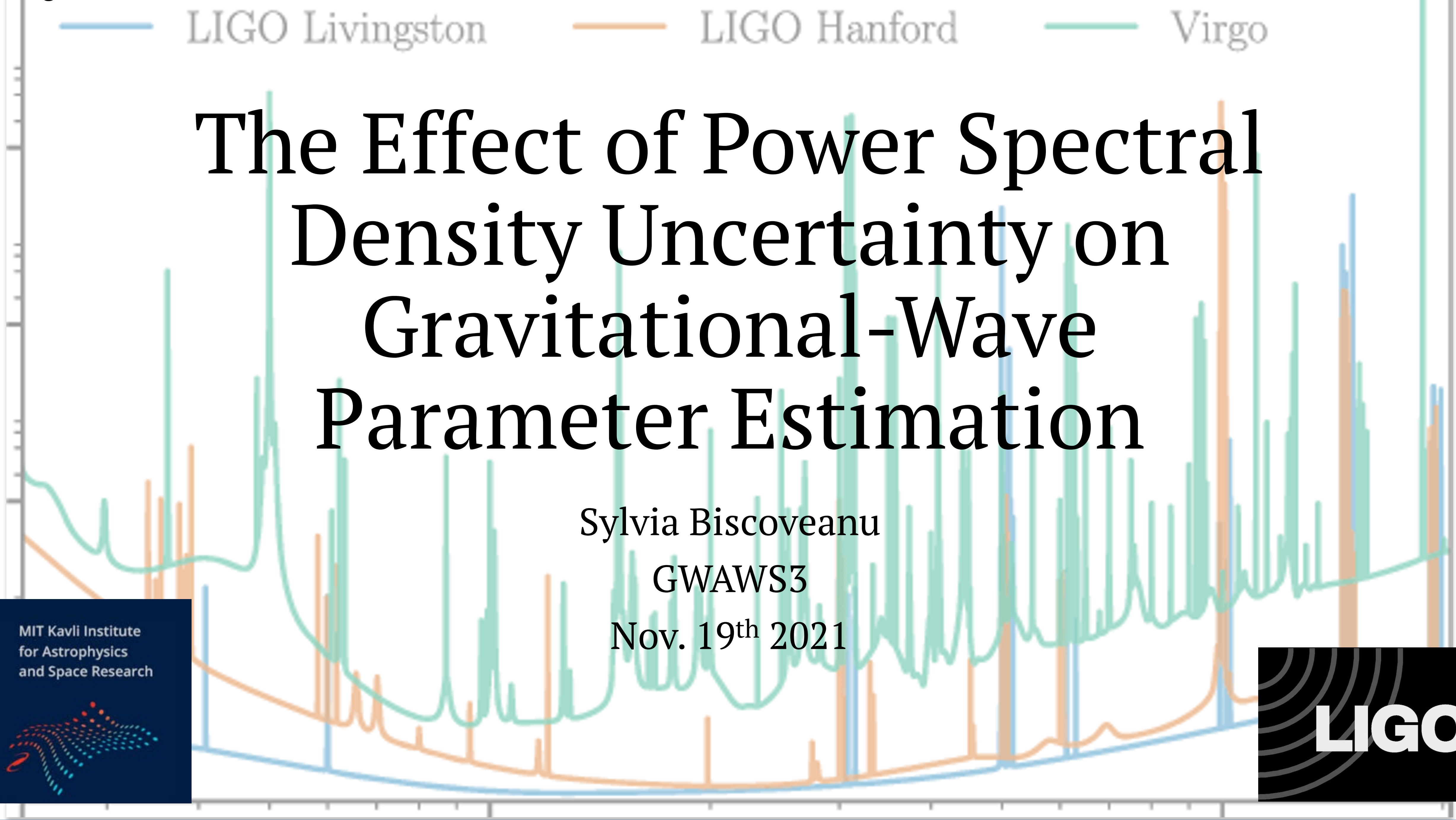
Image: LSC 1805.11579



The Effect of Power Spectral Density Uncertainty on Gravitational-Wave Parameter Estimation

Sylvia Biscoveanu

GWAWS3

Nov. 19th 2021

MIT Kavli Institute for Astrophysics and Space Research

LIGO

LIGO Noise Properties

- The data consists of both a noise contribution and an astrophysical component

$$\tilde{d}(f) = \tilde{n}(f) + \tilde{h}(\boldsymbol{\theta}; f)$$

noise — astrophysical contribution

- The noise is typically assumed to be stationary and Gaussian and is characterized by the power spectral density (PSD)

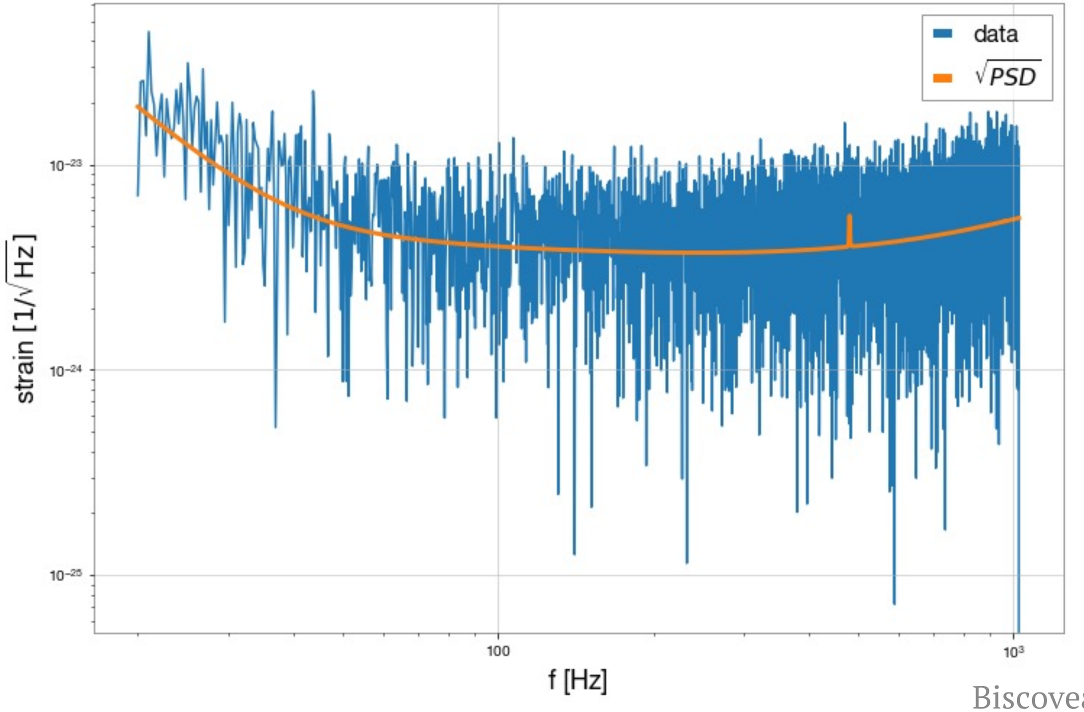
$$\langle \tilde{n}^*(f_i)\tilde{n}(f_j)\rangle = \frac{T}{4} S_n(f)\delta_{ij}$$

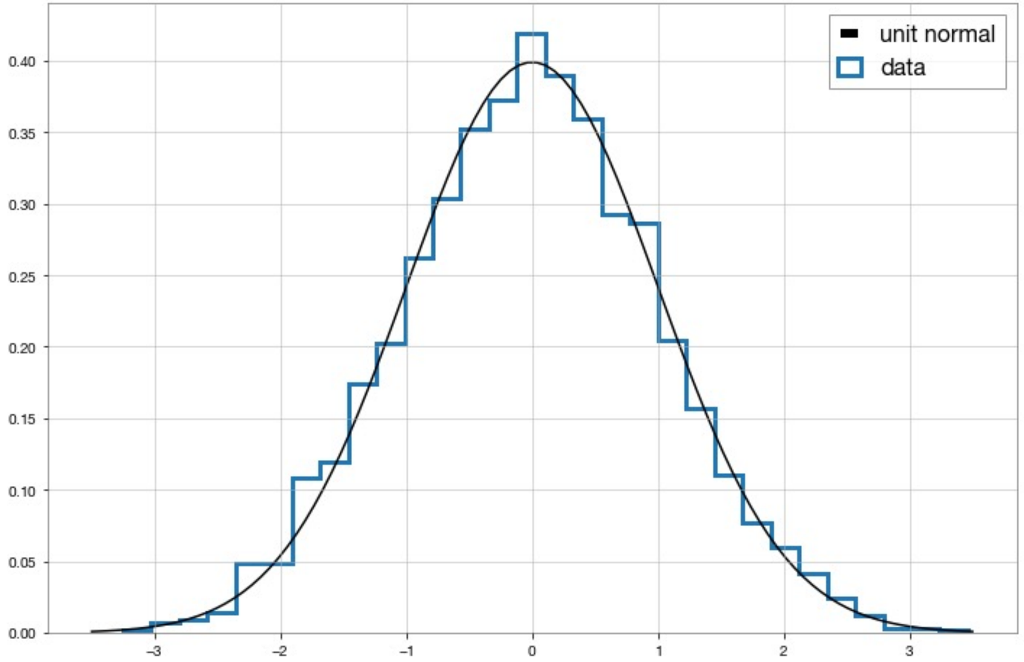
- T is the segment duration, PSD has units of [1/Hz]

LIGO Noise Properties

- For well-behaved noise in the absence of a signal, the real and imaginary parts of the strain each follow a zero-mean Gaussian distribution with variance:

$$\sigma_i^2 = \frac{T S_n(f_i)}{4}$$





The Gravitational-Wave Likelihood

- The residual is the difference between the data and the signal template:

$$\tilde{r}(\boldsymbol{\theta}; f_i) = \tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)$$

- In the presence of a signal, the real and imaginary parts of the residual should also be Gaussian-distributed:

$$p(\Re\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\Re\tilde{r}(\boldsymbol{\theta}; f_i))^2}{2\sigma_i^2}\right)$$

$$p(\Im\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(\Im\tilde{r}(\boldsymbol{\theta}; f_i))^2}{2\sigma_i^2}\right)$$

The Gravitational-Wave Likelihood

- The PSD (and some normalization factors) is the variance of the likelihood:

$$\sigma_i^2 = \frac{TS_n(f_i)}{4}$$

- The total likelihood of the data is the product of the real and imaginary likelihoods:

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}) \equiv p(\Re\tilde{d}(f_i)|\boldsymbol{\theta})p(\Im\tilde{d}(f_i)|\boldsymbol{\theta})$$

The Whittle Likelihood

- The final form of the likelihood used in gravitational-wave data analysis is the Whittle Likelihood:

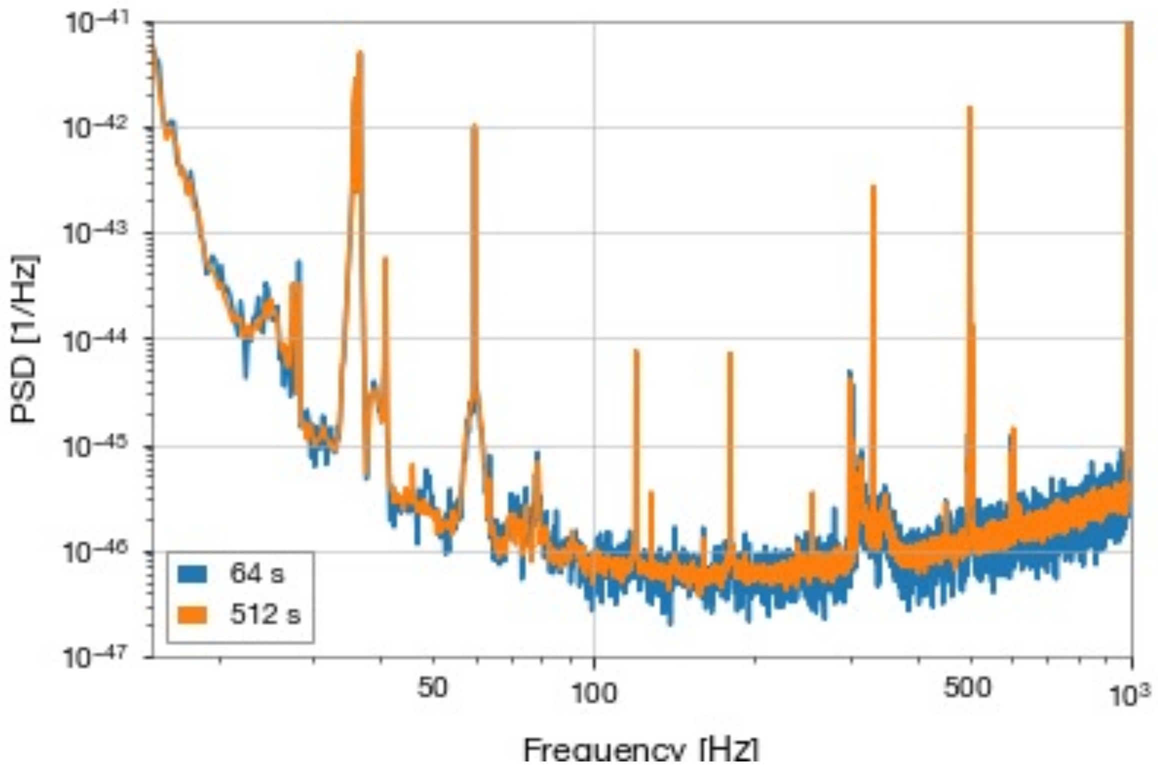
$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}) = \frac{2}{T\pi S_n(f_i)} \exp\left(-\frac{2|\tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)|^2}{T\ S_n(f_i)}\right)$$

$$\mathcal{L}(d|\boldsymbol{\theta}) = \prod_i \mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta})$$

- So far we have assumed that the true PSD, $S_n(f_i)$, is known.
- We can only get an uncertain *estimate* of the true PSD, typically calculated using one of two methods.

Calculating the PSD – Off-source

- Off-source method
  - Also called the periodogram method or Welch method
  - Use a long stretch of data either before or after but always excluding the analysis segment
  - Split the data into short segments and calculate $|\tilde{d}(f_i)|^2$ for each segment after windowing the data
  - Take the median or mean of the periodograms from each short data segment



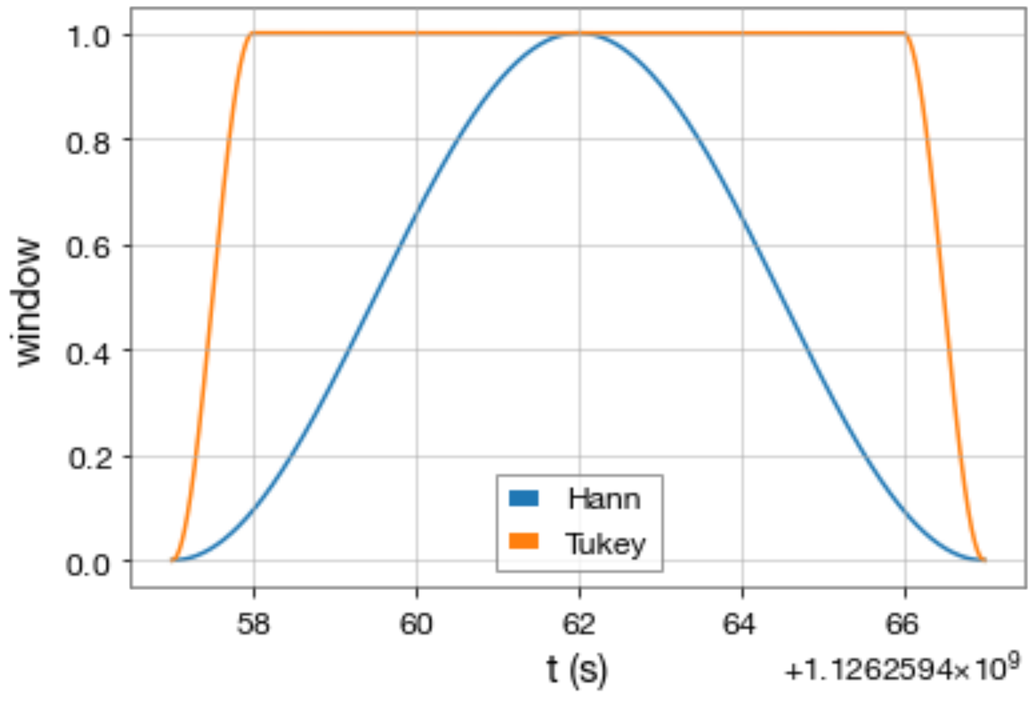
Off-Source Method

- Two corrections need to be applied:
  - Window factor to correct for power lost to window, w_t:

$$W = \frac{1}{N_t} \sum_{t=0}^{N_t - 1} w_t^2$$

  - Median correction, where ℓ is the segment index for odd number of segments

$$\alpha = \sum_{\ell=1}^{N_s} \frac{(-1)^{\ell+1}}{\ell}$$



Allen+ 0509116

Uncertainty in the PSD – Off source

- Define a normalized periodogram for a single segment, Q_ℓ:

$$Q_\ell = 4|\tilde{d}_\ell(f_i)|^2/TS_n(f_i)$$

- Q_ℓ is the quadrature sum of two independent standard normal random variables (the normalized real and imaginary parts of the data)
- Probability of Q_ℓ given by the Chi-squared distribution with two degrees of freedom
- Want to know the probability of the true PSD given the estimated PSD, $\pi(S_n(f_i)|\,\widehat{S_i})$

Off source uncertainty – single segment

- If we were to take the periodogram of a single segment as the estimator of the true PSD, $\widehat{S}_i = Q_\ell S_n(f_i) = 4\left|\tilde{d}(f_i)\right|^2/T$

$$Q \equiv \hat{S}_i / S_n(f_i)$$

Jacobian

$$\pi(S_n(f_i)|\hat{S}_i) = \pi(Q|\hat{S}_i)|dQ/dS_n(f_i)|$$

$$= \chi_2^2(Q)\hat{S}_i/S_n^2(f_i)$$

Off source uncertainty – mean

- Now define the estimator of the true PSD to be the mean of the periodograms across several segments:

$$\hat{S}_i = \frac{1}{N_\ell} \sum_{\ell}^{N_\ell} \frac{4|\tilde{d}_\ell(f_i)|^2}{T} = \frac{S_n(f_i)}{N_\ell} \sum_{\ell}^{N_\ell} Q_\ell$$

- Now the probability of QN_ℓ is a Chi-squared distribution with $2N_\ell$ degrees of freedom, so

$$\pi(S_n(f_i)|\hat{S}_i) = \chi^2_{2N_\ell}(QN_\ell)N_\ell\hat{S}_i/S_n^2(f_i)$$

Off source uncertainty – median

- The mean is known to be an unstable estimator of the PSD for gravitational-wave data because it is more sensitive to outliers
- Typically use the median instead:

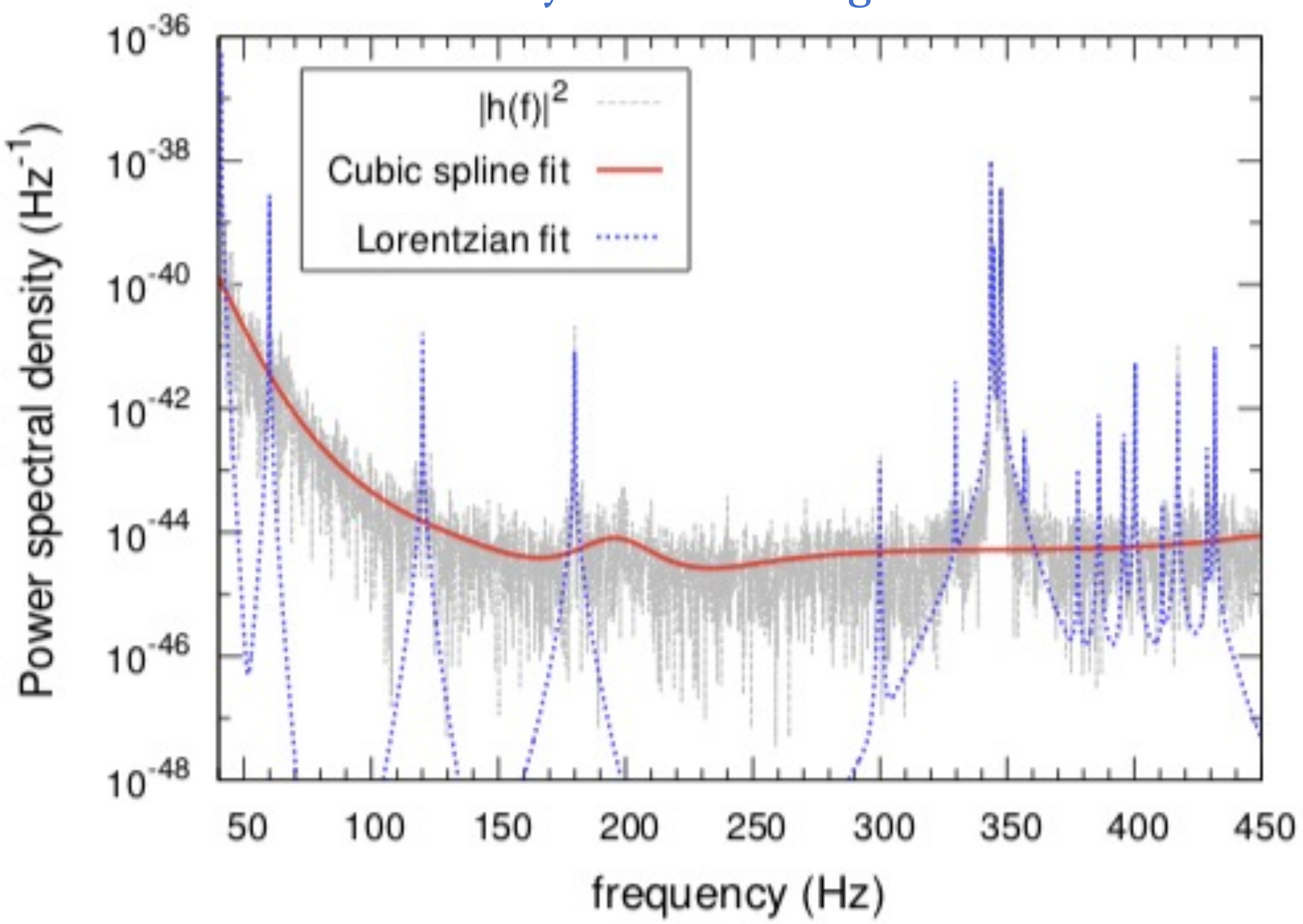
$$\hat{S}_i = \frac{S_n(f_i)}{\alpha}\text{median}(Q_\ell)$$

- If the number of segments is odd, use order statistics to obtain $\pi(S_n(f_i)|\,\widehat{S}_i)$ from the known distribution for a single segment, $\pi(Q_\ell|\widehat{S}_i)$ – Chi-squared with 2 degrees of freedom

Calculating the PSD – On-source

* Tyson Littenberg's talk

- On-source method
  - Model the PSD as a sum of a broadband spline and narrowband Lorentzians using the BayesLine algorithm
  - Using only the data from the analysis segment, infer the spline and Lorentzian parameters that best characterize the PSD
  - Requires significantly less data → more likely that it will be stationary and Gaussian over a shorter period of time

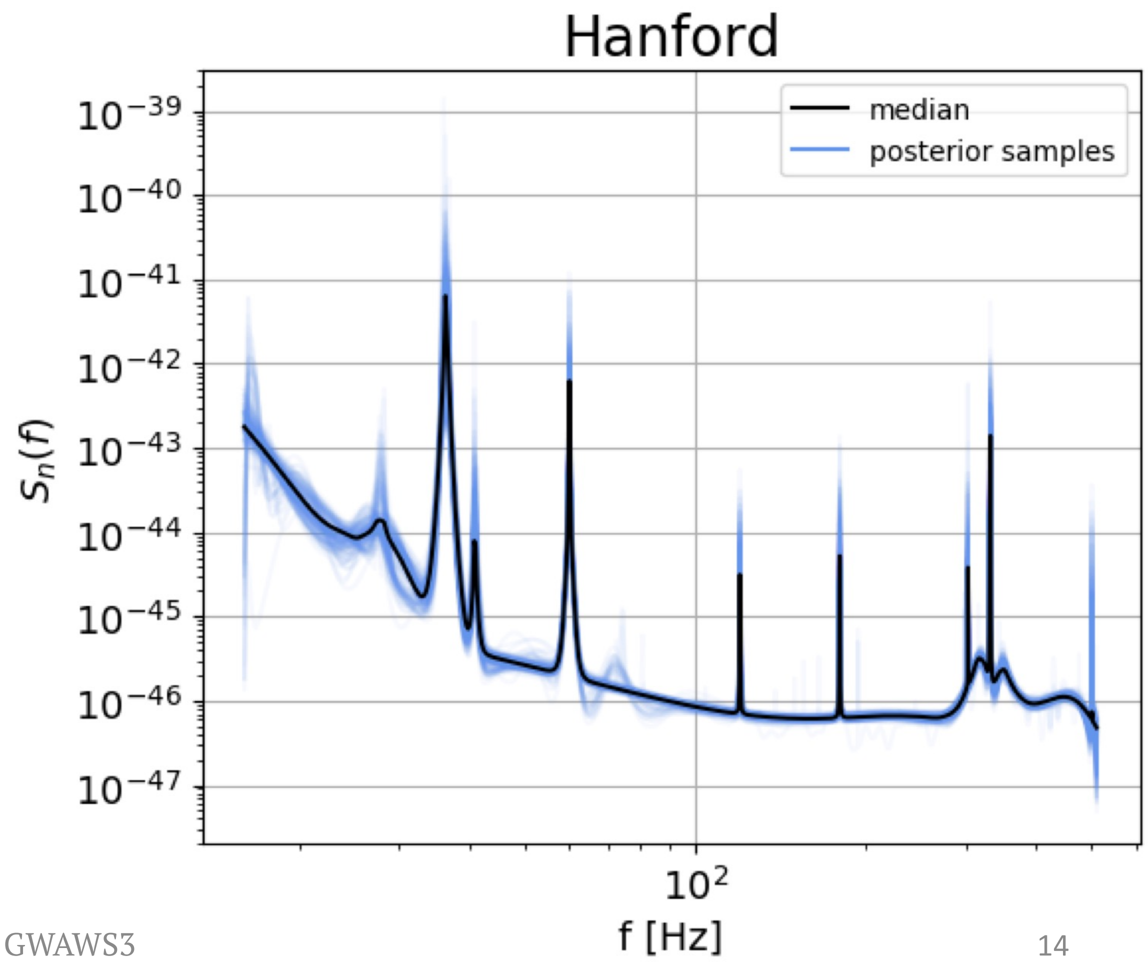


Littenberg and Cornish 1410.3852

On-source uncertainty

- For each of the posterior samples on the spline and Lorentzian parameters, construct a posterior PSD
- Typically ignore the uncertainty and just choose the median, but the true PSD is equally likely to correspond to any of the posterior PSD curves

GW151012



Alternative parameterization

- Hybrid between on-source and off-source approaches, developed in Littenberg+ 1307.8195, Veitch+ 1409.7215
- Add scale parameters multiplying the off-source PSD estimate at fixed, logarithmically-spaced frequency points spanning N_j frequency bins

$$S_n(f_i) \rightarrow \eta_j \hat{S}_i,\ i_j < i \leq i_{j+1}$$

- Prior on η_j is a normal distribution with mean 1 and variance $1/N_j$
- Recover mean off-source uncertainty in the limit that there is one scale parameter per frequency bin and the Gaussian prior is replaced by Chi-squared distribution

PSD marginalization

- Include the PSD as a parameter in the likelihood and marginalize over it:

Whittle likelihood

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta},\hat{S}_i) = \int dS_n(f_i)\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta},S_n(f_i))\pi(S_n(f_i)|\hat{S}_i)$$

- Alternatively if you already have a posterior on the PSD:

$$p(\boldsymbol{\theta}|\tilde{d}(f_i)) = \int p(\boldsymbol{\theta},S_n(f_i)|\tilde{d}(f_i))dS_n(f_i)$$

$$= \int p(\boldsymbol{\theta}|\tilde{d}(f_i),S_n(f_i))p(S_n(f_i)|\tilde{d}(f_i))dS_n(f_i)$$

The Student-Rayleigh Distribution

- Analytically marginalize the Whittle likelihood over the uncertainty for the mean off-source PSD
  - Student-Rayleigh Distribution (Röver+ 0804.3853, Röver 1109.0442)
  - Student-t distribution with $2N_\ell$ degrees of freedom, like a Gaussian but with heavier tails

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}, \hat{S}_i) = \frac{2}{T\pi\hat{S}_i}\left[1 + \frac{2|\tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)|^2}{TN_\ell\hat{S}_i}\right]^{-(1+N_\ell)}$$

The median-marginalized distribution

- Analytically marginalize over the uncertainty in the median PSD estimate given by the median order statistic assuming you have $m = (N_\ell - 1)/2$ measurements less than Q and m measurements greater than Q (Talbot and Thrane 2006.05292):

$$\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}, \hat{S}_i) = \sum_{k=0}^{m} \binom{m}{k} \frac{2(-1)^k}{T\pi\hat{S}_i} \frac{\left(m + k + 1 + \frac{2|\tilde{d}(f_i) - \tilde{h}(\boldsymbol{\theta}; f_i)|^2}{\alpha T \hat{S}_i}\right)^{-2}}{\mathrm{B}(m+1, m+1)}$$

- B is the Beta function

The on-source-marginalized distribution

- Perform a Monte Carlo integral over the PSD posteriors (Biscoveanu+ 2004.05149):

$$p(\boldsymbol{\theta}|\tilde{d}(f_i)) = \int p(\boldsymbol{\theta}|\tilde{d}(f_i), S_n(f_i))p(S_n(f_i)|\tilde{d}(f_i))dS_n(f_i)$$

$$= \frac{1}{N_j}\sum_j p(\boldsymbol{\theta}|\tilde{d}(f_i), S_{n,j}(f_i))$$

- The PSD-marginalized posterior on the binary parameters is the combination of an equal number of posterior samples obtained using each posterior PSD

Caveats

- All approaches to marginalization still assume the data used to calculate the PSD estimator or posterior are stationary and Gaussian
- Analytic off-source marginalization:
  - Requires a longer stretch of data to calculate $\widehat{S}_i$
  - Does not incorporate off-diagonal elements of the frequency-domain noise covariance matrix due to i.e. windows
  - Does not incorporate cross-power between detectors in the joint likelihood
- Numerical on-source marginalization:
  - Requires obtaining posterior samples on the binary parameters independently for hundreds of PSDs
  - Cannot obtain PSD-marginalized evidence estimate using the current outputs of Bayeswave

Aside: likelihood reweighting

- Use likelihood reweighting to obtain posterior samples and evidences under the marginalized likelihoods at a reduced computational cost (Payne+ 1905.05477)

Weights

$$w(\boldsymbol{\theta}) = \frac{\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}, \mathcal{H}_1)}{\mathcal{L}(\tilde{d}(f_i)|\boldsymbol{\theta}, \mathcal{H}_0)}$$

“Target” likelihood (numerator)

”Proposal” likelihood (denominator)

Evidence

$$\mathcal{Z}(\mathcal{H}_1) = \mathcal{Z}(\mathcal{H}_0) \sum_{i}^{\boldsymbol{\theta_i} \sim p(\boldsymbol{\theta}|\tilde{d}(f_i), \mathcal{H}_0)} w(\boldsymbol{\theta_i})$$

Posterior

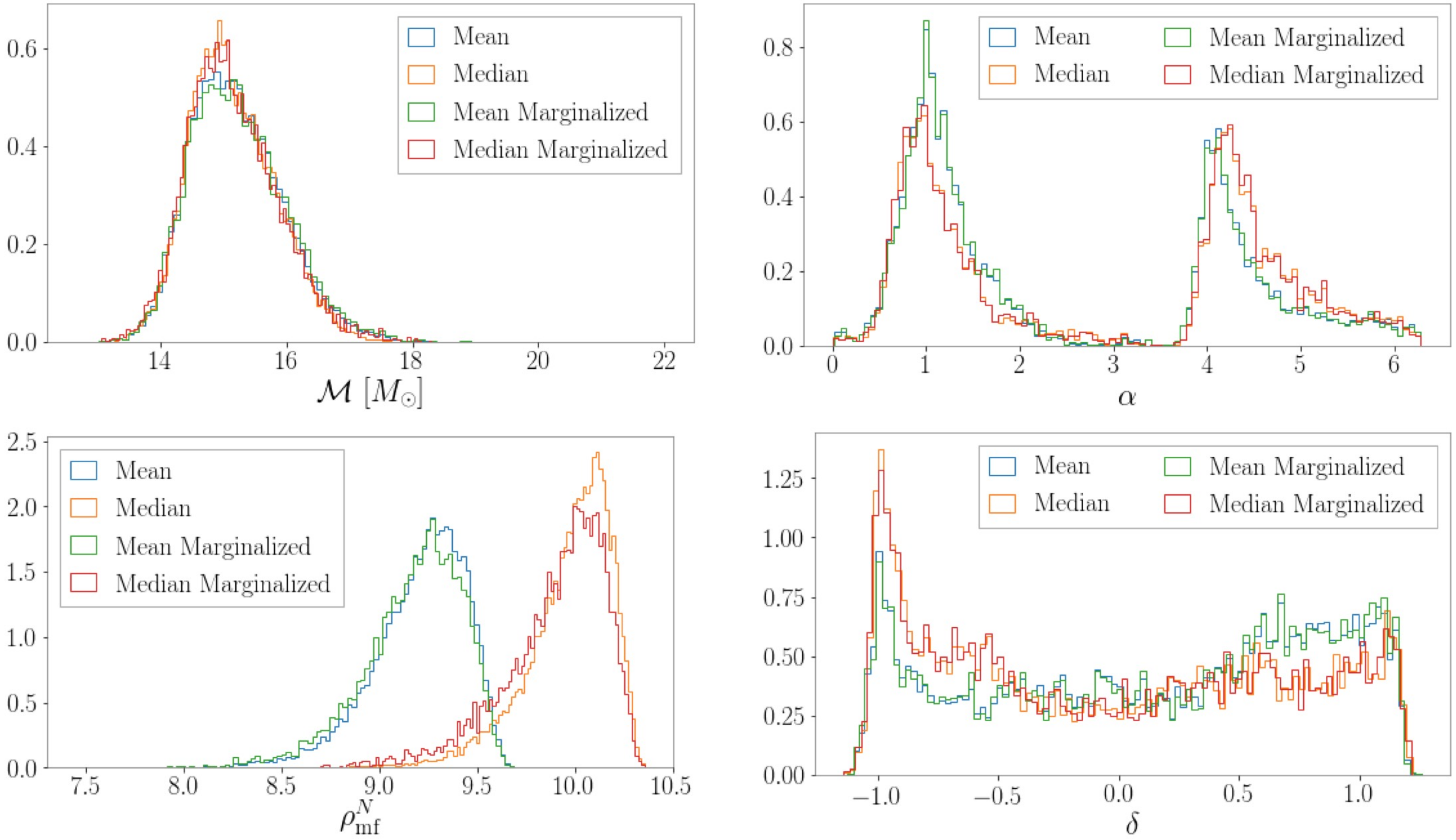
$$p(\boldsymbol{\theta}|\tilde{d}(f_i), \mathcal{H}_1) = \frac{w(\boldsymbol{\theta})}{\sum_i w(\boldsymbol{\theta})} p(\boldsymbol{\theta}|\tilde{d}(f_i), \mathcal{H}_0)$$

* Poster by Bassel Saleh

Posterior comparison – off-source

PSD	Marg vs no marg
Mean	19.26
Median	91.67

Natural log bayes factors show strong preference for marginalized likelihood!

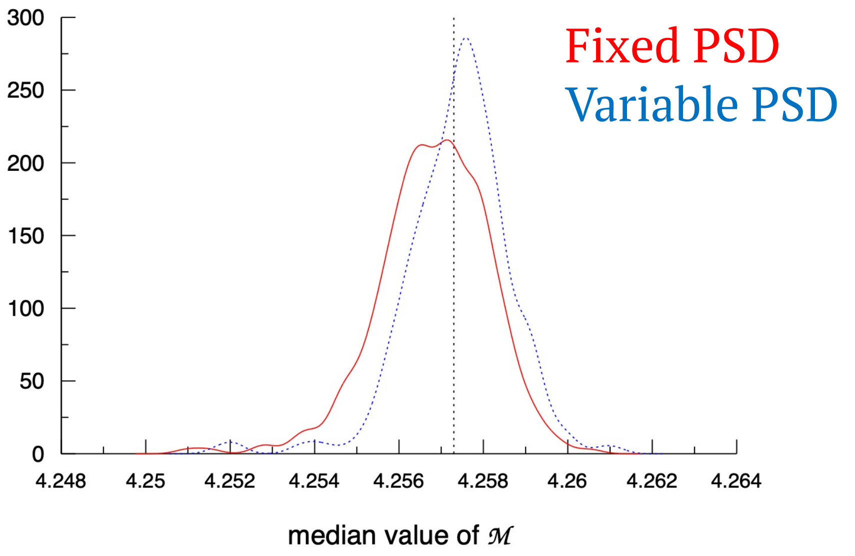


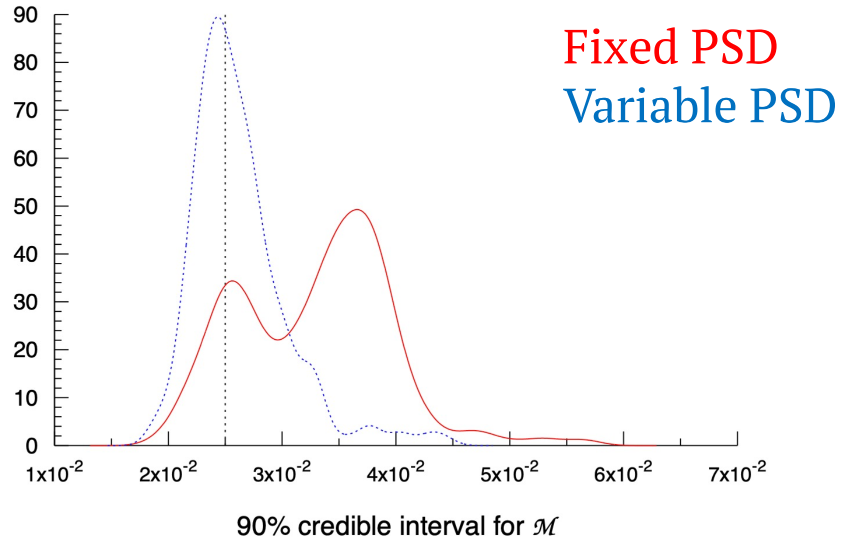
GW151012 results from Talbot and Thrane 2006.05292

Posterior comparison – hybrid approach

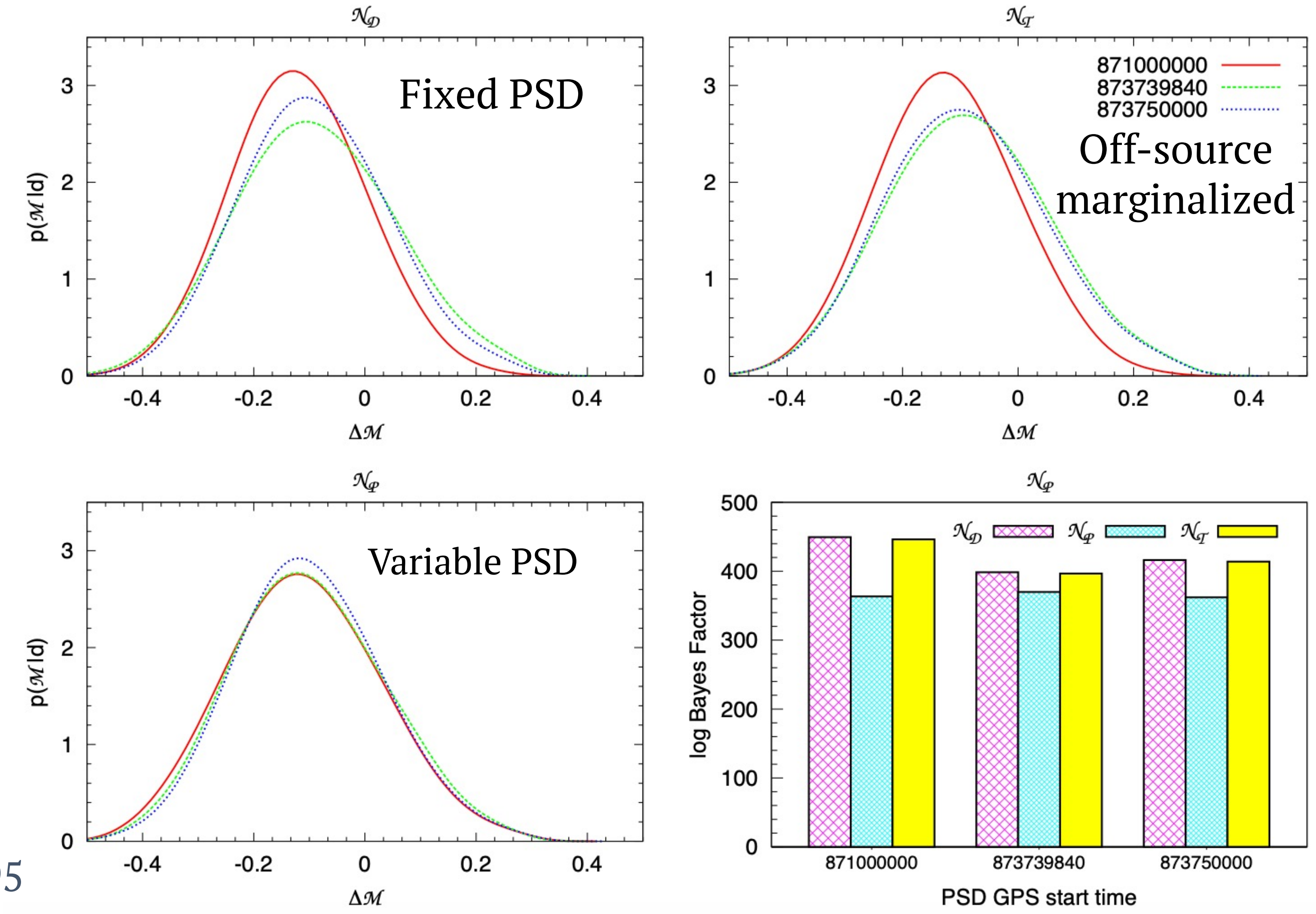
- Perform parameter estimation on the same compact binary signal using 300 different PSDs estimated from different segments of real initial LIGO data
- Compare the distributions of the medians and 90% credible intervals for the chirp mass with and without PSD marginalization

Littenberg+ 1307.8195





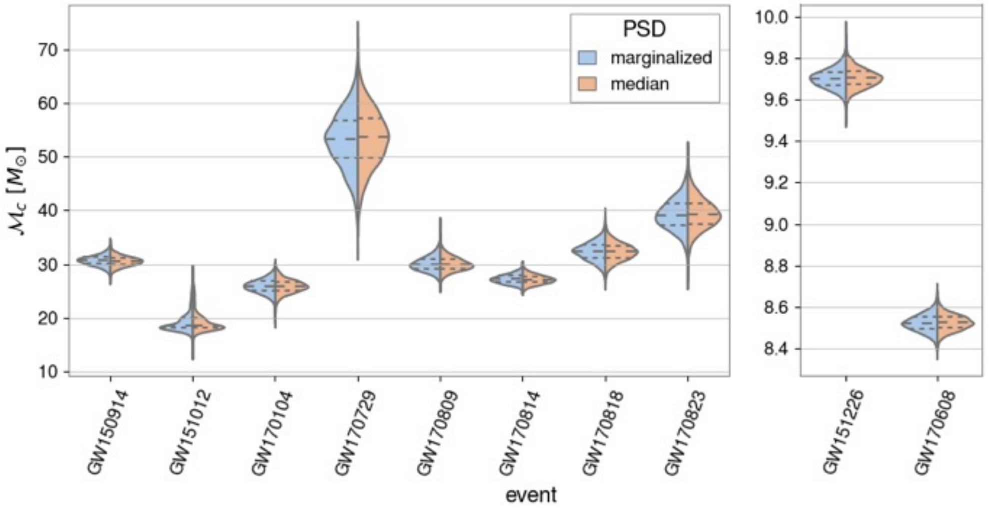
Posterior comparison - hybrid approach

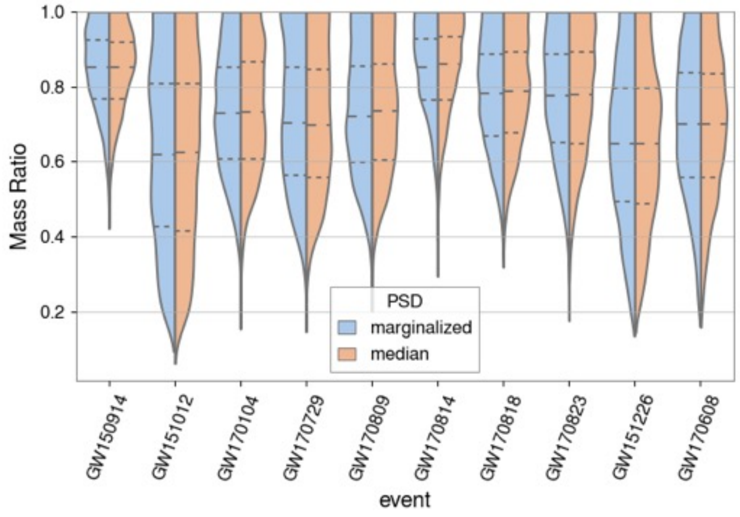


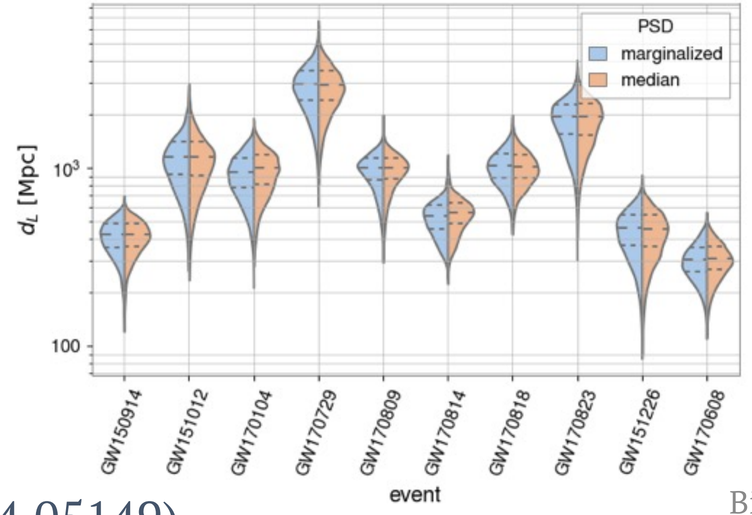
Fixed PSD
Variable PSD
Off-source marginalized

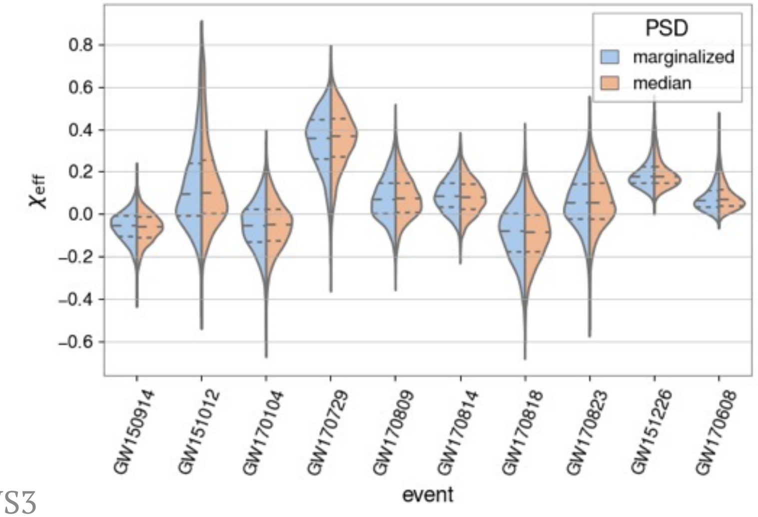
Littenberg+ 1307.8195

Posterior comparison – on-source









(Biscoveanu+ 2004.05149)

Posterior comparison – on-source

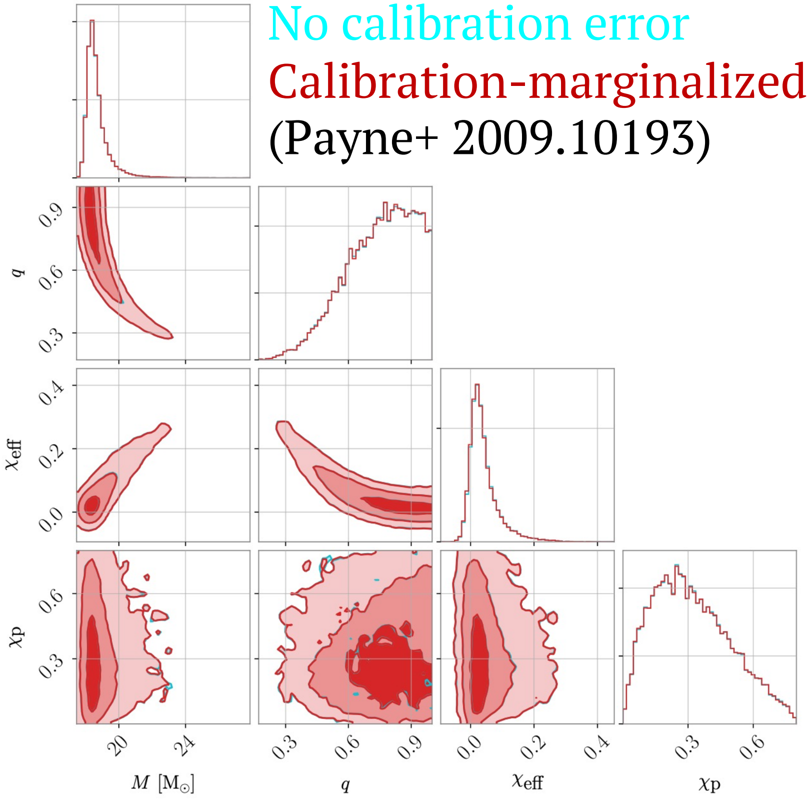
- Change in the width of the 90% credible interval is on the order of 10%
- Posterior variation does not depend on SNR of signal
- Larger variation than when marginalizing over the off-source uncertainty

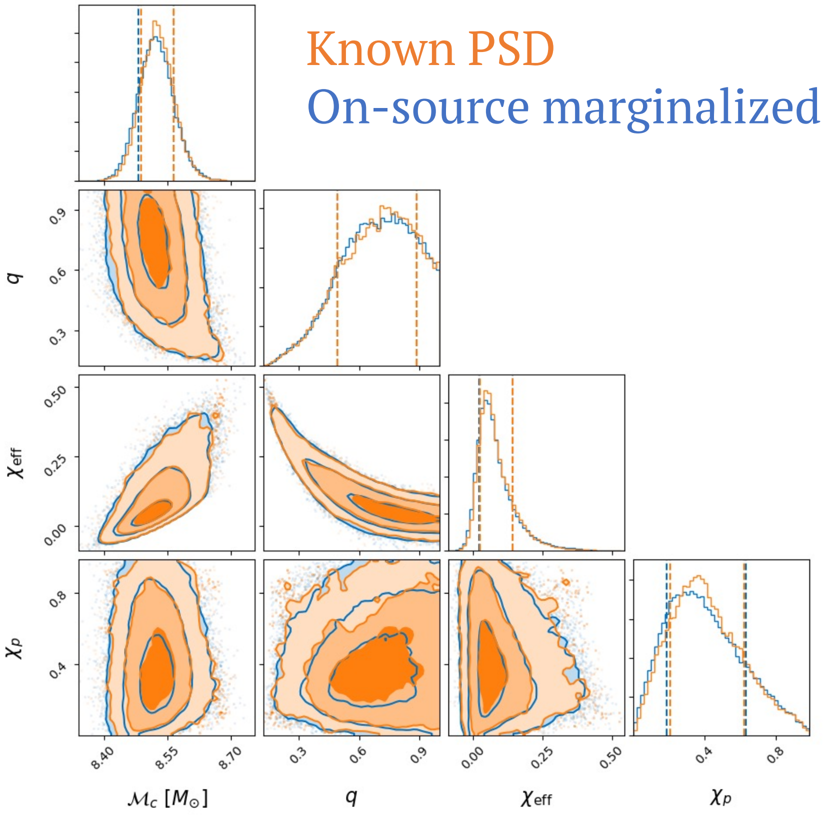
Event	$\Delta\Omega_{50}$ (%)	$\Delta\Omega_{90}$ (%)	$\Delta\Omega_{90}$ (deg^2)
GW150914	13.7	12.3	21
GW151012	1.5	-1.1	-20
GW151226	-9.6	-6.7	-93
GW170104	15.3	7.5	77
GW170608	-12.0	1.9	8
GW170729	19.0	10.4	136
GW170809	-12.1	2.9	9
GW170814	0	-18.6	-24
GW170817	28.6	25.9	7
GW170818	11.1	6.5	2
GW170823	2.9	-0.9	-14

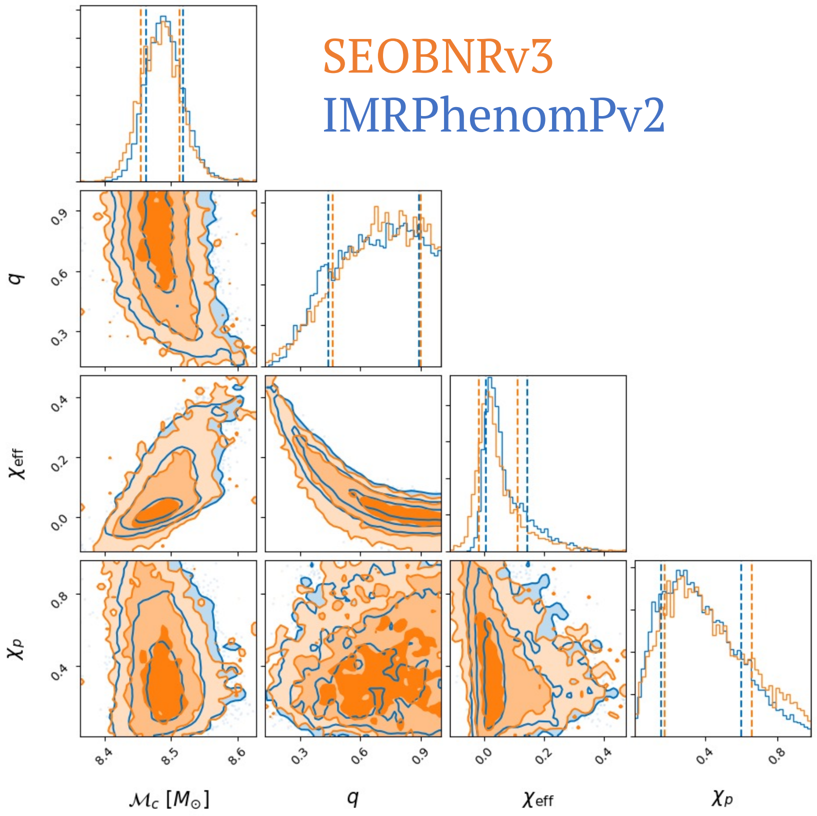
(Biscoveanu+ 2004.05149)

Comparison with other systematics

Ex: GW170608



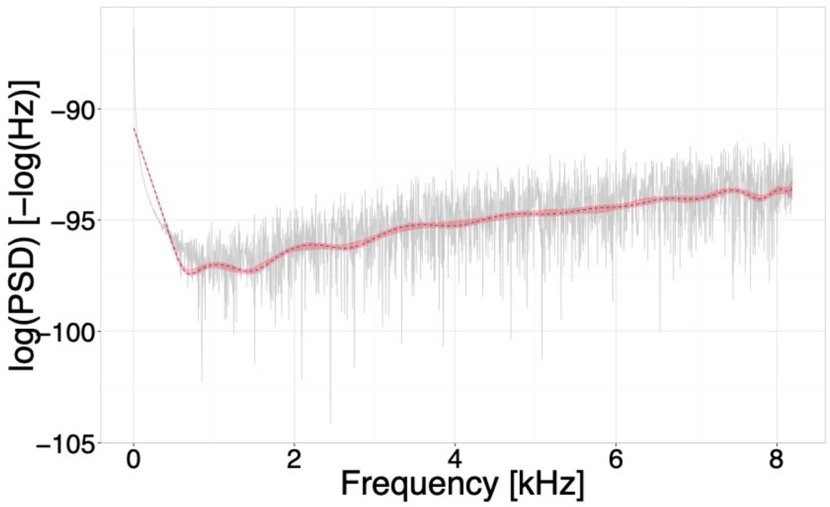




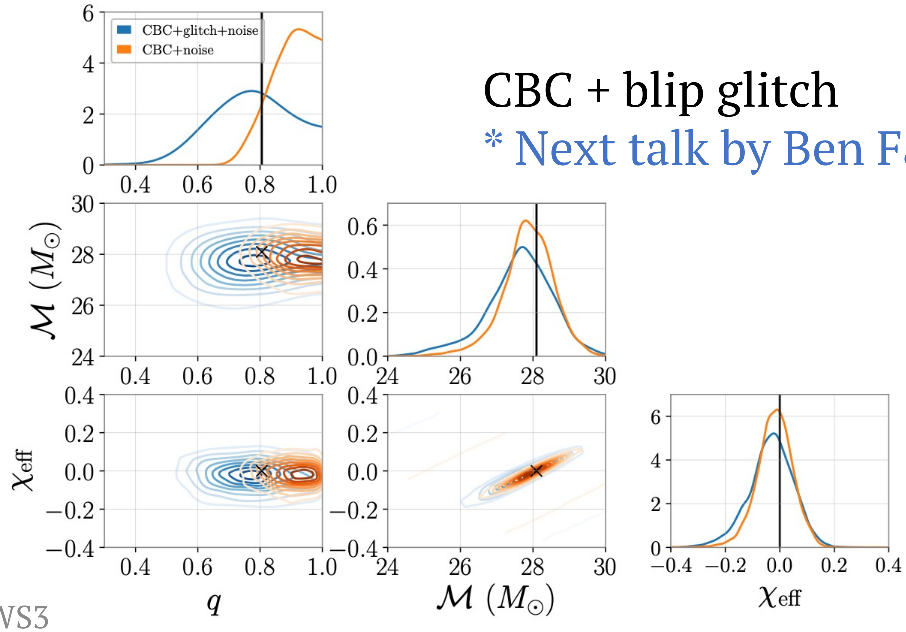
* Geraint Pratten's talk

Alternative approaches – simultaneous fit

- Edwards+ 1506.00185 use a nonparametric Bernstein polynomial prior on the PSD
  - On-source method that does not require assumption of stationary, gaussian noise



- Chatziioannou+ 2101.01200 modify the BayesWave algorithm to fit PSD, CBC signal, and glitches all at once



CBC + blip glitch

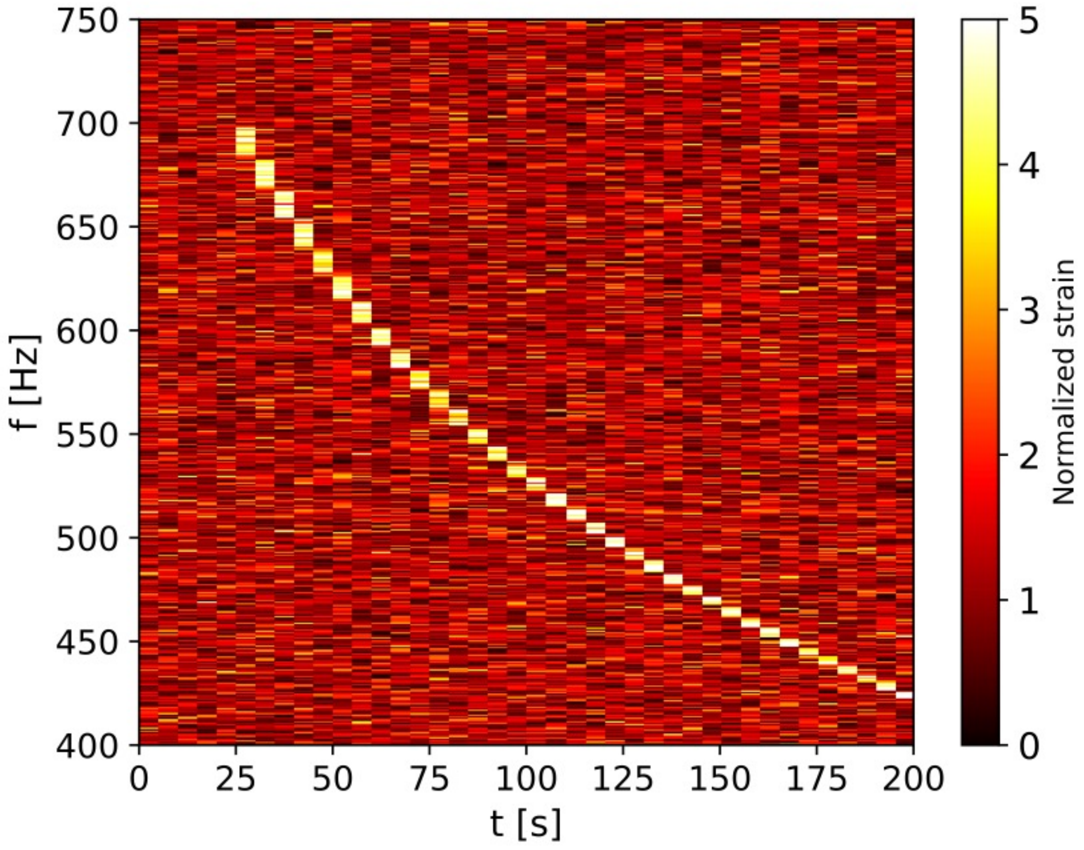
* Next talk by Ben Farr

Applications

- We have seen that with three different methods for marginalizing over the uncertainty in the PSD, the effect on the binary parameter posteriors is small, ~10%
- The Bayes factors seem to be more sensitive to the particular noise model chosen for the off-source and hybrid marginalization approaches
- Incorporating the uncertainty is the PSD is required for analyses that stack bayes factors for individual events or for analyses of individual long-duration transients

Ex: BNS postmerger signal

- Banagiri+ 1909.01934 demonstrate a method to search for long-duration BNS postmerger signals from a spinning-down millisecond magnetar using time-frequency maps
- Likelihood is a function of both the time and frequency indices
- Find a bias in the recovered model parameters when not accounting for the uncertainty in the PSD using the Student-t likelihood



Ex: Bayesian Coherence Ratio

- BCI – Bayes factor between a coherent signal in multiple detectors vs incoherent glitches (Veitch and Vecchio 0911.3820)
- Bayes Coherence Ratio (BCR) – odds comparing the coherent signal hypothesis to the incoherent signal or Gaussian noise hypotheses (Isi+ 1803.09783)

Signal evidence

$$\mathrm{BCI} = \frac{\mathcal{Z}(\mathcal{S})}{\mathcal{Z}(\mathcal{G}_1)\mathcal{Z}(\mathcal{G}_2)}$$

Individual glitch evidences

Prior signal odds

$$\mathrm{BCR} = \frac{\hat{\pi}(\mathcal{S})\mathcal{Z}(\mathcal{S})}{\prod_{i=1,2} \hat{\pi}(\mathcal{G}_i)\mathcal{Z}(\mathcal{G}_i) + (1 - \hat{\pi}(\mathcal{G}_i))\mathcal{Z}(\mathcal{N})}$$

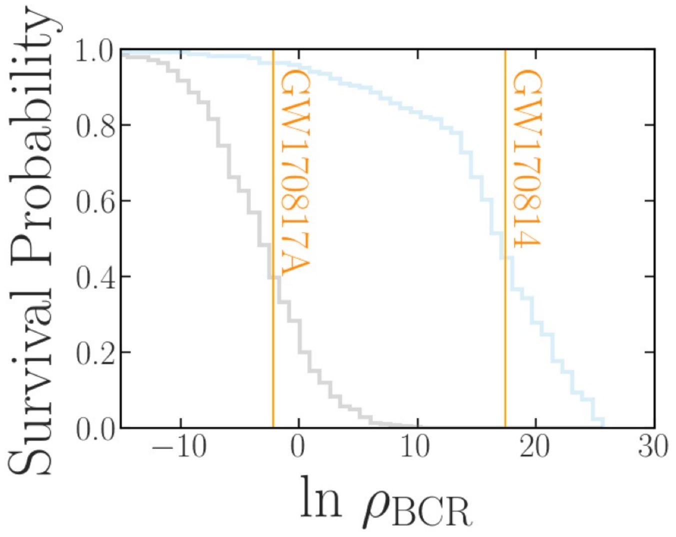
Prior glitch odds

Noise evidence

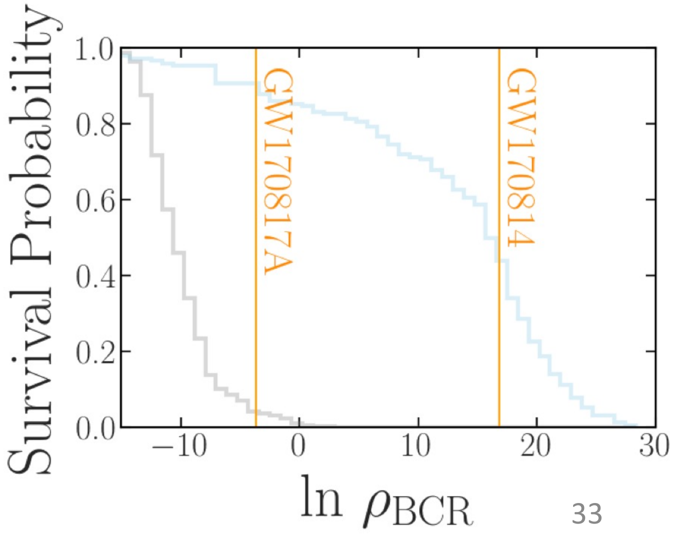
Ex: Bayes Coherence Ratio

- Vajpeyi+ 2107.12109 present a search for intermediate-mass black hole signals in data from LIGO's third observing run using the BCR as a ranking statistic
- Tune the prior signal and glitch odds to separate the background and foreground BCR distributions

Known PSD



Off-source marginalized



Ex: Astrophysical odds

- Infer the signal and glitch prior odds using Bayesian inference (Ashton+ 1909.11872, Ashton and Thrane 2006.05039)

"Duty cycle" → $\langle\xi\rangle$

$$\mathcal{O}^{S_j}_{N_j} \approx \frac{\langle\xi\rangle \mathcal{Z}_j(\mathcal{S})}{\int \mathcal{Z}_j(\mathcal{NG}, \Lambda_{\mathcal{NG}})\pi(\Lambda_{\mathcal{NG}}|d_{j\neq k})d\Lambda_{\mathcal{NG}}}$$

Noise model is a mixture model of Gaussian noise and glitch in one or both detectors

Noise model hyper-parameters including distribution of glitch "masses" and "spins" and glitch rate

Ex: Astrophysical odds

- Ashton and Thrane 2006.05039 calculate the astrophysical odds of three candidate signals from LIGO's first observing run
- Find that ignoring PSD uncertainty produces false positive signals with odds > 1 in time-slid data

Known PSD

Event	GstLAL	PyCBC	1-OGC	2-OGC	IAS	$\langle \xi \rangle$	$\hat{\xi}_g^{\mathrm{H}}$	$\hat{\xi}_g^{\mathrm{L}}$	$\ln B^{G}_{\mathrm{S/N}}$	$\ln B_{\mathrm{S/N}}$	$\ln B_{\mathrm{coh,inc}}$	ln BCR	$\ln \mathcal{O}$	$1 - p_{\mathrm{astro}}$
GW150914	$< 10^{-3}$	$< 10^{-3}$	$< 8\times10^{-4}$	$< 10^{-3}$	–	7.4×10^{-4}	0.0094	0.013	307	205	12.5	14.3	16.2	9×10^{-8}
GW151012	0.001	0.04	0.024	$< 10^{-3}$	–	7.4×10^{-4}	0.031	0.021	28.2	13.2	9.63	5.64	5.74	0.003
GW151216	–	–	0.997	0.82	0.29	7.4×10^{-4}	0.022	0.016	12.7	3.70	3.10	-3.53	-3.50	0.97

Off-source marginalized

Ex: Non-gaussian stochastic background

- Estimate the fraction of segments that contain a signal ("duty cycle", ξ) by performing Bayesian parameter estimation on each segment of data (Smith and Thrane 1712.00688)

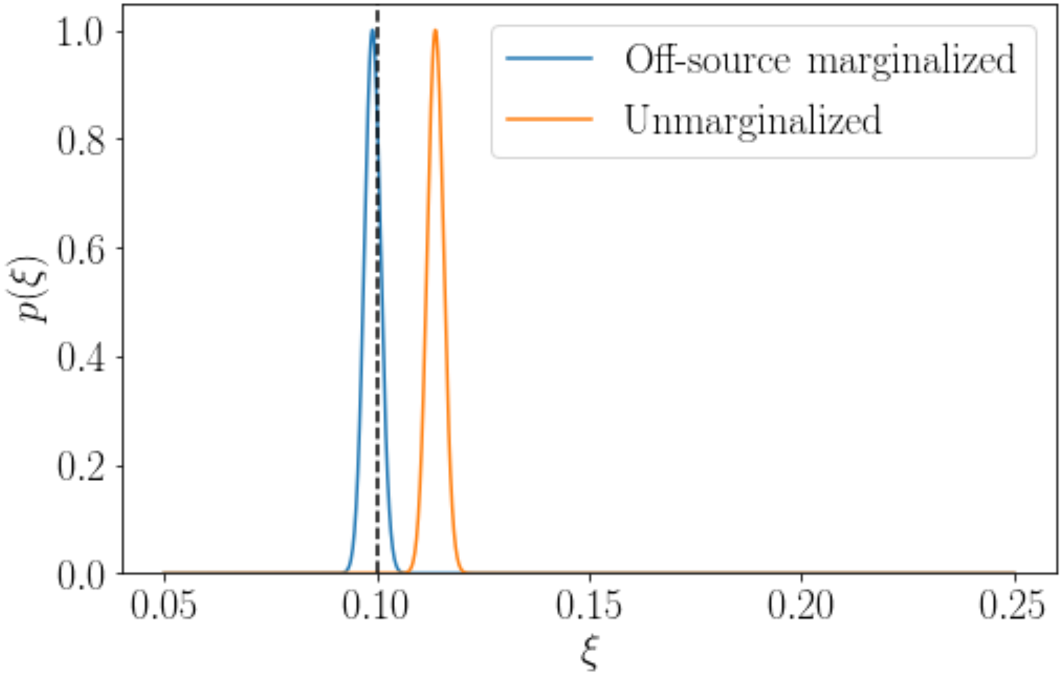
$$\mathcal{L}(\{d\}|\xi) = \prod_j (\xi \mathcal{Z}_j(\mathcal{S}) + (1-\xi)\mathcal{Z}_j(\mathcal{N}))$$

- The statistically optimal method to search for a background of BBH mergers, which occur every ~200 seconds in the universe

* Arianna Renzini's talk

Ex: Non-gaussian stochastic background

- Need accurate evidence estimates to be sensitive to the weakest signals
- Bias in the recovered duty cycle without using PSD marginalization
- Also need to account for off-diagonal elements of the PSD covariance matrix when applied to windowed data



Plot: Colm Talbot

Summary

- Two main methods for calculating and marginalizing over the uncertainty in the PSD
  - Off-source, analytic marginalization
  - On-source, numerical marginalization
- Variation in the posterior on the order of 10%, bayes factors more sensitive
- Cannot neglect this effect when analyzing:
  - Long signals with time-frequency maps
  - An ensemble of data segments to determine Bayesian-based significance or signal probability

Uncertainty in the PSD – Off source

- Reminder: the real and imaginary parts of the data in a given segment are individual zero-mean Gaussian random variables with $\sigma_i^2 = TS_n(f_i)/4$
- The off-source PSD is constructed by averaging the periodogram across several segments

$$|\tilde{d}_\ell(f_i)|^2 = \Re\tilde{d}_\ell(f_i)^2 + \Im\tilde{d}_\ell(f_i)^2$$

- Define a normalized periodogram for a single segment, Q_ℓ:

$$Q_\ell = 4|\tilde{d}_\ell(f_i)|^2/TS_n(f_i)$$

Bayesian Model Selection

- Bayes factor: evidence ratio

$$\mathrm{BF}^S_N = \frac{\mathcal{Z}_S}{\mathcal{Z}_N}$$

- Odds ratio: bayes factor weighted by prior odds

$$\mathcal{O}^S_N = \mathrm{BF}^S_N \frac{\pi(S)}{\pi(N)}$$