Reliable detection of small long-period planets in Kepler data

Oryna Ivashtenko, Barak Zackay

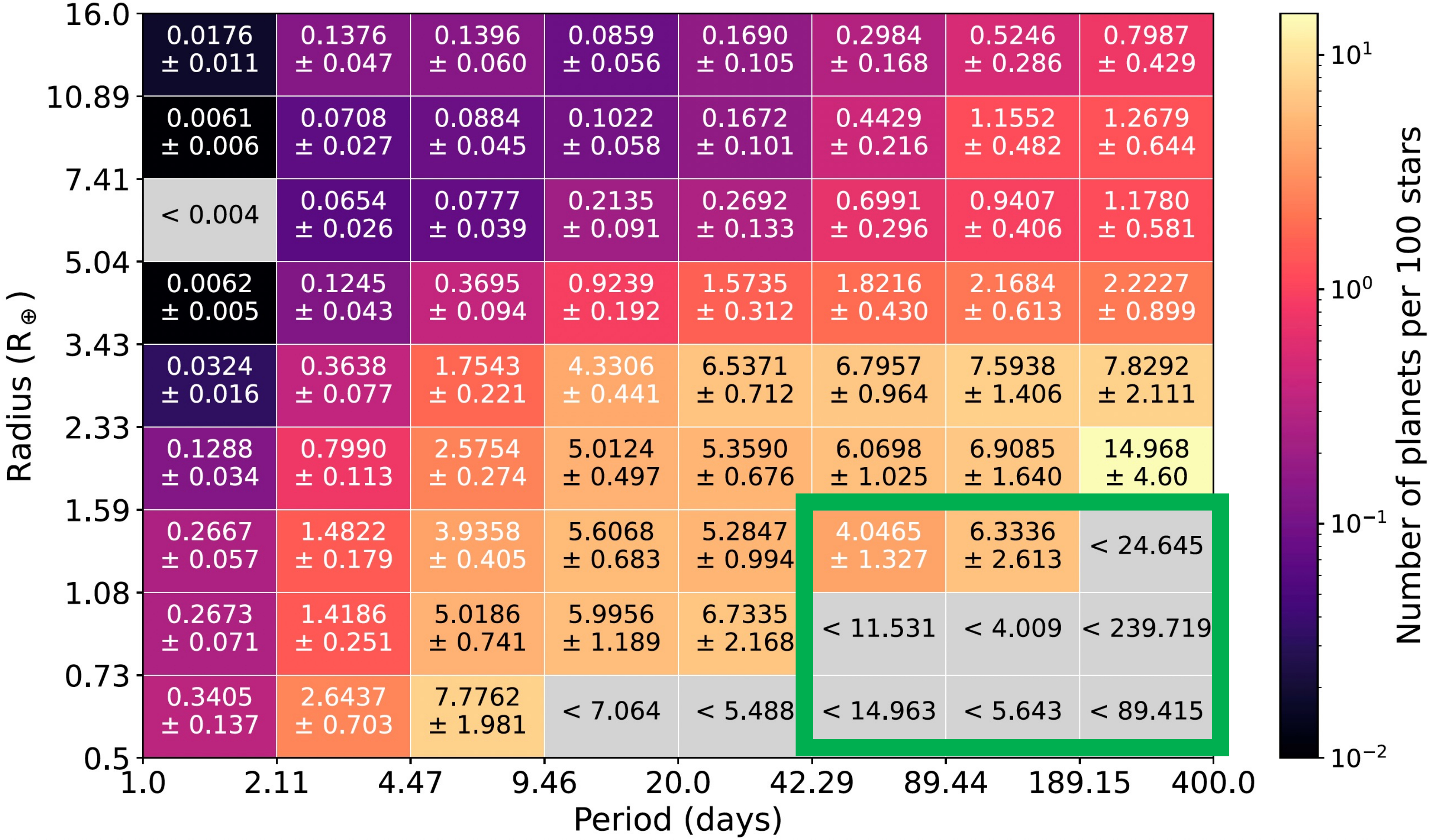
Weizmann Institute of Science

June 2025

Current exoplanet population from *Kepler*

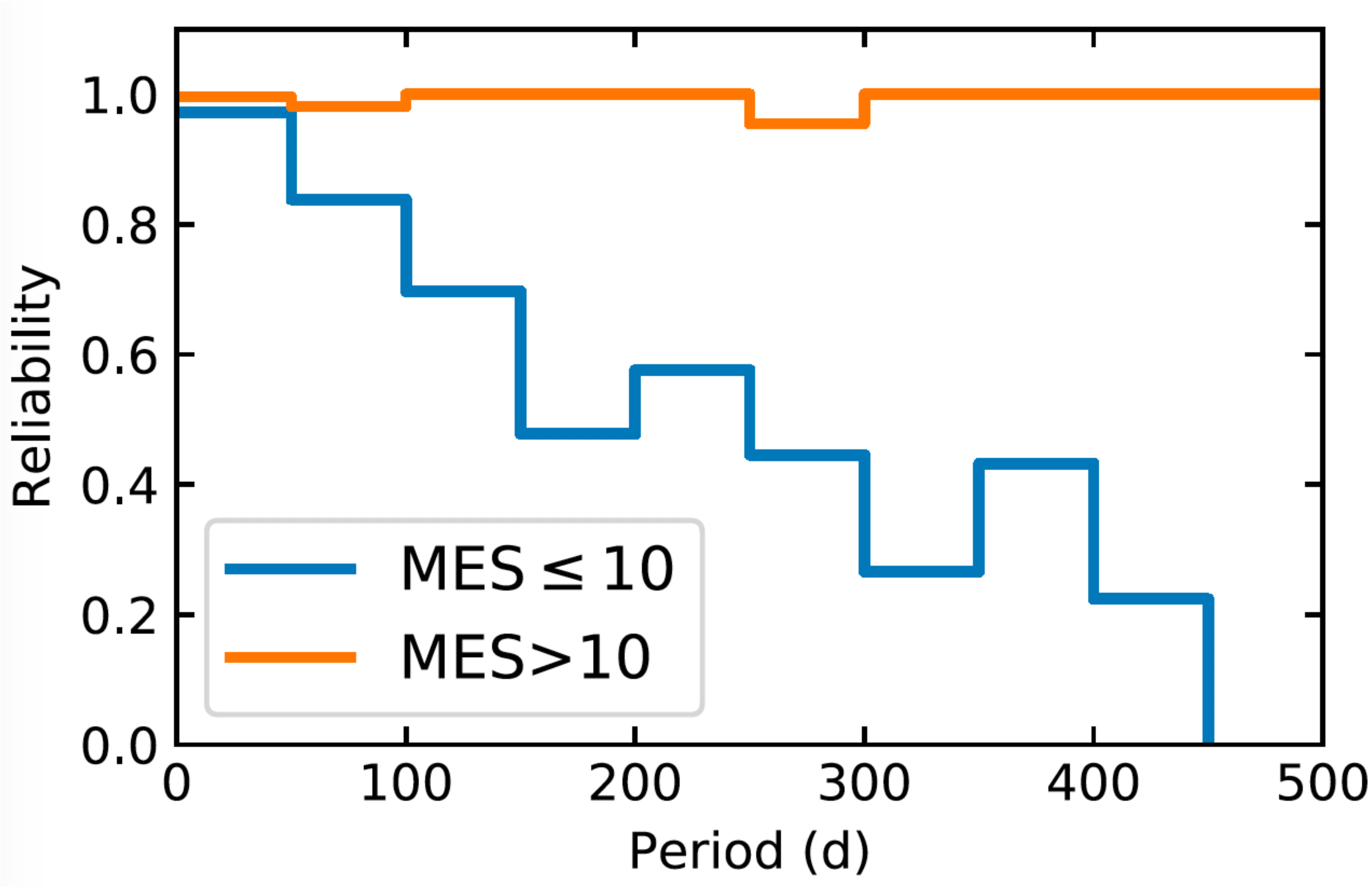
Kepler Exoplanetary occurrence rates around FGK stars:

[Dattilo, Batalha, Bryson 2023]



Reliability challenge of *Kepler* Catalog

- **Reliability:** Fraction of catalog events that are real planets

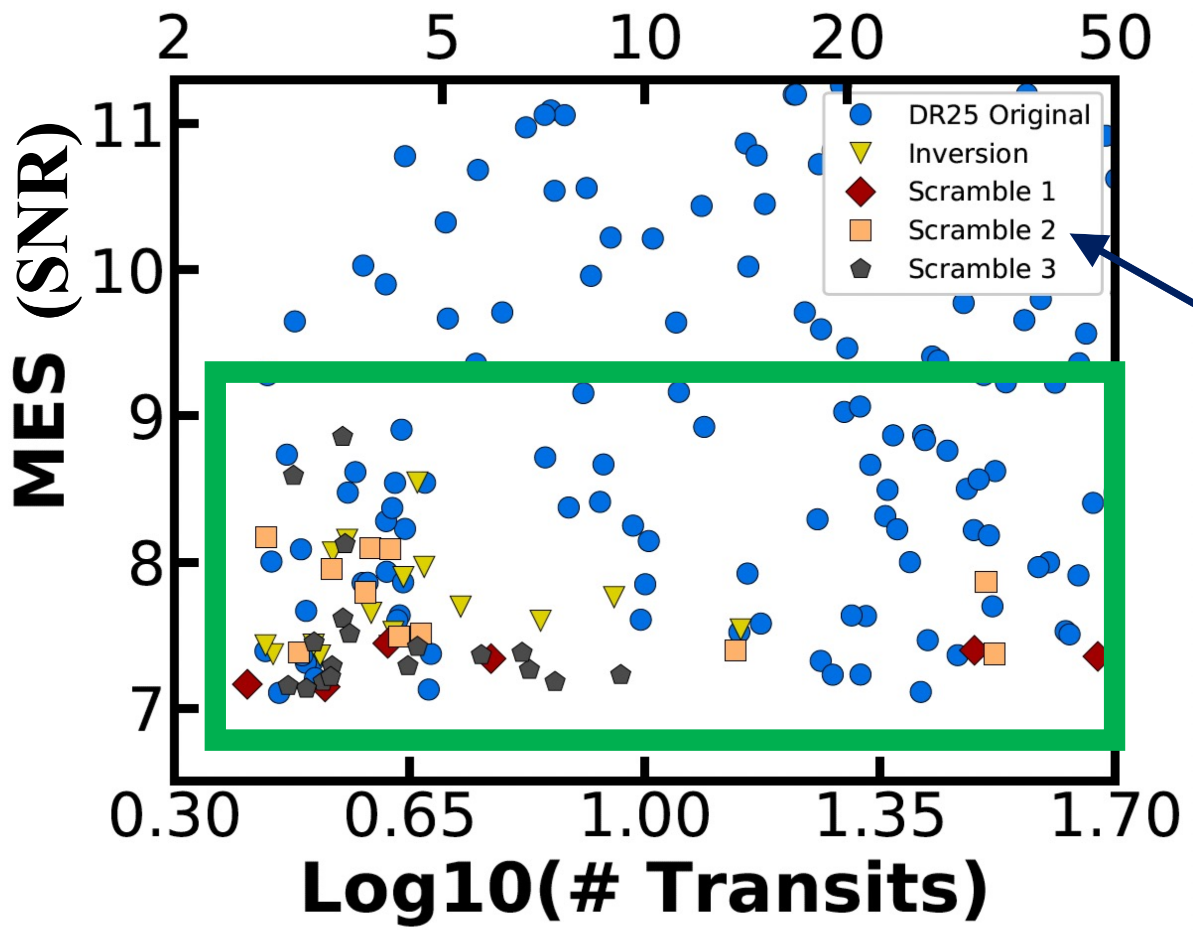


[Thompson 2018]

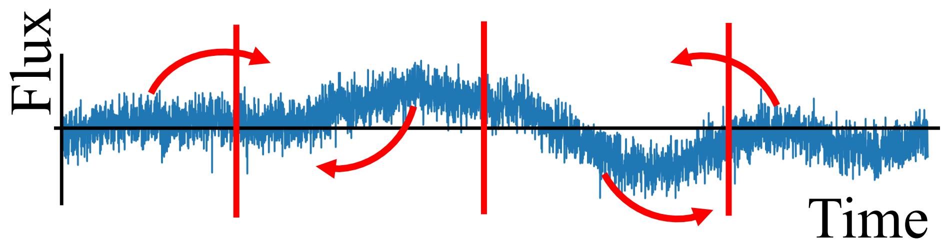
Challenge: false alarms at low MES, long period

- **MES**: Multiple-event statistic (SNR of multiple-transit signal)

[Burke 2019]



Scrambling:

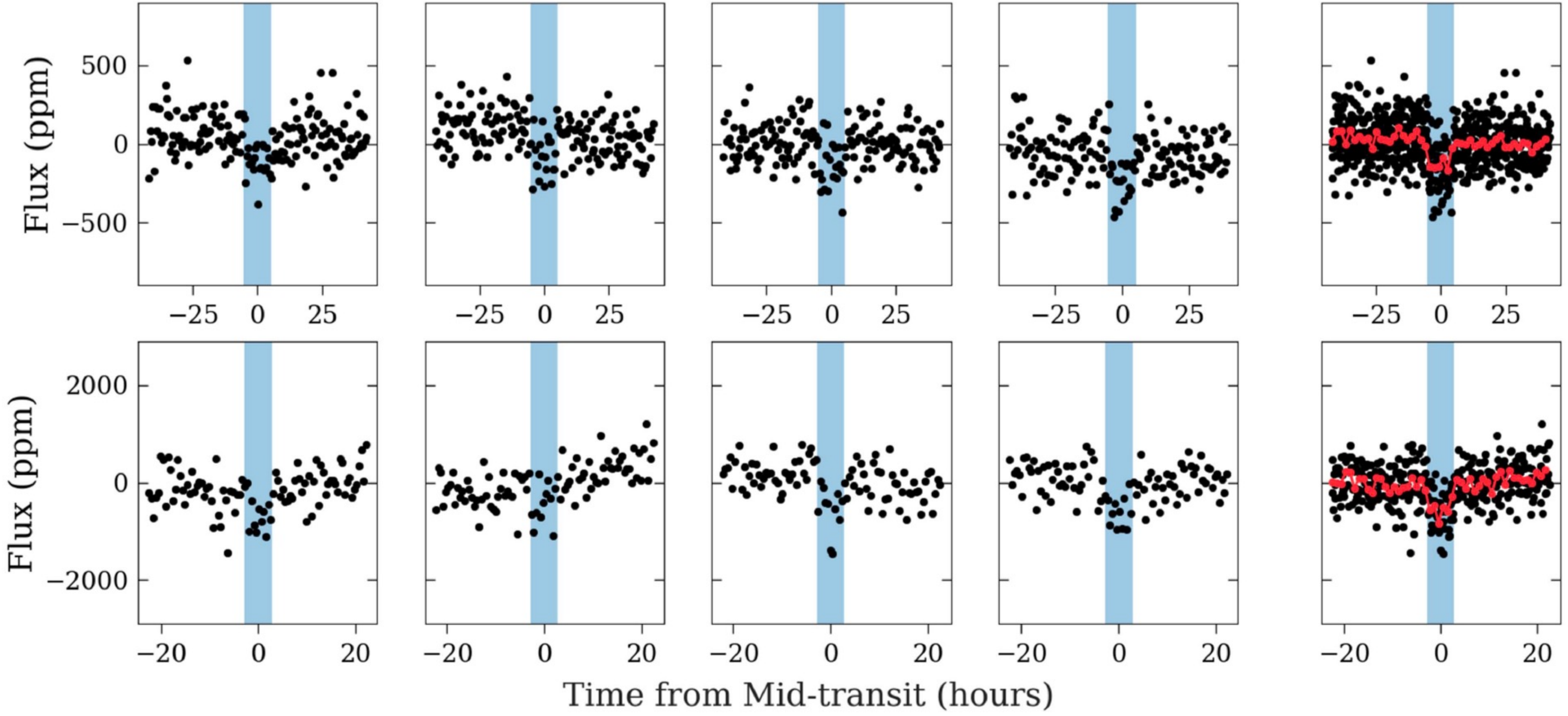


- Destroy periodicity
- Preserve noise properties

False Alarms

[Source: Mullaly 2018]

Example of how a false alarm in inverted light curve can mimic a confirmed planet (the reader was invited to guess which is which)

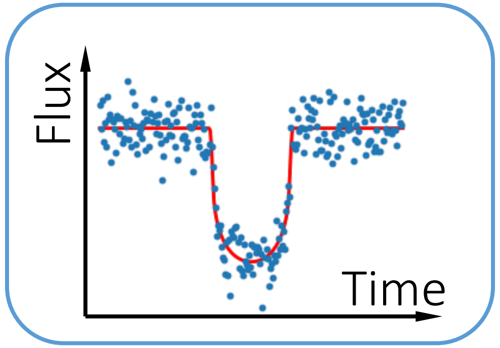


Our Independent search pipeline

- **Goal:** Estimate the population of small long-period planets with *Kepler*
- **Focus of the pipeline:** Search reliability (control false alarms)
- **Main methods:**
  - + Exact detection statistic
  - + Background distribution control
  - + Empirical significance estimation

Pipeline methods

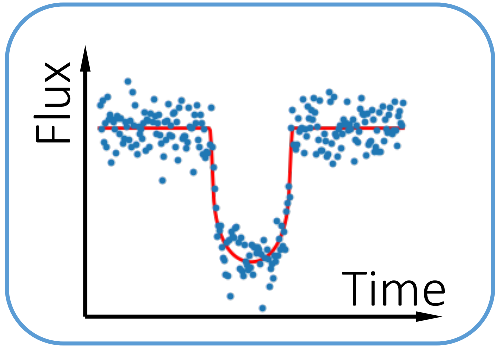
Single-transit detection statistic



Data model: $\vec{d} = A\vec{h} + \mathcal{N}(0, \boldsymbol{C})$

- $\vec{d}$: Flux
- $\vec{h}$: Template
- $\boldsymbol{C}$: Covariance matrix

Single-transit detection statistic



Data model: $\vec{d} = A\vec{h} + \mathcal{N}(0, \boldsymbol{C})$

Flux

Template

Covariance matrix

$\Downarrow$

Optimal detection test:

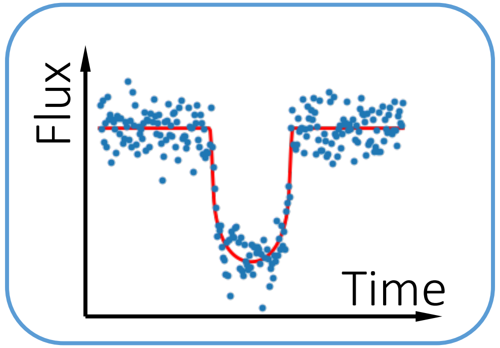
$$\log \frac{\mathcal{L}(\vec{d}|\mathcal{H}_{\text{planet}})}{\mathcal{L}(\vec{d}|\mathcal{H}_{\text{noise}})}$$

$\Longleftarrow$

Single-event statistic (SES):

$$\rho_{\text{SES}} = \vec{h}^T \boldsymbol{C}^{-1} \vec{d}$$

Calculating Single-event statistic



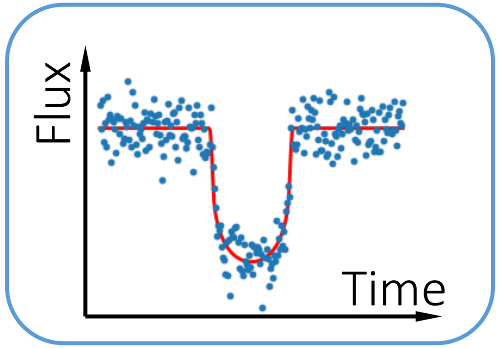
$$\rho_{\mathrm{SES}} = \vec{h}^{T} \boldsymbol{C}^{-1} \vec{d}$$

Template (**use template bank**)

Inverse covariance matrix

Data

Calculating Single-event statistic



Time domain

Fourier domain

$$\rho_{\mathrm{SES}} = \vec{h}^T \boldsymbol{C}^{-1} \vec{d} = \sum_f \frac{\hat{h}^\dagger(f)\hat{d}(f)}{\mathrm{PSD}(f)}$$

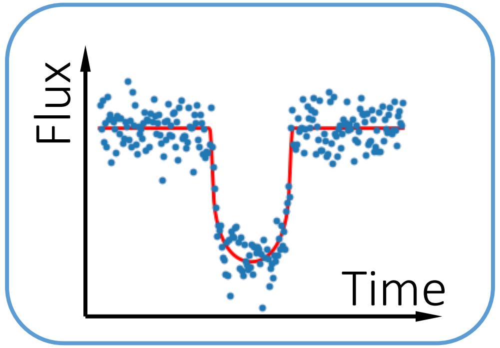
Template (**use template bank**)

Inverse covariance matrix

Data

Power spectral density (**Measure from the data**)

Calculating Single-event statistic



Time domain

Fourier domain

$$\rho_{SES} = \vec{h}^T \boldsymbol{C}^{-1} \vec{d} = \sum_f \frac{\hat{h}^\dagger(f)\hat{d}(f)}{\mathrm{PSD}(f)}$$

Template (**use template bank**)

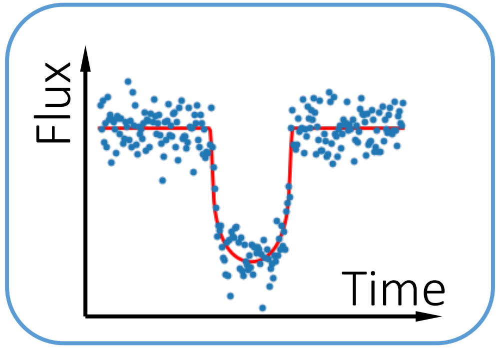
Inverse covariance matrix

Data

Power spectral density (**Measure from the data**)

Efficiently treats stellar variability noise

Calculating Single-event statistic

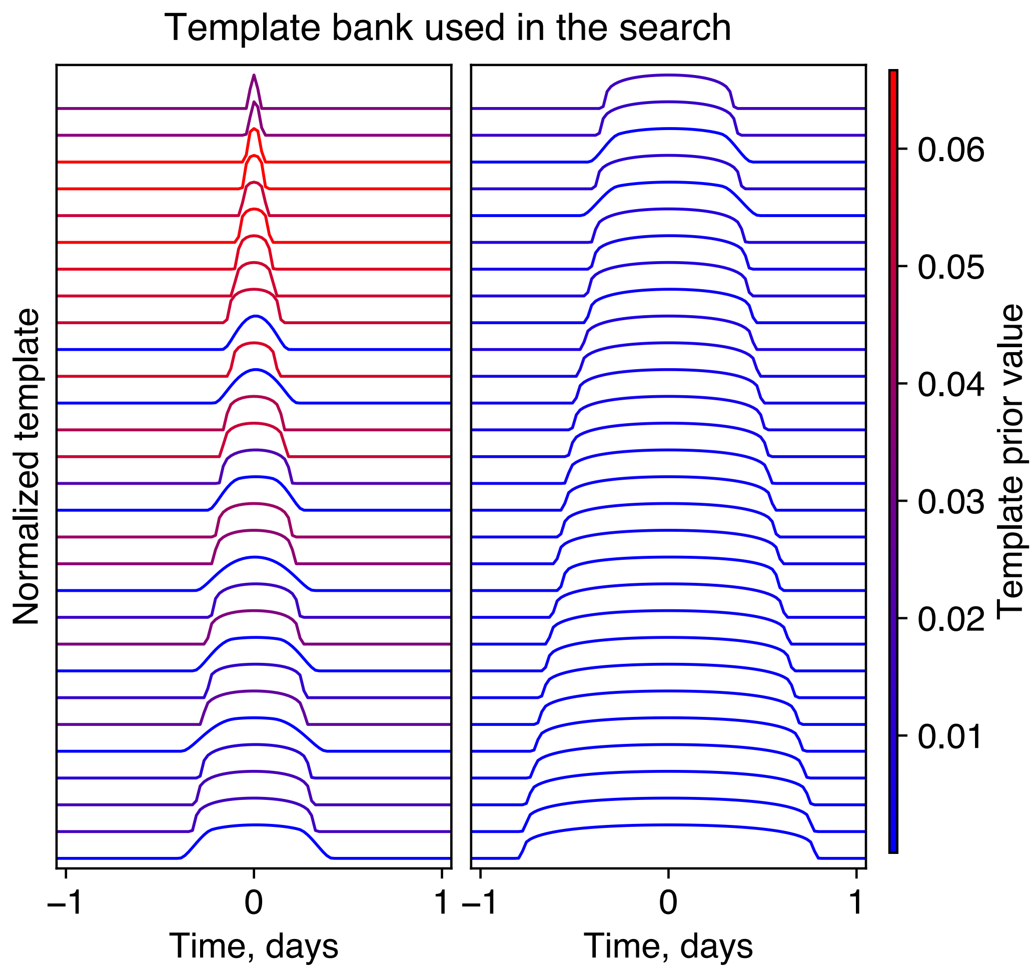


$$\rho_{SES} = \vec{h}^T \boldsymbol{C}^{-1} \vec{d}$$

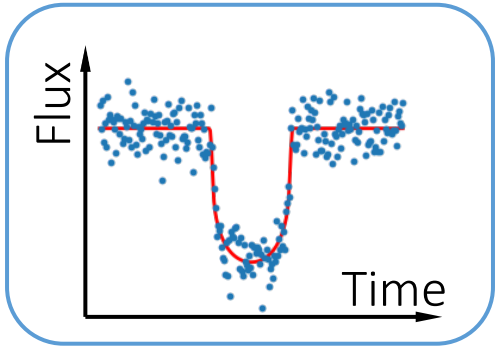
Template (**use template bank**)

Inverse covariance matrix

Data



Calculating Single-event statistic



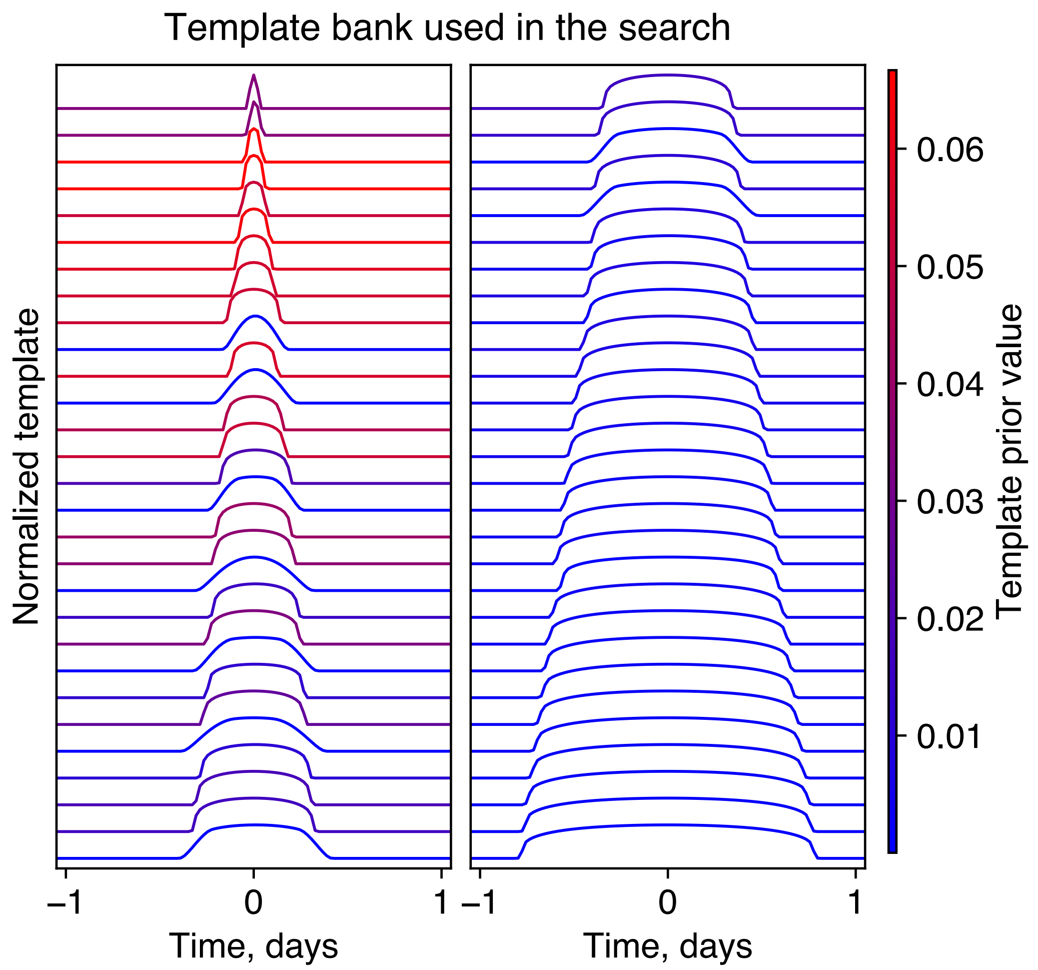
Provides more SNR for faint signals

$$\rho_{\text{SES}} = \vec{h}^T \boldsymbol{C}^{-1} \vec{d}$$

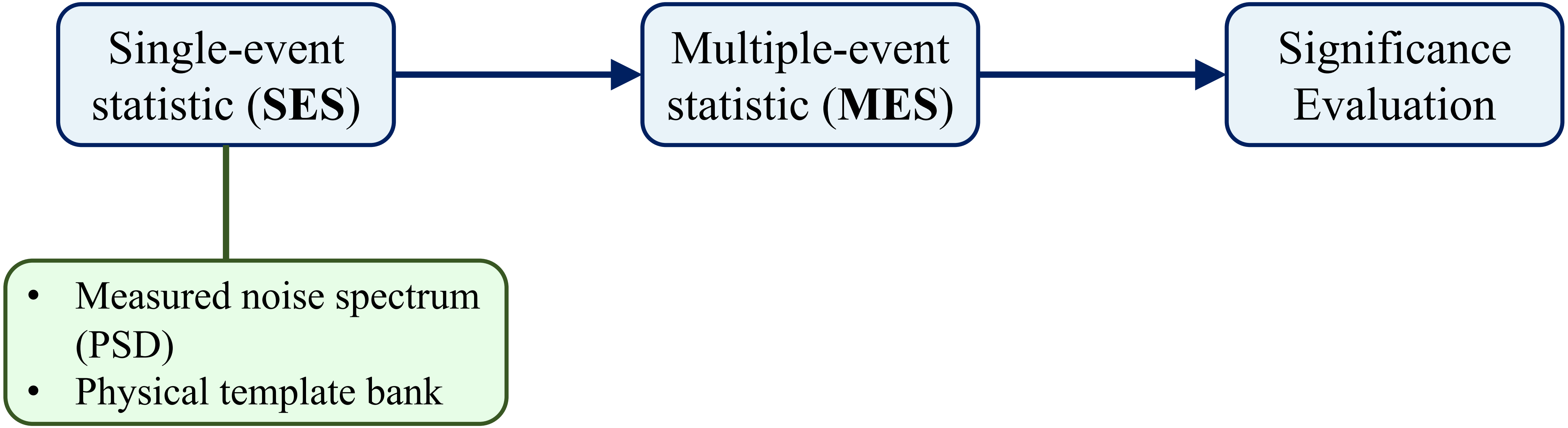
Template (**use template bank**)

Inverse covariance matrix

Data



Search scheme



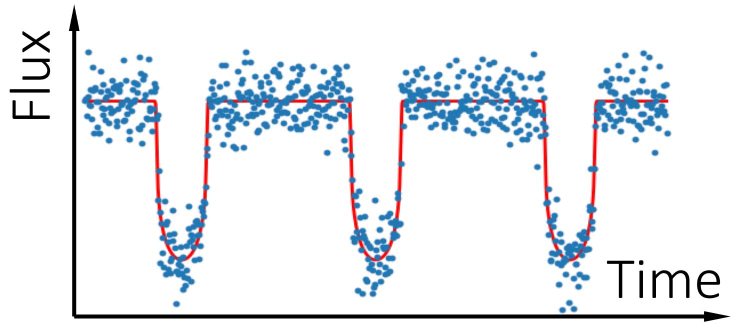
Search scheme

Single-event statistic (**SES**) → Multiple-event statistic (**MES**) → Significance Evaluation

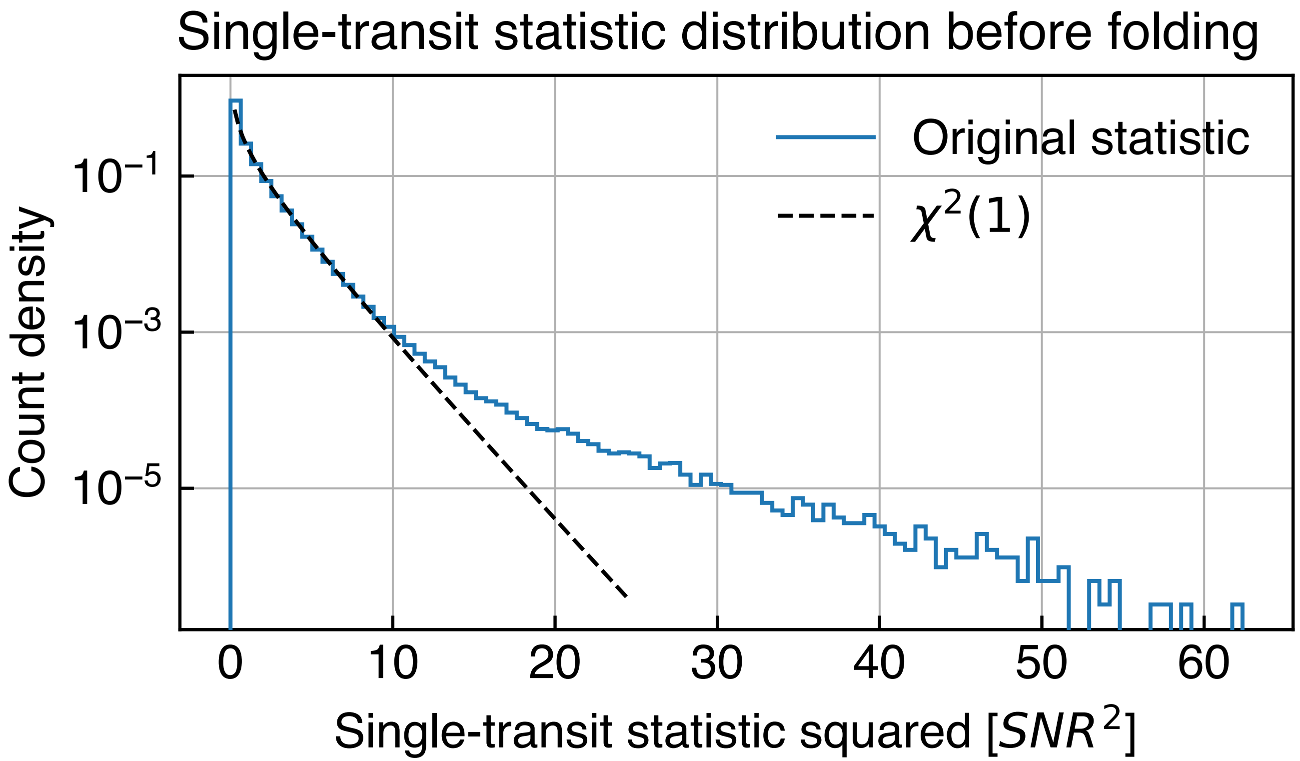
- Measured noise spectrum (PSD)
- Physical template bank

Periodicity search:

$$\rho_{\mathrm{MES}} = \frac{\sum_i \rho_{\mathrm{SES},i}}{\sqrt{\sum_i \mathrm{Var}(\rho_{\mathrm{SES},i})}}$$

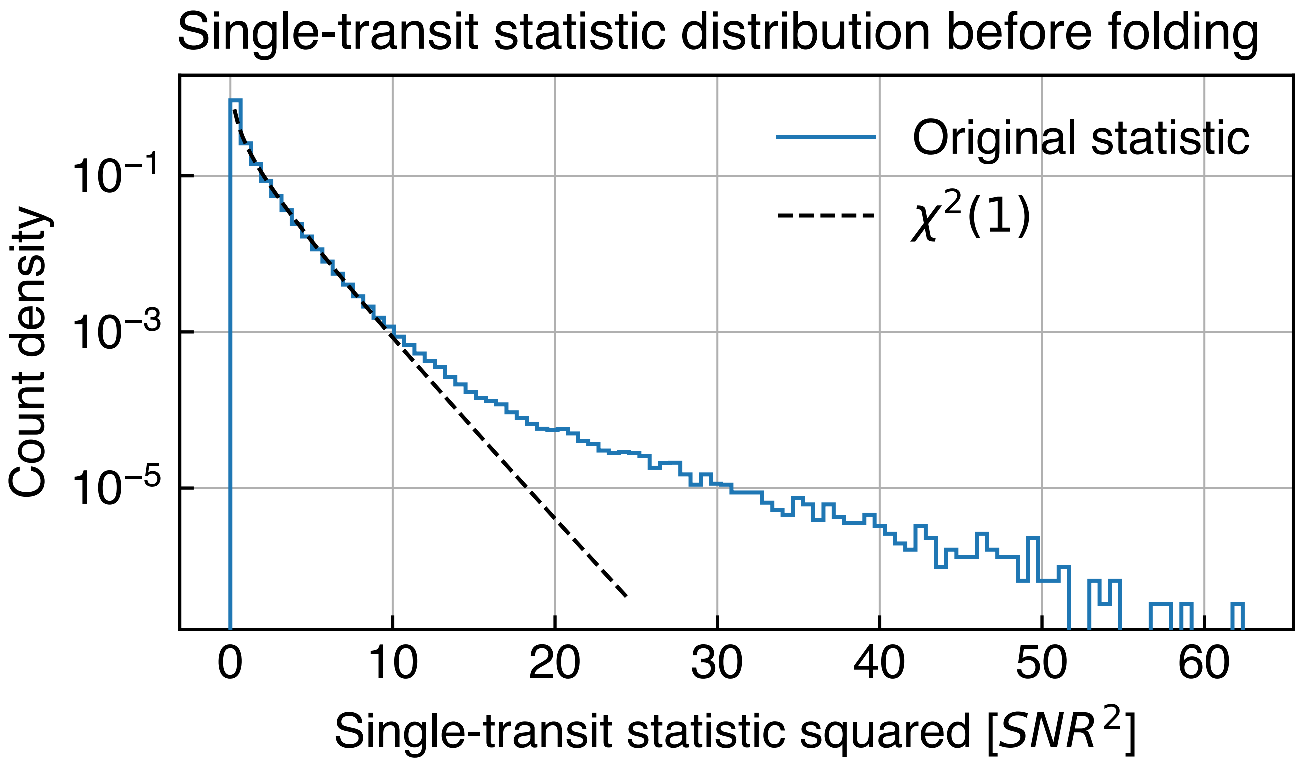


Non-Gaussianity of single-event statistic (SES) distribution



(ρ_{SES}^2)

Non-Gaussianity of single-event statistic (SES) distribution

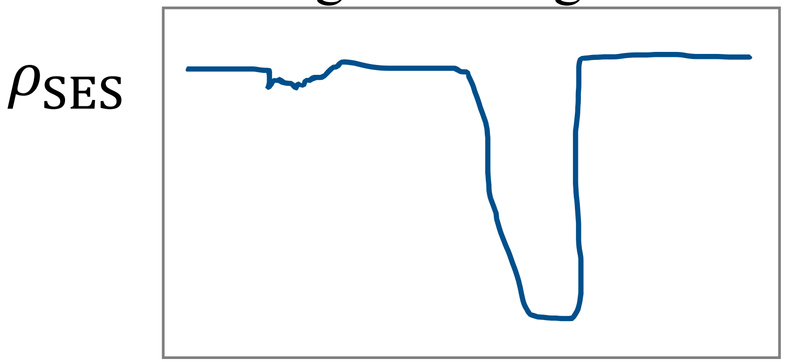


(ρ^2_{SES})

How it contaminates periodic score:

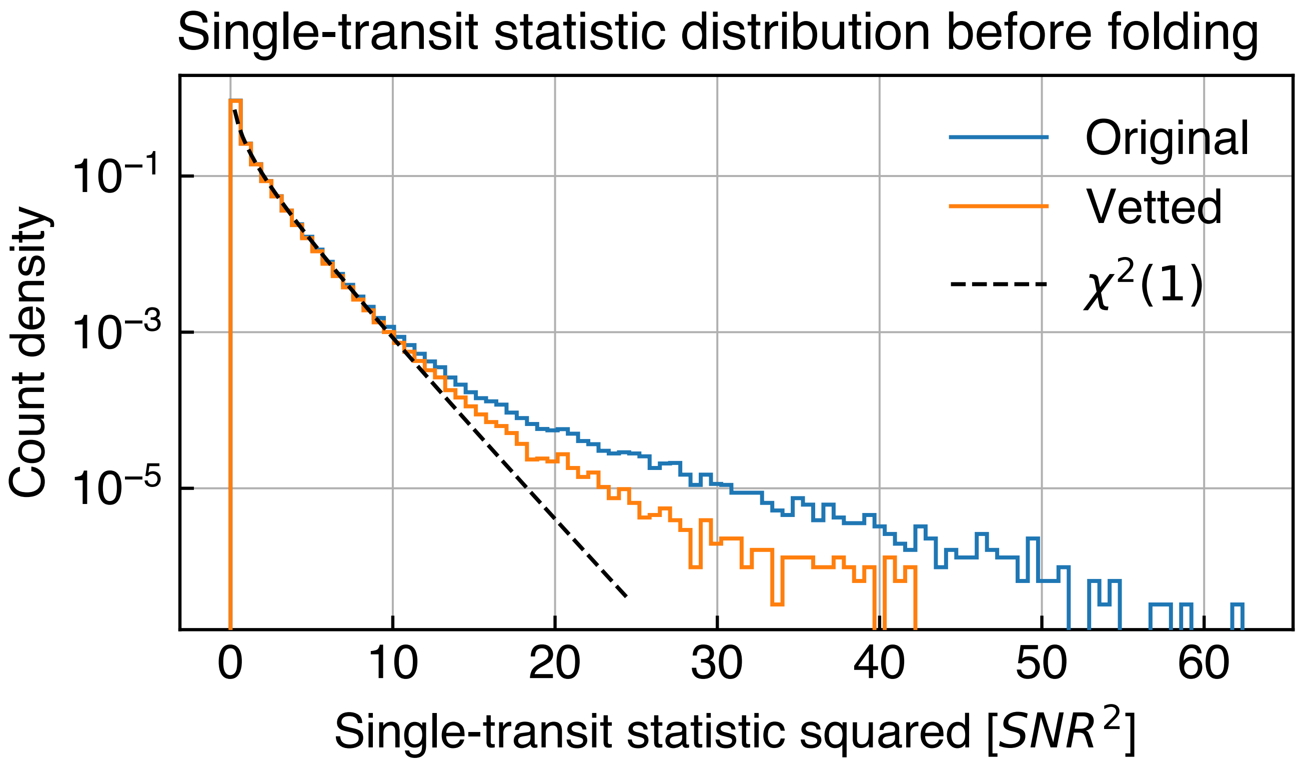
Good planet – big score

Bad glitch – big score

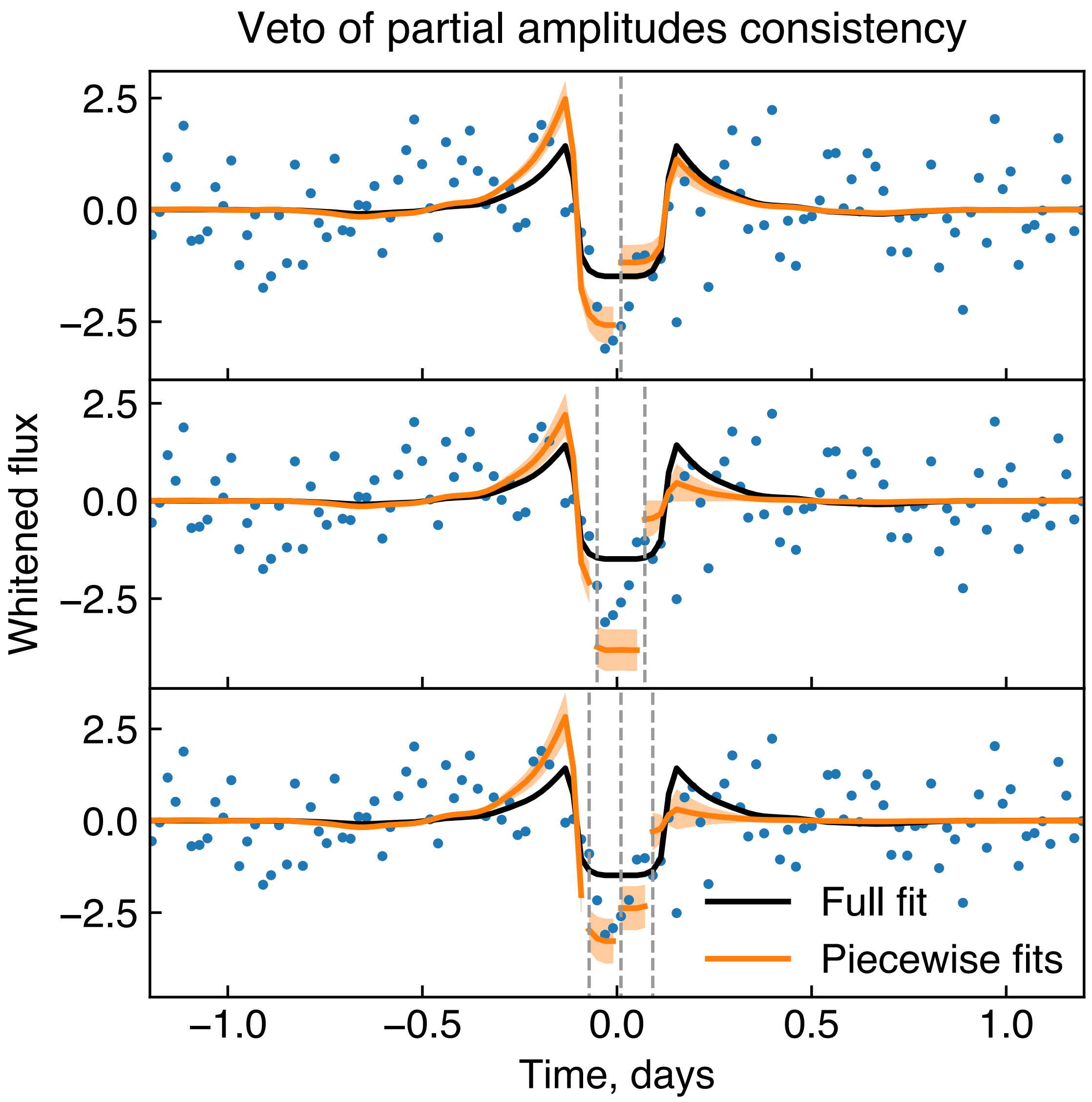


$$\rho_{\mathrm{MES}} = \frac{\sum_i \rho_{\mathrm{SES},i}}{\sqrt{\sum_i \mathrm{Var}(\rho_{\mathrm{SES},i})}}$$

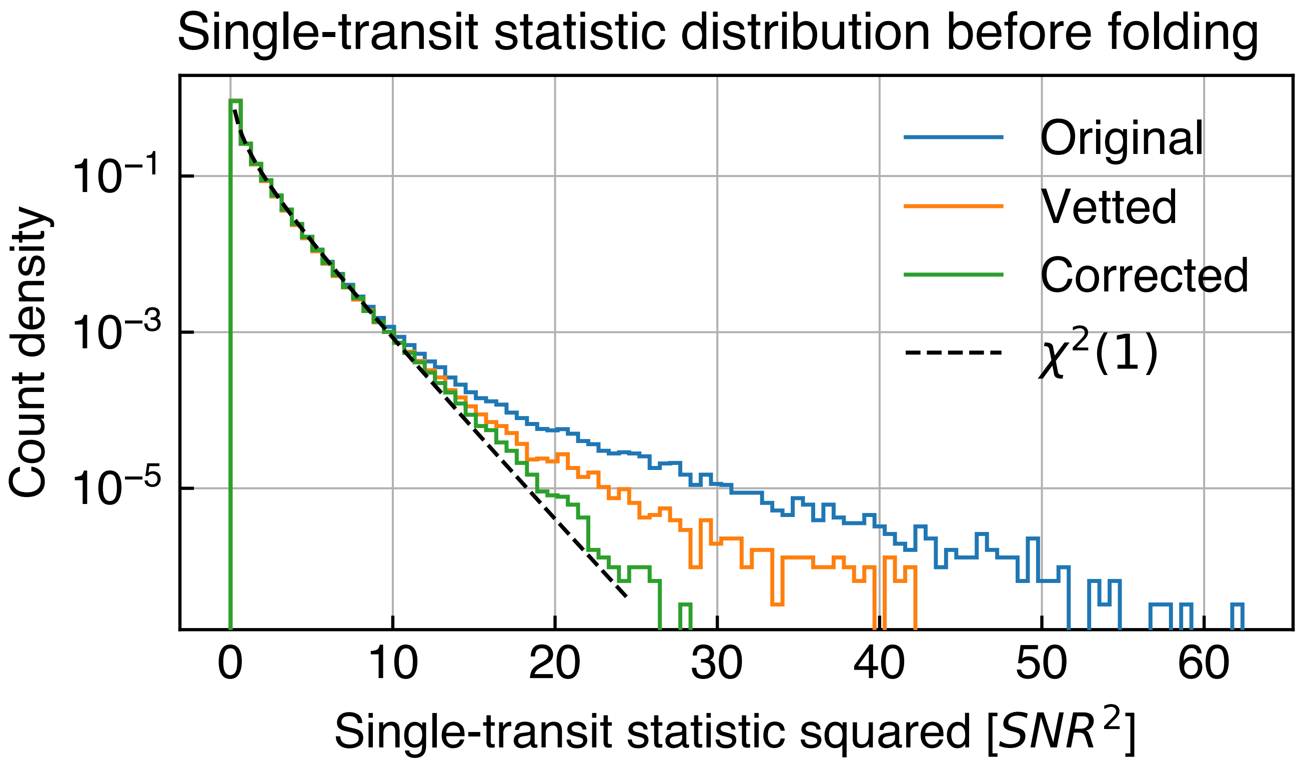
Transit shape veto



(ρ^2_{SES})



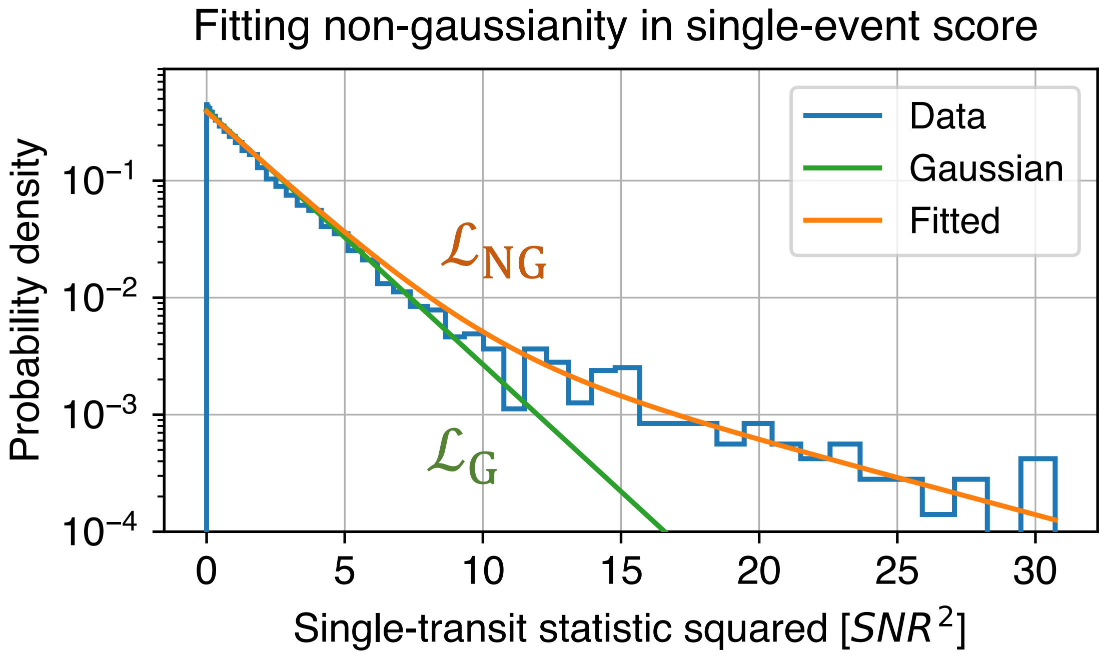
Non-Gaussianity correction



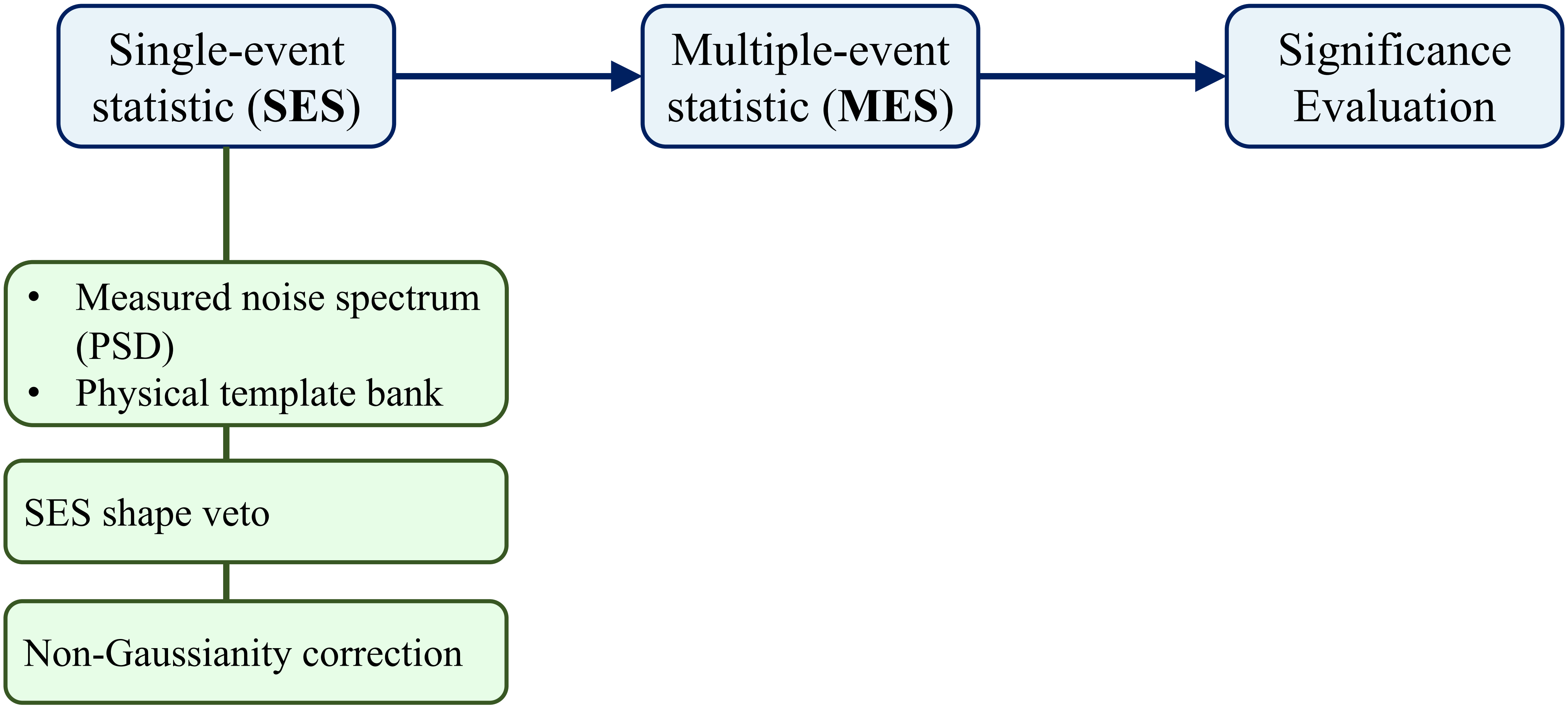
(ρ_{SES}^2)

Adding the non-Gaussianity correction to the score:

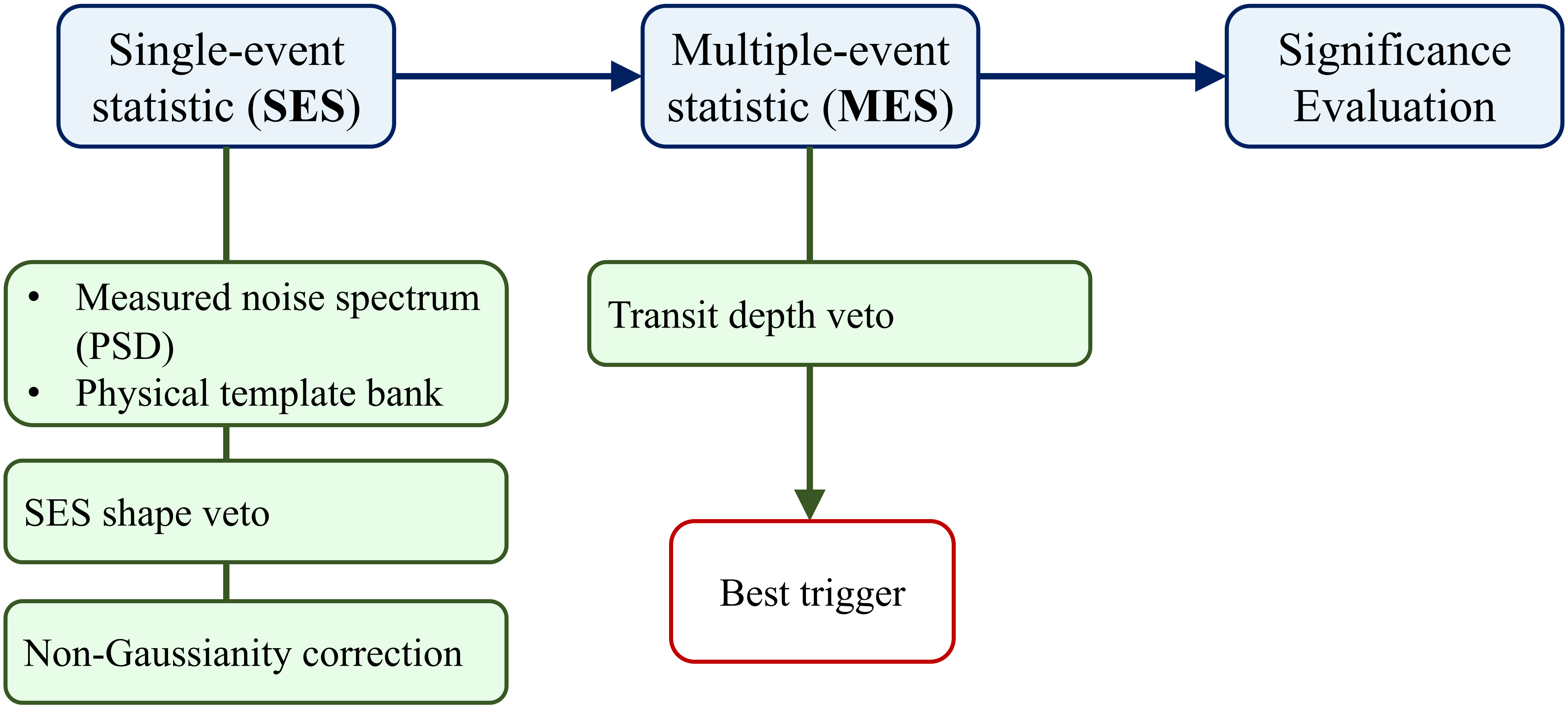
$$\xi = 2\log\frac{\mathcal{L}_{\mathrm{NG}}(\rho_{\mathrm{SES}}|\mathcal{H}_{\mathrm{noise}})}{\mathcal{L}_{\mathrm{G}}(\rho_{\mathrm{SES}}|\mathcal{H}_{\mathrm{noise}})}$$



Search scheme



Search scheme



Integral statistic score

Orbital period

First transit time

Template parameters

$$\mathcal{L} = \int dp \int dt_0 \int d\boldsymbol{\theta}\, \pi\left(\boldsymbol{\theta}, p, t_0\right) \mathcal{L}\left(\mathbf{d} | \mathbf{h}\left(\boldsymbol{\theta}\right), p, t_0\right)$$

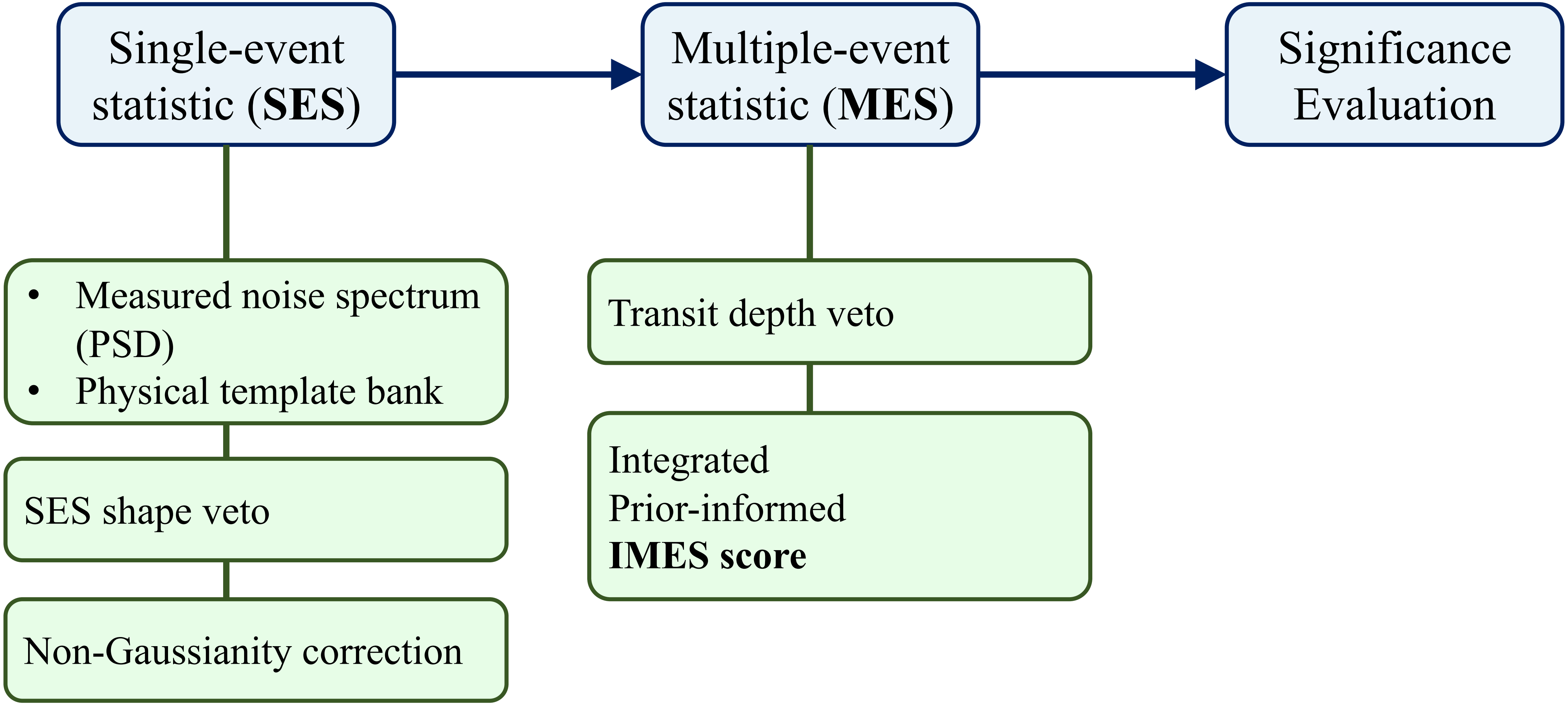
Full likelihood

Prior

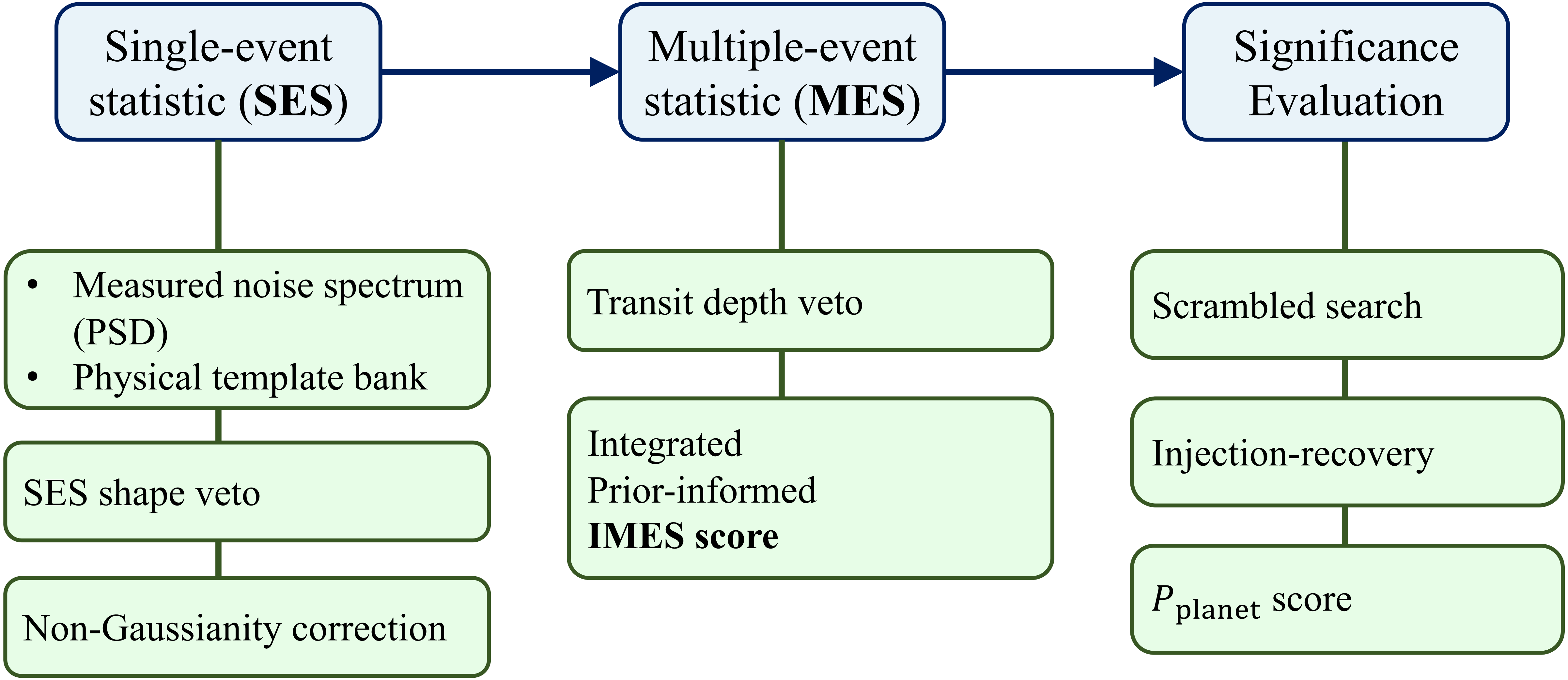
Point likelihood

- **IMES**: Integral multiple-event statistic: $\rho_{\mathrm{IMES}} \propto \log \mathcal{L}$

Search scheme



Search scheme



P_{planet} significance metric

What is the probability that this event is planet and not noise?

$$P_{\text{planet}} = \frac{\pi_{\text{p}} \mathcal{L}(\rho|\mathcal{H}_{\text{planet}})}{\pi_{\text{p}} \mathcal{L}(\rho|\mathcal{H}_{\text{planet}}) + (1 - \pi_{\text{p}}) \mathcal{L}(\rho|\mathcal{H}_{\text{noise}})}$$

Planetary distribution

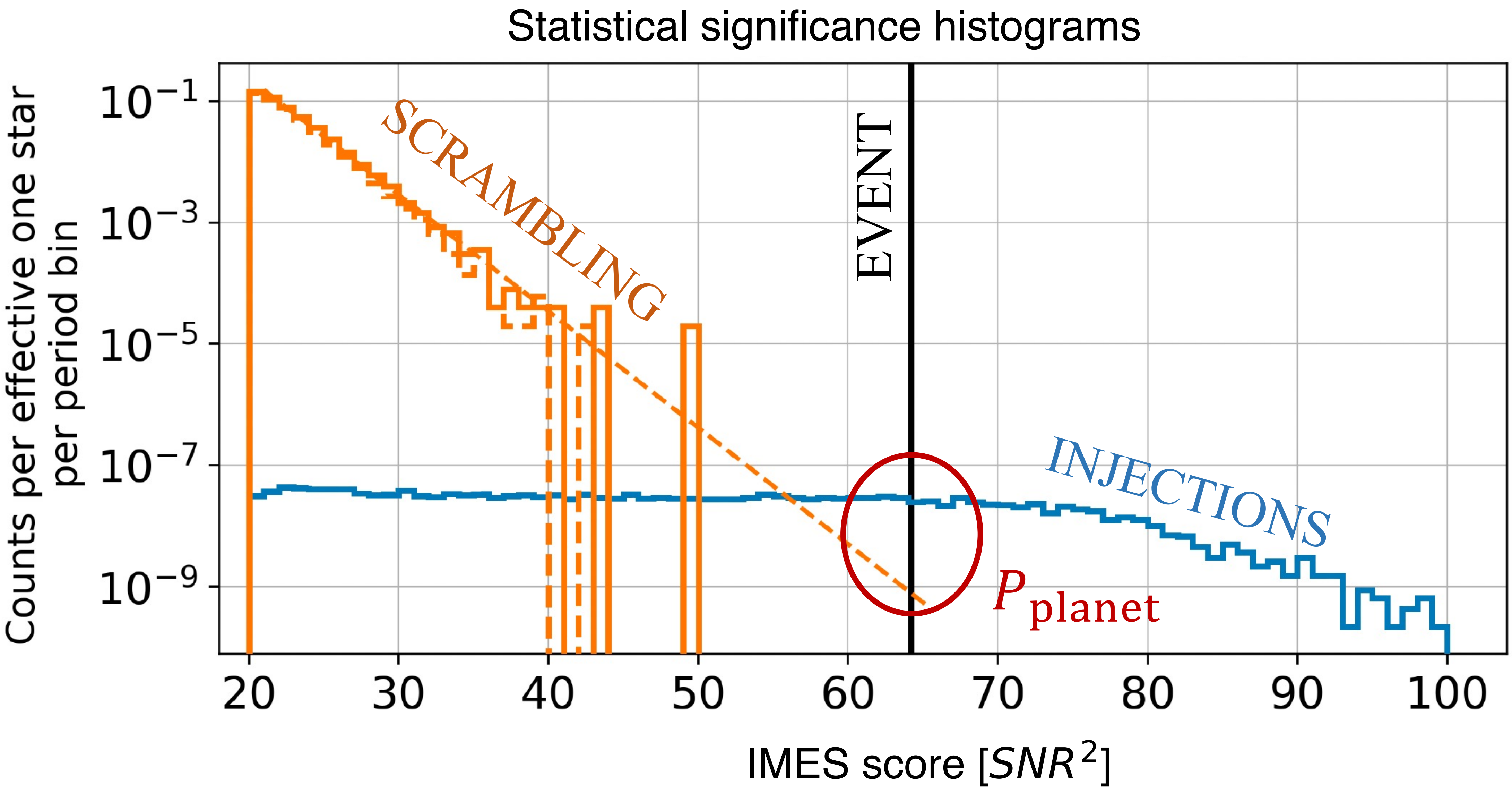
INJECTIONS

Prior

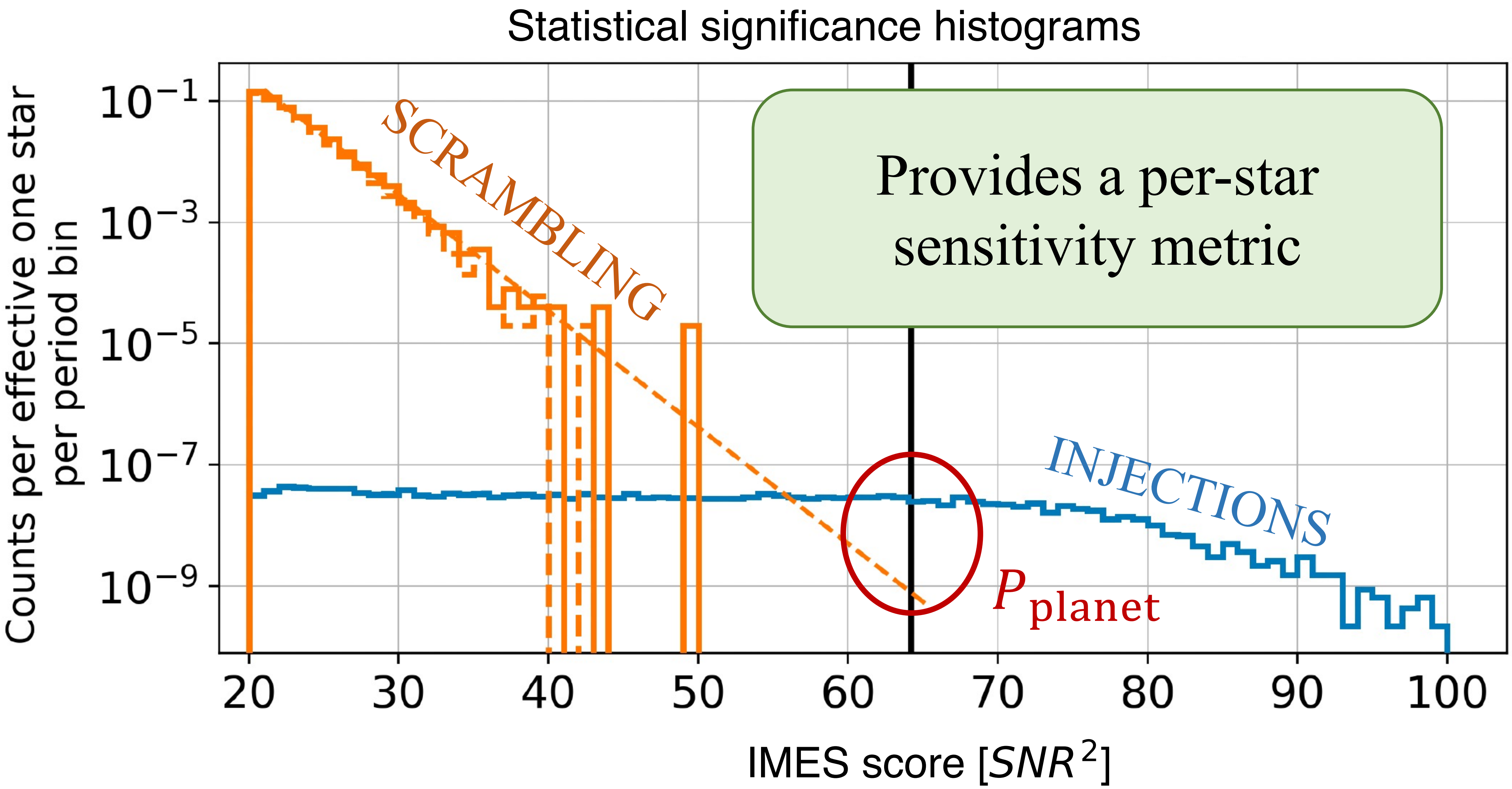
Background distribution

SCRAMBLING

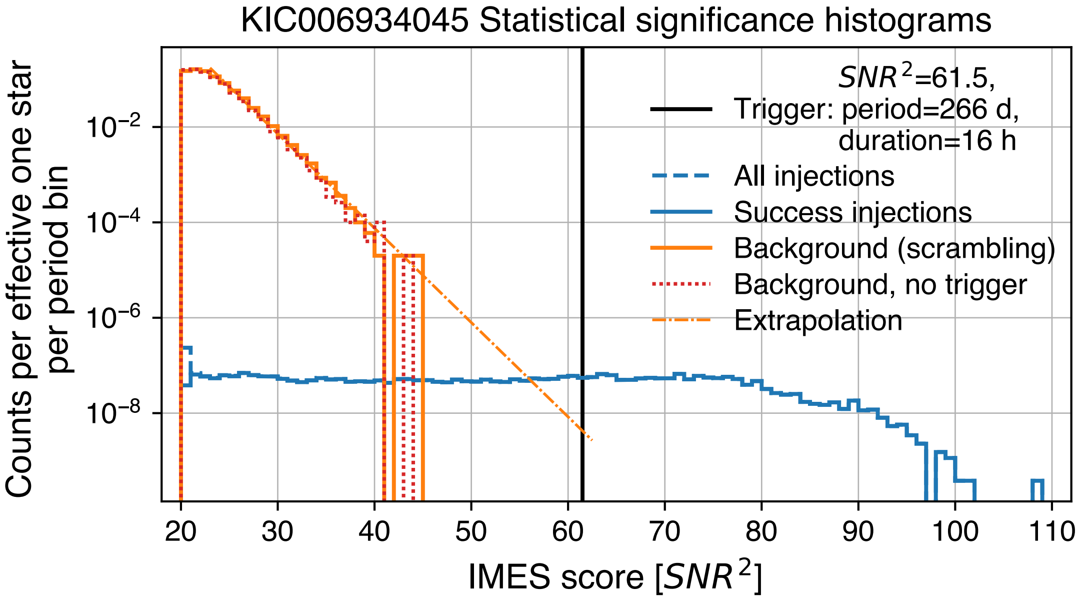
Significance estimation

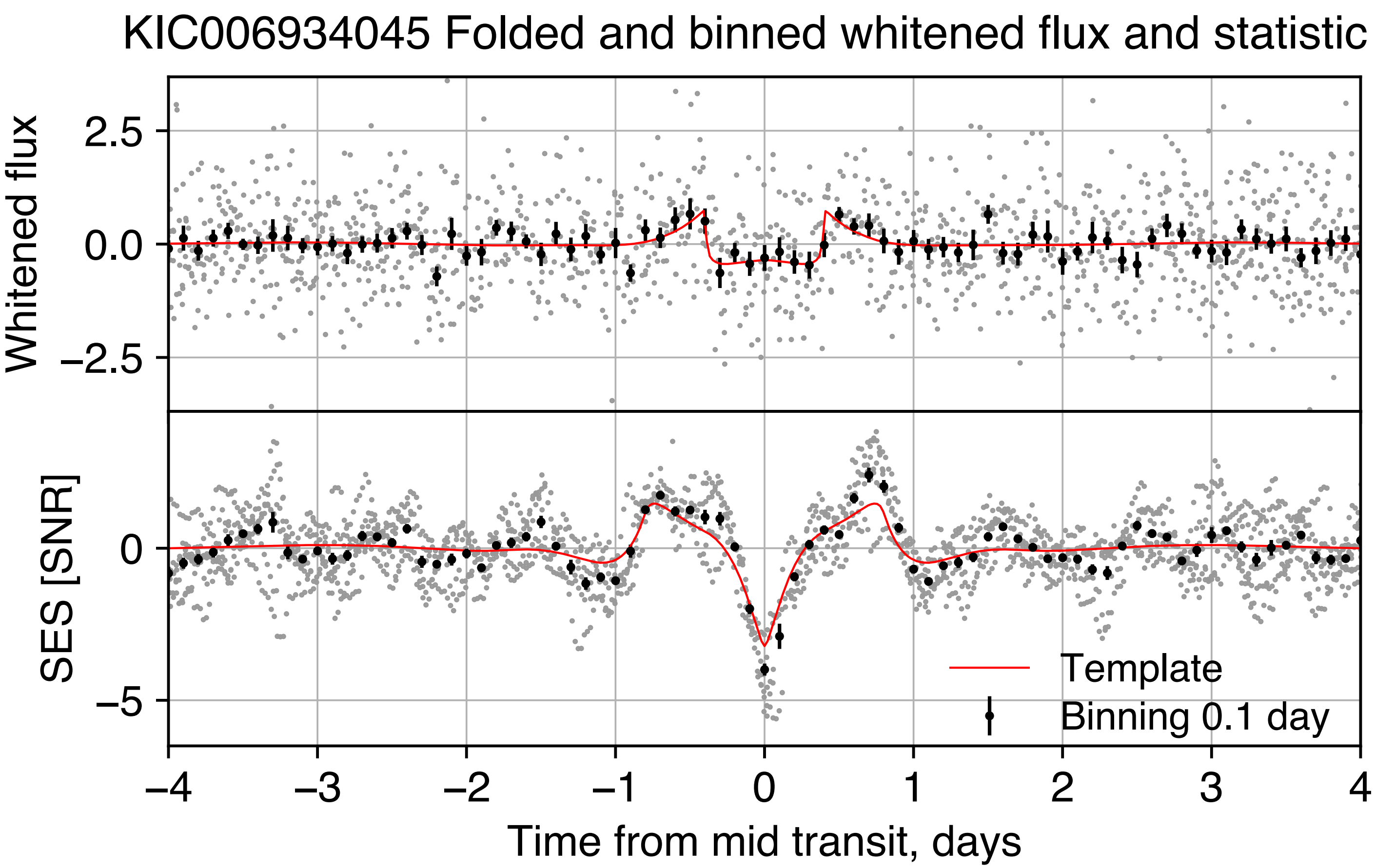


Significance estimation



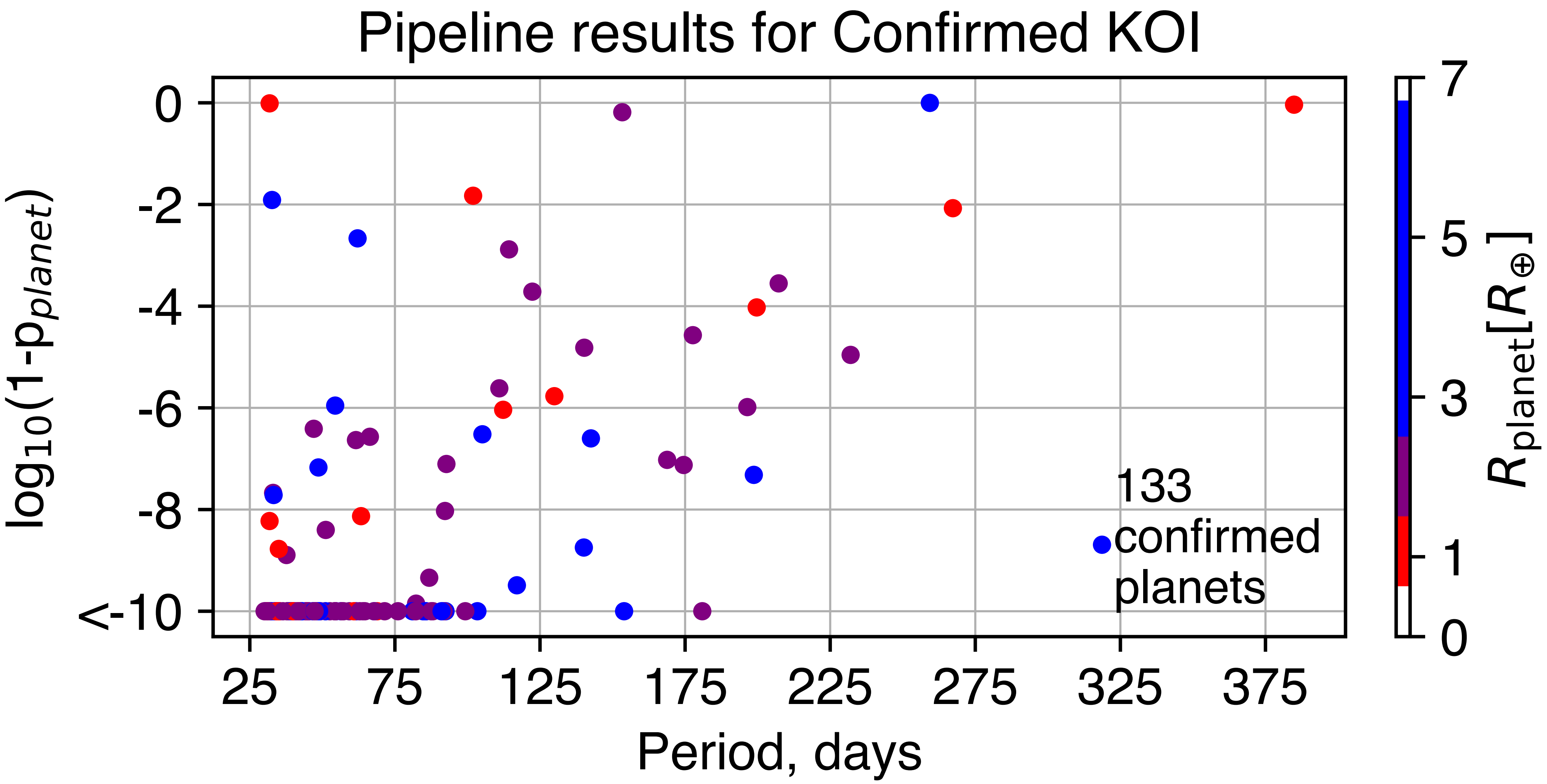
Example of pipeline output



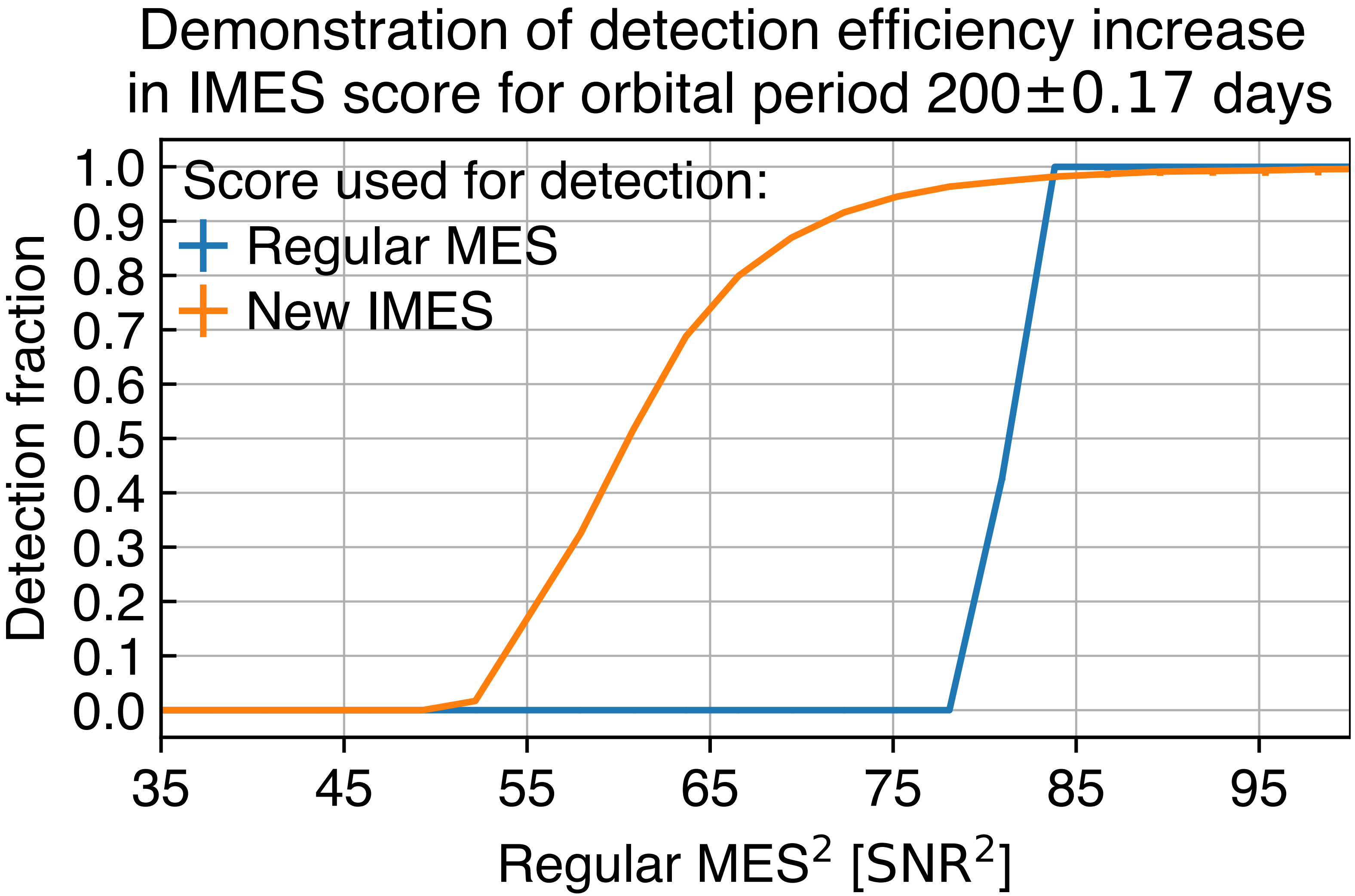


Pipeline performance evaluation

Test 1: recovering faint *Kepler* confirmed planets

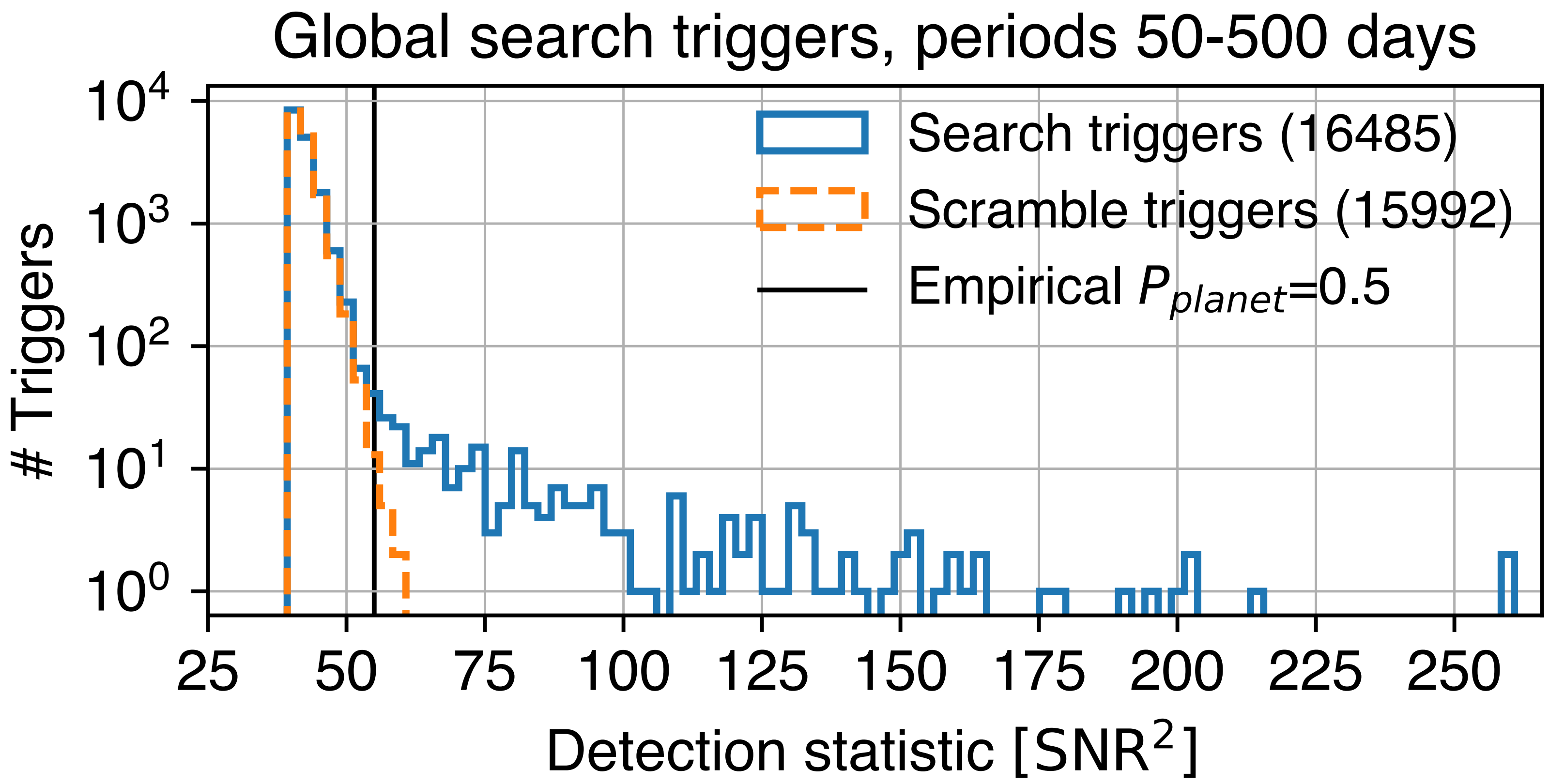


Detection efficiency increase of the new IMES score

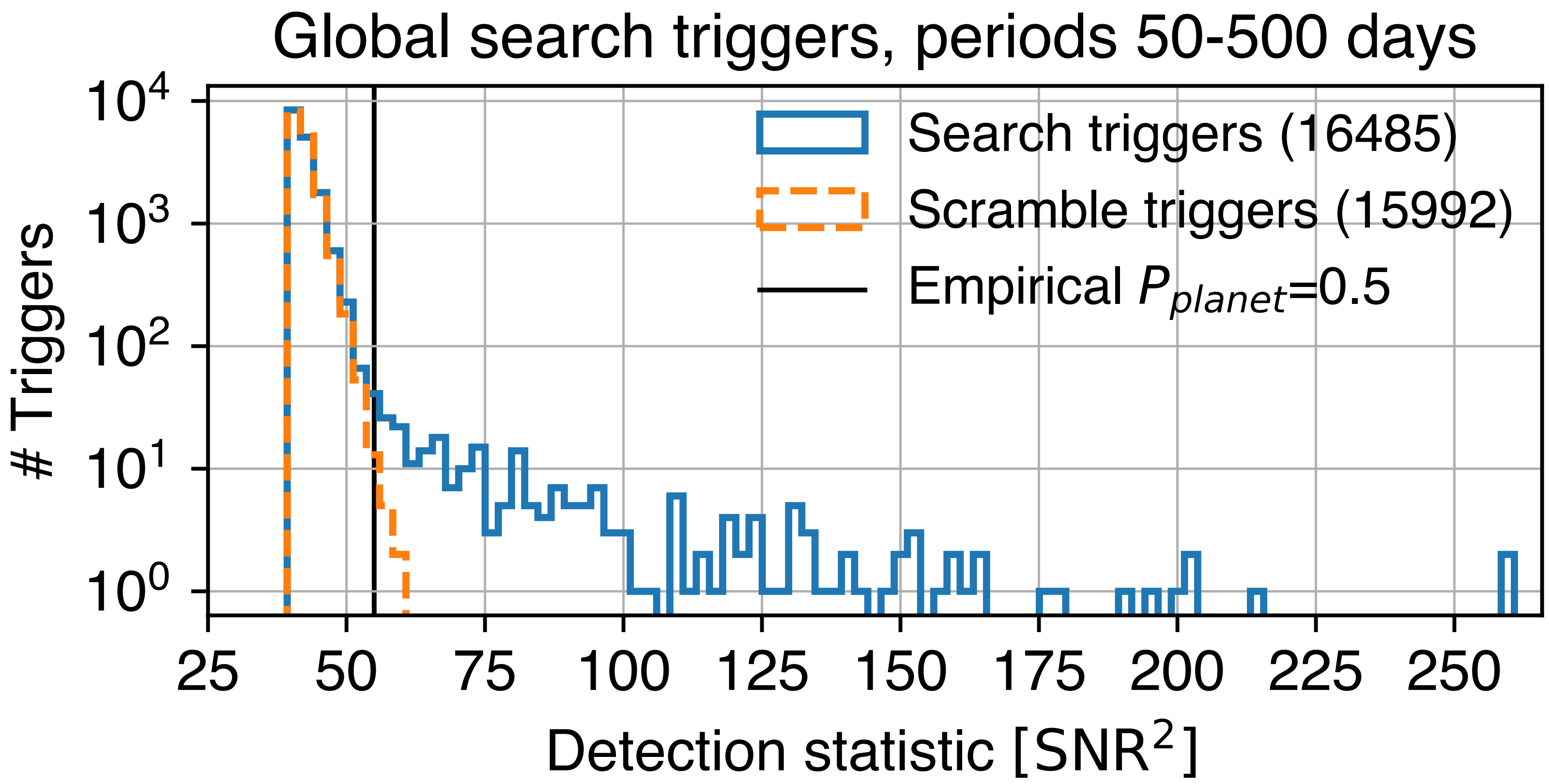


Running the pipeline on the *Kepler* data

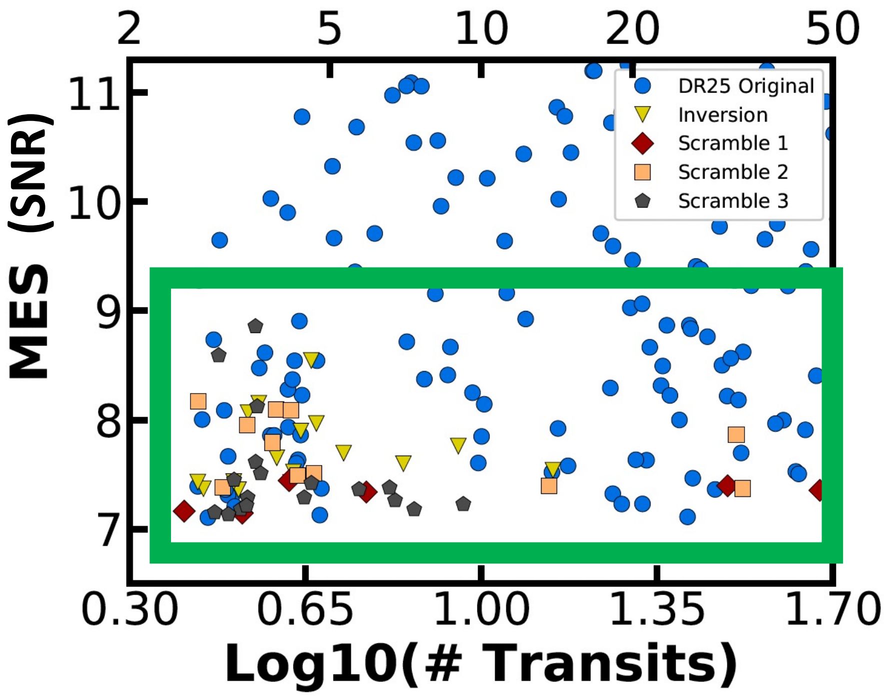
Global search trigger distribution



Global search trigger distribution



[Burke 2019]



Preliminary catalog from the search