Effect of calibration errors on Bayesian parameter estimation for gravitational wave signals from inspiral binary systems in the advanced detectors era: Further investigations

S Vitale[1]**, W Del Pozzo**[1]**, T G F Li**[1]**, C Van Den Broeck**[1]**,
B Aylott**[2] **and J Veitch**[1]

[1] Nikhef, Science Park 105, 1098 XG Amsterdam, The Netherlands
[2] School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

E-mail: `salvatore.vitale@ligo.org`

Abstract. By 2015, the advanced versions of the gravitational wave detectors Virgo and LIGO will be online. They will collect data in coincidence with enough sensitivity to potentially deliver multiple detections of gravitational waves from inspirals of compact-object binaries. In a previous work, we have studied the effects introduced in the estimation of the physical parameters of the source by uncertainties in the calibration of the interferometers. Our bias estimator for parameter errors introduced by calibration uncertainties consisted of two terms: A genuine bias due to the calibration errors, and a contribution coming from the limited number of samples used to explore the parameter space. In this article, we have focused on this second term, and we have shown how it is smaller than the former (about 10 times smaller), and how it decreases as the signal-to-noise ratio increases.

1. Introduction

In a recent paper [1], we have shown that the bias introduced by realistic calibration errors (CEs) in the parameter estimation process for gravitational wave (GW) signals are usually smaller than the uncertainty due to the noise in the instruments. As parameters are estimated using numerical algorithms, in our case, the Nested Sampling algorithm [2] as implemented and described in [3], the biases we have found were really made up of two contributions: A real bias, driven by the size and shape of the calibration errors, and an error coming from the finite sampling of the parameter space. In the rest of this work, we have referred to the first term as *calibration bias* and to the second as *sampling errors*. These are not related to CEs, and in fact, have been presented in any results obtained using Monte Carlo or Nested Sampling based algorithms. In this work, we have quantified the contribution of this term and we have shown that (i) it is much smaller than the calibration bias, (ii) it should not affect in any important way the parameter estimation (PE) process, and (iii) it is smaller for loud signals.

This article is organized as follows: In Section 2, we have described the method used to estimate the bias, in Section 3, we have reported the main results and discussed them, and in Section 4, we have summarized and concluded. We refer to [1] for the main methodology and

results on calibration induced PE biases, to [2, 3] for the Nested Sampling methodology and implementation, and to [6] for a description of the waveform and the parameters on which it depends.

2. Method

In [1], we have estimated the bias introduced by mock, but realistic, calibration errors by running a Nested Sampling algorithm on three catalogues of simulated GW signals. The analysis was performed twice: A first time on the "exact" signals, and then on signals with added calibration errors, while keeping fixed all the relevant parameters. The runs were set up to use two chains to explore the parameter space in parallel to make optimal use of computer clustering, which were then combined to provide a single estimate for the posterior distributions of the parameters and the Bayes factor. Because of the randomness in the exploration of the parameter space, and the finite sample size, running the code on the same stretch of data but starting the exploration from a different point would result in finding slightly different posterior distributions. As mentioned before, this effect is a general one, and is not related to the presence of calibration errors.

Note, however, that in our previous study, calibration errors were the cause for this randomness to arise. In fact, adding calibration errors results in a slight change of the profile of the likelihood function, which is the real driving engine of the Nested Sampling algorithm. As a consequence, even if the two chains in the runs with CE started from the same point of the parameter space as the two chain in the run without CE (which is, indeed, the case, as we had kept *all* parameters and settings to be the same while adding calibration errors, and that includes the seed for the generation of the chain) they would ultimately follow different paths in their exploration of the parameter space. This effect has to be added to the genuine bias (the calibration bias), which would be there even in the case of perfect sampling with an infinite number of iterations, and its magnitude will depend on the particular setting of the run (number of live points and MCMC iterations) and the GW event. Questions which were left unanswered in [1] are: What can be told about the relative weight of these two biases? How does the sampling errors depend on the signal-to-noise ratio (SNR) of the injected signals? In this article, we have answered these questions by analysing in greater details the sampling errors, their typical magnitude and shape.

In particular (a) Take the set of binary systems composed of one neutron star and one black hole (BHNS) used in [1], limiting ourselves to the signals which had passed the SNR cut described therein, (b) Analyse each signal using 12 parallel chains (each run used 1700 live points and 210 MCMC iterations, see [3] for details). These chains are then combined into pairs, always using different chains, to yield 6 independent estimations of the source parameters, and (c) Calculate the differences in the estimated parameters one obtains by using a pair of chains instead of another one. To be more precise, for each event, we have calculated the quantity (somehow similar to the one used in [1]) as:

$$\Sigma_{ij}^{\alpha} \equiv \frac{\overline{\alpha}_i - \overline{\alpha}_j}{\sqrt{\frac{1}{2}\left[\sigma_n(\alpha_i)^2 + \sigma_n(\alpha_j)^2\right]}}, \qquad \text{with } 6 \geq i > j \geq 1 \qquad (1)$$

where $\overline{\alpha}_k$ is the median of the parameter α as calculated by the k-th run ($k = 1, 2, .., 6$) and $\sigma_n(\alpha_k)$ is the corresponding standard deviation. The advantage of normalizing by a symmetric combination of the standard deviations is that Σ will be a dimensionless quantity of the order of 1 or less (while the various parameters on which the waveform depends have very different ranges of variation as well as different units). Nevertheless, we have found occasionally useful to work with the unweighted bias, for which we have used the symbol $\Delta_{ij}^{\alpha} \equiv \overline{\alpha}_i - \overline{\alpha}_j$.
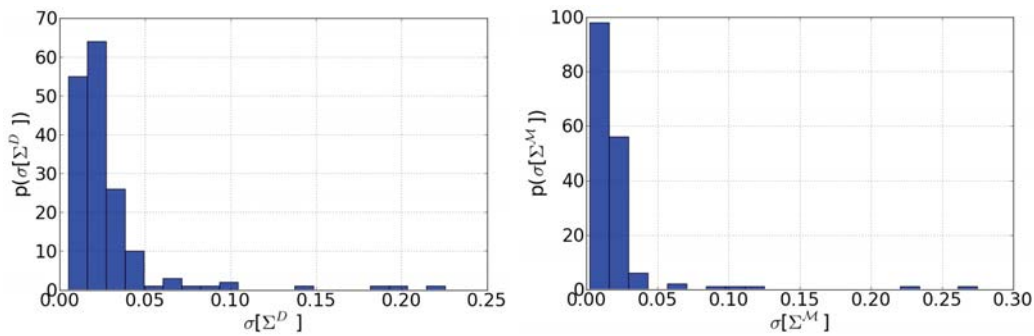
Figure 1. *Left*: The distribution of the standard deviation of the weighted bias among the signals in the catalogue for the distance estimation. *Right*: A similar plot for the chirp mass estimation

3. Results

For each event and each parameter, one can build a distribution of Σ_{ij}^{α}, with $6 \geq i > j \geq 1$. Its standard deviation can give an idea of the typical sampling errors one might get and we have labelled it $\sigma[\Sigma]$. One can then show how the width of the sampling errors distributions varies by building an histogram of $\sigma[\Sigma]$ for all the signals in the catalogue. This is shown in Figure 1 for the chirp mass and the distance (see [6] for a description of the waveform and parameters).

The results for all the parameters are summarized in Table 1, where we have given the mean and standard deviation of $\sigma[\Sigma]$ among the 167 events, as well as the 5th and 95th percentiles (we have left out the statistics for the polarization and the coalescence phase as those are, in practice, not estimable with Advanced LIGO and Virgo). In the last column, we have reported the results we had found in [1] for the standard deviation of the weighted bias due to the joint effect of calibration bias and sampling errors.

	$\overline{\sigma[\Sigma]}$	$\Delta\{\sigma[\Sigma]\}$	5th	95th	$\overline{\sigma[\Sigma]}$ [CB+SE]
$\mathcal{M}$	$2.02 \cdot 10^{-2}$	$2.96 \cdot 10^{-2}$	$7.02 \cdot 10^{-3}$	$3.36 \cdot 10^{-2}$	$1.02 \cdot 10^{-1}$
η	$1.99 \cdot 10^{-2}$	$2.99 \cdot 10^{-2}$	$6.71 \cdot 10^{-3}$	$3.67 \cdot 10^{-2}$	$1.28 \cdot 10^{-1}$
RA	$5.84 \cdot 10^{-2}$	$2.22 \cdot 10^{-1}$	$2.15 \cdot 10^{-3}$	$5.32 \cdot 10^{-2}$	$3.87 \cdot 10^{-1}$
dec.	$5.54 \cdot 10^{-2}$	$1.88 \cdot 10^{-1}$	$4.60 \cdot 10^{-3}$	$1.11 \cdot 10^{-1}$	$4.49 \cdot 10^{-1}$
t_0	$5.68 \cdot 10^{-2}$	$2.07 \cdot 10^{-1}$	$4.90 \cdot 10^{-3}$	$5.74 \cdot 10^{-2}$	$4.05 \cdot 10^{-1}$
D	$2.75 \cdot 10^{-2}$	$2.98 \cdot 10^{-2}$	$8.81 \cdot 10^{-3}$	$6.88 \cdot 10^{-2}$	$2.80 \cdot 10^{-1}$
ι	$6.03 \cdot 10^{-2}$	$1.40 \cdot 10^{-1}$	$8.99 \cdot 10^{-3}$	$2.98 \cdot 10^{-1}$	$3.75 \cdot 10^{-1}$

Table 1. The mean $\overline{\sigma[\Sigma]}$, standard deviation $\Delta\{\sigma[\Sigma]\}$, 5th and 95th percentiles of the standard deviation of Σ for all the parameters. The last column shows the value of the standard deviation that was found in [1], for the bias due to the joint effect of calibration errors and sampling errors. These numbers take into account all 167 events.

On average, we can say that using a different pair of chains to explore the parameter space has little effect on the estimated parameters. From Table 1, we have seen, in fact, that the averaged $\sigma[\Sigma]$ is smaller than 6% of the noise error for all the parameters (and only $\sim 2\%$ for the intrinsic parameters). Moreover, the distributions are quite compact: For the declination and inclination angle, 90% of the times the standard deviation is contained in a region corresponding, respectively to $\sim 11\%$ and 30% of the noise error; for the other parameters, this region is even narrower (less than 7%). Moreover, comparing the first and last columns, we can say that the
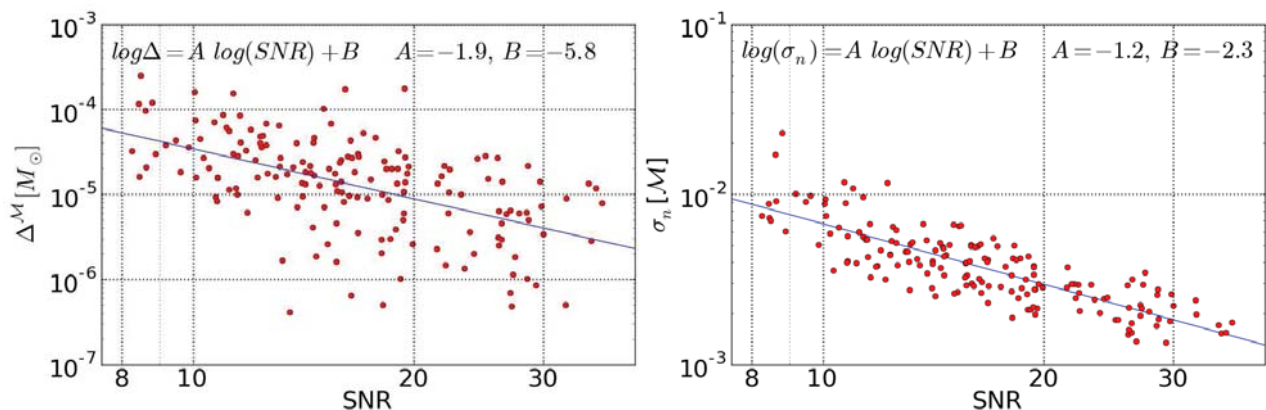
Figure 2. *Left*: The sampling errors in the estimation of the median of the chirp mass for all the signals in our catalogue together with a best fit. *Right*: The width of the posterior for the chirp mass against the SNR together with a best fit (see the text for more details).

sampling errors had a small weight in the biases we have found in [1]. Generally speaking, the results of Table 1 are good as they confirm that the results found by parameter estimation codes are robust against the particular pattern followed to explore the parameter space, at least with the set up we have used.

We have briefly investigated the dependence of the sampling errors on the SNR of the simulated signals. In the left panel of Figure 2, we have shown for each event the mean of the bias $\Delta^{\mathcal{M}} \equiv \frac{1}{15} \sum_{6 \geq i > j \geq 1} \Delta_{ij}^{\mathcal{M}}$ among the 15 combined runs plotted against the SNR. If the samples used to build the posterior distributions are independent, we expect the errors to go like $\sigma_n/\sqrt{N}$, where σ_n is the width of the posterior distribution (due to the noise), and N is the number of samples in the chain. From Fisher Information studies [4, 5, 6], we expect that at least for medium-high SNRs (≥ 15), $\sigma_n \propto 1/\text{SNR}$ so that the bias we measure should also depend on the inverse of SNR. The fit in the left panel of Figure 2 (blue line), shows, instead, that the bias dependence on the SNR is more close to an inverse quadratic relationship, and that there is a non-negligible scattering. On the other hand, the noise standard deviation $\sigma_n[\mathcal{M}] \equiv \frac{1}{15} \sum_{6 \geq i > j \geq 1} \sqrt{\frac{1}{2} [\sigma_n(\mathcal{M}_i)^2 + \sigma_n(\mathcal{M}_j)^2]}$ for the chirp mass is quite close to an inverse relationship with the SNR, and the points are close to the best fit line. The reasons why the bias seems not to follow the expected distribution, as well as a check on the dependence of the Nested Sampling algorithm on the number of samples will be the subject of a forthcoming work.

4. Conclusions

In this work, we have investigated the typical contributions of statistical fluctuations due to limited posterior sample size to the fluctuation of the mean estimator of posterior parameter values, and we have compared it with the bias introduced by calibration errors, which we had quantified in [1]. We have built a catalogue of 167 events, injected into simulated Advanced LIGO/Virgo noise, and we have run the parameter estimation code using 12 parallel chains to explore the parameter space. For each event, the chains were combined into 6 independent pairs, and the results delivered by each pair were confronted with those of the others. We have found that the code is robust against the statistical fluctuations arising from using a pair of chain than another, the typical shifts in the estimated parameters being a tiny fraction of the random error due to the noise in the instruments: For all parameters but inclination and declination, 90% of the signals had a spread in the estimation smaller than 4% of the noise random error. We also

have verified that these errors decrease as the SNR of the event increases, even though they do not seem to follow the expected inverse relationship.

References

[1] Vitale S *et al* 2012 *Phys. Rev. D* **85** 064034
[2] Skilling J 2004 in *AIP Conf. Proc.: 24th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* **735** 395
[3] Veitch J and Vecchio A 2010 *Phys. Rev. D* **81** (6) 062003 [gr-qc/0911.3820]
[4] Zanolin M, Vitale S and Makris N 2010 *Phys. Rev. D* **81** 124048
[5] Vitale S and Zanolin M 2010 *Phys. Rev. D* **82** 124065
[6] Vitale S and Zanolin M 2011 *Phys. Rev. D* **84** 104020