Modelling coloured residual noise in gravitational-wave signal processing

# Modelling coloured residual noise in gravitational-wave signal processing

**Christian Röver[1], Renate Meyer[2] and Nelson Christensen[3]**

[1] Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut) and Leibniz Universität Hannover, Hannover, Germany
[2] Department of Statistics, The University of Auckland, Auckland, New Zealand
[3] Physics and Astronomy, Carleton College, Northfield, MN, USA

E-mail: christian.roever@aei.mpg.de

## Abstract

We introduce a signal processing model for signals in non-white noise, where the exact noise spectrum is *a priori* unknown. The model is based on a *Student's t distribution* and constitutes a natural generalization of the widely used normal (Gaussian) model. This way, it allows for uncertainty in the noise spectrum, or more generally is also able to accommodate outliers (heavy-tailed noise) in the data. Examples are given pertaining to data from gravitational-wave detectors.

PACS numbers: 04.80.Nn, 02.30.Zz, 05.45.Tp

## 1. Introduction

The measurement of gravitational radiation holds great promise for exciting astronomical observations [1, 2]. Around the world, efforts are under way to construct and improve detectors for gravitational waves; among these are the LIGO detectors in the USA [3], GEO [4] and Virgo [5] in Europe, and TAMA in Japan [6], some of which have already started taking data. Plans for future detectors include advanced LIGO [7], LCGT in Japan [8], the Laser Interferometer Space Antenna (LISA) [9] and the Einstein Telescope (ET) [10]. As existing instruments are becoming increasingly sensitive and future instruments are approaching completion, sophisticated signal processing methods are required in order to detect and accurately interpret these often weak signals within a noisy environment. For example, signal detection is usually implemented via matched filtering [11–13], while parameter estimation is commonly done using Bayesian methods [14, 15]. Many of the data analysis procedures employed to date are based on the assumption of the noise being Gaussian with a known power spectral density [11, 16]. While these methods have proven to be computationally efficient and very powerful at discriminating rare, weak signals within noise, some more flexibility or robustness is sometimes desired. One example is the analysis of data to be expected from LISA, where the measurement noise originates partly from instrumental as well as astrophysical sources, and the noise spectrum is only vaguely known *a priori* [17].

We introduce an approach that was developed in the context of the latter case, where, along with the signal parameters to be inferred, the noise's power spectrum needs to be incorporated as an unknown into the model [18, 19]. The model developed here turns out to be computationally convenient, as the additional noise parameters may be analytically integrated out, leading to a Student-*t* likelihood expression instead of the original normal formulation. We expect the same approach to be useful in many related signal processing contexts, as it allows us to specify the prior information about the power spectrum, i.e. its expected magnitude and the associated certainty, in a straightforward manner. In this way, it should e.g. also be able to account for nonstationarities in the data, in the sense that the power spectrum does not necessarily need to be assumed to be exactly the same as at an earlier measurement; other fluctuations, like outliers, could also be accommodated in a similar manner. In fact, the resulting model actually falls into a class of generalizations to the standard Gaussian that have been proposed for their robustness properties: models accommodating heavier-tailed noise, or outliers, by either ('directly') implementing a non-Gaussian noise model or ('indirectly') substituting the corresponding least-squares procedures by less outlier-sensitive methods. Instead of the derivation based on an unknown power spectrum, the same model may be motivated by assuming the power spectrum itself to be random, so that the resulting noise *mixture distribution* exhibits a greater variability. In the limiting case of decreasing spectrum variability, the model again simplifies to the Gaussian. So the model will also be applicable as an ad hoc alternative tunable robust model, with clearly interpretable 'tuning parameters'.

The organization of the paper is as follows. The time series setup is introduced in section 2.1. In the subsequent sections, the probabilistic modelling is described, including its time-domain counterpart, the likelihood, prior distributions, posterior distribution, marginal likelihood and some implications. Section 3 describes the approach using two illustrative examples of simulated time series with and without a signal. Section 4 puts the new approach into context. An appendix explicating the discrete Fourier transform conventions used in this paper is attached.

## 2. The time series model

### 2.1. The setup

Consider a time series $x_1, \ldots, x_N$ of $N$ real-valued observations sampled at constant time intervals $\Delta_t$, so that each observation $x_i$ corresponds to time $t_i = i \Delta_t$. This set of $N$ observations can equivalently be expressed in terms of sinusoids of the Fourier frequencies:

$$x_i = \frac{1}{\sqrt{N \Delta_t}} \sum_{j=0}^{\lfloor N/2 \rfloor} a_j \cos(2\pi f_j t_i) + b_j \sin(2\pi f_j t_i) \tag{1}$$

where the variables $a_j$ and $b_j$ each correspond to Fourier frequencies $f_j = j \Delta_f = \frac{j}{N \Delta_t}$. The summation in (1) runs from $j = 0$ to $j = \lfloor N/2 \rfloor$, with $\lfloor N/2 \rfloor$ denoting the largest integer less than or equal to $N/2$. By definition, $b_0$ is always zero, and $b_{\lfloor N/2 \rfloor}$ is zero if $N$ is even; going over from $x_i$'s to $a_j$'s and $b_j$'s again yields the same number ($N$) of non-zero figures. The set of ($N$) *frequency-domain* coefficients $a_j$ and $b_j$ and the *time-domain* observations $x_i$ are related to each other through a discrete Fourier transform (and appropriate scaling; see the appendix). The set of trigonometric functions in (1) constitutes an orthonormal basis of the sample space, so that there is a unique one-to-one mapping of the observations in time and in frequency

domain. Instead of the two amplitudes $a_j$ and $b_j$, the definition in (1) may equivalently be expressed in terms of a single amplitude and a phase parameter at each frequency:

$$x_i = \frac{1}{\sqrt{N\Delta_t}} \sum_{j=0}^{\lfloor N/2 \rfloor} \lambda_j \, \sin(2\pi f_j t_i + \varphi_j), \tag{2}$$

where $\lambda_j = \sqrt{a_j^2 + b_j^2}$ and

$$\varphi_j = \begin{cases} \arctan\left(\frac{b_j}{a_j}\right) & \text{if } a_j > 0, \\ \arctan\left(\frac{b_j}{a_j}\right) \pm \pi & \text{if } a_j < 0. \end{cases} \tag{3}$$

For each Fourier frequency $f_j$, let $\kappa_j$ be the number of Fourier coefficients not being zero by definition, i.e.

$$\kappa_j = \begin{cases} 1 & \text{if } (j = 0) \text{ or } (N \text{ is even and } j = N/2) \\ 2 & \text{otherwise} \end{cases} \tag{4}$$

so that $\sum_{j=0}^{\lfloor N/2 \rfloor} \kappa_j = N$. Note that generally one may often simplify to $\kappa_j = 2 \; \forall j$ without introducing a noticeable discrepancy, but in the following we will stick to the accurate notation of $\kappa$ depending on the index $j$.

For a *given* time series (either in terms of $x_i$ or $a_j$ and $b_j$), we define the functions of the Fourier frequencies

$$p_1(f_j) = \frac{a_j^2 + b_j^2}{\kappa_j} \qquad \text{and} \tag{5}$$

$$p_2(f_j) = \frac{a_j^2 + b_j^2}{\kappa_j^2} = \frac{p_1(f_j)}{\kappa_j} = \frac{\Delta_t}{N} |\tilde{x}(f_j)|^2 \tag{6}$$

for $j = 0, \ldots, \lfloor N/2 \rfloor$, where $\tilde{x}$ denotes the discretely Fourier-transformed time series $x$, as defined in the appendix. These are the empirical, discrete analogues of the *one-sided* and *two-sided* spectral power. The set of $p_1(f_j)$ is also known as the *periodogram* of the time series [20].

Now consider the case where the observations $x_i$, and consequently the $a_j$ and $b_j$, correspond to the random variables $X_i$, $A_j$ and $B_j$, respectively. This may mean that these are realizations of a random process, where the random variables' probability distributions describe the *randomness* in the observations, or that they are merely unknown, where the probability distributions describe a *state of information*, or it may also be a mélange of both [21].

The time series has a zero mean if and only if the expectation of all frequency-domain coefficients is zero as well:

$$\text{E}[X_i] = 0 \quad \forall i \qquad \Leftrightarrow \qquad \text{E}[A_j] = \text{E}[B_j] = 0 \quad \forall j. \tag{7}$$

For the probabilistic time series, the spectral power ($p_1(f_j)$ or $p_2(f_j)$) consequently also is a random variable ($P_1(f_j)$ or $P_2(f_j)$, respectively). We denote its expectation value by $S^*(f_j)$; if the mean is zero as in (7), it is given by

$$S_1^*(f_j) = \kappa_j S_2^*(f_j) = \text{E}[P_1(f_j)] = \text{E}\left[\frac{A_j^2 + B_j^2}{\kappa_j}\right] \stackrel{(7)}{=} \frac{\text{Var}(A_j) + \text{Var}(B_j)}{\kappa_j}. \tag{8}$$

It is important to note that while the above expectation is closely related to a time series' power spectral density, it is yet quite different. Up to here we have considered the discretely

Fourier transformed data and some of its basic properties. The figure $S^*(f_j)$ refers to data sampled at a particular resolution and sample size and is defined only at a discrete set of Fourier frequencies $f_j$. One might want to refer to $S^*$ as the *discretized* power spectrum. *If* in fact the data are a realization from a stationary random process with (continuous) power spectral density $S(f)$, then the expectation $S^*(f_j)$ is related to $S(f)$ through a convolution, depending on the sample size $N$. Only in the limit of an infinite sample size both of them are equal [20, 22]. While both continuous and discretized spectra are commonly used as an approximation to or in place of one another, it is crucial to realize that in the following inference will be done with reference to the case of a finite sample size and the discretized, convolved spectrum $S^*(f_j)$.

A related useful figure in this context is the *integrated spectrum* with respect to some frequency range $[f_1, f_2]$ [23], which here we define as

$$\mathcal{I}_{[f_1, f_2]} = \Delta_f \sum_{j=j_1}^{j_2} \frac{\kappa_j}{2} S_1^*(f_j) \tag{9}$$

where the summation ranges over the corresponding Fourier frequencies ($j_1 = \min\{k : f_k > f_1\}$, $j_2 = \max\{k : f_k \leqslant f_2\}$, where $f_2 - f_1 \geqslant \Delta_f$). The integrated spectrum allows us to investigate or compare (discrete) spectra independent of the particular sample size $N$.

## 2.2. The normal model

Suppose the moments as defined in (7) and (8) are given. We may set up a corresponding normal time series model by assuming all the (frequency-domain) observables to be stochastically independent, have zero mean and

$$\text{Var}(A_j) = \sigma_j^2, \tag{10}$$

$$\text{Var}(B_j) = (\kappa_j - 1)\,\sigma_j^2 \quad \text{for} \quad j = 0, \dots, \lfloor N/2 \rfloor, \tag{11}$$

where

$$\sigma_j^2 = S_1^*(f_j) = \kappa_j\,S_2^*(f_j) \quad \text{for } j = 0, \dots, \lfloor N/2 \rfloor. \tag{12}$$

While other choices of variance settings matching the assumptions would be possible, the assumption of equal variances for $A_j$ and $B_j$ is the only one that makes the joint density of $(A_j, B_j)$ a function of overall amplitude $\lambda_j$ alone, independent of the phase $\varphi_j$, and with that leaves the model invariant with respect to time shifts. The same model may also be motivated via the *maximum entropy principle*, as the (zero mean, equal variance, independent) normal distribution maximizes the entropy, given the constraints on the moments given in (7) and (8) above [24]. In that way, the normal model constitutes a most conservative model setup under the given assumptions [21, 22, 24].

The joint normal distribution derived here has actually been commonly used earlier (see e.g. [25]); the normality assumption may not only appear as a natural choice, but will also turn out computationally convenient in the following. In the context of Fourier-domain data in particular, the normality assumption may also be motivated via asymptotic arguments, by considering the limit of an infinite observation time [26–28]. The resulting normal model also is exactly the same as the one underlying the so-called Whittle likelihood [29, 30], or the one at the basis of *matched filtering* [11–13]. While intuitively in other contexts it may often be sufficient to point out the asymptotic equality $S^*(f_j) \approx S(f_j)$ for $N \to \infty$, here it is crucial to appreciate what exactly the $\sigma_j^2 = S_1^*(f_j)$ stand for in the case of a finite sample size $N$. The difference between the *exact* and *approximate* ('Whittle') model is explicated in more detail in [30].

### 2.3. The corresponding time-domain model

An immediate consequence of the normality assumption for the frequency-domain coefficients $(A_j, B_j)$ is that the time-domain variables $(X_i)$, being linear combinations of the frequency-domain variables, also follow a normal distribution. The exact joint distribution of the $X_i$ is completely determined by their variance/covariance structure, which may be expressed in terms of the autocovariance function. The covariance for any pair of time-domain observations $X_m$ and $X_n$ (with $m, n \in \{1, \ldots, N\}$, and corresponding to times $t_m$ and $t_n$) is given by

$$\text{Cov}(X_m, X_n) \;=\; \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \left( S_1^*(f_j) \tfrac{\kappa_j}{2} \cos(2\pi f_j (t_n - t_m)) \right). \tag{13}$$

Note that the autocovariance $\text{Cov}(X_m, X_n) = \gamma(t_n - t_m)$ depends on $t_m$ and $t_n$ only via their time lag $t_n - t_m$. Remarkably (though not surprisingly), following either the invariance or the maximum entropy argument to motivate the model structure in section 2.2, the result is a *strictly stationary* model for data [31]. Note also that $\gamma(0) = \text{Var}(X_i) = \mathcal{I}_{[0, f_{N/2}]}$.

   The autocovariance function allows one to express the distribution of the $X_i$ in terms of their ($N \times N$) covariance matrix $\Sigma_X$, which is the symmetric, square Toeplitz matrix

$$\Sigma_X \;=\; \begin{pmatrix} \gamma(0) & \gamma(\Delta_t) & \gamma(2\Delta_t) & \cdots & \gamma(\Delta_t) \\ \gamma(\Delta_t) & \gamma(0) & \gamma(\Delta_t) & \cdots & \gamma(2\Delta_t) \\ \gamma(2\Delta_t) & \gamma(\Delta_t) & \gamma(0) & \cdots & \gamma(3\Delta_t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(\Delta_t) & \gamma(2\Delta_t) & \gamma(3\Delta_t) & \cdots & \gamma(0) \end{pmatrix}, \tag{14}$$

since $\gamma$ is periodic such that $\gamma(i\Delta_t) = \gamma((N-i)\Delta_t)$. The periodic formulation in (1) makes the first and last observations $X_1$ and $X_N$ 'neighbours', just as $X_1$ and $X_2$ are. This may seem odd (depending on the context, of course), but is not an unusual problem, as it arises for any conventional spectral analysis of discretely sampled data in the form of spectral leakage [32]; it may not always be as obvious in its consequences. The above time-domain expression sheds some light on the exact shortcomings of the Whittle likelihood approximation; you can see for example that it will be a poor approximation in case of predominantly low-frequency noise and a small number of observations. The problem may be tackled via windowing of the data, or it will also lessen with an increasing sample size; again, the features of this approximation are discussed in more detail in [30].

### 2.4. Incorporating an unknown spectrum

*2.4.1. The likelihood function.* Now suppose the spectrum is unknown and to be inferred from the data $x_1, \ldots, x_N$. An unknown spectrum here is equivalent to the variance parameters $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$ being unknown. Assuming normality of $A_j$ and $B_j$, the likelihood function (as a function of the parameters $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$) is

$$\begin{aligned} p\big(x_1, \ldots, x_N \mid \sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2\big) \\ = \; \prod_{j=0}^{\lfloor N/2 \rfloor} \Bigg[ \frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left( -\frac{a_j^2}{2\sigma_j^2} \right) \\ \times \left( \frac{1}{\sqrt{2\pi}\,\sigma_j} \right)^{\kappa_j - 1} \exp\left( -\frac{b_j^2}{2\sigma_j^2} \right) \Bigg] \end{aligned} \tag{15}$$

$$= \exp\left( -\frac{N}{2}\log(2\pi) - \sum_{j=0}^{\lfloor N/2 \rfloor} \left[ \kappa_j \log(\sigma_j) + \frac{a_j^2 + b_j^2}{2\sigma_j^2} \right] \right) \qquad (16)$$

$$\propto \exp\left( -\sum_{j=0}^{\lfloor N/2 \rfloor} \frac{\kappa_j}{2} \left[ \log\big(S_2^*(f_j)\big) + \frac{\frac{\Delta_t}{N}|\tilde{x}_j|^2}{S_2^*(f_j)} \right] \right). \qquad (17)$$

The term $\tilde{x}_j$ here denotes the $j$th element of the Fourier-transformed data vector, as defined in the appendix. The above likelihood (17) is exactly equivalent to the so-called Whittle likelihood, which is an approximate expression for stationary time series [29, 30].

If the noise spectrum is *a priori* known, the '$\log\big(S_2^*(f_j)\big)$' term may also be irrelevant (see (31)), and the model again reduces to the normal model described e.g. in [11] that is also the basis for matched filtering [12, 13]. In that case, the logarithmic likelihood may be conveniently expressed in terms of an inner product of vectors, allowing for a geometric interpretation that is especially useful in the case of linear signal models [11, 16, 33].

*2.4.2. The conjugate prior distribution.* For the model defined above, the conjugate prior distribution for each of the $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$ is the *scaled inverse $\chi^2$-distribution* with scale parameter $s_j^2$ and degrees-of-freedom parameter $\nu_j$:

$$\sigma_j^2 \; \sim \; \text{Inv-}\chi^2\big(\nu_j, s_j^2\big) \qquad (18)$$

with the density function

$$f_{\nu_j, s_j^2}\big(\sigma_j^2\big) \; = \; \frac{\left(\frac{\nu_j s_j^2}{2}\right)^{\nu_j/2}}{\Gamma(\nu_j/2)} \big(\sigma_j^2\big)^{-(1+\nu_j/2)} \exp\left( -\frac{\nu_j s_j^2}{2\sigma_j^2} \right) \qquad (19)$$

[34]. The degrees-of-freedom here denote the precision in the prior distribution, while the scale determines its order of magnitude. For increasing $\nu_j$ the distribution's variance goes towards zero, and for $\nu_j \to 0$ the density converges towards the non-informative (and improper) distribution with density $f\big(\sigma_j^2\big) = \frac{1}{\sigma_j^2}$, that is uniform on $\log\big(\sigma_j^2\big)$, and which also constitutes the corresponding *Jeffreys prior* for this problem [35]. If $\sigma_j^2$ follows an Inv-$\chi^2\big(\nu_j, s_j^2\big)$ distribution, then its expectation and variance are given by

$$\text{E}\big[\sigma_j^2\big] = \frac{\nu_j}{\nu_j - 2} s_j^2 \quad \text{and} \quad \text{Var}(\sigma_j^2) = \frac{2\nu_j^2}{(\nu_j - 2)^2(\nu_j - 4)} s_j^4, \qquad (20)$$

and the mean and variance are finite for $\nu_j > 2$ and $\nu_j > 4$, respectively [34].

In addition to the Jeffreys prior (with $\nu = 0$) already mentioned above, other improper prior distributions may be implemented as special cases of an Inv-$\chi^2(\nu, s^2)$ distribution. A uniform prior distribution on $\sigma$ corresponds to ($\nu = -1, s^2 = 0$), and a uniform prior on $\sigma^2$ corresponds to ($\nu = -2, s^2 = 0$). Generally, a prior with density $f(\sigma^2) \propto \frac{1}{(\sigma^2)^k}$ (for $k \geqslant 0$) corresponds to an Inv-$\chi^2(\nu = 2(k-1), s^2 = 0)$ distribution [36]. As usual, care must be taken when using such improper priors, as the resulting posterior may not be a proper probability distribution [34].

*2.4.3. The posterior distribution.* Due to the use of a *conjugate* prior distribution, the posterior distribution of the $\sigma_j^2$ for given data again is of the same family. The posterior

density is defined by the product of prior (19) and likelihood (16):

$$
p\big(\sigma_j^2 \,\big|\, \{x_1, \ldots, x_N\}\big) \propto \big(\sigma_j^2\big)^{-(1+\nu_j/2)} \, \exp\!\left(-\frac{\nu_j \, s_j^2}{2\,\sigma_j^2}\right)
$$

$$
\times \; \big(\sigma_j^2\big)^{-\kappa_j/2} \, \exp\!\left(-\frac{a_j^2 + b_j^2}{2\,\sigma_j^2}\right) \tag{21}
$$

$$
= \big(\sigma_j^2\big)^{-(1+\frac{\nu_j+\kappa_j}{2})} \, \exp\!\left(-\frac{\nu_j\, s_j^2 + a_j^2 + b_j^2}{2\,\sigma_j^2}\right) \tag{22}
$$

which again can be recognized as a scaled inverse $\chi^2$-density (cp. (19)):

$$
\sigma_j^2 \,\big|\, \{x_1, \ldots, x_N\} \;\sim\; \text{Inv-}\chi^2\!\left(\nu_j + \kappa_j, \; \frac{\nu_j s_j^2 + a_j^2 + b_j^2}{\nu_j + \kappa_j}\right), \tag{23}
$$

where all the different parameters corresponding to different frequencies are mutually independent. Comparing prior and posterior parameters ((18), (23)), and the way the prior and likelihood are combined (21), one can see that the prior distribution might be thought of as providing the information equivalent to $\nu_j$ observations (of coefficients $a_j$ or $b_j$) with an average squared deviation $s_j^2$ [34]. The prior is essentially of the same functional form as the likelihood (21), modulo a $\frac{1}{\sigma^2}$ term, which resembles an overall (uninformative) *Jeffreys prior* prefactor [35]. The use of the conjugate prior distribution hence is not only a computationally convenient choice, but may also appear as a 'natural' way of expressing prior information in this context.

The use of the conjugate prior distribution leads to a convenient expression for the posterior distribution of the variance parameters $\sigma_j^2$, and with that of the complete discrete spectrum. Also, if $\sigma_j^2 \sim \text{Inv-}\chi^2\big(\nu_j, s_j^2\big)$, then, since $s_j^2$ is a scale parameter, it follows that the distribution of the two-sided spectrum $S_2^*(f_j) = \sigma_j^2/\kappa_j$ simply is $\text{Inv-}\chi^2\big(\nu_j, s_j^2/\kappa_j\big)$. There is a deterministic relationship between given variance parameters $\sigma_j^2$ and the implied autocorrelation function $\gamma(t)$ (see (13)), and so for random $\sigma_j^2$ the distribution of the corresponding $\gamma(t)$ may be numerically explored via Monte Carlo sampling from the distribution of the $\sigma_j^2$. The expectation and variance of $\gamma(t)$ may also be derived analytically: the expected autocovariance (13), with respect to the distribution of the $\sigma_j^2$, is

$$
\text{E}[\gamma(t)] \;=\; \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \left(\text{E}\big[\sigma_j^2\big] \frac{\kappa_j}{2} \, \cos(2\pi f_j t)\right), \tag{24}
$$

which is finite as long as $\text{E}\big[\sigma_j^2\big]$ is finite for all $j$. Similarly, the variance of the autocorrelation is

$$
\text{Var}(\gamma(t)) \;=\; \frac{1}{N^2\Delta_t^2} \sum_{j=0}^{\lfloor N/2 \rfloor} \left(\text{Var}\big(\sigma_j^2\big) \frac{\kappa_j^2}{4} \, \cos(2\pi f_j t)^2\right), \tag{25}
$$

which again is finite as long as $\text{Var}\big(\sigma_j^2\big)$ is finite for all $j$.

*2.4.4. The marginal likelihood.* The use of the conjugate $\text{Inv-}\chi^2$ prior (with $\nu_j > 0$ degrees of freedom) for the variance parameters $\sigma_j^2$ allows us to integrate the unknown noise spectrum

out of the likelihood expression (17). Then, the *marginal* likelihood function is

$$
p\big(x_1, \ldots, x_N \,\big|\, v_0, \ldots, v_{\lfloor N/2 \rfloor}, s_0^2, \ldots, s_{\lfloor N/2 \rfloor}^2\big)
$$

$$
= \prod_{j=0}^{\lfloor N/2 \rfloor} \int_0^\infty p\big(a_j, b_j \big| \sigma_j^2\big) \, p\big(\sigma_j^2 \big| v_j, s_j^2\big) \, \mathrm{d}\sigma_j^2 \tag{26}
$$

$$
= \prod_{j=0}^{\lfloor N/2 \rfloor} \frac{(2\pi)^{-\frac{\kappa_j}{2}} \left(\frac{v_j s_j^2}{2}\right)^{\frac{v_j}{2}}}{\left(\frac{v_j s_j^2 + \left(a_j^2 + b_j^2\right)}{2}\right)^{\frac{v_j + \kappa_j}{2}}} \frac{\Gamma\left(\frac{v_j + \kappa_j}{2}\right)}{\Gamma\left(\frac{v_j}{2}\right)} \tag{27}
$$

$$
\propto \prod_{j=0}^{\lfloor N/2 \rfloor} \left(1 + \frac{\kappa_j^2 \frac{\Delta_t}{N} |\tilde{x}_j|^2}{v_j s_j^2}\right)^{-\frac{v_j + \kappa_j}{2}} \tag{28}
$$

$$
= \exp\left(-\sum_{j=0}^{\lfloor N/2 \rfloor} \frac{v_j + \kappa_j}{2} \log\left(1 + \frac{\kappa_j^2 \frac{\Delta_t}{N} |\tilde{x}_j|^2}{v_j s_j^2}\right)\right), \tag{29}
$$

which constitutes a product of Student-*t* densities with $(v_j + \kappa_j - 1)$ degrees of freedom [34]. Using the uninformative, improper Jeffreys prior (with $v_j = 0$ degrees of freedom and density $p(\sigma_j^2) \propto \frac{1}{\sigma_j^2}$) yields the marginal likelihood

$$
p(x_1, \ldots, x_N) \propto \prod_{j=0}^{\lfloor N/2 \rfloor} \left(\kappa_j^2 \frac{\Delta_t}{N} |\tilde{x}_j|^2\right)^{-\frac{\kappa_j}{2}}. \tag{30}
$$

One may then also get a mixture of the above expressions (28) and (30) in the case of a prior setting of partly zero and non-zero prior degrees of freedom.

Note that the corresponding analogue likelihood expression in the case of an *a priori known* spectrum (as e.g. in [11, 29, 30], see also (17)) was

$$
p\big(x_1, \ldots, x_N \big| \sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2\big) \propto \exp\left(-\sum_{j=0}^{\lfloor N/2 \rfloor} \frac{\kappa_j^2 \frac{\Delta_t}{N} |\tilde{x}_j|^2}{2\,\sigma_j^2}\right), \tag{31}
$$

so that the (normal) sum-of-squares expression (31) in the case of a known spectrum generalizes to the Student-*t* expression (29) once one takes uncertainty about the spectrum's scale into consideration. The normal model in turn is the limiting case for increasing degrees of freedom.

## 2.5. Robust inference via the Student-t model

The Student-*t* marginal likelihood expression (28) suggests that considering noise as normal but with unknown spectrum is technically equivalent to viewing the noise itself as being *t*-distributed. In that way, the Student-*t* model may also be useful as an alternative for robust modelling, as the wider family of *t*-distributions includes the normal and Cauchy distributions as special or limiting cases. In other contexts the *t*-distribution is commonly used as a generalization of the normal model in order to accommodate heavier-tailed errors (see e.g. [37–40]). In contrast to the above derivation of the *t*-distributed noise based on prior/posterior distributions, the noise may also be considered as a *scale mixture* of normal distributions, that is, normal noise with a randomly varying scale (variance) [38]. Both relations may actually be used to motivate the use of the Student-*t* distribution for noise

exhibiting certain nonstationarities (e.g. a noise spectrum that is slightly fluctuating over time), be it to model the 'variability' or the 'uncertainty' in the spectrum, both of which are mathematically the same here (except that uncertainty is reduced through learning, while the random variation is not).

The use of heavier-tailed noise models already has repeatedly been advocated in the gravitational wave detection context, in order to gain robustness against outliers and similar deviations from the commonly assumed normality. Creighton [41] for example suggests the use of a mixture of Gaussian noise and a (uniform) burst component in order to account for wide-band noise burst events. Allen *et al* [42] also illustrate the use of a two-component Gaussian mixture in order to reduce the influence of outliers in the data, but more generally they advocate an approach also known as *M-estimation*, namely the explicit downweighting or ignorance of outlier observations falling far into the tail of the noise distribution [43, 44]. Application of the above Student-*t* model may also be considered as a special case of robust M-estimation, based on a clearly interpretable noise model and associated parameters [39, 40].

In the case of an *a priori* known power spectrum, maximization of the normal likelihood (31) is the basis for the *matched filtering* approach commonly used in signal detection problems, when looking for signals of parameterized shape in noise [12, 13, 16]. In Gaussian noise, the likelihood maximization then is equivalent to a least-squares approach. A filter based on the Student-*t* likelihood (28) may therefore be useful for cases of an unknown noise spectrum or non-Gaussian noise.

Note that since the likelihood (28) implies *independent* (as opposed to merely uncorrelated) errors, the likelihood maximization here is different from least-squares estimation [38, 45, 46], and the likelihood might in fact exhibit multiple modes [47]. The (independent) Student-*t* model not only leads to a less drastic fall-off of the likelihood for extreme values, and hence reduced leverage of outliers, but it also implies non-spherical density contours for the joint distribution of the noise, effectively allowing for a fraction of scrambled *individual* noise residuals. A similar effect was pointed out by Creighton [41] when implementing a robust Gaussian/uniform mixture noise model: the approach would allow for excess noise in individual interferometers, so that a noise burst would essentially be automatically 'vetoed' if it is only measured in one of several interferometers. Similarly, the non-spherical density contours of the Student-*t* model make it robust against odd data values at individual frequencies.

## 2.6. Defining the prior distribution's parameters

Depending on the particular application and context in mind, there may be different ways to sensibly specifying prior parameters. Firstly, there are the supposedly uninformative priors; the *Jeffreys prior* [35] with $\nu_j = 0$ degrees of freedom, and priors that are uniform on $\sigma_j$ or on $\sigma_j^2$, were already mentioned above (see section 2.4.2). Care needs to be taken here though, as these priors (with $\nu_j \leqslant 0$) are improper distributions. These may lead to improper marginal likelihoods, as in the case of (30), which does not correspond to a normalizable probability distribution for the noise. The resulting posterior distribution of signal or noise parameters then *may* or *may not* be a proper probability distribution [34].

If the prior information on the spectral parameters is in the form of *measurements* (or *samples*) of the spectrum (of same size and resolution), then this may be expressed in terms of an equivalent sample size and corresponding prior degrees of freedom, as suggested in section 2.4.3. Choosing e.g. the $\nu_j$ to be twice the number of measurements of the spectrum taken (i.e. equal to the number of observed Fourier coefficients) would then be equivalent to

*initially* assuming an uninformative Jeffreys prior and then using the posterior based on the measurements as the prior for the actual (signal) analysis. This will of course only make sense if one assumes the spectrum unknown but stationary.

When using the Student-*t* model as a robust model accommodating for heavy-tailed noise, i.e. when the noise itself is assumed to actually follow a *t*-distribution that one can sample from, one can use sample estimates for the Student-*t* parameters. Moment estimators for $s_j^2$ and $\nu_j$ (based on sample variance and kurtosis) are given e.g. in [48, 49].

When specifying prior parameters that are supposed to reflect information and/or variability, it may be helpful to consider the implied moments or quantiles, for individual frequency bins or frequency bands. The expressions for the prior's moments in (20) may be inverted to

$$\nu_j \;=\; 4 + 2 \frac{\mathrm{E}\big[\sigma_j^2\big]^2}{\mathrm{Var}\big(\sigma_j^2\big)} \quad \text{and} \quad s_j^2 \;=\; \frac{\nu_j - 2}{\nu_j}\,\mathrm{E}\big[\sigma_j^2\big], \tag{32}$$

which allows one to specify the scale $s_j^2$ and degrees of freedom $\nu_j$ based on pre-defined prior expectation and variance of $\sigma_j^2$, respectively. Note that the degrees of freedom $\nu_j$ then are simply a function of the prior variation coefficient $\sqrt{\mathrm{Var}(\sigma_j^2)}/\mathrm{E}[\sigma_j^2]$. A specification of $s_j^2$ independent of $j$ means *a priori* white noise, and specifying individual $\nu_j$ for different $j$ indicates varying prior certainty across the spectrum. A sensible definition of the prior certainties for the individual spectrum parameters may be complicated by the fact that the exact meaning of this discrete set of parameters depends on the sample size *N*. In that case it may be helpful to instead consider the *integrated spectrum* (9) and its *a priori* properties. The (prior) moments of the integrated spectrum are given by

$$\mathrm{E}[\mathcal{I}_{[f_1,f_2]}] \;=\; \Delta_f \sum_{j=j_1}^{j_2} \frac{\kappa_j}{2}\,\mathrm{E}[\sigma_j^2] \;=\; \Delta_f \sum_{j=j_1}^{j_2} \frac{\kappa_j}{2} \frac{\nu_j}{\nu_j - 2} s_j^2 \tag{33}$$

(if all $\nu_j > 2$) and

$$\mathrm{Var}(\mathcal{I}_{[f_1,f_2]}) = \Delta_f^2 \sum_{j=j_1}^{j_2} \frac{\kappa_j^2}{4} \frac{2\nu_j^2}{(\nu_j - 2)^2(\nu_j - 4)} s_j^4 \tag{34}$$

(if all $\nu_j > 4$). The (prior) variation coefficient for the power within any frequency range then is

$$\frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_1,f_2]})}}{\mathrm{E}[\mathcal{I}_{[f_1,f_2]}]} \;=\; \frac{\sqrt{\sum_{j=j_1}^{j_2} \frac{\kappa_j^2}{4} \frac{2\nu_j^2}{(\nu_j-2)^2(\nu_j-4)} s_j^4}}{\sum_{j=j_1}^{j_2} \frac{\kappa_j}{2} \frac{\nu_j}{\nu_j-2} s_j^2} \tag{35}$$

which simplifies in case all degrees-of-freedom parameters are taken to be equal ($\nu_j \equiv \nu$):

$$\frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_1,f_2]})}}{\mathrm{E}[\mathcal{I}_{[f_1,f_2]}]} \;=\; \sqrt{\frac{2}{\nu - 4}} \frac{\sqrt{\sum_{j=j_1}^{j_2} \frac{\kappa_j^2}{4} s_j^4}}{\sum_{j=j_1}^{j_2} \frac{\kappa_j}{2} s_j^2} \tag{36}$$

and simplifies further in case all scale parameters are taken to be equal ($s_j^2 \equiv s^2$):

$$\frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_1,f_2]})}}{\mathrm{E}[\mathcal{I}_{[f_1,f_2]}]} \;=\; \sqrt{\frac{2}{\nu - 4}} \frac{\sqrt{\sum_{j=j_1}^{j_2} \frac{\kappa_j^2}{4}}}{\sum_{j=j_1}^{j_2} \frac{\kappa_j}{2}}. \tag{37}$$

In the general case of $(j_1 > 0,\ j_2 < \lfloor N/2 \rfloor)$ this is

$$\frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_1, f_2]})}}{\mathrm{E}[\mathcal{I}_{[f_1, f_2]}]} = \sqrt{\frac{2}{\nu - 4}} \frac{1}{\sqrt{j_2 - j_1 + 1}} \tag{38}$$

and similarly, in the case $(j_1 = 0,\ \text{and}\ j_2 = N/2,\ N\ \text{even})$ it is:

$$\frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_0, f_{N/2}]})}}{\mathrm{E}[\mathcal{I}_{[f_0, f_{N/2}]}]} = \sqrt{\frac{2}{\nu - 4}} \frac{\sqrt{\frac{N-1}{2}}}{\frac{N}{2}}. \tag{39}$$

This may be useful if one wants to specify piecewise constant prior settings with given constraints on the overall power *per frequency range*; this would then lead to the degrees-of-freedom settings

$$\nu = 4 + \frac{2}{j_2 - j_1 + 1} \left( \frac{\mathrm{E}[\mathcal{I}_{[f_1, f_2]}]}{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_1, f_2]})}} \right)^2 \tag{40}$$

$$\text{or} \quad \nu = 4 + 2 \frac{N-1}{N^2} \left( \frac{\mathrm{E}[\mathcal{I}_{[f_0, f_{N/2}]}]}{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_0, f_{N/2}]})}} \right)^2 \tag{41}$$

respectively, for the above two cases ((38) and (39)).

For example, if one wants the spectrum to be *a priori* white with some scale $s^2$, such that the marginal prior mean and variation coefficient of the integrated power $\mathcal{I}_{[f_0, f_{N/2}]} = \mathrm{Var}(X_i) = \gamma(0)$ are given by

$$\mathrm{E}[\mathcal{I}_{[f_0, f_{N/2}]}] = \varsigma^2 \quad \text{and} \quad \frac{\sqrt{\mathrm{Var}(\mathcal{I}_{[f_0, f_{N/2}]})}}{\mathrm{E}[\mathcal{I}_{[f_0, f_{N/2}]}]} = c, \tag{42}$$

then a setting of

$$\nu_j = 4 + \frac{N-1}{N^2} \frac{2}{c^2} \quad \text{and} \quad s_j^2 = 2 \Delta_t \frac{\nu - 2}{\nu} \varsigma^2 \tag{43}$$

independent of $j$ will yield an *a priori* white spectrum with constant expectation and variation in the overall variance $\mathcal{I}_{[f_0, f_{N/2}]}$ for any sample size $N$. Piecewise constant settings across different frequency bands may be implemented analogously. In case all $\nu_j > 4$, the integrated power, being a sum of random variables with finite mean and variance, will be asymptotically normally distributed.

Instead of using moments, the prior may also be specified in terms of quantiles, for example by first deciding on the degrees-of-freedom settings and then aiming the prior median at a certain value. The quantiles of an Inv-$\chi^2(\nu, s^2)$ distribution may be derived based on the quantiles of a $\chi^2$-distribution; the $p$-quantile is then given by

$$\nu s^2 / \chi^2_{\nu; 1-p} \tag{44}$$

where $\chi^2_{\nu; 1-p}$ is the $(1 - p)$-quantile of a $\chi^2$-distribution with $\nu$ degrees of freedom.

Finally, in the M-estimation context, the *t*-distribution's degrees-of-freedom parameter may also be set based on the shape of the corresponding *influence function* [43, 44].

## 3. Examples

### 3.1. Noise only

Consider a time series $n(t)$ of $N = 100$ data points sampled at times $t_i = \frac{i}{100}$. As a simple example of non-white noise, the data are generated from an autoregressive (AR) process

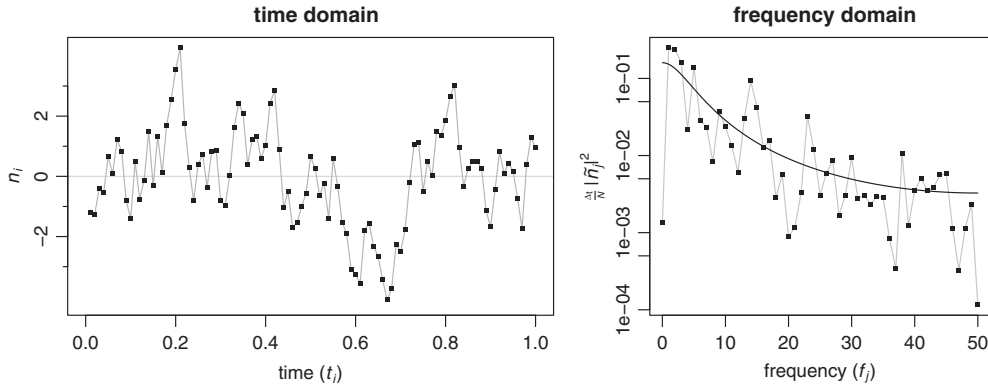$$n(t_i) = \tfrac{3}{4} n(t_{i-1}) + x(t_i), \tag{45}$$

**Figure 1.** The example data $n_i$ in the time domain, and its (empirical) power $\frac{\Delta t}{N}|\tilde{n}_j|^2$ in the Fourier domain. The solid line in the right panel shows the theoretical (two-sided) power spectral density in comparison.

where the innovations $x(t_i)$ are drawn independently from a uniform distribution across the interval $[-\sqrt{3}, \sqrt{3}]$, so that they have zero mean and unit variance. The overall variance of the process defined in this way is $\text{Var}(n(t_i)) = 2.29$; due to the positive correlation of subsequent samples it has a higher power at low frequencies and less at high frequencies. Figure 1 shows a noise sample generated using the above prescription, together with its theoretical power spectral density.

We will now apply the noise model introduced above and derive the posterior distribution of the noise parameters. Assuming one has a rough idea of the noise variance, we set the prior scale parameters $s_j^2$ so that the noise is *a priori* white with a prior expectation of $\text{E}[\text{Var}(n(t_i))] = 2.50$. We use $\nu_j = 3$ degrees of freedom, so that the noise parameters' (and with that, the overall power's) prior expectations are finite, while the variances are not. Prior scale and prior expectation then are $s_j^2 = 0.0166$ and $\text{E}[\sigma_j^2] = 0.05$. Figure 2 illustrates the resulting posterior distribution for all 51 noise parameters (23) in comparison to the case of using the uninformative (and improper) Jeffreys prior. The Jeffreys prior (with $\nu_j = 0$ degrees of freedom for each frequency bin) does not depend on the scale parameters $s_j^2$, and the resulting posterior with $\leqslant 2$ degrees of freedom at each frequency does not have finite expectation values. Note also that the posterior distributions corresponding to the first and last frequency bins (zero and Nyquist frequencies, $\sigma_0^2$ and $\sigma_{50}^2$) are wider than the others in both cases, as they have one less degree of freedom.

Figure 3 shows the posterior distributions of autocovariance and variance, which are functions of the individual spectrum parameters $\sigma_j^2$. The variance may either be considered as the zero-lag autocovariance $\gamma(0)$ (see (13)), or as the integrated power $\mathcal{I}_{[0, f_{N/2}]}$ across the whole frequency range (see (9)).

### 3.2. A signal with additive noise

In the following example, we will consider a time series where the primary interest is in a *signal* component, while the additive *noise* component will be modelled using the approach introduced above. As in the previous section, we will again consider $N = 100$ data points $y(t_1), \ldots, y(t_{100})$ that are modelled as

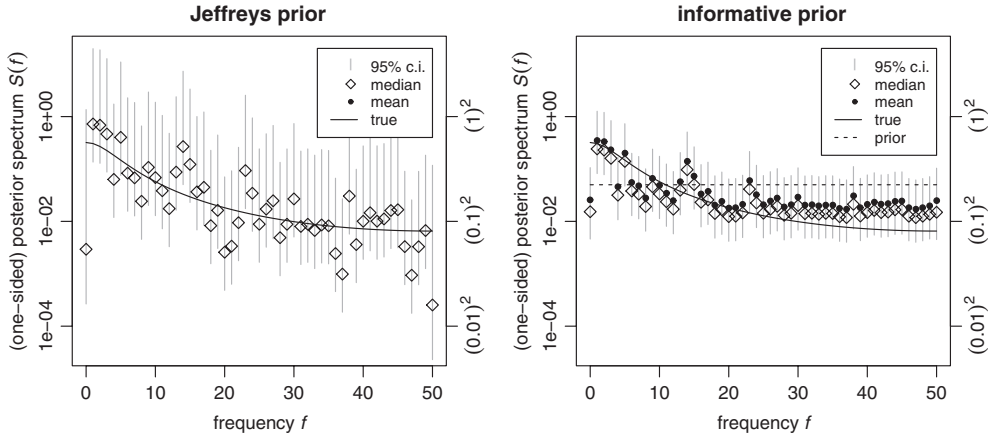$$y(t_i) = g_{f,\dot{f},a,\phi}(t_i) + n_{\tilde{\sigma}}(t_i), \tag{46}$$

**Figure 2.** Posterior distributions of the 51 spectrum parameters $\sigma_j^2$ based on the data shown in figure 1. The left plot shows the posterior corresponding to the uninformative (and improper) Jeffreys prior. Posterior expectations do not exist in this case. The right plot corresponds to assuming *a priori* white noise with $\nu_j = 3$ degrees of freedom for each frequency bin; the dashed line marks the prior expectation value.
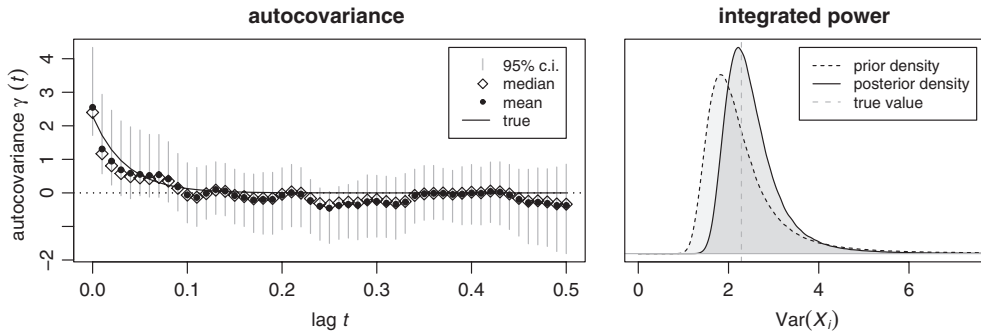


**Figure 3.** Posterior distributions of autocovariance and variance; the distributions shown here were derived via Monte Carlo integration.

where $n_{\vec{\sigma}}(t_i)$ is non-white noise of unknown spectrum, and $g_{f,\dot{f},a,\phi}(t)$ is a 'chirping' signal waveform of increasing frequency:

$$g_{f,\dot{f},a,\phi}(t) = a \sin(2\pi(f + \dot{f}t)t + \phi) \tag{47}$$

where $f$ and $\dot{f}$ are the *frequency* and *frequency derivative*, $a$ is the *amplitude*, and $\phi$ is the *phase*. The noise is again generated in the same way as in the previous example, by simulating an AR process. The signal's 'size' relative to the noise is given by the signal-to-noise ratio (SNR: $\varrho = \sqrt{4 \sum_j \frac{\frac{\Delta_t}{N}|\tilde{g}(f_j)|^2}{S_1(f_j)}}$), which here is 15.

We define the signal parameters' prior as uniform for phase, frequency and amplitude ($\phi \in [0, 2\pi]$, $f \in [1, 50]$ and $a \in [0, 10]$) and normal for the frequency derivative $\dot{f}$ (zero mean and standard deviation 5). The prior distribution for the noise parameters $\sigma_0^2, \ldots, \sigma_{50}^2$ is set exactly as in the previous section 3.2. For comparison, we analyse the data two more times, once assuming the true noise spectrum to be known, and once assuming the noise to be
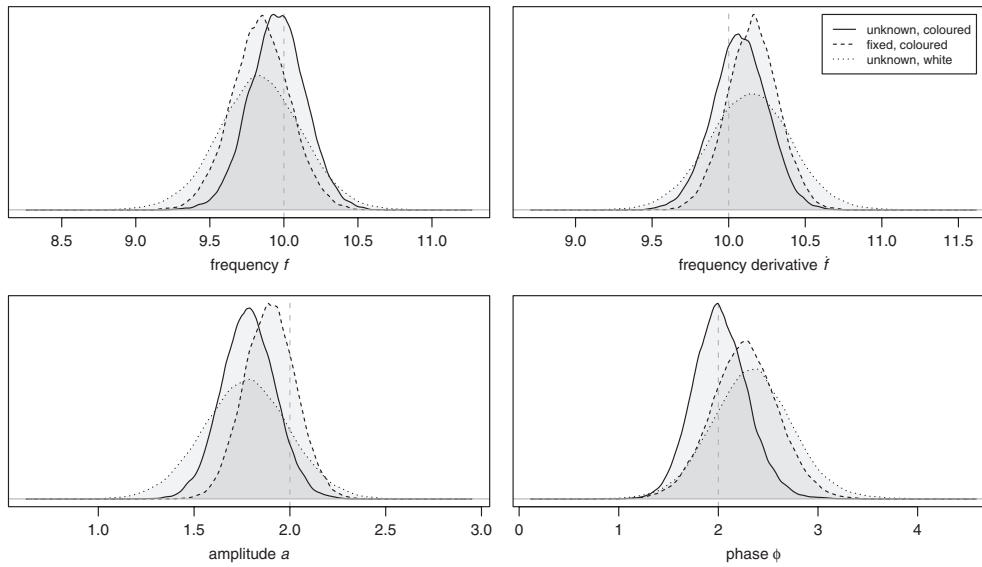
**Figure 4.** Marginal posterior densities of the four individual *signal* parameters. The three densities in each panel result from three different models applied for the noise term. The '*unknown, coloured*' case corresponds to the model introduced above; the '*fixed, coloured*' case assumes the true noise spectrum to be *a priori* known; and in the '*unknown white*' case the noise is modelled as white with unknown variance. The vertical dashed lines indicate the true parameter values.

white, but with an unknown variance. In the latter case, we again assume a (conjugate) Inv-$\chi^2$ prior with an expectation of 2.5 and 3 degrees of freedom for the variance parameter.

The posterior distributions of signal and noise parameters may now be derived via Monte Carlo integration; we implemented a Metropolis sampler to simulate draws from the joint posterior probability distribution of all parameters [34]. For the model including the coloured noise spectrum as unknown, the signal parameters $(f, \dot{f}, a, \phi)$ may be sampled based on the marginal likelihood expression (29), while samples of the 51 noise parameters $(\sigma_0^2, \ldots, \sigma_{50}^2)$ may then be sampled in an additional step via the conditional distribution of noise parameters for given signal parameters and the corresponding implied vector of noise residuals (23). Similarly, for the known spectrum model, sampling of the four signal parameters may be based on the likelihood expression (31), while sampling for the white noise model may be done using a Gibbs sampler alternately sampling from the conditional distributions of four signal parameters and the noise parameter [34].

Figure 4 shows the resulting marginal posterior distributions of the four signal parameters in comparison. One can see that application of the more flexible coloured noise model for the error term yields a more precise posterior distribution than the white noise model, and in fact the posterior is more similar to the case when the true noise spectrum is plugged into the model. Looking at the (marginal) posterior distributions of the noise parameters in figure 5, one can see that although the noise spectrum is only recovered with great uncertainty, this does not seem to harm the estimation of the signal parameters of actual interest. The overall variance is better recovered by the single-parameter white noise model, whereas the adaptability of the coloured noise model to the predominantly low-frequency noise (which *is* reflected in the posterior) seems to be the greater advantage. While the relative accuracy of the different
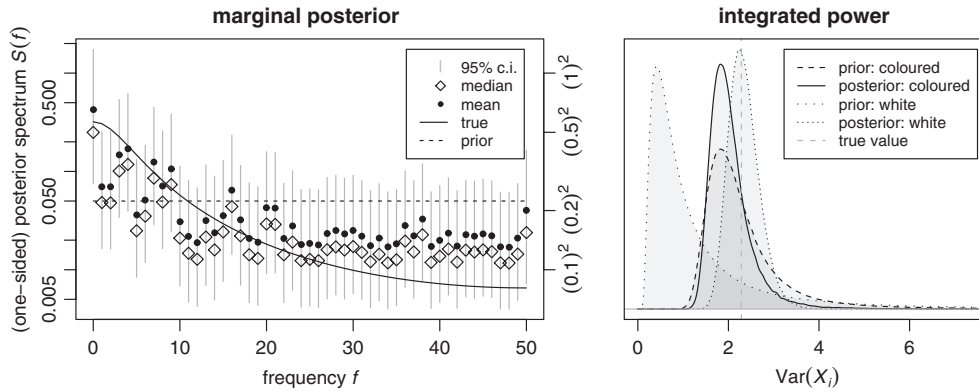
**Figure 5.** Marginal posterior distributions of the noise parameters. The left panel shows the posterior distribution for the 51 individual parameters of the coloured noise model. The right panel shows the prior and posterior distributions of the integrated power in comparison with the corresponding variance parameter in the white noise model.

methods is subject to a multitude of circumstances, one can see that despite its considerably greater complexity, the coloured noise model seems to perform competitively here.

## 4. Discussion

This work originated from the *Mock LISA Data Challenges (MLDC)* [50], a gravitational wave parameter estimation effort in the context of the planned *Laser Interferometer Space Antenna (LISA)*. Here, parameters of a signal were to be inferred, where the signal was buried in noise known to be non-white and interspersed with a host of individual emission lines [17]. So the problem was to model a non-white, non-continuous spectrum where the shape of the spectrum was only vaguely known in advance [18, 19]. Also, the spectrum itself was not of primary interest, but rather a nuisance parameter that still needed to be accounted for along with the actual signal. Application of the method described here solved the problem and allowed the implementation of a Markov chain Monte Carlo (MCMC) algorithm based on a straightforward generalization of the commonly utilized Gaussian noise assumption, where modelling the spectrum complicated the analysis only slightly.

Several approaches to the simultaneous estimation of signal and noise parameters have been proposed before. An obvious way to model non-white noise would be to assume a particular time series formulation that allows for some flexibility in the resulting spectrum, for example an AR model. Representations of this kind have been applied in various signal processing contexts, e.g. for detecting sinusoidal signals and estimating their parameters when these are buried in coloured noise [51, 52], or in the context of musical pitch estimation, where the signals to be modelled again are sinusoidal, but include higher harmonics and time-varying amplitudes [53]. If one is only interested in a narrow frequency band of the data, it may also be appropriate to model the noise spectrum as constant across the concerned range [54]. Regarding data treated in their Fourier-domain representation, there is a rich literature concerned with the *measurement* of spectral densities (see e.g. [26] and references therein). However, there the aim is usually to produce consistent and smooth estimates of the spectrum, and the approaches applied include e.g. averaging [32, 55], smoothing via splines [56] or Bayesian model fitting [57, 58].

The approach to modelling the noise spectrum introduced here is different in that the spectrum *per se* is not of interest, or only of interest as far as it enters into likelihood computations. Smoothness or interpolation therefore are not primarily aimed for. What in fact is of concern is properly accounting for an *a priori* uncertainty in the (discretized, convolved) spectrum, in the frequency resolution that is given by the numerically Fourier transformed data, which then consequently entails the necessity for updating our knowledge about the spectrum as data are being processed. The resulting approach generalizes the model underlying the commonly used *Whittle likelihood* [11, 29, 30], and is hence essentially based on the 'plain' periodogram of the noise time series, which for other purposes is commonly dismissed due to its unfavourable large-sample convergence behaviour. The model is very flexible as it is built upon a 'binned' spectrum estimate without introducing any extra assumptions on the shape of the underlying noise spectral density. Its generality and simplicity make it useful for modelling residual noise at very little computational cost. In the way it is defined, the model is very general; it is a generalization of the model that constitutes the basis for *matched filtering* (e.g. [12, 13]) and that is commonly applied in signal processing problems (e.g. [11]), which then in turn constitutes the special case of an *a priori* known spectrum. The additional feature of the approach introduced here is that it allows us to specify corresponding uncertainties in addition to the (prior) scale of the noise spectrum. Marginalization over the uncertainty in the noise spectrum yields a Student-*t* model for the Fourier-domain data as a natural generalization of the common normal model. Due to the straightforward interpretability of the model and its computationally convenient form, we expect it to be particularly useful for modelling an unknown noise spectrum that constitutes a nuisance parameter rather than being of interest in itself. In that sense, the approach is not primarily aimed at gaining information about an unknown spectrum, but rather at properly accounting for uncertainty, and avoiding bias from supposedly precise *a priori* knowledge of the spectrum.

Alternatively, the Student-*t* model may also be viewed as a generalized, robust model accommodating for heavier-tailed, non-Gaussian, or non-stationary noise. Related approaches are commonly used in many other applications in the context of *robust statistics* [37–40, 43, 44]. In fact, the use of similar methods for accommodating noise outliers has already been advocated in the context of gravitational wave signal processing [41, 42]. Along similar lines, Clark *et al* [59] implemented a (sine–Gaussian) noise-glitch component into the noise model. Principe and Pinto [60] suggested the use of a dictionary of glitch 'atoms' in order to account for burst-like noise events. Veitch and Vecchio [61] implemented a coherence test based on a model selection procedure that is able to discriminate noise glitches (that appear independently in data) from actual astrophysical events (that appear coherently in several data streams). Similarly, Littenberg and Cornish [62] extend their model to allow for excess noise that is isolated in time and frequency via a wavelet approach. Middleton [63] treated the general case of Gaussian noise with superimposed Poisson-distributed impulse noise bursts. We are planning to investigate the performance and sensitivity of a Student-*t* model for robust detection and parameter estimation in real interferometer noise.

We expect the approach to be especially useful as a model component properly accounting for non-white residual noise, at little computational cost and without introducing overly restrictive assumptions about the noise. The interpretability of the Student-*t* model in the context of robust modelling and M-estimation also makes it straightforwardly useful for robust inference. In that way, it may particularly be useful in signal processing contexts [13, 64], but it should be applicable in any case where a model for (residual) noise is needed [23, 65]. The fact that the model represents noise properties in terms of its power spectral density and is specified through physically meaningful parameters may make it particularly appealing in physical or engineering applications, where modelling is commonly based on Fourier-domain

descriptions, and a time-domain formulation might be hard to incorporate or motivate. The advantage from application of the more flexible Student-*t* model will very much depend on the particular inference problem at hand. The common normal model will obviously be optimal when its assumptions are met, and the degree to which one will outperform the other will depend on the particular departure(s) from Gaussianity, or on the imperfect prior knowledge of the spectrum, even if the data are perfectly Gaussian. While the range of possible deviations from the standard Gaussian model is infinite, some insight into the behaviour of a Student-*t* model may be gained from the *robust statistics* literature; the properties of such an approach for location estimation have been investigated via Monte Carlo studies [39, 66] as well as theoretical considerations of key figures like *relative efficiency* or *breakdown point* [39, 43, 44]. A case study in the regression context may be found in [38].

The above basic model may in future be extended by introducing smoothness constraints on the spectrum. This might for example be approached by considering correlations between neighbouring spectral bins, or by assuming a piecewise constant spectrum, similar to what was done in [54]; the latter approach would in fact be a compromise between two models used in the example discussed above (section 3.2), namely a flat spectrum and individually modelled frequency bins. Another interesting extension would be the incorporation of cross-spectra [23, 67]. Some of the methods described here have been coded as an R software extension package and are available for free download [68].

## Acknowledgments

## Appendix

### A.1. Discrete Fourier transform (DFT)

The Fourier transform convention used in this paper is specified below; it is defined for a real-valued function $h$ of time $t$, sampled at $N$ discrete time points, at a sampling rate of $\frac{1}{\Delta_t}$, and it maps from

$$\{h(t) \in \mathbb{R} : \ t = 0, \Delta_t, 2\Delta_t, \ldots, (N-1)\Delta_t\} \tag{A.1}$$

to a function of frequency $f$

$$\{\tilde{h}(f) \in \mathbb{C} : \ f = 0, \Delta_f, 2\Delta_f, \ldots, (N-1)\Delta_f\}, \tag{A.2}$$

where $\Delta_f = \frac{1}{N\Delta_t}$ and

$$\tilde{h}(f) \ = \ \sum_{j=0}^{N-1} h(j\Delta_t) \exp(-2\pi \mathrm{i} j \Delta_t f). \tag{A.3}$$

The inverse DFT then is given by

$$h(t) \ = \ \frac{1}{N} \sum_{j=0}^{N-1} \tilde{h}(j\Delta_f) \exp(2\pi \mathrm{i} j \Delta_f t) \tag{A.4}$$

[22].

*A.2. Relationship between DFT and time series model*

Let

$$\alpha_j = \mathrm{Re}\big(\tilde{h}(f_j)\big) \quad \text{and} \quad \beta_j = \mathrm{Im}\big(\tilde{h}(f_j)\big), \qquad (A.5)$$

i.e. $\tilde{h}(f_j) = \alpha_j + \beta_j \mathrm{i}$. For simplicity, in the following, $N$ is assumed to be even; for uneven $N$, the derivation is similar. The inverse DFT was defined as (A.4):

$$h(t) = \frac{1}{N} \sum_{j=0}^{N-1} \tilde{h}(f_j) \exp(2\pi \mathrm{i} f_j t) \qquad (A.6)$$

$$= \frac{1}{N} \sum_{j=1}^{\frac{N}{2}-1} \big[\big(2\alpha_j \cos(2\pi f_j t) + 2(-\beta_j)\sin(2\pi f_j t)\big)\big]$$

$$+ \frac{1}{N}\alpha_0 + \frac{1}{N}\alpha_{N/2}\cos(2\pi \mathrm{i} f_{N/2} t) \qquad (A.7)$$

where $t \in \{0, \Delta_t, 2\Delta_t, \ldots, (N-1)\Delta_t\}$, and $f_j = j\Delta_f = \frac{j}{N\Delta_t}$ are the Fourier frequencies. So, comparing to (1), one can see that the realizations of $a_0, \ldots, a_{\lfloor N/2 \rfloor}$ and $b_0, \ldots, b_{\lfloor N/2 \rfloor}$ are derived from a given time series by Fourier transforming and then setting

$$a_j = \kappa_j \sqrt{\frac{\Delta_t}{N}}\,\alpha_j \quad \text{and} \quad b_j = -\kappa_j \sqrt{\frac{\Delta_t}{N}}\,\beta_j \qquad (A.8)$$

for $j = 0, \ldots, \lfloor N/2 \rfloor$, which especially implies that

$$a_j^2 + b_j^2 = \kappa_j^2 \frac{\Delta_t}{N}\big(\alpha_j^2 + \beta_j^2\big) = \kappa_j^2 \frac{\Delta_t}{N}\big|\tilde{h}(f_j)\big|^2. \qquad (A.9)$$

## References

[1] Thorne K S 1987 *Gravitational radiation* ed S W Hawking and W Israel (Cambridge: Cambridge University Press) pp 330–58 300 Years of Gravitation chapter 9
[2] Schutz B F 1999 Gravitational wave astronomy *Class. Quantum Grav.* **16** A131–56
[3] Sigg D *et al* 2008 Status of the LIGO detectors *Class. Quantum Grav.* **25** 114041
[4] Grothe H *et al* 2010 The GEO 600 status *Class. Quantum Grav.* **27** 084003
[5] Accadia T *et al* 2010 Commissioning status of the Virgo interferometer *Class. Quantum Grav.* **27** 084002
[6] Arai K *et al* 2009 Status of Japanese gravitational wave detectors *Class. Quantum Grav.* **26** 204020
[7] Harry G *et al* 2010 Advanced LIGO: the next generation of gravitational wave detectors *Class. Quantum Grav.* **27** 084006
[8] Kuroda K *et al* 2010 Status of LCGT *Class. Quantum Grav.* **27** 084004
[9] Danzmann K and Rüdiger A 2003 LISA technology—concept, status, prospects *Class. Quantum Grav.* **20** S1–9
[10] Punturo M *et al* 2010 The third generation of gravitational wave observatories and their science reach *Class. Quantum Grav.* **27** 084007
[11] Finn L S 1992 Detection, measurement, and gravitational radiation *Phys. Rev.* D **46** 5236–49
[12] Turin G L 1960 An introduction to matched filters *IRE Trans. Inf. Theory* **6** 311–29
[13] Wainstein L A and Zubakov V D 1962 *Extraction of Signals from Noise* (Englewood Cliffs, NJ: Prentice-Hall)
[14] Röver C, Meyer R and Christensen N 2007 Coherent Bayesian inference on compact binary inspirals using a network of interferometric gravitational wave detectors *Phys. Rev.* D **75** 062004
[15] Searle A C, Sutton P J and Tinto M 2009 Bayesian detection of unmodeled bursts of gravitational waves *Class. Quantum Grav.* **26** 155017
[16] Jaranowski P and Królak A 2005 Gravitational-wave data analysis. Formalism and sample applications: the Gaussian case *Living Rev. Rel.* **8** 3 (http://www.livingreviews.org/lrr-2005-3)
[17] Barack L and Cutler C 2004 Confusion noise from LISA capture sources *Phys. Rev.* D **70** 122002
[18] Röver C 2007 Bayesian inference on astrophysical binary inspirals based on gravitational-wave measurements *PhD Thesis* The University of Auckland (http://hdl.handle.net/2292/2356)

[19] Röver C *et al* 2007 Inference on inspiral signals using LISA MLDC data *Class. Quantum Grav.* **24** S521–7
[20] Jenkins G M 1961 General considerations in the analysis of spectra *Technometrics* **3** 167–90
[21] Jaynes E T 2003 *Probability Theory: The Logic of Science* (Cambridge: Cambridge University Press)
[22] Gregory P C 2005 *Bayesian Logical Data Analysis for the Physical Sciences* (Cambridge: Cambridge University Press)
[23] Jenkins G M and Watts D G 1968 *Spectral Analysis and its Applications* (San Francisco: Holden Day)
[24] Bretthorst G L 1999 The near-irrelevance of sampling frequency distributions ed W v d Linden *et al* (Dordrecht: Kluwer) pp 21–46 Maximum Entropy and Bayesian Methods
[25] Fisher R A 1929 Tests of significance in harmonic analysis *Proc. R. Soc.* A **125** 54–9
[26] Brillinger D R 1981 *Time Series: Data Analysis and Theory* (New York: McGraw-Hill)
[27] Champeney D C 1987 *A Handbook of Fourier Theorems* (Cambridge: Cambridge University Press)
[28] Kawata T 1966 On the Fourier series of a stationary stochastic process *Z. Wahrscheinlichkeitstheor. Verwandte Geb.* **6** 224–45
[29] Whittle P 1957 Curve and periodogram smoothing *J. R. Stat. Soc.* B **19** 38–63
[30] Choudhuri N, Ghosal S and Roy A 2004 Contiguity of the Whittle measure for a Gaussian time series *Biometrika* **91** 211–8
[31] Taqqu M S 1988 Weak stationarity ed S Kotz and N L Johnson *Encyclopedia of Statistical Sciences* (New York: Wiley)
[32] Harris F J 1978 On the use of windows for harmonic analysis with the discrete Fourier transform *Proc. IEEE* **66** 51–83
[33] Mood A M, Graybill F A and Boes D C 1974 *Introduction to the Theory of Statistics* 3rd edn (New York: McGraw-Hill)
[34] Gelman A, Carlin J B, Stern H and Rubin D B 1997 *Bayesian Data Analysis* (Boca Raton: Chapman and Hall/CRC)
[35] Jeffreys H 1946 An invariant form for the prior probability in estimation problems *Proc. R. Soc.* A **186** 453–61
[36] Bolstad W M 2007 *Introduction to Bayesian Statistics* 2nd edn (New York: Wiley) pp 297–316 chapter 15
[37] Lange K L, Little R J A and Taylor J M G 1989 Robust statistical modeling using the *t* distribution *J. Am. Stat. Assoc.* **84** 881–96
[38] Geweke J 1993 Bayesian treatment of the independent Student-*t* linear model *J. Appl. Econometrics* **8** S19–40
[39] Divgi D R 1990 Robust estimation using Student's *t* distribution *CNA Research Memorandum* CRM 90-217 Center for Naval Analyses, Alexandria, VA, USA
[40] McDonald J B and Newey W K 1988 Partially adaptive estimation of regression models via the generalized *t* distribution *Econometric Theory* **4** 428–57
[41] Creighton J D 1999 Data analysis strategies for the detection of gravitational waves in non-Gaussian noise *Phys. Rev.* D **60** 021101
[42] Allen B, Creighton J D E, Flanagan É É and Romano J D 2002 Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise: frequentist analyses *Phys. Rev.* D **65** 122002
[43] Hampel F R, Ronchetti E M, Rousseeuw P J and Stahel W A 1986 *Robust Statistics: The Approach Based on Influence Functions* (New York: Wiley)
[44] Huber P J and Ronchetti E M 2009 *Robust Statistics* 2nd edn (New York: Wiley)
[45] Kelejian H H and Prucha I R 1985 Independent or uncorrelated disturbances in linear regression: an illustration of the difference *Economics Lett.* **19** 35–8
[46] Breusch T S, Robertson J C and Welsh A H 2001 The emperor's new clothes: a critique of the multivariate *t* regression model *Statistica Neerlandica* **51** 269–86
[47] Mäkeläinen T, Schmidt K and Styan G P H 1981 On the existence and uniqueness of the maximum likelihood estimate of a vector-valued parameter in fixed-size samples *Ann. Stat.* **9** 758–67
[48] Sutradhar B C and Ali M M 1986 Estimation of the parameters of a regression model with a multivariate *t* error variable *Commun. Stat. - Theory Methods* **15** 429–50
[49] Singh R S 1988 Estimation of error variance in linear regression models with errors having multivariate Student-*t* distribution with unknown degrees of freedom *Economics Lett.* **27** 47–53
[50] Arnaud K A *et al* 2006 An overview of the Mock LISA data challenges *Laser Interferometer Space Antenna: 6th International LISA Symposium AIP Conf. Proc.* vol 873 ed S M Merkowitz and J C Livas pp 619–24
[51] Chatterjee C, Kashyap R L and Boray G 1987 Estimation of close sinusoids in colored noise and model discrimination *IEEE Trans. Acoust. Speech Signal Process.* **35** 328–37
[52] Cho C-M and Djurić P M 1995 Bayesian detection and estimation of cisoids in colored noise *IEEE Trans. Signal Process.* **43** 2943–52

[53] Godsill S and Davy M 2002 Bayesian harmonic models for musical pitch estimation and analysis *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* (Orlando, FL, USA) vol 2, pp 1769–72

[54] Dupuis R J and Woan G 2005 Bayesian estimation of pulsar parameters from gravitational wave data *Phys. Rev. D* **72** 102002

[55] Welch P D 1967 The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms *IEEE Trans. Audio Electroacoust.* **15** 70–3

[56] Wahba G 1980 Automatic smoothing of the log periodogram *J. Am. Stat. Assoc.* **75** 122–32

[57] Gangopadhyay A K, Mallik B K and Denison D G T 1999 Estimation of spectral density of a stationary time series via an asymptotic representation of the periodogram *J. Stat. Plan. Inference* **75** 281–90

[58] Choudhuri N, Ghosal S and Roy A 2004 Bayesian estimation of the spectral density of a time series *J. Am. Stat. Assoc.* **99** 1050–9

[59] Clark J, Heng I S, Pitkin M and Woan G 2007 Evidence-based search method for gravitational waves from neutron star ring-downs *Phys. Rev. D* **76** 043003

[60] Principe M and Pinto I M 2009 Locally optimum network detection of unmodelled gravitational wave bursts in an impulsive noise background *Class. Quantum Grav.* **26** 045003

[61] Veitch J and Vecchio A 2010 Bayesian coherent analysis of in-spiral gravitational wave signals with a detector network *Phys. Rev. D* **81** 062003

[62] Littenberg T and Cornish N 2010 Separating gravitational wave signals from instrument artifacts arXiv:1008.1577

[63] Middleton D 1951 On the theory of random noise: phenomenological models *J. Appl. Phys.* **22** 1143–1163

[64] McDonough R N and Whalen A D 1995 *Detection of Signals in Noise* 2nd edn (New York: Academic)

[65] Bloomfield P 2000 *Fourier Analysis of Time Series: An Introduction* 2nd edn (New York: Wiley)

[66] Andrews D F *et al* 1972 *Robust Estimates of Location: Survey and Advances* (Princeton, NJ,: Princeton University Press)

[67] Nofrarias M, Röver C, Hewitson M, Monsky A, Heinzel G, Danzmann K, Ferraioli L, Hueller M and Vitale S 2010 Bayesian parameter estimation in the second LISA Pathfinder Mock Data Challenge arXiv:1008.5280 [gr-qc]

[68] Röver C 2008 bspec: Bayesian spectral inference, R package. URL: http://cran.r-project.org/package=bspec