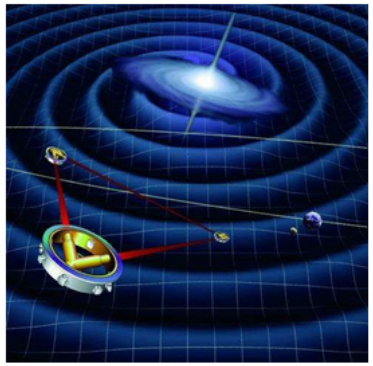
DE LA RECHERCHE À L'INDUSTRIE

CEA - Saclay

Sparse data inpainting for gapped data in LISA

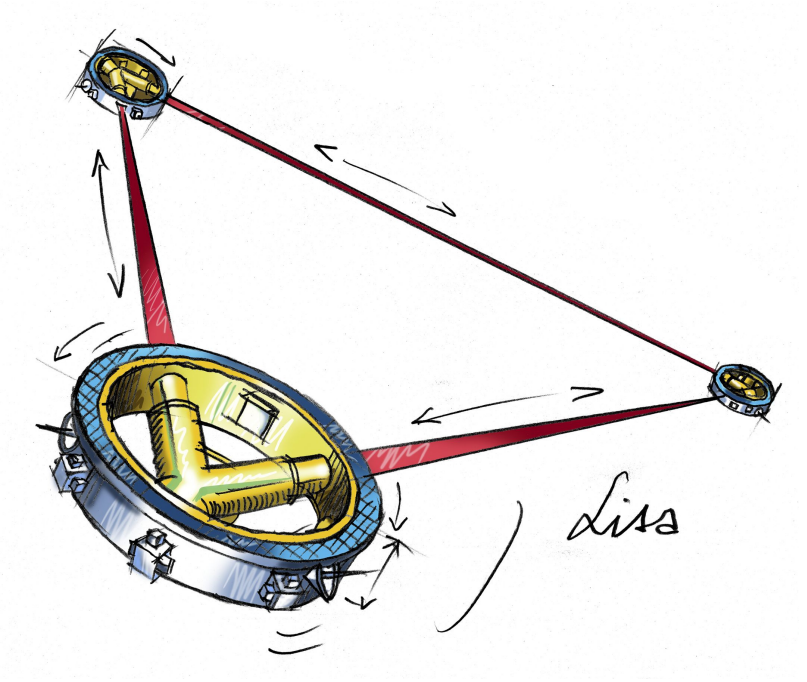


Aurore BLELLY

March 30, 2021

www.cea.fr

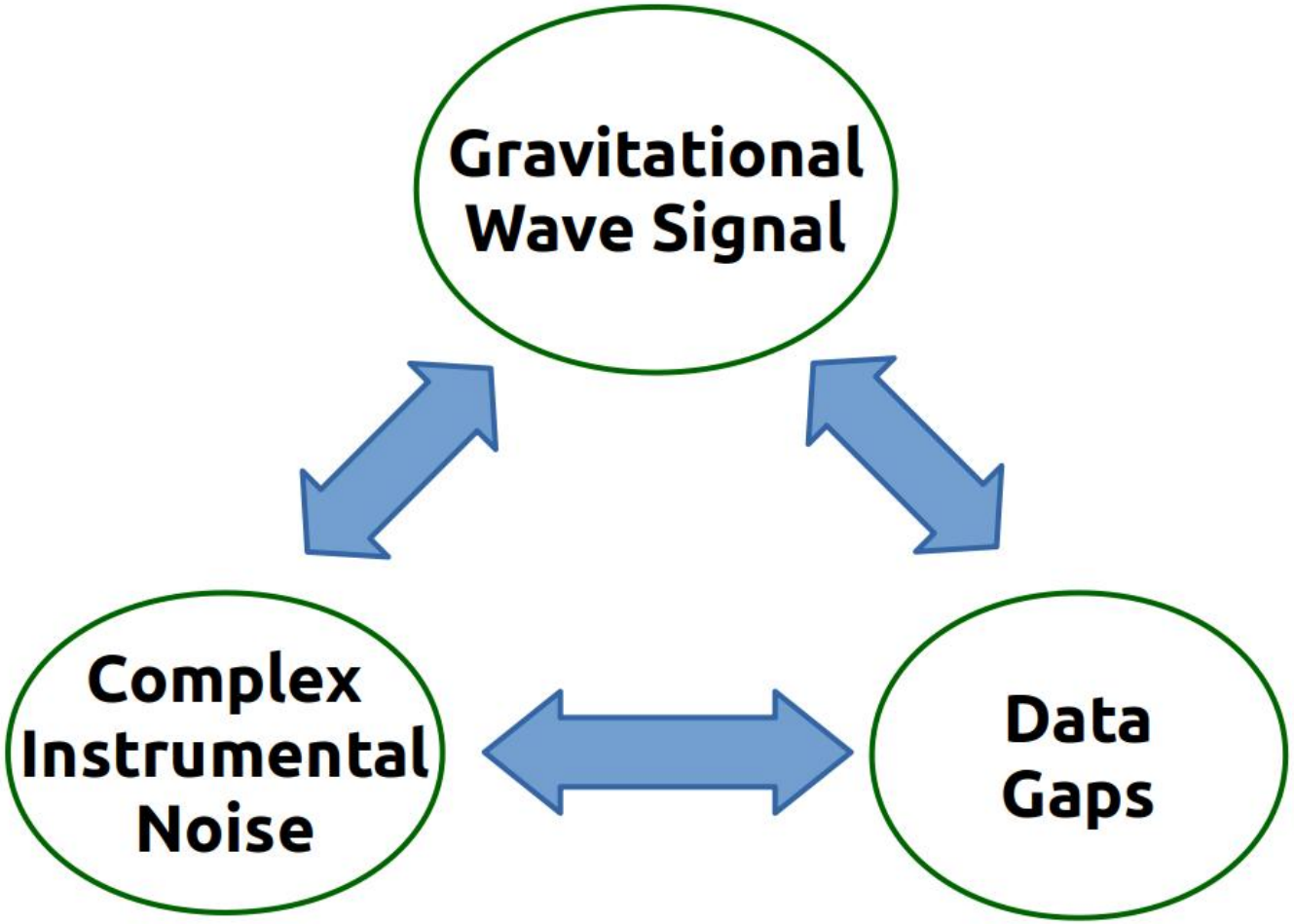
- 3 satellites 2.5 millions km one from another
- Follows the movement of the Earth on its orbit
- Probes a **frequency range that is still unexplored** $\Rightarrow$ high potential for **new discoveries**
- Launching: 2034



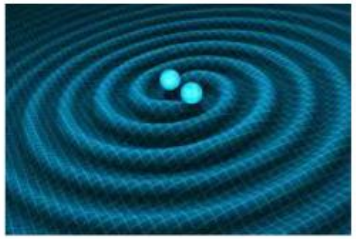
[1]

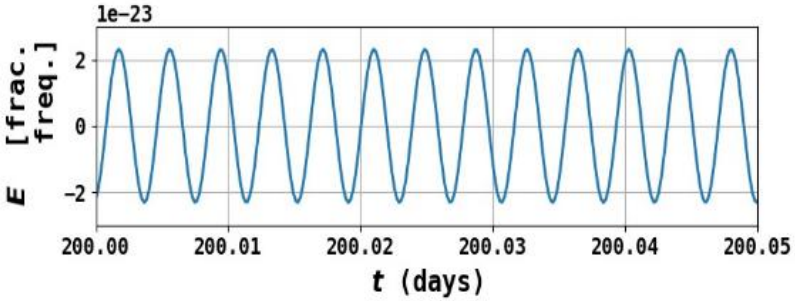
[1] https://www.esa.int/Science_Exploration/Space_Science/LISA

Data Characteristics



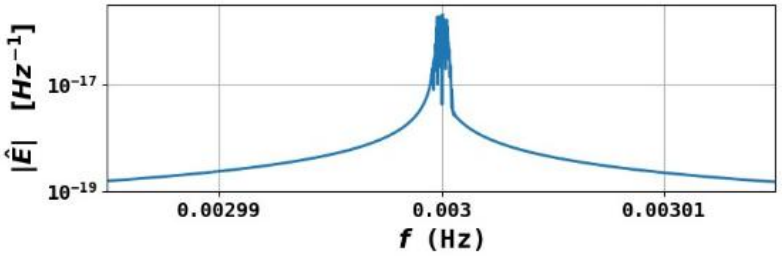
Gravitational Wave Signal



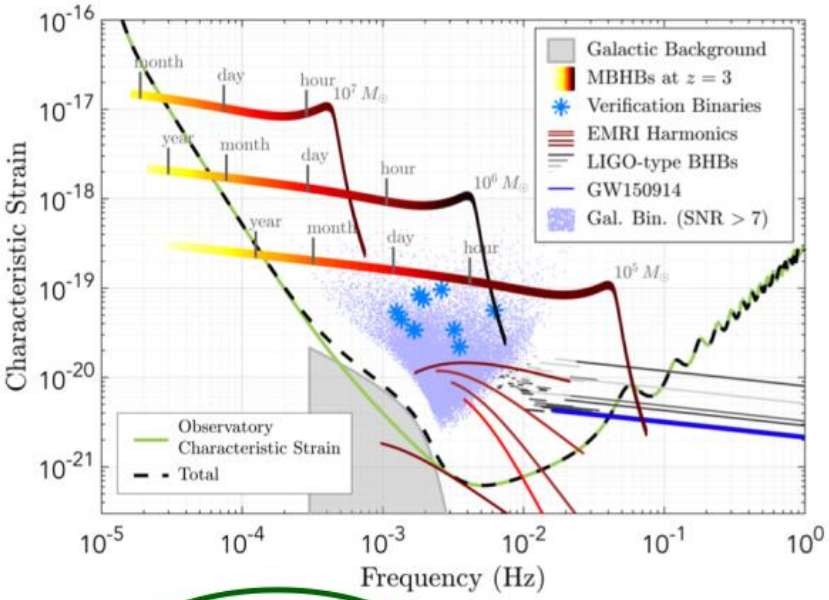


Focus on GBs :

- 10^6 expected sources
- Goal : identify ~ 20.000
- Low SNR = long observation required
- **Neary mono-frequency signal**



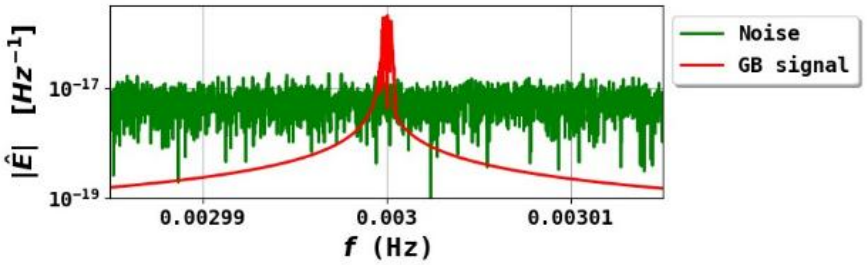
Data Characteristics



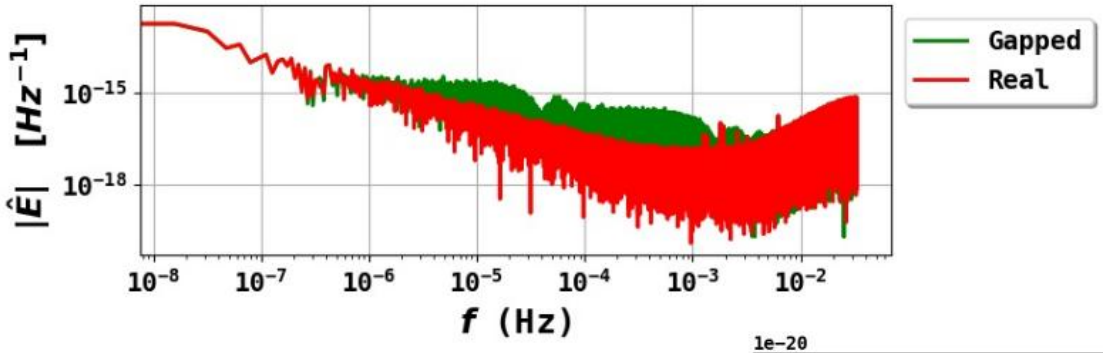
Noise caracteristics :

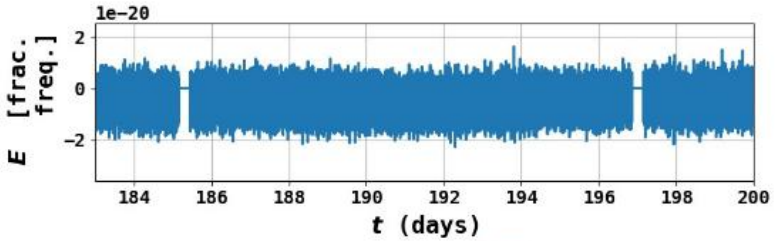
- **Known expected distribution**
- For now, simplified hypothesis (Gaussianity)
- Noise level impacts the ability to detect a source

Complex Instrumental Noise



Data Characteristics





Data gaps :

- Maintenance or unplanned
- Various durations/frequencies :
  - Planned : 7 hours / 2 weeks
  - Unplanned : a few seconds => several days
- **Impact expected noise distribution**
- **Impact** quality of identification

Data Gaps

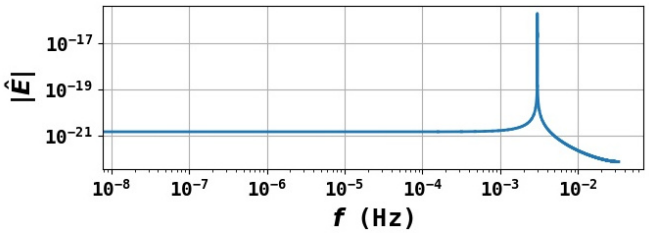
- **Goal:** Mitigating the impact of gaps in LISA data by filling the gaps with as little new information as possible.
- **State of the art:** Data augmentation method [2] : based on sampling and MCMC.
- **Strategy:** Using the **waveform expected structure** instead of **parametric description**.

$\Rightarrow$ **Non-parametric** method, no GB identification.

[2] Q. Baghi, I. Thorpe, J. Slutsky, J. Baker, T. Dal Canton, N. Korsakova, N. Karnesis, *Gravitational-wave parameter estimation with gaps in LISA: a Bayesian data augmentation method*, 2019

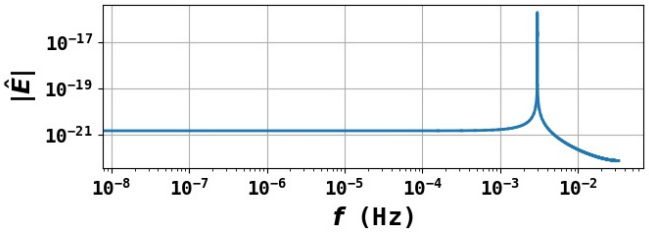
Prior hypothesis:

- Simple representation in Fourier basis (**Sparsity prior**)

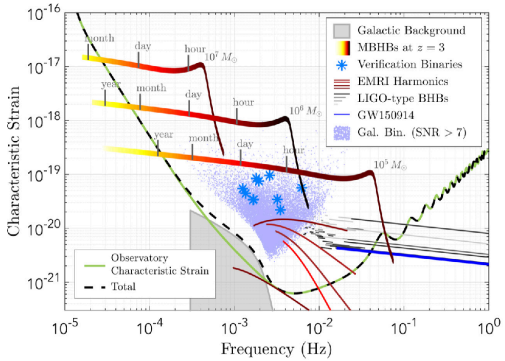


Prior hypothesis:

- Simple representation in Fourier basis (**Sparsity prior**)

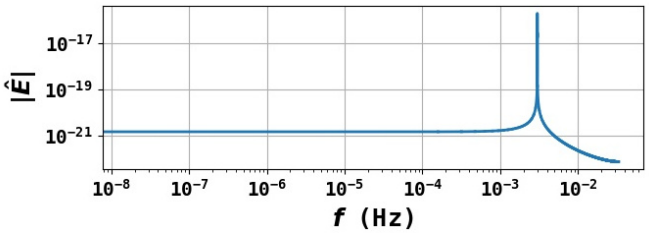


- Noise is Gaussian with known PSD Σ **for ungapped data**

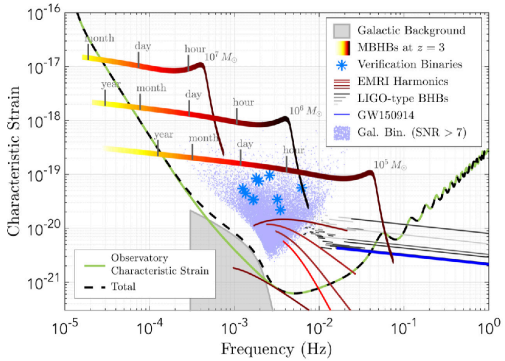


Prior hypothesis:

- Simple representation in Fourier basis (**Sparsity prior**)



- Noise is Gaussian with known PSD Σ **for ungapped data**



- Gap time is known:

Sparse Data Inpainting

Sparse Data Inpainting Method [3]

Cost function with data gaps

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) =$$

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

Sparse Data Inpainting Method [3]

Cost function with data gaps

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) = \underbrace{\underset{Y=M(S+N)}{\mathrm{Argmin}}}_{\text{Data Adequacy}} \Bigg[$$

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

Sparse Data Inpainting Method [3]

Cost function with data gaps

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) = \underbrace{\underset{Y=M(S+N)}{\mathrm{Argmin}}}_{\text{Data Adequacy}} \Bigg[\underbrace{\left\| \gamma \odot \Sigma^{-1/2} \widehat{S} \right\|_{12}}_{\text{Solution prior Sparsity}} +$$

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) = \underbrace{\underset{Y=M(S+N)}{\mathrm{Argmin}}}_{\substack{\text{Data}\\ \text{Adequacy}}} \left[\underbrace{\left\| \gamma \odot \Sigma^{-1/2} \widehat{S} \right\|_{12}}_{\substack{\text{Solution prior}\\ \text{Sparsity}}} + \underbrace{\frac{1}{2} \widehat{N}^T \Sigma^{-1} \widehat{N}}_{\substack{\text{Noise prior}\\ \text{Gaussian dist.}}} \right]$$

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) = \underbrace{\underset{Y=M(S+N)}{\mathrm{Argmin}}}_{\substack{\text{Data}\\ \text{Adequacy}}} \left[\underbrace{\left\| \gamma \odot \Sigma^{-1/2} \widehat{S} \right\|_{12}}_{\substack{\text{Solution prior}\\ \text{Sparsity}}} + \underbrace{\frac{1}{2} \widehat{N}^T \Sigma^{-1} \widehat{N}}_{\substack{\text{Noise prior}\\ \text{Gaussian dist.}}} \right]$$

With Y the gapped measurement; Σ the noise PSD and γ the detection threshold.

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

- M is the data mask: 

- (S, N) the joint estimate of **ungapped** signal and noise

$$(S, N) = \underbrace{\underset{Y=M(S+N)}{\mathrm{Argmin}}}_{\substack{\text{Data}\\ \text{Adequacy}}} \left[\underbrace{\left\| \gamma \odot \Sigma^{-1/2} \widehat{S} \right\|_{12}}_{\substack{\text{Solution prior}\\ \text{Sparsity}}} + \underbrace{\frac{1}{2} \widehat{N}^T \Sigma^{-1} \widehat{N}}_{\substack{\text{Noise prior}\\ \text{Gaussian dist.}}} \right]$$

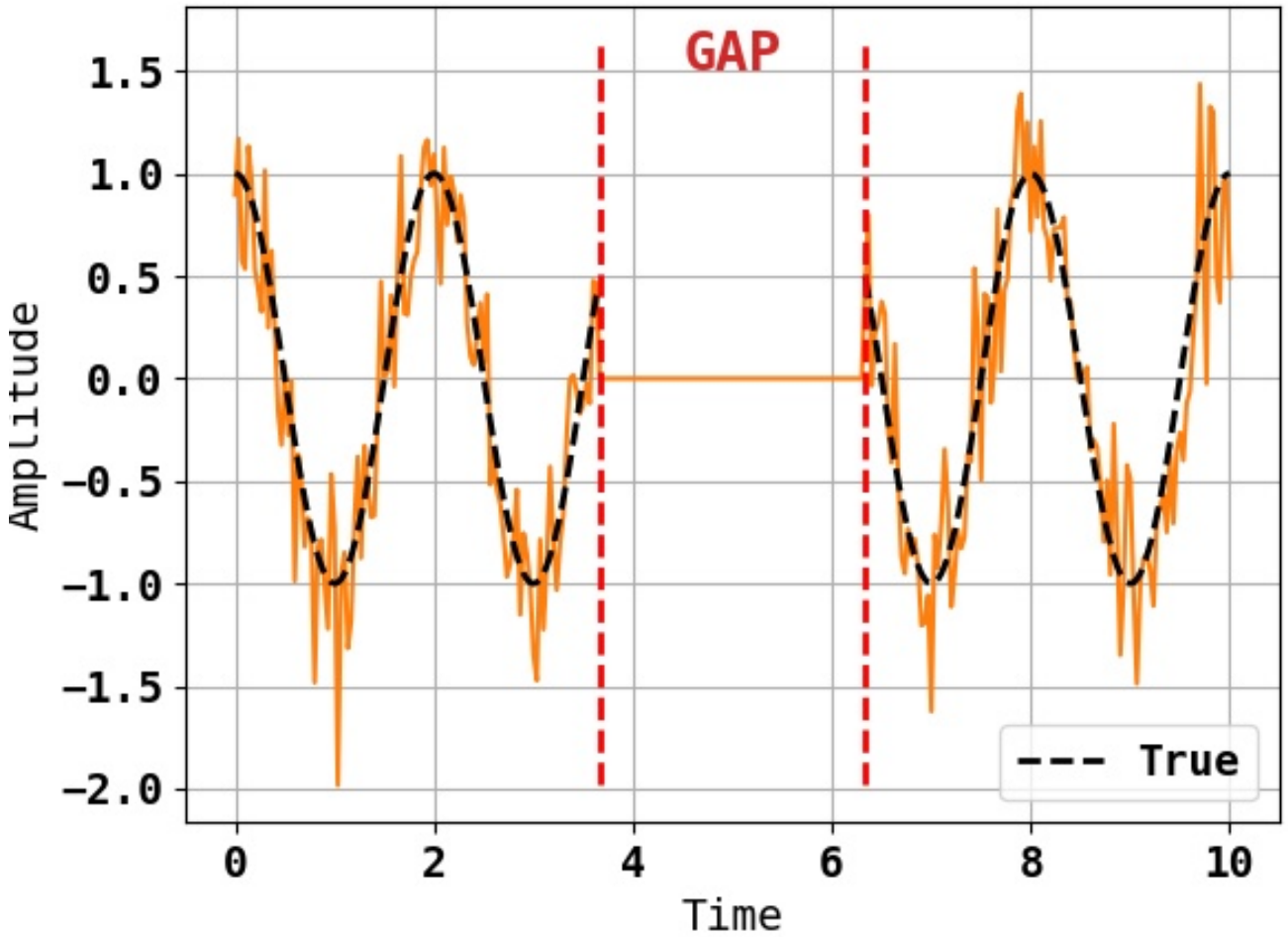
With Y the gapped measurement; Σ the noise PSD and γ the detection threshold.

- **Resolution:** alternates between signal estimation and noise estimation.

[3] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]

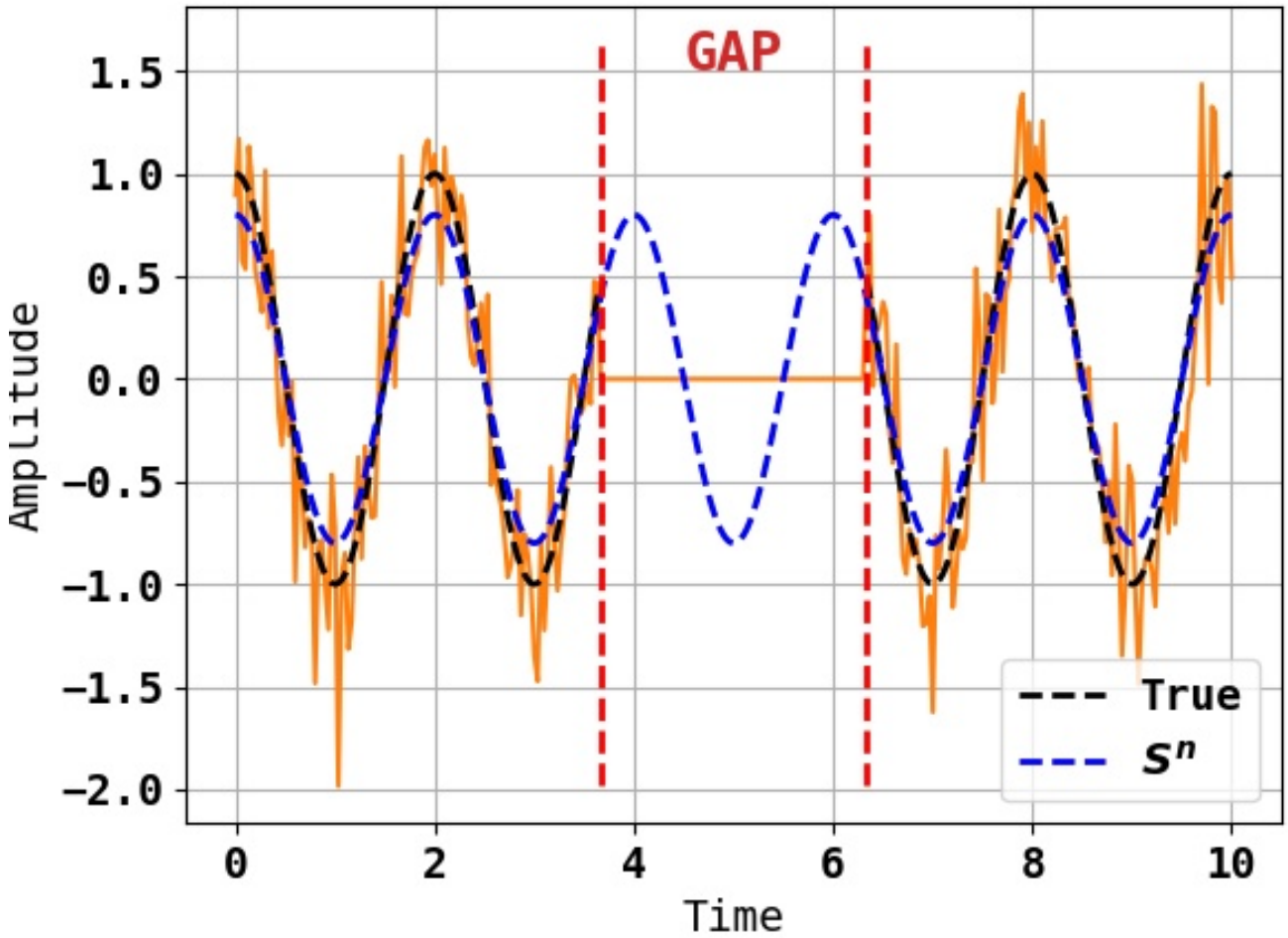
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



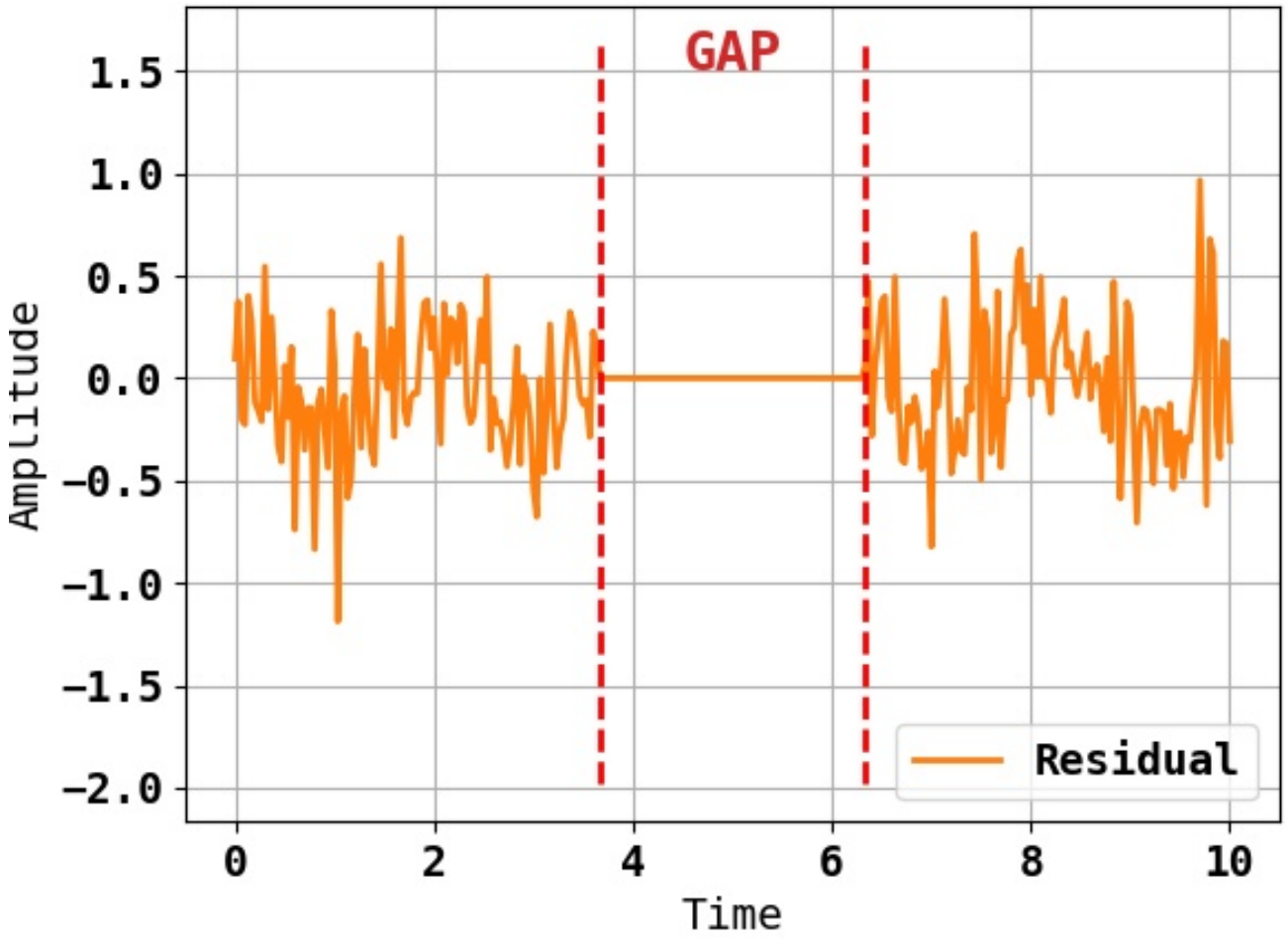
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



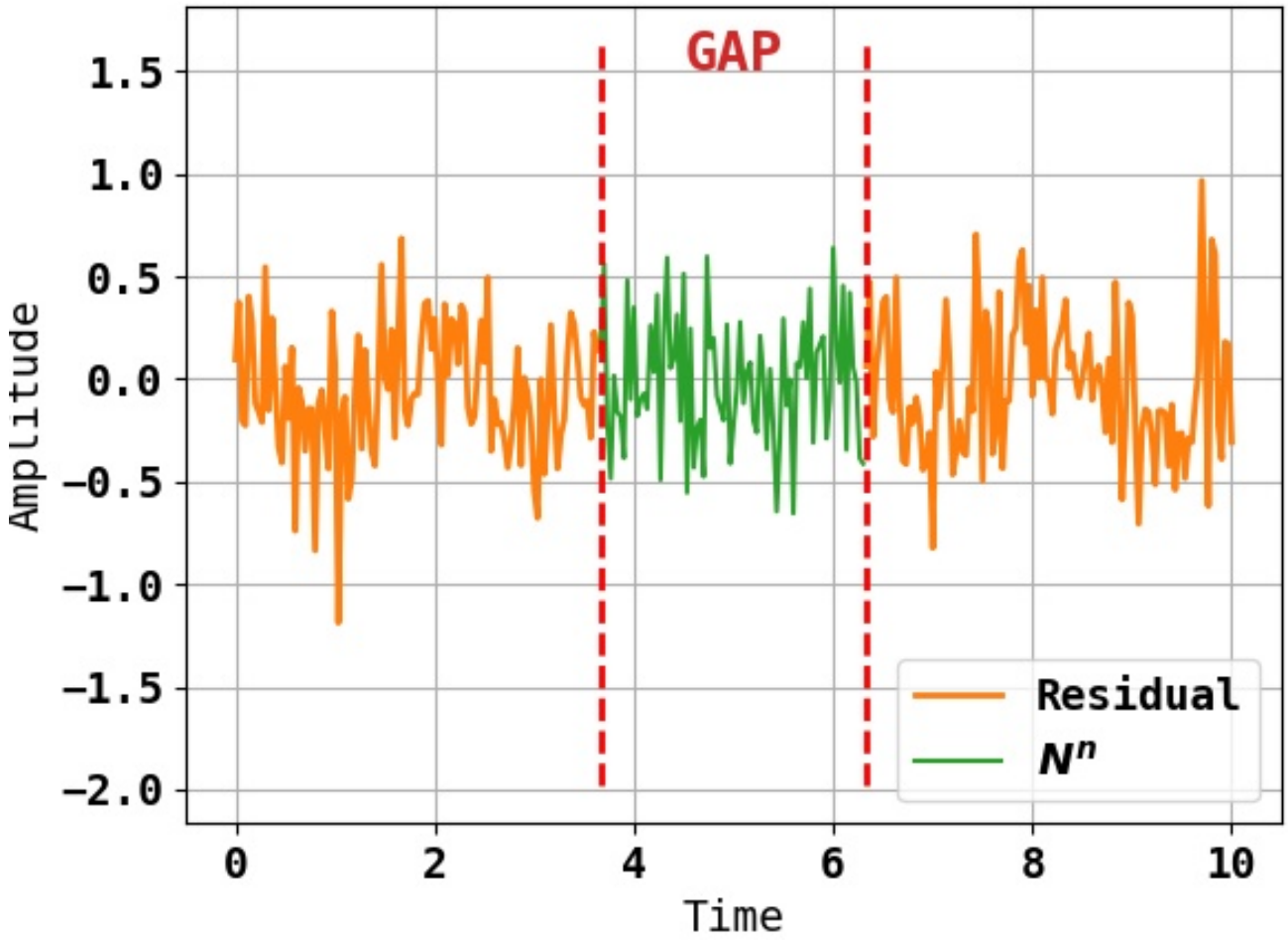
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



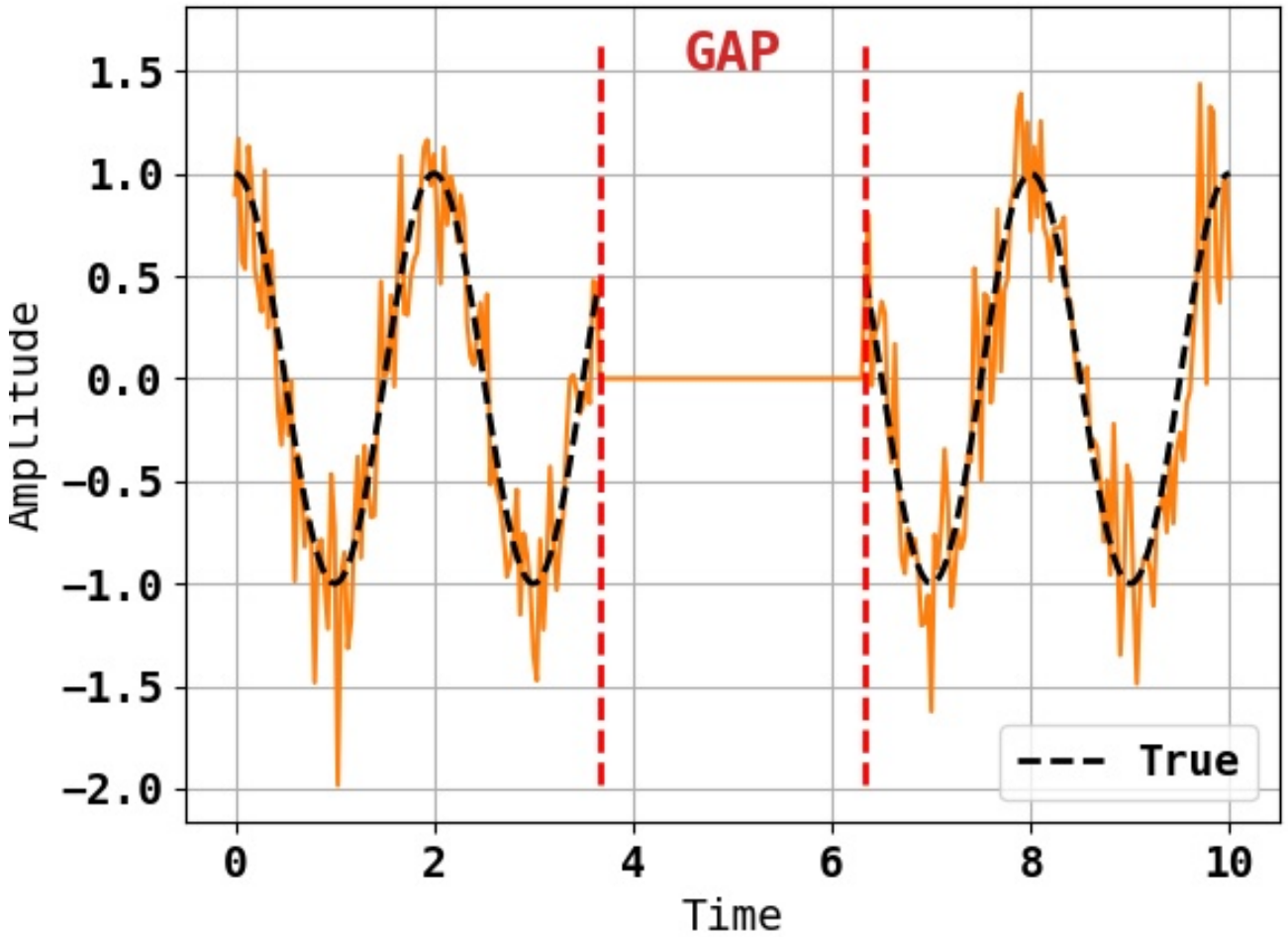
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



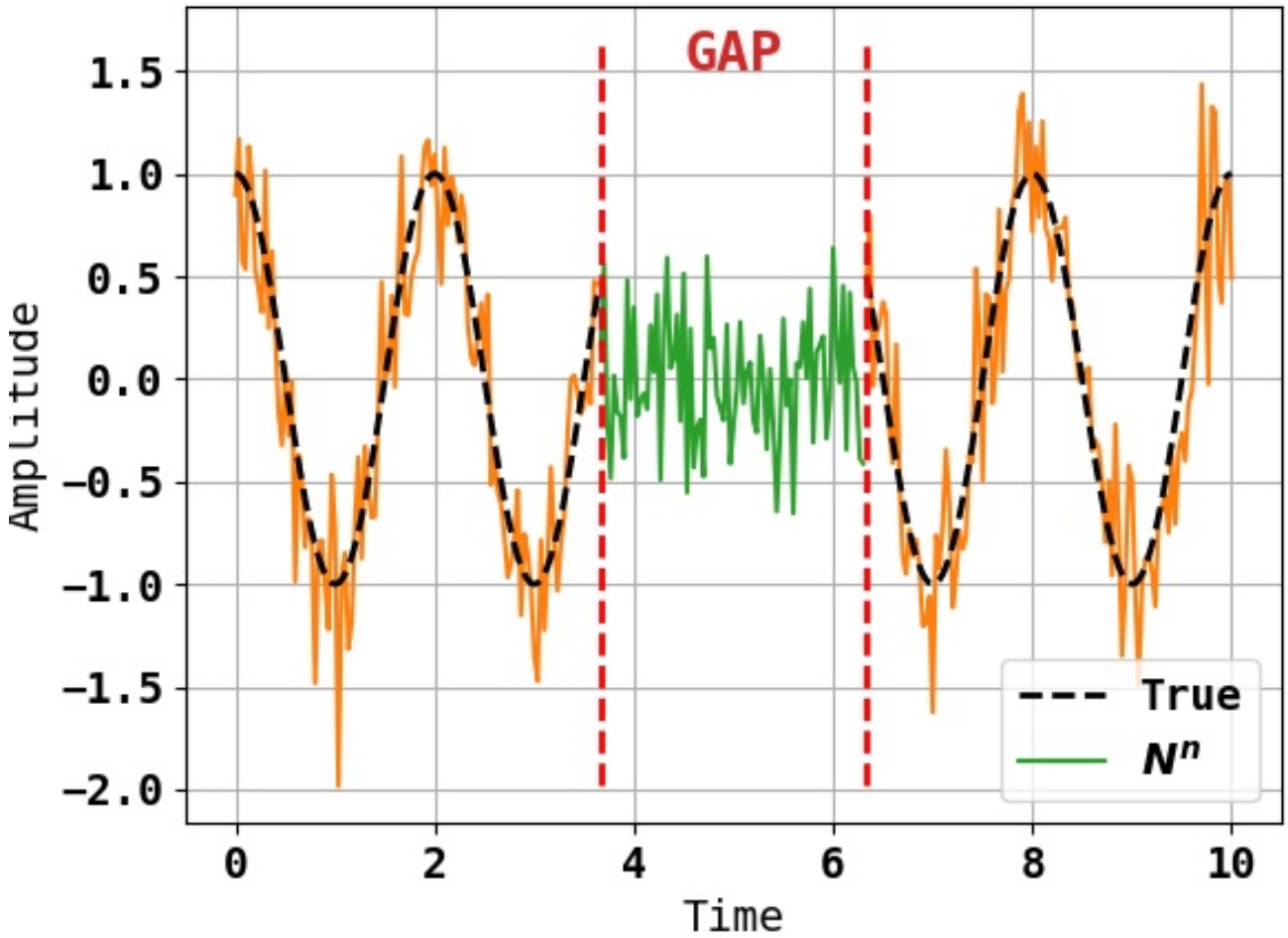
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



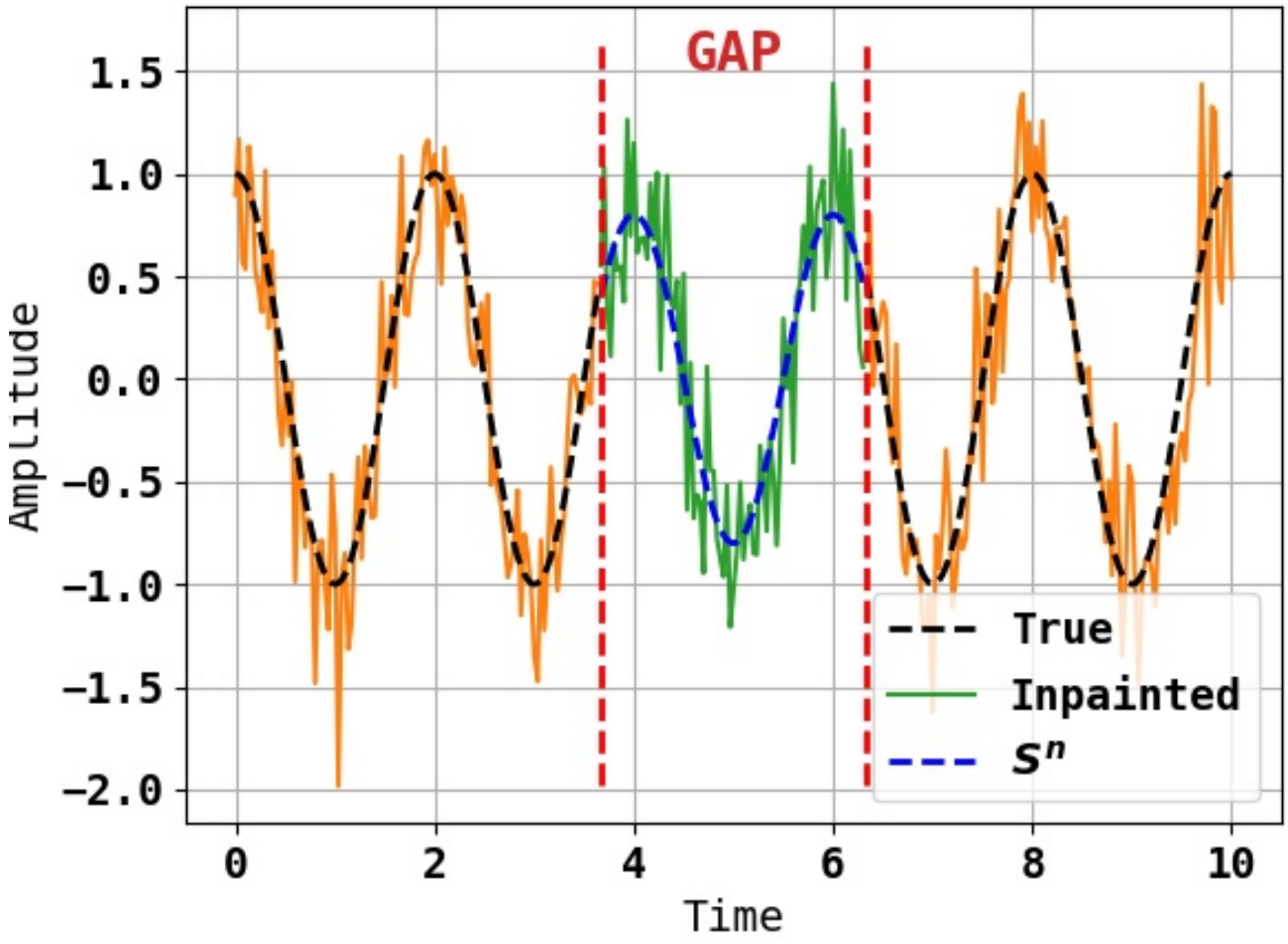
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



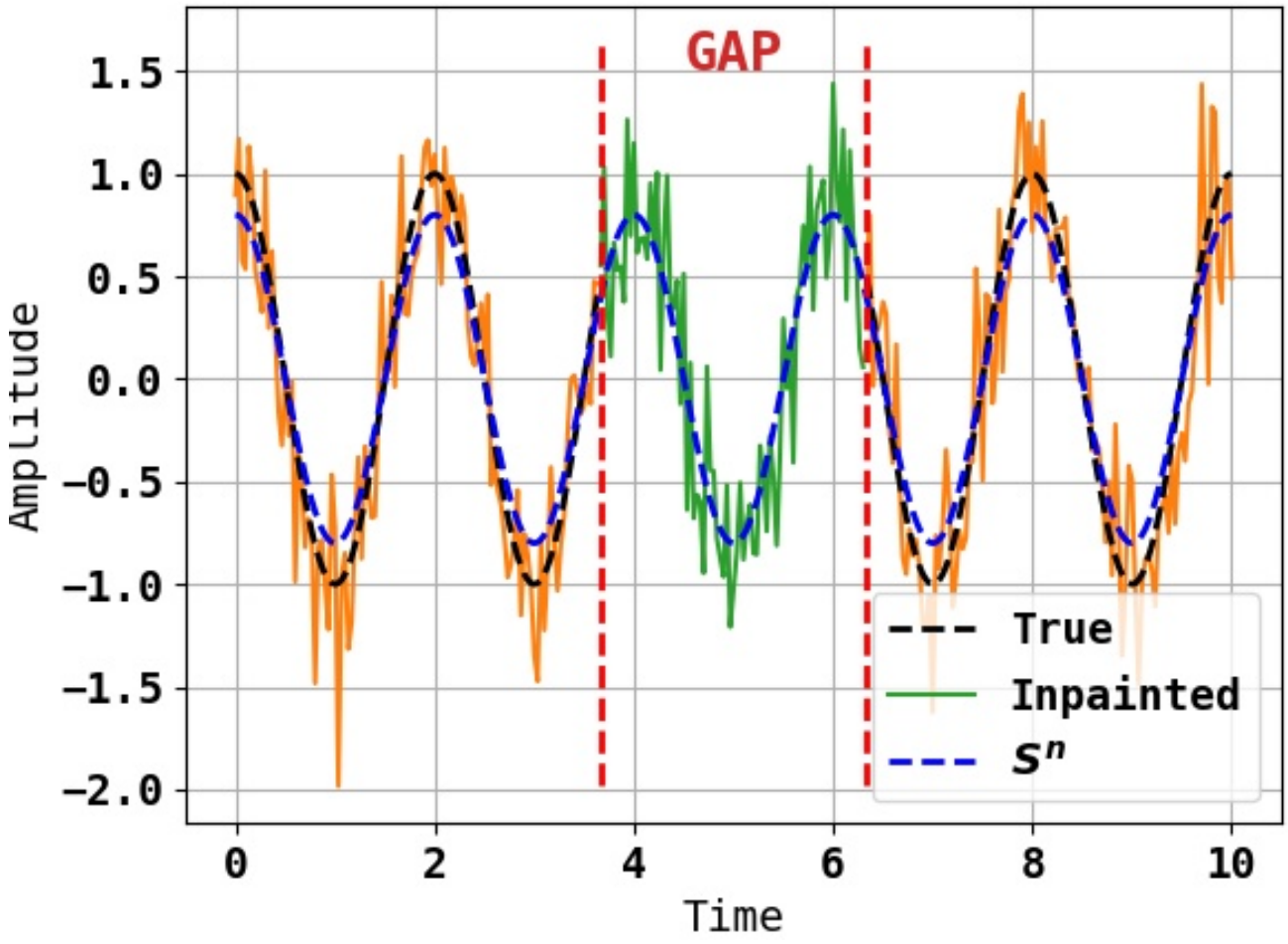
Iterative Inpainting Algorithm

Progress of an iteration on a simple case



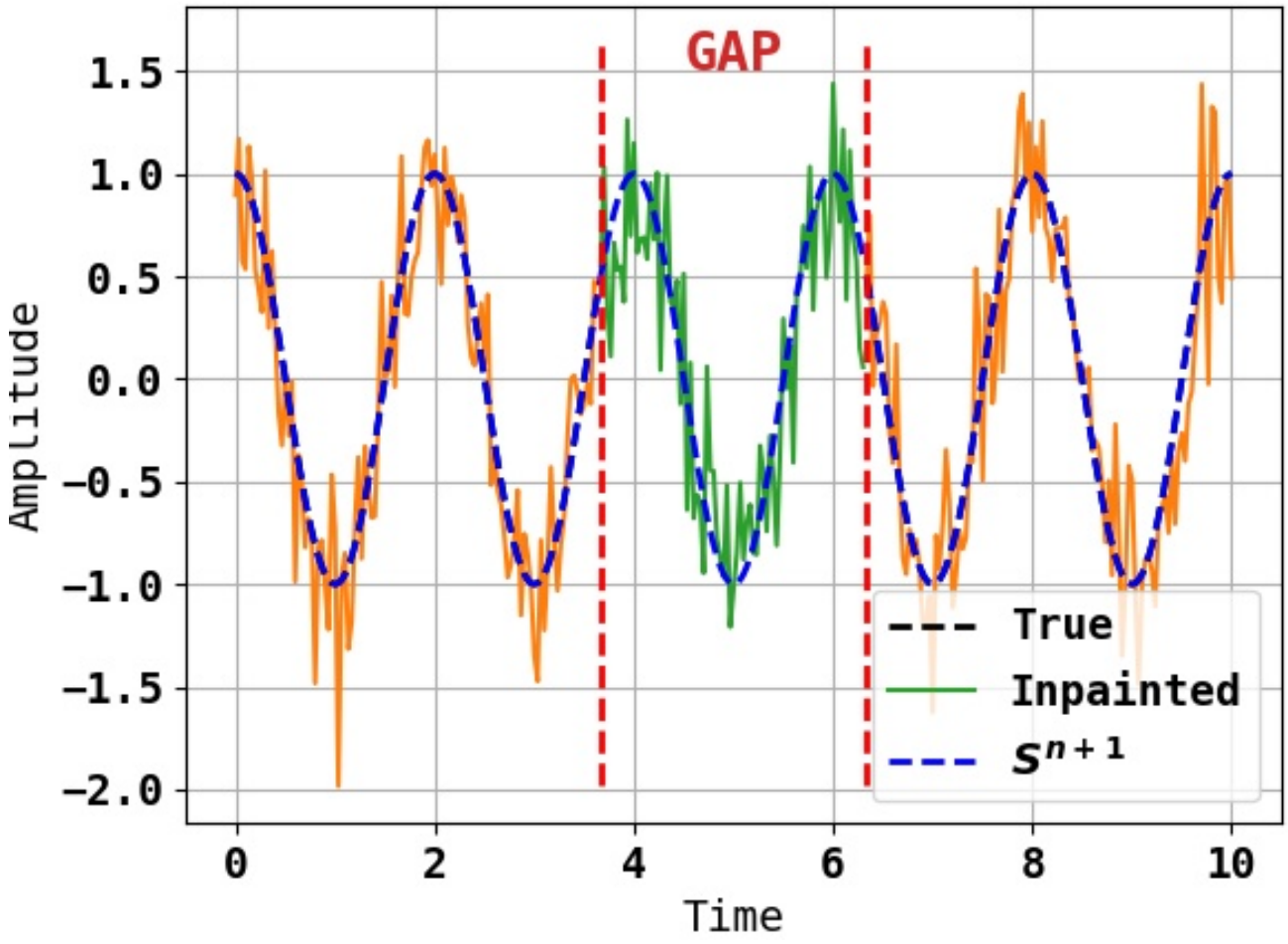
Iterative Inpainting Algorithm

Progress of an iteration on a simple case

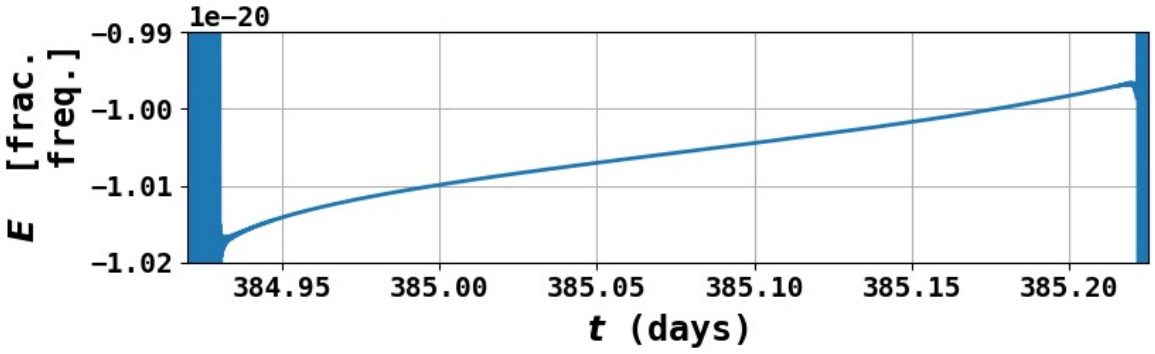


Iterative Inpainting Algorithm

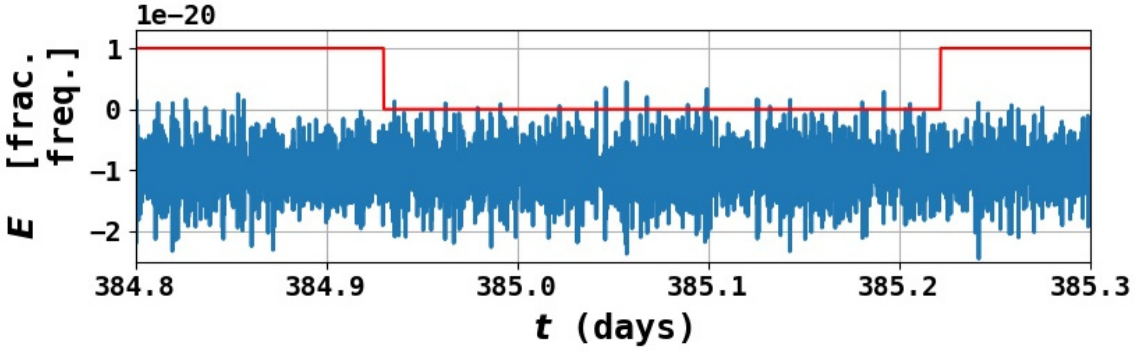
Progress of an iteration on a simple case



Classical Inpainting (**C.I.**): only retrieves lower-frequency noises



Modified Inpainting (**M.I.**): restores higher frequency noises using a stochastic component



Results

Performance Assessment

The performances of our algorithms are assessed on three main outcomes:

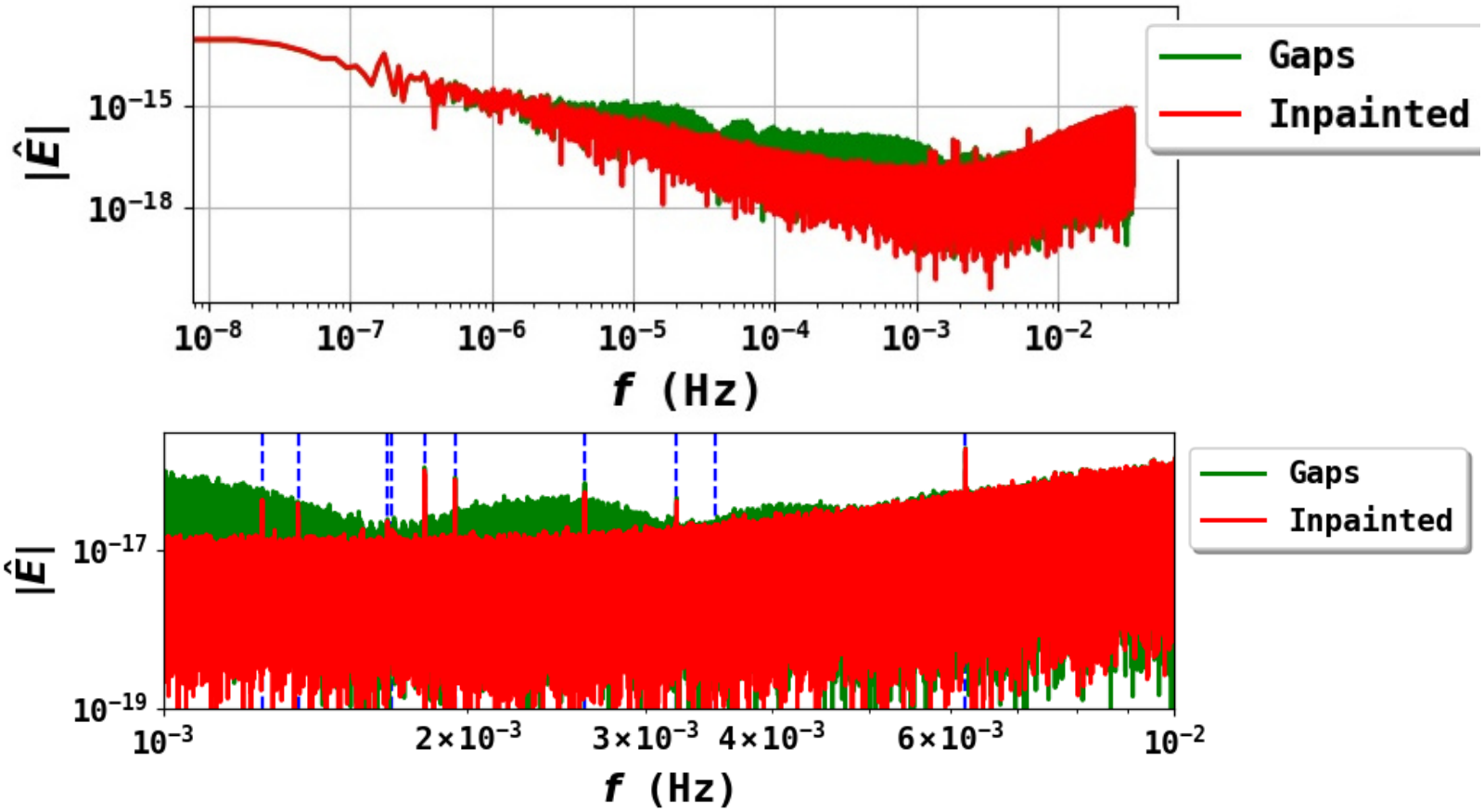
1. the detection capacity (comparatively to the ungapped case)
2. the recovered noise statistics
3. the quality of extracted signal

On 3 types of gaps:

Small (**S**): 10min/day
Medium (**M**): 7hours/2 weeks
Large (**L**): 3 days / 12 days

Sparse Data Inpainting Method

Impact on input signal spectrum - LDC1-3 [4]

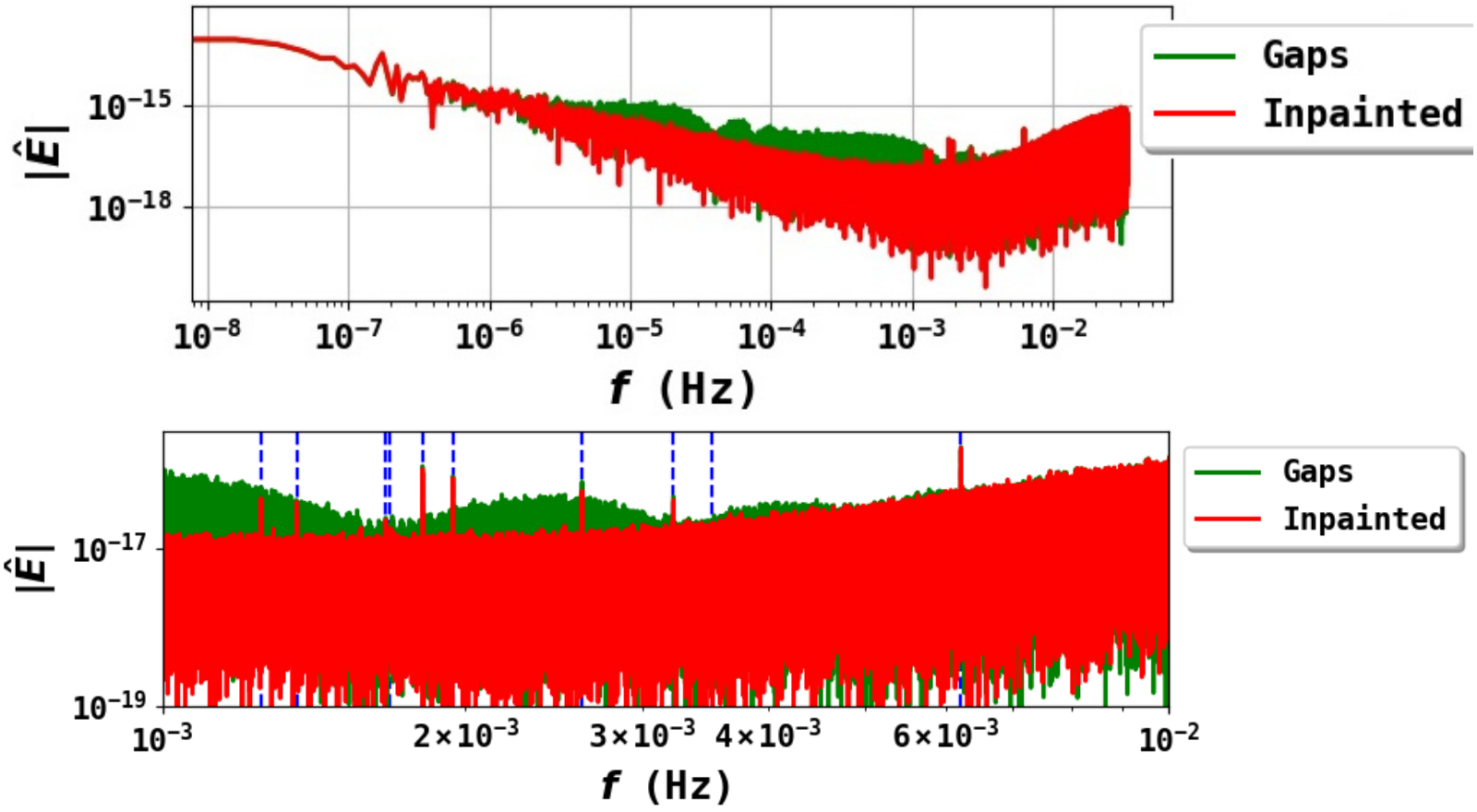


- **LDC 1-3:** 10 Verification galactic binaries

[4] https://lisa-ldc.lal.in2p3.fr/

Sparse Data Inpainting Method

Impact on input signal spectrum - LDC1-3 [4]



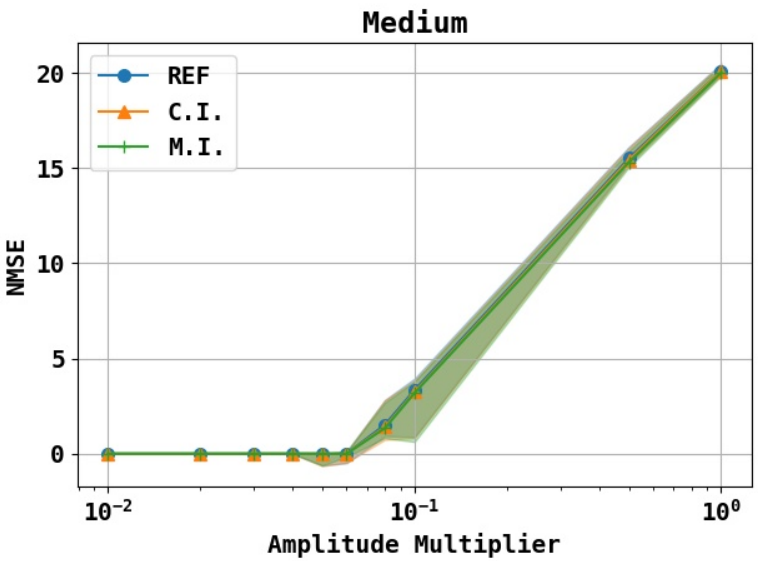
⇒ Impact of the gaps on the data is strongly mitigated

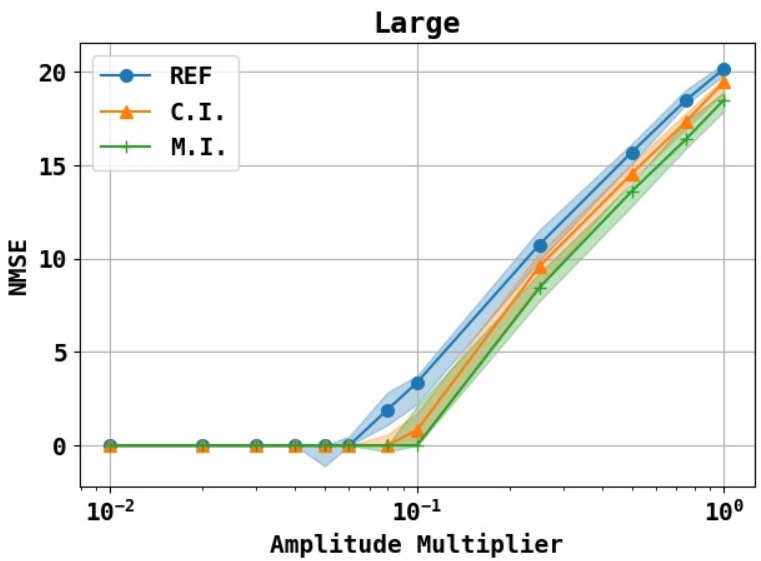
⇒ Contamination of the signal by noise is limited

[4] https://lisa-ldc.lal.in2p3.fr/

Normalized Means Square Error

Quality of recovered signal



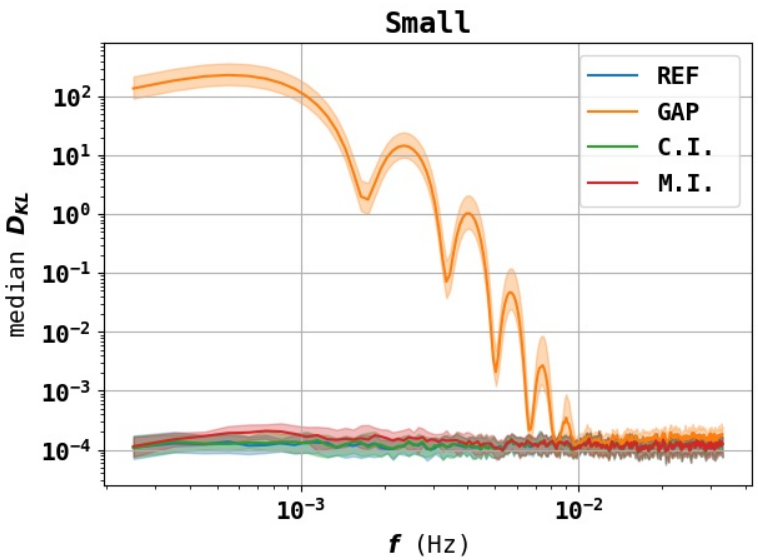


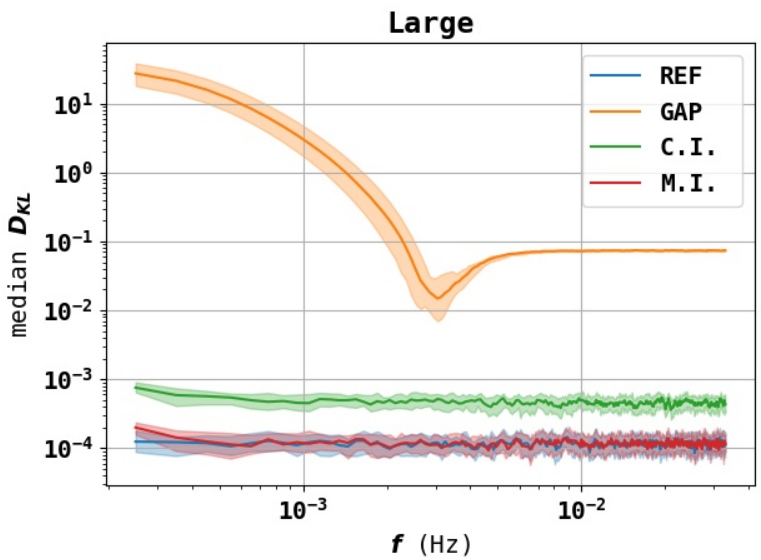
Single GB, quality of extracted signal with input amplitude

	S/M	L
C.I.	Same as ungapped	Quality loss
M.I.	Same as ungapped	Quality loss

Kullback-Leibler divergence

Quality of recovered noise distribution





The Kullback-Leibler Divergence measures a discrepancy between the expected and the recovered noise distributions.

	S/M	L
C.I.	Good	Cannot make up for power loss
M.I.	Good	Good

Conclusion

Conclusion

- **Efficient reconstruction** of signal and noise
- Works as well with a **large unknown number of separated sources**.
  - Worked on LDC1-3 with 28% of missing data [5]
- **Flexible framework** which could be adapted to:
  - Other types of sources.
  - Sources separation.
  - Noise distribution estimation.
- Fast: ~ 1 hour for 2 years of data on a laptop

Future work: improving waveform representation (machine learning, non-parametric)

[5] *Sparse data inpainting for the recovery of Galactic-binary gravitational wave signals from gapped data*, A. Blelly, J. Bobin, M. Moutarde [IN PROGRESS]