On Filtering Time Series with Monotonically Time Varying Frequencies

Krista B. Cohlmia[*], Henry L. Gray[†]
Wayne A. Woodward[‡] and Huiping Jiang[§]

November 9, 2005

Abstract

In this paper we introduce a new approach for filtering data whose periodic structure changes approximately monotonically in time. The paper focuses primarily on the linear case, i.e. the case in which periods change approximately like $at + b$.

The linear case depends on a class of processes referred to as Euler processes, introduced by Gray and Zhang (1988) and more recently extended by Gray, Vijverberg, and Woodward (2004). Filtering in the more general case is based on the $G(\lambda)$ processes introduced by Jiang, Gray and Woodward (2004), i.e. for data whose frequencies change asymptotically in time like αt^{β}. Some simulated and real data examples are given.

[*]Odessa College, Odessa, Texas
[†]Southern Methodist University, Dallas, Texas
[‡]Southern Methodist University, Dallas, Texas
[§]Columbia University, New York, New York

1. Introduction

Gray and Zhang (1988) introduced continuous M-stationary processes for the purpose of analyzing long memory data. Their work has been recently extended by Vijverberg (2002) and Gray, Vijverberg, and Woodward (2004). In those papers it is shown that M-stationary processes are very good models for data whose periodic structure changes approximately linearly or equivalently, processes whose frequencies change like $(at + b)^{-1}$. Examples of such data include bat echolocation signals, whale calls, and many other physical signals. Several approaches for analyzing data with time varying frequencies such as wavelets, short term Fourier transforms (SFT), Wigner distributions (WD), etc., have been proposed in the literature. In this paper we show that the method of *time deformation* when applied to M-stationary processes and their generalization, $G(\lambda)$ processes, is particularly useful for filtering applications when the realization has a monotonically changing frequency structure.

In Section 2 we briefly review some of the properties of M-stationary processes and their *dual processes*. In Section 3 we describe techniques for filtering M-stationary processes and demonstrate the procedure using examples. Although our focus is on M-stationary processes we also show how the method can be extended to processes whose frequencies change like αt^{β}. In Section 4 we briefly discuss $G(\lambda)$ processes and provide examples to illustrate filtering in this more general setting. Software for M-stationary and $G(\lambda)$ modeling is available at http://faculty.smu.edu/hgray/research.htm.

2. The Discrete M-Stationary Process

Definition 1: A continuous stochastic process, $\{X(t); t\epsilon(0,\infty)\}$ is said to be multiplicative stationary (M-stationary) if the following hold for every $\tau \in (0,\infty)$:

i) $\mathrm{E}[X(t)] = \mu$

ii) $\mathrm{Var}(X(t)) = \sigma^2 < \infty$

iii) $\mathrm{E}[(X(t) - \mu)(X(t\tau) - \mu)] = R_X(\tau)$

Examples of continuous processes that are M-stationary include

$$X(t) = A\cos(\beta \ln t + \phi), \tag{1}$$

where $t \in (0,\infty)$ and $\phi \sim U(0, 2\pi)$ and the pth order Euler process

$$t^p X^{(p)}(t) + \phi_1 t^{p-1} X^{(p-1)}(t) + ... + \phi_p(X(t) - \mu) = \epsilon(t) \tag{2}$$

where $\epsilon(t)$ is M-white noise (see Gray and Zhang, 1988).

Definition 2: Let $X(t)$ be an M-stationary process. Then the process $\{Y(u) : -\infty < u < \infty\}$, where $Y(u) = X(t)$, with $t = e^u$ is called the dual of $X(t)$.

It is easily shown that $X(t)$ is M-stationary if and only if $Y(u)$ is stationary, and that the dual of the Euler process in (2) is a continuous pth order autoregressive (AR) process,

$$Y^{(p)}(t) + \alpha_1 Y^{(p-1)}(t) + ... + \alpha_p(Y(t) - \mu) = a(t). \tag{3}$$

See Gray and Zhang (1988) .

Remark: The derivatives in (2) and (3) are considered formally as is the usual treatment for continuous AR processes. These statements can be made mathematically more rigorous by introducing $X(t)$ as a general linear process, but this approach is not as attractive from an instructive or physical perspective. See Priestley (1981).

Definition 3: Let $h > 1$ and $S = \{t : t = h^k, k = 0, \pm 1, \pm 2, ...\}$. Then the discrete stochastic process $\{X(t) : t \epsilon S\}$ is said to be a discrete M-stationary process if $X(t)$ satisfies Definition 1 for every $t \in S$ and $\tau \in S$.

Definition 4: The discrete process $Y_k = X(h^k)$, $k = 0, \pm 1, \pm 2, ...$ is called the dual of $X(h^k) = X(t)$.

If $C_Y(k)$ denotes the autocovariance of Y, then $C_Y(k) = R_X(h^k)$ and $\{X(t) : t \in S\}$ is M-stationary if and only if $\{Y_k; k = 0, \pm 1, \ldots\}$ is stationary. From this observation we see that $R_X(h^k) = R_X(h^{-k})$.

Examples of discrete M-stationary processes are

$$X(h^k) = \cos(\beta \ln h^k + \phi) \tag{4}$$

where $\phi \sim \text{Uniform}(0, 2\pi)$, $k = 0, \pm 1, \ldots$, and

$$(X(h^k) - \mu) - \phi_1(X(h^{k-1}) - \mu) \ldots - \phi_p(X(h^{k-p}) - \mu) = a(h^k), \tag{5}$$

$k = 0, \pm 1, \ldots$

The process in (5) is called a discrete pth order Euler process and its dual, Y_k, is the pth order discrete autoregressive process given by

$$(Y_k - \mu) - \phi_1(Y_{k-1} - \mu) - \ldots - \phi_p(Y_{k-p} - \mu) = Z_k, \tag{6}$$

where $Z_k = a(h^k)$. Henceforth, we will take $\mu = 0$. For more details see Gray, Vijverberg and Woodward (2004).

In Vijverberg (2002), it was shown that the limit, with probability one, as $h \to 1$ of a discrete Euler process is the continuous Euler process defined by (2). Additionally in Choi, Gray, and Woodward (2003), it was shown that if one samples a continuous Euler process at $t = h^k$, the resulting process is a mixed discrete Euler process. However if this process is invertible it can then be well represented by a discrete Euler process. In the remainder of this paper we will therefore consider only the discrete process, but it should be kept in mind that the discrete Euler process will be viewed as a sampled continuous process.

Definition 5: The discrete M-spectrum is defined as the discrete Mellin transform of $R_X(h^k)$, i.e.

$$\begin{aligned} G_X(f^*) &= \sum_{k=-\infty}^{\infty} h^{-2\pi i f^* k} R_X(h^k), \\ &= \sum_{k=-\infty}^{\infty} e^{-2\pi i f^* k \ln h} R_X(h^k) \end{aligned} \tag{7}$$

where $h > 1$, $|f^* \ln h| \leq \frac{1}{2}$, and f^* is referred to as M-frequency.

Since $C_Y(k) = R_X(h^k)$, it is clear from (7) that $G_X(f^*) = S_Y(f)$, where $f = f^* \ln h$ and $S_Y(f)$ is the usual spectrum of the dual process Y_k.

If $X(t)$ is a discrete Euler process, from (6) it follows that the M-spectrum is given by

$$\begin{aligned} G_X(f^*) &= \frac{\sigma_a^2}{|\phi(e^{-2\pi i f^* \ln h})|^2} \\ &= \frac{\sigma_a^2}{|1 - \phi_1^{-2\pi i f^* \ln h} - \ldots - \phi_p e^{-2\pi i p f^* \ln h}|^2}, \end{aligned} \tag{8}$$

and the M-spectral density is

$$g_X(f^*) = \frac{\sigma_a^2}{\sigma_X^2 |\phi(e^{-2\pi i f^* \ln h})|^2} \quad (9)$$

where $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ is the autoregressive operator.

In addition to the M-spectrum and M-frequency, f^*, for processes with time-varying frequencies we are most often interested in the instantaneous period, instantaneous frequency and instantaneous spectrum, i.e., the period, frequency and spectrum at any given t.

To this end Jiang, Gray, and Woodward (2004) and Gray, Vijverberg and Woodward (2004) introduce the instantaneous period, instantaneous frequency and instantaneous spectrum of a $G(\lambda)$ process and an M-stationary process respectively. Their definitions are such that the instantaneous period, frequency and spectrum are precisely the period, frequency, and spectrum at any given instant. In Jiang, Gray and Woodward (2004), it is shown that for M-stationary processes these definitions yield a more accurate representation of the period, frequency, and spectrum at a given time than do similar notions resulting from the common practice by the engineering community of defining the instantaneous frequency to be the derivative of the phase. Jiang, Gray, and Woodward (2004) then extend these notions to the instantaneous frequency and spectrum of $G(\lambda)$ processes. In that paper they show that such processes furnish a good model for data whose frequencies change asymptotically like αt^β, for $\alpha > 0$ and $-\infty < \beta < \infty$. We will return to this in Section 4.

Definition 6: A function, $g(t)$, is said to be M-periodic over I with multiplicative period (M-period), δ, if $\delta > 1$ is the minimum value of $\delta \in I$ such that $g(t) = g(t\delta)$ for all $t \in I$. The associated M frequency, f^*, is given by $f^* = (\ln \delta)^{-1}$ which gives $\delta = e^{1/f^*}$.

As an example consider $g(t) = \cos(\beta \ln t)$. In this case $\delta = e^{2\pi/\beta}$ since $g(t\delta) = \cos(\beta \ln(te^{2\pi/\beta})) = \cos(\beta \ln t + 2\pi) = g(t)$. Also $f^* = \beta/2\pi$.

From the definition, it follows that for each fixed t, an M-periodic function returns to the value $g(t)$ at the distance $t\delta - t = t(\delta - 1)$. Thus when viewed on the "regular time" scale, $g(t)$ has a linearly lengthening period. For a discrete Euler process, it can be shown that for each complex root of the characteristic equation of the dual process, the M-autocorrelation has a damped (undamped if the root is on the unit circle) periodic component that elongates linearly. A similar result can be shown for the continuous case.

Before formally defining the instantaneous frequency and spectrum of an M-stationary process, we need to introduce one more concept. It should be noted that in our application, the origin of the process, zero, will not usually be the origin of the observations. Therefore if t' denotes the actual time (from the origin), t_0 denotes the origin of the observations, and t denotes the distance from t_0 to t', then $t_0 + t = t'$. The observation origin, t_0, will be called the *offset*, t' will be called the *process index* and t will be called the *observation index*. In the discrete case we will typically observe data at equally-spaced time points. Thus, we observe values at $h^j + il,\ i = 1, \ldots, n$, where $h^j + nl = h^j h^n$. For a given sample multiple h, we will denote the *offset* by $\Delta = h^j - 1$ and the *observation Euler index* by h^k, where the observations are obtained by interpolation at h^{j+k}, $k = 1, \ldots, n$. In general the offset will not be known and must be estimated.

Note that in the discrete case, if we denote the M-period by $\delta = h^m$, and note that $f^* = (\ln \delta)^{-1} = (\ln h^m)^{-1}$. Then $h^m = e^{1/f^*}$ as in the continuous case. We now give the following definition first given by Gray, Vijverberg and Woodward (2004).

Definition 7. The instantaneous period, $\mathcal{P}$, and instantaneous frequency, f, at h^k with offset Δ of a discrete M-stationary process are defined as

$$\mathcal{P}(h^k; f^*, h^j) = h^j h^k (e^{1/f^*} - 1)$$

and

$$f(h^k; f^*, h^j) = [h^j h^k (e^{1/f^*} - 1)]^{-1}, \tag{10}$$

where h^j and h^k are as defined in the above.

From Definition 7 and (9) we can define the instantaneous spectrum given in Jiang, Gray and Woodward (2003) and Gray, Vijverberg and Woodward (2004). Definition 7 is extended to the continuous case by replacing h^{j+k} by $t_0 + t$. Solving for f^* in (10) in terms of f, h^j and h^k leads to the following definition.

Definition 8: The instantaneous spectrum, $S(f; h^k, h^j)$, of a discrete M-stationary process $\{X(h^k)\}$ is defined by

$$S(f; h^k, h^j) = G_X(f^*), \tag{11}$$

where

$$f^* = \left[\ln\left(\frac{1 + f h^j h^k}{f h^j h^k}\right)\right]^{-1}. \tag{12}$$

If the process is continuous the instantaneous spectrum is defined by

$$S(f; t; t_0) = G_X(f^*)$$

where

$$f^* = \left[\ln \left(\frac{1 + (t_0 + t)f}{(t_0 + t)f} \right) \right]^{-1}.$$

Remark: From (10) and (11), it should be clear that the instantaneous spectrum gives the spectrum at any given t. Thus it shows the contribution to the variance by each frequency at any given time.

Example 1. To demonstrate the previous results, consider a realization of length $n = 400$ generated from the Euler(4) model

$$\phi_1(B)\phi_2(B)X(h^k) = a(h^k) \tag{13}$$

where

$$\begin{aligned} \phi_1(B) &= 1 - 1.97B + .98B^2 \\ \phi_2(B) &= 1 - 1.7B + .99B^2, \end{aligned} \tag{14}$$

$\sigma_1^2 = 1$, $h = 1.005$ and the offset $h^j = 15$.

Using the author's GWS software, which can be downloaded from http://faculty.smu.edu/hgray/research.htm, an AR(18) and an Euler(11) were chosen as the best AR and Euler models respectively when $0 \leq p \leq 20$. For the Euler(11) model, $\hat{h} = 1.0082$ and $\hat{h}^j = 15$.

Remark: Throughout this paper the procedure will be to determine the "best" AR(p) fit to the dual, from which we obtain the corresponding Euler(p) as our M-stationary model. In later sections, we however generalize this to $G(\lambda)$-process which will be defined in Section 4.

The residuals of the Euler(11) passed a Ljung-Box test for white noise while the AR(18) did not. Figure 1 shows a plot of the data, the sample autocorrelations (ACF), as well as the sample spectrum and an AR(18) spectral estimate. From the data there appears to be some cyclic nature to the data, although it is not suggested from the sample ACF or either the sample spectrum or the AR(18) spectral estimate. This is not surprising since, from the data, it is clear that the frequencies are changing with time to a degree that stationary methods seem inappropriate.

Table 1 shows the factors of the Euler(11) model. That is, we show the irreducible first and second order factors of the Euler(11) model fit to the data. Additionally, the table shows $|r_i^{-1}|$ where the r_i is the root of the characteristic equation associated with the i-th factor. The closer $|r_i^{-1}|$ is to unity, the higher the power associated with the corresponding frequency, and therefore the spectrum will tend to be quite peaked at frequencies for which $|r_i^{-1}|$ is close to unity. One can see from the table that most of the power or variance is contained in the M-frequencies 17.42 and 3.223 which are associated with the factors

$$\begin{aligned} \hat{\phi}_1(B) &= 1 - 1.234B + .968B^2 \\ \hat{\phi}_2(B) &= 1 - 1.918B + .945B^2. \end{aligned} \tag{15}$$

It should be noted that the M-frequencies associated with the factors $\phi_1(B)$ and $\phi_2(B)$ in (14) are 17.44 and 3.19 respectively which are quite close to the estimates shown in Table 1. As a result (11) will yield an excellent estimate of the instantaneous spectrum.

Table 1. Factors of the Euler(11) Model fit to the data in Figure 1(a)

Absolute Reciprocal	M-Frequency	Dual Frequency	Factors
0.984	17.42	0.142	1 - 1.234B + 0.968B^2
0.972	3.22	0.026	1 - 1.918B + 0.945B^2
0.802	48.07	0.392	1 + 1.250B + 0.643B^2
0.787	35.65	0.291	1 + 0.400B + 0.619B^2
0.770	61.28	0.5	1 + 0.770B
0.646	6.17	0.050	1 - 1.227B + 0.417B^2

Figure 2 shows the dual data and again a strong periodic appearance is seen. Moreover, unlike the sample spectrum and AR(18) spectral estimate, the sample M-spectrum and Euler(11) spectral estimates show two very strong peaks, one at $f_1^* = 17.42$ and the other at $f_2^* = 3.22$ as is suggested in Table 1. Therefore the dual does have a strong cyclic nature and hence the autocorrelations of the dual have a slowly damped periodic nature that is reflected by the sample ACF of the dual shown in Figure 2.

The Euler(11) spectrum shown in Figure 2 indicates that the original data should exhibit a frequency behavior that varies like $(a + bt)^{-1}$. This can be clearly seen by the instantaneous spectrum shown in Figure 3. The estimated instantaneous frequency is shown for $t \in (0, 400)$. Figure 4 shows "snapshots" of the estimated instantaneous spectrum at $t = 1, 200, 300$, and 400. The decreasing frequency structure is again apparent.

Although the changing period and frequency are properties of the autocorrelation, for roots as close to the unit circle as these, this same behavior is to a great degree imposed on the data. In that regard it is interesting to estimate the instantaneous frequency, using Figure 3 or Equation 12, and observe the extent to which these cycles are present in the data. That is, from (12) we can write

$$\hat{f}^* = \left[\ln\left(\frac{1+\hat{f}\hat{h}^j\hat{h}^k}{\hat{f}\hat{h}^j\hat{h}^k}\right)\right]^{-1},$$

so that

$$\hat{f} = \frac{1}{\hat{h}^j\hat{h}^k(e^{1/\hat{f}^*}-1)}. \tag{16}$$

Note that $\hat{h}^j\hat{h}^k \sim \hat{h}^j + t$ where t is the observation index of the equally spaced data. So, denoting $\hat{f}$ by $\hat{f}(t; f^*)$ to emphasize that $\hat{f}$ is a function of t for each f^*, we have

$$\hat{f}(t; f^*) \sim \frac{1}{(\hat{h}^j + t)(e^{1/\hat{f}^*}-1)} \tag{17}$$

where $\hat{h}^j = 15$ mentioned previously. We will examine the low frequency component to determine to what degree this changing frequency can be observed in the data. From (16) we see that $\hat{f}(1; 3.22) = 1/5.8$, $\hat{f}(7, 3.22) = 1/8$, and $\hat{f}(150, 3.22) = 1/60$. Thus at the beginning of the realization the autocorrelation has a period between 5 and 6 while by $t = 150$ that period has lengthened to 60. Figures 5a and 5b show the data from 1 to 25 and 150 to 209. The data are clearly very representative of the correlation, as can be seen from these figures. That is in Figure 5a we count 6 points in what appears to be the first cycle associated with this frequency, while we count 8 points in the second cycle as predicted by $\hat{f}$. Similarly in Figure 5b it appears that the 60 points is approximately a cycle. Thus the instantaneous spectrum seems to describe the changing cyclic nature of the data quite well.

Finally, Figure 6 shows the forecasts of the last 20 points using an AR(18) model and an Euler(11). The Euler(11) forecasts are dramatically better as would be expected. The AR(18) forecasts emphasize the fact that ignoring the fact that the frequencies seem to be changing in time and proceeding with standard methods can result in exceedingly poor results.

3. Filtering M-Stationary Processes

In this section we introduce a new approach to filtering processes with time varying frequencies that change like $(at+b)^{-1}$. First we introduce the discrete M-linear process.

Gray and Zhang (1988) defined the continuous causal M-linear process $\{X(t)\}$ by

$$X(t) - \mu = \int_1^{\infty} h(u)a(\frac{t}{u})d(\ln u), \tag{18}$$

where $a(t)$ is M-white noise. This definition is extended by Gray, Vijverberg and Woodward (2004) to the discrete M-linear process by the following

$$X(t) - \mu = \sum_{j=0}^{\infty} \psi_j a(\frac{t}{h^j}) = \sum_{j=0}^{\infty} \psi_j a_{h^{k-j}}, \tag{19}$$

for $k = 0, \pm 1, ...$, and where a_t is white noise with $E(a_t) = 0$ and $\mathrm{Var}(a_t) = \sigma_a^2 < \infty$. We generalize $X(t)$ in (18) and (19) by replacing $a(t)$ and a_{h^k} by M-linear processes. In that case we will refer to $X(t)$ in (18) as a continuous M-linear filter and to $X(t)$ in (19) as a discrete M-linear filter. To be specific we have the following definition.

Definition 9. Let $h > 1$ and $t \in S$, where $S = \{t : t = h^k, k = 0, \pm 1, \ldots\}$, and let $\{X(t)\}$ and $\{Z(t)\}$ be discrete M-linear processes. Then we define $X(t)$ to be a discrete M-linear filter if

$$X(h^k) - \mu = \sum_{j=0}^{\infty} \psi_j Z\left(\frac{h^k}{h^j}\right) = \sum_{j=0}^{\infty} \psi_j Z(h^{k-j}). \tag{20}$$

The dual process of $X(t)$ is then defined by $Y_k = X(h^k)$. Thus

$$Y_k - \mu = \sum_{j=0}^{\infty} \psi_j W_{k-j} = \psi(B) W_k, \tag{21}$$

where $W_k = Z(h^k)$ is the dual of Z. Now let S_U denote the power spectrum or spectral density of a process $\{U(t)\}$. Then it is well known that

$$S_Y(f) = |\psi(e^{-2\pi i f})|^2 S_W(f). \tag{22}$$

As mentioned earlier, $G_X(f^*) = S_Y(f)$ where $f = f^* \ln h$ and where X_t is an M-stationary process and Y is its dual. So, for $\{X(t)\}$ defined by Equation (20), it follows that

$$\begin{aligned} G_X(f^*) = S_Y(f) &= |\psi(e^{-2\pi i f})|^2 S_W(f) \\ &= |\psi(e^{-2\pi i f^* \ln h})|^2 S_W(f) \\ &= |\psi(e^{-2\pi i f^* \ln h})|^2 G_Z(f^*). \end{aligned} \tag{23}$$

From (22) and (23) it follows that an M-stationary process can be filtered by using standard methods to filter its stationary dual. The result in (23) can easily be shown to hold when the corresponding input and output processes are continuous. Finally it should be remarked that when the roots of the characteristic equation associated with the dual in (6) lie outside the unit circle, $X(t)$ in (19) can always be written as a convergent M-linear process.

In this section we will demonstrate the application of the previous results to filtering. In general it is our intent to demonstrate that standard filtering methods can be employed once the proper time transformation is made. For a discussion of filtering methods based on stationarity see Shumway and Stoffer (2000) and Hamming (1998). However, ignoring the time transformation and applying such methods may lead to unacceptable results. Of course one could also apply a window-based filtering method. However here we focus on time transformation methods which, when the frequencies change monotonically like a power function, tend to out-perform window-based methods. See Gray, Vijverberg and Woodward (2004) and Jiang, Gray, and Woodward (2004). In Section 4 we will give similar examples for a more general class of time transformations. We will first give two simulated examples and then apply the results to a well studied bat echolocation signal.

Example 2. Consider once again the process in Example 1. That is we consider a realization of length 100 (shown in Figure 1) generated from the discrete Euler model

$$\phi_1(B)\phi_2(B)X(h^k) = a(h^k),$$

with $E[a(h^k)] = 0$ and $\phi_1(B)$ and $\phi_2(B)$ given in (14). From the spectrum in Figure 1, the data appear to be primarily low frequency. To remove the higher frequencies one would assume that one could apply a low-pass filter.

We now use the ideas presented above to low-pass filter the data in Figure 1 by first applying a low-pass filter to the dual data in Figure 2. From the M-sectrum in Figure 2 and Table 1, it is clear that it would be reasonable to pass the dual data through a low-pass filter with cutoff .08. The dual data were therefore passed through a 4th order Butterworth filter (see eg. Hamming, 1998) to obtain a filtered dual data set, which in turn produces a filtered M-stationary process at $\hat{h}^k$. These values are then interpolated to give the filtered equally spaced observations from the underlying continuous process. The result is shown in Figure 8b. The filter has essentially extracted only that part of the signal associated with $f^* = 3.22$, i.e. the lower of the two dominant M-frequencies.

To illustrate the problems associated with applying standard filters directly to the data in Figure 1, we applied several different fourth order Butterworth filters (Hamming, 1998) with cutoff frequencies .06, .07, .08 and .09. Figure 7 shows the results. Note that none of these filters successfully filtered out the higher frequency over the entire record. Comparison of Figure 7a with the data in Figure 1 shows that the cutoff was too low to pass what appears to be the first 3 cycles. On the other hand increasing the cutoff to .09 does not really help because by $n = 300$ the higher frequency component has decreased to the point that it is below the filter cutoff and thus both frequencies pass through the filter. This can be seen in Figure 7d. In Figure 7d the front end of the data is filtered better in that the higher frequency is removed without removing the entire signal. However inspection of that figure shows that while the beginning may be a bit better, the end is worse, in that now the higher frequency component has decreased to the point that it is below the cutoff threshold. Similar problems occur with high pass filters, band-pass-filters, etc. Thus standard filtering procedures cannot successfully remove the frequencies associated with $f^* = 17.42$.

The source of the problem with the above filtering can be vividly seen in Figures 3 and 4. Note that at $t = 400$ the higher instantaneous frequency has been reduced to .04 which is substantially below the lower frequency, .17 at $t = 1$. Thus, any low pass filter applied to the data must either pass frequencies associated with $\hat{f}^* = 17.42$ near $t = 400$ or it must filter out frequencies associated with $f^* = 3.22$ near $t = 1$. So if the goal is to filter out the signal associated with $f^* = 17.42$, it cannot be done using any filters based on stationarity.

The M-filtering procedure is summarized below. Filtering data to remove frequencies associated with specified M-frequencies, i.e. frequencies varying in time like $(a + bt)^{-1}$, will be referred to as M-filtering.

M-Filtering Summary

1. Estimate h, the offset, and the best M-stationary process, $\hat{\phi}(B)\hat{X}(\hat{h}^k - \overline{X}) = a(h^k)$.
   See Vijverberg (2002) or Gray, Vijverberg and Woodward (2004) for details. See also http://faculty.smu.edu/hgray/research.htm.

2. Let $Y_k = X(\hat{h}^k) - \overline{X}$ to obtain the dual.

3. Filter the dual by an appropriate filter. Denote the resulting filtered dual data as $\{F(Y_k)\}$.

4. Define $F(X(h^k)) = F(Y_k)$.

5. Interpolate $\{F(X(h^k))\}$ to obtain $\{F(X_k)\}$, i.e. the M-filtered data at the original time points.

Steps 1 and 2 can be accomplished by making use of the authors' available software. This will produce the dual data which can then be filtered by stationary methods to obtain $\{F(X(h^k)\}$. Then $\{F(X_k)\}$ can be obtained by interpolation. In this paper we have used linear interpolation.

Example 3. Consider now a realization from the M-stationary continuous process $\{X(t)\}$ given by

$$X(t) = \cos(36\pi(\ln(t + 175) + \psi)) + .5\cos(100\pi\ln(t + 175) + \psi) + .1n(t)$$

where $n(t) = N(0, 1)$ and $\psi \sim$ Uniform $[0, 2\pi]$. See Gray and Zhang (1988). Figure 9a shows a realization of length 400 from this process where $\psi = 0$. There it can be seen that the data show two time-varying frequencies, a lower frequency component associated with $\cos(36\pi\ln t)$ and a higher frequency component associated with $\cos(100\pi\ln t)$. Figure 10a shows the corresponding estimated M-spectrum based on an Euler(18) fit to the data. Figures 9c and 9d show the results of applying a 4th order Butterworth lowpass filter with cutoffs .12 and .3 respectively. As in the previous example, it is clear that the two signals cannot be separated by filtering the data using stationary methods since the frequencies corresponding to the $\cos(36\pi\ln t)$ term at the beginning of the realization are higher than the frequencies associated with $\cos(100\pi\ln t)$ at the end of the data. This can be clearly seen from Figures 10b and 10c that show the instantaneous spectra based on the Euler(18) fit and snapshots of the instantaneous spectrum respectively. Finally, Figure 9b shows the original data filtered by transforming time and filtering the dual. Note that except for a small amount of noise in the amplitude, the signal $\cos(36\pi\ln t)$ is almost perfectly recovered except at the end points. This is due to the Butterworth filter since no effort was made to window the data. This will be the case in all of our examples, i.e. we will not window the dual in any way.

Example 4. In this example, we consider echolocation data from a large brown bat. The data were obtained courtesy of Al Feng of the Beckman Institute at the University of Illinois. The entire data set is shown in Figure 11a while close-ups of the first 100 and the last 60 points are shown in Figure 11b and 11c, respectively. The data consist of 381 data points taken at 7-microsecond intervals with a total duration of .00266 seconds. Unlike the previous examples, the instantaneous period does not visually appear to be linear over the entire signal. In fact, the signal appears to be made up of possibly two different signals. However, numerous studies have confirmed that such bat signals as this have a frequency structure that changes with time like $(a+bt)^{-1}$ (e.g. Masters, Jacobs, and Simmons, 1991). Thus the data appear to be a good candidate for modeling as an M-stationary process. In fact among the class of models whose frequencies change asymptotically as t^{β} (see Section 4), the GWS software selects $\beta = -1$, i.e., the M-stationary case, as the best model. We will therefore consider the application of the Euler model to this data set and compare its usefulness with the autoregressive model. Based on the AIC, an AR(20) was fit to the data using standard methods, and using the methodology described here, an Euler(11) with offset equal to 203 was determined to be the best Euler model. In each case a maximum model of order 20 was considered. Tables 2 and 3 show the factors of the AR(20) and Euler(11) models along with the corresponding frequencies and their proximity to the unit circle. Figures 12a and 12b show the sample ACF and the AR(20) spectral estimator while Figures 12c and 12d show the sample M-ACF and Euler(11) spectral estimator for the bat data. The lack of an indication of a periodic component in the sample ACF and AR(20) spectral estimator are quite surprising in view of the cyclic appearance of the data. This is due to the fact that although there clearly is a cyclical nature to the data, the cycle is lengthening slightly with time. As a result the usual spectrum is spread and the correlation is lengthening slightly with time. Thus, the correlation changes with time resulting in the sample estimates shown in Figures 12a and 12c. It should be pointed out that even though the Euler process has an elongating period, the M-ACF does not depend on time, nor does the M-spectrum. Figures 12c and 12d clearly indicate the cyclic behavior of the data on the log scale. It is important to note that the energy in the signal is primarily concentrated at approximately the M-frequencies $53k$, k = 0,1,2,3. We will refer to the cases $k = 1, 2,$ and 3 as the fundamental and its M-harmonics. Figures 13a and 13b show the residuals. Clearly most of the variation in the data has been accounted for in the case of the EAR(11), unlike the AR(20) fit.

Table 2: Factor Table for AR(20) Fit for Large Brown Bat Data

Absolute Reciprocal	Frequency	Factors
.997	.0	1-.997B
.960	.149	1-1.141B+.922B^2
.950	.097	1-1.557B+.903B^2
.930	.178	1-.817B+.864B^2
.919	.500	1+.919B
.905	.258	1+.094B+.819B^2
.892	.225	1-.274B+.795B^2
.890	.299	1+.540B+.792B^2
.887	.441	1+1.652B+.787B^2
.884	.389	1+1.357B+.782B^2
.878	.341	1+.952B+.771B^2

Table 3: Factor Table for Euler(11) Fit for Large Brown Bat Data

Absolute Reciprocal	M-Frequency	Dual Frequency	Factors
.997	53.2	.148	1-1.195B+.993B^2
.996	105.8	.294	1+.539B+.991B^2
.970	0.0	.0	1-.97B
.955	156.7	.435	1+1.753B+.911B^2
.706	61.1	.170	1-.684B+.498B^2
.658	135.9	.377	1+.944B+.433B^2

Figure 14 shows the instantaneous spectral estimate, $\hat{S}(f, h^k; h^j)$, for $0 \leq f \leq .5$ associated with the $G(11, 0; 0)$ model. Again, the instantaneous frequencies appear to be decreasing, indicating that the periods are lengthening. Thus at the beginning of the data the major source of the variation is at frequencies .26 and above while at the end of the data the variation is concentrated at frequencies .27 and below. In this regard it is interesting to note that the instantaneous frequency, .26, associated with $f^* = 53$ at the initial observation is almost exactly equal to the instantaneous frequency, .27, associated with $f^* = 157$ at the final data point $t = 381$. Also it should be noted that the sampling rate is not quite fast enough at the beginning of the data. That is, the Nyquist frequency for the equally spaced data obtained at $\Delta = 7$ microsecond units is $1/2\Delta$, so the highest instantaneous frequency that can be detected at a given t, is $f = f(h^k, f^*) \leq 1/2$. It can be shown that at $t = 9$ the instantaneous frequency associated with $f^* = 106$ is approximately .5 so that beginning at about the 9th data value, we can detect the M-frequency $f^* = 106$. This is visually displayed in Figure 11 in the sense that initially we cannot visually detect the frequencies associated with M-frequencies 106 and 157. However, by the 9th data point we begin to see the appearance of $f^* = 106$. Also, the instantaneous frequency associated with $f^* = 157$ cannot be fully detected until about the 100th observation in Figure 11. This explains the unusual appearance of the data beginning around the 100th data point. Up to this point, this highest frequency of the underlying signal has been too high to detect at this sample rate, and as a result has been completely aliased.

In Figure 15a we show the modulus of the continuous wavelet transform, in Figure 15b we show the Gabor transform, and in Figure 15c we show the Wigner-Ville time frequency distribution. Figures 15 a and b were obtained using The Rwave package, and involve transformed versions of the time and frequency axes. In the case of the wavelet transform the vertical axis is based on “scale” which is an inverted version of frequency. These window-based presentations of the time-varying spectral content also tend to show a portion of the fundamental frequency and its “harmonics.” However, in these representations, the periodic behavior toward the beginning and end of the realization are not seen and the zero frequency is not visible. However, as described above, examination of the data shows that the instantaneous spectrum does an excellent job of describing the frequency behavior throughout the entire realization. The improvement is obtained since, in contrast to the window-based methods, M-stationary analysis uses the entire data set to estimate spectral information at each frequency.

Finally suppose we wish to filter the data to remove the aliased data, i.e. the data associated with the M-frequency 157. There is no way to accomplish this with stationary-based methods. From Figure 12d it is clear that the M-filtering, i.e. filtering the dual and reinterpolating, should be effective in this case. As mentioned earlier, one could alternatively apply window-based methods. However, the time deformation approach is preferred here since these data are so well modeled as M-stationary. Figure 16b shows the result of applying a fourth order low-pass Butterworth filter to the dual using a cutoff of .4 which is suggested by the Euler(11) M-spectrum in Figure 12d. Clearly the high M-frequency behavior has been removed and the data are no longer aliased.

4. Filtering $G(\lambda)$-Stationary Processes

To this point we have focused on M-stationary processses. However, Jiang (2003) has generalized the M-stationary process to accommodate a wide range of time deformations. This new class is referred to as $G(\lambda)$-stationary processes, and in this case, the time deformation is the Box-Cox transformation defined by

$$u_\lambda(t) = \frac{t^\lambda - 1}{\lambda}, \quad -\infty < \lambda < \infty. \tag{24}$$

Since $\lim_{\lambda \to 0} u_\lambda(t) = \ln t$, the limiting case, denoted by $\lambda = 0$, is the M-stationary process. It is shown by Jiang, Gray and Woodward (2004) for $\lambda < 1$, that $G(\lambda)$ processes have an elongating cycle length while if $\lambda > 1$ they have contracting length. If $\lambda = 1$ the process is the standard stationary case, i.e. the frequencies are fixed in time.

Additionally Jiang (2003) shows that the instantaneous frequencies of a $G(\lambda)$ processes vary asymptotically like αt^β as $t \to \infty$, where $\alpha > 0$ and $\beta = \lambda - 1$. The theoretical development of $G(\lambda)$ processes represents a considerable generalization of M-stationary processes, and given a data set, the GWS software referenced earlier can be used to find the $G(\lambda; p)$ model that best fits the data. A $G(\lambda; p)$ model is a $G(\lambda)$ model whose dual is a pth order autoregressive model. See Jiang (2003).

Just as in the M-stationary case, the first step in processing $G(\lambda; p)$ stationary data is to estimate λ and transform to the appropriate dual process. Thus $G(\lambda)$ filtering can be performed in the same manner as in the M-stationary process. Since each transformation defines a different sampling scheme, the only difference in the filtering is the locations on the time axis at which the data are sampled.

The steps for $G(\lambda)$-filtering are the same as those listed in the previous summary for M-filtering and the first two steps can be performed with the GWS software. That is if $\{Y_k(\lambda)\}$ denotes the dual process, then we filter this stationary data appropriately to obtain $\{F[Y_k(\lambda)]\} = \{F[X(t_k(\lambda)]\}$. This produces the filtered data at the points $t_k(\lambda)$ which are then re-interpolated to obtain the filtered data at the original data points.

As noted in (24), $G(\lambda)$ processes are based on the Box-Cox transformation of time. Thus if we let $u = u_\lambda(t)$ we have

$$u = \frac{t^\lambda - 1}{\lambda} \Rightarrow t = (u\lambda + 1)^{1/\lambda}, \quad t > 0. \tag{25}$$

Then $X(t) = X(u\lambda + 1)^{1/\lambda} = Y(u)$, where $Y(u)$ is referred to as the dual of $X(t)$ as before. Sampling the dual at $(k + \xi)\Lambda$, where ξ is fixed, results in sampling $X(t)$ at $t_k(\lambda; \xi) = [(k - 1 + \xi)\Lambda\lambda + 1]^{1/\lambda}$ (and visa versa) as mentioned in the above. The quantity $\Delta = [\xi\Lambda\lambda + 1]^{1/\lambda} - 1$ is called the "offset" or "realization origin". Note if $\xi = 0$ there is no offset.

The dual is modeled as an AR process (Jiang, 2003). We should also note the relationship between the frequency, $f_d(\lambda)$, of the dual process and the frequency, $f_G(\lambda)$, of the $G(\lambda;p)$ process. As in the case of the M-stationary process, the only difference in the spectrum of the dual and the $G(\lambda)$-spectrum is the scale. As noted earlier $f_d(0) = f^* \ln h = f_G(0)$. For $\lambda \neq 0$ it can be shown that $f_d(\lambda) = f^*\Lambda = f_G(\lambda)$. In the M-stationary case, clearly $\Lambda = \ln h$.

Among the more common types of data from $G(\lambda)$ processes (or at least asymptotically $G(\lambda)$ processes) are chirps and Doppler signals. In both cases the problem of filtering one such signal from a background of several is a difficult problem of much interest. For example, Xu, Durand, and Pibarot (2000) describe a new approach, based on the time-frequency representation of transient nonlinear chirp signals, for modeling the aortic and the pulmonary components of the second heart sound. They demonstrate that each component is a narrow-band signal with decreasing instantaneous frequency. As another example, a Doppler signal results from the back scattering of an ultrasound beam by moving red blood cells. Flow disturbances and changes in the velocity waveform result in an increase in Doppler spectral width which is used to detect atherosclerotic lesions in arteries. Due to time-varying frequencies, current instruments are only marginally effective in detecting such lesions. See Baston, Fish and Vaz (1999).

These problems can however be addressed by noting that chirp and Doppler signals are well modeled as $G(\lambda)$ processes where asymptotically $\lambda = 2, 3$ for a linear and quadratic chirp respectively while $\lambda \sim -1$ for many Doppler-type signals. In the final two examples we demonstrate how "$G(\lambda)$-filtering" can be used to separate such signals even when their "G-frequencies" are quite close to each other.

Example 5. Consider now a realization of size 400 from the quadratic chirp given by

$$X(t) = \cos[2\pi(\frac{t}{250} + .25)^3] + 2\cos[7\pi(\frac{t}{250} + .25)^3] + .1n(t), \tag{26}$$

where $n(t)$ is normal (0,1) noise.

Figure 17 shows the data and its two deterministic components while Figure 18 shows the data along with its sample ACF and sample spectrum overlaid with an AR(15) Burg spectral estimate. The spectrum is clearly spread as is typical of chirp signals. It provides little evidence of the fact that the data are made up of two superimposed chirps and even less suggestion of how it might be filtered to separate the two signals.

A $G(\lambda)$ model is fit to the data and we obtain $\hat{\lambda} = 3, \hat{\Lambda} = 81127$ and $\hat{\Delta} = 60$. Therefore the observed data are taken at $\hat{\Delta} + 1 = 61, \hat{\Delta} + 2 = 62, \ldots, \hat{\Delta} + 400 = 460$. But ,

$$\begin{aligned} t_k(\lambda;\xi) &= [(k-1+\xi)\Lambda\lambda + 1]^{1/\lambda} \\ &= [((\xi\Lambda\lambda+1)^{1/\lambda})^{\lambda} + (k-1)\Lambda\lambda]^{1/\lambda} \\ &= [(\Delta+1)^{\lambda} + (k-1)\Lambda\lambda)]^{1/\lambda}. \end{aligned}$$

Therefore

$$t_k(3;\hat{\xi}) = [(\hat{\Delta}+1)^{\hat{\lambda}} + (k-1)\hat{\Lambda}\hat{\lambda}]^{1/\hat{\lambda}} = [(61)^3 + 3(81127)(k-1)]^{1/3}.$$

So, since the data are given at 61, 62, ... , 460, we interpolate to obtain the data at $t_1(3;\hat{\xi}) = 61,\ t_2(3;\hat{\xi}) = 77.77,\ t_3(3;\hat{\xi}) = 89.37, \ldots,\ t_{399}(3;\hat{\xi}) = 459.62,\ t_{400}(3;\hat{\xi}) = 460$ in order to compute the dual process.

The resulting dual realization is shown in Figure 19a. Also shown in Figure 19 are the dual sample ACF and the dual sample spectrum overlaid with an AR(20) spectral estimator. Note the dual clearly does not exhibit any time varying frequency behavior. Moreover the data appear to be made up primarily of two cycles of very distinct frequencies.

Figure 20 shows the instantaneous spectrum and snapshots at $t = 1$, 100, 200, and 400. From these graphs it can be seen that the data are made up primarily of two signals with monotonically increasing frequencies. At $t = 1$ all the frequencies are near zero while by $t = 400$ the highest frequency with significant power is around .15. From the snapshots it is clear that neither stationarity based high-pass or low-pass filtering will be effective in identifying and separating the signals since the frequencies with any power are all below .02 at the beginning and all above .02 by the end of the data. Figure 21 shows the data alongside the results of using a 4th order low-pass filter directly on the data at cutoff frequencies .05 and .1 and $G(\lambda)$-filtering using a 4th order Butterworth filter with a .03 cutoff. As predicted the direct application of the Butterworth filter was totally ineffective. On the other hand low-pass filtering the dual and reinterpolating to obtain the filtered data produces Figure 21b which is almost a perfect recovery of the lower "frequency" chirp, except for the end point effects. Figure 22 shows similar results for a fourth order high-pass Butterworth filter. Again the $G(\lambda)$ method is quite good, but high-pass filtering the data directly gives poor results as is seen in Figures 22c and d.

Example 6. In this final example we consider a Doppler-type signal of the form

$$X(t) = \sin\left(\frac{840\pi}{t+50}\right) + .5\sin\left(\frac{2415\pi}{t+50}\right) + .05n(t), \tag{27}$$

where $n(t)$ is normal (0,1) noise.

Figure 23 shows a realization of size 200 and a plot of the two deterministic components given in (27), Figure 24 shows the data along with the sample ACF and the sample spectrum with the AR(8) spectral estimate overlaid. As in the case of the chirp signal, the spectrum is quite spread and is of no help in separating the signals. Using a $G(\lambda)$ analysis we obtain $\hat{\lambda} = -1.7$, $\hat{\Delta} = 90$, and $\hat{\Lambda} = 1.1888 \times 10^{-6}$. Therefore the data were interpolated to obtain values at

$$\begin{aligned} t_k(-1.7;\hat{\xi}) &= [(\hat{\Delta}+1)^{\hat{\lambda}} + (k-1)\hat{\Lambda}\hat{\lambda}]^{1/\hat{\lambda}} \\ &= [(91)^{-1.7} - 1.7\hat{\Lambda}(k-1)]^{-1/1.7} \end{aligned}$$

$t_k(-1.7;\hat{\xi}) = 91, 91.23, 91.47, \ldots, 284.83, 290$. Figure 25 shows the dual which appears to have fixed frequencies. This is confirmed by Figure 25 and Table 4 which shows that the dual is primarily composed of signals at the two frequencies .095 and .034 on the dual scale. The remaining part of the spectrum is mostly side lobes due to the noise. This suggests that the signals can be separated by low-pass filtering the dual with a cutoff of around .05 to .07.

Table 4. Factor Table for $G(-1.7)$ Fit to Doppler Data in Figure 24(a)

Absolute Reciprocal	G(-1.7)-Frequency	Dual Frequency	Factors
0.996	79570	.095	1 - 1.6497B + 0.9914B^2
0.974	28940	.034	1 - 1.9031B+ 0.9492B^2
0.904	150700	.179	1 - 0.7785B + 0.817B^2
0.883	374500	.445	1 + 1.6619B + 0.7795B^2
0.861	302100	.359	1 + 1.0901B + 0.7414B^2
0.857	219500	.261	1 + 0.117B + 0.7345B^2

Figure 26a shows the instantaneous spectrum where it can also be seen that the data are primarily composed of two time varying frequencies. It is again of interest to consider the snapshots shown in Figure 26b where it is very clear that at the beginning of the data almost all of the power in the spectrum is above the frequency .1 while by $t = 100$ it is almost all below .1. For this reason filtering the data by stationary methods "as is" is doomed to failure. Figure 27 confirms this observation by showing the results of a low-pass and a high-pass 4th order Butteworth filter applied directly to the data. As in previous examples, these filtering results are poor. However, if the data are $G(\lambda)$-filtered, i.e. the dual is filtered and transformed back to original scale, the results are quite good. Figure 28 shows these results where it can be seen that the two components are almost perfectly recovered.

Finally in order to demonstrate that the $G(-1.7; 12)$ process fit to the data is, in fact an excellent fit, we show the forecasts obtained by fitting the best AR(p) to the raw data and the $G(-1.7; 12)$ forecasts. The $G(-1.7; 12)$ forecasts are in fact quite good, whereas the AR(8) forecasts essentially move directly to the mean and stay there.

Concluding Remarks

We have shown that the difficult problem of filtering time-varying frequencies can in many cases be accomplished using time deformation to transform to a signal with fixed frequencies. Then standard filtering techniques can be used to remove the desired signal before transforming back to the original time scale. We have shown how this technique can be successfully applied to data sets for which frequencies are monotonically increasing or decreasing.

References

[1] Baston, C., Fish, P., Vaz, F. (1999). Spectrum of Doppler Ultrasound Signals from Nonstationary Blood Flow, *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, **46**, 1201-1217.

[2] Choi, E., Gray, H. L., and Woodward, W. A. (2003). EARMA Processes with Spectral Analysis of Non-Stationary Time Series, SMU Tech Report, #313.

[3] Gray, H. L. and Zhang, N. F. (1988). On a Class of Nonstationary Processes, *Journal of Time Series Analysis*, **9**, 133-154.

[4] Gray, H. L., C. P. C. Vijverberg and Woodward, W. A. (2004). Nonstationary Data Analysis by Time Deformation, to appear in *Communications in Statistics A*.

[5] Hamming, R. W. (1998). *Digital Filters*, Dover Publications: Mineola, NY.

[6] Jiang, H., Gray, H. L. and Woodward, W. A. (2004). Theory and Applications of G(λ) Processes to Time Frequency Analysis, Southern Methodist University, Statistical Science Dept., Technical Report #311.

[7] Jiang, H. (2003). Time-Frequency Analysis: G(λ) Stationary Processes. Ph.D. Dissertation, Southern Methodist University.

[8] Masters, W. M., Jacobs, S. C., Simmons, J. A. (1991). The structure of echolocation sounds used by the big brown bat. *Eptesicus fuscus*: Some

consequences for echoprocessing. *Journal of the Acoustical Society of American*, **89**, 1402-1413.

[9] Priestly, M. B. (1981). *Spectral Analysis and Time Series.* London: Academic Press.

[10] Vijverberg, C. P. C. (2002). *Discrete Multiplicative Stationary Processes.* Ph.D. Dissertation, Department of Statistical Science, Southern Methodist University.

[11] Xu, J., Durand, L., and Pibarot, P. (2001). Extraction of the Aortic and Pulmonary Components of the Second Heart Sound Using a Nonlinear Transient Chirp Signal Model, *IEEE Transactions in Biomedical Engineering*, **48**, 277-283.