ET-0211A-22
Contact: q.hu.2@research.gla.ac.uk

Systematic error accumulation in testing GR with ET

Qian Hu and John Veitch @ University of Glasgow

ET-osb-da Telecon

10 Oct 2022

Overview

- Predicting systematic errors in PE: overlap signals and inaccurate waveforms
- Mock catalog for ET
- Systematic error accumulation – using parameterized post-Newtonian (PPN) coefficients tests of GR as an example

Excess strains

Where systematic errors originate (arXiv:0707.2982, 2104.01897)

- Likelihood in PE: assuming noise is stationary and Gaussian, and d-h=n

$$\ln L(\vec{\theta}) = -\frac{1}{2}(d-h|d-h) = -\frac{1}{2}(n|n),$$

- Maximum likelihood (ML) estimator:

$$\partial_i \ln L \mid_{\vec{\theta}=\vec{\theta}_{\rm ML}} = (\partial_i h|d-h) \mid_{\vec{\theta}=\vec{\theta}_{\rm ML}} = 0,$$

- The statistical error caused by noise:

$$\Delta\theta^i_{\rm stat} \approx (\Gamma^{-1})^{ij}(\partial_j h_{\rm m}|n)$$

$$< \Delta\theta^i_{\rm stat} >= 0$$ Unbiased

$$< \Delta\theta^i_{\rm stat}\Delta\theta^j_{\rm stat} >= (\Gamma^{-1})^{ij}$$ Fisher matrix forecast

- If there are strains other than noise (d-h=n+δH), the ML will be biased:

$$\Delta\theta^i_{\rm sys} = (\Gamma^{-1})^{ij}(\partial_j h_{\rm m}|\delta H)$$

Our concerns

$$\Delta\theta^i_{\mathrm{stat}} \approx (\Gamma^{-1})^{ij}(\partial_j h_{\mathrm{m}}|n) \qquad \Delta\theta^i_{\mathrm{sys}} = (\Gamma^{-1})^{ij}(\partial_j h_{\mathrm{m}}|\delta H)$$

- Waveform systematics: $\delta H_{waveform} = h_{real} - h_{model}$
- Overlap signals:
  - Detected overlaps: $\delta H_{do} = h'_{real}(\theta_{true}) - h'_{model}(\theta_{ML}) \approx \Delta\theta'^{i}_{stat}\partial_i h'_m + \delta H'_{waveform}$
  - Undetected overlaps (SNR<8): $\delta H_{\mathrm{uo}} = h''_{real}(\theta_{true})$
- $\Delta\theta^i_{stat} \propto 1/\mathrm{SNR}$, $\Delta\theta^i_{sys}$ doesn't . When SNR increases, systematic errors may dominate
- Different types of systematic errors may have impacts on each other

PPN, waveforms, and parameters

- Parametrized post-Newtionial (PPN) coefficient tests of GR:
  - IMRPhenomPv2 phase is characterized by a set of parameters $\{p_i\}$
    - Inspiral regime parameters: $\{\varphi_0, \dots, \varphi_7\}, \{\varphi_{5l}, \varphi_{6l}\}$
    - Phenomenological coefficients: $\{\beta_0, \dots, \beta_3\}$
    - Merger-ringdown parameters: $\{\alpha_0, \dots, \alpha_5\}$
  - $p_i \to (1 + \delta\hat{p}_i)p_i$, the $\delta\hat{p}_i$ is the testing parameter. $\delta\hat{p}_i = 0$ returns to GR

$$\varphi_{\mathrm{PN}}(f) = 2\pi f t_c - \varphi_c - \frac{\pi}{4} + \frac{3}{128\eta}\left(\pi\tilde{f}\right)^{-5/3}\sum_{i=0}^{7}\left[\varphi_i + \varphi_{il}\log(\pi\tilde{f})\right]\left(\pi\tilde{f}\right)^{i/3}.$$

- We choose $\delta\widehat{\varphi_0}$ as an example testing parameter in this work
- We perturbate $\delta\widehat{\beta_2}$ as the inaccurate waveform parameter:
  - $\delta\widehat{\beta_2} = 0$: model waveform
  - $\delta\widehat{\beta_2} = 5\times10^{-2}$: "real" waveform for current waveform modelling, mismatch ~ $10^{-4} - 10^{-3}$
  - $\delta\widehat{\beta_2} = 5\times10^{-4}$: "real" waveform for future waveform modelling, mismatch ~ $10^{-7} - 10^{-6}$
- Parameters in GR: chirp mass, mass ratio, merger time. Other parameters are set to be perfectly known

Catalog simulation

- We use an analytical approximation for merger rate density (Oguri 2018)

$$\frac{R_{\rm GW}(z)}{\rm Gpc^{-3}yr^{-1}} = \frac{a_1 e^{a_2 z}}{e^{a_3 z} + a_4}, \qquad R_{\rm GW}^{\rm obs}(z) = R_{\rm GW}(z)\frac{dV_c}{dz}(z).$$

- According to local merger rate estimation, we scale a_1 to generate different merger rates: low, median, high
- BBH
  - mass: power law + peak model in GWTC-2 population inference
  - zero spin
- BNS
  - All BNS have the same intrinsic parameters
  - 1.45+1.4 solar mass, zero spin, tidal deformability parameter=425
- Isotropic inclination and sky direction
- We will test GR with all BBH events. BNS only appears as overlap background

Catalog simulation

BBH and BNS: summary

(per year)

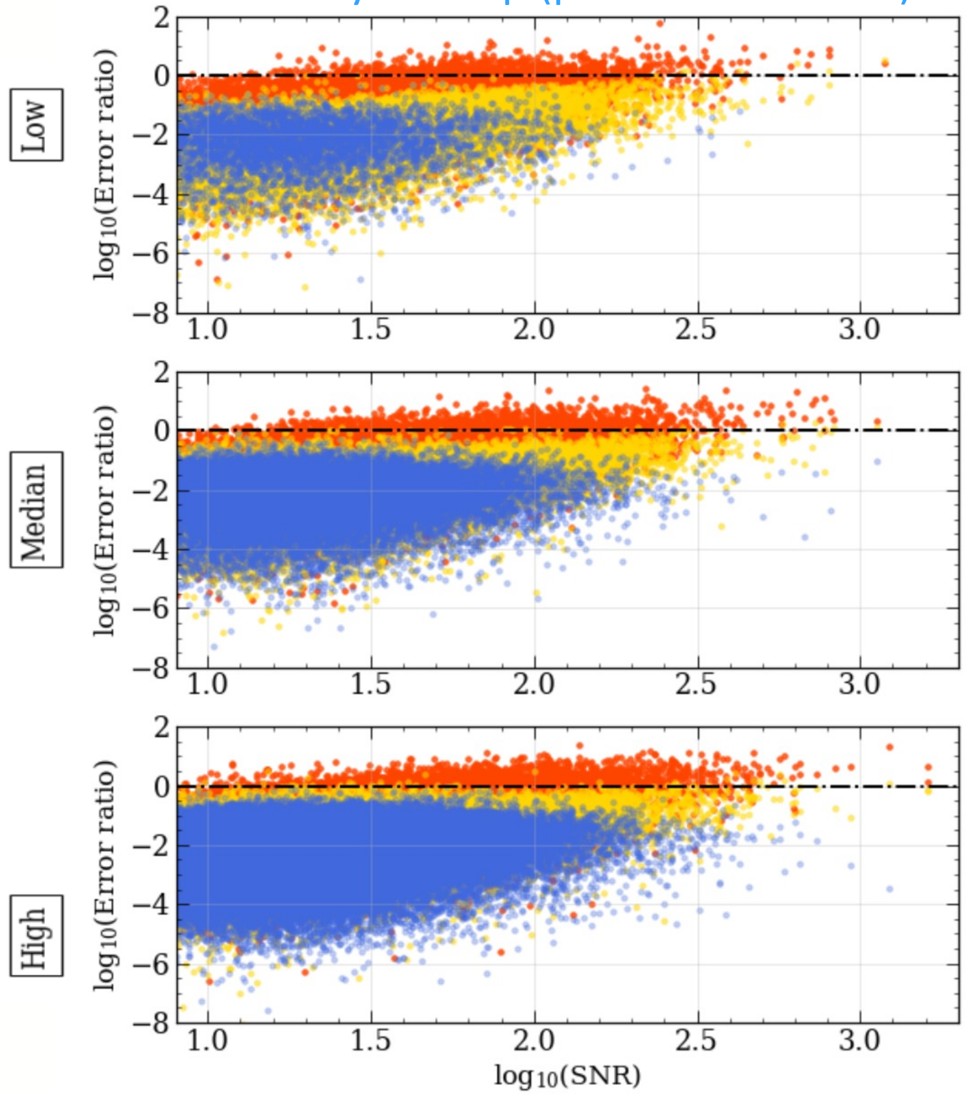
Overlap:= $|\Delta t|<4s$

	# of observable binaries		Detected overlaps on BBH events		Undetected overlaps on BBH events	
	BBH	BNS	# of overlaps	# (fraction) of events	# of overlaps	# (fraction) of events
Low	56526	286088	0	48380 (98%)	0	45991 (93%)
			1	937 (1.9%)	1	3217 (6.5%)
			2	11 (0.022%)	2	119 (0.24%)
					3	1 (0.0020%)
Median	88300 (with ~80000 SNR>8)	1144354	0	73224 (95%)	0	58921 (77%)
			1	3574 (4.6%)	1	15658 (20%)
			2	74 (0.096%)	2	2108 (2.7%)
			3	1 (0.0010%)	3	174 (0.23%)
					4	10 (0.013%)
					5	2 (0.0030%)
High	143349	2896647	0	112745 (90%)	0	63932 (51%)
			1	11496 (9.2%)	1	42931 (34%)
			2	589 (0.47%)	2	14143 (11%)
			3	19 (0.015%)	3	3208 (2.6%)
					4	540 (0.43%)
					5	86 (0.069%)
					6	6 (0.0050%)
					7	3 (0.0020%)

Overview of systematic errors

- Calculate error ratio of each event caused by different waveform models
- Error ratio increases when SNR goes higher
- Generally, systematic error < statistic error, but exceptions exist
- Statistic error decrease when combining multiple events. Could systematic error surpass statistic at some point?

Red: current waveform + overlap
Yellow: future waveform + overlap
Blue: only overlap (perfect waveform)



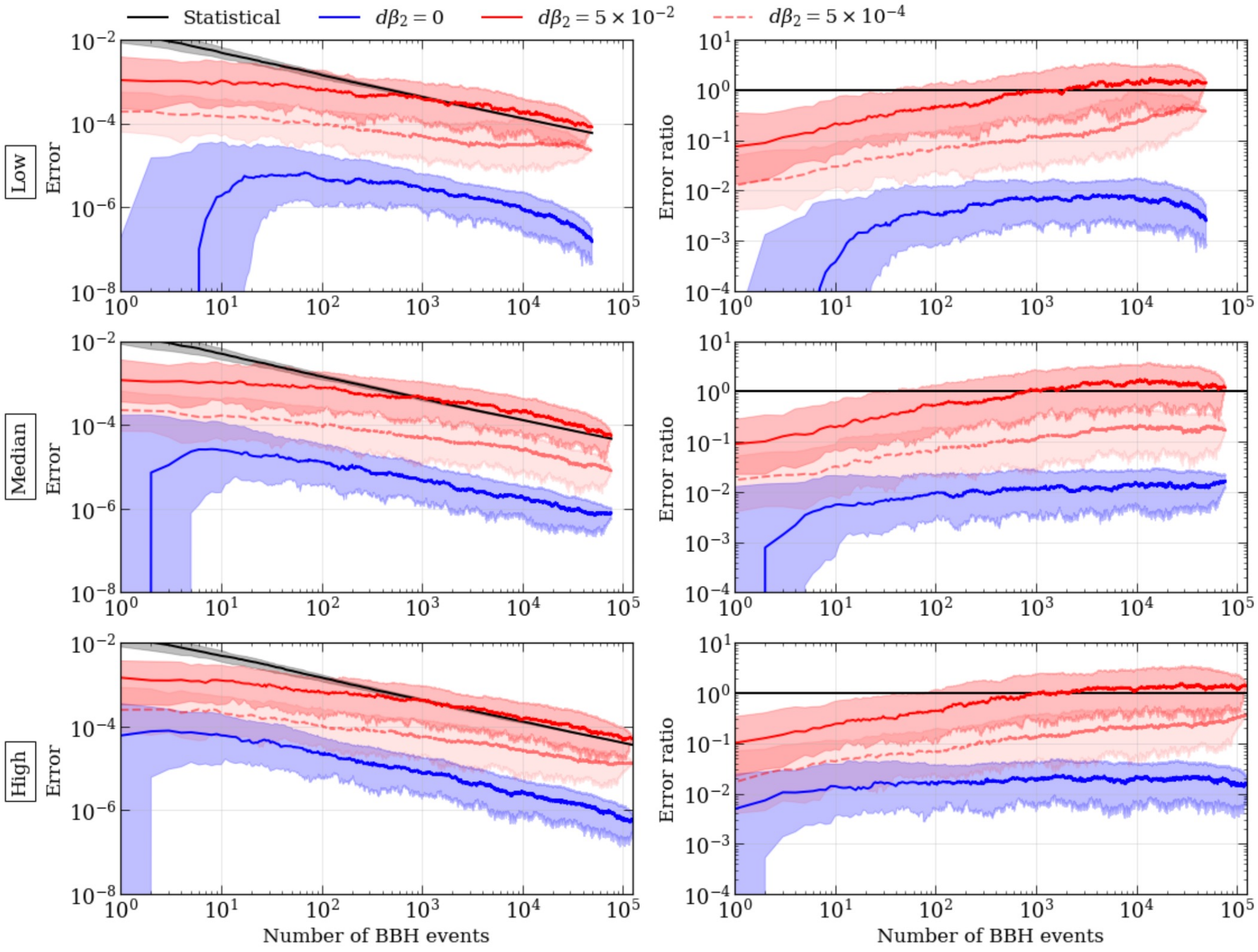
Catalog tests I

Multiplying likelihood (=posterior)

- Posterior of testing parameter is Gaussian
- Multiplying two Gaussian distributions:

$$\sigma_f^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \qquad \mu_f = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

- Statistical error ~ N^(-1/2)
- Systematic error ~N^(-1/2+ ε) , The ε comes from μ≠0 when multiplying posteriors.
- Error ratio may increase to 1 as number of events increases
- Waveform systematic is more problematic than overlap signals



Catalog tests II

Multiplying Bayes factors

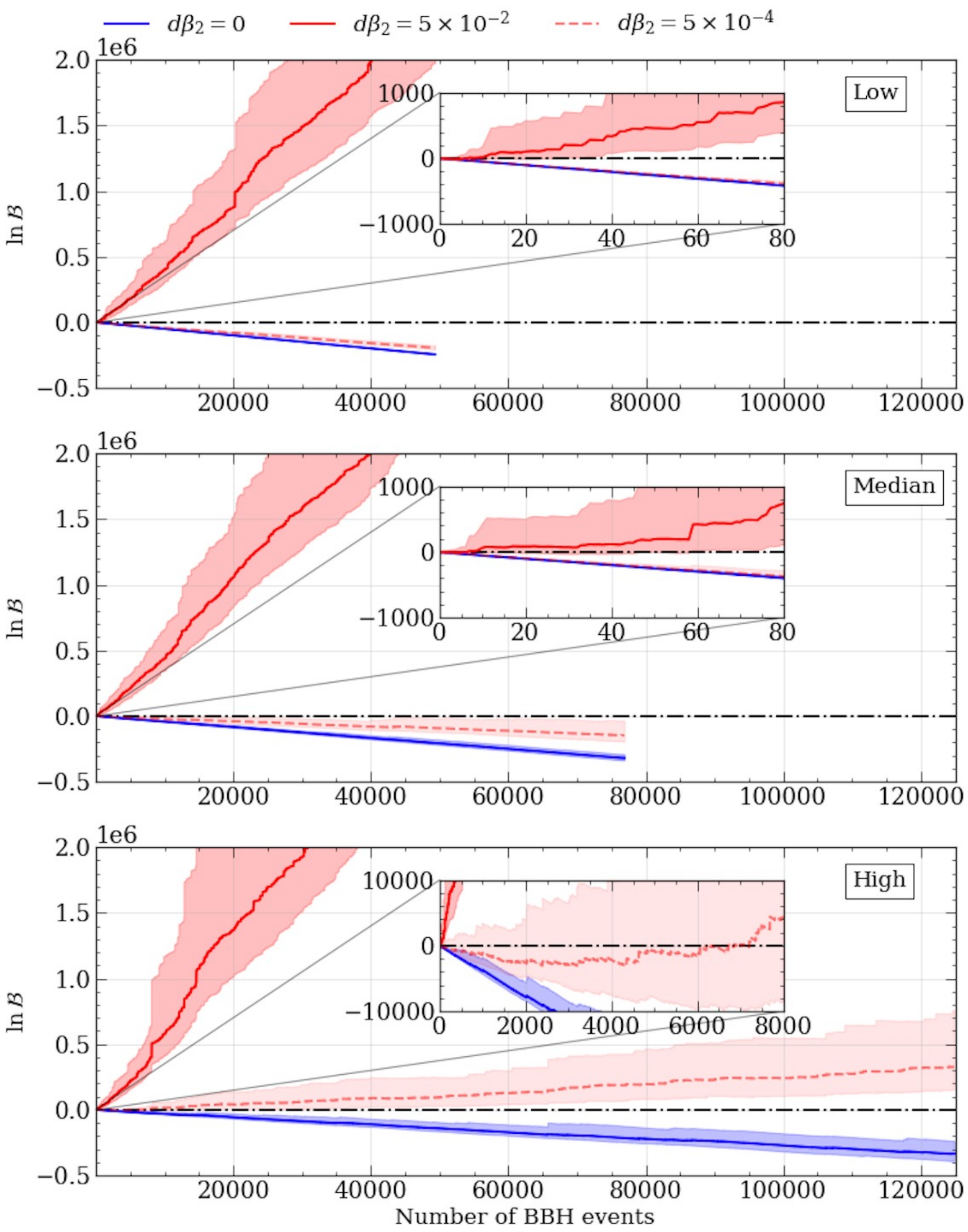
- Gaussian posterior. Calculate Bayes factor and multiply BFs of multiple events

$$Z_{\mathrm{nonGR}} \equiv \int \mathrm{d}\lambda\, \mathcal{L}(\lambda)$$

$$Z_{\mathrm{GR}} \equiv \int \mathrm{d}\theta\, \mathcal{L}_{\mathrm{GR}}(\theta)$$

$$\mathcal{B} \equiv \frac{\Pi}{A} \frac{Z_{\mathrm{nonGR}}}{Z_{\mathrm{GR}}} \rightarrow \log B \sim a\, \mu^2 + b$$

- Much faster towards a false deviation, for systematic error appears as quadratic term
- Waveform systematics would deteriorate with the increase of overlap rates (more inaccurate subtractions)



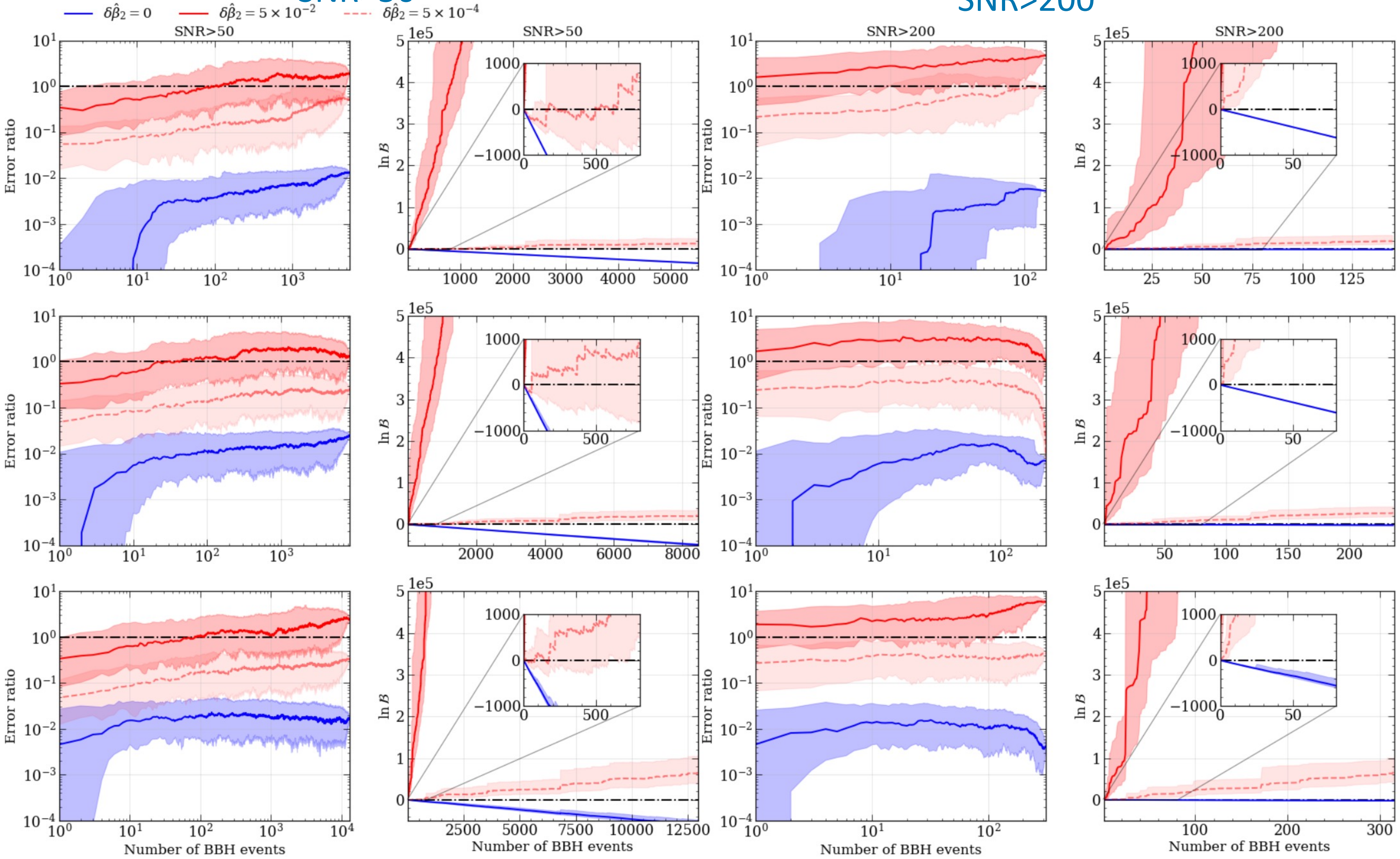
Golden events

Testing GR with selected high SNR and clean events

- SNR>50 or 200 & No detected overlapping signal
- No qualitative difference between different merger rates – as expected
- More vulnerable to systematic errors!

SNR>50

SNR>200



Summary

- Waveform systematic vs (detected/undetected) overlap signals:
  - Detected overlapping signals magnifies waveform systematics errors
  - Overlapping signals do not always exist, it is inaccurate waveform that keeps contributing to systematic error
  - Undetected overlaps are more frequent than the detected! Confusion background
- Catalog tests of GR:
  - Systematic errors do accumulate when we combine results from multiple events
  - Multiplying posteriors and BFs could both lead to false deviations. The latter is more sensitive
  - Golden events are more vulnerable to false deviations: systematic error dominates at high SNR scenario
- The idea of error accumulation is universal: testing GR, cosmology, population inference, EoS of neutron stars, …

Contact: q.hu.2@research.gla.ac.uk