

Towards a data and detector characterization robot for gravitational wave detectors

Soumya D Mohanty and Soma Mukherjee

Albert Einstein Institut, Max Planck Institut für Gravitationsphysik, Am Mühlenberg 1,
D-14476 Golm, Germany

E-mail: mohanty@aei-potsdam.mpg.de and soma@aei-potsdam.mpg.de

Received 2 October 2001, in final form 30 October 2001

Published 11 March 2002

Online at stacks.iop.org/CQG/19/1471

Abstract

A change of non-astrophysical origin in the detector state or in the statistical nature of data while an interferometer is in lock reflects an abnormality. The change can manifest itself in many forms: transients, drifts in noise power spectral density, change in cross correlation between channels, etc. We advance the idea of an algorithm for detecting such change points whose design goal is *reliable* performance, i.e. a known false alarm rate, even when statistically unmodelled data such as those from the physical environmental monitors are included. Reliability is important since following up on such change points could be fairly labour intensive. Such an algorithm need not be simply a collection of isolated independent monitors running in parallel. We present the first design steps towards building this detector characterization robot along with some preliminary results and outline some possibilities for the future.

PACS numbers: 0705K, 0480N, 9575W

1. Introduction

As far as their design is concerned, gravitational wave (GW) detectors are supposed to behave stably and data produced by them is supposed to be stationary. Thus, any *change* in the state of a detector when it is operating is an abnormality. Most such changes are likely to be caused by corresponding changes in the terrestrial environment of the detector and thus show up in the physical environmental monitor (PEM) channels. Changes in the environment recorded in a PEM channel which should have changed the detector state, but actually do not, are also interesting since they reflect a gap in our understanding of the interaction between the detector and its terrestrial environment.

One way to study the changes in the detector state is to actively induce them either in some detector component or in the terrestrial environment. However, this is not always practical

and only simulates some aspects of the full problem. Some kind of passive monitoring of *all the data channels* is also required.

Thus, in principle, one would like to study all change points in all data channels. Clearly, given the large data rate from the detectors, an *automated system* is required for detecting change points. We address here the issues in *algorithm design* for such a system and present the first steps towards the design of such an algorithm which we call the detector characterization robot (DCR).

2. Meaning of ‘change points’ and requirements for interferometric GW detectors

Some examples of what is meant by *change points* are (i) changes in the power spectral density of a channel, (ii) changes in the transfer function connecting one channel to another, (iii) presence of transients, (iv) change in the nature of phase and amplitude evolution of a narrowband noise feature, (v) change in the rate of transients and (vi) change in the time delay distribution between transients of a certain type occurring in coincidence in two different channels, etc.

Since one is dealing mostly with noise in all channels (as opposed to predictable signals), there will be *false alarms* and *false dismissals* in the detection of change points. Even if we were to continuously monitor only a few channels for change points and follow up each such point with more intensive study, the workload can become overwhelming if the false alarm rate is not under control. Conversely, a poor false dismissal probability can bias our understanding since a large fraction of genuine change points would be missed.

Thus, not only do we need an automated system for detecting change points, we also want it to be *reliable*, i.e. it must have a known false alarm rate or at least a robust upper limit on false alarms. This insistence on reliability as a constraint on algorithm design for such a system is a key point.

Given that this system must deal with PEM data also, which can be highly non-stationary/non-Gaussian or generally difficult to model, achieving a known reliability becomes a non-trivial task. In addition, reliability will not be a simple function of the performance of individual tools alone. As an example, line tracking tools can be influenced by the presence of transients and the same is true for the converse. Thus the design of the system must account for such interdependences from the start and also be robust, where required, against the different statistical character of data from different channels.

3. The DCR concept

The DCR algorithm seeks to achieve the aims outlined in the previous section. The design of the DCR (figure 1) as it stands now is aimed at achieving a restricted task, namely, reliably mark segments of data in each channel such that the noise floor between the consecutive marks has the same statistics up to the second-order moments (i.e. the noise is *wide sense stationary*). In this approach, all further processing is supposed to be done only with such segments. For example, parameters for data conditioning algorithms are constructed from each such segment separately. Confidence level estimation via fake signal injection should be done only within one segment. One could of course combine segments provided one knew what to do about the particular type of non-stationarity connecting any two segments.

Here is a brief explanation of the current design (figure 1).

- (i) Blocks (1) and (2) are devoted to the cleaning of line features, detection of transients in all channels and noise floor tracking. The line estimation algorithm should (i) work in

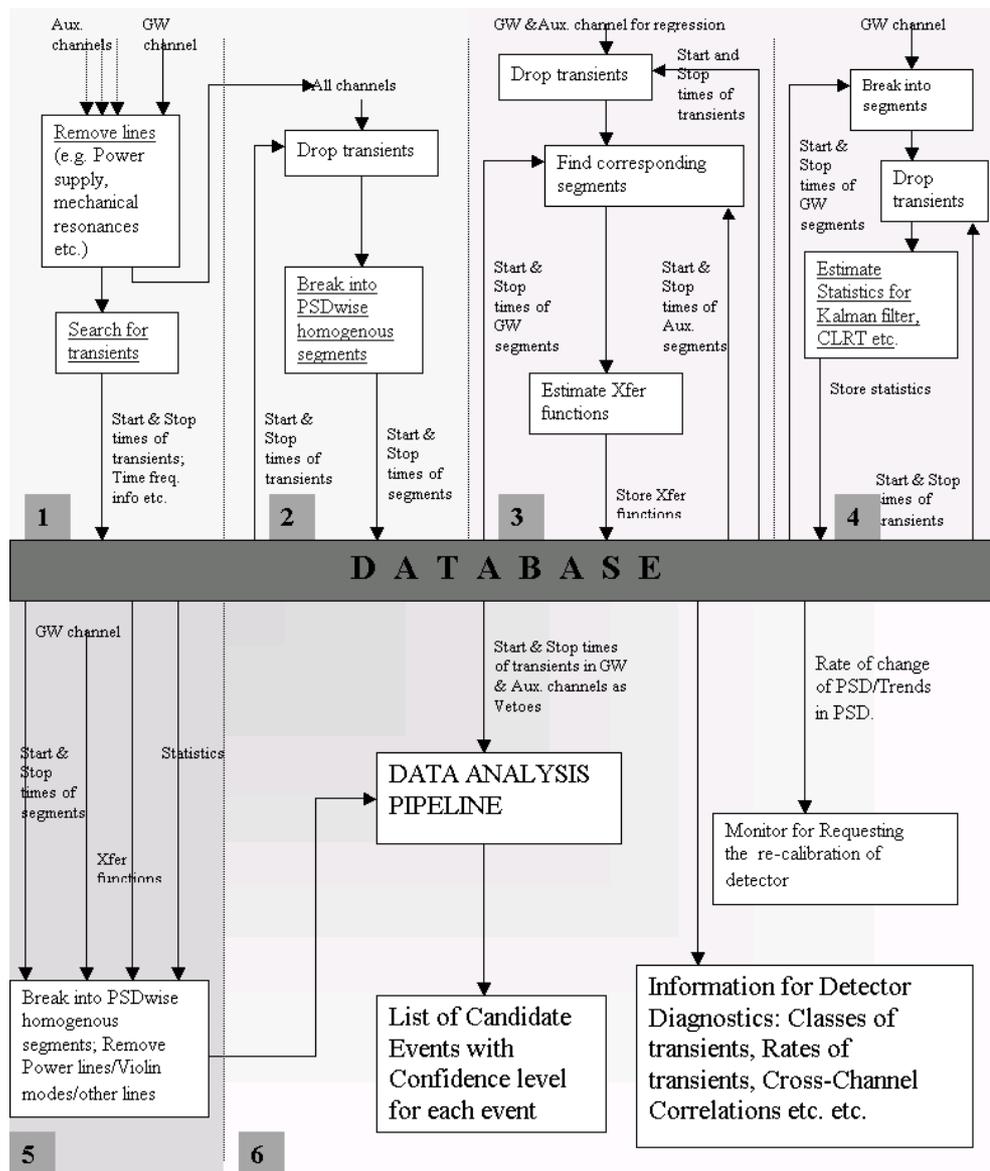


Figure 1. A schematic diagram of the DCR algorithm.

the time domain and (ii) should be robust against transients so that line removal does not reduce the power in a transient substantially. The transient detection algorithm should have a false alarm rate that is weakly dependent on the statistical nature of the data. Removal of lines substantially reduces false alarm in transient detection algorithms.

Candidate algorithms which are designed to meet the above requirements are described in [1] (transient detection) and [2] (line removal).

- (ii) The noise floor tracking algorithm is based on ARMA (adaptive/non-adaptive) modelling of the residual after line removal (cf section 5).

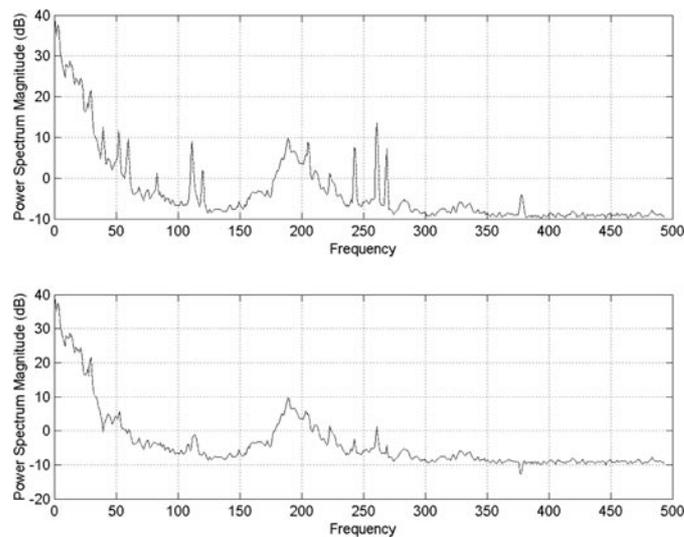


Figure 2. Caltech 40 m 1994 data run: PSD after line removal of IFO_Seis.1 channel before (top) and after (bottom) line removal using MBLT.

- (iii) Blocks (3), (4) and (5) use the wide sense stationary segments found earlier to estimate transfer functions and some essential parameters for more specialized line removal (such as [3] and [4]) and transient detection algorithms [5].
- (iv) The database block in the centre of the diagram indicates that information about change points and summary information about stationary segments will be recorded and available for other analyses.
- (v) Block (6) indicates the possible uses of the information stored in the database.

4. Implementation

The DCR is being implemented within the GEO600 project [6] in the C++ language. It will be based on a general purpose C++ digital signal processing library which is now under construction¹. Since a number of DCR components will be operating in the time domain, the computational requirements could be large. The available computational resources will play a significant role in deciding between alternative DCR designs.

5. Preliminary results of analysis

A short segment of data from the Caltech 40 metre prototype interferometer's 1994 data run is used for demonstration of our preliminary results. The line removal is done by a model-independent time domain method called the median based line tracker (MBLT) [2]. Figure 2 shows the performance of this technique for a seismic channel. Note that since this is a model-independent algorithm, it is possible to suppress lines throughout the Nyquist frequency range of the data irrespective of any assumptions about the line source. It has also been shown that the transients are not affected substantially in the seismic channel by the line removal method.

¹ A DSP library is under construction by the authors and members of the gravitational waves group at Cardiff University, UK.

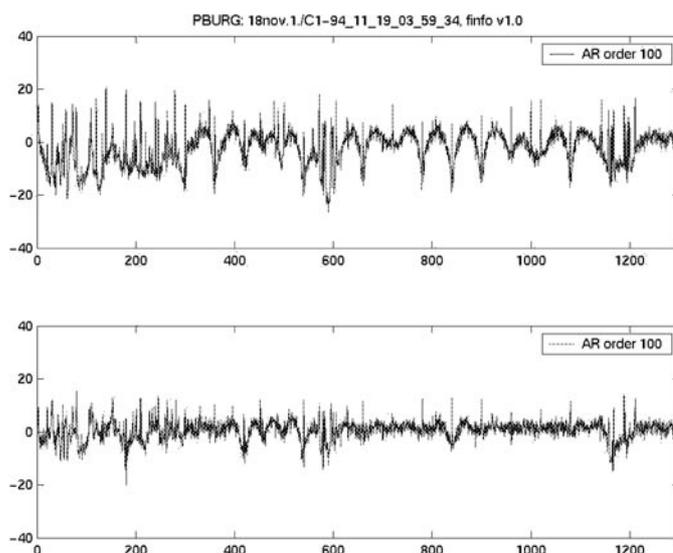


Figure 3. Caltech 40 m 1994 data run: each plot shows the ratio of PSD (IFO_DMRO channel) obtained from Welch's method to that obtained from an AR(100) model: (top) without line removal and (bottom) with line removal.

We obtain power spectral density (PSD) estimates by *time series modelling* [7]. Given a stochastic time series, one can model it (up to second-order moments) as being the result of white noise filtered through a zero-pole filter (i.e., an ARMA model). The zeros, poles and the gain of the filter can be estimated directly from the data. Once the filter parameters have been obtained, the estimate of the PSD (in an ensemble sense) can be obtained from the modulus square of the filter transfer function itself. This is the essence of *parametric* spectral estimates as opposed to *non-parametric* methods such as Welch's overlapped segment averaging method [7]. The advantage of parametric fit to the data is that if the data do indeed follow the general model outlined above, the signal-to-noise ratio of the estimated PSD is much better. However, in the case of interferometric data, the presence of line features prevents the full exploitation of the power of time series analysis. We amend for it by using the MBLT. Figure 3 shows that the PSD estimate is closer to the true value (estimated from a much longer data segment by Welch's method) after the lines have been removed by the MBLT.

6. Evolution of DCR and future direction

Evolution of the DCR *must* be dictated by the actual data that stably operating detectors will produce. Here is an example of a possible scenario: semi-automation of the first few tasks that a human analyst would do when looking more deeply at a change point. For instance, classify the change point based on past experience and reanalyse the data with tests that are more finely tuned for that class of change points. Thus, if the inter-relationship of some set of channels were to change, DCR will redo an analysis of variance (ANOVA) [8] on that set but check the statistics (Gaussian or non-Gaussian) of the data at the moment before choosing between parametric or non-parametric methods. It is possible to think about many such new directions, but it must be emphasized that the necessity of any of these will become evident only with time and experience with actual data.

Acknowledgment

We gratefully acknowledge the use of LIGO 40 metre prototype (1994) data for demonstration of the preliminary results.

References

- [1] Mohanty S D 2000 *Phys. Rev. D* **61** 122001
- [2] Mohanty S D 2002 *Proc. of the 4th Edoardo Amaldi Conf. on Gravitational Waves (Perth, Western Australia, 8–13 July 2001)* *Class. Quantum Grav.* **19** 1513
- [3] Mukherjee S and Finn L S 2000 *Proc. 3rd Amaldi Conference (Pasadena)* ed S Meshkov (New York: AIP) p 362
Finn L S and Mukherjee S 2001 *Phys. Rev. D* **63** 062004
- [4] Sintes A M and Schutz B 1998 *Phys. Rev. D* **58** 122003
- [5] Arnaud N *et al* 1999 *Phys. Rev. D* **59** 082002
- [6] Danzmann K *et al* 1995 *Proc. Edoardo Amaldi Conf.* (Singapore: World Scientific) p 86
- [7] Percival D B and Walden A T 1993 *Spectral Analysis for Physical Applications* 1st edn (Cambridge: Cambridge University Press)
- [8] Rao C R 1973 *Linear Statistical Inference and its Applications* (New York: Wiley)