Class. Quantum Grav. **23** (2006) S829–S836 doi:10.1088/0264-9381/23/19/S21

Normal/independent noise in VIRGO data

F Acernese[1], P Amico[2], M Alshourbagy[3], F Antonucci[4], S Aoudia[5], S Avino[1], D Babusci[6], G Ballardin[7], F Barone[1], L Barsotti[3], M Barsuglia[8], F Beauville[9], S Bigotta[3], S Birindelli[3], M A Bizouard[8], C Boccara[10], F Bondu[5], L Bosi[2], C Bradaschia[3], S Braccini[3], A Brillet[5], V Brisson[8], L Brocco[4], D Buskulic[9], E Calloni[1], E Campagna[11], F Cavalier[8], R Cavalieri[7], G Cella[3], E Cesarini[11], E Chassande-Mottin[5], C Corda[3], F Cottone[2], A-C Clapson[8], F Cleva[5], J-P Coulon[5], E Cuoco[7], A Dari[2], V Dattilo[7], M Davier[8], R De Rosa[1], L Di Fiore[1], A Di Virgilio[3], B Dujardin[5], A Eleuteri[1], D Enard[7], I Ferrante[3], F Fidecaro[3], I Fiori[3], R Flaminio[7,9], J-D Fournier[5], O Francois[7], S Frasca[4], F Frasconi[3,7], A Freise[7], L Gammaitoni[2], F Garufi[1], A Gennai[3], A Giazotto[3], G Giordano[6], L Giordano[1], R Gouaty[9], D Grosjean[9], G Guidi[11], S Hebri[7], H Heitmann[5], P Hello[8], L Holloway[7], S Karkar[9], S Kreckelbergh[8], P La Penna[7], M Laval[5], N Leroy[8], N Letendre[9], M Lorenzini[11], V Loriette[10], M Loupias[7], G Losurdo[11], J-M Mackowski[12], E Majorana[4], C N Man[5], M Mantovani[3], F Marchesoni[2], F Marion[9], J Marque[7], F Martelli[11], A Masserot[9], M Mazzoni[11], L Milano[1], C Moins[7], J Moreau[10], N Morgado[12], B Mours[9], A Pai[4], C Palomba[4], F Paoletti[3,7], S Pardi[1], A Pasqualetti[7], R Passaquieti[3], D Passuello[3], B Perniola[11], F Piergiovanni[11], L Pinard[12], R Poggiani[3], M Punturo[2], P Puppo[4], K Qipiani[1], P Rapagnani[4], V Reita[10], A Remillieux[12], F Ricci[4], I Ricciardi[1], P Ruggi[7], G Russo[1], S Solimeno[1], A Spallicci[5], R Stanga[11], R Taddei[7], M Tonelli[3], A Toncelli[3], E Tournefier[9], F Travasso[2], G Vajente[3], D Verkindt[9], F Vetrano[11], A Viceré[11], J-Y Vinet[5], H Vocca[2], M Yvert[9] and Z Zhang[7]

[1] INFN, sezione di Napoli and/or Università di Napoli 'Federico II' Complesso, Universitario di Monte SAngelo, and/or Università di Salerno, Fisciano (Sa), Italy
[2] INFN, Sezione di Perugia and/or Università di Perugia, Perugia, Italy
[3] INFN, Sezione di Pisa and/or Università di Pisa, Pisa, Italy
[4] INFN, Sezione di Roma and/or Università 'La Sapienza', Roma, Italy
[5] Departement Artemis—Observatoire de la Côte d'Azur, BP 42209 06304 Nice, Cedex 4, France
[6] INFN, Laboratori Nazionali di Frascati, Frascati (Rm), Italy
[7] European Gravitational Observatory (EGO), Cascina (Pi), Italy
[8] Laboratoire de l'Accélérateur Linéaire (LAL), IN2P3/CNRS-Université de Paris-Sud, Orsay, France
[9] Laboratoire d'Annecy-le-Vieux de Physique des Particules, Annecy-le-Vieux, France
[10] ESPCI, Paris, France
[11] INFN, Sezione di Firenze/Urbino, Sesto Fiorentino, and/or Università di Firenze, and/or Università di Urbino, Italy
[12] LMA, Villeurbanne, Lyon, France

E-mail: Fabio.Garufi@na.infn.it

Received 29 March 2006
Published 15 September 2006
Online at stacks.iop.org/CQG/23/S829

Abstract
The analysis of data taken during the C7 VIRGO commissioning run showed strong deviations from Gaussian noise. In this work, we explore a family of distributions, derived from the hypothesis that heavy tails are an effect of a particular kind of nonstationarity, heterocedasticity (i.e. nonuniform variance), that appear to fit VIRGO noise better than a model based on the assumption of Gaussian noise. To estimate the parameters of the noise process (including the heterogeneous variance) we derived an expectation-maximization algorithm. We show the consequences of non-Gaussianity on the fitting of autoregressive filters and on the derivation of test statistics for matched filter operation. Finally, we apply the new noise model to the fitting of an autoregressive filter for whitening of data.

PACS numbers: 04.80.Nn, 07.05.Kf, 07.60.Ly, 05.40.Ca

(Some figures in this article are in colour only in the electronic version)

1. Introduction

During the analysis of VIRGO data, one of the procedures applied to data before the detection algorithms is the whitening filter [1]. The filter currently used is based on an autoregressive (AR) parametrization for the noise process and the residuals of the AR fit are called the whitened data. If the noise were Gaussian, then also the residual (i.e. the whitened data) should be Gaussian. What is observed in VIRGO commissioning run C7 data, instead, is that the distributions of residuals are not Gaussian and show, in most cases, large tails and high peaks, that could be due to non-stationary noise. It appears that Student-like distributions better fit the data.

It is well known that if data are extracted from two or more different Gaussian distributions with different variances, the resulting sum process shows heavy tails, and can be used for outlier modelling. Since the causes of the deviations from the Gaussian distribution are not known, a more general mechanism must be investigated. This work aims at generalizing the case of a known number of variances to an indeterminate number, which can be modelled by a generalized Student-t distribution.

2. The Gaussian model

Autoregressive modelling of the h-reconstructed data stream[13] shows that the whitened data are not Gaussian-distributed, as expected if the detector noise were Gaussian, which is the hypothesis underlying the whitening procedures applied in VIRGO. Strong deviations from Gaussianity can be observed in normal probability plots, with very fat tails and leptokurticity (lower kurtosis with respect to Gaussian) accounting for most of the deviation.

Let us consider an AR(M) model:

$$y(n) = \sum_{k=1}^{M} a_k y(n-k) + \epsilon_n(\theta) \tag{1}$$

[13] h-reconstructed data are time series calibrated in the time domain, obtained processing interferometer (photodiode) output and control signals with a series of time-domain filters to give the strain amplitude $h(t)$.

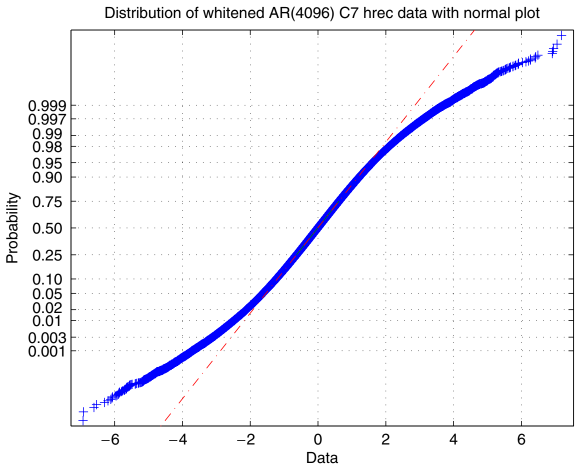

Figure 1. AR(4096)-whitened and normalized C7 data segment probability plot (5 min, h-reconstructed channel sampled at 20 kHz, GPS=810777660, Science Mode) versus probability plot of the sample under the hypothesis of Gaussian noise.

where $\epsilon_n(\theta)$ is a random variable characterized by a parameter vector θ and a_k are the AR coefficients. The common assumption in the application of whitening algorithms is that ϵ is a normal distributed random variable: $\epsilon_n \sim \mathcal{N}(a_0, \sigma^2), \theta = (a_0, \sigma)$. In figure 1 we show the probability plot of VIRGO C7 calibrated data whitened with the Burg algorithms [2]. This plot points out the non-Gaussianity of the data.

The application of fits based on the Gaussianity of data to non-Gaussian data is clearly sub-optimal and the methods based on linear predictors result biased. In the following we find that the assumption that data come from several Gaussian distributions with different variances (heterocedasticity), results in a distribution of the extracted data capable of reproducing the observed non-Gaussian tails and higher peaks.

3. A normal/independent model

Let us consider the family of normal/independent (N/I) random variates [3]:

$$Y|U \sim \mathcal{N}(\mu(\theta), \sigma^2/U), \tag{2}$$

where U is a random variable with nonnegative support. The above equation expresses Y as a *conditionally Gaussian* process, whose variance changes locally. Every draw from the random variable Y implies a draw from the random variable U, so that with every observation drawn from Y we may associate a different variance. Note that to allow regression modelling, the mean μ is dependent on the parameter vector θ. The density of the conditionally Gaussian process will be denoted by $p(y|u)$.

Since the variance of Y depends on a random variable U, fitting observations from Y would require estimation in an infinite-dimensional space. Such problems are generally ill-posed and quite complex. A viable alternative is to fix a parametrized distribution for U, $p(u)$, and evaluate the joint distribution of y and u:

$$p(y, u) = p(y|u)p(u). \tag{3}$$

To get the unconditional density of y, we must marginalize with respect to u:

$$p(y) = \int p(y, u)\,\mathrm{d}u = \int p(y|u)p(u)\,\mathrm{d}u. \tag{4}$$

The resulting unconditional density will no longer be Gaussian. Equation (4) clearly shows that Y can be represented as a *hidden mixture* of Gaussian processes, i.e. Y results from the superposition of a very large (formally infinite) number of Gaussian processes with the same mean and different variances.

To choose a suitable parametric form for U, we should take into account some practical requirements:

- it should have as few parameters as possible;
- the integral in equation (4) should be tractable;
- the expression for $p(y)$ should be a generalization of a Gaussian density, so that the same model could be used also for strictly Gaussian observations.

Clearly, any choice which meets the above requirements but does not correctly fit observations, can be discarded *ex post*.

One possible candidate is given by the *chi-square* distribution:

$$p(u) = \chi^2_\nu/\nu \tag{5}$$

where the parameter ν is usually called *degrees of freedom*, although in our setting it can be any real positive number.

It can be shown that integration of equation (4) gives:

$$p(y) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)(\nu\sigma^2\pi)^{1/2}}\left(1 + \frac{(y-\mu(\theta))^2}{\nu\sigma^2}\right)^{-(\nu+1)/2}, \tag{6}$$

which is called a $t(\mu(\theta), \sigma^2, \nu)$ (or Student) density, $\Gamma(\nu)$ being the Euler gamma function of parameter ν.

Furthermore, we have

$$\lim_{\nu\to+\infty} t(\mu(\theta), \sigma^2, \nu) = \mathcal{N}(\mu(\theta), \sigma^2). \tag{7}$$

Numerically, the two distributions can be practically considered equal when $\nu \geqslant 150$.

The mean and variance of Y are, respectively,

$$\mathbb{E}[Y] = \mu(\theta) \qquad \text{for} \quad \nu > 1, \tag{8}$$

$$\mathbb{E}[(Y-\mu(\theta))^2] = \frac{\nu\sigma^2}{\nu-2} \qquad \text{for} \quad \nu > 2. \tag{9}$$

It is worth mentioning that the more common ϵ-contamination modelling approach of summing two Gaussian random variates with weights ϵ and $1-\epsilon$ is a particular case of equation (4), when the measure of U is atomic, and the distribution collapses to two point masses. Such a model, while formally simpler, may not be flexible enough.

As we shall see, the use of the distribution of equation (6) to build an autoregressive filter to whiten the data, is not trivial and computationally very expensive. Work is in progress to find a less expensive fitting algorithm.

4. Implications of *t*-distributed noise

In the following we shall examine the case of AR model fitting and detection statistics when we apply methods designed for Gaussian noise on data that follow a t distribution.

4.1. AR model fitting

Given a data stream y of length N, let us define the data matrix X and the regression vector v:

$$X_{k,:} = [y(k-1), y(k-2), \ldots, y(k-M)], \qquad v_k = y(k), \tag{10}$$

where $X_{k,:}$ is the kth row of X.

In the following, we shall consider an AR(M) model, expressed as a linear predictor of the form

$$\mu(\theta) = \theta' X_{k,:} \tag{11}$$

where the prime symbol denotes the transposition operation. We shall specialize the coefficient vector θ for the case of Gaussian and t noise models.

The best linear predictor under the Gaussian noise model assumption is given by the solution of a least-squares problem [2]:

$$\theta_G = (X'X)^{-1} X' v. \tag{12}$$

Fitting a linear predictor under the t noise model instead involves the solution of a weighted least-squares problem [4]:

$$\theta_t = (X'WX)^{-1} X' W v. \tag{13}$$

The elements of the diagonal weight matrix W are estimates of the inverse variances u of the samples.

By using the above expressions, the solution we get by applying algorithms designed for Gaussian noise when the noise is t-distributed can then be written as

$$\theta_G = \theta_t + \zeta, \qquad \zeta = (X'X)^{-1} X' v - (X'WX)^{-1} X' W v. \tag{14}$$

The Gaussian solution is therefore *biased* with respect to the optimal linear solution θ_t. It should be noted that the effects of the bias cannot be estimated *a priori* since the weight matrix W is unknown. Estimation of the W matrix is a complex task, and is part of fitting the t model. The practical significance of the bias is that it produces wrong estimates of the noise parameters.

4.2. Detection statistic

Gravitational wave detection algorithms in the current literature, like the matched filter, assume that the instrumental noise is Gaussian distributed. This hypothesis implies that the signal-to-noise ratio, used as the detection statistic, is distributed as a χ^2 random variable. It is possible to show that under the hypothesis of a t-distributed noise model, the detection statistic follows an F distribution. Thus we must be careful in the estimation of false alarm and detection probability, since they depend on the signal-to-noise ratio distribution.

5. Parameter estimation

Estimation of parameters in a N/I model is not easy especially if we are interested in evaluation of the hidden scale factors u, since they cannot be obtained by maximum likelihood approaches. Furthermore, $\mu(\theta)$ and σ cannot be directly estimated, but some iterative process is required. It should be noted that direct maximum likelihood estimation of μ, σ and ν cannot even be attempted, since it can be shown [4] that the likelihood function in such cases is unbounded.

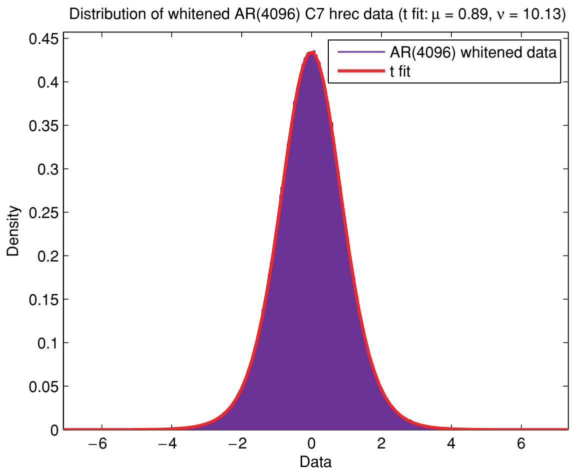

Figure 2. AR(4096)-whitened and normalized C7 data segment histogram (5 min, h-reconstructed channel, sampled at 20 kHz, GPS = 810777660, Science Mode) versus density of the sample under the Student-*t* hypothesis. The scale is $\sigma = 0.89$ and the degrees of freedom is $\nu = 10.13$.

We estimated the parameters of the model (including scale factors u, which are treated as missing data) by an iterative expectation-maximization (EM) algorithm [5]. At iteration n, for each observation $i \in \{1, \ldots, N\}$:

- E-step: compute the elements of the diagonal matrix W:

$$w_{i,n} = \mathbb{E}[u_i | y_i, \theta_n, \sigma_n] = \frac{\nu_n}{\nu_n + (y_i - \mu(\theta_n))^2/\sigma_n^2}. \tag{15}$$

- M-step: find the weighted least-squares solution of the problem:

$$\theta_{n+1} = \arg\min_{\theta} \sum_i w_{i,n}(y_i - \mu(\theta))^2; \tag{16}$$

then, calculate

$$\sigma_{n+1}^2 = \frac{1}{n} \sum_i w_{i,n}(y_i - \mu(\theta_{n+1}))^2 \tag{17}$$

$$\nu_{n+1} = \arg\max_{\nu} \left[\frac{N\nu}{2} \ln\left(\frac{\nu}{2}\right) - N \ln\Gamma\left(\frac{\nu}{2}\right) + \left(\frac{\nu}{2} - 1\right) \sum_i \mathbb{E}[\ln u_i | y_i, \theta_n, \sigma_n, \nu_n] - \frac{\nu}{2} \sum_i w_{i,n} \right]. \tag{18}$$

It can be shown [5] that the EM algorithm always monotonically converges in a finite number of steps.

6. Experiment

To assess the goodness of our model, we fitted an AR(4096) model by using the EM algorithm on the same data set shown in figure 1. The algorithm provided estimates for the AR parameters, the scale σ and the degrees of freedom ν. The results are shown in figures 2 and 3. We note in particular:

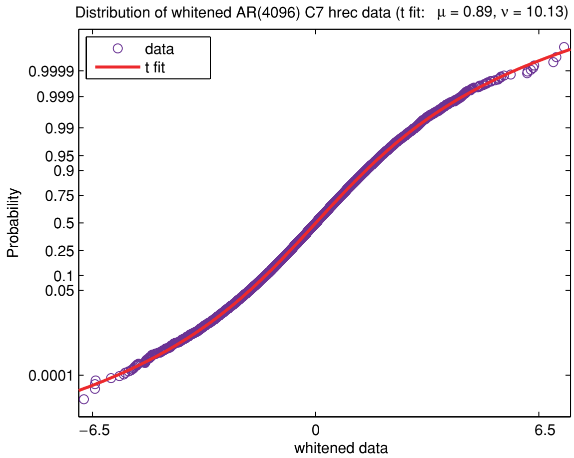

Figure 3. AR(4096)-whitened and normalized C7 data segment probability plot (5 min, h-reconstructed channel, sampled at 20 kHz, GPS = 810777660, Science Mode) versus probability plot of the sample under the Student-t hypothesis. The scale is $\sigma = 0.89$ and the degrees of freedom is $\nu = 10.13$.

(i) The estimated $\nu \approx 10$; for a Gaussian model to hold, ν should diverge (numerically, it should be larger than 150).
(ii) The likelihood of the data under the t error density is statistically different (and larger) than the likelihood of the same data under the Gaussian error density.
(iii) The heavy tails and high peak of the data are successfully modelled.

We point out that evaluation of the bias in the AR estimates cannot be done *a priori* since it depends on the W matrix (see equation (14)), which in turn depends on the mean, scale and degrees of freedom of the model (see section 5). However, from the above results, it is evident that the correction of the bias allows better data modelling.

7. Conclusions

In this paper we have shown how the heavy tails observed in whitened data streams from the C7 VIRGO commissioning run can be explained by a process with heterogeneous variance. We used the N/I model with χ^2-distributed inverse variance since it gives a numerically tractable problem, and it provides a good fit to the data. We have shown that common AR estimates and detection statistics based on the Gaussian noise model can give biased results when the noise is t-distributed. We derived an EM algorithm to fit model parameters, and have shown an application of the algorithm to VIRGO data of the C7 run.

At present work is in progress on the methods for the estimation of parameters and on exhaustive studies on the analysis of the detection using the information on the noise distribution.

References

[1] Cuoco E *et al* 2001 *Class. Quantum Grav.* **18** 1727–51
[2] Kay S 1998 *Fundamentals of Statistical Signal Processing, vol I Estimation Theory* (Englewood Cliffs, NJ: Prentice-Hall)

[3] Andrews D F and Mallows C L 1974 Scale mixtures of normal distributions *J. R. Stat. Soc.* B **36** 99–102

[4] 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

[5] Dempster A P, Laird N M and Rubin D B 1977 Maximum likelihood form incomplete data via the EM algorithm *J. R. Stat. Soc.* B **39** 1–38