A Bayesian Framework for Exoplanet Direct Detection and Non-detection

Jean-Baptiste Ruffio[1] , Dimitri Mawet[2,3] , Ian Czekala[1,7] , Bruce Macintosh[1] , Robert J. De Rosa[4] , Garreth Ruane[2,8] , Michael Bottom[3] , Laurent Pueyo[5] , Jason J. Wang[4] , Lea Hirsch[4] , Zhaohuan Zhu[6] , and Eric L. Nielsen[1]

[1] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA 94305, USA; jruffio@stanford.edu
[2] Department of Astronomy, California Institute of Technology, Pasadena, CA 91125, USA
[3] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA
[4] Astronomy Department, University of California, Berkeley; Berkeley, CA 94720, USA
[5] Space Telescope Science Institute, Baltimore, MD 21218, USA
[6] Department of Physics and Astronomy, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA

Abstract

Rigorously quantifying the information in high-contrast imaging data is important for informing follow-up strategies to confirm the substellar nature of a point source, constraining theoretical models of planet–disk interactions, and deriving planet occurrence rates. However, within the exoplanet direct imaging community, non-detections have almost exclusively been defined using a frequentist detection threshold (i.e., contrast curve) and associated completeness. This can lead to conceptual inconsistencies when included in a Bayesian framework. A Bayesian upper limit is such that the true value of a parameter lies below this limit with a certain probability. The associated probability is the integral of the posterior distribution with the upper limit as the upper bound. In summary, a frequentist upper limit is a statement about the detectability of planets while a Bayesian upper limit is a statement about the probability of a parameter to lie in an interval given the data. The latter is therefore better suited for rejecting hypotheses or theoretical models based on their predictions. In this work we emphasize that Bayesian statistics and upper limits are more easily interpreted and typically more constraining than the frequentist approach. We illustrate the use of Bayesian analysis in two different cases: (1) with a known planet location where we also propose to use model comparison to constrain the astrophysical nature of the point source and (2) gap-carving planets in TW Hya. To finish, we also mention the problem of combining radial velocity and direct imaging observations.

Key words: instrumentation: adaptive optics – instrumentation: high angular resolution – methods: statistical – planetary systems – planet–disk interactions – planets and satellites: detection

1. Introduction

Direct imaging is a method that can spatially resolve exoplanets' light from their host star using large telescopes, adaptive optics, coronagraphs and sophisticated data processing. With ground-based telescopes, this technique currently allows the detection of young ($< 300\,\mathrm{Myr}$), massive ($> 2\,M_{\mathrm{Jup}}$), self-luminous exoplanets at host-star separations that have not yet been probed by indirect methods ($a > 5\,\mathrm{au}$). Direct imaging surveys of previously unobserved stars mostly produce non-detections ($> 98\%$ of stars do not have a detectable planet with current instruments). A wide range of science can be drawn from these null results, but thinking about the definition of upper limits is important. In this paper, we will discuss some of these applications.

First, we consider cases in which the position of the object is known. Most detected point sources are background stars, therefore confirmation of their planetary nature requires follow-up observations (Black 1980). Several strategies can be adopted depending on the information at hand. It is common practice to use the upper limit from the non-detection and/or the frequencies of the different astrophysical signals to make the case for a planet and to reject the background or foreground hypothesis (Meshkat et al. 2013; Macintosh et al. 2015; Wagner et al. 2016; Chauvin et al. 2017; Nielsen et al. 2017). When a point source has been detected in one of two spectral

bands, for example, the upper limit on the second band can place limits on its color and in some cases we can reject the possibility of it being a star. Generally, a red object or an object showing significant spectral features characteristic of low-temperature atmospheres will favor lower masses and this will increase the likelihood of it being bound. This can be a powerful tool to prioritize follow-up observations. Fomalhaut b is another interesting example of the use of upper limits in determining the nature of an object. The signal was discovered in the optical (Kalas et al. 2008) but all subsequent follow-up observations in the infrared yielded non-detections, which cast doubts on its planetary nature (Kalas et al. 2008; Marengo et al. 2009; Currie et al. 2012, 2013; Janson et al. 2012). The lack of infrared emission suggests that the signal comes from starlight scattered by a disk surrounding a planetary body. A more formal statistical approach, such as the one derived in this work, could be used to set tighter limits on the mass of a self-luminous planet or to compare different dust formation hypotheses (Kenyon et al. 2014).

Another application of upper limits is to rule out models where an undetected planet perturbs some visible source, such as by clearing a gap in a circumstellar dust disk. Ruane et al. (2017) used direct imaging data of the TW Hya system to constrain the masses or the accretion rates of hypothetical gap-carving planets. The mass upper limit is derived from the flux constraint using a planet formation model (Baraffe et al. 2003; Allard et al. 2012). In this case, the exact position of the planet is not known but the shape of the gap defines its orbit.

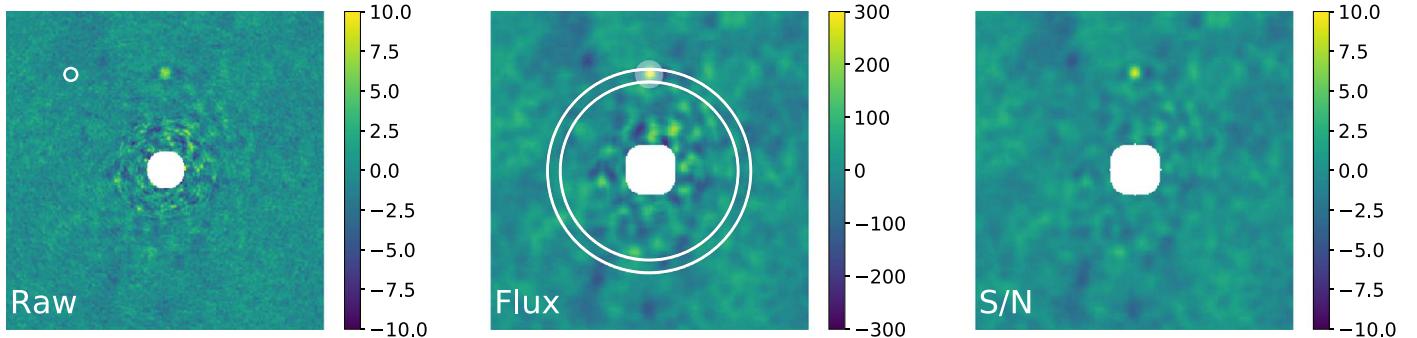**Figure 1.** Processed images of a Keck II/NIRC2 observation of TW Hya (Ruane et al. 2017) including a simulated 10σ point source located north of the center star. The left-hand image is the combined dataset after the speckles were removed (the units are arbitrary). The middle image is a flux map resulting from cross correlating a 10 pixel wide aperture (white empty circle in the left image) across the image. For each position, the sample standard deviation is computed from a 10 pixel wide annulus at the same separation after the surroundings of the pixel of interest have been masked out (white filled circle). The right-hand image is the resulting signal-to-noise ratio (S/N) map.

A final example of the use of upper limits is to combine radial velocity (RV) measurements and direct imaging observations. As time baselines keep growing and sensitivity improves, the overlap between the accessible mass and semimajor axis parameter space of the two methods keeps increasing. When a direct and a RV detection is available, it can be combined to infer the dynamical masses of a binary system. Spectroscopic binary stars that are spatially resolved can be used in this way to constrain the age of moving groups (Nielsen et al. 2016). A direct imaging non-detection can still bring useful constraints on the mass of a bound companion (Joergens et al. 2012; Hardy et al. 2015; Vaccaro et al. 2015).

The use of non-detections in the derivation of exoplanet occurrence rates is also extremely important (Cumming et al. 2008; Nielsen & Close 2010; Brandt et al. 2014; Bowler 2016; Galicher et al. 2016; Vigan et al. 2017). In this context, combining RV and direct imaging can also help to yield better estimates of planet frequency (Bryan et al. 2016; Lannier et al. 2017). However, this problem is complex and outside of the scope of this work. Bayesian based occurrences are very sensitive to the accuracy of the noise model, the characterization of which remains an on-going effort for high-contrast imaging.

Planet flux upper-limits are commonly defined using a frequentist approach from a detection threshold (e.g., contrast curve), usually 5σ with σ the standard deviation of the noise (Meshkat et al. 2013; Macintosh et al. 2015; Mesa et al. 2017; Nielsen et al. 2017). In principal, the detection threshold should be derived to set an acceptable false positive rate (Wahhaj et al. 2013) but in practice it is often set to the traditional 5σ limit. The upper-limit can also be thought of in terms of true positive fraction (i.e., completeness). By definition, the detection threshold corresponds to a 50% completeness. The fundamental conceptual differences between frequentist and Bayesian upper limits have been detailed in the context of gravitational wave detection (Finn 1998; Abbott et al. 2004; Brady et al. 2004; Röver et al. 2011). For example, Finn (1998) emphasizes that Bayesian analysis makes a measure of our degree of belief in a proposition while the frequentist analysis addresses our confidence in the ability of a procedure to decide if a signal is present or absent, making the Bayesian analysis better suited for the study of individual events. Röver et al. (2011) notes that a frequentist upper limit requires the maximization of the likelihood while the Bayesian upper limits

requires integration and also argues that the latter is more easily interpretable.

Although a detection threshold does provide a measure of the depth of the observation (i.e., its sensitivity to faint point sources), it is not a statement about the degree of belief in a given point-source flux given the data. While the detection upper limit indicates which planets would have been detected with a given completeness, a Bayesian upper limit is a statement about the probability of the planet flux given the data. A Bayesian upper limit f_{lim} is defined from the planet flux posterior and a fixed probability of the true flux to be smaller than this value given the data, $\mathcal{P}(f < f_{lim}|d)$, which is referred to as the cutoff probability in the following. The cutoff probability is the value of the cumulative distribution of the posterior at the position of the upper limit. Note that the posterior needs to be carefully defined as a function of the question that is asked to the data and the assumptions made. Additionally, using a detection threshold in all circumstances makes the interpretation of the results more difficult. The existing examples of combining RV and direct imaging measurements treat the RV data in a Bayesian framework while using the frequentist approach for direct imaging upper limits. Using the concepts presented here, Mawet et al. (2018) will make a step towards treating both data types in a consistent Bayesian framework.

We will illustrate this approach for direct imaging by revisiting practical cases of non-detection. Section 2 assumes that the location of the planet is known and it proposes to use Bayesian model comparison to decide the most probable nature of a candidate. Section 3 considers the case of a known orbit for the companion but no information of its precise location. This case is applicable to constraining the mass of an undetected accreting planet in a disk gap, for example. In Section 4, we look at combining RV and direct imaging measurements to constrain planet mass and orbital parameters. We conclude in Section 5.

2. Companion at a Known Location

2.1. Bayes' Rule and Upper Limit

In this section, we address the simple problem of defining an upper limit for the flux of a point source at a known location. A typical example can be found when the planet was clearly detected in only one of two spectral bands. In the literature, the
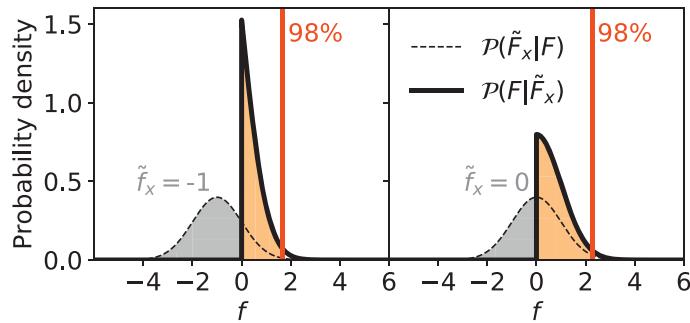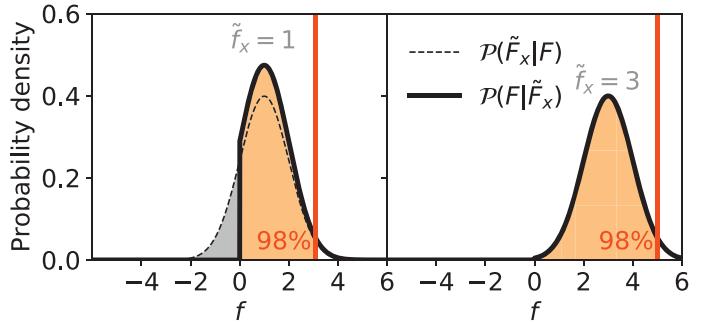
Figure 2. Flux posterior (solid line) and upper-limit as a function the estimated flux. The dashed line is the likelihood centered on the estimated flux, for which we assume a unit standard deviation. We assume a different S/N for each plot (-1σ, 0σ, 1σ and 3σ). The gray area represents the values of the flux that are rejected by the positive prior. The upper limit is defined from a 97.7% cutoff probability represented by the orange area under the curve. The different y-scales are due to the necessary normalization of the posteriors. For the 3σ case, the likelihood is hidden in the width of the line from the posterior because the effect of the prior is negligible.

quoted upper limits are often defined as the detection threshold (commonly 5σ) at the location of the object.

Figure 1 shows a typical example of a high-contrast image[9] in which a simulated $\approx 10\sigma$ point source was injected north of the center star. The left-hand image is the result of combining the single exposures after the speckle noise was removed using a principal component based approach (Soummer et al. 2012; Wang et al. 2015). The simulated point source was injected in the individual 357 frames before the speckle subtraction is performed. The simplest approach to compute the flux of a point source is aperture photometry, which consists of integrating the flux inside an aperture with a diameter that is equal to the width of the point spread function (PSF; Mawet et al. 2014). Such a flux map is equivalent to the cross correlation[10] of that aperture with the image (middle panel in Figure 1). A more accurate flux could be estimated using a matched filter (Ruffio et al. 2017), as shown in Section 3.1. For each position, the error bar σ on the flux estimate can be defined as the sample standard deviation computed from a 10 pixel wide annulus at the same separation from the primary star after the surroundings of the pixel of interest have been masked out (white filled circle) (Marois et al. 2008; Mawet et al. 2014). This is the same standard deviation that is used to define the detection threshold as a function of separation. A signal-to-noise ratio (S/N) map is simply the flux map divided by the standard deviation at each separation (right-hand image in Figure 1). Any signal brighter than 5σ is generally flagged as a possible candidate. The threshold can be defined to yield a reasonable number of false positives over a fixed field of view. In a direct imaging survey, this can be limited by the number of candidates on which follow-up observations can be performed or more generally by the follow-up strategy that maximizes the science return of the survey.

This detection threshold is fundamentally not a statement about the parameter space that was ruled out by the data. An upper-limit is better understood in the context of Bayesian inference. One first needs to define a probability, hereafter cutoff probability, of the true planet flux to fall below the upper limit. The Bayesian upper limit is then defined as the value for which the cumulative distribution of the posterior is equal to the cutoff probability. From now on, we will make a distinction between such an upper limit and a detection threshold. For a Gaussian noise with known standard deviation σ, an

unbounded uniform prior and a 97.7% cutoff probability, the upper limit is 2σ above the estimated flux (see Figure 2). The flux can be estimated even if the planet is not formally detected as long as its position is known. In the unlucky event of a very negative noise sample (e.g., lower than -2σ) at the location of the planet, the estimated flux and the upper-limit could become negative. This might be unsettling because we know that a flux is strictly positive but this will be corrected by a more informative prior, which will forbid negative values of the flux. For a given noise distribution, it is important to note that the upper-limit is a function of the data, here the estimated flux, while the detection threshold is a property of the noise. As we said, the choice of the threshold is also somewhat arbitrary, whereas a posterior is entirely defined by the properties of the noise, the data and the choice of a prior. Consequently, using the detection threshold in place of an upper limit does not make optimal use of the data.

Let F be the true planet flux random variable and $\tilde{F}_x$ be its estimate at the position x based on the observation. We denote random variables and random vectors with an upper case and their realization with a lower case. The posterior $\mathcal{P}(F|\tilde{F}_x)$ is the probability of the point-source flux given its estimated value from the observation. In this context, we define the data as being the flux map and following Bayes' rule (Sivia & Skilling 2006),

$$\mathcal{P}(F|\tilde{F}_x) = \frac{\mathcal{P}(\tilde{F}_x|F)\mathcal{P}(F)}{\mathcal{P}(\tilde{F}_x)}, \qquad (1)$$

where $\mathcal{P}(\tilde{F}_x|F)$ is the likelihood, $\mathcal{P}(F)$ the flux prior and $\mathcal{P}(\tilde{F}_x)$ acts as a normalization factor to ensure the posterior distribution has a unit integral. If the noise is Gaussian, the likelihood $\mathcal{P}(\tilde{F}_x|F)$ is given by the Gaussian distribution,

$$\mathcal{P}(\tilde{F}_x = \tilde{f}_x|F = f) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\frac{(f - \tilde{f}_x)^2}{\sigma^2}\right\}. \qquad (2)$$

The term $\mathcal{P}(\tilde{F}_x)$ is also called the marginal likelihood (or model evidence) and can be written as,

$$\mathcal{P}(\tilde{F}_x) = \int_{f=-\infty}^{\infty} \mathcal{P}(\tilde{F}_x|F = f)\mathcal{P}(F = f)df. \qquad (3)$$

The marginal likelihood is used in Bayesian model comparison. In the following, we use a simple uniform positive prior,

$$\mathcal{P}(F = f) = \begin{cases} \alpha, & \text{if } f > 0, \text{ with } \alpha \text{ constant} \\ 0, & \text{if } f \leqslant 0 \end{cases}. \qquad (4)$$

The uniform prior corresponds to the objective Jeffreys' prior (i.e., square root of the determinant of the Fisher information) for location parameters such as the mean of a normal distribution (Kass & Wasserman 1996). The uniform prior is improper (i.e., its integral is infinite) but the likelihood will be very constraining for large values of the flux f, so an upper bound for the prior of the companion flux will not make a significant difference. The inverse prior $\mathcal{P}(f) \propto 1/f$ is a common choice for positive real valued parameters because it is the only prior invariant by rescaling, i.e., $\mathcal{P}(f)df = \mathcal{P}(\beta f)d(\beta f)$. This means that the inverse prior does not favor any particular scale of the parameter but it requires a strictly positive lower bound to produce a normalizable posterior. The problem is that the previous likelihood does not constrain null values, which means that the inverse prior will dominate at small values of the flux. Consequently, due to the divergence of the integral of the inverse function, the upper limit will be extremely dependent on the choice of the lower bound in the prior. For example, the upper limit would tend to zero for an infinitely small lower bound. The prior could also be derived from our current knowledge of planet population. This would suggest a log-uniform planet mass prior Cumming et al. (2008) and the corresponding flux prior after a change of variable, but this would lead to the same difficulty near zero. In this context, the uniform prior remains a conservative choice for the definition of an upper limit. Indeed, a more relevant prior based on a planet population model would give more weights to lower fluxes and, therefore, decrease the upper limit. The uniform prior also facilitates the comparison of upper limits resulting from different works because it does not include a user-defined parameter. Note that in the absence of a prior, the posterior is simply equal to the likelihood centered on the estimated flux.

We will assume Gaussian noise in the following but the previous statements apply irrespective of the choice of noise distribution. We now define the cumulative distribution of a Gaussian distribution with standard deviation σ and mean $\tilde{f}$ as

$$\mathcal{C}_{\tilde{f},\sigma}(f_0) = \int_{f=-\infty}^{f_0} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\frac{(f-\tilde{f})^2}{\sigma^2}\right\}df. \quad (5)$$

and its inverse $\mathcal{Q}$, also known as the quantile function, such that $f_0 = \mathcal{Q}(\mathcal{C}(f_0))$.

The flux upper limit $f_{\lim}$ is the value for which the cumulative distribution of the posterior is equal to a cutoff probability η (for example 97.7%); i.e.,

$$\eta = \int_{f=-\infty}^{f_{\lim}} \mathcal{P}(f|\tilde{f}_x)df = \int_{f=-\infty}^{f_{\lim}} \frac{\mathcal{P}(\tilde{f}_x|f)\mathcal{P}(f)}{\mathcal{P}(\tilde{f}_x)}df, \quad (6)$$

with $\tilde{f}_x$ and σ_x the estimated flux and standard deviation at the position x. We also used Equation (1) in the second equality.

Then, we note that using Gaussian statistics $\mathcal{P}(\tilde{f}_x) = 1 - \mathcal{C}_{\tilde{f}_x,\sigma_x}(0)$ and using a positive uniform prior $\int_{f=-\infty}^{f_{\lim}} \mathcal{P}(\tilde{f}_x|f)\mathcal{P}(f)df = \mathcal{C}_{\tilde{f}_x,\sigma_x}(f_{\lim}) - \mathcal{C}_{\tilde{f}_x,\sigma_x}(0)$, which leads to

$$\eta = \frac{\mathcal{C}_{\tilde{f}_x,\sigma_x}(f_{\lim}) - \mathcal{C}_{\tilde{f}_x,\sigma_x}(0)}{1 - \mathcal{C}_{\tilde{f}_x,\sigma_x}(0)}. \quad (7)$$

After rearranging the terms and using the quantile function, one finds

$$f_{\lim} = \mathcal{Q}_{\tilde{f}_x,\sigma_x}(\eta + (1 - \eta)\mathcal{C}_{\tilde{f}_x,\sigma_x}(0)). \quad (8)$$

Figure 2 illustrates the posterior and upper-limit for different measured fluxes assuming a unit standard deviation. Note that if we drop the positivity constraint on the prior $\mathcal{P}(F)$, or if the estimated flux $\tilde{f}_x$ is large, then the 97.7% upper-limit does correspond to 2σ above the measured value. Figure 2 highlights how the upper limit closely depends on the realization of the noise, which is not true of the detection threshold. We encourage the direct imaging community to consider quoting upper-limits based on this definition instead of a detection threshold.

We have assumed that the standard deviation of the flux estimate was known. In practice, the standard deviation is estimated from the data, which means that it needs to be marginalized over when it is poorly constrained (e.g., when the number of noise realizations in the annulus of Figure 1 is small). The planet flux posterior can be defined as the two-sample t-test in which one of the samples has one element (corresponding to the location of the planet) and the other sample contains all of the pixels of a region with similar noise properties but free of astrophysical signal (Mawet et al. 2014). The second sample is often defined as the pixels taken at the same projected separation but located one resolution element apart from each other. The goal of a two-sample t-test, with unequal sample sizes but equal variance, is to estimate the difference between the means of two samples given that the variance is unknown and must be estimated from the data itself. In this case, the differences of the means is no other that the planet flux and its posterior must be marginalized over the uncertainty of the sample means and standard deviation, which results in a Student-t distribution:

$$\mathcal{P}(F = f|D) \propto \left[\frac{t_f}{N-1} + 1\right]^{-((N-1)+1)/2}, \quad (9)$$

with

$$t_f = \frac{(f - \tilde{f}_x)^2}{\left(\sum_k \tilde{f}_k^2 / (N-1)\right)\sqrt{1/N+1}}, \quad (10)$$

where N is the number of elements without an astrophysical signal. Equation (8) can still be used after redefining $\mathcal{C}$ and $\mathcal{Q}$ using Equation (9) in place of the normal distribution.

2.2. Bayesian Model Comparison: Is It a Star?

A recurring problem in direct imaging is to constrain the astrophysical nature of a candidate given a set of concurrent observations in different spectral bands. In this context, the nature of a point-like source can, for example, be a background star ($\mathcal{H}_\star$), a galaxy, a brown dwarf (background/foreground or gravitationally bound), a planet or a false positive ($\varnothing$), which is also known-as a null hypothesis. Point-source detections are a common occurrence in direct imaging surveys due to the prevalence of background stars, which is why we would like to prioritize their follow-up observations to optimize the discovery of new planets. For the purpose of this section, we will assume that we have two broadband observations, one of which could be a non-detection. With a single detection, it is not possible to compare the astrometric measurements of the

Table 1
Evidence Against $\mathcal{H}_1$ Compared to $\mathcal{H}_2$ Given the Value of the Bayes Factor, $B = \mathcal{P}(\mathcal{H}_2|D)/\mathcal{P}(\mathcal{H}_1|D)$, from Kass & Raftery (1995)

$\log_{10}(B)$	B	Evidence against $\mathcal{H}_1$
0–1/2	1–3.2	Not worth more than a bare mention
1/2–1	3.2–10	Substantial
1–2	10–100	Strong
>2	>100	Decisive

candidate to that of a background object. It is common practice to use the upper limit from the non-detection and/or the frequencies of the different astrophysical signals to make the case for a planet and reject the background or foreground hypothesis (Meshkat et al. 2013; Macintosh et al. 2015; Wagner et al. 2016; Chauvin et al. 2017; Nielsen et al. 2017).

Deciding between alternative hypotheses, which are here the possible classes of an astrophysical signal, can be done more generally using Bayesian model comparison. Given an hypothesis $\mathcal{H}$, data D and model parameters Θ, the marginal probability $\mathcal{P}(D|\mathcal{H})$ is the probability of obtaining the data assuming that an hypothesis is true. This is also defined as the normalization factor in the denominator of the Bayes' rule:

$$\mathcal{P}(\Theta|D, \mathcal{H}) = \frac{\mathcal{P}(D|\Theta, \mathcal{H})\mathcal{P}(\Theta|\mathcal{H})}{\mathcal{P}(D|\mathcal{H})}. \quad (11)$$

This marginal probability also appears in the expression of the posterior probability of an hypothesis $\mathcal{P}(\mathcal{H}|D) \propto \mathcal{P}(D|\mathcal{H})\mathcal{P}(\mathcal{H})$. The ratio of the posterior probability of two hypotheses $\mathcal{H}_1$ and $\mathcal{H}_2$, called the Bayes factor, is then given by

$$\frac{\mathcal{P}(\mathcal{H}_2|D)}{\mathcal{P}(\mathcal{H}_1|D)} = \frac{\mathcal{P}(D|\mathcal{H}_2)\mathcal{P}(\mathcal{H}_2)}{\mathcal{P}(D|\mathcal{H}_1)\mathcal{P}(\mathcal{H}_1)}, \quad (12)$$

with $\mathcal{P}(\mathcal{H}_1)$ and $\mathcal{P}(\mathcal{H}_2)$ the prior probability that each hypothesis to be true. The hypothesis $\mathcal{H}_2$ is preferred when the ratio is large. The significance of a given Bayes factor can be read out of published tables (see Table 1, Jeffreys 1961; Kass & Raftery 1995; Robert et al. 2011).

In the same way as we defined $\tilde{F}_x$, we write $\tilde{G}_x$ as the estimated photometry in a second spectral band. The probability of the null hypothesis $\mathcal{H} = \varnothing$ given the observations, $\tilde{F}_x$ and $\tilde{G}_x$, is given by

$$\mathcal{P}(\varnothing|\tilde{F}_x, \tilde{G}_x) = \frac{\mathcal{P}(\tilde{F}_x, \tilde{G}_x|\varnothing)\mathcal{P}(\varnothing)}{\mathcal{P}(\tilde{F}_x, \tilde{G}_x)}, \quad (13)$$

with $\mathcal{P}(\varnothing)$ the prior probability of not having an astrophysical signal. Given that the two observations are independent (i.e., $\mathcal{P}(\tilde{F}_x, \tilde{G}_x|\varnothing) = \mathcal{P}(\tilde{F}_x|\varnothing)\mathcal{P}(\tilde{G}_x|\varnothing)$) and assuming Gaussian distributions, the likelihood given the null hypothesis is defined as

$$\mathcal{P}(\tilde{F}_x = \tilde{f}_x, \tilde{G}_x = \tilde{g}_x|\varnothing)$$
$$= \frac{1}{2\pi\sigma_{fx}\sigma_{gx}} \exp\left\{-\frac{1}{2}\frac{\tilde{f}_x^2}{\sigma_{fx}^2} - \frac{1}{2}\frac{\tilde{g}_x^2}{\sigma_{gx}^2}\right\}. \quad (14)$$

We define m_1 and m_2 as the magnitude in each spectral band and $m_{1/2} = m_1 - m_2$ as the color. If $\mathcal{H}$ represents a given

hypothesis of the astrophysical nature of a point source, then

$$\mathcal{P}(\mathcal{H}|\tilde{F}_x, \tilde{G}_x) = \int_{m_1, m_{1/2}} \mathcal{P}(\mathcal{H}, m_1, m_{1/2}|\tilde{F}_x, \tilde{G}_x)dm_1 dm_{1/2},$$
$$= \int_{m_1, m_{1/2}} \mathcal{P}(\tilde{F}_x, \tilde{G}_x|\mathcal{H}, m_1, m_{1/2})$$
$$\times \mathcal{P}(m_1, m_{1/2}|\mathcal{H})dm_1 dm_{1/2}\frac{\mathcal{P}(\mathcal{H})}{\mathcal{P}(\tilde{F}_x, \tilde{G}_x)}. \quad (15)$$

From right to left, $\mathcal{P}(\mathcal{H})$ is the prior probability of the hypothesis, $\mathcal{P}(m_1, m_{1/2}|\star)$ is the prior probability of the color–magnitude of an object defined by $\mathcal{H}$, $\mathcal{P}(\tilde{F}_x, \tilde{G}_x|\star, m_1, m_{1/2})$ is the likelihood, and to finish, $\mathcal{P}(\tilde{F}_x, \tilde{G}_x)$ is the marginal likelihood. $\mathcal{P}(\mathcal{H})$ is therefore defined as the frequency of objects $\mathcal{H}$ in a small arbitrary box at the position x on the detector and $\mathcal{P}(m_1, m_{1/2}|\star)$ is the color–magnitude distribution of these objects. The likelihood is given by

$$\mathcal{P}(\tilde{F}_x = \tilde{f}_x, \tilde{G}_x = \tilde{g}_x|\mathcal{H}, m_1, m_{1/2})$$
$$= \frac{1}{2\pi\sigma_{fx}\sigma_{gx}} \exp\left\{-\frac{1}{2}\frac{(\tilde{f}_x - f)^2}{\sigma_{fx}^2} - \frac{1}{2}\frac{(\tilde{g}_x - g)^2}{\sigma_{gx}^2}\right\}, \quad (16)$$

with $f = 10^{-m_1/2.5}$ and $g = 10^{-m_2/2.5}$. σ_{fx} and σ_{gx} are the error bars respectively for the flux estimates $\tilde{f}_x$ and $\tilde{g}_x$.

To carry on our hypothetical example, we assume that a first observation was made with the 4.4 μm filter (F444W) of the NIRCAM instrument on-board the *James Webb Space Telescope* (*JWST*). The follow-up observation is done in H band with the Gemini Planet Imager (GPI). The goal is to identify the most likely nature of a candidate given the *JWST* detection and the GPI data. For the sake of simplicity, we will only consider two stellar populations (high mass noted $\mathcal{H}_{\star,\text{high}}$, low mass noted $\mathcal{H}_{\star,\text{low}}$) and the null hypothesis to illustrate the classification method. However, we would like to emphasize that this framework is in no way restricted to this example and should in practice at least include models of planets or brown dwarfs.

The prior probability distributions of finding a background star as a function of their position in a color–magnitude diagram, $\mathcal{P}(m_1, m_{1/2}|\mathcal{H}_{\star,\text{high}})$ and $\mathcal{P}(m_1, m_{1/2}|\mathcal{H}_{\star,\text{low}})$, can be calculated from the Besançon model of stellar populations (Robin et al. 2003). We generated a galactic population model[11] within a solid angle of 0.23 deg^2 in the vicinity of p Puppis (HD 60863, $l = 242°.96$, $b = -3°.87$). The area was chosen to generate approximately 2×10^5 stars within 125 kpc and with an apparent magnitude of $K < 28$. White dwarfs were filtered out of the catalog based on their surface gravity. The magnitudes for each star within the simulation were calculated using a grid of stellar atmospheres (Castelli & Kurucz 2004) and the properties of each star reported in the catalog; the distance, effective temperature, surface gravity, metallicity, radius, and extinction (using $R_V = 3.1$, $A_H/A_V = 0.184$, and $A_{4.4}/A_V = 0.0$). Figure 3 shows the color–magnitude diagram of the catalog in the *JWST*/F444W and GPI/H filters. We identify two families of stars, low and high mass (198,027 and 32,292 stars), with a boundary at 0.65 $M_\odot$, which corresponds to the soft boundary between the two modes of the two-dimensional histogram. A low-mass star explanation will be

---

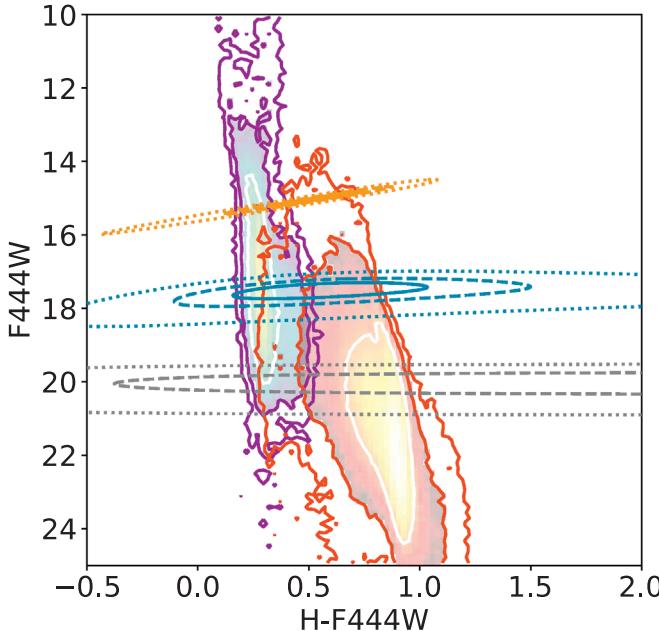[11] Using http://model.obs-besancon.fr.

Figure 3. Color and apparent magnitude diagram of stars compared to observations. The colormaps correspond to the density of low-mass (red) and high-mass stars (>0.65 $M_\odot$, purple), which is only drawn up to a 95% confidence levels. The inner (outer) contour represents the 68% (99.7%) confidence level contours. The elongated contours represent the 68% (solid), 95% (dashed) and 99.7% (dotted) confidence levels of the likelihood for the three scenarios described in Table 2. The probabilities of each candidate to be, respectively, a low-mass or high-mass star is shown in Table 3.

preferred when the ratio $\mathcal{P}(\mathcal{H}_{\star,\mathrm{low}}|\tilde{F}_x, \tilde{G}_x)/\mathcal{P}(\mathcal{H}_{\star,\mathrm{high}}|\tilde{F}_x, \tilde{G}_x)$ is large. The priors $\mathcal{P}(m_1, m_{1/2}|\mathcal{H}_{\star,\mathrm{high}})$ and $\mathcal{P}(m_1, m_{1/2}|\mathcal{H}_{\star,\mathrm{low}})$ are, respectively, the normalized two-dimensional histograms in color–magnitude resulting from the stellar population model.

The prior probabilities $\mathcal{P}(\mathcal{H}_{\star,\mathrm{high}})$ and $\mathcal{P}(\mathcal{H}_{\star,\mathrm{low}})$ are defined as the frequency of such stars in a patch of the sky corresponding to a GPI resolution element (a $\approx$50 mas diameter circle). The probability of the null hypothesis is defined such that $\mathcal{P}(\varnothing) + \mathcal{P}(\mathcal{H}_{\star,\mathrm{high}}) + \mathcal{P}(\mathcal{H}_{\star,\mathrm{low}}) = 1$. The size of the patch of sky will not impact the ratios of probabilities of low-mass and high-mass stars but it will influence the relative probability of the null hypothesis, which is why the probability of the null hypothesis should not be confused with a false positive rate.

The likelihood is a function of the fluxes and associated error bars of the candidate in the different spectral bands. Assuming that some sources were detected in *JWST* at 5σ and subsequently observed by GPI, we describe three possible scenarios corresponding to three candidates. In these scenarios, we vary the sensitivity of the *JWST* observation and the S/N of the GPI follow up, assuming a clear detection (20σ) and two non-detections (3σ, −1σ). The contours of the likelihood for each of the candidates are drawn in Figure 3, which is therefore an illustration of the integrand of Equation (15), $\mathcal{P}(\tilde{F}_x, \tilde{G}_x|\mathcal{H}_\star, m_1, m_{1/2})\mathcal{P}(m_1, m_{1/2}|\mathcal{H}_\star)$. The details of the parameters are described in Table 2 and the results are presented in Table 3.

We can conclude that the first candidate (orange) is most likely to be a high-mass star. The probability of a low-mass star is still high because low-mass star are five times more common that high-mass stars in this catalog. The nature of the second candidate is undecided because the probabilities are too similar. However it is still unlikely to be a false positive. The follow-up

Table 2
5σ Sensitivity in Apparent Magnitude, S/N and Apparent Magnitude of the Point Source for three Hypothetical Candidates Detected with *JWST*/NIRCAM with the F444W 4.4 μm Filter and Followed-up with GPI in H Band

Candidate	F444W (NIRCAM)			H (GPI)		
	5σ	S/N	Mag	5σ	S/N	Mag
1 (Orange)	15	5	15	17.5	30	15.6
2 (Blue)	17.5	5	17.5	17.5	3	18.1
3 (Gray)	20	5	20	17.5	−1	N/A

Note. These candidates are compared to a population of stars in Figure 3.

Table 3
Model Probabilities for the Three Scenarios Described in Table 2 and Illustrated in Figure 3

Candidate	$\mathcal{H}_{\star,\mathrm{high}}$	$\mathcal{H}_{\star,\mathrm{low}}$	$\varnothing$	$\mathcal{H}_{\star,\mathrm{low}}/\mathcal{H}_{\star,\mathrm{high}}$
$\mathcal{P}(\mathcal{H})$ prior[a]	10^{-4}	5×10^{-4}	0.9994	5
1 (orange)	0.81	0.19	0.0	0.20
2 (Blue)	0.50	0.50	0.002	0.84
3 (Gray)	0.04	0.88	0.08	16.2

Note.
[a] The priors $\mathcal{P}(\mathcal{H}_{\star,\mathrm{high}})$ and $\mathcal{P}(\mathcal{H}_{\star,\mathrm{low}})$ are defined from the frequency of the corresponding kind of stars from the Besançon stellar population model (Robin et al. 2003).

epoch of the third candidate would be classified as a non-detection and yielded a negative flux, which means that it is entirely dominated by the noise. However, stars are still located in the 2σ region of the likelihood. The first epoch 5σ detection still makes it somewhat unlikely for the candidate to be pure noise and the low-mass star hypothesis is preferred.

3. Companion on a Known Orbit

3.1. Definition

In this section, we will assume that we know the orbit of a planet projected onto the sky plane but that we do not know its position along it. For example, this situation can arise when trying to constrain the mass of a planet in the gap of a proto-planetary disk (Ruane et al. 2017), where the geometry of the gap defines its orbit (Dong & Fung 2017). Future astrometric discoveries of planets could also provide orbits of unseen planets (Perryman et al. 2014).

We define the data D, or observation, as the random vector representing the pixel values of the image. The point-source parameters are its position on the projected orbit defined as the curvilinear abscissa S and its flux F. We also define N as a Gaussian random vector with zero mean and covariance matrix Σ. In practice, the noise is assumed to be independent, in which case Σ becomes diagonal. Data, signal and noise are related through,

$$D = Fm + N, \qquad (17)$$

with $m = m(s)$ the planet model in the direct imaging data, which is effectively a function of the planet position S. We assume that m (i.e., the shape of the PSF) is independent of the flux F. When the planet is located outside the field of view or inside the focal plane mask, the planet model m is simply null.

The flux posterior given the data is

$$\mathcal{P}(F|D) = \int_S \mathcal{P}(F, S|D)\, dS,$$

$$= \int_S \frac{\mathcal{P}(D|F, S)\mathcal{P}(F, S)}{\mathcal{P}(D)}\, dS,$$

$$= \int_S \frac{\mathcal{P}(D|F, S)\mathcal{P}(F)\mathcal{P}(S)}{\mathcal{P}(D)}\, dS, \qquad (18)$$

where $\mathcal{P}(F)$ and $\mathcal{P}(S)$ are the priors for the point-source flux and position.

The likelihood is defined as,

$$\mathcal{P}(D = \boldsymbol{d}|F = f, S = s)$$

$$= \frac{1}{\sqrt{2\pi|\boldsymbol{\Sigma}|}} \exp\left\{ -\frac{1}{2}(\boldsymbol{d} - f\boldsymbol{m})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{d} - f\boldsymbol{m}) \right\}, \qquad (19)$$

which is the matrix version of a Gaussian likelihood seen in Equation (2).

For simplicity, we assume a uniform positive prior over flux,

$$\mathcal{P}(F = f) = \begin{cases} \alpha, & \text{if } f > 0, \text{ with } \alpha \text{ constant} \\ 0, & \text{if } f \leqslant 0 \end{cases}. \qquad (20)$$

The probability of finding the companion at any position in the orbit is proportional to the time spent around that position according to Kepler's laws. We can write:

$$\mathcal{P}(S) = \frac{1}{T v_{\text{proj}}} \qquad (21)$$

where T is the orbital period and $v_{\text{proj}} = v_{\text{proj}}(s)$ is the projected velocity of a companion at position s on the ellipse representing the projected orbit. v_{proj} is constant in a face-on and circular orbit. If $\mathcal{P}(S)$ is a delta function, we get the example from Section 2.1.

We have not yet specified the data term d for the likelihood in Equation (19). Although it is possible to choose the final combined image, the planet model might be poorly known when over- and self-subtraction from the speckle subtraction applies. Instead, we directly define the likelihood from the individual speckle-subtracted images, which are for example defined by their exposure number and wavelength for an integral field spectrograph. The speckles are here subtracted using a principal component analysis based algorithm called Karhunen–Loéve Image Projection (KLIP, Soummer et al. 2012). KLIP consists in subtracting to each image its own projection on a subset containing K elements of the principal components z_k. Defining the matrix $\boldsymbol{Z}_K = [z_1, z_2, ..., z_K]^\top$, the speckle subtraction takes the form

$$\boldsymbol{i}_{\text{sub}} = \boldsymbol{i} - \boldsymbol{Z}_K^\top \boldsymbol{Z}_K \boldsymbol{i}, \qquad (22)$$

with $\boldsymbol{i}$ the science image and $\boldsymbol{i}_{\text{sub}}$ the speckle-subtracted image. Generally, the model of the signal is a function of the speckle-subtraction algorithm used. When using a KLIP framework, point sources are distorted by the speckle subtraction. Pueyo (2016) derived a linearized approximation of this distorted PSF, which will be referred to as the forward model of the signal. Indeed, the existence of a faint point source in the data induces a perturbation on the principal components denoted $\Delta \boldsymbol{Z}_K$. We refer the reader to Pueyo (2016) for the analytical expression of $\Delta \boldsymbol{Z}_K$, which is outside the scope of this paper. If $\boldsymbol{a}$ is the

vectorized normalized planet signal in the science image, f the planet flux and n_i the associated noise containing the speckles, such that $\boldsymbol{i} = f\boldsymbol{a} + \boldsymbol{n}_i$, then the normalized forward model then can be written as

$$\boldsymbol{m} = \boldsymbol{a} - \boldsymbol{Z}_K^\top \boldsymbol{Z}_K \boldsymbol{a} - (\boldsymbol{Z}_K^\top \Delta \boldsymbol{Z}_K + (\boldsymbol{Z}_K^\top \Delta \boldsymbol{Z}_K)^\top)\frac{\boldsymbol{i}}{f}. \qquad (23)$$

The linear approximation is valid when the planet signal is faint relative to the speckles (i.e., f is small) and when there is little spatial overlap between the planet signal in the different reference images used in the principal components calculation. The diagonal terms of the covariance matrix $\boldsymbol{\Sigma}$ are directly estimated from the data as the empirical variance of each pixel. With the assumption of Gaussian noise, the likelihood can be written directly as a function of the estimated flux $\tilde{f}_x$ and associated error bar σ_x as defined in Ruffio et al. (2017) or Cantalloube et al. (2015), with:

$$\tilde{f}_x = \frac{\boldsymbol{d}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}{\boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}, \qquad (24)$$

and

$$\sigma_x^2 = \frac{1}{\boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}. \qquad (25)$$

The estimated flux $\tilde{f}_x$ is defined as the value maximizing the likelihood from Equation (19). The terms $\boldsymbol{\Sigma}$ and σ_x should not be confused—the former characterizes the noise in the uncombined data and the latter represents the noise in the estimated flux map. We can also write the theoretical matched filter S/N as

$$\mathcal{S}_x = \frac{\boldsymbol{d}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}{\sqrt{\boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}}}. \qquad (26)$$

These quantities are the final products of matched-filter based data reduction (Cantalloube et al. 2015; Ruffio et al. 2017), whose goal is to find the location of a known signal in noisy data. This is a maximum likelihood approach and it consists in maximizing the S/N from Equation (26) as a function of the position of the planet. By substituting Equations (19)–(21), (24), and (25) in Equation (18), we get the posterior

$$\mathcal{P}(F = f|D = \boldsymbol{d}) \propto \mathcal{H}(f)$$

$$\times \int_{s=s_i}^{s_f} \exp\left\{ -\frac{1}{2}(f^2 \boldsymbol{m}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m} - 2f\boldsymbol{d}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{m}) \right\} \frac{1}{T v_{\text{proj}}}\, ds,$$

$$\propto \mathcal{H}(f) \int_{s=s_i}^{s_f} \exp\left\{ -\frac{1}{2\sigma_s^2}(f^2 - 2f\tilde{f}_s) \right\} \frac{1}{T v_{\text{proj}}}\, ds. \qquad (27)$$

We have used the fact that $\boldsymbol{d}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{d}$ does not depend on the position s, so we can factor it out of the integral as a proportionality constant.

In practice, the assumptions that were just made result in a biased estimate of the flux and the standard deviation. For example, the covariance $\boldsymbol{\Sigma}$ is not truly diagonal. The forward model is also not a perfect model of the planet, which underestimates the flux. Consequently, we define the algorithm throughput as the ratio between the measured flux and the true flux of a point source and we denote it as μ_x. A common
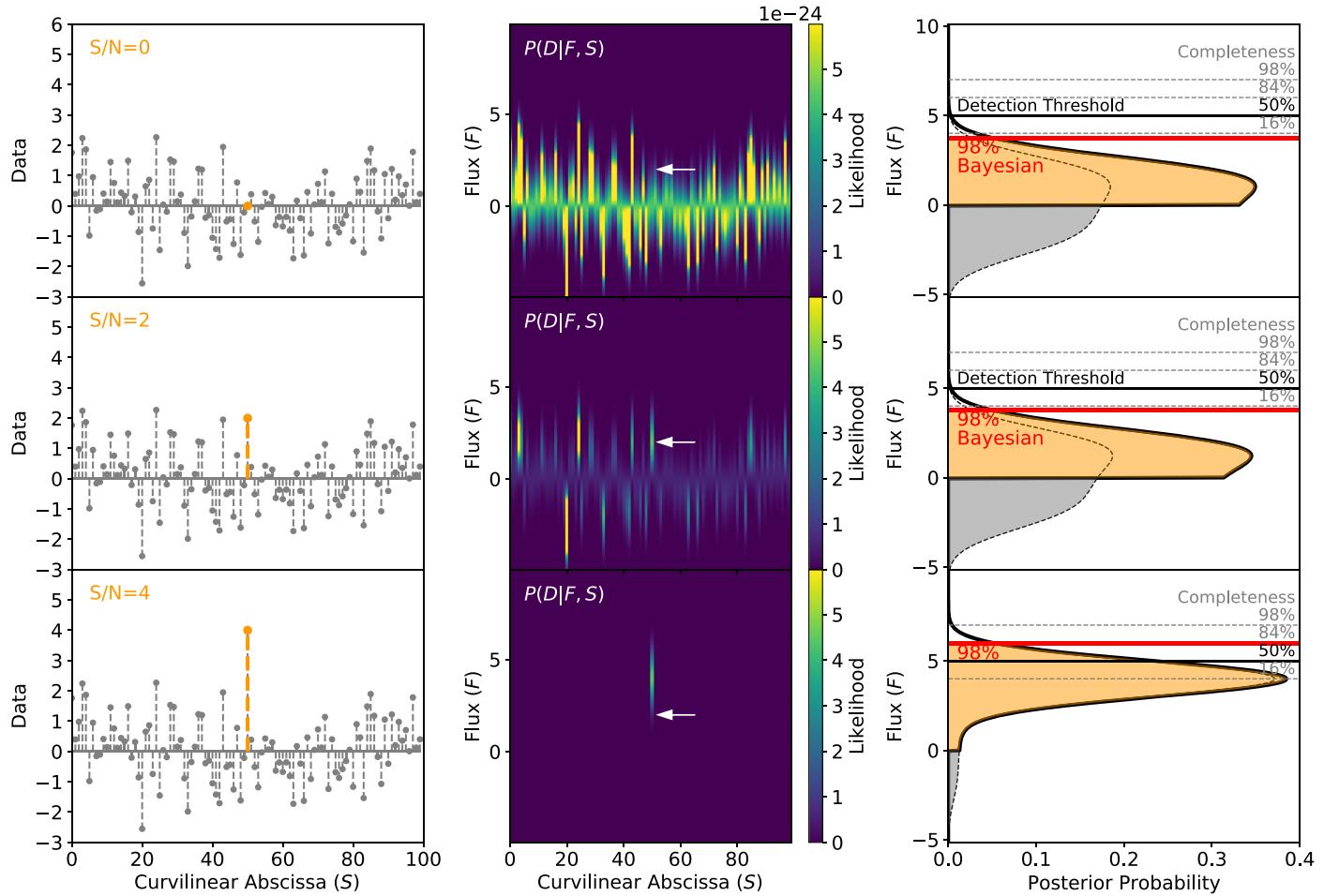
Figure 4. Toy simulation illustrating Bayesian upper limits for planets on a known orbit but unknown location. The left-hand column features the observed flux as a function of the curvilinear abscissa of the orbit path. Each row corresponds to a given S/N of the pixel with curvilinear abscissa 50 highlighted by the vertical orange dashed line. The central column shows the joint likelihood as a function of flux and position of the planet in arbitrary units. The likelihood globally decreases as the signal increases because it becomes harder to explain with pure noise. The white arrow does not move in each plot and it highlights the location where the likelihood changes. The right-hand column plots the flux posterior (black solid curve) of the planet compared to the completeness contours (gray horizontal lines) of a 5σ detection threshold (black horizontal lines). The dashed curve shows the posterior before the positive prior rejects the negative values of the flux (gray area). The Bayesian upper limit with 98% cutoff probability is drawn as a horizontal solid red line. The posterior is marginalized over S, which corresponds to a horizontal integral of the joint likelihood in the middle panels.

practice to mitigate the standard deviation bias (Cantalloube et al. 2015; Ruffio et al. 2017) is to re-normalize the standard deviation to yield a S/N map with unit standard deviation ($\mathcal{S}_x/\eta_x \longrightarrow \mathcal{S}_x$, with the η_x the standard deviation of the $\mathcal{S}_x$ map). The flux calibration is done with simulated planet injection and recovery ($\tilde{f}_x/\mu_x \longrightarrow \tilde{f}_x$), which is also known as algorithm throughput correction. Equation (27) now becomes:

$$\mathcal{P}(F=f|D=\boldsymbol{d}) \propto \mathcal{H}(f)$$
$$\times \int_{s=s_i}^{s_f} \exp\left\{-\frac{1}{2\eta_s^2\sigma_s^2}((\mu_s f)^2 - 2(\mu_x f)\tilde{f}_s)\right\}\frac{1}{Tv_{\mathrm{proj}}}ds.$$
(28)

Figure 4 features a toy simulation of the approach assuming 100 independent samples representing the pixel values along the orbit path, a unit standard deviation and a Dirac-like planet model (all of the flux contained in one pixel). The value of the middle-data point was fixed to a given S/N to show the effect of outliers on the upper limit. Similarly to Figure 2, Figure 4 shows that the upper limit should truly be a function of the data

—in other words, a function of the realization of the noise, which is not true of the detection threshold.

Note that this method is even valid when there is a strong (or weak) signal in the gap, which could be either a rare occurrence of the noise or an astrophysical object. This would not be true for a detection threshold based approach because the latter does not depend on the actual measurement.

3.2. Effect of the Size of the Orbit

Intuitively, the upper limit is to the first order defined by the brightest signal in the data. As the number of elements increases, it becomes more likely to draw high S/N signals. Therefore, the larger the uncertainty on the location of the planet, the more realizations of the noise have to be considered and the poorer the upper limit will be. Figure 5 shows examples of posteriors when varying the number of samples in the Figure 4 simulation and assuming pure noise data with no real signal. Unless the data contains an outlier—defined as an unlikely event considering the number of realizations—the posterior flattens out as the number of samples increases. At the limit of an infinite number of elements, the data loses any
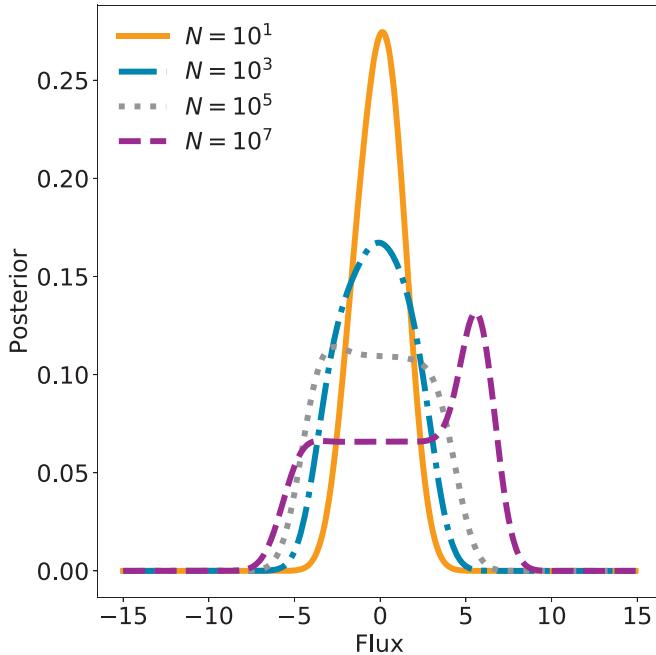
Figure 5. Flux posterior as a function of the number of elements using a simple simulation. The $N = 10^7$ samples case highlights the effect of an outlier in the data.

Figure 6. Histogram of the speckle-subtracted uncombined images for three separations for the NIRC2 observation of TW Hya. A resolution element is defined as $1.22\lambda/D \approx 0\rlap{.}''1$ ($\lambda = 4\ \mu$m and $D = 10$ m).

constraining power on the flux of the planet and the posterior becomes a constant. When defining the upper limit from a detection threshold, this effect can be partially accounted for by modifying the detection threshold to yield the same false positive rate at any distance of the star (Ruane et al. 2017; Jensen-Clem et al. 2018).

3.3. Qualitative Effect of Non-Gaussianity and Correlation

We have assumed that the noise was Gaussian and uncorrelated. Figure 6 shows the histogram of the pixel values in uncombined speckle-subtracted images for three different separations. While the distribution approximates a Gaussian for the largest two separations, it has a very large tail in the innermost case. In this section, we simulate the effects of the correlation and the non-Gaussianity on the final upper limit in a simplified example. We assume a discrete orbit made of hundreds of elements similar to Figure 4, containing pure noise. To explore the effect of the size of the planet with respect to non-Gaussian noise and non-diagonal correlation matrix, we consider two models: a small Dirac-like PSF and a large PSF that is five-elements wide. In Figure 7, we estimate the error made on the upper limit when we erroneously assume that the noise is Gaussian or independent. To do so, we compare the upper limits derived from the true properties of the noise with the upper limits derived with the assumption of Gaussian and independent noise. The non-Gaussianity is first tested using a Student-t distribution with 10 degrees of freedom and normalized to unit standard deviation (top panels of Figure 7), which was a good fit to the pixel distribution of the inner-most annulus in Figure 6. In this context, the Student-t distribution is used because of its wider tails compared to a
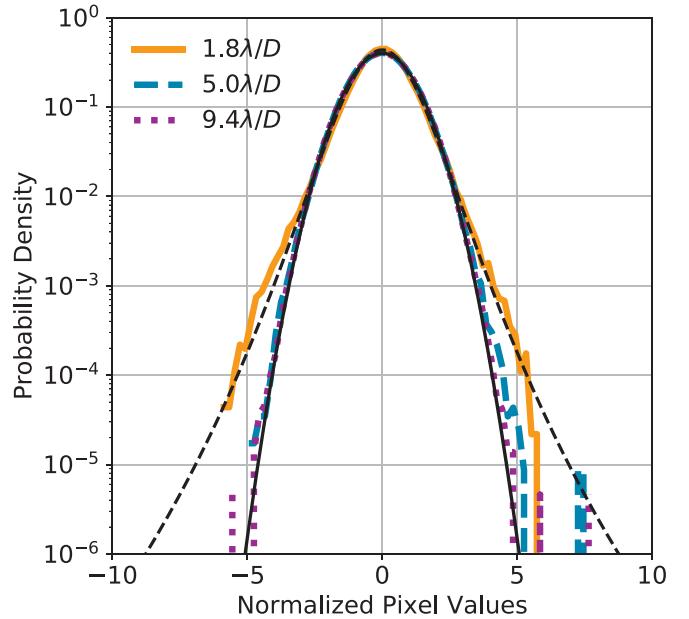
Gaussian distribution but not because of its relation to small sample statistics. We then try a correlated Gaussian noise with a circulant covariance matrix[12] (bottom panels of Figure 7). The correlation profile is defined as a Gaussian with a standard deviation equal to two elements. The upper limits are calculated in a similar fashion as for Figure 4 by replacing the likelihood from Equation (18) to include a non-diagonal covariance matrix or to use a Student-t distribution. The Student-t joint likelihood is computed as the product of the individual likelihood for each pixel, which is granted by the independence of the noise. Despite the simplicity of these simulations, we can draw some general principles from them. First, it is necessary to account for the non-Gaussian tail of the noise only when considering high cutoff probabilities (e.g., >0.999, which is equivalent to 3σ). It is therefore good practice to quote upper-limits derived from reasonable cutoff probabilities if the distribution of the noise is poorly known. In this crude simulation, the non-Gaussian noise is uncorrelated, which means that the effect is mitigated when the planet PSF is large. Indeed, the noise becomes more Gaussian when combining several pixels together because of the central limit theorem. In practice, the non-Gaussian noise comes from the correlated speckle noise with a correlation length equal to the PSF size, which means that a larger PSF will not help. However, the noise will be made more Gaussian thanks to the observing strategies—Angular Differential Imaging (Marois et al. 2006) and Spectral Differential Imaging (Marois et al. 2000; Sparks & Ford 2002)—where the quasi-static speckle are subtracted and the displacement of the planet relative to them is used.

Second, when neglected, correlated noise can create the illusion of a signal, which results in inflated upper limits. However, overestimated upper limits are a conservative choice, which makes it acceptable. This effect is partially corrected

---

[12] A circulant matrix is a matrix for which each row vector is shifted by one element to the right relative to the preceding row vector. It is a special case of Toeplitz matrix for which the diagonals are constant. The circulant matrix is used here to express the periodicity of the projected orbit.
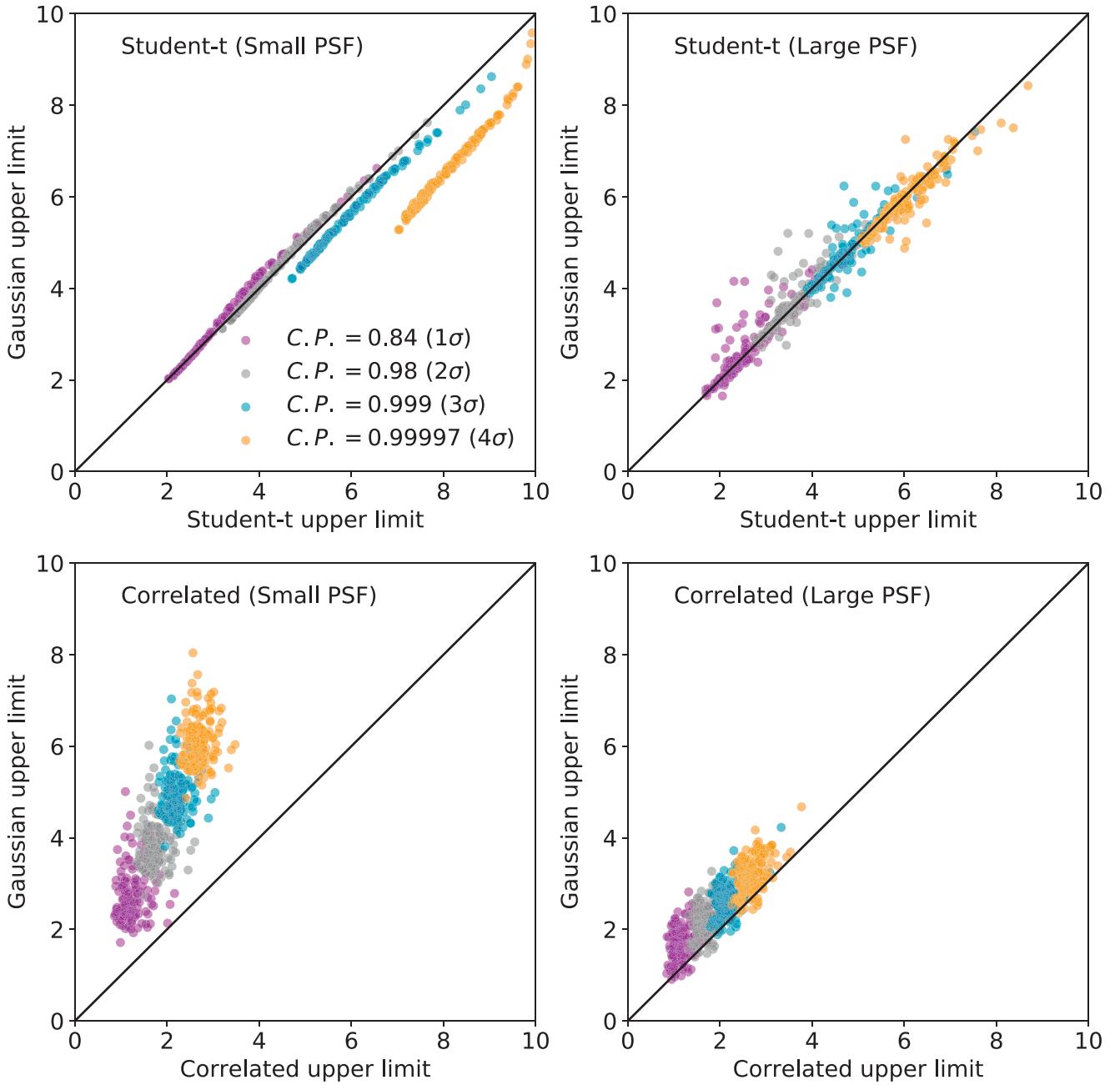
Figure 7. Comparison of the upper limits derived from the true properties (x-axis) of the noise with the upper limits derived with the assumption of Gaussian and independent noise (y-axis). The color corresponds to different cutoff probability (CP). The diagonal represents a correct estimation of the upper limit despite the approximation that is made. Points below the diagonal show that the upper limits has been underestimated leading to over-confident constraints. Points above the diagonal represent over-estimated upper limits leading to conservative results. Units are arbitrary.

when we calibrate the standard deviation as discussed at the end of Section 3.1. Note that a large PSF here again mitigates the effects of the correlation. Increasing the correlation length and the PSF size in tandem is equivalent to reducing the number of independent realizations of the noise.

3.4. Constraining the Mass of a Planet in the TW Hya Protoplanetary Disk

We apply the previous framework to the Keck-NIRC2 observations of TW Hya at L′ presented in Ruane et al. (2017). TW Hya features a proto-planetary disk in which gaps have been detected (Weinberger et al. 2002; Akiyama et al. 2015; Rapson et al. 2015; Andrews et al. 2016; Debes et al. 2016; van

Boekel et al. 2017). A possible explanation for the gaps could be the presence of accreting planets carving them (Dong & Fung 2017). The goal is to use the high-contrast observation of the system to set an upper limit on the mass of the hypothetical planets, which can be compared to the predictions of theoretical models of planet formation.

We will compare the different definitions of upper limit. As a reminder, our Bayesian upper limits are defined from the flux posterior and the probability of the true planet flux (or mass) to be smaller than the limit, which we called cutoff probability. The frequentist definition of upper limit relies on the detection threshold and its associated completeness. A detection occurs when the measured flux of a point source falls above the
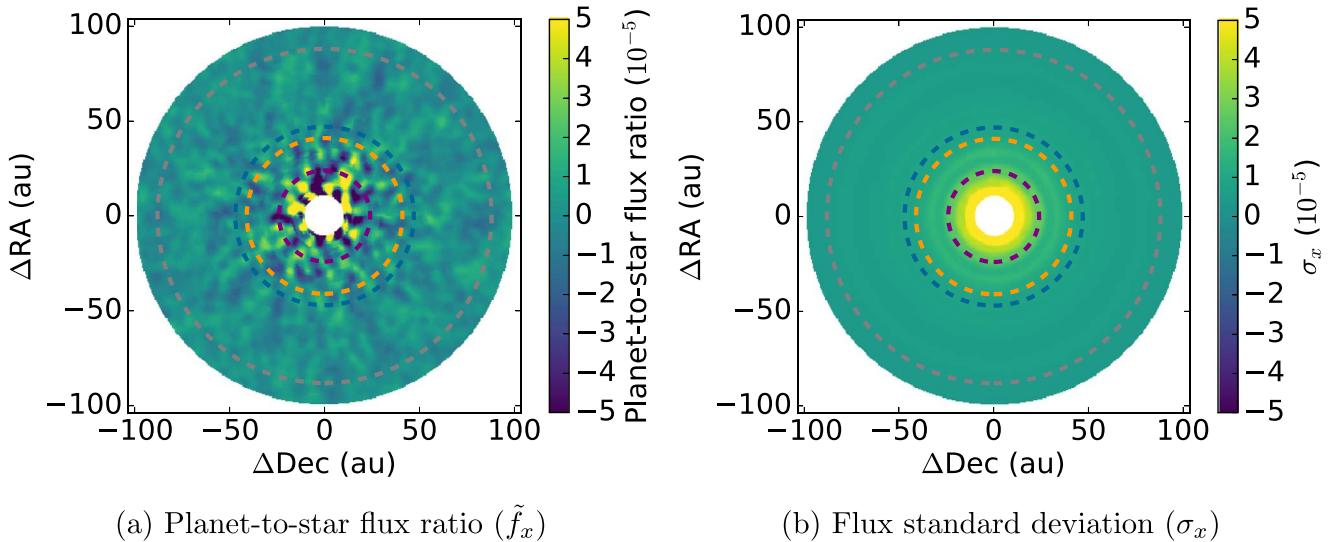
(a) Planet-to-star flux ratio ($\tilde{f}_x$)

(b) Flux standard deviation (σ_x)

Figure 8. (a) Planet-to-star flux ratio of TW Hya and (b) its corresponding standard deviation maps. The maps were calculated according to Equations (24) and (25) using a forward model matched filter described in Ruffio et al. (2017). The dashed lines represent the known gaps in the protoplanetary disk at 24, 41, 47, and 88 au.

detection threshold. The completeness is the probability that a planet of a given flux (or mass) is detected. The completeness is therefore always 50% for planets with a true flux equal to the detection threshold. Otherwise, it will, for example, be equal to 16% if the true flux of the planet is 1σ below the detection threshold and 84% if it is 1σ above the detection threshold. To summarize, the frequentist upper limit makes a statement about the detectability of a planet, while the Bayesian upper limit makes a statement about the probability of the planet flux given the data.
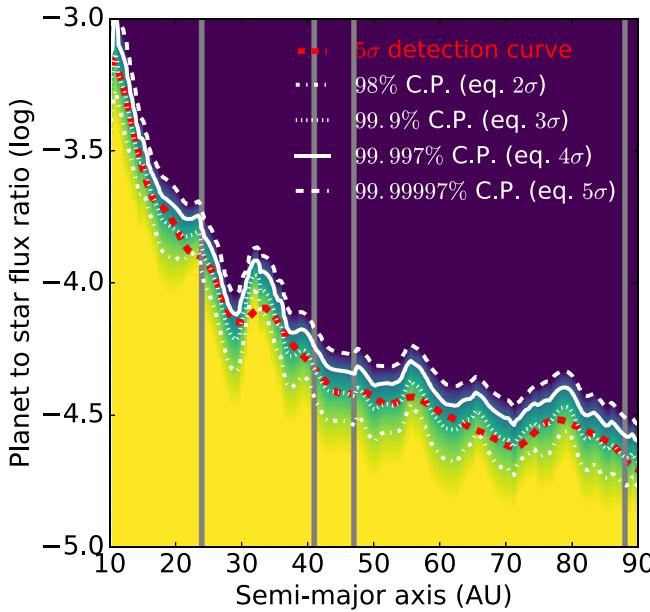
The units of the planet flux have not yet been defined and will depend on the normalization of the planet model. In the following, we will express the flux as the planet-to-star flux ratio. Figure 8 shows the planet-to-star flux ratio of TW Hya and its corresponding standard deviation maps, which are used in the calculation of the likelihood in Equation (27). Figure 9 shows the resulting planet-to-star flux ratio and the mass upper limits for different cutoff probabilities as a function of the planet semimajor axis (white lines). It also features the 5σ detection threshold as a function of separation (dashed red line). The location of the four gaps in TW Hya proto-planetary disk are marked with gray lines. The planet to star flux ratio to mass conversion was performed using the AMES-Cond model (Baraffe et al. 2003; Allard et al. 2012) and an age of 10 Myr for the star (Bell et al. 2015), which only accounts for photospheric emissions and neglect possible effects of accretion. We assume a uniform positive prior in flux or mass for the calculation of the flux or mass posterior, respectively. This means that the priors in both cases, mass or flux, are not equivalent. We could use a more informed mass prior based on observational results but it is still poorly constrained for giant planets at large separation and the choice of a constant is conservative.

Figure 9 does not allow for an easy comparison of the different upper limits. Figure 10 features the posterior and its cumulative distribution, as well as the completeness for each gap, which are in substance vertical cuts through Figure 9. For example, on the one hand, the detection threshold corresponds to a $\approx 1 M_J$ planet at the 88 au gap (gray lines). A possible definition of mass upper limit would be the mass of a planet that would be detected 95% of the time, which is $\approx 1.3 M_J$ for
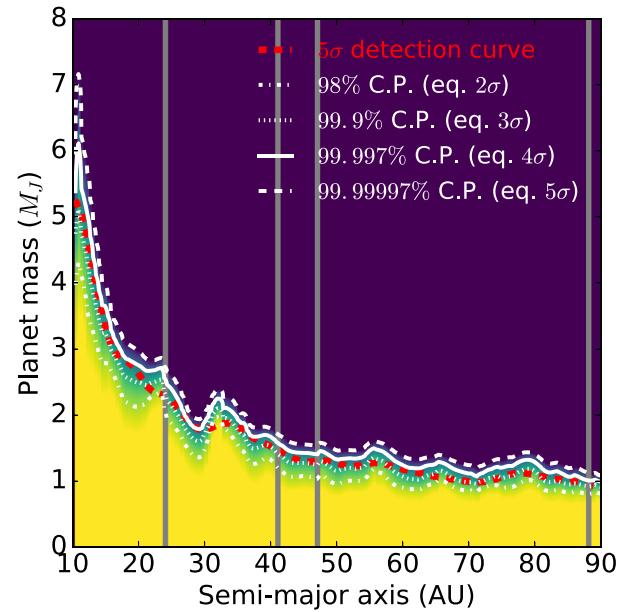
that gap. On the other hand, the value of the cumulative distribution of the posterior at the detection threshold is 0.9995 ($\approx 3\sigma$), which already highlight a comfortable degree of confidence in the fact that the mass of the planet is lower. In hindsight, it seems unnecessary to look for such a high completeness in this case. It is also common practice to quote the detection threshold itself as the upper limit. This approach can lead to confusion because the 5σ of the detection threshold can be easily mistaken for a cutoff probability, when it is really associated to a false positive rate. A 5σ detection threshold is not equivalent to a Bayesian upper limit with a "5σ" cutoff probability, but is truly $\approx 3\sigma$ in the specific example. An upper limit should always be accompanied by a statement about its probability, in which a sole detection threshold is lacking.

We have already mentioned other caveats coming from a detection-based upper limit. First, the definition of the threshold is somewhat arbitrary. The field has widely adopted a 5σ threshold but it yields many false positives in practice due to the non-Gaussianity of the noise. If we chose a larger threshold to mitigate this issue, then should the upper limit change or remain the same? Second, a detection threshold only indirectly depends on the data through the sample standard deviation, while the Bayesian upper limit fully expresses the information contained in the data. The latter will be highly sensitive to outliers. This dependence to the data is illustrated in Figure 9, where the Bayesian upper limit varies strongly as a function of the semimajor axis while the detection threshold is smoother. Finally, and related to the previous point, an advantage of the Bayesian upper limit is that it is valid regardless of the strength of the signal (detection or non-detection), while the detection threshold framework requires the absence of outliers. In addition, the Bayesian approach is less sensitive to the assumed noise distribution than the frequentist approach due to the dominant impact of the strongest signal in the data.

In Ruane et al. (2017), the upper limit is calculated for a 95% completeness and a detection threshold that is defined to yield 0.01 false positives within $1''$ of the host star. The threshold varies from 8.1σ to 4.5σ with increasing separation to the star accounting for both the larger area available at larger separation and small sample statistics using a Student-t distribution (Mawet et al. 2014). The most conservative upper limits for the

(a) Flux ratio

(b) Mass

Figure 9. Cumulative distribution of the planet-to-star flux ratio (respectively mass) posterior for TW Hya as a function of the semimajor axis of the putative planet to the star. The Bayesian upper limit contours are drawn for different probabilities and can be compared to the conventional 5σ detection threshold. The equivalent significance of each cutoff probability is also written in terms of sigmas in the legend. The location of the gaps are marked with a vertical gray solid line. The flux ratio to mass conversion was performed using the AMES-Cond model and a 10 Myr old star.

mass of a companion around TW Hya assume an age of 10 Myr and the AMES-Cond model. The reported upper limits for each gap in the system are: 2.3 M_J at 24 au, 1.6 M_J at 41 au, 1.5 M_J at 47 au, and 1.2 M_J at 88 au. Using a Bayesian analysis and 0.999 cutoff probability (3σ), this work finds mass upper limits equal to 2.4 M_J at 24 au, 1.5 M_J at 41 au, 1.3 M_J at 47 au, and 0.9 M_J at 88 au. Note that this work uses a different reduction algorithm compared to Ruane et al. (2017), which has not been optimized and might explain the limited gains. Figure 10 should be used for a fair comparison of the frequentist and Bayesian approach. The upper limits for the planet-to-star flux ratio are 1.4×10^{-4} at 24 au, 4.9×10^{-5} at 41 au, 3.4×10^{-5} at 47 au, and 2.2×10^{-5} at 88 au. The corresponding absolute magnitudes are, respectively 12.9, 14.0, 14.3, and 14.9, where we have assumed a TW Hya distance of 60.1 ± 0.15 pc (Gaia Collaboration et al. 2016, 2018) and a 7.01 apparent magnitude at Wise W1 band (proxy for Keck-NIRC2 L' band) (Wright et al. 2010).

The previous mass upper limits do not consider possible accretion of the planet, which is most likely to occur for protoplanetary disks like TW Hya. The absolute magnitude of a circumplanetary disk is a function of the product of the planet mass with the accretion rate, $M\dot{M}$, and the inner radius of the circumplanetary disk, R_{in} (Zhu 2015). For small enough planets, which is most often the case, the intrinsic flux of the planet can be neglected because the accretion appears much brighter. We can compute the probability corresponding to each set of parameters $(M\dot{M}, R_{in})$ using Zhu (2015) model predictions for the circumplanetary disk absolute magnitude in L band and the posterior distribution of Figure 10. We compare the map of cutoff probabilities and completeness for each gap in Figure 11. On a logarithmic scale, the 99.9% Bayesian cutoff probability leads to marginally better constraints on $(M\dot{M}, R_{in})$ than the 5σ completeness based approach. However, we argue that it is easier to statistically interpret the Bayesian constraints.

Dong & Fung (2017) argues that hypothetical gap opening planets should have mass of the order of 0.1M_J in the 20 and 80 au gaps. The thermal emission of such planets would be invisible at our current sensitivity and, therefore, can be neglected. Consequently, assuming a 0.1M_J planet, we can set constraints on the accretion rate $\dot{M}$ by adding one unit to the y-logarithmic-scale in Figure 11. In conclusion, given the data, we are 99.9% confident that the accretion rate of a 0.1M_J planet in the TW Hya gaps, if it exists, is below 9.3×10^{-7} M_J yr^{-1} at 24 au, 5.0×10^{-7} M_J yr^{-1} at 41 au, 4.5×10^{-7} M_J yr^{-1} at 47 au, and 3.4×10^{-7} M_J yr^{-1} at 88 au.

4. Combining RV and Direct Imaging Observations

Combining RV data with direct imaging is another very promising avenue for constraining the masses of non-transiting wide-orbit planets detected with Doppler measurements. An example of application of this method can be found in the case of ϵ Eridani in Mawet et al. (2018) for which this section describes the theoretical concepts.

Radial velocity only provides a lower limit on the mass of the planet due to the $M \sin(i)$ mass-inclination degeneracy. Direct imaging non-detection can provide a mass upper bound and therefore reject the lower inclinations. We define D_{RV} as the time series of RVs and D_{DI} as the direct imaging observation. The model parameters Θ to be inferred include the orbital elements and the mass of the planet as well as the star. We also define Θ' such that $\Theta = \{M_p, \Omega, \Theta'\}$ with M_p the mass of the planet and Ω the position angle or longitude of ascending node.
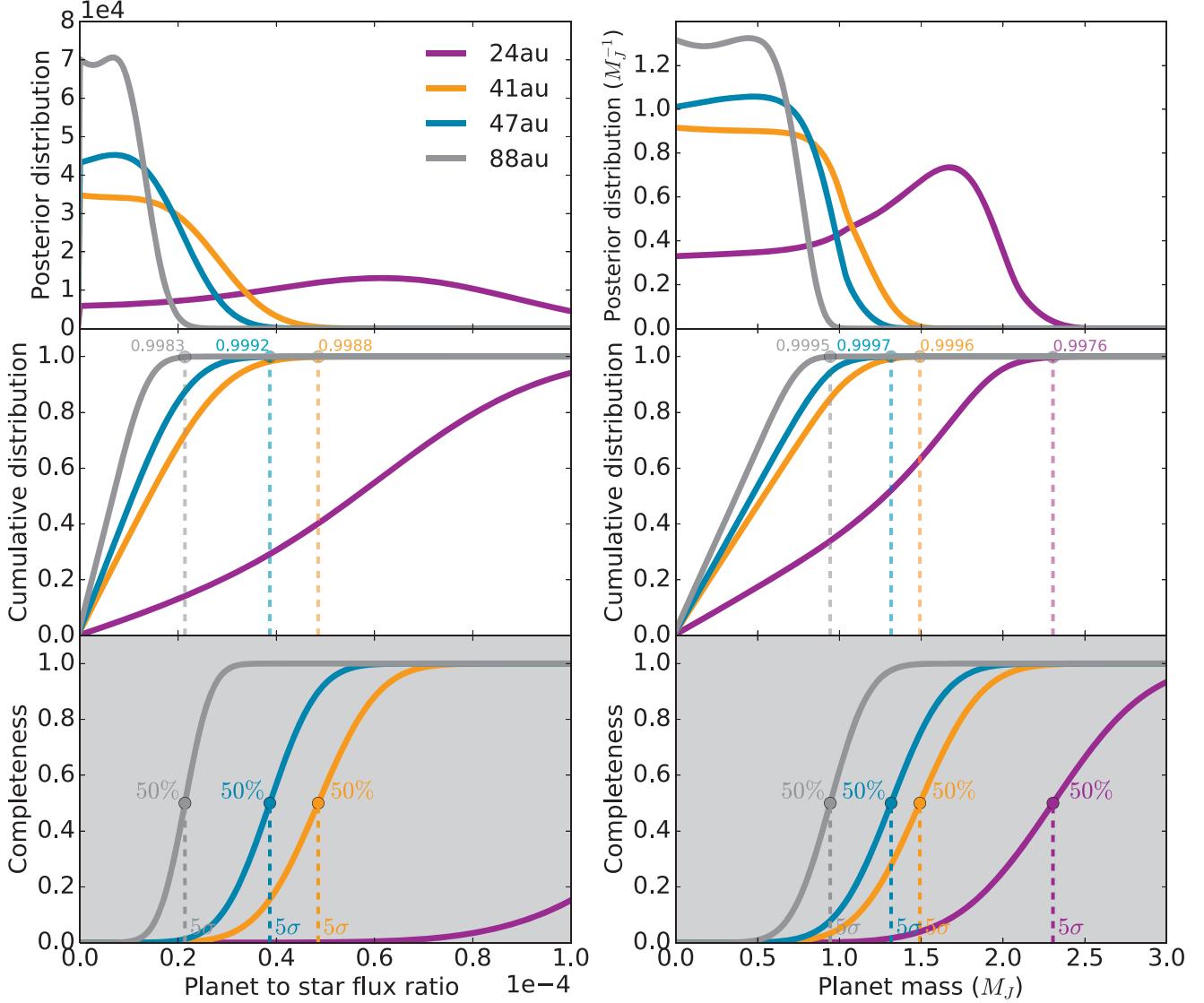
Figure 10. Posterior, cumulative distribution assuming a positive prior and completeness for each gap of Figure 9. The first column features the distributions as a function of the planet-to-star flux ratio while the second column uses the planet mass. The completeness is calculated for a 5σ detection threshold, which itself corresponds to a 50% completeness. The value of the cumulative distribution of the posterior (i.e., the cutoff probability) corresponding to the detection threshold is indicated above the curve for each gap.

D_{RV} and D_{DI} are independent so the posterior can be written as,

$$\mathcal{P}(\Theta|D_{\mathrm{RV}}, D_{\mathrm{DI}}) = \frac{\mathcal{P}(D_{\mathrm{RV}}, D_{\mathrm{DI}}|\Theta)\mathcal{P}(\Theta)}{\mathcal{P}(D_{\mathrm{RV}}, D_{\mathrm{DI}})},$$
$$= \frac{\mathcal{P}(D_{\mathrm{RV}}|\Theta)\mathcal{P}(D_{\mathrm{DI}}|\Theta)\mathcal{P}(\Theta)}{\mathcal{P}(D_{\mathrm{RV}}, D_{\mathrm{DI}})}. \quad (29)$$

The RV log-likelihood can be written as (Howard et al. 2014)

$$\log \mathcal{P}(D_{\mathrm{RV}}|\Theta)$$
$$= -\sum_i \left[\frac{(v_i - v_m(t_i))^2}{2(\sigma_i^2 + \sigma_j^2)} + \log \sqrt{2\pi(\sigma_i^2 + \sigma_j^2)} \right], \quad (30)$$

where v_i are the measured RVs at the times t_i and $v_m(t_i)$ are the corresponding projected Keplerian velocities. The standard deviations σ_i and σ_j are, respectively, the internal uncertainty for each measurement and the instrument-specific jitter term.

As was shown in Section 3.1, the direct imaging log-likelihood can be written as

$$\log \mathcal{P}(D_{\mathrm{DI}}|F = f, X = x) \propto -\frac{1}{2\sigma_x^2}(f^2 - 2f\tilde{f}_x). \quad (31)$$

The planet flux f is a function of the planet mass and stellar age, and the position x is determined by the orbital parameters.

The posterior on Θ can be inferred using a Markov Chain Monte Carlo. Keeping both spatial dimensions in the direct imaging likelihood, such as in Mawet et al. (2018), requires a much longer Markov chain to converge. It is possible to make the problem more tractable by marginalizing the problem over the position angle of the planet with minimal loss of information. Indeed, the RV cannot constrain this parameter and the image has little valuable information about the position
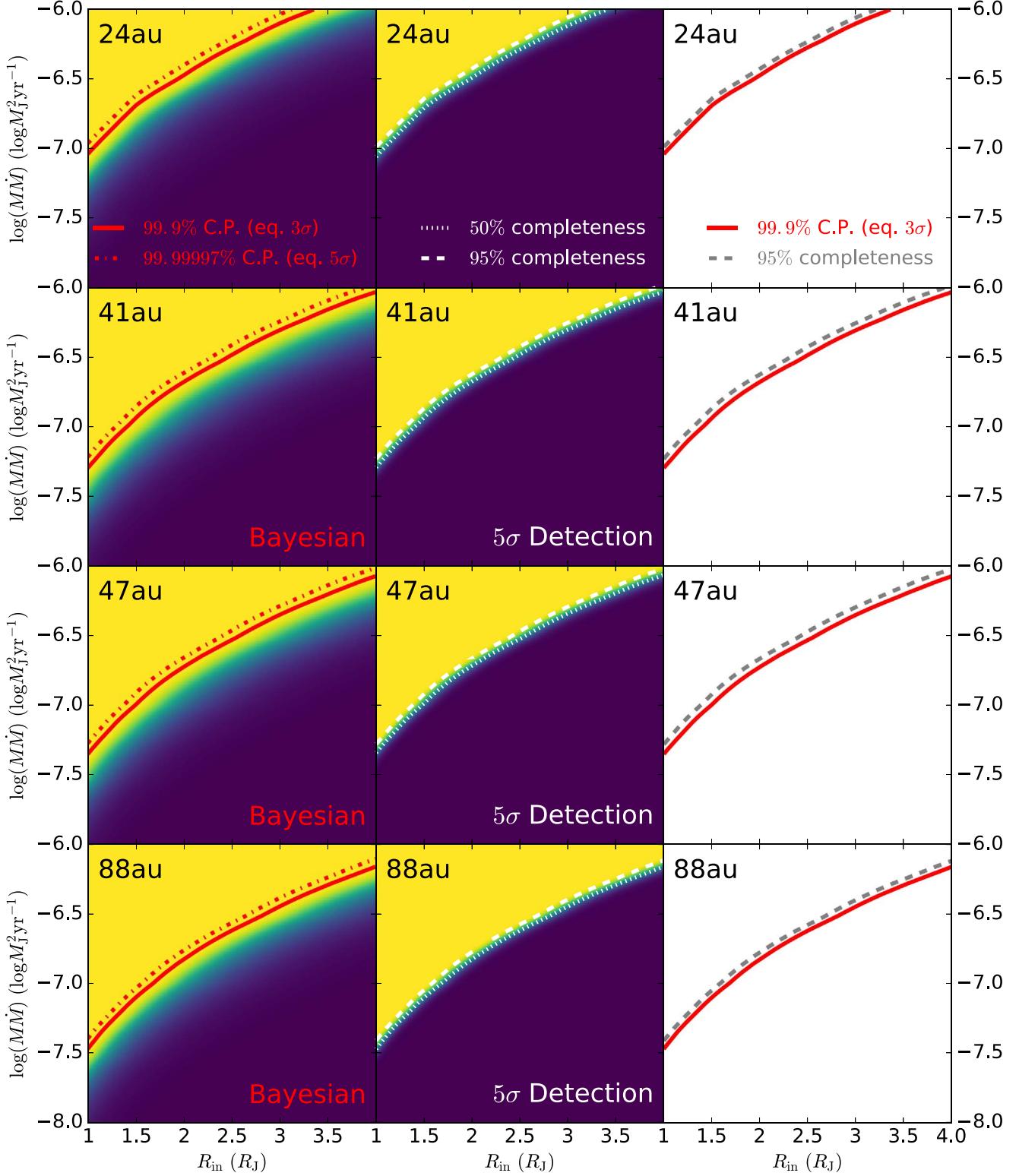
Figure 11. Cutoff probabilities and detection completeness as a function of $(M\dot{M}, R_{in})$ for each gap (rows) in TW Hya protoplanetry disks. The L'-band absolute magnitudes for each value of $(M\dot{M}, R_{in})$ were linearly interpolated from the table in Zhu (2015). The left-hand images show the cutoff probability, which is the value of the cumulative distribution of the posteriors in Figure 10 at the flux of the planet. The middle images show the corresponding completeness for a 5σ detection threshold. The right-hand plots compare the 99.9% cutoff probability with the 95% completeness contours.

angle in the absence of an obvious outlier.

$$
\begin{aligned}
\mathcal{P}(M_p, \Theta'|D_{\mathrm{RV}}, D_{\mathrm{DI}}) &= \int_\omega \mathcal{P}(\Theta|D_{\mathrm{RV}}, D_{\mathrm{DI}}) \\
&= \int_\omega \frac{\mathcal{P}(D_{\mathrm{RV}}|\Theta)\mathcal{P}(D_{\mathrm{DI}}|\Theta)\mathcal{P}(\Theta)}{\mathcal{P}(D_{\mathrm{RV}}, D_{\mathrm{DI}})} \\
&\propto \mathcal{P}(D_{\mathrm{RV}}|M_p, \Theta')\mathcal{P}(\Theta'')\mathcal{P}(M_p) \\
&\quad \times \left[\int_\omega \mathcal{P}(D_{\mathrm{DI}}|F, r, \omega)\mathcal{P}(\omega) \right] \quad (32)
\end{aligned}
$$

with $\mathcal{P}(\Omega) = 1/2\pi$ the prior for the position angle and r, ω are the cylindrical coordinates of the planet.

Equation (32) shows that the RV likelihood can be factored out of the integral, which leaves an integral over the sole direct imaging likelihood. The calculation of the integral is then equivalent to Section 3.1 with a circular face-on orbit. Note that it is always possible to derive an upper limit, even when the existence of a planet is still in doubt. For example, one could use this framework to constrain masses for RV trends.

5. Conclusion

In this work we have addressed the differences between frequentist and Bayesian definitions of planet flux upper limits in the context of exoplanet direct imaging. The frequentist upper limit makes a statement about the detectability of a planet, while the Bayesian upper limit is about the probability of a given planet flux. While upper limits are often thought of in a Bayesian way, they are mostly quoted as detection threshold in our field. This makes the interpretation of detection-based upper limit more challenging. The detection threshold, or contrast curve, is somewhat arbitrary and only a property of the noise (σ), which means it is not an optimal use of the data.

Our goal is to provide a conceptual framework for the analysis of direct imaging data, as well as informative examples rather than an explicit analysis recipe or formalism. Clearly defining a problem and its statistical representation before any calculation is extremely important. Our conceptual framework can also be applied to other cases, such as estimating planet occurrence rates, or combining multiple measured quantities, such as relative motion. Here, we have illustrated three typical cases:

1. Deriving an upper limit on planet flux when the location of the planet is known, which happens when, for example, the planet is robustly detected in a set of filters but not in others. The set of observations can be used to inform the astrophysical nature of a candidate (planet, brown dwarf, star, galaxies and so on) using Bayesian model comparison.
2. Constraining the mass of a hypothetical planet carving a gap in a protoplanetary disk (Ruane et al. 2017). We have shown that the data contains more information and is typically more constraining than the sole detection threshold suggests. Illustrating our method on the TW Hya system, there is a 99.9% probability given the data that the mass of hypothetical non-accreting planets in the gaps are below $2.4\,M_J$ at 24 au, $1.5\,M_J$ at 41 au, $1.3\,M_J$ at 47 au, and $0.9\,M_J$ at 88 au. With the same probability, the accretion rate of a $0.1\,M_J$ planet with a $1\,R_J$ circumplanetary disk inner radius, if they exist, is below $9.3 \times 10^{-7}\,M_J\,\mathrm{yr}^{-1}$ at 24 au, $5.0 \times 10^{-7}\,M_J\,\mathrm{yr}^{-1}$ at

41 au, $4.5 \times 10^{-7}\,M_J\,\mathrm{yr}^{-1}$ at 47 au, and $3.4 \times 10^{-7}\,M_J\,\mathrm{yr}^{-1}$ at 88 au.

3. We also introduced the problem of combining the RV and direct imaging measurement, where a joint Bayesian likelihood brings out the power of the two methods and can be used to infer the mass and orbital parameters of a planet (see Mawet et al. 2018).

ORCID iDs

Jean-Baptiste Ruffio ⓘ https://orcid.org/0000-0003-2233-4821
Dimitri Mawet ⓘ https://orcid.org/0000-0002-8895-4735
Ian Czekala ⓘ https://orcid.org/0000-0002-1483-8811
Bruce Macintosh ⓘ https://orcid.org/0000-0003-1212-7538
Robert J. De Rosa ⓘ https://orcid.org/0000-0002-4918-0247
Garreth Ruane ⓘ https://orcid.org/0000-0003-4769-1665
Michael Bottom ⓘ https://orcid.org/0000-0003-1341-5531
Jason J. Wang ⓘ https://orcid.org/0000-0003-0774-6502
Lea Hirsch ⓘ https://orcid.org/0000-0001-8058-7443
Zhaohuan Zhu ⓘ https://orcid.org/0000-0003-3616-6822

References

Abbott, B., Abbott, R., Adhikari, R., et al. 2004, PhRvD, 69, 082004
Akiyama, E., Muto, T., Kusakabe, N., et al. 2015, ApJL, 802, L17
Allard, F., Homeier, D., Freytag, B., & Sharp, C. M. 2012, in EAS Publications Ser. 57, Low-Mass Stars and the Transition Stars/Brown Dwarfs, ed. C. Reylé, C. Charbonnel, & M. Schultheis (Les Ulis: EDP Sciences), 3
Andrews, S. M., Wilner, D. J., Zhu, Z., et al. 2016, ApJL, 820, L40
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Baraffe, I., Chabrier, G., Barman, T. S., Allard, F., & Hauschildt, P. H. 2003, A&A, 402, 701
Bell, C. P. M., Mamajek, E. E., & Naylor, T. 2015, MNRAS, 454, 593
Black, D. C. 1980, NASSP, 436, 191
Bowler, B. P. 2016, PASP, 128, 102001
Brady, P. R., Creighton, J. D. E., & Wiseman, A. G. 2004, CQGra, 21, S1775
Brandt, T. D., McElwain, M. W., Turner, E. L., et al. 2014, ApJ, 794, 159
Bryan, M. L., Knutson, H. A., Howard, A. W., et al. 2016, ApJ, 821, 89
Cantalloube, F., Mouillet, D., Mugnier, L. M., et al. 2015, A&A, 582, A89
Castelli, F., & Kurucz, R. L. 2004, arXiv:astro-ph/0405087
Chauvin, G., Desidera, S., Lagrange, A.-M., et al. 2017, A&A, 605, L9
Cumming, A., Butler, R. P., Marcy, G. W., et al. 2008, PASP, 120, 531
Currie, T., Cloutier, R., Debes, J. H., Kenyon, S. J., & Kaisler, D. 2013, ApJL, 777, L6
Currie, T., Debes, J., Rodigas, T. J., et al. 2012, ApJ, 760, L32
Debes, J. H., Jang-Condell, H., & Schneider, G. 2016, ApJL, 819, L1
Dong, R., & Fung, J. 2017, ApJ, 835, 146
Finn, L. S. 1998, in Gravitational Waves, Second Edoardo Amaldi Conf., ed. E. Coccia, G. Veneziano, & G. Pizzella (Singapore: World Scientific), 180

---

[13] Documentation available at http://pyklip.readthedocs.io/en/latest/.
[14] http://www.astropy.org
[15] https://matplotlib.org

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, arXiv:1804.09365

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1

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

Hardy, A., Schreiber, M. R., Parsons, S. G., et al. 2015, ApJL, 800, L24

Howard, A. W., Marcy, G. W., Fischer, D. A., et al. 2014, ApJ, 794, 51

Hunter, J. D. 2007, CSE, 9, 90

Janson, M., Carson, J. C., Lafrenière, D., et al. 2012, ApJ, 747, 116

Jeffreys, H. 1961, Theory of Probability (Oxford: Oxford Univ. Press)

Jensen-Clem, R., Mawet, D., Gomez Gonzalez, C. A., et al. 2018, AJ, 155, 19

Joergens, V., Janson, M., & Müller, A. 2012, A&A, 537, A13

Kalas, P., Graham, J. R., Chiang, E., et al. 2008, Sci, 322, 1345

Kass, R. E., & Raftery, A. E. 1995, J. Am. Stat. Assoc., 90, 773

Kass, R. E., & Wasserman, L. 1996, J. Am. Stat. Assoc., 91, 1343

Kenyon, S. J., Currie, T., & Bromley, B. C. 2014, ApJ, 786, 70

Lannier, J., Lagrange, A. M., Bonavita, M., et al. 2017, A&A, 603, A54

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

Marengo, M., Stapelfeldt, K., Werner, M. W., et al. 2009, ApJ, 700, 1647

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

Mawet, D., Hirsch, L., Lee, E. J., et al. 2018, ApJ, in press (arXiv:1810.03794)

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

Mesa, D., Zurlo, A., Milli, J., et al. 2017, MNRAS, 466, L118

Meshkat, T., Bailey, V., Rameau, J., et al. 2013, ApJL, 775, L40

Nielsen, E. L., & Close, L. M. 2010, ApJ, 717, 878

Nielsen, E. L., De Rosa, R. J., Rameau, J., et al. 2017, AJ, 154, 218

Nielsen, E. L., De Rosa, R. J., Wang, J., et al. 2016, AJ, 152, 175

Perryman, M., Hartman, J., Bakos, G. Á., & Lindegren, L. 2014, ApJ, 797, 14

Pueyo, L. 2016, ApJ, 824, 117

Rapson, V. A., Kastner, J. H., Millar-Blanchaer, M. A., & Dong, R. 2015, ApJL, 815, L26

Robert, C. P., Cornuet, J.-M., Marin, J.-M., & Pillai, N. S. 2011, PNAS, 108, 15112

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Röver, C., Messenger, C., & Prix, R. 2011, arXiv:1103.2987

Ruane, G., Mawet, D., Kastner, J., et al. 2017, AJ, 154, 73

Ruffio, J.-B., Macintosh, B., Wang, J. J., et al. 2017, ApJ, 842, 14

Sivia, D. S., & Skilling, J. 2006, Data Analysis, A Bayesian Tutorial (Oxford: Oxford Univ. Press)

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

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

Vaccaro, T. R., Wilson, R. E., Van Hamme, W., & Terrell, D. 2015, ApJ, 810, 157

van Boekel, R., Henning, T., Menu, J., et al. 2017, ApJ, 837, 132

Vigan, A., Bonavita, M., Biller, B., et al. 2017, A&A, 603, A3

Wagner, K., Apai, D., Kasper, M., et al. 2016, Sci, 353, 673

Wahhaj, Z., Liu, M. C., Biller, B. A., et al. 2013, ApJ, 779, 80

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

Weinberger, A. J., Becklin, E. E., Schneider, G., et al. 2002, ApJ, 566, 409

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Zhu, Z. 2015, ApJ, 799, 16