



MAX-PLANCK-GESELLSCHAFT

**Workshop on LISA data analysis: from
classical to machine learning methods**



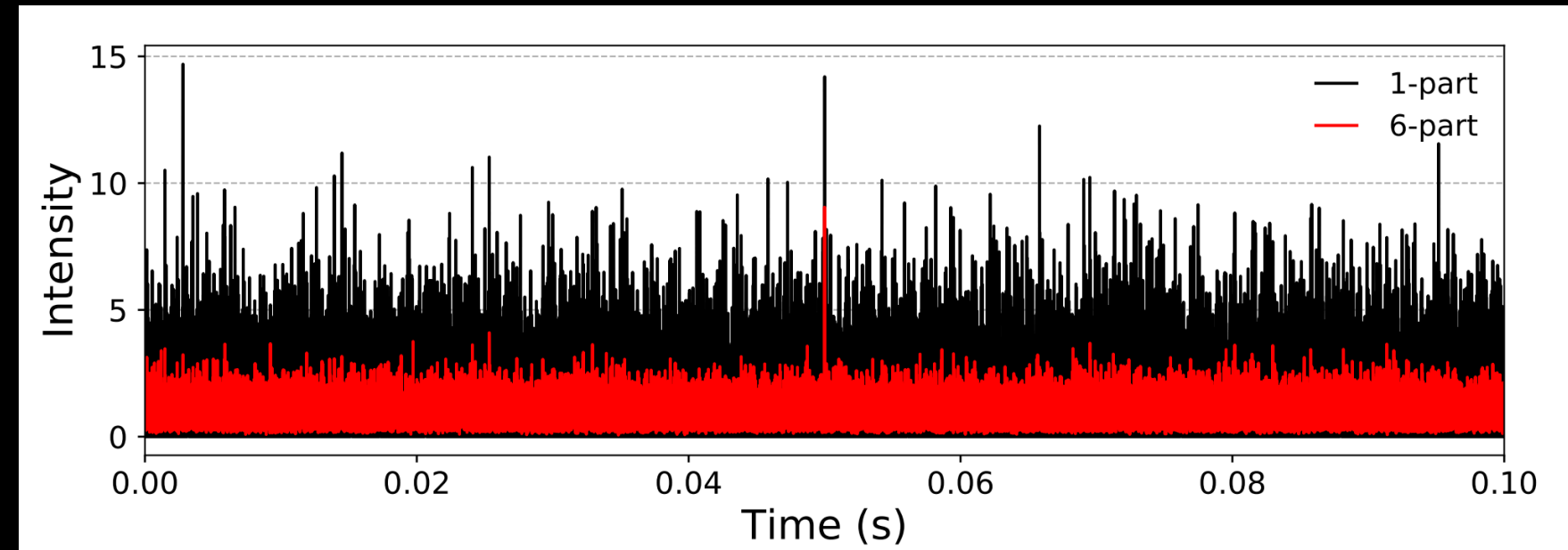
Henri Martin

Toulouse, France
November 23, 2022

Introduction - current radio projects

Eigenfiltering - new technique to search for ultra-fast transients

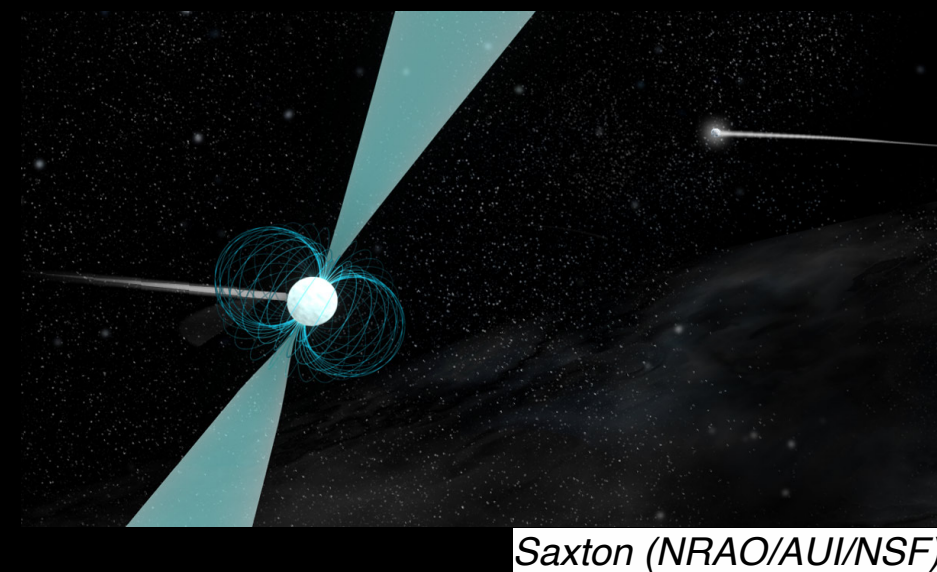
[10.1088/1361-6382/ab95e4](https://doi.org/10.1088/1361-6382/ab95e4)



MeerTime acceleration project - imprint of source acceleration on random Gaussian light emission

- binary pulsars

- quasars



VAMPIRA - provenance collection from a computationally intensive pipelines

[arXiv:2109.10759](https://arxiv.org/abs/2109.10759)

(astronomy lead)

Introduction - current GW+radio projects

EPTA collaboration



- pulsar timing

-absorption of GWs by the IGM

[10.1088/1361-6382/ac5376](https://doi.org/10.1088/1361-6382/ac5376)

- emission of GWs by SMBH accretion disks

-Effelsberg observing

MeerKAT



- check if new binaries discovered by MeerKAT are detectable by LISA (student project)

Introduction - current GW projects

LISA

- data analysis, parameter estimation
from MCMCs

- code development

<https://github.com/tlittenberg/ldasoft>

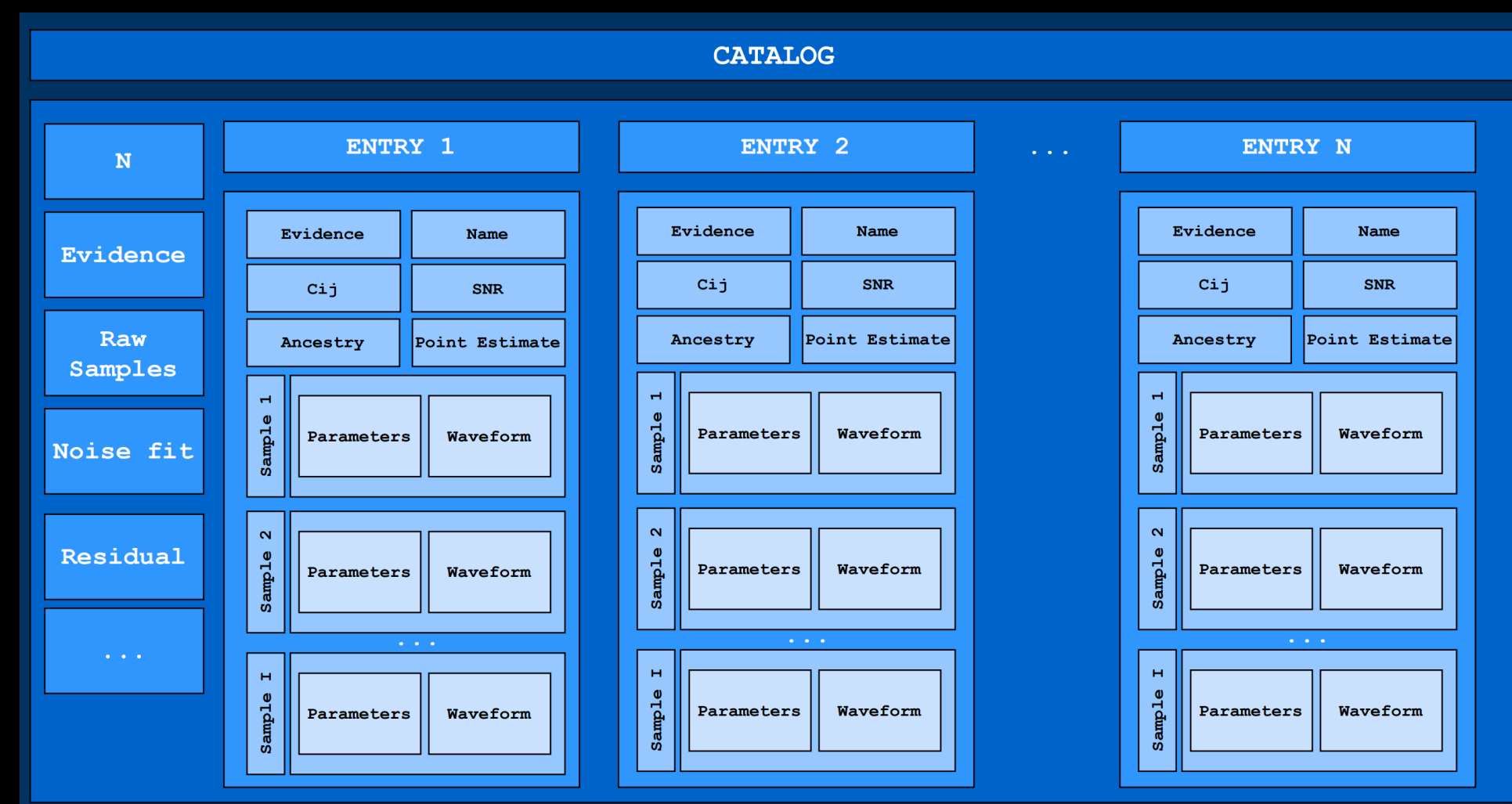
```
00000 00000 00000000 1111100000
0000000 1111111 0100000 110000
00000000 11
000001111 111
0000 1111 111
1111 00000011
11100000000000
000000000000
000000000000
000000000000
0000000000
000000
```

=====
GBMCMC Version: =====
Git commit: f5ba624059e5f51db6467e1f74eb0029630857a1

- catalog creation

[10.1103/PhysRevD.101.123021](https://arxiv.org/abs/10.1103/PhysRevD.101.123021)

- using ML to sort MCMC
chain data into a catalog



LISA Data Challenge

The Global Fit problem

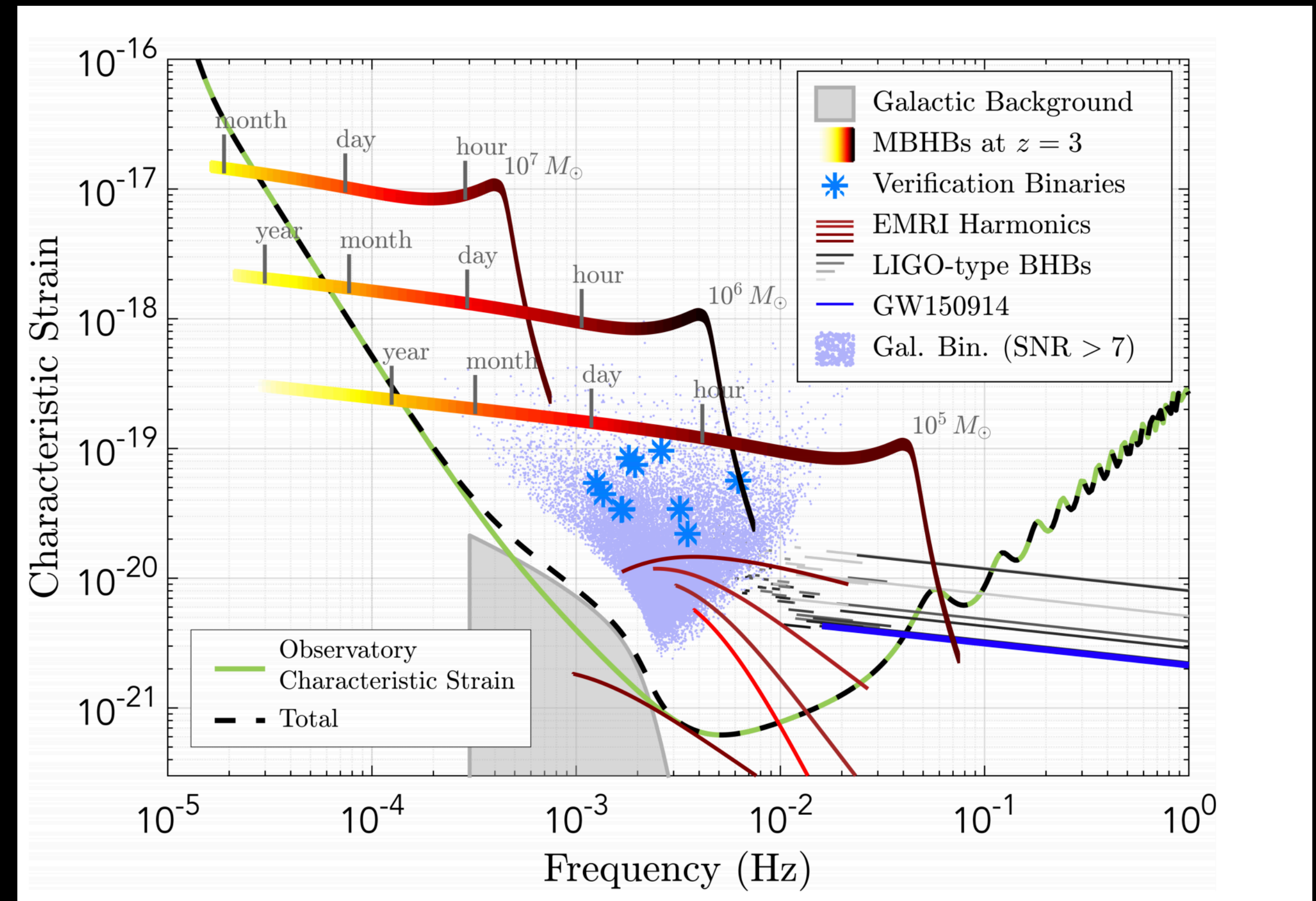
1. Number of detectable sources is an unknown—and is **LARGE**.
2. Sources are overlapping in time and frequency.
3. Individual overlaps between pairs are small. But see 1 above.

Massive black hole binaries

Galactic binaries

Extreme mass ratio inspirals

Stochastic GW background (cosmological and astrophysical origin)



Astro white paper: Amaro-Seoane et al., 2022

https://indico.in2p3.fr/event/27706/contributions/116314/attachments/74022/106498/lejeune_intro_lisada_.pdf

MCMC

MCMC

Given some **model** for the data: M

MCMC

Given some **model** for the data: M

...with **parameters**: X

MCMC

Given some **model** for the data: M

...with **parameters**: X

...the **posterior probability density** for the parameters is:

$$p(x | d, M) = \frac{p(d | x, M)p(x | M)}{p(d | M)}$$

Likelihood = “Goodness of Fit” for parameters

Prior = Previously known values for parameters

Marginalized Likelihood = “Goodness of Fit” for model

MCMC

Given some **model** for the data: M

...with **parameters**: X

...the **posterior probability density** for the parameters is:

$$p(x | d, M) = \frac{p(d | x, M)p(x | M)}{p(d | M)}$$

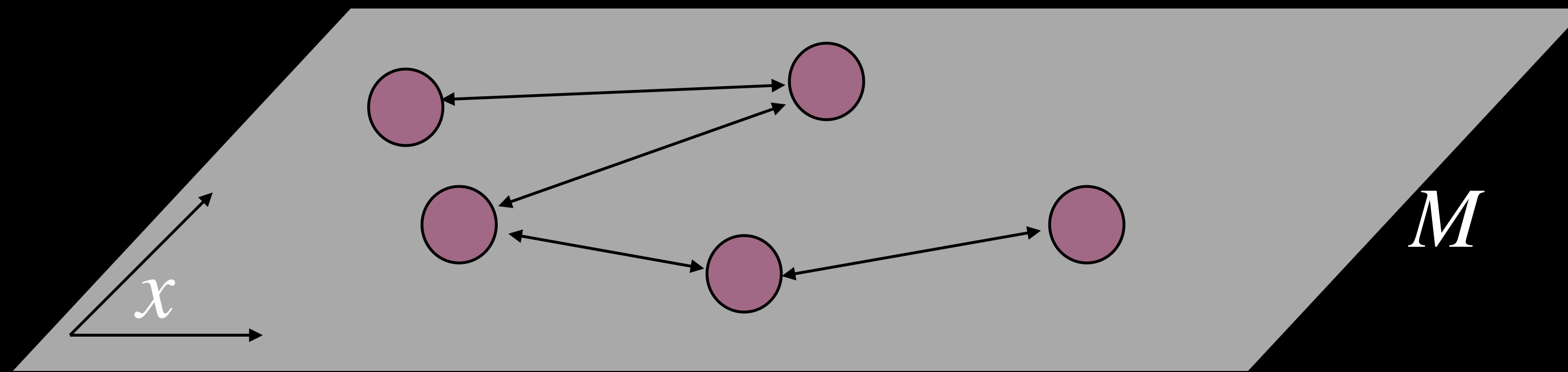
Likelihood = “Goodness of Fit” for parameters

Prior = Previously known values for parameters

Marginalized Likelihood = “Goodness of Fit” for model

MCMC produces independent samples from $p(x | d, M)$

MCMC

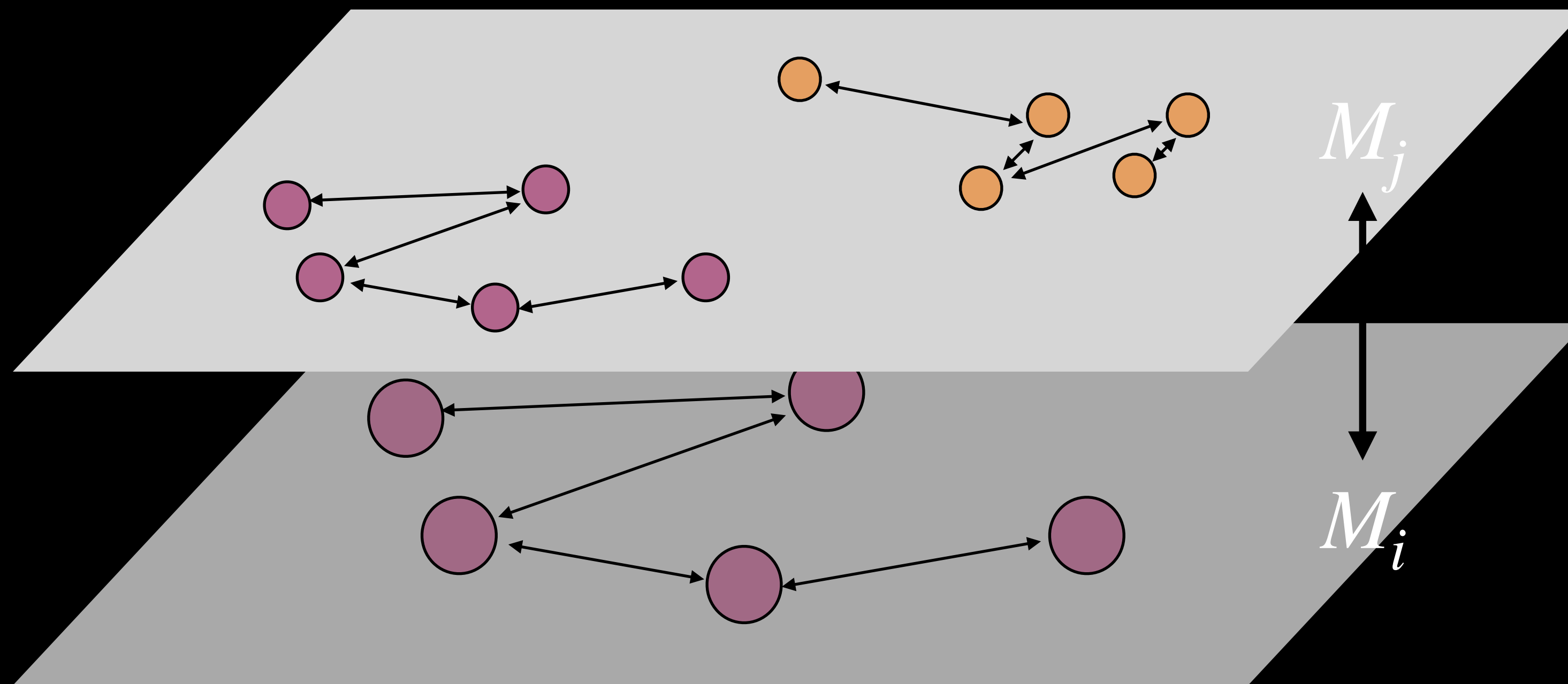


$$p(x | d, M) = \frac{p(d | x, M)p(x | M)}{p(d | M)}$$

- I. Stochastically sample large and complicated parameter spaces
- II. Always converges, usually faster than grid-based approaches when parameter space is **LARGE**.
- III. Detection, characterization, and quantifying confidence
- IV. Stochastically sample between models with RJMCMC

Why MCMC?

Why MCMC?



I. Stochastically sample large and complicated parameter spaces

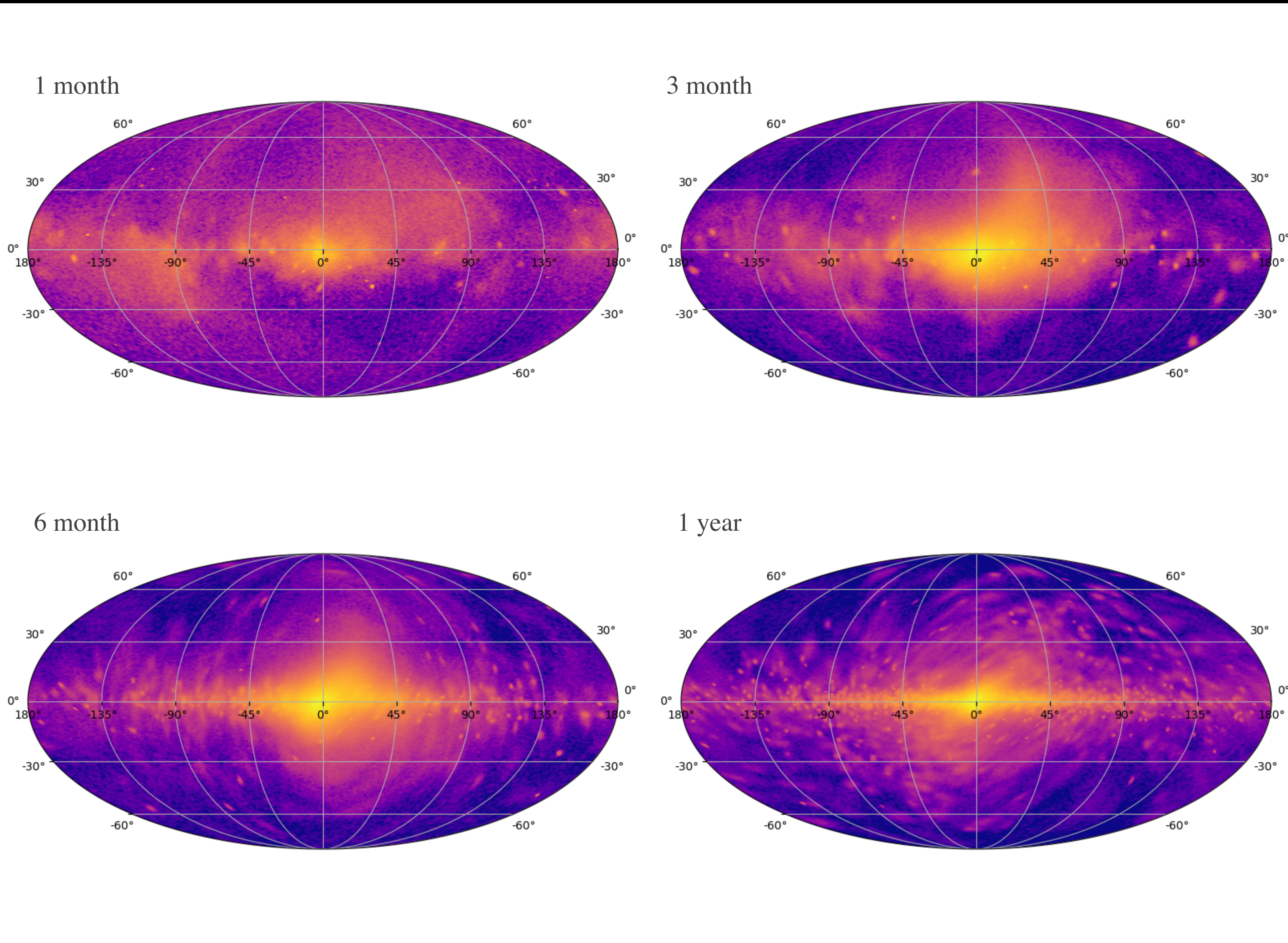
II. Always converges, usually faster than grid-based approaches when parameter space is **LARGE**.

III. Detection, characterization, and quantifying confidence

IV. Stochastically sample between models with RJMCMC

$$p(x, M | d) = \frac{p(d, M | x)p(x, M)}{p(d, M)}$$

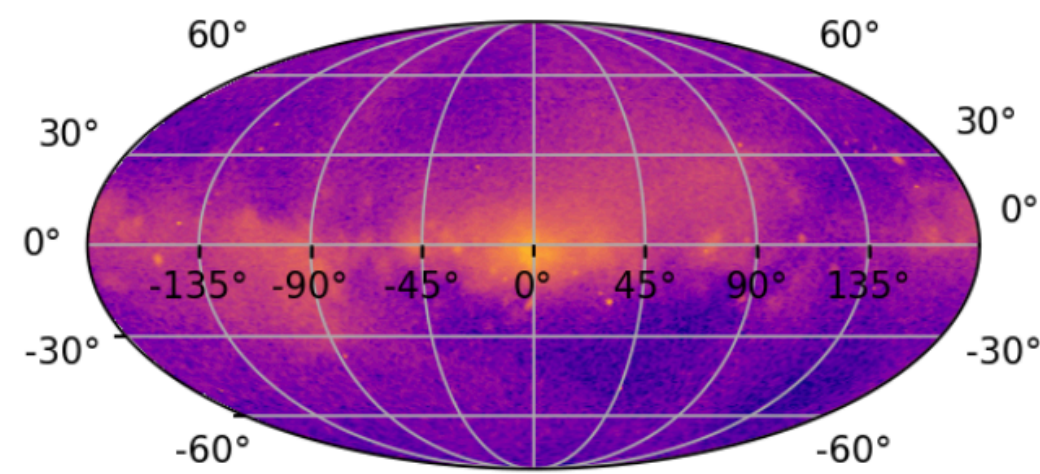
LDC - Challenge 1: *Radler*



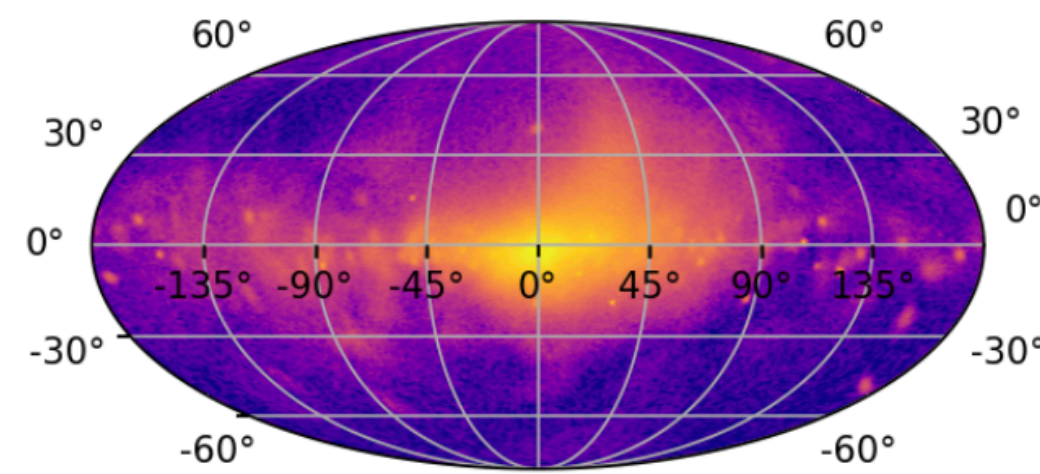
LDC - Challenge 1: *Radler*

Observation time (months)	GBMCMC Catalog detections
1.5	1998
3.0	2758
6.0	6196
12.0	10027

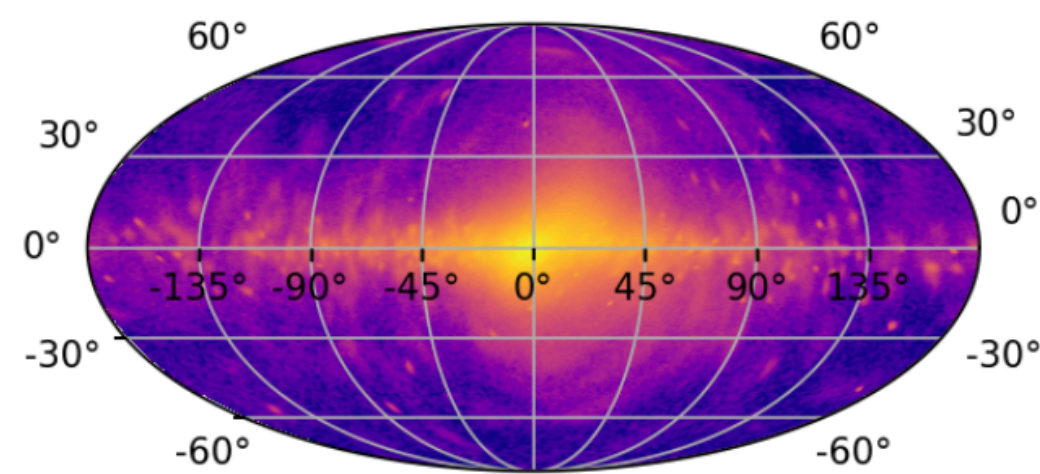
1 month



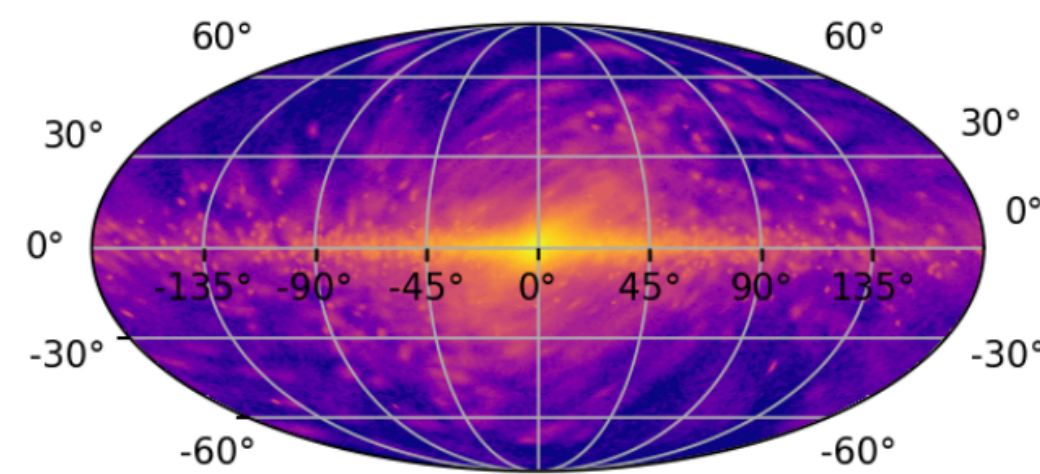
3 month



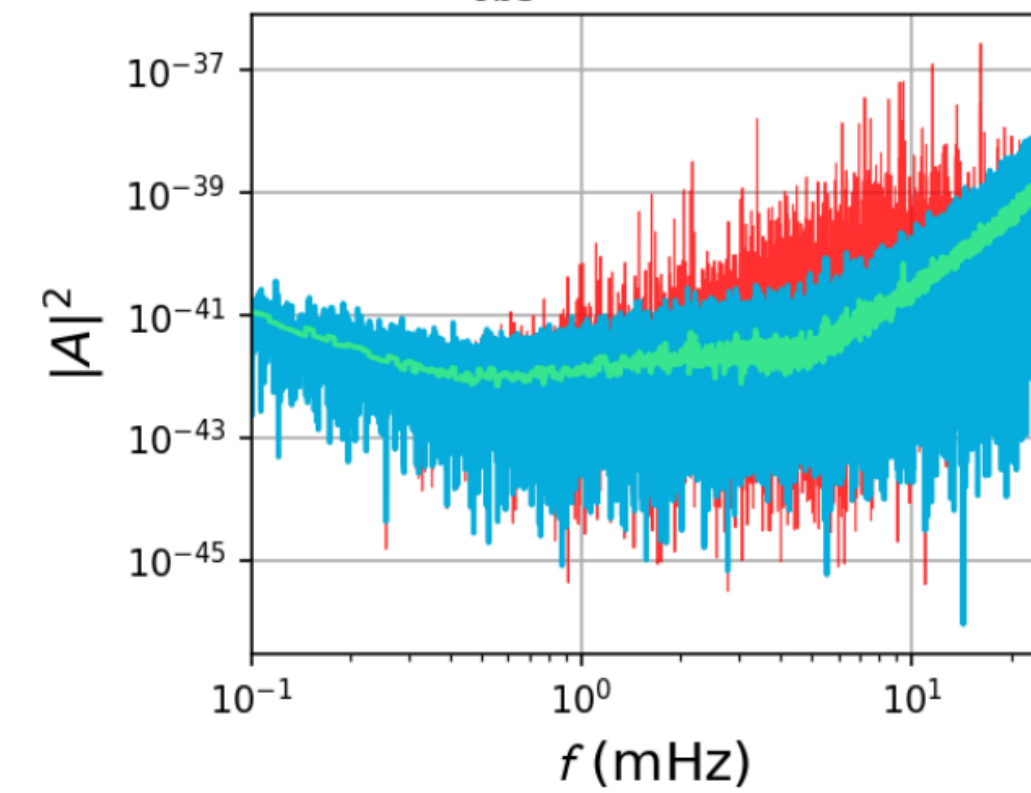
6 month



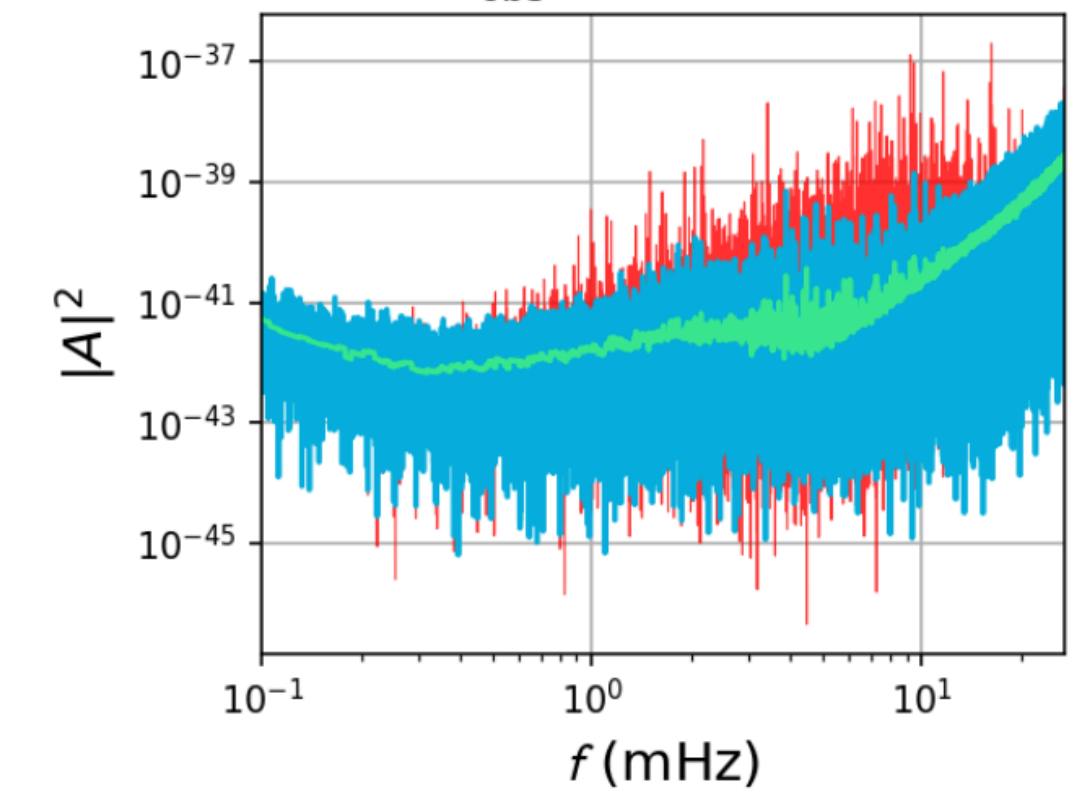
1 year



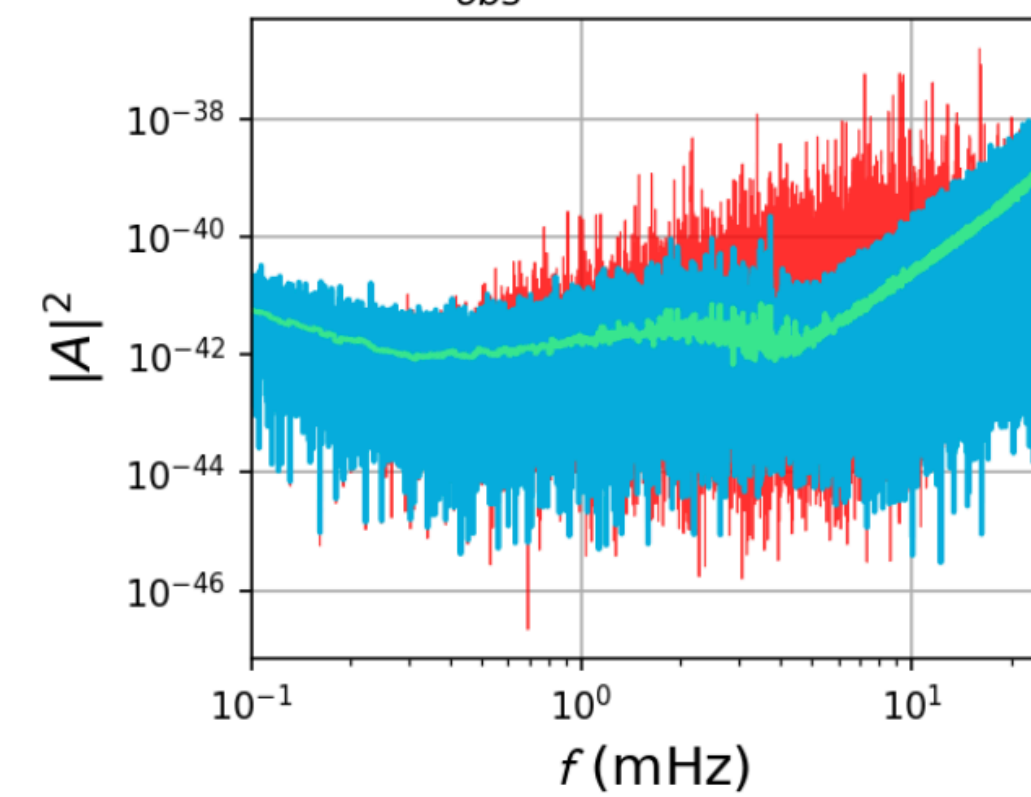
$T_{obs}=3932160$ s



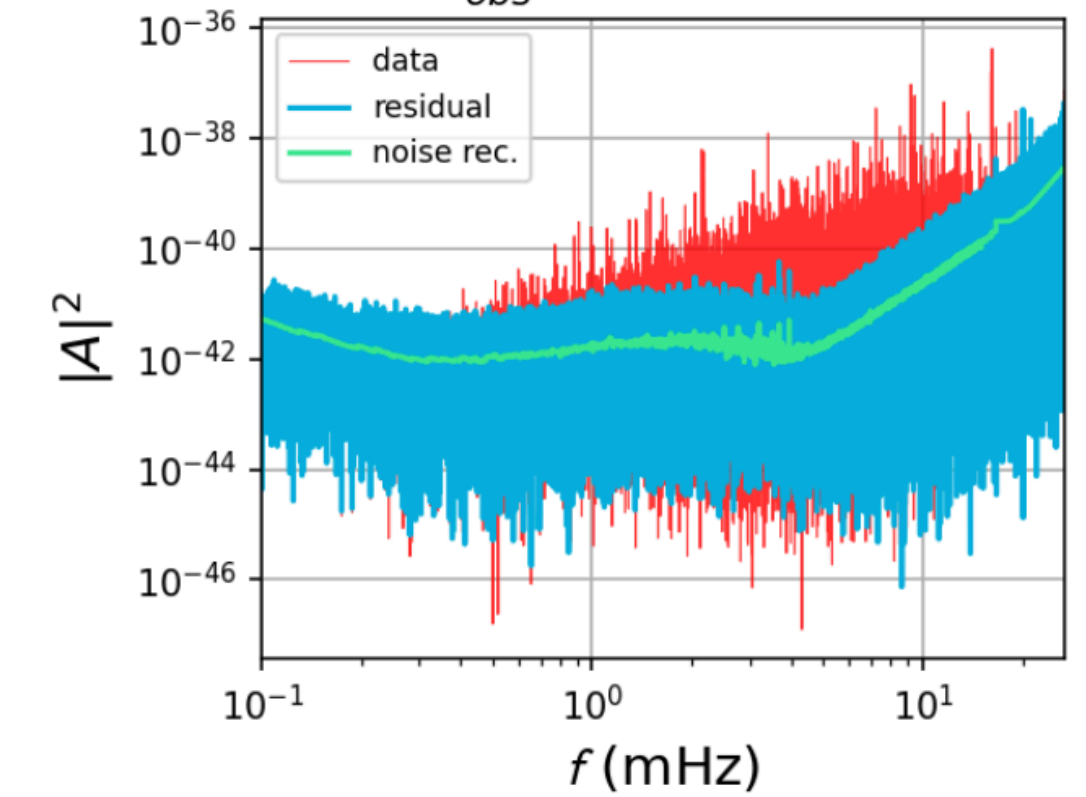
$T_{obs}=7864320$ s



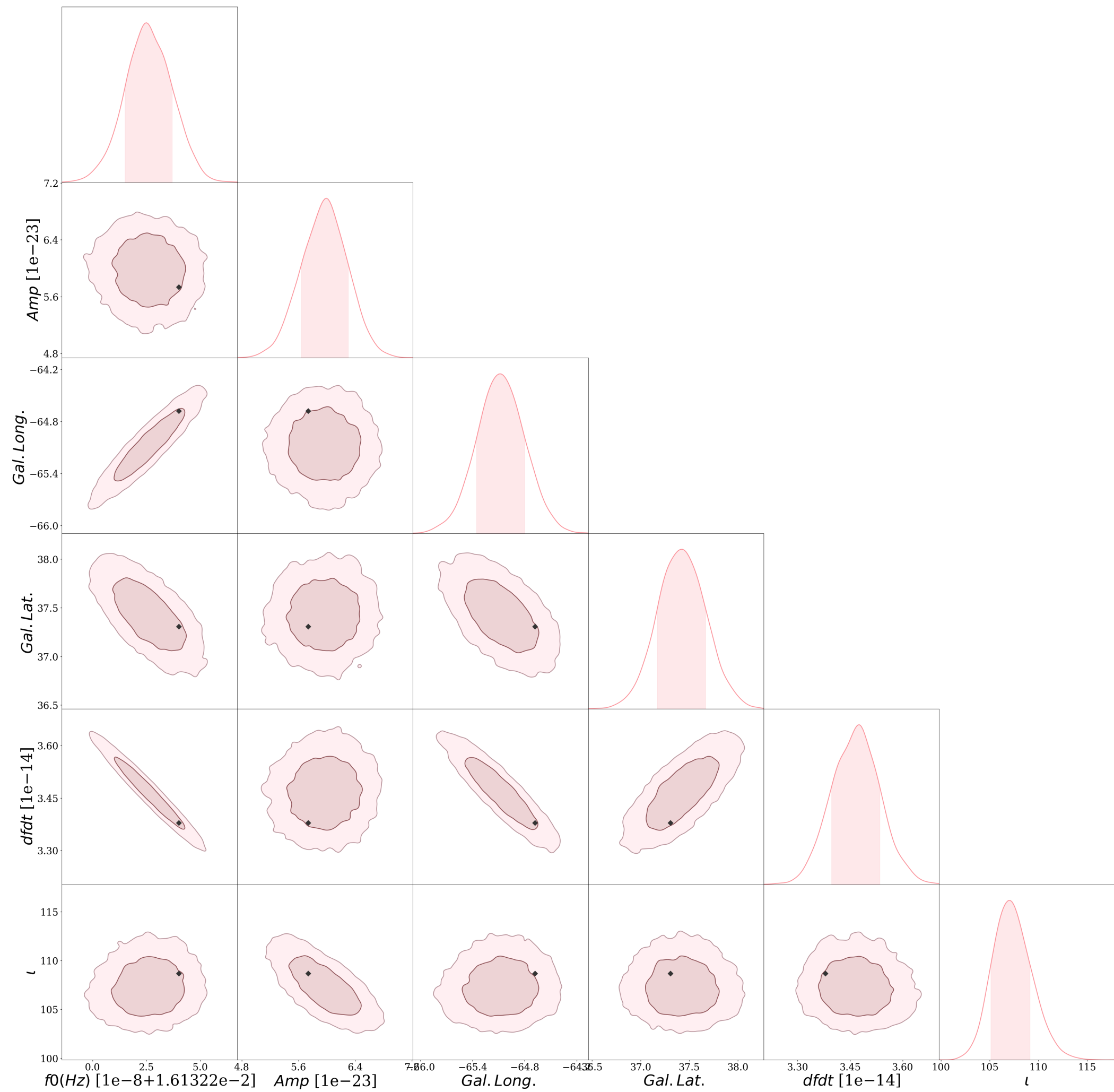
$T_{obs}=15728640$ s



$T_{obs}=31457280$ s



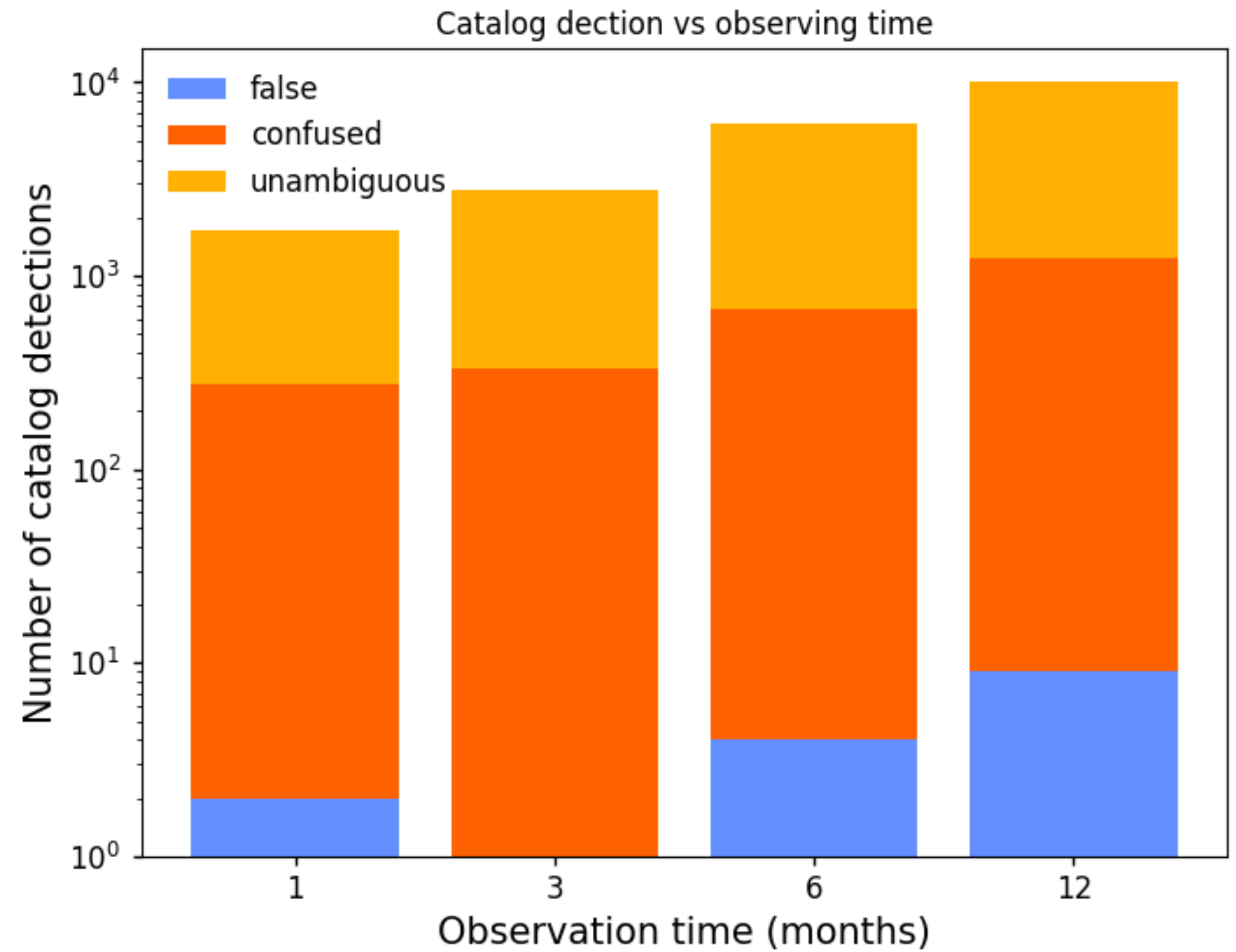
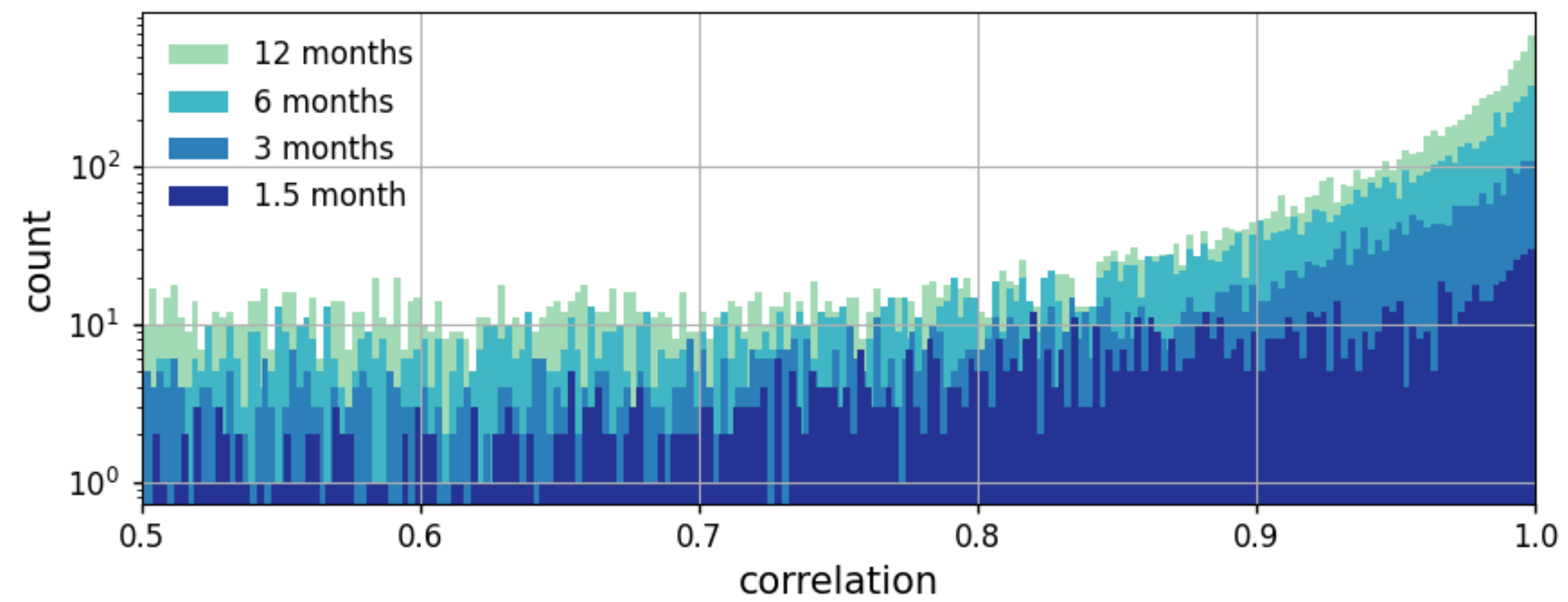
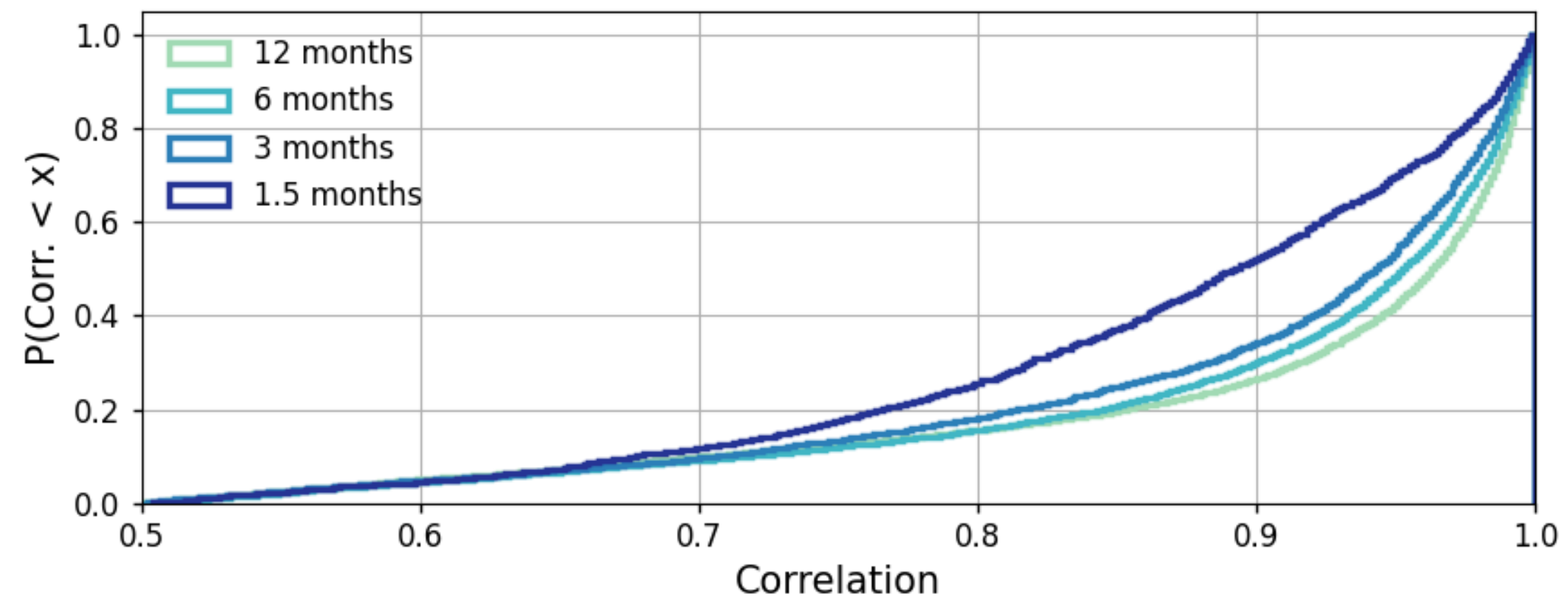
LDC - Challenge 1: *Radler*



Observation time	sky localisation 10 sq. deg.	$\dot{f} > 0$	$\dot{f} < 0$	eclipsing	eclipsing with $\dot{f} > 0$	eclipsing and well-localised
1.5	2	279	5	91	16	0
3.0	3	553	1	372	48	0
6.0	25	1950	10	1027	275	4
12.0	441	4239	20	1873	735	58

LDC - Challenge 1: *Radler*

Observation time	percent unambiguous	percent confused	percent false alarm
1.5	84	16	0.12
3.0	88	12	0
6.0	89	11	0.07
12.0	88	12	0.09



LDC - Challenge 1: *Radler*

06 mo

12 mo

24 mo

LDC00521426070

LDC00521426233

LDC00521426302

LDC00520996521

LDC00520996456

LDC00520996812

LDC00521130528

LDC00521130989

LDC00521131694

LDC00521496198

LDC00521499036

LDC00521501630

LDC00521437777

LDC00521440627

LDC00521440337

LDC00520862145

LDC00520863596

LDC00520863048

LDC00521031121

LDC00521029147

LDC00521030958

LDC00521323156

LDC00521322257

LDC00521323001

LDC00521567820

LDC00521568593

LDC00521568503

LDC00521631910

LDC00521631036

LDC00521631952

LDC00520931391, S/R = 14.8132, Match1 = 0.982422, Match2 = --

LDC00520925144, S/R = 18.7571, Match1 = 0.384154, Match2 = 0.413555

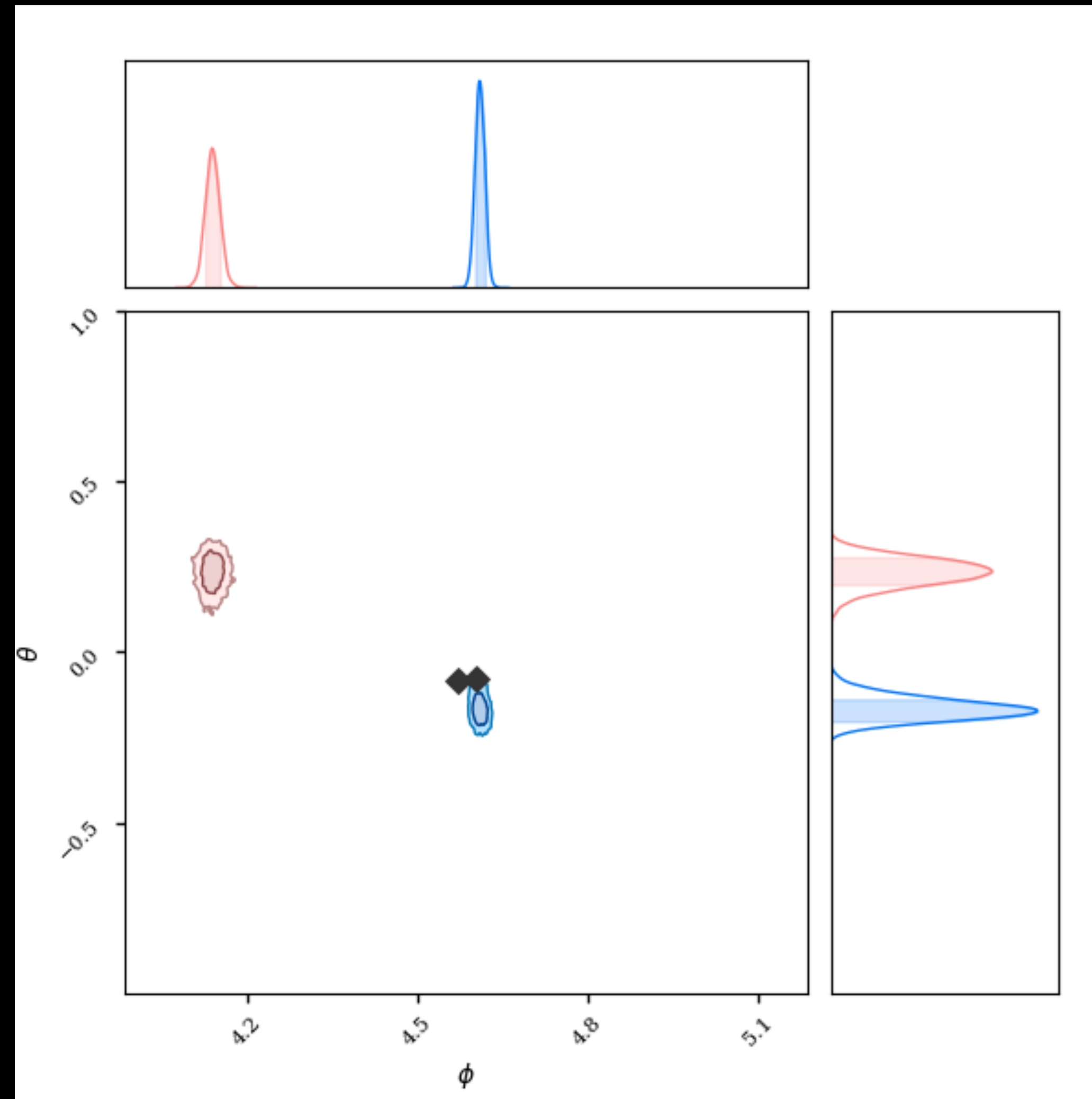
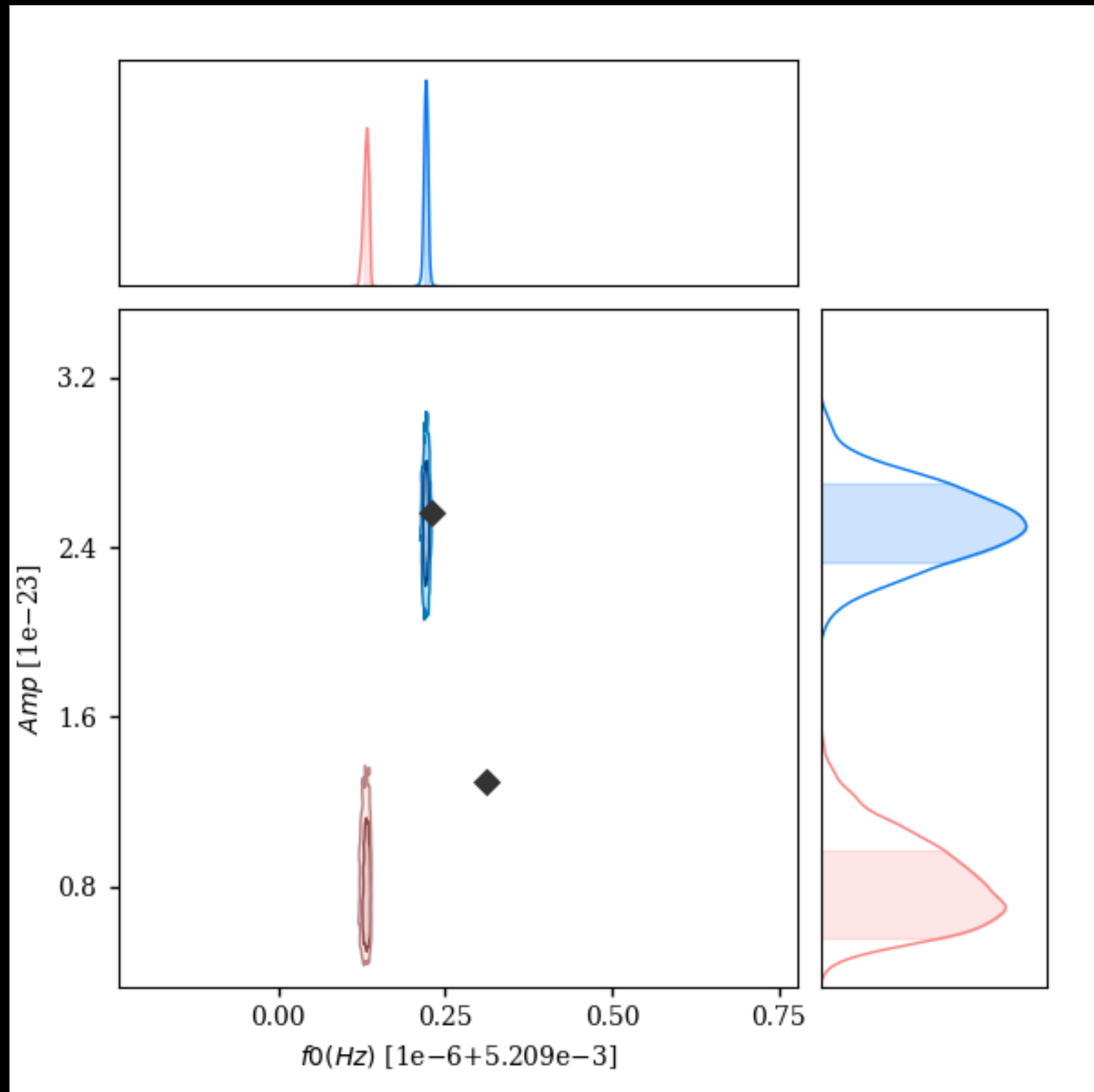
LDC00520922189, S/R = 28.5803, Match2 = 0.924785, Match1 = --

LDC00520927782, S/R = 17.1364, Match1 = 0.440143, Match2 = 0.144022

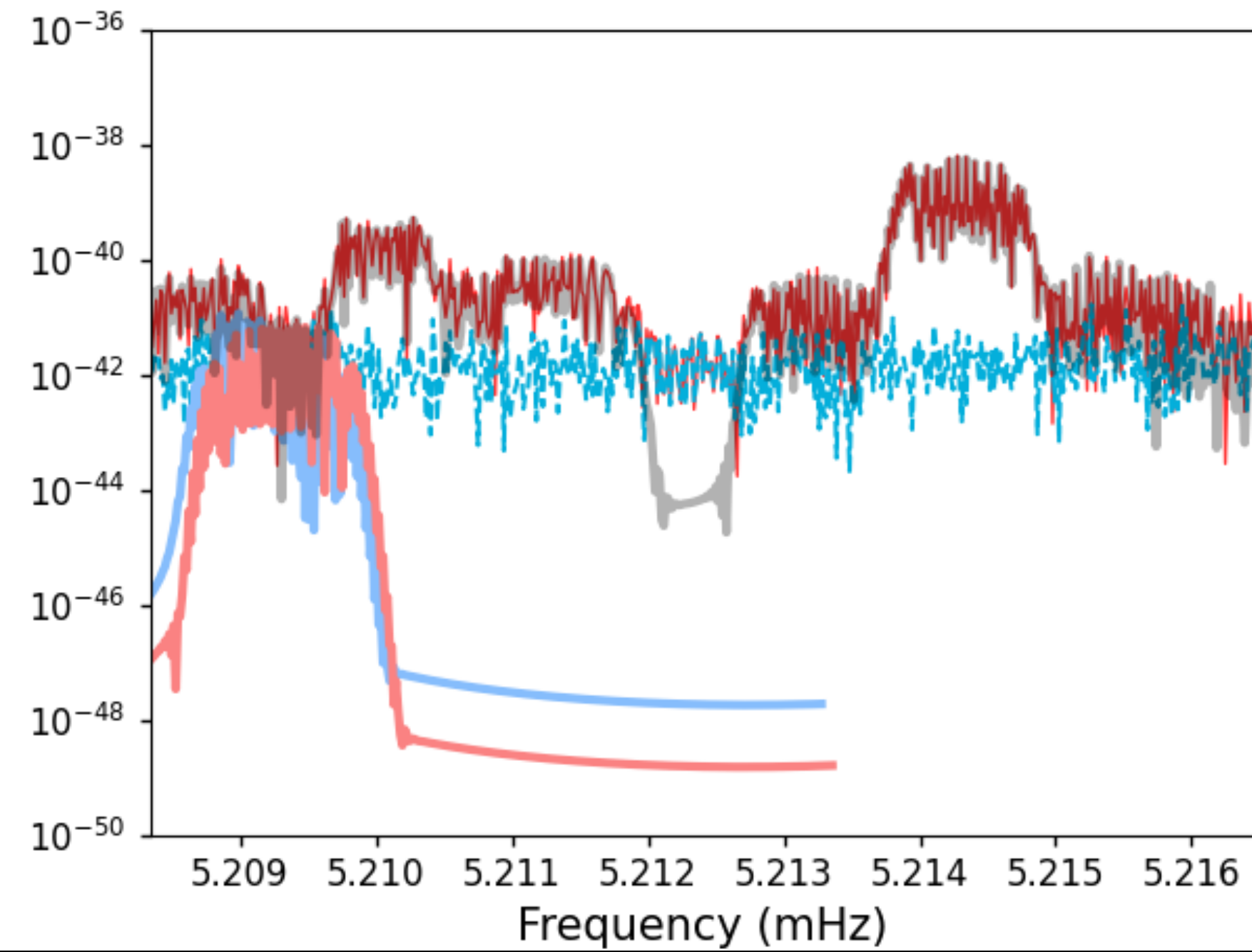
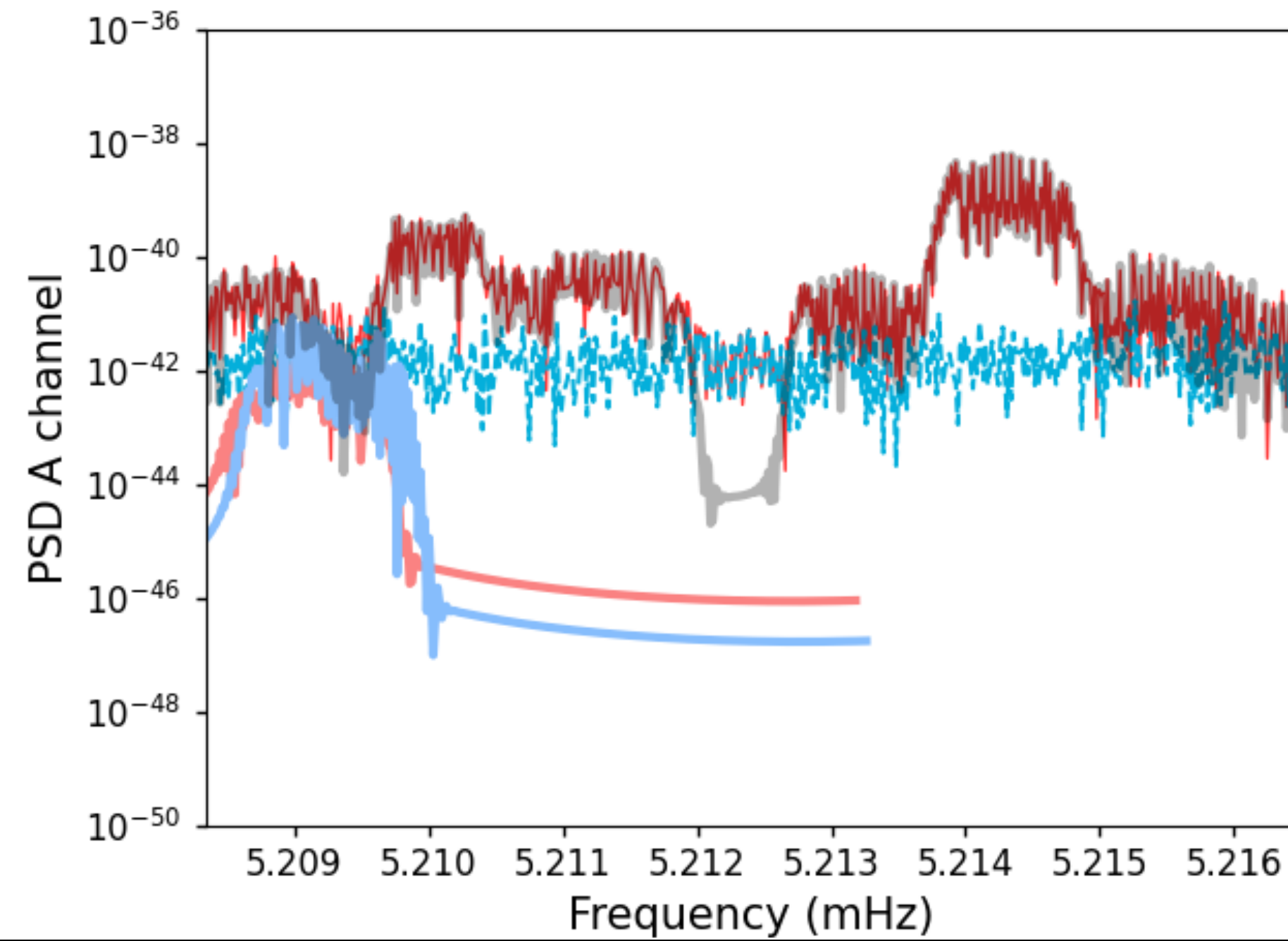
LDC00520913250, S/R = 17.2258, Match1 = 0.413048, Match2 = --

LDC00521646740

LDC - Challenge 1: *Radler*



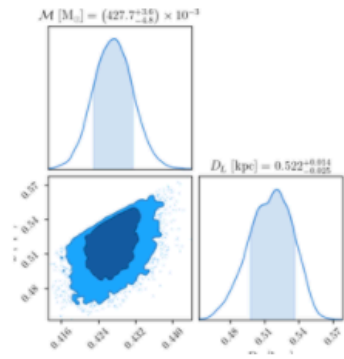
LDC - Challenge 1: *Radler*



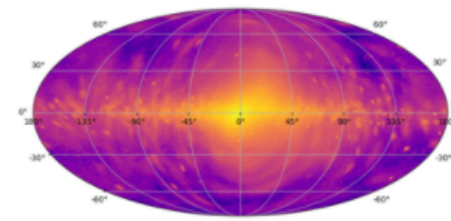
UCB Use Case Gallery

lisacattools

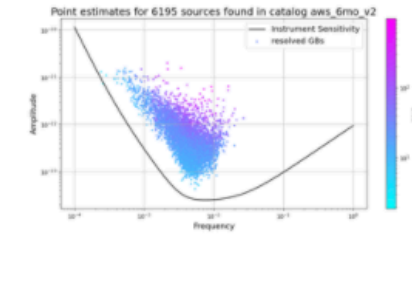
Below are examples using UCB catalogs. Catalog data products are formatted as `pandas` data frames and stored in `HDF5` files. Top level catalog files have the list of all candidate detections, point estimates, etc. In addition, posterior samples for each candidate are stored as separate data frames and grouped together by frequency segments. There are different catalog products for different LISA observing times.



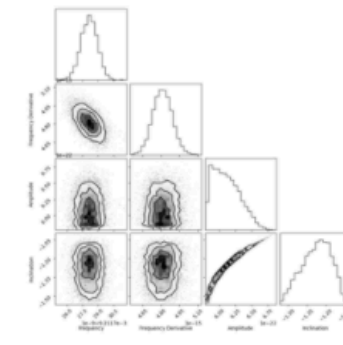
Resampling posteriors



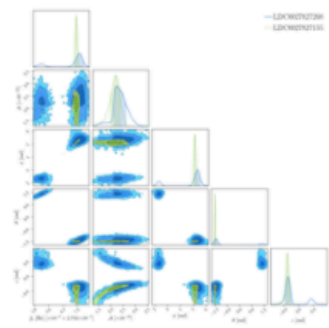
Joint PDF of sky location



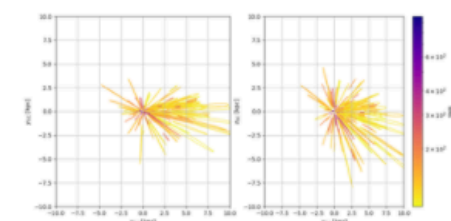
Scatter plots



Corner plots



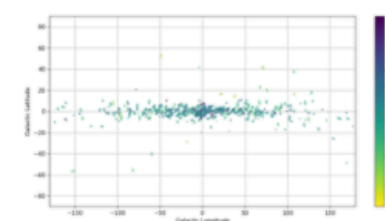
Connecting catalogs



3D map of the galaxy



Parameter tables



Sky Localization Ellipses

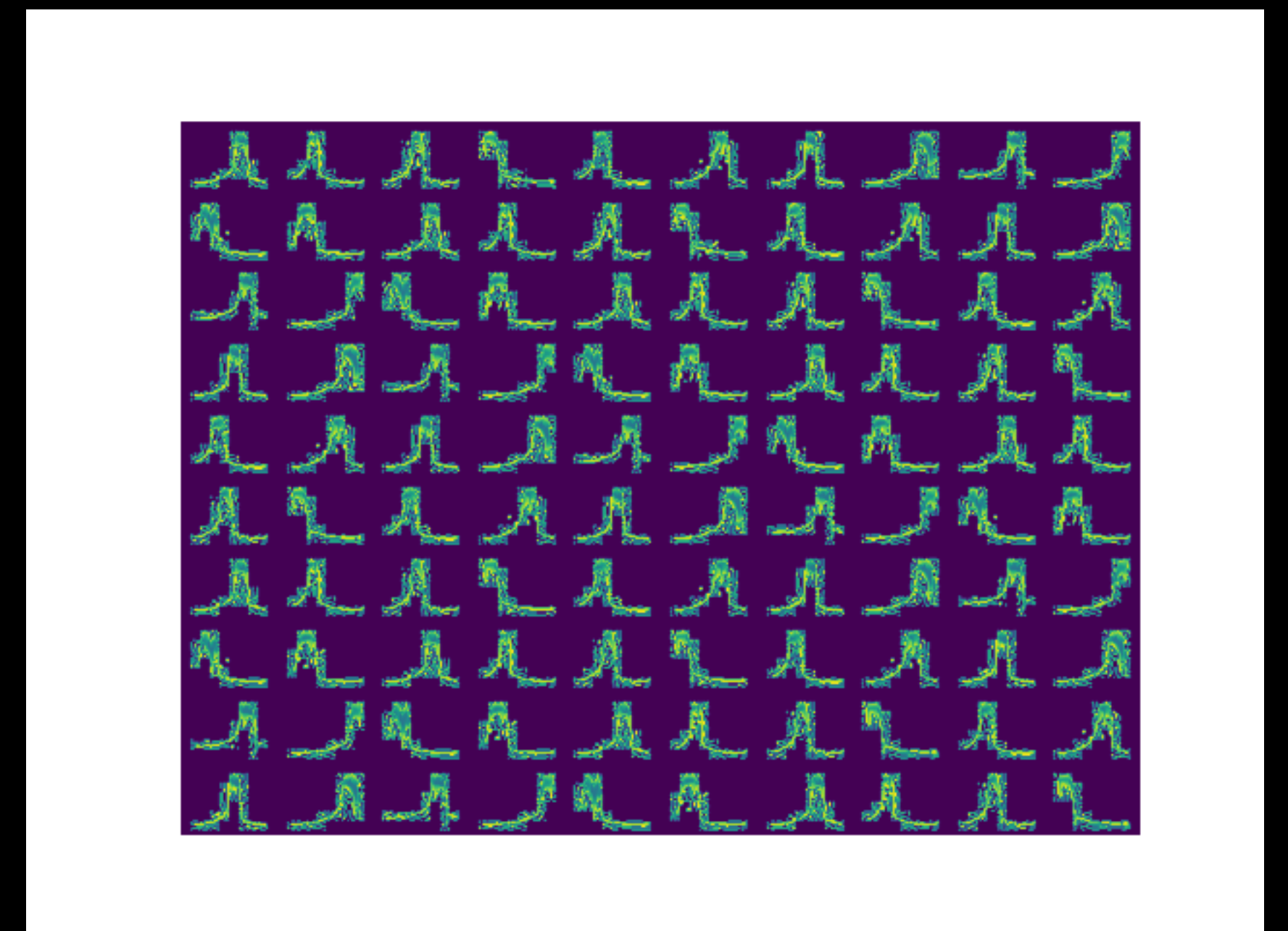
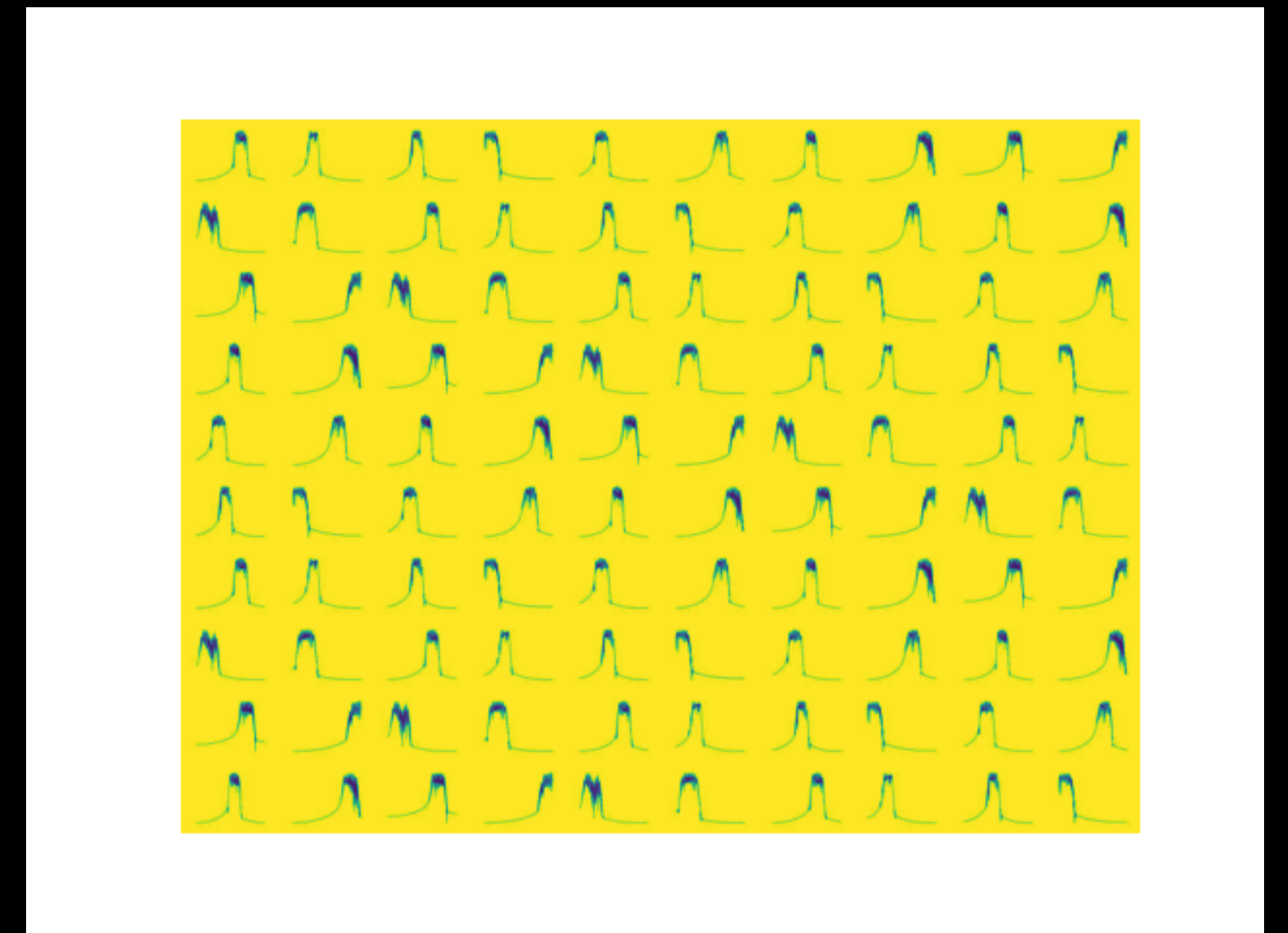
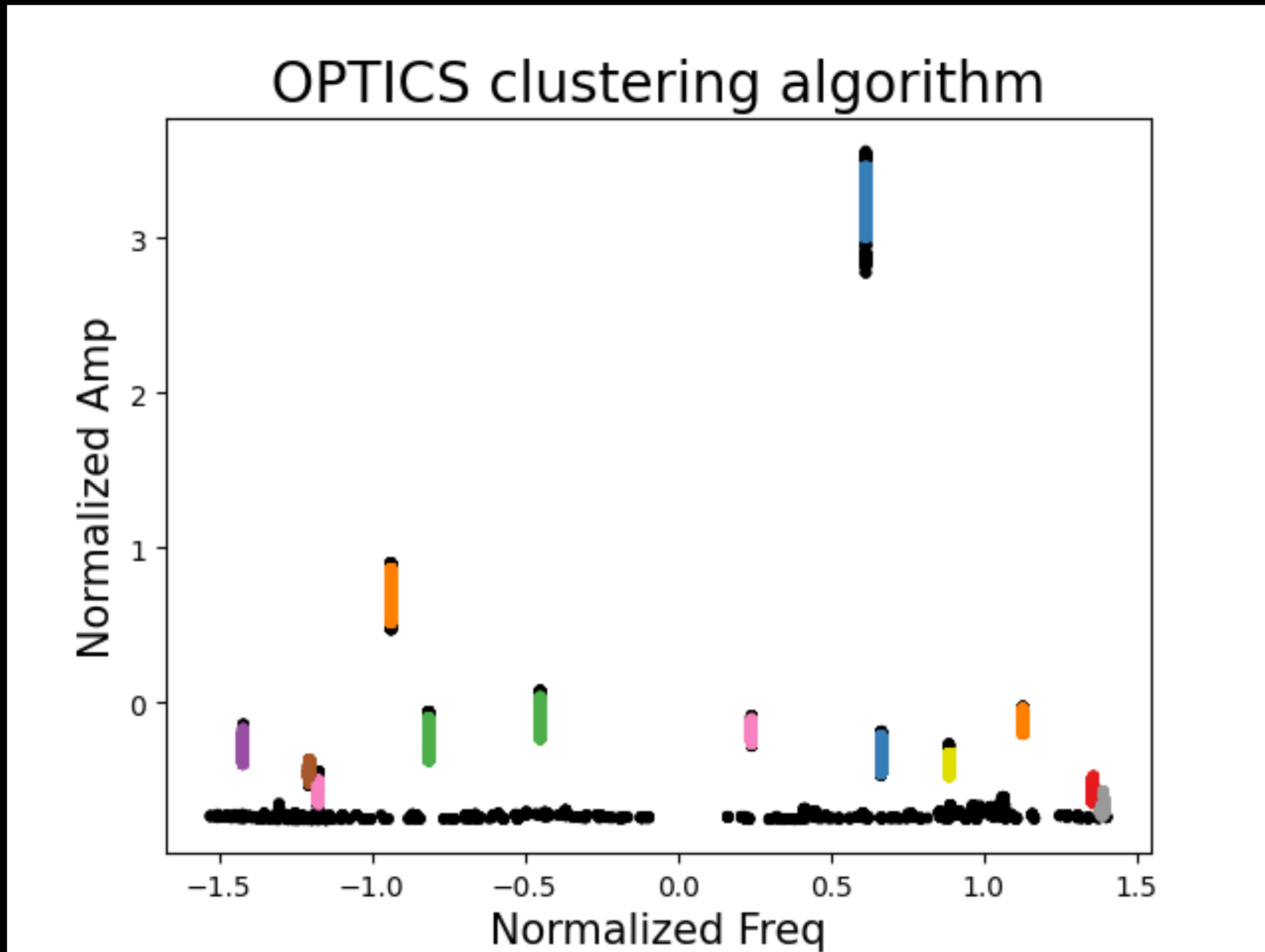
Download all examples in Python source code: [examples_ucb_python.zip](#)

Download all examples in Jupyter notebooks: [examples_ucb_jupyter.zip](#)

For the MCMC code see <https://github.com/tlittenberg/lidasoft.git>



Building catalogs with ML
L2 -> L3



Merci



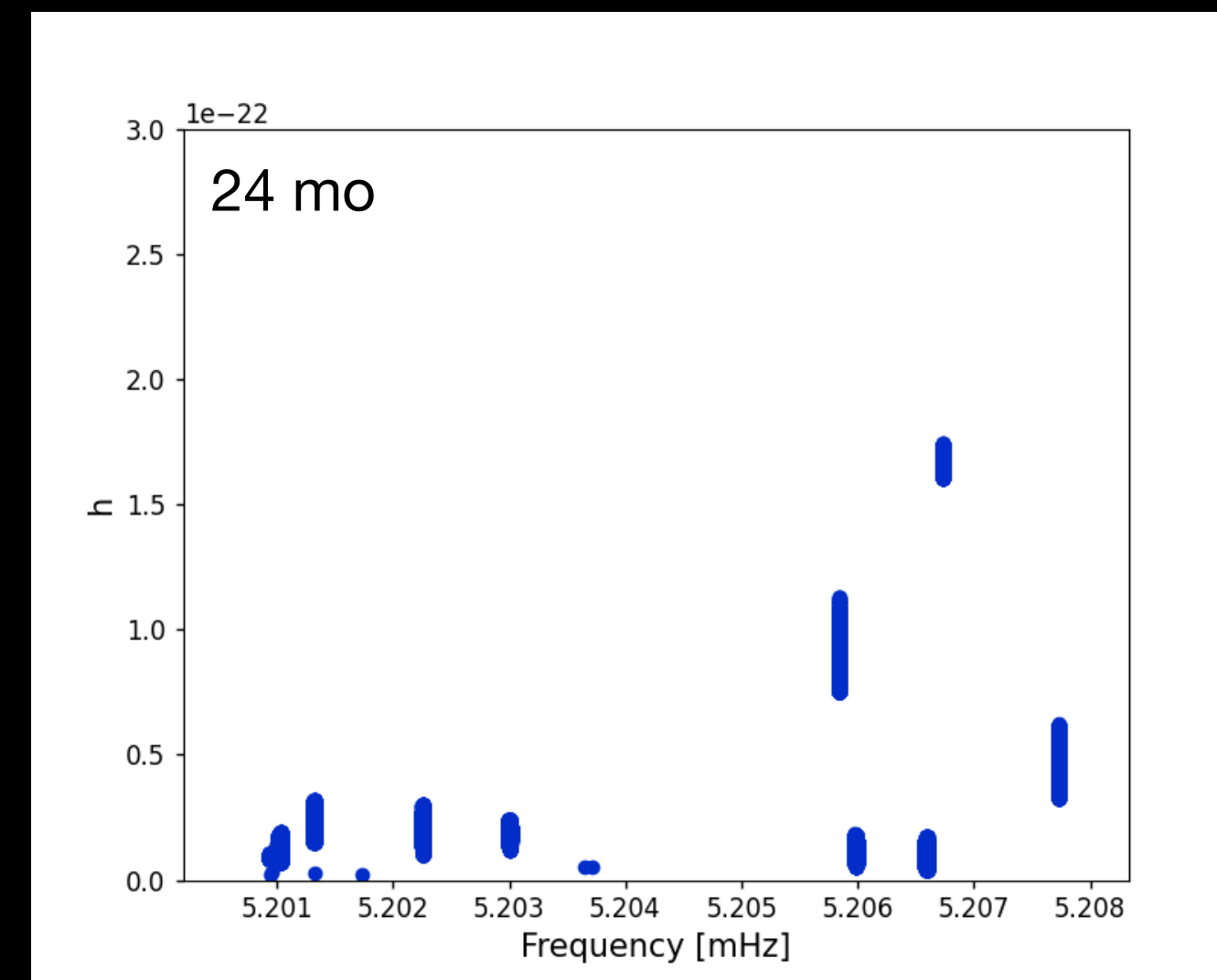
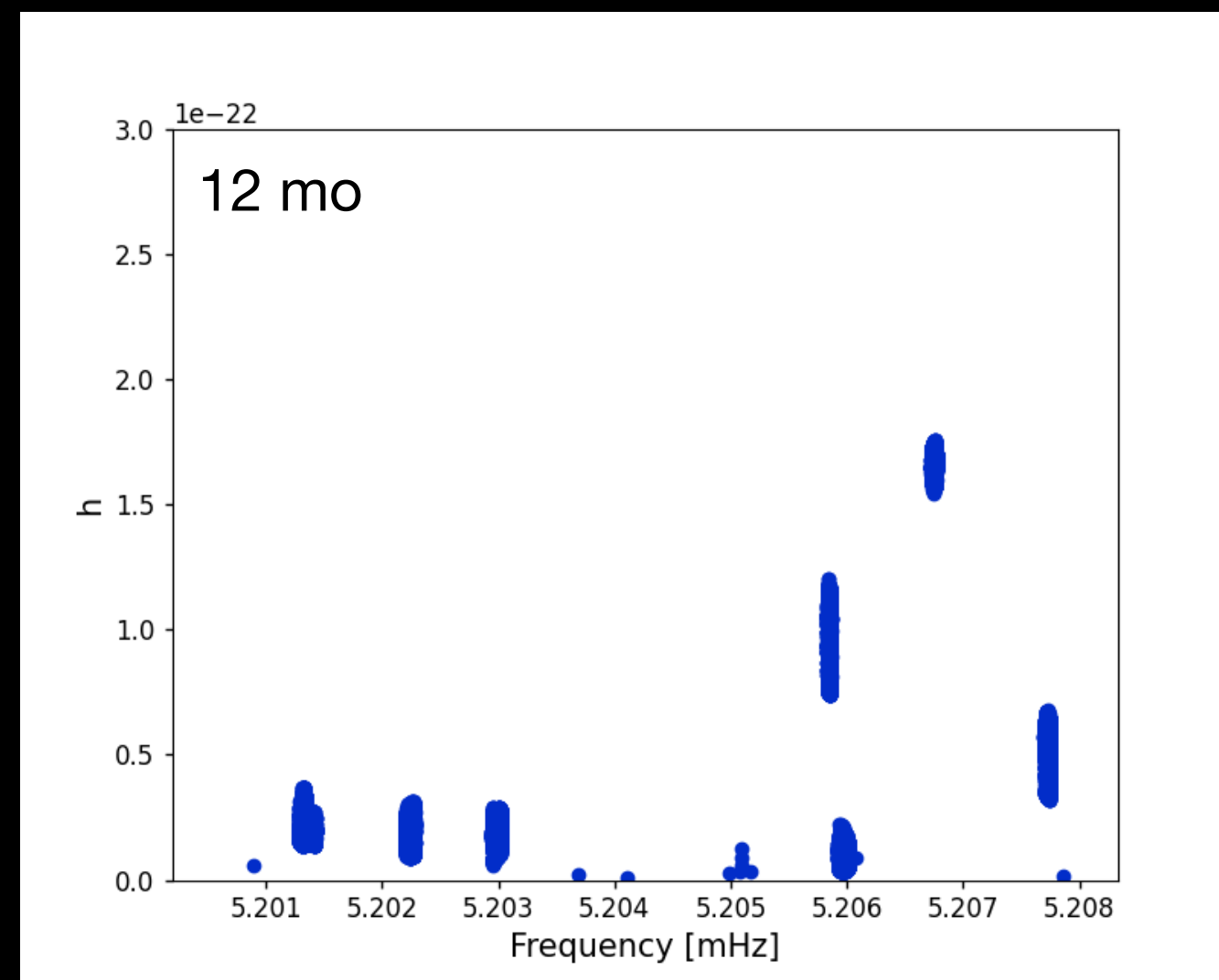
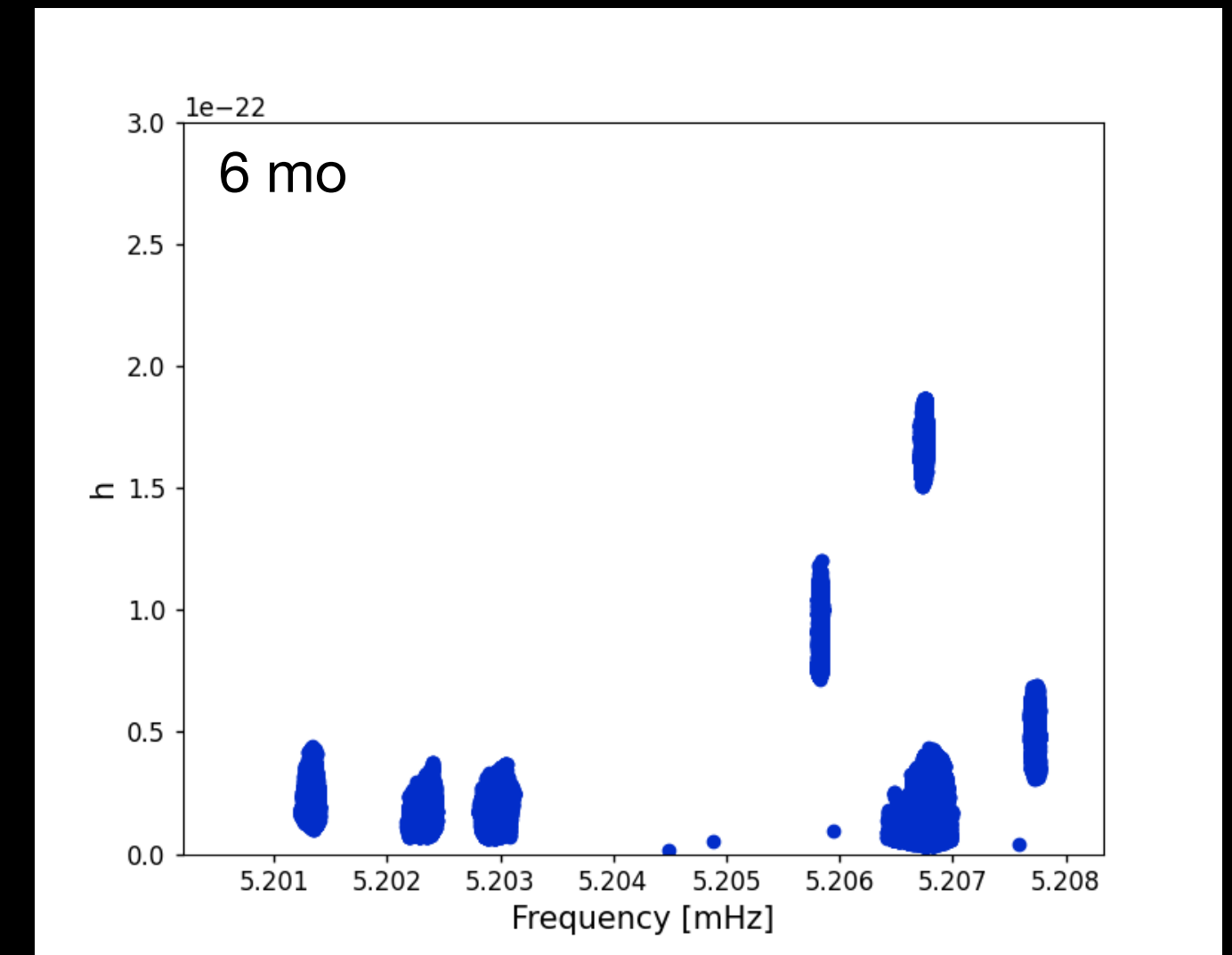
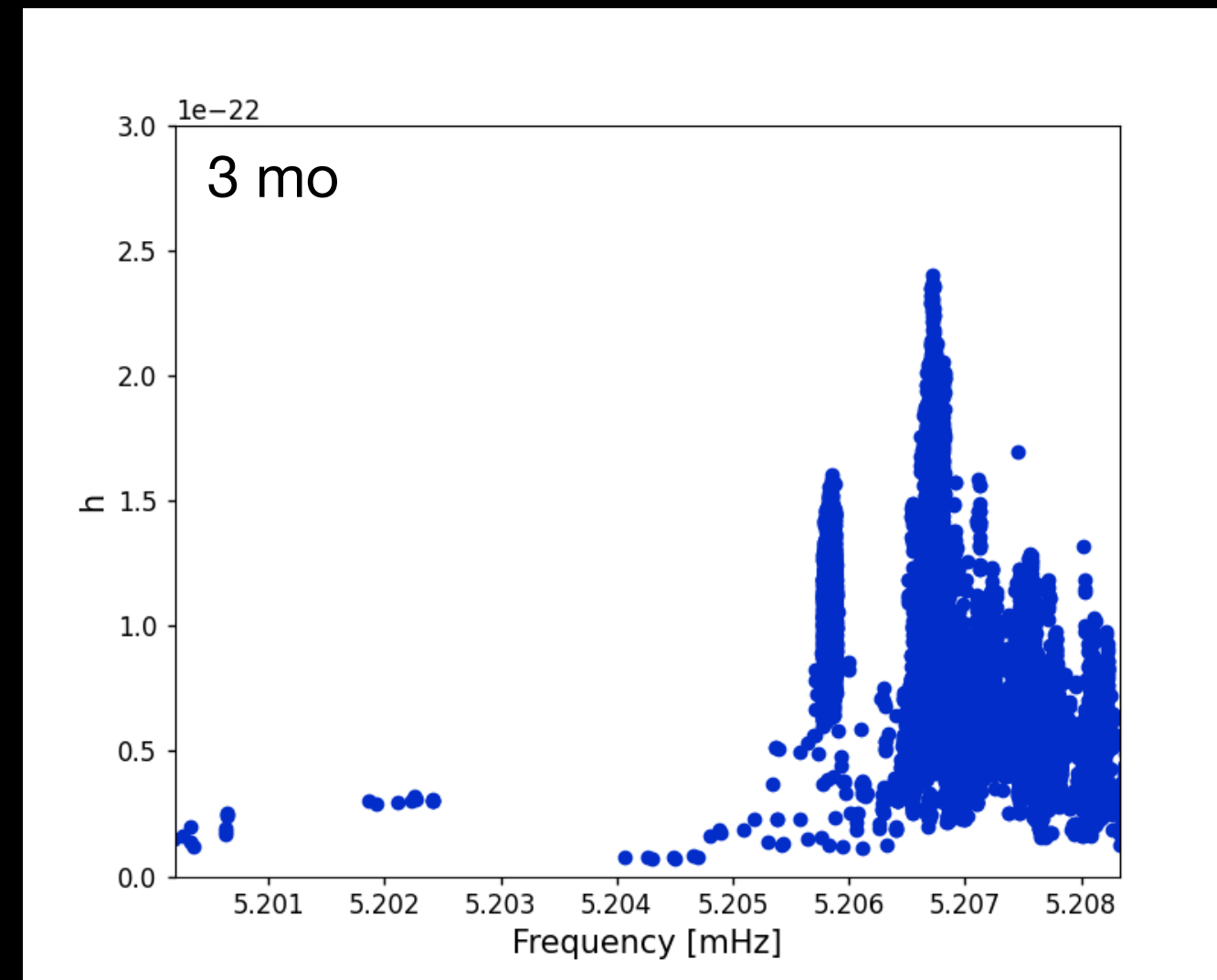
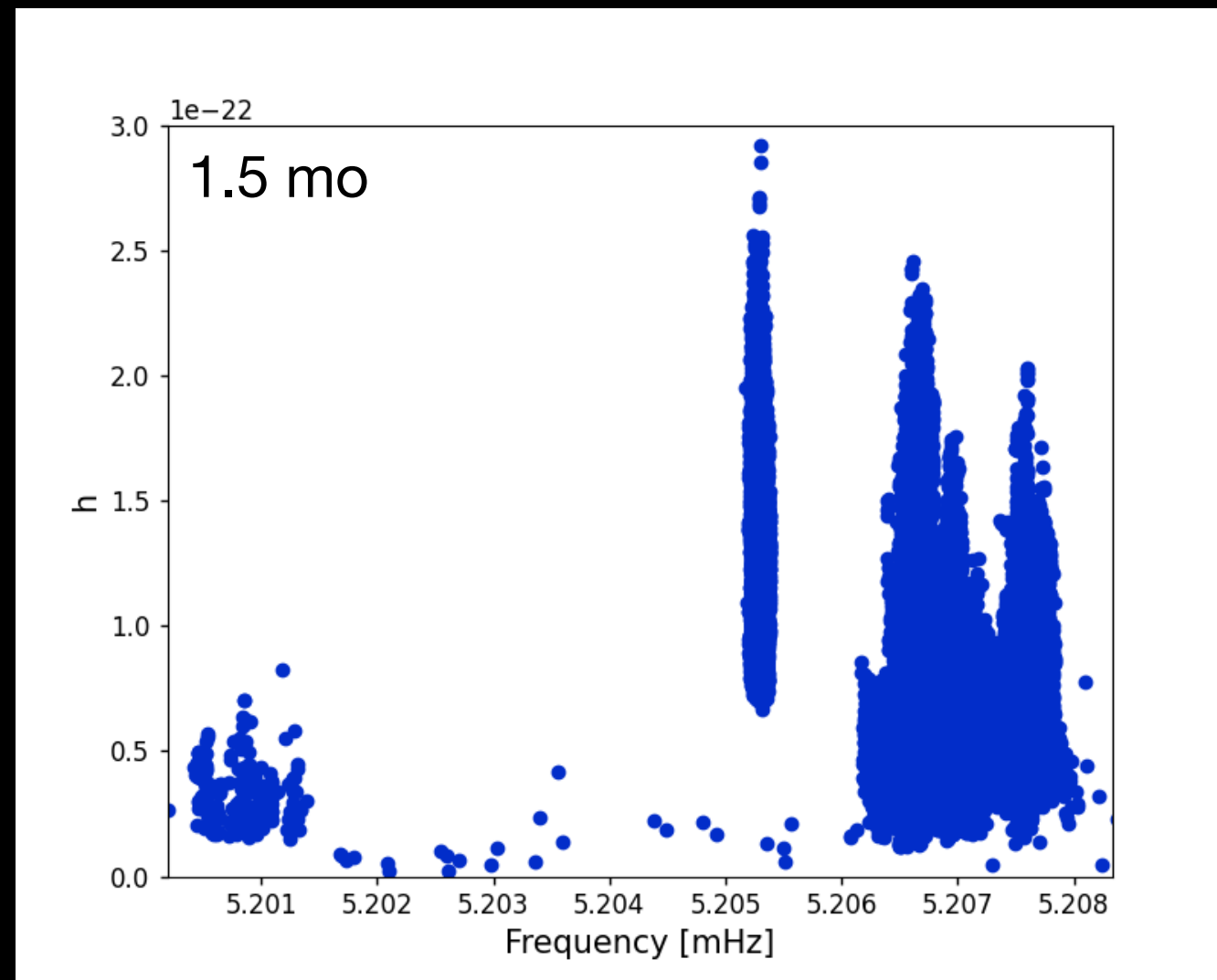
Parametrisation of the problem

Parameterization of Galactic binary signal

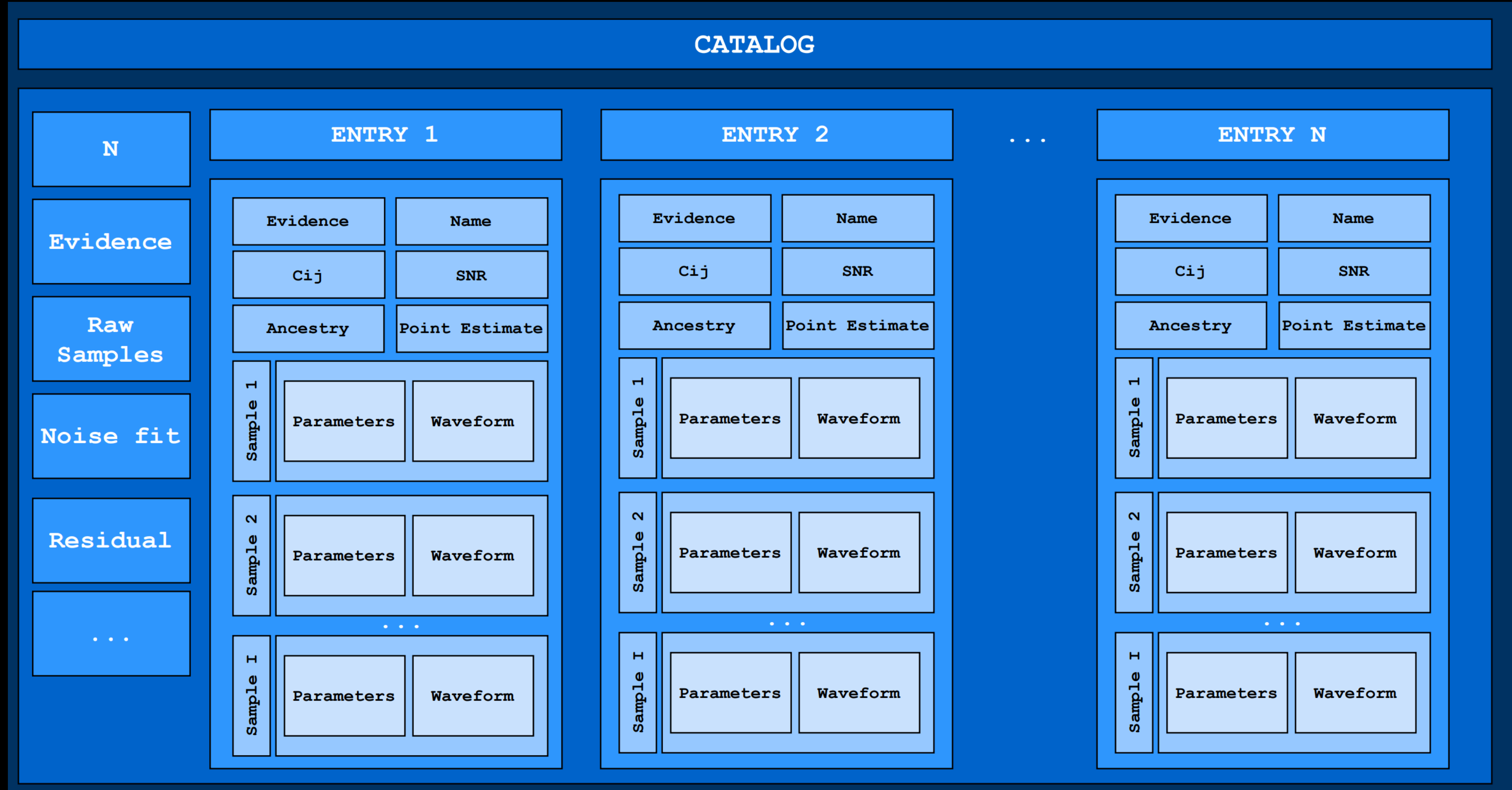
Parameter	Notation	units
ϑ	EclipticLatitude	Radian
φ	EclipticLongitude	Radian
\mathcal{A}	Amplitude	strain
f_0	Frequency	Hz
\dot{f}_0	FrequencyDerivative	Hz^2
ι	Inclination	Radian
ψ	Polarization	Radian
ϕ_0	InitialPhase	Radian
T_{obs}	ObservationDuration	Seconds
Δt	Cadence	Seconds

LDC - Challenge 1: *Radler*

Raw samples

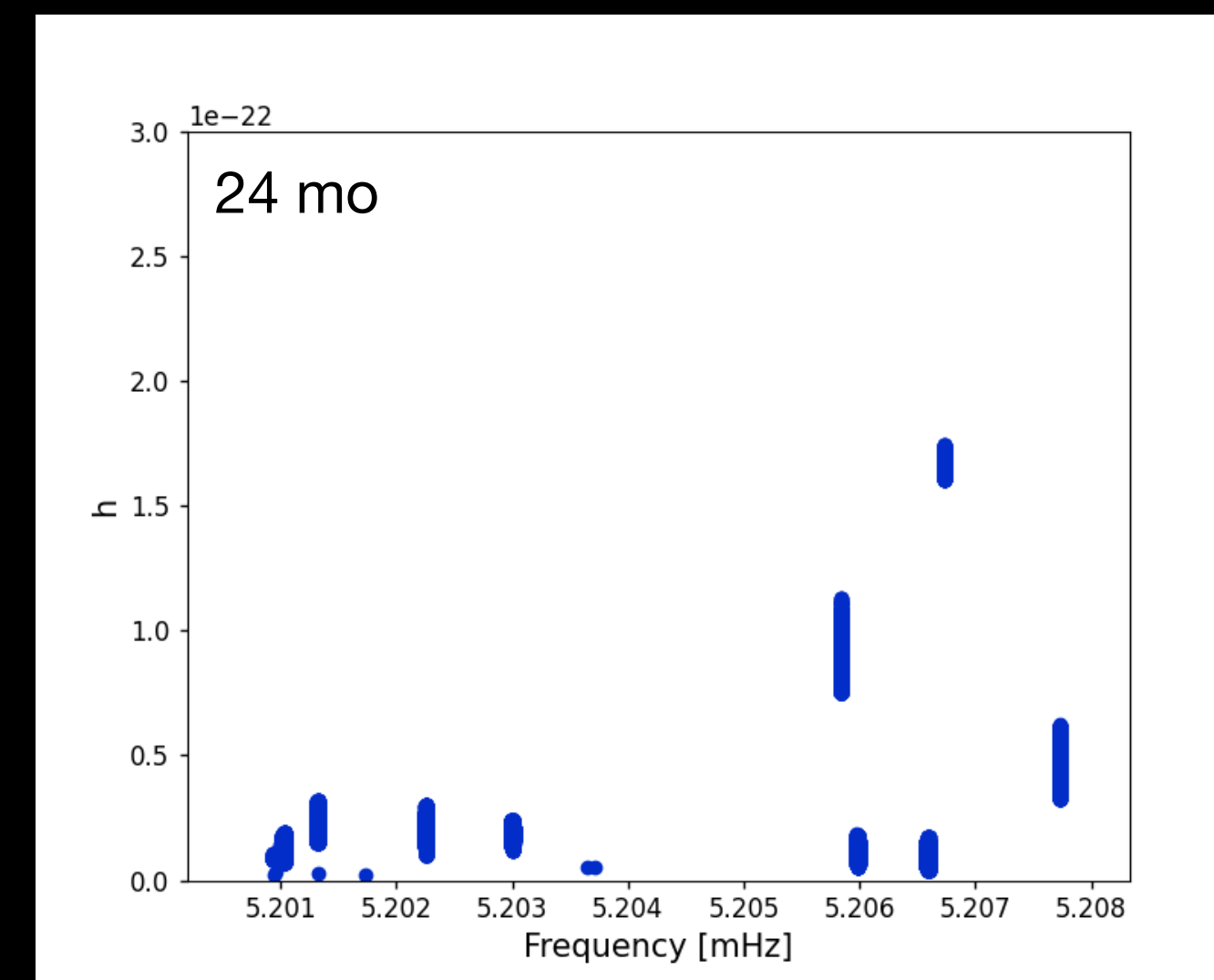
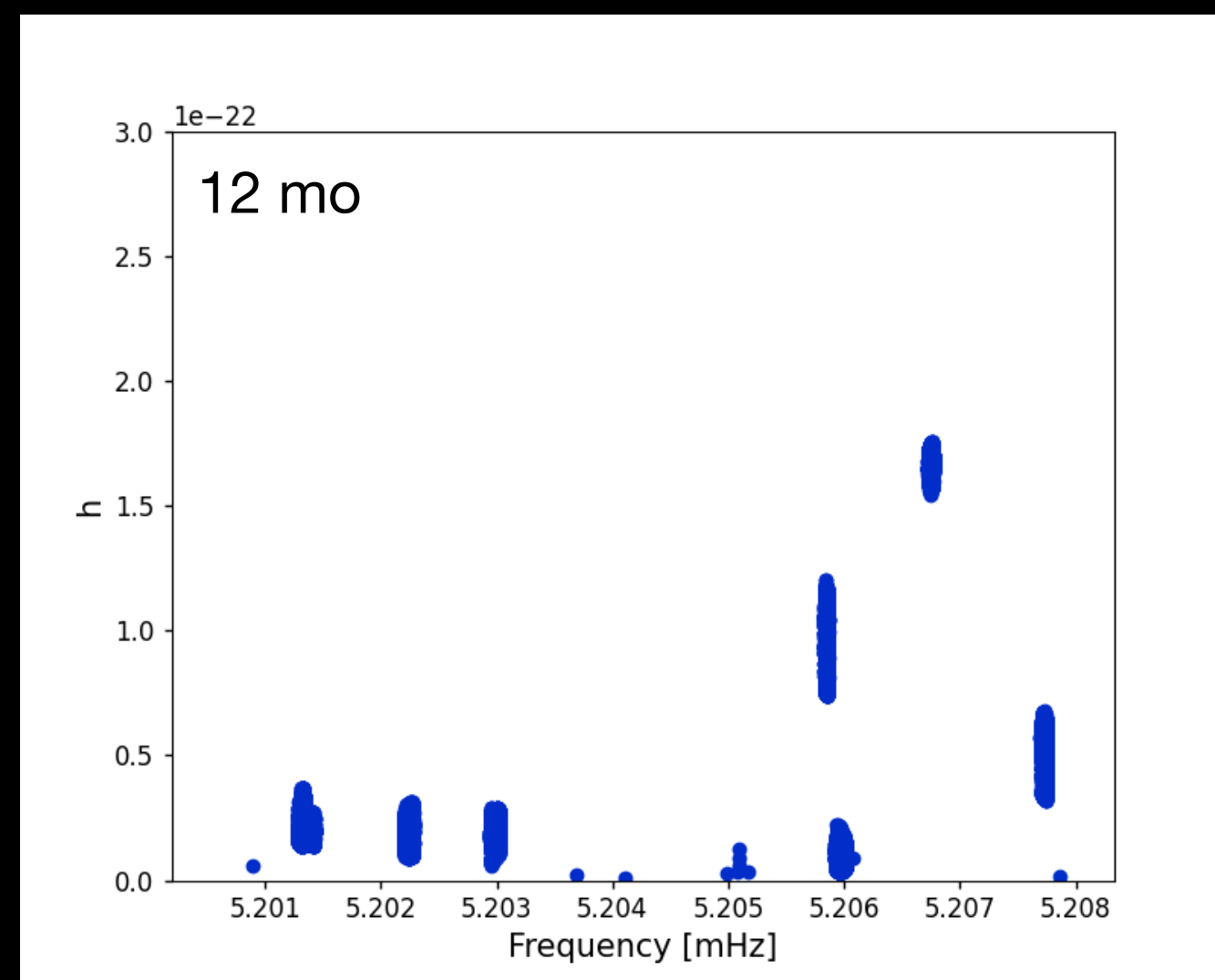
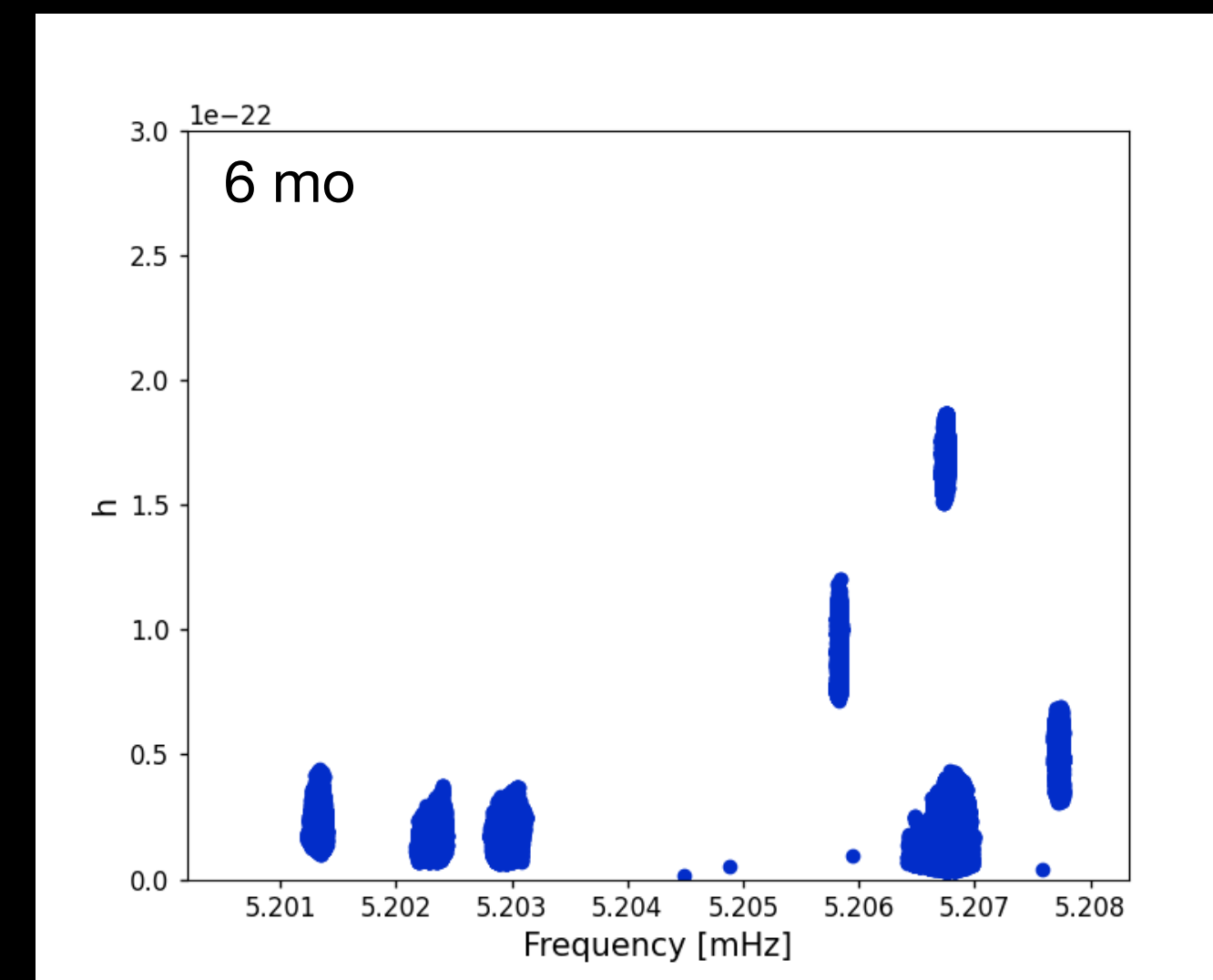
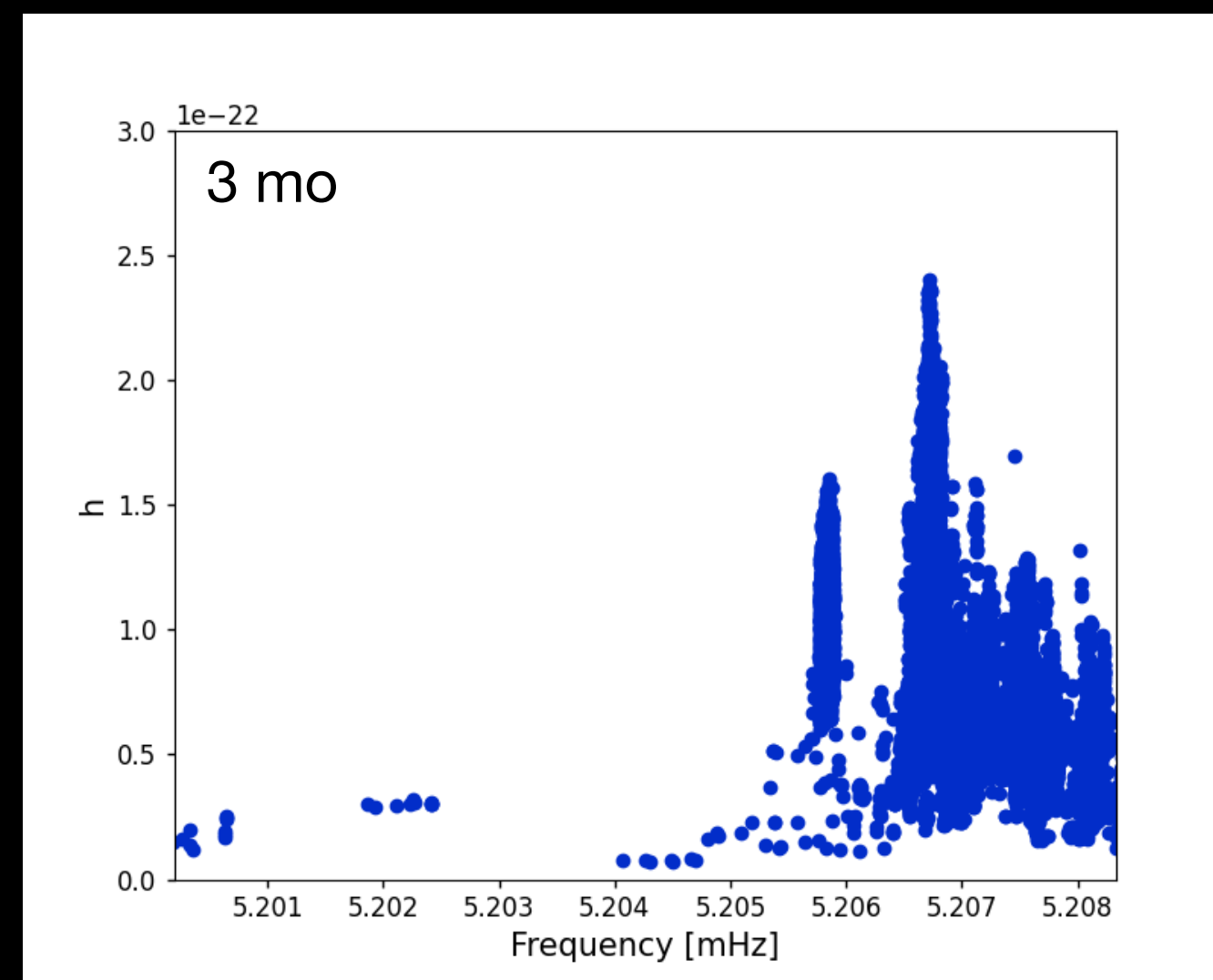
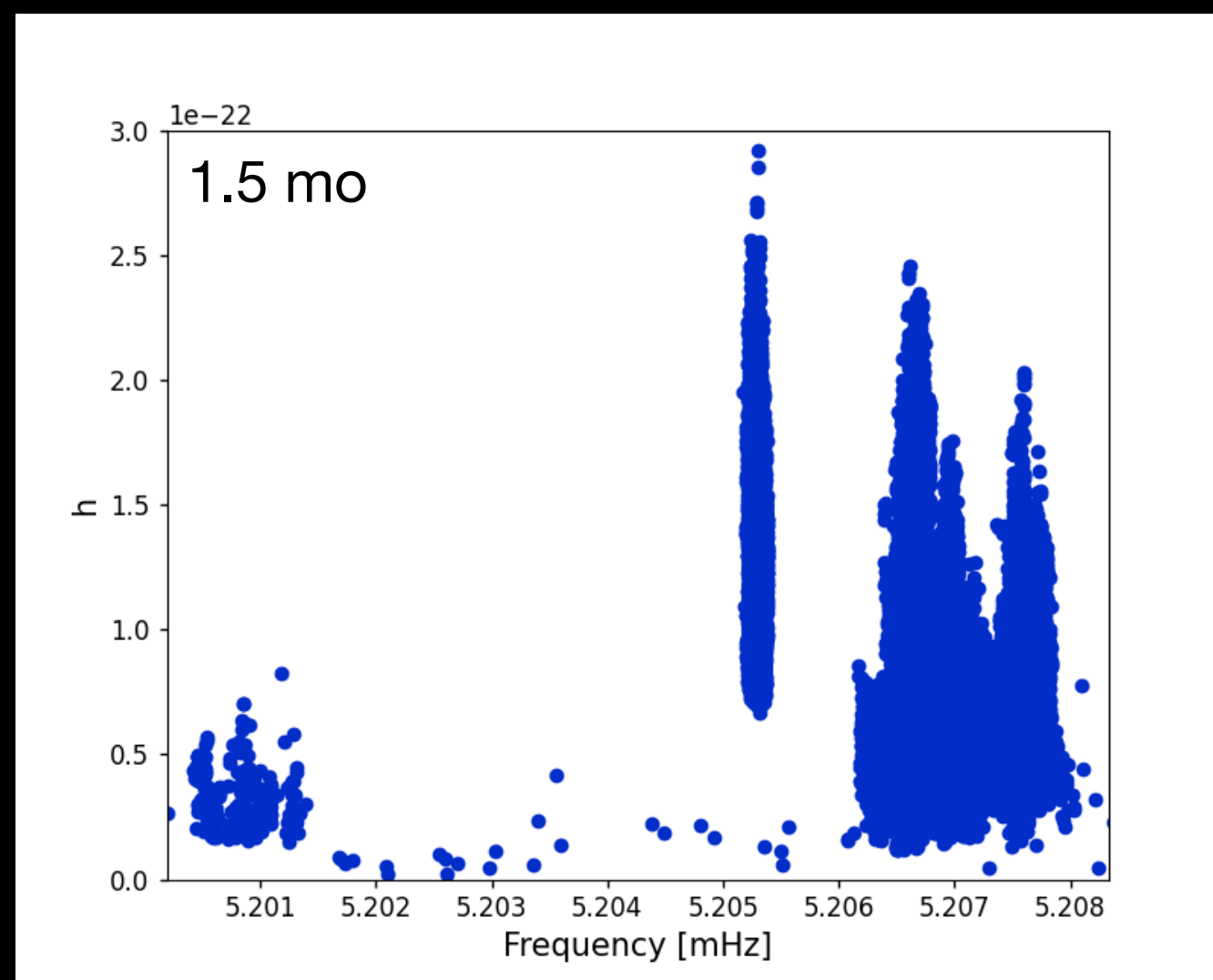


Catalog production



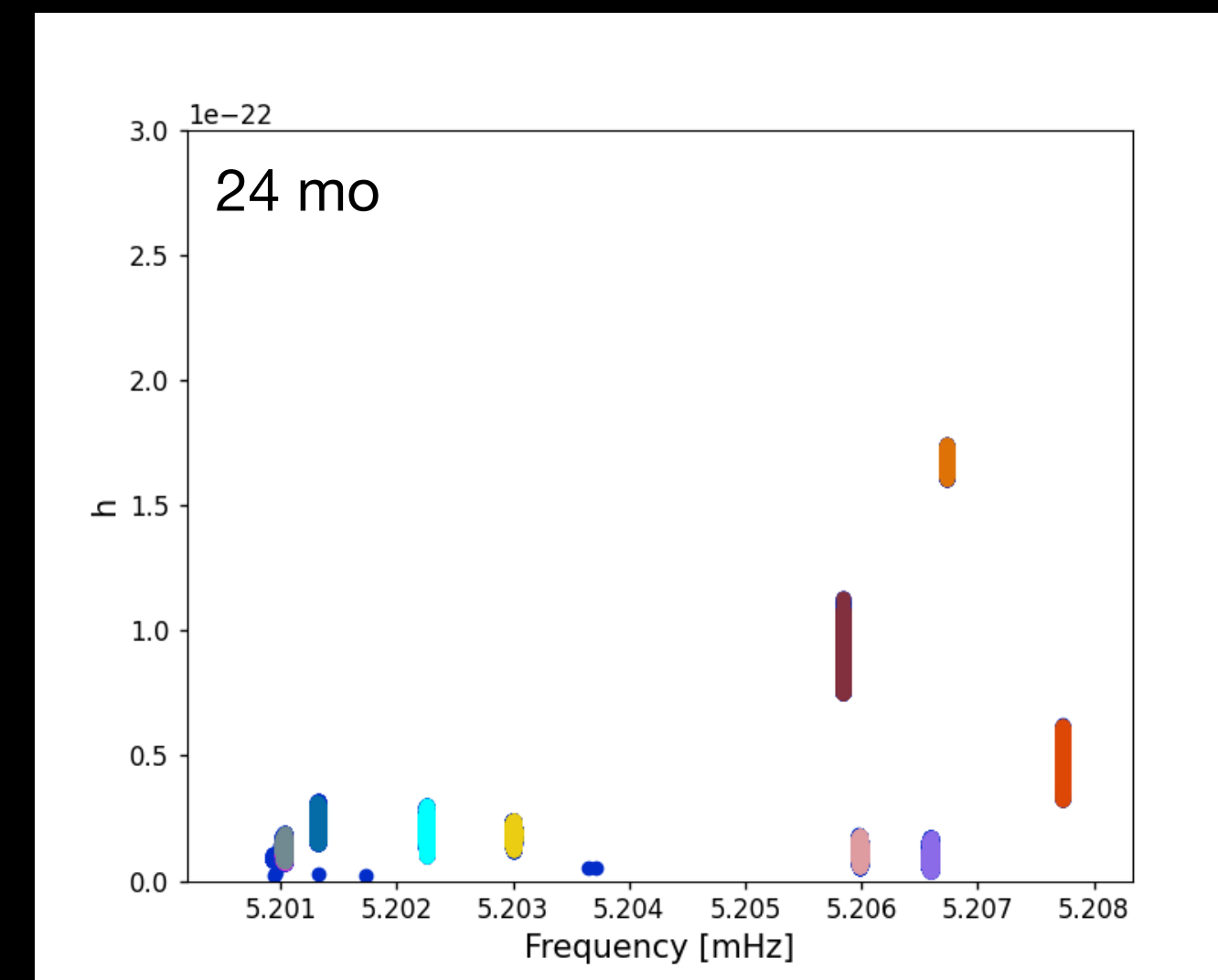
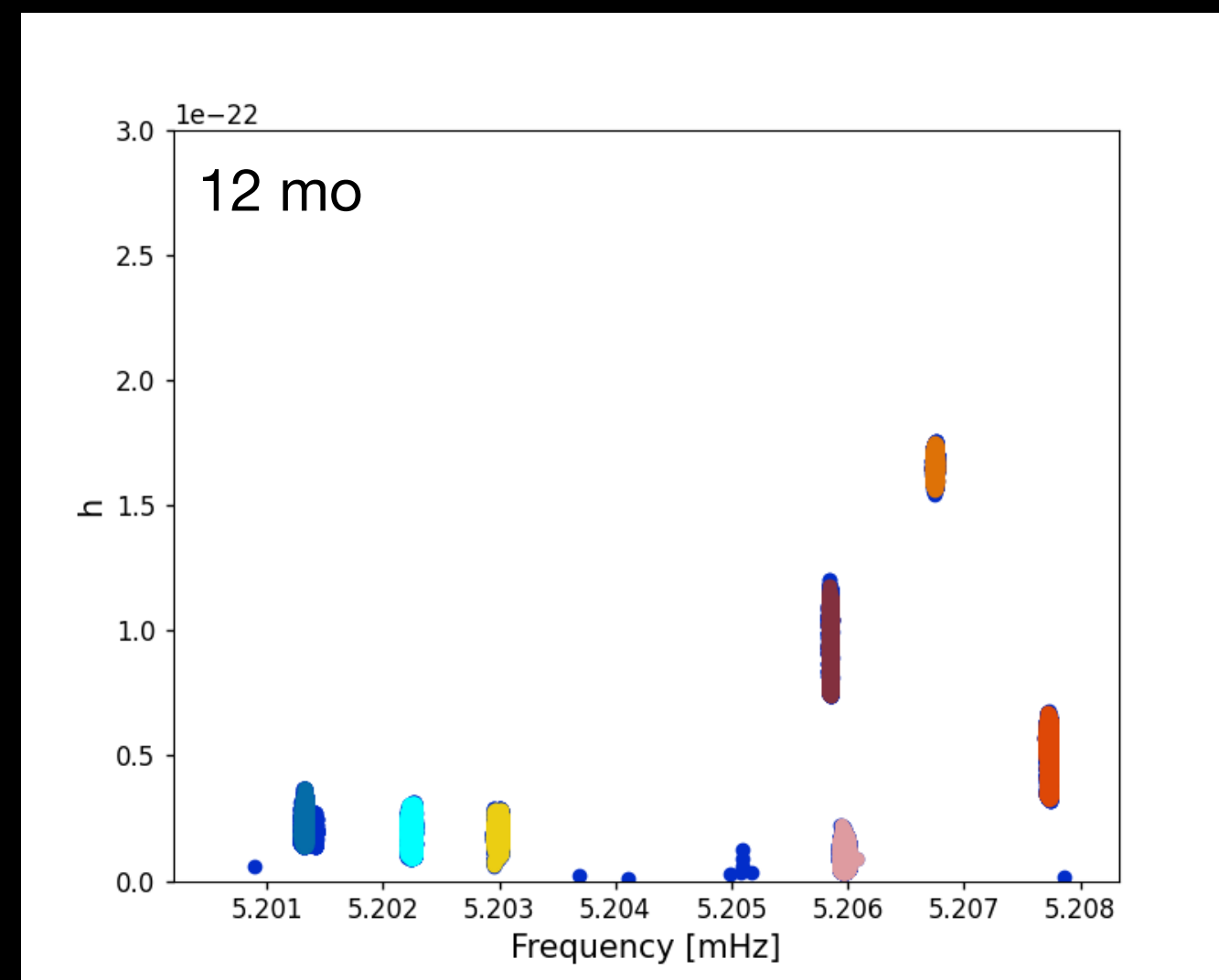
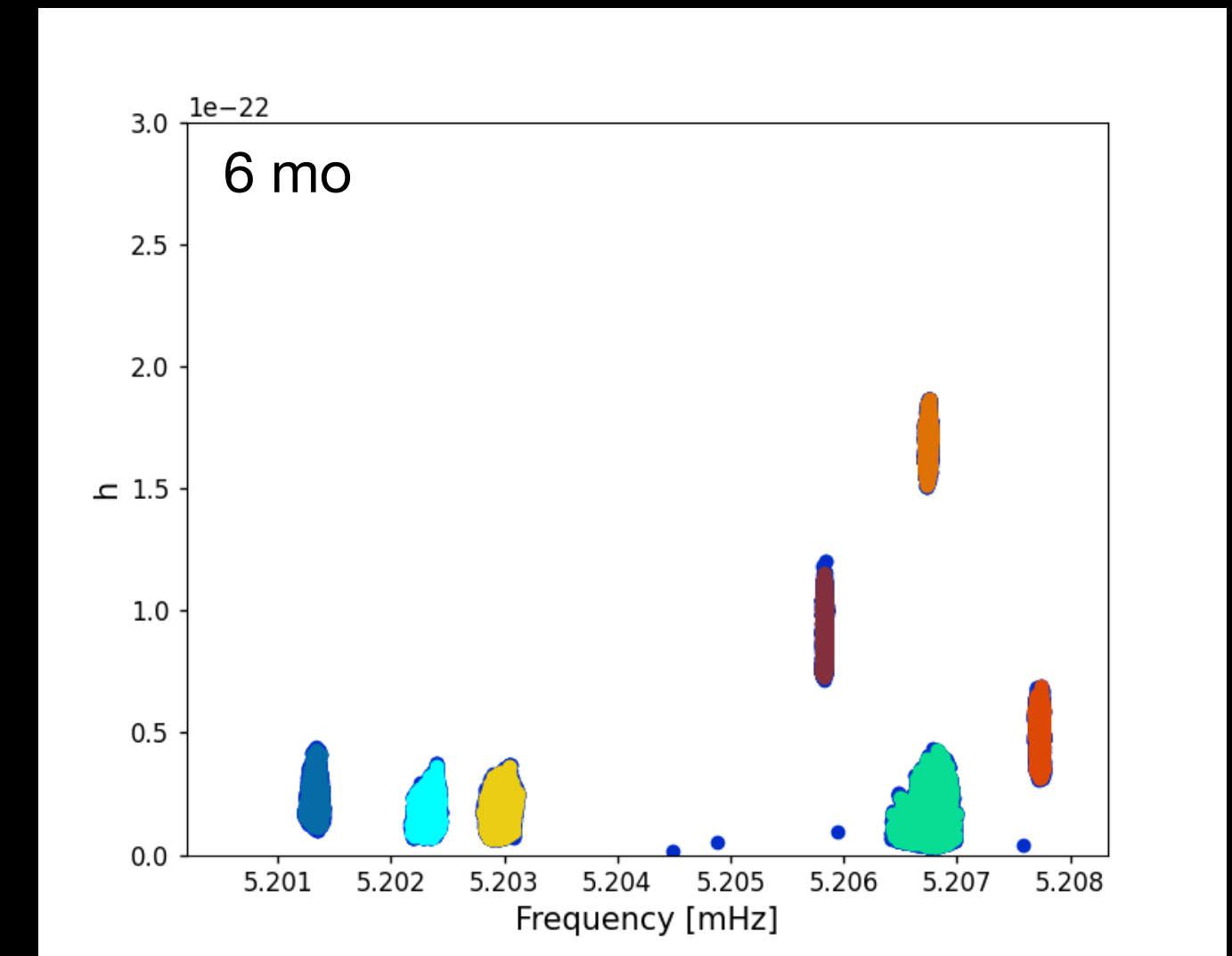
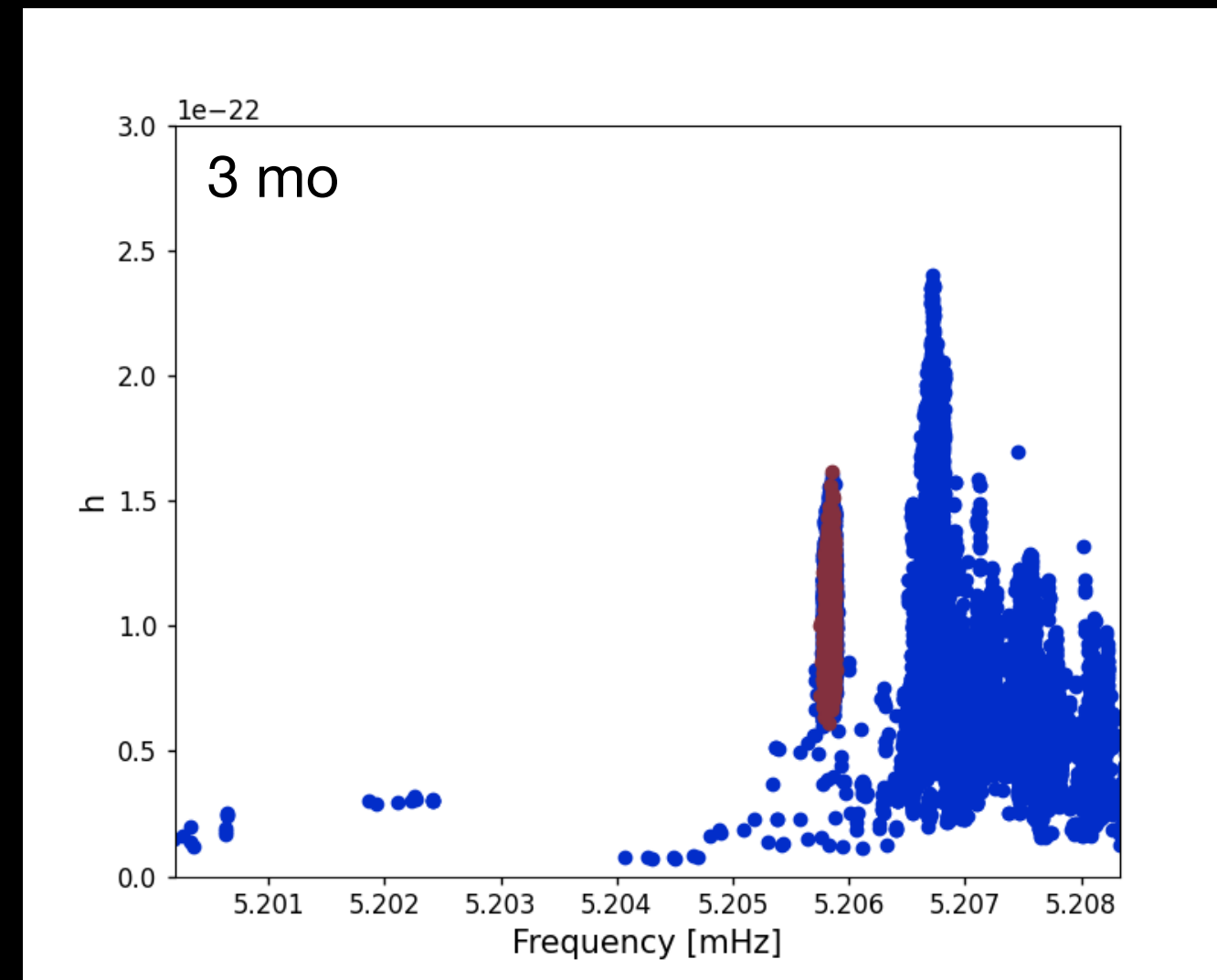
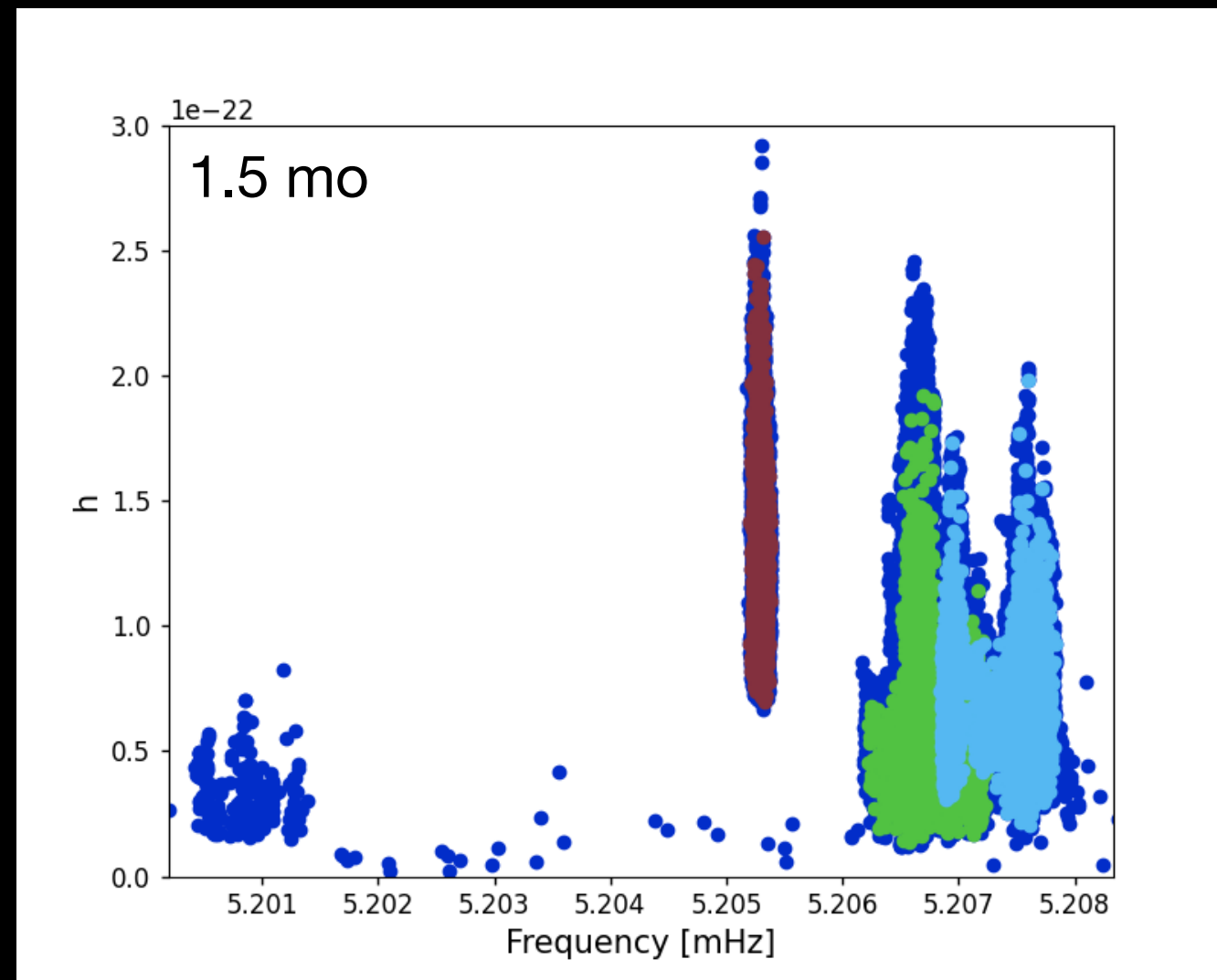
LDC - Challenge 1: *Radler*

Raw samples



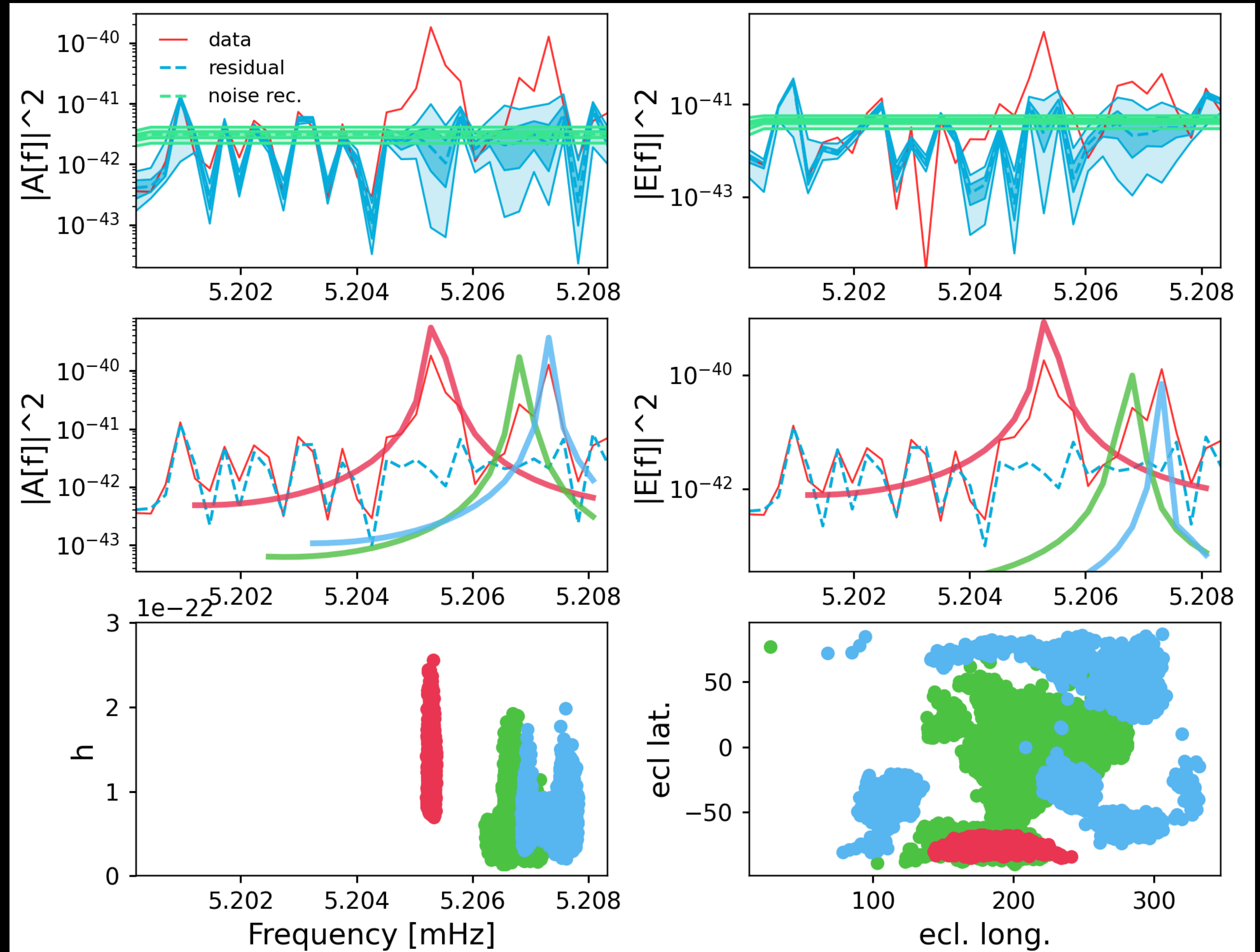
LDC - Challenge 1: *Radler*

Catalog detections



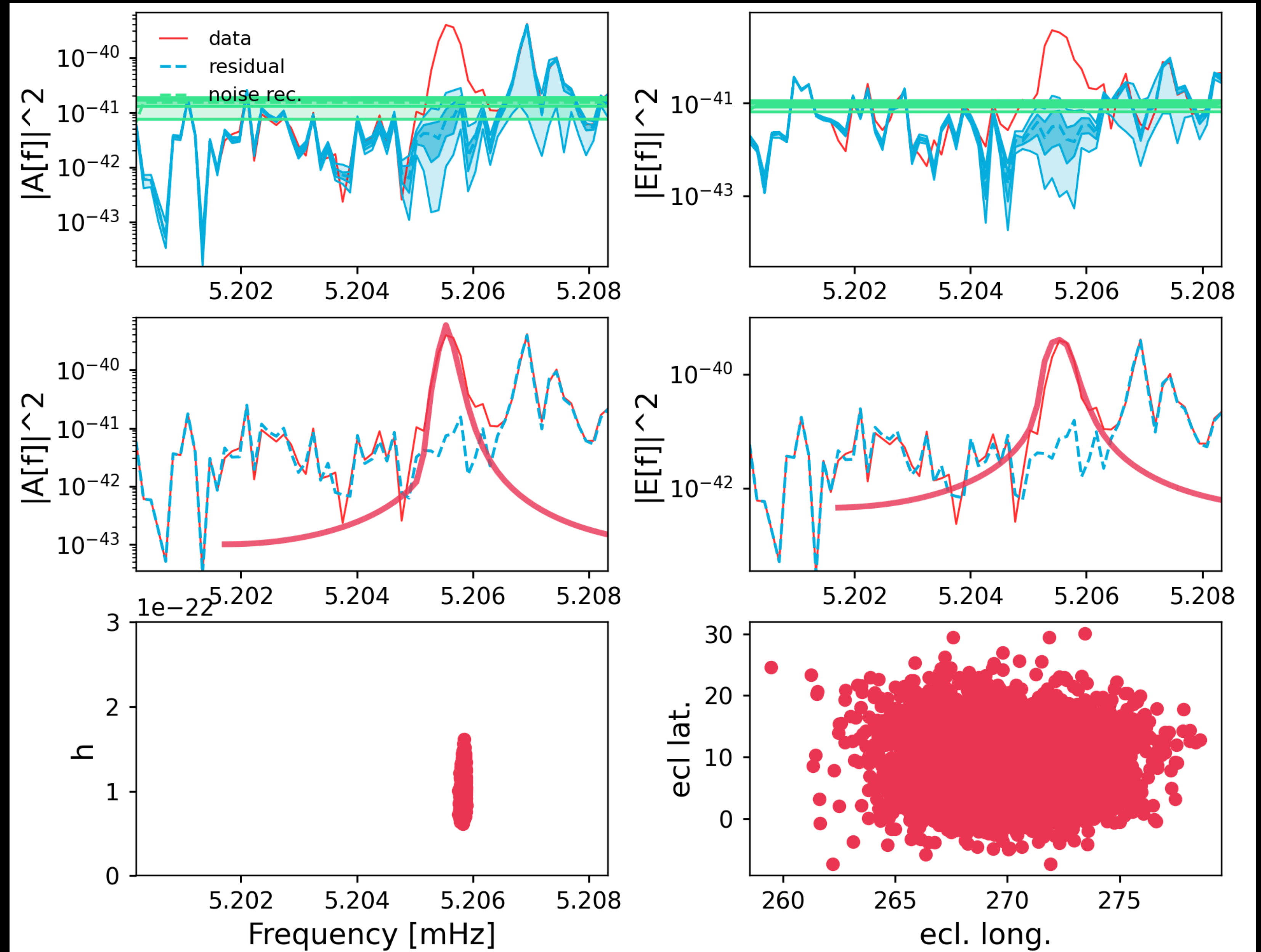
LDC - Challenge 1: *Radler*

1 month



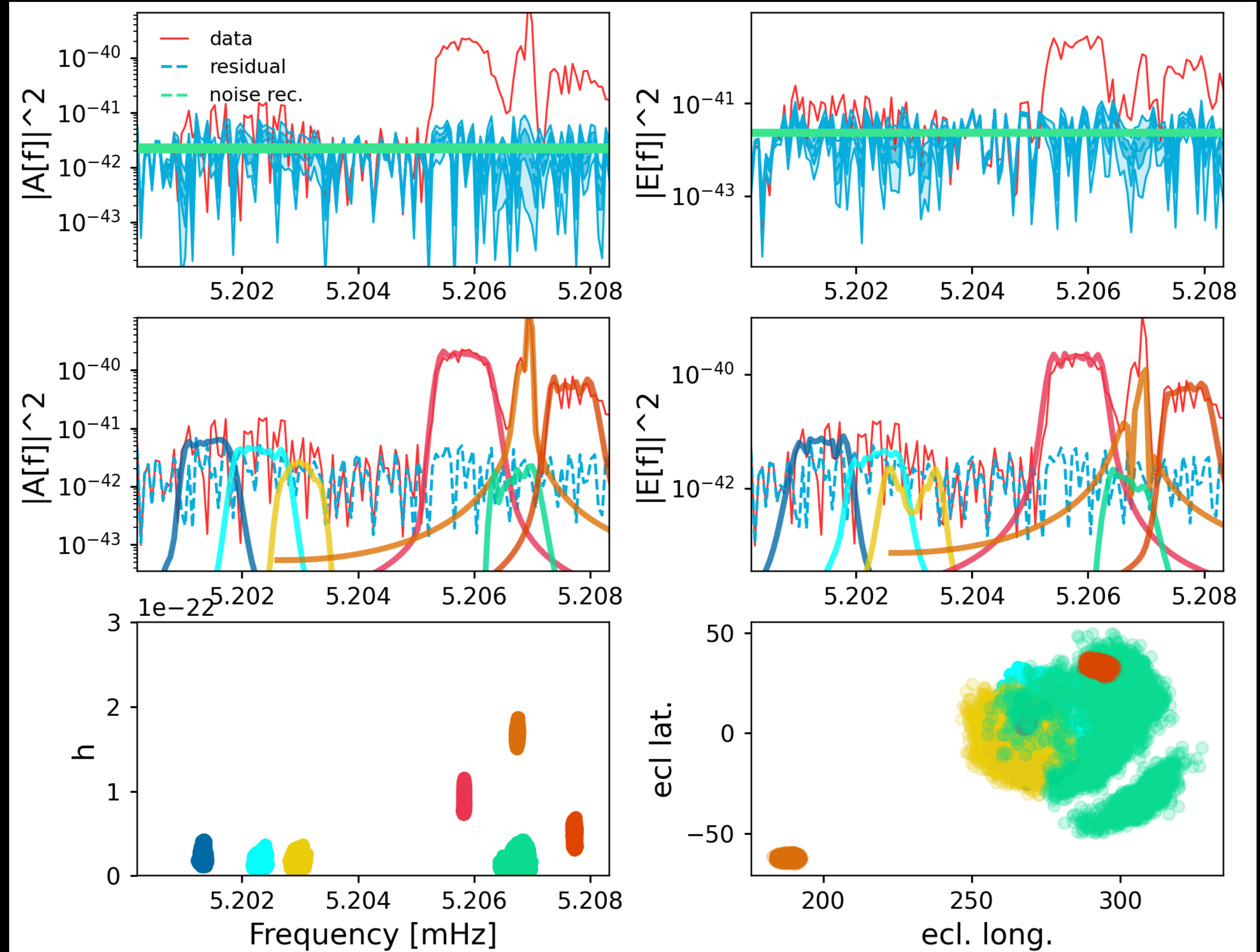
LDC - Challenge 1: *Radler*

3 months



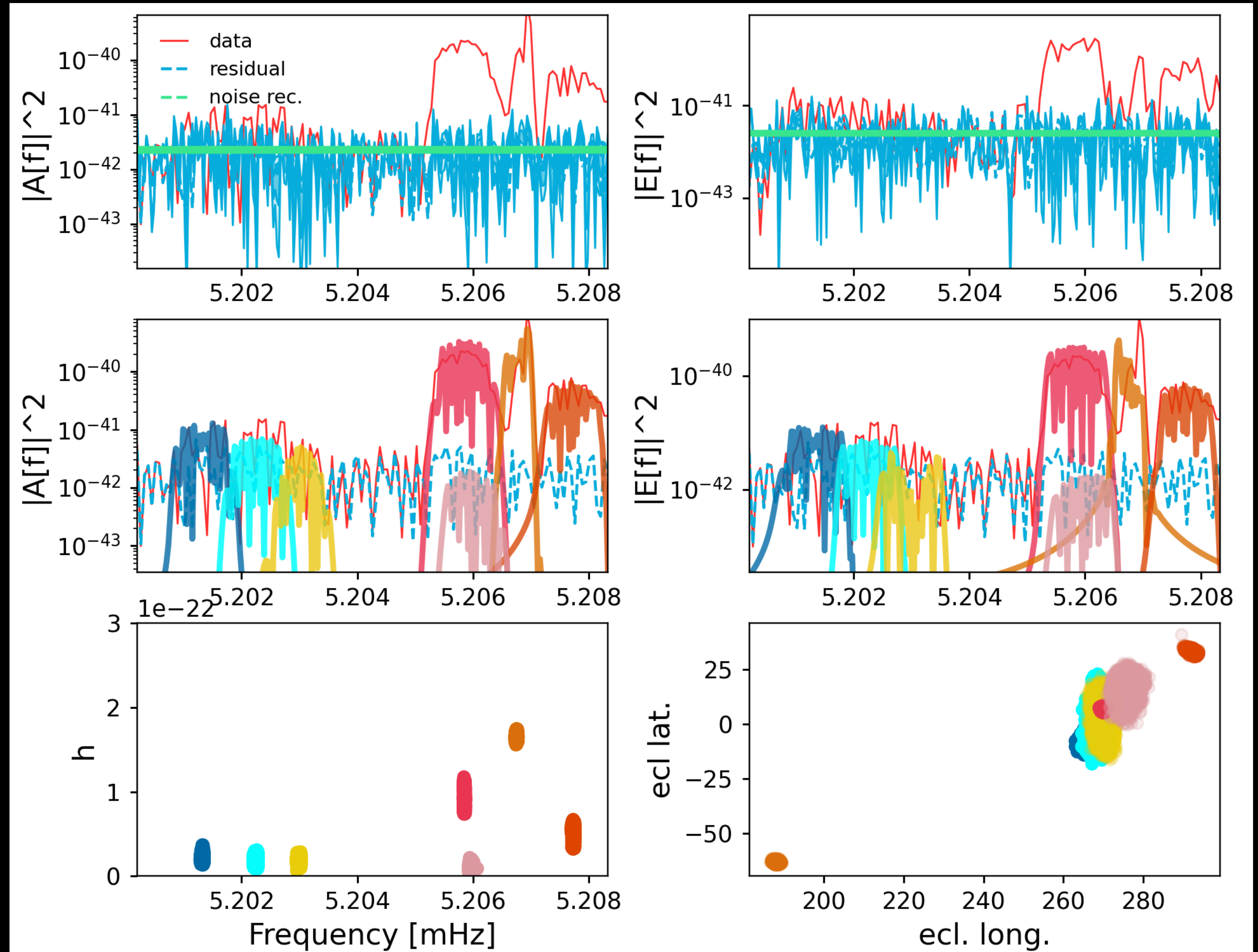
LDC - Challenge 1: *Radler*

6 months



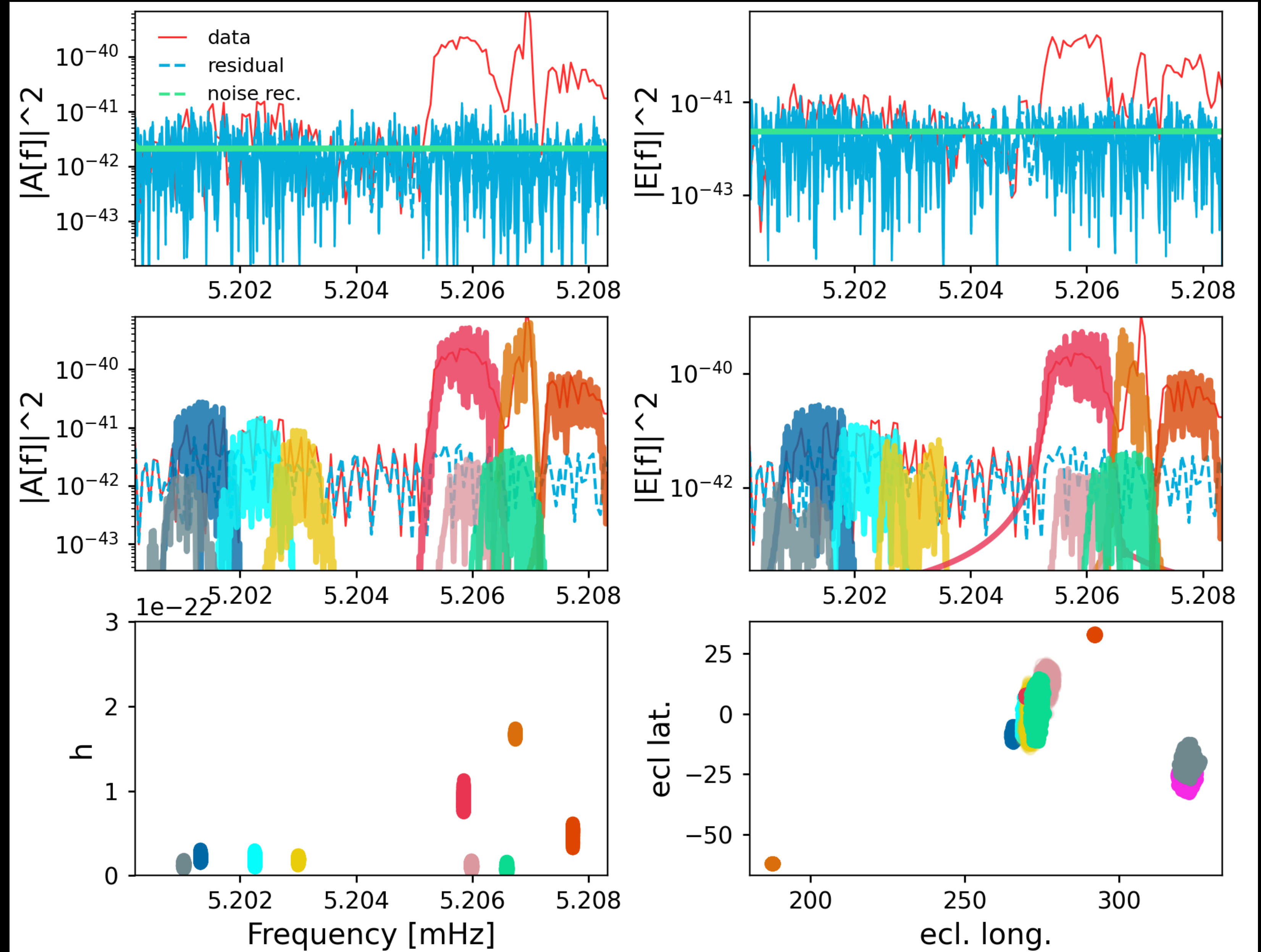
LDC - Challenge 1: *Radler*

12 months



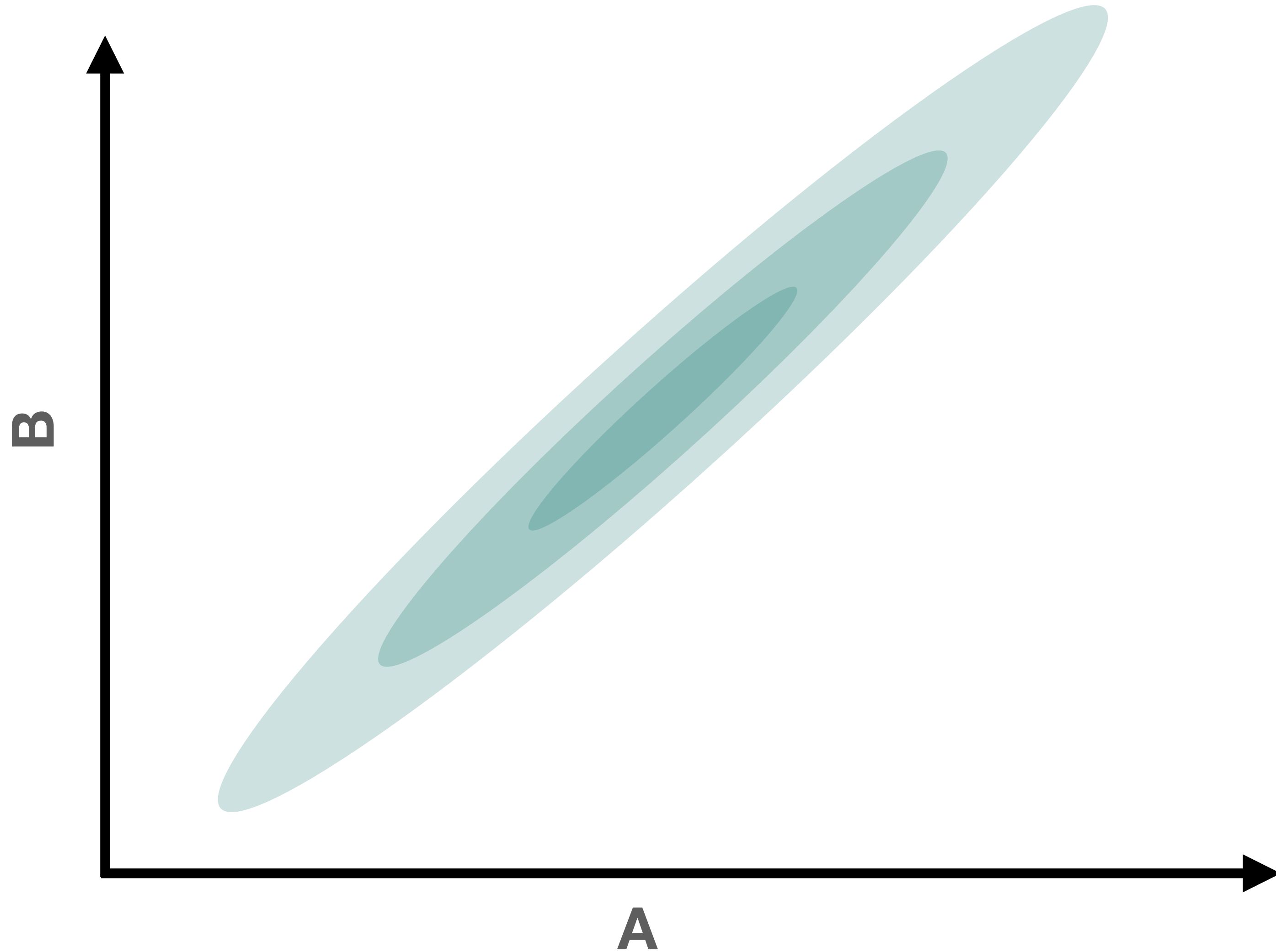
LDC - Challenge 1: *Radler*

24 months

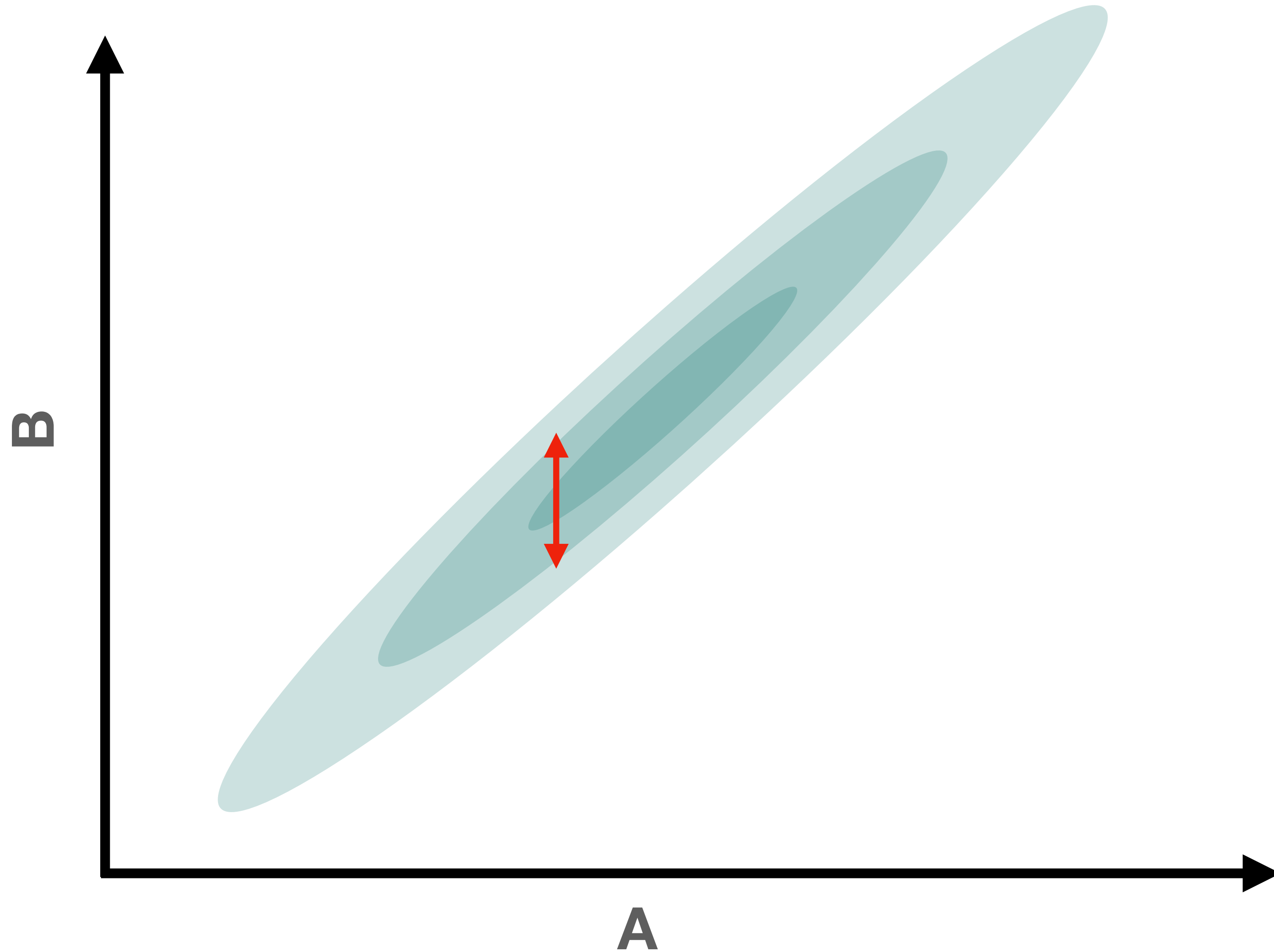


BLOCKED
GIBBS
SAMPLING

What is Gibbs Sampling?

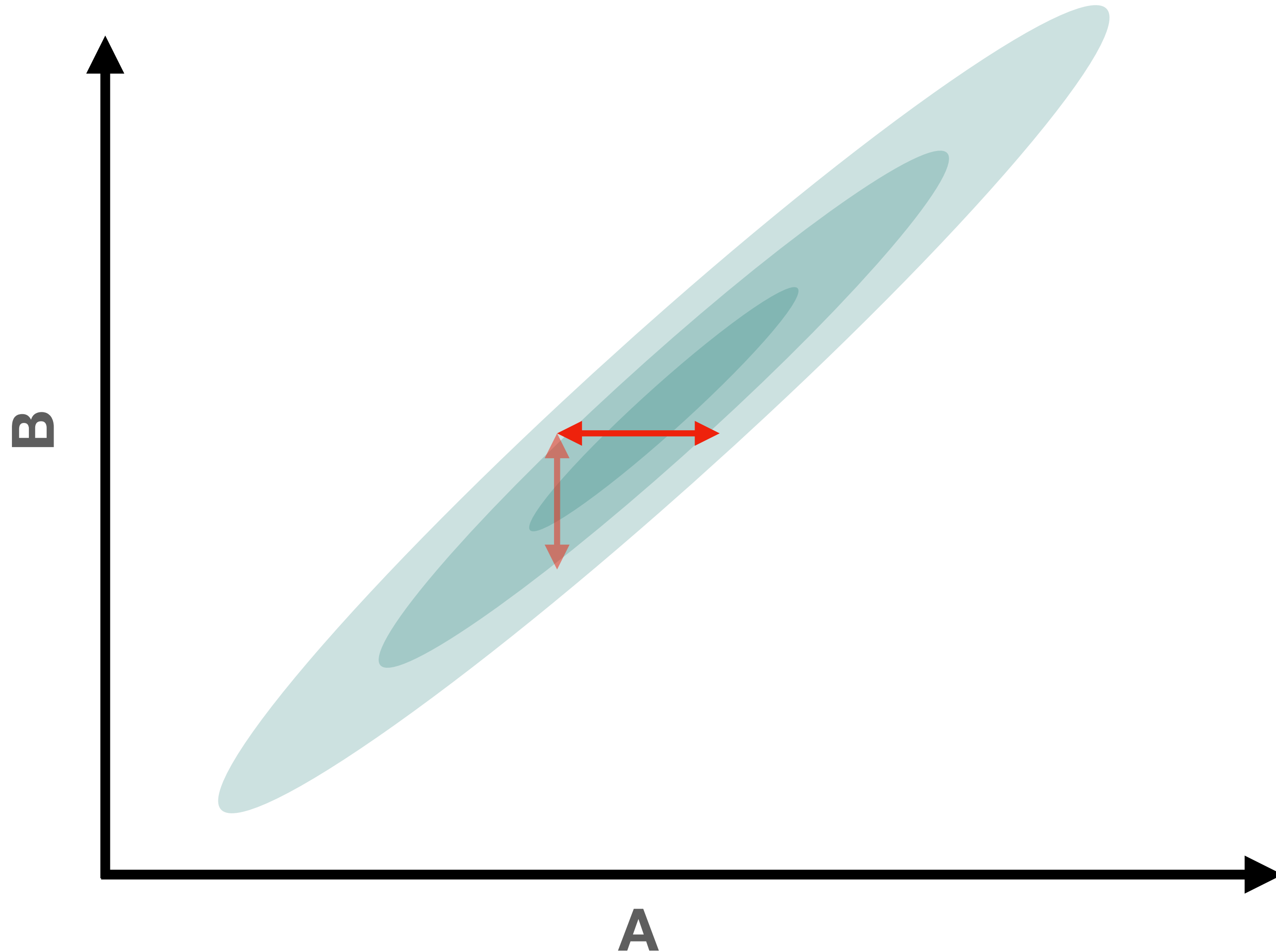


What is Gibbs Sampling?



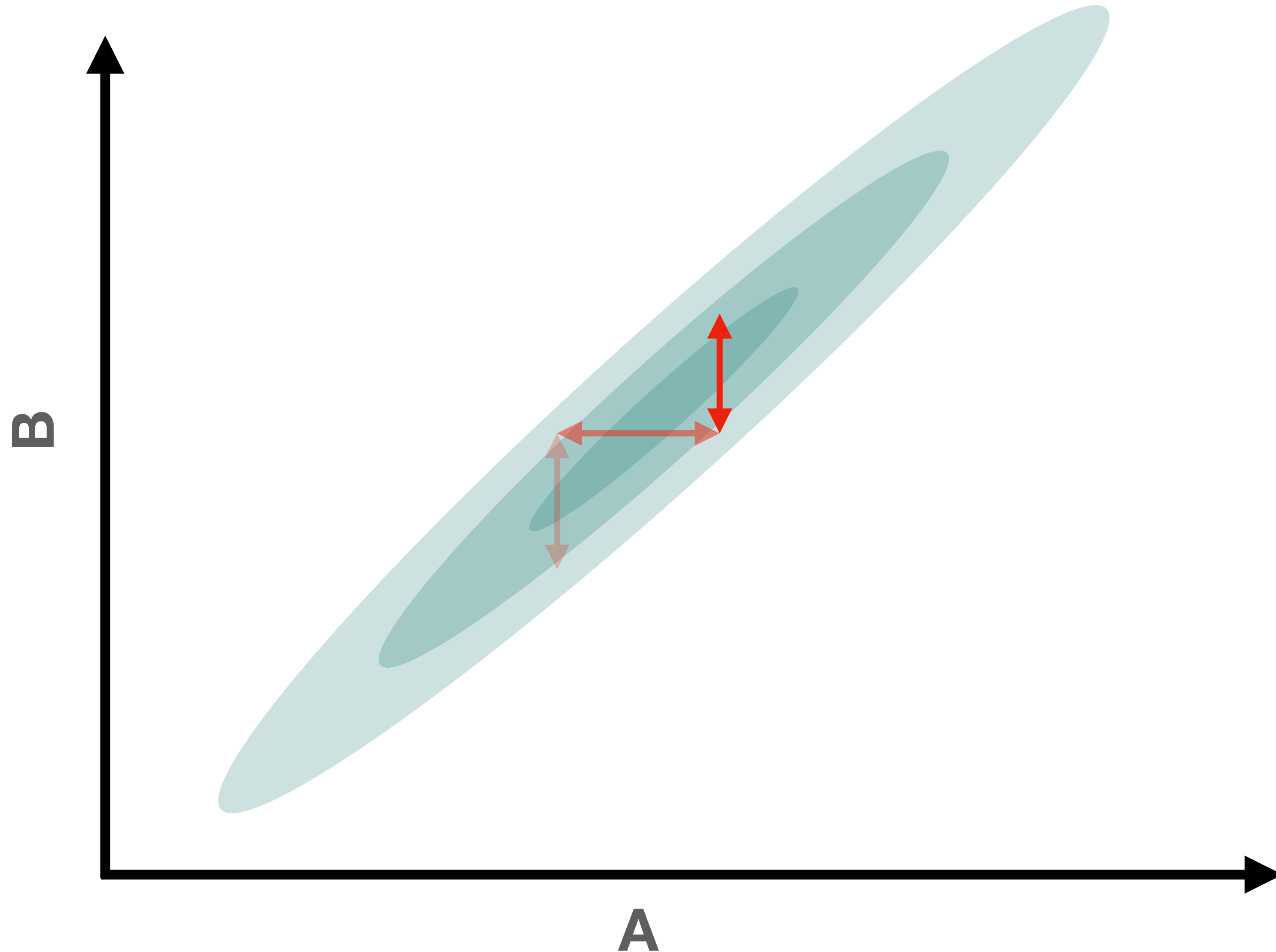
I. Hold Block A fixed and sample over B

What is Gibbs Sampling?



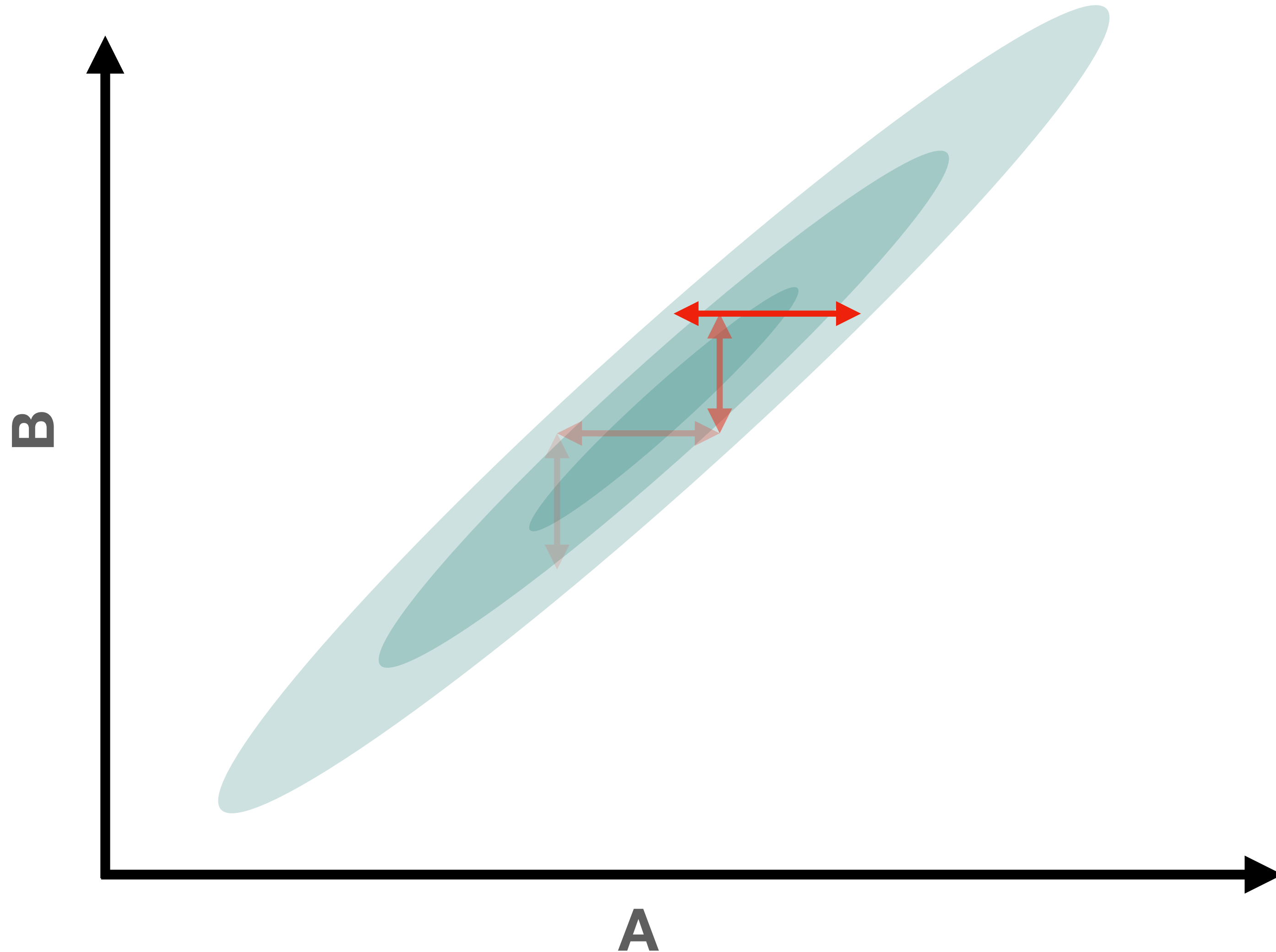
- I. Hold Block A fixed and sample over B
- II. Hold Block B fixed and sample over A.

What is Gibbs Sampling?



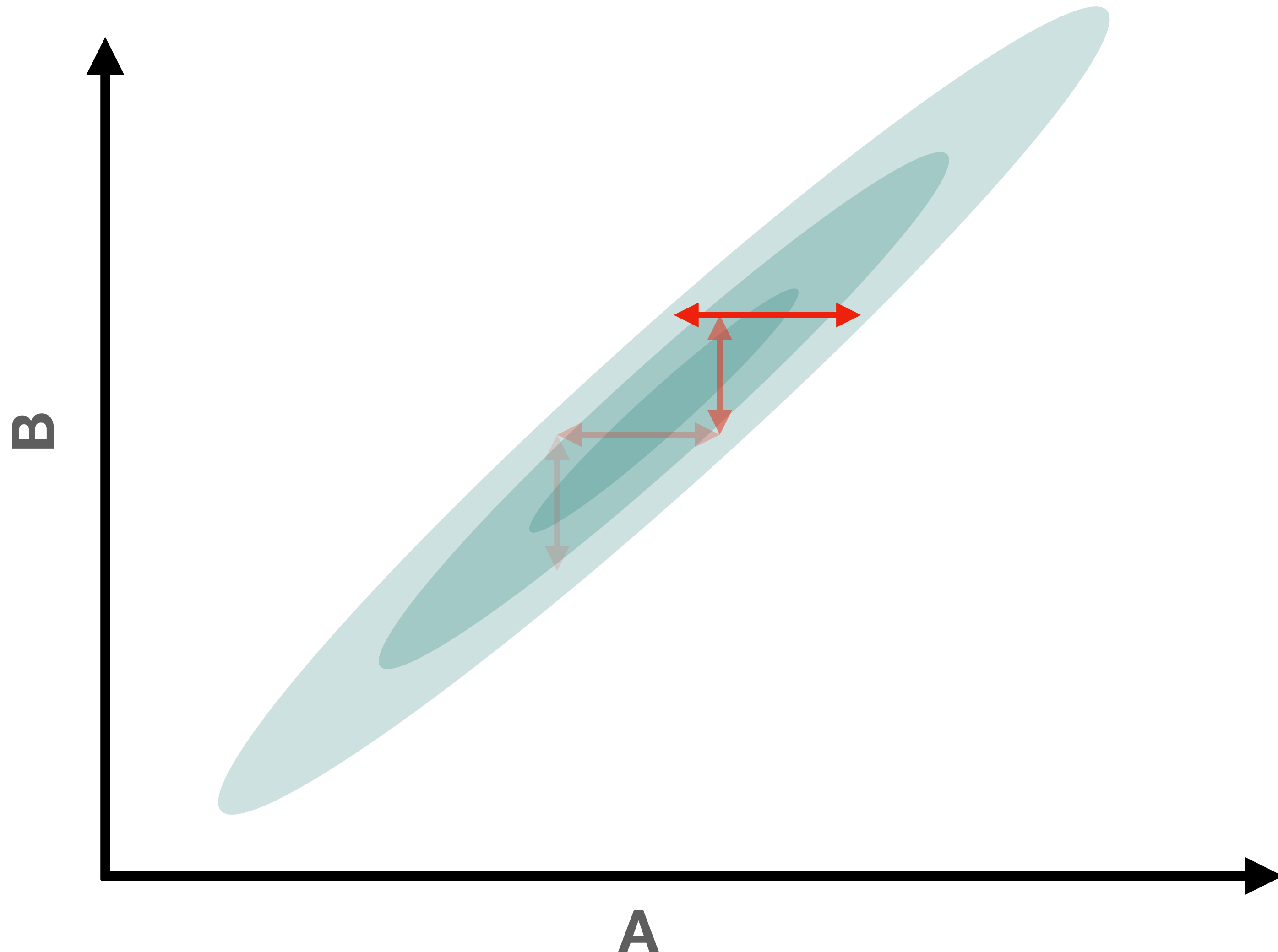
- I. Hold Block A fixed and sample over B
- II. Hold Block B fixed and sample over A.
- III. Repeat

What is Gibbs Sampling?



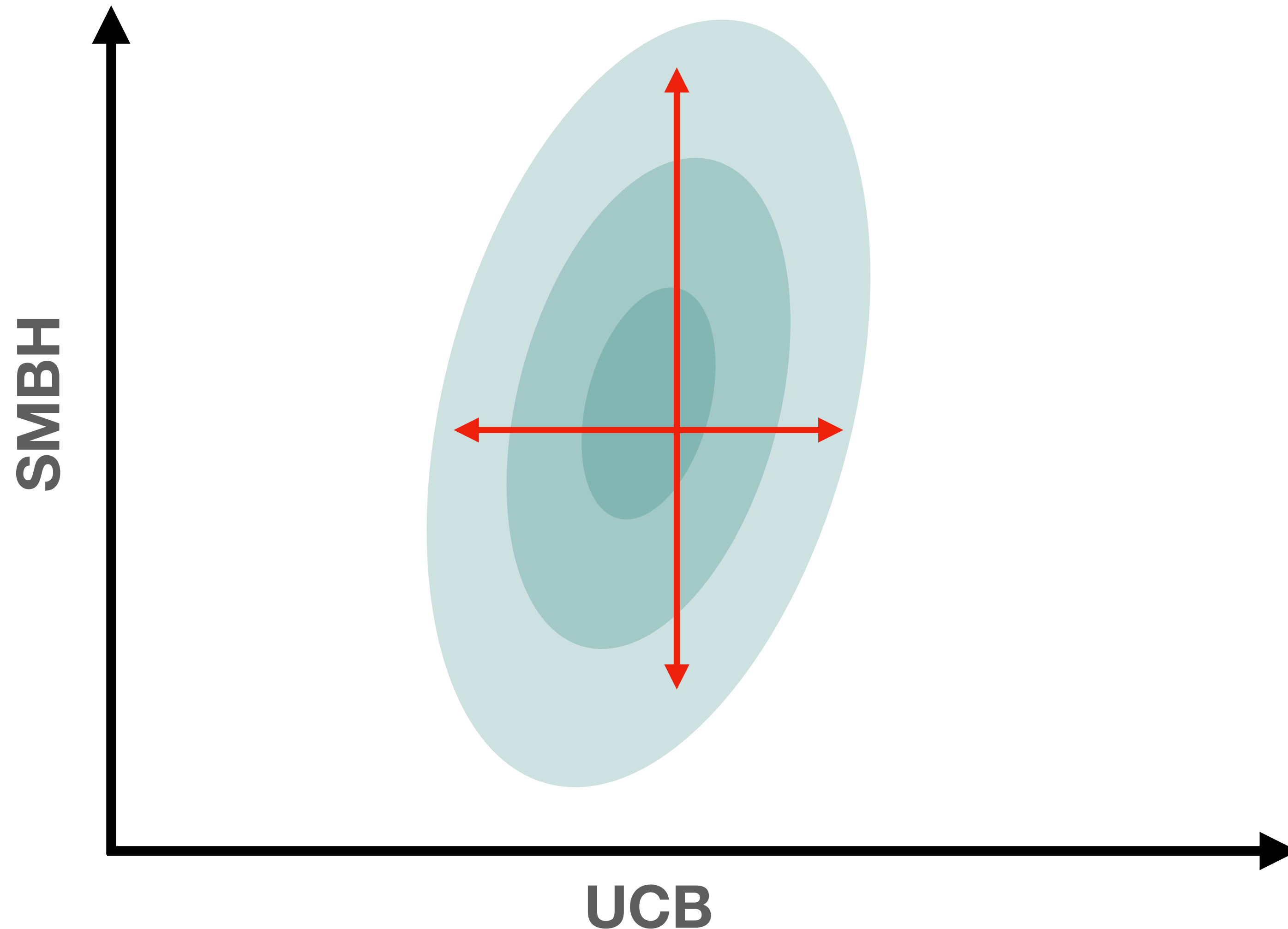
- I. Hold Block A fixed and sample over B
- II. Hold Block B fixed and sample over A.
- III. Repeat
- IV. Keep repeating...

What is Gibbs Sampling?



- I. Hold Block A fixed and sample over B
- II. Hold Block B fixed and sample over A.
- III. Repeat
- IV. Keep repeating...
- V. Converges miserably for correlated blocks

Why Gibbs Sampling?



- I.** Correlations between different source types are small.
- II.** “Blocks” are developed independently w/ common API.
- III.** With clever scheduling is nicely parallelized.
- IV.** Blocks can be inserted/removed without disrupting the workflow.

Global Fit Design

Data Shared Between Blocks (e.g. Residuals, Noise Cov. Matrix, etc.)

