

# Data Analysis

Thea Quistgaard

Master Thesis 2020/2021

## 1 Spectral Analysis

### 1.1 Power Spectral Densities

The focus of this project is to investigate if the Signal-to-Noise Ratio(SNR) of isotopic ratio depth series can be increased by reducing the noise via different protocols. One way to be investigate the signal-to-noise ratio(SNR) is by converting the time series(here: depth series) to power spectral densities. A spectral density of a time series can be estimated through many different nonparametric(e.g. periodograms) and parametric methods(e.g. Fast Fourier Transform(FFT) and Maximum Entropy Method (MEM)). During this project the main focus have been on using and MEM[2][?] to generate estimates of Power Spectral Densities. The Power Spectral Density(from now on PSD) is the Fourier Transform of the autocorrelation of a stationary, real signal,  $x(t)$ . The autocorrelation of a such signal is defined as

$$R_x(\tau) = E[x(t)x(t + \tau)], \quad (1)$$

which tells you how correlated the points in the data are with each other, depending on how far separated in time they are. As known, the Fourier transform of a function is

$$H(f) = \int_{-\infty}^{\infty} h(t)e^{2\pi ift} dt, \quad (2)$$

or

$$H(\omega) = \int_{-\infty}^{\infty} h(t)e^{i\omega t} dt, \quad (3)$$

in terms of angular frequency, where  $\omega = \frac{f}{2\pi}$ . Thus the PSD of a real, stationary signal is given by:

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau)e^{-2\pi if\tau} d\tau = \lim_{T \rightarrow \infty} \frac{1}{T} E[|X_T(f)|^2]. \quad (4)$$

PSDs have units of  $\frac{\text{signal units}^2}{\text{Hz}}$  - or in case of a depth series  $\frac{\text{signal units}^2}{1/m} = \text{signal units}^2 \cdot m$ , and has some important properties: due to  $S_x$  being an average of the magnitude squared of the Fourier transform of the time series, then  $S_x \in \mathbb{R}$  and  $S_x \geq 0$  for all frequencies  $f$ .

### 1.2 Maximum Entropy Method

The following section has been written on the basis of the work of Tad J. Ulrych in [2].

An estimate of PSDs of stationary uniformly sampled processes can be given by the spectrum computed by the Maximum Entropy Method(from here on: MEM). The spectrum is determined as the one which maximizes the given process' entropy, which also is an intuitively satisfying answer, as it is logical to search to maximize the entropy of a process. To reach the end goal of a function which can compute the PSD of a process, one must first visit information and entropy theory. If we consider a set of  $M$  things that can happen with each a probability  $p_i$ , then the probability of the occurrence of an event is related to some of the information in the system:

$$I = k \cdot \ln\left(\frac{1}{p_i}\right), \quad (5)$$

where  $k$  is 1 if  $\log_2$  is used. If we then observe the system over a long period of time,  $T$ , we expect (if  $T$  is large)  $p_1 \cdot T$  of  $m_1$  things,  $p_2 \cdot T$  of  $m_2$  things and so on, which all adds up to the total information of:

$$I_{tot} = k \cdot \sum_{i=1}^M (p_i T \ln(\frac{1}{p_i})) \quad (6)$$

Commonly, the average information per time is referred to as entropy:

$$H = \frac{I_{tot}}{T} = k \cdot \sum_{i=1}^M (p_i \ln(\frac{1}{p_i})) = -k \cdot \sum_{i=1}^M (p_i \ln(p_i)) \quad (7)$$

Entropy can be viewed as a measure of uncertainty or variance described by a set of probabilities. That is, for example, if all  $p_i$  except one equals zero, then the system is deterministic and no uncertainty exists. In all other cases the entropy will be positive,  $H > 0$ . Now, the question left to answer is how this theory connects to estimation of PSDs. To answer this consider Jaynes' Principle of maximum entropy[?]:

*The prior probability assignment that describes the available information but is maximally noncommittal with regard to the unavailable information is the one with maximum entropy.*

This principle states that by using a maximum entropy method to determine the prior probability, it is possible to consider the information of the given data and at the same time give the largest uncertainty to the unavailable information. This eliminates some of the constraints which have conventionally been demanded by spectral analysis, where unavailable information prior have been considered to must be either zero or cyclical, which can give rise to errors in the spectral estimates. Now, the method of determining this maximum entropy prior probability distribution is as follows:

- The considered process  $x_t$  can take the values  $x_1, x_2, \dots, x_n$ . The information available about this process is assumed in the form of average values of several functions,  $\langle f_1(x_t) \rangle, \langle f_2(x_t) \rangle, \dots, \langle f_m(x_t) \rangle$ , where  $m < n$ .

The probability distribution  $p_i = p(x_i)$  which is consistent with this infor is the one which maximizes the entropy,  $H = -\sum_{i=1}^n (p_i \log_2(p_i))$  with the following constraints:

$$\sum_{i=1}^n p_i = 1 \text{ and } \langle f_k(x_t) \rangle = \sum_{i=1}^n (p_i f_k(x_i)), k = 1, 2, \dots, m \quad (8)$$

The solution to this problem is well known:

$$p_i = \frac{1}{Z(\lambda_1, \dots, \lambda_m)} \exp(\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)) \quad (9)$$

where the partition function  $Z$  is:

$$Z(\lambda_1, \dots, \lambda_m) = \sum_{i=1}^n \exp(\lambda_1 f_1(x_i) + \dots + \lambda_m f_m(x_i)) \quad (10)$$

The Lagrange multipliers  $\lambda_k$  are determined from the constraints and reduces to:

$$\langle f_k(x_i) \rangle = \frac{\partial}{\partial \lambda_k} \log(Z(\lambda_1, \dots, \lambda_m)) \quad (11)$$

Now, what is really important to notice is that "The probability distribution which maximizes the entropy,  $H$ , is numerically identical to the frequency distribution."[?]. So computing the entropy maximizing probability distribution also leaves us with a way to the PSD of interest.

First we need to take a look at the relation between entropy and the spectral density  $S(f)$  of a stationary gaussian process, whic we can assume the one in this report is approximate to:

$$H = \frac{1}{4f_N} \int_{-f_N}^{f_N} \log(S(f)) \quad (12)$$

and in terms of the autocorrelation,  $\phi(k)$

$$H = \frac{1}{4f_N} \int_{-f_N}^{f_N} \log\left(\sum_{-\infty}^{\infty} \phi(k) \exp(-2i\pi f k \Delta t)\right), \quad (13)$$

where  $\Delta t$  is the uniform sampling rate and  $f_N = \frac{1}{\Delta t}$  is the Nyquist critical frequency.

Maximizing the entropy with respect to the unknown autocorrelation  $\phi(k)$  with the constraint that  $S(f)$  is consistent with the known autocorrelations,  $\phi(0), \dots, \phi(M-1)$  results in exactly the Maximum Entropy Method spectral estimate. As mentioned, this is interpreted as a solution that expresses the maximum uncertainty with respect to the unknown information while simultaneously being consistent with the known information.

All this finally leads us to the generally known expression for a MEM spectral density estimate for a linear process:

$$P_E(f) = \frac{P_M}{f_N |1 + \sum_{j=1}^{M-1} (\lambda_j \exp(-2\pi f j \Delta t))|^2}, \quad (14)$$

where  $P_M$  is a constant and  $\lambda_j$  is the prediction error coefficients which is determined from the data. For a description on how to determine the coefficients see [?].

### 1.3 Noise of PSD and variance of depth series

Now, the actual purpose of computing a PSD estimate from the given depth data series is to determine the noise in the data. The goal is to be able to separate the true signal,  $s(t)$  or  $S(f)$ , from the noise,  $n(t)$  or  $N(f)$ , by analyzing a single measured signal,  $c(t)$  or  $C(f)$ [?]. What is actually measured, when converting to spectral densities, is  $|C|^2$ , and what is desired to find is  $|N|^2$  and  $|S|^2$ . For an example of the approximated noise level see figure ??.

From the noise it is possible to determine the standard deviation of the time series as[?][?][1]:

$$\sigma^2 = \int_{-f_N}^{f_N} |N(f)|^2 df. \quad (15)$$

$|N(f)|^2$  is obtained by calculating an average of the last  $n$  points in the flattening high frequency range of the spectrum. The number of points used for this average is determined from an analysis of the graphical data and differs for  $\delta D(N_{points} = 144)$ ,  $\delta^{18}O(N_{points} = 594)$  and  $\delta^{17}O(N_{points} = 594)$ .

## 2 Signal restoration/Resolution enhancement

Due to diffusion in firn and ice, some of the water isotopic signal is lost. Some of this signal can be restored by investigating the diffusion process, and through filtering and deconvolution techniques(REFERENCES). For the data of this thesis two different restoration techniques are presented: a spectral method, determining the effect of mixing and diffusion as a spectral filter(REFERENCES), and a kernel restoration method much like the ones used to restore pixel resolution in images (REFERENCES).

## 2.1 Wiener Filtering

Through spectral analysis it is possible to treat the noise of the signal consistently. The goal is to create spectral filters which enhances the signal while minimizing the effect of the noise, thus increasing the signal-to-noise ratio (SNR).

Theoretically, without any diffusion, the change in isotopic concentration would be described through a step function, going from one constant concentration to another. This step function can be described by the Heaviside function:

$$D(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases} \quad (16)$$

In reality, a number of different mixing processes change this step function, and the measured signal will be a smooth curve,  $s(t)$ , which corresponds to the convolution of  $S(t)$  with the mixing response function  $M(\tau)$

$$d(t) = \int_{-\infty}^{\infty} D(\tau) \cdot M(t - \tau) d\tau \quad (17)$$

## 2.2 Kernel Estimation

As is well known, in the spectral domain, convolution is multiplication and the mixing is described as the multiplication between the Fourier transform of  $S$  and  $M$ :

$$\tilde{d} = \tilde{D} \cdot \tilde{M} \quad (18)$$

By differentiation with respect to time, the mixing filter  $M$  is unaffected, and differentiation of the measured system response, the Heaviside function,  $S'$  is a delta function, which Fourier transformed is unity, leading to:

$$\tilde{d}' = \tilde{D}' \cdot \tilde{M} = \tilde{M} \quad (19)$$

The mixing filter can thus be determined by measuring the system response to a step function, differentiating performing Fourier transform of the result  $d'$ .

After determination of the mixing filter  $\tilde{M}$ , the unmixed signal  $D$  can be estimated in theory by inverse Fourier transform of

$$\tilde{D} = \tilde{d} \cdot \tilde{M}^{-1} \quad (20)$$

During the mixing, cycles with short wavelengths are heavily washed out, and through the restoration in Eq. 20, the amplitudes corresponding to these wavelengths are heavily amplified by the filter. This method though has a drawback, which is that when the measurements contain noise, the restored signal will be dominated by high-frequency noise, greatly amplified by the mixing filter. Thus it is a problem of retaining as much (short wavelength) signal as possible while simultaneously attempting to amplify the high-frequency noise as little as possible. This optimal trade-off can be found by creating an optimum filter for the considered measured isotopic signal:

$$\delta_M(z) = \delta_m(z) + \eta(z) \quad (21)$$

This optimal (Wiener) filter  $\tilde{F}$ , defined for each wave number  $k = 2\pi\omega$ , is presented as the ratio between pure signal and pure signal plus noise described in Power Spectral Densities as:

$$\tilde{F}(k) = \frac{|\tilde{\delta}_m(\omega)|^2}{|\tilde{\delta}_m(\omega)|^2 + |\tilde{\eta}(\omega)|^2} \quad (22)$$

In this work, the power spectral densities of the signal and the noise, respectively, are determined through analysis of the power spectral density of the combined signal/noise PSD.

The PSD of the noise free measured signal,  $|\tilde{\delta}_m(\omega)|^2$ , is assumed describe as

$$|\tilde{\delta}_m(\omega)|^2 = P_0 e^{-k^2 \sigma_{\text{tot}}^2} \quad (23)$$

where  $\sigma_{\text{tot}}^2$  describes the total estimated diffusion length of the mixing.

The noise is assumed to be red noise, described by an autoregressive process of first order, AR1:

$$|\tilde{\eta}(\omega)|^2 = \frac{\sigma_{\eta}^2 \Delta z}{|1 + a_1 \exp(-2\pi i \omega \Delta z)|^2} \quad (24)$$

where  $\sigma_{\eta}^2$  is the variance of the red noise,  $a_1$  is the AR1 coefficient and  $\Delta z$  is the resolution of the time/depth data. It is then possible to estimate the parameters  $P_0$ ,  $\sigma_{\text{tot}}^2$ ,  $\sigma_{\eta}^2$  and  $a_1$  by curve fitting, separately, the two expressions in Eq. 23 and 24 to the data. The estimated parameters are varied to find the optimal guess to use for the filter.

## References

- [1] Vasileios Gkinis, Trevor J Popp, Sigfus J Johnsen, Thomas Blunier, M Bigler, S Schüpbach, E Kettner, and Sigfus J Johnsen. Water isotopic ratios from a continuously melted ice core sample. *Atmospheric Measurement Techniques*, 4(4):2531–2542, 2011.
- [2] Tad J Ulrych and Thomas N Bishop. Maximum entropy spectral analysis and autoregressive decomposition. *Reviews of Geophysics*, 13(1):183–200, 1975.