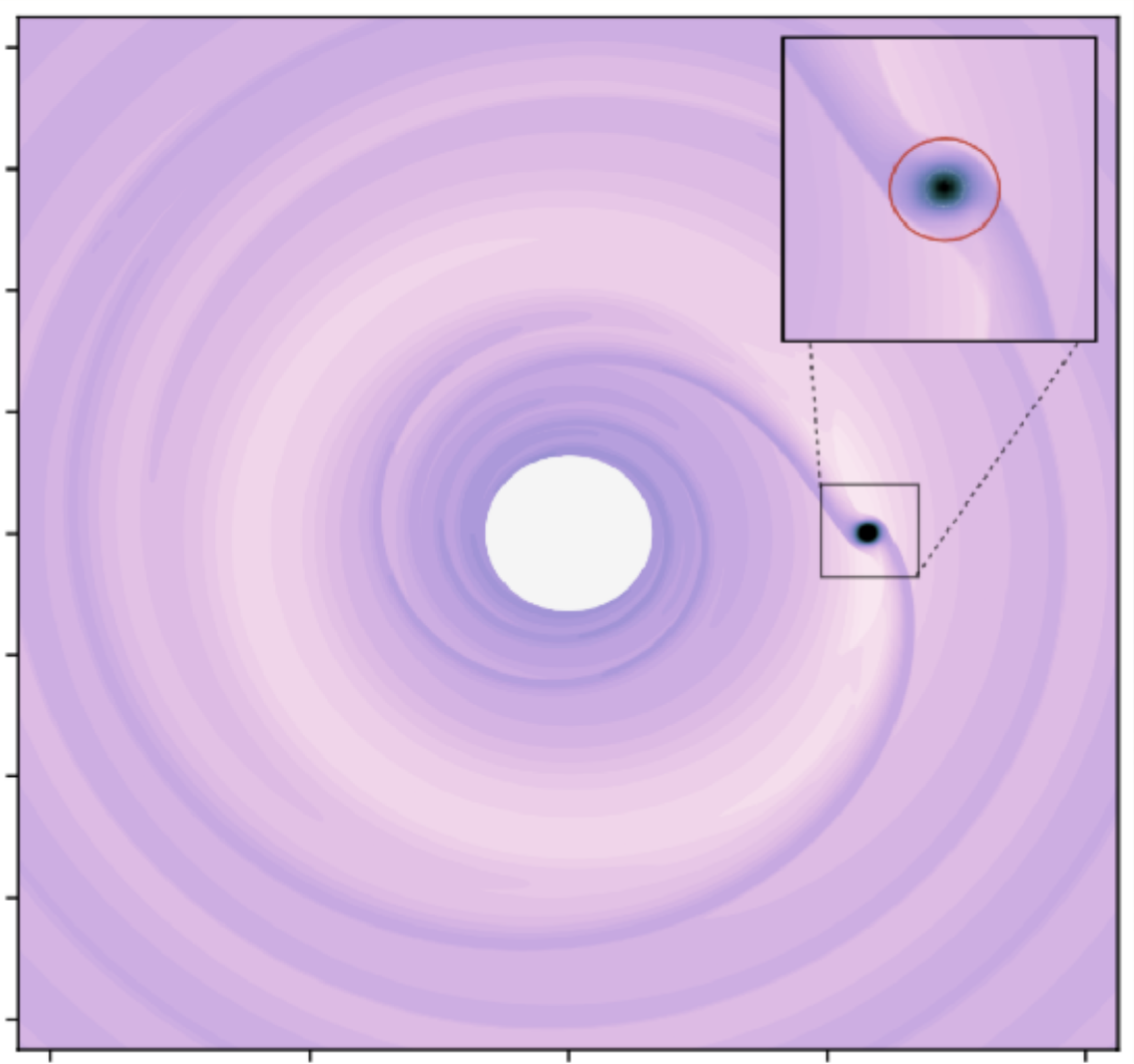
LISA data analysis challenges for constraining environmental effects

Jonathan Gair, Max Planck Institute for Gravitational Physics, Potsdam



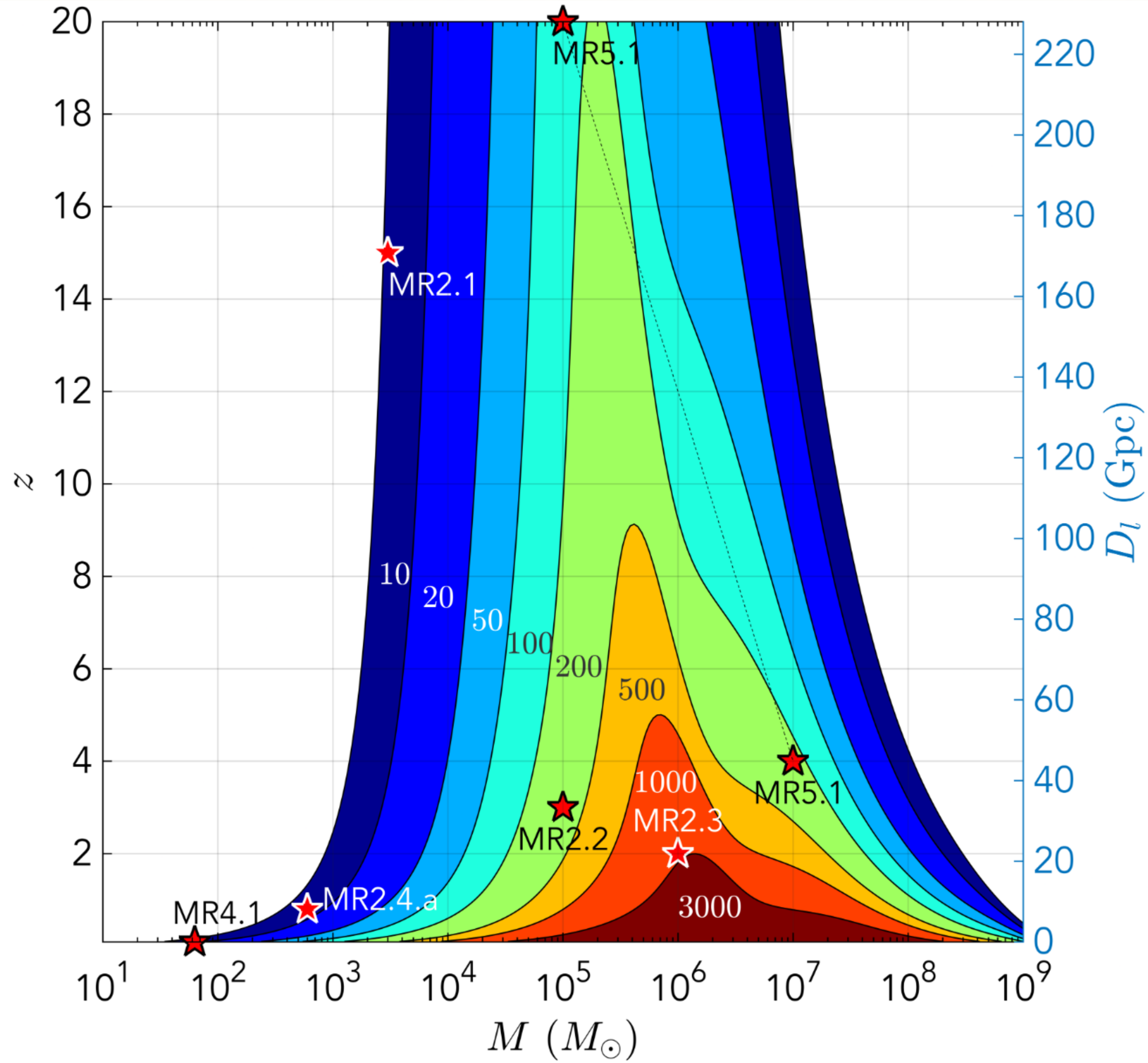
Talk outline

- Quick overview of LISA sources
- State of the art of LISA Data Analysis
- Outstanding Challenges and their potential impact on environmental constraints.
  - EMRI search and parameter estimation
  - Instrumental glitches
  - Data gaps
  - Lack of noise knowledge
  - Source confusion

Sources: massive black hole mergers

- Expected to occur following mergers of the host galaxies. LISA can observe gravitational waves from these with very high signal-to-noise ratio.

Sources: massive black hole mergers

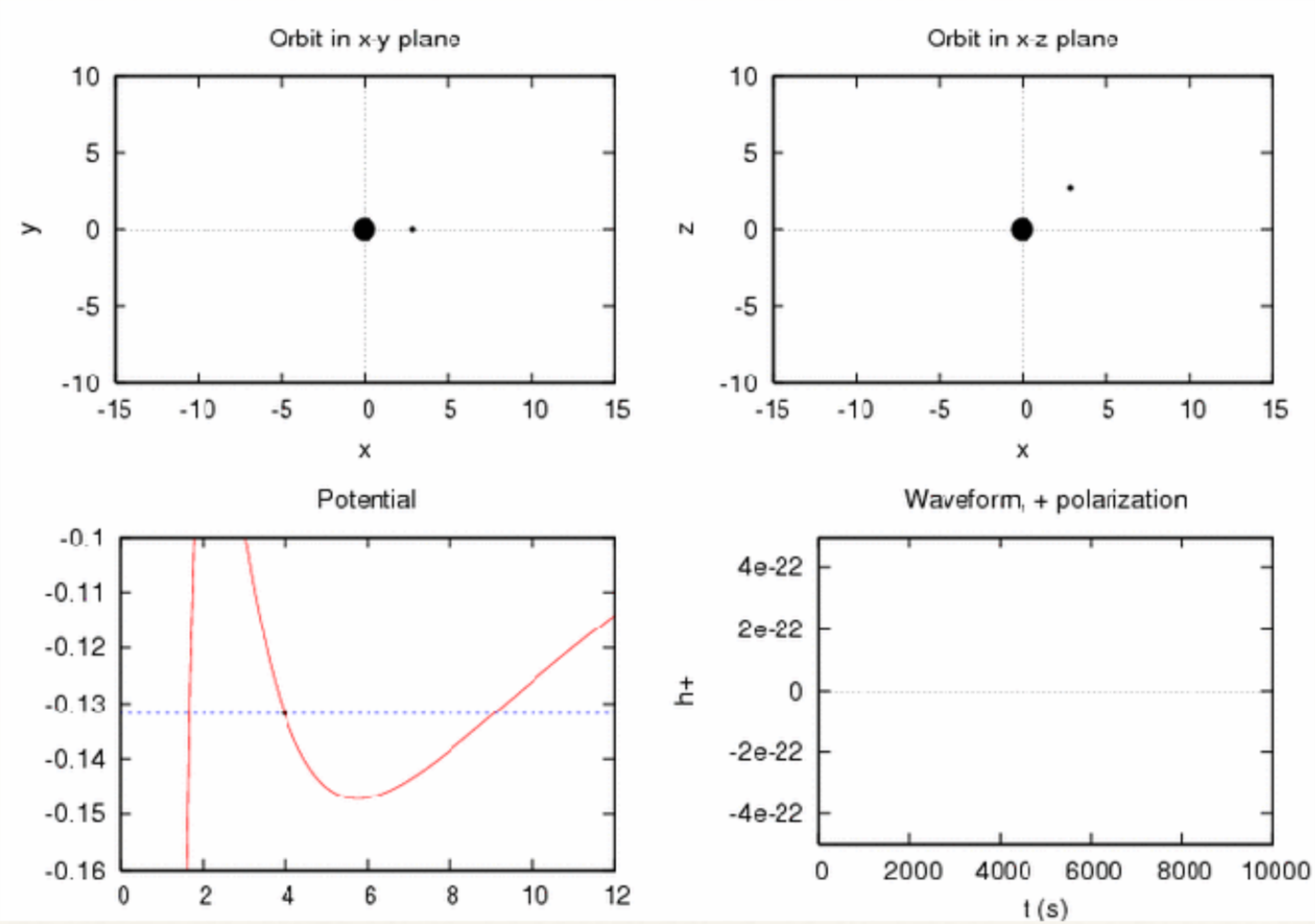


Sources: massive black hole mergers

- Expected to occur following mergers of the host galaxies. LISA can observe gravitational waves from these with very high signal-to-noise ratio.
- Expected event rate depends on assumptions about black hole population (Klein+, 2016)
  - Light pop-III seed model: expect to see ~350 events.
  - Heavy seed model, no delay in binary formation: ~550 events.
  - Heavy seed model, with delays: ~50 events.
- LISA observations expected too provide mass measurements to ~ 0.1-1%, spin measurements to 1-10%, sky location to ~tens of square degrees and luminosity distance to ~10%.

Sources: extreme-mass-ratio inspirals

- The inspiral of a compact object into a massive black hole in the centre of a galaxy.
- Form as a result of scattering in dense galacto-centric stellar clusters.
- Orbits are expected to be both eccentric and inclined - rich waveform structure.



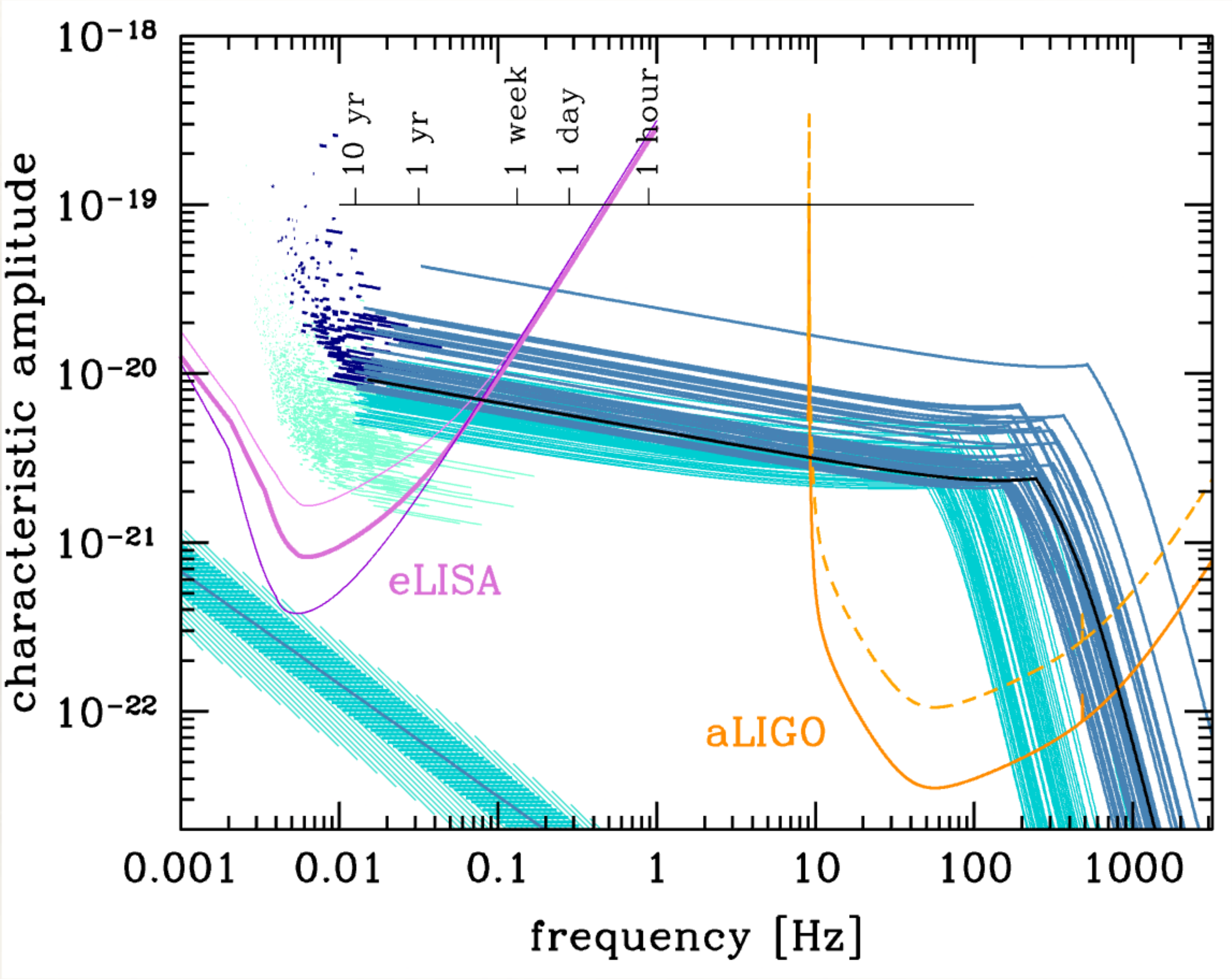
Sources: extreme-mass-ratio inspirals

- There are large astrophysical uncertainties, but expect to see between a few tens and a few hundreds of events.

Model	Mass function	MBH spin	Cusp erosion	M–σ relation	N_{p}	CO mass $[M_\odot]$	EMRI rate $[\mathrm{yr}^{-1}]$ Total	Detected (AKK)	Detected (AKS)
M1	Barausse12	a98	yes	Gultekin09	10	10	1600	294	189
M2	Barausse12	a98	yes	KormendyHo13	10	10	1400	220	146
M3	Barausse12	a98	yes	GrahamScott13	10	10	2770	809	440
M4	Barausse12	a98	yes	Gultekin09	10	30	520 (620)	260	221
M5	Gair10	a98	no	Gultekin09	10	10	140	47	15
M6	Barausse12	a98	no	Gultekin09	10	10	2080	479	261
M7	Barausse12	a98	yes	Gultekin09	0	10	15800	2712	1765
M8	Barausse12	a98	yes	Gultekin09	100	10	180	35	24
M9	Barausse12	aflat	yes	Gultekin09	10	10	1530	217	177
M10	Barausse12	a0	yes	Gultekin09	10	10	1520	188	188
M11	Gair10	a0	no	Gultekin09	100	10	13	1	1
M12	Barausse12	a98	no	Gultekin09	0	10	20000	4219	2279

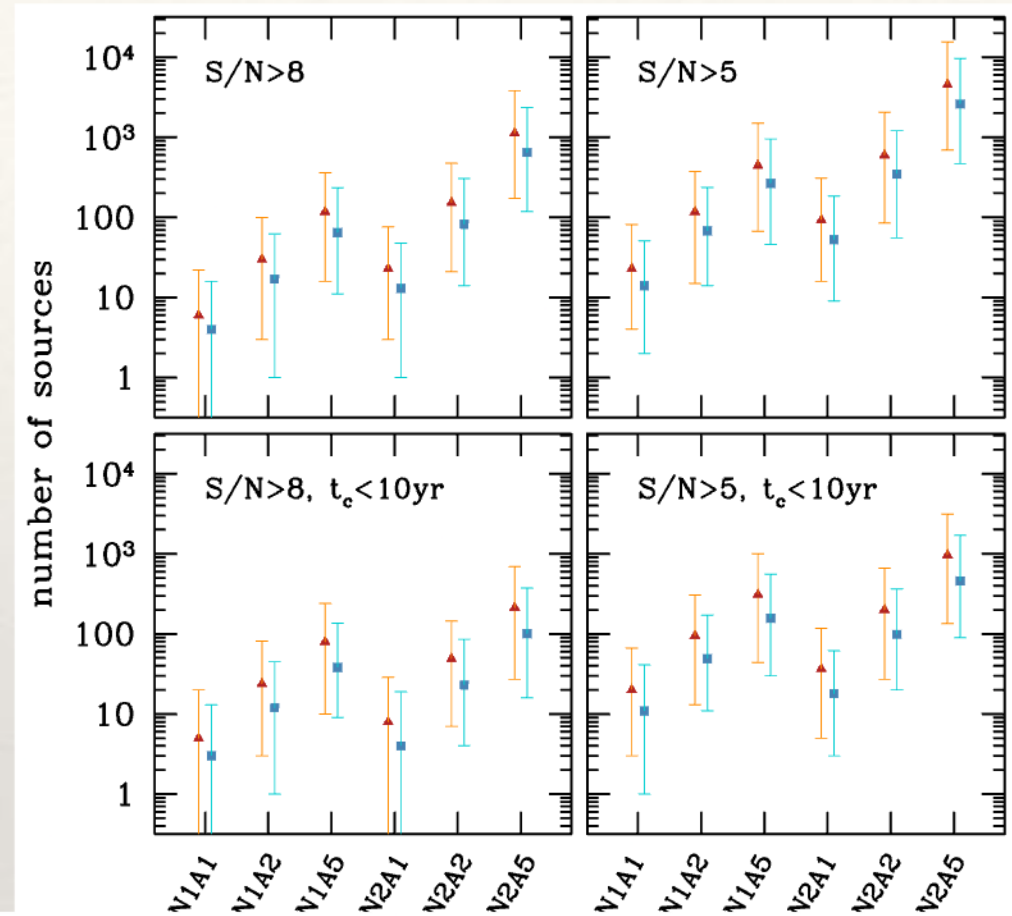
Stellar-origin black hole binaries

- GW150914 would have been observable by LISA ~5 years before being observed by LIGO, with S/N~10 in a 5yr observation. (Sesana 2016)
- LISA provides sky location to ~0.few square degrees and time of coalescence to ~few s.
- Number of events could be high (several tens) but there are significant uncertainties.



Stellar-origin black hole binaries

- GW150914 would have been observable by LISA ~5 years before being observed by LIGO, with S/N~10 in a 5yr observation.   (Sesana 2016)
- LISA provides sky location to ~0.few square degrees and time of coalescence to ~few s.
- Number of events could be high (several tens) but there are significant uncertainties.

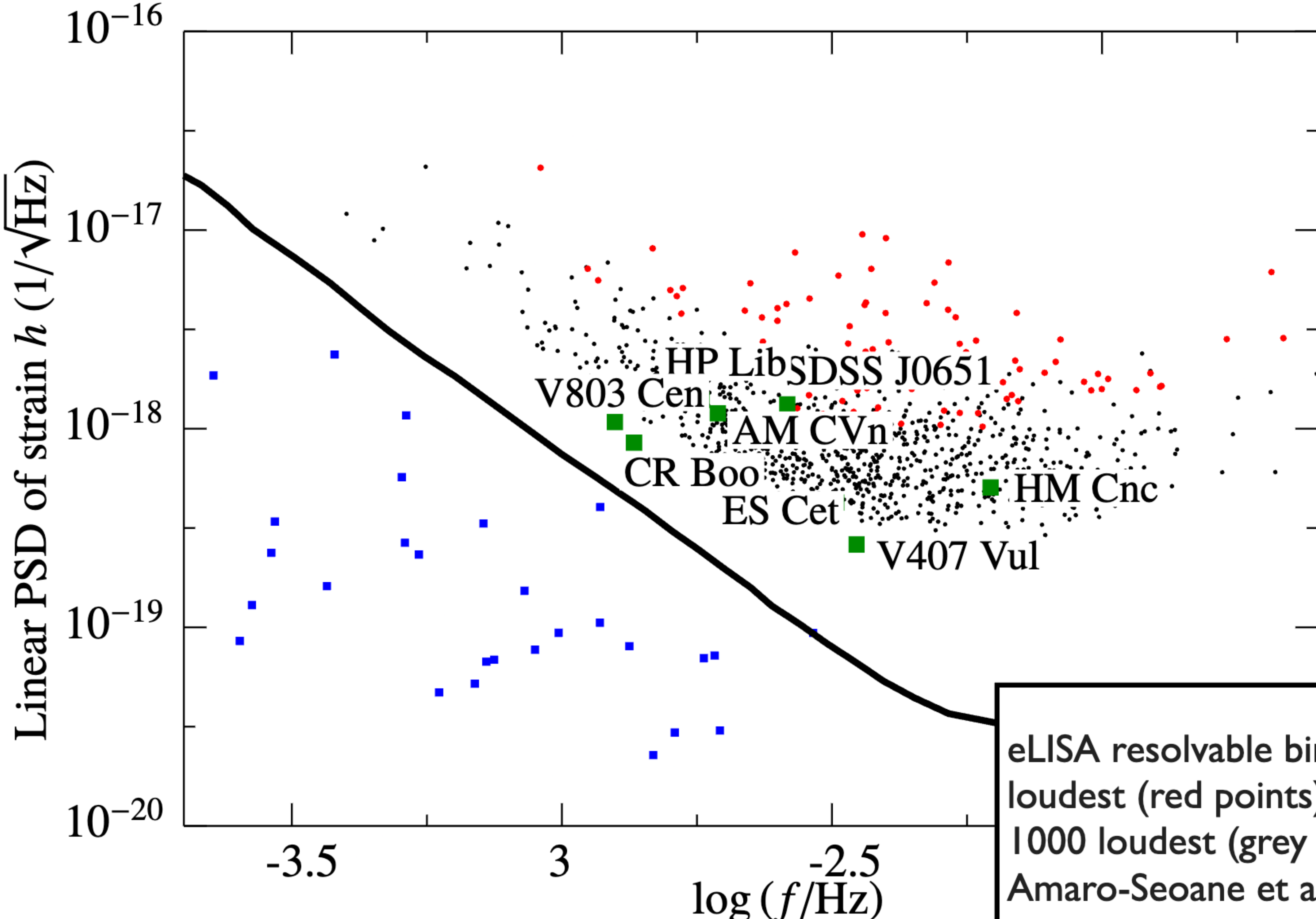


Mass distribution	$R/(\mathrm{Gpc}^{-3}\mathrm{yr}^{-1})$		
	PyCBC	GstLAL	Combined
Event based			
GW150914	$3.2^{+8.3}_{-2.7}$	$3.6^{+9.1}_{-3.0}$	$3.4^{+8.6}_{-2.8}$
LVT151012	$9.2^{+30.3}_{-8.5}$	$9.2^{+31.4}_{-8.5}$	$9.4^{+30.4}_{-8.7}$
GW151226	35^{+92}_{-29}	37^{+94}_{-31}	37^{+92}_{-31}
All	53^{+100}_{-40}	56^{+105}_{-42}	55^{+99}_{-41}
Astrophysical			
Flat in log mass	31^{+43}_{-21}	30^{+43}_{-21}	30^{+43}_{-21}
Power Law (−2.35)	100^{+136}_{-69}	95^{+138}_{-67}	99^{+138}_{-70}

Other sources

- Compact binaries in the Milky Way
  - Binaries of stellar remnants (white dwarfs or neutron stars) with orbital periods of ~1 hour.
  - Known (verification) and unknown sources.
  - Signals almost monochromatic.

Other sources

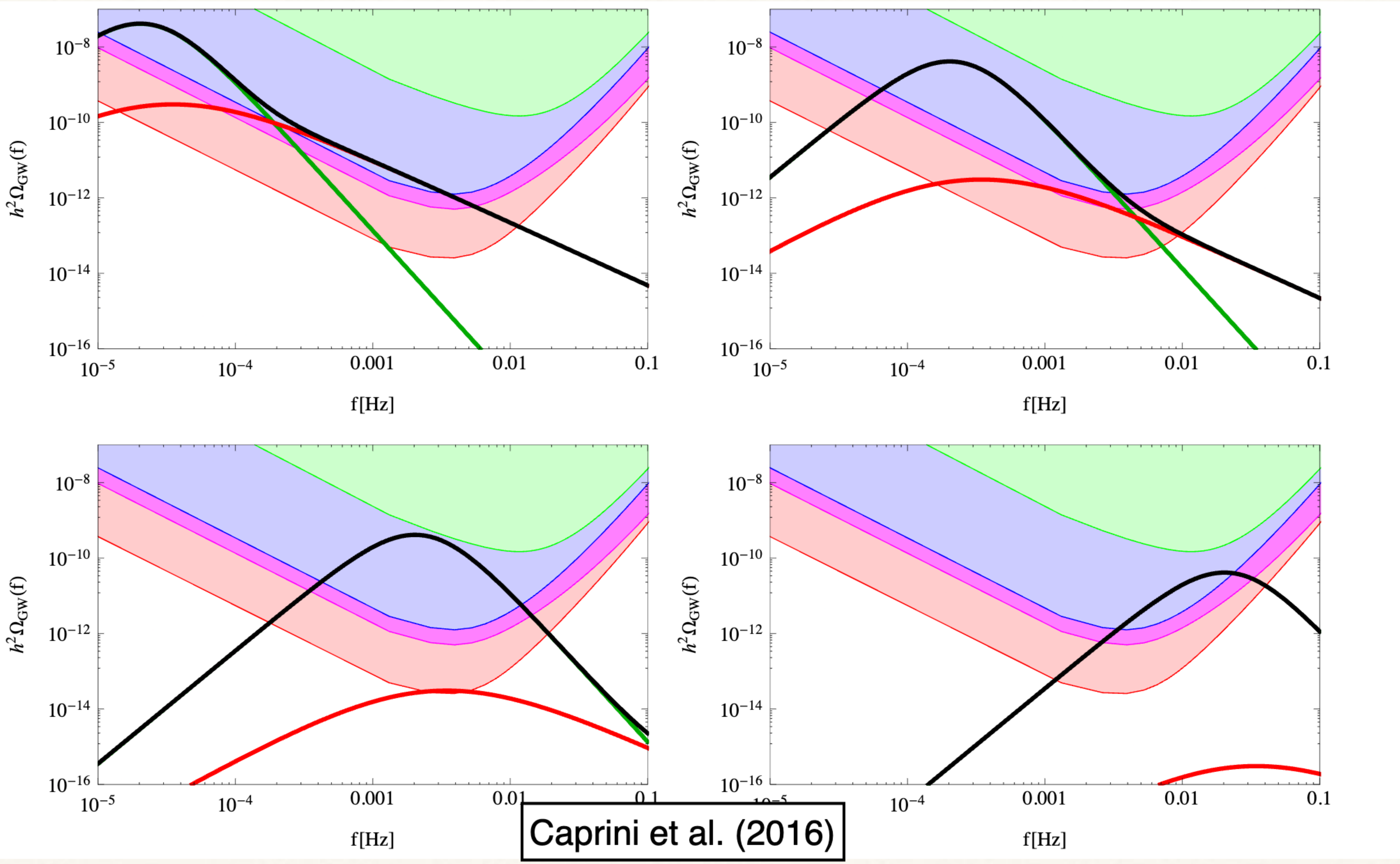


eLISA resolvable binaries: 100 loudest (red points)
1000 loudest (grey points)
Amaro-Seoane et al. (2013)

Other sources

- Compact binaries in the Milky Way
  - Binaries of stellar remnants (white dwarfs or neutron stars) with orbital periods of ~1 hour.
  - Known (verification) and unknown sources.
  - Signals almost monochromatic.
  - LISA expected to detect >10000 binaries with $S/N > 7$.
  - LISA should determine 2D/3D location for 4500/1250 sources, measure df/dt for several thousand and d^2f/dt^2 for a handful.
- Cosmological sources
  - Processes occurring at the TeV scale in the early Universe could generate a mHz stochastic gravitational wave background.
  - Cosmic string networks could produce both individual burst events and a stochastic background.

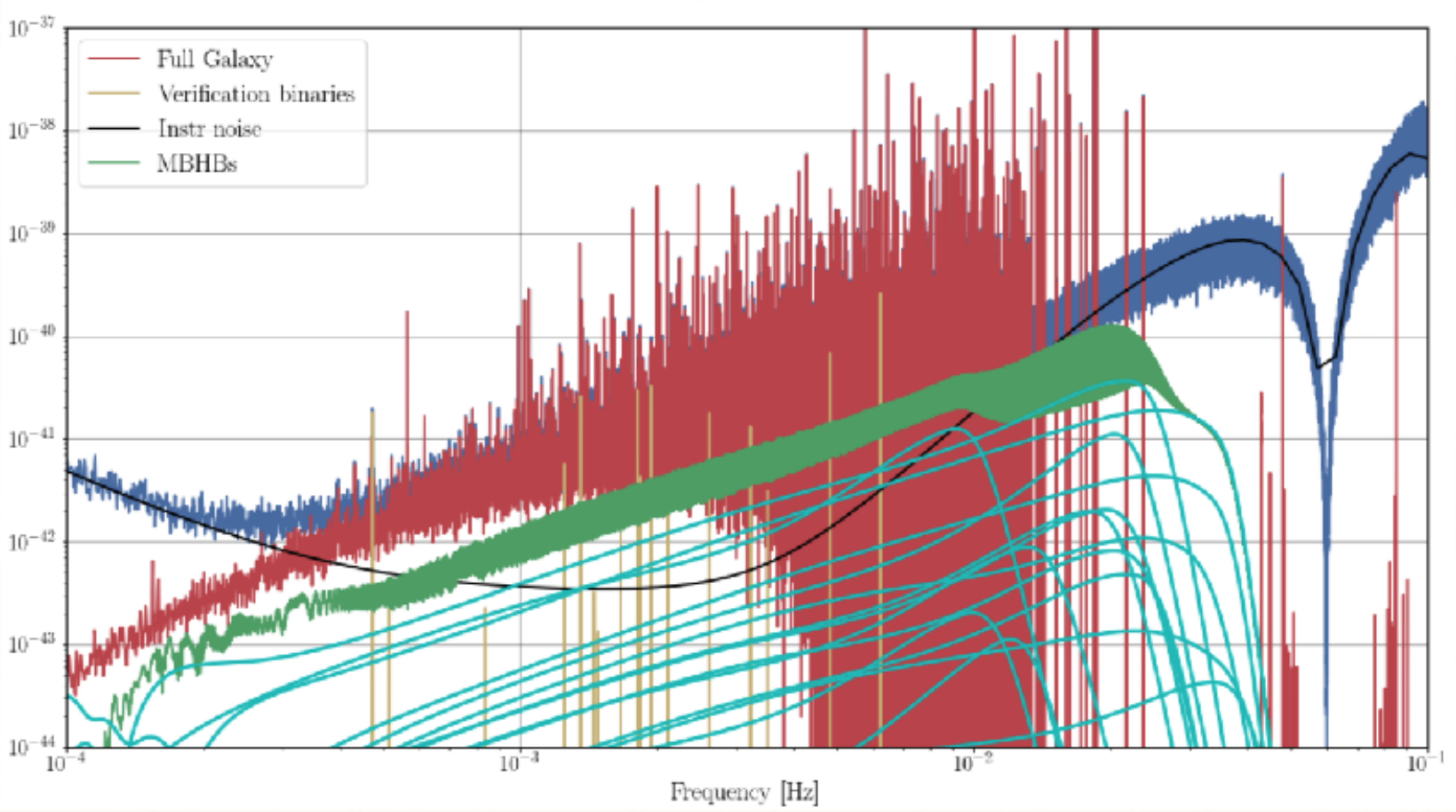
Other sources



Caprini et al. (2016)

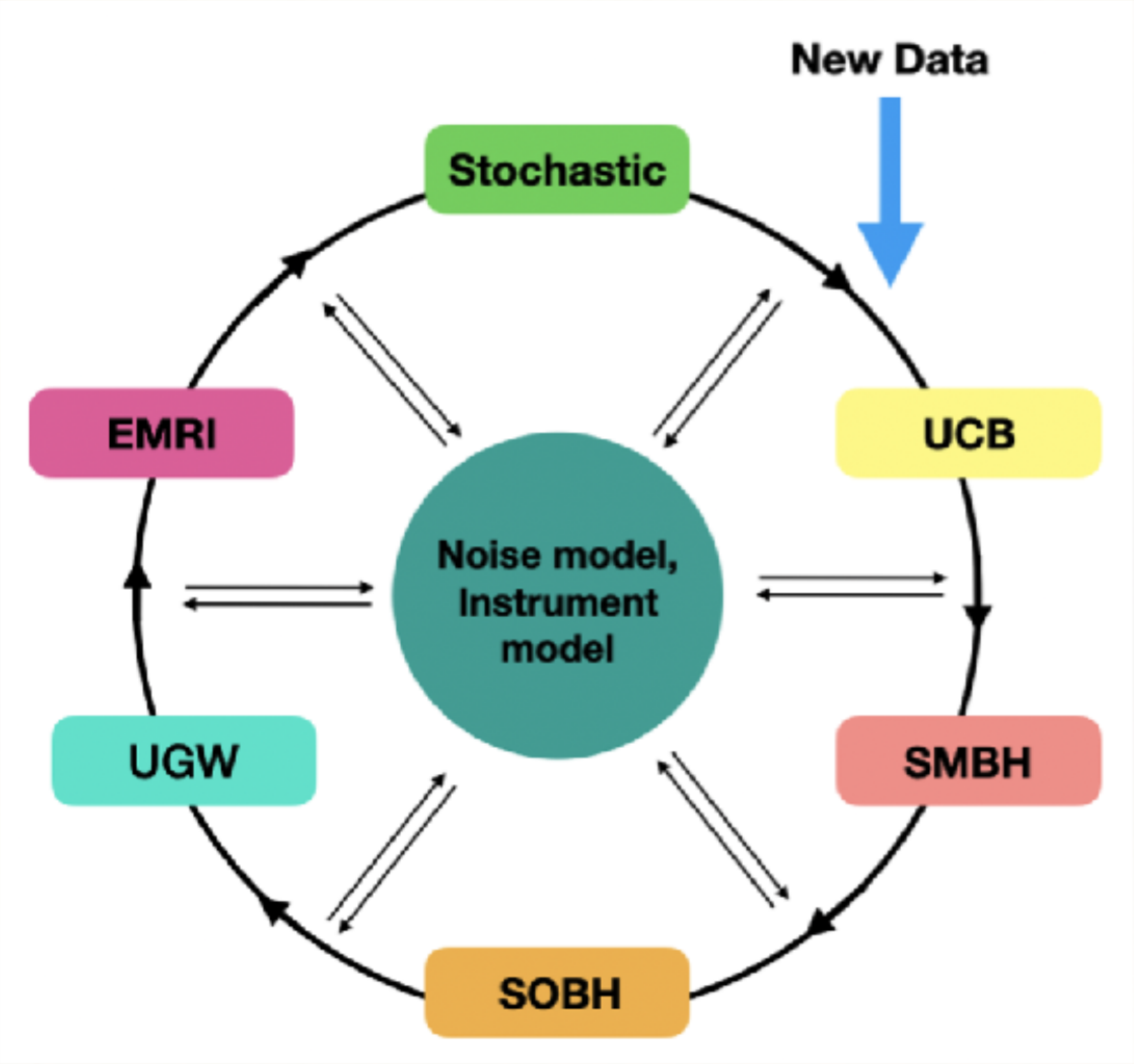
LISA Data Analysis

- Primary challenge in LISA Data Analysis is the need for a *Global Fit* of an unknown number of sources of all of these different types.

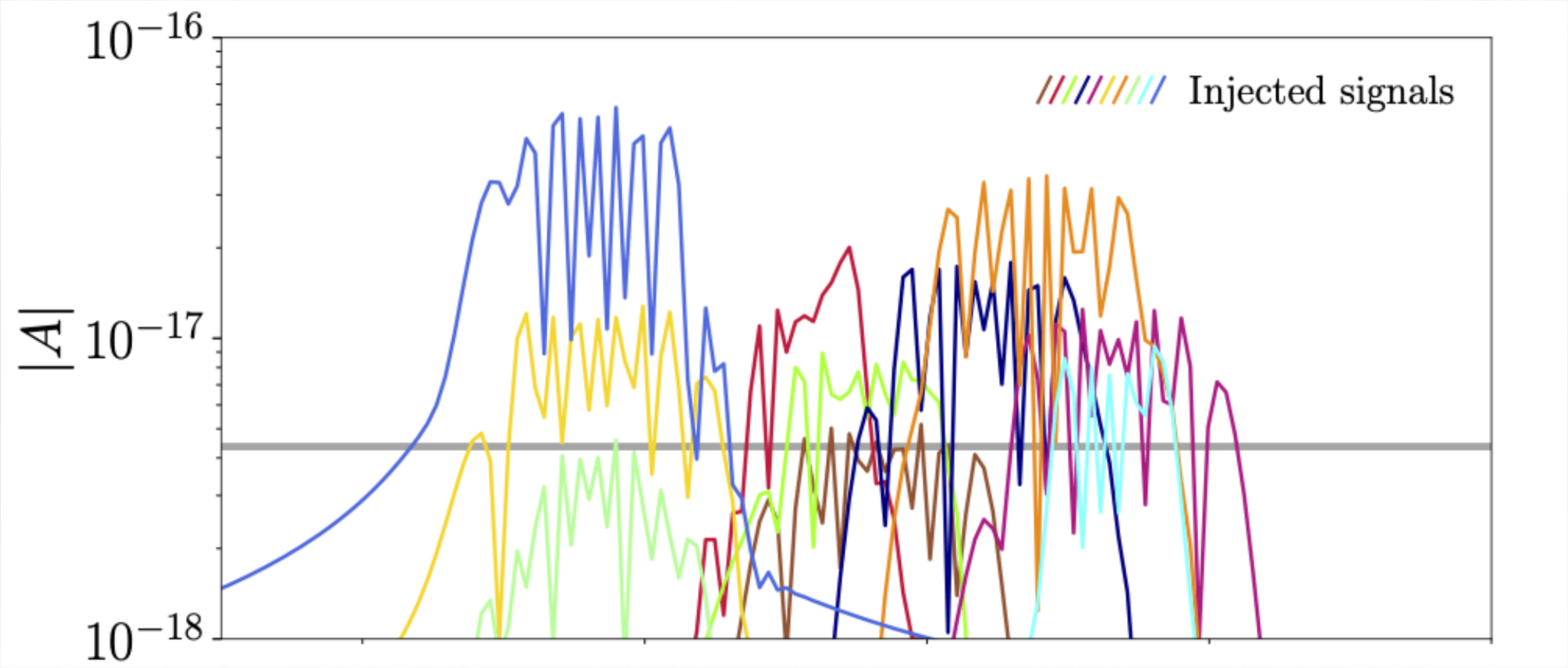


LISA Data Analysis

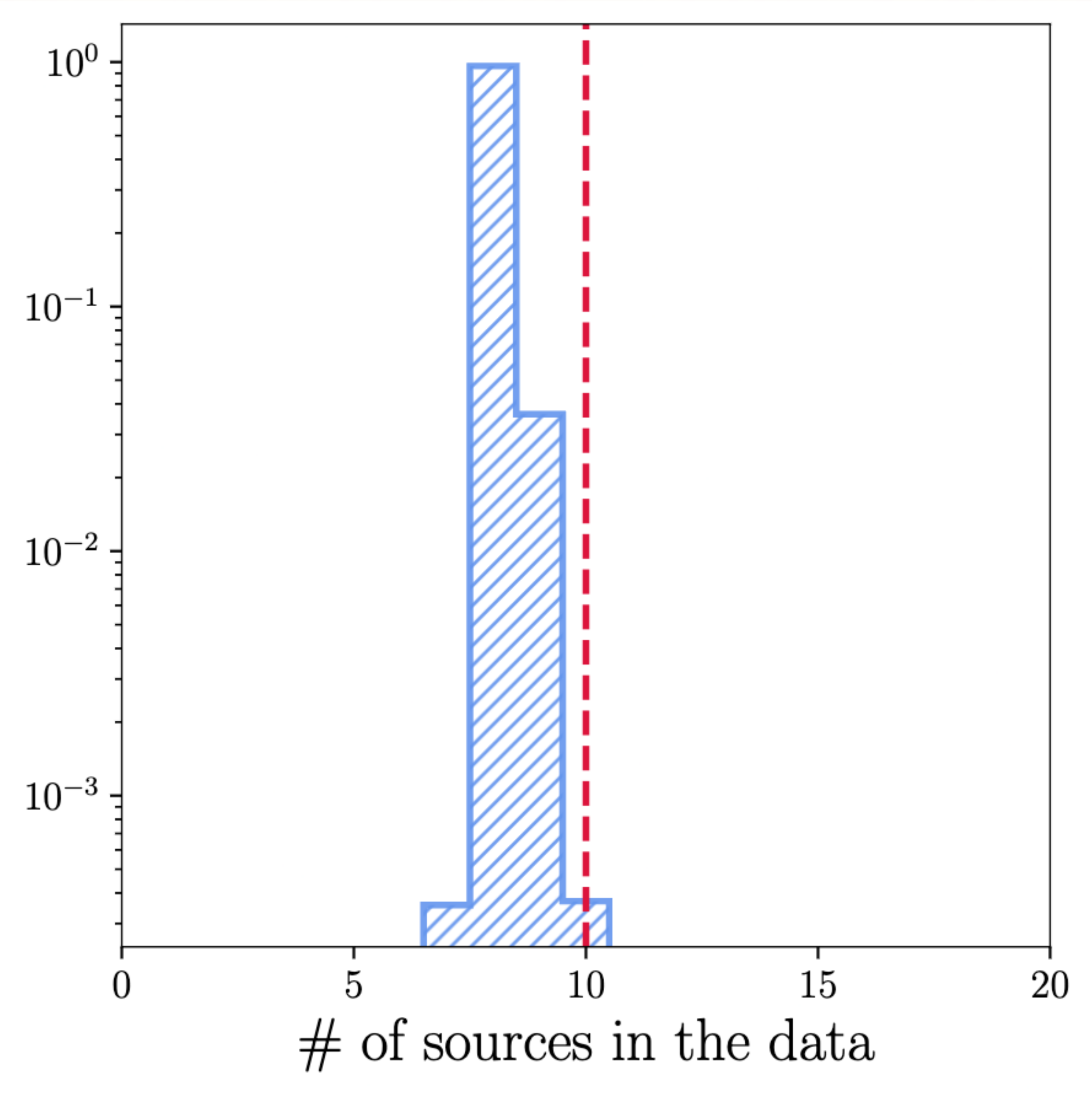
- Typical strategy adopted is to iteratively update the solution for one source type and then move to the next.
- Solution will be continuously refined as new data is added. Plan is for multiple data releases during the mission.
- A key component of the analysis is *variable dimensionality*. Techniques like *reversible jump MCMC* are necessary.



LISA Data Analysis

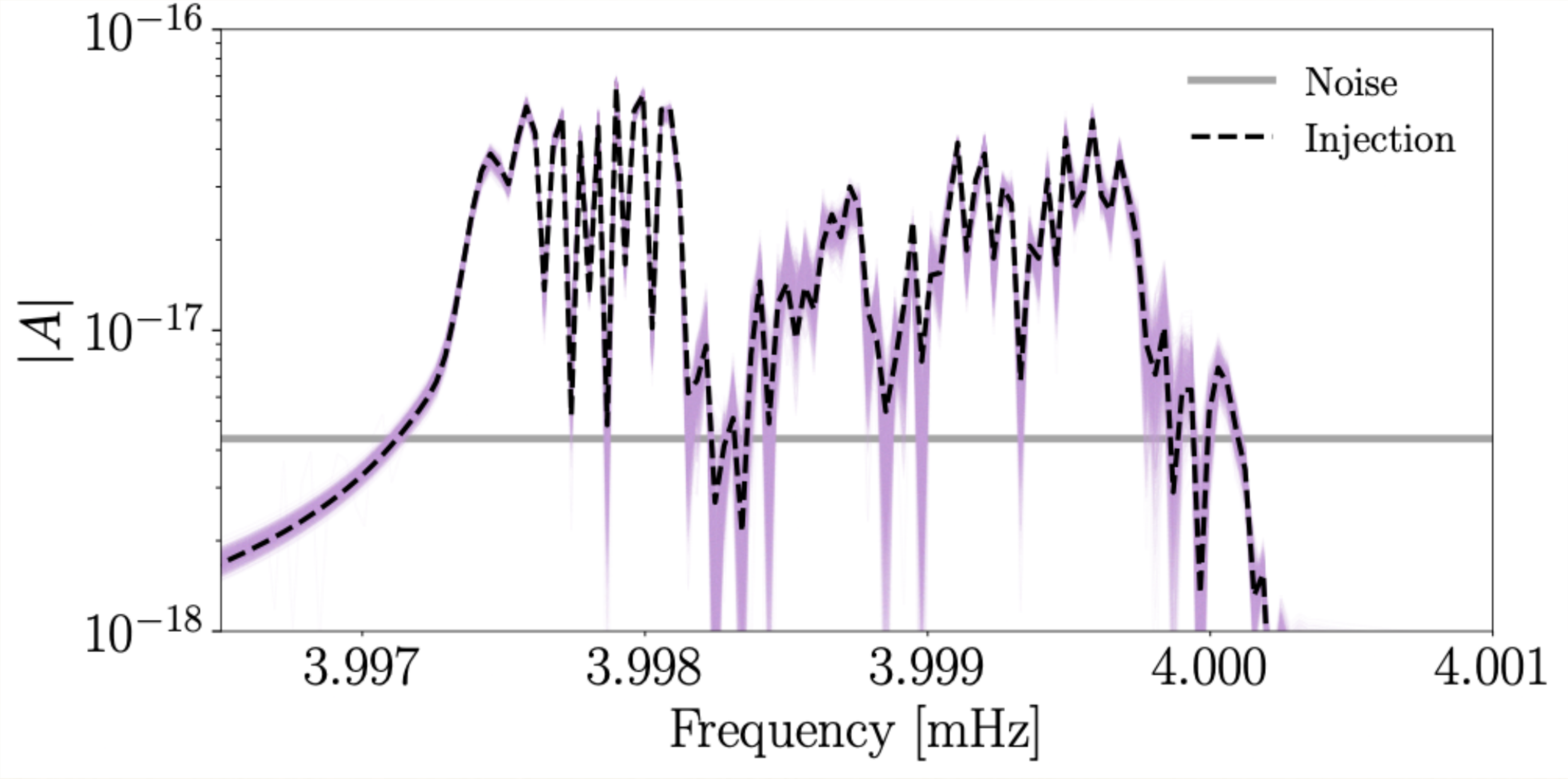


LISA Data Analysis



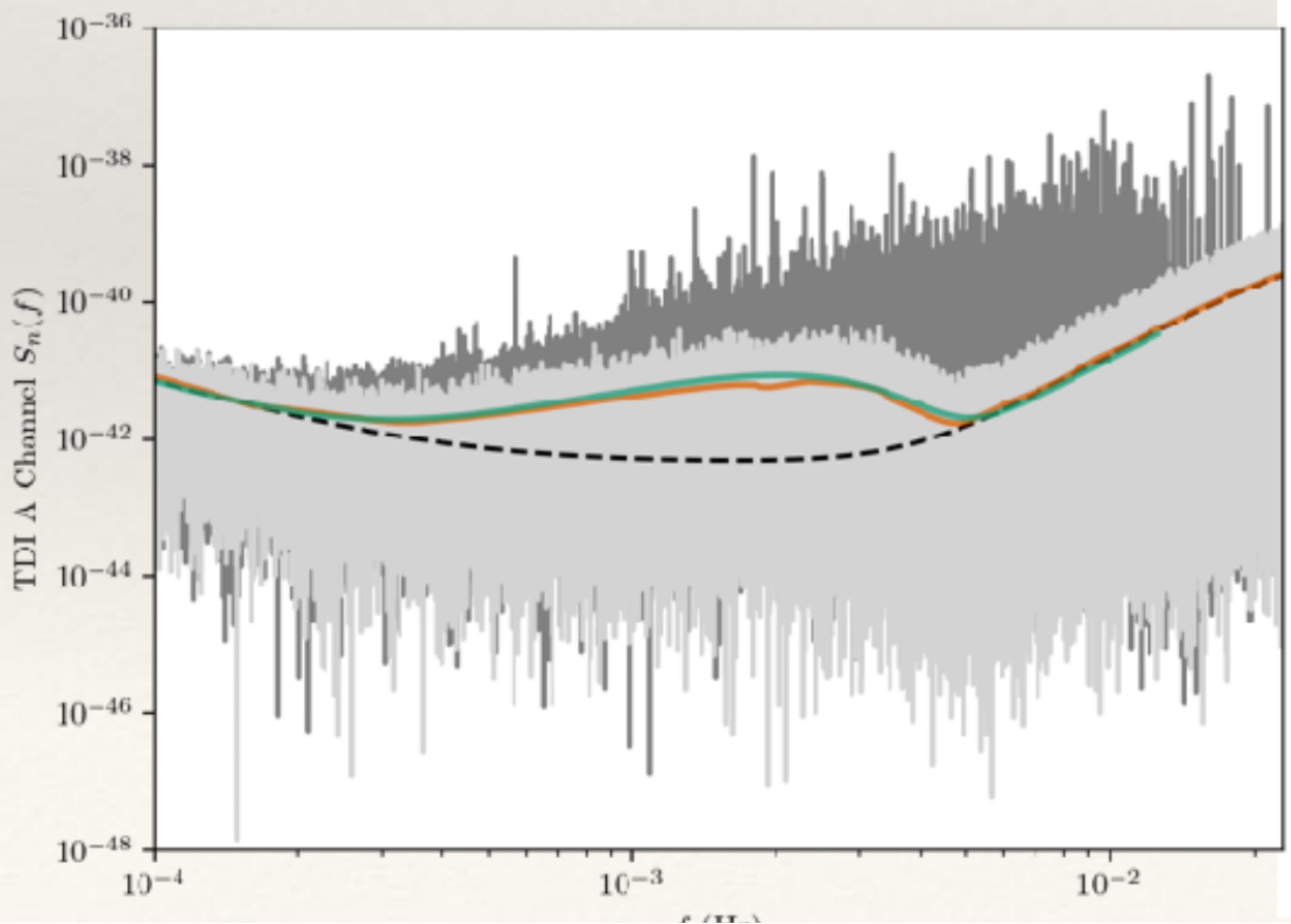
- In RJMCMC, both *within-model* and *between-model* moves are proposed.
- Between-model moves add or remove one (or more) sources from the model.
- These moves are represented by a reversible mapping of the current point $\mathbf{x}$ and a set of random variables $\mathbf{u}$ into the new variables $\mathbf{x}'$ and $\mathbf{u}'$.
- Acceptance probability: $$\alpha = \min\left(1, \frac{p(\boldsymbol{\theta}'|\mathbf{x})q(\boldsymbol{u}')}{p(\boldsymbol{\theta}|\mathbf{x})q(\boldsymbol{u})}\left|\frac{\partial(\boldsymbol{\theta}', \mathbf{u}')}{\partial(\boldsymbol{\theta}, \mathbf{u})}\right|\right)$$

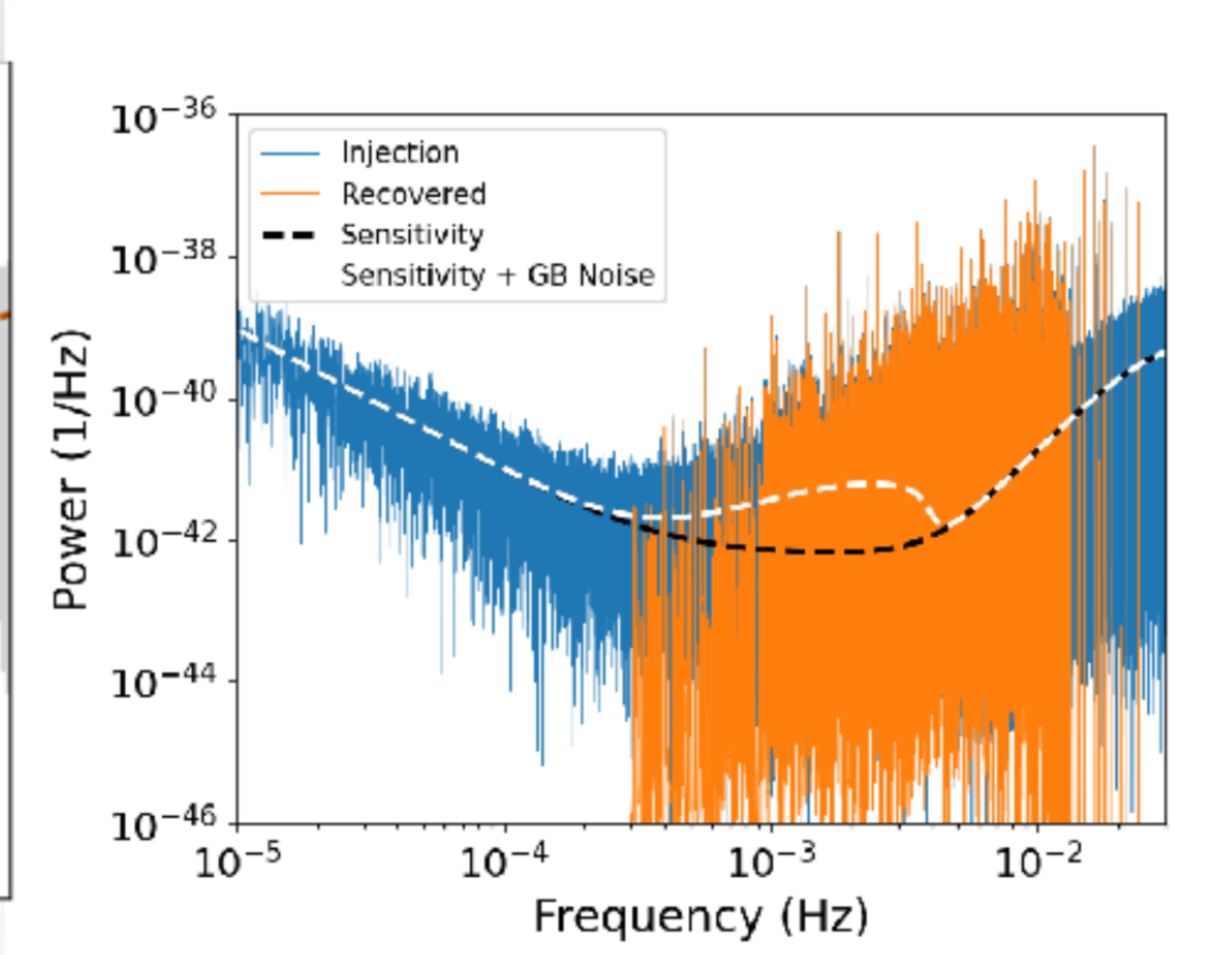
LISA Data Analysis



LISA Data Analysis: state of the art

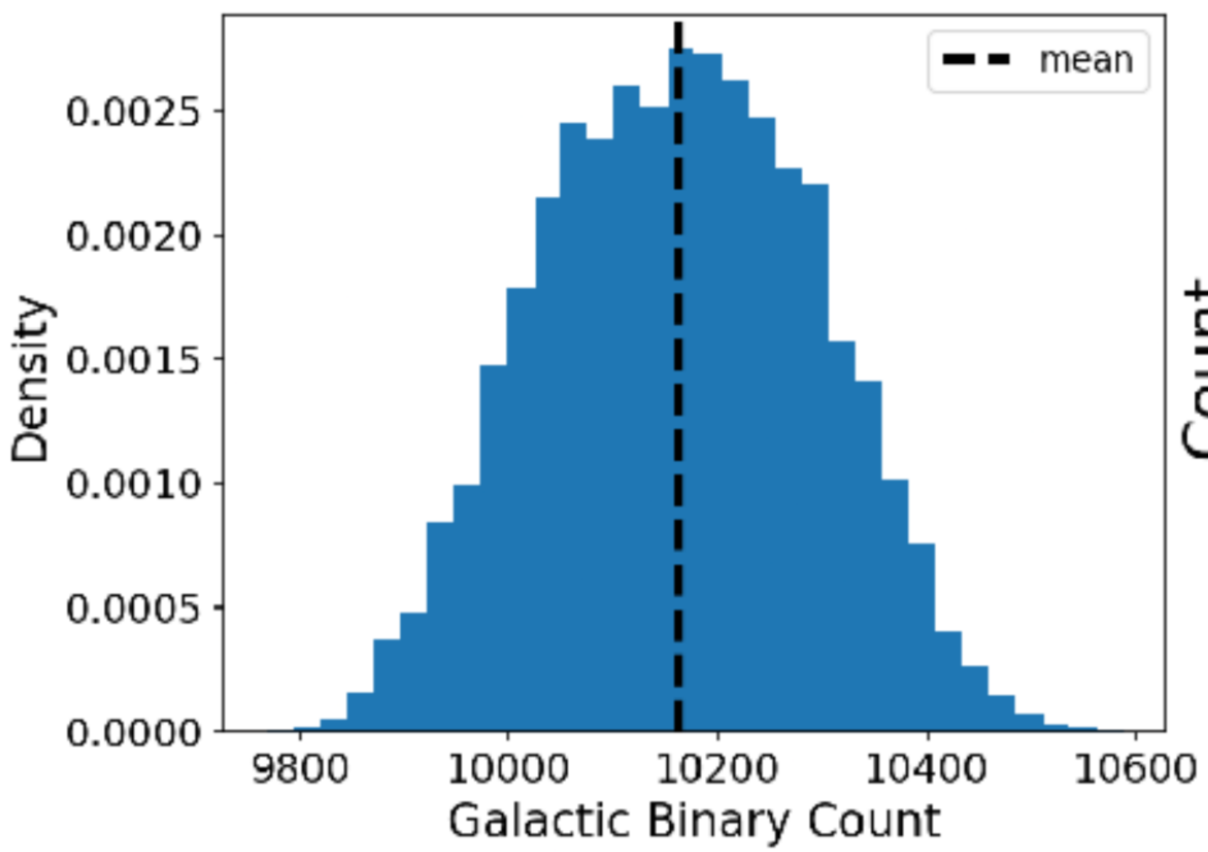
- Data analysis development progressing in LISA Data Challenge group.
- State of the art: successful analysis of data sets containing a galaxy of white dwarf binaries, several MBH mergers and unknown instrumental noise.
- Next stage: additional source types, increasing instrumental realism.

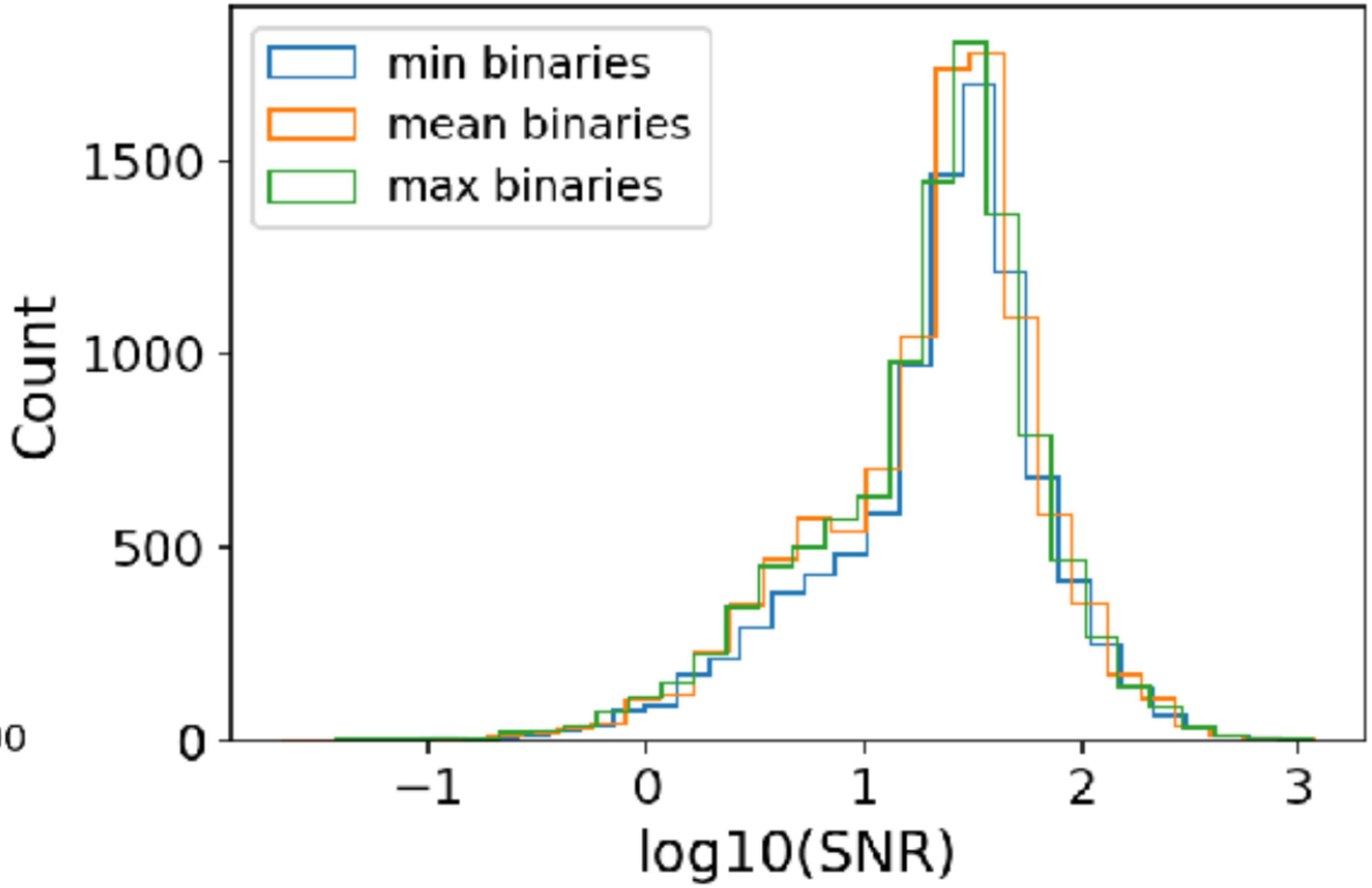




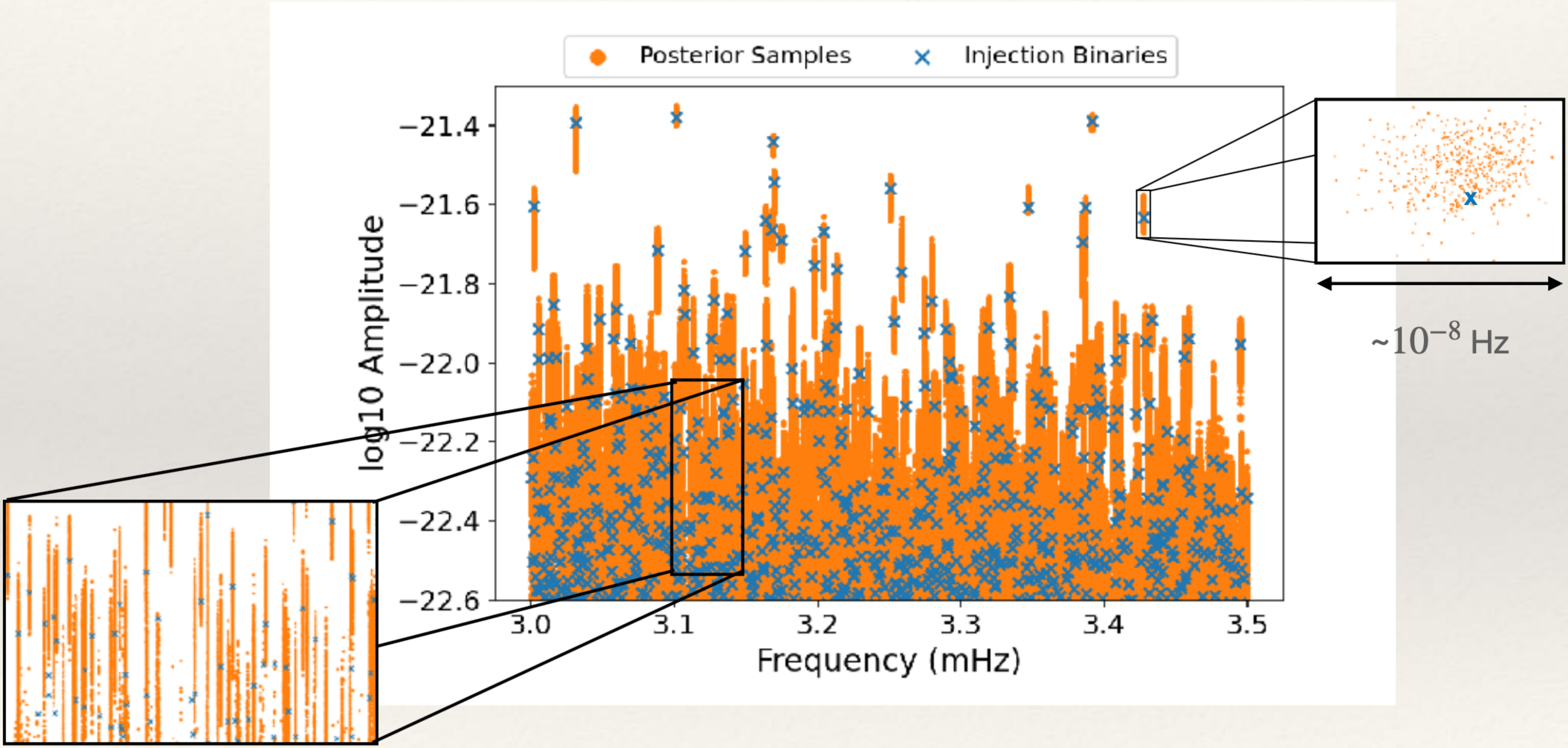
LISA Data Analysis: state of the art

- Recover ~10,000 galactic binaries.
- Total number of binaries is uncertain at ~1% level.

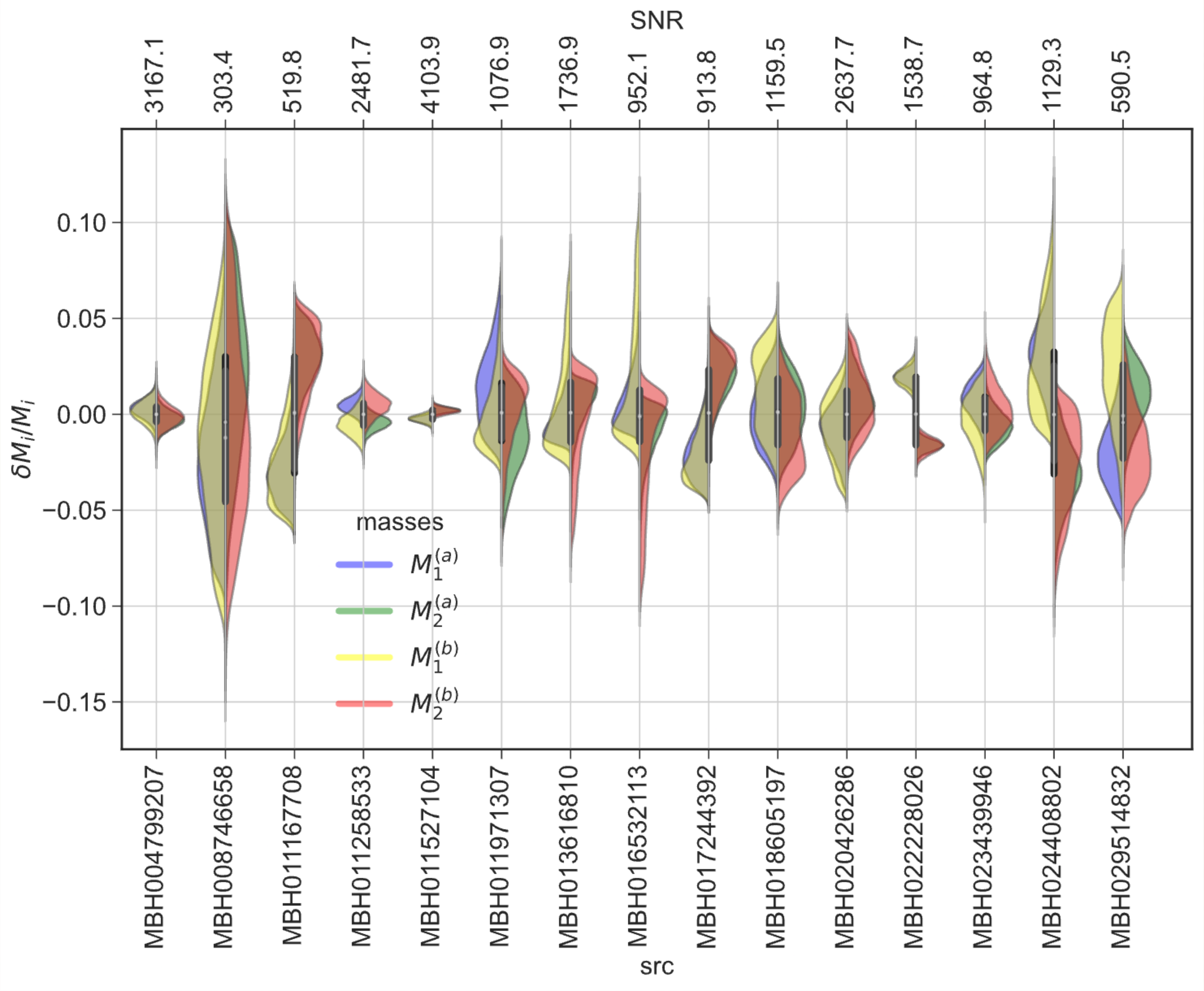




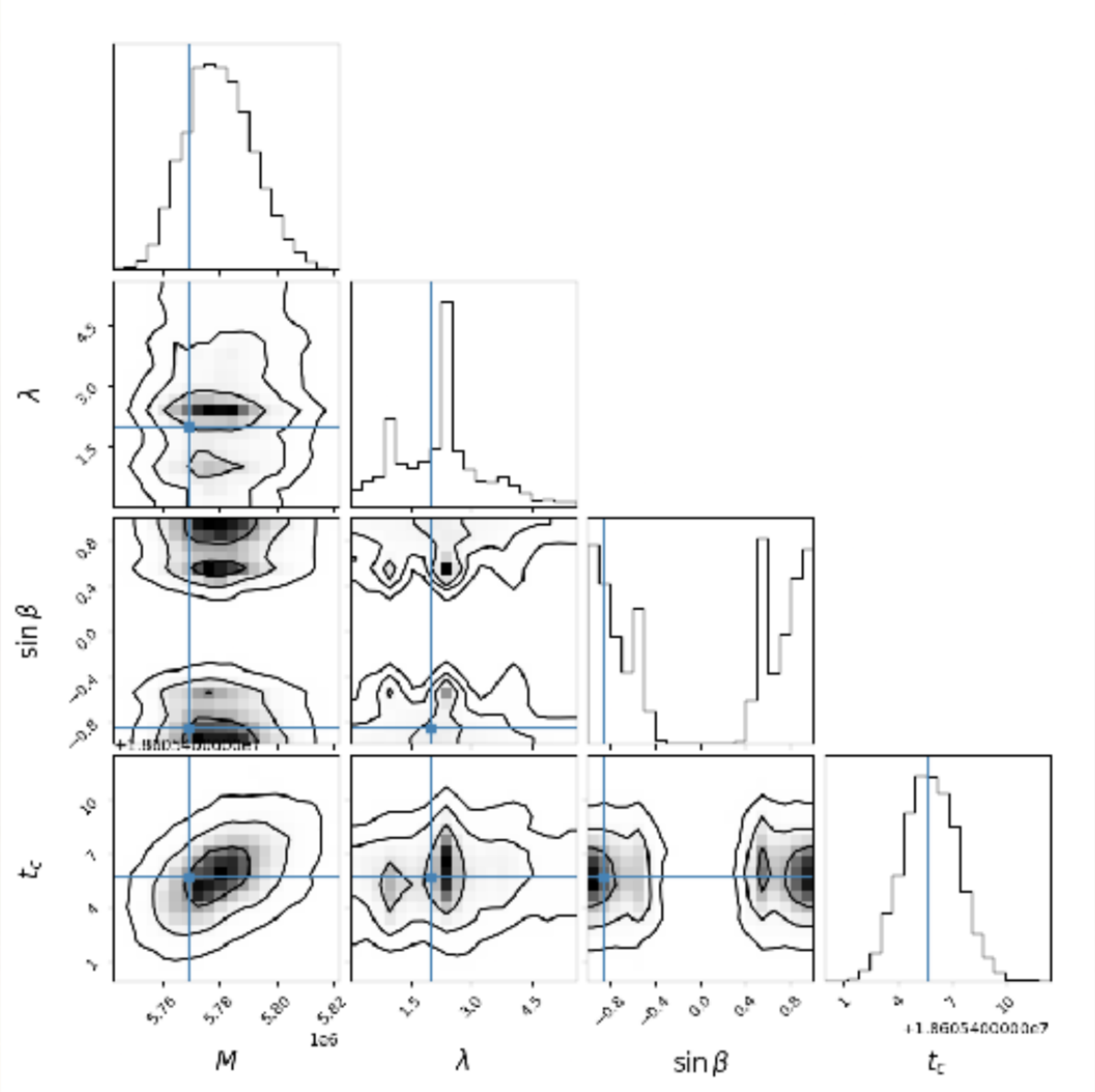
LISA Data Analysis: state of the art

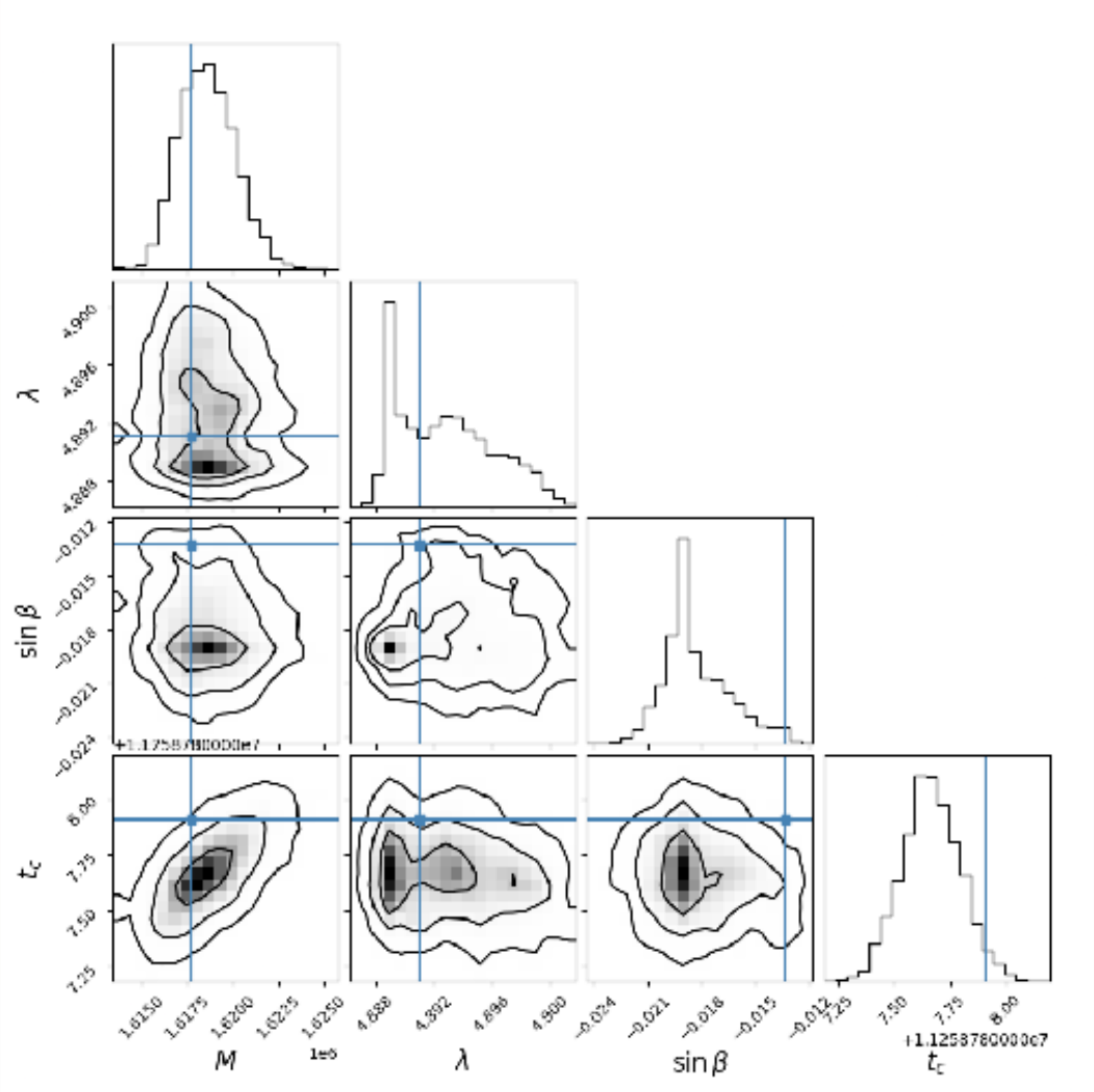


LISA Data Analysis: state of the art

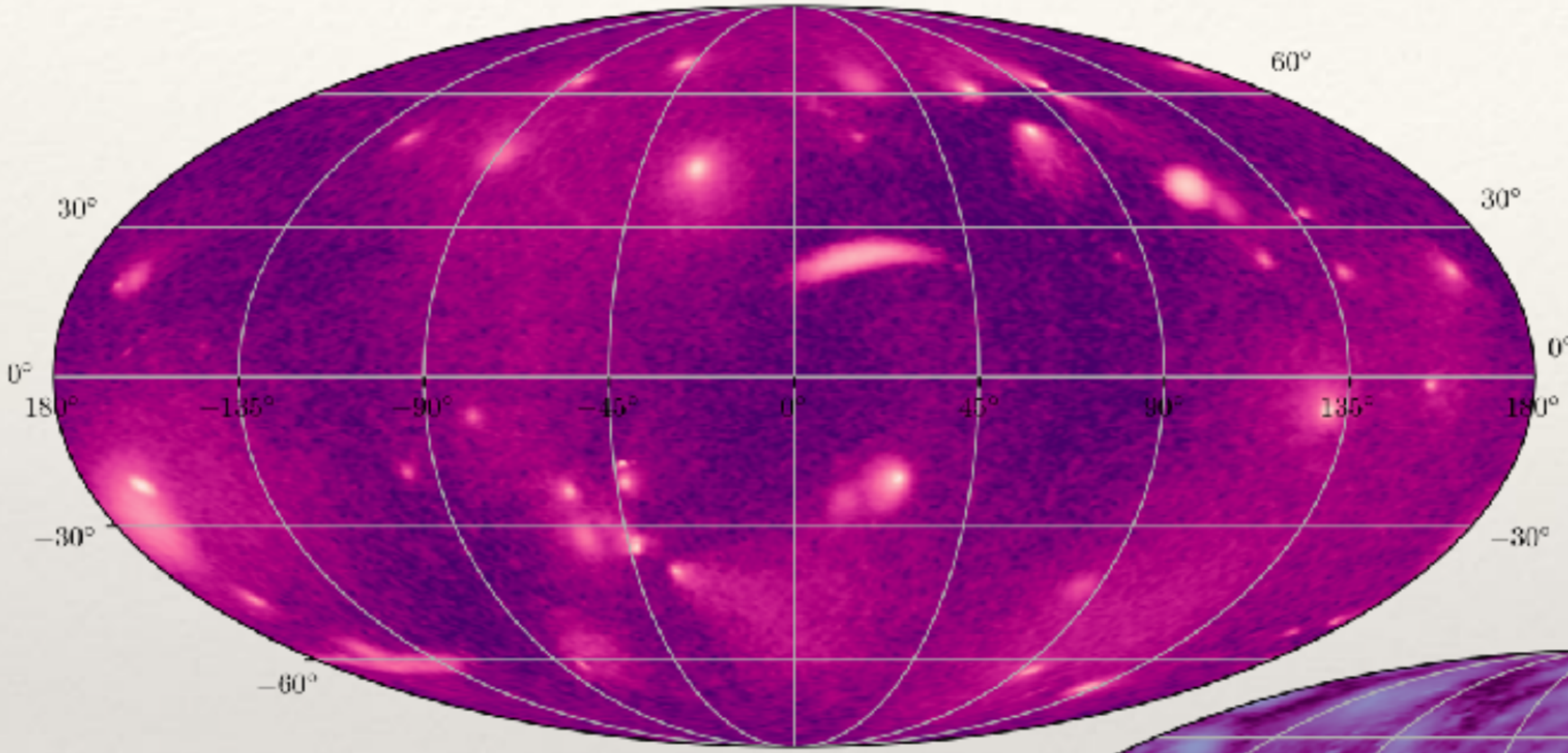


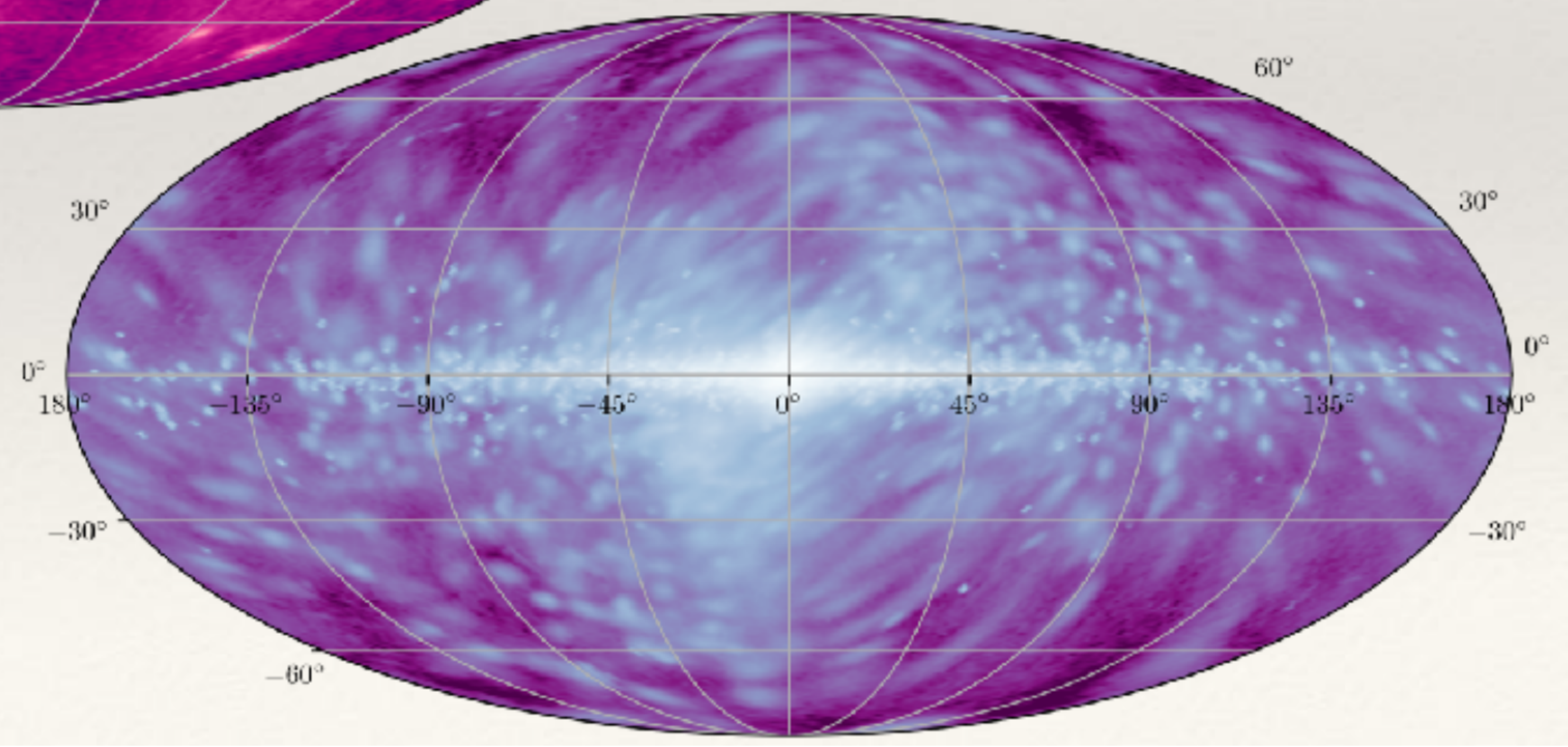
LISA Data Analysis: state of the art





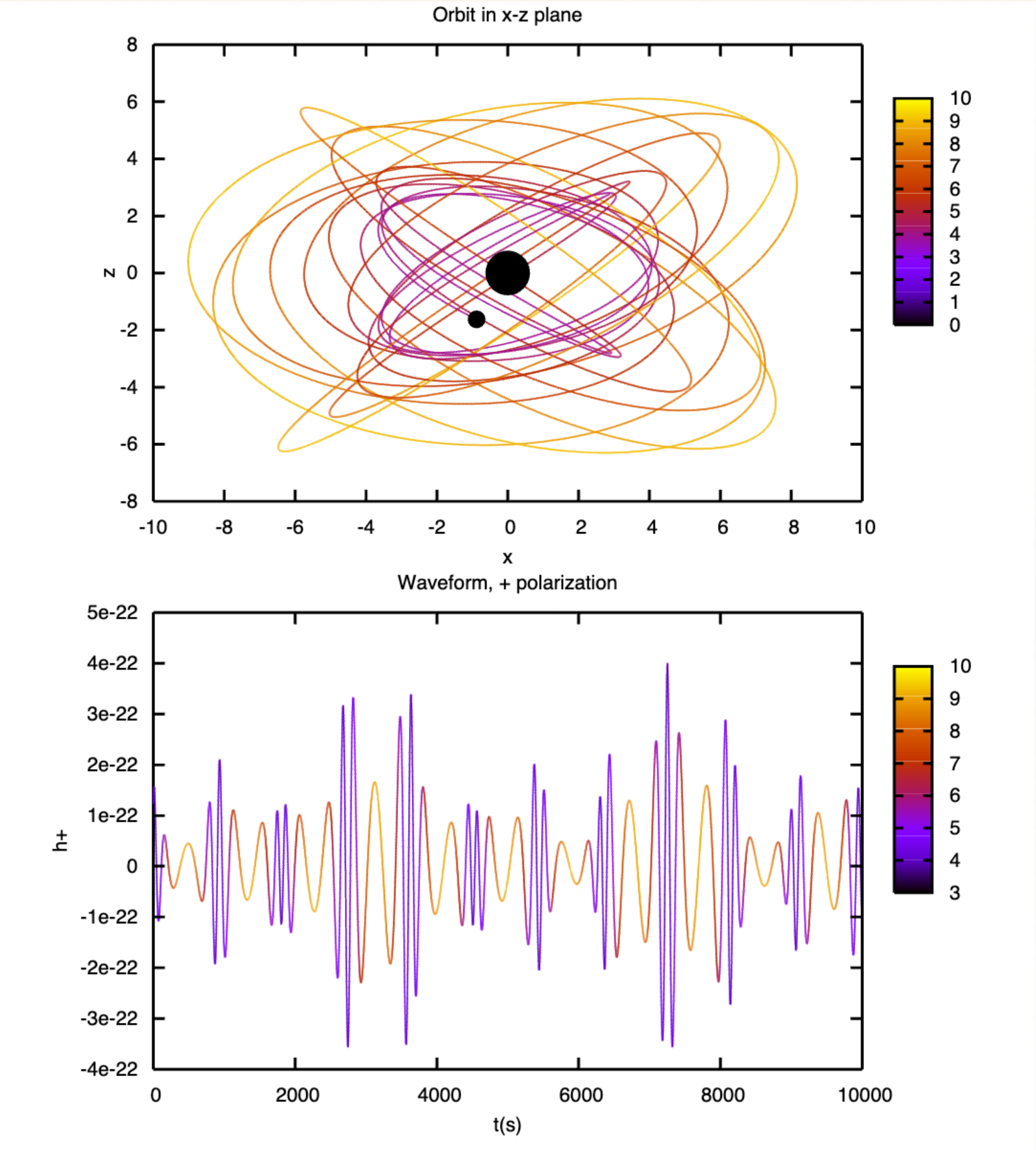
LISA Data Analysis: state of the art



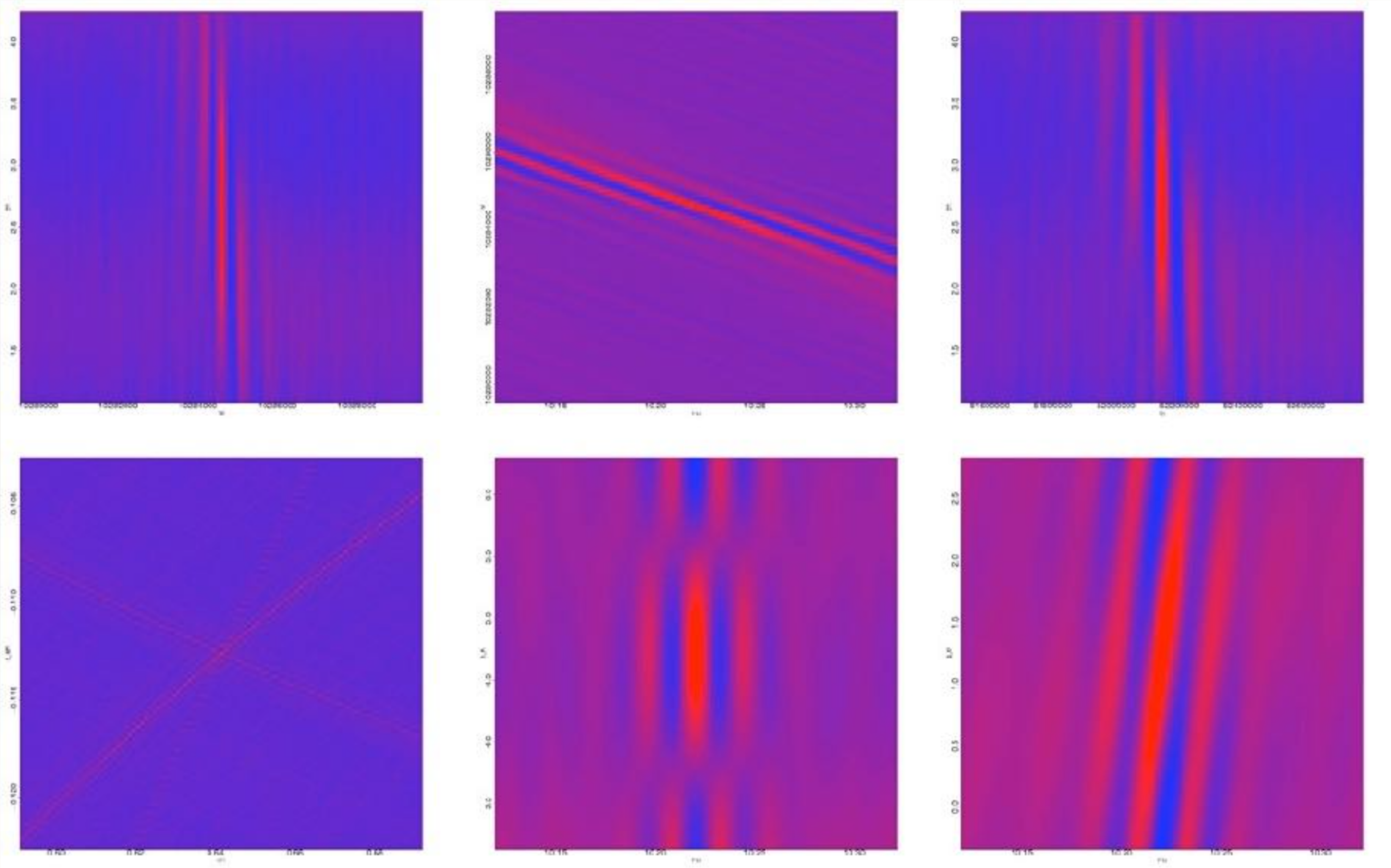


Outstanding challenges: EMRIs

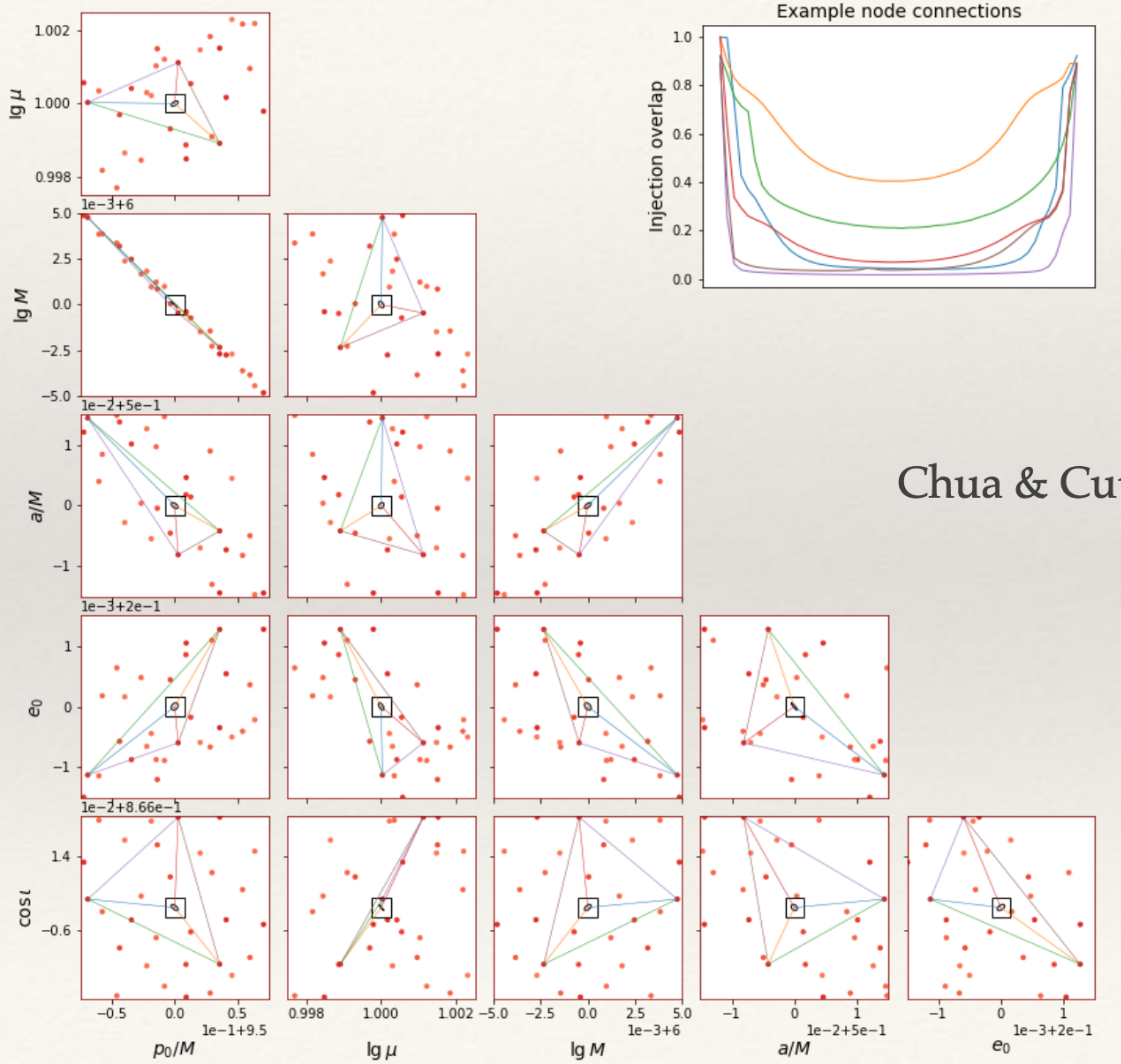
- EMRI waveforms show a rich structure which is composed of harmonics of three fundamental frequencies.
- EMRIs generate $O(10^5)$ cycles in strong field region close to central black hole.
- ***In principle***: high precision measurements of system properties, including environmental effects.
- ***In practice***: narrow mode in big parameter space, many secondaries.



Outstanding challenges: EMRIs

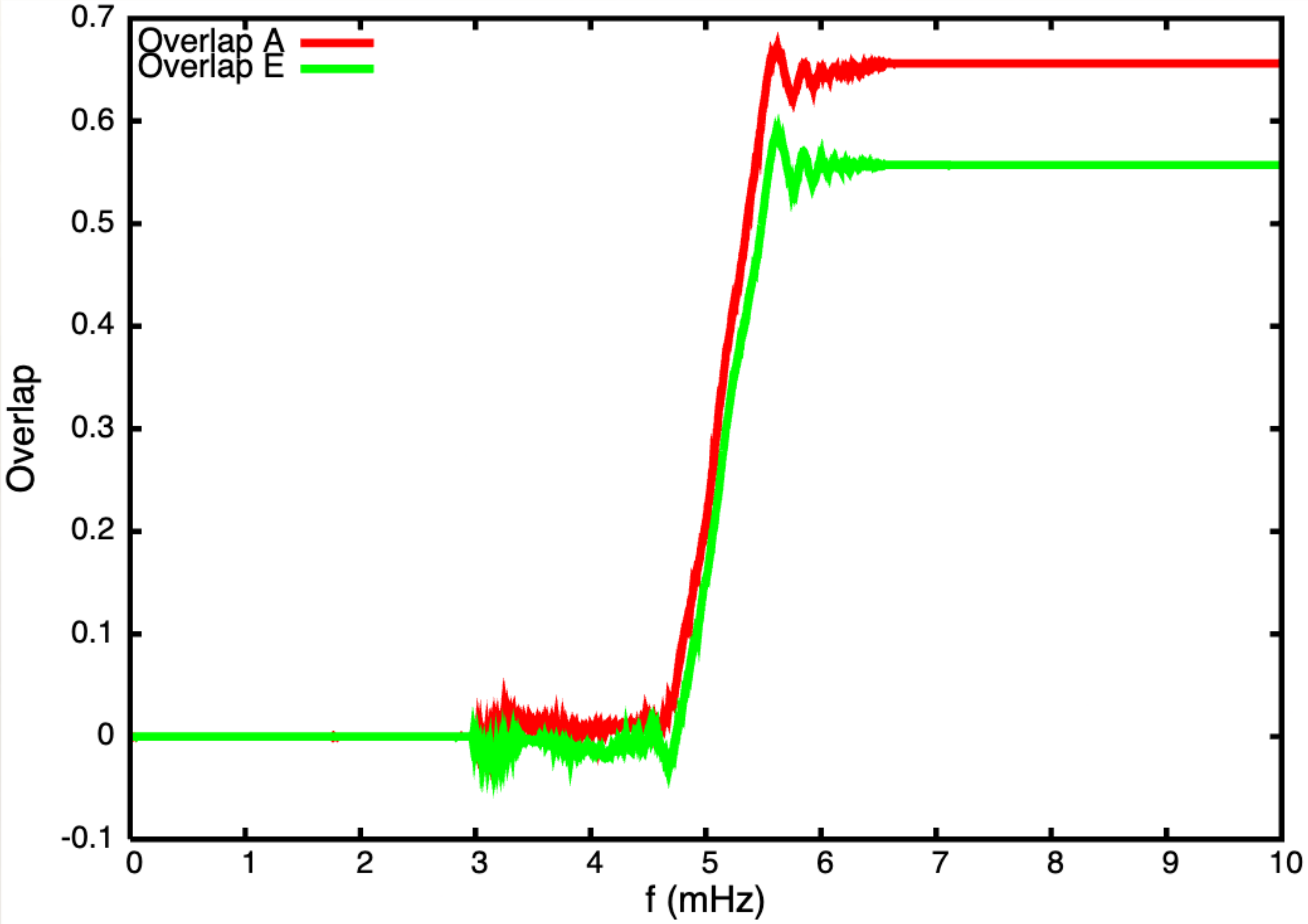


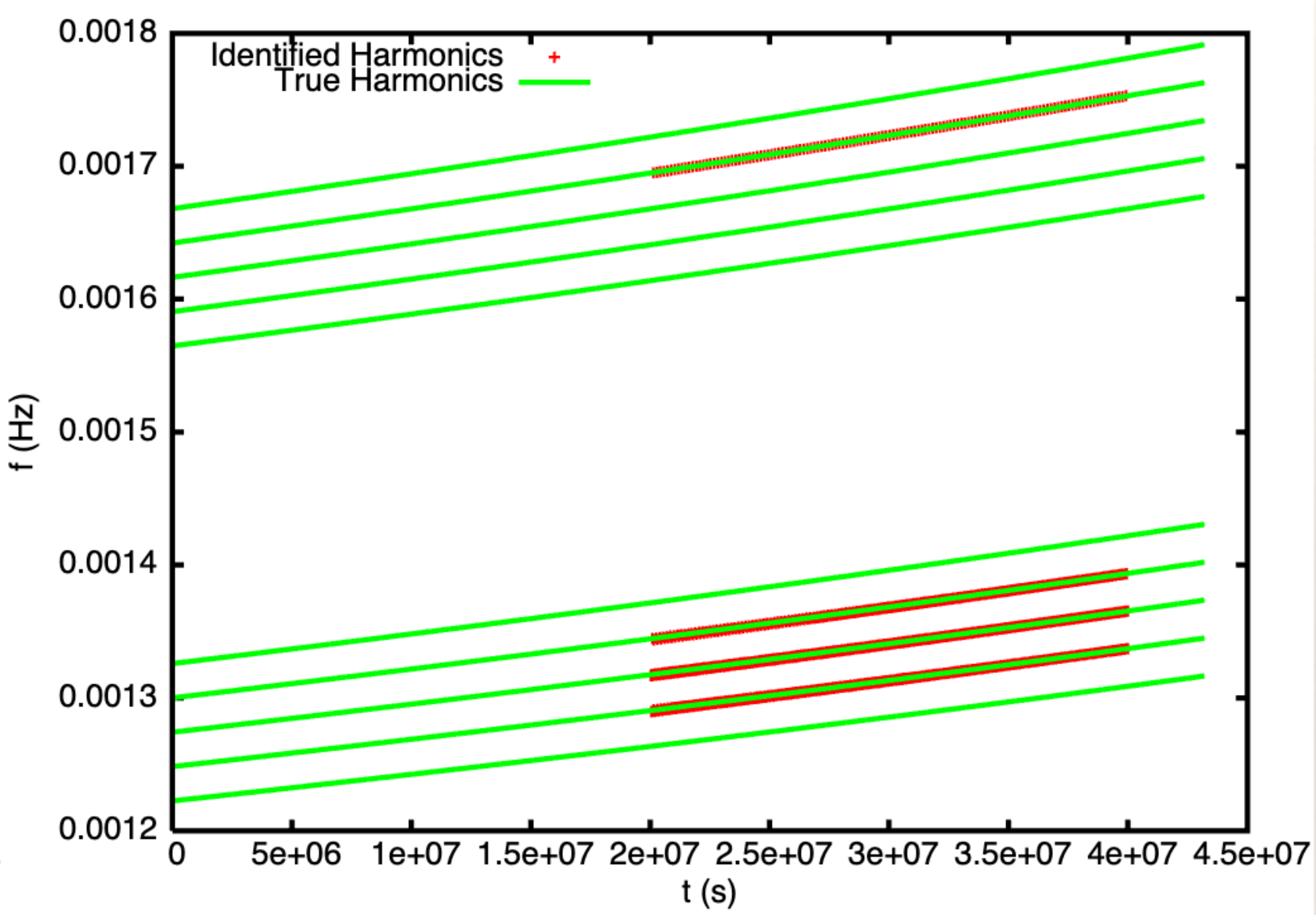
Outstanding challenges: EMRIs



Chua & Cutler (2021)

Outstanding challenges: EMRIs





- Secondaries arise from matching a subset of harmonics for a certain period of time.

Outstanding challenges: EMRIs

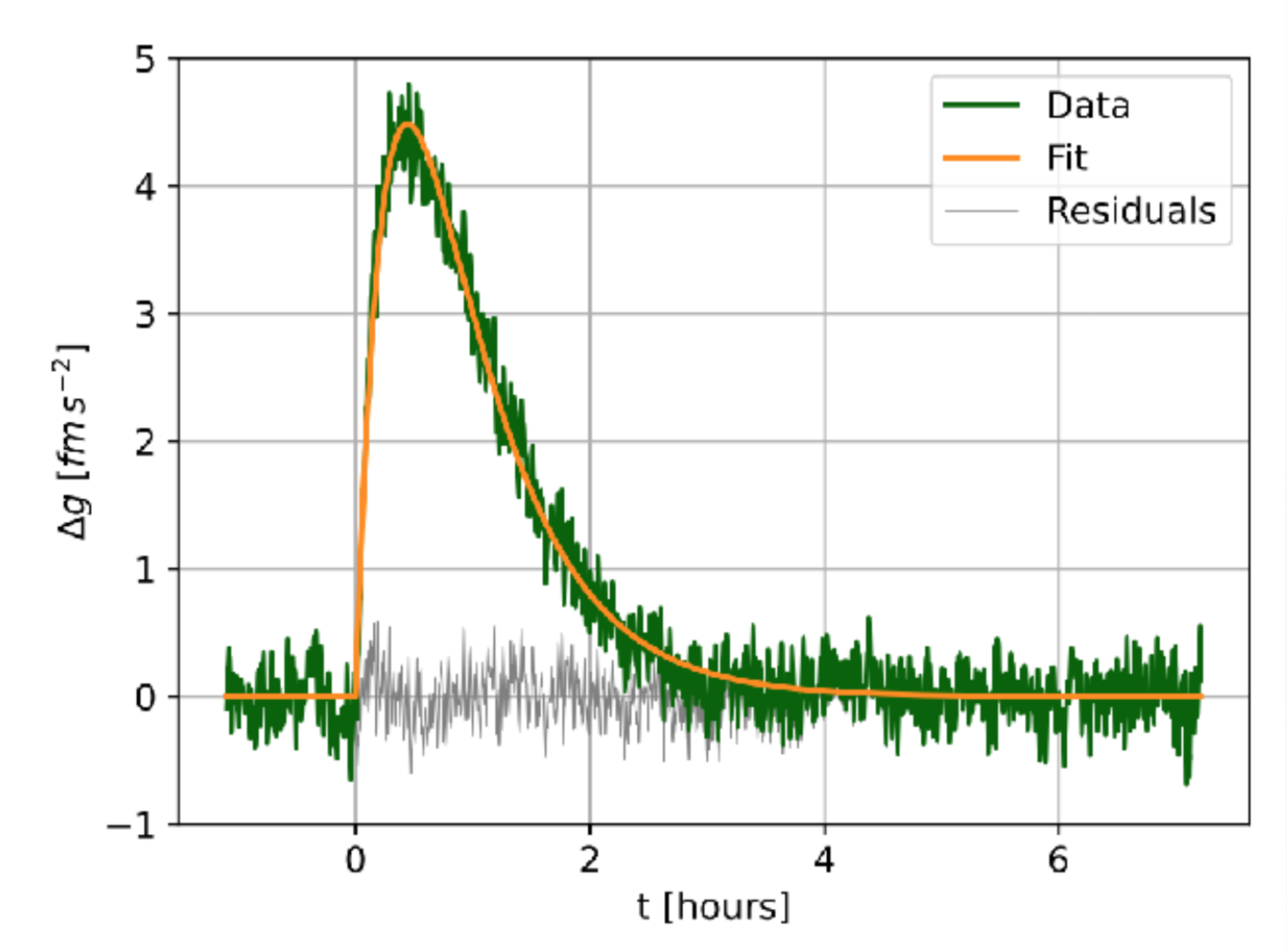
- Not completely hopeless! Ability to find EMRIs was demonstrated in previous MLDCs, but assumptions were highly simplified.

type[1]	ν (mHz)	$\mu/M_\odot$	$M/M_\odot$	e_0	θ_S	φ_S	λ	a/M^2	SNR
True	0.1920421	10.296	9517952	0.21438	1.018	4.910	0.4394	0.69816	120.5
Found	0.1920437	10.288	9520796	0.21411	1.027	4.932	0.4384	0.69823	118.1
True	0.34227777	9.771	5215577	0.20791	1.211	4.6826	1.4358	0.63796	132.9
Found	0.34227742	9.769	5214091	0.20818	1.172	4.6822	1.4364	0.63804	132.8
True	0.3425731	9.697	5219668	0.19927	0.589	0.710	0.9282	0.53326	79.5
Found	0.3425712	9.694	5216925	0.19979	0.573	0.713	0.9298	0.53337	79.7
True	0.8514396	10.105	955795	0.45058	2.551	0.979	1.6707	0.62514	101.6
Found	0.8514390	10.106	955544	0.45053	2.565	1.012	1.6719	0.62534	96.0
True	0.8321840	9.790	1033413	0.42691	2.680	1.088	2.3196	0.65829	55.3
Found	0.8321846	9.787	1034208	0.42701	2.687	1.053	2.3153	0.65770	55.6
Blind									
True	0.1674472	10.131	10397935	0.25240	2.985	4.894	1.2056	0.65101	52.0
Found	0.1674462	10.111	10375301	0.25419	3.023	4.857	1.2097	0.65148	51.7
True	0.9997627	9.7478	975650	0.360970	1.453	4.95326	0.5110	0.65005	122.9
Found	0.9997626	9.7479	975610	0.360966	1.422	4.95339	0.5113	0.65007	116.0

Babak, JG & Porter (2009)

Outstanding challenges: glitches

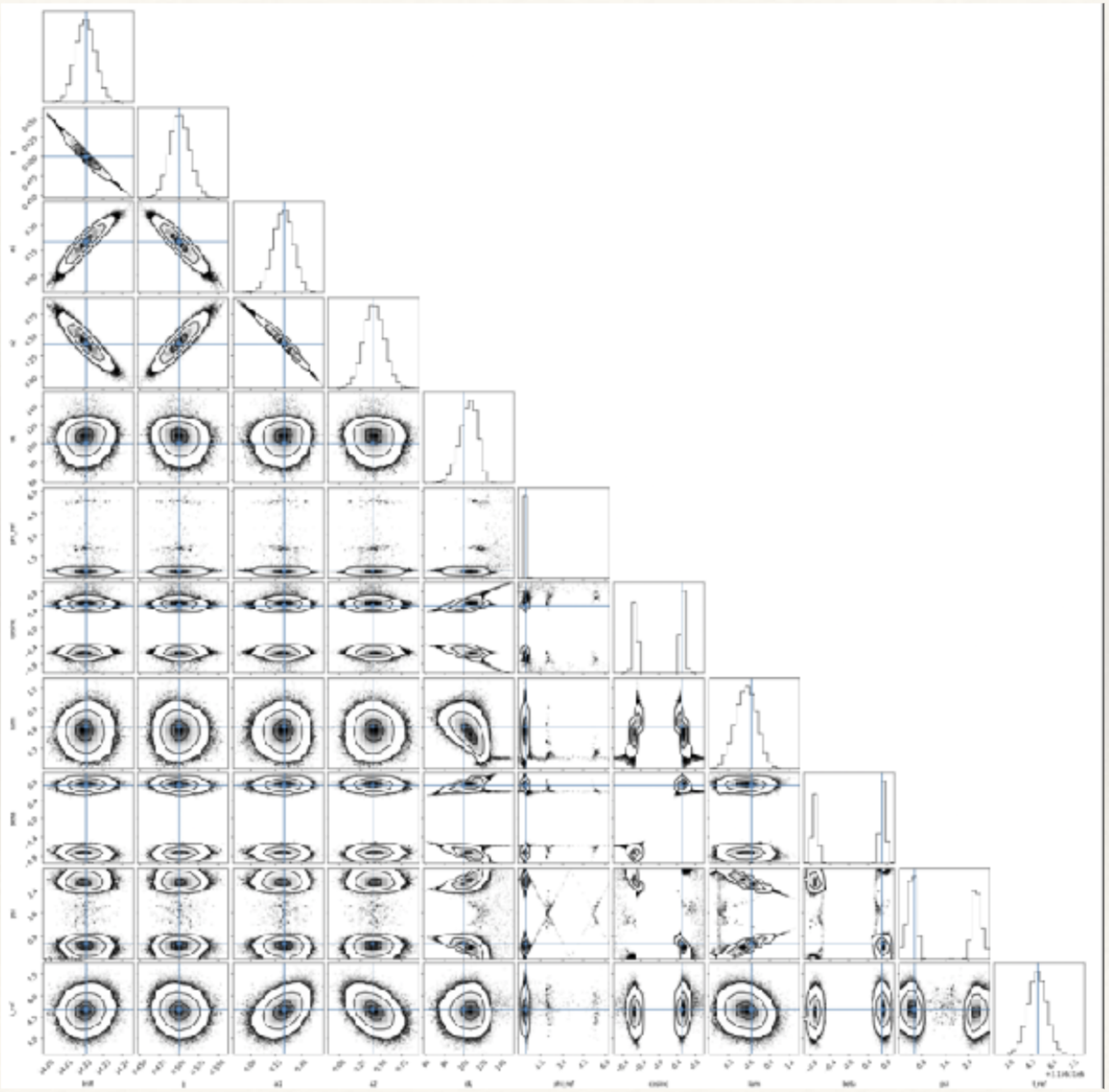
- LISA Pathfinder observed glitches at a rate of 1 / day. Expect glitches in LISA too.
- Pathfinder glitches well described by a single exponential



$$h_1(t) = \frac{\Delta v}{\tau^2}\, t'\, e^{-t'/\tau}\, \Theta(t'), \quad t' = t - t_0$$

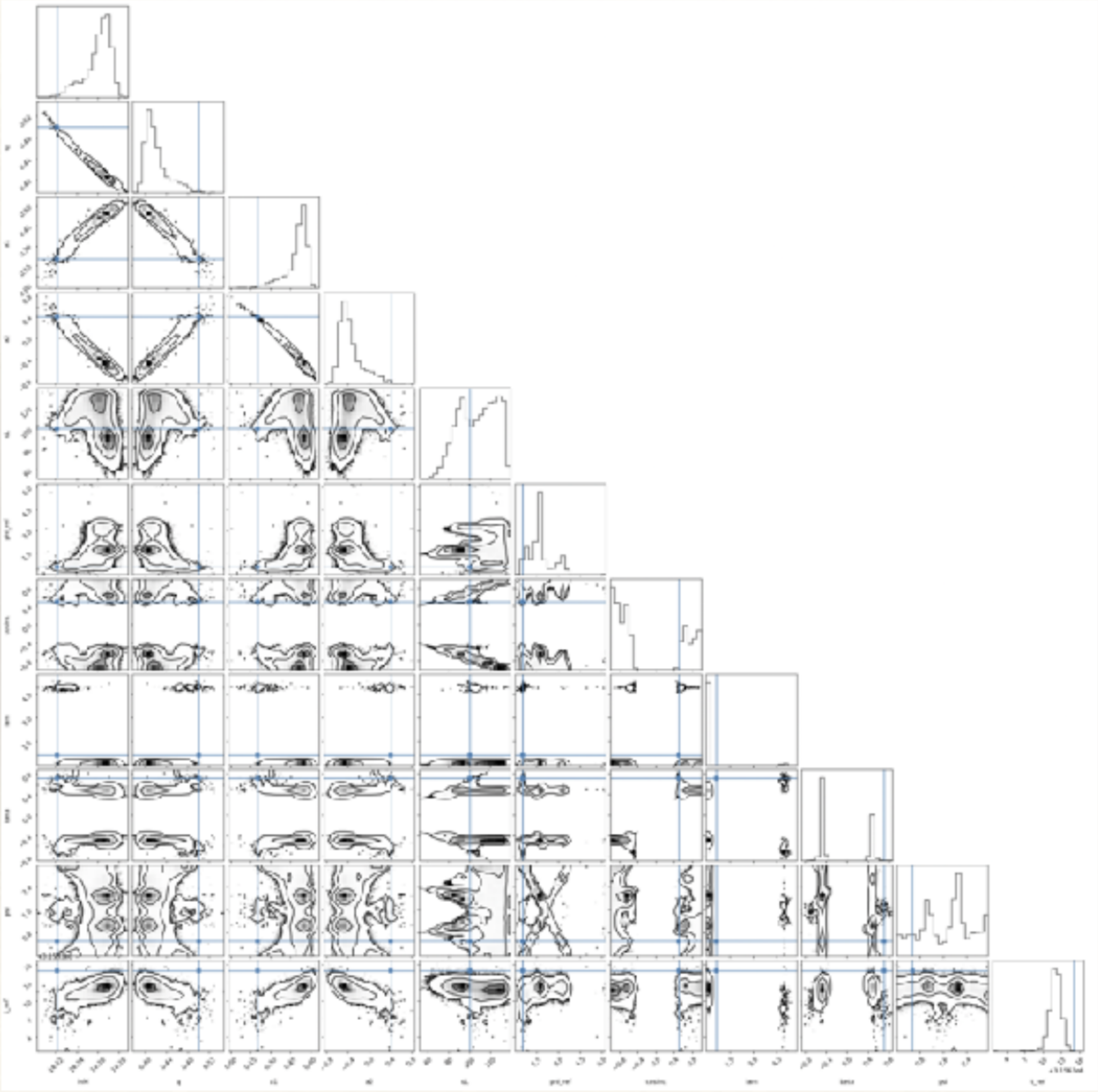
Outstanding challenges: glitches

- Glitches don't look much like GW signals.
- No impact on PE if glitch occurs during inspiral.



Outstanding challenges: glitches

- If glitch overlaps merger, can get biases.
- Avoid biases by fitting for glitch simultaneously with signal parameters.
- Need reliable glitch model.
- Model error could mimic environmental effect.
- But, glitches arise on spacecraft. So, at population level, glitches follow a different distribution.



Outstanding challenges: gaps

- Many possible causes of gaps in the LISA data stream, of both known and unknown origin.

Gap type	Frequency	Duration	Total loss (hr/yr)
Antenna repointing	every 2 weeks	3.3h	1%
PAAM angle adjust	3 per day	100s	0.3%
TM stray pot. est.	2/yr	1 day	0.56%
TTL coupling est.	4/yr	2 days	2.22%
Unplanned: platform	3/yr	2.5 days	2%
Unplanned: payload	4/yr	2.75 days	3%
Unplanned: micro-meteorites	30/yr	1 day	8%

Outstanding challenges: gaps

- Various approaches to dealing with gaps: gap filling, noise filtering, time-frequency analysis etc. Results depend critically on assumptions about noise behaviour across gap.
- Treating gap as **missing data**

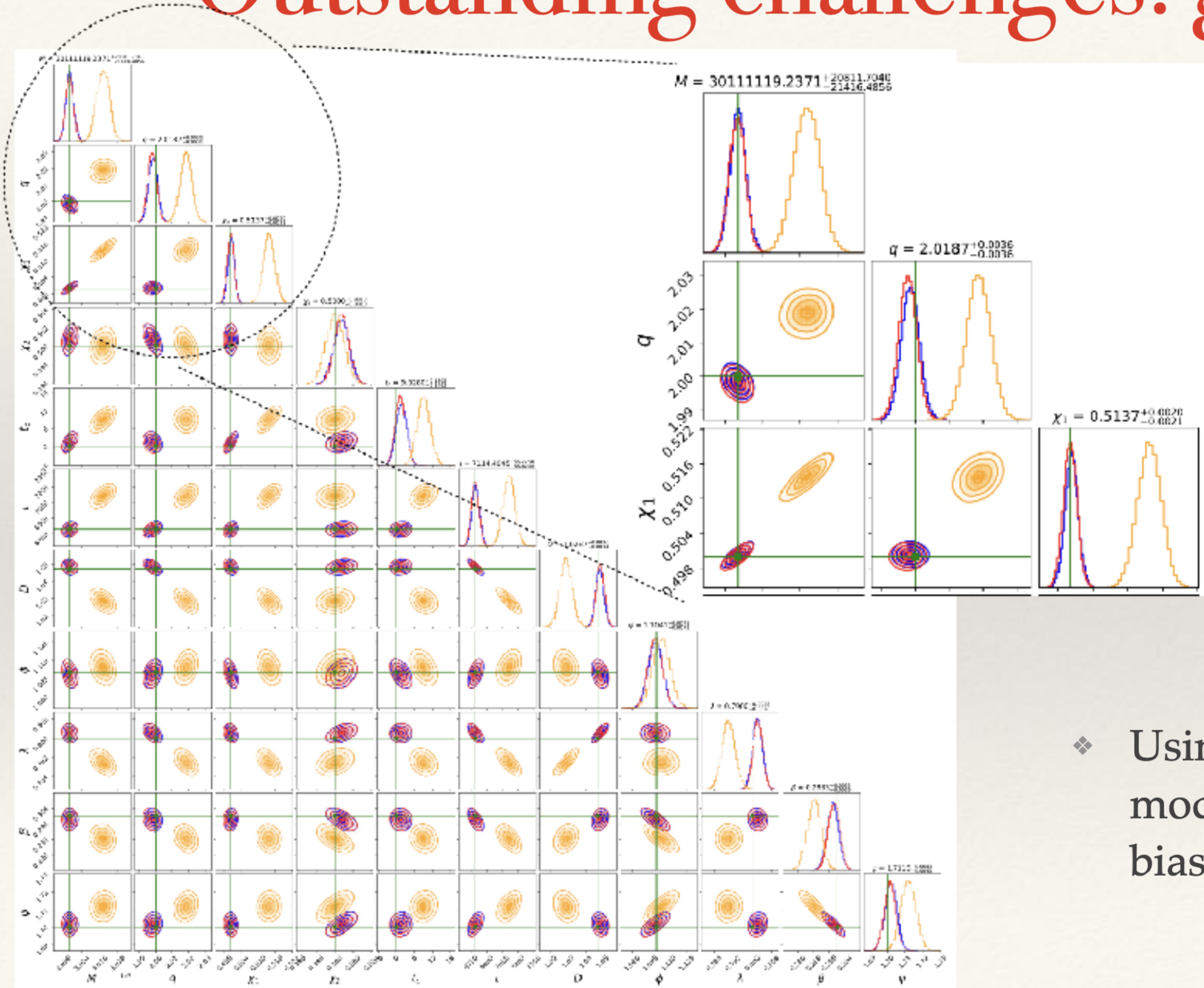
$$D(t) = w(t)h(t;\boldsymbol{\theta}) + w(t)n(t) = H(t;\boldsymbol{\theta}) + N(t)$$

$$\log p(\boldsymbol{D}|\boldsymbol{\theta},\Sigma_N) \propto -\frac{1}{2}(D(t) - H(t;\boldsymbol{\theta})|D(t) - H(t;\boldsymbol{\theta}))_{\Sigma_N} = -(\hat{\boldsymbol{D}} - \hat{\boldsymbol{H}})^{\dagger}\Sigma_N^{-1}(\hat{\boldsymbol{D}} - \hat{\boldsymbol{H}})$$

$$(\Sigma_N)_{ij} \approx \frac{\Delta f}{2} \sum_{p=0}^{\lfloor N/2+1 \rfloor} \hat{w}^{\star}(f_i - v_p)\hat{w}(f_j - v_p)S_n(v_p)$$

- Treating noise as **independent** in each between-gap segment: likelihood is product of likelihoods for each segment.

Outstanding challenges: gaps



- Using the wrong model leads to biases

Outstanding challenges: gaps

- Average bias is zero

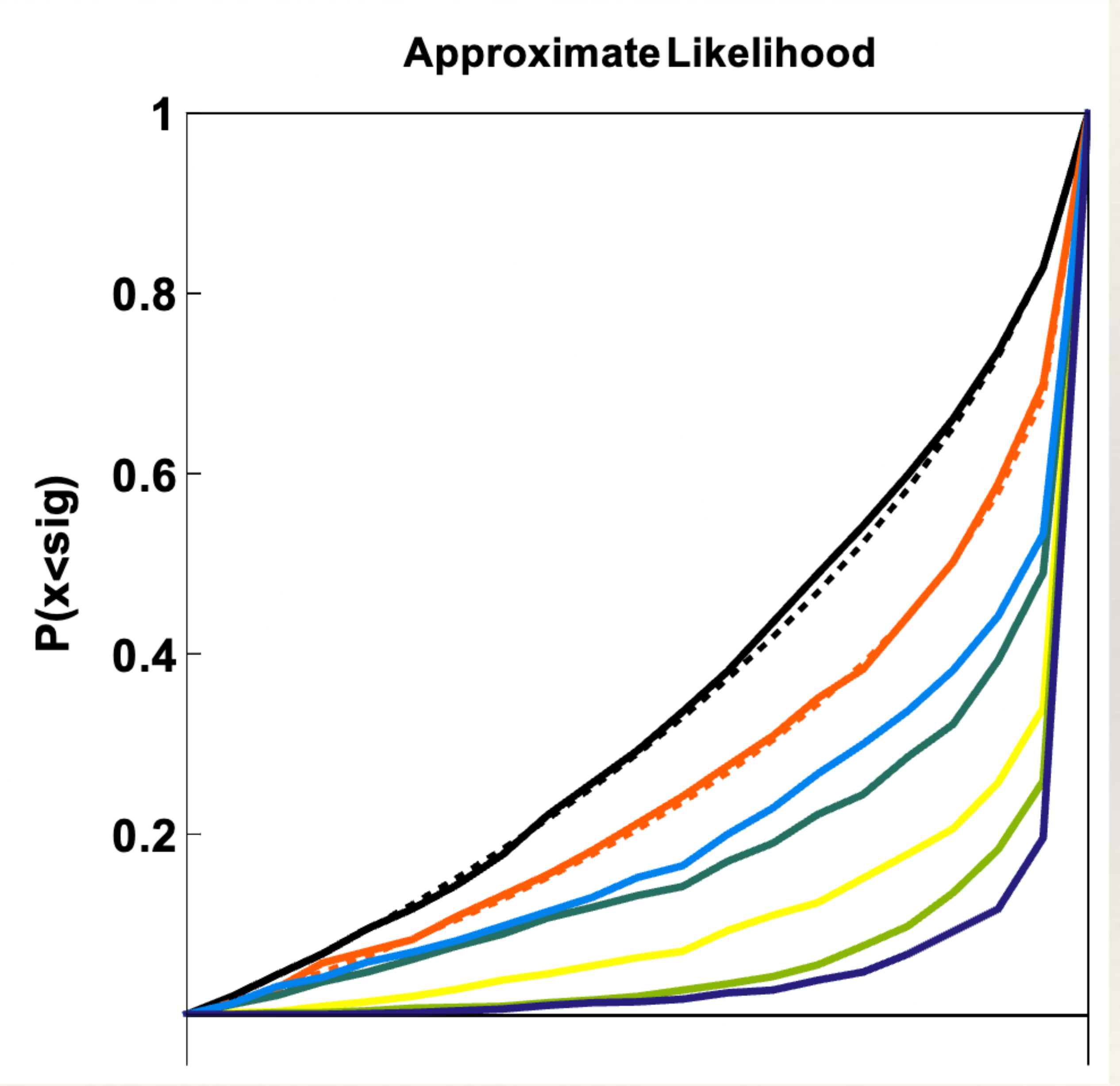
$$\widehat{\Delta\theta^i_{\rm bf}} = (\Gamma^{-1})^{ik}_{\Sigma}(\partial_k H|N)_{\Sigma}$$

$$\Gamma^{(\Sigma)}_{ij} = (\partial_i H|\partial_j H)_{\Sigma} \approx 2\partial_i \hat{\boldsymbol{H}}^{\dagger}\Sigma^{-1}\partial_j \hat{\boldsymbol{H}}$$

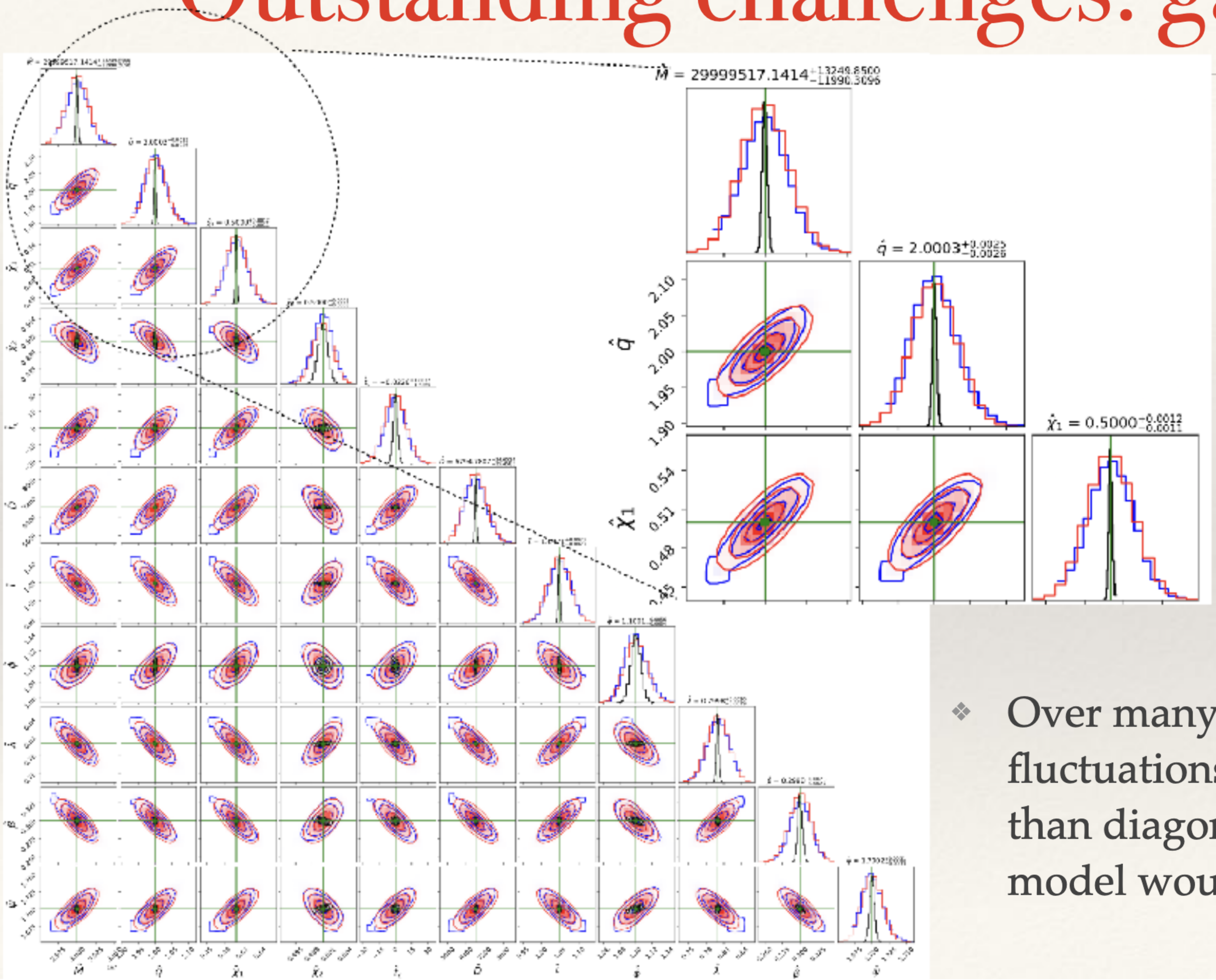
- But, variance inconsistent with likelihood

$$\mathbb{E}[\widehat{\Delta\theta}^i_{\rm bf}\widehat{\Delta\theta}^j_{\rm bf}] = 2(\Gamma^{-1})^{ik}_{(\Sigma)} \cdot \mathrm{Re}\left(\boldsymbol{\gamma}_k^{\dagger}\Sigma_N\boldsymbol{\gamma}_p\right)(\Gamma^{-1})^{pj}_{(\Sigma)}$$
$$\boldsymbol{\gamma}_k = \Sigma^{-1}\partial_k\hat{\boldsymbol{H}}^{\dagger}\,.$$

- See “sag” in a p-p plot for analysis of many events.



Outstanding challenges: gaps



- Over many realisations. fluctuations much greater than diagonal-noise model would predict.

Outstanding challenges: source confusion

- Presence of other sources in the data impacts parameter estimates for sources of interest.

- For resolved sources, assess impact using joint Fisher matrix

$$\Gamma = \begin{pmatrix} \Gamma^{(1)} & \Gamma^{\rm mix} \\ (\Gamma^{\rm mix})^T & \Gamma^{(2)} \end{pmatrix}; \quad \Gamma^{-1} = \begin{pmatrix} \Gamma^{-1}_{11} & \Gamma^{-1}_{12} \\ (\Gamma^{-1}_{12})^T & \Gamma^{-1}_{22} \end{pmatrix}$$

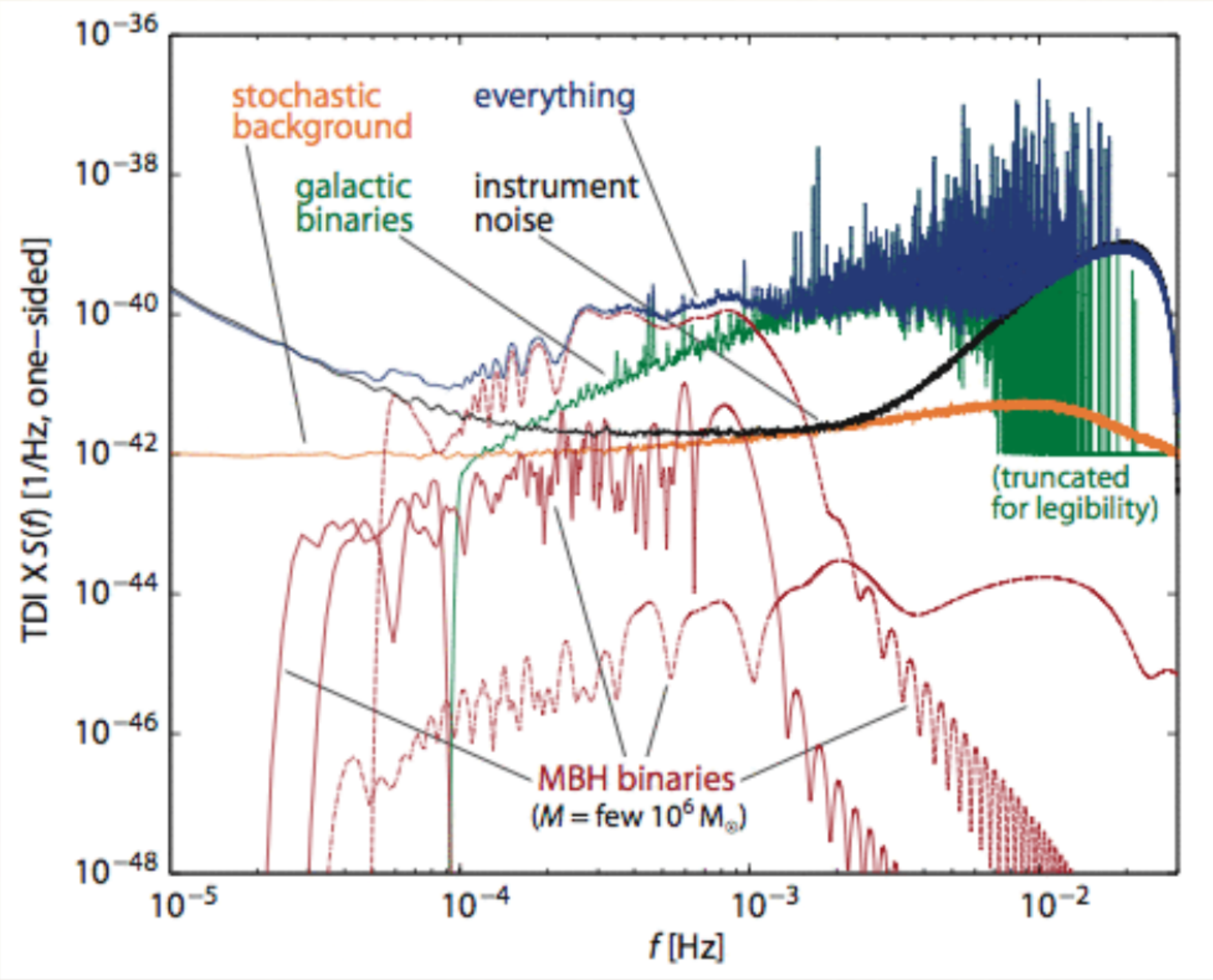
$$\Gamma^{-1}_{11} = \left(\Gamma^{(1)} - \Gamma^{\rm mix}(\Gamma^{(2)})^{-1}(\Gamma^{\rm mix})^T\right)^{-1}$$

$$\Gamma^{-1}_{22} = \left(\Gamma^{(2)} - (\Gamma^{\rm mix})^T(\Gamma^{(1)})^{-1}\Gamma^{\rm mix}\right)^{-1}$$

$$\Gamma^{-1}_{12} = -\Gamma^{-1}_{11}\,\Gamma^{\rm mix}(\Gamma^{(2)})^{-1}\,.$$

- Assuming near-orthogonality

$$\Gamma^{-1}_{11} \approx (\Gamma^{(1)})^{-1} + (\Gamma^{(1)})^{-1}\Gamma^{\rm mix}(\Gamma^{(2)})^{-1}(\Gamma^{\rm mix})^T(\Gamma^{(1)})^{-1}$$



- Can interpret as noise from residuals

$$\Delta\theta^{(1),i}_{\rm sys} = (\Gamma^{(1)})^{-1}_{ij}(\partial_j h^{(1)}_m | \boldsymbol{h}^{(2)})$$

$$\langle \Delta\theta^{(1),i}_{\rm sys}\Delta\theta^{(1),j}_{\rm sys}\rangle = (\Gamma^{(1)})^{-1}_{ik}(\partial_k h^{(1)}_m|\partial_l \boldsymbol{h}^{(2)})\langle\Delta\theta^l_2\Delta\theta^m_2\rangle$$
$$(\partial_n h^{(1)}_m|\partial_m \boldsymbol{h}^{(2)})(\Gamma^{(1)})^{-1}_{jm}$$

Outstanding challenges: source confusion

- Unresolved sources also leave a residual in the data

$$\Delta H_{\rm conf}(t;\boldsymbol{\theta}^{(i)}) = \sum_{i=1}^{N} h_e^{(i)}(t;\boldsymbol{\theta}^{(i)})$$

- which can lead to parameter biases

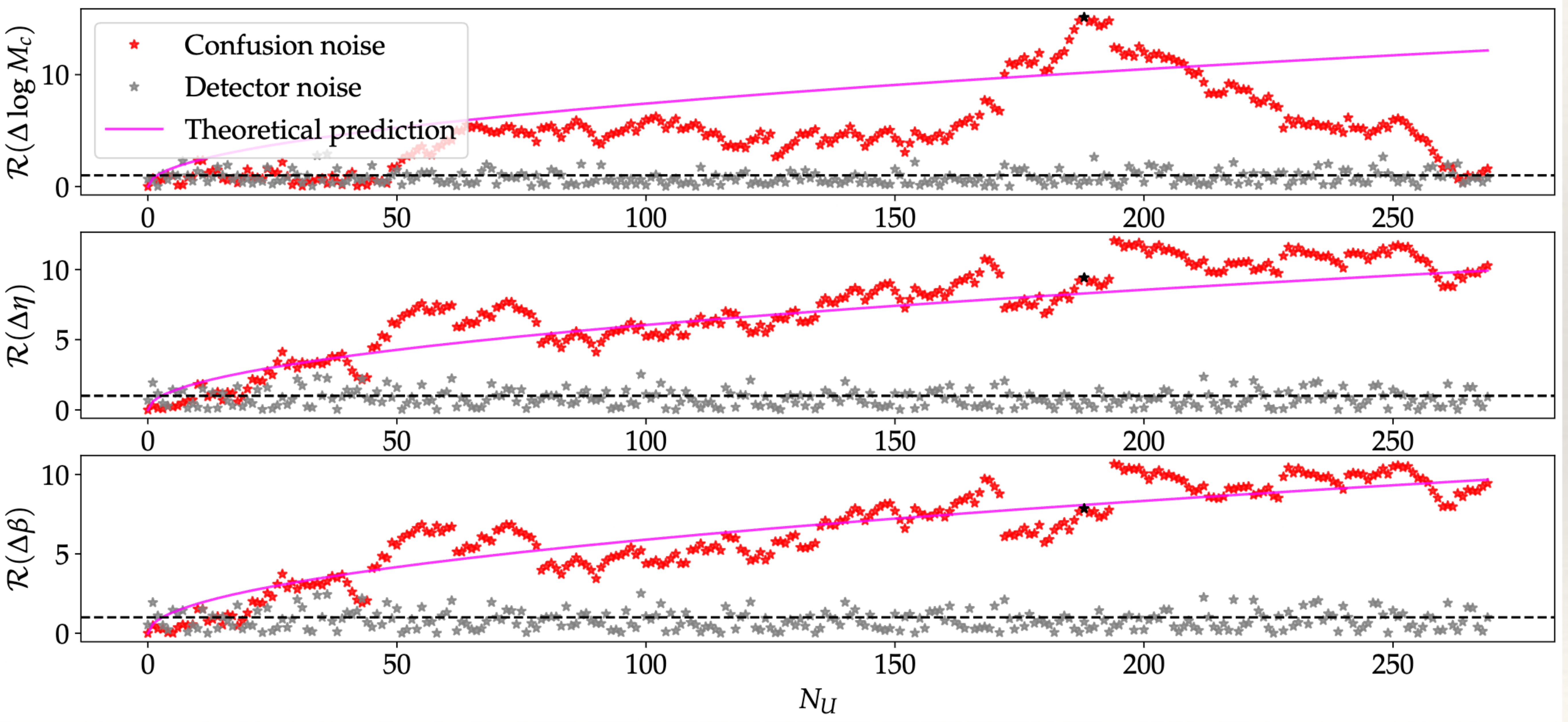
$$\Delta\theta^i_{\rm conf} = (\Gamma^{-1})^{ij}(\partial_j h_m|\Delta H_{\rm conf})$$

- characterise by a mean bias and excess noise variance

$$\mu^i_{\rm conf} = \int (\Gamma^{-1})^{ij}(\partial_j h_m|h_e(\boldsymbol{\theta}_{\rm conf}))\, p_{\rm pop}(\boldsymbol{\theta}_{\rm conf})\, {\rm d}\boldsymbol{\theta}_{\rm conf}\,,$$

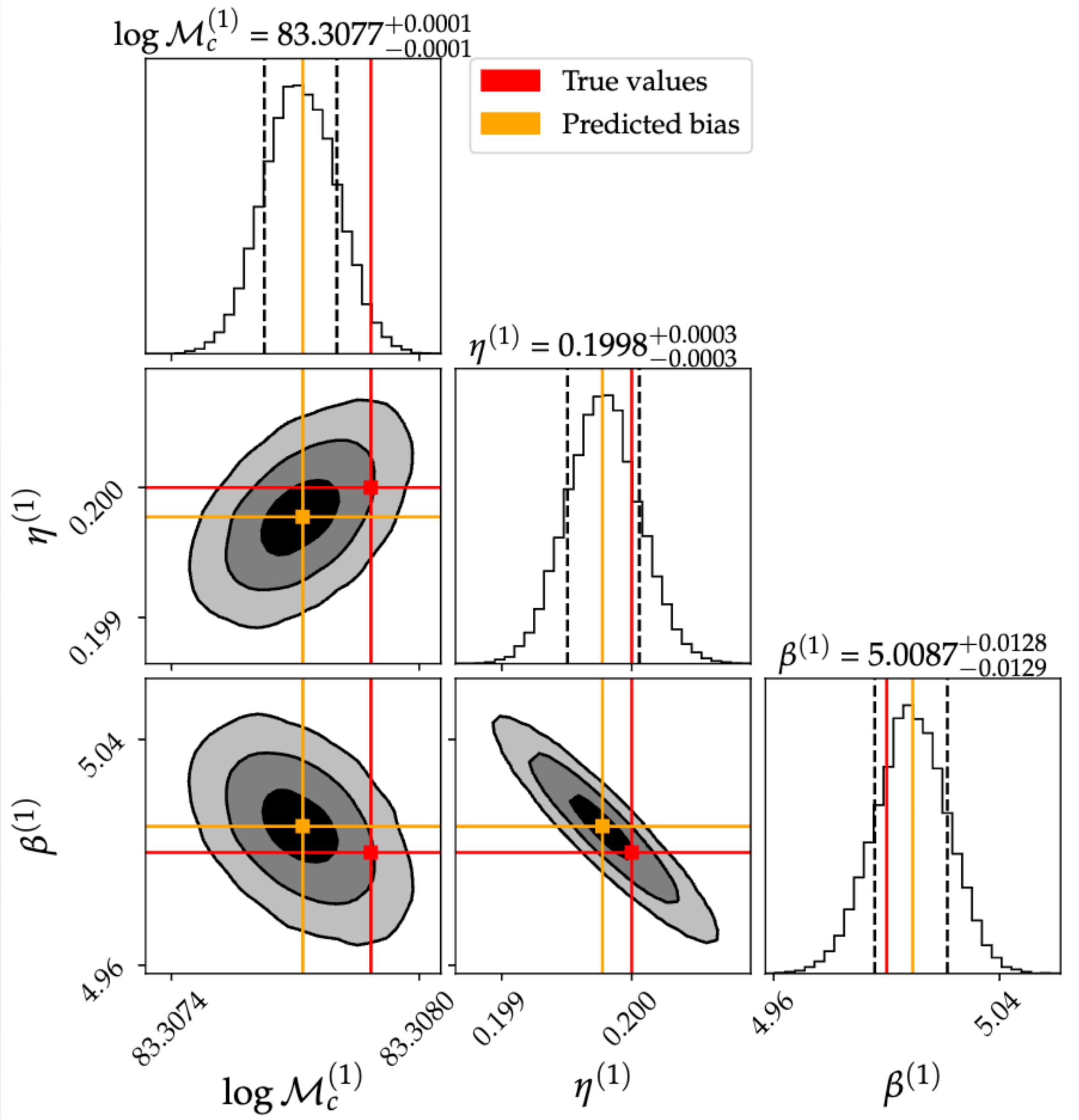
$$\Sigma^{ij}_{\rm conf} = \int (\Gamma^{-1})^{ik}(\partial_k h_m|h_e(\boldsymbol{\theta}_{\rm conf}))\times (\Gamma^{-1})^{jl}(\partial_l h_m|h_e(\boldsymbol{\theta}_{\rm conf}))\, p_{\rm pop}(\boldsymbol{\theta}_{\rm conf})\, {\rm d}\boldsymbol{\theta}_{\rm conf} - \mu^i_{\rm conf}\mu^j_{\rm conf}.$$

Outstanding challenges: source confusion



Outstanding challenges: source confusion

- Get additional errors from mismodelling of sources, e.g., ignoring environmental effects.
- Could limit ability to detect/be misinterpreted as an environmental effect or deviation from GR.
- For example, fitting for a single source of type A in the presence of residuals from a population of type B.

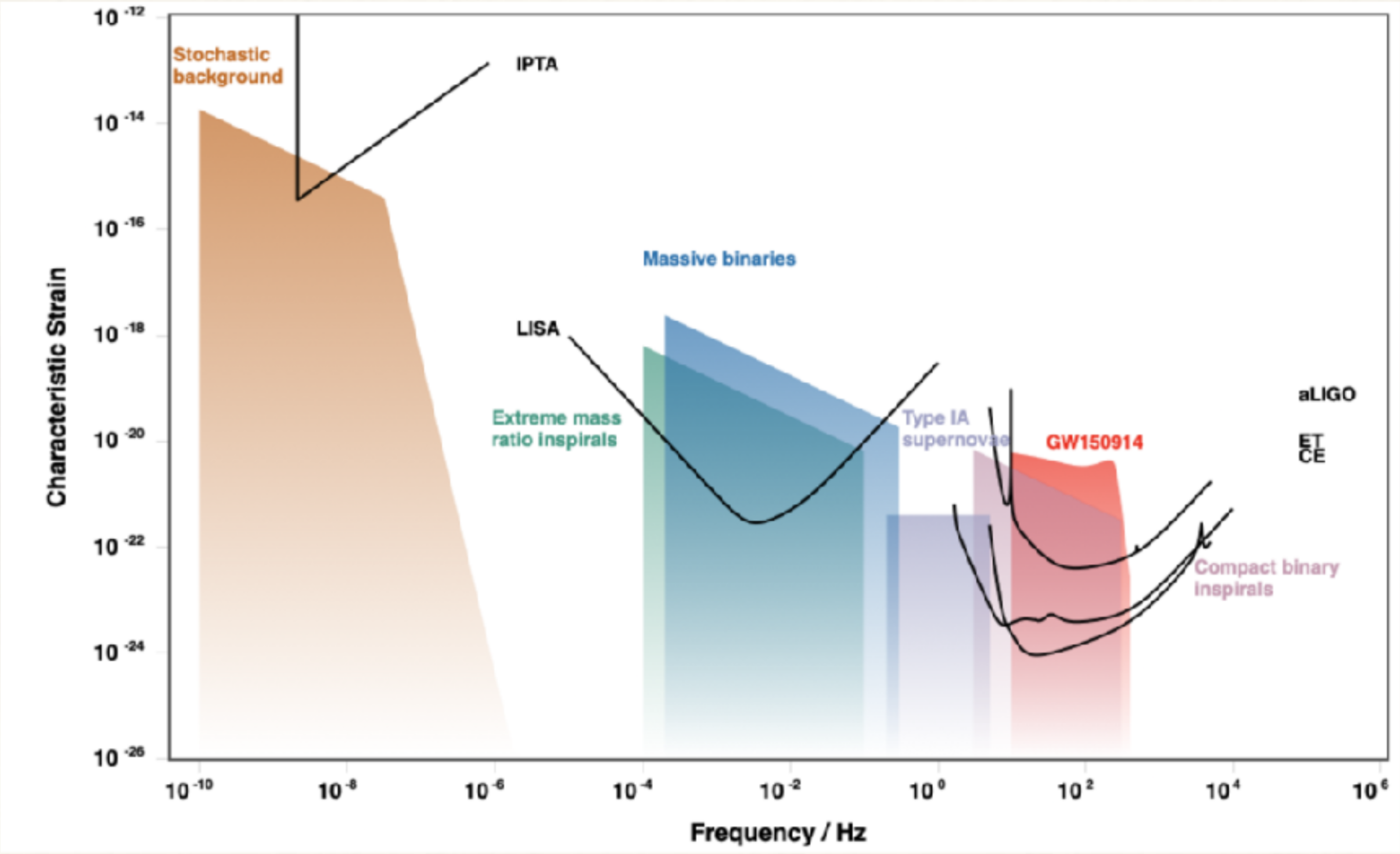


$$\Delta\theta_i^A = -\sum_{a=1}^{N_B} \left(\Gamma^A\right)^{-1}_{ij} \left[(\partial_j h^A | \delta h^B(\boldsymbol{\theta}_a^B)) + \left(\Gamma^{\rm mix}(\boldsymbol{\theta}_a^B)\right)_{jk} \left(\Gamma^B(\boldsymbol{\theta}_a^B)\right)^{-1}_{kl} \sum_{b=1}^{N_B} (\partial_l h^B(\boldsymbol{\theta}_a^B) | \delta h^B(\boldsymbol{\theta}_b^B)) \right]$$

$$\left(\Gamma^{\rm mix}(\boldsymbol{\theta}_a^B)\right)_{ij} = (\partial_i h^A | \partial_j h^B(\boldsymbol{\theta}_a^B)), \quad \Gamma^A_{ij} = (\partial_i h^A | \partial_j h^A), \quad \left(\Gamma^B(\boldsymbol{\theta}_a^B)\right)_{ij} = (\partial_i h^B(\boldsymbol{\theta}_a^B) | \partial_j h^B(\boldsymbol{\theta}_a^B))$$

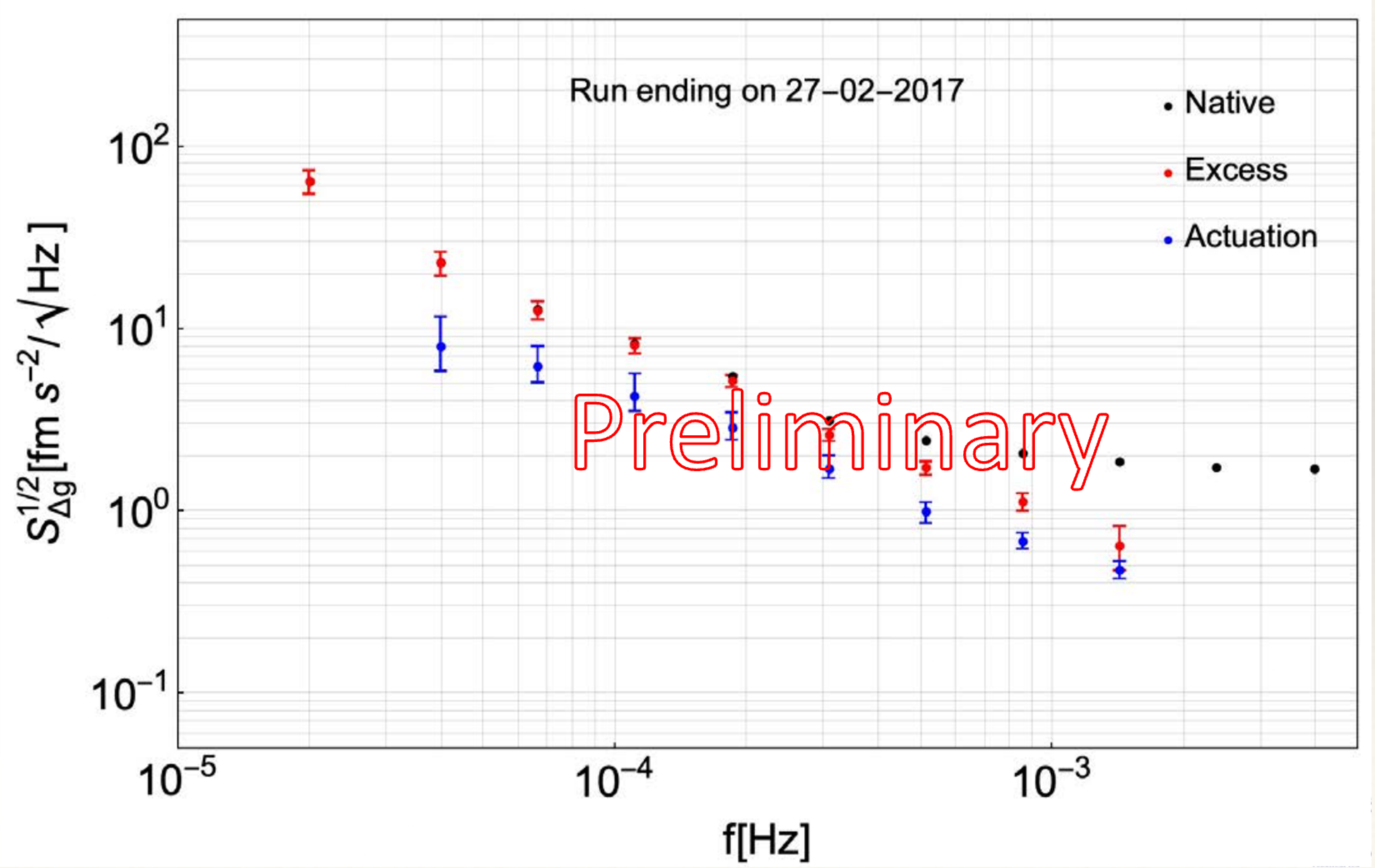
Outstanding challenges: lack of noise knowledge

- Typically assume a known sensitivity when assessing mission performance.



Outstanding challenges: lack of noise knowledge

- Reality is different. In LISA Pathfinder only 25% of total noise power was explained by measured noise sources.



Outstanding challenges: lack of noise knowledge

- Fortunately, noise estimation is independent of signal estimation at leading order. The Fisher matrix is defined in general by

$$\Gamma_{ij} = \mathbb{E}_p\left[\frac{\partial \ln p}{\partial \theta^i}\frac{\partial \ln p}{\partial \theta^j}\right]$$

- The necessary derivatives can be found to be

Even

$$\frac{\partial \ln p}{\partial \theta^i} = \sum_{k=1}^{n_f} \mathrm{Re}\left[\frac{\boxed{(\tilde{d}_k - \tilde{h}_k(\vec{\theta}))}}{S_h(f_k|\vec{\lambda})}\frac{\partial \tilde{h}_k^*}{\partial \theta^i}\mathrm{d}f\right] \qquad \frac{\partial \ln p}{\partial \lambda^i} = \sum_{k=1}^{n_f}\left[\frac{\boxed{|\tilde{d}_k - \tilde{h}_k(\vec{\theta})|^2}}{2S_h^2(f_k|\vec{\lambda})}\mathrm{d}f - \frac{\boxed{1}}{S_h(f_k|\vec{\lambda})}\right]\frac{\partial S_h(f_k)}{\partial \lambda^i}$$

Odd

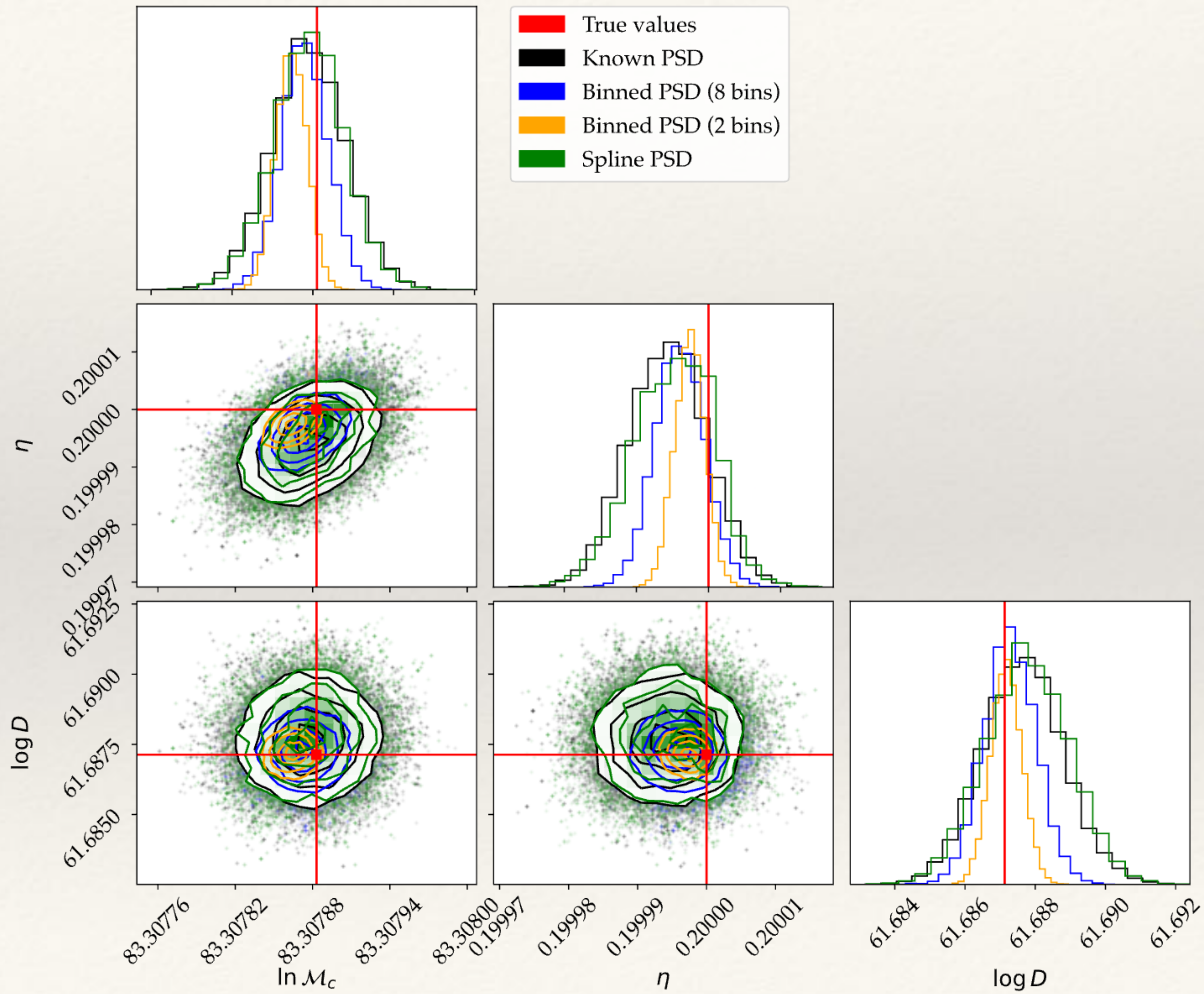
- from which we deduce

$$\mathbb{E}_{\mathcal{L}}\left[\frac{\partial l}{\partial \theta^i}\frac{\partial l}{\partial \lambda^j}\right] = 0$$

- Conclude that estimation of noise and signal are independent - no requirement this way. But what if the noise is highly uncertain?

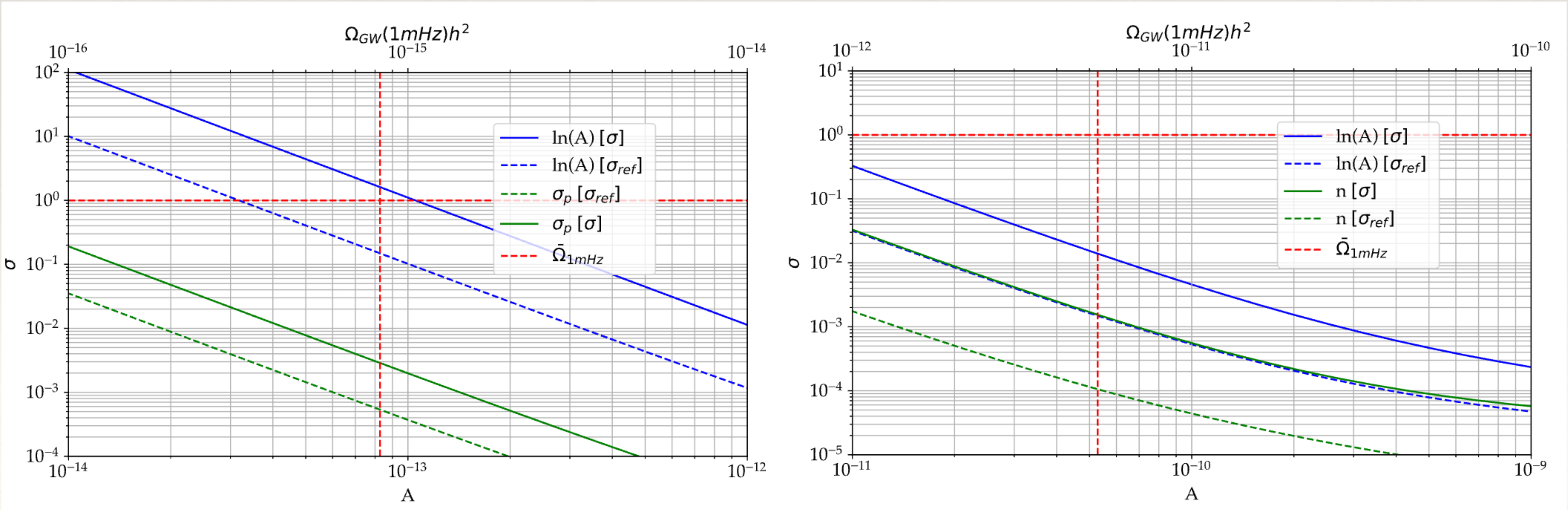
Outstanding challenges: lack of noise knowledge

- Confirmed by MCMC.
- Also seen in analysis of LDC data sets.



Outstanding challenges: lack of noise knowledge

- More problematic for backgrounds -> need amplitude to be a factor of a few bigger for confident detection.
- Could impact our ability to make statistical statements about environmental effects at the population level.
- Also has implications for residual tests of GR.



Take away lessons

- Starting point for understanding observability of environmental effects should still be standard tools.

- **Fisher Matrix:**

$$\Gamma_{ij} = \left(\frac{\partial h}{\partial \lambda_i} \middle| \frac{\partial h}{\partial \lambda_j} \right) \qquad (a|b) = \int_{-\infty}^{\infty} \frac{\tilde{a}^*(f)\tilde{b}(f) + \tilde{a}(f)\tilde{b}^*(f)}{S_n(f)} \, \mathrm{d}f$$

- **Systematic error < statistical error :**

$$\Delta\theta^i_{\rm sys} = (\Gamma^{-1})_{ij}(\mathbf{h}_e - \mathbf{h}_m | \partial_j \mathbf{h}_m) \lesssim \sqrt{(\Gamma^{-1})_{ii}} = \Delta\theta^i_{\rm stat}$$

- **[Lindblom criterion:**

$$(h_e - h_m | h_e - h_m) < 1 \qquad (h_e - h_m | h_e - h_m) < 2\rho^2 \epsilon_{\rm max}$$

- **Total phase error:** $\delta\phi < 1/\rho$]

Take away lessons

- If an effect looks detectable under these idealised assumptions, then should do things more carefully
  - Include other signals in Fisher matrix -> account for confusion, instrumental glitches, (PSD uncertainties) etc.
  - Modifying overlap to allow for non-diagonal covariance, as in the case of gaps

$$(a|b) = (\mathbf{a}^T \ \mathbf{a}^\dagger)\Sigma^{-1} \begin{pmatrix} \mathbf{b} \\ \mathbf{b}^* \end{pmatrix}$$

  - We have LDC data sets including many sources (and soon including instrumental artefacts). Use these to inject sources with environmental effects and assess measurability.

Summary

- LISA data analysis is complicated, with inference of all sources highly coupled.
- Rapid progress is being made, driven by the LDCs. Several groups have working prototype global-fit pipelines.
- Various complications still exist
  - EMRI search problem not yet solved
  - Need to account for and mitigate the impact of instrumental artefacts, including gaps and glitches.
  - Need to simultaneously fit unknown instrumental noise component of the data stream.
  - Confusion from resolved, or unresolved, other sources and residuals from mismodelling of the signal templates will limit PE capabilities.
- All of these complications will impact our ability to measure/constrain environmental effects, and this needs to be properly assessed.