ICRAR

International Centre for Radio Astronomy Research

Gaussian process models for identifying variables and transients in large astronomical surveys

Shih Ching Fu

Co-supervisors:

Dr Arash Bahramian, Dr Aloke Phatak

Associate Supervisors:

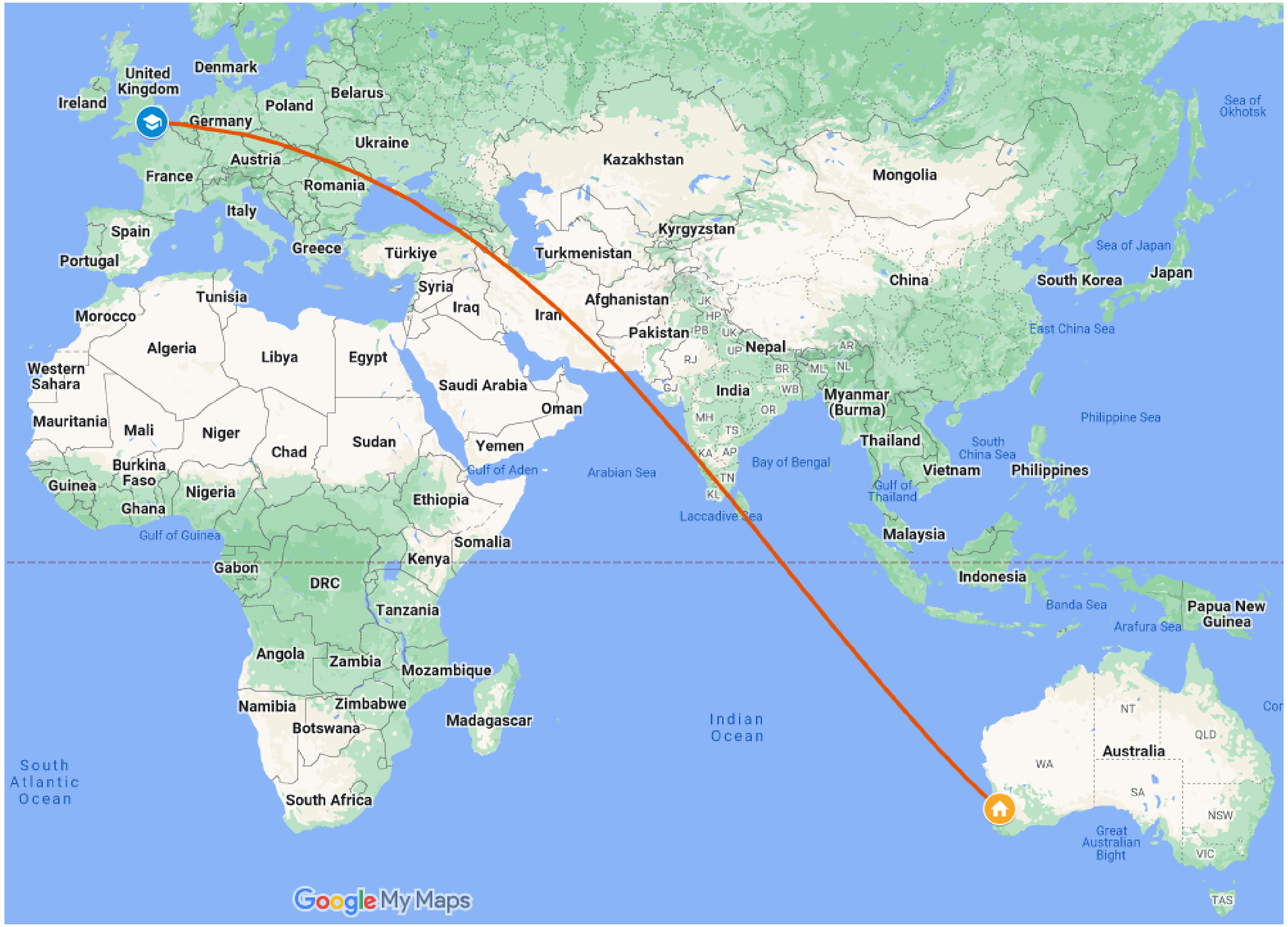
Dr James Miller-Jones, Dr Suman Rakshit

Curtin University

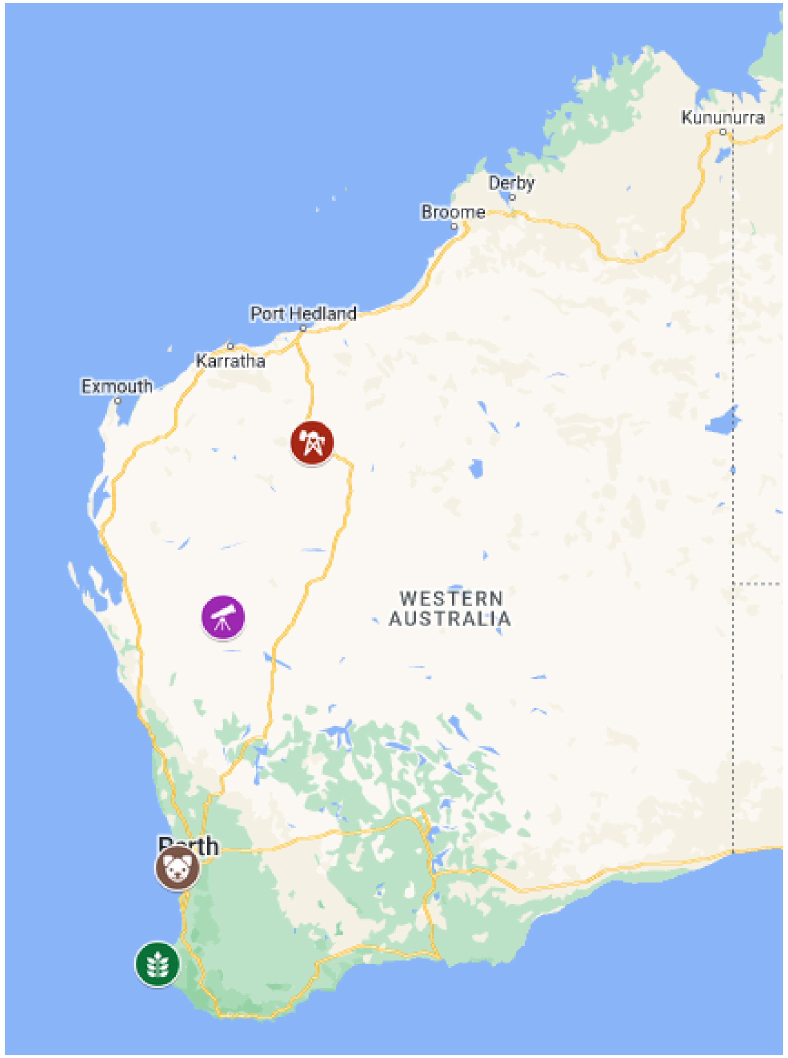
THE UNIVERSITY OF WESTERN AUSTRALIA

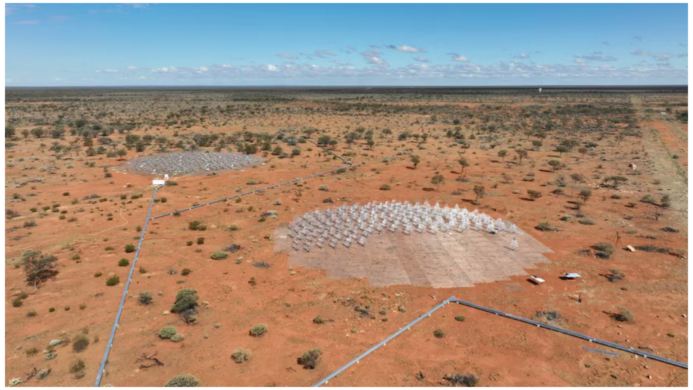
Statistical Society of Australia

Curtin to Imperial ≈ 14 500 km, 18h flight time



Western Australia ≈ 2.6M km^2





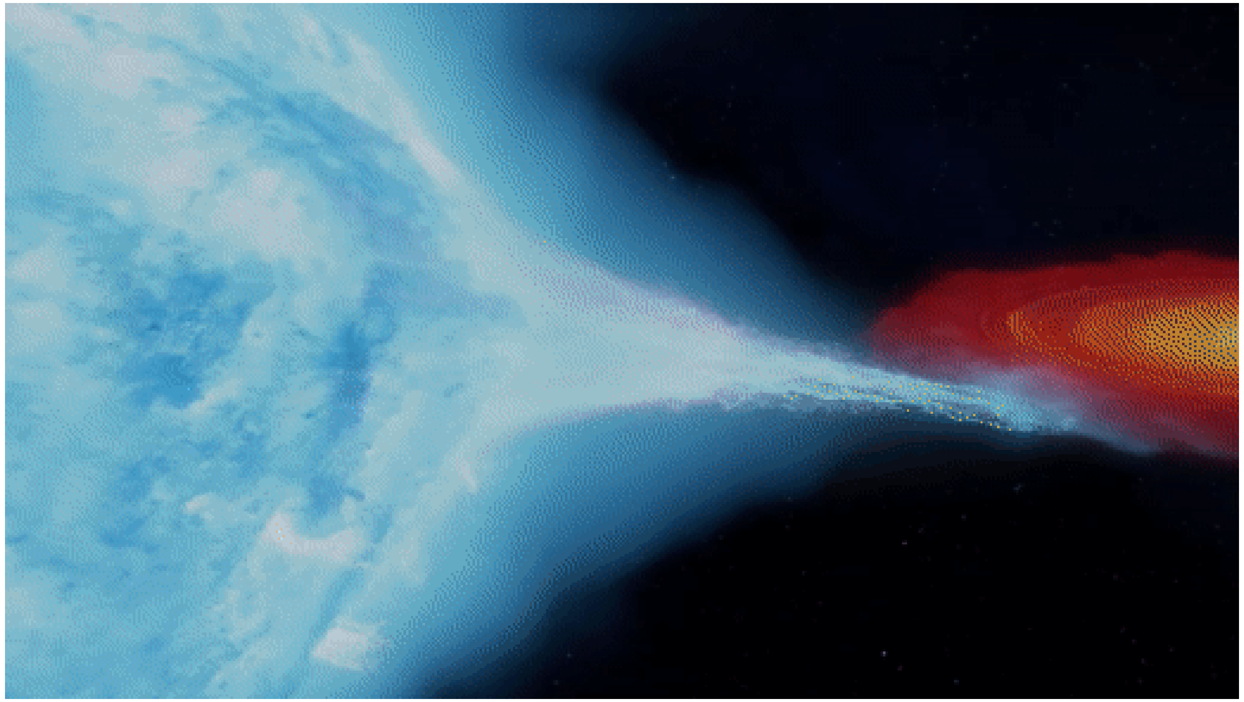
Curtin Institute of Radio Astronomy (CIRA)

- Formed in 2017 (1 of 2 ICRAR nodes)
- 27 academic, 28 technical, 36 HDRs
- Science
  - ***Accretion, Jets, and Slow Transients***
  - Epoch of Reionisation science
  - Extragalactic Radio Astronomy
  - Pulsars and Transients
- Engineering
- Murchison Widefield Array (MWA)
- Australian SKA Pathfinder (ASKAP)

Twinkle twinkle…

A *transient* is an astrophysical phenomenon whose brightness changes "meaningfully" over observable time.

- Supernovae
- Variable stars, e.g., pulsating, eclipsing binaries.
- Gamma-ray bursts (GRBs)
- Fast radio bursts (FRBs)
- Transiting planets
- Active galactic nuclei (AGN)
- Accreting blackholes
- and lots more…

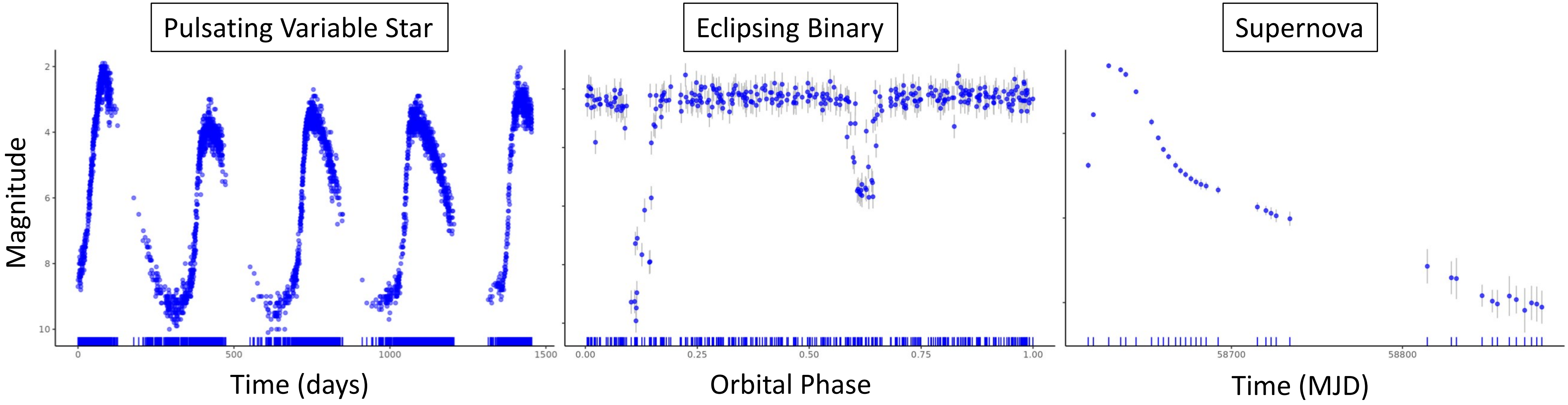


Artist's impression of the Cygnus X-1 system. Credit: ICRAR

Light Curves

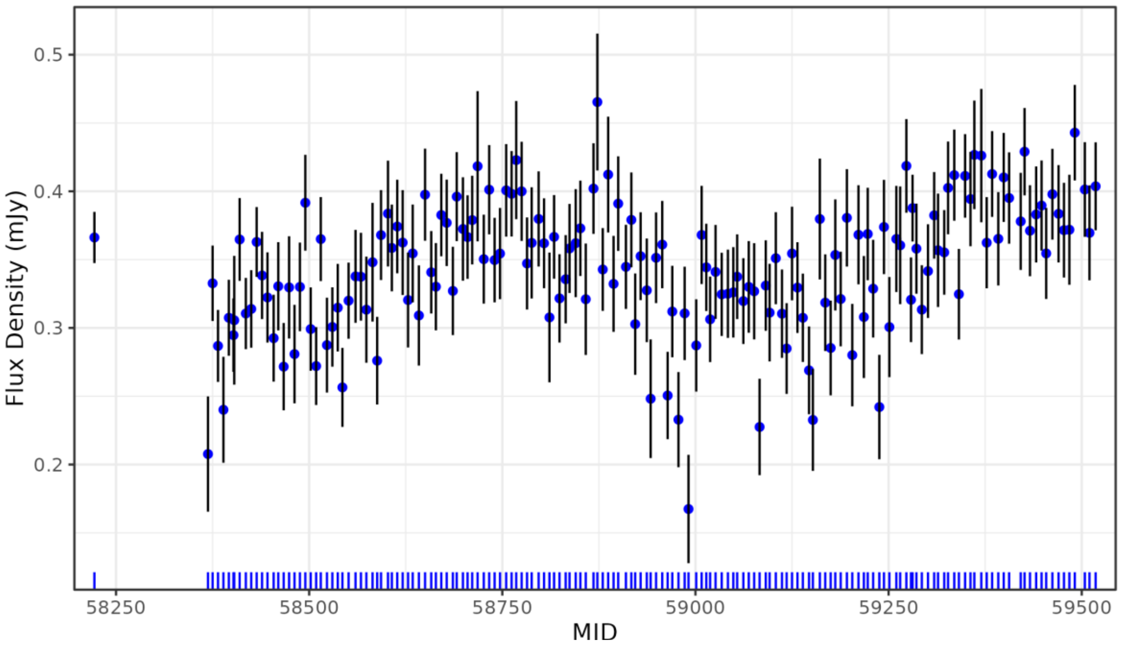
Light curves are time series describing the brightness of a source over time.

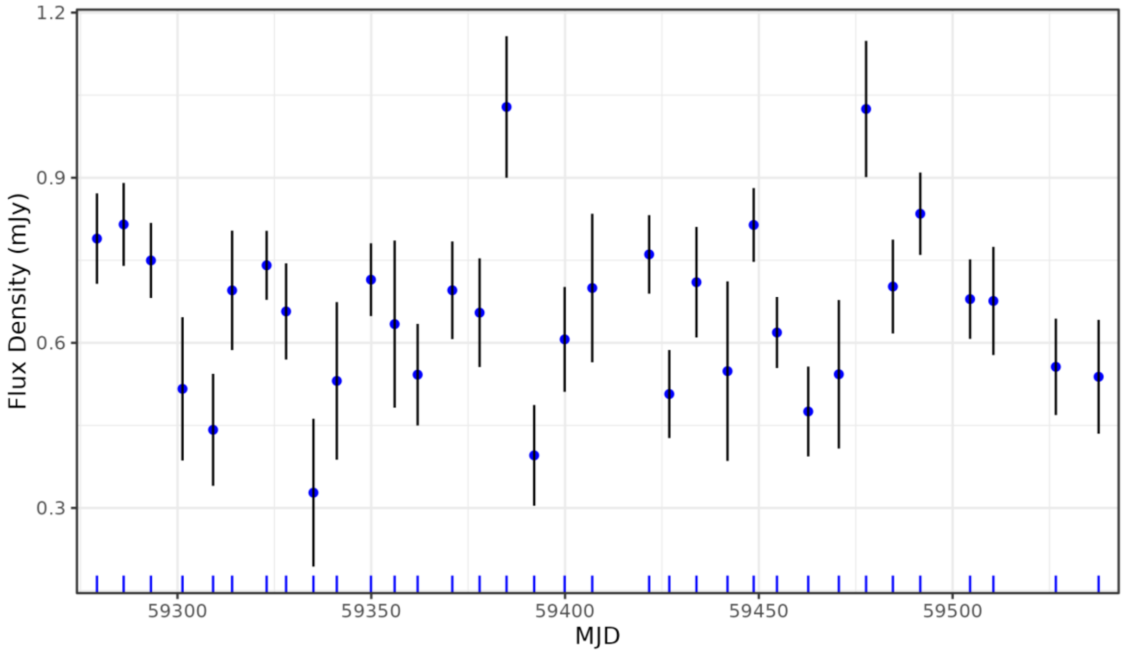
- The shape of a light curve can reveal the type of object or event.
- Variability in brightness can reveal information about the processes underlying the observed phenomenon.

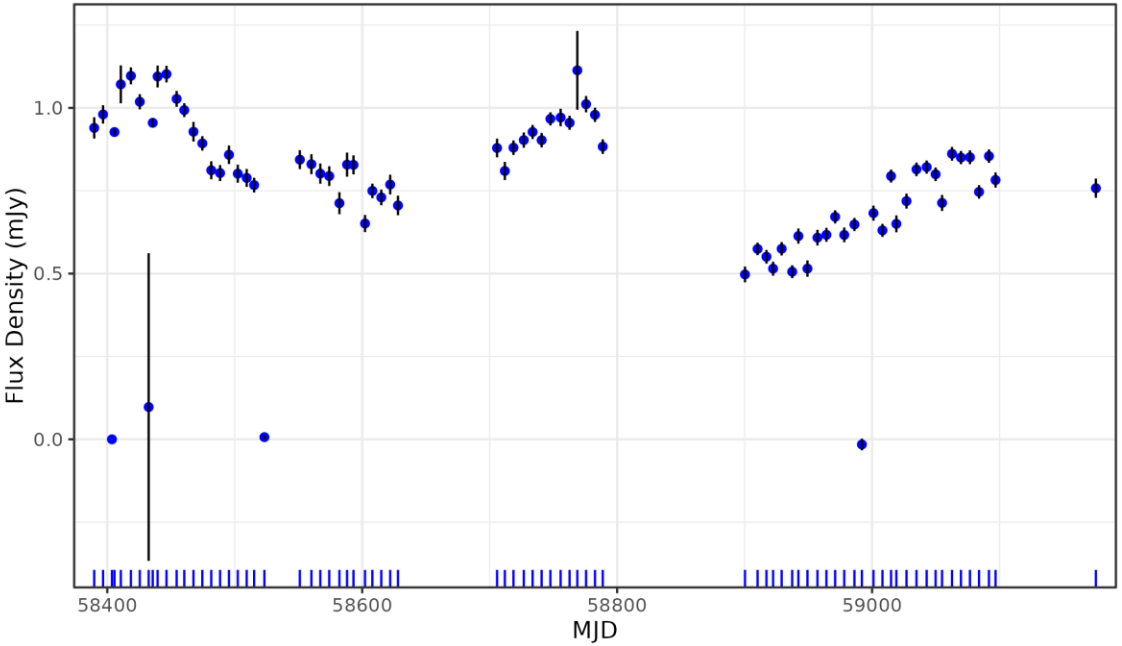


Inconsistent Data

- Different cadences
- Sparse observations
- Irregular sampling
- Varying noise levels



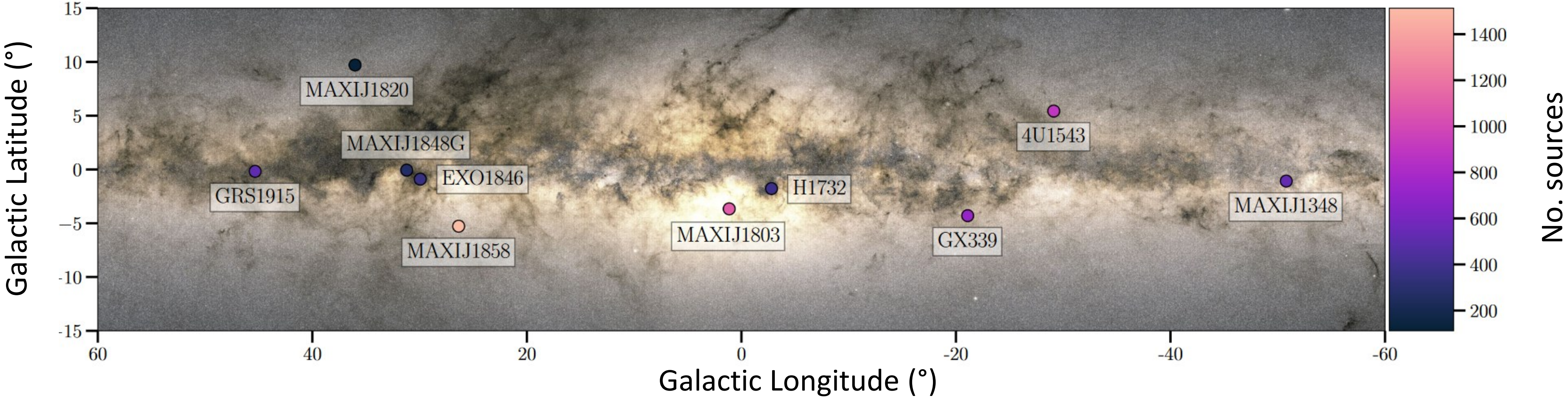


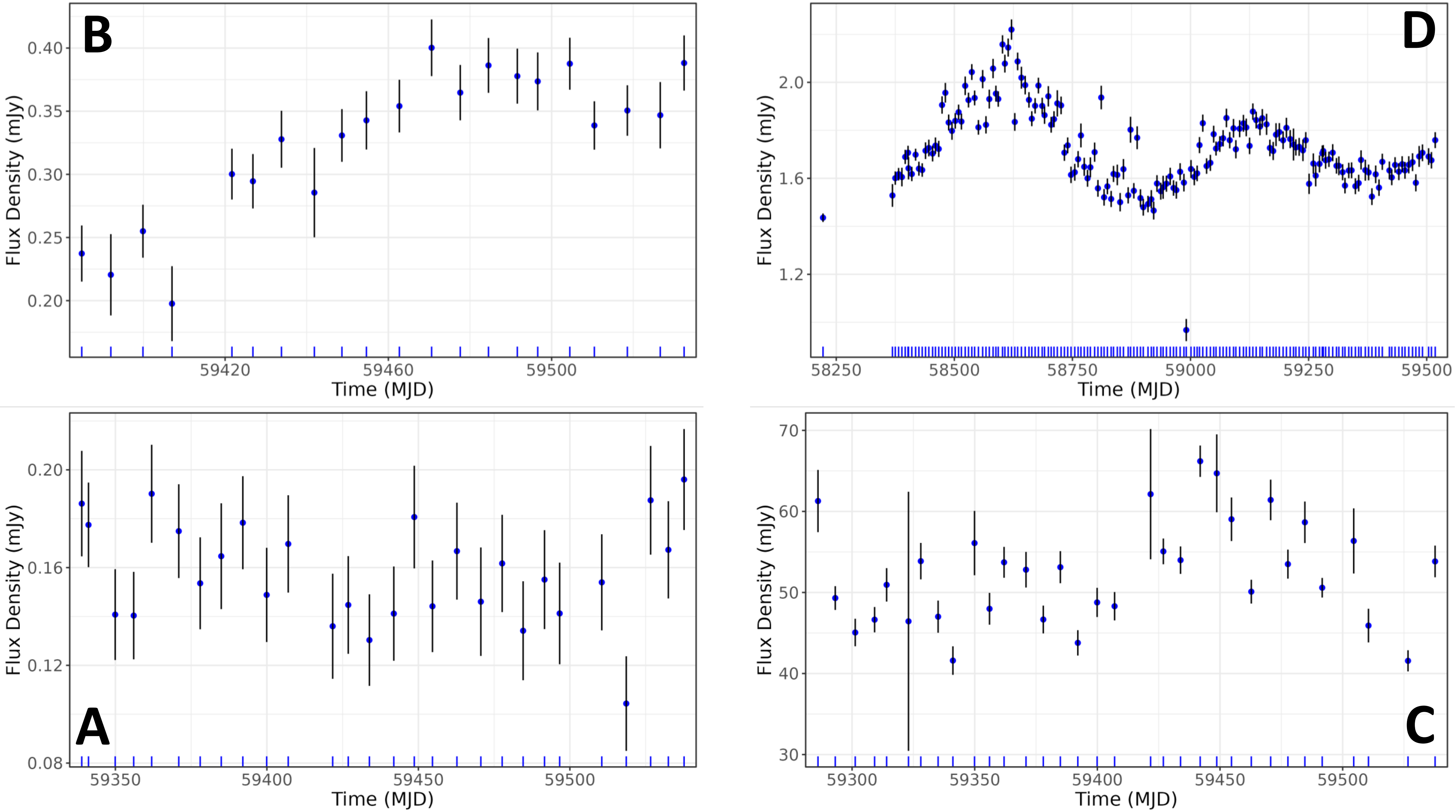


ThunderKAT Survey

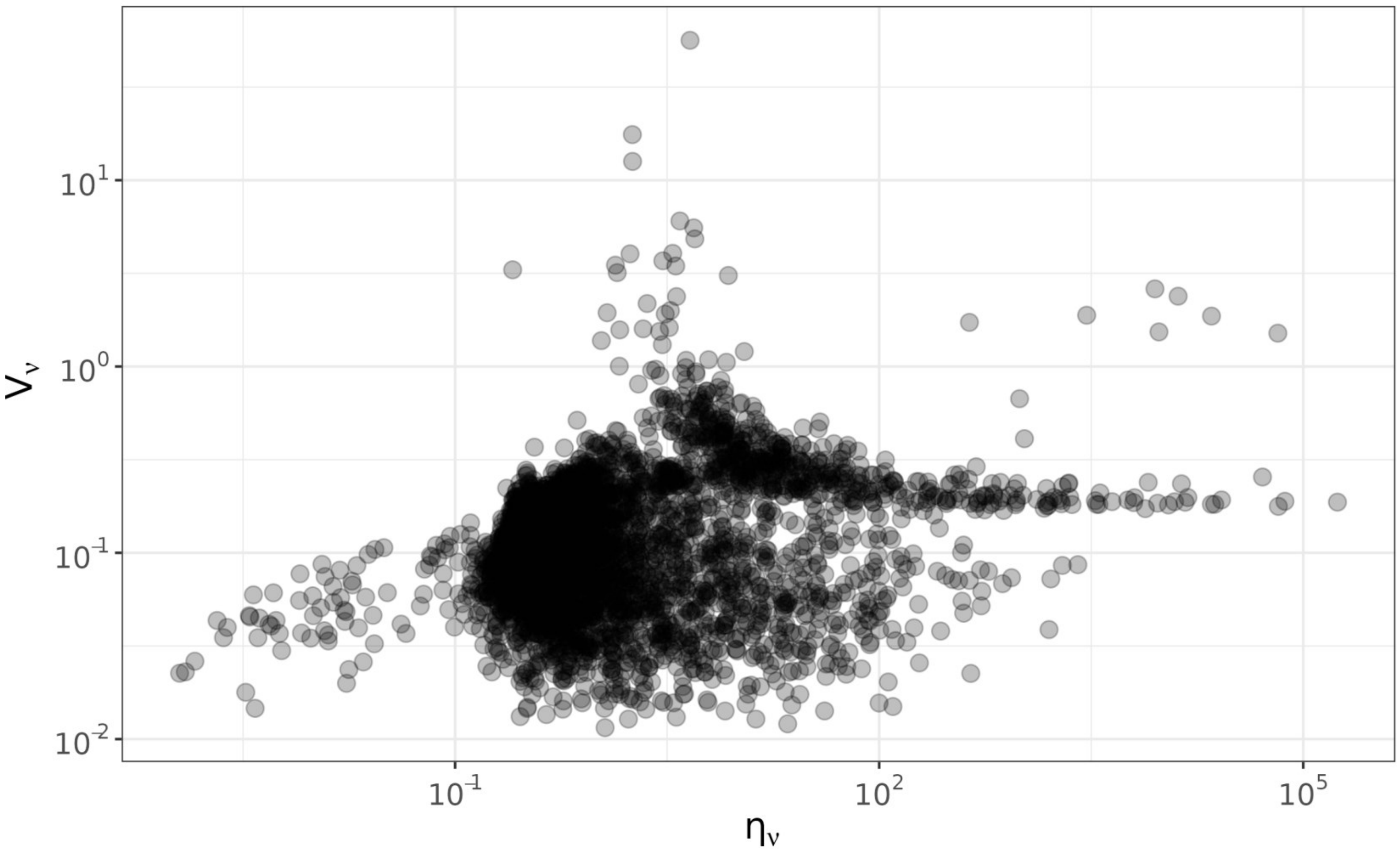
- The HUNt for Dynamic and Explosive Radio transients with MeerKAT
- Field of view of ≈ 1 square degree
- 6,394 radio light curves over 10 fields
- Flux density measurements + standard errors

MeerKAT Radio Telescope (Credit: SARAO)




B
Flux Density (mJy)
0.40
0.35
0.30
0.25
0.20
59420
59460
59500
Time (MJD)
D
Flux Density (mJy)
2.0
1.6
1.2
58250
58500
58750
59000
59250
59500
Time (MJD)
A
Flux Density (mJy)
0.20
0.16
0.12
0.08
59350
59400
59450
59500
Time (MJD)
C
Flux Density (mJy)
70
60
50
40
30
59300
59350
59400
59450
59500
Time (MJD)

Variability Statistics: η_ν and V_ν



(Data courtesy of Andersson et al., 2023)

$$\eta_\nu = \frac{1}{N}\sum\left(\frac{\text{Obs.} - \text{Wt. Mean}}{\text{Std. Error}}\right)^2$$

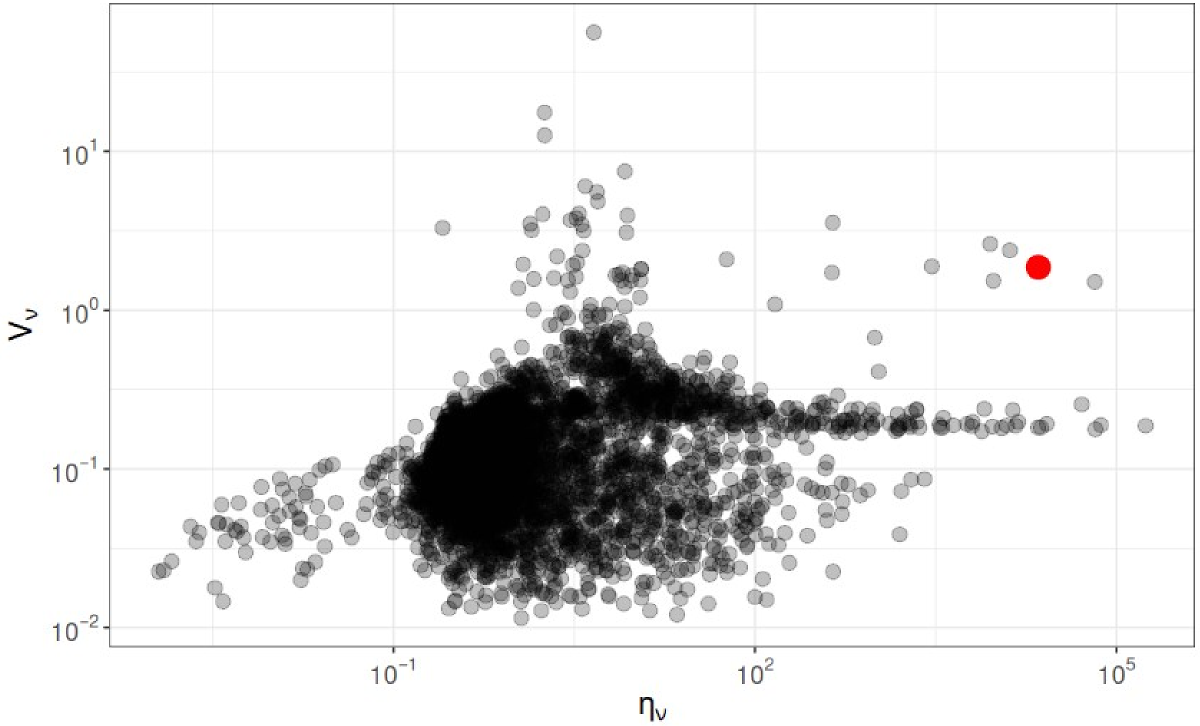
$$\sim \chi^2_{N-1}$$

$$V_\nu = \frac{\text{Standard Deviation}}{\text{Mean}}$$

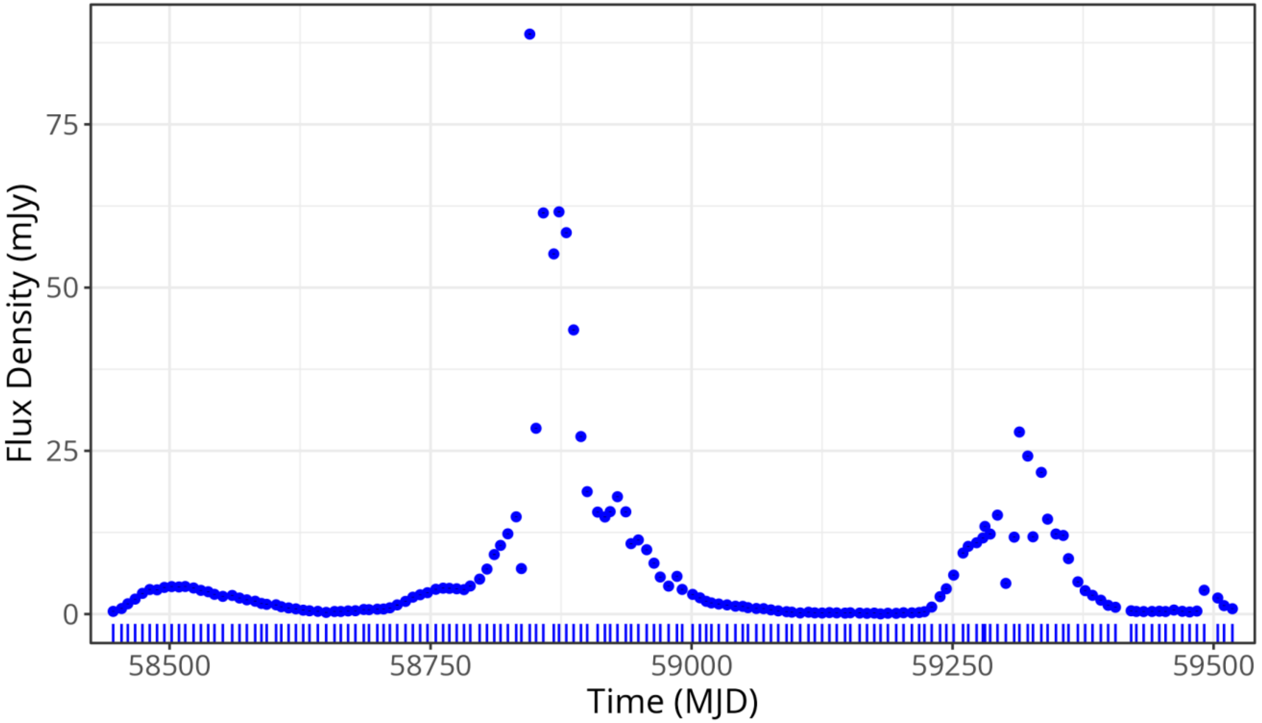
As $\eta_\nu \to \infty$ and $V_\nu \to \infty$

Source is likely transient

Variability Statistics: η_ν and V_ν



GX 339-4



$$(\eta_\nu = 22427.6, V_\nu = 1.86)$$

Characterising Light Curves

Oversimplified

- Fewer parameters
- Scales easily
- High information loss

Overspecified

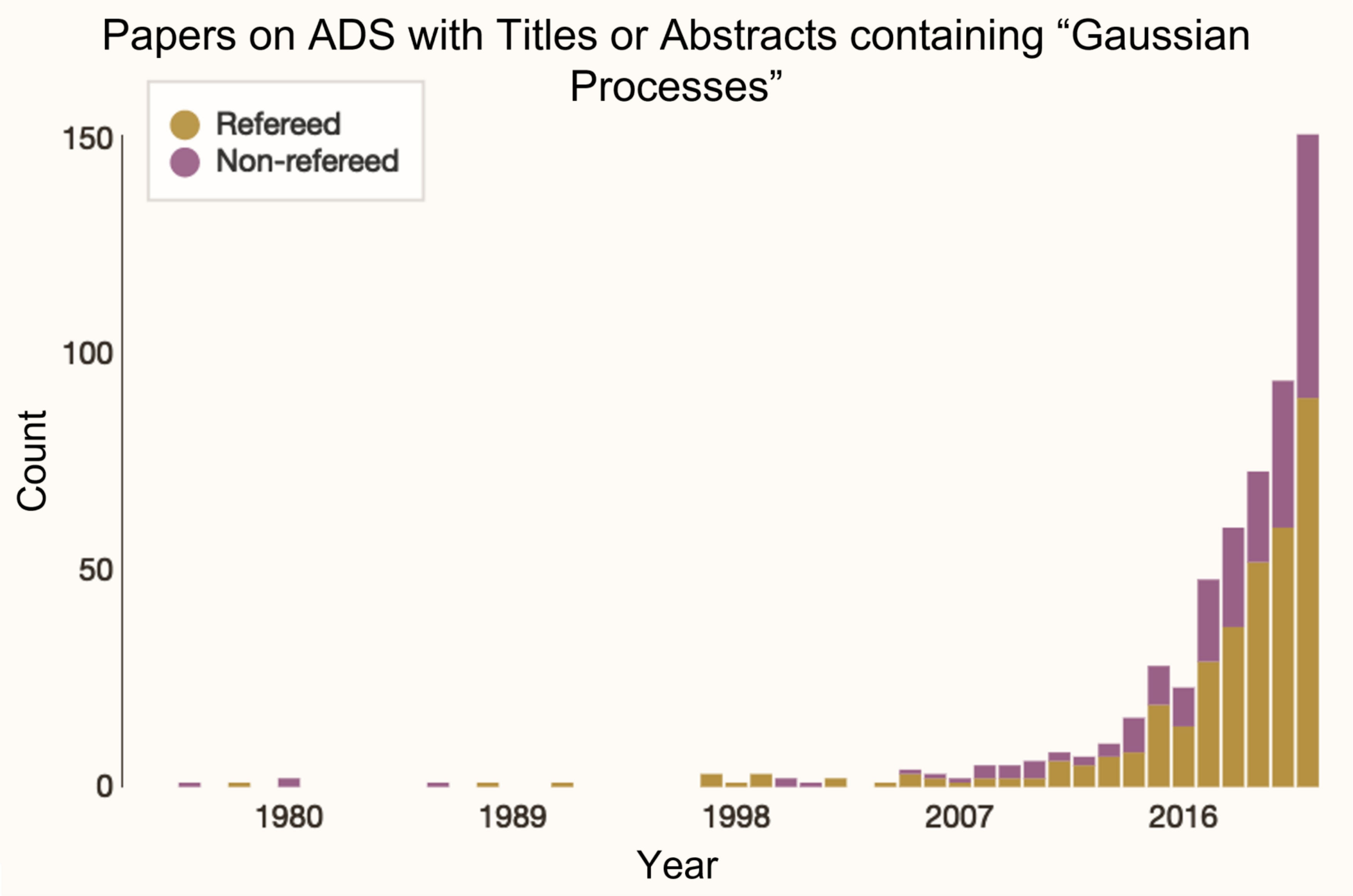
- Many parameters
- High discriminatory power
- Overfitting

Model light curves as a Gaussian Process (GP)

Astronomical & Statistical Objectives

1. Find the missing stellar mass black holes
   - Estimated population is $> 10^5$ but only found < 100.
   - New large-scale astronomical surveys, e.g., LSST, SKA.
   - Need techniques to analyse these large datasets.
2. Advance the use of GPs for time-series astronomy
   - Statistically justified and astrophysically meaningful representation of transients.
   - Handle sparse, unevenly sampled data.
   - Go beyond interpolation and point estimates.

GPs in Astronomy

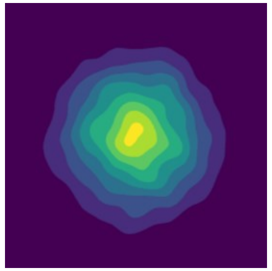


Multivariate Normal $\mathbf{Y} \sim \mathrm{MVN}(\mathbf{0}, \mathbf{\Sigma}_{n\times n})$

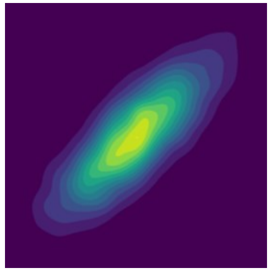
$\boldsymbol{Y}$ is a vector of n Gaussian random variables.

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \boldsymbol{Y} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{n\times n}), \qquad \boldsymbol{\Sigma}_{n\times n} = \begin{bmatrix} \Sigma_{11} & \cdots & \Sigma_{1n} \\ \vdots & \ddots & \vdots \\ \Sigma_{n1} & \cdots & \Sigma_{nn} \end{bmatrix}$$

where $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_n)$ and $\boldsymbol{\Sigma}$ is a $n \times n$ covariance matrix.



$$\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\Sigma = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

- Symmetric, positive semi-definite matrix.
- Linear combinations of covariance matrices are also valid covariance matrices.

Gaussian Processes (GPs)

Extend multivariate Gaussian to ‘infinite’ dimensions.

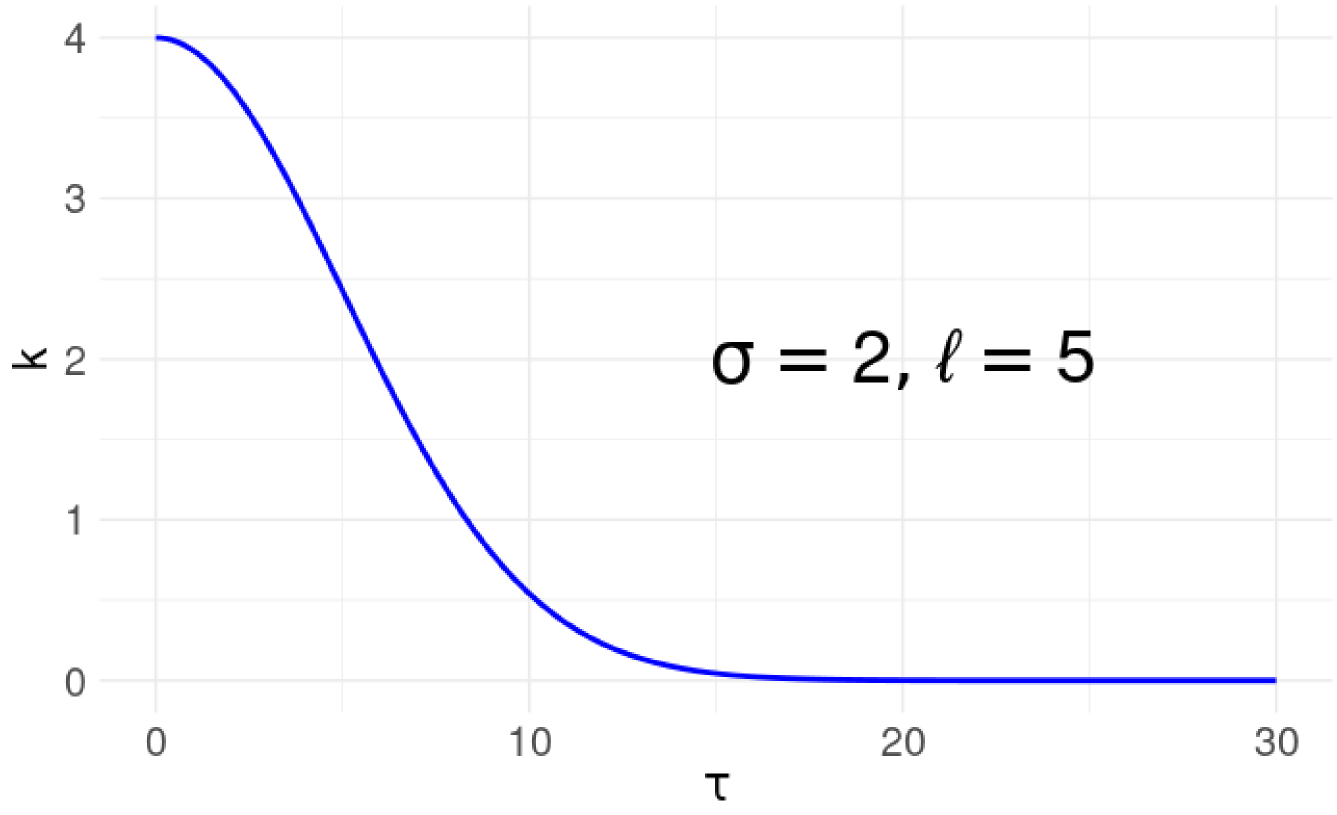
- Mean function, $\mu(t)$
- Covariance or **kernel function**, $\boldsymbol{\kappa(t,t)}$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \end{bmatrix} = \boldsymbol{Y} \sim GP(\mu(t), \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu} = \mu(t_i)$ and $\Sigma_{ij} = \boldsymbol{\kappa(t_i, t_j)}$, for $i, j = 1, 2, \ldots$

Rather than specifying a fixed covariance matrix with fixed dimensions, compute covariances using the kernel function.

Squared Exponential Kernel



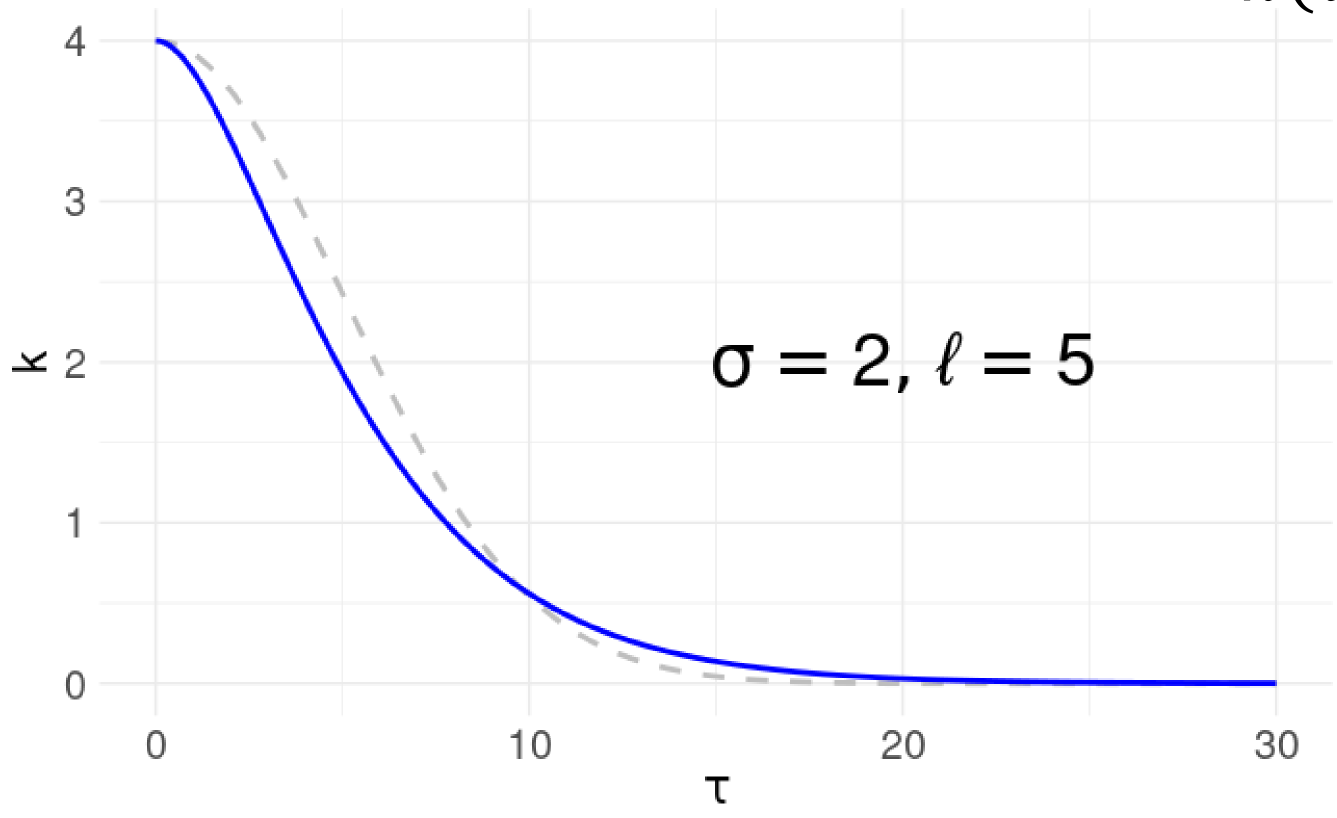
$$k(\tau; \sigma, \ell) = \sigma^2 \exp\left\{-\frac{1}{2}\left(\frac{\tau}{\ell}\right)^2\right\}$$

$$\sigma, \ell > 0$$

$$\tau = |t_r - t_c|$$

- As distance (in time) increases ↗ the covariance decreases ↘
- Stationary time-series

Matern 3/2 Kernel



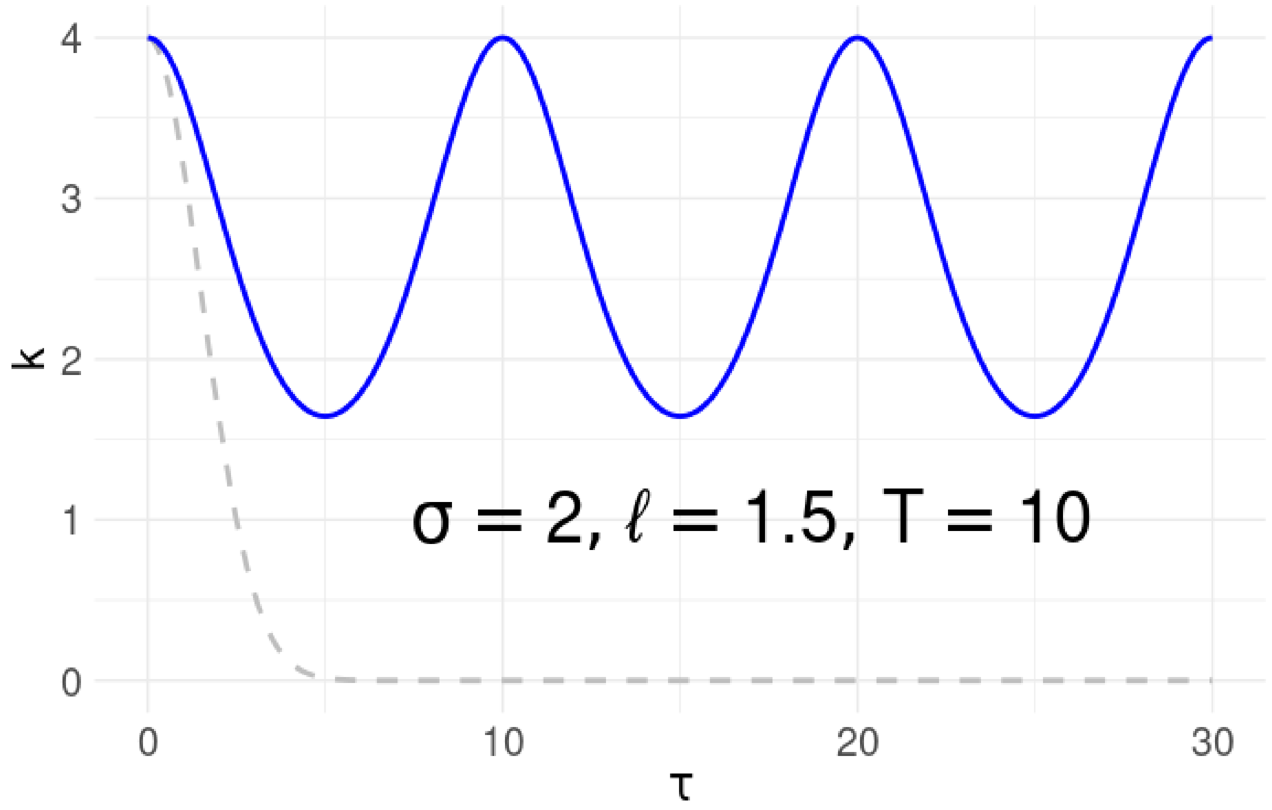
$$k(\tau; \sigma, \ell) = \sigma^2 \left(1 + \sqrt{3}\frac{\tau}{\ell}\right) \exp\left\{-\sqrt{3}\frac{\tau}{\ell}\right\}$$

$$\sigma, \ell > 0$$

$$\tau = |t_r - t_c|$$

- Decays faster than SE kernel
- Converges on SE as order, i.e., 3/2, 5/2, etc, increases
- More jagged curves

Periodic Kernel



$$k(\tau;\sigma,\ell) = \sigma^2 \exp\left\{-\frac{2}{\ell^2}\sin^2\left(\pi\frac{\tau}{T}\right)\right\}$$

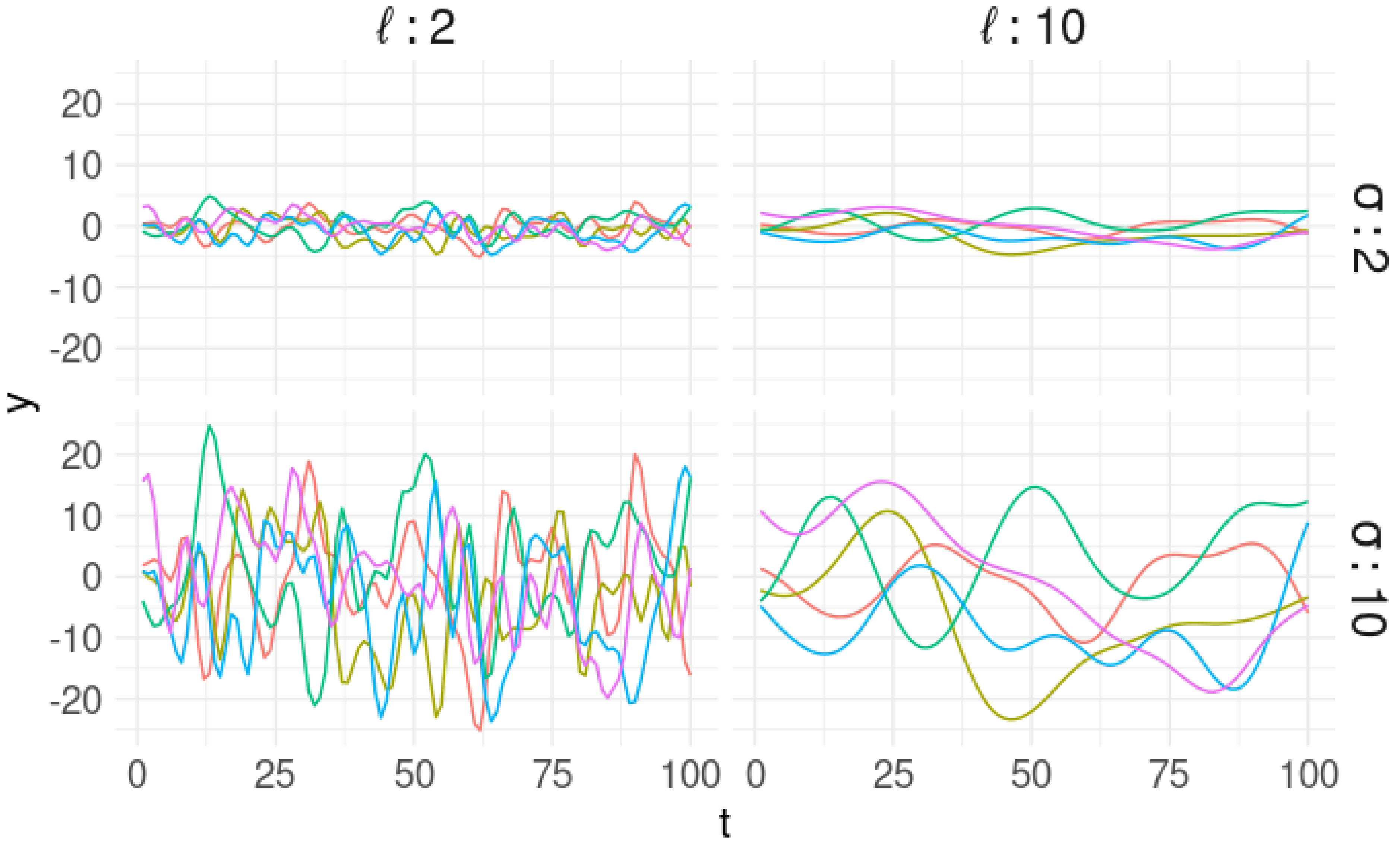
$$\sigma, \ell, T > 0$$

$$\tau = |t_r - t_c|$$

- Additional hyperparameter
- Covariance might never decay to zero

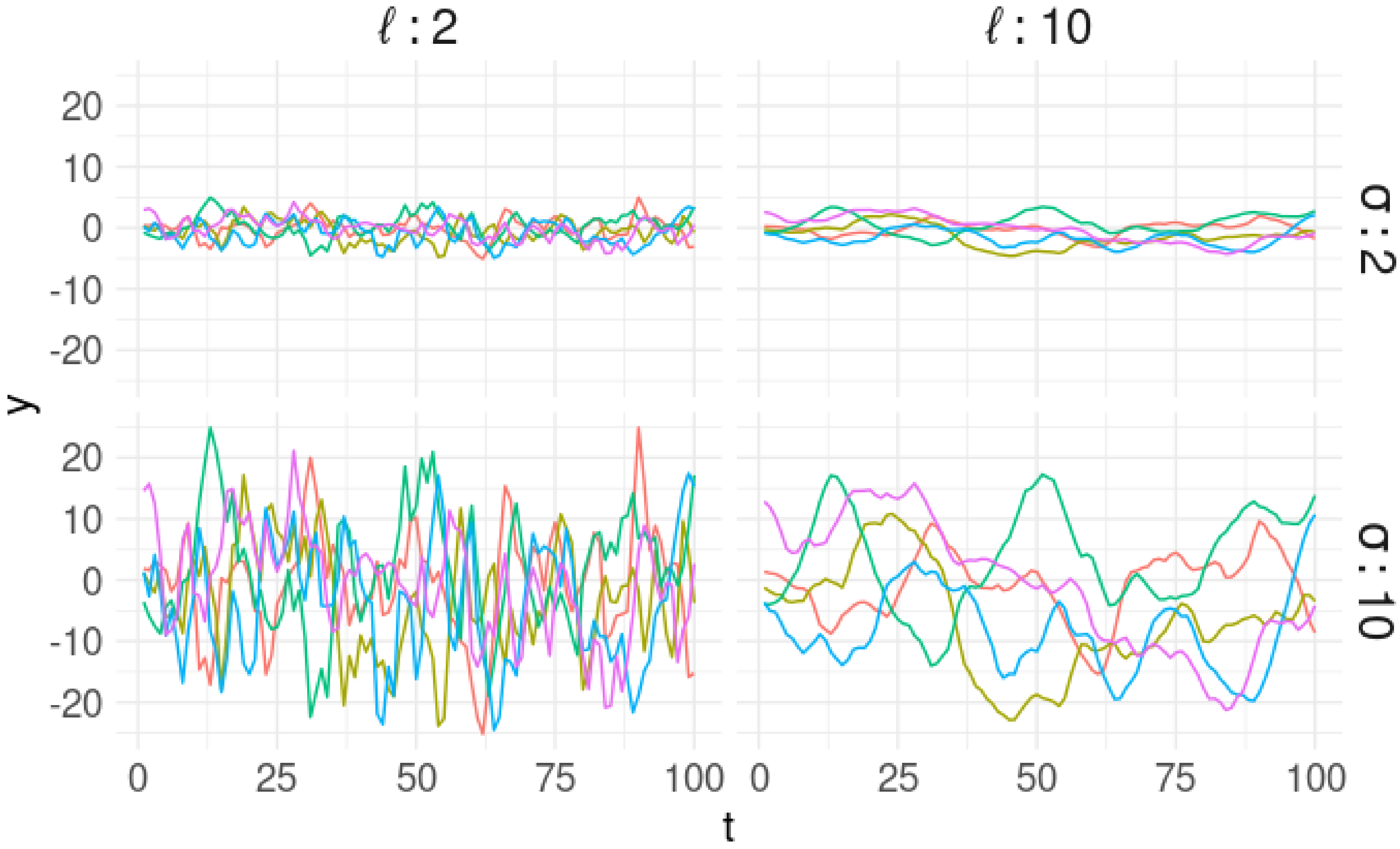
Squared Exponential Kernel

$$\kappa(\tau; \sigma, \ell) = \sigma^2 \exp\left\{-\frac{1}{2\ell^2}\tau^2\right\}$$



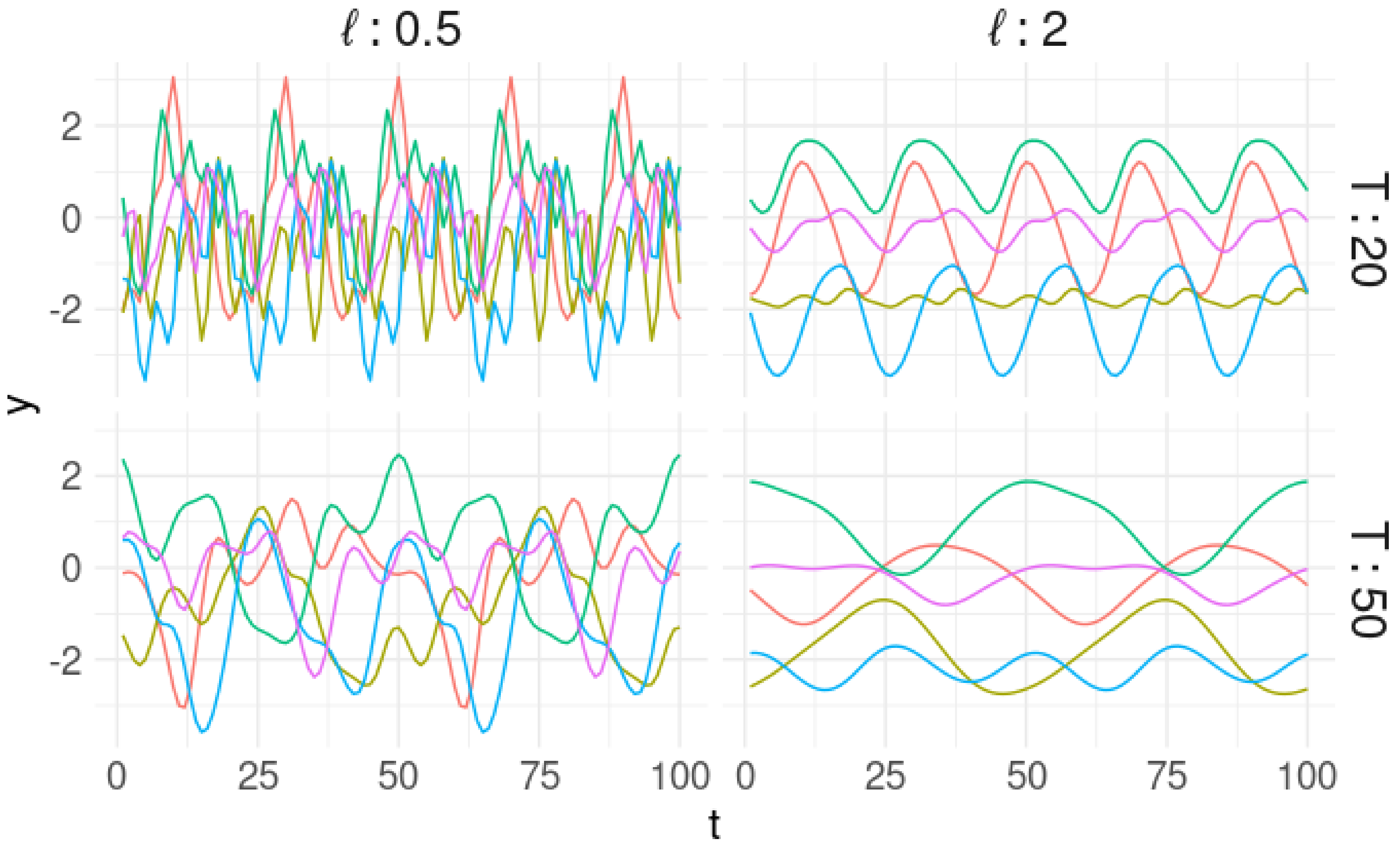
Matern 3/2 Kernel

$$\kappa(\tau;\sigma,\ell) = \sigma^2\left(1+\sqrt{3}\,\frac{\tau}{\ell}\right)\exp\left\{-\sqrt{3}\,\frac{\tau}{\ell}\right\}$$

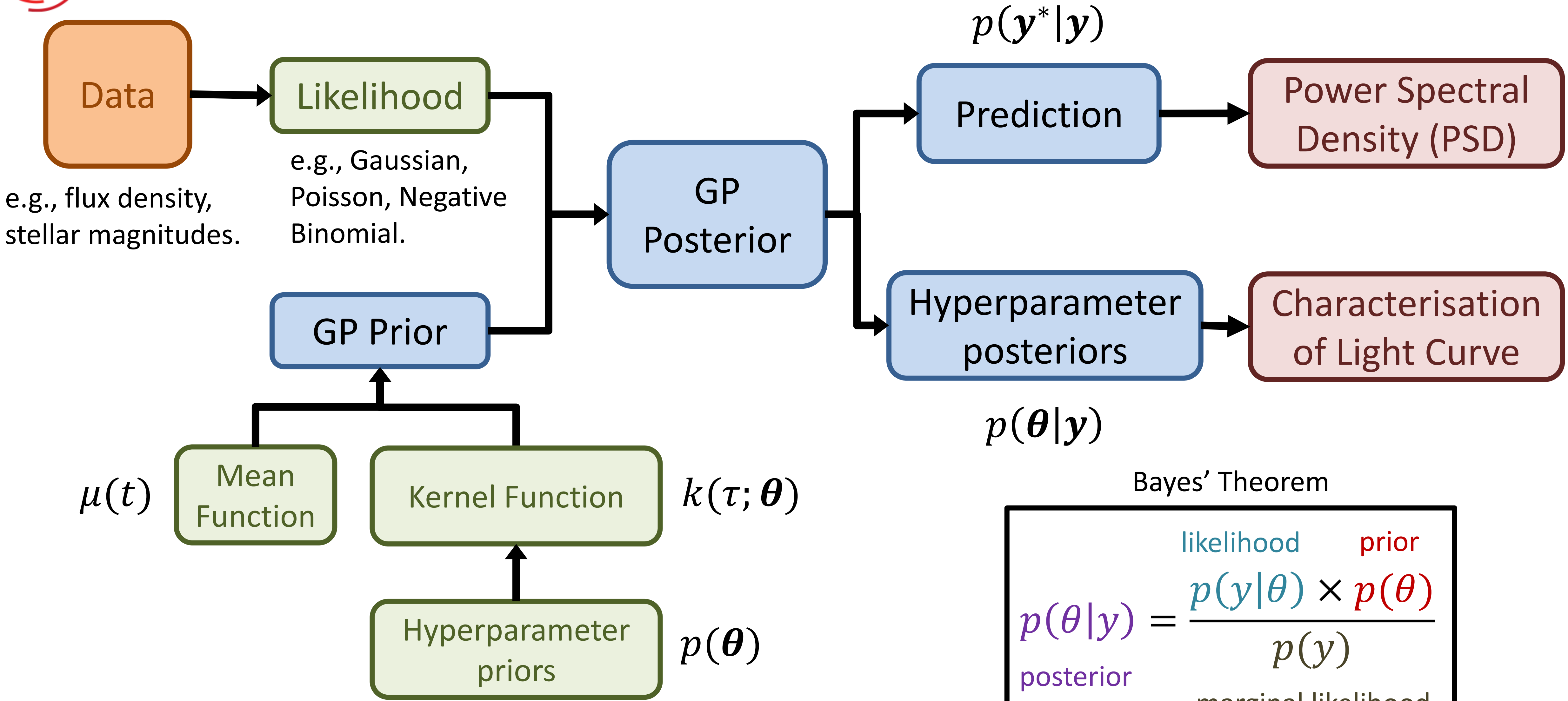


Periodic Kernel

$$\kappa(\tau;\sigma,\ell) = \sigma^2 \exp\left\{-\frac{2}{\ell^2}\sin^2\left(\pi\frac{\tau}{T}\right)\right\}$$



Modelling Workflow



Bayes' Theorem

$$\underbrace{p(\theta|y)}_{\text{posterior}} = \frac{\overbrace{p(y|\theta)}^{\text{likelihood}} \times \overbrace{p(\theta)}^{\text{prior}}}{\underbrace{p(y)}_{\text{marginal likelihood}}}$$

Bayesian Hierarchical Model

Data Model $\mathbf{Y} \sim \mathrm{MVN}(\mathbf{f}, \hat{\boldsymbol{e}}^2)$

Process Model $\mathbf{f} \sim \mathrm{GP}(\mathbf{0}, \boldsymbol{K}_{N\times N})$

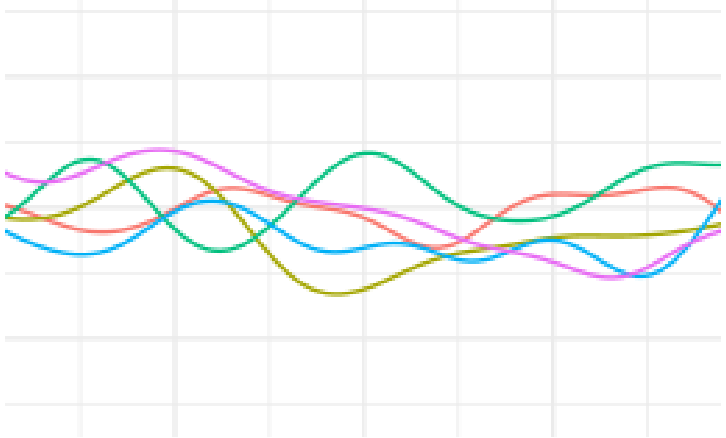
$$r, c = 1, \ldots, N.$$

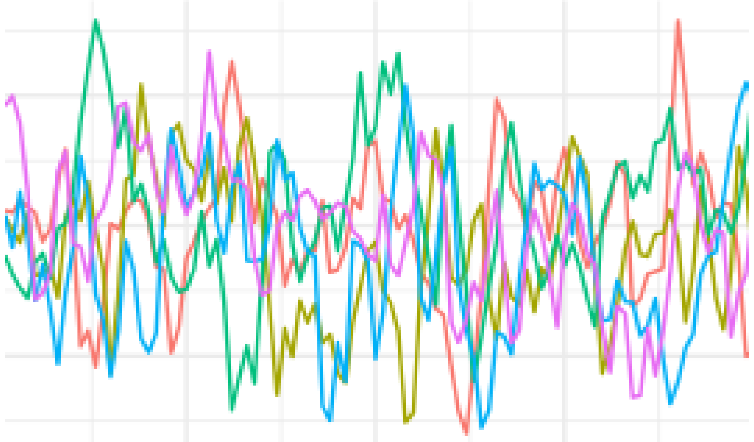
$$\boldsymbol{\theta} = (\sigma_{SE}, \ell_{SE}, \sigma_{M32}, \ell_{M32}, \sigma_P, \ell_P, T)$$

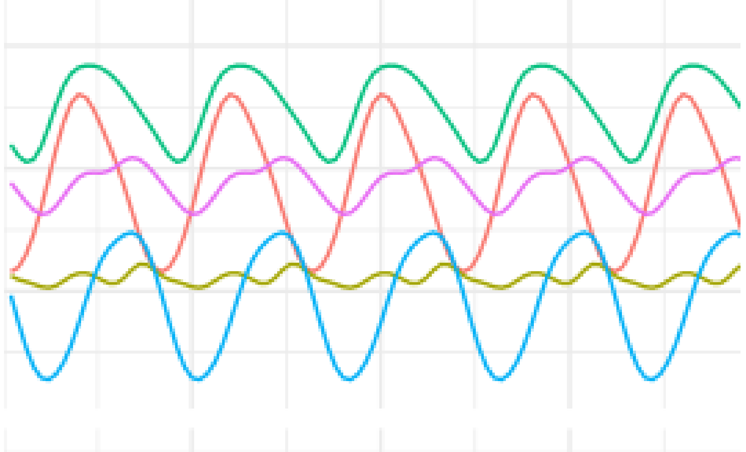
$$[\boldsymbol{K}]_{rc} = \kappa(t_r, t_c | \boldsymbol{\theta})$$

$$= \boxed{\underbrace{\kappa_1(\tau; \sigma_{SE}, \ell_{SE})}_{\text{Squared Exponential}}} + \boxed{\underbrace{\kappa_2(\tau; \sigma_{M32}, \ell_{M32})}_{\text{Matern 3/2}}} + \boxed{\underbrace{\kappa_3(\tau; \sigma_P, \ell_P, T)}_{\text{Periodic}}}$$

Covariance Kernel







Hyperparameter Model

Minimal density near short scales

Standardised flux densities

$$\sigma_{SE}, \sigma_{M32}, \sigma_P \sim N^+(0,1)$$

$$\ell_{SE}, \ell_{M32}, \ell_P \sim \mathrm{InverseGamma}\left(\alpha = 3, \beta = \frac{1}{2}\mathrm{range}(t)\right)$$

$$T \sim \mathrm{Uniform}\left[2 \times \min(\Delta t),\ \frac{1}{4}\mathrm{range}(t)\right]$$

SE kernel to fit longer term trends than M32 kernel

$$\ell_{SE} > \ell_{M32} > 0$$

$$\min(\Delta t) < \ell_{\cdot}$$

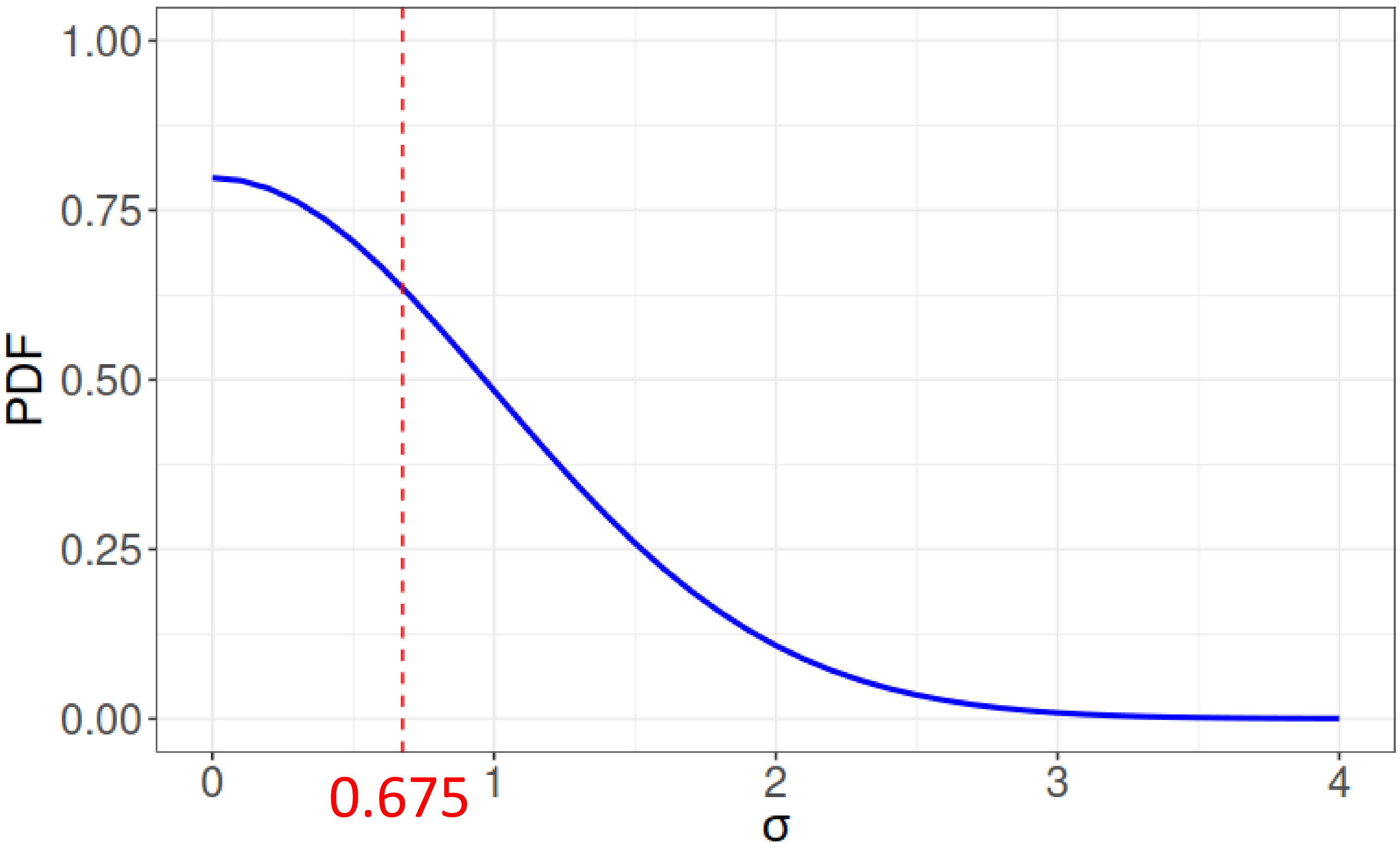
No more than half of total duration

Constrain length scale to be at least as wide as the narrowest gap in light curve

Periods bounded by Nyquist rate

Observe at least four cycles of any periodicity

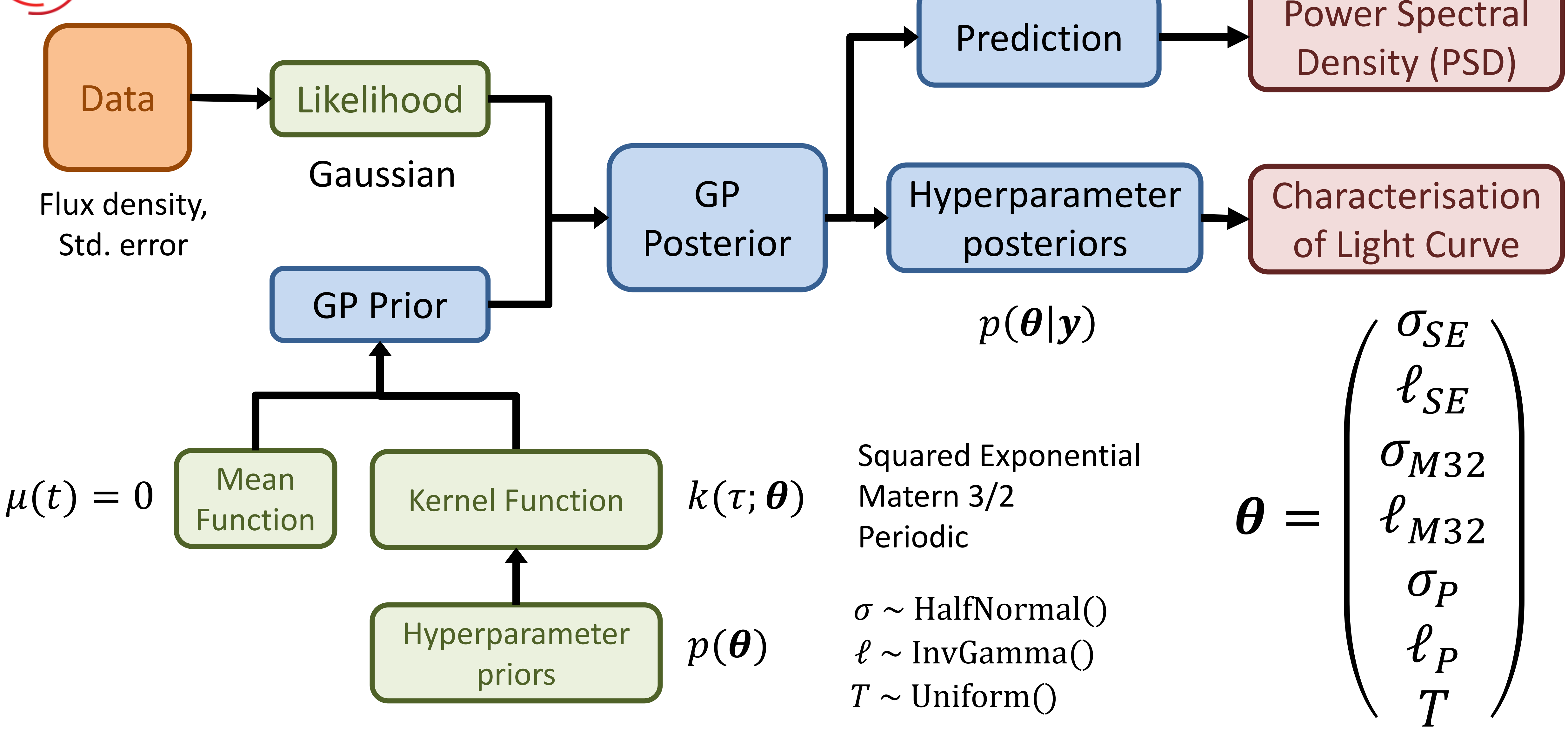
Half-Normal Distribution



$$\sigma \sim N^{+}(0,1)$$

- Truncated and rescaled standard Normal distribution.
- Use median = **0.675** as a naive threshold.

Modelling Workflow

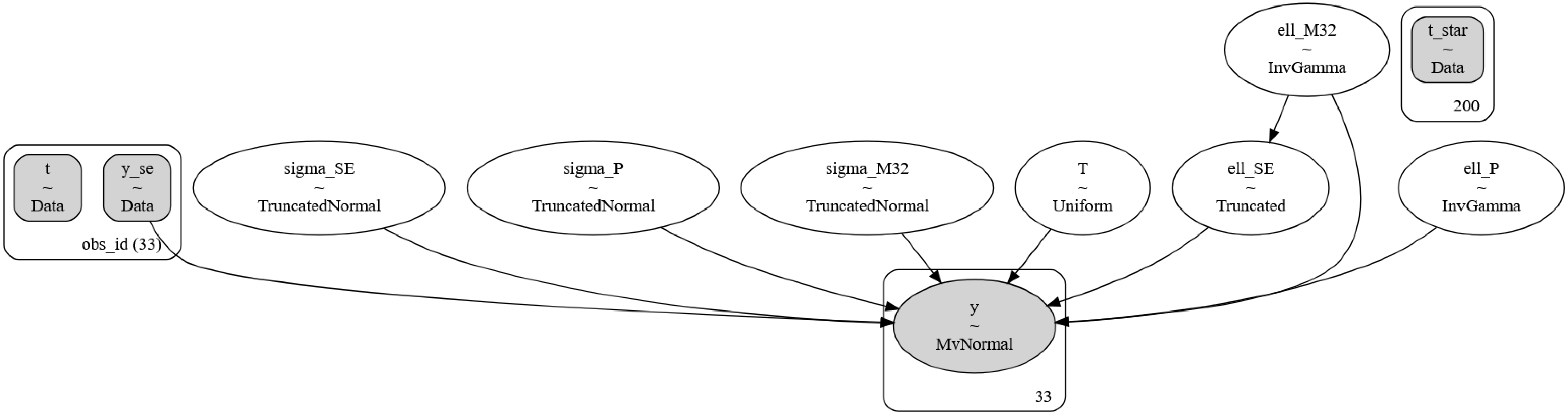


Tools

- Implemented in `Python`[1] (v3.11) and `PyMC`[2] (v3.16.2)
    - Accessible to astronomers
    - Probabilistic programming framework
    - Well-maintained open-source software
- Repeated analyses in `R`[3] (v4.3.1) and `Stan`[4] (v2.34)
- Also considered: `celerite2`[5], `george`[6].

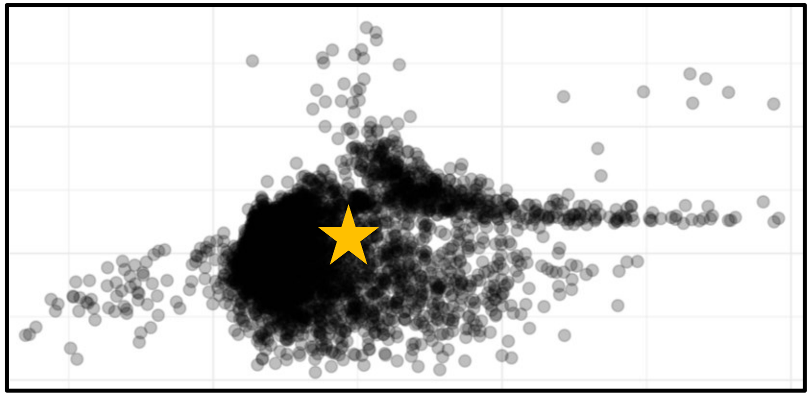
1. https://www.python.org
2. https://www.pymc.io
3. https://cran.r-project.org/
4. https://mc-stan.org/
5. https://celerite2.readthedocs.io/en/latest/
6. https://george.readthedocs.io/en/latest/

PyMC generated DAG

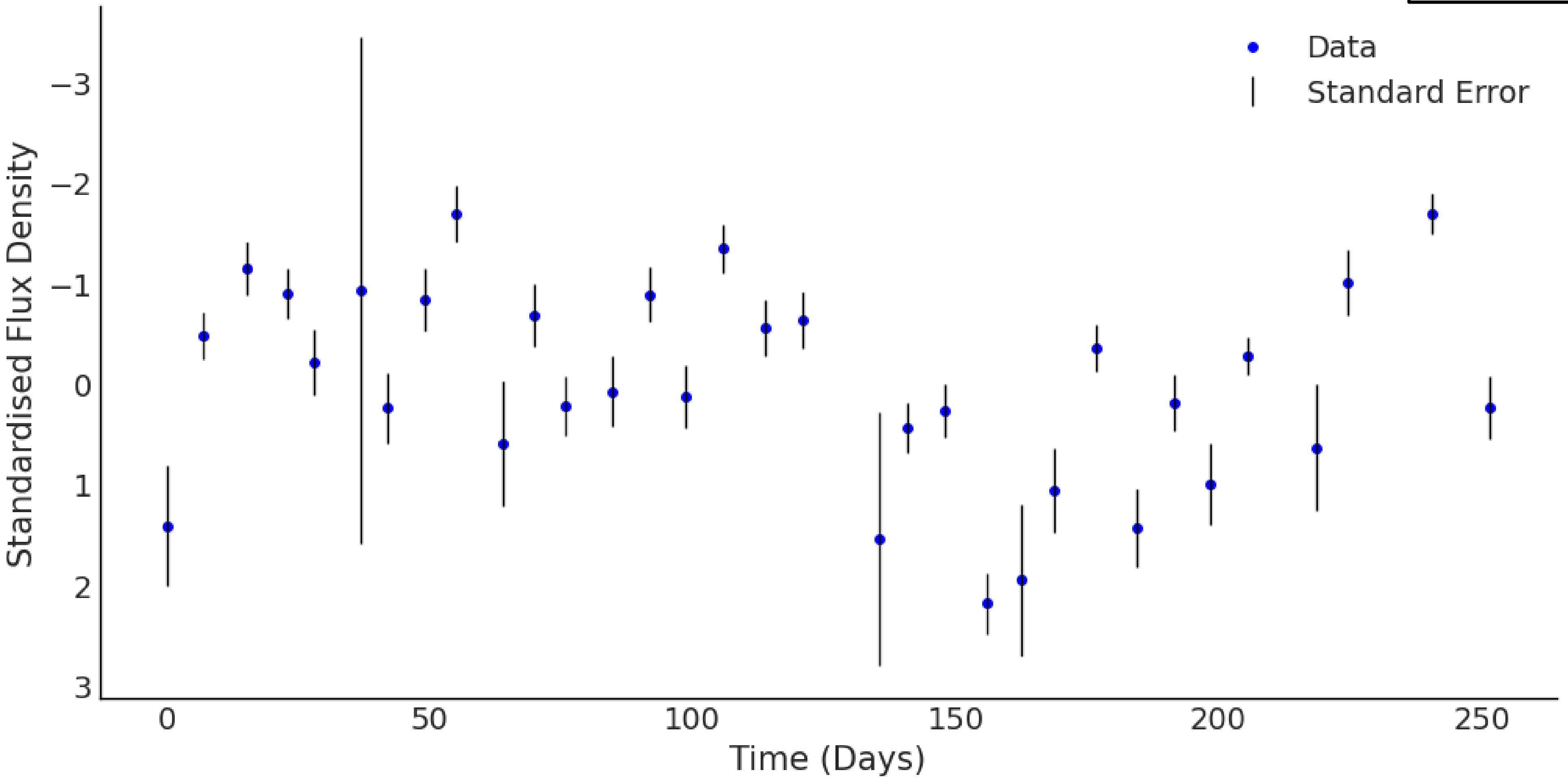


Marginal GP implementation

Example

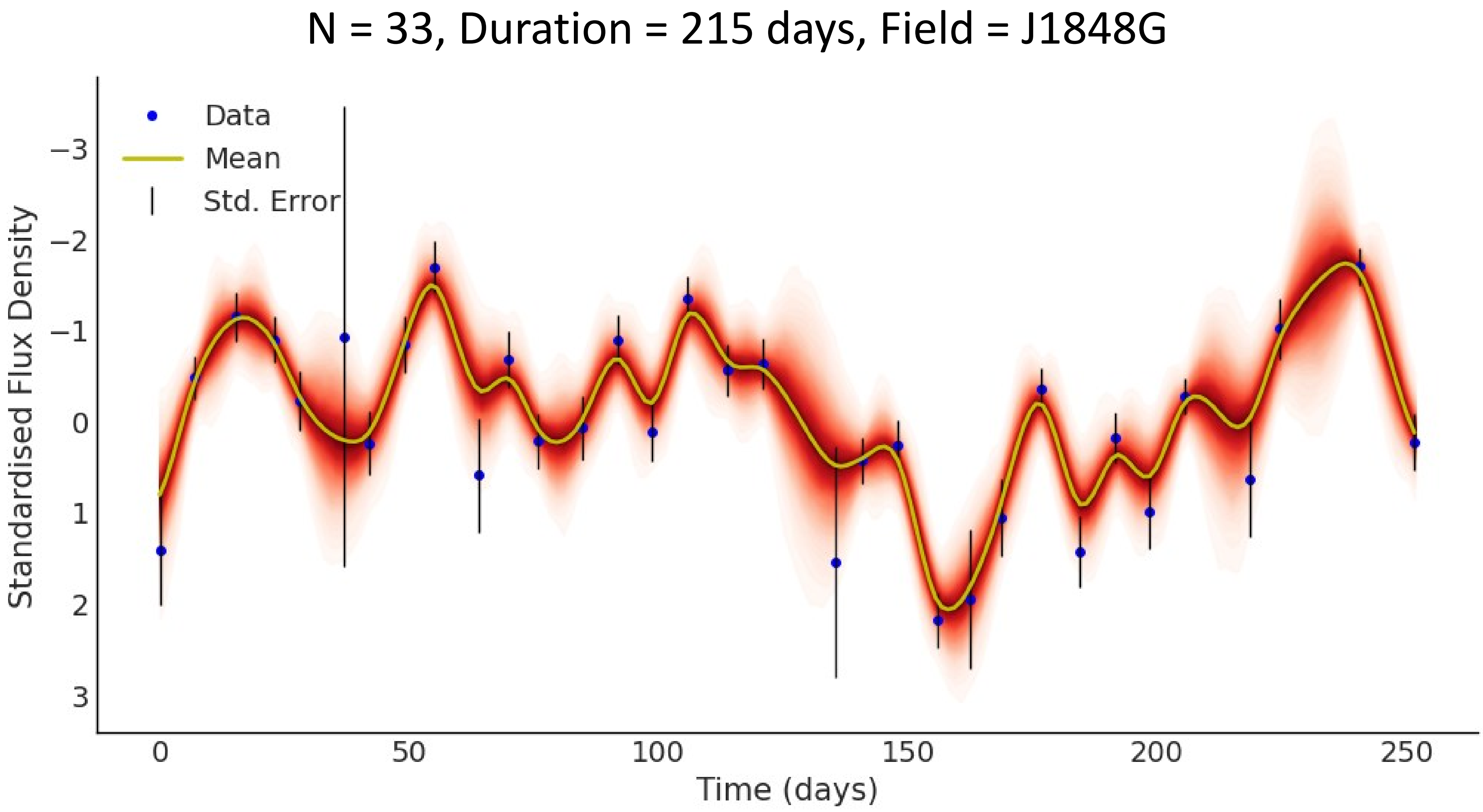


N = 33, Duration = 215 days, Field = J1848G



$$\eta_\nu = 2.91$$
$$V_\nu = 0.12$$

Posterior Predictive Samples



Posterior Medians

$\sigma_{\mathrm{SE}} = 0.39$

$\boldsymbol{\sigma_{\mathrm{M32}} = 1.26}$

$\sigma_{\mathrm{P}} = 0.50$

$\ell_{\mathrm{SE}} = 50.0$

$\ell_{\mathrm{M32}} = 11.9$

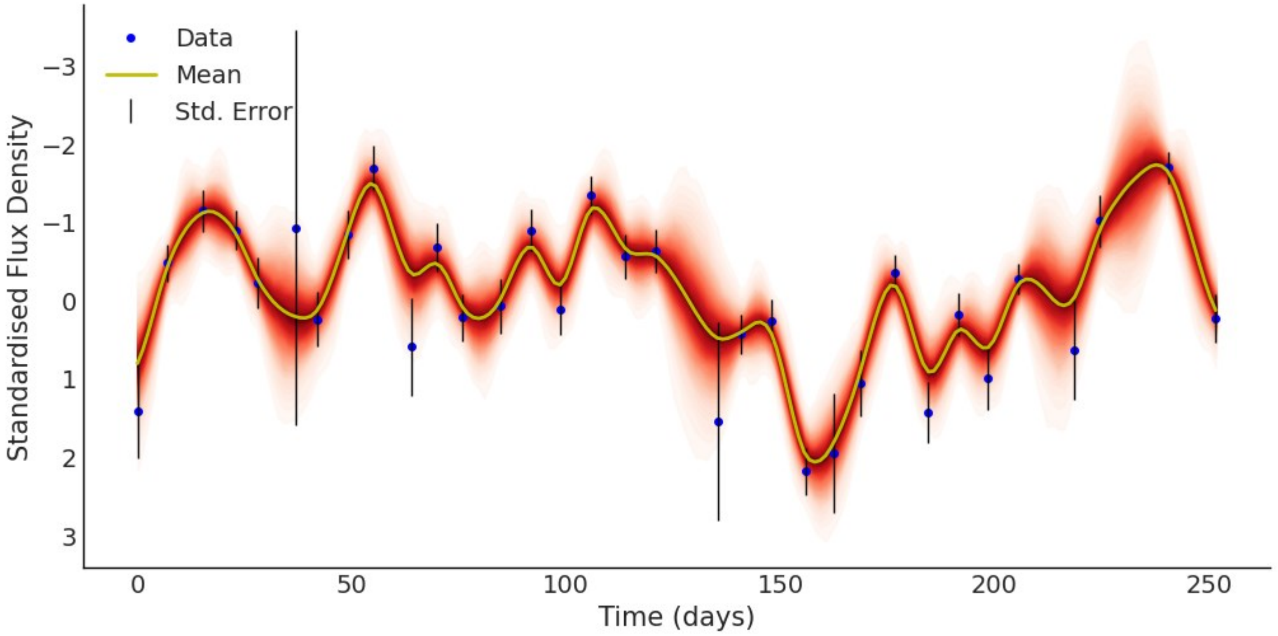
$\ell_{\mathrm{P}} = 46.7$

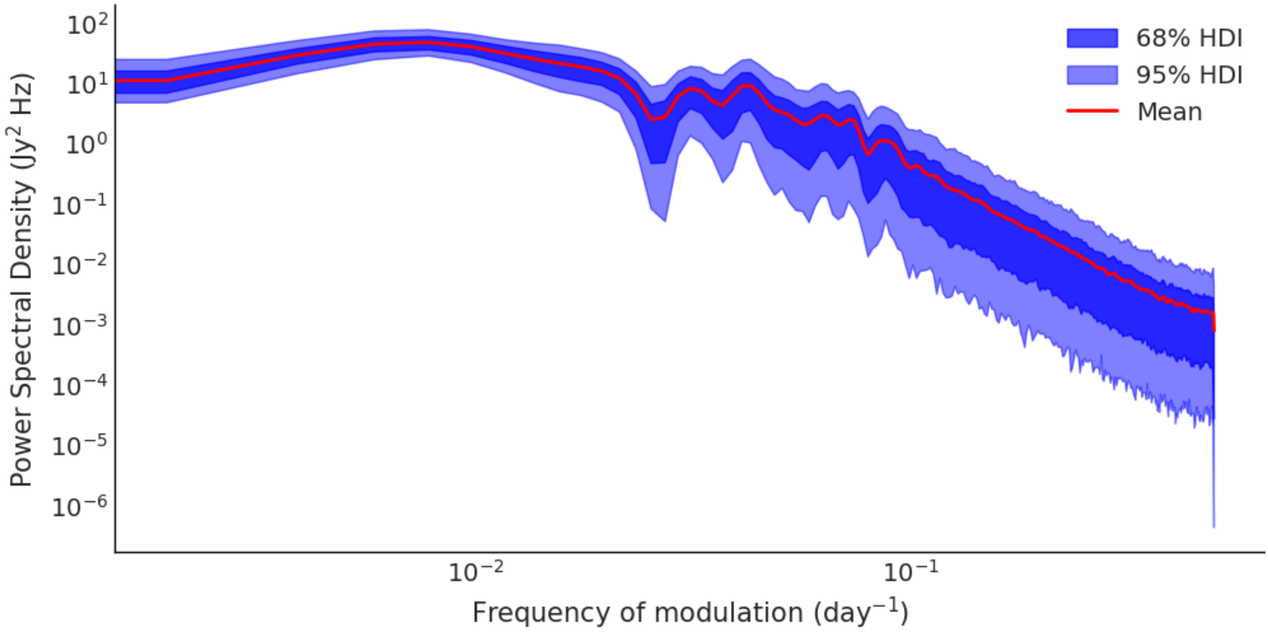
$T = 41.1$

$\eta_{\nu} = 2.91$

$V_{\nu} = 0.12$

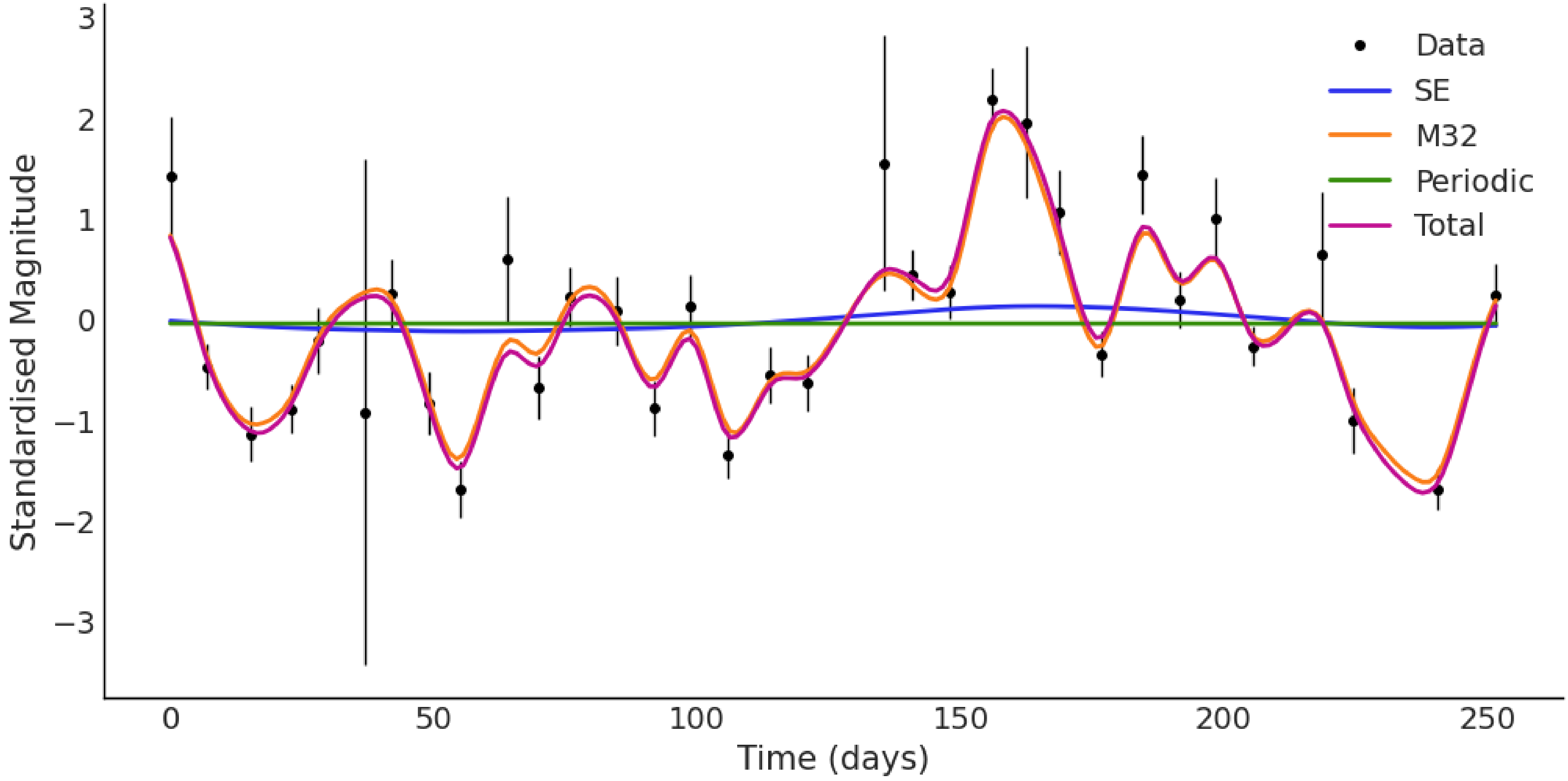
Power Spectral Density (PSD)



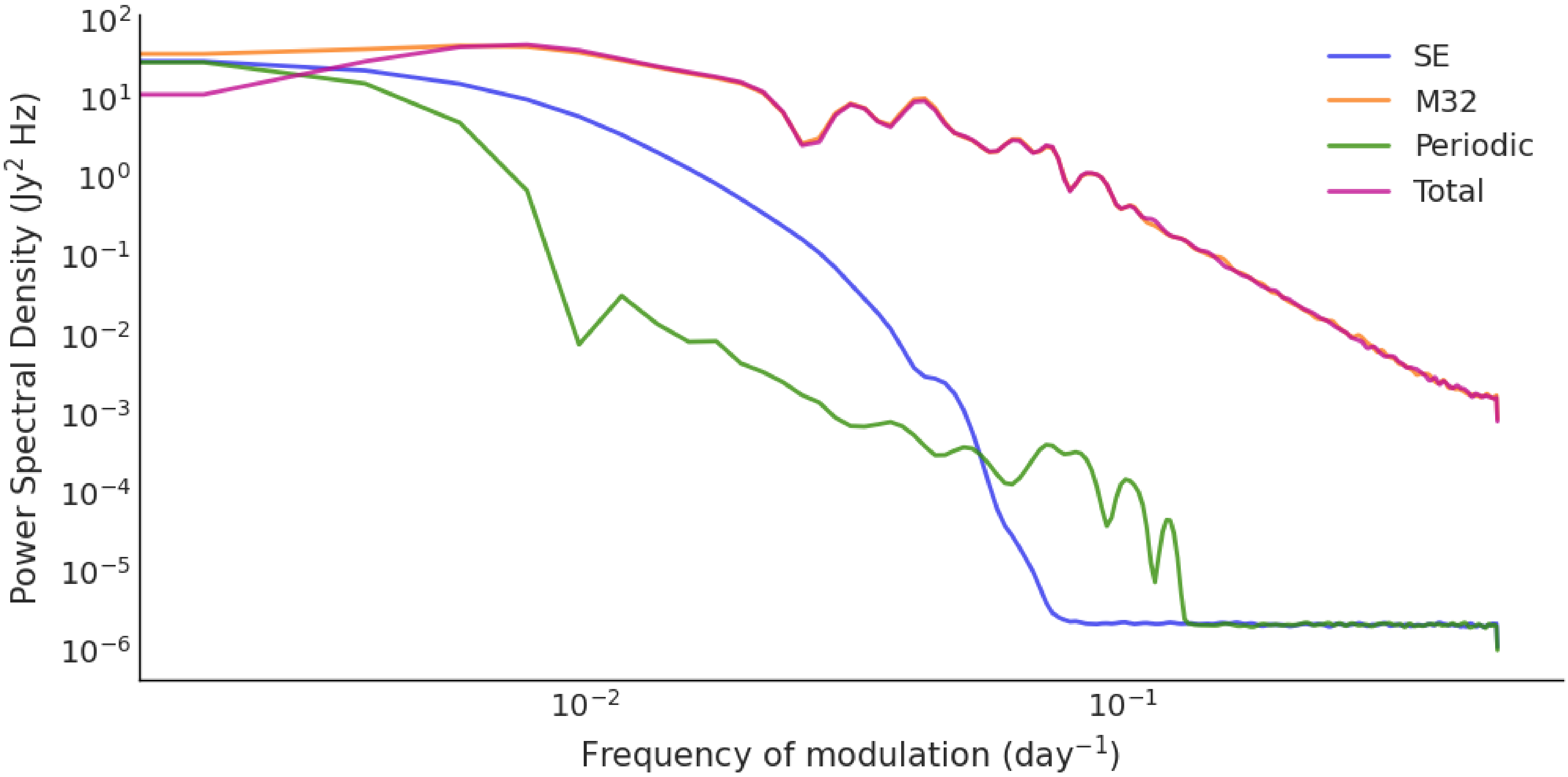


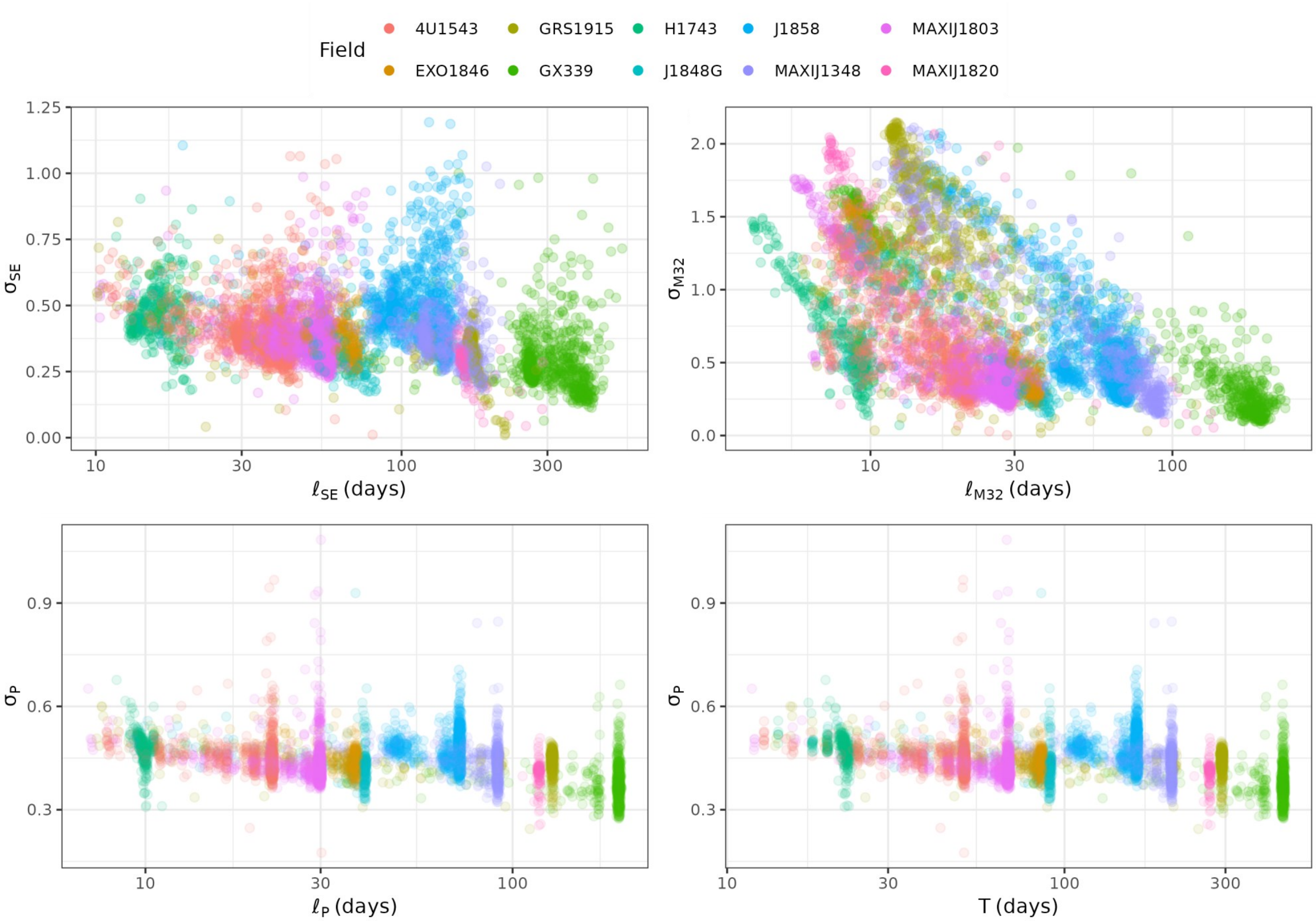
- Compute PSD of each posterior predictive sample
- Typical correlated (red) noise spectrum

Additive Components (Posterior Predictive)



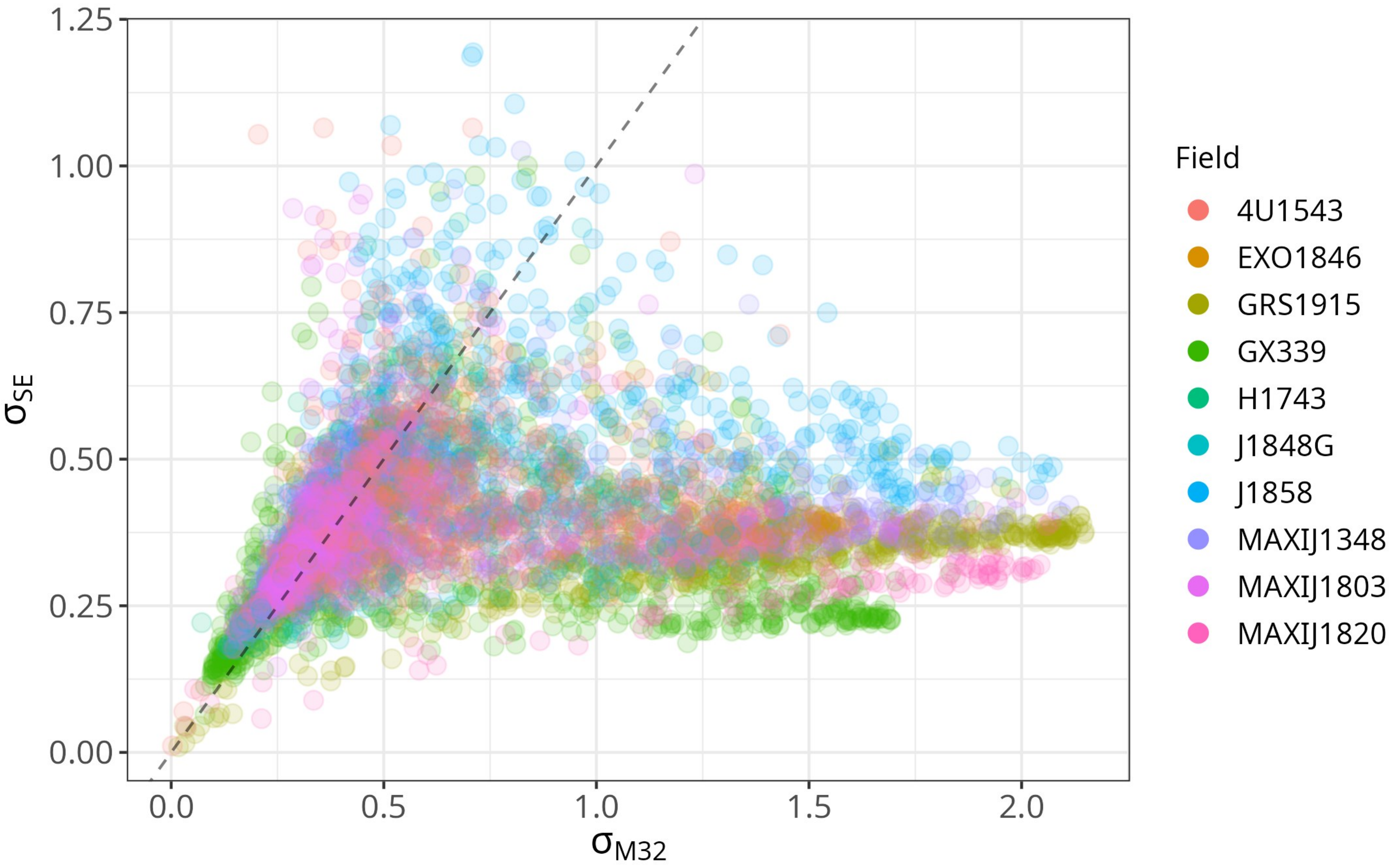
Additive Components (PSD)



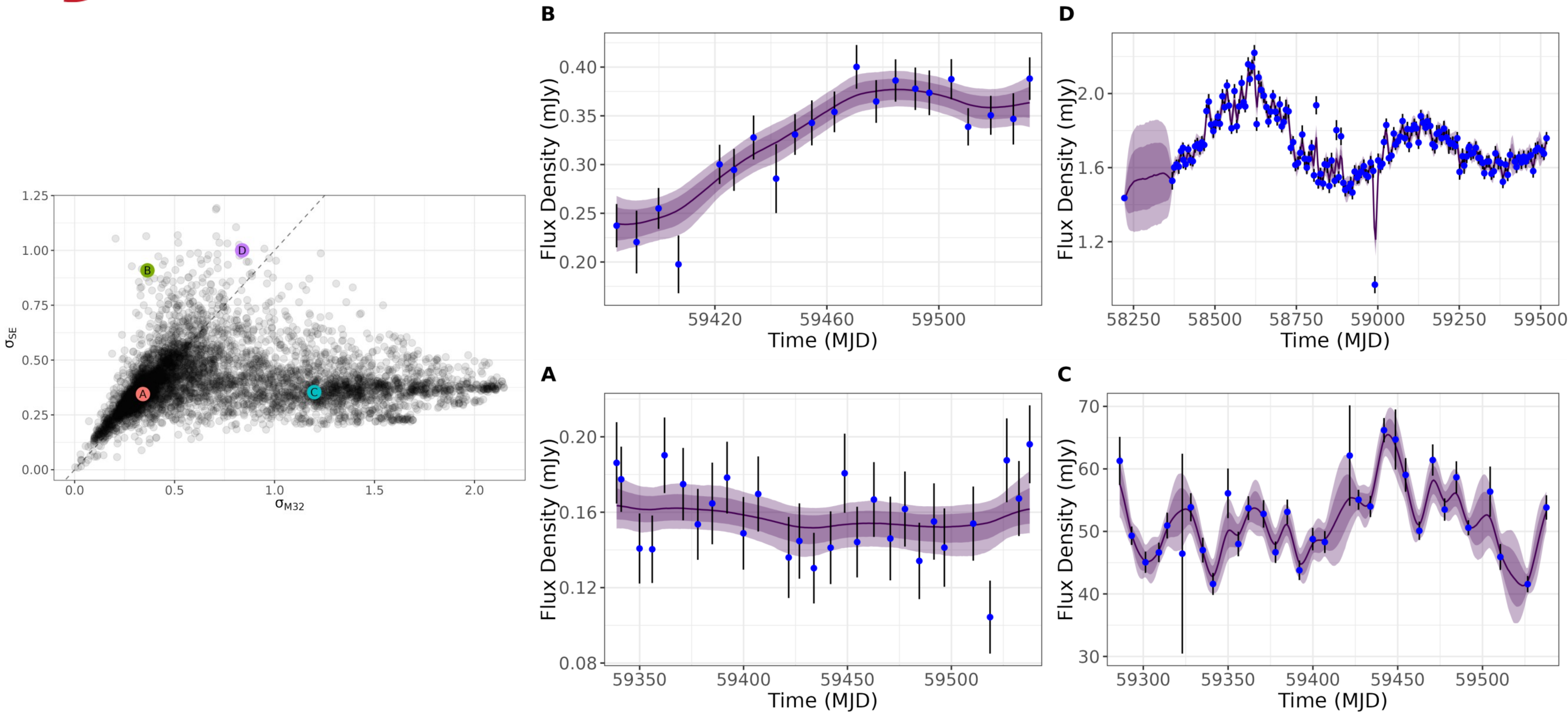


Field
4U1543
GRS1915
H1743
J1858
MAXIJ1803
EXO1846
GX339
J1848G
MAXIJ1348
MAXIJ1820
σ_{SE}
ℓ_{SE} (days)
σ_{M32}
ℓ_{M32} (days)
σ_P
ℓ_P (days)
σ_P
T (days)

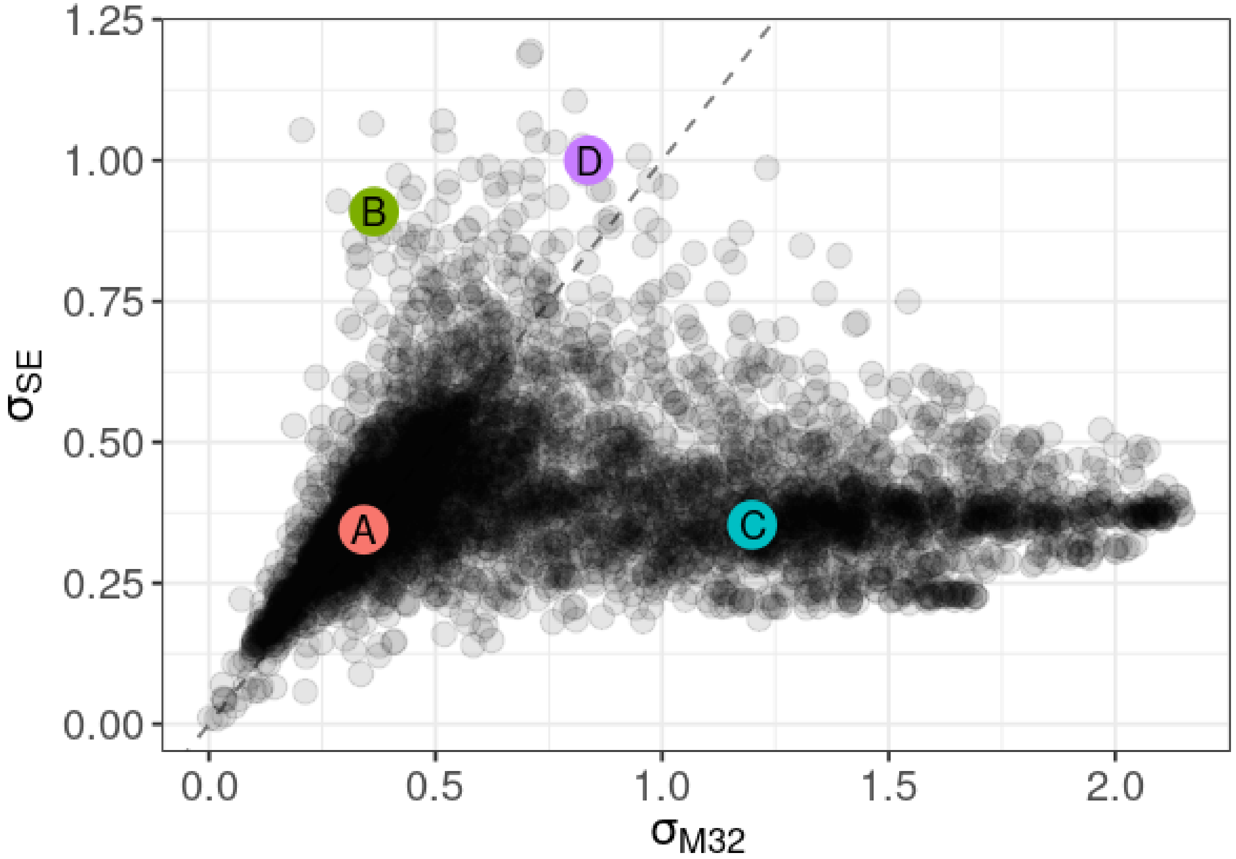
Amplitude Hyperparameter

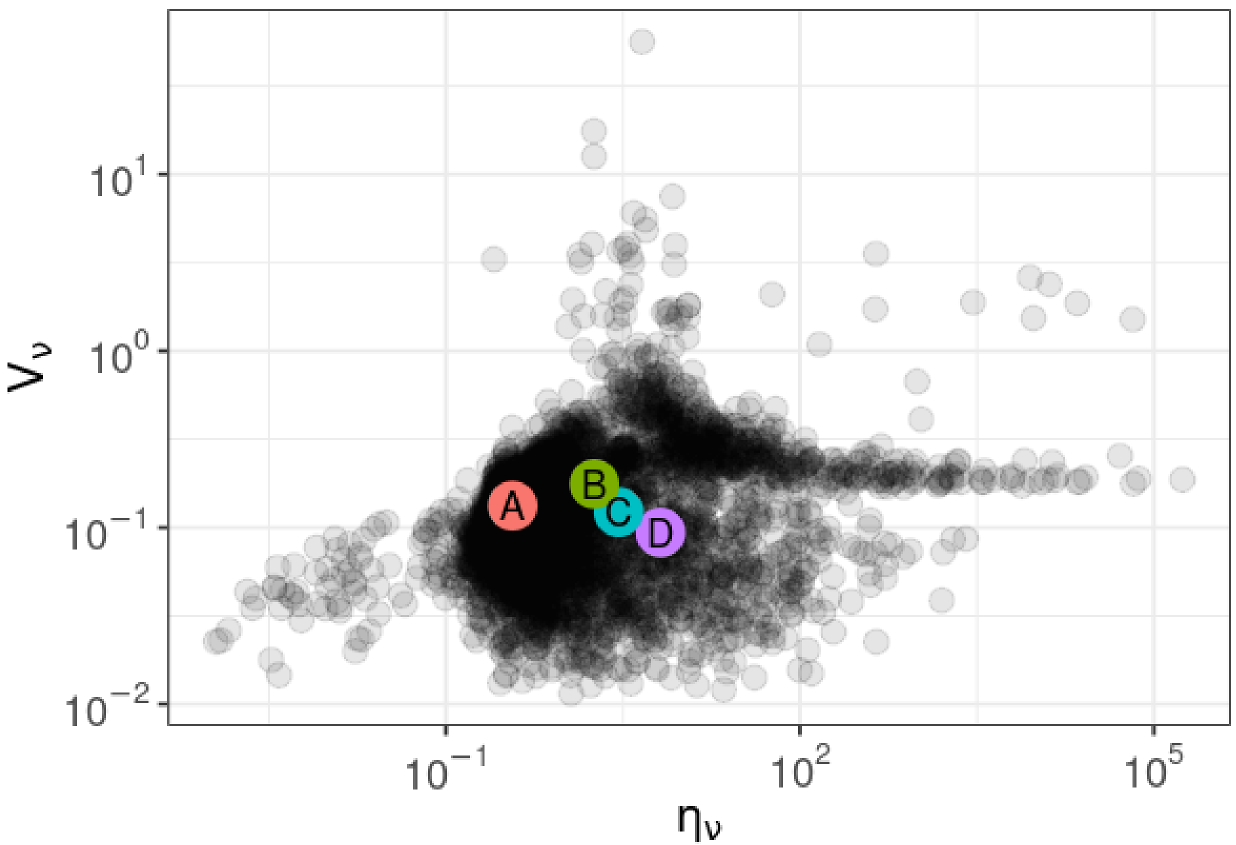


Explore the hyperparameter space



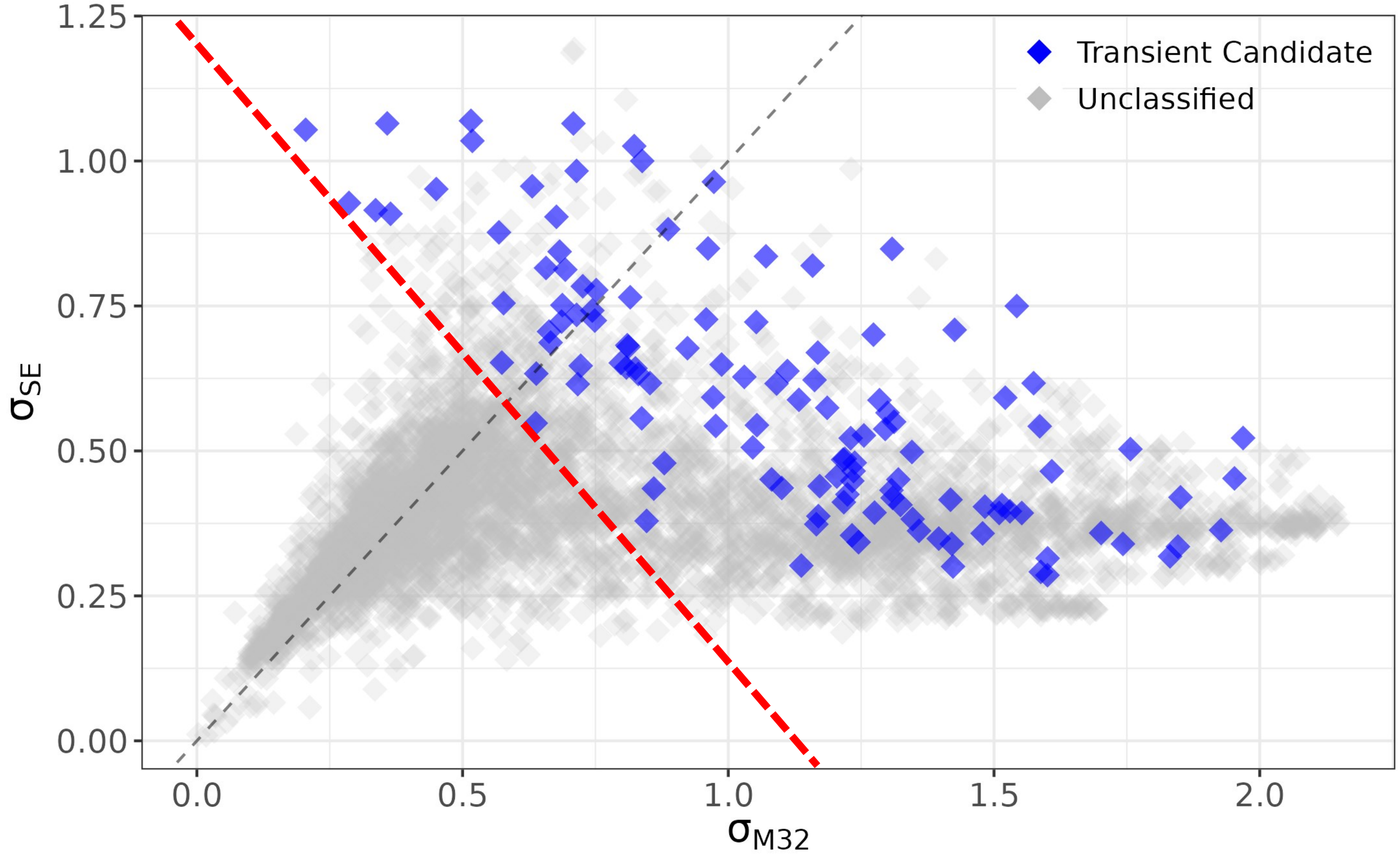
Comparison with V_ν vs η_ν





Interpreting Amplitude as Transcience

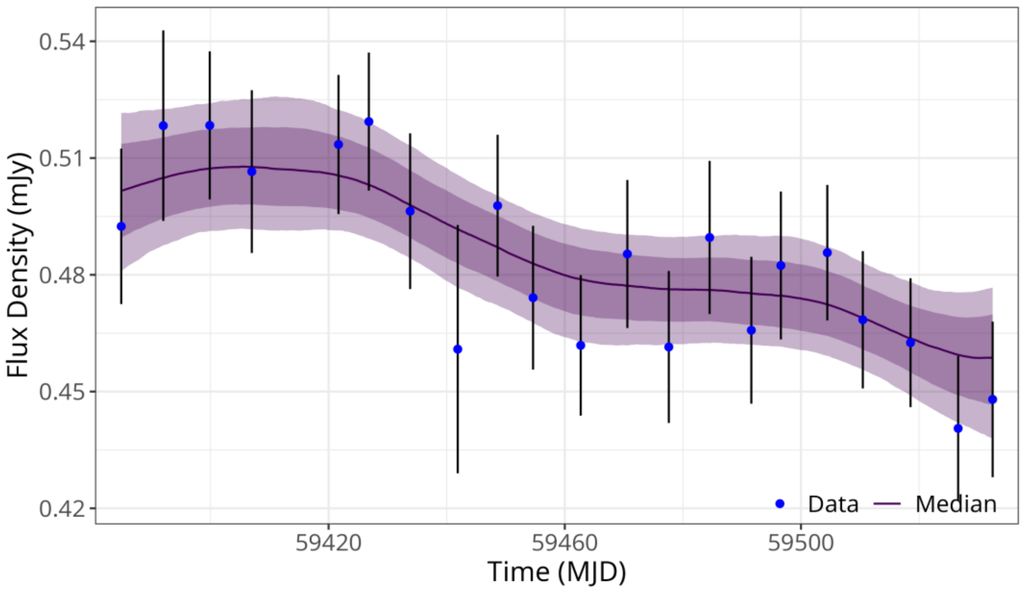
- Transience seems to manifest as large values in **amplitude**, σ.
- Previously identified transient candidates all seem to lie the upper right of this parameter space.

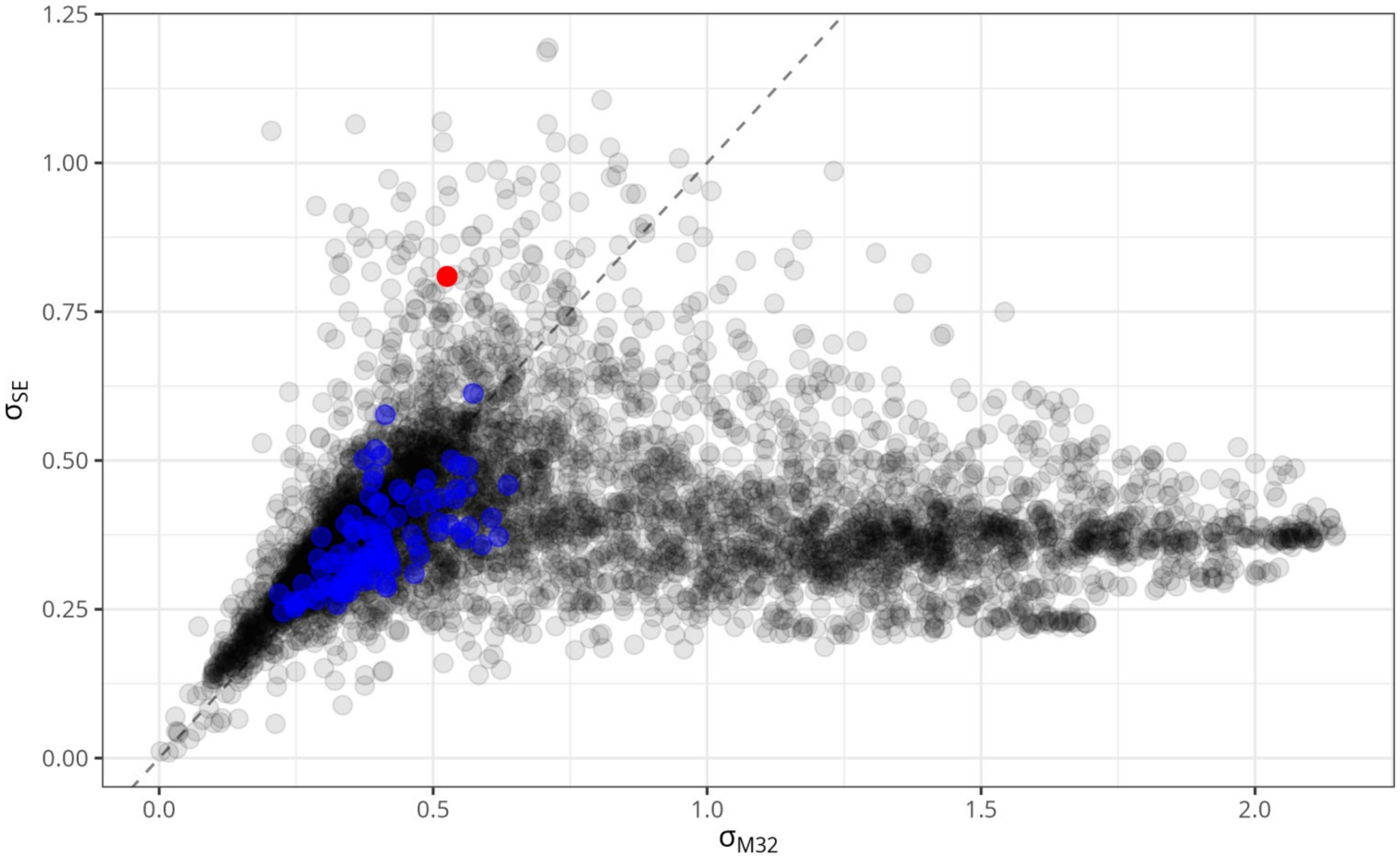


Data: Andersson et al. (2023)

Figure: Fu et al. (in prep.)

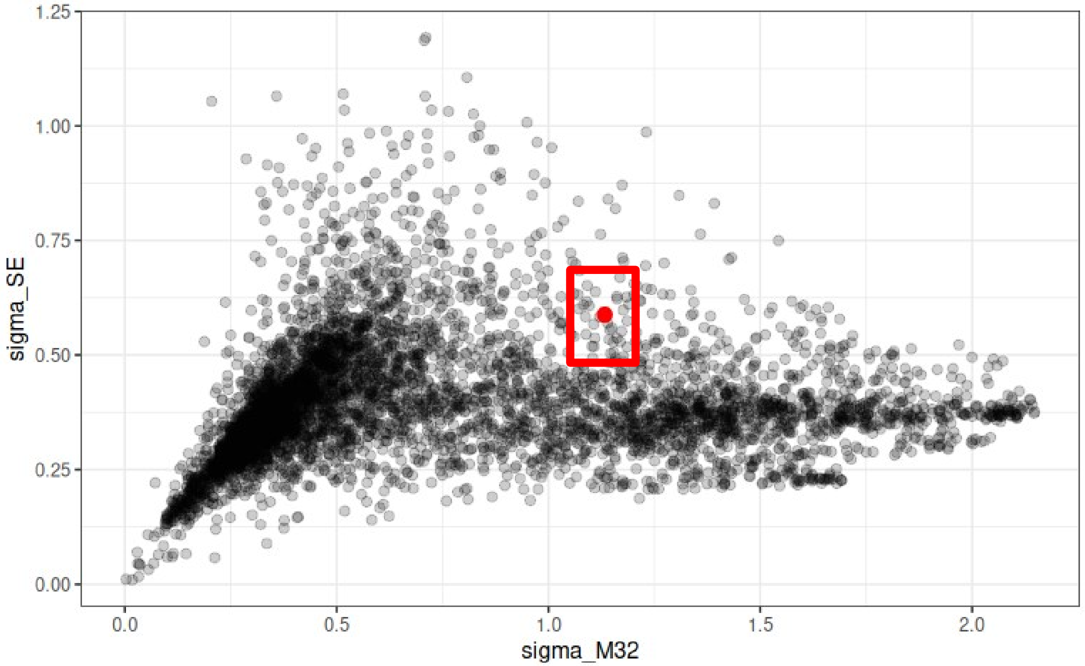
Permutation Test

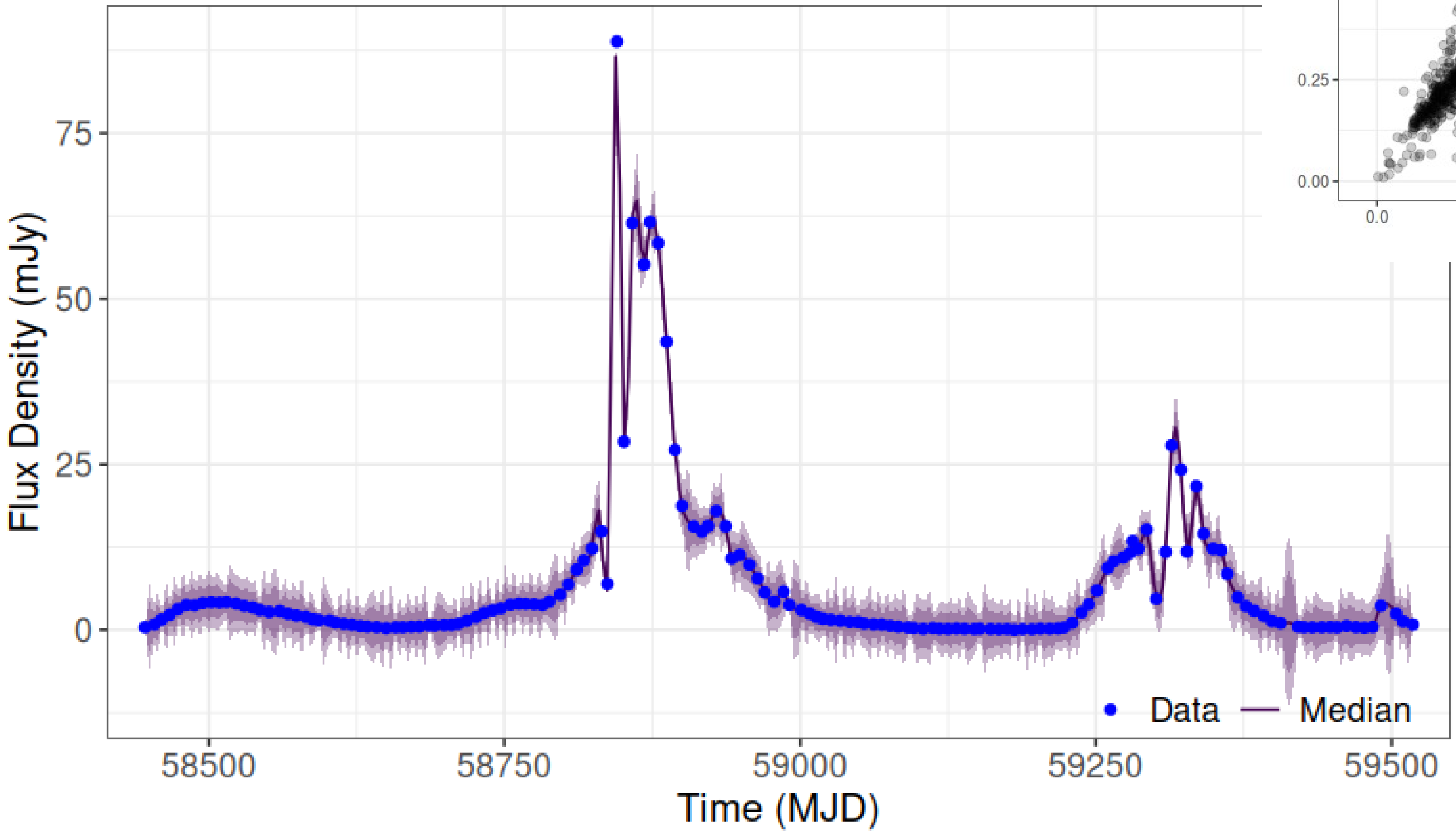




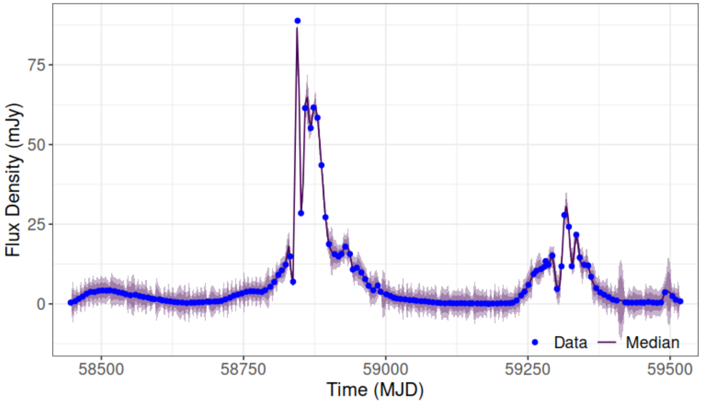
- Destroys any auto-correlation structures
- “null” distribution of joint posterior means marginalised over other hyperparameters

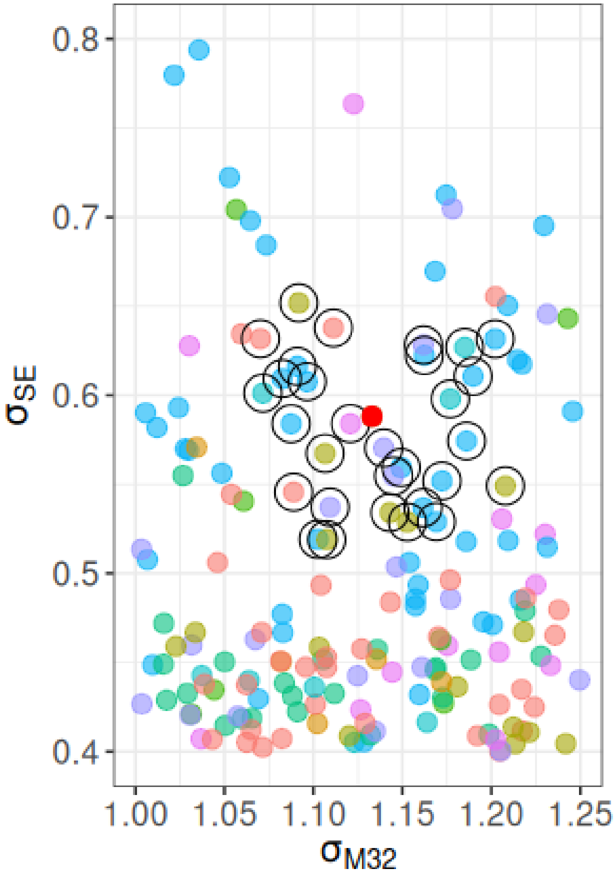
XRB: GX 339-4

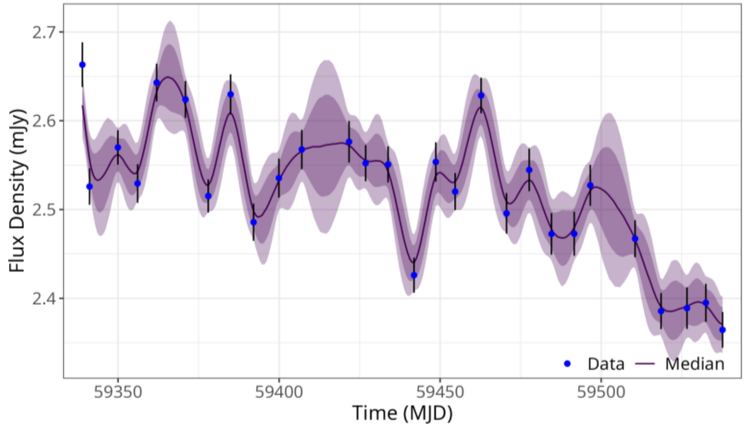


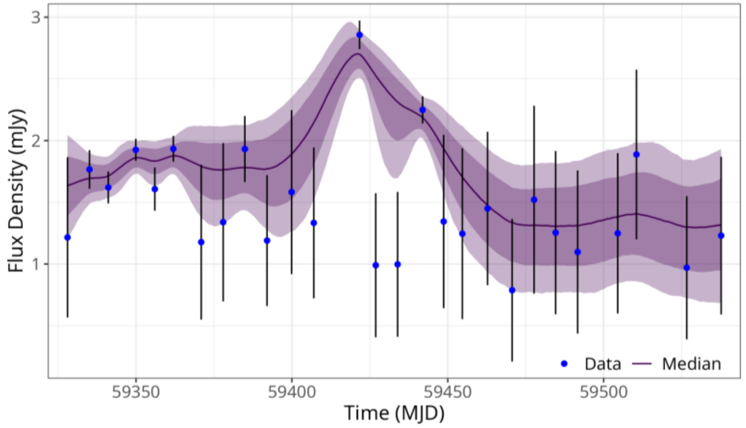


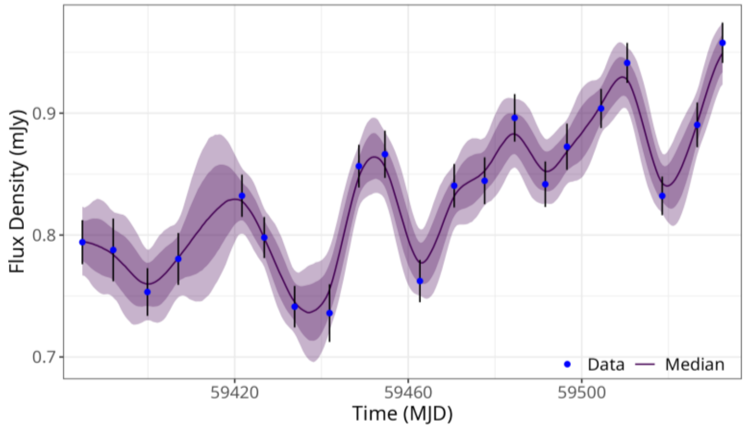
- Known X-ray binary
- Very bright peaks
- Transient source

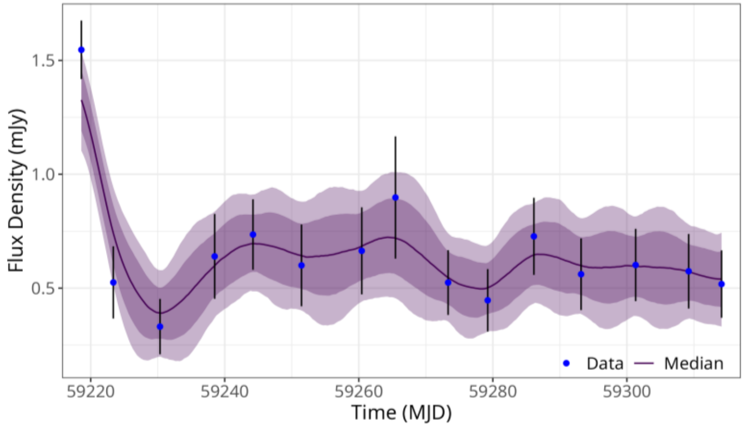

Flux Density (mJy)
75
50
25
0
58500
58750
59000
59250
59500
Time (MJD)
Data
Median

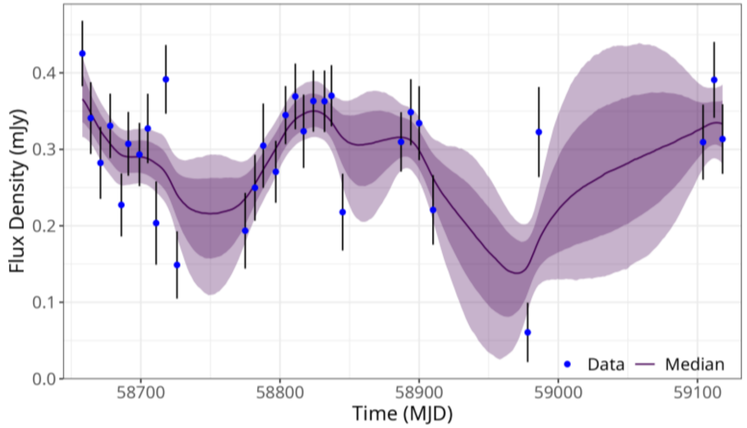

σ_{SE}
0.8
0.7
0.6
0.5
0.4
1.00
1.05
1.10
1.15
1.20
1.25
σ_{M32}

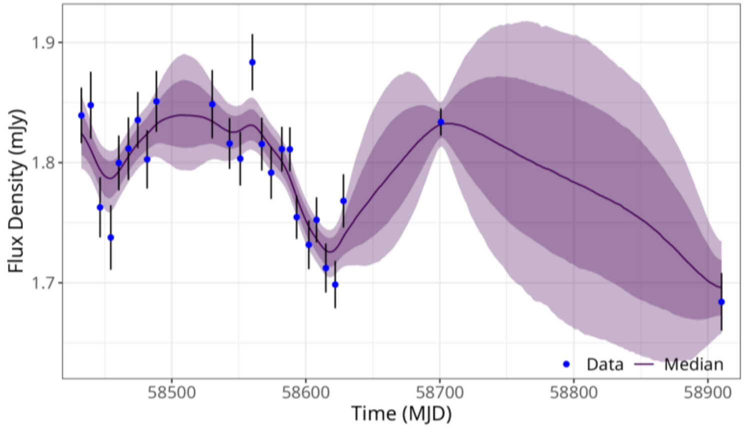

Flux Density (mJy)
2.7
2.6
2.5
2.4
59350
59400
59450
59500
Time (MJD)
Data
Median


Flux Density (mJy)
3
2
1
59350
59400
59450
59500
Time (MJD)
Data
Median

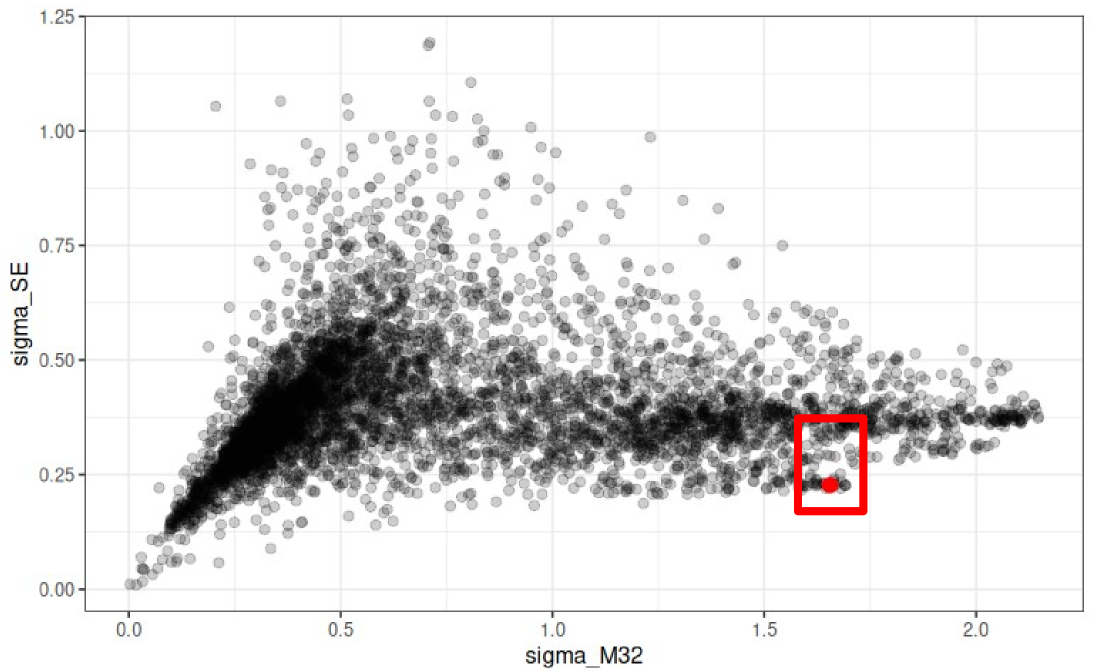

Flux Density (mJy)
0.9
0.8
0.7
59420
59460
59500
Time (MJD)
Data
Median

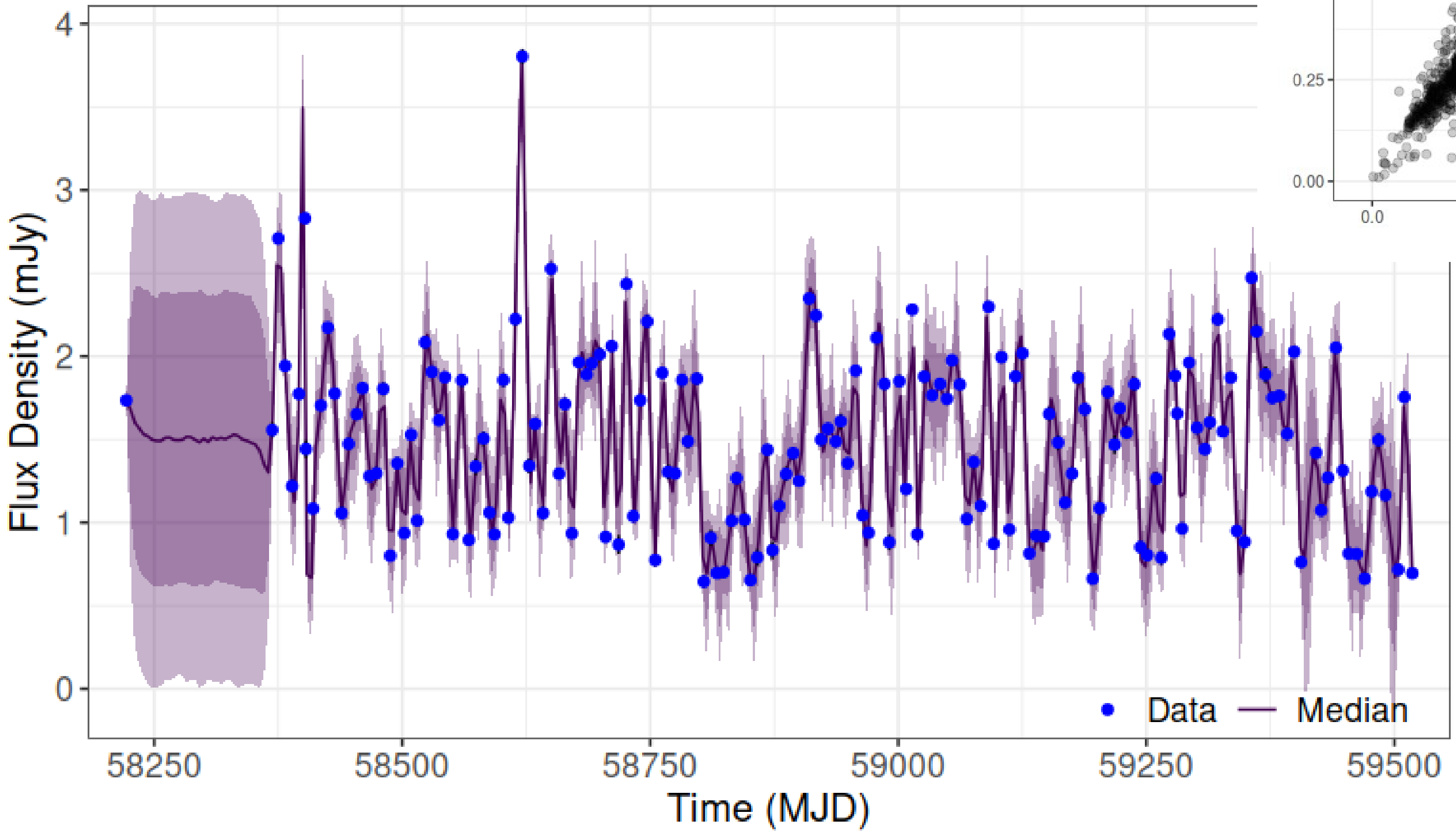

Flux Density (mJy)
1.5
1.0
0.5
59220
59240
59260
59280
59300
Time (MJD)
Data
Median


Flux Density (mJy)
0.4
0.3
0.2
0.1
0.0
58700
58800
58900
59000
59100
Time (MJD)
Data
Median

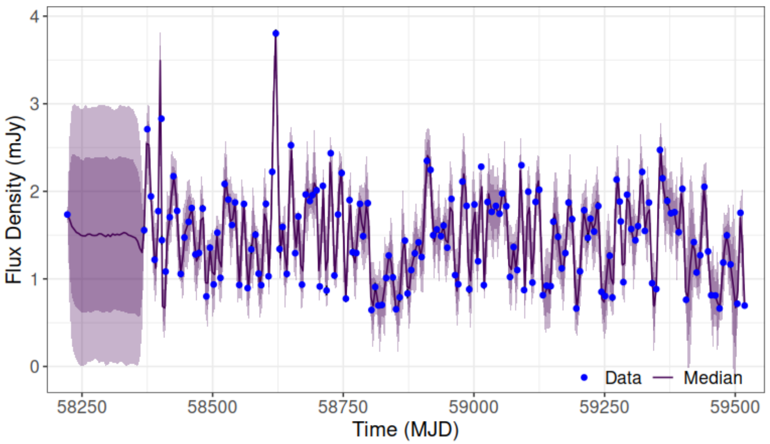

Flux Density (mJy)
1.9
1.8
1.7
58500
58600
58700
58800
58900
Time (MJD)
Data
Median

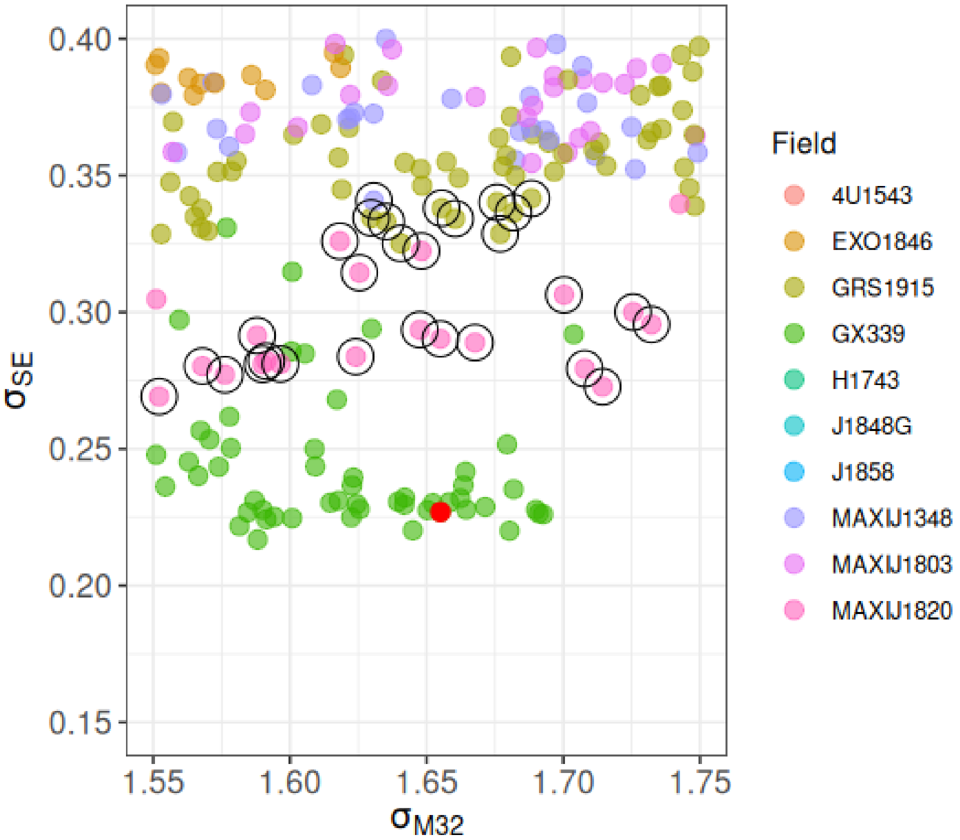
Pulsar: PSR J1703–4851

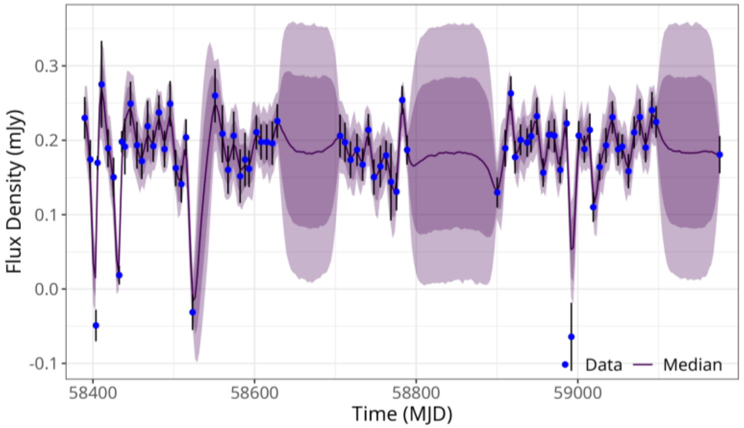


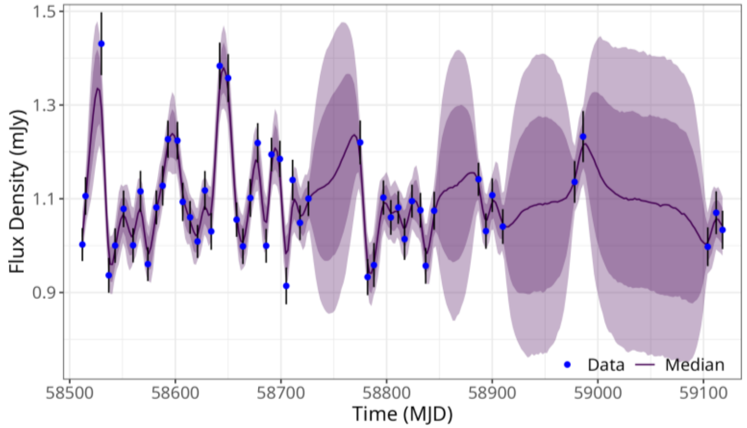


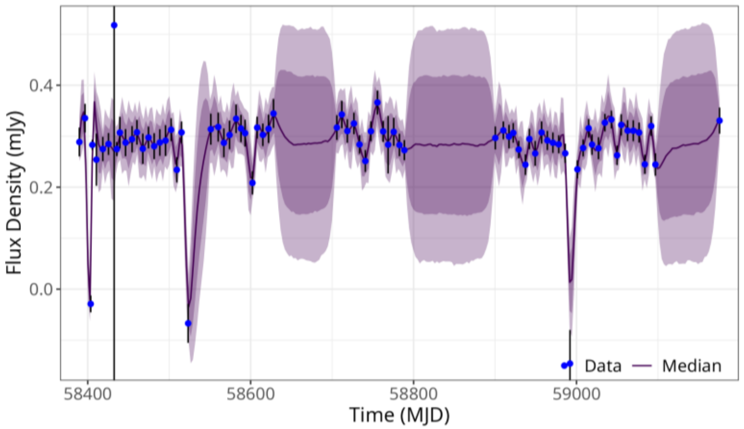
- Known pulsar
- Very high frequency variability
- Long term “variable”

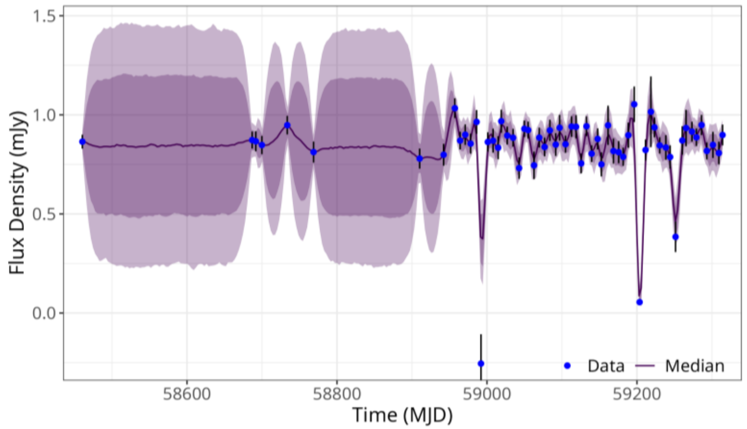

Flux Density (mJy)
Time (MJD)
Data
Median

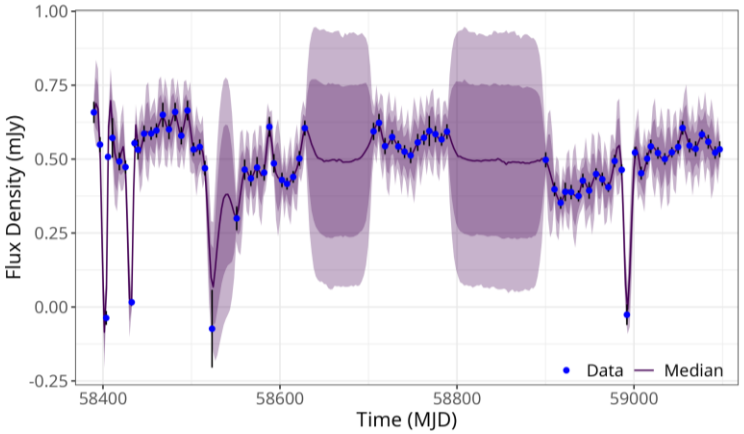

σ_{SE}
σ_{M32}
Field
4U1543
EXO1846
GRS1915
GX339
H1743
J1848G
J1858
MAXIJ1348
MAXIJ1803
MAXIJ1820

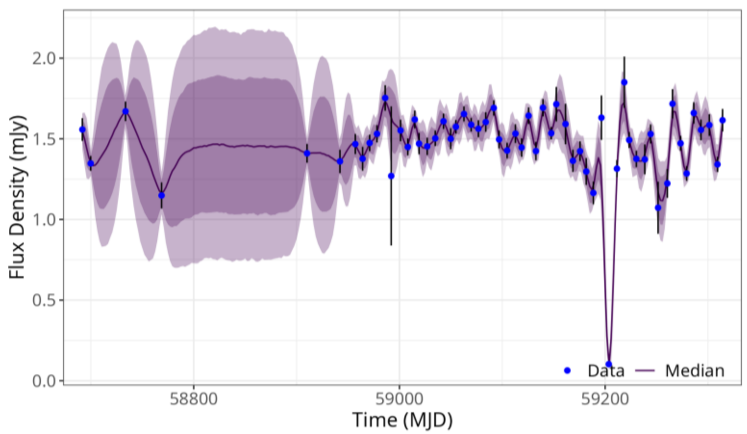

Flux Density (mJy)
Time (MJD)
Data
Median


Flux Density (mJy)
Time (MJD)
Data
Median


Flux Density (mJy)
Time (MJD)
Data
Median


Flux Density (mJy)
Time (MJD)
Data
Median


Flux Density (mJy)
Time (MJD)
Data
Median


Flux Density (mJy)
Time (MJD)
Data
Median

Summary so far

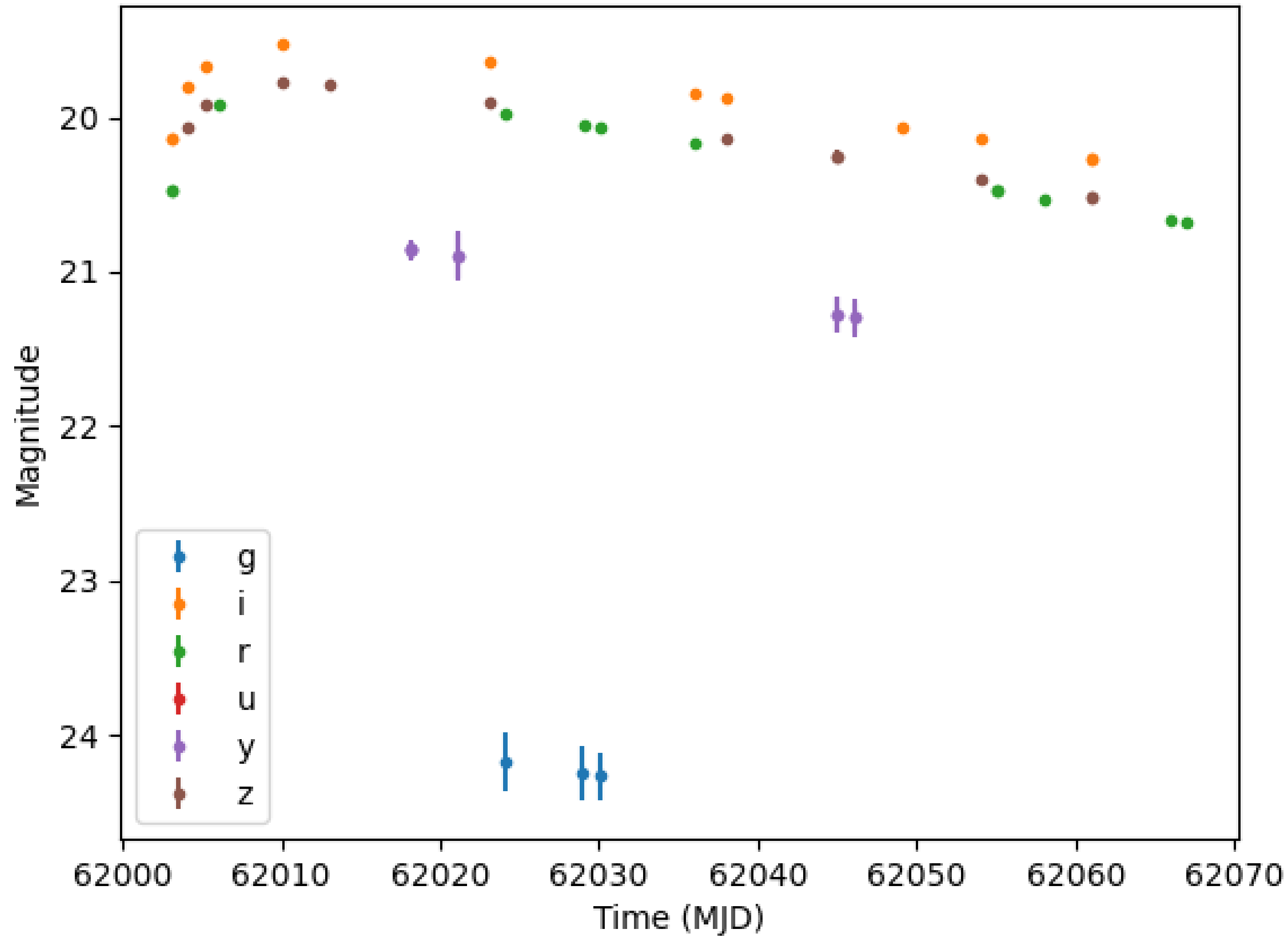
- Developed models and code suitable for fitting univariate GPs to the light curves of a large radio survey, i.e., ThunderKAT.
- GP amplitude hyperparameters are a better descriptor of variability than more commonly used statistics.
- GPs can be used to perform inference as well as interpolation in time-domain astronomy.

GPs: not only a means to an end but an end to only means.

Upcoming

	Astronomy	Statistics
Modelling radio light curves from ThunderKAT	Identifying transient and variable candidates in commensal radio surveys in the SKA era.	• Univariate Gaussian Processes • Gaussian likelihood • Sparse, unevenly sampled
Guidance for GPs in time-domain astronomy	Mean function, covariance kernel, and hyperprior choice for inference not just curve smoothing in astronomy.	
Modelling LSST light curves	Identifying transient candidates in multi-wavelength light curves across the optical band.	• Multi-output Gaussian Process regression • High noise and nuisance artefacts
Modelling light curves from large X-ray surveys (eROSITA, Swift)	Characterisation of black hole accretion through light curve modelling.	• Non-Gaussian likelihood • Non-Gaussian noise

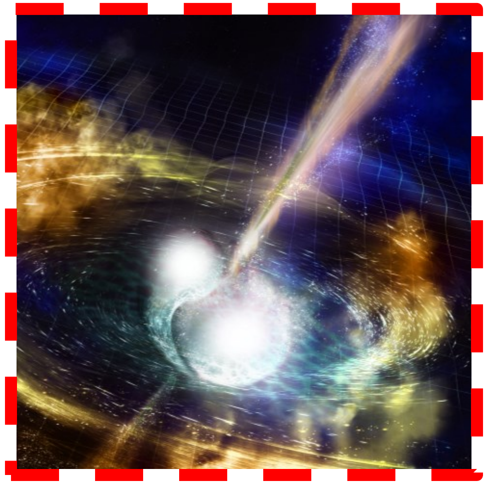
Multi-band Optical Light Curves



(Credit: S. Sett, 2024)

- LSST light curves may have measurements in multiple bands.
- Expect each band to be correlated.
- Sparsity and sampling will differ between bands.
- Multi-output GPs with different noise model.

Twinkle twinkle little star…



Exotic phenomena

Large-scale survey

Raw Data Processing → Identify → Classify

10^3 to 10^6 light curves

Transient candidates

Black holes, supernova, eclipsing binary, GRB, FRB, AGN, etc, …

… a Gaussian Process is what you are!