PROCEEDINGS OF SPIE

Exoplanet imaging data challenge: benchmarking the various image processing methods for exoplanet detection

Cantalloube, F., Gomez-Gonzalez, C., Absil, O., Cantero, C., Bacher, R., et al.

SPIE.

Event: SPIE Astronomical Telescopes + Instrumentation, 2020, Online Only

Exoplanet Imaging Data Challenge: benchmarking the various image processing methods for exoplanet detection

Cantalloube F.[a], Gomez-Gonzalez C.[b], Absil O.[c], Cantero C.[c,d], Bacher R.[e], Bonse J. M.[f], Bottom M.[g], Dahlqvist C.-H.[c], Desgrange C.[a,h], Flasseur O.[i], Fuhrmann T.[j], Henning Th.[a], Jensen-Clem R.[k], Kenworthy M.[l], Mawet D.[m], Mesa D.[n], Meshkat T.[m], Mouillet D.[e], Müller A.[a], Nasedkin E.[a], Pairet B.[o], Pierard S.[d], Ruffio J.-B.[m], Samland M.[p], Stone J.[q], and Van Droogenbroeck M.[d]

[a]Max-Planck-Institut für Astronomie, Königstuhl 17, Heidelberg 69117, Germany
[b]Barcelona Supercomputing Center, Spain
[c]Space Sciences, Technologies & Astrophysics Research (STAR) Institute, Université de Liège, Allée du Six Août 19c, B-4000 Liège, Belgium
[d]Montefiore Institute, Université de Liège, 4000 Liège, Belgium
[e]Université Grenoble Alpes, CNRS, IPAG, 38000, Grenoble, France
[f]Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang-Pauli-Strasse 27, 8093, Zurich, Switzerland
[g]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, United States
[h]Department of Physics, Ecole Normale Supérieure de Lyon, 69364, Lyon, France
[i]Université de Lyon, Université Lyon1, ENS de Lyon, CNRS, Centre de Recherche Astrophysique de Lyon UMR 5574, F-69230, Saint-Genis-Laval, France
[j]Ostbayerische Technische Hochschule, Regensburg, Germany
[k]Astronomy Department, University of California Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA
[l]Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands
[m]Department of Astronomy, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA
[n]INAF-Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, I-35122 Padova, Italy
[o]ISPGroup, ELEN/ICTEAM, UCLouvain, B-1348 Louvain-la-Neuve, Belgium
[p]Department of Astronomy, Stockholm University, Stockholm, Sweden
[q]Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721-0065, USA

ABSTRACT

The *Exoplanet Imaging Data Challenge* is a community-wide effort meant to offer a platform for a fair and common comparison of image processing methods designed for exoplanet direct detection. For this purpose, it gathers on a dedicated repository (Zenodo), data from several high-contrast ground-based instruments worldwide in which we injected synthetic planetary signals. The data challenge is hosted on the CodaLab competition platform, where participants can upload their results. The specifications of the data challenge are published on our website https://exoplanet-imaging-challenge.github.io/. The first phase, launched on the 1st of September 2019 and closed on the 1st of October 2020, consisted in detecting point sources in two types of common data-set in the field of high-contrast imaging: data taken in pupil-tracking mode at one wavelength

Further author information: (Send correspondence to F.C)
F.C.: E-mail: cantalloube@mpia.de

(subchallenge 1, also referred to as ADI) and multispectral data taken in pupil-tracking mode (subchallenge 2, also referred to as ADI+mSDI). In this paper, we describe the approach, organisational lessons-learnt and current limitations of the data challenge, as well as preliminary results of the participants' submissions for this first phase. In the future, we plan to provide permanent access to the standard library of data sets and metrics, in order to guide the validation and support the publications of innovative image processing algorithms dedicated to high-contrast imaging of planetary systems.

Keywords: Exoplanet detection; High-contrast imaging; Adaptive Optics; Coronagraphy; Post-processing techniques; Data challenge

1. INTRODUCTION

As of today, there exist a large number of algorithms to detect exoplanets within high-contrast images. During the last few years, this field of research was very active in order to reach fainter signals and accurately recover their properties, leading to a few publications per year. However, in each publication of a new algorithm, the authors choose to demonstrate their method on a different data set and using different metrics compared to previous publications. In order to avoid confusion in the community, and to offer a catalog of existing methods adapted to various science cases (e.g. large surveys or in-depth characterisation), we propose to homogenize the comparison between methods through the *exoplanet imaging data challenge* (EIDC). The EIDC offers a common ensemble of representative data sets from different high-contrast instruments, as well as a common set of metrics to compare and validate current and new methods.

This data challenge is the outcome of a collaboration between several institutes worldwide and is an open-source community effort, allowing anyone to contribute. The underlying idea is to share knowledge and spark new collaborations to make the best use of the data we acquired from various telescopes. The ensemble of data sets are hosted on a *Zenodo*[*] open-access repository. The participants are invited to submit their results via the *CodaLab*[†] competition platform. Upon reception, CodaLab computes various metrics that we have defined and calculates a score publicly displayed on a leader-board. At last, relevant information about the exoplanet imaging data challenge is gathered on a dedicated website https://exoplanet-imaging-challenge.github.io/.

We communicated about this data challenge via various media (mailing-lists, social networks, instrument consortia and advertisement during conferences). The first phase was officially launched on the 1st of September 2019, with a preliminary deadline on the 30th of January 2020. We discussed the outcome of this preliminary phase during a dedicated workshop, called *Post-processing for high-contrast imaging of exoplanets and circumstellar disks*, which took place at the end of January 2020 at the Max Planck society Harnack-haus (Berlin, Germany). During this workshop we discussed the feedback from participants and the relevance of the metrics used. At the end of the workshop we decided to form a working group (contact: exoimg.datachallenge@gmail.com), to add more information about the data sets on the website and to reveal the results from one single data set (https://exoplanet-imaging-challenge.github.io/preresults/). The latter proved useful for several reasons, in particular: (1) verifying the convention used when running the algorithms on the data, (2) making sure that the intention of the EIDC is not to reduce the methods to a single grade, but to broadly discuss the variety of existing algorithms and (3) encouraging other teams to participate to the data challenge. Following this, we extended the deadline for the first phase to the 1st of October 2020.

In the following, we first describe the data sets provided to the participants via the Zenodo repository (Sect. 2) and then describe the evaluation procedure used on the results collected during the first phase of the EIDC (Sect. 3). We then discuss the submissions we received and the rankings obtained when applying our evaluation procedure (Sect. 4). At last, we propose a path-forward for the continuation of the EIDC and for leveraging the limitations noted during this first phase.

---

[*]https://zenodo.org/record/3361544#.X8pxK9so9GE
[†]https://competitions.codalab.org/competitions/20693#participate

2. DATA SETS

Several typical high-contrast data sets are available on the EIDC *Zenodo* repository. Participants must submit the results of a single algorithm applied to the complete ensemble of data sets described in Sect. 2.1. It is possible for participants to provide several submissions, each resulting from one single algorithm applied to the ensemble of data sets. However, the *CodaLab* competition platform only displays one result per participant (selected by the participant) on the public leader-board. We therefore warn users to not rely on the results shown on the *CodaLab* EIDC competition web-page. In each image cube, we injected synthetic planetary signals following the procedure described in Sect. 2.2.

2.1 Pre-reduced data provided to the participants

The performance of an image processing algorithm can be highly dependent upon the type of data through the various temporal and spatial distribution of the starlight residuals (generally referred to as speckles) they can exhibit. To mitigate this effect and to discuss which method is best suited to which type of data, we used data from various high-contrast instruments installed on 8-m class telescopes around the world. In addition, the observing conditions shape the type of starlight residuals prevailing in the images. Again, to mitigate this effect, we selected several data sets from the same instrument, but taken under various observing conditions. In this section, we describe the data selected for both the ADI subchallenge (Sect. 2.1.1) and the ADI+mSDI subchallenge (Sect. 2.1.2). The data have been kindly provided by collaborators from the instruments consortia.

2.1.1 Subchallenge 1: angular diversity

The most common observing mode to search for exoplanet candidates is to use the pupil tracking mode of the telescope (providing the telescope has an alt-az mount). In this mode the circumstellar signals rotate at a deterministic speed (following the parallactic angles), while the optical aberrations, responsible for the starlight residuals observed in the field of view, are stabilized. This angular diversity makes it possible to use image processing techniques based on Angular Differential Imaging[1] (ADI).

In this first subchallenge, we provided temporal image cubes from three different instruments: VLT/SPHERE-IRDIS,[2,3] Keck/NIRC2,[4] and LBT/LMIRCam.[5] For each instrument, three representative data sets are provided, making a total of nine data sets. The VLT/SPHERE instrument is a second generation instrument equipped with an extreme adaptive optics[6] (AO) and an apodized Lyot coronagraph[7] (APLC), feeding light into three subsystems, including the IRDIS dual band imager,[8] working from the Y2-band (1.02 μm) to the K2-band (2.25 μm). The Keck/NIRC2 instrument is equipped with an adaptive optics system[9] and an Annular Groove Phase Mask (AGPM) vector vortex coronagraph,[10] optimized for observations in the Lp-band[11] (3.78 μm). The LBT/LMIRCam instrument is equipped with an extreme adaptive optics system[12] and the data for the EIDC are taken without coronagraph, in the Lp-band (3.78 μm).

Each data set is constituted of four files (all given in *.fits* format[13]): (1) the image cube, (2) the corresponding parallactic angles variation corrected from true North, (3) the pixel scale of the detector and (4) a non-coronagraphic or non-saturated point spread function (PSF) of the instrument, taken either before or after the observing sequence, used to calibrate the detection in terms of contrast and which can be used as a model of a planetary signal (e.g. for forward modelling or supervised machine learning methods). The images provided are pre-reduced in terms of bad pixels removal, dark current and background subtraction, flat fielding and normalization of the flux for the PSF. We assume that the users do not perform further frame selection and that all the images are exploited. The VLT/SPHERE-IRDIS data were pre-reduced with the SPHERE Data Center,[14] using the SPHERE Data Reduction and Handling pipeline.[15] The Keck/NIRC2 data were pre-reduced by the dedicated pre-processing pipeline for AGPM images[16‡]. The LBT/LMIRCam data were pre-reduced by the pipeline developed for the LEECH exoplanet survey.[17] The properties of the nine image cubes are summarized in Tab. 1.

---

‡https://github.com/vortex-exoplanet/NIRC2_Preprocessing

Table 1. Description of the nine data sets provided for the EIDC ADI subchallenge. D_{tel} is the effective telescope diameter (downstream the coronagraph when any); λ_{obs} is the observation wavelength; Plate-scale is the pixel scale of the detector used; N_{img} is the size of each image frame; N_t is the number of frames along the temporal axis; Δ_{field} is the total field rotation of circumstellar objects (or parallactic angle variation).

Instrument	ID	D_{tel} [m]	λ_{obs} [μm]	Plate-scale [mas/pixel]	N_{img} [px] $\times$ [px]	N_t	Δ_{field} [°]
VLT/SPHERE-IRDIS	sph1	7.87	1.625 ± 0.29	12.255	160×160	252	40.3
	sph2	7.87	1.593 ± 0.052	12.255	160×160	80	31.5
	sph3	7.87	1.593 ± 0.052	12.255	160×160	228	80.5
Keck/NIRC2	nrc1	8.72	3.776 ± 0.70	20.00	321×321	29	53.0
	nrc2	8.72	3.776 ± 0.70	20.00	321×321	40	37.3
	nrc3	8.72	3.776 ± 0.70	20.20	321×321	50	166.9
LBT/LMIRCam	lmr1	8.40	3.780 ± 0.10	10.70	200×200	4838	153.4
	lmr2	8.40	3.780 ± 0.10	10.70	200×200	3219	60.6
	lmr3	8.40	3.780 ± 0.10	10.70	200×200	4620	91.0

2.1.2 Subchallenge 2: angular and spectral diversity

The second generation of instrument dedicated to exoplanet imaging, such as VLT/SPHERE and Gemini-S/GPI, are equipped with integral field units (IFU), with relatively low spectral resolving power ($R < 100$), producing images taken simultaneously at different wavelengths. In the images, the starlight residuals move radially with the wavelength, while the circumstellar signals remain fix in the field of view whatever the wavelength. This additional spectral diversity makes it possible to use image processing techniques based on simultaneous multiple Spectral Differential Imaging[18,19] (mSDI). In practice the observations are carried out in pupil tracking mode to combine ADI and mSDI techniques.

In this second subchallenge, we provided 4D image cubes from two different instruments: VLT/SPHERE-IFS,[2,20] and Gemini-S/GPI.[21] For each instrument, five representative data sets are provided, making a total of ten data sets. The VLT/SPHERE-IFS has two wavelength ranges, either covering the YJH bands (0.95 to 1.64 μm) or the YJ bands (0.95 to 1.33 μm), with spectral resolving power of 30 and 50 respectively. Gemini-S/GPI is similarly equipped with an extreme AO system and an APLC, and works from the J-band (0.95 μm) to the K-band (2.12 μm) with a resolving power of up to 80 in the latter case.

Each multispectral data set is constituted of five files (again given in *.fits* format): The same four files as for the ADI data sets, plus a vector containing the wavelengths corresponding to each spectral channel of the IFU. Again, the provided images are pre-reduced. The VLT/SPHERE-IFS data were pre-reduced with the SPHERE Data Center,[14] using the SPHERE Data Reduction and Handling pipeline.[15] The Gemini-S/GPI data were pre-reduced with the GPI Data Cruncher,[22] making use of the GPI reduction pipeline.[23] The properties of the ten image cubes are summarized in Tab. 2.

2.2 Injected planetary signals

Injection procedure Within each data cube we injected from 0 to 5 point sources using the VIP package.[24] To avoid interfering with potential real planetary signals or extended sources in the images, we injected the point sources using the opposite parallactic angles. The latter procedure keeps the temporal statistical behavior of the starlight residuals, while smearing out the potential circumstellar signals. We randomly picked separations from the inner working angle of the coronagraph to the end of the AO correction zone, at a random position angle, making sure that the injected signals will not overlap. For a given position, we randomly picked a contrast value in a range of $\pm 3\,\sigma$ from the $5\,\sigma$ contrast curve provided when running the chosen baseline algorithm, which is a regular annular Principal Component Analysis[25,26] (PCA), using the VIP package implementation. In the 9 ADI subchallenge data sets, we injected a total of 20 planetary signals, including two data sets without injection. In the 10 ADI+mSDI subchallenge data sets, we injected a total of 23 planetary signals, including two data sets without injection.

Table 2. Description of the ten data sets provided for the EIDC ADI+mSDI subchallenge. D_{tel} is the effective telescope diameter (downstream the coronagraph); λ_{range} is the range of observation wavelength; Plate-scale is the pixel scale of the detector used; N_{img} is the size of each image frame; N_t is the number of frames along the temporal axis; N_λ is the number of channels along the spectral axis; Δ_{field} is the total field rotation of circumstellar objects.

Instrument	ID	D_{tel} [m]	λ_{range} [μm]	Plate-scale [mas/pixel]	N_{img} [px] $\times$ [px]	N_t	N_λ	Δ_{field} [°]
VLT/SPHERE-IFS	ifs1	7.87	$0.957 - 1.329$	7.46	200×200	128	39	25.6
	ifs2	7.87	$0.957 - 1.636$	7.46	200×200	112	39	131.9
	ifs3	7.87	$0.957 - 1.636$	7.46	200×200	109	39	43.5
	ifs4	7.87	$0.957 - 1.636$	7.46	200×200	112	39	46.2
	ifs5	7.87	$0.957 - 1.329$	7.46	200×200	80	39	31.8
Gemini-S/GPI	gpi1	7.57	$1.495 - 1.797$	14.17	161×161	35	37	54.7
	gpi2	7.57	$1.495 - 1.797$	14.17	161×161	41	37	22.5
	gpi3	7.57	$1.495 - 1.797$	14.17	161×161	38	37	12.6
	gpi4	7.57	$1.495 - 1.797$	14.17	161×161	37	37	13.8
	gpi5	7.57	$1.495 - 1.797$	14.17	161×161	44	37	24.4

Limitations of the injection procedure The injected synthetic planetary signals are modeled with the non-coronagraphic (or unsaturated) PSF used for calibration, which is included in each data set. This injection procedure is therefore limited: (i) the flux of the synthetic planetary signal is constant throughout the observing sequence, (ii) the center of the image (position of the host star) is fixed when injecting the signals, that is to say potential mis-centering of the star throughout the sequence is not taken into account, and (iii) we do not take into account the azimuthal smearing of the signal that could appear for long exposure times and/or at large angular separation from the star. For the injection in the multispectral data (ADI+mSDI subchallenge), we used the trend from typical L-type and T-type brown dwarf spectra obtained from the SpexPrism[27] data base.

3. EVALUATION PROCEDURE

The first phase of the exoplanet imaging data challenge is primarily focused on the point source detection capabilities of the classical and latest image processing methods dedicated to high-contrast imaging. We therefore decided to analyze the submitted material as a classification problem, that is to say in terms of counting detections and non-detections. The EIDC participants were requested to submit their results by including three types of files (in *.fits* format): (1) the detection maps for each data set, (2) one threshold common to all the data set, and optionally (3), the full width at half maximum (FWHM) of the expected planetary signal in the detection map. The results of the first phase of the EIDC are available upon request if one wants to apply different evaluation tools to the detection maps. In the future, we plan to apply more advanced metrics that take into consideration aspects such as the working nature of each algorithm, the difficulty of a given detection (e.g. in terms of contrast and separation to the star) and a proper ranking procedure (Piérard et al., in prep.).

With this data challenge, we do not intend to reduce the performance of a given algorithm to a single number, which would be pointless since the efficiency of each method is highly dependent on the type of instrument and observing conditions. Instead, we aim at providing the community with homogeneous information about the different algorithm capabilities so that they can build their own opinion based on what type of application they need. In this section, we propose to evaluate the results by using a set of metrics, with known short-comings, in order to guide the comparison, from both a qualitative and quantitative standpoint.

3.1 Detection considerations

Based on the planetary signal injections in the nine (resp. ten) data sets available for the ADI subchallenge (resp. ADI+mSDI subchallenge), we counted the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) at the detection threshold provided by the participants. We then repeated this procedure for a range of thresholds discussed below. A TP is defined as a detection within one resolution

element from the position of an injected companion, while a FP is a detection at any other location. The FN and TN regroups respectively the non-detections at the position of injected companions and the non-detections at any other location considering the FWHM aperture grid applied to the image.

Any signal above the chosen threshold is considered as a detection: we do not take into account neither visual detections (detections that are obvious visually but below the threshold) nor any type of post-sorting (signals that are above threshold but obviously artefacts from a visual inspection or automatic rejection as in Ref.[28]). In the future, we will alleviate this limitation by asking the participants for their detection parameters (e.g. position, signal-to-noise ratio or probability of presence).

For the non-detections, we counted the number of possible detections in the image and subtract the number of detections. We decided not to use the FWHM submitted by participants since it could (dis)favor some algorithms. Instead, we used the instrumental PSF FWHM (or resolution element) computed from the observation wavelength, the telescope effective diameter and the detector pixel scale. In practice, a planetary signal should always be of about this size, as none of the submitted algorithm perform any type of deconvolution.

In addition, each algorithm has its own inner and outer working angle (smallest and largest angular separation processed). To be fair, we applied to all the detection maps received for a given data set, a unique binary mask to select the smallest inner working angle and largest outer working angle submitted for this data set (see the images galleries in Appendix, bottom-right).

In order to guarantee that the results are not biased by the person analyzing the data, and to minimize the errors, we analyzed the data in double blind. Two authors evaluated the data without knowing their reference, and analyzed the results with their own tools to make sure that the results are consistent.

3.2 Metrics

From these counts, we compute different metrics:

1. the true positive rate (also called sensitivity or recall): $TPR = TP/(TP + FN)$,

2. the false positive rate (also called fall-out): $FPR = FP/(FP + TN)$,

3. the false discovery rate (also called precision): $FDR = FP/(FP + TP)$,

4. the F1-score (harmonic mean of precision and sensitivity): F1-score $= 2 \cdot TP/(2 \cdot TP + FP + FN)$.

True positive rate, TPR: If the number of injection is non-null, the TPR as a function of the threshold is a decreasing step function. An ideal algorithm should provide a TPR of 1 whatever the threshold (up to the maximum value of the detection map). We additionally compute the area under the curve (AUC) of the TPR as a function of various threshold values, which must be as close as possible to one.

False positive rate, FPR: On the contrary, the FPR as a function of the threshold is decreasing monotonically. An ideal algorithm should provide a FPR of 0 whatever the threshold. This quantity characterizes the residual noise in the detection map. At a threshold of zero, this FPR indicates the fraction of resolution elements above zero in the detection map. For a threshold value corresponding to the minimum of the detection map, the FPR is equal to one.

False discovery rate, FDR: The FDR as a function of the threshold is a decreasing step function. An ideal algorithm should also provide a FDR of 0 whatever the threshold. This figure of merit gives information about the detections, regardless of the true negatives, which is more important for our science case as we want to detect planetary signals (and we are not interested in knowing how we cannot detect them). We also compute the AUC of the FDR as a function of various threshold values, which must be as close as possible to zero.

As a result, we chose to use three figures of merit: (1) the F1-score computed at the submitted threshold, (2) the AUC of the TPR (must be close to 1), and (3) the AUC of the FDR (must be close to 0). The three scores have values between 0 and 1. If no synthetic planetary signals have been injected into the data set, these three scores are undefined. We computed these scores for each data set separately, and averaged them for each instrument, then finally averaged them for all the data sets of a given subchallenge, as gathered in Tab. 3, Tab. 4, Tab. 5 and Tab. 6. We ranked each submission according to these three scores, not only to verify the relevance of the F1-score, but also to unveil more information about the performance of each algorithm (Fig. 5 and Fig. 8).
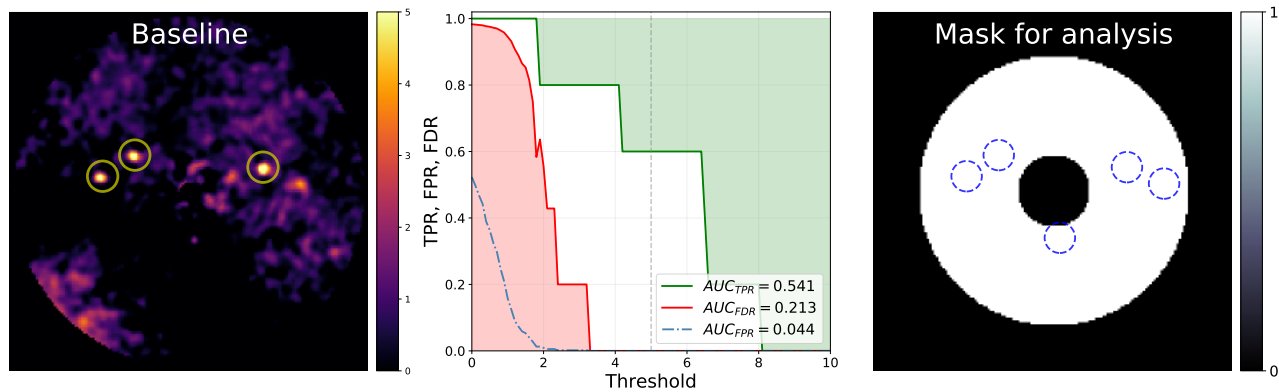
Figure 1. Example of the results displayed for the chosen baseline algorithm applied on the third VLT/SPHERE-IRDIS data set (sph3). Left: Detection map with a color-bar ranging from 0 to the submitted threshold. True positives at the submitted threshold are indicated with yellow circles (the diameter of the circles correspond to the instrumental PSF FWHM). False positives at the submitted threshold are indicated with red squares. Middle: True Positive Rate (TPR, in green), False Discovery Rate (FDR, in red) and False Positive Rate (FPR, dash-dotted blue line), for a range of threshold varying from 0 to twice the submitted threshold. Right: Common mask applied to all the detection maps of this specific sph3 data set. The dashed blue circles indicate the positions of the 5 synthetic planetary signals injected in this data set.

3.3 Graphical comparison

To support the scores mentioned above, we also provide with qualitative and graphical comparisons. First, the detection map shows visually the type of residuals (not quantified in the current paper), as well as the inner and outer working angles of each submitted algorithm (see Fig. 6, left). As mentioned previously, we apply a common binary mask to all the detection maps for counting the detections and non-detections within the same region for all the algorithms (see Fig. 6, right). In addition, we plot the TPR, FDR and FPR as a function of a given range of thresholds to visualize the trend of these three quantities (see Fig. 6, middle). We arbitrarily chose a range of threshold from 0 to twice the submitted threshold. However for some type of algorithms (e.g. the ones providing close to a binary detection map), this range of threshold can (dis)favor the AUC of the TPR. We therefore separated the algorithms based on their underlying concept and compared them with each other.

4. SUBMISSIONS RESULTS

4.1 Subchallenge 1: ADI

For this challenge, we received 22 valid submissions from 12 participants. We splitted these submissions into three main families on which they are based: (1) speckle subtraction techniques, (2) inverse problem approaches, and (3) supervised machine learning techniques. In the following, we show the detailed plots for the third VLT/SPHERE-IRDIS data set (sph3), which contains 5 injected planetary signals (indicated with blue circles in Fig. 6, right). The 22 submitted detection maps obtained for the nine data sets of this subchallenge can be found in App. A.

4.1.1 Speckle subtraction techniques

We received 12 submissions based on speckle subtraction techniques, which is the most commonly used technique, consisting of three steps: (i) estimating the stellar light residuals (or speckle field) directly from the image cube, (ii) subtracting the model(s) of the stellar light residuals to each frame of the cube and (iii) de-rotating the images to align the circumstellar signals with the true North and combine all these images to produce a so-called residual map. In addition to or instead of this last step, some techniques build detection maps where the pixel values correspond to a probability of presence of a planetary signal.

cADI$_{SpeCal}$ stands for classical Angular Differential Imaging, as implemented in the SpeCal pipeline[29] dedicated to process VLT/SPHERE images. LOCI stands for Locally Optimized Combination of Images,[30] and is the version implemented in the VIP package.[24] TLOCI$_{SpeCal}$ is an improved version of LOCI called Template

LOCI,[31] as implemented in the SpeCal pipeline. PCA_{SpeCal} is a Principal Component Analysis (PCA[25, 26]), as implemented in the SpeCal pipeline, using a default value of 5 principal components. PCA_{MPIA} is a full-frame PCA used for the VLT/NACO ISPY[32] survey. PCA_{Padova}[33] is an improved PCA used for the VLT/SPHERE SHINE[34] survey and also included in the SpeCal pipeline. Unknown stands for a reference-less submission we received. $STIM_{fullframe}$ is the Standardized Trajectory Intensity Mean (STIM) detection map[35] computed from a fullframe PCA subtraction. $STIM_{annuli}$ is the STIM detection map computed from an annular PCA subtraction. $STIM_{hpf}$ is the STIM detection map obtained from a fullframe PCA subtraction applied on previously high-pass filtered images (using a filtering fraction of 0.25). SLIMask[36] for STIM Largest Intensity Mask, is an improvement of the STIM detection map where several PCA with different number of PCs are ran to produce a mask and discard unlikely signals. RSM[37§], standing for Regime-Switching Model detection map, traces the deviation form a noise to a noise+planet regime in previously speckle-subtracted images with various speckle subtraction techniques, using in this case annular PCA, Non-negative Matrix Factorization (NMF[38]), and low rank plus sparse decomposition (LLSG[39]).

For these post-processing methods, the images and corresponding counts of detections and non-detections as a function of the threshold are displayed in Fig. 2 for the case of the third VLT/SPHERE-IRDIS data set (sph3). The scores proposed in Sect. 3 computed on the 9 data sets from this ADI subchallenge are gathered in Tab. 3. The final ranking is shown in Fig. 5, in red for the residual maps and in orange for the detection maps.

4.1.2 Inverse problem techniques:

We received 5 submissions based on inverse problem approaches, which consist in modeling the expected planetary signal and tracking it in the image cube. By construction, this method estimates the contrast of the potential planetary signal, at any position in the field of view, via a maximum likelihood estimation, itself relying upon a specific noise distribution. The provided detection map is the signal-to-noise ratio map computed as the estimated contrast divided by the uncertainty on this estimated contrast.

ANDROMEDA[28] first performs a pair-wise image subtraction before tracking the planetary signal under the assumption that the residual noise (after the subtraction) is white and Gaussian. ANDROMEDA is included in the SpeCal pipeline and implemented in the VIP package. FMMF[40¶] performs a PCA subtraction before matching a forward-modeled planetary template to the data, assuming that the residual noise is white and Gaussian. FMMF is included in the pyKLIP pipeline[41] and has been used to process the GPI survey demographic analysis data (GPIES[42]). PACO[43] does not perform any subtraction prior to tracking the planetary signal, and assumes that the noise is a multivariate Gaussian noise, self-calibrated on the data, accounting for the spatial covariances of the speckles at the scale of small patches. pyPACO is the python version of PACO, currently under implementation in the VIP package. At last, TRAP[44‖] estimates both the starlight residuals and the planetary signal along the temporal axis (contrary to other methods modeling the signal for each frame) and models the starlight residuals as principal components derived from specific areas that do not include the planetary signal but share similar noise statistical properties.

The images and corresponding counts of detections and non-detections as a function of the threshold are displayed in Fig. 3 for the case of the third VLT/SPHERE-IRDIS data set (sph3). The scores proposed in Sect. 3, computed on the 9 data sets from this ADI subchallenge, are gathered in Tab. 4. The ranking of the different methods based on inverse problem is shown in Fig. 5 (blue).

4.1.3 Supervised machine learning techniques:

We received 5 submissions based on supervised machine learning techniques. SODIRF and SODINN[45] are based on the random forest technique and neural networks, respectively. Both algorithms are trained on hundreds of thousands of two class Multi-level Low-rank Approximation Residual (MLAR) samples. MLAR samples are square patch sequences of twice the instrument PSF FWHM size, centered on each pixel of the ADI sequence. Each of these patches on a MLAR sequence corresponds to a low-rank level (PCA-principal components). The full MLAR sequence shares the temporal dimension of the ADI sequence. Therefore, one MLAR class represents

---

§https://github.com/chdahlqvist/RSMmap

¶https://bitbucket.org/pyKLIP/pyklip/

‖https://github.com/m-samland/trap

the evolution of an exoplanet light source along the principal component space and the other class the evolution of residual noise. Once trained, SODIRF and SODINN generate a normalized detection map in which each pixel has a probability to belong to one of these classes.

$SODIRF_{original}$[**] is the original random forest algorithm published in Ref.,[45] using the same set of hyper-parameters. $SODIRF_{adapted}$ is a modified version in which the optimal hyper-parameter values have been found for each particular dataset by means of applying a grid search and cross-validation with three folds. For the case of neural networks, $SODINN_{3D}$ and $SODINN_{LSTM}$ are the original versions presented in Ref.,[45] and use respectively a 3D convolutional layers and long short term memory (LSTM) layers. $SODINN_{BiLSTM}$ preserves the full architecture of $SODINN_{LSTM}$ except that it uses bidirectional-LSTM layers instead.

The images and corresponding counts of detections and non-detections as a function of the threshold are displayed in Fig. 4 for the case of the third VLT/SPHERE-IRDIS data set (sph3). The scores proposed in Sect. 3, computed on the 9 data sets from this ADI subchallenge, are gathered in Tab. 5. The ranking of the different methods based on supervised machine learning is shown in Fig. 5 (green).

4.1.4 Ranking the ADI subchallenge results

Based on the three scores defined in Sect. 3, we compute the F1-score, the AUC of the TPR and AUC of the FDR for all data sets. The different rankings for these three metrics and the five families of methods described are shown in Fig. 5. Based on the ADI subchallenge results, we can already draw a few conclusions.

(1) The three rankings based on each of the three scores give consistent results: for a given method, a high F1-score indeed corresponds to a high AUC of the TPR and low AUC of the FDR. However, it is possible that the AUC of the FDR differs a lot from other scores. For instance when some algorithms are more robust to false alarms (high AUC of the FDR) but are not very sensitive to faint signals (low AUC of the TPR). Depending on the science case (such as deep search for a point source or homogeneous processing of a large sample of data) one may use the most relevant metric to choose some methods over the others. For further analysis, the results of the data challenge are available upon request to the first author, in agreement with the EIDC collaboration.

(2) As expected, we observe that globally the speckle subtraction based techniques (in red in Fig. 5) do not perform as good as more advanced techniques: the most recent techniques indeed show better detection capabilities than former techniques.

(3) The exception goes to the supervised machine learning techniques (in green in Fig. 5) for which we observe numerous false positives. Applying supervised machine learning techniques to the challenging field of high-contrast imaging is very recent and investigations are on-going to improve this type of methods. Another bias in the current ranking of these techniques stems from the type of detection map they deliver. The detection maps are a binary classifier, normalized to one (see detection maps Fig. 4, top row) and the TPR is abruptly dropping to zero for thresholds greater than the submitted threshold. Therefore calculating the AUC of the TPF for a range of threshold from 0 to twice the submitted threshold does not make sense and this specific score disfavor these methods.

(4) Figure 5 also highlights that the performance are indeed dependent on the type of instrument. Indeed coronagraphic images (light and medium colors) do not show the same type of stellar light residuals as non-coronagraphic images (dark colors). In addition, the wavelength and therefore the exposure time (number of frames) do play a role as some algorithms are working along the temporal axis. For instance, it is expected from Ref.[44] that TRAP will not perform as good on temporally binned data, which is the case for the Keck/NIRC2 data sets (medium color).

(5) At first glance we can see that the latest algorithms, RSM and SLIMask for the speckle subtraction techniques, and FMMF, PACO and TRAP for the inverse problem approaches, are providing with the best performance along with the three scores used here. Note that in this phase of the EIDC we did not account for practical aspects such as the running time or the number of user-parameters to fine-tune. In future phases of the EIDC, we intend to take more factors into account to better guide users towards understanding better the performance of the various methods dedicated to exoplanet detection in high-contrast images.

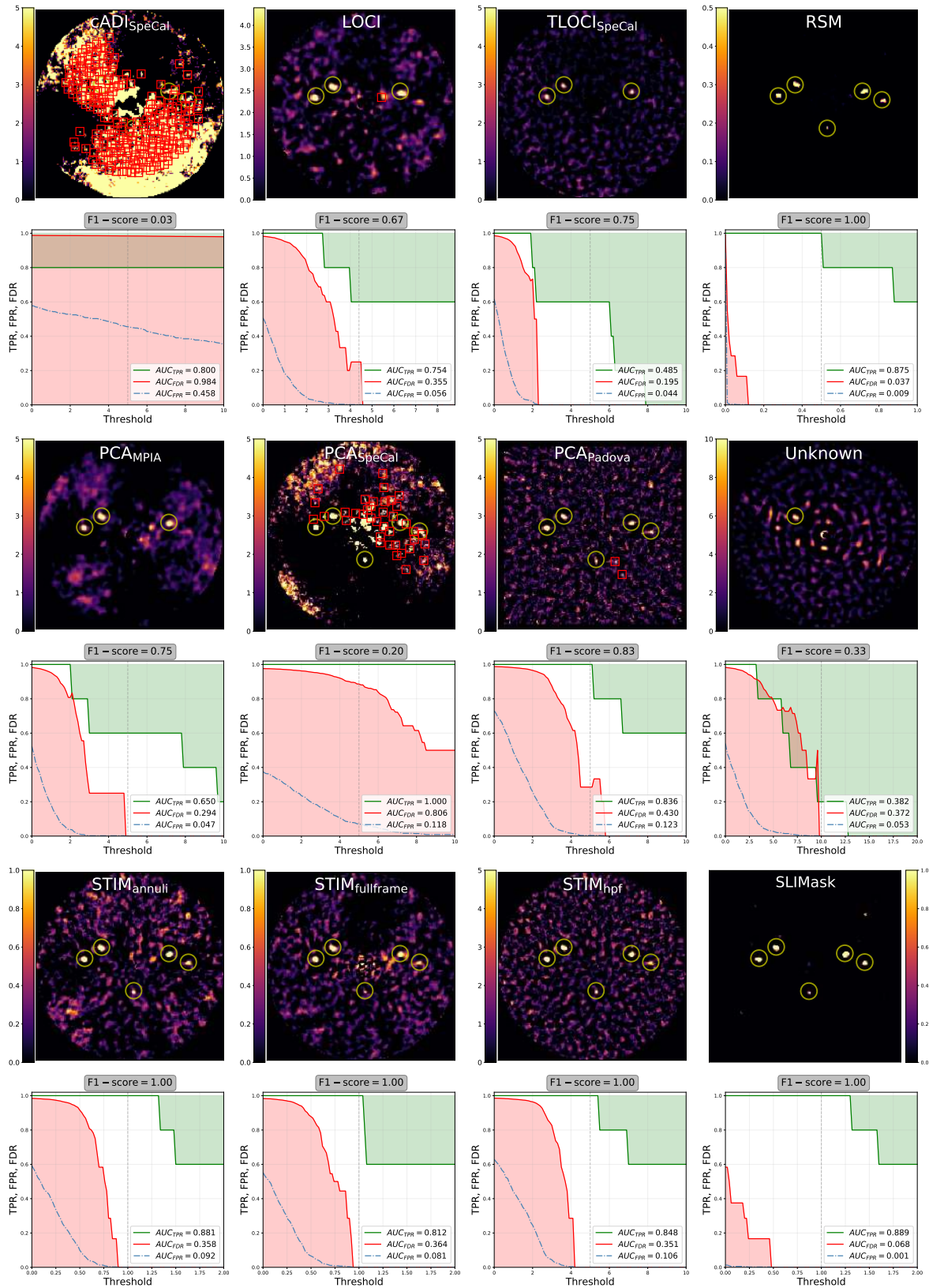[**]https://github.com/carlgogo/sodinn

Figure 2. Results of the ADI subchallenge for the sph3 VLT/SPHERE-IRDIS dataset: speckle subtraction techniques.
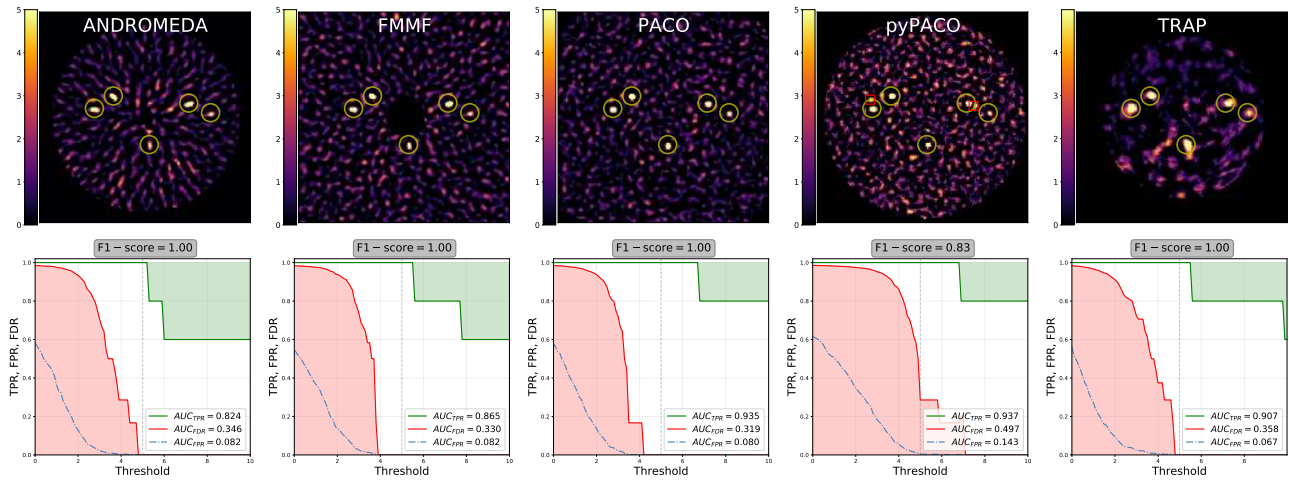
Figure 3. Results of the ADI subchallenge for the sph3 VLT/SPHERE-IRDIS dataset: inverse problem approaches.
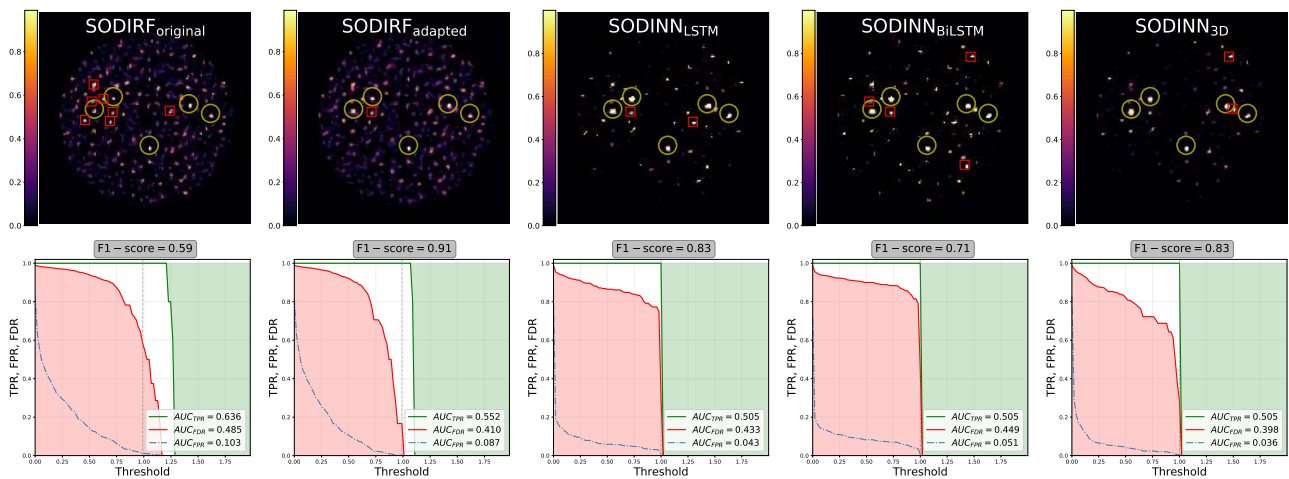

Figure 4. Results of the ADI subchallenge for the sph3 VLT/SPHERE-IRDIS dataset: supervised machine learning.

Table 3. Results from the first subchallenge (ADI) of the EIDC: speckle subtraction based techniques. For each algorithm and each data set, the F1-score computed at the submitted threshold is in black (1st line), the AUC of the TPR is in green (2nd line) and the AUC of the FDR is in red (3rd line). When there are no injected planetary signals, the scores cannot be computed and are therefore undefined (indicated with the letter u). The last column is the mean of the scores for the seven data sets that include injected planetary signal(s).

Algorithm	VLT/SPHERE-IRDIS				Keck/NIRC2				LBT/LMIRCam				All
	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	
Baseline	1.00	u	0.75	0.88	0.00	0.00	u	0.00	0.67	0.50	0.00	0.39	0.42
	0.52	u	0.54	0.53	0.35	0.24	u	0.29	0.50	0.49	0.17	0.38	0.40
	0.42	u	0.21	0.32	0.29	0.54	u	0.41	0.32	0.49	0.31	0.37	0.37
cADI$_\mathrm{SpeCal}$	0.00	u	0.03	0.02	0.00	0.00	u	0.00	0.00	0.00	0.00	0.00	0.01
	0.00	u	0.80	0.40	0.01	0.01	u	0.01	0.00	0.00	0.01	0.01	0.14
	1.00	u	0.98	0.99	0.01	0.01	u	0.01	0.01	0.01	0.01	0.01	0.34
LOCI	1.00	u	0.67	0.83	0.86	0.50	u	0.68	0.67	0.50	0.80	0.66	0.72
	0.74	u	0.75	0.75	0.48	0.30	u	0.39	0.65	0.55	0.60	0.60	0.58
	0.42	u	0.35	0.39	0.36	0.43	u	0.40	0.39	0.43	0.36	0.39	0.39
TLOCI$_\mathrm{SpeCal}$	0.00	u	0.75	0.38	0.00	0.00	u	0.00	0.00	0.00	0.00	0.00	0.12
	0.39	u	0.49	0.44	0.21	0.08	u	0.15	0.00	0.25	0.00	0.08	0.22
	0.27	u	0.20	0.23	0.37	0.39	u	0.38	0.82	0.98	0.79	0.86	0.49
PCA$_\mathrm{SpeCal}$	0.00	u	0.20	0.10	0.00	0.00	u	0.00	0.00	0.00	0.00	0.00	0.03
	0.00	u	1.00	0.50	0.01	0.01	u	0.01	0.00	0.00	0.01	0.00	0.17
	1.00	u	0.81	0.90	0.01	0.01	u	0.01	0.01	0.01	0.01	0.01	0.31
PCA$_\mathrm{MPIA}$	1.00	u	0.75	0.88	0.67	0.00	u	0.33	0.67	0.80	0.50	0.66	0.62
	0.68	u	0.65	0.67	0.41	0.22	u	0.31	0.50	0.71	0.45	0.55	0.51
	0.31	u	0.29	0.30	0.32	0.37	u	0.34	0.33	0.42	0.32	0.36	0.33
PCA$_\mathrm{Padova}$	0.50	u	0.83	0.67	0.00	0.00	u	0.00	0.40	0.50	0.22	0.37	0.35
	0.66	u	0.84	0.75	0.16	0.17	u	0.17	0.66	0.53	0.54	0.58	0.50
	0.48	u	0.43	0.46	0.22	0.25	u	0.23	0.64	0.50	0.56	0.57	0.42
Unknown	0.00	u	0.33	0.17	0.00	0.00	u	0.00	0.00	0.00	0.00	0.00	0.06
	0.00	u	0.38	0.19	0.15	0.14	u	0.15	0.00	0.12	0.07	0.06	0.13
	1.00	u	0.37	0.69	0.60	0.37	u	0.48	0.39	0.38	0.43	0.40	0.52
STIM$_\mathrm{fullframe}$	1.00	u	1.00	1.00	0.00	0.50	u	0.25	0.00	0.67	0.50	0.39	0.55
	0.79	u	0.88	0.84	0.36	0.19	u	0.28	0.23	0.48	0.62	0.44	0.52
	0.40	u	0.36	0.38	0.44	0.39	u	0.41	0.45	0.42	0.40	0.42	0.41
STIM$_\mathrm{annuli}$	0.67	u	1.00	0.83	0.00	0.00	u	0.00	0.67	0.67	0.67	0.67	0.50
	0.84	u	0.81	0.83	0.21	0.22	u	0.21	0.26	0.68	0.67	0.54	0.53
	0.49	u	0.36	0.43	0.52	0.38	u	0.45	0.37	0.42	0.43	0.40	0.43
STIM$_\mathrm{hpf}$	0.67	u	1.00	0.83	0.00	0.50	u	0.25	0.00	0.00	0.50	0.17	0.42
	0.81	u	0.85	0.83	0.23	0.21	u	0.22	0.33	0.35	0.43	0.37	0.47
	0.48	u	0.35	0.42	0.49	0.41	u	0.45	0.39	0.50	0.30	0.40	0.42
SLIMask	1.00	u	1.00	1.00	0.67	0.50	u	0.58	0.67	0.80	1.00	0.82	0.80
	0.88	u	0.89	0.89	0.39	0.27	u	0.33	0.67	0.64	0.71	0.67	0.63
	0.37	u	0.07	0.22	0.34	0.40	u	0.37	0.36	0.36	0.31	0.34	0.31
RSM	1.00	u	1.00	1.00	0.86	0.50	u	0.68	0.67	1.00	1.00	0.89	0.86
	0.84	u	0.88	0.86	0.55	0.21	u	0.38	0.45	0.62	0.73	0.60	0.61
	0.09	u	0.04	0.06	0.06	0.02	u	0.04	0.14	0.29	0.08	0.17	0.09

Table 4. Results from the first subchallenge (ADI) of the EIDC: inverse problem based techniques.

Algorithm	VLT/SPHERE-IRDIS				Keck/NIRC2				LBT/LMIRCam				All
	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	
ANDROMEDA	1.00	u	1.00	1.00	0.00	0.50	u	0.25	0.00	1.00	0.00	0.33	0.53
	0.53	u	0.82	0.68	0.30	0.40	u	0.35	0.18	0.63	0.21	0.34	0.46
	0.38	u	0.35	0.36	0.36	0.39	u	0.38	0.38	0.39	0.38	0.38	0.37
FMMF	0.67	u	1.00	0.83	0.67	0.50	u	0.58	0.67	0.80	1.00	0.82	0.75
	1.00	u	0.87	0.93	0.51	0.27	u	0.39	0.65	0.80	0.85	0.77	0.70
	0.43	u	0.33	0.38	0.39	0.39	u	0.39	0.43	0.43	0.33	0.40	0.39
PACO	1.00	u	1.00	1.00	0.60	0.40	u	0.50	0.67	0.57	0.06	0.43	0.64
	1.00	u	0.93	0.97	0.44	0.34	u	0.39	0.29	0.58	0.51	0.46	0.61
	0.39	u	0.32	0.36	0.43	0.48	u	0.45	0.30	0.51	0.90	0.57	0.46
pyPACO	0.08	u	0.83	0.46	0.00	0.00	u	0.00	0.07	0.09	0.12	0.09	0.18
	1.00	u	0.94	0.97	0.23	0.10	u	0.16	0.78	1.00	1.00	0.93	0.69
	0.70	u	0.50	0.60	0.40	0.69	u	0.55	0.91	0.75	0.68	0.78	0.64
TRAP	1.00	u	1.00	1.00	0.00	0.50	u	0.25	0.67	0.80	0.80	0.76	0.67
	0.68	u	0.91	0.80	0.33	0.33	u	0.33	0.50	0.70	0.61	0.60	0.58
	0.37	u	0.36	0.36	0.32	0.35	u	0.33	0.32	0.45	0.33	0.37	0.35

Table 5. Results from the first subchallenge (ADI) of the EIDC: supervised machine learning based techniques.

| Algorithm	VLT/SPHERE-IRDIS				Keck/NIRC2				LBT/LMIRCam				All
	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	(1)	(2)	(3)	Mean	
sodirf$_{\text{original}}$	0.40	u	0.59	0.49	0.22	0.00	u	0.11	0.18	0.33	0.86	0.46	0.35
	0.70	u	0.64	0.67	0.36	0.10	u	0.23	0.28	0.58	0.52	0.46	0.45
	0.55	u	0.48	0.52	0.66	0.62	u	0.64	0.52	0.53	0.47	0.50	0.55
sodirf$_{\text{adapted}}$	0.33	u	0.91	0.62	0.22	0.00	u	0.11	0.33	0.80	0.86	0.66	0.47
	0.71	u	0.55	0.63	0.42	0.22	u	0.32	0.30	0.57	0.54	0.47	0.48
	0.56	u	0.41	0.48	0.64	0.62	u	0.63	0.51	0.44	0.45	0.47	0.53
sodinn$_{\text{LSTM}}$	0.67	u	0.83	0.75	0.33	0.00	u	0.17	0.33	0.57	0.55	0.48	0.47
	0.50	u	0.50	0.50	0.19	0.16	u	0.17	0.50	0.50	0.50	0.50	0.39
	0.47	u	0.43	0.45	0.43	0.34	u	0.38	0.47	0.47	0.48	0.48	0.44
sodinn$_{\text{BiLSTM}}$	0.40	u	0.71	0.56	0.00	0.50	u	0.25	0.11	0.33	0.20	0.21	0.34
	0.50	u	0.50	0.50	0.08	0.17	u	0.13	0.50	0.50	0.50	0.50	0.38
	0.48	u	0.45	0.47	0.45	0.22	u	0.34	0.49	0.49	0.49	0.49	0.43
sodinn$_{\text{3D}}$	0.29	u	0.83	0.56	0.25	0.29	u	0.27	0.50	1.00	0.86	0.79	0.54
	0.84	u	0.50	0.67	0.29	0.35	u	0.32	0.47	0.50	0.50	0.49	0.50
	0.65	u	0.40	0.53	0.70	0.68	u	0.69	0.48	0.46	0.48	0.47	0.56
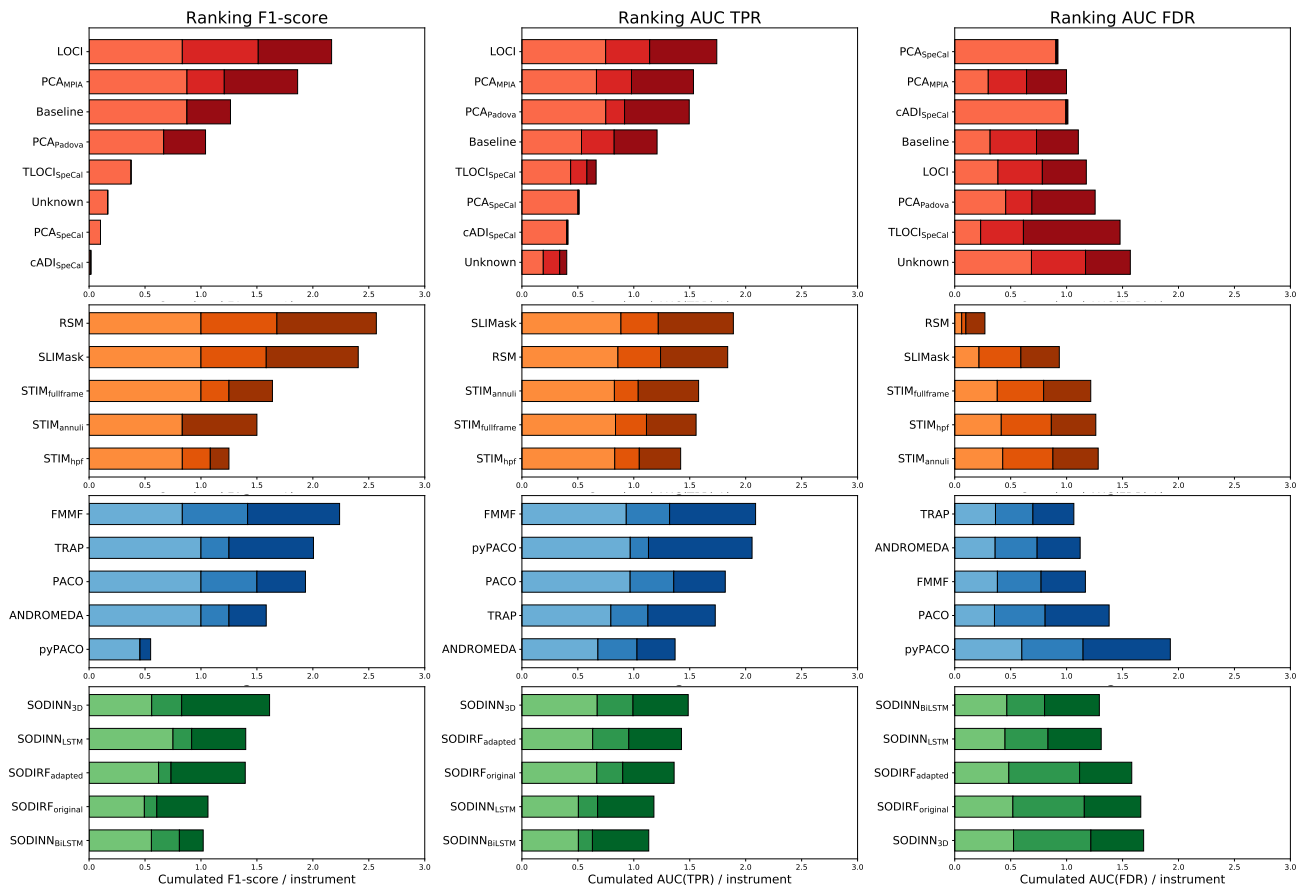
Figure 5. Ranking of the different families of submitted algorithms for the EIDC ADI subchallenge. Left column: Ranking based on the F1-score. Middle column: Ranking based on the AUC of the TPR. Right column: Ranking based on the AUC of the FDR. From top to bottom, the 5 different families of algorithms: classical speckle subtraction providing residual maps (red), advanced speckle subtraction building detection maps (orange), inverse problems (blue) and supervised machine learning (green). The light, medium and dark colors correspond to the three VLT/SPHERE-IRDIS, Keck/NIRC2, and LBT/LMIRCam data sets respectively.

4.2 Subchallenge 2: ADI+mSDI

For this challenge, we received 6 valid submissions from 4 participants. Two algorithms belong to the speckle subtraction family (PCA_{Padova} and $STIM_{ADI}$). The other four methods belong to the inverse problem approach family ($ANDROMEDA_{ADI}$, $ANDROMEDA_{ASDI}$, $FMMF_{ASDI}$, and $PACO_{ASDI}$). In the following, we show the detailed plots for the fifth Gemini-S/GPI data set (gpi5), which contains 4 injected planetary signals (indicated with blue circles in Fig. 6, right). On this data set, the result of the baseline algorithm is shown in Fig. 6 (detection map on the left and analysis in the middle). The 6 submitted detection maps obtained for the ten data sets of this subchallenge can be found in App. B.
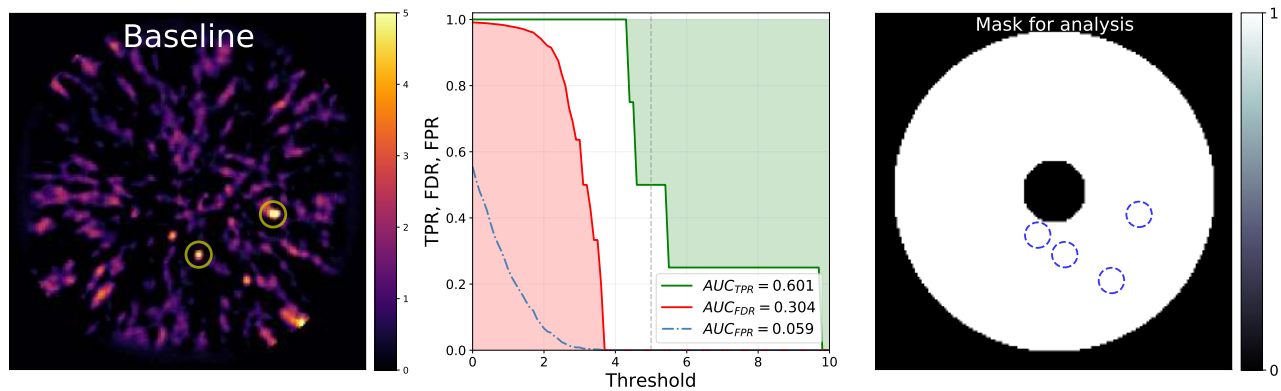
Figure 6. Example of the results displayed for the chosen baseline algorithm (annular PCA) applied on the fifth Gemini-S/GPI data set (gpi5). Left: Detection map with a color-bar ranging from 0 to the submitted threshold. True positives at the submitted threshold are indicated with yellow circles (the diameter of the circles correspond to the instrumental PSF FWHM). False positives at the submitted threshold are indicated with red squares. Middle: True Positive Rate (TPR, in green), False Discovery Rate (FDR, in red) and False Positive Rate (FPR, dash-dotted blue line), for a range of threshold varying from 0 to twice the submitted threshold. Right: Common mask applied to all the detection maps of this specific gpi5 data set. The dashed blue circles indicate the positions of the 4 synthetic planetary signals injected in this data set.

PCA_{Padova} builds the library of speckle field models by applying a principal component analysis on the whole 4D cube. This method is implemented in the SpeCal pipeline and used for analysing the SPHERE guaranteed time observations data. $STIM_{ADI}$, consists in building a STIM detection map for each spectral channel (using the high-pass filtered version and a full frame PCA) and then combining the detection maps of each spectral channel by summing the squared images. $ANDROMEDA_{ADI}$, consists in running ANDROMEDA on each spectral channel and then combining the signal-to-noise ratio maps of each channel by summing the squared images. $ANDROMEDA_{SADI}$, performs a simple spectral differential imaging (SDI) of the data cube before running ANDROMEDA on each spectral channel and then combining the signal-to-noise ratio maps of each channel by summing the squared images. The reference channels to be rescaled and subtracted (used as a SDI starlight residuals model) are chosen according to typical expected exoplanet spectrum. $FMMF_{SADI}$ consists in taking into account the spectral dimension included in FMMF and weights each channel with a typical substellar spectral template, as described in Ref.[40] $PACO_{ASDI}$[46] uses the PACO framework to model both the temporal and spectral variation of the stellar residuals and track the planetary signal.

The images and corresponding counts of detections and non-detections as a function of the threshold are displayed in Fig. 7 for the case of the fifth Gemini-S/GPI data set (gpi5). The scores proposed in Sect. 3 computed on the 10 data sets from this ADI+mSDI subchallenge are gathered in Tab. 6. The ranking of this subchallenge is shown in Fig. 8.

4.2.1 Ranking the ADI+mSDI subchallenge results

Based on this ADI+mSDI subchallenge, we can already draw a few conclusions. (1) Exploiting properly the spectral information does help to detect fainter sources. As an example, stacking each spectral channel after processing each spectral channel separately using ADI-based techniques, does not allow one to detect the lower
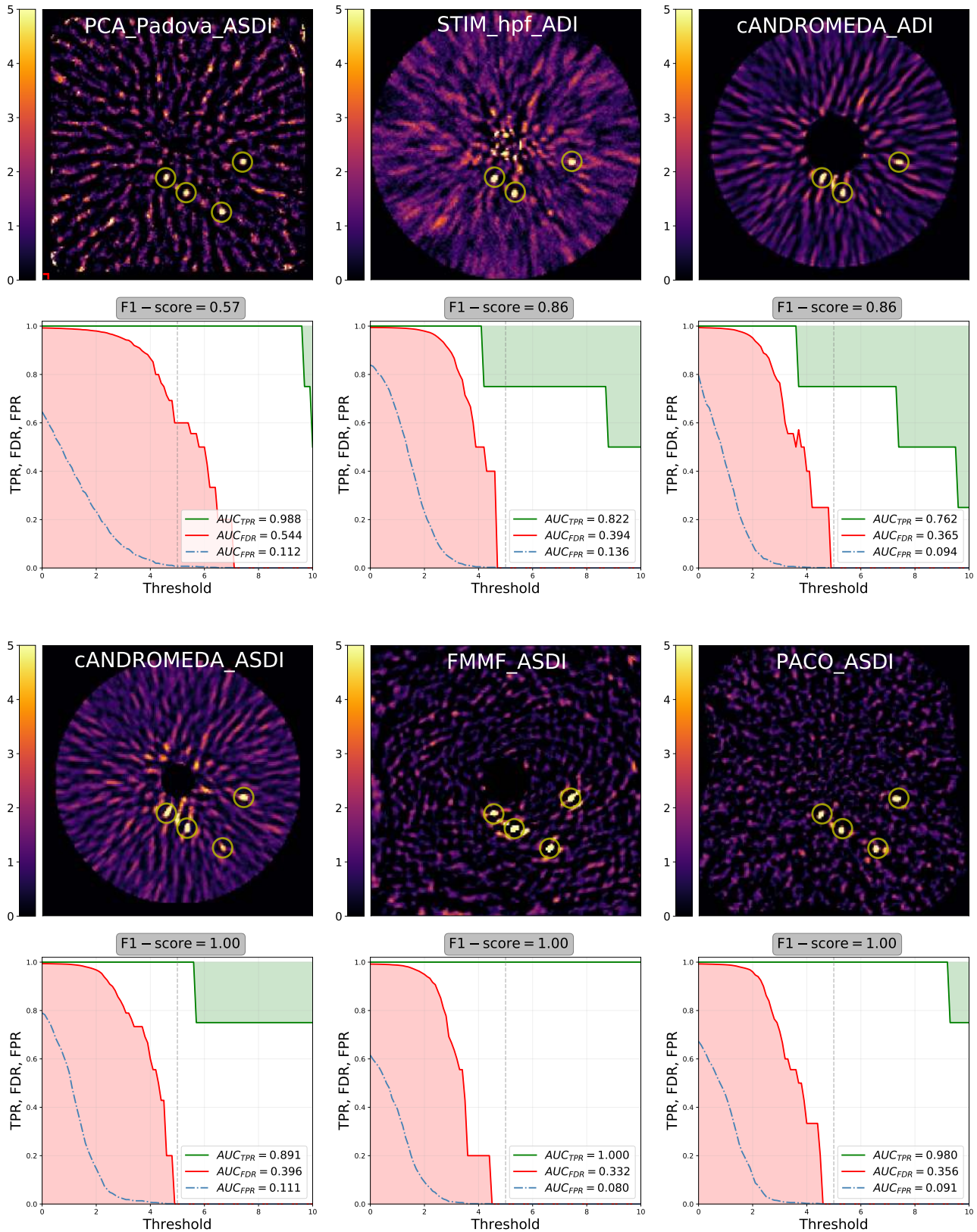
Figure 7. Results of the ADI+mSDI subchallenge for the fifth Gemini-S/GPI dataset (gpi5).

Table 6. Results from the second subchallenge (ADI+mSDI) of the EIDC. The 'X' symbol means that the detection map has not been submitted. The corresponding algorithm, PACO-ASDI, is not be taken into account in the ranking.

Algorithm	VLT/SPHERE-IFS						Gemini-S/GPI						All
	(1)	(2)	(3)	(4)	(5)	Mean	(1)	(2)	(3)	(4)	(5)	Mean	
Baseline	0.80	0.75	0.40	0.00	u	0.49	0.00	0.00	u	0.00	0.67	0.17	0.33
	0.78	0.77	0.63	0.21	u	0.60	0.34	0.24	u	0.19	0.60	0.34	0.47
	0.57	0.36	0.71	0.98	u	0.65	0.46	0.39	u	0.46	0.30	0.40	0.53
PCA$_{Padova,ASDI}$	0.50	0.25	0.18	0.17	u	0.27	0.00	0.15	u	0.22	0.57	0.24	0.26
	0.94	0.87	0.73	0.50	u	0.76	0.46	0.44	u	0.41	0.99	0.57	0.67
	0.59	0.78	0.83	0.58	u	0.69	0.60	0.97	u	0.61	0.54	0.68	0.69
STIM$_{ADI}$	1.00	0.86	0.67	0.40	u	0.73	0.00	0.00	u	0.00	0.86	0.21	0.47
	0.86	0.80	0.65	0.42	u	0.68	0.39	0.29	u	0.24	0.82	0.44	0.56
	0.32	0.30	0.29	0.29	u	0.30	0.36	0.38	u	0.30	0.39	0.36	0.33
ANDROMEDA$_{ADI}$	0.80	1.00	0.67	0.40	u	0.72	1.00	0.00	u	0.29	0.86	0.54	0.63
	0.75	0.81	0.66	0.44	u	0.66	0.73	0.26	u	0.38	0.76	0.53	0.60
	0.46	0.36	0.39	0.41	u	0.41	0.38	0.44	u	0.55	0.36	0.43	0.42
ANDROMEDA$_{ASDI}$	1.00	1.00	0.50	0.33	u	0.71	1.00	0.00	u	0.29	1.00	0.57	0.64
	0.81	0.76	0.65	0.53	u	0.69	0.51	0.35	u	0.34	0.89	0.52	0.60
	0.44	0.32	0.44	0.44	u	0.41	0.37	0.53	u	0.58	0.40	0.47	0.44
FMMF$_{ASDI}$	1.00	1.00	0.80	1.00	u	0.95	0.00	0.80	u	0.00	1.00	0.45	0.70
	1.00	0.97	1.00	0.98	u	0.99	0.37	0.55	u	0.33	1.00	0.56	0.77
	0.39	0.34	0.40	0.33	u	0.36	0.43	0.38	u	0.39	0.33	0.38	0.37
PACO$_{ASDI}$	1.00	0.86	0.67	0.86	u	0.85	x	x	u	x	1.00	x	x
	1.00	0.86	0.57	0.78	u	0.81	x	x	u	x	0.98	x	x
	0.34	0.37	0.38	0.33	u	0.36	x	x	u	x	0.36	x	x
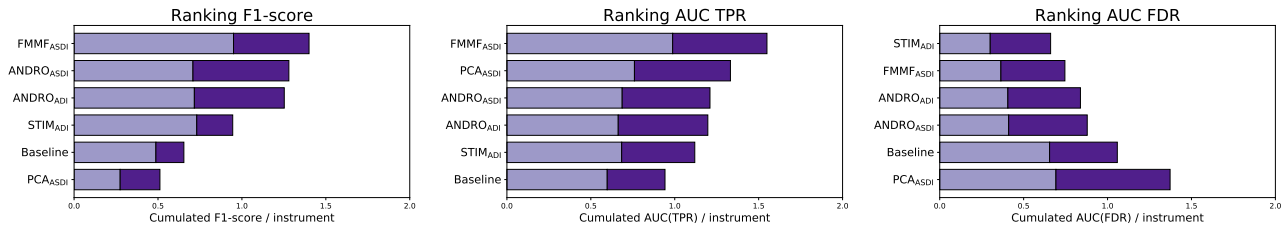
Figure 8. Ranking of the 5 valid algorithms submitted for the EIDC ADI+mSDI subchallenge. Left column: Ranking based on the F1-score. Middle column: Ranking based on the area under curve of the TPR. Right column: Ranking based on the area under curve of the FDR. The light and dark colors correspond to the five VLT/SPHERE-IFS and five Gemini-S/GPI data sets respectively.

faint companion shown in Fig. 7 with STIM$_{ADI}$ compared to PCA$_{Padova,ASDI}$ for the speckle subtraction techniques, and with ANDROMEDA$_{ADI}$ compared to ANDROMEDA$_{ASDI}$ for the inverse problem approaches. (2) The spectral template used as a prior for the planetary signal in some methods plays an important role. For instance, in Fig. 19, PACO$_{ASDI}$ uses a uniform spectral template prior and does not detect one of the four planetary signal. However by using a non-flat spectrum as a prior, the signal of the fourth planet is above the submitted threshold. (3) As a general rule, based on the image galleries in App. B, the more advanced/recent techniques perform better at detecting planetary signals but the results are not homogeneous. This shows once again that applying various algorithms based on different concept is the most appropriate way of validating or invalidating a candidate based on a single data set. (4) As for the ADI subchallenge, inverse problem approaches are the only methods providing directly with the contrast of the planetary signal candidates. This allows to additionally help disentangling the candidate from a stellar residual, by comparing the extracted spectrum of the candidate with the extracted spectrum of a residual speckle.

5. CONCLUSION & PERSPECTIVES

In this paper, we presented the Exoplanet Imaging Data Challenge and the results from its first phase, consisting in assessing the detection performance of various post-processing techniques dedicated to exoplanet imaging. For this first phase, we gathered typical data sets from several instruments offering a high-contrast imaging mode, in which we injected synthetic planetary signals. The data sets are available permanently on a *Zenodo* repository, and are described on the EIDC website[††]. Participants were invited to submit their results, consisting in one detection map for each data set and one single threshold value for all the data sets, on our competition platform hosted on *CodaLab*. For the analysis of the submitted detection maps, we opted for counting the detections and non-detections at a given threshold and derive relevant scores to compare the different submissions. By organizing and maintaining this data challenge, we intend to provide the user-community with a global vision of all the existing post-processing methods and their different capabilities. Within the developer-community, we intend to trigger new collaborations and ideas, while providing with a platform for a common comparison between the algorithms.

The first phase of the EIDC presented in this paper is limited in several aspects: the data chosen (via, for instance, the planetary signal injection procedure) and the evaluation method proposed (by counting detections and non-detections) are not perfectly representative and might favor or disfavor some methods. Further work will make use of the data challenge results and apply evaluation methods more adapted to the planet detection problem. On the practical aspect, we encountered a few difficulties with the CodaLab platform, in particular because it is not designed for participants to submit more than one algorithm products (it displays only one result per participant in the leader-board). In addition, the position of the injections are revealed to the participants after their first submission, potentially biasing the following submissions (one participant can submit as many results as they wanted). At last, the leader-board is currently displaying incorrect results since the backend calculating the F1-score is erroneous and it was unfortunately not possible to modify the backend while the competition is open.

In the future, we intend to increase the number of data sets available. For instance, we put a link towards the NEAR campaign[47] data on the EIDC website[‡‡], to allow anyone to test advanced algorithms on this challenging observation sequence. On top of using comparison metrics that are more adapted, we also plan to launch new phases of the data challenge, adding, for example, the characterisation of the detections (position and contrast with their uncertainties) or adding extended sources to be detected. We also plan to offer alternative data set enabling the direct detection of exoplanets, such as high spectral resolution images.[48,49] We aim to regularly publish the results of the EIDC at every future SPIE astronomical telescopes+instrumentation conference.

APPENDIX A. GALLERY OF DETECTION MAPS: ADI SUBCHALLENGE

In this appendix, we display the detection maps of the 22 submissions for the nine data sets of the ADI subchallenge. Each detection map is shown with a color-bar ranging from 0 to the submitted threshold. The yellow circles indicate true positives at the given threshold, while red squares indicate false positives. The first image (top left) shows the detection map for the chosen baseline algorithm (annular PCA). The last image (bottom right) shows the mask used to conduct the analysis of each map, with blue circles indicating the location of the injected synthetic planetary signals.

APPENDIX B. GALLERY OF DETECTION MAPS: ADI+MSDI SUBCHALLENGE

In this appendix, we display the detection maps of the 6 submissions for the ten data sets of the ADI+mSDI subchallenge. Each detection map is shown with a color-bar ranging from 0 to the submitted threshold. The yellow circles indicate true positives at the given threshold, while red squares indicate false positives. The first image (top left) shows the detection map for the chosen baseline algorithm (annular PCA). The last image (bottom right) shows the mask used to conduct the analysis of each map, with blue circles indicating the location of the injected synthetic planetary signals.

[††]https://exoplanet-imaging-challenge.github.io/datasets/
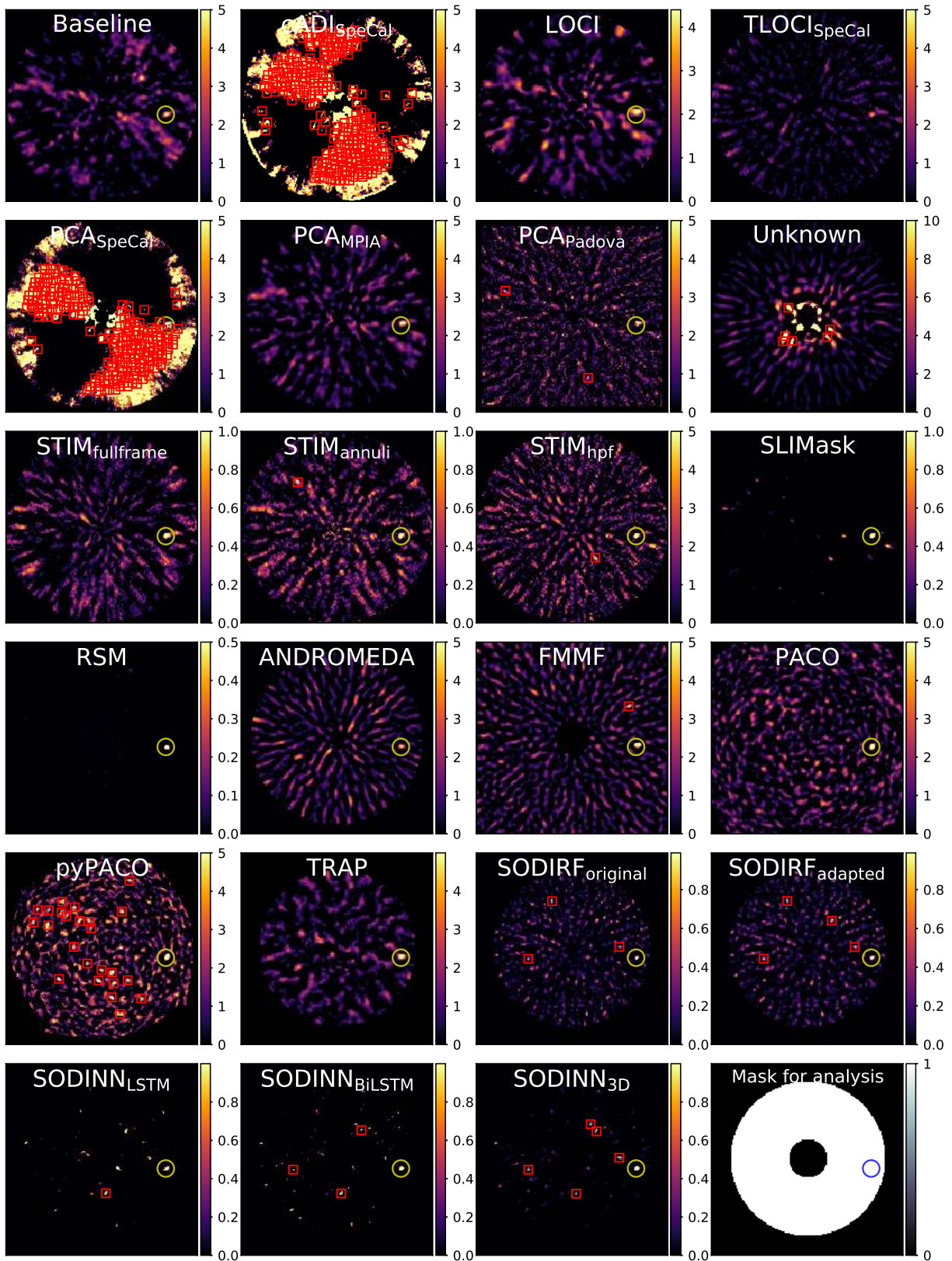[‡‡]https://exoplanet-imaging-challenge.github.io/near/

Figure 9. Results of the ADI subchallenge for the first VLT/SPHERE-IRDIS dataset (sph1).
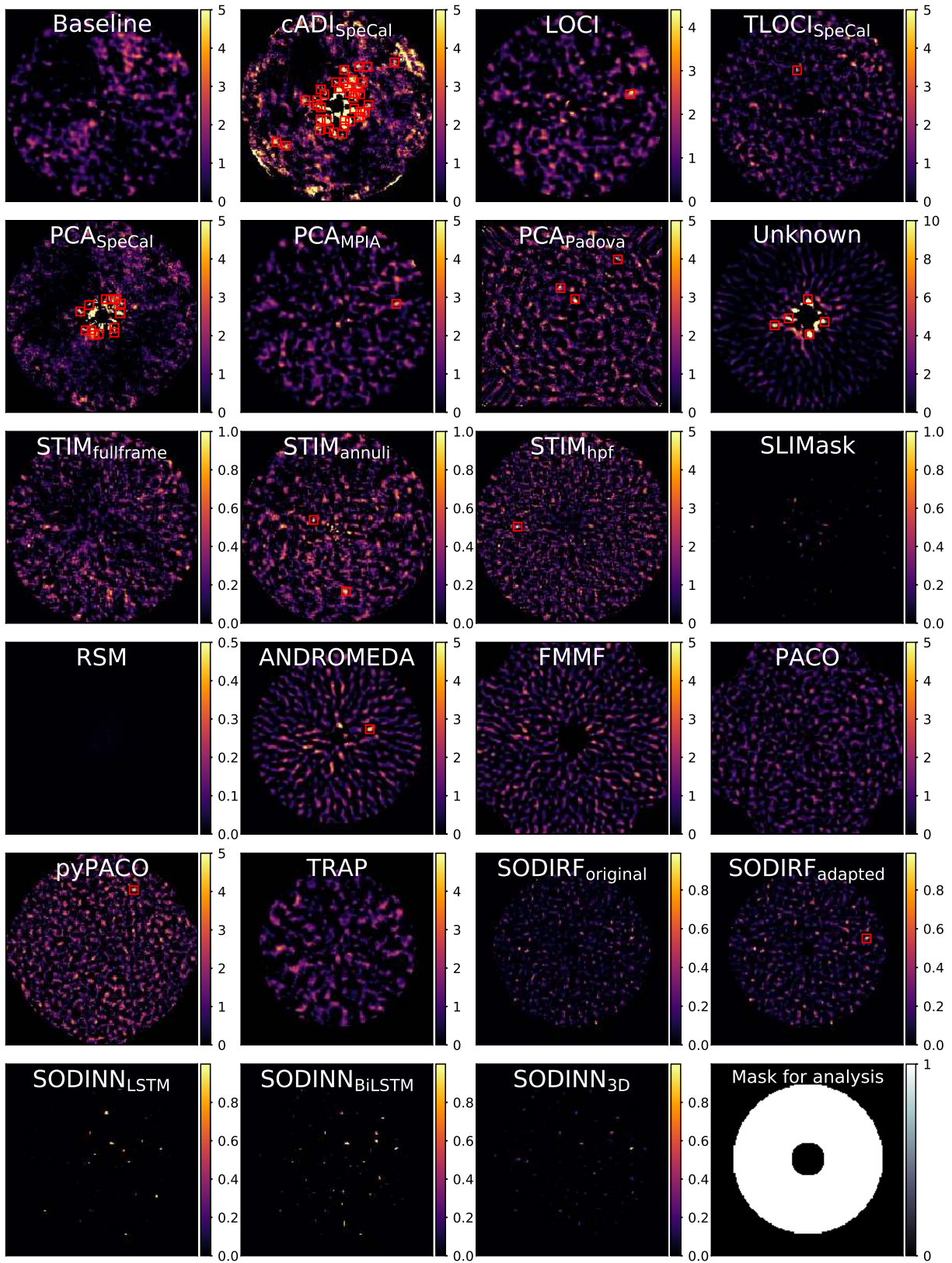
Figure 10. Results of the ADI subchallenge for the second VLT/SPHERE-IRDIS dataset (sph2).
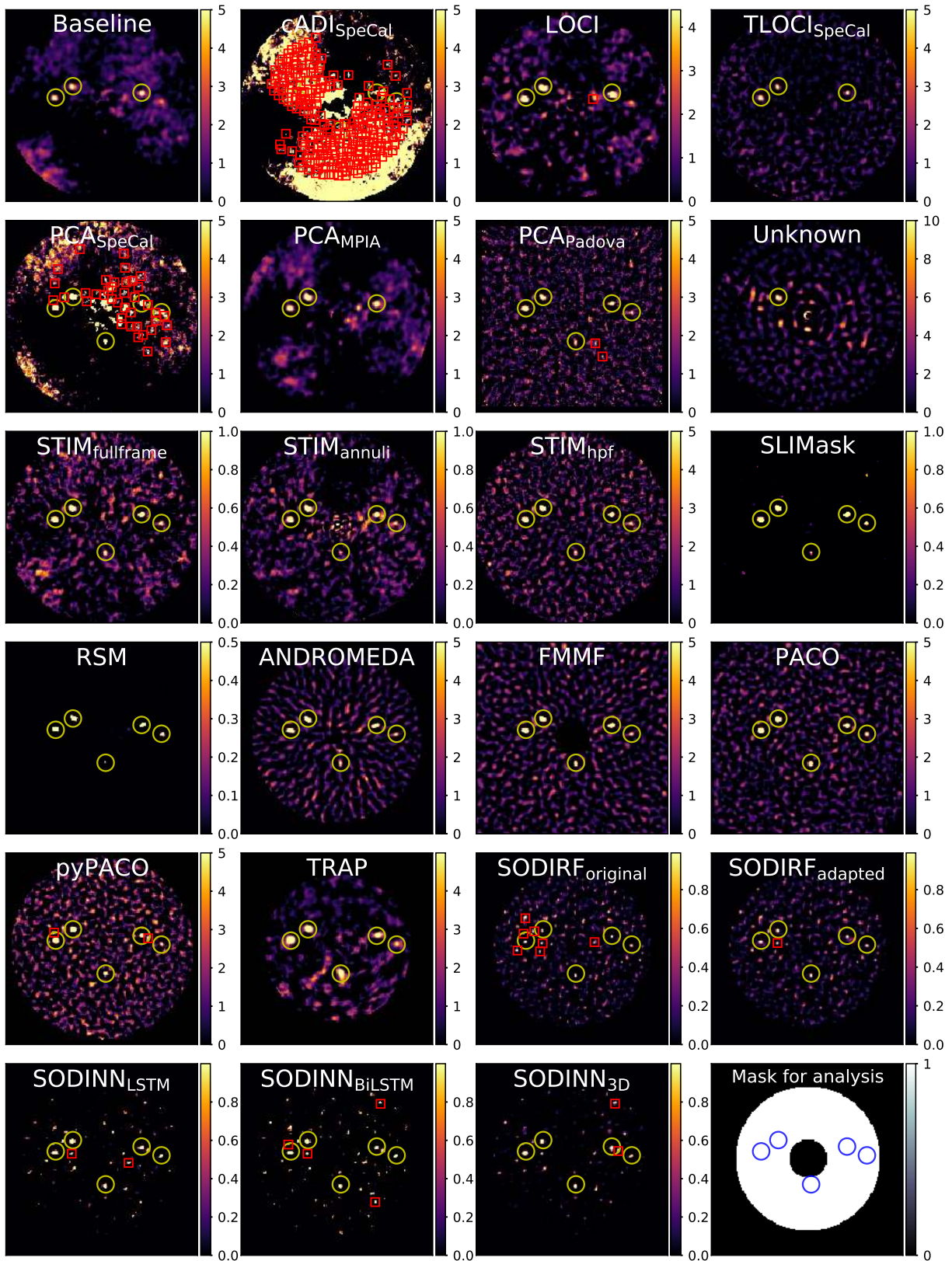
Figure 11. Results of the ADI subchallenge for the third VLT/SPHERE-IRDIS dataset (sph3).
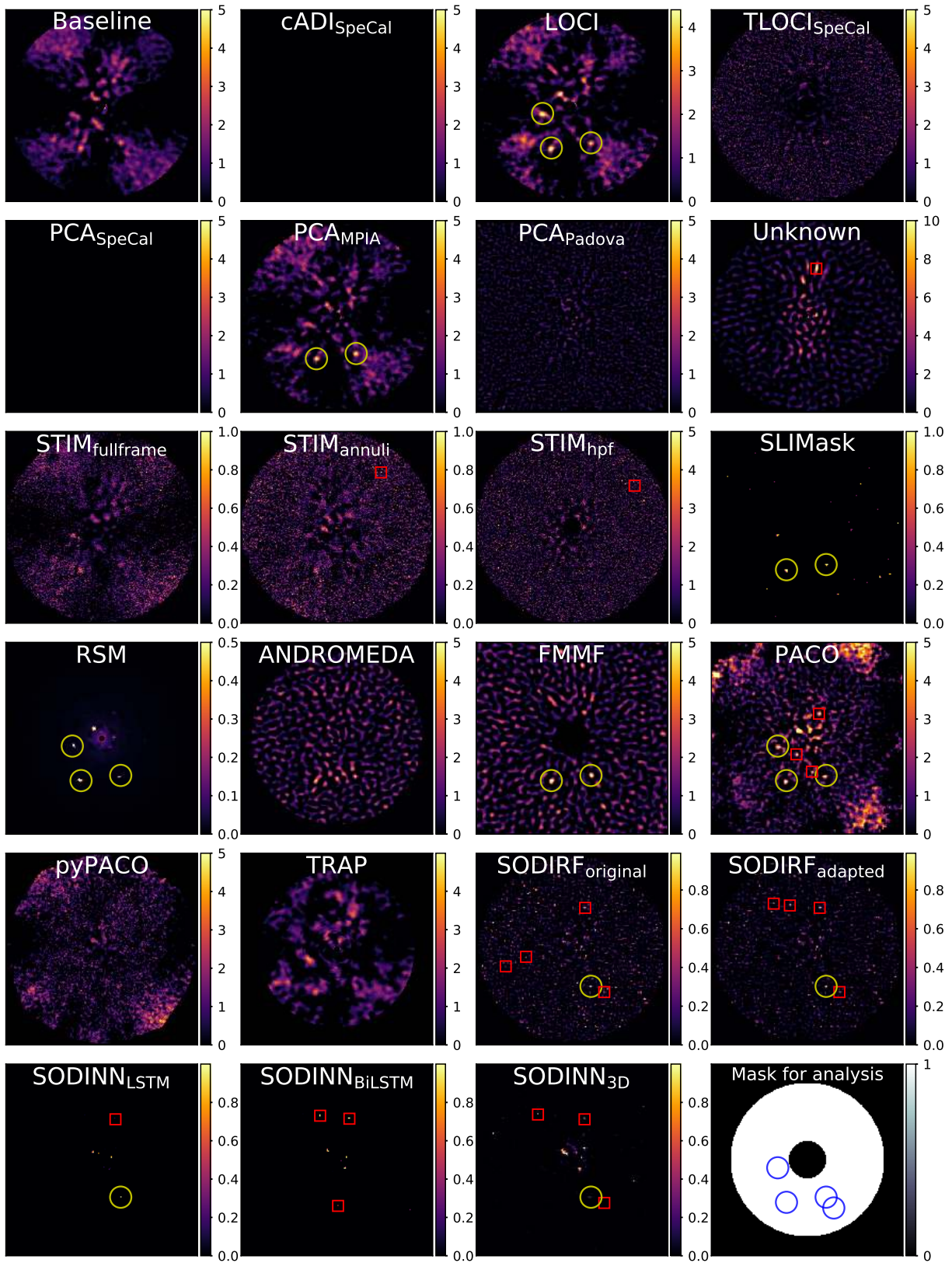
Figure 12. Results of the ADI subchallenge for the first Keck/NIRC2 dataset (nrc1).
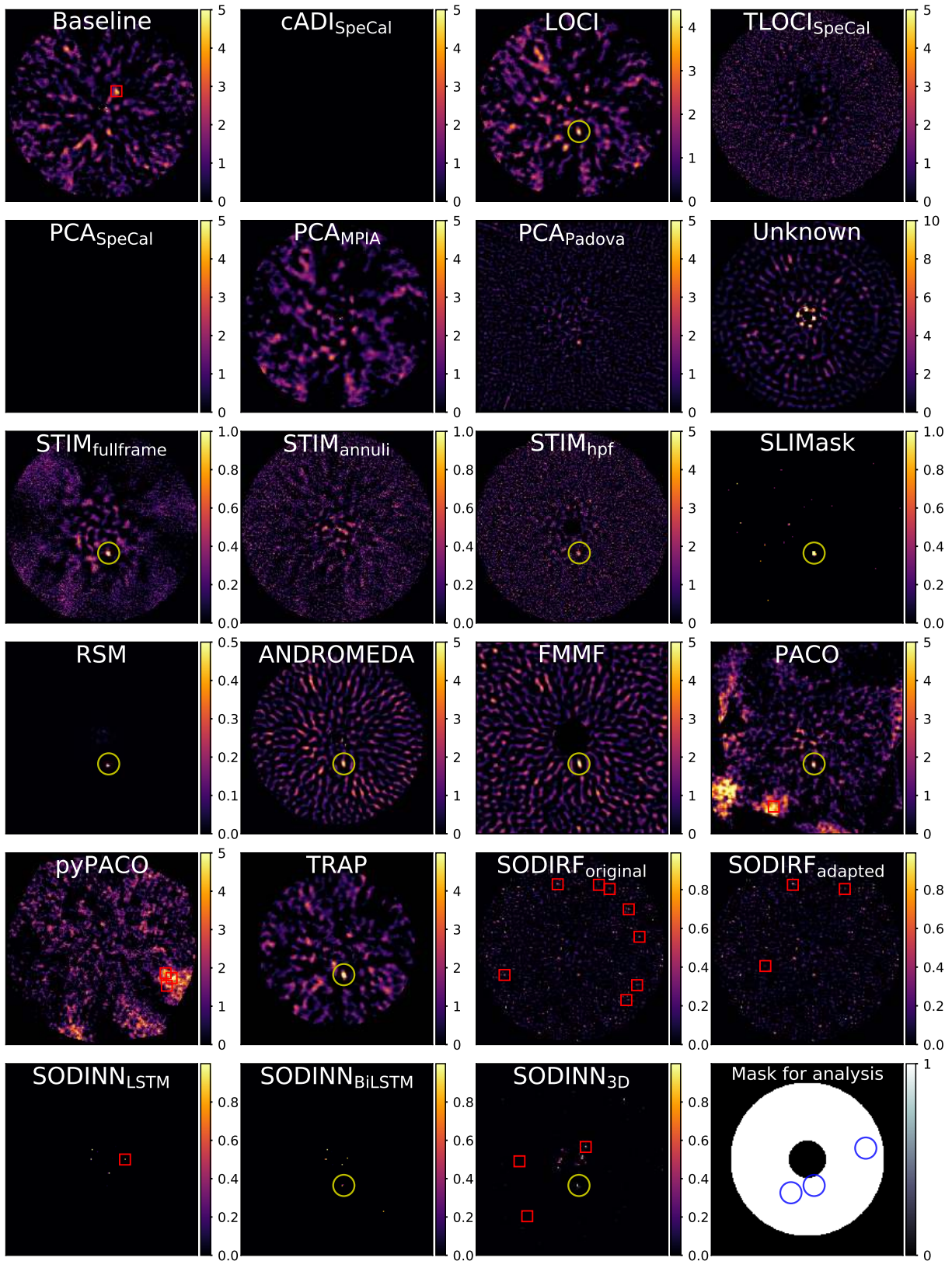
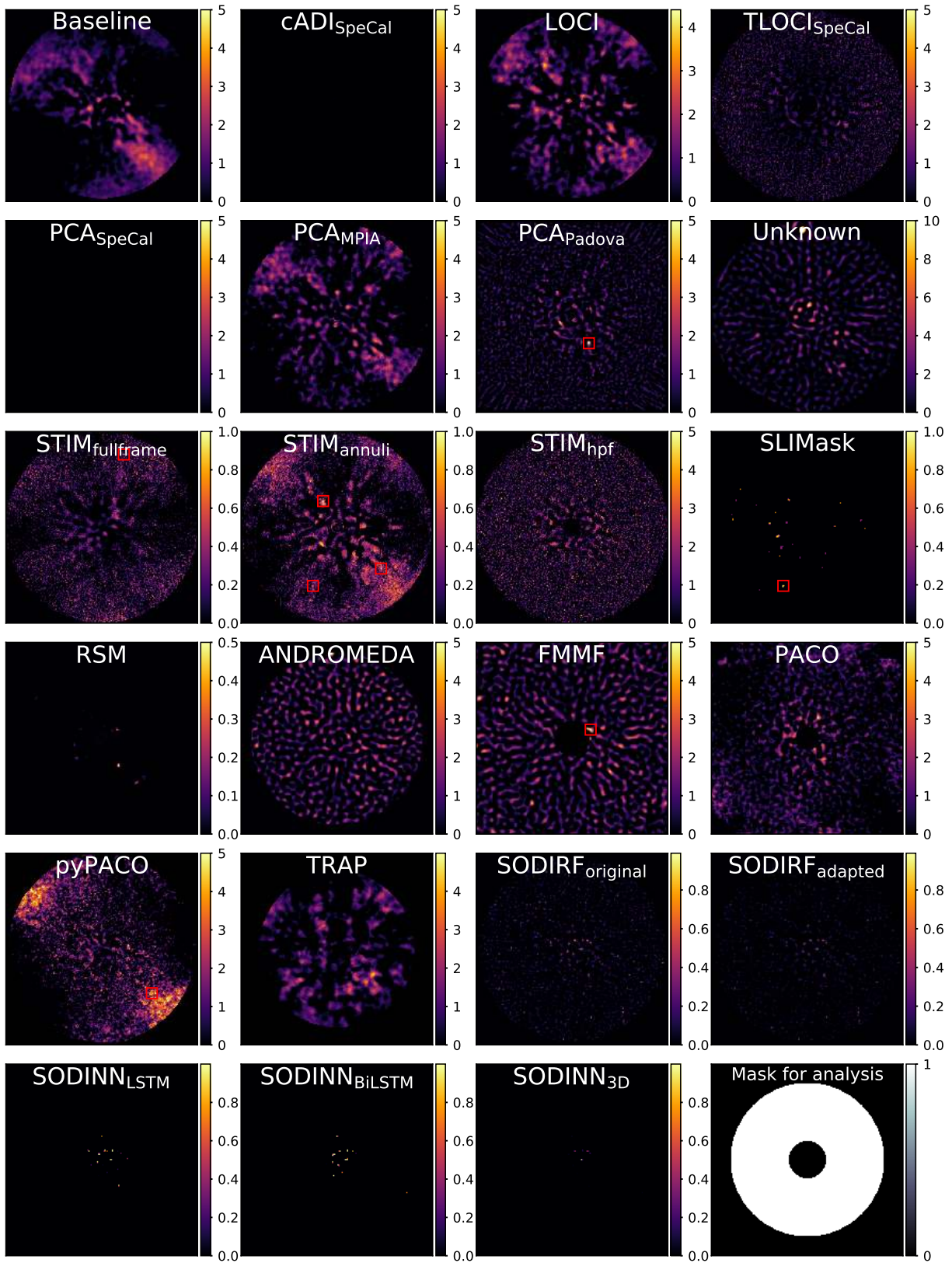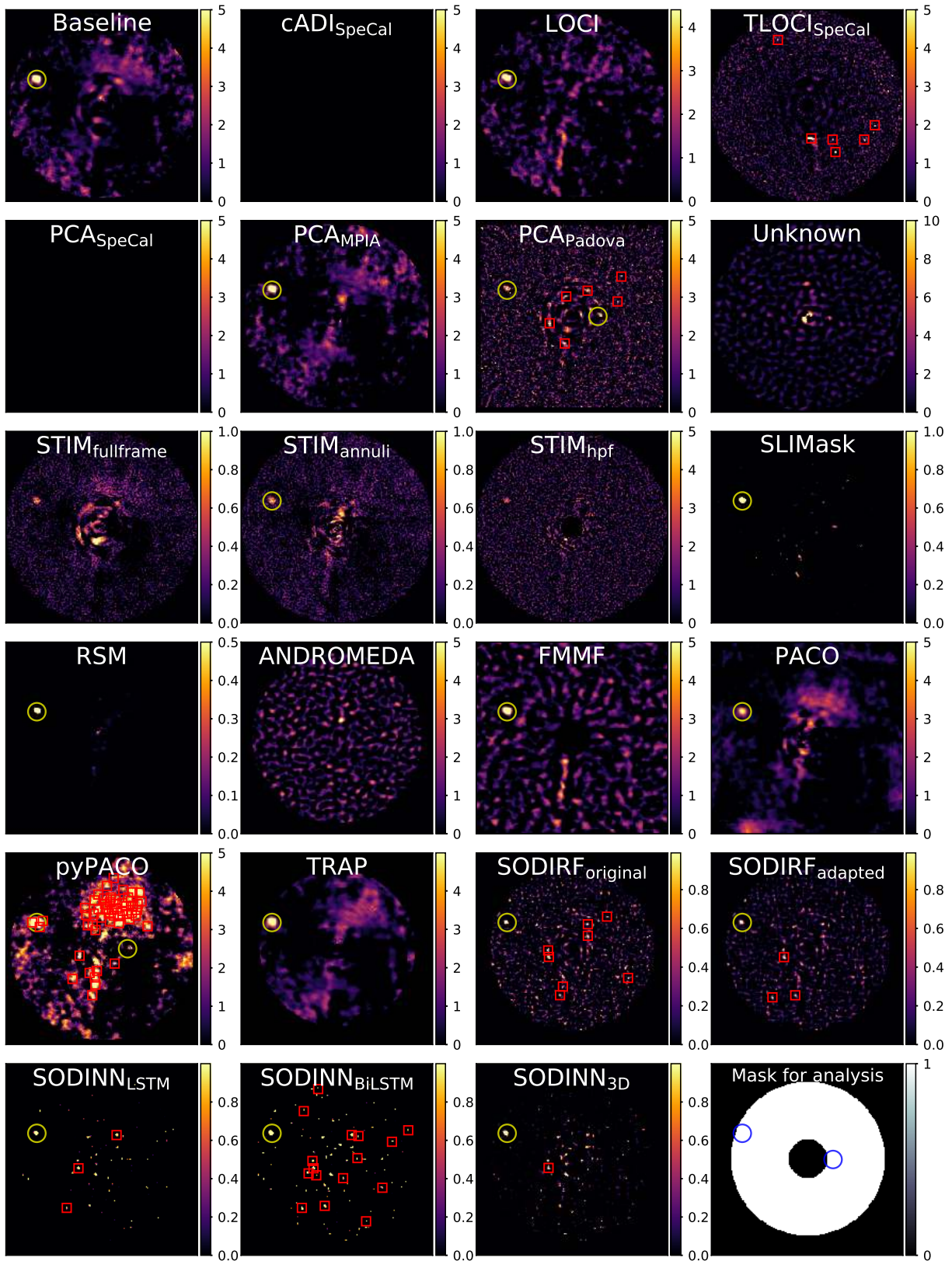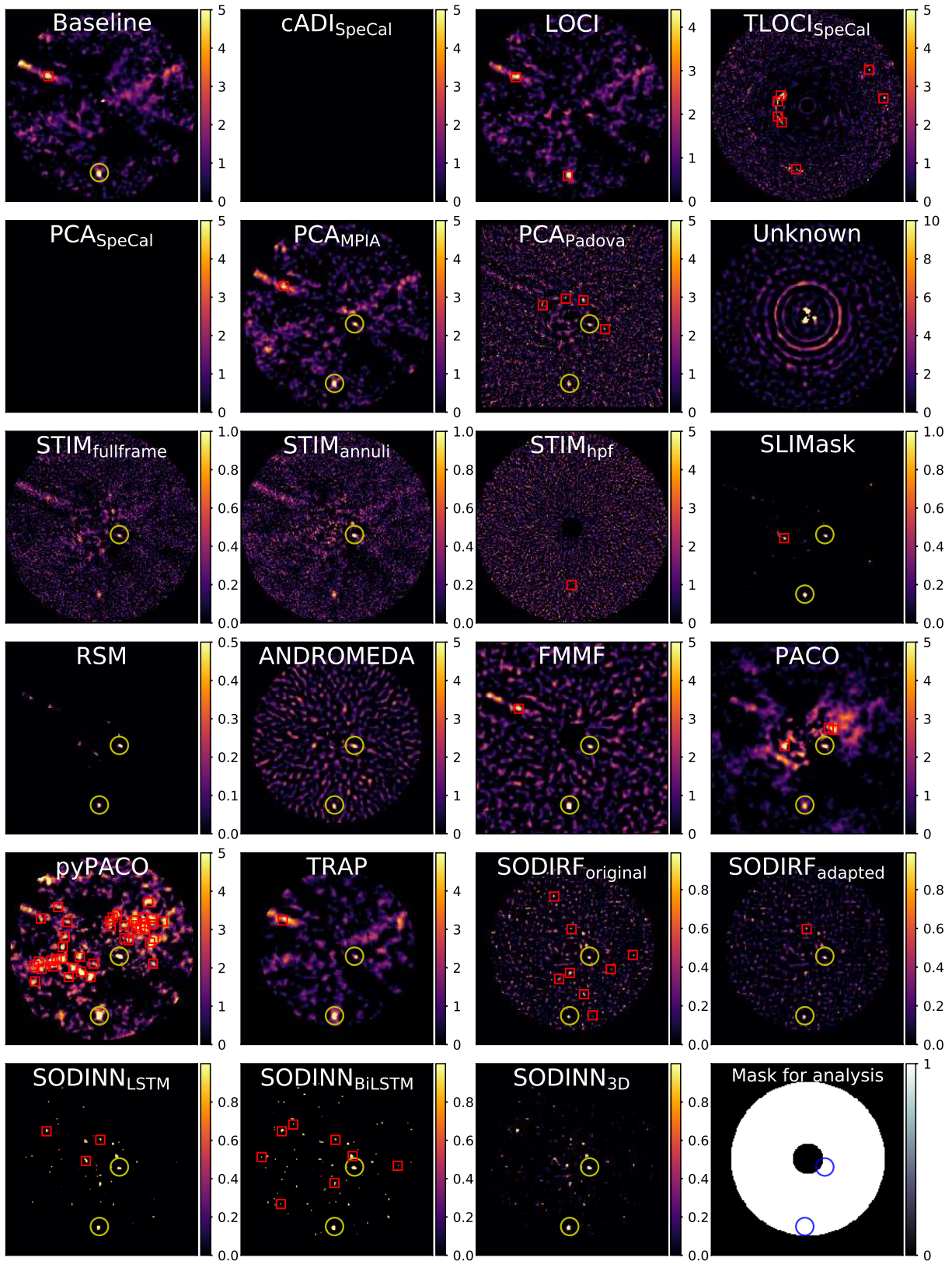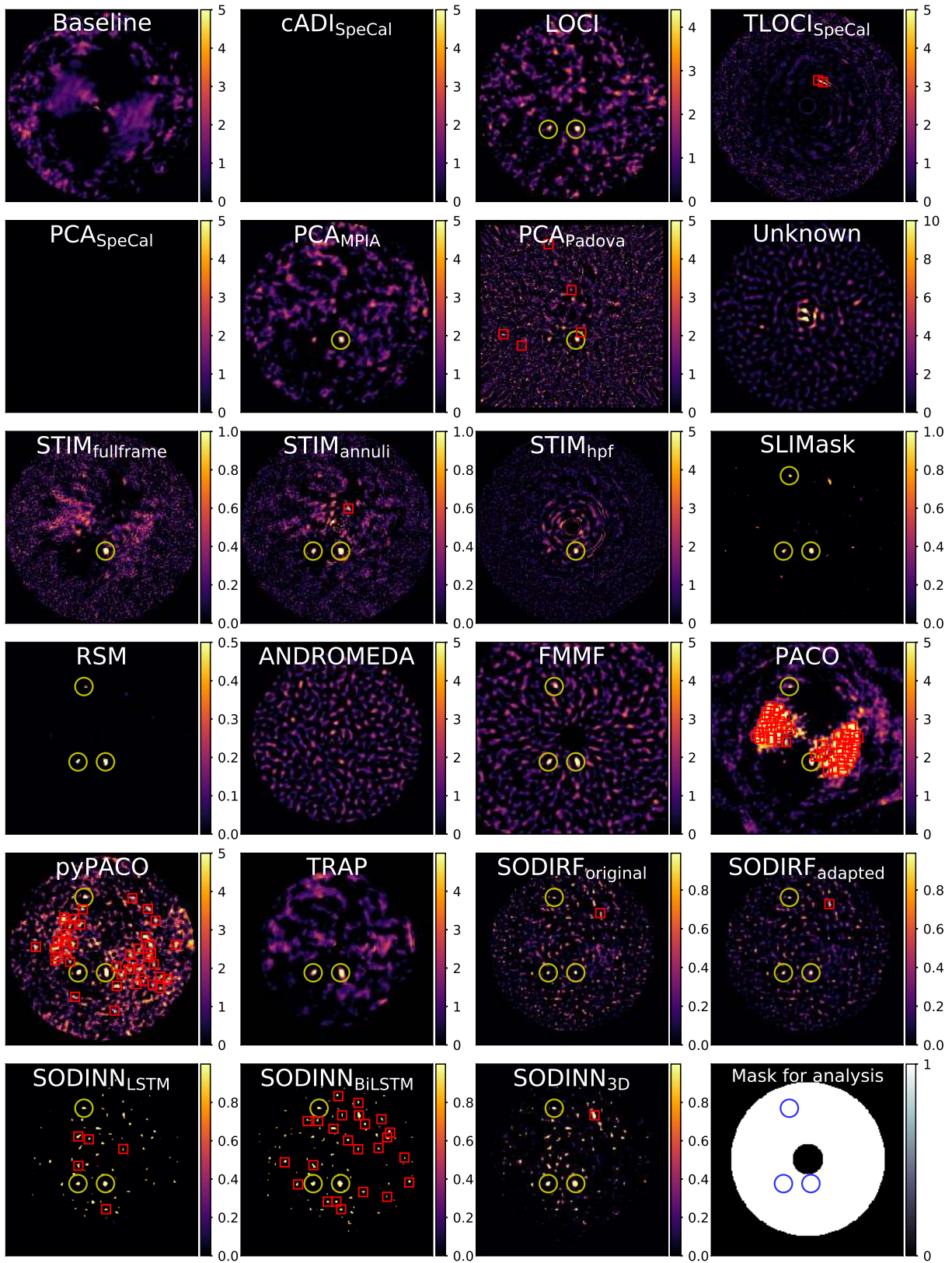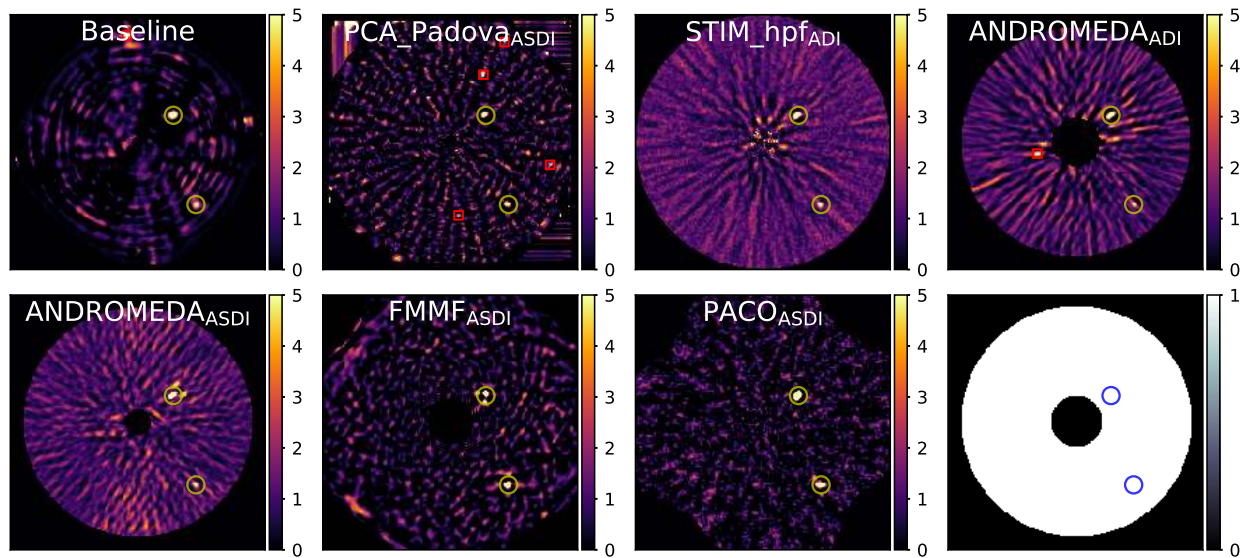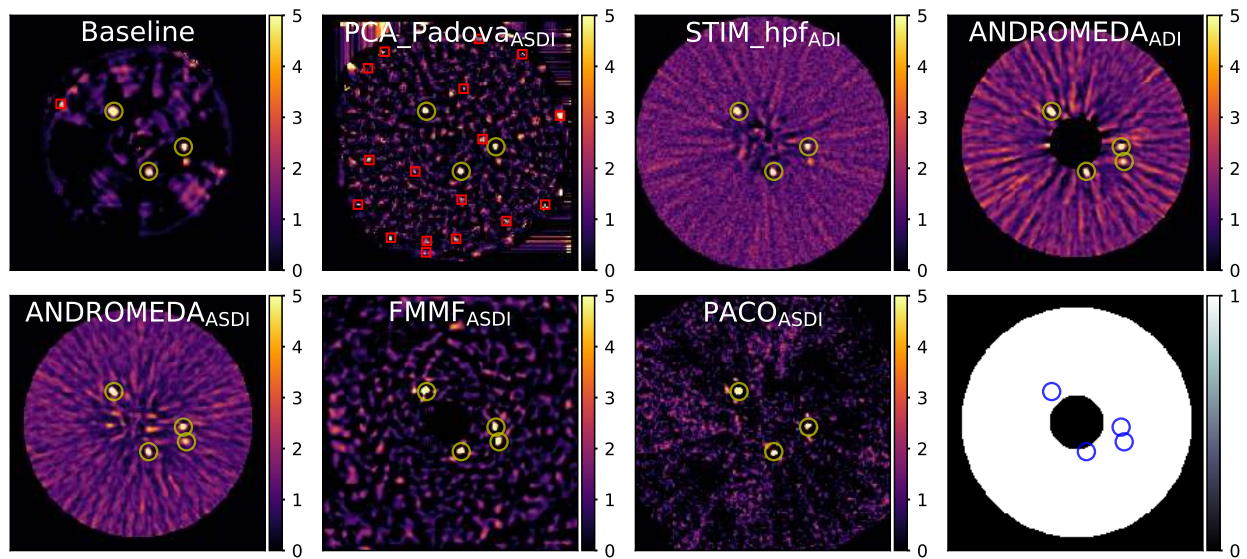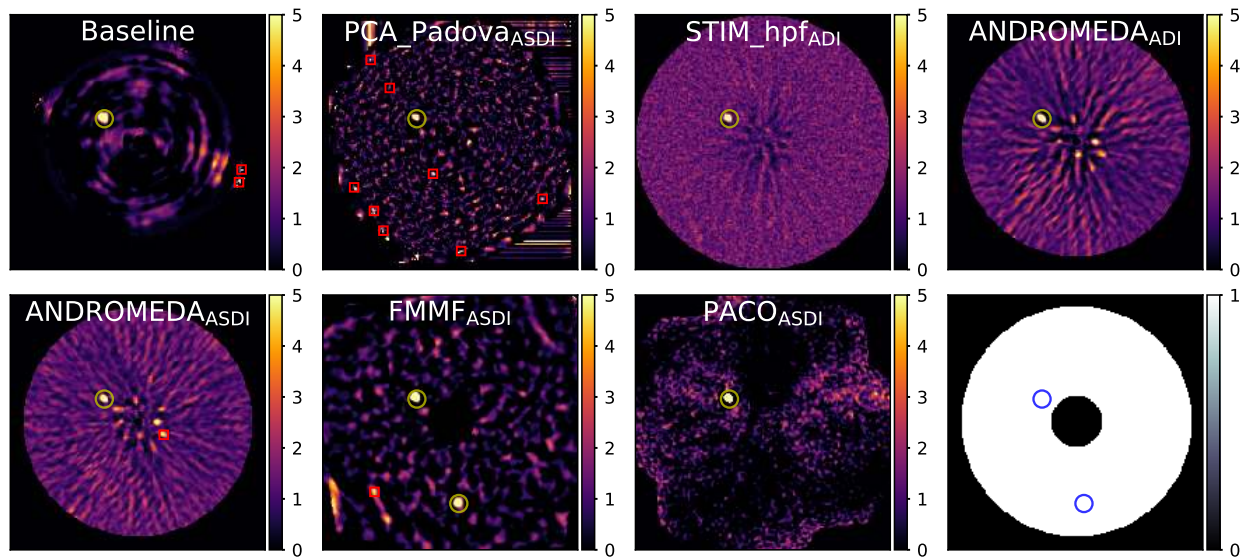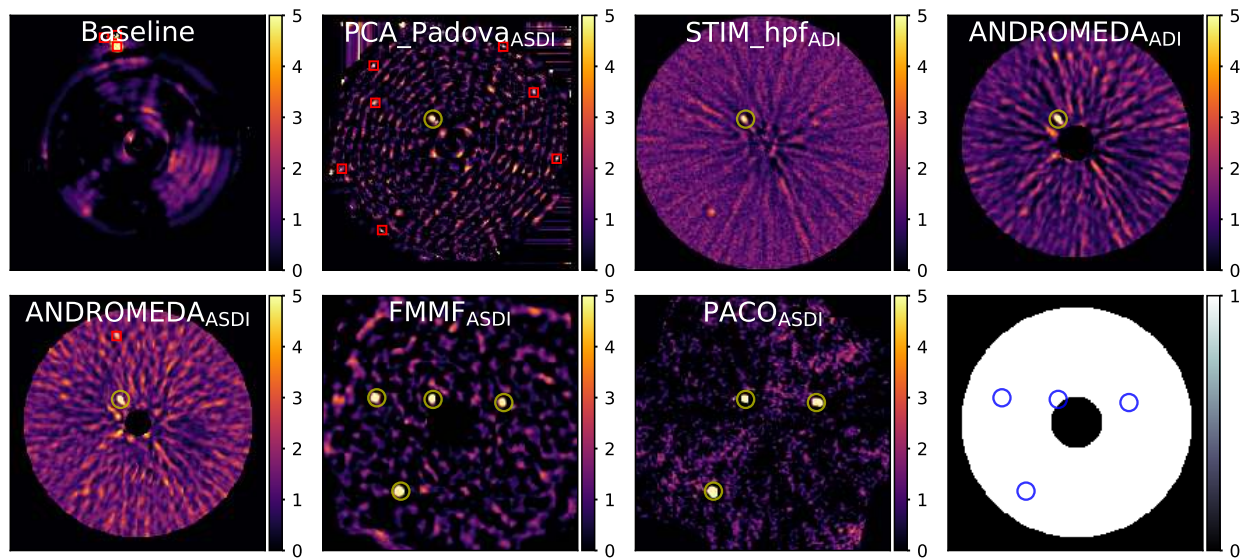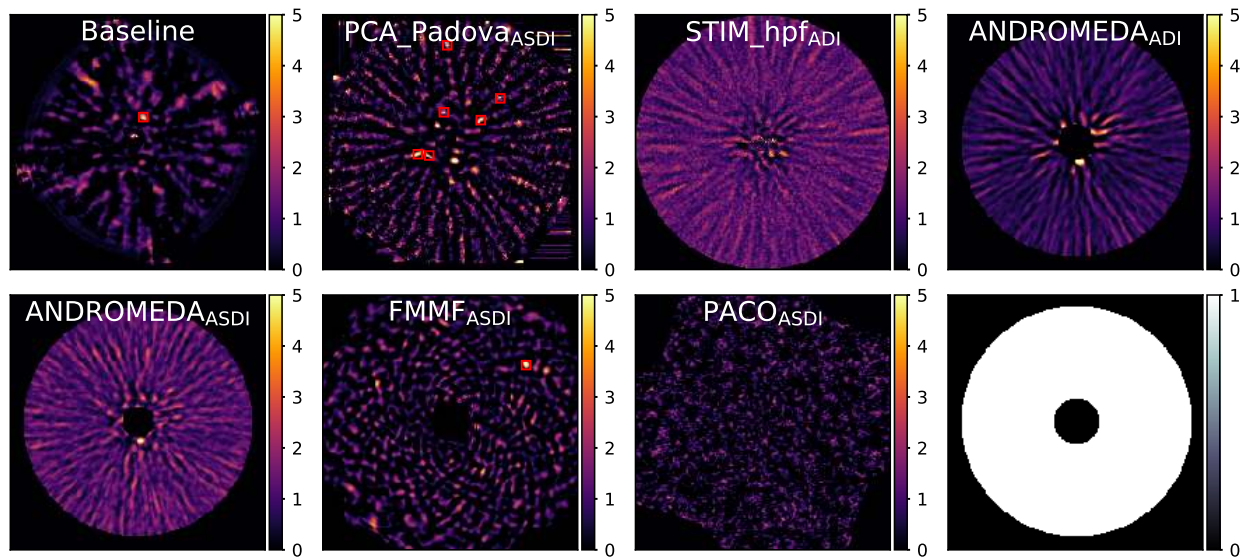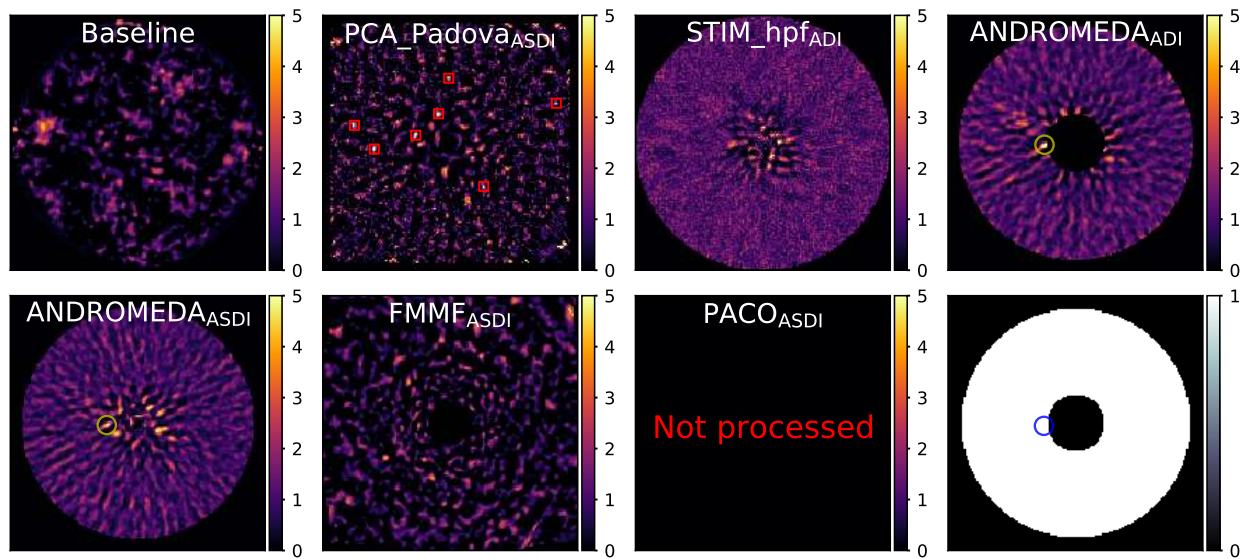Figure 13. Results of the ADI subchallenge for the second Keck/NIRC2 dataset (nrc2).

Figure 14. Results of the ADI subchallenge for the third Keck/NIRC2 dataset (nrc3).

Figure 15. Results of the ADI subchallenge for the first LBT/LMIRCam dataset (lmr1).

Figure 16. Results of the ADI subchallenge for the second LBT/LMIRCam dataset (lmr2).

Figure 17. Results of the ADI subchallenge for the third LBT/LMIRCam dataset (lmr3).

Figure 18. Results of the ADI+mSDI subchallenge for the first VLT/SPHERE-IFS dataset (ifs1).
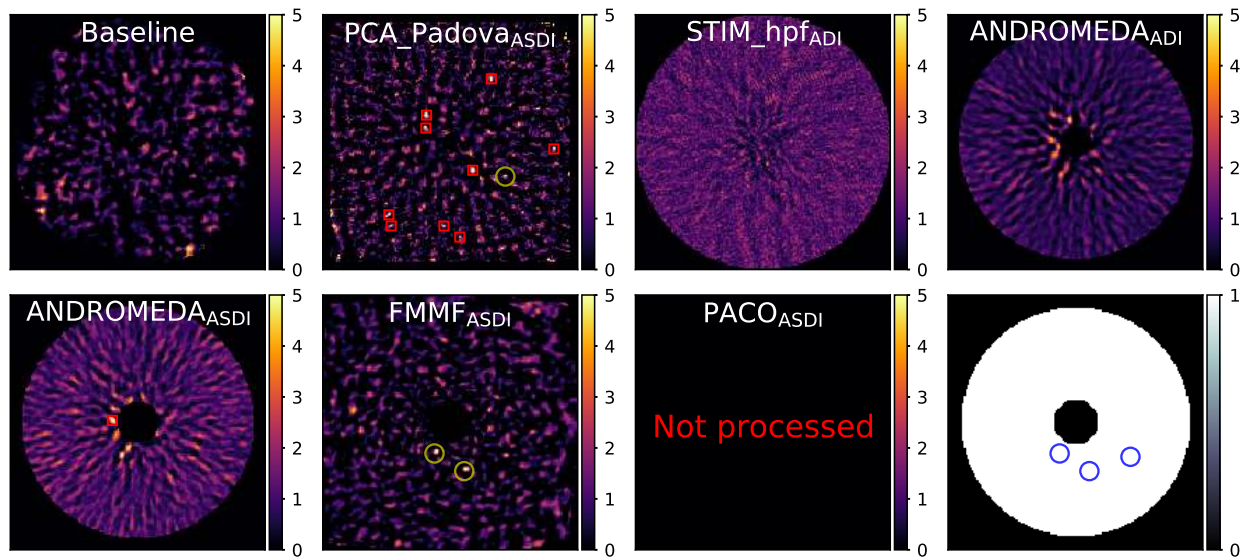

Figure 19. Results of the ADI+mSDI subchallenge for the second VLT/SPHERE-IFS dataset (ifs2).

Figure 20. Results of the ADI+mSDI subchallenge for the third VLT/SPHERE-IFS dataset (ifs3).

Figure 21. Results of the ADI+mSDI subchallenge for the fourth VLT/SPHERE-IFS dataset (ifs4).

Figure 22. Results of the ADI+mSDI subchallenge for the fifth VLT/SPHERE-IFS dataset (ifs5).

Figure 23. Results of the ADI+mSDI subchallenge for the first Gemini-S/GPI dataset (gpi1).

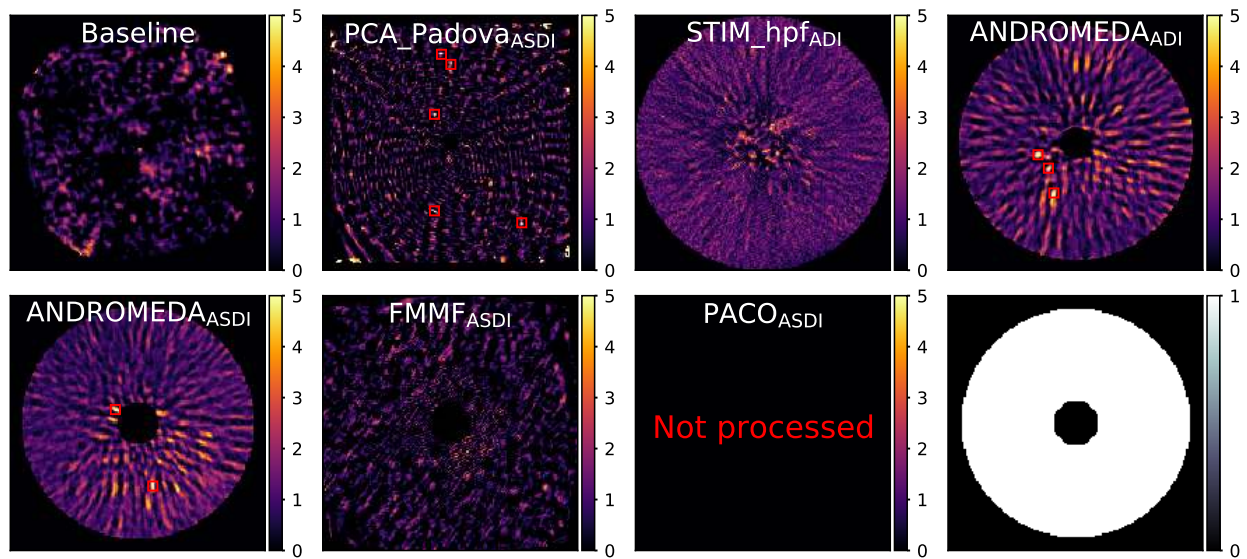Figure 24. Results of the ADI+mSDI subchallenge for the second Gemini-S/GPI dataset (gpi2).

Figure 25. Results of the ADI+mSDI subchallenge for the third Gemini-S/GPI dataset (gpi3).
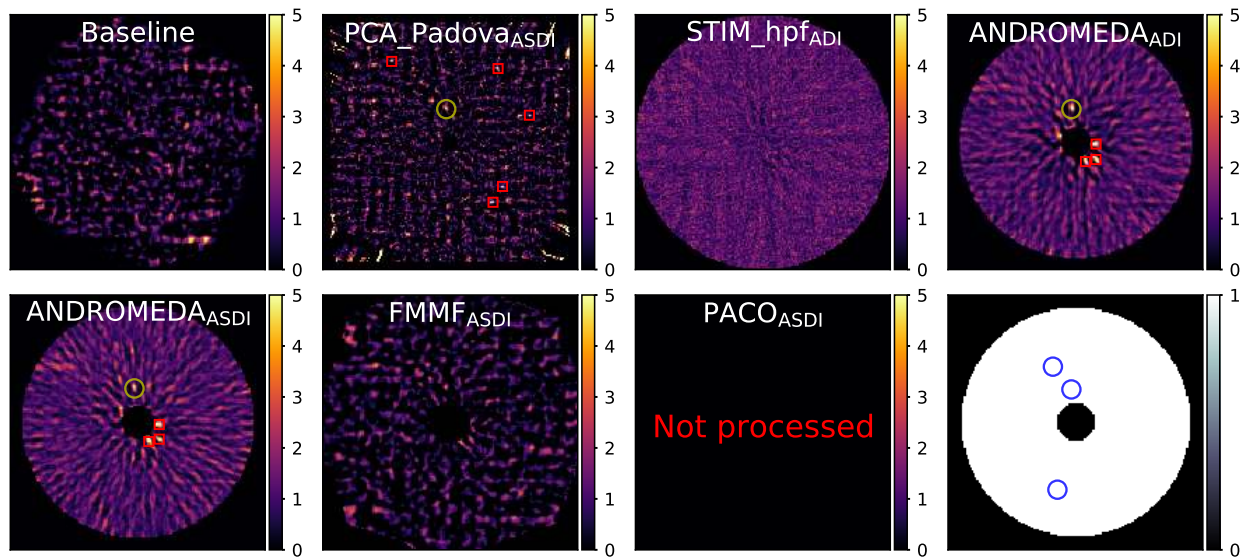
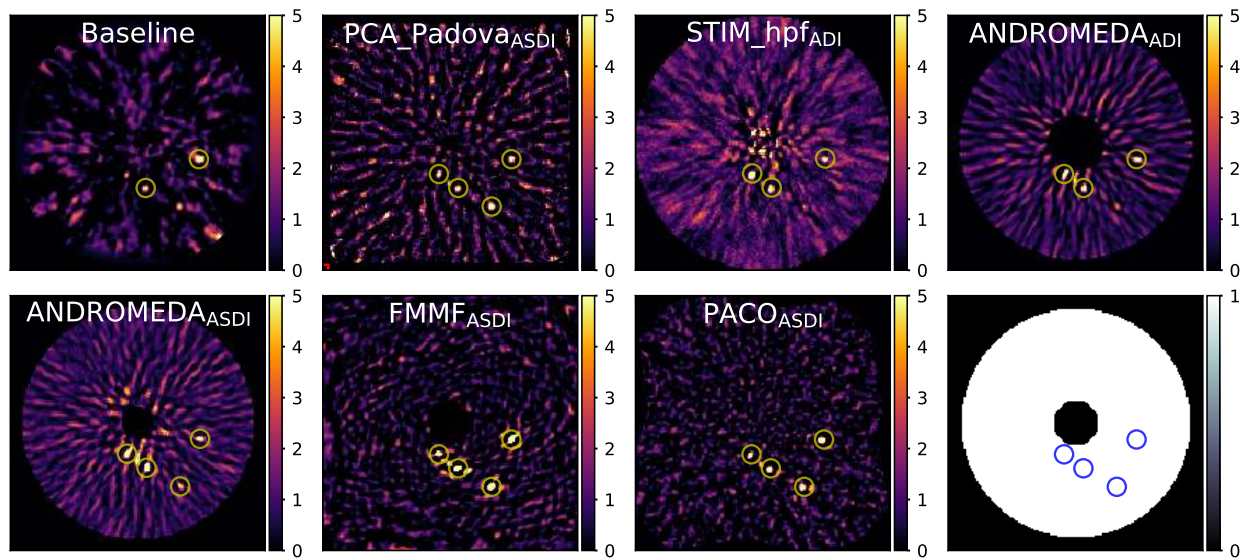Figure 26. Results of the ADI+mSDI subchallenge for the fourth Gemini-S/GPI dataset (gpi4).

Figure 27. Results of the ADI+mSDI subchallenge for the fifth Gemini-S/GPI dataset (gpi5).

ACKNOWLEDGMENTS

REFERENCES

[1] Marois, C., Lafrenière, D., Doyon, R., Macintosh, B., and Nadeau, D., "Angular Differential Imaging: A Powerful High-Contrast Imaging Technique," **641**, 556–564 (Apr. 2006).

[2] Beuzit, J. L., Vigan, A., Mouillet, D., Dohlen, K., Gratton, R., Boccaletti, A., Sauvage, J. F., Schmid, H. M., Langlois, M., Petit, C., Baruffolo, A., Feldt, M., Milli, J., Wahhaj, Z., Abe, L., Anselmi, U., Antichi, J., Barette, R., Baudrand, J., Baudoz, P., Bazzon, A., Bernardi, P., Blanchard, P., Brast, R., Bruno, P., Buey, T., Carbillet, M., Carle, M., Cascone, E., Chapron, F., Charton, J., Chauvin, G., Claudi, R., Costille, A., De Caprio, V., de Boer, J., Delboulbé, A., Desidera, S., Dominik, C., Downing, M., Dupuis, O., Fabron, C., Fantinel, D., Farisato, G., Feautrier, P., Fedrigo, E., Fusco, T., Gigan, P., Ginski, C., Girard, J., Giro, E., Gisler, D., Gluck, L., Gry, C., Henning, T., Hubin, N., Hugot, E., Incorvaia, S., Jaquet, M., Kasper, M., Lagadec, E., Lagrange, A. M., Le Coroller, H., Le Mignant, D., Le Ruyet, B., Lessio, G., Lizon, J. L., Llored, M., Lundin, L., Madec, F., Magnard, Y., Marteaud, M., Martinez, P., Maurel, D., Ménard, F., Mesa, D., Möller-Nilsson, O., Moulin, T., Moutou, C., Origné, A., Parisot, J., Pavlov, A., Perret, D., Pragt, J., Puget, P., Rabou, P., Ramos, J., Reess, J. M., Rigal, F., Rochat, S., Roelfsema, R., Rousset, G., Roux, A., Saisse, M., Salasnich, B., Santambrogio, E., Scuderi, S., Segransan, D., Sevin, A., Siebenmorgen, R., Soenke, C., Stadler, E., Suarez, M., Tiphène, D., Turatto, M., Udry, S., Vakili, F., Waters, L. B. F. M., Weber, L., Wildi, F., Zins, G., and Zurlo, A., "SPHERE: the exoplanet imager for the Very Large Telescope," **631**, A155 (Nov 2019).

[3] Dohlen, K., Langlois, M., Saisse, M., Hill, L., Origne, A., Jacquet, M., Fabron, C., Blanc, J.-C., Llored, M., Carle, M., Moutou, C., Vigan, A., Boccaletti, A., Carbillet, M., Mouillet, D., and Beuzit, J.-L., "The infra-red dual imaging and spectrograph for sphere: design and performance," (2008).

[4] Serabyn, E., Huby, E., Matthews, K., Mawet, D., Absil, O., Femenia, B., Wizinowich, P., Karlsson, M., Bottom, M., Campbell, R., et al., "The wm keck observatory infrared vortex coronagraph and a first image of hip 79124 b," *The Astronomical Journal* **153**(1), 43 (2017).

[5] Skrutskie, M., Jones, T., Hinz, P., Garnavich, P., Wilson, J., Nelson, M., Solheid, E., Durney, O., Hoffmann, W., Vaitheeswaran, V., et al., "The large binocular telescope mid-infrared camera (lmircam): final design and status," in [*Ground-based and Airborne Instrumentation for Astronomy III*], **7735**, 77353H, International Society for Optics and Photonics (2010).

[6] Fusco, T., Rousset, G., Sauvage, J.-F., Petit, C., Beuzit, J.-L., Dohlen, K., Mouillet, D., Charton, J., Nicolle, M., Kasper, M., Baudoz, P., and Puget, P., "High-order adaptive optics requirements for direct detection of extrasolar planets: Application to the SPHERE instrument," *Optics Express* **14**, 7515 (2006).

[7] Soummer, R., "Apodized Pupil Lyot Coronagraphs for Arbitrary Telescope Apertures," **618**, L161–L164 (Jan. 2005).

[8] Vigan, A., Moutou, C., Langlois, M., Allard, F., Boccaletti, A., Carbillet, M., Mouillet, D., and Smith, I., "Photometric characterization of exoplanets using angular and spectral differential imaging," **407**, 71–82 (Sept. 2010).

[9] Wizinowich, P., Acton, D., Shelton, C., Stomski, P., Gathright, J., Ho, K., Lupton, W., Tsubota, K., Lai, O., Max, C., et al., "First light adaptive optics images from the keck ii telescope: a new era of high angular resolution imagery," *Publications of the Astronomical Society of the Pacific* **112**(769), 315 (2000).

[10] Mawet, D., Riaud, P., Absil, O., and Surdej, J., "Annular Groove Phase Mask Coronagraph," **633**, 1191–1200 (Nov. 2005).

[11] Castellá¡, B. F., Serabyn, E., Mawet, D., Absil, O., Wizinowich, P., Matthews, K., Huby, E., Bottom, M., Campbell, R., Chan, D., Carlomagno, B., Cetre, S., DefrÃšre, D., Delacroix, C., Gonzalez, C. G., Jolivet, A., Karlsson, M., Lanclos, K., Lilley, S., Milner, S., Ngo, H., Reggiani, M., Simmons, J., Tran, H., Catalan, E. V., and Wertz, O., "Commissioning and first light results of an L'-band vortex coronagraph with the Keck II adaptive optics NIRC2 science instrument," in [*Adaptive Optics Systems V*], Marchetti, E., Close, L. M., and VÃ©ran, J.-P., eds., **9909**, 697 – 710, International Society for Optics and Photonics, SPIE (2016).

[12] Esposito, S., Riccardi, A., Fini, L., Puglisi, A. T., Pinna, E., Xompero, M., Briguglio, R., Quirós-Pacheco, F., Stefanini, P., Guerra, J. C., et al., "First light ao (flao) system for lbt: final integration, acceptance test in europe, and preliminary on-sky commissioning results," in [*Adaptive Optics Systems II*], **7736**, 773609, International Society for Optics and Photonics (2010).

[13] Wells, D. C., Greisen, E. W., and Harten, R. H., "FITS - a Flexible Image Transport System," **44**, 363 (June 1981).

[14] Delorme, P., Meunier, N., Albert, D., Lagadec, E., Coroller, H. L., Galicher, R., Mouillet, D., Boccaletti, A., Mesa, D., Meunier, J.-C., et al., "The sphere data center: a reference for high contrast imaging processing," *arXiv preprint arXiv:1712.06948* (2017).

[15] Pavlov, A., Möller-Nilsson, O., Feldt, M., Henning, T., Beuzit, J.-L., and Mouillet, D., "SPHERE data reduction and handling system: overview, project status, and development," in [*Advanced Software and Control for Astronomy II*], **7019**, 701939 (July 2008).

[16] Xuan, W. J., Mawet, D., Ngo, H., Ruane, G., Bailey, V. P., Choquet, É., Absil, O., Alvarez, C., Bryan, M., Cook, T., Femenía Castellá, B., Gomez Gonzalez, C., Huby, E., Knutson, H. A., Matthews, K., Ragland, S., Serabyn, E., and Zawol, Z., "Characterizing the Performance of the NIRC2 Vortex Coronagraph at W. M. Keck Observatory," **156**, 156 (Oct. 2018).

[17] Stone, J. M., Skemer, A. J., Hinz, P. M., Bonavita, M., Kratter, K. M., Maire, A.-L., Defrere, D., Bailey, V. P., Spalding, E., Leisenring, J. M., Desidera, S., Bonnefoy, M., Biller, B., Woodward, C. E., Henning, T., Skrutskie, M. F., Eisner, J. A., Crepp, J. R., Patience, J., Weigelt, G., De Rosa, R. J., Schlieder, J., Brandner, W., Apai, D., Su, K., Ertel, S., Ward-Duong, K., Morzinski, K. M., Schertl, D., Hofmann, K.-H., Close, L. M., Brems, S. S., Fortney, J. J., Oza, A., Buenzli, E., and Bass, B., "The LEECH Exoplanet Imaging Survey: Limits on Planet Occurrence Rates under Conservative Assumptions," **156**, 286 (Dec. 2018).

[18] Racine, R., Walker, G. A. H., Nadeau, D., Doyon, R., and Marois, C., "Speckle Noise and the Detection of Faint Companions," **111**, 587–594 (May 1999).

[19] Sparks, W. B. and Ford, H. C., "Imaging Spectroscopy for Extrasolar Planet Detection," **578**, 543–564 (Oct. 2002).

[20] Antichi, J., Gratton, R. G., Claudi, R. U., Giro, E., Mesa, D., Desidera, S., Scuderi, S., Bruno, P., Turatto, M., Cascone, E., De Caprio, V., Beuzit, J.-L., Puget, P., and Dohlen, K., "Sphere ifs optical concept description and design overview," (2008).

[21] Macintosh, B. A., Graham, J. R., Palmer, D. W., Doyon, R., Dunn, J., Gavel, D. T., Larkin, J., Oppenheimer, B., Saddlemyer, L., Sivaramakrishnan, A., Wallace, J. K., Bauman, B., Erickson, D. A., Marois, C., Poyneer, L. A., and Soummer, R., "The Gemini Planet Imager: from science to design to construction," in [*Adaptive Optics Systems*], **7015**, 701518 (July 2008).

[22] Wang, J. J., Perrin, M. D., Savransky, D., Arriaga, P., Chilcote, J. K., De Rosa, R. J., Millar-Blanchaer, M. A., Marois, C., Rameau, J., Wolff, S. G., Shapiro, J., Ruffio, J.-B., Maire, J., Marchis, F., Graham, J. R., Macintosh, B., Ammons, S. M., Bailey, V. P., Barman, T. S., Bruzzone, S., Bulger, J., Cotten, T., Doyon, R., Duchêne, G., Fitzgerald, M. P., Follette, K. B., Goodsell, S., Greenbaum, A. Z., Hibon, P., Hung, L.-W., Ingraham, P., Kalas, P., Konopacky, Q. M., Larkin, J. E., Marley, M. S., Metchev, S., Nielsen, E. L., Oppenheimer, R., Palmer, D. W., Patience, J., Poyneer, L. A., Pueyo, L., Rajan, A., Rantakyrö, F. T., Schneider, A. C., Sivaramakrishnan, A., Song, I., Soummer, R., Thomas, S., Wallace, J. K., Ward-Duong, K., and Wiktorowicz, S. J., "Automated data processing architecture for the Gemini Planet Imager Exoplanet Survey," *Journal of Astronomical Telescopes, Instruments, and Systems* **4**, 018002 (Jan. 2018).

[23] Perrin, M. D., Ingraham, P., Follette, K. B., Maire, J., Wang, J. J., Savransky, D., Arriaga, P., Bailey, V. P., Bruzzone, S., Chilcote, J. K., De Rosa, R. J., Draper, Z. H., Fitzgerald, M. P., Greenbaum, A. r. Z., Hung, L.-W., Konopacky, Q., Macintosh, B., Marchis, F., Marois, C., Millar-Blanchaer, M. A., Nielsen, E., Rajan, A., Rameau, J., Rantakyro, F. T., Ruffio, J.-B., Ward-Duong, K., Wolff, S. G., and Zalesky, J., "Gemini Planet Imager observational calibrations XI: pipeline improvements and enhanced calibrations after two years on sky," in [*Ground-based and Airborne Instrumentation for Astronomy VI*], Evans, C. J., Simard, L., and Takami, H., eds., *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **9908**, 990837 (Aug. 2016).

[24] Gonzalez, C. A. G., Wertz, O., Absil, O., Christiaens, V., Defrère, D., Mawet, D., Milli, J., Absil, P.-A., Van Droogenbroeck, M., Cantalloube, F., et al., "Vip: Vortex image processing package for high-contrast direct imaging," *The Astronomical Journal* **154**(1), 7 (2017).

[25] Amara, A. and Quanz, S. P., "PYNPOINT: an image processing package for finding exoplanets," **427**, 948–955 (Dec. 2012).

[26] Soummer, R., Pueyo, L., and Larkin, J., "Detection and Characterization of Exoplanets and Disks Using Projections on Karhunen-Loève Eigenimages," **755**, L28 (Aug. 2012).

[27] Burgasser, A. J., "The SpeX Prism Library: 1000+ low-resolution, near-infrared spectra of ultracool M, L, T and Y dwarfs," in [*Astronomical Society of India Conference Series*], *Astronomical Society of India Conference Series* **11**, 7–16 (Jan. 2014).

[28] Cantalloube, F., Mouillet, D., Mugnier, L. M., Milli, J., Absil, O., Gomez Gonzalez, C. A., Chauvin, G., Beuzit, J. L., and Cornia, A., "Direct exoplanet detection and characterization using the ANDROMEDA method: Performance on VLT/NaCo data," **582**, A89 (Oct. 2015).

[29] Galicher, R., Boccaletti, A., Mesa, D., Delorme, P., Gratton, R., Langlois, M., Lagrange, A.-M., Maire, A.-L., Le Coroller, H., Chauvin, G., et al., "Astrometric and photometric accuracies in high contrast imaging: The sphere speckle calibration tool (specal)," *Astronomy & Astrophysics* **615**, A92 (2018).

[30] Lafrenière, D., Marois, C., Doyon, R., Nadeau, D., and Artigau, É., "A New Algorithm for Point-Spread Function Subtraction in High-Contrast Imaging: A Demonstration with Angular Differential Imaging," **660**, 770–780 (May 2007).

[31] Marois, C., Correia, C., Galicher, R., Ingraham, P., Macintosh, B., Currie, T., and De Rosa, R., "Gpi psf subtraction with tloci: the next evolution in exoplanet/disk high-contrast imaging," in [*Adaptive Optics Systems IV*], **9148**, 91480U, International Society for Optics and Photonics (2014).

[32] Launhardt, R., Henning, T., Quirrenbach, A., Ségransan, D., Avenhaus, H., van Boekel, R., Brems, S. S., Cheetham, A. C., Cugno, G., Girard, J., et al., "Ispy-naco imaging survey for planets around young stars-survey description and results from the first 2.5 years of observations," *Astronomy & Astrophysics* **635**, A162 (2020).

[33] Mesa, D., Gratton, R., Zurlo, A., Vigan, A., Claudi, R. U., Alberi, M., Antichi, J., Baruffolo, A., Beuzit, J. L., Boccaletti, A., Bonnefoy, M., Costille, A., Desidera, S., Dohlen, K., Fantinel, D., Feldt, M., Fusco, T., Giro, E., Henning, T., Kasper, M., Langlois, M., Maire, A. L., Martinez, P., Moeller-Nilsson, O., Mouillet, D., Moutou, C., Pavlov, A., Puget, P., Salasnich, B., Sauvage, J. F., Sissa, E., Turatto, M., Udry, S., Vakili, F., Waters, R., and Wildi, F., "Performance of the VLT Planet Finder SPHERE. II. Data analysis and results for IFS in laboratory," **576**, A121 (Apr. 2015).

[34] Chauvin, G., Desidera, S., Lagrange, A.-M., Vigan, A., Feldt, M., Gratton, R., Langlois, M., Cheetham, A., Bonnefoy, M., and Meyer, M., "Shine, the sphere infrared survey for exoplanets," in [*SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics*], 331–335 (2017).

[35] Pairet, B., Cantalloube, F., Gomez Gonzalez, C. A., Absil, O., and Jacques, L., "STIM map: detection map for exoplanets imaging beyond asymptotic Gaussian residual speckle noise," **487**, 2262–2277 (Aug. 2019).

[36] Pairet, B., *Signal processing methods for high-contrast observations of planetary systems*, PhD thesis, Ph. D. thesis, Université de Louvain-la-Neuve, ICTEAM/ELEN (2020).

[37] Dahlqvist, C.-H., Cantalloube, F., and Absil, O., "Regime-switching model detection map for direct exoplanet detection in adi sequences," *Astronomy & Astrophysics* **633**, A95 (2020).

[38] Ren, B., Pueyo, L., Zhu, G. B., Debes, J., and Duchêne, G., "Non-negative matrix factorization: robust extraction of extended structures," *The Astrophysical Journal* **852**(2), 104 (2018).

[39] Gonzalez, C. G., Absil, O., Absil, P.-A., Van Droogenbroeck, M., Mawet, D., and Surdej, J., "Low-rank plus sparse decomposition for exoplanet detection in direct-imaging adi sequences-the llsg algorithm," *Astronomy & Astrophysics* **589**, A54 (2016).

[40] Ruffio, J.-B., Macintosh, B., Wang, J. J., Pueyo, L., Nielsen, E. L., De Rosa, R. J., Czekala, I., Marley, M. S., Arriaga, P., Bailey, V. P., et al., "Improving and assessing planet sensitivity of the gpi exoplanet survey with a forward model matched filter," *The Astrophysical Journal* **842**(1), 14 (2017).

[41] Wang, J. J., Ruffio, J.-B., De Rosa, R. J., Aguilar, J., Wolff, S. G., and Pueyo, L., "pyKLIP: PSF Subtraction for Exoplanets and Disks," (June 2015).

[42] Nielsen, E. L., De Rosa, R. J., Macintosh, B., Wang, J. J., Ruffio, J.-B., Chiang, E., Marley, M. S., Saumon, D., Savransky, D., Ammons, S. M., Bailey, V. P., Barman, T., Blain, C., Bulger, J., Burrows, A., Chilcote, J., Cotten, T., Czekala, I., Doyon, R., Duchêne, G., Esposito, T. M., Fabrycky, D., Fitzgerald, M. P., Follette, K. B., Fortney, J. J., Gerard, B. L., Goodsell, S. J., Graham, J. R., Greenbaum, A. Z., Hibon, P., Hinkley, S., Hirsch, L. A., Hom, J., Hung, L.-W., Dawson, R. I., Ingraham, P., Kalas, P., Konopacky, Q., Larkin, J. E., Lee, E. J., Lin, J. W., Maire, J., Marchis, F., Marois, C., Metchev, S., Millar-Blanchaer, M. A., Morzinski, K. M., Oppenheimer, R., Palmer, D., Patience, J., Perrin, M., Poyneer, L., Pueyo, L., Rafikov, R. R., Rajan, A., Rameau, J., Rantakyrö, F. T., Ren, B., Schneider, A. C., Sivaramakrishnan, A., Song, I., Soummer, R., Tallis, M., Thomas, S., Ward-Duong, K., and Wolff, S., "The Gemini Planet Imager Exoplanet Survey: Giant Planet and Brown Dwarf Demographics from 10 to 100 au," **158**, 13 (July 2019).

[43] Flasseur, O., Denis, L., Thiébaut, É., and Langlois, M., "Exoplanet detection in angular differential imaging by statistical learning of the nonstationary patch covariances. The PACO algorithm," **618**, A138 (Oct. 2018).

[44] Samland, M., Bouwman, J., Hogg, D. W., Brandner, W., Henning, T., and Janson, M., "TRAP: A temporal systematics model for improved direct detection of exoplanets at small angular separations," *arXiv e-prints* , arXiv:2011.12311 (Nov. 2020).

[45] Gonzalez, C. G., Absil, O., and Van Droogenbroeck, M., "Supervised detection of exoplanets in high-contrast imaging sequences," *Astronomy & Astrophysics* **613**, A71 (2018).

[46] Flasseur, O., Denis, L., Thiébaut, É., and Langlois, M., "Paco asdi: an algorithm for exoplanet detection and characterization in direct imaging with integral field spectrographs," *Astronomy & Astrophysics* **637**, A9 (2020).

[47] Kasper, M., Arsenault, R., Käufl, U., Jakob, G., Leveratto, S., Zins, G., Pantin, E., Duhoux, P., Riquelme, M., Kirchbauer, J. P., Kolb, J., Pathak, P., Siebenmorgen, R., Soenke, C., Fuenteseca, E., Sterzik, M., Ageorges, N., Gutruf, S., Kampf, D., Reutlinger, A., Absil, O., Delacroix, C., Maire, A. L., Huby, E., Guyon, O., Klupar, P., Mawet, D., Ruane, G., Karlsson, M., Dohlen, K., Vigan, A., N'Diaye, M., Quanz, S., and Carlotti, A., "NEAR: First Results from the Search for Low-Mass Planets in α Cen," *The Messenger* **178**, 5–9 (Dec. 2019).

[48] Hoeijmakers, H. J., Schwarz, H., Snellen, I. A. G., de Kok, R. J., Bonnefoy, M., Chauvin, G., Lagrange, A. M., and Girard, J. H., "Medium-resolution integral-field spectroscopy for high-contrast exoplanet imaging. Molecule maps of the β Pictoris system with SINFONI," **617**, A144 (Oct. 2018).

[49] Haffert, S. Y., Bohn, A. J., de Boer, J., Snellen, I. A. G., Brinchmann, J., Girard, J. H., Keller, C. U., and Bacon, R., "Two accreting protoplanets around the young star PDS 70," *Nature Astronomy* **3**, 749–754 (June 2019).