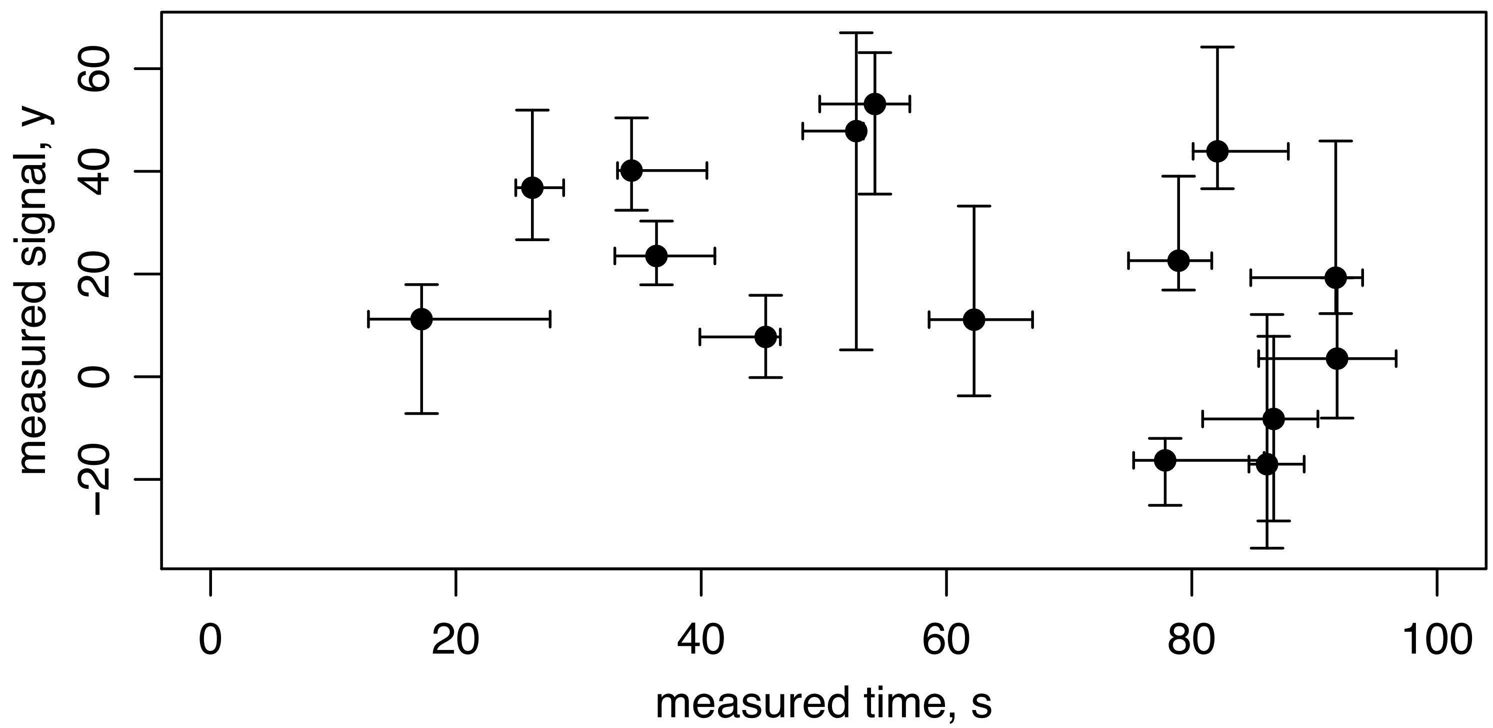
A Bayesian method for the analysis of deterministic and stochastic time series

Coryn Bailer-Jones
Max Planck Institute for Astronomy, Heidelberg

DPG, Berlin, March 2015

Time series modelling

- heteroscedastic, asymmetric noise on time and signal
- non-uniform time sampling



Measured data $D_j = (s_j, y_j)$ and uncertainties $\sigma_j = (\sigma_{s_j}, \sigma_{y_j})$

Model M with parameters θ

Likelihood of single data point: integrate over unknown true time (t) and signal (z)

$$P(D_j|\sigma_j,\theta,M) = \int_{t_j,z_j} \underbrace{P(D_j|t_j,z_j,\sigma_j)}_{\text{Measurement model}} \underbrace{P(t_j,z_j|\theta,M)}_{\text{Time series model}} dt_j dz_j$$

Model comparison

Likelihood of all data points is $$P(D|\sigma,\theta,M) = \prod_j P(D_j|\sigma_j,\theta,M)$$

Evidence is the likelihood marginalized over the parameter prior

$$P(D|\sigma,M) = \int_\theta \underbrace{P(D|\sigma,\theta,M)}_{\text{likelihood}} \; \underbrace{P(\theta|M)}_{\text{prior}} \; d\theta$$

More robust alternative is the *leave-one-out cross validation likelihood*

$$P(D_j|D_{-j},\sigma,M) = \int_\theta \underbrace{P(D_j|\sigma_j,\theta,M)}_{\text{likelihood}} \underbrace{P(\theta|D_{-j},\sigma_{-j},M)}_{\text{posterior}} d\theta$$

$$L_{\mathrm{CV}} = \prod_{j=1}^{j=J} P(D_j|D_{-j},\sigma,M)$$

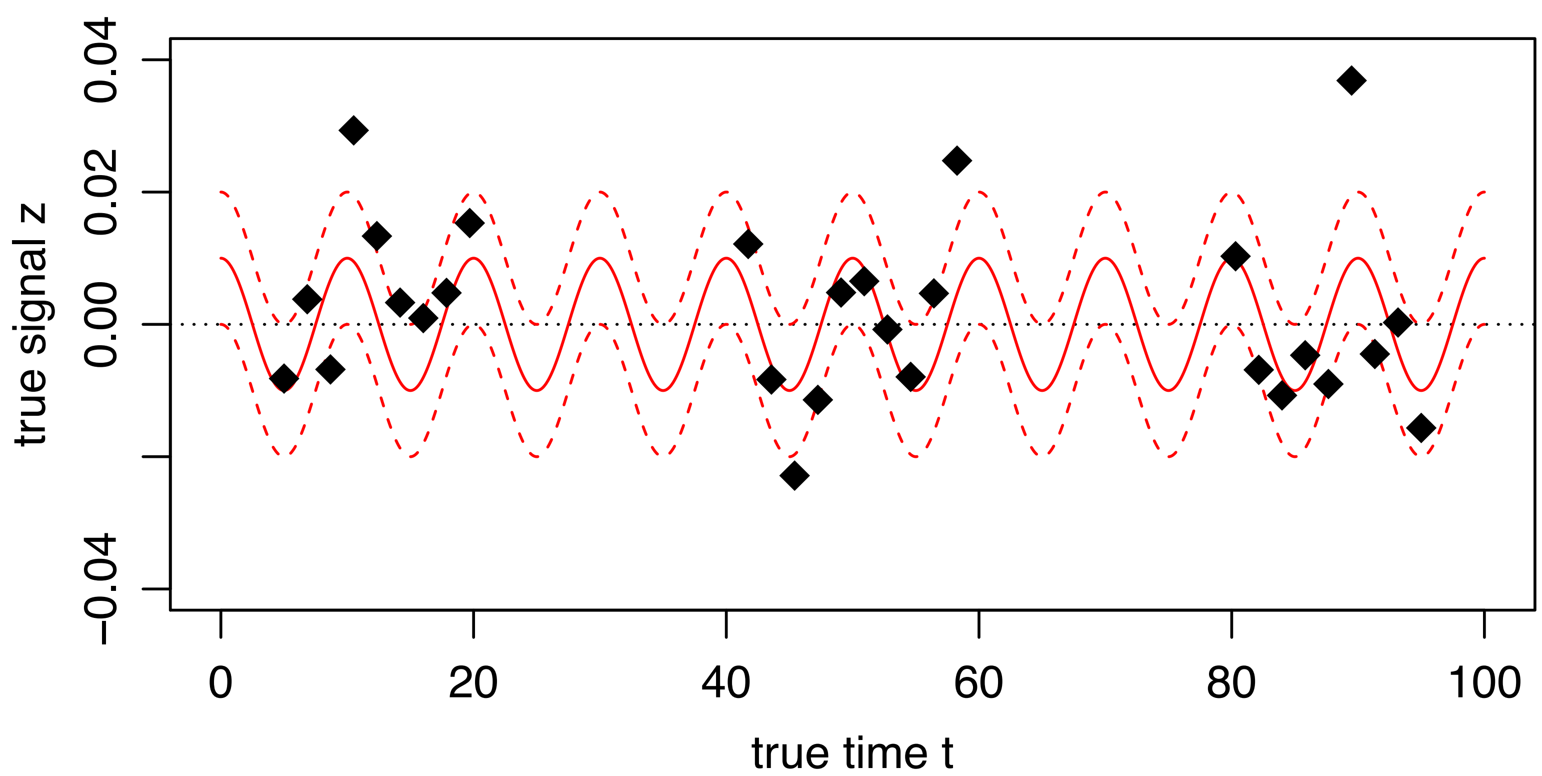
Calculate integrals by MCMC sampling of posterior

Time series model

Deterministic mean plus stochastic variation of constant variance

$$P(z_j|t_j, \theta, M) = \frac{1}{\sqrt{2\pi}\omega} e^{-(z_j - \eta(t_j))^2/2\omega^2} \quad \text{Gaussian}$$

$$\eta(t_j) = \frac{a}{2}\cos[2\pi(\nu t + \phi)] + b \quad \text{sinusoidal}$$



- red solid: deterministic component
- red dashed: standard deviation of stochastic component
- black: true data

Coryn Bailer-Jones, MPI for Astronomy, Heidelberg

Time series model

Ornstein-Uhlenbeck process

A Stationary, Markov, Gaussian process

$$dz(t) = -\frac{1}{\tau} z(t)dt + c^{1/2}\mathcal{N}(t; 0, dt)$$

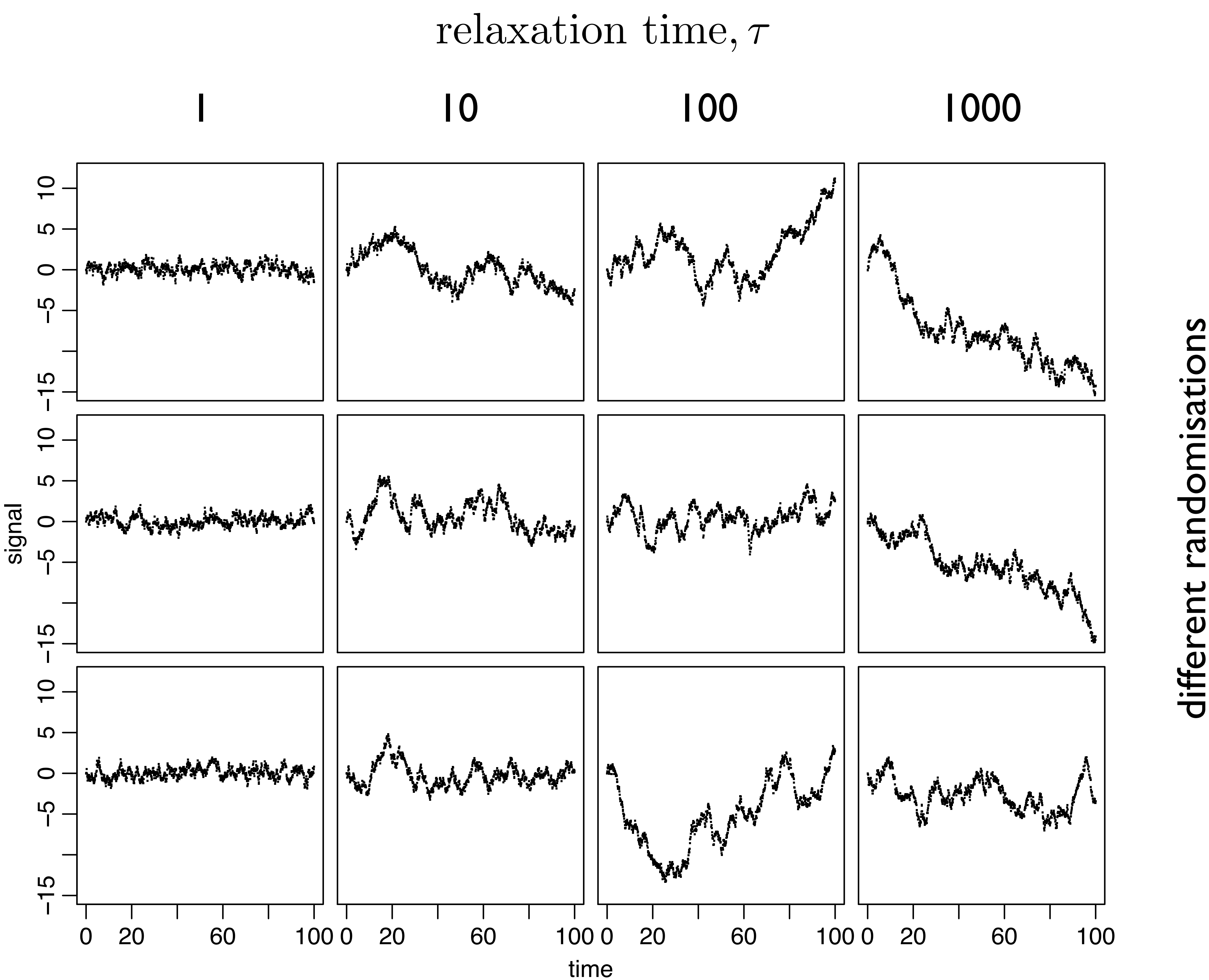
τ relaxation time

c diffusion constant

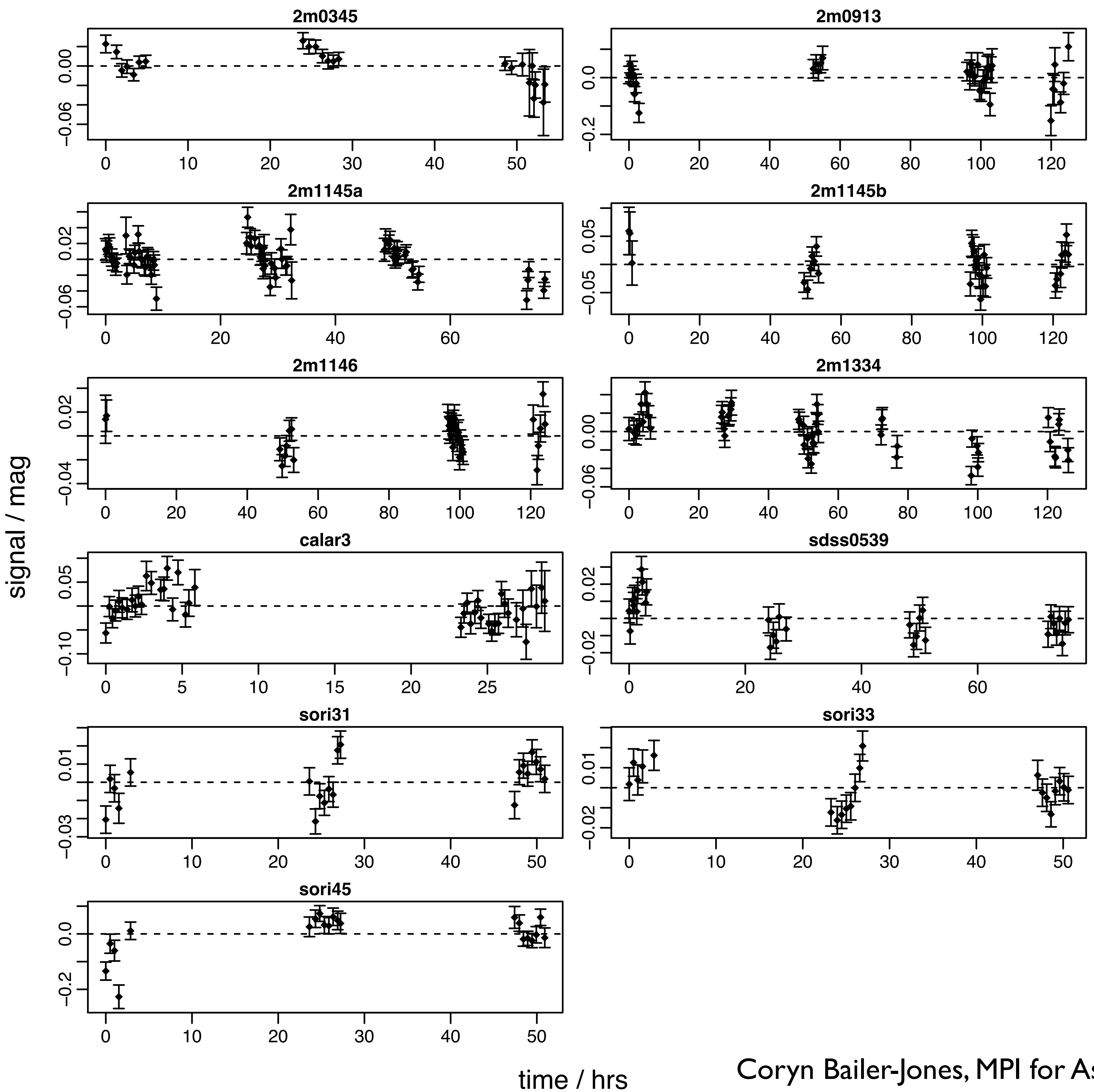
$$P(z_j|t_j, \theta, M) = \frac{1}{\sqrt{2\pi V_z}} e^{-(z_j-\mu_z)^2/2V_z} \quad \text{with}$$

$$\left.\begin{aligned} \mu_z &= z_0 \upsilon \\ V_z &= \frac{c\tau}{2}(1-\upsilon^2) \end{aligned}\right\} \quad \text{where} \quad \upsilon = e^{-(t-t_0)/\tau} \quad \text{for} \quad t > t_0$$

Examples of OU process realizations



Luminosity variations in ultra cool dwarf stars



Coryn Bailer-Jones, MPI for Astronomy, Heidelberg

Luminosity variations in ultra cool dwarf stars

Models compared:

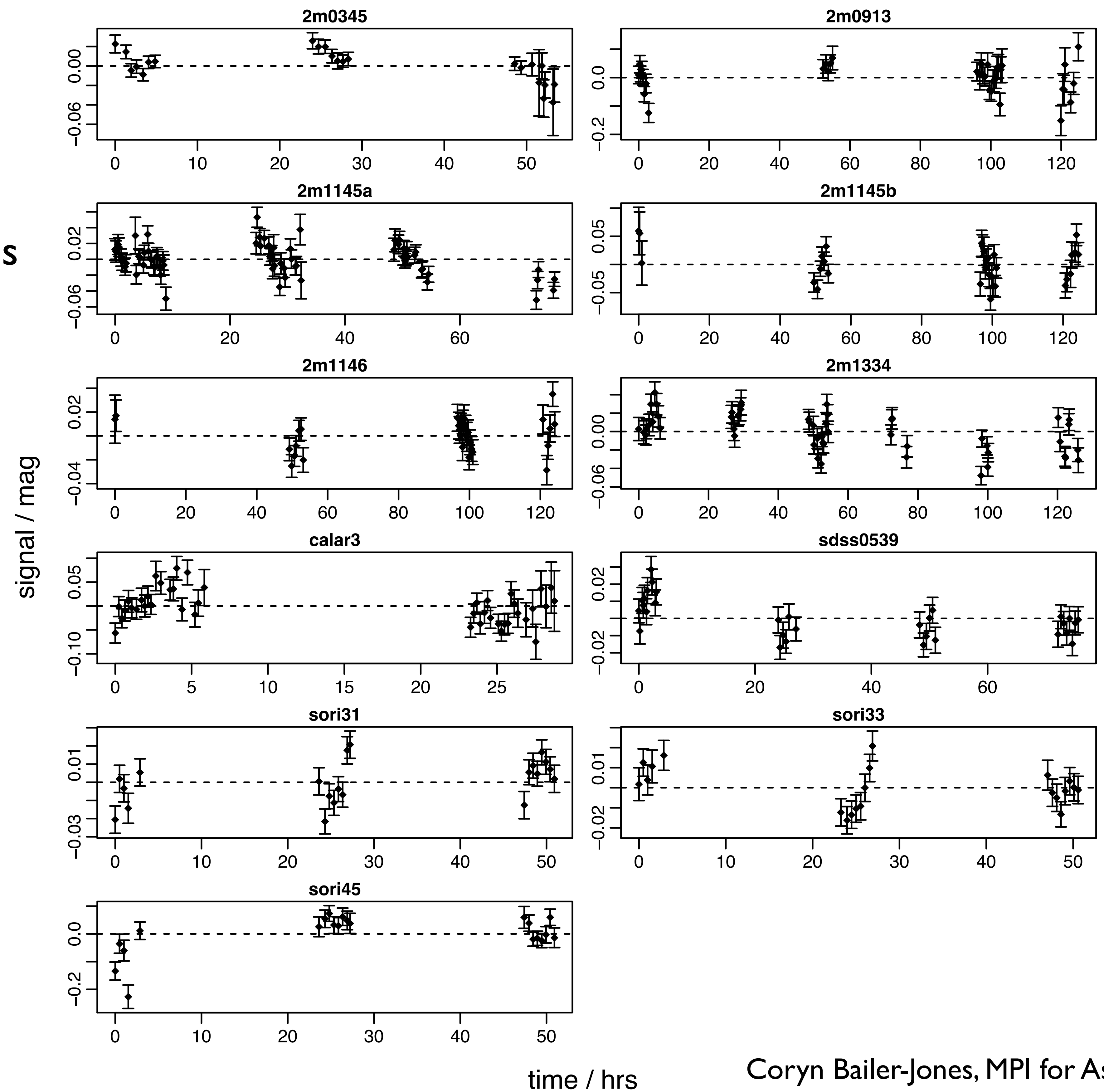
- constant (variability just due to measurement noise)
- constant with Gaussian stochastic component
- sinusoid with Gaussian stochastic component
- OU process

Luminosity variations in ultra cool dwarf stars

OU process

Sinusoid + stochastic

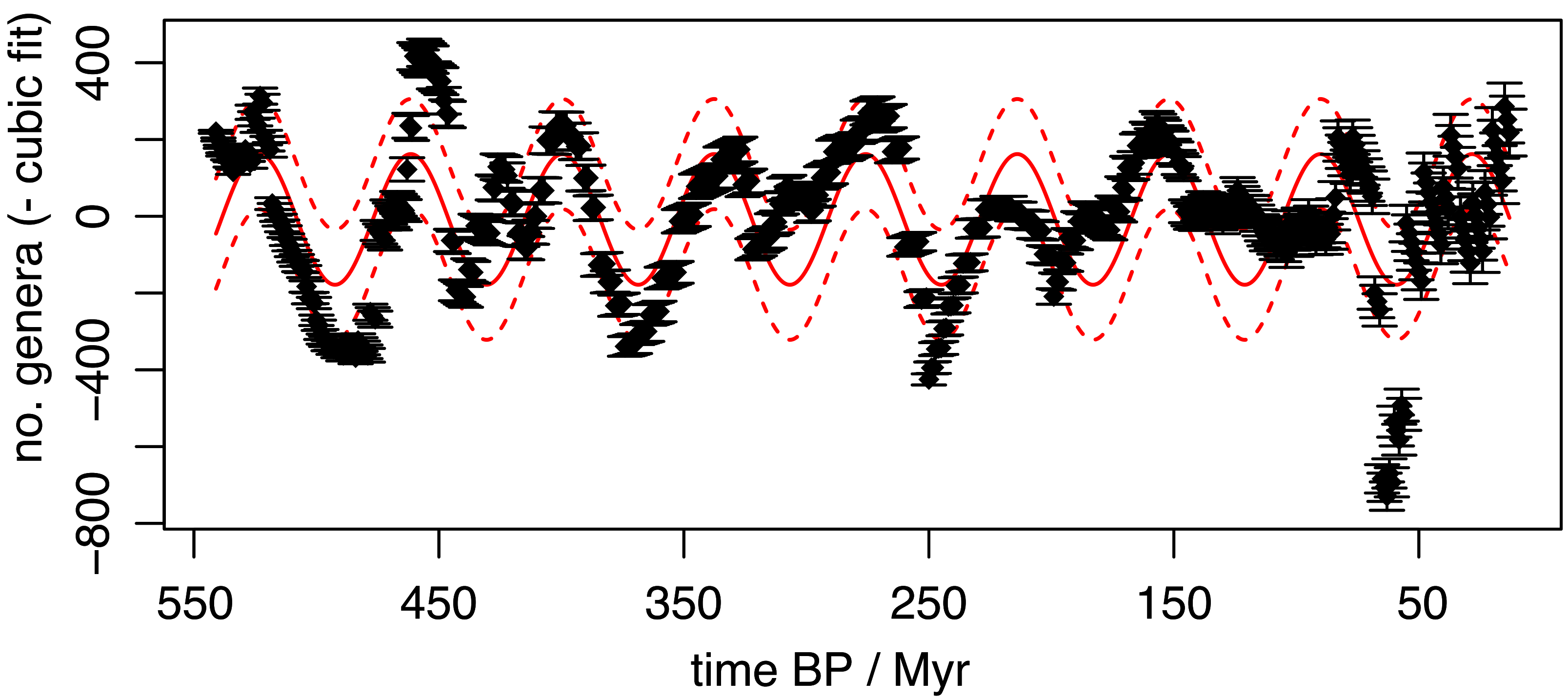
Sinusoid (8.3h, 13.3h)



Coryn Bailer-Jones, MPI for Astronomy, Heidelberg

Periodicity in biodiversity over past 550 Myr?

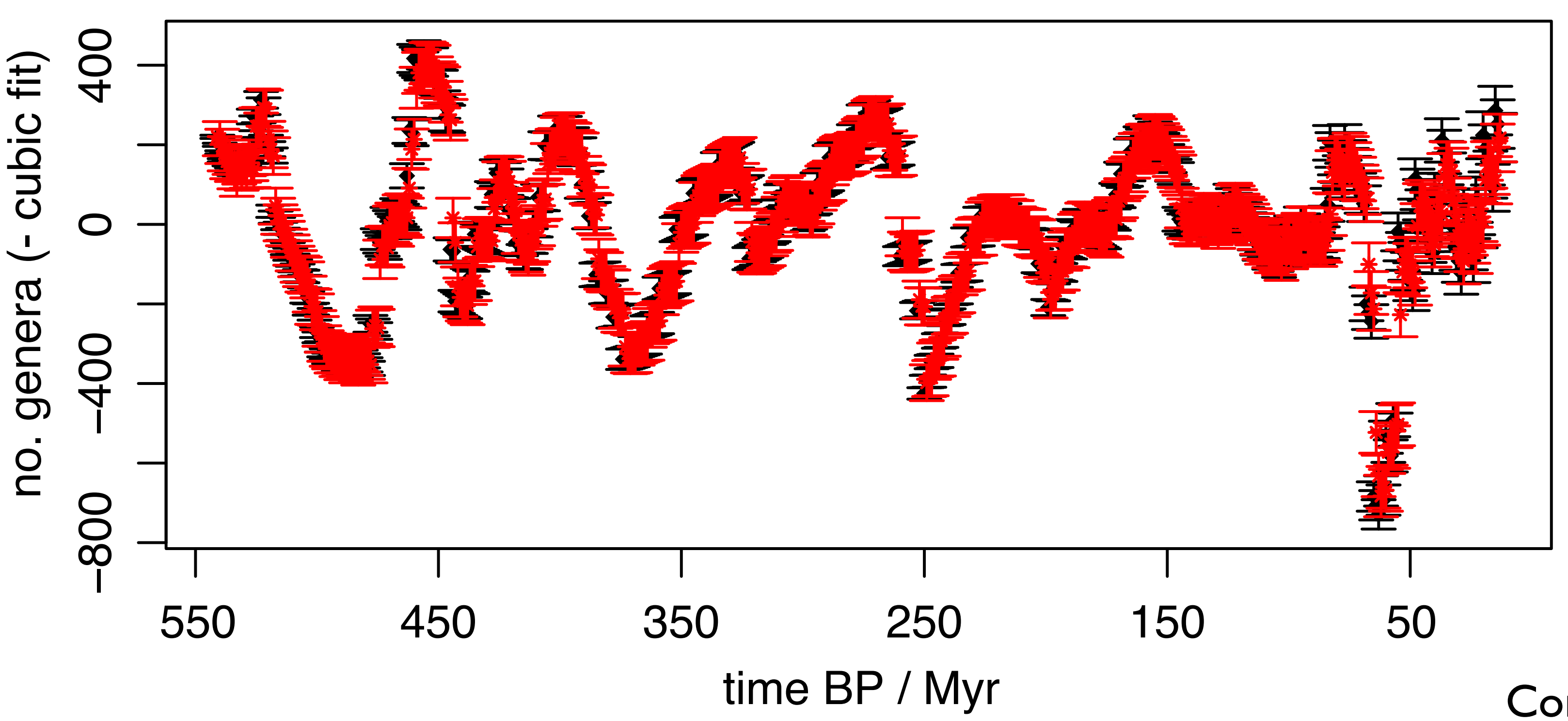
Rohde & Muller 2005



periodic model with additional fitted Gaussian noise

black = data

red = model fit



stochastic process (OU process)

CV likelihood is *much* higher for this model

Coryn Bailer-Jones, MPI for Astronomy, Heidelberg

Summary

- a Bayesian method for modelling times series
  - arbitrary time sampling and error models
  - deterministic and stochastic times series
  - use of cross-validation likelihood, a robust alternative to the evidence
- applications
  - light curves of some very cool stars (and quasars) evolve stochastically
  - no evidence for periodic variation of biodiversity over past 550 Myr
- more information and software: `tinyurl.com/ctsmod`

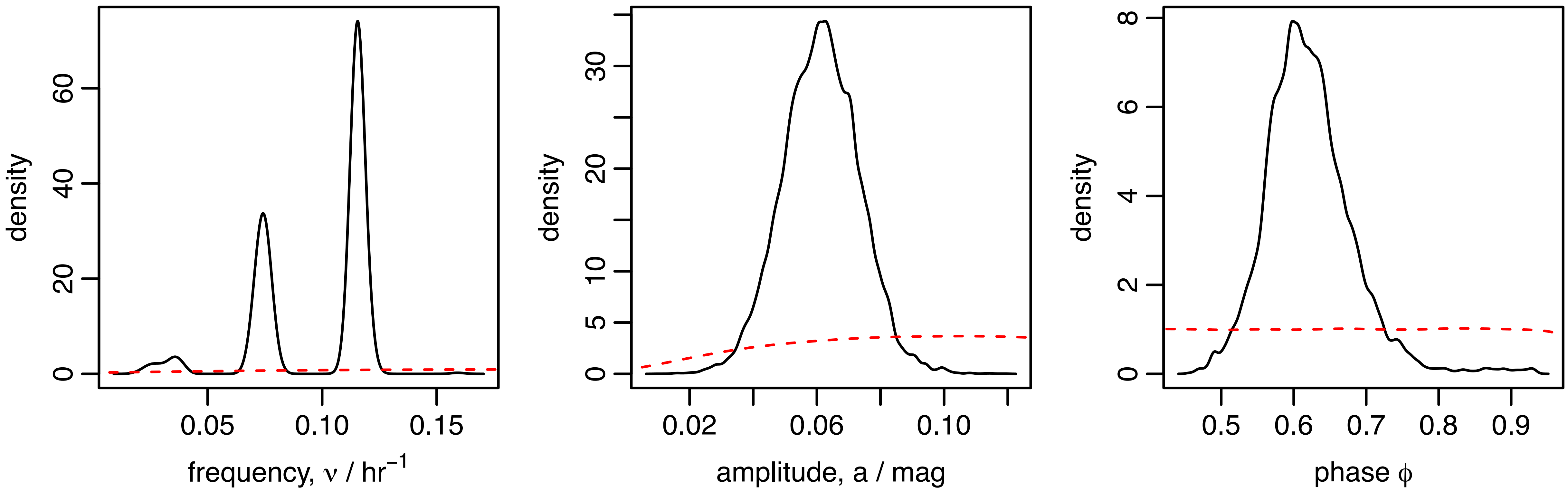
Ultra cool dwarf model comparison results

Table 4. Log (base 10) LOO-CV likelihood of each model relative to that for the no-model for each light curve ($\log L_{\mathrm{LOO-CV}} - \log L^{\mathrm{NM}}$).

Light curve	OUprocess	Off+Stoch	Sin	Sin+Stoch	Off+Sin+Stoch	No-model	p-value
2m0345	3.26	2.07	0.15	2.06	2.66	–13.60	4e-4
2m0913	0.44	0.72	0.23	0.97	0.10	–53.39	7e-4
2m1145a	15.23	8.59	3.01	12.26	11.70	–63.83	<1e-9
2m1145b	–0.73	1.96	2.00	2.69	2.95	–39.71	1e-3
2m1146	0.67	0.56	–0.08	0.21	1.17	–26.83	3e-3
2m1334	14.95	12.82	4.06	16.86	16.12	–65.88	1e-9
sdss0539	5.50	1.99	4.93	4.48	4.67	–19.62	3e-5
calar3	3.60	1.43	5.65	5.11	4.28	–28.06	6e-4
sori31	2.04	2.12	1.02	2.59	1.90	–11.16	4e-5
sori33	1.49	0.66	2.14	1.85	2.12	–8.39	2e-3
sori45	6.70	4.32	5.08	6.23	6.32	–29.93	5e-9

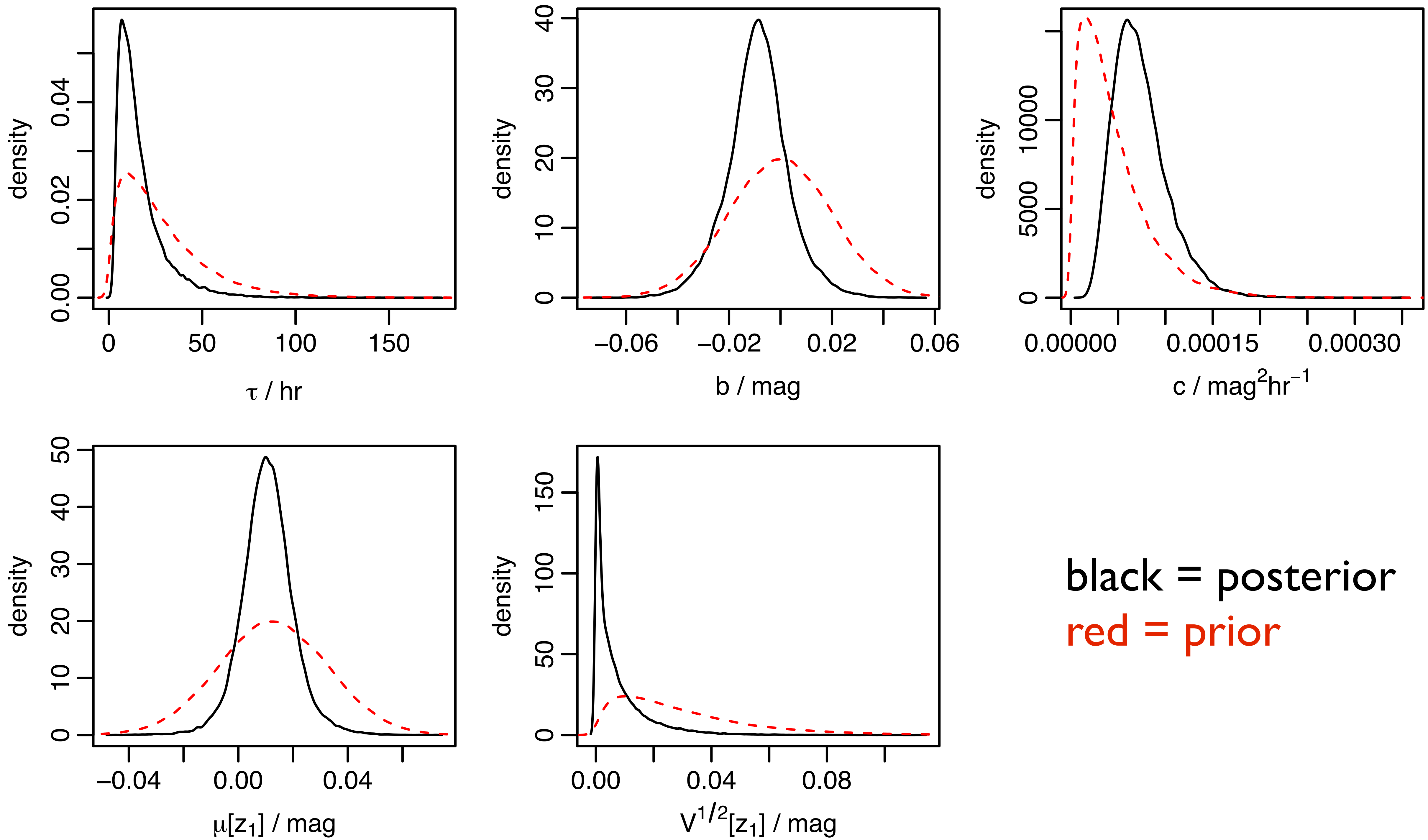
Notes. The penultimate column gives the value of the log likelihood for the no-model, $\log L^{\mathrm{NM}}$. The last column is the p-value for the hypothesis test from BJM.

Parameter posterior PDFs:



black = posterior
red = prior

Parameter posterior PDFs: 2m1145a

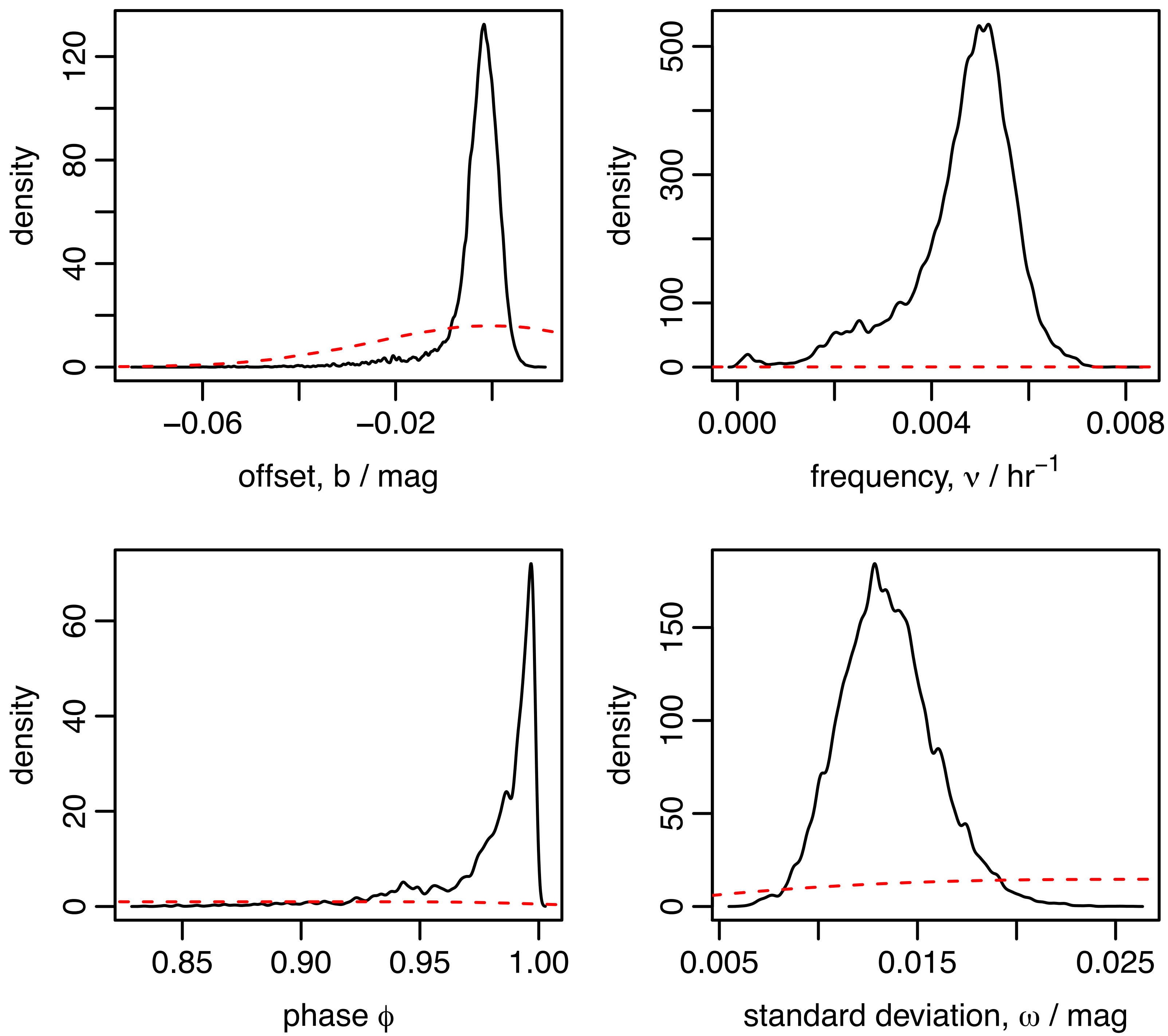


black = posterior

red = prior

Coryn Bailer-Jones, MPI for Astronomy, Heidelberg

Parameter posterior PDFs: 2m1334



black = posterior
red = prior