Study of Periodicity in Blazar Light Curves with a Machine Learning Approach

Paolo Cristarella Orestano[1,2]

Prof. Gino Tosti[1,2], Dr. Sara Cutini[2], Dr. Stefano Germani[1,2]

[1] University of Perugia

[2] INFN Perugia

Outline

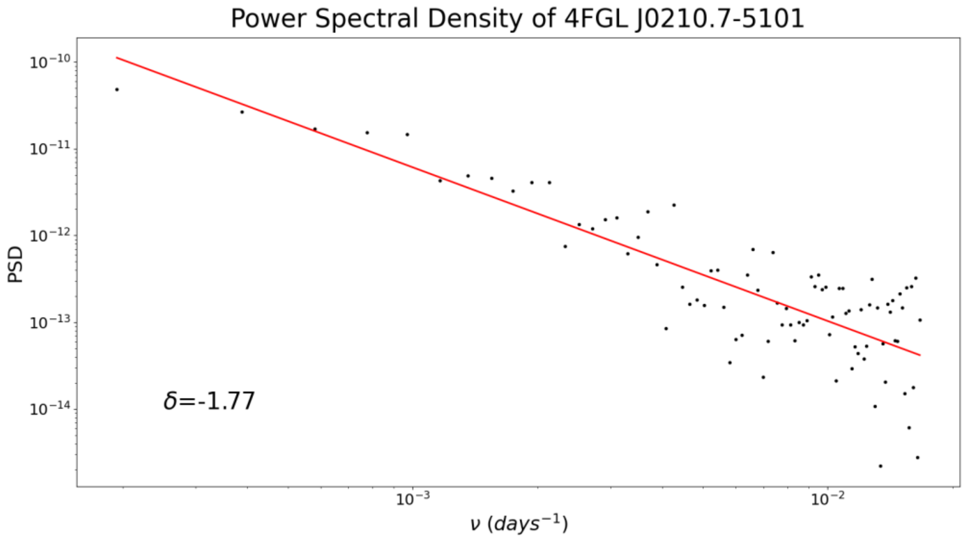
- **Blazar Variability and Periodicity**
- **Sources**
- **Analysis**
    - Weighted Wavelet Z-transform
    - Lomb Scargle Periodogram
    - Time Series and Noise Simulations
    - Emmanoulopoulos Simulations
- **Machine Learning**
- **Outlook**

Blazar Variability and Periodicity

Large variability at various wavelengths.

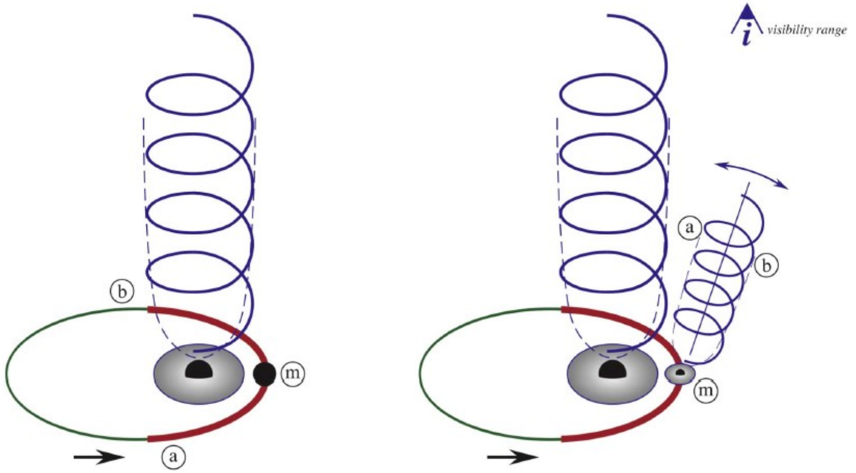
Pink or **red noise** type power spectrum $\propto f^{\delta}$.

The parameter δ goes from $\delta = -0.5$ to $\delta = -2$.



Long-term periodicities, or Quasi Periodic Oscillation, QPO, (founded in previous works [1][2]) could be related to **binary black holes** [3]:

- Intensity modulation.
- Precession, deflection or curvature of the jet changing the viewing angle.
- Not one but two jets.

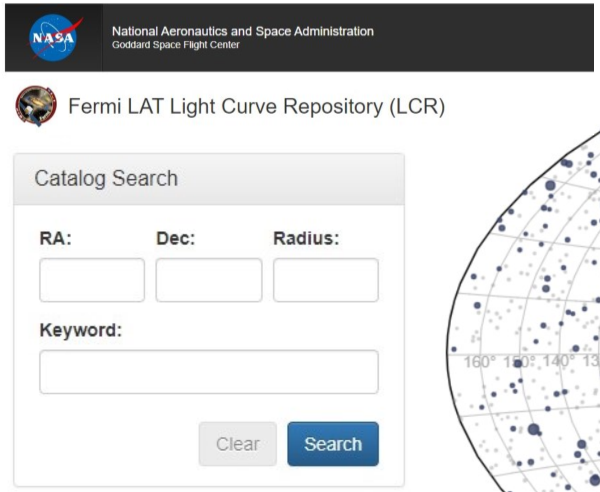


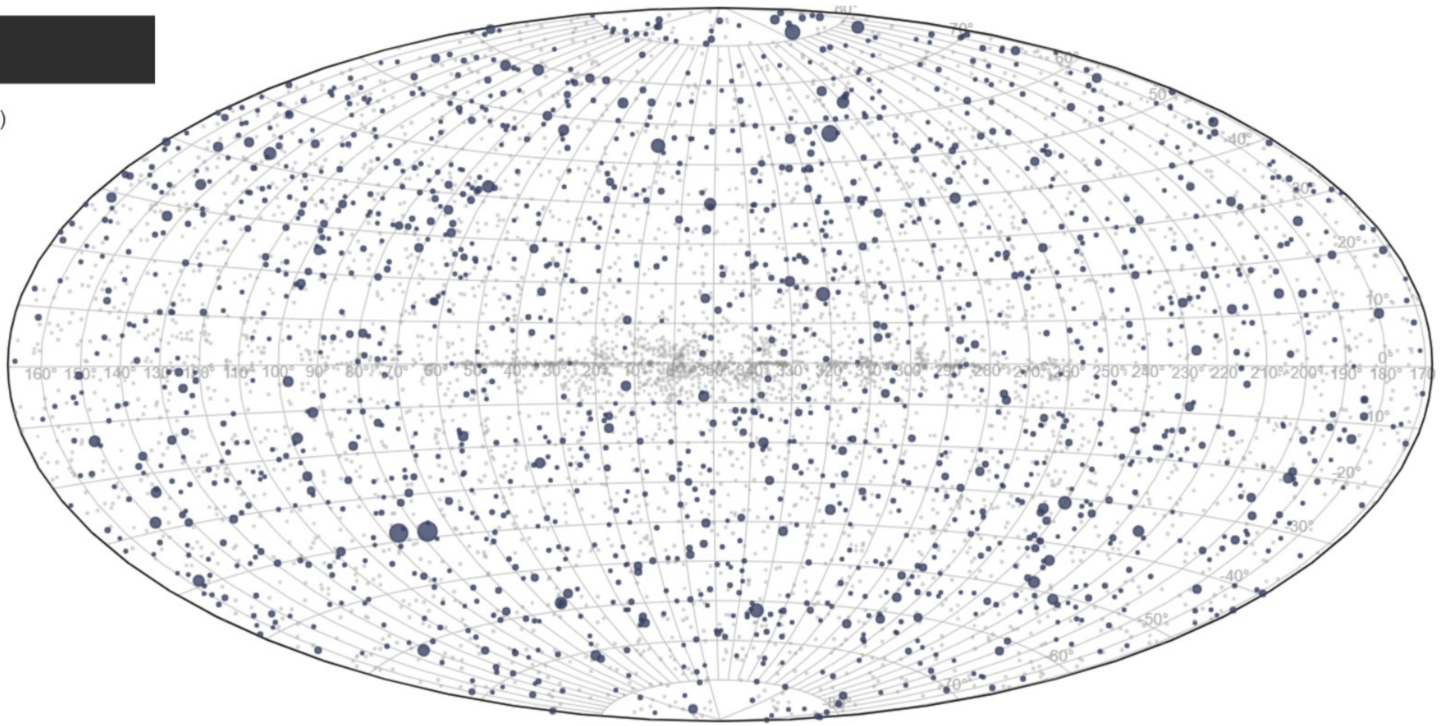
Tavani et al., 2018 [4]

Sources

From the **Light Curve Repository**[1] **1525** γ-ray **sources** analyzed: 571 FSRQ, 476 BLL, 371 unkown types of blazars, 107 other sources.

Six light curves **data types**: Energy flux (free index), Photon Flux (fixed index), 30d, 7d and 3d sampling.

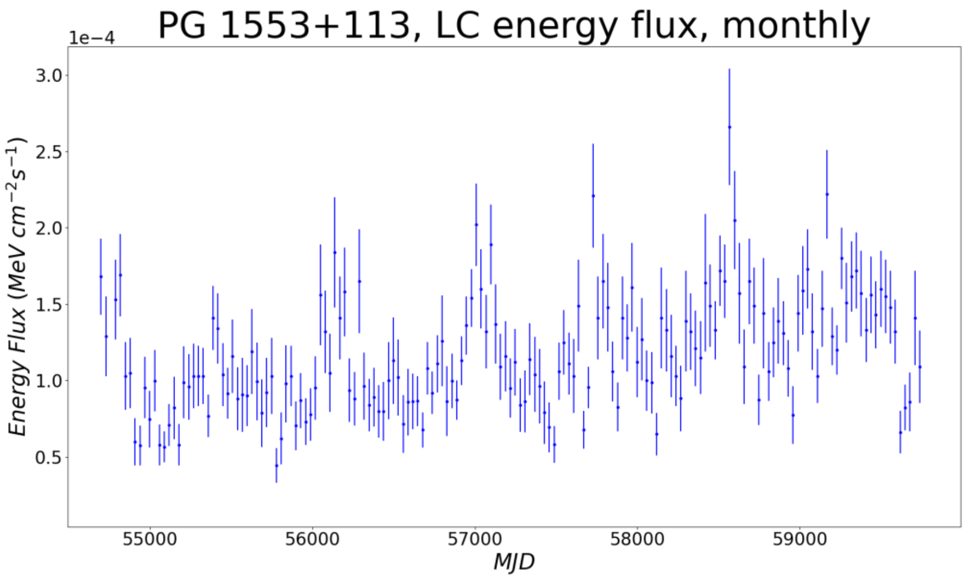


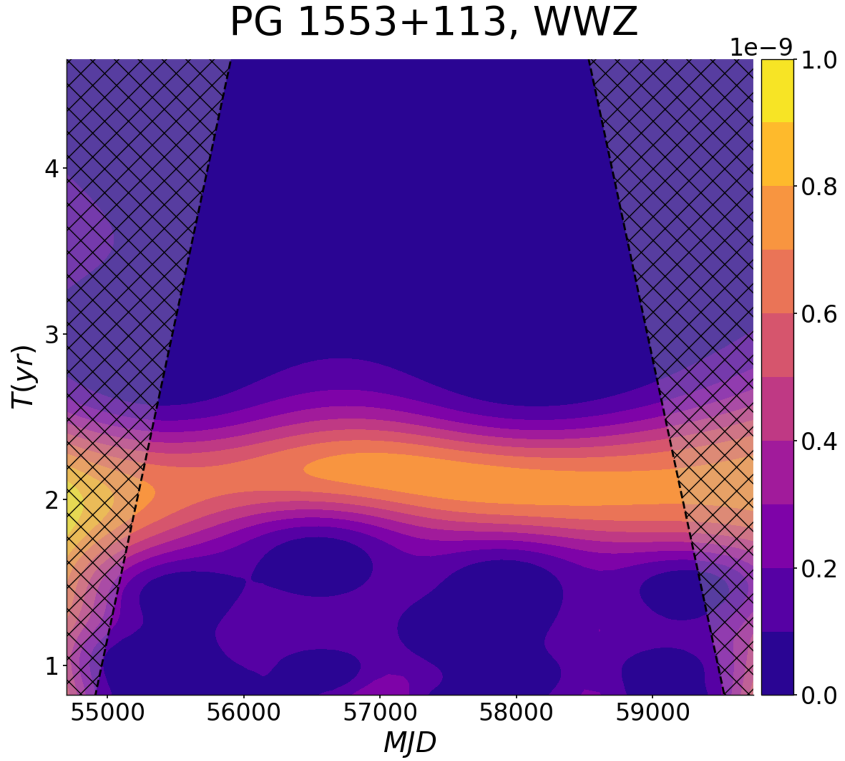


[1] https://fermi.gsfc.nasa.gov/ssc/data/access/lat/LightCurveRepository/

Analysis Weighted Wavelet Z-transform (WWZ)

Time series projection onto a model function (**Morlet**) [5], to detects transient periodicities, studies temporal evolution of data and signal parameters.

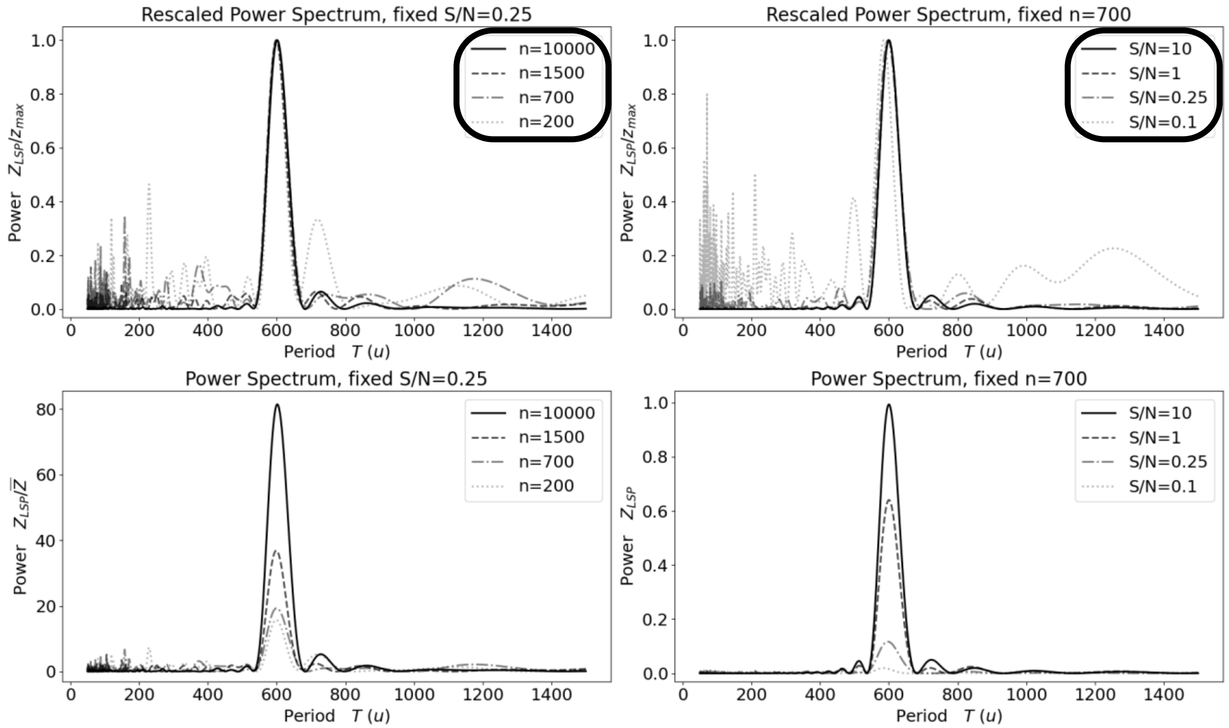




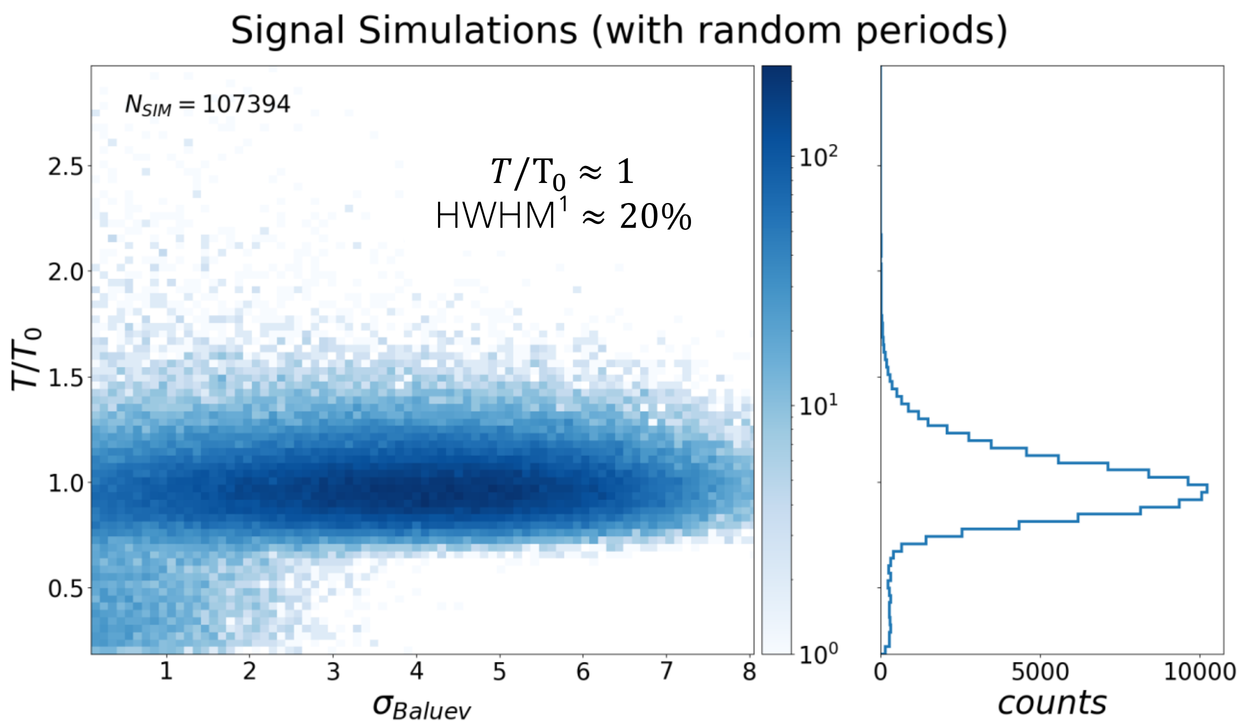
Analysis Lomb-Scargle Periodogram (LSP)

Square module of the Discrete Fourier transform. In addition to the changes by Scargle [6] we take into account some consideration by Vanderplas [7].

The number of points n and the signal-to-noise ratio S/N do not affect the width of the peak but only its **height**.



Analysis
Simulations



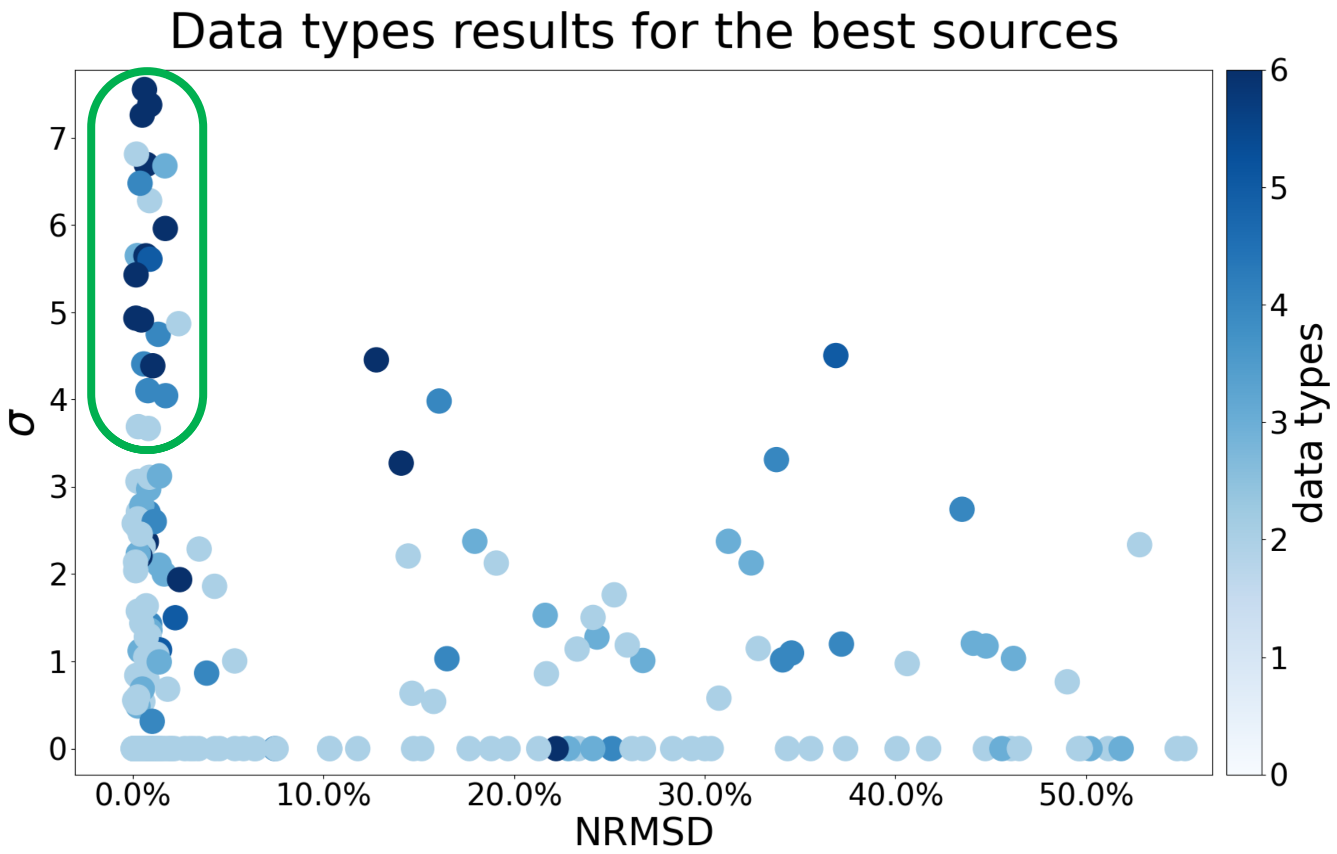
Preliminary significance estimation with **False Alarm Probability** (FAP) [7], comparing peak height with white noise background.

[1] Half Width at Half Maximum

Analysis

First Results from the Repository

Significant periodicities found in **23 blazars**. Many data types, $> 3\sigma$, and NRMSD[1] $< 3\%$. Cutoff 40% UL.

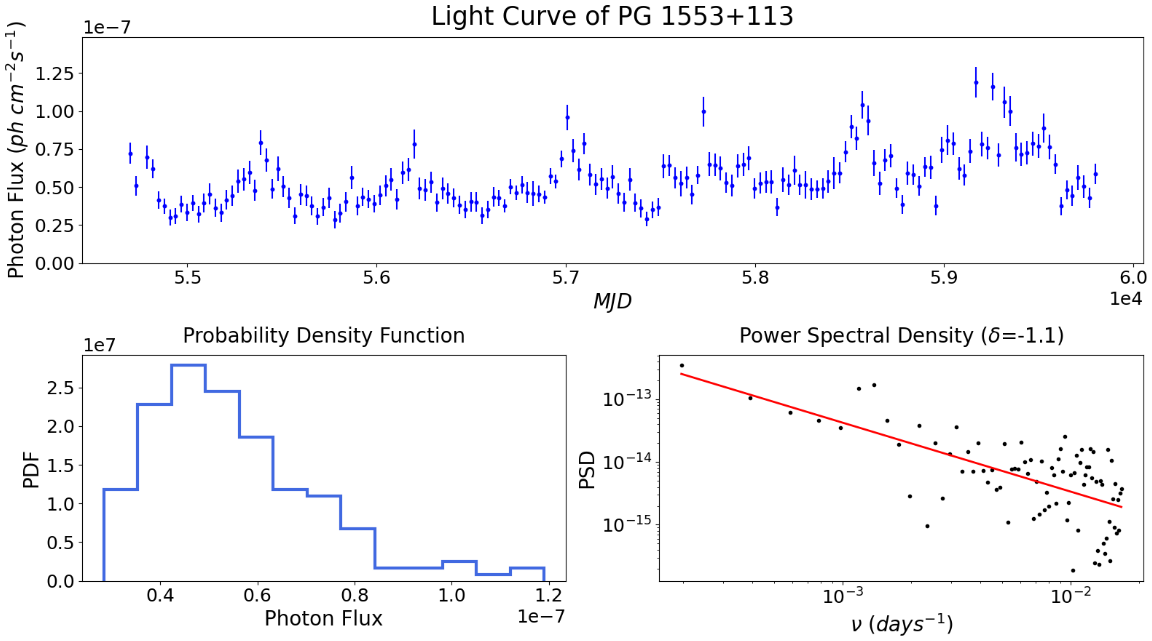


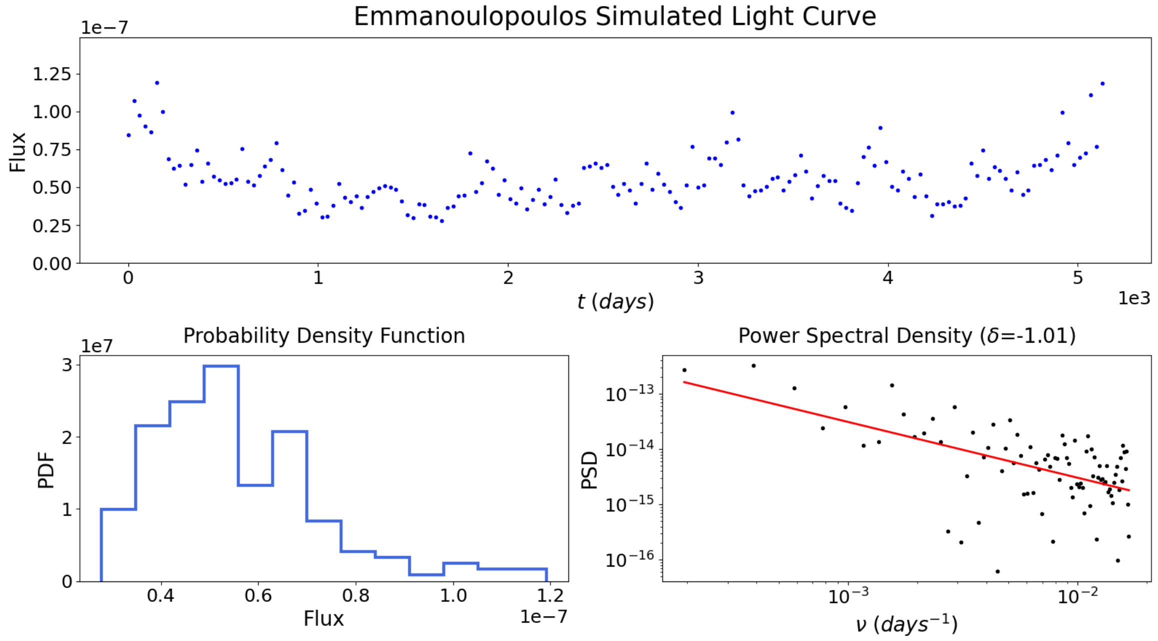
[1] Normalized Root Mean Square Deviation

Analysis

Emmanoulopoulos Simulations

Simulations with the Emmanoulopoulos algorithm [8] to study light curves with similar characteristics (**P**robability **D**ensity **F**unction and **P**ower **S**pectral **D**ensity).





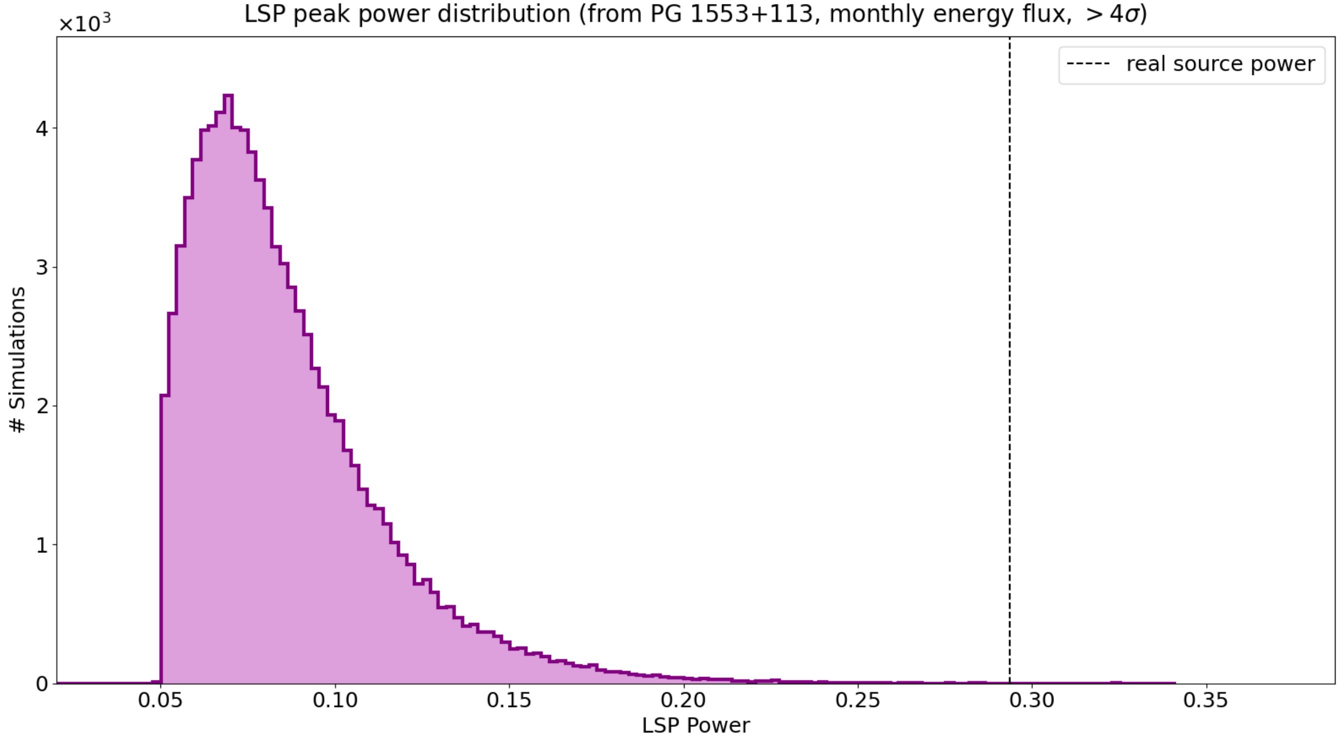
https://github.com/PaoloCO42/Emmanoulopoulos-Light-Curve-Simulations.git

Analysis

Best Results and Golden Sample

Selected a subsample, for each source (and each sampling and flux type) we perform **1 000 000 simulations**.

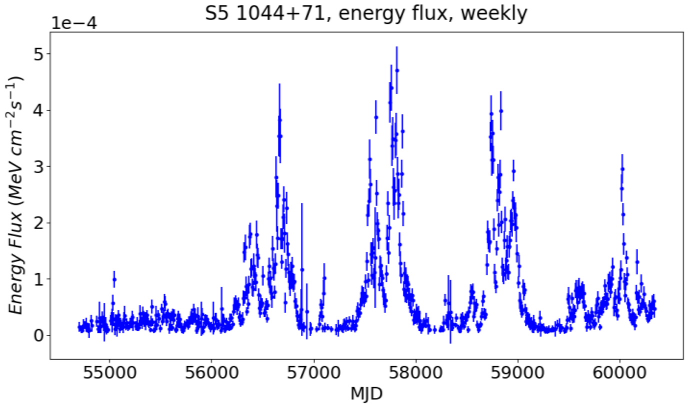
The *significance* is given by the number of times higher peak values are found in the LSP of the simulations.

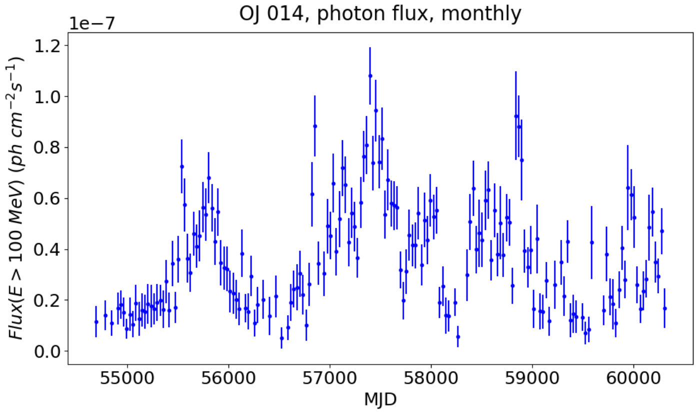


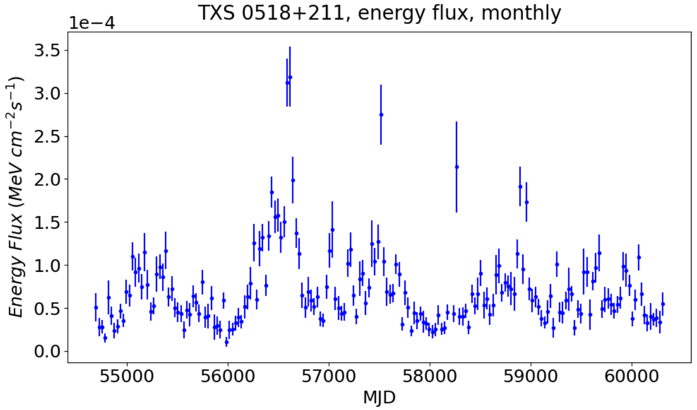
Source Name	P (yr)	Nσ
TXS 0059+581	4.2	3.4σ
MG1 J021114+1051	4.0	2.7σ
PKS 0215+015	3.4	2.1σ
PKS 0405-385	2.6	2.2σ
PKS 0426-380	3.7	2.9σ
PKS 0454-234	3.4	3.0σ
TXS 0518+211	3.0	4.3σ
S5 0716+71	2.4	1.4σ
OJ 014	4.1	$> 4.7\sigma$
S4 0814+42	2.2	4.1σ
S5 1044+71	3.1	$> 4.7\sigma$
S4 1144+40	3.3	$> 4.7\sigma$
B2 1215+30	2.8	3.0σ
4C +21.35	3.7	1.8σ
PG 1553+113	2.1	$> 4.7\sigma$
PKS 2155-304	1.6	1.5σ

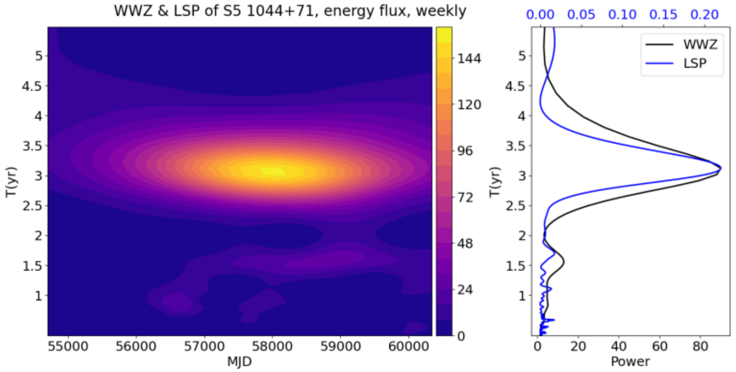
Analysis

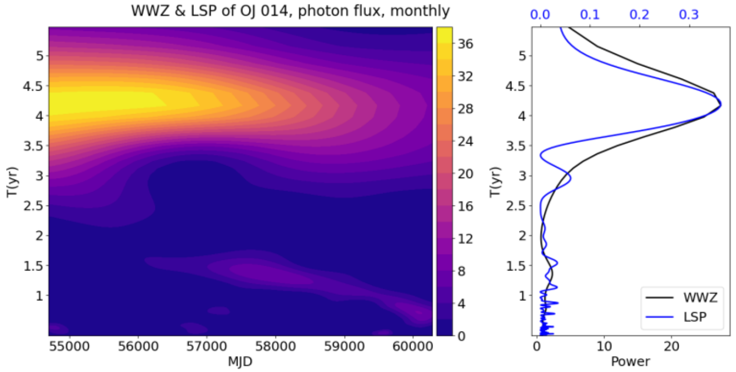
Best Results and Golden Sample

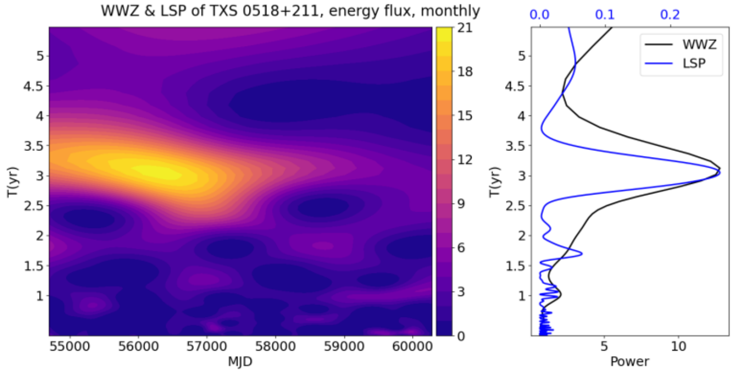












Machine Learning

t-Stochastic Neural Embedded neighbor algorithm unsupervised [9] (**t-SNE**):

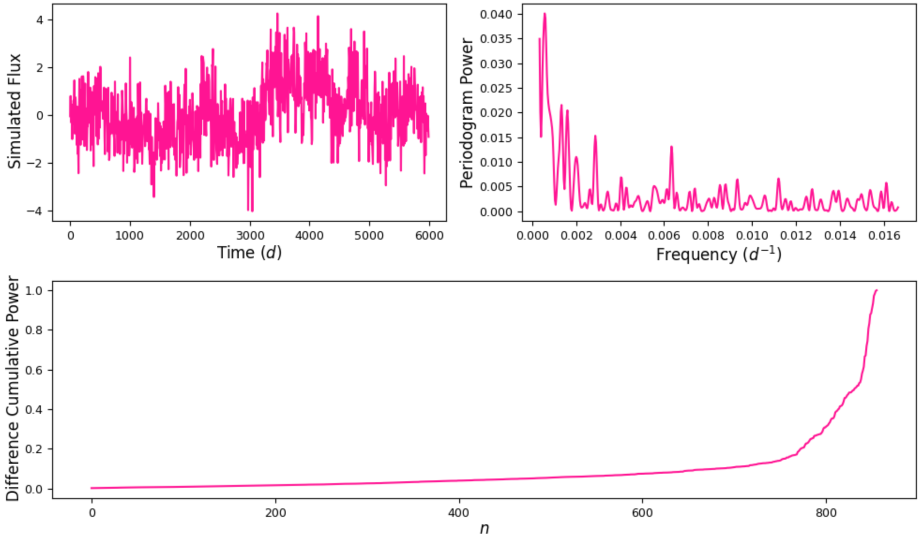
Calculates the probability that each high vectors $\{x_j\}$ should be consider a neighbor of $\{x_i\}$ from the set of Euclidean distances $\{|\vec{x_j} - \vec{x_i}|\}$.

Since the vectors $\{x\}$ are compared point to point we opted for the normalized **Difference Cumulative** of the sorted periodogram **Power**.

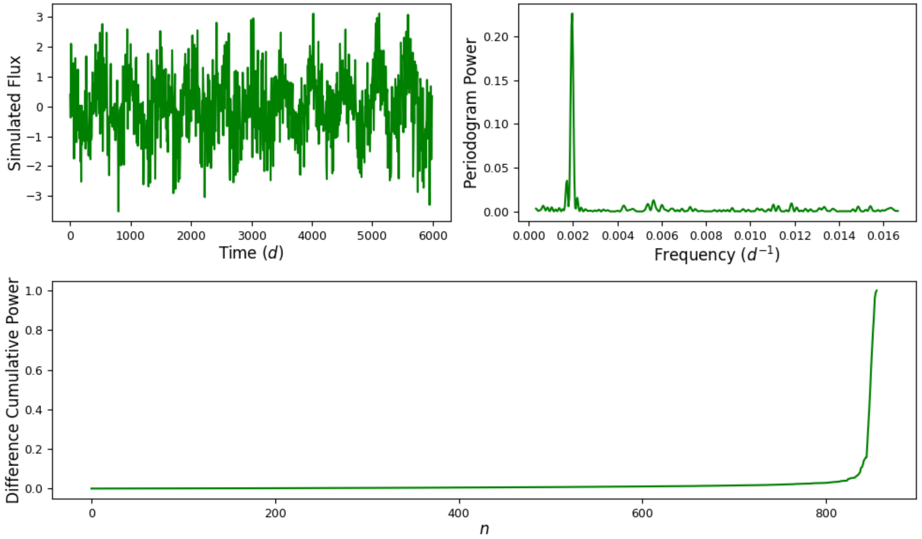
LSP power → sort → cumulative →

→ difference $(a_{i+1} - a_i)$ → normalize

Pink / Red Noise Simulation



Signal Simulation



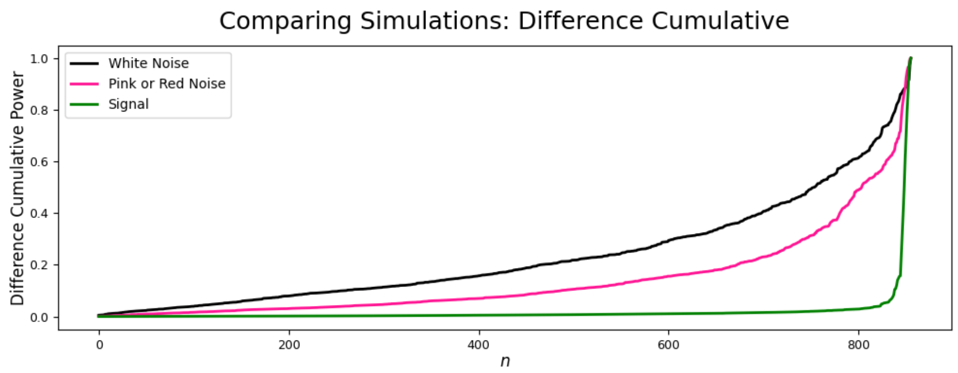
Machine Learning

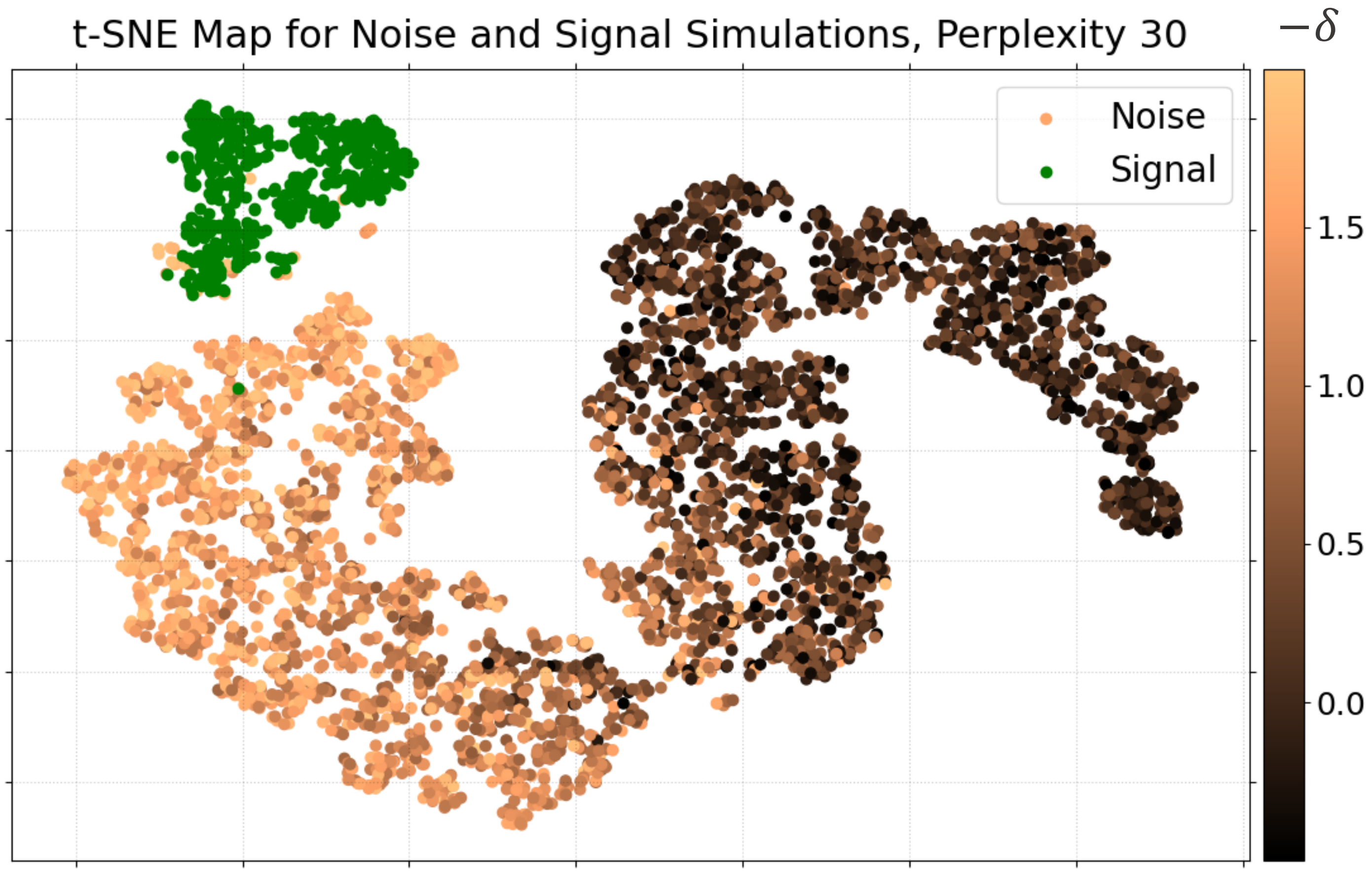
t-SNE produces a simplified map where the axes of this 2D space have no proper labels or meaning.

With this type of input the simulations can be distinguished.

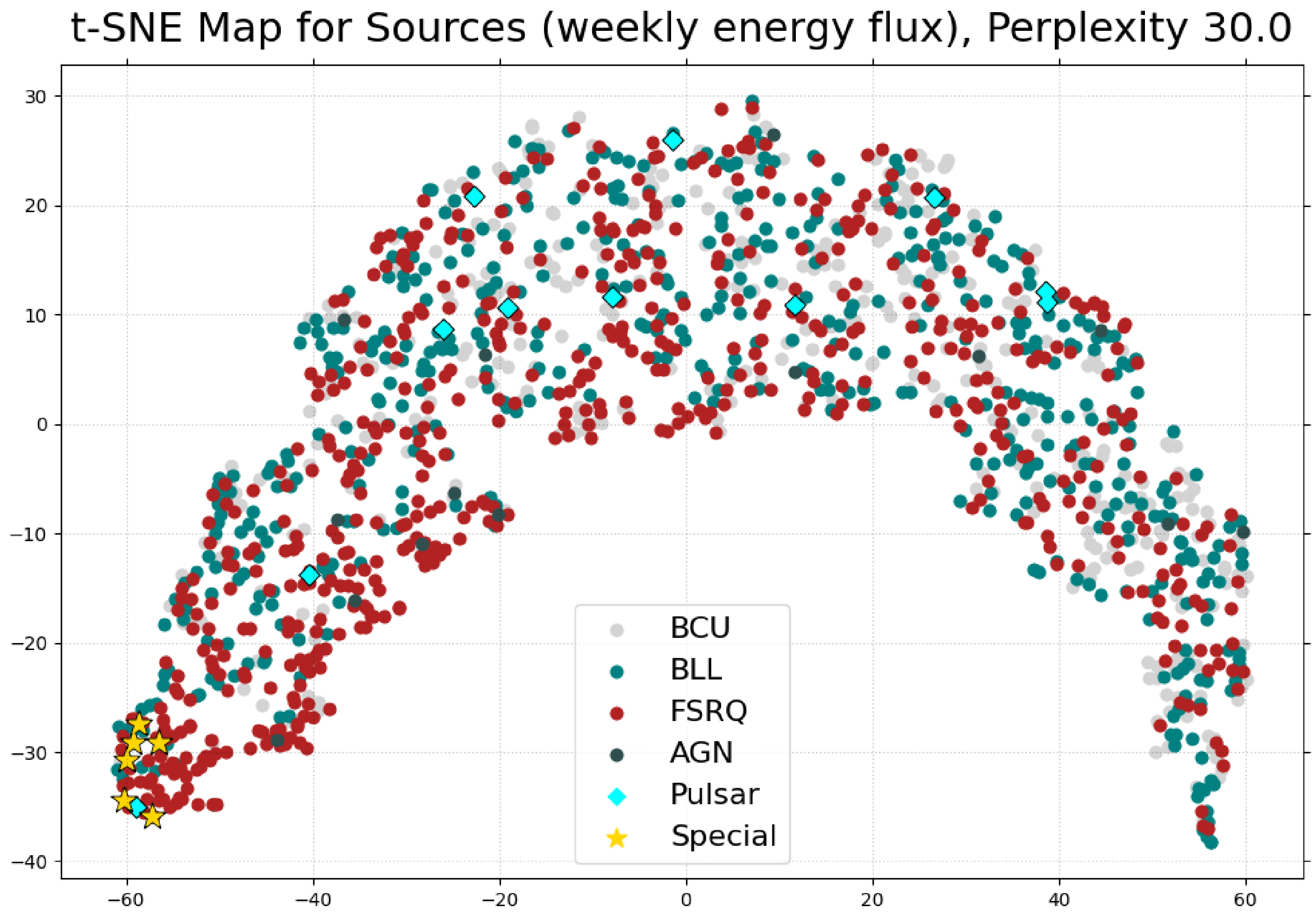
Characteristics of the simulations:

SNR = 1, $-2.0 < \delta_{noise} < +0.5$





Machine Learning



With the sources around 60 of them are clustered near the **golden sample**.

Outlook

- Periodicities in 1% (*similar sources in a previous work of Peñil et al*. [2]) of variable sources with 16 years data from LCR, and a **golden sample of 6 sources**.
- **t-SNE** method seems to bring out a larger cluster, but not separated enough.
- Increase the complexity of noise and signal simulation to obtain similar morphology in the map and try adding different types of **input data** (also multiwavelength).
- We do not exclude other machine learning methods, that may be better suited to the data and the goal.

Outlook

- Periodicities in 1% (*similar sources in a previous work of Peñil et al.* [2]) of variable sources with 16 years data from LCR, and a **golden sample of 6 sources**.
- **t-SNE** method seems to bring out a larger cluster, but not separated enough.
- Increase the complexity of noise and signal simulation to obtain similar morphology in the map and try adding different types of **input data** (also multiwavelength).
- We do not exclude other machine learning methods, that may be better suited to the data and the goal.

?!

Outlook

- Periodicities in 1% (*similar sources in a previous work of Peñil et al*. [2]) of variable sources with 16 years data from LCR, and a **golden sample of 6 sources**.

- **t-SNE** method seems to bring out a larger cluster, but not separated enough.

- Increase the complexity of noise and signal simulation to obtain similar morphology in the map and try adding different types of **input data** (also multiwavelength).

- We do not exclude other machine learning methods, that may be better suited to the data and the goal.

New outlook (more work) thanks to Fermi Summer School

- Comparing spectral models with a Bayesian approach for the golden sample for a more robust analysis.

Bibliography

[1] M. Ackermann, et al., “*Multiwavelength evidence for quasi-periodic modulation in the gamma-ray blazar pg 1553+113*”, The Astrophysical Journal, vol. 813, p. L41,Nov 2015.

[2] P. Peñil, et al., “*Systematic Search for γ-Ray Periodicity in Active Galactic Nuclei Detected by the Fermi Large Area Telescope*”, ApJ, vol. 896, p. 134, June 2020.

[3] S. Komossa, “*Observational evidence for binary black holes and active double nuclei*”, Memorie della Società Astronomica Italiana, vol. 77, p. 733, 2006.

[4] M. Tavani, Marco, et al. "*The blazar PG 1553+ 113 as a binary system of supermassive black holes*" , ApJ 854.1 (2018): 11.

[5] G. Foster, “*Wavelets for period analysis of unevenly sampled time series*”, AJ, vol. 112, pp. 1709–1729, Oct. 1996.

[6] J. D. Scargle, “*Studies in astronomical time series analysis. II. Statistical aspects of spectral analysis of unevenly spaced data*”, ApJ, vol. 263, pp. 835–853, Dec. 1982.

[7] J. T. VanderPlas, “*Understanding the Lomb-Scargle Periodogram*”, ApJS, vol. 236, p. 16, May 2018.

[8] D. Emmanoulopoulos, I. M. McHardy, and I. E. Papadakis. "*Generating artificial light curves: revisited and updated.*" Monthly Notices of the Royal Astronomical Society 433.2 (2013): 907-927.

[9] L. van der Maaten and G. Hinton, “*Visualizing Data using t-SNE*”, Journal of Machine Learning Research 9 (86), 2579–2605, 2008.

Backup Slides

Weighted Wavelet Z-transform (WWZ)

Time series projection onto a model function (Morlet) [3].

$$W = \sum_{k}^{N} x(t_k) F^*(t_k)$$

With the S-matrix $S_{ab} = \langle F_a | F_b \rangle$ as our metric tensor with statistical weights w in it, then the inner product of two function is:

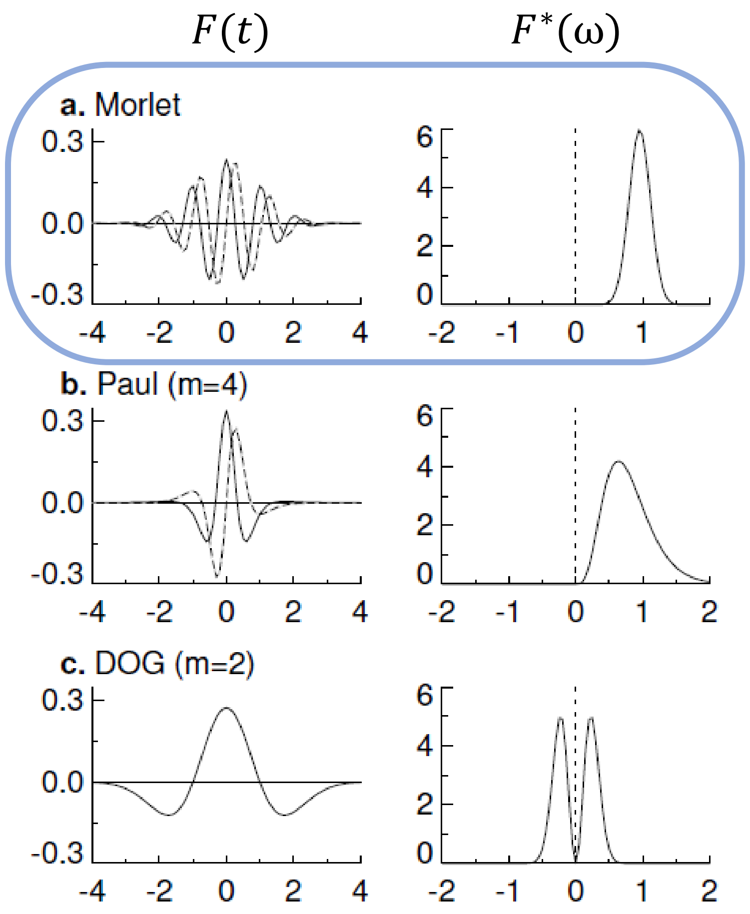
$$\langle f | g \rangle = \sum_a \sum_b S_{ab}\, f_a\, g_b$$

Hence, using the variations $V_x = \langle x|x \rangle - \langle 1|x \rangle^2$ and $V_R = \langle R|R \rangle - \langle 1|R \rangle^2$ where the first is variations of the data and the second of the residual vector.

The WWZ is defined as:

$$Z = \frac{(N_{eff} - 3) V_R}{2(V_x - V_R)}$$

Where $\mathrm{N_{eff}} = (\sum w_\alpha)^2 / \sum w_\alpha^2$.



https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Lomb-Scargle Periodogram (LSP)

Square module of the Discrete Fourier transform. In addition to the changes by Scargle [4] we take into account some consideration by Vanderplas [5].

Signal and noise: $x_j = x(t_j) = x_{S\,j} + r_j$ Classic periodogram: $P_x(\omega) = \frac{1}{N}|DFT_x(\omega)|^2 = \frac{1}{N}\left|\sum_{j=1}^{N} x(t_j) e^{-i\omega t_j}\right|^2$

Not useful in case of noisy data. Spectral Leakage.

Discrete Fourier Transform:

$$DFT_x(\omega) = \sqrt{N/2}\sum_{j=1}^{N} x(t_j)(A(\omega)\cos\omega t_j + iB(\omega)\sin\omega t_j)$$

If $A(\omega) = B(\omega) = \sqrt{2/N}$ we obtaing the classic periodogram. Lomb-Scargle changes:

$$A(\omega) = \left(\sum_j \cos^2 \omega t_j\right)^{-1/2} \quad ; \quad B(\omega) = \left(\sum_j \sin^2 \omega t_j\right)^{-1/2}$$

Modified periodogram:

$$P_x(\omega) = \frac{1}{2}\left(\frac{\left(\sum_j x_j \cos\omega(t_j - \tau)\right)^2}{\sum_j x_j \cos^2\omega(t_j - \tau)} + \frac{\left(\sum_j x_j \sin\omega(t_j - \tau)\right)^2}{\sum_j x_j \sin^2\omega(t_j - \tau)}\right) \quad \text{with} \quad \tan 2\omega\tau = \frac{\sum_j \sin 2\omega t_j}{\sum_j \cos 2\omega t_j}$$

False Alarm Probability (FAP)

To compare **peak height** with **background** and **spurious peaks**.

Probability that a peak of a certain height Z will be found from a data set consisting of white noise.

❖ Naive

The cumulative probability of observing a value less than Z with white noise is $P(Z) = 1 - e^{-Z}$.

$$FAP_{Naive}\,(Z) = 1 - \left(P(Z)\right)^{N_{eff}}$$

❖ Bootstrap

Randomization of the time series, high computational cost. Simulations required: N=10/r, for a false positive rate r.

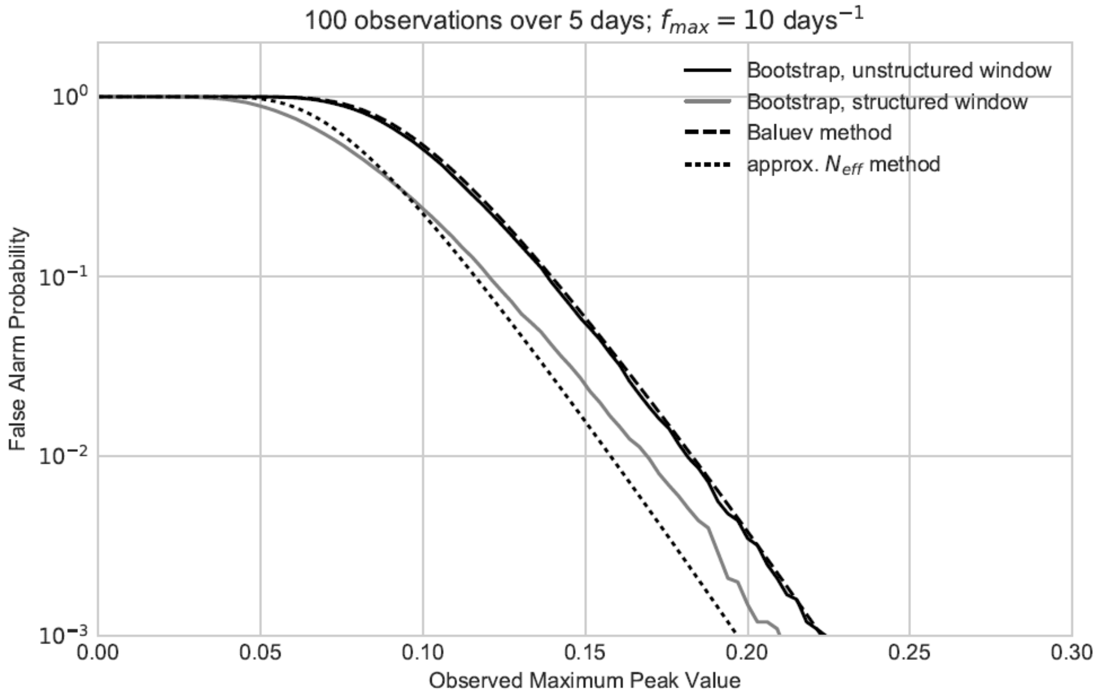
❖ Baluev

Extreme value theory for random processes. Upperbound for the FAP.

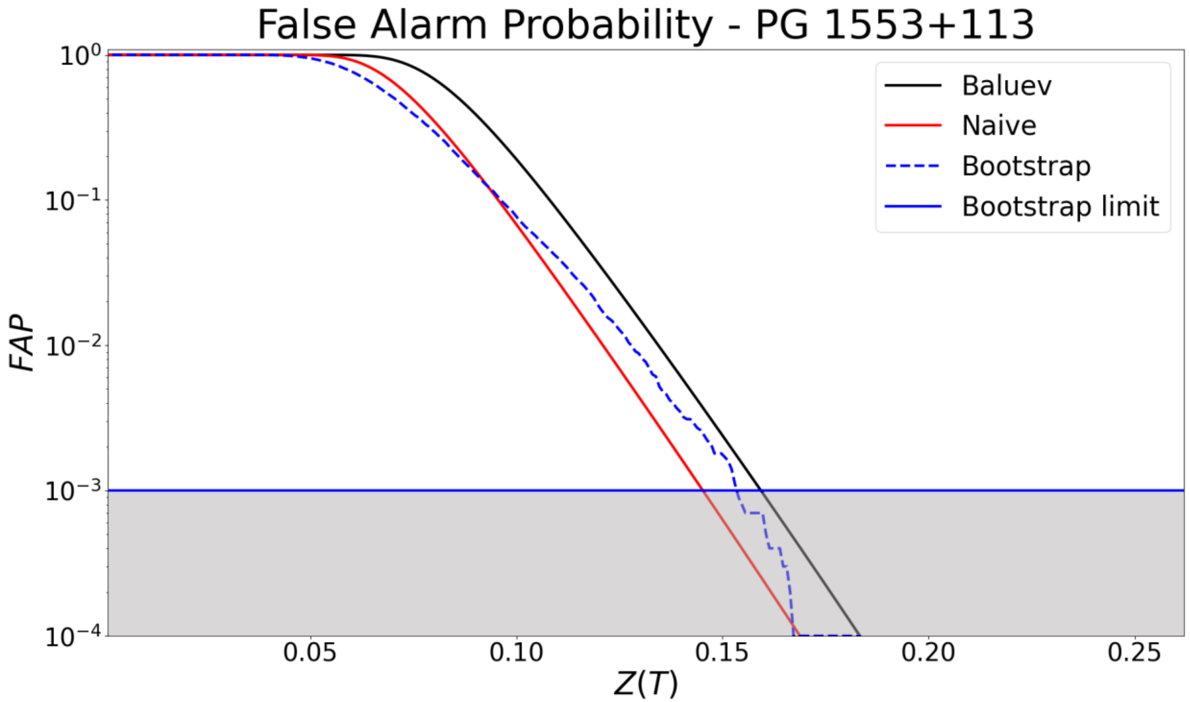
False Alarm Probability (FAP)

Bootstrap is a the most reliable method, **Naive** overestimates the significance, while **Baluev** underestimates it.

From the FAP we extrapolate a fictitious number of σ.



From Vanderplas, 2016 [5]



Baluev FAP

Two hypothesis: noise H and periodic signal K. FAP comes from the probability distribution of maxima in periodograms under H hypothesis.

Least-squares periodogram:

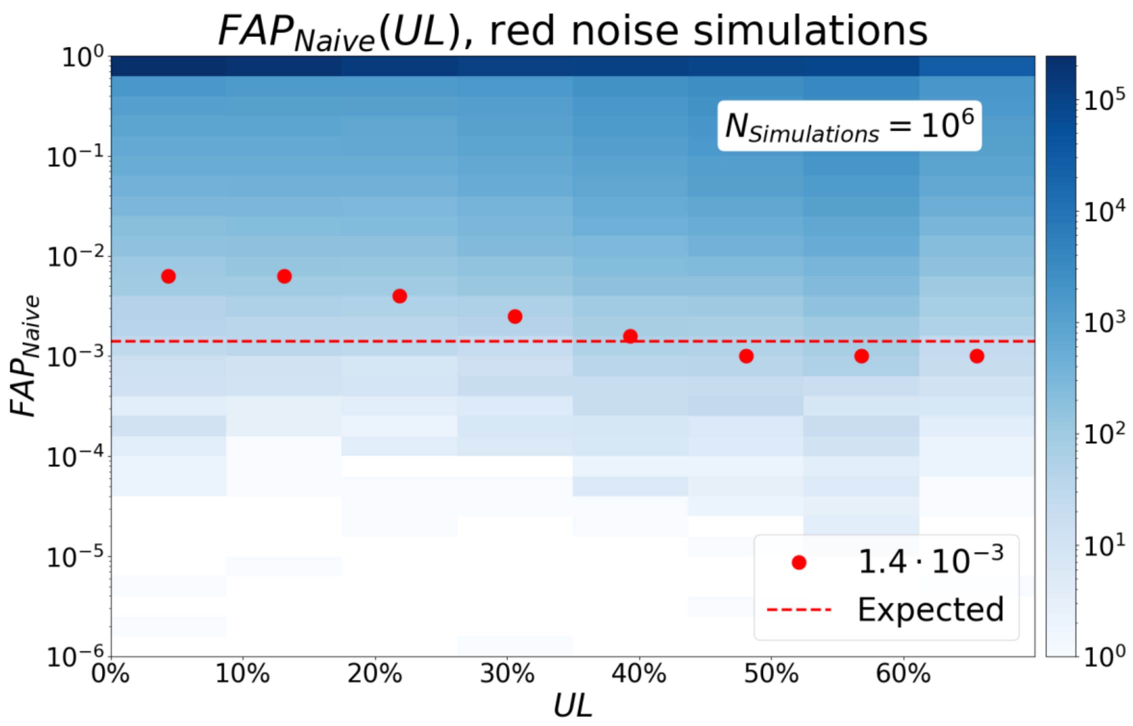
$$z(f) = \frac{\chi_H^2 - \chi_K^2(f)}{2}$$

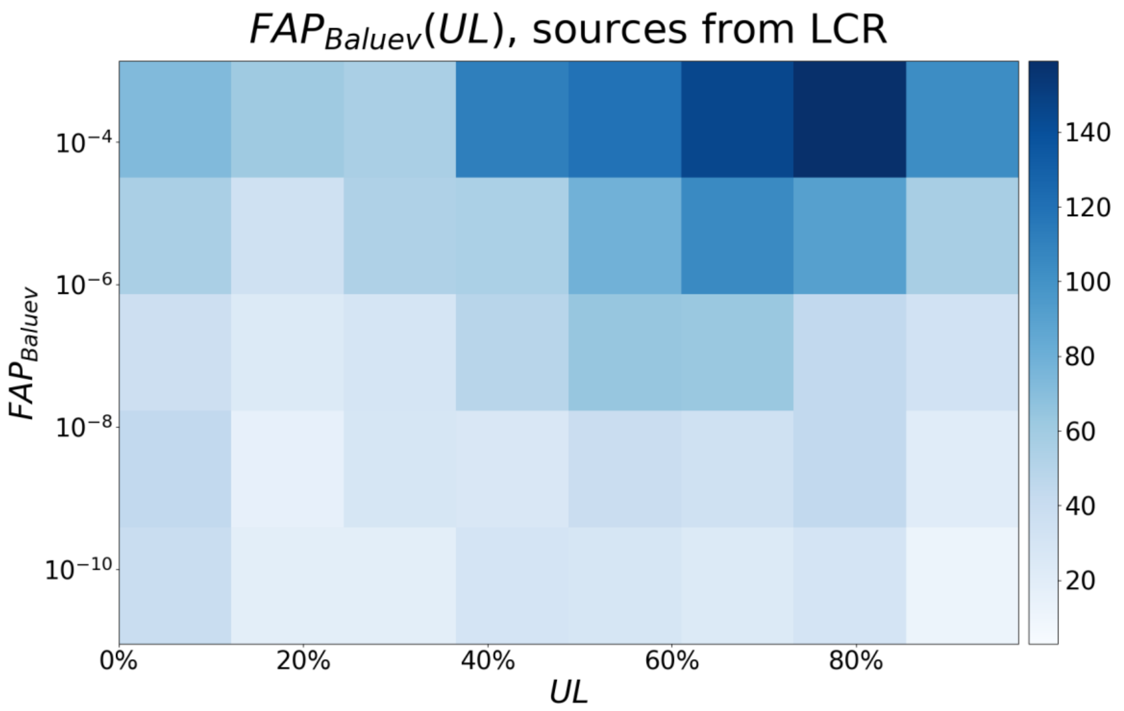
With the theory of random processes we estimate the FAP:

$$FAP_{Baluev}(z, f_{max}) = 1 - \exp(-W e^{-z} \sqrt{z}\,)$$

$$W = f_{max} T_{eff} \quad ; \quad T_{eff} = \sqrt{4\pi Dt} \quad ; \quad Dt = \overline{t^2} - \bar{t}^2$$

False Alarm Probability (FAP)

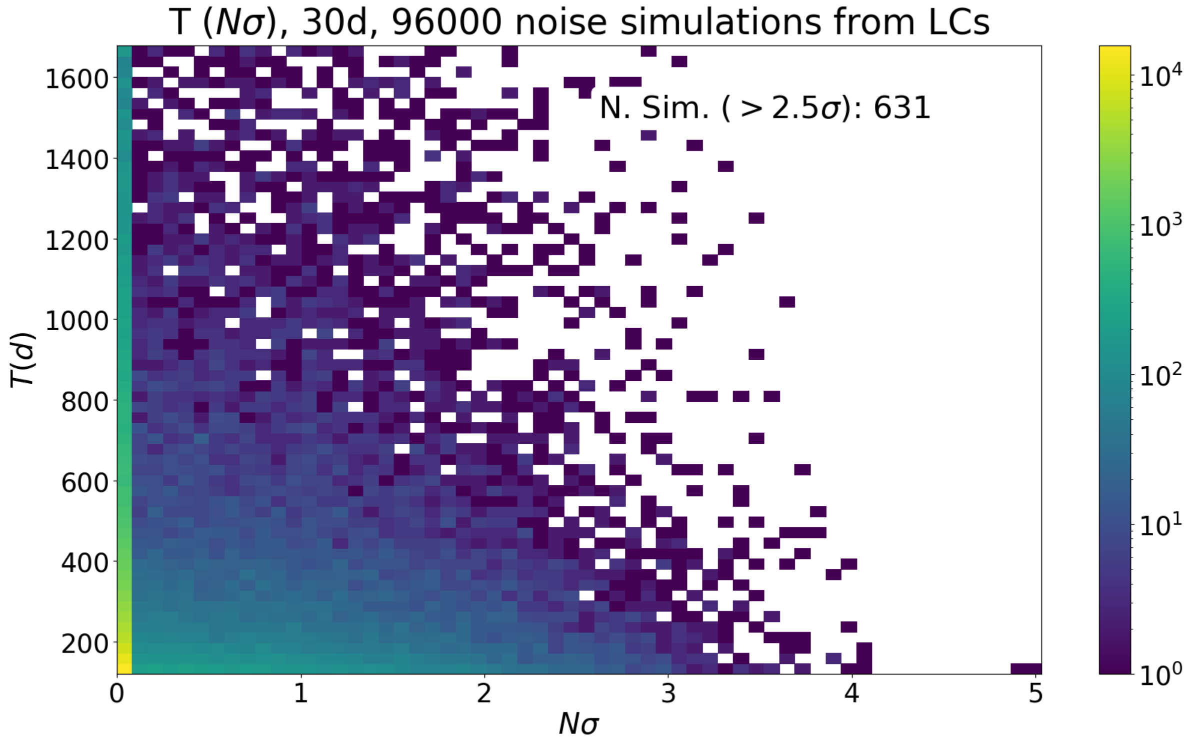


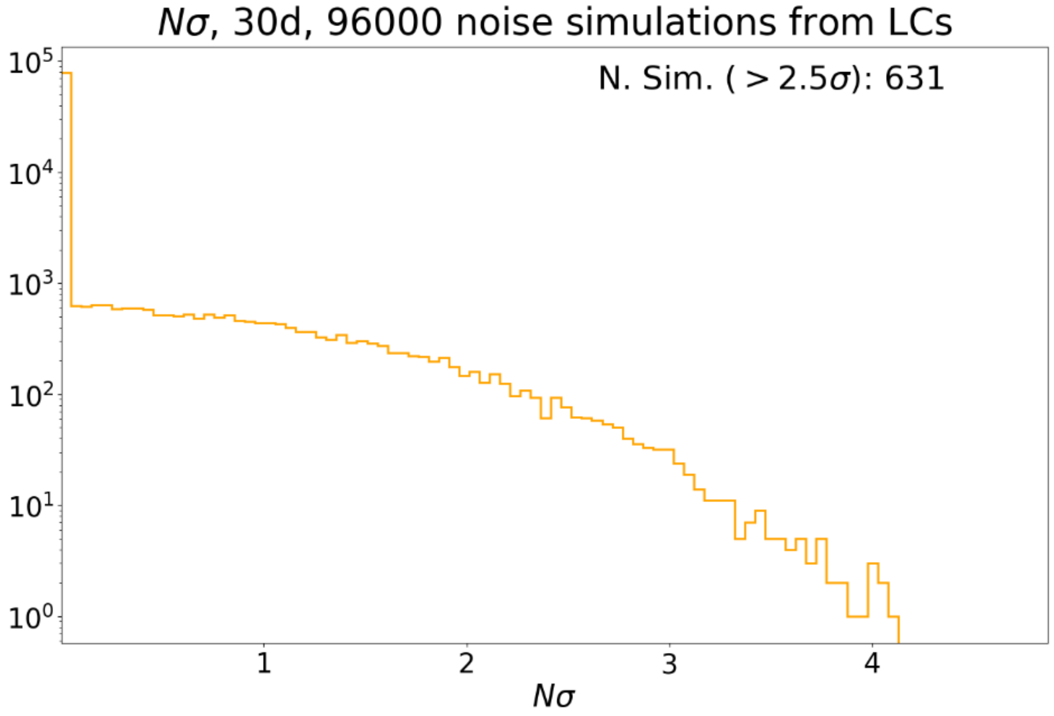


Time Series and Noise Simulations

White Noise simulations as uniform randomization of Blazar light curves.

A period with significance $> 2.5\sigma$ is found in less then 1% simulations.





[1] Half Width at Half Maximum

Emmanoulopoulos Simulations

Simulations with the Emmanoulopoulos algorithm [6] to study light curves with similar characteristics (**P**robability **D**ensity **F**unction and **P**ower **S**pectral **D**ensity).

First step:
From the PSD make Timmer-König simulation: $x_{TK}\,(t)$, perform the DFT (Discrete Fourier Function) on it and take the amplitude A_{TK}.

Second step:
Take $x_0(t)$, White Noise simulation obtained from the PDF, perform the DFT and take phase ϕ_0.

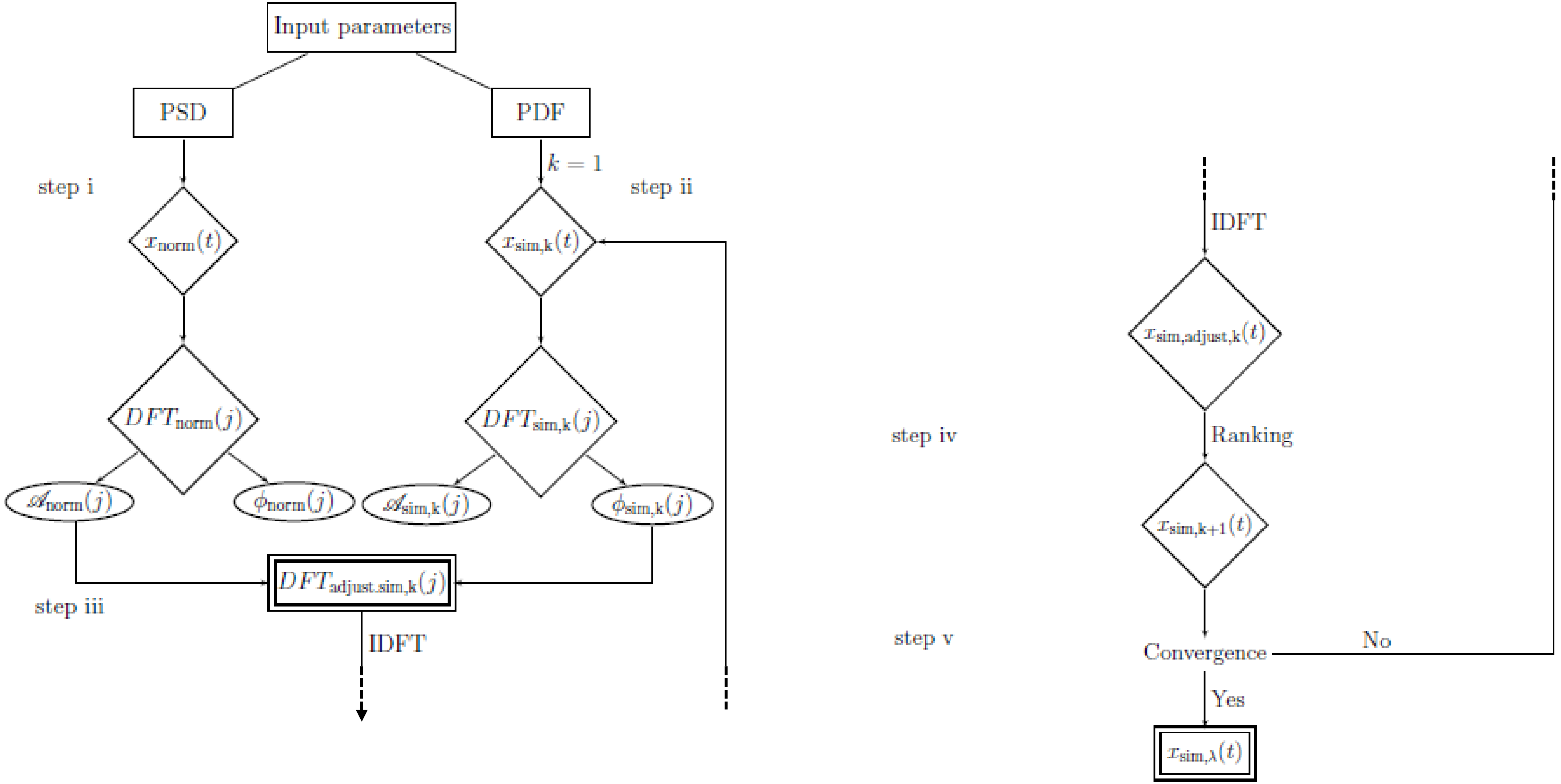
Third step:
Combine A_{TK} and ϕ_0 to have a $X(j)$ in the frequency domain, perform the IDFT (Inverse) obtaining $x_C(t)$.

Fourth step:
Sort $x_C(t)$ in descending order and replace the values x_C with x_0, also sorted in descending order, obtaining $x_i(t)$, where i is the number of iterations.

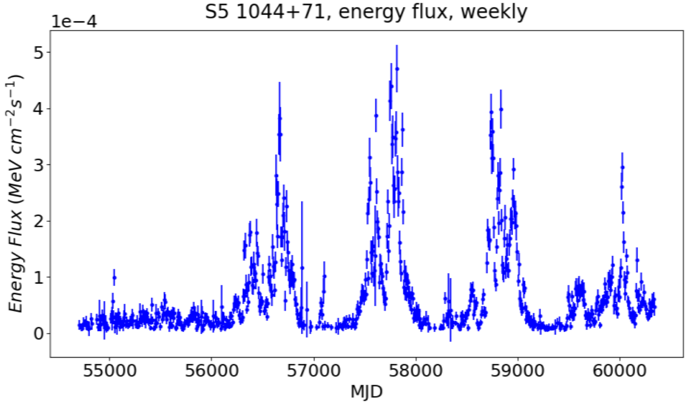
Fifth step:
Replacing $x_0(t)$ with the new $x_i(t)$, repeat the process from the second step until it converges i.e. when $x_i(t)$ and $x_{i-1}(t)$ have the same PSD.

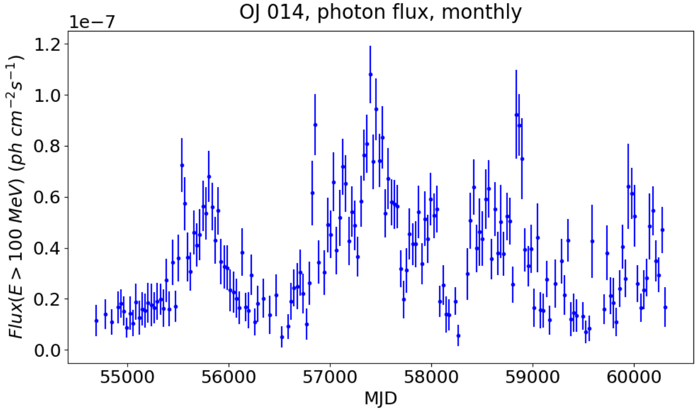
https://github.com/PaoloCO42/Emmanoulopoulos-Light-Curve-Simulations.git

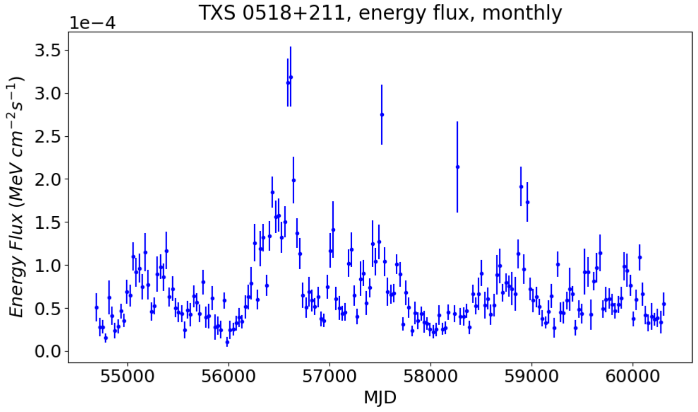
Emmanoulopoulos Simulations

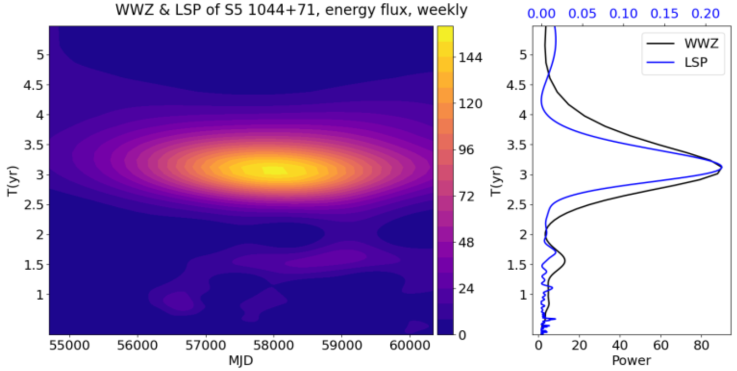


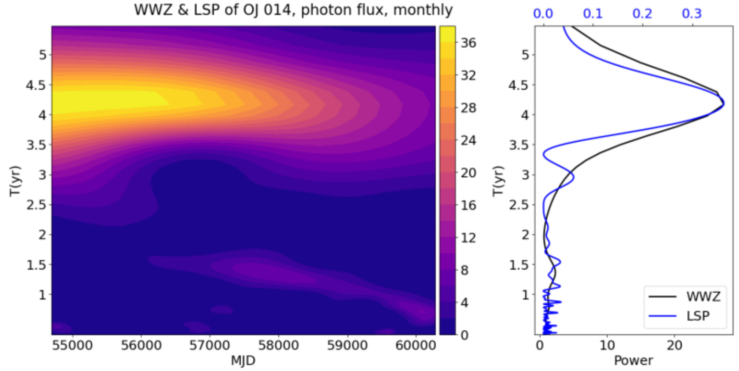
Golden Sample

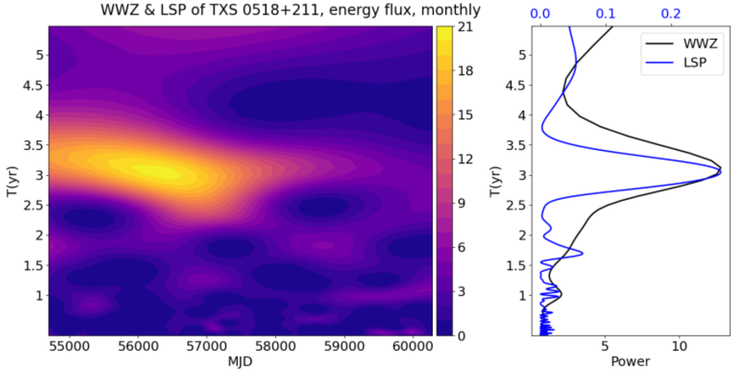




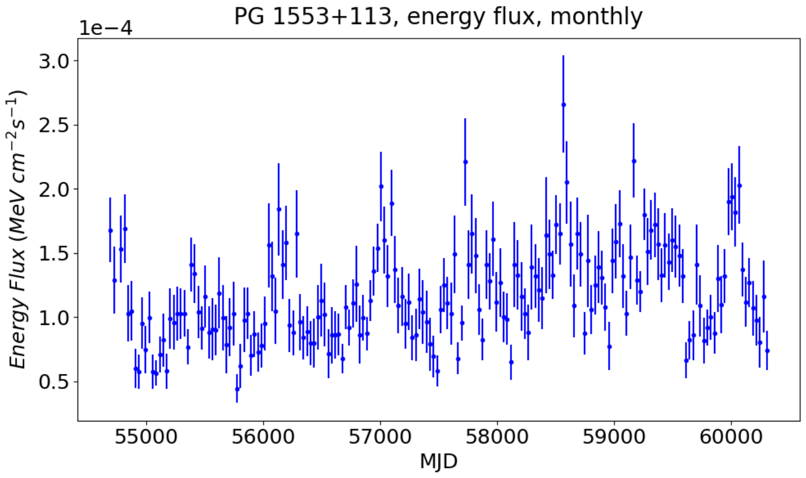


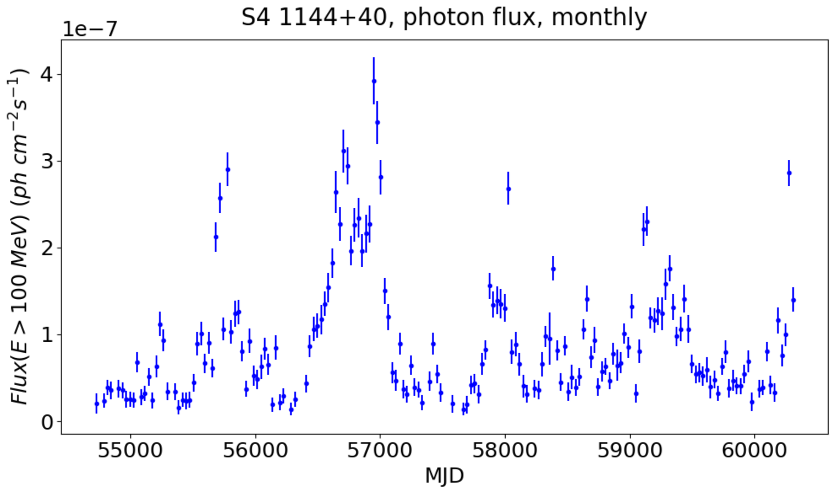


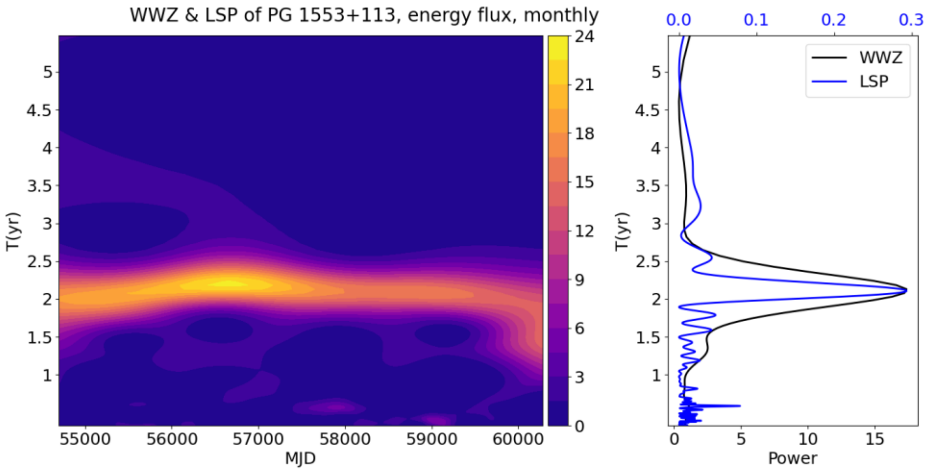


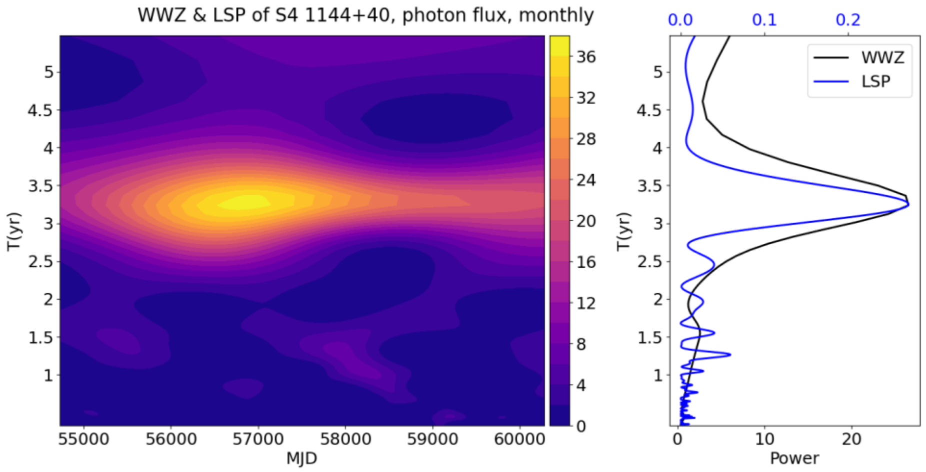


Golden Sample

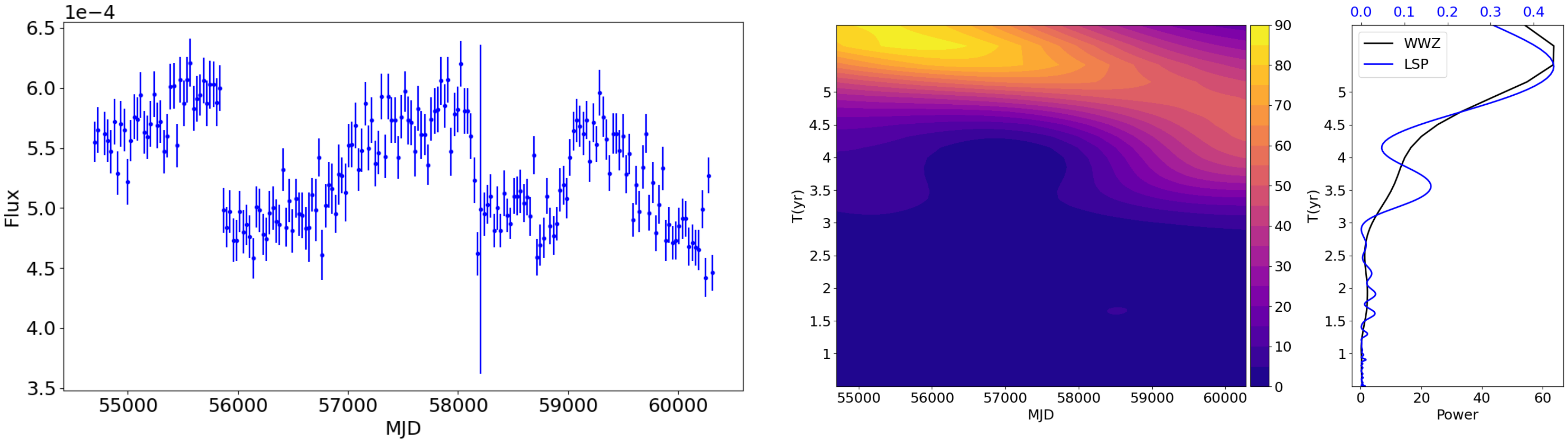








Pulsar



1e−4
6.5
6.0
5.5
5.0
4.5
4.0
3.5
Flux
55000
56000
57000
58000
59000
60000
MJD
5
4.5
4
3.5
3
2.5
2
1.5
1
T(yr)
55000
56000
57000
58000
59000
60000
MJD
90
80
70
60
50
40
30
20
10
0
0.0
0.1
0.2
0.3
0.4
WWZ
LSP
T(yr)
0
20
40
60
Power

Look-elsewhere effect

Apparent statistically significant observation, which has arisen from searching a large parameter space.

The **Trial Factor** (**N**) is used to account this effect. It is the ratio between the probability of observing a possible excess at some fixed point, to the probability of observing it anywhere in the range.

The local significance is reduced to the global significance (local>global).

$$p_{GLOBAL} = 1 - (1 - p_{LOCAL})^N$$

In this case someone considers $N = P \cdot B$, where P is the number of independent periods and B is the number of blazars.

In a previous work on the catalog: $N = 35 \cdot 351 = 12{,}285$, so a local $5.5\,\sigma$ become $2.8\,\sigma$ global significance). While in a paper about PG 1553+113, is $N = P \cdot 1 = 43$.

t-SNE, theory

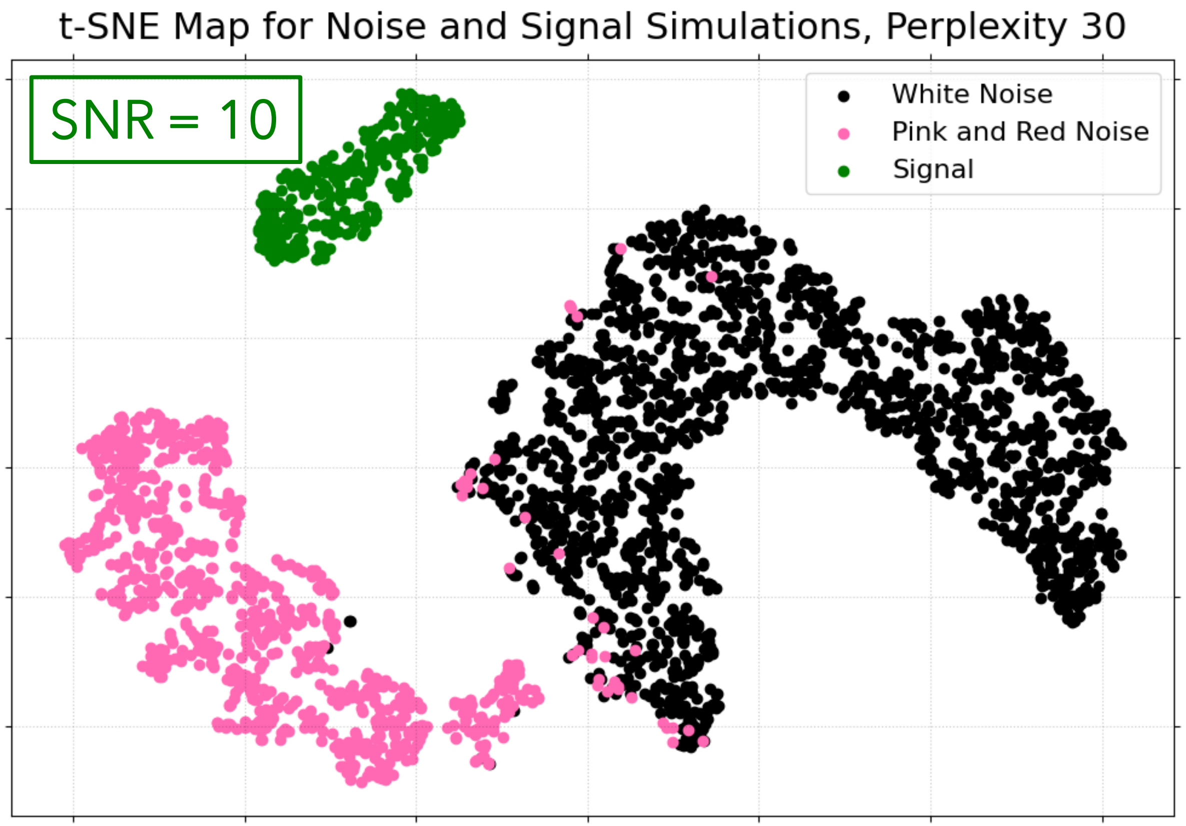
Algorithm 1: Simple version of t-Distributed Stochastic Neighbor Embedding.

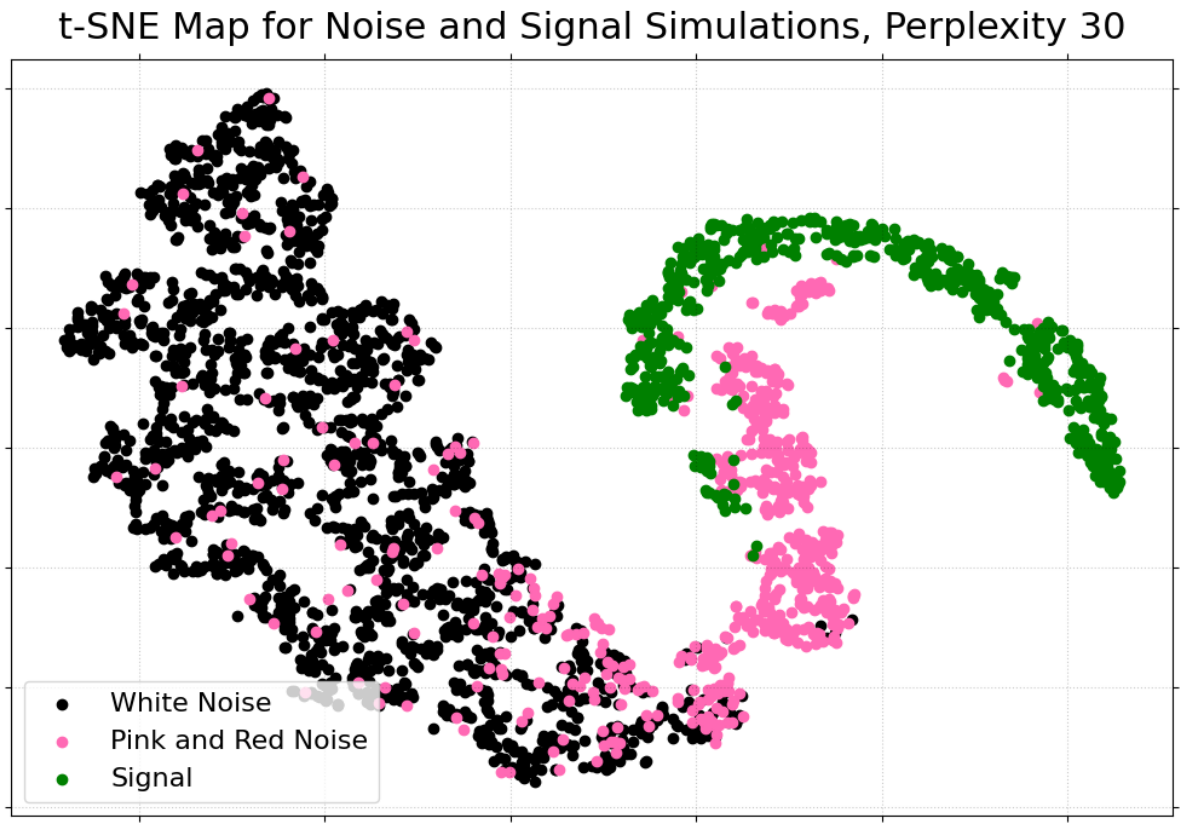
Data: data set $X = \{x_1, x_2, ..., x_n\}$,
cost function parameters: perplexity $Perp$,
optimization parameters: number of iterations T, learning rate η, momentum $\alpha(t)$.
Result: low-dimensional data representation $\mathcal{Y}^{(T)} = \{y_1, y_2, ..., y_n\}$.
begin
compute pairwise affinities $p_{j|i}$ with perplexity $Perp$ (using Equation 1)
set $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$
sample initial solution $\mathcal{Y}^{(0)} = \{y_1, y_2, ..., y_n\}$ from $\mathcal{N}(0, 10^{-4}I)$
for *t=1* **to** T **do**
compute low-dimensional affinities q_{ij} (using Equation 4)
compute gradient $\frac{\delta C}{\delta \mathcal{Y}}$ (using Equation 5)
set $\mathcal{Y}^{(t)} = \mathcal{Y}^{(t-1)} + \eta \frac{\delta C}{\delta \mathcal{Y}} + \alpha(t) \left(\mathcal{Y}^{(t-1)} - \mathcal{Y}^{(t-2)}\right)$
end
end

$$\epsilon(P_i) = 2^{H(P_i)},$$

where P_i represents the conditional probability distribution over all other $\{x_j\}$ given $\{x_i\}$, and $H(P_i)$ is the Shannon entropy of P_i in bits [202].

t-SNE, Simulations





Slightly different input type

t-SNE, Sources