Statistics and Its Interface Volume 16 (2023) 97–108

AutoSpec: detection of narrowband frequency changes in time series

David S. Stoffer

Most established techniques that search for structural breaks in time series have a difficult time identifying small changes in the process, especially when looking for narrowband frequency changes. The problem is that many of the techniques assume very smooth local spectra and tend to produce overly smooth estimates. The problem of oversmoothing tends to produce spectral estimates that miss slight frequency changes because frequencies that are close together will be lumped into one frequency. The goal of this work is to develop techniques that concentrate on detecting slight frequency changes by requiring a high degree of resolution in the frequency domain.

Keywords and phrases: Spectral analysis, Structural breaks, Whittle likelihood, Minimum description length.

1. INTRODUCTION

The focus here is the detection of slight frequency changes in time series as opposed to, for example, the detection of level shifts, for which many methods exist (e.g., see Aminikhanghahi and Cook, 2017). Many time series are realizations of nonstationary random processes, hence estimating their time varying spectra may provide insight into the physical processes that give rise to these time series. For example, EEG time series are typically nonstationary, and estimating the time varying spectra based on the EEG of epilepsy patients may lead to methods capable of predicting seizure onset (e.g., see Aksenova, Volkovych and Villa, 2007). Similarly, analyzing the time varying spectrum of the Southern Oscillation Index (SOI) may further our knowledge of the frequency of the El Niño Southern Oscillation (ENSO) phenomenon and its impact on global climate (e.g., see An and Wang, 2000).

Most current techniques that search for structural breaks in time series spectra, however, may not be able to identify slight frequency changes at the resolution of interest. While resolution depends on the particular application, a problem is that established techniques generally assume smooth local spectra that produce overly smooth estimates. In turn, the smooth estimates can miss slight frequency changes because peaks in the spectrum at frequencies that are close together will be combined by the estimator; see Section 2 for further details. The goal of this work is to develop techniques that concentrate on detecting narrowband frequency changes by requiring a high degree of resolution in the frequency domain.

The basic assumptions here are that, conditional on the location and number of segments, the time series is piecewise stationary with each piece having a spectral density. A detailed description of the model is given in Section 3. In addition to representing time series that have regime changes, the model can be used to approximate slowly varying processes; e.g., see Adak (1998), which uses dyadic segmentation to find the approximate location of breakpoints. The approach taken in Davis, Lee and Rodriguez-Yam (2006) was to fit piecewise AR models using minimum description length and a genetic algorithm for solving the difficult optimization problem. Ombao et al. (2001) proposed nonparametric estimators based on dyadic segmentation and smooth local exponential functions. Rosen, Wood and Stoffer (2009) estimated the log of the local spectrum using a Bayesian mixture of splines. The basic idea of this approach is to first partition the data into small sections. It is then assumed that the log spectral density of the evolutionary process in any given partition is a mixture of individual log spectra. A mixture of smoothing splines model with time varying mixing weights is used to estimate the evolutionary log spectrum. Later, Rosen, Wood and Stoffer (2012) improved on the technique of Rosen, Wood and Stoffer (2009) by adaptively selecting breakpoints.

For background, note that spectral analysis has to do with partitioning the variance of a stationary time series into components of oscillation indexed by frequency, ω, and measured in cycles per unit of time. Given a time series sample, $\{X_t; t = 1, ..., n\}$, that has been centered by its sample mean, the sample spectral density (or *periodogram*) is defined in terms of frequency ω:

$$I_n(\omega_{k_n}) = \left| n^{-1/2} \sum_{t=1}^{n} X_t \, \mathrm{e}^{-2\pi i \omega_{k_n} t} \right|^2, \tag{1}$$

where $\omega_{k_n} = k_n/n$ for $k_n = 1, \ldots, n-1$. The periodogram is essentially the squared-correlation of the data with sines and cosines that oscillate at frequency ω.

The spectral density, $f(\omega)$, of a stationary time series can be defined as the limit ($n \to \infty$ and $\omega_{k_n} \to \omega$) of $E[I_n(\omega_{k_n})]$, provided that the limit exists; details can be found in Shumway and Stoffer (2017, Ch. 4). It is worth-

while to note that $f(\omega) \geq 0$, $f(\omega) = f(-\omega)$, and

$$\int_{-1/2}^{1/2} f(\omega)\, d\omega = 2\int_{0}^{1/2} f(\omega)\, d\omega = \sigma^2, \tag{2}$$

where $\sigma^2 = \text{var}(X_t) < \infty$. Thus, the spectral density can be thought of as the variance density of a time series relative to frequency of oscillation. That is, for positive frequencies between 0 and 1/2, the proportion of the variance that can be attributed to oscillations in the data at frequency ω is roughly $2f(\omega)d\omega$. If the time series X_t is *white noise*, that is, $E(X_t) = 0$ and $\text{cov}(X_s, X_t) = 0$ for all $s \neq t$, then $f(\omega) \equiv \sigma^2$. The designation white originates from the analogy with white light and indicates that all possible periodic oscillations are present with equal strength.

Under mild conditions, for large samples ($n \to \infty$ and $\omega_{k_n} \to \omega$),

$$\frac{2\, I_n(\omega_{k_n})}{f(\omega)} \dot{\sim} \chi^2_2\,, \tag{3}$$

where χ^2_2 is the chi-squared distribution with 2 degrees of freedom. It follows from (3) that

$$E[I_n(\omega_{k_n})] \approx f(\omega) \quad \text{and} \quad \text{Var}[I_n(\omega_{k_n})] \approx f^2(\omega).$$

Thus, the periodogram is not consistent for the spectral density. To get a consistent estimate, we can use nonparametric kernel smoothing. Most estimates can be written in the form

$$\widehat{f}(\omega) = \int_{-1/2}^{1/2} K_n(\omega - \lambda) I_n(\lambda) d\lambda, \tag{4}$$

where $K_n(\omega)$ is the spectral window. The integral is typically approximated by a sum, and the amount of smoothing of the periodogram is controlled by the bandwidth, B, of the window. The bandwidth determines essentially the number of parameters being estimated; i.e., the number of distinct bands of frequencies for which $f(\omega)$ is estimated. We mention that Hannan and Rissanen (1988) presented a method based on Rissanen (1978) to choose the bandwidth B via stochastic complexity and minimum description length (MDL), which of interest in this paper. The idea will be discussed further in the Appendix.

A parametric approach is to fit an AR(p) to the data,

$$X_t - \mu = \sum_{j=1}^{p} \phi_j (X_{t-j} - \mu) + Z_t\,,$$

where Z_t is white noise with variance σ^2, and the fitted order p is determined by a model selection criteria such as AIC or BIC. The spectral density of an AR(p) model is given by

$$f(\omega) = \sigma^2_w\, |\phi(e^{-2\pi i\omega})|^{-2}, \tag{5}$$

where

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p.$$

If $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p$ and $\hat{\sigma}^2_w$ are the estimates from an AR(p) fit to data, then an estimate of $f(\omega)$ is attained by substituting the estimates for the parameters in (5). The development of autoregressive spectral estimators has been summarized by Parzen (1983).

Obtaining confidence intervals for spectra is difficult in this case. Most techniques rely on unrealistic assumptions. An interesting fact about spectra of this form is that any spectral density can be approximated, arbitrarily close, by the spectrum of an AR process (see Shumway and Stoffer, 2017, §4.5). Unfortunately, the order of the model, p, may be very large to achieve a decent approximation.

2. RESOLUTION

The problem of frequency resolution was discussed in the literature in the latter half of the twentieth century and culminated in the early 1980s with the extensive work on resolution in Kay and Marple (1981) and Marple (1982). Further discussion of resolution may be found in texts such as Bloomfield (2004) and Brillinger (1981).

When considering resolution, the basic rule of thumb is that the achievable frequency resolution should be approximately (depending on the signal-to-noise ratio) the reciprocal of the observational time interval of the data. That is, if most of the signal energy is concentrated in an interval of Δt units of time, then the Fourier transform of the signal will have most of its energy concentrated in a frequency interval of $\Delta\omega$ cycles per unit of time, where

$$\Delta\omega \approx \frac{1}{\Delta t}\,.$$

This relationship is the basis of being able to distinguish between two narrow band signals (e.g., sinusoids). Two signals can be as close as $1/\Delta t$ apart before there is significant overlap in the transform and the separate peaks are no longer distinguishable.

To illustrate this idea for our problem, we generated a time series of length 2000 where

$$X_t = \begin{cases} X_{1t} & 1 \leq t \leq 1000\,, \\ X_{2t} & 1001 \leq t \leq 2000\,, \end{cases} \tag{6}$$

with

$$X_{1t} = 2\cos(2\pi\omega\, t)\cos(2\pi\delta\, t) + Z_{1t}\,,$$
$$X_{2t} = \cos(2\pi\omega\, t) + Z_{2t}\,,$$

and $\omega = 1/25$, $\delta = 1/150$, and Z_{it} for $i = 1, 2$ are independent i.i.d. standard normals. The difference between the two halves of the data is that X_{1t} is a modulated version of X_{2t}. Modulation is a common occurrence in many signal processing applications, e.g., EEG (see Novak, Lepicovska and Dostalek, 1992). In addition, note that

$$X_{1t} = \cos(2\pi[\omega + \delta]\, t) + \cos(2\pi[\omega - \delta]\, t) + Z_{1t}\,,$$

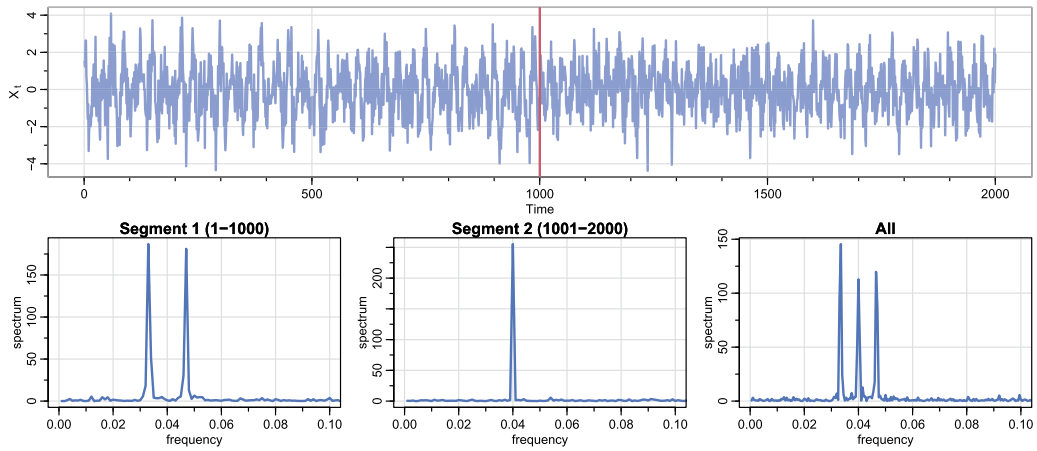

Figure 1. TOP: *Realization of* (6) *showing the true breakpoint as a solid vertical line.* BOTTOM: *Individual periodograms of the first and second halves, and all of the data shown on the top.*

so that X_{1t} is distinguishable at the sampling rate by twin peaks in the frequency domain. We note that in this example, X_t does not have a spectral density but there is a spectral distribution that is a mix of discrete and absolutely continuous components.

Figure 1 shows a realization of X_t on the top with the changepoint marked. The two signals are distinguishable at this sampling rate as pointed out by the individual periodograms of the first half and second half of the data (bottom of the figure).

3. SOME EXISTING METHODS

Let a time series $\{X_t;\, t = 1, \ldots, n\}$ consist of an unknown number of segments, m, and let ξ_j be the unknown location of the end of the jth segment, $j = 0, 1, \ldots, m$, with $\xi_0 = 0$ and $\xi_m = n$. Then conditional on m and $\boldsymbol{\xi} = (\xi_0, \ldots, \xi_m)'$, assume that the process $\{X_t\}$ is piecewise stationary. That is,

$$X_t = \sum_{j=1}^{m} X_{t,j}\, \delta_{t,j}\,, \tag{7}$$

where, for $j = 1, \ldots, m$, the processes $X_{t,j}$ have spectral density $f_j^\theta(\omega)$ that may depend on parameters θ, and $\delta_{t,j} = 1$ if $t \in [\xi_{j-1} + 1, \xi_j]$ and 0 otherwise.

Consider a realization $\boldsymbol{x} = (x_1, \ldots, x_n)$ from process (7), where the breakpoints are known. Let n_j be the number of observations in the jth segment. We assume that each n_j is large enough for the local Whittle likelihood (Whittle, 1957) to provide a good approximation to the likelihood. Given a partition of the time series $\boldsymbol{x}$, the jth segment consists of the observations $\boldsymbol{x}_j = \{x_t : \xi_{j-1}+1 \le t \le \xi_j\}$, $j = 1, \ldots, m$, with underlying spectral densities f_j^θ and periodograms I_j, evaluated at frequencies $\omega_{k_j} = k_j/n_j$, $0 \le k_j \le n_j - 1$. For a given partition $\boldsymbol{\xi}$, the approximate likelihood of the time series is given by

$$L(f_1^\theta, \ldots, f_m^\theta \mid \boldsymbol{x}, \boldsymbol{\xi}) \approx \prod_{j=1}^{m} (2\pi)^{-n_j/2} \times \prod_{k_j=0}^{n_j-1} \exp\Big\{-\frac{1}{2}\Big[\log f_j^\theta(\omega_{k_j}) + I_j(\omega_{k_j})/f_j^\theta(\omega_{k_j})\Big]\Big\}. \tag{8}$$

Note that in setting up the model, most items depend on the number of regimes, m. For ease, that dependence is understood and dropped from the notation.

3.1 AdaptSpec

The frequency domain approach used in Rosen, Wood and Stoffer (2012) is a Bayesian method that incorporates (8) with a linear smoothing spline prior on the $\log f_j^\theta(\omega)$ for $j = 1, \ldots, m$. In addition, a uniform prior is placed on the breakpoints, $\Pr(\xi_j = t \mid m) = 1/p_j$, for $j = 1, \ldots, m-1$, where p_j is the number of available locations for split point ξ_j, as is the prior on the number of segments, $\Pr(m = k) = 1/M$ for $k = 1, \ldots, M$ and M is some large but fixed number. The approach uses reversible jump Markov chain Monte Carlo (RJ-MCMC) methods to evaluate the posteriors. The technique is available in an R package called `BayesSpec`.

To keep the exposition simple, we concentrate on estimation of the spectrum in one particular segment j and drop it from the notation for now. Let $y_n(\omega_k) = \log I_n(\omega_k)$ and $g(\omega_k) = \log f(\omega_k)$, the representation (3) suggests the log-linear model

$$y_n(\omega_k) = g(\omega_k) + \epsilon_k, \tag{9}$$

where the ϵ_k's are independent, $\epsilon_k \sim \log(\chi^2_2/2)$. Note that $y_n(k/n) = y_n(1 - k/n)$. Representation (9) was used by

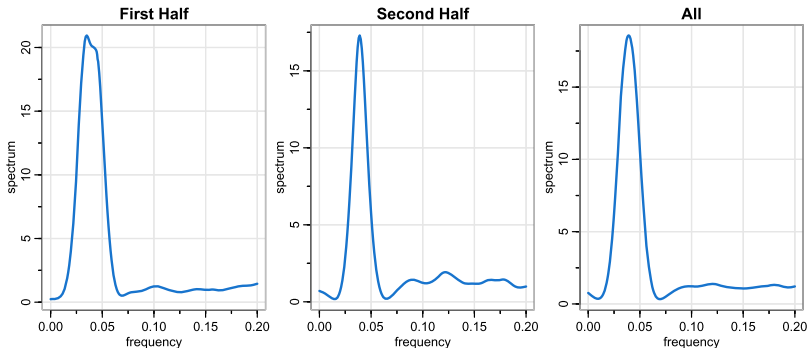

Figure 2. AdaptSpec does not find breakpoint in the full sequence specified in (6) and displayed in Figure 1. When the breakpoint is given to the algorithm, the figure displays individual AdaptSpec spectral estimates of X_{1t}, X_{2t}, and X_t.

a number of authors for nonparametric estimation of the log spectral density; e.g., Wahba (1980); Carter and Kohn (1997).

The prior on $g(\omega_k)$ follows Wahba (1990, p. 16), which expresses $g(\omega_k)$ as the sum of its linear and nonlinear components, so that

$$g(\omega_k) = \alpha_0 + \alpha_1\omega_k + h(\omega_k), \tag{10}$$

where $h(\omega_k)$ is the nonlinear component. A linear smoothing spline prior was put on the vector $\boldsymbol{h} = (h(\omega_0), \ldots, h(\omega_{n/2}))'$, i.e., $\boldsymbol{h} \sim N(\mathbf{0}, \tau^2\Omega)$, where τ^2 is a smoothing parameter with prior $U(0, c_{\tau^2})$, and $(\Omega)_{ij} = \min(\omega_i, \omega_j)$. The parameters α_0 and α_1 are the values of $g(\omega)$ and its first derivative at $\omega = 0$, respectively. The symmetry and periodicity of the spectral density mean that $(\partial g(\omega)/\partial\omega)|_{\omega=0} = 0$. Accordingly, α_1 is set to be identically zero, and the prior on α_0 is $N(0, \sigma_\alpha^2)$, for some large σ_α^2.

Prior distributions are placed on the number of segments, m, and the partition, $\boldsymbol{\xi}_m$. Without going into explicit details, the partitions given m are assumed a priori to be uniform over the number of time points in an interval (while guaranteeing a reasonable number of observations in a segment), and the number of breakpoints m is also assumed a priori to be uniform over M (large) values. Explicit details are given in Rosen, Wood and Stoffer (2009).

3.2 AutoParm

Although this method, which is described in Davis, Lee and Rodriguez-Yam (2006), is a time domain approach, it is well-known that AR models are dense in the space of bounded spectral densities and can thus be used in the frequency domain. The basic assumption is that, in each segment, the time series $\{X_{t,j}\}$ is a stationary AR(p_j), $j = 1, \ldots, m$. Then, minimum description length (MDL) as described in Rissanen (1983) is used to find the best combination of the number of segments, m, the breakpoints ξ_j (or segment sizes n_j), and the orders/estimates of the piecewise AR processes. The idea is to minimize the code length (CL) necessary to store the data (i.e., the amount of memory required to encode the data), which leads to a BIC-type criterion to find the model that minimizes the negative of the log-likelihood penalized by the number of segment and the number of parameters in each segment. Details may be found in Davis, Lee and Rodriguez-Yam (2006) and and the criterion is similar to the one for our proposed method, which is described in the Appendix.

Because of the Markov structure of AR models, the likelihood has a simple form; see Brockwell and Davis (2013, Prob. 8.7) for details. Fitting the model has to be done via numerical optimization, which is accomplished via a genetic algorithm (a derivative free smart search for minimization based on evolutionary biology concepts). Basic information on genetic algorithms may be obtained from Mathworks (2018), which provides informative videos. We discuss some details in the Appendix.

4. THE PROBLEM

The problem with the techniques presented in Section 3.1 and Section 3.2 is that they are not designed to detect narrowband frequency shifts and tend to over-smooth the spectral estimates. Local autoregressions (AutoParm) will rarely be able to resolve narrowband shifts unless the order is allowed to be very large, which is prohibitive in this computationally expensive scenario. Also, as packaged, AdaptSpec (via `BayesSpec`) puts a smooth prior on the spectra. It may, however, be possible to change AdaptSpec by putting a prior on the spectra that is not smooth (assuming one was a priori certain), perhaps using techniques in Bretthorst (2013, Ch. 2).

For the example in Section 2, the breakpoint was given to the methods, but the techniques cannot distinguish between the spectral distributions of each half of the data. Figure 2 shows the individual AdaptSpec estimates of X_{1t}, X_{2t} and of the entire sequence X_t. There is a slight difference between the spectral estimates of the first and second halves, but the technique indicates that the spectra of the combined halves

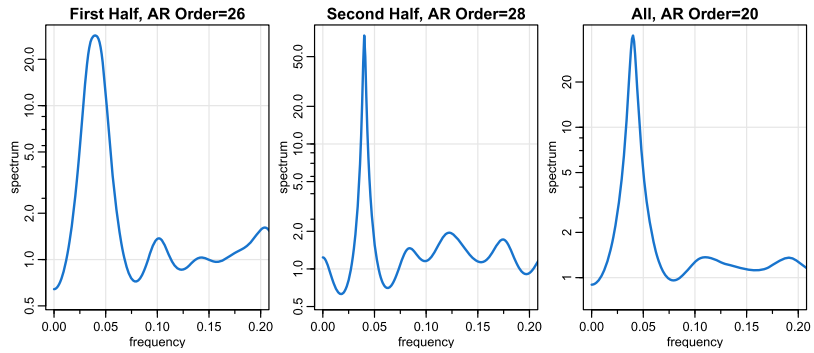

Figure 3. AutoParm does not find breakpoint in the full sequence specified in (6) and displayed in Figure 1. When the breakpoint is given to the algorithm, the figure displays individual AutoParm spectral estimates of X_{1t}, X_{2t}, and X_t.

is explaining the process better than dividing the process into two parts.

Also, Figure 3 shows the AutoParm estimates of X_{1t}, X_{2t} and of the entire sequence X_t. The results are similar to the AdaptSpec results, but notice that the AR spectra are perhaps the smoothest of the estimates. Again, there is a slight difference between the spectra of the two halves of the data, but the algorithm does not detect the difference.

5. AUTOSPEC

Because of the problem with some of the existing methods is resolution, we consider a method that allows for the possibility of narrowband spectral estimators in a fully nonparametric technique. First, consider a triangular (Bartlett) kernel, $\{h_\ell;\, \ell = 0, \pm1, \ldots, \pm b\}$ with $h_\ell \propto 1 - |\ell|/(b+1)$ such that $\sum h_\ell = 1$. Then, to nonparametrically evaluate the likelihood, (8), in each segment $j = 1, \ldots, m$, use

$$\hat{f}_j(\omega_{k_j}) = \sum_{\ell=-b_j}^{b_j} h_\ell \, I_j^{\text{tpr}}(\omega_{k_j+\ell}) \tag{11}$$

for $b_j = 0, 1, 2, \ldots$, where $B_j = 2b_j + 1$ are the bandwidths for each segment. Here, $I_j^{\text{tpr}}(\cdot)$ represents the periodogram of the fully cosine tapered data in segment j for $j = 1, \ldots, m$.

For example, if $\{x_{t,j}\}$ represents the data in segment $j = 1, \ldots, m$, and $t = 1, \ldots, n_j$, then they are preprocessed as $y_{t,j} = c_{t,j} x_{t,j}$ where $c_{t,j}$ is the cosine bell taper favored by Blackman and Tukey (1958),

$$c_{t,j} = .5\left[1 + \cos\left(\frac{2\pi(t - \overline{t}_j)}{n_j}\right)\right],$$

$\overline{t}_j = (n_j + 1)/2$. In this case, the periodogram is of the preprocessed data, $y_{t,j}$. Figure 4 shows an example of the Bartlett window with $b = 4$; the corresponding spectral window (see Shumway and Stoffer, 2017, §4.4) of the Bartlett kernel is not very good unless the data are tapered. The spectral window corresponding to the Bartlett kernel (known as the Fejér kernel) with tapering is also displayed in Figure 4.

The AutoSpec procedure is to use minimum description length (MDL) as derived in (15) and a Genetic Algorithm, which are discussed in the Appendix to fit the model. The MDL term is, of course, a BIC-type term that balances accuracy and complexity. In this case, the spectral estimates governed by (15) can choose very small bandwidths. This concept is contrary to the fitting of local autoregressions or smoothing splines as in AutoParm or AdaptSpec, respectively, where the local spectra are assumed to be smooth.

Figure 5 shows the results of running AutoSpec described on the simulated data X_t, as described in Section 2. As seen from the figure, the procedure is able to distinguish between the two processes (with a breakpoint at $t = 1002$).

6. ENSO AND CLIMATE CHANGE

El Niño and La Niña are the warm and cool phases of a recurring climate pattern across the tropical Pacific, also referred to as the El Niño–Southern Oscillation (ENSO). The pattern can shift back and forth irregularly every two to seven years, and each phase triggers predictable disruptions of temperature, precipitation, and winds.

The Southern Oscillation Index (SOI) measures changes in air pressure related to sea surface temperatures in the central Pacific Ocean. The central Pacific warms every two to seven years due to the El Niño effect, which has been blamed for various global extreme weather events. Very early on, Hansen and Lebedeff (1987) concluded that, "A strong warming trend between 1965 and 1980 raised the global mean temperature to the highest level in the period of instrumental records." These changes disrupt the large-scale air movements in the tropics, triggering a cascade of global side effects. More recently, Wang et al. (2017, 2019) concluded that, "Since the 1970s, El Niño has changed its origination from the eastern Pacific to the western Pacific, along with increased strong El Niño events due to a background warming in the western Pacific warm pool. This suggests

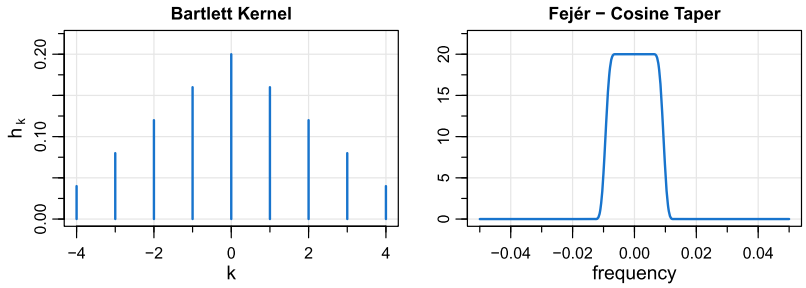

Figure 4. Example of the Bartlett kernel and the corresponding Fejér spectral window when a taper is applied.

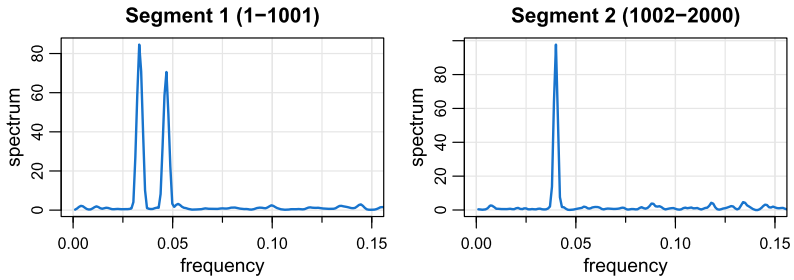

Figure 5. AutoSpec finds a breakpoint at $t = 1002$ for the simulated data specified in (6) and displayed in Figure 1. The estimated spectra for each data piece is exhibited and can be compared to estimates in Figure 1. The estimates are of the form given in (11) each with an optimal bandwidth (chosen by MDL) with $b = 3$.

the controlling factors that may lead to increased extreme El Niño events in the future. If the observed background changes continue under future anthropogenic forcing, more frequent extreme El Niño events will induce profound socioeconomic consequences."

Monthly values of the SOI are displayed in Figure 6 for years 1866–2019 (Climatic Research Unit, 2018); additionally, the data have been filtered to exhibit the ENSO cycle. Also shown in Figure 6 are the AutoParm results (vertical dashed line) and AdaptSpec results (vertical solid line) when applied to the SOI series. AutoParm prefers a breakpoint around 1920, whereas AdaptSpec is indicating there are no breakpoints because $\Pr(\text{break} \mid \text{data}) = .3$. However, assuming that there is one structural break, the posterior distribution of the breakpoints (with a vertical line at the mean) is displayed in the figure.

Figure 6 also shows the estimated spectra for each segment for the AutoParm technique; i.e., AR spectra. The gray swatch shows the 2–7 year cycle known ENSO cycle (see McPhaden, Zebiak and Glantz, 2006). Both methods indicate that in the second segment, the ENSO cycle is much more broad, including both slower and faster frequencies than the usual ENSO cycle. One thing that is clear from both methods is that the estimated spectra are too smooth (broad) to reveal if there has been a decisive frequency shift in the ENSO cycle. In addition, no experts have pointed to a suspected shift around 1920 so it is not clear why there would be a breakpoint then. It is possible that the air pressure data collection or definition of SOI changed around that time.

Finally, Figure 7 displays the analysis based on AutoSpec, which finds three regimes with breakpoints at mid-1880 and the spring of of 1975. The bandwidths, as defined below (11), of the selected Bartlett windows in each segment had $B_j = 3, 15, 9$, respectively. The bottom row of the figure shows the estimated spectra for each segment. The data in each segment are standardized so that the spectra are comparable. It is difficult to interpret the first breakpoint because there is not much discussion about ENSO during that period. However, as previously mentioned, the second breakpoint has been discussed in the climate literature (Wang et al., 2019).

The difference between the segment 2 and 3 spectra is that the segment 2 spectrum has more power at the lower frequencies, whereas the segment 3 spectrum has less power there, and consequently more power at higher frequencies.

7. AN AWAY GAME

In this section, we use a simulation to see how AutoSpec competes with AdaptSpec and AutoParm while playing on AutoParm's field; i.e., when the true models are autoregressions. In this case, the family of densities is known and once breakpoints are located, AutoParm only has to estimate a few parameters to estimate the spectral density. As previously mentioned, AdaptSpec will use smoothing splines to

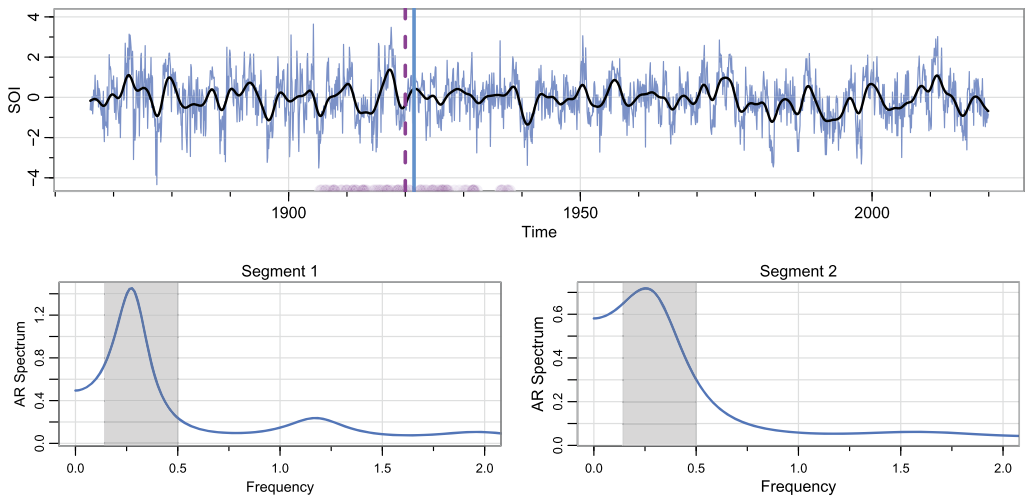

Figure 6. TOP: *Monthly values of SOI for years 1866–2019 with breakpoints (vertical lines) determined by AutoParm (|) and by AdaptSpec (¦). The solid smooth line is the filtered series that exhibits the ENSO cycle. AutoParm finds a breakpoint in June 1921 whereas AdpatSpec finds one at January 1920. For AdaptSpect, however,* $\Pr(a\ break \mid data) = .3$ *indicates there is probably not a breakpoint. The dots at the bottom of the graph indicate the distribution of breakpoints found by AdaptSpec when one is found; the line shows the mean location.* BOTTOM: *The segmented AR spectral estimates with a breakpoint at 1921. The gray swatch shows the 2–7 year cycle known ENSO cycle range.*

estimate spectra and AutoSpec will use kernel density estimation.

Here, we take an example from Davis, Lee and Rodriguez-Yam (2006, §4.1) but slightly simplified. The time series is given by two AR(2)s,

(12)
$$X_t = \begin{cases} 1.69X_{t-1} - .81X_{t-2} + Z_t & \text{for } 1 \le t \le 500\,, \\ 1.32X_{t-1} - .81X_{t-2} + Z_t & \text{for } 501 \le t \le 1000\,, \end{cases}$$

where the Z_t are i.i.d. standard normals (we note that there is no overlap in the generation of the data, we simply generated two AR(2)s, each of length 500, and put them together). Both segments have complex roots; the first has a somewhat broad spectrum with a peak at approximately $\omega = .055$, while the second segment is more narrow band with a peak at about twice the frequency of the first segment. The simulated data are shown in Figure 8. We note that the location of the breakpoint is obvious, as is the fact that the data in first segment have a lower frequency of oscillation than that of the second segment.

As expected, AutoParm does very well, finding a breakpoint at $t = 498$, choosing an AR(2) for each segment, and with parameter estimates very near the truth (the standard error of any AR parameter estimate in this game is about $[1 - (-.81)^2]/\sqrt{500} \approx .015$, so we expect the estimates to be very close). The resulting spectral estimates via (5) are shown in Figure 9 and are, as expected, very close to the true spectra, which are also displayed in the figure.

AdaptSpec does a good job capturing the dynamics of the example. The posterior probability of one breakpoint is about .46, whereas the probabilities of zero or two breakpoints are about .32 and .22, respectively. Figure 10 shows the distribution of breakpoint locations for 500 runs. Although the distribution is wide, model averaging puts the mean location at $t = 530$ (the median at $t = 526$). The estimated spectra (posterior means) are displayed in Figure 11 and compared to the true values. The estimated spectra are close to the truth but with the peak of the second segment being slightly shifted toward the lower frequencies.

Finally, Figure 12 shows the results of AutoSpec, which finds one breakpoint at $t = 499$. The estimated spectra are based on bandwidths $B_1 = 23$ and $B_2 = 5$, respectively. The results are similar to the AdaptSpec results wherein the first segment estimate is close to the truth, but the second segment has a peak that is slightly shifted toward a lower frequency.

8. CONCLUDING REMARKS

There are a number of directions to take this work. It seems as though using a nonparametric approach via MDL in conjunction with the Whittle likelihood is a viable solution to the problem of detecting small frequency shifts in time series. It would be worthwhile to explore the use of more general kernels in addition to the Bartlett kernel. Also, it may be important to restrict the Whittle likelihood approximation to frequency regions of interest. For example, in the ENSO cycle problem, the interest is in frequencies well below the annual cycle. Hence, we may restrict the second product in (8) to be over frequencies in some small band of

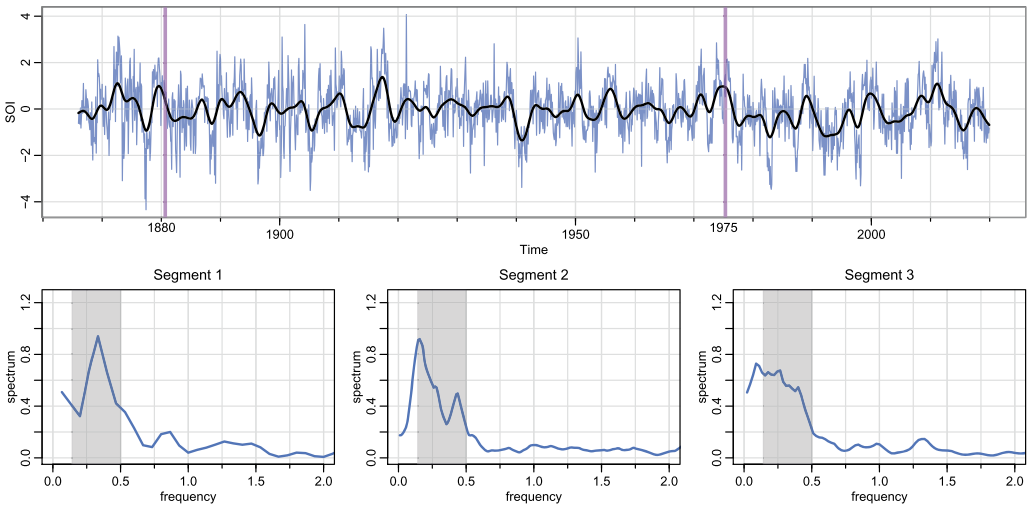

Figure 7. TOP: *Monthly values of SOI for years 1866–2019 with breakpoints (vertical lines) determined by AutoSpec. The solid smooth line is the filtered series that exhibits the ENSO cycle. AutoSpec finds a breakpoint at the beginning of 1880 and the beginning of 1975.* BOTTOM: *The segmented AutoSpec spectral estimates for each segment. The gray swatch shows the 2–7 year cycle known ENSO cycle range.*

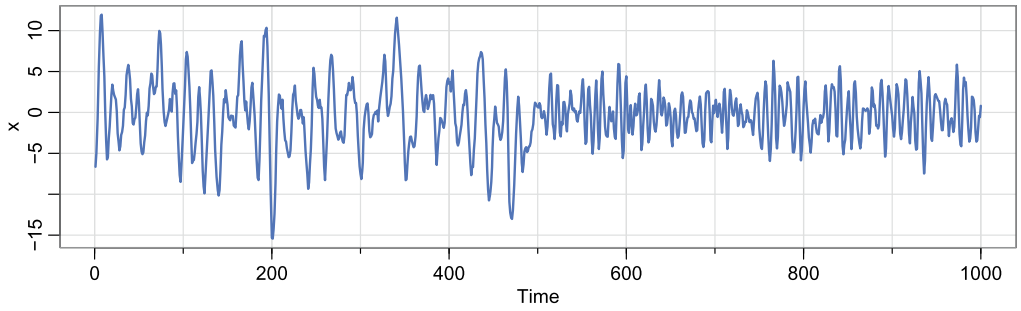

Figure 8. Simulated AR(2) data with an obvious breakpoint at $t = 500$ and an obvious increase in frequency (in fact it is a two-fold increase) in the second half of the data.

frequencies of interest. In addition, as discussed in Bloomfield (2004, Ch. 9), tapering increases the variance of a spectral estimate and only tapering the 10% or 20% ends of the data is a good tradeoff to protect against leakage with only modest increases in variance. Consequently, the technique might try various (discrete) levels of tapering.

Another approach in the univariate case would be to use MDL along with a wavelet spectrum such as discussed in Chiann and Morettin (1998). Finally, in many cases, the interest is in nonstationary multivariate time series. For example, there are various measurement series for the sea surface temperature of the Pacific Ocean; e.g., see L'Heureux, Tippett and Barnston (2015). Similarly, when monitoring EEG signals, several locations on the scalp are typically used; e.g., see Aksenova, Volkovych and Villa (2007). Although it has not yet been tried, it should be easy to extend the fully nonparametric approach of Section 5 to the multivariate case.

APPENDIX: AUTOSPEC MDL AND OPTIMIZATION

In this section, we derive a minimum description length (MDL) criterion for choosing the best fitting model from the AUTOSPEC procedure. The "best" model is defined as the one that enables the best compression of the observed series $\boldsymbol{x} = \{x_1, \ldots, x_n\}$.

There are various versions of the minimum description length principle as put forth by Rissanen (Rissanen, 1978, 1989) and the version adopted here is a two-part code. In this case, the first part, denoted by $\mathcal{C}$, represents the *complexity* of the fitted model, and the second part, denoted by $\mathcal{A}$, which represents the *accuracy* of the fitted model. The idea of the minimum description length principle is to find the best pair of $\mathcal{C}$ and $\mathcal{A}$ so that via encoding (or compressing) $\mathcal{C}$ and $\mathcal{A}$, $\boldsymbol{x}$ can be transmitted (or stored) with the least amount of codelength (or memory). To quantify this

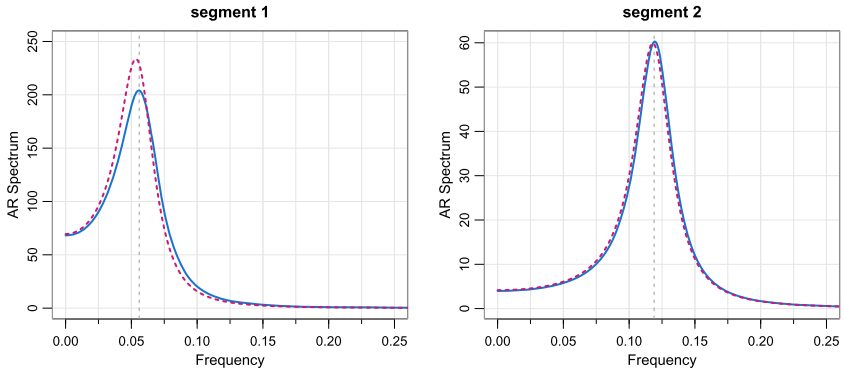

Figure 9. Estimated spectra for Figure 8 data using AutoParm (solid line) compared to the true spectra (dashed line) for each of the two segments. AutoParm finds a breakpoint at $t = 498$. In this and similar figures, the vertical dashed line indicates the locations of the peak of each true spectrum (at approximate frequencies of .055 and .11).

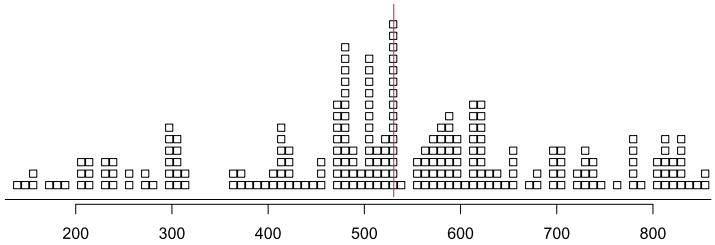

Figure 10. AdpatSpec posterior distribution of a breakpoint location for 500 runs for Figure 8 data in. The solid vertical line shows the posterior mean location of a breakpoint at $t = 530$.

idea, let $\mathrm{CL}(\cdot)$ denote the codelength of an object based on a given model. Then we have the decomposition

$$\mathrm{CL}(\boldsymbol{x}) = \mathrm{CL}(\mathcal{C}) + \mathrm{CL}(\mathcal{A} \mid \mathcal{C}) \tag{13}$$

for the data $\boldsymbol{x}$. This approach leads to familiar concepts such as AIC and BIC where accuracy is measured by the negative of the log-likelihood evaluated at the estimated parameters, and the complexity is a penalty based on the number of parameters in the model and possibly the sample size.

For the complexity term in (13), we must consider the various parameters of the model, which includes the number of segments, m, the change points, $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)$, and the individual bandwidths in each segment, $B_1, \ldots, B_m$ as defined in (11). In this case we have

$$\begin{aligned} \mathrm{CL}(\mathcal{C}) = \mathrm{CL}(m) + \mathrm{CL}(\xi_1, \ldots, \xi_m \mid m) \\ + \mathrm{CL}(B_1, \ldots, B_m \mid m, \boldsymbol{\xi})\,. \end{aligned} \tag{14}$$

To evaluate (14), the codelength for an integer m is $\log_2 m$ bits. For the second term, we note that knowledge of the breakpoints, ξ_j, is equivalent to knowledge of the number of observations in segment j, namely n_j. Noting that the n_j are bounded by the number of observations, n, we have a bound, $\mathrm{CL}(n_j) = \log_2 n$ so that

$$\mathrm{CL}(\xi_1, \ldots, \xi_m \mid m) = \mathrm{CL}(n_1, \ldots, n_m \mid m) = m \log_2 n\,.$$

Each bandwidth value will cost about $\log_2 B_j$ bits. In addition, the bandwidth in each segment $j = 1, \ldots, m$ is determined by maximizing the likelihood based on the segment data of n_j observations. For this, we can use a result of Rissanen that states a maximum likelihood estimate of a parameter computed from n_j observations can be effectively encoded with $\frac{1}{2}\log_2 n_j$ bits, making the third term

$$\mathrm{CL}(B_1, \ldots, B_m \mid m, \boldsymbol{\xi}) = \tfrac{1}{2}\sum_{j=1}^{m} \log_2(n_j\, B_j^2)\,.$$

For the second term in (13), it is shown in Rissanen (1989) that the codelength of the accuracy term, $\mathcal{A}$, is the negative of the $\log_2$ likelihood of the fitted model $\mathcal{C}$. In our case, we use the Whittle likelihood approximation; see (8).

Combining the results and working with natural log instead of base 2, we obtain an approximation to the MDL of the model,

$$\begin{aligned} \mathrm{MDL} = \log m + m \log n + \tfrac{1}{2}\sum_{j=1}^{m} \log(n_j\, B_j^2) \\ + \sum_{j=1}^{m}\Big\{ \frac{n_j}{2}\log(2\pi) + \tfrac{1}{2}\sum_{k_j=0}^{n_j-1}\Big[\log \hat{f}_j(\omega_{k_j}) + \frac{I_j(\omega_{k_j})}{\hat{f}_j(\omega_{k_j})}\Big]\Big\}\,. \end{aligned} \tag{15}$$

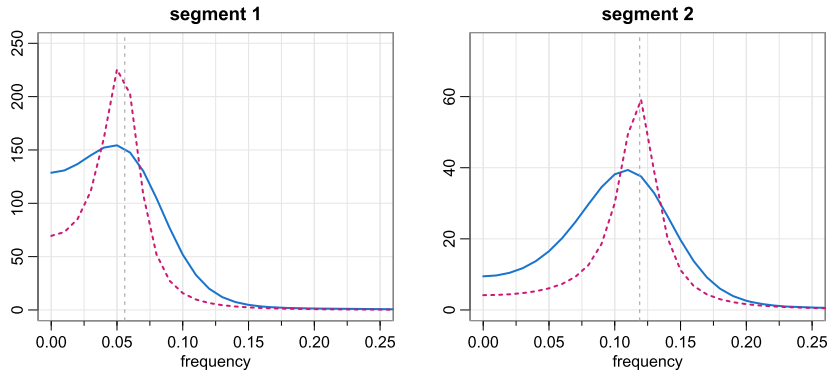

Figure 11. Estimated spectra (posterior means) for Figure 8 data using AdaptSpec (solid line) compared to the true spectra (dashed line) for each of the two segments. AdaptSpec finds a breakpoint (posterior mean) at $t = 530$.

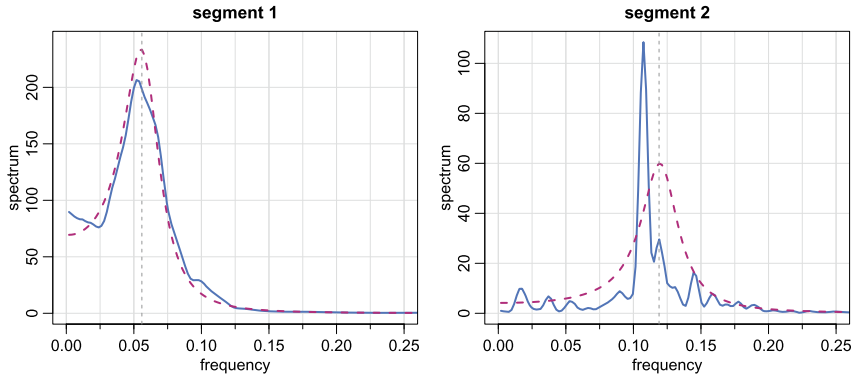

Figure 12. Estimated spectra for Figure 8 data using AutoSpec (solid line) compared to the true spectra (dashed line) for each of the two segments. AutoSpec finds a breakpoint at $t = 499$.

Because the search space is enormous, optimization is a nontrivial task. As in Aue et al. (2014) or Davis, Lee and Rodriguez-Yam (2006) and many others, we use a genetic algorithm (GA) to effectively tackle the problem. Briefly, genetic algorithms are a class of iterative optimization methods that use the principles of evolutionary biology. The algorithm typically begins with some initial randomly chosen population and each generation afterwards produces an offspring population using genetic operators. Genetic operators include selection, crossover, and mutation, which are based on the principle of natural selection to find the best solution while using the principle of diversity to avoid convergence to a local minima. *Selection* operators are used to select which offspring survive to the next generation. It is crucial that the fitter individuals are not kicked out of the population, while at the same time diversity should be maintained in the population. Truncation is the simplest selection operator which simply chooses the fittest individuals from the parent and offspring population. Tournament selection is another selection operator that randomly sorts the individuals into blocks and chooses the best individual from each block. In Age-Based Selection, there is not a notion of a fitness but it is based on the premise that each individual is allowed in the population for a finite generation where it is allowed to reproduce and then it is kicked out of the population no matter how fit. In Fitness Based Selection, the children tend to replace the least fit individuals in the population. The selection of the least fit individuals may be done using a variation of any of the selection policies described before; e.g., tournament selection. *Recombination* operators, often referred to as *crossover*, are used to mix two or more parents to produce similar, but slightly different offspring. Most crossover operators convert the individual into binary representation to perform the operations. One-point crossover crosses the binary digits at some crossover point of two parents to create two new individuals. *Mutation* operators are used to further preserve the diversity of a population to ensure convergence to an optimum. A simple type of mutation involves the addition of a number chosen from a standard normal. Another type of mutation known as flip bit also performs operations on the binary representation of a number where each bit in the representation has some probability of being mutated. A flow chart of a genetic algorithm is shown in Figure 13.

Received 25 January 2021

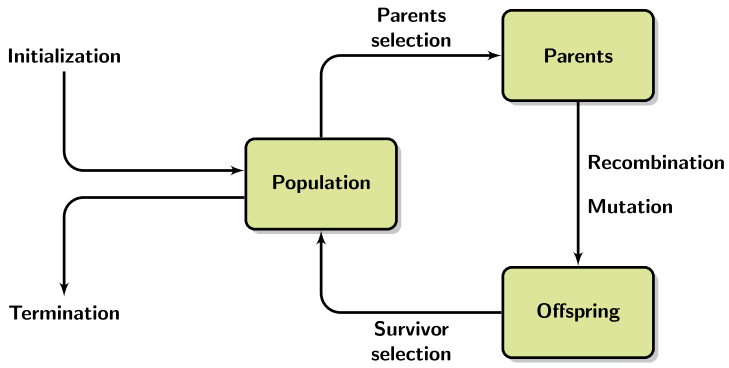

Figure 13. Flow chart of a Genetic Algorithm. The algorithm typically begins with an initial randomly chosen population. Afterwards, each generation produces an offspring population using genetic operators. Selection operators are used to select which offspring survive to the next generation. Recombination operators, often referred to as Crossover are used to mix two or more parents to produce similar, but slightly different offspring. Mutation operators are used to further preserve the diversity of a population to ensure convergence to an optimum.

REFERENCES

Adak, S. (1998). Time-dependent Spectral Analysis of Nonstationary Time Series. *Journal of the American Statistical Association* **93** 1488–1501. MR1666643

Aksenova, T., Volkovych, V. and Villa, A. (2007). Detection of spectral instability in EEG recordings during the preictal period. *Journal of Neural Engineering* **4** 173–178.

Aminikhanghahi, S. and Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and Information Systems* **51** 339–367.

An, S.-I. and Wang, B. (2000). Interdecadal change of the structure of the ENSO mode and its impact on the ENSO frequency. *Journal of Climate* **13** 2044–2055.

Aue, A., Cheung, R. C. Y., Lee, T. C. M. and Zhong, M. (2014). Segmented Model Selection in Quantile Regression Using the Minimum Description Length Principle. *Journal of the American Statistical Association* **109** 1241–1256. MR3265694

Blackman, R. B. and Tukey, J. W. (1958). The measurement of power spectra from the point of view of communications engineering. *Bell System Technical Journal* **37** 185–282. MR0102897

Bloomfield, P. (2004). *Fourier Analysis of Time Series: An Introduction*. John Wiley & Sons. MR0654511

Bretthorst, G. L. (2013). *Bayesian Spectrum Analysis and Parameter Estimation* **48**. Springer Science & Business Media. MR0971023

Brillinger, D. R. (1981). *Time series: Data Analysis and Theory* **36**. SIAM. MR1853554

Brockwell, P. J. and Davis, R. A. (2013). *Time Series: Theory and Methods*, 2nd ed. Springer Science & Business Media.

Carter, C. K. and Kohn, R. (1997). Semiparametric Bayesian inference for time series with mixed spectra. *Journal of the Royal Staistical Society B* **59** 255–268. MR1436567

Chiann, C. and Morettin, P. A. (1998). A wavelet analysis for time series. *Journal of Nonparametric Statistics* **10** 1–46. MR1708134

Climatic Research Unit (2018). Southern Oscillation Index (SOI). https://crudata.uea.ac.uk/cru/data/soi/.

Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. (2006). Structural breaks estimation for nonstationary time series models. *Journal of the American Statistical Association* **101** 223–239. MR2268041

Hannan, E. and Rissanen, J. (1988). The width of a spectral window. *Journal of Applied Probability* 301–307. MR0974589

Hansen, J. and Lebedeff, S. (1987). Global trends of measured surface air temperature. *Journal of Geophysical Research: Atmospheres* **92** 13345–13372.

Kay, S. M. and Marple, S. L. (1981). Spectrum analysis—a modern perspective. *Proceedings of the IEEE* **69** 1380–1419.

L'Heureux, M. L., Tippett, M. K. and Barnston, A. G. (2015). Characterizing ENSO coupled variability and its impact on North American seasonal precipitation and temperature. *Journal of Climate* **28** 4231–4245.

Marple, S. (1982). Frequency resolution of Fourier and maximum entropy spectral estimates. *Geophysics* **47** 1303–1307.

Mathworks (2018). Genetic Algorithm – MATLAB The Mathworks, Inc., Natick, Massachusetts. https://www.mathworks.com/discovery/genetic-algorithm.html.

McPhaden, M. J., Zebiak, S. E. and Glantz, M. H. (2006). ENSO as an integrating concept in earth science. *Science* **314** 1740–1745.

Novak, P., Lepicovska, V. and Dostalek, C. (1992). Periodic amplitude modulation of EEG. *Neuroscience Letters* **136** 213–215.

Ombao, H., Raz, J., Von Sachs, R. and Malow, B. (2001). Automatic statistical analysis of bivariate nonstationary time series. *Journal of the American Statistical Association* **96** 543–560. MR1946424

Parzen, E. (1983). Autoregressive Spectral Estimation. *Handbook of Statistics* **3** 221–247. MR0749788

Rissanen, J. (1978). Modeling by shortest data description. *Automatica* **14** 465–471.

Rissanen, J. (1983). A universal prior for integers and estimation by minimum description length. *The Annals of Statistics* 416–431. MR0696056

Rissanen, J. (1989). Stochastic Complexity in Statistical Inquiry. *World Scientific Series in Computer Science* **15** 79–93. MR1082556

Rosen, O., Wood, S. and Stoffer, D. S. (2009). Local spectral analysis via a Bayesian mixture of smoothing splines. *Journal of the American Statistical Association* **104** 249–262. MR2504376

Rosen, O., Wood, S. and Stoffer, D. S. (2012). AdaptSPEC: Adaptive spectral estimation for nonstationary time series. *Journal of the American Statistical Association* **107** 1575–1589. MR3036417

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*, 4th ed. Springer, New York. MR3642322

Wahba, G. (1980). Automatic smoothing of the log periodogram. *Journal of the American Statistical Association* **75** 122–132.

Wahba, G. (1990). *Spline Models for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics* **59**. SIAM, Philadelphia. MR1045442

Wang, G., Cai, W., Gan, B., Wu, L., Santoso, A., Lin, X., Chen, Z. and McPhaden, M. J. (2017). Continued increase of extreme El

Niño frequency long after 1.5 °C warming stabilization. *Nature Climate Change* **7** 568.
Wang, B., Luo, X., Yang, Y.-M., Sun, W., Cane, M. A., Cai, W., Yeh, S.-W. and Liu, J. (2019). Historical change of El Niño properties sheds light on future changes of extreme El Niño. *Proceedings of the National Academy of Sciences* **116** 22512–22517.
Whittle, P. (1957). Curve and periodogram smoothing. *Journal of the Royal Statistical Society B* **19** 38–47. MR0092331

David S. Stoffer
Department of Statistics
University of Pittsburgh
United States
E-mail address: stoffer@pitt.edu
url: www.stat.pitt.edu/stoffer