The Reliability of Period Detections in Regularly Spaced Data

Chris Koen

Department of Statistics

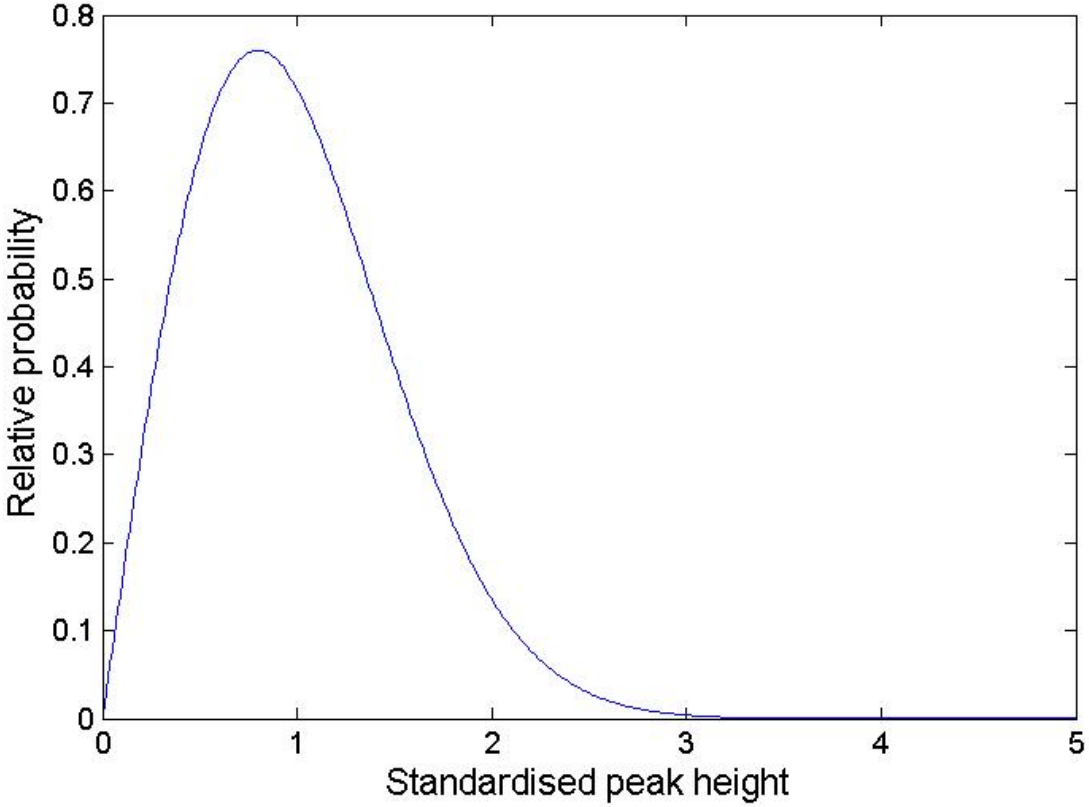
University of the Western Cape

Assumed throughout:

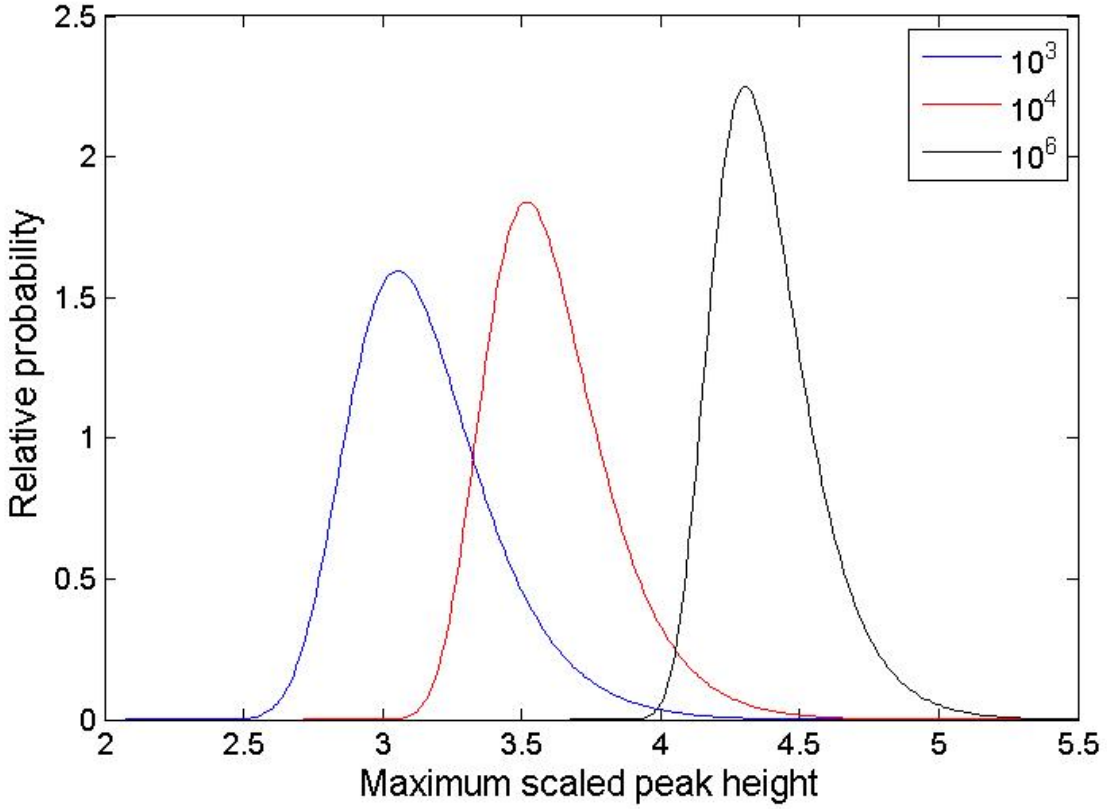
- Regular data spacing
- Amplitude, rather than power spectra
- Spectra calculated for all frequencies, i.e.

  $0 \leq f \leq$ Nyquist frequency
- Spectra fully oversampled
- Spectra standardised by division by its mean
- Large N (thousands)
- Noise is uncorrelated Gaussian
- Signal is a single sinusoid

Noise only: distribution of individual peak heights



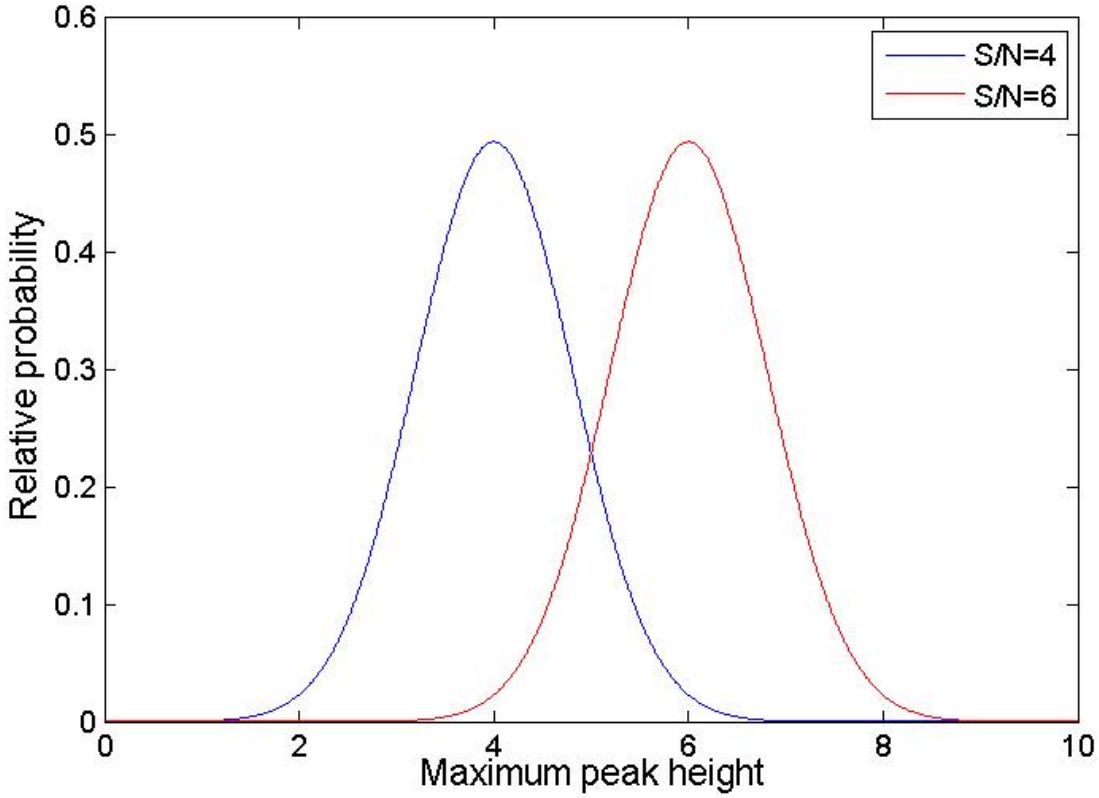
Distribution of maximum peak heights, over entire spectra; noise only



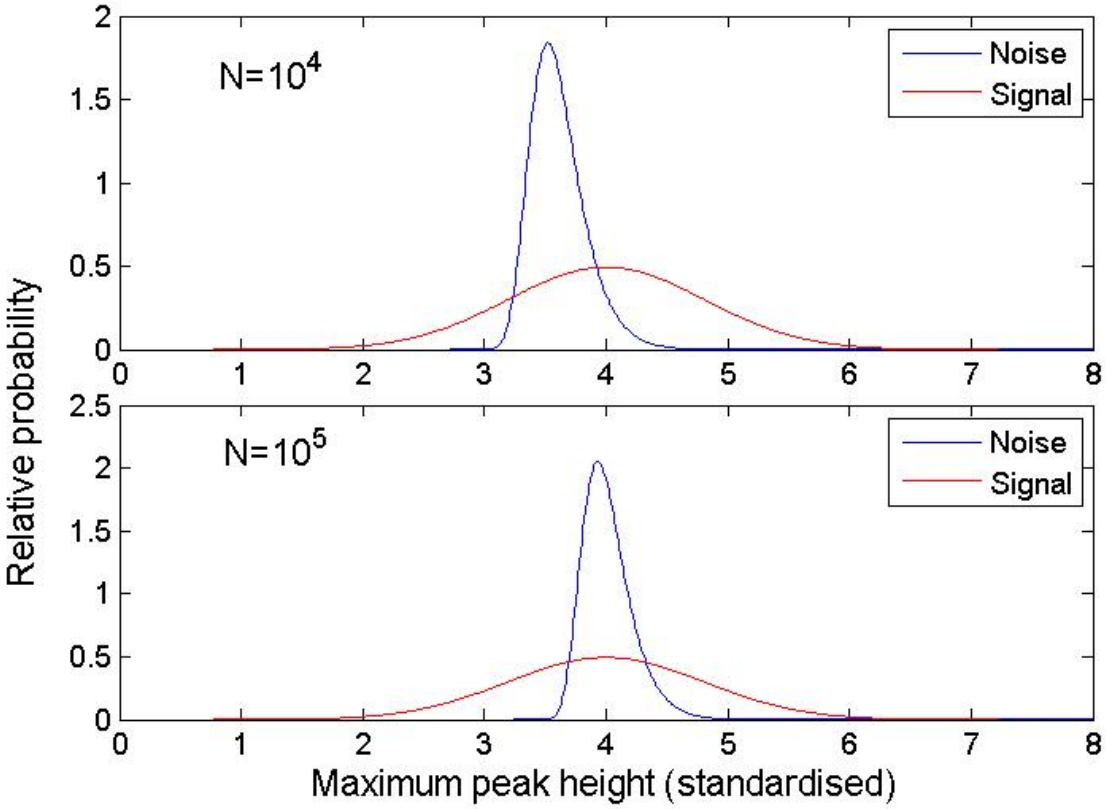
Signal peaks

- Single sinusoid
- Maximum height of spectral peak affected by noise

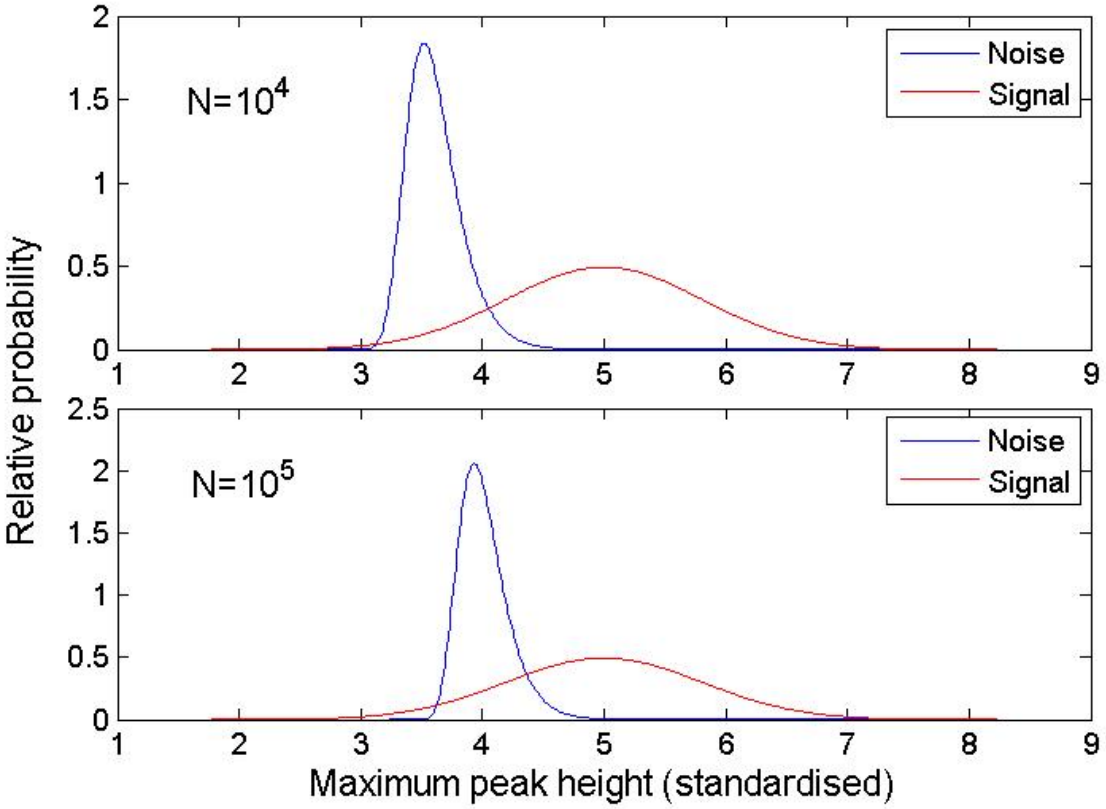
Distribution of (standardised) maximum peak height: centred on S/N, fixed width



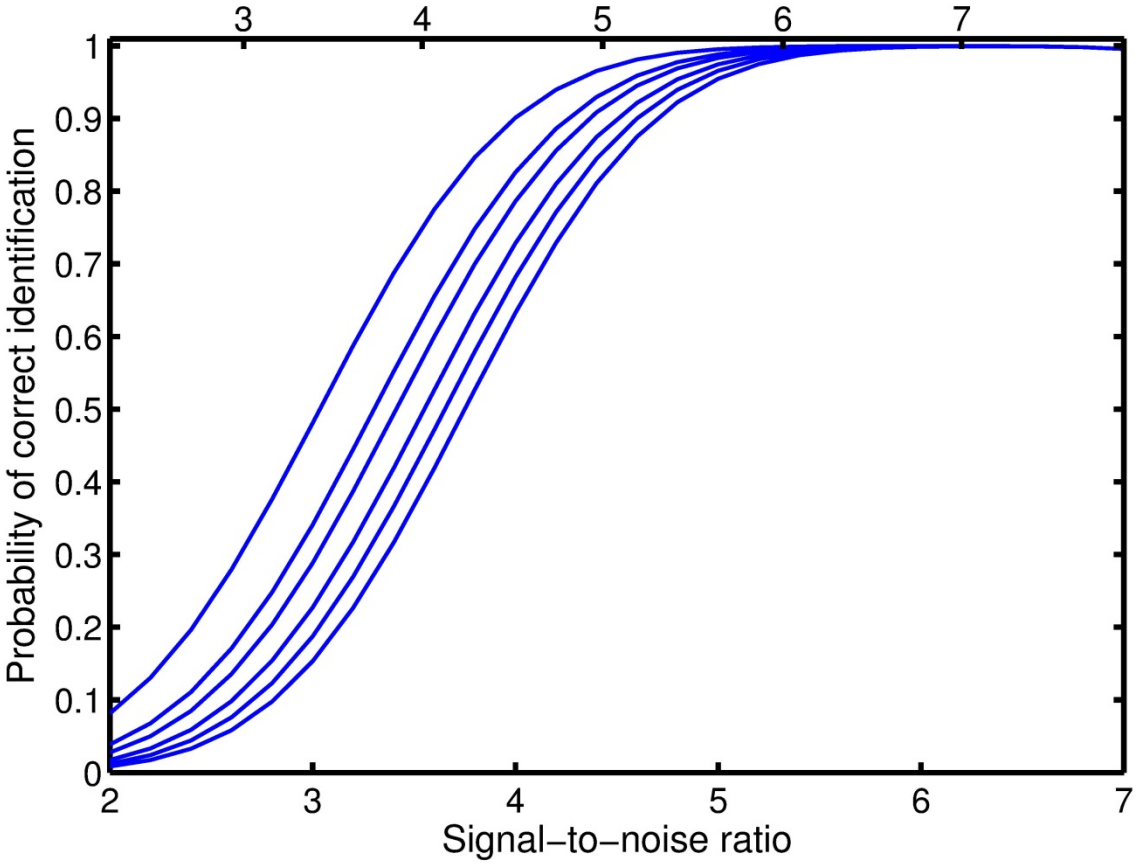
Comparison of largest peak distributions, S/N=4



Comparison of largest peak distributions, S/N=5



$$N = 10^4, 5 * 10^4, 10^5, 2.5 * 10^5, 5 * 10^5, 10^6$$



Details 1

- Distribution of noise spectrum maximum is Generalised Extreme Value (GEV) form

$$f(y) = \frac{\pi y}{2.208} \exp[-V - e^{-V}]$$

with $V = [\frac{\pi y^2}{4} - 1.05 \log N\ ]/1.04$

Details 2

- Distribution of signal spectrum maximum is Gaussian

$$f(x) = \frac{1}{1.43\sqrt{2}} \exp[-\frac{1}{2}U^2\ ]$$

with $U = \left[x - \left(\frac{\mathcal{S}}{\mathcal{N}}\right)\right] / (\frac{1.43}{\sqrt{\pi}})$

Details 3

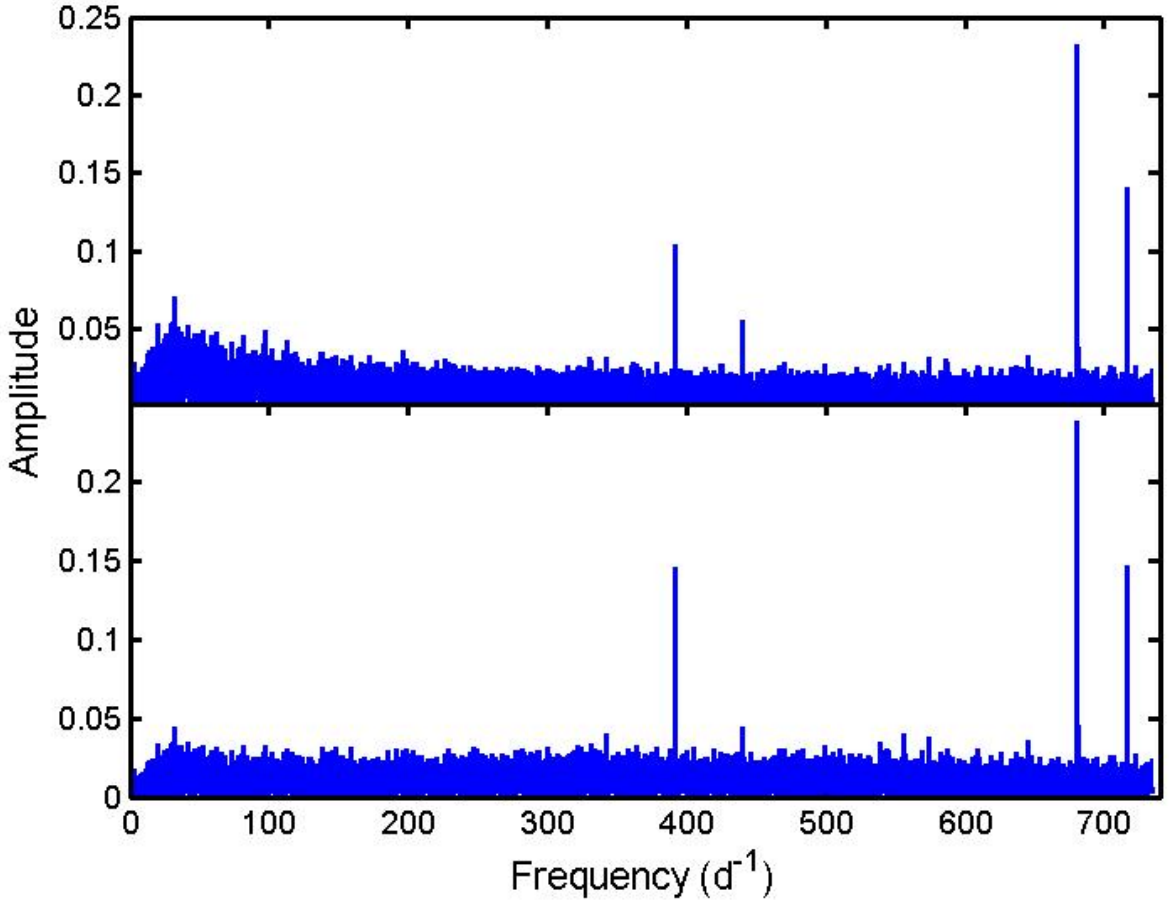
- Probability signal correctly identified

= Probability (signal peak>largest noise peak)

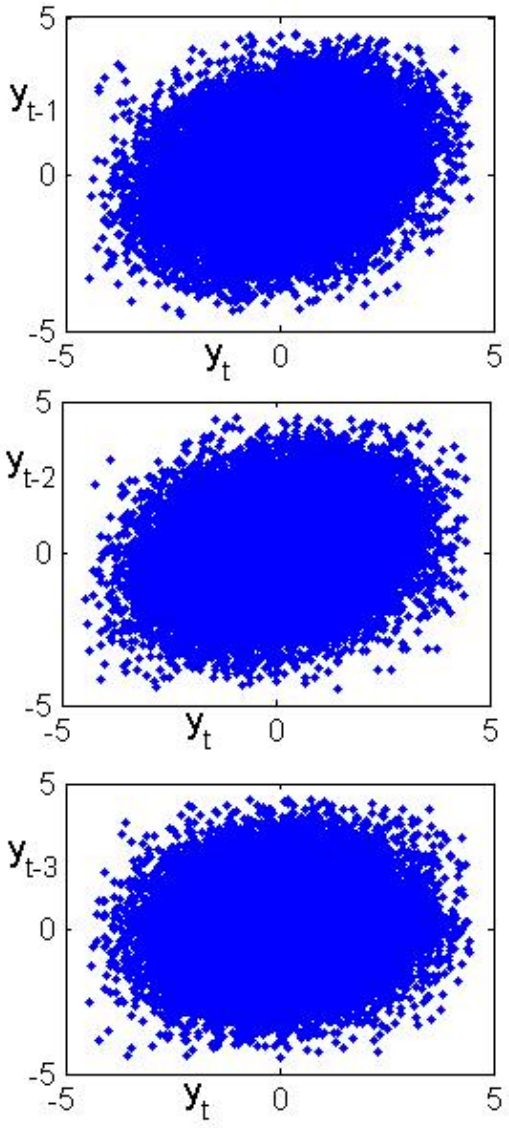
=P(x>y)

$=\int_0^{\infty} f(x)dx \int_0^{x} g(y)dy$

KIC8008067 (N=40 000) – note slight power excess at low frequencies in top panel



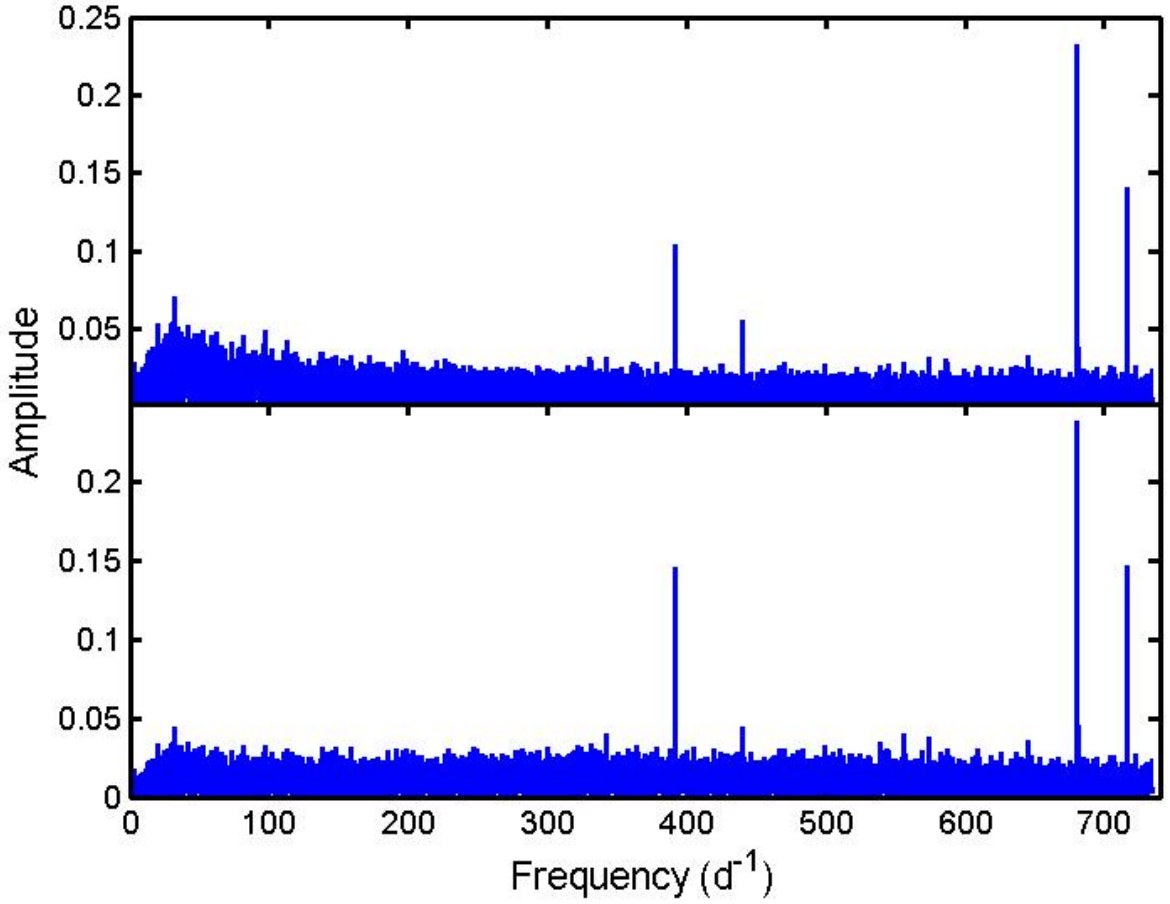
The origin of the low frequency power excess – autocorrelation at lags 1 and 2 in the noise



Filter data to remove autocorrelation:

- $y(t) = \alpha_1 y(t-1) + \alpha_2 \mathrm{y}(\mathrm{t}-2)$
- $r(t) = y(t) - \alpha_1 y(t-1) - \alpha_2 \mathrm{y}(\mathrm{t}-2)$
- Amplitudes and phases of sinusoids affected, not frequencies

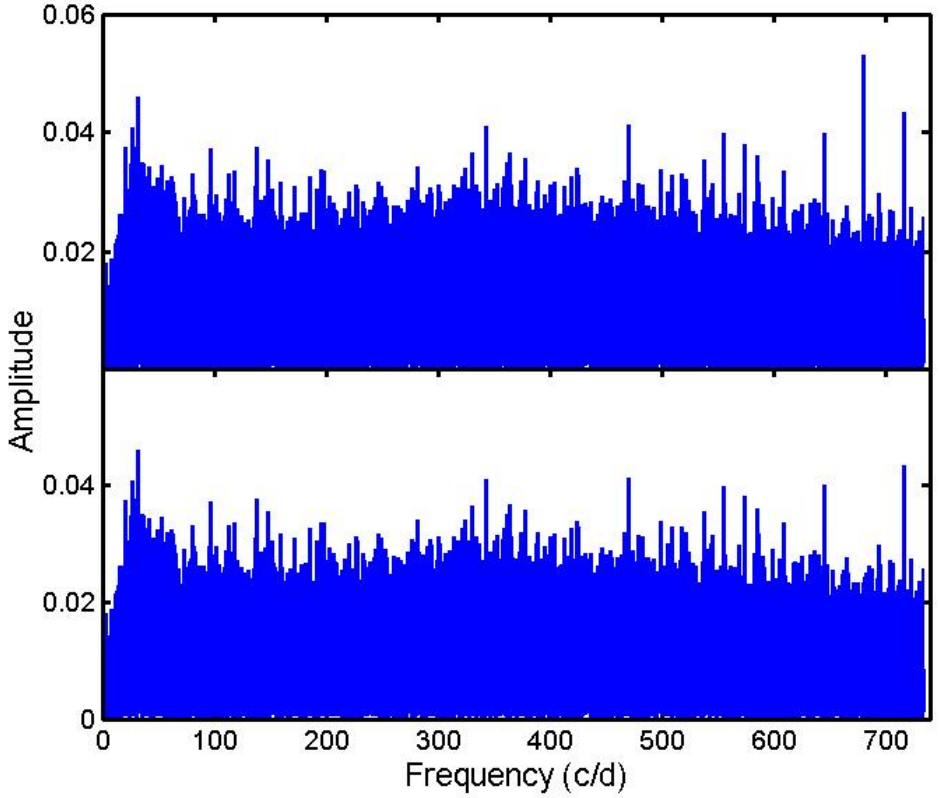
KIC8008067 (N=40 000): spectrum of original series in top panel,spectrum of filtered series in bottom panel



Successive prewhitening: according to Baran (2013) first 8 frequencies are Kepler artefacts, i.e."real" features in the data

	Frequency (d^{-1})	Maximum	Mean	Estimated $\mathcal{R}$	p
1	680.7609	0.2392	0.0104	20.5795	1.000
2	680.7242	0.1746	0.0104	15.0194	1.000
3	716.3978	0.1739	0.0104	14.9612	1.000
4	391.5798	0.1503	0.0103	12.9308	1.000
5	680.8013	0.1285	0.0103	11.0521	1.000
6	680.8343	0.0698	0.0103	6.0054	1.000
7	440.5286	0.0670	0.0103	5.7683	1.000
8	680.6948	0.0531	0.0103	4.5660	0.959
9	32.1529	0.0459	0.0103	3.9511	0.821

Last “significant” peak (#8), first “non-significant” peak (#9)



References

- Baran A., 2013, Acta Astron., 63, 203
- Baran A., Koen C., Pokrzywka B., 2015, MNRAS, 448, L16
- Koen C., 2015a, MNRAS, 449, 1098
- Koen C., 2015b, MNRAS, in press

Thank you, organisers of sdOB7

Thank you, funding agencies (University of the Western Cape, South African National Research Foundation).