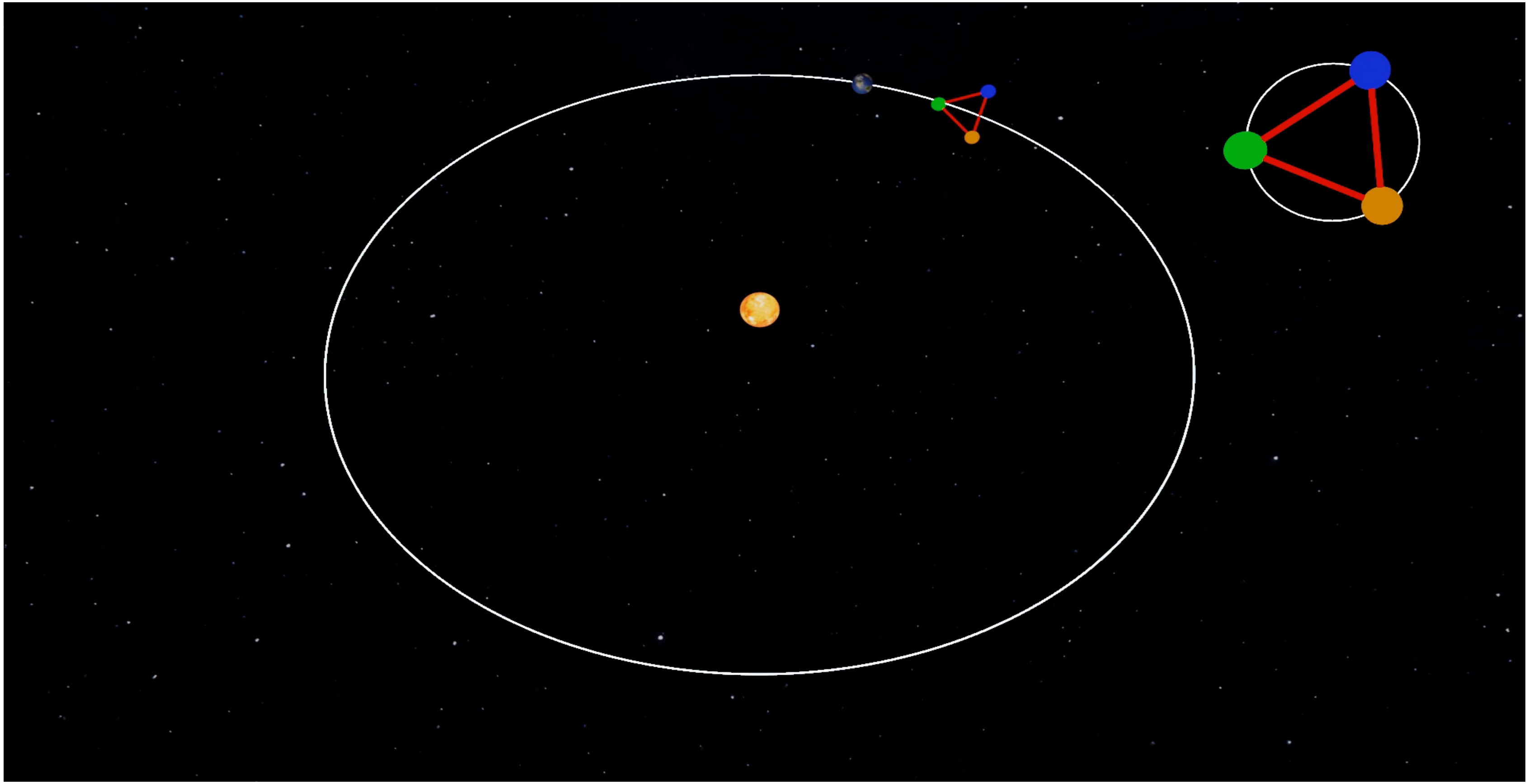
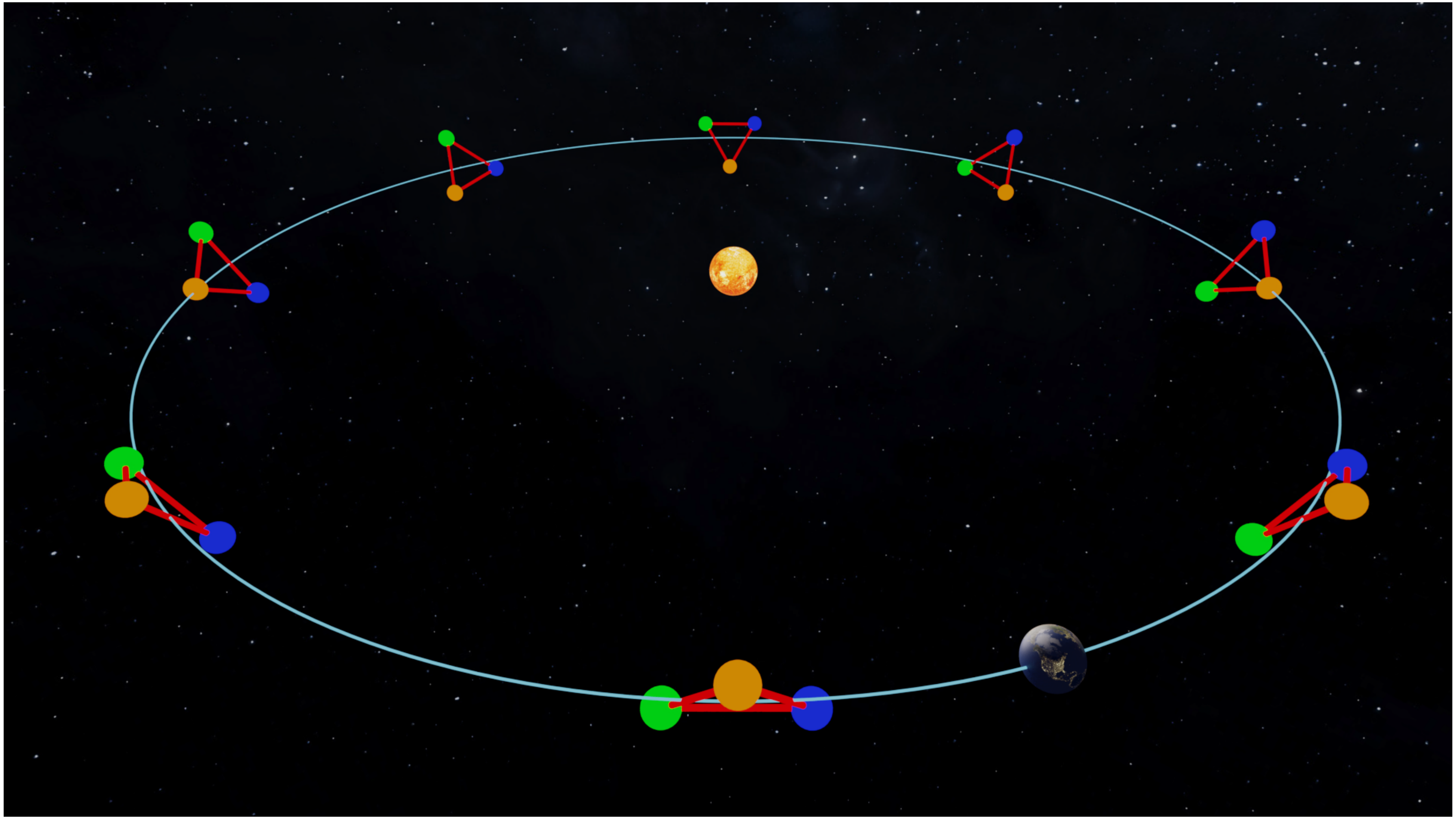
Accelerating parameter estimation of Galactic binaries in the full LISA frequency band using Gaussian Process Regression



https://zenodo.org/record/6761175

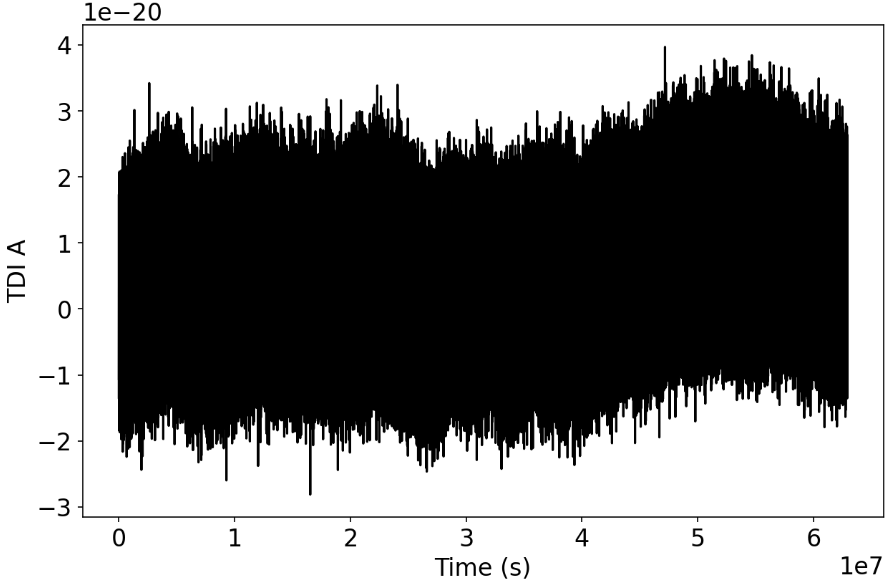
Stefan Strub, Luigi Ferraioli, Cedric Schmelzbach, Simon Stähler, Domenico Giardini

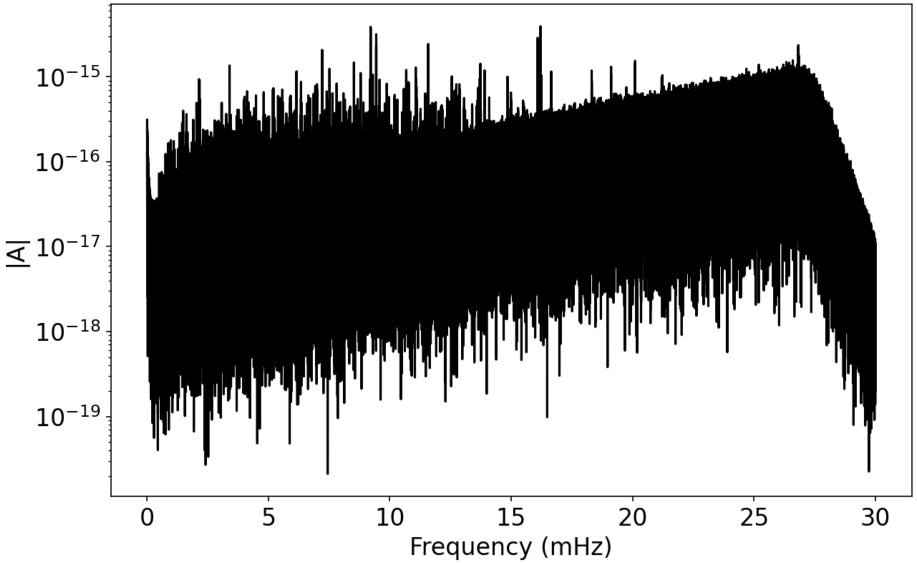
Accelerating parameter estimation of Galactic binaries in the full LISA frequency band using Gaussian Process Regression



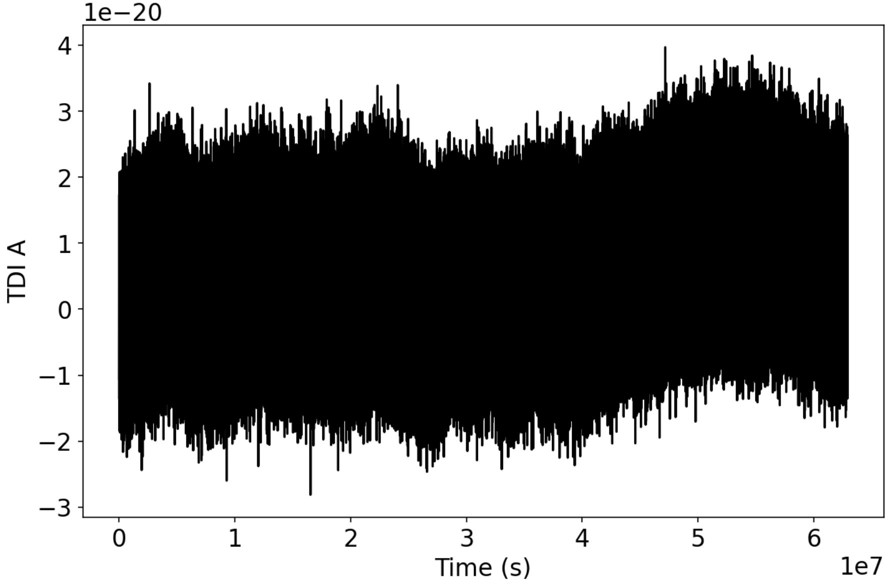
Stefan Strub, Luigi Ferraioli, Cedric Schmelzbach, Simon Stähler, Domenico Giardini

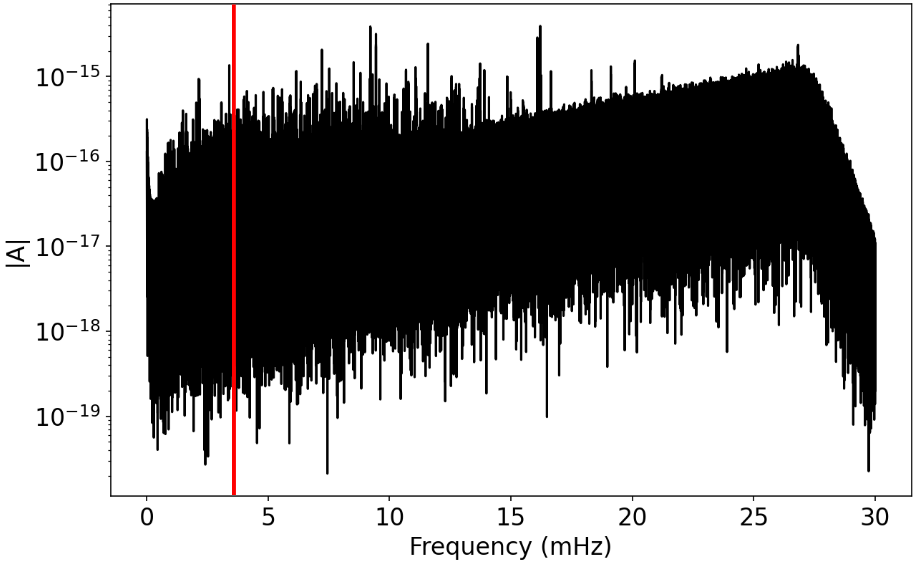
LDC1-4 (Radler)



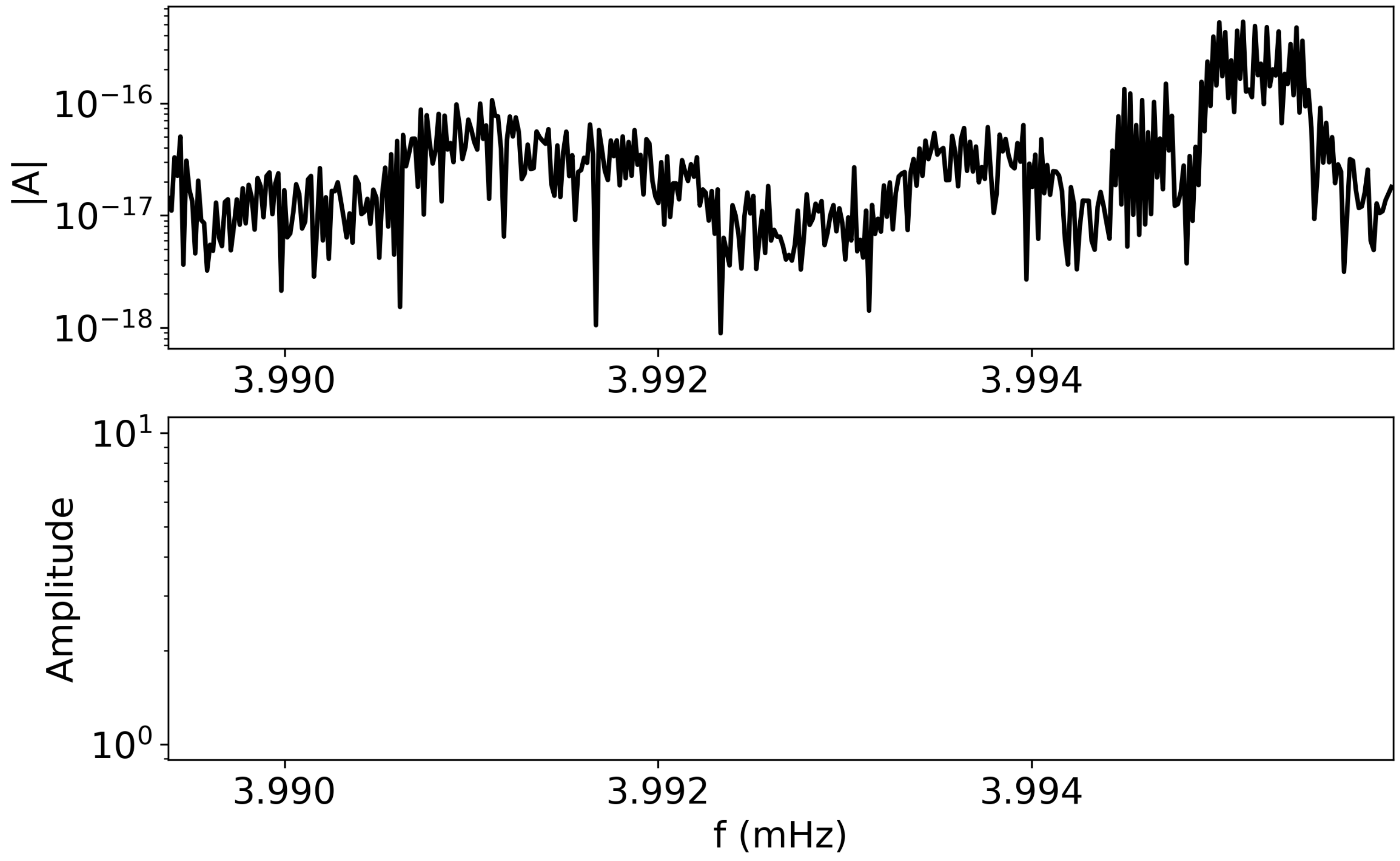


LDC1-4 (Radler)

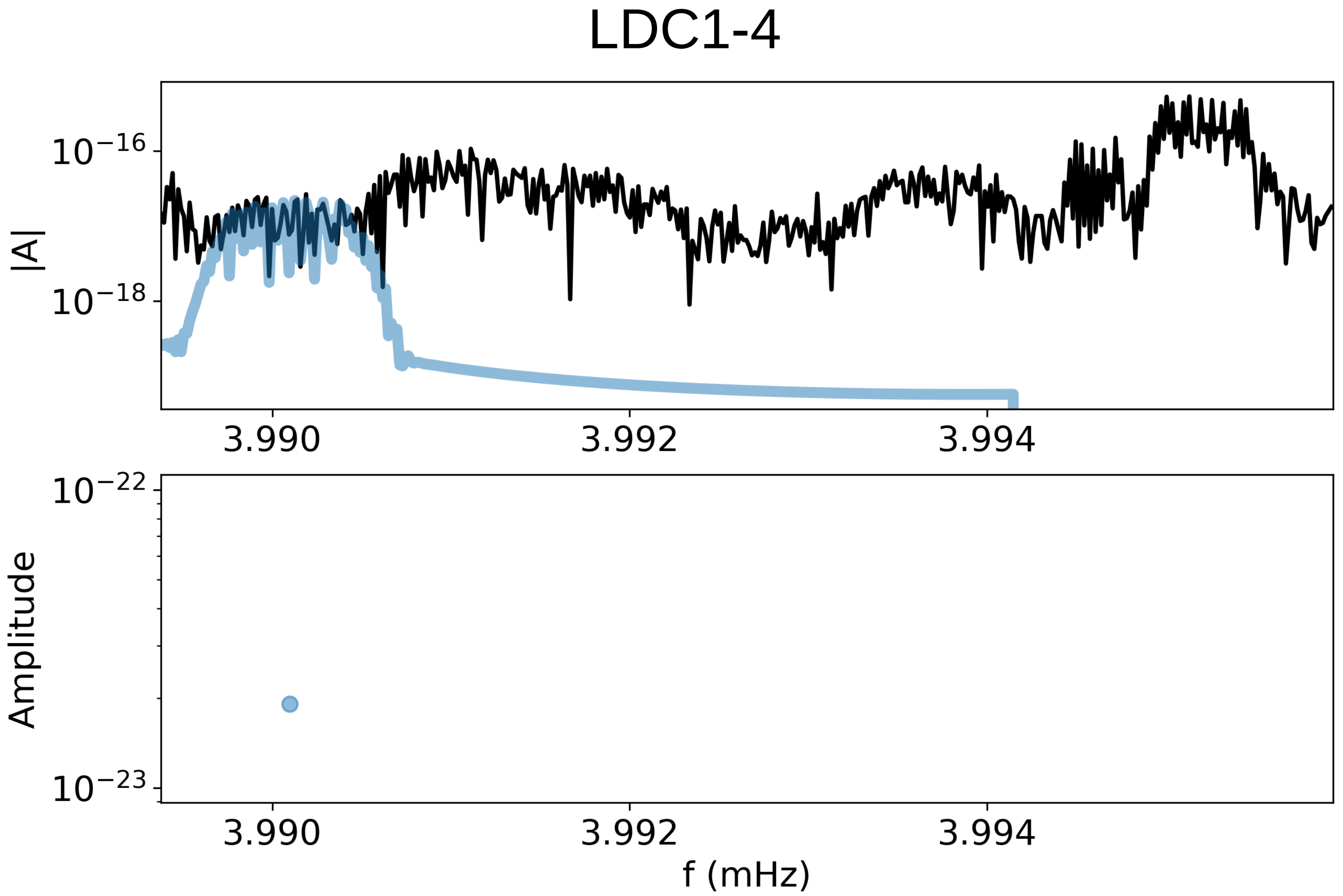


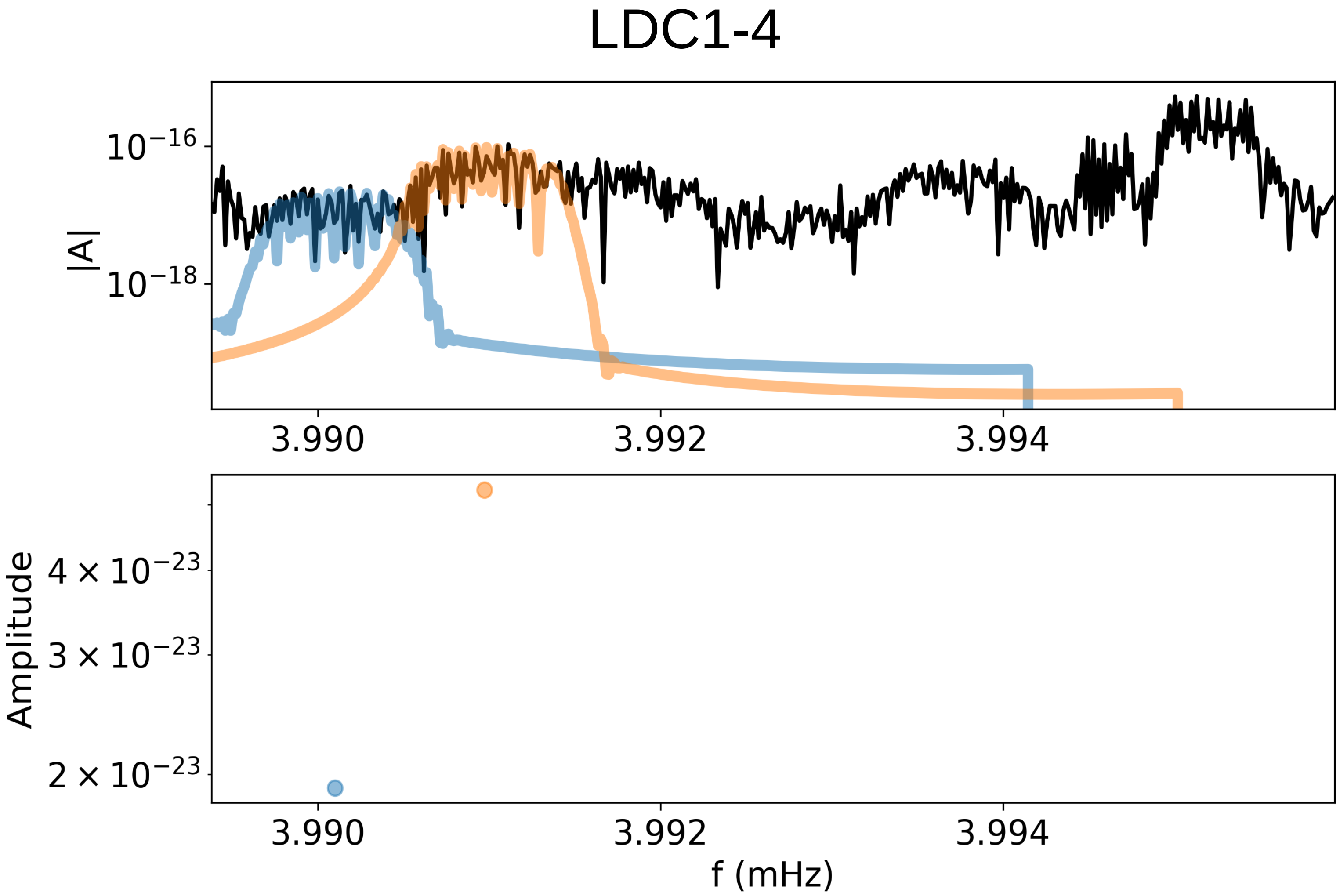


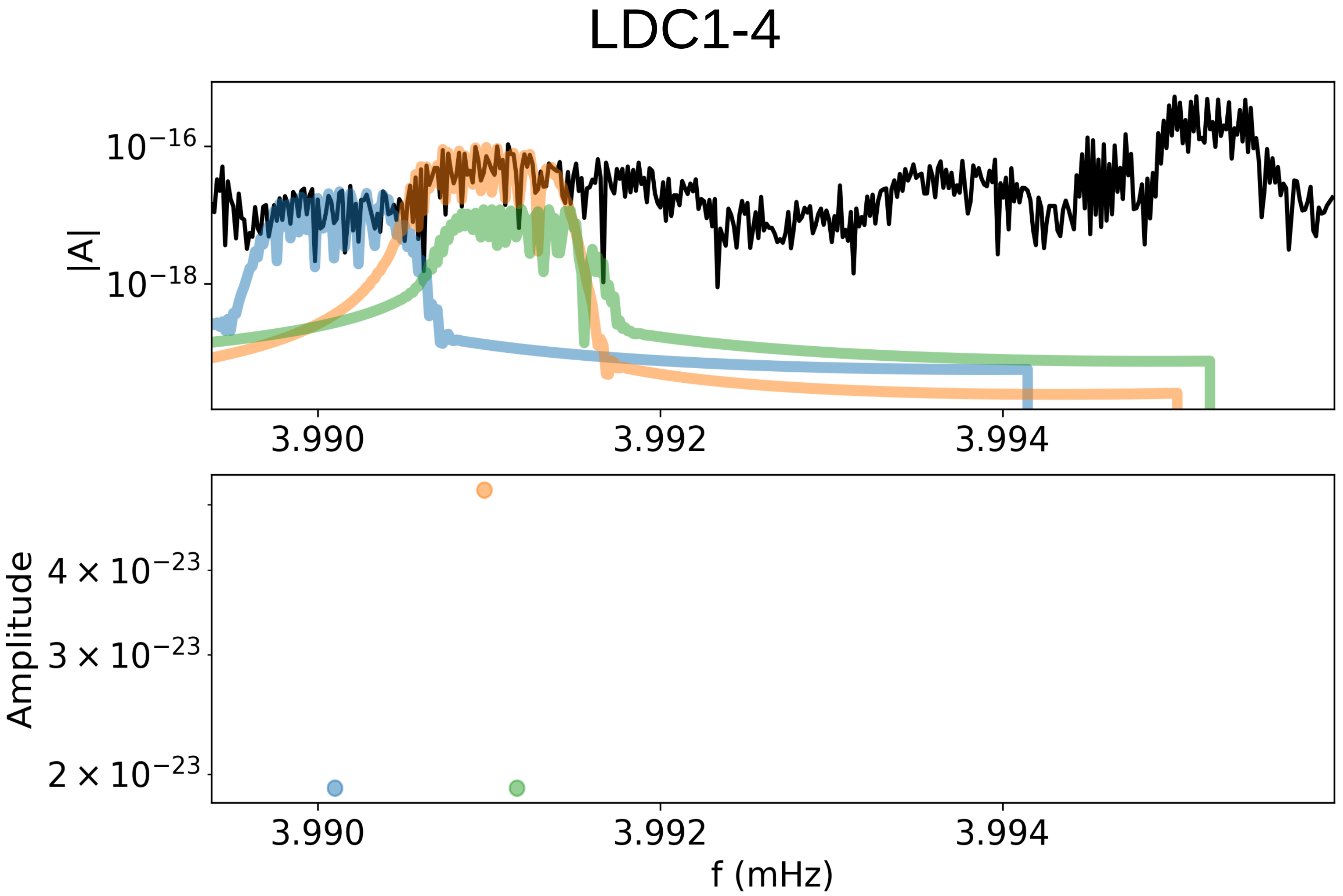
LDC1-4



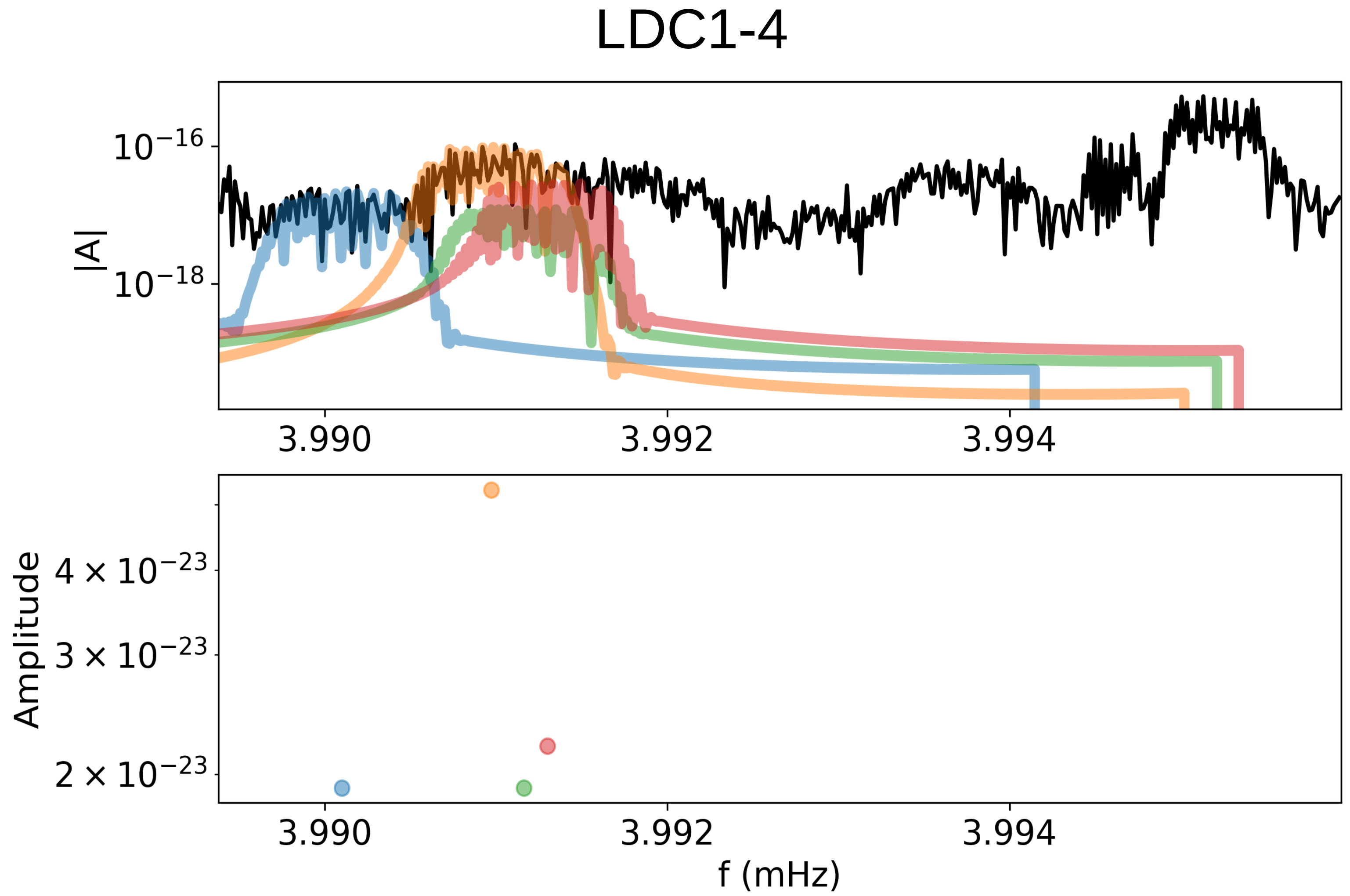
|A|
10^{-16}
10^{-17}
10^{-18}
3.990
3.992
3.994
Amplitude
10^{1}
10^{0}
3.990
3.992
3.994
f (mHz)


LDC1-4
10^{-16}
10^{-18}
|A|
3.990
3.992
3.994
10^{-22}
10^{-23}
Amplitude
3.990
3.992
3.994
f (mHz)

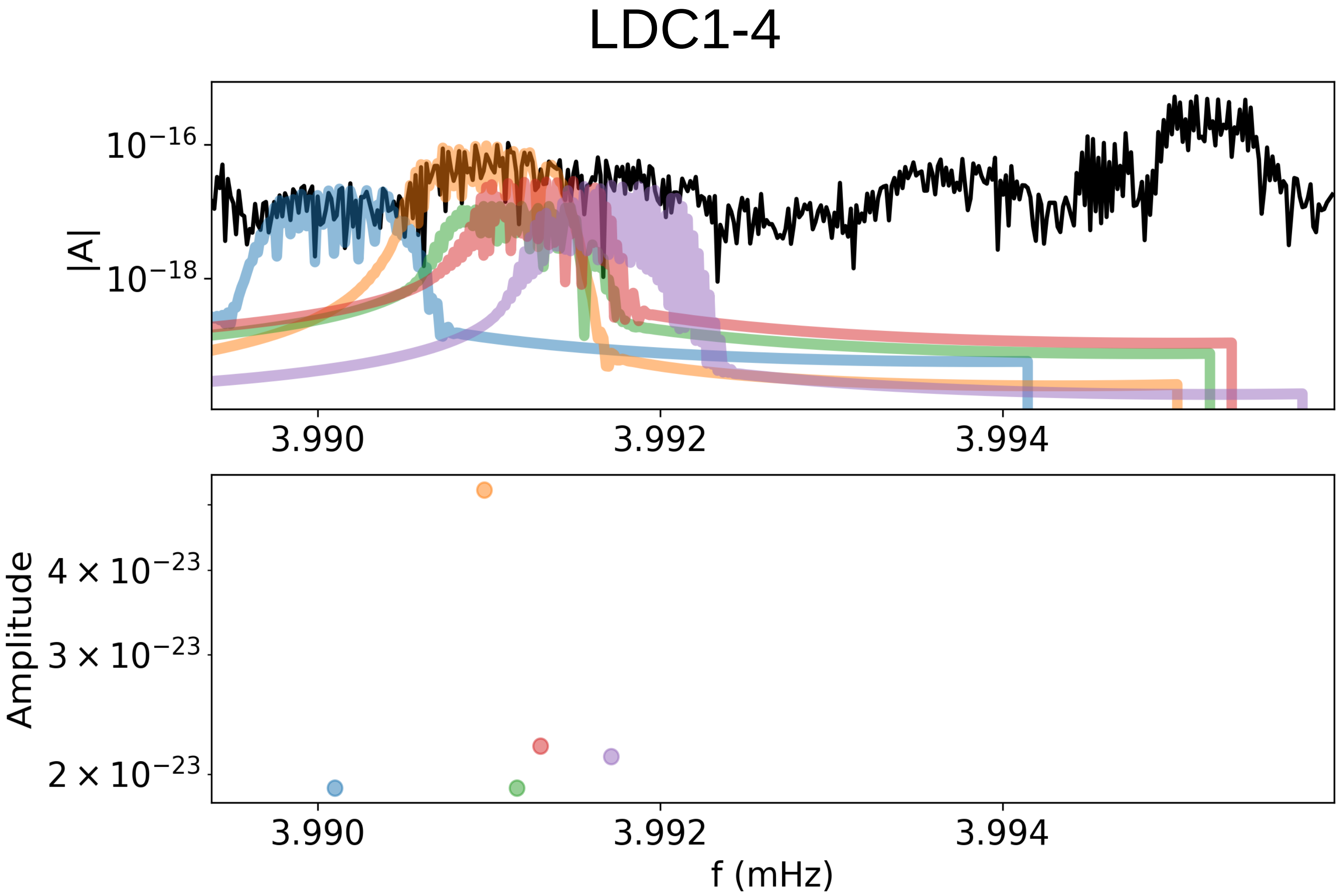

LDC1-4
10^{-16}
10^{-18}
|A|
3.990
3.992
3.994
Amplitude
4×10^{-23}
3×10^{-23}
2×10^{-23}
3.990
3.992
3.994
f (mHz)

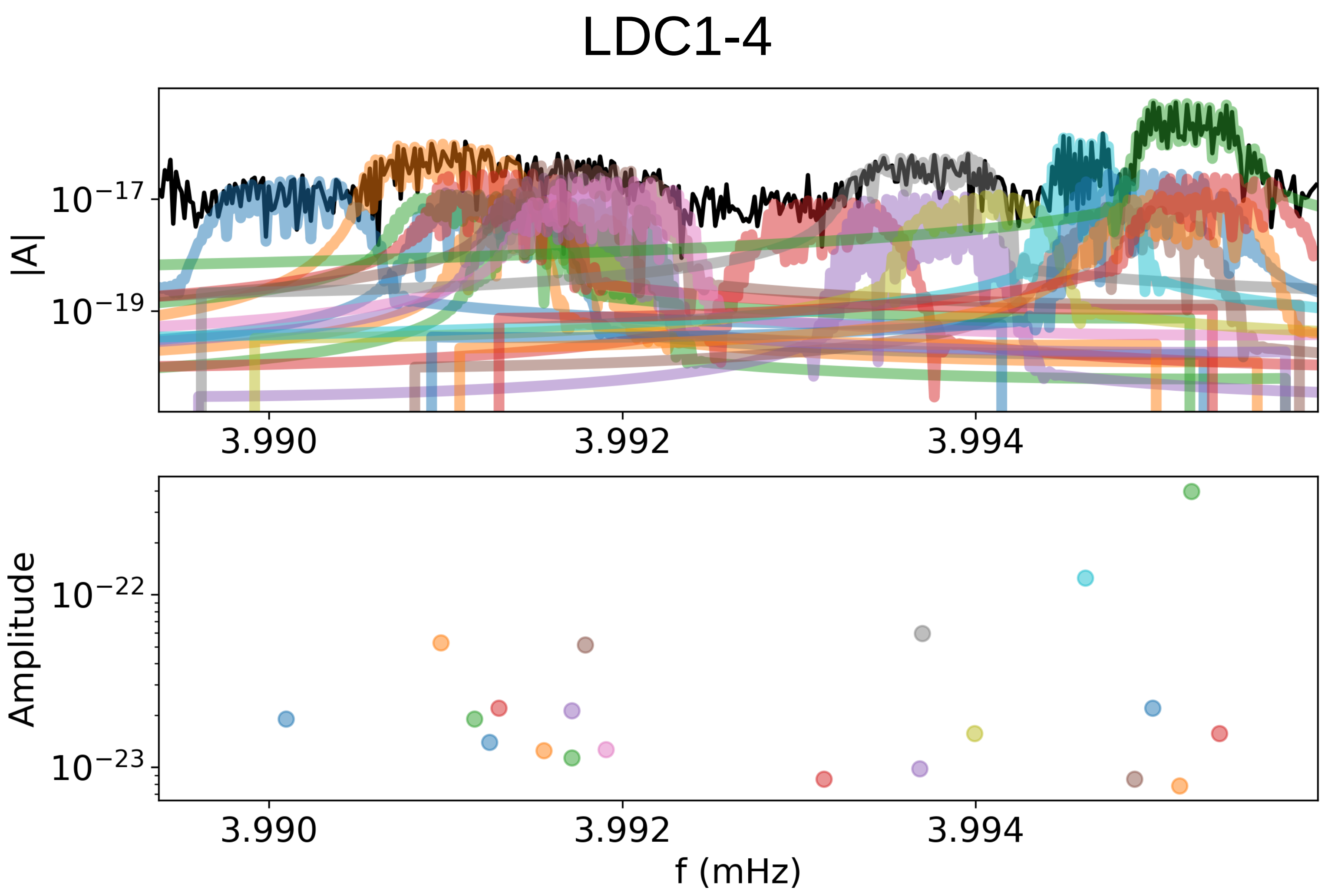


LDC1-4
|A|
10^{-16}
10^{-18}
3.990
3.992
3.994
Amplitude
4×10^{-23}
3×10^{-23}
2×10^{-23}
3.990
3.992
3.994
f (mHz)



LDC1-4
10^{-16}
10^{-18}
|A|
3.990
3.992
3.994
Amplitude
4×10^{-23}
3×10^{-23}
2×10^{-23}
3.990
3.992
3.994
f (mHz)


LDC1-4
|A|
10^{-16}
10^{-18}
3.990
3.992
3.994
Amplitude
4×10^{-23}
3×10^{-23}
2×10^{-23}
3.990
3.992
3.994
f (mHz)


LDC1-4
|A|
10^{-17}
10^{-19}
3.990
3.992
3.994
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)

Fast best fit and slow posterior distribution: The best of both worlds

	Best fit	Bayesian posterior distribution
Possible method	Genetic Search Algorithm	Markov Chain Monte Carlo
Number of signal simulations	20’000 Fast	100’000 – 1’000’000 Slow
Uncertainty estimate	No	Yes

Log-likelihood:

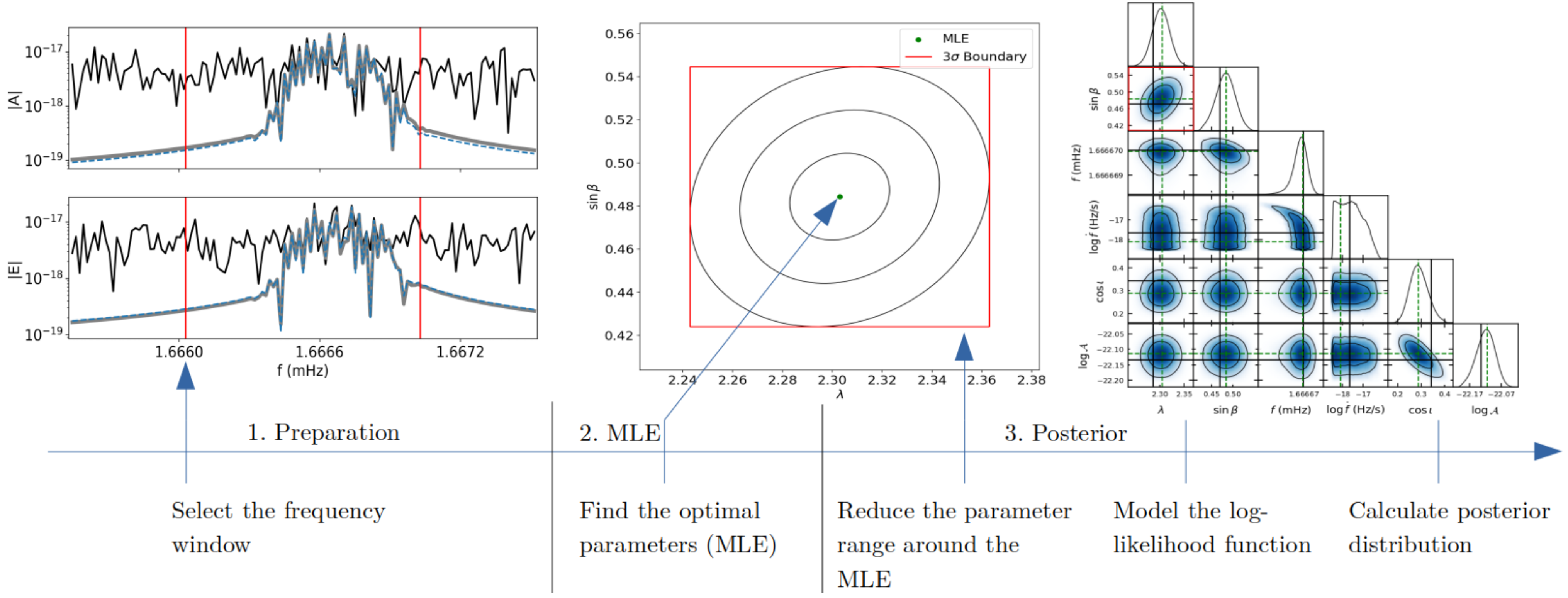
$$\log p(d \mid \theta) = -\frac{1}{2}\langle d - s(\theta) \,| d - s(\theta)\rangle$$

θ : parameter set

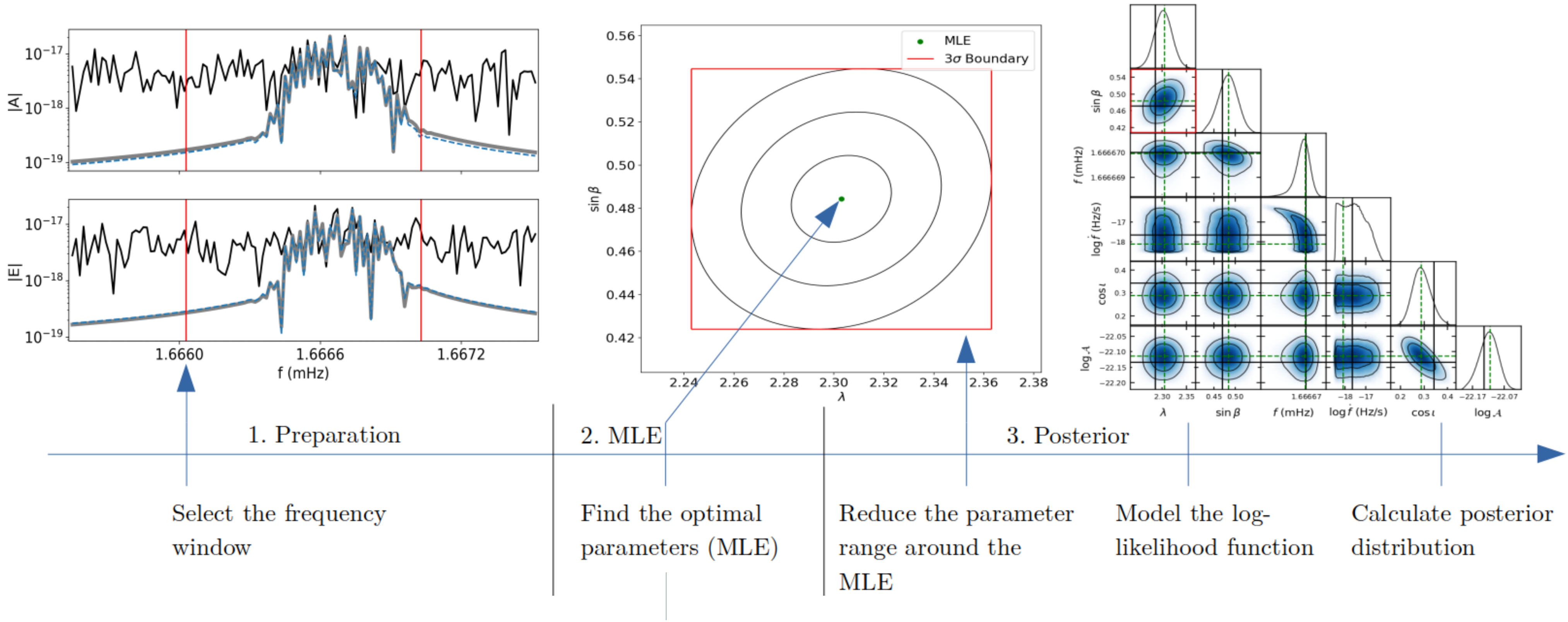
Scalar product:

$$\langle x(t) \,| y(t)\rangle = 4\mathcal{R}\left(\int_0^\infty \frac{\tilde{x}(f)\,\tilde{y}^*(f)}{S(f)}\,df\right)$$

The pipeline

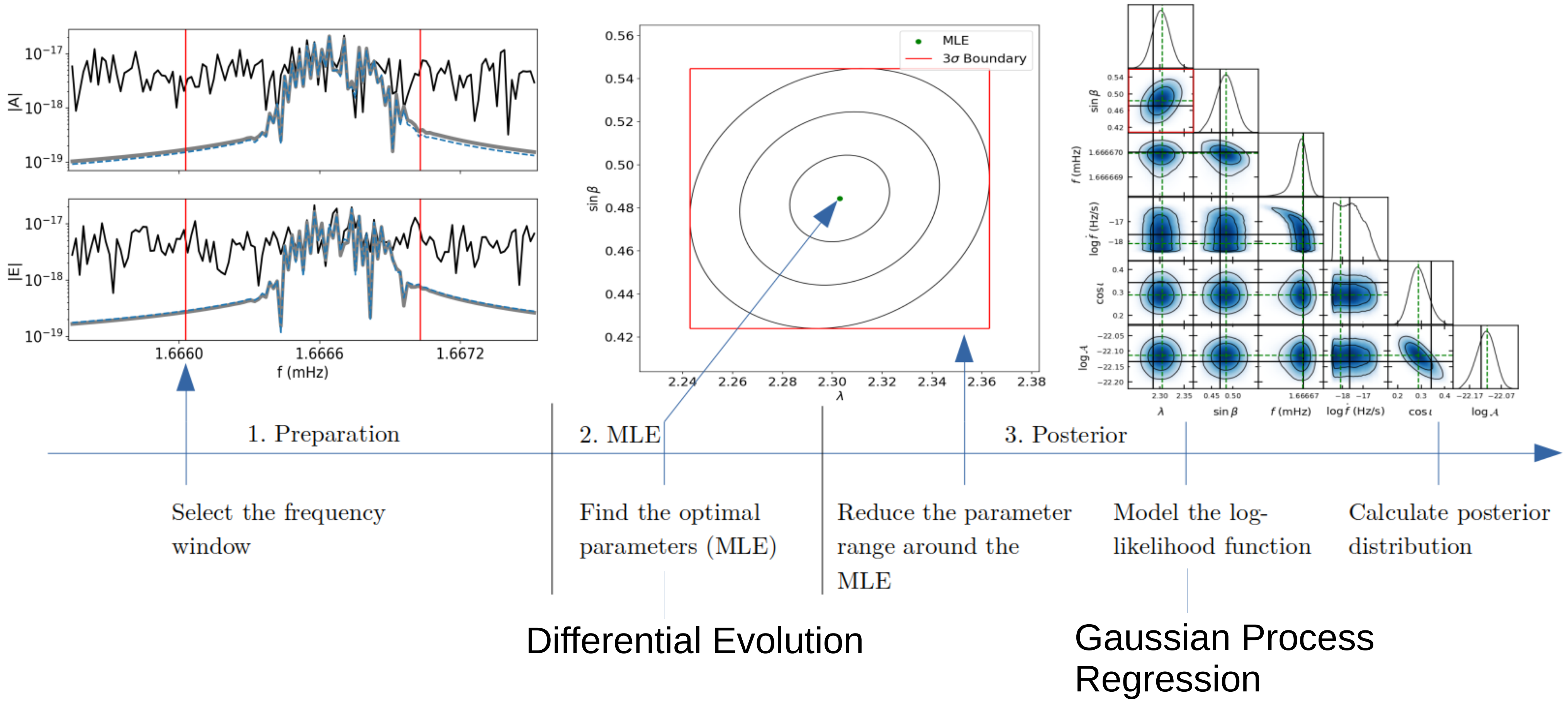


The pipeline



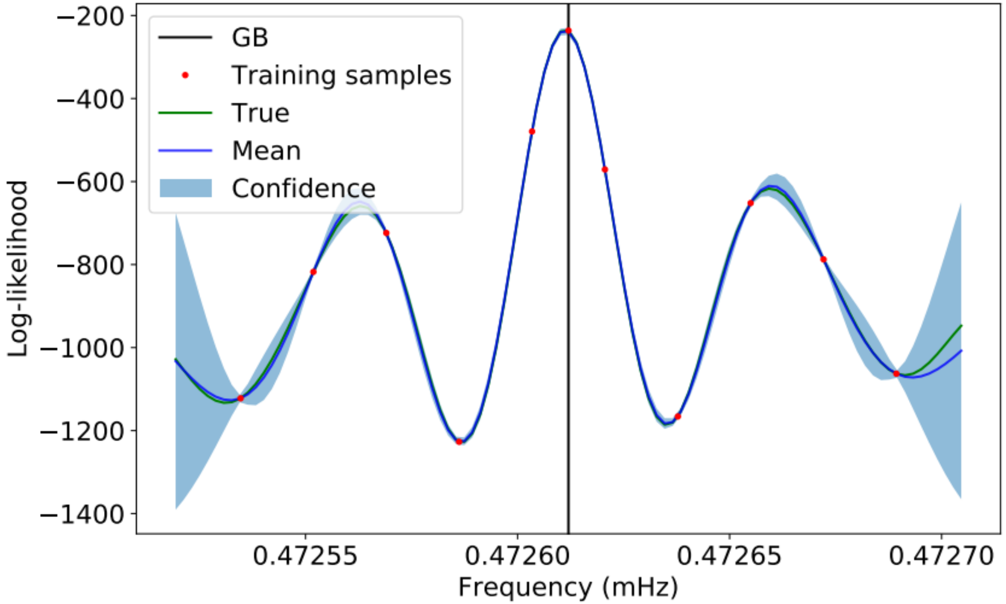
Differential Evolution

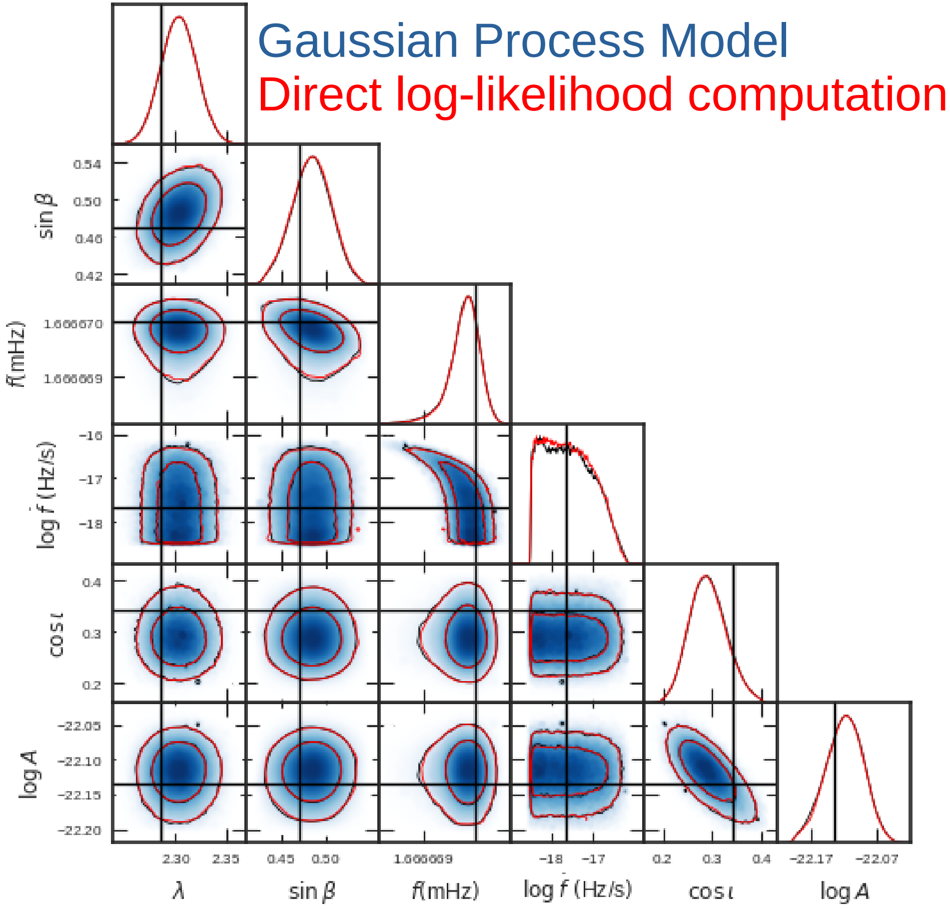
The pipeline



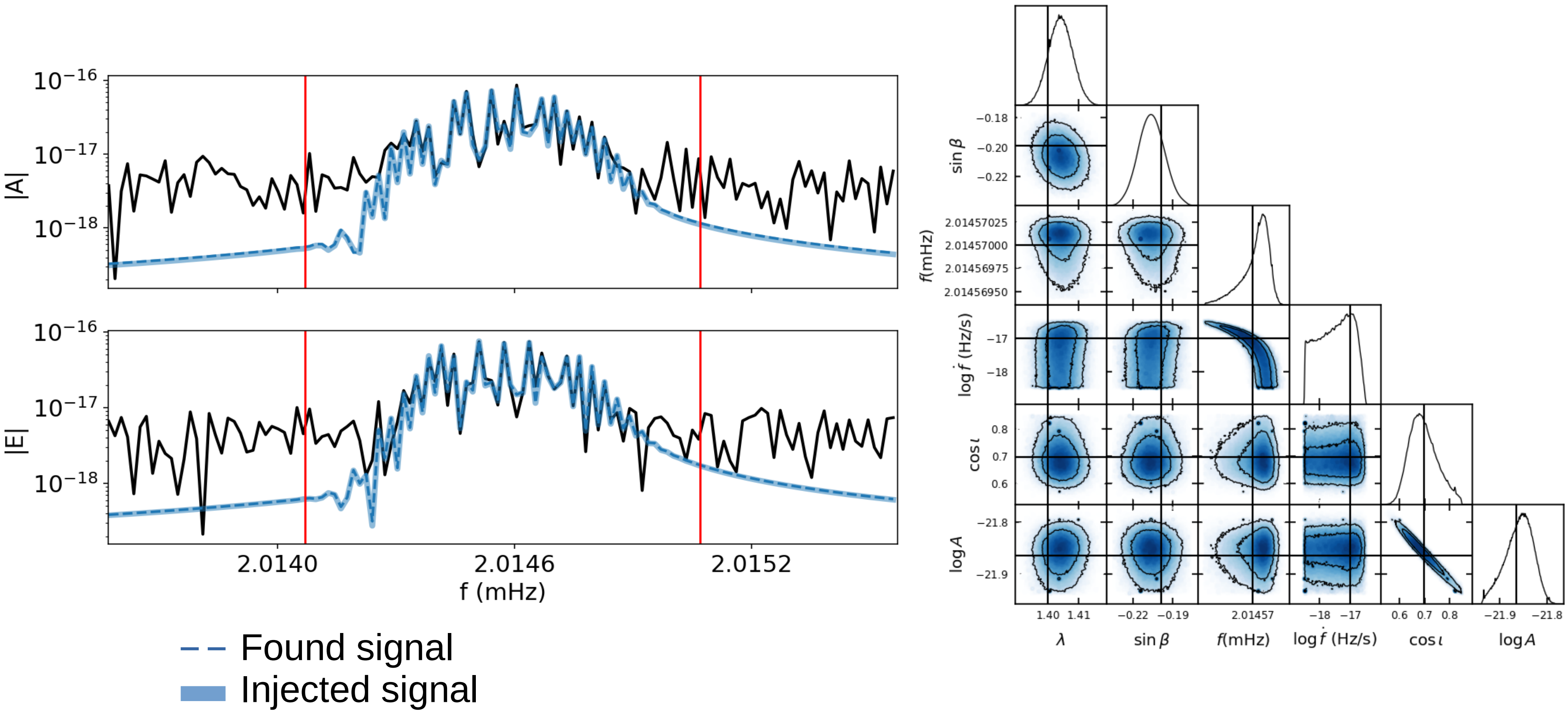
Step 3: Gaussian Process Regression to model the likelihood for rapid MCMC sampling

Redo plot

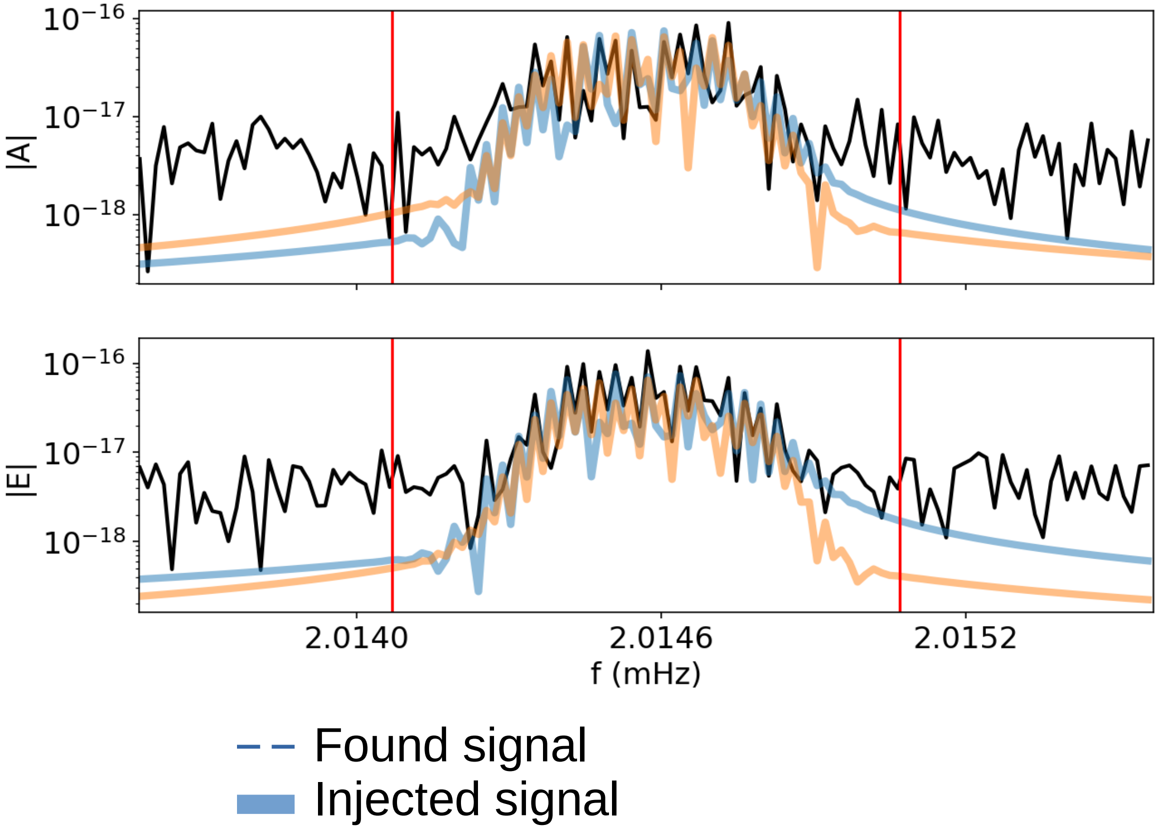




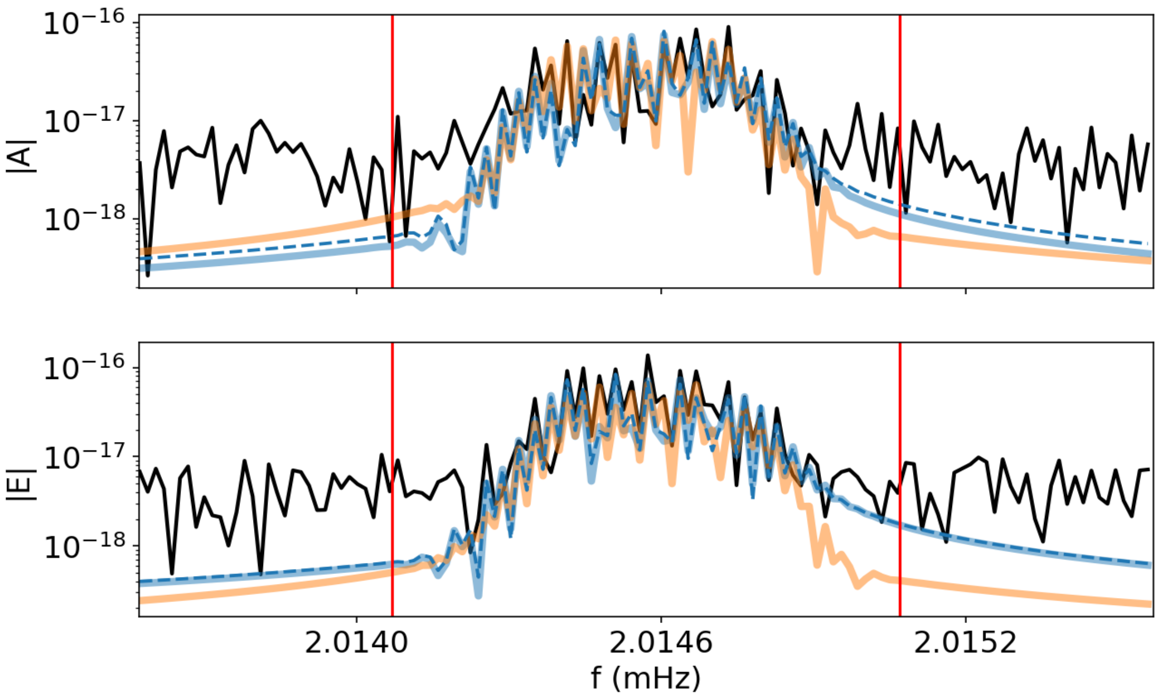
From single to overlapping signals



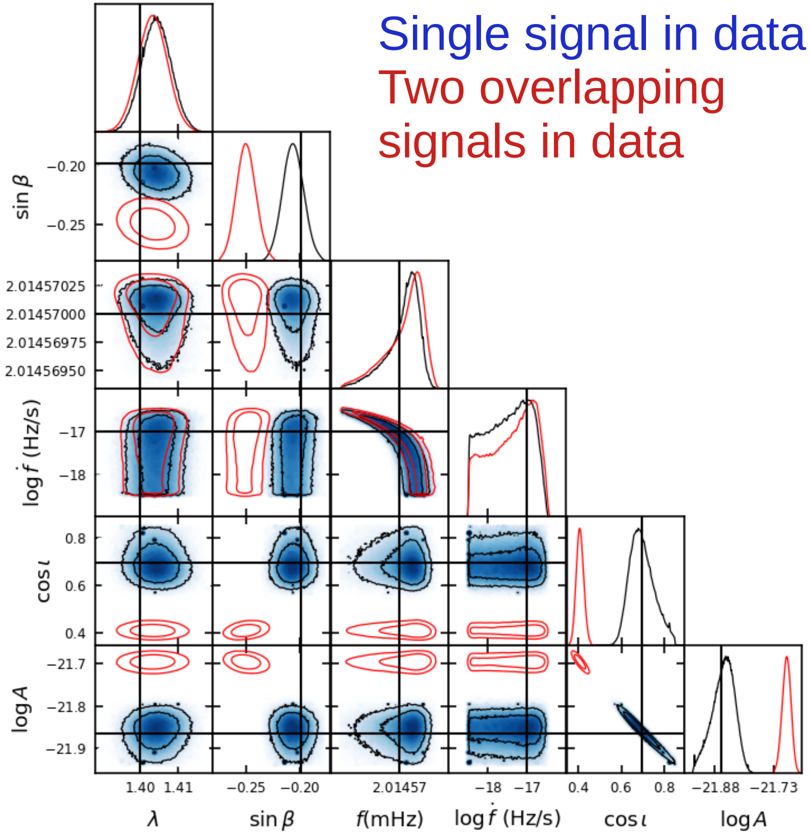
Two overlapping signals which only differ at sky location



Finding only 1 signal has large bias

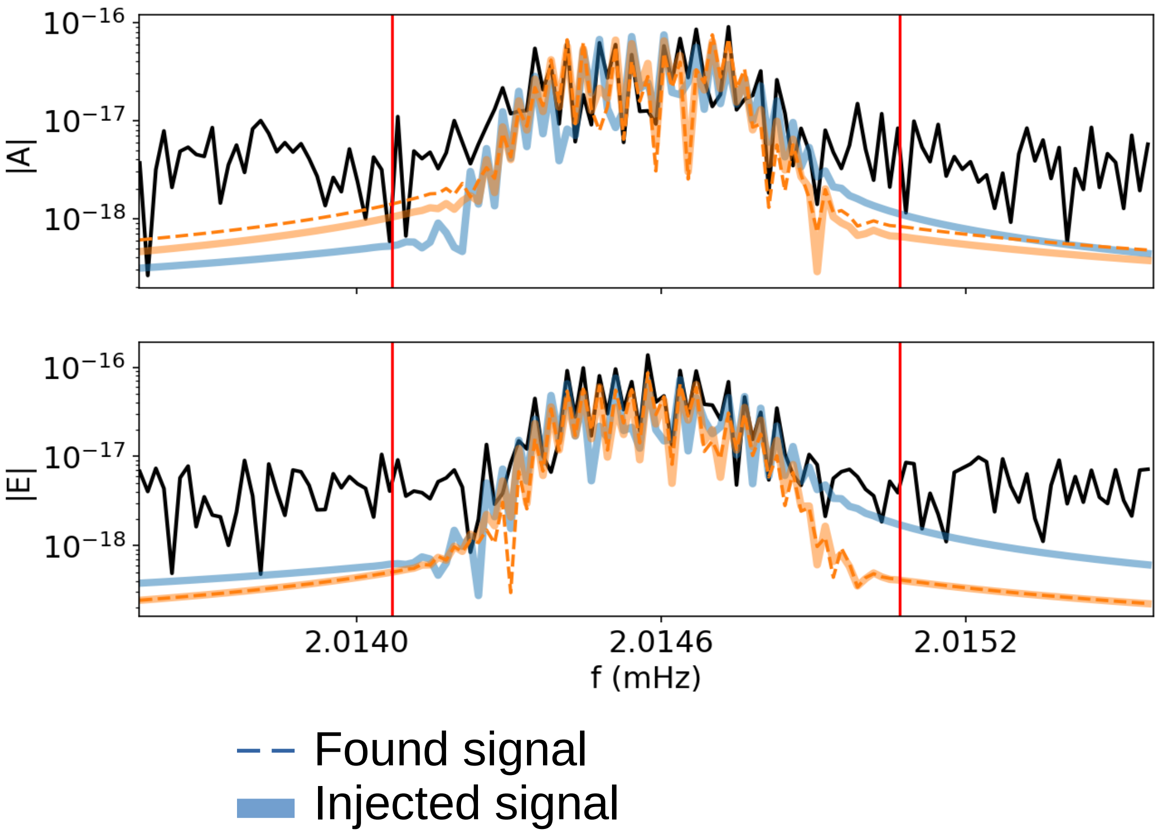


Found signal
Injected signal

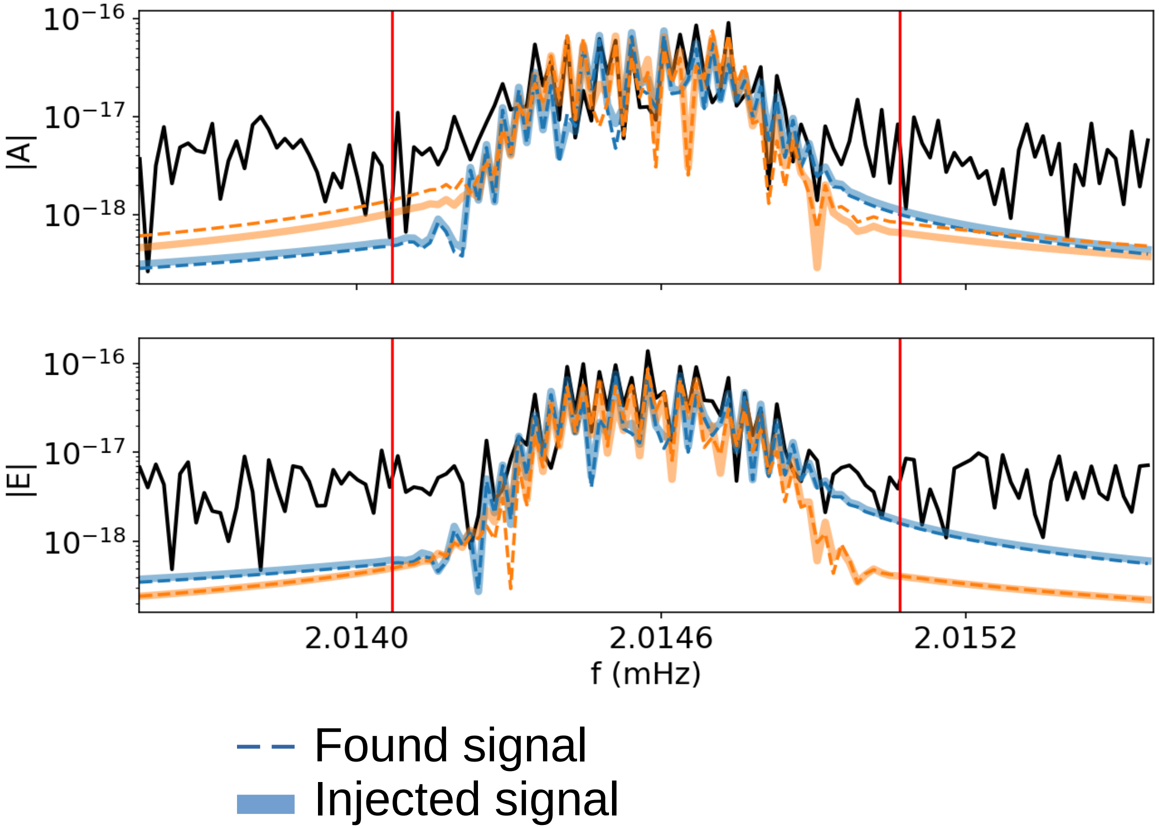


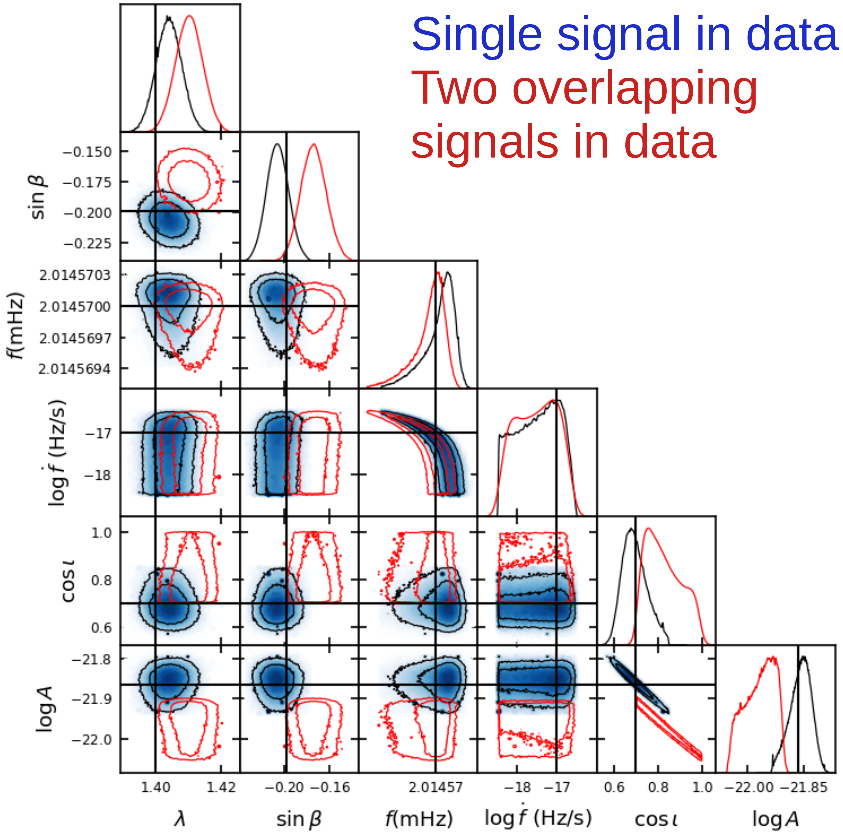
Single signal in data
Two overlapping signals in data

First find the orange signal and subtract



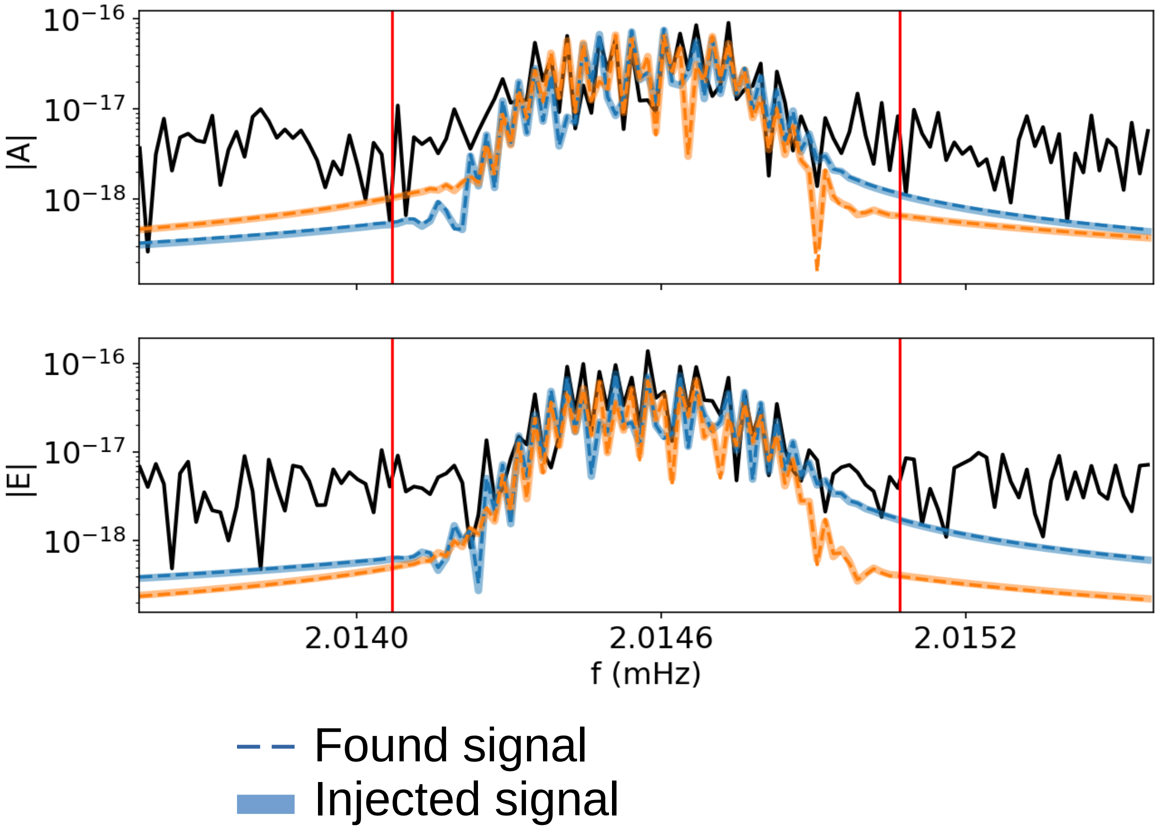
The bias due to the orange signal is reduced but still there

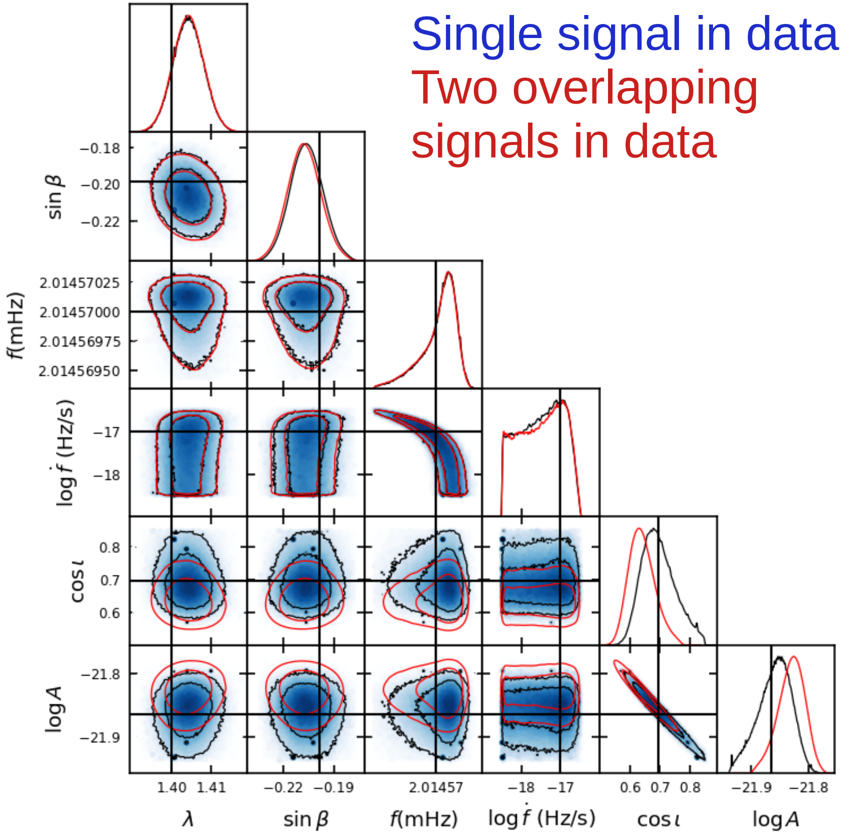




Single signal in data
Two overlapping signals in data

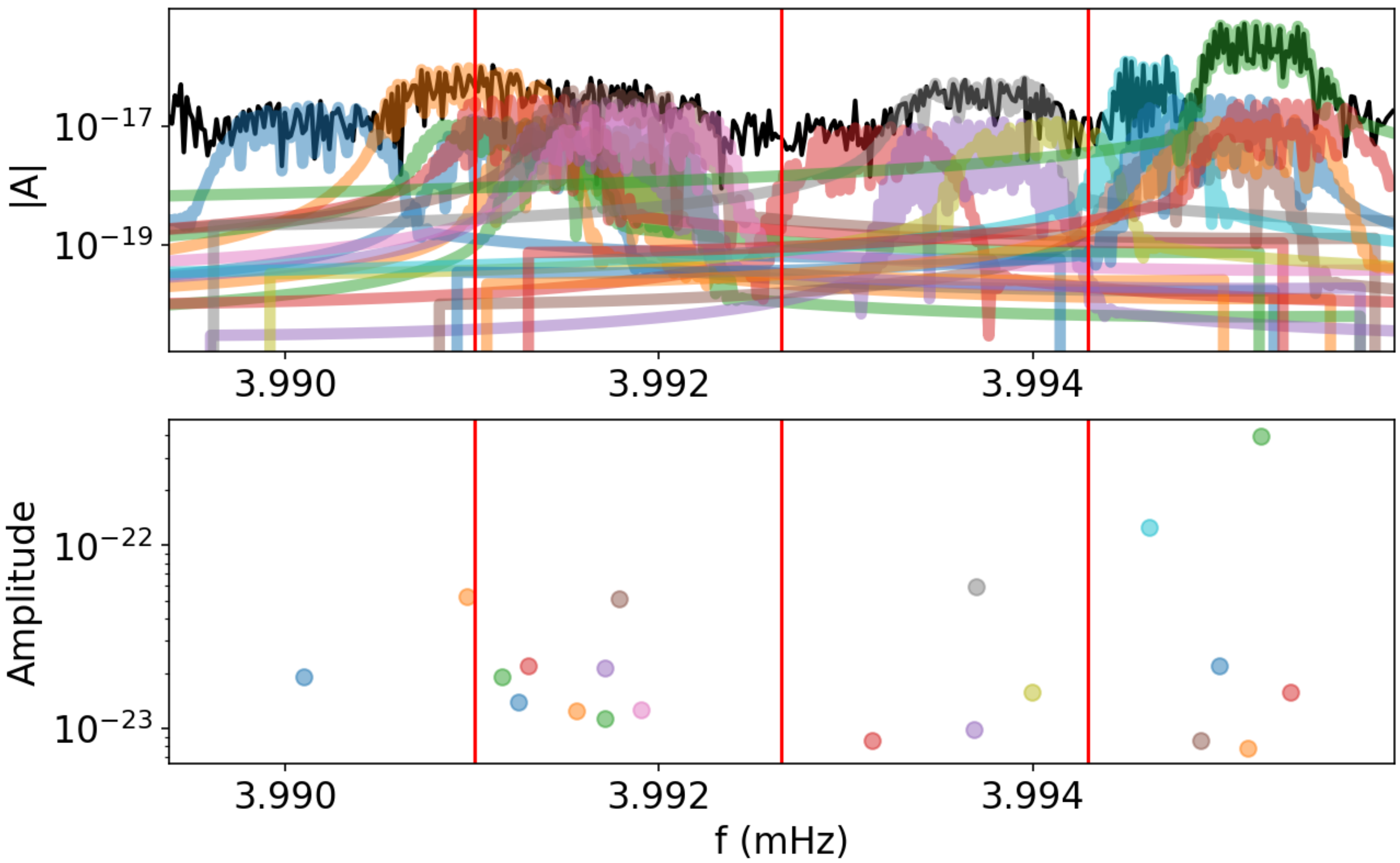
Solution: Perform a global fit among both signals with initial guess from former biased solution



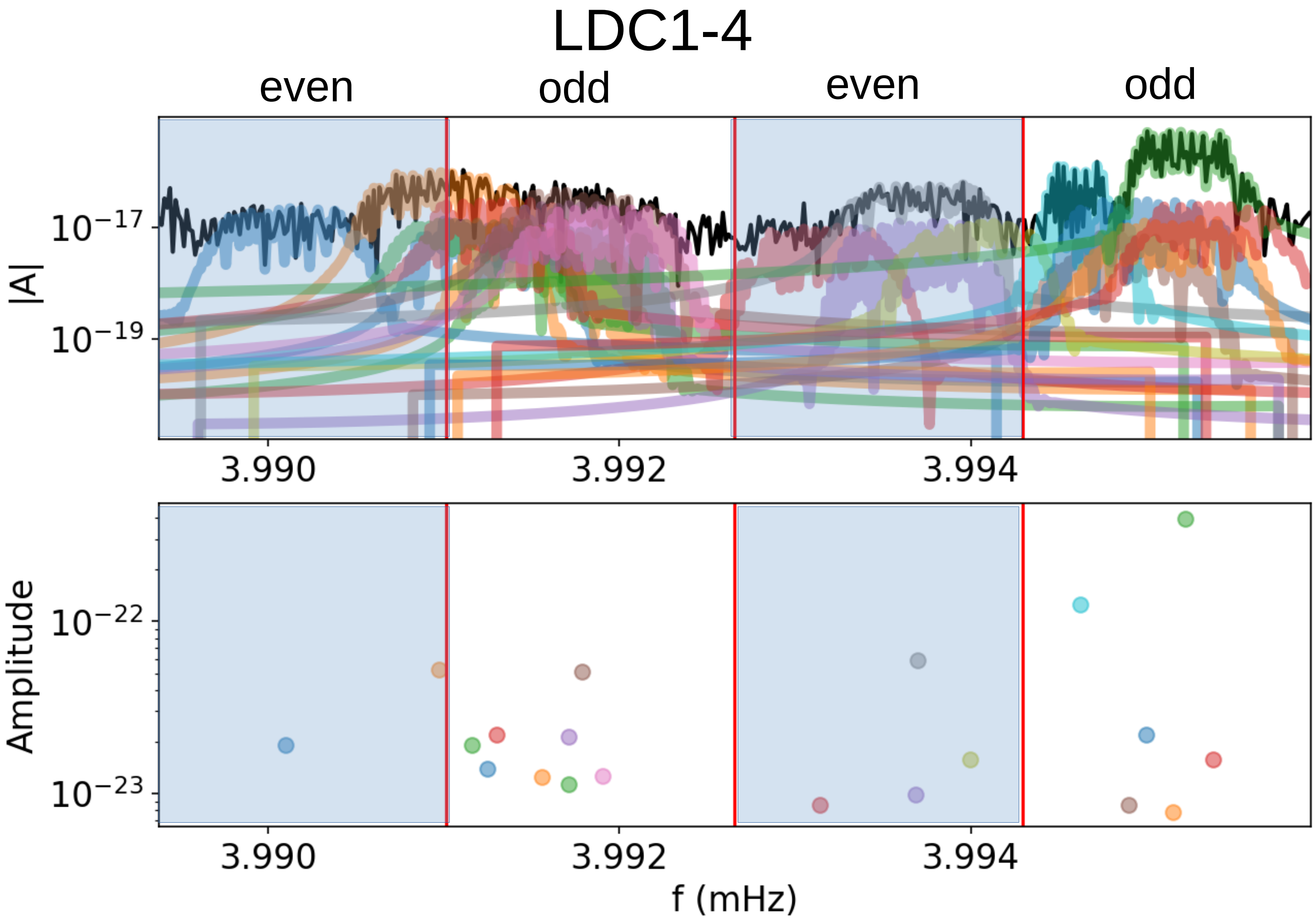


Single signal in data
Two overlapping signals in data

LDC1-4



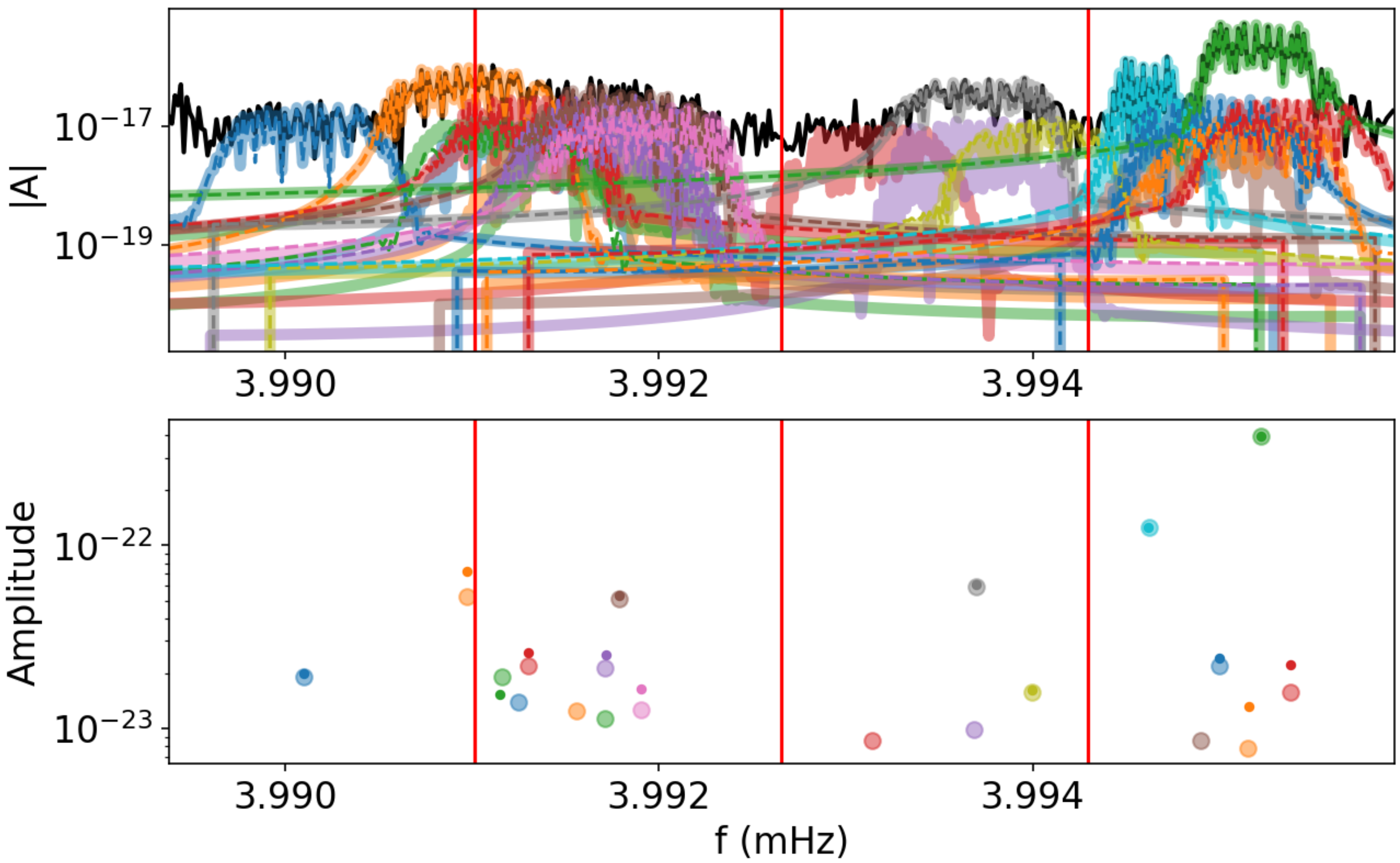
10^{-17}
10^{-19}
|A|
3.990
3.992
3.994
10^{-22}
10^{-23}
Amplitude
3.990
3.992
3.994
f (mHz)



LDC1-4
even
odd
even
odd
|A|
10^{-17}
10^{-19}
3.990
3.992
3.994
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)

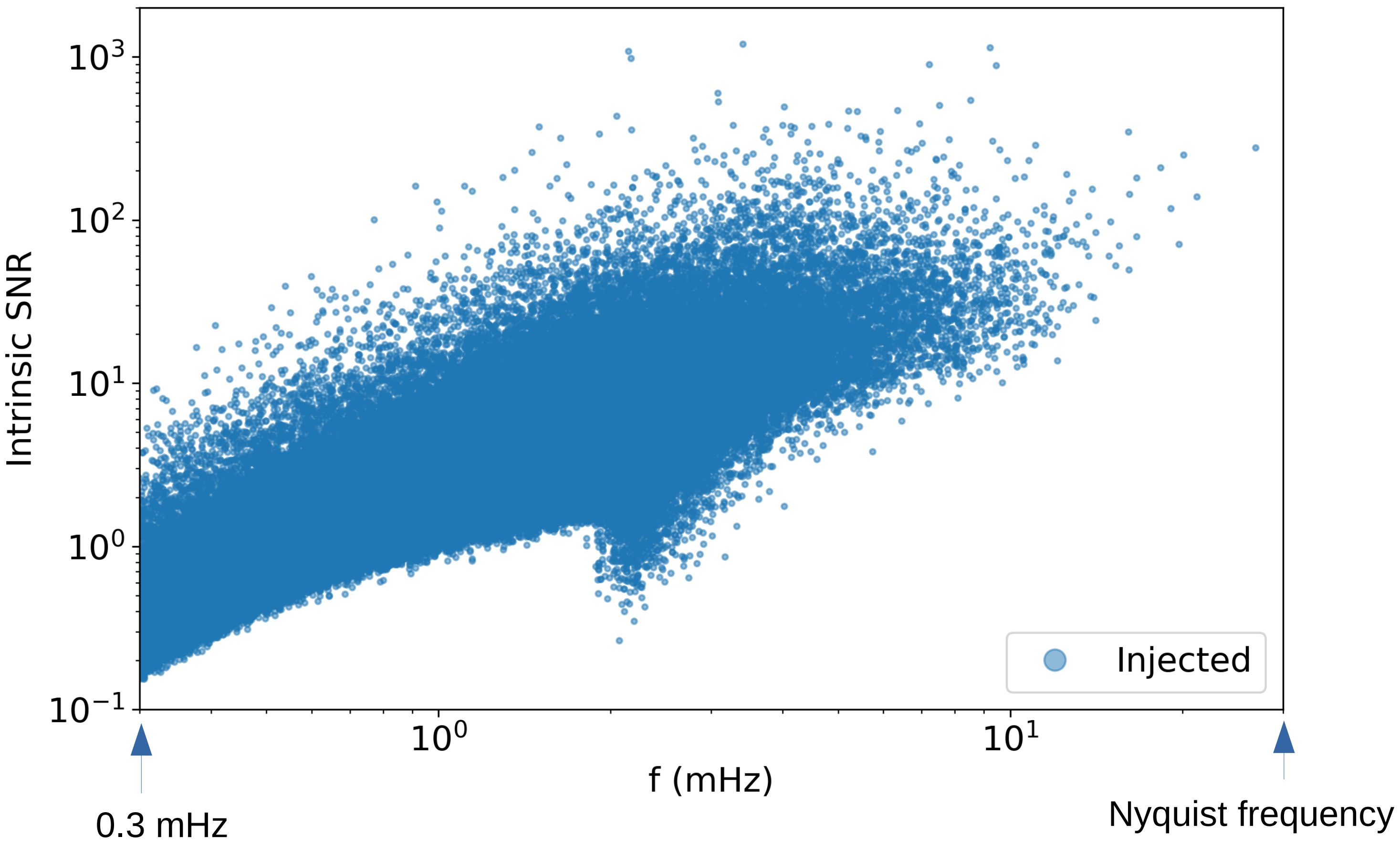
14 out of 20 signals found

LDC1-4



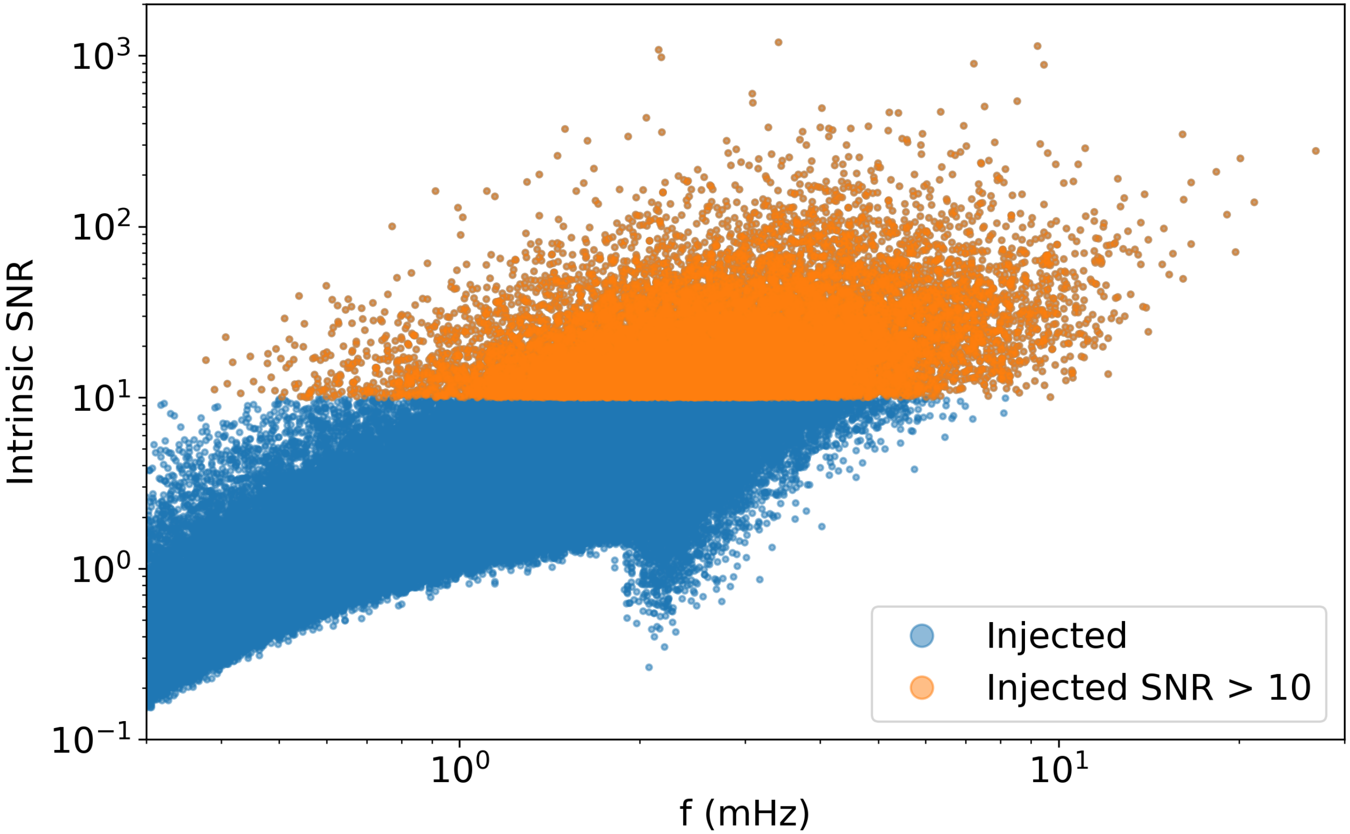
	Number of signals
Injected	6’836’835

LDC 1-4 – 2 years of data



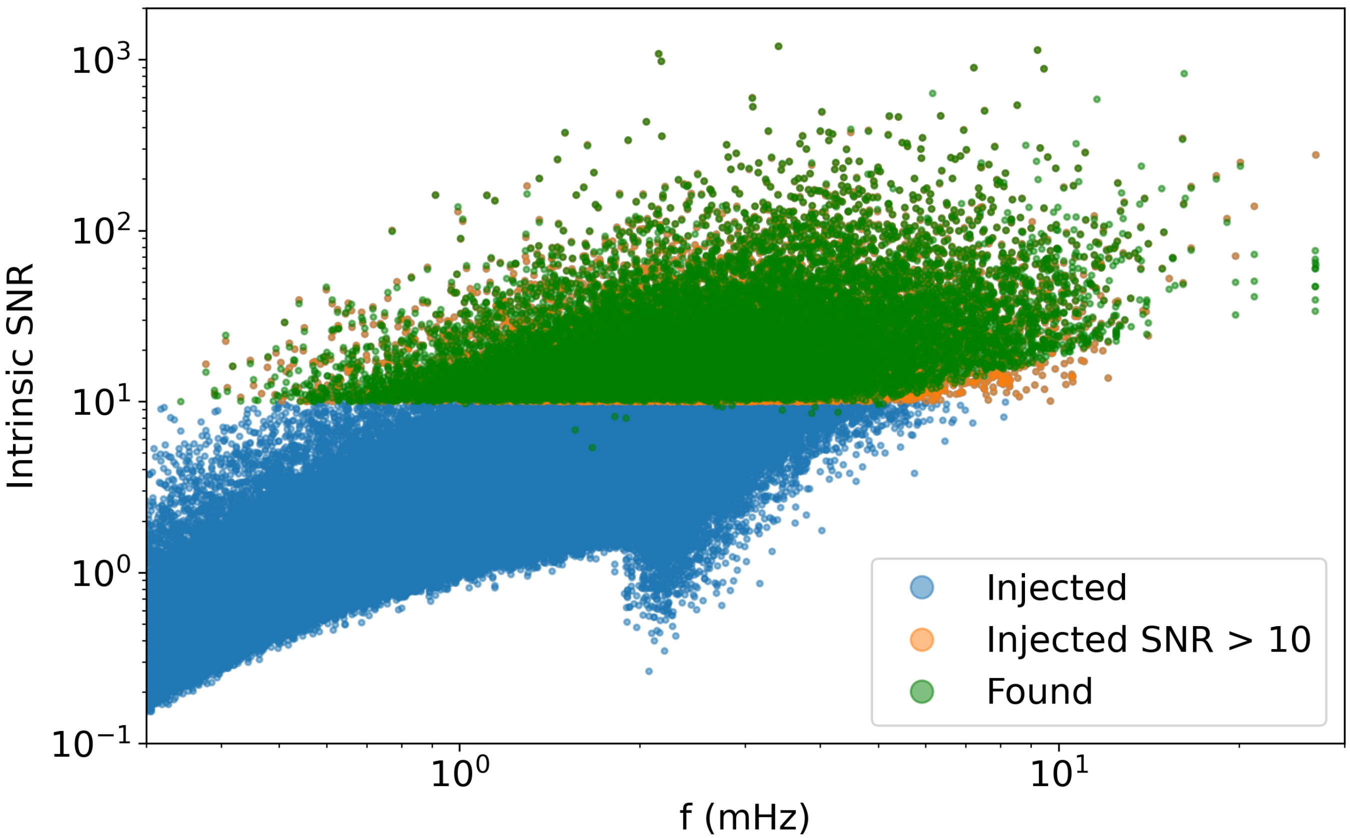
LDC 1-4 – 2 years of data

	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954



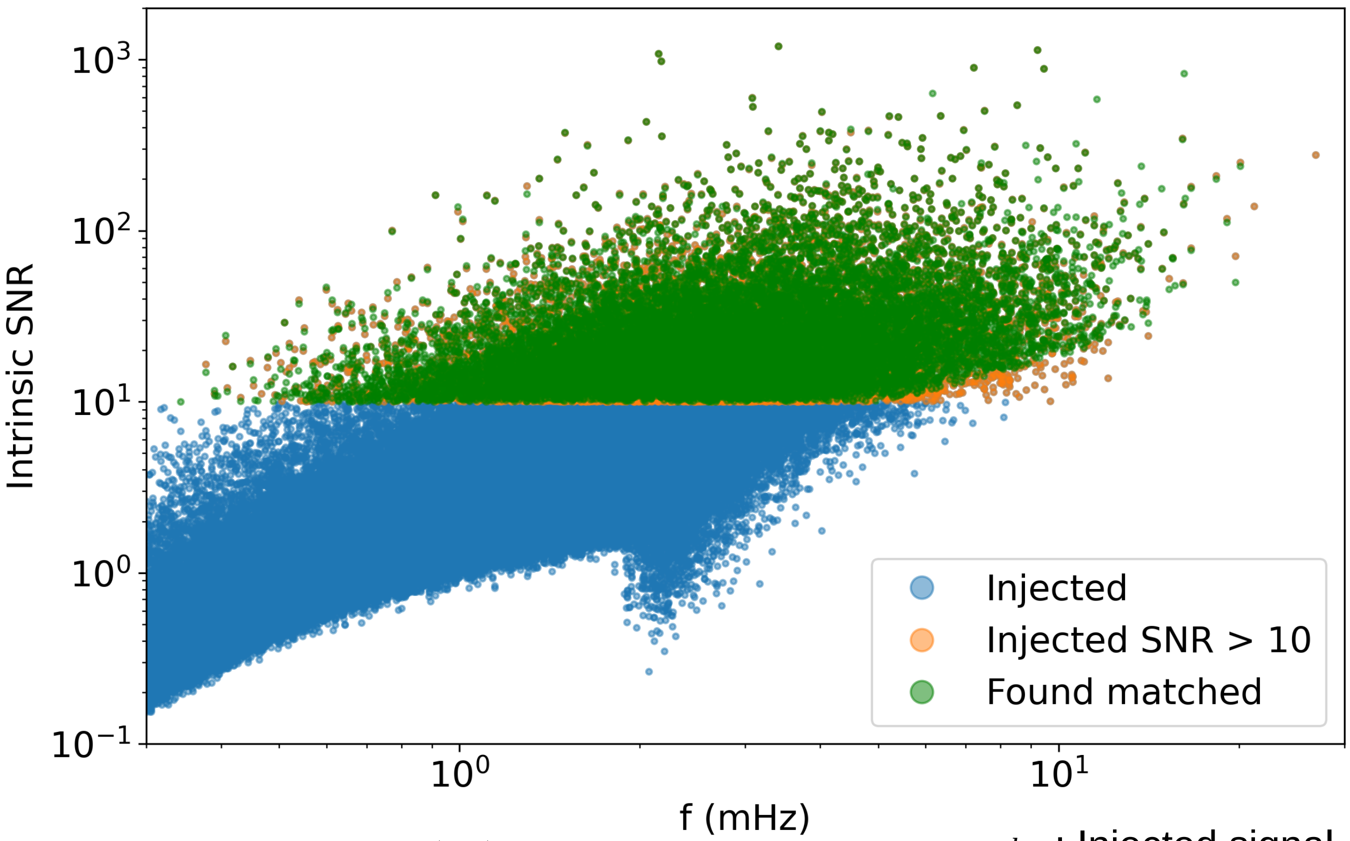
LDC 1-4 – 2 years of data

	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954
Found	18’901



	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954
Found	18’901
Found matched	16’621

LDC 1-4 – 2 years of data



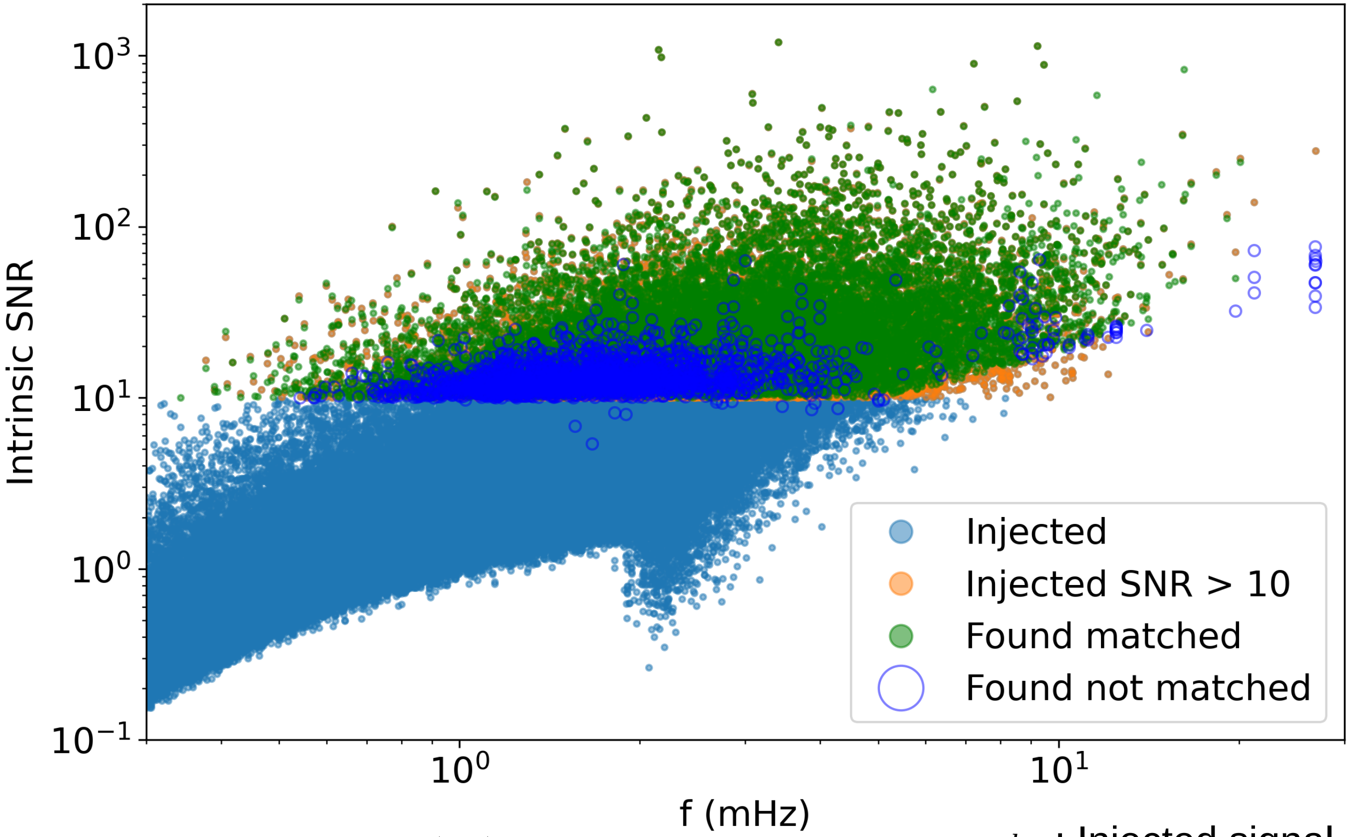
Match if: $\frac{\langle h|s\rangle}{\sqrt{\langle s|s\rangle\langle h|h\rangle}} > 0.5$

h : Injected signal

s : Found signal

LDC 1-4 – 2 years of data

	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954
Found	18’901
Found matched	16’621
Found not matched	2’280



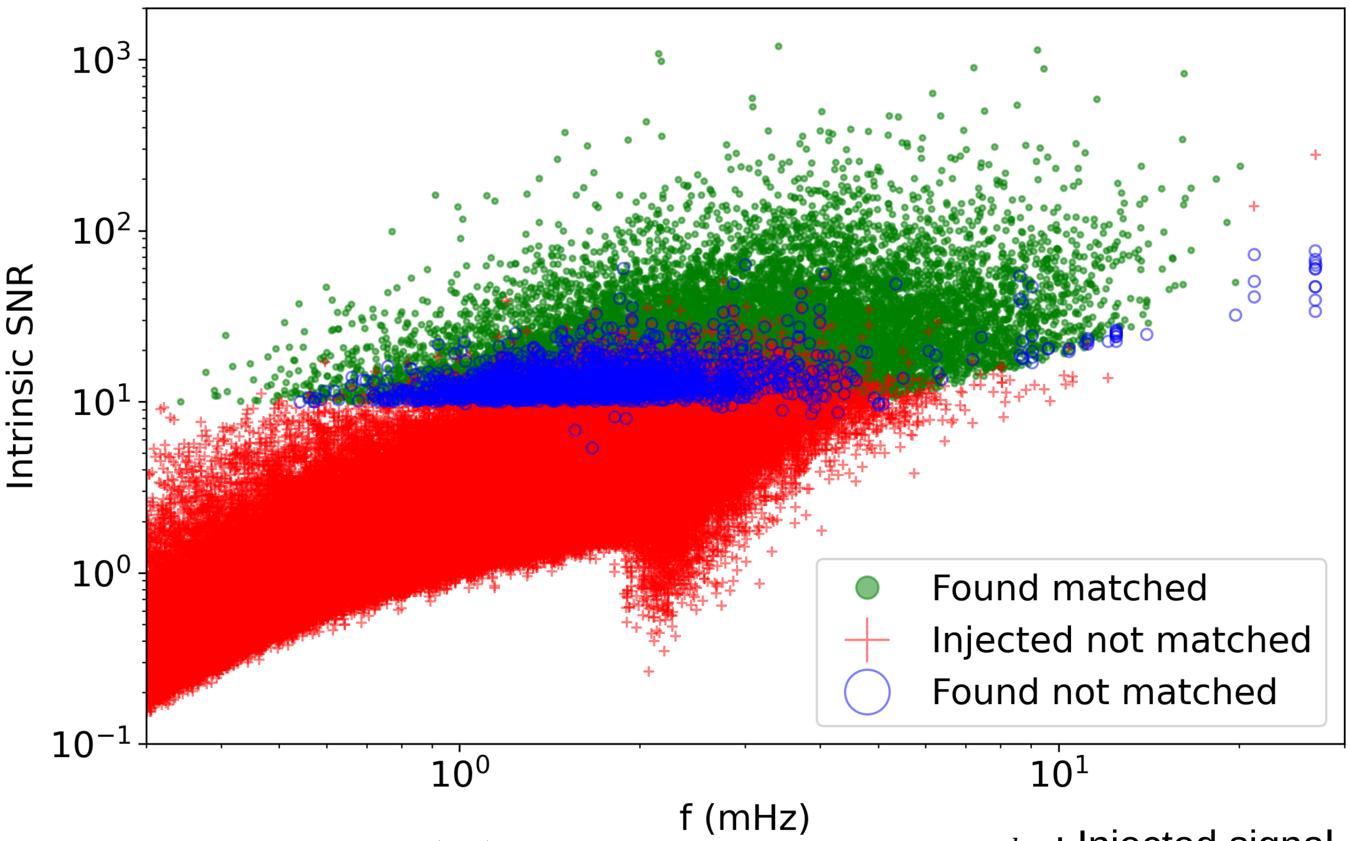
Match if: $\frac{\langle h|s\rangle}{\sqrt{\langle s|s\rangle\langle h|h\rangle}} > 0.5$

h : Injected signal

s : Found signal

	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954
Found	18’901
Found matched	16’638
Found not matched	2’263
Match rate Found matched vs Found	88%

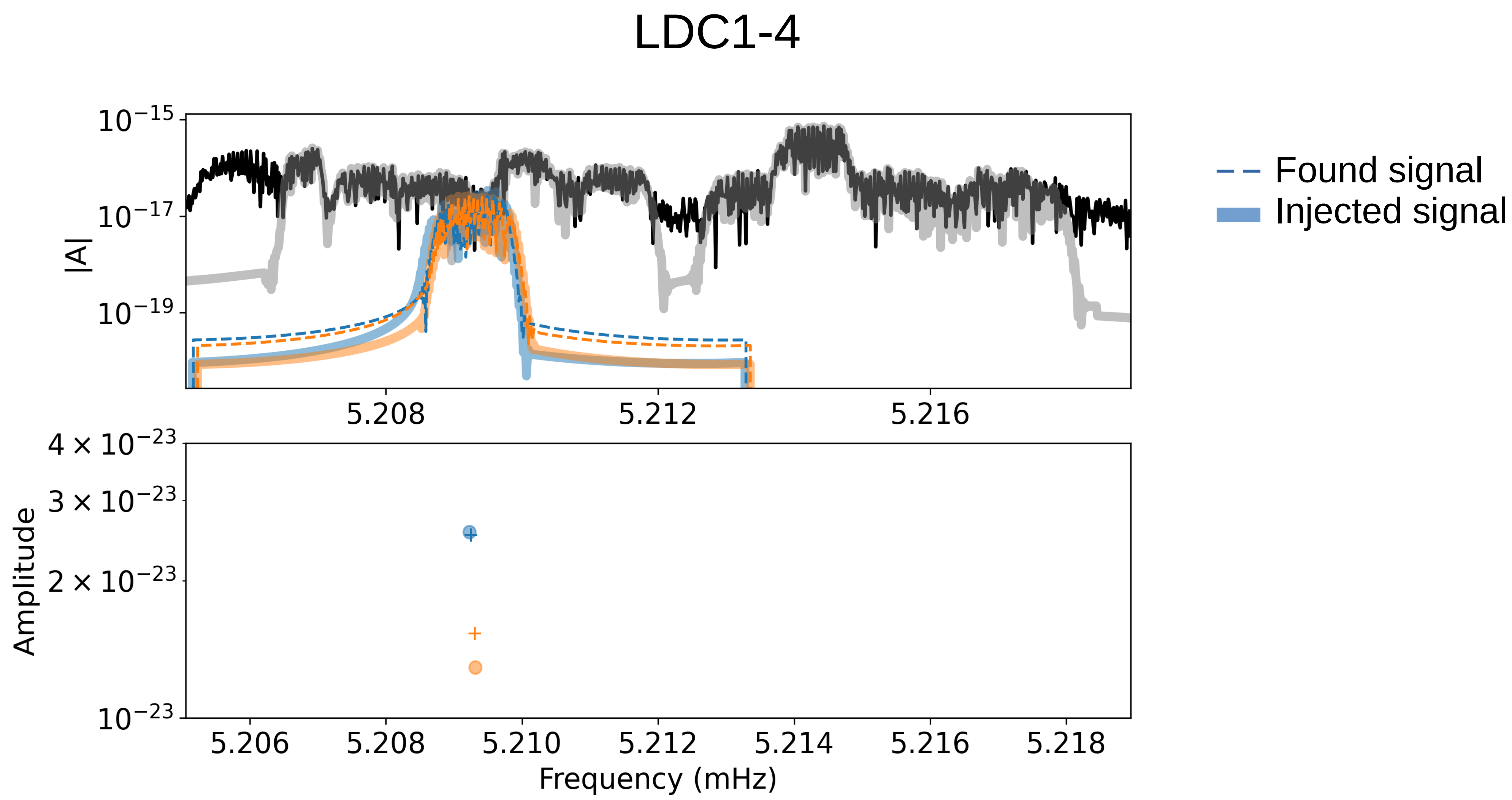
LDC 1-4 – 2 years of data



Match if: $\frac{\langle h|s\rangle}{\sqrt{\langle s|s\rangle\langle h|h\rangle}} > 0.5$

h : Injected signal

s : Found signal



LDC1-4
Found signal
Injected signal
|A|
10^{-15}
10^{-17}
10^{-19}
5.208
5.212
5.216
Amplitude
4×10^{-23}
3×10^{-23}
2×10^{-23}
10^{-23}
5.206
5.208
5.210
5.212
5.214
5.216
5.218
Frequency (mHz)

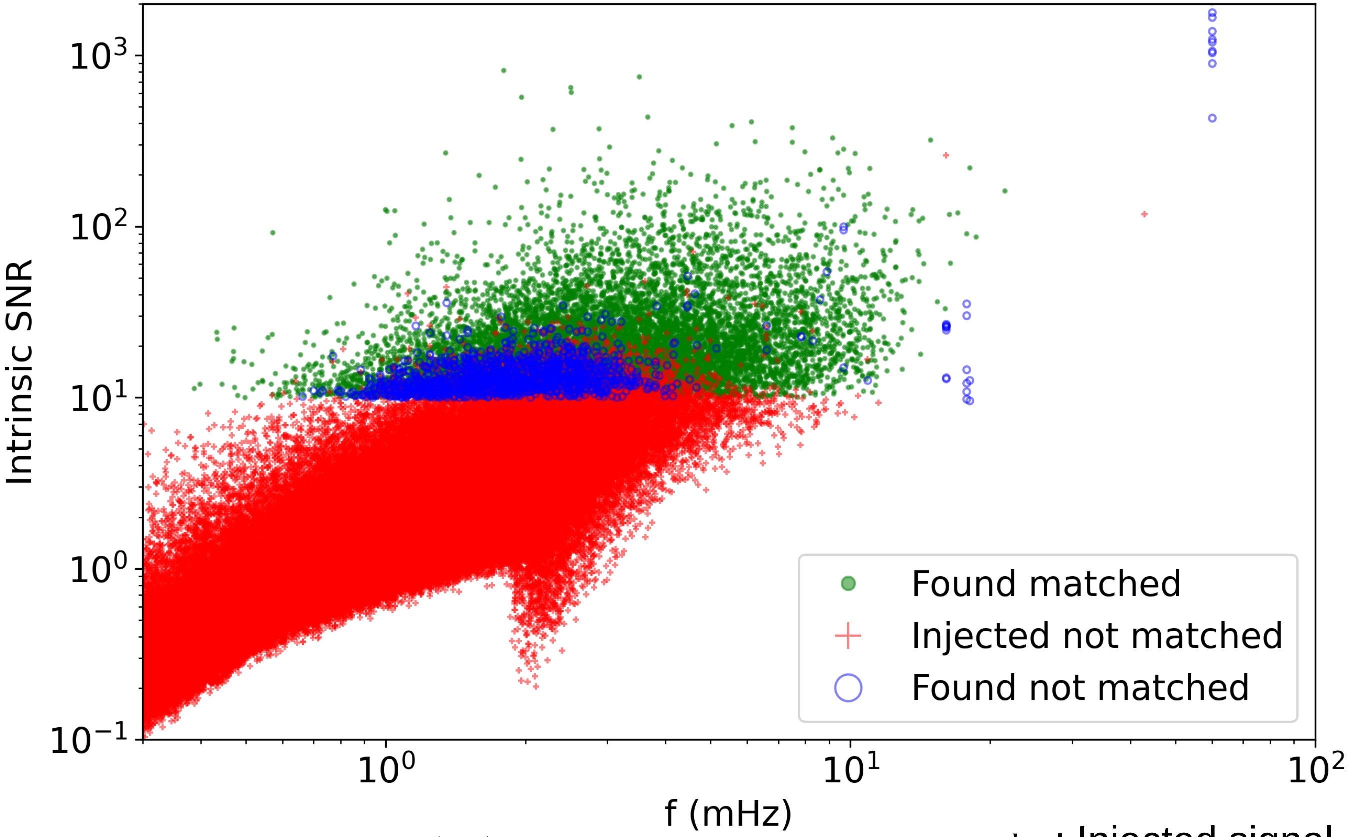
LDC 1-4 – 2 years of data CPU time

Windows	Number of Windows	Max number of signals	CPU Time [h]	CPU Time Parallelized [min]
Even	3485	3	1.6K	27
Odd	3485	10	3.1K	53
Even	3485	10	2.5K	42
Total	10455		7.2K	122

From Maudes' introduction slides: Time estimate was ~200K CPU-hours

	Number of signals
Injected	6’357’082
Injected SNR > 10	11’600
Found	12’338
Found matched	10’933
Found not matched	1’405
Match rate Found matched vs Found	89%

LDC 2 Sangria – 1 year of data



Match if: $\frac{\langle h|s\rangle}{\sqrt{\langle s|s\rangle\langle h|h\rangle}} > 0.5$

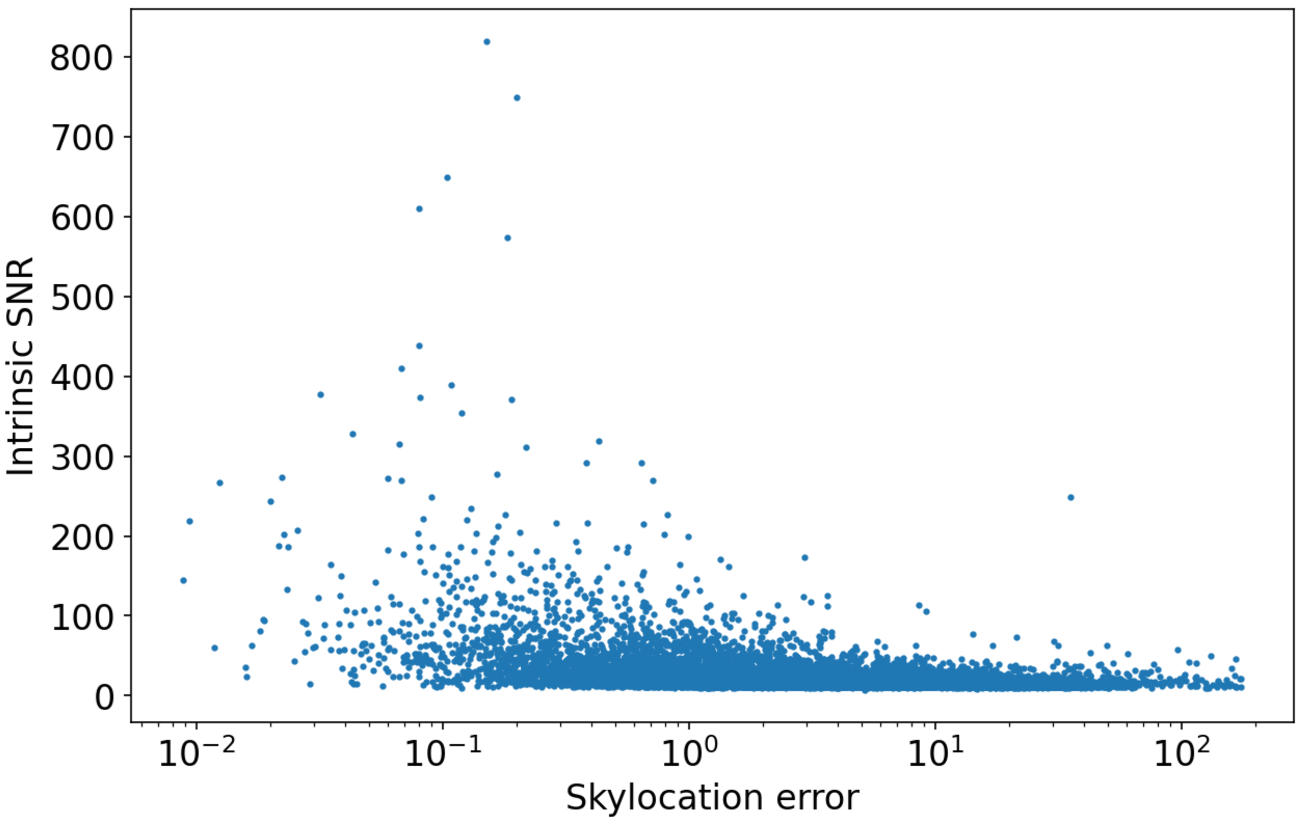
h : Injected signal

s : Found signal

Conclusion

- Fast Bayesian parameter estimation of Galactic Binaries
  (Strub et al. 2022, 10.1103/PhysRevD.106.062003)
- Genetic Algorithms are efficient in finding the best fit solution
- Gaussian Process Regression accurately models the log likelihood function
  (so far only for unimodal posteriors, faint signals with SNR < 7 could be multimodal)
- Robust pipeline to extract GBs from the full Galaxy
- Low computational cost: Only 7.2K CPU-hours for 2 years of data
- Low latency detection due to parallelization: ~2h for full galaxy with 3’500 CPU threads

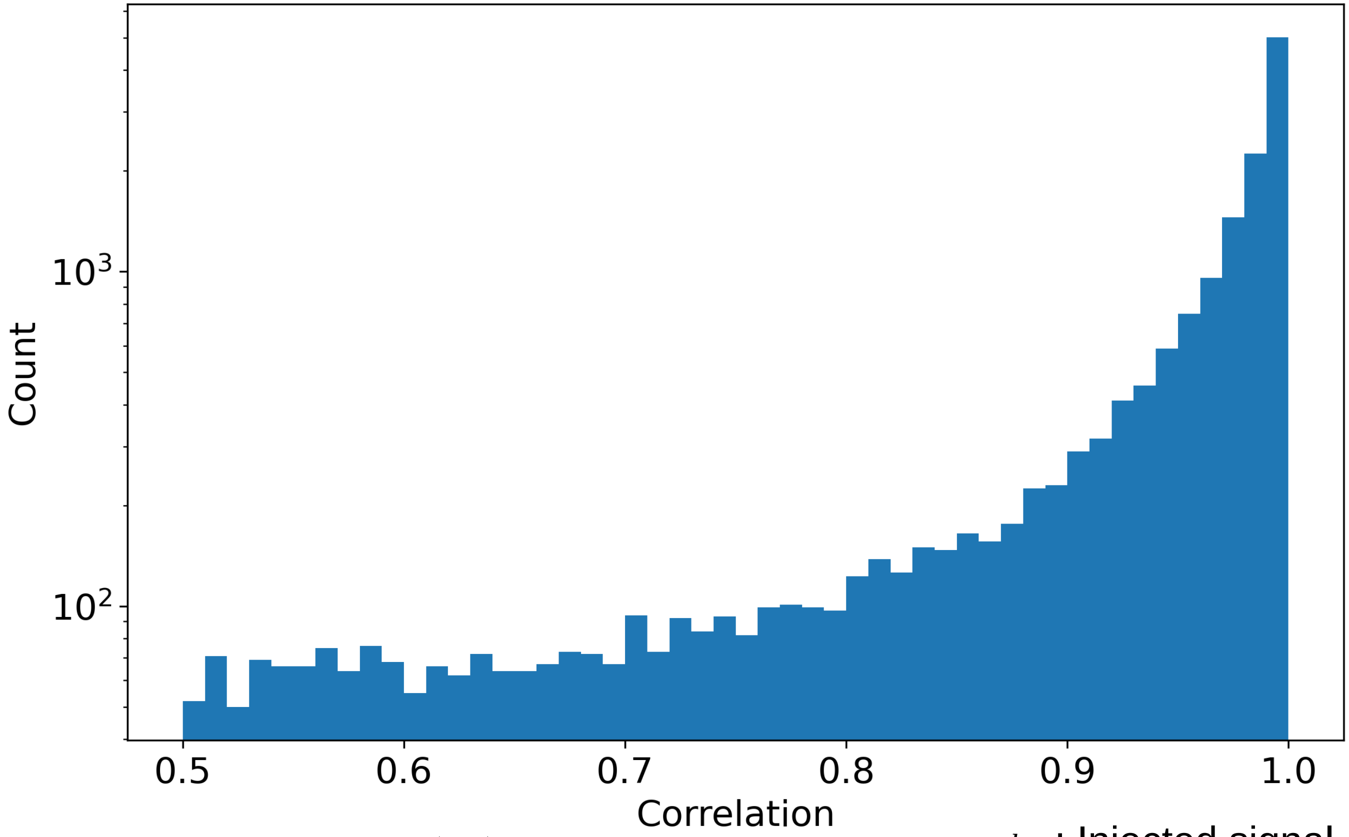
Errors



Intrinsic SNR
800
700
600
500
400
300
200
100
0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
Skylocation error

LDC 1-4 – 2 years of data

	Number of signals
Injected	6’836’835
Injected SNR > 10	17’954
Found	18’901
Found matched	16’621
Found ambiguous	2’280
Match rate Found matched vs Found	88%

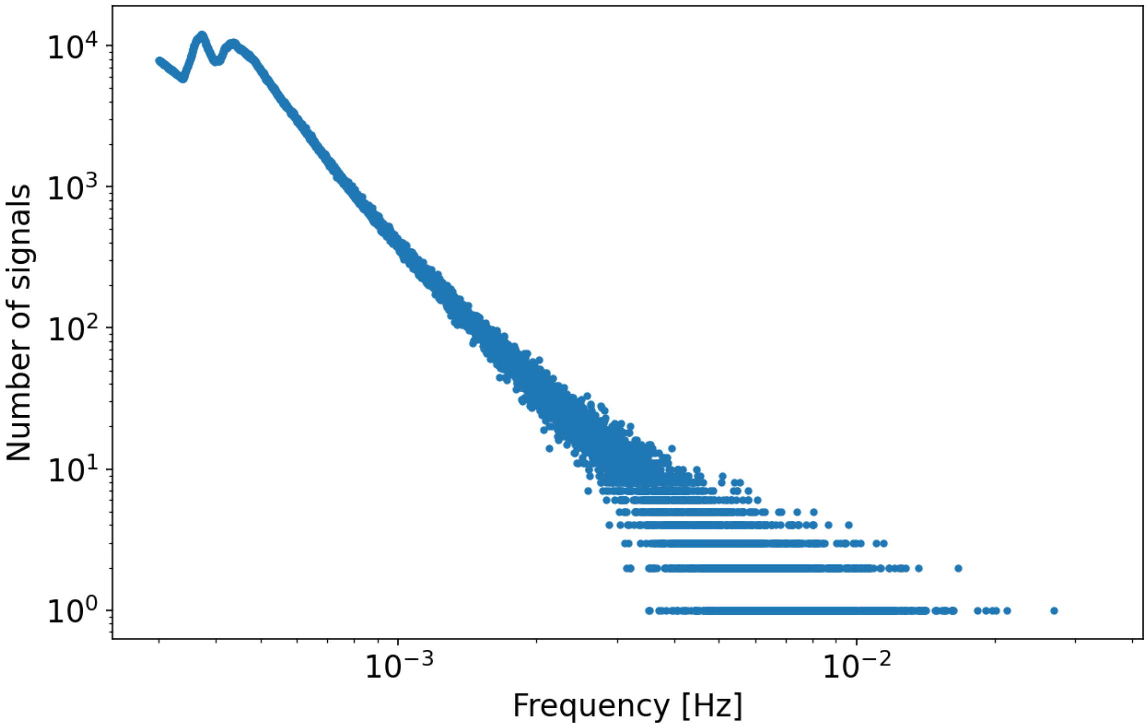


Match if: $\frac{\langle h|s\rangle}{\sqrt{\langle s|s\rangle\langle h|h\rangle}} > 0.5$

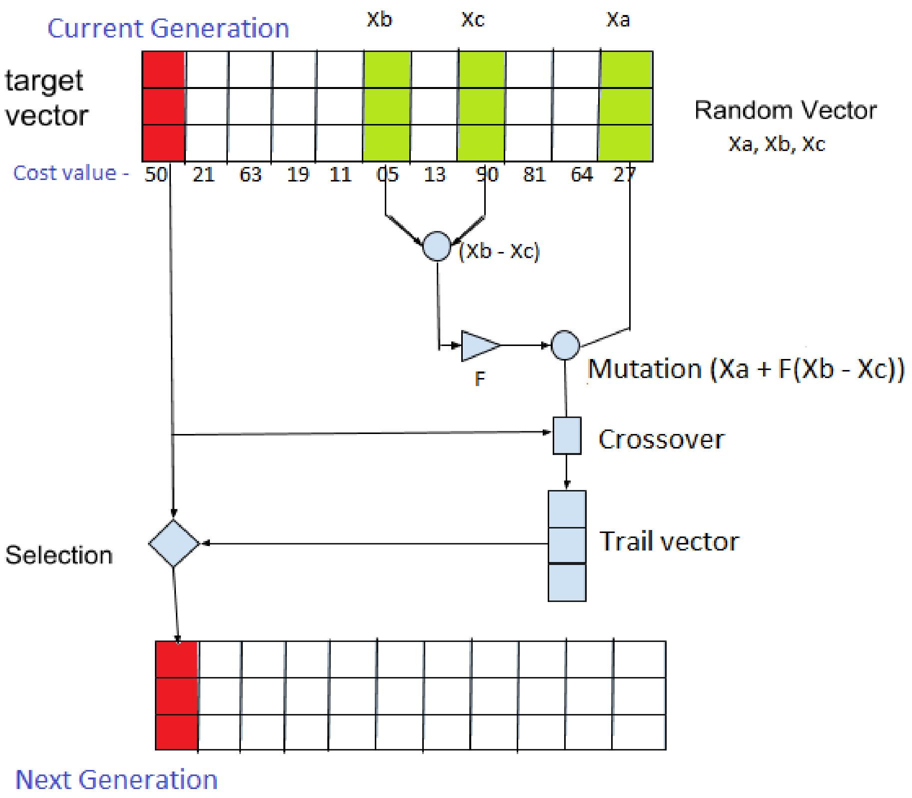
h : Injected signal

s : Found signal

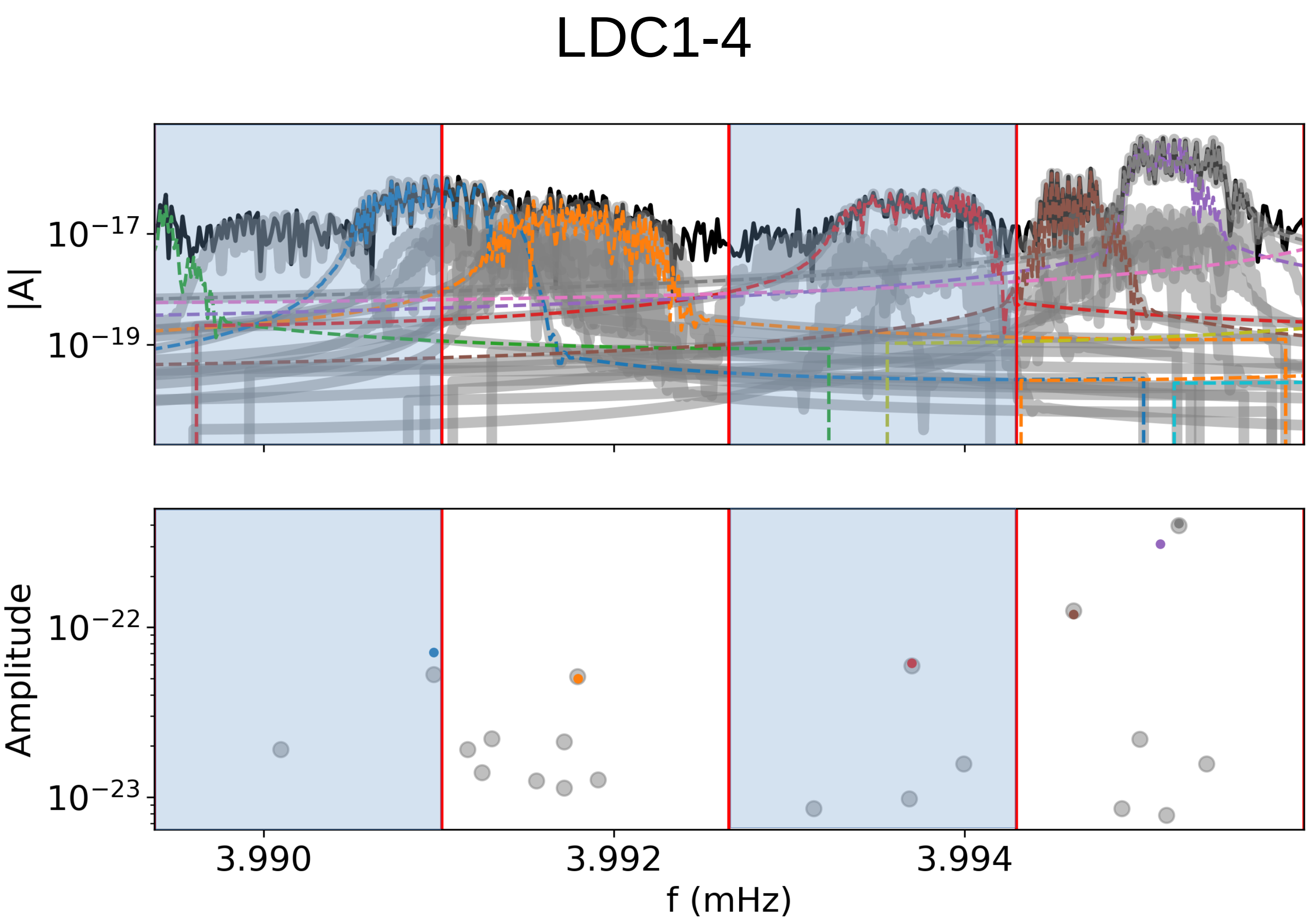
LDC 1-4 Number of signals per frequency window



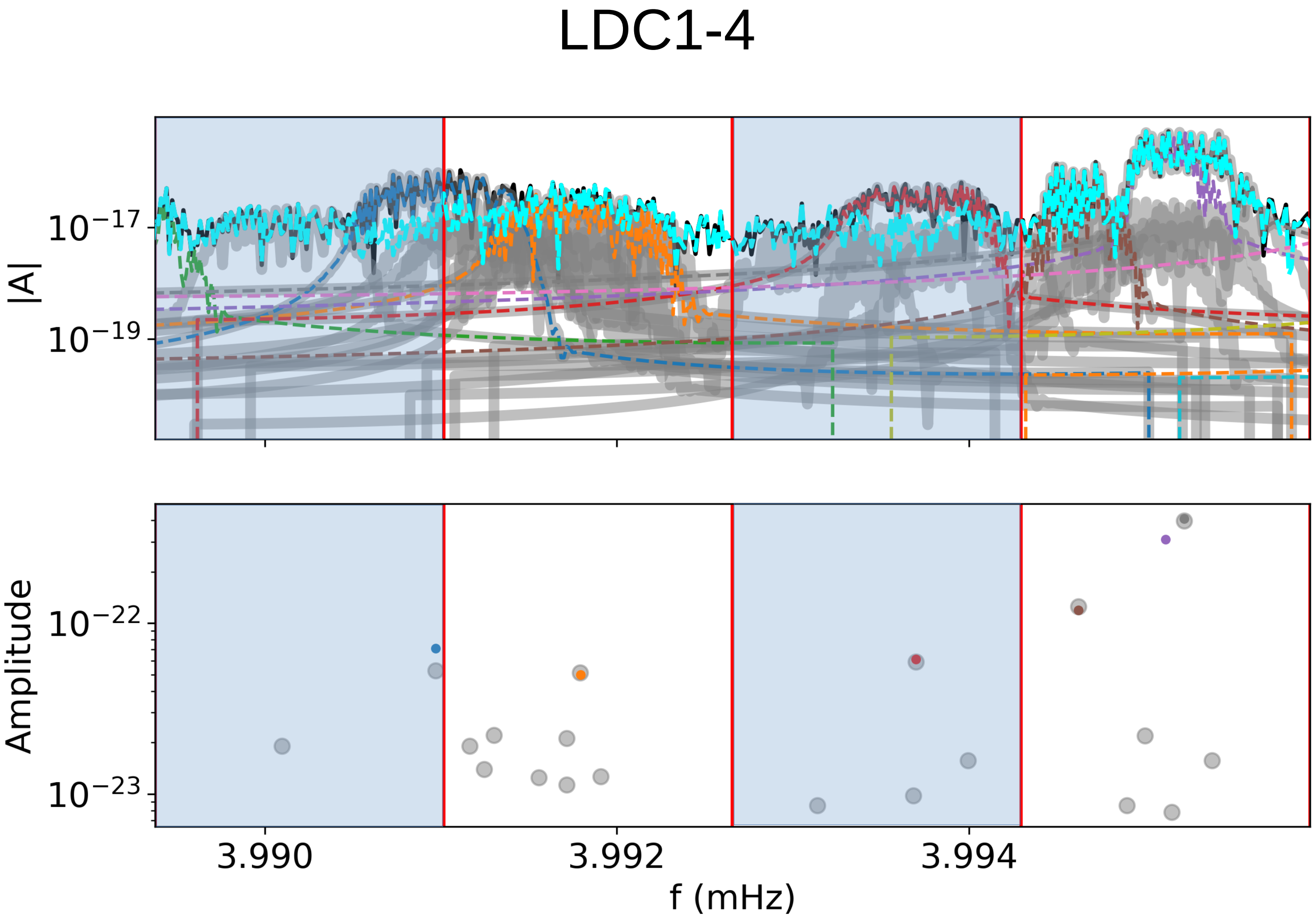
Differential Evolution



Source: Medium, Abhishek Patel

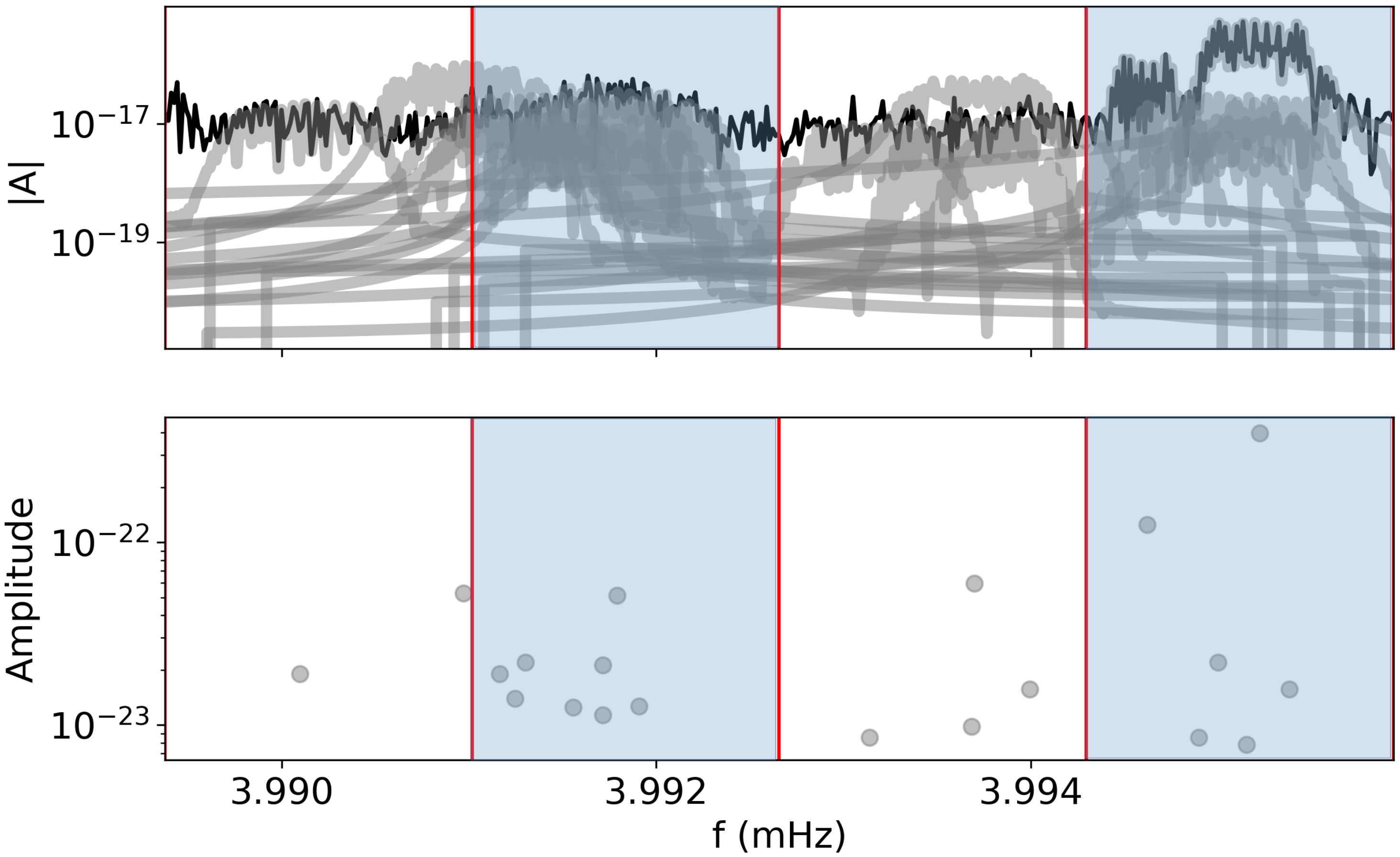


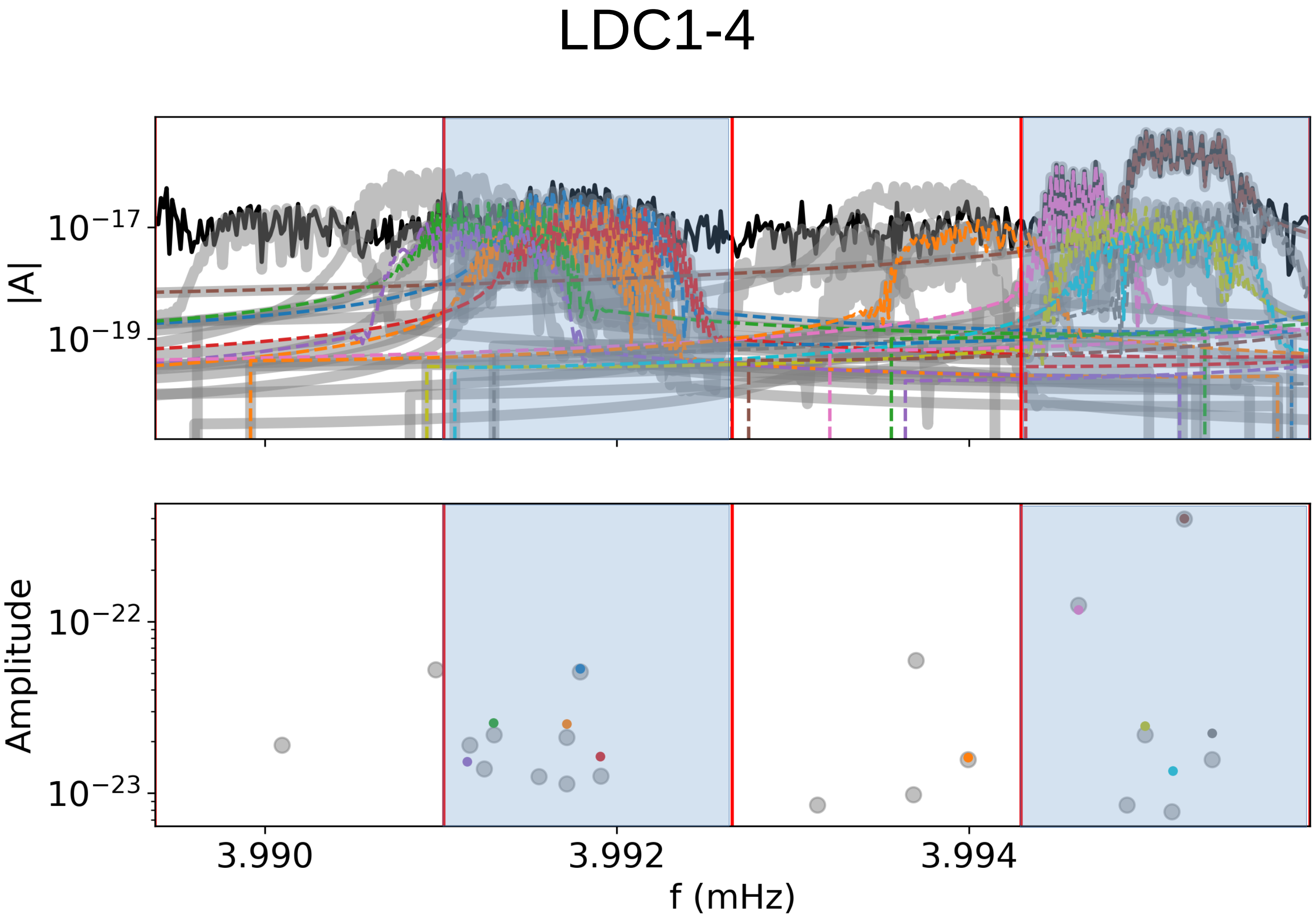
LDC1-4
|A|
10^{-17}
10^{-19}
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)



LDC1-4
|A|
10^{-17}
10^{-19}
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)

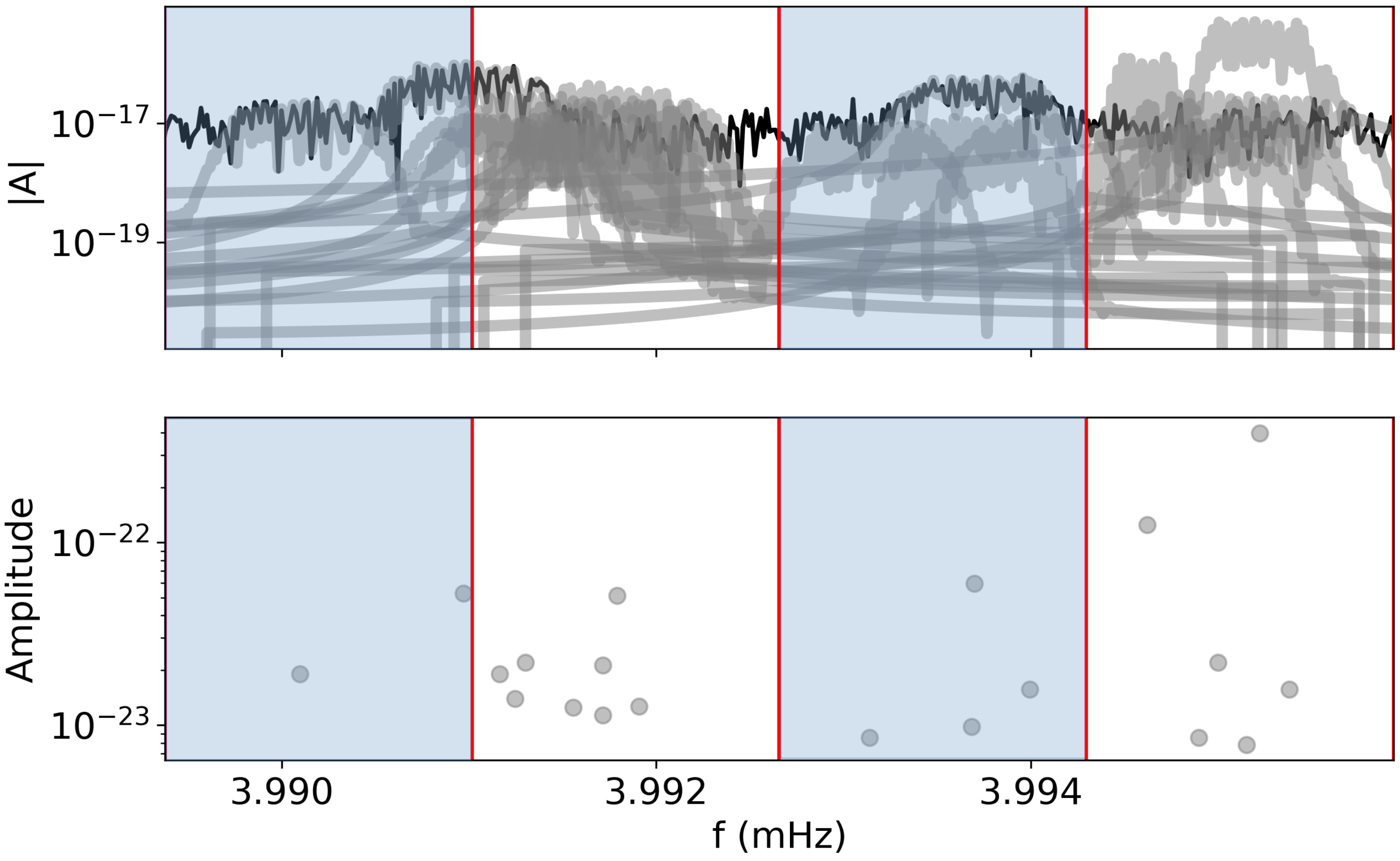
LDC1-4

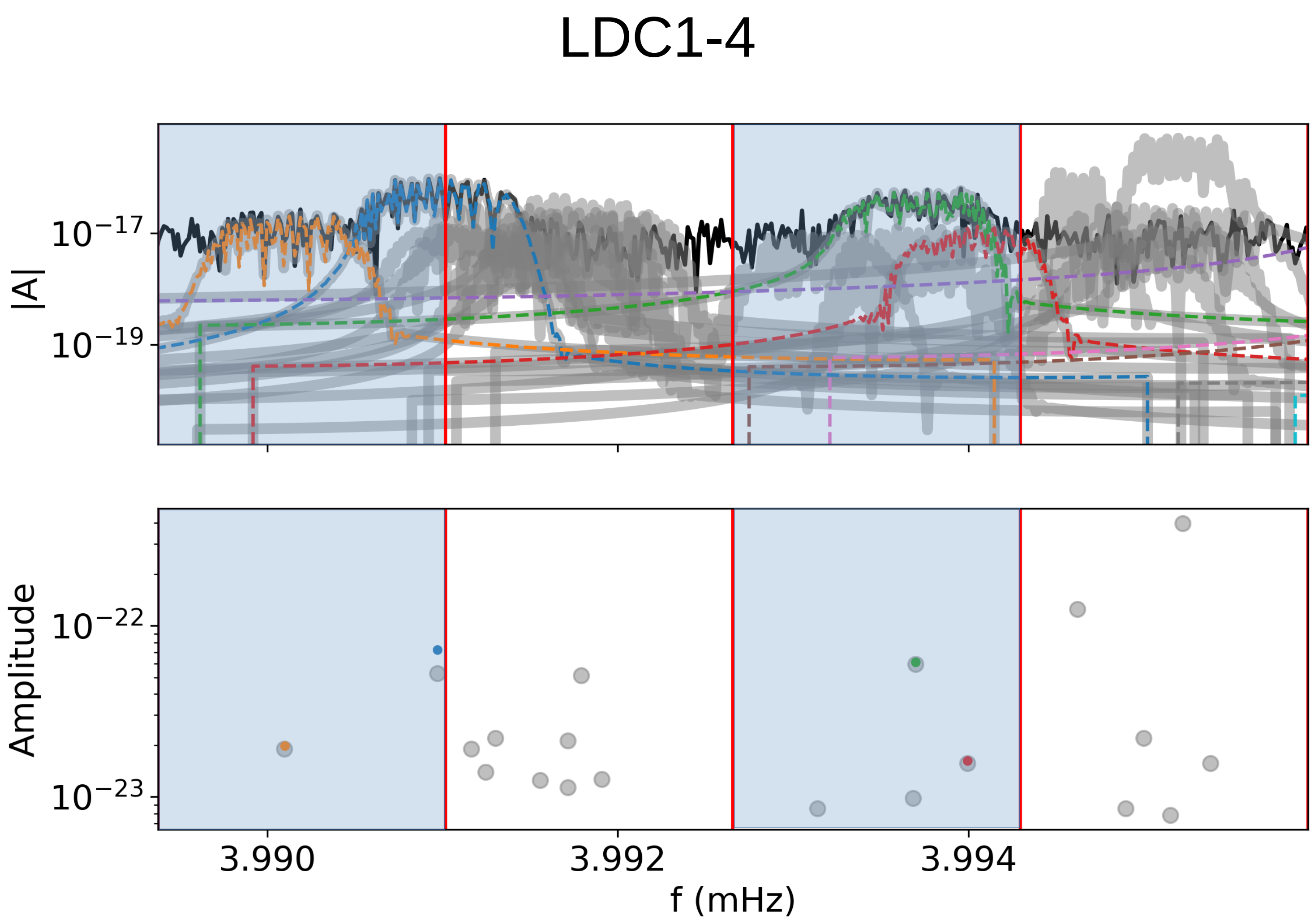




LDC1-4
|A|
10^{-17}
10^{-19}
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)

LDC1-4





LDC1-4
|A|
10^{-17}
10^{-19}
Amplitude
10^{-22}
10^{-23}
3.990
3.992
3.994
f (mHz)