IEEE Instrumentation and Measurement
Technology Conference
Anchorage, AK, USA, 21-23 May 2002

Generating Data with Prescribed Spectral Density

P.M.T. Broersen and S. de Waele

Department of Applied Physics
Delft University of Technology
P.O.Box 5046, 2600 GA Delft, The Netherlands
phone +31 15 2786419, fax +31 15 2784263, email broersen @ tn.tudelft.nl

Abstract: ***Time series models are suitable for the generation of data with prescribed covariance or spectral characteristics. If the required covariance function or spectral density is defined by a time series model, data generation is straightforward. An arbitrary prescribed spectral density will be approximated by a finite number of equidistant samples in the frequency domain. This approximation becomes accurate by taking more and more samples. Those samples can be inversely Fourier transformed into a covariance function of finite length. The covariance in turn is used to compute a long autoregressive (AR) process with the Yule-Walker relations. Data can be generated with this long AR process. Unfortunately, the most general prescribed spectral densities belong to infinitely wide covariance functions. Therefore, finite covariance representations are necessarily approximations. It is possible to derive objective rules to choose a minimal finite order for the generating AR process. This order depends on the number of observations to be generated. The criterion is that the spectrum of those observations cannot be distinguished from the prescribed spectrum.***

Keywords: Spectral analysis, time series models, order selection, ARMA process, linear filtering

1. INTRODUCTION

Measurement noise is present in the observations of many experiments. It will certainly be a problem in sensitive satellite measurement data [1]. Often, a good characterization of the noise spectrum is known from previous experiments or from a detailed physical description of the sensor and its environment. In such circumstances, it is possible to test the signal processing sequence in advance by using a noise realization of a stochastic process that has the known spectral density. The purpose may be to verify that the proposed processing allows for the desired accuracy. One specific application is the new gravity field and ocean circulation explorer mission, to be launched in 2006 [1]. This mission aims at the development of an improved model of the earth's gravity field. The presence of colored observation noise in a huge number of observations leads to a difficult numerical regression problem, demanding a weighted least squares solution. The weighting matrix is the inverse of the covariance matrix of the noise, which is extremely large here. The spectrum of the colored noise is concentrated in the low frequency part. A time series description of the known and given noise spectral density gives the possibility to generate noise realizations. The same time series model can be used to design an inverse filter for the colored noise; for real time implementation it is important that the filter has a low order. The filtered noise becomes uncorrelated, which is an important advantage in the remaining regression problem [1]. Other examples for the generation of data emerge in turbulence, where the spectrum is proportional to $f^{-5/3}$ and in other physical problems where the $1/f$ noise plays a role.

Time series modeling is a parametric description of spectral densities. The three model types that can be used for time series models are autoregressive (AR), moving average (MA) and combined ARMA models. All stationary stochastic processes can be characterized by AR or by MA models of infinite order [2]. AR models are more suitable for spectral peaks; MA models are better for valleys. In practice, most processes can be described adequately by AR(p), MA(q) or ARMA(p,q) processes of finite orders p and/or q [3].

Generating stationary data for some given ARMA process can be troublesome. Using zeros or any arbitrary values as initial conditions, generated signals become stationary after the length of the impulse response. Unfortunately, that length is only finite for a MA process. It is infinitely long for AR and ARMA processes. Therefore, this primitive method of data generation without care for initial conditions will only be exact for MA processes. It is at best an approximation for AR or ARMA processes. A better method is found by separating the generation into an AR and a MA part. Consider the joint probability density function of a finite number of AR observations which describes the correlation. Data can be generated that obey that prescribed AR correlation. A realization of the ARMA process is obtained by filtering AR data with the MA part.

So far, it has been assumed that the requirements are already formulated as a time series model. However, prescriptions for data can also be given in terms of correlation functions or power spectral densities. The treatment of prescribed spectra also uses time series models. A solution has been given to fit an AR, MA or an ARMA model to a finite number of sampled spectrum values [4]. It uses the inverse Fourier transform of the spectrum as the desired covariance function. An AR model of high order is calculated then from the covariances as a basis for other time series models. As AR algorithms are linear and hence much simpler than the non-linear MA computations, long AR models are preferred. Models can often be simplified by using the long AR model as input for a reduced statistics analysis [4].

This paper describes the joint probability density function of an arbitrary number of observations of an AR(p) process. That is the basis for the generation of data. Prescribed

0-7803-7218-2/02/$10.00 ©2002 IEEE

spectral densities can be based on estimated or on exact knowledge. It is shown that order selection criteria can be adapted to the reliability or the accuracy of the prescribed spectra. Furthermore, it is shown that the generating process for the prescribed spectral density can depend on the number of observations that will be generated.

2. ARMA MODELS

An ARMA(p,q) process is defined as [2]:

$$x_n + a_1 x_{n-1} + \cdots a_p x_{n-p} = \varepsilon_n + b_1 \varepsilon_{n-1} + \cdots + b_q \varepsilon_{n-q} \tag{1}$$

where ε_n represents a series of independent, identically distributed, zero mean white noise observations. The process is AR for q=0 and MA for p=0. The ARMA process can also be written with polynomials of AR and of MA parameters as:

$$A(z)x_n = B(z)\varepsilon_n\,,\; z^{-1}x_n = x_{n-1} \tag{2}$$

with $A(z)=1+a_1z^{-1}+\ldots+ a_pz^{-p}$ and $B(z)= 1+b_1z^{-1}+\ldots+ b_qz^{-q}$. Models may have estimated polynomials of arbitrary orders, not necessarily equal to the true p and q. Processes and models are stationary if the estimated roots of $A(z)$ are inside the unit circle and invertible if the zeros, the roots of $B(z)$ are inside. The spectral density is given by [2]:

$$h(\omega) = \frac{\sigma_\varepsilon^2}{2\pi}\frac{\left|B(e^{j\omega})\right|^2}{\left|A(e^{j\omega})\right|^2} \tag{3}$$

This shows that the parameters of a time series model, together with the variance of the exciting white noise, determine the spectral density. The covariance function is the inverse integral Fourier transform of (3). Computer calculations replace this integral by a summation and will generally only give an approximation of the true covariance function, of finite length. That approximation can be made as close as desired by taking more and more samples of (3) in the summation, but this Fourier relation between spectrum and covariance remains an approximation. However, it is also possible to obtain an exact covariance function for a given ARMA(p,q) process by direct computations in the time domain [2]. Therefore, the parameters of a time series model are a good representation for the characteristics of a process, exact in both the time and in the frequency domain.

The quality of estimated ARMA(p',q') models is measured with the model error ME [5]. This measure can be used in simulations where an omniscient experimenter knows the true process ARMA(p,q) parameters that generated the observations. Likewise, it can be used to evaluate the difference between a prescribed spectral density, expressed as an ARMA process and the approximating ARMA spectrum. The ME is a scaled transformation of the expectation of the squared error of prediction PE:

$$ME\left(\frac{\hat{B}_{q'}(z)}{\hat{A}_{p'}(z)}, \frac{B(z)}{A(z)}\right) = N\left(PE/\sigma_\varepsilon^2 - 1\right). \tag{4}$$

The model error is asymptotically equal to N times the spectral distortion SD, defined as

$$SD = \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\ln\hat{h}(\omega) - \ln h(\omega)\right)^2 d\omega\,, \tag{5}$$

where the ^ denotes an approximating spectrum. In this way, the ME is also defined for spectral densities that are not given by time series models. The expectation of ME is for <u>*unbiased*</u> models independent of N and of the variance of the signal. Only true and estimated parameter values are required to compute the ME with (4). The expectation of the ME of all unbiased ARMA(p',q') models that contain at least all truly non-zero parameters is equal to the number $p'+q'$ of estimated parameters.

3. PROBABILITY DENSITY OF AR PROCESS

The normal distribution of a variable x with mean μ and variance σ^2 is given by:

$$\mathbb{N}(x,\mu,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{6}$$

The joint probability density function (pdf) of N observations $\boldsymbol{X}$, where $\boldsymbol{X}$ is a jointly normally distributed vector stochastic variable

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_N \end{pmatrix} \tag{7}$$

with

$$E\{X\} = \mu_x \tag{8}$$
$$E\left\{(X-\mu_x)(X-\mu_x)^T\right\} = R_{xx}$$

is given by

$$f(X) = \frac{1}{2\pi^{N/2}\left|R_{xx}\right|^{1/2}}exp\left[-\frac{1}{2}\left\{(X-\mu_x)^T R_{xx}^{-1}(X-\mu_x)^T\right\}\right] \tag{9}$$

This is a general expression that can be applied to ARMA processes by expressing the Toeplitz matrix R_{xx} in the parameters. It is often better and it has computational advantages for ARMA processes to use as much as possible the white noise innovations with diagonal covariance function in the probability density [2]. The probability density function of N observations of an AR(p) process with polynomial $A_p(z)$, with normally distributed zero mean innovations ε_n can also be written as a conditional product of the last N-p observations, given the first p:

$$f(x_1,x_2,\cdots,x_N) = f(x_{p+1},\cdots,x_N \mid x_1,\cdots,x_p)f(x_1,\cdots,x_p) \tag{10}$$

The first part of the right-hand side for the last N-p observations is given by [2, p. 347]

$$f(x_{p+1},\cdots,x_N \mid x_1,\cdots,x_p)=\left(\frac{1}{2\pi\sigma_\varepsilon^2}\right)^{\frac{N-p}{2}}\exp\left(\frac{1}{2\sigma_\varepsilon^2}\sum_{n=p+1}^{N}A_p(z)x_n\right) \tag{11}$$

The second part describing the first p observations is also known [2, p 350]. However, a recursive form of that expression is a better starting point for the generation of data. By using conditional densities, it follows that

$$f(x_1,x_2,\cdots,x_p)=f(x_p \mid x_1,\cdots,x_{p-1})f(x_1,\cdots,x_{p-1}). \tag{12}$$

Generally, for all observations q, with $q < p$,

$$f(x_1,x_2,\cdots,x_q)=f(x_q \mid x_1,\cdots,x_{q-1})f(x_1,\cdots,x_{q-1}). \tag{13}$$

With those intermediate results for q, (12) can be written as

$$f(x_1,x_2,\cdots,x_p)=\left(\prod_{q=2}^{p}f(x_q \mid x_1,\cdots,x_{q-1})\right)f(x_1). \tag{14}$$

Elements of the Levinson-Durbin algorithm [6] can be used to evaluate this expression. This algorithm recursively computes parameter sets of increasing order $q < p$ from reflection coefficients, as well as variance expressions. The last parameter of an AR model of order q is always equal to the reflection coefficient k_q. The algorithm starts with:

$$\begin{aligned}&a_1^1=-R(1)/R(0)=k_1\\&\sigma_1^2=R(0)\left(1-k_1^2\right)\\&A^1(z)=1+a_1^1z^{-1}\end{aligned} \tag{15}$$

where $R(q)$ is the covariance of the process at lag q. The recursion for q=2,3,...,p is given by

$$\begin{aligned}&a_q^q=-\left[R(q)+\sum_{i=1}^{q-1}a_i^{q-1}R(q-i)\right]/\sigma_{q-1}^2\\&k_q=a_q^q\\&a_i^q=a_i^{q-1}+k_qa_{q-i}^{q-1},\ 1\le i<q\\&\sigma_q^2=\sigma_{q-1}^2\left(1-k_q^2\right)\\&A^q(z)=1+a_1^qz^{-1}+\cdots+a_q^qz^{-q}\end{aligned} \tag{16}$$

At the final stage p of the recursion, the polynomial $A^p(z)$ is equal to $A(z)$ in (1). The polynomial $1\text{-}A^{q-1}(z)$ is the best linear predictor of order $q\text{-}1$, or the best linear combination of $q\text{-}1$ previous observations to predict the next observation x_q. Hence, the conditional probability density of x_q, conditional on $q\text{-}1$ previous observations has as expectation $[1\text{-}A^{q-1}(z)]x_q$ with the variance σ_{q-1}^2. Using this in conditional expectations gives with (6)

$$f(x_q \mid x_1,\cdots,x_{q-1})=\mathbb{N}\left(x_q,\left(1-A_{q-1}(z)\right)x_q,\sigma_{q-1}^2\right). \tag{17}$$

The recursive variance relation can be expressed with increasing or decreasing index:

$$\sigma_x^2=R(0)=\frac{\sigma_\varepsilon^2}{\prod_{i=1}^{p}\left(1-k_i^2\right)} \tag{18}$$

which gives

$$\sigma_q^2=\frac{\sigma_\varepsilon^2}{\prod_{i=q+1}^{p}\left(1-k_i^2\right)}=\sigma_x^2\prod_{i=1}^{q}\left(1-k_i^2\right) \tag{19}$$

and (17) becomes with (6) (20)

$$f(x_q \mid x_1,\cdots,x_{q-1})=\left(\frac{\prod_{i=q}^{p}\left(1-k_i^2\right)}{2\pi\sigma_\varepsilon^2}\right)^{\frac{1}{2}}\exp\left(-\frac{\prod_{i=q}^{p}\left(1-k_i^2\right)\left(A_{q-1}(z)x_q\right)^2}{2\sigma_\varepsilon^2}\right).$$

The substitution of (20) in (14) and that result together with (11) in (10) is merely an exercise with as result sums of products. However, as all ingredients for the probability density function $f(x_1, x_2,..., x_N)$ are given, the derivation is sufficient to be used in generating data for an AR(p) process.

4. FROM SPECTRUM TO ARMA MODEL

The prescribed spectral density will at the end be defined by the parameters of an ARMA model. If they are already given as spectral characterization, they can be used immediately.

The prescribed spectrum can also be given as an exact continuous function $h(\omega)$, $0 \le \omega \le \pi/T$. Without loss of generality, the sampling time T is taken to be 1. The first step is to find an approximation for the covariance function. The inverse integral Fourier transform of the continuous function $h(\omega)$ is the exact covariance function $R(k)$:

$$R(k)=\int_{-\pi}^{\pi}h(\omega)e^{j\omega k}d\omega. \tag{21}$$

with possibly infinite length. By defining $h(\omega)$ periodic with period 2π, the integral in (21) can also be taken from 0 to 2π. As ω in (21) is continuous, the usual discrete time inverse Fourier computer transformations can only be approximations for $R(k)$, unless specific assumptions can be made about the interpolated values between the discrete frequency values $h(\omega_k)$ that are used in the computation. However, the covariance function of an AR or an ARMA process is infinitely wide. Approximating the integral of $h(\omega)$ in (21) by a summation is equivalent with sampling in the frequency domain. This causes the equivalent of aliasing of the covariance function, in the time domain. Therefore, some care in the covariance calculation with inverse Fourier transform is required. Taking the mid range value of the integrand $h(\omega)$ as representative for an integration interval $\Delta\omega$, a correlation function $R(k)$ can be approximated from discrete values $h(m\Delta\omega)$ as

$$\begin{aligned}R(k)&=\int_0^{2\pi}h(\omega)e^{j\omega k}d\omega=\int_{-\Delta\omega/2}^{2\pi-\Delta\omega/2}h(\omega)e^{j\omega k}d\omega\\&\approx\sum\nolimits_m h(m\Delta\omega)\int_{(m-0.5)\Delta\omega}^{(m+0.5)\Delta\omega}e^{j\omega k}d\omega\\&=\frac{2}{k\Delta\omega}\sin\left(\frac{k\Delta\omega}{2}\right)\sum\nolimits_m h(m\Delta\omega)e^{j\Delta\omega km}\Delta\omega\ .\end{aligned} \tag{22}$$

For small $\Delta\omega$, this summation is almost equal to the usual summation result in (22) because $sin(y)/y$ has the limiting value 1 for small y. The derivation of (22) for $R(k)$ is preferred above the usual inverse FFT if only a few samples of a given $h(\omega)$ are used in the inverse transformation. For one sided spectra defined for ω between 0 and π, the first step is to sample $h(\omega)$ at intervals $\Delta\omega$, with $\Delta\omega=\pi/L$. The L samples $h(m\Delta\omega)$ are made to a symmetric spectrum by adding the L-2 samples for m=2, ..., L-1 in reversed order. Taking $L=2^k+1$ gives a convenient length for the FFT. After the inverse Fourier transform of the elongated sampled spectrum, the first L, ($L=\pi/\Delta\omega$) points describe the desired covariance function. L should be chosen high enough, such that the values of the covariance are effectively zero for lags beyond L. If that turns out to be impossible for a prescribed spectrum because the covariance function is to elongated, L should be chosen at least as high as the number of observations N that has to be generated. The larger L, the better the estimated covariance. This is an operational advice for the length L and it should always be validated that the correlation is negligible beyond L, or that $L > N$. The first K lags of the covariance function are transformed to an AR(K-1) model with the Yule-Walker relations [6]. If the prescribed $h(\omega)$ is exact, taking L and K greater will generally give a better approximation; the improvement may be small.

If the true process has a finite AR order p, no serious problems arise as long as the correlation length L is taken much greater than p. The true AR order of the prescribed spectrum can be infinite, however, and it will be infinite if the true spectrum would be MA or ARMA or any arbitrary function of ω. The question is if some value M of the AR order less than K can be considered high enough for an approximation of sufficient accuracy. It may be advantageous to have a low order process to generate data. It is still more important to design a low order filter to undo the coloring of some given spectral density to whiten the noise [1]. Any distortion of the true $h(\omega)$ will finally become noticeable if more and more data are generated. Therefore, the minimally required order M depends on the number of observations N that has to be generated. A natural choice is to allow distortions that are much smaller than the expected statistical uncertainty if the generated data are analyzed. The distortion can be quantified as the bias contribution of all omitted true AR reflection coefficients to the ME (4), or to the spectral distortion (5). If the bias contribution of the omitted reflection coefficients is smaller than the variance contribution of estimating a single parameter, the higher order parameters can be omitted without the possibility that the omission can be detected from N generated observations. This is no tight boundary but it is a sensible compromise. This gives an allowed ME of 1 for the truncated AR(M) model with respect to the true AR(∞) process for a given N. With (4) this gives:

$$ME\left(\frac{1}{A_M(z)},\frac{1}{A_\infty(z)}\right)=1\,.$$

This is the smallest M for which AR(M) model gives:

$$\sigma_M^2=\frac{\sigma_\varepsilon^2}{\prod_{i=M+1}^{\infty}\left(1-k_i^2\right)}=\sigma_x^2\prod_{i=1}^{M}\left(1-k_i^2\right)<\sigma_\varepsilon^2\left(1+\frac{1}{N}\right) \qquad (23)$$

In practice, the AR(∞) process in (23) is replaced by the finite order AR(K) process, with K chosen at least so high that the difference between the AR(K) and the AR($K/2$) model has a ME difference smaller than say 0.1 for the given N. This can be verified in practice and K can be chosen higher until this condition is met. Simulations will indicate that orders K and M are often much smaller than N. It is possible to use the AR(K) model as input for a reduced statistics algorithm [4] that computes AR, MA and ARMA models of various orders. Moreover, it will select the model type and the model order that are closest to the given AR(K) process. Furthermore, the lowest order MA(q) and the lowest order ARMA(r,r-1) model can be found that have a ME value less than 1 with respect to the AR(K) model, for given N. If the true process $h(\omega)$ would be a finite order MA(q) or ARMA(p,q) process, that would be detected with the reduced statistics estimator. If this MA or ARMA model require less parameters than the AR(M) model described before, it may be worthwhile to use the most parsimonious time series model to generate or to filter data. However, it should be realized that in inverse filtering the MA and ARMA models have long transients.

It is also possible to define a prescribed spectrum with an estimated periodogram, instead of with a continuous function $h(\omega)$. In those cases, special care is required, because estimated periodograms contain a lot of spurious details. If N_0 observations are transformed with the FFT and the absolute values are squared, the resulting function is the raw periodogram, say $P(\omega)$. The inverse Fourier transform of $P(\omega)$ should be the covariance function, but it is an aliased version which has as expectation [2]:

$$E\left[\hat{R}(s)\right]=\frac{N_0-s}{N_0}R(s)+\frac{s}{N_0}R(N_0-s)\,. \qquad (24)$$

with $R(s)=E\{x_n x_{n+s}\}$. Using tapers and windows will cause more distortions of the covariance. The use of periodograms cannot be advised. Periodograms are of the same accuracy as spectral prescription as some very high order estimated AR model. However, a much better solution exists. If one wants to use measured data as prototype, it is advised to estimate the time series model for those data with ARMAsel [7,8]. This automatic computer program selects the best spectral time series model for given data with robust algorithms and criteria, which select only statistically significant details.

Another possibility is that the information in $h(\omega)$ is not exact, at least not for all frequencies. For colored satellite noise, the spectral density at some discrete, not equidistant, frequencies has been given and the continuous function $h(\omega)$ is determined with interpolation [1]. The proposed method to deal with those circumstances is largely the same as dealing with exact $h(\omega)$, but with a completely different critical value for ME. Variation of the interpolation methods and spline solutions produces different continuous spectra, different

covariances and hence different time series models. The difference between those interpolation solutions can be expressed in ME (4) for a given N as number of data to be generated. If no information about the best interpolation is available, all solutions are equally well obeying the prescriptions. The largest difference found in ME between the different interpolation results can be used as a critical ME value, instead of the critical value ME=1 that was derived before for a given exact true $h(\omega)$.

The criterion for data generation is always: generate N observations with a time series spectrum that cannot be distinguished from the true given spectrum. It might be possible that the model is good for the generation of N observations, but the difference can perhaps be seen if the same model was used to generate $100N$ or more observations. Therefore, the generating model may depend on N.

5. DATA GENERATION

Data generation is strictly separated in MA and AR generation. For ARMA processes, it is essential that the AR part is used first, because the derivation of section 3 used white noise as input, e.g. in (11). The principle is for ARMA processes given by:

$$\begin{aligned} A(z)v_n &= \varepsilon_n, \\ x_n &= B(z)v_n. \end{aligned} \tag{25}$$

which together combines to the result of (2):

$$A(z)x_n = A(z)B(z)v_n = B(z)\varepsilon_n. \tag{26}$$

AR data

The purpose is to generate N observations with the probability density function $f(v_1, v_2, \ldots, v_N)$ like in (10). The first observation v_1 has the only requirement that it has expectation zero and variance $R(0)$ or σ_v^2. That is found with a random number generator with normal or Gaussian density and the prescribed variance. The second observation v_2 follows from (15) and (17) as a normally distributed random variable with mean $-a_1^1 v_1$ and variance σ_1^2. The third observation v_3 uses (16) and (17) with mean $-a_1^2 v_1 - a_2^2 v_2$ and variance σ_2^2. The first p observations are generated in this way. According to (11), all further observations can be generated with the regime:

$$v_n + a_1 v_{n-1} + \cdots + a_p v_{n-p} = \varepsilon_n, \quad n = p+1, \cdots, N \tag{26}$$

This is a filter procedure with a Gaussian random white noise signal as input signal and the first p observations as initial conditions. For AR processes, the p initial observations and the N-p filter results with (26) are together the N observations.

MA or ARMA data

For MA data, the input signal v_n is a Gaussian white noise ε_n, for ARMA processes the input to the MA filter will be the output v_n of the AR filter. The data x_n are computed with:

$$x_n = v_n + b_1 v_{n-1} + \cdots + b_q v_{n-q}. \tag{27}$$

The first q data require negative input index. Therefore, to generate N MA or ARMA observations with this method, the input sequence has to be $N+q$ long. The first q points of the filter output are disregarded.

6. SIMULATIONS

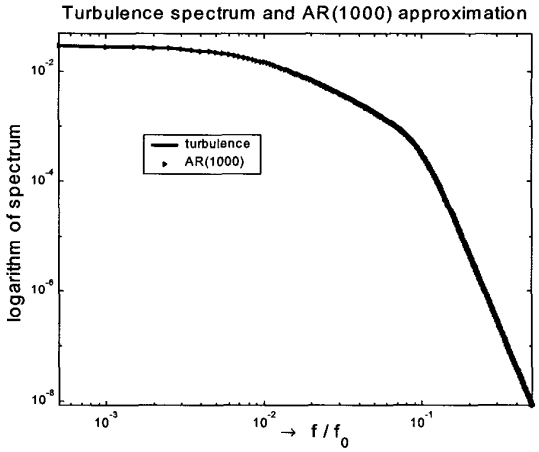

Fig.1 Turbulence prototype spectrum and approximating AR(1000) model for L=1001.

To show the possibilities of the data generation, it is applied to an example where broken exponents determine the desired spectral shape. It is some prototype spectrum for turbulence data. The spectrum consists of 2 declining slopes after a low frequency constant. The first slope descends at a rate of $\sim f^{5/3}$ from $0.01 f_0$ and the second slope of $\sim f^7$ starts at $0.1 f_0$.

$$h(\omega) = \frac{1}{1+\left(\frac{\omega}{0.02\pi}\right)^{\frac{5}{3}}} * \frac{1}{1+\left(\frac{\omega}{0.2\pi}\right)^{\frac{16}{3}}} \tag{25}$$

Fig.1 shows no visible difference between the true $h(\omega)$ and the AR(1000) approximation, obtained with L=1001. The ME of the AR(500) truncated model is $4.10^{-9}N$, which is less than 0.1 for $N < 2.5*10^7$. According to the rules developed in section 4, data can be generated with sufficient accuracy if the ME of the truncated model is less than 1. Table 1 presents those orders. It is clear that the data generation can be carried out with a very high accuracy. Taking much higher values for L, the length of the Fourier transform, has no influence on the minimum order M. The values for M obtained with $L=2^{20}$ are equal to those in Table 1. This demonstrates that this covariance function is already damped out at lag 1000.

Table1 Lowest allowed truncated AR order M with ME less than 1 for the generation of N observations, as a function of N for two prescribed spectral densities; L=1001, K=1000.

$\Rightarrow$ N	10^2	10^3	10^4	10^5	10^6	10^7
turbulence	7	11	19	37	73	155
1/f	8	38	113	319	883	977

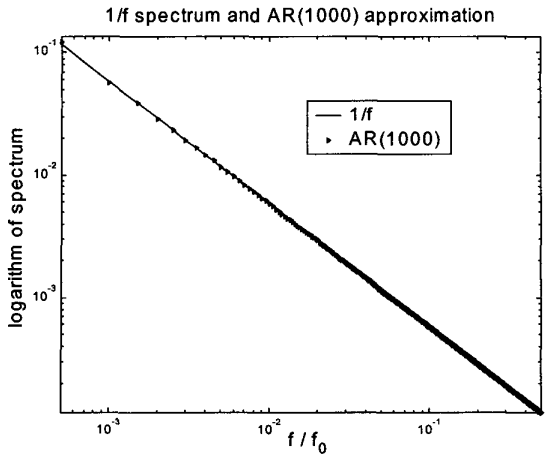

Fig.2 1/f prototype spectrum and approximating AR(1000) model for L=1001.

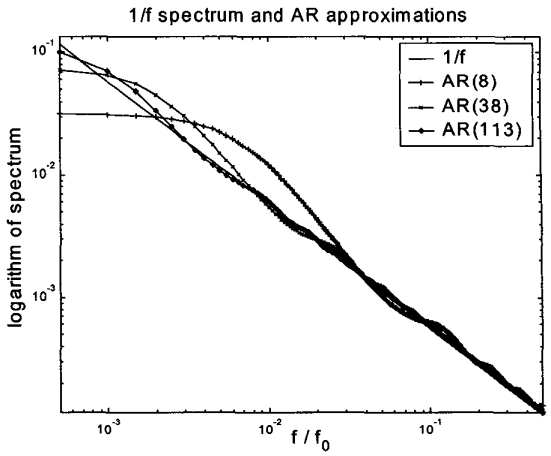

Fig.3 1/f prototype spectrum and lower order AR approximations, obtained with L=1001, K=1000.

An interesting spectral density in physics is the $1/f$ spectrum in Fig.2. Due to the singularity at f=0, it is difficult or even impossible to recognize or to verify this shape from data at frequencies lower than say $1/N_0$, if N_0 is the number of observations. This is not visible in Fig.2 because $f<1/N_0$ will be beyond the displayed frequency range. In approximating this true $1/f$ spectrum, a choice has to be made for f=0, because the true value of the spectrum becomes infinite and leads to numerical problems. Therefore, some extrapolation from the first two non-zero sampled values for $h(\omega)$ will be used. Those are at the frequencies $f_0/2L$ and f_0/L. A parabola through the spectrum of those 2 points, with the additional demand that the derivative of the spectrum at f=0 equals 0 yields: $h(0)=7/6h(f_0/2L)$. The ME of the AR(500) truncated model is $9.10^{-6}N$ for L=1001, which is less than 0.1 for $N < 1.1*10^4$. According to the rules, a higher value for L is required if N is greater than 10000. This is also clear in Table 1 because the required AR order becomes too close to L for the $1/f$ spectrum. Moreover, it follows already from a visual inspection of the covariance function that it is not damped out for L=1001. Nevertheless, the accuracy of the AR models of order M from Table 1 are shown in Fig.3. Using this poor design, the AR(1000) model of Fig.2 is still very good in the displayed frequency range and the low order models in Fig.3 are still reasonable for frequencies higher than $0.03f_0$. A much higher length L should be chosen. For $L=2^{20}$, the computed minimum order M for N=1000 is 146 for K=10000 and 147 for K=50000. This demonstrates that it is possible to find a satisfactory time series model for any N. However, in cases with a singularity at f=0, the required minimum model order M is rather high.

The first example shows that smooth spectral shapes can be modeled easily with time series models. The second example deals with the problem that the estimated covariance will always depend on the arbitrary length L. That gives the sampling distance in the frequency domain and therefore the lowest non-zero frequency which is taken into account.

7. CONCLUSIONS

Data with prescribed spectral density can be generated in a simple way with time series models. The first step is to determine a finite order AR model that has a spectrum close enough to the prescribed spectrum. Several tools are available to facilitate the search for a parsimonious time series model. Use of higher orders only improves the accuracy.

By using an exact description of the probability density function of N autoregressive observations, it is possible to efficiently generate autoregressive data. The generation of MA or ARMA data is solved with a simple filter operation.

References

[1] R. Klees and P.M.T. Broersen, "How to handle colored observation noise in large-scale least-squares problems? Part II, Building the optimal filter." Submitted to *Journal of Geodesy*, 2002.

[2] M.B. Priestley. *Spectral Analysis and Time Series*, London, Ac. Press, 1981.

[3] P.M.T. Broersen, "Automatic spectral analysis with time series models". *IEEE Trans. on Instrum. and Meas.*, vol. 51, no. 2, April, 2002.

[4] P.M.T. Broersen and S. de Waele, "Selection of order and type of time series models from reduced statistics", *IMTC 2002 Conference*, Anchorage, 2002.

[5] P.M.T. Broersen, "The quality of models for ARMA processes", *IEEE Trans. on Signal Process.*, vol. 46, pp. 1749-1752, June, 1998.

[6] S.M. Kay and S.L. Marple, "Spectrum analysis-a modern perspective". *Proc. IEEE*, vol. 69, pp. 1380-1419, 1981.

[7] P.M.T. Broersen, "Facts and fiction in spectral analysis", *IEEE Trans. on Instrum. and Meas.*, vol. 49, pp. 766-772, August, 2000.

[8] P.M.T. Broersen, *ARMASA Toolbox*, freely available from http://www.tn.tudelft.nl/mmr/downloads.