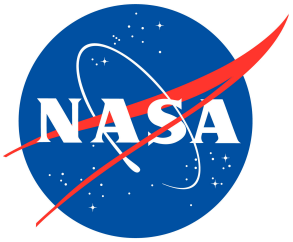


Time Series Analysis of Data With Gaps

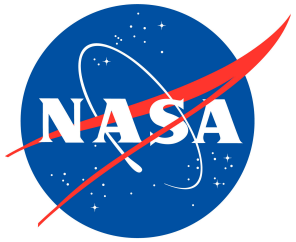
Jeff Scargle

Space Science Division
NASA Ames Research Center

Jeffrey.D.Scargle@nasa.gov

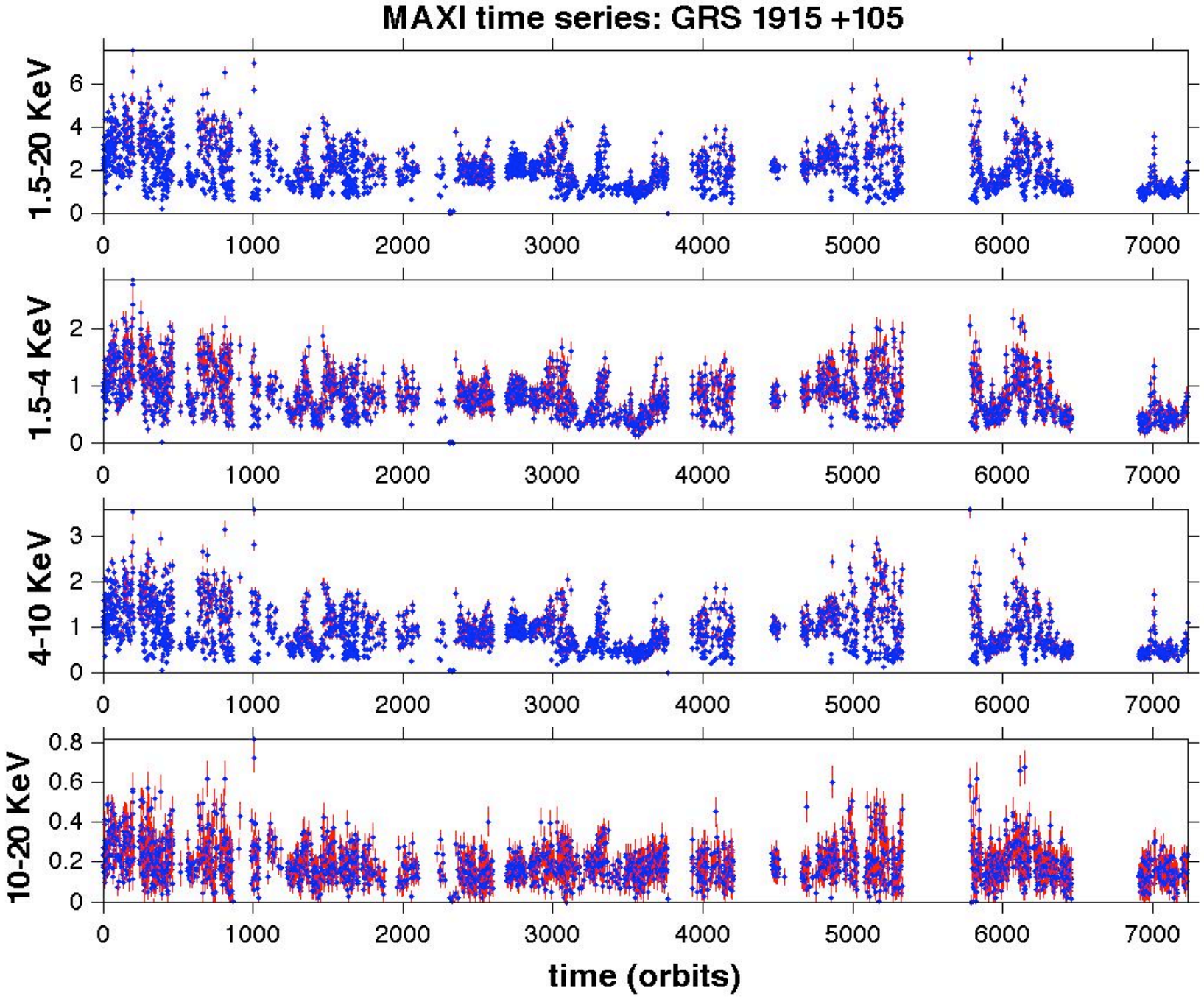
The First Year of MAXI: Monitoring variable X-ray sources

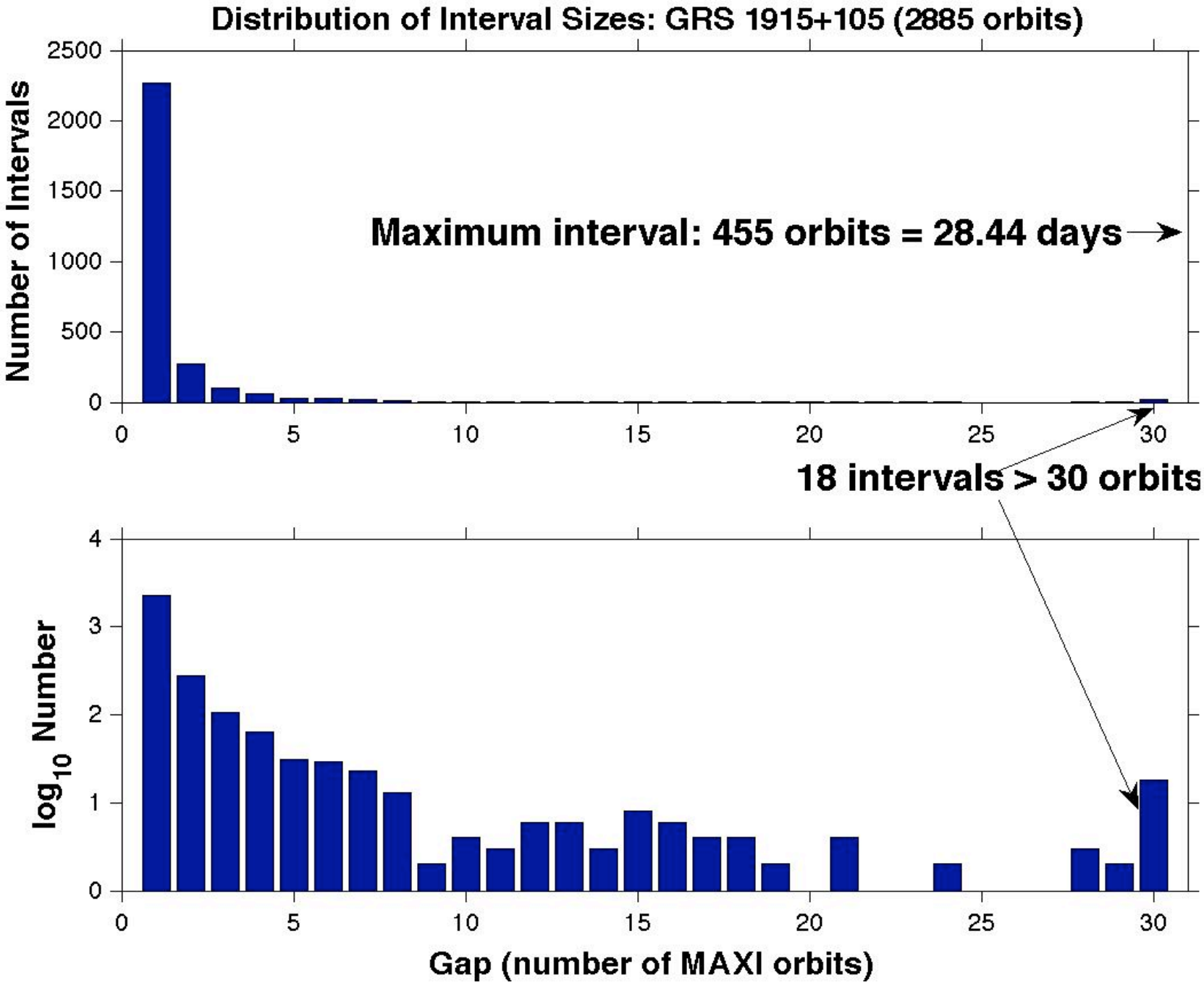
Special thanks to Tatehiro Mihara-san

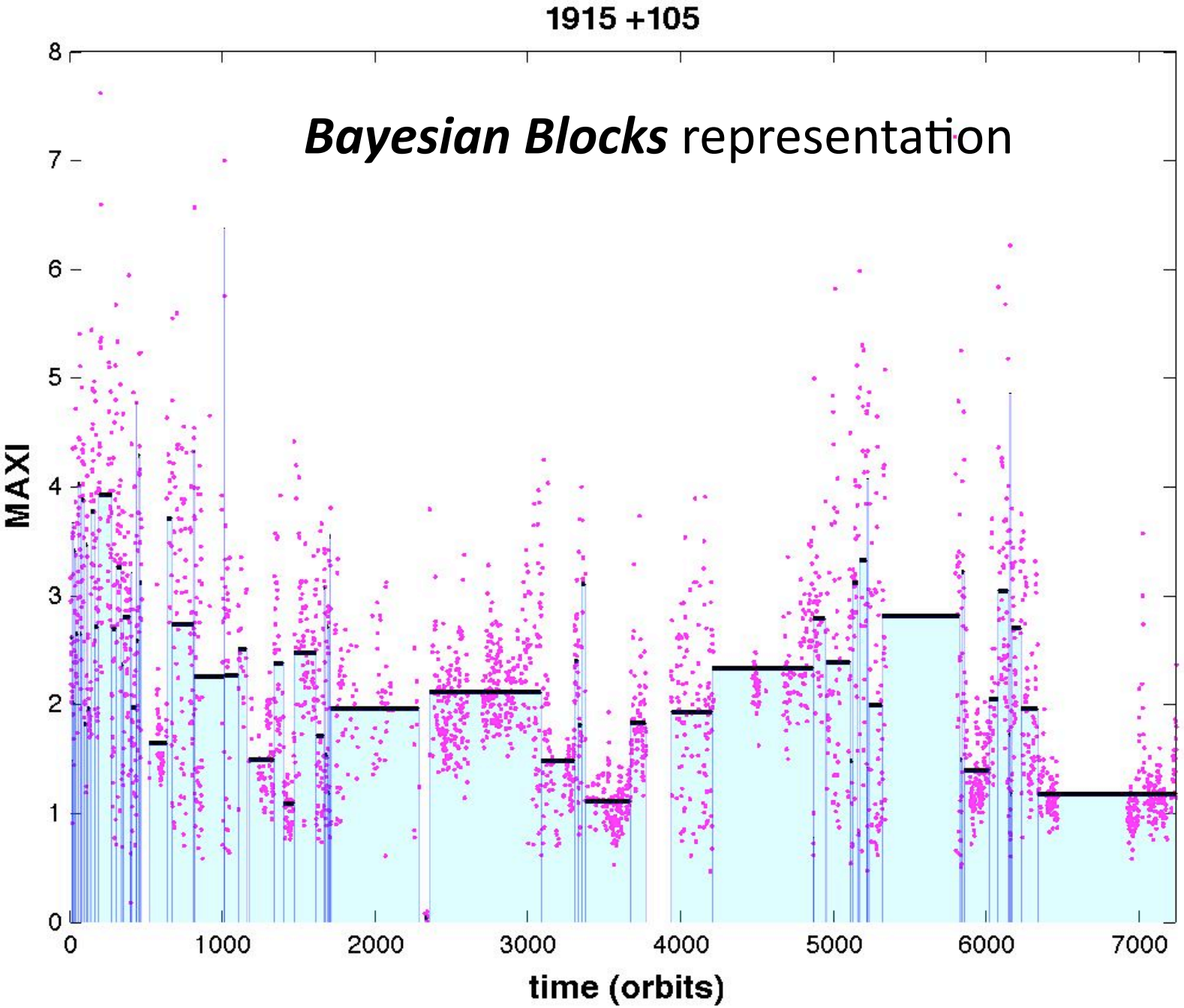


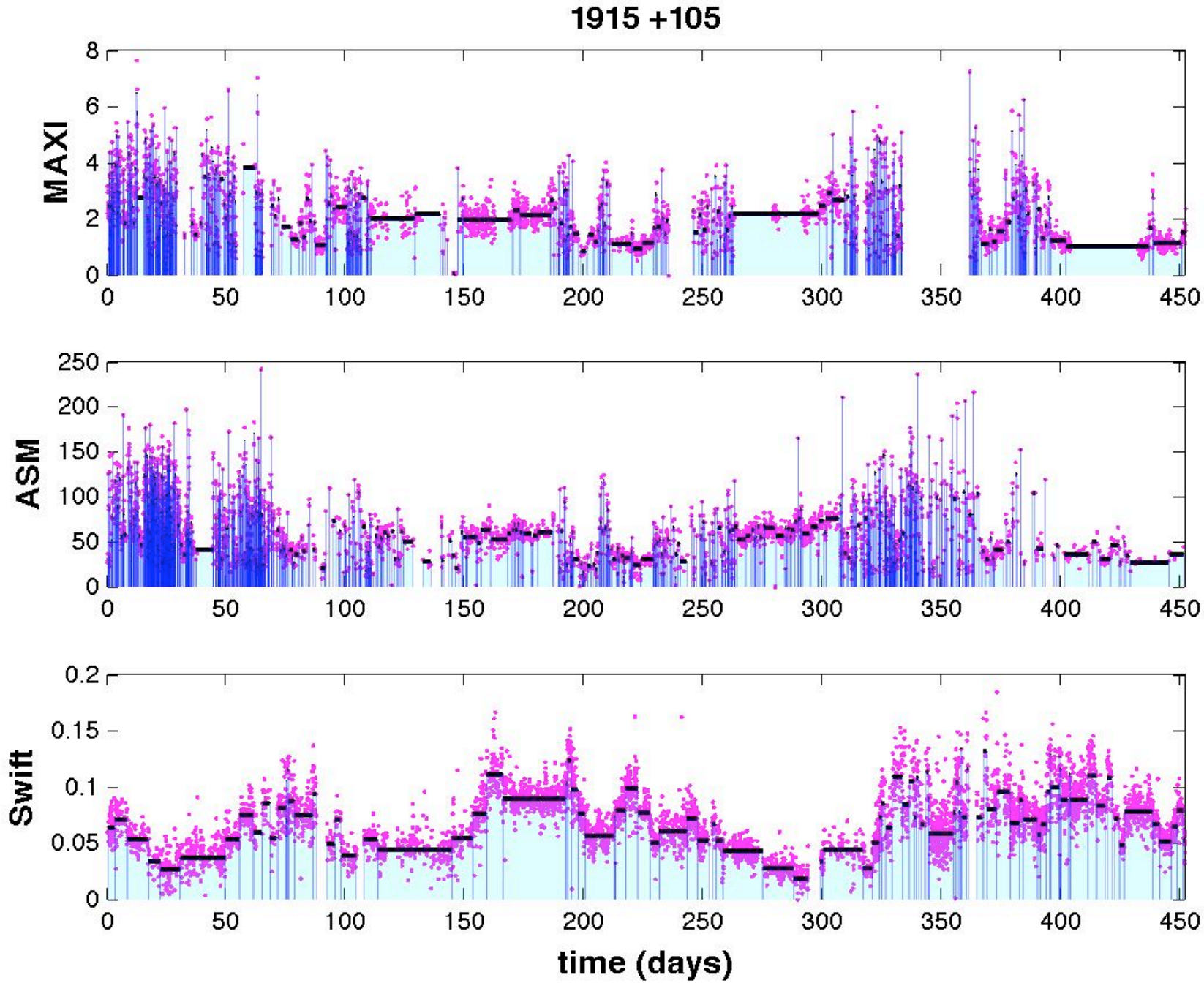
Practical Time Series Methods

- Data Issues: Sampling Intervals and Gaps
- Light Curve Representations (Data Cells)
- Scatter Plots
- Correlation Functions (Edelson and Krolik algorithm)
- Spectral analysis:
  - Amplitude (Power)
  - Phase
  - Wavelet Transform (Scalogram)
  - Wavelet Power (Scalegram)
  - Structure Functions
  - Time-Scale/Time-Frequency Analysis
- Cautions: “stationarity”, “nonlinearity”, “correlations”, …

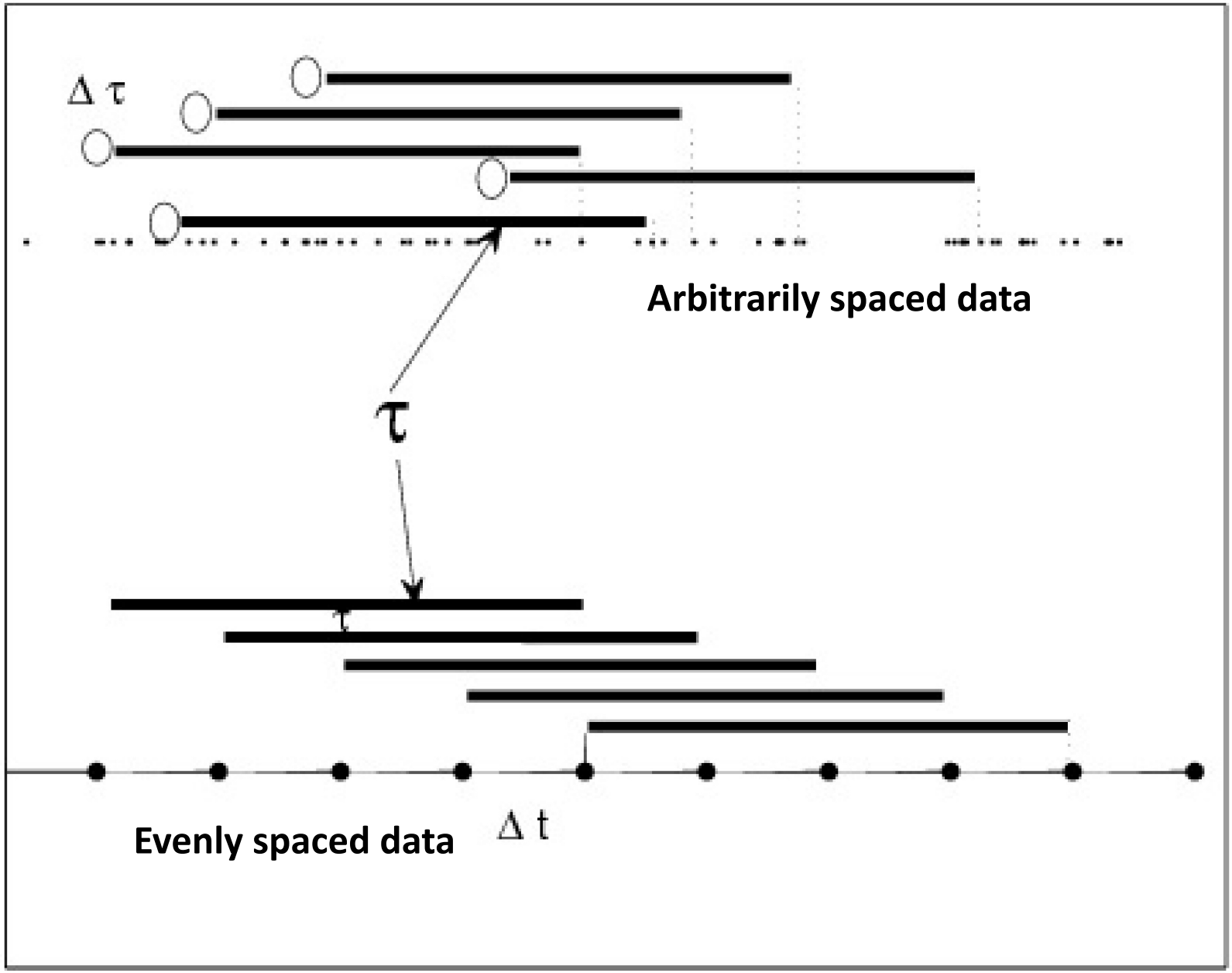

MAXI time series: GRS 1915 +105
1.5-20 KeV
1.5-4 KeV
4-10 KeV
10-20 KeV
time (orbits)
0
1000
2000
3000
4000
5000
6000
7000

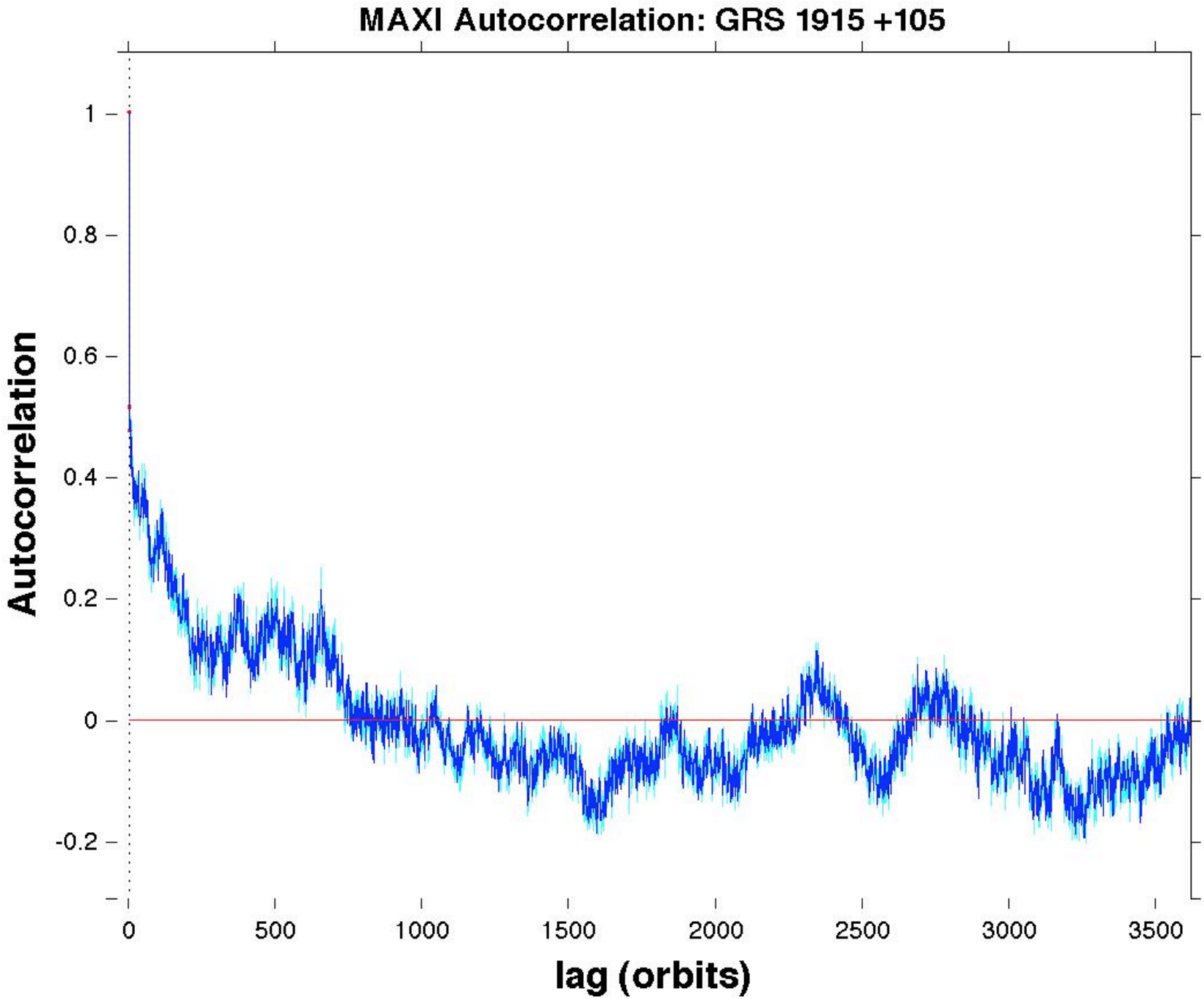

Distribution of Interval Sizes: GRS 1915+105 (2885 orbits)
Number of Intervals
2500
2000
1500
1000
500
0
0
5
10
15
20
25
30
Maximum interval: 455 orbits = 28.44 days
18 intervals > 30 orbits
log10 Number
4
3
2
1
0
0
5
10
15
20
25
30
Gap (number of MAXI orbits)

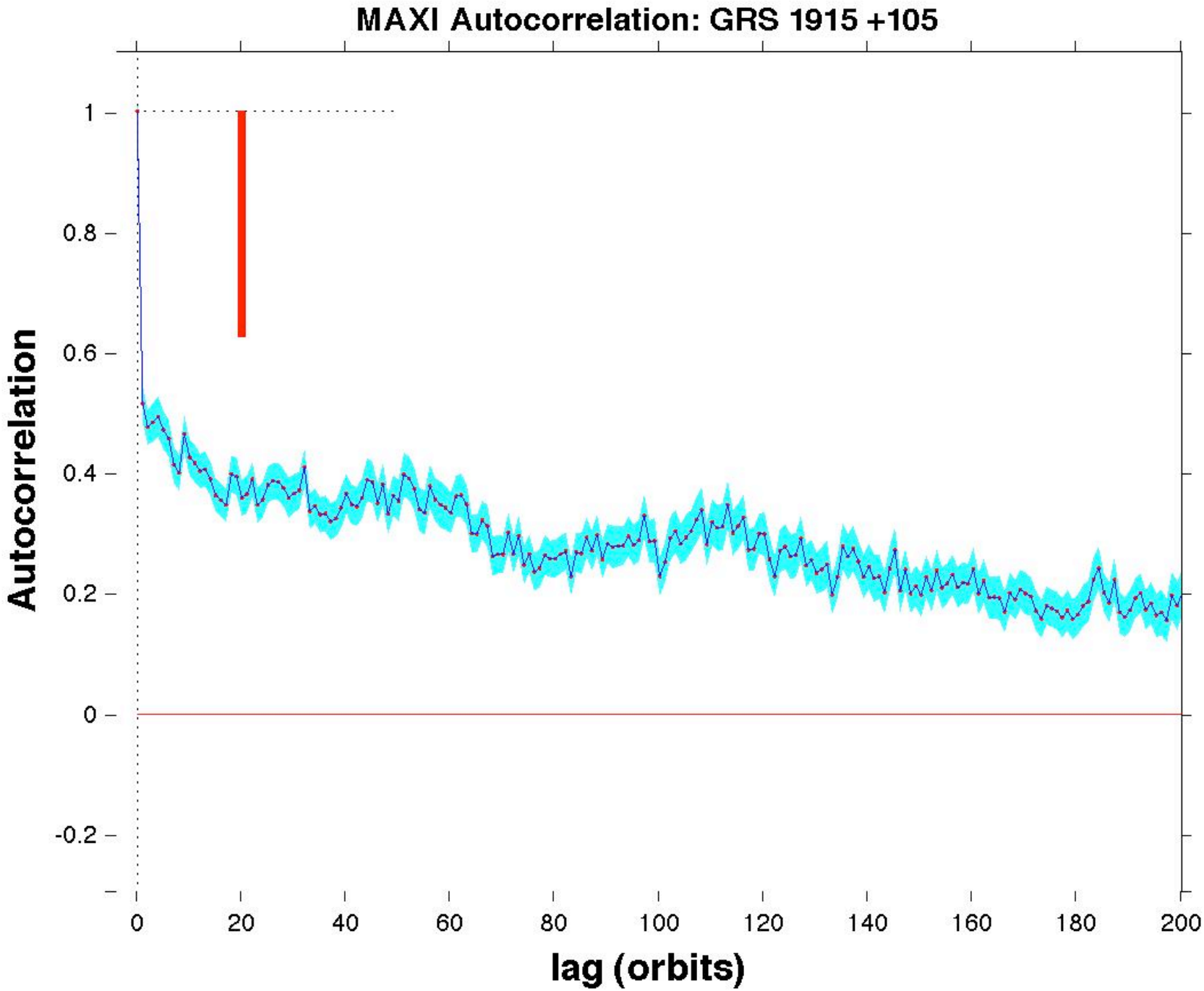

1915 +105
Bayesian Blocks representation
MAXI
time (orbits)
0
1
2
3
4
5
6
7
8
0
1000
2000
3000
4000
5000
6000
7000

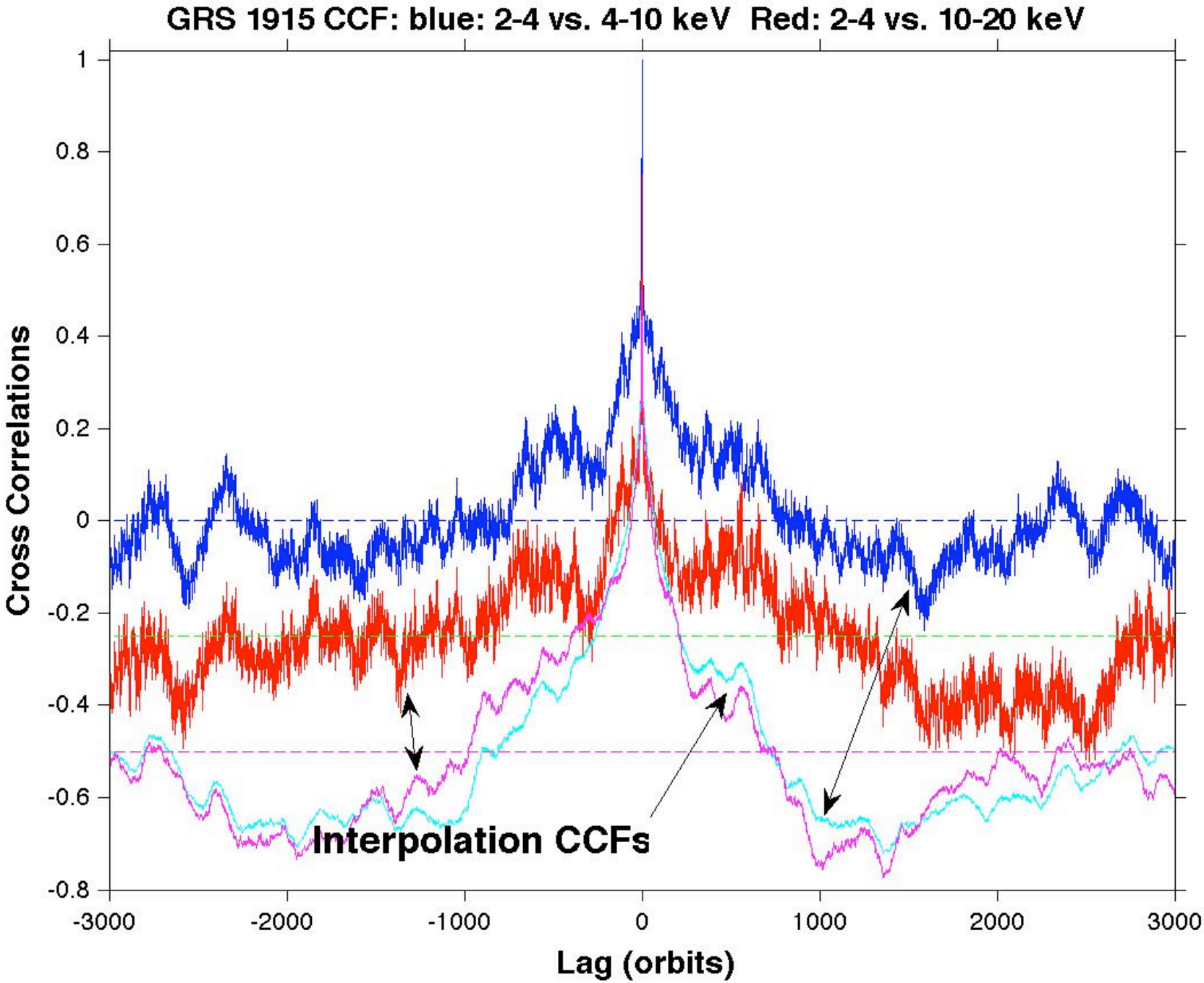

1915 +105
MAXI
ASM
Swift
time (days)
0
50
100
150
200
250
300
350
400
450
2
4
6
8
0.05
0.1
0.15
0.2

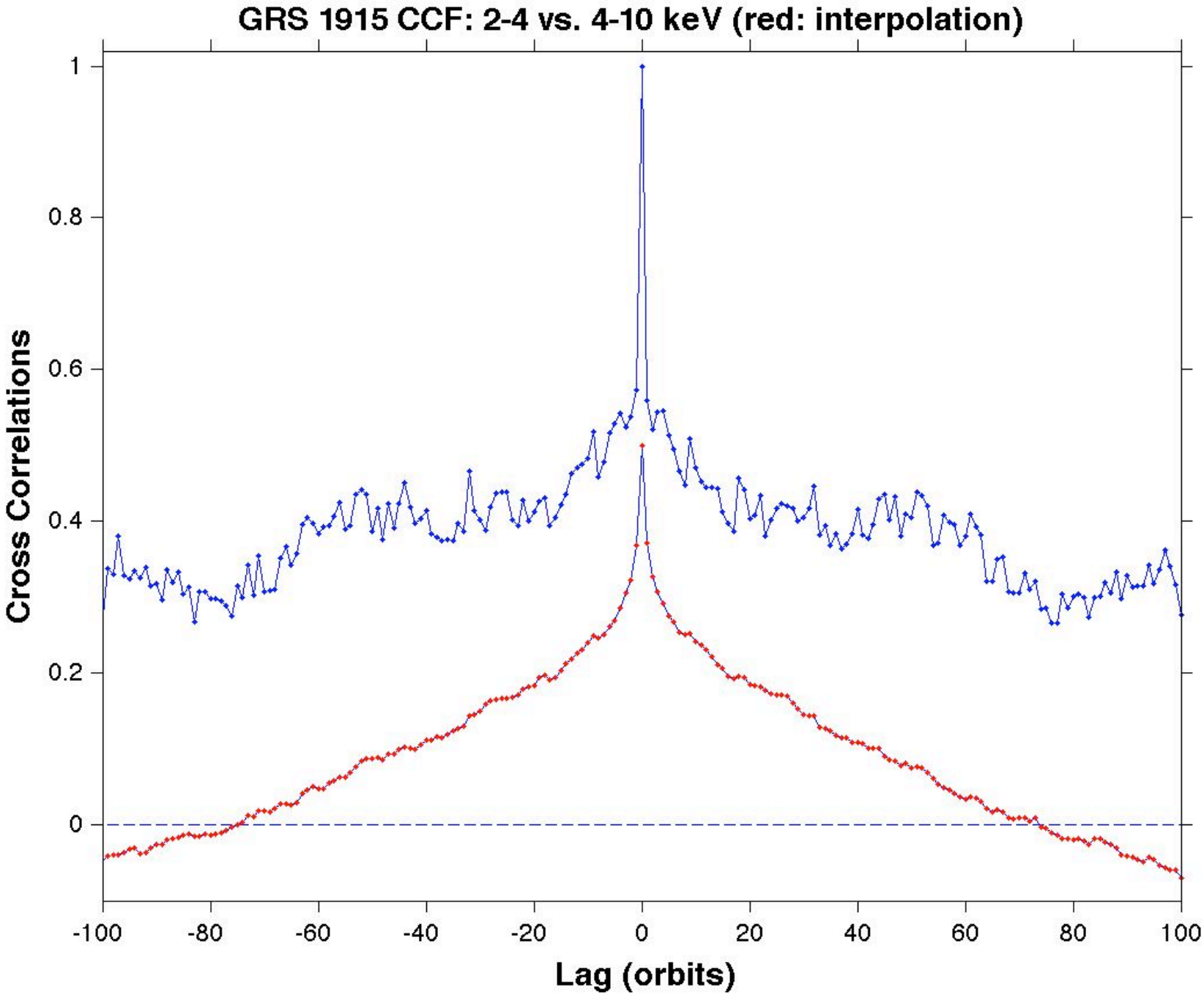
Edelson and Krolik: The Discrete Correlation Function: a New Method for Analyzing Unevenly Sampled Variability Data, Ap. J. 333, 1988, 646 - starting point for all else!

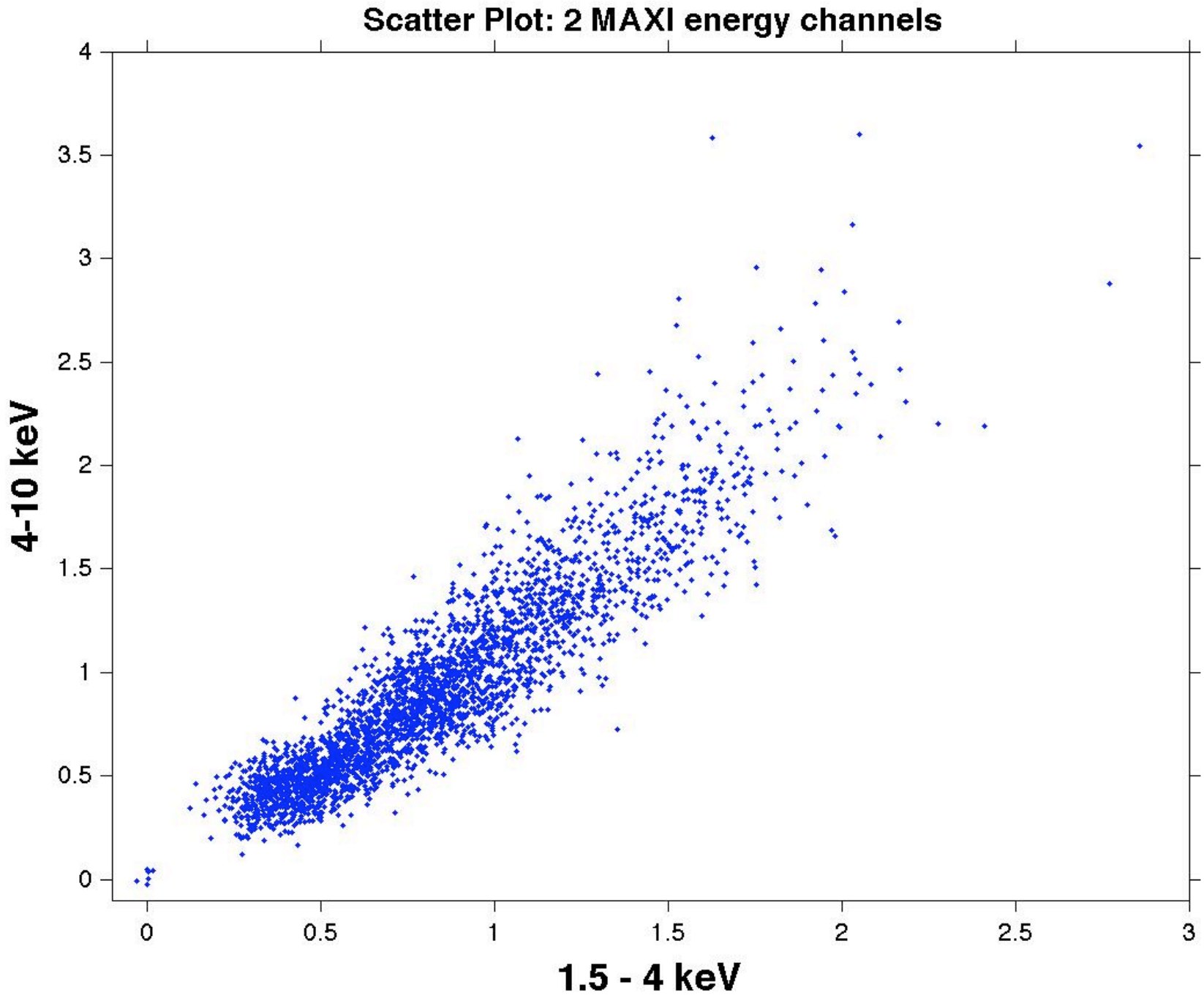


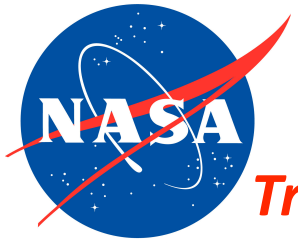

MAXI Autocorrelation: GRS 1915 +105
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
0
500
1000
1500
2000
2500
3000
3500
lag (orbits)


MAXI Autocorrelation: GRS 1915 +105
Autocorrelation
1
0.8
0.6
0.4
0.2
0
-0.2
0
20
40
60
80
100
120
140
160
180
200
lag (orbits)


GRS 1915 CCF: blue: 2-4 vs. 4-10 keV Red: 2-4 vs. 10-20 keV
Cross Correlations
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-3000
-2000
-1000
0
1000
2000
3000
Lag (orbits)
Interpolation CCFs


GRS 1915 CCF: 2-4 vs. 4-10 keV (red: interpolation)
Cross Correlations
1
0.8
0.6
0.4
0.2
0
-100
-80
-60
-40
-20
0
20
40
60
80
100
Lag (orbits)


Scatter Plot: 2 MAXI energy channels
4-10 keV
1.5 - 4 keV
0
0.5
1
1.5
2
2.5
3
3.5
4

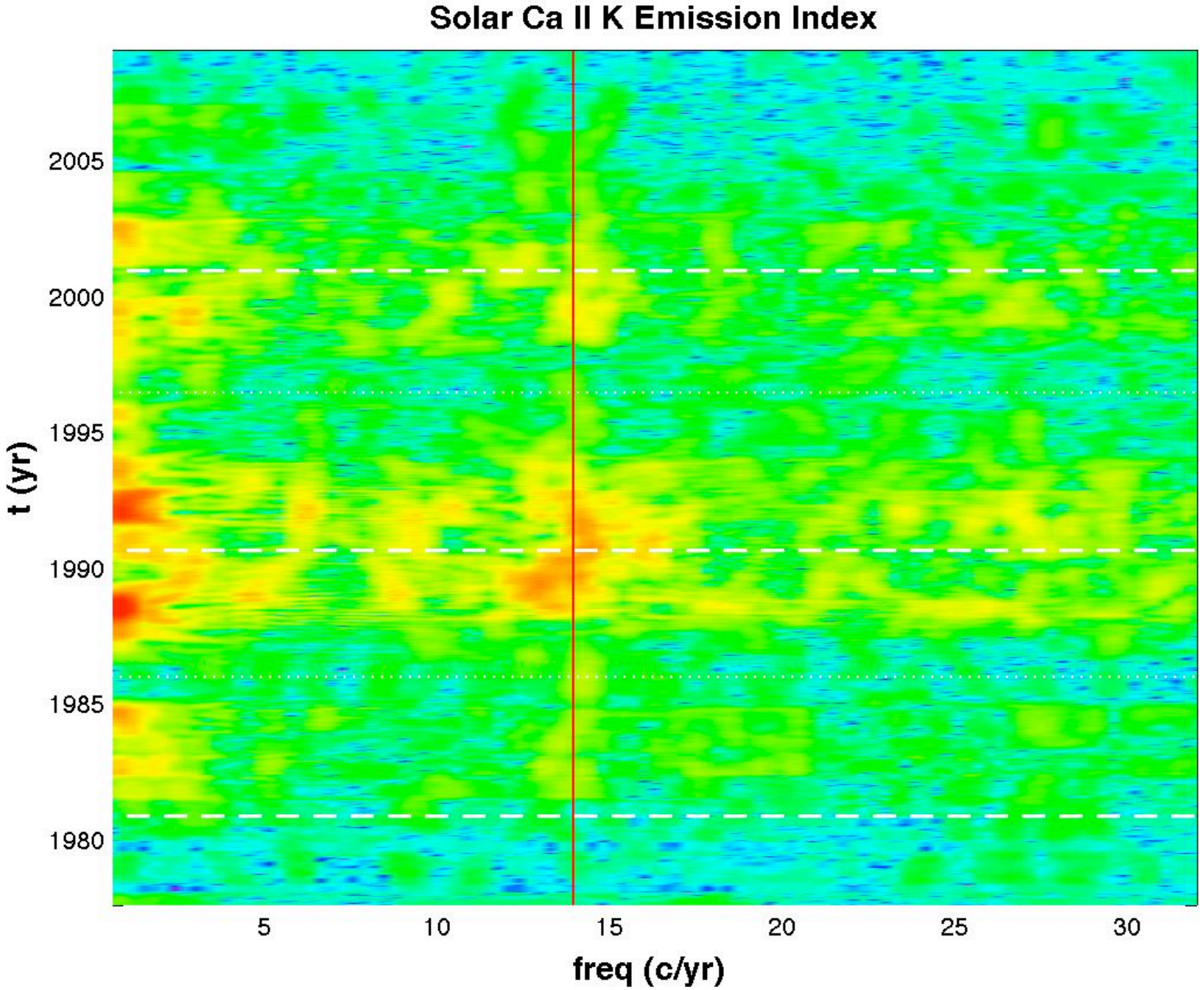


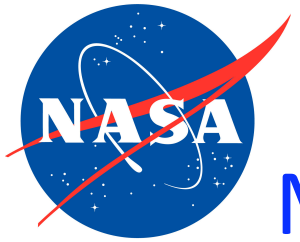
Time-Frequency/Time-Scale Analysis

***Transform to a new view of the time series information*.**

- A Reality in joint time & frequency (or scale) representation
- Atomic decomposition
  - Time-frequency atoms
  - Over-complete representations
  - Optimal Basis Pursuit (Mallat), etc.
- Uncertainty Principle: T-F resolution tradeoff
- Non-stationary processes
  - Flares
  - Trends and Modulations
  - Statistical change-points
- Instantaneous Frequency
- Local vs. Global structure
- Interference (cross-terms in bi-linear representation)

Time-Frequency/Time-Scale Analysis (Temps-Fréquence) Patrick Flandrin http://perso.ens-lyon.fr/patrick.flandrin/publis.html; A Wavelet tour of Signal Processing (Une Exploration des Signaux en Ondelettes) Stéphane Mallat

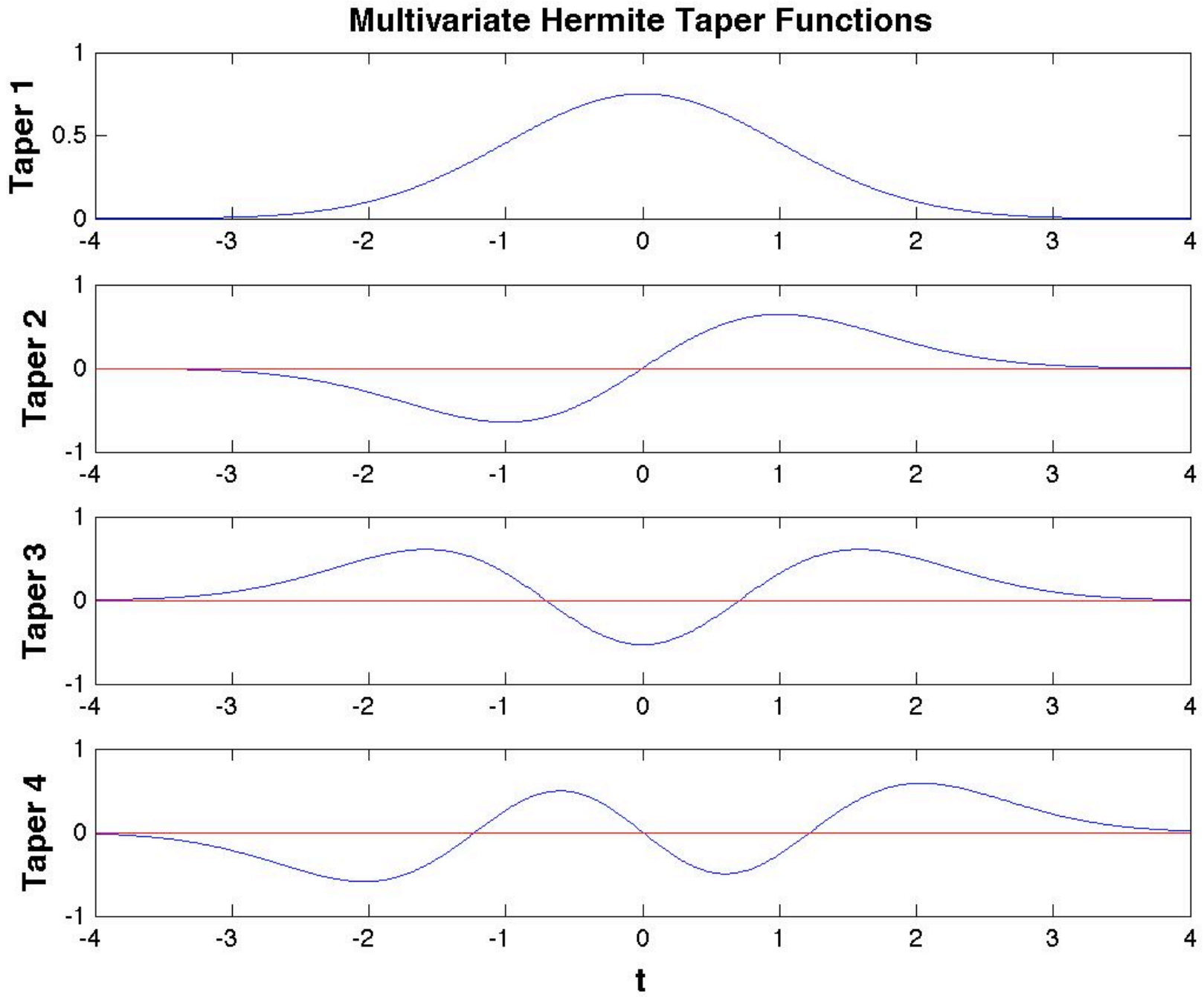

Solar Ca II K Emission Index
2005
2000
1995
1990
1985
1980
t (yr)
5
10
15
20
25
30
freq (c/yr)

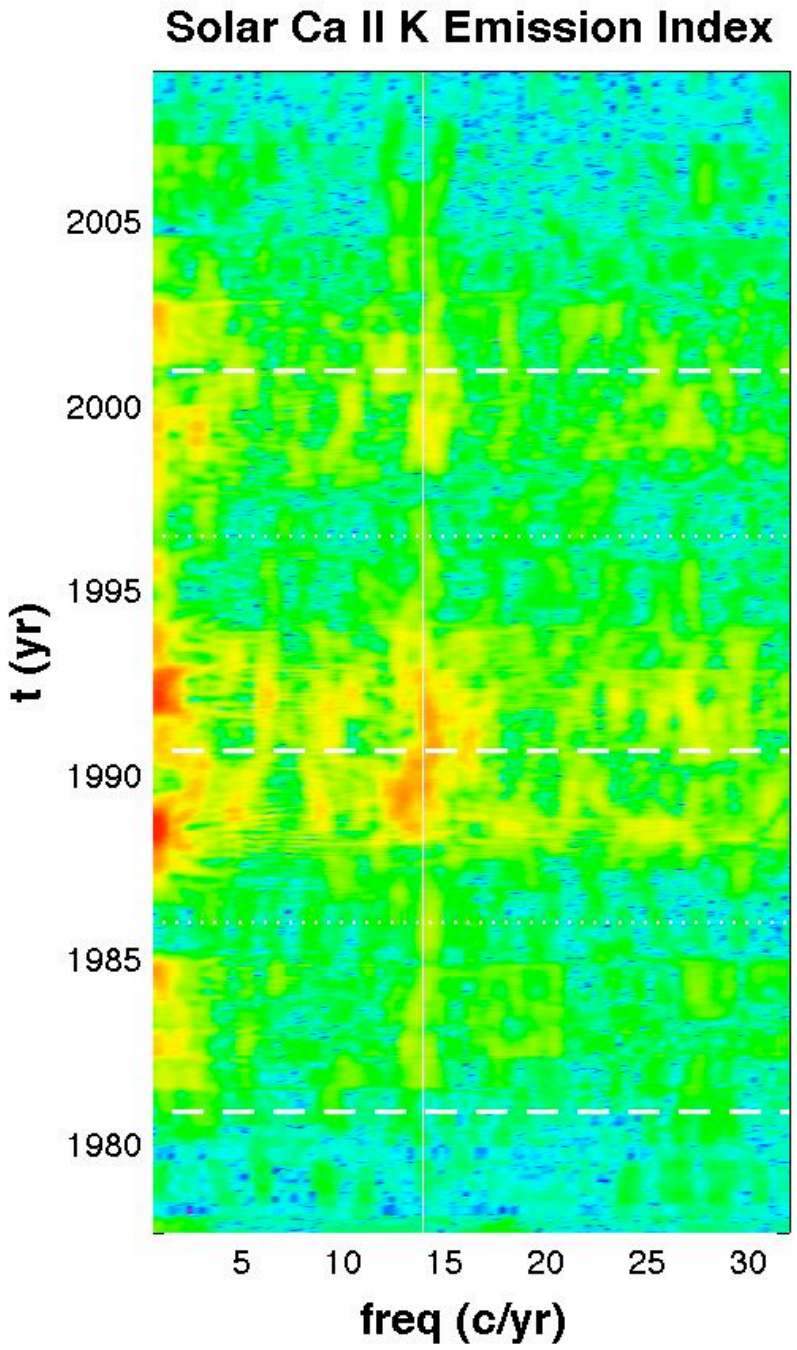


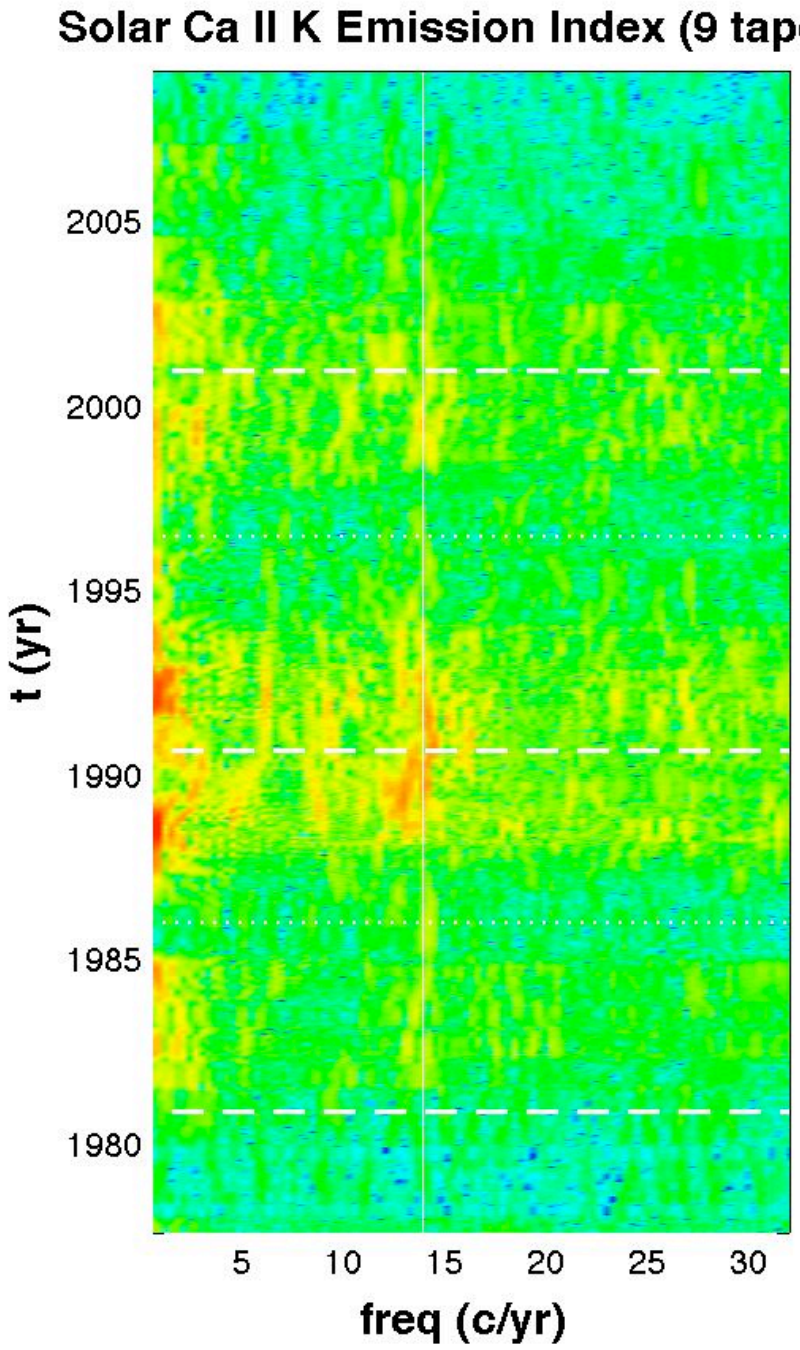
Multi-taper Analysis (Thomson 1982)

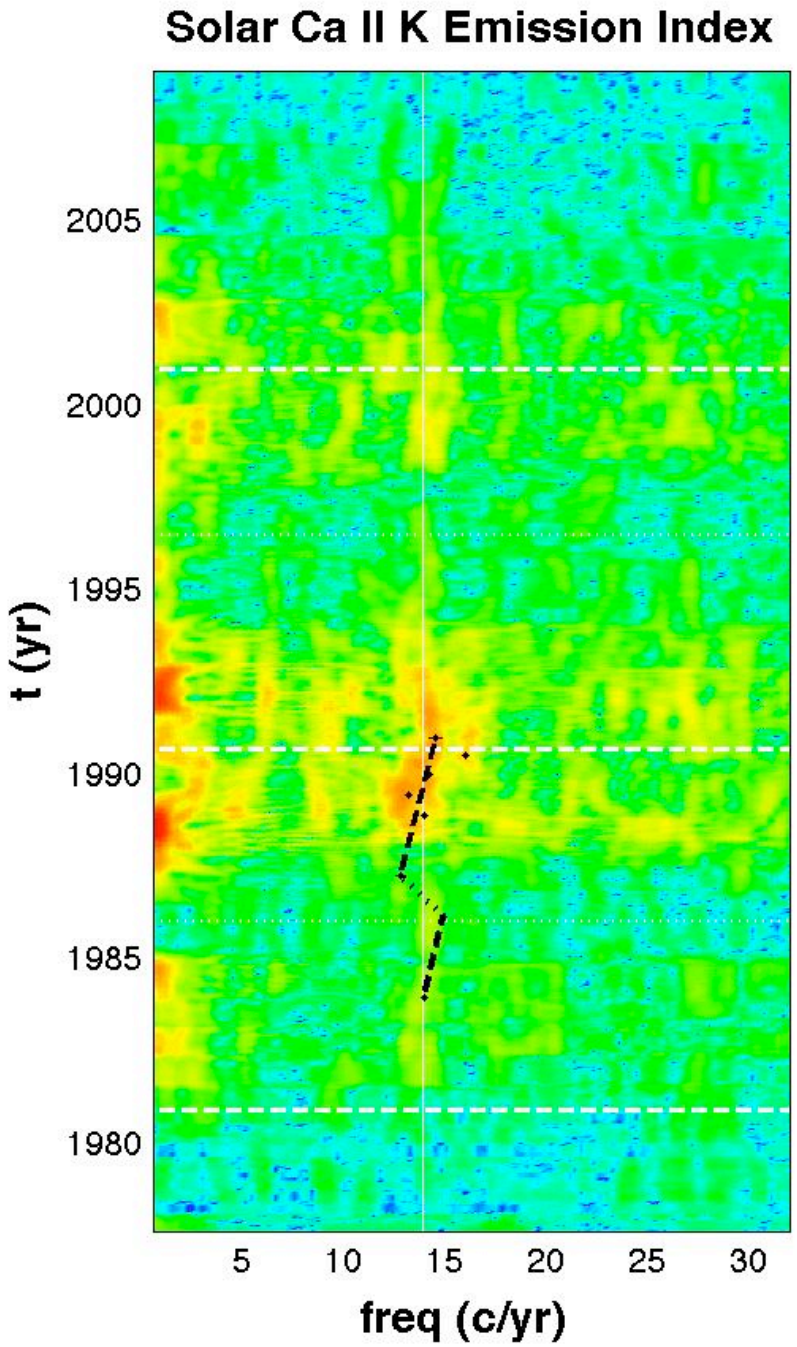
- Tapers (windows) reduce sidelobe leakage = bias
- Incomplete use of data → loss of information
- Multitapers recover this information
- Leakage minimization = eigenvalue problem
  - Eigenfunctions: efficient window functions
  - Eigenvalues
    - measure effectiveness
    - determine how many terms to include

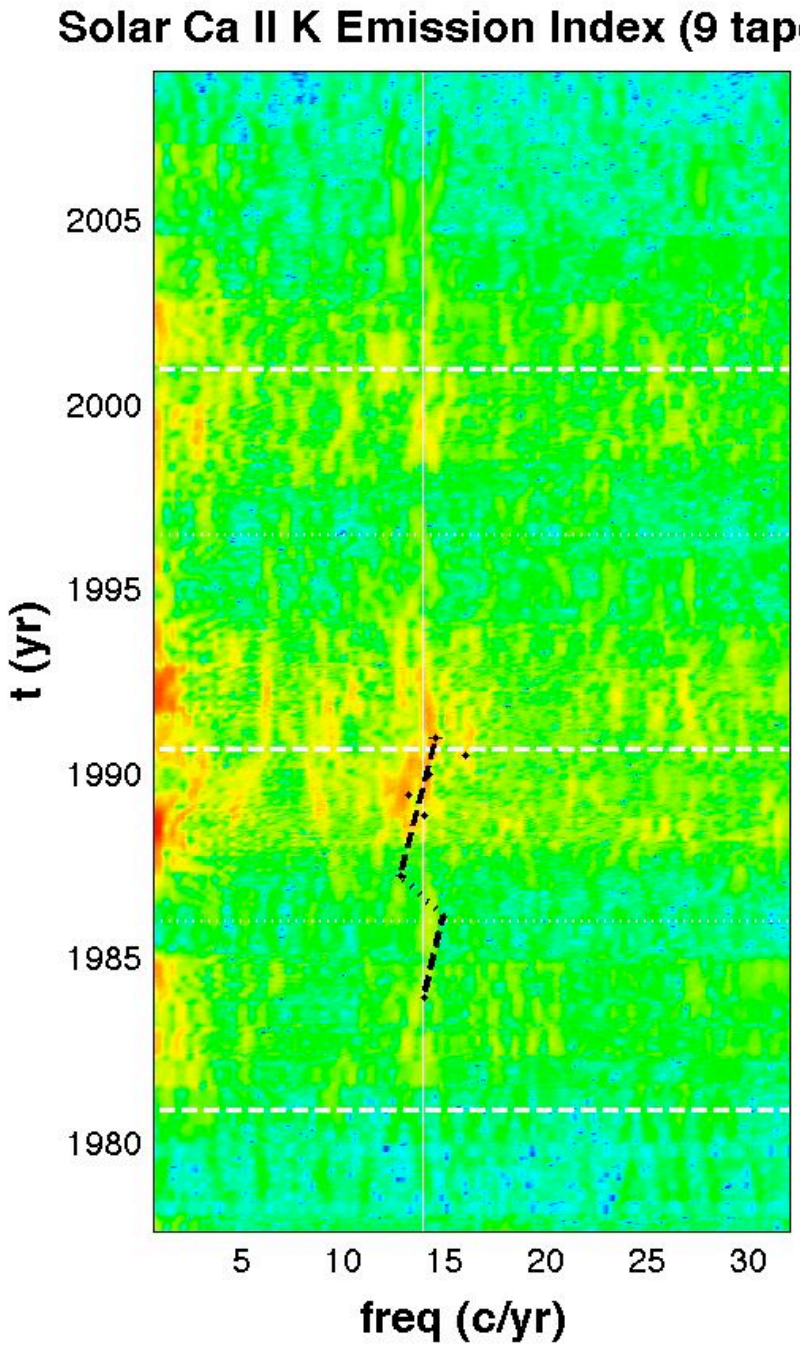
Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques, Don Percival and Andrew Walden (1993)


Multivariate Hermite Taper Functions
Taper 1
Taper 2
Taper 3
Taper 4
t

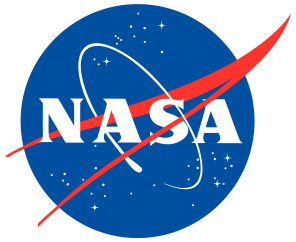

Solar Ca II K Emission Index
2005
2000
1995
1990
1985
1980
t (yr)
5
10
15
20
25
30
freq (c/yr)


Solar Ca II K Emission Index (9 tapers)
2005
2000
1995
1990
1985
1980
t (yr)
5
10
15
20
25
30
freq (c/yr)


Solar Ca II K Emission Index
2005
2000
1995
1990
1985
1980
t (yr)
5
10
15
20
25
30
freq (c/yr)


Solar Ca II K Emission Index (9 tapers)
2005
2000
1995
1990
1985
1980
t (yr)
5
10
15
20
25
30
freq (c/yr)

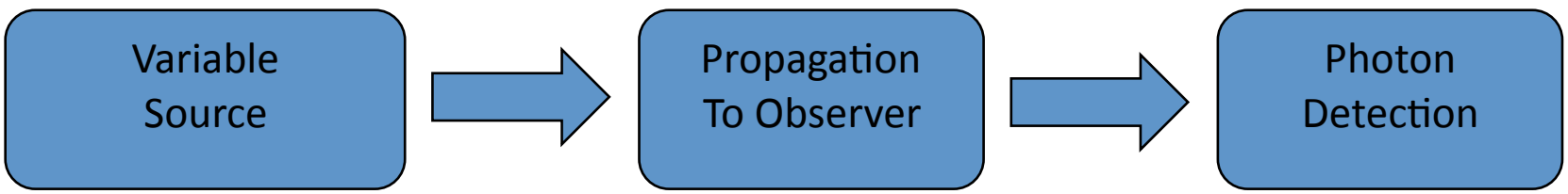
Function	Domain	Range	Auto-	Cross-	Physical Interp
Bayesian blk. Light Curve	Time	Flux	✔	✔ multivar. BB	Flares, events etc.
Scatter Plot	Flux 1	Flux 2		✔	Dependency (not just cor.)
Correlation	Lag	$<X^2>$ <XY>	✔	✔	Correlated behavior/lags
Spectrum	Frequency	Power	✔	✔	Periodicity 1/f noise …
		Phase	✔	✔	Shifts, lags
Structure	Lag	$<X^2>$ <XY>	✔	✔	Correlated behavior/lags
Scalogram	Scale/Time	Power	✔	✔	Dynamic behavior
Scalegram	Scale	Power	✔	✔	1/f noise QPOs
Distribution	Time/scale/ frequency	Power	✔	✔	Dynamic behavior



Practical Suggestions

(somewhat exaggerated)

- Study distribution of sample intervals $dt_n = t_{n+1} - t_n$
- Never subtract mean of time series
- Edelson and Krolik CF is the source of all other analysis
- Use self terms in E&K CF to assess observational errors
- Don't confuse: source randomness/observational noise
- H_0: AGNs are identical stochastic dynamical systems
- Stationarity is a local property
- Any stationary random process is exactly shot noise (random pulses; the Wold Decomposition Theorem)
- Linearity is a physical property, not one of time series
- Do not bin data

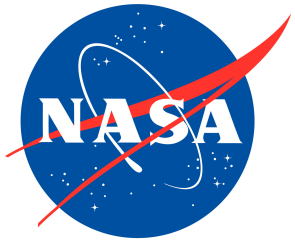


- Luminosity: random or deterministic
- Photon Emission Independent Random Process (Poisson)

- Random Scintillation, Dispersion, etc.?

- Random Detection of Photons (Poisson)

Correlations in source luminosity do not imply correlations in time series data!



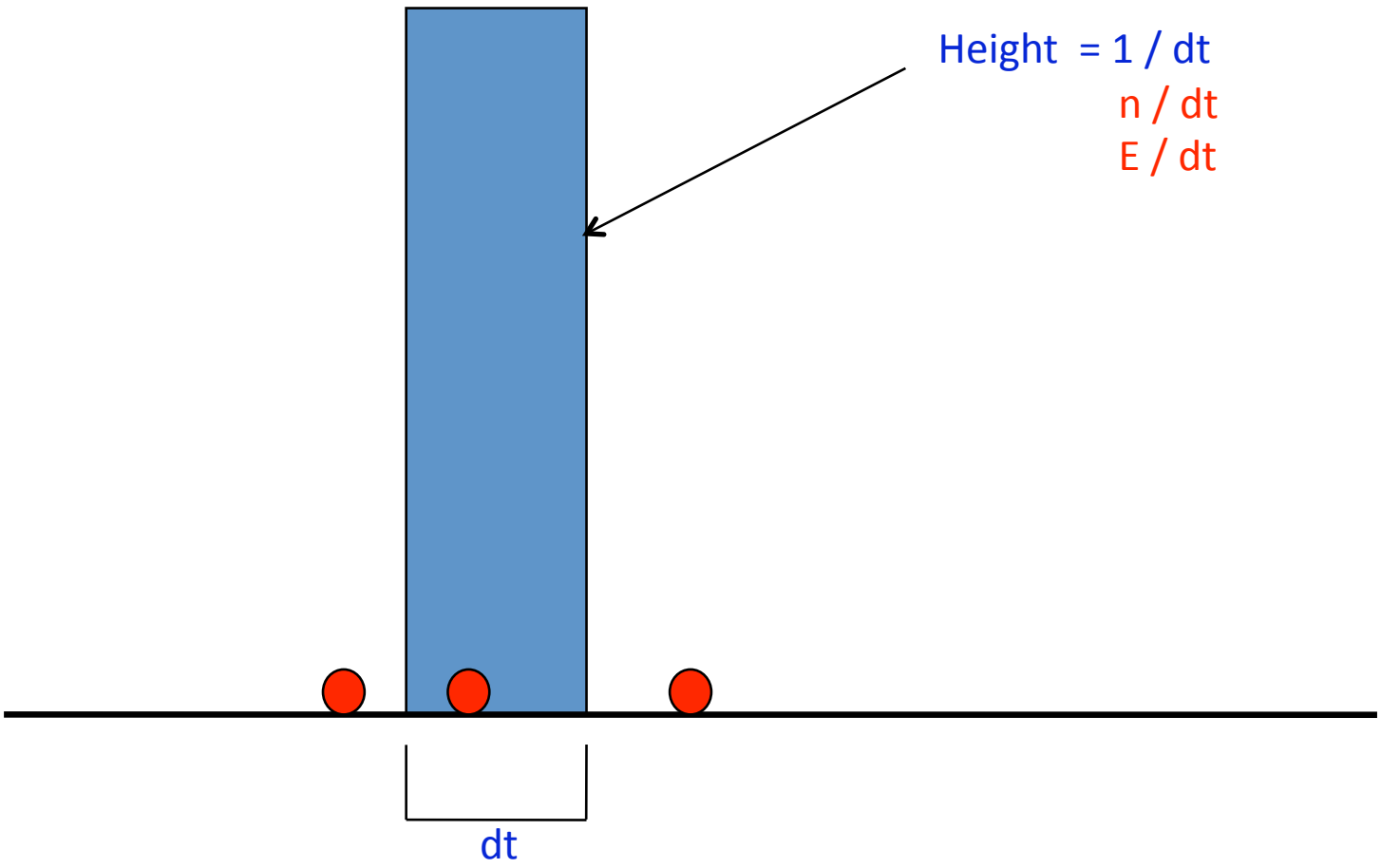
All of this will be in the

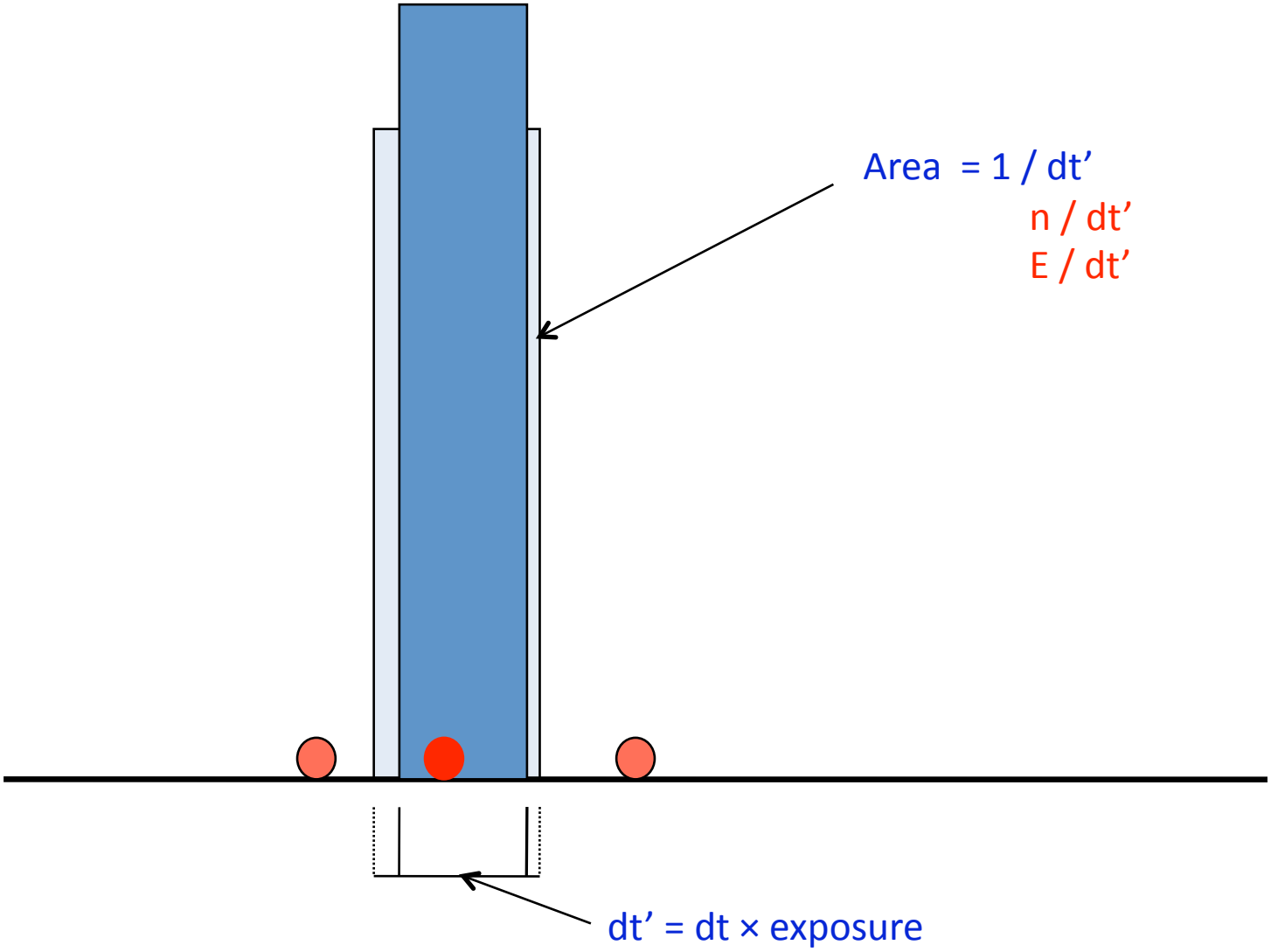
Handbook of Statistical Analysis of Event Data

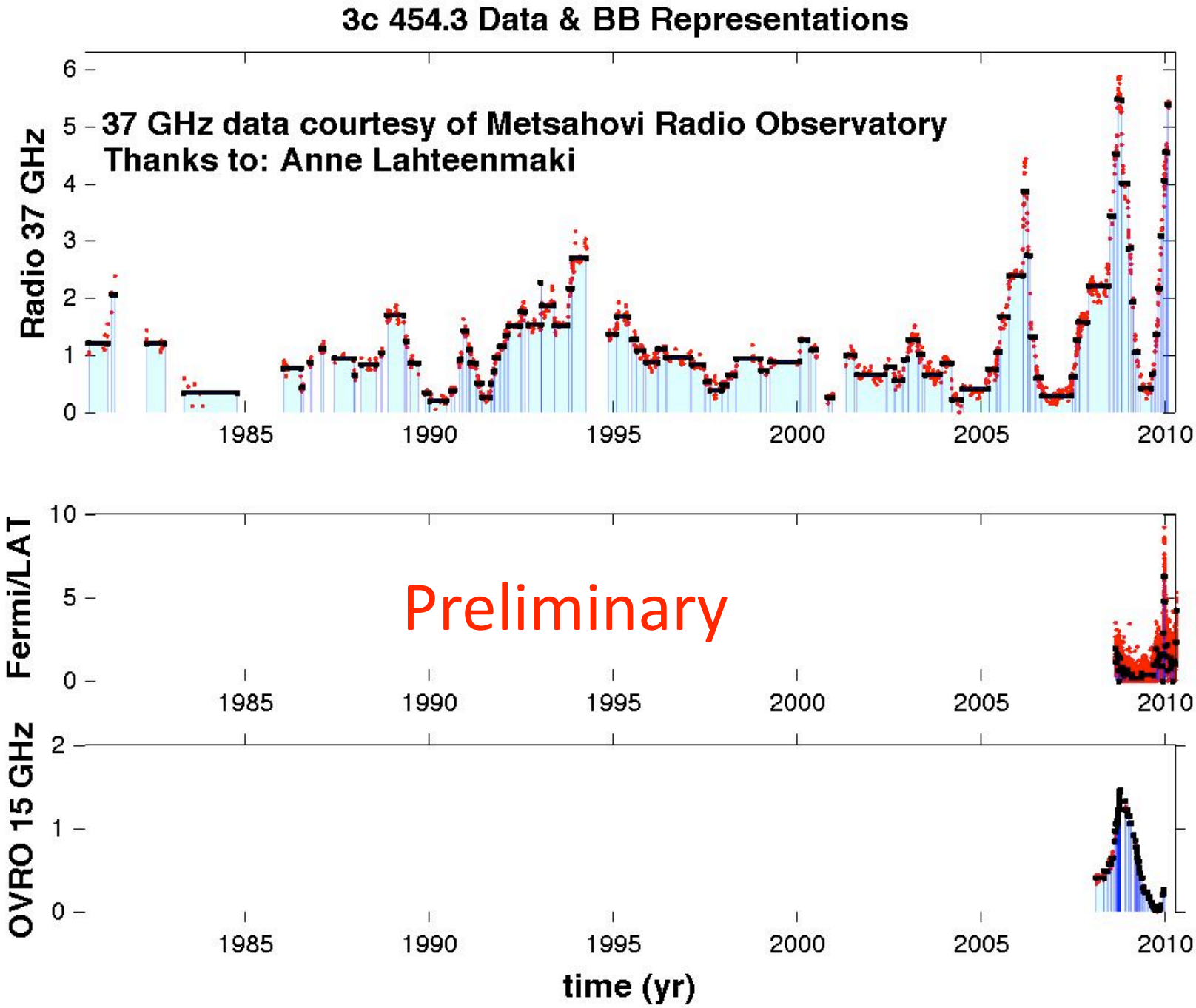
… funded by the NASA AISR Program

MatLab Code
Documentation
Examples
Tutorial

Contributions welcome!

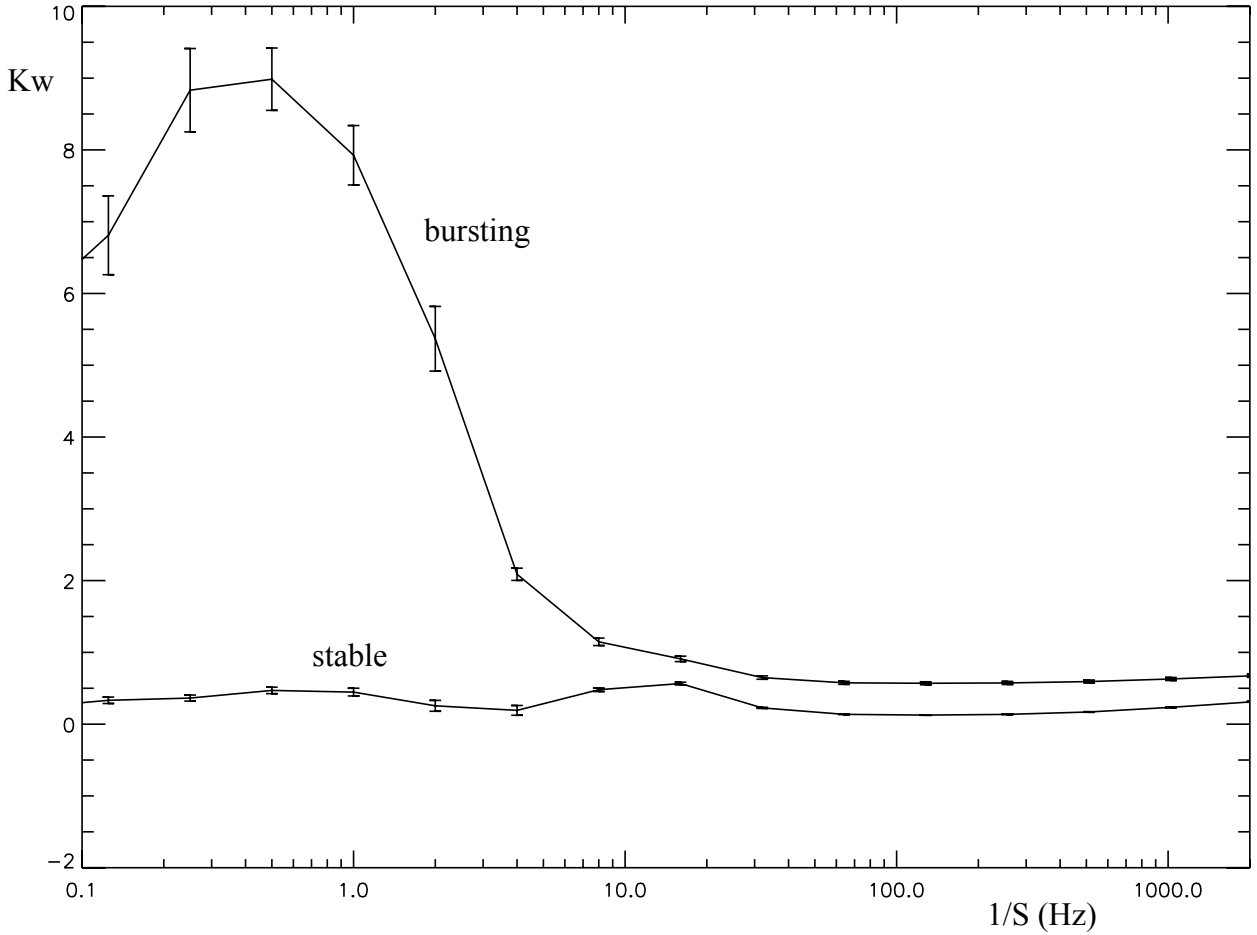

Height = 1 / dt
n / dt
E / dt
dt


Area = 1 / dt'
n / dt'
E / dt'
dt' = dt × exposure

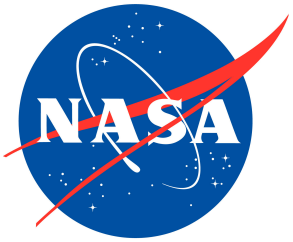

3c 454.3 Data & BB Representations
37 GHz data courtesy of Metsahovi Radio Observatory
Thanks to: Anne Lahteenmaki
Radio 37 GHz
Fermi/LAT
OVRO 15 GHz
Preliminary
time (yr)
1985
1990
1995
2000
2005
2010

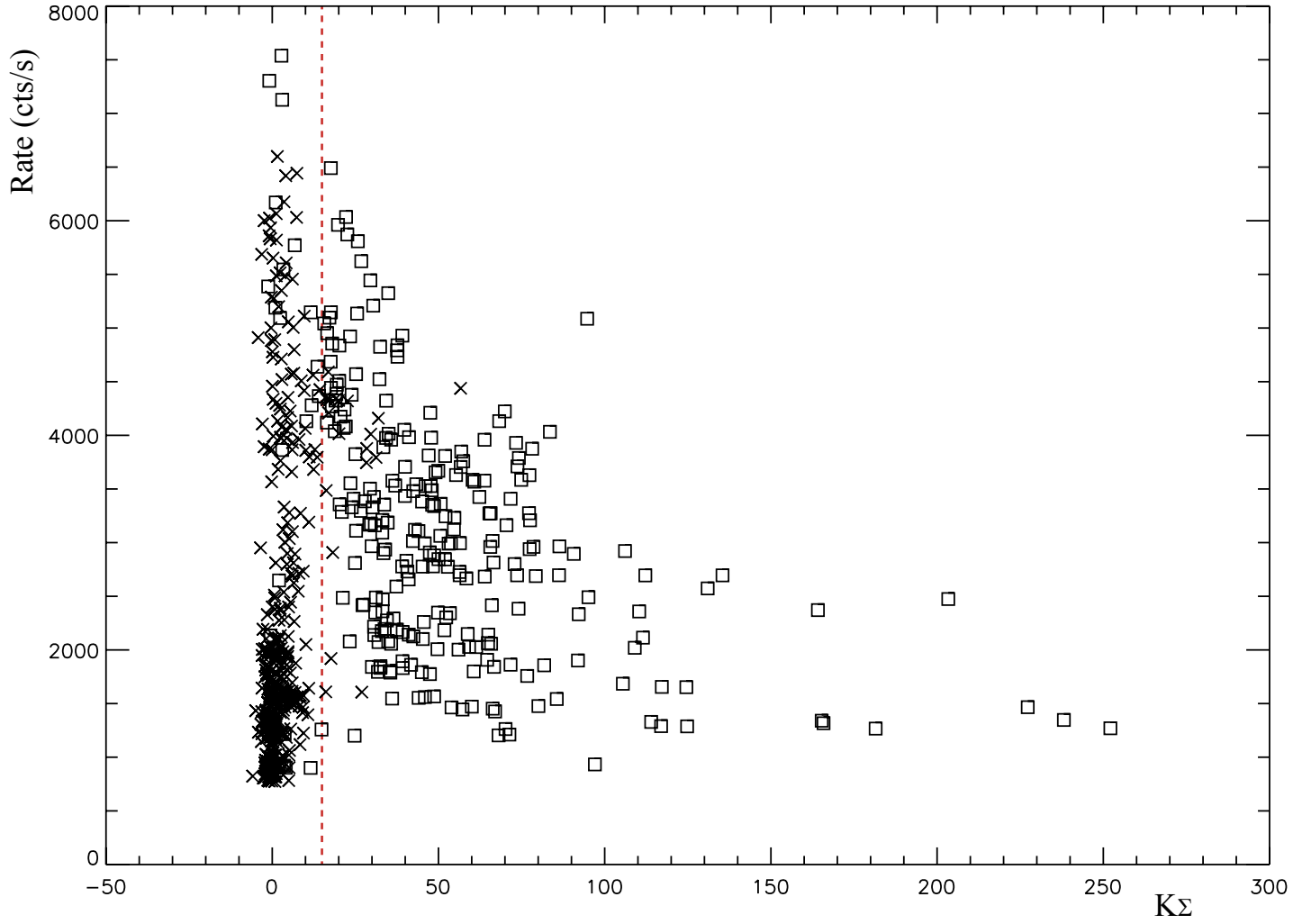
Wavelet Kurtosis

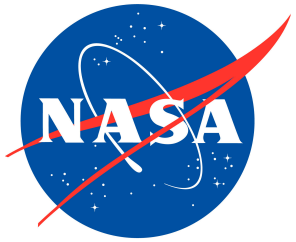
New Statistic to Detect and Characterize Intermittency



Daniel Engavatov, Elliott Bloom, JS; SLAC PhD Thesis



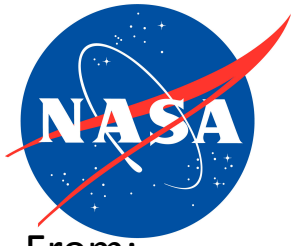

Rate (cts/s)
8000
6000
4000
2000
0
−50
0
50
100
150
200
250
300
Kσ



Stationarity vs. Non-Stationarity

- Formal definition requires infinite amount of data
- Local stationarity depends on scale
- Construct stationarity measure S[x(t)]
  - E.g. variance of TF distribution vs. time marginal
  - Any such measure has statistical fluctuations
  - Simulate surrogate data: scramble Fourier phase
- Construct distribution of S(surrogate data)

Testing Stationarity with Time-Frequency Surrogates, Jun Xiao, Pierre Borgnat, and Patrick Flandrin



From:
Flandrin & Borgnat
"Revisiting and testing stationarity," 2008

… interpreted as "stationary" or "nonstationary" depending on the observation scale …

TL: nonstationary

TR: stationary (periodic)

BL: nonstationary

BR: stationary (homogeneous texture)