A Bayesian Method for Oscillator Stability Analysis

François Vernotte Gilles Zalamansky

Abstract

The power spectral density of frequency fluctuations of an oscillator is generally modeled as a sum of power laws with integer exponents (from -2 to +2). However, a power law with a fractional exponent may exist. We propose a method for measuring the level of such a noise process and determining the probability density of the exponent. This yields a criterion for compatibility with an integer exponent. This method is based upon a Bayesian approach called the reference analysis of Bernardo-Berger. The application to a sequence of frequency measurement from a quartz oscillator illustrates this paper.

Index Terms

oscillator characterization, time stability, noise analysis, Bayesian analysis

I. Introduction

IT is commonly assumed that $S_y(f)$, the power spectral density (PSD) of frequency deviation of an oscillator, may be modeled as the sum of 5 power laws, defining 5 types of noise:

$$S_y(f) = \sum_{\alpha=-2}^{+2} h_\alpha f^\alpha \qquad (1)$$

where h_α is the level of the f^α noise. But it may be noticed that models with non-integer exponents are occasionally used.

The estimation of the noise levels is mainly achieved by using the Allan variance [1], which is

F. Vernotte is with the Observatoire de Besançon - UPRESA 6091, 41 bis av. de l'Observatoire, BP 1615, F-25010 Besançon Cedex. Phone: +33-3-81.66.69.22, email: francois@obs-besancon.fr

G. Zalamansky is with the Université de Metz, Campus Bridoux, F-57000 Metz

defined versus the integration time τ as:

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle \left(\overline{y}_{k+1} - \overline{y}_k\right)^2 \right\rangle . \tag{2}$$

In the frequency domain, the Allan variance may be considered as a filter. If the Allan variance versus the integration time τ is plotted, the graph exhibits different slopes, each slope corresponding to a type of noise:

$$\sigma_y^2(\tau) = C_\mu \tau^\mu \qquad \Leftrightarrow \qquad S_y(f) = h_\alpha f^\alpha \qquad \text{and} \qquad \alpha = -\mu - 1. \tag{3}$$

The estimation of C_μ yields an estimation of the noise level h_α.

However, this curve may exhibit an exponent μ which seems to be non-integer. Does this mean that the corresponding PSD is not compatible with the 5 power law model? In this paper, we propose a method for estimating the most probable value of this exponent in order to solve this ambiguity. This method is applied to an example of stability measurement.

II. Classical Stability Analysis of an Oscillator

A. *Sequence of frequency measurements*

[Fig. 1 about here.]

Figure 1 shows average frequency measurements $\overline{\nu}_k$ of a 10 MHz quartz oscillator compared to a Cesium clock. The sampling period is 10 s and the integration time of each frequency measurement is also 10 s (sampling without dead time).

In order to obtain dimensionless $\overline{y}_k$ samples, we must subtract the nominal frequency ν_0 (10 MHz) from the frequency measurements and normalize by ν_0:

$$\overline{y}_k = \frac{\overline{\nu}_k - \nu_0}{\nu_0}. \tag{4}$$

B. Variance analysis

[Fig. 2 about here.]

Figure 2 is a log-log plot of the Allan variance of the quartz $\overline{y}_k$ samples versus the integration time τ. A least squares fit of these variance measurements (solid line), weighted by their uncertainties, detects only two types of noise: a white noise and an f^{-2} noise. The corresponding noise level estimations are:

$$h_0 = (2.2 \pm 0.4) \cdot 10^{-19} \mathrm{s} \qquad \text{at } 1\sigma \text{ (68\% confidence)}$$

$$h_{-2} = (2.3 \pm 0.6) \cdot 10^{-26} \mathrm{s}^{-1} \qquad \text{at } 1\sigma \text{ (68\% confidence)}$$

(for the assessment of the h_α noise levels and their uncertainties, we used the multivariance method described in [2]).

However, for large τ values (corresponding to low frequencies), the variance measurements move away from the fitted curve. Two explanations are possible:

- instead of an f^{-2} noise, there is a noise whose non-integer exponent is comprised between -2 and -3 ;
- since the uncertainty domains of the variance measurements contain the fitted curve, this apparent divergence may be due to a statistical effect.

In order to choose between these two explanations, we decided to estimate the probability density of the exponent with a Bayesian approach.

III. Bayesian Approach

A. Principle

A.1 The Bayes theorem

The goal of all measurement is the estimation of an unknown quantity θ from measurements ξ, i.e. determining $p(\theta|\xi)$, the density of probability of the quantity θ knowing the measurements ξ. The Bayesian theory is based on the following equality [3]:

$$p(\theta|\xi) \propto p(\xi|\theta)\pi(\theta) \tag{5}$$

where $p(\xi|\theta)$ is the distribution of the measurements ξ for a fixed value of the quantity θ and $\pi(\theta)$ is the *a priori* density of probability of the quantity θ, i.e. before performing any measurement.

The determination of this a priori density, called the *prior*, is generally one of the main difficulties of this approach (particularly in the case of total lack of knowledge!).

A.2 Properties of a prior

1. If several measurements $\{\xi_1, \ldots, \xi_N\}$ are performed in order to estimate a quantity θ, the result of the estimation must be independent of the ordering of the measurements.
2. The result must be invariant under reparametrization of the problem: if a bijection transforms θ into $\theta' = f(\theta)$, the prior must satisfy:

   $$\pi(\theta') = \pi(\theta)\,|J(\theta)|^{-1} \tag{6}$$

   where $J(\theta)$ is the Jacobian of the transformation.
3. It must be possible to normalize the posterior distribution to one.

For example, Laplace considered that a uniform prior ($\pi(\theta) = 1$) should be chosen if no information is available (1812). Unfortunately, this choice is not invariant under reparametrization (condition 2) and does not always lead to a proper posterior distribution (condition 3).

Jeffrey proposed to use the square root of the determinant of the Fisher information matrix as a prior which respects these 3 conditions (1961). The Fisher information matrix $I(\theta)$ is defined by:

$$[I(\theta)]_{ij} = -\left\langle \frac{\partial^2 \ln(p(\xi|\theta)}{\partial\theta_i \partial\theta_j} \right\rangle . \tag{7}$$

See [3] and [4] for a more thorough discussion about the choice of a prior.

In this paper we use the Bernardo-Berger prior which is an extension of the Jeffrey's prior when the parameters may be divided into nuisance parameters (here h_α) and informative parameters (here α) [3].

B. Spectral density and covariance matrix

Let us define the vector y whose components are the N $\overline{y}_k$ samples. We assume that y is a Gaussian vector. The probability distribution of y is:

$$p(y) = \frac{\exp\left(-\frac{y^t C^{-1} y}{2}\right)}{(2\pi)^{\frac{N}{2}} \sqrt{|C|}} \tag{8}$$

where C is the covariance matrix. Since $S_y(f)$ is the Fourier transform of $\mathcal{R}_y(\tau)$, the autocorrelation function of the frequency deviation, the general term of C is:

$$C_{ij} = 2\int_{f_l}^{f_h} S_y(f) \cos\left(2\pi f(t_i - t_j)\right) df \tag{9}$$

where f_l and f_h are respectively the low cut-off frequency and the high cut-off frequency.

Equation (9) reveals the key role played by the spectral density of the noise in the expected fluctuation. We will present a general method for estimating the parameters of the model for $S_y(f)$.

C. Assumed model for the spectral density

We assume that the sequence of frequency measurements contains a white noise, whose level h_0 is known, and a red noise whose level h_α is unknown. This yields the following model for $S_y(f)$:

$$S_y(f) = h_0 + h_\alpha \cdot f^\alpha \qquad \text{with } -3 \le \alpha \le -1 \tag{10}$$

where $h_0 = 2.2 \cdot 10^{-19}$s, h_α and α are the unknown parameters.

Denote by y_n the noise vector normalized by the square root of the white level h_0:

$$y_n = \frac{y}{\sqrt{h_0}} = \frac{y}{\sigma_w \sqrt{2\tau_0}} \tag{11}$$

where σ_w is the actual standard deviation of the white noise (σ_w is easily estimated from high sampling rate frequency measurements) and τ_0 the sampling period ($\tau_0 = 10$ s in this case).

This normalized vector y_n may be rewritten as the sum of a white noise component y_w and a red noise component y_r:

$$y_n = y_w + y_r. \tag{12}$$

The corresponding normalized PSD $S_n(f)$ is:

$$S_n(f) = 1 + H \cdot u_\alpha \cdot f^\alpha \tag{13}$$

where H is the normalized red noise level and u_α is an amplitude factor whose meaning will be explained below (see equation (26)).

D. Statistical model

The part of the spectral density due to the red noise y_r may be written:

$$S_r(f) = H \cdot u_\alpha \cdot f^\alpha. \tag{14}$$

We use the Bernardo-Berger analysis [3], [5] for estimating the unknown parameter $\theta = (\alpha, H)$.

D.1 Construction of the estimators

Let us introduce the orthonormal basis of $\Re^N$, $\{p_0, \ldots, p_j, \ldots, p_{N-1}\}$, defined such as the i^{th} component of p_j is:

$$p_{ij} = \widetilde{p}_j(t_i) \tag{15}$$

where t_i is the time of the i^{th} frequency measurement and $\tilde{p}_j(t)$ is a polynomial of degree j, satisfying the orthonormality condition [6]:

$$\sum_{i=0}^{N-1} \tilde{p}_j(t_i) \cdot \tilde{p}_k(t_i) = \delta_{jk}. \tag{16}$$

It can be shown than the expectation of the square of the scalar product of a vector p_j by the noise vector y is an estimate of the spectral density $S_y(f)$ for a given frequency f_j [7]. Denote by ξ_j the scalar product of the estimator p_j and the normalized noise vector y_n:

$$\xi_j = p_j \cdot y_n. \tag{17}$$

Thus, ξ_j may be considered as an estimate of the noise spectrum for the frequency f_j.

Practically, we limited to 16 the number of estimators p_j (from degrees 0 to 15) for limiting the computation time and because the highest degrees, estimating the high frequency content, are less informative for a red noise.

Moreover, in order to ensure convergence for very low frequencies (even if the low cut-off frequency tends towards 0), the polynomials must satisfy the moment condition [6], [7]: the minimum degree j_{min} of a polynomial to ensure convergence up to an exponent α is:

$$j_{min} \geq \frac{-\alpha - 1}{2}. \tag{18}$$

Since we have assumed $\alpha \geq -3$, the first 2 estimators (p_0 and p_1) must be removed. Thus we have $n = 14$ estimators $\{p_2, \ldots, p_{15}\}$ and $n = 14$ estimates $\{\xi_2, \ldots, \xi_{15}\}$.

D.2 Construction of the priors

The covariance matrix defined in relationship (9) is an ensemble average of the different estimate products over an infinite number of realizations of this process:

$$C = \left\langle \xi \cdot \xi^t \right\rangle \tag{19}$$

$$C_{ij} = \left\langle \xi_i \cdot \xi_j \right\rangle. \tag{20}$$

As for the noise vector y_n, the estimate vector ξ may be split into two terms, according to equations (12) and (17):

$$\xi = \xi_w + \xi_r. \tag{21}$$

The covariance matrix may also be split:

$$C = \left\langle \xi_w \xi_w^t \right\rangle + \left\langle \xi_r \xi_r^t \right\rangle = I_n + H \cdot u_\alpha \cdot V(\alpha) \tag{22}$$

where I_n is the identity matrix in $\Re^n$. The general term of the matrix $V(\alpha)$ is:

$$[V(\alpha)]_{ij} = 2 \int_{1/T}^{f_h} f^\alpha P_i(f) \cdot \overline{P_j(f)} df \tag{23}$$

where $P_i(f)$ and $\overline{P_j(f)}$ respectively denote the discrete Fourier transform of $p_i(t)$ and the complex conjuguate of the discrete Fourier transform of $p_j(t)$:

$$P_i(f) = \sum_{k=0}^{N-1} p_i(t_k) \exp(-i2\pi f t_k). \tag{24}$$

The high cut-off frequency f_h in (23) is the Nyquist frequency and T is the total duration of the sequence.

Let $e_i(\alpha)$ denote the i^{th} eigenvector of $V(\alpha)$ and $\gamma_i(\alpha)$ its i^{th} eigenvalue ($i \in \{0, \ldots, n-1\}$). The averaged quadratic norm of the estimate vector ξ is:

$$\left\langle \|\xi\|^2 \right\rangle = n + H \cdot u_\alpha \sum_{i=0}^{n-1} \gamma_i(\alpha) = \left\langle \|\xi_w\|^2 \right\rangle + \left\langle \|\xi_r\|^2 \right\rangle. \tag{25}$$

The factor u_α is chosen in such a way that, for $H = 1$, the averaged quadratic norms $\left\langle \|\xi_w\|^2 \right\rangle$ and $\left\langle \|\xi_r\|^2 \right\rangle$ are equal. It follows:

$$u_\alpha = \frac{n}{\sum_{i=0}^{n-1} \gamma_i(\alpha)}. \tag{26}$$

The direct problem is now solved since ξ is a vector of $\Re^n$ with a probability distribution given the parameter θ equal to:

$$p(\xi|\theta) = \frac{1}{(2\pi)^{n/2}\sqrt{|C|}} \exp(-\frac{1}{2}\xi^t C^{-1} \xi). \tag{27}$$

Denoting "Tr(M)" the trace of a matrix M and X the matrix defined as:

$$X = u_\alpha \cdot V(\alpha) \tag{28}$$

the Fisher information matrix $I(\theta)$ is (see [5]):

$$I(\theta) = \frac{1}{2}\begin{pmatrix} H^2 \mathrm{Tr}\left(C^{-1}\frac{dX}{d\alpha}C^{-1}\frac{dX}{d\alpha}\right) & H\mathrm{Tr}\left(C^{-1}XC^{-1}\frac{dX}{d\alpha}\right) \\ H\mathrm{Tr}\left(C^{-1}XC^{-1}\frac{dX}{d\alpha}\right) & \mathrm{Tr}\left(C^{-1}XC^{-1}X\right) \end{pmatrix}. \tag{29}$$

The Jeffrey's prior $\pi(\theta)$ is defined as:

$$\pi(\theta) = \sqrt{|I(\theta)|}. \tag{30}$$

The parameter θ is a two-dimensional parameter composed of the exponent parameter α and of the amplitude parameter H. Since we are mostly interested in α, H is called a nuisance parameter.

In presence of nuisance parameter, Bernardo and Berger suggested that α should first be fixed and the conditional prior $\pi(H|\alpha)$ computed for that value. The full prior is then:

$$\pi(\theta) = \pi(H|\alpha) \cdot \pi(\alpha). \tag{31}$$

The conditional prior $\pi(H|\alpha)$ is given by:

$$\pi(H|\alpha) = \sqrt{|[I(\theta)]_{22}|} \tag{32}$$

where $[I(\theta)]_{11}$, $[I(\theta)]_{12} = [I(\theta)]_{21}$ and $[I(\theta)]_{22}$ are the elements of the Fisher information matrix $I(\theta)$.

The prior for α may be computed as:

$$\pi(\alpha) = c \cdot \exp\left(\int \pi(H|\alpha) \ln |k(\alpha, H)|^{1/2} \, dH\right) \tag{33}$$

where c is a normalization coefficient ensuring that $\int \pi(\alpha) d\alpha = 1$ and:

$$k(\alpha, H) = [I(\theta)]_{11} - \frac{[I(\theta)]_{12}^2}{[I(\theta)]_{11}}. \tag{34}$$

This prior is plotted in Figure 3.

[Fig. 3 about here.]

D.3 Construction of the posteriors

According to the Bayes theorem, the posterior probability distribution is given by:

$$p(\theta|\xi) = \frac{p(\xi|\theta)\pi(\theta)}{\int p(\xi|\theta')\pi(\theta')d\theta'}. \qquad (35)$$

The posterior probability distribution for α is then given by:

$$p(\alpha|\xi) = \frac{\int p(\xi|\alpha, H)\pi(H|\alpha)\pi(\alpha)dH}{\int\int p(\xi|\alpha', H')\pi(H'|\alpha')\pi(\alpha')dH'd\alpha'}. \qquad (36)$$

IV. Results and Discussion

A. *Compatibility with an integer exponent*

[Fig. 4 about here.]

Figure 4 shows the posterior probability distribution for the exponent α of the red noise using the Bernardo-Berger prior.

The median value obtained for the exponent, just as for the maximum of the distribution, is $\alpha = -2.2$ (the maximum of $p(\alpha|\xi)$ is -2.18, the median value is -2.19 and the mean value is -2.21). The estimation result is :

$$\alpha = -2.2 \pm 0.5 \qquad \text{at } 1\sigma \text{ } (68\% \text{ confidence}).$$

Therefore $\alpha = -2$ is fully compatible with this posterior distribution. Thus we may conclude that the apparent divergence between the variance measurements and the fitted curve in Figure 2 is probably due to a statistical bias of the data. The spectral density $S_y(f)$ is then compatible with the following model:

$$S_y(f) = h_0 + h_{-2}f^{-2}. \qquad (37)$$

B. *Noise level estimation*

[Fig. 5 about here.]

Selecting an exponent value $\alpha = -2$, we obtained the posterior probability distribution plotted in Figure 5. As in the variance analysis, we chose a confidence interval of 68% (16% probability that h_{-2} is smaller than the low bound and 16% probability that h_{-2} is greater than the high bound):

$$h_{-2} = \left(2.5 \begin{array}{cc} + & 2.6 \\ - & 1.1 \end{array}\right) \cdot 10^{-26}\mathrm{s}^{-1} \qquad \text{at } 1\sigma \text{ (68\% confidence)}$$

The difference between the median value ($h_{-2} = 2.47 \cdot 10^{-26}\mathrm{s}^{-1}$) and the variance analysis value ($h_{-2} = 2.29 \cdot 10^{-26}\mathrm{s}^{-1}$) is only 7%.

However, the confidence intervals given by these two methods are noticeably different. The main difference concerns the symmetry of the variance analysis interval: in this case, we do not take into account the fact that the noise levels are positive, whereas the prior of the Bayesian approach is null for negative values of h_{-2}.

Moreover, the variance analysis interval seems to be a bit underestimated relative to the Bayesian approach taken here.

V. Conclusion

The variance analysis is an useful tool for a quick estimation of the noise levels in the output signal of an oscillator. However, a negative estimate of a noise level may occur. Generally, in this case, this value is rejected and the corresponding noise level is assumed to be null. On the other hand, although the Bayesian method is a bit more difficult, it takes properly into account the a priori information, and gives a more reliable estimation of these noise levels and especially of their confidence intervals.

However, the main advantage of the Bayesian method concerns the verification of the validity of the power law model of spectral density. Each time the model is suspected, such an approach should be used in order to estimate the exponent of the power law. In particular, this method should be very interesting for the study of the f^{-1} and f^{+1} noise, whose origin remains mysterious [8].

REFERENCES

[1] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leesson, T. E. McCuningal, J. A. Mullen, W. L. Smith, R. L. Sydnor, R. Vessot, and G. M. R. Winkler, "Characterization of frequency stability," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, pp. 105–120, 1971.

[2] F. Vernotte, E. Lantz, J. Groslambert, and J.J. Gagnepain, "Oscillator noise analysis : multi-variance measurement," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-42, no. 2, pp. 342–350, Avril 1993.

[3] J. Bernardo and A. Smith, *Bayesian theory*, Wiley & Son, 1993.

[4] C. Robert, *The Bayesian Choice*, Springer-Verlag.

[5] G. Zalamansky and C. Robert, "Bayesian analysis with application to the timing residuals of a pulsar," in *MaxEnt'98*, Garching, Germany, July 1998.

[6] J. E. Deeter and P. E. Boynton, "Techniques for the estimation of red power spectra. i. context and methodology," *The Astrophysical Journal*, vol. 261, pp. 337–350, October 1982.

[7] F. Vernotte, G. Zalamansky, M. McHugh, and E. Lantz, "Cut-off frequencies and noise power law model of spectral density : adaptation of the multi-variance method for irregularly spaced timing data using the lowest mode estimator approach.," *IEEE Transactions on Ultrasonics, ferroelectrics, and Frequency*, vol. 43, no. 3, pp. 403–409, 1996.

[8] M. Planat, V. Giordano, G. Marianneau, F. Vernotte, M. Mourey, C. Eckert, and J. Miehe, "Is the frequency noise of an oscillator of deterministic origin ?," *IEEE Transactions on Ultrasonics, ferroelectrics, and Frequency*, vol. IM-43, no. 2, pp. 326–330, 1996.

LIST OF FIGURES

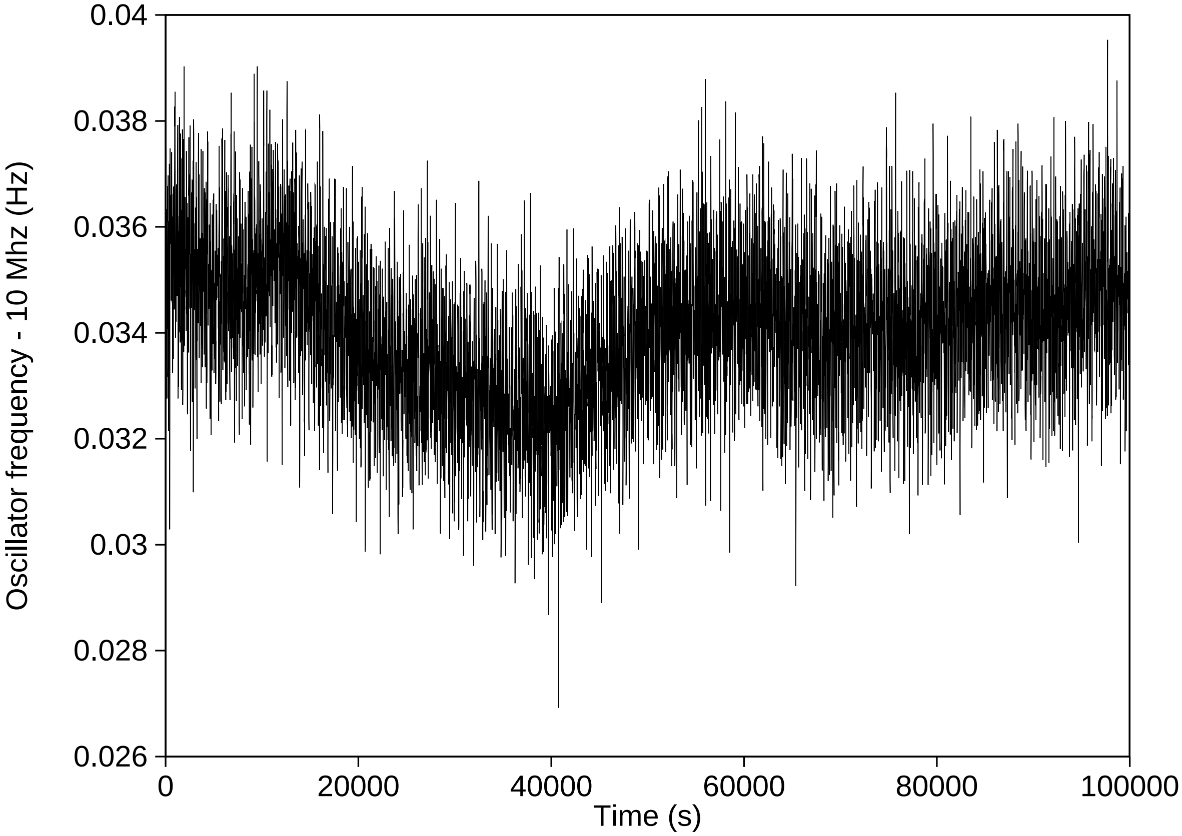

Fig. 1. Sequence of frequency measurements

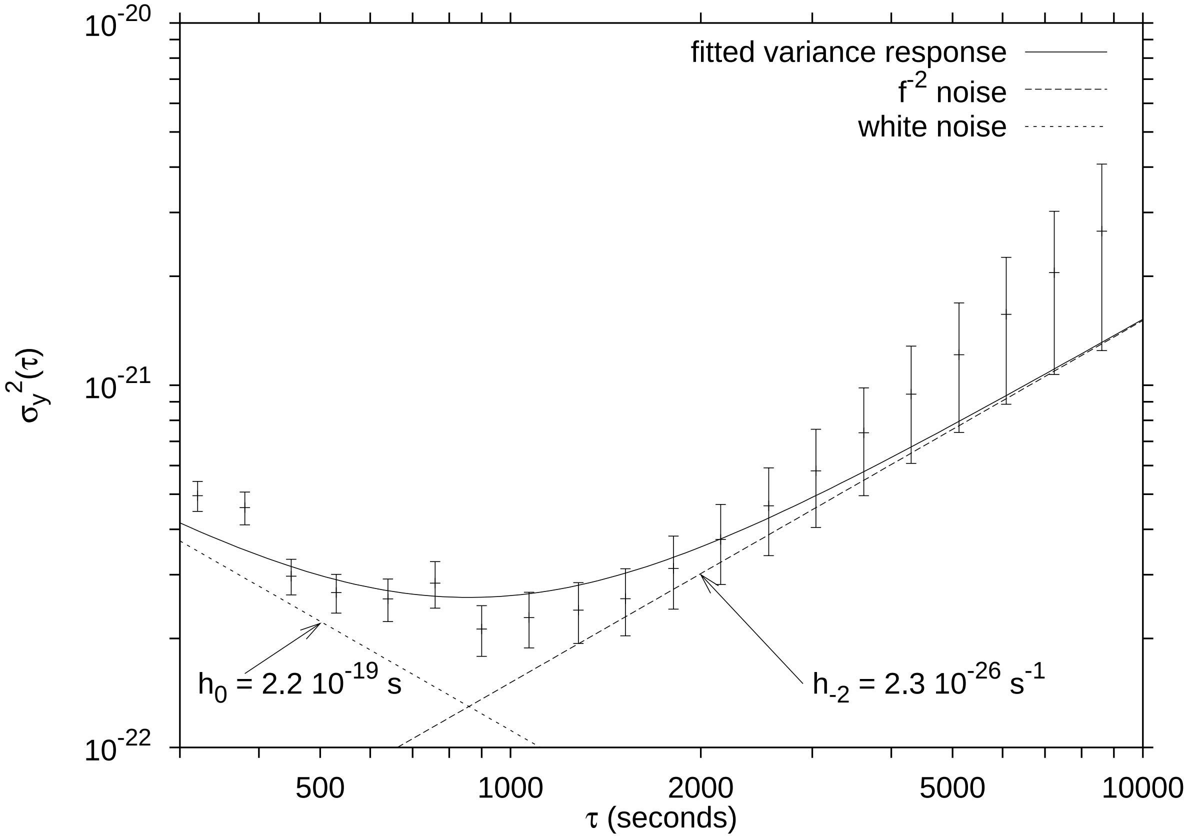

Fig. 2. Allan variance of the sequence of frequency measurements

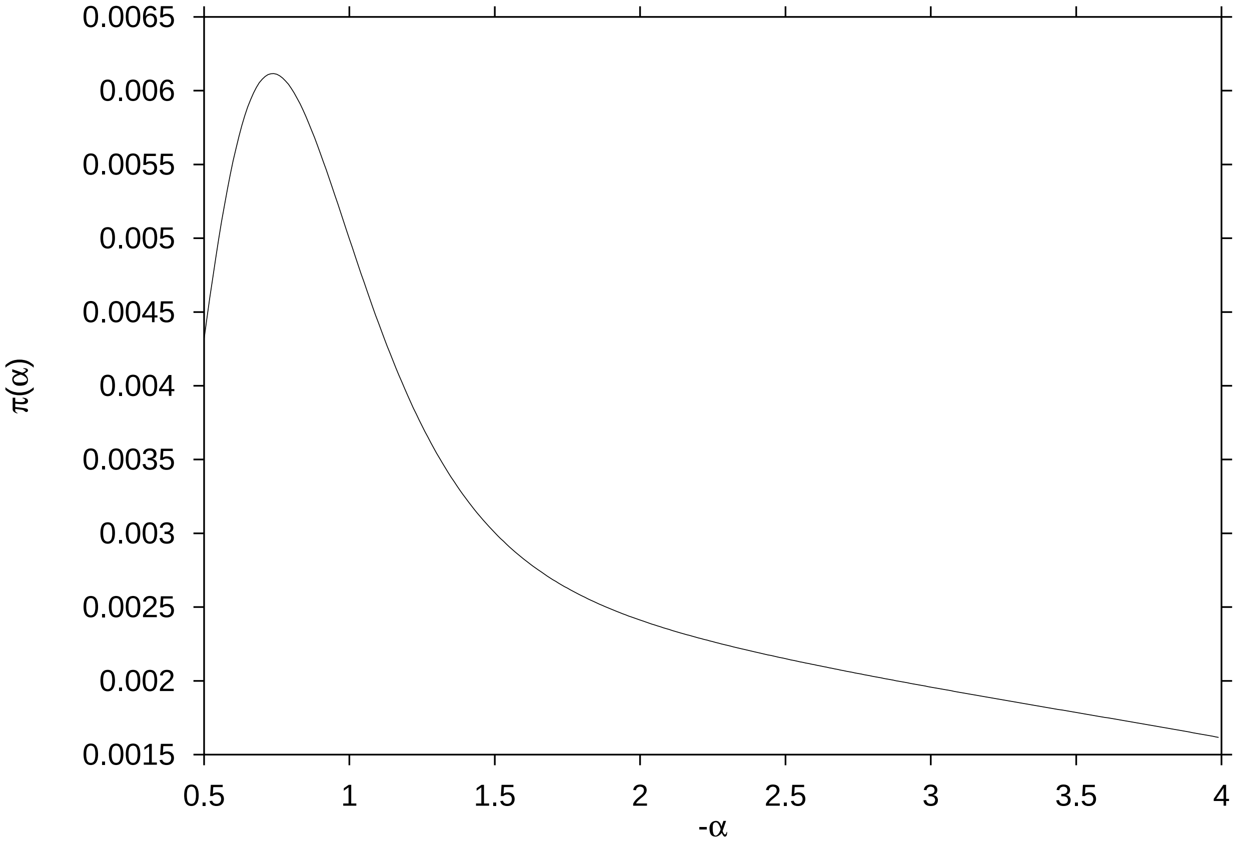

Fig. 3. Reference prior for the power α

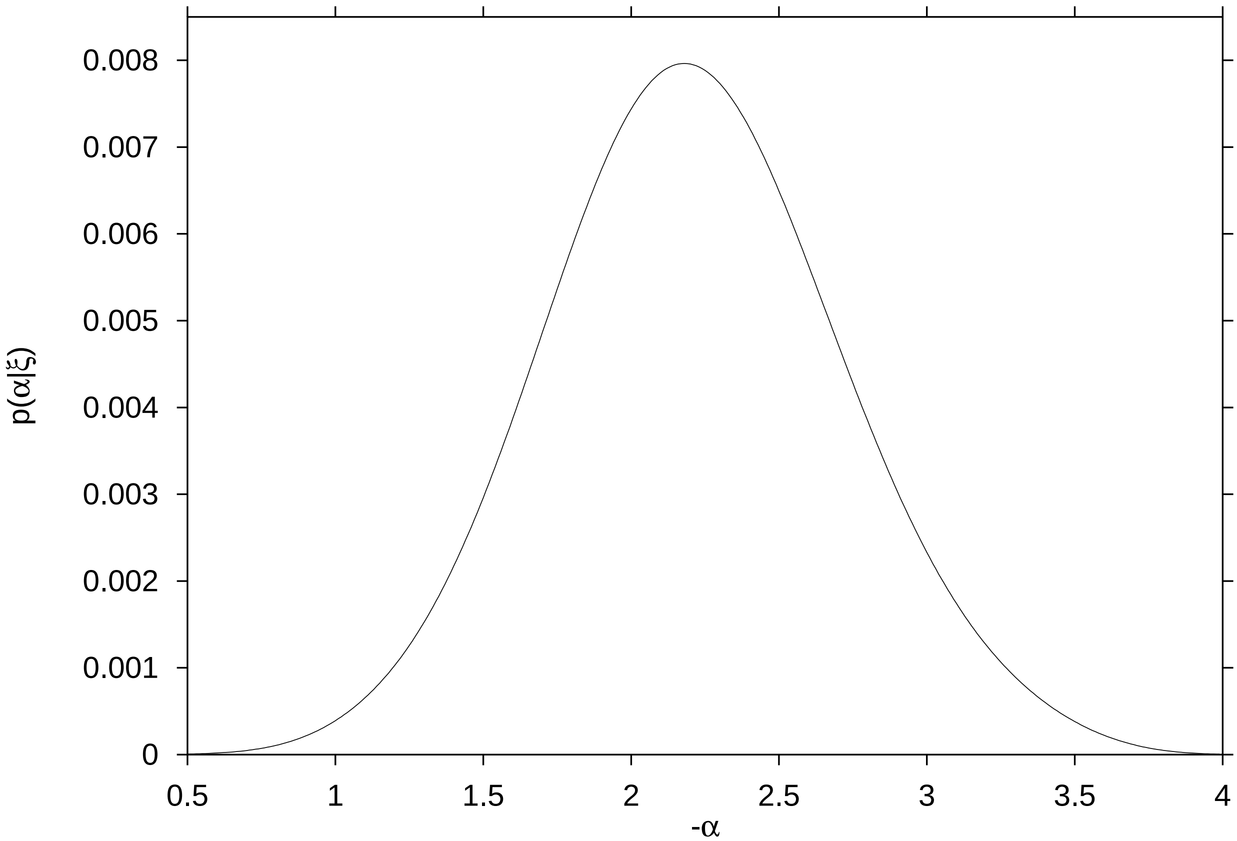

Fig. 4. Posterior probability density for the power α

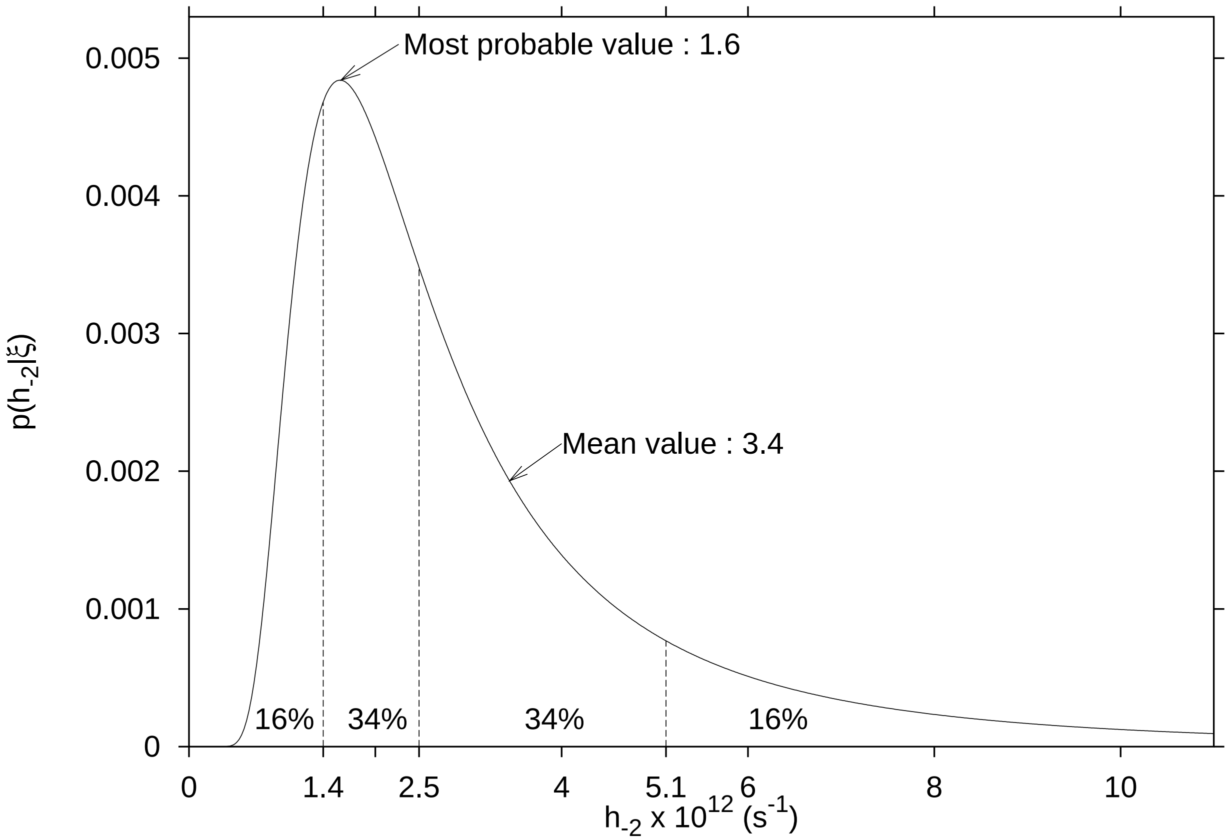

Fig. 5. Posterior probability density for the noise level h_{-2}