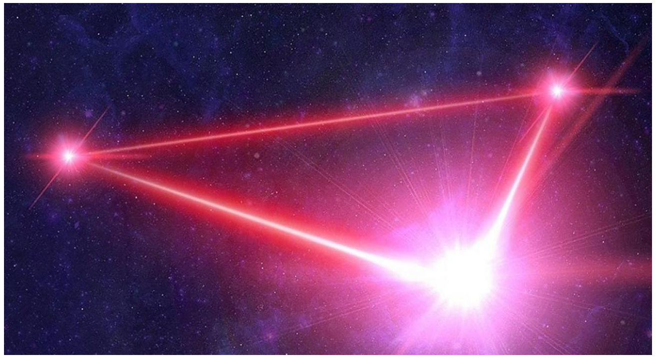
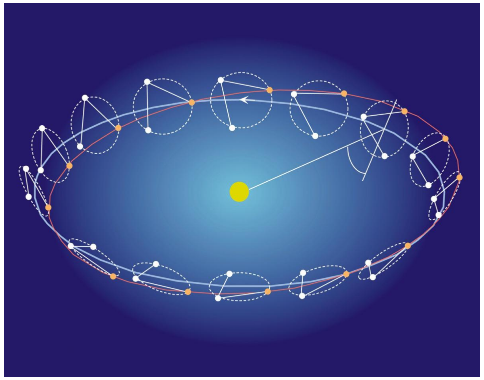
DE LA RECHERCHE À L'INDUSTRIE

DATA ANALYSIS : STATUS, PROJECTS AND PLANS





LISA Data Analysis in France workshop | Quentin Baghi, Marc Besançon, Jérôme Bobin, Hervé Moutarde, Antoine Petiteau

19 MAY 2022

www.cea.fr

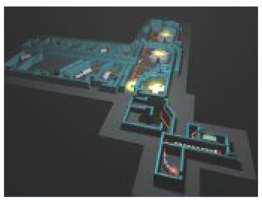
PRESENTATION

ABOUT IRFU

Institute of Research into the Fundamental laws of Universe
Anne-Isabelle Etienvre

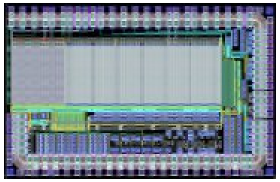
DAp
Astrophysics and Space Technologies
Pierre-Olivier Lagage

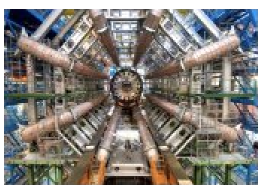
DACM
Accelerators, Cryogenics and Magnetism
Pierre VEDRINE



DPhN
Nuclear Physics
Franck SABATIE

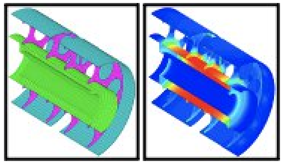
DEDIP
Detectors, Electronics and Computing
Eric DELAGNES





DPhP
Particle Physics
Nathalie Besson

DIS
MechanDcal Design and System Engineering
Pierre Manil



GANIL
Accelerators, Cryogenics and Magnetism
Patricia Roussel-Chomaz

200 physicists
300 engineers
190 technicians
100 PhD students
150 CDD and postdocs

LISA DATA ANALYSIS GROUP AT IRFU

* 4 Divisions involved in LISA data analysis:
  - DPhP: Département de Physique des Particules
  - DPhN: Département de Physique Nucléaire
  - DAp: Département d'Astrophysique
  - DEDIP: Département d'Electronique des Détecteurs et d'Informatique pour la Physique

* IRFU Data Analysis Group:
  - 4 Permanent staff members:
    - Marc Besançon (DPhP)
    - Jérôme Bobin (DEDIP)
    - Hervé Moutarde (DPhN)
    - Antoine Petiteau (DPhP)
  - 2 Post-docs:
    - Quentin Baghi (DPhP)
    - Adrien Bourgoin (DAp)
  - 3 PhDs :
    - Aurore Blelly (DPhN)
    - Elie Leroy (DPhN)
    - Chantal Pitte (DPhP)
  - 2 intern students in average over the 4 divisions

POSSIBLE CONTRIBUTIONS

POSSIBLE CONTRIBUTIONS

- Science Softwares
  - L0->L1 SW
    - INREP 1
    - INREP 2
    - ?
  - L1->L2 Alerts SW
    - Low Latency Alert Pipeline 1
    - Low Latency Alert Pipeline 2
    - Deep Analysis Alert Pipeline
  - L1->L2 Deep Analysis
    - Global Fit 1 pipeline
    - Global Fit 2 pipeline
    - Global Fit 3 pipeline
  - L1->L2 Mono-Blocks
    - MBHB
    - EMRIs
    - SOBBH
    - Galactic Binaries
    - Unmodelled Sources
    - Stochastic Background
    - Instrumental Noise Cleaning
  - L2->L3 SW

- Simulation
  - Simulateur Instrumental
  - Simulation Workflow
- Common SW and Tools
  - Comparison Methods
  - Wave Form generator (Fast Approximate and High fidelity)
    - WFG - MBHB
      - WFG - MBHB - FA
      - WFG - MBHB - HF
    - WFG - Galaxy
      - WFG - Galaxy -FA
      - WFG - Galaxy - HF
    - WFG - EMRIs
      - WFG - EMRIs - FA
      - WFG - EMRIs - HF
    - WFG - SOBBH
      - WFG - SOBBH - FA
      - WFG - SOBBH - HF
    - WFG - Stochastic Background
  - Time series Management tool
  - AI model
  - Generic Research Methods
- External Preparatory Data
- Data Quality Monitoring

POSSIBLE CONTRIBUTIONS

* **Possible leading role:**
  - Low Latency Pipeline
  - MBHB mono-bloc
  - GB mono-bloc
  - Waveforms for GB fast and high-fidelity

* **Contributions:**
  - Global fit
  - SWGB
  - Instrumental Noise Cleaning
  - Catalogue facilities
  - Instrumental simulation
  - Simulation workflow
  - Model for SGWB
  - AI
  - Generic Research Methods

PROJECTS

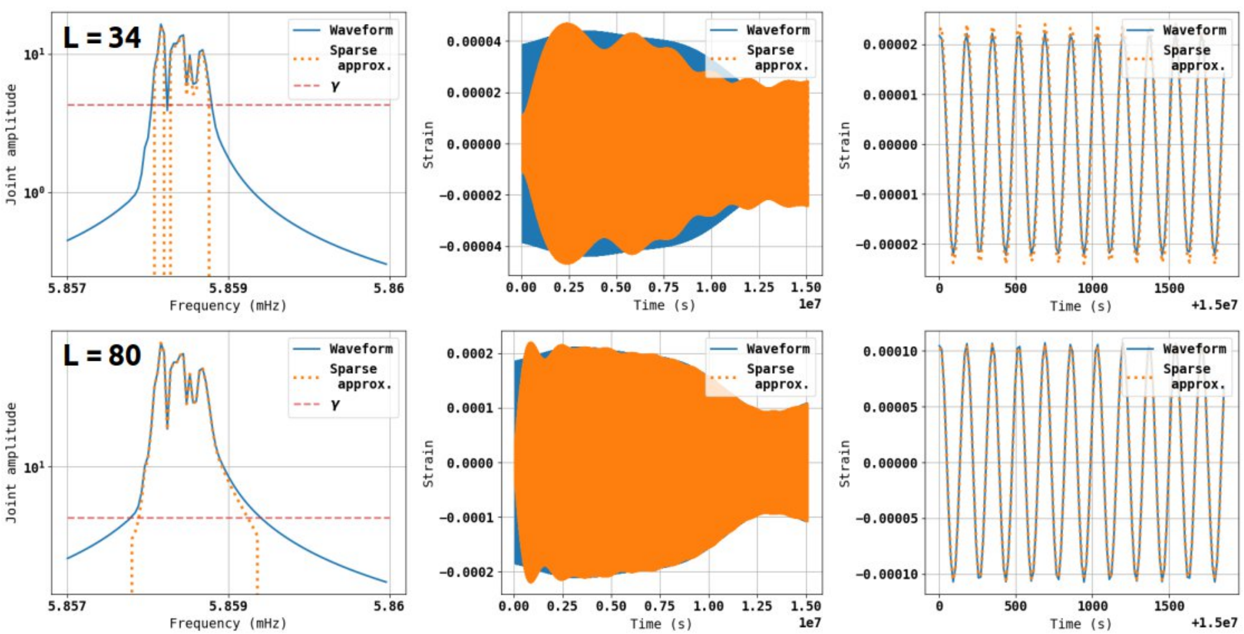
PAST AND CURRENT PROJECTS

* Non-parametric methods
* PCA
* SGWB / Noise
* QNM
* Glitches & Gaps

PROJECTS:
NON PARAMETRIC METHODS

Sparse modeling for LISA data analysis

- Develop fast non-parametric methods for **event detection** or **recovery**.
- Method : builds upon **sparse modeling** and iterative solvers
- No knowledge about the waveform beyond the **sparsity assumption over a specific dictionary**.
- Used first for galactic binaires, but in principle can be used for other sources.



Sparsity for galactic binaries in Fourier basis
(A. Blelly PhD dissertation)

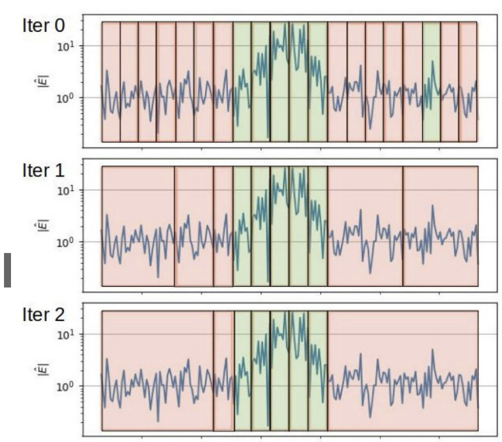
Sparse modeling algorithms : features

$$\hat{\mathbf{h}}_I = \underset{\mathbf{v}}{\mathrm{Argmin}} \left[\sum_k \gamma[k] \mathcal{A}_{\mathbf{s}_I}(\mathbf{v}_I)[k] + \frac{1}{2}\langle \mathbf{d}_I - \mathbf{v}_I, \mathbf{d}_I - \mathbf{v}_I \rangle_{\mathbf{s}_I} \right]$$

Signal estimate Threshold Data PSD

– Direct relation between **threshold** and **false positive rate**.

$$\hat{\mathbf{h}}_I[k] = \begin{cases} \dfrac{\mathcal{A}_{\mathbf{s}_I}(\mathbf{d}_I)[k] - \gamma[k]}{\mathcal{A}_{\mathbf{s}_I}(\mathbf{d}_I)[k]} \mathbf{d}_I[k] & \text{if } \mathcal{A}_{\mathbf{s}_I}(\mathbf{d}_I)[k] \geq \gamma[k] \\ 0 & \text{otherwise .} \end{cases}$$



– Can **simultaneously process** all TDI channels.
– Can be extended to frequency blocks to account for **signal substructures**.
– **Modular** structure: can be plugged in a bigger pipeline.
– Can work with **joint PSD estimation**.

NON PARAMETRIC METHODS (3/3)

Sparse modeling algorithms : applications

.

– Used in LDC1 for verification binaries.

$$(\widehat{\mathbf{h}}, \widehat{\mathbf{d}_m}) = \underset{\substack{(\mathbf{v},\mathbf{u}) \\ M\mathbf{u}=0}}{\mathrm{Argmin}} \left[\|\gamma \odot \mathcal{A}_S(\mathbf{v})\|_1 + \frac{1}{2} \langle \mathbf{d}_g + \mathbf{u} - \mathbf{v}, \mathbf{d}_g + \mathbf{u} - \mathbf{v} \rangle_{\mathbf{s}} \right]$$

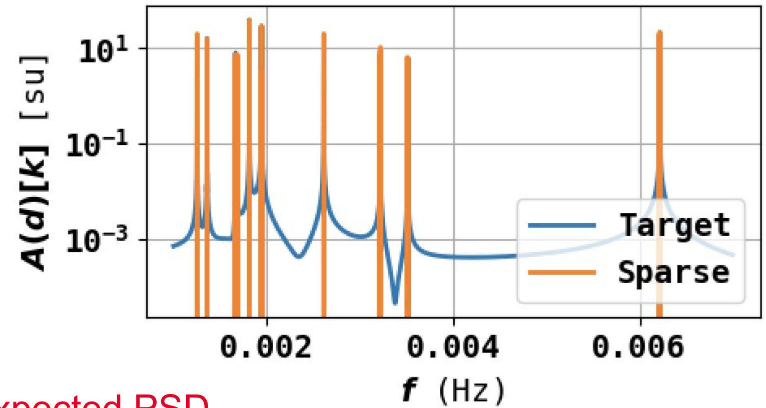
Noise and signal estimates

Mask

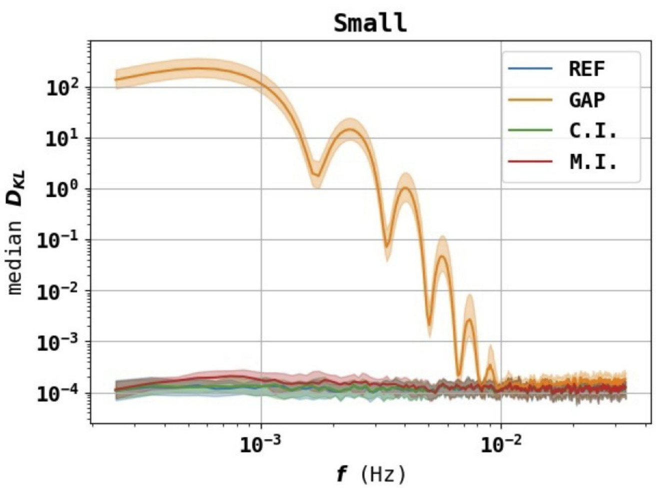
Threshold

Gapped data

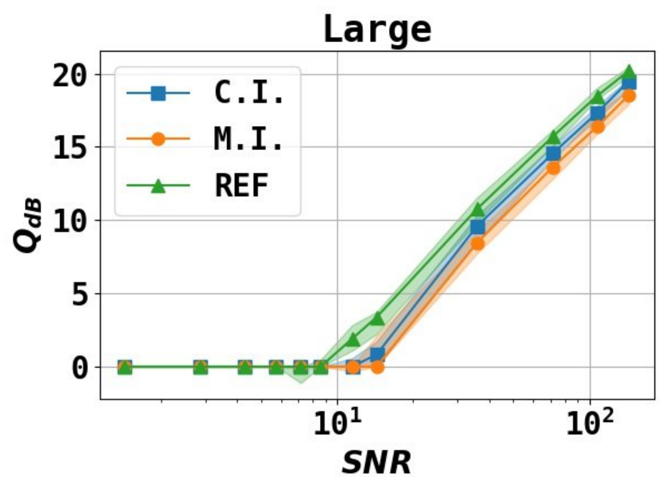
Expected PSD



– Gap mitigation (small, medium and large gaps).



Kullback Leibler divergence

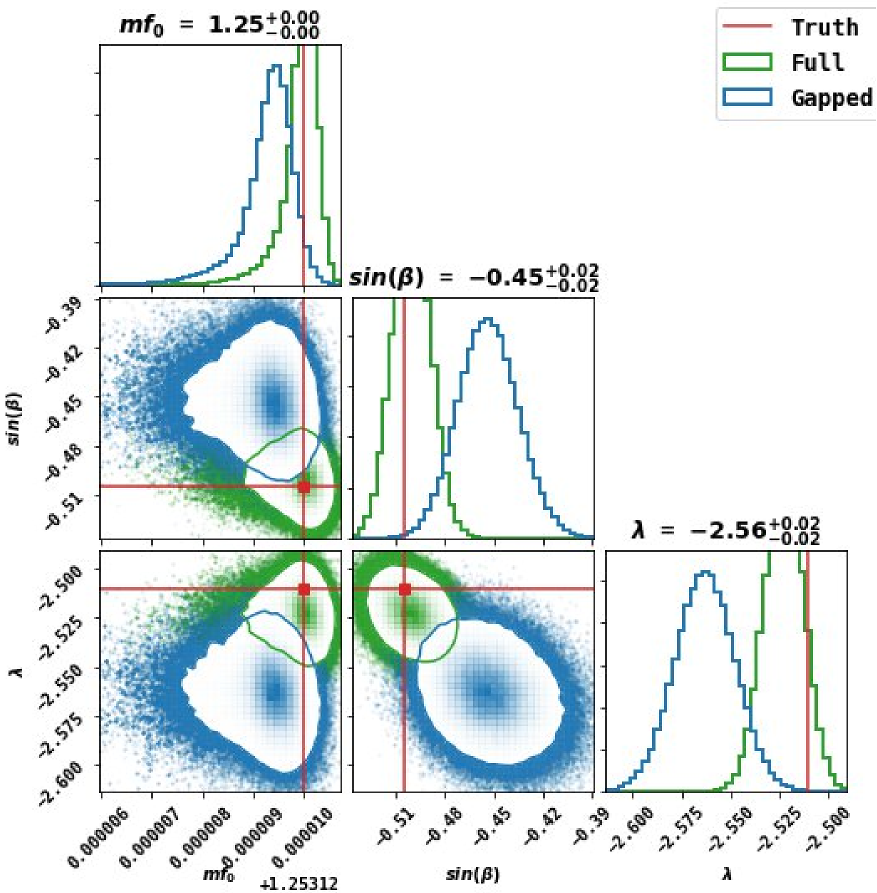


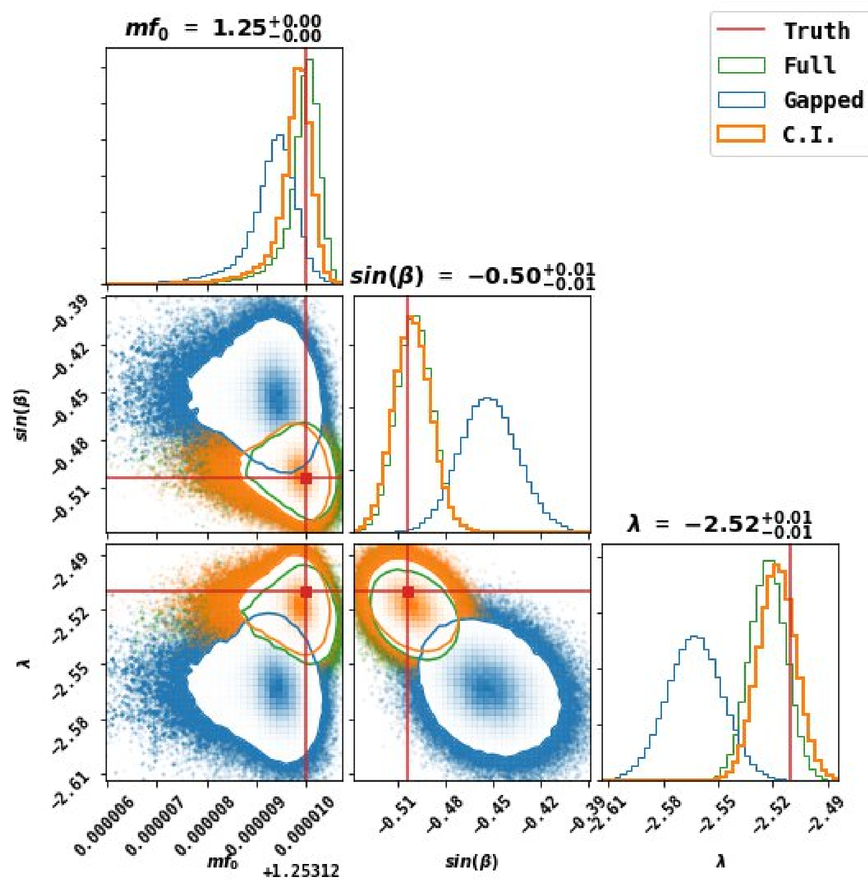
Quality of signal reconstruction

(A. Blelly PhD dissertation)

INTERFACE WITH PARAMETER ESTIMATION

Parameter estimation w/o inpainting (28% data loss)





(A. Blelly PhD dissertation)

NON PARAMETRIC METHODS (1/3)

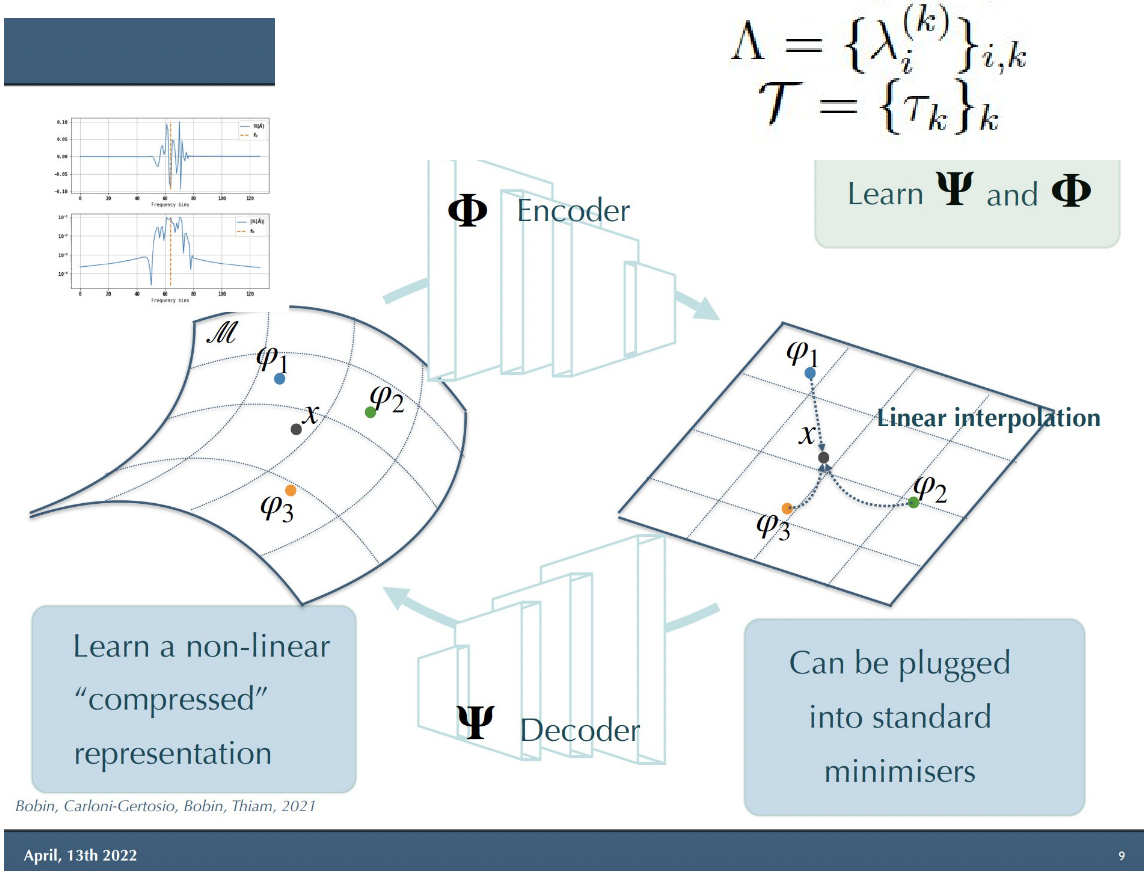
Follow-up on Jérôme's talk

$$\mathcal{A}(\mathrm{s}),\Lambda(\mathrm{s}),\mathcal{T} = \underset{\substack{\mathcal{A}=\{\mathcal{A}^{(k)}\}\\ \Lambda=\{\lambda_i^{(k)}\}_{i,k}\\ \mathcal{T}=\{\tau_k\}_k}}{\mathrm{Argmin}} \left\| \mathrm{s} - \sum_{k=1}^{K} \mathcal{D}_{\tau_k}\left[\mathcal{A}^{(k)} \cdot \Psi\left(\{\lambda_i^{(k)}\}_i\right)\right]\right\|_2^2$$

Shift operator

Amplitude

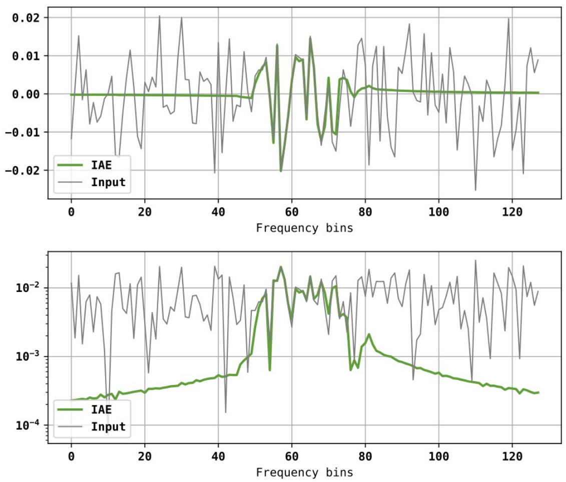
Barycentric coordinates



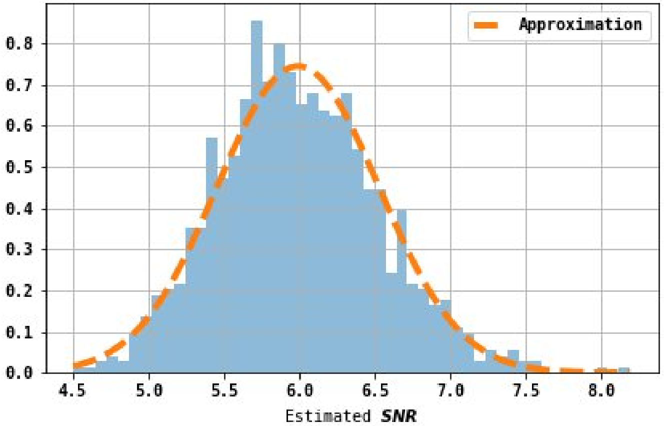
- First attempt : approach can (will ?) be modified later
- Look for exactly K>1 sources

NON PARAMETRIC METHOD (2/3)

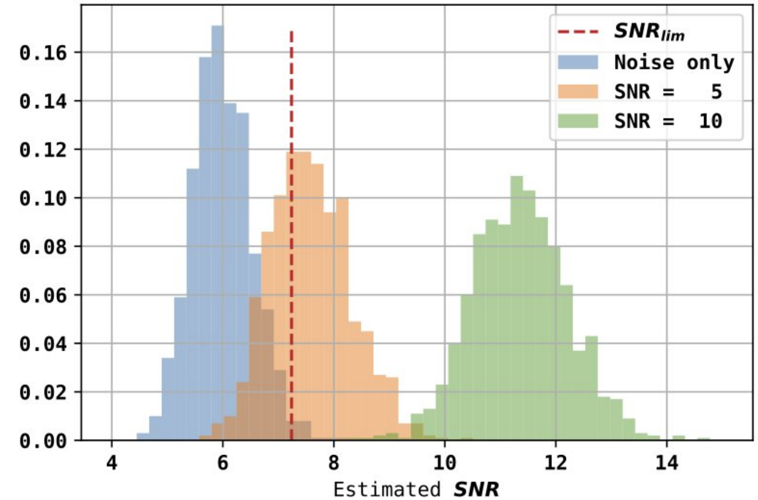
IAE in absence of signal



- Overfitting with IAE : define SNRest and obtain its distribution.
- Define criterion to test presence of signal.
- Detect all signals with SNR>7.
- Marked improvement compared to sparse modeling for low SNR.



(A. Blelly PhD dissertation)



(A. Blelly PhD dissertation)

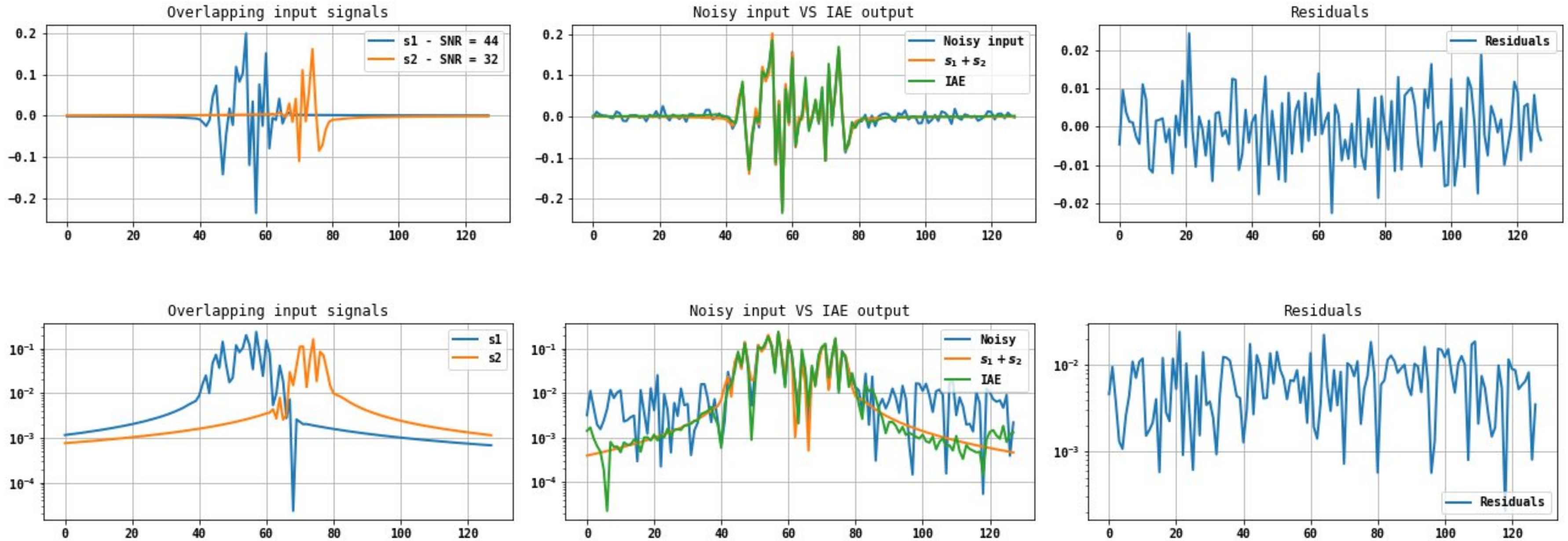
NON PARAMETRIC METHODS (3/3)

Detecting two GBs with IAE

- Consider a signal composed of two GBs of differents SNRs

$$\mathbf{d} = h_0^1(SNR_1)\cdot\mathbf{h}_1 + h_0^2(SNR_2)\cdot\mathbf{h}_2 + \mathbf{n}$$

- Here SNR1 = 44 and SNR2 = 32.



(A. Blelly PhD dissertation)

PROJECTS:
PCA

PCA (Principal Component Analysis) Project

- Idea: develop a complementary, data-driven INREP approach
- Method: singular spectral analysis of the single-link data
- No knowledge of the measurement model nor the delays
- Collaboration with NASA Goddard

arXiv:2010.07224
arXiv:2110.06024

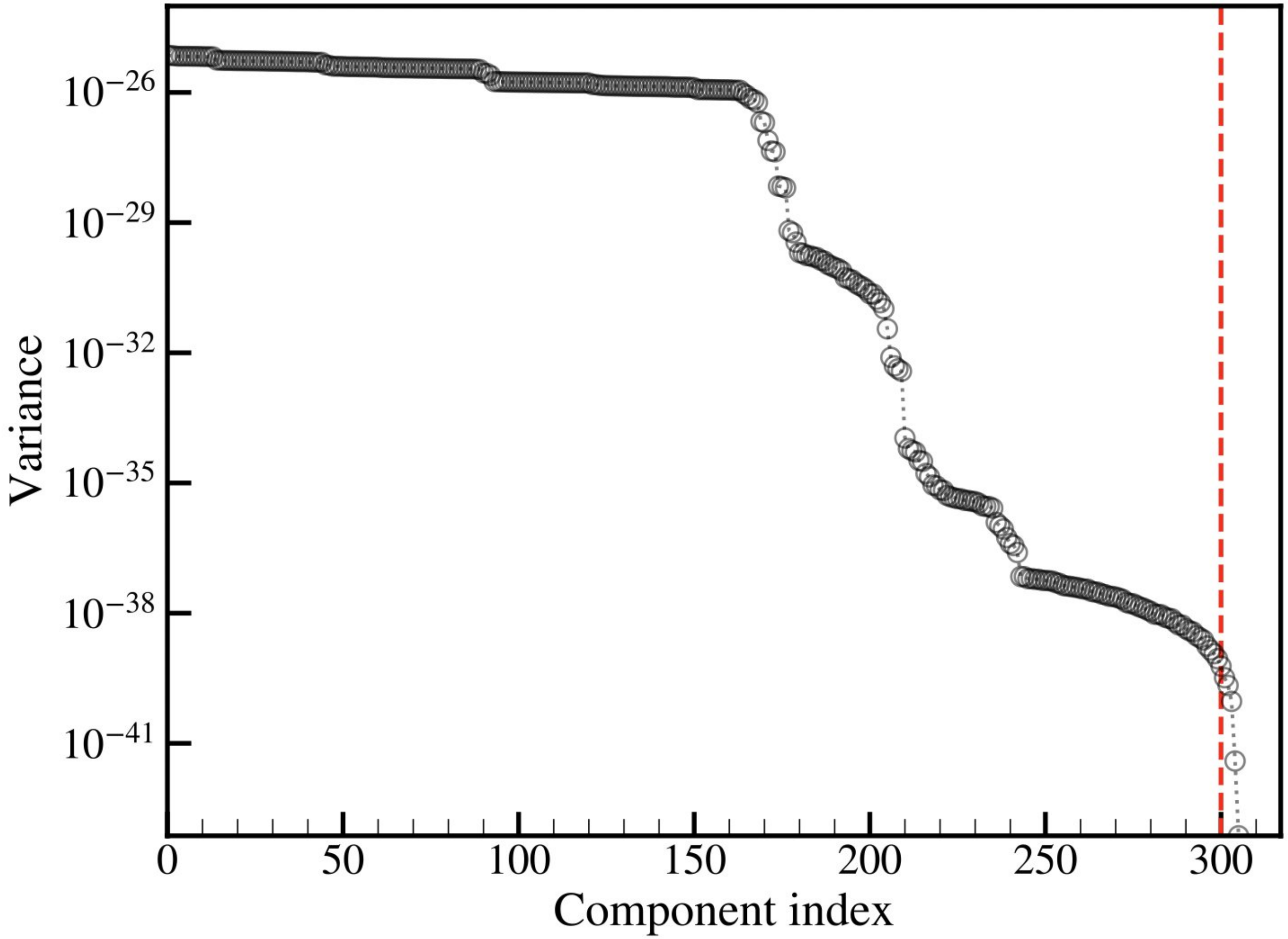
1. $\boldsymbol{Y} \equiv (\text{Science}, \text{Test-mass}, \text{Reference})$ — Stack data

2. $\boldsymbol{X} \equiv (\boldsymbol{D}_{-n_h}\boldsymbol{Y}, \boldsymbol{D}_{-n_h+1}\boldsymbol{Y}, \ldots, \boldsymbol{D}_{n_h}\boldsymbol{Y})$ — Form a data matrix with copies of shifted measurements

3. $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\intercal}$ — Compute PCA

4. $\boldsymbol{T} \equiv \boldsymbol{X}\boldsymbol{V}$ — Project the data onto the singular vectors

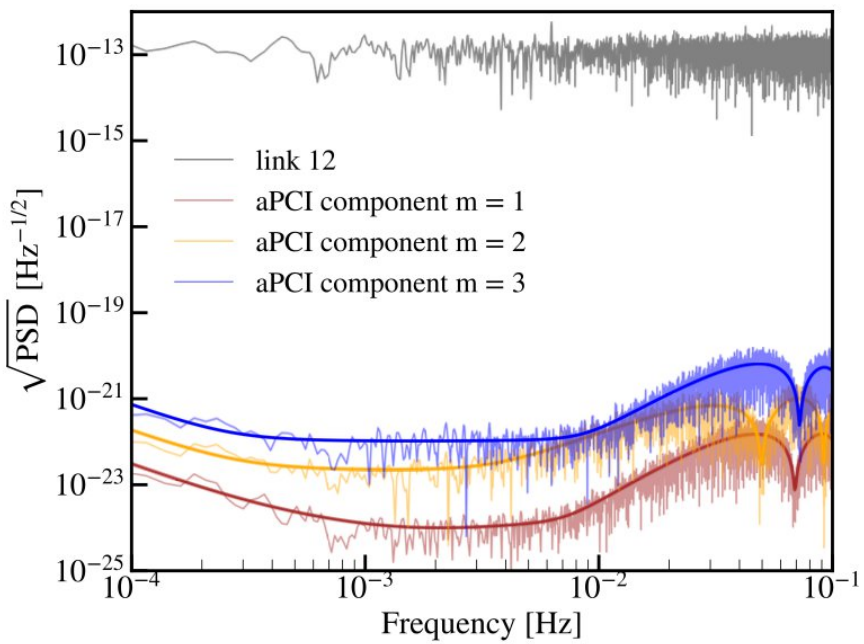
PCA

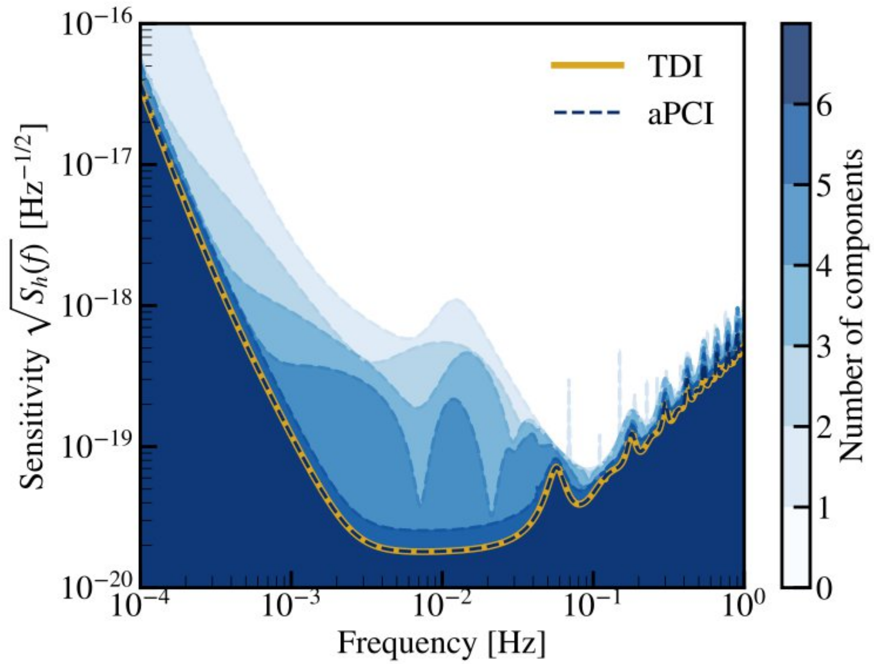


Keep only the lowest variance components

Previous activities and results

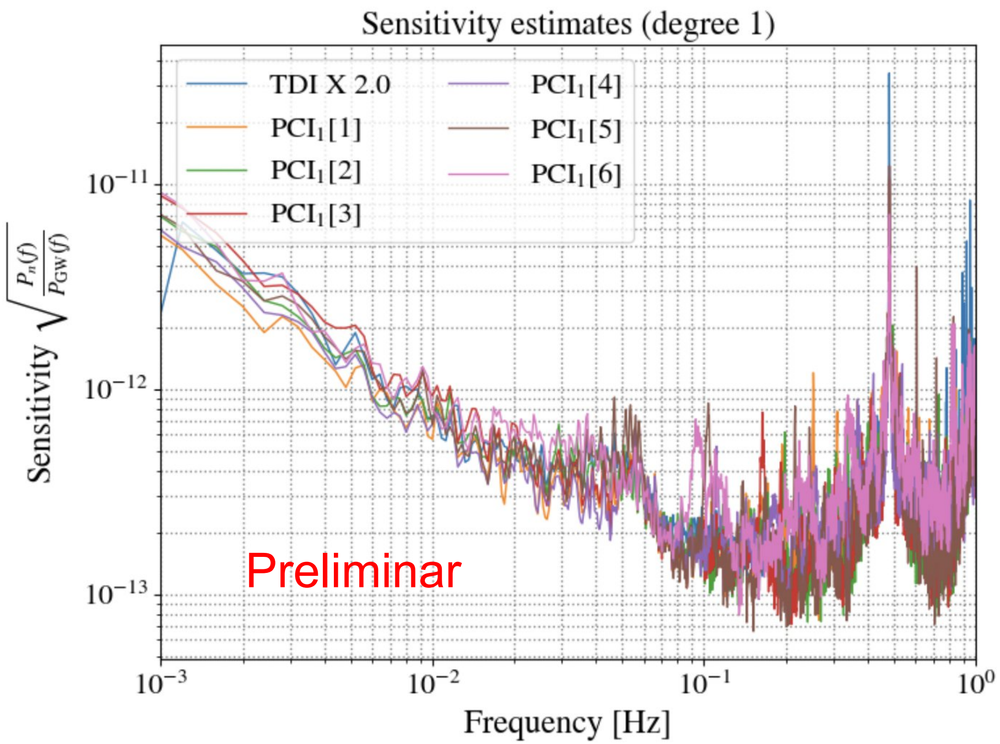
- We demonstrate:
  - Successful cancellation of laser noise for fixed arm lengths
  - Same sensitivity as TDI 1.5





Continuation

- Extend the method to time-varying armlengths
- Method : same as before, but now including higher orders in time



PROJECTS:
SGWB / NOISE PROJECT

Method

- Idea : assess the detectability of SGWBs with minimal assumptions on the instrumental noise shape and level
- Method : describe the noise PSD with B-splines at the single-link level
- Take advantage of GW & noise transfer functions knowledge
- Collaboration with Nikos, Jean-Baptiste, Henri

SGWB term

Noise term

$$\tilde{\boldsymbol{d}}(f_k) = \boldsymbol{R}(f_k)\tilde{\boldsymbol{h}}(f_k) + \boldsymbol{T}(f_k)\tilde{\boldsymbol{n}}(f_k)$$

GW response

Noise transfer function

$$S_h(f) = \Omega_m(f)\frac{3H_0^2}{4\pi^2 f^3}$$

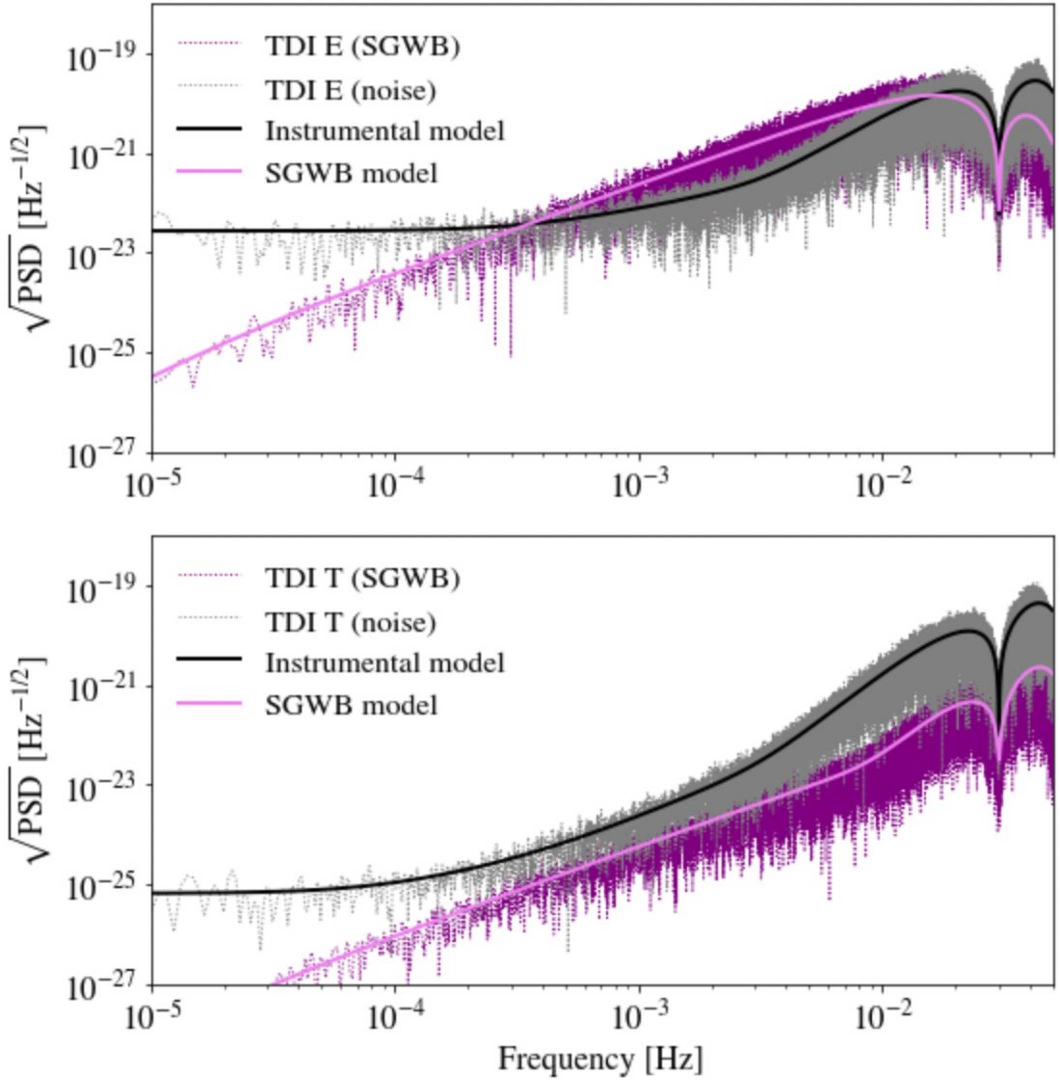
Power law

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

Cubic splines

SGWB / NOISE PROJECT

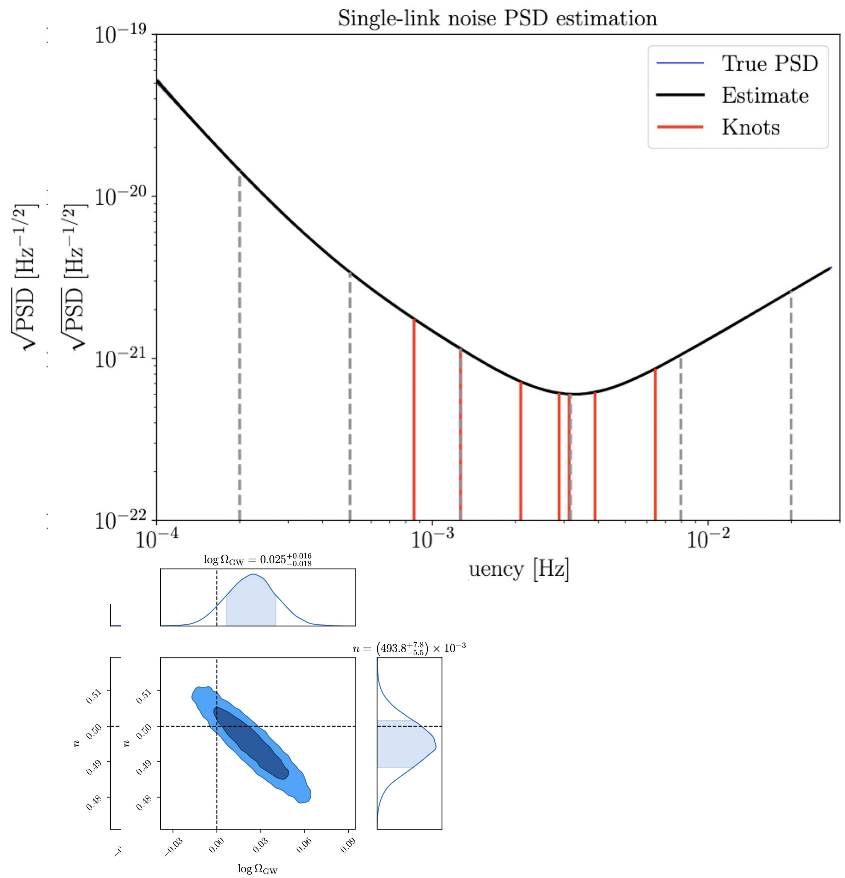
Single-link simulation using LISA GW Response + pyTDI 2.0

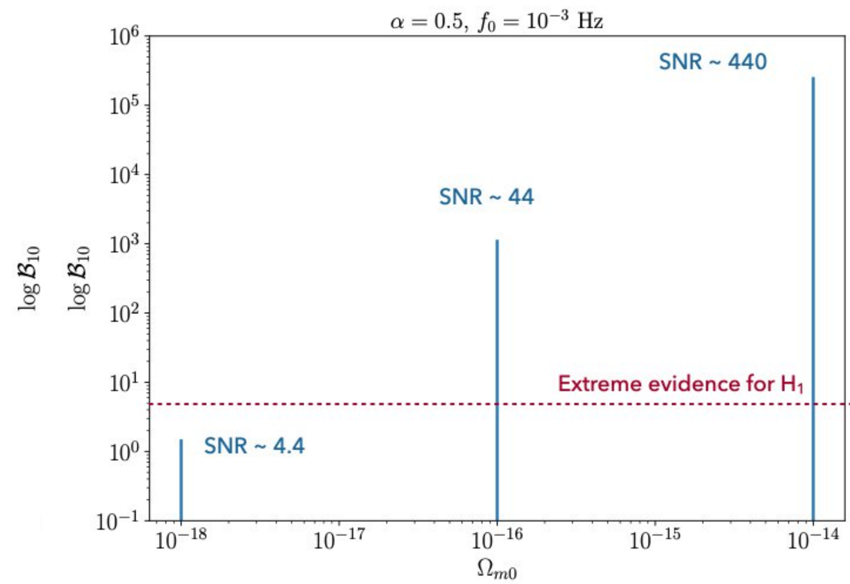


SGWB / NOISE PROJECT

Results

- Demonstrated Bayesian spline noise estimation
- Demonstrated detectability of SGWB in high-SNR cases





SGWB / NOISE PROJECT

Going forwards

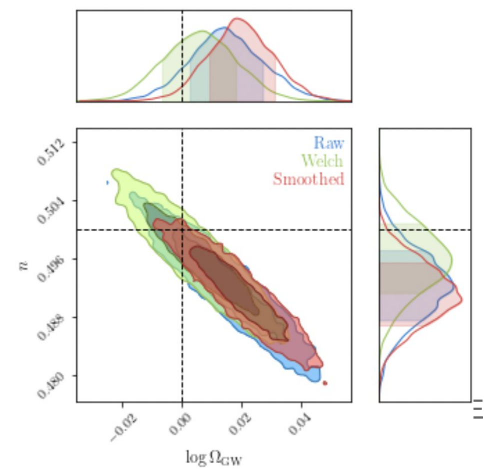
- Provide detectability plot for a larger space of SGWB parameters
- Increase noise realism in simulation + analysis model (collaboration with Olaf Hartwig, Martina Muratore, Nam Dam Quang, Marc Lilley)

$$C(f) = \begin{pmatrix} S_A & 0 & 0 \\ 0 & S_E & 0 \\ 0 & 0 & S_T \end{pmatrix} \rightarrow \begin{pmatrix} S_{XX} & S_{XY} & S_{XZ} \\ S_{YX} & S_{YY} & S_{YZ} \\ S_{ZX} & S_{ZY} & S_{ZZ} \end{pmatrix}$$

- Increase computational efficiency using sufficient statistics

$$\boldsymbol{y} \rightarrow T\left(\boldsymbol{y}\right)$$

$$p\left(\boldsymbol{\theta} | T(\boldsymbol{y})\right) \approx p\left(\boldsymbol{\theta} | \boldsymbol{y}\right)$$

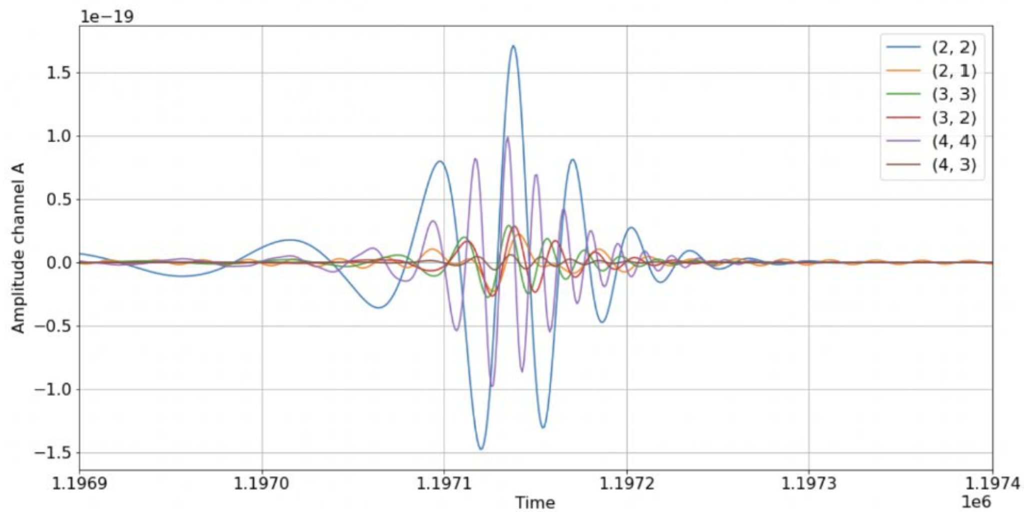


PROJECTS:
QUASI-NORMAL MODE

QNM project

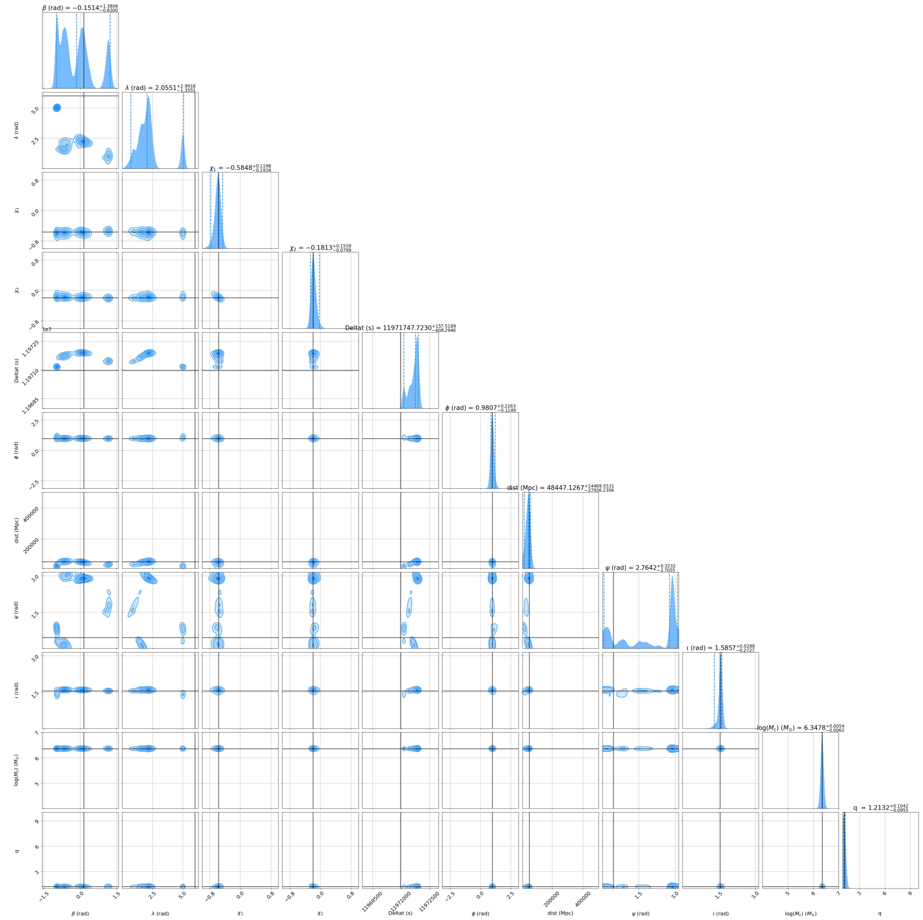
- Idea : assess detectability of quasi-normal modes emitted by merged MBHB
- Evaluate our ability to test the no-hair theorem
- Method: use IMRPhenomHM and Bayesian model comparison
- Collaboration with Sylvain Marsat

$$h(t) = \sum_{l,m,n} A_{lmn} e^{-t/\tau_{lmn}} cos(w_{lmn} t + \phi_{lmn})$$



QNM project

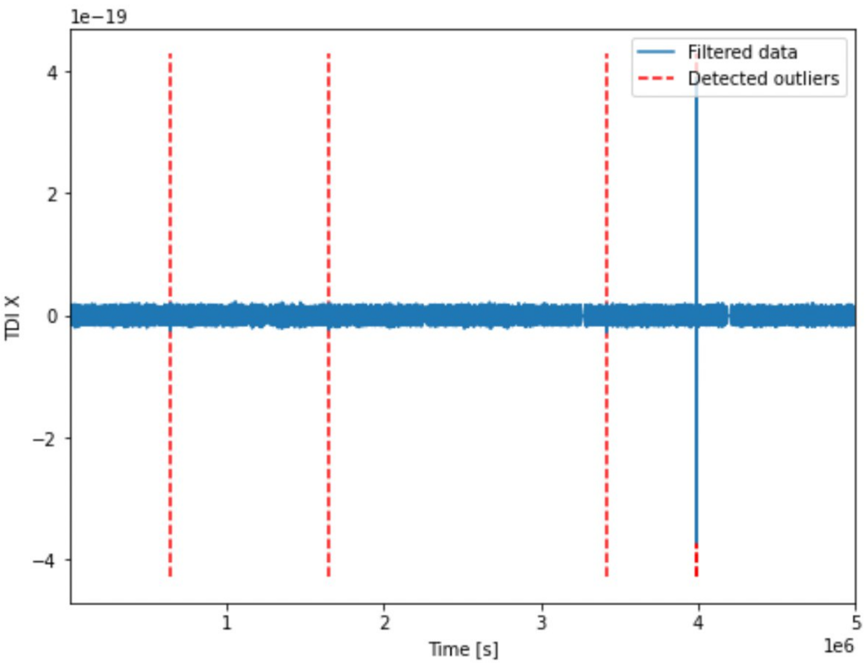
- Analyse de data with different models including 1, 2, ..., n modes
- Injection has 6 modes

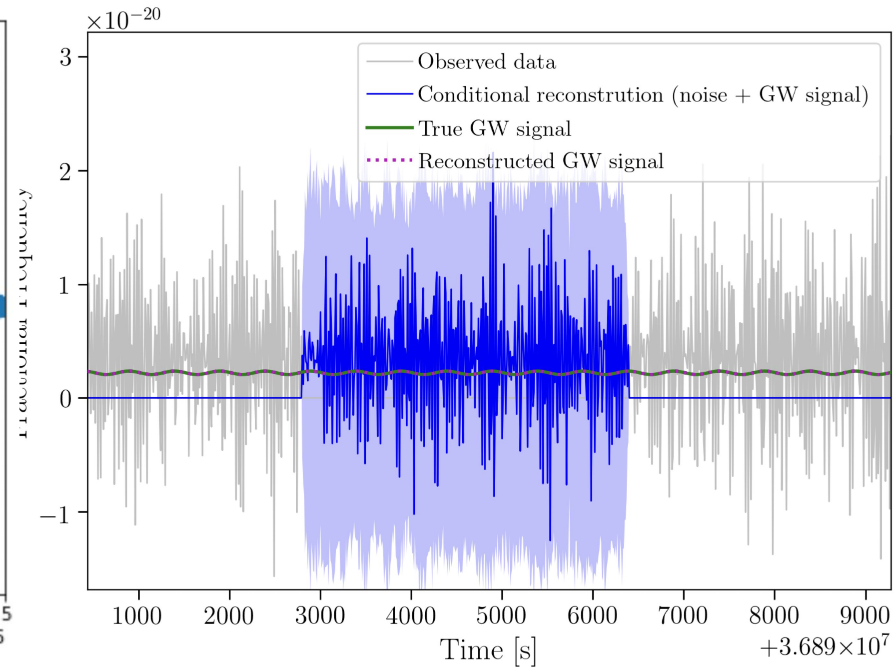


PROJECTS:
GLITCHES & GAPS

Analysis of LDC Spritz data

- Starting project
- Idea : tackle the MBHB + gaps & glitches problem
- Method : mask glitch-perturbed segments and apply data augmentation





FORESEEN EVOLUTION

FORESEEN EVOLUTION

- DA for data with artefacts applied to LDC
  - Sparse / ML approach, gaps mitigation, glitch detection/removal
  - Bayesian augmentation
- Low latency :
  - Sparse / ML approach for source detection/subtraction/identification
- Separation of sources / global fit
  - Fast raw estimates of key parameters to build intermediate priors
  - Refined Bayesian parametric inference
- Development of GB waveforms?
  - tidal effects, magnetisms, ...

Commissariat à l’énergie atomique et aux énergies alternatives
Centre de Saclay | 91191 Gif-sur-Yvette Cedex
T. +33 (0)1 69 08 42 67 | F. +33 (0)1 69 08 50 13

DRF
IRFU

Etablissement public à caractère industriel et commercial | R.C.S Paris B 775 685 019