MODELLING COLOURED NOISE

By Christian Röver, Renate Meyer and Nelson Christensen

Max-Planck-Institut für Gravitationsphysik, The University of Auckland and Carleton College

This paper introduces a novel approach to modelling non-white residual noise in discrete time series. We present a Markov chain Monte Carlo (MCMC) algorithm for combined posterior inference on signal and noise parameters. By choosing a conjugate prior distribution for the noise parameters, the additional Gibbs sampling steps have a particularly simple form and are easy to implement as well as fast to run. Furthermore, the sampling-based approach allows for easy inference on the autocovariance function. The model is illustrated using the well-known sunspot dataset as well as a simulated dataset of a chirp signal embedded in non-Normal, coloured noise where the spectrum is regarded as a nuisance parameter.

1. Introduction. In many statistical applications one is faced with the problem of inferring 'model' or 'signal' parameters along with the properties of noise contaminating the data. A very simple case would be a regression problem where the noise is assumed to follow a Normal distribution with known variance. Likelihood computations here reduce to computing residual sums-of-squares that depend on the regression model parameters via the implied set of noise residuals. If, on the other hand, the noise is assumed to be Normal but non-white with a known power spectral density, an approximate likelihood can be computed based on the Fourier-transformed set of residuals and the power spectral density of the noise (Finn, 1992). The approximation is due to the fact that the Fourier-transformed data are discrete, while the matched power spectrum is continuous. In this case, the likelihood is a weighted sum-of-squares, summed over frequencies and weighted by the power spectral density. If the noise is known to be white Normal noise but with unknown variance, the variance enters the likelihood as an additional parameter to be inferred from the data. If Bayesian posterior computation is performed via MCMC sampling, inference on the variance parameter can easily be implemented by an additional Gibbs sampling step. In this paper we present a generalisation of the above approaches, addressing the case where the noise is coloured, but the noise spectrum is not a priori known.

One way to model non-white noise would be to assume a particular time series form that allows for some flexibility in the resulting spectrum. A popular approach along these lines is the use of autoregressive (AR) models for the coloured noise component. Representations of this kind have been applied in various signal processing contexts, e.g. for detecting sinusoidal signals and estimating their parameters when they are buried in coloured noise (Chatterjee et al., 1987; Cho and Djurić, 1995), 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 (Godsill and Davy, 2002). Similarly, in the following a parametric model is set up, but in a flexible way, such that a spectrum of arbitrary shape may be characterized, while invoking the maximum entropy principle ensures that the model is based on a minimal set of assumptions. Starting with a discrete Fourier transformation of the discrete time series data, our approach aims to model the discrete analogue of the power spectral density, and with that circumventing the approximation taken e.g. by Finn (1992). The choice of a conjugate prior allows for flexible prior specification as well as very efficient implementation.

This work originated from an astrophysical application, where parameters of a signal were to be inferred, and the signal was buried in noise known to be non-white and interspersed with a host of individual emission lines. So the problem was to model a non-white, non-continuous spectrum where the shape of the spectrum was only vaguely known in advance (Röver et al., 2007; Röver, 2007). Also, the spectrum itself was not of primary interest, but rather a nuisance parameter that still needed to be modelled along with the actual signal. Application of the method described here solved the problem and allowed the implementation of an MCMC algorithm where modelling the spectrum only complicated the algorithm by an additional Gibbs step.

The organization of the paper is as follows. The time series setup is presented in Section 2. In the subsequent Section 3, the set-up of the Bayesian model is described, including the prior distributions, likelihood, posterior distribution and some implications. Section 4 illustrates the modelling approach with three example applications. Section 5 closes with a discussion, and conclusions. An appendix on formulae for the Discrete Fourier Transform, autocovariance function and tempering is attached.

2. The time series model. Consider a time series $x_1, \ldots, x_N$ of N real-valued observations sampled at constant time intervals Δ_t, so that each observation x_i corresponds to time $t_i = i\Delta_t$. This set of N observations can

equivalently be expressed as sinusoids of the Fourier frequencies:

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

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 Fourier transform (and appropriate scaling; see appendix 6.2). 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 the time and in the frequency domain. The definition in (1) can equivalently be expressed as:

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

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

For each Fourier frequency f_j, let $\kappa(j)$ define the indicator function for coefficient b_j not being zero by definition, i.e.:

$$(3) \qquad \kappa(j) = \begin{cases} 0 & \text{if } (j = 0) \text{ or } (N \text{ is even and } j = N/2) \\ 1 & \text{otherwise} \end{cases}$$

so that $\sum_{j=0}^{\lfloor N/2 \rfloor}(1 + \kappa(j)) = N$.

For a *given* time series (either in terms of x_i or a_j and b_j), the spectral power is a function of the Fourier frequencies. The *one-sided* spectrum is given by

$$(4) \qquad p_1(f_j) = \frac{a_j^2 + b_j^2}{1 + \kappa(j)} \qquad \text{for } j = 0, \ldots, \lfloor N/2 \rfloor,$$

and the *two-sided* spectrum is given by

$$(5) \qquad p_2(f_j) = \frac{a_j^2 + b_j^2}{(1 + \kappa(j))^2} = \frac{p_1(f_j)}{1 + \kappa(j)}$$

$$(6) \qquad = \frac{\Delta_t}{N} |\tilde{x}(f_j)|^2, \qquad \text{for } j = 0, \ldots, \lfloor N/2 \rfloor$$

where $\tilde{x}$ denotes the discretely Fourier-transformed time series x, as defined in appendix 6.2. The set of $p_1(f_j)$ is also known as the *periodogram* of the time series.

Now consider the case where the observations x_i and consequently the a_j and b_j correspond to random variables X_i, A_j and B_j, respectively. This may mean that these are realisations of a random process, where the random variables' probability distributions describe the *randomness* in the observations, or that they are fixed, but unknown, where the probability distributions describe a *state of information*, or it may also be a mélange of both (Jaynes, 2003).

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

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

For the probabilistic time series, the spectral power consequently is a random variable ($P_1(f_j)$ or $P_2(f_j)$) and the *(power) spectral density* is defined as the expectation of the spectral power (Blackman and Tukey, 1958). If the mean is zero as in (7), then the one-sided and two-sided power spectral densities $S_1(f_j)$ and $S_2(f_j)$ are determined by

$$S_1(f_j) \;=\; (1+\kappa(j))S_2(f_j) \;=\; \mathrm{E}[P_1(f_j)] \;=\; \mathrm{E}\!\left[\tfrac{A_j^2+B_j^2}{1+\kappa(j)}\right] \;\overset{(7)}{=}\; \tfrac{\mathrm{Var}(A_j)+\mathrm{Var}(B_j)}{1+\kappa(j)}.$$
(8)

Note that strictly speaking P_1 and P_2, as defined above are not exactly the power spectral density, but a discrete analogue; one might rather want to refer to these as the *discrete* power spectrum. On the other hand, it is exactly the quantity required in the denominator of the likelihood (see following section), where typically the actual *continuous* PSD (or estimates thereof) are used as a substitute, as e.g. by Finn (1992). So in some sense the discrete spectrum might actually be regarded as the more relevant figure. Describing the spectrum as defined above means modelling the discrete spectrum of the observed (discrete and finite) time series, corresponding to a discrete Fourier transform. *If* in fact there is a *continuous* spectrum corresponding to the observations, for example because these are samples from some continuous function, or they constitute a finite sample from a random process with known properties, then that (continuous) spectrum and the above discrete spectrum are related by a convolution; see e.g. Gregory (2005, appendix B).

3. Bayesian inference.

3.1. *The maximum entropy model for a given power spectrum.* If all that is known about the time series are its mean of zero and its (discrete) power spectrum, then the *maximum entropy principle* provides a way to determine a sampling distribution to be assigned for modelling the observations. This will ensure that these, and only these assumptions enter into the following analysis, and in that sense yields the most conservative possible model specification (Bretthorst, 1999; Jaynes, 2003; Gregory, 2005). For a given power spectral density (S_1 or S_2, as in (8)), the maximum entropy distribution for the observables is a Normal distribution for all A_j and B_j, where all the (frequency domain) observables are stochastically independent, have zero mean and

$$\text{(9)} \qquad \text{Var}(A_j) \;=\; \sigma_j^2,$$
$$\text{(10)} \qquad \text{Var}(B_j) \;=\; \kappa(j)\,\sigma_j^2 \qquad \text{for } j = 0,\dots,\lfloor N/2 \rfloor$$

(Bretthorst, 1999), where

$$\text{(11)} \quad \sigma_j^2 \;=\; S_1(f_j) \;=\; (1 + \kappa(j))\,S_2(f_j) \qquad \text{for } j = 0,\dots,\lfloor N/2 \rfloor.$$

3.2. *Some implications.* 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 then 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 $m, n \in \{1,\dots,N\}$, and corresponding to times t_m and t_n) is given by:

$$\text{(12)}\,\text{Cov}(X_m, X_n) \;=\; \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \left(S_1(f_j)\,(1 + \kappa(j))\,\tfrac{1}{2}\cos\!\left(2\pi f_j(t_n - t_m)\right) \right)$$
$$=\; \gamma(t_n - t_m)$$

(see section 6.3 for an explicit derivation). 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 $t_n - t_m$. This means that, remarkably though not surprisingly, the application of the maximum entropy principle yields a *strictly stationary* model (Taqqu, 1988).

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

$$
(13) \quad \Sigma_X = \begin{pmatrix}
\gamma(0) & \gamma(\Delta_t) & \gamma(2\Delta_t) & \cdots & \gamma(2\Delta_t) & \gamma(\Delta_t) \\
\gamma(\Delta_t) & \gamma(0) & \gamma(\Delta_t) & \cdots & \gamma(3\Delta_t) & \gamma(2\Delta_t) \\
\gamma(2\Delta_t) & \gamma(\Delta_t) & \gamma(0) & \cdots & \gamma(4\Delta_t) & \gamma(3\Delta_t) \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(2\Delta_t) & \gamma(3\Delta_t) & \gamma(4\Delta_t) & \cdots & \gamma(0) & \gamma(\Delta_t) \\
\gamma(\Delta_t) & \gamma(2\Delta_t) & \gamma(3\Delta_t) & \cdots & \gamma(\Delta_t) & \gamma(0)
\end{pmatrix}
$$

since γ is periodic such that $\gamma(i\Delta_t) = \gamma((N+1-i)\Delta_t)$. 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 a new problem, and it arises for any conventional spectral analysis of discretely sampled data (Harris, 1978); it may just not always be as obvious in its consequences.

Note that the Normal distribution of the variables A_j and B_j implies a (scaled) χ^2-distribution with 2 degrees of freedom for the power $P_1(f_j)$ (or $P_2(f_j)$). This distribution is equal to an Exponential distribution, which again is the maximum entropy distribution for a quantity that is positive and has a given expectation value. So applying the maximum entropy principle at the level of A_j and B_j also yields the corresponding maximum entropy distribution for the power, on which the restriction was originally imposed.

3.3. *Inferring an unknown spectrum.* Now suppose the spectrum is unknown and to be inferred from the data $x_1, \ldots, x_N$. An unknown spectrum is equivalent to the variance parameters $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$ being unknown. 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 χ^2-distribution* with scale parameter s_j^2 and degrees-of-freedom parameter ν_j:

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

(Gelman et al., 1997). The degrees-of-freedom here denote the precision in the prior distribution, while the scale determines its order of magnitude. For increasing ν_j the distribution's variance goes towards zero, and for $\nu_j \to 0$ the density converges toward the non-informative (and improper) distribution with density $f(\sigma^2) = \frac{1}{\sigma^2}$, that is uniform on $\log(\sigma^2)$, and which also constitutes the corresponding *Jeffreys prior* for this problem (Jeffreys, 1946). If σ_j^2 follows an Inv-$\chi^2(\nu_j, s_j^2)$ distribution, then its expectation and variance

are given by

$$(15) \quad \mathrm{E}[\sigma_j^2] \;=\; \frac{\nu_j}{\nu_j - 2}\, s_j^2 \qquad \text{and} \qquad \mathrm{Var}(\sigma_j^2) \;=\; \frac{2\nu_j^2}{(\nu_j - 2)^2(\nu_j - 4)}\, s_j^4,$$

and the mean and variance are finite for $\nu_j > 2$ and $\nu_j > 4$, respectively (Gelman et al., 1997). Alternatively, the expressions in (15) may be inverted to

$$(16) \qquad \nu_j \;=\; 4\left(\frac{\mathrm{E}[\sigma_j^2]^2}{\mathrm{Var}(\sigma_j^2)} + 1\right) \qquad \text{and} \qquad s_j^2 \;=\; \frac{\nu_j - 2}{\nu_j}\,\mathrm{E}[\sigma_j^2],$$

which allows one to specify the scale s_j^2 and degrees-of-freedom ν_j based on prior expectation and variance of σ_j^2, respectively. A specification of s_j^2 independent of j means a priori white noise, and specifying individual ν_j for different j indicates varying prior certainty across the spectrum.

Assuming Normality of A_j and B_j, the likelihood function (as a function of the parameters $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$) can be expressed as:

$$p(x_1, \ldots, x_N \,|\, \sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2)$$

$$(17) \quad = \quad \prod_{j=0}^{\lfloor N/2 \rfloor} \left[\frac{1}{\sqrt{2\pi}\,\sigma_j} \exp\left(-\frac{a_j^2}{2\sigma_j^2}\right) \times \left(\frac{1}{\sqrt{2\pi}\,\sigma_j}\right)^{\kappa(j)} \exp\left(-\frac{b_j^2}{2\sigma_j^2}\right) \right]$$

$$(18) \quad = \quad \exp\left(-\frac{N}{2}\log(2\pi) + \sum_{j=0}^{\lfloor N/2 \rfloor} \left[-(1 + \kappa(j))\log(\sigma_j) - \frac{a_j^2 + b_j^2}{2\sigma_j^2} \right]\right)$$

$$(19) \quad \propto \quad \exp\left(\sum_{j=0}^{\lfloor N/2 \rfloor} -\frac{1 + \kappa(j)}{2} \left[\log(S_2(f_j)) + \frac{\frac{\Delta_t}{N}|\tilde{x}_j|^2}{S_2(f_j)} \right]\right).$$

The term $\tilde{x}_j$ here denotes the jth element of the Fourier-transformed data vector, as defined in appendix 6.1 and 6.2. An equivalent expression could also be derived in terms of the time-domain observations $x_1, \ldots, x_N$ via their covariance matrix implied by $\sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2$, which was given in (13).

If one has more data available than just a single time series $x_1, \ldots, x_N$, in particular, if there are k independent time series of the same kind, $x_1^{(1)}, \ldots, x_N^{(1)}$, $x_1^{(2)}, \ldots, x_N^{(2)}, x_1^{(3)}, \ldots, x_N^{(k-1)}, x_1^{(k)}, \ldots, x_N^{(k)}$, then the likelihood function generalises to

$$p(x_1^{(1)}, \ldots, x_N^{(k)} \,|\, \sigma_0^2, \ldots, \sigma_{\lfloor N/2 \rfloor}^2)$$

$$(20) \quad = \quad \exp\left(-\frac{kN}{2}\log(2\pi) + \sum_{j=0}^{\lfloor N/2 \rfloor} \left[-k(1 + \kappa(j))\log(\sigma_j) - \frac{v_j}{2\sigma_j^2} \right]\right),$$

where

$$(21) \qquad v_j \;=\; \sum_{\ell=1}^{k}\Bigl(\bigl(a_j^{(\ell)}\bigr)^2 \,+\, \bigl(b_j^{(\ell)}\bigr)^2\Bigr)$$

is the total sum of all squared amplitudes corresponding to frequency f_j.

For given data, the posterior distribution of the σ_j^2 is again a scaled inverse χ^2-distribution:

$$(22) \quad \sigma_j^2 \,\Big|\, \{x_1,\ldots,x_N\} \;\sim\; \text{Inv-}\chi^2\left(\nu_j + 1 + \kappa(j),\; \frac{\nu_j s_j^2 \,+\, a_j^2 + b_j^2}{\nu_j + 1 + \kappa(j)}\right)$$

for a single time series, and

$$(23)\sigma_j^2 \,\Big|\, \{x_1^{(1)},\ldots,x_N^{(k)}\} \;\sim\; \text{Inv-}\chi^2\left(\nu_j + k(1 + \kappa(j)),\; \frac{\nu_j s_j^2 \,+\, v_j}{\nu_j + k(1 + \kappa(j))}\right)$$

for several time series, where all the different parameters corresponding to different frequencies are mutually independent. Comparing prior (14) and posterior (23), one can see that the prior distribution might be thought of as providing the information equivalent to ν_j observations (of coefficients a_j or b_j) with average squared deviation s_j^2 (Gelman et al., 1997).

3.4. *Some more implications.* The use of the conjugate prior distribution leads to a convenient expression for the posterior distribution of the variance parameters σ_j^2, and with that of the complete discrete spectrum. Also, if $\sigma_j^2 \sim \text{Inv-}\chi^2(\nu_j, s_j^2)$, then, since s_j^2 is a scale parameter, it follows that the distribution of $S_2(f_j) = \sigma_j^2/(1 + \kappa(j))$ is $\text{Inv-}\chi^2(\nu_j, s_j^2/(1 + \kappa(j)))$. There is a deterministic relationship between given variance parameters σ_j^2 and the implied autocorrelation function $\gamma(t)$ (see (12)), and so for random σ_j^2 the distribution of the corresponding $\gamma(t)$ may be numerically explored via sampling from the distribution of the σ_j^2. The expectation and variance of $\gamma(t)$ may also be derived analytically. The expected autocovariance (12), with respect to the distribution of the σ_j^2, is:

$$(24) \qquad \text{E}[\gamma(t)] \;=\; \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \Bigl(\text{E}[\sigma_j^2]\,(1 + \kappa(j))\,\tfrac{1}{2}\cos(2\pi f_j t)\Bigr),$$

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

$$(25) \quad \text{Var}(\gamma(t)) \;=\; \frac{1}{N^2\Delta_t^2} \sum_{j=0}^{\lfloor N/2 \rfloor} \Bigl(\text{Var}(\sigma_j^2)\,(1 + \kappa(j))^2\,\tfrac{1}{4}\cos(2\pi f_j t)^2\Bigr),$$

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

The moments given in (24) and (25) are exact, but they do not provide a complete picture of the distribution of $\gamma(t)$, since the distribution is generally rather skewed (see the examples in section 4); in particular, a Normal approximation for $\gamma(t)$ based on mean and variance would be inappropriate. A way to still make use of mean and variance may be to derive conservative (but exact) confidence bounds for $\gamma(t)$ via the Chebyshev inequality:

$$(26) \qquad \mathrm{P}\Big(|\gamma(t) - \mathrm{E}[\gamma(t)]| \geq \frac{1}{\sqrt{\alpha}}\sqrt{\mathrm{Var}(\gamma(t))}\Big) \leq \alpha.$$

Looking at the likelihood function (19) (or the related expression (29) in section 3.5), one can see that once one assumes the discrete spectrum to be known a priori, the model is equivalent to the one described by Finn (1992), except that there the actual continuous power spectral density is used in place of and as an approximation of the discrete analogue.

3.5. *Implementation as part of a larger model.* The greatest practical use for the method described above is probably when implemented as part of a larger model, e.g. in order to account for additive noise in the measurement of some signal. The interrelations of signal and noise parameters will of course add to the complexity of the posterior distribution of noise parameters, but their *conditional* posterior distribution will still be as described above for a given set of signal parameter values. This makes inference easy when MCMC methods are used for posterior inference. Dividing the set of parameters into subsets corresponding to signal and noise in particular allows to easily generate draws from the conditional posterior distribution of noise parameters, for any given set of signal parameter values. This means that the noise model may easily be incorporated in an MCMC algorithm using a Gibbs step for the noise parameters. Consider for example the measurement of a time series $y(t)$, sampled at equidistant time points t_i, $i = 1, \ldots, N$, that is assumed to be a sum of a signal, which is some function of time $f_{\vec{\theta}}(t)$, and coloured noise $n_{\vec{\sigma}^2}(t)$ as defined above:

$$(27) \qquad y(t_i) = f_{\vec{\theta}}(t_i) + n_{\vec{\sigma}^2}(t_i).$$

The prior then is defined so that it factors out as $p(\vec{\theta}, \vec{\sigma}^2) = p(\vec{\theta}) \times p(\vec{\sigma}^2)$, where $p(\vec{\sigma}^2)$ again is defined as a product of Inv-$\chi^2(\nu_j, s_j^2)$ distributions as in (14). The likelihood, a function of the composite parameter vector $\left(\begin{smallmatrix}\vec{\theta}\\\vec{\sigma}^2\end{smallmatrix}\right)$, depends on the data $y(t_i)$ via the vector of noise residuals that are implied by the signal parameters $\vec{\theta}$:

$$(28) \qquad r_{\vec{\theta}}(t_i) = y(t_i) - f_{\vec{\theta}}(t_i)$$

and is computed based on the Fourier-transformed residual vector $\tilde{r}_{\vec{\theta}}$ and the noise parameters $\vec{\sigma}^2$ as in equations (18) and (19):

$$(29) \quad p(y|\vec{\theta},\vec{\sigma}^2) \;\propto\; \exp\left(\sum_{j=0}^{\lfloor N/2\rfloor}\left[-\frac{1+\kappa(j)}{2}\log(\sigma_j^2) - \frac{\frac{\Delta_t}{N}|\tilde{r}_{\vec{\theta}}(f_j)|^2}{2\sigma_j^2}\right]\right).$$

MCMC sampling from the posterior distribution may now be done by implementing a Gibbs sampler that alternatingly samples from the conditional distributions of $\vec{\sigma}^2|\vec{\theta}$ and $\vec{\theta}|\vec{\sigma}^2$. One way to implement this would be to define some initial value $\vec{\theta}_{(0)}$, then set $t := 1$ and

1.) determine the noise residuals $r_{\vec{\theta}_{(t-1)}}$ corresponding to $\vec{\theta}_{(t-1)}$ and compute their Fourier transform $\tilde{r}_{\vec{\theta}_{(t-1)}}$

2.) generate a draw $\vec{\sigma}^2_{(t)}$ from the conditional distribution of $\vec{\sigma}^2|\vec{\theta}_{(t-1)}$, i.e. draw each individual $\sigma^2_{j\,(t)}$ from an Inv-$\chi^2\left(\nu_j+1+\kappa(j),\frac{\nu_j s_j^2 + \frac{\Delta_t}{N}|\tilde{r}_{\vec{\theta}_{(t-1)}}(f_j)|^2}{\nu_j+1+\kappa(j)}\right)$ distribution (as in equation (22)), where ν_j and s_j^2 are the corresponding prior degrees-of-freedom and scale parameters

3.) generate a draw $\vec{\theta}_{(t)}$ from the conditional distribution of $\vec{\theta}|\vec{\sigma}^2_{(t)}$, e.g. by generating a suitable Metropolis-proposal $\vec{\theta}^*$, computing the ratio of posterior densities

$$(30) \qquad r \;=\; \frac{p(\vec{\theta}^*,\vec{\sigma}^2_{(t)}|y)}{p(\vec{\theta}_{(t-1)},\vec{\sigma}^2_{(t)}|y)} \;=\; \frac{p(\vec{\theta}^*)\,p(y|\vec{\theta}^*,\vec{\sigma}^2_{(t)})}{p(\vec{\theta}_{(t-1)})\,p(y|\vec{\theta}_{(t-1)},\vec{\sigma}^2_{(t)})},$$

and setting

$$(31) \qquad \vec{\theta}_{(t)} := \begin{cases} \vec{\theta}^* & \text{with probability } \min(1,r) \\ \vec{\theta}_{(t-1)} & \text{otherwise} \end{cases}$$

4.) set $t := t + 1$ and continue at 1.)

Note that the likelihood evaluations in (30) again require the Fourier transforms of the noise residuals. For more details on the use of MCMC methods for posterior computations see e.g. Gelman et al. (1997). Application of the above method is actually not restricted to modelling of additive noise.

In the context of an MCMC application (such as in the example shown in section 4.3) it is often helpful to extend the basic algorithm by using 'tempered' versions of the posterior distribution; examples of such cases are the *simulated annealing* or *parallel tempering* algorithms (Röver, 2007). If part of the model specification is a method as described above, it turns out

that the 'tempered' (conditional) posterior distribution again is an Inv-χ^2-distribution, which is very convenient for a practical implementation. More details about these tempered distributions are given in section 6.4.

4. Examples.

4.1. *Sunspots: little data.*   The following two examples are standards for this kind of analysis, the *sunspot data* (Andrews and Herzberg, 1985), which give monthly relative numbers of observed sunspots for the years 1749–1983. These data are illustrated in Fig. 1. At first only a small subset of the data are considered, the last decade covering the years 1974-1983. There are $N = 120$ observations in that data subset, sampled at a rate of $\frac{1}{\Delta_t} = 12$ per year. Consequently, there are $\lfloor N/2 + 1 \rfloor = 61$ parameters $\sigma_0^2, \ldots, \sigma_{60}^2$ corresponding to the Fourier frequencies $f_0 = 0, f_1 = \frac{1}{10}, \ldots, f_{60} = 6$.
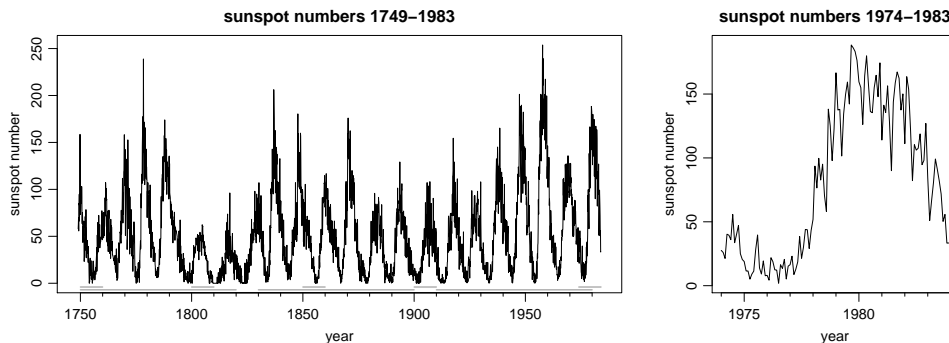

FIG 1. *The sunspots data, the complete data set (left) and the final decade (right). The grey lines in the left plot indicate the different data segments used in the examples below.*

We will consider two prior settings here, firstly a prior where the prior scale s_j^2 for each frequency is estimated by computing an empirical spectrum estimate for the four decades starting 1750, 1800, 1850 and 1900 and averaging over these; the prior degrees-of-freedom ν_j are then all set to one. The second prior setting is the non-informative prior with zero degrees of freedom.

The resulting posterior spectra for these two cases are illustrated in Fig. 2. In the top plot, the prior scale is indicated by the dotted line; the non-informative prior used for the bottom plot does not depend on its scale parameter. Note that the posterior distributions corresponding to the first and (since the number of data samples N was even) last frequency (σ_0^2 and σ_{60}^2) are wider than the others, as they have one less degree of freedom (see (22)). The expectation does not exist for these two parameters in case
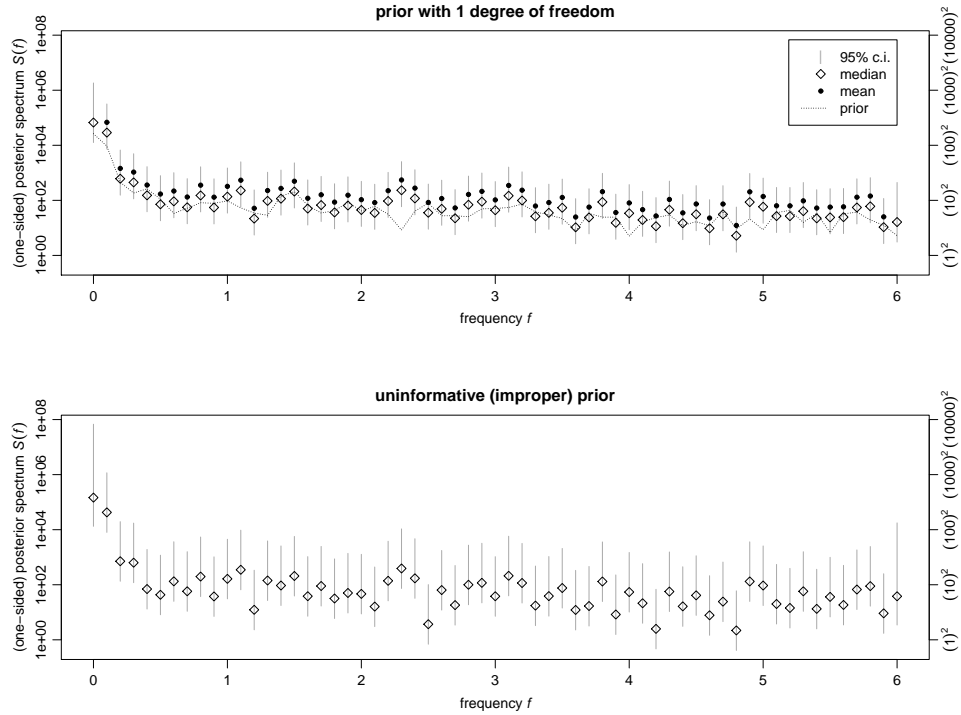
FIG 2. *Posterior power spectra for the last decade of the sunspots data, years 1974–1983. The vertical grey lines indicate the central posterior interval that each of the parameters $\sigma_j^2 = S_2(f_j)$ falls into with 95% probability. The diamonds and dots indicate median and mean for each of the distributions. For the top plot a prior with one degree of freedom for each parameter was used (dotted line). For the bottom plot a non-informative, improper Jeffreys prior with zero degrees of freedom was used.*

of the one-degree-of-freedom prior, and it does not exist for any of the parameters in the case of the noninformative prior. Looking at the width of the posterior, one can see that for both prior settings the uncertainty in each individual parameter (indicated by the vertical lines) is greater than an order of magnitude.

Concerning the analysed data, the plots only reveal that something appears to be going on at low frequencies (ignoring the very first parameter σ_0^2 here, as it corresponds to zero frequency, i.e., a constant 'intercept' term). Looking at the complete data (Fig. 1) one can see that there is an obvious periodicity, and using a more appropriate analysis for that question one can determine that period to be at about 11 years (Bretthorst, 1988), slightly above the observation period for this subset of data. Note that the model described here is actually neither intended nor appropriate for modelling

single periodic features in the data. If one is interested in determining the number or frequencies and amplitudes of individual periodic contributions, for instance, there are other related but distinctly different methods available (Bretthorst, 1988; Umstätter et al., 2005). The posterior distributions derived here are intended for robust inference on the *discrete* spectrum of a finite time series.
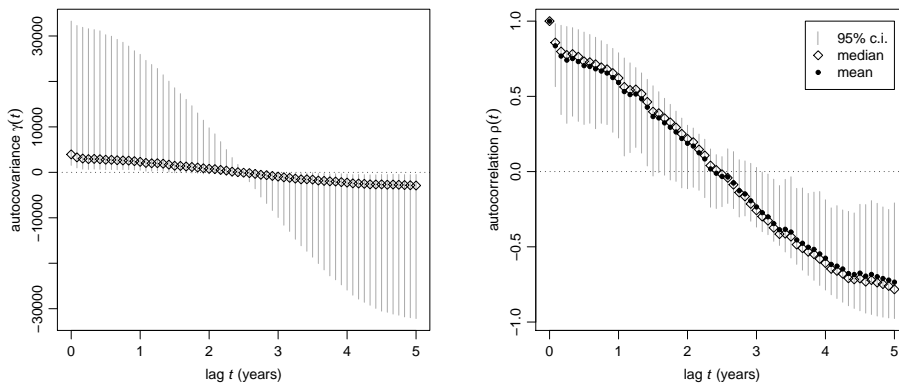

FIG 3. *Posterior distributions of the autocovariance function $\gamma(t)$ and autocorrelation function $\rho(t) = \frac{\gamma(t)}{\gamma(0)}$, corresponding to the posterior spectrum in the top plot of Fig. 2, and derived by Monte Carlo simulation. The expectations of the $\gamma(t)$ are not finite here.*

Using the relations derived in section 3.2, one can now also infer the posterior distribution of the autocovariance function $\gamma(t)$, and with that the autocorrelation function $\rho(t) = \frac{\gamma(t)}{\gamma(0)}$. Fig. 3 illustrates these distributions as derived by Monte Carlo simulation. Given the posterior distribution of all the σ_j^2, these can easily be sampled and then the autocovariance function γ corresponding to a set of sampled σ_j^2 parameters may be derived. Repeated sampling of these autocovariance functions then allows to estimate the posterior distribution of each $\gamma(t)$ and $\rho(t)$. In the example shown here, the intercept term σ_0^2 was deliberately set to zero as the data obviously do not have zero mean.

The autocovariance function here has 61 different elements, as many as there are Fourier frequencies, and due to the model specification in (1) continues periodically beyond that. The autocovariances tend to have highly asymmetric distributions, and for the number of posterior degrees-of-freedom in the given example, their posterior expectations do not exist. Note that the distributions of the individual $\gamma(t)$ are highly correlated, as

opposed to those of the individual σ_j^2, which are mutually independent.

Relating these results back to the data, the plots indicate a positive correlation between nearby samples, and a negative correlation for a lag of some 5 years, which makes sense given the appearance of the data (Fig. 1) and the additional knowledge that the complete data set indeed exhibits a periodicity of about 11 years.

4.2. *Sunspots: more data.*   In the following example more data are considered, in particular *several* pieces of *longer* observation time; as a consequence the observation period is significantly longer than 11 years, and the higher moments (mean and variance) of the posterior autocovariance exist. Here we will examine three segments of the sunspots data, of a length of 70 years each, and starting in the years 1750, 1830 and 1910 respectively (see Fig.1). We will assume that these three segments are independent of each other, and incorporate them as described in (23). The very last datum of each segment is dropped here so that the resulting individual sample size $N = 839$ is odd and hence the posteriors for all the variance parameters σ_j^2 have 6 degrees of freedom, except for the very first parameter σ_0^2, which is then omitted as well, as this corresponds to an (obviously non-zero) intercept term.
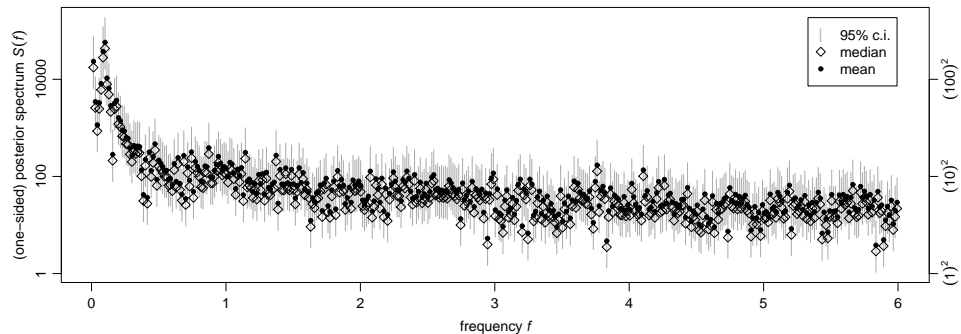

FIG 4. *The posterior spectrum based on three 70-year segments of the sunspots data that are assumed independent.*

Fig. 4 shows the resulting posterior spectrum. The widths of the individual distributions are much smaller than in the previous example shown in Fig. 2; now they "only" span one order of magnitude. Now that the observation time is significantly larger than the periodicity in the data, a peak at frequencies around $\frac{1}{11}$ becomes apparent. As the degrees-of-freedom of the posterior distributions are all equal to 6, expectations and variances now exist for all of them.
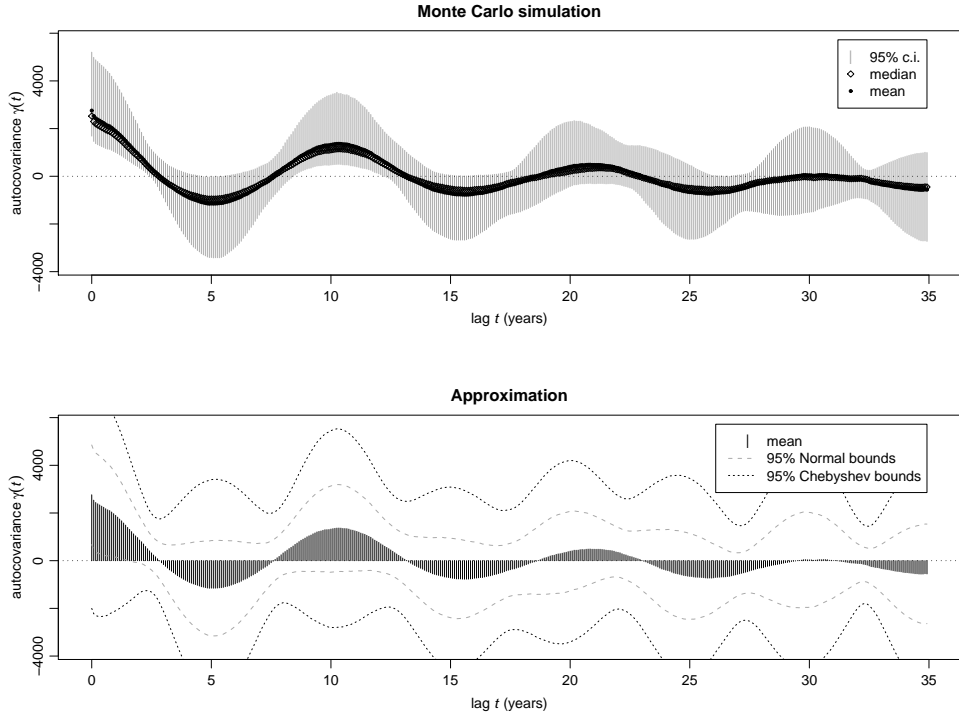
Fig 5. *Posterior distributions of the autocovariance function $\gamma(t)$ corresponding to the spectrum shown in Fig. 4. The top panel is based on Monte Carlo simulation, and the bottom panel gives the exact posterior mean and approximative bounds based on the (exact) variance.*

Fig. 5 illustrates the corresponding posterior distribution of the autocovariance function $\gamma(t)$. The top panel shows Monte Carlo estimates of mean, median and quantiles as in Fig. 3. Since mean and variance of *all* the individual σ_j^2 are finite, the exact posterior mean and variance of $\gamma(t)$ can be derived as described in section 3.4. The dashed lines in the bottom plot show 95% confidence bounds based on a (rather inappropriate) Normal approximation and an (extremely conservative) approximation based on the Chebyshev inequality (26). Comparing both plots at the peak at a lag of around 5 years one can clearly see how the Normal distribution does not account for the skewness in the distribution: the posterior distribution is skewed towards low values of $\gamma(t)$, and hence both upper and lower bounds are overestimated. Especially in the top panel of Fig. 5 one can see two prominent periodicities in the autocovariance function: one that completes roughly 3.5 cycles within 35 years and one that completes 3 cycles. These periods correspond to the two spectral parameters that are at the low-frequency peak in Fig. 4.

4.3. *MCMC example.* The following example demonstrates the use of the approach described above for modelling a non-white noise component in a more complex model when MCMC methods are used for inference. The explicit closed-form expression of the spectrum's posterior distribution (23) allows one to easily incorporate the noise model as described in section 3.5.

Consider a time-series where the $N = 100$ sampled time points each are at $t_i = \frac{i}{100}$. The data $y(t_1), \ldots, y(t_{100})$ are modelled as

$$(32) \qquad y(t_i) \; = \; f_{f,\dot{f},a,\phi}(t_i) \, + \, n_{\vec{\sigma}}(t_i),$$

where $n_{\vec{\sigma}}(t_i)$ is non-white noise of supposedly unknown spectrum, and $f_{f,\dot{f},a,\phi}(t)$ is a "chirping" signal waveform of increasing frequency:

$$(33) \qquad f_{f,\dot{f},a,\phi}(t) \; = \; a \, \sin(2\pi(f + \dot{f}t)t + \phi)$$

where f and $\dot{f}$ are the *frequency* and *frequency derivative*, a is the *amplitude*, and ϕ is the *phase.* Fig. 6 shows an example of such data, where the signal
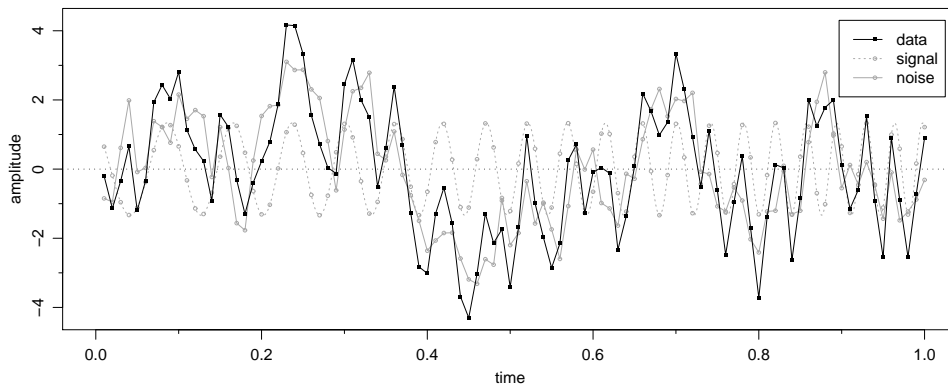

FIG 6. *The (simulated) data used in the MCMC example, and both signal waveform and noise alone.*

parameters are $f = 10$, $\dot{f} = 10$, $a = 1.335$ and $\phi = 2$. The noise that is added to the signal here is generated from an autoregressive process

$$(34) \qquad n(t_i) \; = \; \tfrac{3}{4} n(t_{i-1}) + x(t_i),$$

where the innovations $x(t_i)$ are drawn independently from a uniform distribution across the interval $[-\frac{\sqrt{12}}{2}, \frac{\sqrt{12}}{2}]$, so that they have zero mean and unit variance. The noise process defined this way is clearly not Normal and has (due to the positive correlation of subsequent samples) a higher power

at low frequencies. The signal parameters are chosen such that the resulting signal-to-noise ratio (SNR) is unity.

Given a prior for the signal parameters f, $\dot{f}$, a and ϕ, one can use the spectrum model (1), with corresponding prior settings, in order to infer the four signal parameters and the 50 noise parameters. In the following the prior is defined 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). Assuming one has a rough idea of the noise
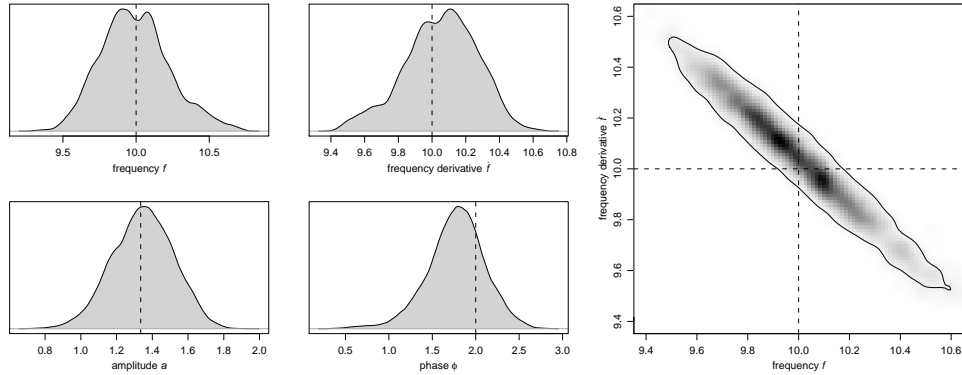

FIG 7. *Marginal posterior distributions of the signal parameters; the left four panels illustrate the distributions of each individual parameter, while the right panel shows the joint distribution of the two frequency parameters. The dashed lines indicate the true values.*

variance (which for the definition in (34) would be at 2.29), a corresponding white noise spectrum (with $S_1(f) = 0.0457$) is assumed for the prior scale of the noise parameters σ_j^2. The prior degrees-of-freedom parameters are all set to 0.5. The posterior distribution of all (signal *and* noise) parameters may then be inferred using a simple Metropolis sampler within which the noise parameters are drawn in an additional Gibbs step, as described in section 3.5.

Figures 7 and 8 illustrate the posterior distributions for both signal and noise parameters in the above example data as derived by such an MCMC algorithm. Fig. 7 shows the marginal distributions of the four individual signal parameters, and the joint marginal distribution of the two frequency parameters. Fig. 8 illustrates the marginal posterior distributions of the 51 noise parameters σ_0^2, ..., σ_{50}^2 that are estimated along with the signal parameters. The prior scale and the actual spectrum of the noise process defined in (34) are shown in the same plot. The individual noise parameters' posterior distributions indicate the excess of low-frequency noise (as opposed
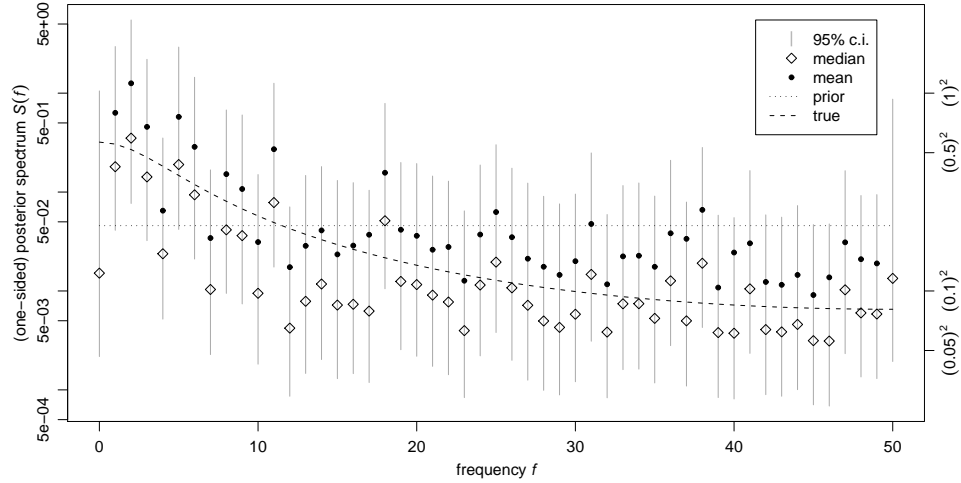
FIG 8. *The posterior noise spectrum as inferred along with the signal parameters shown in Fig. 7. A priori, the noise is assumed to be white (dotted line), while the actual noise is dominated by low frequencies (dashed line).*

to the prior scale), while at higher frequencies they are noticeably skewed towards the (greater) prior noise scale.

5. Conclusions. A general time series model for inference on the (discrete) power spectrum of a time series is introduced, which is robust in the sense that inference is only based on a few explicit presumptions. The model assumptions are (i) zero mean of the time series, (ii) a finite spectral density, and (iii) the expression of prior information about the spectrum in terms of a scaled inverse χ^2-distribution. The robustness and generality of the model comes at the price of a large number of parameters that need to be modelled: half as many parameters as there are samples in the time series, where each parameter corresponds to a Fourier frequency. The prior mean and variance of each of the parameters may be defined independently. As the number of parameters grows with the length of the time series, the consideration of additional data only improves posterior certainty with respect to the overall power within a fixed frequency interval, or when considering an increasing number of independent time series.

The inferred spectrum needs be understood in the way it is defined, as the expectation of summed and squared Fourier coefficients. Consequently, what is explicitly modelled here is the discrete analogue of a continuous power spectral density, the result of a discrete Fourier transform, and the convolution that implicitly comes with the transform. The model is in par-

ticular not appropriate for estimating frequencies or amplitudes of single, isolated spectral lines; if, for example, the time series' spectrum is known to be somewhat smooth, then the approach would still be correct, but suboptimal. Due to the computationally convenient conjugate form of the posterior distribution it is particularly useful for modelling an unknown noise spectrum, that constitutes a nuisance parameter rather than being of interest itself, as this can be handled "on the fly" within a Bayesian MCMC application. Some of the methods described here have been coded as an extension package for R and are available for free download (Röver, 2008).

6. Appendix.

6.1. *Discrete Fourier transform (DFT) definition.* Due to the various differing conventions that are commonly used, the discrete Fourier transform is defined here explicitly, following the definitions given e.g. in Gregory (2005). The discrete Fourier transform is defined for a real-valued function h of time t, sampled at N discrete time points, at a sampling rate of $\frac{1}{\Delta_t}$. The transform maps from

$$(35) \qquad \{h(t) \in \mathbb{R} : \ t = 0, \Delta_t, 2\Delta_t, \ldots, (N-1)\Delta_t\}$$

to a function of frequency f

$$(36) \qquad \{\tilde{h}(f) \in \mathbb{C} : \ f = 0, \Delta_f, 2\Delta_f, \ldots, (N-1)\Delta_f\},$$

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

$$(37) \qquad \tilde{h}(f) \ = \ \sum_{j=0}^{N-1} h(j\Delta_t)\exp(-2\pi \mathrm{i}j\Delta_t f).$$

Since h is real-valued, the elements of $\tilde{h}$ are symmetric (and with that redundant) in the sense that $\tilde{h}(i\Delta_f)$ and $\tilde{h}((N-i)\Delta_f)$ are complex conjugates. The back-transformation to the time domain, the *inverse DFT*, is given by

$$(38) \qquad h(t) \ = \ \frac{1}{N}\sum_{j=0}^{N-1} \tilde{h}(j\Delta_f)\exp(2\pi \mathrm{i}j\Delta_f t).$$

6.2. *Relationship between DFT and time series model.* Let

$$(39) \qquad \alpha_j \ = \ \mathrm{Re}(\tilde{h}(f_j)) \quad \text{and} \quad \beta_j \ = \ \mathrm{Im}(\tilde{h}(f_j)),$$

i.e.: $\tilde{h}(f_j) = \alpha_j + \beta_j \mathrm{i}$ and $f_j = j\Delta_f$. For simplicity and clarity, in the following derivation N is assumed to be even; this requires the special treatment of

first ($j = 0$) and last ($j = N/2$) purely real Fourier coefficients that can be seen starting from (41) below. For odd N, the derivation is similar except for that last coefficient.

The inverse DFT was defined as (38):

$$(40)(t) \quad = \quad \frac{1}{N} \sum_{j=0}^{N-1} \tilde{h}(f_j) \exp(2\pi \mathrm{i} f_j t)$$

$$= \quad \frac{1}{N} \sum_{j=1}^{\frac{N}{2}-1} \left[\tilde{h}(f_j) \exp(2\pi \mathrm{i} f_j t) + \overline{\tilde{h}(f_j)} \exp(2\pi \mathrm{i} f_{N-j} t) \right]$$

$$(41) \qquad + \tfrac{1}{N} \tilde{h}(f_0) \exp(2\pi \mathrm{i} f_0 t) \; + \; \tfrac{1}{N} \tilde{h}(f_{N/2}) \exp(2\pi \mathrm{i} f_{N/2} t)$$

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

$$+ (\alpha_j - \beta_j \mathrm{i})(\cos(-2\pi \; f_{N-j} t) - \sin(-2\pi f_{N-j} t)\mathrm{i}) \Big]$$

$$(42) \qquad + \tfrac{1}{N} \alpha_0 \cos(-2\pi \mathrm{i} f_0 t) \; + \; \tfrac{1}{N} \alpha_{N/2} \cos(-2\pi \mathrm{i} f_{N/2} t)$$

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

$$+ (\alpha_j - \beta_j \mathrm{i})(\cos(-2\pi f_j t) + \sin(-2\pi f_j t)\mathrm{i}) \Big]$$

$$(43) \qquad + \tfrac{1}{N} \alpha_0 \; + \; \tfrac{1}{N} \alpha_{N/2} \cos(-2\pi \mathrm{i} f_{N/2} t)$$

$$= \quad \frac{1}{N} \sum_{j=1}^{\frac{N}{2}-1} \Big[(\alpha_j \cos(\ldots) + \beta_j \sin(\ldots)) + (-\alpha_j \sin(\ldots) + \beta_j \cos(\ldots))\mathrm{i}$$

$$+ (\alpha_j \cos(\ldots) + \beta_j \sin(\ldots)) + (\alpha_j \sin(\ldots) - \beta_j \cos(\ldots))\mathrm{i} \Big]$$

$$(44) \qquad + \tfrac{1}{N} \alpha_0 \; + \; \tfrac{1}{N} \alpha_{N/2} \cos(-2\pi \mathrm{i} f_{N/2} t)$$

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

$$(45) \qquad + \tfrac{1}{N} \alpha_0 \; + \; \tfrac{1}{N} \alpha_{N/2} \cos(-2\pi \mathrm{i} f_{N/2} t)$$

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

$$(46) \qquad + \tfrac{1}{N} \alpha_0 \; + \; \tfrac{1}{N} \alpha_{N/2} \cos(2\pi \mathrm{i} f_{N/2} t)$$

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 this with the definition in (1), one can see that in general the realisations 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

$$(47) \qquad a_j = (1 + \kappa(j)) \sqrt{\tfrac{\Delta_t}{N}} \, \alpha_j \quad \text{and} \quad b_j = -(1 + \kappa(j)) \sqrt{\tfrac{\Delta_t}{N}} \, \beta_j$$

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

$$(48) \quad a_j^2 + b_j^2 = (1 + \kappa(j))^2 \tfrac{\Delta_t}{N} (\alpha_j^2 + \beta_j^2) = (1 + \kappa(j))^2 \tfrac{\Delta_t}{N} |\tilde{h}(f_j)|^2.$$

6.3. *Autocovariance derivation.* Here the autocovariance stated in equation (12) is derived explicitly.

$$\mathrm{Cov}(X_m, X_n)$$

$$= \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \sum_{k=0}^{\lfloor N/2 \rfloor} \mathrm{E}\Big[\big(\sqrt{A_j^2 + B_j^2} \, \sin(2\pi f_j t_m + \varphi_j)\big)$$

$$(49) \qquad\qquad\qquad \times \big(\sqrt{A_k^2 + B_k^2} \, \sin(2\pi f_k t_n + \varphi_k)\big)\Big]$$

$$= \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \mathrm{E}\Big[\big(\sqrt{A_j^2 + B_j^2} \, \sin(2\pi f_j t_m + \varphi_j)\big)$$

$$(50) \qquad\qquad\qquad \times \big(\sqrt{A_j^2 + B_j^2} \, \sin(2\pi f_j t_n + \varphi_j)\big)\Big]$$

$$(51) = \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \mathrm{E}\Big[(A_j^2 + B_j^2) \sin(2\pi f_j t_m + \varphi_j) \sin(2\pi f_j t_n + \varphi_j)\Big]$$

$$(52) = \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \mathrm{E}\Big[A_j^2 + B_j^2\Big] \mathrm{E}\Big[\sin(2\pi f_j t_m + \varphi_j) \sin(2\pi f_j t_n + \varphi_j)\Big]$$

$$(53) = \frac{1}{N\Delta_t} \sum_{j=0}^{\lfloor N/2 \rfloor} \Big(S_1(f_j) (1 + \kappa(j)) \tfrac{1}{2} \cos(2\pi f_j (t_n - t_m))\Big) = \gamma(t_n - t_m)$$

where
$$(54) \qquad \mathrm{E}[A_j^2 + B_j^2] = (1 + \kappa(j)) \, \sigma_j^2 = (1 + \kappa(j)) \, S_1(f_j)$$

and

$$(55) \qquad \mathrm{E}\Big[\sin(2\pi f_j t_m + \varphi_j) \sin(2\pi f_j t_n + \varphi_j)\Big]$$

$$(56) \quad = \int \sin(2\pi f_j t_m + \varphi_j) \sin(2\pi f_j t_n + \varphi_j) \, \mathrm{dP}_{\varphi_j}$$

$$(57) \quad = \quad \int_0^{2\pi} \sin(2\pi f_j t_m + \varphi) \, \sin(2\pi f_j t_n + \varphi) \tfrac{1}{2\pi} \, \mathrm{d}\varphi$$

$$(58) \quad = \quad \int_0^{2\pi} \tfrac{1}{2\pi} \sin(\varphi) \, \sin(2\pi f_j(t_n - t_m) + \varphi) \, \mathrm{d}\varphi$$

$$(59) \quad = \quad \tfrac{1}{2\pi} \left(\tfrac{1}{2}\varphi \cos(2\pi f_j(t_n - t_m)) - \tfrac{1}{4} \sin(2\pi f_j(t_n - t_m) + 2\varphi) \right) \Big|_0^{2\pi}$$

$$(60) \quad = \quad \tfrac{1}{2} \cos(2\pi f_j(t_n - t_m)).$$

6.4. *Tempering the posterior distribution.* In the context of MCMC applications as in section **4.3**, it is often desirable to apply *tempering* to the (posterior) distribution of interest. This is supposed to make the distribution numerically more tractable, in the sense that it is easier to move through parameter space, or find its modes. Examples where tempering is utilised are *simulated annealing, parallel tempering* or *evolutionary MCMC* algorithms. Tempering may be done by specifying a certain 'temperature' $T > 0$, that is used to manipulate the original posterior distribution $p(\vartheta|y)$, in order to then use the tempered distribution $p_T(\vartheta)$ instead. In the most general case, the tempering may be applied by attaching an exponent of $\frac{1}{T}$ to the posterior density and defining:

$$(61) \qquad p_T(\vartheta) \; \propto \; p(\vartheta|y)^{\frac{1}{T}} \; = \; (p(\vartheta)\,p(y|\vartheta))^{\frac{1}{T}},$$

where $p(\vartheta)$ and $p(y|\vartheta)$ are prior and likelihood, respectively. Alternatively, the tempering may be applied to the likelihood part only by setting

$$(62) \qquad p_T(\vartheta) \; \propto \; p(\vartheta)\,p(y|\vartheta)^{\frac{1}{T}}$$

(Röver, 2007).

With respect to the model described here, where the posterior is an Inv-$\chi^2(\nu, s^2)$-distribution, tempering applied as in the former case (61) would again yield an Inv-χ^2-distribution:

$$(63) \qquad \text{Inv-}\chi^2 \left(\frac{\nu + 2}{T} - 2, \; \frac{\nu\,s^2}{2 + \nu - 2T} \right)$$

that is only defined as long as $T < \frac{\nu+2}{2}$.

If the tempering is only applied to the likelihood part of the posterior, as in (62), the resulting tempered distribution is also an Inv-χ^2-distribution. If the prior was defined as Inv-$\chi^2(\nu_0, s_0^2)$, with prior degrees of freedom ν_0 and prior scale s_0^2, then the regular (un-tempered) posterior is an

$$(64) \qquad \text{Inv-}\chi^2 \left(\nu_0 + n, \; \frac{\nu_0 s_0^2 + nv}{\nu_0 + n} \right)$$

distribution, where n is the sample size and v is the observed mean squared deviation (see also (23)). The tempered version of this is simply

$$(65) \qquad \text{Inv-}\chi^2 \left(\nu_0 + \tfrac{n}{T}, \ \frac{\nu_0 s_0^2 + \tfrac{n}{T} v}{\nu_0 + \tfrac{n}{T}} \right),$$

so implicitly the 'weight' of the observed data in the posterior is down-weighted by a factor of $\frac{1}{T}$ from n to $\frac{n}{T}$.

References.

Andrews, D. F. and A. M. Herzberg (1985). *Data: A collection of problems from many fields for the student and research worker.* New York: Springer-Verlag.

Blackman, R. B. and J. W. Tukey (1958). The measurement of power spectra from the point of view of communications engineering (Parts I and II). *The Bell System Technical Journal 37*, 185–282, 485–569.

Bretthorst, G. L. (1988). *Bayesian spectrum analysis and parameter estimation*, Volume 48 of *Lecture Notes in Statistics.* Berlin: Springer-Verlag.

Bretthorst, G. L. (1999). The near-irrelevance of sampling frequency distributions. In W. v. d. Linden et al. (Eds.), *Maximum Entropy and Bayesian Methods*, pp. 21–46. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Chatterjee, C., R. L. Kashyap, and G. Boray (1987, March). Estimation of close sinusoids in colored noise and model discrimination. *IEEE Transactions on Acoustics, Speech, and Signal Processing ASSP-35*(3), 328–337.

Cho, C.-M. and P. M. Djurić (1995, December). Bayesian detection and estimation of cisoids in colored noise. *IEEE Transactions on Signal Processing 43*(12), 2943–2952.

Finn, L. S. (1992, December). Detection, measurement, and gravitational radiation. *Physical Review D 46*(12), 5236–5249.

Gelman, A., J. B. Carlin, H. Stern, and D. B. Rubin (1997). *Bayesian data analysis.* Boca Raton: Chapman & Hall / CRC.

Godsill, S. and M. Davy (2002). Bayesian harmonic models for musical pitch estimation and analysis. In *Proceedings of the IEEE international conference on acoustics, speech and signal processing (ICASSP) 2002*, Volume 2, pp. 1769–1772. IEEE.

Gregory, P. C. (2005). *Bayesian logical data analysis for the physical sciences.* Cambridge: Cambridge University Press.

Harris, F. J. (1978, January). On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceedings of the IEEE 66*(1), 51–83.

Jaynes, E. T. (2003). *Probability theory: The logic of science.* Cambridge: Cambridge University Press.

Jeffreys, H. (1946, September). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London, Series A 186*(1007), 453–461.

Röver, C. (2007). *Bayesian inference on astrophysical binary inspirals based on gravitational-wave measurements.* Ph. D. thesis, The University of Auckland. URL `http://hdl.handle.net/2292/2356`.

Röver, C. (2008). bspec: Bayesian spectral inference. R package. URL `http://cran.r-project.org/package=bspec`.

Röver, C., A. Stroeer, E. Bloomer, N. Christensen, J. Clark, M. Hendry, C. Messenger, R. Meyer, M. Pitkin, J. Toher, R. Umstätter, A. Vecchio, J. Veitch, and G. Woan (2007, October). Inference on inspiral signals using LISA MLDC data. *Classical and Quantum Gravity 24*(19), S521–S527.

Taqqu, M. S. (1988). Weak stationarity. In S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of statistical sciences*. New York: Wiley & Sons.

Umstätter, R., N. Christensen, M. Hendry, R. Meyer, V. Simha, J. Veitch, S. Vigeland, and G. Woan (2005, July). Bayesian modeling of source confusion in LISA data. *Physical Review D 72*(2), 022001.

MAX-PLANCK-INSTITUT FÜR GRAVITATIONSPHYSIK
(ALBERT-EINSTEIN-INSTITUT)
CALLINSTRASSE 38
30167 HANNOVER
GERMANY

DEPARTMENT OF STATISTICS
THE UNIVERSITY OF AUCKLAND
PRIVATE BAG 92019
AUCKLAND 1142
NEW ZEALAND

CARLETON COLLEGE
PHYSICS AND ASTRONOMY DEPARTMENT
ONE NORTH COLLEGE STREET
NORTHFIELD, MN 55057
USA