On the construction of synthetic time series
Version 1.3

T.Appourchaux

April 5, 2002

1 Goal

The purpose of this document is to set a reference for generating artificial time series simulating stellar power spectrum.

2 Background

2.1 Methodology

The generation of time series is based upon the formalism developed for digital electronic processing. For instance, it is well known that for reconstructing a given signal in time, the minimum sampling frequency is given by the spectral content that we wish to recover. For example, let's assume that the maximum frequency is ν_{max}, the sampling time corresponding is $1/2/\nu_{max}$, or the inverse of the Nyquist frequency (=$2\nu_{max}$). Other technical considerations such as aliasing, may come into play but we leave that to the digital-reconstruction-signal expert, or to the next version of the document.

When the Nyquist frequency is known, the number of points in the time series will define the frequency resolution from which we can generate the sampled spectrum. So for reconstructing a signal, it is not needed (it is superfluous...), to generate a highly resolved spectrum (corresponding to a long time series) when the time series itself is much shorter. The reconstruction will not be better.

2.2 Convention of Parseval's theorem

This is basically a theorem addressing energy conservation. Typically, projecting a function of time into an expansion of orthogonal function of time (sine and cosine) will not modify the energy (or the norm) of this function.

$$\int p d\nu = \sigma^2 \tag{1}$$

where p is the spectral density power of the process (in ppm$^2/\mu$Hz for example) and σ is the root mean square of the process in time.

For instance for white noise, we can write:

$$P_0 \Delta\nu = \sigma^2 \tag{2}$$

where P_0 is the power spectra density of the white noise (assumed to be independent of the observation), $\Delta\nu$ is the bandwidth over which this random process is observed and σ is the standard deviation of the random process in time.

In the same spirit, for a sine wave we have that the energy of the wave is given by:

$$\frac{A^2}{4} + \frac{A^2}{4} = \frac{A^2}{2} \tag{3}$$

The left hand side of the equation represents the sum of the 2 terms that appear in the positive and negative part of the power spectrum, while the right hand side is just the expression of the energy of the wave (or the square of the rms amplitude of the wave).

Now if we assume a lorentzian profile for a p mode peaking at A ppm$^2/\mu$Hz and a FWHM of Γ then the integrated power is $P = 2 \times \pi A \Gamma/2$. The factor 2 in front comes from the fact that we need to take into account the negative and positive side of the spectrum.

2.3 Mode amplitude in the literature

The theoretical mode amplitude computed by Houdek (1999) is the rms velocity or intensity amplitude derived from the energy in the mode. Therefore the peak amplitude in the power spectrum can be derived from the rms amplitude using Parseval's theorem and we have:

$$A = V_{rms}^2/\pi/\Gamma \tag{4}$$

This is assuming that the profile is lorentzian. It should be noted that the V_{rms} is usually derived from just $A \times \Gamma$ and the π seems to be forgotten... unless I made a mistake. To be discussed and verified.

3 Recipe for time series generation

Here are the steps needed for generating a time series obeying a prescribed power spectrum:

- Assume a Nyquist frequency (ν_{max}) or a sampling time ($\Delta t = 1/\nu_{max}$).
- Assume a length for the time series or the number of points in the power spectrum (N)

- The power spectrum is derived from whatever is known or assumed about the instrument or the star. The units of choice here are $\mathrm{ppm}^2/\mu\mathrm{Hz}$
- Generate the power spectrum for half the number of points ($N/2$).
- Multiply the spectral density by the size of the frequency bin ($1/N/\Delta t$) expressed in μHz if this is the unit of choice. This takes into account the length of the time series.
- Take the square root of the power to get the amplitude
- Multiply the amplitude spectrum by a complex random variable with each component having a normal distribution where the real (or imaginary part) has a mean of 0 and an rms of 0.5. The real and imaginary parts are independent of each other.
- The amplitude spectrum now obtained needs to be symmetrized because the time series is real, i.e. $F(-\nu) = F^*(\nu)$. We get the final Fourier spectrum.
- Invert the Fourier spectrum
- Check that the imaginary part is indeed negligible compared to the real part.

The final check is done by comparing the results given by Parceval's theorem. Let's assume that a noise of 1 $\mathrm{ppm}^2/\mu\mathrm{Hz}$ has been generated over a bandwidth of 16666 μHz (1 over 60 s). The expected rms value of the corresponding time series should be $\sqrt{1 \times 16666}$=129 ppm. So if $A = \Gamma = 1$, then the rms is $\sqrt{\pi}$=1.77 ppm

4 The Corot noise level

It does make sense to relate the 0.6 ppm in 5 days given in the Corot literature to 'real' amplitude in the time series. The reference document is 'Corot: scientific program and specifications, March 1998'.

The reference document mentions a signal-to-noise ratio of 15 as a baseline for reaching a frequency precision close to what would be obtained with a large signal-to-noise ratio (say 100). The signal-to-noise ratio of 15 combined with a typical solar-like amplitude of 2.5 ppm gives a noise level to be reached in 5 days (typical solar-like mode lifetime) of 0.18 $\mathrm{ppm}^2/\mu\mathrm{Hz}$, The typical noise is therefore in a 5-day frequency bin 0.64 ppm (0.18×2.3).

Now what does it mean for a mode? The peak amplitude in the power spectrum is 0.18×15=2.7 $\mathrm{ppm}^2/\mu\mathrm{Hz}$. Then the total power of this mode with a 5-day lifetime is $\pi \times 2.7 \times 2.3 = 19.5$ ppm^2, or 4.4 ppm in the time series. This latter number is nothing less but the 'usually quoted amplitude of the mode' time the square root of humble π ($4.4 = 2.5 \times \sqrt{\pi}$). Hope this clarify the situation now.

5 A remark...

Please bear with me as this report was written under the influence of a flu, an insomnia, an amnesia and other neurotic behaviours.