The A-Star Puzzle
Proceedings IAU Symposium No. 224, 2004
J. Zverko, J. Žižňovský, S.J. Adelman, & W.W. Weiss, eds.

DOI: 10.1017/S1743921305009762

SigSpec – reliable computation of significance in Fourier space

Piet Reegen

Institut für Astronomie, Türkenschanzstraße 17, 1180 Vienna, Austria
email: reegen@astro.univie.ac.at

Abstract. SigSpec is a new method to compute the significance of the amplitude levels in the frequency domain, based on the false-alarm probability associated with a peak in the amplitude spectrum. The underlying probability density function (PDF) of the amplitude spectrum generated by pure noise may explicitly be derived if treated as frequency and phase-dependent. A comparison of the analytical solution with the results of extensive numerical calculatgions provides excellent agreement. In addition, the SigSpec software has already demonstrated clear advantages compared with the commonly used Fourier methods in various respects. A few examples for ground-based as well as space photometry are presented.

Keywords. Techniques: photometric, methods: data analysis, methods: statistical

1. From equidistant to non-equidistant sampling

The estimation of peak significance in the amplitude spectrum of a nonequidistantly sampled time series is commonly believed analytically unsolvable. This is mainly due to the following issues arising with the transition from equidistant to nonequidistant sampling in the time domain:

- The Nyquist frequency is not uniquely defined. This invokes complications when attempting to find the "highest" peak in the "entire" amplitude spectrum.
- Writing the amplitude spectrum in the form

$$A\left(\omega\right) = \frac{1}{K}\sqrt{\sum_{k=0}^{K-1}\sum_{l=0}^{K-1} x_k\, x_l \cos\omega\left(t_k - t_l\right)}\,, \tag{1.1}$$

where K denotes the total number of time series data, and considering the measurements x_k, x_l to be statistically independent (noise) and hence uncorrelated in any measurable respect, the expected value of amplitudes may be written according to

$$\left\langle A\right\rangle\left(\omega\right) = \frac{1}{K^2}\sqrt{\left(\sum_{k=0}^{K-1}\sum_{l=0}^{K-1} x_k\, x_l\right)\left[\sum_{k=0}^{K-1}\sum_{l=0}^{K-1}\cos\omega\left(t_k - t_l\right)\right]}\,. \tag{1.2}$$

The expected amplitude level is not unique for all frequencies in the spectrum (Fig. 1). The highest amplitude need not refer to the most significant peak.

- In general, it is impossible to find orthogonal Fourier coefficients for nonequidistant sampling, whence the Fourier spectrum is usually considered as a continuous function, and the frequencies are oversampled to achieve a more accurate determination of peaks. Furthermore, the tendency of peak frequencies to reflect the properties of sampling rather than the signal is frequently corrected by combining Fourier analysis with least-squares fitting.

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press

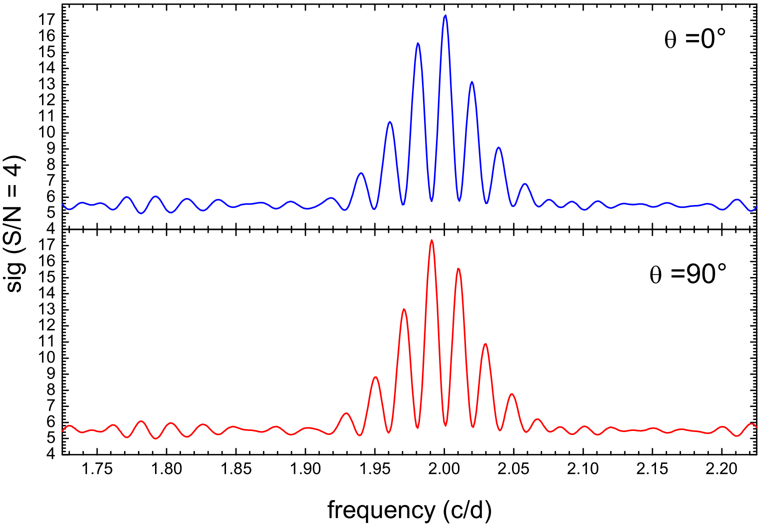

Figure 1. Significance vs. frequency ($S/N = 4$) for a time series sampled according to Johnson V measurements of IC 4996 #89 by Zwintz *et al.* (2005). Daily data gaps invoke aliases close to 2 d^{-1} handled by frequency-dependent analysis. The two graphs refer to different phases and give an impression of the dependence of significance on the phase angle in Fourier space, θ.

• For equidistantly sampled real observables in the time domain, the Fourier sine coefficients disappear, and the amplitude is entirely defined by the cosine coefficient. The Fourier space phase angle θ is zero. For nonequidistant sampling, this angle does not necessarily vanish, which is taken as a motivation to introduce the phase information in the frequency domain into a statistical analysis (Fig. 2).

The consequence of these features is to reword the question for the right peak. Instead of asking for the probability of the peak with the highest signal-to-noise ratio (S/N) in the amplitude spectrum to be generated by observational noise, the present study investigates the probability of an amplitude level *at a given frequency and Fourier space phase angle* to be generated by observational noise.

2. The PDF of the amplitude spectrum

Another issue to be taken into account when deriving the frequency- and phase-dependent amplitude PDF is that the observable is usually adjusted to zero mean before computing the Fourier spectrum. The mean of a finite set of random variates is free to scatter about the expected value of the underlying random process. The mean in our applications is shorn of this freedom, which affects the amplitude PDF.

The frequency- and phase-dependent amplitude PDF is an elliptical Gaussian,

$$\phi_A(\omega,\theta) = \frac{\sqrt{2A}}{\rho}\,\mathrm{e}^{-\frac{A^2}{\rho}} \tag{2.1}$$

with

$$\rho(\omega,\theta) := \frac{2\,\sigma^2}{K^3}\,\frac{ab}{b^2\,\cos(\theta-\theta_0) + a^2\,\sin(\theta-\theta_0)}\,, \tag{2.2}$$

where the standard deviation of the measurements in the time domain is denoted σ. The variables a and b denote the semi-major and semi-minor axes of the elliptical standard

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press

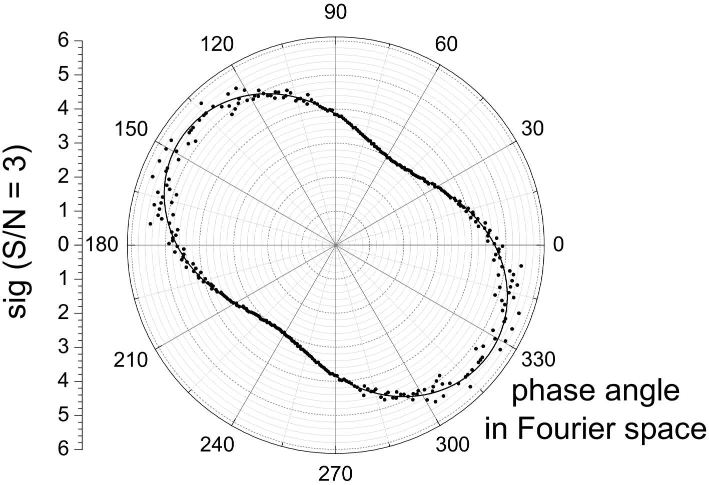

Figure 2. Significance vs. phase ($S/N = 3$) for the time series used in Fig. 1 at a frequency of 1.956 d^{-1}. The solid line represents the theoretical result. The dots are the output of a numerical simulation based on 250 million synthetic time series with the given sampling.

deviation and θ_0 is the phase angle of the semi-major axis. These quantities are

$$\tan 2\,\theta_0(\omega) := \frac{K\sum_{k=0}^{K-1}\sin 2\,\omega t_k - 2\left(\sum_{k=0}^{K-1}\cos\omega t_k\right)\left(\sum_{k=0}^{K-1}\sin\omega t_k\right)}{K\sum_{k=0}^{K-1}\cos 2\,\omega t_k - \left(\sum_{k=0}^{K-1}\cos\omega t_k\right)^2 + \left(\sum_{k=0}^{K-1}\sin\omega t_k\right)^2}\,, \tag{2.3}$$

$$a\,(\omega,\theta) := \left| K\sum_{k=0}^{K-1}\cos^2(\theta-\theta_0) - \left[\sum_{k=0}^{K-1}\cos(\theta-\theta_0)\right]^2\right|\,, \tag{2.4}$$

$$b\,(\omega,\theta) := \left| K\sum_{k=0}^{K-1}\sin^2(\theta-\theta_0) - \left[\sum_{k=0}^{K-1}\sin(\theta-\theta_0)\right]^2\right|\,. \tag{2.5}$$

For $a = b$ and averaged over frequency, eq. (2.1) reduces to the classical S/N estimate (Scargle 1982), which appears as a rather poor approximation especially in the low frequency region or in case of strong aliasing due to periodic data gaps.

Since representing a weighted sum of random variates, and thanks to the Central Limit Theorem (e.g., Stuart & Ord 1994), eq. (2.1) will also hold for a sufficiently large number of time series data even if the noise in the time domain is assumed nonGaussian. This is in good agreement to numerical simulations.

3. False-alarm probability and significance

Integration of eq. (2.1) over amplitude returns the cumulative distribution function (CDF) giving the probability for an amplitude generated by pure Gaussian noise to be lower than a given amplitude level. The complementary defines the false-alarm probability

$$\Phi_{\mathrm{FA}}(A) := \mathrm{e}^{-\frac{A^2}{\rho}}\,. \tag{3.1}$$

Then the significance of an amplitude,

$$\mathrm{sig}\,(A) := -\lg\Phi_{\mathrm{FA}}\,(A)\;, \tag{3.2}$$

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press

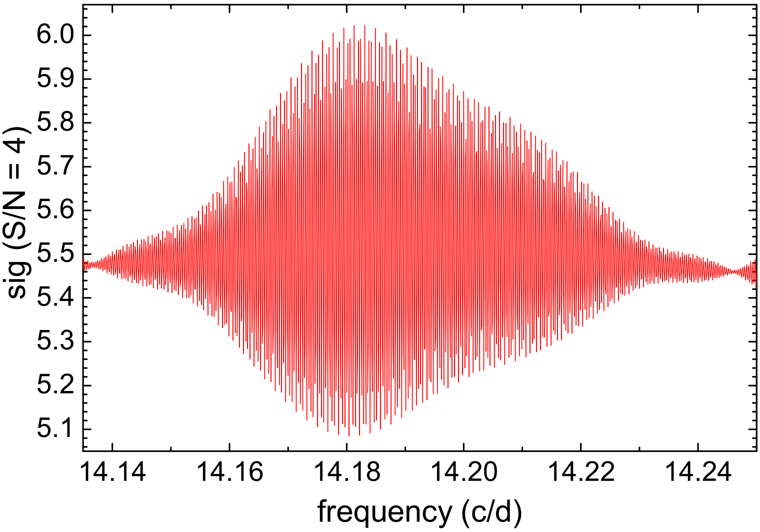

Figure 3. Results of SigSpec for MOST observations of the commissioning target δ Cet with a duty cycle of 99.9%. A coarse straylight correction and the removal of data points contaminated by cosmics (South Atlantic Anomaly) produce periodical gaps in the dataset. The diagram shows the significance associated to a S/N of 4 for frequencies close to the orbital period of the satellite. The false-alarm probability varies by a factor ≈ 8.

is a measure for the number of noise datasets to be analysed on average to obtain one amplitude at least as high as A. Eq. (3.2) immediately evaluates to

$$\mathrm{sig}\,(A) := \frac{A^2}{\rho}\,\lg \mathrm{e}\,. \tag{3.3}$$

Example. The expected value of significance for an amplitude S/N

$$\frac{AK}{\sigma}\,\frac{\sqrt{\pi}}{2} =: 4\,, \tag{3.4}$$

is 5.46, which represents the estimate by Breger *et al.* (1993). Using eq. (3.2), $\mathrm{sig}\,(A) = 5.46$ yields an associated false-alarm probability of $\approx 1 : 300000$.

4. Tests with synthetic data

A software package based on these results is currently being tested by comparing the theoretically evaluated significance with extensive numerical simulations. Applications of the program SigSpec to test the data illustrate the improvement provided by frequency- and phase-dependent significance computations. Fig. 2 gives an impression of the dependence of significance on phase angle and, in addition, illustrates the perfect agreement between theory and numerical analysis.

5. Real life

SigSpec has already been applied to real photometric data from ground-based as well as space observations. Fig. 3 – 7 present a few results obtained with this new method.

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press

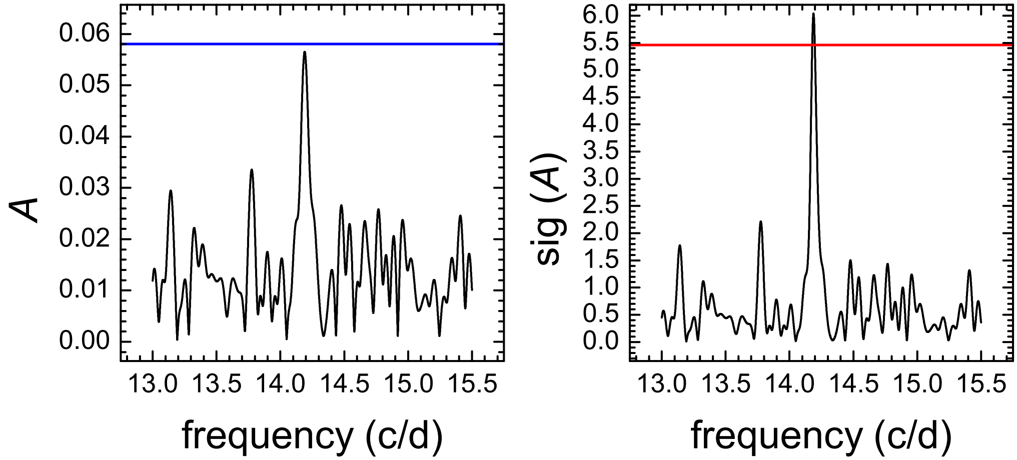

Figure 4. Sampling of δ Cet with synthetic data: a single sinusoidal signal with a period close to the orbital, plus Gaussian noise with $AK\frac{\sqrt{\pi}}{2} = 4.3\,\sigma$. The amplitude spectrum (*left*) shows a peak below a signal-to-nosie ratio of 4. The significance spectrum (*right*) produces a peak consistent with the initial specifications. Even a 99.9% duty cycle cannot guarantee a good Fourier analysis!

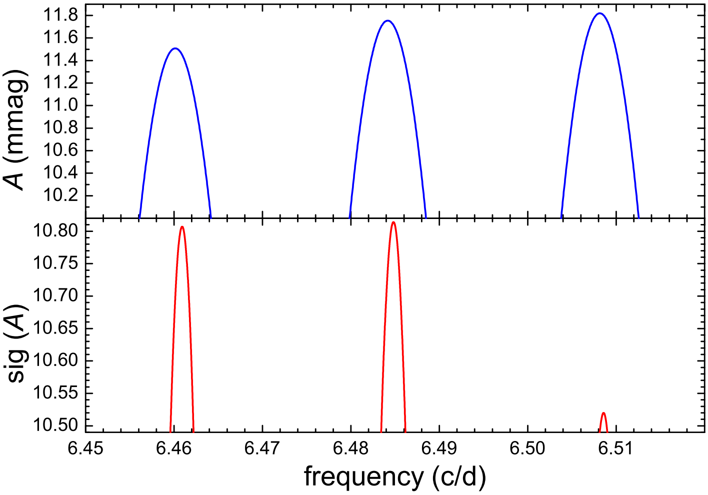

Figure 5. For V589 Mon = NGC 2264 W20 (Peña *et al.* 2002), Fourier analysis (*top*) returns the highest amplitude at 6.51 d^{-1}. The significance spectrum for the same data (*bottom*) is 6.485 d^{-1}, in excellent agreement with the new campaign data by Kallinger & Zwintz (2004).

5.1. *Periodical gaps and aliasing*

One clear result of the analyses performed on real astronomical data is that the S/N estimate frequently fails for low frequencies or in case of periodical changes in the time domain sampling. The latter are known to produce aliases at the frequency of the sampling variations (and integer multiples).

Fig. 3 gives an example for periodical gaps in the MOST measurements of δ Cet, which are invoked by the rejection of images contaminated by cosmics and/or strong straylight. Fig. 4 illustrates the systematic error in the recovery of S/N for a synthetic signal with a period close to the orbital period of the satellite. The comparison to the significance spectrum gives an impression of the clear advantages of SigSpec in this respect.

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press

Table 1. *Left:* Fourier analysis and consecutive prewhitening for IC 4996 #89 return 5 peaks with an amplitude S/N above 4. Indication for a new Pre-Main Sequence variable? *Right:* In contradiction, SigSpec provides only two significant components. The first is considered as a 1 d^{-1} alias, the second is probably due to background variations, whence the star turns out to be constant. The corresponding light curve is shown in Fig. 6.

#	f (c/d)	A (mmag)	phase	#	f (c/d)	A (mmag)	phase
1	3.133	1.90	0.45	1	2.976	1.57	0.74
2	2.982	1.96	0.53	2	3.132	2.12	0.76
3	3.997	1.85	0.15				
4	5.407	1.51	0.87				
5	17.368	1.25	0.91				

5.2. *Which alias?*

The question whether the highest amplitude (or S/N) is associated to the most significant peak becomes important especially in case of aliases. Fig. 5 contains the amplitude and significance spectra for V589 Mon photometry (Peña *et al.* 2002). The frequency spacing of aliases is 0.024 d^{-1}, corresponding to the total time interval of the measurements (42 nights). The Fourier spectrum shows the highest amplitude at 6.51 d^{-1}, whereas maximum significance is obtained at 6.485 d^{-1}. Recent measurements (Kallinger & Zwintz, 2004) confirm the SigSpec result.

5.3. *Overestimation of significance by S/N*

A practical consequence of the differences between S/N estimation and SigSpec is shown for V measurements of star #89 in the young open cluster IC 4996 (Zwintz *et al.* 2005). Table 5.4 compares five frequencies obtained by consecutive prewhitening using the classical method to the corresponding SigSpec result. The fits to the light curve are displayed in Fig. 6 which illustrates the enhanced reliability of the SigSpec analysis.

In the 5-frequency solution by means of the S/N method (Table 5.4), one would tend to interpret the frequencies 4 and 5 as an indication of stellar pulsation. These two components are not recovered by SigSpec. Instead, two frequencies are found with the new method, the first probably representing a 1 d^{-1} alias. The second frequency, 3.132 d^{-1}, is also found in the data of other stars in the cluster, indicating variations in the cluster background. Finally, the SigSpec result is in perfect agreement with the analysis of the B data, where the components 3 to 5 in Table 5.4 are not confirmed. The conclusion for this star is that no evidence of stellar variability is found.

5.4. *Underestimation of significance by S/N*

In 5.3, an example for frequencies erroneously interpreted significant by means of S/N is given. On the other hand, measurements of the HST-FGS on GSC 09137-03505 (Kallinger *et al.* 2004) provide complementary results. Fourier analysis, S/N estimation, and consecutive prewhitening return three significant components, whereas SigSpec provides 37 frequencies. The amplitude and significance spectra are shown in Fig. 7.

6. Conclusions and outreach

SigSpec exceeds the diagnostic capabilities of classical Fourier analysis. The program is based on an analytically clean determination of the amplitude probability density function. It is the first technique in astronomical time series analysis to use both frequency and phase angle to compute instead of an estimate, a false alarm probability. Tests on

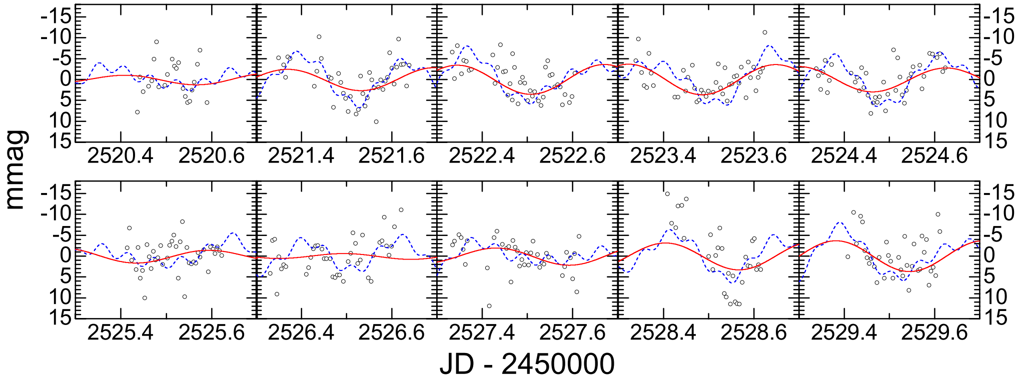

Figure 6. V measurements of IC 4996 #89. The dashed line represents the fit produced by classical Fourier analysis (Table 5.4, *left*), the solid curve refers to SigSpec (Table 5.4, *right*). IC 4996 photometry has been performed by Zwintz *et al.* (2005)

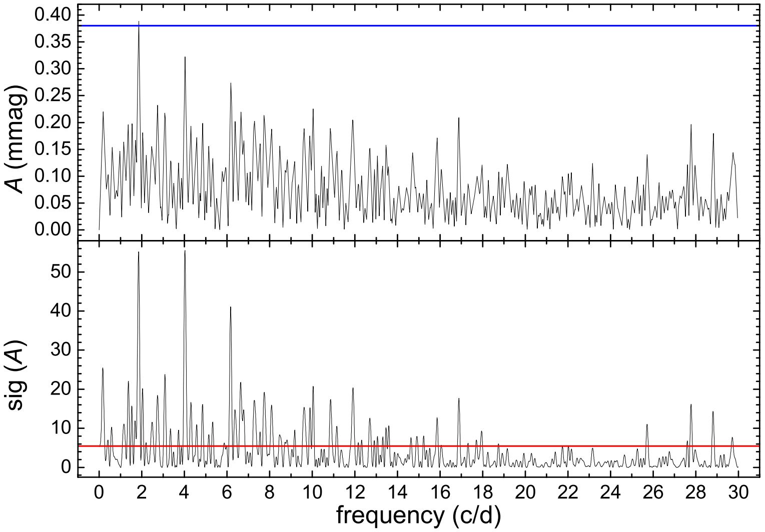

Figure 7. The amplitude spectrum (*top*) of HST-FGS data on GSC 09137-03505 shows a single peak reaching a S/N of 4. After consecutive prewhitening, three significant components are identified (Kallinger *et al.* 2004). SigSpec yields 37 frequencies with a significance > 5.5 (*bottom*).

synthetic and real data confirm that the SigSpec analysis is superior to classical signal-to-noise ratio estimates.

There is indication that frequencies obtained by SigSpec are practically as accurate as the least-squares fits. Since the latter become extremely time consuming especially in case of multifrequency analysis of long datasets, the program SigSpec may be suspected to be considerably faster than classical methods.

Further development focuses on using SigSpec for a more intelligent averaging strategy than the commonly used zero mean correction and on the implementation into a fully automatic data reduction and analysis package for MOST photometry.

Acknowledgements

It is a pleasure to thank Michel Breger, Werner W. Weiss and Maria G. Firneis for support and valuable discussion, as well as Thomas Kallinger and Konstanze Zwintz for testing the software in real life and providing a bunch of precious output.

Financial support was kindly granted by the IAU, the Austrian Space Agency (ASA), the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF), and the Österreichische Forschungsgemeinschaft (ÖFG).

References

Breger, M., Stich, J., Garrido, R., Martin, B., Jiang S.-Y., Li Z.-P., Hube, D.P., Ostermann, W., Paparo, M., Scheck, M. 1993, *A&A*, 271, 482

Kallinger, Th., Zwintz, K. 2004, *Private communication*

Kallinger, Th., Zwintz, K., Pamyatnykh, A.A., Guenther, D.B., Weiss, W.W. 2004, *A&A*, submitted

Peña, J.H., Peniche, R., Cervantes, F., Parrao, L. 2002, *Rev. Mex. de Astron. y Astrofis.*, 38, 31

Scargle, J.D. 1982, *ApJ*, 263, 835

Stuart, A., Ord, J.K. 1994, *Kendall's Advanced Theory of Statistics*, vol. 1 *Distribution Theory*, 6th ed. (London: Arnold), p. 310f

Zwintz, K., Marconi, M., Kallinger, Th., Weiss, W.W. 2005, *These Proceedings*, 353

https://doi.org/10.1017/S1743921305009762 Published online by Cambridge University Press