Searching in the Dark Chasing Magnum Opus

Jacob Lange, Chi Nguyen, Daniel Wysocki *Rochester Institute of Technology*

Statistical Methods for Astrophysical Sciences (ASTP-611)
Spring 2016

The detection of gravitational waves (GW) on September 14, 2015 (Abbott et al. 2016), opened a new window into the Universe. Joint observations of electromagnetic (EM) counterparts to the GW signals will result in a deeper understanding of the physics involved. We anticipate EM counterparts to binary sources involving matter, i.e. neutron star–neutron star and neutron star–black hole binaries. In an effort to measure these joint signals, GW detectors are working in conjunction with traditional observatories.

This report is organized as follows. In §1, we describe the estimation of the chirp mass distribution. In §2 we describe an EM followup classifier based on the data.

1 Chirp Mass Distribution

1.1 Density Estimation

We use a histogram to estimate the merger rate as a function of $\mathcal{M}_c$, using Knuth's rule to determine the bin size (See Figure 1). Since we are interested in the intrinsic rate, not just that of detected events, we weigh each point by the inverse of the spacetime volume in which we are sensitive to it, $w = 1/VT$. Binaries with a higher chirp mass are easier to detect, so we do not want to count them as heavily. For a given chirp mass, we are sensitive out to a distance

$$D(\mathcal{M}_c) = 200\,\mathrm{Mpc}\big(\mathcal{M}_c/1.2\,\mathrm{M}_\odot\big)^{5/6} \tag{1}$$

which corresponds to a volume

$$V(\mathcal{M}_c) = \frac{4}{3}\pi D^3(\mathcal{M}_c). \tag{2}$$

Multiplying this by the time spent observing, $T = 0.6\,\mathrm{yr}$, gives us the spacetime volume $V(\mathcal{M}_c)T$.

To obtain uncertainties in our histogram, we take the square root of the sum-of-squares of the weights within that bin, i.e.

$$\sigma_k = \sqrt{\sum_i w_i^2}, \tag{3}$$

which was taken from Wiss (1995). This reduces to $\sqrt{N}$ in the case of an unweighted histogram, as $w_i = 1$, so $\sum_i w_i^2 = N$.

We also over-plot a pure power law. To do this, we employ Bayesian linear regression, fitting a straight line to $\log r$ versus $\log \mathcal{M}_c$, and transforming back to linear space. This is also shown in Figure 1.

1.2 The Likelihood of Fitting Parameters

For our fit, we assume a 9^{th} degree polynomial, and use a smoothing prior to prevent over-fitting. We fit this model to the points from the histogram rate, $r(x)$, obtained in §1.1. The data likelihood is

$$\ln p(\{r(x_k)\}_k|\lambda) = -\frac{1}{2}\sum_k \left\{ \frac{[r(x_k) - F_\alpha(x_k)\lambda_\alpha]^2}{\sigma^2_{r(x_k)}} \right\}, \tag{4}$$

where $F_\alpha(x) = x^\alpha$.

$\ln p_{\mathrm{smooth}}$ is the smoothing prior, defined to be:

$$\ln p_{\mathrm{smooth}}(\lambda) = -\gamma \int_{x_{\min}}^{x_{\max}} \left[\frac{\mathrm{d}^n F_\alpha(x)}{\mathrm{d}x^n} \lambda_\alpha \right]^2 \mathrm{d}x \quad (5)$$

In theory, p_{smooth} can be any n^{th} derivative. To make our code robust, we define a function that takes n as an argument. The function then calls `numpy.polynomial.polynomial.polyder()` to find the n^{th} derivative. Next, we square the n^{th} derivative and integrate it between the minimum and maximum of x. Here we choose $n = 3$, and let $\gamma = 0.5$.

1.3 Model fitting

We employed a Markov Chain Monte Carlo (MCMC), using `emcee` (Foreman-Mackey et al. 2013), to fit the coefficients of a polynomial model for $r(\mathcal{M}_c)$. The MCMC uses the smoothing prior as defined in §1.2. We performed a least squares fit first in order to obtain the initial guess for the coefficients of the model.

The model we used is a polynomial of degree 9. Because of the smoothing function, using higher order does not change the shape of the fit in a significant way. The MCMC best fit can be seen in Figure 2. The MCMC walkers and triangle plot are in Figures 5 and 6.

The best fit model is

$$r(\mathcal{M}_c) = \sum_{k=0}^{9} \alpha_k (\log \mathcal{M}_c)^k, \quad (6)$$

where the coefficients, α_k, are listed in Table 1.

Table 1: Coefficients from the MCMC fit.

k	α_k
0	-1.903×10^{-6}
1	1.842×10^{-5}
2	9.999×10^{-1}
3	1.748×10^{-4}
4	-2.441×10^{-4}
5	2.122×10^{-4}
6	-1.130×10^{-4}
7	3.470×10^{-5}
8	-5.416×10^{-6}
9	3.470×10^{-7}

2 Classification

2.1 Overview

LIGO can provide very rapid mass estimates of candidate GW events. Since most of these detections are binary black holes, and EM followup is extremely expensive, very few events are expected to have confirmed EM counterparts.

We have trained a classifier to determine if an event will have a EM counterpart. In §2.2 we discuss the method used, and in §2.3 we discuss our results.

2.2 Method

We trained this classifier on the first half of the data, simply taking the mid-way point in $\mathcal{M}_c$ between the population of events with EM counterparts and without. This is demonstrated in Figure 7.

We took a slightly different approach for the 2D mass distribution, as part of the **500% extra credit** problem. Here we trained a linear SVM, using `sklearn.svm.LinearSVC`, with `C=100`, and the masses transformed into log-space. Again, we performed the training on half of the data, and made correct predictions on the full data set. The results are shown in Figure 9.

2.3 Results

The classifier correctly separated the two groups with 100% completeness and zero contamination, even when the training was done using only half of the data set. Of course, this is potentially sensitive to precisely *which* half of the data set was used, so one can imagine a likely scenario where this failed.

Looking at Figure 1, you can see that the decision boundary occurs in one of the histogram bins with zero observations. In this sense, it does correlate with structure in the data.

In Figure 8, we have plotted that same decision boundary over the 2D mass distribution. To do this, we had to derive an expression for $m_2(m_1, \mathcal{M}_c)$, which we accomplished using `Mathematica`. The expression is

$$m_2(m_1, \mathcal{M}_c) = \frac{(2/3)^{1/3} \mathcal{M}_c^5}{X} + \frac{X}{2^{1/3} \cdot 3^{2/3} \cdot m_1^3}, \quad (7)$$

where

$$X \equiv \left[9 m_1^7 \mathcal{M}_c^5 + \sqrt{3 m_1^9 \mathcal{M}_c^{10} \left(27 m_1^5 - 4 \mathcal{M}_c^3\right)} \right]^{1/3}. \quad (8)$$

As you can see from Figure 8, this line corresponds to the division between the events with and without counterparts.

A similar result is obtained using the linear SVM on the 2D mass distribution directly. The result is shown in Figure 9.

3 Conclusions

By observing compact binary mergers with gravitational wave detectors, and following up with EM counterparts, we can begin to understand the true mass distribution of neutron stars and black holes in the Universe. In this paper, we have outlined some rudimentary methods that can be used, once a large number of detections have been made. The mass distribution estimation described in §1 can be used to understand the true distribution, and the classifier described in §2 can be used to reduce resource usage, by restricting EM followup to events known to have them.

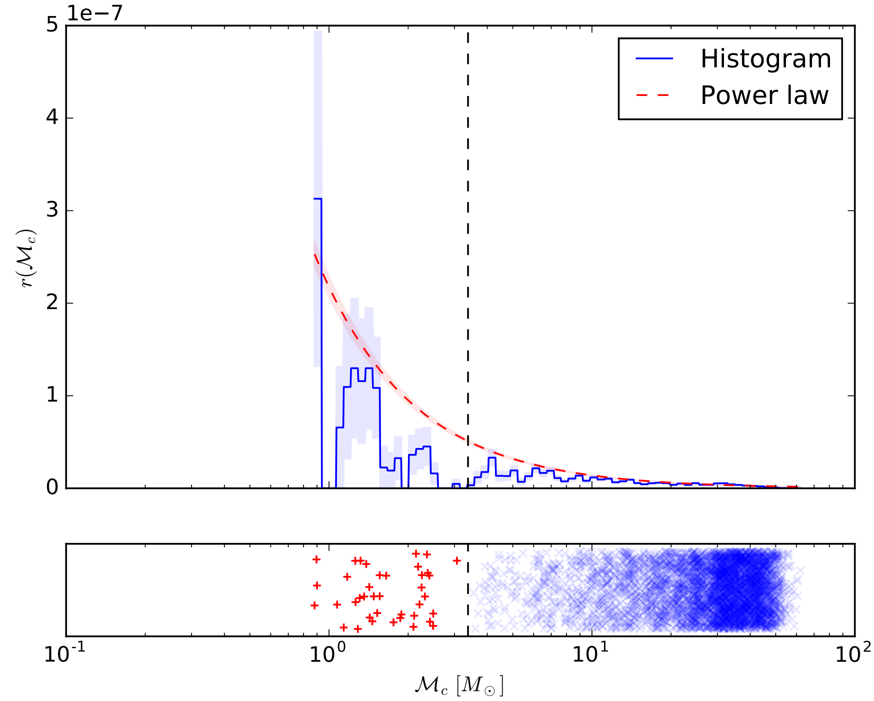
By including uncertainties in our rate estimates, these methods can be applied to a much smaller number of detections, such as the two currently reported. Of course, to have more confidence in the results, we would need a more sophisticated approach, with a more cautious treatment of uncertainty.

References

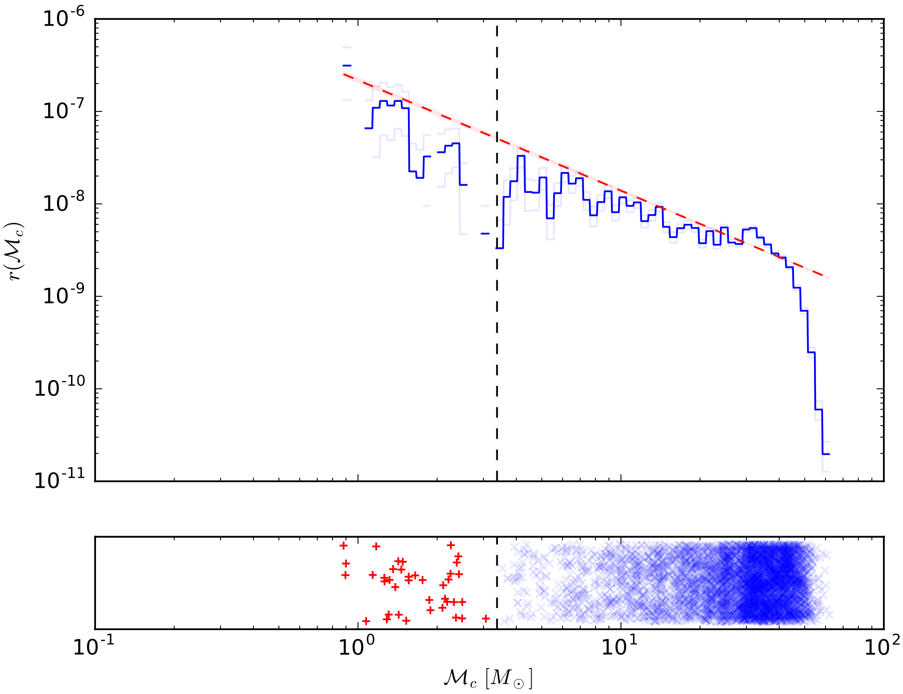
[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al., "Observation of Gravitational Waves from a Binary Black Hole Merger", Physical Review Letters **116**, 061102, 061102 (2016).

[2] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC Hammer", PASP **125**, 306–312 (2013).

[3] J. Wiss, *Errors on weighted averages, and weighted histograms*, (1995) `www.hep.uiuc.edu/e687/memos/weight_err.PS`.



(a)



(b)

Figure 1: Estimated rate of compact binary mergers, based on 5000 synthetic observations. Rate is shown in (a) linear and (b) log scale. Blue line is weighted histogram fit. Red curve is power law fit. Shaded regions are 1-σ error bars. Vertical dashed line is boundary between events with counterparts and without.

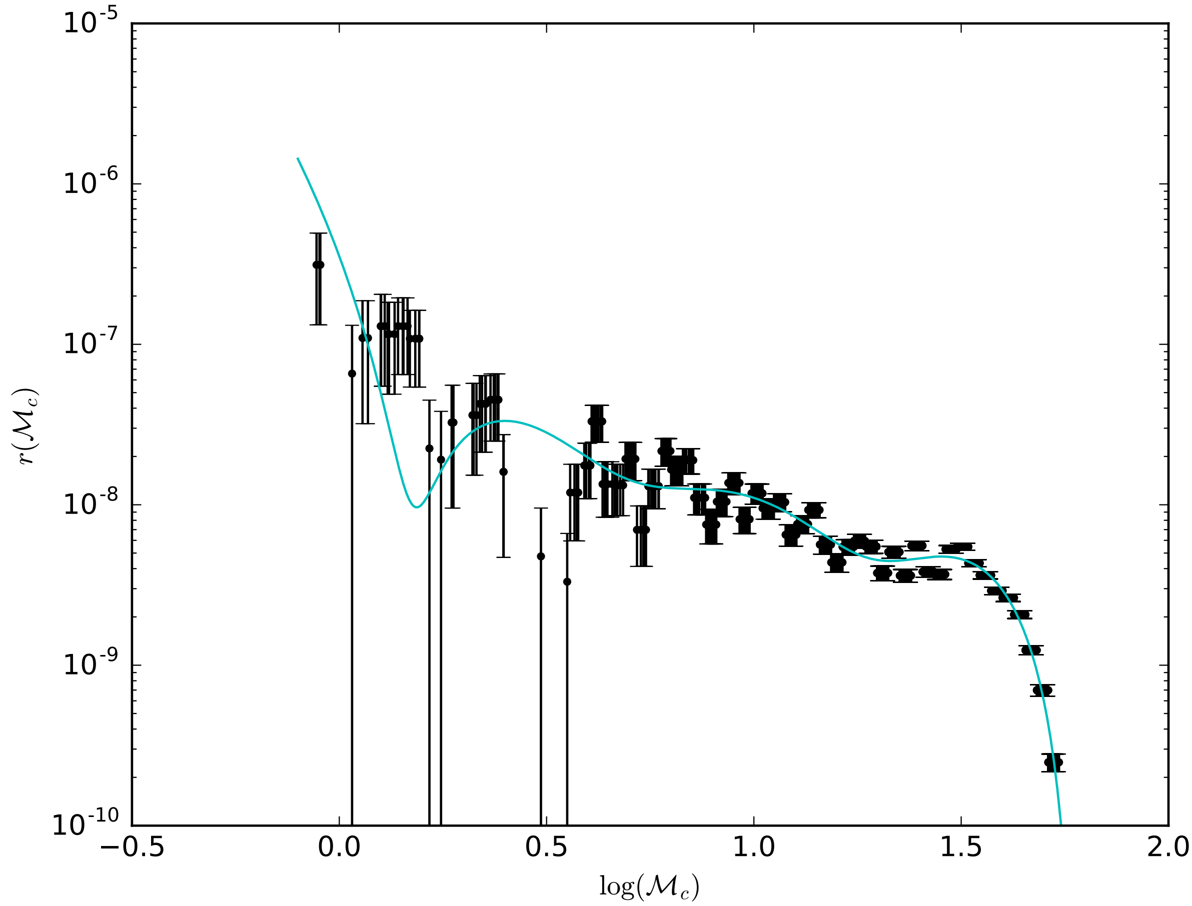


Figure 2: The cyan line is an example of a fit from MCMC.

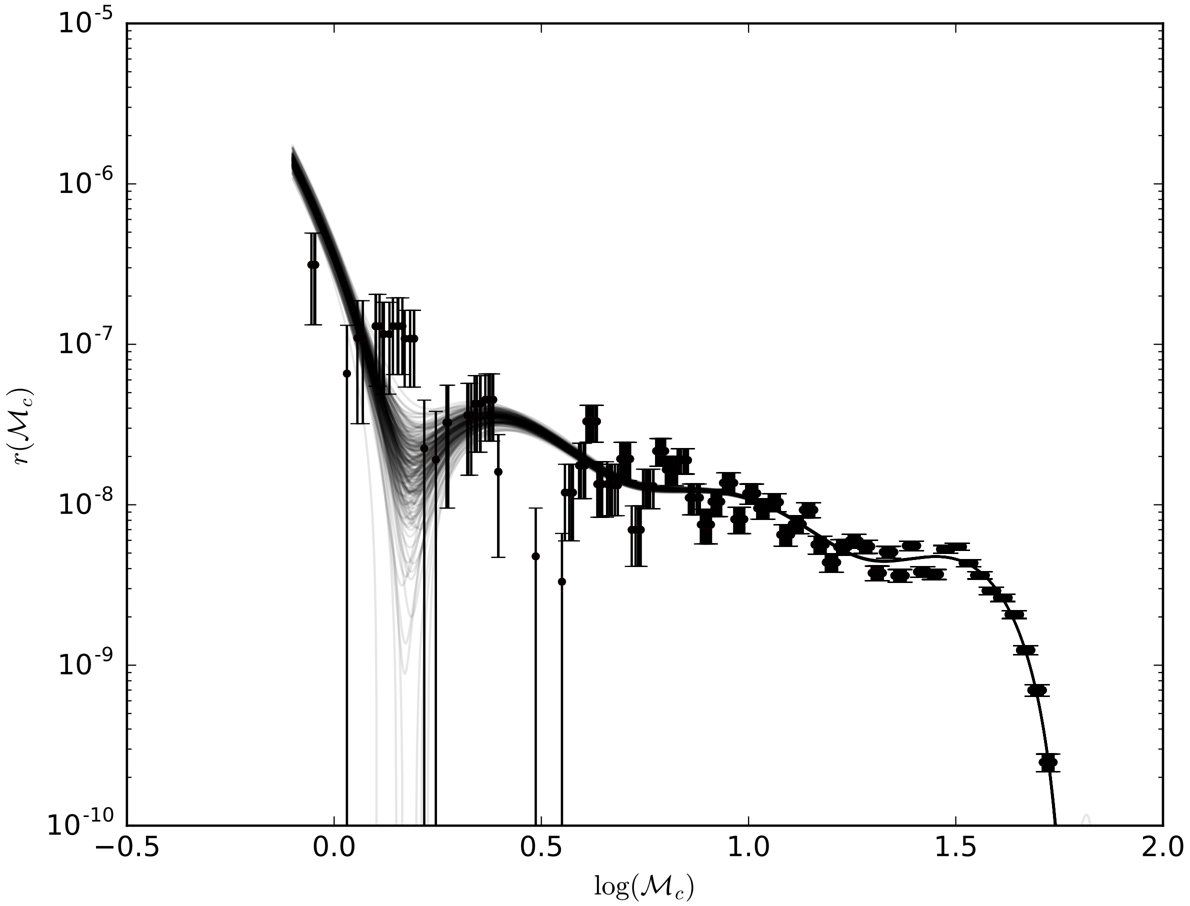


Figure 3: 2000 randomly chosen fits from the MCMC.

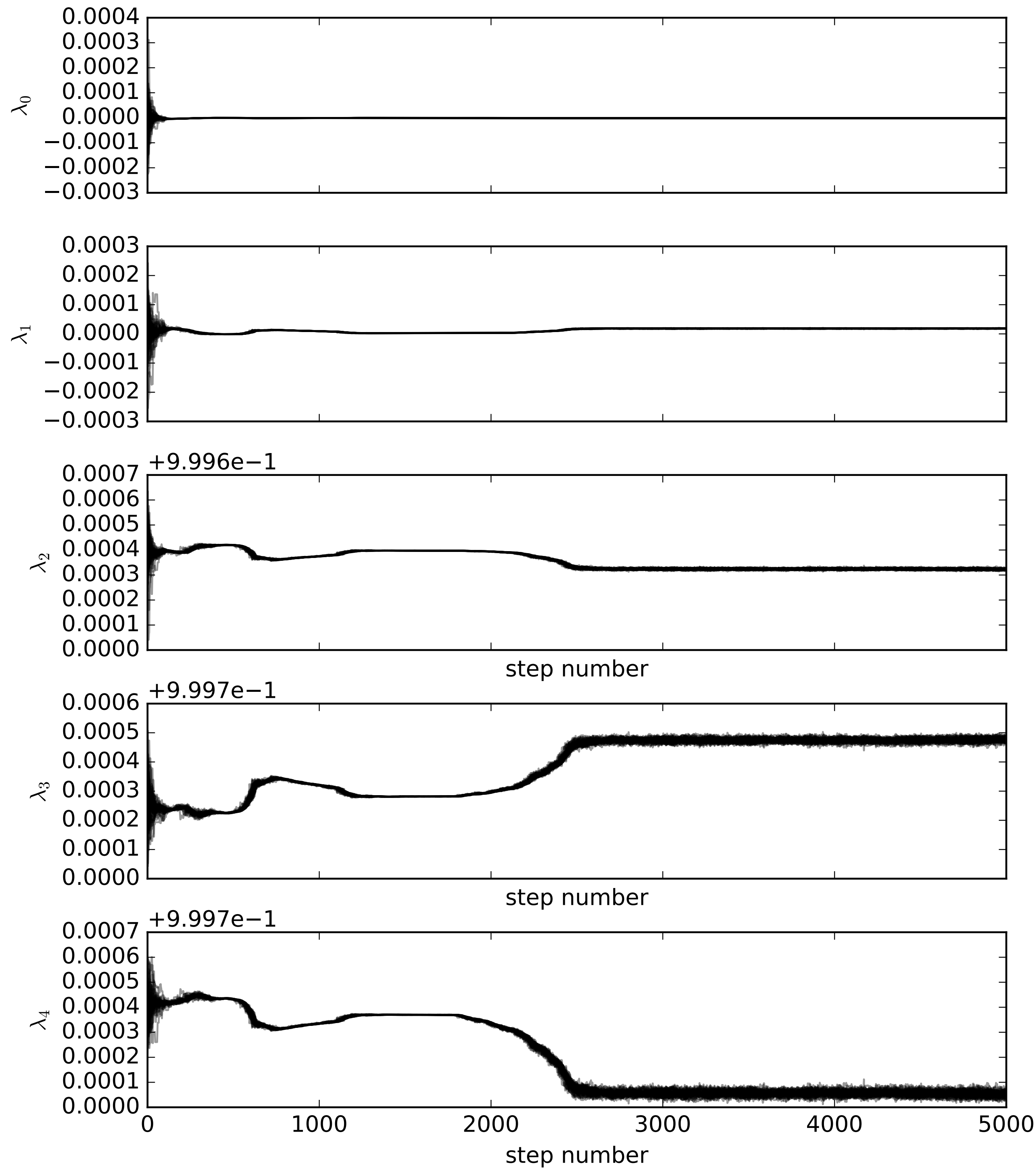


Figure 4: Visualization of MCMC worker locations over the large number of iterations performed for coefficients λ_0 to λ_4.

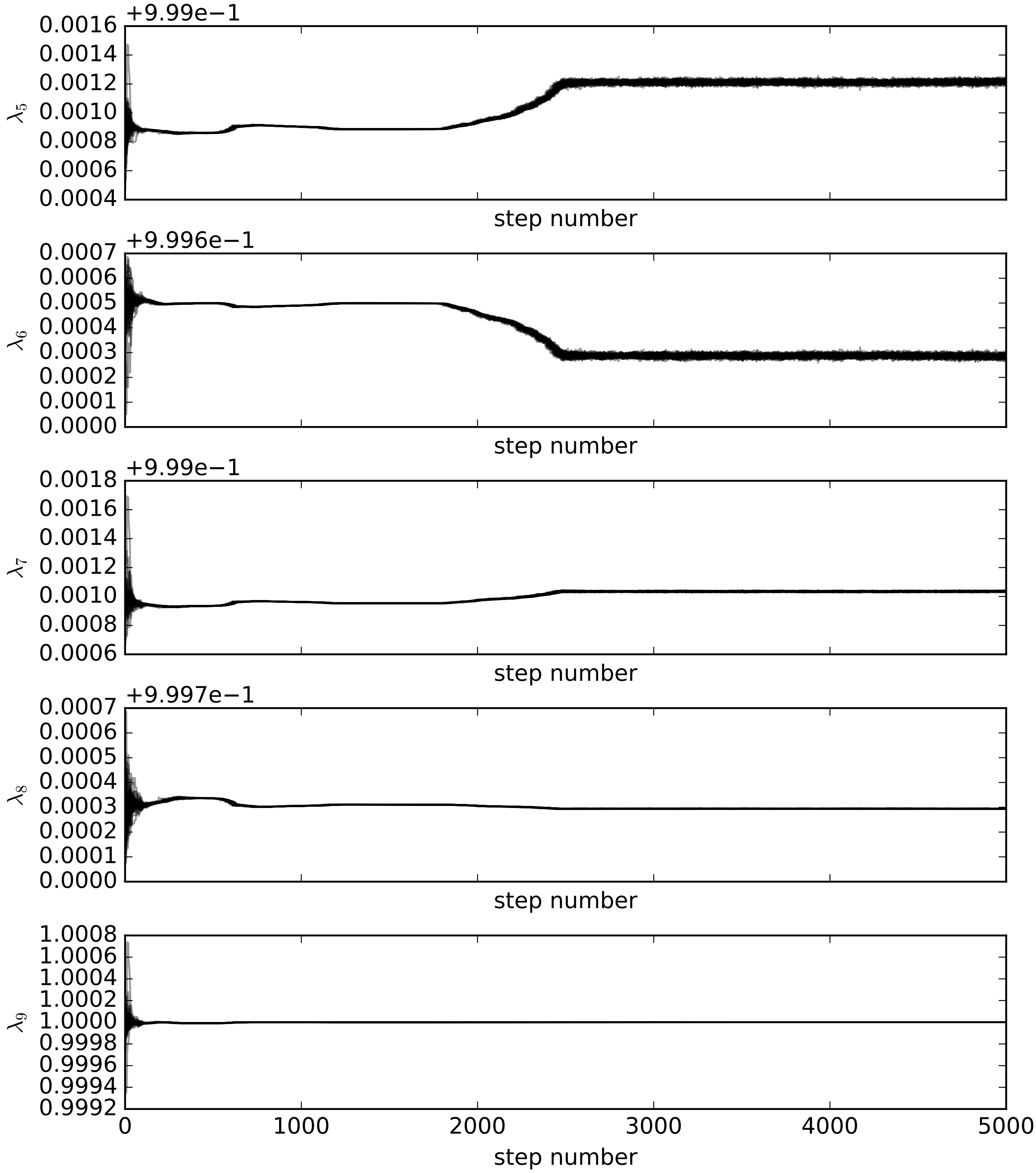


Figure 5: Visualization of MCMC worker locations over the large number of iterations performed for coefficients λ_5 to λ_9.

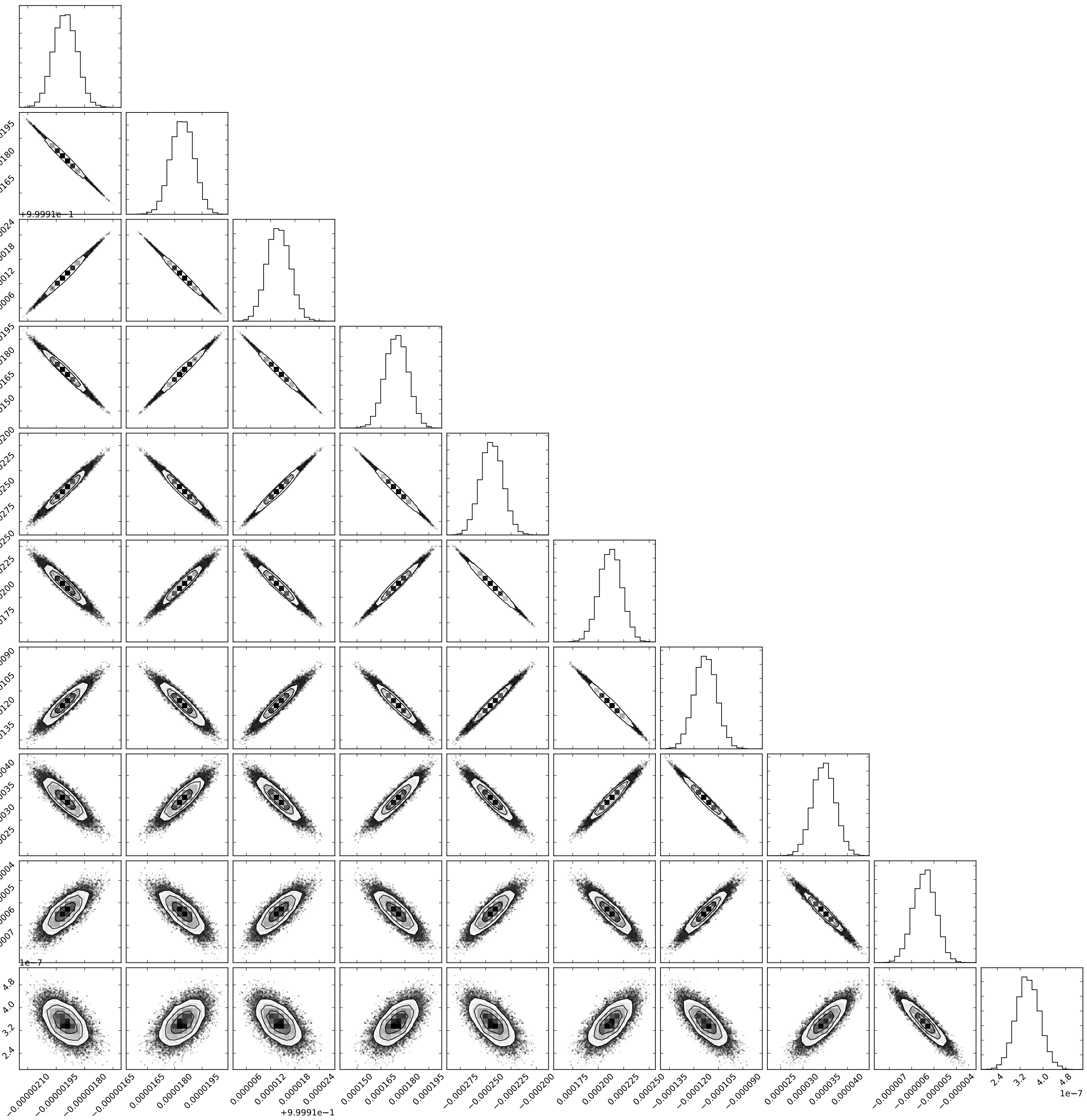


Figure 6: Corner plot of MCMC fit coefficients.

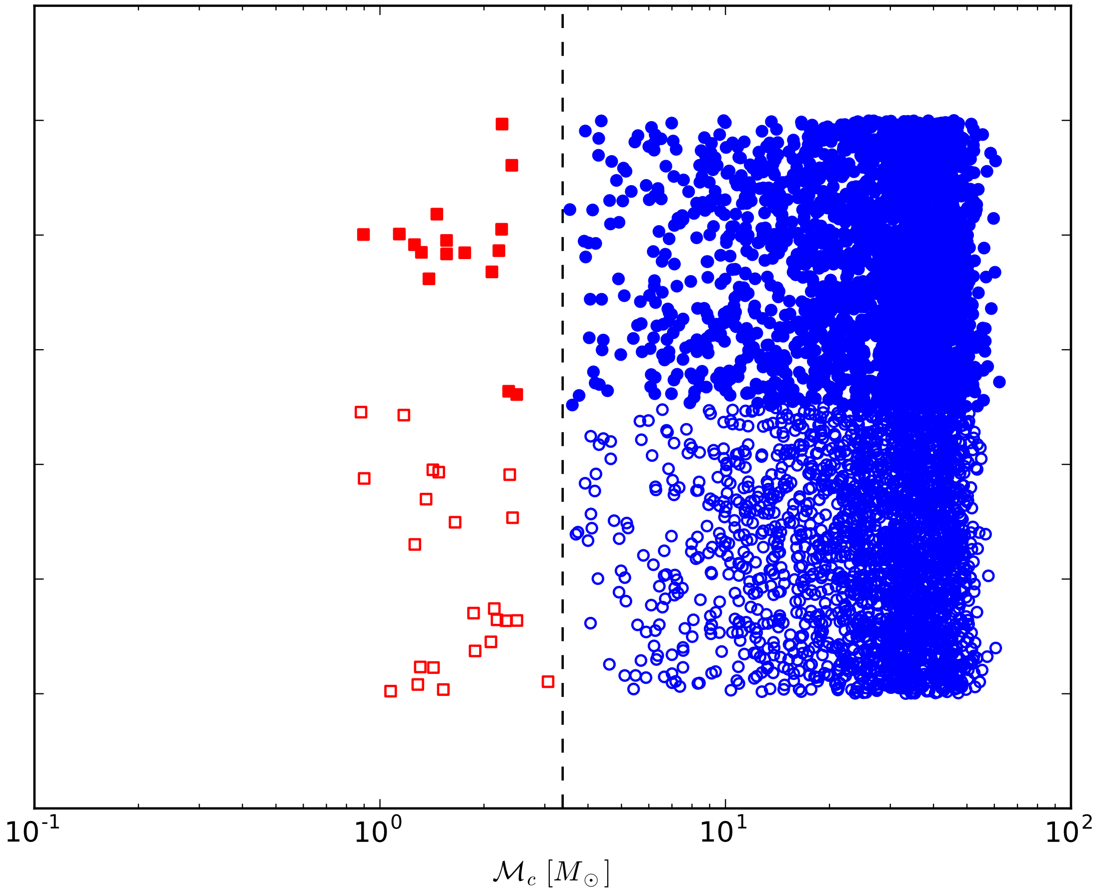


Figure 7: Visualization of 1D classifier performance. Red points have observed EM counterparts, while blue do not. We trained the classifier only on the data with solid points, and obtained a decision boundary shown by the dashed line. The open points are the remaining half of the data, which, as you can see, were correctly classified by this decision boundary.

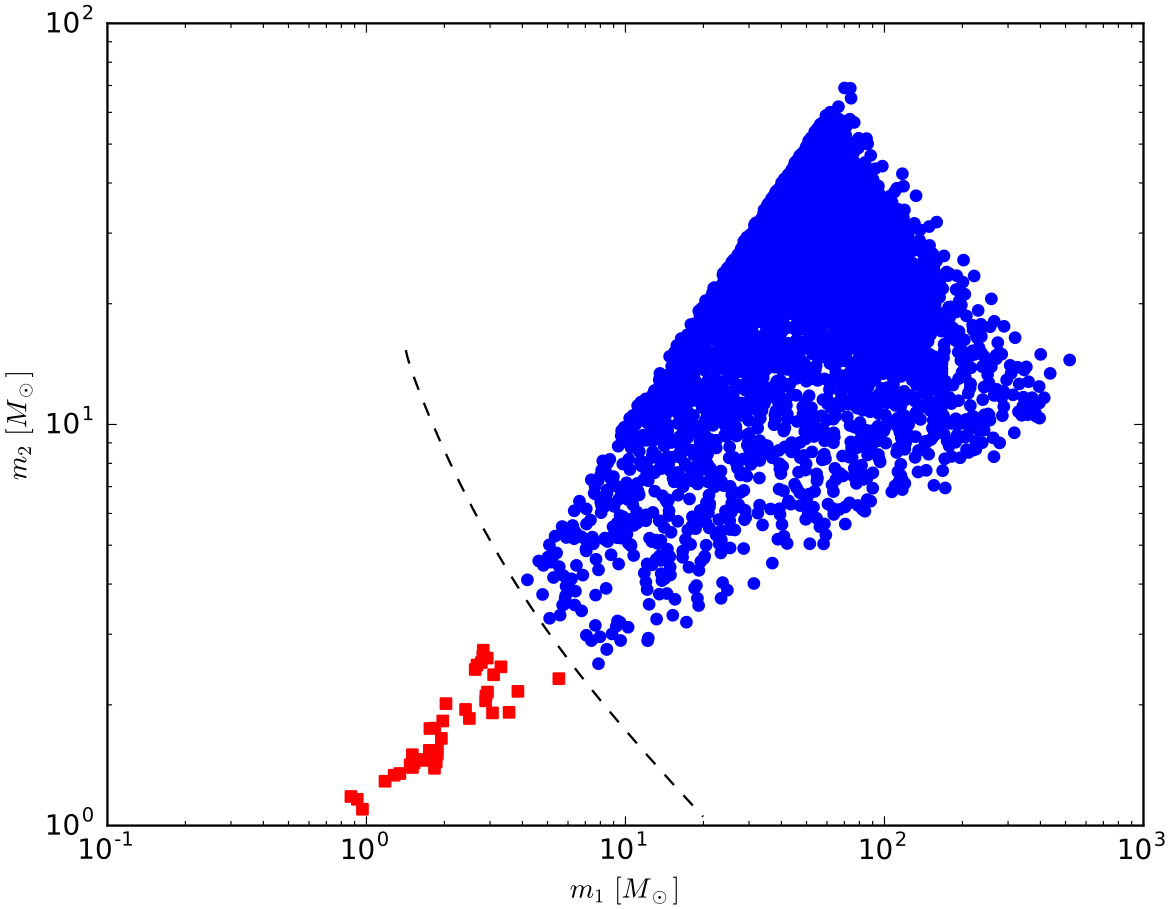


Figure 8: Decision boundary from 1D classifier trained on $\mathcal{M}_c$, overlayed on the 2D mass distribution. Red points have EM counterparts, while blue do not.

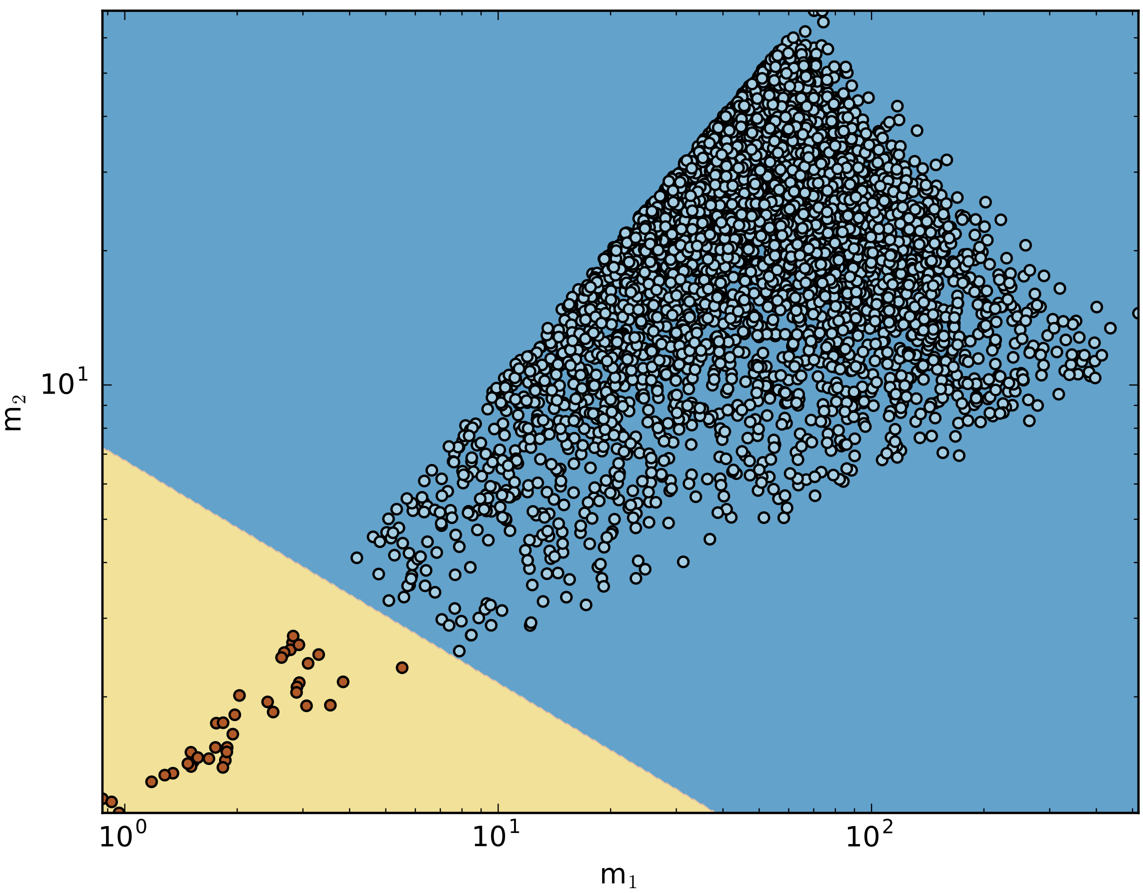


Figure 9: Results of linear SVM classifier, trained on half of the data set, displayed here to perfectly classify the entire data set. Red points represent events with EM counterparts, and blue points represent the events without. The predictions of the classifier are shown as the yellow and blue shaded regions.