The LISA Data Challenges

Ivan Martin Vilchez

Institut de Ciències de l'Espai (ICE, CSIC and IEEC)

Iberian GW Meeting, June 9-11 2021

The LISA Data Challenges (LDCs)

- Continuation/reboot of the old Mock LDCs (MLDCs).

The LISA Data Challenges (LDCs)

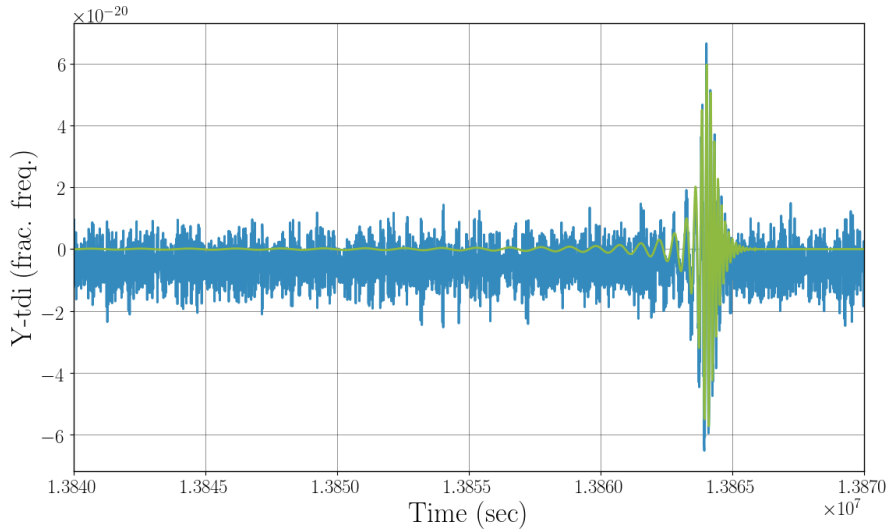
- Continuation/reboot of the old Mock LDCs (MLDCs).
- Goals:
  - Develop and test new DA algorithms and pipelines for LISA, in a "realistic" scenario.
  - Test current waveform and instrument simulations, incentivize development of new, faster, more accurate ones.
  - Create a common toolkit for LISA DA.

The LISA Data Challenges (LDCs)

- Continuation/reboot of the old Mock LDCs (MLDCs).
- Goals:
  - Develop and test new DA algorithms and pipelines for LISA, in a "realistic" scenario.
  - Test current waveform and instrument simulations, incentivize development of new, faster, more accurate ones.
  - Create a common toolkit for LISA DA.
- You can participate! `https://lisa-ldc.lal.in2p3.fr/ldc`

Source Types and their challenges

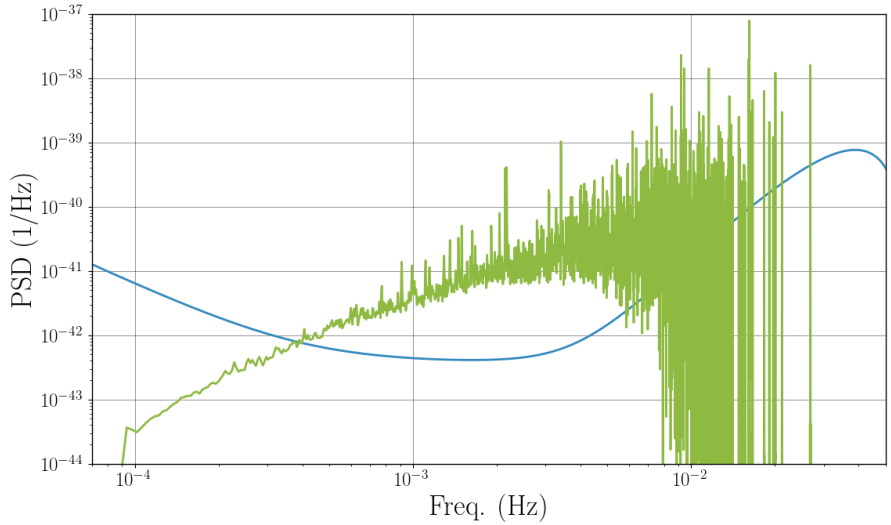
- Massive Black Hole Binaries (MBHBs):
  - Very high SNR, low number of events expected.
  - High dimensionality in parameters, likely multimodally distributed (degeneracies).



Credit: *LISA Data Challenges*
`https://lisa-ldc.lal.in2p3.fr/ldc`

Source Types and their challenges

- Galactic Binaries (GBs):
  - Very simple and fast models exist.
  - There are a lot of them, and overlap in a way many will be unresolvable.



Credit: *LISA Data Challenges*
`https://lisa-ldc.lal.in2p3.fr/ldc`

Source Types and their challenges

- Extreme Mass-Ratio Inspirals (EMRIs):
  - Rich harmonic signature is characteristic.
  - Harmonics have relatively low SNR, waveform models need more development.
- Stellar-Origin Binary Black Holes (SOBBHs):
  - They can have overlap with LIGO-Virgo, opportunity for multi-band astrophysics.
  - There are many of them, with low SNR.
- Stochastic Gravitational Wave Background (SGWB):
  - Of cosmological origin, should be isotropic.
  - Unknown model beyond that (?)

LDC 1: Radler

Several separate datasets:

- 1 Massive Black Hole Binary (MBHB)
- 1 Extreme Mass-Ratio Inspiral (EMRI)
- 30 million Galactic Binaries (GBs)
  - Secondary dataset: **10 Verification Galactic Binaries (VGBs)**
- 21721 Stellar Origin Black Hole Binaries (SOBBHs)
- Stochastic GW background (SGWB) signals

Parallel-Tempered MCMC[1]

- Our goal: estimate parameter values $\vec{\boldsymbol{\theta}}$ from data $\mathbf{x}$.
- Done by sampling $p(\vec{\boldsymbol{\theta}}|\mathbf{x}) \propto L(\mathbf{x}|\vec{\boldsymbol{\theta}})\ \pi(\vec{\boldsymbol{\theta}})$.
- Some succesful MLDC Data Analysis[2] was based on Parallel-Tempered MCMC (PTMCMC):
  - Run multiple MCMC chains in parallel
  - Each chain has a temperature T
  - What is being sampled is $p_T(\vec{\boldsymbol{\theta}}|\mathbf{x}) \propto L(\mathbf{x}|\vec{\boldsymbol{\theta}})^{1/T}\ \pi(\vec{\boldsymbol{\theta}})$
  - Chains with higher T explore larger regions of parameter space.
  - Swaps between adjacent temperature chains are proposed at set intervals with acceptance probability $A_{i,j} = \min\left\{\left(\frac{L(\mathbf{x}|\vec{\boldsymbol{\theta}}_i)}{L(\mathbf{x}|\vec{\boldsymbol{\theta}}_j)}\right)^{\beta_j-\beta_i}, 1\right\}$.

[1] For an introductory MCMC text, see e.g. W.R. Gilks, S. Richardson, D. Spiegelhalter; *Markov Chain Monte Carlo in Practice* (1996)

[2] See for instance: MLDC Task Force (2010); arXiv:0912.0548

PTMCMC for VGBs

- We want to do as basic of a search as possible, with potential to be refined later.
- Use the `ptemcee`[1] Python package as our backend.
- `FastGB` is used for waveform simulations $\vec{\boldsymbol{\theta}} \rightarrow \mathbf{h}(\vec{\boldsymbol{\theta}})$.
- Our priors $\pi(\vec{\boldsymbol{\theta}})$ are flat for all parameters[2].
- Likelihood:

$$\log L(\mathbf{x}|\vec{\boldsymbol{\theta}}) = \sum_{n=A,E,T} 2(\mathbf{h}(\vec{\boldsymbol{\theta}})|\mathbf{x})_n - (\mathbf{h}(\vec{\boldsymbol{\theta}})|\mathbf{h}(\vec{\boldsymbol{\theta}}))_n,$$

with

$$(\mathbf{g}|\mathbf{h})_n = 2\int_0^\infty \mathrm{d}f\, \frac{\tilde{g}(f)\tilde{h}^*(f) + \tilde{g}^*(f)\tilde{h}(f)}{S_n(f)}.$$

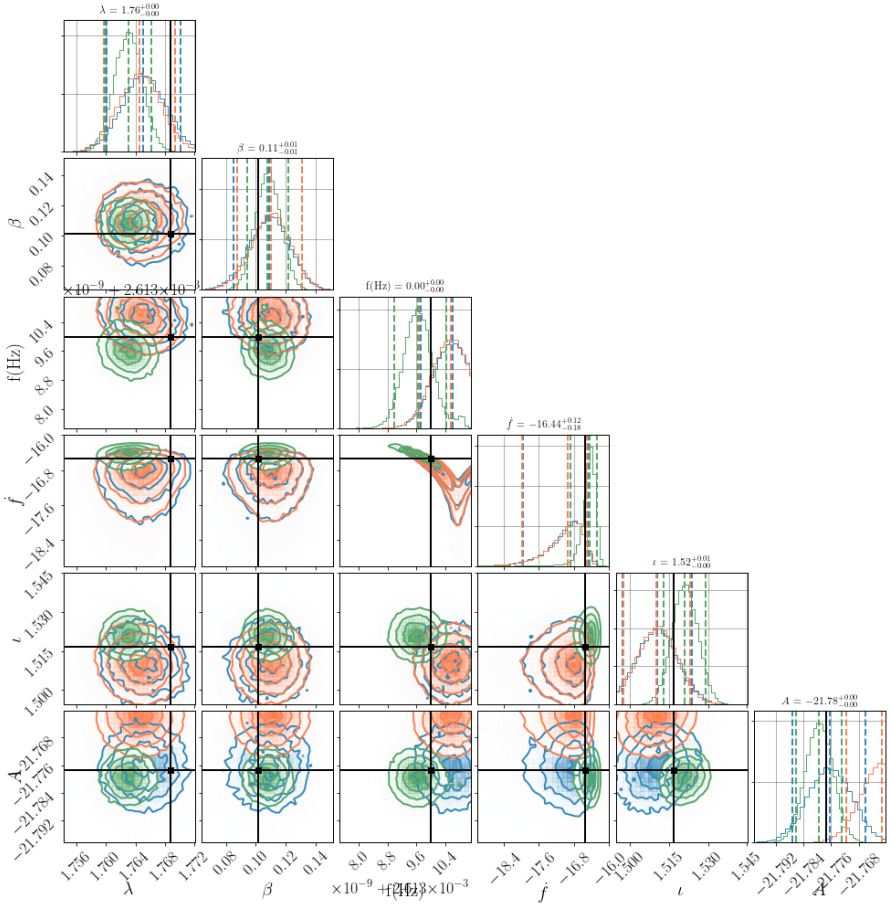
[1] W. Vousden, W.M. Farr, and I. Mandel (2015); arXiv:1501.05823

[2] we also experimented with priors uniform in the $\log_{10}$ for $\mathcal{A}$ and $\dot{f}_0$.

Current LDC status

- Radler evaluation:
  - Currently, 7 submissions for MBHBs, 4 for VGBs, 2 for GBs, 3 for SGWB, (1) for SOBBH, (1) for EMRIs.
  - Wide variety of algorithms, tools and assumptions used for each source type.
- Computational resources:
  - ESA now needs an assessment of the computational infrastructure that will be needed for the ground segment work (INREP & Data Analysis) in the LISA mission.
  - Information from the LDC is being used to estimate costs and write a report.

Radler VGB Evaluation (preliminary!)



Preliminary results. Credit: S. Babak *(LDC1 Evaluation)*

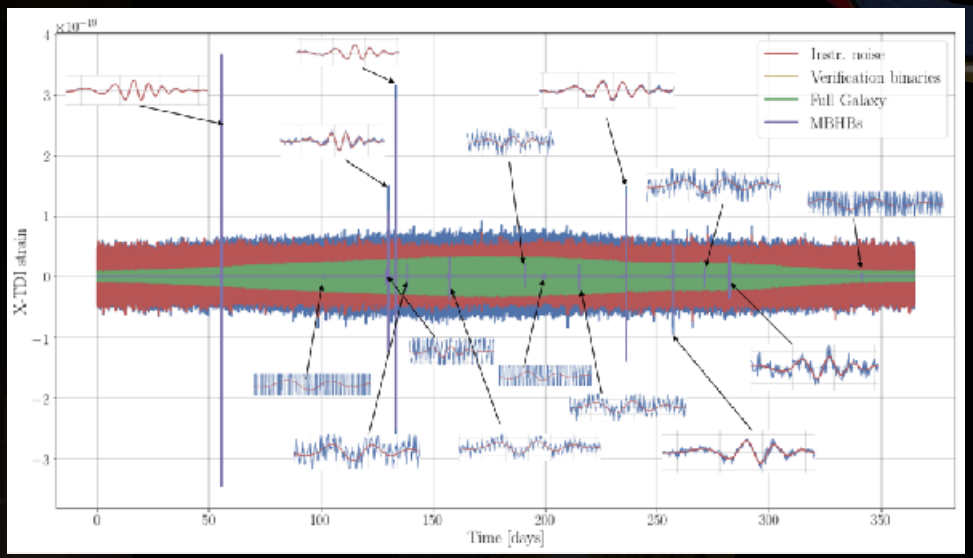
- Notation:
  - Birmingham[1]
  - Marshall-Montana[2]
  - Barcelona
- General agreement among submissions on most parameters
- Some discrepancies that could be explained
- Evaluation still ongoing

[1] R. Buscicchio; University of Birmingham

[2] T. Littenberg, K. Lackeos; NASA Marshall Space Flight Center; arXiv:2004.08464

WG3 - LDC2a (Sangria)

- Data set containing mlns Galactic binaries and merging massive black hole binaries
- Blind data challenge



11

Towards Consortium Ground-Segment Deliverables - LDPG - LISACon8 - 1st to 3rd Dec 2020

Credit: A. Petiteau, S. Babak

WG3 - LDC2b (Spritz)

- Instruments:
  - Glitches ingestion
  - Gaps
  - TDI 2.0, residual laser noise, pseudo-realistic orbits, non-stationary noise by reduced Galactic background
- Two datasets
  - Verification GB data set : 1 year of data
  - MBHB data set: a single event, 1 month of data

12

Towards Consortium Ground-Segment Deliverables - LDPG - LISACon8 - 1st to 3rd Dec 2020

Credit: A. Petiteau, S. Babak

WG3 - LDC1b (Yorsh)

- Requested data sets
- Two datasets:
  - SBBH: 2 data sets 2 years long each:
    - Gaussian instrumental noise + 1 SBBH of SNR 25, measurable evolution in frequency (can take one of LVC events)
    - Gaussian instrument noise + 1 SBBH of SNR 10.
  - EMRI: data set with "realistic" EMRI signal of SNR 40
    - 1 EMRI (typical, taken from catalogue), Augmented Analytic kludge, Gaussian instrumental noise. 2 years
    - 1 EMRI (non-rotating BH), fully relativistic model, Gaussian instrumental noise, 2 years

13

Towards Consortium Ground-Segment Deliverables - LDPG - LISACon8 - 1st to 3rd Dec 2020

Credit: A. Petiteau, S. Babak

Our plans

- Keep participating in the challenges.
- Switch to MBHBs.
- New parameter estimation methods: currently experimenting with Machine Learning.
- Work on developing a fast EMRI waveform model.

Comments and questions

Thank you for your attention.