LISA DATA ANALYSIS: FROM MEASUREMENTS TO DISCOVERIES

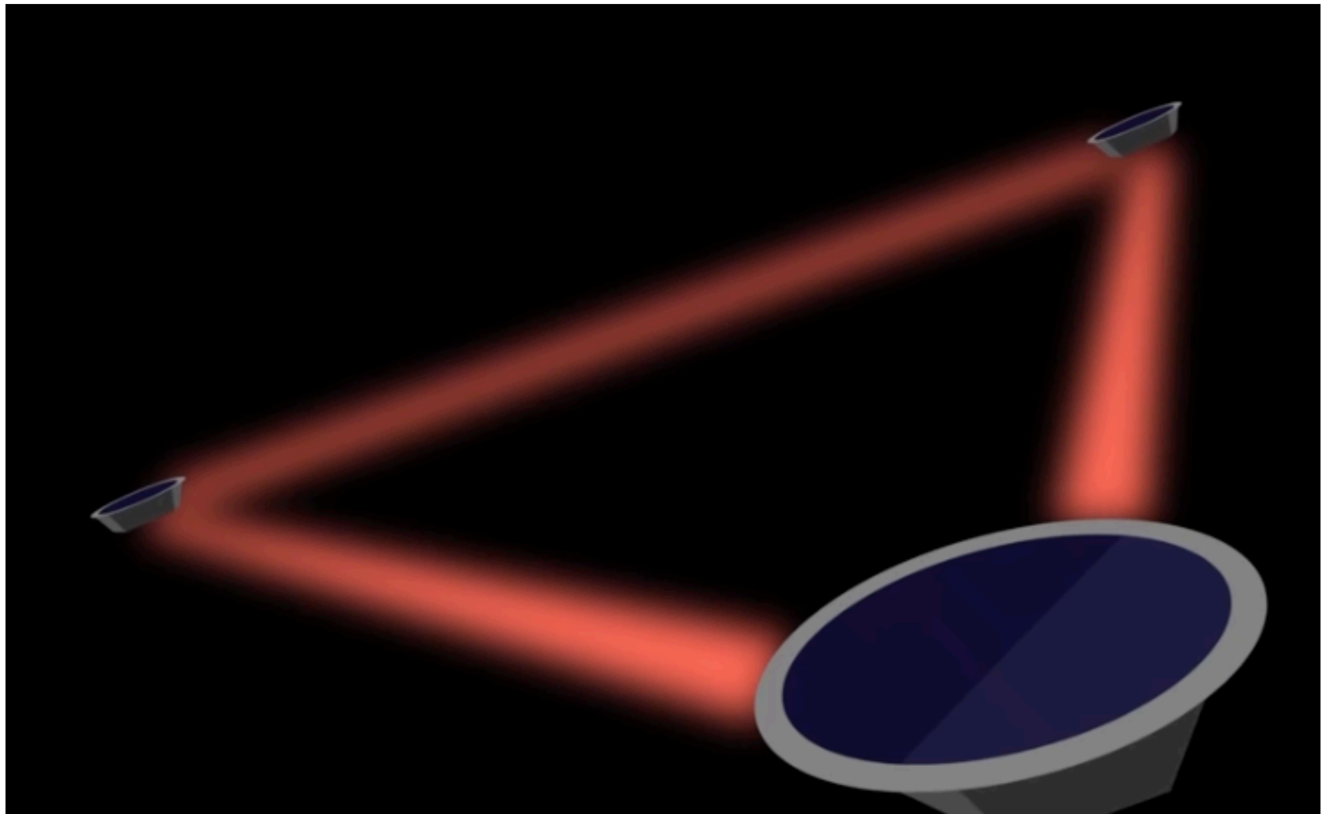
Quentin Baghi, CEA Paris-Saclay

Tuesday, June 6th, 2023

10th LISA Cosmology Working Group Workshop

Layout

1. Challenges of LISA data analysis
2. The global fit strategy - state of the art
3. Looking for the unknown
4. Towards the future



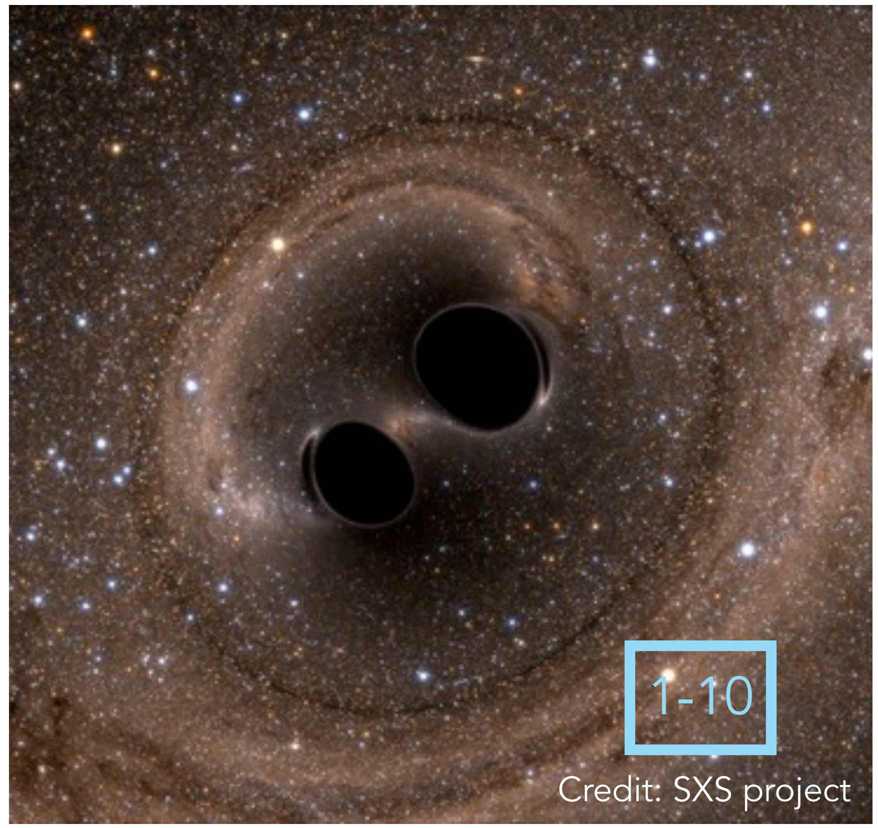
Credit: D. Walliman, Domain of Science

GW sources in LISA band

What makes gravitational sounds in the millihertz band?

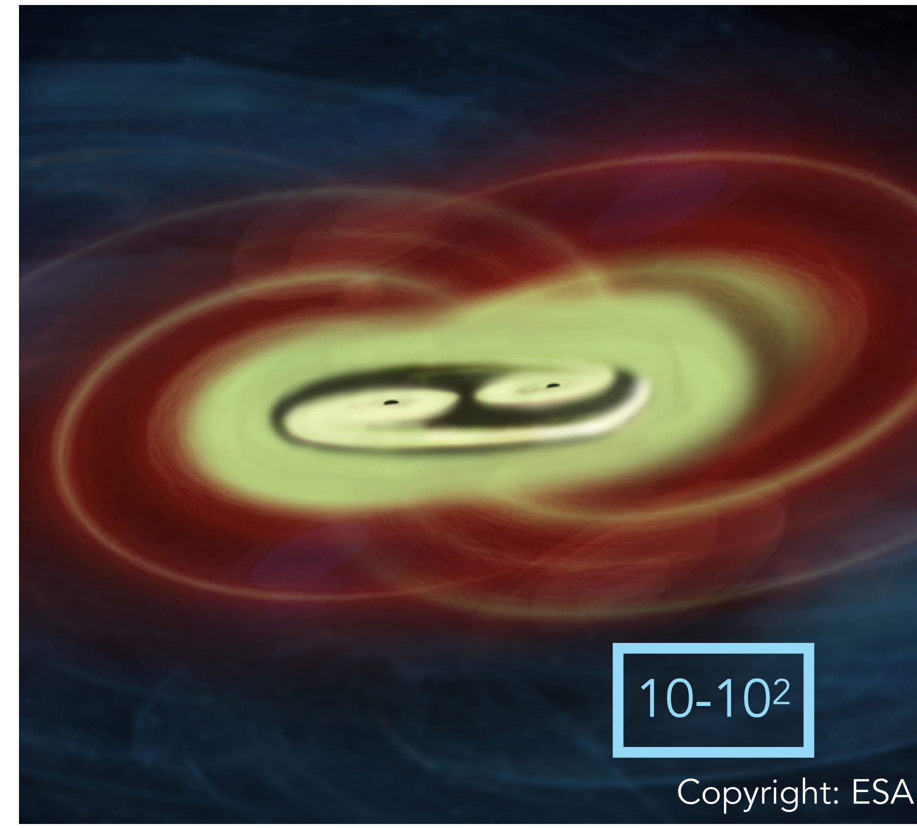
Copyright: Jorge Lugo

In our galaxy: pairs of orbiting white dwarfs



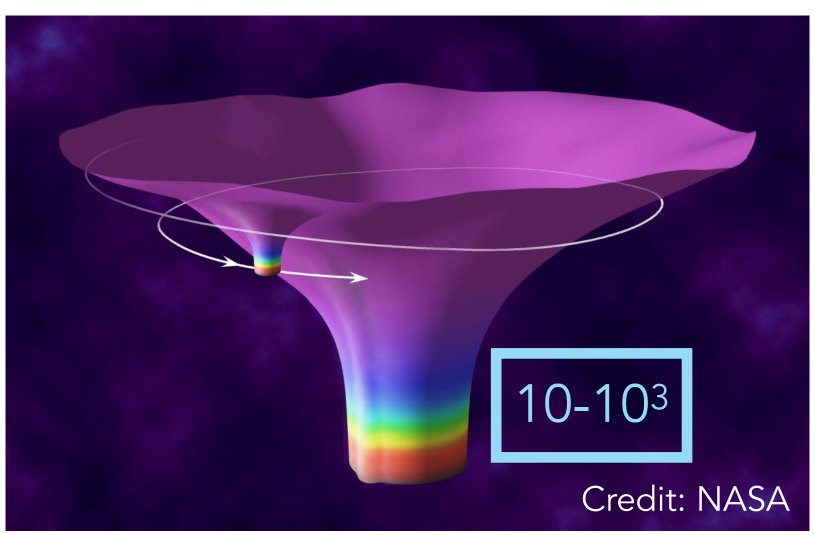
Credit: SXS project

One billion light-years away: collision of stellar-mass black holes



Copyright: ESA

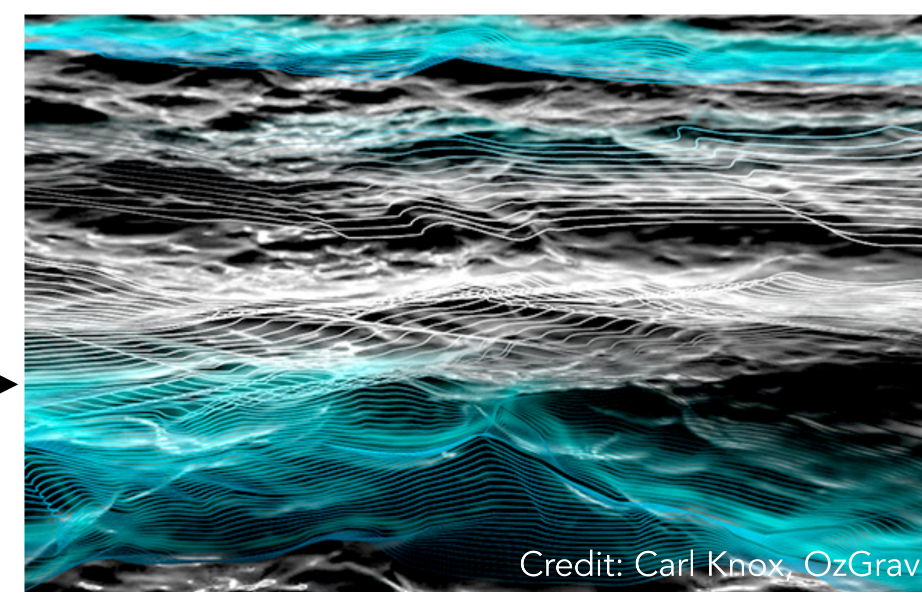
Possibly farther away: merging supermassive black holes



Credit: NASA

◀ Extreme-mass-ratio inspirals: a smaller compact object orbiting a supermassive black hole

In the entire universe: a cosmic gravitational wave background? ▶


Credit: Carl Knox, OzGrav

- The analysis of LISA data will be drastically **different from current ground-based detection**:
  - ✦ Numerous superimposed sources ≠ isolated events
  - ✦ Different time scales, larger waveform cycles observed
  - ✦ Signal-dominated measurement ≠ noise-dominated
  - ✦ Unique detector ≠ network of detectors

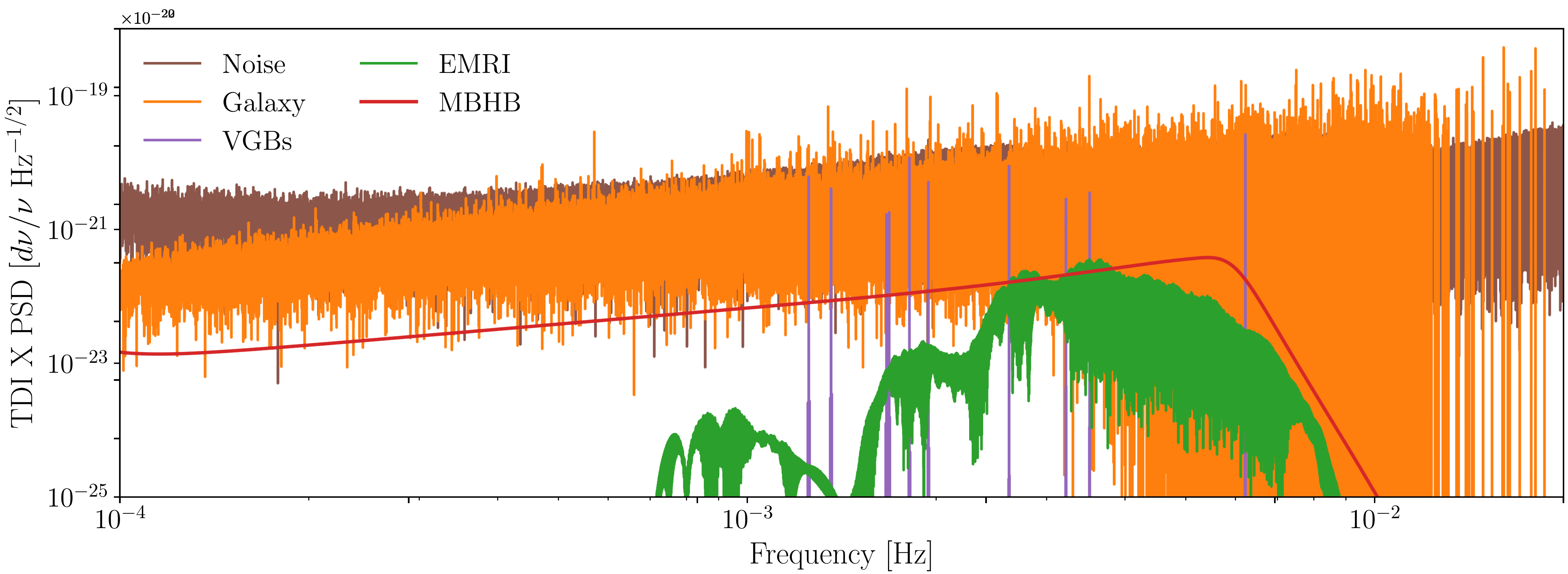
  } Research problem

- Additional difficulties, similar to ground-based detection:
  - ✦ Stochastic noise
  - ✦ Instrumental transients (glitches)
  - ✦ Non-stationarities
  - ✦ Spectral lines
  - ✦ Data gaps

  } Disturbances

- What kind of data will LISA measure?
  - ✦ Fractional frequency deviations (relative doppler shifts) from 27 interferometers
  - ✦ Times series sampled at 4 Hz, observed over 4.5+ years with 82% duty cycle
  - ✦ Dominated by laser frequency noise
  - ✦ After pre-processing, obtain 3 time-delay interferometry (TDI) data streams (X, Y, Z)

VGBs + EMRI + MBHB + Galaxy + noise



- What is the strategy to analyse the data?
  - ✦ Bayesian framework: probe the parameters + number of model components posterior

$$p\left(\boldsymbol{\theta}, k \,|\, \boldsymbol{d}\right) = \frac{p\left(\boldsymbol{d} \,|\, \boldsymbol{\theta}, k\right) p(\boldsymbol{\theta}, k)}{p(\boldsymbol{d})}$$

Model parameters > 10^5

Number of model components ~ several 10^4

Data vector. For example d=(X, Y, Z)

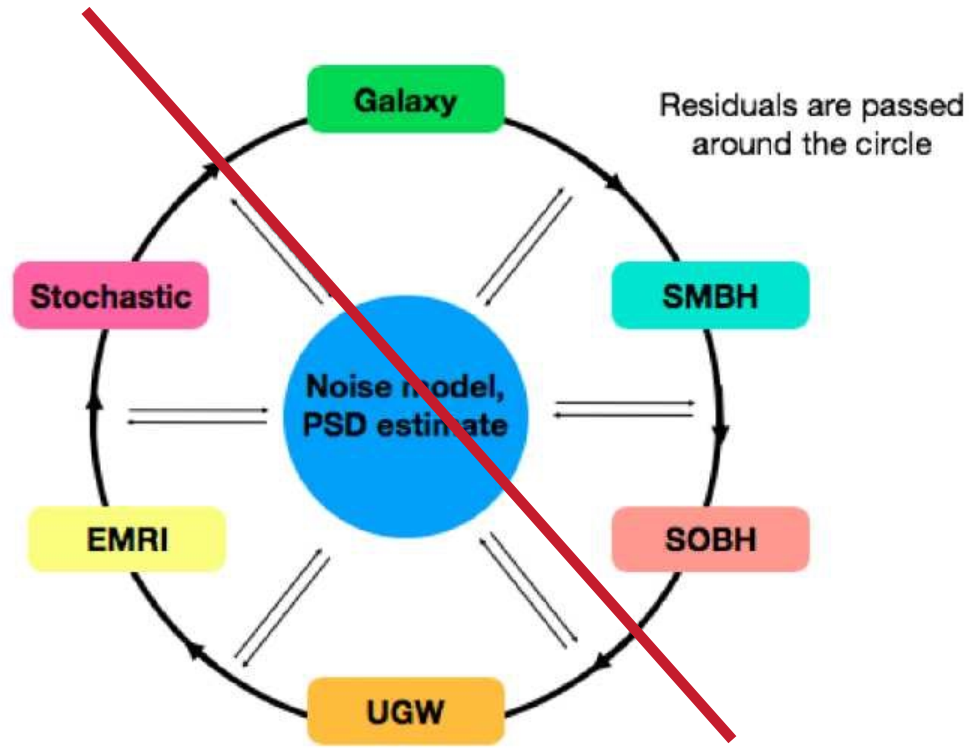
  - ✦ Define a likelihood function: e.g. Gaussian

$$p\left(\boldsymbol{d} \,|\, \boldsymbol{\theta}, k\right) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}(\boldsymbol{\theta})|}} \exp\left\{ \left(\boldsymbol{d} - \boldsymbol{h}(\boldsymbol{\theta}, k)\right)^{\dagger} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \left(\boldsymbol{d} - \boldsymbol{h}(\boldsymbol{\theta}, k)\right) \right\}$$

GW signals: $\boldsymbol{h}(\boldsymbol{\theta}, k) = \sum_{j=1}^{k} \boldsymbol{h}_j(\boldsymbol{\theta}_j)$

Stochastic processes: $\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sum_{i=1}^{p} \boldsymbol{\Sigma}_i(\boldsymbol{\theta}_i)$

- In practice, we need to **partition the parameter space** in several sub-problems
- Example of the blocked Gibbs scheme (iterative sampling)



Simultaneous parallel sampling of all sources is statistically incorrect in a Gibbs scheme

Sequential sampling is necessary for convergence:

1. Sample for $p\left(\boldsymbol{\theta}_{\mathrm{GB}} | \boldsymbol{y}, \boldsymbol{\theta}_{\mathrm{others}}\right)$
2. Sample for $p\left(\boldsymbol{\theta}_{\mathrm{MBHB}} | \boldsymbol{y}, \boldsymbol{\theta}_{\mathrm{others}}\right)$
3. Sample for $p\left(\boldsymbol{\theta}_{\mathrm{EMRI}} | \boldsymbol{y}, \boldsymbol{\theta}_{\mathrm{others}}\right)$

...

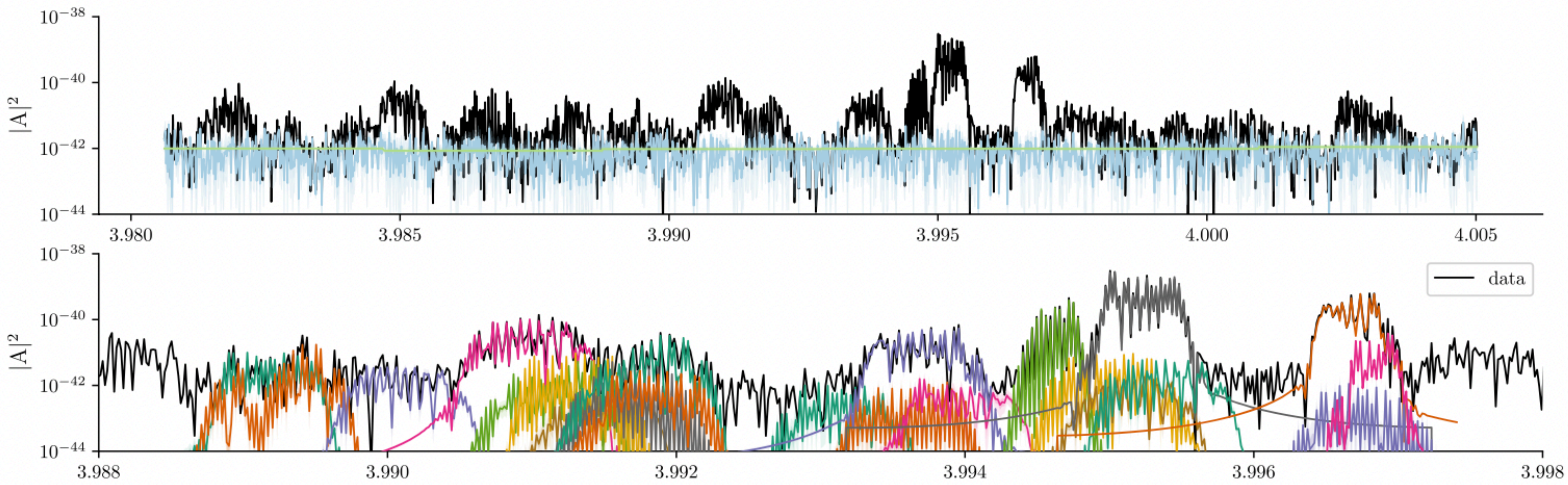
K. Sample for $p\left(\boldsymbol{\theta}_{\mathrm{noise}} | \boldsymbol{y}, \boldsymbol{\theta}_{\mathrm{others}}\right)$

- The number of overlapping sources (especially Galactic binaries) is not know in advance
- Need to estimate the optimal number of sources

[Littenberg et al., 2020, arXiv:2004.08464]

Data

Waveform posteriors

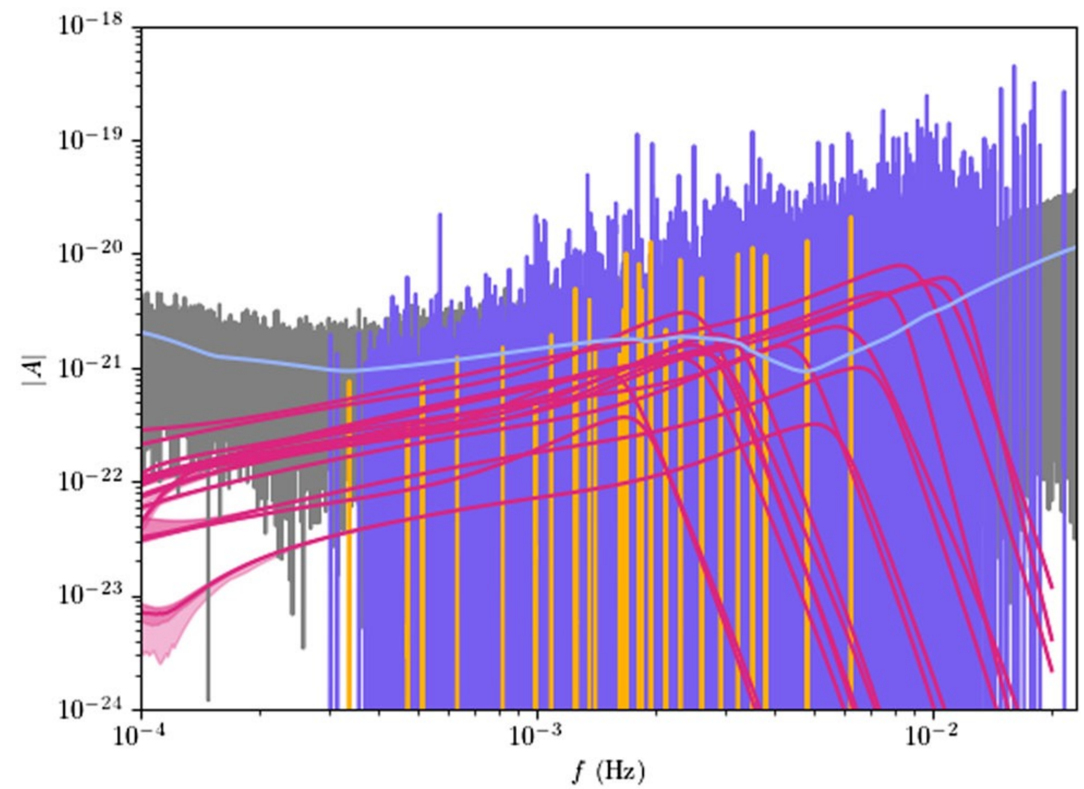


Radler challenge

- Algorithm: reverse-jump Markov-chain Monte Carlo (RJMCMC)
- Data segmentation: allow for parallel computing by splitting the frequency-domain data into segments
- Possible acceleration: Gaussian process modelling of the likelihood [Strub+ 2022]

- Interplay between continuous and transient sources

[Littenberg & Cornish., 2023, arXiv:2301.03673]



MBHB search phase
GB are treated as noise spikes

F-statistics
Hierarchical (high SNR sources first)

Parameter estimation refinement for each MBHB source

Full MCMC exploration

Initialisation

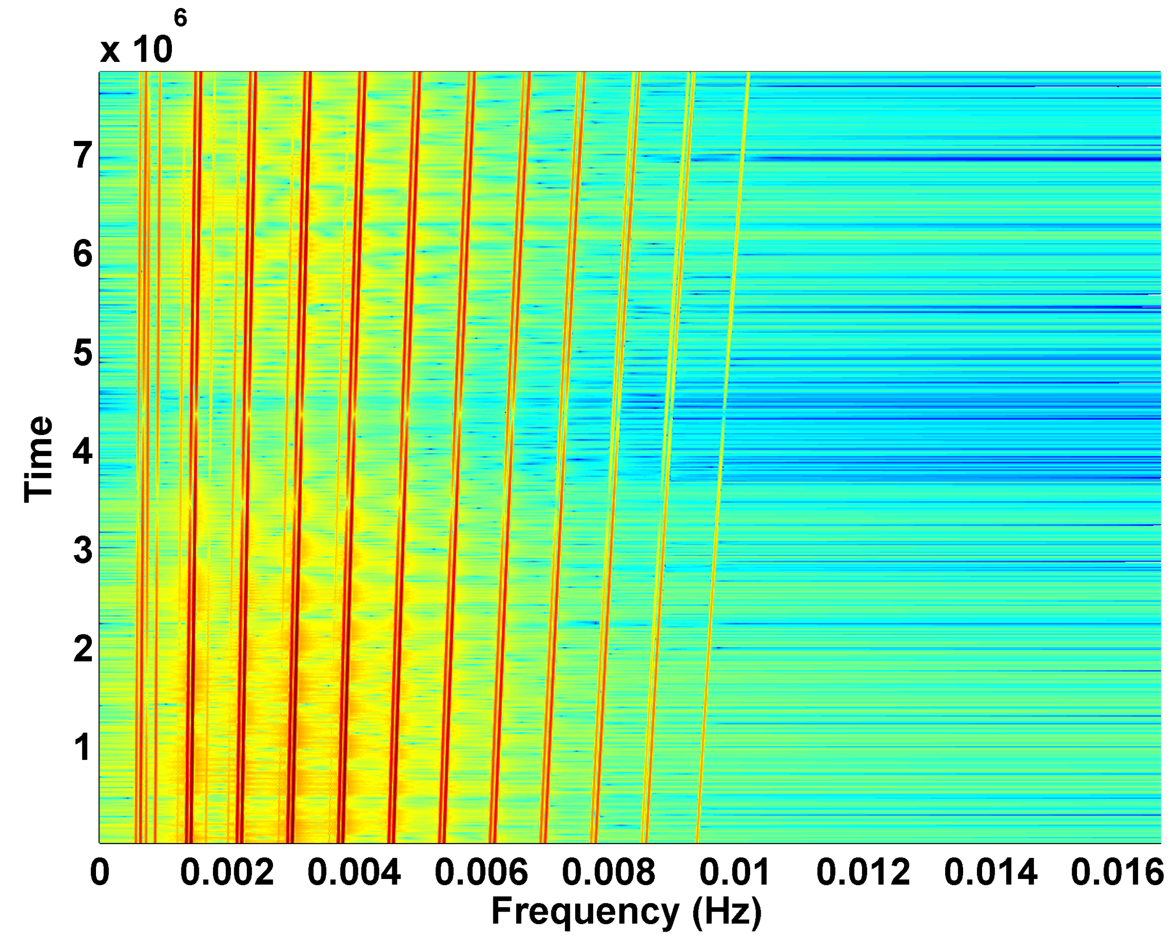
Global fit

Galactic binaries → Noise → MBHB

Sangria challenge

- Other Sangria submissions to be published soon: APC, Katz et al, Strub et al.
- Possible acceleration: GPU-accelerated likelihood computation [Katz+ 2020]

The global fit: 3 extreme-mass ratio inspirals



[Babak et al., 2017]

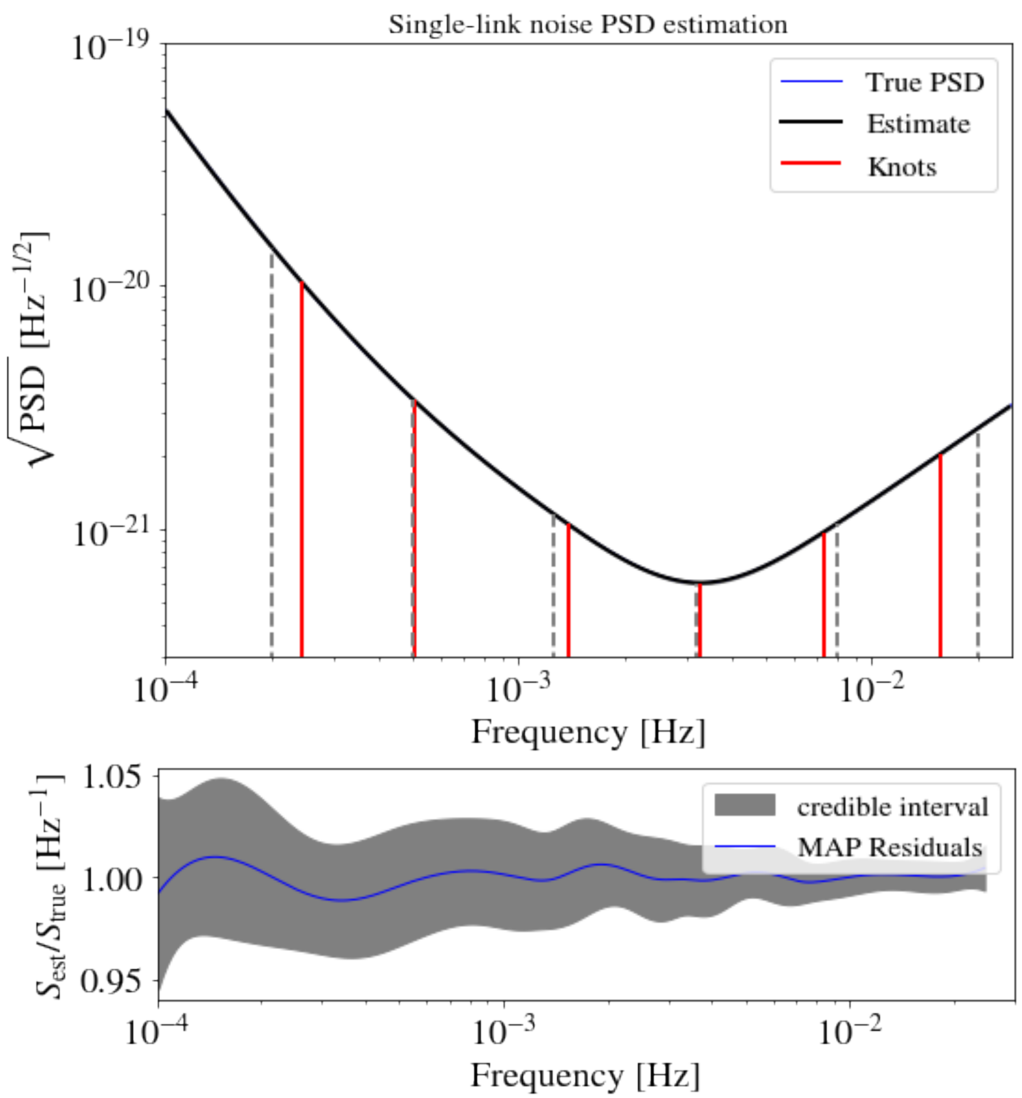
- **Hard problem**:
  - ✦ Complicated orbit through many thousands of cycles
  - ✦ Lots of harmonics
  - ✦ Need for waveform that are both fast and accurate
- Recent improvements in waveform developments
  - ✦ Self-force model: increased accuracy
  - ✦ Kludge models: increased speed
- **Detection and parameter estimation: still a long way to go**
  - ✦ First detection and PE algorithm for single EMRI in Gaussian noise: [Babak, Gair, Porter, 2009]
  - ✦ Full Bayesian MCMC inference (but no blind detection) [Ali et al., 2013; Katz et al., 2021]
- **Yet unaddressed questions**
  - ✦ Detection in **confused environment** ?
  - ✦ With up-to-date waveforms ?

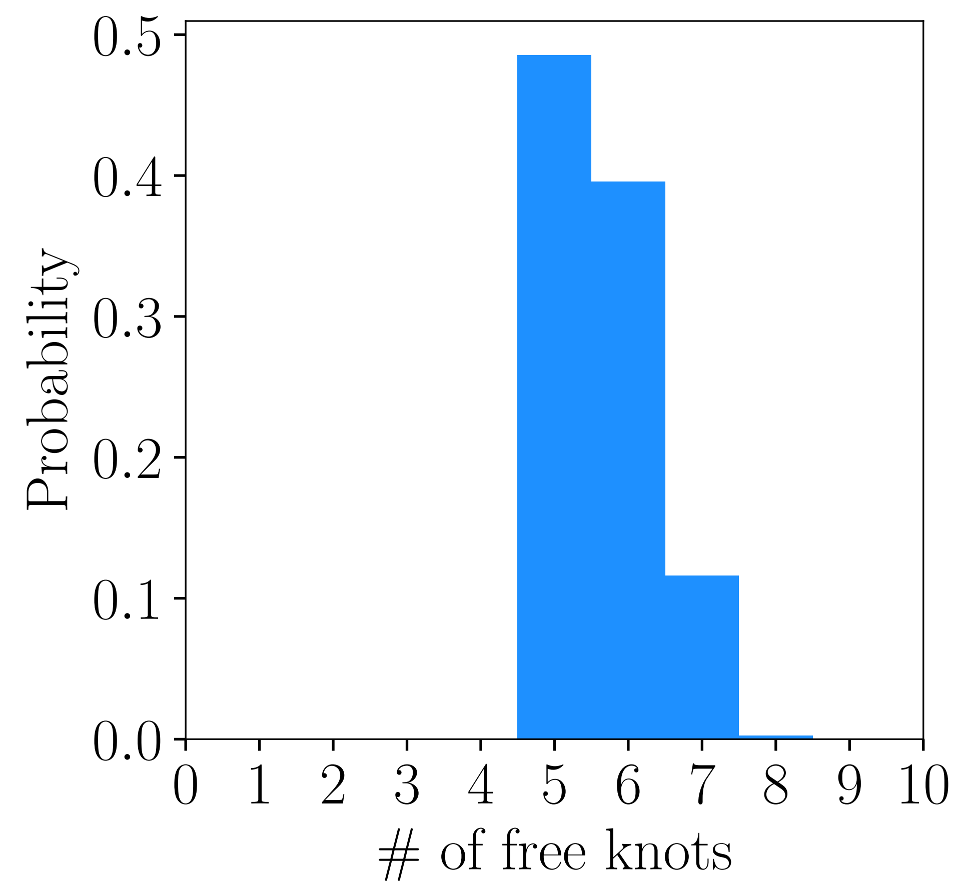
Upcoming

Yorsh challenge

- The noise power spectral density must be estimated consistently across the full frequency band
- Example of BayesLine for LIGO-Virgo [Cornish & Littenberg, 2015]
- Spline noise modelling [Baghi+ 2023]

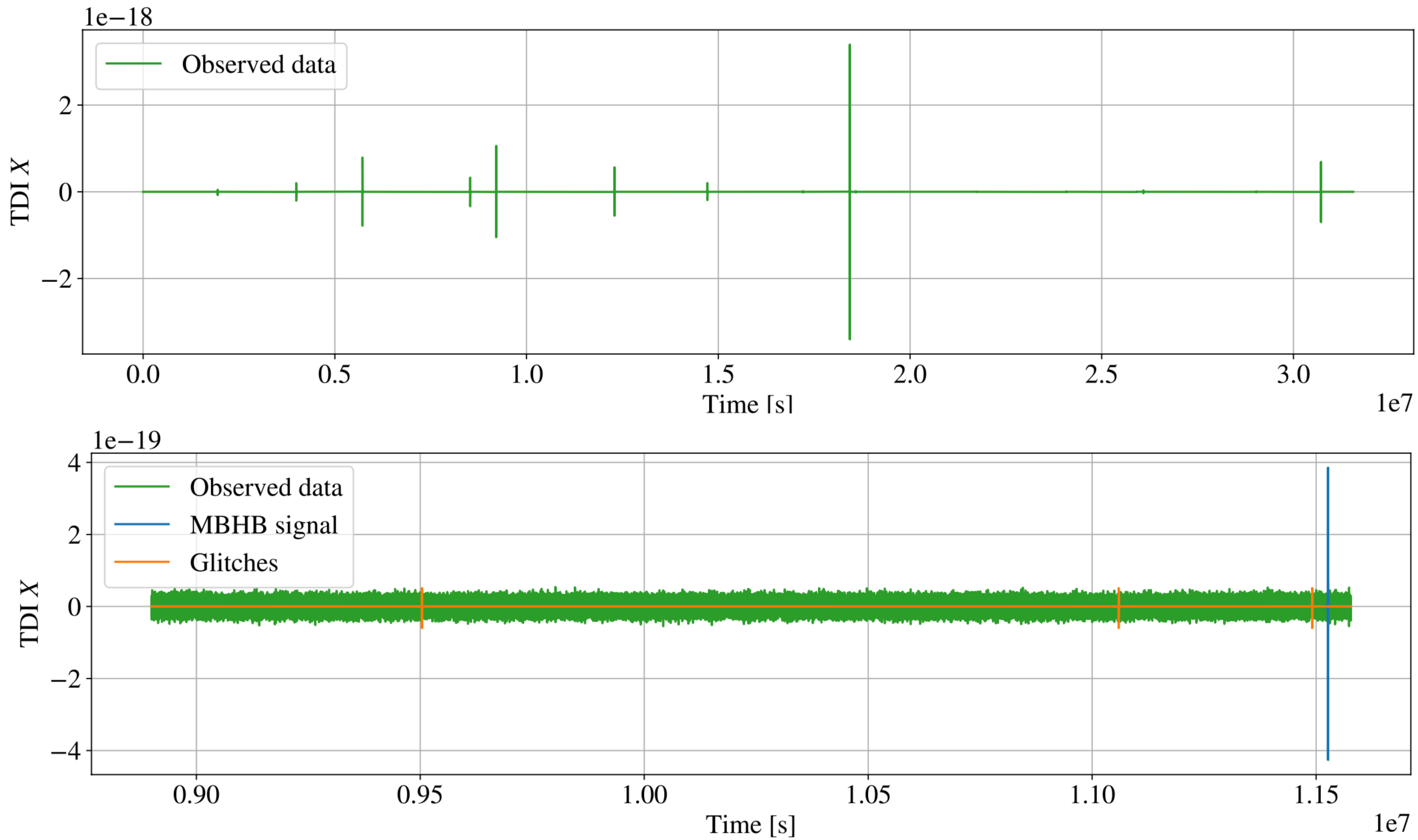
$$\log S_n(f) = \sum_{j=0}^{K-1} c_j B_j \left(f, \boldsymbol{\xi}\right)$$



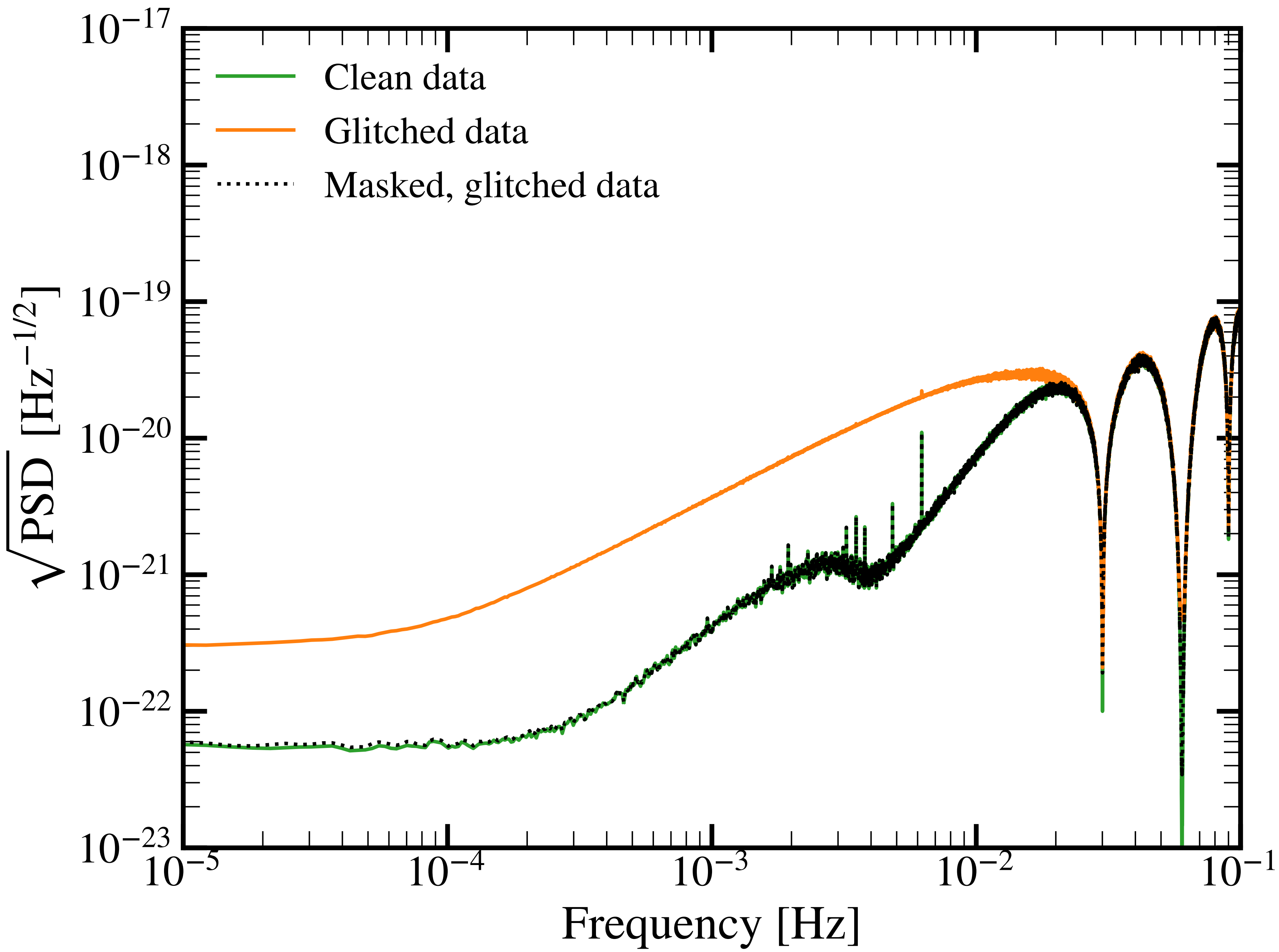


Global fit: transient disturbances (glitches)

- It's almost certain: they will be there! Cf. LISA Pathfinder measurements → Spritz LDC dataset



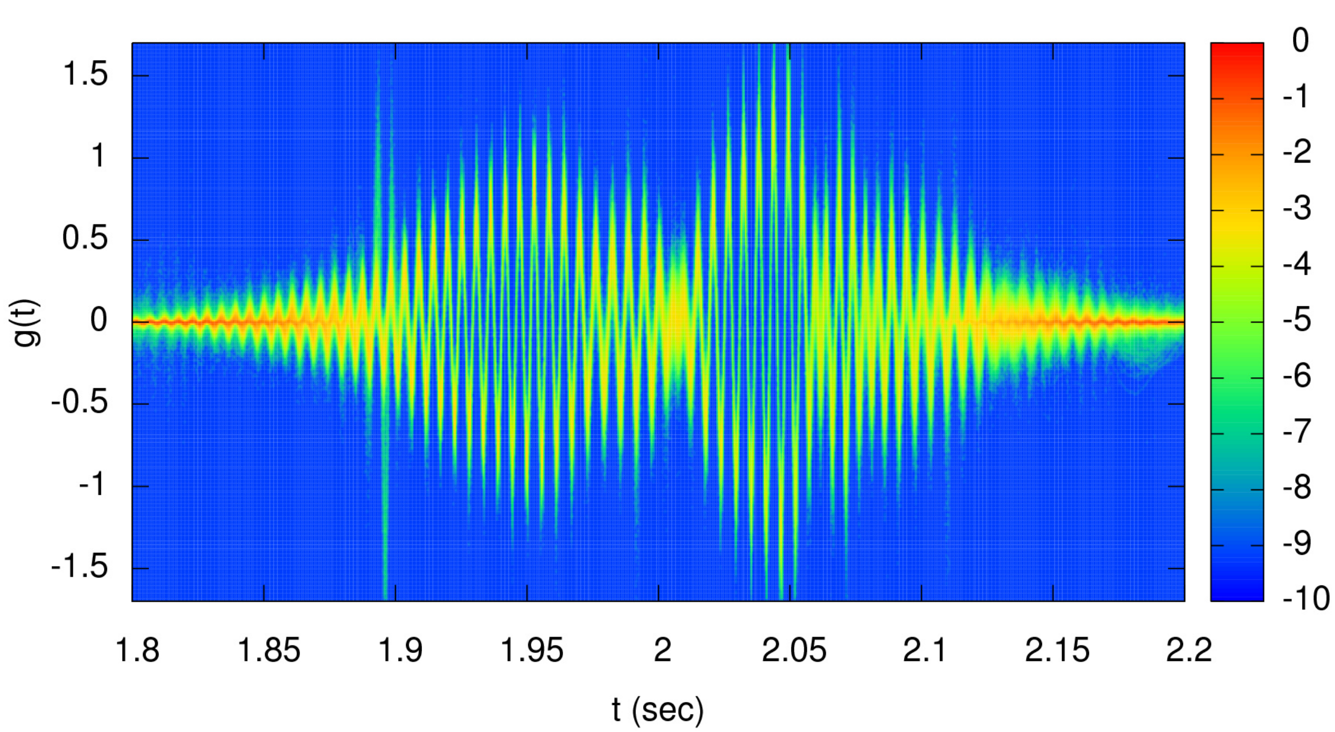
- And we need to do something about them…



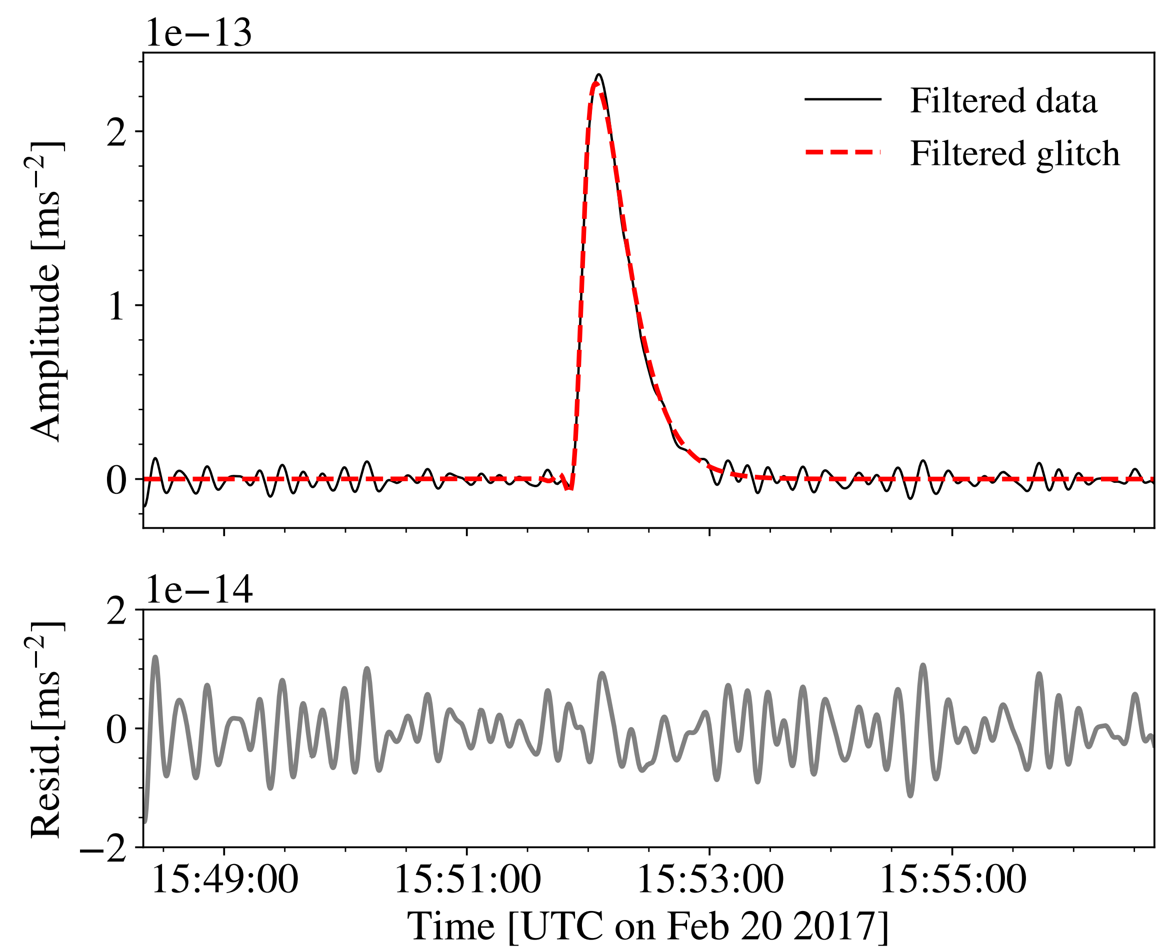
[LISA data analysis robustness report, in prep.]

Global fit: transient disturbances (glitches)

- They should be fitted together with the other signals
  - Morlet-Gabor wavelets [Cornish & Littenberg, 2015]
  - Shapelets [QB et al., 2022]



Glitch model reconstructed with BayesWaves for LIGO/Virgo, arXiv:1410.3835



Glitch model reconstructed with shapelets in LISA Pathfinder, arXiv:2112.07490

Looking for the unknown

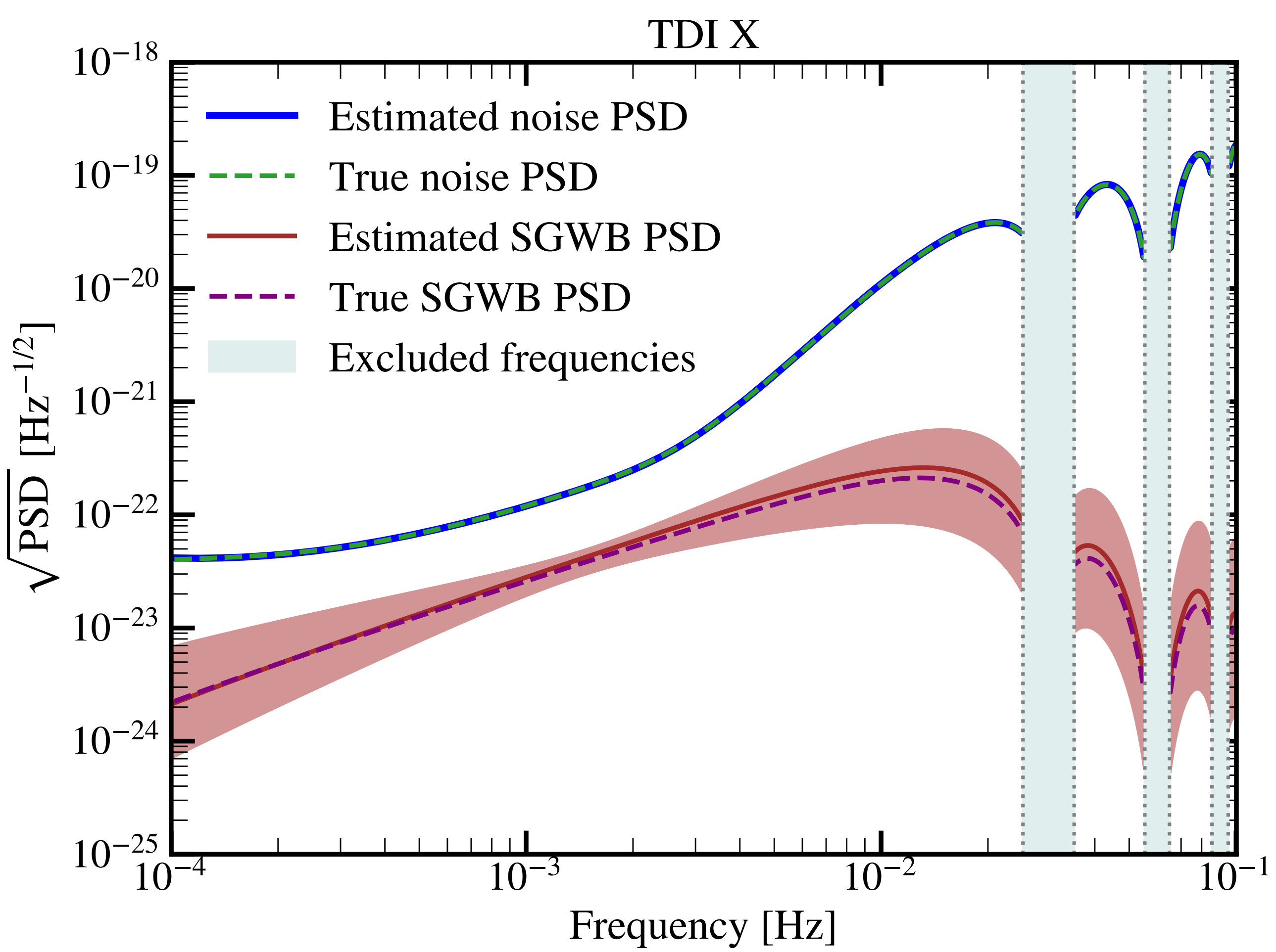
- Possibly the most difficult task: **searching for poorly modelled sources** in the gravitational cacophony
  - Stochastic gravitational-wave backgrounds
  - Cusps and kinks of cosmic strings
  - Unmodelled sources

- Requires **strategies to distinguish between noise and signal**
  - Accurately characterise the stationary, Gaussian noise power spectrum
  - Account for long-term non-stationaries
  - Model short-term non-Gaussianities: glitches

- Searching for **stochastic GW backgrounds** with LISA: up to now there are 3 kinds of searches

Noise \ Signal	Template-based	Template-free
Template-based	Adams and Cornish, 2010 Adams and Cornish, 2013 Boileau et al., 2021 Boileau et al., 2022 Boileau et al., 2023	Caprini et al., 2019 Karsesis et al., 2020 Pieroni and Barauss, 2020 Flauger et al., 2021 Banagiri et al., 2021
Template-free	QB et al., 2023	?

- **Detecting** a modelled stochastic **GW background** has been shown to be possible **when the noise spectral shape is weakly constrained**



[QB, Karnesis, Bayle, Besançon, Inchauspé, 2023 arXiv:2302.12573]

- But here we assumed all noises transformed like interferometric phase noise (OMS)
- May impact SGWB detectability, see Martina's talk!
- **A more general independent component analysis is needed**

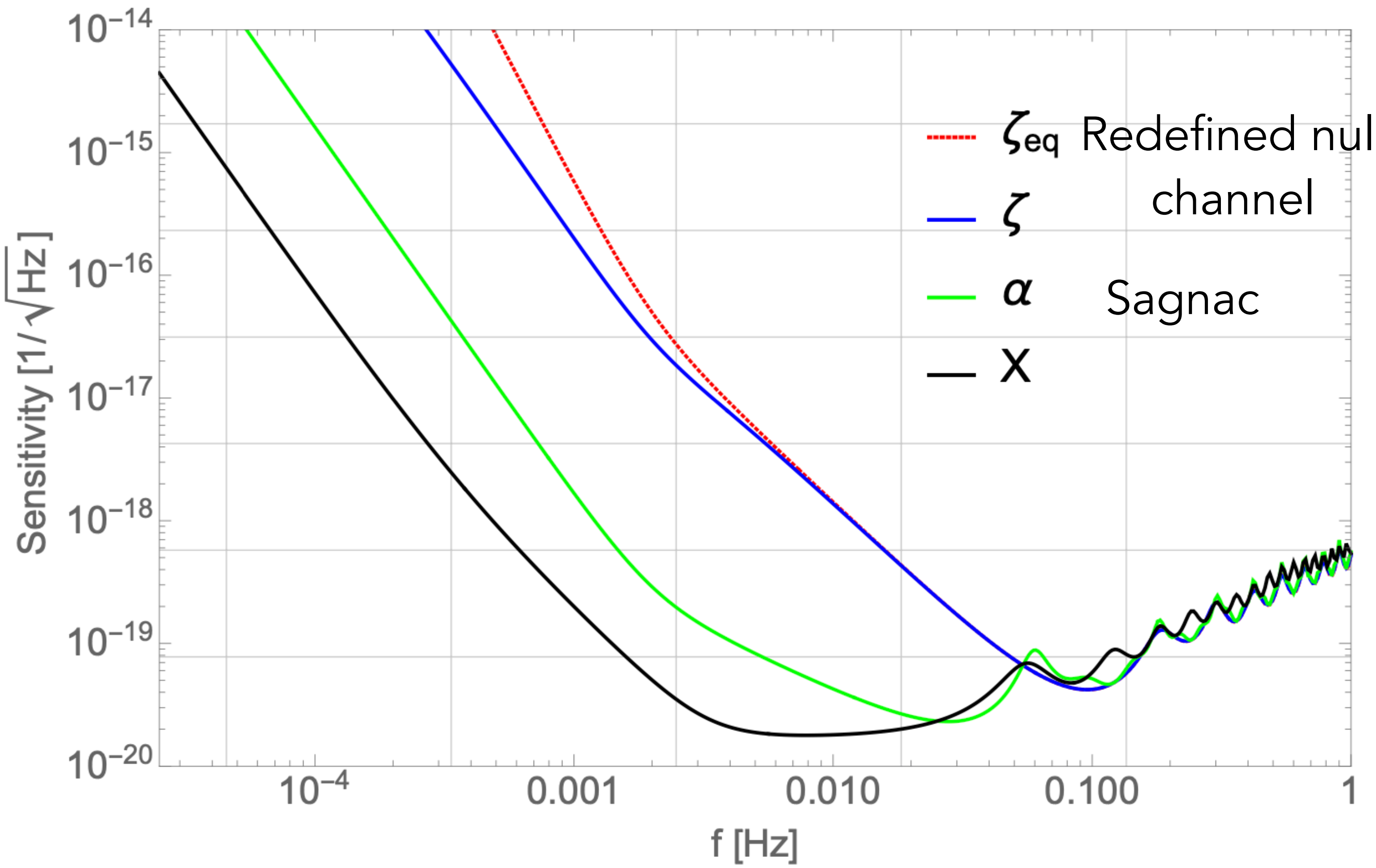
Looking for the unknown

- We used to say:

> *"Having the ability to use the Sagnac (or null-stream) TDI channels to veto such instrumental events will play a crucial role in the exploration of this discovery space"*

[ESA L3 mission concepts proposal, 2017]

- But is it possible?
  - **Need to redefine the null channel**
  - Test-mass noise is still weak in the null channel
  - Hard to distinguish with signal: **see Olaf Hartwig's talk**!



[Muratore et al. 2023, arxiv:2207.02138]

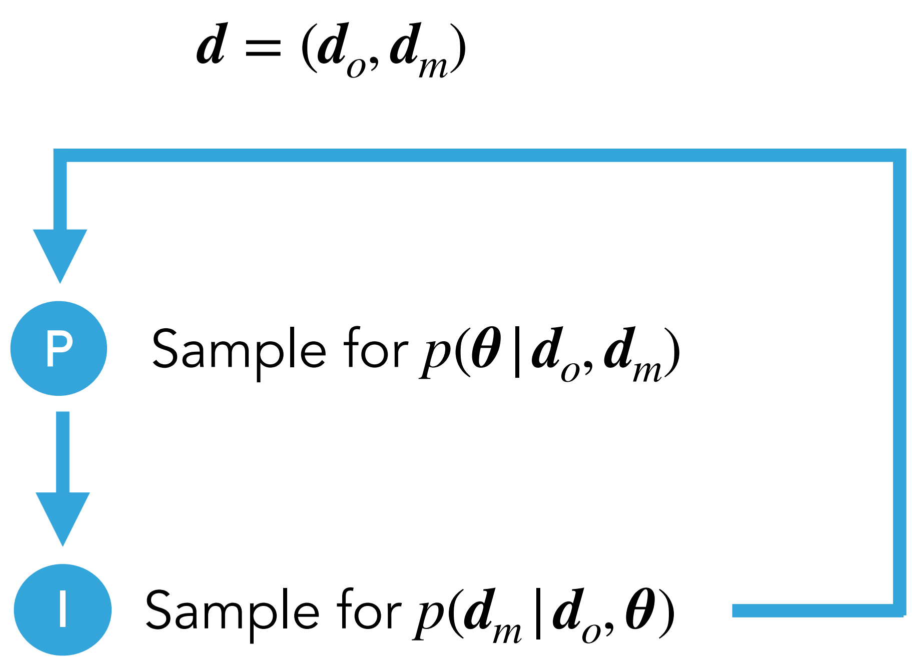
- Solving the LISA global fit is a **ongoing research problem**
- There are already challenges with known GW sources. WE NEED:
  - ✦ Accurate & fast waveform models particularly for MBHBs and EMRIs
  - ✦ To seriously tackle the EMRI detection problem
  - ✦ Computationally efficient methods : low-latency pipeline
- To look for the unknown, including SGWB, WE NEED:
  - ✦ **Robust and accurate noise models + realistic time-domain simulations**
  - ✦ **Fully orthogonal TDI variables:** see Marc Lilley's talk
  - ✦ **Address the confusion problem:** see Robert Rosati's talk
- Framework for research: **the LISA Data Challenges**
  - ✦ Collaborative playground https://lisa-ldc.lal.in2p3.fr/
  - ✦ Progressively increases the number of source types in "enchiladas" + instrumental realism

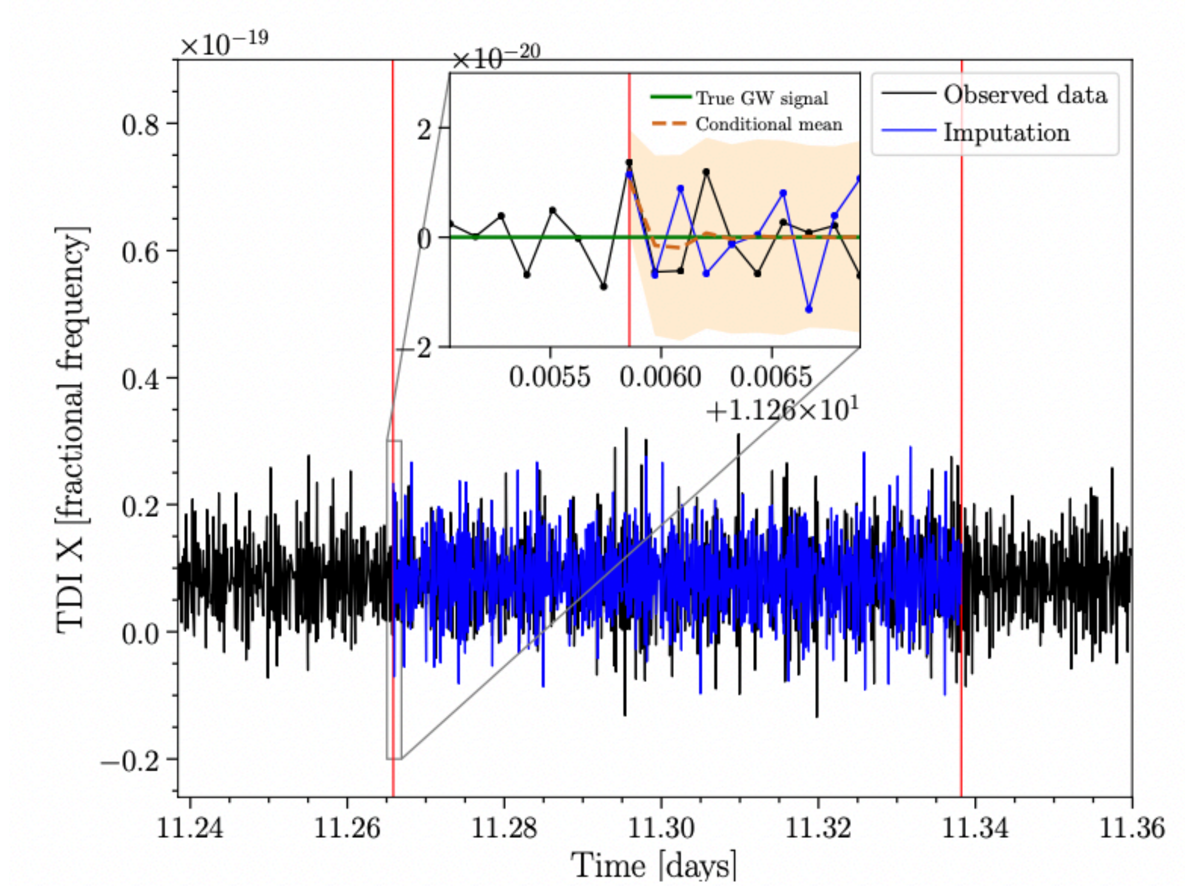
Thank you for your attention !

BACKUP SLIDES

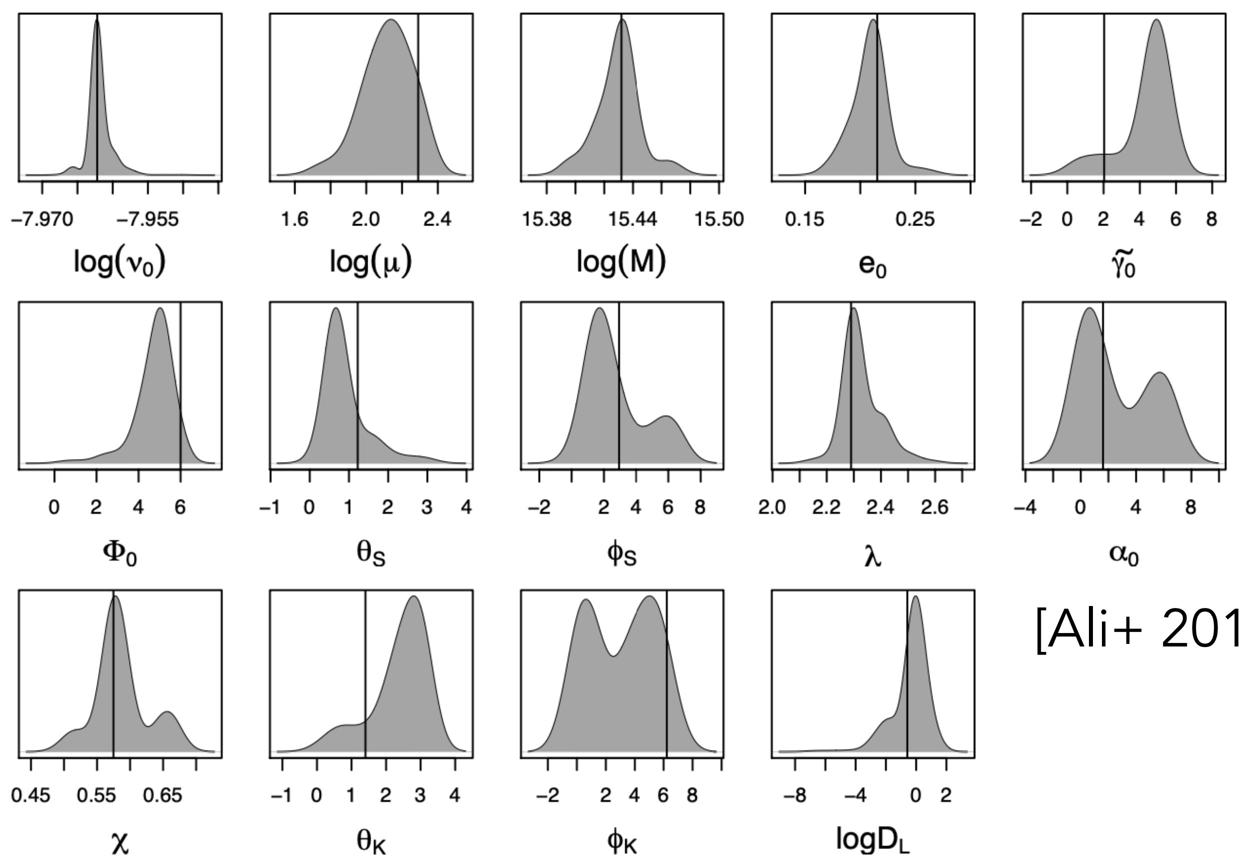
The global fit: data gaps

- Missing data points or gaps: 82% duty cycle!
- Example: interrupted science data due to antenna repointing
- Consequence: both the signal and the covariance become expensive to compute
- One strategy is data augmentation [Baghi et al, 2019]





- Recent improvements in waveform developments
  - ✦ Self-force model: increased accuracy [van de Meent + 2018, Pound+ 2020, Warburton+ 2021]
  - ✦ Kludge models: increased speed [Babak+ 2007, Chua & Gair 2015, Chua+ 2021, Katz+ 2021]
- **Detection and parameter estimation: still a long way to go**
  - ✦ First detection and PE algorithm for single EMRI in Gaussian noise: [Babak, Gair, Porter 2009]
  - ✦ Full Bayesian MCMC inference (but no blind detection) [Ali+ 2013, Katz+ 2021]



[Ali+ 2013]