Time Series Explorer

Jeff Scargle
NASA Ames Research Center
Jeffrey.D.Scargle@nasa.gov

Thanks: Brad Jackson, Jay Norris, Jim Chiang, Tom Loredo
George Djorgovski, Ashish Mahabal

Applied Information Systems Research Program (NASA)
Keck Institute for Space Studies (Caltech)
Institute for Pure and Applied Mathematics (UCLA)
Center for Applied Mathematics and Computer Science (SJSU)
Statistical and Mathematical Sciences Institute (Duke)

Data source: Catalina Sky Survey AGN and Blazar data

AstroInformatics 2012:
Redmond, WA, Sep 10-13, 2012

In a nutshell …

- There are many algorithms, old and new,
  for astronomical time series data

- Analysis results can be easily computed and displayed in large data bases

- Automated exploration of time series surveys?

Algorithm Categories …

- Time Domain (local)
  - Nonparametric models
  - Bayesian Blocks (piecewise constant, linear, or exponential)
  - Symbolic Representation
  - Parametric model fitting

- Frequency/Scale Domain (global)
  - Power Spectrum
  - Correlation Function
  - Structure Function
  - Wavelet Spectrum

- Time-Frequency/Time Scale Domain (hybrid)
  - Time-frequency distribution
  - Wavelet transform

Barriers to Astronomical Time Series Analysis

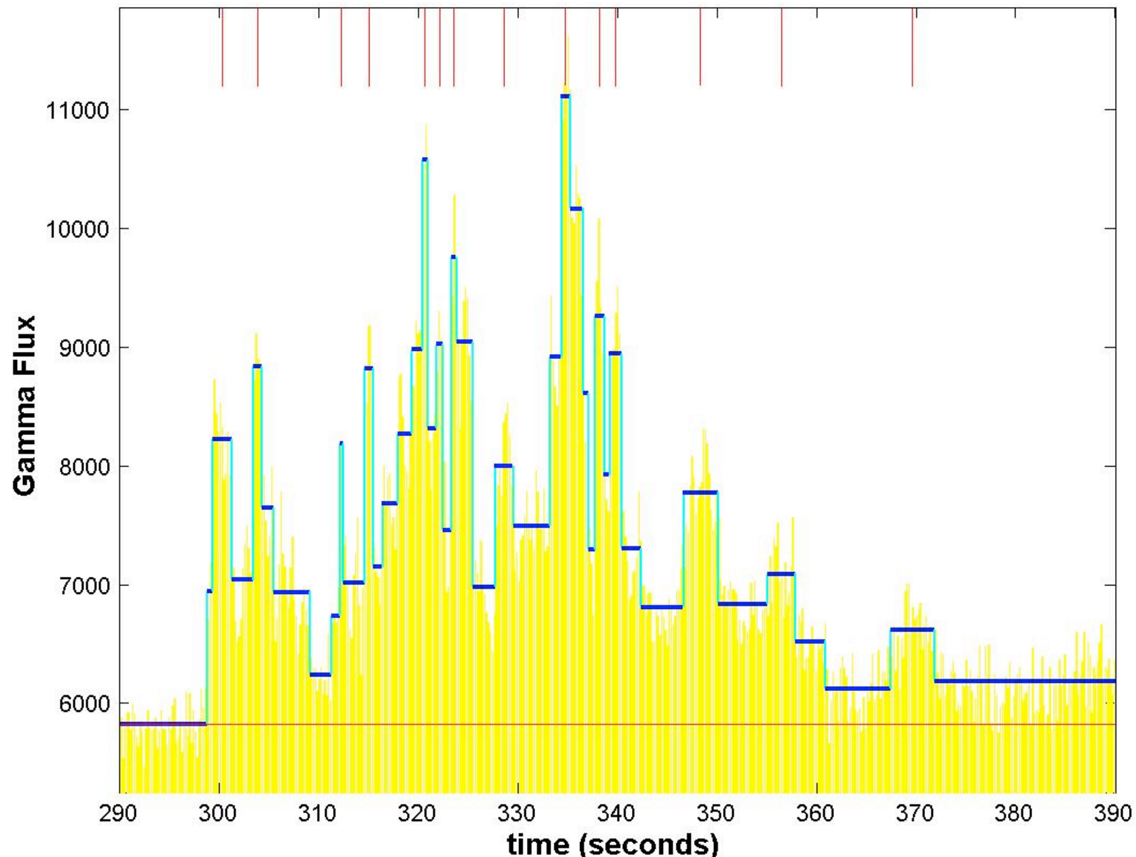
- Uneven Sampling
- Data Gaps
- Observational Errors
- Variable Observational Errors (heteroskedastic)
- Exposure Variation
- Variety of Data Modes:
  - Events (photons)
  - Counts in bins
  - Point measurements …

Barriers to Astronomical Time Series Analysis

- Uneven Sampling
- Data Gaps
- Observational Errors
- Variable Observational Errors (heteroskedastic)
- Exposure Variation
- Variety of Data Modes:
  - Events (photons)
  - Counts in bins
  - Point measurements …

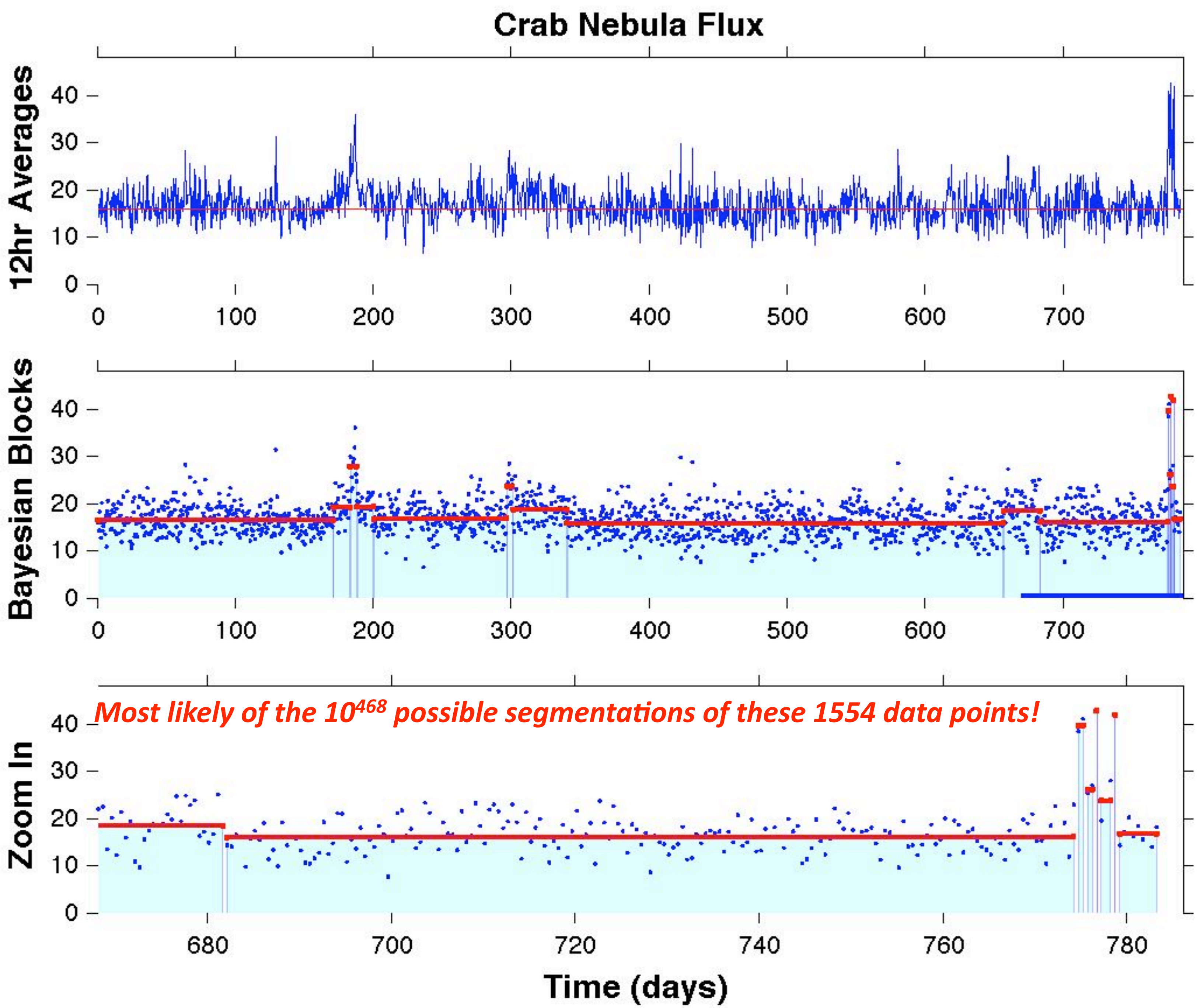
Problem Solved

Studies in Astronomical Time Series Analysis. VI. Bayesian Block Representations
Jeffrey D. Scargle, Jay P. Norris, Brad Jackson, James Chiang arxiv.org/abs/1207.5578
Listed in Astrophysics Source Code Library: http://asci.net



Bellman, R. 1961, On the approximation of curves by line segments using dynamic programming, Communications of the ACM, 4, 284.

15 Years of Reproducible Research in Computational Harmonic Analysis. Donoho, D. et al. 2009, Computing in Science and Engineering, 11, 8
stats.stanford.edu/~donoho/Reports/2008/15YrsReproResch-20080426.pdf



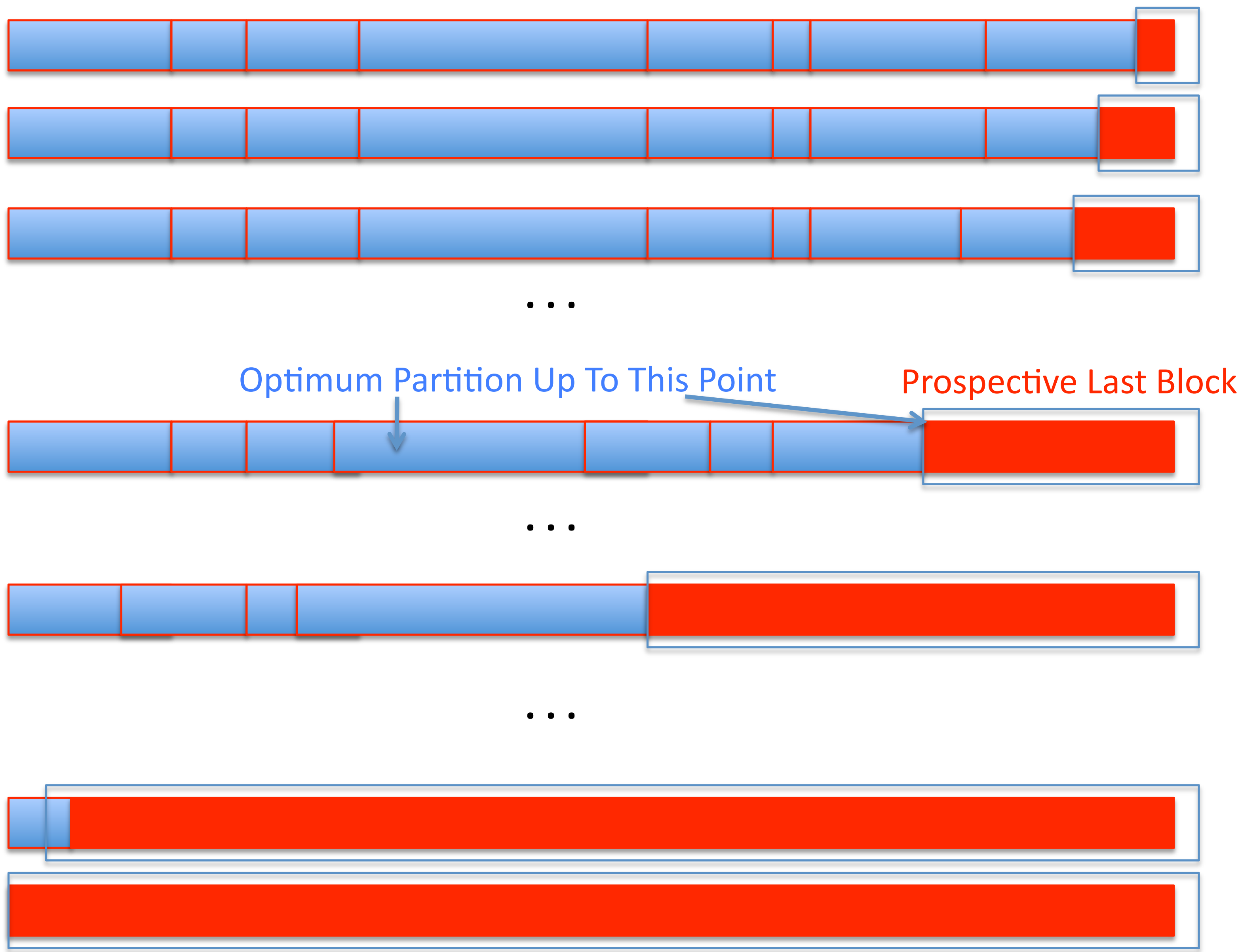
Crab Nebula Flux
12hr Averages
Bayesian Blocks
Zoom In
Most likely of the 10^468 possible segmentations of these 1554 data points!
Time (days)

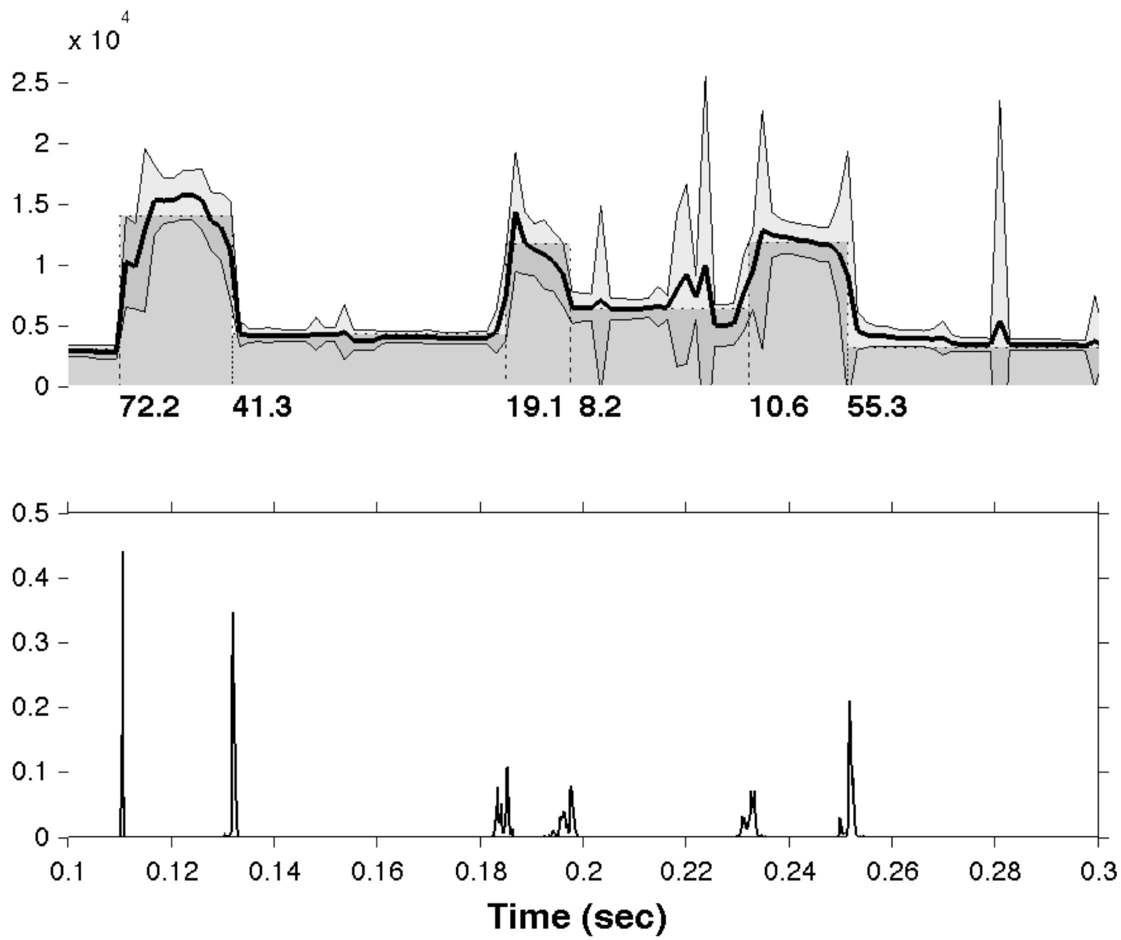
The Optimizer

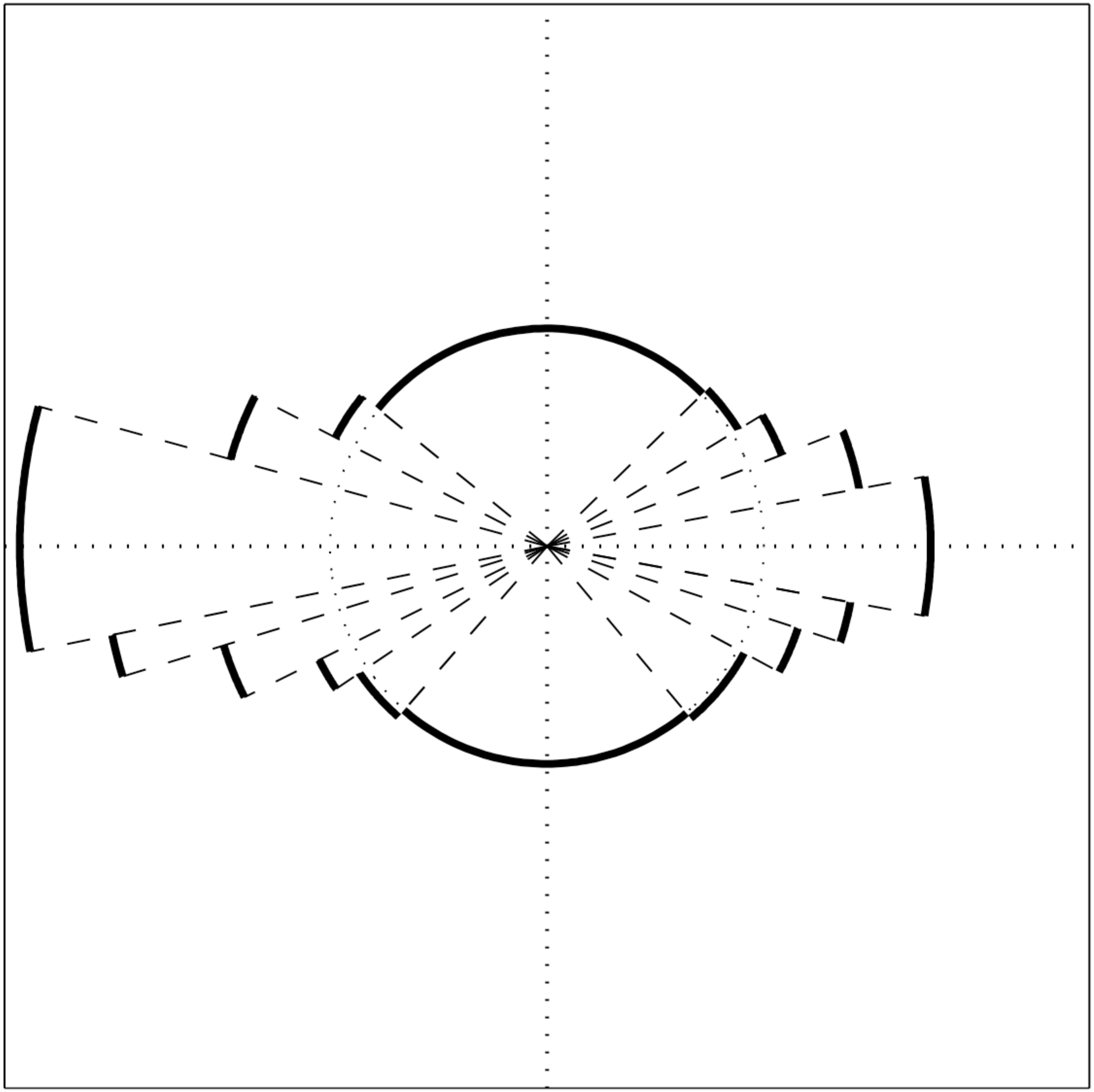
```
best = []; last = [];
for R = 1:num_cells
   [ best(R), last(R) ] = max( [0 best] + ...
          reverse( log_post( cumsum( data_cells(R:-1:1, :) ), prior,  type )  ) );

     if first > 0 & last(R) > first  % Option: trigger on first significant block
          changepoints = last(R); return
     end
end

% Now locate all the changepoints
index = last( num_cells );
changepoints = [];
while index > 1
   changepoints = [ index changepoints ];
    index = last( index - 1 );
end
```

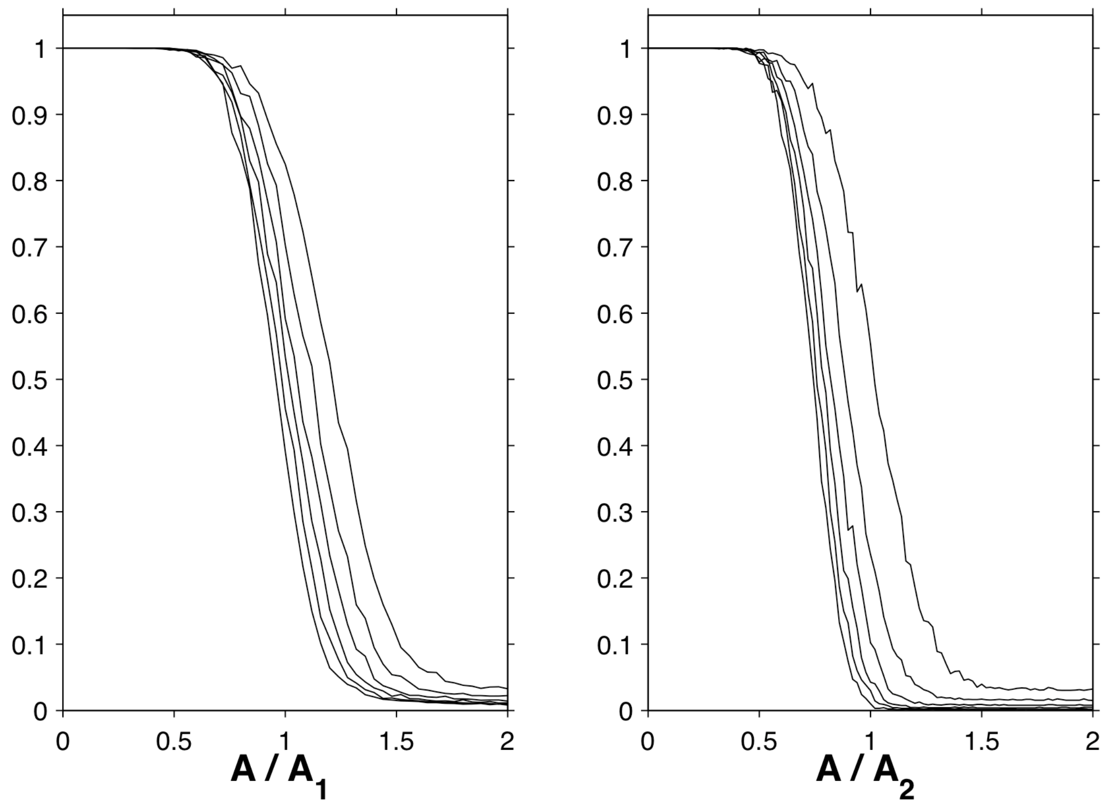

Optimum Partition Up To This Point
Prospective Last Block


x 10^4
2.5
2
1.5
1
0.5
0
72.2
41.3
19.1
8.2
10.6
55.3
0.5
0.4
0.3
0.2
0.1
0
0.1
0.12
0.14
0.16
0.18
0.2
0.22
0.24
0.26
0.28
0.3
Time (sec)



BB Nearly Achieves Theoretical Detection Limit

Detection error rate vs. signal amplitude in units of asymptotic result.



Arias-Castro, E., , Donoho, D., & Huo, X. 2003, Near-Optimal Detection of Geometric Objects by Fast Multiscale Methods
IEEE Transactions on Information Theory, 51, 2402-2425

Cross- and Auto- Correlation Functions for unevenly spaced data

Edelson and Krolik: "The Discrete Correlation Function: a New Method for Analyzing Unevenly Sampled Variability Data"
Astrophysical Journal 333 (1988) 646

$\rho_{xy}(\tau) = (1/N_k) \; \Sigma \, X(t_n) \, Y(t_m) \quad \text{for} \;\; t_n - t_m \text{ in } \tau \text{ bin}$

Data Mode

- Photon events
- Time-to-Spill
- Counts in bins
- Flux measurements
- Any Mode/Sampling!

Universal Time Series Analysis Machine

Auto-

- Correlation Function
- Fourier Power Spectrum
- Fourier Phase Spectrum
- Wavelet scalgram
- Wavelet scaleogram
- Structure Function
- Time-Frequency Distribution
- Time-Scale Distribution
- …

Extension of Edelson & Krolik
Algorithm for Correlation Function
of Unevenly Sampled Data

Jeff Scargle

Data Mode

- Photon events
- Time-to-Spill
- Counts in bins
- Flux measurements
- Any Mode/Sampling!

any combination

- Photon events
- Time-to-Spill
- Counts in bins
- Flux measurements
- Any Mode/Sampling!

Universal Time Series Analysis Machine

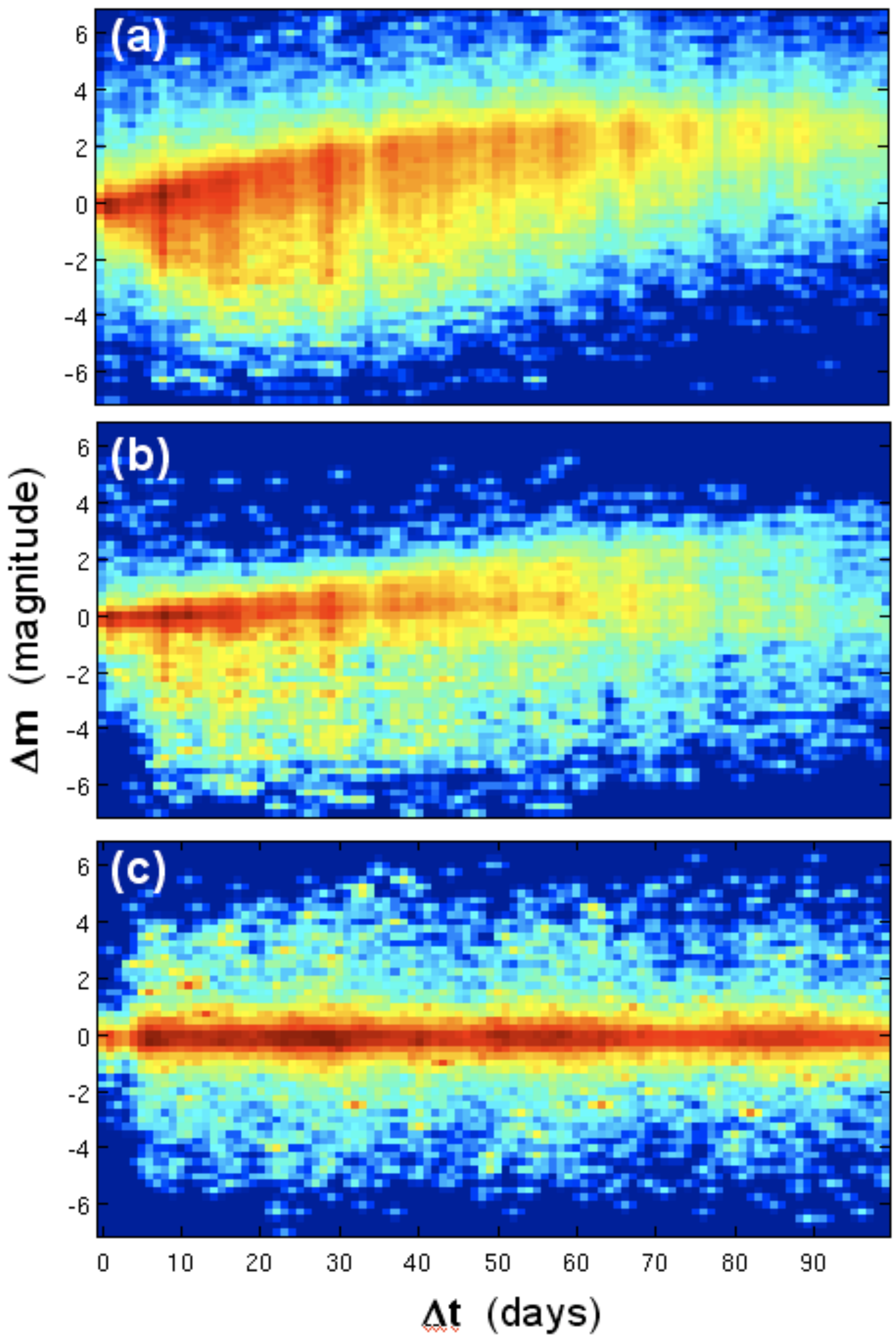
Cross-

- Correlation Function
- Fourier Power Spectrum
- Fourier Phase Spectrum
- Wavelet scalgram
- Wavelet scaleogram
- Structure Function
- Time-Frequency Distribution
- Time-Scale Distribution
- …

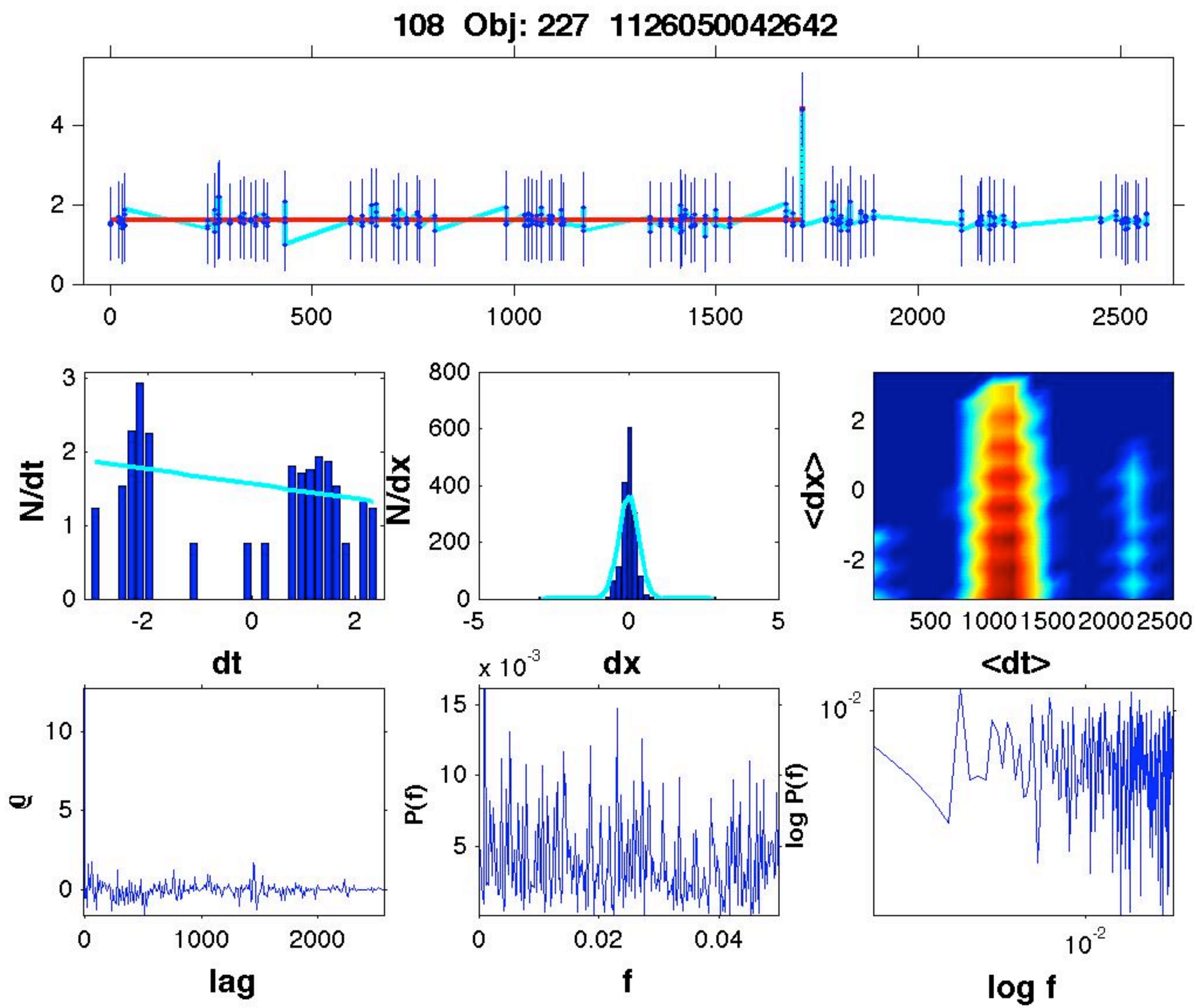
Extension of Edelson & Krolik Algorithm for Correlation Function of Unevenly Sampled Data

Jeff Scargle

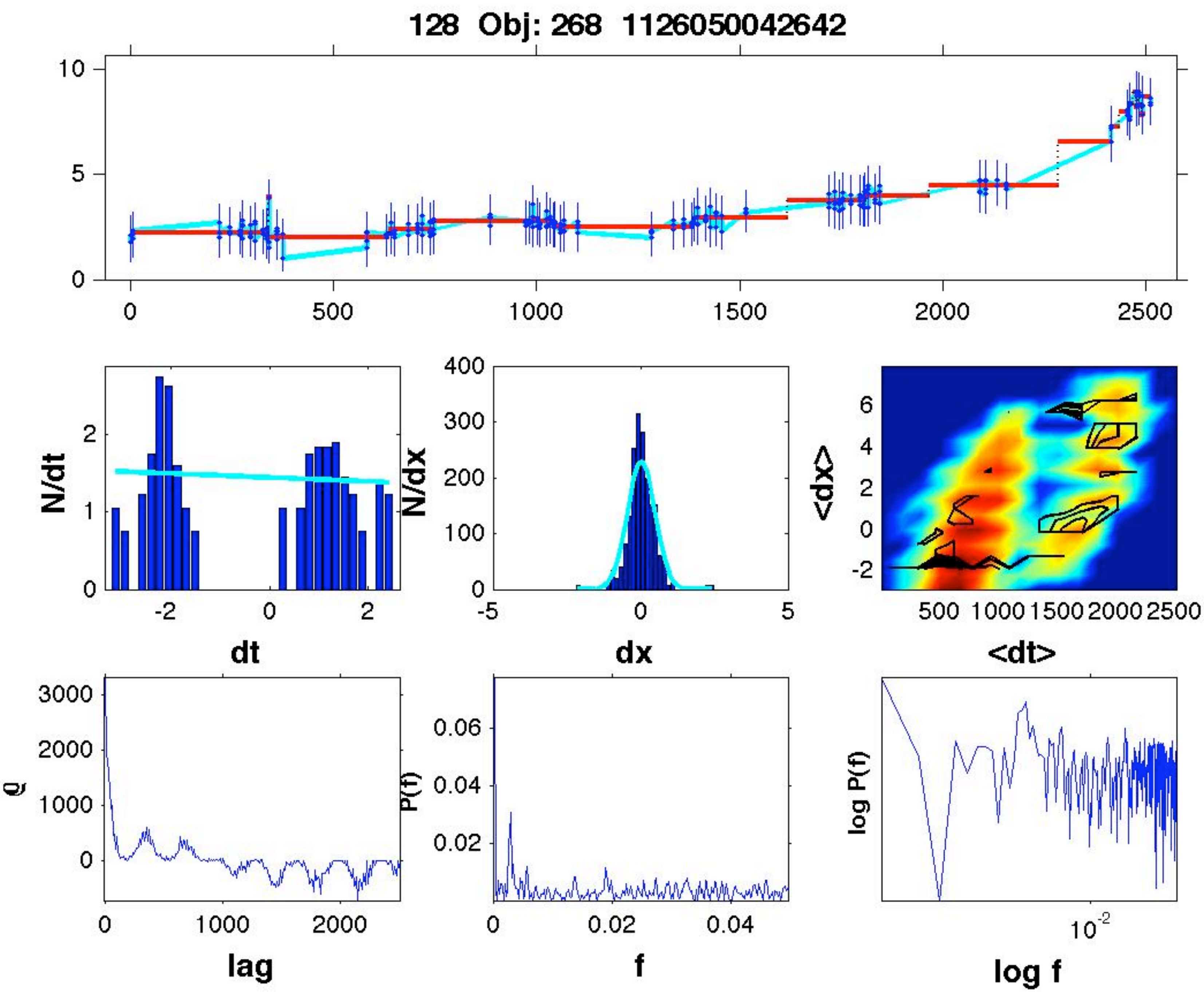
Function	Domain	Range	Auto-	Cross-	Physical Interp
Bayesian blk. Light Curve	Time	Flux	✔	✔ multivar. BB	Flares, events etc.
Scatter Plot	Flux 1	Flux 2		✔	Dependency (not just cor.)
Correlation	Lag	$<X^2>$ $<XY>$	✔	✔	Correlated behavior/lags
Spectrum	Frequency	Power	✔	✔	Periodicity 1/f noise …
		Phase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



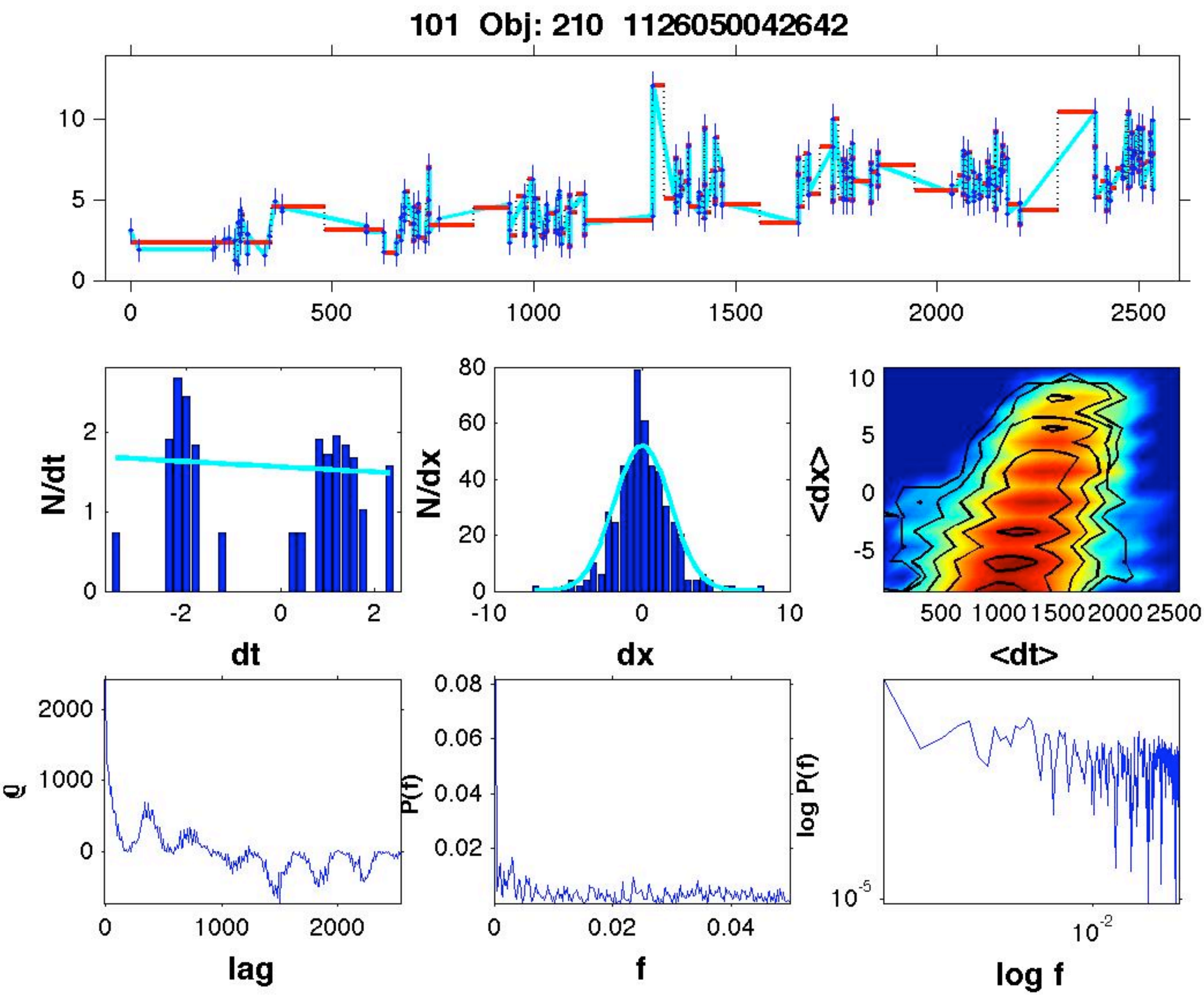
(a)
(b)
(c)
Δm (magnitude)
Δt (days)
6
4
2
0
-2
-4
-6
0
10
20
30
40
50
60
70
80
90



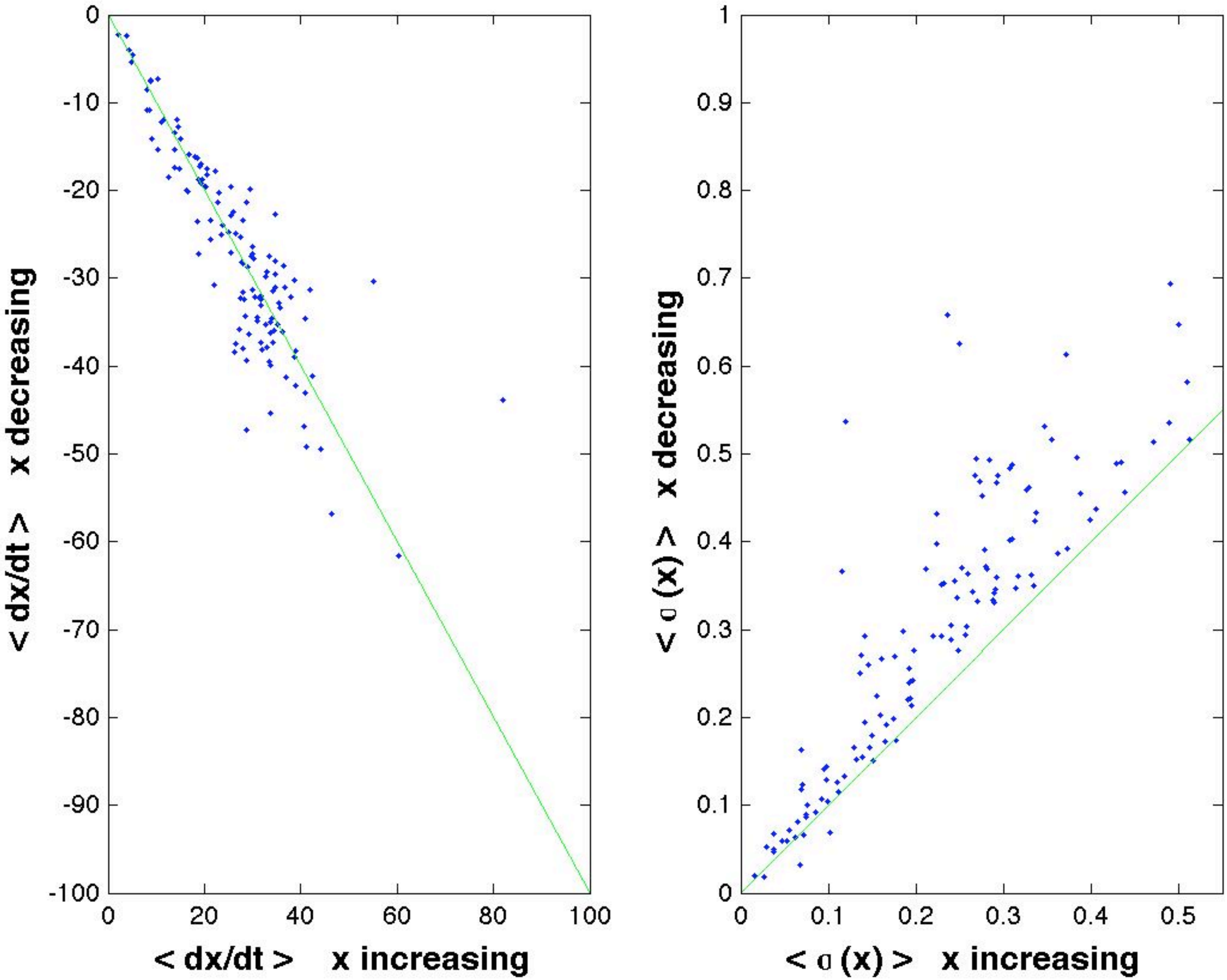
108 Obj: 227 112605004264 2
N/dt
dt
N/dx
dx
<dx>
<dt>
lag
x 10^-3
P(f)
f
log P(f)
log f



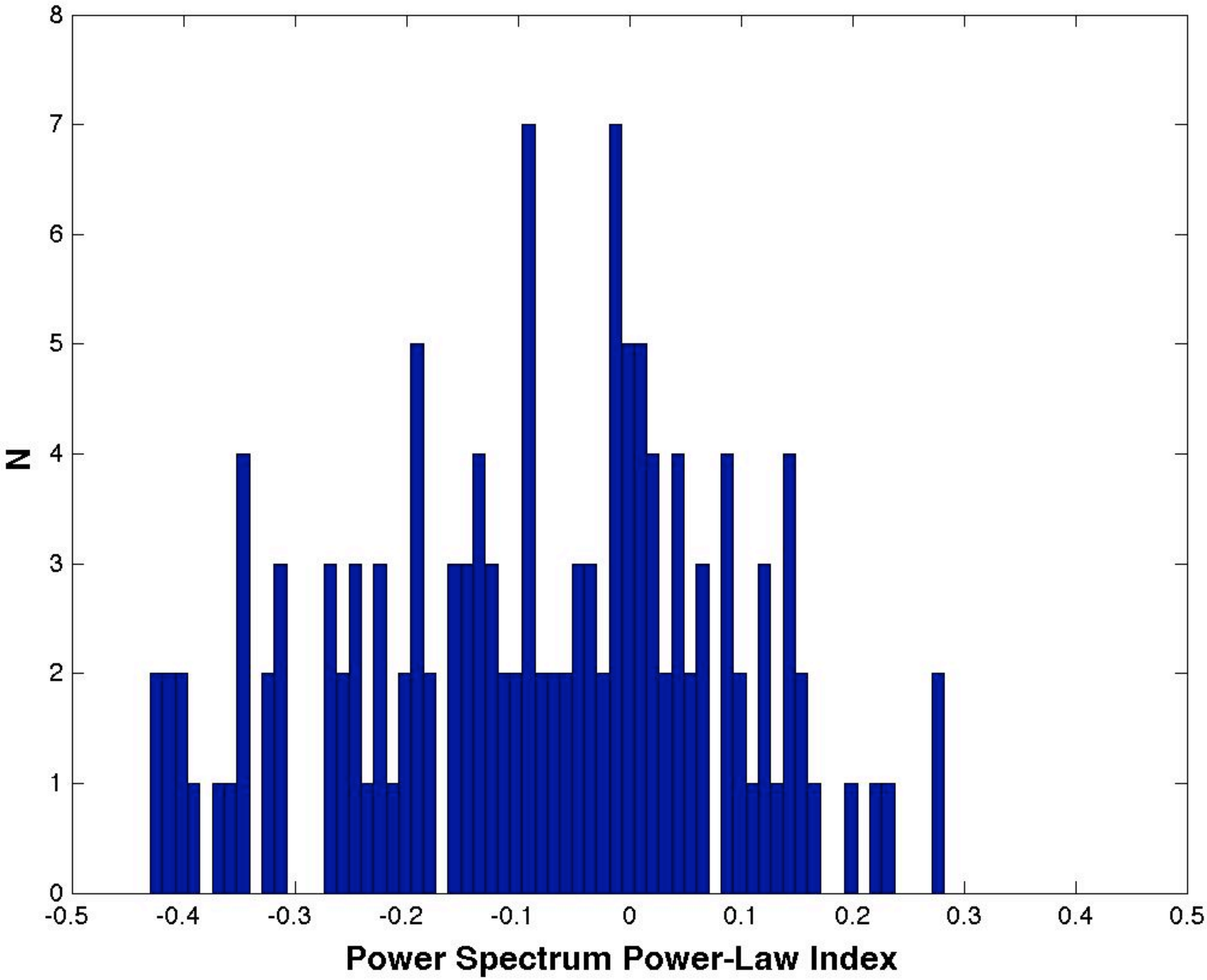
128 Obj: 268 1126050042642
N/dt
dt
N/dx
dx
<dx>
<dt>
lag
P(f)
f
log P(f)
log f

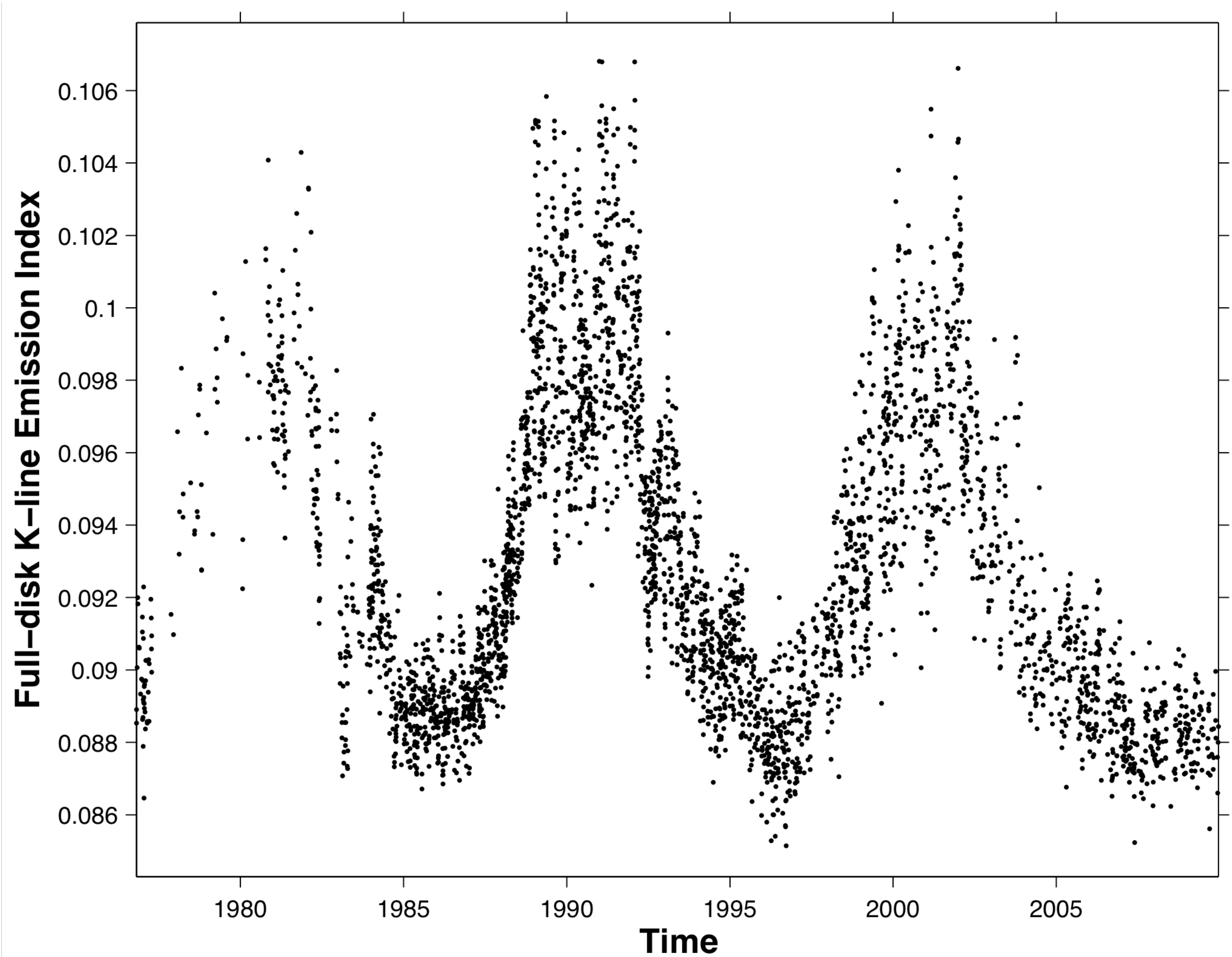


101 Obj: 210 1126050042642
N/dt
dt
N/dx
dx
<dx>
<dt>
lag
P(f)
f
log P(f)
log f

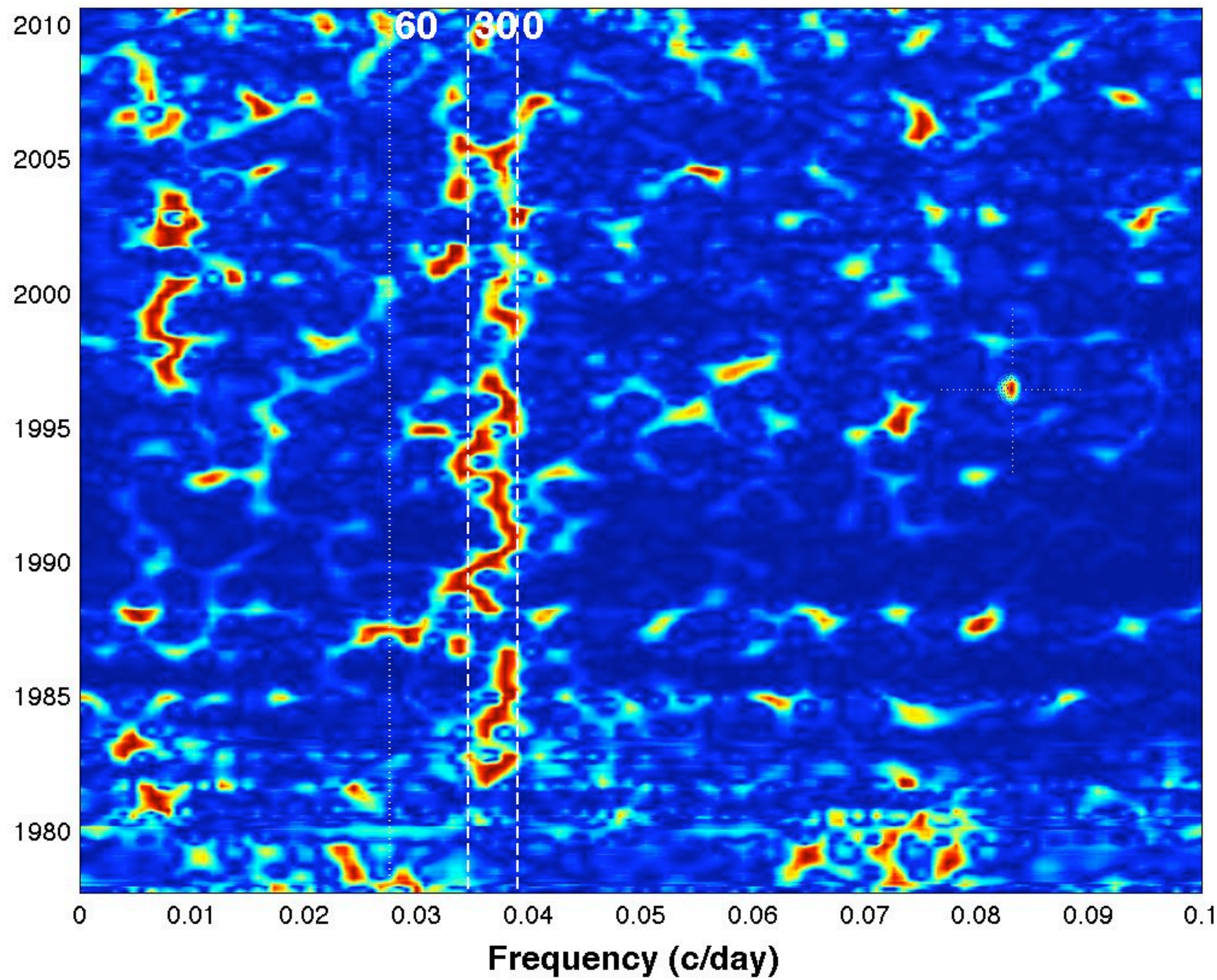


< dx/dt > x decreasing
< dx/dt > x increasing
0
-10
-20
-30
-40
-50
-60
-70
-80
-90
-100
0
20
40
60
80
100
< σ (x) > x decreasing
< σ (x) > x increasing
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5

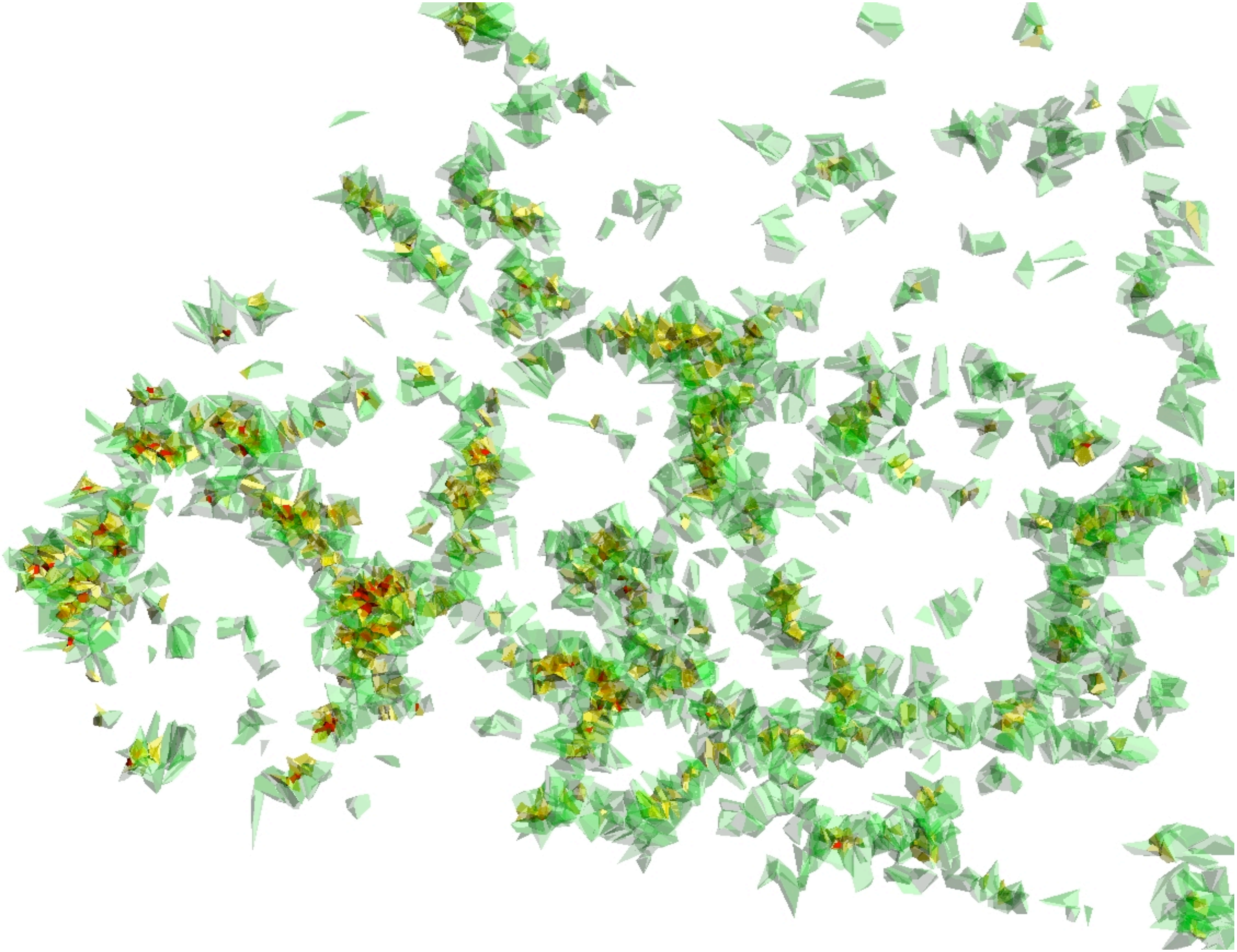

N
Power Spectrum Power-Law Index
0
1
2
3
4
5
6
7
8
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5



Full–disk K–line Emission Index
0.106
0.104
0.102
0.1
0.098
0.096
0.094
0.092
0.09
0.088
0.086
1980
1985
1990
1995
2000
2005
Time



60
30
0
2010
2005
2000
1995
1990
1985
1980
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
Frequency (c/day)



Summary:

We have an efficient algorithm for exact global optimization of segmented representations of time series data -- with any data mode and time sampling.
Data gaps, variable exposure, multivariate data are no problem.

For Streaming Data:

Everything described here can be implemented in a real-time or trigger mode:

Return the Next Significant Change-Point in Any Streaming Time Series

New Book:

Way, M.J., J.D. Scargle, K. Ali, and A.N. Srivastava (Eds.), 2012:
Advances in Machine Learning and Data Mining for Astronomy.
Data Mining and Knowledge Discovery Series. Chapman and Hall/CRC.

Conference:

Progress on Statistical Issues in Searches
June 4 - 6, 2012
Kavli Auditorium
SLAC National Accelerator Laboratory, Menlo Park, CA