A New Statistical Model of Star Speckles for Learning to Detect and Characterize Exoplanets in Direct Imaging Observations

Supplementary Material

A. Technical details

A.1. Estimation of the statistical parameters

In this section, we provide additional information regarding the estimation of the statistical parameters of the Gaussian distribution (see Sec. 4.1). These parameters are estimated by maximum likelihood:

$$\widehat{\boldsymbol{m}}_j = \frac{1}{T}\sum_t \boldsymbol{y}_t^{(j)},\ \widehat{\mathbf{S}}_j = \frac{1}{T}\sum_t \left(\boldsymbol{y}_t^{(j)} - \widehat{\boldsymbol{m}}_j\right)\left(\boldsymbol{y}_t^{(j)} - \widehat{\boldsymbol{m}}_j\right)^\top,$$

where $\widehat{\mathbf{S}}_j$ in $\mathbb{R}^{p\times p}$ is the sample covariance matrix. Given that $\widehat{\mathbf{S}}_j$ usually suffers from a large variance due to small sample statistics (i.e., $p \geq T \simeq 70$), we use a shrinkage estimator leading to the final covariance estimator $\widehat{\mathbf{C}}_j$:

$$\widehat{\mathbf{C}}_j = (1-\widehat{\rho}_j)\,\widehat{\mathbf{S}}_j + \widehat{\rho}_j\,\widehat{\mathbf{D}}_j\,, \tag{23}$$

where $\widehat{\mathbf{D}}_j$ in $\mathbb{R}^{p\times p}$ is the low-variance/high-bias diagonal matrix formed with diagonal elements of $\widehat{\mathbf{S}}_j$, and $\widehat{\rho}_j$ in $[0,\,1]$. Hyper-parameter $\widehat{\rho}_j$ can be estimated optimally in a data-driven fashion by risk minimization between the true (unknown) covariance $\mathbf{C}_j$ and its shrunk counterpart $\widehat{\mathbf{C}}_j$. This leads to the closed-form expression [29]:

$$\widehat{\rho}_j = \frac{\operatorname{tr}(\widehat{\mathbf{S}}_j^2) + \operatorname{tr}^2(\widehat{\mathbf{S}}_j) - 2\operatorname{tr}(\widehat{\mathbf{S}}_j \circ \widehat{\mathbf{S}}_j)}{(T+1)(\operatorname{tr}(\widehat{\mathbf{S}}_j^2) - \operatorname{tr}(\widehat{\mathbf{S}}_j \circ \widehat{\mathbf{S}}_j))}\,, \tag{24}$$

where $\circ$ stands for Hadamard (element-wise) product.

A.2. Dense detection map computation details

In this section, we provide additional details on the computation of the dense detection map introduced in Sec. 4.1. To obtain a full-frame detection map, $\widehat{\alpha}$ and $\widehat{\sigma}_\alpha$ must be computed for all spatial coordinates, which can be formalized by:

$$\forall \boldsymbol{x}_i \in \mathcal{G}_p,\ \widehat{\alpha}_i = \arg\max_{\alpha_i} \ell(\alpha_i, \boldsymbol{x}_i)\,. \tag{25}$$

Note that it corresponds to solve independent optimization problems at each location i. To efficiently solve Eq. (25), we employ a fast approximation to compute the terms $a(\boldsymbol{x}_t)$ and $b_t(\boldsymbol{x}_t)$ at each time step and for every initial position. First, the terms a_t and b_t are computed for all positions in the grid $\mathcal{G}_p$ of spatial pixels. In the following, we note $b_{ti} = b_t(\boldsymbol{x}_i)$, $a_i = a(\boldsymbol{x}_i)$ and $\boldsymbol{h}_{ji} = \boldsymbol{h}^{(j)}(\boldsymbol{x}_i)$ for all $\boldsymbol{x}_i$ in $\mathcal{G}_p$, such that:

$$\forall (i,t), b_{it} = \sum_{j\in S(\boldsymbol{x}_i)} w_j\,\boldsymbol{h}_{ji}^\top\,\widehat{\mathbf{C}}_j^{-1}\left(\boldsymbol{y}_t^{(j)} - \widehat{\boldsymbol{m}}_j\right)\,, \tag{26}$$

$$\forall i, a_i = \sum_{j\in S(\boldsymbol{x}_i)} w_j\,\boldsymbol{h}_{ji}^\top\,\widehat{\mathbf{C}}_j^{-1}\,\boldsymbol{h}_{ji}\,. \tag{27}$$

With our convolutional approach, these terms can be computed efficiently by Cholesky factorization of the precision matrix, as detailed in Sec. A.3. Let $\boldsymbol{b}_t$ in $\mathbb{R}^{H\times W}$ and $\boldsymbol{a}$ in $\mathbb{R}^{H\times W}$ represent the resulting quantities. Then, $b_t(\boldsymbol{x}_t)$ and $a(\boldsymbol{x}_t)$ can be approximated by spatial interpolation of $\boldsymbol{b}_t$ and $\boldsymbol{a}$. However, the trajectories of all exoplanets in the observation are governed by a common parallactic rotation vector $\boldsymbol{\phi}$ in $\mathbb{R}^T$. So in practice, this interpolation can be computed efficiently for each time-step t, by rotating the frames $\boldsymbol{a}$ and $\boldsymbol{b}_t$ by an angle $-\phi_t$. Thus, a dense detection map $\widehat{\boldsymbol{\gamma}}$ in $\mathbb{R}^{H\times W}$ can be obtained with:

$$\widehat{\boldsymbol{\gamma}} = \frac{\widehat{\boldsymbol{\alpha}}}{\widehat{\boldsymbol{\sigma}}} = \frac{\sum_t \mathbf{R}_{-\phi_t}(\boldsymbol{b}_t)}{\sqrt{\sum_t \mathbf{R}_{-\phi_t}(\boldsymbol{a})}}\,, \tag{28}$$

where $\mathbf{R}_{-\phi_t} : \mathbb{R}^{H\times W} \to \mathbb{R}^{H\times W}$ is the frame rotation operator with angle $-\phi_t$, and $\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\sigma}}$ in $\mathbb{R}^{H\times W}$ such that:

$$\widehat{\boldsymbol{\alpha}} = \frac{\sum_t \mathbf{R}_{-\phi_t}(\boldsymbol{b}_t)}{\sum_t \mathbf{R}_{-\phi_t}(\boldsymbol{a})} \quad \text{and} \quad \widehat{\boldsymbol{\sigma}} = \frac{1}{\sqrt{\sum_t \mathbf{R}_{-\phi_t}(\boldsymbol{a})}}\,. \tag{29}$$

A.3. Fast computation of b_t and a

In this section, we detail an efficient strategy to derive a dense two-dimensional detection map $\widehat{\boldsymbol{\gamma}}$ serving as the final quantity to detect exoplanets, see Sec. 4.1.

The terms $\boldsymbol{b}_t$ and $\boldsymbol{a}$ in $\mathbb{R}^{H\times W}$ involved in $\widehat{\boldsymbol{\gamma}}$ (see Eqs. (26)–(28)) can be computed efficiently through:

$$\boldsymbol{b}_t = \sum_j w_j\,\mathbf{Q}_j(\boldsymbol{b}_t^{(j)})\,, \tag{30}$$

$$\boldsymbol{a} = \sum_j w_j\,\mathbf{Q}_j(\boldsymbol{a}^{(j)})\,, \tag{31}$$

where $\mathbf{Q}_j$ is the linear operator that places the patch indexed by j in the correct position in the image, and $\boldsymbol{b}_t^{(j)}$, $\boldsymbol{a}^{(j)}$ are in $\mathbb{R}^p$. The individual terms $\boldsymbol{a}^{(j)}$ and $\boldsymbol{b}_t^{(j)}$ are given by:

$$\boldsymbol{b}_t^{(j)} = \mathbf{P}^\top\,\widehat{\mathbf{C}}_j^{-1}\left(\boldsymbol{y}_t^{(j)} - \widehat{\boldsymbol{m}}_j\right)\,, \tag{32}$$

$$\boldsymbol{a}^{(j)} = \operatorname{diag}\left(\mathbf{P}^\top\,\widehat{\mathbf{C}}_j^{-1}\mathbf{P}\right)\,, \tag{33}$$

where $\mathbf{P}$ in $\mathbb{R}^{p\times p}$ is a circulant matrix, with each column representing the flattened PSF centered on a different pixel. Equation (32) can be implemented efficiently with a convolution operation. Implementing Eq. (33) directly is however computationally challenging as it requires storing the matrix $\mathbf{P}$, which is highly inefficient, especially for large patches. To circumvent this issue, we propose to decompose $\boldsymbol{a}^{(j)}$ as:

$$\boldsymbol{a}^{(j)} = \mathbf{N}(\widehat{\mathbf{L}}_j\mathbf{P})\,, \tag{34}$$

where $\mathbf{N} : \mathbb{R}^{p\times p} \to \mathbb{R}^p$ is the operator that computes the column-wise squared ℓ_2–norm, and the matrix $\widehat{\mathbf{L}}_j$ in $\mathbb{R}^{p\times p}$ is the lower triangular matrix obtained via Cholesky decomposition of the precision matrix (i.e., $\widehat{\mathbf{C}}_j^{-1} = \widehat{\mathbf{L}}_j\widehat{\mathbf{L}}_j^\top$). The formulation of Eq. (34) allows implementing the matrix multiplication with $\mathbf{P}$ efficiently as a convolution operation.

Algorithm 1 Iterative procedure for flux and position estimation.

Require: $\boldsymbol{y} \in \mathbb{R}^{T\times H\times W}$, $\boldsymbol{z} \in \mathbb{R}^3$
  for $l = 0 \cdots L-1$ **do**
    $\alpha, \boldsymbol{x}_0 \leftarrow \boldsymbol{z}$
    for $t = 0 \cdots T-1$ **do** ▷ Update nuisance estimate
      $\boldsymbol{x}_t \leftarrow r(\boldsymbol{x}_0, \phi_t)$
      $\boldsymbol{s}_t \leftarrow \boldsymbol{y}_t - \boldsymbol{h}(\boldsymbol{x}_t)$
    end for
    $\boldsymbol{g}, \mathbf{H} \leftarrow \mathbf{0}_{\mathbb{R}^3}, \mathbf{0}_{\mathbb{R}^{3\times 3}}$
    for $t = 0 \cdots T-1$ **do** ▷ Gradient and Hessian
      for $j \in S(\boldsymbol{x}_t)$ **do**
        $\widehat{\boldsymbol{m}}_j, \widehat{\mathbf{C}}_j \leftarrow \text{Estim}(\{\boldsymbol{s}_t^{(j)}\}_t)$ ▷ Eqs. (23)–(24)
        $\boldsymbol{g} \leftarrow \boldsymbol{g} + w_j \frac{\partial \ell_j}{\partial \boldsymbol{z}}(\boldsymbol{z}, \widehat{\boldsymbol{m}}_j, \widehat{\mathbf{C}}_j)$
        $\mathbf{H} \leftarrow \mathbf{H} + w_j \frac{\partial^2 \ell_j}{\partial \boldsymbol{z}^2}(\boldsymbol{z}, \widehat{\boldsymbol{m}}_j, \widehat{\mathbf{C}}_j)$
      end for
    end for
    $\boldsymbol{z} \leftarrow \text{Proj}_Z(\boldsymbol{z} - \mathbf{H}^{-1}\boldsymbol{g})$ ▷ Projection on feasible set
  end for
  return $\boldsymbol{z}$

A.4. Iterative optimization scheme for astrometry

We detail in this section the characterization procedure mentioned in Sec. 4.1 for precisely estimating the flux α and the two-dimensional sub-pixel position $\boldsymbol{x}_0$ of an exoplanet. This procedure is described by Algorithm 1 in pseudo-code form, where $\boldsymbol{z} = [\alpha, \boldsymbol{x}_0]$ in $\mathbb{R}^3$ is the variable to be refined. The position is initialized to the coordinates of the pixel with the highest detection score from the detection step, while the flux is initialized using the value extracted from the flux map $\boldsymbol{\alpha}$ at that location. We decompose the involved loss function ℓ as:

$$\ell(\boldsymbol{z}, \{\boldsymbol{m}_j, \mathbf{C}_j\}_j) = \sum_t \sum_{j\in S(\boldsymbol{x}_t)} w_j \ell_j(\boldsymbol{z}, \boldsymbol{m}_j, \mathbf{C}_j)\,, \tag{35}$$

where:

$$\ell_j([\alpha, \boldsymbol{x}_0], \boldsymbol{m}_j, \mathbf{C}_j) = \left\| \mathbf{L}_j \left(\boldsymbol{y}_t^{(j)} - \alpha \boldsymbol{h}^{(j)}(r(\boldsymbol{x}_0, \phi_t)) - \boldsymbol{m}_j \right) \right\|_2^2 , \tag{36}$$

such that $\mathbf{C}_j^{-1} = \mathbf{L}_j \mathbf{L}_j^\top$. We also define the feasible set Z such that $\alpha > 0$ (enforcing a non-negativity constraint on the exoplanet flux) and the estimated position remains within a fixed radius of the initial estimate. We denote by Proj_Z the projection associated with it. In Algorithm 1, we also denote by Estim(.) the procedure described in Eqs. (23)–(24) to estimate the statistical parameters of the nuisance component.

A.5. Prior information on the exoplanet spectrum

The channel-wise detection maps $\widetilde{\boldsymbol{\gamma}}_c = \widetilde{\boldsymbol{b}}_c / \sqrt{\widetilde{\boldsymbol{a}}_c}$ are combined into a single detection map by averaging $\widetilde{\boldsymbol{a}}_c$ and $\widetilde{\boldsymbol{b}}_c$ with weights e_c in $\mathbb{R}^+$ satisfying $\sum_c e_c = 1$ such that:

$$\widetilde{\boldsymbol{a}}_C = \sum_c e_c \widetilde{\boldsymbol{a}}_c \quad \text{and} \quad \widetilde{\boldsymbol{b}}_C = \sum_c e_c \widetilde{\boldsymbol{b}}_c\,. \tag{37}$$

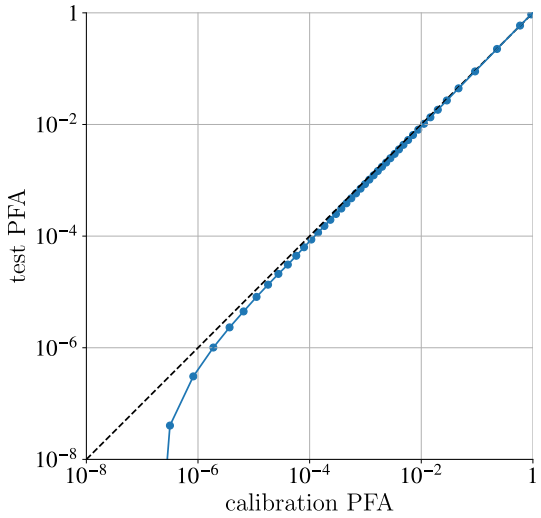


Figure F. Calibrated PFA versus test PFA: the calibration procedure allows obtaining a reliable approximation of the PFA, except in the extreme low-PFA regime (high detection score under $\mathcal{H}_0$) where samples are exceedingly rare.

The final detection score, spectrally combined over C channels, is $\widetilde{\boldsymbol{\gamma}}_C = \widetilde{\boldsymbol{b}}_C / \sqrt{\widetilde{\boldsymbol{a}}_C}$ in $\mathbb{R}^{H\times W}$. The weights e_c reflect prior assumptions about the exoplanet spectrum, and can be used (if available) to boost the detection of exoplanets having similar spectrum, see e.g., [12, 26]. For all experiments and baselines, we assumed a non-informative flat prior (i.e., $e_c = 1/C$). If additional information about the exoplanet's spectrum is available, the prior can be adjusted accordingly.

A.6. Calibration procedure

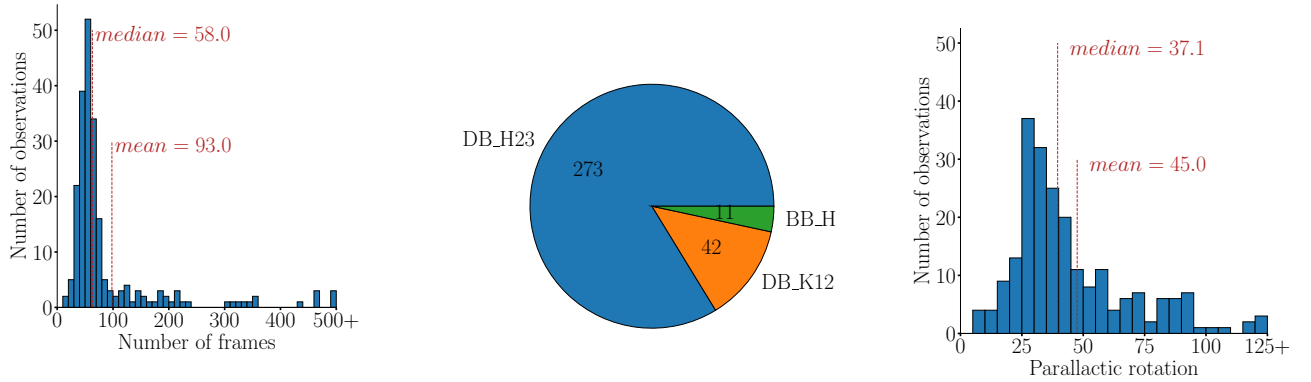
This section outlines the calibration procedure for the full model described in Sec. 4.3, following the approach proposed in [5], which we briefly summarize here.

We aim to calibrate the predicted probability of false alarm (PFA), a critical metric for astronomers to quantify detection uncertainty. This requires characterizing the output distribution of our method under the null hypothesis $\mathcal{H}_0$ (i.e., when no source is present), which we denote $\widetilde{\gamma}|\mathcal{H}_0$. For that purpose, we use a calibration dataset $\mathcal{D}_{\text{calib}}$, consisting of 15 observations excluded from training, validation, and testing. Temporal shuffling of the frames and parallactic rotation reversal are applied to augment the dataset and ensure no real source is present. These custom data-augmentations disrupt the temporal coherence of any (potentially unknown) real exoplanets, preventing them from biasing the training and calibration processes. We apply the model to 10,000 cubes from $\mathcal{D}_{\text{calib}}$, concatenate all output pixels into a single vector $\widetilde{\boldsymbol{\gamma}}_{\text{calib}}$ of length n_{calib}, and compute the empirical cumulative distribution function (eCDF):

$$\forall\, \tau \in \mathbb{R}, \quad \widehat{F}_{\widetilde{\gamma}|\mathcal{H}_0}(\tau) := \frac{1}{n_{\text{calib}}} \sum_{i=0}^{n_{\text{calib}}-1} \mathbf{1}_{\widetilde{\gamma}_{\text{calib},i} \le \tau}\,. \tag{38}$$

where $\mathbf{1}$ is the indicator function. The estimated probability of false alarm is related to the eCDF through:

$$\forall\, \tau \in \mathbb{R}, \quad \widehat{\text{PFA}}(\tau) = \widehat{\mathbb{P}}(\widetilde{\gamma} > \tau | \mathcal{H}_0) = 1 - \widehat{F}_{\widetilde{\gamma}|\mathcal{H}_0}(\tau)\,. \tag{39}$$



(a) Distribution of the number of frames per observation

(b) Repartition of the filters used

(c) Distribution of the parallactic rotation per observation

Figure G. Additional details on the F150 dataset

In Fig. F, we evaluate the PFA across a range of detection thresholds using the calibration set and apply the same thresholds to the test set for comparison. The results show that the calibrated PFA closely matches the test PFA, except in the extremely low-PFA regime. This slight discrepancy is expected, as high detection scores under $\mathcal{H}_0$ are exceedingly rare, leading to increased noise in PFA estimation. In this regime, our estimate is nevertheless conservative, i.e., the experienced PFA is slightly smaller than predicted one.

A.7. Implementation details

Our model implementation and training are based on the PyTorch framework. We optimize the neural network weights using the Adam optimizer with a learning rate of 5×10^{-4}, a batch size of 16, and 50,000 iterations. We apply data augmentation techniques, including random 90° rotations, vertical/horizontal flipping, and frame shuffling. Bicubic interpolation is used for spectral and exoplanet alignment. For ADI, the model is trained on 64 frames, while for ASDI, data cubes have 32 frames with two channels each. All evaluations and tests are done with 64 frames in both ADI and ASDI. To ensure accurate results, pixels containing real known exoplanets are masked out during training before the temporal summation in Eq. (29). The loss is calculated on the final output for pixels with at least 8 valid frames, unaffected by real exoplanets.

B. Dataset description

SHINE-F150 is a subset of the SPHERE large survey, with 322 data cubes ($20,462$ images in total) from 150 stars. The spatial resolution is 12 milliarcseconds/pixel. The spectral resolution λ/Δ_λ is 20. Detailed description is in [39]. We followed [5] to exclude corrupted data, leaving 220 datasets, with 165 used for training. The main statistics of the F150 dataset are shown in Fig. G.

C. Comparison of state-of-the-art methods

Method	nuisance model	P1	P2	P3	P4	P5	P6	P7
PACO	statistical on data patches	X	✓	✓	✓	X	✓	✓
SODINN	learned on PCA residuals	X	X	X	✓✓	X	X	X
deep PACO	learned after PACO whitening	X	X	✓	✓✓	X	✓	X
Super-RDI	PCA	✓	✓	X	✓✓	X	X	✓
MWIN5-RB	learned on PCA residuals	✓	✓	X	✓✓	✓	X	✓✓
ConStruct	generative, learned on data	✓	✓	X	✓✓	✓	X	✓✓✓
MODEL&CO	statistical on learned features	✓	X	X	✓✓✓	✓✓	X	✓✓✓
proposed	multi-scale stat. on learned features	✓	✓	✓	✓✓✓	✓✓✓	✓	✓✓✓

Table D. Comparison of advanced exoplanet imaging algorithms based on key desirable properties. P1=observation-independence of the model, P2=physical interpretability, P3=multi-spectral, P4=detection sensitivity, P5=accuracy near the star, P6=position and flux estimation, P7=practicality on large-scale surveys.

Table D compares the key properties of state-of-the-art algorithms for exoplanet detection via direct imaging.

D. Additional results

In this section we provide additional visual results, and show detection maps obtained with synthetic exoplanets. Consistent with the experimental findings in the main body of the paper, our method performs on par or better than baselines in ADI, and outperforms other approaches in ASDI.

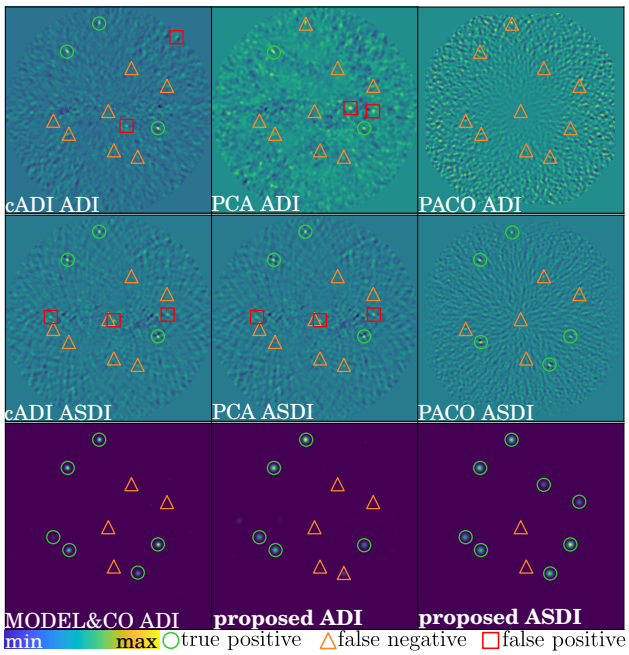


Figure H. Detection maps on observations of HD 102647 star with synthetic exoplanets.

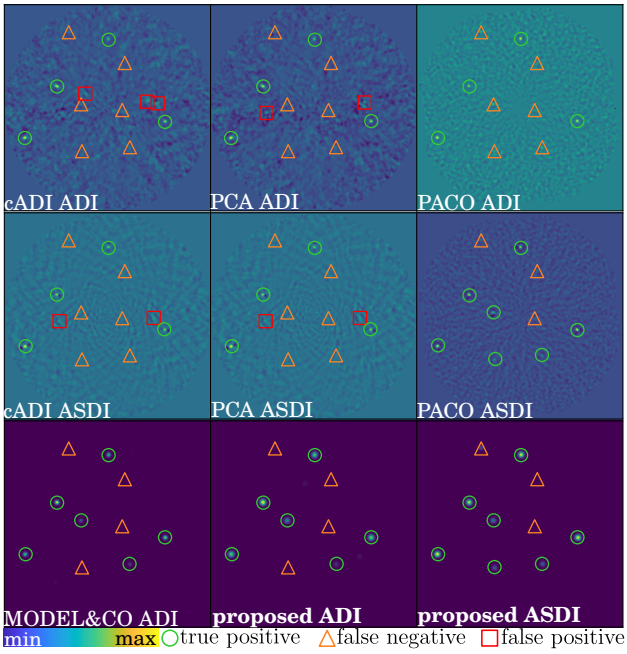


Figure J. Detection maps on observations of HD 206860 star with synthetic exoplanets.

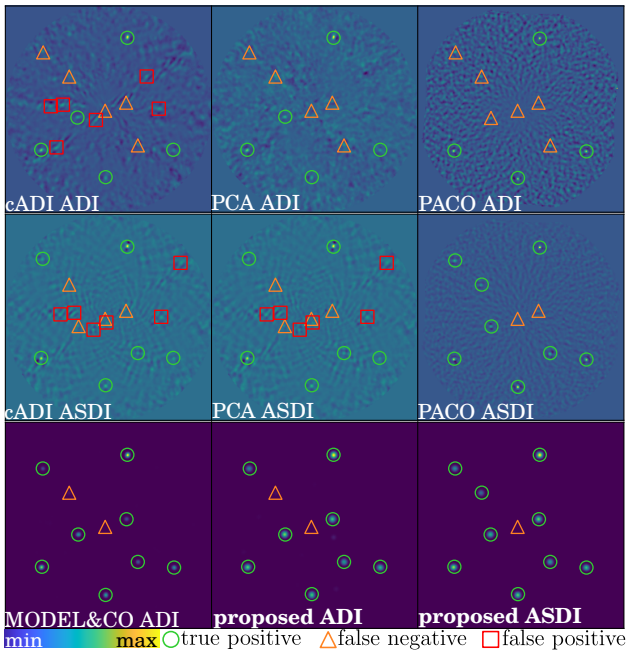


Figure I. Detection maps on observations of HD 188228 star with synthetic exoplanets.

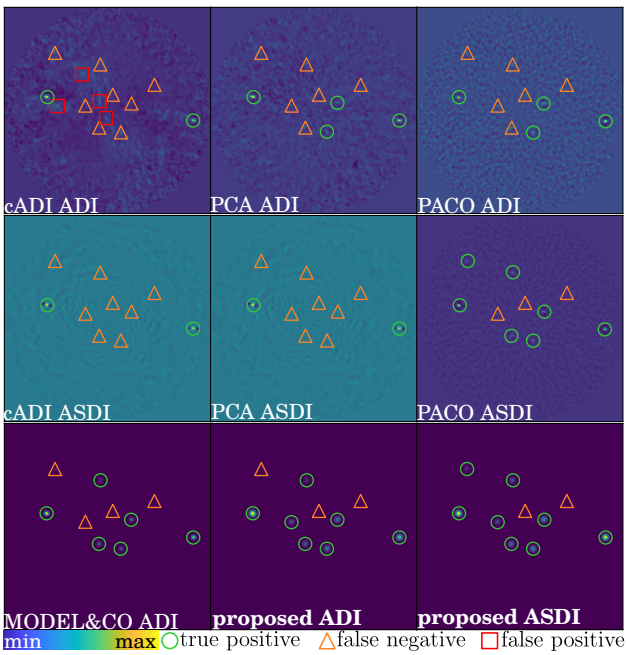


Figure K. Detection maps on observations of HD 216803 star with synthetic exoplanets.