

Eliminating Systematic Bias in Bright-Siren Multimessenger Observations of Compact Object Mergers from Inclination-Dependent Emission

WILL M. FARR^{1,2} AND FRIENDS

¹*Department of Physics and Astronomy, Stony Brook University, Stony Brook NY 11794, USA*

²*Center for Computational Astrophysics, Flatiron Institute, New York NY 10010, USA*

ABSTRACT

This article and associated code explores systematics associated with inclination dependence in detection of bright counterparts to GW mergers. The TL;DR is that, by fitting out the inclination dependence of EM detections with fairly simple models, we can remove what otherwise would be a bias that could prevent percent-level expansion history measurements with bright sirens.



1. INTRODUCTION

2. OBSERVATIONAL MODEL

We use a simplified observational model that reproduces the essential aspects of gravitational wave distance measurements without requiring generating and inferring full gravitational waveforms. We imagine that each gravitational wave event has a (cosine) inclination $x = \cos \iota$ and a distance, and these combine to give an amplitude in the left- and right-handed polarization components of

$$A_{R/L} = \frac{(1 \pm x)^2}{d} \quad (1)$$

in some units. We assume that a detector network measures a linear projection of these these polarization amplitudes with unit-amplitude Gaussian noise:

$$A_{R/L,\text{obs}} = f_{R/L} A_{R/L} + N(0, 1). \quad (2)$$

For a detector network with two or more arbitrarily oriented, nearly equally-sensitive detectors, we can set $f_R = f_L = 1$; for two nearly-aligned equally-sensitive detectors (e.g. the two LIGO detectors in Washington and Louisiana), we can set $f_R \ll f_L = 1$ to reflect that one polarization component is measured much more precisely than

the other. From here onward we assume events used for bright sirens have small localization volumes and are therefore observed in three or more detectors and we set $f_L = f_R = 1$.

The S/N of an observation in our model is given by

$$\rho^2 = \frac{A_{R,\text{obs}}^2 + A_{L,\text{obs}}^2}{2}. \quad (3)$$

We impose a Euclidean prior on d , $p(d) \propto d^2$, appropriate for the nearby universe, and a flat prior on x , $p(x) = 1/2$, appropriate for an intrinsically isotropic population of GW mergers. Inferring d and x from $A_{R,\text{obs}}$ and $A_{L,\text{obs}}$ in our model results in posterior densities that reproduce the distance-inclination correlation observed in analysis of actual GW events, as well as the (fractional) distance uncertainty and inclination uncertainty observed in actual GW events at comparable S/N. Figure 1 shows the inferred distance and inclination from our model for a GW170817-like event ($\rho \simeq 30$, $x \simeq -1$); our model's inferences compare directly to those for GW170817 itself (Abbott et al. 2017).

2.1. Effect of EM Selection

We assume that each of our observations also has a detection of an electromagnetic counterpart that gives a perfect measurement of redshift, thus constraining H_0 . The full likelihood for both GW and EM data for an observation factorizes because the measurement noise for the observations are independent:

$$p(d_{\text{GW}}, d_{\text{EM}} | x, d, H_0) = p(d_{\text{GW}} | x, d) p(d_{\text{EM}} | x, d, H_0). \quad (4)$$

The first term is the likelihood representing the measurement model described above; the second term relates the EM measurements to the inclination, distance, and (through the distance-redshift relation) H_0 . If we were willing to model the dependence of the EM data on x , we could use it to further constrain the inclination measurement and thereby improve the H_0 measurement. But here we suppose that this is not possible—either we feel that the inclination modeling for the EM data is systematics-dominated, or we are unable to even produce a model of the inclination-dependent emission being measured. In this case we can marginalize the EM data out of the likelihood, retaining only (1) the redshift measurement and (2) the knowledge that the EM emission was detectable. Let Ω be the set of EM data consistent with these criteria. Then

$$\int_{d_{\text{EM}} \in \Omega} dd_{\text{EM}} p(d_{\text{EM}} | x, d, H_0) \propto \delta(z - H_0 d) P_{\text{det,EM}}(x, d, H_0), \quad (5)$$

where $P_{\text{det,EM}}(x, d, H_0)$ is the fraction of mergers with inclination x , at distance d , and with Hubble constant H_0 that produce detectable EM emission (Mandel et al. 2019).

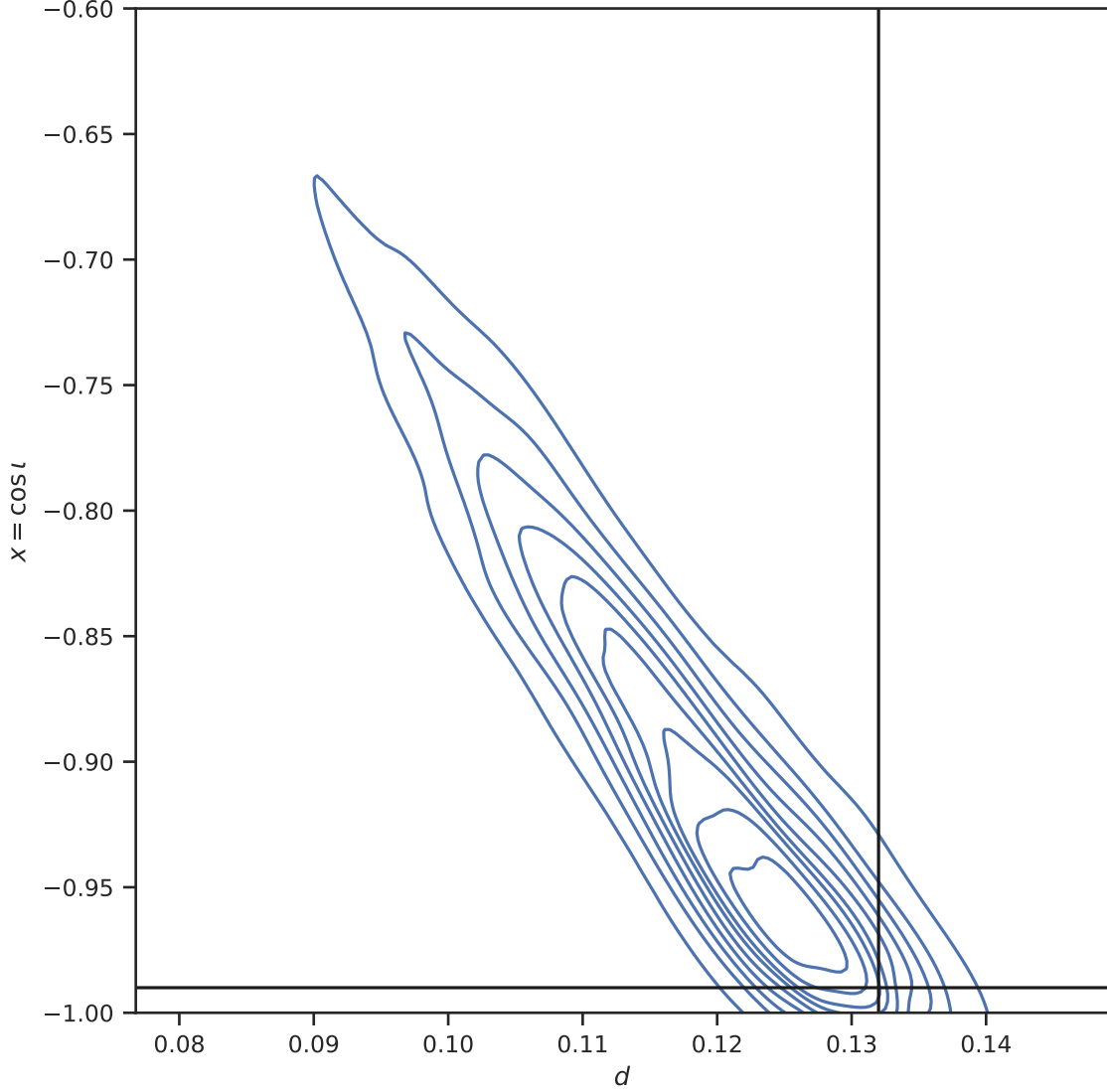


Figure 1. Posterior on distance d and $x = \cos \iota$ for a GW170817-like event ($\rho \simeq 30$) from our observational model. Contour levels correspond to successive 10% credible intervals; true parameters are indicated by the black lines. Compare to Figure 2 of [Abbott et al. \(2017\)](#). The fractional uncertainty on d after marginalizing over x with an isotropic (flat) prior is $\sigma_d / \langle d \rangle \simeq 0.14$; for GW170817 the same quantity is also $\sigma_d / \langle d \rangle \simeq \sigma_{H_0} / \langle H_0 \rangle \simeq 0.14$ ([Abbott et al. 2017](#)). Similarly, our model has $x_{84\%} \simeq -0.82$, while for GW170817 the equivalent 1- σ upper limit on $x \simeq -0.81$ ([Abbott et al. 2017](#)). Our model thus does a reasonable job reproducing the correlation between distance and inclination and the relevant uncertainties at the S/N of GW170817; in both our model and a full analysis of a GW event uncertainties will scale inversely with S/N.



If we had a good model for the EM emission, we could *calculate* $P_{\text{det,EM}}$; but if we do not, we can treat it as a parameterized function, and learn its shape from the GW data. (Intuitively: the GW data carry some information about the inclination of each event; considering the set of events with coincident EM detections lets us learn the population of inclinations for *just these* events, which is proportional to $P_{\text{det,EM}}$.)

The term $P_{\text{det,EM}}$ acts like a modification of the prior applied to x , d , and H_0 in the overall analysis. After integrating out the EM data in Ω , we obtain a modified GW likelihood:

$$p(d_{\text{GW}}, z \mid x, d, H_0) = \int_{d_{\text{EM}} \in \Omega} dd_{\text{EMP}}(d_{\text{GW}} \mid x, d) p(d_{\text{EM}} \mid x, d, H_0) \propto \delta(z - H_0 d) p(d_{\text{GW}} \mid x, d) P_{\text{det,EM}}(x, d, H_0). \quad (6)$$

Applying a d^2 prior and integrating across the δ function, we obtain

$$p(d_{\text{GW}}, z \mid x, H_0) \propto \int dd p(d_{\text{GW}}, z \mid x, d, H_0) d^2 = \frac{z^2}{H_0^3} P_{\text{det,EM}}(x, d = z/H_0, H_0) p(d_{\text{GW}} \mid x, d = z/H_0). \quad (7)$$

Here we see that $P_{\text{det,EM}}$ acts as an additional multiplicative factor, appearing as a modification to the isotropic prior that we would naturally apply for the inclination of our gravitational wave population.

TODO: Flesh this out, and discuss GW selection (mechanical, and our model, with $\rho > 10$); and then get specific about our model for $P_{\text{det,EM}}$. Also discuss “joint” catalogs, like combining GRB-selected and kilonova-selected data sets.

2.2. Functional Form of $P_{\text{det,EM}}$

Here for simplicity we assume

- $P_{\text{det,EM}}$ does not depend on d wherever GW events are detectable. That is: the distance-dependence of selection effects is dominated by the GW selection-function.
- $P_{\text{det,EM}}$ is independent of H_0 . This will be the case whenever the choice of EM survey strategy / catalog selection does not depend on the redshift of the EM sources. For example, a flux-limited survey of GW localization volumes would have a detection efficiency that is independent of H_0 ; a survey that focuses on objects with spectroscopically-measured redshifts consistent with the GW d would have a selection function that is strongly-dependent on H_0 .

With these assumptions, $P_{\text{det,EM}}$ varies only with x . We propose to use an expansion of $P_{\text{det,EM}}$ in Legendre polynomials to model and learn the EM selection function. Since $P_{\text{det,EM}}$ must always be positive, we write

$$\log P_{\text{det,EM}}(x, d, H_0) = \sum_{l=1}^{N_l} A_l P_l(x) \quad (8)$$

(The normalization of $P_{\text{det,EM}}$ does not affect the likelihood; here we have fixed the normalization by omitting the constant $P_0(x) = 1$ term from the sum.) The $A =$

$\{A_l \mid l = 1, \dots, N_l\}$ are additional parameters of our population model that we will infer.

Some anticipated inclination-dependent selection effects fit naturally into this form (e.g. kilonovae are expected to have significant emission into all viewing angles [TODO: citations](#)); others do not (e.g. GRB detections require a small viewing angle [TODO: citations](#)). We will show below that even for the latter case, where $P_{\text{det,EM}}$ varies strongly—even discontinuously—with x , the smooth Legendre polynomial interpolation is sufficient to reduce systematics below the percent-level in a realistic measurement.

3. EXAMPLES

3.1. *Fitting Weakly Variable EM Selection*

Let

$$\log P_{\text{det,EM}}(x) = P_2(x) = \frac{1}{2} (3x^2 - 1), \quad (9)$$

and $N_l = 2$. We generate a mock catalog of $N = 128$ GW+EM detections with S/N $\rho > 10$, and $H_0 = 0.7$, and calculate the joint posterior on H_0 , $A = \{A_1, A_2\}$, and each event's x . Figure 2 shows that we recover the intrinsic inclination distribution of these events well.

Ultimately, fitting the EM selection effects with this dataset results in recovery of $H_0 = 0.7027^{+0.0057}_{-0.0055}$ (median and 68% credible interval, i.e. sub-percent uncertainty). Failing to account for the inclination-dependent EM selection effect introduces significant bias in the H_0 estimate, recovering $H_0 = 0.7131^{+0.0058}_{-0.0056}$ ([Chen 2020](#)). Figure 3 illustrates these effects.

3.2. *Fitting a Strongly Variable EM Selection*

Different EM counterparts will exhibit different dependence on inclination angle. In this subsection, we model an GW+EM catalog composed of two components. One component has EM detectable emission only very close to on/off axis, with $P_{\text{det,EM}} = 0$ unless $|x| \geq x_{\text{min}}$ with $x_{\text{min}} = \cos 15^\circ$. The other component has a (weak) preference for on/off axis emission, with detectability varying quadratically from a maximum at $|x| = 1$ to and 1/3 of its maximum value at $x = 0$. The two components are mixed equally in the population of binary mergers, so that the overall EM detection efficiency is

$$P_{\text{det,EM}}(x) \propto \frac{1}{2} \left(\frac{H(|x| - x_{\text{min}})}{x_{\text{min}}} + \frac{3(x^2 + 1/2)}{5} \right), \quad (10)$$

with H the Heaviside step function. We choose $N_l = 4$. The logarithm of this function is not expressible as any finite sum of Legendre polynomials, so our model is incapable of exactly reproducing the shape of this detection efficiency for any choice of N_l . Nonetheless, we will see that we can still obtain sub-percent accurate, unbiased estimates of H_0 , even with only four additional parameters to describe the EM selection function.

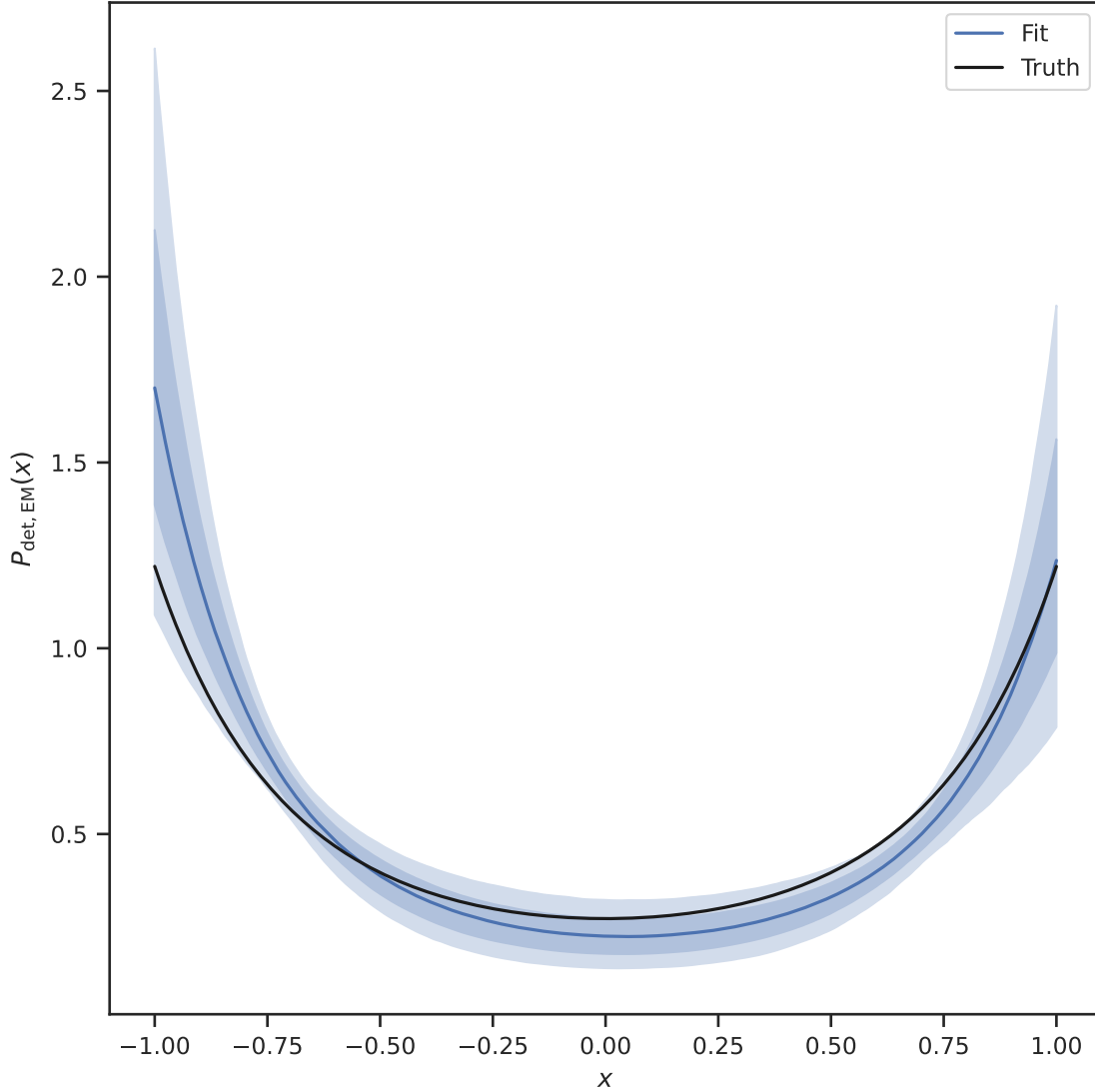


Figure 2. The posterior on the EM detection efficiency inferred from 128 mock GW+EM detections with the EM detection efficiency described in § 3.1. The blue line gives the posterior median; bands give the 68% and 95% credible intervals. The black line is the true detection efficiency used to generate the mock data set (Eq. (9)).



Once again, we generate a catalog of $N = 128$ GW+EM mock detections with S/N $\rho > 10$ using $H_0 = 0.7$. We recover a $P_{\text{det,EM}}$ shape that is indicative of the EM detection efficiency, but the true efficiency lies well outside the posterior 95% credible interval over much of the range $-1 < x < 1$. Figure 4 shows the posterior on $P_{\text{det,EM}}$.

Even though the posterior on the detection efficiency does favors curves very different to the true detection efficiency, the posterior on H_0 is not biased. The imperfect measurements of inclination from the GW data “soften” the bias in the recovery of the detection efficiency enough to recover a reasonable posterior on H_0 . This is shown in Figure 5, which also shows that failing to account for the inclination-dependent selection effects here leads to very significant bias in the recovered H_0 value.

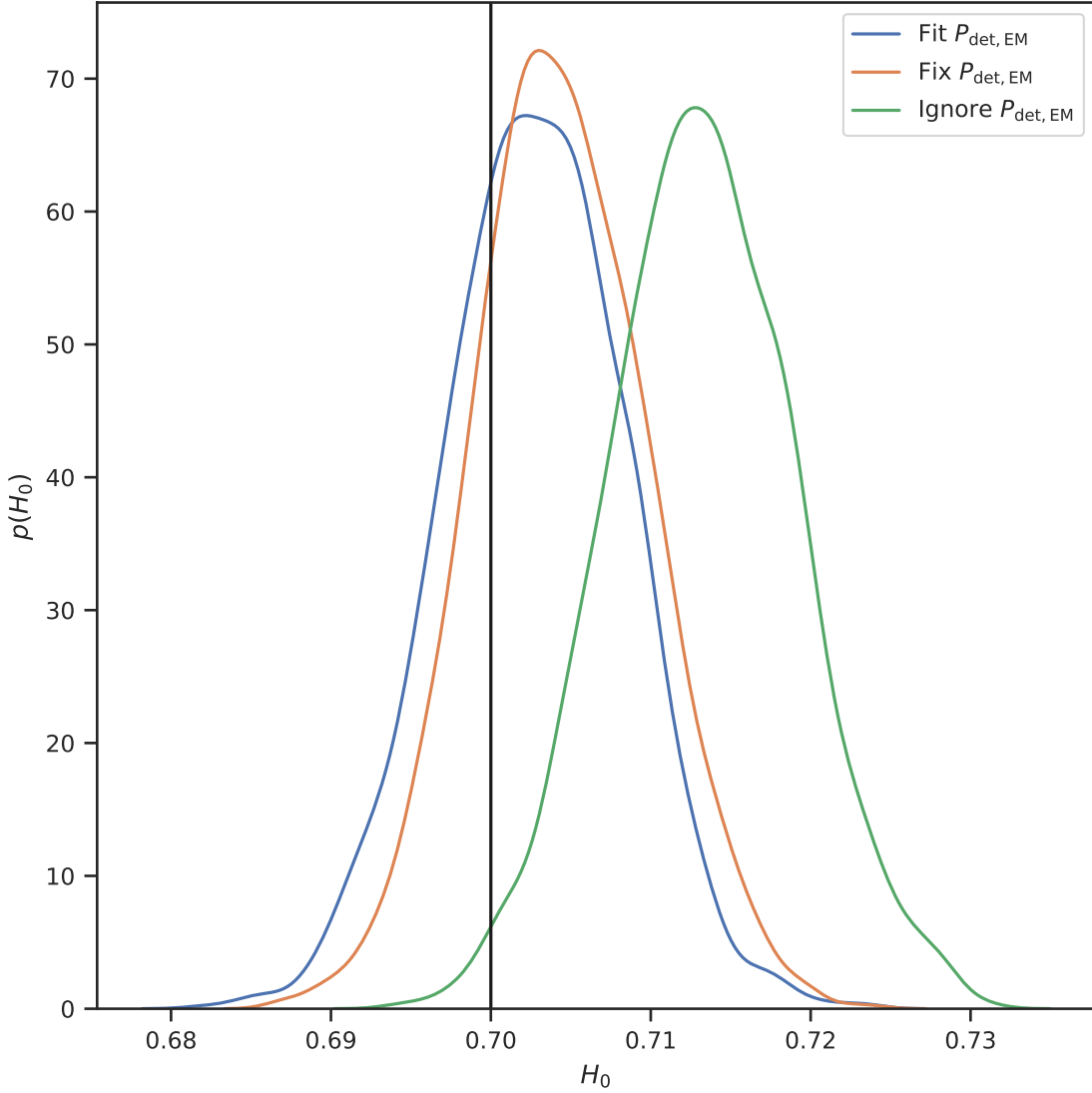


Figure 3. Posterior on H_0 for the mock catalog discussed in § 3.1. The blue curve is the posterior when fitting $P_{\text{det,EM}}$ with $N_l = 2$. We find $H_0 = 0.7027^{+0.0057}_{-0.0055}$ (median and 68% credible interval). The orange curve is the posterior when $P_{\text{det,EM}}$ is fixed to its true value (i.e. we assume we know the EM selection effect perfectly). We find $H_0 = 0.7042^{+0.0057}_{-0.0050}$, a 5 percent reduction in uncertainty. The green curve is the posterior ignoring inclination-dependent EM selection effects, which results in $H_0 = 0.7131^{+0.0058}_{-0.0056}$, a $\sim 2\sigma$ bias. The black vertical line is the true H_0 used to generate the catalog.



With this choice of EM detection efficiency, we observe a counterpart in about 15% of the GW-detectable systems. It is interesting to ask whether the improvement in inclination estimation from the structured EM emission (and therefore improved distance estimation from the GW observations) outweighs the loss of sources compared to a scenario where the EM emission is uniform. (In other words, would it be better to employ a search strategy that selects sources that are likely to have anisotropic EM emission in order to improve distance estimates?) For this distribution, a cat-

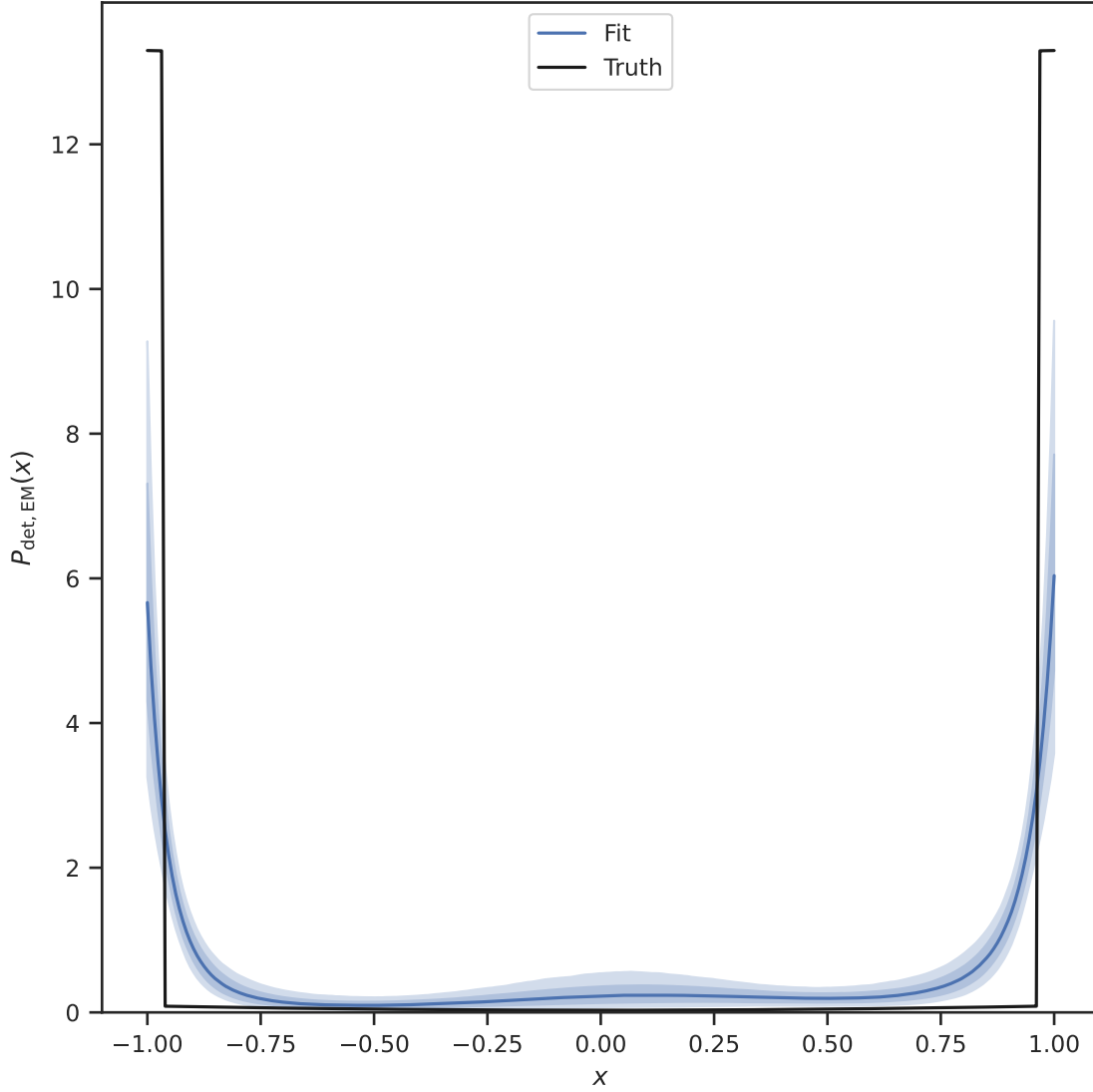



Figure 4. Posterior on $P_{\text{det,EM}}$ from fitting the mock catalog described in § 3.2. The catalog is an equal mixture of an on/off axis component ($|x| > \cos 15^\circ$) and a quadratically-varying component; the true $P_{\text{det,EM}}$ is shown by the black curve. The blue line shows the posterior median, and the dark and light bands the posterior 68% and 95% credible intervals. For most of the interval $-1 < x < 1$ the true detection efficiency is well outside the posterior 95% credible interval due to the mismatch between the Legendre polynomial fitting function, Eq. (8) with $N_l = 4$, and the sharply-varying detection efficiency. 

alog with $1/0.15 \sim 7$ times as many sources with isotropic EM emission provides overall a better Hubble parameter measurement, as shown in Figure 6. (For both mock catalogs, we employ $N_l = 4$, allowing for the same possibility of structure in the EM emission; the inferred $P_{\text{det,EM}}$ for the catalog with isotropic emission is shown in Figure 7.)

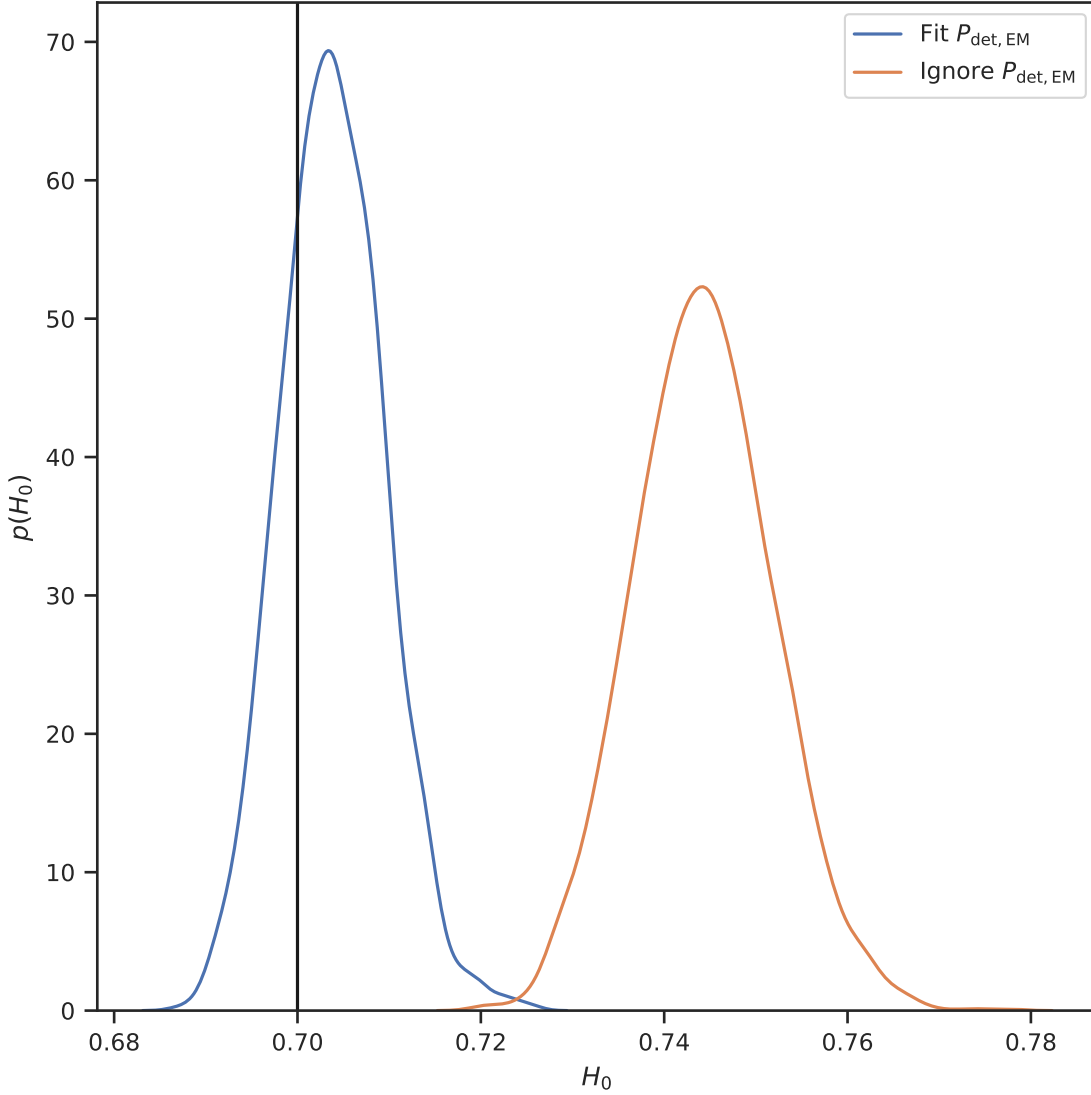


Figure 5. Posterior on H_0 from fitting the mock catalog described in § 3.2. The vertical black line shows $H_0 = 0.7$, the value used to prepare the catalog. The blue curve shows the posterior when fitting $P_{\text{det,EM}}$ using Eq. (8) with $N_l = 4$. In this case we find $H_0 = 0.7037^{+0.0055}_{-0.0055}$ (median and 68% credible interval). The orange curve shows the posterior when fitting without accounting for inclination-dependent EM selection. In this case we find $H_0 = 0.7442^{+0.0077}_{-0.0075}$, with $\sim 6\sigma$ bias.



2 **TODO: We thank people for stuff.** This work was begun at the KICP workshop
 3 “The Quest for Precision Gravitational Wave Cosmology” in September 2022; the
 4 authors thank the KICP and workshop organizers for an exceptionally stimulating
 5 environment.

Software: `arviz` (Kumar et al. 2019), `matplotlib` (Hunter 2007), `pymc` (Salvatier et al. 2016), `seaborn` (Waskom 2021), `showyourwork` (Luger et al. 2021)

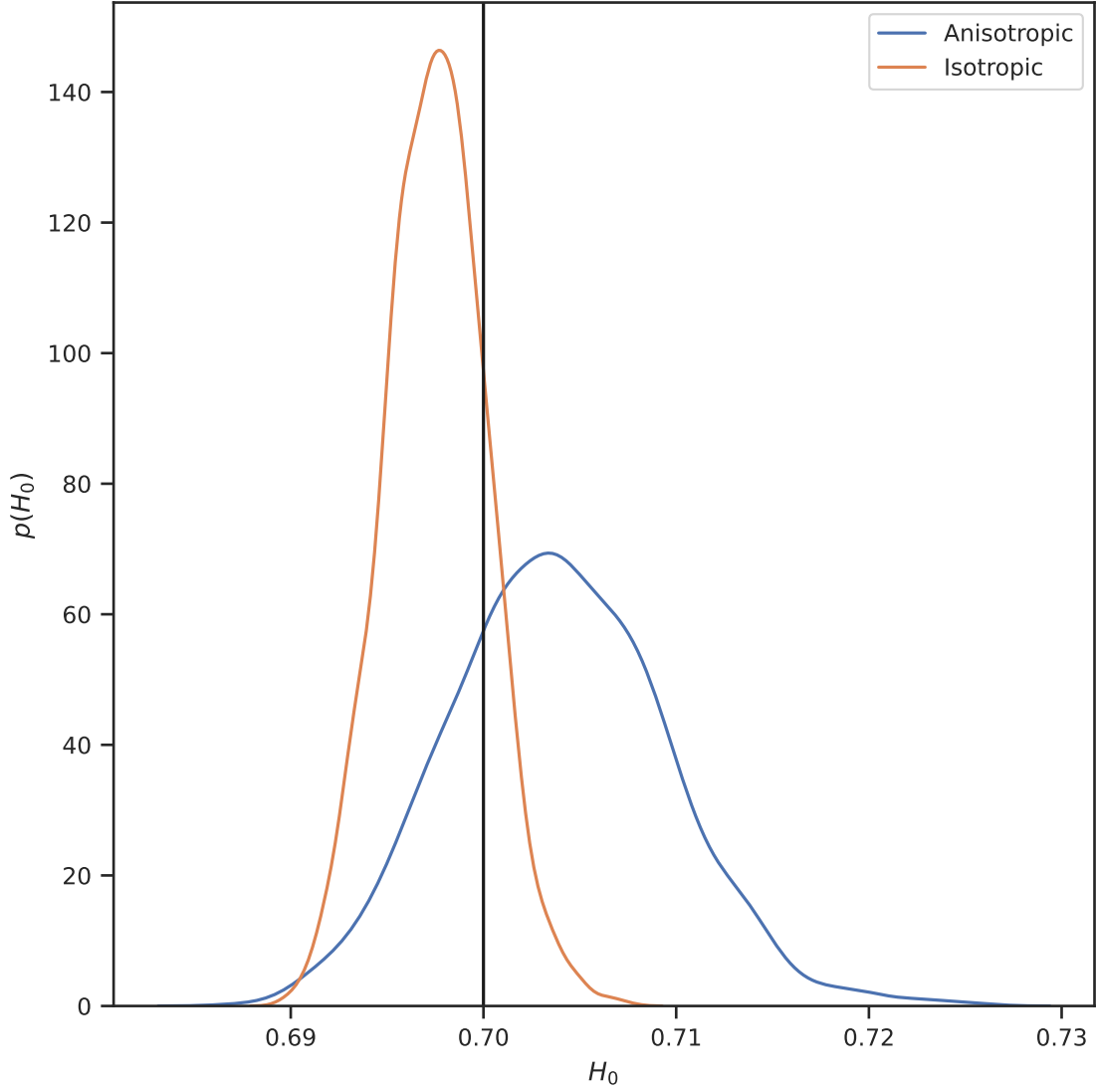


Figure 6. Comparison between H_0 inferred from a mock catalog with the anisotropic EM emission discussed in § 3.2 (blue curve) with a catalog with an equivalent GW merger rate but isotropic emission (and therefore ~ 7 times as many detections; orange curve). The improved catalog size from the isotropic emission overcomes the improved distance estimates from the knowledge of inclination gained from the anisotropic emission, and the overall measurement is tighter. We find the posterior median and 68% credible interval $H_0 = 0.7037^{+0.0055}_{-0.0055}$ from the catalog with anisotropic EM emission, and $H_0 = 0.6975^{+0.0026}_{-0.0026}$ from the larger catalog with isotropic emission.



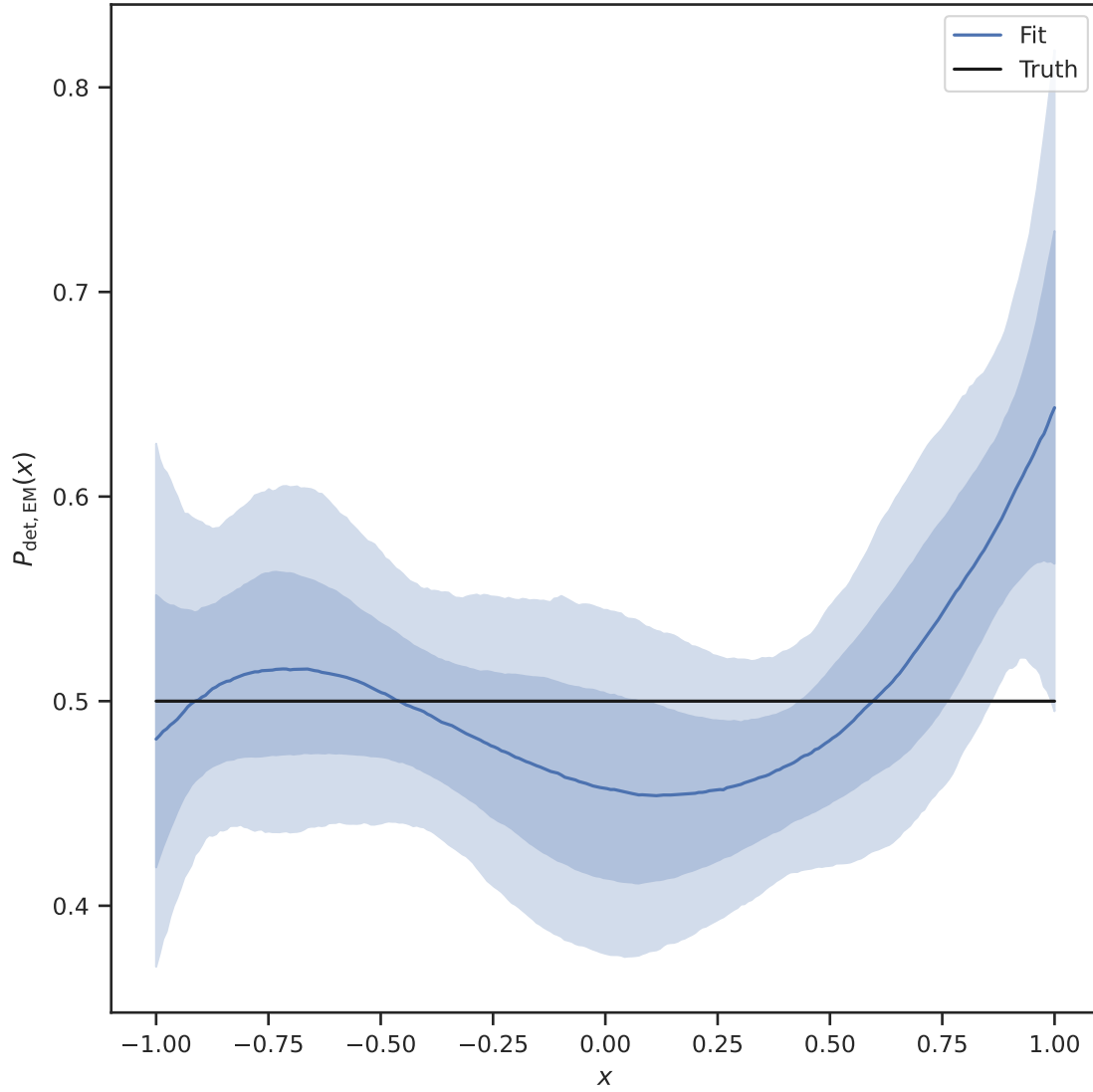


Figure 7. The inferred inclination distribution for the catalog with isotropic EM emission and equivalent GW merger to the anisotropic emission described in § 3.2. The blue line gives the posterior median and the dark and light bands the posterior 68% and 95% credible interval. Even though the EM emission is isotropic, we fit the detection efficiency with $N_l = 4$ for a fair comparison with the (smaller) catalog of anisotropic sources whose inferred inclination distribution is shown in Figure 4.



REFERENCES

- Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, *Nature*, 551, 85, doi: [10.1038/nature24471](https://doi.org/10.1038/nature24471)
- Chen, H.-Y. 2020, *PhRvL*, 125, 201301, doi: [10.1103/PhysRevLett.125.201301](https://doi.org/10.1103/PhysRevLett.125.201301)
- Hunter, J. D. 2007, *Computing in Science & Engineering*, 9, 90, doi: [10.1109/MCSE.2007.55](https://doi.org/10.1109/MCSE.2007.55)
- Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, *Journal of Open Source Software*, 4, 1143, doi: [10.21105/joss.01143](https://doi.org/10.21105/joss.01143)
- Luger, R., Bedell, M., Foreman-Mackey, D., et al. 2021, arXiv e-prints, arXiv:2110.06271. <https://arxiv.org/abs/2110.06271>
- Mandel, I., Farr, W. M., & Gair, J. R. 2019, *MNRAS*, 486, 1086, doi: [10.1093/mnras/stz896](https://doi.org/10.1093/mnras/stz896)
- Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. 2016, *PeerJ Computer Science*, 2, e55. <https://arxiv.org/abs/1507.08050>
- Waskom, M. L. 2021, *Journal of Open Source Software*, 6, 3021, doi: [10.21105/joss.03021](https://doi.org/10.21105/joss.03021)