Bayesian Periodograms

Phil Gregory

University of British Columbia

Nov. 2011

Introduction

Science is concerned with identifying and understanding structures or patterns in nature.

Periodic patterns have proven especially important. This is particular evident in the field of astronomy where the study of periodic phenomena yield:

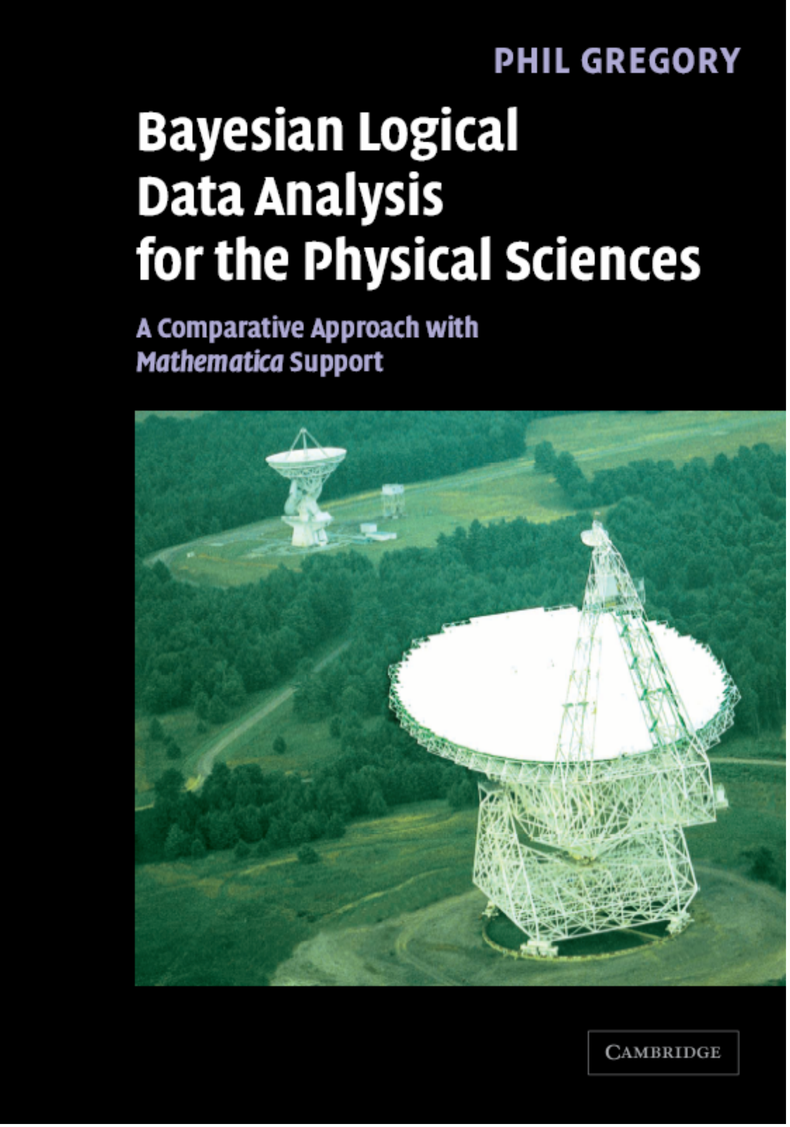
- Fundamental properties like mass and distance
- Interior structure of stars (stellar seismology)
- Extra solar planets
- Exotic states of matter (neutron stars & BH)
- Fundamental tests of physics

Any significant advance in our ability to detect periodic phenomena will profoundly affect our capability of unlocking nature's secrets.

The purpose of this talk is to describe such an advance and provide an illustration of its power through several examples in physics and astronomy.

Outline (Lecture based on Chapter 13 of my book)

1. Bayesian primer 1
2. A Bayesian revolution in spectral analysis
    - Fourier power spectrum, new Bayesian insight 2
    - Bayesian spectral analysis with strong prior information about the signal model 3
        - Bretthorst periodogram 4
        - Kepler periodogram 5
    - Bayesian spectrum analysis with weak prior information about the signal model 6
        - Gregory-Loredo periodogram 7
3. Conclusions 8



Chapters

1. Role of probability theory in science
2. Probability theory as extended logic
3. The how-to of Bayesian inference
4. Assigning probabilities
5. Frequentist statistical inference
6. What is a statistic?
7. Frequentist hypothesis testing
8. Maximum entropy probabilities
9. Bayesian inference (Gaussian errors)
10. Linear model fitting (Gaussian errors)
11. Nonlinear model fitting
12. Markov chain Monte Carlo
13. Bayesian spectral analysis
14. Bayesian inference (Poisson sampling)

Resources and solutions
This title has free Mathematica based support software available

Introduces statistical inference in the larger context of scientific methods, and includes 55 worked examples and many problem sets.

A Bayesian Revolution in Spectral Analysis

1) Fourier Power Spectrum (Schuster periodogram 1905)

The use of the Discrete Fourier Transform (DFT) is ubiquitous in spectral analysis as a result of the FFT introduced by Cooley and Tukey in 1965.

$$\text{periodogram} = \mathrm{C}(\mathrm{f}_n) = \frac{1}{\mathrm{N}} \left| \sum_{k=1}^{\mathrm{N}} \mathrm{d}_k e^{-i2\pi\, n\Delta f\, k\Delta t} \right|^2$$

$$= \frac{1}{\mathrm{N}} \left| \mathrm{FFT} \right|^2$$

2) New Insights on the periodogram from Bayesian Probability Theory (BPT)

In 1987 E. T. Jaynes derived the DFT and periodogram directly from the principles of BPT and showed that the periodogram is an optimum statistic for the detection of a single stationary sinusoidal signal in the presence of independent Gaussian noise.

He showed that the probability of the frequency of a periodic signal is given to a very good approximation by,

$$\mathrm{p}(\mathrm{f}_n \mid \mathrm{D}, \mathrm{I}) \propto \exp\left\{ \frac{\mathrm{C}(\mathrm{f}_n)}{\sigma^2} \right\}$$

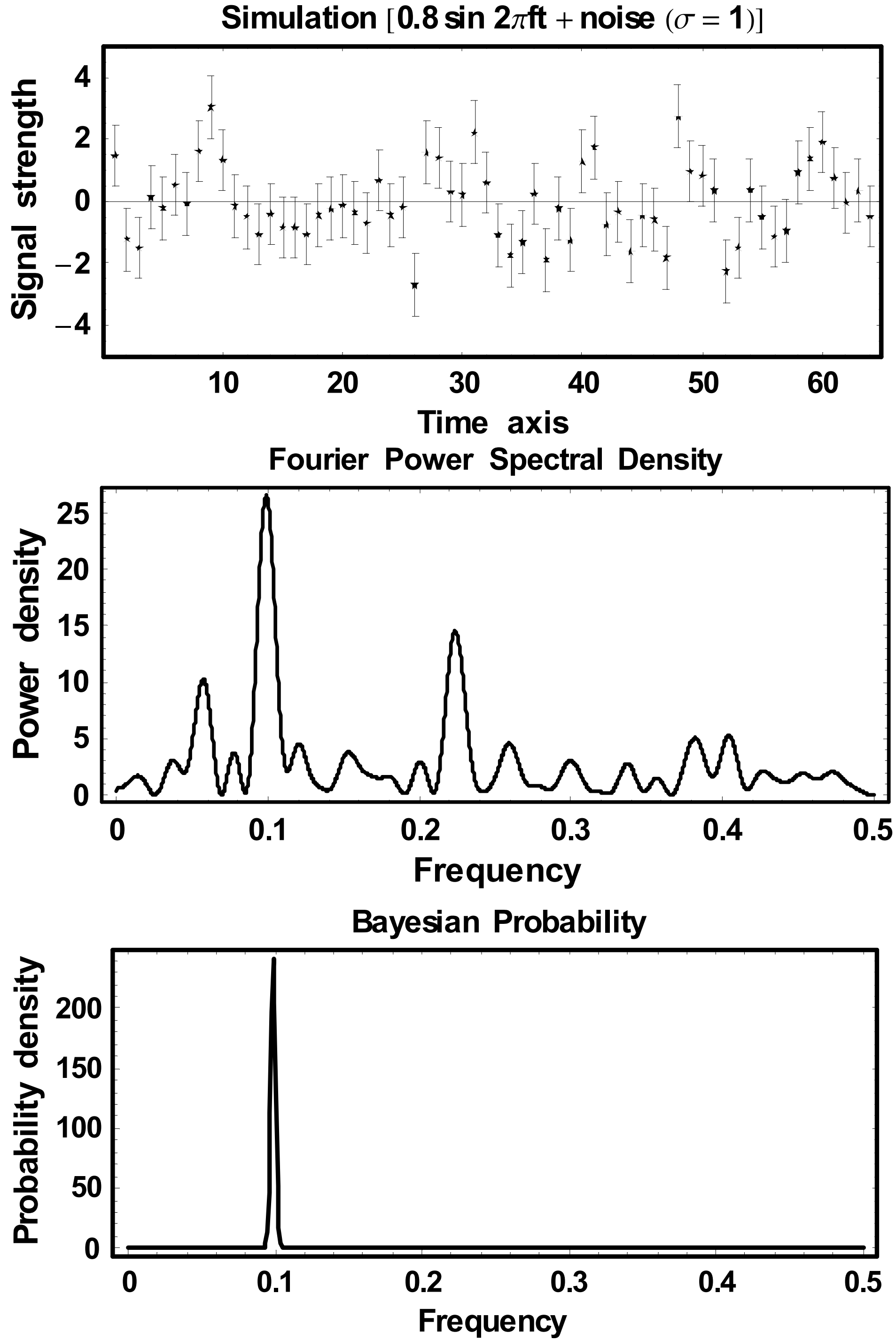
$$p(f_n \mid D, I) \propto \exp\left\{\frac{C(f_n)}{\sigma^2}\right\}$$

Thus $C(f_n)$ is indeed fundamental to spectral analysis but not because it is itself a satisfactory spectrum estimator.

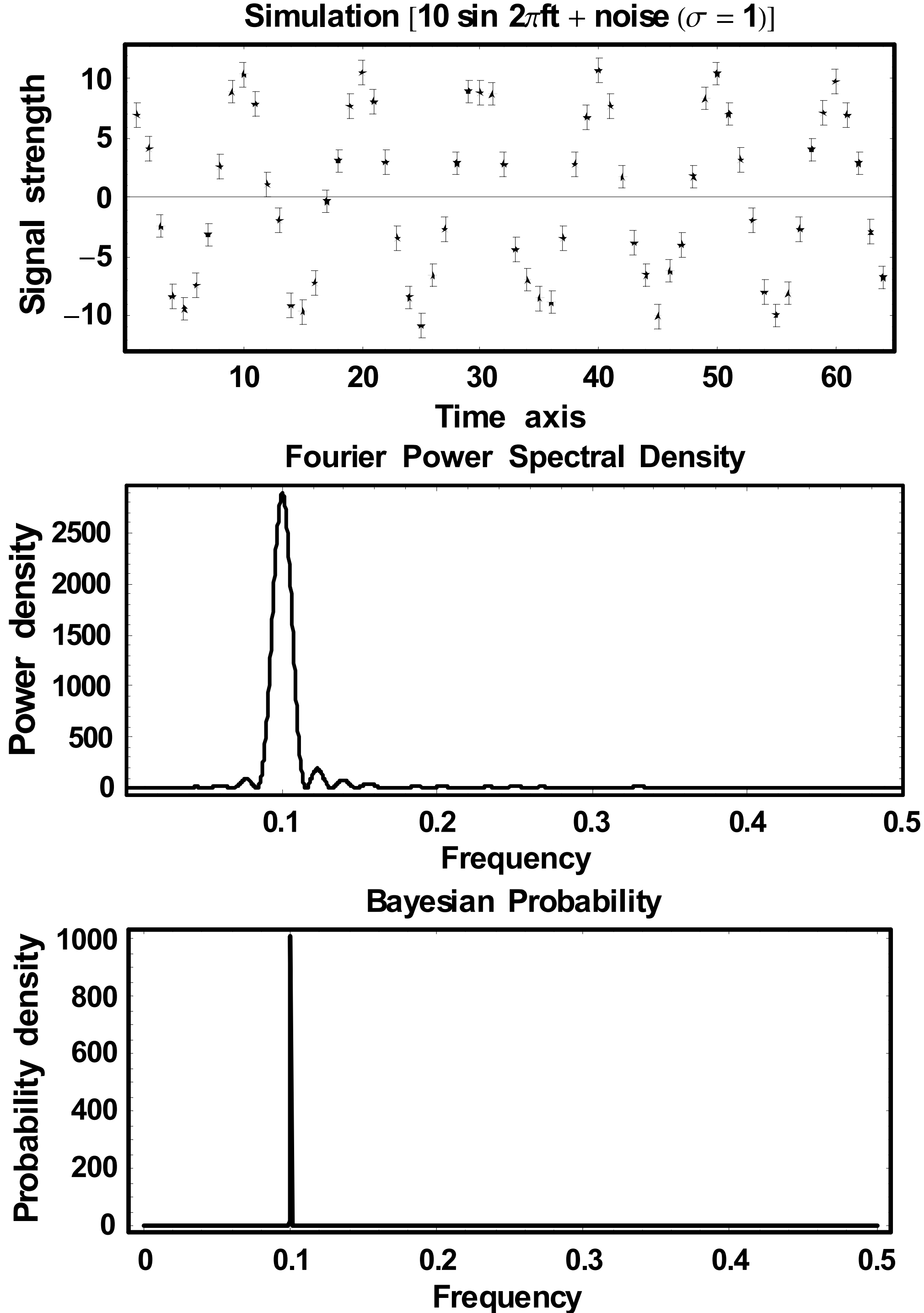
The proper algorithm to convert $C(f_n)$ to $p(f_n|D,I)$ involves first dividing $C(f_n)$ by the noise variance and then exponentiation.

This naturally suppresses spurious ripples at the base of the periodogram as well as linear smoothing; but does it by attenuation rather than smearing, and therefore does not lose any resolution.

The Bayesian nonlinear processing of $C(f_n)$ also yields, when the data give evidence for them, arbitrarily sharp spectral peaks.



Simulation [0.8 sin 2πft + noise (σ = 1)]
Signal strength
Time axis
Fourier Power Spectral Density
Power density
Frequency
Bayesian Probability
Probability density
Frequency


Simulation [10 sin 2πft + noise (σ = 1)]
Signal strength
10
5
0
−5
−10
10
20
30
40
50
60
Time axis
Fourier Power Spectral Density
Power density
2500
2000
1500
1000
500
0
0.1
0.2
0.3
0.4
0.5
Frequency
Bayesian Probability
Probability density
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
Frequency

What if σ is unknown

$$p(f_n \mid D, I) \propto \exp\left\{\frac{C(f_n)}{\sigma^2}\right\}$$

This equation assumes that the noise variance is a known quantity.

In some situations, the noise is not well understood, i.e., our state of knowledge is less certain. Even if the measurement apparatus noise is well understood, the data may contain a greater complexity of phenomena than the current signal model incorporates.

Again, Bayesian inference can readily handle this situation by treating the noise variance as a nuisance parameter with a prior distribution reflecting our uncertainty in this parameter. We need to integrate over this parameter to compute $p(f_n|D,I)$.

The resulting posterior can be expressed in the form of a Student's *t* distribution. The corresponding result for estimating the frequency of single sinusoidal signal (Bretthorst 1988, Bayesian Spectrum Analysis and Parameter Estimation, Springer) is given approximately by

$$p(f_n \mid D, I) \propto \left[1 - \frac{2C(f_n)}{N\overline{d^2}}\right]^{\frac{2-N}{2}}$$

where

$$\overline{d^2} = \frac{1}{N}\sum_j d_i^2$$

Bayesian Spectrum Analysis with Strong Prior Information of the Signal Model

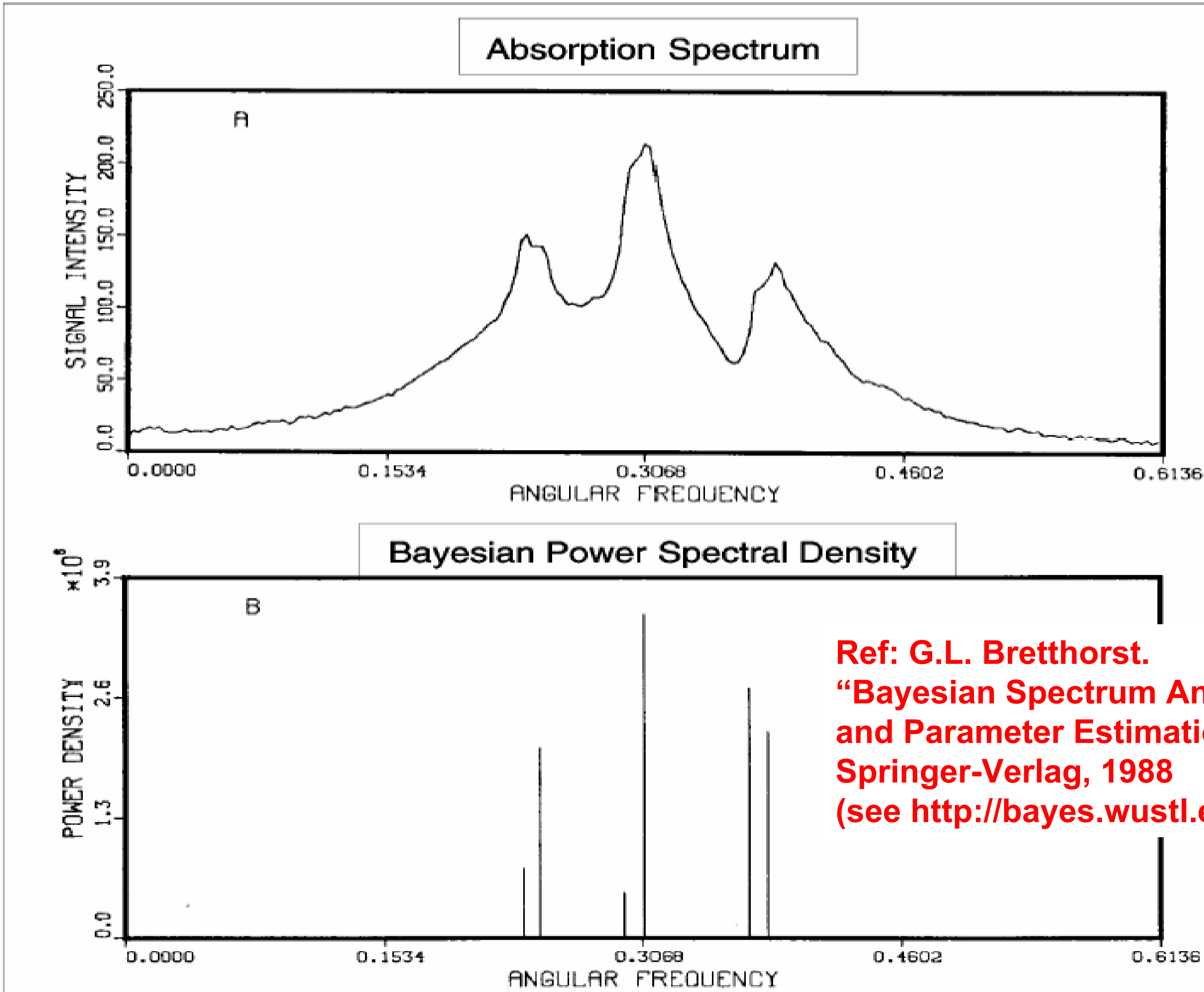
Larry Bretthorst (Jaynes' last PhD student) extended Jaynes' work to more complex signal models with additive Gaussian noise and revolutionized the analysis of Nuclear Magnetic Resonance (NMR) signals.

Here one is dealing with multiple damped sinusoids.

See http://bayes.wustl.edu/ for a copy of Larry Bretthorst's papers and book.

Varian Corporation now offer an expert analysis package with their new NMR machines based on Bretthorst's Bayesian algorithm.

Analysis of Nuclear Magnetic Resonance Free Induction Decay Data



Ref: G.L. Bretthorst. "Bayesian Spectrum Analysis and Parameter Estimation", Springer-Verlag, 1988 (see http://bayes.wustl.edu)

The Bretthorst periodogram:

Bretthorst generalized Jaynes' insights to a broader range of single-frequency and multi-frequency estimation problems and sampling conditions.

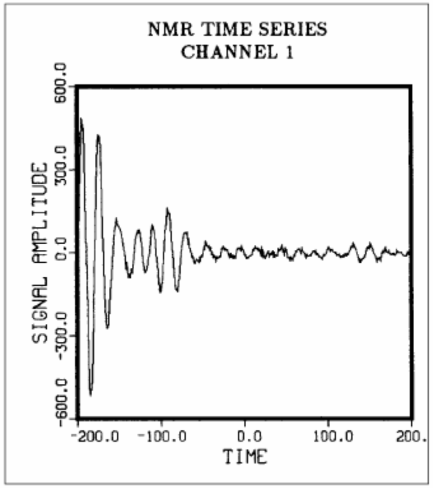
In the single-frequency case which we examine here, he established a connection between the Bayesian results and an existing frequentist statistic known as the Lomb-Scargle periodigram, which is a widely used replacement for the Schuster periodogram in the case of non-uniform sampling.

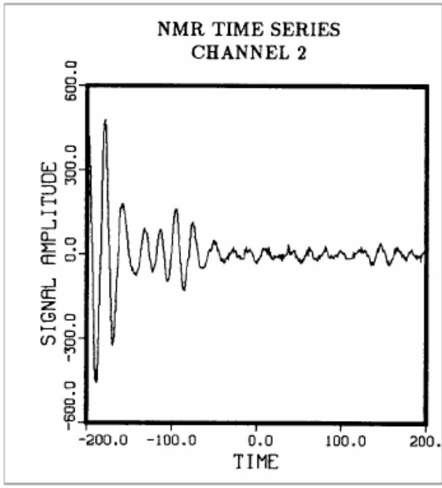
Bretthorst, G.L. (2001), American Institute of Physics Conference Proceedings, 568, pp. 241.

Bretthorst's analysis allows for the following complications:

1. Either real or quadrature data sampling. Quadrature data involves measurements of the real and imaginary components of a complex signal.

The figure show an example of quadrature signals occurring in NMR.





The Brethorst periodogram

outline

2. Allows for uniform or non-uniform sampling and for quadrature data with non-simultaneous sampling.

The analysis does not require the real and imaginary data samples to be simultaneous and successive samples can be unequally spaced in time.

3. Allows for non-stationary single sinusoid model of the form

(real channel)

$$d_R(t_i) = A\cos(2\pi f\, t_i - \theta)\, Z(t_i) + B\sin(2\pi f\, t_i - \theta)\, Z(t_i) + e_R(t_i),$$

(imaginary channel)

$$d_I(t_j') = A\cos(2\pi f\, t_j' - \theta)\, Z(t_j') + B\sin(2\pi f\, t_j' - \theta)\, Z(t_j') + e_I(t_j'),$$

The function $Z(t_i)$ describes an arbitrary modulation of the amplitude, e.g., exponential decay as exhibited in NMR signals.

In this analysis $Z(t_i)$ is assumed to be known, but in other analysis he allows it to have unknown parameters.

The Brretthorst periodogram

The angle θ is defined in such a way as to make the cosine and sine functions orthogonal on the discretely sampled times. In general, θ is frequency dependent.

Note: if the data are simultaneously sampled, $t_i = t'_j$, then the orthogonal condition is automatically satisfied so $\theta = 0$.

4. The noise terms $e_R(t_i)$ and $e_I(t'_j)$ are assumed to be IID Gaussian with an unknown σ. Thus, is a nuisance parameter, which is assumed to have a Jeffreys prior. By marginalizing over σ, any variability in the data that is not described by the model is assumed to be noise.

The Bretthorst periodogram

In this problem the main parameter of interest is the frequency f. To compute p (f |D,I) we need to marginalize over the two amplitude parameters A, B, and σ.

$$p(f \mid D, I) = \int dA\, dB\, d\sigma \;\; p(f, A, B, \sigma \mid D_R, D_I, I)$$

The RH side of this equation can be factored using Bayes' theorem and the product rule to yield

$$p(f \mid D, I) \propto \int dA\, dB\, d\sigma \;\; p(f \mid I)\, p(A \mid I)\, p(B \mid I)\, p(\sigma \mid I) \times$$

$$p(D_R \mid f, A, B, \sigma, I)\, p(D_I \mid f, A, B, \sigma, I)$$

Bretthorst assigns uniform priors for f, A & B and a Jeffreys prior for σ.

The Bretthorst periodogram

It turns out that the triple integral can be performed analytically using simple changes in the variables.

The final Bayesian expression for *p* (*f* |*D,I*), after marginalizing over amplitudes *A, B* & *σ* is given by

$$p(f|D,I) \propto \frac{1}{\sqrt{C(f)S(f)}}\left[N\overline{d^2} - \overline{h^2}\right]^{\frac{2-N}{2}},$$

where $$\overline{h^2} = \frac{R(f)^2}{C(f)} + \frac{I(f)^2}{S(f)},$$

The Bretthorst periodogram

where

$$R(f) \equiv \sum_{i=1}^{N_R} d_R(t_i)\cos(2\pi f t_i - \theta) Z(t_i) - \sum_{j=1}^{N_I} d_I(t'_j)\sin(2\pi f t'_j - \theta) Z(t'_j),$$

$$I(f) \equiv \sum_{i=1}^{N_R} d_R(t_i)\sin(2\pi f t_i - \theta) Z(t_i) + \sum_{j=1}^{N_I} d_I(t'_j)\cos(2\pi f t'_j - \theta) Z(t'_j),$$

$$C(f) \equiv \sum_{i=1}^{N_R} \cos^2(2\pi f t_i - \theta) Z(t_i)^2 + \sum_{j=1}^{N_I} \sin^2(2\pi f t'_j - \theta) Z(t'_j)^2$$

and

$$S(f) \equiv \sum_{i=1}^{N_R} \sin^2(2\pi f t_i - \theta) Z(t_i)^2 + \sum_{j=1}^{N_I} \cos^2(2\pi f t'_j - \theta) Z(t'_j)^2.$$

$$\theta = \frac{1}{2}\tan^{-1}\left[\frac{\sum_{i=1}^{N_R} \sin(4\pi f t_i) Z(t_i)^2 - \sum_{j=1}^{N_I} \sin(4\pi f t'_j) Z(t'_j)^2}{\sum_{i=1}^{N_R} \cos(4\pi f t_i) Z(t_i)^2 - \sum_{j=1}^{N_I} \cos(4\pi f t'_j) Z(t'_j)^2}\right]$$

Result

$$P(f|DI) \propto \frac{1}{\sqrt{C(f)S(f)}} \left[N\overline{d^2} - \overline{h^2}\right]^{\frac{2-N}{2}}$$

where the sufficient statistic $\overline{h^2}$ is given by $\overline{h^2} = \frac{R(f)^2}{C(f)} + \frac{I(f)^2}{S(f)}$

Simplifications

1. **When the data are real and the sinusoid is stationary, the sufficient statistic for single frequency estimation is the Lomb-Scargle periodogram; not the Schuster periodogram (power spectrum). However, the Schuster periodogram is often an excellent approximation and is much faster to compute.**

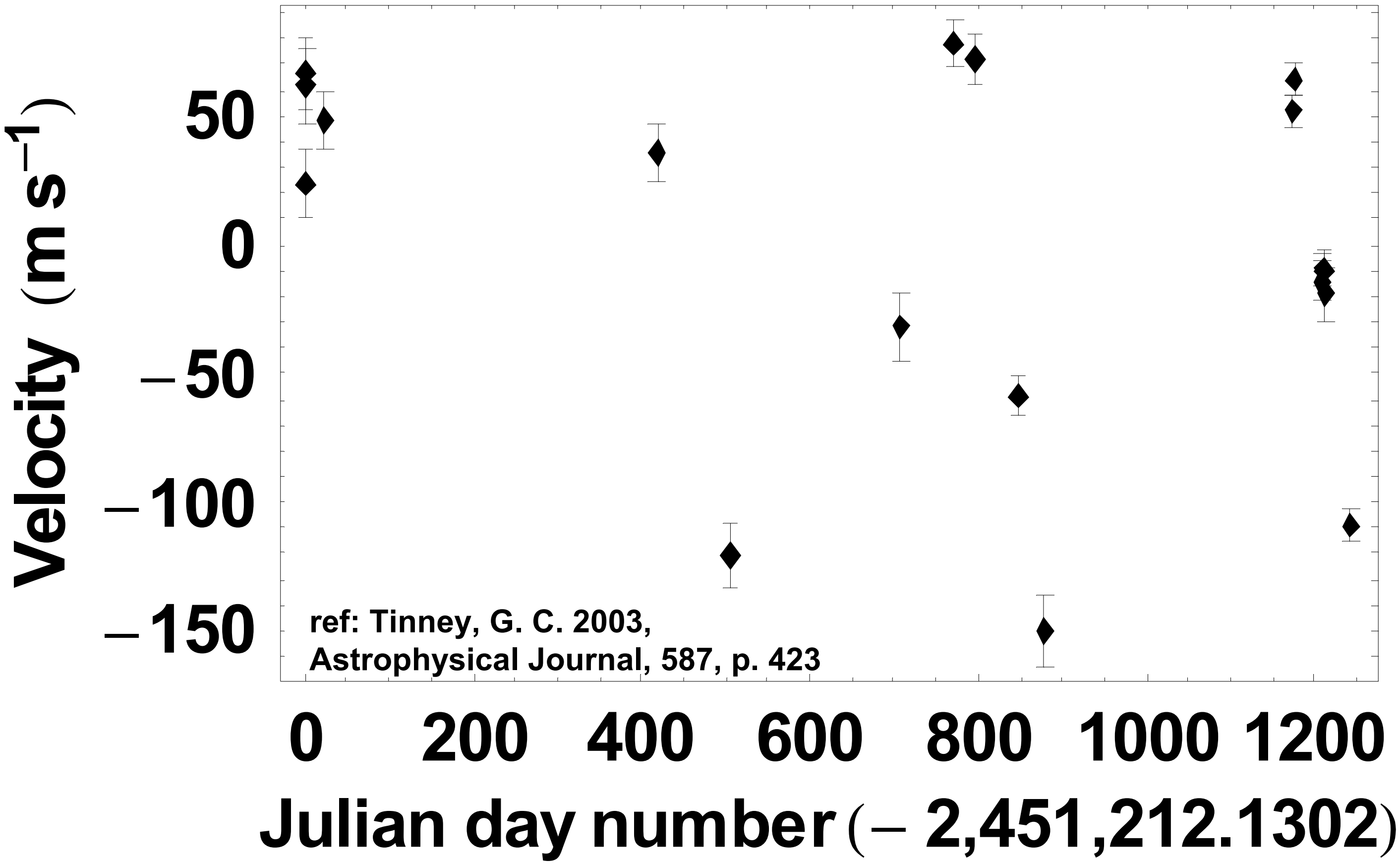
2. **When the data are real, but *Z(t)* is not constant, then $\overline{h^2}$ generalizes the Lomb-Scargle periodogram in a very straightforward manner to account for the decay of the signal.**

3. **For uniformly sampled quadrature data when the sinusoid is stationary, $\overline{h^2}$ reduces to a Schuster periodogram of the data.**

$$p(f_n \mid D, I) \propto \left[1 - \frac{2C(f_n)}{N\overline{d^2}}\right]^{\frac{2-N}{2}}$$

Application to extra-solar planet data

Precision radial velocity measurements for HD 73526

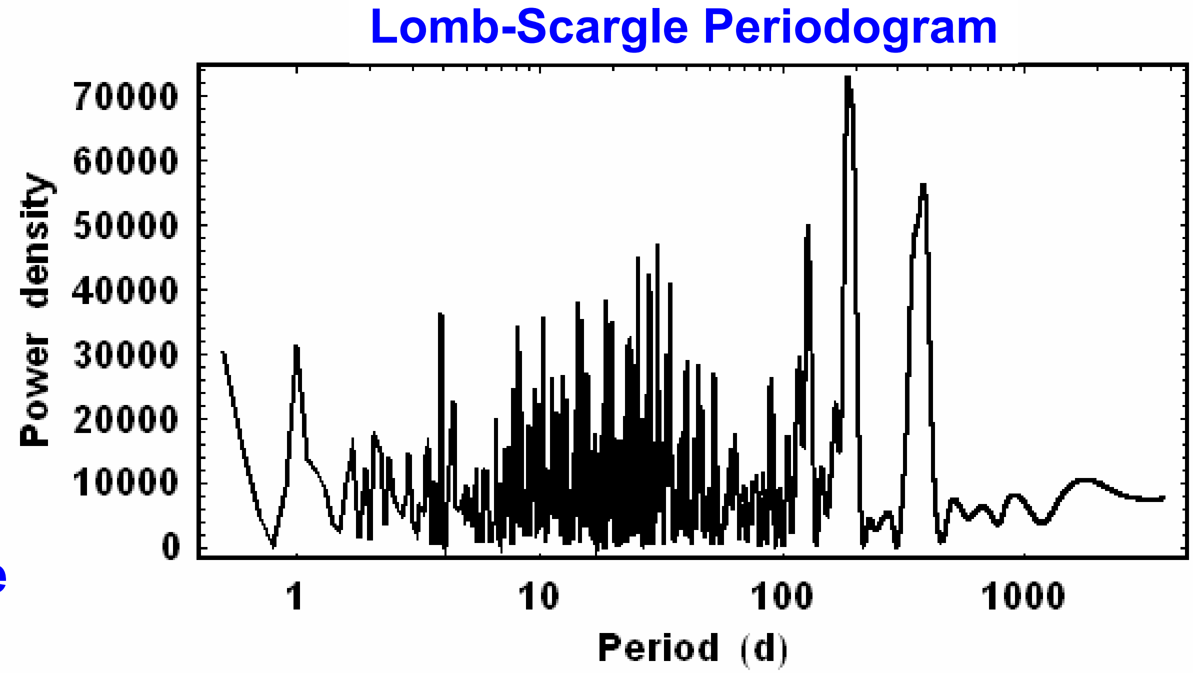


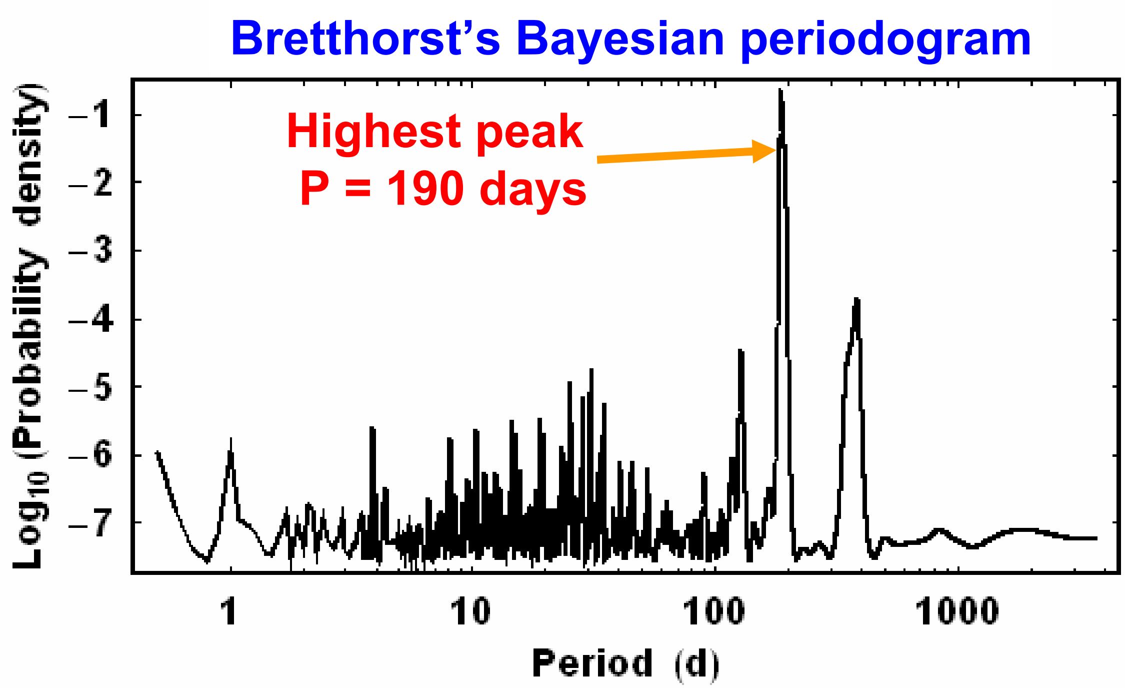
Gregory, P. C., Astrophysical J., 631, 1198, 2005
"A Bayesian Analysis of Extrasolar Planet Data for HD 73526",

Conventional Nonlinear least-squares analysis requires a good initial guess at the parameter values.

Need to use some form of periodogram to estimate the orbital period.

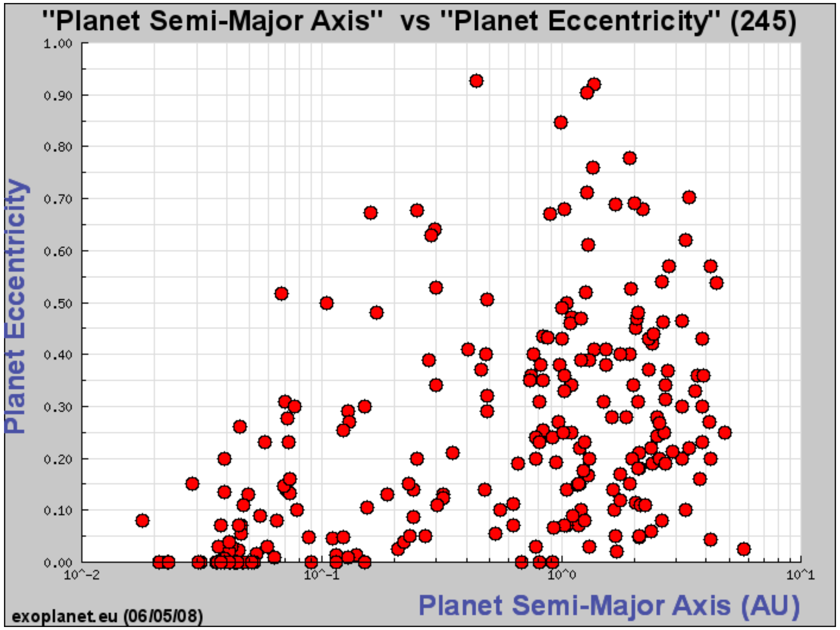
Here is a comparison of the Lomb-Scargle and Bretthorst's Bayesian generalization.





Problem

The Lomb-Scargle periodogram, and Bretthorst's Bayesian generalization, assume a sinusoidal signal which is only optimum for circular orbits.



Why not develop a Bayesian Kepler periodogram designed for all Kepler orbits?

In 2005, Eric Ford and I independently published Kepler Periodograms based on a Markov chain Monte Carlo (MCMC) approach.

This was the subject of Lecture 1.

Bayesian Spectrum Analysis with Weak Prior Information of the Signal Model

In the early 1990's, Tom Loredo and I attached the question of

"How to detect a signal of unknown shape in a time series"

We developed a useful Bayesian solution to this problem.

We were initially motivated by the problem of detecting periodic signals in X-ray astronomy data. In this case the time series consisted of individual photon arrival times where the appropriate sampling distribution is the Poisson distribution.

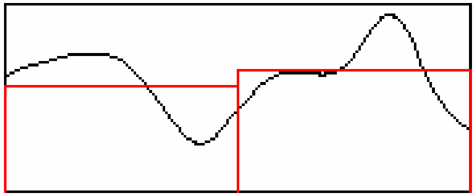
X-ray pulsars exhibit a wide range of pulse shapes.

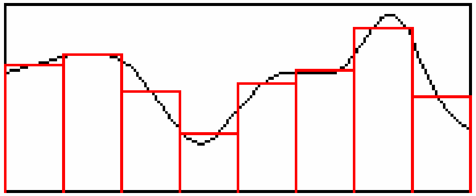
Gregory-Loredo (GL) Method (Astrophysical J. 398,1992)

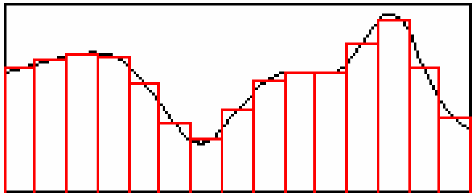
To address the detection problem we compute the ratio of the probabilities (odds) of two models M_{Per} and M_1 .

Model M_{per} is a family of periodic models capable of describing a background + periodic signal of arbitrary shape.

Each member of the family is a histogram with m bins where m > or = 2. Three examples are shown .



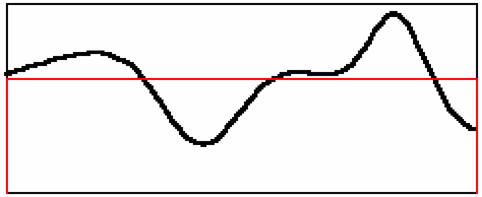




M_m represents the periodic model with m bins.

Model M_1 assumes the data is consistent with a constant event rate A.

M_1 is a special case of M_m with m = 1 bin



GL Method continued

Search parameters for one of the m bin periodic models

period, phase, and m shape parameters

Special feature of the histogram models is that the search in the m shape parameters can be carried out analytically, permitting the method to be computationally tractable.

The Bayesian posterior probability for a particular periodic model, M_m, contains a term which quantifies Occam's razor, penalizing successively more complicated periodic models (increasing m) for their greater complexity.

The calculation thus balances model simplicity with goodness of fit, allowing us to determine both:

- **whether there is evidence for a periodic signal, and**
- **the optimum number of bins for describing the structure in the data.**

The Bayesian solution is very satisfying:
In the absence of knowledge about the shape of the signal the method identifies the most organized (minimum entropy) significant periodic structure in the model parameter space. What structure is significant is determined through built in quantified Occam's penalties in the calculation.

Re-analysis of 50 ms X-ray Pulsar in Large Magellanic Cloud

1977 - first discovery by Seward et al. with Einstein telescope

FFT analysis of photon arrival times (21,000 events). Not on optimum detection strategy but widely used because of its speed. The pulsar was the second highest peak in power spectrum which repeated using different data sets.

$$P = 0.050208679\,\mathrm{s} \pm 3\mathrm{ns}$$

$$\dot{P} = 4.79 \times 10^{-13}\,\mathrm{ss}^{-1}$$

Using a small sample of 3850 events, the Bayesian GL algorithm yielded a period with an uncertainty 0.3 ns

1990 - pulsar re-observed with ROSAT satellite

FFT analysis failed to detect the pulsar.

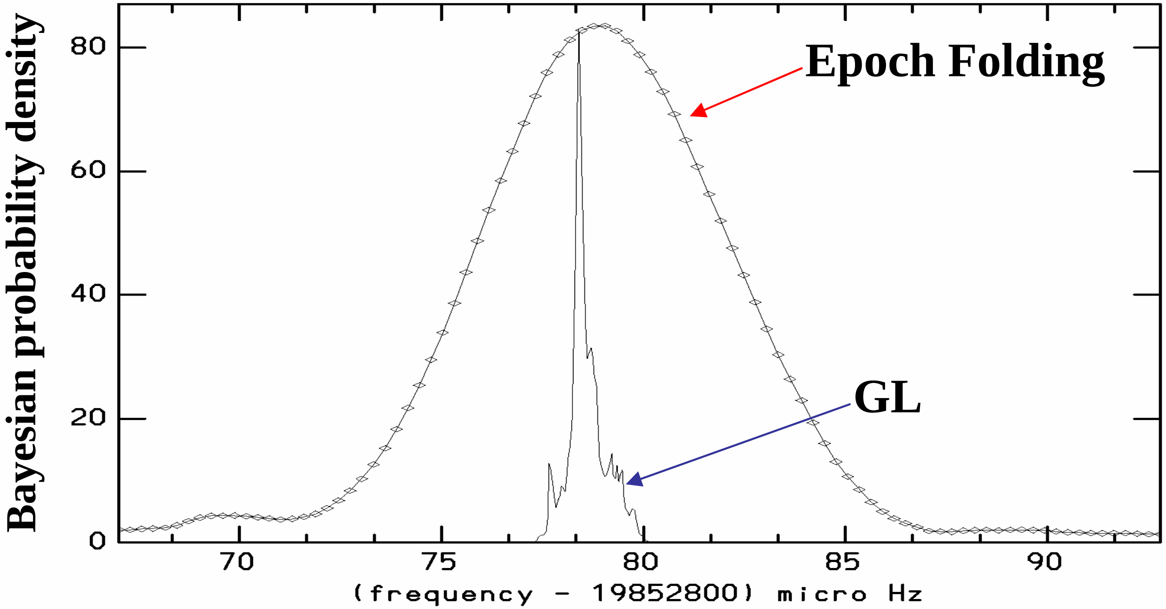
a) Using prior knowledge of P and P derivative GL yielded a detection with an

$$\mathrm{Odds} = \frac{\text{probability of periodic signal}}{\text{probability of no periodic signal}} = 10^{11}$$

b) Ignoring the prior detection, and assuming a prior P range from 1 ms to 1/2 data duration=16h.

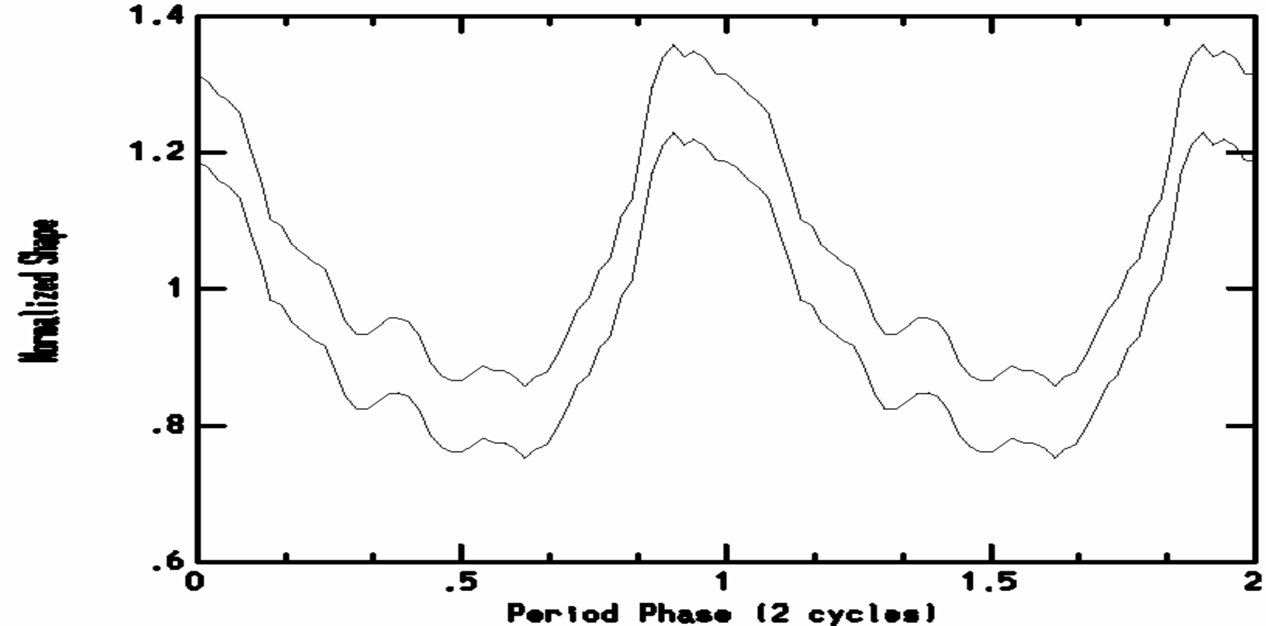
GL yielded a detection with an Odds = 5 x 10^5

Pulse frequency determination, GL versus Epoch Folding



Ref: Gregory, P.C., & Loredo, T.J., "Bayesian Periodic Signal Detection: Analysis of ROSAT Observations of PSR 0540-693", Astrophysical J., 473, 1059, 1996

Pulsar light curve from GL (mean + & - 1 standard deviation)

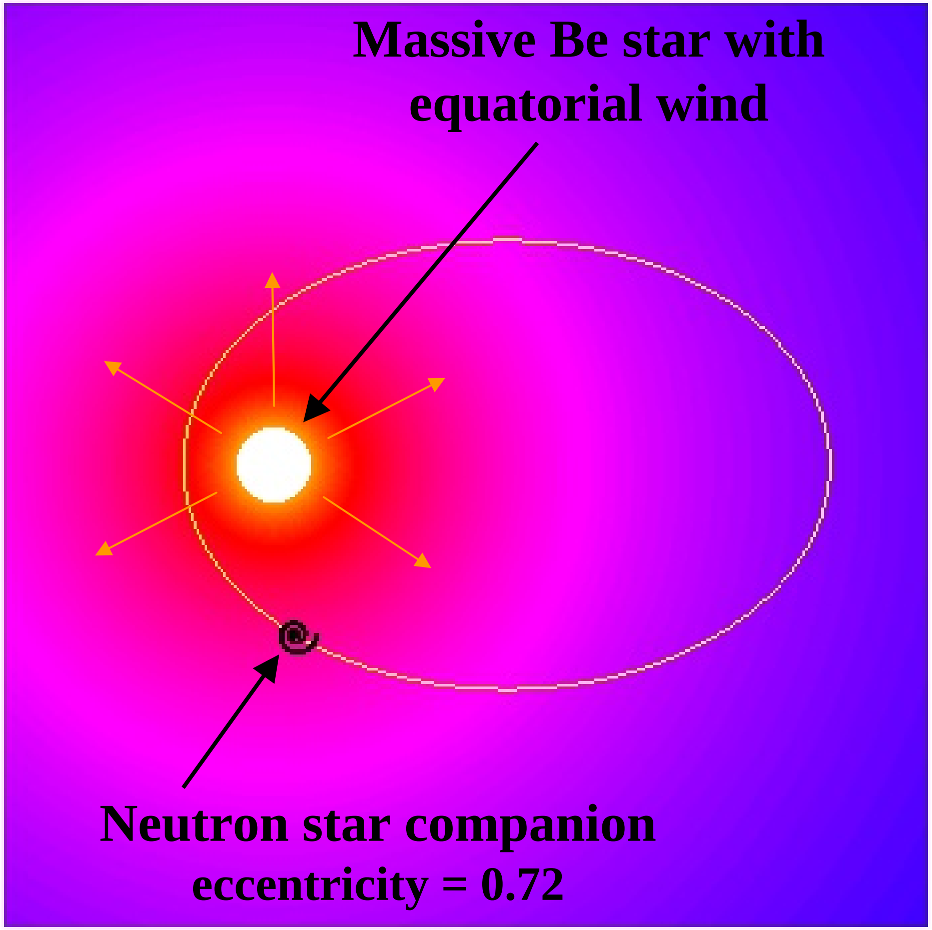


Bayesian Spectrum Analysis with Weak Prior Information of the Signal Model

In 1999 I extended the GL algorithm to the Gaussian noise case and applied it to a radio astronomy data set.

Gregory, Astrophysical J. 580, 1133, 2002

This resulted in the discovery of a new periodic phenomena in the radio and X-ray binary LS I+61^{0} 303.

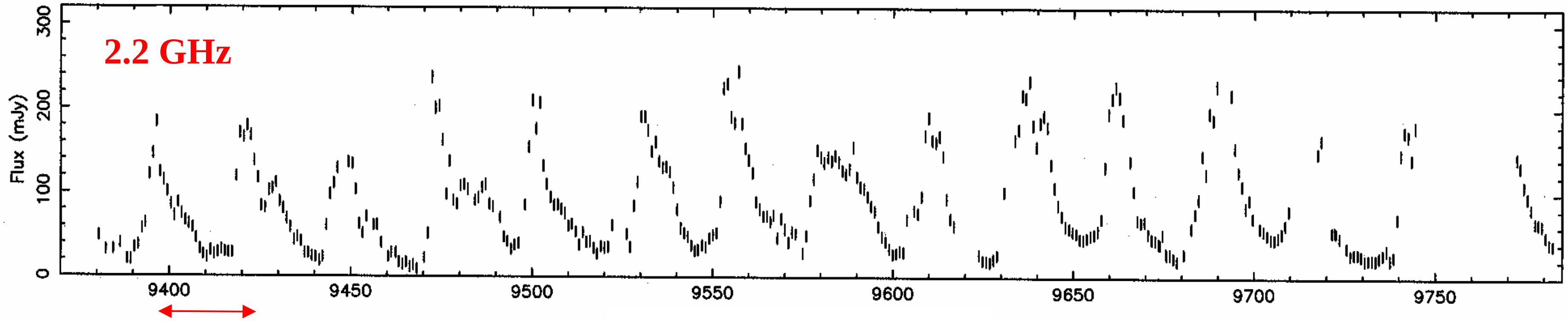


References:
(1) Gregory, Peracaula, & Taylor, Astrophysical J., 520, 376,1999

(2)Gregory, Astrophysical J., 575, 427,2002

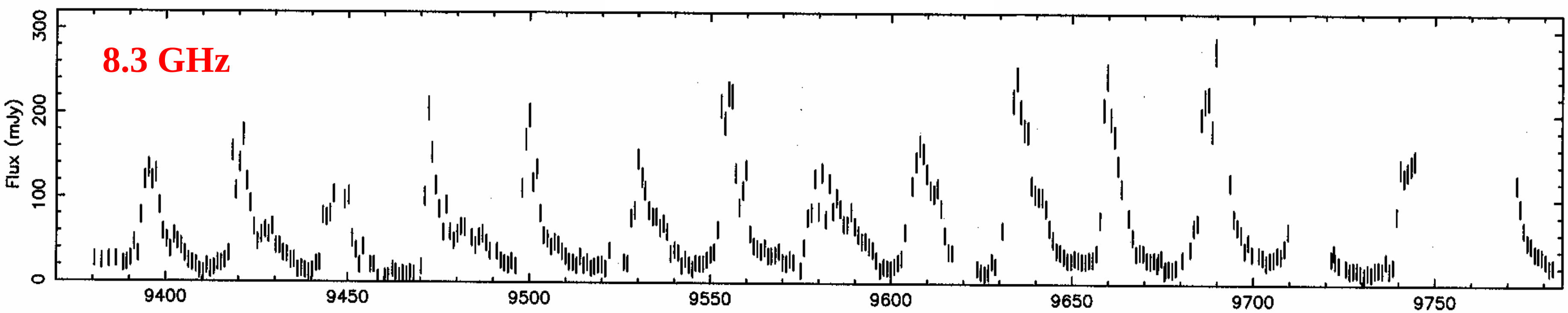
Periodic Radio Outbursts from LS I+61^0 303

Green Bank Interferometer-NASA-Naval Observatory Radio Monitoring



Julian Day - 2440000.0

P_1~26.5 d



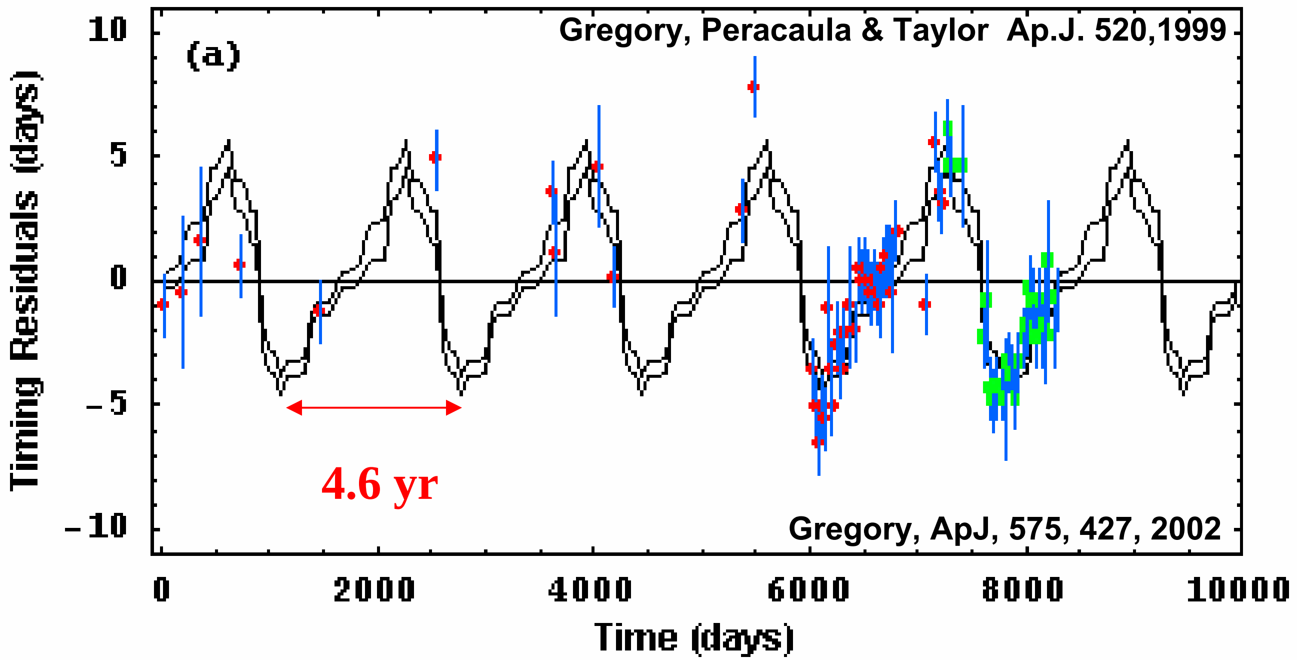
Julian Day - 2440000.0

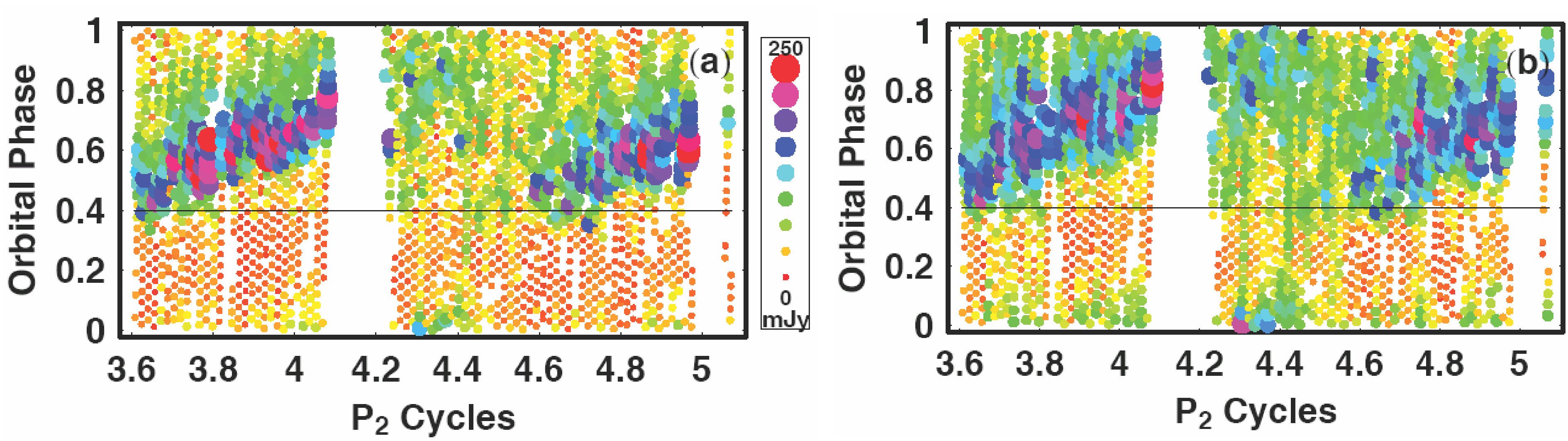
1 year

23 years of data since discovery in 1977

The last 6 years are from the GBI radio monitoring program

Discovery of a 4.6 year saw-tooth modulation of the phase of the radio outbursts





Plots of the daily averaged radio flux densities at, (a) 8.3 GHz, and (b) 2.2 GHz versus orbital phase and 4.6 year modulation cycle.

Wind Probe Analysis

In a subsequent paper Gregory & Niesh showed how to extract information about the Be star wind density and velocity from the radio flux density measurements.

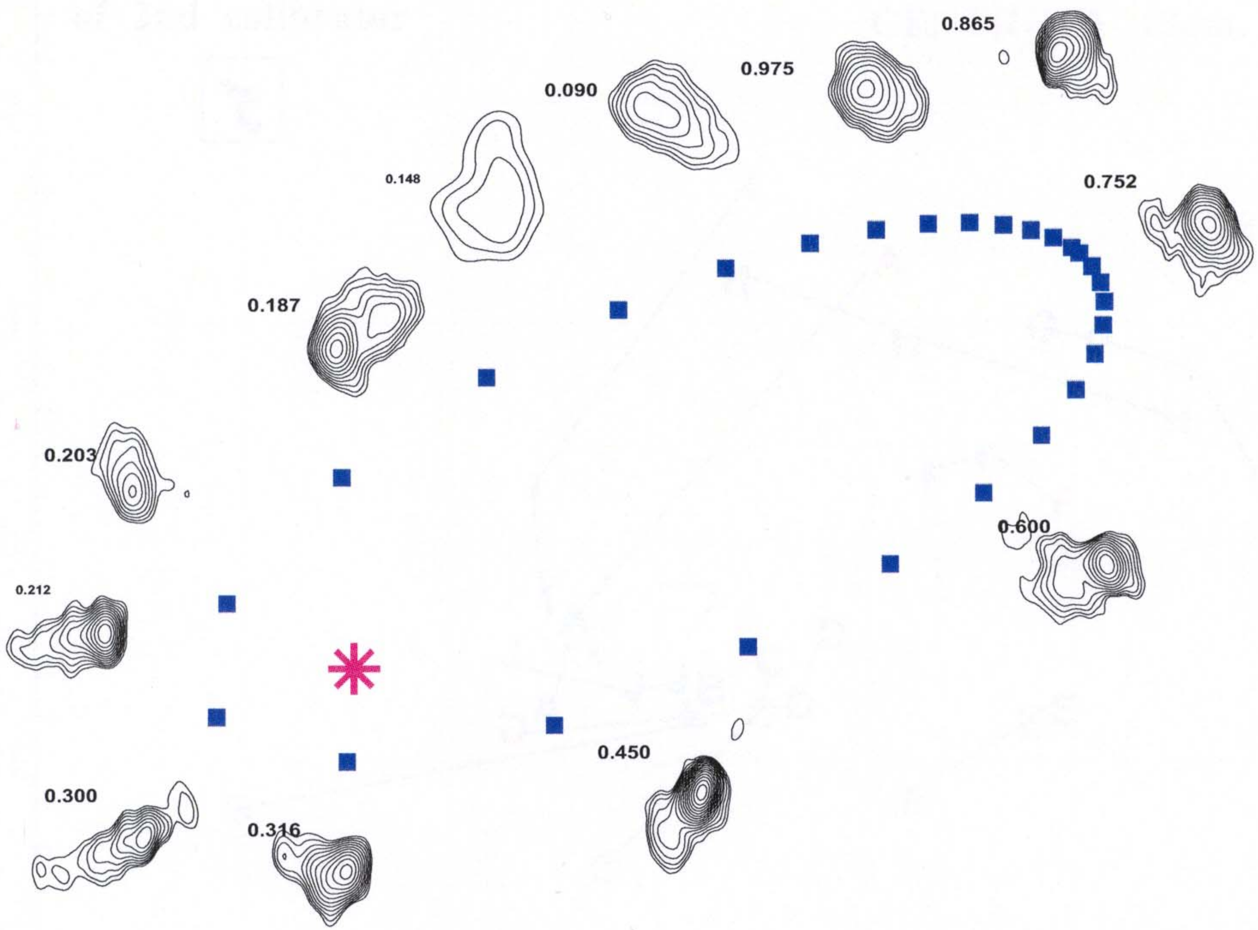
Conclusion: find evidence for a periodic outward moving density enhancement in the Be star equatorial disk.

Density Movie

Gregory & Niesh Ap. J. 580, p. 1133, 2002)

LS I +61 303 as a Be-Pulsar binary

The structure of AU-scale radio imaging with the VLBA appears cometary, typical of shock-accelerated particles from the interaction of a pulsar wind with the dense equatorial wind from the Be star.



Dhawan, Vivek, Mioduszewski, Amy, Rupen, M. Proceedings of the VI Microquasar Workshop: Microquasars and Beyond, 2006, Como, Italy., p.52.1

Conclusions

1. In these examples I have tried to demonstrate the power of the Bayesian approach in the arena of spectral analysis.
2. A Bayesian analysis can sometimes yield orders of magnitude improvement in model parameter estimation, by the inclusion of valuable prior information such as our knowledge of the signal model.
3. For some problems a Bayesian analysis may lead to a familiar statistic. Even in this situation (as we saw with the periodogram) it often leads to new insights concerning the interpretation and generalization of the statistic.
4. As a theory of extended logic it can be used to address the question of what is the optimal answer to a particular scientific question for a given state of knowledge, in contrast to a numerical recipe approach.
5. With all useful theoretical advances we expect to gain powerful new insights. This has already been amply demonstrated in the case of BPT and it is still early days.

One of these important new insights :

BPT provides a means of assessing competing theories at the forefront of science by quantifying Occam's razor,