STAT 854 - Final Project

Dave Riegert & Aaron Springford

April 11, 2012

Part 2 - Programming

For all of the following questions we have used the first column of the Black Forest Observatory barometric data that was gathered from the BMG microbarometer.

Question 1 - Plot raw data, mean & median, Gaussian?:

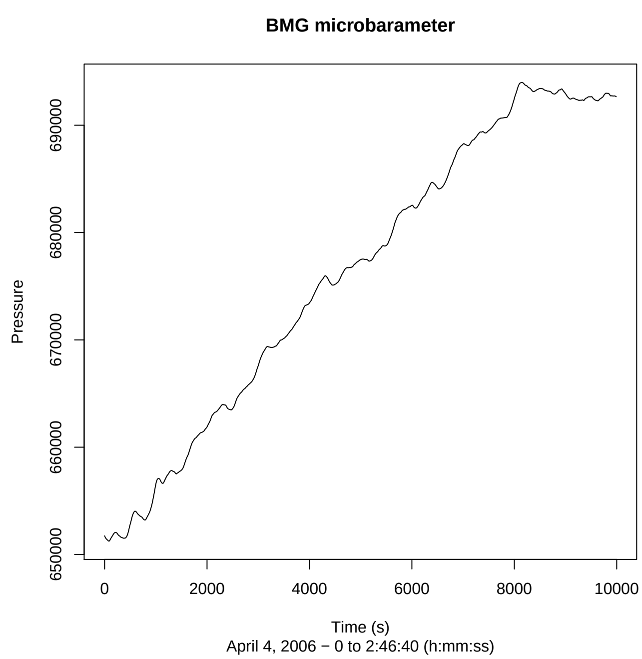

Figure 1: BMG microbarometer raw data

Estimating the mean and median of the data we obtained:

Mean: 677364

Median: 676209.583

Having the mean and the median so close together suggests that the distribution of the data is reasonably symmetric.

In order to determine if the data is Normal we first calculated the second through fourth moments as recommended:

$$
\begin{aligned}
\text{Second Moment: } & 0.0181 \\
\text{Third Moment: } & -0.0007 \\
\text{Fourth Moment: } & 0.00059
\end{aligned}
$$

Based on these moments we would conclude that the data is not Gaussian. A Normal quantile-quantile plot suggested the same conclusion (figure 2; notice the truncated tails of the data).

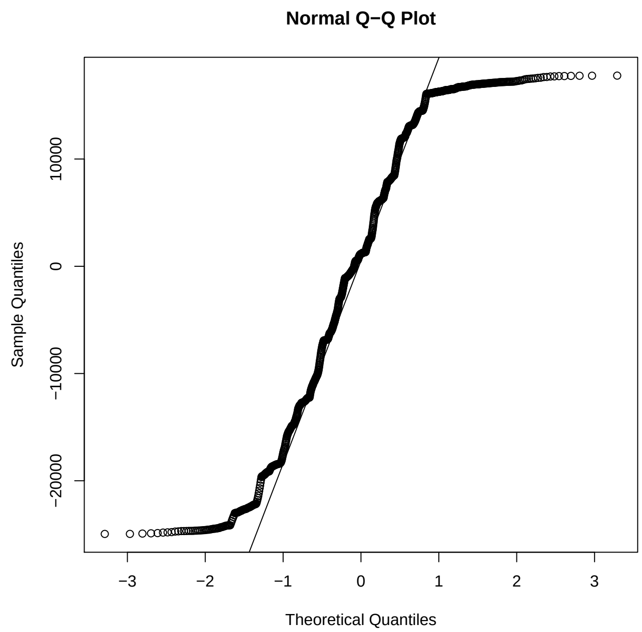

Figure 2: qqnorm plot of the data - qualitative test for normality

Finally, a Shapiro-Wilk normality test was performed. The test statistic value $W = 0.9182$ had a corresponding p-value of $p < 2.2 \times 10^{16}$, strongly rejecting the hypothesis that the data is Gaussian.

Question 2 - Bartlett Autocorrelations:

(See figure 3.) The standard Bartlett autocorrelations might lead some to conclude that this is a *long memory process*. However, this is a poor estimate of the autocorrelations, and gives a very different estimate than the multitaper-derived autocorrelations, as we will see later (question 6).

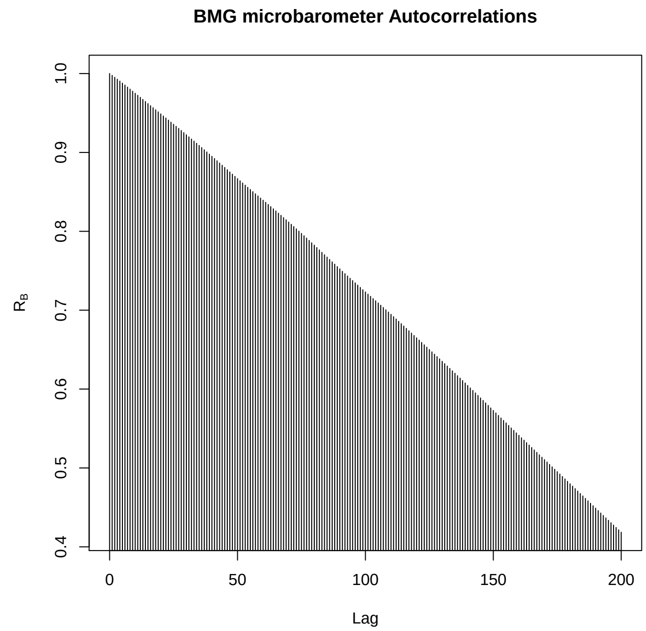

Figure 3: Standard Bartlett autocorrelations of demeaned data

Question 3 - Multitaper spectrum, sectioning with Hanning window:

The following figures plot the spectrum of the first 1000 samples, zero padded to $M = 2048$. We estimated the spectrum using the multitaper estimate, NW = 6, K = 5 (figure 4). With the Hanning windows we tried the arithmetic mean (figure 5), median (figure 6), and geometric mean (figure 7). The Hanning sectioning method had 5 data sections offset by 50 percent (i.e. 400 data points per estimate).

Comparing the estimates, we get rather similar results (figure 8).

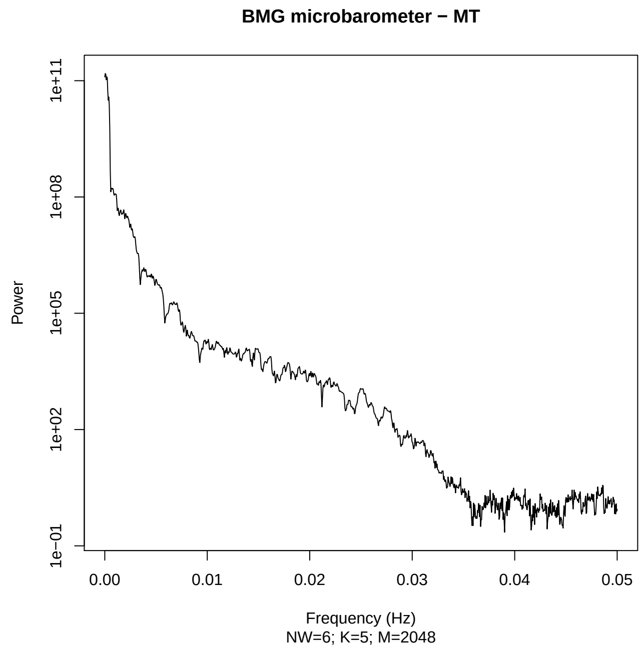

Figure 4: Spectrum using multitaper method

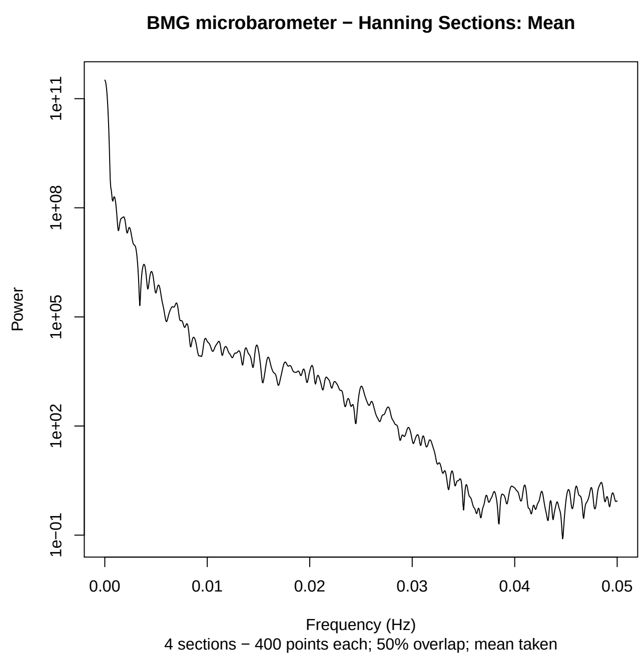

Figure 5: Sectioning approach using Hanning window - mean of all sections

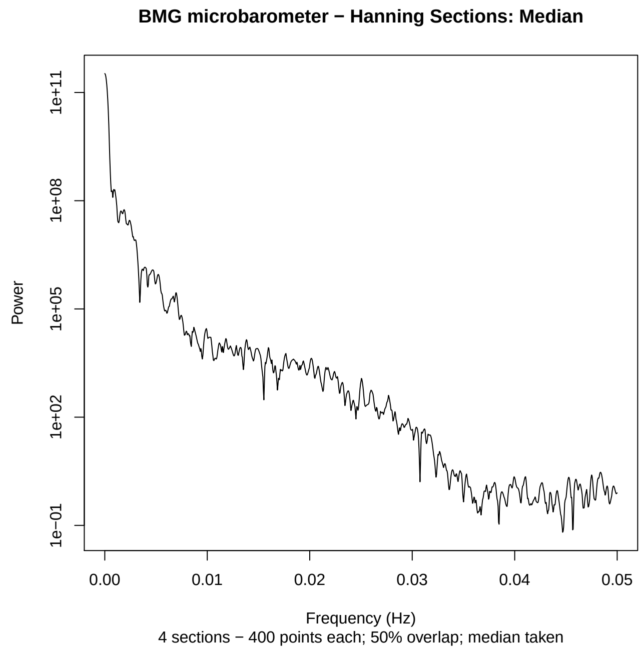

Figure 6: Sectioning approach using a Hanning window - median of each frequency of ensemble

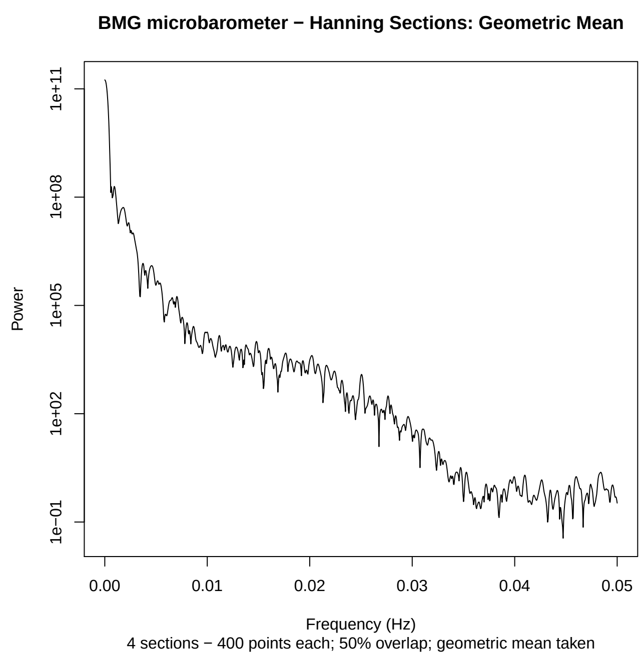

Figure 7: Sectioning approach using a Hanning window - geometric mean taken

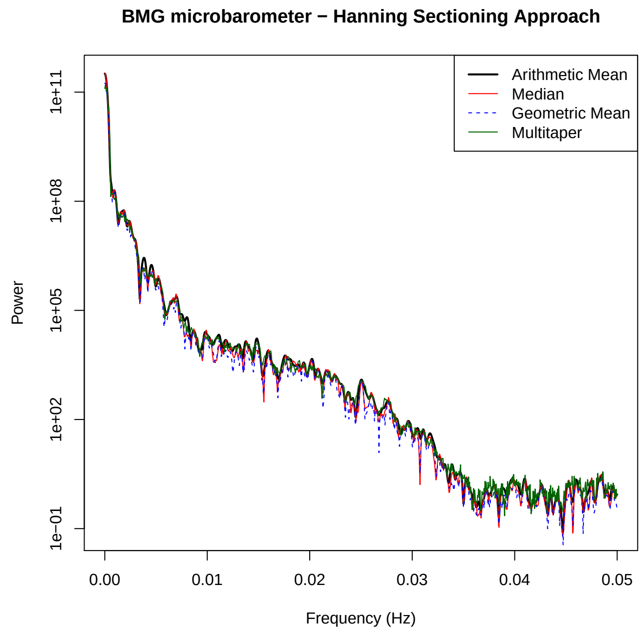

Figure 8: Comparison between multitaper and the three Hanning estimates

Question 4 - Multitaper NW=10, adaptive weighting

A multitaper with a time-bandwidth $NW = 10$, number of tapers $K = 20$, and adaptive weighting applied (figure 9). The weights were on the eigenspectra across all frequencies, rather than having different weights for each frequency-eigenspectrum combination. The adaptive weighting results in more power at low frequencies, a smoother spectrum at higher frequencies, and some interesting details at moderate frequencies.

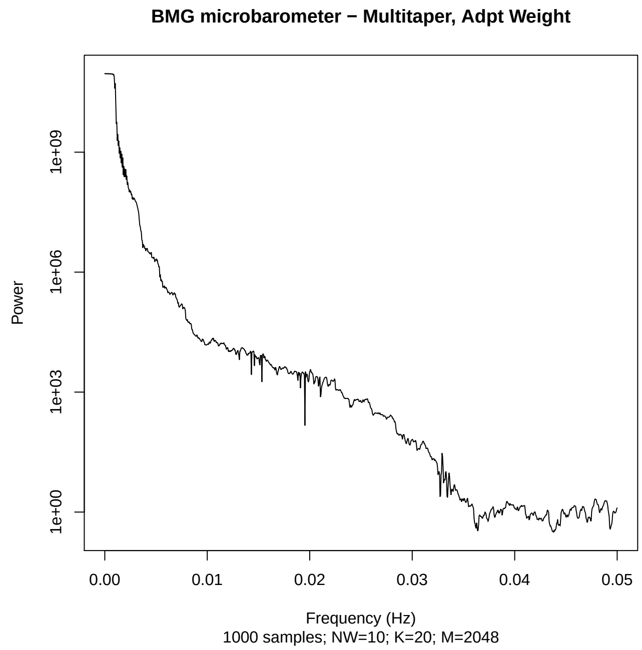

Figure 9: Multitaper with adaptive weighting

Question 5 - Approximate variance, estimate quantization noise power

We computed

$$I_S = \int_{-F_{Nyq}}^{+F_{Nyq}} \hat{S}(f)df$$

by summing over the FFT bins (as suggested) and compared with the sample variance. We also estimated the quantization noise power from the highest fifth of the spectrum frequencies (that is, frequencies > 0.04). We got values of

$$I_s = 1.846 \times 10^8$$
$$\sigma^2 = 1.867 \times 10^8$$
$$\text{QNP} = 0.988$$

We notice that the estimated variance from the spectrum is 98.8% of the sample variance for the first 1000 samples.

Question 6 - Autocorrelation through FFT of spectrum, compare with previous estimates

We took the Fourier transform of the multitaper estimate computed in question #5 and compared it with the original Bartlett estimates computed in question #2. The *long memory process* interpretation seems less likely considering the multitaper estimates (figure 10).

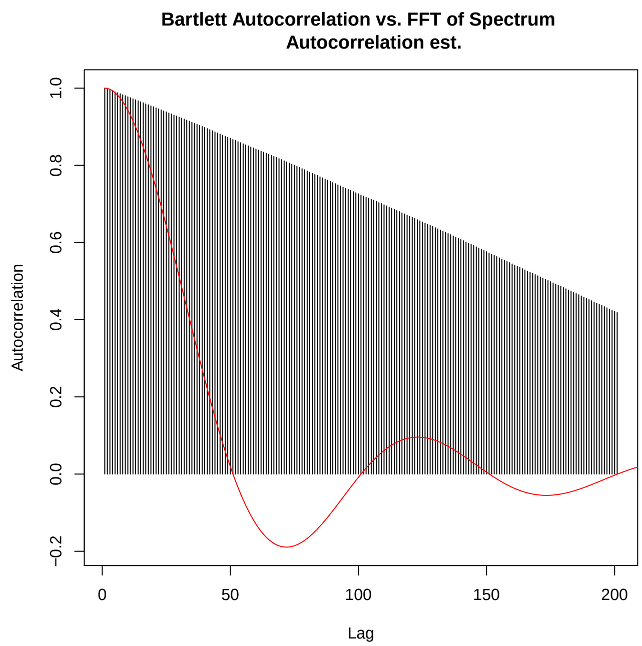

Figure 10: Bartlett Autocorrelations (black) compared with FFT of Multitaper Spectral estimate (red)

Question 7 - AR2 model of the process

We used a Yule-Walker approach in our next-step prediction to obtain prediction residuals of the form

$$y(t) = x(t) - \sum_{j=1}^{P} \alpha_j^{(P)} x(t-j)$$

We used an AR(2) model for prewhitening (that is, P=2). Our estimates of α were: $\alpha_1 = 0.999999999945615$, and $\alpha_2 = -0.000000000000455$. The resulting pre-whitened data (figure 11) have the large scale trend in figure 1 removed.

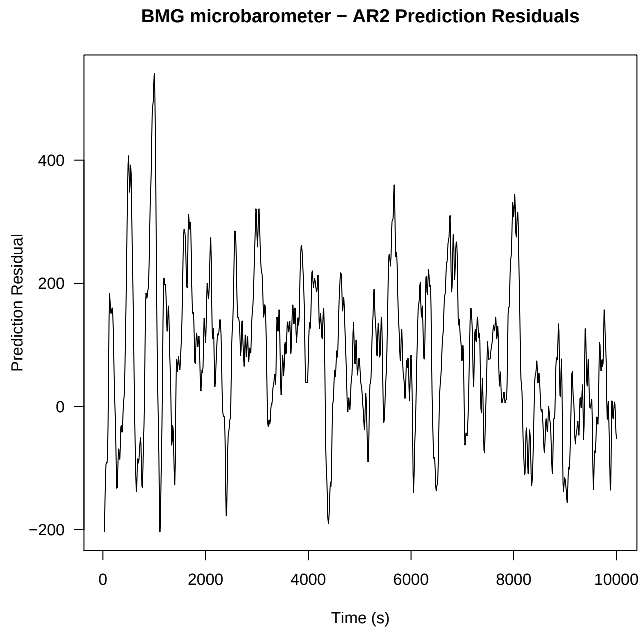

Figure 11: AR2 Prediction Residuals

Question 8 - Repetition of spectrum estimation with prewhitened data

We repeated the adaptive multitaper as in question 4 on the prewhitened data (NW=10, K=20, adaptive). The spectrum of the prediction residuals is in figure 12. We corrected this spectrum estimate by adding back the power attributable to the AR(2) portion of the original time series. If $\hat{S}_y(f)$ is the spectrum of the pre-whitened data, then we obtained the spectrum estimate

$$\hat{S}_x(f) = \frac{\hat{S}_z(f)}{|1 - \sum_{k=1}^{P} a_k e^{-i2\pi kf}|^2}$$

This corrected estimate is in figure 13, along with the original spectrum estimate for comparison. The transfer function for the AR(2) prediction error filter was obtained by dividing the pre-whitened spectrum $\hat{S}_y(f)$ by the corrected spectrum $\hat{S}_x(f)$ (figure 14).

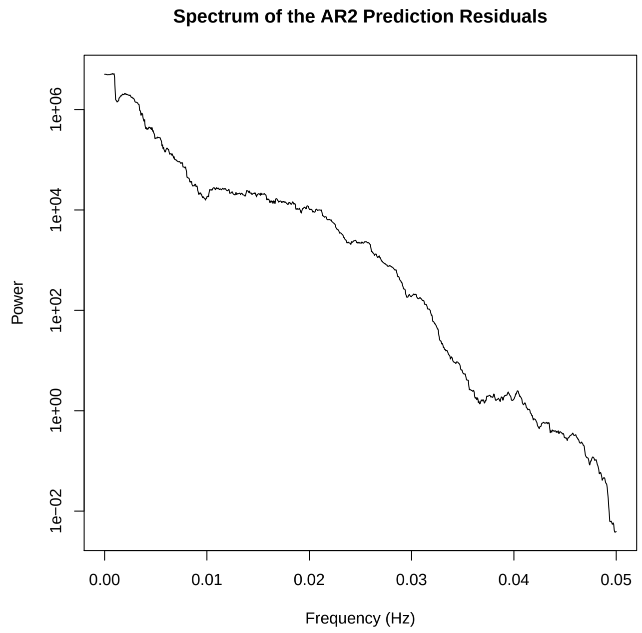

Figure 12: Spectrum of AR2 prediction residuals

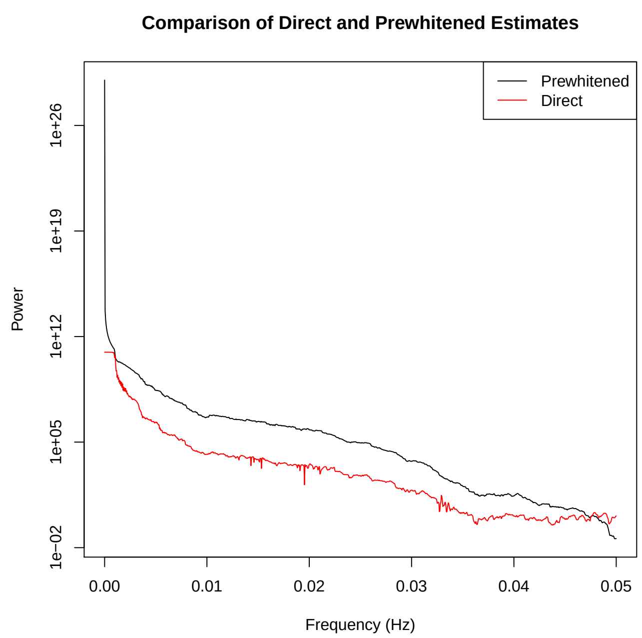

Figure 13: Comparison of the pre-whitened spectrum and the original spectrum

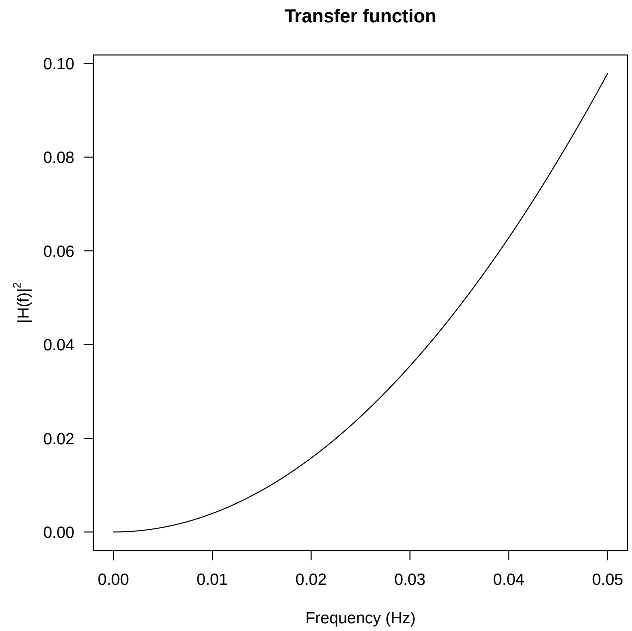

Figure 14: AR(2) transfer function

Question 9: Rough test of stationarity

As a rough test of stationarity, we computed the multitaper spectrum as in question 4 (NW=10, K=20, adaptive) on four non-overlapping data sections (1000 samples each). All of the spectra were plotted together for comparison (figure 15). It appears that on the timescale of 1000 samples, the data are not stationary, with some sections of data trading low-frequency power for mid-frequency power. Over a longer timescale than 1000 samples, the data may appear more stationary (although we did not check this).

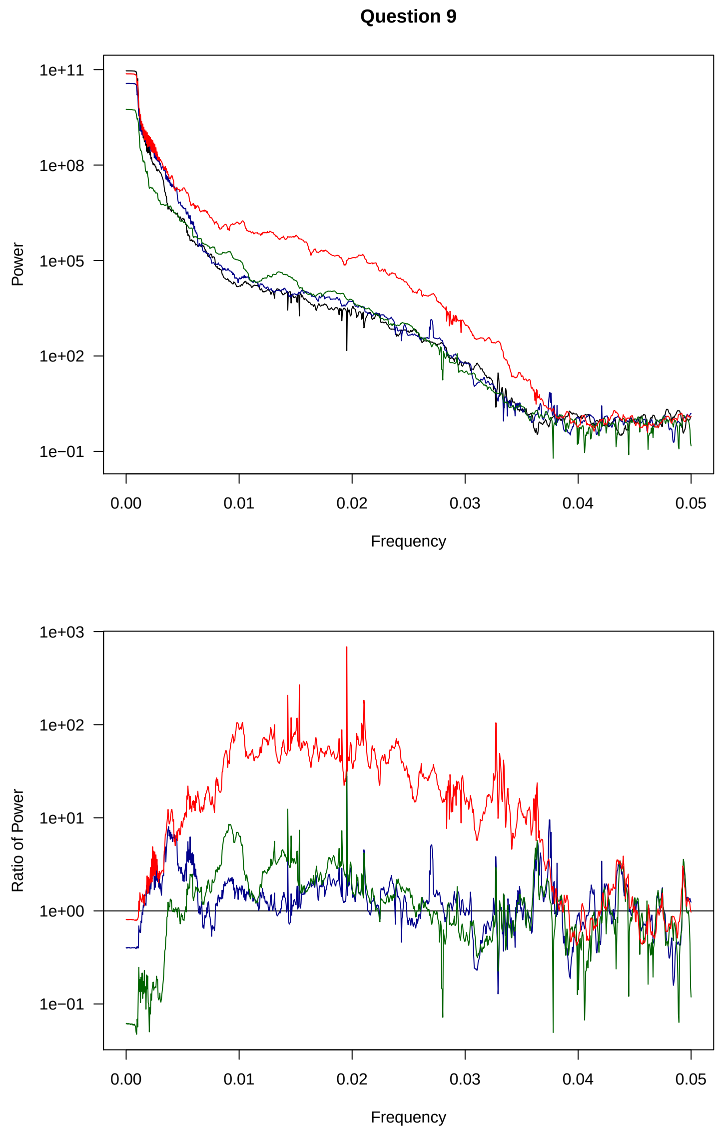

Figure 15: Adaptive multitaper spectrum estimates for four non-overlapping sections of the data (1000 samples each). The top panel are the spectrum estimates, the bottom panel are ratios relative to the spectrum for the first 1000 samples.

Question 10: Estimates of coherence

We used the same AR(2) pre-whitening filter on time series from both of the instruments, and then estimated the coherence of the pre-whitened series. The coherence estimates were done both with the adaptive weights (figure 16) and without (figure 17). The coherency was estimated as

$$\hat{C}_{12}(f) = \frac{\hat{S}_{12}(f)}{\sqrt{\hat{S}_{11}(f)\hat{S}_{22}(f)}} \frac{\frac{1}{\sum_{k=0}^{K-1} d_{1,k}(f)d_{2,k}(f)} \sum_{k=0}^{K-1} d_{1,k}(f)d_{2,k}(f)x_{1,k}(f)x^*_{2,k}(f)}{\sqrt{\hat{S}_{11}(f)\hat{S}_{22}(f)}}$$

where $d_{1,k}(f)$ and $d_{2,k}(f)$ are the adaptive weights for taper k and instrument 1 and 2, respectively; $x_{1,k}(f)$ and $x_{2,k}(f)$ are the eigencoefficients for taper k and instrument 1 and 2, respectively; and $\hat{S}_{11}(f)$ and $\hat{S}_{22}(f)$ are the adaptive multitaper spectrum estimates for taper k and instrument 1 and 2, respectively. The phase was estimated as

$$\phi(f) = \frac{360^\circ}{2\pi} \arctan\left(\frac{\mathrm{Im}\left(\hat{S}_{12}(f)/|\hat{S}_{12}(f)|^2\right)}{\mathrm{Re}\left(\hat{S}_{12}(f)/|\hat{S}_{12}(f)|^2\right)}\right)$$

where $\hat{S}_{12}(f)$ is the cross-spectrum defined above.

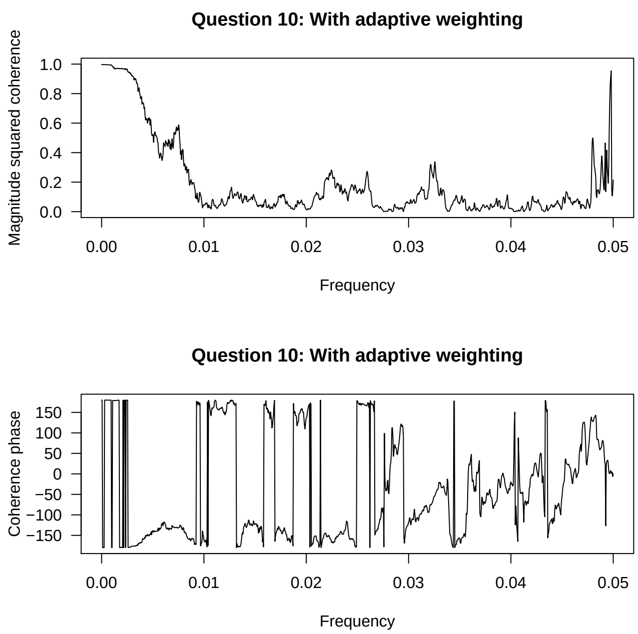

Figure 16: Adaptive multitaper coherence estimates. The top panel are estimates of the magnitude squared coherence. The bottom panel are estimates of the phase of the coherence.

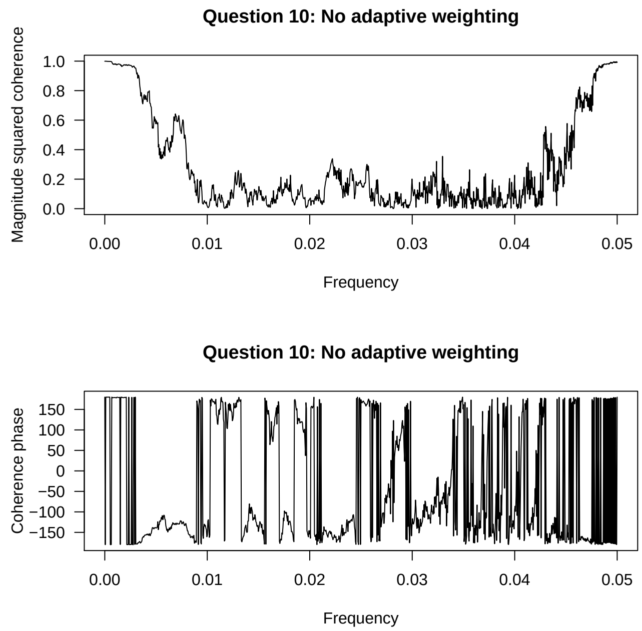

Figure 17: Multitaper coherence estimates (no adaptive weighting). The top panel are estimates of the magnitude squared coherence. The bottom panel are estimates of the phase of the coherence.