Tutorial: When Randomness Helps in Undersampling

Roel Snieder[(1)] and Michael B. Wakin[(2)]

[(1)] *Center for Wave Phenomena, Colorado School of Mines, rsnieder@mines.edu*
[(2)] *Dept. of Electrical Engineering, Colorado School of Mines, mwakin@mines.edu*

ABSTRACT
Signals cannot always be sampled at their full desired resolution. In this tutorial, we explore the benefits of randomly subsampling a signal's frequency spectrum. Whereas uniform subsampling would introduce structural artifacts in the time series, random subsampling introduces a type of noise whose behavior we quantify. This analysis gives insight into the reasons why random sampling is employed in more sophisticated processing techniques such as compressive sensing. The signal processing codes and data used in this work can be downloaded from `https://mines.edu/~mwakin/software`.

Keywords: digital signal processing, statistics, compressive sensing

1 INTRODUCTION

Signals must be digitized if they are to be stored or processed with computers. Commonly, one does this by collecting and storing samples of a function $f(t)$ at discrete times t_j. The resulting time series might be a recording of music, a telecommunication signal, or the output from any type of continuous measurement. An example is shown by the black line in Figure 1 which represents the air pressure recorded on a volcano. This time series consists of a set of discrete measurements that are sampled with a time interval of 0.025s. As we discuss in Section 2, using the discrete Fourier transform (DFT), one can alternatively represent such a time series in the frequency domain, i.e., as a sum of sinusoids having various frequencies, amplitudes, and phases.

The default assumption in signal processing is that one has access to a full collection of N uniform samples in time, or equivalently, a full collection of N frequency components (DFT coefficients). In practical applications, however, it may be useful to violate this assumption in order to reduce the burden of data acquisition, data transmission, or storage. A measurement system may only have access to certain time samples (or frequency components) of a signal, or one may choose to discard certain time (or frequency) samples in order to store less data. When only $M < N$ samples are available in the time or frequency domain, we say that a signal is *undersampled*. One might think that undersampling is best done by leaving out samples at regular intervals, but in this tutorial we give insight into why random undersampling, where one undersamples with random intervals, in general is superior to regular undersampling.

Random or irregular sampling has been exploited in a broad variety of applications. In seismic surveys there are advantages to recording seismic wavefields at random space intervals rather than at regular intervals (Herrmann, 2010). In order to make seismic surveys at sea more efficient, one often uses multiple seismic sources. Firing the different seismic sources at irregular intervals, a process descriptively called *popcorn shooting*, produces clearer seismic images than when the sources fire at regular intervals (Abma & Foster, 2020; Abma & Ross, 2013). In Magnetic Resonance Imaging (MRI), one can accelerate the data acquisition process by collecting incomplete, irregular samples in the frequency domain (Lustig et al., 2008). Other examples of random undersampling in the time or frequency domain have arisen in the context of numerical integration (Caflisch, 1998), compressive sensing (Candès & Wakin, 2008; Foucart & Rauhut, 2013), matched filtering (Eftekhari et al., 2013), and power spectrum estimation (Ariananda & Leus, 2012).

The above examples vary widely in the mechanisms for data acquisition, the steps for processing the available

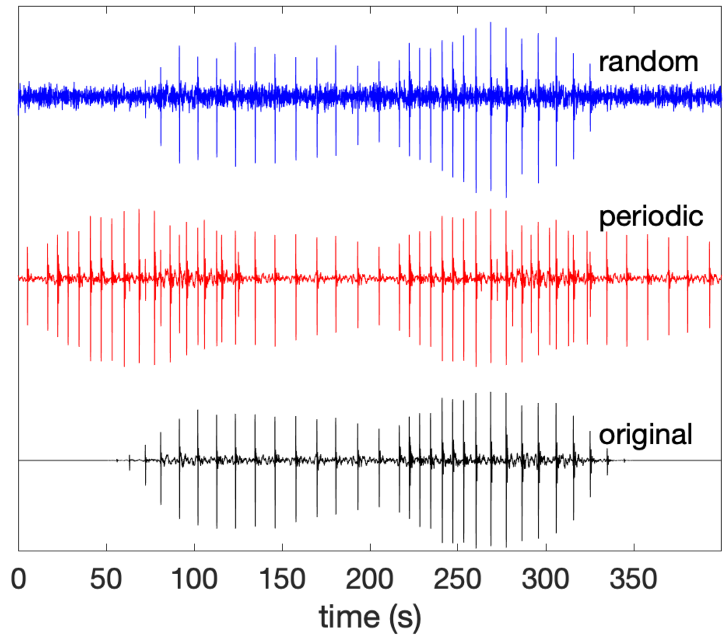

Figure 1. Pressure recorded at Arenal after tapering and adding zeros at both ends (black). The pressure after periodic subsampling in the frequency domain with a factor of 2 (red). The pressure after random subsampling in the frequency domain with a factor of 2 (blue).

samples, and the specific theoretical justifications for random undersampling. In this article, we focus on the problem of reconstructing a signal in the time domain using undersampled frequency components. (We merely omit the missing components from the inverse DFT equation.) While simple, this framework gives a good platform for observing and understanding the benefits of random undersampling compared to regular undersampling. As we discuss in Section 6, the intuitions carry over to more sophisticated processing methods such as those used in compressive sensing.

We illustrate the benefits of random undersampling with an example of the air pressure recorded at Arenal, a volcano in Costa Rica. This volcano has frequent explosions that occur at regular intervals in time. A short sequence of the air pressure fluctuations caused by these explosions is shown by the black time series in Figure 1. This is part of a longer pressure recording that was used by Snieder & Hagerty (2004) to retrieve the elastic response of the volcano to a single explosion. For illustrative purposes we add 50s with zero air pressure at the beginning and end of the time series.

We decompose the pressure signal into its frequency components via the DFT (see Section 2) and use two methods to subsample the data in the frequency domain. The red time series in Figure 1 is obtained by throwing away every other frequency component. This means that the data are subsampled periodically in the frequency domain by a factor of 2, and only $M = N/2$ DFT coefficients are used to reconstruct the time series. Note that the red time series does not resemble the original time series at all. Additional pressure spikes are present in the red time series that do not appear in the original time series. For example, pressure spikes appear in the first and last 50s of the red time series, where the original time series was equal to zero in these intervals. In contrast, the blue time series in Figure 1 is obtained by randomly selecting $M = N/2$ of the frequency components, while discarding the other $N/2$ frequency components, and by taking the real part of the resulting time series. The pressure spikes in the blue time series thus obtained agree well with the pressure spikes in the original time series. The only difference from the original time series is that random noise is present. In summary, we see in this example that periodic subsampling in the frequency domain leads to a structural distortion, while random subsampling in the frequency domain leads to the addition of a type of incoherent random noise. We explain both of these phenomena in this article.

Our treatment thus involves topics such as frequency analysis, aliasing, and convolution which are commonly encountered in undergraduate courses on signal processing or engineering mathematics. Meanwhile, we also draw from concepts in probability and statistics which are rarely discussed at the undergraduate level in the context of frequency analysis, aliasing, and convolution. Section 2 of this article begins by discussing the DFT, its properties, and its connection to convolution. In Section 3, we review the time-domain aliasing effects associated with periodic subsampling of the DFT. In Section 4, we analyze the mean and variance of the signal estimate based on random subsampling of the DFT. We assume the frequency-domain samples are drawn randomly without replacement. Section 5 then extends our treatment to the case of random sampling with replacement. We conclude in Section 6 with a broader view of related problems where the benefits of random subsampling may emerge.

2 THE DISCRETE FOURIER TRANSFORM AND ITS PROPERTIES

The discrete Fourier transform (DFT) allows a length-N discrete-time signal to be represented as a sum of N complex exponentials (Oppenheim & Schafer, 2010). Letting $f_j, j = 0, 1, \ldots, N-1$ represent the time domain samples and $F_n, n = 0, 1, \ldots, N-1$ represent the signal's DFT coefficients, we have

$$f_j = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{i2\pi jn/N} \ , \tag{1}$$

with

$$F_n = \sum_{j=0}^{N-1} f_j e^{-i2\pi jn/N} \ , \tag{2}$$

where $i = \sqrt{-1}$ denotes the imaginary unit. We assume throughout this article that N is even. The signal represented by the Fourier expansion is periodic; this can be seen by evaluating expression (1) for sample $j+N$. In that case

$$f_{j+N} = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{i2\pi jn/N} e^{i2\pi Nn/N} = \frac{1}{N} \sum_{n=0}^{N-1} F_n e^{i2\pi jn/N} = f_j \ , \tag{3}$$

where we used that $e^{i2\pi Nn/N} = e^{i2\pi n} = 1$ for any integer n, and where we used equation (1) again in the last identity. This means that f_j repeats itself after N samples: $f_{j+N} = f_j$.

In our discussion, we will make use of an elementary signal known as the *delta function*

$$e_j := \begin{cases} 1, & j = 0, \\ 0, & j = 1, 2, \ldots, N-1. \end{cases} \tag{4}$$

Relatedly, we introduce notation for the *Kronecker delta* as

$$\delta_{j,k} := \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases} \tag{5}$$

and using this notation, we note that $e_j = \delta_{j,0}$.

Exercise 1. Show that all of the DFT coefficients of the delta function are equal: $E_n = 1$ for $n = 0, 1, \ldots, N-1$.

Finally, we introduce the notion of circular convolution, which arises in the context of filtering discrete-time signals. The circular convolution of two discrete-time signals f_j and h_j is given by

$$g_j = \sum_{\ell=0}^{N-1} h_\ell f_{(j-\ell) \bmod N} = h_0 f_j + h_1 f_{(j-1) \bmod N} + \cdots + h_{N-1} f_{(j-N+1) \bmod N} \ . \tag{6}$$

We denote this process as $g_j = (h * f)_j$ for short. The notation $f_{(j) \bmod N}$ indicates that for a given value of j multiples of N are added or subtracted from j so that the result lies in the interval $0 \leq (j) \bmod N \leq N-1$. In words, circular convolution creates multiple circularly shifted copies of the signal f_j, weights each copy by one entry of h_j, and adds everything. Circular convolution has an important connection with the DFT: convolution in time corresponds to multiplication in frequency. In particular, for the signal g_j defined in (6),

$$G_n = F_n H_n \text{ for } n = 0, 1, \ldots, N-1. \tag{7}$$

Exercise 2. Show that circular convolution with the delta function e_j leaves a signal unchanged, i.e., that $f_j = (f * e)_j$. This can be proved either in the time domain using (6) or in the frequency domain using (7).

3 PERIODIC SUBSAMPLING IN THE FREQUENCY DOMAIN

Now that we have the basic tools of the DFT, we explain why the red time series in Figure 1 differs so much from the original time series shown in black. The red time series is obtained by periodic subsampling of the frequency components. In the example of Figure 1 the periodic subsampling left out every other frequency component. We denote the resulting time series by $\tilde{f}_j$. Leaving out every other frequency component in expression (1) gives

$$\tilde{f}_j = \frac{2}{N} \sum_{n=0}^{N/2-1} F_{2n} e^{i2\pi j(2n)/N} \ . \tag{8}$$

Note that the summation index n is replaced by $2n$ to ensure that only the frequency components $F_0, F_2, \ldots, F_{N-2}$ contribute. The factor of 2 out front is added to compensate for the fact that only half the frequency components are accounted for in this sum.

The frequency components are still given by expression (2), hence

$$\tilde{f}_j = \frac{2}{N} \sum_{n=0}^{N/2-1} \sum_{k=0}^{N-1} f_k e^{-i2\pi k(2n)/N} e^{i2\pi j(2n)/N} = \frac{2}{N} \sum_{k=0}^{N-1} f_k \sum_{n=0}^{N/2-1} e^{i2\pi (j-k)n/(N/2)} \ , \tag{9}$$

where the second equality is obtained by interchanging the double sums and writing $2n/N$ in the exponents as $n/(N/2)$. Expression (9) relates the original time series f_k to the subsampled time series $\tilde{f}_j$. Noting that $e^{i2\pi(j-k)n/(N/2)} = e^{i2\pi((j-k) \bmod N)n/(N/2)}$, with a change of variables $l = j - k$ we can rewrite (9) as

$$\tilde{f}_j = \sum_{\ell=0}^{N-1} f_{(j-\ell) \bmod N} \left(\frac{2}{N} \sum_{n=0}^{N/2-1} e^{i2\pi \ell n/(N/2)} \right) . \tag{10}$$

Comparing the expression in parentheses in (10) to the inverse DFT equation (1), we recognize this expression as the length-$N/2$ (rather than length-N) inverse DFT of a signal whose DFT coefficients are all 1. From Exercise 1, we know that such a signal corresponds to a delta function in the time domain, and from the periodicity discussed in (3), we know that such a signal repeats periodically with period $N/2$. Therefore, we have

$$\tilde{f}_j = \sum_{\ell=0}^{N-1} f_{(j-\ell) \bmod N} (\delta_{\ell,0} + \delta_{\ell,N/2}). \tag{11}$$

In the sum over ℓ, the first delta function only gives a nonzero contribution when $\ell = 0$, and the second delta function only gives a nonzero contribution when $\ell = N/2$, hence

$$\tilde{f}_j = f_j + f_{(j-N/2) \bmod N} . \tag{12}$$

This equation states that the time series obtained by only taking every other frequency component into account is the superposition of original time series and the time series shifted to the right by half the total time interval. (Because all time shifts are modulo N, the parts of the function that are shifted out of the interval come back on the other side.)

This behavior can be seen by comparing the red time series in Figure 1 with the original (black) time series. The red time series is the sum of the black time series plus a version of this time series shifted to the right over half the time interval. For example, the high amplitude peaks between 220s and 320s are shifted to 420s and 520s. These times fall outside the time interval, which runs from 0s to 400s, and so the shifted high amplitude peaks show up in the red time series between 20s and 120s.

The derivation in this section is standard, and we make two remarks here. (i) The phenomenon observed in (12) is commonly known as *aliasing* of the signal f_j in the time domain. Indeed, this form of aliasing arises any time a signal is periodically subsampled in the frequency domain. When the roles of time and frequency are reversed, aliasing occurs in the frequency domain when a signal is periodically subsampled in time (Oppenheim & Schafer, 2010). (ii) Comparing equation (11) with (6), we recognize $\tilde{f}_j$ as the circular convolution of f_j with a sum of two delta functions $h_j := \delta_{j,0} + \delta_{j,N/2}$. Indeed, we can interpret the process of zeroing out the spectrum of f_j as multiplication in the frequency domain by an alternating sequence of 0's and 1's (actually 0's and 2's due to our rescaling). Taking the inverse length-N DFT of this sequence of 0's and 2's, one obtains h_j; recall that multiplication in frequency corresponds to circular convolution in time, and so (11) also follows from (7).

Exercise 3. Our analysis in this section can readily be generalized to the case when one subsamples not by a factor of 2 but by an arbitrary positive integer η that divides N. In this case, show that the retrieved function $\tilde{f}_j$ is the superposition of the original time series f_j and copies of the time series that are shifted over by multiples of N/η.

We conclude from the theory in this section and the data example in Figure 1 that periodic subsampling in the frequency domain leads to the superposition of the original time series and shifted versions of this time series. This leads, in general, to severe distortions of the signal, which is undesirable if one wants the subsampled time series to resemble the original time series.

4 STATISTICAL ANALYSIS OF RANDOM SUBSAMPLING IN THE FREQUENCY DOMAIN

The blue time series in Figure 1 suggests that random subsampling introduces a type of background noise on the subsampled signal but leaves it otherwise intact. In this section we explain why this is the case. We consider the case where the frequency components are randomly subsampled but no frequency component is used twice. This is called *subsampling without replacement.* This type of subsampling is used in the example of figure 1. In Section 5 we relax this constraint and consider the case where the frequencies are sampled independently.

4.1 The mean of the randomly subsampled time series

As shown in equation (1), the DFT allows a length-N discrete-time signal to be represented as a sum of N complex exponentials with DFT coefficients indexed by $n = 0, 1, \ldots, N-1$. Let $n_1, n_2, \ldots, n_M$ denote a random subset of indices drawn without replacement from the set $\{0, 1, \ldots, N-1\}$. We consider using only these indices in the DFT synthesis equation, yielding the subsampled signal

$$\tilde{f}_j = \frac{N}{M} \cdot \frac{1}{N} \sum_{m=1}^{M} F_{n_m} e^{i2\pi j n_m/N} = \frac{1}{M} \sum_{m=1}^{M} F_{n_m} e^{i2\pi j n_m/N} \ . \tag{13}$$

The factor of N/M is introduced to compensate for the fact that only M out of N DFT coefficients are used in the synthesis sum. Because (13) uses only M frequencies to synthesize $\tilde{f}_j$, and because these frequencies are chosen randomly, the effect of the subsampling should be analyzed statistically.

To gain some intuition for the effect of this subsampling, recall the delta function e_j defined in (4). From Exercise 1, we know that this signal has DFT coefficients all equal to 1. The randomly subsampled version of the delta function is thus given by

$$\tilde{e}_j = \frac{1}{M} \sum_{m=1}^{M} E_{n_m} e^{i2\pi j n_m/N} = \frac{1}{M} \sum_{m=1}^{M} e^{i2\pi j n_m/N} \ , \tag{14}$$

which is a sum of complex exponential signals with randomly chosen frequencies. For each time index j, $\tilde{e}_j$ is thus a random variable. We can compute its expectation:

$$\mathbb{E}(\tilde{e}_j) = \mathbb{E}\left(\frac{1}{M} \sum_{m=1}^{M} e^{i2\pi j n_m/N} \right) = \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}(e^{i2\pi j n_m/N}) \ , \tag{15}$$

where the only random quantity inside the expectation is the random index n_m. Although the random indices $n_1, n_2, \ldots, n_M$ are not independent (since repetitions are not allowed), we may assume they all have the same marginal distribution, namely the uniform distribution over $\{0, 1, \ldots, N-1\}$. Therefore $e^{i2\pi j n_m/N}$ takes the possible values $e^{i2\pi j 0/N}, e^{i2\pi j 1/N}, \ldots, e^{i2\pi j(N-1)/N}$, each with probability $1/N$. This is all that is required to compute the expectation:

$$\mathbb{E}\left(e^{i2\pi j n_m/N} \right) = \frac{1}{N} \sum_{n=0}^{N-1} e^{i2\pi j n/N}, \tag{16}$$

which merely equals e_j thanks to (1) and Exercise 1. Inserting in to (15), we obtain

$$\mathbb{E}\left(\tilde{e}_j\right) = \frac{1}{M}\sum_{m=1}^{M}\mathbb{E}(e^{i2\pi j n_m/N}) = \frac{1}{M}\sum_{m=1}^{M} e_j = \frac{1}{M}\cdot M \cdot e_j = e_j \ . \tag{17}$$

The expectation value of the randomly subsampled time series $\tilde{e}_j$ is thus equal to the original time series e_j. This means that the random subsampling is *unbiased*—on average, random subsampling does not alter the time series. Let us reflect on why there is no bias. The crux is expression (16) where the expectation value $\mathbb{E}\left(e^{i2\pi j n_m/N}\right)$ is computed. Because the probability distribution of the random index n_m is uniform, the expectation corresponds to an unweighted sum over all frequency components, which yields the original function e_j.

To understand the effect of random subsampling on f_j, consider the following exercise which establishes a very handy fact.

Exercise 4. Show that $\tilde{f}_j = (f * \tilde{e})_j$; that is, random subsampling of a time series f_j is equivalent to convolution with a randomly subsampled delta function. Hint: Recall Exercise 2 and use the fact that circular convolution in time corresponds to multiplication in the frequency domain.

With the result of Exercise 4 in hand, it follows that

$$\mathbb{E}(\tilde{f}_j) = \mathbb{E}((f * \tilde{e})_j) = (f * \mathbb{E}(\tilde{e}))_j = (f * e)_j = f_j, \tag{18}$$

where we have used (17) and again used the identity from Exercise 2. The fact that subsampling e_j is unbiased thus means that subsampling any time series f_j is unbiased.

A comparison of the blue time series in Figure 1, obtained from random subsampling in the frequency domain by a factor of 2, with the original time series shows that the pressure spikes in the original time series are indeed maintained by the random subsampling. We show in expression (22) below that random subsampling does not change the value of a delta function, hence spikes are preserved by random subsampling. However, as the blue time series shows, the random subsampling also introduces background noise. We show in expression (23) below that this noise is uncorrelated and derive the noise level in equation (24).

4.2 The covariance of the randomly subsampled delta function

In order to understand the introduction of the background noise due to random subsampling it does not suffice to compute the mean of the subsampled function, because the mean is equal to the original function. Rather, to quantify the noise we must also consider the variability of the subsampled time series. Here, we do this by evaluating the covariance of the randomly subsampled delta function $\tilde{e}_j$. This covariance is defined as

$$C_{j,k} = \mathbb{E}\left(\left(\tilde{e}_j - \mathbb{E}(\tilde{e}_j)\right)\left(\tilde{e}_k - \mathbb{E}(\tilde{e}_k)\right)^*\right) \ . \tag{19}$$

In this expression the asterisk denotes complex conjugation. The complex conjugation is introduced because $\tilde{e}_j$ is, in general, complex. For $j = k$ the covariance denotes the variance (squared standard deviation) of $\tilde{e}_j$, and the complex conjugation ensures that the variance is real and nonnegative. The derivation of this covariance for the random subsampling is somewhat lengthy, and we show in Section B of the appendix that

$$C_{j,k} = \frac{(N-M)}{(N-1)}\frac{1}{M}\left(\delta_{j,k} - \delta_{j,0}\delta_{k,0}\right) \ , \tag{20}$$

where N denotes the number of samples in the original time series and M the number of frequency components that are taken into account in the subsampling.

When confronted with a seemingly complicated expression such as (20), it is useful to first consider limiting cases where one knows the answer. For example, when $M = N$, the number of subsampled frequencies equals the number of original frequencies. Since we assumed that all sampled frequency components are distinct, this means that the subsampled frequencies cover all of the original frequencies; that is, there is no subsampling at all, and the subsampled function must actually be equal to the original function. Consequently, there should be no variability in the subsampled signal, and so the covariance should vanish. And indeed, when $M = N$, the covariance in (20) equals zero.

Next we consider the case $j = k = 0$, so that $C_{j,k} = C_{0,0}$ represents the variance of $\tilde{e}_0$. In this case $\delta_{j,k} = \delta_{0,0} = 1$

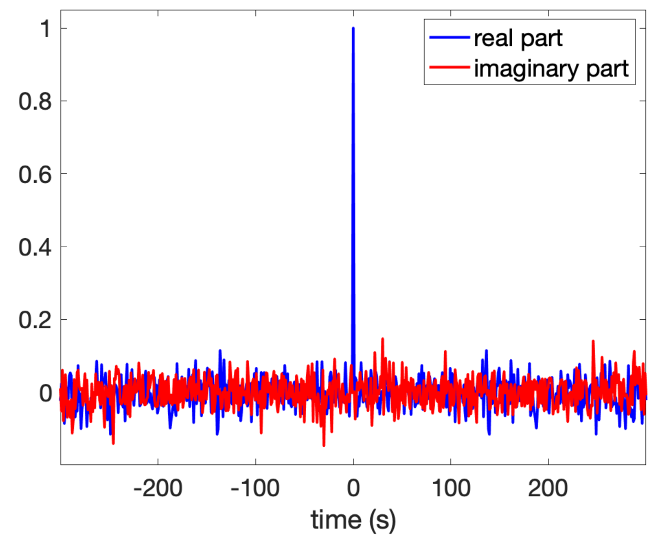

Figure 2. The real part (blue) and imaginary part (red) of a delta function that is randomly subsampled in frequency without replacement by a factor 3. There are $N = 600$ points in the time series.

and $\delta_{j,0} = \delta_{k,0} = \delta_{0,0} = 1$, so equation (20) gives

$$C_{0,0} = 0 \; . \tag{21}$$

Does the variance of $\tilde{e}_0$ truly equal zero? According to equation (14),

$$\tilde{e}_0 = \frac{1}{M}\sum_{m=1}^{M} e^{i2\pi 0 n_m/N} = \frac{1}{M}\sum_{m=1}^{M} e^0 = \frac{1}{M}\sum_{m=1}^{M} 1 = \frac{1}{M}\cdot M = 1. \tag{22}$$

Hence the value of $\tilde{e}_0$ does not depend on the random sampling of the frequencies, and thus the variance of $\tilde{e}_0$ vanishes, as predicted by expression (21).

As an illustration, we show in Figure 2 an example of a randomly subsampled delta function $\tilde{e}_j$. In this example the time series consist of $N = 600$ samples, and one third ($M = 200$) of the frequencies are used in the random subsampling. Since the subsampled function is complex, both the real and imaginary parts are shown in Figure 2. Although it is not visible to the eye, the real part of the subsampled function for $j = 0$ is exactly equal to 1 and the imaginary part vanishes. This confirms that the variance at the time of the spike at $j = 0$ indeed vanishes.

Next we consider $C_{j,k}$ when $j \neq k$. When $j \neq k$, $\delta_{j,k} = 0$ and $\delta_{j,0}\delta_{k,0} = 0$ as well because $\delta_{j,0}\delta_{k,0}$ is only nonzero when both j and k equal zero, but this cannot be satisfied when $j \neq k$. Inserting into expression (20), it follows that

$$C_{j,k} = 0 \quad \text{for } j \neq k \; . \tag{23}$$

Indeed, the covariance for different values of the subsampled time series vanishes, which means that these different sampling points are *uncorrelated*. The example of the randomly subsampled delta function in Figure 2 shows that away from the spike at $j = 0$, the variations behave like random noise.

The amplitude of the noise that is generated by the random subsampling follows by considering the covariance (20) for $k = j \neq 0$. The covariance then reduces to the variance σ_j^2, which is given by

$$\sigma_j^2 = C_{j,j} = \frac{(N-M)}{(N-1)}\frac{1}{M} \quad \text{for } j \neq 0 \; . \tag{24}$$

We restrict ourselves to the case $j \neq 0$ because according expression (21), the variance vanishes for $j = 0$. According to expression (24), the variance σ_j^2 does not depend on j; this can also be seen in Figure 2 where the noise level is roughly constant across time. Taking the derivative of expression (24) with respect to M gives

$$\frac{d\sigma_j^2}{dM} = -\frac{N}{M^2(N-1)} < 0 \; . \tag{25}$$

This means that the variance decreases when M, the number of frequency component that are taken into account, increases.

The standard deviation of the noise follows by taking the square root of expression (24) and is given by

$$\sigma_j^{\text{no replacement}} = \sqrt{\frac{N-M}{N-1}}\frac{1}{\sqrt{M}} \ . \tag{26}$$

The first square root involves the ratio of the number of frequency components that are not taken into account $(N-M)$ relative to the total number of frequency components reduced by one $(N-1)$. The second square root involves the number of frequency components that are taken into account. The $1/\sqrt{M}$ dependence in the standard deviation is the same as for the average of M independent random numbers; for example, see equation (21.41) of Snieder & van Wijk (2015).

The standard deviation of the fluctuations in Figure 2 can be estimated by evaluating $\sigma^2_{\text{estimated}} = \sum_{j=1}^{N-1} |\tilde{e}_j|^2/(N-1)$, where we left out the point $j = 0$ because it has zero variance (expression (20)). The estimated standard deviation is given by $\sigma_{\text{estimated}} = 0.057$. In generating this figure, $N = 600$ and $M = 200$, and so equation (26) predicts for these values that $\sigma_j = 0.057$, which matches the empirical value of the standard deviation.

4.3 Random subsampling for a general time series

In the previous subsection we analyzed the imprint of random subsampling without replacement of the time series e_j. In this subsection, we analyze the case where random subsampling without replacement is applied to a general time series f_j instead of the delta function e_j. For more general signals f_j, the behavior of the subsampled signal $\tilde{f}_j$ defined in (13) can be derived from the treatment of the subsampled delta function. Recall that $\tilde{f}_j = (f * \tilde{e})_j$ as derived in Exercise 4. The subsampled delta function $\tilde{e}_j$ is thus convolved with the signal f_j to yield the subsampled signal estimate $\tilde{f}_j$.

As an example, we consider a Ricker wavelet which is defined as a the second derivative of a Gaussian

$$f_j = \left(1 - 2(t_j/\tau)^2\right)\exp\left(-(t_j/\tau)^2\right) \quad \text{with } \tau = 10 \ s \ , \tag{27}$$

where t_j maps the sample index j to a time interval symmetric about 0. This time series is shown by the black line in Figure 3.

We showed in expression (18) that the subsampled time series is unbiased, so we only need to investigate the covariance of the subsampling noise. The real and imaginary parts of the Ricker wavelet that is randomly subsampled by a factor of 3 are shown in Figure 3 by blue and red lines, respectively. Note that, as in Figure 2, the noise introduced by the subsampling is constant in time. The level of the subsampling noise is, however, larger than is it for the subsampled delta function, even though the number of samples and the subsampling used are identical.

We show in Section C of the appendix that the covariance of the subsampled time series $\tilde{f}_j$ is given by

$$C^{(f)}_{j,k} = \left(\frac{N-M}{N-1}\frac{1}{M}\right)\left(\left\{\sum_{p=0}^{N-1} f_{j-p}f^*_{k-p}\right\} - f_j f^*_k\right) \ , \tag{28}$$

where it is understood that the subscripts of f are evaluated $\bmod N$, that is, integer multiples of N are added or subtracted so that the index lies in the interval $0, 1, \ldots, N-1$. According to expression (23) the subsampling noise for a delta function is uncorrelated for different samples. This is, in general, not the case for an arbitrary time series; expression (28) does not predict that the covariance between different samples in the subsampled time series vanishes. The reason is that a general time series f_j does not contain all frequencies, and because of this band-limitation, the artifacts caused by subsampling are band-limited as well. This can be seen in Figure 3 where the subsampling noise has the same frequency content as the Ricker wavelet.

The variance $\sigma^{(f)2}_{jj}$ of the subsampled wavelet follows by setting $k = j$. In that case the sum in equation (28) is given by $\sum_p f_{j-p}f^*_{j-p} = \sum_p |f_{j-p}|^2$. Since the summation index p ranges from 0 to $N-1$, it holds that $\sum_p |f_{j-p}|^2 = \sum_p |f_p|^2$. Therefore, the variance is given by

$$\sigma^{(f)2}_{jj} = \left(\frac{N-M}{N-1}\frac{1}{M}\right)\left(\left\{\sum_{p=0}^{N-1} |f_p|^2\right\} - |f_j|^2\right) \ . \tag{29}$$

Exercise 5. Verify that when f_j is given by a delta function (i.e., $f_j = e_j$), this variance equals the variance derived in Section 4.2 in the sense that the variance vanishes for $j = 0$ and that it is given by equation (24) for $j \neq 0$.

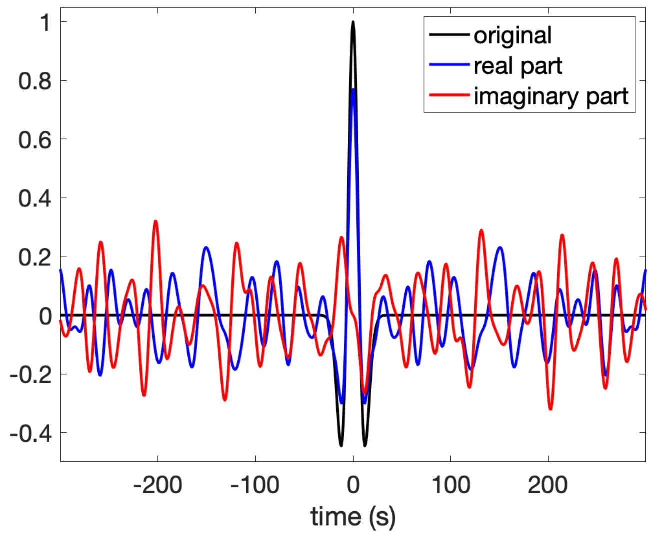

Figure 3. A Ricker wavelet with $\tau = 10$ (black), and the real part (blue) and imaginary part (red) of the Ricker wavelet that is subsampled in frequency without replacement by a factor of 3. There are $N = 600$ points in the time series with sampling interval dt=1 s.

The sum $\sum_p |f_p|^2$ gives the total energy in the time series. For the delta function $f_j = e_j$, the energy is contained in the sample $j = 0$, but for a time series with a longer duration the total energy is much larger than the energy $|f_j|^2$ of any single sample, hence we ignore the contribution $|f_j|^2$ in the expression above. In this approximation the standard deviation of the undersampling noise is given by

$$\sigma_j^{(f)} \approx \sqrt{\frac{N-M}{N-1}} \frac{1}{\sqrt{M}} \sqrt{E} \; , \tag{30}$$

where the total energy in the time series is given by

$$E = \sum_{p=0}^{N-1} |f_p|^2 \; . \tag{31}$$

Note that the right hand side of expression (30) does not depend on j, this means that the subsampling noise is spread out roughly homogeneously across the samples. This behavior is confirmed by the example in Figure 3 and by the subsampled pressure shown by the blue line in Figure 1 where the subsampling noise level is for practical purposes constant in time.

For the Ricker wavelet in expression (27), the energy is given by $E = 9.3686$, and for the number of samples and subsamples used ($N = 600$ and $M = 200$), this leads to a standard deviation of the subsampling noise $\sigma^{(f)} = 0.18$. The standard deviation computed from the noise in Figure 3 is equal to $\sigma \approx 0.15$, which agrees fairly well with the standard deviation predicted by equation (30).

5 RANDOM SUBSAMPLING WITH REPLACEMENT

We next consider the case where the random frequency-domain indices $n_1, n_2, \ldots, n_M$ are drawn independently (with replacement) from the set $\{0, 1, \ldots, N-1\}$. In this case, a frequency may be sampled more than once; we refer to this as *subsampling with replacement.* We again consider the subsampled signal defined in (13), but we note that the summation may include duplicate terms. Despite the possible duplications, the mean of the subsampled time series is again given by (18); that is, random subsampling with replacement is again unbiased.

What does change in the case of subsampling with replacement is the covariance analysis; indeed it is easier than

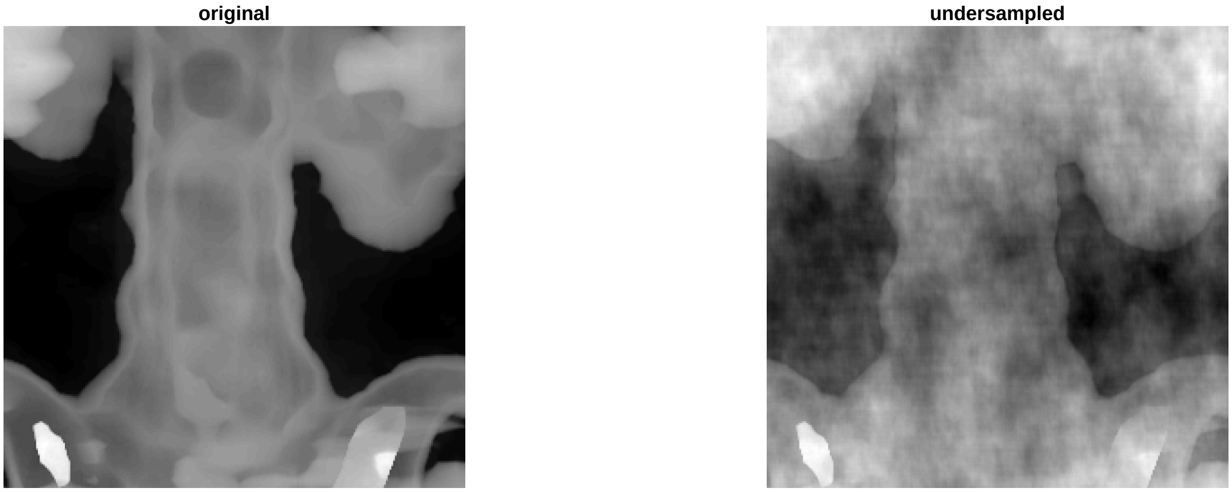

Figure 4. Spine image before (left) and after (right) random undersampling by a factor of 3 in the wavenumber domain.

in the case of subsampling without replacement due to the independence of the sample indices. We show in Section D of the appendix that when the indices $n_1, n_2, \ldots, n_M$ are drawn independently, the covariance of the subsampled delta function $\tilde{e}_j$ is given by

$$C_{j,k} = \frac{1}{M}\left(\delta_{j,k} - \delta_{j,0}\delta_{k,0}\right) \ . \tag{32}$$

This expression contains the same combination of delta functions as expression (20) for the covariance where the chosen frequencies are all different. The only difference is that the factor that multiplies the deltas is different. This means that nearly all of the conclusions listed in Section 4.2 also apply to the case where the frequencies are chosen independently.

Exercise 6. One conclusion that differs from that in Section 4.2 is that when $M = N$ (the number of random indices equals the number of original frequency coefficients), the covariance $C_{j,k}$ in (32) is no longer zero. Why?

The resulting standard deviation of the noise is given by

$$\sigma_j^{\text{with replacement}} = \frac{1}{\sqrt{M}} \ . \tag{33}$$

A comparison of expressions (26) and (33) for the standard deviation of the noise introduced by subsampling without and with replacement shows that

$$\sigma_j^{\text{no replacement}} = \sqrt{\frac{N-M}{N-1}}\sigma_j^{\text{with replacement}} < \sigma_j^{\text{with replacement}} \ , \tag{34}$$

when $M > 1$. When frequencies are drawn independently, the amplitude of the noise introduced by the subsampling thus increases compared to the case of random sampling without replacement. The reason for this is that when frequencies are sampled only once, the noise is caused by the absence of frequencies that are not sampled. But when the frequencies are chosen independently, not only are some frequencies missing in the subsampled time series, but also some frequencies are sampled more than once. This introduces an additional source of error that increases the noise level of the subsampled time series.

The arguments used in Section 4.3 for a general time series that is subsampled without replacement apply equally well when subsampling with replacement is used. The only difference is that in expressions (28) through (30) the factor $(N - M)/(N - 1)$ is replaced by 1.

6 CONCLUSION

This tutorial gives insight into why random undersampling, where one undersamples with random intervals, in general is superior to regular undersampling. Whereas regular undersampling in the frequency domain leads to aliasing in the time domain where multiple signal copies periodically overlap with one another, random sampling in the frequency domain tends to introduce noise in the time domain. Increasing the number of frequency samples merely decreases the variance of this noise and allows more and more signal features to stand out above this noise. Alternatively, it is instructive to view the random undersampling process as the convolution of the original signal with a randomly undersampled (thus noisy) delta function.

It is worth noting that, given an undersampled set of a signal's DFT coefficients, the zero-padded inverse DFT (13) is not the only way that one might attempt to estimate the original signal. Indeed, in the field of compressive sensing, the goal is actually to recover a signal's missing coefficients by exploiting the assumption that the signal is sparse in a complementary domain (Candès & Wakin, 2008; Barranca et al., 2016; Khosravy et al., 2020). In the context of our article, when f_j is sparse (containing few spikes) in the time domain, it is possible to recover its missing DFT coefficients. We omit the details here but note that the behavior of the randomly undersampled delta function, $\tilde{e}_j$, plays an important role in compressive sensing theory. In particular, its maximum value away from the origin, $\max_{j\neq 0} |\tilde{e}_j|$, is known as the coherence (Candès & Wakin, 2008; Foucart & Rauhut, 2013) of the sampling pattern. In general, lower coherence is better.

The reader may have noted that although we have demonstrated random subsampling exclusively using real-valued time series f_j, the resulting time series $\tilde{f}_j$ have been complex-valued. For real-valued signals, one could modify the random subsampling procedure in the frequency domain to exclusively produce real-valued estimates, for example by drawing the coefficients in complex-conjugate pairs. Due to complex-conjugate symmetry in the DFT coefficients of a real-valued signal (Oppenheim & Schafer, 2010), once a coefficient F_n is known, its companion $F_{N-n} = F_n^*$ is also known; this effectively halves the number of frequency samples required to characterize a real signal, which is beneficial for data compression. However, we also note that merely taking the real part of the estimator considered in this paper, $\mathrm{Re}(\tilde{f}_j)$, is unbiased. Moreover, its variance is bounded by that of $\tilde{f}_j$: $\mathrm{Var}(\mathrm{Re}(\tilde{f}_j)) \leq \mathrm{Var}(\tilde{f}_j)$. Indeed, Figure 1 shows only the real part of the complex-valued estimate $\tilde{f}_j$.

The conclusions in this paper carry over naturally if one exchanges the roles of time and frequency, that is, if one samples a signal randomly in time and attempts to estimate its frequency spectrum from the partial time samples. The time (frequency) domain can naturally be replaced with the space (wavenumber) domain for problems involving spatial-domain signals rather than time series. More sophisticated analysis appears in contexts such as power spectrum estimation (Ariananda & Leus, 2012), where one assumes a stochastic model for the randomly observed signal.

Finally, the lessons of this article also carry over to multidimensional data. In Figure 4, we show an image of a spine before and after random undersampling without replacement by a factor of 3 in the two-dimensional DFT domain. (For the image with $N = 256 \times 256 = 65536$ pixels, a total of $M = 21845$ sampling indices were drawn uniformly from among all two-dimensional pairs (m, n) in the wavenumber domain; the real part of the estimated image is displayed.) Although artifacts are again introduced, the prominent features are still visible. For a related discussion of the multidimensional case, we refer the reader to Naghizadeh & Sacchi (2010).

REFERENCES

Abma, R. & Foster, M., 2020. *Simultaneous source seismic acquisition*, Geophysical Developments Series, **18**, Society of Exploration Geophysicists, Tulsa, OK.

Abma, R. & Ross, A., 2013. Popcorn shooting: Sparse inversion and the distribution of airgun array energy over time, in *Proc. of the SEG Houston 2013 Annual Meeting*, pp. 31–35, Society of Exploration Geophysicists, Tulsa, OK.

Ariananda, D. & Leus, G., 2012. Compressive wideband power spectrum estimation, *IEEE Transactions on Signal Processing*, **60**(9), 4775–4789.

Barranca, V., Kovačič, G., Zhou, D., & Cai, D., 2016. Improved compressive sensing of natural scenes using localized random sampling, *Scientific Reports*, **6**, 31976.

Caflisch, R., 1998. Monte Carlo and quasi-Monte Carlo methods, *Acta Numerica*, **7**, 1–49.

Candès, E. & Wakin, M., 2008. An introduction to compressive sampling, *IEEE Sign. Proc. Mag.*, pp. 21–30.

Eftekhari, A., Romberg, J., & Wakin, M., 2013. Matched filtering from limited frequency samples, *IEEE Transactions on Information Theory*, **59**(6), 3475–3496.

Foucart, S. & Rauhut, H., 2013. *A mathematical introduction to compressive sensing*, Springer.

Herrmann, F., 2010. Randomized sampling and sparsity: Getting more informaton from fewer samples, *Geophysics*, **75**, WB173–WB187.

Hogg, R. & Craig, T., 1978. *Introduction to mathematical statistics*, Macmillan, New York, 4th edn.

Khosravy, M., Gupta, N., Patel, N., & Duque, C., 2020. Recovery in compressive sensing: a review, in *Compressive sensing in healthcare*, pp. 25–42, Academic Press.

Lustig, M., Donoho, D., Santos, J., & Pauly, J., 2008. Compressed sensing MRI, *IEEE Signal Processing Magazine*, **25**(2), 72–82.

Naghizadeh, M. & Sacchi, M., 2010. On sampling functions and Fourier reconstruction methods, *Geophysics*, **75**(6), WB137–WB151.

Oppenheim, A. & Schafer, R., 2010. *Discrete-Time Signal Processing, 3rd Edition*, Pearson.

Snieder, R. & Hagerty, M., 2004. Monitoring change in volcanic interiors using coda wave interferometry: Application to Arenal Volcano, Costa Rica, *Geophys. Res. Lett.*, **31**, L09608.

Snieder, R. & van Wijk, K., 2015. *A Guided Tour of Mathematical Methods for the Physical Sciences*, Cambridge Univ. Press, Cambridge, UK, 3rd edn.

APPENDIX A: THE COVARIANCE OF THE RANDOMLY SUBSAMPLED DELTA FUNCTION

Here we derive expressions for the covariance of a randomly subsampled delta function. The original delta function is given by

$$e_j = \frac{1}{N}\sum_{n=0}^{N-1} e^{i2\pi jn/N} = \delta_{j,0} \ . \tag{A1}$$

According to expression (14) of the main text, the randomly subsampled delta function is given by

$$\tilde{e}_j = \frac{1}{M}\sum_{m=1}^{M} e^{i2\pi jn_m/N} \ , \tag{A2}$$

where $n_1, n_2, \ldots, n_M$ denote a random subset of indices drawn without replacement from the set $\{0, 1, \ldots, N-1\}$.

The covariance of the randomly sampled delta function $\tilde{e}_j$ is given by:

$$C_{j,k} = \mathbb{E}\left(\left(\tilde{e}_j - \mathbb{E}(\tilde{e}_j)\right)\left(\tilde{e}_k - \mathbb{E}(\tilde{e}_k)\right)^*\right) \ . \tag{A3}$$

In Sections B and C, we consider the case where distinct frequency components are used, i.e., random sampling without replacement. In Section D, we consider the case of random frequency sampling with replacement. In this analysis we will use the fact that the randomly sampled delta function is unbiased, that is, as described in equation (17) of the main text,

$$\mathbb{E}\left(\tilde{e}_j\right) = e_j \ , \tag{A4}$$

so that the covariance (A3) is equal to

$$C_{j,k} = \mathbb{E}\left(\left(\tilde{e}_j - e_j\right)\left(\tilde{e}_k - e_k\right)^*\right) \ . \tag{A5}$$

The expectation value is not a random variable, hence taking the expectation value again does not change the value: $\mathbb{E}(\mathbb{E}(\tilde{e}_j)) = \mathbb{E}(\tilde{e}_j)$. Using this property and expanding the product in expression (A5) gives

$$C_{j,k} = \mathbb{E}\left(\tilde{e}_j\tilde{e}_k^*\right) - \mathbb{E}(\tilde{e}_j)\mathbb{E}(\tilde{e}_k^*) \ . \tag{A6}$$

Using expression (A4) gives

$$C_{j,k} = \mathbb{E}\left(\tilde{e}_j\tilde{e}_k^*\right) - e_j e_k \ , \tag{A7}$$

where we used that according to (A1), e_k is real, hence $e_k^* = e_k$.

APPENDIX B: THE COVARIANCE FOR RANDOM SAMPLING WITHOUT REPLACEMENT

To compute the covariance we need to specify the joint probability of n_m and $n_{m'}$. In the randomly sampled data only M frequencies are chosen out of the total of N possible frequencies. In this section we consider the case where these no frequency is used twice, hence

$$n_m \neq n_{m'} \quad \text{for } m \neq m' \ . \tag{B1}$$

In the following we need the joint probability distribution of two random indices n_m and $n_{m'}$. According to (B1), for any $\alpha \in \{0, 1, \ldots, N-1\}$, the joint probability that $n_m = n_{m'} = \alpha$ is zero. Meanwhile, there are $N(N-1)$ pairs of nonequal indices drawn from the set $\{0, 1, \ldots, N-1\}$, and each pair is equally likely. Thus, the joint probability distribution of $(n_m, n_{m'})$ is uniform over these pairs; that is, each nonequal pair occurs with probability $1/N(N-1)$. In summary, for $m \neq m'$,

$$\Pr(n_m = \alpha, n_{m'} = \beta) = \begin{cases} 0, & \alpha = \beta, \\ \frac{1}{N(N-1)}, & \alpha \neq \beta. \end{cases} \tag{B2}$$

To compute the covariance in expression (A7) we need to evaluate $\mathbb{E}\left(\tilde{e}_j \tilde{e}_k^*\right)$. Using equation (A2) this expectation value is given by

$$\mathbb{E}\left(\tilde{e}_j \tilde{e}_k^*\right) = \frac{1}{M^2} \sum_{m=1}^{M} \sum_{m'=1}^{M} \mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) \ . \tag{B3}$$

The double sum can be decomposed into terms $m = m'$ and $m \neq m'$: $\sum_{m,m'} = \sum_{m=m'} + \sum_{m \neq m'}$, hence

$$\mathbb{E}\left(\tilde{e}_j \tilde{e}_k^*\right) = \frac{1}{M^2} \sum_{m=1}^{M} \mathbb{E}\left(e^{i2\pi(j-k)n_m/N}\right) + \frac{1}{M^2} \sum_{m \neq m'} \mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) \ . \tag{B4}$$

For the first term in the right hand side we use expression (17) of the main text and equation (A1) of the appendix to give

$$\mathbb{E}\left(e^{i2\pi(j-k)n_m/N}\right) = \delta_{j-k,0} = \delta_{j,k} \ . \tag{B5}$$

This expectation value does not depend on m, hence the sum over m in the first term in the right hand side of equation (B4) amounts to a multiplication with M, hence

$$\sum_{m=1}^{M} \mathbb{E}\left(e^{i2\pi(j-k)n_m/N}\right) = M\delta_{j,k} \ . \tag{B6}$$

For the second term in the right hand side of equation (B4) we compute the expectation using the joint probability distribution of n_m and $n_{m'}$ given in (B2): when $m \neq m'$,

$$\begin{aligned} \mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) &= \sum_{\alpha \neq \beta} e^{i2\pi j \alpha/N} e^{-i2\pi k \beta/N} \cdot \Pr(n_m = \alpha, n_{m'} = \beta) \\ &= \frac{1}{N(N-1)} \sum_{\alpha \neq \beta} e^{i2\pi j \alpha/N} e^{-i2\pi k \beta/N}. \end{aligned} \tag{B7}$$

If all values α, β were included in the sum, we could factorize the sum and write

$$\sum_{\alpha,\beta} e^{i2\pi j \alpha/N} e^{-i2\pi k \beta/N} = \left(\sum_{\alpha=0}^{N-1} e^{i2\pi j \alpha/N}\right) \left(\sum_{\beta=0}^{N-1} e^{i2\pi k \beta/N}\right)^* = N^2 \delta_{j,0} \delta_{k,0} \ , \tag{B8}$$

where expression (A1) is used in the last identity. In order to rewrite the double sum in equation (B7) in the form of equation (B8) we use that $\sum_{\alpha \neq \beta} = \sum_{\alpha,\beta} - \sum_{\alpha=\beta}$, so that expression (B7) can be written as

$$\mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) = \frac{1}{N(N-1)} \sum_{\alpha,\beta} e^{i2\pi j \alpha/N} e^{-i2\pi k \beta/N} - \frac{1}{N(N-1)} \sum_{\alpha=0}^{N-1} e^{i2\pi(j-k)\alpha/N} \ . \tag{B9}$$

The double sum is given by equation (B8), while according to expression (A1) the single sum is equal to $N\delta_{j-k,0} = N\delta_{j,k}$, so that

$$\mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) = \frac{N}{N-1} \delta_{j,0} \delta_{k,0} - \frac{1}{N-1} \delta_{j,k} \ . \tag{B10}$$

This expression is independent of m and m', so that the double sum $\sum_{m\neq m'}$ in equation (B4) corresponds to a multiplication with $M(M-1)$, hence

$$\sum_{m\neq m'} \mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) = M(M-1)\left(\frac{N}{N-1}\delta_{j,0}\delta_{k,0} - \frac{1}{N-1}\delta_{j,k}\right) . \tag{B11}$$

Inserting expressions (B6) and (B11) into equation (B4) gives

$$\mathbb{E}\left(\tilde{e}_j \tilde{e}_k^*\right) = \frac{N-M}{M(N-1)}\delta_{j,k} + \frac{N(M-1)}{M(N-1)}\delta_{j,0}\delta_{k,0} . \tag{B12}$$

Using this in equation (A7), and using expression (A1) for e_j and e_k, gives, after a rearrangement of terms

$$C_{j,k} = \frac{(N-M)}{(N-1)}\frac{1}{M}\left(\delta_{j,k} - \delta_{j,0}\delta_{k,0}\right) , \tag{B13}$$

APPENDIX C: THE COVARIANCE FOR RANDOM SAMPLING WITHOUT REPLACEMENT FOR A GENERAL TIME SERIES

In this section we derive the covariance matrix for a subsampled time series from the covariance of a subsampled delta function. The covariance of $\tilde{f}$ is defined as

$$C_{j,k}^{(f)} = \mathbb{E}\left(\left(\tilde{f}_j - \mathbb{E}(\tilde{f}_j)\right)\left(\tilde{f}_k - \mathbb{E}(\tilde{f}_k)\right)^*\right) \tag{C1}$$

We first use that the time series f_j does not depend on the subsampling, but the subsampled delta function $\tilde{e}_j$ does depend on the subsampling. This means that using expression (18) of the main text, the subsampled time series $\tilde{f}_j$ satisfies

$$\mathbb{E}\left(\tilde{f}_j \tilde{f}_k^*\right) = \mathbb{E}\left(\sum_{p,q} \tilde{e}_p f_{j-p} \tilde{e}_q^* f_{k-q}^*\right) = \sum_{p,q} f_{j-p} f_{k-q}^* \mathbb{E}\left(\tilde{e}_p \tilde{e}_q^*\right) , \tag{C2}$$

where it is understood in this section that the convolutions are cyclical, i.e., integer multiples of N are added or subtracted to the subscripts of f so that the index lies in the interval $0, 1, \ldots, N-1$. The second identity in equation (C2) is due to the fact that f_j is not a statistical variable, but $\tilde{e}_j$ is, hence the original time series f_j is just a multiplicative constant as far as the expectation value is concerned.

The expectation value of the expectation value satisfies $\mathbb{E}(\mathbb{E}(\tilde{f}_j)) = \mathbb{E}(\tilde{f}_j)$, hence the covariance $C_{j,k}^{(f)}$ of the subsampled time series satisfies $C_{j,k}^{(f)} = \mathbb{E}(\tilde{f}_j \tilde{f}_k^*) - \mathbb{E}(\tilde{f}_j)\mathbb{E}(\tilde{f}_k^*)$. Using expression (C2) and using that $\tilde{f}$ is the convolution of f_j with the subsampled delta function $\tilde{e}_j$, so that $\tilde{f}_j = \sum_p \tilde{e}_p f_{j-p}$, and it follows that

$$C_{j,k}^{(f)} = \sum_{p,q} f_{j-p} f_{k-q}^* \left\{\mathbb{E}(\tilde{e}_p \tilde{e}_q^*) - \mathbb{E}(\tilde{e}_p)\mathbb{E}(\tilde{e}_q^*))\right\} . \tag{C3}$$

Following expression (A6) the quantity between curly brackets is the covariance $C_{p,q}$ is the subsampled delta function, hence

$$C_{j,k}^{(f)} = \sum_{p,q} f_{j-p} f_{k-q}^* C_{p,q} . \tag{C4}$$

The covariance of any subsampled time series thus follows from the covariance of the subsampled delta function. The covariance of the delta function $C_{p,q}$ is given by equation (20) of the main text. The first delta functions in expression (20) of the main text gives a contribution $\sum_{p,q} f_{j-p} f_{k-q}^* \delta_{p,q} = \sum_p f_{j-p} f_{k-p}^*$. The second group of delta functions in equation (20) of the main text gives a contribution $\sum_{p,q} f_{j-p} f_{k-q}^* \delta_{p,q}\delta_{p,0}\delta_{q,0} = f_j f_k^*$. Using this in (C4) gives

$$C_{j,k}^{(f)} = \left(\frac{N-M}{N-1}\frac{1}{M}\right)\left(\left\{\sum_{p=0}^{N-1} f_{j-p} f_{k-p}^*\right\} - f_j f_k^*\right) . \tag{C5}$$

APPENDIX D: THE COVARIANCE FOR RANDOM SAMPLING WITH REPLACEMENT

We next consider the case where the M indices $n_1, n_2, \ldots, n_M$ are chosen independently from the uniform distribution on $\{0, 1, \ldots, N-1\}$. Consequently, a frequency may be selected more than once. In this case the treatment of

Sections A and B is unchanged up to equation (B6), with the exception that the joint probability distribution of $(n_m, n_{m'})$ (assuming $m \neq m'$) is given by

$$\Pr(n_m = \alpha, n_{m'} = \beta) = \Pr(n_m = \alpha)\Pr(n_{m'} = \beta) = \frac{1}{N^2} \tag{D1}$$

for any $\alpha, \beta \in \{0, 1, \ldots, N-1\}$.

The fact that the indices n_m are chosen independently simplifies the derivation. Indeed, since the expectation of a product of independent random variables equals the product of expectations (Hogg & Craig, 1978), the expectation in the second term in the right hand side of equation (B4) simplifies to

$$\mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) = \mathbb{E}\left(e^{i2\pi j n_m/N}\right)\mathbb{E}\left(e^{-i2\pi k n_{m'}/N}\right) = \delta_{j,0}\delta_{k,0}. \tag{D2}$$

This expectation value does not depend on m and m'. Since there are $M(M-1)$ terms in the sum $\sum_{m\neq m'}$ the summation over m and m' leads to a multiplication with $M(M-1)$, hence

$$\sum_{m\neq m'} \mathbb{E}\left(e^{i2\pi j n_m/N} e^{-i2\pi k n_{m'}/N}\right) = M(M-1)\delta_{j,0}\delta_{k,0}\ . \tag{D3}$$

Inserting expressions (B6) and (D3) into equation (B4) gives

$$\mathbb{E}\left(\tilde{e}_j \tilde{e}_k^*\right) = \frac{1}{M}\delta_{j,k} + \frac{M-1}{M}\delta_{j,0}\delta_{k,0}\ . \tag{D4}$$

Using this in the covariance (A7) and using equation (A1) gives

$$C_{j,k} = \frac{1}{M}\left(\delta_{j,k} - \delta_{j,0}\delta_{k,0}\right)\ . \tag{D5}$$