Statistical analysis of long term Gamma-Ray data

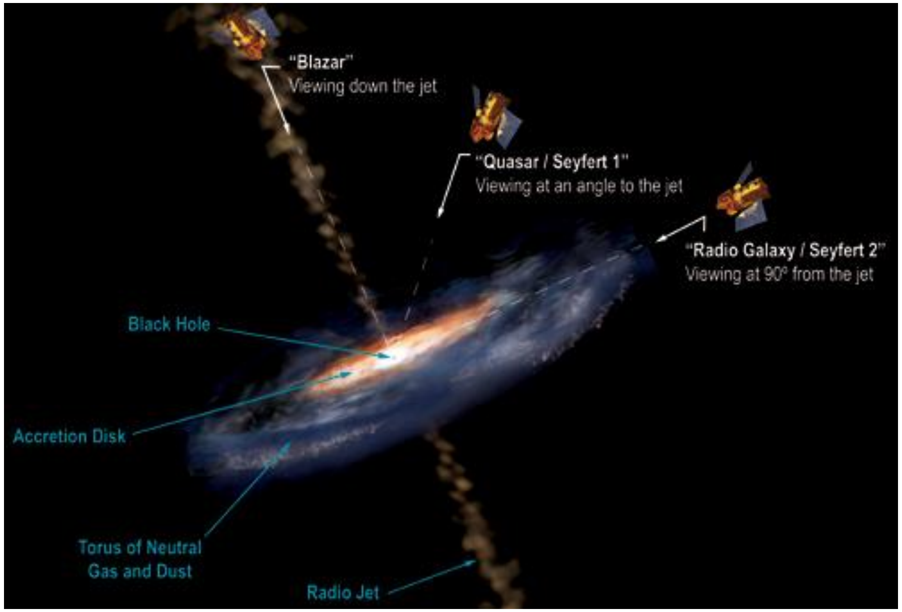
DESY Summer Student Programme 2010

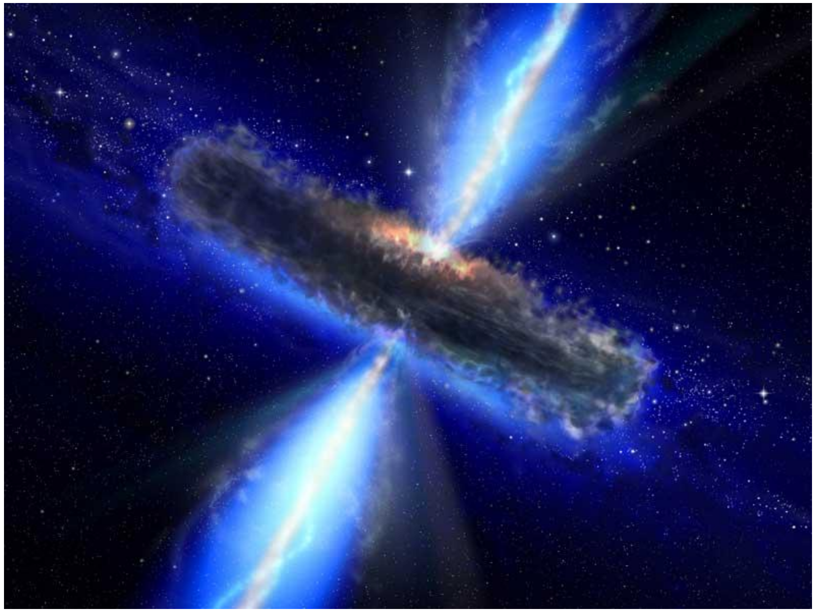
S. Awiphan[1], K. Satalecka[2], E. Bernardini[2]

[1] Department of Physics and Materials Science,
Faculty of Science, Chiang Mai University, Thailand
[2] Multimessenger Young Investigator Group,
DESY Zeuthen, Germany.

Very High Energy Gamma-ray astronomy

- Very High Energy (VHE, E>100GeV) Gamma-Ray astronomy is one of the youngest branches of physics (~30 years old).
- ***Past*** : Discovering the Gamma-Ray flux sources
- ***Now*** : Interest in more detailed and statistical studies is growing
- ***Problems***
  - Light curves of VHE Gamma-rays aren't continuous like in other wavelengths

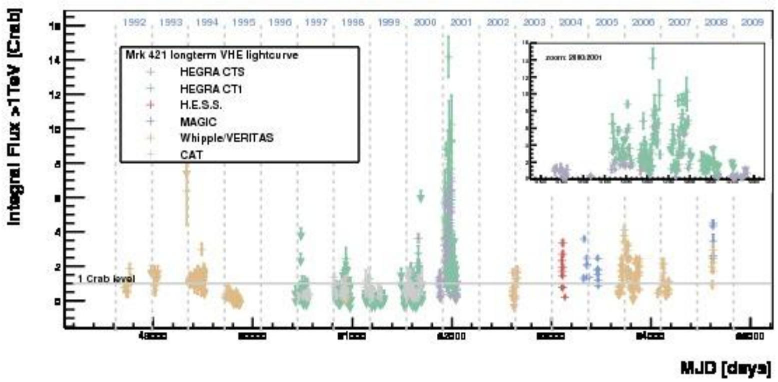




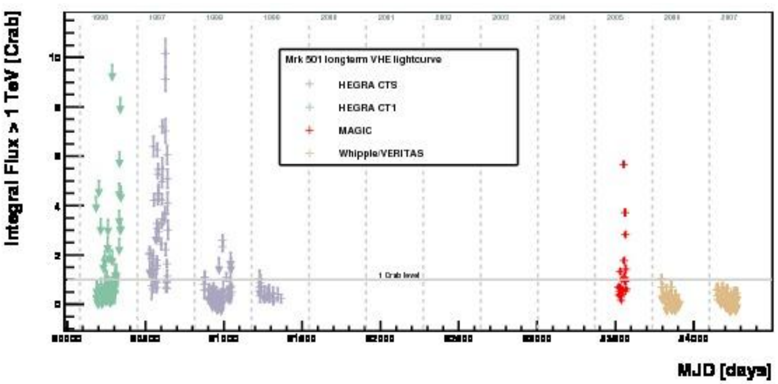
http://www.nasa.gov/

Statistical analysis of long term Gamma-Ray data

Mrk 421



Mrk 501



http://www-zeuthen.desy.de/multi-messenger/GammaRayData/index.html

Long-term collected Gamma-ray data
(Tauczykont M. et al. 2010)

Flux distribution
(Time domain)

Periodogram
(Frequency domain)

The stationarity of light curve
(Comparing PSDs and variance method)

Generated light curve
(Based on statistic results)

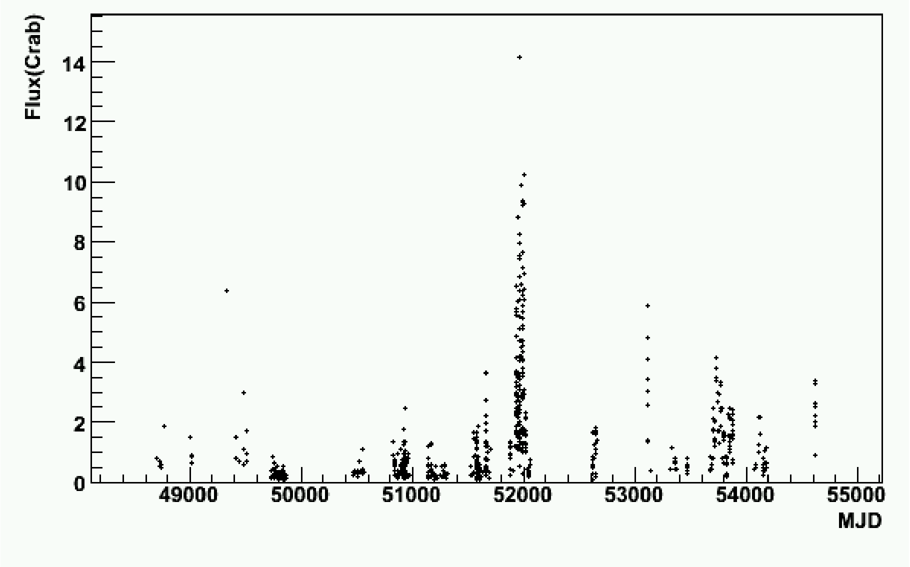
Conclusion

Long term collected Gamma-ray data

Light curve

- Light curve data collected by the multi-messenger group at DESY.
  - Public data from 1992 until today
  - Collected from Whipple, HEGRA, CAT, H.E.S.S., MAGIC, and VERITAS
  - Common threshold of 1 TeV
  - Upper limit excluded
  - (Tluczykont et al. 2010)
- Active Galactic Nuclei (Extragalactic sources)
  - Mrk421
  - Mrk501
  - 1ES1959 +650

- **Light curve of Mrk421**



➢ The Modified Julian Day (MJD) is an abbreviated version of the old Julian Day (JD)

MJD=JD-2400000.5

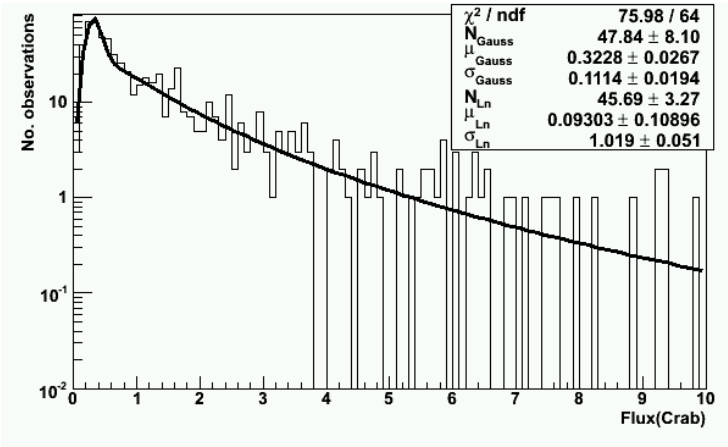
➢ Crab is a unit of the flux observed from the Crab Nebula, the standard candle of Gamma-ray astronomy

Flux distribution

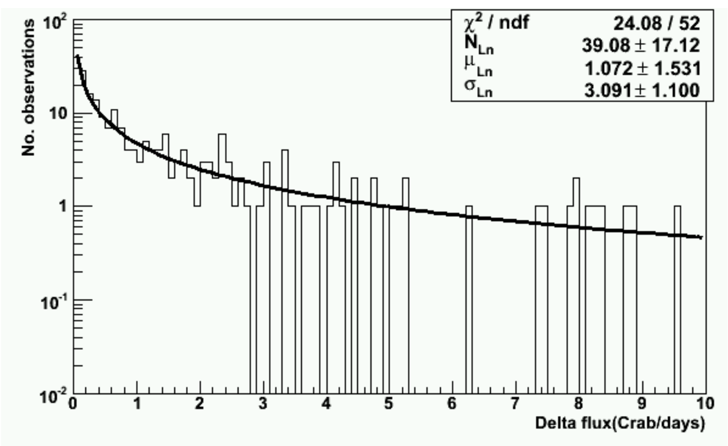
Flux state distribution

- Flux values integrated above 1 TeV in one day bins
- Fit Gaussian and Log-normal distribution
- Integral of distribution is the probability to measure a flux above a certain level

- **Flux state distribution of Mrk421**



- **Delta flux distribution of Mrk421**



Delta flux distribution

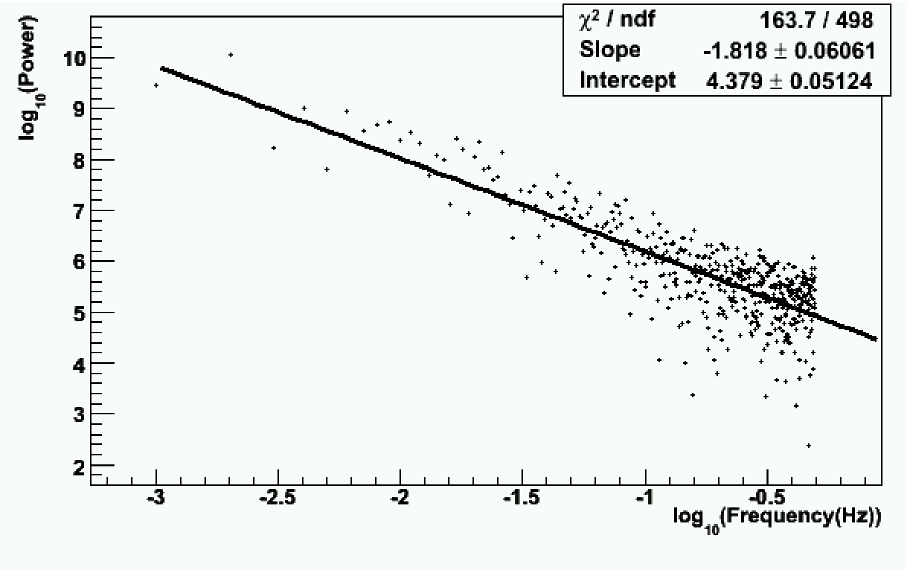
- Time derivative of the observed flux.
- Log-normal distribution

Periodogram

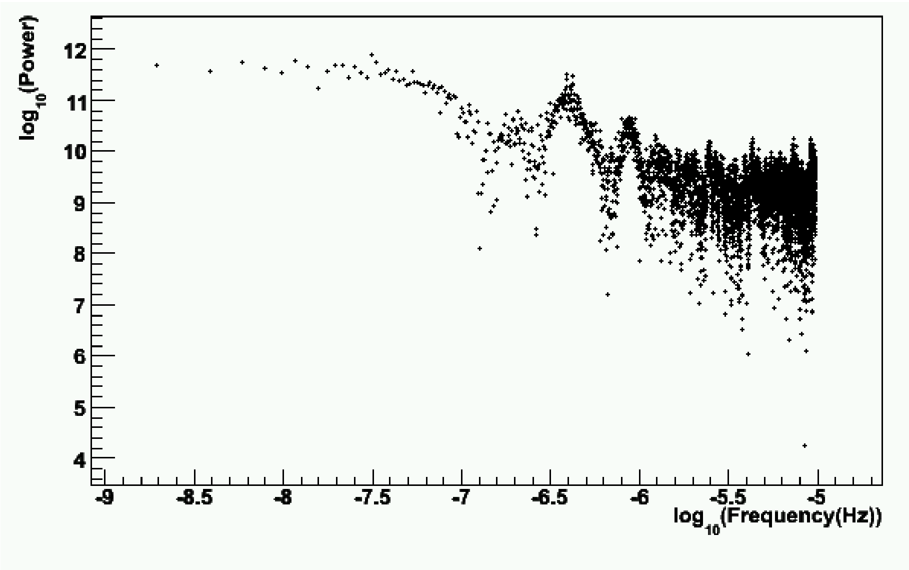
Periodogram

- The modulus-squared of the Discrete Fourier Transform (DFT) of the data
- Use to estimate Power Spectral Density (PSD)
- Show some strange peaks at frequency range 10^{-7} - 10^{-6} Hz
  - Periodic nature of the Gamma-ray emission?
  - Bias of the observation time?

- **Periodogram**



- **Periodogram of Mrk421**



The stationarity of light curve

The stationarity of light curve

- The analysis of periodogram is meaningful only when the underlying processes are statistically stationary
  - Comparing PSDs
  - Comparing variance

Comparing PSDs

- Comparing the PSDs from different periodograms
  - Divide the light curve into 2 parts and calculate the periodogram of both parts.
  - Compare power from periodogram of 2 periodogram
  - Form the test statistic S (normally distributed with zero mean and unit variance)
- The light curves are strongly non-stationary (S=25-30)
- The periodogram of part which contains a large flare show strange peaks

The stationarity of light curve

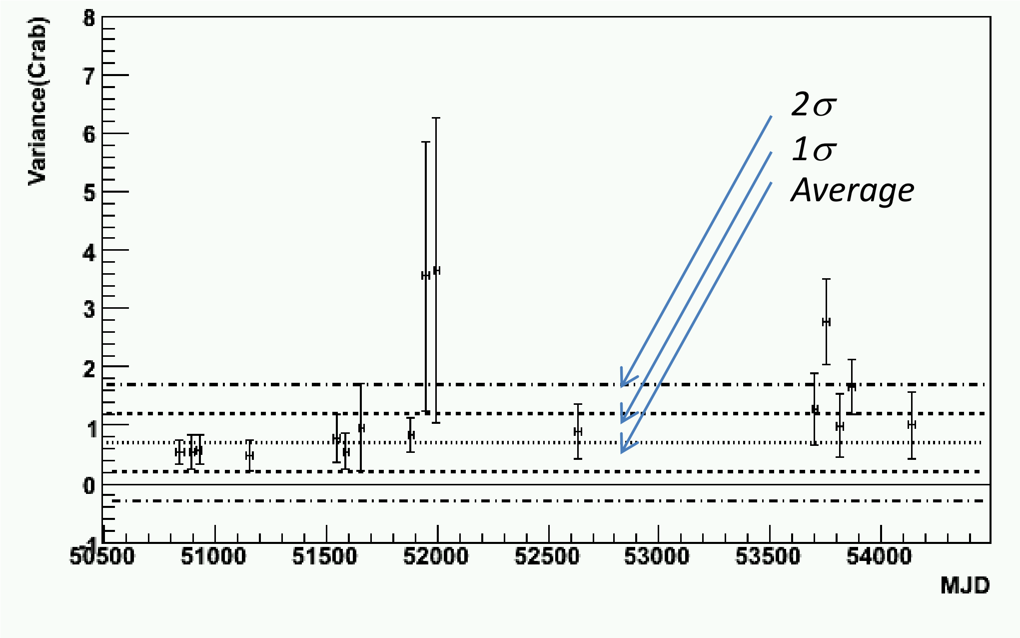
Comparing variances

- Use simple variance

$$S^2 = \frac{1}{N-1}\sum_{i-1}^{N}(x_i - \bar{x})^2$$

- 50 days intervals with at least 10 flux measurements.
- Cannot conclude that the light curves are stationary or not.

- **Variance of the flux as a function of time from Mrk421**



➢ dotted line	average variance
➢ dashed line	1σ confidence interval
➢ dash-dotted line	2σ confidence interval

Simulated light curve

Simulate from flux state distribution

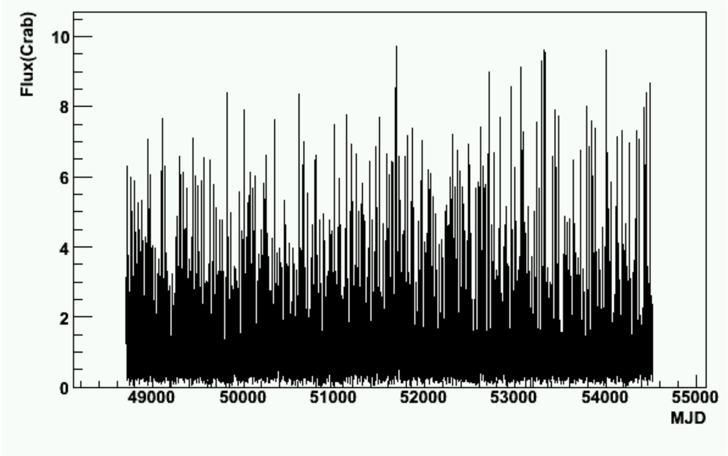
Random method

- Randomly generate flux from flux state distribution
- ***Problem :*** light curve looks too “spiky”.

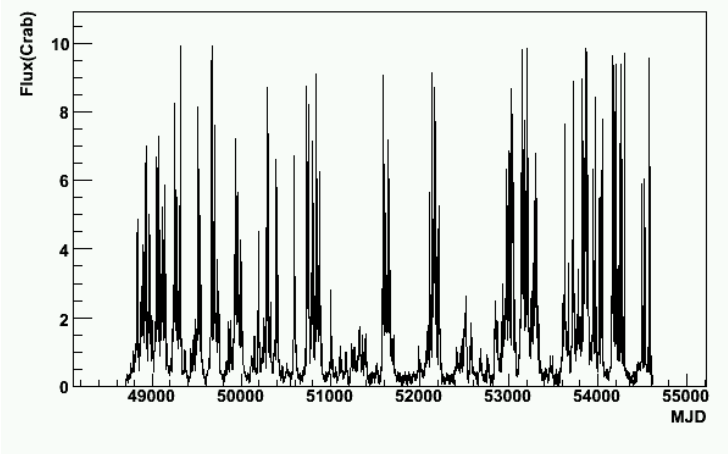
Rearrange method

- Flux value generated in the n-th step is taken from the values with indexes in the range of n±10%n
- ***Problem :*** Don’t know how to realistically model the time evolution

- **Simulated light curve(random method)**



- **Simulated light curve(rearrange method)**



Simulated light curve

Simulate from delta flux distribution

- Randomly choose the change in the flux from the delta flux distribution
- ***Problem :*** Very long and very high flares

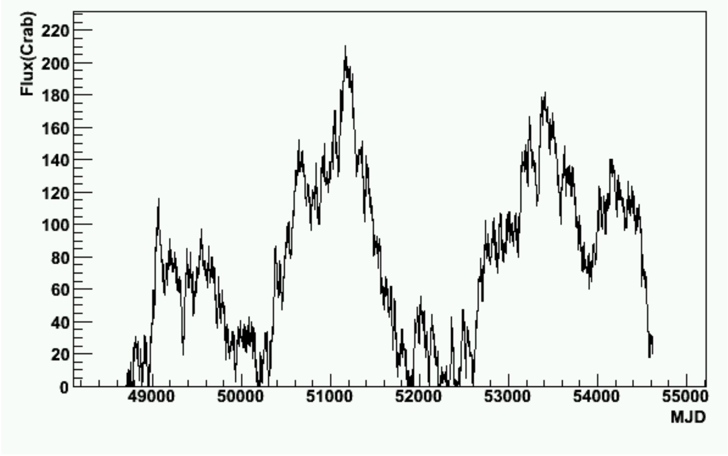
Simulate from periodogram

- Proposed by Timmer & König (1995)
- ***Problem :*** Don't know periodogram slope

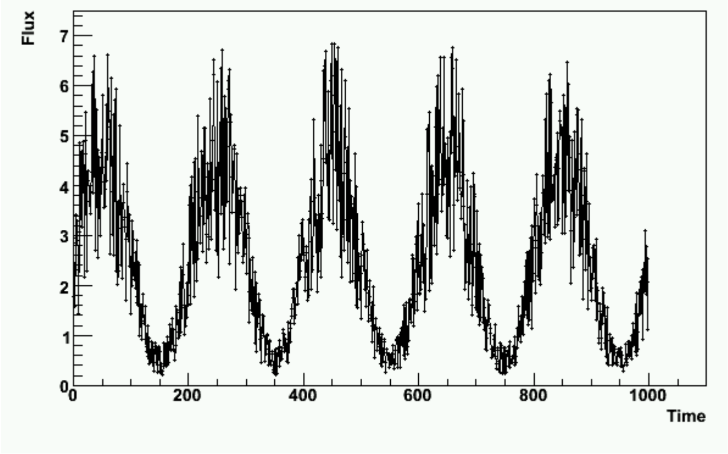
Simulate from periodic function

- Sine wave to describe the mean flux level
- Fluctuation proportional to that mean flux
- ***Problem :*** Very short low state period

- **Simulated light curve(periodogram)**



- **Simulated light curve(periodic function)**



Simulated light curve

Simulate light curve with separate high state and low state

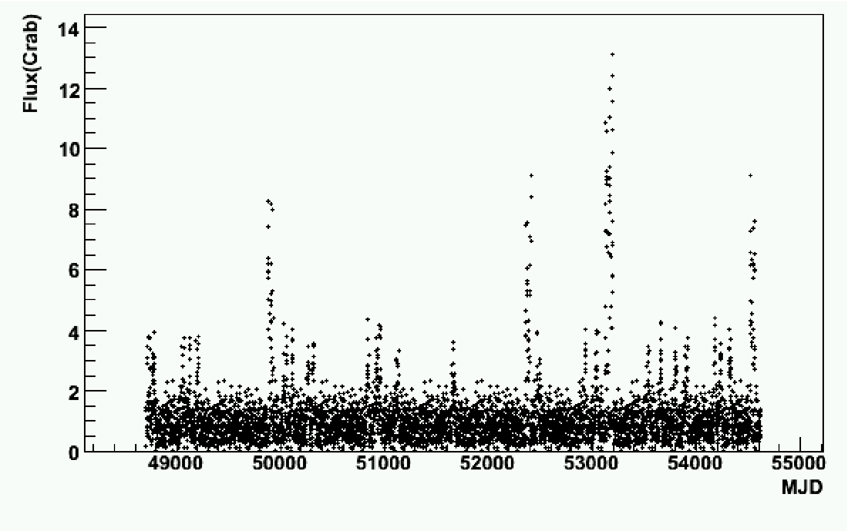
Low state

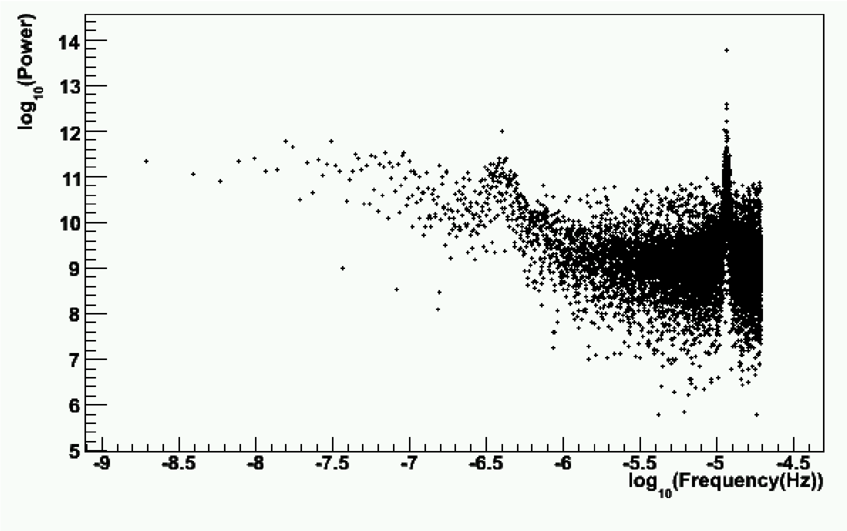
- Generating from periodogram (Power law)
- Power from slope of low state periodogram from real data

High state

- Generating from periodic function
- Period from fit to real data periodogram

- **Simulated light curve with separate high state and low state and corresponding periodogram**





Conclusion

- Work on statistical study of long term VHE Gamma-ray light curves of three sources: Mrk421, Mrk501, and 1ES1959+650
- The periodograms show strange peaks at frequency range 10^{-7} - 10^{-6} Hz.
- The stationarity of the light curve : non-conclusive
- Many different algorithms to generate light curves were developed using the results from our statistical studies and tested
- The light curves simulated with separate high state and low state can very well reproduce the behavior of the real light curve
- The sine function does not exactly describe the structure of the flare. Exponential function will be use to describe the structure of flare in future work

References

[1] Bevington P.R., & Robinson D.K., 1992, Data Reduction and Error analysis for the Physical Sciences, McGraw-Hill (New York)
[2] Bloomfield P., 2000, Fourier Analysis of Time Series, Wiley (New York)
[3] Gaskell C.M., 2004, AJ, 612, 24
[4] Giebels B., & Degrange B., 2009, A&A, 503, 797
[5] Jenkins G.M., 1961, Technometries, 3, 133
[6] Papadakis I.E., & Lawrence A., 1993, MNRAS, 261, 612
[7] Papadakis I.E., & Lawrence A., 1995, MNRAS, 272, 161
[8] Priestley M.B., 1981, Spectral Analysis and Time Series, Academic Press (London)
[9] Punch M. et at., 1992, Nature, 358, 477
[10] Rödig C. et al., 2009, A&A , 501, 952
[11] Timmer J., & König M., 1995, A&A, 300, 700
[12] Tluczykont M. et al., 2010, A&A (Accepted)
[13] van der Kils M., 1989, ARA&A, 27, 517
[14] Vaughan S. et al., 2003, MNRAS, 345, 1271

Thank you for your kind attention.

Statistical analysis of long term Gamma-Ray data (Backup slide)

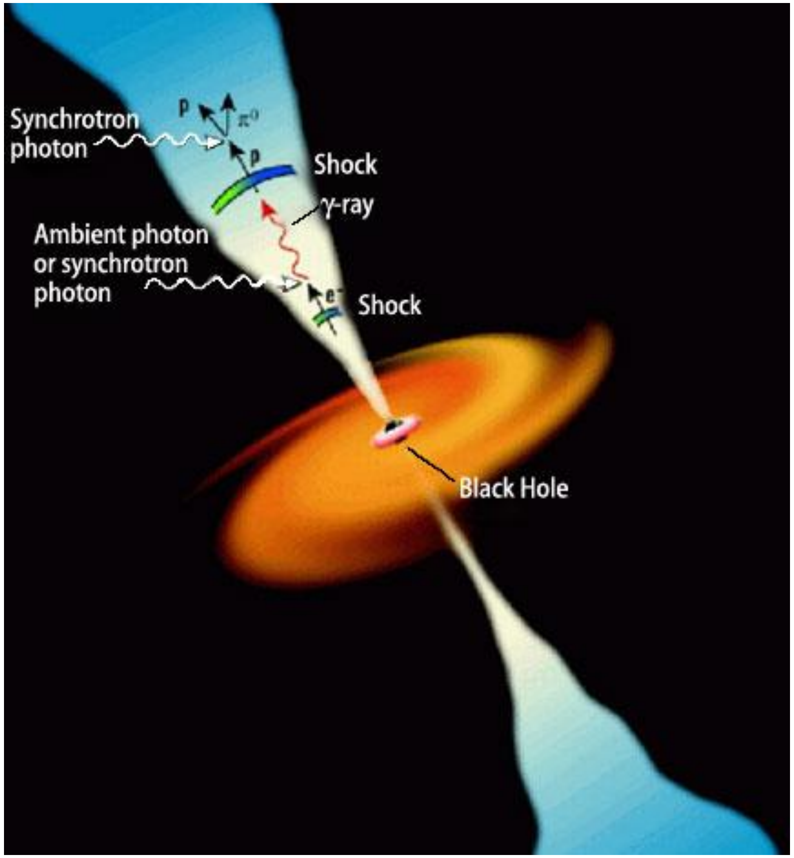
DESY Summer Student Programme 2010

S. Awiphan[1], K. Satalecka[2], E. Bernardini[2]

[1] Department of Physics and Materials Science,
Faculty of Science, Chiang Mai University, Thailand
[2] Multimessenger Young Investigator Group,
DESY Zeuthen, Germany.

Blazar

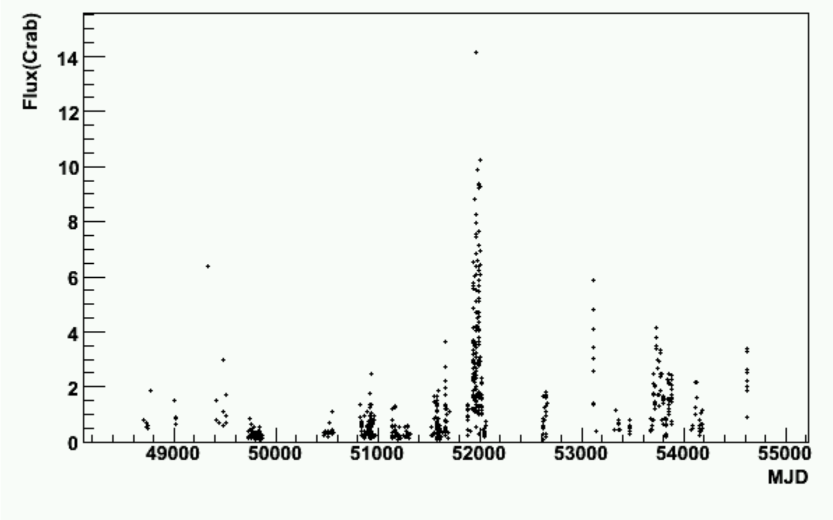
- Relativistic outflows (jets) powered by mass accretion onto a supermassive black hole (10^6 - 10^{10} M_{sun}) in the galactic center.
- The jets are directed at a small angle with respect to the line of sight of the observer.
- A continuous Spectral Energy Distribution (SED) with two peaks
    - *Radio to UV* : synchrotron emission from relativistic electrons in the jet
    - *X-ray to Gamma-ray* : the Synchrotron-Self Compton (SSC) model. Compton scattering of lower energy radiation by the same relativistic electrons which are responsible for the synchrotron emission at lower frequencies

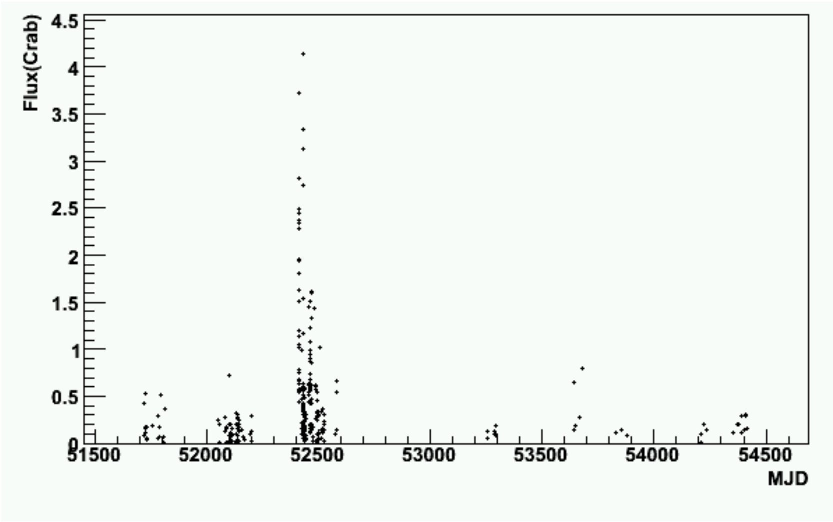


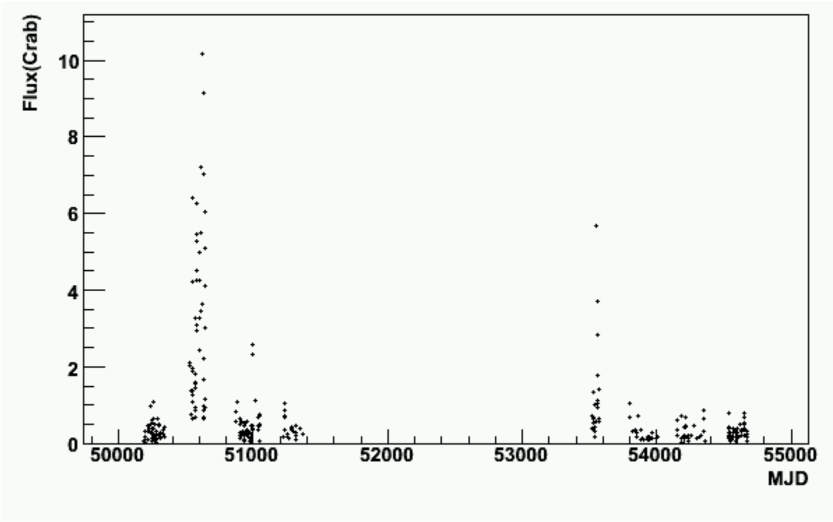
http://www.lanl.gov/milagro/

Light curve

- Mrk421 *(Right top)*
- Mrk501 *(Left bottom)*
- 1ES1959 +650 *(Right bottom)*



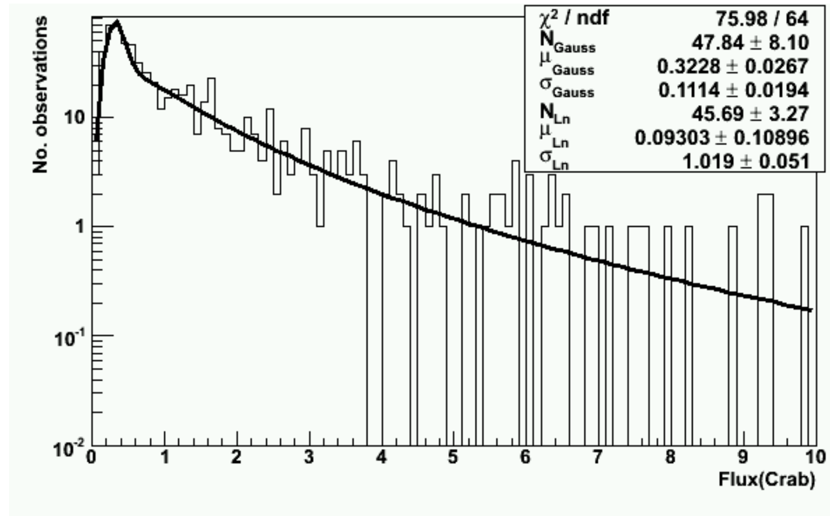


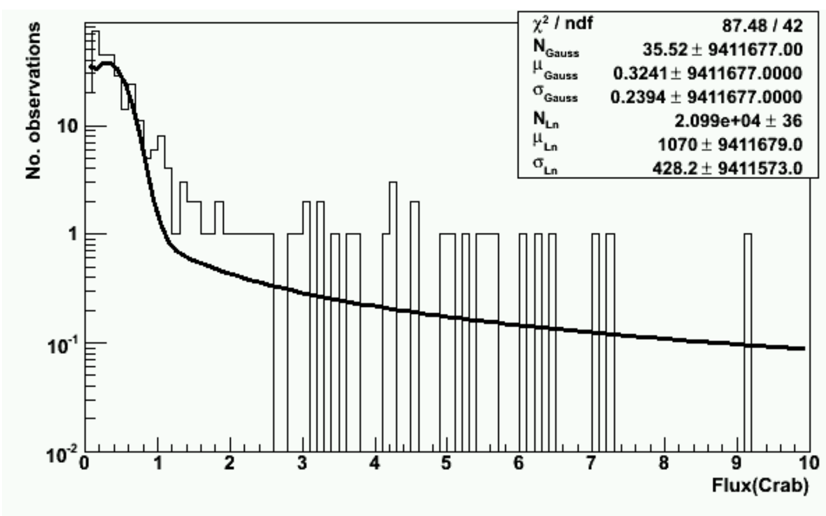


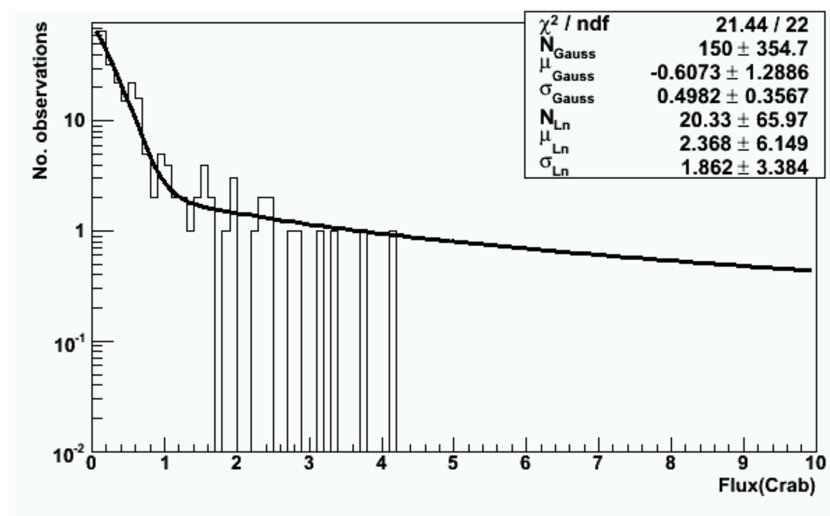
Flux state distribution

Flux state distribution

- ***Low state*** : Gaussian distribution.
    - <0.5 Crab
    - The mean of Gaussian distribution may represent the baseline flux.
- ***High state*** : Log-normal distribution
    - Probably related to accretion disk activity.





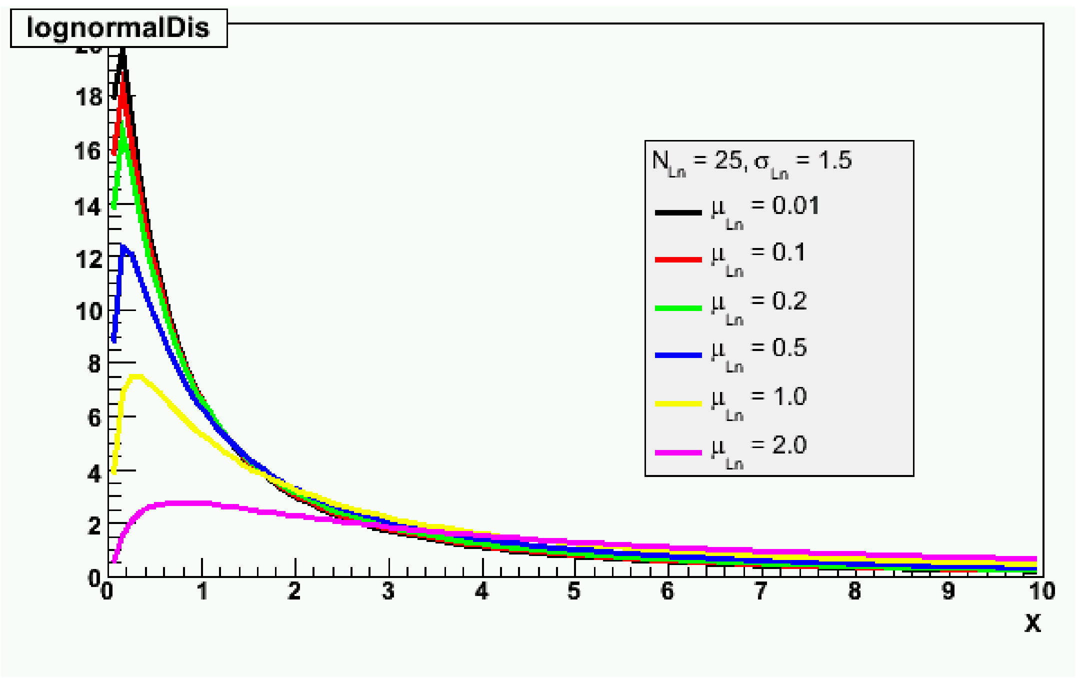


Log-normal distribution

Log-normal distribution

- A random variable whose logarithm is normally distributed

$$f(x:\sigma_{Ln},\mu_{Ln}) = \frac{1}{x\sigma_{Ln}\sqrt{2\pi}}e^{\left(-\frac{1}{2\sigma_{Ln}^2}(\ln(x)-\mu_{Ln})^2\right)}$$



- For x>0
- σ_{Ln} is the mean
- μ_{Ln} is the standard deviation

Probability of high state

Probability of high state

- The probability that the source in a flux state higher than a certain threshold.

$$P_{F>F_{th}} = 1 - \int_0^{F_{th}} f_{Total}(x)dx$$

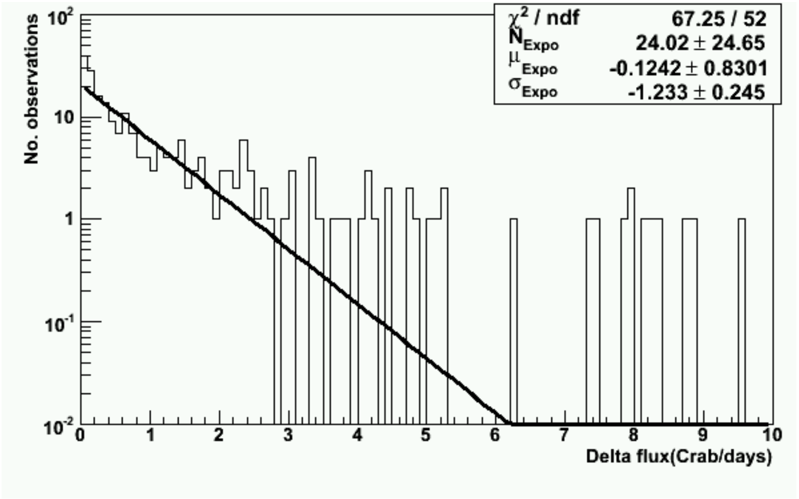
- where $f_{Total}(x)$ is the function fitted to the flux state distribution
- The source as being in high state when the flux level exceed the μ_{Gauss} by σ_{Gauss}

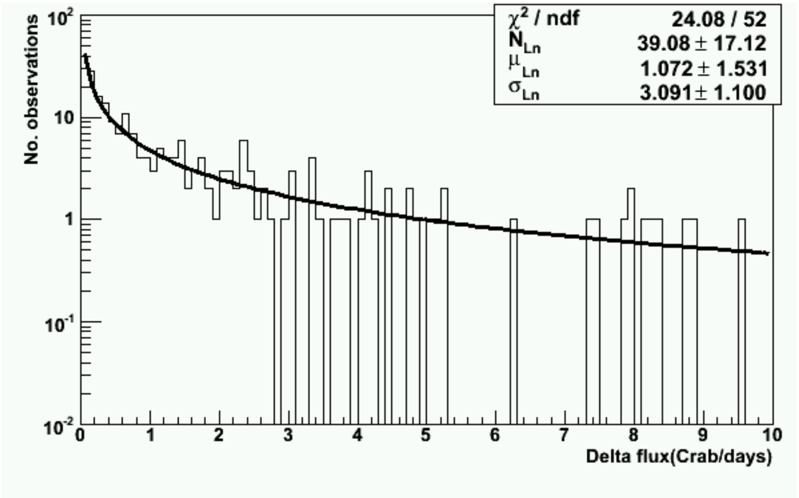
	Mrk421		Mrk501		1ES1959 +650	
	F_{th}(Crabs)	*P(%)*	*F_{th}(Crabs)*	*P(%)*	*F_{th}(Crabs)*	*P(%)*
$5\sigma_{Gauss}$	0.9	45	1.1	16	1.3	9
$10\sigma_{Gauss}$	1.5	32	1.9	11	2.7	2
$20\sigma_{Gauss}$	2.6	16	3.5	6	5.4	-

Delta flux distribution

Delta flux distribution

- “delta flux”, the time derivative of the observed flux.
- Use only 2 nearby flux measurements which have the start time of the observation not more than 2 days apart to avoid the long time gaps present in the data.
- ***Exponential distribution***
    - Reflects a stochastic process of the object
- ***Log-normal distribution***
    - Better fit





Flux distribution

Relation between delta flux and flux

- The delta flux is proportional to flux

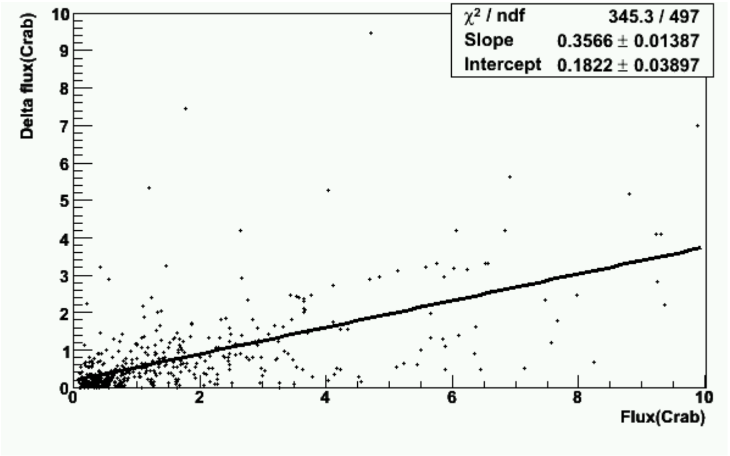
Relation between excess variance and average flux

- Use "excess variance" to estimate the intrinsic source variance

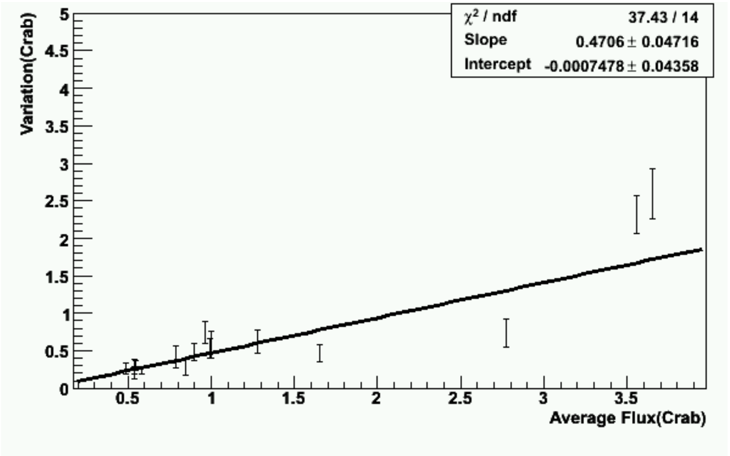
$$\sigma_{xs} = \sqrt{\frac{1}{N}\sum_{i-1}^{N}((x_i - \bar{x})^2 - \sigma_i^2)}$$

- The variability of the flux is directly proportional to the mean flux level
- Confirms the Log-normal variability of the source

- **Delta flux VS flux of Mrk421**



- **Excess variance VS average flux of Mrk421**

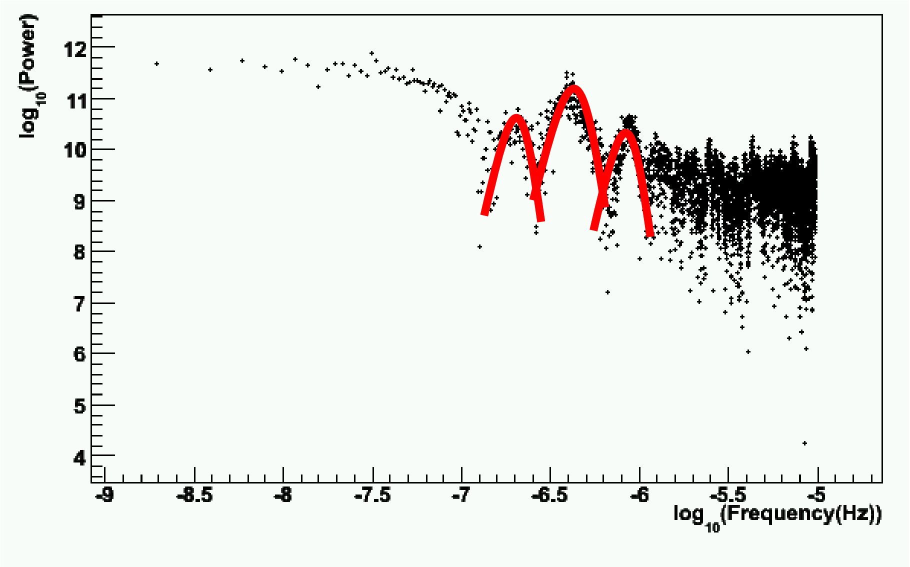


Periodogram

Periodogram

- The modulus-squared of the Discrete Fourier Transform (DFT) of the data
- Use to estimate Power Spectral Density (PSD)
- Show some strange peaks at frequency range 10^{-7} - 10^{-6} Hz
  - Periodic nature of the Gamma-ray emission?
  - Bias of the observation time?

- **Periodogram of Mrk421**

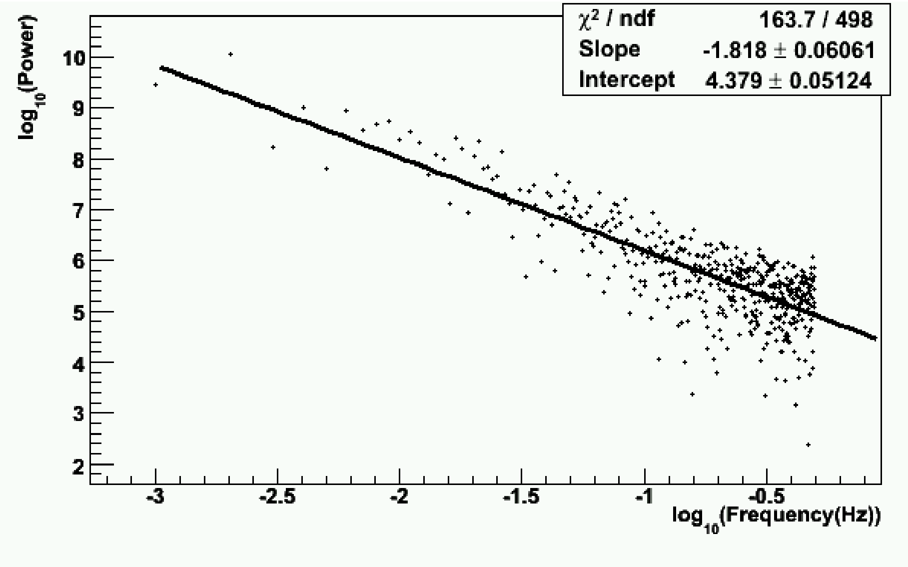


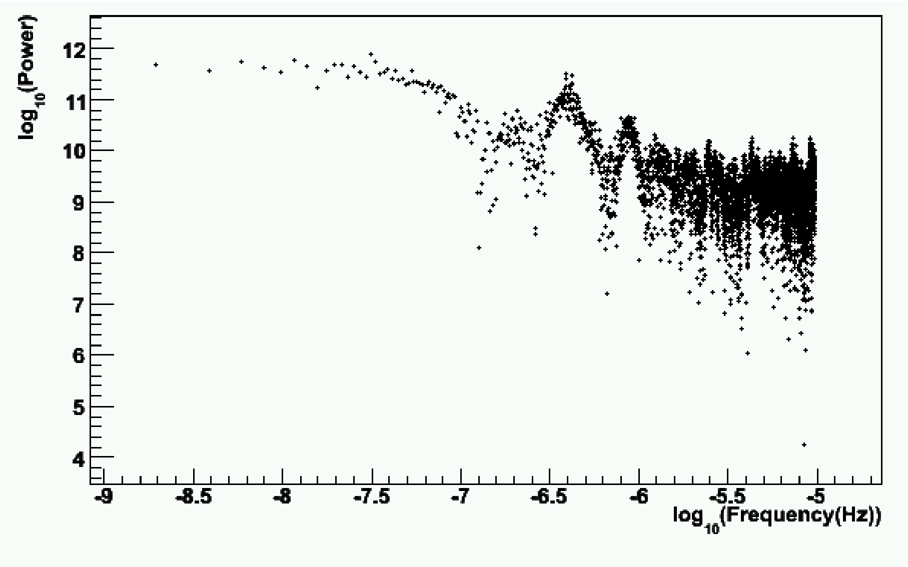
- 1st period 58.9±0.8 days
- 2nd period 29.0±0.1 days
- 3rd period 13.13±0.01 days

Periodogram

- Tools for examining AGN variability
- Represents the amount of variability power as a function of frequency
- Periodogram is the modulus-squared of the Discrete Fourier Transform (DFT) of the data (Press et al. 1996).

$$I\left(v_{p_j}\right) = A\left|DFT\left(v_p\right)\right| = A\{\sum_{i=1}^{N} x_i \cos(2\pi f_j t_i)\}^2 + A\{\sum_{i=1}^{N} x_i \sin(2\pi f_j t_i)\}^2$$

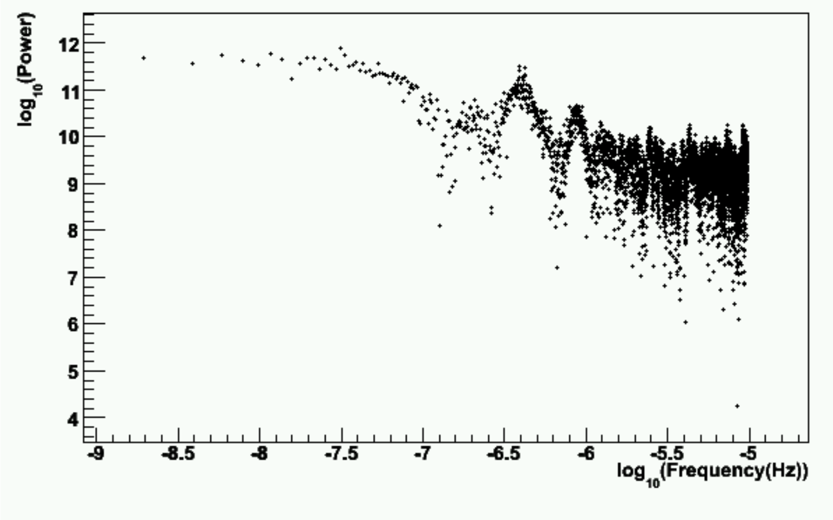


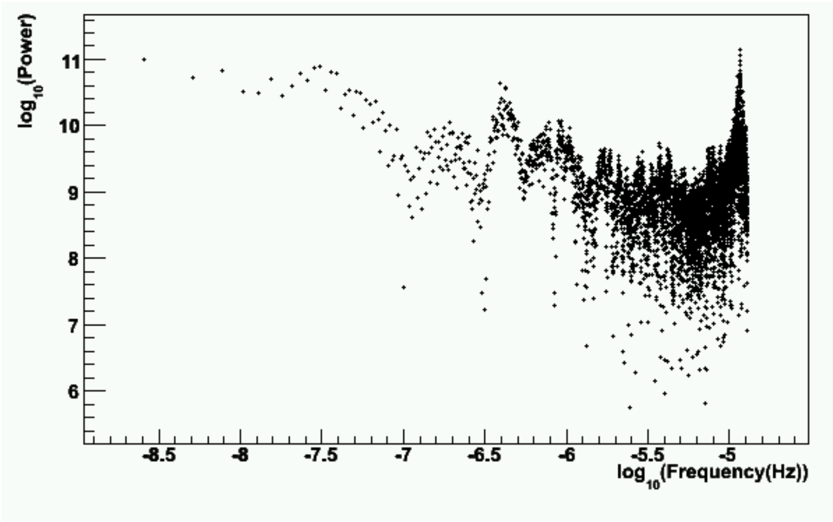


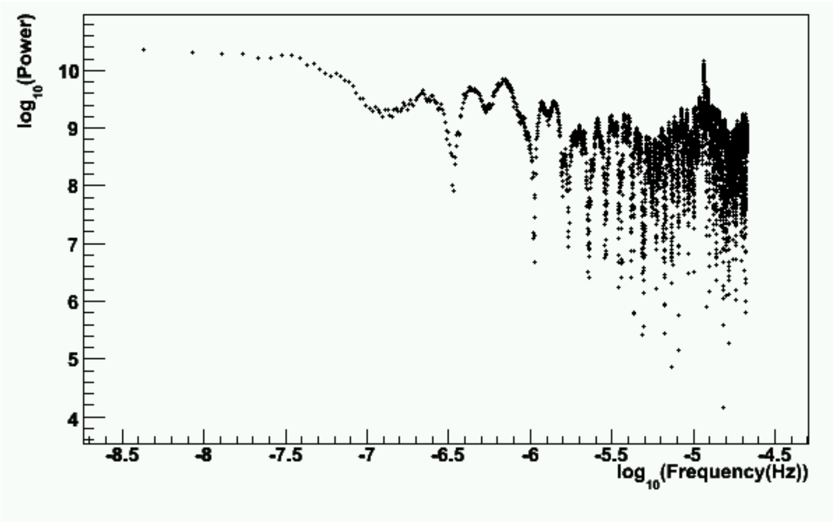
Periodogram

Periodogram

- Mrk421 *(Right top)*
- Mrk501 *(Left bottom)*
- 1ES1959 +650 *(Right bottom)*







Periodogram

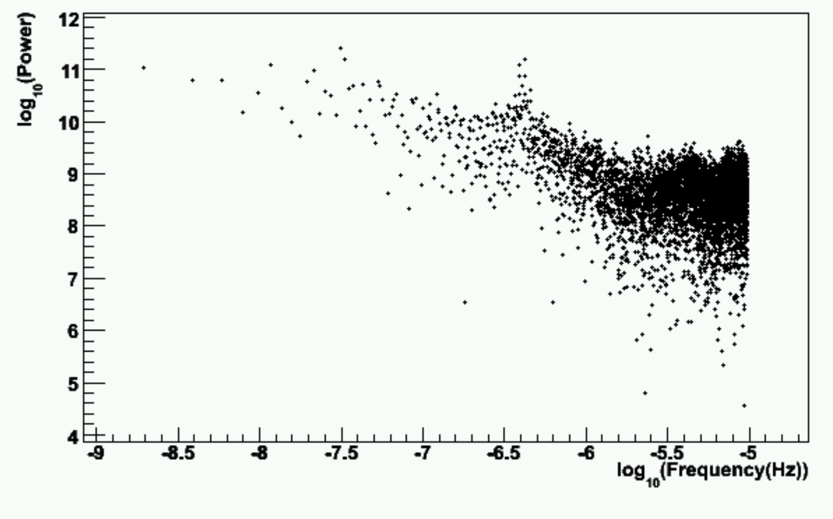
- Calculate the periods by fitting the periodograms with polynomial of 2nd degree
- Use the 1st derivative to find out the position peaks
- Range of fit : compare the χ^2 per number of degree of freedom (NDF)
    - 1st period 10^{-7} - $10^{-6.4}$ Hz
    - 2nd period $10^{-6.6}$- $10^{-6.2}$ Hz
    - 3rd period $10^{-6.3}$ – $10^{-5.9}$ Hz
- Use the range that yielded the smallest value of χ^2 per NDF

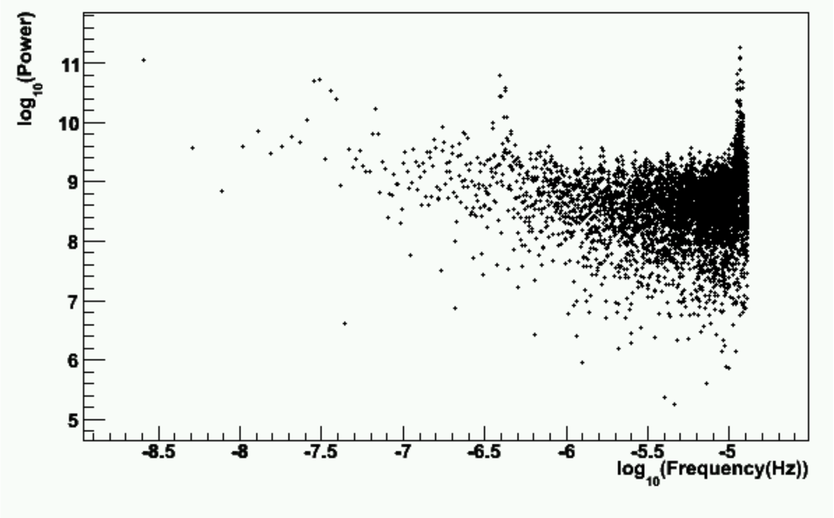
Period (Days)	Mrk421	Mrk501	1ES1959 +650
1st period	58.9±0.8	64.6±0.7	58.8±0.4
2nd period	29.0±0.1	27.46±0.05	25.58±0.02
3rd period	13.13±0.01	16.93±0.04	13.13±0.03

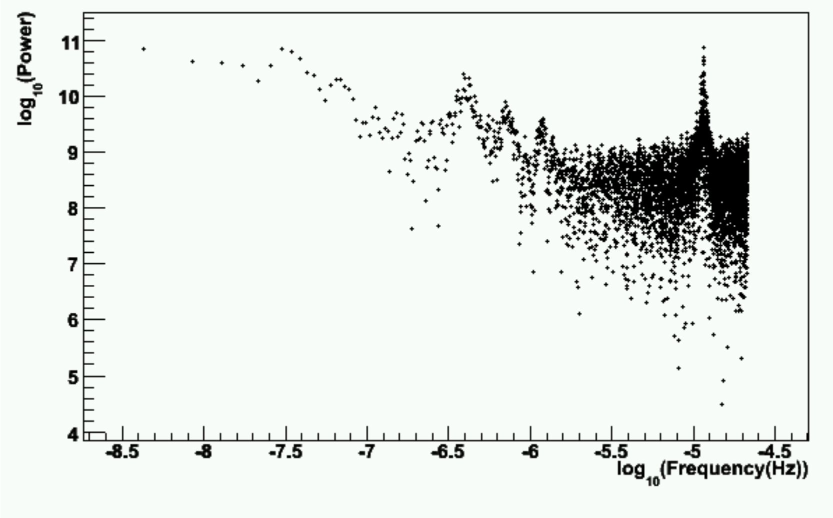
Periodogram

Periodogram of MC generated

- The periodogram of Monte Carlo generated light curve which we sampled at the same observation times as in the original data
- Have peaks in periodogram, but the structures of the peaks are not exactly the same as in the real data.





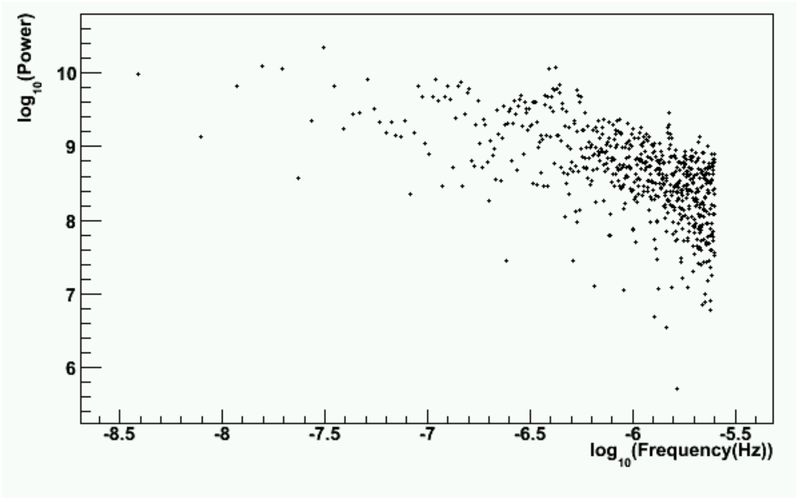


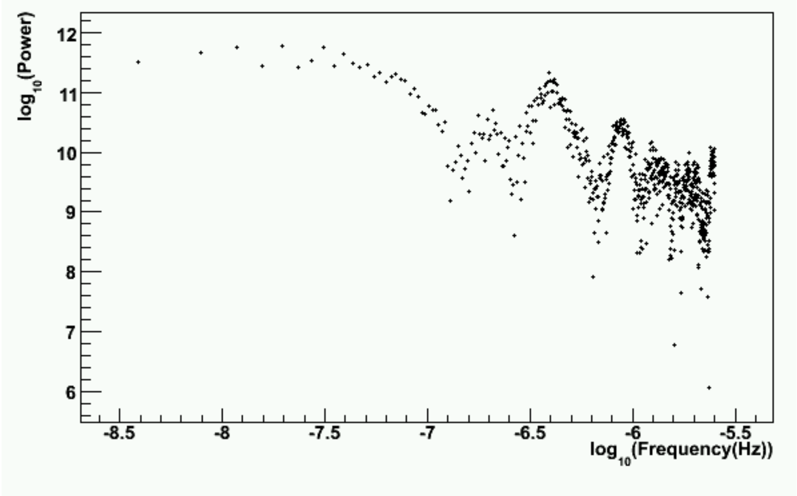
Periodogram

Periodogram of 2 parts

- Only one of the periodograms from each pair shows strange structures (peaks)
  - The part which contains a large flare. From this result

- The large flare generated strange structures on periodograms of all sources and made their light curve look non-stationary





The stationarity of light curve

Comparing PSDs

- Method proposed by Papadakis & Lawrence (1995) based on an original idea by Jenkins (1961)
- Comparing the PSDs from different periodograms. If the PSDs show significant difference, the underlying process can be said to be strongly non-stationary
    - Divide the time series into two parts and calculate the periodogram

$$S(\nu_p) = \frac{\log[I_{PartI}(\nu_p)] - \log[I_{PartII}(\nu_p)]}{\sqrt{var\{\log[I_{PartI}(\nu_p)]\} + var\{\log[I_{PartII}(\nu_p)]\}}}$$

    - Where $var\{\log[I(\nu_p)]\} \approx 0.310$ (Papadakis & Lawrence, 1993)
    - The mean and variance of the random variable $S(v_p)$ have to be 0 and 1.
    - Form the test statistic S

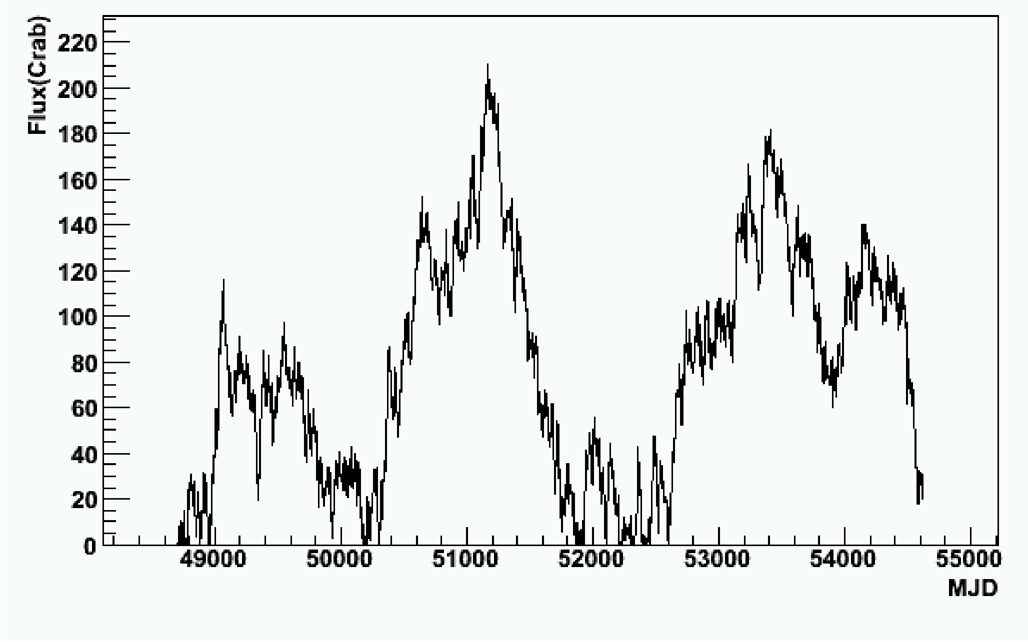
$$S = \frac{1}{\sqrt{p_{max}}} \sum_{p=0}^{p_{max}} S(\nu_p)$$

    - Where p_{max} is total number of frequencies

Simulate from delta flux distribution

Simulate from delta flux distribution

- Introduced a time evolution constraint
- Rearrange the generated fluxes in ascending order
- The flux value generated in the n-th step is then taken from the values with indexes in the range of n±10\%n.



Simulate from periodogram

Simulate from periodogram

- Proposed by Timmer & König (1995).
- This algorithm is based on a main result of the theory of spectral estimation.

$$DFT(v_p) = \mathrm{N}\left(0, \frac{1}{2}S(v_p)\right) + i\mathrm{N}\left(0, \frac{1}{2}S(v_p)\right)$$

- Choose a power law from which we want to generate the light curve.

$$S(v_p) \sim \left(\frac{1}{2\pi v_p}\right)^{\alpha}$$

- For each Fourier frequency v_p, generate two Gaussian distributed random numbers and multiply them by

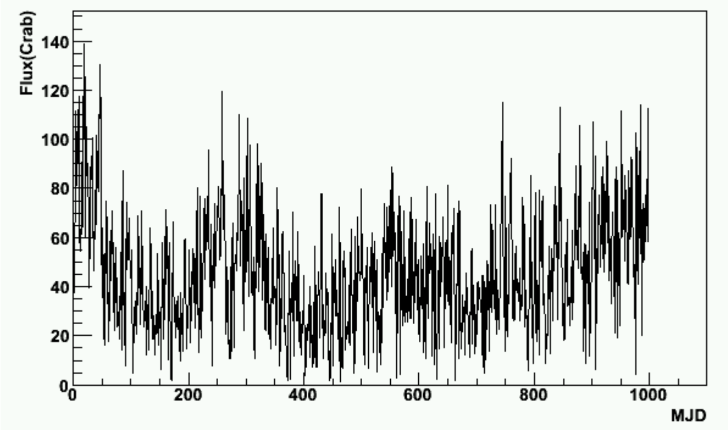
$$\sqrt{\frac{1}{2}S(v_p)} \sim \left(\frac{1}{2\pi v_p}\right)^{\frac{\alpha}{2}}$$

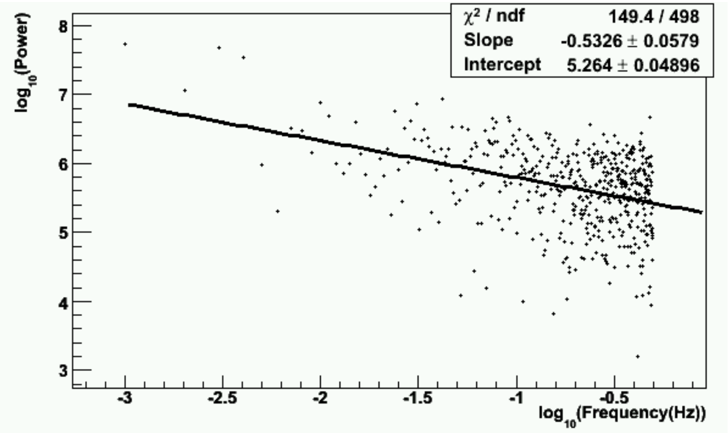
- Use the result as the real and imaginary part of the Fourier transform
- Obtain the time series by using Inverse Discrete Fourier Transform (IDFT)

Simulate from periodogram

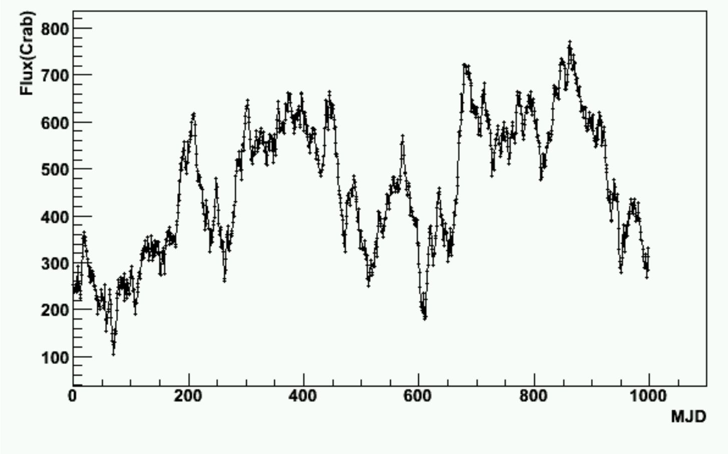
Simulate from periodogram

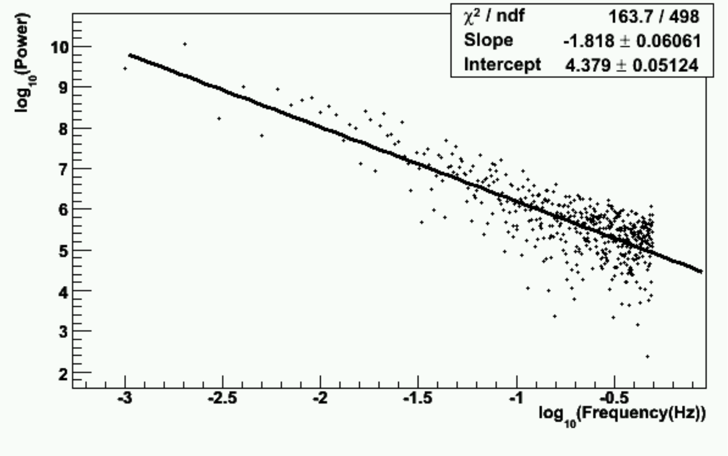
- $\alpha = 1$





- $\alpha = 2$

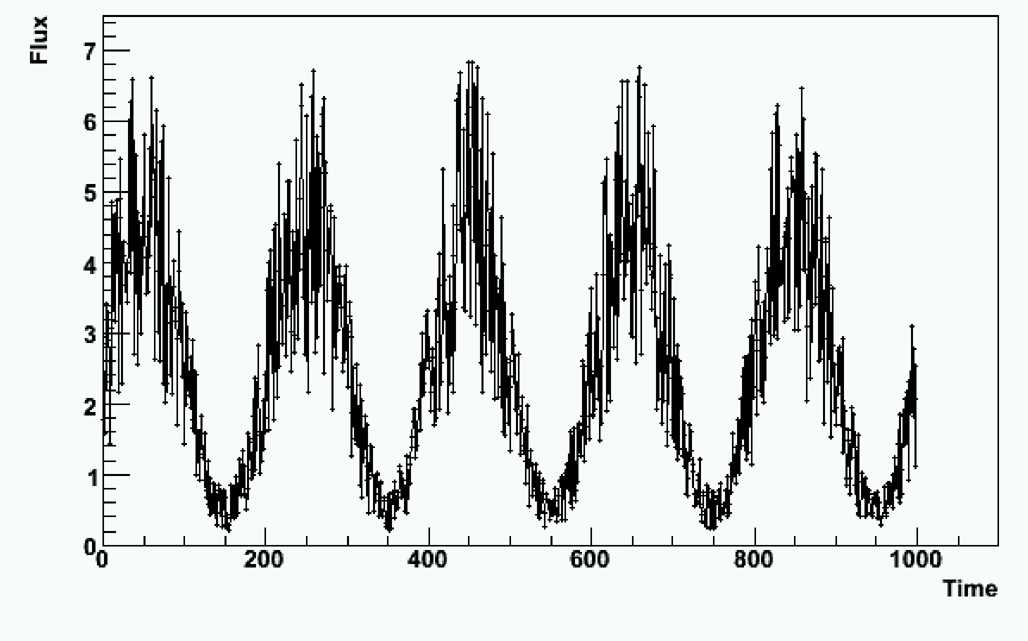




Simulate from periodic function

- Use sine wave to describe the mean flux level and fluctuation proportional to that mean flux level to generate light curve.
- apply the flux fluctuations on top of sine wave

$$\Delta x = \pm(S_d x + I_d) \pm Random\{0, (S_v x + I_v)\}$$



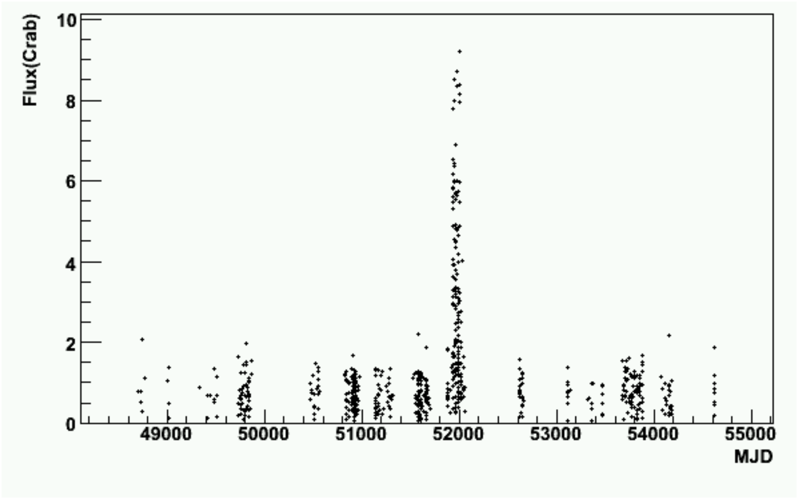
Simulated light curve (separate)

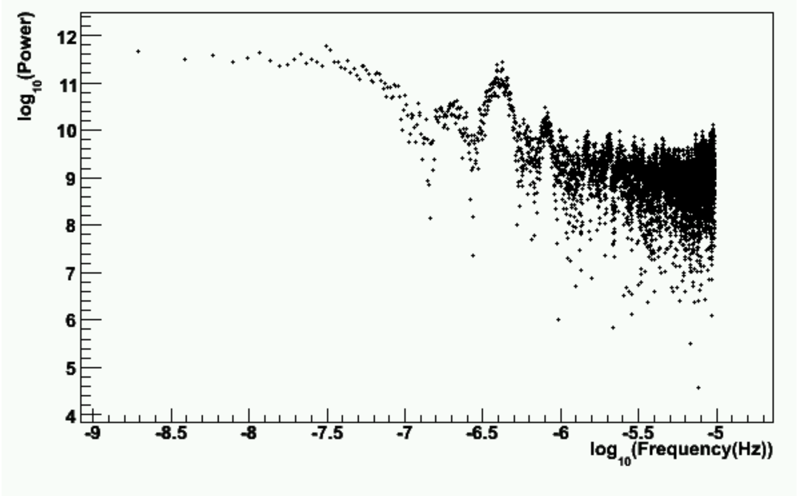
Simulate light curve with separate high state and low state

- Generated a big flare from a periodic function and a low state from periodogram

$$x_i = (\overline{x_F} - \overline{x_L})\sin\left(\frac{2\pi t_i}{T} - \frac{\pi}{2}\right) + \overline{x_F}$$

- Periods of sine wave as obtained from real data
- The shape of the resulting periodograms is very similar to the real ones





Simulated light curve (separate)

Simulate light curve with separate high state and low state

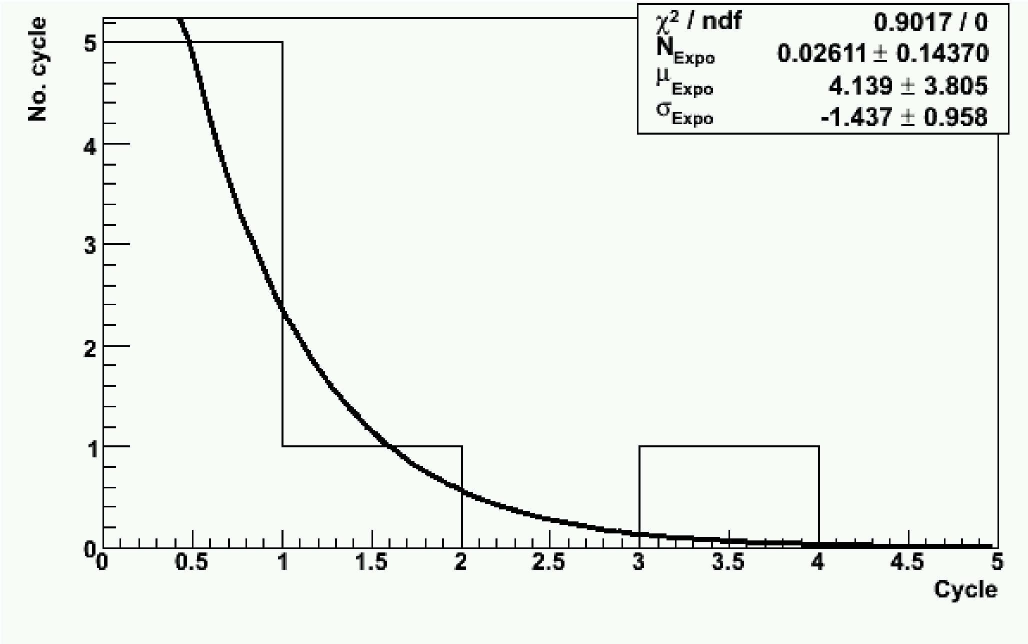
- Simulated 1,000 light curves from this algorithm
- Found the average S-Value between them and the periodogram of original data
- Use second period to generate big flare gives the best average S-Values for all sources.

S-Value	Mrk421	Mrk501	1ES1959 +650
1st period	6.0	1.5	14
2nd period	0.8	1.2	11
3rd period	1.7	1.2	20

Simulated light curve (separate)

Simulate light curve with separate high state and low state

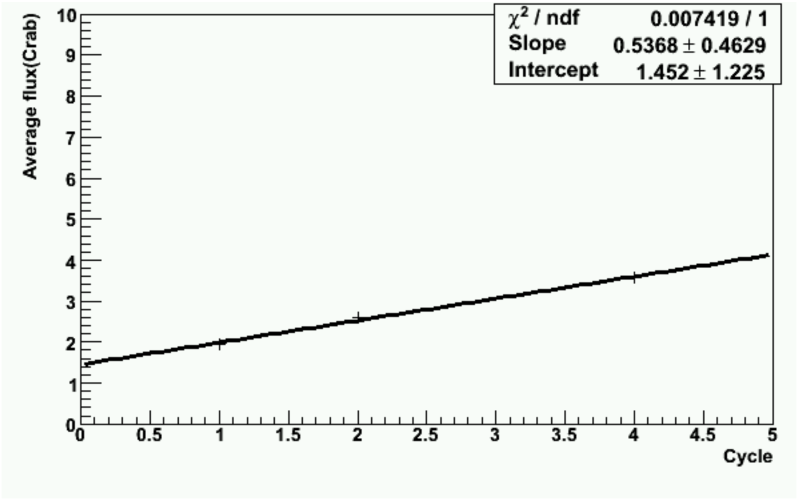
- Ddefine as high state flux which exceeds μ_{Gauss} by $20\sigma_{Gauss}$
- Take the second period from fit to define the standard duration, called "cycle", of a flare
- Calculate how many cycles are observed during the high state period in real data.

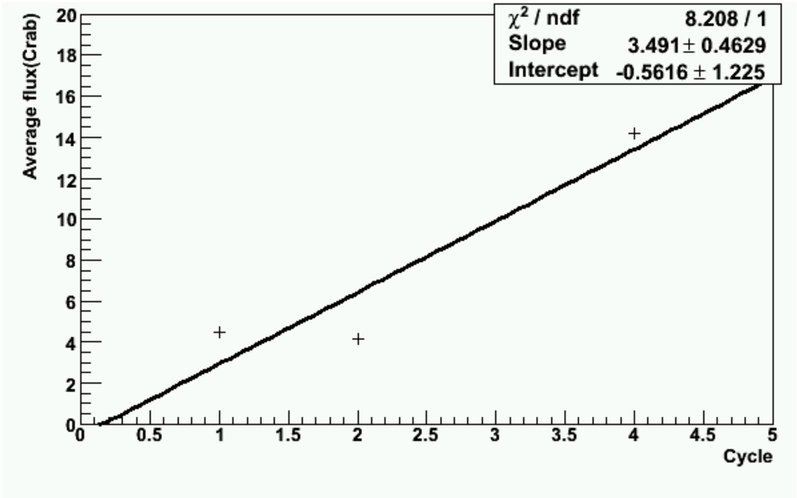


Simulated light curve (separate)

Simulate light curve with separate high state and low state

- Calculated average flux and maximum flux to plot them as a function of the number of cycles
- Average flux and maximum flux in each flare are proportional to the number of cycles.





Simulated light curve

Simulate light curve with separate high state and low state

Low state

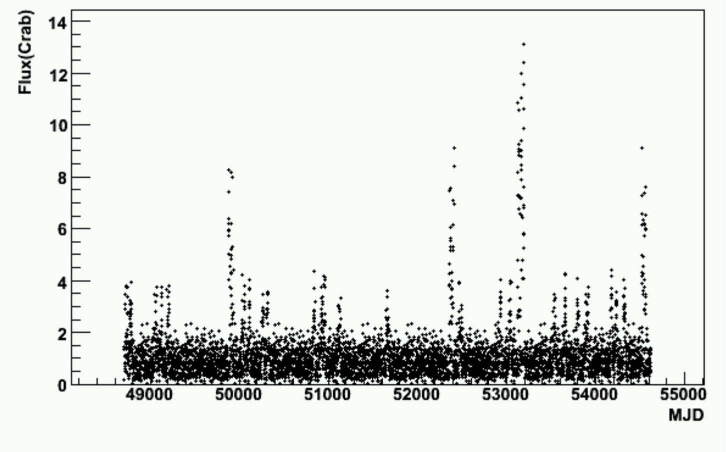
- Generating from periodogram (Power law)
- Power from slope of low state periodogram from real data

High state

- Generating from periodic function
- Period from fit to real data
- Cycle is the standard duration of a flare
- Number of cycles in high state has an Exponential distribution
- Average flux and maximum flux in each flare are proportional to the number of cycles.

$$x_i = (N_c S_{max} + I_{max} - N_c S_{avg} - I_{avg}) \sin\left(\frac{2\pi t_i}{T} - \frac{\pi}{2}\right) + (N_c S_{avg} + I_{avg})$$

- **Simulated light curve(separate)**



- **Corresponding periodogram**