

Application of Autoregressive Spectral Analysis to Missing Data Problems

Piet M. T. Broersen, Stijn de Waele, and Robert Bos

Abstract—Time series solutions for spectral analysis in missing data problems use reconstruction of the missing data, or a maximum likelihood approach that analyzes only the available measured data. Maximum likelihood estimation yields the most accurate spectra. An approximate maximum likelihood algorithm is presented that uses only previous observations falling in a finite interval to compute the likelihood, instead of all previous observations. The resulting nonlinear estimation algorithm requires no user-provided initial solution, is suited for order selection, and can give very accurate spectra even if less than 10% of the data remains.

Index Terms—Autocovariance estimation, missing observations, order selection, parameter estimation, spectral estimation, Vostok data.

I. INTRODUCTION

OBERVED data are often incomplete. In controlled experiments, sensor failure or outliers lead to missing data. In meteorological, hydrological, astronomical, or satellite observations, the weather conditions may disturb the equidistant sampling scheme. In paleoclimatic data, the relation between the chronological time and the physical depth causes an observed time series with missing observations on an equidistant time grid [1], [2]. The pattern of data that are missing is important for the analysis. It will often be random for sensor failure or outliers, it may have large gaps due to weather conditions, and it will not be stationary for paleoclimatic data where recent observations are missing less frequently than very old ones. The paper is mainly concerned with randomly missing data.

The treatment of missing data has two different principles: reconstruct or interpolate the missing data and analyze the new equidistant signal or use only the remaining data. The first group of methods relies on estimation algorithms that have been developed for consecutive equidistant data. Three methods that extract equidistant information from missing data will be distinguished in this category. The first is some sort of static interpolation between the remaining observations, linear, nearest neighbor, sample and hold, or splines. Those methods have been studied extensively. The performance depends heavily on the characteristics of the data. It may be accurate for slowly varying low frequency signals, but no interpolation method gives good results for all types of data [3]. The second method reconstructs the missing data with a model for the covariance structure. This method is derived from the EM algorithm for missing data. The

E step finds the conditional expectation of the missing data given the observed data together with the currently estimated model; the M step computes maximum likelihood estimates for the parameters of the model from all consecutive data, observed and reconstructed [4]. The third reconstruction method is only applicable in simulations because it uses the true process covariance in the reconstruction of the missing data. It can be considered as the ultimate possible reconstruction. If reconstruction with the knowledge of the true process is not sufficient for accurate spectral estimation in comparison with other methods, no reconstruction method can be.

The method of Lomb–Scargle [5], [6] uses only measured data. It computes Fourier coefficients as the least squares fit of sines and cosines to the available observations. The Lomb–Scargle spectrum is accurate in detecting spectral peaks but rather poor in describing slopes in the spectrum [7]. A second idea is to find some covariance estimate for the incomplete data and to use that for further analysis [8]. This technique does not always guarantee that the sample covariance estimate is positive semidefinite [9]. Moreover, the sample covariances are known to be inefficient estimators for the covariance structure and will generally not produce accurate spectral estimates, not even for equidistant data [10]. The third idea selects consecutive segments in the observed data and uses a special segment variant of the Burg algorithm to compute the parameters of an autoregressive (AR) model [11]. It can be successful if only a few parameters have to be computed or if large gaps alternate with longer consecutive segments. Its use is limited to rather small randomly missing fractions, because only uninterrupted data segments can be used. The final type fits a time series model directly to the available observations with maximum likelihood estimation. An exact maximum likelihood approach using Kalman filtering can be calculated [12] and also a finite interval approximation [13]. The computational effort of the exact approach depends on the total observation length of missing and available data; the computation time for the approximation depends mainly on the available number of data.

This paper is organized as follows. An introduction in time series shows that the AR parameters represent the spectrum and the covariance function of measured data. The conditional density of joint normally distributed variables is used in Section III to derive an approximate maximum likelihood estimator for missing data AR models. Some specific choices for a good and efficient numerical performance of this method are described and the resulting spectra are compared with several existing methods for randomly missing data. An investigation into the possibility of order selection is made. Furthermore, the new

Manuscript received June 15, 2003; revised March 31, 2004.

The authors are with the Department of Applied Physics, Delft University of Technology, 2600 GA Delft, The Netherlands (e-mail: broersen@tn.tudelft.nl).

Digital Object Identifier 10.1109/TIM.2004.830597

method is applied to practical paleoclimatic data to demonstrate that spectral estimates can be obtained for important practical problems. In this example, the recent observations are consecutive but many data in the far past are missing.

II. AR THEORY

AR models describe the characteristics of stationary stochastic processes [14]. The power spectrum and the complete covariance function are determined by the parameters of the AR model and by the variance. With consecutive data, the best model *order* can be selected objectively and automatically, based on reliable statistical criteria [10]. The observations x_n of an AR(p) process with the parameters a_i , $i = 1, \dots, p$, can be written as [14]

$$x_n + a_1 x_{n-1} + \dots + a_p x_{n-p} = \varepsilon_n \quad (1)$$

where ε_n is a purely random process with variance σ_ε^2 . Almost any *stationary* stochastic process can be written as a unique AR(∞) process. In practice, estimated finite-order models will include all statistically significant details. Hence, this model type can be applied to random physical, meteorological, or astronomical phenomena. The autocovariance function and the power spectral density of the data can be computed from the estimated model parameters. The power spectral density $H(\omega)$ of the AR(p) model is thus given by [14]

$$H(\omega) = \frac{\sigma_\varepsilon^2}{|A(e^{j\omega})|^2}. \quad (2)$$

All elements of the infinitely long *true* normalized correlation function are determined by the p *true* parameters of (1) with the Yule–Walker relations [15]

$$\rho_n + a_1 \rho_{n-1} + \dots + a_p \rho_{n-p} = 0, \quad n > 0; \quad \rho_{-n} = \rho_n \quad (3)$$

where ρ_n denotes $E(x_i x_{i+n}) / \sigma_X^2$ and σ_X^2 is the variance of the signal x_n . The same relations (3) can be used to compute the *estimated* correlation function, belonging to *estimated* AR parameters. Also the inverse use of the Yule–Walker relations (3), computing p' parameters from p' true or estimated correlations, is often applied.

III. ARFIL ALGORITHM: AR FINITE INTERVAL LIKELIHOOD

AR models for missing data can be estimated with an exact maximum likelihood (ML) technique [12]. However, the exact likelihood theory may give computational problems in practice because it computes the recursive Kalman prediction for each time instant, which is not efficient if too many data are missing. Therefore, an approximate likelihood solution is developed here, with special care for the numerical robustness in each stage of the algorithm.

The general prediction theory of Gaussian data gives a possibility to estimate AR(p) parameters from N randomly available observations. The observations at time $t(i)$, $i = 1, \dots, N$, are denoted $x_{t(i)}$; each $t(i)$ is a multiple nT of the sampling time T . For notational simplicity, the observations are separated in two groups Z and Y , with Z being given by $x_{t(k)}$, the single observation that will be predicted using the $N-1$ dimensional vector

Y that consists of all available observations except $x_{t(k)}$. Z has mean μ_x and variance σ_x^2 , and Z has a joint multivariate normal distribution with the $N-1$ dimensional vector Y , with the same mean value μ_x for each element and with the $(N-1) \times (N-1)$ covariance matrix R_{YY} or $\sigma_x^2 \rho_{YY}$. The cross covariance matrix is denoted R_{ZY} or $\sigma_x^2 \rho_{ZY}$. Then the conditional density of Z for given Y is a one-dimensional normal distribution with mean and variance given by [16, p. 25]

$$\begin{aligned} \mu_{Z|Y} &= \mu_x + R_{ZY} R_{YY}^{-1} (Y - \mu_x) \\ &= \mu_x + \rho_{ZY} \rho_{YY}^{-1} (Y - \mu_x) \\ \sigma_{Z|Y}^2 &= \sigma_x^2 s_{Z|Y}^2 \\ &= \sigma_x^2 [1 - \rho_{ZY} \rho_{YY}^{-1} \rho_{YZ}]. \end{aligned} \quad (4)$$

Hence, $\mu_{Z|Y}$ is the best prediction of Z for given Y for normally distributed observations. In applying (4) to missing data, the unconditional means and variances of Z and Y are the same: the constant mean μ_x and variance σ_x^2 of the signal. All remaining observations are required in the vector Y in (4), from past and future. This number can be limited to only the past observations with a general relation for the joint distribution $f(X)$ of N normally distributed variables x_i [17, p 255]

$$f(X) = \left(\prod_{k=2}^N f(x_k | x_1, \dots, x_{k-1}) \right) f(x_1). \quad (5)$$

The joint distribution $f(X)$ can be written exactly as a product of conditional densities of increasing time index. This product of conditional probabilities in chronological order will be used to construct a prediction with (4) for an available observation $x_{t(k)}$ at time $t(k)$ from only previous available, not consecutive $x_{t(k-i)}$. Minimizing the difference between those predictions and the actual observations carries out the parameter estimation. The covariance matrix of the observations of an AR(p) process is required in (4). That covariance structure is completely determined by the AR parameters with (3). The best prediction from *consecutive* observations requires the previous p observations in (1) [14]. If data are missing, an AR(1) process is still best predicted by the single closest previous observation. However, higher order AR(p) processes require all previous observations for the best prediction if data are missing.

The estimated AR(p) parameter vector $\hat{\mathbf{a}}$ with elements $\hat{a}_1, \dots, \hat{a}_p$ is estimated in the case of missing data by minimizing the negative of the log of the likelihood function defined as $L(X; \hat{\mathbf{a}}) = -2 \log f(X) - N \log 2\pi$, with respect to $\hat{\mathbf{a}}$. The exact algorithm for the likelihood, based on this principle, would still require too much computing time in practice, even for moderate N . That is because the size of the predictor Y would become $N-1$ for the last observation and an $(N-1) \times (N-1)$ matrix has to be inverted. Generally, the prediction accuracy improves most with the nearest previous observations, and observations farther away have almost no influence. Therefore, the maximum length of the predictor with an AR(p) model is limited to a finite time interval

$$t(j) - t(j - K) \leq \frac{2pT}{\gamma} \quad (6)$$

where γ is the remaining fraction of the data. This algorithm is *AR for finite interval likelihood* (ARFIL). The number of observations K within the interval (6) has as average $2p$. It has been verified in several simulations that using a larger interval in (6) would not have a noticeable influence on the estimated parameters, but it would only require longer computing times. Taking a much smaller interval reduces the accuracy. Of course, only the first p predecessors contribute to the prediction accuracy if the observations are consecutive.

To improve the numerical robustness, reflection coefficients k_i are estimated instead of parameters. The ARFIL algorithm uses the unconstrained optimization of $\tan(\pi/2 * k_i)$ for increasing orders p . This guarantees that the estimated reflection coefficient k_i is always in the range $-1 < k_i < 1$. Hence, all models computed by nonlinear numerical optimization routines are stationary. The usual Yule–Walker or Burg algorithms keep previous reflection coefficients constant in computing the new k_p [15]. In contrast with those consecutive algorithms, however, all $k_i, i = 1, \dots, p$, are optimized afresh and simultaneously in ARFIL, for every model order p . Initial solutions for the nonlinear optimization at order p are the reflection coefficients of the $AR(p-1)$ model with an additional zero for k_p .

Order selection can be performed with the criterion

$$\text{GIC}(p, \alpha) = L(\hat{X}; \hat{a}) + \alpha p. \quad (7)$$

The best value for the penalty factor α still has to be investigated in the missing data case; The criterion (7) with $\alpha = 2$ is the famous AIC criterion [15]. For consecutive data, the choice $\alpha = 3$ is used in automatic spectral analysis [10] because it gives a better practical compromise between underfitting and overfitting models.

IV. SIMULATIONS

First, the computational requirements of the ARFIL ML algorithm of Section III have been compared to the exact ML algorithm of Jones, with Kalman prediction [12]. The numerical outcomes are almost always identical for both algorithms. The Kalman prediction time is more or less proportional to the total number N/γ and is faster for $\gamma > 0.05$. The time required for ARFIL is proportional to the number of remaining observations N and ARFIL is faster for smaller fractions γ . Those computing times depend also somewhat on the AR order, where ARFIL is more sensitive. For estimated $AR(10)$ models, both methods require about the same time for $\gamma = 0.05$. The Kalman filtering is roughly ten times faster for $\gamma = 0.5$ and ARFIL is 50 times faster for $\gamma = 0.001$. Therefore, the remaining fraction γ determines which ML algorithm is preferable.

The accuracy of estimated models is evaluated with the model error (ME). This is a relative measure in the frequency domain: the integrated ratio of the spectral difference between model and true process on the one hand and the true spectrum on the other hand. Also a time-domain expression for the model error ME exists as a normalized prediction error (PE) [10]

$$\text{ME} = N \left(\frac{\text{PE}}{\sigma_\varepsilon^2} - 1 \right). \quad (8)$$

The PE is the mean square error of the fit of an estimated model to new consecutive data of the same process. The multiplication

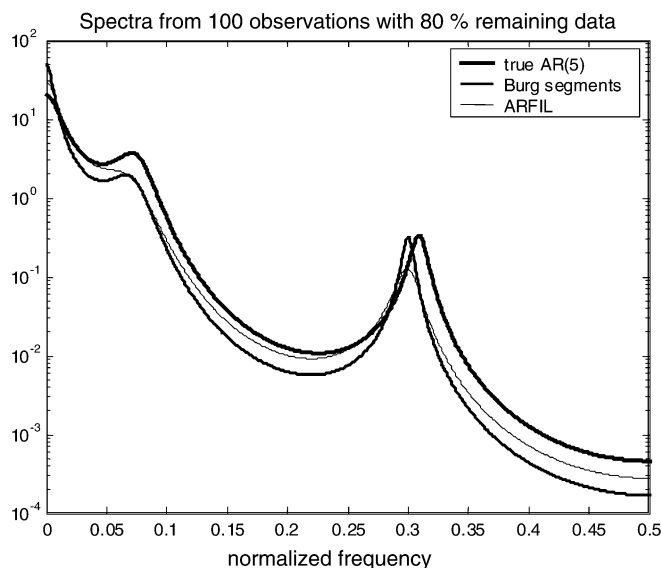


Fig. 1. Spectra of the $AR(5)$ process and of two $AR(5)$ estimates from a realization of 100 remaining observations with 20% missing data. The ME for Burg on segments is 13.2, the ME for ARFIL is 3.7.

with N gives the ME an expectation that is independent of the sample size, for unbiased models estimated from consecutive data, with p' as the minimal expectation of the ME for unbiased $AR(p')$ models.

A comparison of the spectral accuracy of the ML method with the Lomb–Scargle method and with the sample covariance estimate has been made, but the results of those methods do not deserve further mention. The Lomb–Scargle quality measured in the ME is generally a hundred or more times worse, as it is in a comparison of periodograms with time series spectra [10]. In spectra with peaks and deep valleys, the Lomb–Scargle spectra display a serious bias in the frequency range where the spectral density is small. This method is not reliable in describing the global spectral shape; the asset of this method is the capability to detect periodicities against a more or less white background noise.

ML will be compared to Burg on segments and to two reconstruction methods. Those methods reconstruct the missing data and estimate the AR parameters from the newly obtained consecutive signal, consisting of the true observations and the model predictions for the missing data. The reported reconstructions used are the following.

- 1) Linear interpolation to reconstruct the missing data. This was in most examples more accurate than nearest neighbor, cubic, or spline interpolation.
- 2) Reconstruction of the data with the true covariance of the data. It has been verified that this was always the best reconstruction in terms of giving the most accurate spectral estimate. All recursive reconstruction variants like expectation-maximization [10] that alternate estimation and reconstruction always had higher model errors, often much higher.

The true process in the simulations is an $AR(5)$ process with parameters $[1 \ -1.98 \ 1.76 \ -1.78 \ 1.75 \ -0.73]$.

This gives a spectrum with slopes and a strong peak at the frequency $0.3/T$ as is shown in Fig. 1. Also two individual esti-

TABLE I
THE AVERAGE MODEL ERROR ME AS A FUNCTION OF THE REMAINING DATA FRACTION γ . $N = 100$ REMAINING OBSERVATIONS, ESTIMATED FIXED ORDER AR(5) MODELS OF A TRUE AR(5) PROCESS

Method	$\rightarrow \gamma$.95	.9	.8	.5
AR ML		11.8	12.6	8.5	40.7
Burg on segments		15.3	21.5	26.3	1217.9
True reconstructed		13.6	15.3	9.5	167.1
Linear interpolated		45.4	85.0	129.2	200.1

TABLE II
THE AVERAGE MODEL ERROR ME AS A FUNCTION OF THE REMAINING DATA FRACTION γ . $N = 1000$ REMAINING OBSERVATIONS, ESTIMATED FIXED ORDER AR(5) MODELS OF A TRUE AR(5) PROCESS

Method	$\rightarrow \gamma$.95	.9	.8	.5
AR ML		5.1	4.9	4.5	30.2
Burg on segments		7.3	8.9	13.8	139.3
True reconstructed		5.3	6.1	10.7	60.1
Linear interpolated		154.6	384.9	810.3	1886.9

TABLE III
THE AVERAGE MODEL ERROR ME FOR THE TRUE FIXED ORDER ESTIMATED MODELS OF THREE PROCESSES. $N = 500$ REMAINING OBSERVATIONS, REMAINING FRACTION $\gamma = 0.75$

Method	AR(5)	AR(5) _B	AR(2)
AR ML	4.8	13.3	1.9
Burg segments	24.4	106.6	3.3
True reconstructed	8.9	14.8	9.9
Linear interpolated	514.7	3778.1	3.6

mated AR(5) spectra are given in the figure, to demonstrate the accuracy and also to give an impression of the measure ME that is used. Tables I and II give some average numerical results. In all examples and for all fractions γ , the average ML quality is better than Burg on segments and also better than reconstruction with the true covariance. In this example, linear interpolation is not accurate at all. However, some smooth simulation examples have been found where the performance of linear interpolation is reasonable for γ close to one. Finally, all methods except ML and sometimes Burg for segments lose most of their accuracy for γ less than about 0.5. The test results for lower γ are not reported. The results indicate that reconstruction methods may be accurate for few missing data, but not as good as ML estimation. However, if 50% of the data or more is missing, reconstruction is very inaccurate.

The simulations in Table III give results of three processes, for $\gamma = 0.75$ and for $N = 500$. The first process, denoted "AR(5)," is the same as above; the second process "AR(5)_B" is an AR(5) process with a strong spectral peak on a weak and almost white background; and the third process "AR(2)" is a low-order AR(2) process. All conclusions from the first two tables remain applicable. The only exception is that linear interpolation is rather accurate for the AR(2) example. Furthermore, Burg on segments has the advantage that the model order is only two, which gives contributions of all segments of length three and greater to the AR(2) estimation. The consecutive Cramér–Rao lower bound for the ME would be five, five, and two for the three processes, respectively. For all three processes, the quality of ML spectra is the best of the four reported spectra. It should be mentioned, however, that all processes are low-order AR processes with known order.

Taking $N = 10000$, $\gamma = 0.1$ with the first AR(5) example gives the opportunity to test the possibility of order selection in difficult conditions. The highest candidate order was taken as AR(10). The average ME of the AR(5) model was 132; of AR(10) it was 14 697. The inaccuracy measured in the ME increases very much with the model order. The average ME of selected models depends strongly on the penalty factor α in (7). For $\alpha = 2$ and 3, the average ME of the selected model was 2121 and 76, respectively. It is obvious that $\alpha = 2$ gives a poor result, like it does in consecutive data [10]. It is remarkable that the average ME of models selected with $\alpha = 3$ is better than the ME of the average estimated AR(5) model with the fixed true order. The explanation is simple. Sometimes, in individual simulation runs, the nonlinear minimization algorithm did not converge completely for the true order but it did converge for a higher order model that used the lower order as initial condition. In those cases, the selection criteria select the higher order models and order selection helps to improve the quality of the estimated spectrum. The influence of the penalty factor α in the selection criterion (7) is very similar to that in consecutive data. Many other simulations have supported the preference for $\alpha = 3$ for AR order selection, for different missing fractions, for various N , and for many processes.

The very high model error for the AR(10) model demonstrates that the accuracy of the higher order models deteriorates quickly if many data are missing. In consecutive data, the highest AR candidate order can be $N/2$ without any problem, but missing data allow much lower AR orders. About $N\gamma$ sequences of length two are present in the average, $N\gamma^2$ sequences of length three, and so on [7]. Therefore, AR(p) models can be estimated with some reasonable accuracy with the Burg on segments algorithm if $N\gamma^p$ is greater than ten, because then about ten consecutive segments contribute to the p th parameter [7] and more to the lower orders. The requirements for the ML method are much less severe. Often much higher order models can be estimated with ML with sufficient accuracy. This shows that for every value of γ , however small, a minimum remaining sample size N can be found for which an accurate AR(p) model can be estimated.

V. EXAMPLE OF REAL-WORLD DATA

An ice core of length 3310 m is drilled from the ice cap at the Vostok station in East Antarctica [1]. The ice core with sediments provides information over a period of 420 ky. The core is analyzed with a spatial distance of 1 m between the samples. A time scale is linked to the depth by combining theoretical flow and accumulation models for the ice. The equidistant depth scale gives irregular distances at the time scale. Equidistant resampling with a period of four times per ky and accepting only observations within a time slot of 1/8 ky from the resampling instants gives an equidistant missing data temperature signal of length 1474. Almost all time instants in the first part have an observation within the slot, but a lot of data are missing in the eldest period. This can be recognized in Fig. 2 where data are sparser with increasing age.

The ARFIL algorithm and the Burg on segments algorithm have been applied to the paleoclimatic data [1] in Fig. 3. The

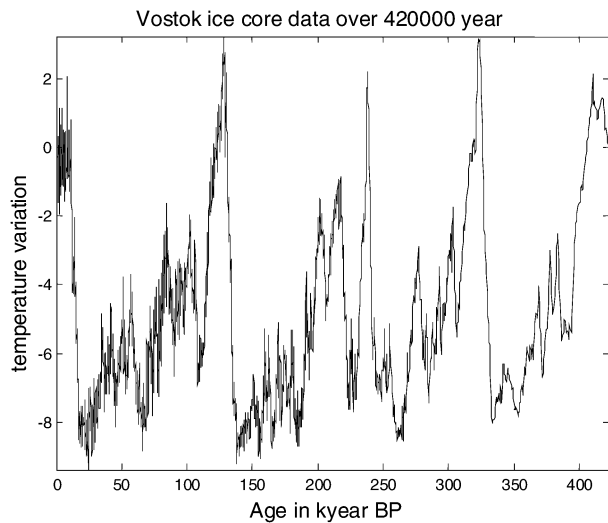


Fig. 2. Paleoclimatic temperature variation data, extracted once per meter from the Antarctica Vostok 3310-m-long ice core data and ordered along an estimated time scale.

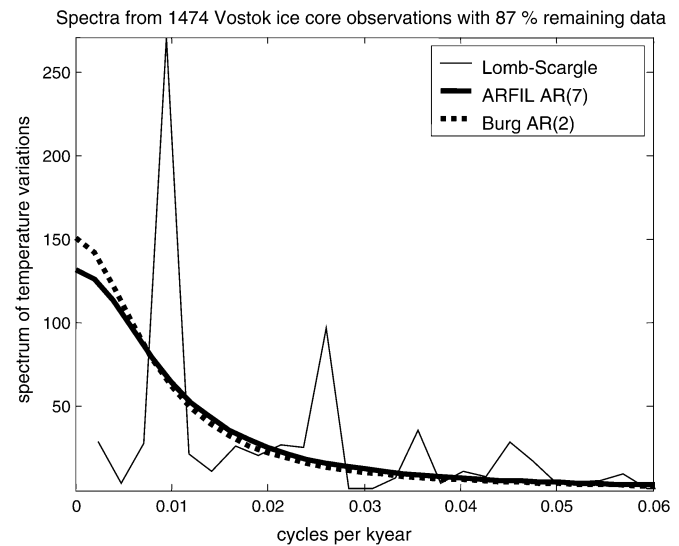


Fig. 4. Comparison between Lomb–Scargle and time series for spectral estimation of paleoclimatic temperature data from the Antarctica Vostok data. Only a small portion of the frequency scale is shown.

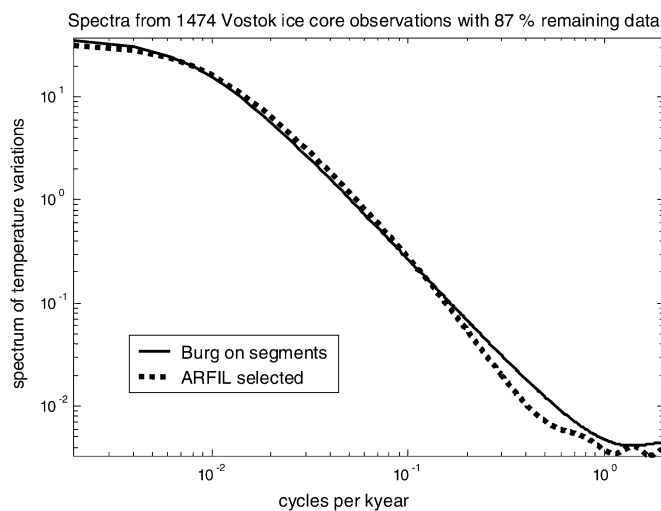


Fig. 3. Estimated spectra of paleoclimatic temperature data from the Antarctica Vostok 3310-m-long ice core data. Burg on segments selected an AR(2) model, ARFIL selected the AR(7) model.

selected order was two for Burg on segments; for ML it was seven. Obviously, ARFIL selects more details as statistically significant. However, the order selection criterion had only a slight preference for order seven; all orders between two and ten give very similar values for $GIC(p, 3)$. The Burg spectrum for order seven looks very much like the ARFIL spectrum in Fig. 3. Taking somewhat higher AR orders, some more small wiggles appear in the high-frequency end, above 0.5 cycles per ky. Taking the estimated ARFIL model of order seven as a reference, the ME value of the Burg model of order seven was two and all Burg models with orders between three and 25 had ME values less than 33. No peaks at low frequencies have been detected with the time series methods.

This conclusion is in contrast with the spectra that have been computed with Fourier techniques like Lomb–Scargle spectra for missing data or modified periodograms for consecutive data

[1]. The Lomb–Scargle spectrum is shown in Fig. 4. The frequency scale is truncated and the vertical scaling is made linear for an easy comparison with previous spectral results [1]. Qualitatively, the appearance of the Lomb–Scargle spectrum is close to the temperature difference spectrum as presented in [1]. The two time series spectra are also close to each other on the linear scale and quite different from the Lomb–Scargle spectrum.

Lomb–Scargle treats the data as periodic with noise; time series methods treat them as a finite number of observations of a stationary stochastic signal. The difference in the estimated spectra is remarkable in Fig. 4. No undisputable preference can be given. If one wants to predict the temperature variations in the future by assuming that the temperature behavior will be stationary over the centuries, the stochastic signal approach is preferable. Only if one wants to treat the data as quasi-deterministic, implying that taking another ice core on another location would have given identical data with perhaps some negligible noise contribution, would the Lomb–Scargle spectrum be proper to use. However, predictions for the future are not justified then. It is peculiar that the appearance of the Lomb–Scargle spectrum just looks like a possible realization of a Lomb–Scargle estimate computed from 1474 AR(7) data. This has been verified by using the AR(7) model as generating process and applying the ARFIL and the Lomb–Scargle estimator to many realizations. ARFIL estimates are close to the generating spectrum in every realization; Lomb–Scargle spectra are very wild and different in every realization.

Fig. 5 shows that the Burg spectrum becomes similar to Lomb–Scargle for an estimated AR(250) model. Even the AR(100) model shows no peaks in the very low frequency range. This demonstrates that although time series have a model similar to Lomb–Scargle among the candidates, order selection will not consider those details as significant. Looking at the data in Fig. 2, one might say that four periods are visible in 400 ky. However, the third period is about 30% shorter than the first. Therefore, the distributed low frequency emphasis of

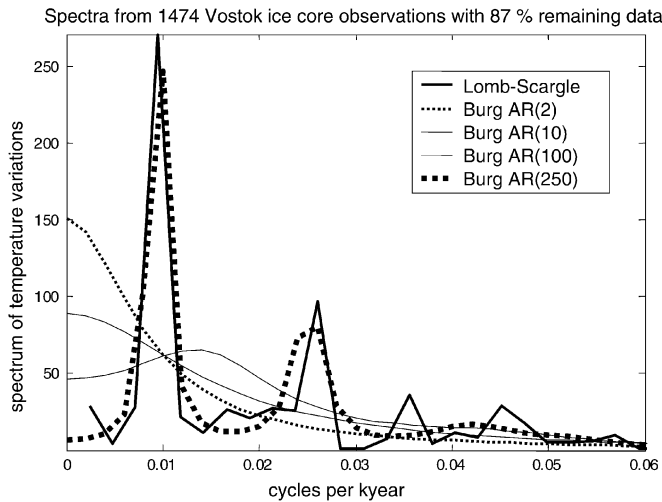


Fig. 5. Spectra of paleoclimatic Vostok data. Burgs on segments spectra for several orders are shown. The AR(2) model was selected for Burg. The AR(250) model looks like Lomb–Scargle, but has many insignificant parameters.

AR spectra looks more reliable for those data than the strong peaks at 0.01 and 0.025 ky in Lomb–Scargle.

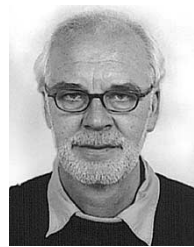
VI. CONCLUDING REMARKS

The finite interval likelihood maximization algorithm ARFIL is numerically stable in estimating AR models from incomplete data. This approximation gives the same result as the exact likelihood method of Jones. ARFIL is computationally attractive if less than 5% of the data remain. For few missing data, the performance of ML methods is better than that of other known methods, including all methods that reconstruct the missing data before the spectral density is estimated. Moreover, ARFIL can also be used for very sparse data. The quality of the estimated model with a selected model order is very good in simulations where the true process is a low-order AR(p) process, often comparable to the Cramér–Rao lower bound for consecutive data.

REFERENCES

- [1] J. R. Petit, “Climate and atmospheric history of the past 420.00 years from the Vostok ice core, Antarctica,” *Nature*, vol. 399, pp. 429–436, 1999.
- [2] P. Yiou, E. Baert, and M. F. Loutre, “Spectral analysis of climate data,” *Surv. Geophys.*, vol. 17, pp. 619–663, 1996.
- [3] S. de Waele and P. M. T. Broersen, “Error measures for resampled irregular data,” *IEEE Trans. Instrum. Meas.*, vol. 49, no. 2, pp. 216–222, April 2000.
- [4] R. J. A. Little and D. B. Rubin, *Statistical Analysis With Missing Data*. New York: Wiley, 1987.
- [5] N. R. Lomb, “Least squares frequency analysis unequally spaced data,” *Astrophys. Space Sci.*, no. 39, pp. 447–462, 1976.
- [6] J. D. Scargle, “Statistical aspects of spectral analysis of unevenly spaced data,” *Astrophys. J.*, ser. Studies in astronomical time analysis II, no. 263, pp. 835–853, 1982.
- [7] R. Bos, S. de Waele, and P. M. T. Broersen, “Autoregressive spectral estimation by application of the Burg algorithm to irregularly sampled data,” *IEEE Trans. Instrum. Meas.*, vol. 51, pp. 1289–1294, Dec. 2002.

- [8] Y. Rosen and B. Porat, “Optimal ARMA parameter estimation based on the sample covariances for data with missing observations,” *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 342–349, 1989.
- [9] S. de Waele and P. M. T. Broersen, “Reliable LDA-spectra by resampling and ARMA-modeling,” *IEEE Trans. Instrum. Meas.*, vol. 48, pp. 1117–1121, Dec. 1999.
- [10] P. M. T. Broersen, “Automatic spectral analysis with time series models,” *IEEE Trans. Instrum. Meas.*, vol. 51, pp. 211–216, Apr. 2002.
- [11] S. de Waele and P. M. T. Broersen, “The Burg algorithm for segments,” *IEEE Trans. Signal Processing*, vol. 48, pp. 2876–2880, Oct. 2000.
- [12] R. H. Jones, “Maximum likelihood fitting of ARMA models to time series with missing observations,” *Technometrics*, vol. 22, no. 3, pp. 389–395, 1980.
- [13] P. M. T. Broersen, S. de Waele, and R. Bos, “Estimation of autoregressive spectra with randomly missing data,” in *Proc. IMTC Conf.*, Vail, CO, 2003, pp. 1154–1159.
- [14] M. B. Priestley, *Spectral Analysis and Time Series*. London, U.K.: Academic, 1981.
- [15] S. M. Kay and S. L. Marple, “Spectrum analysis—a modern perspective,” *Proc. IEEE*, vol. 69, pp. 1380–1419, 1981.
- [16] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
- [17] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*. New York: Springer-Verlag, 1991.



Piet M. T. Broersen was born in Zijdewind, the Netherlands, in 1944. He received the M.Sc. degree in applied physics and the Ph.D. degree from the Delft University of Technology, Delft, the Netherlands, in 1968 and 1976, respectively.

He is currently with the Department of Multi Scale Physics of the Delft University of Technology. His main research interest is in automatic identification on statistical grounds by letting measured data speak for themselves. He developed a practical solution for the spectral and the autocorrelation analysis of stochastic data by the automatic selection of a suitable order and type for a time series model of the data.



Stijn de Waele was born in Eindhoven, the Netherlands, in 1973. He received the M.Sc. degree in applied physics and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 1998 and 2003, respectively.

His thesis was entitled “Automatic model inference from finite time observations of stationary stochastic processes.” Currently, he is a Senior Scientist with Philips Research Laboratories, Eindhoven, the Netherlands, where he works in the field of digital video compression.



Robert Bos was born in Papendrecht, the Netherlands, in 1977. He received the M.Sc. degree in applied physics from the Delft University of Technology, Delft, the Netherlands, in 2001, where he currently is pursuing the Ph.D. degree.

He is with the Delft Center for Systems and Control of the Delft University of Technology. His research is aimed at the development and the application of state estimation techniques for high-order first principle models of complex processes in the process industry.