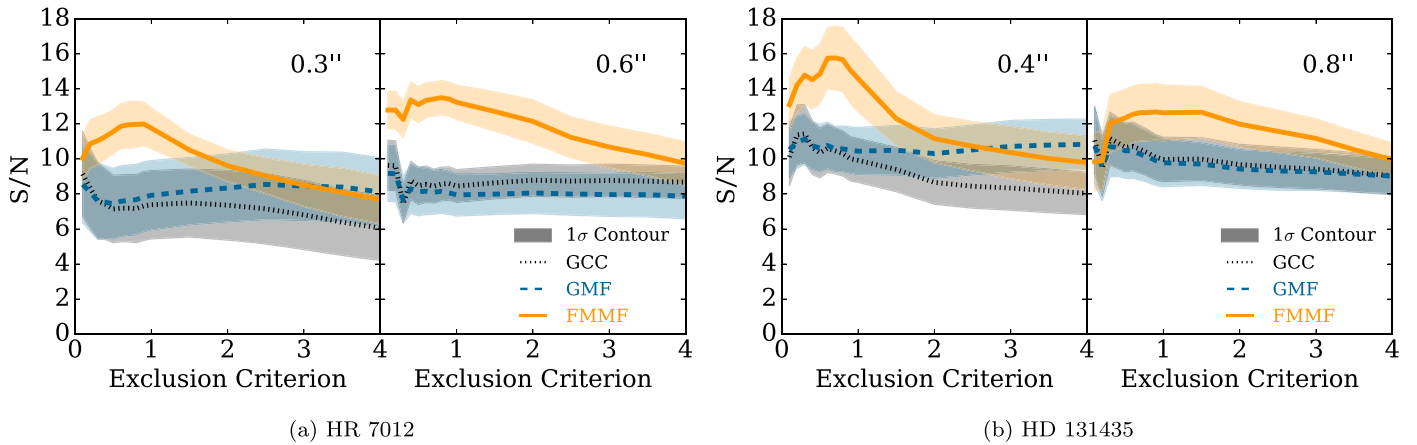
Figure 8. S/N of simulated planets as a function of the exclusion parameter for the three algorithms: FMMF (solid orange), GMF (dashed blue), and GCC (dotted black). The test was performed with two representative data sets at two different separations: (a) HR 7012 at 0″3 and 0″6, and (b) HD 131435 at 0″4 and 0″8. The data sets were chosen to represent different regimes of data quality; HR 7012 is an example of a good data set, while HD 131435 is of medium quality. The S/N was calculated for 35 simulated exoplanets injected at different position angles in seven different copies of the original data set. The shaded region is the 1σ spread of the S/N. A T-type spectrum similar to 51 Eridani b was used for the simulated planets, and a spectrum with a sharper methane signature was used for the reduction.


(a) Planet Spectrum Vs. Reduction Spectrum   (b) H-Band Model Spectra.
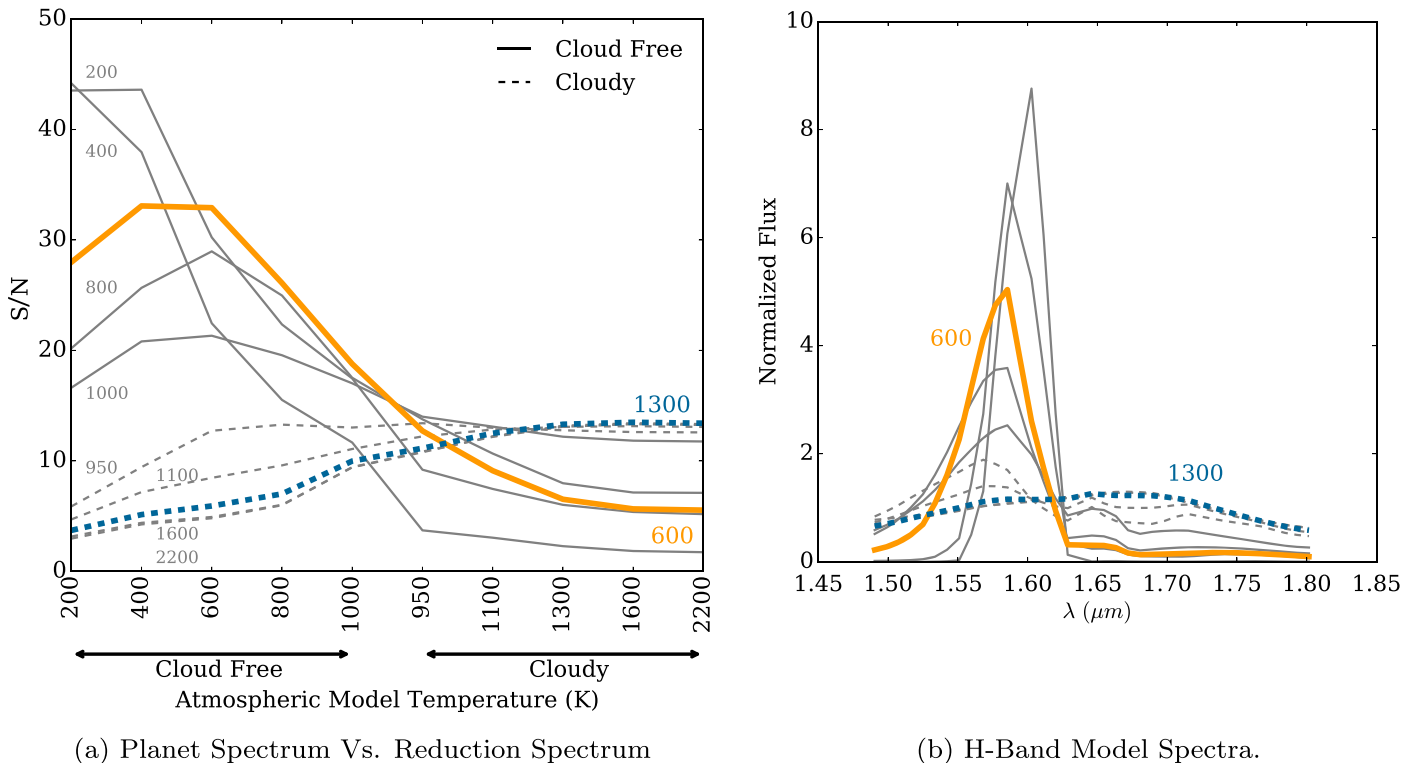
Figure 9. Effect of a mismatch between the planet spectrum and the reduction spectrum. (a) The median S/N (y-axis) of 16 planets injected in a the HD 131435 data set is shown as a function of both the planet spectrum (x-axis) and the reduction spectrum (curves). Figure (b) illustrates the model H-band spectra used in (a). The orange and blue spectra have been selected as they allow the recovery of most spectral types without a significant loss of S/N. The spectra are taken from atmospheric models described in M. S. Marley et al. (2017, in preparation) and Saumon et al. (2012).

T-type and L-type are represented. Note that the presence of clouds changes the temperature at which the methane features appear, but it does not significantly change the shape of the spectra. Then, each simulated data set was reduced with each of the ten reduction spectra, resulting in 100 different final products. For each of the planet spectra, the best reduction spectrum is defined by the one yielding the best median S/N for the 16 simulated planets. Figure 9 shows the median S/N of simulated planets as a function of their spectrum and the reduction spectrum. Surprisingly, the best reduction spectrum

is not the one corresponding to the simulated spectrum. One possible explanation is that the same reduction spectrum is used for the forward model as for the reference library selection through the exclusion criterion. The spectrum is also a way to weight the spectral channels differently, which could effectively correct for a biased estimation of the standard deviation where there is planet signal. However, a deeper exploration of this effect is beyond the scope of this paper. We also conclude that only two spectra, the cloud-free 600 K and the cloudy 1300 K, are necessary to allow the detection of most giant
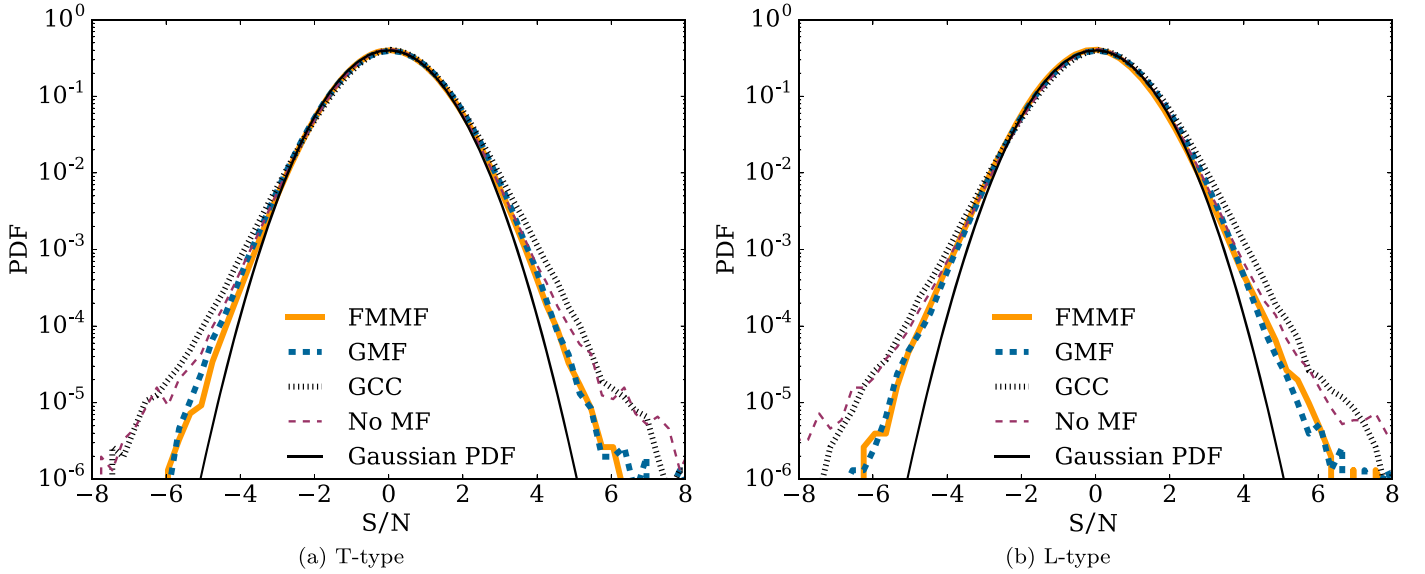
Figure 10. PDF of the GPIES S/N maps for the three algorithms: FMMF (solid orange), GMF (dashed blue), and GCC (dotted black). The matched-filter residuals can be compared to the residuals without any matched filter (dashed purple) and to an ideal Gaussian PDF (solid black). The PDFs are given for both the T-type (a) and the L-type (b) reductions. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.
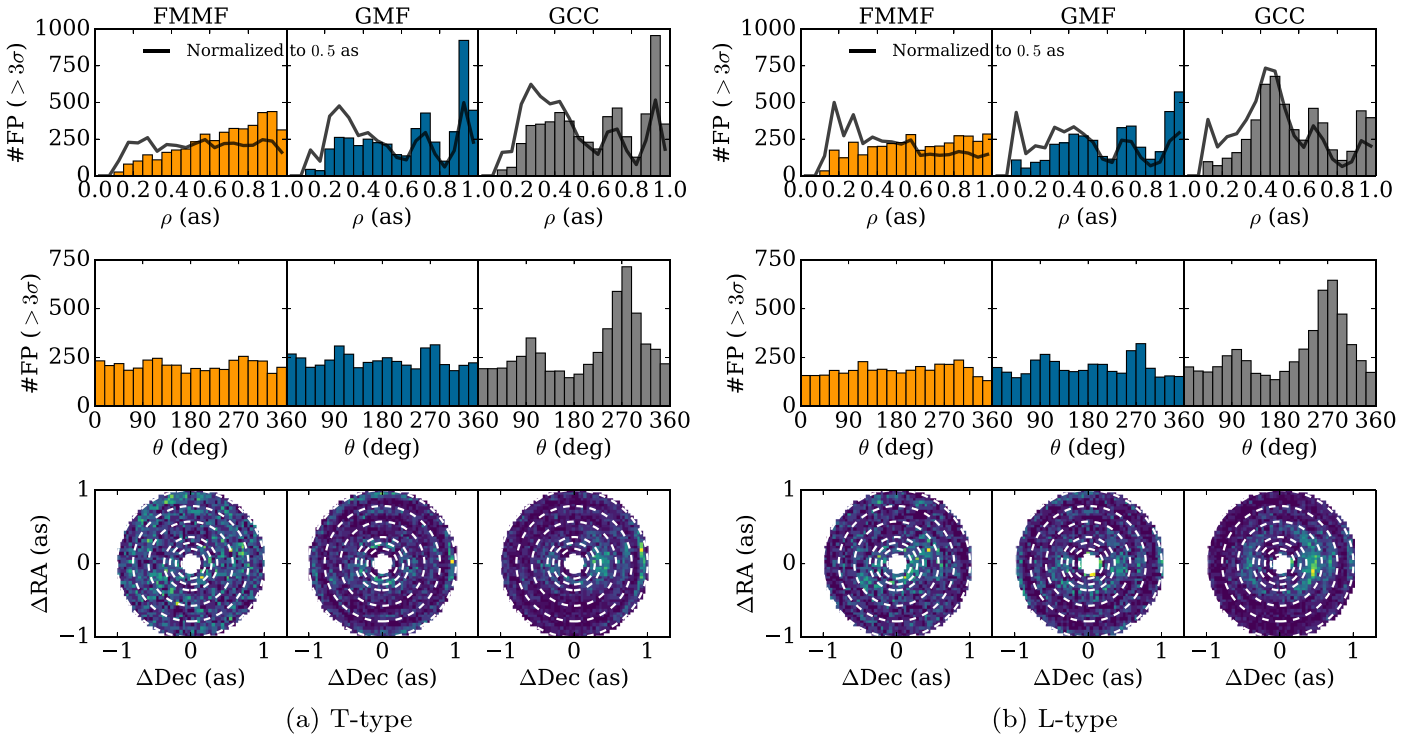
Figure 11. Spatial density of false positives (FP) brighter than 3σ in GPIES for three algorithms: FMMF (first column, orange), GMF (second column, blue), and GCC (third column, gray). The histograms are given for both the T-type (a) and the L-type (b) reductions. The first and second row feature the number of false positives per bin as a function of separation and position angle, respectively. The density of false positives is expected to increase at larger separation because a larger area is available. The solid black line gives the equivalent number of false positives at $0\rlap{.}''5$ after normalizing by the area at each separation. The bottom row shows the two-dimensional density of false positives as a function of declination and right ascension, or equivalently, x, y-axes. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.

planets without a significant loss of S/N. This result is consistent with a similar study in Johnson-Groh (2017) using TLOCI (Marois et al. 2014). Although the 600 K spectrum is the optimal reduction spectrum for T-type objects, the methane-induced peak in H band is unrealistically sharp. It has indeed not yet been observed in a real directly imaged exoplanet. However,

the reduction spectrum and the spectrum of the simulated planets can be different. In order to be more representative of the observations, we use a 1000 K cloud-free model spectrum similar to 51 Eridani b for the T-type injected planet, which has a softer methane-induced peak. The L-type simulated planets use the same 1300 K cloudy model spectrum as for the reduction.
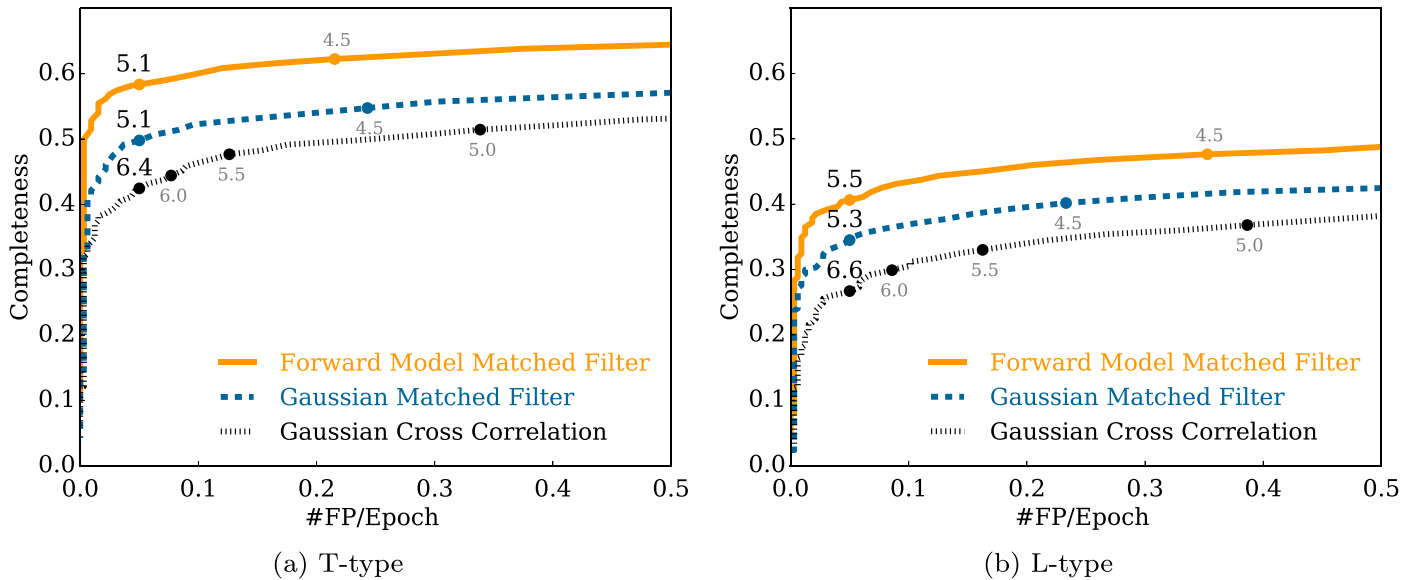
Figure 12. ROC for three different algorithms: FMMF(orange solid line), GMF (blue dashed line), and GCC (gray dotted line). The ROC curves feature the current GPIES completeness to T-type (a) and L-type (b) planets with a 4×10^{-6} contrast integrated over all separations and angles as a function of the number of False Positives (FP) per epoch. A few values of the S/N threshold are annotated in gray on each curve. The threshold corresponding to a fraction of false positives per epoch of 0.05 is written in a larger black font. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.

For the remainder of the paper, the T-type reduction refers to a cloud-free 600 K reduction spectrum, while the L-type reduction refers to a cloudy 1300 K reduction model.

5. Noise Characterization

5.1. Noise Distribution

By combining the 330 GPIES H-band observations after removing any known astrophysical point sources, we have been able to estimate the PDF of the residual noise up to an unprecedented precision in the fully reduced S/N maps. Figure 10 compares the ideal Gaussian PDF with the PDF of the different algorithms calculated from the normalized histograms of the pixel values of all the S/N maps. The statistics of the noise strongly deviates from the Gaussian distribution at occurrence rates much greater than the frequency of planets (Nielsen et al. 2013; Bowler et al. 2015; Galicher et al. 2016; Vigan et al. 2017), demonstrating the high occurrence of false positives with high S/N in direct imaging. The Gaussian cross-correlation has the same PDF as the S/N maps calculated from the speckle-subtracted images with no cross-correlation or matched filter. The GMF and the FMMF both significantly improve the statistics of the residuals but remain quite remote from an ideal Gaussian distribution. The excess of high S/N occurrences can be explained by either a truly non-Gaussian statistics or by a poor estimation of the standard deviation of the noise when calculating the S/N. For example, an underestimated standard deviation for a pixel will result in more high S/N false positives. The relative improvement between the cross-correlation and the matched filter suggests that the latter explanation may still be dominant, which we discuss in more detail in the next paragraph.

Figure 11 shows the spatial densities of 3σ false positives as a function of separation and position angle. Owing to the correlation in the final S/N maps, one should count the number of speckles and not the raw number of pixels above a given threshold in order to count the number of false positives. Indeed, for any high S/N false detection, the size of the bump,

and therefore the number of pixels above the threshold, will depend on the correlation length of the noise, which also depends on the algorithm used. The detection of false positives is therefore done recursively as follows. The highest S/N pixel is flagged, and a 4-pixel area radius is masked around it. This process is repeated until the only false positives left are below a predefined S/N threshold. Both the GCC and the GMF exhibit strong radial variations of their false-positive densities at the position of the sector boundaries, as seen in the top row plots of Figure 11. In addition, the GCC has a significant excess of false positives around 90° and 270° in position angle. This feature can be explained by the excess of speckle noise on both sides of the focal plane mask in the direction of the wind, which we refer to as the wind-butterfly. The pattern is still visible after combining the entire survey because the wind direction overwhelmingly favors the northeast on Cerro Pachon at the Gemini South telescope. The wind-butterfly breaks the assumption of azimuthally identically distributed noise, which we use when estimating the standard deviation in concentric annuli. The wind-butterfly explains why the probability density function of the GCC in Figure 10 is significantly higher than the other matched filters. The GMF does not suffer from this effect because the matched filter includes a normalization with respect to the local standard deviation estimated around each pixel. The FMMF features a similar PDF, meaning that a similar S/N detection threshold should yield the same number of false positives. The real performance of each algorithm is studied in the next sections. While a cross-correlation is a common planet-detection approach, our analysis suggests that it can be ill-suited if the noise varies azimuthally. One should instead use the expression for the matched filter from Equation (2). An alternative approach would be to vary the S/N threshold for each data set and as a function of position, but the lack of local independent samples to estimate a position dependent PDF at low false-positive rates makes this endeavor very challenging. Indeed, one needs to have probed the PDF of the noise at high S/N in order to evaluate the false-positive probability of any detection. For example, 5σ events are

(a) Conversion Factor

(b) Relative Variations
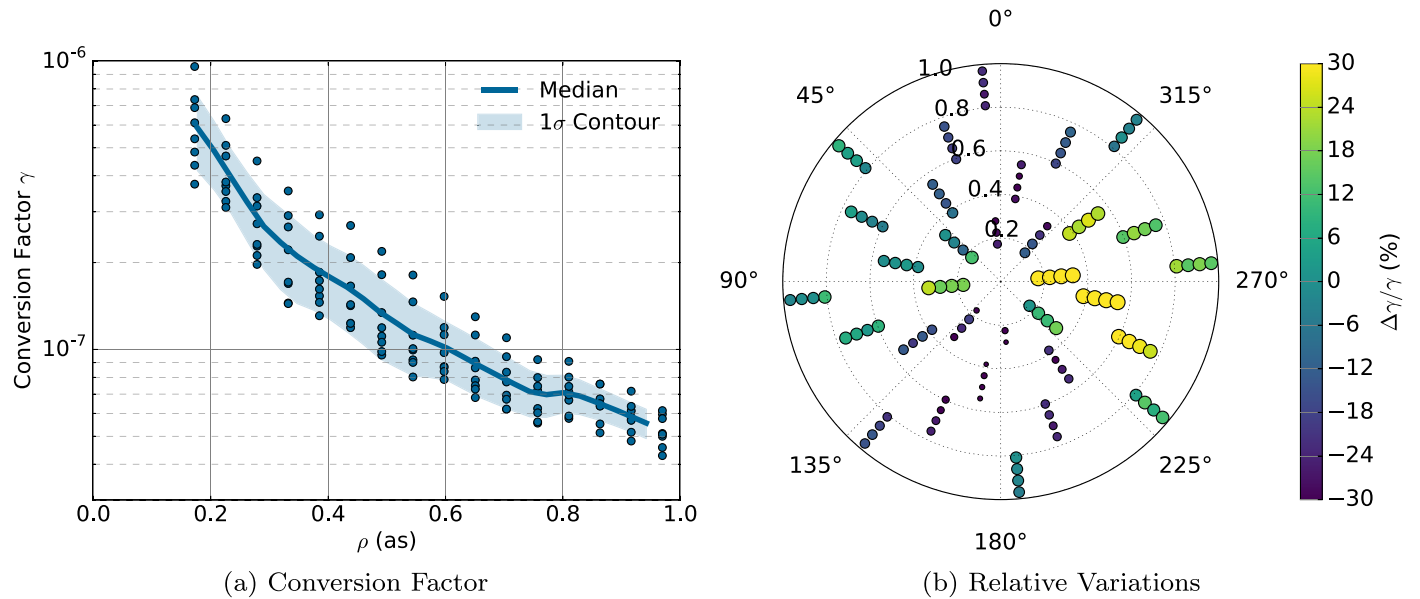
Figure 13. Illustration of a T-type conversion factor calibration with simulated injected planets. (a) The conversion factor is calculated from the median over eight simulated planets at each separation. A total of 128 simulated exoplanets injected in eight different copies of the data set is used to calibrate the conversion factor. (b) The polar plot shows the relative azimuthal variation of the conversion factor (represented by both the color map and the dot scale) as a function of separation in arcseconds and position angle in degrees. The data set has been chosen to feature a strong wind-butterfly effect, which results in a higher conversion factor of around 90° and 270°.

sufficiently rare that their occurrence rate can only be estimated from the data of a entire survey and not from individual images.

The FMMF L-type reduction has significantly more false positives near the mask than a T-type reduction. This is likely because there is a higher density of speckles near the mask and the L-type spectrum is a better match to them than a sharper spectrum. We have also seen in Figure 10 that for the three algorithms, the L-type PDF has wider tails than the T-type PDF, suggesting that the number of L-type false positives will be higher at constant S/N.

### 5.2. Receiver Operating Characteristic

In general, an improvement in the S/N does not guarantee a better detection efficiency because the false-positive rate could increase in the mean time. It is therefore important to compare the number of detected planets to the number of false positives. For example, Figure 8 showed that the Gaussian cross-correlation tends to have slightly higher S/N than the GMF but we will show in this section that the cross-correlation leads to more false positives. For this reason, receiver operating characteristic (ROC) have become increasingly popular in direct imaging to compare different algorithms (Caucci et al. 2007; Choquet et al. 2015; Pairet et al. 2016; Pueyo 2016; R. Jensen-Clem et al. 2017, in preparation). A ROC curve compares the false-positive fraction to the true positive fraction, i.e., completeness, as a function of the detection threshold. Alternatively, we have decided to replace the false-positive fraction by the number of false detections per epoch integrated over the entire image since it is easily translatable to survey efficiency and astrophysical background occurrence rate. We have also chosen to fix the planet contrast and assume a uniform planet distribution in separation and position angle. Although this approach does not address the dependence of the ROC curve on separation and planet contrast, it is sufficient to evaluate the relative performance of each algorithm. Multi-dimensional ROC curves could be used, but the contrast curves

calculated in Section 6 already include most of the relevant information. The ROC curves for the three algorithms are shown in Figure 12. Each ROC curve has been built by injecting 16 simulated exoplanets, using either a 1000 K cloud-free T-type spectrum, reduced with a 600 K cloud-free model spectrum, or a 1300 K cloudy L-type spectrum, reduced with a the same cloudy spectrum. All planets were injected at a 4×10^{-6} contrast in 330 GPIES H-band data sets using the spiral pattern illustrated in Figure 6. The advantage of using a large number of data sets is to marginalize over the conditions in any one particular epoch. Figure 12 shows that FMMF yields a better completeness at any false-positive rate. In addition, the S/N threshold corresponding to a given false-positive rate is always higher in the case of the cross-correlation than for the matched filters because of the larger tail in the PDF. For example, it has a T-type false-positive rate that is roughly six times higher at 5σ than the two other algorithms.

The detection threshold should be defined by the number of false positives that can reasonably be followed up during the survey. We set this false-positive rate at 0.05 per epoch corresponding to 30 false positives for the entire survey. The S/N threshold therefore depends on the algorithm, as shown in Figure 12.

6. Contrast Curve

In this paper, contrast is defined as the broadband flux ratio between the companion and star. The contrast curve is defined as the 50% detection completeness contour assuming a false-positive rate sufficiently low to limit the number of false positives. The false-positive rate can be expressed in terms of an S/N threshold, which is not necessarily 5σ. Using a hard 5σ threshold does not allow for a meaningful comparison of contrast curves, because different algorithms can lead to different numbers of false positives for the same S/N (as demonstrated in Section 5.2). Contrast curve calculations require a calibration step to translate pixel values into planet
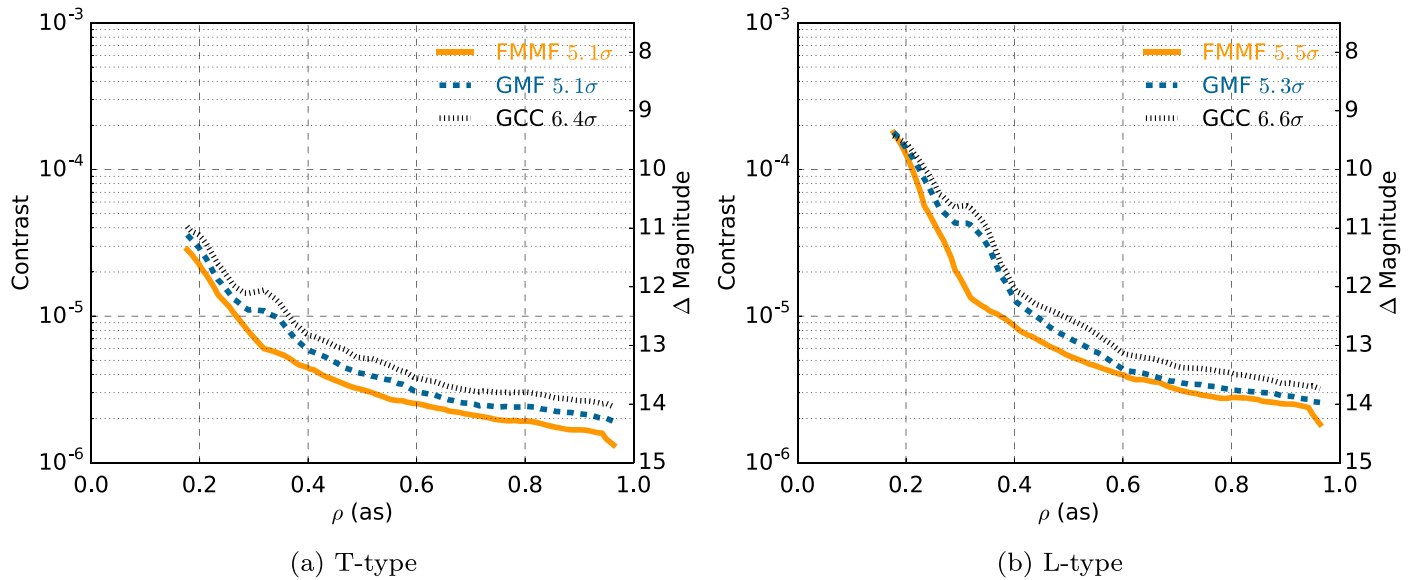
Figure 14. Median GPIES contrast curves for T-type (a) and L-type (b) reductions. Three different algorithms are compared: FMMF (orange solid line), GMF (blue dashed line), and GCC (gray dotted line). The contrast on the y-axis refers to the companion-to host-star brightness ratio with a 50% completeness and a false-positive rate of 0.05 per epoch. The detection threshold, which is indicated in the legend, varies from one algorithm to the other in order to always yield the same number of false positives. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.
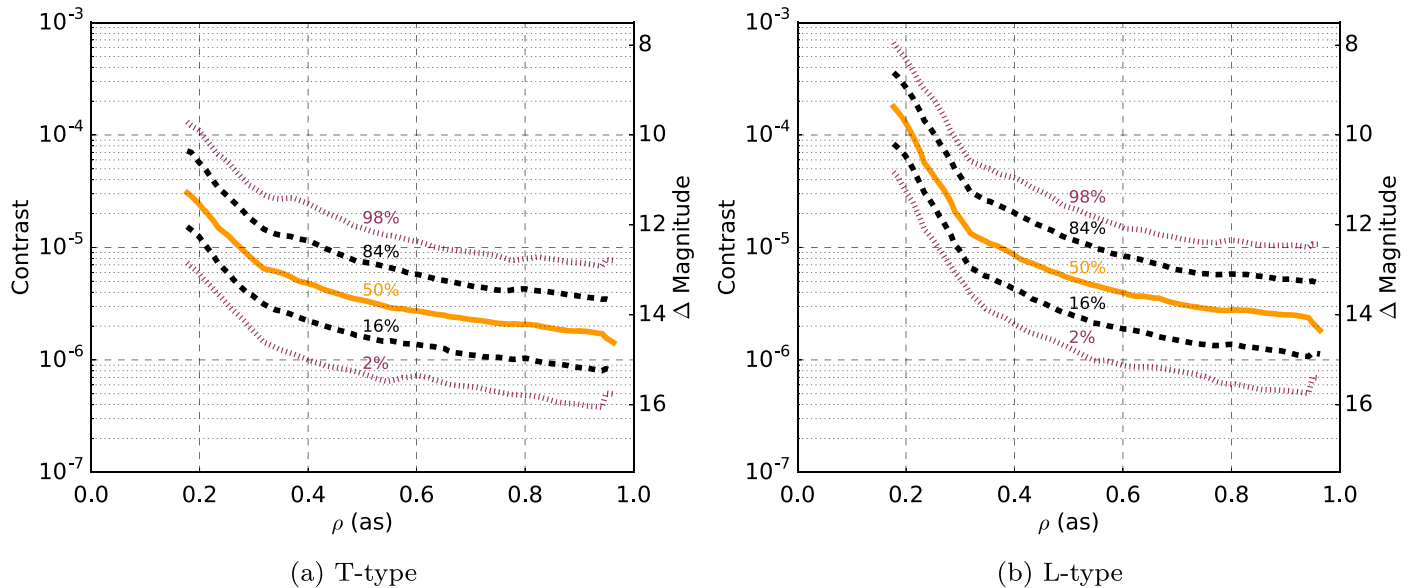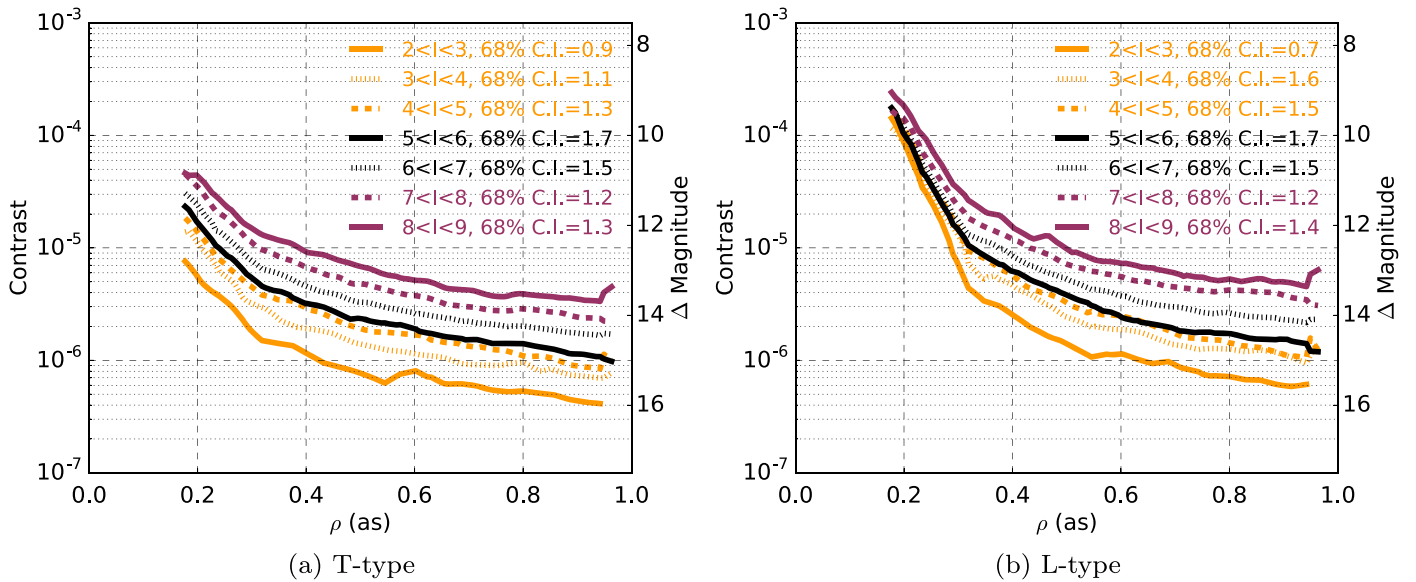

Figure 15. Percentiles of GPIES contrast curves with the FMMF for T-type (a) and L-type (b) reductions. The contrast on the y-axis refers to the companion-to-host-star brightness ratio with a 50% completeness and a false-positive rate of 0.05 per epoch. The detection threshold is set to 5.1σ and 5.5σ for the T-type and L-type reductions, respectively, in order to always yield the same number of false positives. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.

contrast, which is in some cases referred to as throughput correction. Indeed, speckle subtraction algorithms like KLIP are known to oversubtract the signal of the planet and make it appear fainter than it really is. This effect can be calibrated out by inflating the standard deviations by a certain factor, known as throughput, calculated from simulated planet injection. However, the throughput correction only makes sense if the pixel values, and therefore the standard deviation, are in units of contrast, which is not the case in this paper. The matched-filter maps do not directly estimate the contrast of a planet, but rather try to estimate a theoretical S/N as defined in Equation (2). Therefore, the well-known throughput correction is here replaced by a conversion factor to relate the

matched-filter values to actual contrast estimates, which is also calibrated using simulated planets.

As a consequence, we define a contrast curve as $\eta\gamma(\rho)\sigma_S(\rho)$, where η is the S/N threshold, $\gamma(\rho)$ is a median conversion factor between the matched-filter map and the true contrast of the planet, and $\sigma_S(\rho)$ is the standard deviation of the noise in the matched-filter map.

The different thresholds η were determined in Section 5.2 for the three algorithms to yield a false-positive frequency of 0.05 per epoch. The standard deviation $\sigma_S(\rho)$ is calculated in concentric annuli according to Section 3.6. The conversion factor γ is empirically determined by injecting simulated planets with known contrast in each data set. The contrast of

Figure 16. Median GPIES contrast curves as a function of the star I magnitude with the FMMF for T-type (a) and L-type (b) reductions. The legend includes the I magnitude bin used for each curve as well as the size of the 68% confidence interval (CI) at $0\rlap{.}{''}4$ in magnitude units. The contrast on the y-axis refers to the companion-to-host-star brightness ratio with a 50% completeness and a false-positive rate of 0.05 per epoch. The detection threshold is set to 5.1σ and 5.5σ for the T-type and L-type reductions, respectively, in order to always yield the same number of false positives. A total of 330 GPIES H-band data sets were used in which any known astrophysical signal was removed.

the simulated planets is chosen to result in a signal somewhere between 5–15$\sigma(\rho)$ in the final matched-filter map. In total, 128 planets are injected at 16 different separations and position angles in eight different copies of the data set. The original matched-filter map is subtracted from the simulated planet reductions in order to remove the effect of the residual noise. The conversion factor is linearly interpolated from the median of the eight simulated planets at each separation, as shown in Figure 13. Note that the contrast curve is a function of the spectrum of the simulated planets as well as of the reduction spectrum. Ideally, there should be as many contrast curves as there are possible spectra. We have limited our study to objects with either a 1000 K cloud-free T-type spectrum, reduced with a 600 K cloud-free model spectrum, or a 1300 K cloudy L-type spectrum, reduced with the same cloudy spectrum. A caveat is that the contrast curves are therefore only truly valid for planets with the same spectrum as the injected planets.

The GPIES median contrast curve for each algorithm as well as the associated detection threshold for the two spectral type reductions is given in Figure 14. FMMF yields the best median contrast curve at all separations. In the T-type reduction and compared to the GMF, the FMMF contrast enhancement ranges from a median 25% up to more than a factor 2 in some cases. The L-type median contrast enhancement drops from 25% below $0\rlap{.}{''}5$ to 10% at larger separation. The median contrast improvement relative to the cross-correlation is around 50%, which corresponds to a factor 2.3 gain in exposure time assuming a square root increase of the S/N with time. Tests have shown that this assumption generally holds true, with the exception of a slight dependence on observing conditions. The local maximum in the contrast curve for the GCC and the GMF is likely due to the non-optimal definition of the reduction sectors for a regular KLIP reduction. We therefore ignore the FMMF contrast increase in the $0\rlap{.}{''}3$–$0\rlap{.}{''}4$ range as we do not consider it representative of the overall performance.

The sensitivity of any observation is highly dependent on the observing conditions and the brightness of the star. Figure 15

shows the percentile contours of the contrast curves for GPIES using FMMF. Figure 16 shows the FMMF median contrast curves as a function of the I magnitude of the star. There is an order of magnitude ratio between the sensitivity around the faintest stars ($I_{\mathrm{mag}} \approx 9$) of the survey compared to the brightest stars ($I_{\mathrm{mag}} \approx 2$).

7. Candidate Vetting

In order for the contrast curves and the planet completeness to be accurate, any signal above the detection threshold must be properly vetted. This means that any high S/N signal should be confirmed as a true or a false positive. The best approach is always to follow up all the candidates, which can be a significant telescope time investment. In the case of a real astrophysical signal, the second epoch is generally used to determine proper motion and parallax in order to exclude or confirm the possibility of a background object. In the case of low S/N candidates, if it cannot be detected in the second epoch, it is usually classified as a false positive. However, it is then necessary to improve the contrast curve in the second epoch in order to exclude the real signal hypothesis with a high enough significance. We discuss the necessary contrast improvements in this section.

First, the detection threshold can be lowered in the follow-up observation compared to the first epoch, because the constraint on the noise is higher. A spurious signal would need to be in the same position in order for it to be mistakenly considered as a true detection. With a 4σ detection threshold in the second epoch, there is less than one chance in one thousand to incorrectly classify the first detection as a real signal, which seems reasonable. Indeed, there is on the order of 10^3 independent samples in a $1''$ field of view in a GPI image assuming a conservative characteristic correlation length of the residuals of 3 pixels, and a 4σ threshold yields less than one false positive per epoch for both matched filters.

Second, we can estimate the probability of detecting the signal in the new epoch as a function of the original S/N and
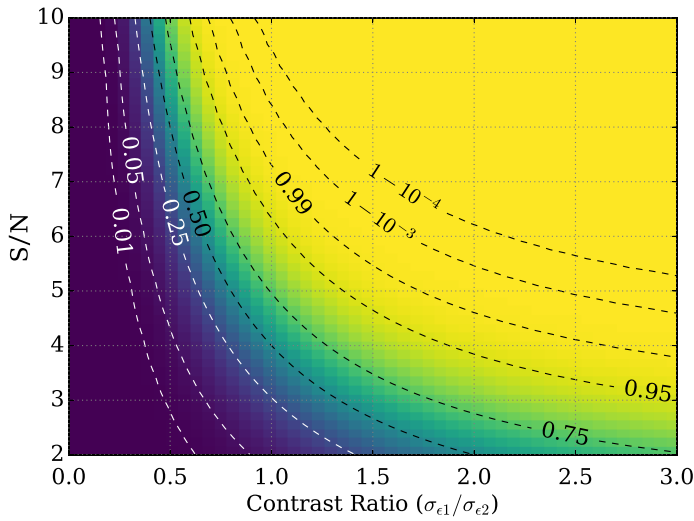
(a) 2nd Epoch 4σ Detection Probability
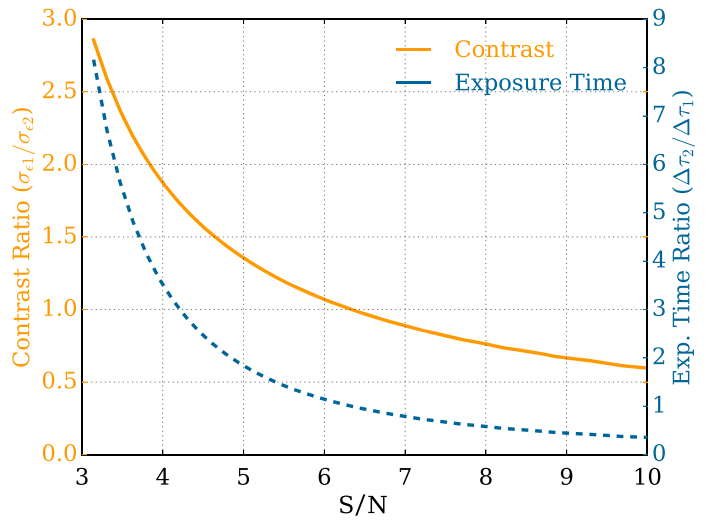
(b) 95% Probability

Figure 17. Detection probability of a point-source candidate in a second epoch. Figure (a) shows the probability of detecting a point source in the follow-up observation (color map and contours) as a function of the contrast ratio $\sigma_{\epsilon1}/\sigma_{\epsilon2}$, where the integer subscript refers to the first or the second epoch and the first-epoch S/N. Figure (b) represents the 95% detection probability contour with the corresponding exposure time ratio $\Delta\tau_2/\Delta\tau_1$. The exposure time calculation assumes that the observing conditions are identical for both epochs.
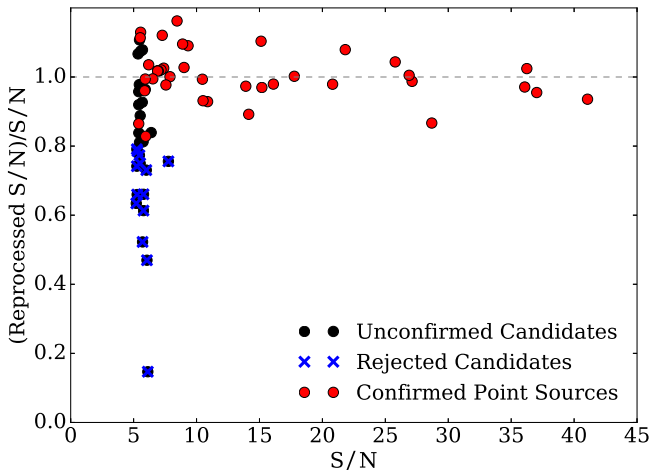
Figure 18. Exclusion of detected candidates based on a second reduction of the data with an annulus centered at their separation. The dots represent the L-type candidates above 5.2σ as a function of their original S/N and the ratio of the S/N between the first and the second reduction. We reject all candidates for which the S/N drops by more than 20% in the new reduction.

the ratio of contrast standard deviation $\sigma_{\epsilon1}/\sigma_{\epsilon2}$. Assuming a known contrast for a point source, the probability of detection is given by the tail distribution of the planet signal at the detection threshold in the second epoch. However, it needs to be marginalized over the planet contrast because of its uncertainty in the first epoch. The detection probability in the second epoch assuming a detection threshold η_2 can be written as

$$P_{\rm detec}\left(S/N, \sigma_{\epsilon1}/\sigma_{\epsilon2}\right) = \int_s$$
$$\times (1 - {\rm CDF}(\eta_2 - s\,\sigma_{\epsilon1}/\sigma_{\epsilon2})){\rm PDF}(s - S/N)ds, \quad (4)$$

with PDF and CDF the probability density function and the cumulative distribution function of a Gaussian distribution with zero mean and unit standard deviation. Figure 17 shows that

the detection probability contours as a function of the original S/N and the noise ratio.

The problem is that candidate follow-ups are expensive in telescope time and a significant fraction of the detections are expected to be false positives. Indeed, a significant number of candidates appear to be reduction artifacts. Some of these artifacts are very sensitive to the edge of the sectors and can therefore be spotted by running a new reduction with an annulus centered at their separation. Real objects are less sensitive to the definition of the sectors, so if the S/N of the candidate drops significantly, it is likely a false positive and should not be reobserved. Figure 18 shows the L-type candidates above 5.2σ as a function of their original S/N and the relative S/N with the new reduction. Because the S/N ratios for the real point sources are always close to unity, we conclude that we can reject all candidates for which this ratio drops below 0.8 without significantly biasing the algorithm completeness. This plot cannot be shown for the T-type reduction because there is only one confirmed T-type object in GPIES, which is 51 Eridani b. Most point sources are indeed background stars, and the few substellar objects feature an L-type spectrum. The lack of control sample for the T-type reduction does not allow us to define a boundary in this case. We acknowledge that rejecting candidates based on an ad hoc criterion is not the best solution and proves that work remains to be done. Ideally, we would wish to combine the results of both reductions into a single detection metric rather than applying different cutoffs sequentially.

8. From Contrast Curve to Completeness

The main goal of a direct-imaging exoplanet survey, after discovering new worlds to characterize, is to place constraints on the population of wide-orbit substellar objects (Nielsen et al. 2013; Bowler et al. 2015; Galicher et al. 2016; Vigan et al. 2017). Deriving the frequency of exoplanets first relies on the completeness of the survey. The most common method to calculate the completeness of an observation is to reduce the same data set with simulated planets at all separations and
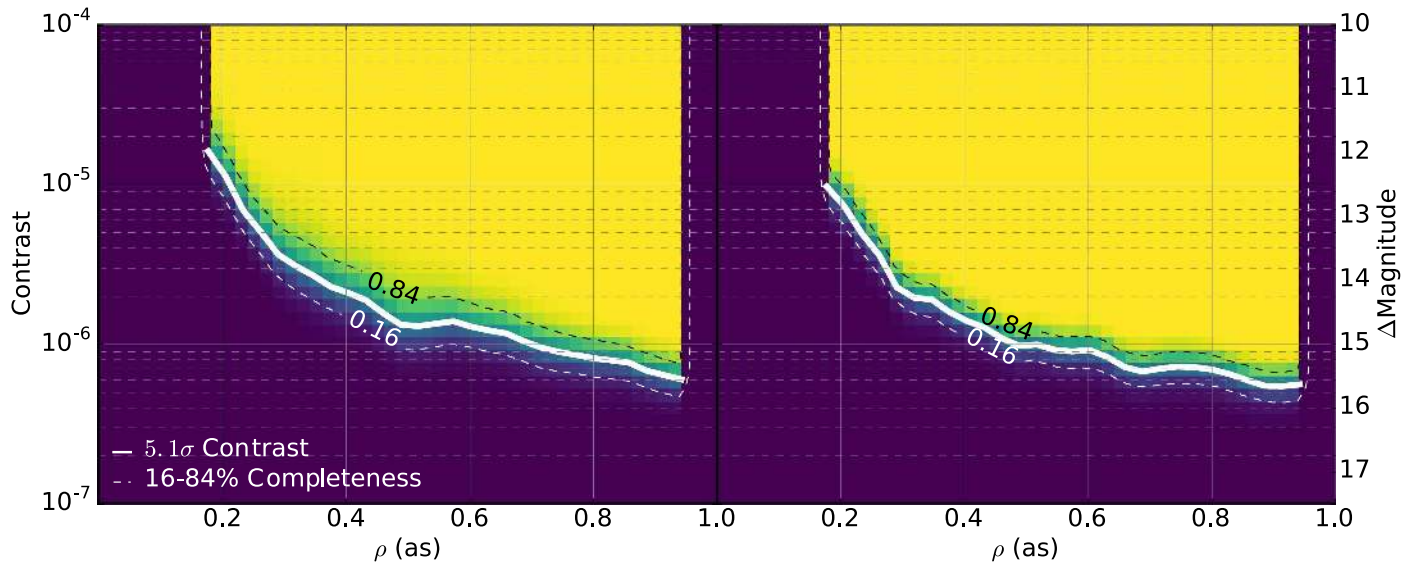
Figure 19. Exoplanet completeness as a function of contrast and separation ρ. The left panel shows an example of strong wind-butterfly resulting in significant variation of the conversion factor as a function of the position angle. The right panel shows an example in which the azimuthal scatter of the conversion factor is negligible. Contrast curve at 5.1σ corresponding to the 50% detection completeness contour for a FMMF T-type reduction (solid line). 16% and 84% completeness contours (dashed line).

contrasts. The completeness at a given point is then the fraction of the fiducial planets detected. This approach requires many reductions for each data set, which is tractable with classical ADI reductions but computationally out of reach for the FMMF. Lafrenière et al. (2007b) showed that when Gaussian noise is assumed, the completeness can be estimated directly from the contrast curve. For example, a planet lying exactly on the contrast curve will have a 50% chance to fall above or below it as a result of the noise. The probability of detecting a planet of a given contrast is given by the tail distribution of a Gaussian distribution centered on the contrast of the planet evaluated at the contrast curve. We have shown that the Gaussian assumption is not verified in the tail of the noise distribution, but this effect would only become important for a completeness of planets far from the contrast curve. In the ideal case, the $(\eta + 1)\gamma(\rho)\sigma_\mathcal{S}(\rho)$ and the $(\eta - 1)\gamma(\rho)\sigma_\mathcal{S}(\rho)$ curves represent the 84% and 16% completeness contour, respectively. One caveat is that the azimuthal variations of the conversion factor (e.g., due to the wind-butterfly) add a scatter term to the measured contrast of the planet that effectively widens the completeness contours. For example, a planet inside the wind-butterfly suffers from a higher conversion factor than a planet outside of it, resulting in a lower detection probability.

The azimuthal variations of the conversion factor due to the wind-butterfly or other artifacts can be interpreted as a noise term with a standard deviation σ_γ. When we assume that the residual noise and the conversion factor variations are independent, the standard deviation of the contrast can be written

$$\sigma_\epsilon(\epsilon) = \sqrt{\gamma^2\sigma_\mathcal{S}^2 + \frac{\epsilon^2}{\gamma^2}\sigma_\gamma^2}, \qquad (5)$$

which is used to estimate the completeness. Figure 19 shows two examples of completeness contours, one with a strong wind-butterfly and another where it is negligible.

9. Conclusion

In order to improve the GPIES planet sensitivity, we implemented a new matched-filter based algorithm using KLIP and a forward-modeled PSF template. This algorithm includes the speckle subtraction efficiency of KLIP, while mitigating the PSF distortion penalty. We also presented a GCC and a GMF for comparison. The cross-correlation requires an identically distributed noise in order to be valid. This assumption is not verified in GPI images and it therefore leads to an increase of false positives in the part of the image, like the wind-butterfly, where the noise is consistently stronger. An accurate matched filter needs to include a normalization by the local noise standard deviation, which improves the statistics of the residual noise in the final S/N maps significantly.

After optimizing the aggressiveness of the algorithm, we showed that two reduction spectra, a T-type and an L-type, are sufficient to recover substellar companion candidates with their expected spectra without a significant loss in S/N.

The uniform reduction of more 330 data sets from GPIES allowed the unprecedented characterization of the PDF of the residual noise, which significantly deviates from a Gaussian distribution. We showed that the FMMF has a uniform spatial distribution of false positives that does not feature any excesses due to the sector boundaries or the wind-butterfly. The improved performance provided by the FMMF also appeared in the ROC curves and the contrast curves that have been shown for GPIES. We built the contrast curves using a different detection threshold for each spectral type and algorithm such that it yields the same number of false positives in each case. The FMMF for example allows the detection of objects that are 50% fainter than a classical cross-correlation. We reaffirm that a hard 5σ detection threshold does not allow a meaningful comparison of the algorithm performance and can overestimate the sensitivity of a survey. In the future, a greater number of diverse algorithms should be compared following a similar recipe to the one presented in this paper.

To complete the study, we showed that the planet completeness can be derived from each contrast curve, which is simply the 50% detection completeness contour. These completeness contours can be used in a future exoplanet population study.

Appendix A
Method

In this appendix, we summarize the key theoretical results used in this paper. First, the KLIP framework and its effect on the planet PSF, i.e., self-subtraction, is introduced in Appendix A.1. Then, the principles of a matched filter are presented in Appendix A.2. Finally, the forward model is defined in Appendix A.3.

Throughout the paper, with some rare exceptions, scalars are lowercase Greek characters, vectors are bold lowercase Latin characters, and matrices are bold uppercase Latin characters. A summary of the notations can be found in Appendix B.

A.1. Karhunen-Loève Image Processing

A.1.1. Formalism

Karhunen-Loève image processing (KLIP) (Soummer et al. 2012) is one of the most widely used speckle subtraction

frameworks. The general concept of speckle subtraction algorithms is to use a library of reference images from which a model of the speckle pattern is computed and then subtracted for each science image, effectively whitening the noise. KLIP calculates the principal components of this reference library and filters out the higher order modes that contain more noise and more planet signal. The reference library is a set of images spatially scaled to the same wavelength and containing random realization of the correlated speckles. We use a general observing strategy by combining ADI and SDI observations, sometimes referred to as ASDI, which is possible with an IFS like GPI. This choice of strategy implies that the astrophysical signal will be found both in the reference images and the science image. As a consequence, the algorithm is subject to self-subtraction, which is discussed in Appendix A.1.2. A detailed description of the mathematical formalism underlying the KLIP algorithm can be found in Savransky (2015), which we briefly summarize here.

The $N_R \times N_{\mathrm{pix}}$ matrix $\boldsymbol{R}$ is defined by concatenating the vectorized mean subtracted reference images $\boldsymbol{r}_j$, which are vectors with N_{pix} elements, such that

$$\boldsymbol{R} = [\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_{N_R}]^\top. \qquad (6)$$

Then, the image-to-image sample covariance matrix describing the correlation of the images in the reference library is described, up to a proportional factor, by

$$\boldsymbol{C} = \boldsymbol{R}\boldsymbol{R}^\top, \qquad (7)$$

with its eigenvectors and eigenvalues $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_{N_R}$ and $\mu_1, \mu_2, ..., \mu_{N_R}$, respectively. The Karhunen-Loève images form the optimal orthonormal basis that best represents any realization of the noise in a least-squares sense. They are defined as

$$\boldsymbol{z}_k = \frac{1}{\sqrt{\mu_k}} \sum_{j=1}^{N_R} \boldsymbol{v}_k[j] \boldsymbol{r}_j. \qquad (8)$$

The images can be written in a more compact form by defining the matrices $\boldsymbol{V}_K = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_K]^\top$, $\boldsymbol{Z}_K = [\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_K]^\top$ and the diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_K)^\top$, with $K \leqslant N_R$ being the number of selected Karhunen-Loève images kept for the subtraction. Then, the collection of horizontally stacked Karhunen-Loève images $\boldsymbol{Z}_K = [\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_K]^\top$ is equivalently written as

$$\boldsymbol{Z}_K = \sqrt{\boldsymbol{\Lambda}^{-1}} \boldsymbol{V}_K \boldsymbol{R}. \qquad (9)$$

The speckle subtraction consists in subtracting the projection of the science image i onto the Karhunen-Loève basis from itself, which is given by

$$\boldsymbol{p} = \boldsymbol{i} - \sum_{k=1}^{K} \langle \boldsymbol{i} | \boldsymbol{z}_k \rangle \boldsymbol{z}_k = \boldsymbol{i} - \boldsymbol{Z}_K^\top \boldsymbol{Z}_K \boldsymbol{i}, \qquad (10)$$

where p refers to the processed image. The science and the reference images are often part of the same data set, and the science image refers to the image of a specific exposure and wavelength from which the speckle pattern is being subtracted. Typically, speckle subtraction algorithms are performed independently on small sectors of the image to account for spatial variations of the speckle noise, instead of using the full image, but the formalism is identical.
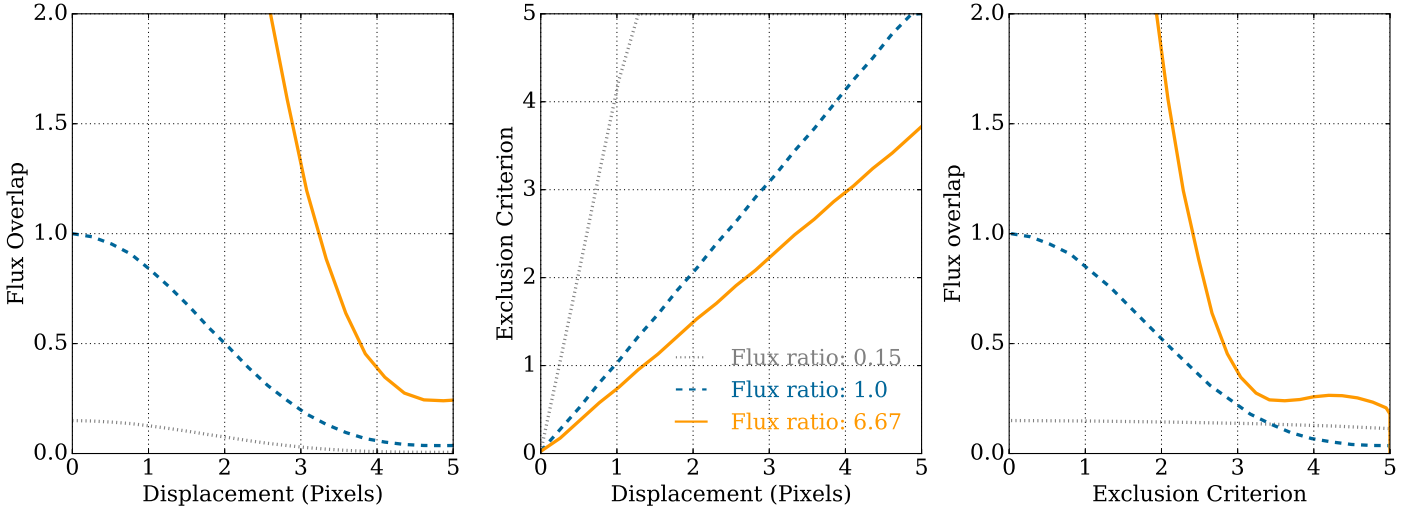
Figure 20. Conversion plots between the exclusion criterion, the PSF flux overlap, and the PSF displacement between two images. The flux ratio is defined as the ratio of the peak value of the PSF in the reference frame over the peak value in the science image. The ratio of 0.15 represents a typical value for the peak to bottom ratio of a spectrum with a strong methane signature.

A.1.2. Self-subtraction

In the this section, we review the effect of an astrophysical signal as a small perturbation to Equation (10), following Pueyo (2016). In this section only, variables defined in Appendix A.1.1 are assumed to be planet free, and the hat indicates the perturbed variable, such that

$$\forall j \in [1, N_R], \quad \hat{r}_j = r_j + \Delta r_j, \tag{11}$$

with $\Delta r_j = \epsilon\, a_j$, where a_j is the normalized planet signal. The planet signal is defined by its amplitude ϵ, a spectrum and the instrumental PSF before speckle subtraction. It is more convenient to normalize the planet signal to the same brightness as the host star such that ϵ is in practice the contrast of the planet. Similarly, we write $\hat{z}_j = z_j + \Delta z_j$, $\hat{Z}_K = Z_K + \Delta Z_K$, $\hat{i} = i + \Delta i = i + \epsilon\, a$, and $\hat{p} = p + \Delta p \approx p + \epsilon\, m$, where m is the planet model that will be used to build the matched-filter template t introduced in Equation (2).

Applying KLIP to the perturbed images gives

$$\hat{p} = \hat{i} - \hat{Z}_K^\top \hat{Z}_K \hat{i},$$
$$p + \Delta p = (i + \epsilon\, a)$$
$$\qquad - (Z_K + \Delta Z_K)^\top \times (Z_K + \Delta Z_K)(i + \epsilon\, a), \tag{12}$$

and expanding Equation (12), one identifies p, which should only contain residual speckles such that

$$p = i - Z_K^\top Z_K i \approx 0. \tag{13}$$

Then, the perturbed processed image is given by

$$\Delta p = \epsilon\, a \quad \text{Original planet signal,}$$
$$\quad - \epsilon Z_K^\top Z_K a \quad \text{Oversubtraction,}$$
$$\quad - (Z_K^\top \Delta Z_K + (Z_K^\top \Delta Z_K)^\top)i \quad \text{Self-subtraction,}$$
$$\quad + \mathcal{O}(\epsilon^2) \quad \text{Neglected second-order terms.} \tag{14}$$

The second and third terms of Equation (14) are the first-order distortions introduced by KLIP on the planet signal, as shown

in Figure 2(b). The oversubtraction comes from the projection of the planet onto the unperturbed Karhunen-Loève modes, while the self-subtraction is the result of the projection of the speckles onto the perturbation of the modes. The former is unavoidable in a least-squares approach. The characteristic patterns for the self-subtraction are negative lobes in the radial and azimuthal direction that are due to the movement of the planet in the reference frames with wavelength and parallactic angle. The self-subtraction vanishes when there is no planet signal in the reference library as $\Delta Z_K = 0$, which is the case when using a reference star differential imaging (RDI) approach. When the locally optimized combination of images (LOCI) of Lafrenière et al. (2007a) is used, another way to avoid self-subtraction is to use optimization and subtraction region that do not overlap (see, Marois et al. 2010). However, the price of these strategies is a possibly less effective speckle subtraction.

A.1.3. Reference Library Exclusion Criterion

Part of the planet signal is lost in the processed image, especially when the planet PSF spatially overlaps in the reference images and in the science image. The distortion of the central peak of the PSF is more important when the overlap is large. A common practice to limit the distortion with ADI and SDI is to only include images in the reference library in which the planet has moved significantly compared to the science image. The importance of the distortion also depends on the relative brightness of the planet in the science image and in the references, which itself depends on the planet spectrum. Therefore, the selection of the reference images can be based on the relative flux overlap with the planet signal in the science image (Marois et al. 2014) instead of the sole consideration of the displacement.

The exclusion parameter chosen in this paper is a hybrid between a flux overlap and a displacement parameter. It corresponds to a pure displacement when comparing images at similar wavelength, but it more or less accepts images in the reference library when the wavelengths are different. This

(a) 20th exposure, 1.60 μm (b) 20th exposure, 1.65 μm
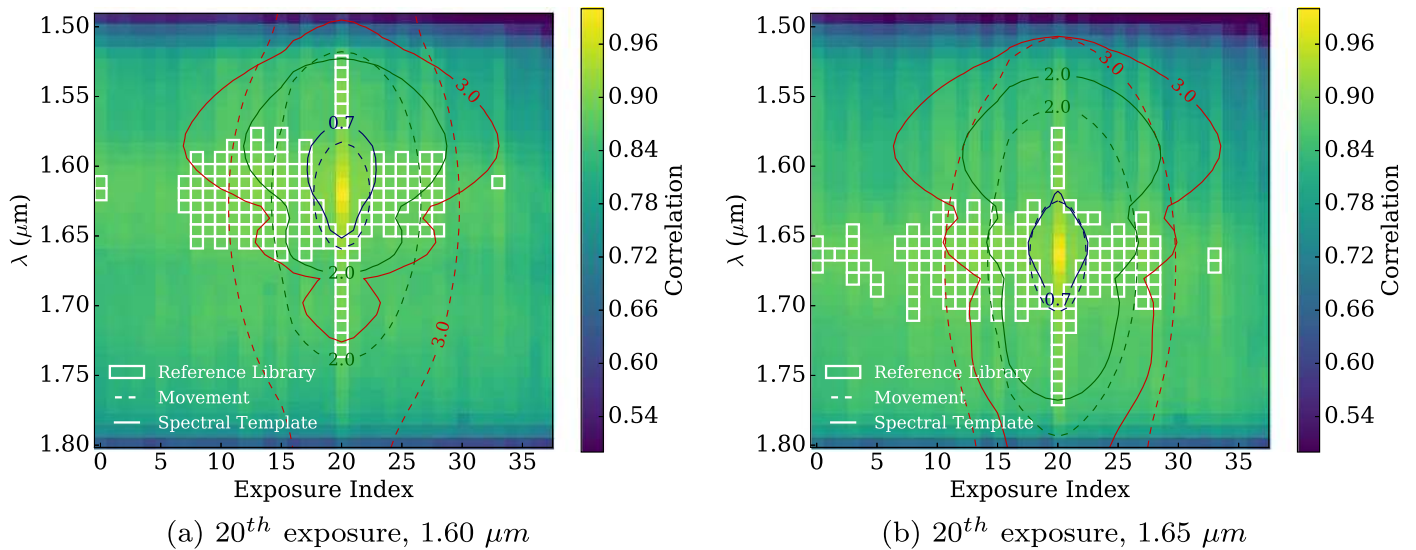
Figure 21. Illustration of the reference library selection for a given science image in a GPI data set. Two examples are shown with science images in the same exposure but at different wavelengths in H band: (a) 1.60 and (b) 1.65 μm. The color map represents the correlation between each image in the data set with the science image. The data set is represented as a function of the exposure (x-axis) and the wavelength (y-axis). The exclusion criterion used in this paper for the reference library selection is represented with solid contours and includes a spectral template. It is a hybrid between a pure displacement in pixels, represented with dashed lines, and a flux overlap criterion. The white squares highlight the 150 most correlated images satisfying a 0.7 exclusion criterion, which will form the reference library. More images tend to be exluded around 1.58 μm because we chose a spectrum with a strong methane signature, resulting in a peak flux at this wavelength. The data set used here is the 51 Eridani b GPIES discovery epoch on 2014 December 18. The signal displacement is calculated for an hypothetical planet at 30 pixel ($\approx$0.42 as) separation, and the total rotation during the full sequence is 24°.

implementation of the exclusion criterion has given a significantly better result than any of the two other methods. The conversion between the three metrics is given for a specific planet PSF flux ratio between the science and the reference images in Figure 20. The selection of the reference images is illustrated in Figure 21 in a typical GPI data set. Only the 150 most correlated images with any given science frame are kept for the speckle subtraction in order to limit the size of the covariance matrix C and speed up the reduction. There is a stronger correlation between images with respect to wavelength than with respect to time, suggesting that SDI will give better results than ADI.

The choice of the exclusion criterion is a trade-off between an effective speckle subtraction and a limited self-subtraction. A low value of the exclusion criterion includes more correlated images in the reference library and results in a more aggressive reduction because of the higher degree of self-subtraction. The aggressiveness is also affected by the number of Karhunen-Loève modes used in the subtraction, but we keep this number fixed to $K = 30$ throughout this paper. We describe how the optimization of the exclusion parameter is performed in Section 4.1.

A.2. Matched Filter

In the field of signal processing, a matched filter is the linear filter maximizing the S/N of a known signal in the presence of additive noise (Kasdin & Braems 2006). If the noise samples are independent and identically distributed, the matched filter corresponds to the cross-correlation of a template with the noisy data. In the context of high-contrast imaging, the pixels are neither independent nor identically distributed (i.e., heteroskedastic). Note that the matched filter is a very general tool that can be applied regardless of the algorithm used for speckle subtraction.

We define d as the vectorized speckle-subtracted data set representing a series of individual exposures of a star. Because GPI

is an integral field spectrograph (IFS), the one-dimensional array d typically contains $N_x N_y N_\lambda N_\tau$ elements, where N_x and N_y are the number of pixels in the spatial dimensions (for a square image, $N_x = N_y$), N_λ is the number of spectral channels, and N_τ is the number of exposures. A typical GPIES data set is the concatenation of all the processed images $\{p_l\}$ from an epoch, with $l = (\tau, \lambda)$, so $N_x N_y N_\lambda N_\tau = 281 \times 281 \times 37 \times 38$. If we assume that the data set contains a single point source, it can be decomposed as

$$d = \epsilon\, t(\rho, \theta) + n. \qquad (15)$$

With $t(\rho, \theta)$ the vectorized template of the planet with separation and position angle (ρ, θ). In practice we conveniently normalize the template to the same flux as the star such that ϵ is again the contrast. It is the concatenation of the planet model PSF for each image $\{m_l\}$ and it has the same dimension as d. The second term n is for now assumed to be a Gaussian random vector assuming a zero mean and a covariance matrix Σ. For conciseness, the (ρ, θ) notation in $t(\rho, \theta)$ is dropped in the remainder of the paper, but the position dependence should be assumed for any point-source-related variables.

A matched filter can also be interpreted as a maximum likelihood $\mathcal{L}(\epsilon, \rho, \theta)$ estimator of the planet amplitude and position assuming a Gaussian noise:

$$\mathcal{L}(\epsilon, \rho, \theta) = \frac{1}{\sqrt{2\pi|\Sigma|}} \exp\left[\frac{-1}{2}(d - \epsilon\, t)^\top \Sigma^{-1}(d - \epsilon\, t)\right]. \qquad (16)$$

The most likely amplitude $\tilde{\epsilon}$ as a function of position is given by

$$\tilde{\epsilon}(\rho, \theta) = \frac{d^\top \Sigma^{-1} t}{t^\top \Sigma^{-1} t}. \qquad (17)$$

Table 1
Definition of Scalar Variables

Symbol	Type	Description
N_x, N_y	$\mathbb{N} \times \mathbb{N}$	Dimensions of the image.
N_{pix}	$\mathbb{N}$	Number of pixels in a sector. If the sector is the entire image, then $N_{\mathrm{pix}} = N_x N_y$.
N_λ	$\mathbb{N}$	Number of wavelength channels in the IFS data cubes.
N_τ	$\mathbb{N}$	Number of exposures in an epoch.
N_R	$\mathbb{N}$	Number of images in the reference library.
K	$\mathbb{N}$	Number of selected Karhunen-Loève images to be used in the speckle subtraction.
ρ	$\mathbb{R}^+$	Separation to the star.
θ	$\mathbb{R}$	Position angle with the origin at the image north.
σ_l	$\mathbb{R}^+$	Local standard deviation of the noise at the position (ρ, θ) in the speckle-subtracted images.
$\mathcal{S}(\rho, \theta)$	$\mathbb{R}$	Theoretical S/N at the position (ρ, θ) assuming independent Gaussian noise. This is not the true S/N of the signal.
ϵ	$\mathbb{R}^+$	Contrast of the point source relative to the host star. Different units could be used depending on the normalization of the planet model.
$\tilde{\epsilon}(\rho, \theta)$ or $\tilde{\epsilon}$	$\mathbb{R}$	Maximum likelihood estimation of the amplitude of the planet signal ϵ at the position (ρ, θ).
σ_ϵ	$\mathbb{R}^+$	Standard deviation of the planet contrast at the position (ρ, θ). $\sigma_{\epsilon 1}$ and $\sigma_{\epsilon 2}$ refer to the first and follow-up observation of a star, respectively.
μ_j	$\mathbb{R}$	Eigenvalues of the matrix $\boldsymbol{C}$.
γ	$\mathbb{R}$	Conversion factor from the contrast matched-filter values $\mathcal{S}$ to the contrast ζ such that $\zeta = \gamma \mathcal{S}$.
η	$\mathbb{R}$	S/N detection threshold. η_2 is the detection threshold in the follow-up observation, which can be lower.
σ_γ	$\mathbb{R}^+$	Standard deviation of n_γ
$\sigma_\mathcal{S}$	$\mathbb{R}^+$	Standard deviation of the noise in the matched-filter map as a function of position.

The theoretical S/N of a point source is given as a function of position by

$$\mathcal{S}(\rho, \theta) = \tilde{\epsilon}/\sigma_{\tilde{\epsilon}} = \frac{\boldsymbol{d}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{t}}{\sqrt{\boldsymbol{t}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{t}}}, \qquad (18)$$

which is a linear function of the square root of the maximized log-likelihood (Röver 2011; Cantalloube et al. 2015), where maximization is carried out over the point-source amplitude.

When we assume that the noise is uncorrelated and that its variance is constant in the neighborhood of the planet, the planet amplitude becomes

$$\tilde{\epsilon}(\rho, \theta) = \frac{\sum_l \boldsymbol{p}_l^\top \boldsymbol{m}_l / \sigma_l^2}{\sum_l \boldsymbol{m}_l^\top \boldsymbol{m}_l / \sigma_l^2}, \qquad (19)$$

where σ_l is the local standard deviation at the position (ρ, θ). Note that the planet model $\boldsymbol{m}_l$ rapidly approaches rapidly when moving away from its center (ρ, θ), allowing one to only consider postage-stamp-sized images containing the putative planet.

The theoretical S/N of the planet becomes

$$\mathcal{S}(\rho, \theta) = \frac{\sum_l \boldsymbol{p}_l^\top \boldsymbol{m}_l / \sigma_l^2}{\sqrt{\sum_l \boldsymbol{m}_l^\top \boldsymbol{m}_l / \sigma_l^2}}. \qquad (20)$$

A detection can be claimed when the S/N is such that the observation cannot be explained by the null-hypothesis.

A.3. KLIP Forward Model

It is common practice to perform a matched filter with a simple Gaussian or a box template. Such templates are not optimal because they at best waste planet signal because of the self-subtraction. This paper uses the KLIP forward model from Pueyo (2016) to improve the matched-filter planet-detection sensitivity.

The perturbation of the processed image $\frac{\Delta p}{\epsilon}$ from Equation (14) should be used as the template for the matched filter (also known as planet signature in Cantalloube et al. 2015 and in the forward model in Pueyo 2016). It can also be thought of as the derivative of the KLIP operator along the direction defined by the planet.

The ideal template $\frac{\Delta p}{\epsilon}$ is a function of the original planet PSF and spectrum. The difficulty is the estimation of the Karhunen-Loève perturbation term $\Delta \boldsymbol{Z}_K$ because it is a nonlinear function of the planet flux. Therefore Pueyo (2016) (E18) derived a first-order approximation of this term that can be used to calculate a close estimation of the planet PSF after speckle subtraction:

$$\frac{\Delta z_k}{\epsilon} = -\left(\frac{\boldsymbol{v}_k^\top \boldsymbol{C}_{AR} \boldsymbol{v}_k}{2\mu_k} \right) z_k + \sum_{j=1, j \neq k}^{N_R} \left(\sqrt{\frac{\mu_j}{\mu_k}} \frac{\boldsymbol{v}_j^\top \boldsymbol{C}_{AR} \boldsymbol{v}_k}{\mu_k - \mu_j} \right) z_j$$
$$+ \frac{1}{\sqrt{\mu_k}} \boldsymbol{A}^\top \boldsymbol{v}_k. \qquad (21)$$

With $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_{N_R}]^\top$ and $\boldsymbol{C}_{AR} = \boldsymbol{A}\boldsymbol{R}^\top + (\boldsymbol{A}\boldsymbol{R}^\top)^\top$.

The forward model of the planet, also called template, is then given by

$$\boldsymbol{m} = \boldsymbol{a} - \boldsymbol{Z}_K^\top \boldsymbol{Z}_K \boldsymbol{a} - (\boldsymbol{Z}_K^\top \Delta \boldsymbol{Z}_K + (\boldsymbol{Z}_K^\top \Delta \boldsymbol{Z}_K)^\top) \frac{\boldsymbol{i}}{\epsilon}, \qquad (22)$$

with $\boldsymbol{m} \approx \frac{\Delta p}{\epsilon}$. As a reminder, we have dropped the (ρ, θ) upper script at the very beginning of this paper, but $\boldsymbol{m}$, $\boldsymbol{a}$ and $\Delta \boldsymbol{Z}_K$ all depend on the position in the image.

Appendix B
Mathematical Notations

We present here a summary of the notations used in this paper. Tables 1, 2, and 3 contain the definition of the scalars, the vectors, and the matrices, respectively.

Table 2
Definition of Vectors

Symbol	Length	Description
r_j	N_{pix}	Vectorized images or sectors from the KLIP reference library.
$\hat{r}_j$	N_{pix}	Vectorized perturbed images or sectors from the KLIP reference library including some planet signal.
Δr_j	N_{pix}	Vectorized perturbation of the images or sectors from the KLIP reference library.
v_j	N_R	Eigenvectors of the matrix C.
z_k	N_{pix}	Karhunen-Loève images or sectors for the reference library matrix R.
$\hat{z}_k$	N_{pix}	Perturbed Karhunen-Loève images or sectors for the reference library matrix R including planet signal.
Δz_k	N_{pix}	Perturbation of the Karhunen-Loève images or sectors for the perturbed reference library.
i	N_{pix}	Vectorized science image or sector before speckle subtraction.
$\hat{i}$	N_{pix}	Vectorized perturbed science image or sector before speckle subtraction including planet signal.
Δi	N_{pix}	Vectorized perturbation of the science image or sector before speckle subtraction.
a_j	N_{pix}	Vectorized planet signal in the reference image or sector $\hat{r}_j$ before speckle subtraction.
a	N_{pix}	Vectorized planet signal in the science image or sector before speckle subtraction.
p	N_{pix}	Vectorized processed image or sector after speckle subtraction.
$\hat{p}$	N_{pix}	Vectorized perturbed processed image or sector after speckle subtraction including planet signal.
Δp	N_{pix}	Vectorized perturbation of the processed image or sector after speckle subtraction.
m	N_{pix}	Vectorized planet model in the processed image or sector after speckle subtraction.
d	$N_x N_y N_\lambda N_\tau$	Vectorized speckle-subtracted data set.
$t(\rho, \theta)$ or t	$N_x N_y N_\lambda N_\tau$	Vectorized planet template for the speckle-subtracted data set.
n	$N_x N_y N_\lambda N_\tau$	Multivariate noise distribution of the speckle-subtracted data set. The noise is assumed to be Gaussian for any mathematical derivation.
p_l	$N_x N_y$	Vectorized speckle-subtracted image with exposure τ and wavelength λ.
$m_l(\rho, \theta)$ or m_l	$N_x N_y$	Vectorized planet template for the speckle-subtracted image with exposure τ and wavelength λ.

Table 3
Definition of Matrices

Symbol	Dimensions	Description
Σ	$(N_x N_y N_\lambda N_\tau)^2$	Covariance matrix of the noise n in the speckle-subtracted image
R	$N_R \times N_{pix}$	Reference library matrix such that $R = [r_1, r_2, ..., r_{N_R}]^\top$.
C	$N_R \times N_R$	Matrix proportional to the time covariance matrix defined as $C = RR^\top$.
V_K	$K \times N_R$	Eigenvector matrix defined as $V_K = [v_1, v_2, ..., v_K]^\top$.
Z_K	$K \times N_{pix}$	Karhunen-Loève image matrix defined as $Z_K = [z_1, z_2, ..., z_K]^\top$.
$\hat{Z}_K$	$K \times N_{pix}$	Perturbed Karhunen-Loève image matrix defined as $\hat{Z}_K = [\hat{z}_1, \hat{z}_2, ..., \hat{z}_K]^\top$ including planet signal.
ΔZ_K	$K \times N_{pix}$	Perturbation of the Karhunen-Loève image matrix defined as $\Delta Z_K = [\Delta z_1, \Delta z_2, ..., \Delta z_K]^\top$.
Λ	$K \times N_{pix}$	Diagonal matrix defined as $\Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_{N_R})^\top$.
A	$N_R \times N_{pix}$	Planet signal component in the reference library R and $A = [a_1, a_2, ..., a_{N_R}]^\top$.
C_{AR}	$N_R \times N_R$	Matrix used in the forward calculation defined as $C_{AR} = AR^\top + (AR^\top)^\top$.

References

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Bloemhof, E. E., Dekany, R. G., Troy, M., & Oppenheimer, B. R. 2001, ApJL, 558, L71

Bloemhof, E. E. 2004, Proc. SPIE, 5553, 281

Bowler, B. P., Liu, M. C., Shkolnik, E. L., & Tamura, M. 2015, ApJS, 216, 7

Cantalloube, F., Mouillet, D., Mugnier, L. M., et al. 2015, A&A, 582, A89

Caucci, L., Barrett, H. H., Devaney, N., & Rodríguez, J. J. 2007, JOSAA, 24, B13

Choquet, É, Pueyo, L., Soummer, R., et al. 2015, Proc. SPIE, 9605, 96051P

Fitzgerald, M. P., & Graham, J. R. 2006, ApJ, 637, 541

Galicher, R., Marois, C., Macintosh, B., et al. 2016, A&A, 594, A63

Hinkley, S., Oppenheimer, B. R., Soummer, R., et al. 2007, ApJ, 654, 633

Ingraham, P., Perrin, M. D., Sadakuni, N., et al. 2014, Proc. SPIE, 9147, 91477O

Johnson-Groh, M. 2017, arXiv:1704.00317

Kasdin, N. J., & Braems, I. 2006, ApJ, 646, 1260

Konopacky, Q. M., Thomas, S. J., Macintosh, B. A., et al. 2014, Proc. SPIE, 9147, 914784

Lafrenière, D., Doyon, R., Marois, C., et al. 2007b, ApJ, 670, 1367

Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., & Artigau, É 2007a, ApJ, 660, 770

Macintosh, B., Graham, J. R., Barman, T., et al. 2015, Sci, 350, 64

Macintosh, B., Graham, J. R., Ingraham, P., et al. 2014, PNAS, 111, 12661

Maire, J., Ingraham, P. J., De Rosa, R. J., et al. 2014, Proc. SPIE, 9147, 914785

Marois, C., Correia, C., Galicher, R., et al. 2014, Proc. SPIE, 9148, 91480U

Marois, C., Doyon, R., Racine, R., & Nadeau, D. 2000, PASP, 112, 91

Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., & Nadeau, D. 2006, ApJ, 641, 556

Marois, C., Lafrenière, D., Macintosh, B., & Doyon, R. 2008, ApJ, 673, 647

Marois, C., Macintosh, B., & Véran, J.-P. 2010, Proc. SPIE, 7736, 77361J

Mawet, D., Milli, J., Wahhaj, Z., et al. 2014, ApJ, 792, 97

Morzinski, K. M., Males, J. R., Skemer, A. J., et al. 2015, ApJ, 815, 108

Nielsen, E. L., Liu, M. C., Wahhaj, Z., et al. 2013, ApJ, 776, 4

Pairet, B., Jacques, L., Gomez Gonzalez, C., & Absil, O. 2016, in Proc. iTWIST16, Int. Traveling Workshop on Interactions Between Sparse Models and Technology (Aalborg: Aalborg Univ.), https://dial.uclouvain.be/pr/boreal/object/boreal:178310

Perrin, M. D., Ingraham, P., Follette, K. B., et al. 2016, Proc. SPIE, 9908, 990837

Perrin, M. D., Sivaramakrishnan, A., Makidon, R. B., Oppenheimer, B. R., & Graham, J. R. 2003, ApJ, 596, 702

Pickles, A. J. 1998, PASP, 110, 863

Pueyo, L. 2016, ApJ, 824, 117

Röver, C. 2011, PhRvD, 84, 122004

Saumon, D., Marley, M. S., Abel, M., Frommhold, L., & Freedman, R. S. 2012, ApJ, 750, 74

Savransky, D. 2015, ApJ, 800, 100

Sivaramakrishnan, A., Lloyd, J. P., Hodge, P. E., & Macintosh, B. A. 2002, ApJL, 581, L59

Sivaramakrishnan, A., & Oppenheimer, B. R. 2006, ApJ, 647, 620

Soummer, R., & Aime, C. 2004, Proc. SPIE, 5490, 495

Soummer, R., Ferrari, A., Aime, C., & Jolissaint, L. 2007, ApJ, 669, 642

Soummer, R., Pueyo, L., & Larkin, J. 2012, ApJL, 755, L28

Sparks, W. B., & Ford, H. C. 2002, ApJ, 578, 543

Vigan, A., Bonavita, M., Biller, B., et al. 2017, A&A, in press (arXiv:1703.05322)

Wang, J. J., Graham, J. R., Pueyo, L., et al. 2016, AJ, 152, 97

Wang, J. J., Ruffio, J.-B., De Rosa, R. J., et al. 2015, pyKLIP: PSF Subtraction for Exoplanets and Disks, Astrophysics Source Code Library, ascl:1506.001

Wolff, S. G., Perrin, M. D., Maire, J., et al. 2014, Proc. SPIE, 9147, 91477H