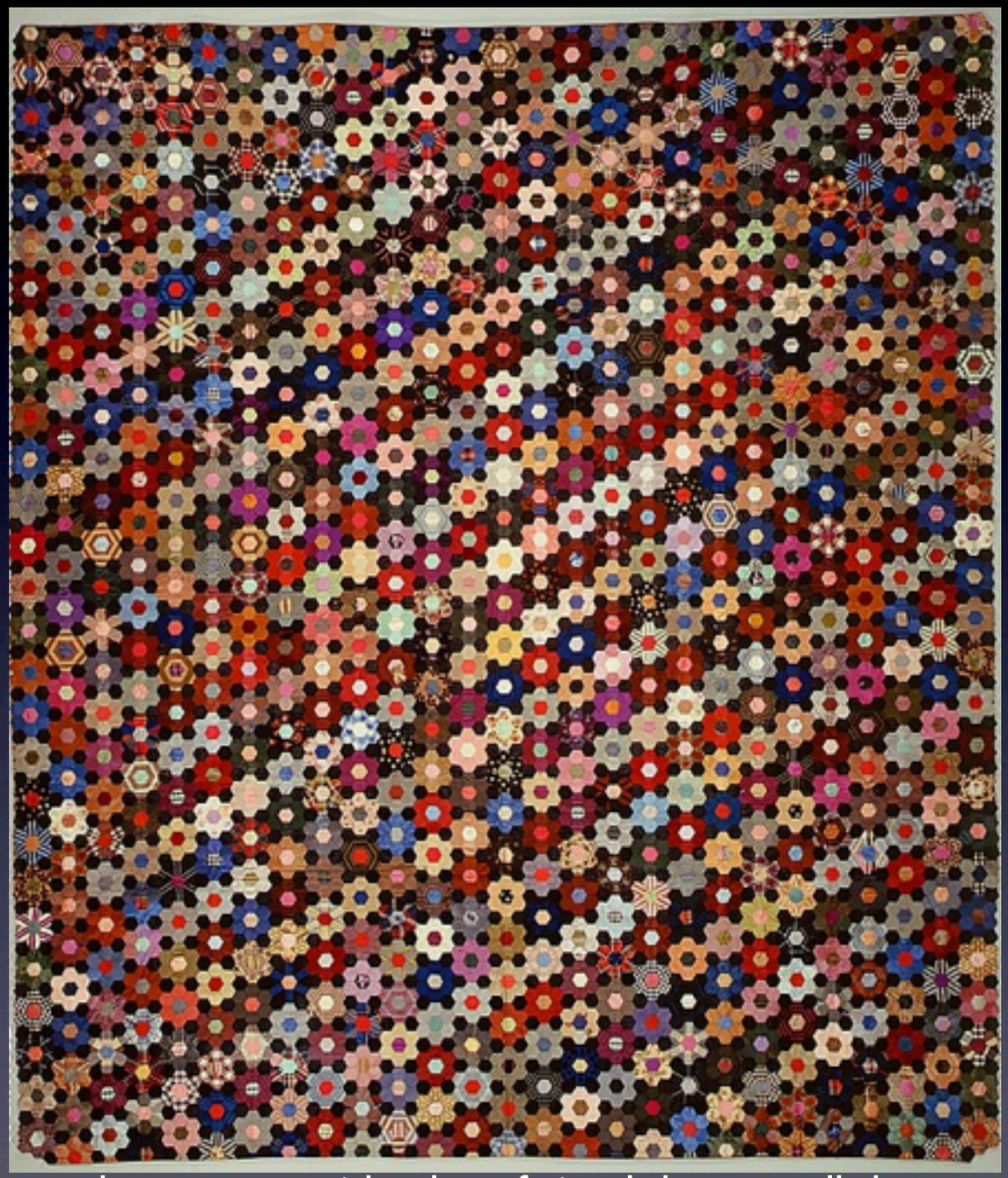
A hybrid method for periodicity detection in time series

Andjelka Kovačević[1], Luka Č. Popović[1,2], Dragana Ilić[1]

[1] Department of astronomy, Faculty of Mathematics, University of Belgrade, Serbia
[2] Astronomical Observatory Belgrade, Serbia


Well, all information looks like noise until you break the code.
Neal Stephenson

What happens when noise becomes information?



If you provide an AI-powered computer with a lot of visual data, say all the cameras on an autonomous car, that computer could filter out different combinations of pixels that are meaningful and categorically save them just as it would a word document.

In other words, AI software can take what looks like a bunch of unmanageable, noisy data and turn it into something useful: signal.

Noise is becoming manageable information, and there is a lot of noise to play with.

Time series processes: periodic

- orbits – planets, binary stars, comets, …
- rotation/cycles – solar cycle, pulsars, Cepheids, …
- identification of new periods→ orbits, rotation, system identification
- estimate parameters: period, amplitude, waveform, (small) perturbations in period,phase

Time series processes: stochastic

- accreting systems (neutron stars, black holes), jets
- cannot predict (time series) data exactly
- statistical comparison between data and model to infer physics of system

Searching for Periodic Signals in Time Series Data I

- Least squares sine fitting
- Discrete Fourier Transform
- Lomb-Scargle Periodogram
- Pre-whitening Data
- Other
- Phase Dispersion Minimization
- Wavelets

Searching for Periodic Signals in Time Series Data II

Fourier Methods are based on the Fourier transform, power spectra, and closely related correlation functions. These methods include the classical or Schuster periodogram (Schuster 1898), the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982), the correlation-based method of Edelson & Krolik (1988), and related approaches (see also Foster 1996, for a discussion of wavelet transforms in this context).

Phase-folding Methods depend on folding observations as a function of phase, computing a cost function across the phased data (often within bins constructed across phase space) and optimizing this cost function across candidate frequencies. Some examples are String Length (Dworetsky 1983), Analysis of Variance (Schwarzenberg-Czerny 1989), Phase Dispersion Minimization (Stellingwerf 1978), the Gregory-Laredo method (Gregory & Loredo 1992), and the conditional entropy method (Graham et al. 2013a). Methods based on correntropy are similar in spirit, but do not always require explicit phase folding (Huijse et al. 2011, 2012).

Least Squares Methods involve fitting a model to the data at each candidate frequency, and selecting the frequency which maximizes the likelihood. The Lomb-Scargle periodogram also falls in this category (see Section 5), as does the Supersmoother approach (Reimann 1994). Other studies recommend statistics other than least square residuals; see, e.g., the orthogonal polynomial fits of Schwarzenberg-Czerny (1996).

Bayesian Approaches apply Bayesian probability theory to the problem, often in a similar manner to the phase-folding and/or least-squares approaches. Examples are the generalized Lomb-Scargle models of Bretthorst (1988), the phase-binning model of Gregory & Loredo (1992), Gaussian process models (e.g. Wang et al. 2012), and models based on stochastic processes (e.g. Kelly et al. 2014).

VanderPlas 2017

our data are red noise

- At optical wavelengths, De Vries et al. (2005) found evidence for the variability of quasar emission following red-noise laws on timescales as long as approximately 40 years.
- problem: how to recover spectrum of very "red" processes?(e.g. Mushotzky et al. 2011)

PROBLEM WITH PERIODOGRAM METHOD

Astronomical Data Analysis Software and Systems XX
ASP Conference Series, Vol. 442
Ian N. Evans, Alberto Accomazzi, Douglas J. Mink, and Arnold H. Rots, eds.

Detection of Periodic Variability in Simulated QSO Light Curves

David B. Westman,[1] Chelsea L. MacLeod,[1] and Željko Ivezić[1,2]

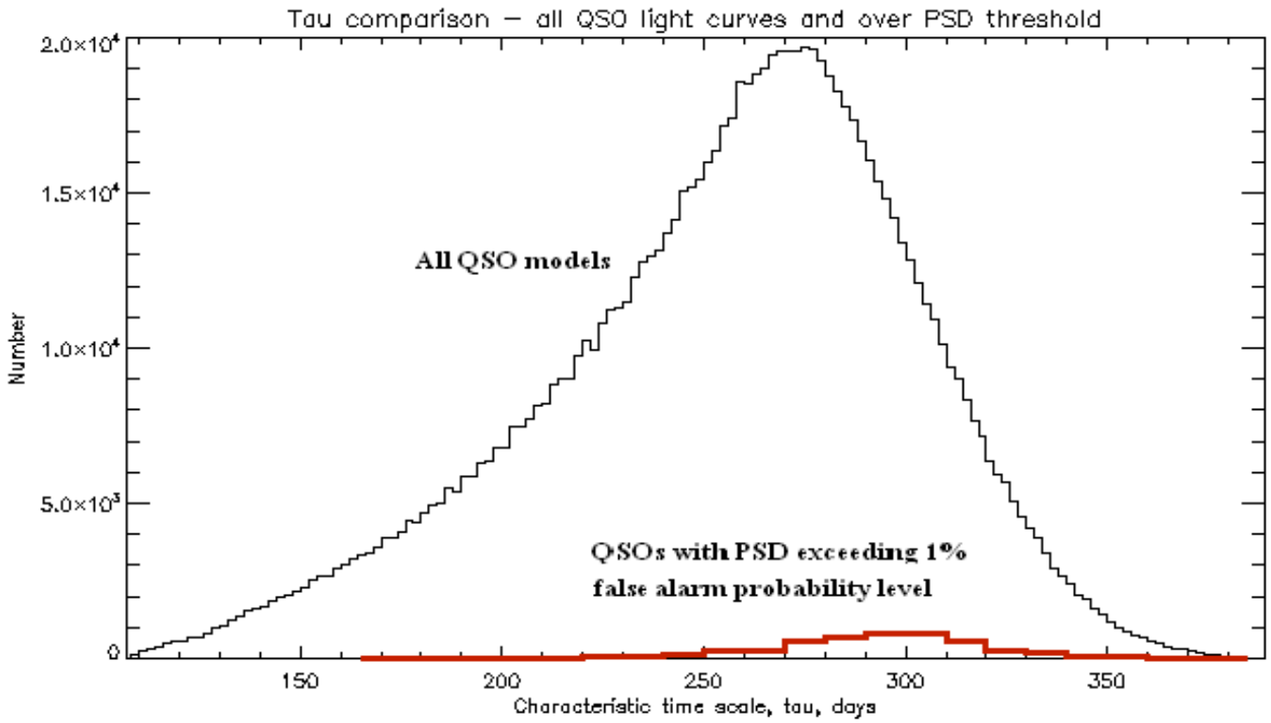


Figure 2. Distribution of characteristic time, τ, for all QSO light curves compared with τ distribution for light curves with a peak SED value exceeding the 1% *fap* level

Figure 2 shows that values of τ for QSO models with PSD exceeding the 1% *fap* level (red histogram) are distributed differently than those for all the QSO models (black histogram). This bias is due to the fact that when τ is long, only a few "oscillations" are observed over the duration of the light curve, causing the periodogram to mistake the damped random walk behavior for a periodic behavior.

PROBLEM WITH PERIODOGRAM METHOD CONT'D

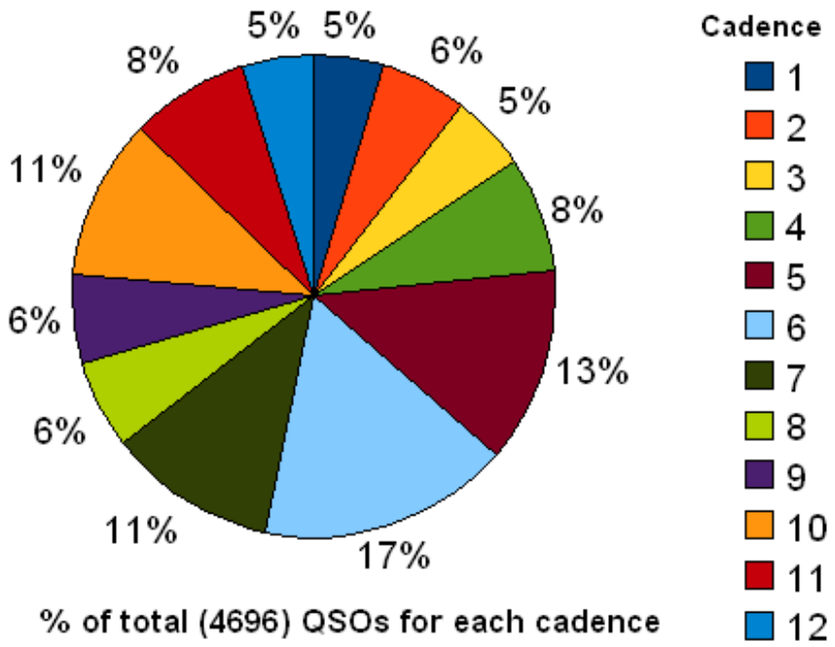


Figure 3. Proportion of QSOs with peak PSD values over 1% *fap* level sorted by LSST cadence

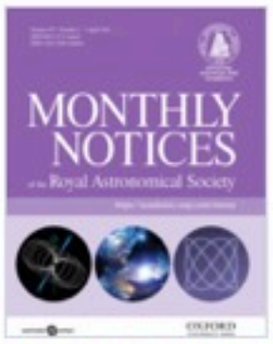
Figure 3 compares the proportions of QSO models found to exceed the 1% *fap* level for each of the cadences. There is a marked difference between the various cadences here, showing that some of the cadences allow many more periodogram results for which the 1% *fap* level is exceeded. This may be because the variations in the sampling pattern and the "windowing" effect of the Lomb-Scargle periodogram can cause a resonance effect at some test frequencies. Therefore, the results obtained by using the Lomb-Scargle periodogram method can be greatly influenced by the pattern of the observations used.

- even when we can "beat down" the intrinsic fluctuations in the periodogram, bias – in the form of leakage and aliasing

Drawbacks of Spectral Analysis

- Time domain usually put the equations at the form of state space, then you're capable of looking "inside" of your system- the internal variables that are invisible when you use frequency domain(transfer function).
- But when you study the system in frequency domain you stabilish a relationship beetween input and output by laplace transform.
- they impose strong restrictions regarding the possible processes underlying the dynamics of the series (e.g. stationarity),
- all information from the time-domain representation is lost in the operation

MONTHLY NOTICES
of the Royal Astronomical Society



Volume 475, Issue 2
April 2018

Oscillatory patterns in the light curves of five long-term monitored type 1 active galactic nuclei

Andjelka B Kovačević ✉, Ernesto Pérez-Hernández, Luka Č Popović ✉, Alla I Shapovalova, Wolfram Kollatschny, Dragana Ilić ✉

Monthly Notices of the Royal Astronomical Society, Volume 475, Issue 2, 1 April 2018, Pages 2051–2066,
https://doi.org/10.1093/mnras/stx3137
Published: 31 January 2018 **Article history** ▾

HYBRID METHOD: 3 STEPS

OUGP=Ornstein–Uhlenbeck Gaussian Process

OUGP(raw tseries1)=tseries1

OUGP(raw tseries2)=tseries2

0 STEP: raw data preprocessing

$$\mathrm{CWT}(a,b) = \frac{1}{\sqrt{2\pi a}} \int_{-\infty}^{+\infty} x(t) \mathrm{e}^{[(t/b)^2/a/2]\beta^2} \mathrm{e}^{\mathrm{i}\omega(t-b)/a} \mathrm{d}t,$$

$$\mathrm{env}(a,b) = \sqrt{\mathrm{Re}\{[\mathrm{CWT}(a,b)]^2\} + \mathrm{Im}\{[\mathrm{CWT}(a,b)]^2\}},$$

1 STEP

SpearmanCorrCoeff(env(tseries1),env(tseries2)) 2 STEP

The next step is to calculate the correlation coefficients of the envelopes of the wavelet coefficients of each light curve at each wavelet scale using the Spearman rank correlation coefficient. The Spearman coefficient measures statistical dependence

APPLIED ON ASTRONOMICAL DATA..:

- Gapped/irregular data
- Diurnal, monthly, season and yearly cycles
- satellite orbital cycles
- telescope allocations
- measurement errors & heteroscedastic

  Signal-to-noise ratio ratio differs from point to point

ARP 102 B

observed: 1987–2010

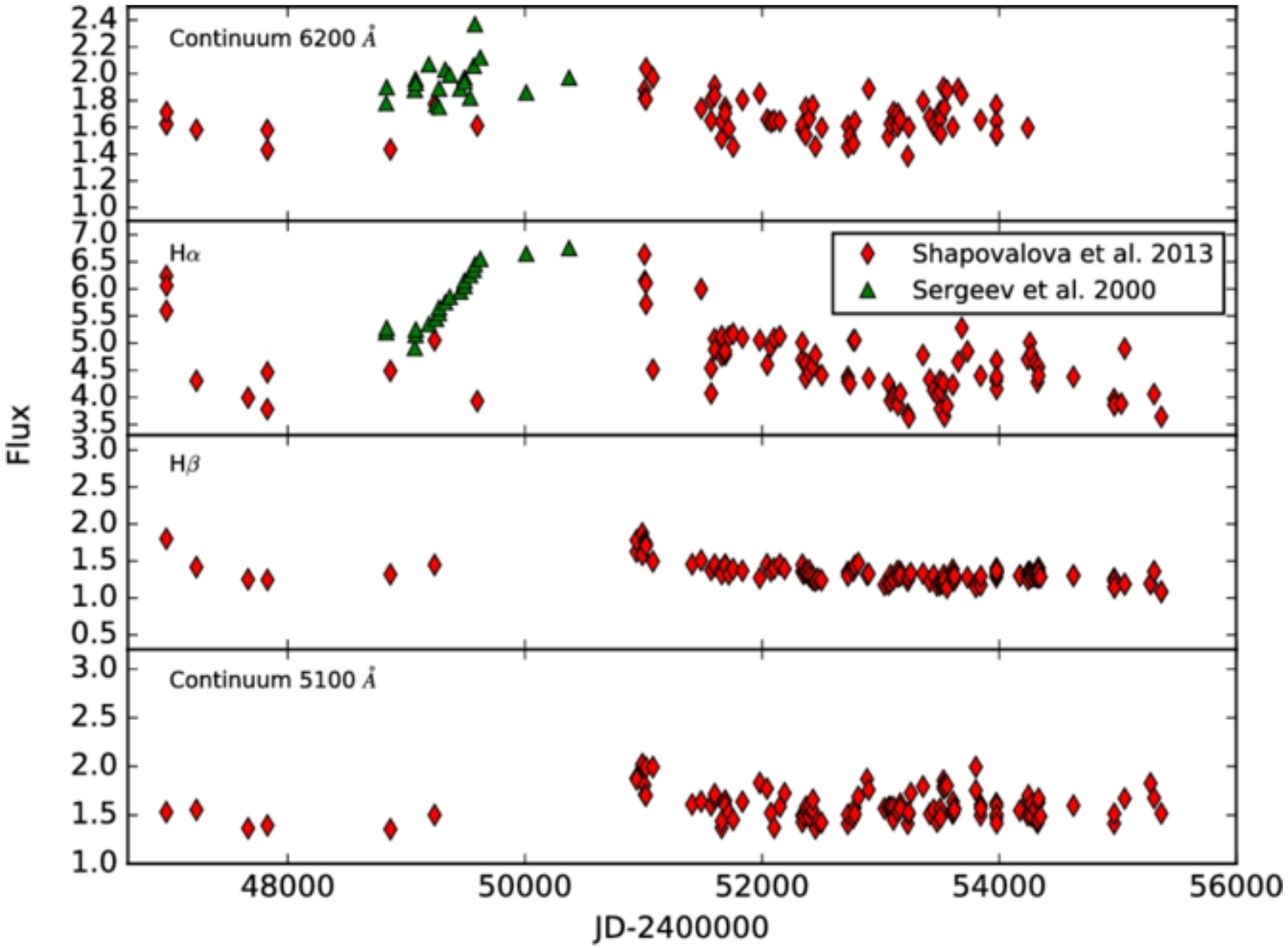
sampling:

78.1days

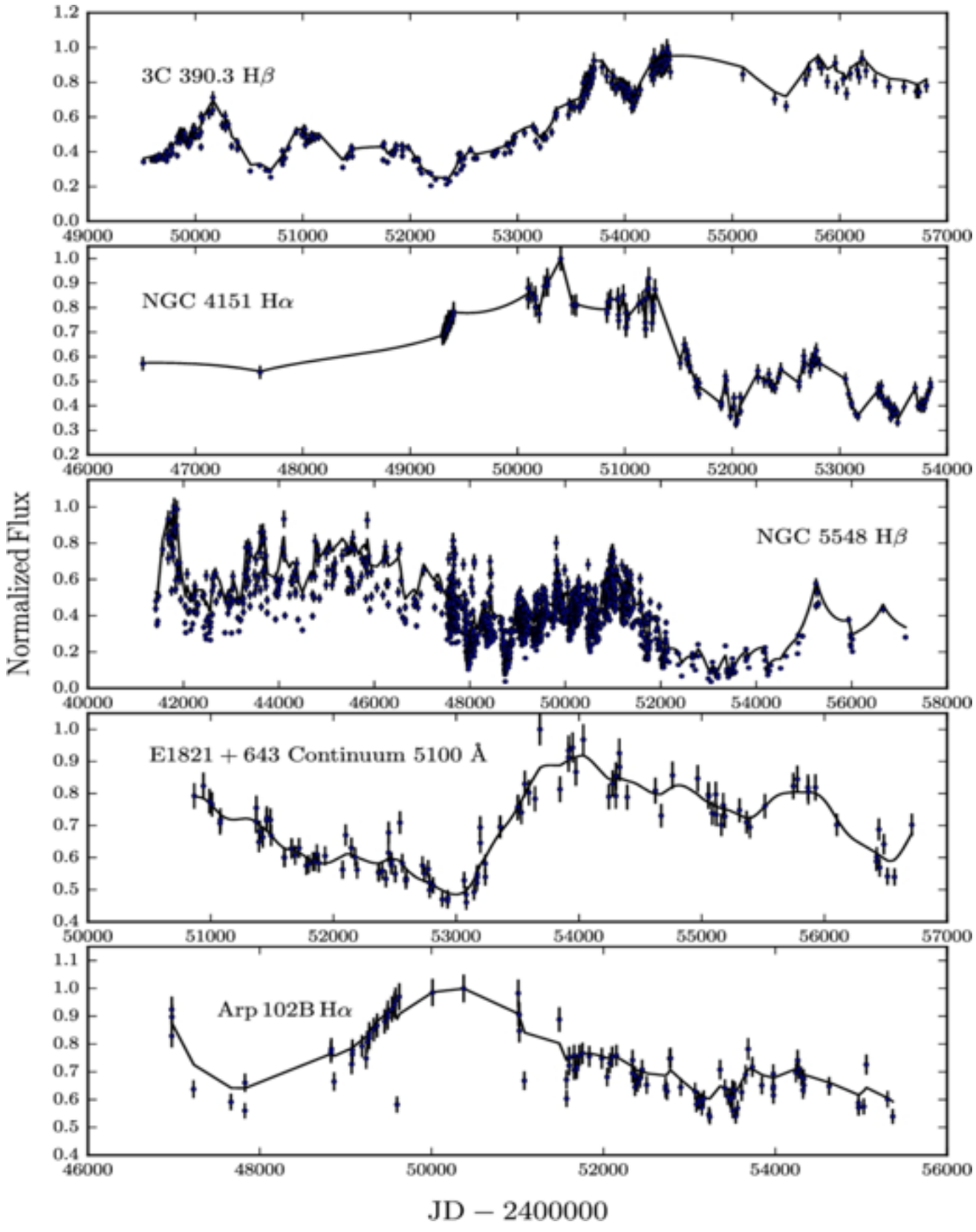
77 days

73 days

60 days

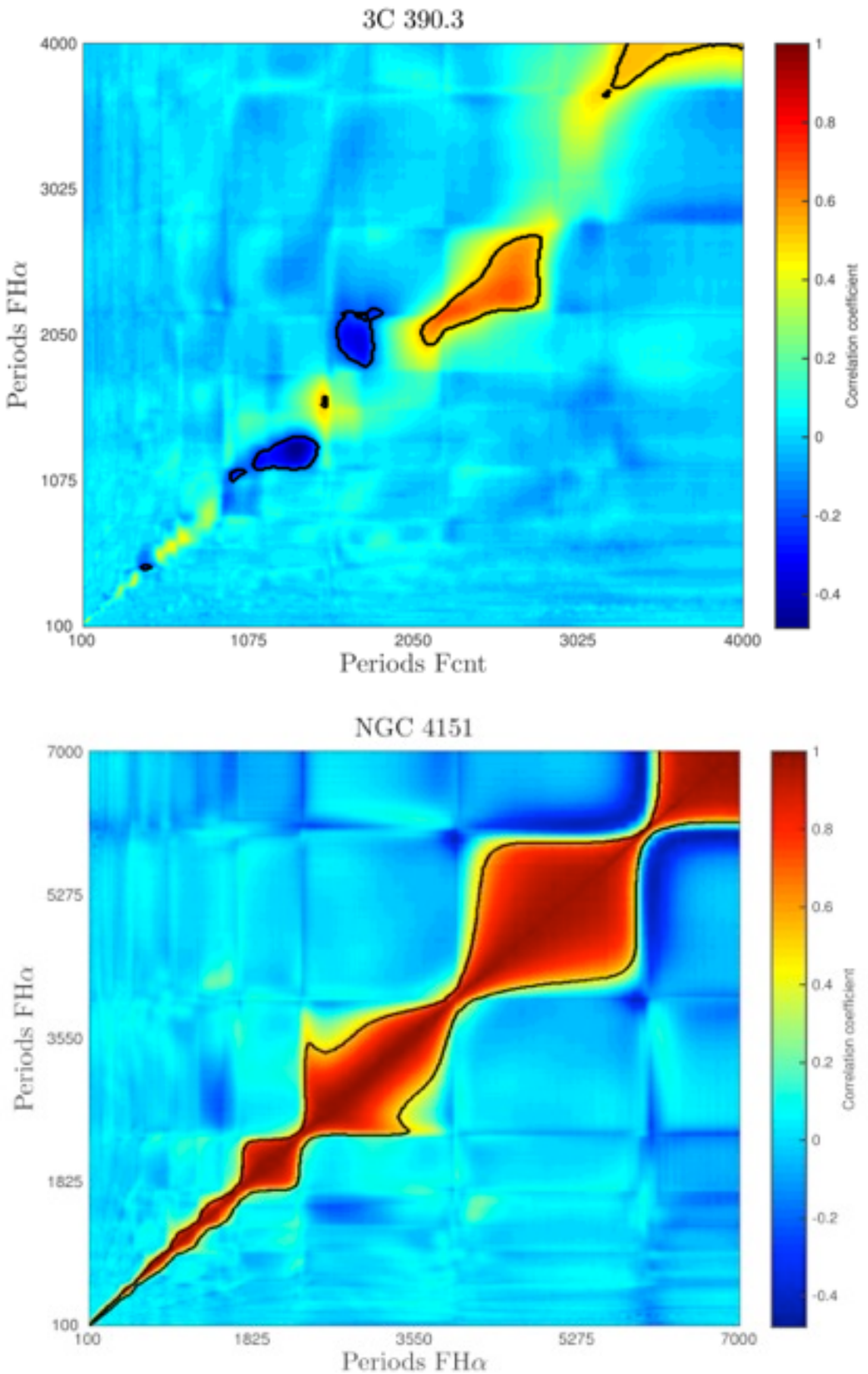


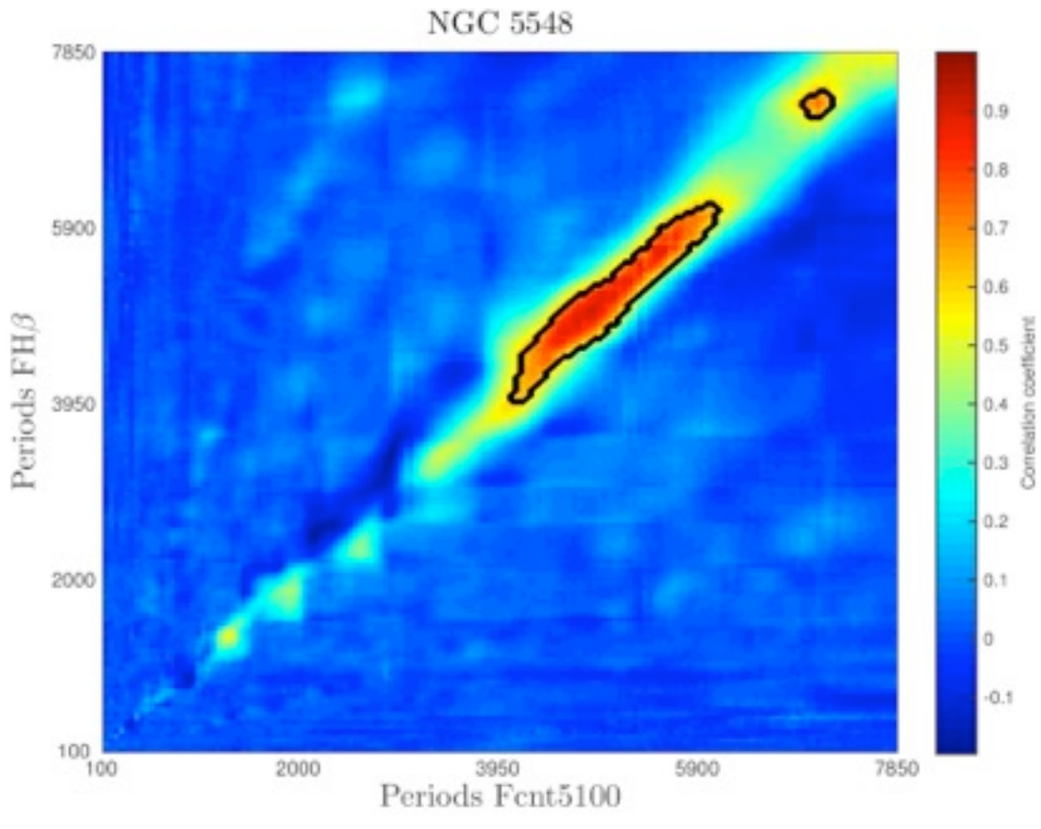
0 STEP: RAW DATA PREPROCESSING

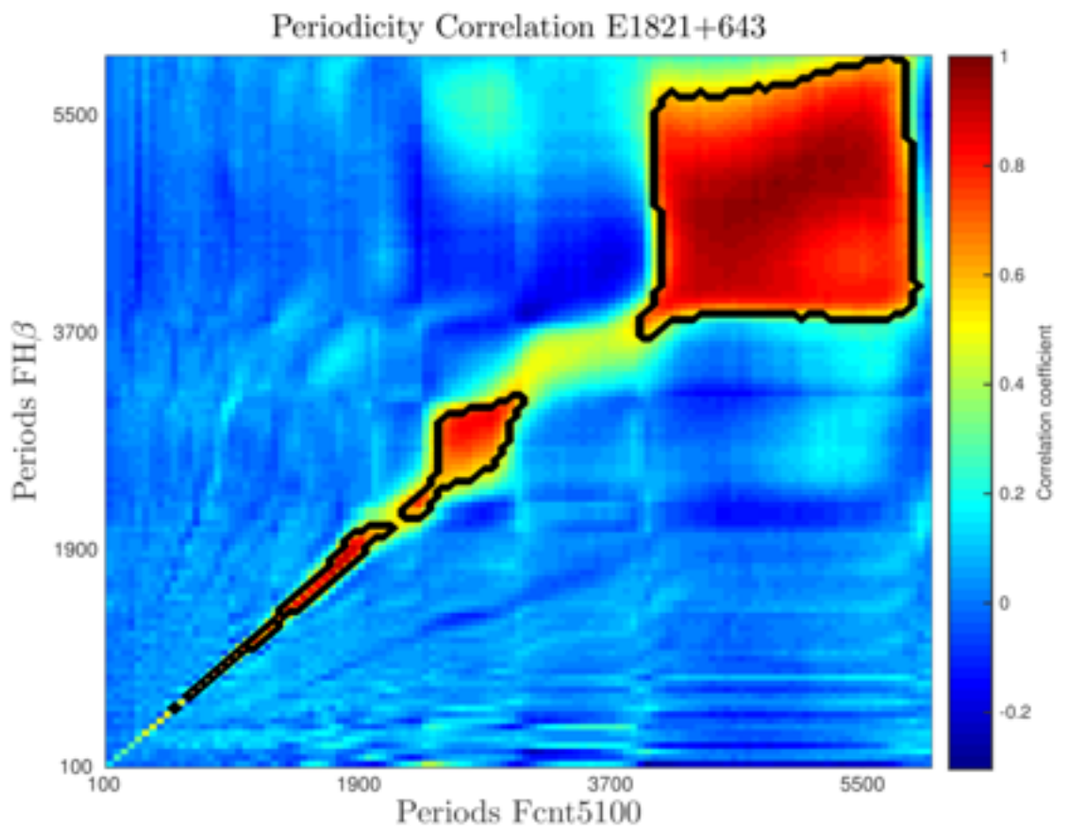


Examples of the comparison between the GP posterior mean (solid line) and observed light curves (dots) for the five objects.

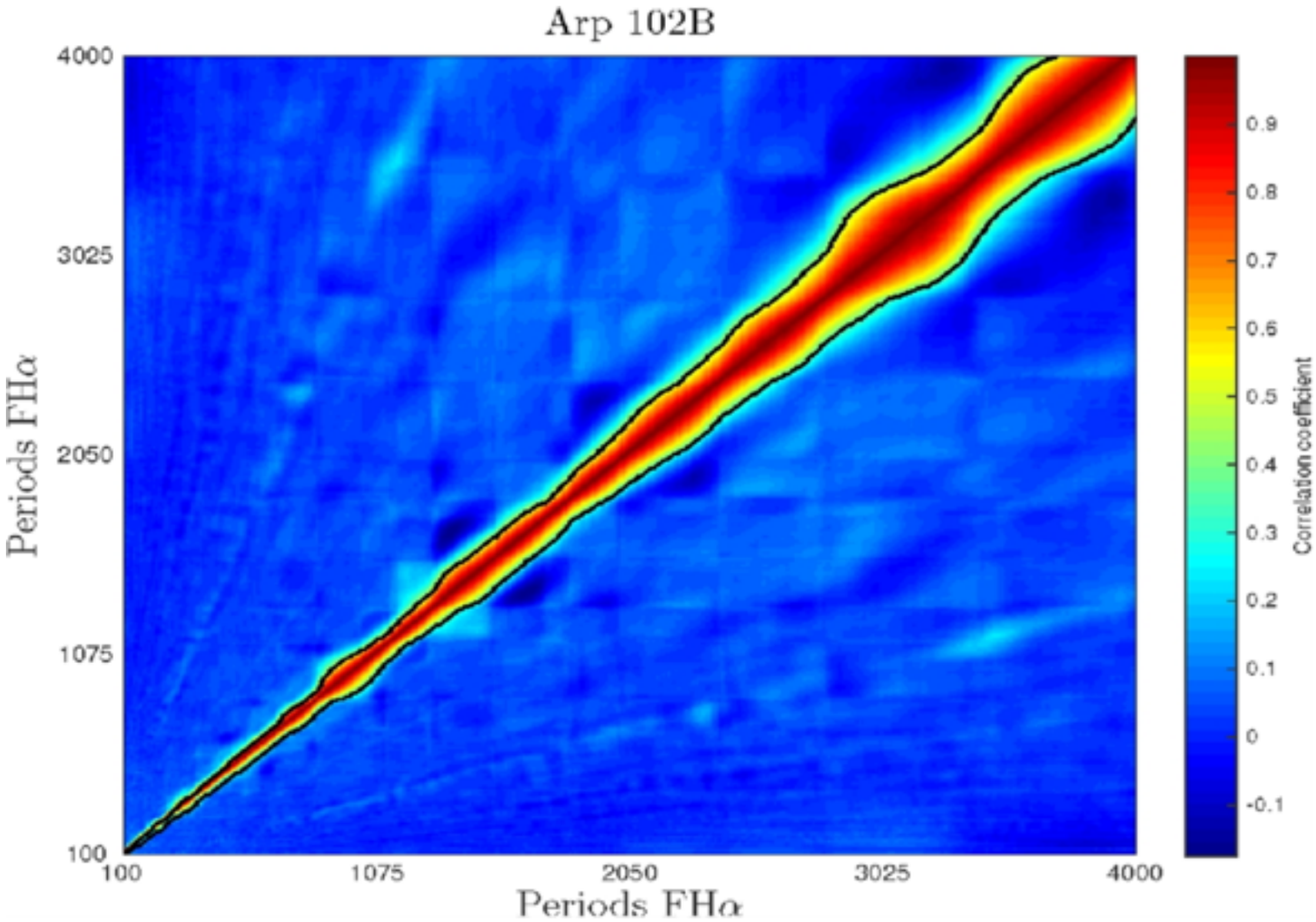
1&2 STEP: RESULTS







Object name	CLC1	CLC2	$P \pm \Delta P$ (yr)	r	95% CI	p
3C 390.3	Continuum 5100 Å	Hα	9.5 ± 0.3	0.5	(0.49,0.51)	<0.00001
			7.2 ± 1.2	0.69	(0.68,0.7)	<0.00001
			6.3 ± 0.9	0.68	(0.67,0.69)	<0.00001
			4.0 ± 0.04	−0.47	(−0.48, −0.45)	<0.00001
			5.44 ± 0.1	−0.35	(−0.37, −0.33)	<0.00001
		Hβ	10.11 ± 0.1	0.77	(0.76,0.78)	<0.00001
			7.67 ± 0.02	0.71	(0.7,0.72)	<0.00001
			6.42 ± 1.6	0.75	(0.74,0.76)	<0.00001
			5.43 ± 0.8	−0.47	(−0.48, −0.45)	<0.00001
			3.6 ± 0.4	−0.33	(−0.35, −0.31)	<0.00001
	Continuum 1370 Å	Lyα	10.34 ± 0.1	−0.47	(−0.49, −0.45)	<0.00001
			7.1 ± 0.02	−0.53	(−0.54, −0.51)	<0.00001
			6.25 ± 1.42	0.77	(0.76,0.78)	<0.00001
		CIV	9.42 ± 0.02	0.85	(0.84,0.86)	<0.00001
			7.84 ± 0.02	−0.6	(−0.61, −0.59)	<0.00001
			6.4 ± 1.22	0.85	(0.84,0.86)	<0.00001
			4.68 ± 0.7	−0.42	(−0.44, −0.40)	<0.00001
			3.4 ± 0.4	0.75	(0.74,0.76)	<0.00001



Note the prominent stationarity of the diagonal correlation line and the absence of correlation clusters.

Object name	CLC1	CLC2	$P \pm \Delta P$ (yr)	r	95% CI	p
Arp 102B	Continuum 6200 Å	Hα	–	–	–	
	Continuum 5100 Å	Hβ	–	–	–	

CHECKING RESULTS ON 2 LEVELS

- **1L**>non linear least square fitting of multisinusoidal models to the observed light curves

$$y = \sum_{i=1}^{n} c_i \sin\left(\frac{2\pi t}{p_i} + \phi_i\right) + B.$$

- **2L**>comparison of dynamics of observed light curves and time-series models

MODEL1 (**linearly** coupled oscillators of 2 units)

$$U_a(t) = A(t)\sin(2\pi f_a t + \phi) + cp_{b\to a} \quad (3)$$

$$\times\, B(t)\sin(2\pi f_b t + 2\pi f_b \tau) + W(t) \quad (4)$$

$$U_b(t) = B(t)\sin(2\pi f_b t) + cp_{a\to b} \quad (5)$$

$$\times\, A(t)\sin(2\pi f_a t + 2\pi f_a \tau + \phi) + W(t). \quad (6)$$

MODEL 2 (**linearly** coupled oscillators of 3 units)

$$U_a(t) = A(t)\sin(2\pi f_a t + \phi) + cp_{b\to a} \quad (7)$$

$$\times\, B(t)\sin(2\pi f_b t + 2\pi f_b \tau) + cp_{c\to a} \quad (8)$$

$$\times\, C(t)\sin(2\pi f_c t + 2\pi f_c \tau_1) + W(t) \quad (9)$$

$$U_c(t) = B(t)\sin(2\pi f_b t) + C(t)\sin(2\pi f_c t) + cp_{a\to b} \quad (10)$$

$$\times\, A(t)\sin(2\pi f_a t + 2\pi f_a \tau + \phi) + cp_{a\to c} \quad (11)$$

$$\times\, A(t)\sin(2\pi f_a t + 2\pi f_a \tau_1 + \phi_1) + W(t). \quad (12)$$

- MODEL3 (**nonlinearly** coupled oscillators of 3 units)

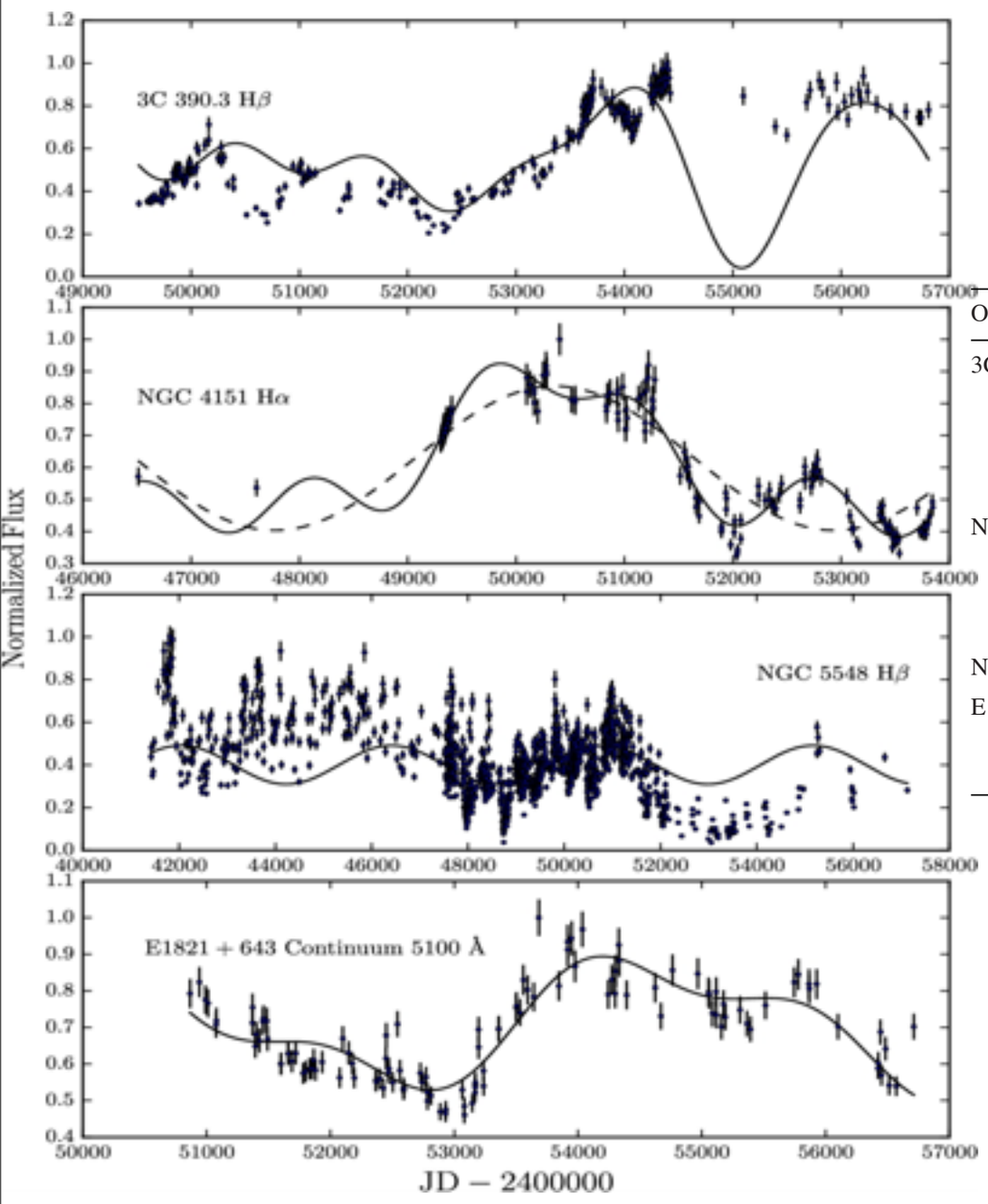
$$U_a(t) = A(t)\sin(2\pi f_a t + \phi) + cp_{b\to a} \quad (16)$$

$$\times\, B(t)\sin(2\pi f_b t + 2\pi f_b \tau) + W(t) \quad (17)$$

$$U_b(t) = B(t)\sin(2\pi f_b t) + cp_{a\to b} \quad (18)$$

$$\times\, U_a(t)^2 + W(t). \quad (19)$$

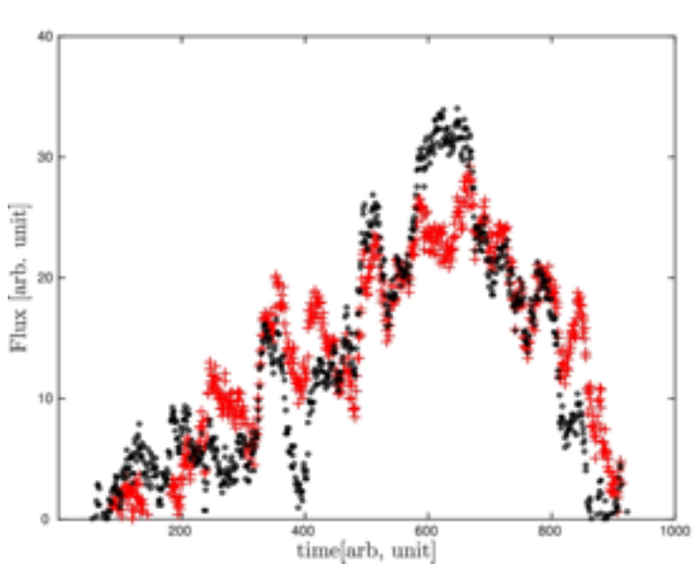
IL of CHECKING:

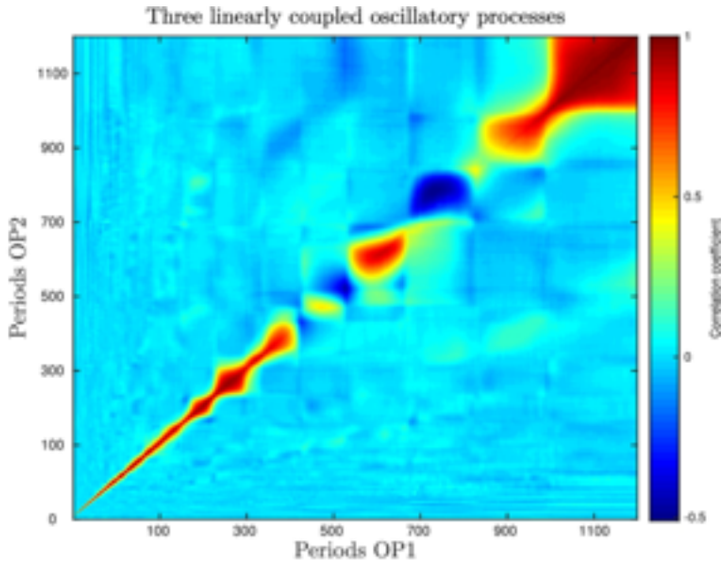


Estimated parameters of sinusoidal best-fitting of normalized observed light curves.

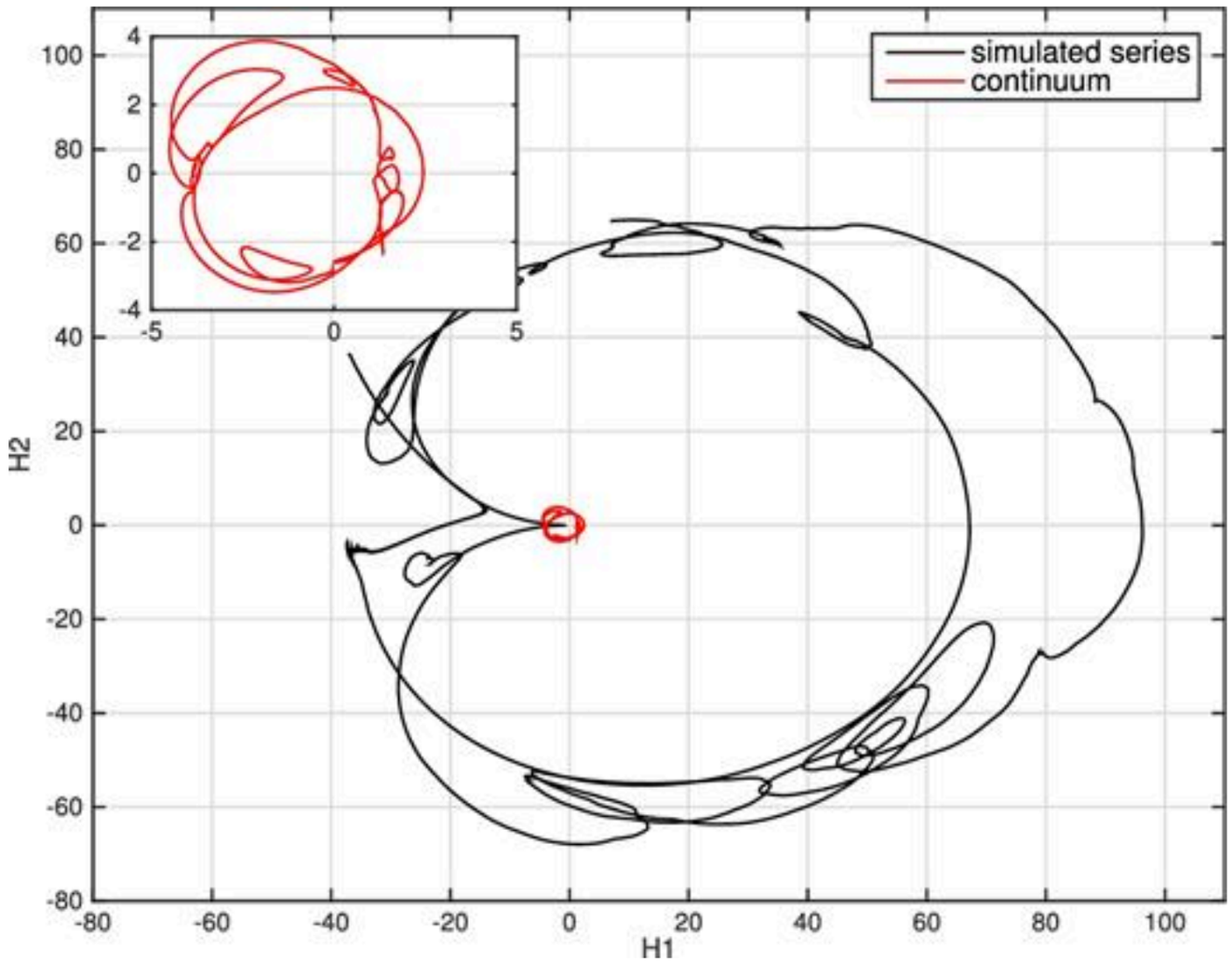
Object name	LC	c_i	p_i (d)	ϕ_i (rad)	B	r	χ^2
3C 390.3	Hβ	0.11 ± 0.02	3760 ± 7	6.02 ± 0.01	0.52 ± 0.01	0.81	4.748
		0.05 ± 0.03	2743 ± 15	5.51 ± 0.03			
		0.29 ± 0.04	2300 ± 2	5.47 ± 0.03			
		0.17 ± 0.03	2000 ± 2	0.17 ± 0.005			
		0.08 ± 0.01	1322 ± 1	−5.24 ± 0.1			
NGC 4151	Hα	0.22 ± 0.01	5580 ± 435	1.52 ± 4.34	0.63 ± 0.02	0.96	0.381
		−0.07 ± 0.02	2730 ± 422	−4.20 ± 5.63			
		−0.08 ± 0.01	1534 ± 28	−4.02 ± 3.82			
	Hα	−0.23 ± 0.01	5165 ± 3	–	0.63 ± 0.01	0.87	1.275
NGC 5548	Hβ	−0.10 ± 0.01	4378 ± 70	−5.35 ± 1.12	0.40 ± 0.004	0.40	32.804
E1821 + 643	Continuum 5100 Å	0.16 ± 0.001	4511 ± 1	0.02 ± 0.005	0.71 ± 0.0	0.87	0.449
		0.50 ± 0.0002	2529 ± 0.005	1.57 ± 0.03			
		0.07 ± 0.005	1977 ± 0.1	1.10 ± 0.002			

2L of CHECKING :simulations



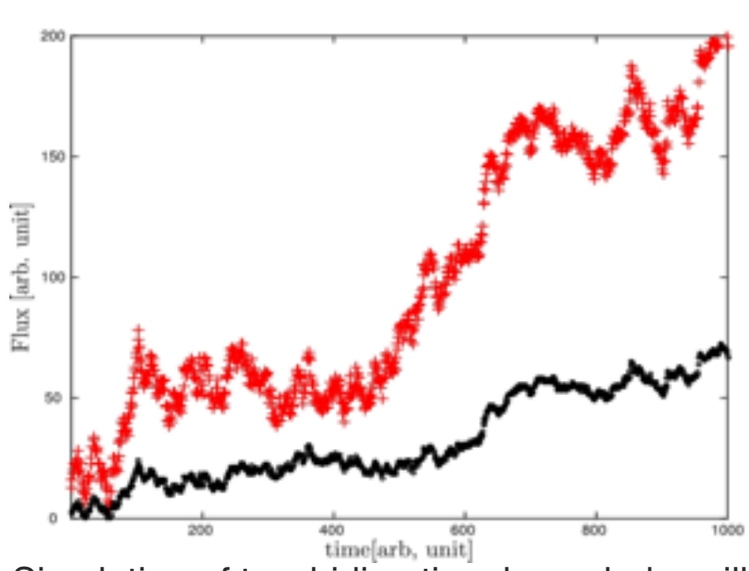


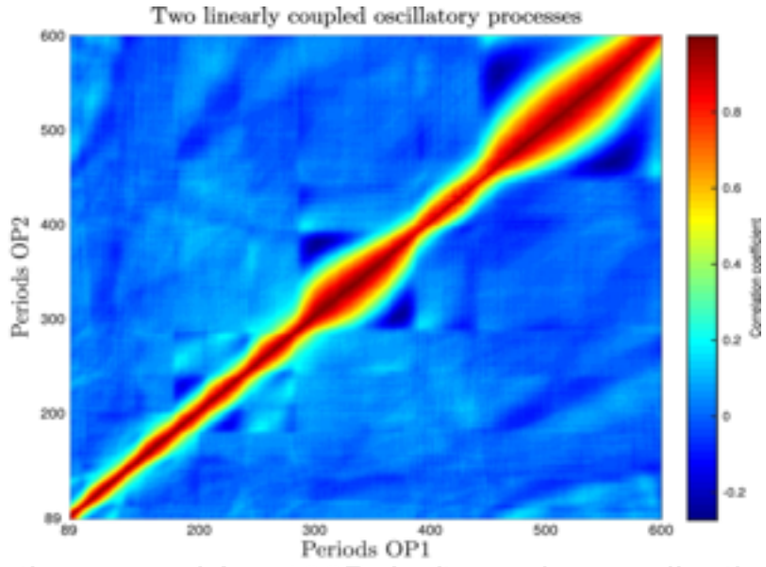
Simulation of bidirectional coupled three-oscillator network for the case of 3C 390.3.



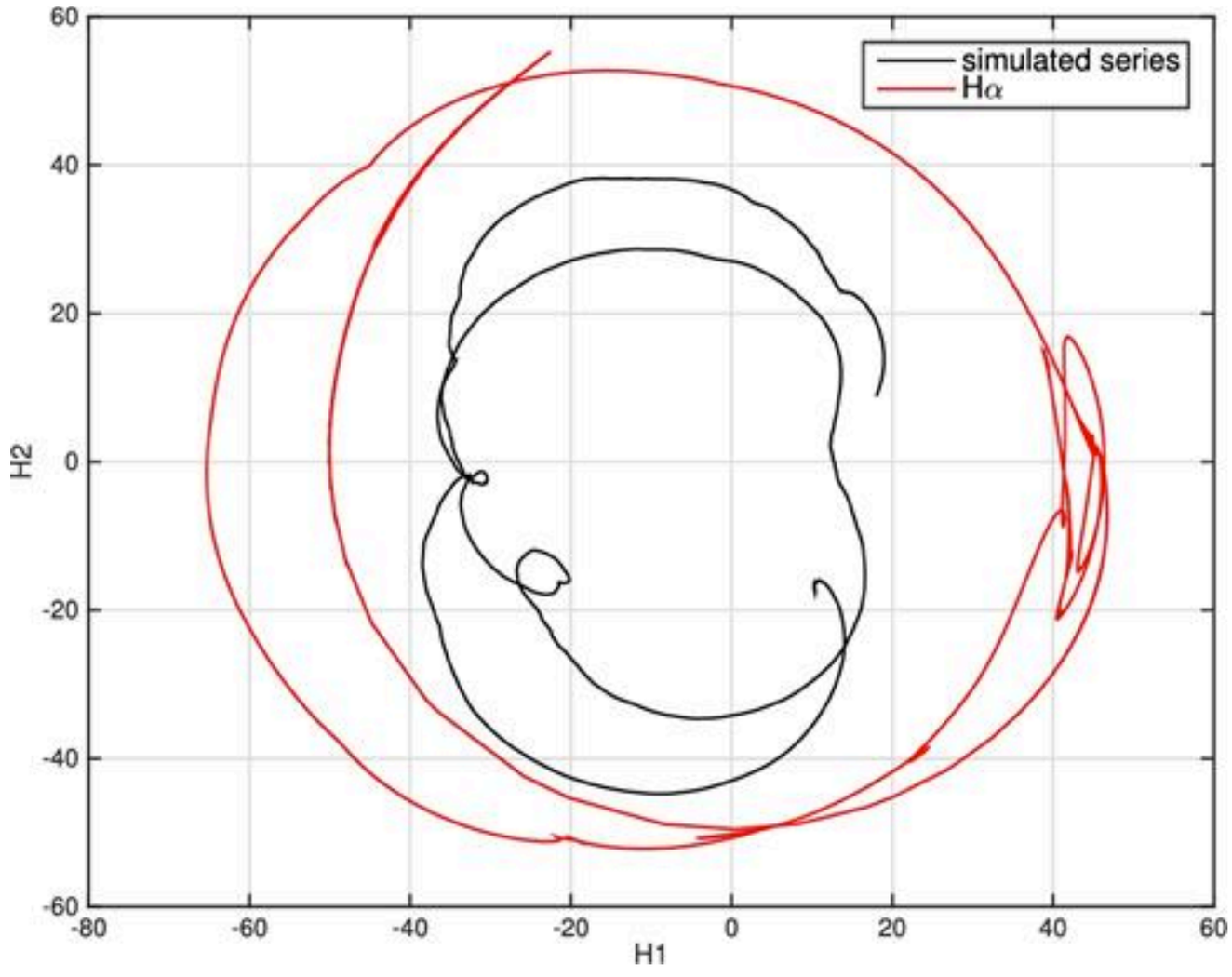
Comparison of the phase trajectories between the continuum of 3C 390.3 and simulated curve OP1 from the oscillatory network model

$$y^{\mathrm{H}}(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{y(t)}{t-\tau} \mathrm{d}\tau$$

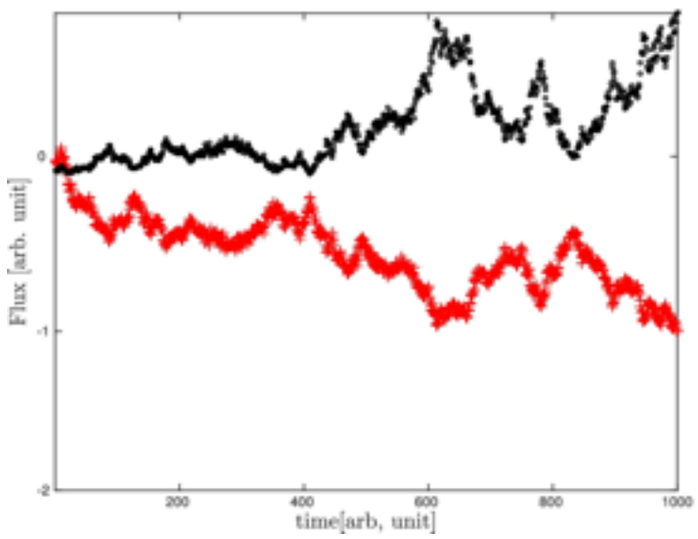


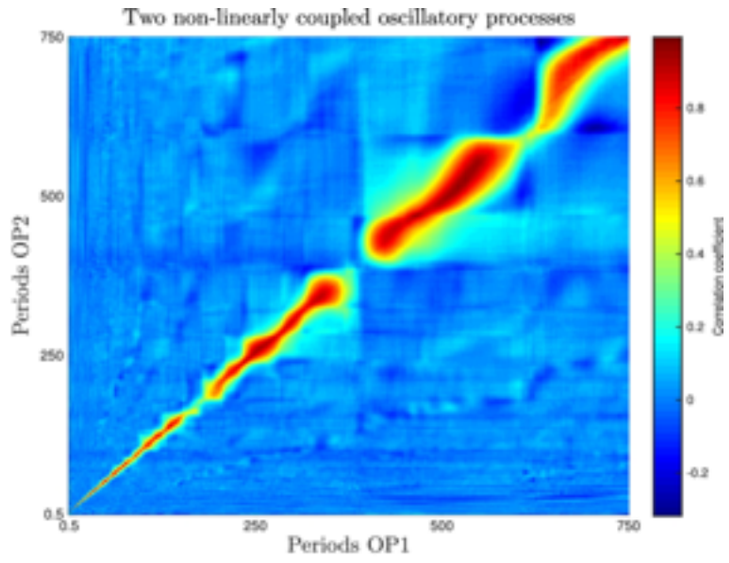


Simulation of two bidirectional coupled oscillators for the case of Arp 102B. Left: random realization of equation from two time series (black is $U_a = OP1$ and red is $U_b = OP2$) of amplitudes $A = 5.29$, $B = 1.99$, phase $\phi = 0.4174$ rad, coupling strengths $cp_{a \to b} = 0.4$, $cp_{b \to a} = 0.2$, time delay is 100 and periods are 500 and 300 arbitrarily chosen time units. Right: corresponding 2D correlation map.

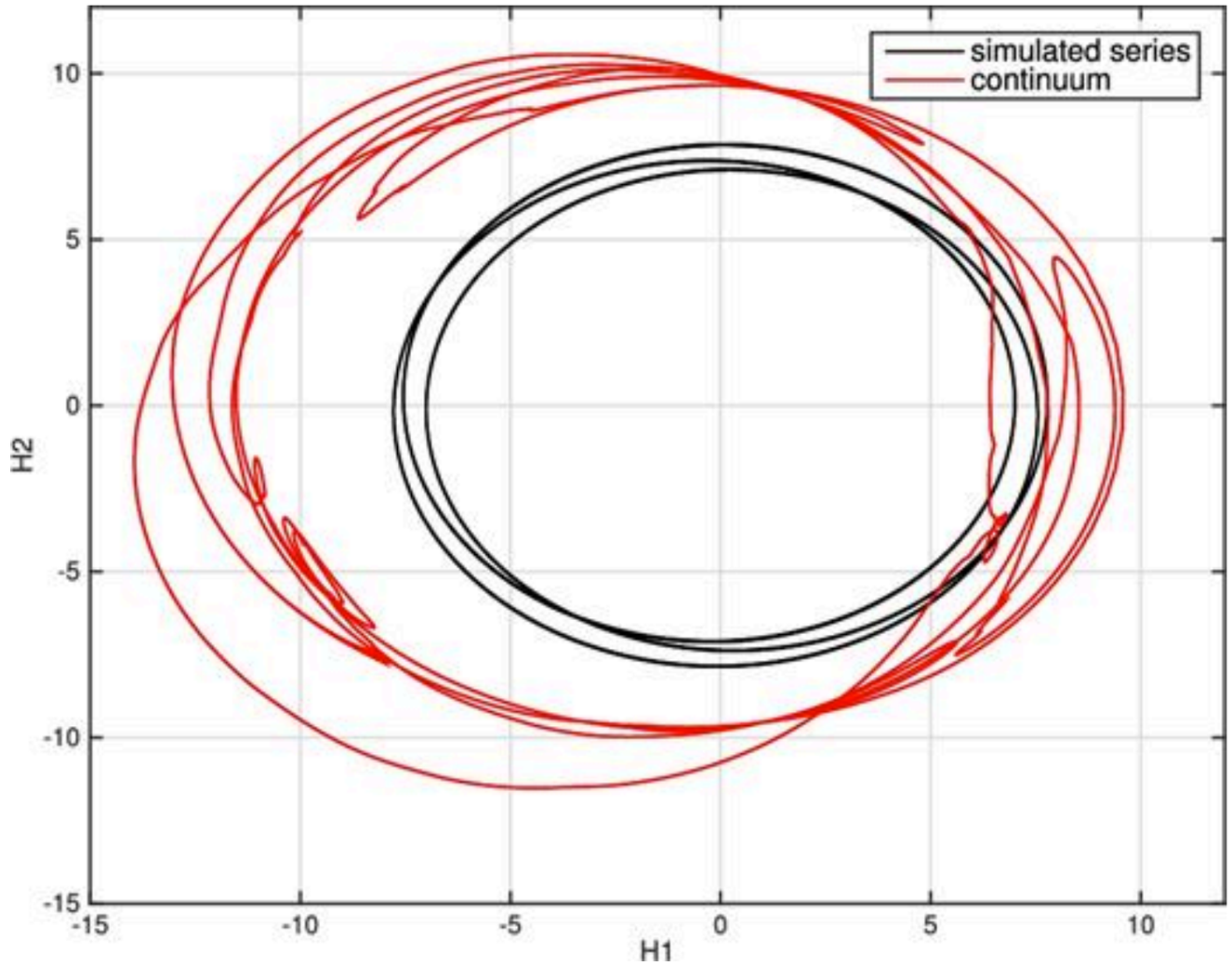


Both curves are similar and non-closed, indicating either weak coupling or the absence of periodicity. They appear to intersect themselves due to projection on to 2D phase space.

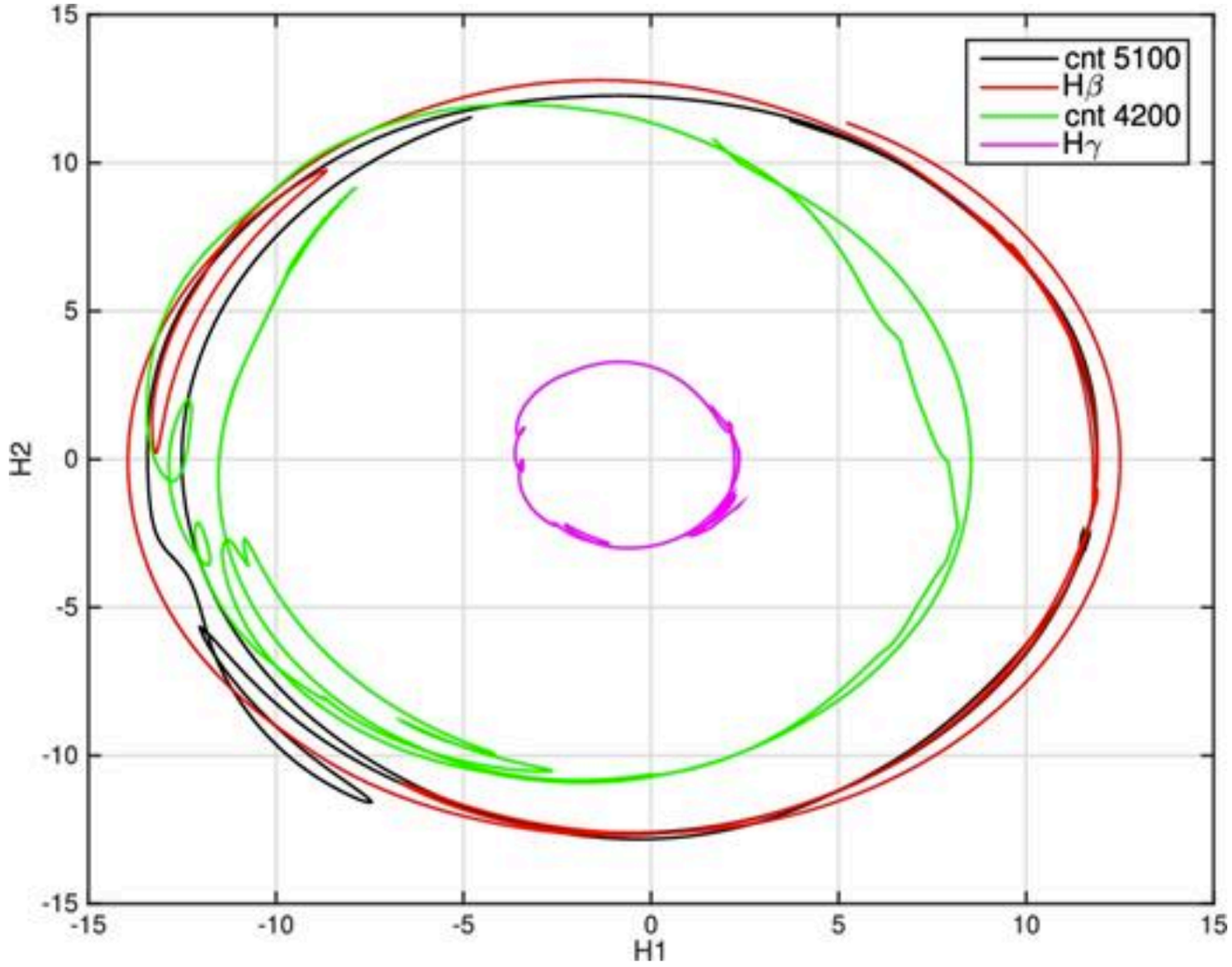




Simulation of two bidirectional coupled oscillators for the case of NGC 5548. Left: random realization of equation (19) from two time series (black is $U_a = OP1$ and red is $U_b = OP2$) of amplitudes $A = 5.92$, $B = 1.27$, phases $\phi = 2.65$ rad, coupling strengths $cp_{a \to b} = 0.7$, $cp_{b \to a} = 0.2$, periods 500 and 300 and time delay is 100 arbitrarily chosen time units



Note the chaotic-like appearance of both curves.



Its 2D correlation maps are similar to the case of NGC 4151. Particularly, if we look at phase portraits of the light curves normal limit cycles are observed in the dynamics of E1821 + 643. They are similar to the phase portrait of regular sinusoids. We note the presence of two smaller elongated loops in all phase curves reflecting two smaller periods.

CONCLUSIONS

- We provided a general hybrid method for mining periodicity which also allows us to discern dynamics of quasars
- This is useful where signals are highly variable or ‘noisy’ and where links are difficult to discern from comparison of the individual standard wavelet transforms
- The method does not require any a priori filtering assumptions or assuming certain values of periods
- Our method automatically provides information (both numerical and visual) on the interactions between oscillators, which is not readily available using other methods