Panel: Model selection with gravitational wave observations

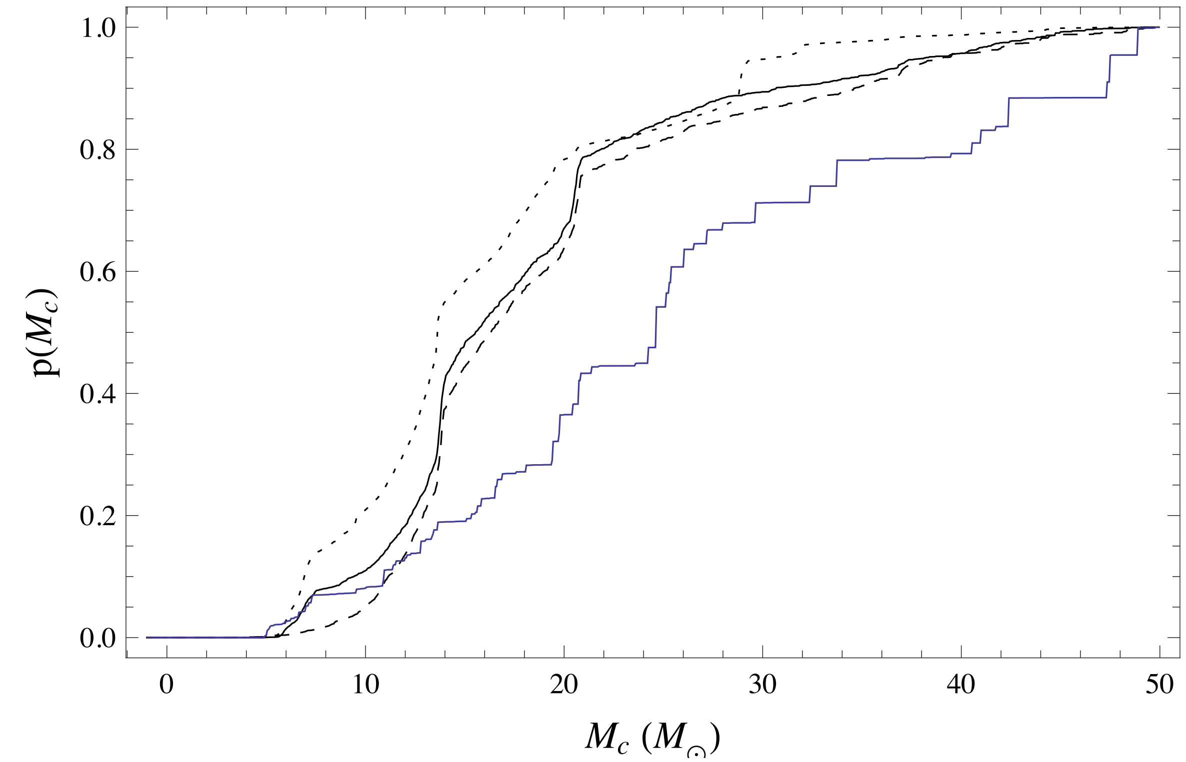
Strong Gravity and Binary dynamics, Oxford, MS
2017-02-27

Panel:
Richard O'Shaughnessy
Salvatore Vitale
Chris Pankow
Simon Stevenson

Rules of thumb

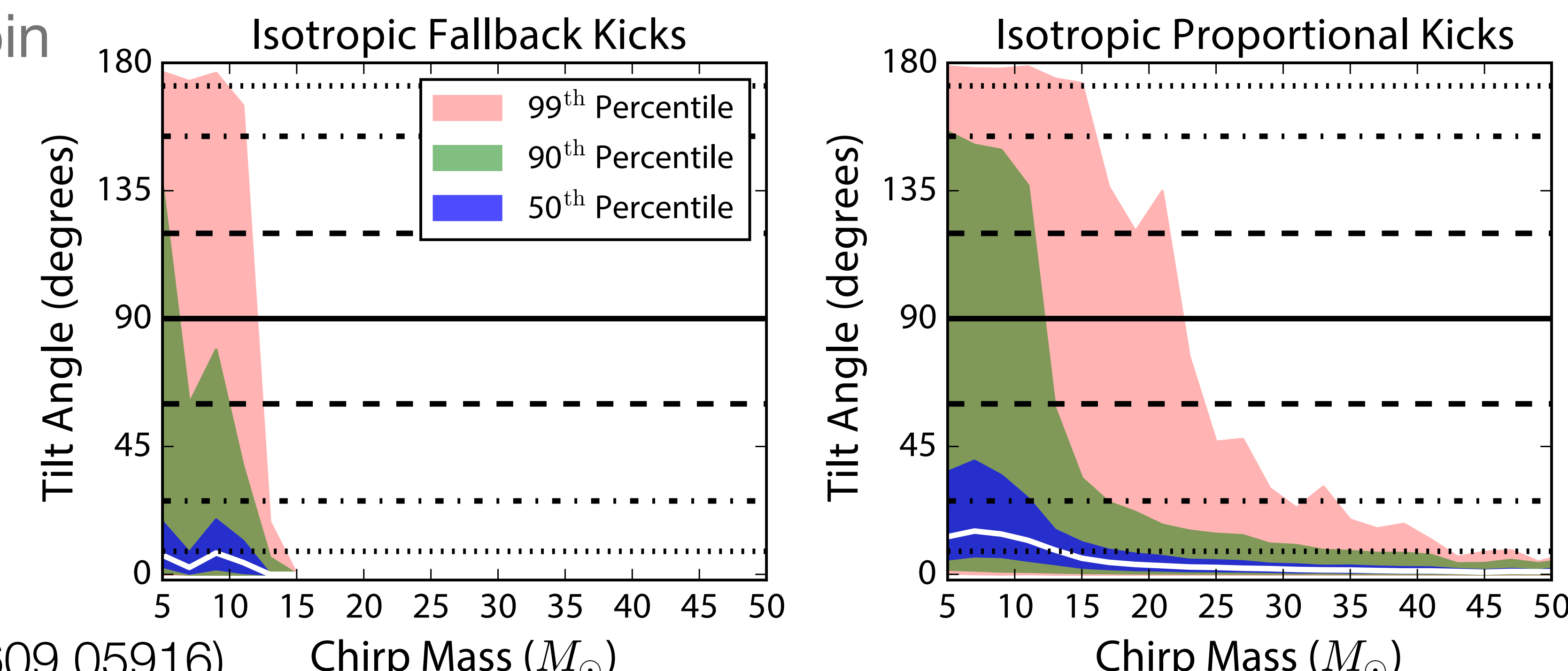
- Distributions more useful than rates
  - Scale tricky (IMF, all past SFR, distribution of conditions, many channels)
  - Functions can encode an **infinite** number of parameters [e.g., ROS PRD 2013]
  - At least one distribution (chirp mass) is easy to measure

(**Detected** distribution)



Dominik et al (2015: 1405.7016)

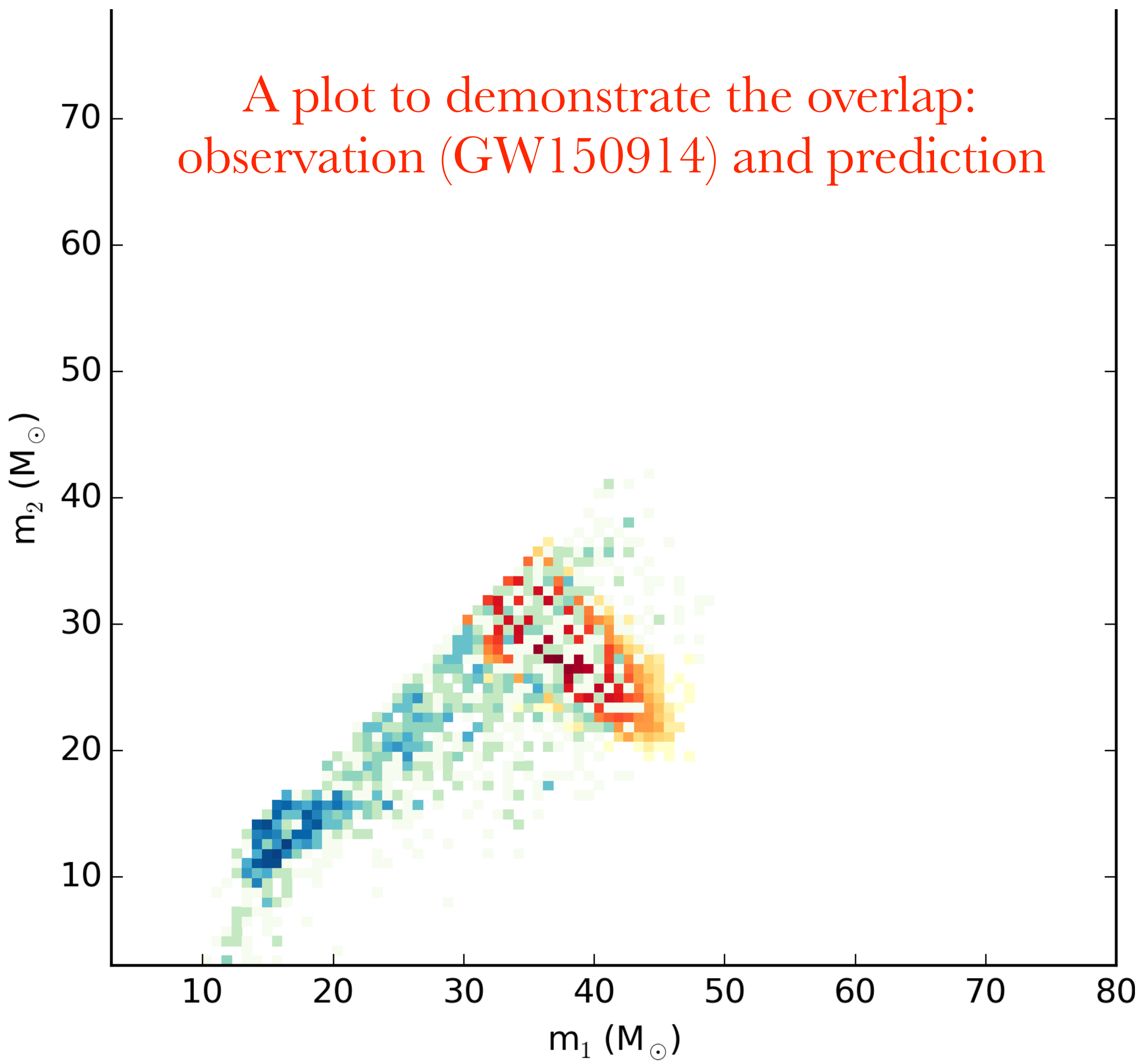
- Robust observables are tricky
  - Cluster formation: strong spin misalignment = $\chi_{\mathrm{eff}} < 0$
  - Mass gaps
    - (e.g., pair instability SN)

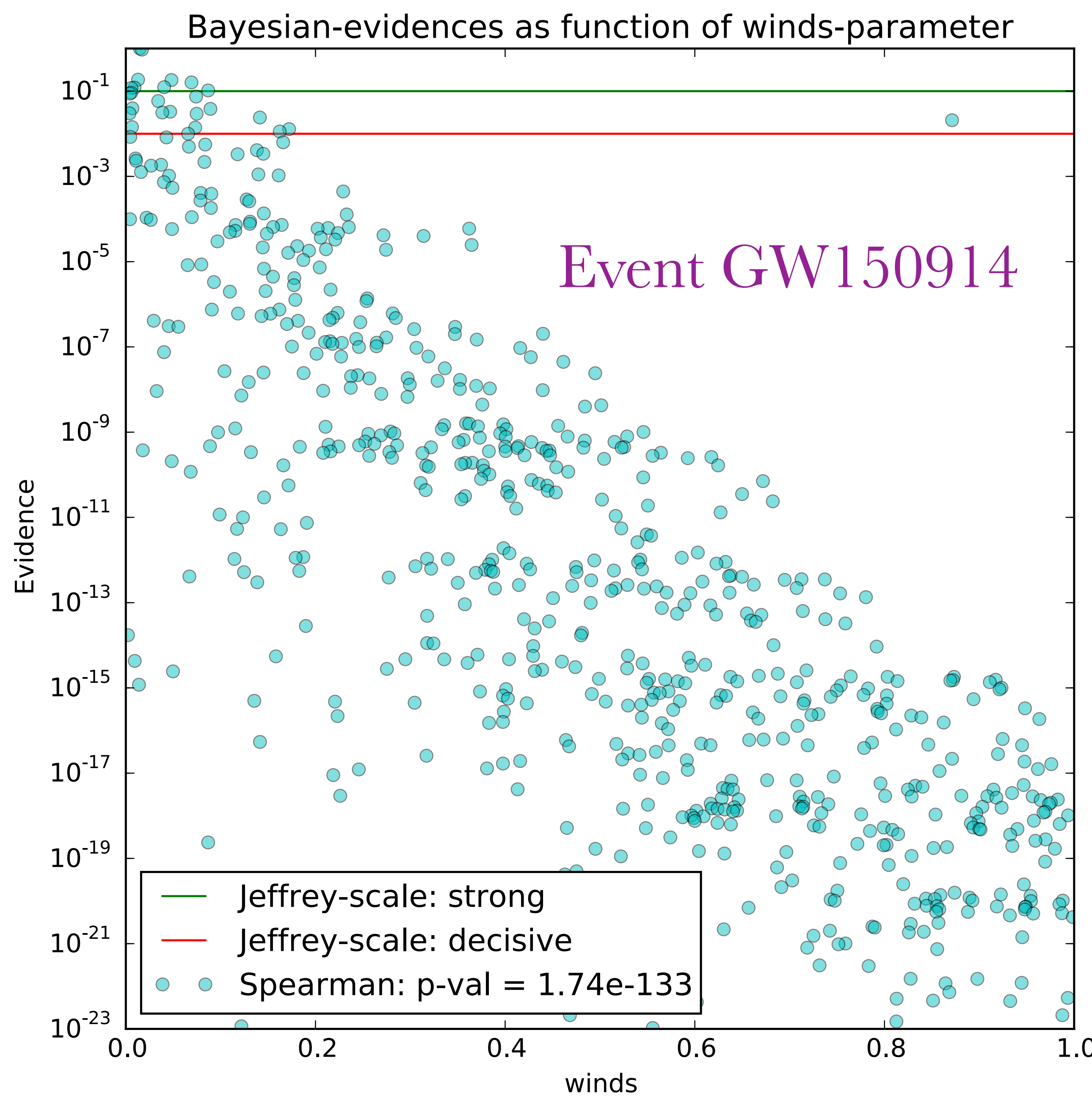


Rodriguez et al (2016: 1609.05916)

Do we need more proof of concept calculations?

- Several for discrete model selection or ad-hoc mixtures, but…
- When do we / how do we measure real parameters?

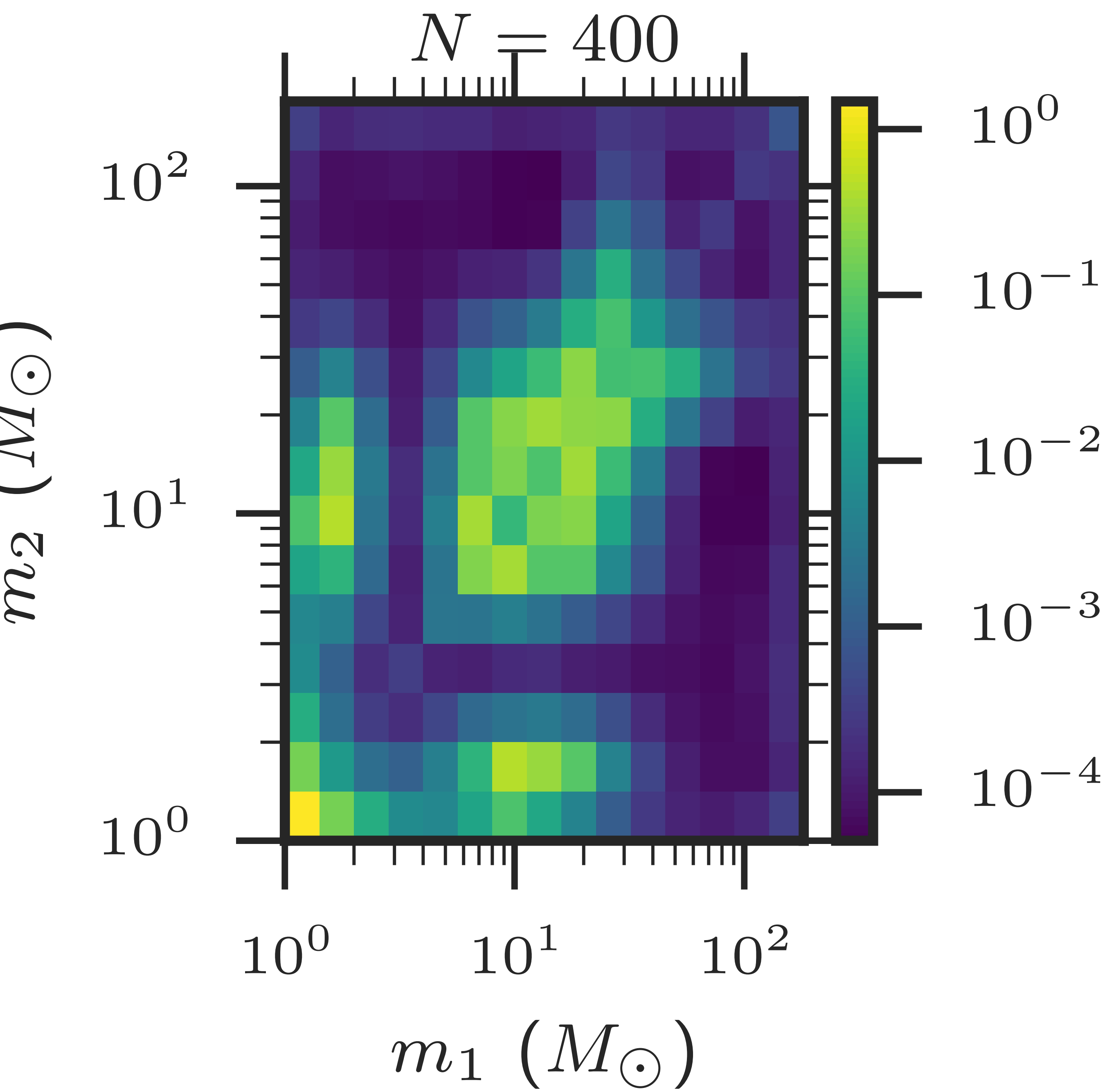




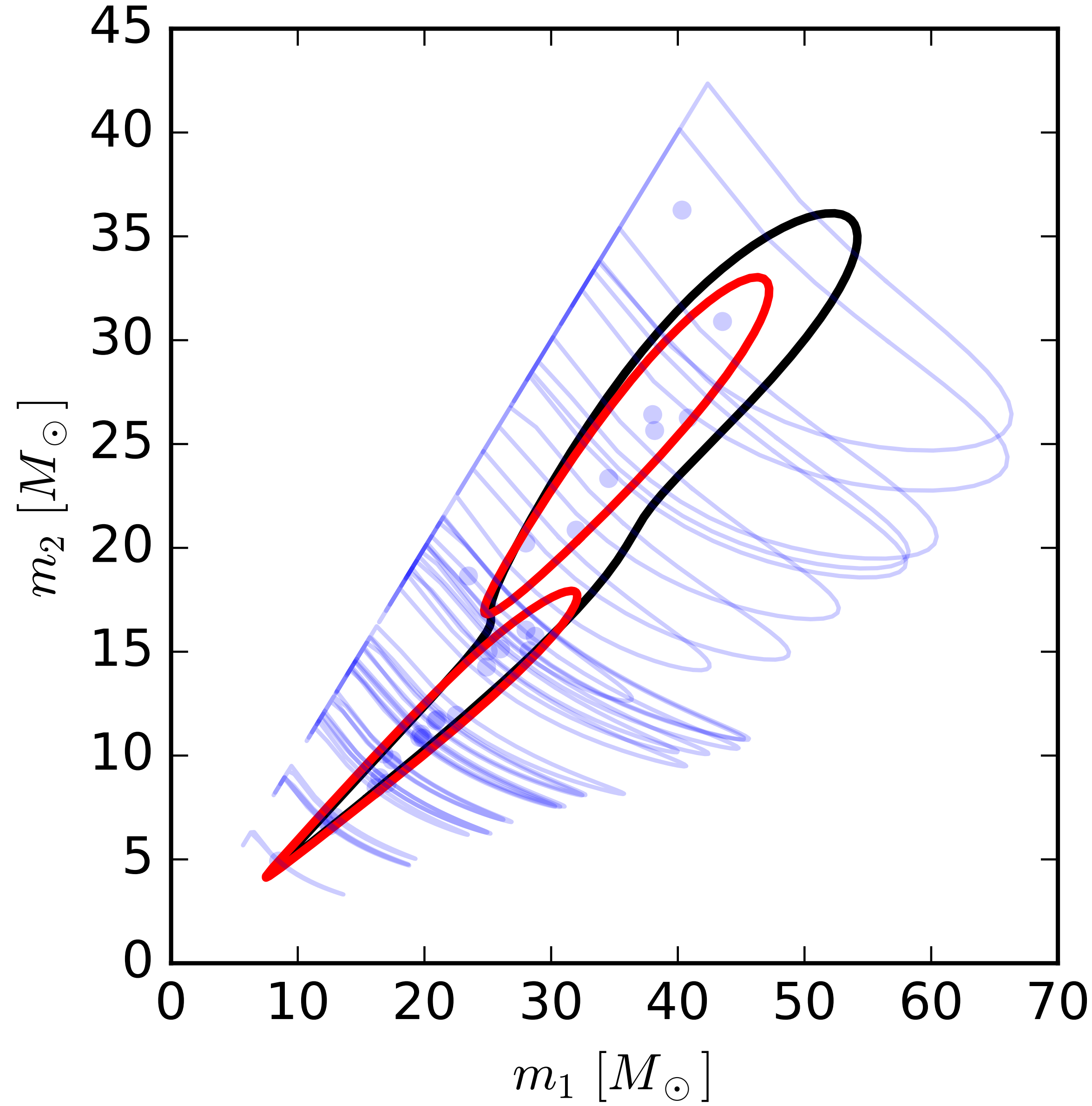
Mukherjee, ROS et al (in prep)
[based on ROS et al 2008,2010]

Reconstructing and reporting the observations

- Density estimation



Mandel et al 2017 MNRAS



Wysocki, ROS (in prep)
[includes spin, real measurement errors]

How many events to distinguish populations?

- KL divergence: unambigous way to compute average information gain per event

$$D_{KL}(p|q) \equiv \int dx p \ln p/q$$

  - Standard tool in probability and statistics
  - Arbitrary dimensions / # of observables. Coordinate-system independent
  - Includes measurement error, selection bias (=apply to observed distribution)

$$\left\langle \hat{L}(X)/n \right\rangle = \left\langle \frac{\ln p(n|\mu)}{n} \right\rangle - [H_{p_*} + D_{KL}(p_*|p)]$$

ROS PRD 2013

- Trivial to use for for toy models (e.g., power laws, gaussians, …)

- Hard part:
  - Evaluating & exploring the model space with sufficient accuracy
  - KL divergence is infinitely sensitive to gaps / exclusions, which are always decisive
  - As written, distinguishes two models (=points in hyperparameter space), not family

Salvo Vitale

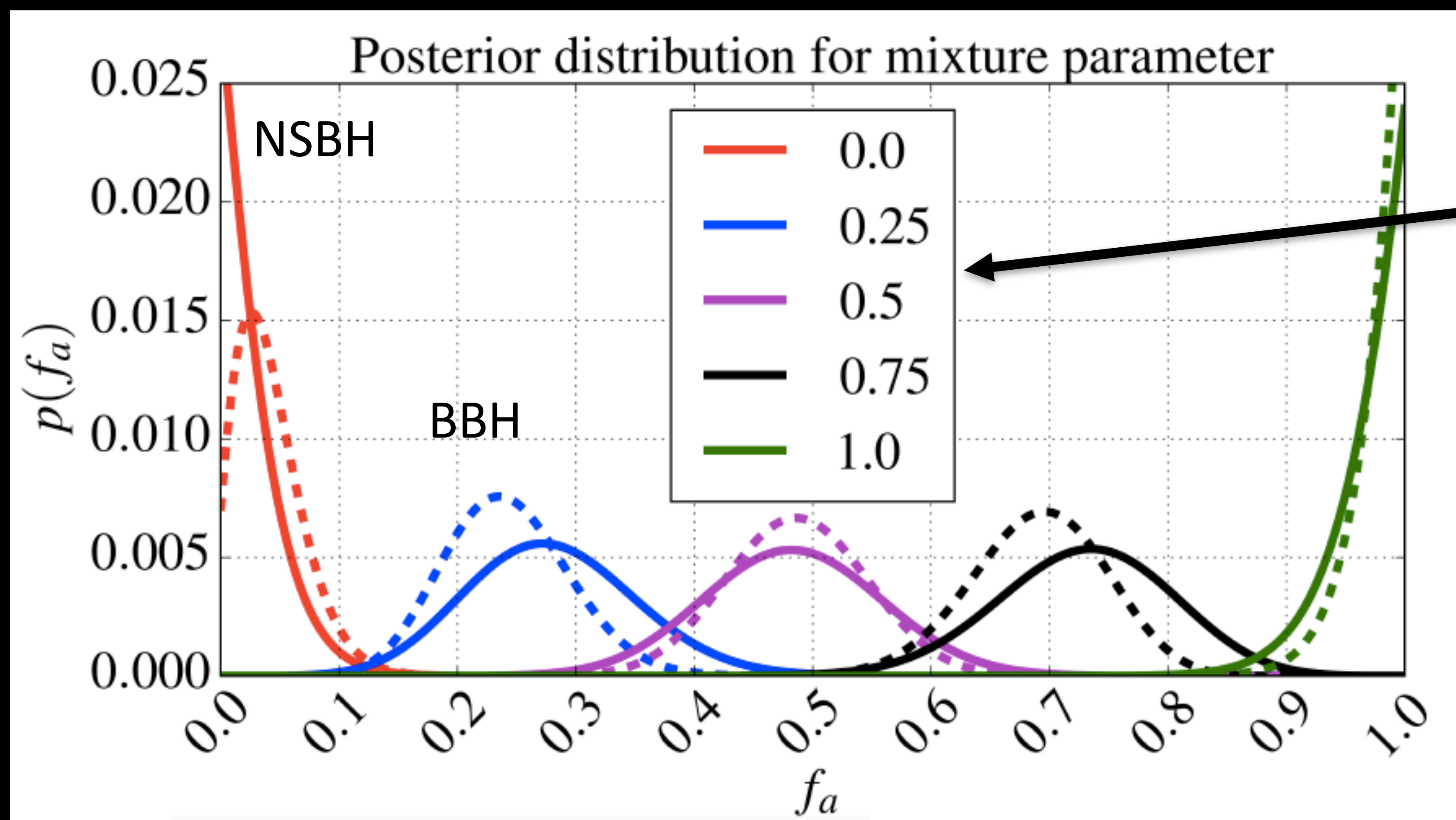
- Questions from Richard & audience
  - Systematics: The approximants are approximate. How do you build confidence in the result given uncertainty in strong merger?
    - What about NR (higher modes)? Precession? Uncertain high PN terms (tides?)
  - Calibration errors: How can we test GR or measure EOS in future instruments, given systematic amplitude and phase errors?
  - Dependence on parameters: What if tides / modified GR effects depend sensitively on nature of binary? How do we stack them?
  - Prior: past infinity or in band?

TIGER - caveats

- Odds in favor of modGR **not** necessarily equivalent to "GR is wrong"
- Could be that waveform model is inappropriate to start with
- Something weird with the data or calibration
- Unaccounted (GR) physics
  - E.g. non-linear NS tides (Essick+ 2016)
- Priors on GR parameters (?)
- Most of these effects shown to be under control in Agathos+ 2013

Measuring the mixture fraction

100 NSBH
200 BBH



True underlying fraction of aligned sources

S. Vitale, Aug 26 2016

27

Caveats – To dos

- Assumed what I called "aligned” is what the universe calls aligned – should include possible prior mismatch
- Can extend the model so that they also take into account mass ratios, eccentricity, or anything else that might be useful to distinguish
- Can include more than 2 models

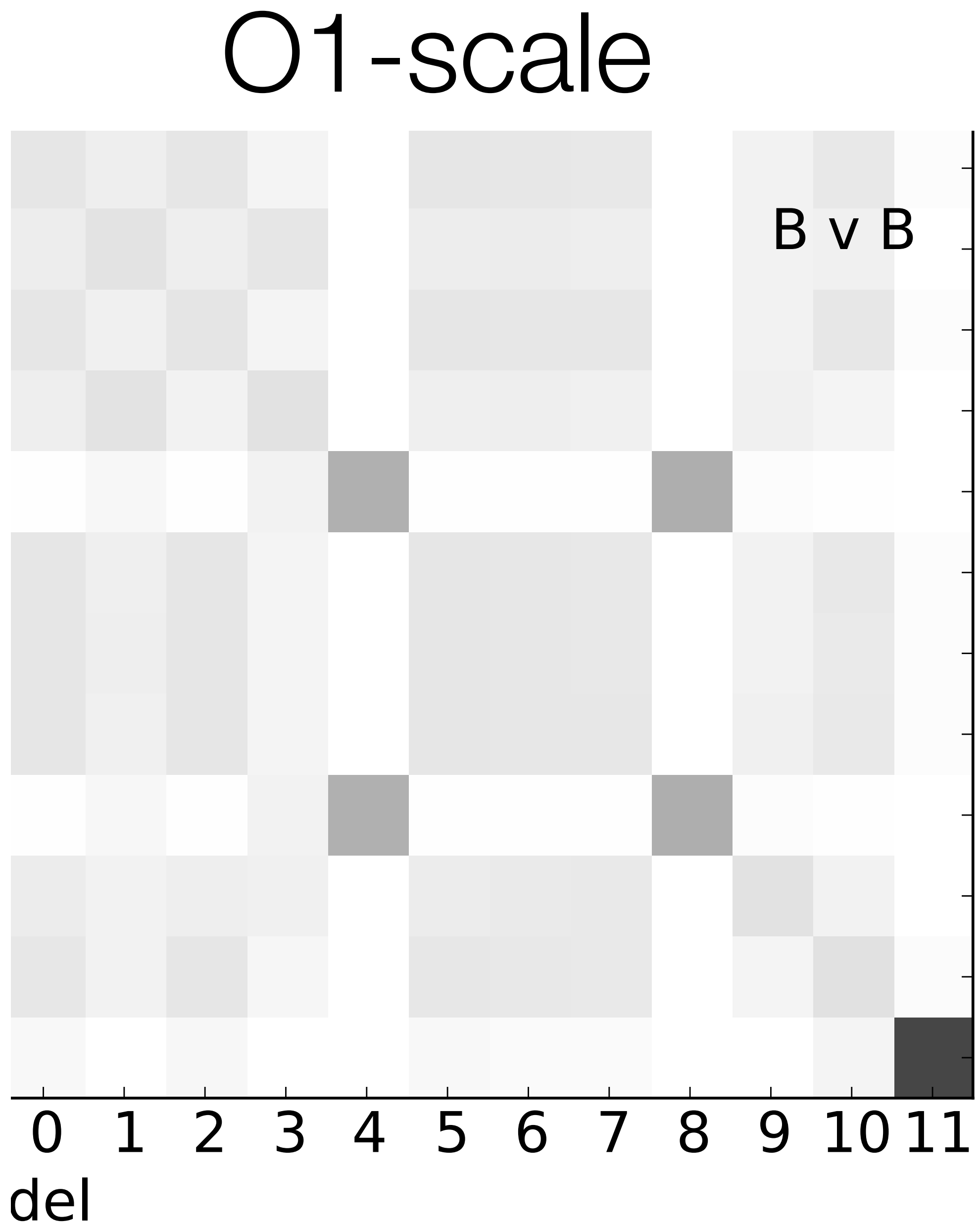
Chris Pankow

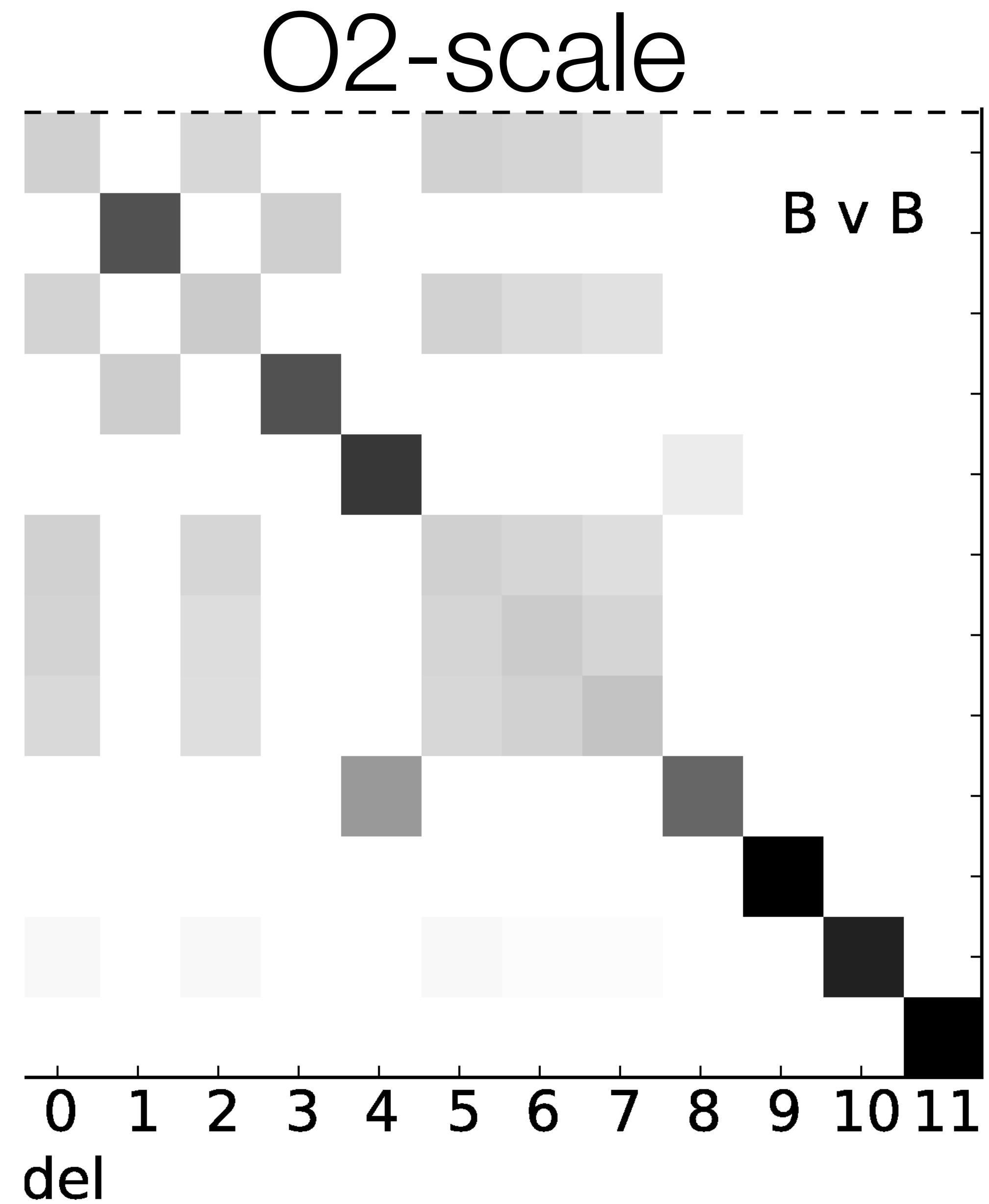
- Questions from Richard and audience
    - Does reweighting posteriors work?
    - How do we deal with selection bias of real searches against interesting things (e.g., precessing; modified GR; …)

Simon Stevenson

- Questions from Richard and audience
  - Joint constraints: How can you do multi-observation constraints with an interpolated model? Interpolate all observations?
  - [Technical] How does interpolation work safely and with high contrast? Basis functions for log(rate)?
  - [Technical] Are you also interpolating <u>observable </u>universe (selection bias-selected) or <u>full universe</u>(including distribution of conditions and z)

Distinguishing a discrete model set straightforward





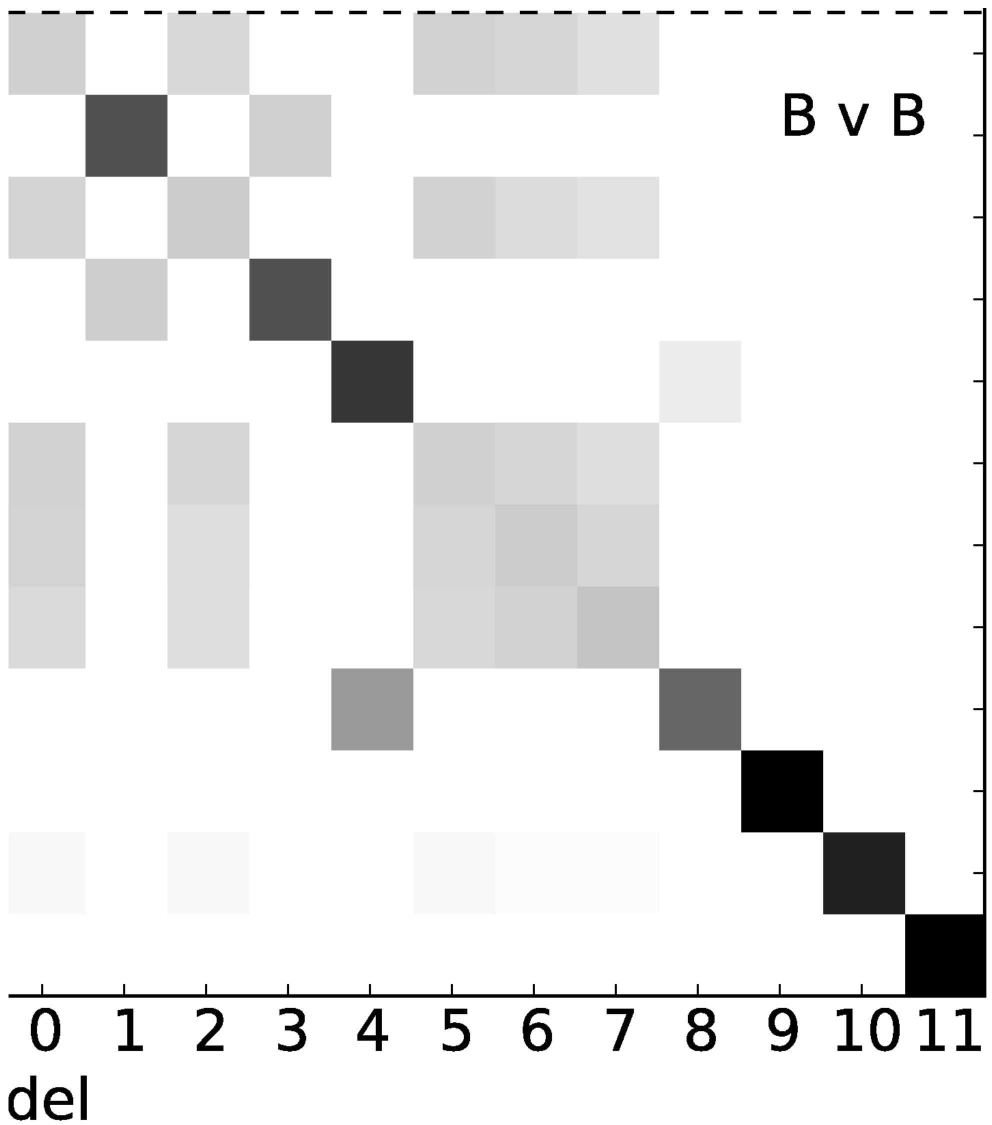
Stevenson, Ohme, Fairhurst (1504.07802), based on Dominik et al 2012
See also Miyamoto et al, GWPAW 2016; Dhani, Mukerjee et al 2016 (LVC meeting)

but this is driven by large rate differences. Rate is highly degenerate with other factors…

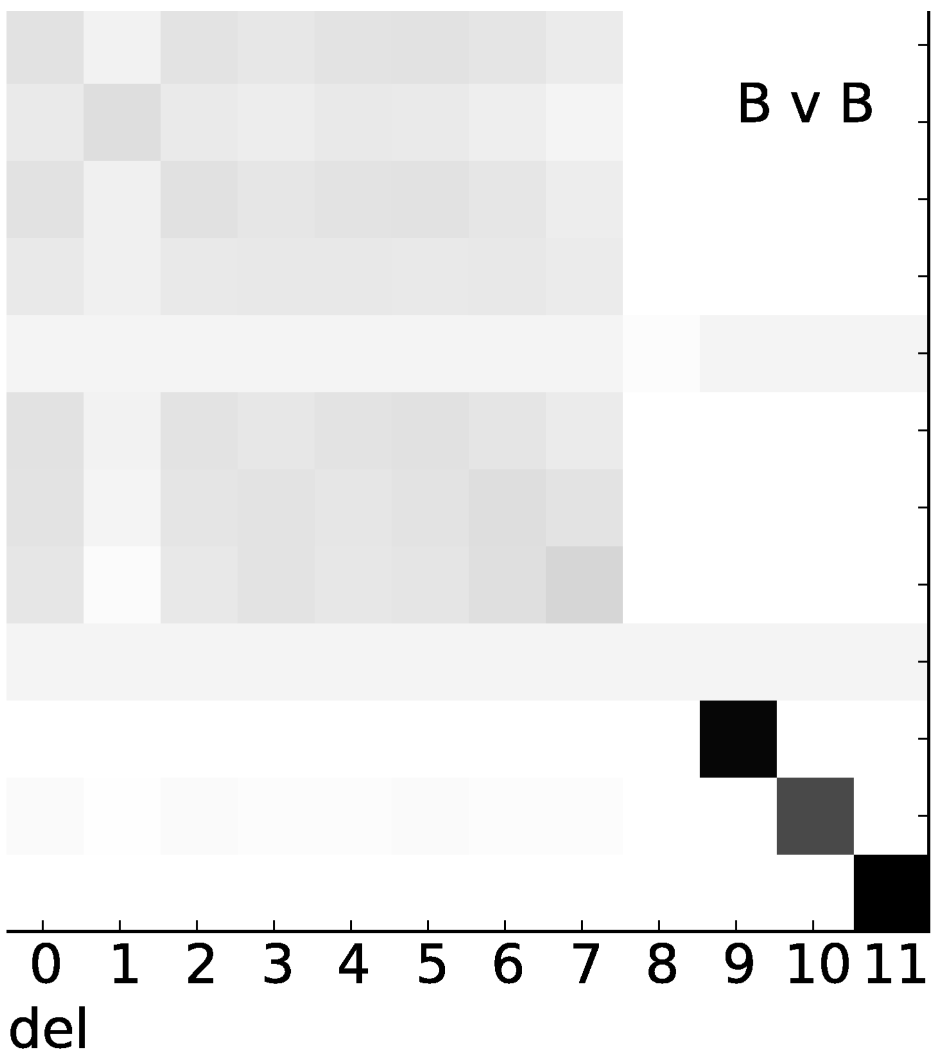
Distinguishing a discrete model set straightforward

- Mass distributions alone are more similar, given measurement error

O2-scale, as before



O2-scale, no rate info



Bayesian Model Selection

- GW PE: (mostly) straightforward application of Bayes' Law — posterior distribution on binary parameters derived from (mostly uninformative, but astrophysically motivated priors) and influenced through the data + waveform model through the likelihood ratio
    - Obtain a set of samples of physical parameters of interest: chirp mass ($\mathcal{M}_c$), mass ratio ($\mathbf{q}$), spin orientations and magnitudes ($\mathbf{s}_1$, $\mathbf{s}_2$), and at some point probably eccentricity (not addressed here)
- Question: ***Given a set of plausible astrophysical formation channels, how do we select a model resembling nature as well as quantify any parameters of that model?***
    - Need to map $\{\mathcal{M}_c, \mathbf{q}, \mathbf{s}_1, \mathbf{s}_2\}$ to mass/spin spectrums, progenitor metallicity, SN kick prescriptions, evolutionary pathways, etc…

Bayesian Hierarchical Modeling

- Foreman-Mackey, et al. 2014 lays out the foundation
- convert p(mod|obs) → p(mod|PE)

$$p(\{h_i\}|\beta) = \prod_i p(h_i) \int \frac{p(\theta|h_i)p(\theta|\beta)}{p(\theta)} d\theta$$

- Integral over model parameters (β) can be evaluated via **importance sampling** using parameter estimation (θ_k) samples

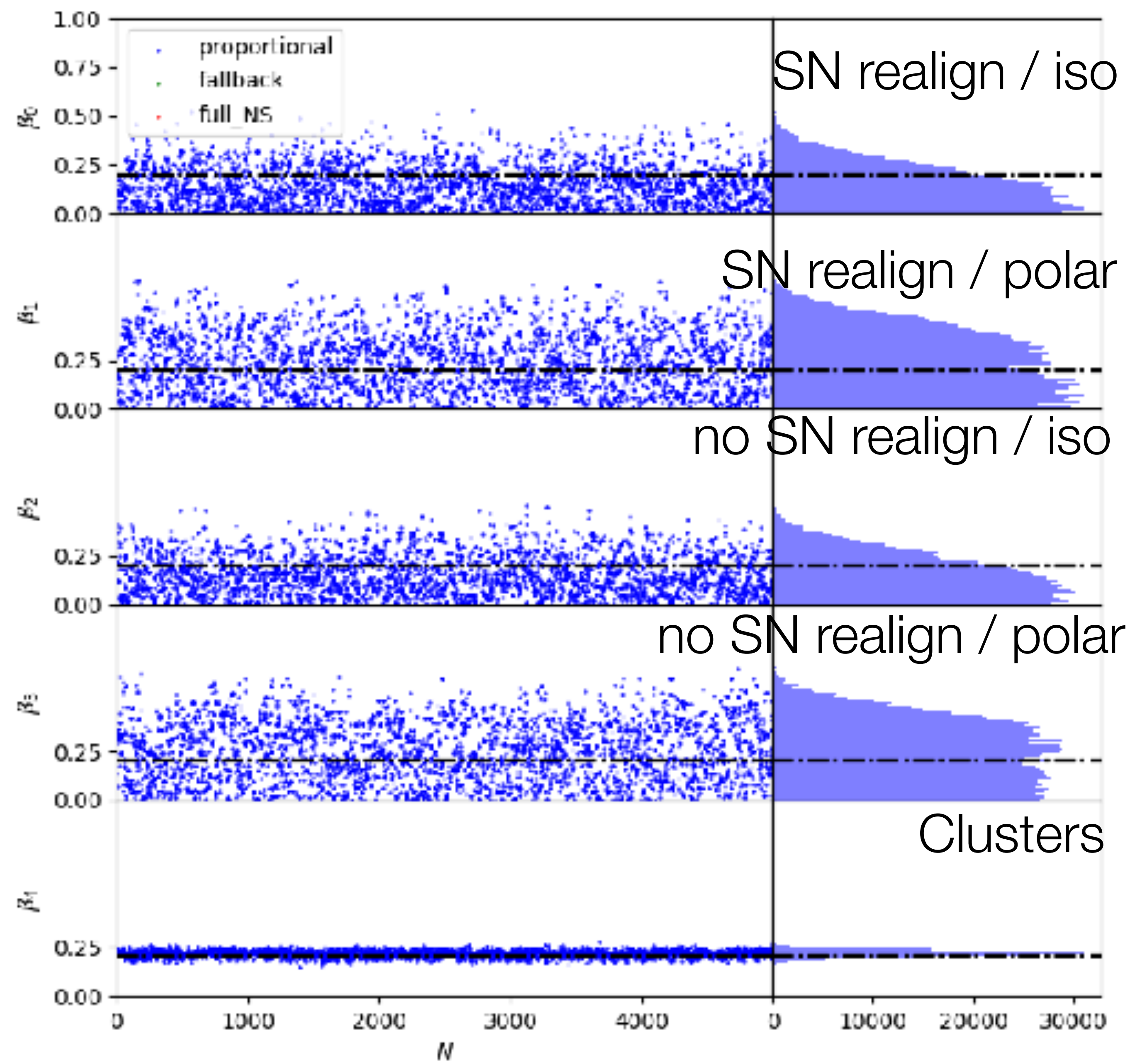
$$\rightarrow p(\beta|\{h_i\}) \propto \prod_i \frac{1}{N} \sum_k \frac{p(\theta_k|\beta)}{p(\theta_k)} p(\beta)$$

- Recasts the problem as a "higher level" parameterization with **no dependence** on original data $\{h_i\}$

Beyond Two Parameter Models

- Are kick *direction* prescriptions (**isotropic / polar**) measurable at the level of mass spectrums?
    - **Spoilers**: *No*. Most mass spectrums are degenerate, and spins (Stevenson, et al. 2017, Rodriguez, et al. 2016) are required

N_{obs} = 10000



Stevenson

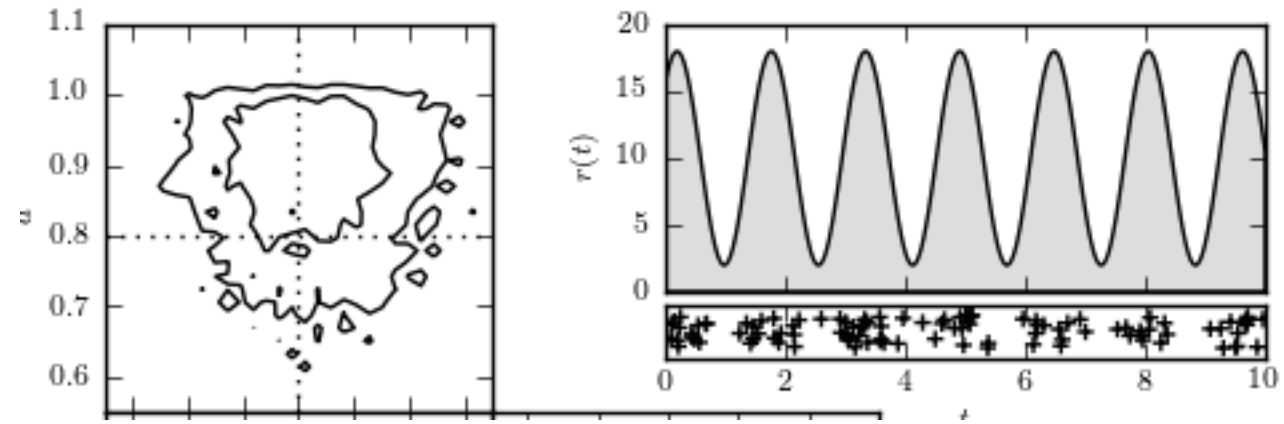
Richard

- Slides from KITP talk, 2016

Familiar statistical challenge

- Inference via Poisson likelihood + bayes

$$L(\Lambda) = e^{-\mu} \frac{\mu^n}{n!} \prod_k \int d\lambda_k p(d_k|\lambda_k) p(\lambda_k|\Lambda)$$



Ivezic et al, *Statistics, data mining, and machine learning in astronomy*
Gregory and Loredo (discrete photon light curves)

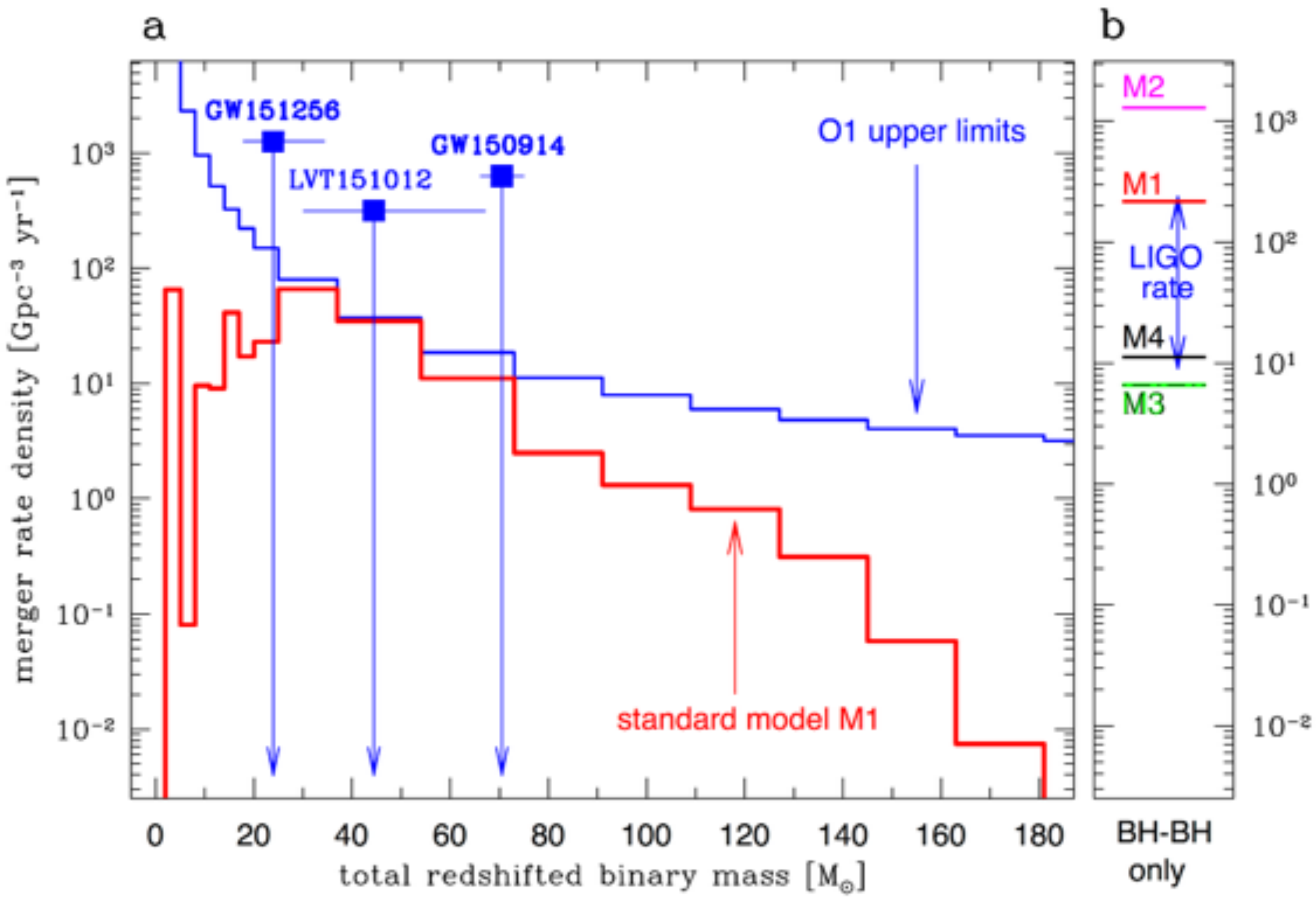
ROS PRD 2013
Hogg and Bovy
W. Farr, LIGO LIGO-T1600562; Mandel, Farr, Gair LIGO-P1600187
ROS LIGO T1600208

  - Same likelihood for nonparametric, parametric, and physical models
  - μ expected n (selection bias)
  - $p(d_k|\lambda_k)$ measurements and error
  - $p(\lambda_k|\Lambda)$ binary parameter distribution, given model parameters

- Informal approaches: weighted histograms (=gaussian mixture models)

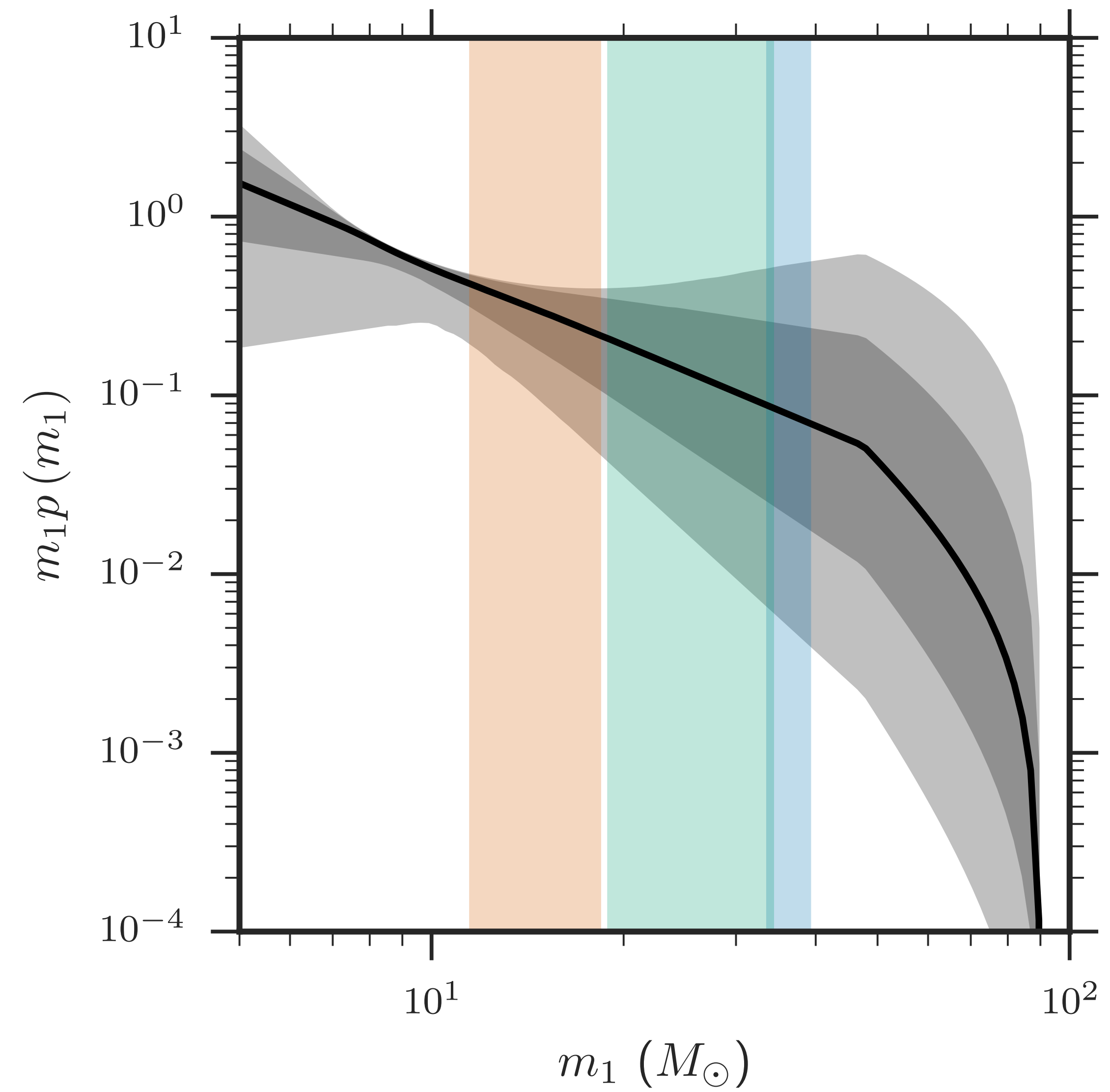
Confronting theory with observations

(Intrinsic distribution)



Belczynski et al *Nature* 2016

(Intrinsic distribution)



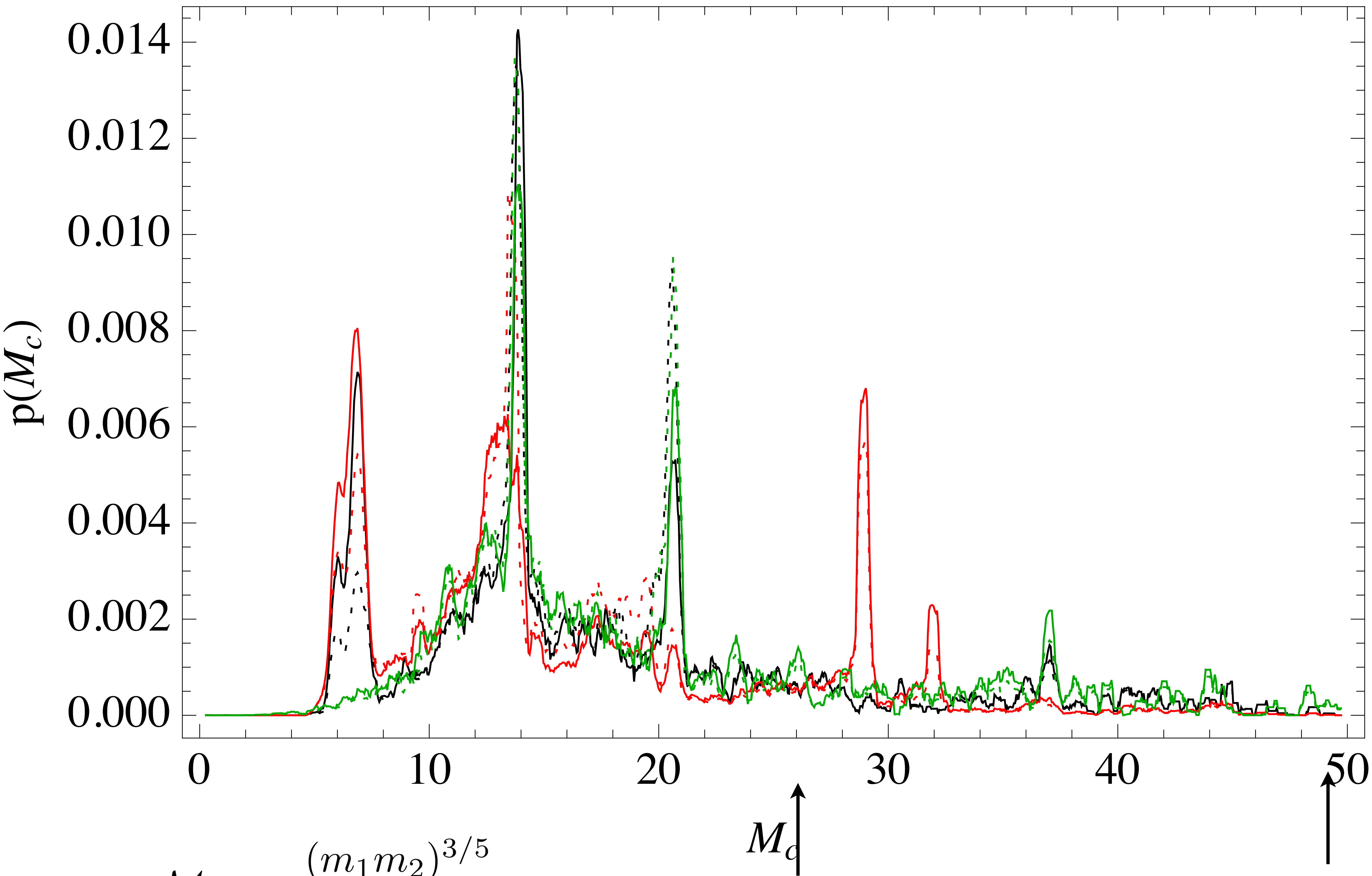
Abbott et al O1 BBH (1606.04856)

A function has infinitely many degrees of freedom

Distributions vary significantly…

(**Detected** distribution)

Dominik et al (2015: 1405.7016)



$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

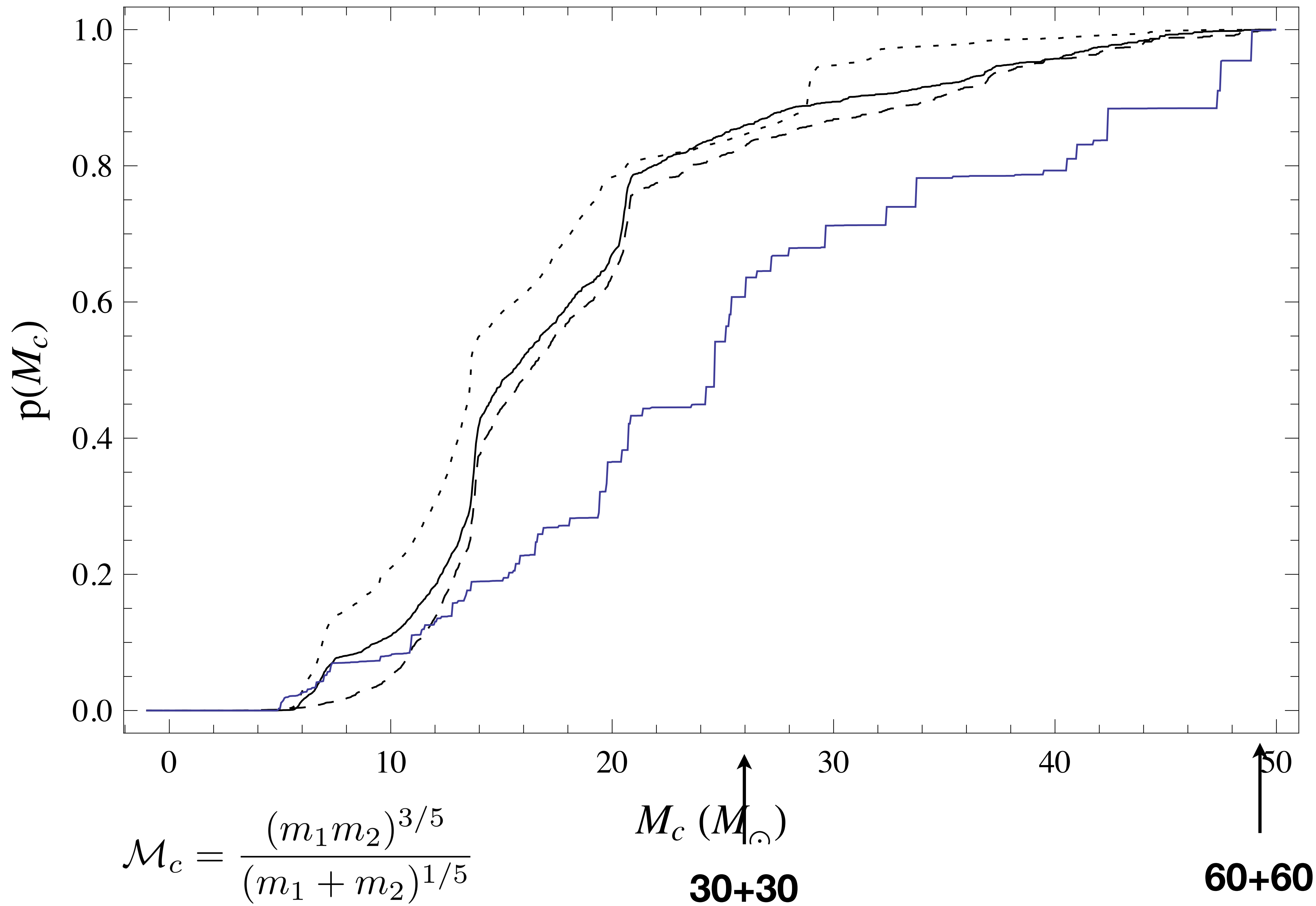
30+30

60+60

Distributions vary significantly…

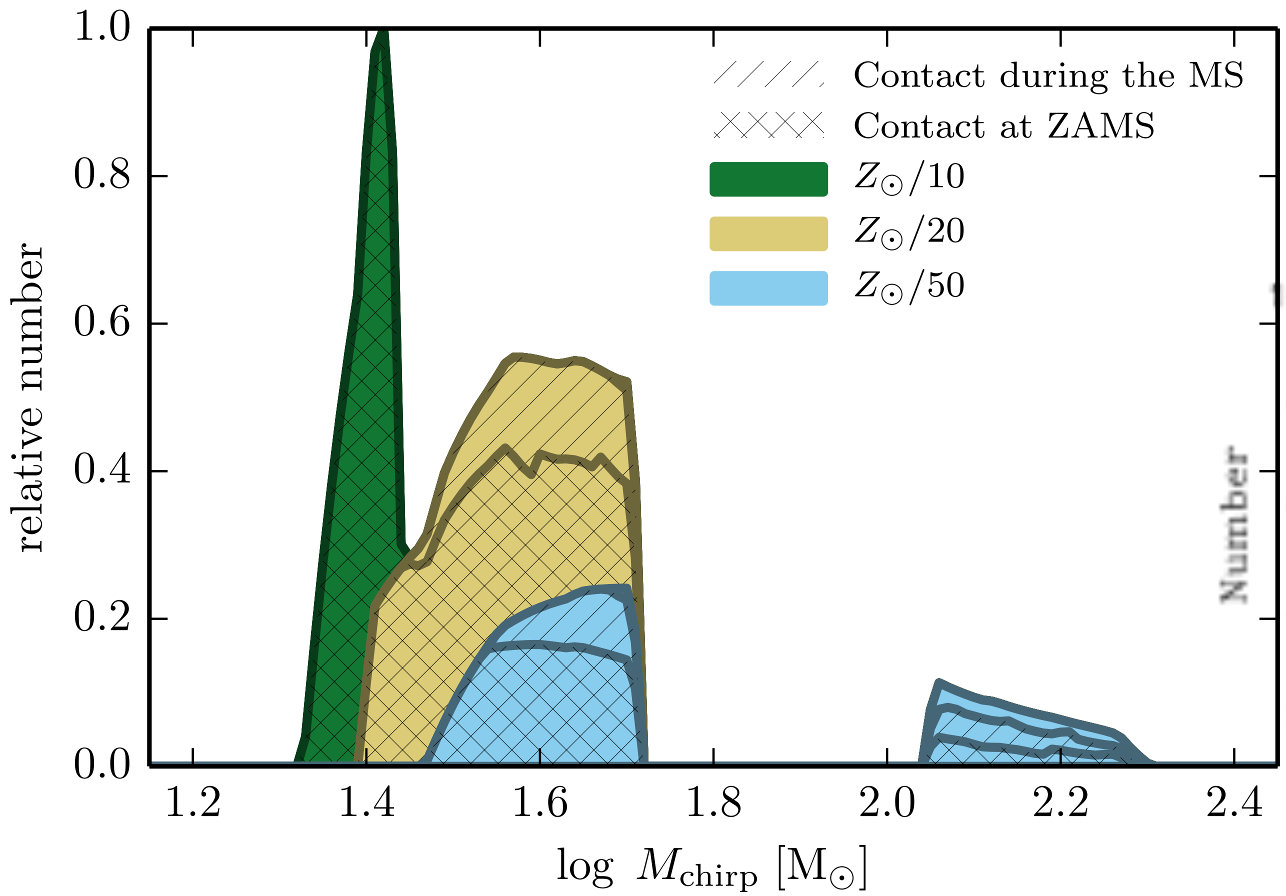
(**Detected** distribution)

Dominik et al (2015: 1405.7016)



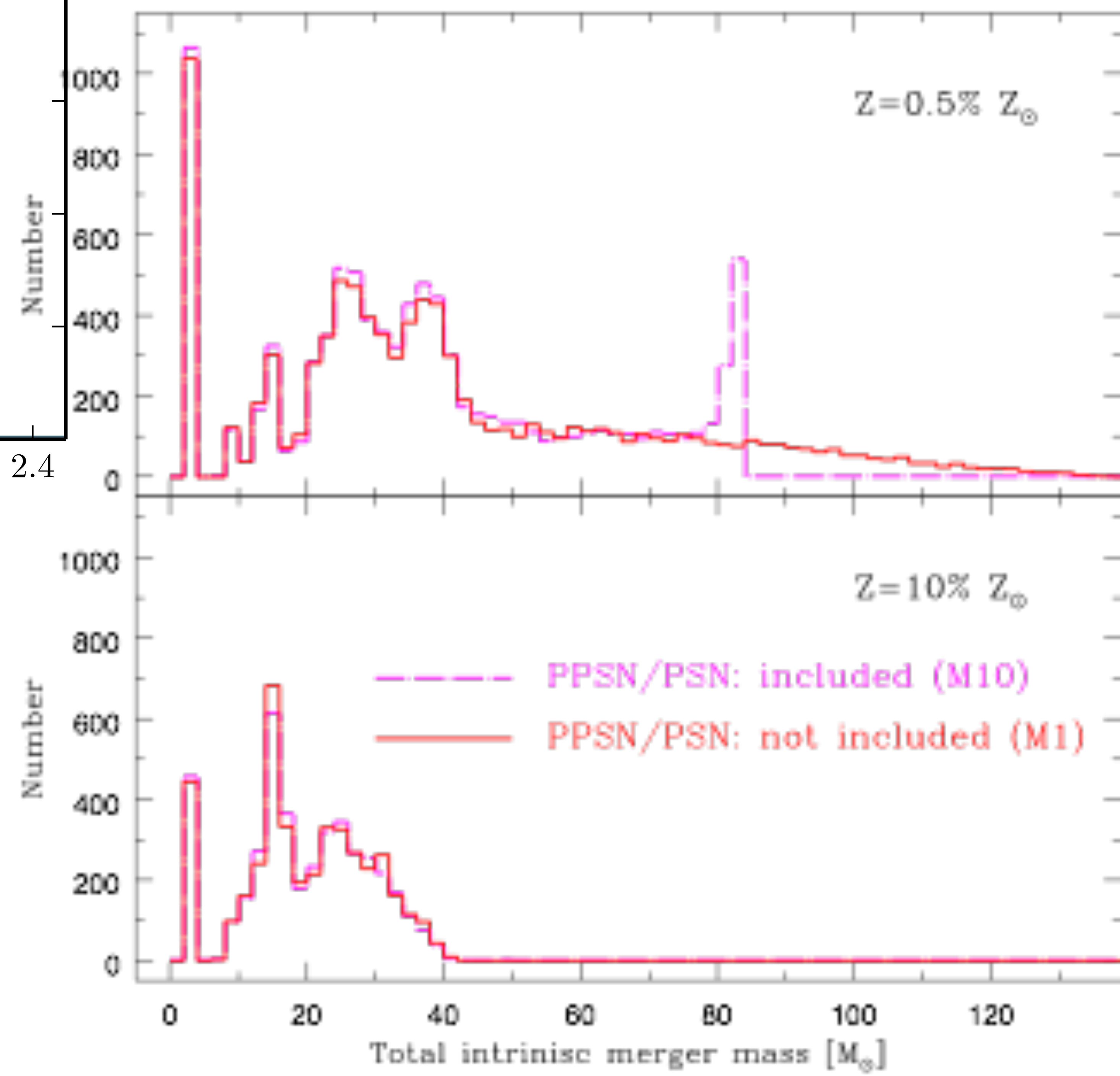
$$\mathcal{M}_c = \frac{(m_1 m_2)^{3/5}}{(m_1 + m_2)^{1/5}}$$

…and for physical reasons, like pair instability



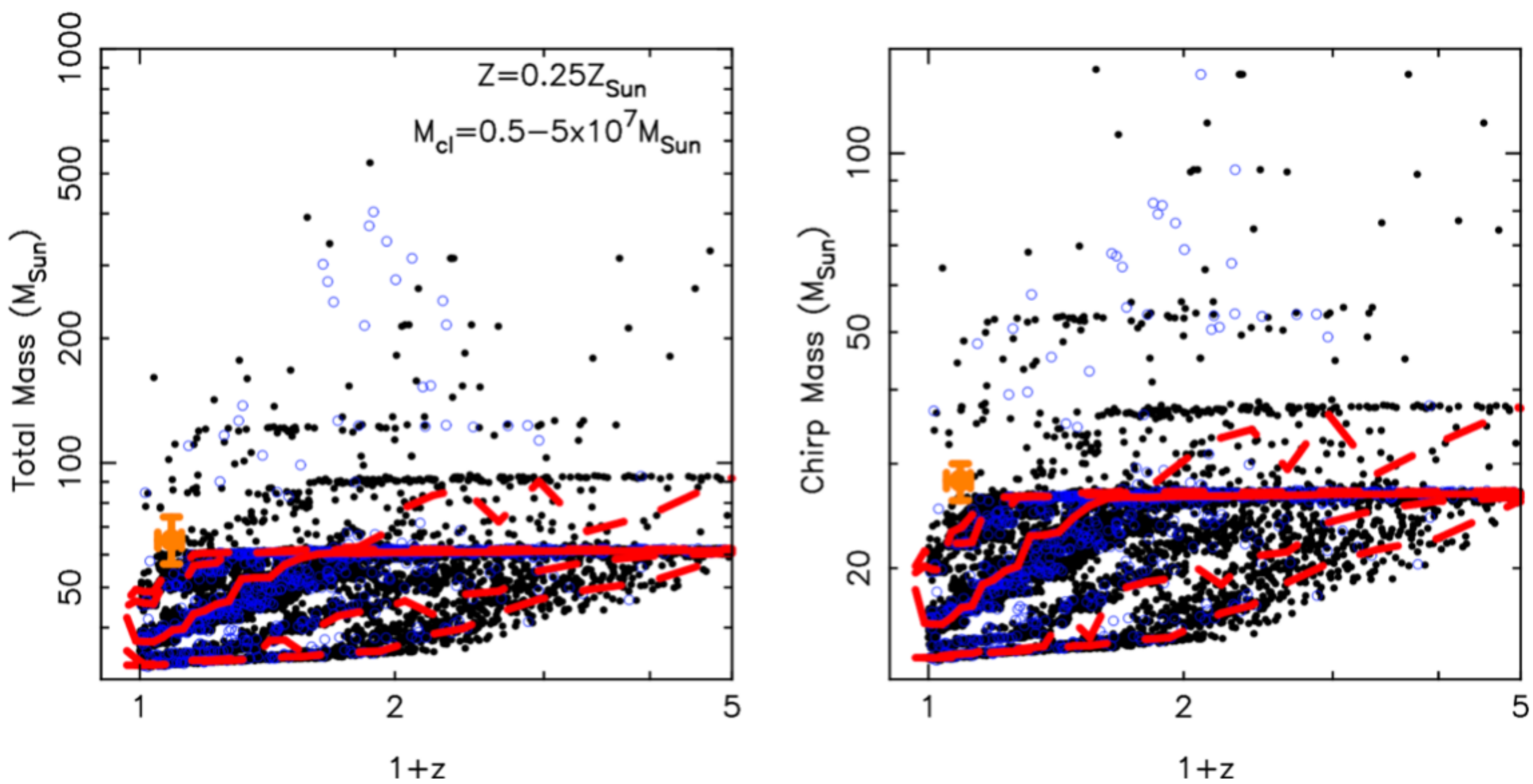
Marchant et al A&A 2016 (1601.03718)

(Intrinsic distribution)



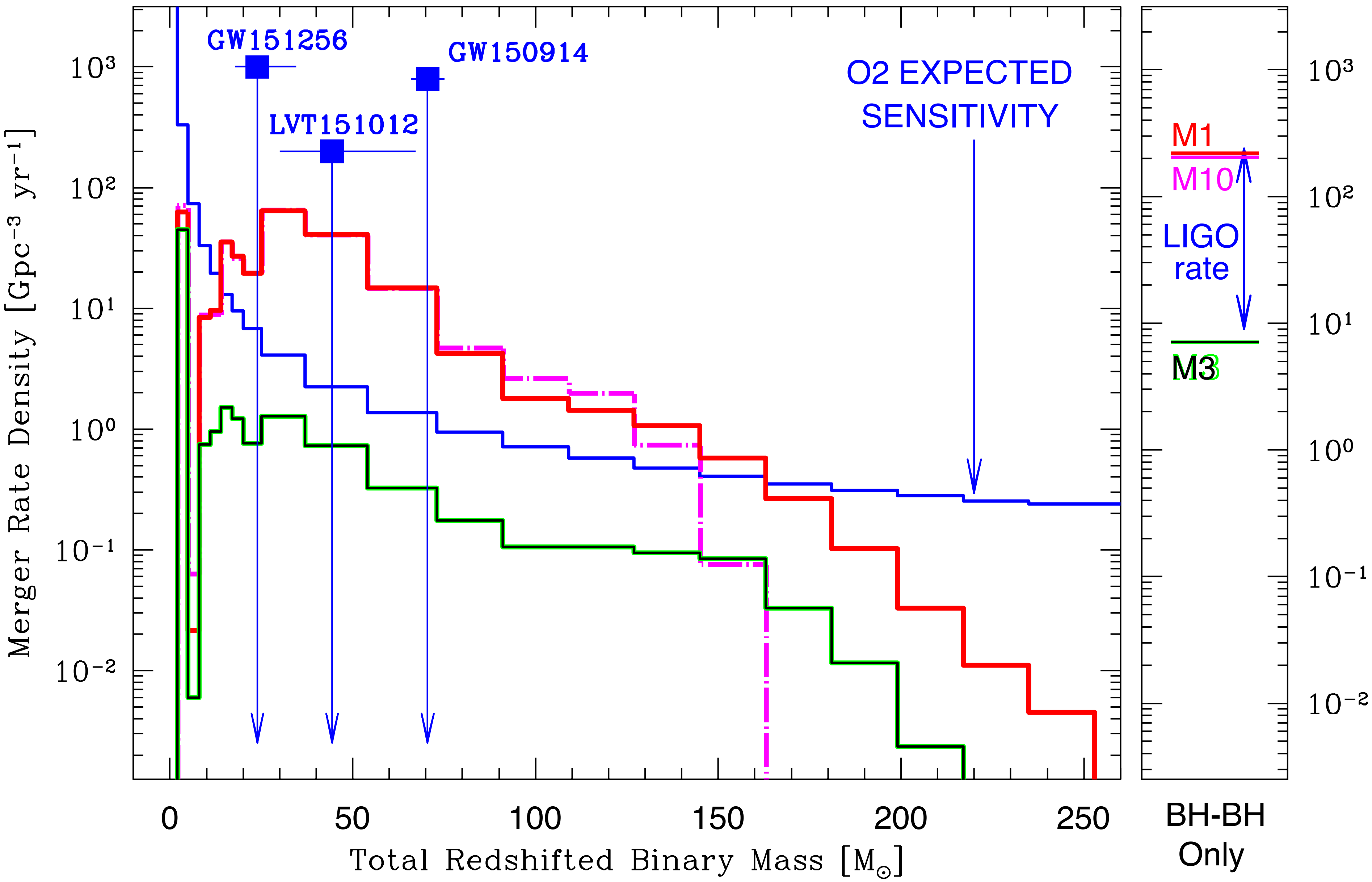
Belczynski et al 1607.03116

…or multiple mergers and single star evolution



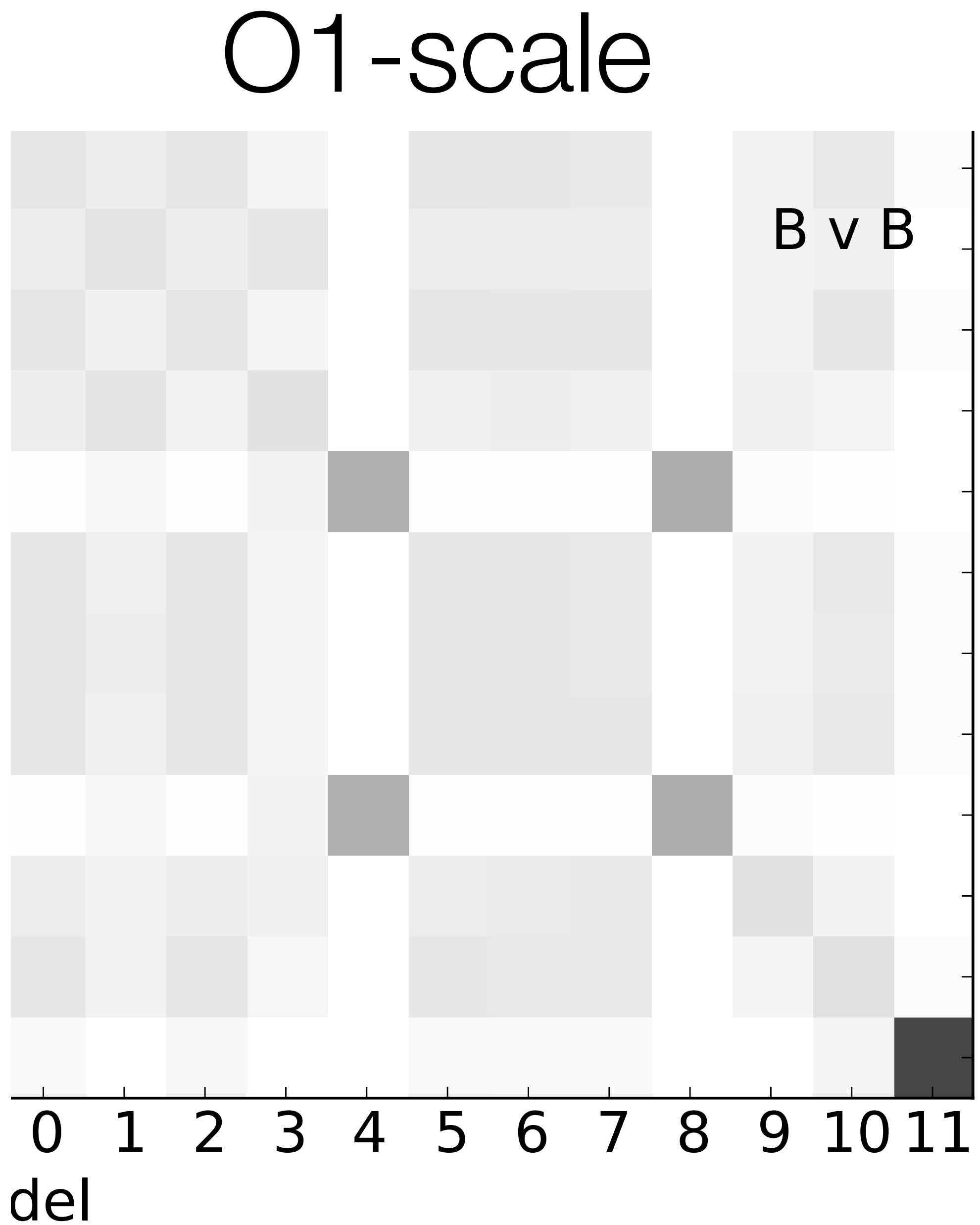
Antonini and Rasio 2016
[see Carl Rodriguez talk]

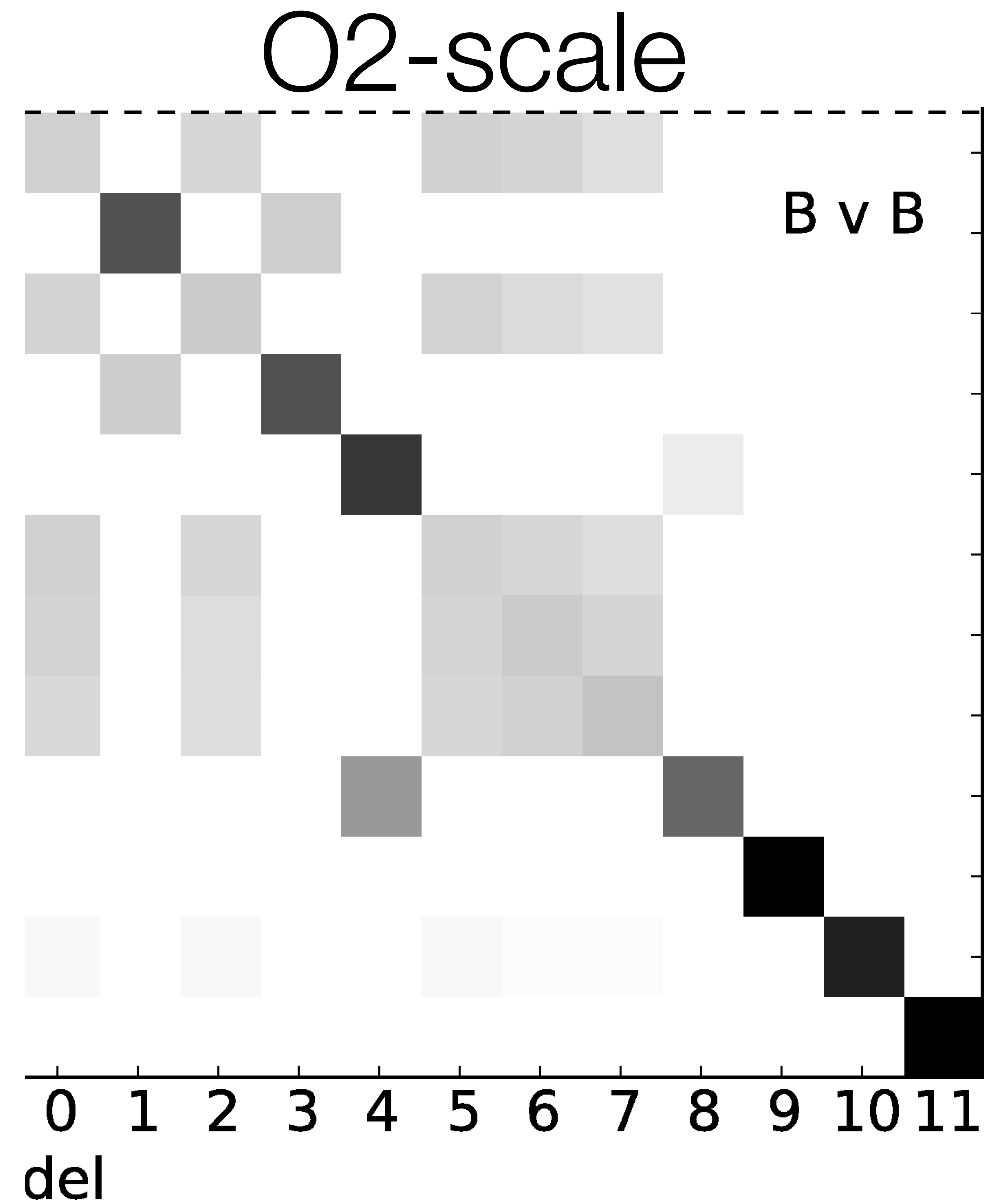
…that may be observationally accessible soon



Belczynski et al 1607.03116

Distinguishing a discrete model set straightforward





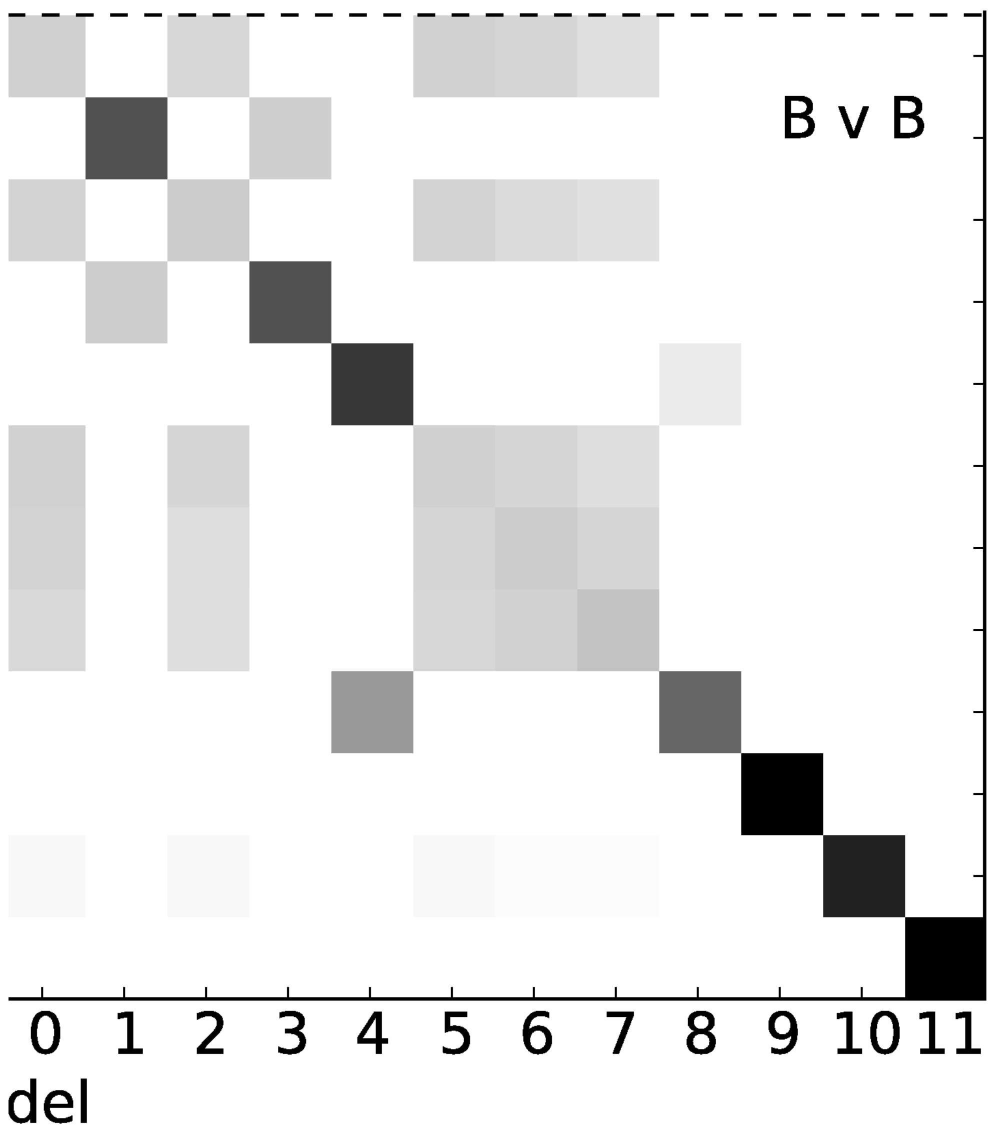
Stevenson, Ohme, Fairhurst (1504.07802), based on Dominik et al 2012
See also Miyamoto et al, GWPAW 2016; Dhani, Mukerjee et al 2016 (LVC meeting)

but this is driven by large rate differences. Rate is highly degenerate with other factors…

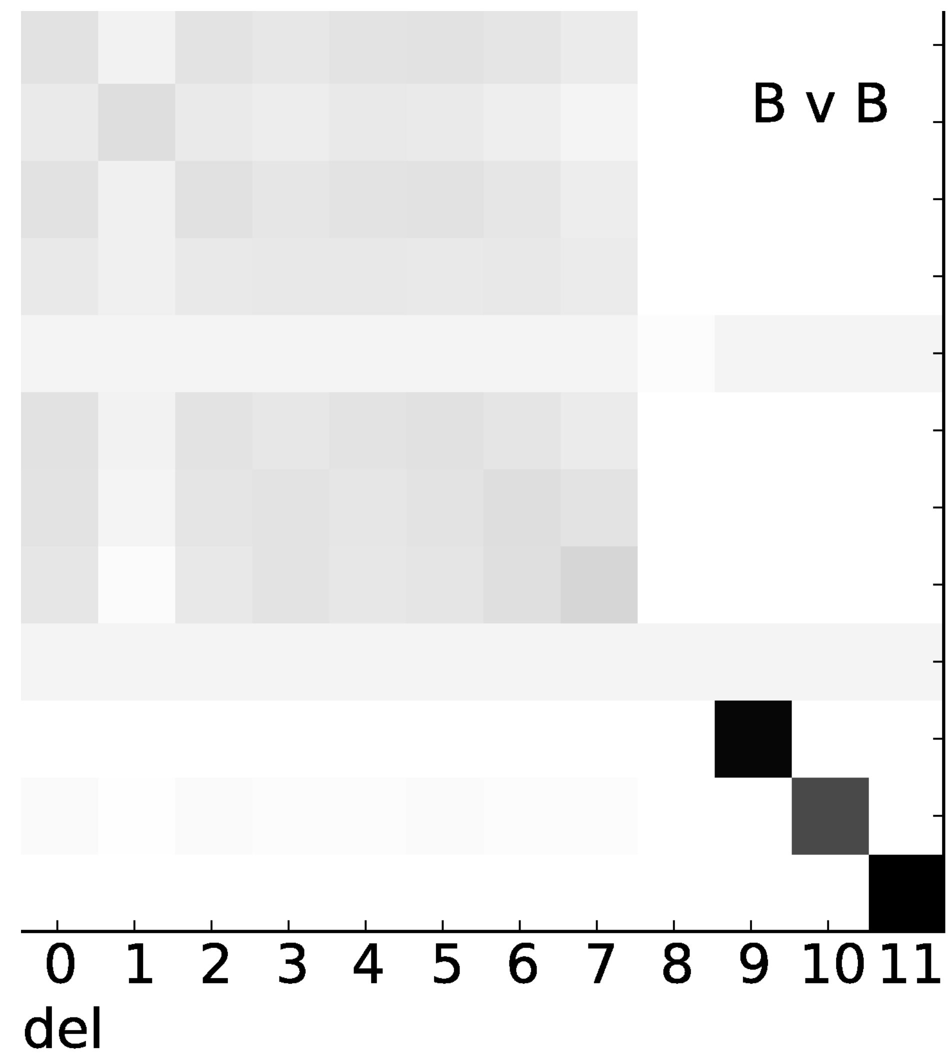
Distinguishing a discrete model set straightforward

- Mass distributions alone are more similar, given measurement error

O2-scale, as before

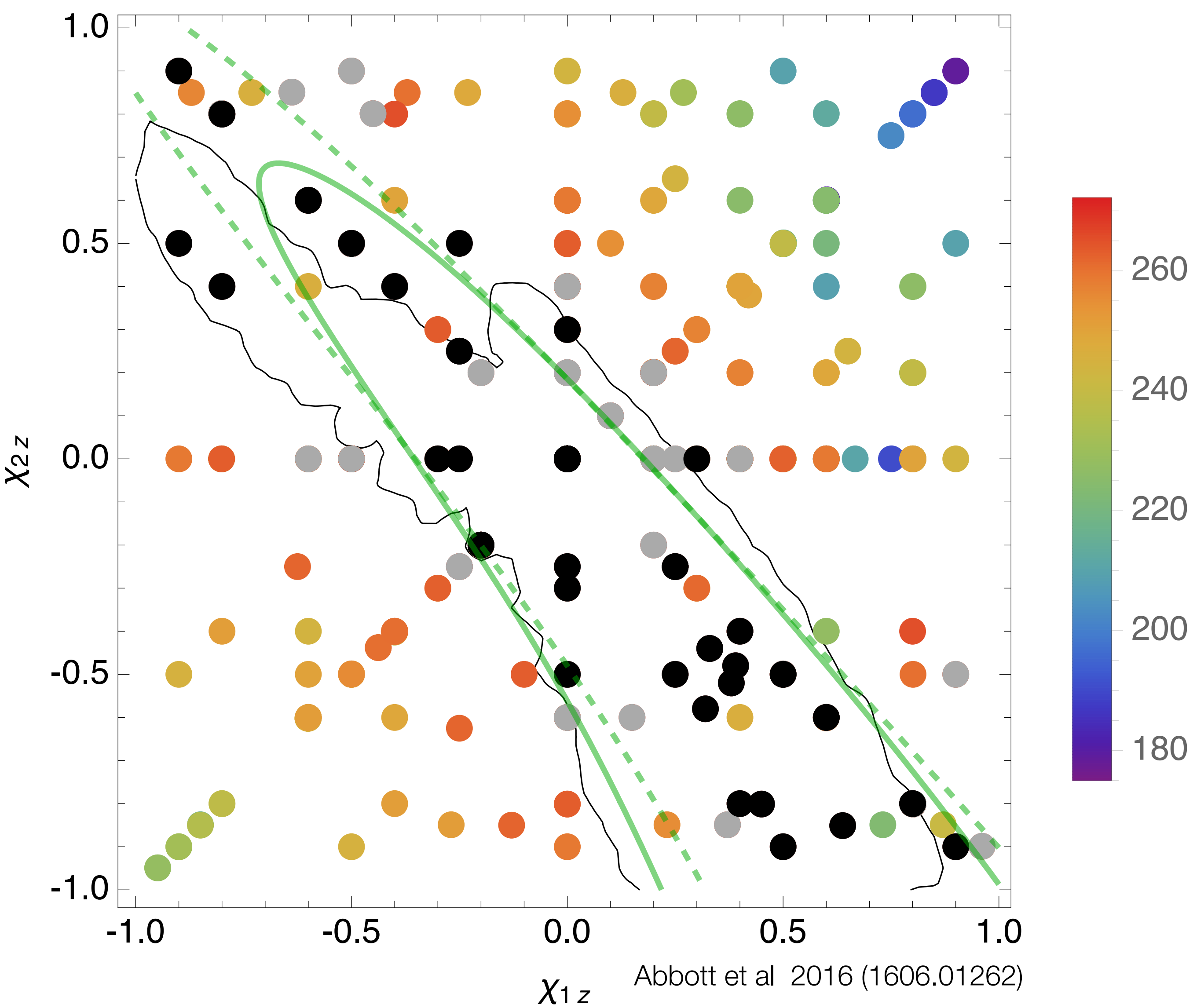


O2-scale, no rate info

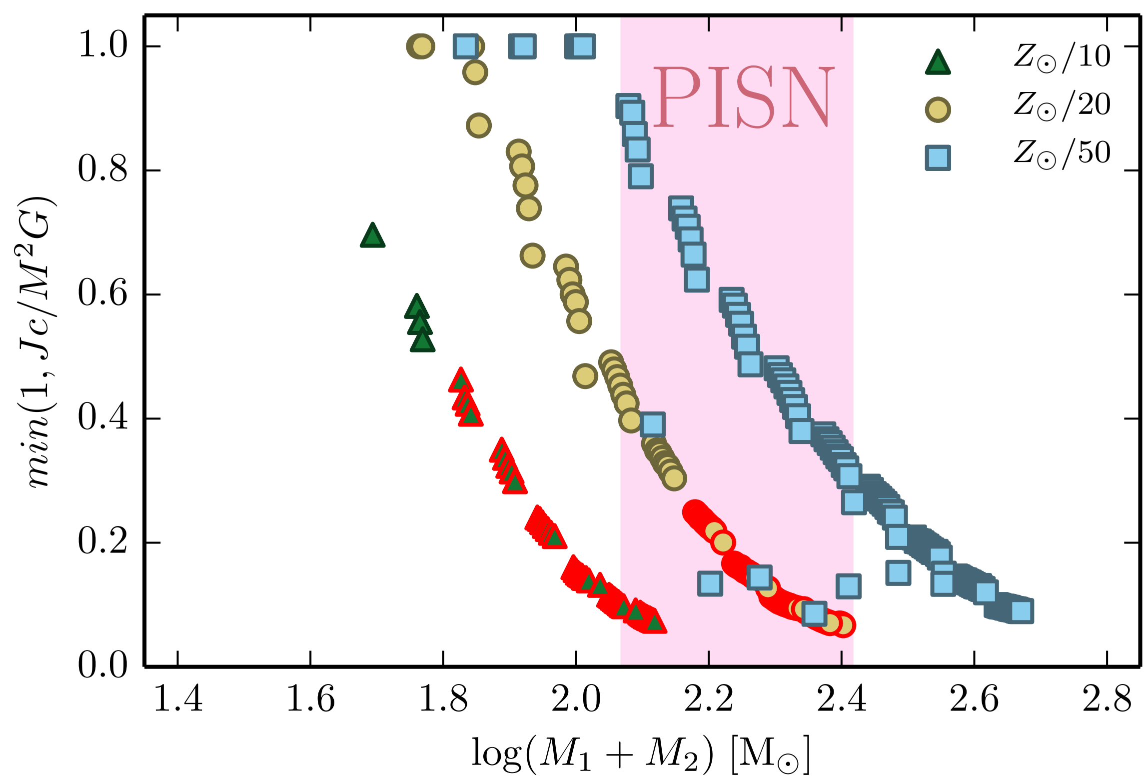


Beyond the mass distribution: Power of spin

- High mass binaries may be strictly and positively aligned (fallback)
- Low spins required for GW150914…possible? [Kushnir et al]
    - Tells us something about how massive stars evolve? About tides?
    - Or favors dynamics?



Abbott et al 2016 (1606.01262)



Marchant et al A&A 2016 (1601.03718)