MAX-PLANCK-GESELLSCHAFT

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

Henri Martin

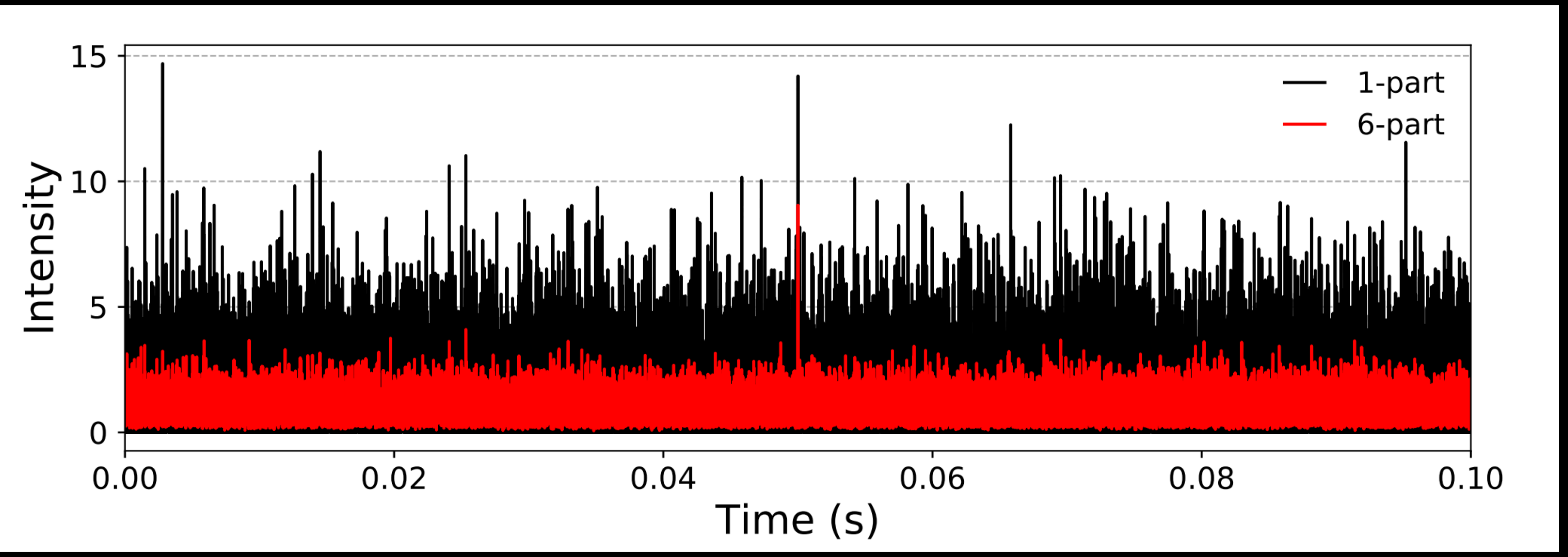
Toulouse, France

November 23, 2022

Introduction - current radio projects

Eigenfiltering - new technique to search for ultra-fast transients

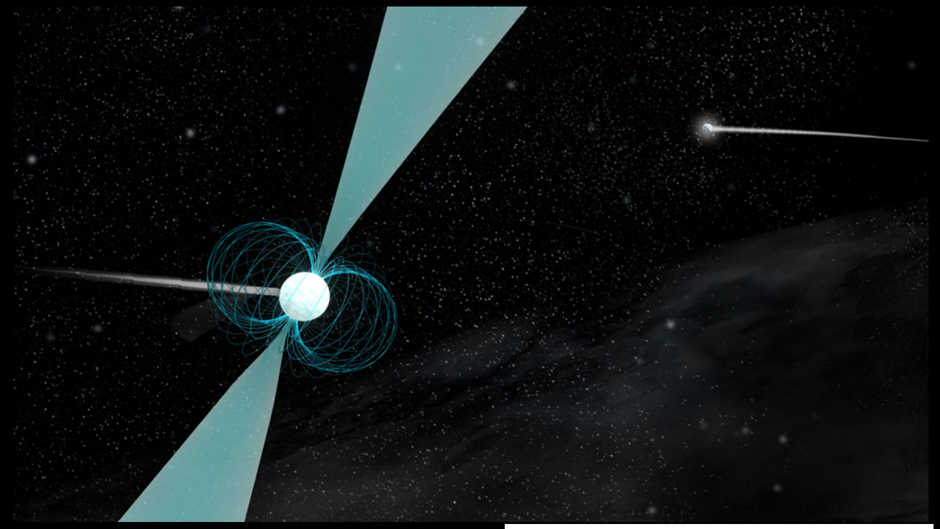
10.1088/1361-6382/ab95e4



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

- binary pulsars

- quasars



Saxton (NRAO/AUI/NSF)

ALMA (ESO/NAOJ/NRAO))

VAMPIRA - provenance collection from a computationally intensive pipelines

arXiv:2109.10759

(astronomy lead)

Introduction - current GW+radio projects

<u>EPTA collaboration</u>

- pulsar timing

-absorption of GWs by the IGM

10.1088/1361-6382/ac5376

- emission of GWs by SMBH accretion disks

-Effelsberg observing

<u>MeerKAT</u>

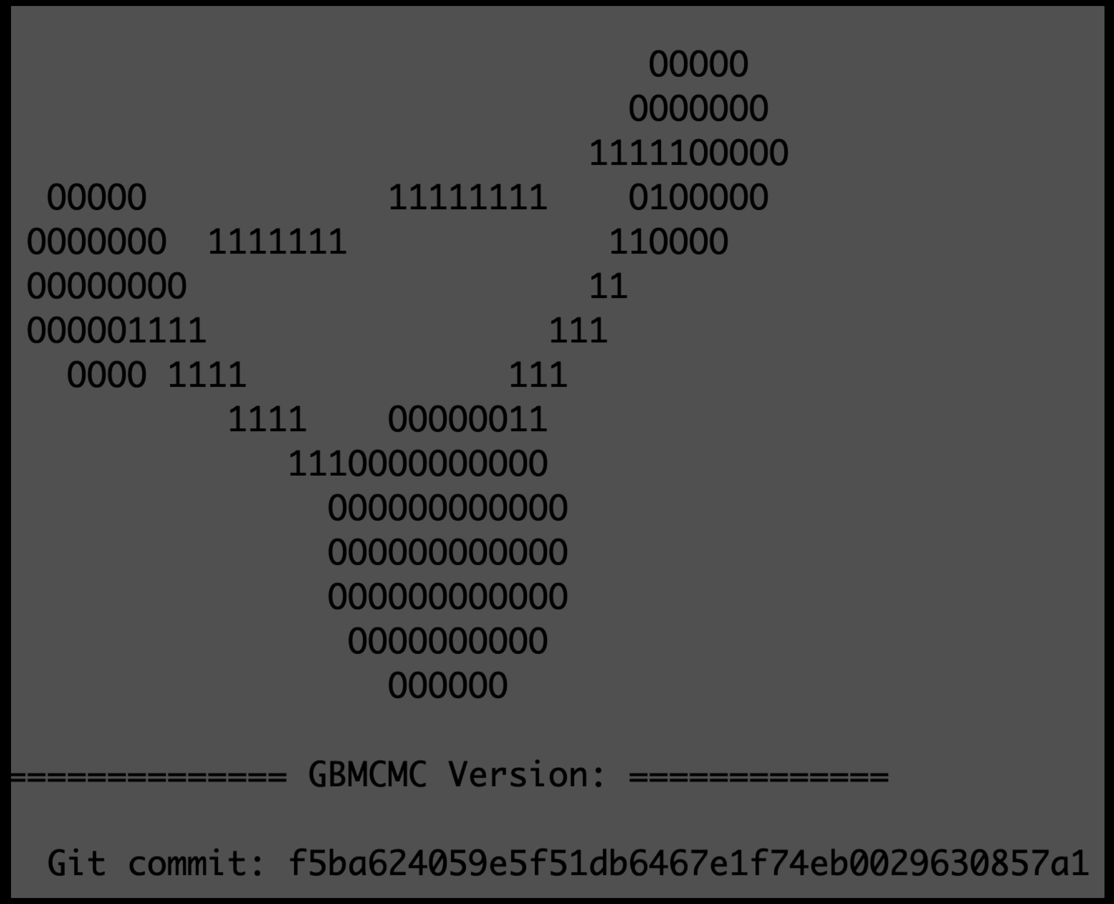
- check if new binaries discovered by MeerKAT are detectable by LISA (student project)

Introduction - current GW projects

LISA

- data analysis, parameter estimation from MCMCs
- code development

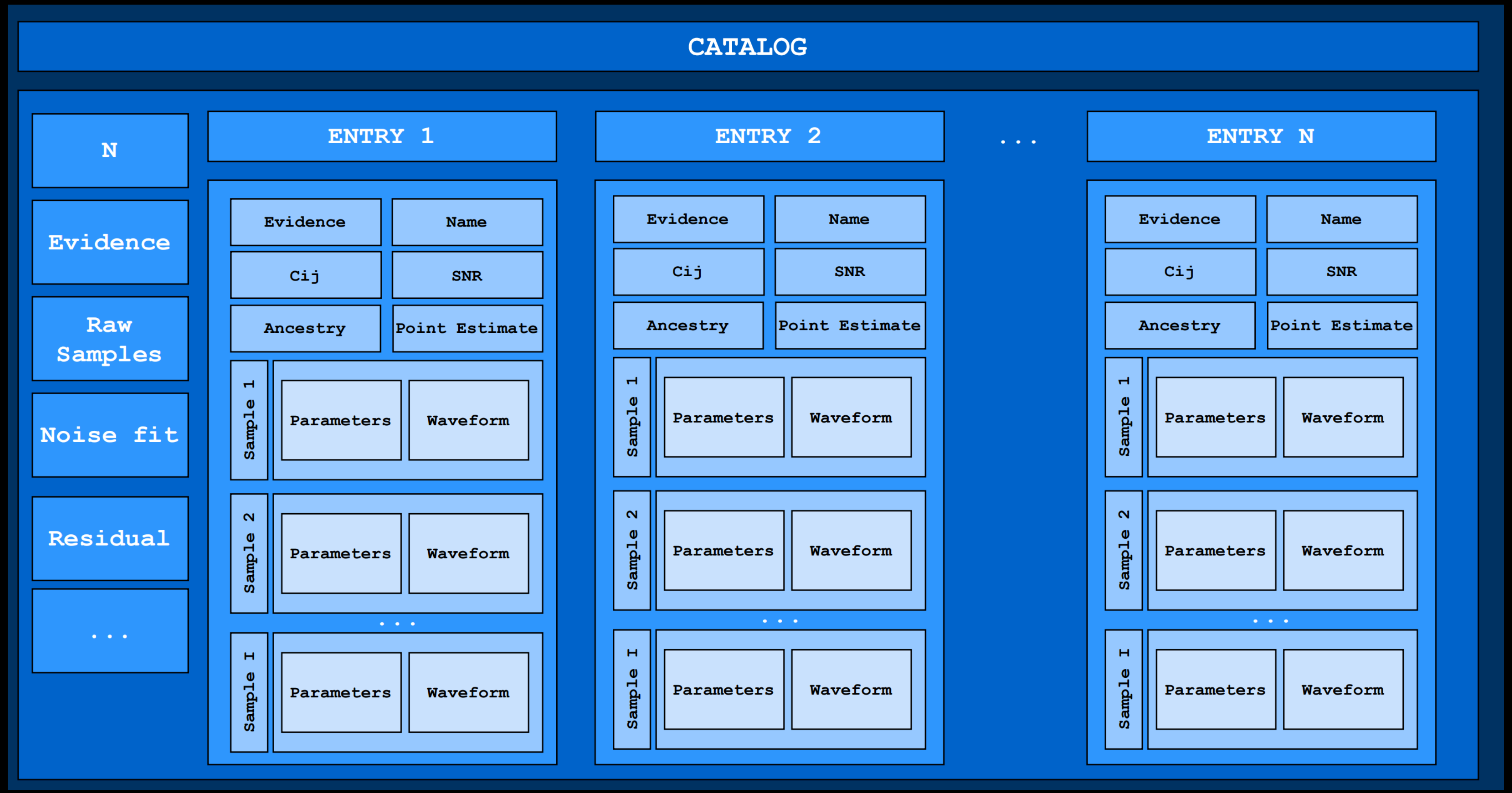
https://github.com/tlittenberg/ldasoft



- catalog creation

10.1103/PhysRevD.101.123021

- using ML to sort MCMC chain data into a catalog



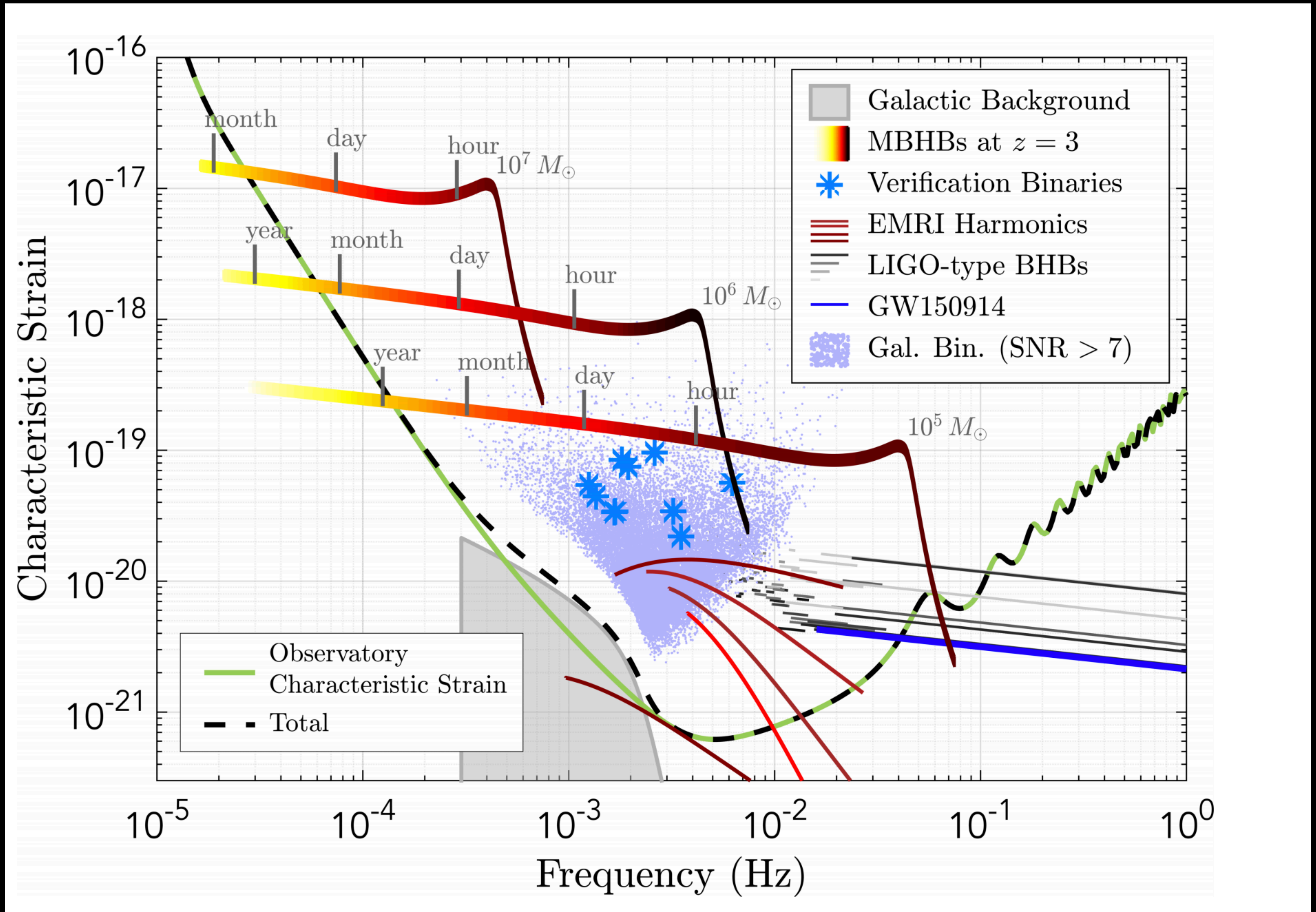
LISA Data Challenge

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.



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

Massive black hole binaries

Galactic binaries

Extreme mass ratio inspirals

Stochastic GW background (cosmological and astrophysical origin)

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:

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

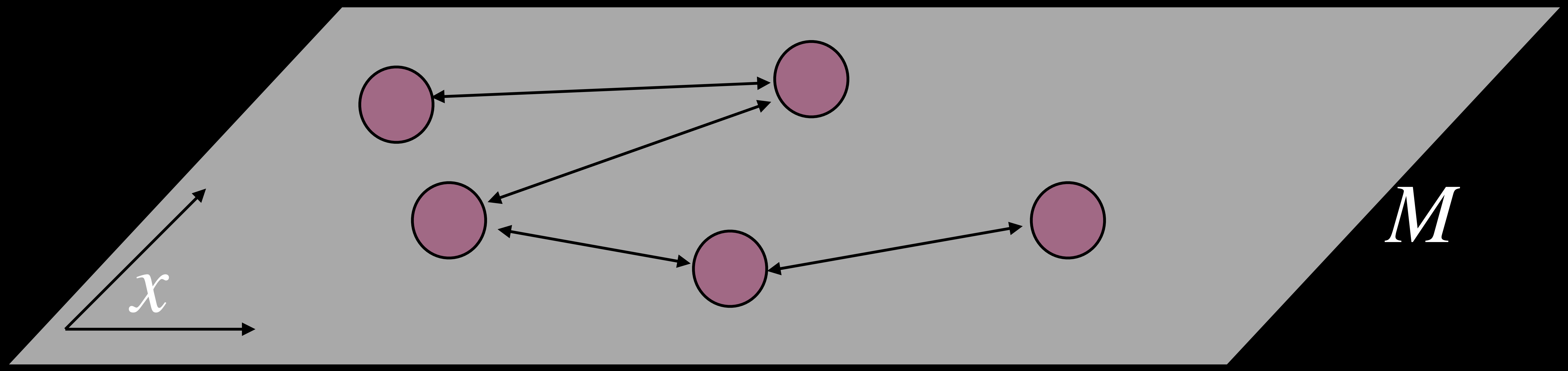
Likelihood = “Goodness of Fit” for parameters

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

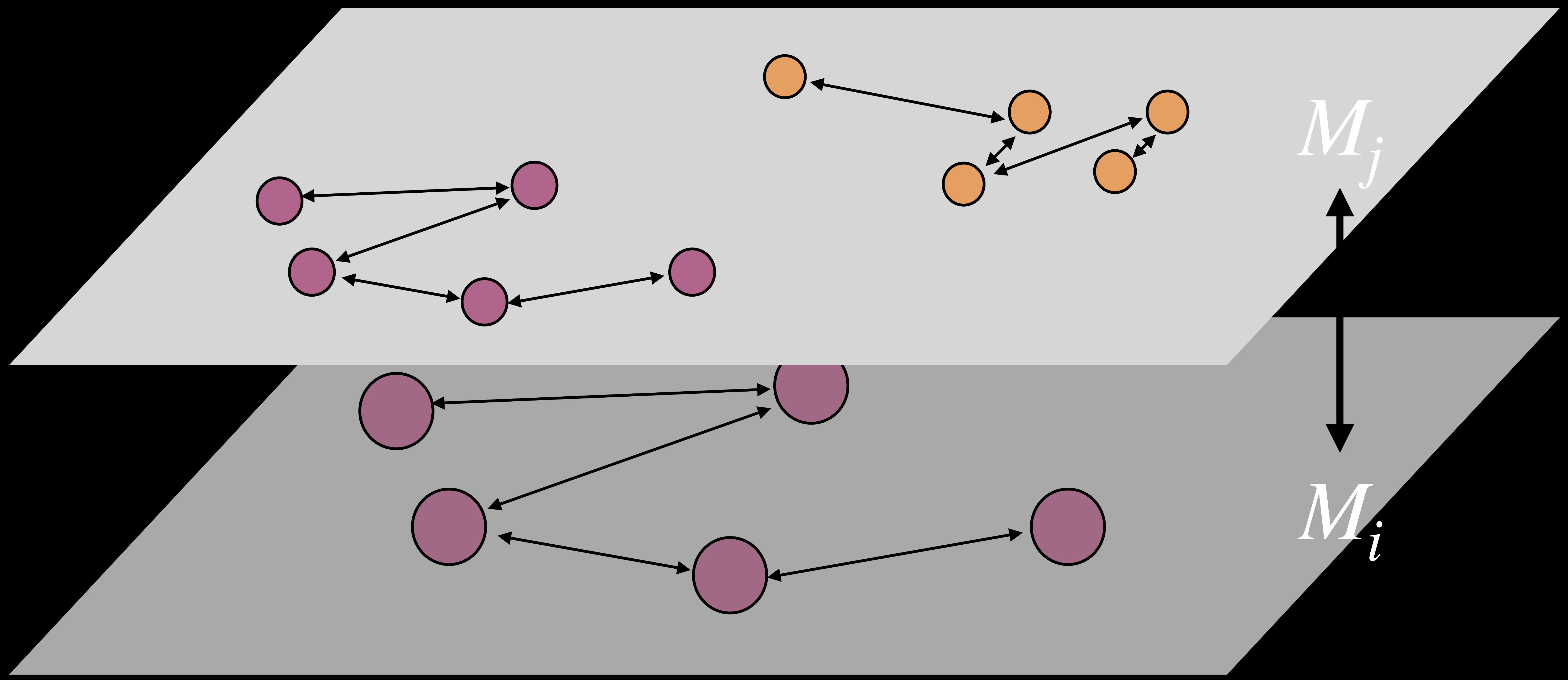
II. Always converges, usually faster than grid-based approaches when parameter space is **LARGE**.

III. Detection, characterization, and quantifying confidence

IV. Stochastically sample between models with RJMCMC

Why MCMC?

Why MCMC?



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

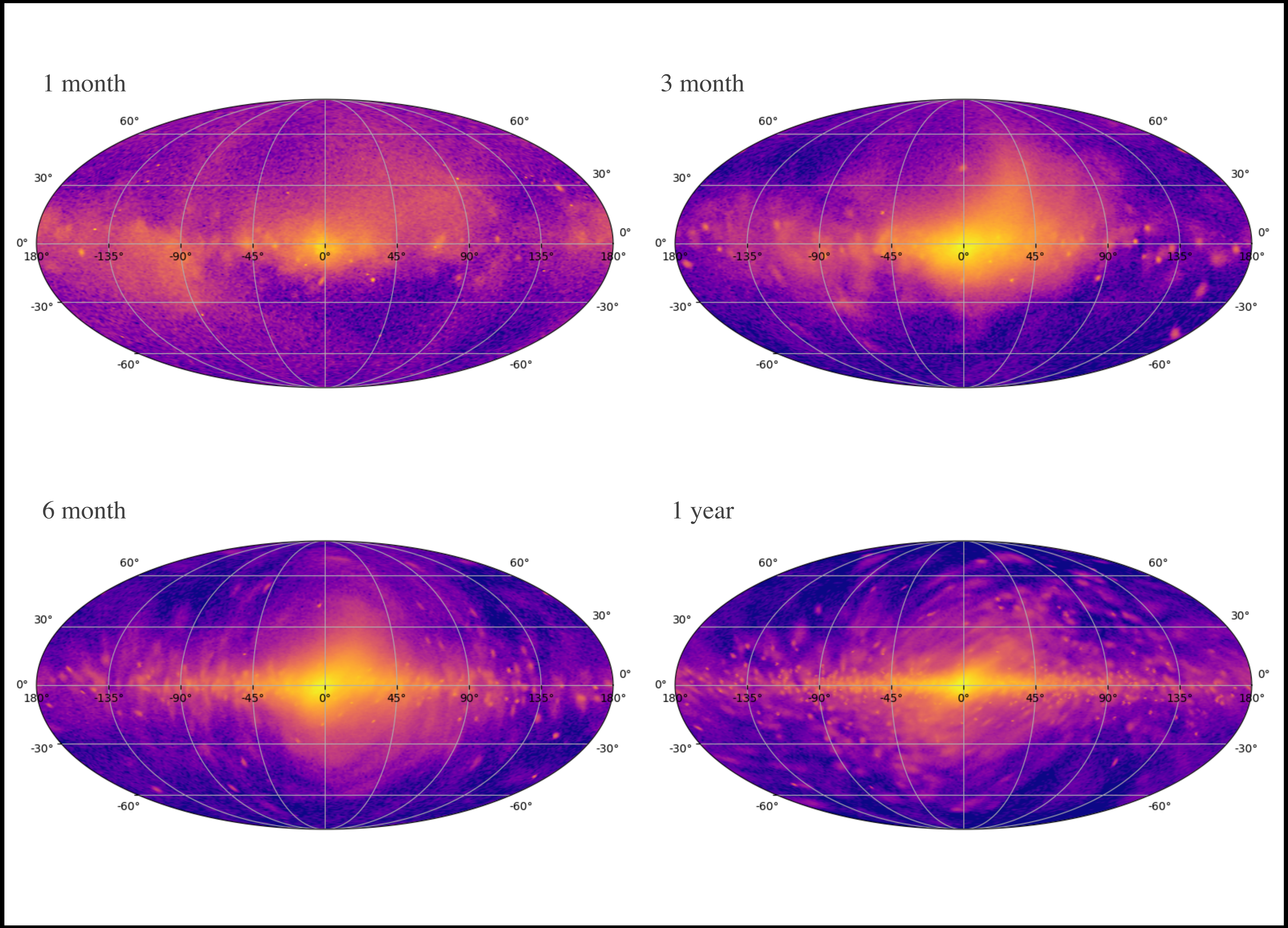
I. Stochastically sample large and complicated parameter spaces

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

III. Detection, characterization, and quantifying confidence

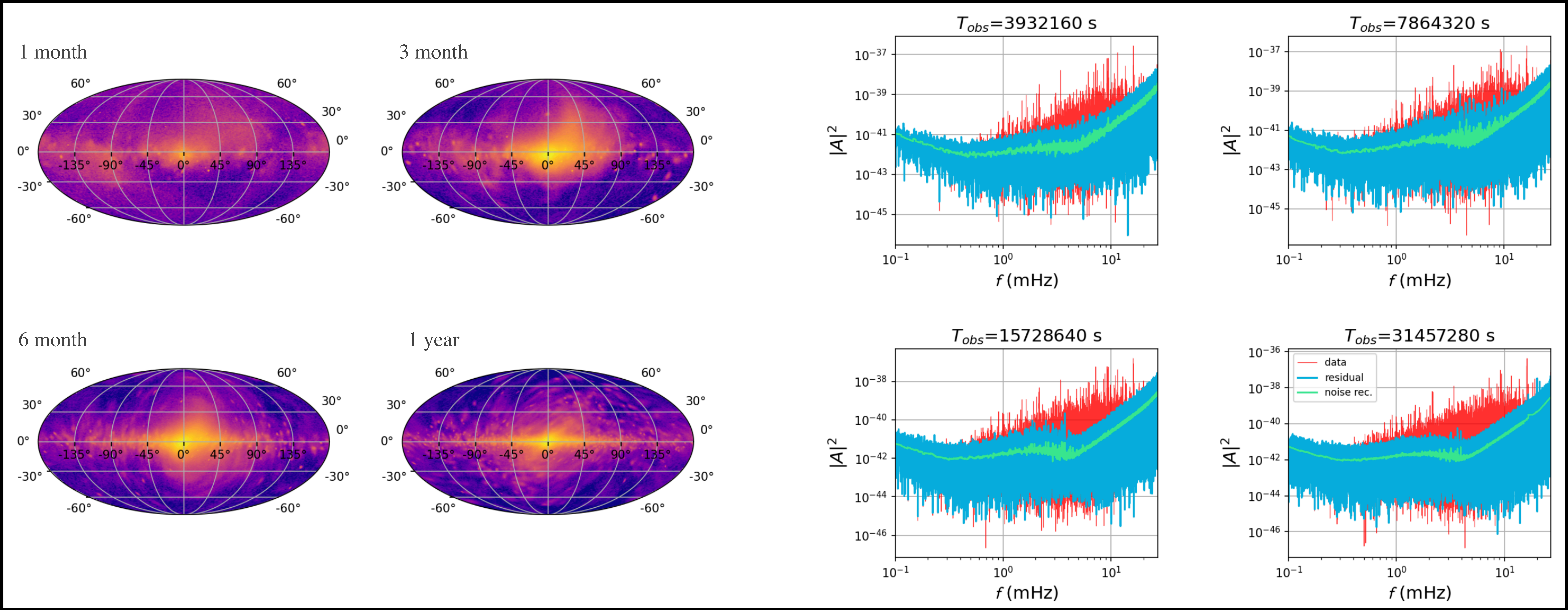
IV. **Stochastically sample between models with RJMCMC**

LDC - Challenge 1: *Radler*

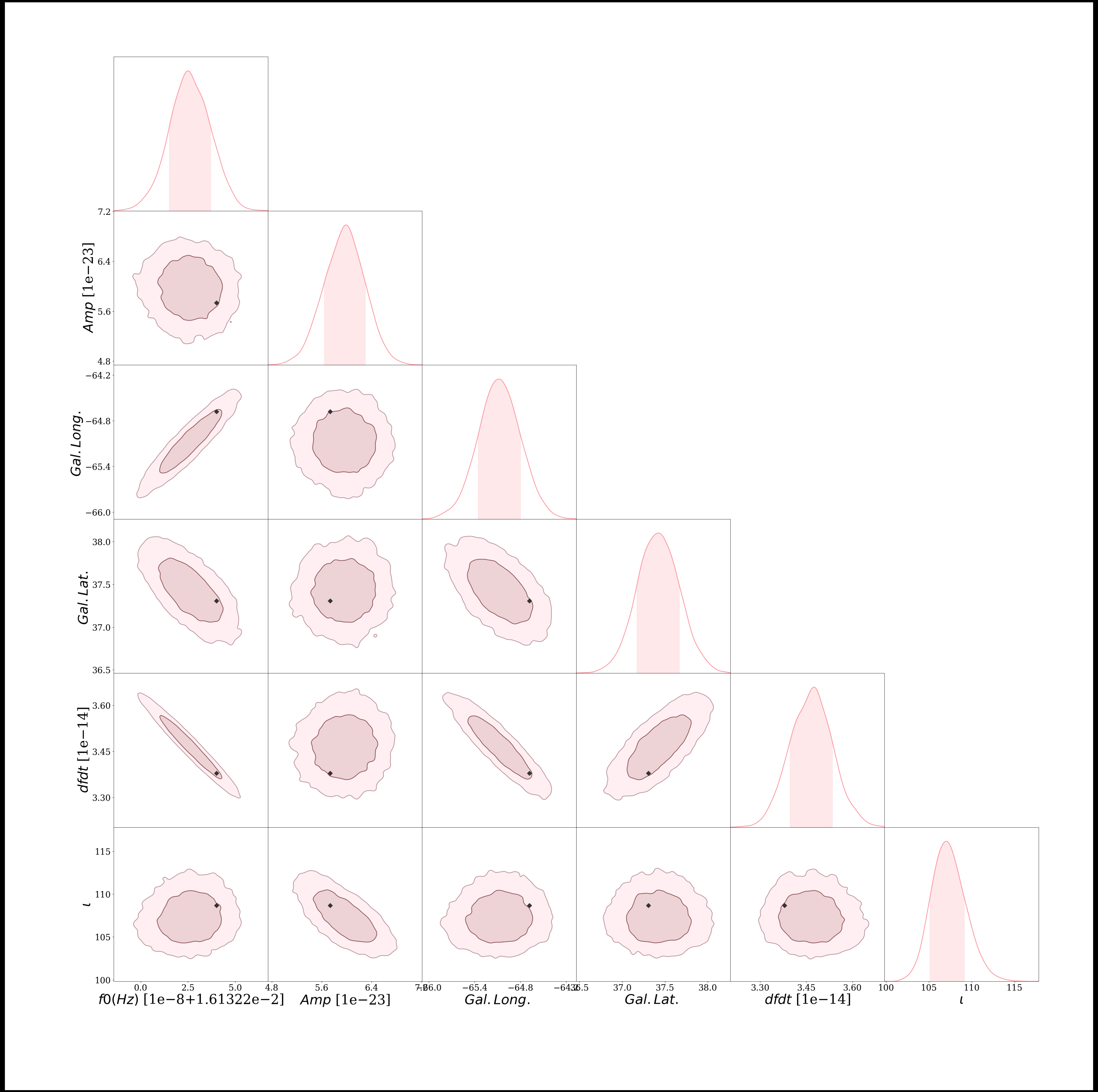


LDC - Challenge 1: *Radler*

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



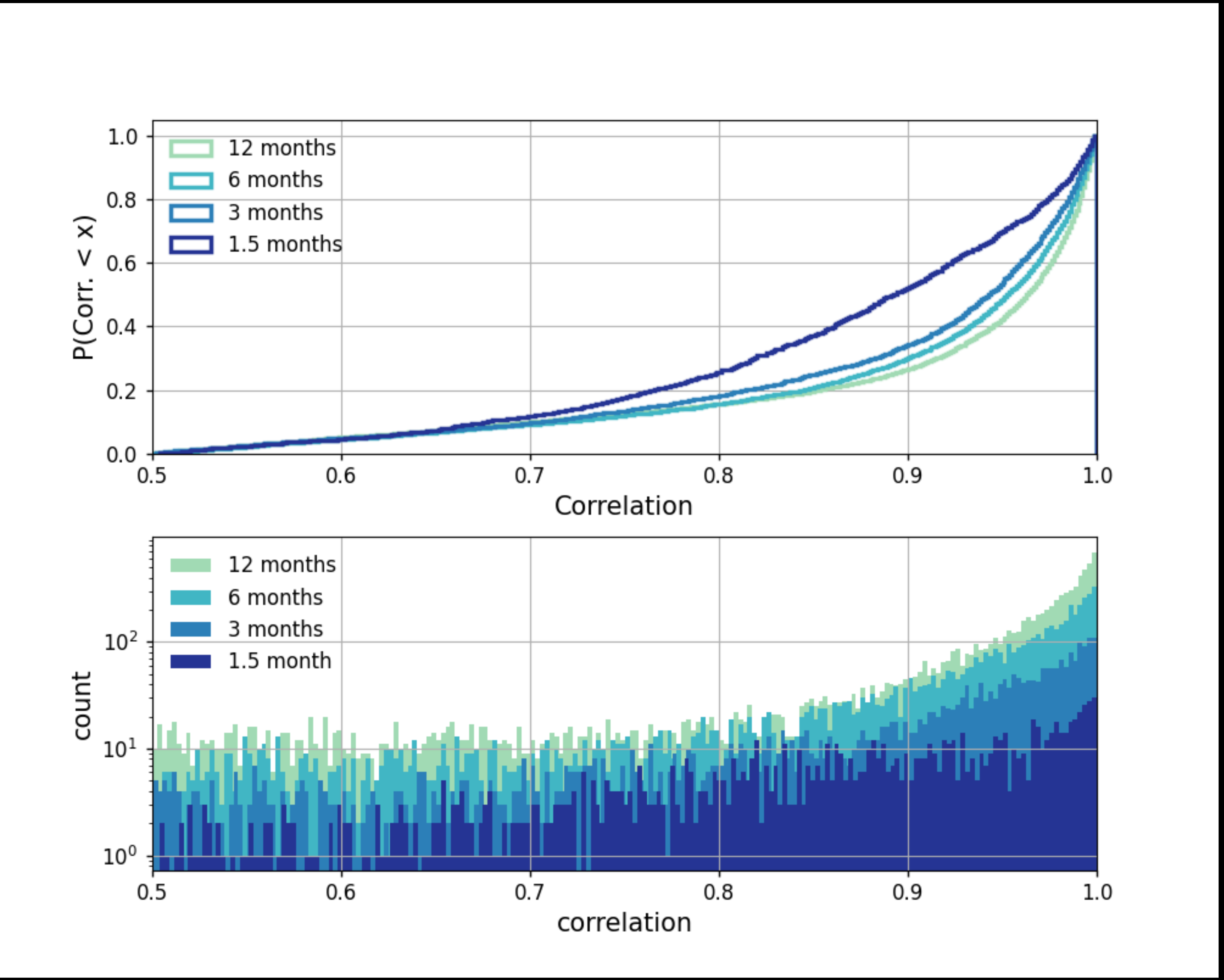
LDC - Challenge 1: *Radler*

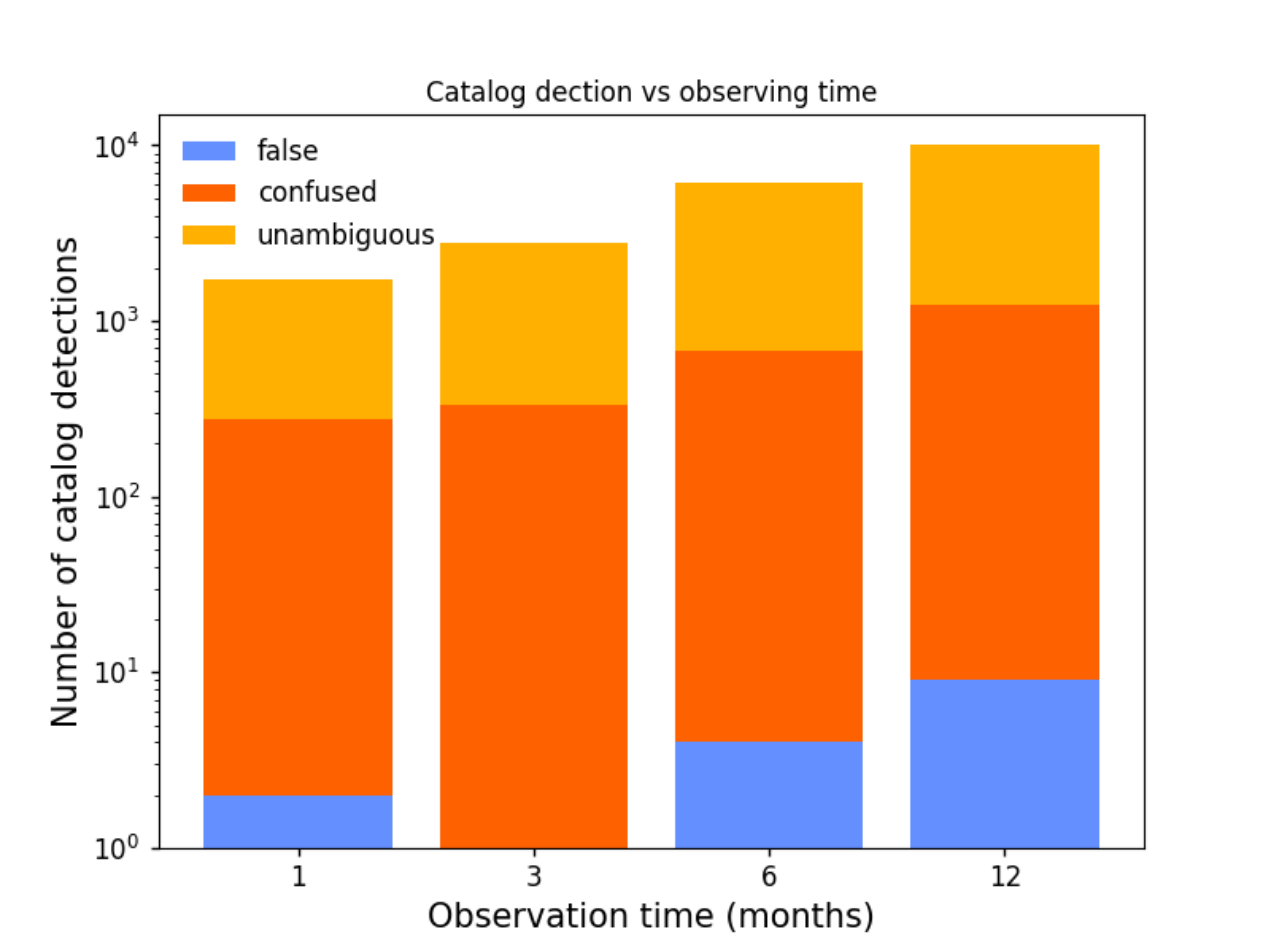


Observation time	sky localisation 10 sq. deg.	$\dot{f}>0$	$\dot{f}<0$	eclipsing	eclipsing with $\dot{f}>0$	eclipsing and well-localised
1.5	2	279	5	91	16	0
3.0	3	553	1	372	48	0
6.0	25	1950	10	1027	275	4
12.0	441	4239	20	1873	735	58

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



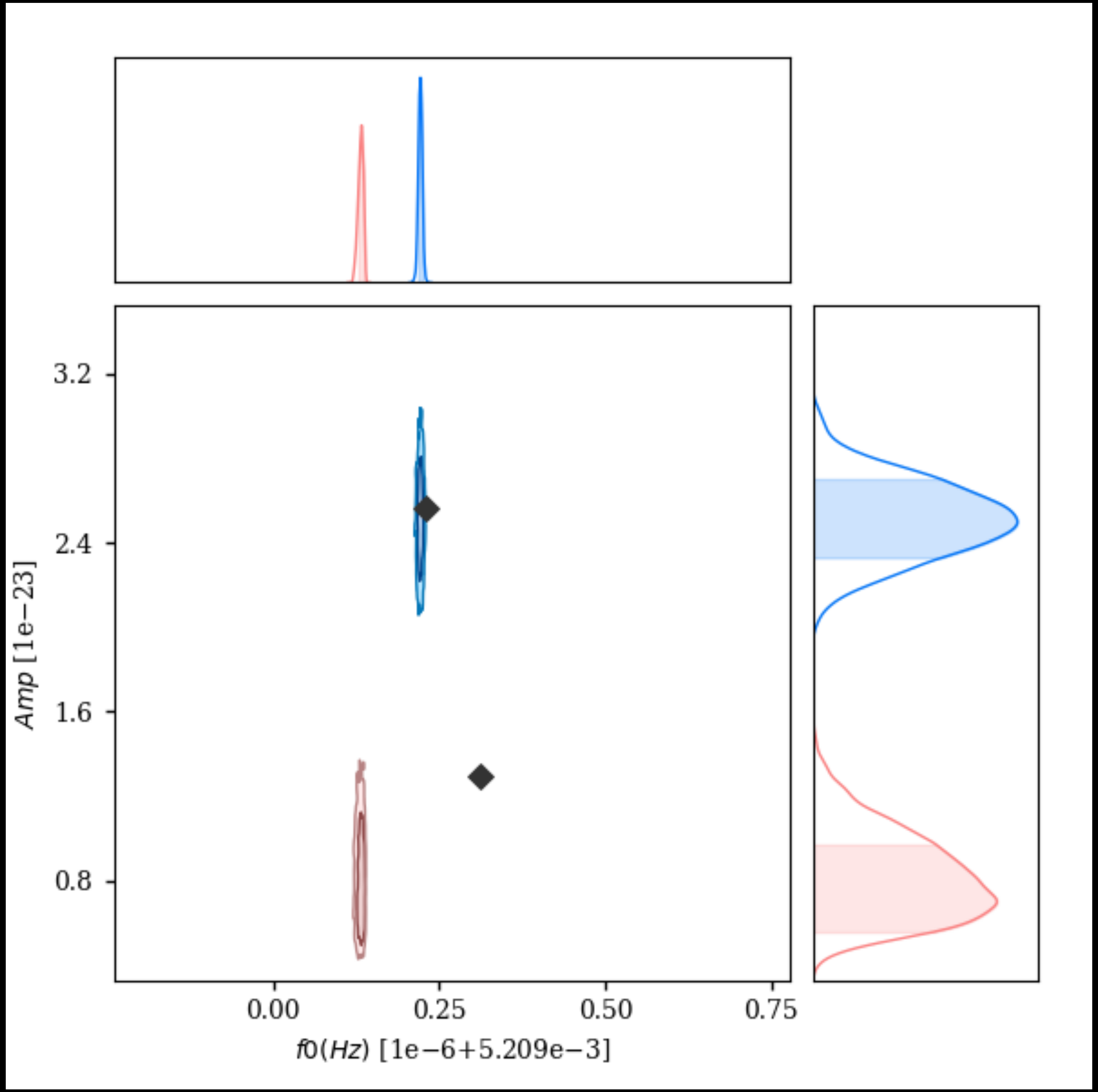


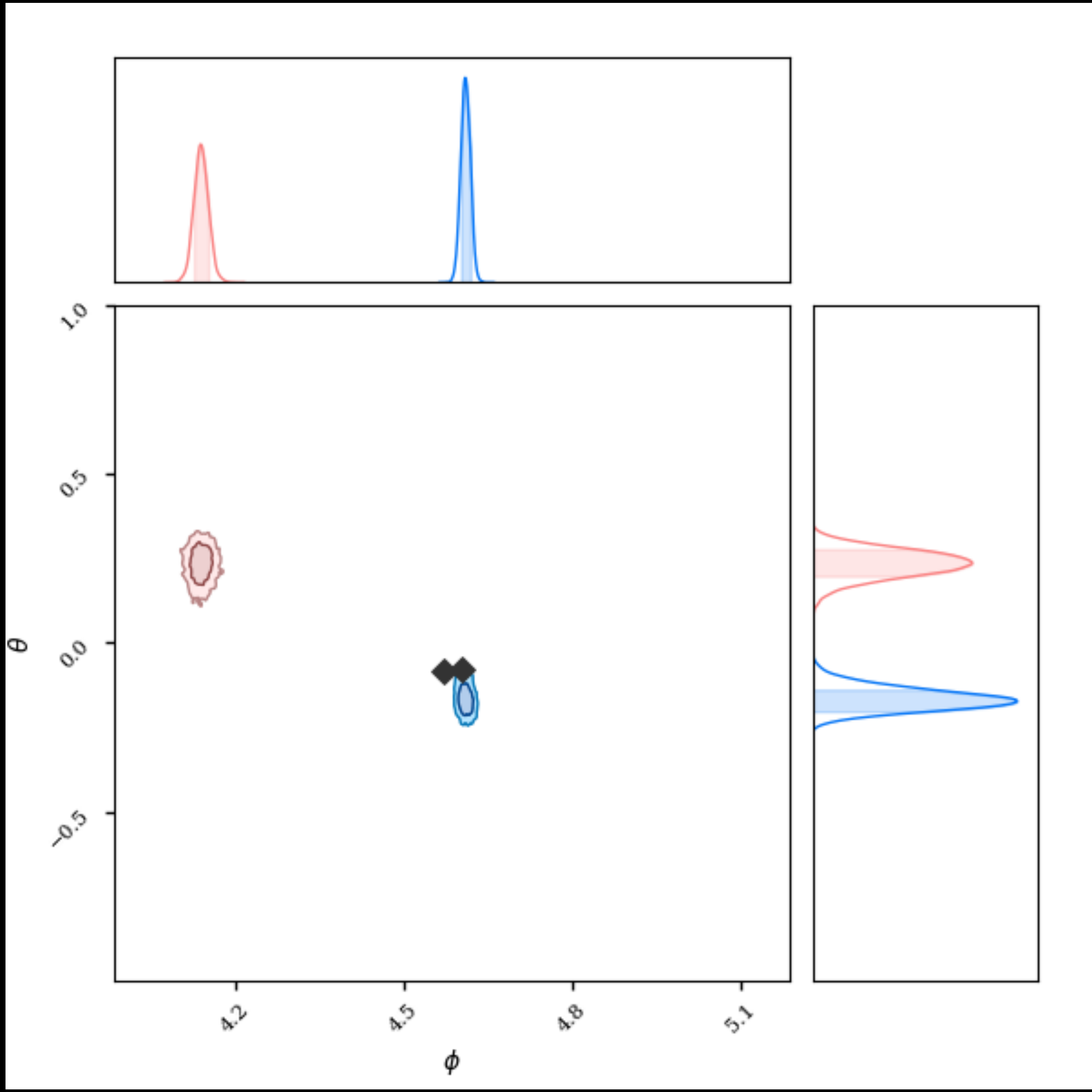
Kristen Lackeos

LDC - Challenge 1: *Radler*

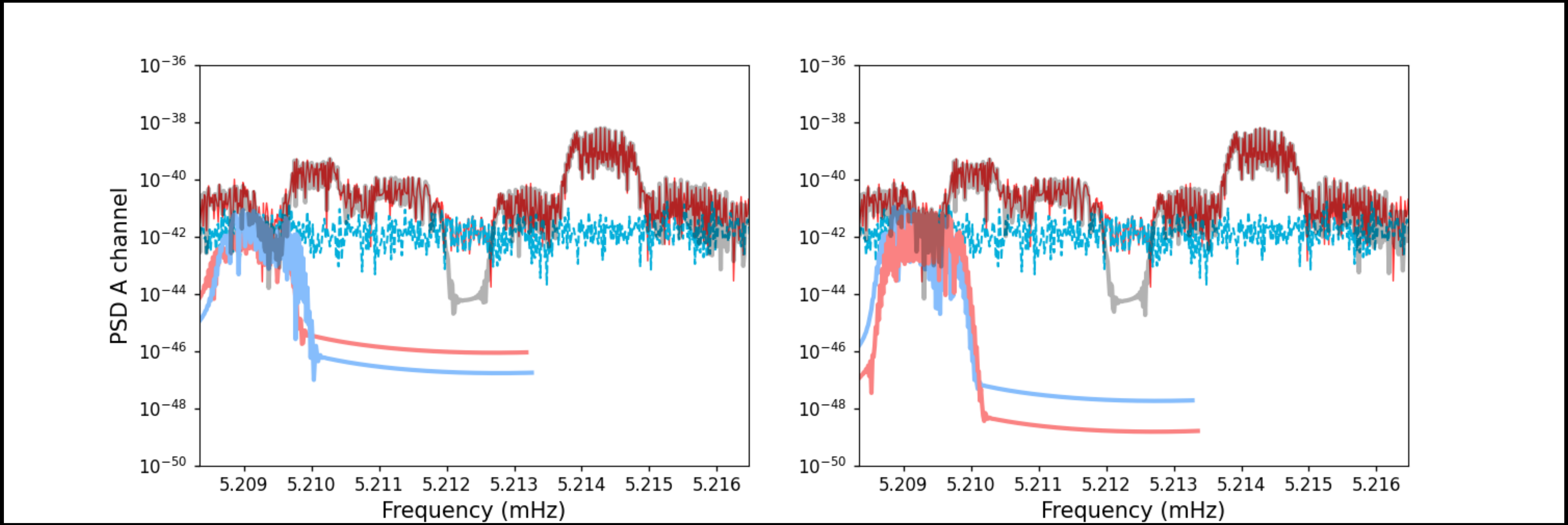
Kristen Lackeos

LDC - Challenge 1: *Radler*





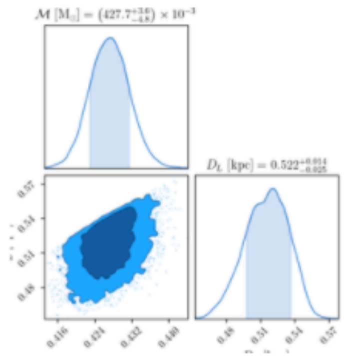
LDC - Challenge 1: *Radler*



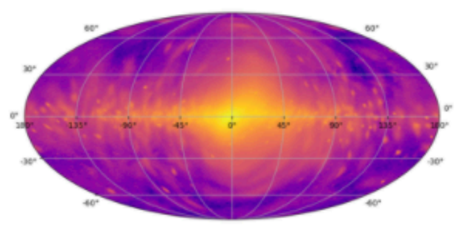
UCB Use Case Gallery

lisacattools

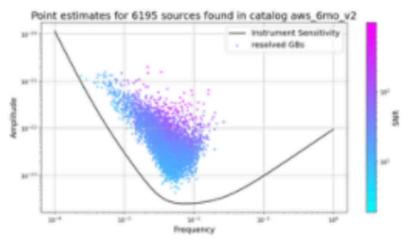
Below are examples using UCB catalogs. Catalog data products are formatted as `pandas` data frames and stored in `HDF5` files. Top level catalog files have the list of all candidate detections, point estimates, etc. In addition, posterior samples for each candidate are stored as separate data frames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



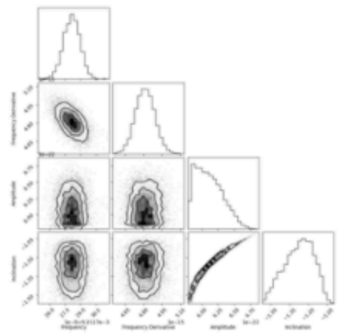
Resampling posteriors



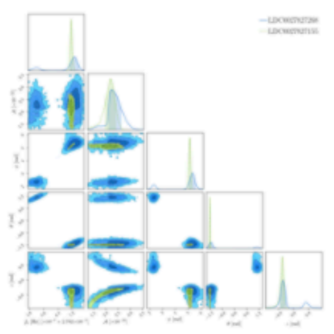
Joint PDF of sky location



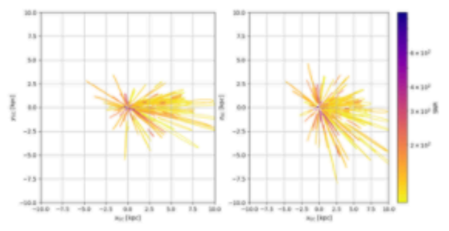
Scatter plots



Corner plots

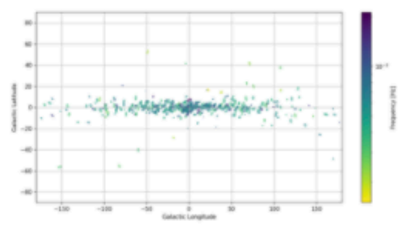


Connecting catalogs



3D map of the galaxy

Parameter tables



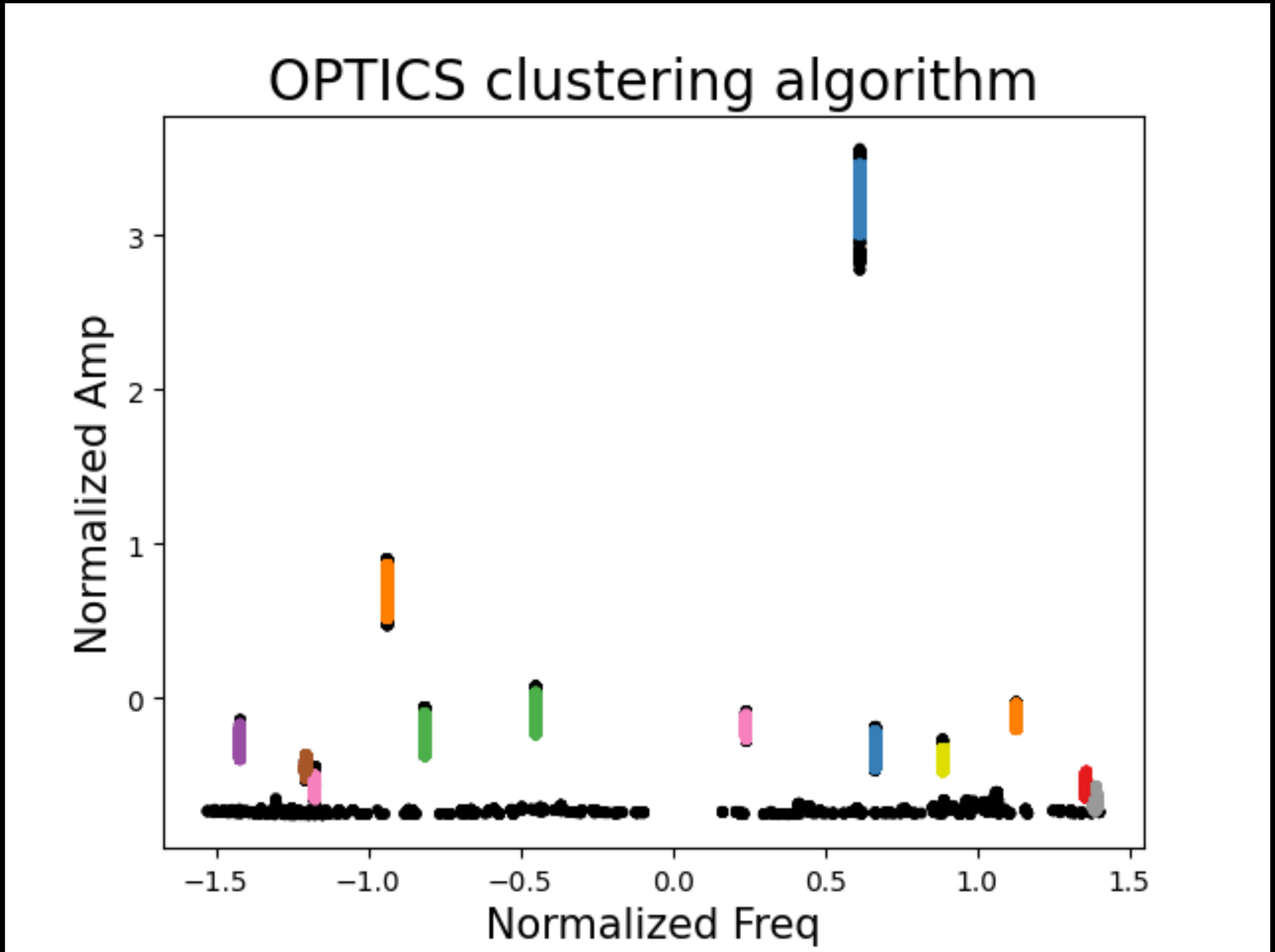
Sky Localization Ellipses

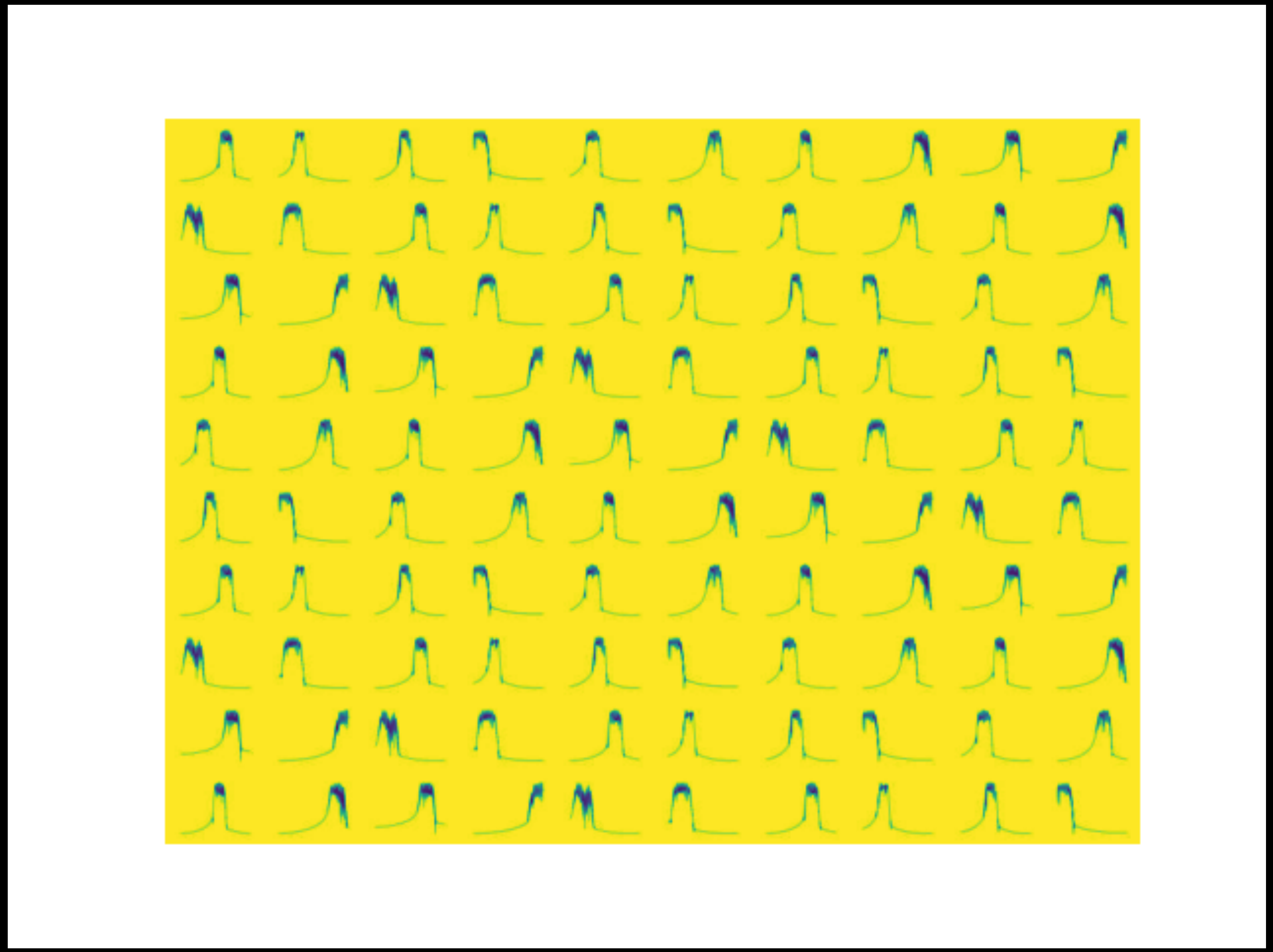
Download all examples in Python source code: examples_ucb_python.zip

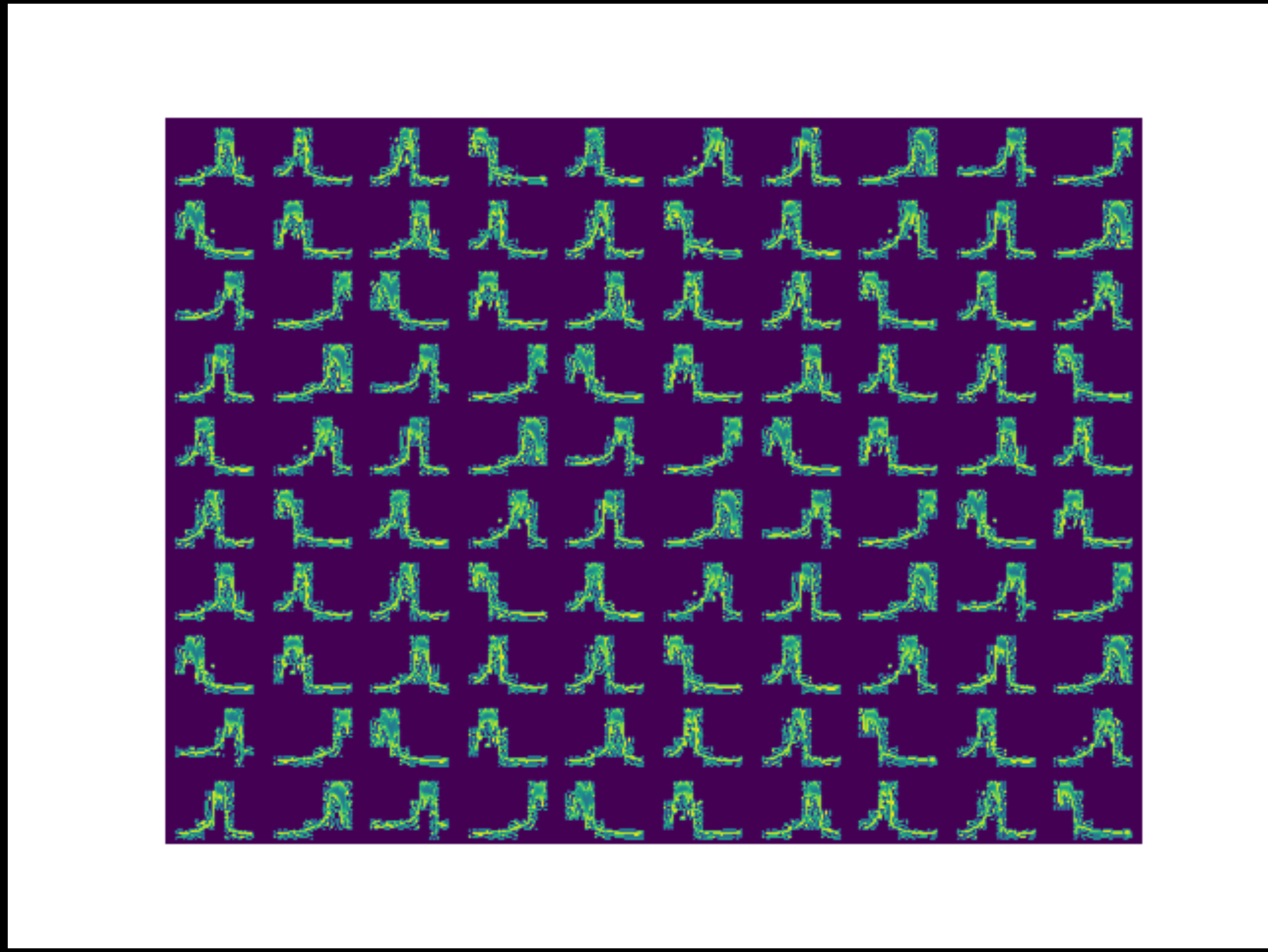
Download all examples in Jupyter notebooks: examples_ucb_jupyter.zip

For the MCMC code see https://github.com/tlittenberg/ldasoft.git

Building catalogs with ML
L2 -> L3







Merci

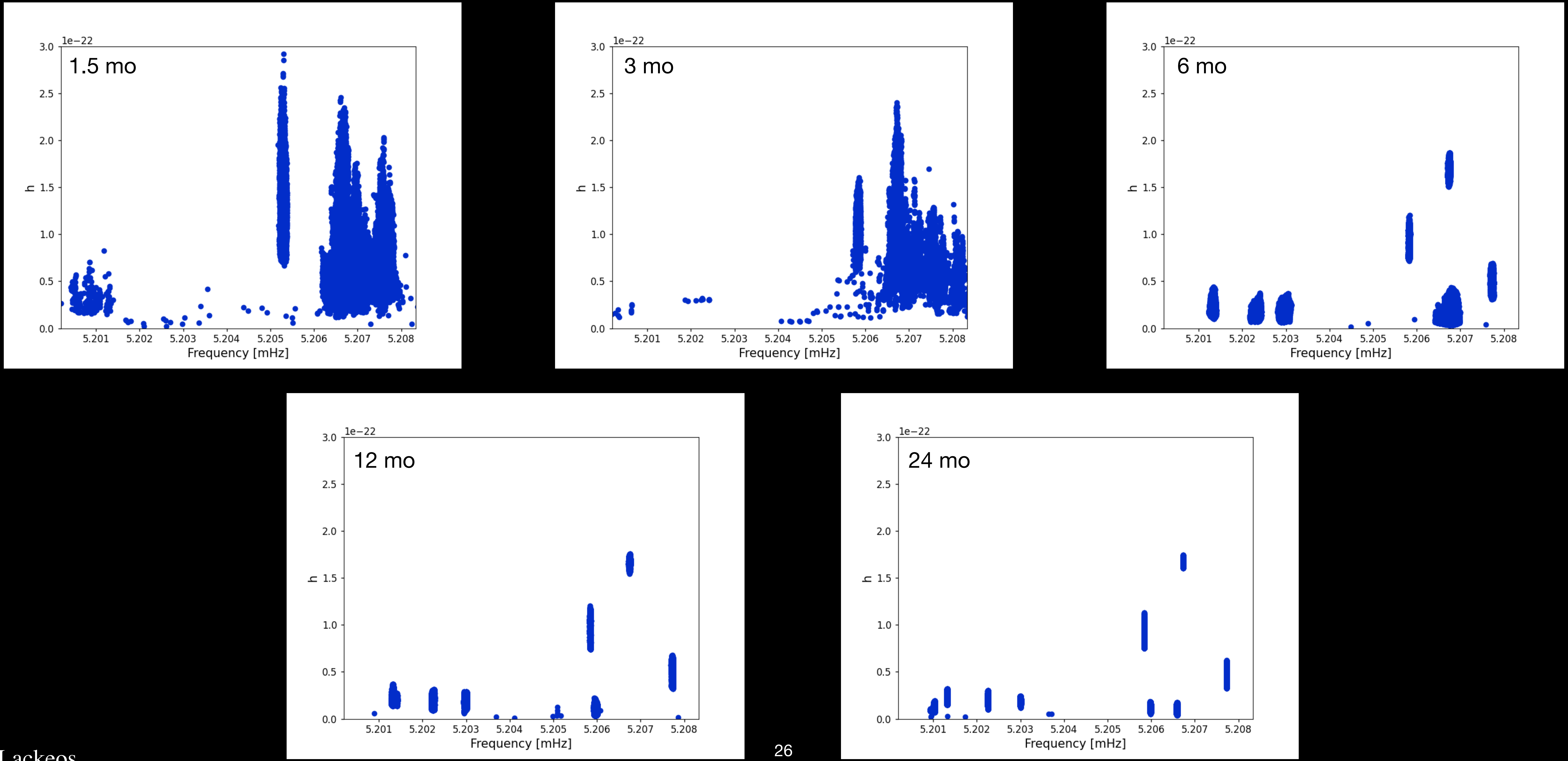
Parametrisation of the problem

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

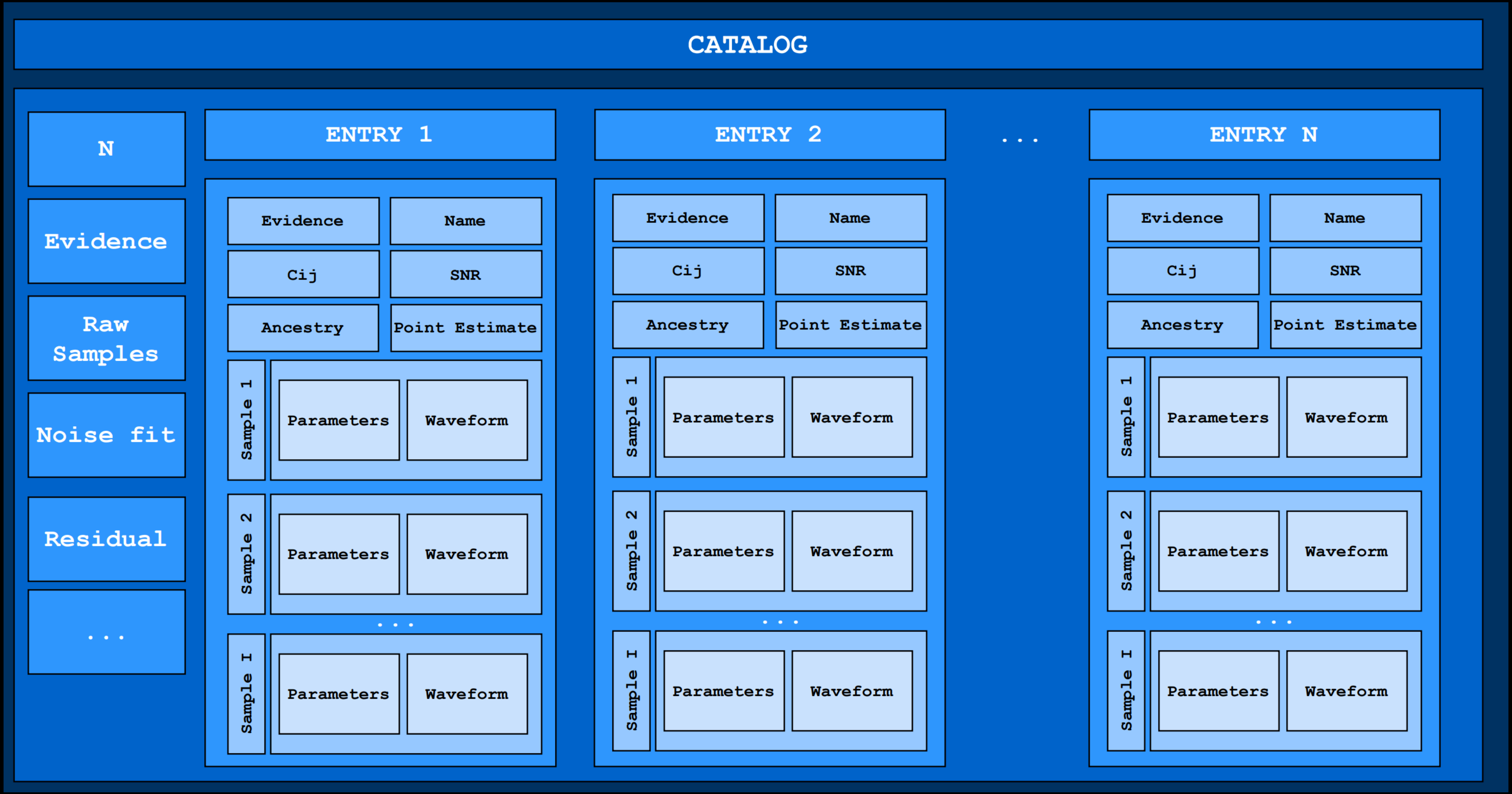
LDC - Challenge 1: *Radler*

Raw samples



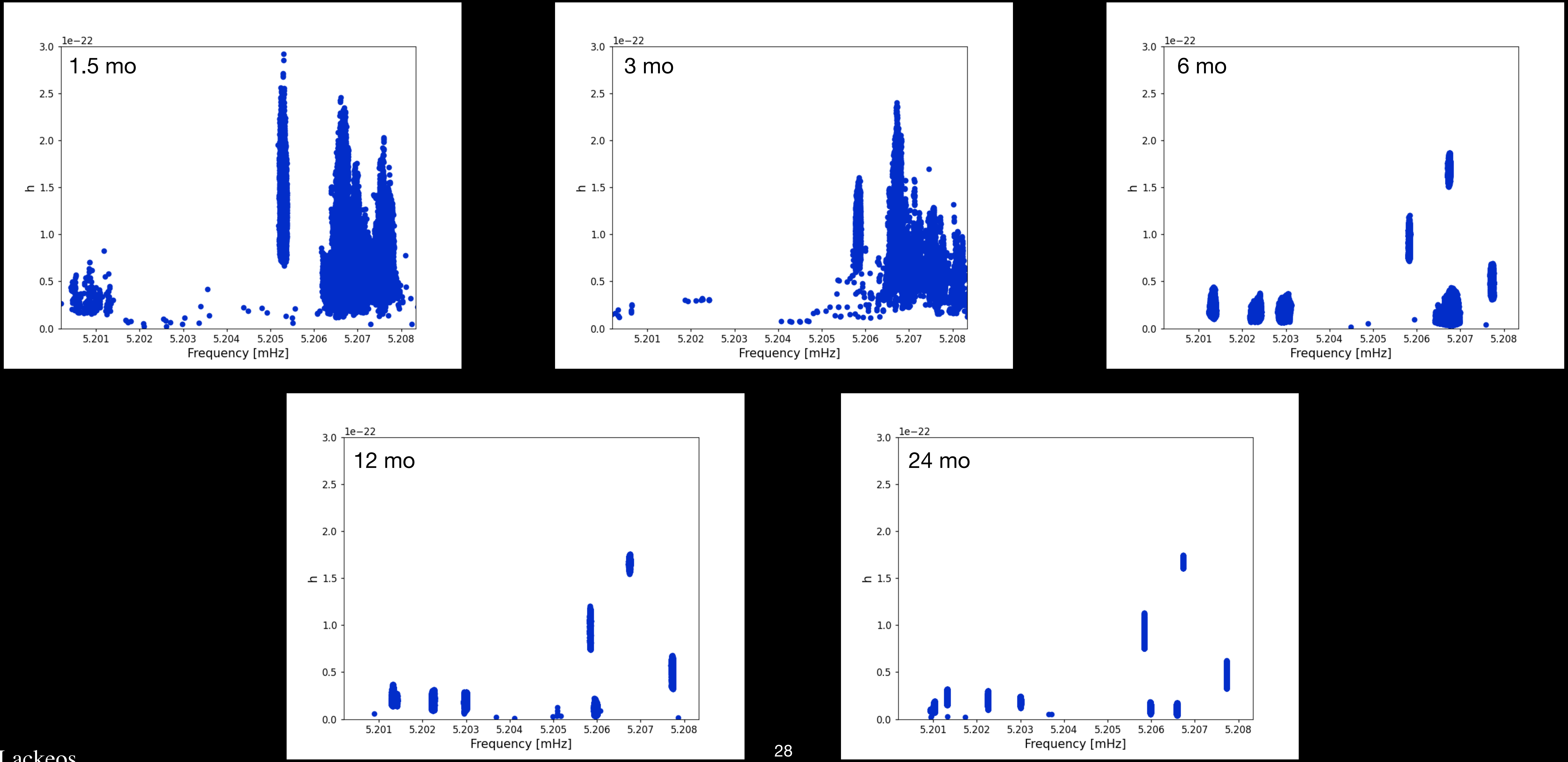
Kristen Lackeos

Catalog production



LDC - Challenge 1: *Radler*

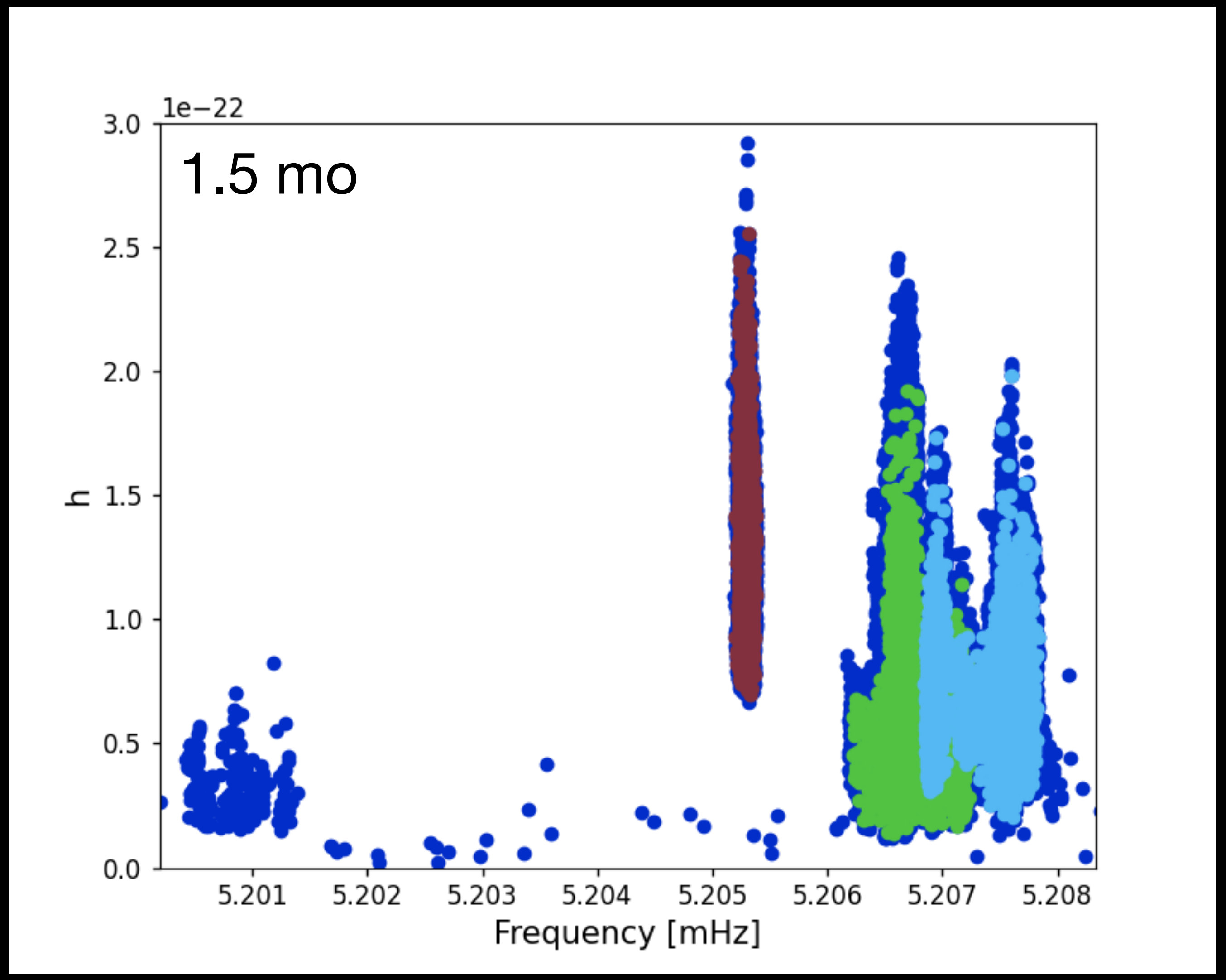
Raw samples

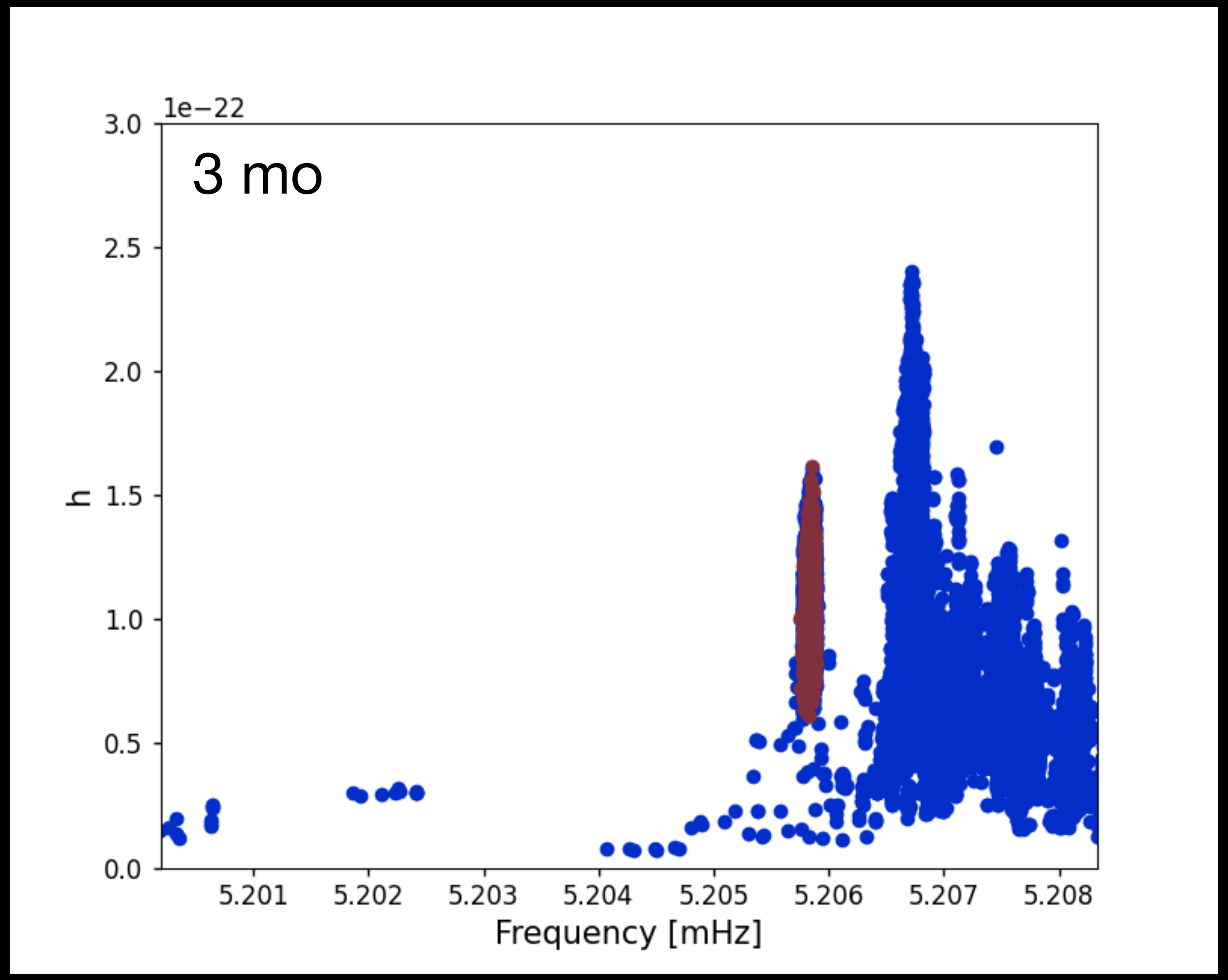


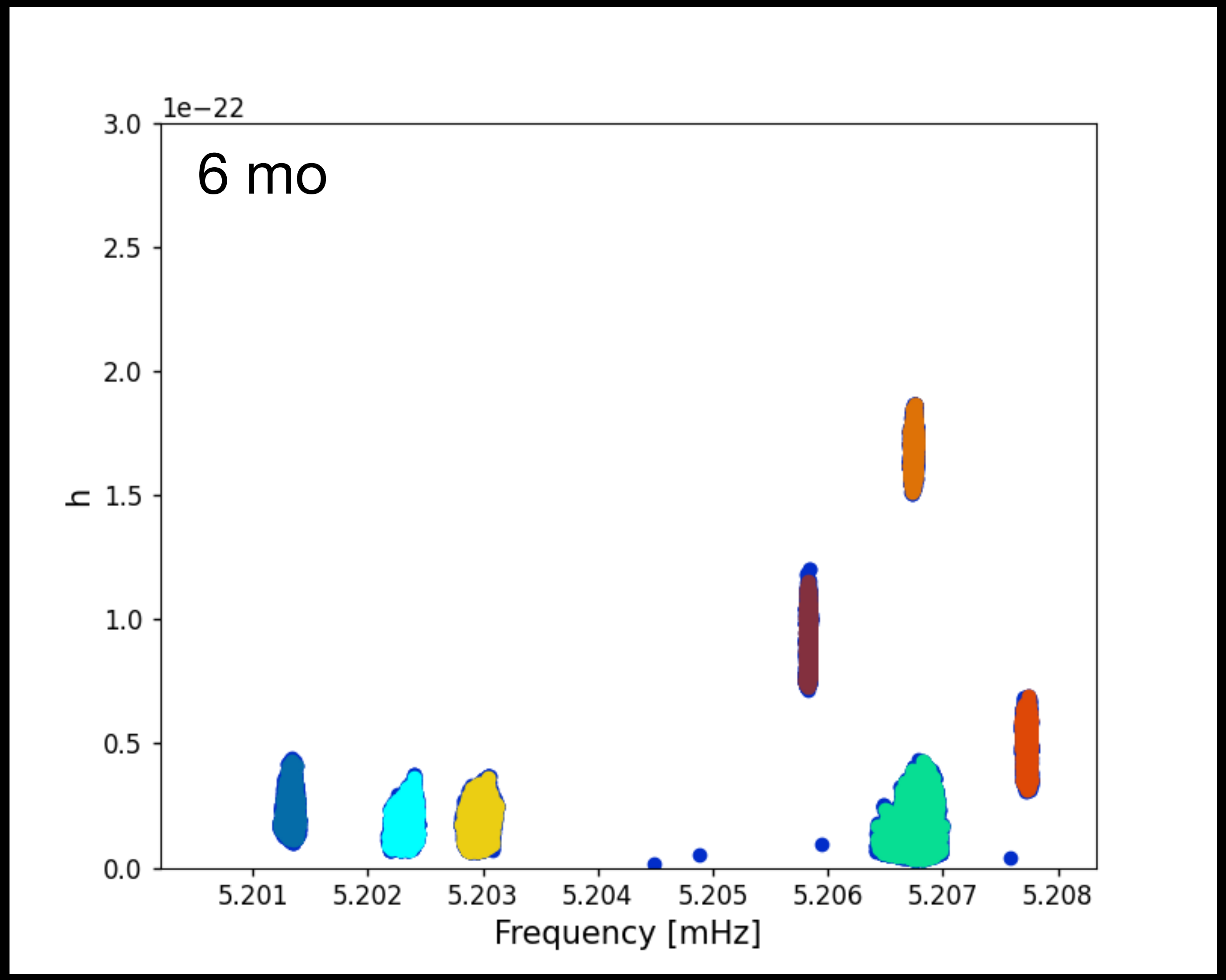
Kristen Lackeos

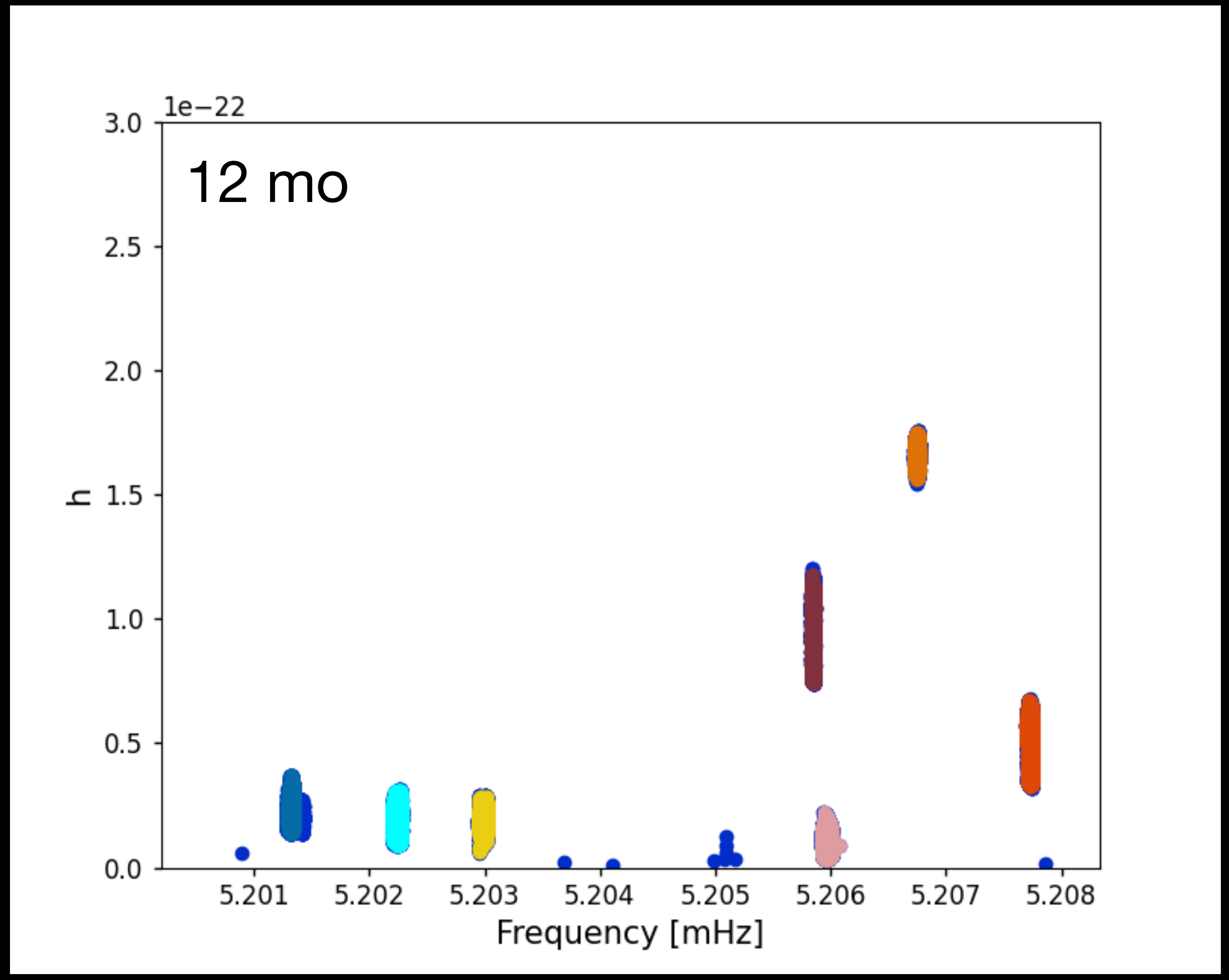
LDC - Challenge 1: *Radler*

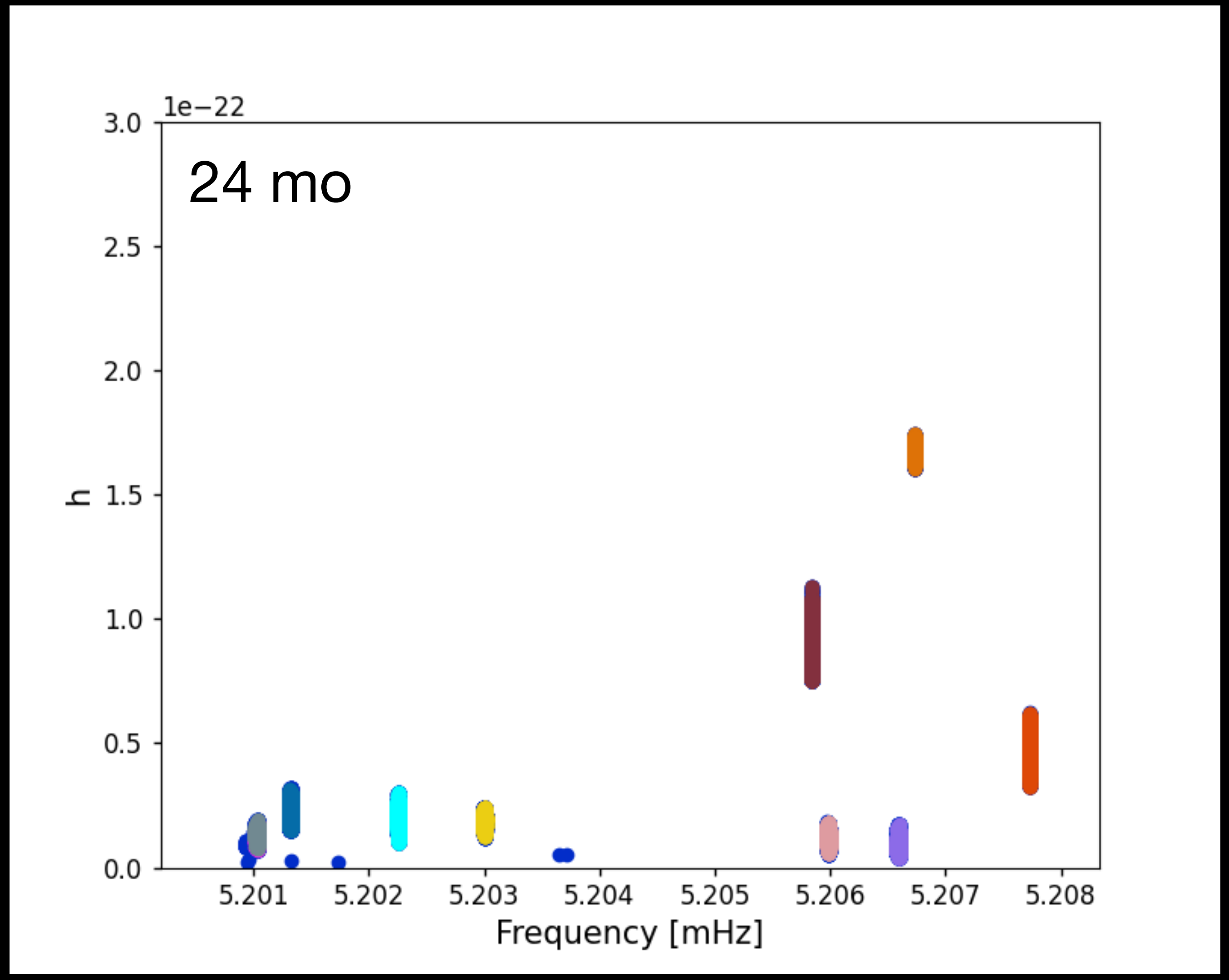
Catalog detections





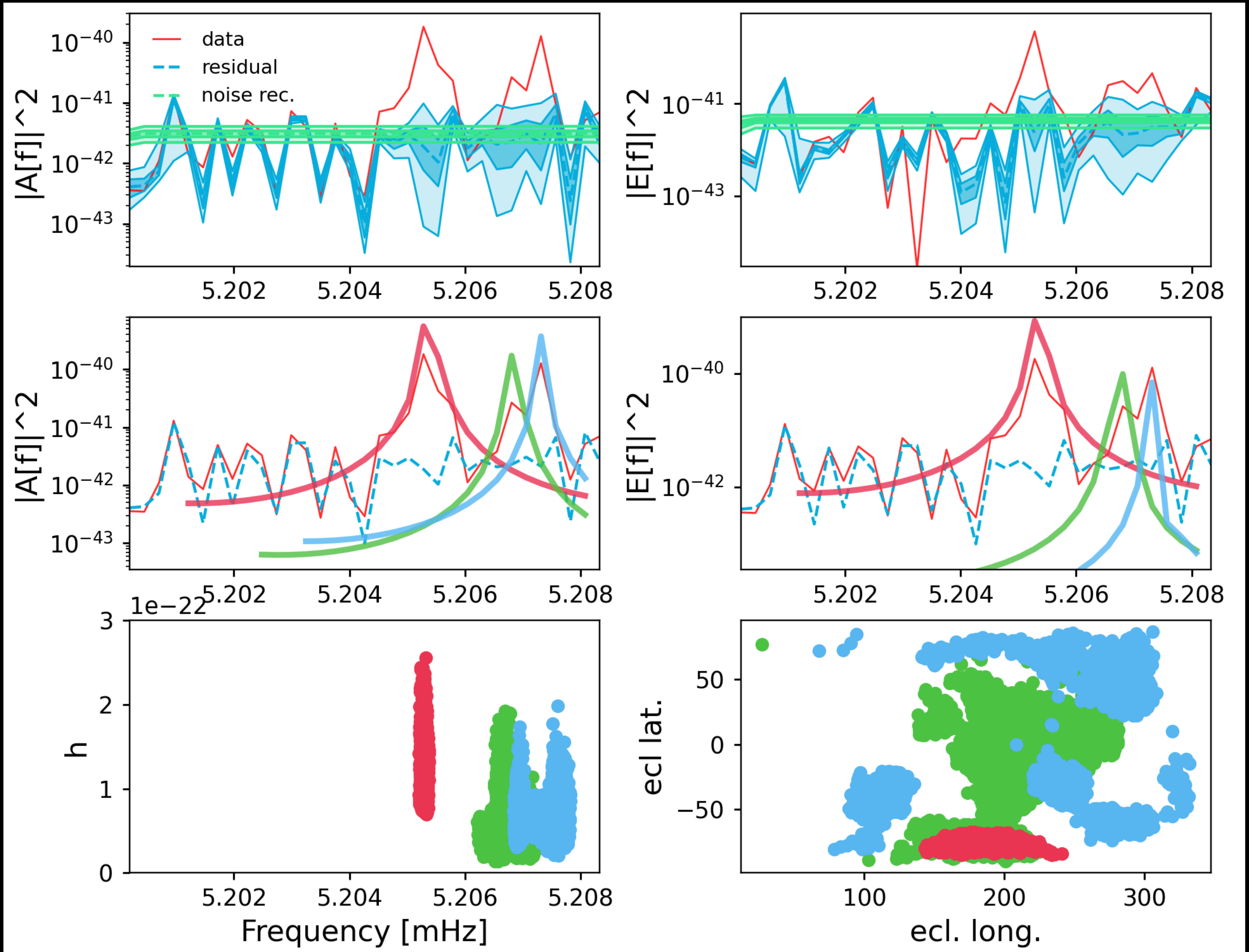






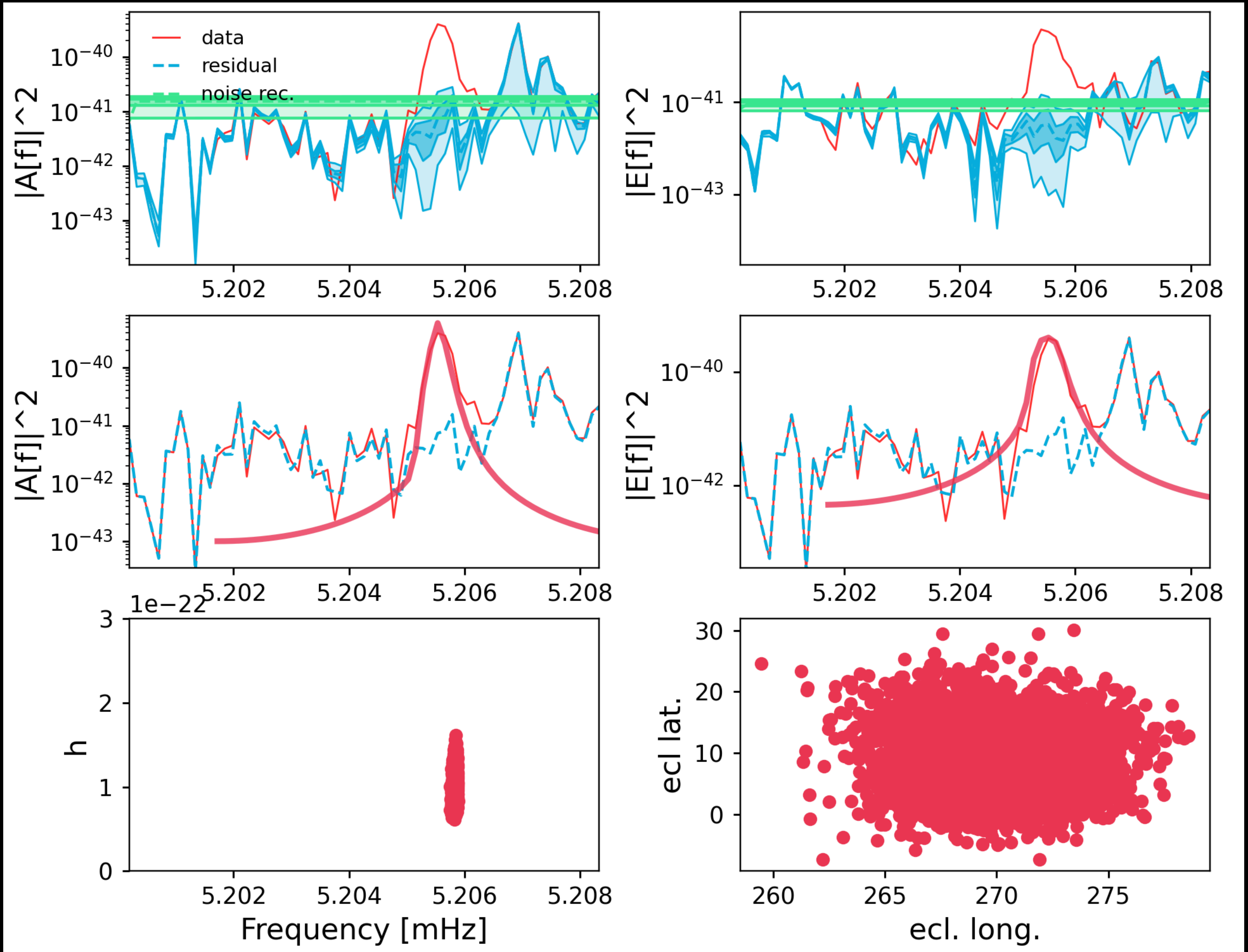
LDC - Challenge 1: *Radler*

1 month



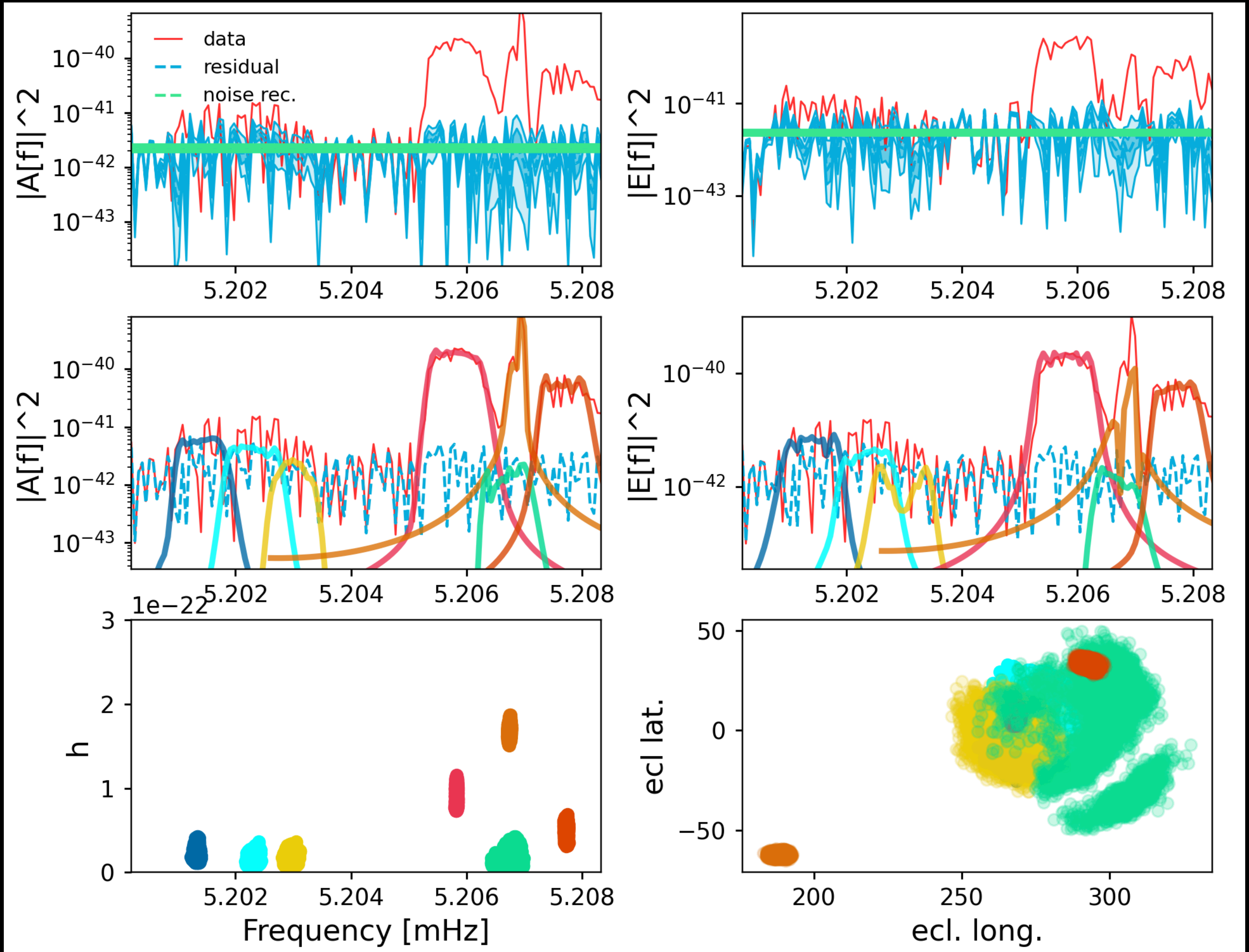
LDC - Challenge 1: *Radler*

3 months



LDC - Challenge 1: *Radler*

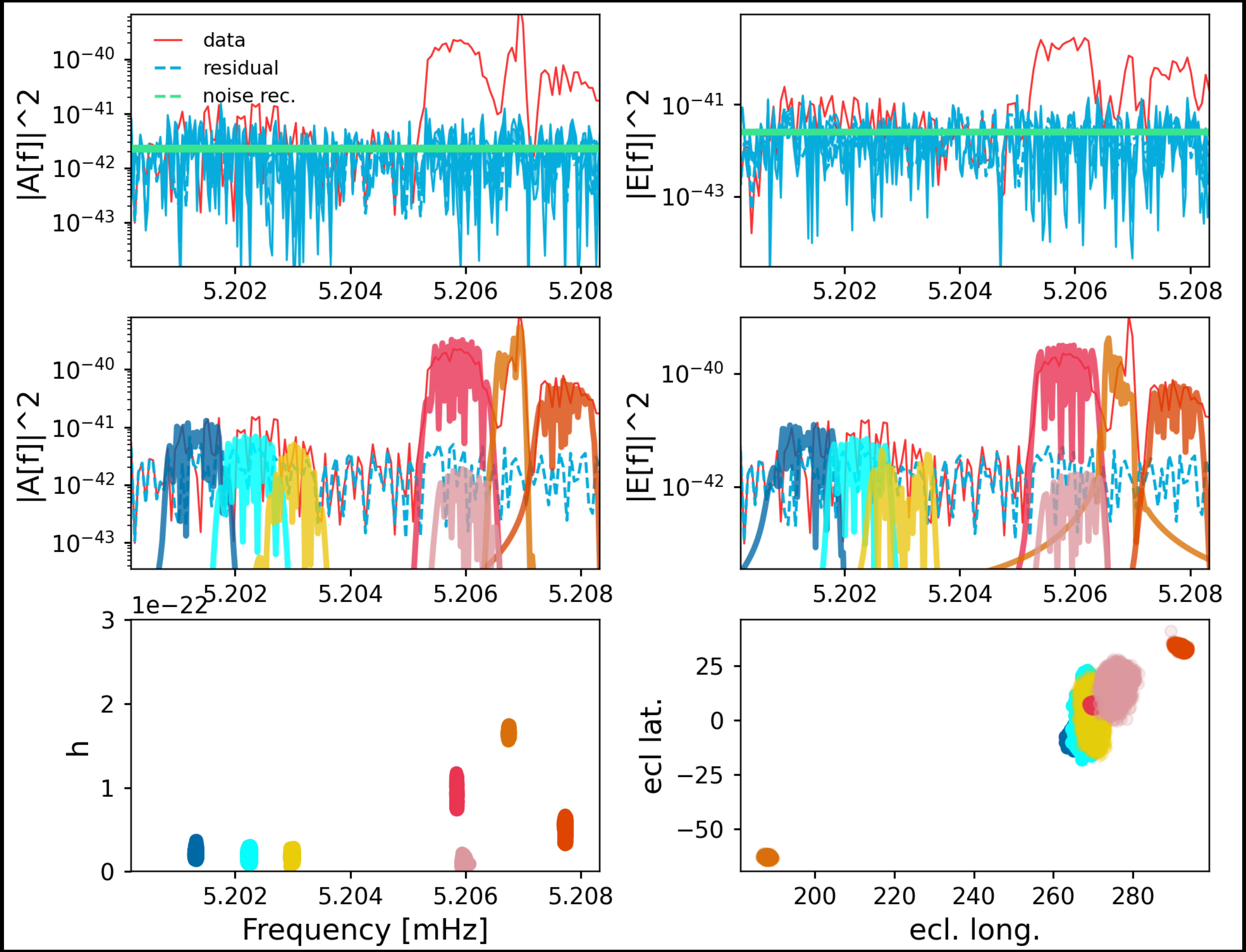
6 months



Kristen Lackeos

LDC - Challenge 1:
Radler

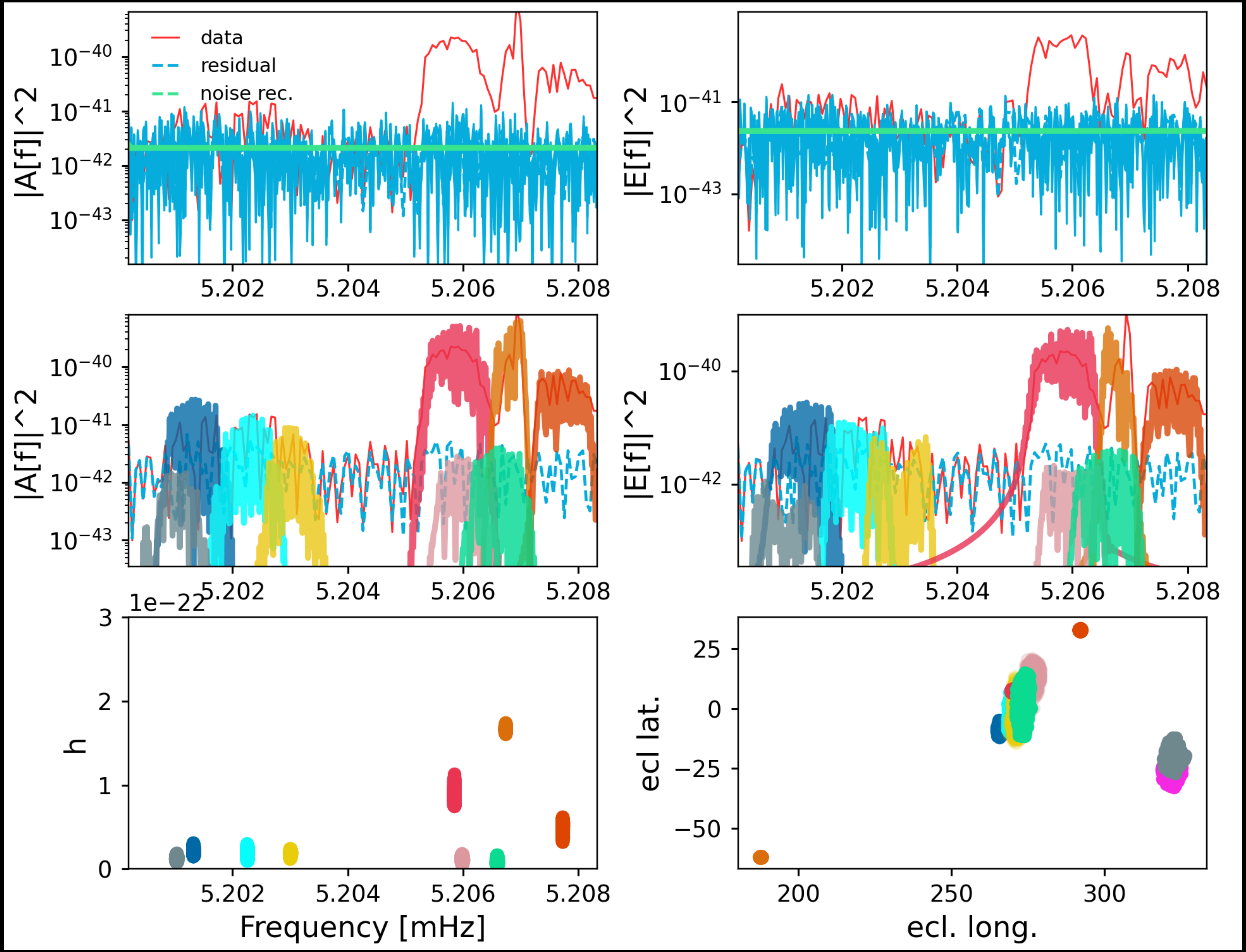
12 months



Kristen Lackeos

LDC - Challenge 1:
Radler

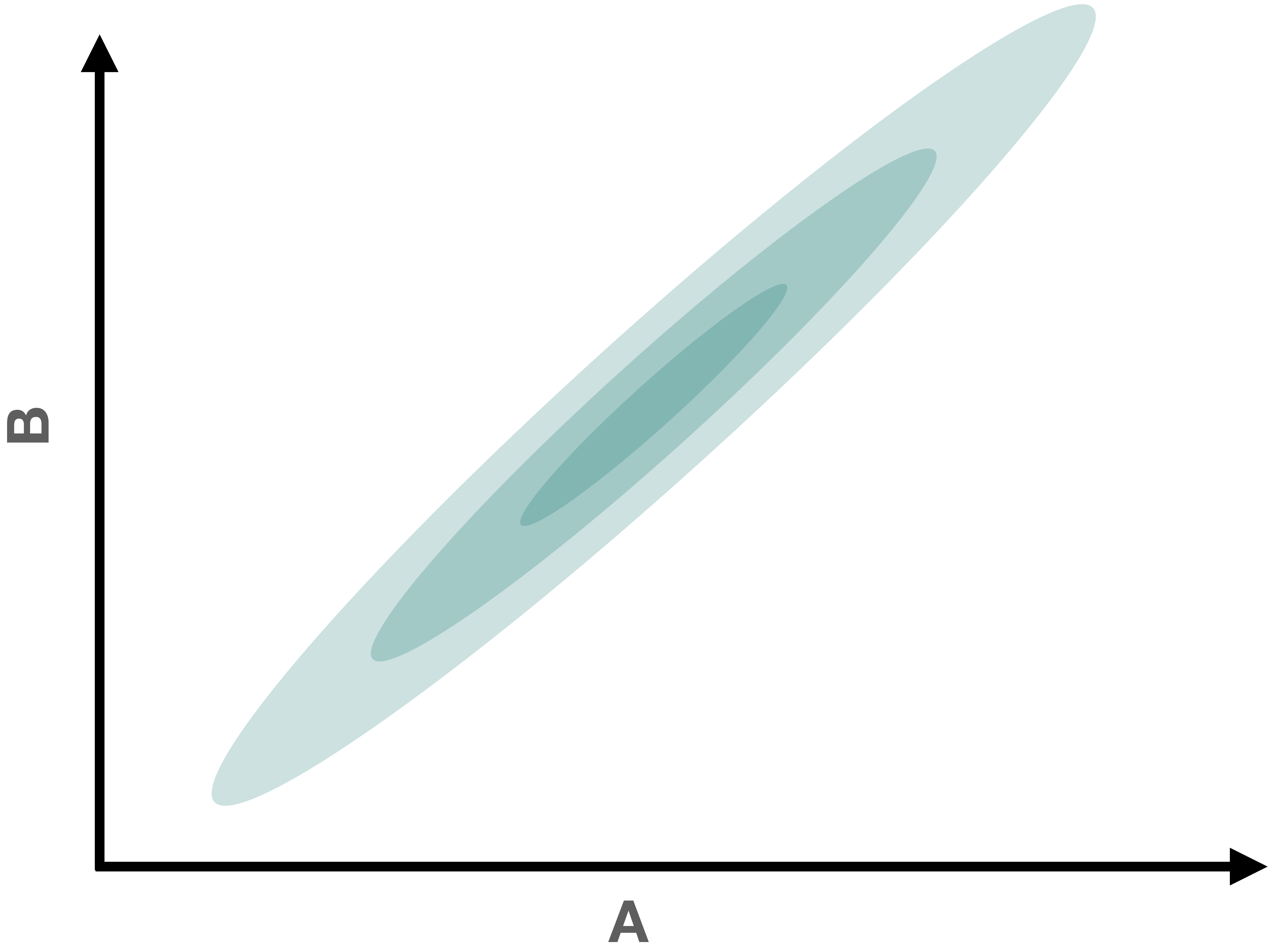
24 months



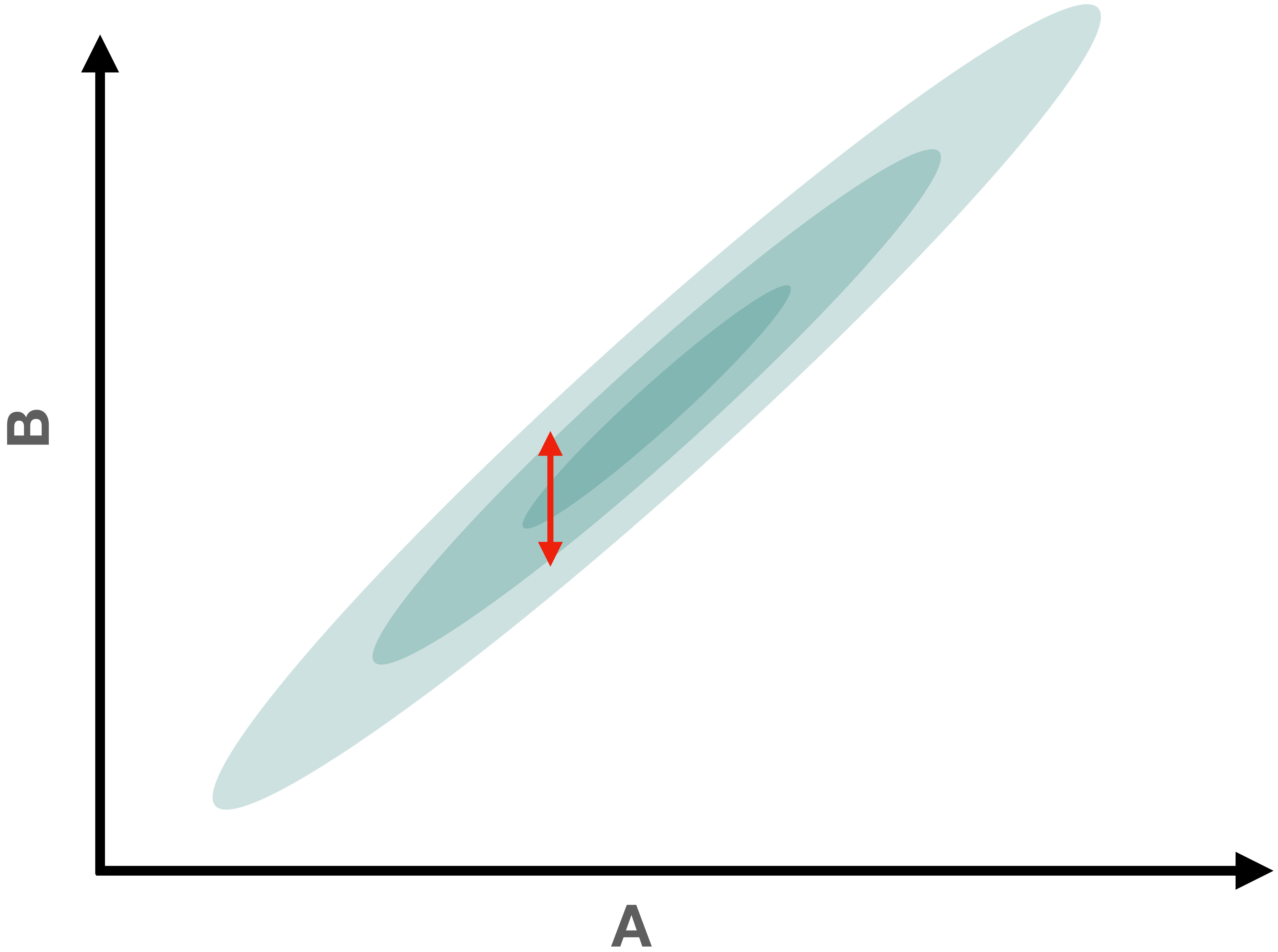
Kristen Lackeos

BLOCKED GIBBS SAMPLING

What is Gibbs Sampling?

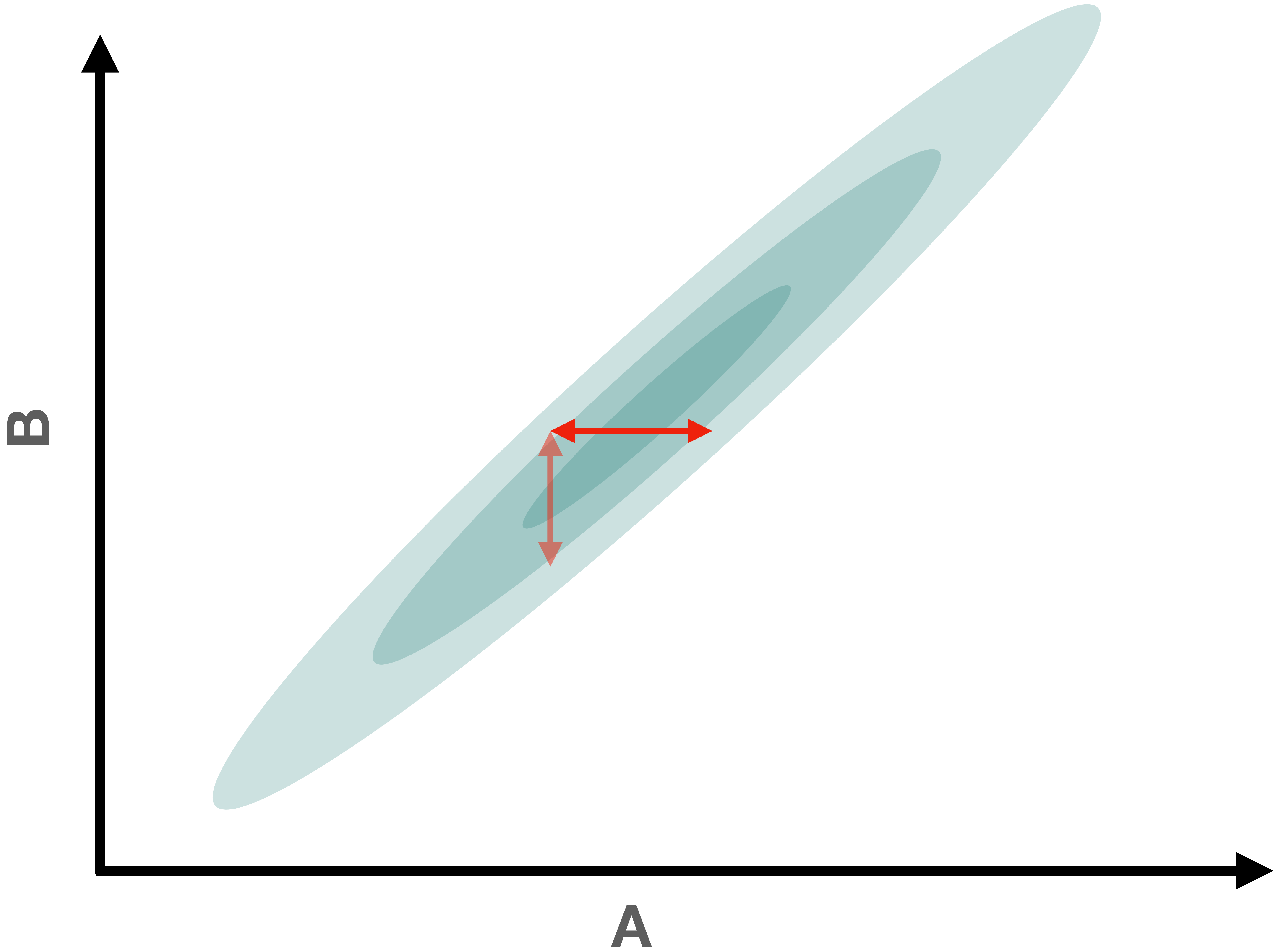


What is Gibbs Sampling?



I. Hold Block A fixed and sample over B

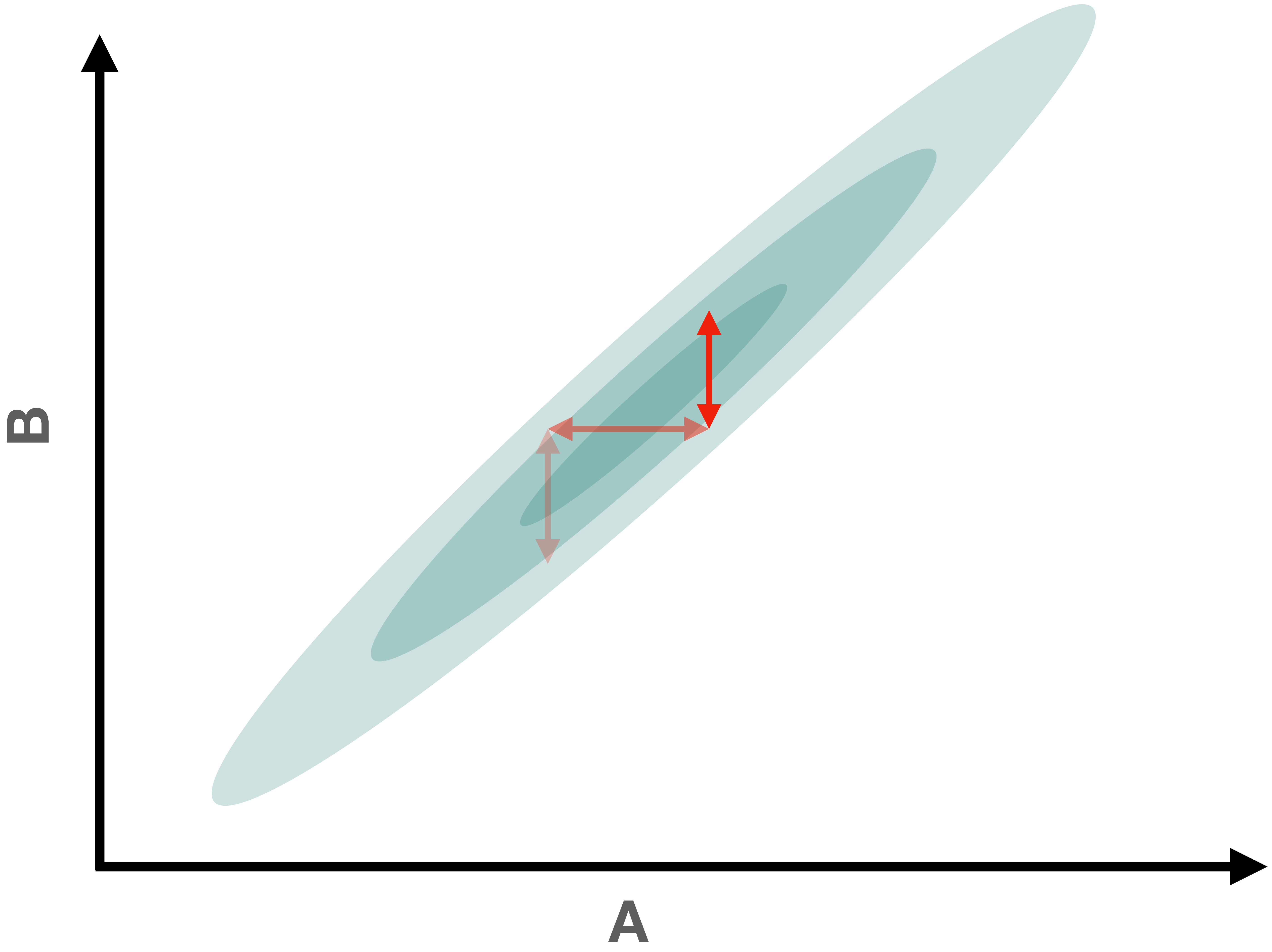
What is Gibbs Sampling?



I. Hold Block A fixed and sample over B

II. Hold Block B fixed and sample over A.

What is Gibbs Sampling?

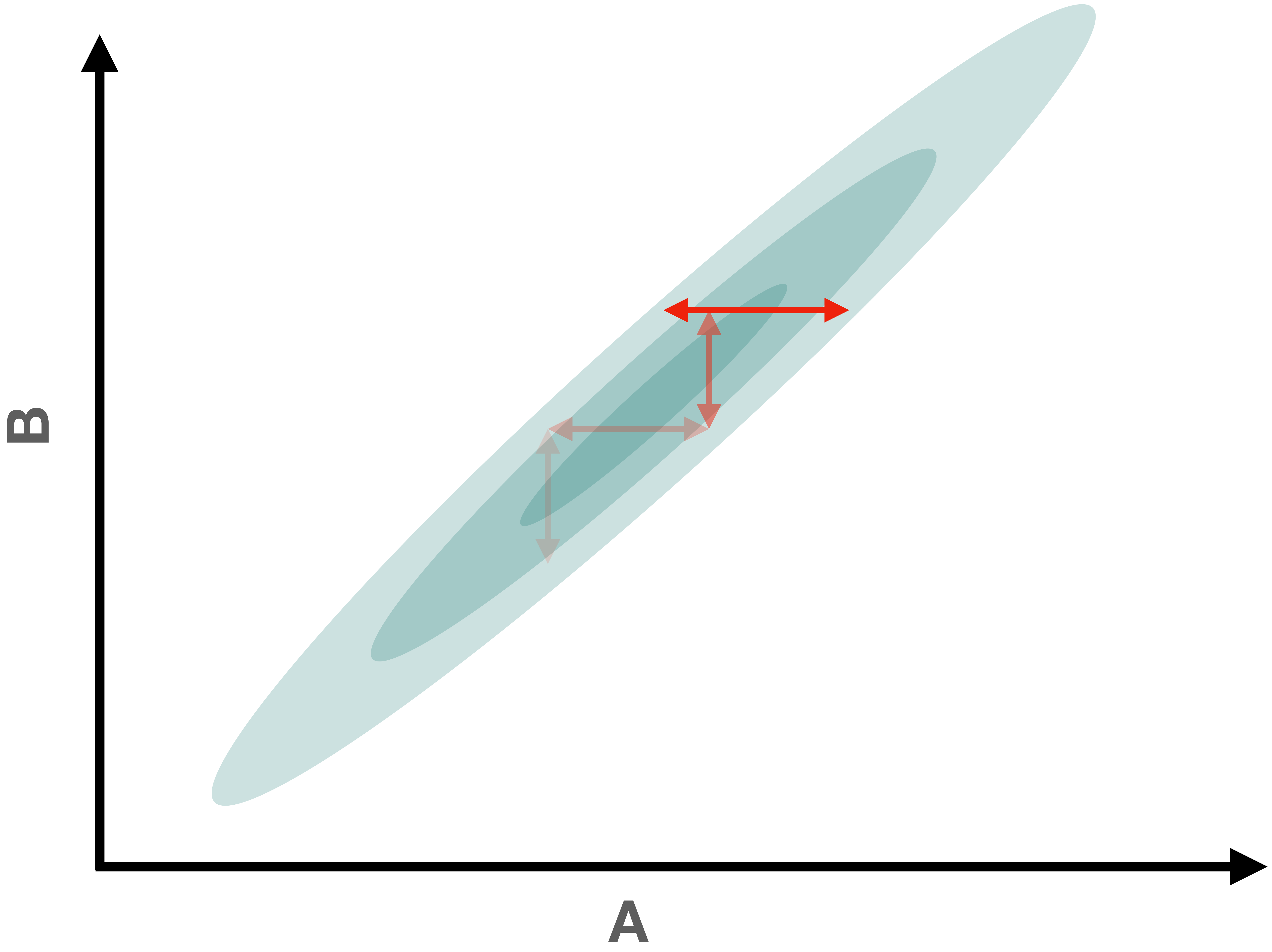


I. Hold Block A fixed and sample over B

II. Hold Block B fixed and sample over A.

III. Repeat

What is Gibbs Sampling?



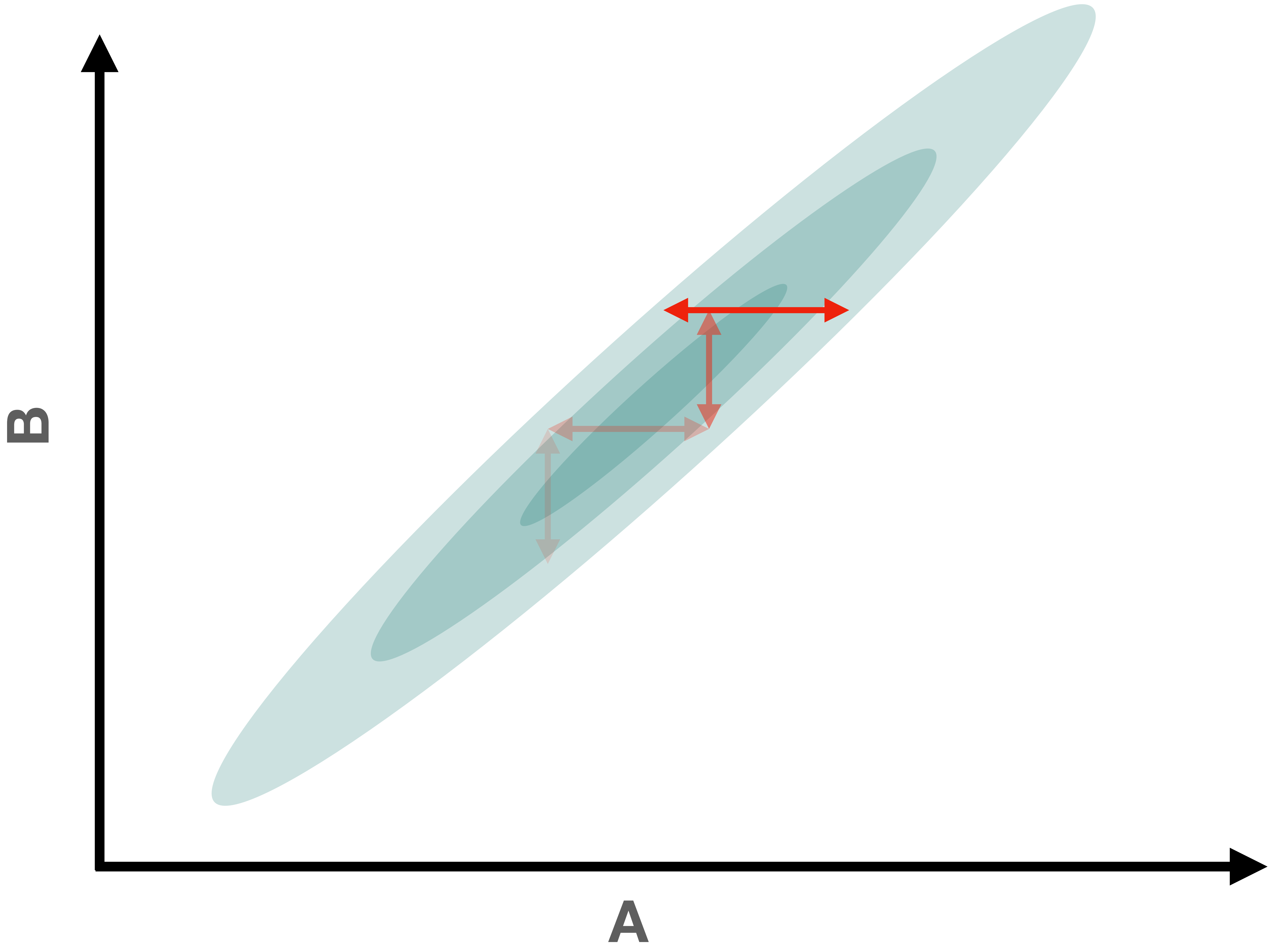
I. Hold Block A fixed and sample over B

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

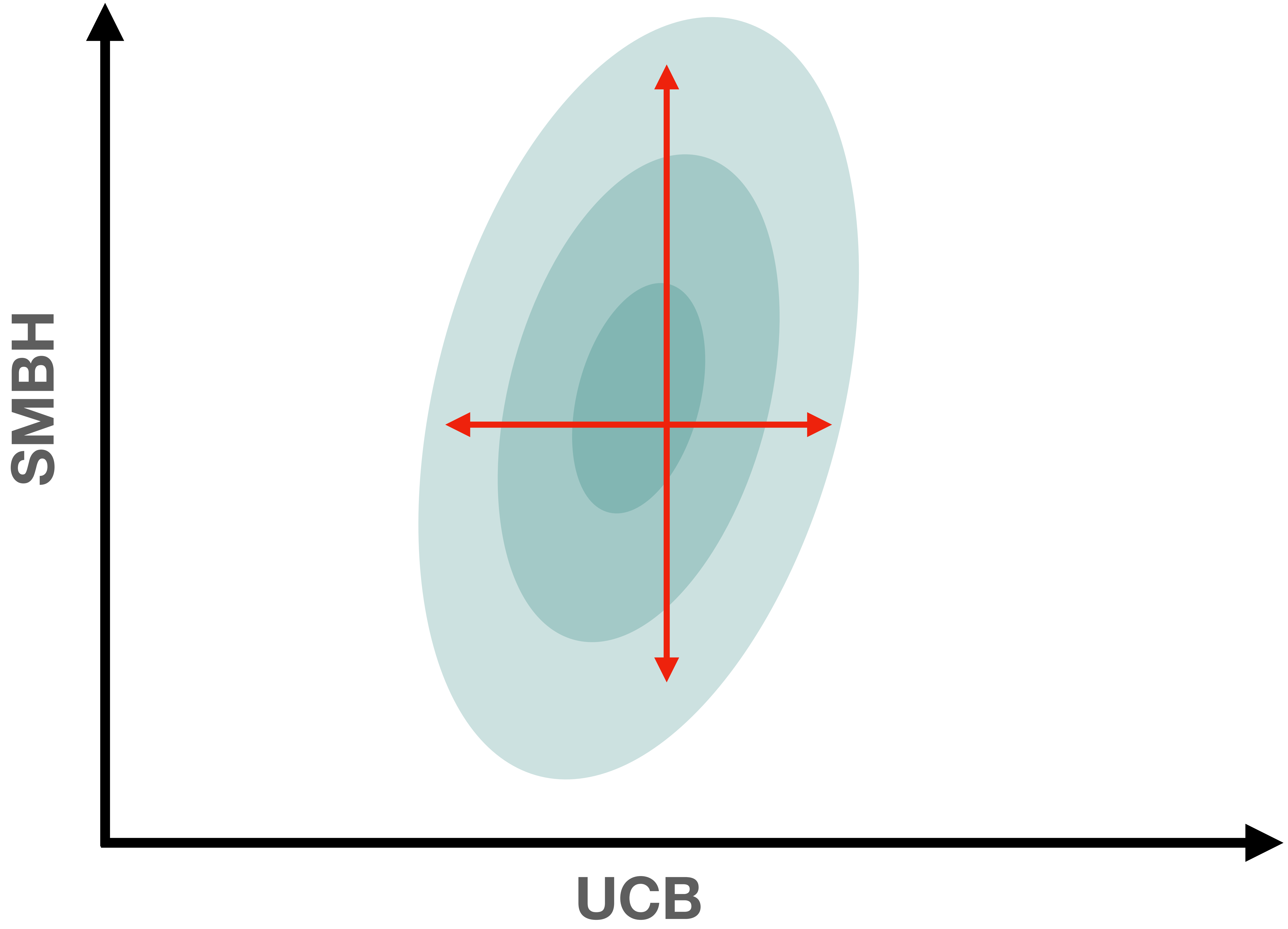
II. Hold Block B fixed and sample over A.

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