THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF PHYSICS

BAYESIAN ANALYSIS TECHNIQUES FOR GRAVITATIONAL WAVE DATA

PATRICK C. BREYSSE
Spring 2012

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Astronomy and Astrophysics and Physics
with honors in Physics

Reviewed and approved by the following[1]:
L. Samuel Finn
Professor of Physics
Thesis Supervisor

Richard Robinett
Professor of Physics
Honors Adviser

[1]Signatures are on file in the Schreyer Honors College

Abstract

Data from gravitational wave detectors suffer from a very low signal to noise ratio, which makes it difficult to find signals using conventional statistical techniques. Fitting a data set with conventional techniques will often fit measurement noise in addition to signal, leading to incorrect results. We explore a solution to this problem using Bayesian model comparison techniques. A given data set is first fit with each of a variety of models using Bayes' law, then the models are ranked based on their evidence, which is a measure of the probability that the model accurately represents the source of the data. The model fitting stage has been shown to be effective on a sample problem of radioactive decay. It was then applied to looking for the signal of black hole ringdown. The evidence ranking was also tested on the ringdown problem, but it has not yet been proven to be effective.

Contents

List of Figures

List of Tables

1 Introduction

One of the most significant ways in which general relativity differs from Newtonian gravity is the prediction of gravitational waves. Einstein's theory predicts that accelerating masses produce oscillations in the space time metric which propagate away from the source at the speed of light, analogous to the electromagnetic waves produced by accelerating charges. The existence of gravitational waves has been inferred indirectly by studying the orbital periods of binary pulsars, but they have yet to be directly detected. Efforts are underway to detect gravitational waves using large Fabry-Perot-Michelson interferometers (Abbott et al. 2009). One of the most prominent efforts is the Laser Interferometer Gravitational Wave Antenna (LIGO), a set of three interferometers in two locations in the United States. Direct detection of gravitational waves has not yet been achieved, largely because - at present detector sensitivity - strong signals are very rare. We have investigated the use of Bayesian statistical analysis techniques for the reliable detection of weak signals, with the goal of applying these methods to the problems of analyzing data from the LIGO detectors.

1.1 Properties of Gravitational Waves

Just as Maxwell's equations describe the relationship between charges and electromagnetic fields, Einstein's equations describe the interactions between masses and spacetime (Abbott et al. 2009). Time dependent vacuum solutions to Maxwell's equations describe electromagnetic waves, and time dependent vacuum solutions to Einstein's equations describe gravitational waves. However, unlike electromagnetic waves, gravitational waves are quadrupolar rather than dipolar, which means that the wave compresses space along one transverse direction and expands it along another. Gravitational radiation is produced by oscillating multiple moments of a system's mass distribution. However, the

principles of conservation of mass and conservation of momentum rule out any asymmetric monopolar or dipolar oscillations, so the quadrupole radiation is thus dominant. Gravitational waves propagating far from any source can be described in a linearized approximation as small perturbations to the special relativistic spacetime metric $\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$ (Hartle 2003). If we write the perturbed metric as

$$g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} \tag{1}$$

then for gravitational waves propagating in the z direction, $h_{\alpha\beta}$ can be written as $h_{\alpha\beta}$ is

$$h_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} f(t-z) + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} g(t-z) \tag{2}$$

where $f(t-z)$ and $g(t-z)$ are any smooth functions of $t-z$ such that $|f|$ and $|g|$ are much less than unity. The metric perturbation $h_{\alpha\beta}$ is typically referred to as the *strain*.

As suggested by the form of the perturbation shown in Eqn. (2), the gravitational wave effect is transverse to the direction of wave propagation (in this case the z direction) and traceless. (Abbott et al. 2009). The strain has two independent polarization states referred to as the + (plus) polarization and the $\times$ (cross) polarization, represented here by the two independent functions f and g.

Gravitational waves in nature are typically very weak. At the linear approximation the strain $h_{\alpha\beta}$ is given by

$$h_{ij} = \frac{2G}{c^2 r} \left| \ddot{Q}_{ij} \right|^{\mathrm{TT}} \tag{3}$$

where i and j are the spatial indices of $h_{\alpha\beta}$, r is the distance to the source, G and c are Newton's gravitational constant and the speed of light, $\ddot{Q}_{ij}$ is the second time derivative of the source's mass quadrupole moment and the superscript TT indicates that one should take the transverse (to the direction of wave propagation) and traceless part of $\ddot{Q}_{ij}$. Dimensionally this equation may be written

$$h \sim \frac{GM}{c^2 r} M\omega^2 L^2 \tag{4}$$

where M is the source mass engaged in quadrupolar acceleration, ω is the characteristic angular frequency of the motion, and L is the source size. For astrophysical systems $(\omega L)^2 \sim v^2 \sim GM/c^2 L$; correspondingly, the strongest sources will be those that pack the greatest amount of mass into the smallest region of space. For example, two stellar mass black holes or neutron stars 1 Mpc away orbiting each other at 100 Hz will lead to a strain on the order of 10^{-21}.

1.2 Detecting Gravitational Waves

Observations of the binary pulsar PSR 1913+16 provided the first indirect evidence for the existence of gravitational waves. General relativity predicts that such a system would gradually lose energy to gravitational wave emission, causing the neutron stars to drift closer and closer together at an ever accelerating rate. Careful measurements of the orbital period of PSR 1913+16 shows that its orbital radius decreases at a rate closely corresponding to the predictions of general relativity. These observations show us the action on this binary system of what is almost certainly gravitational radiation; however, it is not a detection of the gravitational waves themselves.

One way to directly detect the very small strain produced by a gravitational wave is to use a large Fabry-Perot-Michelson interferometer. Figure 1 shows the effect of a passing gravitational wave on a ring of test masses and a Michelson

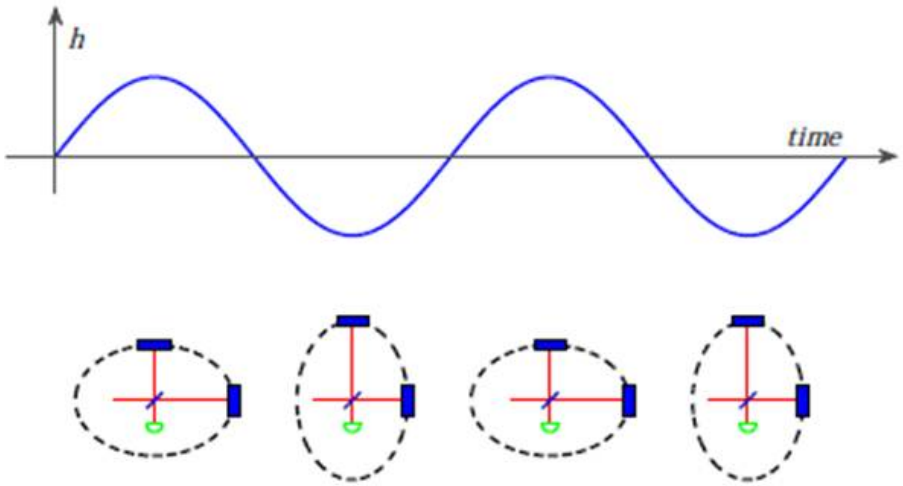

Figure 1: The effects of a gravitational wave passing through a Michelson interferometer and a ring of test masses. The arm lengths of the interferometer and the locations of test masses are altered in a plane perpendicular to the direction the wave is traveling. The change in arm lengths is observed by studying the changes in light intensity at the photodetector (From Abbott et al. 2009)

interferometer. The passing wave changes the relative lengths of the two arms of the interferometer, changing the phase shifts of the beams passing down each arms which can be detected even at very small strain amplitudes.

The interferometers used for LIGO are designed to be sensitive to strains as small as 10^{-21}. With 4 km arms, the resulting change in arm length is only 10^{-18} m, which is only a fraction of the diameter of a proton (Abbott et al. 2009). The strain sensitivity of the LIGO detector corresponds to the signal expected from binary neutron star systems at the distance of the Virgo Cluster of galaxies. Since sources like these are expected not more frequently that once per hundred years, we need more sensitive detectors if we are to detect gravitational astrophysical gravitational wave sources like these. Figure 2 shows the noise curves for both the current generation LIGO interferometers (labeled as LIGO I) as well as the planned upgrades (labeled as LIGO II). Signal curves

for various potential sources are plotted as well. Whenever a signal curve passes above the noise curve, it means that a signal from a random direction with a random orientation is detectable with a false alarm probability of less that 1%.

Studying these noisy data is a nontrivial task. Conventional frequentist statistics are ill suited to finding individual, weak, and rare signals. We propose a method based on Bayesian statistics to better locate and characterize gravitational wave signals in LIGO data.

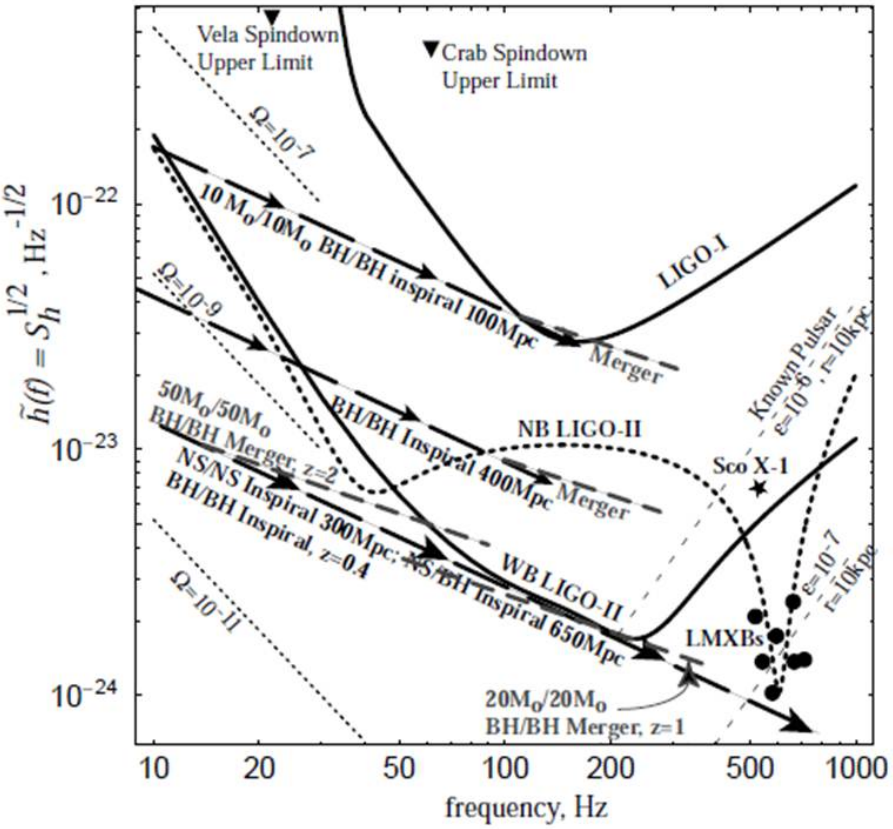

Figure 2: The noise curves of the current generation LIGO I and planned LIGO II interferometers plotted with the signal curve of several different gravitational wave sources. Whenever a signal curve passes above the noise curve, it means that a signal from a random direction with a random orientation is detectable with a false alarm probability of less that 1% (From Cutler & Thorne 2002)

2 Bayesian Model Comparison

Bayesian Model Comparison is a method uniquely suited to the task of interpolating a noisy data set. There are often many possible models which could be used to fit the data, each corresponding to a different possible source of the signal. While it may be fairly straightforward to fit the data with each model using conventional fitting techniques, determining which model provides the "best" fit is much harder to do. The best model is not simply the model that provides the closest fit. Models of increasing complexity will allow themselves to fit the observations ever more closely, even to the point of fitting themselves to the measurement noise.

The model comparison method outlined by MacKay (1991) provides a solution to the problem of over-fitting using Bayesian statistical techniques. Bayesian statistics differ significantly from classical frequentist statistics. Frequentist statistics view probabilities as the frequencies of random events, while Bayesian statistics view probabilities as providing quantification of uncertainty (Bishop 2006). In situations when over-fitting is a major problem, Bayesian techniques much more effectively capture the odds that a particular model is an appropriate representation of the signal embedded in the data.

The easiest way to intuitively understand why is in terms of the principle of Occam's razor. Occam's Razor states that the best solution to a problem is not necessarily the most complex. Bayesian statistics embody the idea that simpler models are preferable to more complex models. However, where Occam's razor is merely an intuition that may or may not be correct, Bayesian analysis is a proper statistical treatment which does not rely on untestable assumptions. These techniques provide quantitative answers to both the problem of model fitting and model comparison without showing undue preference to complex, overparameterized models. The method described here is a more general version

of the method used by Finn and Lommen (2010).

2.1 Model Fitting

Bayesian inference is carried out in two steps: Model fitting and model comparison (MacKay 1991). In the first step, a variety of models H_i are invented and the data are fit with each of them. A "model" in this context refers to a function which predicts data values based on the values of one or more parameters. One way to look at it is that a model H_i gives the probability $P(D|\mathbf{w}, H_i)$ that the parameter values $\mathbf{w}$ will yield the data set D. This probability is often referred to as the likelihood of model H_i with parameters $\mathbf{w}$. Fitting a data set with a model means finding the parameter values that were most likely used to create the data set. This is done by maximizing a quantity called the posterior probability $P(\mathbf{w}|D, H_i)$, which gives the probability that the parameters $\mathbf{w}$ were used to generate data set D given model H_i.

To better understand the concepts of models and posteriors, consider a very simple example of drawing different types of fruit from a bag (Bishop 2006). The different models in this case could be the number of types of fruit present in the bag, while the parameters for each model would be the number of each type of fruit. One model might assume that there are apples, bananas, and oranges in the bag, and the parameters would be the number of apples, the number of bananas, and the number of oranges. The likelihood is the probability of drawing a particular fruit out of the bag given the assumed types of fruit present. The posterior distribution then gives the probability distribution of the numbers of each fruit after one or more have been picked.

The posterior distribution is calculated from the likelihood using a formula called Bayes' Law. Bayes law describes how the conditional probability of an event behaves when the condition and the event are interchanged: i.e., the

relationship between P(A—B) — the probability of event A given event B has occurred — and the probability of event B given that event A has occurred. Bayes law says that

$$P(A|B)P(B) = P(B|A)P(A) \tag{5}$$

where $P(B)$ and $P(A)$ are the probabilities of B and A without any conditions.

For example, suppose that A is the event of rolling two dice that sum to 6 and B is the event that at least one of the two dice has come-up a five. There are ten ways that only one of the dice is a five, and one way that both dice are a five, and thirty-six ways that the dice can turn-up; so $P(B) = 11/36$. Of the 11 ways that at least one can be a five, there are two ways that the sum can be a six; so $P(A|B) = 2/11$. There are also nine ways that the sum of two dice can be 6; so, P(A) = 1/4. Finally, of these nine ways, there are two in which at least one is a five; so, $P(B|A) = 2/9$. From this example we can verify Bayes' law:

$$P(A|B)P(B) = P(B|A)P(A) = 1/18 \tag{6}$$

In Bayesian analysis, the probabilities P(A) and P(B) are known as the prior probabilities, usually shortened to just "priors". The priors represent the information known before any data are taken. The prior distribution over the parameters of a model gives the values that the parameters could plausibly take a priori. For example, in the fruit-in-a-bag problem described above, you might place limits on the total number of fruit inside based on the size of the bag before drawing any out. Priors are a very important concept in Bayesian analysis.

Consider Bayes' Law for the problem of model fitting. For the case of a model $H_i$with parameter values given by the components of the vector $\mathbf{w}$, the posterior

probability distribution over the possible parameter values can be written as

$$P(\mathbf{w}|D,H) = \frac{P(D|\mathbf{w},H)P(\mathbf{w}|H_i)}{P(D|H_i)}. \tag{7}$$

The most probable values of the parameters of H_i are found by determining the maximum of the posterior distribution $P(\mathbf{w}|D,H)$. $P(D|\mathbf{w},H_i)$ is the likelihood of data set D for model $H_i$and parameters $\mathbf{w}$ and $P(\mathbf{w}|H_i)$ is the prior distribution for model H_i. The prior distribution is the probability known before observing the data set (Bishop, 2006). For example, in the classical probability experiment involving drawing different colored marbles from a bag, the prior distribution would be the best guess for the number of marbles of each color before drawing any, The denominator term, $P(D|H_i)$, is called the evidence for model H_i and serves as the normalizing constant for the posterior distribution. This value is usually ignored during the model fitting stage, since all values of $\mathbf{w}$ have the same evidence. However, it becomes very important in the model comparison stage.

2.2 Model Comparison

Model comparison is the task where Bayesian statistics really shine. If there is no reason to assign preference to particular models, models are ranked based on the evidence value $P(D|H_i)$ from the denominator of Eqn (7). This quantity is used to calculate the posterior probability of each model, again from Bayes' Law:

$$P(H_i|D) \propto P(D|H_i)P(H_i). \tag{8}$$

The evidence is simply the normalizing constant for the posterior probability:

$$P(D|H_i) = \int P(D|\mathbf{w},H_i)P(\mathbf{w}|H_i)d\mathbf{w}. \tag{9}$$

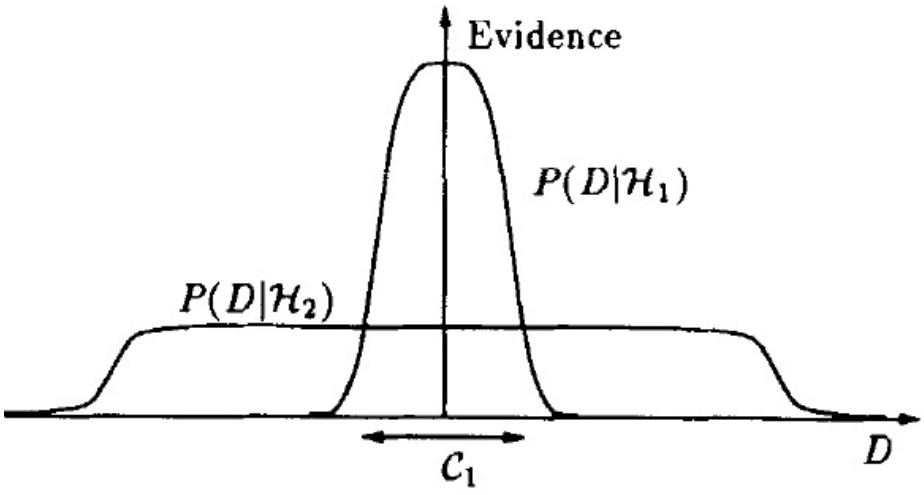

Figure 3: An illustration of how the evidence calculation applies Occam's Razor. The horizontal axis is the space of possible data sets, and the vertical axis is the evidence $P(D|H_i)$. The simpler model, H_1, has fewer free parameters than H_2, so it cannot predict as great a variety of data sets. However, within the range C_1, H_1 predicts the result more strongly than H_2. This means that if the data falls withing C_1, the simpler model will be more probable than the more complex model. (From MacKay 1991).

The quantity $P(H_i)$ is a subjective prior which expresses how probable various models are before examining data. This subjective quantity is typically overwhelmed by the objective evidence term. Thus the evidence is usually the only term that matters when comparing models. Figure 3 shows how this quantity embodies Occam's razor. While a more complicated model can fit a wide variety of data sets, certain data sets will be better fit by a simpler model.

The integral needed to determine the evidence can often be difficult to evaluate. Fortunately, for most problems, the posterior distribution has a sharp peak at the most probable parameters. Thus the evidence is approximately equal to the height of the posterior at $\mathbf{w}_{mp}$ times its width $\Delta\mathbf{w}$:

$$P(D|H_i) \simeq P(D|\mathbf{w}_{MP}, H_i)P(\mathbf{w}_{MP}|H_i)\Delta\mathbf{w} \tag{10}$$

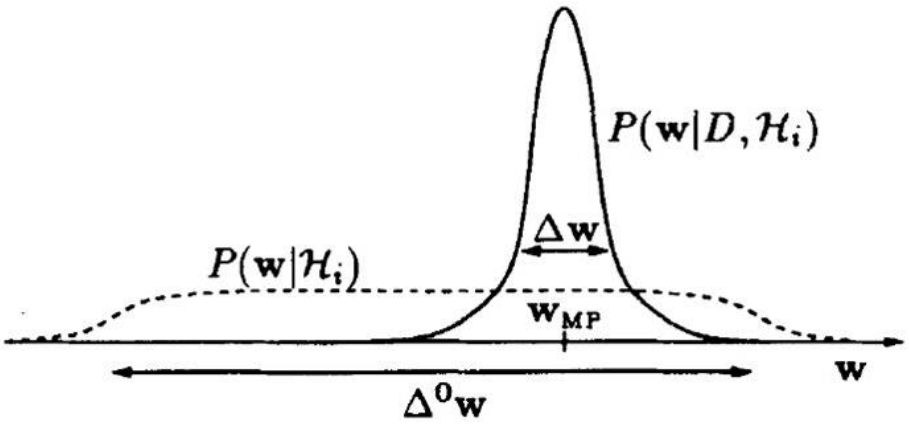

Figure 4: A plot of a sharply peaked posterior distribution $P(\mathbf{w}|D, H_i)$ and its prior $P(\mathbf{w}|H_i)$. The prior is flat for a range of parameter values $\Delta^0\mathbf{w}$. The Occam factor is equal to the ratio of the width of the peak $\Delta\mathbf{w}$ to $\Delta^0\mathbf{w}$ (From MacKay 1991)

The quantity $P(\mathbf{w}_{MP}|H_i)\Delta\mathbf{w}$ is called the Occam factor. This is the term which penalizes the model $H_i$for having the parameter $\mathbf{w}$. The quantity $\Delta\mathbf{w}$ is the posterior uncertainty in $\mathbf{w}$. In the simple case where the prior $P(\mathbf{w}|H_i)$ is uniform on a large interval $\Delta^0\mathbf{w}$ (Figure 4), the prior probability of $\mathbf{w}_{MP}P(\mathbf{w}_{MP}|H_i) = 1/\Delta^0\mathbf{w}$ and the Occam factor is equal to $\Delta\mathbf{w}/\Delta^0\mathbf{w}$. This is the factor by which H_i's hypothesis space collapses when the data are observed. The Occam factor relates to the complexity of the predictions a model makes in data space. Models with many parameters which can vary over a large $\Delta^0\mathbf{w}$ are penalized, as are models which have to be very finely.

If a model has k parameters and is well approximated by a Gaussian, the evidence can be approximated by

$$P(D|H_i) \simeq P(D|\mathbf{w}_{MP}, H_i)P(\mathbf{w}_{MP}|H_i)(2\pi)^{k/2}\det{}^{-1/2}\mathbf{A} \qquad (11)$$

where $\mathbf{A} = -\nabla\nabla \log P(\mathbf{w}|D, H_i)$ The evidence is the product of the best fit likelihood and the Occam factor. This approximation becomes more and more

exact as the number of data points increases as a consequence of the central limit theorem.

3 Sample Radioactive Decay Problem

The data fitting methods outlined in Section 2.1 were first tested on a sample problem which was much simpler than gravitational wave analysis. Consider a sample of an unknown radioactive material. The sample could be made of one or more component substances with different decay rates. The only tool available for analysis is a Geiger counter which clicks every time an atom in the sample decays.

This sample problem can easily be expressed in the terms of Bayesian analysis. The data being studied consist of the times when the Geiger counter records a decay. The models in this case are each characterized by the number of substances present in the sample. One model says that all of the atoms in the sample decay at the same rate, one says there are two different types of atom with unique decay rates, and so on. The parameters of the model are the amounts and decay constants (e-folding times) of each substance.

The first step in fitting this type of data set is to calculate the likelihood of the data set given the parameters. Consider first the case of binned data, where the data set consists of the number of events occurring in each a series of time intervals of length T. Such a data set might look like this: 10 events occurred between $t = 0$ and $t = 1$, 5 events occurred between $t = 1$ and $t = 2$, and so forth. The average decay rate (also the average rate of Geiger counter clicks) at time t is given by:

$$\mu(t) = \sum_i \frac{N_{0i} e^{-t/\tau_i}}{\tau_i} \tag{12}$$

where N_{0k} and τ_k are the initial amount and decay constant for substance i and t is the time since data taking began. The expected number of events ϵ in

between time t and time $t+T$ is

$$\epsilon = \int_t^{t+T} \mu(t')dt' = \int_t^{t+T} \sum_i \frac{N_{0i}e^{-t'/\tau_i}}{\tau_i} dt' \tag{13}$$

$$\epsilon = \sum_i \left(N_{0i}e^{-t/\tau_i} - N_{0i}e^{-(t+T)/\tau_i} \right) \tag{14}$$

From this, it is possible to calculate the likelihood of a single data bin. If n events are measured between t and $t+T$, the likelihood is

$$P(n|N_{0ki}, \tau_i, t, T) = \frac{\epsilon^n}{n!} e^{-\epsilon} \tag{15}$$

which is the probability that n events were measured in that bin given the parameters, the start time of the bin, and the bin size. The likelihood of the whole data set is simply the product of the likelihoods for each bin.

However, the likelihood would be much more accurate if it were calculated using unbinned data. Binned data adds an extra parameter T which can change the likelihood. The most straightforward way to adjust this likelihood calculation for use with unbinned data is to look at a data set with the time of each event and consider the time between each event to be a "bin" with zero events and the infinitesimal time interval around each event to be a "bin" with one event. Then the likelihood is the probability that zero events occur between times t_n and t_{n+1} times the probability that only one event occurs between t_{n+1} and $t_{n+1} + dt$. From Eqn. (15), the former is simply

$$P(0|N_{0i}, \tau_i, t_n, t_{n+1}) = e^{-\epsilon} \tag{16}$$

where ϵ has become

$$\epsilon = \sum_i N_{0i} \left(e^{-t_n/\tau_i} - e^{-t_{n+1}/\tau_i} \right). \tag{17}$$

The probability of only one event occurring within dt of time t_{n+1} is

$$P(1|N_{0i}, \tau_i, t_{n+1}, dt) = \sum_i \frac{N_{0i}}{\tau_i} e^{t_{n+1}/\tau_i}. \tag{18}$$

Therefore, the probability that an event occurs at t_{n+1} given the parameter values and an event at time t is

$$P(t_{n+1}|N_{0i}, \tau_i, t_n) = e^{-\epsilon} \sum_i \frac{N_{0i}}{\tau_i} e^{t_{n+1}/\tau_i}. \tag{19}$$

And again, the likelihood of the entire data set is just the product of the likelihoods of each event.

The next step to calculate the posterior is to determine the prior. There is no obvious reason to assume a certain amount of radioactive material or decay constant a priori, so the prior in this case is an "ignorance prior". The ignorance prior $P(w)$ over parameter w is calculated by finding a distribution which is invariant under the transformation:

$$P(w)dw = P(w')dw' \tag{20}$$

For example, consider scaling the parameter τ_1 by a constant factor α so that $\tau_1' = \alpha\tau_1$. To make this transformation invariant, it must be true that

$$P(\tau_1)d\tau_1 = P(\tau_1')d\tau_1' \tag{21}$$

$$P(\tau_1) = \alpha P(\alpha\tau_1). \tag{22}$$

Differentiating with respect to α gives

$$0 = \alpha\tau_1 \frac{dP(\alpha\tau_1)}{d\alpha} + P(\alpha\tau_1). \tag{23}$$

Since α is just an arbitrary constant, we can set it equal to 1 and solve the differential equation for P. This yields

$$P(\tau_1) = \frac{1}{\tau_1}. \tag{24}$$

Carrying out this process for the other parameters yields the same result, for example, the prior for a two-substance model is

$$P(N_{01}, N_{02}, \tau_1, \tau_2) = \frac{1}{N_{01} N_{02} \tau_1 \tau_2}. \tag{25}$$

From this, the posterior distribution for this two substance model is

$$P(N_{01}, N_{02}, \tau_1, \tau_2 | t_n, t_{n+1}, H) = \prod \left[\frac{e^{-\epsilon}}{N_{01} N_{02} \tau_1 \tau_2} \left(\frac{N_{01}}{\tau_1} e^{-t_{n+1}/\tau_1} + \frac{N_{02}}{\tau_2} e^{-t_{n+1}/\tau_2} \right) \right]. \tag{26}$$

Fitting a data set requires calculating which parameter values maximize this posterior distribution. The distribution calculated above was tested using sample data sets generated with known initial amounts and decay constants. One such sample data set is shown in Figure 5. The data set was generated for 1000 atoms of a single substance with a decay constant of 5. Maximizing the posterior for a one substance model gave an initial amount of 989.79 and a decay constant of 5.15, which are fairly close to the correct values. The fit curve generated with these parameters is plotted over the data in Figure 3.1, and the posterior distribution is shown in Figure 6.

Another sample data set is shown in Figure 7. This one was generated for two substances with initial amounts 100 and 1000 and decay constants 10 and 1. This data set was fit with two different models. Fitting for a single substance yielded an initial amount of 1130.2 and a decay constant of 1.92. The curve for this model is the blue curve in Figure 3.3. As expected, this curve does not fit

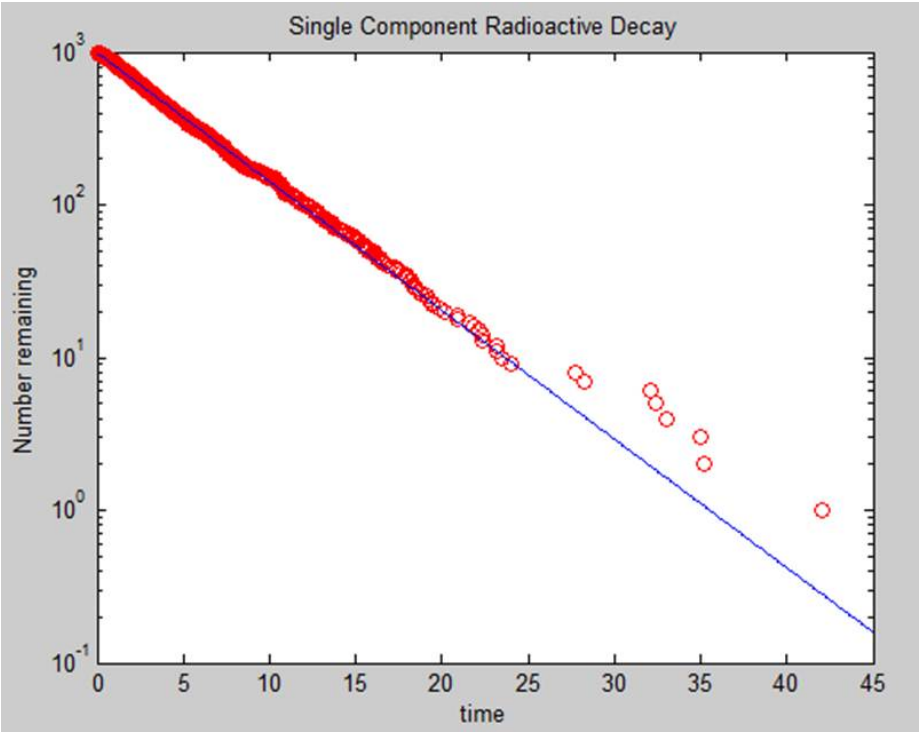

Figure 5: Sample data set for the radioactive decay of 1000 atoms one material with a decay constant of 5. Fit parameters calculated by maximizing the posterior are $N_0 = 989.79$ and $\tau = 5.15$.

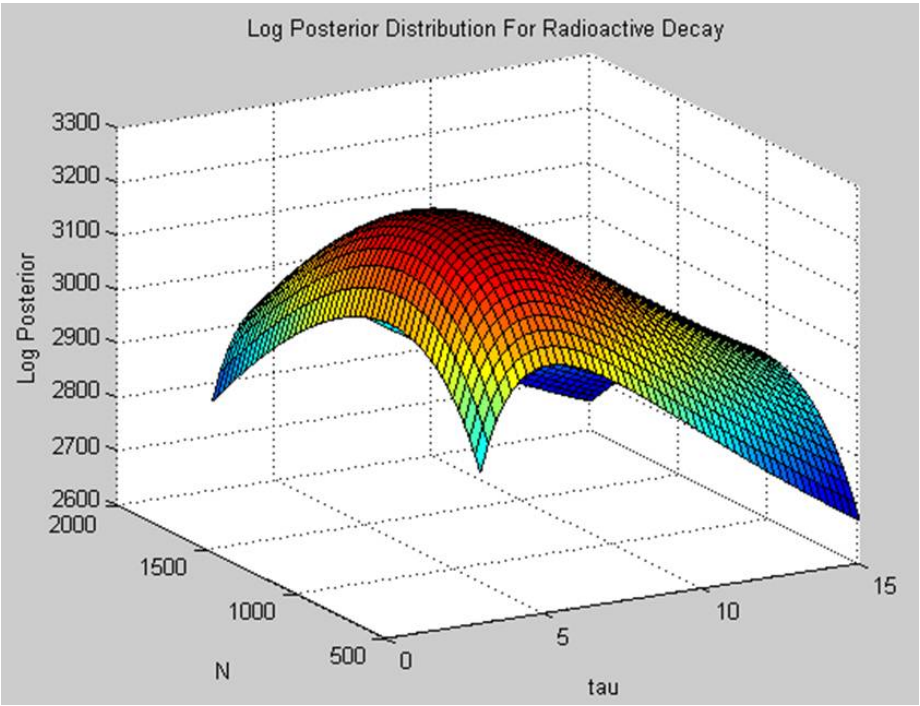

Figure 6: Log of the posterior probability distribution of the one substance model used to fit the data set in Figure 5. The distribution has a maximum at $N = 989.79$ and $\tau = 5.15$

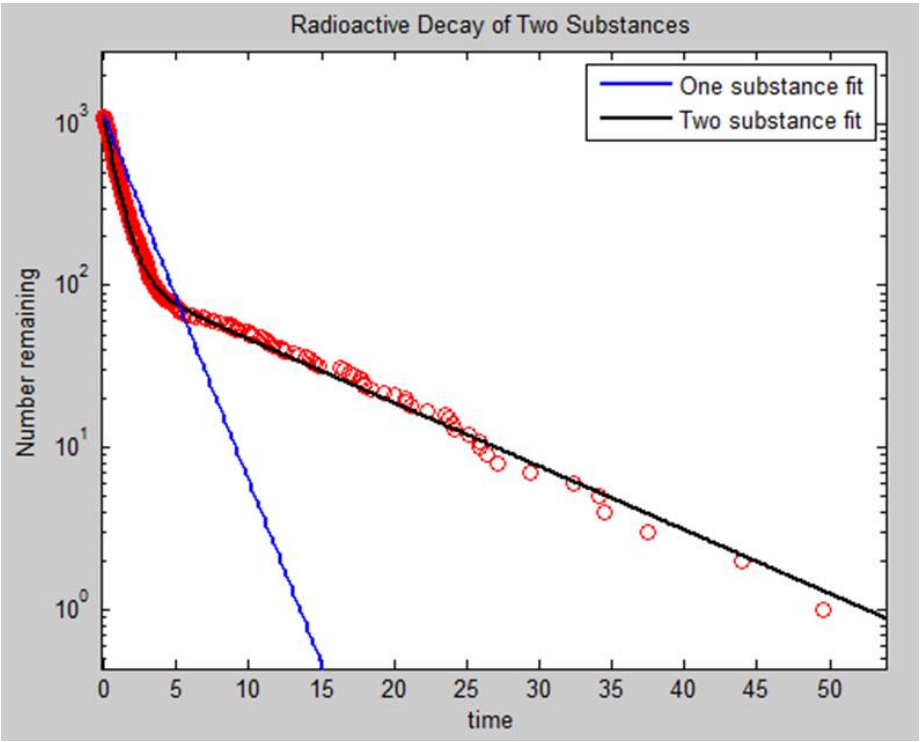

Figure 7: Sample data set for the radioactive decay of 100 atoms with a decay constant of 10 and 1000 atoms with a decay constant of 1. Fit parameters for the one substance fit are $N_0 = 1130.2$ and $\tau = 1.92$. Fit parameters for the two substance fit are $N_{01} = 114.74$, $N_{02} = 952.63$, $\tau_1 = 11.08$, and $\tau_2 = 0.88$.

the data very well, so the single substance model is clearly too simple for these data. Fitting for two substances yielded initial amounts of 114.74 and 952.63 and decay constants of 11.08 and 0.88, and the curve for this model is the black curve in Figure 7. These values closely match the correct values, and the curve fits the data well. These results demonstrate that Bayesian methods are able to fit curves well, and illustrate the effect of an insufficiently complex model.

Unfortunately, since these sample data sets do not include very much noise, they cannot be used to demonstrate the effect of overparameterizing. Attempting to fit the data in Figure 7 with a three substance model yields the the same parameters as the two substances from the two substance fit plus one substance with a negligible amount of atoms which decay almost instantly, so the fit curve is essentially identical to that of the two substance model.

4 Black Hole Ringdown

4.1 Overview

We applied this Bayesian method to search for the gravitational wave signal from black hole ringdown in a data set with Gaussian white noise. When a black hole merges with another or is otherwise perturbed, the perturbations are radiated away from the black hole as gravitational waves (Goggin 2006). The characteristic signal of these ringdown waves is given by

$$h = \sum_k A_k e^{-(t-T_k)/\tau_k} sin(\omega_k(t-T_k)+\phi_k)H(t-T_k) \tag{27}$$

The signal takes the form of a set of quasinormal exponentially decaying sinusoidal modes, each one with amplitude A_k, frequency ω_k, phase ϕ_k, and decay timescale τ_k. $H(t-T_k)$ is the heaviside function, so T_k is the time when each mode "turns on". t is the vector of times for each point. The values of ω and τ are uniquely determined by the mass and spin of the black hole. For this problem, each "model" corresponds to a different number of modes.

The likelihood for a ringdown model is assumed to be a Gaussian distribution. For a single data point, the peak of the Gaussian is located at the point

$$d - h = 0$$

where d is the strain of the data point and h is the strain which would be produced by a given set of parameter values for that model at the time of that point. The width of the Gaussian is determined by the standard deviation of the white noise σ. This quantity is assumed to be known a priori. The likelihood Λ for a single data point is thus

$$\Lambda = e^{-(d-h)^2/2\sigma^2}. \tag{28}$$

For a full data set defined at the times in the vector t, this likelihood is the product of the likelihoods for each point:

$$\Lambda = e^{-(d-h)\bullet(d-h)/2\sigma^2}. \tag{29}$$

For this work, all possible parameter values were assumed a priori to be equally probable, i.e. the prior was assumed to be uniform, so the posterior is equal to the likelihood. The most probable parameter values are found by maximizing the likelihood. This was accomplished using numerical methods, particularly the GlobalSearch function in MatLab.

The evidence for a given model is calculated using the approximate formula from Eqn. (7). We again assume flat priors, i.e. no model is held to be more likely than any other before the data are observed. This yields the following:

$$P(D|H_i) = \Lambda(\mathbf{w}_{MP})(2\pi)^{k/2}\det{}^{-1/2}\mathbf{A} \tag{30}$$

where $\mathbf{A} = -\nabla\nabla\log\Lambda\mathbf{w}$ and k is five times the number of modes in the model. $\Lambda(\mathbf{w}_{MP})$ is the peak of the likelihood distribution. This value is often very small, so it is more convenient to work with the log of the evidence instead:

$$\log(P(D|H_i) = \log\Lambda(\mathbf{w}_{MP}) + \frac{k}{2}\log(2\pi) - \frac{1}{2}\log(\det\mathbf{A}). \tag{31}$$

The Hessian $\mathbf{A}$ was calculated numerically using the derivest package in MatLab.

4.2 Examples

Figure 8 shows a simple data set fit by maximizing the likelihood from Eqn. (29). The data was generated by a single mode, and it was fit by a single mode model. Table 1 gives the parameter values used to generate the data and the

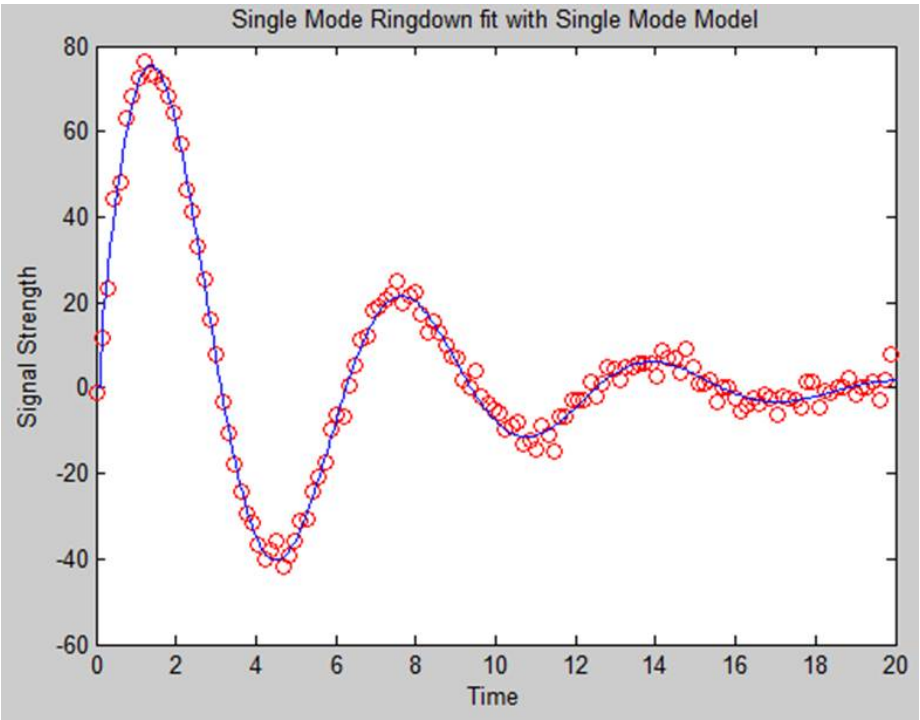

Figure 8: Sample data set generated for single mode ringdown fit with a single mode model. Parameter values used to generate the data and values calculated by maximizing the likelihood are shown in Table 1. The noise used to generate these data had a standard deviation of 2

most probable parameter values calculated by maximizing the likelihood. The resulting values resemble closely the values used to generate the data and the plotted fit curve follows the data well.

The next example fits a data set with several different models and calculates the evidence for each one. The sample data set used for this calculation is shown in Figure 9. It was generated using a two mode model with $\sigma = 10$ white noise. This data set was fit by three different models by maximizing the posterior distribution for each one: one model with one mode, one with two modes, and one with three modes. The three different fits are shown in Figures 10, 11, and 12, and the most probable parameter values are listed in Table 2.

Table 1: Parameters used to generate the data shown in Figure 8 and parameters calculated by maximizing the likelihood for the data set

Parameter	Data Value	Fit Value
Amplitude A	100	98.42
Decay Timescale τ	5	4.99
Frequency ω	1	1.00
Start time T	0	0.15
Phase ϕ	0	0.14

Table 2: Parameters used to generate the data shown in Figure 9 and parameters calculated by maximizing the likelihood for the data set

Parameter	Data Values	One Mode Fit Values	Two Mode Fit Values	Three Mode Fit Values
Amplitude A	100, 200	279.6	113.8, 179.9	364.1, 87.6, 151.5
Decay Timescale τ	8, 5	2.48	7.42, 5.18	2.69, 8.57, 3.64
Frequency ω	1, 2	1.86	1.01, 2.00	1.09, 1.97, 0.58
Start time T	3, 3	4.37	2.86, 3.17	3.47, 3.95, 3.17
Phase ϕ	0, 3.14	6.14	6.00, 3.45	5.91, 5.29, 2.59
Log(Evidence)		-1117	-457.5	-496.1

As expected, the parameter values for the two mode fit are similar those used to generate the data (note that the phase ϕ is periodic from 0 to 2π, so a value of 6.00 is fairly similar to a value of 0). Also, the one mode fit does not follow the data as closely, especially around $t = 12$, while the three mode fit follows the data the closest, particularly around $t = 4$. This matches the predictions of Bayesian analysis well, The single mode model is too simple to fit the data well, and the three mode model is overparameterized and fits the noise in addition to the signal.

Table 2 also lists the log of the evidence calculated for the three fits in Figures 10, 11, and 12. As expected, the two mode fit has the highest evidence. However, these values must be taken with a grain of salt. The MatLab routine

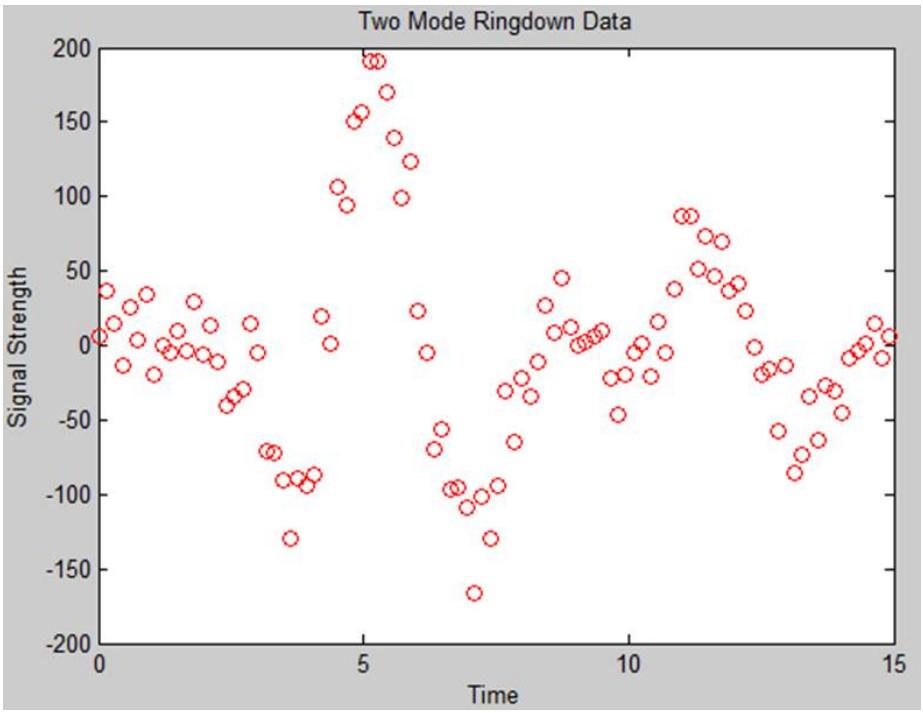

Figure 9: Sample two mode ringdown data set generated to test evidence calculations. Parameter values used to generate model are shown in Table 2

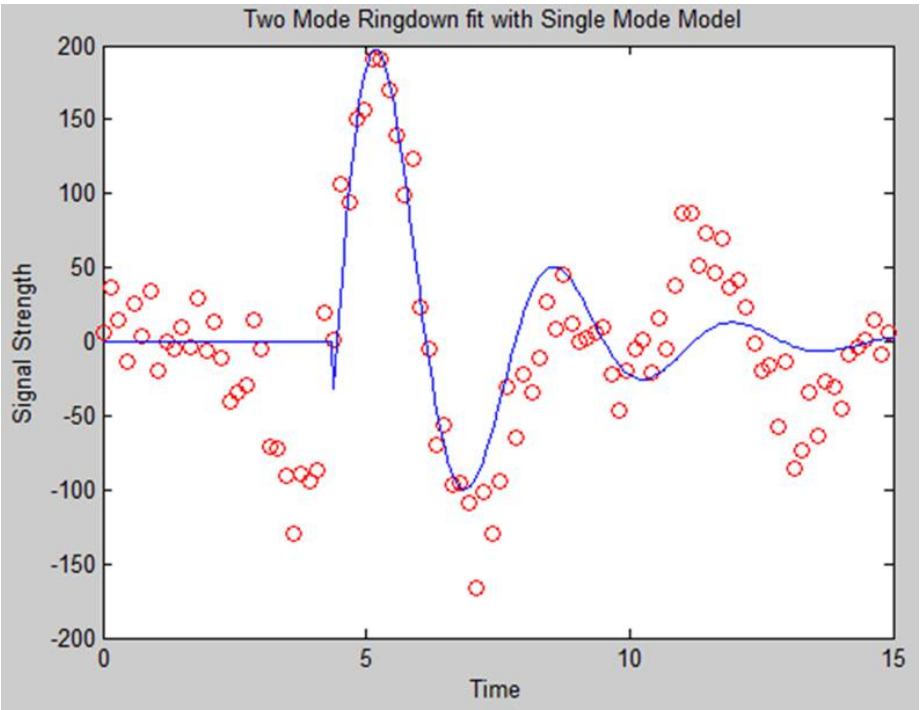

Figure 10: Two mode Data set from Fig. 9 fit with a one mode model. Fit parameter values and the evidence for the model are shown in Table 2.

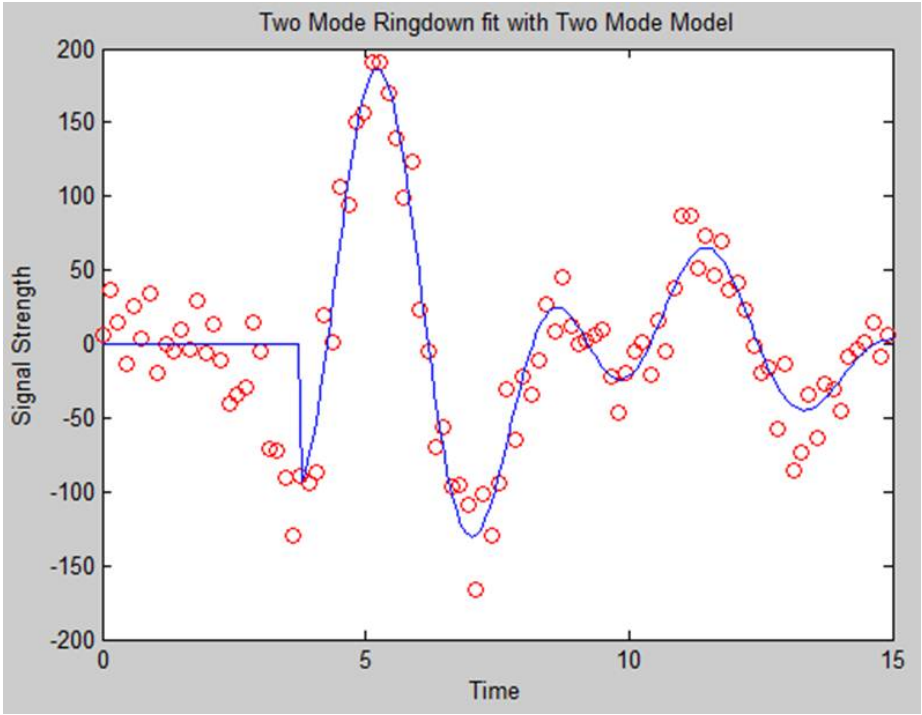

Figure 11: Two mode Data set from Fig. 9 fit with a two mode model. Fit parameter values and the evidence for the model are shown in Table 2.

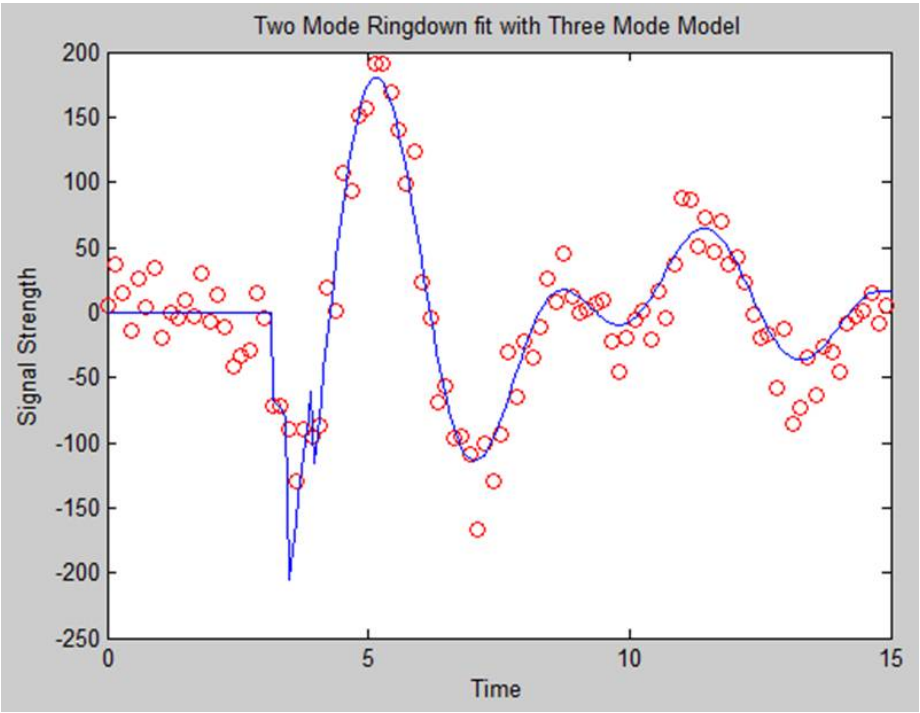

Figure 12: Two mode Data set from Fig. 9 fit with a three mode model. Fit parameter values and the evidence for the model are shown in Table 2.

used to calculate the evidence often returns different evidence values for the same data set with the same model, sometimes with a different model receiving the highest evidence. It remains to be seen if this is a failure of this method or a result of the approximations made to calculate the evidence. In particular, the Hessian $\mathbf{A}$ was calculated numerically rather than analytically, which could be the cause of the discrepancy since the function used to calculate it is not ideal for formulas with a large number of parameters. More testing is necessary to determine the exact cause of this problem and to determine what is necessary to repair it.

5 Conclusion and Future Work

We have investigated a way of analyzing gravitational wave data using Bayesian Model Comparison. This method consists of two stages: model fitting and model comparison. We demonstrated the effectiveness of model fitting by maximizing the posterior distribution on a sample problem of radioactive decay, and were able to successfully reproduce the parameters used to generate sample data sets. The model fitting method was also tested by searching for the gravitational wave signal of black hole ringdown, and it was similarly effective. We also tested the method of model comparison by maximizing evidence on the problem of black hole ringdown. While we were able to select the correct model in many cases, the incorrect model was selected too many times to claim conclusive success. More testing is necessary to determine the exact source of this problem. It may be the result of the numerical methods used to calculate the Hessian $\mathbf{A}$, in which case another way must be found to calculate the Hessian.

The next step for this work will be to calculate posterior distributions for a wide variety of different signals. There is a large effort underway to determine the waveforms of the gravitational wave signals from different sources using both analytical and numerical relativity. Calculating posteriors for these different waveforms is necessary in order to search for other signals besides black hole ringdown.

The work presented here assumes flat prior distributions for the case of black hole ringdown, but that may not be the case when studying real data. In the case of ringdown, since the parameters depend on the mass and angular momentum of the black hole, if certain masses and momenta are more common in nature there will be a nonuniform prior distribution for the parameter values. Similarly, it may be that some numbers of modes may be more common than others, so there would be a nonuniform prior on the evidence values as well.

Also, the presence of other signals in the data could affect the probability. For example, a ringdown signal will always be produced after a binary black hole merges. That means that if a ringdown signal is detected, the probability of a inspiral and merger signals immediately beforehand is increased. Most of these prior effects would need to be derived from the equations of general relativity and cosmic number density estimates.

6 Acknowldgements

The author would like to thank Dr. L. Samuel Finn and the Penn State Gravitational Wave Astronomy group for their help with this project, as well as Dr. Richard Robinett for all of his advice. Thanks also to the authors of the derivest Matlab package which was used for numerical calculations of Hessian matrices. Thanks

7 References

Abbott, B. P., Abbott, R., Adhikari, R., et al. 2009, Reports on Progress in Physics, 72, 076901

Bishop, Christopher M. 2006. Pattern Recognition and Machine Learning
Finn, L. S., & Lommen, A. N. 2010, ApJ, 718, 1400

Goggin, L. M., & LIGO Scientific Collaboration 2006, Classical and Quantum Gravity, 23, 709

Hartle, J. B. 2003, Gravity / James B. Hartle. San Francisco, CA, USA: Addison Wesley, ISBN 0-8053-8662-9, 2003, XXII + 582 pp.,

MacKay, David J.C. 1991, Neural Computation, 4, 415-447

Patrick C. Breysse Academic Vita

11963 Harford Rd. / Glen Arm, MD 21057 / pcb139@psu.edu / 443-631-3445

Education:

Perry Hall High School, 2004-2008

Pennsylvania State University, 2008-present
Senior in the Schreyer Honors College
Double major in Physics and Astrophysics
Projected Graduation in May 2012

Research Experience:

Dr. Robert Cammarata, Electrodeposition of nanomaterials, Johns Hopkins University Department of Materials Science and Engineering
Summer 2009 and 2010
Developed a process for creating thin film nanocomposites with alternating layers of pure metal and metal with embedded ceramic particles. Assisted in sample analysis using X-ray diffraction and SEM.

Dr. Lee Samuel Finn, Gravitational Wave Astronomy, Penn State University Department of Physics
Spring 2010-present
Developing a method to apply Bayesian analysis techniques to gravitational wave data.

Awards and Scholarships:

Dean's List, Fall 2008, Spring 2009, Fall 2009, Spring 2010, Fall 2010, Spring 2011
Astronaut Scholarship, 2011
Penn State Shibley Blue Band Scholarship, 2011
Penn State Physics Department REU, 2011
Schreyer Honors College Scholarship, 2008-present
Penn State Physics Department Scholarship, 2010
National Merit Scholar, 2008-2009
Valedictorian of Perry Hall High School Class of 2008

Publications and Presentations:

Stephen L. Farias, Patrick C. Breysse, Chai-Ling Chien, and Robert C. Cammarata. (2009). Uniform and Functionally Graded Oxide-Dispersion Strengthened Nanocomposite Thin Films. Presented at the 2009 Fall Meeting of the Materials Research Society

Stephen L. Farias, Patrick C. Breysse, Chai-Ling Chien and Robert C. Cammarata (2011). Layered and Functionally Graded Nanocomposite Thin Films with Unique Mechanical Properties. MRS Proceedings, 1304, mrsf10-1304-z11-05 doi:10.1557/opl.2011.610