CWT for Point Source Detection

Created	@March 15, 2023 10:50 AM
Status	Open
Updated	@May 15, 2024 12:03 PM

The Continuous Wavelet Transform

Wavelets in One Dimension

Definition: Suppose $f \in L^2(\mathbb{R})$. Then, its *continuous wavelet transform* (CWT) is given by

$$\begin{aligned} W_\psi(a,b) &= \frac{1}{a^{1/2}} \int dx\ f(x)\ \bar{\psi}\left(\frac{x-b}{a}\right) \\ &\equiv \langle f, \psi^{a,b} \rangle, \end{aligned}$$

where $a \in \mathbb{R}^{+*}$, $b \in \mathbb{R}$ denotes the *scale parameter* and the *translation parameter*. The function $\psi \in L^2(\mathbb{R})$ is referred to as a *mother wavelet*.

Properties of ψ:

- Suppose the mother wavelet satisfies the *admissibility condition*

$$C_\psi = 2\pi \int d\xi\ |\xi|^{-1}\ |\tilde{\psi}(\xi)|^2 < \infty,$$

then f can be reconstructed using the *first inverse CWT* (ICWT):

$$f(x) = C_\psi^{-1} \int_0^\infty \int_{-\infty}^\infty \frac{da\ db}{a^2}\ W_\psi(a,b)\ \psi\left(\frac{x-b}{a}\right).$$

Remark: The integral's convergence to the function f is defined in the *weak* sense in $L^2(\mathbb{R})$.

Spherically Symmetric Wavelets in $\mathbb{R}^n$:

$f, \psi \in \mathbb{R}^n$:

$$\begin{aligned} W_\psi(a,\vec{b}) &= \frac{1}{a^{n/2}} \int d^n x\ f(\vec{x})\ \bar{\psi}\left(\frac{\vec{x}-\vec{b}}{a}\right) \\ &\equiv \langle f, \psi^{a,\vec{b}} \rangle, \end{aligned}$$

$$C_\psi = (2\pi)^n \int d\xi\ |\xi|^{-1}\ |\tilde{\psi}(\xi)|^2 < \infty,$$

$$f(\vec{x}) = C_\psi^{-1} \int_0^\infty \int_{\mathbb{R}^n} \frac{da\ d^n b}{a^{n+1}}\ W_\psi(a,\vec{b})\ \psi\left(\frac{\vec{x}-\vec{b}}{a}\right).$$

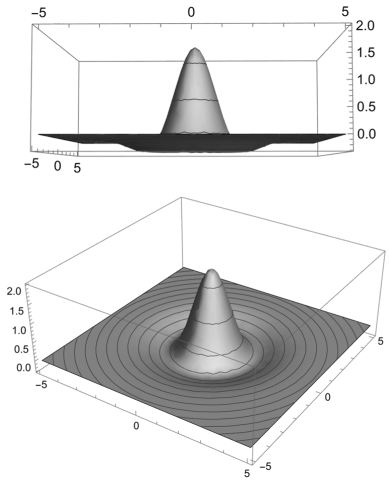
— References: Wikipedia, Daubechies ch. 2

The Two-Dimensional Mexican Hat Wavelet:

The 2D Mexican Hat Wavelet family is defined in terms of any number of Laplacians acting on a Gaussian. We work with the Mexican Hat Wavelet of the first-kind:

$$\psi(\vec{x}) = (2 - x^2)\, e^{-x^2/2},$$

Though it isn't relevant, note that this wavelet is unnormalized: $||\psi|| = 2\pi$. A figure of it is given below.



Two perspectives of the Mexican Hat Wavelet

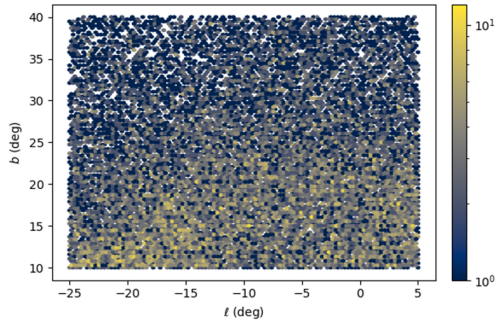
The obvious advantage of this wavelet is its shape. Its central peak makes it suitable for detecting Gaussian-like functions and its negative tails perform a background subtraction. At a deeper level, it is sensitive to the max/min of the convolved functions and filters out constant/linear functions. This follows from the wavelet being the Laplacian of a Gaussian since, through integration by parts, the CWT of a function can be written as the second derivative of the function convolved with a Gaussian filter, due to the wavelet's expression as a Laplacian of a Gaussian.

— References: 0104077, 10.1086/304217

Mexican Hat Wavelets for Source Detection on Small Patch of the Sky

Discriminating Point Sources with the CWT

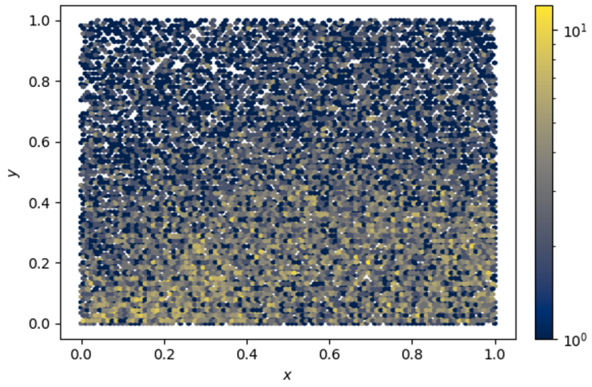
We now test the performance of the CWT as a point source finder. As a simple example, we consider how well the CWT detects point sources that are injected into a section of a synthetic gamma-ray map of the sky, slightly above the Galactic disk. This synthetic gamma-ray map is taken as the background map.



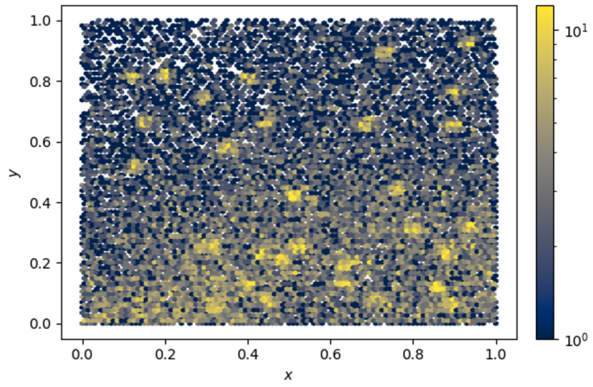
Histogram of counts for the synthetic gamma-ray map

For simplicity, we express the map in terms of an effective longitude and latitude (x, y) such that the data only has support on the unit square.

For illustrative purposes, let's inject 30 point sources (i.e., PSF-scale gaussians) and see how well the CWT performs by eye. These point sources are made up of 100 photons each.



no point sources - count map



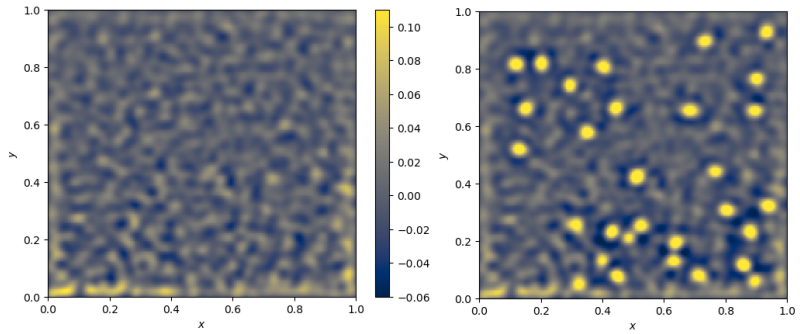
with point sources - count map

Suppose the density corresponding to this map of counts is given by $f(\vec{x})$. We can generate a map of continuous wavelet transforms over a grid of translation parameters $\vec{b}$. For a fixed scale a, the CWT of the density map is given by

$$W_\psi(a, \vec{b}) = \frac{1}{a} \int d^2x \; f(\vec{x}) \; \bar{\psi} \left(\frac{\vec{x} - \vec{b}}{a} \right)$$
$$\sim \frac{1}{a} \times \frac{1}{N} \sum_i \bar{\psi} \left(\frac{\vec{x}_i - \vec{b}}{a} \right),$$

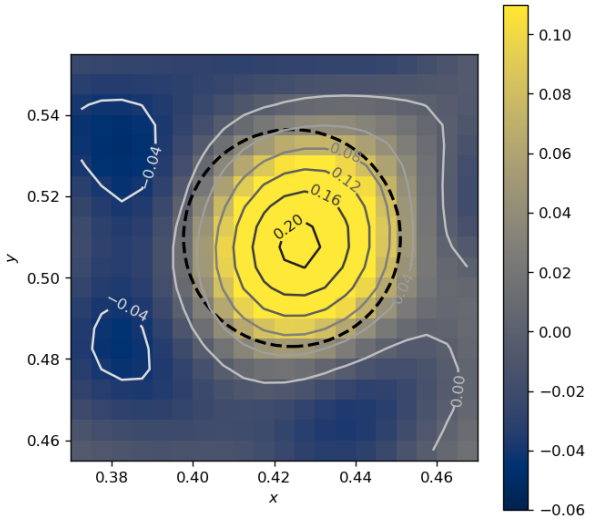
where our sample of photon counts is given by $\{\vec{x}_i\}_{i=1}^N$.

We found the best performance for scales slightly above the psf scale $\sigma_{psf} \sim 0.4^\circ$; specifically, around $a \sim 0.6^\circ$, so let's consider CWTs of this scale. The CWT of the two maps are given by

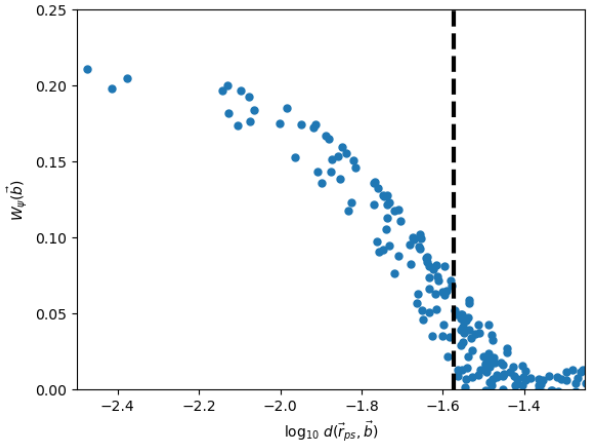


(Left) no point sources - CWT map; (Right) with point sources - CWT map

Therefore, the point sources are clearly visible through their large positive CWTs. To verify this more carefully, let's consider the distribution of CWTs from the background (map with no points), signal + background (map with point sources), and the signal (point sources). To define our signal in this case, consider one of the point sources in this map and zoom in on it. We should find that the CWT falls off as our distance from the center of the point source increases. One such case gives the following two plots:

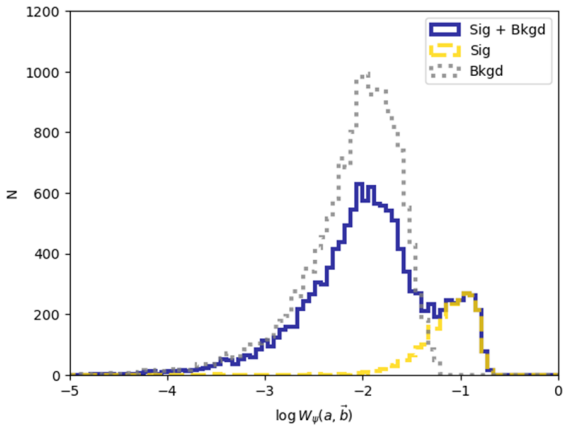


CWT of nearby grid points from the center of point source with contours. The dashed circle encloses the grid points that are within 2\sigma_psf of the point source.



CWT of nearby grid points from the center of point source vs distance from point source; the dashed vertical line corresponds to 2\sigma of the point source

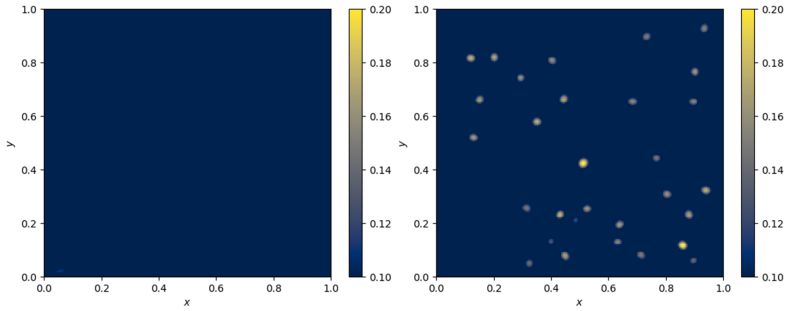
From these plots, it is reasonable to take as our point source as contributing to the CWTs that are less than $2\sigma_{psf}$ away from the center of the point source. With this definition, we can produce the following histogram for the distribution of positively-valued CWTs (after masking contributions from the boundary, where the CWT is expected to fail):



This indicates that our point sources are discriminated by their large S-values.

A Simple Point Source Finder using the CWT

From the above histogram, a grid point $\vec{b}$ corresponds to a point source is $W_\psi(a, \vec{b}) \gtrsim 0.1$. This can be seen clearly by filtering our previous CWT maps to display points where $W_\psi(a, \vec{b}) > 0.1$:

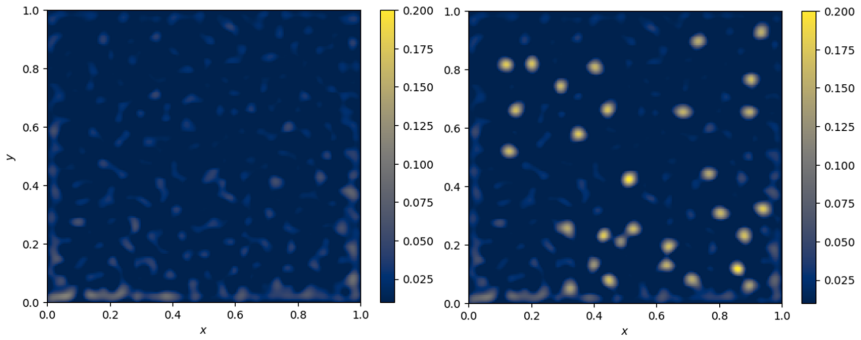


(Left) no point sources - CWT map with values greater than 0.1; (Right) with point sources - CWT map with values greater than 0.1

In this approach, we emphasize that a point source corresponds to a collection of grid points. This motivates the following simple point source finding algorithm:

1. Generate the CWT map $S(\vec{b}) \equiv W_\psi(a, \vec{b})$ for $a = 0.6^\circ$.
2. Filter the map to consider only grid point whose CWTs are greater than some threshold S_0. This defines a set of point source candidate grid points: $\Omega_0 \equiv \{\vec{b} : S(\vec{b}) \geq S_0\}$
3. A point source is identified with its *wavelet peak* in the set Ω_0.

Clearly, this approach correctly identifies all bright point sources. However, for faint point sources, where S_0 must be set too small, leading to many false detections. For example, for S_0 = 0.01, many regions corresponding to background fluctuations would pass the cut and appear as a wavelet peak.

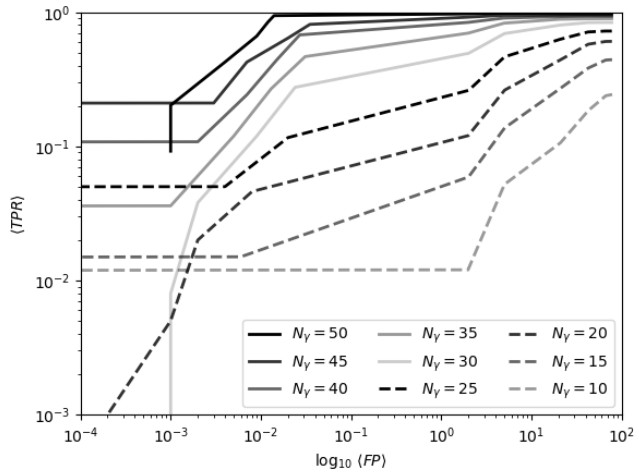


(Left) no point sources - CWT map with values greater than 0.01; (Right) with point sources - CWT map with values greater than 0.01

We measure the algorithm's performance in measuring *isolated point sources* by constructing averaged ROC curves. For the same background map, let's inject a single point source at a random location (sufficiently far from the boundary) 1000 times. For each of these 1000 trials, we use the algorithm to calculate any true/false detections. For each of the trials, we can compute true/false detection rates, defined as

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \qquad\qquad \mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

We can construct ROC curves for each trial or plot a single ROC corresponding to trial-averaged TPRs and FPRs. In this case, our "candidate sources" are the grid points. This leads to the FPR typically being very small since the number of grid points is large relative to the number of positive detections. Therefore, we instead plot the trial-averaged TPR versus the trial-averaged FP.



To reduce FP, we may apply selection criteria to distinguish background fluctuations from true point sources. In general, this will reduce the TPR for faint sources since they are harder to distinguish from background fluctuations. Because of this, such selection criteria are not very useful, but we list two possibilities anyways.

- **Point Source Morphology:** A simple example would be to consider the "Image moment" for the shape of the point source and the size of the point source. This would perform well for sources with a high number of photon counts since the shape it will form is approximately circular. However, faint point sources will more likely appear as smaller, non-circular source that appears more like a background fluctuation.

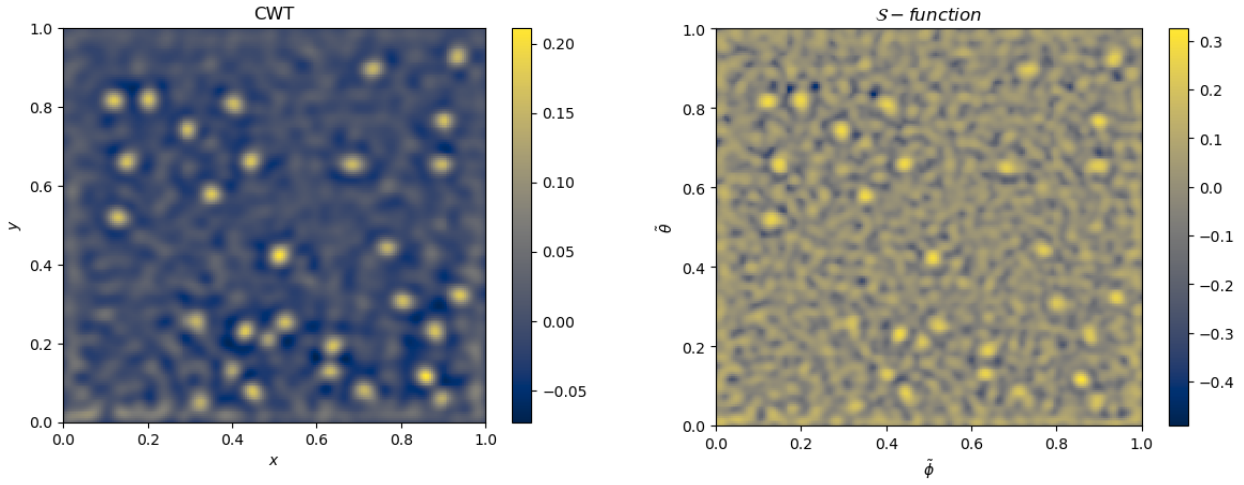
- **Point Source Fit:** A gaussian source has an analytic wavelet coefficient. By considering wavelet coefficients of various scales at the center of the point source, one may perform a fit of the function to the analytical expectation. The fit must be reasonably successful for a wavelet peak to correspond to a point source. [9912471]

Comparison with the Bartels' $\mathcal{S}$-function

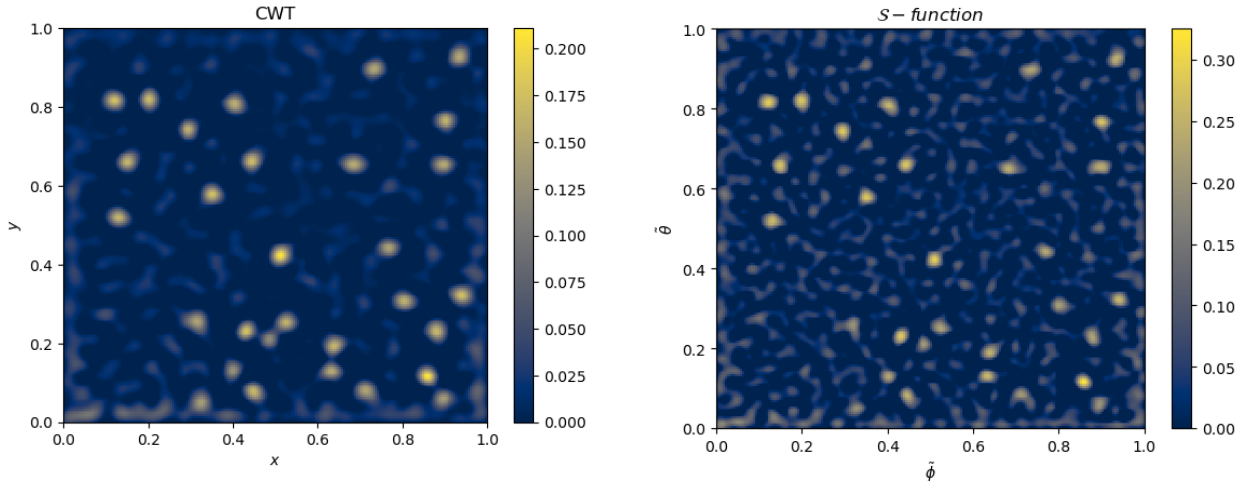
Instead of using the CWT as a figure of merit (fom) for point source detection, Bartels uses the following ratio of convolutions:

$$\mathcal{S}(a,\vec{b}) = \frac{W_\psi(a,\vec{b})}{\sqrt{W_{\psi^2}(a,\vec{b})}},$$

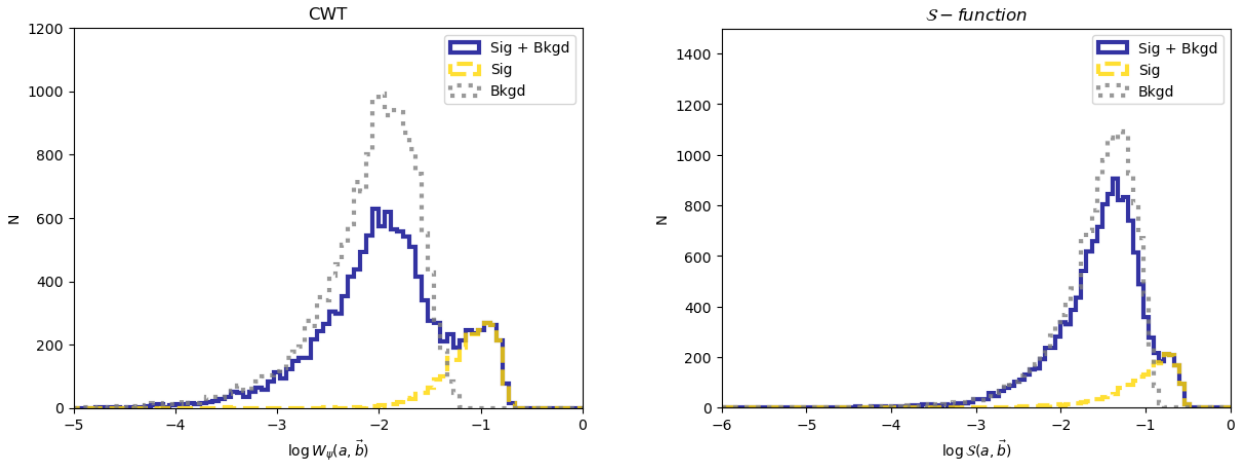
where they take $a = 0.4^\circ \sim \sigma_{psf}$. We compare the performance of $\mathcal{S}$ to the CWT W_ψ in detecting point sources by comparing their averaged ROC curves (as constructed in the previous section). Their maps corresponding to 100-photon point sources considered in the previous section are given below. The colormaps are defined from the minimum value of the fom to the maximum value.



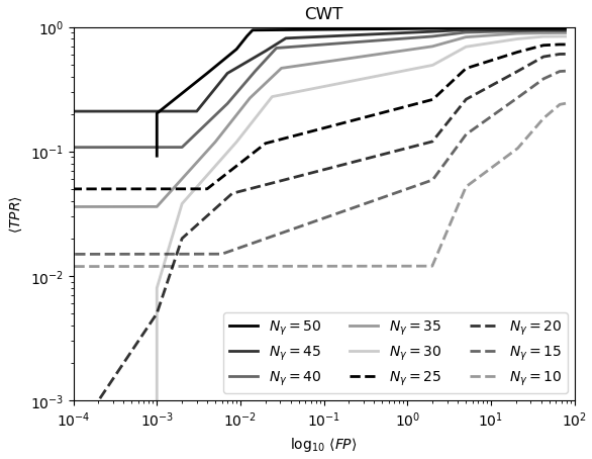
Since we do not consider negative values in our foms to detect point sources, let's define our colormap for only positive values.

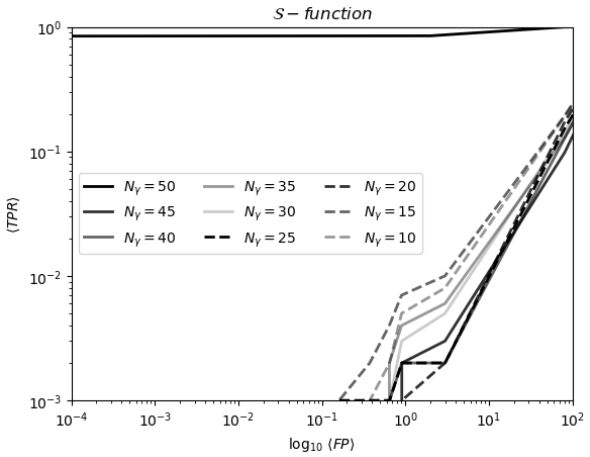


From this plot, we see that the background fluctuations for the $\mathcal{S}$-function map are higher than those of the CWT map. Therefore, the $\mathcal{S}$-function may struggle to detect faint sources that the CWT would consider to still be bright enough for detection. The corresponding histograms (positive-valued with boundaries excluded) are given by:



These histograms simply show that a cut on the $\mathcal{S}$-function can be used as a discriminator of point sources. However, to compare their performance, we perform the same analysis of the previous section. The trial-averaged TPR versus the trial-averaged FP using the same injected isolated point source maps from the previous section is given for each fom below:





These results show that the Bartel's S-function performs worse than the CWT for faint source detection.