A study on the optimum order of autoregressive models for heart rate variability

Article *in* Physiological Measurement · June 2002

DOI: 10.1088/0967-3334/23/2/308 · Source: PubMed

4 authors, including:

Anita Mcgrogan
University of Bath
49 PUBLICATIONS **1,575** CITATIONS

SEE PROFILE

Fernando Soares Schlindwein
University of Leicester
304 PUBLICATIONS **1,682** CITATIONS

SEE PROFILE

Ana Paula Rocha
University of Porto
82 PUBLICATIONS **2,409** CITATIONS

SEE PROFILE

INSTITUTE OF PHYSICS PUBLISHING PHYSIOLOGICAL MEASUREMENT

Physiol. Meas. **23** (2002) 325–336 PII: S0967-3334(02)29245-X

A study on the optimum order of autoregressive models for heart rate variability

Anita Boardman[1], **Fernando Soares Schlindwein**[1], **Ana Paula Rocha**[2] **and Argentina Leite**[3]

[1] Department of Engineering, University of Leicester, Leicester, UK
[2] Department of Applied Mathematics, University of Porto, Portugal
[3] Department of Mathematics, University of Trás-os-Montes e Alto Douro, Portugal

E-mail: ab66@le.ac.uk, fss1@le.ac.uk, aprocha@fc.up.pt and tinucha@utad.pt

Received 28 September 2001, in final form 2 January 2002
Published 8 March 2002
Online at stacks.iop.org/PM/23/325

Abstract

Heart rate variability (HRV) has been used as a non-invasive marker of the activity of the autonomic nervous system and its spectrum analysis gives a measure of the sympatho-vagal balance. If short segments are used in an attempt to improve temporal resolution, autoregressive spectral estimation, where the model order must be estimated, is preferred. In this paper we compare four criteria for the estimation of the 'optimum' model order for an autoregressive (AR) process applied to short segments of tachograms used for HRV analysis. The criteria used were Akaike's final prediction error, Akaike's information criterion, Parzen's criterion of autoregressive transfer function and Rissanen's minimum description length method, and they were first applied to tachograms to verify (i) the range and distribution of model orders obtained and (ii) if the different techniques suggest the same model order for the same frames. The four techniques were then tested using a true AR process of known order $p = 6$; this verified the ability of the criteria to estimate the correct order of a true AR process and the effect, on the spectrum, of choosing a wrong model order was also investigated. It was found that all the four criteria underestimate the true AR order; specifying a fixed model order was then looked at and it is recommended that an AR order not less than $p = 16$, should be used for spectral analysis of short segments of tachograms.

Keywords: heart rate variability, spectrum analysis, autoregressive model order

1. Introduction

Heart rate variability (HRV) describes the beat-to-beat variation in heart rate arising (amongst other causes) from the efferent activity of the sympathetic and parasympathetic branches of

the autonomic nervous system (ANS). The signal used is the tachogram formed from the RR intervals, obtained non-invasively from the electrocardiogram. Power spectral analysis of this signal identifies up to three frequency components that can be associated with identifiable physiological mechanisms: the very low-frequency band (VLF), occurring between 0 and 0.03 Hz caused by thermoregulation and humoral factors; the low-frequency component (LF), centred around 0.1 Hz which comes from baroreflex-related heart rate variability; and the high frequency component (HF), occurring between 0.18 and 0.4 Hz, which arises from the respiratory sinus arrhythmia (RSA) (Cerutti *et al* 1995, van Ravenswaaij-Arts *et al* 1993).

Spectrum analysis is often performed using a Fourier-based technique simply because that is the first-line engineering approach to spectrum analysis and because FFT functions are so widely available. Some of the limitations of the Fourier-based approach are the poor spectral resolution, especially when short time frames are used, and leakage (Kay and Marple 1981).

Autoregressive spectral analysis has been used here mainly because of its better spectral resolution for short frames of data (Marple 1977). An inconvenience of AR-based spectral analysis is that the order for the AR model which best represents the series must be estimated prior to the spectral analysis. Criteria for the estimation of the model order to be used have been published, however, as noted by Kaluzynski (1989) and Schlindwein and Evans (1990), these tend to underestimate the order when applied for AR spectral estimation of Doppler blood flow signals.

This study is concerned with finding adequate AR model orders for spectral analysis of short segments of the time series formed from RR intervals. Jones (1974) noted that Akaike's final prediction error (FPE) and Akaike's information criterion (AIC) tend to predict identical model orders for the same frames of measured data. This will be investigated for HRV data and also compared with the results from Parzen's criterion of autoregressive transfer function (CAT) and Rissanen's minimum description length method (RIS). The resulting spectrum using the overall recommended order will be compared with spectra obtained using different AR model orders (both above and below the estimated 'optimum') and to the spectrum obtained using the non-parametric (Fourier-based) approach to estimate the effect that choosing the wrong order has on the AR spectrum.

When estimating model orders the frequency used to re-sample the tachogram should be taken into account. From Nyquist, this should be at least twice the highest expected frequency component present in the spectrum (the fastest possible change in heart rate is the situation where there is a succession of a short interval followed by a long interval, implying that the fastest frequency is half the mean heart rate), but not so high that the signal is over-sampled. The effect that the re-sampling frequency change has on the estimated model order will be tested by comparing the re-sampling rate of 4 Hz, used throughout the paper, with higher and lower rates of 8 Hz and 2 Hz.

A true AR data series was also generated and all four criteria were used to try to predict the (known) AR order. This allowed an assessment of the abilities of the four criteria used to estimate the correct order of an autoregressive process when frames with $N = 128$ samples are used.

2. Data acquisition and pre-processing

The data was obtained from two sources: the first was a 24 h Holter recording of a normal 19-year-old adult (subject 1) using a Mortara recorder; secondly, six files from the MIT-BIH arrhythmia database (MIT-BIH 1992) were used: five of these files with mainly normal complexes (subjects MIT-100, 101, 112, 113 and 122) and one file with some pre-ventricular complexes (MIT-233). For subject 1 the QRS complexes were detected and a series of

successive RR intervals were produced and recorded as a raw tachogram. Since ectopic beats are independent from the control mechanisms that modulate the heart rate (which is what we are interested in here) we implemented a simple algorithm for rejection of ectopic beats and arrhythmic events which is performed as follows: if $(\mathrm{RR}(k) > 0.7^{*}(\mathrm{RR}(k-1) + \mathrm{RR}(k+1)))$, then $\mathrm{RR}(k) = (\mathrm{RR}(k-1) + \mathrm{RR}(k+1))/2$. For the six records of MIT-BIH data we have used the annotated detection marks as fiducial points for the detection and have *not* implemented the arrhythmia rejection part of the algorithm. This was done to see how much the presence of arrhythmias in subject MIT-233 influenced the estimated AR orders. The raw tachogram consists of samples of data whose time intervals are not uniform, but instead refer to consecutive RR events, that is, it is a time-event series. A cubic spline interpolation was used for re-sampling the series at a constant rate of 4 Hz, to give equal sampling intervals in time and allow traditional spectral estimation with frequency measured in Hz rather than mean heart rate. The data was then sectioned into non-overlapping frames of $N = 128$ samples, each frame was de-trended and analysed individually.

2.1. Autoregressive model

An AR process of order p can be written as

$$x_t = n_t + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} \tag{1}$$

where n_t is the white noise driving signal, which is the innovation of the AR process, and p is the order of the AR model.

To estimate the AR power spectrum density function, the parameters of the filter, $\{a_1, a_2, \ldots, a_p\}$, and the variance that characterizes the white noise, σ^2, must be found. The system of equations to find the AR parameters is linear and its solution is straightforward. Furthermore, it can be solved iteratively, using the Levinson–Durbin equations, thus reducing processing time (Kay and Marple 1981).

The AR power spectrum density estimate is given by

$$P_{AR}(f) = \frac{\sigma^2 \Delta t}{\left| 1 + \sum_{k=1}^{p} a_k \, \mathrm{e}^{-j2\pi f k \Delta t} \right|^2} \tag{2}$$

where σ^2 is the variance of the white noise driving function and Δt is the re-sampling interval.

2.2. Model order estimation

Four criteria were used for the estimation of the optimum model order. The first was Akaike's final prediction error (FPE) criterion (Akaike 1969), described as

$$\mathrm{FPE}_p = \sigma_p^2 \left(\frac{N + p + 1}{N - p - 1} \right) \tag{3}$$

where N is the number of data samples and σ_p^2 is the estimate of the prediction error power for model order p.

The second technique used for estimating the model order was Akaike's information criterion (AIC) (Akaike 1974):

$$\mathrm{AIC}_p = \ln\left(\sigma_p^2\right) + \frac{2(p+1)}{N}. \tag{4}$$

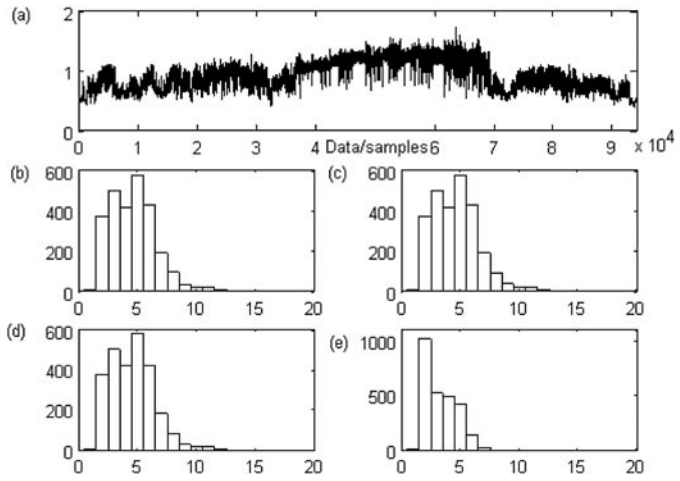

Figure 1. (a) Plot of the tachogram signal from subject 1 in data samples, i.e. each point corresponds to an RR interval, and histograms showing the optimum model order (x-axis) as found by the (b) FPE, (c) AIC, (d) CAT and (e) RIS criteria corresponding to the tachogram shown in (a).

The third criterion was Parzen's criterion autoregressive transfer function (CAT) (Parzen 1975):

$$\mathrm{CAT}_p = \left(\frac{1}{N} \sum_{j=1}^{p} \frac{N-j}{N\sigma_j^2} \right) - \frac{1}{\sigma_p^2} \tag{5}$$

where σ_p^2 is the intermediate prediction error power for $j = 1$ to $j = p$.

The final technique was Rissanen's minimum description length method (RIS) (Rissanen 1984):

$$\mathrm{RIS}_p = \sigma_p^2 \left(1 + \left(\frac{p+1}{N} \right) \ln(N) \right) \tag{6}$$

For all four techniques, the mean square error and bias change for each model order used and the above-defined functions have a minimum point; the model order corresponding to this minimum point is said to be the optimum.

3. Estimating the optimum order for the given data

Each set of data was tested using the four criteria described above and histograms of the optimum model order for each set of data were produced. The data and histograms from subject 1 (normal) are shown in figure 1. It can be seen that overall the model order chosen most frequently for the data over 24 h was $p = 5$ for FPE, AIC and CAT, but for RIS it was $p = 2$, and with a narrower spread in the histogram. For the MIT-BIH subjects the estimated order ranged mostly up to $p = 6$, and in no case was higher than $p = 16$, as illustrated in figure 2 for subject MIT-100. This was similarly true for the other four subjects with mostly 'normal' beats. For the record of subject MIT-233, which contains some pre-ventricular cycles, the range of model orders estimated was wider, up to $p = 22$. The corresponding histogram is shown in figure 3.

We believe all four techniques frequently produce an underestimation of the required AR model order. It was stated in the introduction that, in general, it would be expected that three frequency components might occur, corresponding to very low, low and high frequency, each

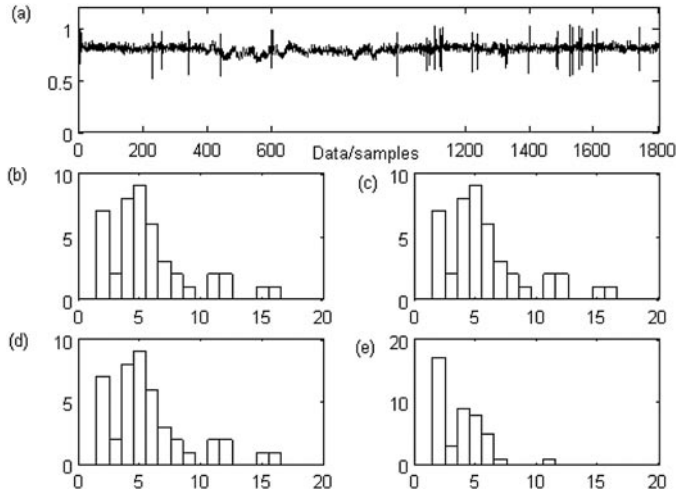

Figure 2. Plot of the tachogram signal (a) from subject MIT-100, in data samples, i.e. each point corresponds to an RR interval, and histograms showing the optimum model order (x-axis) as found by the (b) FPE, (c) AIC, (d) CAT and (e) RIS criteria corresponding to the tachogram shown in (a).

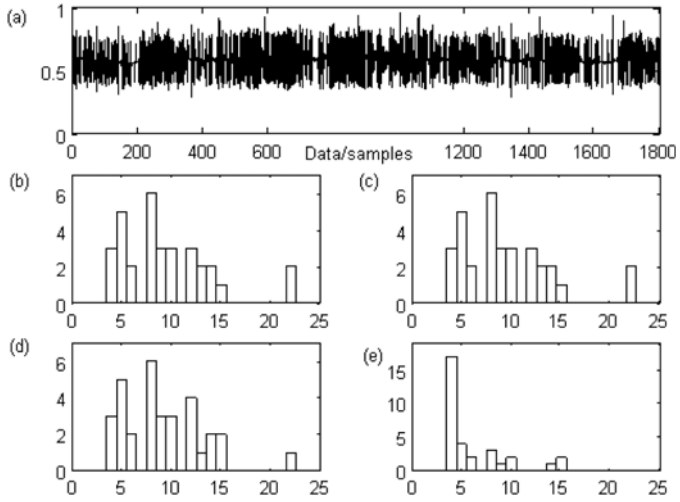

Figure 3. Plot of the tachogram signal (a) from subject MIT-233, in data samples, i.e. each point corresponds to an RR interval, and histograms showing the optimum model order (x-axis) as found by the (b) FPE, (c) AIC, (d) CAT and (e) RIS criteria corresponding to the tachogram shown in (a). The ECG of this subject included some ectopic beats (mostly pre-ventricular cycles) and these were left untouched and included in the processing.

associated to different physiological mechanisms. This situation would require a model order of at least $p = 6$ to correctly estimate the AR-based power spectrum. It must be noted here that due to the very short length of the data frames we use here it is not possible to estimate the VLF component. Although the AR spectral analysis technique fares better than FFT-based approaches for short segments, since each of the segments is 32 s long, we expect to only estimate frequencies above 0.031 Hz correctly.

The underestimation of the correct AR order and its effect on the spectra is illustrated by figure 4(a) where the AR-based power spectra for 500 frames of night time data for subject 1

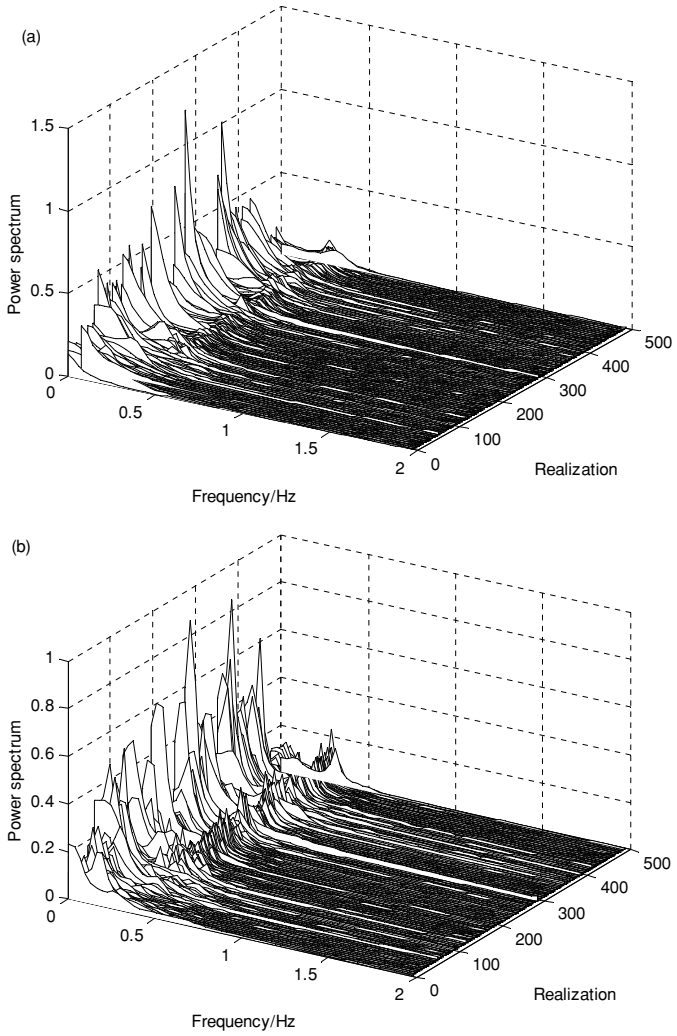

Figure 4. AR power spectra for 500 frames of night time data for subject 1 using the recommended AR order of $p = 5$ (a) and $p = 16$ (b). It can be seen that the increased model order improves spectral resolution.

have been obtained using an order $p = 5$. If this is compared with the power spectra obtained for the same data using an order $p = 16$, shown in figure 4(b), it is found that the RSA peaks in the spectrum are more easily identified for the higher order model. These spectra can also be compared with the corresponding Fourier-based (modified periodogram) spectra shown in figure 5. Note that, due to the fact that short frames of 32 s are used, the resolution of the FFT-based spectrum suffers, and the evidence of this is that the peaks at RSA are quite dull.

3.1. Model order progression

Figure 6 shows how the predicted model order changed over time along the original signal for subject 1. It can be seen that the FPE, AIC and CAT criteria all predict practically the same model order for each frame of data. This was found similarly for the other sets of data. The

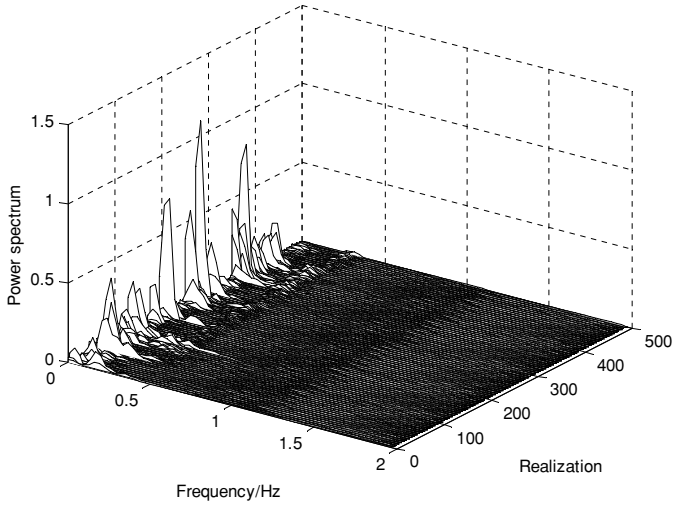

Figure 5. Fourier-based (modified periodogram) power spectra for 500 frames of night time data for subject 1. This allows a comparison between the parametric and non-parametric methods of spectral estimation to be made.

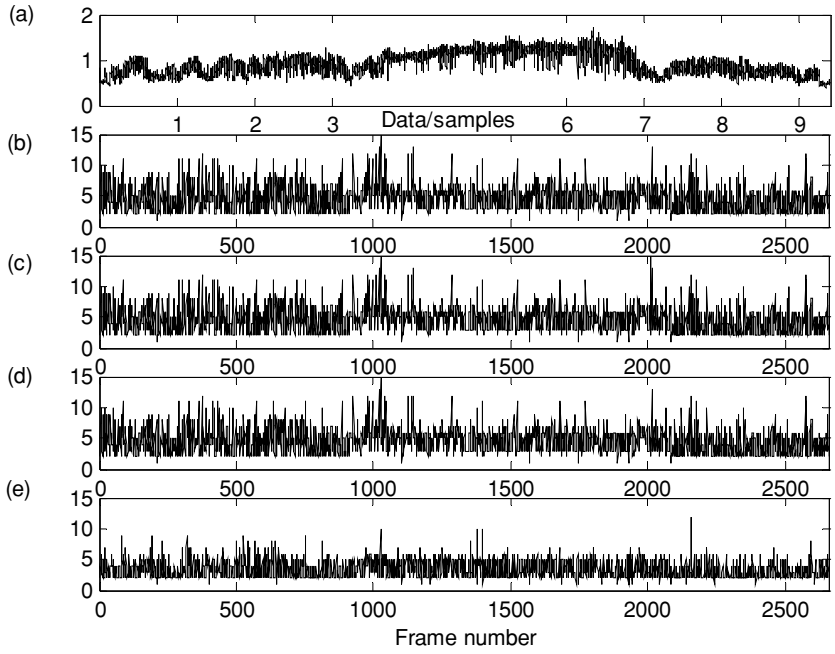

Figure 6. Plot of the signal for subject 1 for 24 h and the progression of the AR model order as predicted by FPE, AIC, CAT and RIS. It can be seen that nearly all the predicted model orders are the same for the first three criteria, but RIS estimates smaller orders.

behaviour of the orders predicted using AIC and FPE agrees with the conclusions of Jones (1974), who stated that these two criteria normally predict identical model orders for the same frames of data. Rissanen's criterion, however, often estimates a lower model order compared to the other three criteria, as seen in figure 6(e).

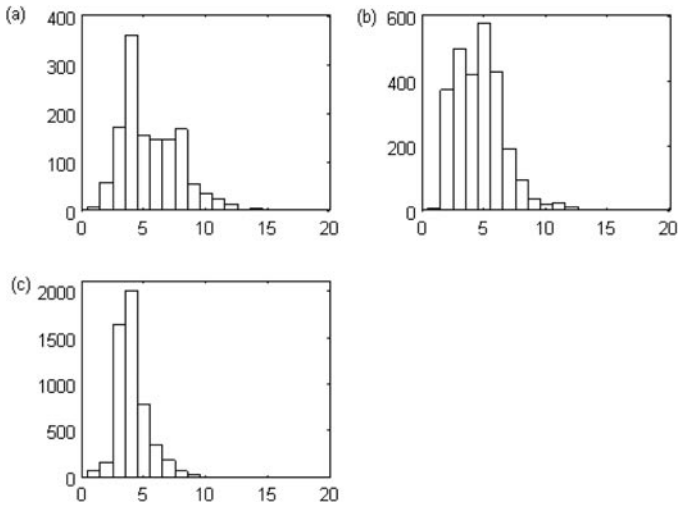

Figure 7. Model orders estimated using FPE for re-sampling frequencies of (a) 2 Hz, (b) 4 Hz and (c) 8 Hz. We expected that the estimated AR model orders would increase with the re-sampling frequency and this happens for the transition from 2 Hz to 4 Hz, but not for the 8 Hz re-sampling frequency.

3.2. Sampling frequency

The model orders estimated by the above four prediction criteria for an autoregressive process cannot be discussed without considering the re-sampling frequency used in conjunction with the cubic spline interpolation. In this part of the study all the data was analysed using re-sampling frequencies of 2, 4 and 8 Hz and the orders estimated using the FPE criterion (AIC and CAT also estimate similar orders) are given in figure 7 for subject 1. It can be seen that for a re-sampling frequency of 2 Hz the model orders estimated range roughly from 2 to 10, with a peak in $p = 4$; for a re-sampling frequency of 4 Hz this distribution shifts slightly to the right, meaning that higher orders are needed, as expected; but for a frequency of 8 Hz this shift occurs in the opposite direction and we did not expect nor can we explain this.

4. Testing the order prediction criteria using a true AR process

The next part of this study allowed the four model order prediction criteria to be tested using a Monte Carlo simulation of a *true* AR process. This was done by creating a true AR signal with a known order, $p = 6$, using coefficients extracted from the actual signal (to preserve its characteristics) and the filter given by equation (1). To allow the system to stabilize, each realization of this test signal was produced from the last 128 of 1024 filtered values; this was repeated to produce 1000 realizations. The filter coefficients were found from one of the frames of the actual data used in the first part of the study for an order 6 system. Order 3 was predicted most often but, since in the general case we would be interested in identifying up to three frequency components, an order of at least $p = 6$ would be required. As mentioned earlier, due to the very short segments used we cannot expect to accurately estimate the VLF, but since LF and HF are relatively close, we still need an AR order larger than 4 (twice the number of frequencies) to estimate the AR-based spectrum.

Histograms were plotted showing the results obtained and these can be seen in figure 8. It was again noted that the four model order prediction criteria more often than not underestimate

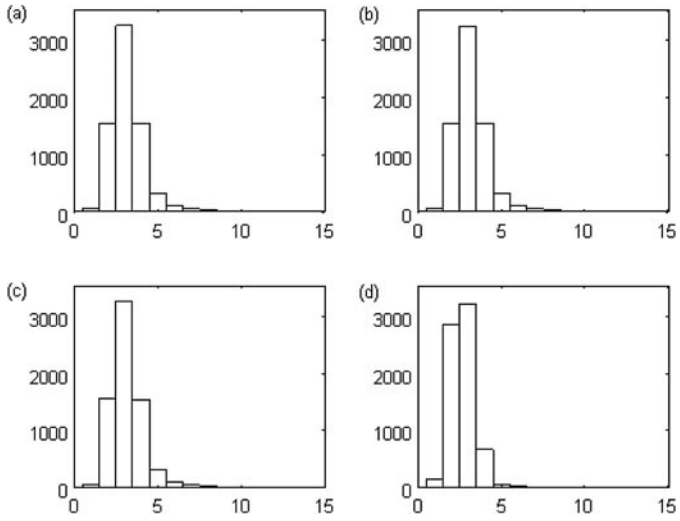

Figure 8. Histograms showing the 'best' orders predicted by (a) FPE, (b) AIC, (c) CAT and (d) RIS, resulting from the analysis of realizations of a true AR signal of order 6, each realization containing $N = 128$ samples. The histograms show that the order is underestimated by all four criteria.

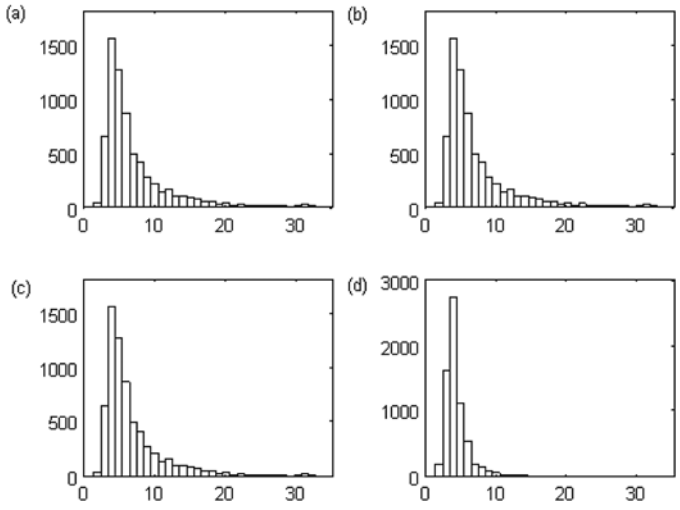

Figure 9. Histograms showing the 'best' orders predicted by (a) FPE, (b) AIC, (c) CAT and (d) RIS resulting from the analysis of realizations of a true AR signal of order $p = 6$, each realization containing $N = 1024$ samples. It can be seen that order 4 is predicted most by all four criteria, but higher orders are predicted more often than for the case where $N = 128$, with RIS tapering sooner than the other three methods.

the correct model order. Increasing the frame length to include $N = 1024$ samples resulted in a peak in the most frequently predicted AR model order at $p = 4$, and with a wider distribution to the right, i.e. higher orders got selected more often than when $N = 128$. The histograms of the distributions for $N = 1024$ can be seen in figure 9.

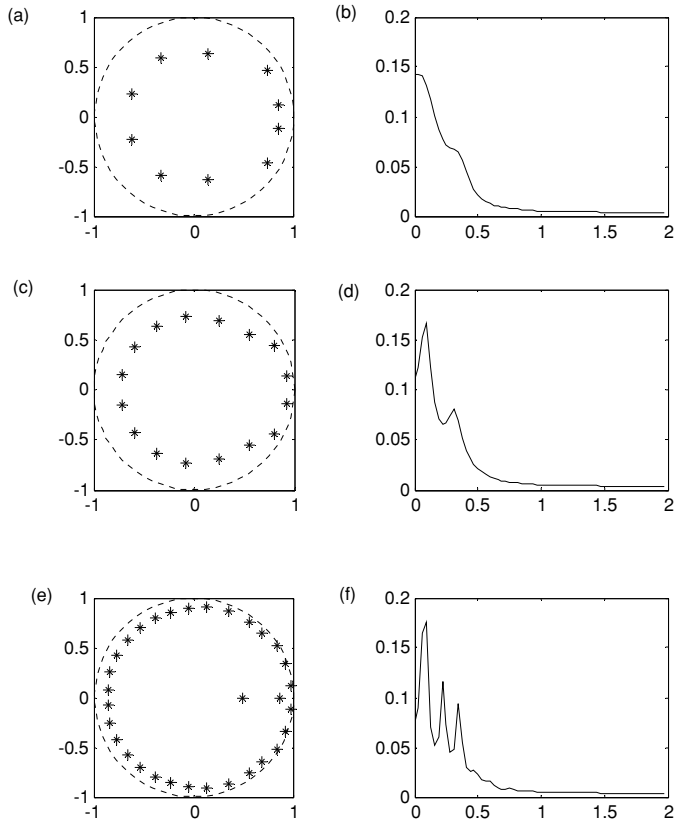

Figure 10. This comparison shows the effect that changing the model order has on the spectra for one specific frame of data. In (a) and (b) the AR model order is $p = 10$; in (c) and (d) the model order is $p = 16$, and for (e) and (f) the model order is $p = 32$. It can be seen that for $p = 16$, the spectrum shows definite spectral peaks without spectral smearing or peak splitting.

5. Specifying a fixed model order

From the results discussed so far it is apparent that for power spectrum analysis of all normal HRV data over short segments, the four prediction criteria do not consistently predict model orders which could be used to find accurate power spectra for the associated data. For short segments ($N = 128$) of known AR processes underestimations of the correct order are produced more often than not, as shown in figure 8. It can therefore be concluded that none of the four criteria is suitable for application to short segments of HRV data. Even with many samples ($N = 1024$) of true AR processes the orders predicted vary considerably and often the estimated order is smaller than the true order, as summarized in figure 9.

It is proposed then, that a fixed model order should be used, the advantage being that it resolves the problem of varying predicted orders while still allowing 'good' spectra to be produced. A survey was conducted using seven frames of data from each subject and producing power spectra for each frame for orders increasing in multiples of 2, from $p = 10$ to $p = 32$. One set of these results is shown in figure 10 along with the relevant pole-zero diagrams.

The pole-zero diagram is a very compact way of displaying the properties of a system and provides an alternative way of visualizing its frequency response. When the poles appear equally close to the edge of the unit circle, it implies that they have roughly equal influence on the spectrum. For a pole away from the unit circle, closer to its centre, the implication is that this pole, or its corresponding filter coefficient, does not contribute significantly to the overall spectrum.

It was found that the lowest model order required to accurately show all spectral components for the different frames behaved in the following way: in the worst case, for $p = 10$ to $p = 14$ the spectra were damped and smeared; for $p > 22$ spurious peaks occurred frequently. Between $p = 16$ and $p = 22$ there was no notable change in the spectra and it is desirable to use the lowest order possible since this parsimony reduces computation time. In all cases, for a model order of $p = 16$, the spectra contained easily resolvable peaks and no spurious peaks or smearing. For these reasons $p = 16$ is recommended as the optimum fixed model order to be used in this type of analysis.

6. Conclusions

Overall it has been found that three of the four criteria (FPE, AIC and CAT) estimate similar values for each set of data tested. However this value is very often an underestimation of the true AR model order when short segments ($N = 128$) are used, and this gives rise to damped power spectra whose true separate frequency components cannot be recognized.

It was found that for longer frames of data ($N = 1024$) FPE, AIC and CAT criteria all gave an improved estimate for the optimum model order when a simulated AR signal of known order was tested.

It was confirmed that underestimating the AR model order has more dramatic effects in the spectral estimation than overestimation, with only gross overestimation producing spurious spectral peaks.

The re-sampling frequency used for the spline interpolation was found to influence the predicted model order and over-sampling the signal would simply result in longer processing time being required, sometimes with higher model orders being necessary, but with no benefits for AR spectral estimation.

To allow an accurate estimation of the power spectrum of RR time series signals re-sampled at 4 Hz, the order obtained from using the prediction criteria should be raised slightly, or alternatively (and we prefer this approach) an overall fixed order around $p = 16$ should be used.

References

Akaike H 1969 Fitting autoregressive models for prediction *Ann. Inst. Stat. Math.* **21** 243–7

Akaike H 1974 A new look at the statistical model identification *IEEE Trans. Autom. Control* **19** 716–23

Akselrod S, Gordon D, Ubel F A, Shannon D C, Berger A C and Cohen R J 1981 Power spectrum analysis of heart rate fluctuations: a quantitative probe of beat-to-beat cardiovascular control *Science* **213** 220–2

Cerutti S, Bianchi A M and Mainardi L 1995 Spectral analysis of the heart rate variability signal *Heart Rate Variability* ed M Malik and A J Camm (Armonk NY: Futura Publishing Company) pp 63–74

Jones R H 1974 Identification and autoregressive spectrum estimation *IEEE Trans. Autom. Control* **19** 894–7

Kaluzynski K 1989 Order selection in Doppler blood flow signal spectral analysis using autoregressive modeling *Med. Biol. Eng. Comput.* **27** 89–92

Karemaker J M 1997 Heart rate variability: why do spectral analysis? *Heart* **77** 99–101

Kay S M and Marple S L 1981 Spectrum analysis—a Modern perspective *Proc. IEEE* **69** 1380–419

Marple L 1977 Resolution of conventional Fourier, autoregressive and special ARMA methods of spectral analysis *IEEE International Conf. on ASSP* pp 74–7

MIT-BIH 1992 Arrhythmia Database Tape Directory and Format Specification, Massachusetts (Harvard University, Massachusetts Institute of Technology, Division of Health Sciences and Technology) (Cambridge MA USA)

Parzen E 1975 Multiple time series: determining the order of approximating autoregressive schemes *Technical Report* No **23** (Buffalo, NY: Statistical Sciences Division, State University of New York)

Rissanen J 1984 Universal coding, information prediction and estimation *IEEE Trans. Inf. Theory* **30** 629–36

Sayers B M 1973 Analysis of heart rate variability *Ergonomics* **16** 17–32

Schlindwein F S and Evans D H 1990 Selection of the order of autoregressive models for spectral analysis of Doppler ultrasound signals *Ultrasound Med. Biol.* **16** 81–91

Ulrych T J and Bishop T N 1975 Maximum entropy and spectral analysis and autoregressive decomposition *Rev. Geophys. Space Phys.* **13** 183–200

van Ravenswaaij-Arts C M A, Kollee L A A, Hopman J C W, Stoelinga G B A and van Geijn H P 1993 Heart rate variability *Ann. Internal Medicine* **118** 4–47