



HAL
open science

Logistic regression to boost exoplanet detection performances

Hadrien Cambazard, Nicolas Catusse, A. Chomez, A.-M. Lagrange

► **To cite this version:**

Hadrien Cambazard, Nicolas Catusse, A. Chomez, A.-M. Lagrange. Logistic regression to boost exoplanet detection performances. 2024. hal-04491635v2

HAL Id: hal-04491635

<https://hal.science/hal-04491635v2>

Preprint submitted on 28 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Logistic regression to boost exoplanet detection performances

Hadrien Cambazard^{1*}, Nicolas Catusse¹, Antoine Chomez^{2,3}, Anne-Marie Lagrange^{2,3}

¹Univ. Grenoble Alpes, CNRS, Grenoble INP, G-SCOP, F-38000 Grenoble, France

²Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique, Observatoire de Paris, Univ. PSL, Sorbonne Univ., Univ. Paris Diderot, France

³Univ. Grenoble Alpes, Institut de Planétologie et d'Astrophysique de Grenoble, France

27 November 2024

ABSTRACT

Direct imaging of exoplanets requires to separate the background noise from the exoplanet signals. Statistical methods have been recently proposed to avoid subtracting any signal of interest as opposed to initial self-subtracting methods based on Angular Differential Imaging (ADI). However, unless conservative thresholds are chosen to claim for a detection, such approaches tend to produce a list of candidates that include many false positives. Choosing high, conservative, thresholds leads to miss the faintest planets. We extend a statistical framework with a logistic regression to filter the list of candidates. Features with physical/optical meaning (in two wavelengths) are used, leading to a very fast and pragmatic approach. The overall method requires a simple edge detection (image processing) and clustering algorithm to work with sub-images. To estimate its efficiency, we apply our approach to targets observed with the ESO/SPHERE high contrast imager, that were previously used as tests for blind surveys. Experimental results with injected signals show that either the number of false detections is considerably reduced or faint exoplanets that would otherwise not be detected can be sometimes found. Typically, on the blind tests performed, we are now able to detect around 50% more of the injected planets with an SNR below 5, and with a very low number of additional candidates.

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

1 INTRODUCTION

High contrast imaging (HCI) of exoplanets has become an important area in astronomy, as it provides direct information of the exoplanets orbital parameters, as well as measurements of their fluxes, which, in turn, can be used to estimate, through comparisons with evolution models, their masses, effective temperatures, and gravity. Coupled to spectroscopy, HCI also allows for atmosphere characterization. It is recognized as a key technique to study Earth twins in the future. Given today instrumental capabilities, only massive giant planets orbiting further than typically 10 au have been detected, and most of them are young (tens to hundreds of Myr).

HCI relies on three pillars: adaptive optics (AO), coronagraphy and data processing. Since the first AO imagers in the nineties on 4m class telescopes, two generations of instruments have been developed on 10m class telescopes, with increased AO and coronagraphy capabilities. The first generation led to imaging of a few exoplanets (see e.g. Chauvin et al. 2004; Marois et al. 2008; Lagrange et al. 2009). Second generation instruments using extreme AO and higher performance coronagraphs (Beuzit et al. 2019; Macintosh et al. 2014) have allowed us to detect a few more planets and provide interesting constraints on the massive giant planet demographics beyond 10 au (Nielsen et al. 2019; Vigan et al. 2021).

While more powerful instruments will allow detecting lighter and closer planets and deriving more accurate and complete information on exoplanets demographics, using better detection algorithms to separate the stellar noise from the exoplanet signal as much as possible can allow for improving exoplanets detection capabilities on today data. A number of techniques have been designed to cancel

out the background speckles by combining images taken at consecutive times. The key idea of the ADI technique Marois et al. (2005); Lagrange et al. (2010) is to subtract the temporal median to each frame and realign the residual images to recover the remaining signals of the companions at the same location. This median acts as a reference Point Spread Function (PSF) that models the speckle patterns to be subtracted. Various alternatives to this basic approach have been proposed to estimate the reference stellar PSF to be removed such as the LOCI algorithm Marois et al. (2013); Absil et al. (2013), the principal component analysis based algorithms Soummer et al. (2012) or LLSG Gomez Gonzalez et al. (2016) which uses a low rank approximation to model the PSF. Yet, the subtraction of the PSF may lead to a loss of the companion signal because it is fitted in the PSF. This so-called *self-subtraction* is a limitation associated to such approaches. Statistically-based methods then appeared from the need to evaluate the confidence of each detection and have been applied after the speckle cancellation step. For instance, AN-DROMEDA (Cantalloube et al. 2015) uses a maximum likelihood estimation to detect planetary signals. This estimation is performed under the assumption of a Gaussian distribution of the noise. This approach produces a Signal to Noise Ratio (SNR) map that can be filtered out with a threshold to identify detected sources. A planetary signal shows an expected pattern typically made of a positive oval lobe surrounded by two negative lobes. But the resulting SNR map still contains a number of artifacts whose signals are above the threshold; three sources of false detections are diagnosed in Cantalloube et al. (2015): a tertiary lobe in the expected pattern, spider diffraction patterns and high speckle noise in close areas to the star. To find and reject the artifacts, a number of tests are performed by fitting the morphological expected shape of the planetary signal. All previously mentioned works rely on unsupervised learning techniques. The lack

* E-mail: hadrien.cambazard@g-scop.grenoble-inp.fr

of a large enough labelled dataset of positive and negative samples is a challenge for supervised learning algorithms. Firstly, the small number of confirmed exo-planets leads to class imbalance with very few guaranteed positive samples. Secondly, images where no candidate planets have been detected can not be used as representative of negative samples since yet undiscovered planets might be present.

Recently, a random forest and neural network classifier were presented in [Gomez Gonzalez et al. \(2018\)](#) where the training data is obtained by injecting fake exo-planets and blurring pre-processed ADI sequence of images. Alternatively, a Generative Adversarial Network is trained in [Yip et al. \(2020\)](#) to get a generative model of the speckle noise pattern and use it to create large negative labeled datasets. Fake planets are then injected in such guaranteed negative samples. A Convolutional Neural Network is then trained and planets are localized using Class Activation Maps (CAM). A post-processing based on half-sibling regression (HSR) is proposed in [Gebhard et al. \(2020\)](#). The authors argue that the speckle pattern can be approximately anti-symmetric across the origin in some circumstances which can help to choose the predictors of the HSR framework. Recently, supervised deep learning has been combined with the statistical model of PACO, leading to a gain of 0.5 mag compared to PACO ([Flasseur et al. 2023](#)).

In this paper, we extend the statistical framework of [Flasseur et al. \(2018\)](#) using a simple logistic regression to filter the list of candidates and simple features expected from a planetary signal and noise. The overall methodology also relies on edge detection and clustering to present a meaningful list of candidate planets to the user. The approach is described in Section 2. We apply it on test cases in Section 3, and analyse the results. We finally use this approach to explore the environment of one iconic system (Section 4).

2 APPROACH

2.0.1 Input data

We use SPHERE/InfraRed Dual-band Imager and Spectrograph (IRDIS) ([Dohlen et al. 2008](#)) data obtained in Angular and Spectral Differential Imaging mode (ASDI). Both H2 and H3 spectral channels are considered. A pre-processing based on a standard pipeline in astronomy is performed with a number of calibration steps (dark, flat, offset, ...) partly dedicated to the instrument. The data sets contain stellar noise close to the star, and background/detector noise further away. The PACO algorithm for (PAtches COvariances) [Flasseur et al. \(2018\)](#) is used to model the noise and to produce an SNR map. In this approach, the background variations are considered to be spatially correlated and non-stationary. A statistical model for these fluctuations is proposed in [Flasseur et al. \(2018\)](#) and the presence of a companion is established using a statistical hypothesis test. More precisely, the background is assumed to follow a multivariate Gaussian distribution locally. The field of view is thus decomposed into small patches that are large enough to contain the core of a point source. A Gaussian distribution is estimated for each patch, to take into account spatial correlation, by using the different images of the patch over time. A hypothesis test is formulated for the pixel at the center of the patch to decide whether a pure background hypothesis can be rejected. The test statistic can be interpreted as a Signal to Noise Ratio (SNR) producing an image or a map that can be inspected (visually or algorithmically) for the presence of companions. Application of PACO to astrophysical sources can be found in [Chomez et al. \(2023b\)](#) and in [Chomez et al. \(2023a\)](#).

A simple threshold of this SNR map can in principle be used

to identify sources. A threshold of 5σ has been used so far when analysing the data with PACO. Using lower thresholds would indeed lead to too many candidates (typically hundreds). [Cantalloube et al. \(2015\)](#) considered slightly lower thresholds by adding a number of dedicated tests to reject artifacts. We propose to enrich the threshold with a limited number of features with optical/physical meanings.

2.1 Principle

We use a simple classifier based on logistic regression to filter the list of source candidates. Simple features are computed on the speckle and planet patterns resulting in a very fast and pragmatic algorithm to classify sub-images of the SNR map.

It is similar to [Gebhard et al. \(2020\)](#) in the sense that scientific knowledge is directly provided to the classifier but it follows a very different implementation. The signal of an exoplanet in the original image is expected to have a specific structure in the SNR map. Typically, an Airy figure in the original image affects the SNR distribution and even if the corresponding structure does not necessarily have the exact characteristics of an Airy Figure, it can still be indicative of the presence of an exoplanet. For sake of simplicity, we shall refer to it as an Airy Figure though this is not technically exact. In addition, a speckle is similar in both wavelengths but appears at different locations in each channel according to a radial shift from the center of the image (the star). A number of such features are computed on small sub-images and a simple classifier is trained to discriminate noise from planetary signals.

2.2 Methodology

The overall process starts from the H2 and H3 SNR maps available after processing the original data cube with the PACO algorithm. The process is summarized [Figure 1](#) in five steps:

(i) (*Training*) Generation of the SNR maps with injected exoplanets or noise. Small sub-images, also referred to as **stamps**, will be extracted from these guaranteed SNR maps. Typically, a stamp containing an injected exoplanet is considered as a positive stamp (of the positive class). Positive stamps are directly produced from the SNR maps injected with planets by centering the stamp at the location of the injection.

(ii) (*Training*) Negative stamps refer to stamps that do not contain a planet and can be centered at any pixel of the noise maps. This would give rise to nearly a million stamps for a single map with a lot of redundancy (overlapping). Since a candidate object can only be detected at a peak of SNR value we use image processing techniques to center the stamps on SNR peaks (see [Section 2.2.2](#)). This step produces a collection of negative stamps extracted from the SNR maps guaranteed with noise only.

(iii) (*Training*) A classifier is trained from the collection of stamps whose class is known. Note that the classifier is thus dedicated to a specific image that was used to produce the SNR maps with planets/noise.

(iv) (*Usage*) For each pixel u of SNR intensity greater than a threshold value (in practice set to 2), a stamp centered on u is submitted to the classifier. This step produces a collection of **candidates**.

(v) (*Usage*) The stamps classified as possible candidates are then clustered (since they can considerably overlap) and the set of clusters is presented to the user.

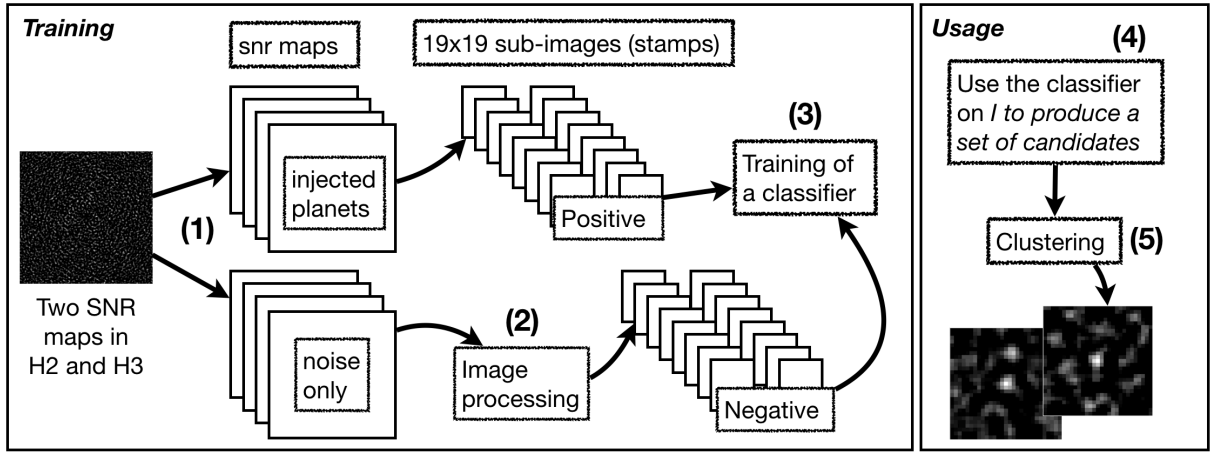


Figure 1. Overview of the methodology.

2.2.1 Step (1): Generation of the training data-set: planet's injections and noise map

A data cube is used to generate a number of SNR maps (see Figure 1) that contains either only noise or signals of previously injected fake giant planets of various masses. The objective is to provide guaranteed positive and negative samples for the machine learning algorithm.

Noise maps. There are two ways to create a "noise map only" from the data set. The first one is to invert the direction of rotation. Because we use ADI-based techniques taking advantage of the rotation, inverting the direction destroys any astrophysical signal, leaving only noise. Another way to create noise only maps is to perform a temporal shuffle of the frames.

Injection of fake planets. Fake planets (FPs) with given masses (expressed in Jupiter masses M_{jup}) are randomly injected in the reduced and centered data cubes provided by the SPHERE data center (Delorme et al. 2017), (Chomez et al. 2023b) taking into account the inverted direction of rotation (see above). The injected masses range from 1 to 5 M_{jup} . A minimum distance between the injections is enforced to avoid any stacked signal. To convert masses into contrasts in the H2 and H3 bands, we use COND (Allard et al. 2001) evolution models, and assume that the planet is coeval with its parent star. Those data cubes are then processed by PACO using ADI and ASDI.

Each SNR map generated from an original image I (made of two SNR maps in H2 and H3) gives a number of sub-images (*stamp*) that are known to contain or not an object (an injected planet/companion). The size of the stamps, 19x19 pixels, is large enough to give the context needed to detect an astrophysical point source (field companion or exoplanet).

2.2.2 Step (2): Negative stamps from the noise maps

The extraction of the stamps is performed using the H2 SNR map. To reduce the number of stamps and center them on SNR peaks, we apply edge detection techniques (Sobel's filter, Duda & Hart 1973) and compute a gradient map of the SNR map. More precisely, each pixel of such a map gives an approximation of the norm of the gradient of the corresponding SNR function. An example of such a gradient map can be seen Figure 2.

A pixel u of an edge is identified as a pixel with a gradient value g_u greater than a threshold (denoted t in Algorithm 1). Consider the

graph of *edges* where vertices are pixels u such that $g_u \geq t$ (edge pixels) and connect two such vertices if they are adjacent pixels in the image. More precisely, consider the graph $G = (V, E)$ with $V = \{u | g_u \geq t, u \in I\}$ and $E = \{(u, v) | u \in V, v \in V, u \text{ and } v \text{ are adjacent}\}$. An area of high SNR in the original image is related to a connected component of G . These components might take a variety of shapes and for sake of simplicity we consider the smallest rectangular **box** containing the component as the relevant area.

The value of t controls the number of components found and Algorithm 1 performs a dichotomy search to reach a given average size of the boxes identifying the areas of interest. In practice we set this *target size* to 12 pixels. Note that the stamps definition could also be done from the H3 SNR map which might contain SNR peaks located differently. The final set of stamps would differ and might be more complete but this remains to be investigated.

Algorithm 1 Stamps generation algorithm.

Input An SNR map I of noise in H2, a *target size* for the areas of interest

Output A collection of stamps

- 1: Set the SNR of each pixel $u \in I$ with a negative SNR at 0
 - 2: Compute an approximation of the norm of the gradient g_u for each pixel $u \in I$
 - 3: $lb \leftarrow 0, ub \leftarrow 3 \times \text{median}(\{g_u | u \in I\}), t \leftarrow \frac{lb+ub}{2}$
 - 4: **do**
 - 5: Build $G = (V, E)$ with $V = \{u | g_u \geq t, u \in I\}$ and $E = \{(u, v) | u \text{ and } v \text{ are adjacent pixels}\}$
 - 6: Compute connected components of G and their boxes
 - 7: **if** average size of boxes $>$ *target* **then**
 - 8: $lb \leftarrow t$
 - 9: **else** $ub \leftarrow t$
 - 10: $t \leftarrow \frac{lb+ub}{2}$
 - 11: **while** $(ub - lb) > 10^{-3}$
 - 12: Build a stamp centered at the pixel of maximum SNR in each box.
-

2.2.3 Step (3): Training

A simple classifier based on Logistic Regression is trained from the collection of stamps whose class is known at this stage. We outline two key points. Firstly, the classifier is dedicated to a specific original

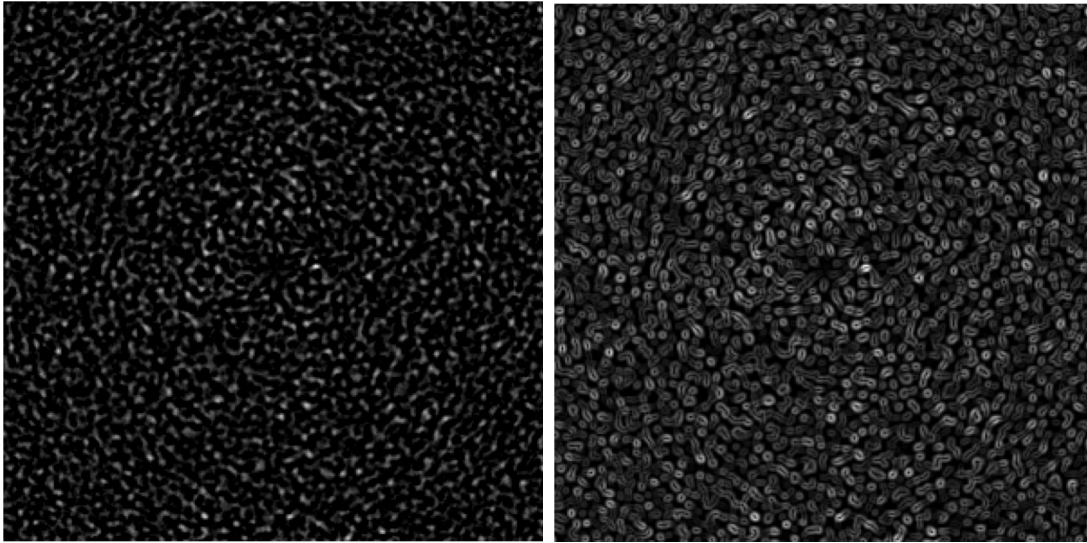


Figure 2. An SNR map (left) and the corresponding gradient map (right).

image. The stamps generated for learning are based on injections and randomization of the original data cube on which the classifier is finally used. So that it potentially takes into account the peculiarities on this data set (e.g weather conditions). Secondly, the learning is not performed on the whole image. On the contrary, a number of limited and meaningful features are computed for each stamp. This typically refers to old-fashioned learning as opposed to deep learning. A dozen of features are computed and detailed in the Appendix. These features relate to high level descriptive statistic of the stamp such as the mean SNR value (**MeanSnr**) or the mean, maximum, and standard deviation of the gradient approximation of the stamp (**MeanGra**, **MaxGra**, **StdevGra**) but also to physical phenomenons. A key feature targets the pattern that can leave an Airy figure on the SNR map (**AiryFig**). Another one (**MeanSpec**) tries to quantify whether the 9 central pixels of a stamp indicate that the SNR signal is due to a speckle. Note that the features are not real physical models of the phenomenon but simple numerical quantities that are correlated with the presence of the phenomenon. An analysis of these correlations is presented Section 3.3. Most of the features are computed on the H2 SNR part of the stamp but a number of them are also computed on the H3 signal (typically the features dealing with the presence of a speckle). In total, a dozen of such features are used. Finally, the logistic regression is trained using cross-validation.

2.2.4 Step (4): Usage

The classifier is then used only on stamps extracted on the original image and whose center is a pixel with a SNR (in H2) greater than 2. We chose this value because below this threshold in H2, it would currently not be considered as a potential candidate by a user (an astrophysicist) in any case. A more elaborate criterion can be used, for instance using the SNR value of multiple wavelengths. But in principle, any pixel can be considered and this step is not computationally time consuming. The features of the stamp centered on that pixel are computed and if the value of the logistic regression function of the classifier is above 0.5, the stamp is kept as a candidate. It is considered to belong to the class of objects. The probability of 0.5 can be tuned and this is discussed in Section 3.7.

2.2.5 Step (5): Usage - Clustering of candidates

Since stamps centered on any pixel are potentially submitted to the classifier, adjacent pixels are likely to be classified similarly. As a result, candidates are made of close clusters that overlap and two closely overlapping stamps are most likely the same object. We therefore present the classification result as clusters of candidates.

The distance between two stamps is defined as the number of distinct pixels. The smaller this distance, the more the stamps overlap. An agglomerative clustering algorithm (Müllner 2011) merges a cluster pair if the minimum of the distances between the stamps of a cluster is lower than a threshold. In practice this threshold is set to 55.5% of the pixels so two stamps of the same cluster overlap over at least 55.5% (200 pixels for 19x19 stamps).

3 TESTS

3.1 Data sets and terminology

We used four IRDIS data sets obtained on HD 108767B, HIP 1993, HIP 12394 and HIP 107345. These stars were chosen as they were also used in a blind test that was performed among the SPHERE SHINE consortium to compare the merits of different algorithms (see below). They are representative of different atmospheric conditions. Table 1 summarize the aforementioned stars properties and Table 3.1 provide the observing conditions.

The four data sets were used to generate **true positive** (TP) stamps (by planet’s injection) and **true negative** (TN) stamps for the learning step (by processing, with algorithm 1, a map of pure noise generated from the original image). The blind tests consist in performing a number of injections in the original image and submitting the resulting modified data set to the classifier. Additional unknown exoplanets might also be present since the original image is used for the test.

In the following, we use the following notations and terminology: #U is the number of stamps extracted on the original image and considered at the Usage step. The classifier is used only on stamps with a SNR greater than 2 in H2 and this number is reported in columns #U_{snr>2}. #TP and #TN refers to numbers of stamps that are known to contain an injected planet or noise *i.e true positives* or *true negatives*. Finally, when considering the results obtained with

Star name	RA (J2000)	DEC (J2000)	Spectral Type	Rmag	Hmag	Age ^a (Myr)
HD108767B	12 29 50.8908	-16 31 15.2081	K 1	8.2	6.3	180 ⁺¹⁷⁰ ₋₈₀
HIP1993	00 25 14.6618	-61 30 48.2527	M0V	10.7	7.9	45 ⁺⁵ ₋₁₀
HIP12394	02 39 35.3612	-68 16 01.0103	B8V	4.1	4.4	45 ⁺⁵ ₋₁₀
HIP107345	21 44 30.1227	-60 58 38.8946	M0	10.5	8.1	45 ⁺⁵ ₋₁₀

Table 1. Targets used for the tests.

Notes. ^a : Ages values extracted from [Desidera et al. \(2021\)](#)

STAR	OBS DATE	FILTER	DIT(s)×Nframe	ΔPA (°) ^a	Seeing (") ^b	Airmass ^b	τ ₀ (ms) ^{a,b}	Program ID
HD108767B	2018-01-24	DB_H23	64x72	94.4	0.59	1.02	8.3	1100.C-0481(D)
HIP1993	2015-11-28	DB_H23	64x64	25.8	1.57	1.26	7.2	096.C-0241(B)
HIP12394	2016-09-15	DB_H23	32x160	29.1	0.42	1.38	9.2	097.C-0865(D)
HIP107345	2015-07-04	DB_H23	64x64	26.0	1.07	1.25	2	095.C-0298(C)

Table 2. Test targets observation logs. **Notes:** ^a: DIT corresponds to the detector integration time per frame, ΔPA is the amplitude of the parallactic rotation, τ₀ corresponds to the coherence time. ^b: Values extracted from the updated DIMM info and averaged over the sequence.

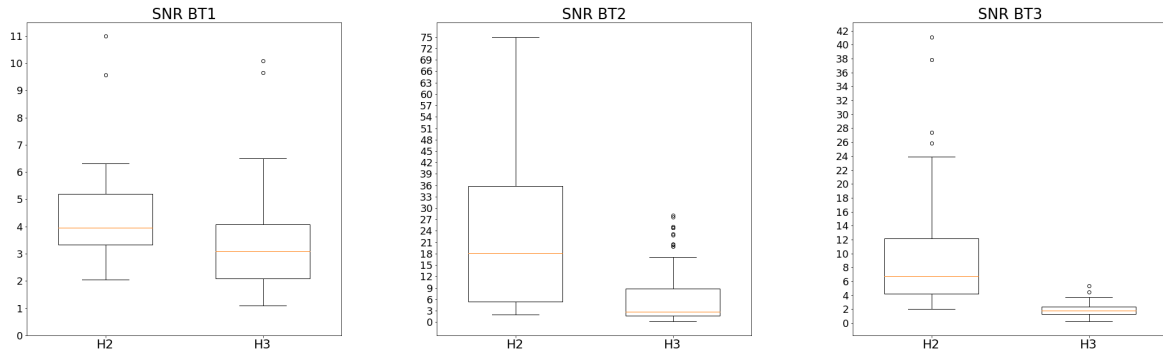


Figure 3. Box plots summarizing the SNR distribution of the injections in H2 and H3 (resp. left and right in each boxplot) for the three Blind Tests: BT1, BT2, BT3. Each boxplot shows the minimum, first quartile, median, third quartile, maximum (by convention, the third quartile plus at most 1.5 of the inter-quartile range) and additional points above.

our algorithm, we note #TPF the number of True Positives Found, and #C the number of remaining Candidates proposed by the algorithm. These candidates are likely **false positive** but since the usage is done on the original image, it could also be real, and so far undetected, exoplanets. Therefore, we refer to them as **candidates**.

The first column of Table 4 reports the number of true positive (column #TP) and negative stamps (column #TN) obtained with our methodology and that have been used to train the classifier (column Learning). The remaining columns gives an overview of the number of stamps generated for the three Blind Tests (BT1, BT2, BT3) detailed below. Note that some injections might lead to SNR values below 2 and thus would not be submitted to the classifier. The results are therefore given only for injected planets giving an SNR greater than 2 in H2 (the exact numbers of such injections are reported in columns titled #TP for each blind test).

3.2 Tests description

We started with a blind test constructed by the SHINE consortium, BT1. Eight fake companions featuring spectral types between early M and late T were injected for each of the stars considered; their contrasts were chosen to provide a TLOCI signal ([Marois et al. 2014](#)) of about 5 sigma (mean of H2 and H3 contrasts). We note that the objects do not necessarily represent realistic planets. Note also that in a few cases, the injected planets are not detected with PACO even

with a low threshold. This happens when the planets are very close to the stars, and their expected contrasts are overestimated. Finally, one out of the 8 planets around HD10767B has an SNR in H2 lower than 2, and two out of the 8 planets around HIP12394 have to SNR in H2 lower than 2. They are therefore not be considered here.

BT2 and BT3 are blind tests generated using the process described in section 2.2.1. The planet’s masses considered are 1, 2, 3, 4 and 5 MJup. The contrasts of the companions wrt the stars are computed according to their masses and to the age of the stars (planets and stars are assumed to be coeval), and using the COND models. Conversely to BT1, the fake planets correspond to realistic cases. We note that the respective contrasts in BT1 are very different from those in BT2 and BT3. Figure 3 shows the distribution of the SNR values of the injections in the three blind tests.

The last test, BT4, is similar to BT2 and BT3. We massively injected FPs close to the star, between 0.25 and 1 arcsec. The FPs were chosen so that their contrasts be along the 5sigma contrast curves to test the capability of the classifier. We chose to use the classifier on HIP12394 with 3Mjup and 4Mjup injected planets because both planets cross the contrast curve between 0 and 1 arcsec.

3.3 Features analysis

We analyse the distribution of six of the features (**MeanSnr**, **MeanGra**, **MaxGra**, **MaxMin**, **AiryFig**, **MeanSpec**) across true positive

star name	Learning			BT1		BT2		BT3	
	#TP	#TN	#TN _{snr≥2}	#TP	#U _{snr≥2}	#TP	#U _{snr≥2}	#TP	#U _{snr≥2}
all	1026	77565	8180	29	14691	140	19017	90	15878
HD108767B	259	15621	1724	7	3577	26	3777	15	3390
HIP1993	253	15228	1271	8	2376	37	3479	24	2810
HIP12394	251	15636	2920	6	6670	39	8226	25	7191
HIP107345	263	31080	2265	8	2068	38	3535	26	2487

Table 3. Number of positive (planet’s injection) and negative stamps (extracted from a pure noise map generated with algorithm 1). The total number as well as the number of negative stamps of SNR H2 greater than 2 are reported.

star name	Learning		BT4	
	#TP	#TN	#TP	#U _{snr≥2}
HIP1993	243	15076	34	2863
HIP12394-4M	214	15540	34	8384
HIP12394-3M	214	15540	34	7707
HIP107345	192	15791	34	2421

Table 4. Number of positive (planet’s injection) and negative stamps (extracted from a pure noise map generated with algorithm 1). The total number as well as the number of negative stamps of SNR H2 greater than 2 are reported.

and true negative stamps. We recall that a feature is simply a real number computed on a stamp and a good feature is correlated to a class (positive/negative). Figure 4 provides, for each feature, two box plots showing the distribution of its values for true positive (box labeled pos) and true negative (box labeled neg) stamps. A box plot gives a summary of the distribution in 5 numbers from bottom to top: minimum, first quartile, median, third quartile and maximum. Typically, the values of the feature for 50% of the stamps lie in the box. The median value is the orange horizontal line within the box. Values considered as outliers (below or above 1.5 the inter-quartile range, which is defined as the maximum) are not shown for sake of clarity. The correlation coefficient (r value) is given for each feature.

We expect useful features to show distinct distributions for positive and negative stamps in order to help the classifier discriminating between the two. The size of the intersection of the two distributions (for the positive and negative class) gives an idea of the discrimination power of the feature.

Four features appear decisive (correlation coefficient $r \geq 0.6$): the mean snr intensity (**MeanSnr**), the gradient features (**MeanGra**, **MaxGra**) as well as the feature related to the presence of an Airy Figure (**AiryFig**).

The feature related to speckles is not strongly correlated to the presence of a companion ($r = 0.29$). This is expected as speckles are only present within the star halo. Restricting the analysis to the halo region, between 30 and 140 pixels (i.e. 370 to 1700 mas), significantly increases the correlation to $r = 0.61$ (see Figure 5). It might therefore be appropriate to build two distinct classifiers, one for the star halo that includes **MeanSpec** and one for the remaining area without it. But this tends to complicate further the overall process. We decided to keep it simple for the moment and used the feature **MeanSpec** on the entire field with an additional 0/1 indicative feature (f_9 in the Appendix) defining whether a stamp is or not in the star halo. In other words, this indicative feature tells when **MeanSpec** is relevant and can be eventually help the classifier.

3.4 Analysis

The results are presented in Table 5 for BT1, BT2, BT3 and BT4. The results of the classifier (column RegL) are compared to the threshold

approach (thresholds of 3 and 5) for two flavors of PACO: ADI and ASDI. In ADI, the SNR is computed considering separately the H2 and H3 channels. So a detection is made if any of the SNR value in H2 or H3 is above the threshold. In ASDI, PACO optimally combines both bands Flasseur, Olivier et al. (2020). The number of planets found (#TPF) and the number of candidates (#C) are reported for each approach (see Section 3.1 for the definition of #TPF and #C).

Overall, ADI, ASDI 3 and RegL retrieve most of the planets. Yet, the former two detect also a huge number of candidates: typically several hundreds candidates are found in ASDI 3 and often about one hundred in ADI 3, while RegL detects a much more limited number of candidates. Such large lists of candidates found by ADI and ASDI 3 prevent from identifying the true positives. This justifies the usual choice of a threshold of 5 when using PACO. Using a threshold of 5 removes nearly all candidates but the detection performances are reduced (in particular for BT2 but also BT3) compared to our classifier. Hence, RegL performs better than ADI or ASDI, giving a better compromise between the number of detections and the number of candidates. We detail the results of each test below.

BT1. RegL finds 23 planets while ADI 3 finds 27. Yet, RegL finds 16 false positives, to be compared to the 715 found by ADI 3. ASDI 5 find 25 planets and avoids any additional candidates. We will discuss this result below.

BT2. RegL retrieves 120 out of the 140 injected planets, identifying one more planet than ADI 3. Noticeably, RegL finds only 16 additional candidates (likely false positive), to be compared to the 543 found with the threshold approach. ASDI 3 finds 114 planets and produces much more additional candidates than RegL. RegL finds more planets than ADI and ASDI 5 (which find 96 planets), with a yet slightly larger number of candidates (16 instead of 0 and 4, respectively).

BT3. RegL does not performs better than ADI or ASDI 3 regarding the number of planets found but the number of candidates with RegL (18) is significantly smaller with with ADI (592) or ASDI (1630). It behaves better when compared to ADI or ASDI 5, with a larger amount of detected planets (69 versus 58 with ADI 5 and 64 with ASDI 5), and a small number of candidates (18 versus 0 and 3, respectively).

BT4. RegL does not perform better than ADI or ASDI 3 regarding

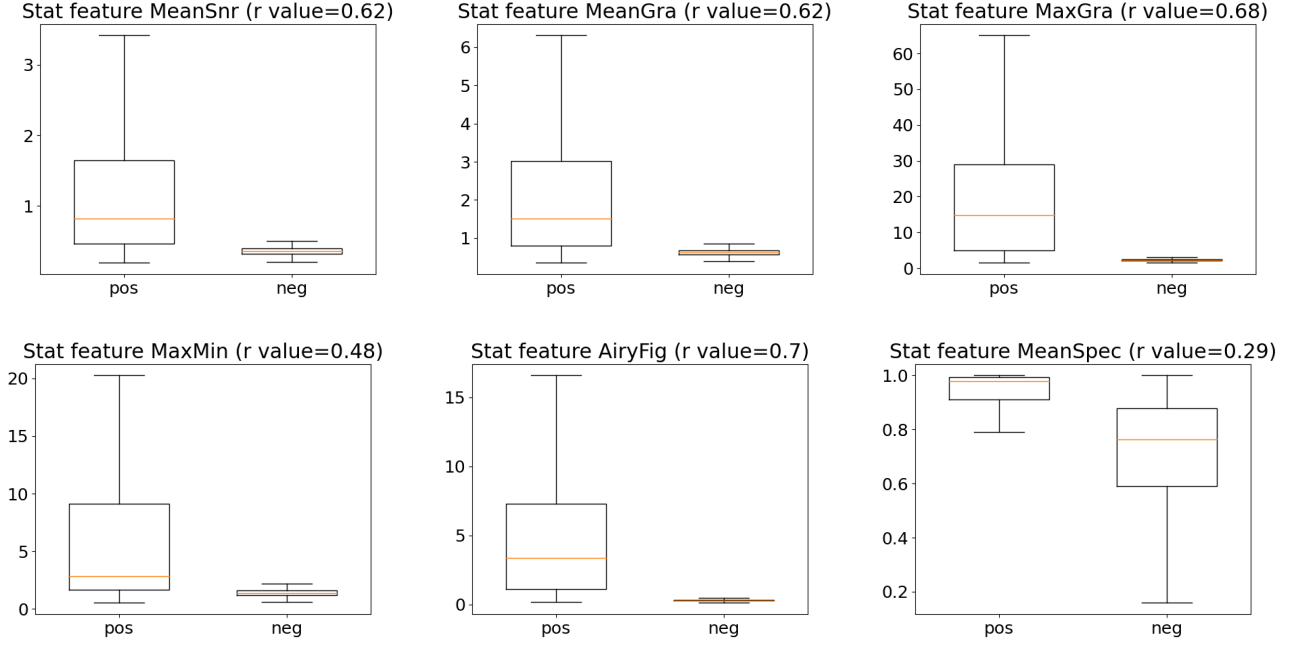


Figure 4. Box plots summarizing the distribution of six features by showing the minimum, first quartile, median, third quartile, maximum. The distribution for positive and negative stamps are shown for each feature to help visualize whether the feature discriminates the two classes. The values are computed over stamps of SNR $H_2 \geq 2$ of the four images (In total 2069 positives, 18074 negatives).

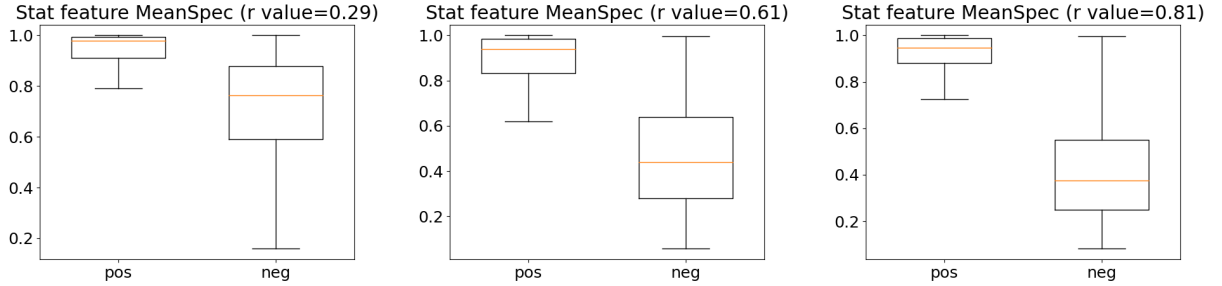


Figure 5. Distribution of the **MeanSpec** feature for three selected subsets of stamps respectively from left to right: (1) all stamps of SNR $H_2 \geq 2$ (2069 positives, 18074 negatives), (2) all stamps of SNR $H_2 \geq 2$ located between 30 and 140 pixels (554 positives, 1969 negatives), (3) all stamps of SNR $H_2 \geq 2.5$ located between 30 and 140 pixels (492 positives, 558 negatives).

the number of planets found but the number of candidates with RegL is much smaller than with ADI 3 or ASDI 3. It is relevant to notice that despite of the little gap between the number TPF and the number of planets injected, RegL finds for HIP1993, HIP12394-4M and HIP107345 almost all the sources with a center SNR > 2 and misses only 5 planets with SNR very close to 2 for HIP12394-3M. In fact, only 26 planets out of the 34 have an SNR above the threshold of RegL. This gap can be explained by the fact that we injected along the 5sigma contrast curves at the constant mass so some planets close to the star are below this curve.

Figure 7 shows the number of planets found (#TPF) (for all blind tests) depending on their projected separations to the star (in $[0, 1715[$ or ≥ 1715 mas), as well as on their SNR value in H_2 (in $[2, 5[$ or ≥ 5). Results for ADI 3 are not reported here since they lead to far too many candidates. As expected, the logistic regression

improves over ADI/ASDI for low H_2 SNR ¹. It can also be noted here that most of the injections of these blind tests have a high SNR and are located outside of the star halo.

Finally, note that the classifier is expected to perform better on BT2/BT3 than on BT1 because the exact same injection process is used at the learning step and the usage step for BT2/BT3. Conversely, the injections of BT1 were performed independently of the present work with different parameters (thus a different opinion on what is most realistic) and can show different patterns that are therefore not learnt by the classifier. In particular, injections of BT1 can have a low SNR in H_2 but a high a SNR in H_3 . This pattern is not typical of the injection process used in the present work (section 2.2.1) for the learning step. As a result, the classifier does not learn to consider

¹ A detection can be made by ASDI 5 below an SNR threshold of 5 in H_2 because PACO combines both H_2 and H_3 bands in ASDI. It can also be made in ADI 5 if the SNR in H_3 alone is above 5.

	RegL			ADI 3		ADI 5		ASDI 3		ASDI 5	
	#Inj	#TPF	#C	#TPF	#C	#TPF	#C	#TPF	#C	#TPF	#C
BT1											
HD108767B	7	5	11	6	153	3	0	6	717	5	0
HIP1993	8	6	1	8	100	3	0	8	724	7	0
HIP12394	6	5	2	5	397	2	0	6	611	5	0
HIP107345	8	7	2	8	65	2	0	8	694	8	0
Total	29	23	16	27	715	10	0	28	2746	25	0
BT2											
HD108767B	26	22	13	21	2	16	0	23	716	18	1
HIP1993	37	33	1	32	91	27	0	31	627	23	1
HIP12394	39	27	1	28	373	20	0	24	620	22	1
HIP107345	38	38	1	38	77	33	0	36	659	33	1
Total	140	120	16	119	543	96	0	114	2622	96	4
BT3											
HD108767B	15	6	10	8	118	2	0	8	440	3	1
HIP1993	24	20	4	19	65	19	0	23	390	18	0
HIP12394	25	19	2	19	348	16	0	20	386	19	1
HIP107345	26	24	2	24	61	21	0	26	414	24	1
Total	90	69	18	70	592	58	0	77	1630	64	3
BT4											
HIP1993	34	30	7	31	100	19	2	32	956	28	1
HIP12394-4MJup	34	33	12	34	392	26	7	33	909	29	0
HIP12394-3MJup	34	21	12	25	388	12	7	26	734	13	0
HIP107345	34	27	2	29	75	18	2	30	951	30	0
Total	136	111	33	119	955	75	18	121	3550	100	1

Table 5. Comparison between the results from the logistic regression approach (Column RegL) and those from threshold detections. Two thresholds are considered: 3 and 5, and two setups are considered: ADI and ASDI. Columns #TPF reports the number of planets found whereas #C is the number of additional candidates produced. #Inj is the number of injections (hidden planets) in the corresponding blind test.

such patterns and is less efficient on BT1. It remains competitive but its performances on BT1 could certainly be improved by including injections at the learning step that are consistent with the one used for the test.

3.5 Image dedicated classifier versus a single classifier.

So far in our study, a classifier is learnt for each image and its usage is dedicated to the original image from which the exoplanets are to be detected. This has the advantage to take into account peculiarities of the image such as the weather conditions at the time it was taken. The alternative is to train a single and common classifier using the four data sets together aiming for better generalization.

The table of Figure 6 gives the result of such a single classifier (first line) trained over all data sets. The two approaches produce very similar results. The single classifier identifies one additional planet on BT3 (70 versus 69) but misses one on BT1 (22 versus 23) and tend to produce the same amount of candidates that are not injected planets. However, these results might considerably depend on the quality and uniformity of the images. We expect that a larger data set sampling a wider range of weather conditions and quality might be needed to investigate this option further.

3.6 Handling speckles with two classifiers.

A speckle moves radially away from the center between the two wavelengths depending on its distance to the center and the ratio $r = \frac{\lambda_{H3}}{\lambda_{H2}}$. One of the feature proposed tries to take advantage of this motion to detect speckles. But since the motion depends on the distance to the star, a speckle near the center might shift less than a single pixel which makes the feature irrelevant. Additionally,

speckles are not present far from the star where another regime occurs and the noise tends to be dominated by photon and instrument noise. Overall, the feature is only valid at a minimum d_{min} and maximum d_{max} distance from the star and associated to an indicator (another feature) that is set to one when the center of the stamp is located in the range $[d_{min}, d_{max}]$ (see the Annexe for details).

Another approach is to train two distinct classifiers. The first one is trained with stamps included in $[d_{min}, d_{max}]$ and includes the speckle feature whereas the second one is trained with stamps within $[0, d_{min}[\cup]d_{max}, +\infty[$ and does not use the speckle feature. The results obtained with two classifiers are shown in the last row of Figure 6 and appear to be very similar or slightly worse. Additional training samples might be required to properly train two classifiers as opposed of one.

3.7 Threshold of the logistic regression.

Logistic regression gives a probability of belonging to a class. By default, a threshold of 0.5 is used to decide whether the object belongs to the class or not. However, it is possible to choose another threshold value and thus obtain more or less samples classified as true. A classical way to determine this threshold is to use a ROC curve (receiver operating characteristic), the true positive rate against the false positive rate. But, in the case of imbalanced data, the precision recall curve is often preferred. Precision is the proportion of relevant items among all the proposed items; recall is the proportion of relevant items proposed among all the relevant items.

To determine a threshold from precision and recall, we use the f-score calculated as follows: $F\beta = (1 + \beta^2) \cdot \frac{\text{precision} \cdot \text{recall}}{(\beta^2 \cdot \text{precision}) + \text{recall}}$

Increasing the value of β increases the weight of the precision.

	BT1			BT2			BT3		
	#Inj	#TPF	#C	#Inj	#TPF	#C	#Inj	#TPF	#C
Single classifier									
4 stars	29	22	15	140	120	13	90	70	16
One classifier per image									
HD108767B	7	5	11	26	22	13	15	6	10
HIP1993	8	6	1	37	33	1	24	20	4
HIP12394	6	5	2	39	26	1	25	19	2
HIP107345	8	7	2	38	38	1	26	24	2
Total	29	23	16	140	120	16	90	69	18
Two classifiers									
4 stars	29	19	14	140	119	6	90	70	14

Figure 6. Comparing a single classifier trained over the 4 data sets to a classifier dedicated to each data set.

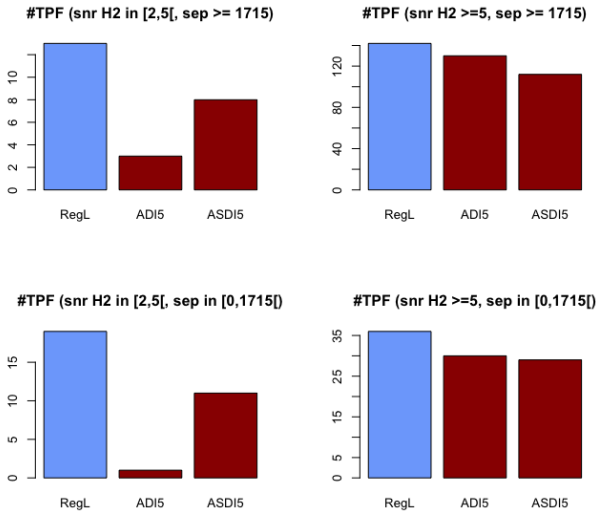


Figure 7. Number of detections (#TPF) achieved by the different methods depending on the SNR and separation from the stars of the targets.

The threshold that maximizes the f -score on the training data is calculated and then used for classification. As can be seen in figure 8, as β increases, the number of candidates proposed can increase, as well as the number of objects found. Comparing it with a threshold set at 0.5 (on the right, Figure 8), we find a larger number of objects, but also a much greater number of candidates.

The classifier can therefore be adjusted to the best compromise (number of true positives versus number of false positives) that suits the user. It depends on how much effort a user can afford to check by hand the candidates in order to increase the chance of a true detection. In the present case, we note that a significant increase of the number of candidates is required to detect only a few more exoplanets.

3.8 Amount of data required.

We evaluated the amount of training data required to reach the performance reported for the single classifier. The injection process can require a non negligible computational effort in practice. It turns out to be more costly than the learning, clustering and classification steps. To evaluate the real need for the injections, we run

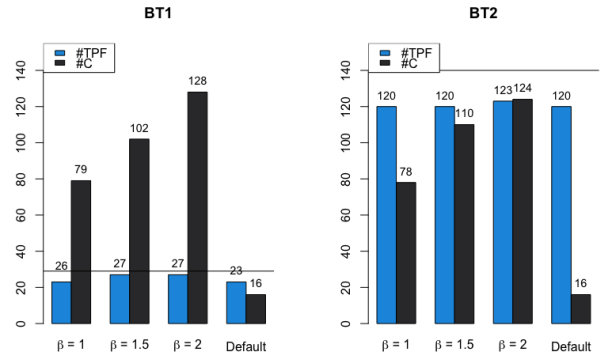


Figure 8. Number of objects found (#TPF in blue) and candidates (#C in black) according to the value of β . The black line represents the number of injections (29 for BT1 and 140 for BT2).

the training for the following numbers of injected planets: $\#pos \in \{25, 50, 100, 150, 200, 250\}$. We also investigate different number of guaranteed noise stamps with $\#neg \in \{1000, 5000, 10000, 15000\}$ to get a sense of the effect of imbalance. This analysis was restricted to BT2.

Overall the 120 planets (actual performance of our approach on BT2) are found using only 150 injected planets in the training stage (whereas 250 injections were initially used). A minimum of 10000 negative samples are needed to avoid too many false positives.

4 APPLICATION TO 51 ERI

4.1 51 Eridani

51 Eridani (HIP 21547) is an F0-type star that hosts one 2-4 M_{Jup} planet imaged in 2014 by the Gemini Planet Imager (Macintosh et al. 2015). We apply the proposed methodology to four IRDIS data sets taken with IRDIS on 25/12/2015, 15/01/2016, 11/12/2016 and 12/12/2016. The observing log and setup of these 4 observations can be found in Table. 4.1. Note that on Dec 2016, the data, obtained under good atmospheric conditions, are affected by the so called

STAR	DATE OBS	FILTER	DIT(s)×Nframe	ΔPA (°) ^a	Seeing (") ^b	Airmass ^b	τ_0 (ms) ^{a,b}	Program ID
HIP 21547	2015-12-25	DB_H23	16x256	37.6	1.18	1.10	1.8	096.C-0241(C)
HIP 21547	2016-01-15	DB_H23	16x256	41.8	1.91	1.08	1.3	096.C-0241(G)
HIP 21547	2016-12-11	DB_H23	64x54	25.3	1.97	1.12	1.5	198.C-0209(C)
HIP 21547	2016-12-12	DB_H23	64x72	45.0	0.84	1.09	5.7	198.C-0209(C)

Table 6. 51 Eri observation logs. **Notes:** ^a: DIT corresponds to the detector integration time per frame, ΔPA is the amplitude of the parallactic rotation, τ_0 corresponds to the coherence time. ^b: Values extracted from the updated DIMM info and averaged over the sequence.

25/12/2015				15/01/2016				11/12/2016				12/12/2016			
SNR H2	SNR ASDI	Found	#C	SNR H2	SNR ASDI	Found	#C	SNR H2	SNR ASDI	Found	#C	SNR H2	SNR ASDI	Found	#C
4.71	4.3	yes	3	5.28	6.6	yes	3	< 2.5	< 4	no	4	2.69	4.1	yes	2

Table 7. Results obtained on 51 Eridani.

low wind effect that occurred when the wind was very low, and considerably degraded the image quality (Milli et al. 2018)².

We consider the four data sets completely independently, as our approach is not informed of the temporality. Table 7 reports the SNR of the planet in each image with PACO ADI and PACO ASDI (column SNR), whether it was found or not by our method (yes/no of column Found) and the number of candidates proposed (column #C).

Our approach allows to find 51 Eri b in 3 out of the 4 datasets, and with an SNR lower than 5 (in the H2 band) in two cases. Moreover, a very small number of additional candidates are proposed (2 or 3). Unfortunately, none of the additional signals found by the classifier seems gravitationally bound to the star (i.e. detected at multiple epoch with a motion compatible with a bound object). We classify them as false positives.

5 CONCLUSIONS

Statistical approaches have proved very efficient to avoid self-subtraction when searching for exoplanets in ADI high contrast images, and to provide means to quantify the confidence of a detection. The SNR maps produced still contain too many artifacts related to background noise to simply identify planets with a threshold of SNR typically lower than 5σ . But planetary signals and noise also leave specific patterns and shapes within the SNR map that can help discriminating them. We have proposed a methodology using simple algorithmic techniques (edge-detection, regression and clustering) to help separating noise and planetary signals in these SNR maps. We demonstrated that the proposed methodology can considerably reduce the number of false positives and even improve detection in some cases (see for instance the case study of 51 Eridani). Moreover, it is well suited to learning with small data-sets (limited number of samples for the learning compared to current need of deep learning techniques) since it relies on dedicated and informative features of the application domain. This also helps explaining the results because the features have a meaning for the user. We now mostly intend to generalize it to spectroscopic data and test it on a larger scale.

ACKNOWLEDGEMENTS

This project is supported in part by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (COBREX; grant agreement n° 885593) as well

as the CNRS (Mission pour les Initiatives Transverses et Interdisciplinaire: MITI).

6 DATA AVAILABILITY STATEMENT

The data as well as detailed results are available on a web page³ and upon request.

REFERENCES

- Absil O., et al., 2013, *Astronomy and Astrophysics*, 559
- Allard F., Hauschildt P. H., Alexander D. R., Tamanai A., Schweitzer A., 2001, *ApJ*, 556, 357
- Beuzit J.-L., et al., 2019, *Astronomy & Astrophysics*, 631, A155
- Cantaloube F., et al., 2015, *Astronomy and Astrophysics*, 582
- Chauvin G., Lagrange A.-M., Dumas C., Zuckerman B., Mouillet D., Song I., Beuzit J.-L., Lowrance P., 2004, *Astronomy & Astrophysics*, 425, L29
- Chomez A., et al., 2023a, *A&A*, 675, A205
- Chomez A., et al., 2023b, *A&A*, 676, L10
- Delorme P., et al., 2017, in Reylé C., Di Matteo P., Herpin F., Lagadec E., Lançon A., Meliani Z., Royer F., eds, SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics. p. Di ([arXiv:1712.06948](https://arxiv.org/abs/1712.06948))
- Desidera S., et al., 2021, *A&A*, 651, A70
- Dohlen K., et al., 2008, in McLean I. S., Casali M. M., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 7014, Ground-based and Airborne Instrumentation for Astronomy II. p. 70143L, [doi:10.1117/12.789786](https://doi.org/10.1117/12.789786)
- Duda R. O., Hart P. E., 1973, Pattern classification and scene analysis. A Wiley-Interscience publication, Wiley, <https://www.worldcat.org/oclc/00388788>
- Flasseur, Olivier Denis, Loïc Thiébaud, Éric Langlois, Maud 2020, *A&A*, 637, A9
- Flasseur O., Denis L., Thiébaud É., Langlois M., 2018, *Astronomy & Astrophysics*, 618, A138
- Flasseur O., Bodrito T., Mairal J., Ponce J., Langlois M., Lagrange A.-M., 2023, *MNRAS*,
- Gebhard T. D., Bonse M. J., Quanz S. P., Schölkopf B., 2020, arXiv preprint arXiv:2010.05591
- Gomez Gonzalez C. A., Absil O., Absil P.-A., Van Droogenbroeck M., Mawet D., Surdej J., 2016, *Astronomy and Astrophysics*, 589, A54
- Gomez Gonzalez C. A., Absil O., Van Droogenbroeck M., 2018, *Astronomy and Astrophysics*, 613, A71
- Lagrange A.-M., et al., 2009, *Astronomy & Astrophysics*, 493, L21
- Lagrange A.-M., et al., 2010, *Science*, 329, 57

² this effect has been taken care of afterwards

³ <https://pagesperso.g-scop.grenoble-inp.fr/~catussen/exoplanet/report/>

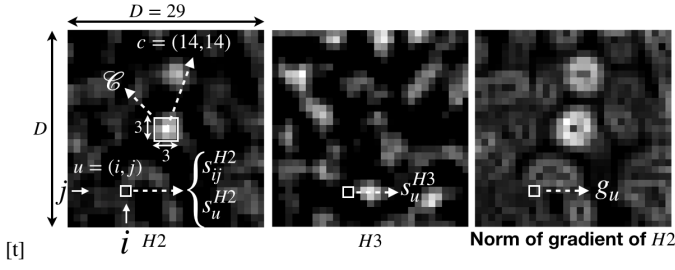


Figure A1. Summary of the notations for a stamp.

- Macintosh B., et al., 2014, Proceedings of the National Academy of Sciences, 111, 12661
- Macintosh B., et al., 2015, *Science*, 350, 64
- Marois C., Lafreniere D., Doyon R., Macintosh B., Nadeau D., 2005, *Astrophysical Journal*, 641
- Marois C., Macintosh B., Barman T., Zuckerman B., Song I., Patience J., Lafreniere D., Doyon R., 2008, *Science*, 322
- Marois C., Correia C., Véran J.-P., Currie T., 2013, Proceedings of the International Astronomical Union, 8, 48–49
- Marois C., Correia C., Véran J.-P., Currie T., 2014, in Booth M., Matthews B. C., Graham J. R., eds, Vol. 299, Exploring the Formation and Evolution of Planetary Systems. pp 48–49, doi:10.1017/S17439213130007813
- Milli J., et al., 2018, in Close L. M., Schreiber L., Schmidt D., eds, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series Vol. 10703, Adaptive Optics Systems VI. p. 107032A (arXiv:1806.05370), doi:10.1117/12.2311499
- Müller D., 2011, Modern hierarchical, agglomerative clustering algorithms (arXiv:1109.2378)
- Nielsen E. L., et al., 2019, *AJ*, 158, 13
- Soummer R., Pueyo L., Larkin J., 2012, *The Astrophysical Journal*, 755, L28
- Vigan A., et al., 2021, *A&A*, 651, A72
- Yip K. H., et al., 2020, in Brefeld U., Fromont E., Hotho A., Knobbe A., Maathuis M., Robardet C., eds, Machine Learning and Knowledge Discovery in Databases. Springer International Publishing, Cham, pp 322–338

APPENDIX A: FEATURES

In the following we denote the maximum, mean and standard deviation over a set of reals numbers $X \subset \mathbb{R}$ by: $\max(X)$, $\mu(X)$ and $\sigma(X)$.

A stamp S is a small sub-image restricted to a $D \times D$ area. This sub-image gives two $D \times D$ matrices of pixels or SNR values extracted from the original image in the two wavelengths $H2$ and $H3$. In practice, we experimented with stamps of size 29×29 and 19×19 ($D = 29$ and $D = 19$). For sake of simplicity, we assume D is always odd so that the center of the stamp is a pixel (and not in between two pixels). To make it easier to generalize D to other contexts, we recall that we were working with 12,270 milli arc-seconds per pixel and two wavelengths at $\lambda_1 = 1.593 \mu\text{m}$ and $\lambda_2 = 1.667 \mu\text{m}$. Consider pixel $u = (i, j)$ in the stamp S ($u \in S$), we denote by:

- s_u^{H2} (resp. s_u^{H3}): the SNR value of pixel u in wavelength $H2$ (resp $H3$).
- g_u : the norm of the SNR gradient at pixel u in wavelength $H2$.
- c : the pixel at the center of the stamp *i.e* $c = (\lfloor D/2 \rfloor, \lfloor D/2 \rfloor)$.
- C : the set of nine pixels in a 3×3 area at the center of the stamp *i.e* $C = \{(i, j) | i \in [\lfloor D/2 \rfloor - 1, \lfloor D/2 \rfloor + 1], j \in [\lfloor D/2 \rfloor - 1, \lfloor D/2 \rfloor + 1]\}$

Note that we refer to pixel $u = (i, j)$ using its i and j coordinates

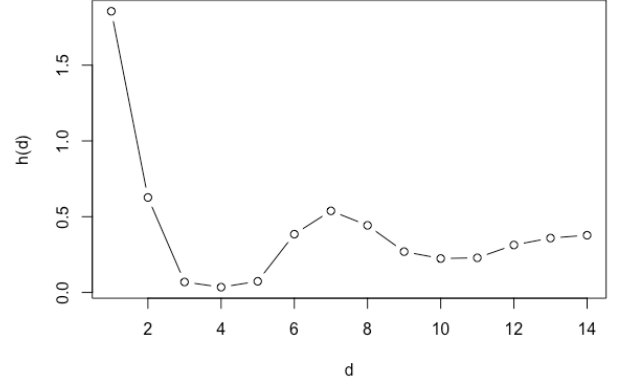


Figure A2. $h_c(d)$ function for the stamp presented figure A1. The first dark ring can be seen around distance 4 whereas the first white ring is around 7.

when needed so that s_u^{H2} can also be written s_{ij}^{H2} . Figure A1 gives a summary of the notations.

The features used are the following:

- (i) (**MeanSnr**). Mean SNR value in $H2$ and $H3$:

$$f_1^{H2} = \mu(\{s_u^{H2} | u \in S\}), \quad f_1^{H3} = \mu(\{s_u^{H3} | u \in S\})$$

- (ii) (**MaxCenteredSnr**). Max SNR value in $H2$ and $H3$ in the center C :

$$f_2^{H2} = \max(\{s_u^{H2} | u \in C\}), \quad f_2^{H3} = \max(\{s_u^{H3} | u \in C\})$$

- (iii) (**MaxGra, MeanGra, StdevGra**). Maximum, Mean and Standard deviation of the gradient in $H2$:

$$f_3 = \max(\{g_u | u \in S\})$$

$$f_4 = \mu(\{g_u | u \in S\})$$

$$f_5 = \sigma(\{g_u | u \in S\})$$

- (iv) (**MaxMin**). Consider the image defined for each pixel u by the minimum SNR value between both wavelengths: $\min(s_u^{H2}, s_u^{H3})$. This *minimum* image only keep the SNR quantities present in both wavelengths. The idea is that a speckle systematically move between $H2$ and $H3$ whereas real companions **can be** present in both $H2$ and $H3$. We use the following feature:

$$f_6 = \max(\{\min(s_u^{H2}, s_u^{H3}) | u \in S\})$$

- (v) (**AiryFig**). To capture the presence of an Airy figure, we rely on an azimuthal mean of SNR values. More precisely, let's denote $h_u(d)$ the mean SNR values in a ring of one pixel's width located at distance d of pixel u . Figure A2 shows $h_c(d)$ *i.e* from the center c of our example stamp.

An Airy figure at the center c of the stamp is expected to have a higher standard deviation $\sigma(c) = \sigma(\{h_c(d) | d \in [\lfloor 1, D/2 \rfloor]\})$ than random noise. This however assumes that the stamp is precisely centered. We therefore relax this constraint and take the best value among the nine central pixels. We use the following feature:

$$f_7 = \max(\{\sigma(u) | u \in C\})$$

Feature f_6 is computed for each wavelength which gives, in practice, two distinct features f_6^{H2} and f_6^{H3} .

- (vi) (**MeanSpec**). A speckle moves radially away from the center (the star) of the image depending on its distance to the center and

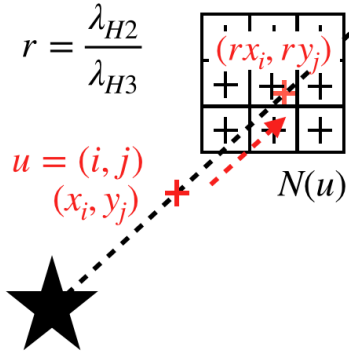


Figure A3. Motion of a speckle from H2 to H3

the ratio $r = \frac{\lambda_{H3}}{\lambda_{H2}} = \frac{1.667}{1.593}$. Assuming the star is the origin of the coordinate system, if pixel $u = (i, j)$ in $H2$ is part of a speckle and its center is located at coordinates (x_i, y_j) in the star system, the point $ru = (rx_i, ry_j)$ in $H3$ is expected to have the same SNR intensity. Since (rx_i, ry_j) does not necessarily match exactly the center of another pixel, we compute a weighted average of its four neighbors. Let $N(u)$ be the nine closest pixels of the real point $ru = (rx_i, ry_j)$, we compute $w(u)$ (where $d'(u, v) = \frac{1}{d(u, v)}$ is the inverse of the distance between pixel u and v):

$$w(u) = \frac{\sum_{v \in N(u)} d'(ru, v) \times s_v^{H3}}{\sum_{v \in N(u)} d'(ru, v)}$$

The following quantity $sp(u)$ is used to quantify how much the SNR intensity s_u^{H2} is found at (rx_i, ry_j) in $H3$ i.e is close to $w(u)$:

$$sp(u) = \frac{|s_u^{H2} - w(u)|}{\max(s_u^{H2}, w(u))}$$

Finally the feature considers the 9 central pixels to identify a speckle:

$$f_8 = \mu(\{sp(u) | u \in C\})$$

The previous characterization of a speckle is only valid at minimum and maximum distance from the star. We therefore introduce an indicative feature:

$$f_9 = \begin{cases} 1 & \text{if } d_{min} \leq d(c, star) \leq d_{max} \\ 0 & \text{otherwise} \end{cases}$$

We used the values (in arc-seconds) of $d_{min} = 0.2''$ and $d_{max} = 1.7''$.

(vii) (**MeanNoise, StdevNoise**). At greater distances from the star ($> d_{max}$) a companion object can sometimes disappear from $H2$ to $H3$. The quantity $n(u)$ measures how much of the intensity in $H2$ at pixel u disappears in $H3$:

$$n(u) = \max(0, \frac{s_u^{H2} - s_u^{H3}}{\max(s_u^{H2}, s_u^{H3})})$$

We use the mean and the standard deviation of this quantity, at the center of the stamp, as features:

$$f_{10} = \mu(\{n(u) | u \in C\}), \quad f_{11} = \sigma(\{n(u) | u \in C\})$$

(viii) (**Dist**). The distance of the center of the stamp to the star is also included in the features and denoted f_{12} .

APPENDIX B: ANNEXE

Figures B1, B2, B3 present the results directly on the SNR maps. The exoplanets injected are located as little green squares, the results

(candidates proposed) of PACO ASDI 5 are shown as little blue circles whereas the results of our approach regL are displayed as large red circles. For instance, exoplanets found only by PACO ASDI 5 therefore appear as a little square within a little circle. Reversely, a square within a large circle is a target identified by regL. Some targets are not found by any of the methods.

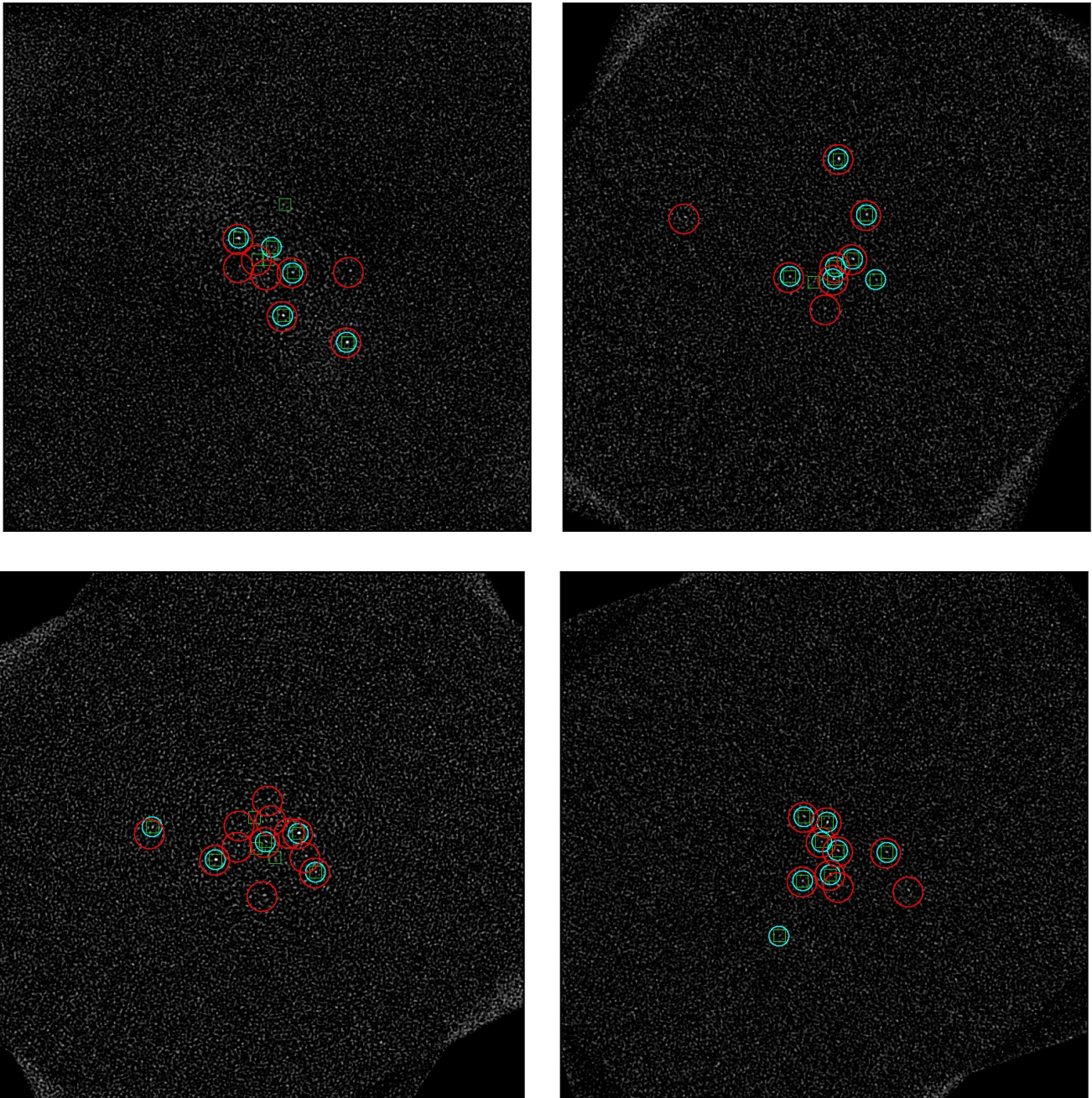


Figure B1. Visualization of the results on BT1. An injected exoplanet is displayed by a green square, a candidate proposed by PACO ASDI 5 is a little blue circle, a candidate proposed by regL is a large red circle. Top left: HD108767B, Top right: HIP1993, Bottom left: HIP12394, Bottom right: HIP107345.

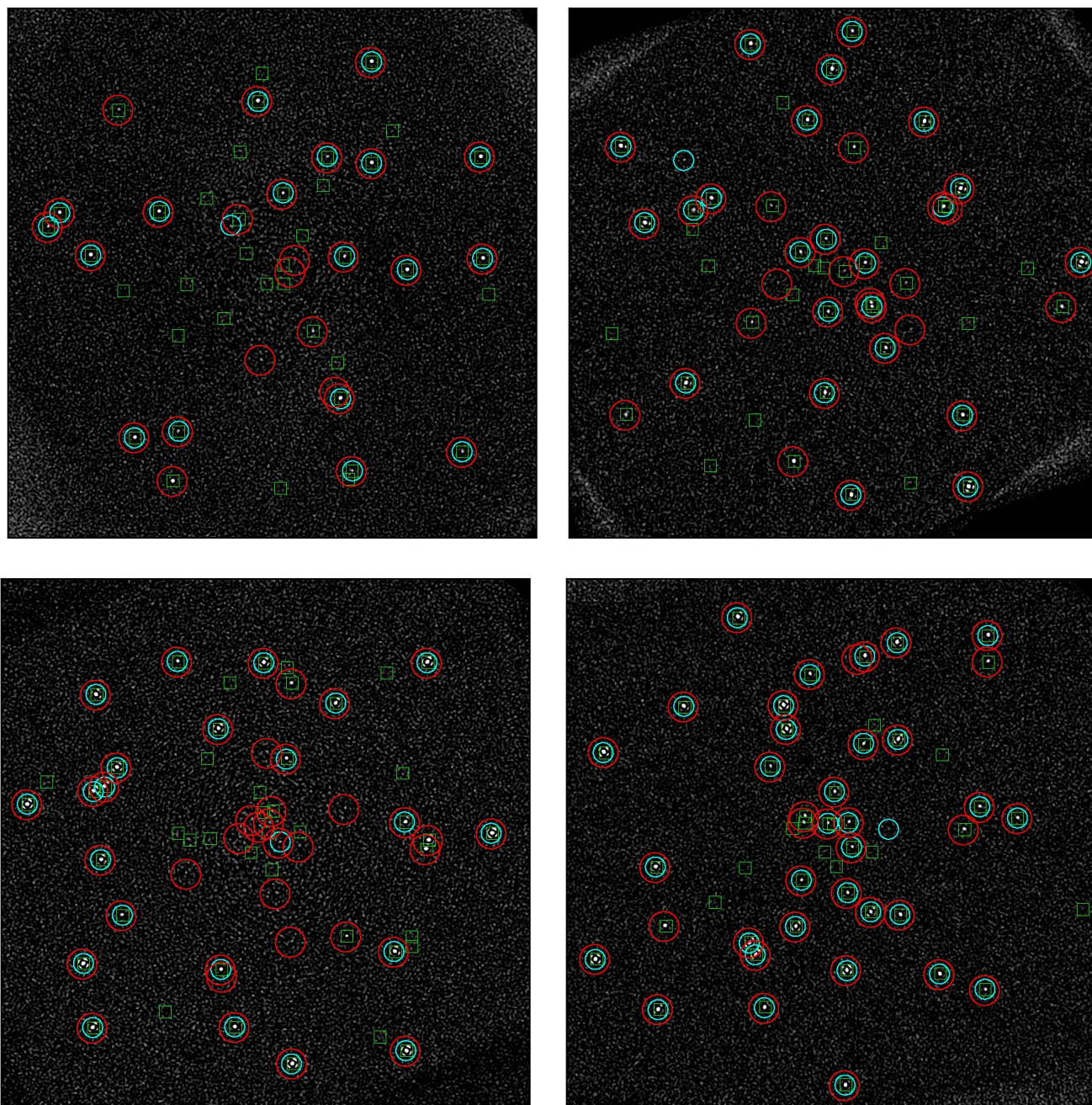


Figure B2. Visualization of the results on **BT2**. An injected exoplanet is displayed by a green square, a candidate proposed by PACO ASDI 5 is a little blue circle, a candidate proposed by regL is a large red circle. Top left: HD108767B, Top right: HIP1993, Bottom left: HIP12394, Bottom right: HIP107345.

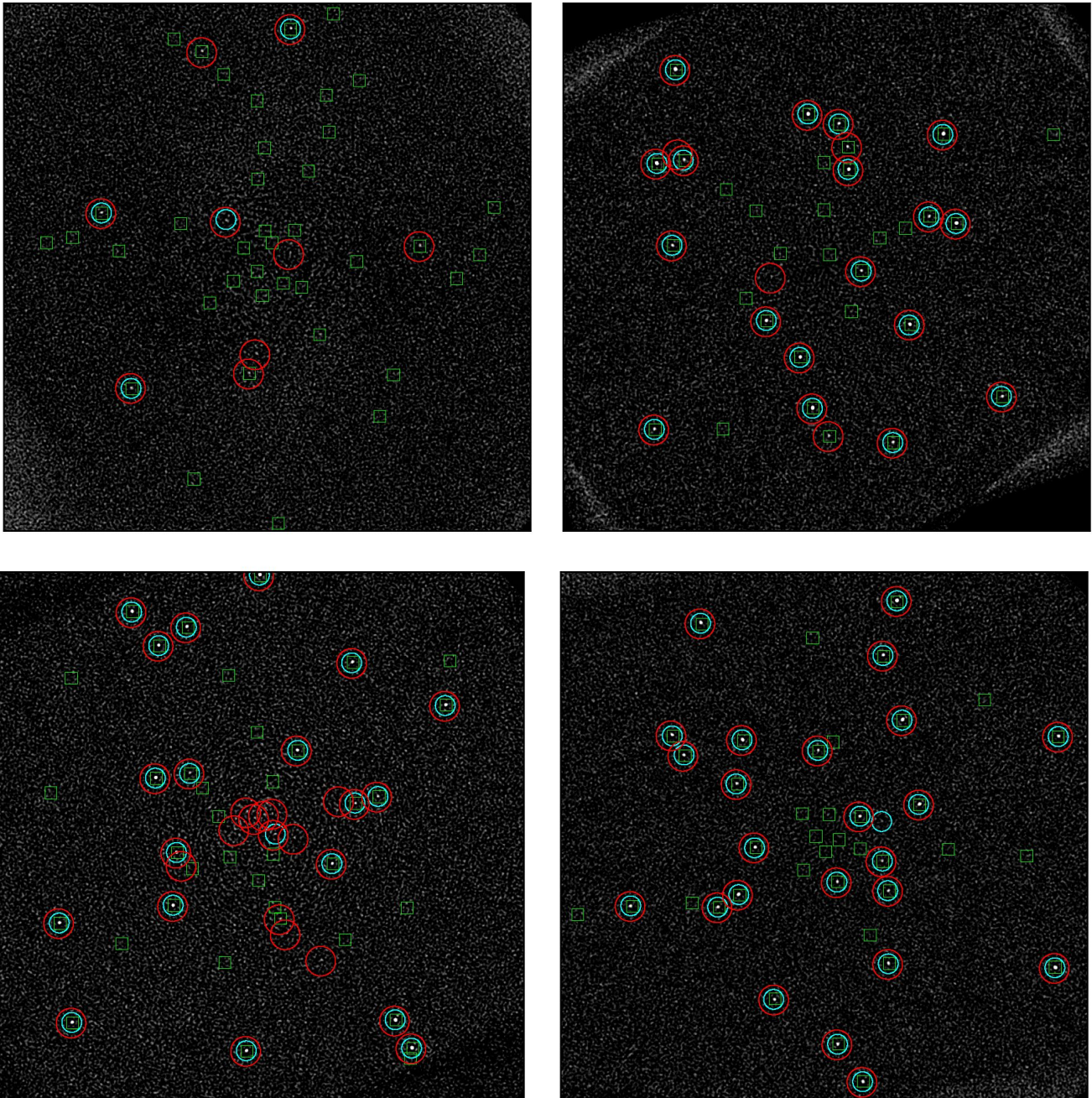


Figure B3. Visualization of the results on **BT3**. An injected exoplanet is displayed by a green square, a candidate proposed by PACO ASDI 5 is a little blue circle, a candidate proposed by regL is a large red circle. Top left: HD108767B, Top right: HIP1993, Bottom left: HIP12394, Bottom right: HIP107345.