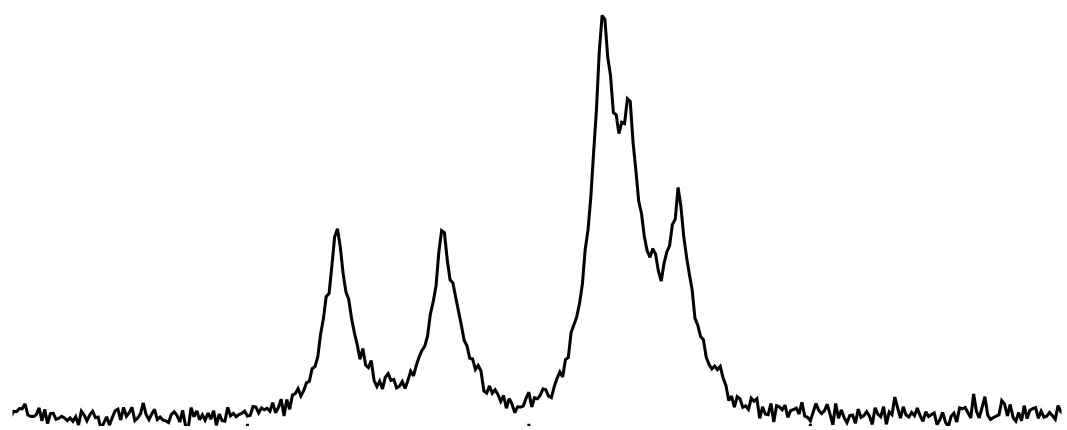
Statistical analysis of multiwavelength light curves



Fermi and Jansky
Our Evolving Understanding of AGN
St. Michaels
10 Nov 2011

Stefan Larsson
Stockholm University
"with ~~a little~~ help of my friends"
in the Fermi collaboration

AIM?

Characterize variability and MW correlations

and/or

Test theoretical models

Also

Discover new phenomena

With

Light curves and statistical tools

Complications

S/N, sampling, time resolution, obs length,

non-stationarity, tac, world economy ...

DATA?

Fermi: Regular sampling, high duty cycle. Low to moderate S/N
(Events or binned?)
Radio: Semi regular at best, but higher S/N

TOOLS?

Variability:

- Variance
- Flare profile fitting
- Flux duty cycles
- Power Density Spectra
- Auto Correlation Function
- Structure Function
- Wavelets

MW Correlation:

- Direct light curve comparison
- Flux - Flux plots, tracks (possibly with time lags)
- Cross Correlation Function
- Cross Spectrum

This talk will focus on practical aspects of Cross Correlation analysis

For a wider overview see:

"Methods for Cross-Analyzing Radio and gamma-ray Time Series Data"

Jeff Scargle

when Fermi met Jansky in 2010

Recipes to calculate the Cross Correlation Function for unevenly sampled light curves:

DCCF, Discrete CCF (Edelson & Krolik, 1988)
ICCF, Interpolated CCF (e.g. Gaskell & Peterson, 1987)
ZCCF, Z-transform CCF (Alexander, 1997)
Inverse FT of PDS (Scargle, 1989)

In most cases we want:

1. Strength and significance of correlation
2. Lag between the two light curves, with uncertainty

Discrete CCF (Edelson & Krolik, 1988)

The classical CCF:

$$\mathrm{CF}(\tau) = \frac{E\{[a(t) - \bar{a}][b(t + \tau) - \bar{b}]\}}{\sigma_a \sigma_b}\,, \qquad (2)$$

For each pair of points in LC 1 and 2 compute their contribution to the CCF at the lag corresponding to their time separation

Pair by pair

Unbinned DCF:

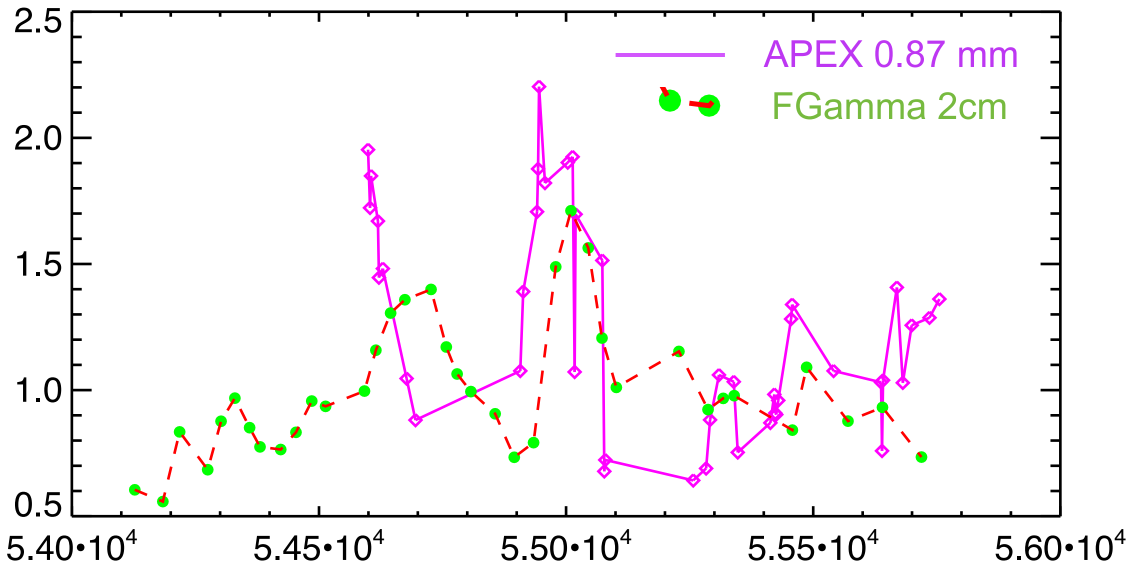
$$\mathrm{UDCF}_{ij} = \frac{(a_i - \bar{a})(b_j - \bar{b})}{\sqrt{(\sigma_a^2 - e_a^2)(\sigma_b^2 - e_b^2)}}\,, \qquad (3)$$

e = rms measurement errors

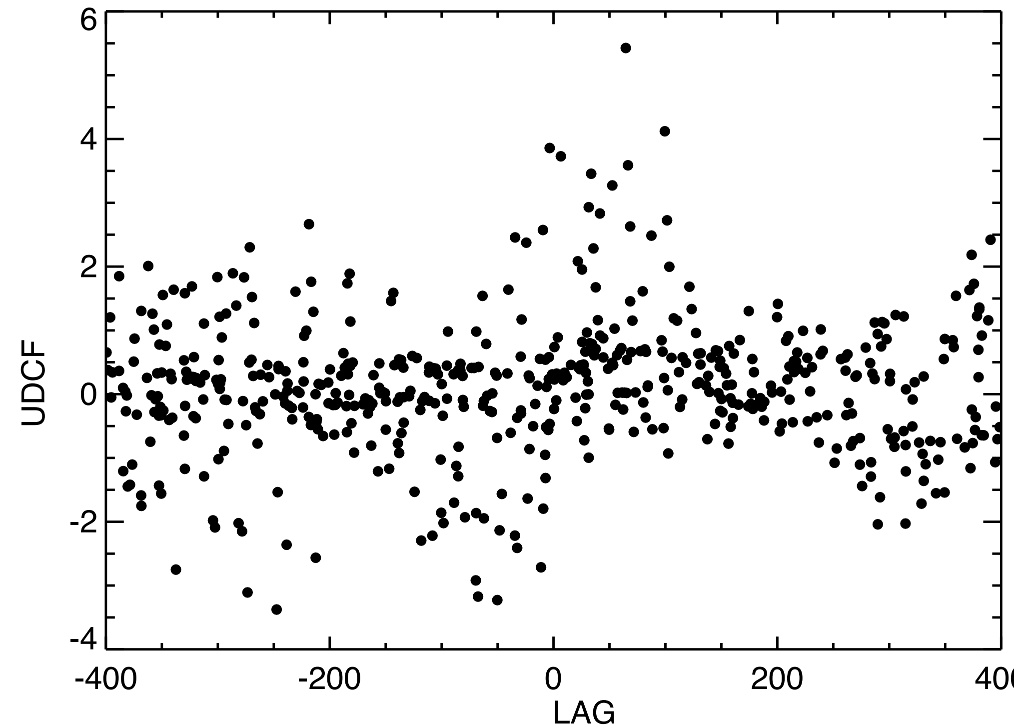
Bin (average) the UDCF => DCCF

DCCF Example

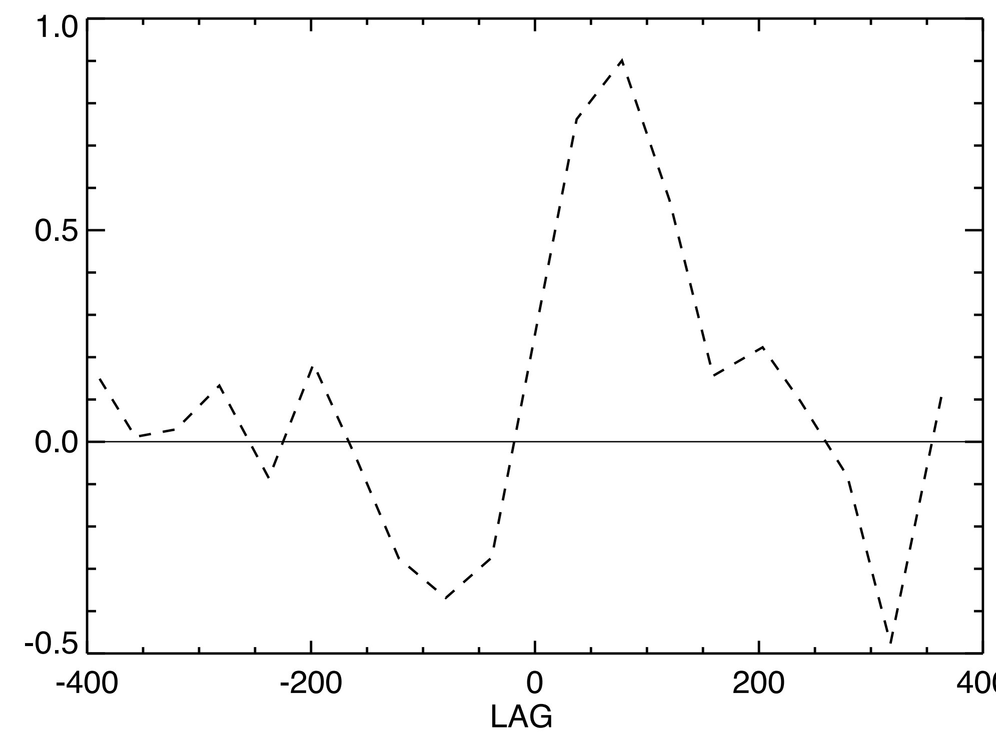
PKS 1510-089



Unbinned DCF:



Bin (average) the UDCF => DCCF



The CCF is affected by

1) Measurement noise

2) Time sampling

3) Stochastic variability

- Different realizations of the same stochastic process will have different PDS/ACF/CCF

- Chance correlations (E.g. 2 independent short LCs with one flare each will show a strong correlation at a lag corresponding to the time shift between the flares

Example of model dependent Monte Carlo method for CCFs

1. Simulate two light curves with some correlation and lag.
2. Sample the two LCs
3. Add errors
4. Compute CCF
5. Determine Lag
6. Repeat N times
7. Compute Lag distribution

Compare to the previous talk where Walter Max-Moerbeck used phase randomization of the Fourier transformed data to estimate correlation significances.

A model independent Monte Carlo method (Peterson et al, 1998) to address error points 1 and 2 (measurement noise and sampling) but NOT 3 (stochastic variations).

1) Add 1 sigma errors to the data

2) Make a bootstrap-like point selection.

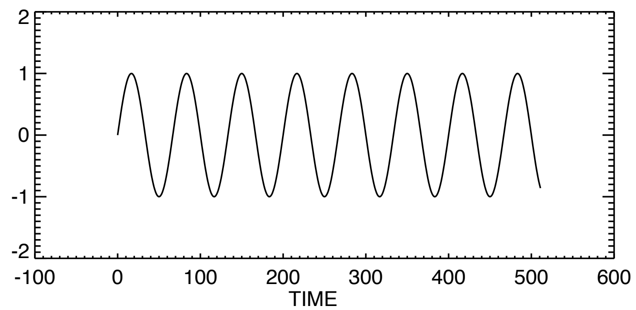
3) Compute CCF and determine Lag of the peak

4) Repeat N times and compute rms(lag)

Note 2: A recipe to compute Lag error not DCCF point errors!
(You can still do that if you keep in mind that the different lag points are correlated).

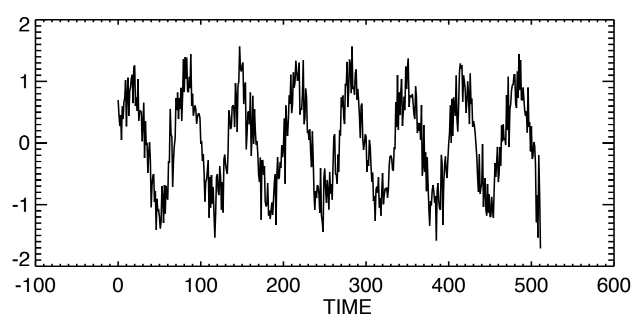
1) Measurement noise

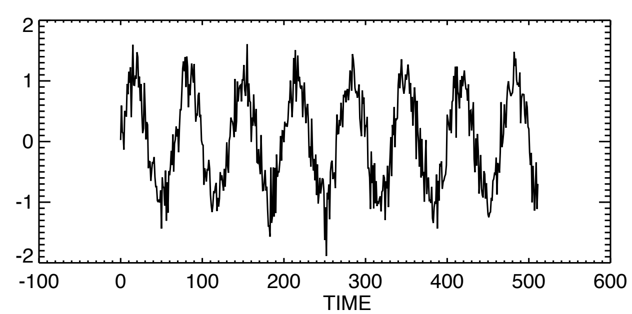
Two “observations” of the same light curve with added noise (independent gaussian noise).

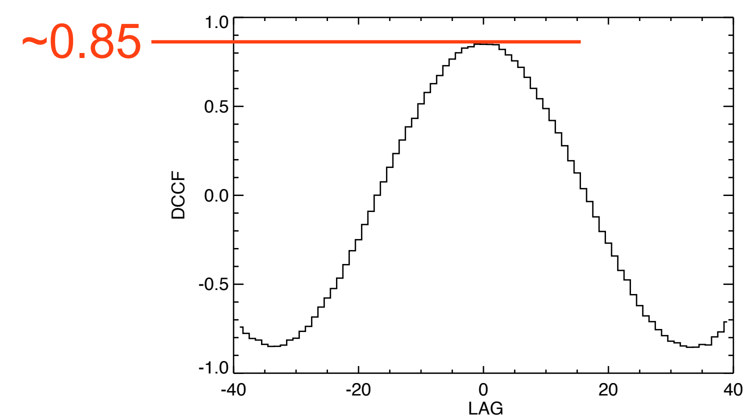


512 points Sine ampl = 1

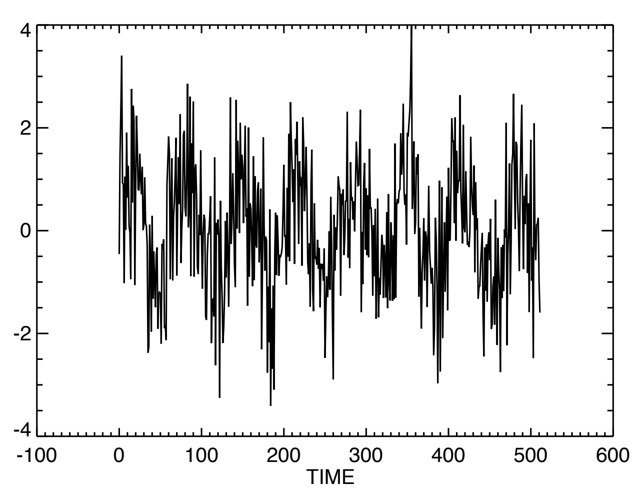
rms noise = 0.3

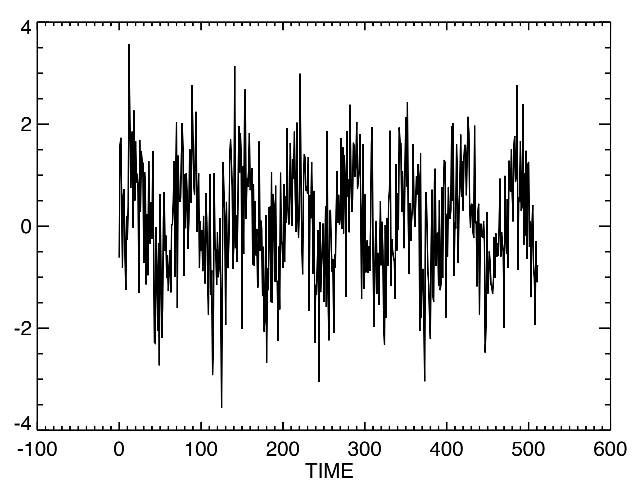


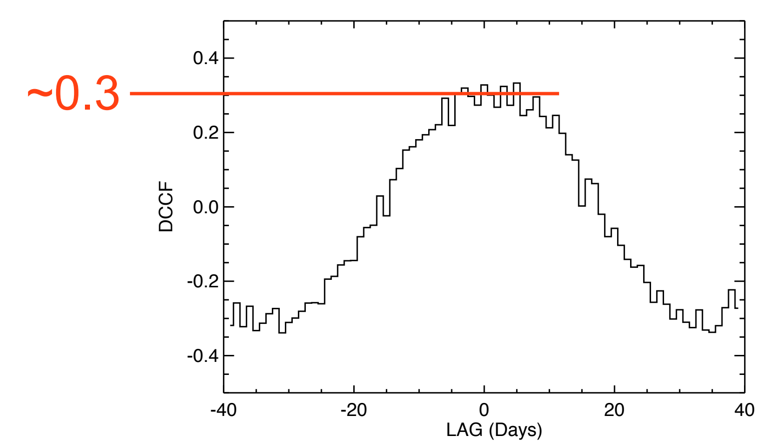




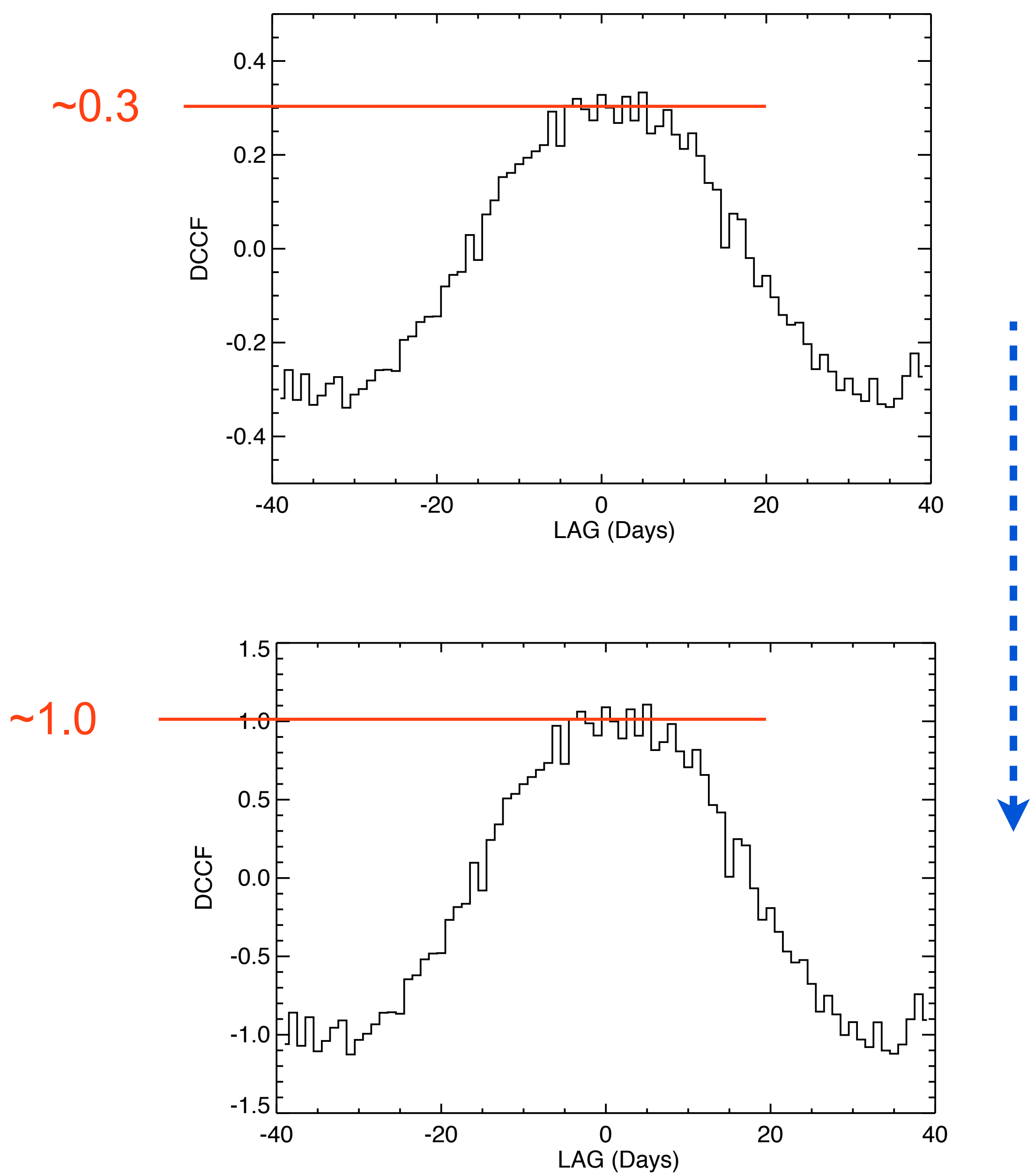
rms noise = 1.0







1) Measurement noise



Correcting for the de-correlation due to white noise

$$\mathrm{UDCF}_{ij} = \frac{(a_i - \bar{a})(b_j - \bar{b})}{\sqrt{(\sigma_a^2 - e_a^2)(\sigma_b^2 - e_b^2)}}\,, \tag{3}$$

e = rms measurement errors

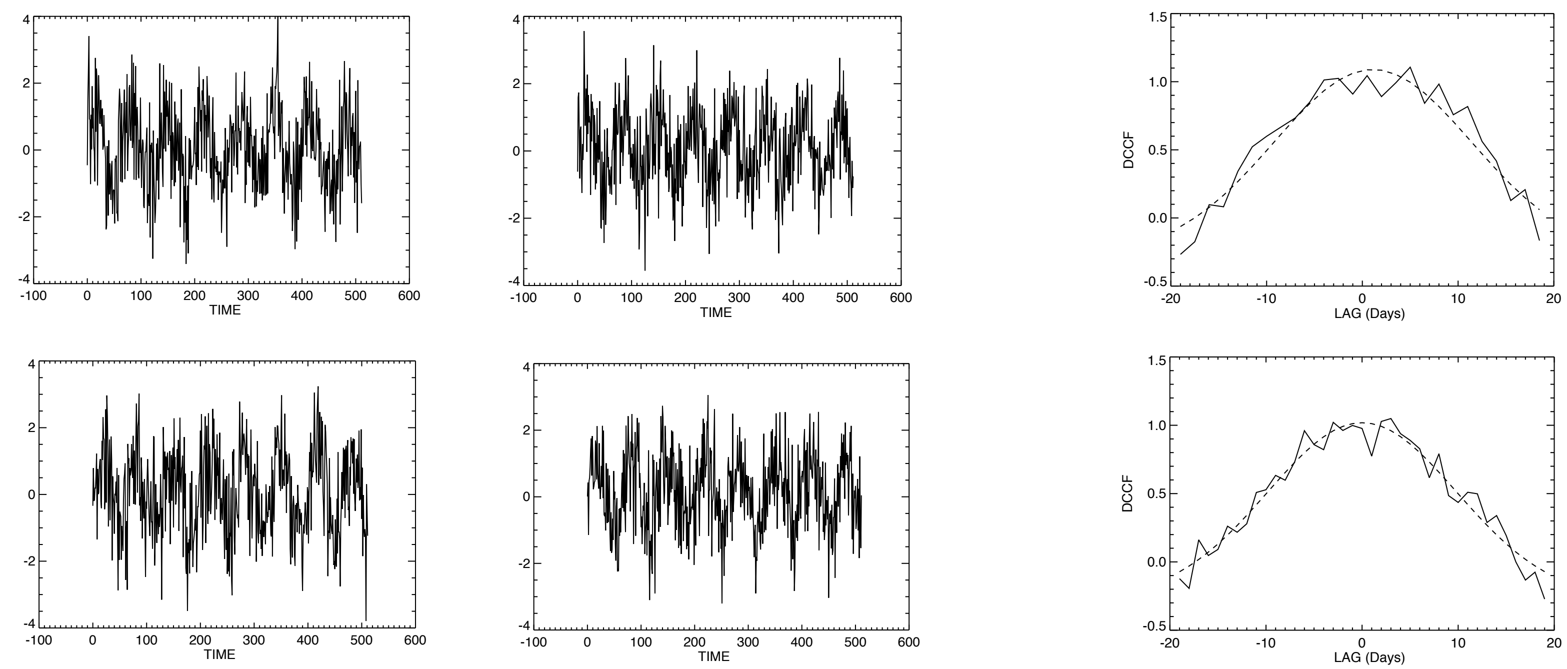
Peterson's recipe 1: Injecting white noise

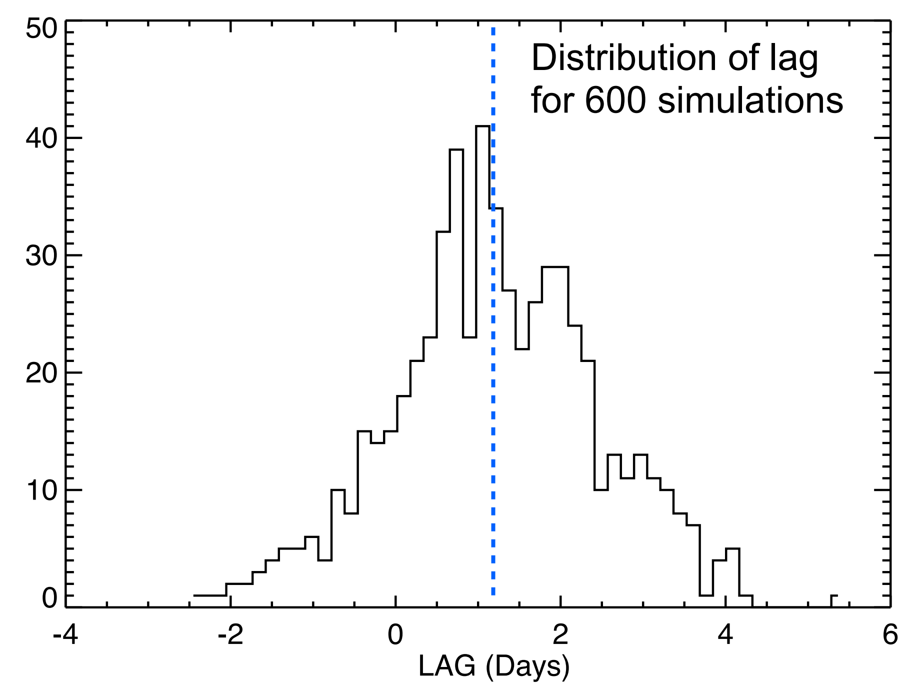
Light curve 1 | Light curve 2 | DCCF

Monte Carlo run 1 — LAG 1

Monte Carlo run 2 — LAG 2

Monte Carlo run N — LAG N





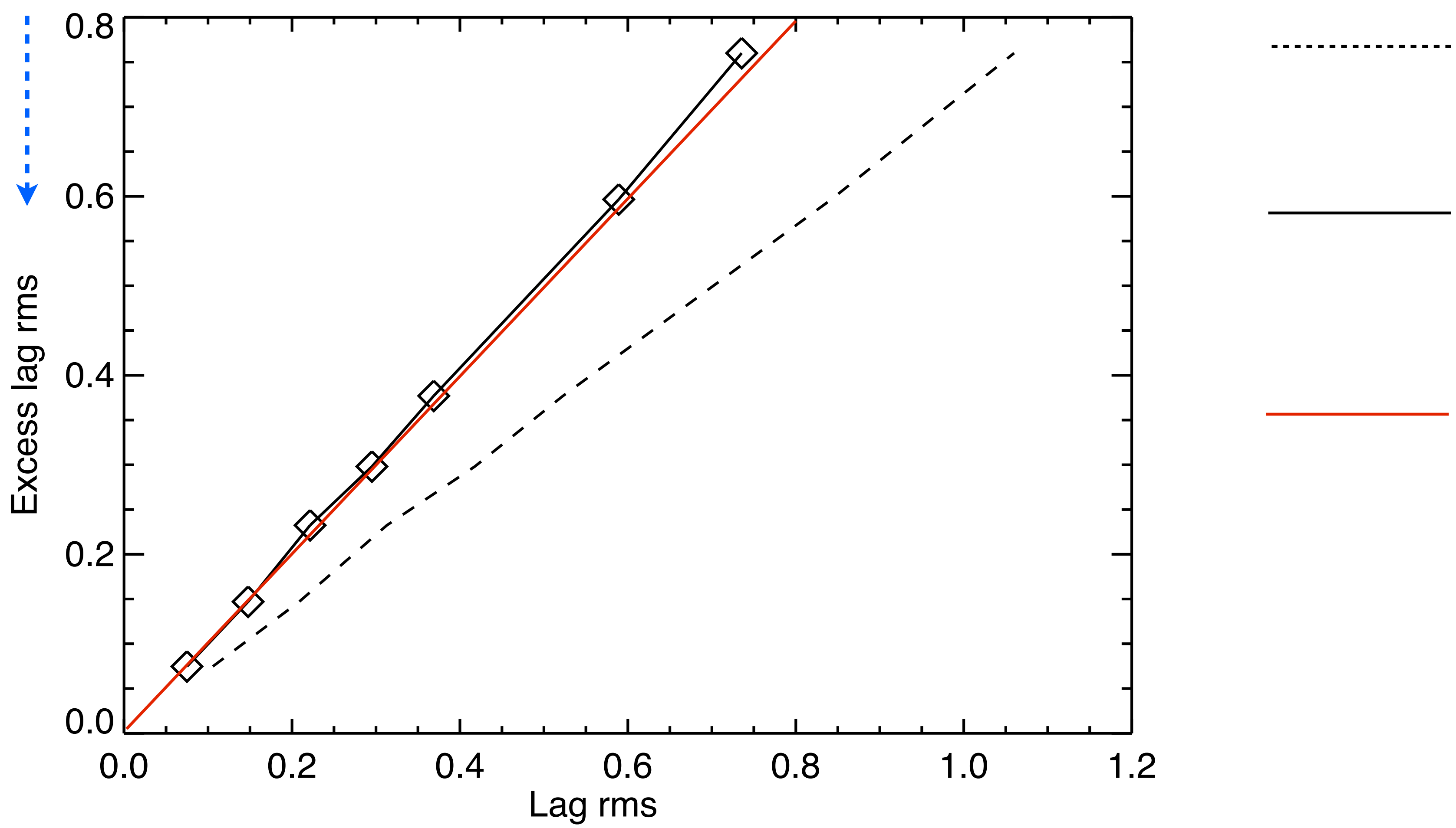
Distribution of Monte Carlo Lags gives uncertainty

"Cross Correlation Peak Distribution" (CCPD), Maoz & Netzer (1989)

Peterson's recipe 1: Injecting white noise

Test of the error estimation with known input (Simulated data with different S/N)

Excess lag rms is relative to the analyzed LC



“True” rms

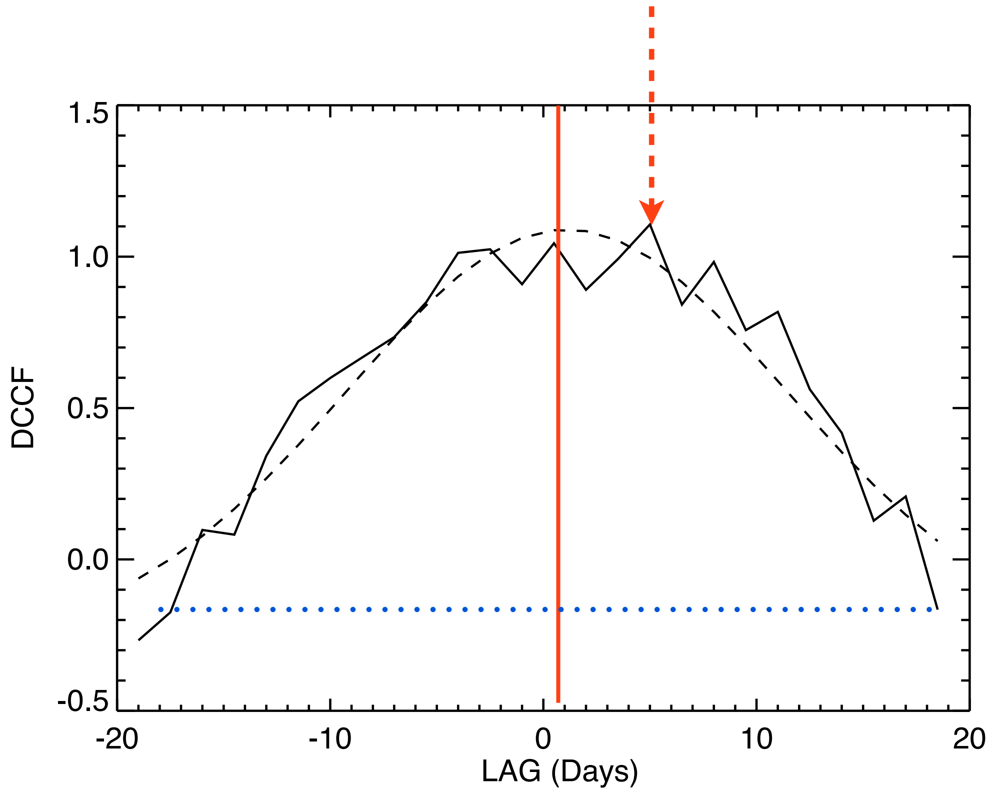
White noise added to one light curve

White noise added to both light curves

Assumption of error linearity (from light curve to lag estimate)

What is the lag at the peak?

- The highest value of the DCCF?
- Centroid?
- Max of a fitted function?



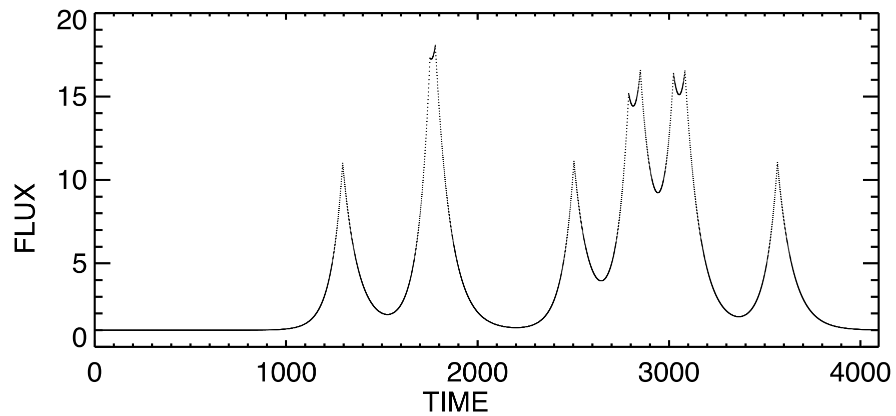
Gaussian fit to estimate correlation peak
[I use gaussian just for convenience]

Wide enough to get a reasonable fit but not so wide that it is determined by the base. (Wade & Horne, 1998)

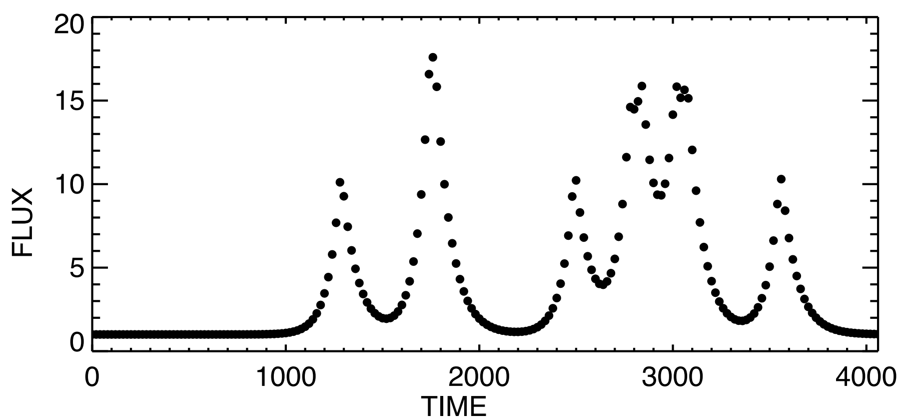
Peterson's recipe 2: Uneven sampling

Let's make a simulation....

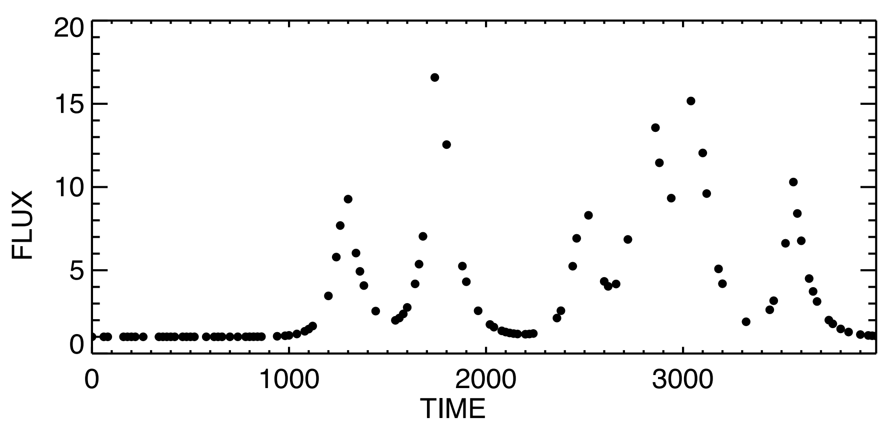
True light curve (no noise)



Evenly sampled 204 points

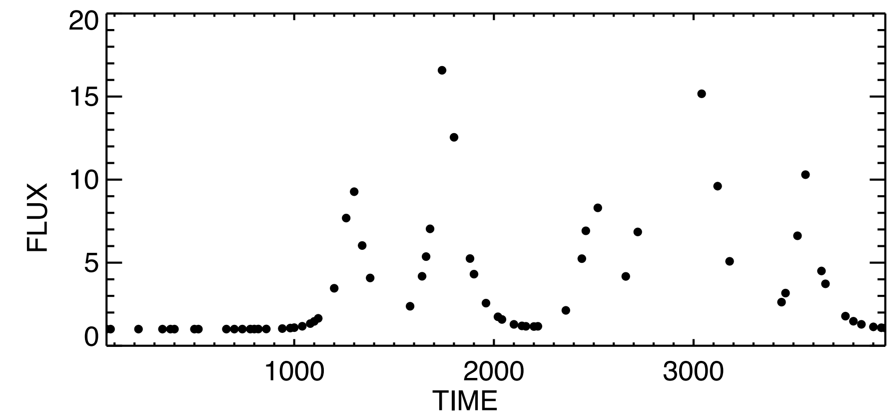


100 random observations of the LC

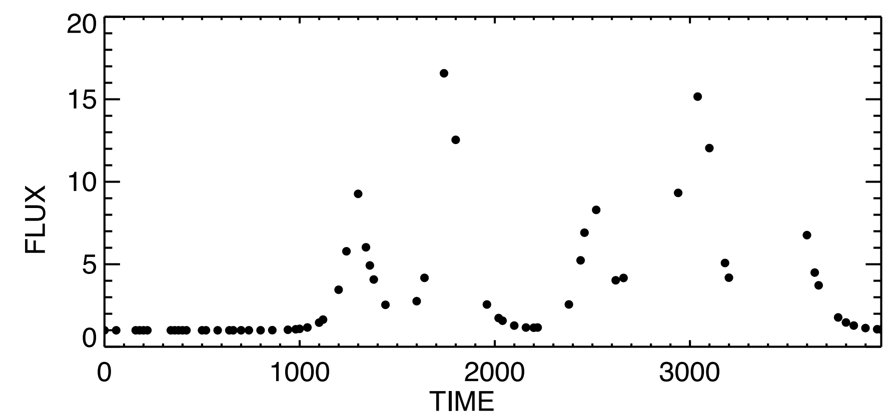


"Bootstrap" resampling

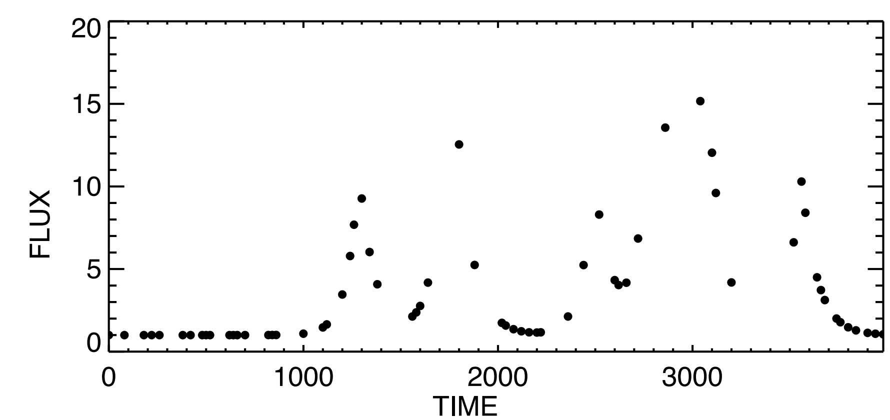
1



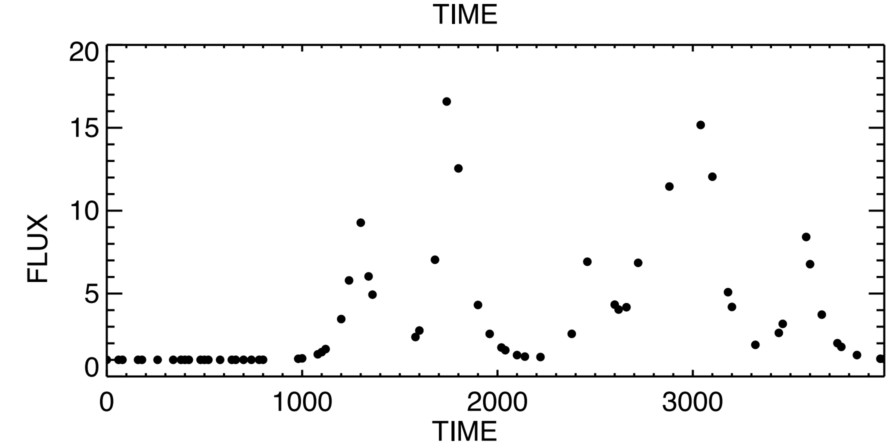
2



3



4



N

Close to detection limit you may have to run separate MCs and
add variances,

Variance (errors LC 1) + Variance (errors LC 2) + Variance Bootstrap

DCCF significances by Mixed Source Correlation

Fermi and Radio monitoring programs are now providing light curves for a large number of sources.

Assuming that all sources have similar variability properties

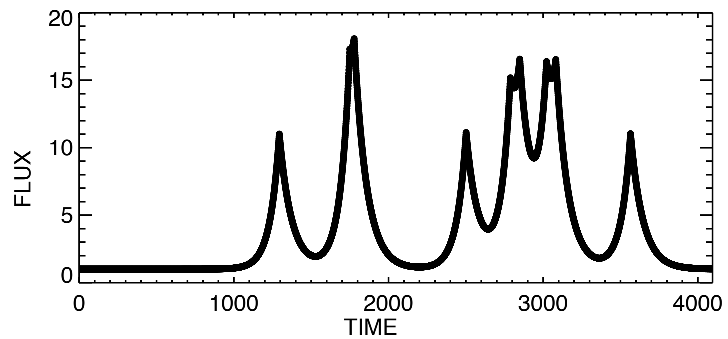
We can estimate the probability of stochastic chance correlations
by correlating each radio light curve with the gamma-ray light curves of all the other sources.
[Gamma is evenly sampled but radio is not!]

Advantage: Requires no characterization of the variability

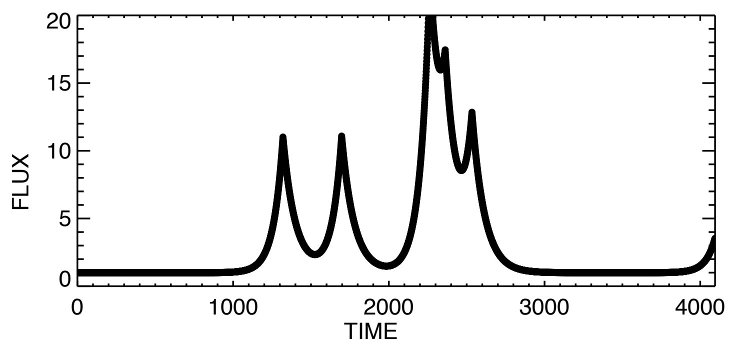
Disadvantage: Limited number of test light curves

DCCF significances by
Mixed Source Correlation

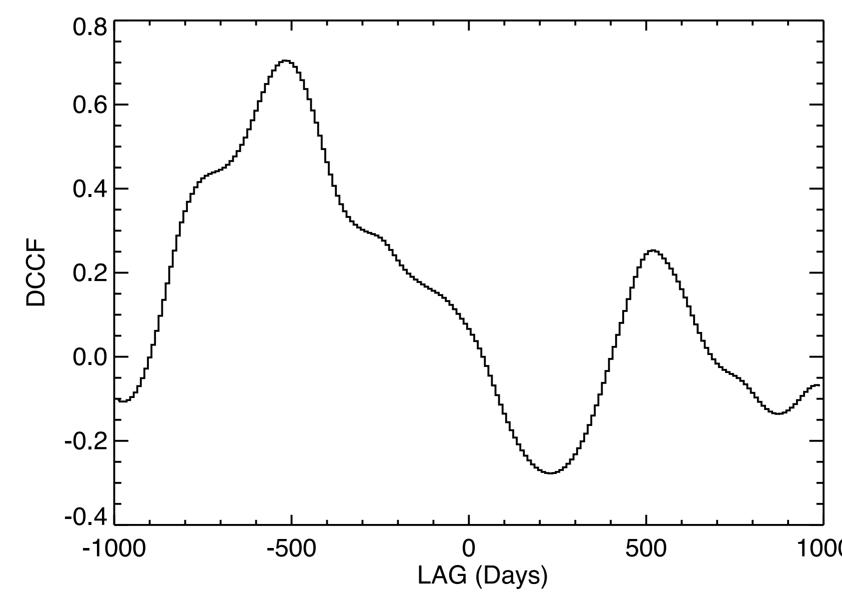
Gamma
JXXXX



Radio
JYYYY

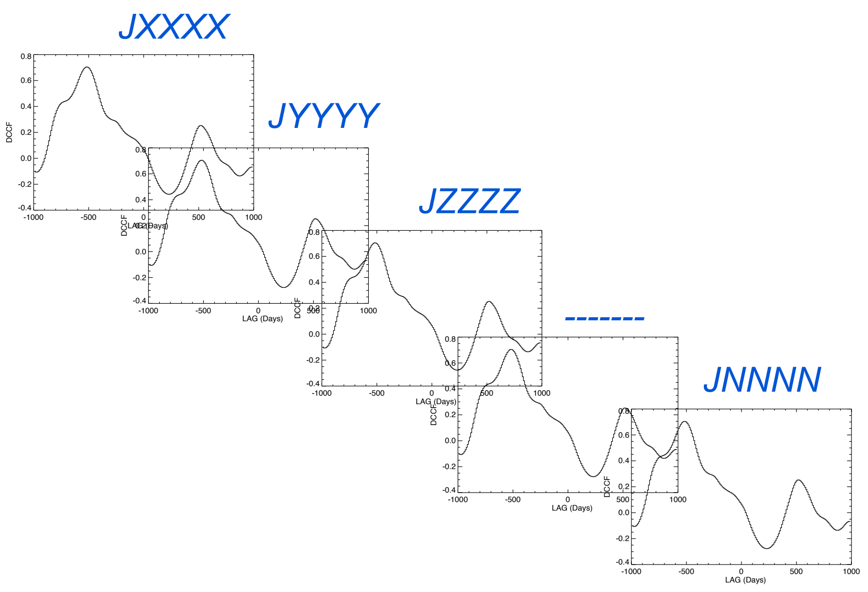


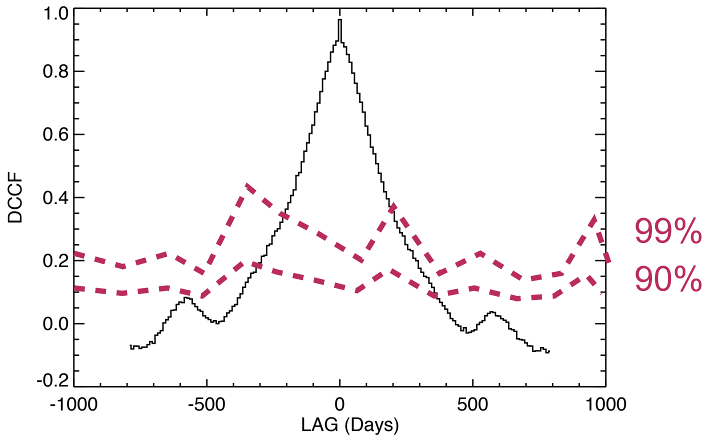
DCCF



Loop over all “JYYYY”
Compare DCCF
distribution with
source DCCF

Average the DCCFs for a sample of sources





AND do the same for the corresponding samples of mixed DCCFs

→ A comparison sample of N-1 DCCFs

Applied to 3 years of Gamma - Radio data (Fermi & FGamma)
[talk by Lars Fuhrmann]

Are correlation properties persistent?

We can analyze segments of the data to

1) Evaluate the significance of the correlation

2) Look for variations in correlation properties

Various observations have revealed variations in MW lags with time.
Such variations can reflect either:

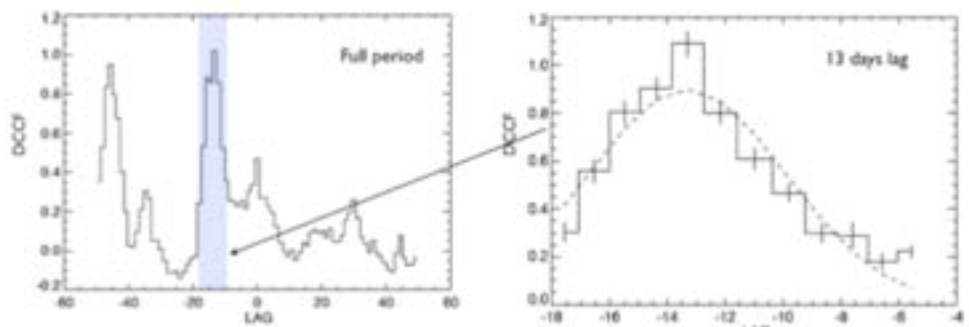
- Real physical changes in the source

- Stochastic variations [Check by ACF/PDS]

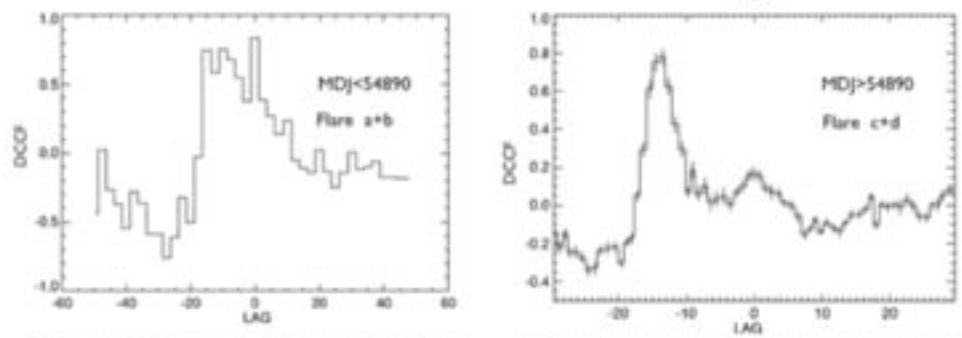
PKS 1510-089 4 major flaring episodes (Sep 2008 - Jun 2009)
Each 3 - 4 weeks long + Smaller sub-flares

Abdo et al, 2010, Ap. J., 721, 1425

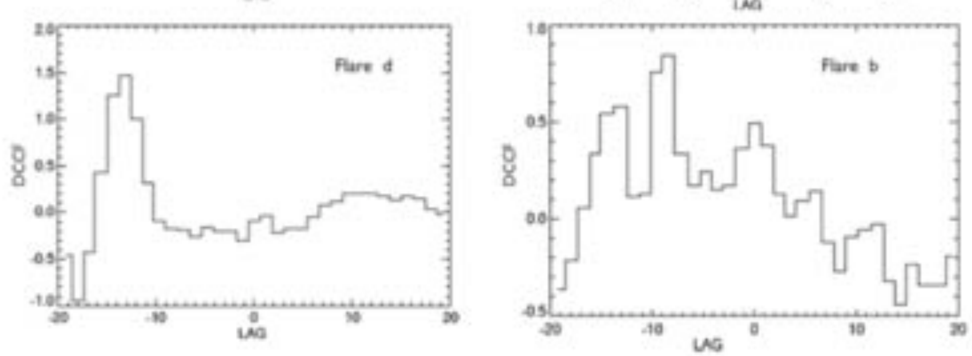
Cross correlations show ~13 day lag (R lagging gamma) both for the total light curve and for the individual flares + correlation on time scales of < 2 days



DCCF for all flares

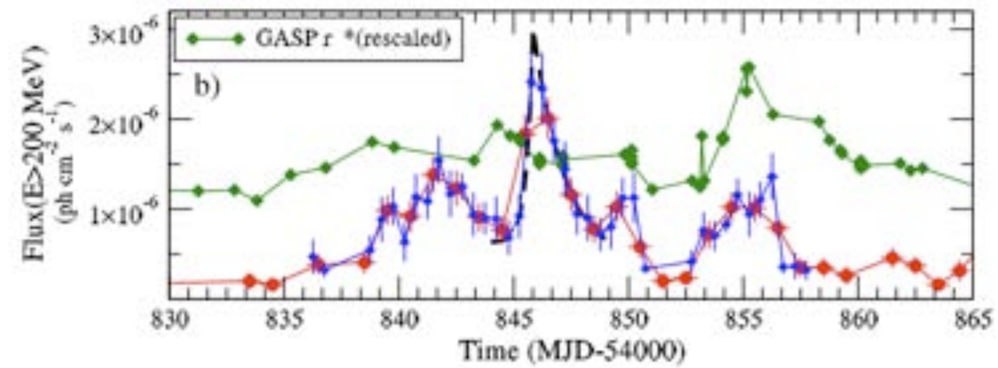


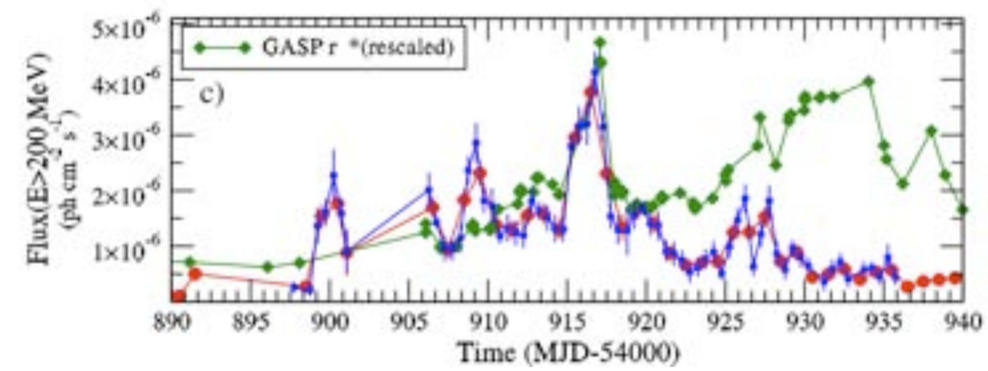
DCCF for a+b and c+d

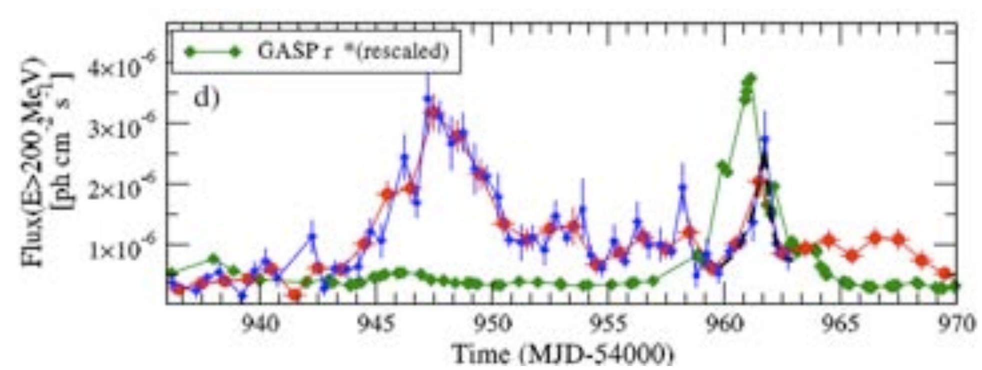


DCCF for d and b

While R and gamma show correlations on time scales < 2 days the ratio of R flux to gamma flux increases towards the end of the flare. This is the cause of the observed 13 day lag.







Detrending

(e.g. polynom subtraction)

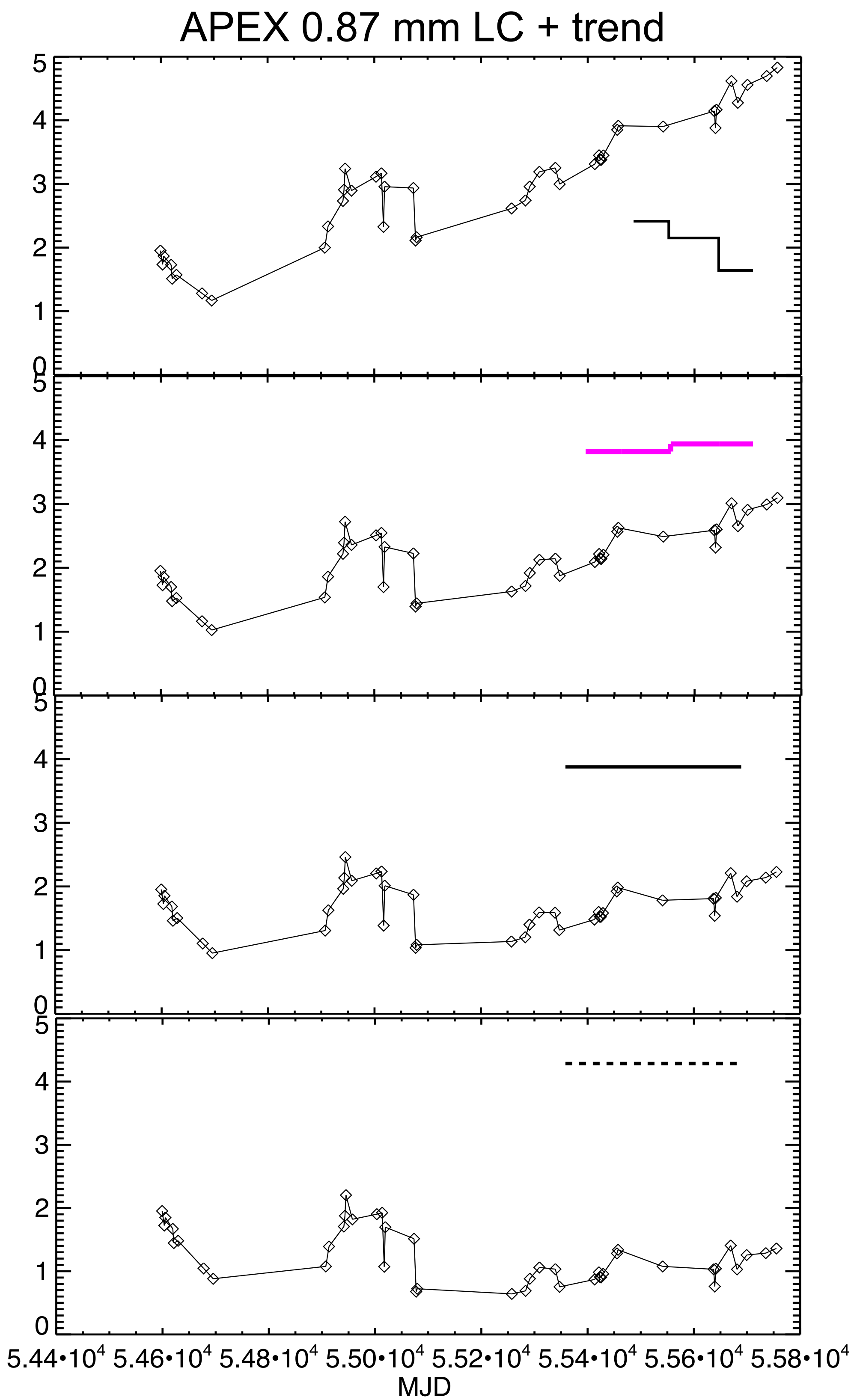
Will:

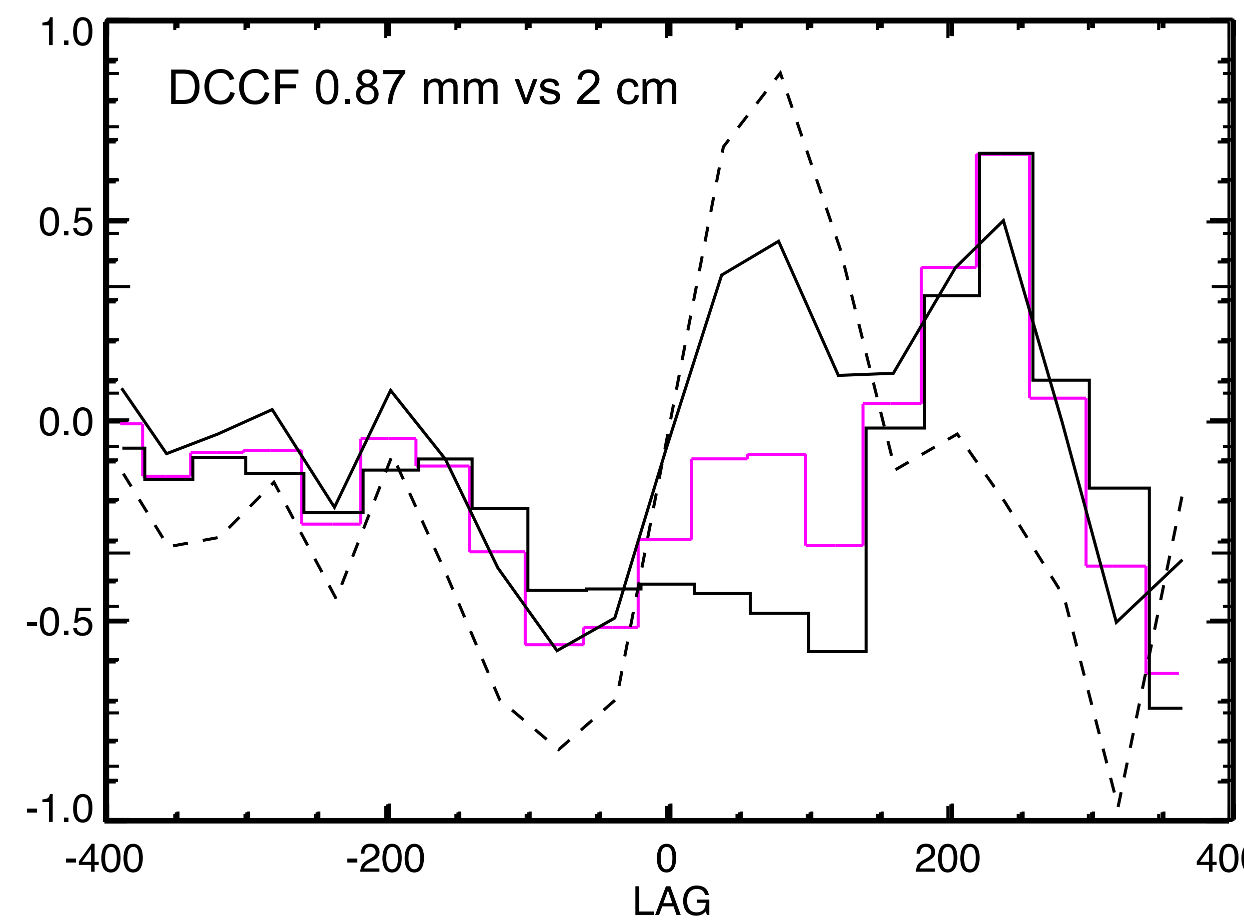
Reduce bias in Lag determination (property of the CCF, see e.g. Welsh, 1999)

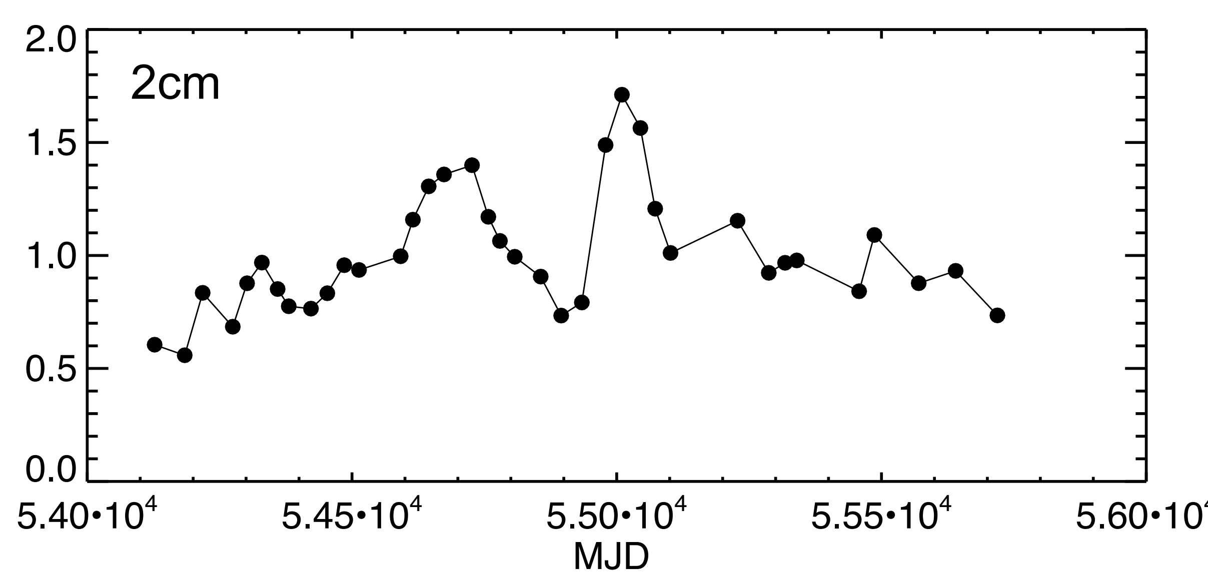
May increase or decrease S/N in lag determination (Long time scales have few points and tend to be noisy but if detrending removes most of the signal we are left with noise)

Reduce the sensitivity on long time scales (Long and short time scales may have different correlation properties).

The effect of adding a trend







The challenge for correlation analysis:

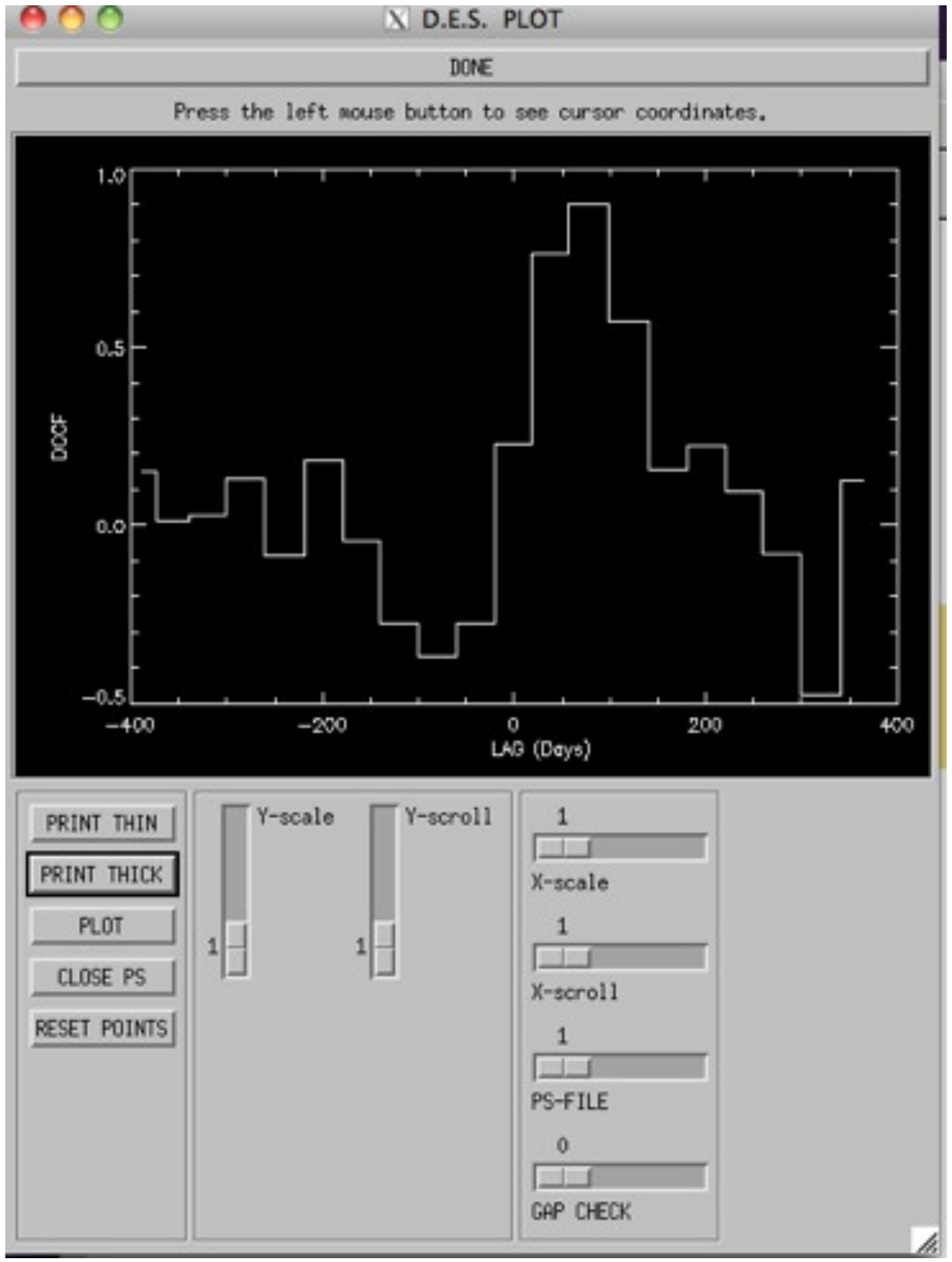
Complexity: More than one simultaneous type of correlation.

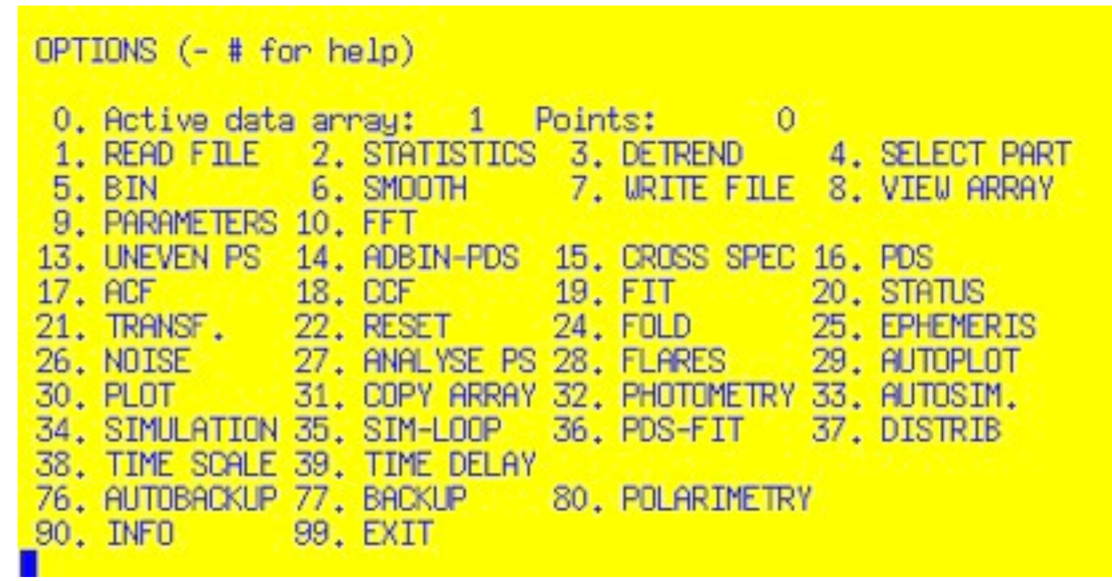
Non-stationarity: Variability and/or correlation properties changing with time. Depends on time scale.

Patience (long LCs) is needed to disentangle these effects.

DES - Data Exploration System

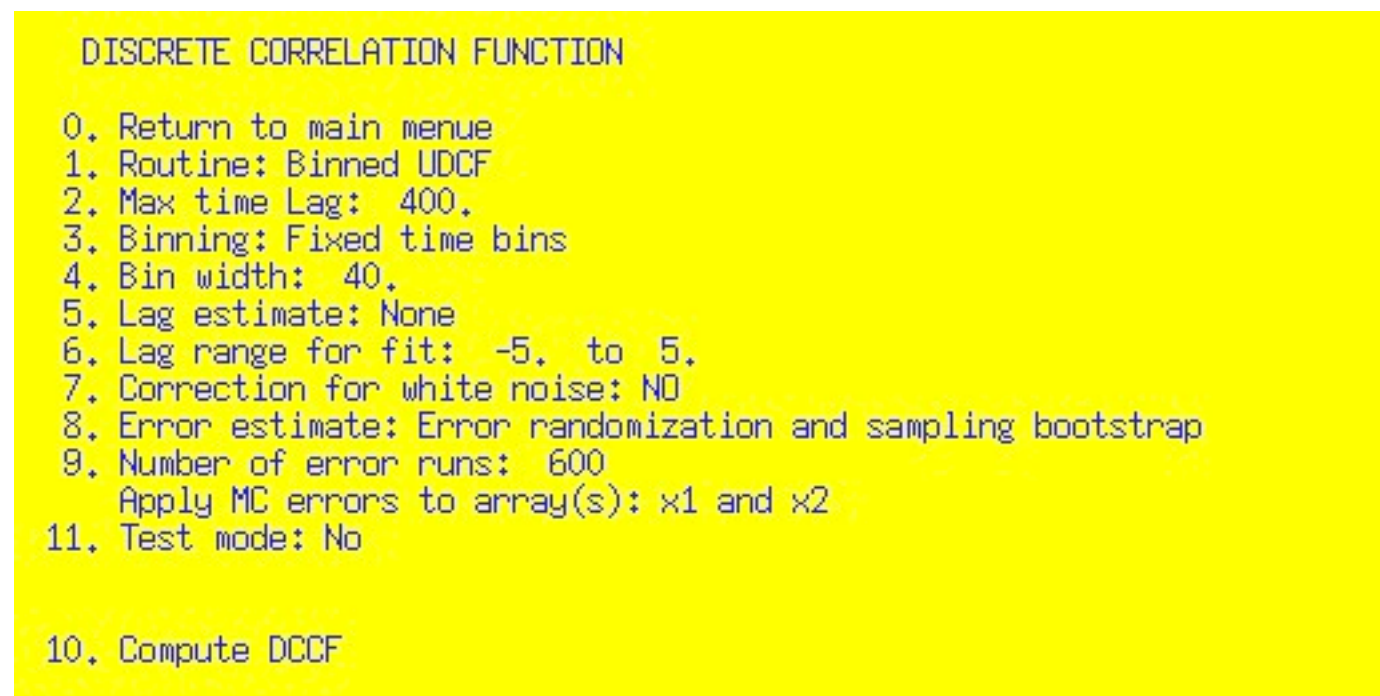
http://ttt.astro.su.se/groups/head/SL/des7.html





```
OPTIONS (- # for help)

 0. Active data array:   1   Points:        0
 1. READ FILE      2. STATISTICS  3. DETREND      4. SELECT PART
 5. BIN            6. SMOOTH      7. WRITE FILE   8. VIEW ARRAY
 9. PARAMETERS    10. FFT
13. UNEVEN PS     14. ADBIN-PDS  15. CROSS SPEC  16. PDS
17. ACF           18. CCF        19. FIT         20. STATUS
21. TRANSF.       22. RESET      24. FOLD        25. EPHEMERIS
26. NOISE         27. ANALYSE PS 28. FLARES      29. AUTOPLOT
30. PLOT          31. COPY ARRAY 32. PHOTOMETRY  33. AUTOSIM.
34. SIMULATION    35. SIM-LOOP   36. PDS-FIT     37. DISTRIB
38. TIME SCALE    39. TIME DELAY
76. AUTOBACKUP    77. BACKUP     80. POLARIMETRY
90. INFO          99. EXIT
```



```
 DISCRETE CORRELATION FUNCTION

 0. Return to main menue
 1. Routine: Binned UDCF
 2. Max time Lag:  400.
 3. Binning: Fixed time bins
 4. Bin width:  40.
 5. Lag estimate: None
 6. Lag range for fit:  -5.  to  5.
 7. Correction for white noise: NO
 8. Error estimate: Error randomization and sampling bootstrap
 9. Number of error runs:  600
    Apply MC errors to array(s): x1 and x2
11. Test mode: No


10. Compute DCCF
```