The Spectral Analysis of Time Series

by
John Reid

The Spectral Analysis of Time Series

Spectral Analysis

The term *spectrum* refers to the distribution of energy or power with frequency or wavelength. It began when Newton passed a beam of white light through a glass prism and saw how it split into its component colours. This happened because different colours have different velocities in glass and so were refracted at different angles. The rainbow effect produced was termed the spectrum after the Latin word for ghost. The splitting of light into its component parts became a very powerful technique, particularly in astronomy where the spectral character of an object is pretty much the only information available.

As every blacksmith or potter knows, the temperature of a hot object can be judged by its colour. As an object is heated it first becomes dull red, then bright red, then orange, then yellow then white, the colour of the Sun's surface. Apart from temperature, the spectrum can also reveal great detail about the chemical composition of the object, but that need not concern us here. The "Quantum Revolution" which overturned classical physics a century ago started with the consideration of the spectra of hot objects as discussed in the previous section.

Clearly the break-down of energy into components associated with different scales is of major interest in the analysis of a physical system because it provides insights into underlying physical processes. In the analogue world, various gadgets such as prisms, diffraction gratings and electronic circuitry are used to perform this task. In this digital age, numerical methods are appropriate.

Scientific measurements of real world quantities do not commonly come singly or in small groups but in very large data sets which may contain, quite literally, millions of values. This is particularly so since the advent of electronic storage methods. One particularly common form that data takes is the *time series*, which a sequence of measurements taken at equal intervals of time or averaged over equal intervals of time. An example would be the Weather Bureau's three hourly readings of the temperature at a particular location.

Drawing conclusions from time series is more complicated than one might expect. Most people like to think they have an intuitive grasp of time series, particularly with regard to the weather: "It was much hotter when I was a kid" and so on. Often these intuitions are quite wrong even when made by supposedly skilled observers. What is needed is a formal discipline for dealing with time series in the Popperian manner. Such a discipline evolved from signal processing theory and Econometrics.

The spectrum is an essential concept in dealing with time series. Rather than the power spectrum of the analogue world it is the variance spectrum which is used. Variance, the mean square of a set of numbers, is the digital analogue of power or energy. Hence it is the *variance spectrum* which must be calculated or estimated. Like the power spectrum, it can be a powerful tool for revealing underlying physical processes. The method is known as *spectral analysis*.

Spectral analysis co-evolved with the rapid development of digital computers in a number of different disciplines. As a consequence, some myths and misconceptions crept in and have become established as Baconian Idols – for example, the belief that the periodogram is not a consistent statistical estimator and must always be smoothed by "windowing" in order to be useful.

Here we develop a rigorous approach to the spectral analysis of time series, one which accommodates conventional methods of statistical inference.

Statistical Inference

Popper's principles are well served by methods of statistical inference developed by Fisher (1934) and others in the early part of the 20th century. These methods formalize and quantify the idea of falsifiability by introducing the concept of the Null Hypothesis, that is, an hypothesis which has been set up to be deliberately falsified using the methods of mathematical statistics. This technique is known as hypothesis testing.

Here is a very simple example:

- **Experiment**: A man tosses a coin 4 times and gets Heads every time. Is there something strange going on, such as a biased coin being tossed?
- **Null Hypothesis**: There is nothing strange going on; it is just due to chance
- **Assumed distribution**: Probability of heads=0.5, probability of tails=0.5.
- **Calculation**: Probability of outcome (4 successive heads) under Null Hypothesis ("It's just chance") $= 0.5^4 = 1/16 = .0625 = 6.25\%$.
- **Conclusion**: There is 1 chance in 16 or a probability of six percent that this could happen by chance so perhaps it did. We cannot reject the Null Hypothesis.

Now consider a second example involving ten tosses rather than four:

- **Experiment**: A man tosses a coin 10 times and gets Heads every time. Is there something strange going on, such as a biased coin being tossed?
- **Null Hypothesis**: There is nothing strange going on; it is just due to chance
- **Assumed distribution**: Probability of heads=0.5, probability of tails=0.5.
- **Calculation**: Probability of outcome (10 successive heads) under the Null Hypothesis $= 0.5^{10} = 1/1024 = .00975$ which is less than one percent.
- **Conclusion**: There is less than 1 chance in a 100 that this could happen by chance so it looks like something strange is going on and that the assumed distribution is wrong. We can reject the Null Hypothesis.

As with all statistical tests, the devil is in the detail. This must be the *only* set of 10 tosses. Obviously if we keep on tossing the coin thousands of times until, sooner or later we do get 10 Heads in a row, then we would have biased our sample to favour our conclusion, and the conclusion would not be valid. Selecting data to fit the hypothesis being tested in this way is termed "cherry picking". We have to trust the experimenters not to cherry-pick.

Note how well Fisher's hypothesis-testing methodology fits Popper's Principles. We don't actually *prove* that the coin is biased; we cannot even be sure that the coin *is* biased – we may be dealing with sleight of hand. We can only say that something highly unlikely has happened which is worthy of further investigation. The only difference between these examples and a real scientific experiment lies in the simplicity of the assumed distribution.

In these examples we make two important assumptions. We assume

1. that the outcome of a normal coin toss is random in that we cannot predict the outcome with certainty, and
2. that successive coin tosses are independent, i.e. that the probability of a coin toss being a Head is 0.5 whatever the outcomes of the preceding tosses may have been.

More precisely, the null hypothesis was that the coin tossing was an independent, random process from a binomial distribution with equal probabilities.

These were examples of statistical inference using "frequentist" statistics and hypothesis testing as a practical example of Popper's ideas. Other statistical methods include "Bayesian" statistics, which have more to do with optimizing strategies when applying established theories to real world problems.

In discussing these ideas, we must distinguish between a *sample* and the *population* or *ensemble* from which it is drawn. The terms population and ensemble designate the same thing: the former comes from the biological sciences, the latter from signal processing.

An ensemble is a set of parallel universes, each with identical statistical properties but differing in random detail. Think of an electronic gadget, a "noise generator", sitting on a laboratory bench making a hissing noise. Now imagine thousands of identical gadgets sitting side by side all making a hissing noise. The hissing noises may all sound the same, but if you make graphs of them with an oscilloscope or a chart recorder, the graphs would all be different; it is only *statistical* properties that we hear.

The average output of all the electronic gadgets at a given time is called the "ensemble", the "expectation" average or "the average across the ensemble". Generally we are not dealing with electronic gadgets nor are we dealing with multiple universes, but we still retain the concept of the ensemble average; it is a useful idea. If x_t is the value of the time series at time t then its ensemble average is written $\mathbb{E}(x_t)$. Note that the ensemble average has the subscript t indicating that it depends on time. We often assume that all ensemble averages of a particular quantity are the same whatever the time, i.e. that

$$\mathbb{E}(x_1) = \mathbb{E}(x_2) = \cdots = \text{constant} \tag{1}$$

If ensemble averages all satisfy this assumption, the time series is termed *stationary*. That is the definition of stationary. Stationarity is an assumption. It is not possible to prove that a time series is stationary, but hypothesis testing can be used to prove that it is not.

In reality we must deal with a *sample* time series. Generally there is only one time series, the sample time series, $x_1, x_2, x_3 ... x_N$, and there is only one type of sample average that can be calculated, the time average. This is written $< x >$ or $\bar{x}$, which is the sum of all the values of the time series divided by the total number of values, i.e. $(x_1 + x_2 + x_3 + ... + x_N)/N$.

For a stationary time series, time averages become closer and closer to ensemble averages as the number of quantities averaged (the "sample size") gets larger. Although intuitively obvious, this is by no means trivial. It is known as the Ergodic Theorem. It forms the basis of statistical inference about time series. Excepting for artificially generated time series, we can never *know* the values of ensemble quantities; we can only estimate them from the data. For example, the sample mean is an estimate of the ensemble mean because of the Ergodic Theorem, i.e. $< x >$ is an estimator of $\mathrm{E}(x)$. Much of statistical theory is concerned with knowing how "good" such an estimate is. When the t-test is used to place 95% confidence limits on a value, we are implying that there is only 1 chance in 20 that the null hypothesis is false, the null hypothesis being that the ensemble mean lies inside the range of the confidence limits.

When estimating ensemble values from sample quantities, there are two important concepts to be considered. They are whether the estimate is *biased* and whether the estimate is *consistent*. An estimate is biased when its expectation value differs from the ensemble value it is supposed to estimate. A consistent estimate is one which converges as the sample size gets larger. This is true if and only if the variance of the estimate tends to zero as the sample size increases.

For notational simplicity only simple averages have been discussed above, and now we shall assume $\mathbb{E}(x_t) \equiv 0$. The same arguments apply to averages of combinations of values. Of particular importance is the product of values which are separated in time by a fixed amount, i.e. $\mathbb{E}(x_t x_{t-L})$

where L is the lag in time between two values of the time series. Its ensemble average, Φ_L, is known as the *covariance* at lag L. When $L = 0$ it is called the *variance*.

The covariance function, Φ, completely summarizes the important statistical properties of a stationary time series. It is a "population parameter", i.e. it is a property of the ensemble. Like the ensemble mean, it can be estimated from the sample data, but placing confidence limits on it is more complicated.

The covariance function does not display well. Covariance functions from entirely different processes can look remarkably similar. Fortunately there is a way around this. The Fourier Transform of the covariance function $S(f)$ displays very well indeed. It shows how variance is distributed with frequency. It has visible features such as peaks, troughs and power law slopes which correspond to underlying physical processes such as blurring, resonance and integration. It is called the *variance density spectral estimate*, the *variance density spectrum*, sometimes the *power spectrum* but usually just the *spectrum*.

Like the covariance function, the ensemble spectrum comprises a set of population parameters which are estimated from the data. There are various methods for estimating spectra which are discussed further below.

Variance density spectra defined in this way are the numerical analogues of the energy density spectra of physics, such as black body radiation spectra or turbulent energy wavenumber spectra mentioned above. Like them, it is crucially dependent on underlying physical processes and so is a powerful analytical tool for exploring such processes.

The Periodogram

The squared modulus of the Fourier Transform of the time series is the simplest, most straightforward method of estimating its spectrum. This estimate is called the *periodogram* and is defined by equations (20) and (24).

Periodogram spectra tend to be noisy compared with other methods of spectral estimation. In the 1950s, when digital computers first started to be used for the spectral analysis of time series, it was widely believed that this was because the periodogram was not a consistent estimator of the ensemble variance density, see, for example, Hannan (1960). This is a fallacy. The length, N, of a sample time series is *not* the sample size. The sample size is 1, and each ordinate has an F-distribution with 2 and $N-1$ degrees of freedom. In this context, the length, N, is the dimension and not the sample size in a statistical sense. A proof that the periodogram is a consistent estimator for a well-behaved (ARMA) process is given below.

This widespread misconception has led to a variety of methods for smoothing or 'windowing' the data, such as the 'Hanning Window' and so on, as in Blackman and Tukey (1958), for example. These may sometimes be convenient for display purposes but are unnecessary and frequently misleading. The noisiness of the periodogram spectral estimate can be attributed to the small number of degrees of freedom, namely 2, associated with each ordinate; the N degrees of freedom of the sample time series are spread out among $N/2$ spectral values. Despite this noisiness, useful features of the periodogram can still be identified.

A narrow, statistically significant peak at frequency f_0 in the periodogram indicates cyclic behaviour in the underlying time series at that frequency. This in turn implies that a deterministic cycle is present in the data, perhaps of astronomical origin. A broad peak in the spectral estimate implies resonant, stochastic behaviour. A flat or 'white' spectrum implies that the time series is unselfcorrelated. So et al. (1999) have shown that the unwindowed, unsmoothed periodogram is the most sensitive method of detecting sinusoids in noisy time series.

Widespread features of time series of physical quantities are power law spectra which are a

consequence of natural processes such as integration, i.e.

$$S = Af^{\nu} \tag{2}$$

where S is the variance density at frequency f and A and ν are constants. For this reason it is convenient to plot $\log(S)$ against $\log(f)$ when displaying spectra. Thus (1) becomes

$$\log(S) = \nu \log(f) + log(A) \tag{3}$$

so that power law relationships appear as straight lines when spectra are plotted using logarithmic scales. Log-log spectra have the additional advantage of conveniently displaying behaviour over a wide range of scales and diminishing the apparent noisiness of the periodogram. Temperature time series often have a power law spectrum with $\nu = -2$ because temperature is commonly dependent on the integration of heat or other forms of energy as discussed in the section on spurious regression and climate.

Of interest is the spectrum obtained by (2) when $\nu = 0$ and $\hat{S}$ =const. In this event, by analogy with light, the spectrum is termed 'white'. It corresponds to a population covariance function which has a positive value at zero lag and is zero elsewhere, i.e. the time series is unselfcorrelated. While this can never be demonstrated conclusively in practice for any particular time series, it is a powerful null hypothesis. If the sample time series is assumed unselfcorrelated then it can be shown that the spectral ordinates are independent of each other and, furthermore, that each ordinate divided by the spectral mean, $\sum \hat{S}/N$, has an F-distribution with 2 and $N-1$ degrees of freedom. This allows confidence limits to be placed on spectral ordinates so that spectral peaks can be tested for significance.

The null hypothesis, that the given time series is the outcome of a white noise process with zero mean, can be rejected when either one or more peaks exceed the confidence level or when the slope is significantly less than zero. There may be other situations in which the null hypothesis can be rejected, e.g. the presence of band-limited noise. The periodogram is a powerful tool for dealing with the behaviour of a time series in the frequency domain. There is no requirement that the data be windowed or filtered in any way. In fact such windowing methods preclude the use of the periodogram as a well-defined statistic.

Maximum Entropy Spectra

The idea of maximizing entropy in order to improve the resolution of an image or spectrum first arose in radio astronomy (Gull and Daniell, 1978). This Maximum Entropy Method (MEM) was rapidly adopted in time series analysis as a more sophisticated method of reducing the inherent noisiness of spectral estimates present in the periodogram of short time series, see, for example, Reid (1979) and references therein. At the time computers were very slow and various tricks were used to speed up calculations. In spectral analysis, one such trick was to compute the sample covariance of the time series out to some maximum lag, N, and then to take the Fourier Transform of this truncated covariance estimate as the spectral estimate.

The resulting spectral estimate has rather low frequency resolution because truncation in the time domain resulted in convolution in the frequency domain. The new Maximum Entropy Method (MEM) improved spectral resolution considerably, or, at least, it appeared to.

MEM assumes that the sample time series $\{x_t,\ t = 1, ..., N\}$ is the realization of a random vector $\boldsymbol{X}$; the $p+1$ ensemble variances and covariances, $\mathbb{E}(X_i X_j)\ \ i, j \leq p$, are fixed; and that the other covariances at higher lags are simply unknown and undetermined. The Gibbs entropy is generalized to the continuous H-function of the probability density distribution of the ensemble.

$$H = -\oint f(\boldsymbol{X}) \ln(f(\boldsymbol{X})) \boldsymbol{dX} \tag{4}$$

This is then maximized under the constraint that the $N+1$ ensemble covariances are fixed at their sample values

$$\oint \boldsymbol{X}\boldsymbol{X}^T f(\boldsymbol{X})\boldsymbol{dX} = [< x_i x_j >] \tag{5}$$

and

$$\oint f(\boldsymbol{X})\boldsymbol{dX} - 1 = 0 \tag{6}$$

where (assuming zero mean) $[< x_i x_j >]$ is the matrix of sample covariances.

The details of the derivation need not concern us here, but it turns out that:

1. the resulting joint distribution is Gaussian and
2. it is an autoregressive process of order p.

MEM spectral analysis was very popular for a while because it gives spectacularly sharp spectral peaks in contrast to the degraded resolution of other spectral estimators. However there are some problems with it.

The number of peaks in the spectral estimate depends on the order chosen for the analysis; there are usually about $p/2$ spectral peaks. This means that choosing the order, p, appropriately is crucial. For this purpose, a parameter known as the Akaike Information Criterion (AIC) is used (Akaike, 1970). The AIC estimates of the relative quality of a statistical model for a given set of data and so provides a means for model selection. The AIC does not provide a test of a model in the sense of testing a null hypothesis. It tells us nothing about the absolute quality of a model, only the quality relative to other models.

But the worst aspect of MEM must surely be the false sense of achievement it gives the researcher. MEM spectra have razor sharp peaks with elegantly sloped skirts; they *look* precise. A researcher enamoured of this method would be well advised to generate a synthetic time series, i.e. a sequence of random numbers of the same length as his/her data, subject it to MEM analysis with the same maximum lag and then compare the results. It is commonly the case that the MEM spectrum of the random numbers looks equally as convincing as that of the data.

These problems arise from the fact that MEM is a purely autoregressive process; it is an *all-pole model*. In cases where a moving average component is present, the assumption that the variance and first p covariances are the only constraints on the data are incorrect; for example, for a pure moving average (MA) process of order q, the covariances from $q+1$ to infinity are all zero. Experience shows that MA components are frequently present in real world data. They are often a consequence of the convolution or "blurring" of data due to the inevitable imperfections of the measurement technique.

The ARMA Spectrum

The Autoregressive Moving Average or $ARMA$ model for spectral estimation specified by equation (19) is discussed in the following sections. It is the best model for spectral estimation. It is the most general linear model of a stochastic time series.

The ARMA model has two advantages over other spectral techniques:

1. The drift parameter estimate, $\hat{c}$, can be used to test whether there is a significant linear trend in the time series.
2. For given order, (p, q), the model itself can be tested for significance.

When the time series exhibits a trend, i.e. when $\hat{c} \neq 0$ in (19), the process is non-stationary and the ensemble spectrum has a singularity at the origin. Such situations are dealt with either by using an ARIMA ("Autoregressive Integrated Moving Average") model, by forming a new time series of first differences or by removing the trend from the data by the usual methods of ordinary least squares linear regression.

As discussed above, the covariance function, Φ, as defined by (33), encapsulates the important statistical properties of a stationary time series. The sequence $\{\epsilon_t\}$ in (19) is assumed to be "white" (i.e. unselfcorrelated) in order that the two sequences of coefficients, $\{a_1, ..., a_p\}$ and $\{b_1, ..., b_q\}$ be estimated from the data. The sequence, $\{\epsilon_t\}$, can be found by iteratively substituting the coefficient estimates $\{\hat{a}_1, ..., \hat{a}_p\}$ and $\{\hat{b}_1, ..., \hat{b}_q\}$ and the original data back into (19), when $\{\epsilon_t\}$ becomes known as the "sequence of residuals".

If the estimation has not been effective, for example because p and q were chosen to be too small, then the sequence of residuals will be self-correlated, i.e. be non-white. Intuitively speaking, information remains behind in the residuals. As it happens there are powerful statistical tests for assessing whether such a sequence of residuals is indeed white. The *Ljung–Box Test* (Ljung and Box, 1978) is one such test.

Most statistical packages return the sequence of residuals when fitting an ARMA model. These can then be input into the package's Ljung–Box procedure. This returns the Ljung–Box Q-parameter which has a χ^2-distribution. Hence a probabilty, α, can be found for the Q-parameter as a function of lag. Should this probability lie below some given threshold at some lag, say $\alpha < 0.05$, then the null hypothesis that the residuals are unselfcorrelated can be rejected.

This is a powerful test of a model, far more powerful than the Akaike Information Criterion mentioned aboved. In practice it is a simple matter to run the ARMA model-fitting procedure with different values of p and q until a fit which satisfies the Ljung–Box test is found. In practice either a number of pairs of values (p, q) are found or none at all. If the former, then the smallest values should be chosen.

The process of finding an ARMA model is similar to factorizing an algebraic expression by inspection: sometimes factors can be found and sometimes not. Indeed it may well be that a linear algebraic solution may be found for factorizing the sample covariance matrix. Sometimes no ARMA model can be found that fits a given sample time series even though it appears stationary. For example, this happens when the sample is a strongly non-linear function of an underlying ARMA process, but there may well be other reasons.

Iterative Processes

The famous Fibonacci series is perhaps the simplest example of an iterative, deterministic process. It goes:

1, 1, 2, 3, 5, 8, 13, 21, 34, ...

i.e. each term is the sum of the preceding two terms. Mathematically

$$x_t = x_{t-1} + x_{t-2} \tag{7}$$

It is very interesting. It grows exponentially, and the ratio of successive terms converges to the Golden Ratio beloved of the Ancient Greeks.

It can be generalized:

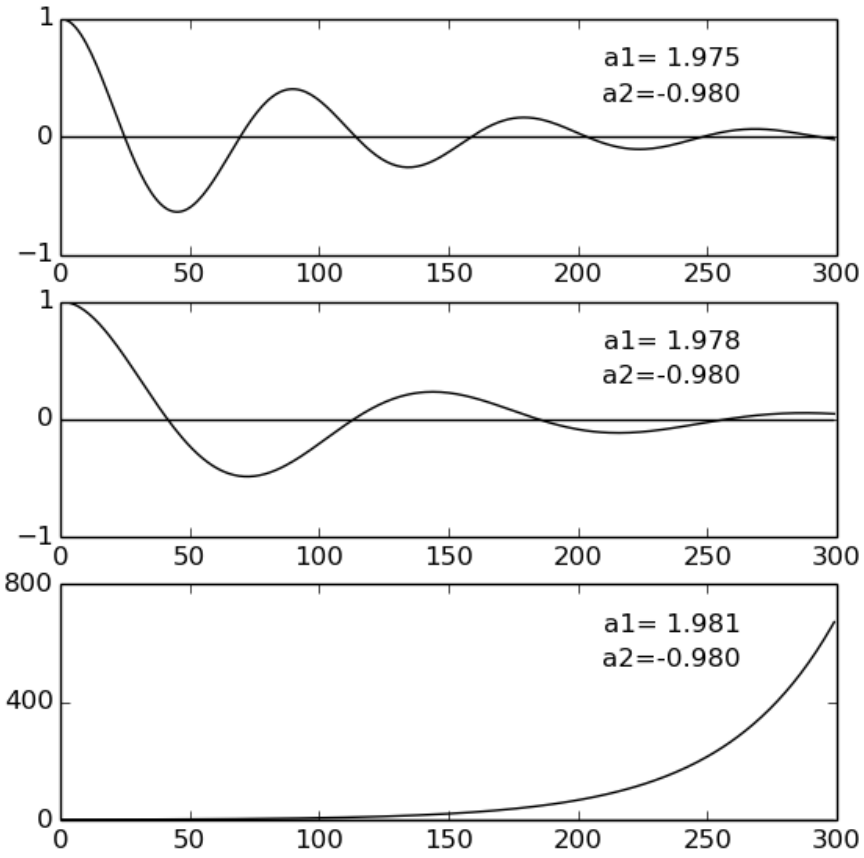
$$x_t = a_1 x_{t-1} + a_2 x_{t-2} \tag{8}$$

so that (7) is just the special case when $a_1 = a_2 = 1$.

Consider what happens if we make $a_1 = 1.183$ and $a_2 = -1$. Starting with (1,1) as before, the sequence goes

1.00, 1.00, 0.18, -0.78, -1.11, -0.53, 0.48, 1.10, 0.82, -0.13, -0.98, -1.02, -0.23, 0.75, 1.12, 0.57, -0.44, -1.09, -0.854, 0.08 ...

This time it doesn't go to infinity; it is cyclic.



Three sequences of the form (8) showing the effect of changing the first coefficient by a small amount.

The gross behaviour of the generated sequence can be strongly dependent on the values of the coefficients a_1 and a_2. The figure shows the 300-long sequences which resulted from three numerical experiments. In each case the second coefficient was set at $a_2 = -0.98$, and the first coefficient varies slightly from run to run, i.e. $a_1 = 1.975, 1.978$ and 1.981. In the first two cases (top two panels), the sequences are damped sinusoids with slightly different frequencies, but in the third case (bottom panel), the sequence increases exponentially like the original (1,1) Fibonacci sequence. A very small change in one coefficient has led to a massive change in the generated sequence.

It turns out that the behaviour of the sequence depends on the *characteristic equation*. Equation (8) can be written using a "shift operator", Z:

$$x_t = a_1 Z x_t + a_2 Z^2 x_t \tag{9}$$

i.e.

$$(a_2 Z^2 + a_1 Z - 1)x_t = 0 \tag{10}$$

where, by definition, $Z x_t = x_{t-1}$. The characteristic equation has the same form as (10) in which the shift operator, Z, becomes a complex number, z:

$$a_2 z^2 + a_1 z - 1 = 0 \tag{11}$$

which has roots $z_{1,2} = \frac{-a_1 \pm \sqrt{a_1^2 + 4a_2}}{2a_2}$

When (11) has complex roots, i.e. when $a_1^2 + 4a_2 < 0$, the generated series is cyclic; otherwise it is exponential. This explains the radical difference between the sequences depicted in the two lower panels of the figure. The characteristic equation (11) has two roots, z_1 and z_2 where

$$z_1 + z_2 = a_1 \tag{12}$$

and

$$z_1 z_2 = -1 \tag{13}$$

When the parabola described by the LHS of 11 does not cross the zero axis, then z_1 and z_2 are complex and $z_2 = z_1^*$. When it does cross the zero axis, z_1 and z_2 are both real.

Equation (8) can be generalized to any finite order, p, viz.:

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} \tag{14}$$

where $\{a_i, i = 1, ..., p\}$ are fixed coefficients.

The shift operator equation is

$$(\sum_{i=1}^{p} a_i Z^i - 1)x_t = 0 \tag{15}$$

and the characteristic equation is

$$\sum_{i=1}^{p} a_i z^i - 1 = 0 \tag{16}$$

The expression on the LHS of (16) is known as the *z-transform* of the model.

Iterative stochastic processes – ARMA

Equation (14) can be generalized still further to

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t \tag{17}$$

where $\{\epsilon_t\}$ is a sequence of random numbers with zero mean. They are assumed to be independent, which implies that they are unselfcorrelated. They are commonly assumed to be Gaussian and identically distributed, but that is not essential.

$\{\epsilon_t\}$ is known as *the innovation*.

As before, (17) can be used to generate sequences iteratively using a random number generator to provide $\{\epsilon_t\}$. There is, however, a major difference. For the same set of starting conditions, every sequence generated by (14) will be the same whereas every sequence generated by (17) will be different. Equation (17) describes a *stochastic* process.

An important feature of (17) is that it is also a regression equation, i.e.

$$X_t = \sum_{i=1}^{p} a_i x_{t-i} + \epsilon_t \tag{18}$$

where at time τ, where $t-1 < \tau < t$, X_t and ϵ_t are random variables, and the x_{t-i} are past fixed sample values. This is the standard regression model with the p coefficients $\{a_i\}$ being the regression coefficients. It means that, given a sequence $\{x_t\}$ and assuming a given order, p, the values of the coefficients can be estimated from the data. For this reason, (17) is referred to as an *autoregression* or AR model since the sequence is regressed on itself, i.e. on its own past values.

The most general linear model of a stationary, stochastic process is the *Autoregressive Moving Average* process or *ARMA* process/model. The output of an ARMA(p,q) process at time t is defined as:

$$\xi_t = c + \epsilon_t + \sum_{i=1}^{p} a_i x_{t-i} + \sum_{j=1}^{q} b_j \epsilon_{t-j} \tag{19}$$

where ξ_t is the random variable output by the process, the x_{t-i} are past realizations of ξ_t, $\{a_i, i = 1,...,p\}$ are known as the autoregression coefficients, the $\{b_j, j = 1,...,q\}$ are known as the moving average coefficients , $\{\epsilon_t\}$ is the zero mean, unselfcorrelated innovation, and c is a constant known as the *drift term* because, when non-zero, it represents an underlying drift in the series values over time. The ordered pair (p,q) is known as the *order* of the process.

Given (p,q), estimates of the $p+q+1$ parameters $\hat{a}_1$, $\hat{a}_2$, ... $\hat{a}_p$, $\hat{b}_1$, $\hat{b}_2$, ... ,$\hat{b}_q$ and $\hat{c}$ can be made from the data. In fact they are estimated from the variance/covariance function of the sample time series. How this is done need not concern us here; suitable packages for ARMA estimation are available in high-level programming languages such as *R, Matlab* and *Python*. ARMA functions in such packages not only return parameter estimates but also complete time series of the *residuals*, i.e. the time series $\{\epsilon_t\}$ obtained when the parameter estimates, $\{\hat{a}_i\}$, $\{\hat{b}_j\}$ and $\hat{c}$, and the sample time series $\{x_t\}$ are substituted in (19). Useful statistics of $\{\epsilon_t\}$ such as its variance σ_ϵ^2 are also returned.

In the definition (19) above, the predicted value of the process, ξ_t, is a random variable. The innovation, ϵ_t, is also a random variable. Consequently the coefficient estimates $\hat{a}_1$ etc. are realizations of random variables with known distributions. Hence confidence limits can be placed on these estimates. This is particularly important with regard to the drift term, $\hat{c}$. All real time series exhibit an apparent long-term trend, either positive or negative. At issue is whether such apparent trends are statistically significant, i.e. whether $\hat{c}$ is significantly different from zero and the trend cannot be attributed to chance alone.

In this way, with an ARMA model, the drift in a sample time series can be tested for significance using accepted methods of statistical inference. While it can never be proven that there is no drift in a population, at least we can show when there is such a drift and at what level of significance. When the drift is so small as to be insignificant, then Occam's Razor allows us to treat the drift as if it were zero.

Name	Symbol	Domain	Sample or Random Variable
t.s. value	x_n	time	sample
t.s. value	y_n	time	sample
F.T of x	X_j	frequency	sample
F.T of y	Y_j	frequency	sample
estimate	$\hat{P}_k$	frequency	sample
t.s. variable	ξ_n	time	random variable
F.T. of ξ	Ξ_j	frequency	random variable
estimator	Π_k	frequency	random variable

Table 1: Symbol conventions

Discrete Time Spectral Analysis

Introduction

A thorough description of Fourier Transforms and Spectral Analysis is given by Bracewell (1986). Hamilton (1994) offers an equally thorough description of time series methods.

In the present development, the variance density spectrum is defined as the ensemble average of the periodogram. The periodogram of the sample is thus a statistic. For an ARMA process, it is a consistent estimator of the ensemble variance density spectrum.

In this way, some key theorems are reproduced without recourse to continuous functions. In practice any real time series occupies a range of time scales, $< \Delta t, N\Delta t >$, where Δt is the sampling interval and N is the number of terms in the series. The behaviour of the underlying continuous process, should there be one, at time scales shorter than the sampling period, Δt, is of no concern. If there is evidence of aliasing such as, for example, a concentration of variance at the high-frequency end of the spectrum, then it is the responsibility of the researcher to change the sampling regime to something more suitable. Likewise the length, N, and span, $N\Delta t$, of the time series are assumed to be finite. These assumptions allow key results to be developed while avoiding the pitfalls of spectral methods based on deterministic, continuous functions of infinite extent.

The Periodogram

We start with a finite sequence of real numbers, $\{x_t, t = 0, ..., N-1\}$, which is assumed to comprise samples of a varying physical quantity separated by equal intervals of time or covering equal intervals of time, i.e. a 'time series'. We are concerned with $\{X_k, k = 0, ..., N-1\}$, the discrete Fourier transform (DFT) of the time series $\{x_n\}$ given by

$$X_k = \sum_{t=0}^{N-1} x_t . e^{-2\pi i k t/N} \ , \ k = 0, 1...N-1 \tag{20}$$

where $i = \sqrt{-1}$. In general, the X_k are complex numbers and, because the x_t are real

$$X_{-k} = X_{N-k} = X_k^* \tag{21}$$

where X_k^* is the complex conjugate of X_k. Because of the cyclical character of the DFT, X_j can be defined for arbitrary k since for any integer m

$$X_{k+mN} = X_k \tag{22}$$

The underlying concept of spectral analysis is encapsulated by Parseval's theorem which, for the finite, discrete case under discussion here, can be expressed as:

$$\sum_{t=0}^{N-1} x_t^2 = \frac{1}{N} \sum_{k=0}^{N-1} \mid X_k \mid^2 . \tag{23}$$

We define a new sequence, the *periodogram*, $\{\hat{P}_k\}$, where

$$\hat{P_k} = 2 \mid X_k \mid^2 /N \ , \ k = 1,2...N/2-1 \tag{24}$$

and, if N is even,

$$\hat{P}_{N/2} = \mid X_{N/2} \mid^2 /N.$$

After substituting from (21), (23) becomes

$$s^2 = \frac{1}{N} \sum_{n=0}^{N-1} x_n^2 = \sum_{k=1}^{N/2} \hat{P}_k. \tag{25}$$

since $\hat{P}_0 = 0$.

For notational convenience (and commonly in practice), the data sequence $\{x_n\}$ has zero mean, and so s^2 is the sample variance. Equation (25) can be regarded as a conservation law which describes variance density as a function of frequency, $\hat{P}_k$ being that component of the variance associated with the frequency $k\Delta f = k/N\Delta t$. Note that $\hat{P}_k$ has the same units as s^2; to make it a variance density, it must be divided by Δf. In the following development, we assume $\Delta t = 1$ without loss of generality.

The ensemble periodogram is given by the equivalent equation to (24), viz.:

$$\Pi_k = 2 \mid \Xi_k \mid^2 /N \ , \ k = 1,2...N/2-1 \tag{26}$$

and, for even N,

$$\Pi_{N/2} = \mid \Xi_{N/2} \mid^2 /N.$$

where

$$\Xi_k = \sum_{t=0}^{N-1} \xi_t . e^{-2\pi ikt/N} \ , \ k = 0,1...N/2 \tag{27}$$

The Variance Density Spectrum

The variance density spectrum S_k is defined as the ensemble average of the periodogram, viz.:

$$S_k = \mathbb{E}(\Pi_k) \quad \text{for} \quad k = 0,1,...N/2 \tag{28}$$

Note that, while this development is discrete in time, it is not, necessarily, discrete in frequency. The discrete frequency forms of (20) and (27) could equally well have been written as continuous functions of frequency, f, viz.:

$$X(f) = \sum_{t=0}^{N-1} x_t . e^{-2\pi ift/f_s} \ , \ 0 < f < f_s \tag{29}$$

and, for the ensemble:

$$\Xi(f) = \sum_{t=0}^{N-1} \xi_t . e^{-2\pi i f t/f_s} \ , \ 0 < f < f_s \tag{30}$$

where $f_s = 1/\Delta t$, Δt being the sampling interval, usually defined to be one. This allows a variance density spectrum, $S(f)$, to be specified as a continuous function of frequency. This is sometimes desirable for display purposes whereas the discrete frequency forms are more easily calculated using the Fast Fourier Transform.

The covariance function

From (26), (27) and (28) and using $|\Xi_k|^2 = \Xi_k \Xi_k^*$:

$$S_k = \frac{1}{N} \mathbb{E}\left[\left(\sum_{m=0}^{N-1} \xi_m . e^{-2\pi i k m/N} \right) \left(\sum_{n=0}^{N-1} \xi_n . e^{2\pi i k n/N} \right) \right] \tag{31}$$

Thus

$$S_k = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \mathbb{E}(\xi_m \xi_n) e^{-2\pi i k (m-n)/N} \tag{32}$$

For a wide-sense stationary process, the ensemble covariance function is defined by

$$\Phi(m-n) = \mathbb{E}(\xi_m \xi_n) = \Phi(n-m) \tag{33}$$

and (32) becomes

$$S_k = \frac{1}{N} \sum_{L=-N}^{N} (N - |L|) \Phi(L) e^{-2\pi i k L/N} \tag{34}$$

so that S is the DFT of the function $(N-|L|)\Phi(L)/N$ on the domain $-N \le L \le N$.

Note that, as defined here, the variance density spectrum is the Fourier Transform not of the covariance function, as is commonly assumed, but of the covariance function multiplied by the triangle function. This comes about because of the finite length of the time series. Consequently the variance spectrum of any finite time series has, in effect, been convoluted with the Fourier Transform of the triangle function, i.e. with the $sinc^2$ function. This is important when sinusoids are involved as discussed further below.

When the time series is unselfcorrelated, i.e. when

$$\mathbb{E}(\xi_m \xi_n) = \Phi(m-n) = \sigma^2 \delta_{mn} \tag{35}$$

where σ^2 is the variance of ξ_n and δ_{mn} is the Kronecker delta, i.e. $\delta_{nn} = 0$ and $\delta_{mn} = 0$ when $m \ne n$, then

$$S_k = \frac{1}{N} \Phi(0) = \frac{\sigma^2}{N} = \text{const.} \tag{36}$$

and the time series is said to be "white" by analogy with the energy spectrum of visible light.

The periodogram of a white time series

The given time series, $\{x_n\}$, can be considered one realization of an ensemble of random sequences, $\{\xi_n\}$. Likewise $\{X_k\}$, the DFT of $\{x_n\}$, is the corresponding realization of the random

sequence $\{\Xi_k\}$, the DFT of $\{\xi_n\}$; and the computed periodogram, $\{P_k\}$, is the realization of the ensemble periodogram, $\{\Pi_k\}$.

As a null hypothesis we assume that the time series has zero mean and is unselfcorrelated. Thus

$$\mathbb{E}(\xi_n) = 0 \tag{37}$$

and

$$\mathbb{E}(\xi_m \xi_n) = \sigma^2 \delta_{mn} \tag{38}$$

The random variable Ξ_k defined by (27) can be written in terms of its real and imaginary components, viz.:

$$\Xi_k = \Re_k + \Im_k i \tag{39}$$

i.e.

$$|\ \Xi_k^2\ |^2 = \Xi_k \Xi_k^* = \Re^2 + \Im^2 \tag{40}$$

where

$$\Re_k = \sum_{n=1}^{N-1} \xi_n \cos(2\pi kn/N) \tag{41}$$

and

$$\Im_k = \sum_{n=1}^{N-1} \xi_n \sin(2\pi kn/N) \tag{42}$$

For large N, by the central limit theorem, both $\Re_k$ and $\Im_k$ must be close to Gaussian whatever the distribution of the ξ_n may be. They are also unselfcorrelated, uncorrelated with each other and have zero mean. The variance of $\Re_k$ is given by

$$\mathbb{E}(\Re_k^2) = \sigma^2 \sum_{m=1}^{N} \sum_{n=1}^{N} \delta_{mn} \cos(2\pi km/N) \cos(2\pi kn/N) \tag{43}$$

$$= \frac{\sigma^2}{2} \left[N + \sum_{n=1}^{N} \cos(4\pi kn/N) \right] \tag{44}$$

$$= \frac{\sigma^2}{2} \left[N + \frac{1}{2} \sum_{n=1}^{N} (z^n + z^{*n}) \right] \tag{45}$$

$$= \frac{\sigma^2}{2} \left[N + \frac{1}{2} \left(\frac{z(1-z^N)}{1-z} + \frac{z^*(1-z^{*N})}{1-z^*} \right) \right] \tag{46}$$

where $z = e^{4\pi ik/N}$ so that

$$z^N = z^{*N} = 1 \tag{47}$$

and

$$\mathbb{E}(\Re_k^2) = \frac{\sigma^2 N}{2} \tag{48}$$

Similarly

$$\mathbb{E}(\Im_k^2) = \frac{\sigma^2 N}{2} \tag{49}$$

It follows that the rescaled random variables $\sqrt{\frac{2}{N\sigma^2}}\Re_k$ and $\sqrt{\frac{2}{N\sigma^2}}\Im_k$ are both normal(0,1). Hence, from (26) the random variable

$$\frac{N\Pi_k}{\sigma^2} = \frac{2 \mid \Xi_k \mid^2}{N\sigma^2} = \frac{2\Re^2}{N\sigma^2} + \frac{2\Im^2}{N\sigma^2} \tag{50}$$

has a χ^2-distribution with two degrees of freedom, and so the scaled spectral estimate of an unselfcorrelated time series, $N\hat{P}_k/\sigma^2$, has a χ^2-distribution with two degrees of freedom.

In practice we do not know σ^2 in (50), which is a population parameter. We only know its estimate $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{k=1}^{N} x_k^2 \tag{51}$$

since $< x_k >= 0$. Assuming that the $\{x_k\}$ are the realization of a sequence of unselfcorrelated random variables $\{\xi_k\}$ which are normally distributed (0,σ), then

$$\frac{\hat{\sigma}^2}{\sigma^2} = \frac{1}{N-1}\sum_{k=1}^{N} \frac{x_k^2}{\sigma^2} \tag{52}$$

has a χ^2-distribution with $N-1$ degrees of freedom. Hence the quantity

$$\frac{N\hat{P}_k}{\hat{\sigma}^2} = \frac{N\Pi_k}{\sigma^2} \Big/ \frac{\hat{\sigma}^2}{\sigma^2} \tag{53}$$

is the realization of a random variable which is the ratio of two random variables having χ^2-distributions with 2 and $N-1$ degrees of freedom, respectively.

If X_1^2 and X_2^2 are independent random variables following chi-square distributions with ν_1 and ν_2 degrees of freedom respectively then the distribution $F = \frac{X_1^2/\nu_1}{X_2^2/\nu_2}$ is said to follow the variance ratio or F-distribution with ν_1 and ν_2 degrees of freedom (see: Abramowitz and Stegun (1965), subsection **26.6.2**). Inserting degrees of freedom of 2 and $N-1$ into (53) gives an F statistic for each value of the periodogram, viz.:

$$F_k = \frac{N(N-1)\hat{P}_k}{2\hat{\sigma}^2} = \frac{N\Pi_k}{2\sigma^2} \Big/ \frac{\hat{\sigma}^2}{(N-1)\sigma^2} \tag{54}$$

and so F_k has an F-distribution with 2 and $N-1$ degrees of freedom.

The p-values for the F-distribution are provided in most statistical software packages. These can be used to test each ordinate of the periodogram for whiteness using the F_k statistic defined by (54). In practice the F-distribution with these degrees of freedom is barely distinguishable from the χ^2-distribution with 2 degrees of freedom for $N > 100$.

Averaged white periodograms

The resolution of periodograms may be improved by averaging a number of them (Bartlett, 1948). In order that sample statistics be calculable, it is essential that the periodograms come from time series that are of the same length and were generated by the same processes or by identical processes and that they be independent. In the present case of time series which are assumed unselfcorrelated under a null hypothesis, it is sufficient that they not overlap in time.

The kth ordinate of the mean of the periodograms of M, N-long time series with ensemble variance σ^2 is given by

$$\bar{\hat{P}}_k = \frac{1}{M}\sum_{m=1}^{M} P_k^m \quad \text{for } k = 1 \cdots N/2 \tag{55}$$

Multiplying both sides by N/σ^2 gives

$$\frac{N\bar{\hat{P}}_k}{\sigma^2} = \frac{1}{M}\sum_{m=1}^{M} \frac{NP_k^m}{\sigma^2} \quad \text{for } k = 1 \cdots N/2 \tag{56}$$

Since each of the terms in the sum has a χ^2-distribution with 2 degrees of freedom, then the sum itself must have a χ^2-distribution with $2M$ degrees of freedom. Hence

$$\text{Var}\left(\frac{N\bar{\hat{P}}_k}{\sigma^2}\right) = \frac{1}{M^2}\text{Var}\left(\sum_{m=1}^{M} \frac{NP_k^m}{\sigma^2}\right) = \frac{2}{M} \tag{57}$$

Hence $\text{Var}\left(\frac{N\bar{\hat{P}}_k}{\sigma^2}\right) \to 0$ as $M \to \infty$, and P_k of a white time series is a consistent estimator of S_k as defined by (28).

Fourier Transform of a Convolution

Measurements are often contaminated or blurred by the physical characteristics of the measuring device. Thermometers have thermal mass and so do not respond instantaneously to temperature changes. Lenses are never perfect and all photographic images are blurred to some degree. Prior to measurement, physical quantities may themselves be blurred by diffusion and mixing. Numerically such blurring is expressed as *convolution*.

Consider an "output" time series $\{y_t, t = 0, ..., N-q-1\}$ formed by convolving an "input" time series $\{x_t, t = 0, .., N-1\}$ with a "filter" function, $\{\beta_j, j = 0, ..., q\}$, i.e.

$$y_t = \sum_{j=0}^{q} \beta_j x_{t-j} \tag{58}$$

Let Y_k, X_k and B_k be the Fourier Transforms of $\{y_t\}$, $\{x_t\}$ and $\{\beta_j\}$, respectively; then from

(20),

$$\begin{aligned}
\sum_{t=0}^{N-1} y_t e^{-2\pi ikt/N} &= \sum_{t=0}^{N-1} \left[\sum_{j=0}^{q} \beta_j x_{t-j} \right] e^{-2\pi ikt/N} \\
&= \sum_{j=0}^{q} \beta_j \left[\sum_{t=0}^{N-1} x_{t-j} e^{-2\pi ik(t-j)/N} \right] e^{-2\pi ikj/N} \\
&= \sum_{j=0}^{q} \beta_j \left[\sum_{t-j=0}^{N-1} x_{t-j} e^{-2\pi ik(t-j)/N} \right] e^{-2\pi ikj/N} \\
&= \sum_{j=0}^{q} \beta_j X_k e^{-2\pi ikj/N} \\
&= \left[\sum_{j=0}^{q} \beta_j e^{-2\pi ikj/N} \right] X_k
\end{aligned}$$

i.e.

$$Y_k = B_k X_k \tag{59}$$

and so the Fourier Transform of convoluted functions is the product of the Fourier Transforms of those functions. The change from t to $t-j$ in the sum is allowable because the kernel, $e^{-2\pi ikt/N}$, is cyclic over the domain of the dummy variable.

Periodogram of an ARMA process

For notational convenience we assume the sampling interval, Δt, to be 1 time unit and the periodogram frequency resolution, $1/N\Delta t$, to be $1/N$ frequency units.

An ARMA(p,q) process is defined by

$$y_t = x_t + \sum_{i=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} b_j x_{t-j} \tag{60}$$

where $\{x_t\}$ is an unselfcorrelated sequence. Equation (60) can be written

$$y_t - \sum_{i=1}^{p} a_i y_{t-i} = x_t + \sum_{j=1}^{q} b_j x_{t-j} \tag{61}$$

We define two new vectors γ and β where

$$\gamma_0 = 1 \quad \text{and} \quad \gamma_i = -a_i \quad \text{for} \quad i > 0 \tag{62}$$

and

$$\beta_0 = 1 \quad \text{and} \quad \beta_j = b_j \quad \text{for} \quad j > 1 \tag{63}$$

Hence

$$\sum_{i=0}^{p} \gamma_i y_{t-i} = \sum_{j=0}^{q} \beta_j x_{t-j} \quad \text{for } t = 1 \cdots N \tag{64}$$

For the ensemble

$$\sum_{i=0}^{p} \gamma_i \upsilon_{t-i} = \sum_{j=0}^{q} \beta_j \xi_{t-j} \quad \text{for } t = 1 \cdots N \tag{65}$$

where x_t and y_t are assumed to be realizations of random variable sequences ξ_t and υ_t respectively. Taking the (continuous) Fourier Transform of both sides of (64) and using (59) gives

$$Y(f)\Gamma(f) = X(f)B(f) \tag{66}$$

where $Y(f)$, $\Gamma(f)$, $X(f)$ and $B(f)$ are the Fourier Transforms of y_t, γ_t, x_t and β_t, respectively. Rearranging, taking the squared modulus of both sides and replacing f with integer value k gives

$$|Y_k|^2 = |X_k|^2 \left| \frac{B_k}{\Gamma_k} \right|^2 \tag{67}$$

By analogy with (24), let Q_k be the periodogram of the ARMA time series, $\{y_k\}$, so that (67) becomes

$$Q_k = P_k \left| \frac{B_k}{\Gamma_k} \right|^2 \tag{68}$$

Now let M independent periodograms, Q_k^m, be averaged, i.e.

$$\bar{\hat{Q}}_k = \frac{1}{M} \sum_{m=1}^{M} Q_k^m \quad \text{for } k = 1 \cdots N/2 \tag{69}$$

Then from (68),

$$\bar{\hat{Q}}_k = \bar{\hat{P}}_k \left| \frac{B_k}{\Gamma_k} \right|^2 \tag{70}$$

where Γ_k and B_k are functions of population parameters such as a_i and b_j, and so are independent of M.

By (57), the variance of $\bar{\hat{P}}_k$ converges as $M \to \infty$. Therefore, by (68), $\bar{\hat{Q}}_k$, the variance of the periodogram of the ARMA time series, $\{y_k\}$, also converges as $M \to \infty$.

Hence the periodogram of an ARMA time series is a consistent estimator of its ensemble spectrum.

The spectral estimate of an ARMA process

From (68),

$$\mathbb{E}(Q_k) = \mathbb{E}(P_k) \left| \frac{B_k}{\Gamma_k} \right|^2 \tag{71}$$

Hence, using (28) and (36), the variance density spectrum, S_k, of an ARMA time series is given by

$$S_k = \frac{\sigma^2}{N} \left| \frac{B_k}{\Gamma_k} \right|^2 \tag{72}$$

and the spectral estimate, $\hat{S}_k$, by

$$\hat{S}_k = \frac{\hat{\sigma}^2}{N} \left| \frac{\hat{B}_k}{\hat{\Gamma}_k} \right|^2 \tag{73}$$

where $\hat{\sigma}^2$ is the variance of the residuals. $\hat{B}_k$ and $\hat{\Gamma}_k$ are given by

$$\hat{\Gamma}_k = 1 - \sum_{j=1}^{p} \hat{a}_i z^j \tag{74}$$

and

$$\hat{B}_k = 1 + \sum_{j=1}^{q} \hat{b}_j z^j \tag{75}$$

where $z = e^{-2\pi ik/N}$, and $\{\hat{a}_i\}$ and $\{\hat{b}_j\}$ are estimated from the data.

The periodogram of a sinusoid

The periodogram of a sinusoid exhibits a narrow peak. The peak is not a δ-function for a time series of finite length as discussed in reference to (34).

Consider a time series $\{c_j, j = 0, ..., N-1\}$, defined explicitly by a cosine function with arbitrary frequency, f_0, and phase, ϕ:

$$c_j = A\cos(2\pi j f_0/N + \phi) = x_j + y_j \tag{76}$$

where

$$x_j = Ae^{2\pi ijf_0/N + \phi i)}/2 \tag{77}$$

and

$$y_j = Ae^{-2\pi ijf_0/N + \phi i)}/2 \tag{78}$$

Substituting x_j into (29) gives

$$X(f) = \frac{A}{2N} \sum_{j=0}^{N-1} e^{-2\pi ij(f-f_0)/N) + \phi i} \tag{79}$$

After multiplying both sides by $e^{-2\pi i(k-f_0)/N}$

$$X(f)e^{-2\pi i(f-f_0)/N} = \frac{A}{2N} \sum_{j=0}^{N-1} e^{-2\pi i(j+1)(f-f_0)/N + \phi i} \tag{80}$$

Subtracting (80) from (79) and transposing gives

$$X(f) = \frac{Ae^{\phi i}\left[1 - e^{2\pi i(f-f_0)}\right]}{2N\left[1 - e^{2\pi i(f-f_0)/N}\right]} \tag{81}$$

$$= \frac{Ae^{\pi i(k-f_0) + \phi i}\left[e^{-\pi i(k-f_0)} - e^{\pi i(k-f_0)}\right]}{2Ne^{\pi i(k-f_0)/N}\left[e^{-\pi i(k-f_0)/N} - e^{\pi i(k-f_0)/N}\right]} \tag{82}$$

i.e.

$$X_k = \frac{Ae^{\pi i(f-f_0) + \phi i}\sin\{-\pi(f-f_0)\}}{2Ne^{\pi i(f-f_0)/N}\sin\{-\pi(f-f_0)/N\}} \tag{83}$$

Likewise, $\{Y(f)\}$, the DFT of $\{y_j\}$ is given by

$$Y(f) = \frac{Ae^{\pi i(f+f_0) + \phi i}\sin\{-\pi(f+f_0)\}}{2Ne^{\pi i(f+f_0)/N}\sin\{-\pi(f+f_0)/N\}} \tag{84}$$

where, from (76) the squared modulus of the DFT of $\{c_j\}$ is given by

$$| C(f) |^2 = X(f)X(f)^* + X(f)Y(f)^* + X(f)^*Y(f) + X(f)^*Y(f)^* \tag{85}$$

From (83) and (84),

$$X(f)X(f)^* = \frac{A^2 \sin^2\{\pi(f-f_0)\}}{4N^2 \sin^2\{\pi(f-f_0)/N\}} \tag{86}$$

$$X(f)Y(f)^* = \frac{A^2 e^{2\pi i f_0 + 2\phi i} \sin\{\pi(f+f_0)\}\sin\{\pi(f-f_0)\}}{4N^2 e^{2\pi i f_0/N} \sin\{\pi(f+f_0)/N\}\sin\{\pi(f-f_0)/N\}} \tag{87}$$

$$X(f)^*Y(f) = \frac{A^2 e^{-2\pi i f_0 - 2\phi i} \sin\{\pi(f+f_0)\}\sin\{\pi(f-f_0)\}}{4N^2 e^{-2\pi i f_0/N} \sin\{\pi(f+f_0)/N\}\sin\{\pi(f-f_0)/N\}} \tag{88}$$

and

$$Y(f)Y(f)^* = \frac{A^2 \sin^2\{\pi(f+f_0)\}}{4N^2 \sin^2\{\pi(f+f_0)/N\}} \tag{89}$$

Adding the cross-modulation terms gives an error term $E(f)$,

$$E(f) = X(f)Y(f)^* + X(f)^*Y(f) \tag{90}$$

Thus

$$E(f) = \frac{A^2 \cos\{2\pi(1-1/N)f_0 + 2\phi\}\sin\{\pi(f+f_0)\}\sin\{\pi(f-f_0)\}}{2N^2 \sin\{\pi(f+f_0)/N\}\sin\{\pi(f-f_0)/N\}} \tag{91}$$

Substituting (21) in (84) gives

$$Y(f)Y(f)^* = X(f)X(f)^* \tag{92}$$

and the periodogram of a cosine function of arbitrary frequency and phase becomes:

$$Q(f) = \frac{A^2 \sin^2\{\pi(f-f_0)\}}{2N^2 \sin^2\{\pi(f-f_0)/N\}} + E(f) \tag{93}$$

Note that the error term, $E(f)$, is dependent on the phase, ϕ. Its absolute value $|E(f)|$ has a maximum value of $0.025A^2$ when $f_0 \approx 1$ and $k = 1$ and is bounded above by $| 1.5A^2/2N \sin(2\pi f/N) |$.

When N is large, (93) becomes

$$Q(f) \approx \frac{A^2 \sin^2\{\pi(f-f_0)\}}{2\{\pi(f-f_0)\}^2} \tag{94}$$

i.e. the $sinc^2$ function. The discrete frequency periodogram is

$$Q_k \approx \frac{A^2 \sin^2\{\pi(k-f_0)\}}{2\{\pi(k-f_0)\}^2} \tag{95}$$

The Spectrum of a Random Walk

Many environmental time series are comprised of temperature measurements or depend in some way on temperature, particularly proxy time series such as isotope ratios in ocean sediment and ice. Temperatures are, in turn, dependent on the transport of heat and other forms of energy by turbulent stochastic processes (Hasselmann, 1976). The relationship between temperature and heat flux is given by Fourier's heat equation:

$$\rho C \frac{\partial T}{\partial t} = -K_0 \frac{\partial T}{\partial x} + Q(x,t) \tag{96}$$

where ρ and C are the density and specific heat of the body being heated, T is its temperature, t is the time, x is a spatial variable with the units of length, K is the conductivity and Q is the flux of heat or other form of energy. In general terms and in discrete form, (96) becomes

$$\rho C \frac{\Delta T}{\Delta t} = -K \frac{\Delta T}{\Delta x} + Q \tag{97}$$

As previously we choose a time scale for which $\Delta t = 1$ gives

$$\rho C(T_t - T_{t-1}) = -\frac{K}{\Delta x}(T_{t-1} - T_{res}) + Q \tag{98}$$

where T_{res} is the temperature of some relatively stable reservoir to and from which heat is being transferred. This becomes

$$T_t = q_t + a_1 T_{t-1} \tag{99}$$

where

$$q_t = Q + \frac{K}{\rho C \Delta x} T_{res} \tag{100}$$

and

$$a_1 = 1 - \frac{K}{\rho C \Delta x} \tag{101}$$

$$= 1 - \alpha \tag{102}$$

where $0 \leq \alpha << 1$.

At sufficiently large time scales, Δt, q_t can be assumed to have zero mean and to be unself-correlated, in which case (99) describes an ARMA(1,0) process for temperature with variance density spectrum given by:

$$S(f) = \frac{\sigma_q^2}{N|1 - a_1 z|^2} \tag{103}$$

where σ_q^2 is the variance of the q_t white noise heating/cooling process,

$$z = e^{-2\pi i f/N} \tag{104}$$

and N is the length of the time series of temperature values. This is a *single-pole model*; there is a single pole (i.e. singularity) in the z-plane. It is on the real axis outside the unit circle at $z_1 = 1/a_1$.

Setting the discrete time angular frequency, ω as $\omega = 2\pi f/N$, (103) can be rewritten

$$\begin{aligned} S(f) &= \frac{\sigma_q^2}{N(1 - 2a_1 \cos\omega + a_1^2)} \\ &= \frac{\sigma_q^2}{N((1 - 2a_1 + a_1^2) + 2(a_1 - \cos\omega))} \\ &= \frac{\sigma_q^2}{N((1 - a_1)^2 + 4a_1 \sin^2 \frac{\omega}{2})} \end{aligned}$$

Hence for small values of f, $\omega << 1$, $\sin\omega = \omega$ and (103) becomes

$$S(f) = \frac{\sigma_q^2}{N(\alpha^2 + a_1\omega^2)} \approx \frac{\sigma_q^2}{N(\alpha^2 + \omega^2)} \tag{105}$$

When $\omega > \alpha$, $S(f)$ is inversely related to ω^2 and so inversely related to f^2.

When $\alpha = 0$ there is a singularity at $S(f) = 0$, and the time series is said to be a *random walk*. Since the variance of a random walk increases with time, (33) does not hold, the time series is not stationary and a variance density spectrum cannot be defined. However, when α is small but non-zero, the time series is stationary, and a variance density spectrum can be defined. Such a time series is called a *centrally biased random walk* in statistics and *low-pass filtered noise* or *red noise* in signal processing. We will use the term *red noise*.

Spurious Regression

Introduction

In statistics, *regression* is the process of fitting a straight line to a collection of points on a graph. The sample correlation coefficient, r, is a measure of "goodness of fit", i.e. of how well the fitted line fits the data points. When the line fits perfectly then $r = \pm 1$. The sign of r indicates the slope of the regression line; a negative value indicates the line slopes downwards, i.e. that the values tend to decrease with increases in the independent variable.

If a system is deterministic then its variables are all single-valued functions of time. Experimental observations of dynamical variables are commonly displayed as functions of time and a regression line fitted to the observations to display the trend or rate of change with time. This is commonplace, something most researchers learned in school.

However, there can be serious problems with this methodology when the system under investigation is stochastic. Nelson and Kang (1984) demonstrated that, for certain stochastic processes such as a *random walk*, the use of time as the explanatory variable can lead to the appearance of a trend even though none was present in the original data. An observed trend obtained by regressing a physical quantity on time may or may not be real, depending on the deterministic or stochastic nature of the system under investigation. A similarly misleading effect is observed when two random walk time series are compared. This was noted nearly a century ago by the then President of the Royal Statistical Society, George Udny Yule and termed "Nonsense-Correlation" by him (Yule, 1926). More recently Granger and Newbold (1974) have described it as *Spurious Regression*. Although not widely known outside the field of Econometrics, the implications of these papers cannot be overestimated. They are little known in the physical sciences.

It is widely assumed that this effect is solely a consequence of the time series being a random walk and so non-stationary, but that is not the case. Here we show that the effect is the result of a high concentration of variance at the low frequency end of the spectrum, i.e. at periods which are long compared with the length of the sample time series, such as the red noise time series described in the previous subsection. Such time series are common in nature, particularly temperature time series.

Time as the independent variable

A useful feature of the ARMA approach is that it provides means for generating synthetic time series with specific statistical properties. A random walk time series can be considered an ARMA(1,0) process with unit autoregression coefficient, i.e. the process whose realization is $\{y_t\}$ given by

$$y_t = a_1 y_{t-1} + x_t \tag{106}$$

where $a_1 = 1$ and x_t is unselfcorrelated.

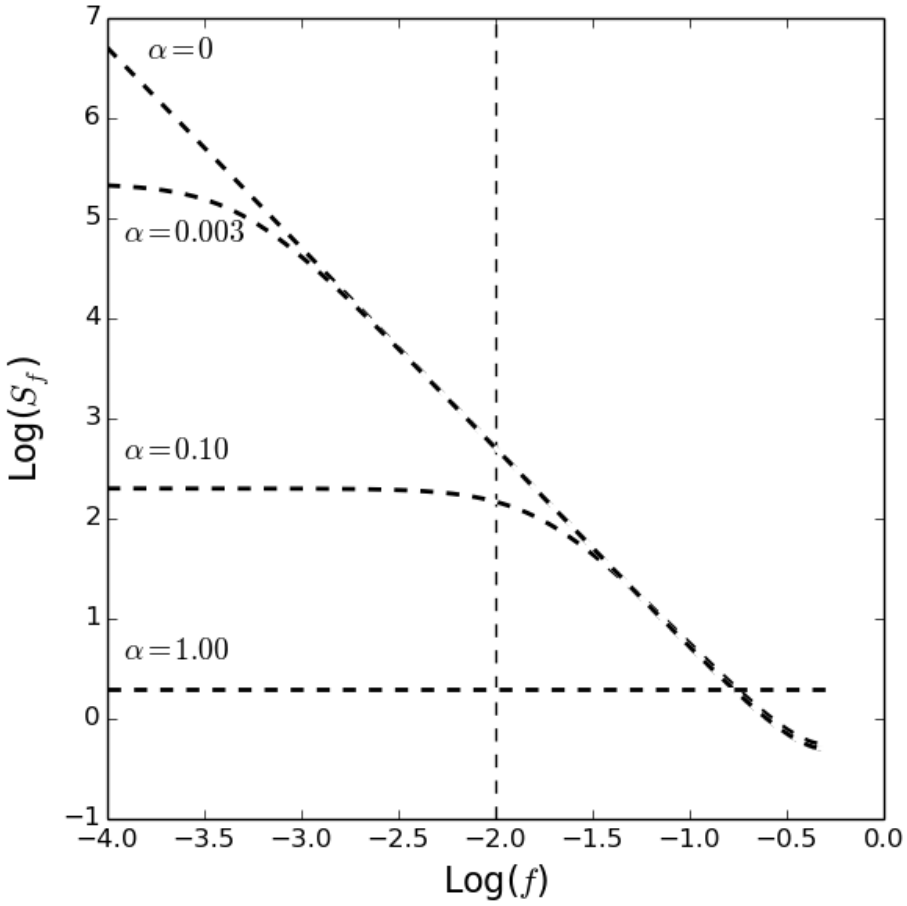


Figure 1: Spectra of four synthetic time series for various values of α

Of interest is the case for which $a_1 \neq 1$ but lies close to 1. Let

$$a_1 = 1 - \alpha \tag{107}$$

where $0 \leq \alpha \leq 1$. Variance spectra for four such time series are shown in Figure 1.

A numerical experiment was performed in which a realization of a 2000-long synthetic time series was generated using equation (106) with $\alpha = .003$. It is shown graphically in the upper diagram of Figure 2. In the diagram, the time series has been subdivided into 20 non-overlapping intervals, each of length 100. In each interval the regression against elapsed time was found. The regression lines for each interval are plotted in the upper diagram, and the values of the correlation coefficient for each interval are plotted in the lower diagram. It can be seen that most of the regression lines fit the data very well indeed and that there is a wide range of correlation coefficient values ranging from close to -1 to close to +1 resulting from the long "period" of the quasi-cyclic behaviour of the time series.

In fact there can be no true correlations or regressions because the time series is purely the outcome of equation (106). These apparent correlations and regression fits are entirely spurious. Likewise neither is there any periodic or cyclic behaviour. The time series displayed in Figure 2 can properly be described as red noise.

It should also be noted that the process described by (106) is stationary: all of its statistical properties are independent of the time, t. Clearly the spurious correlations and spurious regressions observed in this case are not the outcome of the non-stationarity of the process. Tests for stationarity are irrelevant in assessing the validity or otherwise of the regression of a time series against the elapsed time.

What is relevant is the variance density at periods longer than the sample time series under investigation. The reciprocal of the time series length (.01) is shown as the vertical dashed line in Figure 1. Intuitively we "know" about the variance spectrum to the right of this line from the data sample. There is no way to tell much about what is happening to the left of the line because this involves periods greater than the length of the data sample. By using stationarity tests such as Dickey–Fuller, we are trying to assess what is happening at infinitesimal frequencies, i.e. at infinite periods. Common sense would suggest that this is not possible.

How then is spectral shape related to spurious correlation? A further numerical experiment was performed in which the process described by equation (106) was used to generate 100,000 time series realizations, each 100 time units long, and the values correlated with time in each case. Histograms showing the resulting frequency distributions, f, of the correlation coefficient, r, are plotted in Figure 3. In each case the histogram has been normalized by the bin width so that the area under the curve is one, i.e. they are probability distributions.

In Figure 3 the distributions for $\alpha = 0$ (i.e. a true random walk) and $\alpha = .003$ are almost identical. They are bimodal with peaks near $|r| = 0.9$, indicating that in these cases a high absolute value of the correlation coefficient is the most likely outcome when a red noise sample time series is regressed on elapsed time. For $\alpha = .01$ the curve is broadened compared with $\alpha = 0$, indicating that high correlations are possible. For $\alpha = 0$ the time series is white and the displayed histogram is approximately Gaussian as expected. It is the Pearson r-distribution.

Correlations between time series

The above discussion concerns the spurious correlation of red noise sample time series on elapsed time. Spurious correlation also occurs when two red noise time series samples are compared. This is demonstrated by Figure 4. There is a similar broad scatter of correlation coefficients indicating that spurious correlation occurs in this situation as well. Indeed, it was this aspect of spurious correlation or nonsense-correlation that was first noted by Yule (1926).

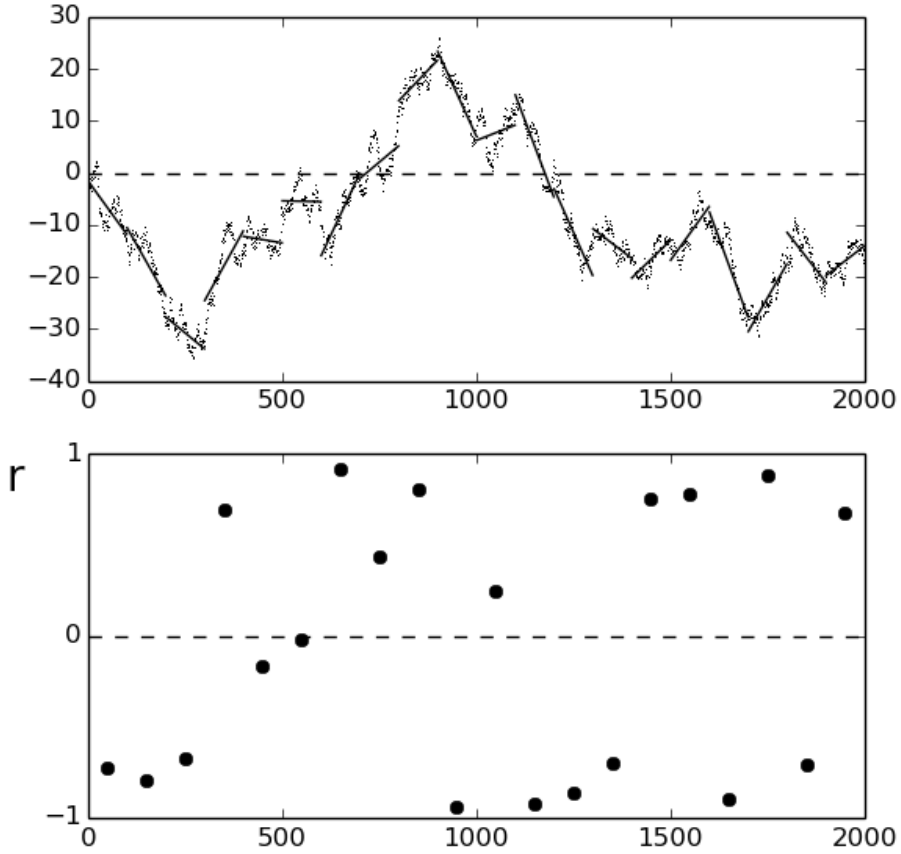


Figure 2: Synthetic time series of an ARMA(1,0) process with $\alpha = .003$. The time series was subdivided into twenty, 100-long, non-overlapping intervals. In each interval the correlation between value and elapsed time was evaluated. The corresponding correlation coefficients are plotted in the lower diagram.

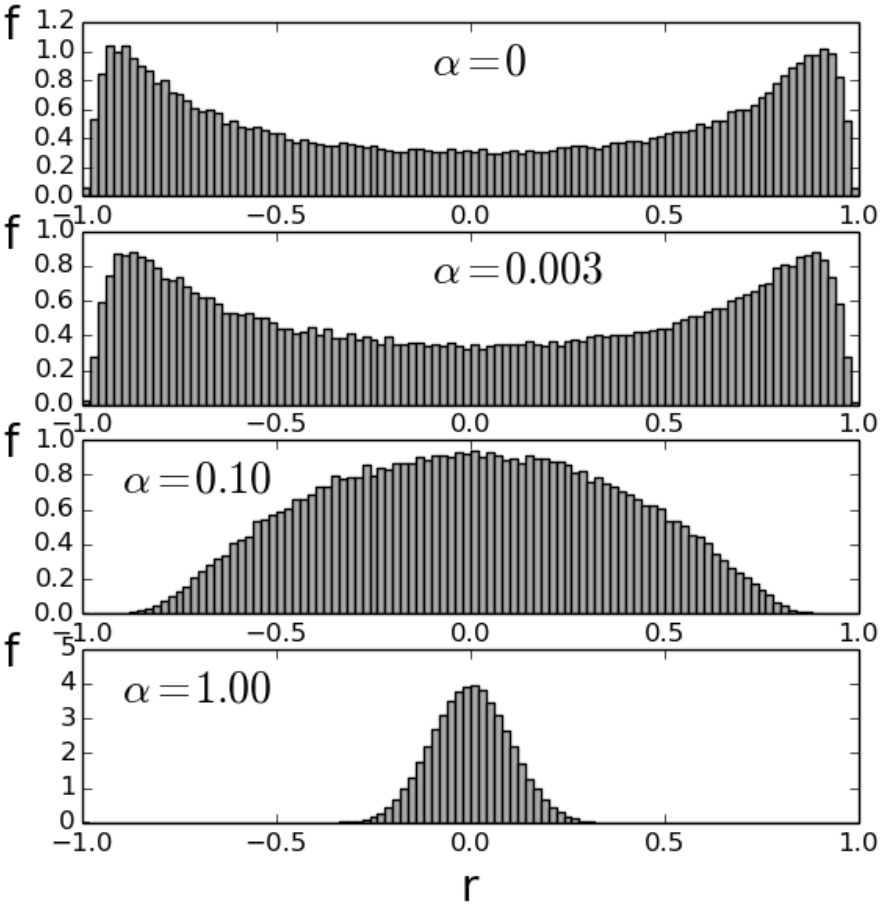


Figure 3: Probability distributions of the sample correlation coefficient, r, for the four time series whose spectra are shown in Figure 1.

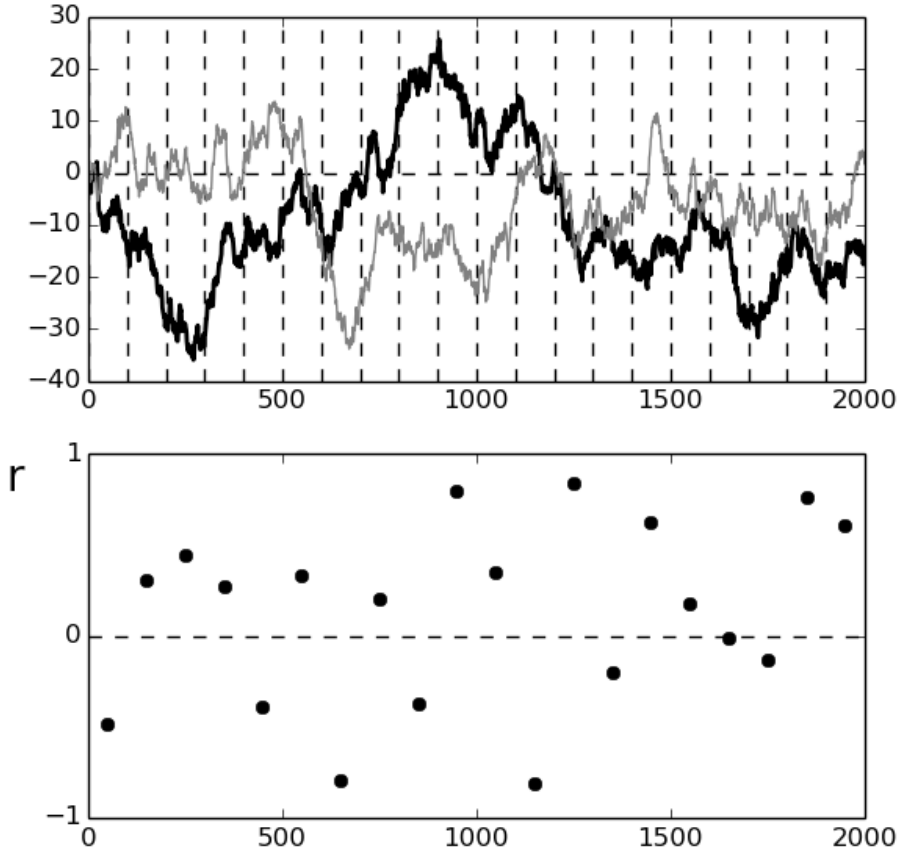


Figure 4: Two synthetic time series of ARMA(1,0) processes with $\alpha = .003$. The 2000-long interval was subdivided into twenty, 100-long, non-overlapping intervals shown by the vertical dashed lines. In each interval the correlation between the two time series samples was evaluated. The corresponding correlation coefficients are plotted in the lower diagram.

Spurious Regression and Climate

This section reproduces the author's paper in Energy and Environment, (Reid, 2017).

Introduction

In recent times much has been made of the apparent rising trend in global average temperature commonly attributed to increased greenhouse gas concentrations in the atmosphere. At issue is whether this trend is a real, deterministic trend or whether the observed variations are merely the outcome of a random process.

Hasselmann (1976) proposed a stochastic model of climate variability wherein slow changes in climate are explained as the integral response to continuous random excitation by short period 'weather' disturbances. Thus intrinsic quantities such as temperature are the outcome of the integration by natural processes of quasi-random, extrinsic quantities such as heat. As a consequence, such measurements can be regarded as the outcome of a stochastic process and can be expected to exhibit a power law spectrum with negative exponent due to such integrating effects. The best known and simplest example of such a process is the 'random walk' obtained when white noise is integrated or summed. It has a power law spectrum with an index of -2 as described by (105).

Here we examine the HadCRUT4 data set of 166 annual values of global average temperature from 1850 to 2015 inclusive (Morice et al., 2012).[1] There are a number of such global temperature data sets available, e.g. those from GISS, NOAA and BEST. Statistically they are almost identical. HadCRUT4 was chosen because it was the longest of these.

The data were analyzed using

1. a deterministic model in which each temperature value is regarded as a deterministic function of elapsed time plus a measurement error, and

2. a stochastic model whereby each temperature value is regarded as a function of preceding values plus an innovation.

The deterministic model

A deterministic model typically comprises a linear function of one or more functions of the explanatory variable plus a random element. In this case the explanatory variable is time. The parameters are estimated by minimizing the sum of squares of the differences between an estimate and the true values of the sample, the residuals. This is the ordinary least squares (OLS) method. It is based on the assumption that a deterministic relationship with the explanatory variable does exist and that the random elements at different times have zero mean and are independent of one another.

The HadCRUT4 time series values were fitted with a straight line by the OLS method of linear regression, i.e. the model

$$y_t = a_0 + a_1 t + \xi_t \tag{108}$$

was fitted to the data and is shown as the straight line in Figure 5(a)

The fitting of a function by OLS regression requires that the sequence of residuals $\{\xi_t\}$ in (108) or $\{\xi'_t\}$ in (108) be unselfcorrelated. Clearly that is not the case for (108) where a sinusoidal function or 'multidecadal oscillation' would remain after removal of the linear trend. For this reason a sinusoid of arbitrary phase was included in the model of equation (109):

[1] Downloaded from `http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time_series/` on 12/4/16.

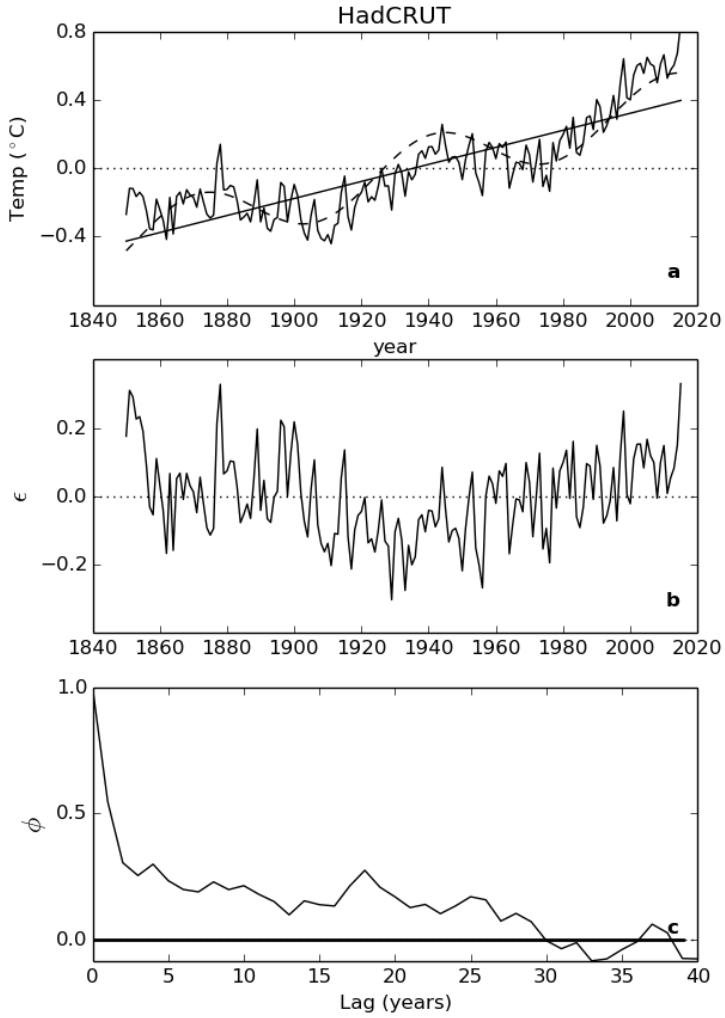


Figure 5: The deterministic model: (a) The HadCRUT time series. The straight solid line shows the linear trend of temperature vs. time. The dashed line shows the multiple regression fit of a linear trend plus a sinusoid. (b) Residuals from the time series of the linear trend plus sinusoid. (c) The autocorrelation function of the residuals, ϕ and the p-value of Ljung–Box Q-statistic (thick line). The p-value is zero at all lags.

	coef	std err	t	P>|t|	95.0% Conf. Int.
$\hat{a}_0$	-0.0149	0.010	-1.508	0.134	<-0.035, 0.005>
$\hat{a}_1$	0.0050	0.000	24.394	0.000	<0.005, 0.005>
$\hat{a}_2$	0.1210	0.014	8.841	0.000	<0.094, 0.148>
$\hat{a}_3$	0.0791	0.014	5.569	0.000	<0.051, 0.107>

Table 2: Coefficient estimates for the deterministic model of equation (109). Standard error, t-value, p-value and 95% confidence intervals are shown for each.

Test		Q-statistic	at lag	probability
Ljung–Box	min	92.665	3	0
Ljung–Box	max	232.052	39	0

Table 3: Testing the residuals of regression model of equation (109) for self-correlation. Minimum and maximum values of the Ljung–Box test statistics and their corresponding probabilities for a maximum lag of 40.

$$y_t = a_0 + a_1 t + a_2 \cos(\omega t) + a_3 \sin(\omega t) + \xi'_t \tag{109}$$

The angular frequency, ω, was chosen by trial and error and corresponded to a period of 70 years. The estimates $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$ and $\hat{a}_3$ of the model parameters a_0, a_1, a_2 and a_3 are shown in Table 2.

The sequence of residuals, $\{\xi'_t\}$, is shown in Figure 5b and its autocorrelation function (ACF)[2] in Figure 5c. Even to the naked eye there appears to be a systematic positive tendency in the ACF out to lag 30. There is a statistical test which can be used to determine whether the non-zero values of the ACF at non-zero lags are significant or just due to chance. It is the Ljung–Box test (Ljung and Box, 1978). The results obtained when fitting equation (109) to the HadCRUT4 data are shown in Table 3.

The probabilities listed in Table 3 are so small that we can reject the null hypothesis that the non-zero ACF values are purely random. Equation (109) is not a good fit to the data. It can be rejected at a very high level of significance.

The stochastic model

An ARMA(1,2) model was fitted to the HadCRUT4 time series using the *Python* statistical package: *statsmodels.tsa.arima-model.ARMA*. The package's *css-mle* option was selected whereby the conditional sum of squares likelihood was maximized and its values used as starting values for the computation of the exact likelihood via a Kalman filter.

The fitted model was thus

$$y_t = a_1 y_{t-1} + \epsilon_t + b_1 \epsilon_{t-1} + b_2 \epsilon_{t-2} + c \tag{110}$$

where a_1, b_1, b_2 and c were parameters to be fitted, and the $\{\epsilon_t\}$ were independent, identically distributed random variables with zero mean. The orders, $p = 1$ and $q = 2$, were found by trial and error, i.e. as the smallest values which resulted in unselfcorrelated residuals.

Note that the parameter c is similar to the coefficient a_1 in (108) and (109). Setting $a_1 = 1$ and the other coefficients in (110) to zero for the moment gives

$$y_t = y_{t-n} + nc \tag{111}$$

[2]The ACF is the covariance function normalized by dividing it by the variance

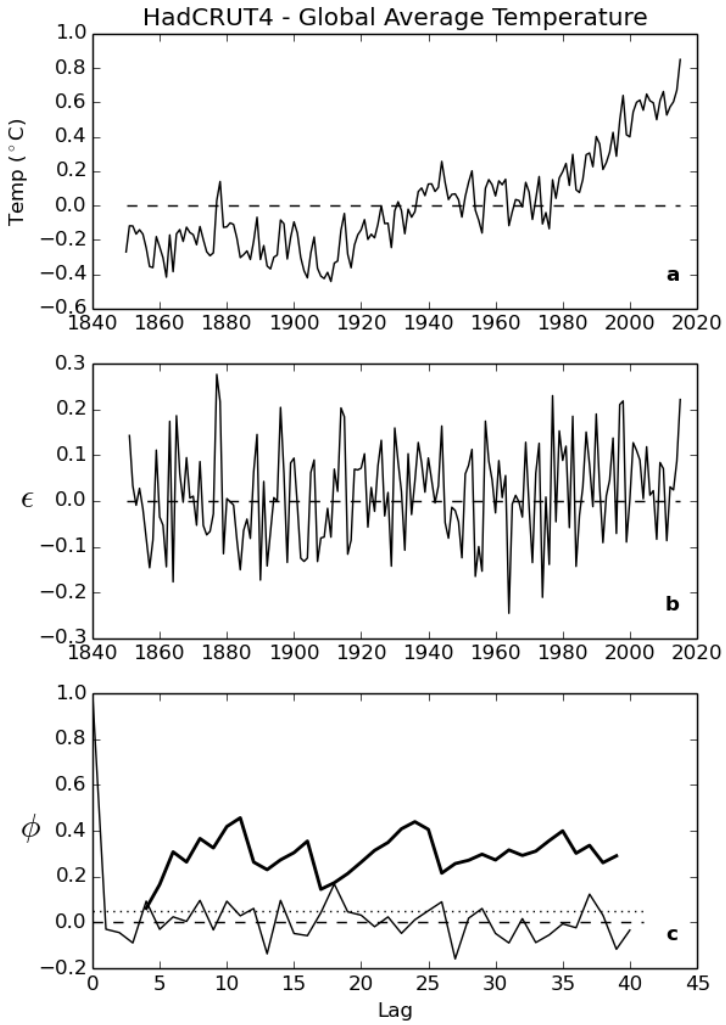


Figure 6: The stochastic method: (a) The HadCRUT time series. (b) Residuals from fitting an ARMA(1,2) model, (c) The autocorrelation function of the residuals, ϕ and the p-value of Ljung–Box Q-statistic (thick line).

	coef	std err	z	P>|z|	95.0% Conf. Int.
$\hat{a}_1$	0.9955	0.006	176.629	0.000	<0.984, 1.007>
$\hat{b}_1$	-0.4068	0.074	-5.490	0.000	<-0.552, -0.262>
$\hat{b}_2$	-0.2276	0.067	-3.379	0.001	<-0.360, -0.096>
$\hat{c}$	0.0736	0.351	0.210	0.834	<-0.615, 0.762>

Table 4: Coefficient estimates for the stochastic model of equation (110). Standard error, z-value, p-value and confidence limits are shown for each.

Test		statistic	at lag	probability
Ljung–Box	min	3.093	3	0.079
Ljung–Box	max	39.843	39	0.345

Table 5: Testing the residuals of the ARMA(1,2) model for self-correlation. Minimum and maximum values of the Ljung–Box test statistics and their corresponding probabilities for a maximum lag of 40.

so that y_t becomes a deterministic linear function of the elapsed time, $n\Delta t$. For this reason c is known as the 'drift term'. It is a deterministic element in an otherwise stochastic model.

The estimates $\hat{a}_1$, $\hat{b}_1$, $\hat{b}_2$ and $\hat{c}$ of the model parameters a_1, b_1, b_2 and c are shown in Table 4.

The most important feature of Table 4 is the small value and large confidence interval of the drift term estimate, $\hat{c}$. It is not significantly different from zero. Unlike the deterministic model, stochastic modelling indicates that there is no significant drift in the HadCRUT4 time series of global average temperature.

The sequence of residuals, $\{\epsilon_t\}$, is shown in Figure 6b and its autocorrelation function in Figure 6c. The ACF values at non-zero lags appear to be randomly distributed on either side of zero. As before, the Ljung–Box test was used to see if the null hypothesis that the residuals are unselfcorrelated can be rejected. The results are shown in Table 5.

None of the probabilities listed in Table 5 lie below the critical value of 0.05 and so there is no reason to reject the null hypothesis that the non-zero values of the ACF are due entirely to chance. The ARMA(1,2) model is a very good fit to the HadCRUT4 time series.

The Frequency Domain

Figure 7 shows the ARMA(1,2) spectral estimate of the HadCRUT4 time series plotted using logarithmic scales (thick dashed line). Also shown is the periodogram of the sample (thin line). The autoregressive coefficient, $\hat{a}_1$, in Table 4 is very close to one. In the nomenclature of the previous section, $\alpha = 1 - \hat{a}_1 = .0045$ – a small value indicating that the time series is red noise. This then is the reason for the f^{-2} trend at low frequencies (dashed line) and for the apparent regression with elapsed time shown by the Deterministic Model above. The HadCRUT4 time series of global average temperature is red noise, and the apparent regression is spurious.

As a consequence of the above-discussed whiteness tests confirming the absence of self-correlation of the residuals, this spectral estimate is optimal. There can be no peak, trough or trend in the spectrum other than those depicted in Figure 6 because this would require further poles and/or zeros in the z-plane which are not included in the ARMA model. Such extra poles or zeros, if unaccounted for, would inevitably lead to self-correlation of the sequence of residuals which would then have failed the Ljung–Box test; the 70-year "multidecadal oscillation" of equation (109) is therefore also spurious.

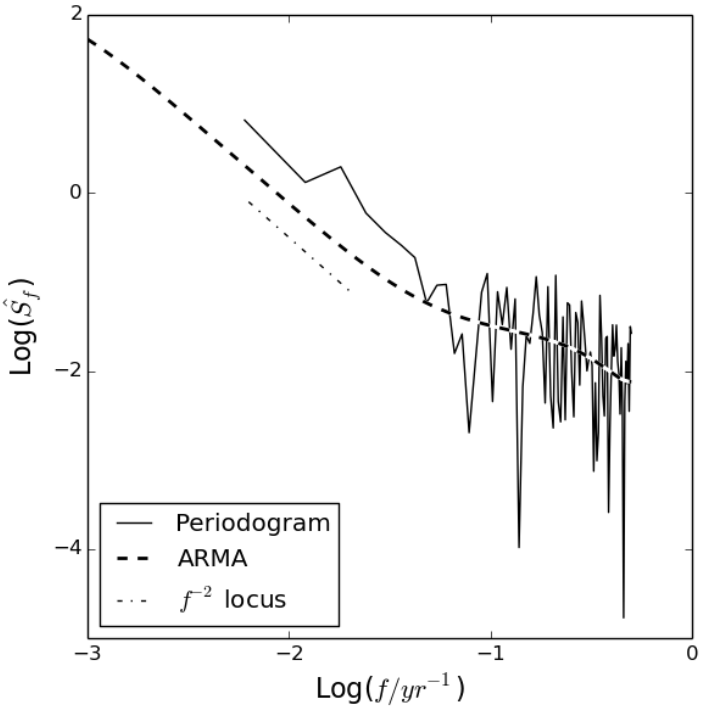


Figure 7: The variance density spectral estimate, $\hat{S}_f$, vs. frequency, f, of the ARMA(1,2) model fitted to the HadCRUT time series (thick dashed line). The thin line shows the periodogram of the time series.

The *Python* code used to generate Figure 6 and Figure 7, *arma.py*, can be found in the Appendix.

Implications

A rigorous, stochastic, discrete time theory of spectral analysis has been developed without making any assumptions about continuity and differentiability. The estimation of the variance density spectrum of a time series only applies to a range of time and frequency scales defined by the sampling period and the span of the sample itself. In this way it is possible to gain insights into underlying processes such as integration and convolution and to detect the significance of sinusoidal components by the accepted hypothesis testing methods of frequentist statistics. In particular the validity of a particular model can itself be assessed by testing the whiteness of the residuals using the Ljung-Box test.

It has been demonstrated (equation 57) that the periodogram of a sample which is the outcome of an ARMA process is a consistent estimator of the ensemble spectrum of the process. This contradicts the widespread assumption that the periodogram is not a consistent estimator and renders traditional methods of windowing unnecessary. In fact any such windowing is likely to subvert whiteness testing of the residuals and should be avoided.

The pitfalls of spurious regression against time and the spurious correlation of sequences with one another have been shown to be due to the concentration of variance at the low frequency end of the spectrum rather than to the non-stationarity of the sample, i.e. to the "redness" of the spectrum. Yule's "nonsense-correlations" are a property of red noise.

The process which gives rise to red noise is widely found in engineering and in nature. In electronics it occurs when noise is fed through an integrator as with the bass control of an audio amplifier. In the natural world it occurs when matter or energy is stored, e.g. when water from random rainfall events is stored in a lake, dam or river catchment. Most importantly, it occurs for temperature whenever heat is stored according to Fourier's equation (96) with a spectrum given by equation (103). Statistically it is a particular sort of Markov process termed a "centrally biased random walk".

The small increase in global average temperature over the last 166 years is not a trend and it is not likely to continue. It is red noise. So-called "multidecadal oscillations" are also red noise. Any correlation between global average temperature and other environmental quantities is likely to be spurious.

Appendix

arma.py

```
""" arma.py fits an ARMA(p,q) model to the time series specified
        by the input function. Values of p and q are optimized by
        trial and error by minimizing the residual variance or the
        Ljung--Box pmin value. These are appended to a 'Summary'
        output file.
        Spectra and time domain graphs are saved in 'spectrum'
        and 'resid' folders.
        Periodograms are saved in numerical form in a
        'periodogram' folder.
        Statistics are saved in a 'stats' folder.
        These four subfolders must be created before running arma.py.
        A text file 'Summary' is created and appended after each run.
        Call: python arma.py p q where (p,q) is the order.
"""
import numpy as np
import matplotlib.pyplot as plt
from matplotlib import gridspec
from scipy.fftpack import fft
from scipy import stats
import statsmodels.tsa.stattools as sts
import statsmodels.tsa.arima_model as sta
import sys

def gat():
        #
        #   read in temperature data
        #
        Ns = 166
        f = open('HadCRUT.4.4.0.0.annual_ns_avg.txt', 'r')
        title = 'HadCRUT'
        x = np.zeros(Ns,float)
        y = np.zeros(Ns,float)
        dy = np.zeros(Ns,float)
        i = 0
        for line in f:
                myline = line.split()
                x[i] = float(myline[0])
                y[i] = float(myline[1])
                i+=1
        f.close()
        dt = 1.
        ylabel=r'Temp ($^\circ$C)'
        xlabel='year'
        title = 'HadCRUT4 - Global Average Temperature'
        """
        fig=plt.figure(num=99,figsize=(6.4,3.2))
        plt.plot(x,y,color='k',linewidth=1)
        plt.hlines(0,x[0],x[-1],colors='k',linestyles='dashed')
        plt.title(title)
        plt.ylabel(ylabel)
        plt.savefig('rawdata')
        plt.show()
        """
        return x,y,Ns,dt,xlabel,ylabel,title

if len(sys.argv)<>3:
        print 'Wrong number of arguments:'
```

```
        print 'Should be: arma.py p q'
        quit()
p = int(sys.argv[1])
q = int(sys.argv[2])
outputcode='{0:1d}{1:1d}'.format(p,q)
ab = [0 for i in range(p+q)]
print "ab = ",ab

x,y,Ns,dt,xlabel,ylabel,title = gat()

df = 1./(Ns*dt)
fs = 1./dt                          #sampling frequency
Nf = 0.5/dt                          #Nyquist frequency
xbar = np.mean(x)
ybar = np.mean(y)
y = y-ybar
yvar = np.var(y)

print title
print 'Ns: {0:4d}, dt:{1:4.0f}, Log(Nf):{2:6.2f}'\
.format(Ns,dt,np.log(Nf)/np.log(10.))
#
#
#
#
fig=plt.figure(num=1,figsize=(6.4,6.4))
gs = gridspec.GridSpec(1, 1, height_ratios=[1])
#
#   prepare log-log periodogram
#
Y = fft(y)
X = np.linspace(0,Nf,Ns/2+1)
Y2 = 2*(np.abs(Y[:Ns/2+1])/Ns)**2
XX = X[1:]
Pgm = Y2[1:]/df
Pgm[Ns/2-1]=0.5*Pgm[Ns/2-1]
lX = np.log(XX)/np.log(10.)
lP = np.log(Pgm)/np.log(10.)

ax=plt.subplot(gs[0])
plt.plot(lX,lP,linewidth='1',color='k',label='Periodogram')
plt.xlabel(r'Log($f/yr^{-1}$)',fontsize=16)
plt.ylabel(r'Log($\hat{S_f}$)',fontsize=16)
#
#        Compute ARMA coefficients using statsmodels
#
model2 = sta.ARMA(y,(p,q))
results2 = model2.fit(start_params=ab, trend='nc', disp = False)
#
resultspq='stats/results{0:s}'.format(outputcode)
f=open(resultspq,'w')
f.write(str(results2.summary(0.05)))
f.close()
#
epsvar = results2.sigma2
#
#        save parameters
#
paramspq ='stats/{0:s}'.format(outputcode)
f = open(paramspq,'w')
for num in results2.arparams: f.write('{0:f} '.format(num))
for num in results2.maparams: f.write('{0:f} '.format(num))
```

```
f.close()
print results2.params
#
#        plot ARMA population spectrum, S, with confidence limits.
#
import cmath
Np = 1000
lf = np.zeros(Np,float)
S = np.zeros(Np,float)
i = 0
deltaf = Nf/Np
integral = 0.
dof = 2
lNf=np.log(Nf)/np.log(10)                #log(Nyquist frequency)
lfarray = np.linspace(-4,lNf, num=Np)
for lf in lfarray:
        f=10.**lf
        phi = 2*np.pi*f/fs
        z = complex(np.cos(phi),np.sin(phi))
        den = 1.
        for ip in range(p):
                den-=results2.arparams[ip]*z**(ip+1)
        num = 1.
        for iq in range(q):
                num+=results2.maparams[iq]*z**(iq+1)
        S[i] = 2*dt*epsvar*(abs(num/den))**2
        integral += S[i]*deltaf
        i+=1
print'yvar, integral = ',yvar, integral
lS = np.log(S)/np.log(10.)
print lfarray.shape,S.shape
plt.plot(lfarray,lS,linewidth=2,color='w')
plt.plot(lfarray,lS,linewidth=2,color='w',linestyle = '-')
plt.plot(lfarray,lS,linewidth=2,color='k',\
linestyle = '--',label='ARMA')
#
#        plot power law locus
#
pllx=np.zeros(2,float)
plly=np.zeros(2,float)
pllx[0]=-2.2
plly[0]= -0.1
pllx[1]=-1.7
plly[1]=-1.1
plt.plot(pllx,plly,color='k',linestyle='-.',label=r'$f^{-2}$ locus')
#
#        Save spectral details
#
periodogrampq='periodograms/{0:s}'.format(outputcode)
f = open(periodogrampq,'w')
f.write('   log f    log P        f        T    log S\n')
for i in range(XX.size):
        phi = 2*np.pi*XX[i]/fs
        z = complex(np.cos(phi),np.sin(phi))
        den = 1.
        for ip in range(p):
                den-=results2.arparams[ip]*z**(ip+1)
        num = 1.
        for iq in range(q):
                num+=results2.maparams[iq]*z**(iq+1)
        S[i] = 2*dt*epsvar*(abs(num/den))**2
        lS[i] = np.log(S[i])/np.log(10.)
```

```
        f.write('{0:8.3f} {1:8.3f} {2:8.5f} {3:8.4f} {4:8.4f}\n'\
.format(lX[i],lP[i],XX[i],1/XX[i],lS[i]))
f.close()
#
#
plt.legend(loc='lower left')
plt.xlim(-3,0)
plt.xticks([-3,-2,-1,0])
plt.ylim(-5,2)
plt.yticks([-4,-2,0,2])
plt.savefig('spectrum/{0:s}'.format(outputcode))
plt.show()
#................................................
#
#        residuals
#
fig=plt.figure(num=2,figsize=(6.4,9.6))
gs = gridspec.GridSpec(3, 1, height_ratios=[1,1,1])
#
#
#
ax=plt.subplot(gs[0])
plt.plot(x,y,color='k',linewidth=1)
plt.hlines(0,x[0],x[-1],colors='k',linestyles='dashed')
plt.title(title)
plt.ylabel(ylabel)
ax.annotate('a',xy=(.95,.1),xycoords='axes fraction', \
xytext=(.95,.1), textcoords='axes fraction',weight='bold')
#
#   find ARMA residuals.
#
ax = plt.subplot(gs[1])
eps = results2.resid
plt.plot(x[1:],eps[1:],color='k')
plt.hlines(0,x[0],x[-1],colors='k',linestyles='dashed')
string = r'$\epsilon$'
ax.annotate(string,xy=(-.12,.48),xycoords='axes fraction', \
xytext=(-.12,.48), textcoords='axes fraction',weight='bold',\
fontsize='20')
ax.annotate('b',xy=(.95,.1),xycoords='axes fraction', \
xytext=(.95,.1), textcoords='axes fraction',weight='bold')
#
#   get acf of residuals
#
nlags = 40
ax=plt.subplot(gs[2])
epsacf,lbvalue,pvalue = sts.acf(eps, unbiased=False, \
nlags=nlags, qstat=True)
iepsacf = range(len(epsacf))
plt.plot(iepsacf,epsacf,linewidth='1',color='k')
plt.hlines(0,0,len(epsacf),linestyles="--")
plt.hlines(0.05,0,len(epsacf),linestyles=":")
plt.xlabel("Lag")
ax.annotate('c',xy=(.95,.1),xycoords='axes fraction',\
xytext=(.95,.1), textcoords='axes fraction',weight='bold')
#
string = r'$\phi$'
ax.annotate(string,xy=(-.12,.48),xycoords='axes fraction', \
xytext=(-.12,.48), textcoords='axes fraction',weight='bold',\
fontsize='20')
#
#               Ljung--Box
```

```
#
truep = np.zeros(nlags,float)
pmin= 1.1
pmax = -1
lagmin=0
ip = range(nlags)
aip = np.array(ip)
for lag in range(p+q,nlags):
        Q = lbvalue[lag]
        tpl=1-stats.chi2.cdf(Q,lag-(p+q))
        if tpl<pmin:
                pmin=tpl
                Qpmin=Q
                lagmin=lag
        if tpl>pmax:
                pmax=tpl
                lagmax=lag
        truep[lag] = tpl
plt.plot(aip[p+q:],truep[p+q:], color='k', linewidth='2')
print 'pmin = {0:5.3f} at lag {1:5d}'.format(pmin,lagmin)
print 'pmax = {0:5.3f} at lag {1:5d}'.format(pmax,lagmax)
plt.savefig('resid/{0:s}'.format(outputcode))
plt.show()
#
#        Create Summary file if none exists and append
#
if True:
        Smmry = 'Summary'
        Exists=True
        try:
                f = open(Smmry,'r')
        except IOError:
                Exists = False
        else:
                f.close()
        if not Exists:
                f = open(Smmry,'w')
                f.write('{0}{1}'.format(title,'\n\n'))
                header1 = '                          Ljung--Box\n'
                f.write(header1)
                header2 = ' pq      resid var         Q      pmin\n'
                f.write(header2)
                f.close()
        f = open(Smmry,'a')
        f.write(' {0:s}  {1:13.3e}{2:10.4f}{3:10.4f}\n'\
        .format(outputcode,epsvar,Qpmin,pmin))
        f.close()
print
print
```

References

Abramowitz, M. and I. A. Stegun (1965). *HANDBOOK OF MATHEMATICAL FUNCTIONS*. 180 Varick Street, New York, N.Y. 10014: Dover Publications, Inc.

Akaike, H. (1970). Statistical predictor identification. *Ann. Inst. Statis. Math. 22*, 233–217.

Bartlett, M. S. (1948). Smoothing periodograms from time-series with continuous spectra. *Nature 161*, 686–687.

Blackman, R. B. and J. W. Tukey (1958). *THE MEASUREMENT OF POWER SPECTRA*. 180 Varick Street, New York, N.Y. 10014: Dover Publications, Inc.

Bracewell, R. N. (1986). *THE FOURIER TRANSFORM AND ITS APPLICATIONS*. McGraw-Hill Book Co-Singapore: McGraw-Hill, Inc.

Fisher, R. (1934). *Statistical Methods for Research Workers*. Tweeddale Court, Edinburgh: Oliver and Boyd.

Granger, C. W. J. and P. Newbold (1974). Spurious regression in econometrics. *Journal of Econometrics 2*, 111–120.

Gull, S. F. and G. J. Daniell (1978). Image reconstruction from incomplete and noisy data. *Nature 272*, 686–690.

Hamilton, E. J. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.

Hannan, E. J. (1960). *TIME SERIES ANALYSIS*. London: Methuen and Co.

Hasselmann, K. (1976). Stochastic climate models, part I, theory. *Tellus XXVIII*(6), 473–485.

Ljung, G. M. and G. E. P. Box (1978). On a measure of the lack of fit in time series models. *Biometrika 65*, 297–303.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 dataset. *Journal of Geophysical Research 117*, D08101.

Nelson, C. R. and H. Kang (1984). Pitfalls in the use of time as an explanatory variable in regression. *Journal of Business and Economic Statistics 2*, 73–82.

Popper, K. R. (1962). *CONJECTURES AND REFUTATIONS*. New York: Basic Books.

Reid, J. (2017). There is no significant trend in global average temperature. *Energy & Environment 28*(3), 302–315.

Reid, J. S. (1979). Confidence limits and maximum entropy spectra. *J. Geophys. Res. 84*(A9), 5289–5301.

So, H. C., Y. T. Chan, Q. Ma, and P. C. Ching (1999). Comparison of various periodograms for sinusoid detection and frequency estimation. *IEEE Trans. Aerospace Electronic Systems 35*(3), 945–951.

Woit, P. (2006). *Not Even Wrong: The Failure of String Theory and the Search for Unity in Physical Law*. New York: Basic Books.

Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series?–a study in sampling and the nature of time series. *J. Royal Statistical Society 89*, 1–63.

Bibtex Citation

```
@inbook{ReidTFC,
        AUTHOR = {Reid, John},
        YEAR = {2019},
        TITLE = {The Fluid Catastrophe},
        PUBLISHER = {Cambridge Scholars Publishing},
        ADDRESS = {Newcastle upon Tyne},
        ISBN = {1-5275-3206-2},
        pages = {31-86}
}
```