Testing the system identification analysis for LISA Pathfinder in case of non-stationary noise processes and data gaps

Nathaniel Strauss
Department of Physics and Astronomy
Carleton College
Northfield, MN 54636, USA

July 31, 2015

Abstract

We utilize LISA Pathfinder's (LPF) already well-developed software package LTPDA (LISA Technology Package Data Analysis) in order to examine how non-stationary noise and gaps in data acquisition affects LPF system identification. We work primarily within a Bayesian framework, developing four different likelihood formulations, each with a different method of dealing with the problem of an unknown noise spectrum. The χ^2 likelihood simply fits residuals to previously measured noise, the noise scaling likelihood allows the measured noise to change in magnitude between measurement and experiment, the student-t likelihood allows the noise to vary more widely than in a Gaussian distribution, and the logarithmic likelihood marginalizes out the noise spectrum entirely. We test these four different theoretical approaches against different non-stationary noise processes, including an overall raised noise level and injected Gaussian bursts. In addition, we test these theoretical approaches in the case of gaps in data acquisition. We perform these tests in three separate simulated experiments: a toy model, simulated LPF data from the Albert Einstein Institute, and simulated data from the European Space Agency. We find that in the case of LPF system identification, all approaches are quite robust against both non-stationary noise and gaps in data acquisition.

...

1 Introduction

Now is an exciting time in the field of gravitational wave research. The ground-based Laser Interferometer Gravitational-wave Observatory (LIGO) detectors are coming online this year and we could see the first direct detection of a gravitational wave at any time. However, ground-based detectors are limited to sensitivity to signals on the order of hundreds of Hertz

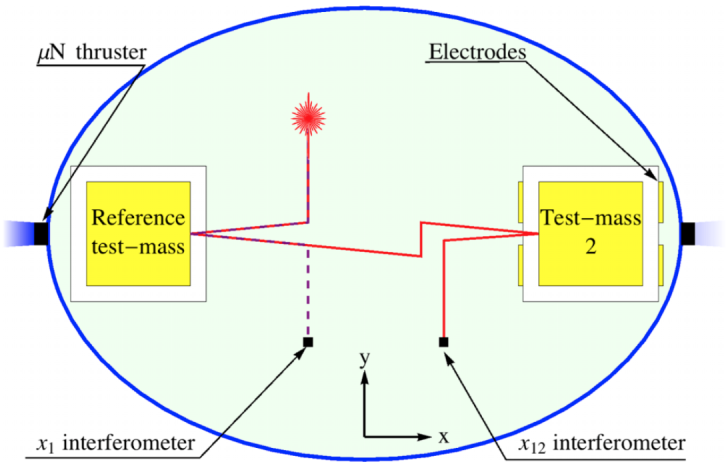


Figure 1: A functional schematic of the LPF spacecraft. The reference test mass (TM1) is kept in freefall via a control loop between interferometer x_1 and the spacecraft thrusters, and the second test mass (TM2) is kept stationary within the spacecraft via a control loop between interferometer x_{12} and the electrodes. [5]

due to seismic noise in lower frequencies [6]. Space-based detectors have significantly lower seismic noise and thus have high sensitivity to signals on the order of mHz.

The goal of the Evolved Laser Interferometer Space Antenna (eLISA) project is to construct such a detector. While a fully functioning space-based gravitational wave detector is more than 20 years away, the eLISA project is finally "getting off the ground" in the form of LISA Pathfinder (LPF), which is a space mission dedicated to demonstrating the technology needed for the eLISA project, as well as confirming Einstein's geodesic motion. The LPF spacecraft consists of two test masses and two interferometers. The spacecraft utilizes its thrusters to remain centered around the first test mass (TM1, also called the reference test mass) such that it is in free fall. The second test mass (TM2) is held stationary in the frame of the spacecraft with electrodes. To accomplish this, interferometer x_1 measures the distance between TM1 and the spacecraft, and interferometer x_{12} measures the distance between the two test masses. Interferometer x_1 feeds information to the thrusters and x_{12} feeds information to the electrodes. (See Figure 1.) The goal of the mission is to measure the relative acceleration between the test masses to the accuracy required for the eLISA project. If we reach the required sensitivity, we can also use the apparatus to measure the geodesic motion of the two test masses.

In this paper we are concerned with the system identification analysis for the LPF apparatus. Our model for the dynamics of the LPF system includes constants such as the spring-like stiffnesses of the test masses or the system delay between interferometer measurement and the application of command forces by the electrodes on the second test mass. The goal of system identification is to determine these constants within our model for LPF experiment. In order to accomplish this, using primarily Bayesian methods on simulated LPF data we compare the model for the residuals to previously measured noise. However, in the future space-based detectors, the gravitational signals we expect to see in our data

cannot be turned off in order to measure the noise directly, which makes it difficult to use traditional Bayesian methods. In addition, there may be non-stationary noise processes, which may change the characteristics of the noise between measurement and experiment. We attempt to preemptively deal with these problems with a variety of theoretical approaches to the Bayesian likelihood. In this paper, we compare these different approaches against different characterizations of the noise, as well as other complications that may arise during data acquisition, such as missing data.

Outline The remainder of this paper is organized as follows:

Section 2 outlines several theoretical approaches to the Bayesian likelihood.

In Section 3, we test the theoretical approaches outlined in Section 2 on a toy model of sinusoidal signals embedded in some noise, with different noise realizations between measurement in experiment.

In Section 4, we perform parameter estimation tests on LPF data simuated by the Albert Einstein Institute and by the European Space Agency, both in the case of varying noise realizations and in the case of missing data.

Finally, Section 5 gives a summary of our findings.

2 Theoretical Approaches

2.1 Bayesian framework

In LPF we must be able to characterize the system in order to extract signals from the noisy data. There are several possible approaches to system identification, but we focus primarily on a Bayesian framework. To utilize Bayesian inference, we assume the following model for the observed data y:

$$y = h(\vec{\theta}) + n, \tag{1}$$

where $h(\vec{\theta})$ is the quantitative result from the system model $\mathcal{M}$ given a set of system parameters $\vec{\theta}$ and n is the noise.

If we assume the quantitative model $\mathcal{M}$ we can assign probabilities to different parameter configurations $\vec{\theta}$ using Bayes' theorem:

$$p(\vec{\theta}|y, \mathcal{M}) = \frac{p(y|\vec{\theta}, \mathcal{M})p(\vec{\theta})}{p(y|\mathcal{M})}. \tag{2}$$

The *posterior distribution* $p(\vec{\theta}|y, \mathcal{M})$ is the probability of the parameters given the observed data. Maximizing this value gives an estimate for the system parameters. The *prior distribution* $p(\vec{\theta})$ gives the domain and weights to the values in the parameter space, and the *evidence* $p(y|\mathcal{M})$ gives the ability of the model to describe the data, but for parameter estimation this is an unimportant normalization constant. Finally, the *likelihood* $p(y|\vec{\theta}, \mathcal{M})$ gives the probability that the data arose from the model [1].

There are many distinct formulations of the likelihood, and the primary difference between the methods we test is how the noise is or is not included in the analysis. Each formulation has its own strengths and weaknesses, and it is the goal of this paper to explore

which formulations are most suitable for LPF data analysis. In the following sections we define four different formulations of the Bayesian likelihood: the χ^2, scaled noise modeling, Student t-distribution, and logarithmic likelihoods.

2.1.1 χ^2 likelihood

We begin with the well-established χ^2 likelihood. Here, the noise is assumed to be Gaussian, uncorrelated, and stationary with zero-mean. Given these assumptions, the random noise process determines a natural inner product $(\cdot|\cdot)$ and associated norm on the vector space of measurements, given by [2]

$$(a|b) = 2\int_0^\infty df \Big[\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)\Big]/\tilde{S}_n(f), \tag{3}$$

where the tilde ($\tilde{}$) denotes operations in the frequency domain, the asterisk (*) denotes the complex conjugate, a and b are time-series data, and $\tilde{S}_n$ is the one-sided power spectral density (PSD) of the noise n in the data. Here we assume the PSD of n has been measured in a previous experiment.

It can then be shown [2] that the probability for n to take a specific value $n_0(t)$ can be expressed as

$$p(n = n_0) \propto \exp\Big[-\frac{1}{2}(n_0|n_0)\Big]. \tag{4}$$

From Equation 1, we see that this is equivalent to $p(y-h(\vec{\theta})) \propto \exp[-1/2(y-h(\vec{\theta})|y-h(\vec{\theta}))]$. We use this equation to define the χ^2 likelihood. For some constant C,

$$p(y|\vec{\theta}, \mathcal{M}, S_n) = C \times \exp\Big[-\frac{1}{2}\Big(y - h(\vec{\theta})\Big|y - h(\vec{\theta})\Big)\Big] = C \times e^{-\chi^2/2}, \tag{5}$$

where

$$\chi^2 = \Big(y - h(\vec{\theta})\Big|y - h(\vec{\theta})\Big). \tag{6}$$

Thus, in essence we fit the PSD of the residuals from the data and model to the PSD of the previously measured noise.

2.1.2 Scaled noise modeling

It may be the case that during an LPF experiment, the noise level on a specific frequency band will change in magnitude. Here we develop a likelihood with the same assumptions as in the χ^2 case, but we allow the noise level in pre-defined frequency bands to fluctuate. We divide $\tilde{S}_n(f)$ into k frequency bands such that

$$\tilde{S}_{n,i}(f) \to \eta_i\, \tilde{S}_{n,i}(f), \quad 1 \leq i \leq k, \tag{7}$$

where $\tilde{S}_{n,i}(f)$ is the i^{th} frequency band of $\tilde{S}_n(f)$ and η_i is the noise scaling factor for that band (See Figure 2).

We now consider these noise coefficients η_i as additional parameters to be estimated along with the system parameters of the model. Thus, some terms in C from Equation 5 are no

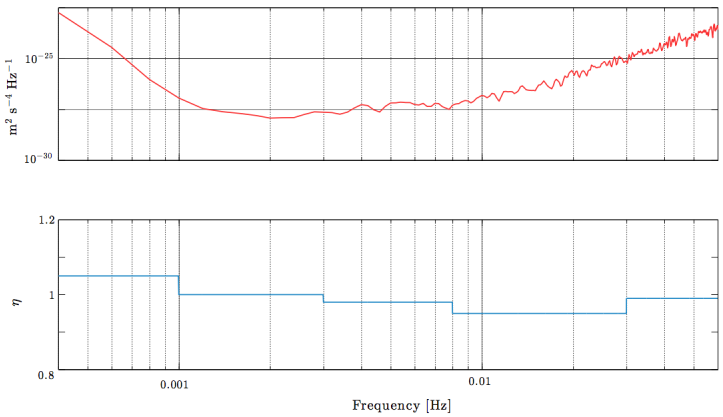


Figure 2: Here we see an example of a possible noise scaling configuration using the noise modeling likelihood. We divide $\tilde{S}_n(f)$ into five frequency bands, each with a different η_i. These two sets of frequency-dependent data are multiplied together point by point and then used as the noise in Equation 3 to formulate the scaled noise modeling likelihood.

longer constant. Pulling these terms out, it can be shown [3] that the likelihood for this new noise model is

$$p(y|\vec{\theta}, \mathcal{M}, S_n) = C' \times \exp\left[-\frac{1}{2}\Big(\chi^2 + \sum_{i=1}^{k} N_i \ln \eta_i\Big)\right], \tag{8}$$

where C' is some new constant, and N_i is the number of Fourier bins in the i^{th} frequency band. Henceforth we call this the *noise modeling* likelihood.

2.1.3 Student t-distribution

Recall the χ^2 likelihood, where we assume the noise to be Gaussian, uncorrelated, and stationary with zero-mean, which requires perfect knowledge of the noise spectrum and therefore its variance σ^2. If we allow the noise spectrum to vary around its measured values, this is equivalent to considering the variance of the noise σ^2 as a random variable, and we can produce a likelihood which marginalizes σ^2 out. It can be shown that this is equivalent to assuming the noise to be t-distributed [4].

First, we now consider the variance of the noise σ^2 to be a random variable drawn from the scaled inverse χ^2 distribution

$$\sigma^2 \sim \text{Inv-}\chi^2(\nu, s^2), \tag{9}$$

where ν is the degrees-of-freedom parameter and s^2 is the scale parameter given by

$$\nu = 4 + 2\frac{\text{E}[\sigma^2]^2}{\text{Var}[\sigma^2]} \tag{10}$$

and

$$s^2 = \frac{\nu - 2}{\nu}\text{E}[\sigma^2]. \tag{11}$$

We take $\mathrm{E}[\sigma^2]$ to be the to be the Gaussian variance σ^2 of the noise, and the ratio $\mathrm{E}[\sigma^2]^2/\mathrm{Var}[\sigma^2]$ to be a fixed parameter in the formulation of this likelihood, which captures how far we allow σ^2 to vary from its mean value.

Using this scaled inverse χ^2 distribution as a prior, we marginalize σ^2 out and arrive at the following marginalized likelihood [4]:

$$p(n = n_0|\nu, s^2) \propto \exp\Big[-\sum_{i=1}^{N} \frac{\nu+2}{2} \ln\big[1 + \frac{|\tilde{n}_{0,i}|^2}{s^2\nu}\big]\Big], \tag{12}$$

where $n_{0,i}$ is the i^{th} data point of the noise residual given by Equation 1, and N is the total number of Fourier bins. We now take this as our Bayesian likelihood:

$$p(y|\vec{\theta}, \mathcal{M}, \nu, s^2) \propto \exp\Big[-\sum_{i=1}^{N} \frac{\nu+2}{2} \ln\big[1 + \frac{|\tilde{y}_i - \tilde{h}_i(\vec{\theta})|^2}{s^2\nu}\big]\Big], \tag{13}$$

where $\tilde{y}_i$ and $\tilde{h}_i(\vec{\theta})$ represent the i^{th} data points in the respective one-sided PSDs. Note how due to the placement of the sum that this is not relatable to the χ^2 quantity from Equation 18. This likelihood shows that this process is equivalent to treating the noise as t-distributed [4], and henceforth will be called the *student-t* likelihood. In the limit of large ν, the Student t-distribution approaches the Gaussian distribution. Thus, for large ν, we should expect this likelihood to be equivalent to the χ^2.

2.1.4 Logarithmic likelihood

It can be shown [5] that the χ^2 likelihood in Equation 5 can be expressed exactly as

$$p(y|\vec{\theta}, \mathcal{M}, S_n) = \prod_{i=1}^{N} \frac{1}{\pi\tilde{n}_i} \exp\Big[-\frac{|\tilde{y}_i - \tilde{h}_i(\vec{\theta})|^2}{\tilde{n}_i}\Big], \tag{14}$$

where $\tilde{n}_i$ is the i^{th} data point in the one-sided PSD of the previously measured noise $\tilde{S}_n$. However, if we partition the time series data into shorter stretches and average, this becomes

$$p(y|\vec{\theta}, \mathcal{M}, S_n) = \prod_{i=1}^{N} \frac{1}{(\pi\tilde{n}_i)^{N_s}} \exp\Big[-N_s\overline{\frac{|\tilde{y}_i - \tilde{h}_i(\vec{\theta})|^2}{\tilde{n}_i}}\Big], \tag{15}$$

where the bar is an average over N_s stretches of data. [5]

Then we marginalize $\tilde{n}_i$ out of Equation 15 by integrating over $\tilde{S}_n$ with a uniform prior in $\ln \tilde{S}_n$. It can then be shown [5] that for high enough N_s, the marginalized likelihood is given by

$$p(y|\vec{\theta}, \mathcal{M}, N_s) \propto \prod_{i=1}^{N} \Big(\overline{|\tilde{y}_i - \tilde{h}_i(\vec{\theta})|^2} \Big)^{-N_s}. \tag{16}$$

The relationship between Equations 15 and 16 is reminiscently logarithmic, so we call this the *logarithmic* likelihood. Because this likelihood is unweighted by the noise, we should expect larger errors in the resulting parameter estimations than in the χ^2 likelihood. It can also be shown that Equation 16 can be derived from the student-t likelihood by taking an improper Jeffrey's prior with $\nu = 0$ [4].

2.2 Iterative Reweighted Least Squares (IRLS) method

The Iterative Reweighted Least Squares (IRLS) method is unlike the other likelihoods in that it is frequentist rather than Bayesian in its approach, and its implementation is iterative in nature. At each iterative step, we maximize the χ^2 likelihood, using the resulting residuals $(y - h(\vec{\theta}))$ as the noise spectrum in the next iteration. At the first iteration, we make an arbitrary guess for the noise spectrum, which we can take to be uniform or just the spectrum of the data. The iteration terminates when the noise residuals cease to change between iterations to some tolerance [5]. Thus, the IRLS likelihood is defined by

$$p(y|\vec{\theta}, \mathcal{M}) = C \times e^{-\chi_n^2/2}, \tag{17}$$

where

$$\chi_n^2 = \sum_{i=1}^{N} \frac{|\tilde{y}_i - \tilde{h}_i(\vec{\theta}_n)|^2}{|\tilde{y}_i - \tilde{h}_i(\vec{\theta}_{n-1})|^2}. \tag{18}$$

Here χ_n^2 is the likelihood at at the n^{th} iteration and $\vec{\theta}_{n-1}$ is the estimation result from the previous iteration.

Unlike all previous likelihoods, which only have to be maximized once, this likelihood must be maximized several times in order to achieve convergence. Thus, in general utilizing the IRLS method is much more computationally expensive than any other likelihood. However, if the model is linear, meaning that the parameters are simply amplitude coefficients, then maximizing this likelihood is reducible to solving a system of linear equations [5], and is such much less computationally intensive than maximizing the above likelihoods.

It can also be shown [5] that the IRLS method converges to the same parameter values as the logarithmic likelihood. However, unlike the logarithmic likelihood, this method is explicitly Gaussian, and may have different error estimates for its parameters.

3 Tests on a Toy Model

In this section we perform our initial tests of the four likelihood formulations and the linear least squares method. The model is a simple superposition of three sinusoids of different frequency, amplitude, and duration. We will estimate the amplitudes A, B, and C of the three signals, assuming knowledge of the frequencies, durations, and start times. To generate the data, we add the sinusoids from the model to a stretch of noise (as in Equation 1), which has been colored similar to LPF-acceleration noise. Also, most likelihood formulations developed above need some prior knowledge of the noise in the data, so we feed them a stretch of noise without an embedded signal (See Figure 14). However, in the LPF experiment, the noise is likely to change between measurement and experiment, so the noise we use in the construction of the data may be different from the noise we use in constructing the likelihoods. We also use Equation 1 as our fitting model, so by design our residuals from the model fit should be exactly the same as our noise, at least in the case of no change in noise between measurement and experiment, as shown in Figure 3. For further confirmation that our model is correct, we performed the Kolmogorov-Smirnov test, which tests if two random variables come from the same distribution. We take the residuals and noise to be from the same distribution and find that this assumption passes the Kolmogorov-Smirnov test.

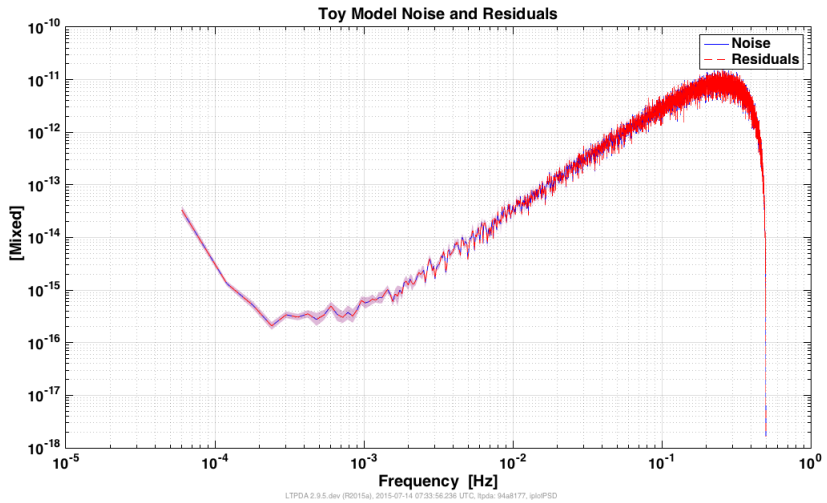
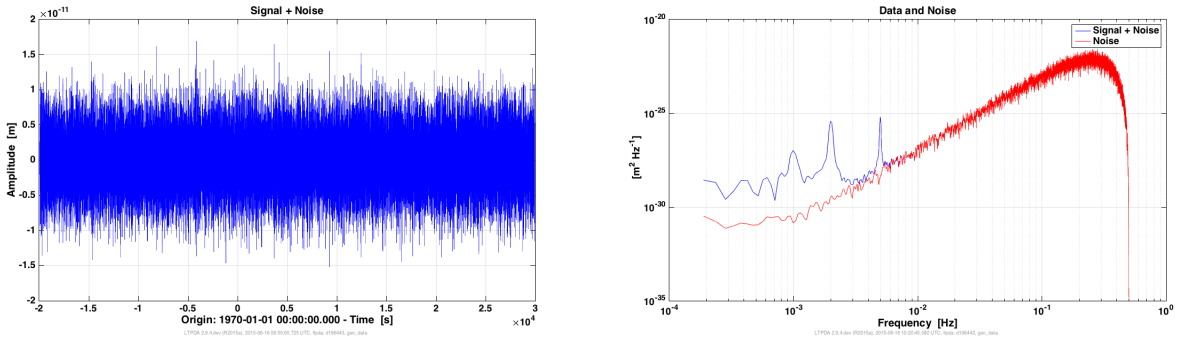


Figure 3: Here we plot the PSDs of both the noise in our data and of our residuals (the model for the signal subtracted from the experimental data) with area errors. Because this is exactly how we construct our data in the first place, we see that the two curves are identical.

In this section we explore how different changes to the noise between measurement and experiment affect the parameter estimation results from the different likelihood formulations and the least squares method.

MCMC tests For all of the parameter estimations outlined in Section 2, we need to maximize the posterior distribution. In order to accomplish this (except for in the case of the IRLS method, where we solve linear equations), we utilize a modified Metropolis-Hastings algorithm. The Metropolis-Hastings algorithm is a Markov Chain Monte Carlo (MCMC) test, where at each iteration, it takes a random step in the parameter space and accepts or rejects that step with a probability based on the difference between the values of the posterior distribution at the start- and end-points in the parameter space. A higher value in the posterior distribution signifies a higher probability that step will be accepted. If the model adequately describes the data, then after many iterations, the posterior distribution will converge to the global maximum, and the input parameters for the global maximum yield the parameter estimation. In addition, if we set our starting point in the parameter space near the maximum, we can map the posterior distribution in that reason to find the width of the maximum and find error estimates for each parameter.

Noise Changes We run three sets of 50 MCMC parameter estimation tests. In each set, we utilize a different noise realization when constructing the experimental data. In the first set of tests, we make no changes to the noise before adding it to the signal (See Figure 4). In the second set of tests, we increase the overall level of the noise in the experimental data by a factor of $\sqrt{2}$ (See Figure 15). In the third set of tests, we embed Gaussian burst signals into the experimental data (See Figures 16 and 17). See the results of the MCMC tests in



(a) Data Time Series

(b) Data and Noise PSDs

Figure 4: The data with no change in noise between measurement and experiment, simply adding the noise to the signal without any modifications. Note in the PSD that at high frequencies the data is indistinguishable from the noise, but we see low frequency differences, including the peaks from the sinusoidal signals.

Figure 5. Here is a list of our observations of the data table:

- In general the parameter estimations are very good in the first two noise realizations, ranging from 0.1% to 3% error in the worst cases.
- The logarithmic estimations have larger errors than the χ^2, as expected due to unweighted noise.
- The student-t estimations have smaller errors than the χ^2, which is unexpected, which means that the wider-tailed Student t-distribution is better suited to characterize the noise.
- The logarithmic and IRLS estimations have very similar results, but different errors, as expected.
- The errors in the noise modeling estimation are in general not very good, even for the change in noise level for which it was designed. This is likely due to improper setup of the MCMC test.
- The errors with the noise level change increase slightly for all but the χ^2 tests.
- The errors with the embedded Gaussian bursts increased by at least an order of magnitude for all but the χ^2 tests, and was truly disastrous for the noise modeling tests. This is likely due to the fact that in the low frequency range, the noise level changed by up to six orders of magnitude (See Figure 17).
- Surprisingly, the errors in the χ^2 tests became smaller with embedded Gaussian bursts, but we see that the spread of the means increased by an order of magnitude, just like the other tests. This could be because the location, size, and frequency of the Gaussian

bursts is randomized, so the convergence value for the parameters is not the same for each MCMC test.

		χ^2			Logarithmic $N_s = 5$			Noise modeling		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	1	1.0006	0.0040	0.0059	1.0038	0.0139	0.0135	0.9878	0.1712	0.2300
B	2	2.0009	0.0049	0.0063	2.0008	0.0073	0.0062	1.9850	0.1326	0.0746
C	2	1.9992	0.0288	0.0281	1.9957	0.0398	0.0404	1.9965	0.1562	0.0332
		Student-t $\nu = 6$			Student-t $\nu = 4$			IRLS		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	1	1.0007	0.0033	0.0061	1.0008	0.0030	0.0061	0.9978	0.0139	0.0128
B	2	2.0013	0.0039	0.0067	2.0007	0.0035	0.0061	2.0008	0.0100	0.0101
C	2	2.0030	0.0235	0.0439	2.0000	0.0211	0.0285	2.0007	0.0247	0.0243

(a) No Change

		χ^2			Logarithmic $N_s = 5$			Noise modeling		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	1	1.0011	0.0039	0.0106	1.0003	0.0301	0.0228	0.9819	0.1818	0.2425
B	2	2.0030	0.0048	0.0100	2.0034	0.0102	0.0099	2.0124	0.1790	0.1173
C	2	1.9882	0.0281	0.0503	2.0010	0.0581	0.0577	1.9946	0.2469	0.0599
		Student-t $\nu = 6$			Student-t $\nu = 4$			IRLS		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	1	0.9997	0.0036	0.0056	0.9995	0.0033	0.0057	0.9976	0.0220	0.0303
B	2	2.0027	0.0045	0.0095	2.0029	0.0040	0.0094	2.0023	0.0140	0.0178
C	2	1.9906	0.0262	0.0473	1.9905	0.0234	0.0471	1.9934	0.0356	0.0396

(b) Level Change

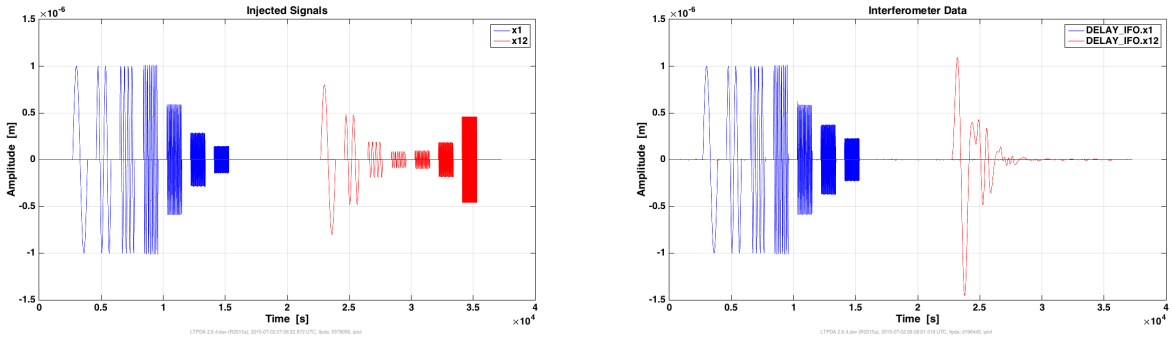
		χ^2			Logarithmic $N_s = 5$			Noise modeling		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	5	5.0001	0.0008	0.3956	4.9483	1.2287	0.8138	8.985	13.892	45.404
B	10	9.9432	0.0010	0.2207	9.9708	0.2161	0.2085	14.803	7.218	31.973
C	10	9.9969	0.0058	0.1546	9.9684	0.2527	0.2438	10.451	2.710	3.125
		Student-t $\nu = 6$			Student-t $\nu = 4$			IRLS		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
A	5	4.9596	0.1085	0.5083	5.0314	0.1190	0.7186	4.9752	1.1235	1.1034
B	10	9.9344	0.0459	0.2193	9.9342	0.0534	0.2185	9.9781	0.3019	0.3501
C	10	9.9900	0.0418	0.1526	9.9907	0.0522	0.1562	9.9780	0.1231	0.1111

(c) Gaussian Bursts

Figure 5: A table of the results from 50 estimations of the toy model amplitude parameters A, B, and C, from each the three different noise realizations during experiment. The "True" column gives the actual values of the parameters, $\hat{\mu}$ is the average of the mean values from the 50 the MCMC results, σ_{rms} is the root-mean-square of the standard deviations, and $\sigma_{\hat{\mu}}$ is the standard deviation of the mean values. Given a consistent, Gaussian spread of the parameters, σ_{rms} and $\sigma_{\hat{\mu}}$ should be identical.

4 Tests on LPF Models

Now we perform the same sorts of tests, but with simulated LPF data and an associated model instead of the toy sinusoidal signal and model. Recall the LPF system configuration from Figure 1. In order to simulate real LPF data for our analysis, we utilize a LPF simulator from the Albert Einstein Institute. To perform a noise measurement, we Let the simulator run without any injected signals. To obtain the data for our experiment, we perform signal injections. This means that we effectively "lie" to the controllers in the simulator, indicating interferometer readout that has no physical cause, then the spacecraft thrusters and electrodes around the second test mass activate to compensate, resulting in more dynamic data. You can see our injection scheme and the resulting data in Figure 6.



(a) Injection Signals

(b) Experiment Data

Figure 6: The signals injected into the interferometer readout, and the resulting data from the LPF simulation. On the left are the injected interferometer signals fed to the controller, x_{12} in blue and x_1 in red. On the right are the resulting interferometer outputs.

The main measurement of interest is the acceleration between the two test masses. Thus, the dynamics of the LPF system are captured in the following equation for the acceleration noise residuals:

$$n(t) = \frac{\partial^2}{\partial t^2} x_{12}(t) - \left(\delta_{IFO} \frac{\partial^2}{\partial t^2} x_1(t) + A_{sus} \frac{F_{cmd,2}(t-\tau)}{m_2} + \left(\omega_2^2 - \omega_1^2\right) x_1(t) + \omega_2^2 \left(x_{12}(t) - \delta_{IFO}\, x_1(t)\right) \right) \quad (19)$$

where $F_{cmd,2}$ is the net of the command forces applied to the second test mass, A_{sus} is an amplitude coefficient for the command suspension forces (expected to be close to unity), m_2 is the mass of the second test mass, τ characterizes the system delay between the measurement of test mass acceleration and the application of the command forces, ω_1 and ω_2 are the spring-like stiffnesses of the respective test masses, and δ_{IFO} is a coupling coefficient between interferometers x_1 and x_{12}. To confirm that Equation 19 accurately describes the noise residuals, we generate some data and plot the PSDs of previously measured noise with the residuals in Figure 7. As with the toy model, for further confirmation we performed the Kolmogorov-Smirnov test, and we find that our model passes.

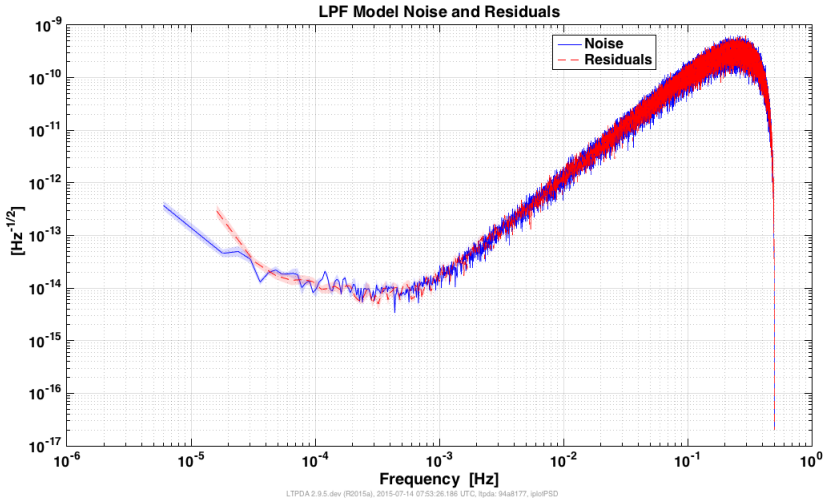


Figure 7: Here we plot the PSDs of the model results from Equation 19 in the case of no injections (blue) and with the injected signals shown in Figure 6 (red) along with their errors as shaded areas. Ideally they would be the same, but we see differences in the low-frequency range, probably due to the fact that we measured the noise for longer than we took data.

4.1 Noise Changes

Because the LPF simulator allows us to generate the signal and noise separately, we can perform a similar analysis to that with the toy model. Like with the toy model, we perform three experiments with different noise realizations, including no change in the noise, an overall change in noise magnitude, and embedding Gaussian bursts. See the estimation results in Figure 8. Here is a list of our observations of the data table:

- These parameter estimations are much better than with the toy model, with correct estimations ranging from 10^{-4} to 1 percent error.
- Each likelihood presents nearly identical errors.
- There is a slight overall increase in the errors in the case of a level change in the noise.
- The errors with each likelihood do not change much with the noise realization, meaning that all of these likelihood formulations are well-suited for simple LPF system identification.

		χ^2			Logarithmic $N_s = 5$			Noise modeling			Student-t $\nu = 54$			Student-t $\nu = 4$		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
τ	0.3	0.3000	3e-4	2e-4	0.3000	3e-4	3e-4	0.3000	3e-4	2e-4	0.3000	3e-4	2e-4	0.3000	2e-4	3e-4
A_{sus}	1	0.999998	5e-6	4e-6	0.999998	4e-6	4e-6	0.999998	5e-6	4e-6	0.999998	5e-6	4e-6	0.999998	4e-6	4e-6
ω_1^2 (e-6)	1.935	1.9349	9e-4	3e-4	1.9349	3e-4	2e-4	1.9349	9e-4	3e-4	1.9349	9e-4	3e-4	1.9350	7e-4	3e-4
ω_2^2 (e-6)	2	2.0000	2e-4	1e-4	1.9999	2e-4	1e-4	2.0000	2e-4	1e-4	2.0000	2e-4	1e-4	2.0000	2e-4	2e-4
δ_{IFO} (e-4)	1	1.0005	0.0038	0.0024	1.0005	0.0026	0.0029	1.0005	0.0038	0.0024	1.0005	0.0037	0.0024	1.0004	0.0029	0.0026

(a) No Change

		χ^2			Logarithmic $N_s = 5$			Noise modeling			Student-t $\nu = 54$			Student-t $\nu = 4$		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
τ	0.3	0.3000	3e-4	3e-4	0.2999	4e-4	4e-4	0.3000	4e-4	3e-4	0.3000	3e-4	3e-4	0.3000	3e-4	4e-4
A_{sus}	1	0.999997	5e-6	6e-6	0.999996	6e-6	6e-6	0.999997	6e-6	6e-6	0.999997	5e-6	6e-6	0.999997	4e-6	6e-6
ω_1^2 (e-6)	1.935	1.9351	9e-4	5e-4	1.9349	4e-4	4e-4	1.9351	0.0011	5e-4	1.9351	9e-4	5e-4	1.9350	8e-4	6e-4
ω_2^2 (e-6)	2	1.9999	2e-4	3e-4	1.9999	3e-4	3e-4	1.9999	2e-4	3e-4	1.9999	2e-4	3e-4	1.9999	2e-4	3e-4
δ_{IFO} (e-4)	1	0.9989	0.0038	0.0045	0.9992	0.0037	0.0047	0.9890	0.0050	0.0044	0.9989	0.0038	0.0045	0.9992	0.0034	0.0052

(b) Level Change

		χ^2			Logarithmic $N_s = 5$			Noise modeling			Student-t $\nu = 54$			Student-t $\nu = 4$		
θ	True	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$	$\hat{\mu}$	σ_{rms}	$\sigma_{\hat{\mu}}$
τ	0.3	0.3000	3e-4	2e-4	0.3000	3e-4	3e-4	0.3000	3e-4	1e-4	0.3000	3e-4	2e-4	0.3000	2e-4	3e-4
A_{sus}	1	0.999998	5e-6	4e-6	0.999998	4e-6	5e-6	0.999998	5e-6	4e-6	0.999998	5e-6	4e-6	0.999998	4e-6	4e-6
ω_1^2 (e-6)	1.935	1.9350	9e-4	4e-4	1.9350	3e-4	3e-4	1.9350	9e-4	4e-4	1.9350	9e-4	4e-4	1.9350	7e-4	4e-4
ω_2^2 (e-6)	2	2.0000	2e-4	2e-4	2.0000	1e-4	2e-4	2.0000	2e-4	2e-4	2.0000	2e-4	2e-4	2.0000	1e-4	2e-4
δ_{IFO} (e-4)	1	1.0000	0.0038	0.0026	1.0001	0.0026	0.0028	1.0000	0.0040	0.0026	1.0000	0.0037	0.0026	1.0000	0.0029	0.0026

(c) Gaussian Bursts

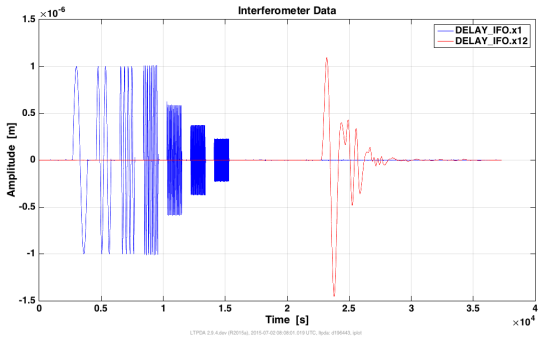
Figure 8: A table of the results from 50 estimations of the LPF parameters with the three different noise realizations.

4.2 Gaps in Data

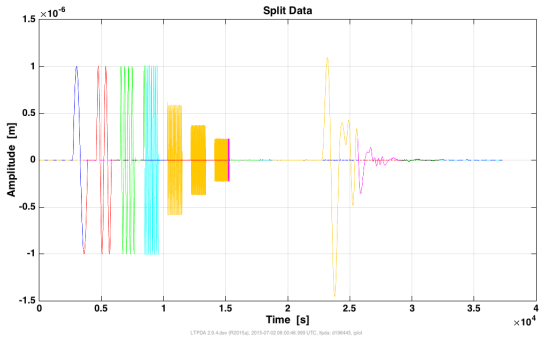
Another concern in the LPF experiment is missing data. Here we performed a single MCMC test on each of four situations: no gaps in the data, ten 5-second gaps in the data, two one-hour gaps in the data eliminating low-frequency injections, and two one-hour gaps in the data eliminating high-frequency injections. In computing the likelihood, we take each uninterrupted segment of data to be a separate experiment. (See Figure 9.)

We also chose this run to test variations in the fixed parameter N_s in the logarithmic likelihood. Recall in the logarithmic likelihood, we split the time series data into N_s stretches and average the PSDs of these stretches. The parameter N_s must be high enough such that the MCMC test converges to the true value, but also low enough to keep the errors at a minimum. See the numerical results in Figure 10 and the accompanying plots in Figures 18 and 19. Here is a list of our observations of the results:

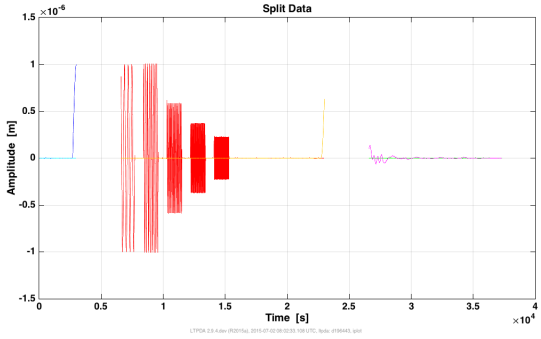
- The parameter estimations are still very good overall, with between $10^{-4}\%$ and 4% error.
- The ten five-second gaps made little difference in using the χ^2 likelihood, but caused a factor of two increase in the errors of the test mass stiffnesses with the logarithmic likelihoods.
- Missing data during low-frequency injections results in an order of magnitude increase in all errors
- Missing data during high-frequency injections causes only a slight increase in the logarithmic errors, and a decrease in the χ^2 errors.
- In the case of no data gaps, setting $N_s = 5$ is an acceptable choice, as we have been doing in previous estimations.
- In the case of gaps in the data, we are analyzing shorter segments, so we would expect to need a smaller N_s. We see that approximately $N_s = 3$ gives minimum errors in these cases.



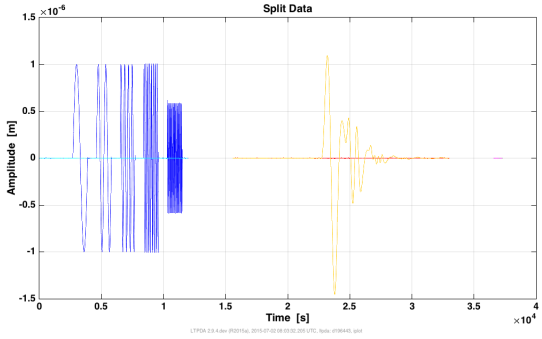
(a) No gaps



(b) Ten five-second gaps



(c) Two one-hour gaps (during low-frequency injections)



(d) Two one-hour gaps (during high-frequency injections)

Figure 9: Examples of the four different types of gaps we introduced into the data. Each color reperesents a different uninterupted segment of data. Note that the two interferometer channels are super-imposed in these graphs.

		χ^2		Logarithmic $N_s = 5$		$N_s = 4$		$N_s = 3$		$N_s = 2$	
θ	True	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
τ	-	0.2997	3e-4	0.2998	1e-4	0.2996	3e-4	0.2997	3e-4	0.2997	2e-4
A_{sus}	1	0.999992	5e-6	0.999988	3e-6	0.999985	5e-6	0.999982	5e-6	0.999991	4e-6
ω_1^2 (e-6)	1.935	1.9342	9e-4	1.9346	2e-4	1.9346	2e-4	1.9343	4e-4	1.9347	3e-4
ω_2^2 (e-6)	2	1.9998	2e-4	1.9998	2e-4	1.9997	2e-4	1.9995	3e-4	1.9998	2e-4
δ_{IFO} (e-4)	1	1.0030	0.0038	1.0022	0.0028	1.0026	0.0023	1.0022	0.0018	1.0018	0.0028

(a) No gaps

		χ^2		Logarithmic $N_s = 5$		$N_s = 4$		$N_s = 3$		$N_s = 2$	
θ	True	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
τ	-	0.3002	3e-4	0.3002	4e-4	0.3004	3e-4	0.3004	3e-4	0.3002	3e-4
A_{sus}	1	1.000006	8e-6	0.999997	8e-6	1.000001	5e-6	0.999998	7e-6	0.999999	5e-6
ω_1^2 (e-6)	1.935	1.9365	0.0012	1.9349	4e-4	1.9354	4e-4	1.9352	7e-4	1.9353	4e-4
ω_2^2 (e-6)	2	2.0010	0.0012	1.9999	4e-4	2.00005	3e-4	2.0000	6e-4	2.0003	3e-4
δ_{IFO} (e-4)	1	1.0029	0.0032	1.0015	0.0018	1.0007	0.0020	1.0008	0.0017	1.0054	0.0031

(b) Ten 5-second gaps

		χ^2		Logarithmic $N_s = 5$		$N_s = 4$		$N_s = 3$		$N_s = 2$	
θ	True	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
τ	-	0.3013	0.0010	0.2995	0.0012	0.3003	6e-4	0.3009	5e-4	0.3018	0.0010
A_{sus}	1	0.99993	8e-5	0.99993	6e-5	0.99997	4e-5	1.00004	2e-5	0.99992	5e-5
ω_1^2 (e-6)	1.935	1.9485	0.0131	1.9294	0.0121	1.9276	0.0177	1.9362	0.0127	1.9294	0.0131
ω_2^2 (e-6)	2	2.0144	0.0127	1.9947	0.0120	1.9933	0.0176	2.0044	0.0126	1.9962	0.0132
δ_{IFO} (e-4)	1	1.0013	0.0038	0.9971	0.0041	1.0013	0.0037	1.0071	0.0037	1.0068	0.0054

(c) Two one-hour gaps during low-frequency injections

		χ^2		Logarithmic $N_s = 5$		$N_s = 4$		$N_s = 3$		$N_s = 2$	
θ	True	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
τ	-	0.3002	2e-4	0.3000	4e-4	0.3002	3e-4	0.2999	4e-4	0.3003	3e-4
A_{sus}	1	1.000004	3e-6	0.999990	7e-6	0.999997	7e-6	1.000004	6e-6	0.999994	7e-6
ω_1^2 (e-6)	1.935	1.9351	2e-4	1.9348	5e-4	1.9346	6e-4	1.9353	3e-4	1.9350	4e-4
ω_2^2 (e-6)	2	2.0000	1e-4	1.9996	4e-4	1.9994	5e-4	2.0001	3e-4	1.9997	4e-4
δ_{IFO} (e-4)	1	1.0011	0.0038	1.0038	0.0031	1.0013	0.0031	1.0009	0.0025	0.9989	0.0031

(d) Two one-hour gaps during high-frequency injections

Figure 10: Results from the LPF parameter estimation with gaps in the data. Here μ is the estimate, and σ is the estimated error for the respective parameter.

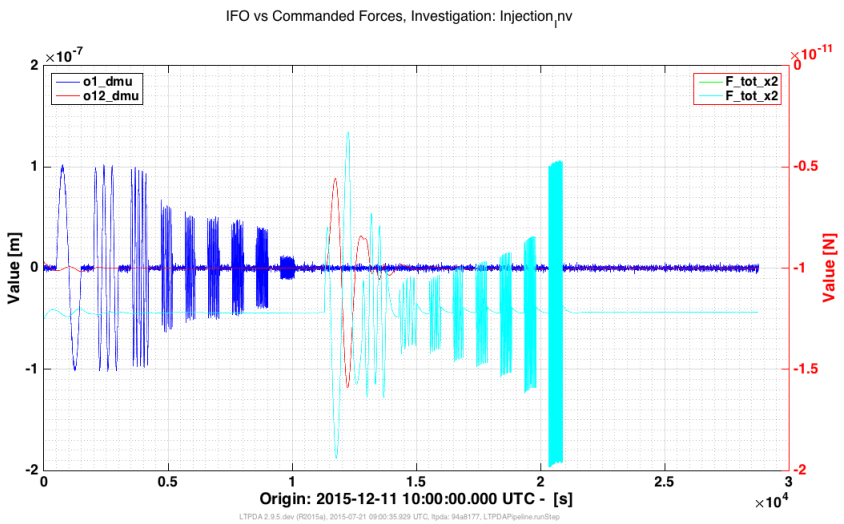


Figure 11: Here you can see the LPF data simulated by the European Space Agency. From the interferometer output in blue and red, we can see that they are using an injection scheme very similar to our own. We can also see the command forces on the second test mass in cyan.

4.3 Black box simulations

In order to verify our results made with our own LPF simulators, we perform similar analyses to data simulated by the European Space Agency, of which we do not have knowledge of the specifics of the model or the true values of any parameters used in generating the data. You can see plots of the data in Figure 11.

In order to estimate these unknown parameters, we test a slightly different linearized version of the model in Equation 19. Firstly, we omit the coupling parameter δ_{IFO} because we know it was not taken into account the data generation. In addition, the delay parameter τ is not linear, but for small values of τ, we can simulate a delay by taking a Taylor expansion of the command forces and adding coefficient parameters. We include the first two derivatives with respective coefficients C_1 and C_2:

$$n(t) = \frac{\partial^2}{\partial t^2}x_{12}(t) - \left(A_{sus}\frac{F_{cmd,2}(t)}{m_2} + C_1\frac{\partial}{\partial t}\frac{F_{cmd,2}(t)}{m_2} + C_2\frac{\partial^2}{\partial t^2}\frac{F_{cmd,2}(t)}{m_2} + \delta\omega^2 x_1(t) + \omega_2^2 x_{12}(t)\right) \quad (20)$$

where we estimate the parameter $\delta\omega^2 = \omega_2^2 - \omega_1^2$ instead of ω_1^2 on its own.

See the results of the parameter estimations in Figure 13 and the accompanying residuals in Figure 12. We notice that these parameter estimations are much less precise than those made with our own LPF data generation, with percent errors ranging from 10^{-2} to 25 percent error. The most imprecise measurement is that of $\delta\omega^2$, except in the case of the logarithmic likelihood and the IRLS method, both of which have much smaller error estimates in this parameter than the other likelihoods.

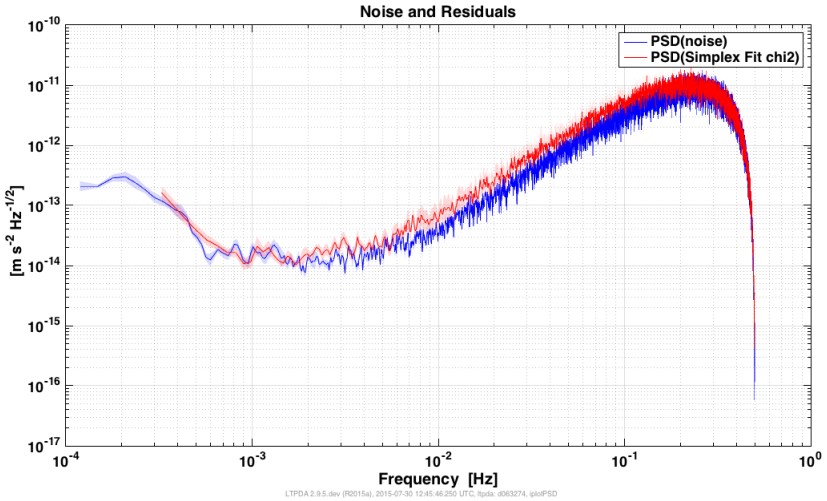


Figure 12: These are the residuals from the external generated data and the model evaluated at the estimated values. As you can see, this model does not do quite as well at eliminating the noise as in the case of the LPF simulations by the Albert Einstein Institute. It appears that there is a level change in the mid-frequency range which is not accounted for in our model.

	χ^2		Logarithmic $N_s = 5$		Noise modeling		Student-t $\nu = 14$		IRLS	
θ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
C_1	-0.9266	0.0057	-0.9287	0.0077	-0.9264	0.0090	-0.9257	0.0056	-0.9260	0.0074
C_2	-0.47	0.12	-0.585	0.125	-0.46	0.19	-0.45	0.12	-0.59	0.12
A_{sus}	1.0000	1e-4	1.0000	2e-4	1.0000	2e-4	0.9999	1e-4	1.0000	2e-4
$\delta\omega^2$ (e-7)	-8.409	1.383	-6.614	0.065	-8.430	2.194	-8.619	1.392	-6.573	0.068
ω_2^2 (e-7)	-9.642	0.059	-9.684	0.095	-9.642	0.094	-9.624	0.058	-9.571	0.118

Figure 13: Results from the LPF parameter estimation with data generated by the European Space Agency. Here μ is the estimate, and σ is the estimated error for the respective parameter.

5 Conclusions

In general all methods of parameter estimation work well in most cases. With both the toy model and the LPF simulations, all likelihoods prove to be accurate in recovering the system parameters. Even when we introduce noise not taken into account in the likelihood formulations, we still recover the correct parameter values with small relative error. Even with numerous or large gaps artificially introduced into the data, all methods remain robust in their estimations.

That said, each likelihood requires a fair bit of fine-tuning before we arrive at such nice results. For example, we must estimate the covariance in the parameters in order to decide the step sizes of our random walk in the parameter space, but each likelihood may need slightly different sizes of steps. Additionally, the student-t and level change likelihoods require manual tuning of the additional parameters introduced in those models. Finally,

the estimation results highly depend on the model used for the acceleration noise residuals. Omitting or adding a parameter can drastically affect the errors in the estimations. This is evidenced by our results from the externally generated data. Thus, one must take care in choosing the model for the fit.

Topics for further investigation include finding a way to determine the likelihood-specific parameters one should use before beginning the parameter estimation (e.g. the number of scaling parameters in the level change likelihood, the number of degrees of freedom in the student-t likelihood, or the number of averages in the logarithmic likelihood). This would greatly speed up the process, since currently these parameters have to be determined manually by trial and error. Further investigation should also include a comparison of parameter estimation results using different models to determine which parameters are important.

References

[1] D. S. Sivia. Data Analysis: A Bayesian Tutorial. Clarendon (Oxford Univ. Press), Oxford, 1996 (ISBN: 0-19-851762-9 or 0-19-851889-7 in paperback).

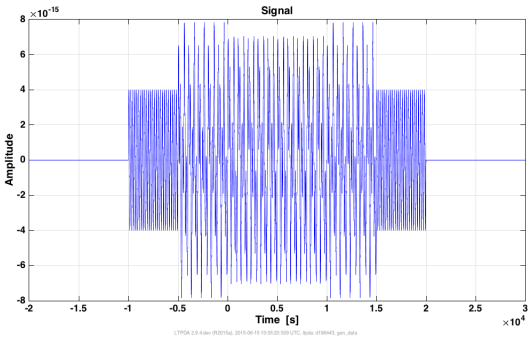
[2] Curt Cutler and Eanna E. Flanagan. Gravitational waves from merging compact binaries: How accurately can one extract the binary's parameters from the inspiral waveform? Phys. Rev. D, 49:2658?2697, Mar 1994.

[3] Tyson B. Littenberg and Neil J. Cornish. Bayesian approach to the detection problem in gravitational wave astronomy. Phys. Rev. D, 80:063007, Sep 2009.

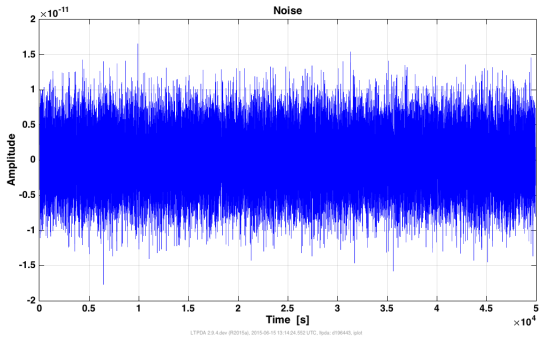
[4] Christian Röver, Renate Meyer, and Nelson Christensen. Modelling coloured residual noise in gravitational-wave signal processing. Classical and Quantum Gravity, 28(1):015010, 2011.

[5] Stefano Vitale, Giuseppe Congedo, Rita Dolesi, Valerio Ferroni, Mauro Hueller, Daniele Vetrugno, William Joseph Weber, Heather Audley, Karsten Danzmann, Ingo Diepholz, Martin Hewitson, Natalia Korsakova, Luigi Ferraioli, Ferran Gibert, Nikolaos Karnesis, Miquel Nofrarias, Henri Inchauspe, Eric Plagnol, Oliver Jennrich, Paul W. McNamara, Michele Armano, James Ira Thorpe, and Peter Wass. Data series subtraction with unknown and unmodeled background noise. Phys. Rev. D, 90:042003, Aug 2014.

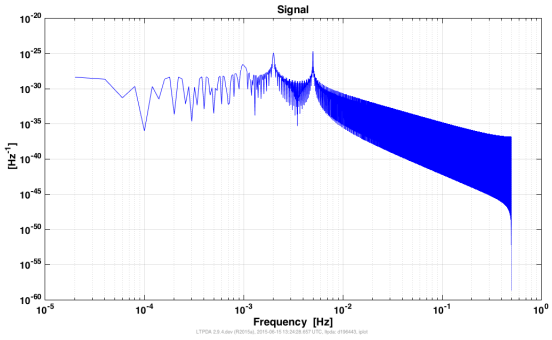
[6] J. Aasi, J. Abadie, B. P. Abbott, et al. Characterization of the LIGO detectors during their sixth science run. Classical and Quantum Gravity, 32(11):115012, 2015.



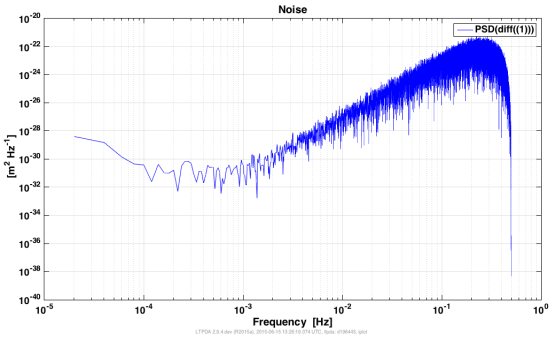
(a) Signal Time Series



(b) Measured Noise Time Series

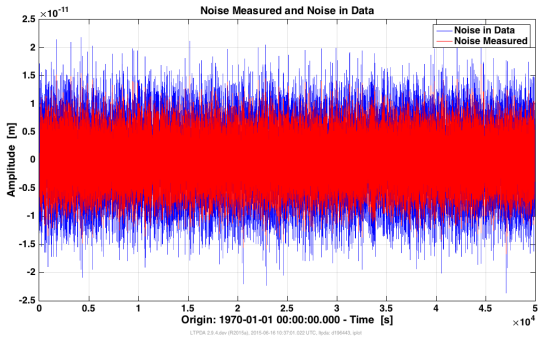


(c) Signal PSD

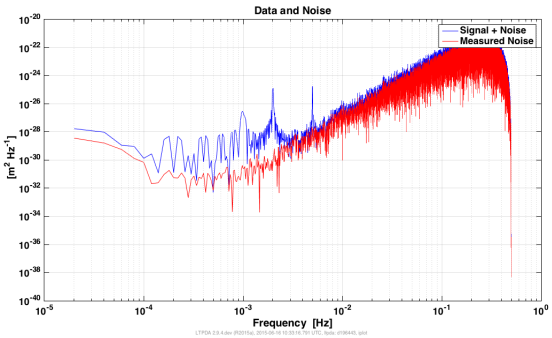


(d) Measured Noise PSD

Figure 14: The time series and power spectral densities (PSDs) of the signal and noise. On the left is the signal we recover using MCMC tests. Note the peaks from the sinusoidal signals in the PSD at 10^{-3}, 2×10^{-3}, and 5×10^{-3} Hz. On the right is a sample of LPF-like acceleration noise without any embedded signal. This is the noise we use to calculate the likelihoods in *all* MCMC tests.



(a) Noise Time Series



(b) Data and Noise PSDs

Figure 15: Noise and data in the case of increase in noise magnitude between measurement and experiment. Note the overall increase in magnitude of the data's PSD and time series.

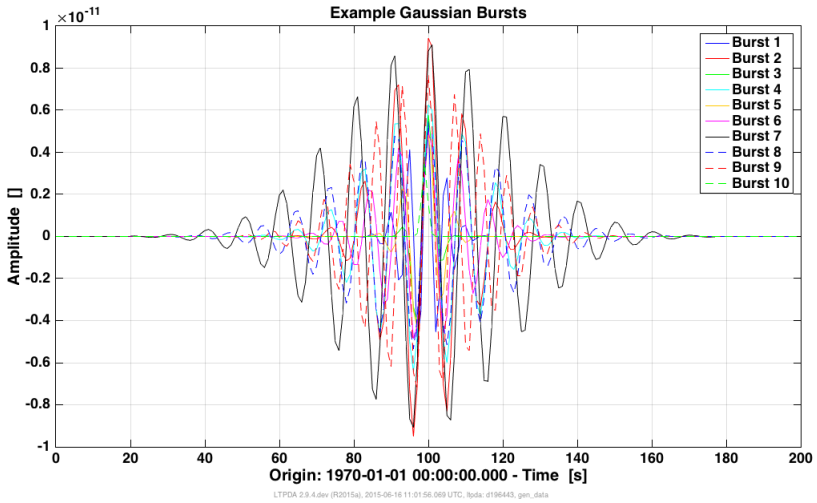
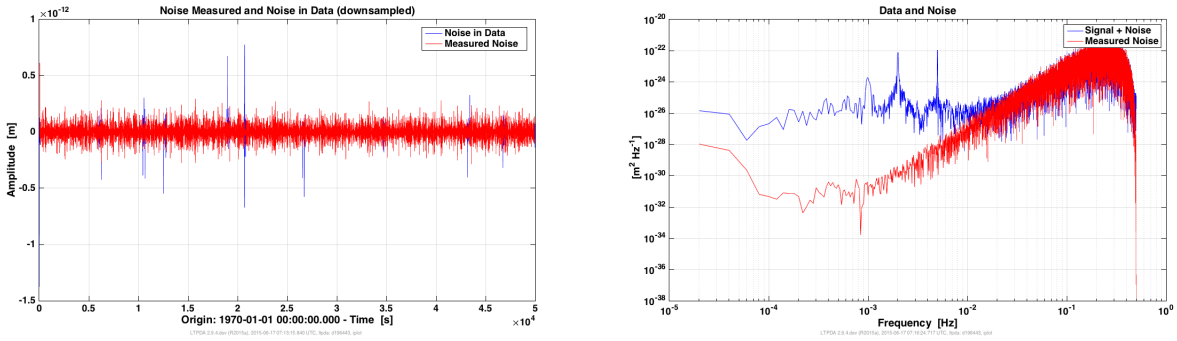


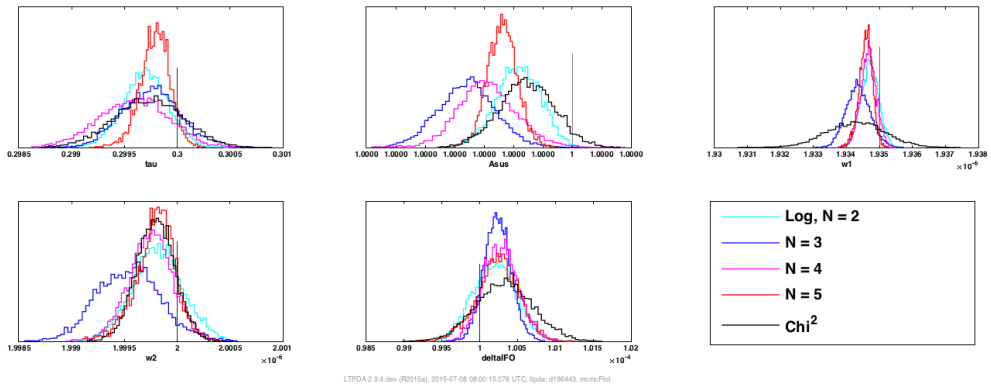
Figure 16: Gaussian burst examples



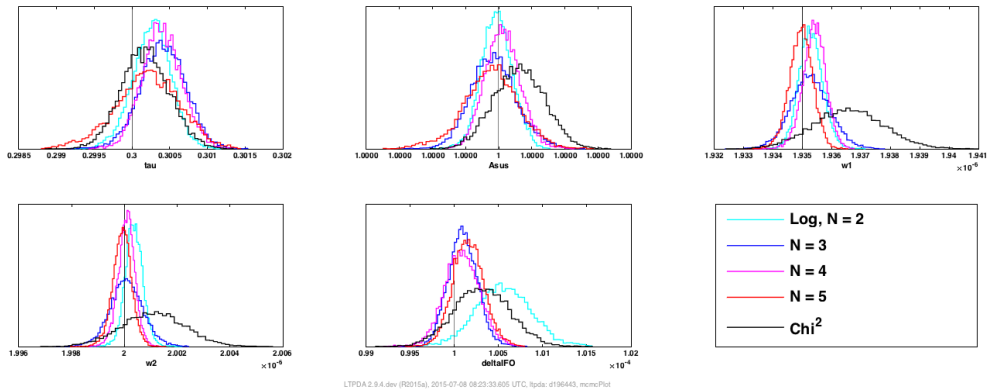
(a) Downsampled Noise Time Series

(b) Data and Noise PSDs

Figure 17: Noise and data in the case of Gaussian bursts embedded in the noise between measurement and experiment. On the left we see the noise time series', which have been downsampled to expose the short-duration Gaussian bursts in the noise added to the signal. On the right we see that the high-frequency data is indistinguishable from the measured noise, while at lower frequencies There is a much higher noise level. However, we can still see the frequency peaks of the sinusoidal signals.

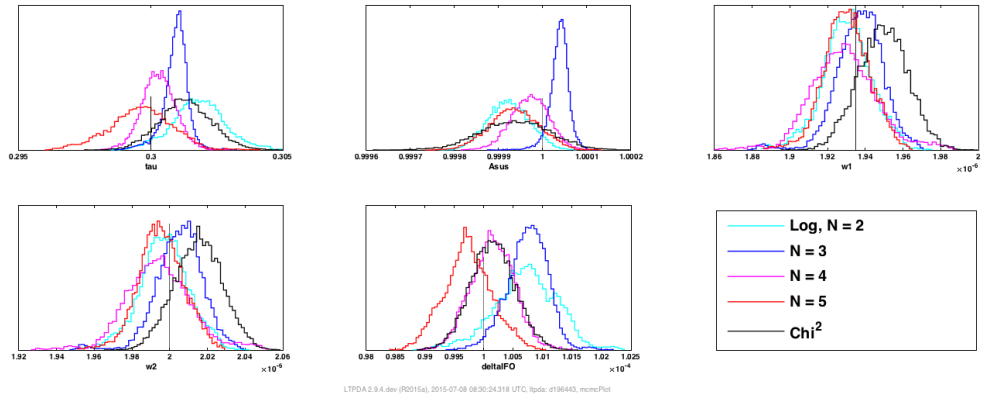


(a) No gaps

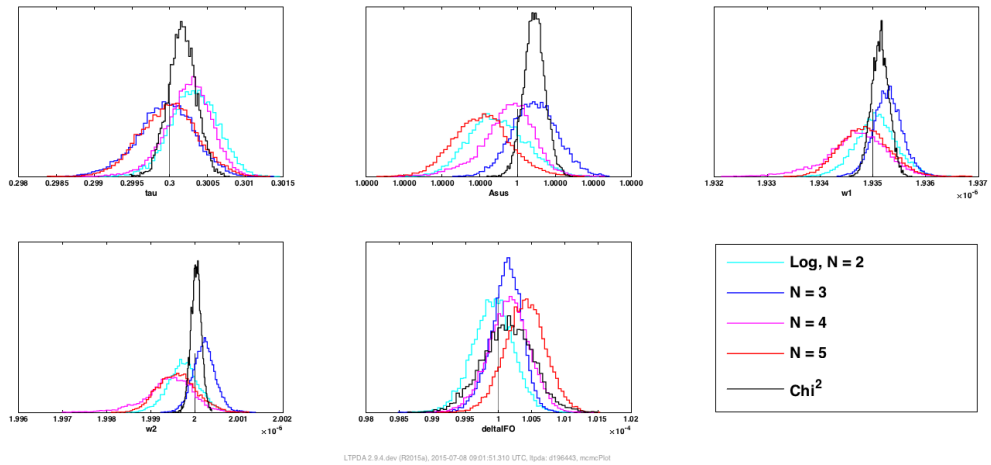


(b) Ten five-second gaps

Figure 18: The sample distributions from the MCMC tests in the case of no gaps and ten five-second gaps. We tested four different values of N_s for the logarithmic likelihood, as well as the χ^2 likelihood.



(a) Two one-hour gaps (during low-frequency injections)



(b) Two one-hour gaps (during high-frequency injections)

Figure 19: The sample distributions from the MCMC tests in the case of hour-long gaps covering low frequency and high frequency injections. We tested four different values of N_s for the logarithmic likelihood, as well as the χ^2 likelihood.